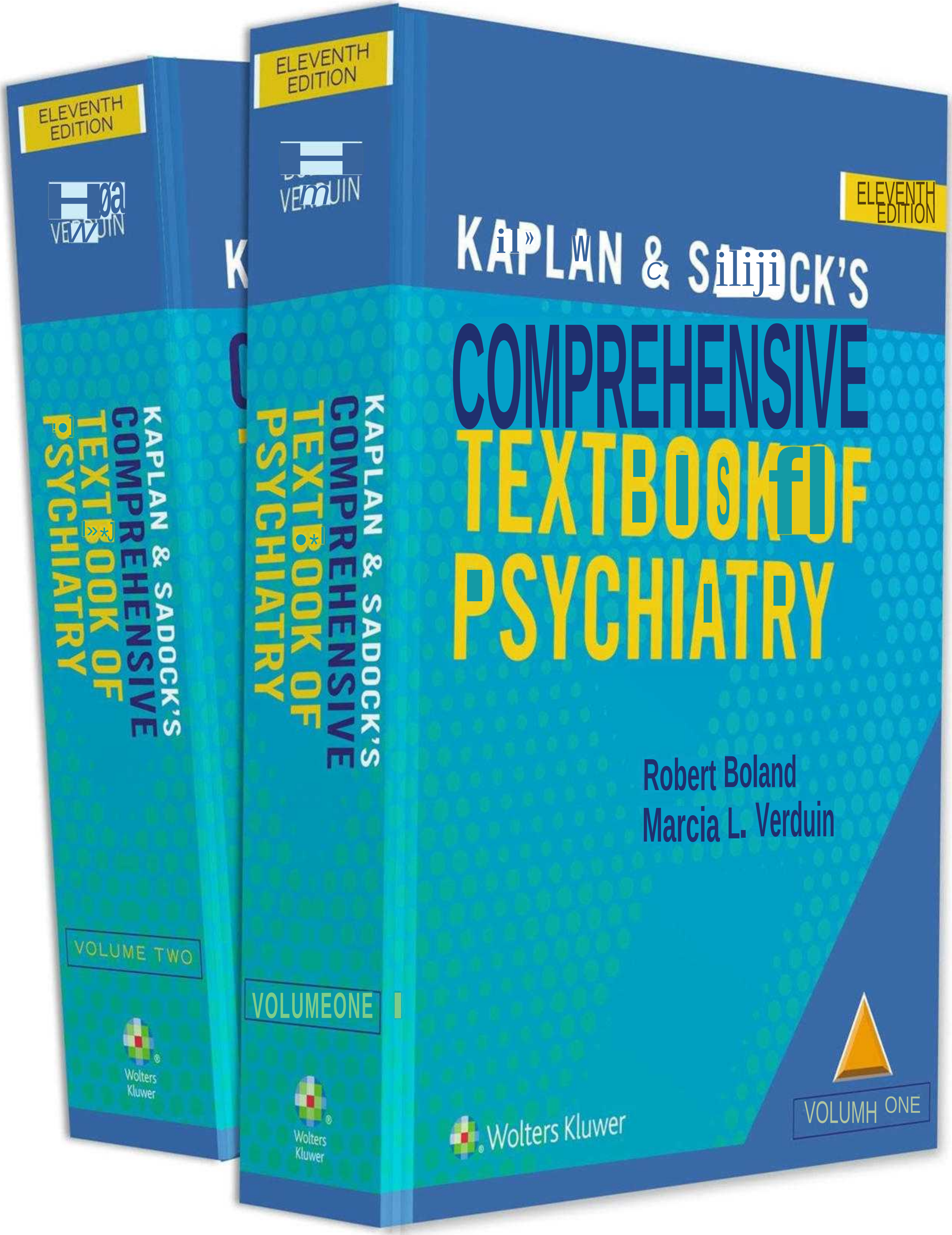

ELEVENTH EDITION

KAPLAN & SADOCK'S COMPREHENSIVE TEXTBOOK OF PSYCHIATRY

Robert Boland
Marcia L. Verduin

VOLUME ONE

Wolters Kluwer

ELEVENTH EDITION
BOLAND VERDUIN
KAPLAN & SADOCK'S COMPREHENSIVE TEXTBOOK OF PSYCHIATRY
VOLUME ONE
Wolters Kluwer

ELEVENTH EDITION
BOLAND VERDUIN
KAPLAN & SADOCK'S COMPREHENSIVE TEXTBOOK OF PSYCHIATRY
VOLUME TWO
Wolters Kluwer

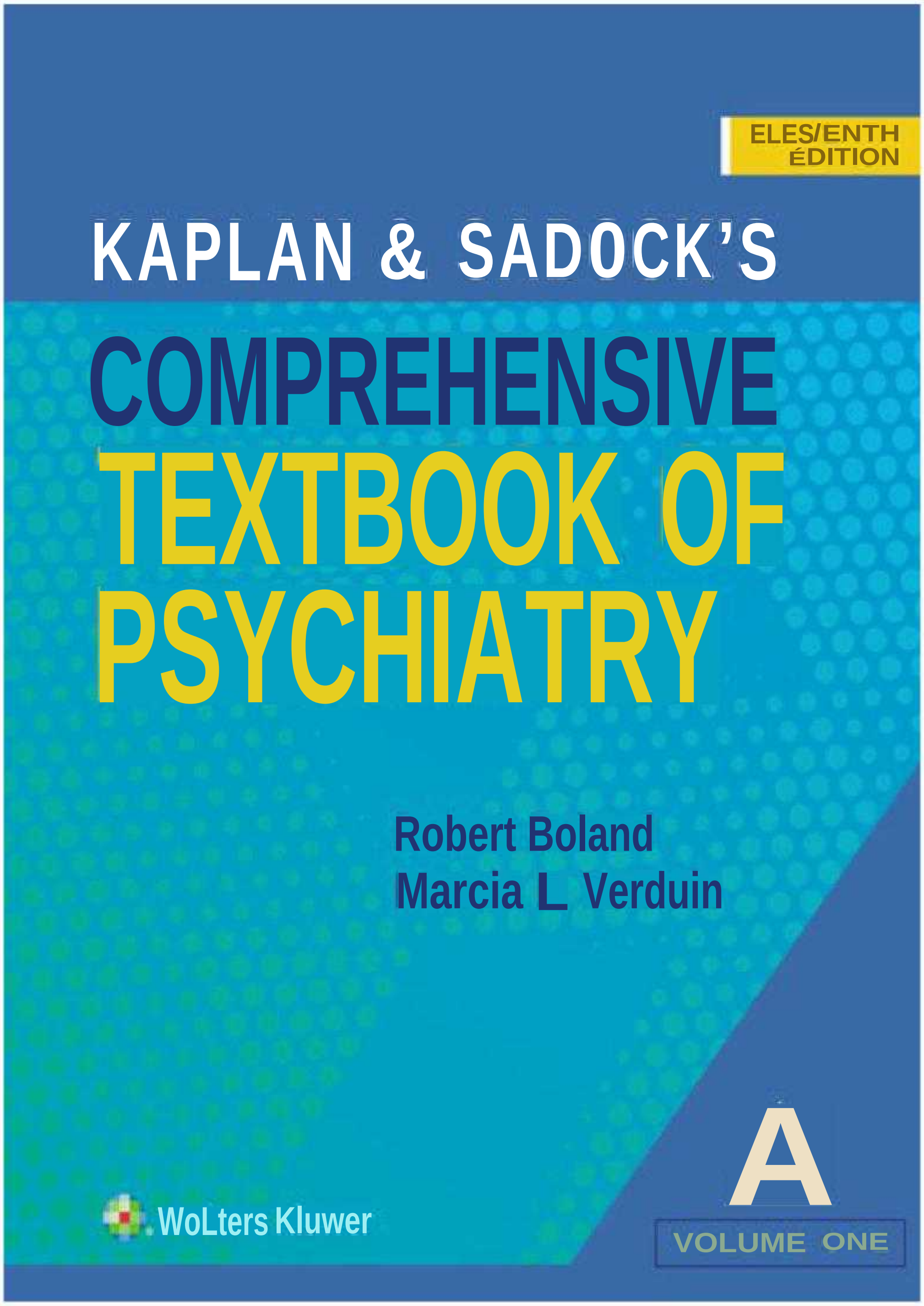

ELES/ENTH
ÉDITION

KAPLAN & SADOCK'S

COMPREHENSIVE TEXTBOOK OF PSYCHIATRY

Robert Boland
Marcia L Verduin

A

VOLUME ONE

WoLters Kluwer

Contributing Editors

Iqbal Ahmed, M.D., F.R.C.Psych.

Clinical Professor, Department of Psychiatry and Geriatric Medicine, University of Hawaii Honolulu, Hawaii; Clinical Professor, Department of Psychiatry, Uniformed Services University of the Health Sciences, F. Edward Hébert School of Medicine, Bethesda, Maryland

Gaston Baslet, M.D.

Chief, Division of Neuropsychiatry, Department of Psychiatry, Brigham and Women’s Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts

Carlyle H. Chan, M.D.

Professor and Vice Chair for Professional Development and Education Outreach, Department of Psychiatry and Behavioral Medicine, Medical College of Wisconsin, Milwaukee, Wisconsin

Dilip V. Jeste, M.D.

Distinguished Professor, Department of Psychiatry and Neurosciences, University of California, School of Medicine, San Diego, California

John H. Krystal, M.D.

Robert L. McNeil, Jr., Professor of Translational Research, Professor of Psychiatry, Neuroscience, and Psychology; Chair, Department of Psychiatry; Co-Director, Yale Center for Clinical Investigation, Yale University, School of Medicine; Chief of Psychiatry and Behavioral Health, Yale-New Haven Hospital, New Haven, Connecticut

Roberto Lewis-Fernández, M.D., M.T.S.

Professor of Clinical Psychiatry at Columbia University, Vagelos College of Physicians and Surgeons; Director of the New York State Center of Excellence for Cultural Competence; Director of the

Hispanic Treatment Program, and Research Area Leader for Anxiety, Mood, Eating and Related Disorders at New York State Psychiatric Institute, Psychiatry (Columbia) and Anxiety, Mood, Eating and Related Disorders (NYS Psychiatric Institute), Columbia University and NYS Psychiatric Institute, New York, New York

Kathleen Ries Merikangas, Ph.D.

Distinguished Investigator, Intramural Research Program, National Institute of Mental Health, Bethesda, Maryland

Christine Yu Moutier, M.D.

Chief Medical Officer, American Foundation for Suicide Prevention, New York, New York

Caroly S. Pataki, M.D.

Health Sciences Clinical Professor, Department of Psychiatry and Biobehavioral Sciences, University of California, David Geffen School of Medicine, Los Angeles, California

Joji Suzuki, M.D.

Director, Division of Addiction Psychiatry, Brigham and Women’s Hospital, Boston, Massachusetts

Carol A. Tamminga, M.D.

Professor and Chair, Department of Psychiatry, University of Texas Southwestern Medical School, Dallas, Texas

Sasson Zemach, M.D.

Doctor of Psychiatry, Eitanim Campus, The Jerusalem Mental Health Center, Affiliated With the Hebrew University and the Hadassah School of Medicine, Jerusalem, Israel

Joseph Zohar, M.D.

Professor, Psychiatrist, Post-Trauma Center, Sheba Medical Center Ramat Gan, Israel

KAPLAN & SADOCK'S
COMPREHENSIVE TEXTBOOK OF PSYCHIATRY

VOLUME I

ELEVENTH EDITION

EDITORS

Robert J. Boland, M.D.

Senior Vice President and Chief of Staff, The Menninger Clinic
Professor and Vice Chair, Menninger Department of Psychiatry and Behavioral Sciences
The Brown Foundation Endowed Chair in Psychiatry
Baylor University, College of Medicine
Houston, Texas

Marcia L. Verduin, M.D.

Associate Dean for Students, Professor of Psychiatry
Departments of Medical Education and Clinical Sciences
University of Central Florida College of Medicine
Orlando, Florida

Wolters Kluwer

Philadelphia • Baltimore • New York • London
Buenos Aires • Hong Kong • Sydney • Tokyo

Acquisitions Editor: Chris Teja
Development Editor: Ariel S. Winter
Editorial Coordinators: Sean Hanrahan and Erin Hernandez
Editorial Assistant: Jaida Lively
Marketing Manager: Kristen Watrud
Production Project Manager: Bridgett Dougherty
Manager, Graphic Arts & Design: Stephen Druding
Manufacturing Coordinator: Lisa Bowling
Prepress Vendor: Aptara, Inc.

11th edition

Copyright © 2025 Wolters Kluwer.

Copyright © 2017 Wolters Kluwer. Copyright © 2009, 2005 by Lippincott Williams & Wilkins. All rights reserved. This book is protected by copyright. No part of this book may be reproduced or transmitted in any form or by any means, including as photocopies or scanned-in or other electronic copies, or utilized by any information storage and retrieval system without written permission from the copyright owner, except for brief quotations embodied in critical articles and reviews. Materials appearing in this book prepared by individuals as part of their official duties as U.S. government employees are not covered by the above-mentioned copyright. To request permission, please contact Wolters Kluwer at Two Commerce Square, 2001 Market Street, Philadelphia, PA 19103, via email at permissions@lww.com, or via our website at shop.lww.com (products and services).

9 8 7 6 5 4 3 2 1

Printed in China

Library of Congress Cataloging-in-Publication Data available upon request.

978-1-9751-7573-3

This work is provided “as is,” and the publisher disclaims any and all warranties, express or implied, including any warranties as to accuracy, comprehensiveness, or currency of the content of this work.

This work is no substitute for individual patient assessment based upon healthcare professionals’ examination of each patient and consideration of, among other things, age, weight, gender, current or prior medical conditions, medication history,

laboratory data and other factors unique to the patient. The publisher does not provide medical advice or guidance and this work is merely a reference tool. Healthcare professionals, and not the publisher, are solely responsible for the use of this work including all medical judgments and for any resulting diagnosis and treatments.

Given continuous, rapid advances in medical science and health information, independent professional verification of medical diagnoses, indications, appropriate pharmaceutical selections and dosages, and treatment options should be made and healthcare professionals should consult a variety of sources. When prescribing medication, healthcare professionals are advised to consult the product information sheet (the manufacturer's package insert) accompanying each drug to verify, among other things, conditions of use, warnings and side effects and identify any changes in dosage schedule or contraindications, particularly if the medication to be administered is new, infrequently used or has a narrow therapeutic range. To the maximum extent permitted under applicable law, no responsibility is assumed by the publisher for any injury and/or damage to persons or property, as a matter of products liability, negligence law or otherwise, or from any reference to or use by any person of this work.

shop.lww.com

DEDICATION

To my family, Jonathan, Max, Sarah, Marcia, Brandy, Ken, and Ginger. Thank you for always encouraging and supporting me.—Marcy Verduin

To the people who most inspire me, my family: Susan, Rich and Greg, Jordan, Wyatt, Richard, Heather, Lisa, and Phillip.—Bob Boland

To all persons who work to improve the lives of persons with mental illness.—Drs. Verduin and Boland

About the Editors

ROBERT J. BOLAND, M.D.

Dr. Boland is the chief of staff and senior vice president at The Menninger Clinic in Houston, Texas. He is also a vice-chair of the Menninger Department of Psychiatry and Behavioral Sciences at Baylor College of Medicine (BCM) and the Brown Foundation Endowed Chair in Psychiatry at BCM.

Before coming to Houston, he served as vice-chair of education and director of the psychiatry residency program at Brigham and Women's Hospital in Boston, Massachusetts, and as an associate professor of psychiatry at Harvard Medical School. He is board-certified in psychiatry with expertise in medical education, consultation-liaison, and geriatric psychiatry. He is an alumnus of Georgetown University, where he earned his undergraduate and medical degrees.

Before joining Brigham and Women's Hospital, Dr. Boland had an 18-year tenure at the Alpert School of Medicine at Brown University, Providence, Rhode Island. For more than 30 years, he has been an active member of the American Psychiatric Association. Dr. Boland is the former president of the American Association of Directors of Psychiatric Residency Training, the Academy of Consultation-Liaison Psychiatry, and the Association for Academic Psychiatry. He is currently a Director of the American Board of Psychiatry and Neurology and a member of the Board of Regents for the American College of Psychiatrists.

Dr. Boland has authored or co-authored over 100 publications and is a frequent speaker at grand rounds and national meetings. In addition, he is the co-editor of the Kaplan and Sadock series of textbooks, which are the most popular texts in psychiatry. Dr. Boland is also the co-host of the Menninger Clinic's Mind Dive Podcast.

MARCIA L. VERDUIN, M.D.

Dr. Verduin is the Associate Dean for Students and a Professor of Psychiatry at the University of Central Florida College of Medicine in Orlando, Florida. She graduated from the University of Florida College of Medicine and completed her residency training and addiction psychiatry fellowship at the Medical University of South Carolina. She is a Fellow of the American College of Psychiatrists and a Distinguished Fellow of both the American Psychiatric Association and the Association for Academic Psychiatry. In 2007, Dr. Verduin became the first psychiatry faculty member at UCF's new medical school, where she played an integral role in developing the overarching undergraduate medical education curriculum and securing LCME accreditation.

In addition to her roles at the University of Central Florida College of Medicine, Dr. Verduin is a Past President for the Association for Academic Psychiatry, Past Chair of the Association of American Medical College's (AAMC) Southern Group on Student Affairs, prior Associate Editor of the Psychiatry Residency In-Training Examination, and has served as a member of the Board of Regents for the American College of Psychiatrists and an examiner for the American Board of Psychiatry and Neurology. She is currently the First Vice President of the American College of Psychiatrists and the Chair for the AAMC's Group on Student Affairs.

Dr. Verduin has a special interest in combatting the burden of stigma associated with mental health disorders, particularly in faith-based communities. She has been invited to speak around the world, including in Australia, New Zealand, Fiji, the Philippines, Southeast Asia, and Nigeria. Her contributions to advancing mental health care in those areas of the world have been covered in various U.S. and international media.

Contents

About the Editors
Contributors
Preface
Foreword

VOLUME I

1 Neural Sciences

- **1.1 The Neuroscience of Psychiatry**
  John H. Krystal, M.D.
- **1.2 Functional Neuroanatomy**
  Darlene S. Melchitzky, M.S., and David A. Lewis, M.D.
- **1.3 Functional Genomics of Human Brain Development**
  Kartik Pattabiraman, M.D., Ph.D., Sirisha Pochareddy, Ph.D., John C. Silbereis, Ph.D., John N. Constantino, M.D., and Nenad Sestan, M.D., Ph.D.
- **1.4 Neural Development and Neurogenesis**
  Emanuel DiCicco-Bloom, M.D., and Courtney R. McDermott, B.A.
- **1.5 Cellular and Synaptic Basis of Neural Signaling**
  Charles F. Zorumski, M.D., and Steven Mennerick, Ph.D.

1.6 Amino Acid Neurotransmitters
Joseph T. Coyle, M.D.
1.7 Biogenic Amine Neurotransmitters
John D. Elsworth, Ph.D.
1.8 Neuropeptides: Biology, Regulation and Role in Neuropsychiatric Disorders
Michael J. Owen, B.Sc. (Hons.), M.B., Ch.B., Ph.D., F.R.C.P.sych., Shaun P. Brothers, Ph.D., M.B.A., Larry J. Young, Ph.D., and Charles B. Nemeroff, M.D., Ph.D.
1.9 Novel Neurotransmitters
Thomas W. Sedlak, M.D., Ph.D., Bindu D. Paul, Ph.D., and Adam I. Kaplin, M.D., Ph.D.
1.10 Neurotrophic Factors
Jihye Kim, M.D., Francis S. Lee, M.D., Ph.D., and Moses V. Chao, Ph.D.
1.11 Intraneuronal Signaling
John A. Gray, M.D., Ph.D.
1.12 Psychoneuroendocrinology
Victor I. Reus, M.D.
1.13 Immune–Brain Interactions in Psychiatry
Ebrahim Haroon, M.D., Andrew H. Miller, M.D., and Charles L. Raison, M.D.
1.14 Chronobiology, Circadian Rhythm, and Psychiatry
Chester Wu, M.D., and Joshua J. Rodgers, M.D.
1.15 Basic Science of Sleep
Ruth M. Benca, M.D., Ph.D., Chiara Cirelli, M.D., Ph.D., and Giulio Tononi, M.D., Ph.D.
1.16 Pain Systems: Interface with Affective and Motivational Mechanisms
John Lotfi, M.D., J.D., and Jon-Kar Zubieta, M.D., Ph.D.
1.17 Basic Science of Appetite
Nori Geary, Ph.D., and Timothy H. Moran, Ph.D.

1.18 Genome, Transcriptome, and Proteome: The Molecular Genetics, Biochemistry, and Multi-Omics Underlying the Neurobiology of Mental Disorders

Evan E. Afshin, M.D., and Christopher E. Mason, Ph.D.

1.19 Gene Mapping Investigations of Psychiatric Disorders

Scott C. Fears, M.D., Ph.D.

1.20 Epigenetics in Psychiatry: The Promise for New Biomarkers and Treatments

Arie Kaffman, M.D., Ph.D.

1.21 Pharmacogenetics

Julio Licinio, M.D., Ph.D., M.B.A., M.S., and Ma-Li Wong, M.D., Ph.D.

1.22 Animal Models in Psychiatry

Janna K. Moen, B.A., Ph.D., and Paul J. Kenny, Ph.D.

1.23 Basic Systems Neuroscience

André A. Fenton, Ph.D., and Dino Dvorak, Ph.D.

1.24 Computational Modeling Approaches to Psychiatry

John D. Murray, Ph.D., and Alan Anticevic, Ph.D.

1.25 Nuclear Magnetic Resonance Imaging and Spectroscopy: Basic Principles and Recent Findings in Neuropsychiatric Disorders

Graeme F. Mason, Ph.D., John H. Krystal, M.D., and Gerard Sanacora, M.D., Ph.D.

1.26 Electroencephalography in Psychiatry

Jason K. Johannesen, Ph.D., and Patrick D. Skosnik, Ph.D.

1.27 Radiotracer Imaging with Positron Emission Tomography and Single Photon Emission Computed Tomography: Fundamental Principles, Methodology and Role in Neuropsychiatric Research

Mark Slifstein, Ph.D., John H. Krystal, M.D., and Anissa Abi-Dargham, M.D.

1.28 Functional Brain Connectivity and Psychopathology

Deanna M. Barch, Ph.D.
- 1.29 Neuropsychiatry of Self
  David R. Vago, Ph.D., Debra A. Gusnard, M.D., and David A. Silbersweig, M.D.
- 1.30 Learning Theory and Psychosis, Anxiety, and Addiction
  Philip R. Corlett, Ph.D.
- 1.31 Neuroscience of Substance Use Disorders
  Khaled Moussawi, M.D., Ph.D.
2 Neuropsychiatry and Behavioral Neurology
- 2.1 The Neuropsychiatric Approach to the Patient
  Fred Ovsiew, M.D.
- 2.2 Neuropsychiatric Aspects of Cerebrovascular Disorders
  Stanley Lyndon, M.D., and Juan Carlos Urizar, M.D.
- 2.3 The Neuropsychiatry of Brain Tumors
  Alasdair G. Rooney, MB.ChB., M.D., Ph.D.
- 2.4 Neuropsychiatric Aspects of Epilepsy and Psychogenic Nonepileptic Seizures
  Mario F. Mendez, M.D., Ph.D., Leila Parand, M.D., and Gaston Baslet, M.D.
- 2.5 Neuropsychiatric Consequences of Traumatic Brain Injury
  Melissa B. Jones, M.D., Amy R. Corcoran, M.D., Joanne A. Byars, M.D., and Ricardo E. Jorge, M.D.
- 2.6 Neuropsychiatric Aspects of Movement Disorders
  Laura Marsh, M.D., and Russell L. Margolis, M.D.
- 2.7 Neuropsychiatric Aspects of Multiple Sclerosis and Other Demyelinating Disorders
  Joshua J. Rodgers, M.D.
- 2.8 Neuropsychiatric Aspects of HIV Infection and AIDS
  Glenn Jordan Treisman, M.D., Ph.D., Andrew F. Angelino, M.D., Jeffrey Hsu, M.D., Anna Moukouri Kouoh, C.R.N.P.,

and Nicholas P. Schweizer, Ed.D., L.C.A.D.C., L.C.P.C.

2.9 Neuropsychiatric Aspects of Other Infectious Diseases (Non-HIV)

Jessica Harder, M.D., Rishab Gupta, M.B.B.S., M.D., Ruth McCann, M.D., and Irene Gonsalvez, M.D.

2.10 Neuropsychiatric Aspects of Prion Disease

Brian Giunta, M.D., Ph.D., Alireza Minagar, M.D., M.B.A., Paul Shapshak, Ph.D., Nadejda Alekseeva, M.D., and Francisco Fernandez, M.D.

2.11 Neuropsychiatric Aspects of Headache

Kathleen Ries Merikangas, Ph.D., Tarannum Lateef, M.D., M.P.H., and James R. Merikangas, M.D.

2.12 Neuropsychiatric Aspects of Neuromuscular Diseases

Timothy P. Wuerz, D.O., M.Sc., Rajeet Shrestha, M.D., and Alan J. Lerner, M.D.

2.13 Psychiatric Aspects of Child Neurology

Jeffrey E. Max, M.D., M.B.B.Ch., Sarah Mattson, Ph.D., Yvonne E. Vaucher, M.D., M.P.H.J., Sharon Nichols, Ph.D., and Mark P. Nespeca, M.D.

2.14 Neuropsychiatry of Neurometabolic and Neuroendocrine Disorders

Mark Walterfang, M.B.B.S. (Hons), Ph.D., F.R.A.N.Z.C.P., Ramon Mocellin, M.B.B.S., M.Sc., F.R.A.N.Z.C.P., Olivier Bonnot, M.D., Ph.D., and Dennis Velakoulis, M.B.B.S., M.P.M., Dip.Crim., D.Med.Sci., F.R.A.N.Z.C.P.

2.15 Catatonia

Joshua J. Rodgers, M.D., and Thanh Thuy Truong, M.D.

3 Contributions of the Psychological Sciences

3.1 Contributions of the Psychological Sciences: Sensation, Perception, and Cognition

Aldrich Chan, Psy.D., Louis J. Cozolino, Ph.D., and Daniel J. Siegel, M.D.

- 3.2 Piaget and Cognitive Development: History, Theory, Extensions, and Modifications
  Harrell Woodson, Ph.D.
- 3.3 Learning Theory
  Mark E. Bouton, Ph.D.
- 3.4 Biology of Memory
  Shruti Dave, Ph.D., and Joel L. Voss, Ph.D.
- 3.5 Normality and Mental Health
  George E. Vaillant, M.D.
- 3.6 Neuroscience of Psychoanalytic Object Relations Theory
  Dragan M. Svrakic, M.D., Ph.D., Charles F. Zorumski, M.D., and Otto F. Kernberg, M.D.

4 Contribution of the Social Sciences
- 4.1 Neurocentrism: Implications for Addiction and the Courtroom
  Sally L. Satel, M.D., and Stephen J. Morse, J.D., Ph.D.
- 4.2 Transcultural Psychiatry
  Kenneth Fung, M.D., F.R.C.P.C., M.Sc., and Robert Kohn, M.D.
- 4.3 Evolutionary Foundations for Psychiatric Research and Practice
  Randolph M. Nesse, M.D.

5 Quantitative and Experimental Methods in Psychiatry
- 5.1 Quantitative and Experimental Methods in Psychiatry: Epidemiology
  Diana E. Clarke, Ph.D., M.Sc., Lamyaa Yousif, M.D., Ph.D., M.Sc., Nusrat Rahman, Ph.D., Adrienne Grzenda, M.D., Ph.D., M.S., and William E. Narrow, M.D., M.P.H.
- 5.2 Statistics and Experimental Design
  Eugene Laska, Ph.D., Carole E. Siegel, Ph.D., and Isaac Galatzer-Levy, Ph.D.

6 Theories of Personality and Psychopathology

6.1 Psychoanalytic Theories and Treatment: Perspective from North America and Great Britain

Sylvia Gonzalez Cruz, M.D., Shweta Sharma, Psy.D., Cynthia Mulder, L.C.S.W.S., and Deeba Ashraf, M.D.

6.2 Erik H. Erikson

Dorian S. Newton, Ph.D., and Sudhir Kakar, Ph.D.

6.3 Other Psychodynamic Schools

Kristine L. Mccaslin, Psy.D., and Adam M. Brenner, M.D.

6.4 Approaches Derived from Philosophy and Psychology

Paul T. Costa, Jr., Ph.D., and Ralph L. Piedmont, Ph.D.

7 Diagnosis and Psychiatry: Examination of the Psychiatric Patient

7.1 Psychiatric Interview, History, and Mental Status Examination of the Adult Patient

Rory P. Houghtalen, M.D., and John S. McIntyre, M.D.

7.2 The Psychiatric Report and Outline for a Psychiatric Examination

Benjamin J. Sadock, M.D.

7.3 Practice Guidelines in Psychiatry

Laura J. Fochtmann, M.D., M.B.I., and John S. McIntyre, M.D.

7.4 Neuropsychological Assessment of Adults

Joseph R. Sadek, Ph.D., A.B.P.P.-C.N., Owen Gaasedelen, Ph.D., Rex M. Swanda, Ph.D., A.B.P.P.-C.N., and Kathleen Y. Haaland, Ph.D., A.B.P.P.-C.N.

7.5 Personality Assessment of Adults

Russell L. Adams, Ph.D., A.B.P.P-NP., C.L., and David L. Adams, Psy.D., B.C.B.A.-D.

7.6 Medical Assessment and Laboratory Testing in Psychiatry

Jennifer L. Kruse, M.D., and John O. Brooks, III, Ph.D., M.D.

7.7 Principles and Applications of Quantitative Electroencephalography in Psychiatry

Leslie S. Prichep, Ph.D.

7.8 Psychiatric Rating Scales

Samuel Justin Sinclair, Ph.D., Mark A. Blais, Psy.D., and Deborah Blacker, M.D., Sc.D.

7.9 Electronic Media in Psychiatry

Zebulon Taintor, M.D.

7.10 Cultural Competency in the Organization and Delivery of Psychiatric Care

Carole E. Siegel, Ph.D., Helena B. Hansen, M.D., Ph.D., and Jonathan Michel Metzl, M.D., Ph.D.

7.11 Medical Error

Nora C. Oberfield, M.D.

8 Clinical Manifestations of Psychiatric Disorders

Anu A. Matorin, M.D., Andres A. Avellaneda Ojeda, M.D., Asim A. Shah, M.D., and Pedro Ruiz, M.D.

9 Classification in Psychiatry

9.1 Present and Future of Classification Systems for Mental Disorders

Javier I. Escobar, M.D., M.Sc., and Humberto Marin, M.D.

9.2 The Classification of Mental Disorders in the International Classification of Diseases (ICD-11)

Javier I. Escobar, M.D., M.Sc., and Carlos López-Jaramillo, M.D., Ph.D.

10 Neurocognitive Disorders

10.1 Cognitive Disorders: Introduction

Iqbal Ahmed, M.D., F.R.C.Psych.

10.2 Delirium

Mandayam Sujata, M.D., Marie Anne Gebara, M.D., Tanya Fabian, Pharm.D., Ph.D., and LalithKumar Solai, M.D.

10.3 Major Neurocognitive Disorders (Dementia)

Julie Graziane, M.D., and Iqbal Ahmed, M.D., F.R.C.Psych.

10.4 Mild Cognitive Impairment and Amnestic Disorders
Daniel Varon, M.D., and Carmen Andreescu, M.D.
10.5 Cognitive Disorders Due to Another Medical Condition or Substance
Robert J. Boland, M.D., and Iqbal Ahmed, M.D., F.R.C.Psych.
11 Substance-Related Disorders
11.1 Substance Use Disorders: Introduction
Joji Suzuki, M.D.
11.2 Alcohol-Related Disorders
Marc A. Schuckit, M.D.
11.3 Stimulant-Related Disorders
Rocco A. Iannucci, M.D., and Roger D. Weiss, M.D.
11.4 Caffeine-Related Disorders
Laura M. Juliano, Ph.D., and Mary Margaret Sweeney, Ph.D.
11.5 Cannabis-Related Disorders
Janni Leung, Ph.D., Wayne D. Hall, Ph.D., and Louisa Degenhardt, Ph.D.
11.6 Hallucinogen-Related Disorders
Brian S. Barnett, M.D., and Joji Suzuki, M.D.
11.7 Inhalant-Related Disorders
Cameron J. Davidson, Ph.D., D.W. Svenson, M.A., and Scott E. Bowen, Ph.D.
11.8 Tobacco-Related Disorders
Smita Das, M.D., Ph.D., M.P.H., Brian Isakson, Ph.D., and Douglas M. Ziedonis, M.D., M.P.H.
11.9 Opioid Use and Related Disorders: From Neuroscience to Treatment
Sean X. Luo, M.D., Ph.D., and Adam Bisaga, M.D.
11.10 Sedative-, Hypnotic-, or Anxiolytic-Related Disorders
John J. Mariani, M.D.

11.11 Anabolic-Androgenic Steroid-Related Disorders
Harrison G. Pope, Jr., M.D., and Gen Kanayama, M.D., Ph.D.
12 Schizophrenia and Other Psychotic Disorders
12.1 Introduction and Overview
Carol A. Tamminga, M.D.
12.2 Phenomenology of Schizophrenia
Stephen F. Lewis, B.S., M.D., Rodrigo Escalona, M.D., and Samuel J. Keith, M.D.
12.3 Global Burden of Schizophrenia
Daniel Mamah, M.D., M.P.E.
12.4 Genetics of Schizophrenia
Kimberley M. Kendall, B.Sc. (Hons.), M.B.B.Ch., Ph.D., M.R.C.Psych., Elliott Rees, B.Sc. (Hons.), Ph.D., and Michael J. Owen, B.Sc. (Hons.), M.B., Ch.B., Ph.D., F.R.C.P.sych.
12.5 The Clinical Epidemiology of Psychosis Spectrum Disorder
Jim van Os, M.D., Ph.D., and Ulrich Reininghaus, Ph.D.
12.6 Cellular and Molecular Neuropathology of Schizophrenia
Ali S. Imami, M.B.B.S., M.S., Nicholas D. Henkel, Ph.D., Justin F. Creeden, Ph.D., and Robert E. McCullumsmith, M.D., Ph.D.
12.7 Structural Brain Imaging in Schizophrenia
Amanda E. Lyall, Ph.D., Marek Kubicki, M.D., Ph.D., and Martha E. Shenton, Ph.D.
12.8 Functional Brain Imaging in Schizophrenia
Raquel E. Gur, M.D., Ph.D., and Ruben C. Gur, Ph.D.
12.9 Molecular Brain Imaging in Schizophrenia
Dean F. Wong, M.D., Ph.D., Ganesh B. Chand, Ph.D., Gerhard Gründer, M.D., Jimmy E. Jakobsson, Ph.D., and Albert Gjedde, M.D., D.Sc.

12.10 Neurocognition in Schizophrenia
Philip D. Harvey, Ph.D.
12.11 Phenotypes of Psychosis
Brett Clementz, Ph.D.
12.12 Schizophrenia: Pharmacologic Treatment
John M. Kane, M.D., and Christoph U. Correll, M.D.
12.13 Psychiatric Rehabilitation
Robert E. Drake, M.D., Ph.D., Gary R. Bond, Ph.D., and Dror Ben-Zeev, Ph.D.
12.14 Medical Health in Schizophrenia
Robert J. Boland, M.D.
12.15 Recovery in Schizophrenia
Dina Hooshyar, M.D., M.P.H., Chacku Mathai, A.A.S., Drew Musa, C.P.S., and Mark Ragins, M.D.
12.16 Psychosis as a Defining Dimension in Schizophrenia
Elena I. Ivleva, M.D., Ph.D., and Carol A. Tamminga, M.D.
12.17 Other Psychotic Disorders
Ramin Mojtabai, M.D., Ph.D., M.P.H., Laura J. Fochtmann, M.D., M.B.I., and Evelyn J. Bromet, Ph.D.
13 Mood Disorders
13.1 Mood Disorders: Historical Introduction and Conceptual Overview
Robert J. Boland, M.D.
13.2 Mood Disorders: Epidemiology
Kathleen Ries Merikangas, Ph.D., and Zoltán Rihmer, M.D., Ph.D., D.Sc.
13.3 Mood Disorders: Clinical Features
Robert J. Boland, M.D.
13.4 Mood Disorders: Intrapsychic and Interpersonal Aspects
John C. Markowitz, M.D., and Barbara L. Milrod, M.D.
13.5 Mood Disorders: Suicidal Behavior

Maurizio Pompili, M.D., Ph.D., and Zoltán Rihmer, M.D., Ph.D., D.Sc.
13.6 Mood Disorders: Pharmacologic Treatment of Depression and Bipolar Disorders
Lakshmi N. Yatham, M.B.B.S., F.R.C.P.C.S., M.R.C.Psych(UK), M.B.A.(Exec), Sidney H. Kennedy, M.D., F.R.C.P.C., F.R.C.Psych., Ayal Schaffer, M.D., F.R.C.P.C., and Wegdan Rashad, M.D.
13.7 Mood Disorders: Psychotherapy
John R. McQuaid, Ph.D., and Michael E. Thase, M.D.
13.8 Mood Disorders: Neurobiology
Michael E. Thase, M.D., Vladimir Maletic, M.D., M.S., and Charles L. Raison, M.D.
14 Anxiety Disorders
14.1 Anxiety Disorders: Introduction and Overview
Kathleen Ries Merikangas, Ph.D.
14.2 Clinical Features of the Anxiety Disorders
Ryan J. Kimmel, M.D., Joseph M. Cerimele, M.D., M.P.H., and Peter Roy-Byrne, M.D.
14.3 Epidemiology of Anxiety Disorders
Kathleen Ries Merikangas, Ph.D., and Michelle M. Theodory, B.A.
14.4 Anxiety Disorders: Neurobiology and Neuroscience
Sara Costi, M.D., Adriana Feder, M.D., Brian Iacoviello, Ph.D., James W. Murrough, M.D., Ph.D., and Dennis S. Charney, M.D.
14.5 Neuroimaging and the Neuroanatomical Circuits Implicated in Anxiety Disorders
Elizabeth R. Kitt, M.S., M.Phil., Emily M. Cohodes, M.S., M.Phil., and Dylan G. Gee, Ph.D.
14.6 Anxiety Disorders: Cognitive-Behavioral Therapy
Jonathan D. Huppert, Ph.D., and Edna B. Foa, Ph.D.
14.7 Somatic Therapies for Anxiety Disorders

Lakshmi N. Ravindran, M.D., F.R.C.P.C., and Murray B. Stein, M.D., M.P.H.

15 Obsessive-Compulsive and Related Disorders

Dan J. Stein, M.D., Ph.D., and Christine Lochner, Ph.D.

16 Pathological and Problem Gambling: Gambling Disorder

Harvey Roy Greenberg, M.D.

17 Post-Traumatic Stress Disorder

Arieh Y. Shalev, M.D., and Charles R. Marmar, M.D.

18 Somatic Symptom and Related Disorders

Javier I. Escobar, M.D., M.Sc., and Joel E. Dimsdale, M.D.

19 Factitious Disorder

Davin K. Quinn, M.D., Dora Wang, M.D., M.A., Seth Powsner, M.D., and Stuart J. Eisendrath, M.D.

20 Dissociative Disorders

Richard J. Loewenstein, M.D., Roberto Lewis-Fernández, M.D., M.T.S., and Paul Frewen, Ph.D., C.Psych.

21 Normal Sexuality and Sexual Disorders

21.1 Normal Human Sexuality and Sexual Dysfunctions

Virginia A. Sadock, M.D.

21.2 Homosexuality, Gay and Lesbian Identities, and Homosexual Behavior

Jack Drescher, M.D., and William M. Byne, M.D., Ph.D.

21.3 Paraphilic Disorders

Reneé M. Sorrentino, M.D.

21.4 Gender Identity, Gender Diversity, Gender Dysphoria/Incongruence

Jack Drescher, M.D., and William M. Byne, M.D., Ph.D.

21.5 Sexual Addiction

Marcy Verduin, M.D.

22 Feeding and Eating Disorders

Christine C. Call, Ph.D., Evelyn Attia, M.D., and B. Timothy Walsh, M.D.

23 Sleep Disorders

Amir Sharafkhaneh, M.D., Ph.D., Habibolah Khazaie, M.D., and Max Hirshkowitz, Ph.D.

24 Impulse-Control Disorders

24.1 Intermittent Explosive Disorder

Taylor A. Ochalek, Ph.D., and F. Gerard Moeller, M.D.

24.2 Pyromania

Taylor A. Ochalek, Ph.D., and F. Gerard Moeller, M.D.

24.3 Kleptomania

Taylor A. Ochalek, Ph.D., and F. Gerard Moeller, M.D.

25 Adjustment Disorders

Jeffrey W. Katzman, M.D., and Cynthia Geppert, M.D., Ph.D., D.P.S., M.A., M.P.H., M.S.B., M.S.J., F.A.C.L.P., D.F.A.P.A., F.A.S.A.M., H.E.C-C.

26 Personality Disorders

C. Robert Cloninger, M.D., Ph.D., and Nigel Lester, M.D.

VOLUME II

27 Consultation-Liaison Psychiatry and Psychosomatic Medicine

27.1 History and Current Trends

Carol L. Alter, M.D., and Steven A. Epstein, M.D.

27.2 Cardiovascular Disorders

Judith Restrepo, M.D., Margo C. Funk, M.D., M.A., F.A.C.L.P., and Scott Beach, M.D.

27.3 Gastrointestinal Disorders

Ashwini Nadkarni, M.D.

27.4 Psychosomatic Medicine: Obesity

Megan E. Lavery, Psy.D., and Ninoska D. Peterson, Ph.D.

27.5 Respiratory Disorders

Thomas Ritz, Ph.D., Hannah O. Nordberg, M.A., Ph.D., Chelsey A. Werchan, Ph.D., and E. Sherwood Brown, M.D., Ph.D., M.B.A.

27.6 Diabetes: Psychosocial Issues and Psychiatric Disorders

Paul Ciechanowski, M.D., M.P.H., and Robert J. Boland, M.D.

27.7 Endocrine and Metabolic Disorders

Katherine E. Williams, M.D., Kathleen Watson, Ph.D., and Natalie Rasgon, M.D., Ph.D.

27.8 Psycho-Oncology

Daniel McFarland, D.O., M.S., Yesne Alici, M.D., and William S. Breitbart, M.D., F.A.C.L.P., D.F.A.P.A., F.A.P.O.S.

27.9 End-of-Life and Palliative Care

Xiomara Rocha-Cadman, M.D., and Mindy Greenstein, Ph.D.

27.10 Stress and Psychiatry

Traci J. Speed, M.D., Ph.D., Paul S. Nestadt, M.D., and Elizabeth J. Prince, D.O.

27.11 Psychiatric Morbidity Following Critical Illness

Christian Romanchek, M.D., Dimitry S. Davydow, M.D., M.P.H., F.A.C.L.P., and O. Joseph Bienvenu, M.D., Ph.D.

27.12 Psychocutaneous Disorder

Ravenna Khanna, M.D., and Jason P. Caplan, M.D.

27.13 Organ Transplantation

Andrea Frances DiMartini, M.D., Mary Schwab Seligman, M.D., M.A.R., Ph.D., Catherine Chang Crone, M.D., and Gerald Scott Winder, M.D., M.Sc.

27.14 Psychiatric Care of the Burned Patient

Martha C. Gamboa, M.D., Ejine Okoroafor, M.D., F.A.P.A., Reena Baharani, M.D., Yvette Smolin, M.D., M.Ed., and Michael Blumenfield, M.D.

27.15 Management of Chronic Pain

J. Gregory Hobelmann, M.D., M.P.H., and Michael R. Clark, M.D., M.P.H., M.B.A.

28 Additional Conditions That May Be a Focus of Clinical Attention

28.1 Malingering

Frank John Ninivaggi, M.D.

28.2 Adult Antisocial Behavior, Criminality, and Violence

Dorothy Otnow Lewis, M.D., F.A.C.P., Harrell Woodson, Ph.D., and Robert J. Boland, M.D.

28.3 Borderline Intellectual Functioning and Academic or Educational Problem

Frank John Ninivaggi, M.D.

28.4 Relational Problems

William Bernet, M.D.

29 Cultural Concepts of Distress and Assessment

29.1 Cultural Concepts of Distress

Roberto Lewis-Fernández, M.D., M.T.S., Laurence J. Kirmayer, M.D., and Brandon A. Kohrt, M.D., Ph.D.

29.2 Cultural Assessments in Psychiatry

Neil Krishan Aggarwal, M.D., M.B.A., M.A., and Roberto Lewis-Fernández, M.D., M.T.S.

30 Special Areas of Interest

30.1 Psychiatry and Reproductive Medicine

Helen Bradshaw, M.D., Marika Toscano, M.D., and Lauren M. Osborne, M.D.

30.2 Genetic Counseling for Psychiatric Conditions

Jehannine Austin, Ph.D., M.S., F.C.A.H.S., C.G.C., and Holly L. Peay, Ph.D., M.S.

30.3 Physical Violence on Adults

Shannon F. Stromberg, M.D., Kyle Owen, D.O., and Nancy Martin, M.D.

30.4 Complementary, Alternative, and Integrative Approaches in Mental Health Care

James H. Lake, M.D.

30.5 Micronutrient Monitoring and Supplementation in Psychiatric Illness

Jessica L. Principe, M.D., M.S., and Norman K. Pollock, Ph.D.

30.6 Disaster Psychiatry: Disasters, Terrorism, and War

David M. Benedek, M.D., Joshua C. Morganstein, M.D., James C. West, M.D., and Robert J. Ursano, M.D.

30.7 Psychiatry and Spirituality

Shweta Kapoor, M.D., Ph.D.

30.8 Physician and Medical Student Mental Health

Neeta Jain, M.D., and Cynthia M. Stonnington, M.D.

30.9 Computers in Psychiatry

John Luo, M.D.

31 Psychiatric Emergencies

31.1 Suicide: Overview of Epidemiology, Risk and Protective Factors, Treatment, and Prevention

Danuta Wasserman, M.D., Ph.D., and Vladimir Carli, M.D., Ph.D.

31.2 Suicide Treatment

Katrina A. Rufino, Ph.D.

31.3 Other Psychiatric Emergencies

Robert T. Cobb, M.D., and Katherine Camfield, M.D., M.P.H.

32 Psychotherapies

32.1 Psychotherapy Integration

Adam M. Brenner, M.D., and Amy S. Brenner, M.S.W., L.C.S.W.

32.2 Supportive Psychotherapy

John Battaglia, M.D.

32.3 Psychoanalysis and Psychoanalytic Psychotherapy

Adam Critchfield, M.D., Christina Kitt Garza, M.D., and Philip R. Muskin, M.D., M.A., D.L.F.A.P.A., L.F.A.C.L.P.

32.4 Psychoanalytic Treatment of Anxiety, Obsessive-Compulsive and Trauma-Related Disorders

Eric M. Plakun, M.D.

32.5 Intensive Short-Term Psychodynamic Psychotherapies

Manuel Trujillo, M.D.

32.6 Individual Psychodynamic Psychotherapy and Cognitive-Behavioral Therapy for Psychosis

Michael D. Garrett, M.D.

32.7 Treatment of Borderline Personality Disorder

Lois Choi-Kain, M.D., M.Ed., Evan Iliakis, B.A., M.D./Ph.D., Grace Murray, B.A., and Ana Rodriguez-Villa, M.D., M.B.A.

32.8 Behavior Therapy

Jenessa S. Price, Ph.D.

32.9 Cognitive Behavior Therapy

Judith S. Beck, Ph.D., and Robert Hindman, Ph.D.

32.10 Interpersonal Psychotherapy

Robert W. Guynn, M.D.

32.11 Dialectical Behavior Therapy

M. Zachary Rosenthal, Ph.D., Jacqueline Trumbull, B.A., Ph.D., and Rachel Guetta, B.A.

32.12 Solution-Focused Brief Therapy

Anne Bodmer Lutz, B.S.N., M.D.

32.13 Single-Session Therapy

Michael F. Hoyt, Ph.D.

32.14 Mentalization-Based Treatment

Anthony W. Bateman, M.A., F.R.C.Psych., and Peter Fonagy, O.B.E., F.Med.Sci., F.B.A., F.Ac.S.S.

32.15 Narrative Psychiatry

Bradley Lewis, M.D., Ph.D.

32.16 Positive Psychology

Nansook Park, Ph.D.

32.17 Acceptance and Commitment Therapy
Courtney Barry, Psy.D., M.S., Zabrina Ebert, D.O., and Joanna Petrides, M.B.S., Psy.D.
32.18 Hallucinogen-Assisted Psychotherapy
Erika Steinbrenner, M.D.
32.19 Hypnosis
Allan David Axelrad, M.D., Daniel P. Brown, Ph.D., and Harold J. Wain, Ph.D.
32.20 Family and Couples Therapy
Henry I. Spitz, M.D., D.F.A.P.A., D.F.A.G.P.A., and Susan Spitz, A.C.S.W., C.G.P.
32.21 Group Psychotherapy
Henry I. Spitz, M.D., D.F.A.P.A., D.F.A.G.P.A.
32.22 Psychodrama, Sociodrama, Sociometry, and Sociatry
Edward J. Schreiber, T.E.P., and Scott Giacomucci, D.S.W., L.C.S.W., B.C.D., C.G.P., F.A.A.E.T.S., T.E.P.
32.23 Other Methods of Psychotherapy
Adam M. Brenner, M.D., and Aaron Van Wright, M.D.
32.24 Evaluating Psychotherapy: Research Methods, Findings, and Future Directions
Lindsay E. Holly, Ph.D., Megan L. Petrik, Ph.D., and Stephen M. Saunders, Ph.D.
33 Biologic Therapies
  33.1 Preface
    33.1a General Principles of Psychopharmacology
    Olusola Ajilore, M.D., Ph.D., and Mark M. Rasenick, Ph.D.
    33.1b Drug Development and Approval Process in the United States
    Amir Kalali, M.D.
    33.1c Classification of Psychotropics: Neuroscience-Based Nomenclature (NbN)

Joseph Zohar, M.D., Sasson Zemach, M.D., and Daniel Minkin Levy, M.D.

33.2 Dopamine Blockers (“Antipsychotics”)

33.2a Dopamine Antagonists

Martin T. Strassnig, M.D., and Philip D. Harvey, Ph.D.

33.2b Dopamine-Serotonin Antagonists

Stephen R. Marder, M.D., and Walter Dunn, M.D., Ph.D.

33.2c Dopamine Blockers and Neuromotor Adverse Effects

Philip G. Janicak, M.D., and Kumail Hussain, M.D.

33.3 Dopamine Partial agonists

33.3a Dopamine Partial Agonists

Stephen R. Marder, M.D., and Walter Dunn, M.D., Ph.D.

33.4 Dopamine Enhancers

33.4a Dopamine Receptor Agonists

Carlos A. Zarate, Jr., M.D., and Lawrence T. Park, A.M., M.D.

33.4b Amphetamines, Methylphenidate, Atomoxetine, and Modafinil

Manpreet Kaur Singh, M.D., M.S.

33.5 Monoamine Enhancers (“Antidepressants”)

33.5a Reuptake Inhibitors

33.5a.1 Serotonin Reuptake Inhibitors (SRIs)

Simon Lafrenière, M.D., M.Sc., and Pierre Blier, M.D., Ph.D.

33.5a.2 Serotonin-Norepinephrine Reuptake Inhibitors (SNRIs)

Pierre Blier, M.D., Ph.D., and Michael E. Thase, M.D.

33.5a.3 Tricyclics and Tetracyclics

J. Craig Nelson, M.D.
33.5a.4 **Dopamine-Noradrenaline Reuptake Inhibitors (Bupropion)**
Charles DeBattista, D.M.H., M.D., and Alan F. Schatzberg, M.D.
33.5b **5HT$_{1A}$ Receptor Partial Agonists (Buspirone)**
Pierre Blier, M.D., Ph.D., and Cary Kogan, Ph.D., M.S.C.P.
33.5c **Vilazodone and Vortioxetine**
Pierre Blier, M.D., Ph.D., and Charles Desfossés, M.D.
33.5d **Monoamine Oxidase Inhibitors**
Ethan F. Karsen, M.D., and Paul E. Holtzheimer, M.D., M.S.
33.6 **Monoamine Modulators**
33.6a **Trazodone**
John T. Vernon, M.D., and Susan G. Kornstein, M.D.
33.6b **Nefazodone**
Amir A. Khan, M.D., and Susan G. Kornstein, M.D.
33.6c **Mirtazapine**
Mousa Sobhy Azer Botros, M.D., Radu V. Saveanu, M.D., and Michael E. Thase, M.D.
33.7 **Serotonin Antagonists/Inverse Agonists**
33.7a **Pimavanserin**
Joan M. Striebel, M.D., Jennifer A. O’Day, M.D., Nekisa Haghighat, M.D., M.P.H., and Stephen M. Stahl, M.D., Ph.D., D.Sc. (Hon.)
33.8 **Noradrenaline Modulators**
33.8a **α_2-Adrenergic Receptor Agonists: Clonidine, Clonidine Extended Release, Guanfacine, and Guanfacine Extended Release**
Eric Hollander, M.D., and Migena Dervishi, M.D.

  - 33.8b **β-Adrenergic Receptor Antagonists**
    Migena Dervishi, M.D., and Eric Hollander, M.D.
- 33.9 **GABA Enhancers**
  - 33.9a **Barbiturates and Similarly Acting Substances**
    Steven L. Dubovsky, M.D., and Dori Marshall, M.D.
  - 33.9b **Benzodiazepine Receptor Agonists and Antagonists**
    Steven L. Dubovsky, M.D., and Dori Marshall, M.D.
- 33.10 **Cation Function Modulators and Lithium (“Mood Stabilizers”)**
  - 33.10a **Calcium Channel Inhibitors**
    Steven L. Dubovsky, M.D., and Dori Marshall, M.D.
  - 33.10b **Gabapentin, Topiramate, Tiagabine, Levetiracetam, Zonisamide, and Pregabalin**
    Hiroyuki Uchida, M.D., Ph.D.
  - 33.10c **Carbamazepine**
    Robert M. Post, M.D., and Mark A. Frye, M.D.
  - 33.10d **Lamotrigine**
    Robert M. Post, M.D.
  - 33.10e **Valproate**
    Robert M. Post, M.D., and Mark A. Frye, M.D.
  - 33.10f **Lithium**
    Robert H. Belmaker, M.D.
- 33.11 **Acetylcholine Blockers and Enhancers**
  - 33.11a **Anticholinergics and Amantadine**
    Samoon Ahmad, M.D.
  - 33.11b **Cholinesterase Inhibitors and Memantine**
    Samir A. Sabbag, M.D., Zelde Espinel, M.D., M.P.H., Luminita Luca, M.D., Ph.D., and Elizabeth A. Crocco, M.D.
- 33.12 **Histamine Modulators**
  - 33.12a **Histamine Blockers and Enhancers**

Sasson Zemach, M.D., and Abraham Weizman, M.D.
33.13 Pharmacological Treatments of Substance Abuse
  33.13a Opioid Receptor Modulators
    33.13a.1 Opioid Maintenance Treatments: Methadone and Buprenorphine
      Jonathan R. Buchholz, M.D., Kelly Barth, D.O., and Andrew J. Saxon, M.D.
    33.13a.2 Opioid Receptor Antagonists: Naltrexone and Nalmefene
      Oluwole O. Jegede, M.D., M.P.H., Stephanie S. O'Malley, Ph.D., Tiffany Arnold, B.S., M.D., Katherine S. Pier, M.D., and Srinivas B. Muvvala, M.D., M.P.H.
  33.13b Disulfiram and Acamprosate
    Katherine S. Pier, M.D.
33.14 Fast-Acting Pharmacologic Treatments
  33.14a Esketamine and Ketamine
    Carlos A. Zarate, Jr., M.D., and Lawrence T. Park, A.M., M.D.
  33.14b MDMA, Psilocybin, Ayahuasca, and Ibogaine
    David Nutt, F.R.C.P., F.R.C.Psych., F.B.Ph.S., F.Med.Sci., D.Laws., and David Erritzoe, M.D., Ph.D., M.R.C.Psych.
33.15 Hormone Receptor Enhancers
  33.15a Melatonin Receptor Agonists
    Tawny L. Smith, Pharm.D., B.c.p.p., and d. jeffrey Newport, m.d., m.s., M.Div.
  33.15b Thyroid Hormones
    Rena Cooper-Kazaz, M.D., and Bernard Lerer, M.D.
33.16 Combination Pharmacotherapy

Charles DeBattista, D.M.H., M.D., and Alan F. Schatzberg, M.D.

33.17 Sex and Psychopharmacology

- **33.17a Sex Differences in Psychopharmacology**
  Anna E. Bauer, Ph.D., M.P.H., and Samantha Meltzer-Brody, M.D., M.P.H.
- **33.17b Reproductive Hormonal Therapy: Theory and Practice**
  David R. Rubinow, M.D., and Peter J. Schmidt, M.D.
- **33.17c Pharmacologic Treatment of Sexual Dysfunction**
  Caridad Ponce Martinez, M.D., Saria El Haddad, M.D., and Anita H. Clayton, M.D.

33.18 Psychopharmacology Across the Life Cycle

- **33.18a Pediatric Psychopharmacology**
  Jeffrey R. Strawn, M.D., and Laura B. Ramsey, Ph.D.
- **33.18b Biologic Therapies: Geriatric Psychopharmacology**
  Ari B. Cuperfain, M.D., M.Sc., and Bruce G. Pollock, M.D., Ph.D., F.R.C.P.C.

33.19 Biologic Therapies Around the Globe

- **33.19a Essential List of Medicines**
  Sasson Zemach, M.D., Daniel Minkin Levy, M.D., and Joseph Zohar, M.D.
- **33.19b Biologic Therapies in Africa**
  Oye Gureje, Ph.D., D.Sc., F.R.C.Psych., F.W.A.C.P., and Olatunde Ayinde, M.B.B.S., M.S., F.W.A.C.P.
- **33.19c Biologic Therapies in India**
  E. Mohandas, M.B.B.S., M.D., Sandeep Grover, M.B.B.S., M.D., and Ajit Avasthi, M.B.B.S., M.D.

        33.19d **Biologic Therapies in Latin America**
            Rodolfo Furlan Damiano, M.D., Ph.D., and André Russowsky Brunoni, M.D., Ph.D.
    33.20 **Brain Stimulation**
        33.20a **Electroconvulsive Therapy**
            Joan Prudic, M.D.
        33.20b **Brain Stimulation Methods**
            Stefan B. Rowny, M.D., M.F.A., and Lawrence Amsel, M.D., M.P.H.
        33.20c **Neurosurgical Treatments**
            Benjamin D. Greenberg, M.D., Ph.D., and Nicole C.R. McLaughlin, Ph.D.
34 **Child Psychiatry**
    34.1 **Introduction and Overview**
        Caroly Sue Pataki, M.D.
    34.2 **Child Development**
        Gabrielle Shapiro, M.D., D.F.A.A.C.A.P., D.F.A.P.A., Robert J. Jaffe, M.D., Alexander Kolevzon, M.D., and Tresha A. Gibbs, M.D.
    34.3 **Adolescent Development**
        Ardis C. Martin, M.D., and Caroly Sue Pataki, M.D.
35 **Child Psychiatric Examination**
    35.1 **Psychiatric Examination of the Infant, Child, and Adolescent**
        Robert A. King, M.D., Prudence W. Fisher, Ph.D., Michael H. Bloch, M.D., M.S., Mary Schwab Seligman, M.D., M.A.R., Ph.D., Walter S. Gilliam, Ph.D., and Michael Vitulano, Ph.D.
    35.2 **Psychiatric Assessment of Preschool Children**
        Helen Link Egger, M.D., and Caroly Sue Pataki, M.D.
    35.3 **Neuropsychological and Psychoeducational Assessment of Children**

Evangelina Barnard, Psy.D., and Patricia D. Walshaw, Ph.D.
35.4 Psychiatric Emergencies in Children and Adolescents
Alexandra Junewicz, M.D., and Ruth S. Gerson, M.D.
36 Genetics in Child Psychiatry
Gil D. Hoftman, M.D., Ph.D., James T. McCracken, M.D., and Erika L. Nurmi, M.D., Ph.D.
37 Neuroimaging in Psychiatric Disorders of Childhood
Frank P. Macmaster, Ph.D., Kayla D. Stone, Ph.D., Sneha Chenji, Ph.D., and David R. Rosenberg, M.D.
38 Temperament: Risk and Protective Factors for Child Psychiatric Disorders
David C. Rettew, M.D.
39 Intellectual Disability
Evan Taniguchi, M.D., Soo-Jeong Kim, M.D., and Karen E. Toth, Ph.D.
40 Specific Learning Disorder
Oren Boxer, Ph.D.
41 Developmental Coordination Disorder
Vijay Vishwanath, M.D., Ph.D., and Wendy G. Mitchell, M.D.
42 Communication
42.1 Language Disorder
Tuana Kant, M.Sc., Emiko Koyama, Ph.D., Mahliqa Ashraf, H.B.Sc., and James L. Kennedy, M.D., F.R.C.P.C., F.R.S.C.
42.2 Social (Pragmatic) Communication Disorder
Mirelle Kass, B.A., Young Shin Kim, M.D., M.S., M.P.H., Ph.D., and Bennett L. Leventhal, M.D.
42.3 Speech Sound Disorder
Emiko Koyama, Ph.D., Mahliqa Ashraf, H.B.Sc., Tuana Kant, M.Sc., and James L. Kennedy, M.D., F.R.C.P.C., F.R.S.C.
42.4 Stuttering

Emiko Koyama, Ph.D., Mahliqa Ashraf, H.B.Sc., James L. Kennedy, M.D., F.R.C.P.C., F.R.S.C., and Tuana Kant, M.Sc.

42.5 Unspecified Communication Disorder

Tim Bressmann, Ph.D.

43 Autism Spectrum Disorders

Fred R. Volkmar, M.D., Giacomo Vivanti, Ph.D., Ryan K. C. Yuen, Ph.D., and Kevin Pelphrey, Ph.D.

44 Attention-Deficit Disorders

44.1 Attention-Deficit/Hyperactivity Disorder

Lily T. Hechtman, M.D., F.R.C.P., A.B.P.N., Margaret D. Weiss, M.D., Ph.D., and Philippe Hwang, M.B., B.Ch., B.A.O., M.Sc.

44.2 Adult Manifestations of Attention-Deficit/Hyperactivity Disorder

James J. McGough, M.D., M.S.

45 Disruptive Behavior Disorders in Children and Adolescents

Erin P. Vaughan, M.A., Eva R. Kimonis, Ph.D., M.S., and Paul J. Frick, Ph.D.

46 Feeding and Eating Disorders of Infancy and Early Childhood

Irene Chatoor, M.D., and Loredana Lucarelli, Psy.D.

47 Tic Disorders

Rahil R. Jummani, M.D., Elizabeth A. Deckler, M.D., and Barbara J. Coffey, M.D., M.S.

48 Elimination Disorders

Edwin Jens Mikkelsen, M.D.

49 Other Disorders of Infancy, Childhood, and Adolescence

49.1 Reactive Attachment Disorder and Disinhibited Social Engagement Disorder

Karin M. Best, Ph.D., Mary Margaret Gleason, M.D., and Charles H. Zeanah, M.D.

49.2 Stereotypic Movement Disorders in Children

Robert Lloyd Doyle, M.A., D.D.S., M.D.

50 Mood Disorders in Children and Adolescents

- 50.1 Depressive Disorders and Suicide
  Karen Dineen Wagner, M.D., Ph.D., Tami Benton, M.D., and Catharyn A. Turner, II, M.D., M.Ed.
- 50.2 Early-Onset Bipolar Disorder
  Gabrielle A. Carlson, M.D., Manpreet Kaur Singh, M.D., M.S., Danella Hafeman, M.D., Ph.D., and Caroly Sue Pataki, M.D.

51 Anxiety Disorders in Children

- 51.1 Obsessive-Compulsive Disorder in Childhood
  Adam B. Lewin, Ph.D., A.B.P.P., and John Piacentini, Ph.D., A.B.P.P.
- 51.2 Posttraumatic Stress Disorder in Children and Adolescents
  Judith A. Cohen, M.D.
- 51.3 Separation Anxiety, Generalized Anxiety, and Social Anxiety
  Benjamin N. Schneider, M.D., Ruben G. Martinez, Ph.D., and Tara S. Peris, Ph.D.
- 51.4 Selective Mutism
  R. Lindsey Bergman, Ph.D., and Araceli Gonzalez, Ph.D.

52 Early-Onset Psychotic Disorders

Kaitlin Boyle, M.D., Linmarie Sikich, M.D., M.A., and Tara Chandrasekhar, M.D.

53 Child Psychiatry: Psychiatric Treatment

- 53.1 Psychodynamic Psychotherapy (Child Psychiatry)
  David L. Kaye, M.D., and Michael Shapiro, M.D.
- 53.2 Cognitive-Behavioral Psychotherapy for Children and Adolescents
  Anne Marie Albano, Ph.D., and Sandra S. Pimentel, Ph.D.
- 53.3 Group Psychotherapy
  Margo L. Thienemann, M.D.

53.4 Family Therapy
John Sargent, M.D., and Neha Sharma, D.O.
53.5 Pediatric Psychopharmacology
Jeffrey R. Strawn, M.D., and Laura B. Ramsey, Ph.D.
53.6 Inpatient Psychiatric, Partial Hospitalization, and Residential Treatment for Children and Adolescents
Cynthia Wilson, M.D., and Andrés Martin, M.D., Ph.D.
53.7 Community-Based Treatment
Andrés J. Pumariega, M.D.
53.8 The Treatment of Adolescents
Khadijah Booth Watkins, M.D., M.P.H., Steven C. Schlozman, M.D., and Eugene V. Beresin, M.D., M.A.
54 Child Psychiatry: Special Areas of Interest
54.1 Adoption and Foster Care
Dale Aaron Peeples, M.D., and Sandra B. Sexson, M.D.
54.2 Child Maltreatment
Heather Williams, M.D., and Bradley Freeman, M.D.
54.3 Children's Reaction to Illness and Hospitalization
Susan Beckwitt Turkel, M.D., Julienne Jacobson, M.D., and Maryland Pao, M.D.
54.4 Psychiatric Sequel of HIV and AIDS
Mark DeAntonio, M.D.
54.5 Adolescent Substance Use Disorders
Sean R. LeNoue, M.D., and Paula D. Riggs, B.A., M.A., M.D.
54.6 Forensic Child and Adolescent Psychiatry
Peter Ash, M.D., and Elissa P. Benedek, M.D.
54.7 Ethical Issues in Child and Adolescent Psychiatry
Adrian N. Sondheimer, M.D., and Takahiro Soda, M.D., Ph.D.
54.8 School Consultation and Collaborating With the Educational System

Sheryl Kataoka, M.D., M.S.H.S., Ilaina Blum, M.A., M.F.T., Edward Castelo, M.D., and Roya Ijadi-Maghsoodi, M.D., M.S. H.P.M.

54.9 Prevention of Psychiatric Disorders in Children and Adolescents

William R. Beardslee, M.D.

54.10 Child Mental Health Services Research

Bonnie T. Zima, M.D., M.P.H., and Regina Bussing, M.D., M.S.H.S.

54.11 Impact of Terrorism and School Shootings on Children

Nayla M. Khoury, M.D., M.P.H., and Wanda P. Fremont, M.D.

54.12 Impact on Parents of Raising a Child with Psychiatric Illness and/or Developmental Disability

Alice R. Mao, M.D., Matthew N. Brams, M.D., Kathryn Forneris, D.O., M.P.H., and Pieter Joost Van Wattum, M.D., M.A.

54.13 Pediatric Sleep Disorders

Jess P. Shatkin, M.D., M.P.H., and Anna Ivanenko, M.D., Ph.D.

54.14 Internet Gaming Disorder in Youth

Michael Feldmeier, M.D., Michael Ogata, M.D., Caroly Sue Pataki, M.D., and Roxy Szeftel, M.D.

54.15 Social Media and Youth

Christina Guest, M.D., Yeun J. Lim, M.D., Michael Feldmeier, M.D., Roxy Szeftel, M.D., and Caroly Sue Pataki, M.D.

54.16 Gay, Bisexual, Transgender, Queer/Questioning, Intersex, and Asexual (LGBTQIA) Youth

Natalia Ramos, M.D., M.P.H., Serena Chang, M.D., B.A., and Mikiko Thelwell, M.D., M.P.H.

54.17 **Racism and Discrimination’s Impact on the Health and Well-Being of Developing Children and Adolescents**
Wanjikũ F.M. Njoroge, M.D., Ewurama Sackey, M.D., Nana Asabere, M.D., and Eraka Bath, M.D.
54.18 **Forensic Evaluations of Child and Adolescent Migrants**
Olivia Shadid, M.D., and Shawn Singh Sidhu, M.D.
55 **Geriatric Psychiatry**
55.1 **Introduction to Geriatric Psychiatry**
Dilip V. Jeste, M.D., Dylan J. Jester, Ph.D., M.P.H., and Ellen E. Lee, M.D.
55.2 **Assessment**
55.2a **Psychiatric Assessment of the Older Patient**
Davangere P. Devanand, M.D., and Andrea Delgado-Peraza, M.D.
55.2b **Complementary and Integrative Geriatric Psychiatry**
Sarah A. Nguyen, M.D., Hanadi Ajam Oughli, M.D., Harris A. Eyre, M.B.B.S., Ph.D., and Helen Lavretsky, M.D., M.S.
55.2c **The Aging Brain**
Robert J. Boland, M.D., and Iqbal Ahmed, M.D., F.R.C.Psych.
55.2d **Psychological Changes With Normal Aging**
Tina L. Mayes, Ph.D., Maya A. Reiter, M.S., Ph.D., and Jennifer J. Dunkin, Ph.D.
55.2e **Neuropsychological Evaluation**
Gauri Savla, Ph.D., and Barton W. Palmer, Ph.D.
55.2f **Genetics of Late-Life Neurodegenerative Disorders**
Seethalakshmi Ramanathan, M.D., M.P.A., and Stephen J. Glatt, Ph.D.
55.3 **Psychiatric Disorders of Late Life**

55.3a Assessment of Functioning
David J. Moore, Ph.D., Maulika Kohli, M.S., and Pariya L. Fazeli, Ph.D.
55.3b Psychiatric Problems in the Medically Ill Geriatric Patient
Anna Borisovskaya, M.D., Beth Ann Labardi, M.D., Matthew Alan Schreiber, M.D., Ph.D., and Marcella Pascualy, M.D.
55.3c Sleep Disorders in Older Adults
Jana R. Cooke, M.D., and Sonia Ancoli-Israel, Ph.D.
55.3d Anxiety Disorders in Older Adults
Helmet T. Karim, Ph.D., and Carmen Andreescu, M.D.
55.3e Geriatric Mood Disorders
Talya Shahal, M.D., Jennie Davidow, M.D., Jennifer R. Gatchel, M.D., Ph.D., and Brent P. Forester, M.D., M.Sc.
55.3f Alzheimer Disease and Other Neurocognitive Disorders
Gary W. Small, M.D., and Stephen T. Chen, M.D.
55.3g Delirium
Jayson Varughese, M.D., M.S., and Susan W. Lehmann, M.D.
55.3h Schizophrenia and Other Psychotic Disorders
Carl I. Cohen, M.D., and Paul D. Meesters, M.D.
55.3i Personality Disorders
Marc E. Agronin, M.D.
55.3j Alcohol and Substance Abuse in Older Adults
Steve Koh, M.D., M.P.H., M.B.A., Hilary Gould, Ph.D., Omar Ghosn, M.D., and Michael Hann, M.D., M.B.A., M.S., C.P.E., F.A.C.H.E., F.A.P.A.
55.4 Treatment of Psychiatric Disorders

55.4a Geriatric Psychopharmacology: General Principles
Bruce G. Pollock, M.D., Ph.D., F.R.C.P.C.
55.4b Antidepressants and Mood Stabilizers
Carl Salzman, M.D.
55.4c Antianxiety Drugs
Alastair J. Flint, M.B., and Kathleen S. Bingham, M.D., Ph.D.
55.4d Antipsychotic Drugs
Kelly C. Lee, Pharm.D., M.A.S., B.C.P.P., and Vivian Hook, Ph.D.
55.4e Antidementia Drugs
Iqbal Ahmed, M.D., F.R.C.Psych.
55.4f Electroconvulsive Therapy and Neurostimulation Treatments
Mustafa M. Husain, M.D., Tarek K. Rajji, M.D., F.R.C.P.C., and Robert Garrett, M.D.
55.4g Psychosocial Factors in Psychotherapy of the Older Adult s
Joel I. Sadavoy, M.D., F.R.C.P.(C)
55.4h Individual Psychotherapy
Joel I. Sadavoy, M.D., F.R.C.P.(C), and Lawrence W. Lazarus, M.D.
55.4i Cognitive-Behavioral Therapy
John R. McQuaid, Ph.D., Eric Granholm, Ph.D., and Patricia A. Areán, Ph.D.
55.4j Family Intervention and Therapy With Older Adults
Lauren N. DeCaporale-Ryan, Ph.D., Ann Elizabeth Cornell, Psy.D., C.E.A.P., Jessica Goodman, Ph.D., and Carol Ann Podgorski, Ph.D., M.P.H., M.S.
55.4k Group Therapy

Kenneth M. Schwartz, M.D., F.R.C.P.C., Ingrid Söchting, Ph.D., R.Psych., C.G.P., Shayne Vitemb, M.A., L.M.F.T., C.G.P., and Alvin Keng, M.D., F.R.C.P.C.

55.4l Counseling and Support Needs of Dementia Caregivers

Liat Ayalon, Ph.D., and Patricia A. Areán, Ph.D.

55.5 Financial Issues in the Delivery of Geriatric Psychiatric Care

Steve Koh, M.D., M.P.H., M.B.A., Helen H. Kyomen, M.D., M.S., Michael Hann, M.D., M.B.A., M.S., C.P.E., F.A.C.H.E., F.A.P.A., and Hilary Gould, Ph.D.

55.6 Special Areas of Interest

55.6a Psychiatric Aspects of Long-Term Care

Abhilash K. Desai, M.D., and George T. Grossberg, M.D.

55.6b Forensic Aspects

David Naimark, M.D., Keith A. Davidson, J.D., Ariana Nesbit, M.D., M.B.E., and Ansar M. Haroun, M.D. (Retired)

55.6c Ethical Issues

Nathan Fairman, M.D., M.P.H.

55.6d Gender Issues

Helen H. Kyomen, M.D., M.S.

55.6e The Legal and Ethical Analysis of Older Adult Abuse

Paul Richard Greenwood, L.L.B. (Hons.), Ansar M. Haroun, M.D. (Retired), David Naimark, M.D., and Ariana Nesbit, M.D., M.B.E.

55.6f Geriatric Psychiatry: Sexuality and Aging

Abesh Kumar Bhattacharjee, M.D., Ph.D., Niraj K. Asthana, M.D., and Daniel D. Sewell, M.D.

55.6g HIV and Aging

Laura M. Campbell, M.S., María Marquine, Ph.D., and Raeanne C. Moore, Ph.D.

55.6h Technology for Seniors

Avery Quynh, B.S., Vanessa Scott, B.S., Colin A. Depp, Ph.D., and Ipsit V. Vahia, M.D.

55.6i Positive Psychiatry of Aging

Barton W. Palmer, Ph.D., and Dilip V. Jeste, M.D.

56 Public Psychiatry

56.1 Public and Community Psychiatry

Nidal Moukaddam, M.D., Ph.D., and Asim A. Shah, M.D.

56.2 Reforming Health Care

Nidal Moukaddam, M.D., Ph.D., and Asim A. Shah, M.D.

56.3 The Role of the Hospital in the Care of Persons with Mental Illness

Erica Taylor, M.D., M.P.H.

56.4 Mental Health Services Research

Lisa B. Dixon, M.D., M.P.H., and Milton Wainberg, M.D.

56.5 The Psychiatric Hospitalist

Angel Cienfuegos, M.D.

56.6 Psychiatric Rehabilitation

Lujain Alhajji, M.D., Vanessa Padilla, M.D., and Radu V. Saveanu, M.D.

56.7 A Sociocultural Framework for Mental Health and Substance Abuse Service Inequities

Margarita Alegría, Ph.D., Bernice A. Pescosolido, Ph.D., Kiara Alvarez, Ph.D., and Glorisa Canino, Ph.D.

56.8 Criminalization of Persons With Serious Mental Illness

Linda E. Weinberger, Ph.D., and H. Richard Lamb, M.D.

57 Psychiatric Education

57.1 Graduate Psychiatric Education

George Alan Keepers, M.D.

57.2 Examining Professional Patients

M. Sagman Kayatekin, M.D., George Bombel, Ph.D., and Kim-Lan Tran Czelusta, M.D.

58 Ethics and Forensic Psychiatry

58.1 Clinical-Legal Issues in Psychiatry

Clarence Watson, J.D., M.D., Spencer Eth, M.D., and Gregory B. Leong, M.D.

58.2 Ethics in Psychiatry

Laura Weiss Roberts, M.D., M.A.

58.3 Correctional Psychiatry

Arya Shah, M.D., and Jhilam Biswas, M.D.

59 History of Psychiatry

Edward Shorter, Ph.D., F.R.S.C.

60 World Aspects of Psychiatry

Sam Gnanapragasam, B.Sc., M.B.B.S., M.R.C.Psych., Max Pemberton, M.B.B.S., B.Sc., M.R.C.Psych., Pedro Ruiz, M.D., Antonio Ventriglio, M.D., Ph.D., and Dinesh Bhugra, C.B.E., M.A., M.Sc., M.B.B.S., F.R.C.P., F.R.C.P.E., F.R.C.Psych., F.F.P.H., M.Phil., Ph.D., F.R.C.Psych. (Hon), F.H.K.C.Psych. (Hon), F.A.C.P. (Hon), F.A.M.S. (Singapore), F.I.M.S.A. (Hon)

Appendix A : History of Psychiatry Table

Appendix B : Glossary of Psychiatry and Psychology Terms

Appendix C : Comparative Diagnoses

Index

Contributors

Anissa Abi-Dargham, M.D. Professor and Chair, Department of Psychiatry and Behavioral Health, Stony Brook University, Renaissance School of Medicine, Stony Brook, New York
1.27 Radiotracer Imaging with Positron Emission Tomography and Single Photon Emission Computed Tomography: Fundamental Principles, Methodology and Role in Neuropsychiatric Research

David L. Adams, Psy.D., B.C.B.A.-D. Adjunct Core Faculty, Department of Applied Behavior Analysis, School Psychology, and Educational Counseling, National University, Orange, California
7.5 Personality Assessment of Adults

Russell L. Adams, Ph.D., A.B.P.P-NP., C.L. David Ross Boyd Professor, Department of Psychiatry and Behavioral Sciences, University of Oklahoma, College of Medicine, Oklahoma City, Oklahoma
7.5 Personality Assessment of Adults

Evan E. Afshin, M.D. Clinical Research Fellow, Department of Physiology and Biophysics, Institute of Computational Biomedicine, Weill Cornell Medical College, New York, New York
1.18 Genome, Transcriptome, and Proteome: The Molecular Genetics, Biochemistry, and Multi-Omics Underlying the

Neurobiology of Mental Disorders

Neil Krishan Aggarwal, M.D., M.B.A., M.A. Assistant Professor of Clinical Psychiatry, Department of Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons, New York, New York
29.2 Cultural Assessments in Psychiatry

Marc E. Agronin, M.D. Senior Vice President, Department of Behavioral Health, Miami Jewish Health, Miami, Florida
55.3i Personality Disorders

Samoon Ahmad, M.D. Clinical Professor, Department of Psychiatry, New York University, Grossman School of Medicine, New York, New York
33.11a Anticholinergics and Amantadine

Iqbal Ahmed, M.D., F.R.C.Psych. Clinical Professor, Department of Psychiatry and Geriatric Medicine, University of Hawaii, at Manoa, John A Burns School of Medicine Honolulu, Hawaii; Clinical Professor, Department of Psychiatry, Uniformed Services University of the Health Sciences, F. Edward Hébert School of Medicine, Bethesda, Maryland
10.1 Cognitive Disorders: Introduction, 10.5 Cognitive Disorders Due to Another Medical Condition or Substance, 10.3 Major Neurocognitive Disorders (Dementia), 55.2c The Aging Brain, 55.4e Antidementia Drugs

Olusola Ajilore, M.D., Ph.D. Professor, Department of Psychiatry, University of Illinois College of Medicine, Chicago, Illinois
33.1a General Principles of Psychopharmacology

Anne Marie Albano, Ph.D. Professor, Department of Psychiatry, Columbia University Vagelos College of Physicians and Surgeons, New York, New York
53.2 Cognitive-Behavioral Psychotherapy for Children and Adolescents

Margarita Alegría, Ph.D. Chief, Disparities Research Unit, Department of Medicine, Massachusetts General Hospital; Professor, Departments of Medicine and Psychiatry, Harvard University, Harvard Medical School, Boston, Massachusetts
56.7 A Sociocultural Framework for Mental Health and Substance Abuse Service Inequities

Nadejda Alekseeva, M.D. Psychiatrist, Mental Health Department, Overton Brooks VA Medical Center, Shreveport, Louisiana
2.10 Neuropsychiatric Aspects of Prion Disease

Lujain Alhajji, M.D. Assistant Professor, Division of Psychiatry and Behavioral Sciences, University of Miami Leonard M. Miller School of Medicine, Miami, Florida
56.6 Psychiatric Rehabilitation

Yesne Alici, M.D. Associate Attending Psychiatrist/Vice Chair of Clinical Operations, Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, New York
27.8 Psycho-Oncology

Carol L. Alter, M.D. Professor and Associate Chair Clinical Affairs, Department of Psychiatry, University of Texas, Dell Medical School, Austin, Texas
27.1 History and Current Trends

Kiara Alvarez, Ph.D. Bloomberg Assistant Professor of American Health, Department of Health, Behavior and Society, Johns Hopkins University, School of Medicine, Bloomberg School of Public Health, Baltimore, Maryland
56.7 A Sociocultural Framework for Mental Health and Substance Abuse Service Inequities

Lawrence Amsel, M.D., M.P.H. Associate Professor of Psychiatry, Department of Psychiatry, Columbia University, Vagelos College of

Physicians and Surgeons, New York, New York
33.20b Brain Stimulation Methods

Sonia Ancoli-Israel, Ph.D. Professor Emeritus, Department of Psychiatry, University of California, School of Medicine, San Diego, California
55.3c Sleep Disorders in Older AdultsCarmen Andreescu, M.D. Associate Professor of Psychiatry, Department of Psychiatry, University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania

10.4 Mild Cognitive Impairment and Amnestic Disorders, 55.3d Anxiety Disorders in Older Adults

Andrew F. Angelino, M.D. Professor of Clinical Psychiatry, Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
2.8 Neuropsychiatric Aspects of HIV Infection and AIDS

Alan Anticevic, Ph.D. Associate Professor, Department of Psychiatry, Yale University, School of Medicine, New Haven, Connecticut
1.24 Computational Modeling Approaches to Psychiatry

Patricia A. Areán, Ph.D. Professor, Department of Psychiatry and Behavioral Sciences, University of Washington, School of Medicine, Seattle, Washington
55.4i Cognitive-Behavioral Therapy, 55.4l Counseling and Support Needs of Dementia Caregivers

Tiffany Arnold, B.S., M.D. Public Psychiatry Fellow, Department of Psychiatry, University of California, School of Medicine, San Francisco, California
33.13a.2 Opioid Receptor Antagonists: Naltrexone and Nalmefene

Nana Asabere, M.D. Child and Adolescent Psychiatry Fellow, Department of Child and Adolescent Psychiatry, New York

Presbyterian—Weill Cornell Medical College/Columbia University, Vagelos College of Physicians and Surgeons, New York, New York
54.17 Racism and Discrimination's Impact on the Health and Well-Being of Developing Children and Adolescents

Peter Ash, M.D. Professor and Director, Psychiatry and Law Service, Department of Psychiatry and Behavioral Sciences, Emory University, School of Medicine, Atlanta, Georgia
54.6 Forensic Child and Adolescent Psychiatry

Mahliqa Ashraf, H.B.Sc. Student Researcher, Molecular Brain Science, Campbell Family Mental Health Research Institute, Center for Addiction and Mental Health (CAMH); Department of Human Biology: Neuroscience and Department of Physiology, University of Toronto, Toronto, Ontario, Canada
42.1 Language Disorder, 42.3 Speech Sound Disorder, 42.4 Stuttering

Deeba Ashraf, M.D. Clinical Assistant Professor of Psychiatry, Menninger Department of Psychiatry and Behavioral Sciences, Baylor University, College of Medicine/Center for Psychoanalytic Studies, Houston, Texas
6.1 Psychoanalytic Theories and Treatment: Perspective from North America and Great Britain

Niraj K. Asthana, M.D. Resident Physician, Department of Psychiatry, University of California, School of Medicine, San Diego, California
55.6f Geriatric Psychiatry: Sexuality and Aging

Evelyn Attia, M.D. Professor of Psychiatry, Department of Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons/Irving Medical Center, New York, New York
22 Feeding and Eating Disorders

Jehannine Austin, Ph.D., M.S. Professor of Psychiatry and Medical Genetics, Departments of Psychiatry and Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada
30.2 Genetic Counseling for Psychiatric Conditions

Ajit Avasthi, M.B.B.S., M.D. Retired Senior Professor and Head, Department of Psychiatry, Post Graduate Institute of Medical Education and Research, Chandigarh, India
33.19c Biologic Therapies in India

Allan David Axelrad, M.D. Clinical Associate Professor, Department of Psychiatry, Baylor University, College of Medicine, Houston, Texas
32.19 Hypnosis

Liat Ayalon, Ph.D. Professor, School of Social Work, Bar Ilan University, Ramat Gan, Israel
55.4l Counseling and Support Needs of Dementia Caregivers

Olatunde Ayinde, M.B.B.S., M.S., F.W.A.C.P. Lecturer in Psychiatry/Honorary Consultant Psychiatrist, Department of Psychiatry, College of Medicine, University of Ibadan/University College Hospital, Ibadan, Oyo, Nigeria
33.19b Biologic Therapies in Africa

Reena Baharani, M.D. Consultation Liaison Psychiatry Attending, Department of Psychiatry, Westchester Medical Center/New York Medical College, Valhalla, New York
27.14 Psychiatric Care of the Burned Patient

Deanna M. Barch, Ph.D. Professor of Psychological and Brain Sciences, Psychiatry, and Radiology; Vice Dean of Research, Arts & Sciences; Gregory B. Couch Professor of Psychiatry; Editor-in-Chief, Biological Psychiatry: Global Open Science, Washington University in St. Louis, School of Medicine, St. Louis, Missouri
1.28 Functional Brain Connectivity and Psychopathology

Evangelina Barnard, Psy.D. Postdoctoral Fellow in Pediatric Neuropsychology, Division of Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
35.3 Neuropsychological and Psychoeducational Assessment of Children

Brian S. Barnett, M.D. Assistant Professor of Psychiatry, Department of Psychiatry and Psychology, Cleveland Clinic Lerner College of Medicine, Cleveland, Ohio
11.6 Hallucinogen-Related Disorders

Courtney Barry, Psy.D., M.S. Assistant Professor, Department of Psychiatry and Behavioral Medicine, Medical College of Wisconsin, Milwaukee, Wisconsin
32.17 Acceptance and Commitment Therapy

Kelly Barth, D.O. Professor, Department of Psychiatry and Behavioral Sciences, Internal Medicine, Medical University of South Carolina, College of Medicine, Charleston, South Carolina
33.13a.1 Opioid Maintenance Treatments: Methadone and Buprenorphine

Gaston Baslet, M.D. Chief, Division of Neuropsychiatry, Department of Psychiatry, Brigham and Women’s Hospital/ Harvard University, Harvard Medical School, Boston, Massachusetts
2.4 Neuropsychiatric Aspects of Epilepsy and Psychogenic Nonepileptic Seizures

Anthony W. Bateman, M.A., F.R.C.Psych. Consultant Psychiatrist and Visiting Professor, Department of Clinical, Educational and Health Psychology, Anna Freud Centre and University College London, London, United Kingdom
32.14 Mentalization-Based Treatment

Eraka Bath, M.D. Clinical Professor of Psychiatry, Department of Psychiatry, Division of Population Behavioral Health, University of California, David Geffen School of Medicine, Los Angeles, California
54.17 Racism and Discrimination's Impact on the Health and Well-Being of Developing Children and Adolescents

John Battaglia, M.D. Clinical Adjunct Professor of Psychiatry, Department of Psychiatry, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin
32.2 Supportive Psychotherapy

Anna E. Bauer, Ph.D., M.P.H. Assistant Professor, Center for Women's Mood Disorders, Department of Psychiatry, University of North Carolina School of Medicine, Chapel Hill, North Carolina
33.17a Sex Differences in Psychopharmacology

Scott Beach, M.D. Associate Professor of Psychiatry, Division of Psychiatry and Medicine, Massachusetts General Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
27.2 Cardiovascular Disorders

William R. Beardslee, M.D. Director, Baer Prevention Initiatives and Distinguished Gardner/Monks, Professor of Child Psychiatry, Department of Psychiatry, Boston Children's Hospital, Boston, Massachusetts
54.9 Prevention of Psychiatric Disorders in Children and Adolescents

Judith S. Beck, Ph.D. Clinical Professor of Psychology in Psychiatry, Department of Psychiatry, University of Pennsylvania, Raymond and Ruth Perelman School of Medicine, Philadelphia, Pennsylvania
32.9 Cognitive Behavior Therapy

Robert H. Belmaker, M.D. Professor Emeritus of Psychiatry, Division of Psychiatry, Faculty of Medicine, Ben-Gurion University, Be'er-Sheva, Israel

33.10f Lithium

Dror Ben-Zeev, Ph.D. Professor of Psychiatry and Behavioral Sciences, Psychiatry and Behavioral Sciences, University of Washington, School of Medicine, Seattle, Washington
12.13 Psychiatric Rehabilitation

Ruth M. Benca, M.D., Ph.D. Professor and Chair, Department of Psychiatry and Behavioral Medicine, Wake Forest University, School of Medicine, Atrium Health Wake Forest Baptist, Winston-Salem, North Carolina
1.15 Basic Science of Sleep

Elissa P. Benedek, M.D. Adjunct Clinical Professor, Department of Psychiatry, University of Michigan Medical School; Psychiatric Resident Teaching, Forensic Fellow Teaching, Medical Student Teaching, Ann Arbor, Michigan; Clinical Professor, Department of Psychiatry, Michigan State University, College of Human Medicine, East Lansing, Michigan; Clinical Professor, Department of Psychiatry and Behavior Neurosciences, Wayne State University School of Medicine, Detroit, Michigan
54.6 Forensic Child and Adolescent Psychiatry

David M. Benedek, M.D. Professor and Chair, Department of Psychiatry, School of Medicine, Uniformed Services University of the Health Sciences, F. Edward Hébert School of Medicine, Bethesda, Maryland
30.6 Disaster Psychiatry: Disasters, Terrorism, and War

Tami Benton, M.D. Chair, Department of Child and Adolescent Psychiatry and Behavioral Sciences, Department of Psychiatry, University of Pennsylvania, Raymond and Ruth Perelman School of Medicine, The Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania
50.1 Depressive Disorders and Suicide

Eugene V. Beresin, M.D., M.A. Executive Director, The Clay Center for Young Healthy Minds; Director, Division of Professional and Public Education, Department of Psychiatry, Massachusetts General Hospital; Professor of Psychiatry, Harvard University, Harvard Medical School, Department of Psychiatry, Massachusetts General Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
53.8 The Treatment of Adolescents

R. Lindsey Bergman, Ph.D. Division of Child and Adolescent Psychiatry, University of California, David Geffen School of Medicine, Los Angeles, California
51.4 Selective Mutism

William Bernet, M.D. Professor Emeritus, Department of Psychiatry and Behavioral Sciences, Vanderbilt University School of Medicine, Nashville, Tennessee
28.4 Relational Problems

Karin M. Best, Ph.D. Assistant Adjunct Professor, Department of Psychiatry and Biobehavioral Sciences, Division of Child and Adolescent Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, David Geffen School of Medicine, Los Angeles, California
49.1 Reactive Attachment Disorder and Disinhibited Social Engagement Disorder

Abesh Kumar Bhattacharjee, M.D., Ph.D. Associate Professor, Department of Psychiatry, University of California, School of Medicine/VA San Diego Healthcare system, San Diego, California
55.6f Geriatric Psychiatry: Sexuality and Aging

Dinesh Bhugra, C.B.E., M.A., M.Sc., M.B.B.S., F.R.C.P., F.R.C.P.E., F.R.C.Psych., F.F.P.H., M.Phil., Ph.D., F.R.C.Psych. (Hon), F.H.K.C.Psych. (Hon), F.A.C.P. (Hon), F.A.M.S. (Singapore),

F.I.M.S.A. (Hon) Emeritus Professor of Mental Health and Cultural Diversity, Health Service & Population Research, Institute of Psychiatry, Psychology and Neuroscience at King's College London, London, United Kingdom
60 World Aspects of Psychiatry

O. Joseph Bienvenu, M.D., Ph.D. Judy Yin Shih, PhD Professor in Anxiety Disorders, Psychiatry and Behavioral Sciences, Johns Hopkins Hospital/Johns Hopkins University School of Medicine, Baltimore, Maryland
27.11 Psychiatric Morbidity Following Critical Illness

Kathleen S. Bingham, M.D., Ph.D. Assistant Professor of Psychiatry, Department of Psychiatry, University Health Network/University of Toronto, Toronto, Ontario, Canada
55.4c Antianxiety Drugs

Adam Bisaga, M.D. Professor of Psychiatry at CUMC, Department of Psychiatry, New York State Psychiatric Institute/Columbia University Vagelos College of Physicians and Surgeons, New York, New York
11.9 Opioid Use and Related Disorders: From Neuroscience to Treatment

Jhilam Biswas, M.D. Director of the Psychiatry, Law and Society Program, Department of Psychiatry, Brigham and Women's Hospital, Newton Highlands, Massachusetts
58.3 Correctional Psychiatry

Deborah Blacker, M.D., Sc.D. Professor of Psychiatry, Harvard University, Harvard Medical School; Professor and Deputy Chair, Harvard T.H. Chan School of Public Health, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts
7.8 Psychiatric Rating Scales

Mark A. Blais, Psy.D. Associate Professor of Psychology (Psychiatry), Massachusetts General Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
7.8 Psychiatric Rating Scales

Pierre Blier, M.D., Ph.D. Professor, Department of Psychiatry, University of Ottawa, Ottawa, Ontario, Canada
33.5a.2 Serotonin-Norepinephrine Reuptake Inhibitors (SNRIs)
33.5a.1 Serotonin Reuptake Inhibitors (SRIs), 33.5b $5HT_{1A}$ *Receptor Partial Agonists (Buspirone), 33.5c Vilazodone and Vortioxetine*

Michael H. Bloch, M.D., M.S. Associate Professor, Child Study Center, Department of Psychiatry, Yale University, School of Medicine, New Haven, Connecticut
35.1 Psychiatric Examination of the Infant, Child, and Adolescent

Ilaina Blum, M.A., M.F.T. Education Consultant, Division of Child/Adolescent Psychiatry, UCLA Resnick Neuropsychiatric Hospital (RNPH), University of California, David Geffen School of Medicine, Los Angeles, California
54.8 School Consultation and Collaborating With the Educational System

Michael Blumenfield, M.D. Sidney E. Frank Distinguished Professor Emeritus of Psychiatry and Behavioral Sciences, Department of Psychiatry, New York Medical College, Valhalla, New York
27.14 Psychiatric Care of the Burned Patient

Robert J. Boland, M.D. Senior Vice President and Chief of Staff, The Menninger Clinic; Professor and Vice Chair, Menninger Department of Psychiatry and Behavioral Sciences, The Brown Foundation Endowed Chair in Psychiatry, Baylor University, College of Medicine, Houston, Texas

10.5 Cognitive Disorders Due to Another Medical Condition or Substance, 12.14 Medical Health in Schizophrenia, 13.1 Mood Disorders: Historical Introduction and Conceptual Overview, 13.3 Mood Disorders: Clinical Features, 27.6 Diabetes: Psychosocial Issues and Psychiatric Disorders, 28.2 Adult Antisocial Behavior, Criminality, and Violence, 55.2c The Aging Brain

George Bombel, Ph.D. Assistant Professor/Senior Clinical Psychologist, Department of Psychiatry and the Behavioral Sciences, Baylor University, College of Medicine/The Menninger Clinic, Houston, Texas
57.2 Examining Professional Patients

Gary R. Bond, Ph.D. Senior Research Associate, Westat, Lebanon, New Hampshire
12.13 Psychiatric Rehabilitation

Olivier Bonnot, M.D., Ph.D. Full Professor of Child and Adolescent Psychiatry, Child and Adolescent Department, Barthelemy Durand Hospital/Paris-Saclay School of Medicine, Kremlin-Bicetre, France
2.14 Neuropsychiatry of Neurometabolic and Neuroendocrine Disorders

Anna Borisovskaya, M.D. Clinical Associate Professor of Psychiatry, Department of Psychiatry and Behavioral Sciences, University of Washington, School of Medicine, Seattle Neuropsychiatric Treatment Center, Seattle, Washington
55.3b Psychiatric Problems in the Medically Ill Geriatric Patient

Mousa Sobhy Azer Botros, M.D. Assistant Professor of Clinical Psychiatry, Psychiatry and Behavioral Sciences, University of Miami Leonard M. Miller School of Medicine, Miami, Florida
33.6c Mirtazapine

Mark E. Bouton, Ph.D. University Distinguished Professor and Robert B. Lawson Green & Gold Professor of Psychology,

Department of Psychological Science, University of Vermont, Robert Larner MD College of Medicine, Burlington, Vermont
3.3 Learning Theory

Scott E. Bowen, Ph.D. Professor and Department Chair, Psychology/Behavioral, Cognitive Neuroscience, Wayne State University, School of Medicine/College of Liberal Arts & Sciences, Detroit, Michigan
11.7 Inhalant-Related Disorders

Oren Boxer, Ph.D. Clinical Neuropsychologist, Department of Psychiatry, Insight Collective, South Pasadena, California; University of California, David Geffen School of Medicine, Los Angeles, California
40 Specific Learning Disorder

Kaitlin Boyle, M.D. Child and Adolescent Psychiatry Fellow, Department of Psychiatry and Behavioral Sciences, Duke University Health System, Durham, North Carolina
52 Early-Onset Psychotic Disorders

Helen Bradshaw, M.D. Resident Physician, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
30.1 Psychiatry and Reproductive Medicine

Matthew N. Brams, M.D. Clinical Professor of Psychiatry, Department of Psychiatry, Baylor University, College of Medicine, Houston, Texas
54.12 Impact on Parents of Raising a Child with Psychiatric Illness and/or Developmental Disability

William S. Breitbart, M.D., F.A.C.L.P., D.F.A.P.A., F.A.P.O.S. Chairman, Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, New York
27.8 Psycho-Oncology

Adam M. Brenner, M.D. Distinguished Teaching Professor and Vice Chair for Education, Department of Psychiatry Department, University of Texas Southwestern Medical School, Dallas, Texas
6.3 Other Psychodynamic Schools, 32.1 Psychotherapy Integration, 32.23 Other Methods of Psychotherapy

Amy S. Brenner, M.S.W., L.C.S.W. Volunteer Clinical Faculty, Department of Psychiatry, University of Texas Southwestern Medical School, Dallas, Texas
32.1 Psychotherapy Integration

Tim Bressmann, Ph.D. Associate Professor, Speech-Language Pathology, Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
42.5 Unspecified Communication Disorder

Evelyn J. Bromet, Ph.D. Distinguished Professor Emerita, Department of Psychiatry, Stony Brook University, Renaissance School of Medicine, Stony Brook, New York
12.17 Other Psychotic Disorders

John O. Brooks, III, Ph.D., M.D. Professor-in-Residence, Department of Psychiatry and Biobehavioral Sciences, University of California, David Geffen School of Medicine, Los Angeles, California
7.6 Medical Assessment and Laboratory Testing in Psychiatry

Shaun P. Brothers, Ph.D., M.B.A. Associate Professor, Department of Psychiarty and Behavioral Sciences, University of Miami, Leonard M. Miller School of Medicine, Miami, Florida
1.8 Neuropeptides: Biology, Regulation and Role in Neuropsychiatric Disorders

E. Sherwood Brown, M.D., Ph.D., M.B.A. Professor and Vice Chair of Clinical Research, Department of Psychiatry, University of Texas Southwestern, Medical School, Dallas, Texas
27.5 Respiratory Disorders

André Russowsky Brunoni, M.D., Ph.D. Associate Professor, Psychiatry, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil
33.19d Biologic Therapies in Latin America

Jonathan R. Buchholz, M.D. Assistant Professor, Department of Psychiatry and Behavioral Sciences, VA Puget Sound Health Care System and University of Washington, School of Medicine, Seattle, Washington
33.13a.1 Opioid Maintenance Treatments: Methadone and Buprenorphine

Regina Bussing, M.D., M.S.H.S. Professor, Department of Psychiatry, University of Florida, College of Medicine, Gainesville, Florida
54.10 Child Mental Health Services Research

Joanne A. Byars, M.D. Assistant Professor of Psychiatry, Department of Psychiatry and Behavioral Health, Geisinger/Geisinger Commonwealth School of Medicine, Danville, Pennsylvania
2.5 Neuropsychiatric Consequences of Traumatic Brain Injury

William M. Byne, M.D., Ph.D. Professor of Clinical Psychiatry, Department of Psychiatry, Columbia University Vagelos College of Physicians and Surgeons, New York, New York
21.2 Homosexuality, Gay and Lesbian Identities, and Homosexual Behavior, 21.4 Gender Identity, Gender Diversity, Gender Dysphoria/Incongruence

Christine C. Call, Ph.D. Postdoctoral Fellow, Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania
22 Feeding and Eating Disorders

Katherine Camfield, M.D., M.P.H. Attending Physician, Psychiatric Emergency Services, Department of Behavioral Health, Denver Health Medical Center, Denver, Colorado
31.3 Other Psychiatric Emergencies

Laura M. Campbell, M.S. Predoctoral Fellow, Department of Psychiatry, San Diego State University/University of California, School of Medicine, Joint Doctoral Program in Clinical Psychology, San Diego, California
55.6g HIV and Aging

Glorisa Canino, Ph.D. Professor, Director Behavioral Sciences Research Institute, Department of Academic Affairs, University of Puerto Rico, School of Medicine, San Juan, Puerto Rico
56.7 A Sociocultural Framework for Mental Health and Substance Abuse Service Inequities

Jason P. Caplan, M.D. Professor of Psychiatry, Department of Psychiatry, St. Joseph's Hospital and Medical Center, Phoenix, Arizona/Creighton University School of Medicine, Omaha, Nebraska
27.12 Psychocutaneous Disorder

Vladimir Carli, M.D., Ph.D. Associate Professor of Psychiatry, National Centre for Suicide Research and Prevention of Mental Ill-Health (NASP), Karolinska Institute, Stockholm, Sweden
31.1 Suicide: Overview of Epidemiology, Risk and Protective Factors, Treatment, and Prevention

Gabrielle A. Carlson, M.D. Professor of Psychiatry and Pediatrics, Department of Psychiatry and Behavioral Health, Stony Brook University, Renaissance School of Medicine, Stony Brook, New York
50.2 Early-Onset Bipolar Disorder

Edward Castelo, M.D. Fellow Physician, Department of Child and Adolescent Psychiatry, Jane and Terry Semel Institute for

Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
54.8 School Consultation and Collaborating With the Educational System

Joseph M. Cerimele, M.D., M.P.H. Associate Professor, Department of Psychiatry and Behavioral Sciences, UW Medicine/University of Washington School of Medicine, Seattle, Washington
14.2 Clinical Features of the Anxiety Disorders

Aldrich Chan, Psy.D. Adjunct Professor Course Lead, Neuropsychologist, Graduate School of Education and Psychology, Pepperdine University, Los Angeles, California
3.1 Contributions of the Psychological Sciences: Sensation, Perception, and Cognition

Ganesh B. Chand, Ph.D. Assistant Professor of Radiology, Department of Radiology, Washington University in St. Louis, School of Medicine, St. Louis, Missouri
12.9 Molecular Brain Imaging in Schizophrenia

Tara Chandrasekhar, M.D. Assistant Professor of Psychiatry and Behavioral Sciences, Division of Child & Family Mental Health & Community Psychiatry, Duke Health/Duke University School of Medicine, Durham, North Carolina
52 Early-Onset Psychotic Disorders

Serena Chang, M.D., B.A. Director of Psychiatry, Department of Behavioral Health, Callen-Lorde Community Health Center, New York, New York
54.16 Gay, Bisexual, Transgender, Queer/Questioning, Intersex, and Asexual (LGBTQIA) Youth

Moses V. Chao, Ph.D. Professor, Departments of Cell Biology, Neuroscience and Physiology, and Psychiatry, New York University Grossman School of Medicine, New York, New York

1.10 Neurotrophic Factors

Dennis S. Charney, M.D. Dean, Office of the Dean, Icahn School of Medicine at Mount Sinai/Mount Sinai Health System, New York, New York
14.4 Anxiety Disorders: Neurobiology and Neuroscience

Irene Chatoor, M.D. Professor of Psychiatry and Pediatrics, Department of Psychiatry and Behavioral Sciences, Children's National Medical Center and George Washington University, School of Medicine and Health Sciences, Washington, DC
46 Feeding and Eating Disorders of Infancy and Early Childhood

Stephen T. Chen, M.D. Health Sciences Clinical Professor, Department of Psychiatry and Biobehavioral Sciences, University of California, David Geffen School of Medicine, Los Angeles, California
55.3f Alzheimer Disease and Other Neurocognitive Disorders

Sneha Chenji, Ph.D. Postdoctoral Research Associate, Psychiatry, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
37 Neuroimaging in Psychiatric Disorders of Childhood

Lois Choi-Kaisn, M.D., M.Ed. Director, Gunderson Personality Disorders Institute, Division of Women's Mental Health, McLean Hospital, Belmont, Massachusetts
32.7 Treatment of Borderline Personality Disorder

Paul Ciechanowski, M.D., M.P.H. Clinical Associate Professor, Department of Psychiatry and Behavioral Sciences, University of Washington, School of Medicine, Seattle, Washington
27.6 Diabetes: Psychosocial Issues and Psychiatric Disorders

Angel Cienfuegos, M.D. Associate Clinical Professor, Department of Psychiatry and Biobehavioral Sciences, Los Angeles County

Department of Mental Health, University of California, David Geffen School of Medicine, Los Angeles, California
56.5 The Psychiatric Hospitalist

Chiara Cirelli, M.D., Ph.D. Professor, Department of Psychiatry, University of Wisconsin, School of Medicine and Public Health, Madison, Wisconsin
1.15 Basic Science of Sleep

Michael R. Clark, M.D., M.P.H., M.B.A. Associate Professor, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
27.15 Management of Chronic Pain

Diana E. Clarke, Ph.D., M.Sc. Director of Research and Senior Epidemiologist/Research Statistician, Division of Research, American Psychiatric Association, Washington, DC
5.1 Quantitative and Experimental Methods in Psychiatry: Epidemiology

Anita H. Clayton, M.D. Wilford W. Spradlin Professor and Chair, Department of Psychiatry and Neurobehavioral Sciences, University of Virginia School of Medicine, Charlottesville, Virginia
33.17c Pharmacologic Treatment of Sexual Dysfunction

Brett Clementz, Ph.D. Distinguished Research Professor, Departments of Psychology and Neuroscience, University of Georgia, Athens, Georgia
12.11 Phenotypes of Psychosis

C. Robert Cloninger, M.D., Ph.D. Professor Emeritus of Psychiatry, Department of Psychiatry, Washington University in St. Louis, School of Medicine, St. Louis, Missouri
26 Personality Disorders

Robert T. Cobb, M.D. Associate Professor of Clinical Psychiatry, Department of Psychiatry and Behavioral Science, University of Southern California, Keck School of Medicine, Los Angeles, California
31.3 Other Psychiatric Emergencies

Barbara J. Coffey, M.D., M.S. Professor and Chairman, Department of Psychiatry and Behavioral Sciences, University of Miami Leonard M. Miller School of Medicine, Miami, Florida
47 Tic Disorders

Carl I. Cohen, M.D. State University of New York, Distinguished Service Professor, Division of Geriatric Psychiatry and Center of Excellence for Alzheimer's Disease, State University of New York, Downstate Health Sciences University, College of Medicine, Brooklyn, New York
55.3h Schizophrenia and Other Psychotic Disorders

Judith A. Cohen, M.D. Professor of Psychiatry, Psychiatry and Behavioral Health Institute, Allegheny Health Network, Drexel University College of Medicine, Pittsburgh, Pennsylvania
51.2 Posttraumatic Stress Disorder in Children and Adolescents

Emily M. Cohodes, M.S., M.Phil. Ph.D. Candidate in Clinical Psychology, Department of Psychology, Yale University, School of Medicine, New Haven, Connecticut
14.5 Neuroimaging and the Neuroanatomical Circuits Implicated in Anxiety Disorders

John N. Constantino, M.D. Liz and Frank Blake Chair for Children's Behavioral and Mental Health; Chief, Center for Behavioral and Mental Health, Children's Healthcare of Atlanta; Vice Chair, Department of Pediatrics; Professor of Psychiatry and Behavioral Sciences, Emory University School of Medicine; Adjunct Clinic Professor of Psychiatry and Behavioral Sciences, Morehouse College, School of Medicine, Atlanta, Georgia

1.3 Functional Genomics of Human Brain Development

Jana R. Cooke, M.D. Clinical Assistant Professor, Pulmonary Disease/Adult and Pediatric Sleep Medicine, Department of Medicine, Washington State University, Elson S. Floyd School of Medicine, Spokane, Washington; The Everett Clinic, Part of Optum Care, Everett, Washington
55.3c Sleep Disorders in Older Adults

Rena Cooper-Kazaz, M.D. Senior Lecturer (Clinical), Department of Psychiatry, Clalit Health Services, Tel Aviv Brull Community Mental Health Center (Ramat Chen), Affiliated With Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
33.15b Thyroid Hormones

Amy R. Corcoran, M.D. Assistant Professor of Psychiatry and Neurology, Department of Psychiatry and Behavioral Sciences, The Ken & Ruth Davee Department of Neurology, Northwestern University, Feinberg School of Medicine, Chicago, Illinois
2.5 Neuropsychiatric Consequences of Traumatic Brain Injury

Philip R. Corlett, Ph.D. Associate Professor, Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut
1.30 Learning Theory and Psychosis, Anxiety, and Addiction

Ann Elizabeth Cornell, Psy.D., C.E.A.P. Associate Professor of Psychiatry, Division of Collaborative Care and Wellness, University of Rochester, School of Medicine and Dentistry/Medical Center, Rochester, New York
55.4j Family Intervention and Therapy With Older Adults

Christoph U. Correll, M.D. Professor of Psychiatry and Molecular Medicine, Department of Psychiatry and Molecular Medicine; Professor of Child and Adolescent Psychiatry, Department of Child and Adolescent Psychiatry; Hofstra University/Northwell Health,

Donald and Barbara Zucker School of Medicine, Hempstead, New York; Charité—Universitätsmedizin Berlin, Berlin, Germany
12.12 Schizophrenia: Pharmacologic Treatment

Paul T. Costa, Jr., Ph.D. Adjunct Professor, Department of Geriatrics, Duke University School of Medicine, Durham, North Carolina
6.4 Approaches Derived from Philosophy and Psychology

Sara Costi, M.D. Clinical Instructor, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York; Wellcome Trust Doctoral Fellow, Department of Psychiatry, University of Oxford; Honorary Consultant, Oxford Health NHS Foundation Trust, Oxford, United Kingdom
14.4 Anxiety Disorders: Neurobiology and Neuroscience

Joseph T. Coyle, M.D. Eben S. Draper Professor of Psychiatry and Neuroscience (Emeritus), Department of Psychiatry, McLean Hospital, Belmont, Massachusetts/Harvard University, Harvard Medical School, Boston, Massachusetts
1.6 Amino Acid Neurotransmitters

Louis J. Cozolino, Ph.D. Professor of Psychology, Graduate School of Education and Psychology, Pepperdine University, Los Angeles, California
3.1 Contributions of the Psychological Sciences: Sensation, Perception, and Cognition

Justin F. Creeden, Ph.D. Graduate Research Assistant, Department of Neurosciences, University of Toledo, College of Medicine and Life Sciences, Toledo, Ohio
12.6 Cellular and Molecular Neuropathology of Schizophrenia

Adam Critchfield, M.D. Assistant Clinical Professor of Psychiatry, Department of Psychiatry, Columbia University, Vagelos College of

Physicians and Surgeons/Irving Medical Center, New York, New York
32.3 Psychoanalysis and Psychoanalytic Psychotherapy

Elizabeth A. Crocco, M.D. Medical Director, UM Memory Disorders Clinic, Center for Cognitive Neuroscience and Aging; Professor and Chief, Division of Geriatric Psychiatry, Department of Psychiatry and Behavioral Sciences, University of Miami Leonard M. Miller School of Medicine; Geriatric Psychiatry Training Director, Jackson Memorial Hospital, Miami, Florida
33.11b Cholinesterase Inhibitors and Memantine

Catherine Chang Crone, M.D. Director of Education, Department of Behavioral Health, Inova Health System, Falls Church, Virginia
27.13 Organ Transplantation

Ari B. Cuperfain, M.D., M.Sc. Psychiatry Resident, Department of Psychiatry, Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
33.18b Biologic Therapies: Geriatric Psychopharmacology

Kim-Lan Tran Czelusta, M.D. Associate Professor of Psychiatry, Menninger Department of Psychiatry and Behavioral Sciences, Baylor University, College of Medicine, Houston, Texas
57.2 Examining Professional Patients

Rodolfo Furlan Damiano, M.D., Ph.D. Collaborator Professor, Department of Psychiatry, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil
33.19d Biologic Therapies in Latin America

Smita Das, M.D., Ph.D., M.P.H. Clinical Associate Professor, Psychiatry and Behavioral Sciences, Stanford University, School of Medicine, Palo Alto, California
11.8 Tobacco-Related Disorders

Shruti Dave, Ph.D. Postdoctoral Fellow, Department of Medical Social Sciences, Northwestern University, Feinberg School of Medicine, Chicago, Illinois
3.4 Biology of Memory

Jennie Davidow, M.D. Instructor in Psychiatry and Staff Psychiatrist, Department of Psychiatry, Beth Israel Deaconess Medical Center/Harvard University, Harvard Medical School, Boston, Massachusetts
55.3e Geriatric Mood Disorders

Cameron J. Davidson, Ph.D. Postdoctoral Fellow, Psychology/Behavioral, Cognitive Neuroscience, Wayne State University, School of Medicine/College of Liberal Arts and Sciences, Detroit, Michigan
11.7 Inhalant-Related Disorders

Keith A. Davidson, J.D. Managing Partner, Albertson & Davidson, LLP, San Diego, California
55.6b Forensic Aspects

Dimitry S. Davydow, M.D., M.P.H., F.A.C.L.P. Psychiatrist and Chief Medical Officer, Comprehensive Life Resources, Tacoma, Washington
27.11 Psychiatric Morbidity Following Critical Illness

Mark DeAntonio, M.D. Professor of Clinical Psychiatry, Department of Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, California
54.4 Psychiatric Sequel of HIV and AIDS

Charles DeBattista, D.M.H., M.D. Professor of Psychiatry and Behavioral Sciences, Department of Psychiatry, Stanford University, School of Medicine, Palo Alto, California

33.5a.4 Dopamine-Noradrenaline Reuptake Inhibitors (Bupropion), 33.16 Combination Pharmacotherapy

Lauren N. DeCaporale-Ryan, Ph.D. Associate Professor and Clinical Psychologist, Department of Psychiatry, Medicine, and Surgery, University of Rochester, School of Medicine and Dentistry/Medical Center, Rochester, New York
55.4j Family Intervention and Therapy With Older Adults

Elizabeth A. Deckler, M.D. Resident Physician, Psychiatry and Behavioral Sciences, University of Miami, Leonard M. Miller School of Medicine/Jackson Memorial Hospital, Miami, Florida
47 Tic Disorders

Louisa Degenhardt, Ph.D. Scientia Professor and Deputy Director, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, New South Wales, Australia
11.5 Cannabis-Related Disorders

Andrea Delgado-Peraza, M.D. Attending Physician, Geriatric Psychiatrist, Department of Psychiatry, VA Caribbean Healthcare System, San Juan/Ponce Health Sciences University, School of Medicine, Ponce, Puerto Rico
55.2a Psychiatric Assessment of the Older Patient

Colin A. Depp, Ph.D. Professor, Department of Psychiatry, University of California, School of Medicine, San Diego, California
55.6h Technology for Seniors

Migena Dervishi, M.D. Research Fellow, Department of Psychiatry and Behavioral Sciences, Albert Einstein College of Medicine, New York, New York
33.8a α_2-Adrenergic Receptor Agonists: Clonidine, Clonidine Extended Release, Guanfacine, and Guanfacine Extended Release, 33.8b β-Adrenergic Receptor Antagonists

Abhilash K. Desai, M.D. Adjunct Associate Professor, Psychiatry and Behavioral Sciences, University of Washington, School of Medicine, Seattle, Washington
55.6a Psychiatric Aspects of Long-Term Care

Charles Desfossés, M.D. Clinical Professor of Psychiatry, Department of Psychiatry and Neuroscience, Institut universitaire en santé mentale de Québec/Université Laval, Québec City, Québec, Canada
33.5c Vilazodone and Vortioxetine

Davangere P. Devanand, M.D. Professor of Psychiatry and Neurology, Area Brain Aging and Mental Health, Department of Psychiatry, New York State Psychiatric Institute and Columbia University, Vagelos College of Physicians and Surgeons/Irving Medical Center, New York, New York
55.2a Psychiatric Assessment of the Older Patient

Emanuel DiCicco-Bloom, M.D. Professor of Neuroscience and Cell Biology/Pediatrics, Department of Neuroscience and Cell Biology/Pediatrics, Division of Child Neurology and Neurodevelopmental Disabilities, Rutgers University, Robert Wood Johnson Medical School, Piscataway, New Jersey
1.4 Neural Development and Neurogenesis

Andrea Frances DiMartini, M.D. Professor of Psychiatry and Surgery, Department of Psychiatry, University of Pittsburgh Medical Center/University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania
27.13 Organ Transplantation

Joel E. Dimsdale, M.D. Distinguished Professor Emeritus, Department of Psychiatry, University of California, School of Medicine, San Diego, California
18 Somatic Symptom and Related Disorders

Lisa B. Dixon, M.D., M.P.H. Edna L. Edison Professor of Psychiatry, Department of Psychiatry, Division of Behavioral Health Services and Policy Research, New York State Psychiatric Institute/Columbia University Vagelos College of Physicians and Surgeons, New York, New York

56.4 Mental Health Services Research

Robert Lloyd Doyle, M.A., D.D.S., M.D. Assistant Medical Director, Department of Child and Adolescent Inpatient Unit, McLean Hospital, Brighton, Massachusetts

49.2 Stereotypic Movement Disorders in Children

Robert E. Drake, M.D., Ph.D. Professor of Clinical Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons, New York, New York

12.13 Psychiatric Rehabilitation

Jack Drescher, M.D. Clinical Professor, Department of Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons, New York, New York

21.2 Homosexuality, Gay and Lesbian Identities, and Homosexual Behavior, 21.4 Gender Identity, Gender Diversity, Gender Dysphoria/Incongruence

Steven L. Dubovsky, M.D. Professor and Chair, Department of Psychiatry, University at Buffalo, Jacobs School of Medicine and Biomedical Sciences, Buffalo, New York

33.9a Barbiturates and Similarly Acting Substances, 33.9b Benzodiazepine Receptor Agonists and Antagonists, 33.10a Calcium Channel Inhibitors

Jennifer J. Dunkin, Ph.D. Clinical Professor (Voluntary), Department of Psychiatry and Biobehavioral Sciences, University of California, School of Medicine, San Diego, California

55.2d Psychological Changes With Normal Aging

Walter Dunn, M.D., Ph.D. Assistant Clinical Professor, Department of Psychiatry, University of California, David Geffen School of Medicine, Los Angeles, California
33.2b Dopamine-Serotonin Antagonists, 33.3a Dopamine Partial Agonists

Dino Dvorak, Ph.D. Postdoctoral Fellow; Senior Director, Data Science and Platform Technologies, Center for Neural Science, New York University, Grossman School of Medicine; Gilgamesh Pharmaceuticals Inc., New York, New York
1.23 Basic Systems Neuroscience

Zabrina Ebert, D.O. Chief Resident, Department of Psychiatry and Behavioral Medicine, Medical College of Wisconsin Affiliated Hospitals, Milwaukee, Wisconsin
32.17 Acceptance and Commitment Therapy

Helen Link Egger, M.D. Chief Medical Officer, Little Otter Health; Adjunct Associate Professor at Duke University, School of Medicine/Duke University Medical Center, Department of Psychiatry and Behavioral Sciences, Durham, North Carolina
35.2 Psychiatric Assessment of Preschool Children

Stuart J. Eisendrath, M.D. Professor of Psychiatry Emeritus, Department of Psychiatry, University of California, School of Medicine, San Francisco, California
19 Factitious Disorder

John D. Elsworth, Ph.D. Senior Research Scientist, Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut
1.7 Biogenic Amine Neurotransmitters

Steven A. Epstein, M.D. Physician Executive Director/Professor, Department of Psychiatry, MedStar Health/Georgetown University School of Medicine, Washington, DC

27.1 History and Current Trends

David Erritzoe, M.D., Ph.D., M.R.C.Psych. Clinical Senior Lecturer and Consultant Psychiatrist, Centres for Neuropsychopharmacology and Psychedelic Research, Department of Brain Sciences, Division of Psychiatry, Imperial College London, London, United Kingdom
33.14b MDMA, Psilocybin, Ayahuasca, and Ibogaine

Rodrigo Escalona, M.D. Professor and Attending Psychiatrist, Department of Psychiatry and Behavioral Sciences, University of New Mexico School of Medicine and New Mexico VA Health Care System, Albuquerque, New Mexico
12.2 Phenomenology of Schizophrenia

Javier I. Escobar, M.D., M.Sc. Professor Emeritus, Departments of Psychiatry and Family Medicine, Rutgers University, Robert Wood Johnson Medical School, Piscataway, New Jersey; Professor, Departments of Psychiatry and Health Humanities and Society, Florida International University, Herbert Wertheim College of Medicine, Miami, Florida; Associate Dean for Global Health, Office of Global Health, Rutgers University, Robert Wood Johnson Medical School, Piscataway, New Jersey
9.1 Present and Future of Classification Systems for Mental Disorders, 9.2 The Classification of Mental Disorders in the International Classification of Diseases (ICD-11), 18 Somatic Symptom and Related Disorders

Zelde Espinel, M.D., M.P.H. Assistant Professor of Psychiatry, Psychiatry and Behavioral Sciences, University of Miami Leonard M. Miller School of Medicine, Miami, Florida
33.11b Cholinesterase Inhibitors and Memantine

Spencer Eth, M.D. Professor and Chief of Mental Health, Miami VA Medical Center, University of Miami Leonard M. Miller School of Medicine, Miami, Florida
58.1 Clinical-Legal Issues in Psychiatry

Harris A. Eyre, M.B.B.S., Ph.D. Fellow in Brain Health, Center for Health and Bioscience, Baker Institute for Public Policy, Rice University, Houston, Texas
55.2b Complementary and Integrative Geriatric Psychiatry

Tanya Fabian, Pharm.D., Ph.D. Associate Professor and Director of Pharmacy, Department of Pharmacy and Therapeutics and Psychiatry, University of Pittsburgh, School of Medicine and University of Pittsburgh Medical Center Western Psychiatric Hospital, Pittsburgh, Pennsylvania
10.2 Delirium

Nathan Fairman, M.D., M.P.H. Associate Health Sciences Clinical Professor, Department of Psychiatry and Behavioral Sciences, University of California-Davis School of Medicine, Sacramento, California
55.6c Ethical Issues

Scott C. Fears, M.D., Ph.D. Professor of Clinical Health Sciences, Department of Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
1.19 Gene Mapping Investigations of Psychiatric Disorders

Adriana Feder, M.D. Professor, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York
14.4 Anxiety Disorders: Neurobiology and Neuroscience

Michael Feldmeier, M.D. Psychiatrist, Department of Psychiatry, University of California, David Geffen School of Medicine, Los Angeles, California
54.14 Internet Gaming Disorder in Youth, 54.15 Social Media and Youth

André A. Fenton, Ph.D. Professor of Neural Science, Center for Neural Science, New York University, Grossman School of

Medicine/Faculty of Arts and Science, New York, New York
1.23 Basic Systems Neuroscience

Francisco Fernandez, M.D. Founding Dean, Professor Emeritus, Department of Psychiatry, University of Texas Rio Grande Valley, School of Medicine, Harlingen, Texas
2.10 Neuropsychiatric Aspects of Prion Disease

Prudence W. Fisher, Ph.D. Associate Professor of Clinical Psychiatric Social Work (in Psychiatry)/Research Scientist, Department of Psychiatry/Child & Adolescent Psychiatry, Columbia University, Vagelos College of Physicians & Surgeons/New York State Psychiatric Institute, New York, New York
35.1 Psychiatric Examination of the Infant, Child, and Adolescent

Alastair J. Flint, M.B. Professor of Psychiatry, Department of Psychiatry, University Health Network/University of Toronto, Toronto, Ontario, Canada
55.4c Antianxiety Drugs

Edna B. Foa, Ph.D. Professor, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Raymond and Ruth Perelman School of Medicine, Philadelphia, Pennsylvania
14.6 Anxiety Disorders: Cognitive-Behavioral Therapy

Laura J. Fochtmann, M.D., M.B.I. State University of New York, Distinguished Service Professor, Department of Psychiatry, Pharmacological Sciences, and Biomedical Informatics, Stony Brook University, Renaissance School of Medicine, Stony Brook, New York
7.3 Practice Guidelines in Psychiatry, 12.17 Other Psychotic Disorders

Peter Fonagy, O.B.E., F.Med.Sci., F.B.A., F.Ac.S.S. Professor of Psychoanalysis and Developmental Science, Division of Psychology and Language Sciences, University College London, London, United Kingdom

32.14 Mentalization-Based Treatment

Brent P. Forester, M.D., M.Sc. Chief, Division of Geriatric Psychiatry, McLean Hospital; Senior Medical Director, Value-Based Care Solutions, Population Health Management, Mass General Brigham; Associate Professor of Psychiatry, Harvard University, Harvard Medical School, Boston, Massachusetts
55.3e Geriatric Mood Disorders

Kathryn Forneris, D.O., M.P.H. Resident Physician, Department of Psychiatry, Huntsman Mental Health Institute/ University of Utah, Spencer Fox Eccles School of Medicine, Salt Lake City, Utah
54.12 Impact on Parents of Raising a Child with Psychiatric Illness and/or Developmental Disability

Bradley Freeman, M.D. Associate Professor of Clinical Psychiatry, Department of Psychiatry and Behavioral Sciences, Vanderbilt University, School of Medicine/Medical Center, Nashville, Tennessee
54.2 Child Maltreatment

Wanda P. Fremont, M.D. Professor of Psychiatry and Family Medicine, Department of Psychiatry, State University of New York, Upstate Medical University, Alan and Marlene Norton College of Medicine, Syracuse, New York
54.11 Impact of Terrorism and School Shootings on Children

Paul Frewen, Ph.D., C.Psych. Associate Professor, Department of Psychiatry, Psychology, and Neuroscience, Western University, Canada, London, Ontario, Canada
20 Dissociative Disorders

Paul J. Frick, Ph.D. Professor, Department of Psychology, Louisiana State University, Baton Rouge, Louisiana
45 Disruptive Behavior Disorders in Children and Adolescents

Mark A. Frye, M.D. Psychiatrist, Mayo Clinic, Alix School of Medicine, Rochester, Minnesota
33.10c Carbamazepine, 33.10e Valproate

Kenneth Fung, M.D., F.R.C.P.C., M.Sc. Professor and Director of Global Mental Health, Department of Psychiatry, Toronto Western Hospital, University Health Network/University of Toronto, Toronto, Ontario, Canada
4.2 Transcultural Psychiatry

Margo C. Funk, M.D., M.A., F.A.C.L.P. Vice Chair for Education, Residency Program Director, Director of Cardiovascular Psychiatry, Department of Psychiatry, Brigham and Women's Hospital, Boston, Massachusetts
27.2 Cardiovascular Disorders

Owen Gaasedelen, Ph.D. Neuropsychologist, Independent Practice, Albuquerque, New Mexico
7.4 Neuropsychological Assessment of Adults

Isaac Galatzer-Levy, Ph.D. Research Assistant Professor, Department of Psychiatry, New York University, Grossman School of Medicine, New York, New York
5.2 Statistics and Experimental Design

Martha C. Gamboa, M.D. Assistant Professor of Psychiatry and Behavioral Sciences Department of Psychiatry, New York Medical College/Westchester Medical Center, Valhalla, New York
27.14 Psychiatric Care of the Burned Patient

Michael D. Garrett, M.D. Professor Emeritus of Clinical Psychiatry, Division of Psychiatry, State University of New York, Downstate Health Sciences University, College of Medicine/Medical Center, Brooklyn, New York
32.6 Individual Psychodynamic Psychotherapy and Cognitive-Behavioral Therapy for Psychosis

Robert Garrett, M.D. Adjunct Assistant Professor in Psychiatry, Department of Psychiatry, University of Texas Southwestern Medical School, Dallas, Texas
55.4f Electroconvulsive Therapy and Neurostimulation Treatments

Christina Kitt Garza, M.D. Assistant Professor of Psychiatry, Department of Psychiatry, Columbia University Irving Medical Center/Columbia University Vagelos College of Physicians and Surgeons, New York, New York
32.3 Psychoanalysis and Psychoanalytic Psychotherapy

Jennifer R. Gatchel, M.D., Ph.D. Assistant Professor of Psychiatry, Harvard University, Harvard Medical School, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts
55.3e Geriatric Mood Disorders

Nori Geary, Ph.D. Professor (Retired), Department of Psychiatry, Weill Cornell Medical College, New York, New York
1.17 Basic Science of Appetite

Marie Anne Gebara, M.D. Assistant Professor of Psychiatry, Department of Psychiatry, University of Pittsburgh Medical Center/University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania
10.2 Delirium

Dylan G. Gee, Ph.D. Associate Professor, Departments of Psychology and Psychiatry, Yale University, School of Medicine, New Haven, Connecticut
14.5 Neuroimaging and the Neuroanatomical Circuits Implicated in Anxiety Disorders

Cynthia Geppert, M.D., Ph.D., D.P.S., M.A., M.P.H., M.S.B., M.S.J., F.A.C.L.P., D.F.A.P.A., F.A.S.A.M., H.E.C-C. Senior Ethicist and Director of Ethics Education, VA National Center for Ethics in Health Care; Professor, Departments of Psychiatry and Internal Medicine;

Director of Ethics Education, University of New Mexico School of Medicine; Adjunct Professor of Bioethics, Alden March Bioethics Institute, Albany, New York
25 Adjustment Disorders

Ruth S. Gerson, M.D. Clinical Associate Professor, Department of Child and Adolescent Psychiatry, New York University Grossman School of Medicine, New York, New York
35.4 Psychiatric Emergencies in Children and Adolescents

Omar Ghosn, M.D. Assistant Clinical Professor of Psychiatry, Division of Geriatric Psychiatry, University of California, School of Medicine, San Diego, California
55.3j Alcohol and Substance Abuse in Older Adults

Scott Giacomucci, D.S.W., L.C.S.W., B.C.D., C.G.P., F.A.A.E.T.S., T.E.P. Director, Founder, and Owner; Research Associate and Adjunct Professor, Phoenix Center for Experiential Trauma Therapy, Bryn Mawr College Graduate School of Social Work and Social Research, Philadelphia. Pennsylvania
32.22 Psychodrama, Sociodrama, Sociometry, and Sociatry

Tresha A. Gibbs, M.D. Assistant Clinical Professor of Child and Adolescent Psychiatry, Department of Psychiatry, New York University Grossman School of Medicine, New York, New York
34.2 Child Development

Walter S. Gilliam, Ph.D. Executive Director and Richard D. Holland Presidential Chair in Early Childhood Development, Buffet Early Childhood Institute, University of Nebraska, Omaha, Nebraska
35.1 Psychiatric Examination of the Infant, Child, and Adolescent

Brian Giunta, M.D., Ph.D. Associate Professor, Department of Psychiatry and Behavioral Neurosciences, University of South Florida, Morsani School of Medicine, Tampa, Florida
2.10 Neuropsychiatric Aspects of Prion Disease

Albert Gjedde, M.D., D.Sc. Emeritus Professor of Neurobiology and Pharmacology, Department of Neuroscience, University of Copenhagen, Copenhagen, Denmark
12.9 Molecular Brain Imaging in Schizophrenia

Stephen J. Glatt, Ph.D. Professor, Department of Psychiatry and Behavioral Sciences; Professor, Department of Neuroscience and Physiology; Professor, Department of Public Health and Preventive Medicine; Director, Psychiatric Genetic Epidemiology & Neurobiology Laboratory (PsychGENe Lab), State University of New York, Upstate Medical University, Alan and Marlene Norton College of Medicine, Syracuse, New York
55.2f Genetics of Late-Life Neurodegenerative Disorders

Mary Margaret Gleason, M.D. Kay W. Abiouness Professor of Pediatric Psychiatry, Department of Pediatrics, Eastern Virginia Medical School/Children's Hospital of The King's Daughters, Norfolk, Virginia
49.1 Reactive Attachment Disorder and Disinhibited Social Engagement Disorder

Sam Gnanapragasam, B.Sc., M.B.B.S., M.R.C.Psych. NIHR Academic Clinical Fellow in Psychiatry, Mental Health & Psychological Sciences, King's College London, London, United Kingdom
60 World Aspects of Psychiatry

Irene Gonsalvez, M.D. Associate Psychiatrist, Department of Psychiatry, Brigham and Women's Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
2.9 Neuropsychiatric Aspects of Other Infectious Diseases (Non-HIV) Araceli Gonzalez, Ph.D. Associate Professor, Department of Psychology, California State University, Long Beach, Long Beach, California

51.4 Selective Mutism

Sylvia Gonzalez Cruz, M.D. Psychiatrist, Psychotherapist, Adjunct Faculty, Center for Psychoanalytic Studies, Houston, Texas
6.1 Psychoanalytic Theories and Treatment: Perspective from North America and Great Britain

Jessica Goodman, Ph.D. Senior Instructor, Department of Psychiatry, University of Rochester, School of Medicine and Dentistry, Rochester, New York
55.4j Family Intervention and Therapy With Older Adults

Hilary Gould, Ph.D. Assistant Clinical Professor, Department of Psychiatry, University of California, School of Medicine, San Diego, California
55.3j Alcohol and Substance Abuse in Older Adults, 55.5 Financial Issues in the Delivery of Geriatric Psychiatric Care

Eric Granholm, Ph.D. Professor of Psychiatry, Department of Psychiatry, University of California, School of Medicine, San Diego, California
55.4i Cognitive-Behavioral Therapy

John A. Gray, M.D., Ph.D. Associate Professor of Neurology, Center for Neuroscience, University of California-Davis, School of Medicine, Sacramento, California
1.11 Intraneuronal Signaling

Julie Graziane, M.D. Assistant Professor, Department of Psychiatry and Behavioral Health, Pennsylvania Psychiatric Institute/Penn State University, College of Medicine, Hershey, Pennsylvania
10.3 Major Neurocognitive Disorders (Dementia)

Benjamin D. Greenberg, M.D., Ph.D. Professor, Department of Psychiatry and Human Behavior, Butler Hospital/VA Providence Healthcare System/Brown University, Warren Alpert Medical School, Providence, Rhode Island
33.20c Neurosurgical Treatments

Harvey Roy Greenberg, M.D. Clinical Professor of Psychiatry, Albert Einstein College of Medicine at Yeshiva University, Bronx, New York.
16 Pathological and Problem Gambling: Gambling Disorder

Mindy Greenstein, Ph.D. Director, Psychological Consultation, MaxFint, Inc., New York, New York
27.9 End-of-Life and Palliative Care

Paul Richard Greenwood, L.L.B. (Hons.) Deputy District Attorney San Diego (Retired); CEO, Greenwood Law Corp, San Diego, California
55.6e The Legal and Ethical Analysis of Older Adult Abuse

George T. Grossberg, M.D. Professor and Director, Department of Psychiatry and Behavioral Neuroscience, Division of Geriatric Psychiatry, Saint Louis University, School of Medicine, St. Louis, Missouri
55.6a Psychiatric Aspects of Long-Term Care

Sandeep Grover, M.B.B.S., M.D. Professor of Psychiatry, Department of Psychiatry, Post Graduate Institute of Medical Education and Research, Chandigarh, Chandigarh, India
33.19c Biologic Therapies in India

Gerhard Gründer, M.D. Professor of Psychiatry, Department for Molecular Neuroimaging, Central Institute of Mental Health, Mannheim, Germany
12.9 Molecular Brain Imaging in Schizophrenia

Adrienne Grzenda, M.D., Ph.D., M.S. Health Sciences Clinical Assistant Professor, Department of Psychiatry and Biobehavioral Sciences, University of California, David Geffen School of Medicine, Los Angeles, California
5.1 Quantitative and Experimental Methods in Psychiatry: Epidemiology

Christina Guest, M.D. Los Angeles County and University of Southern California, Child and Adolescent Psychiatry Fellow, Division of Psychiatry and Behavioral Science, University of Southern California, Keck School of Medicine, Los Angeles, California
54.15 Social Media and Youth

Rachel Guetta, B.A. Clinical Psychology Ph.D. Candidate, Department of Psychology and Neuroscience, Duke University, School of Medicine, Durham, North Carolina
32.11 Dialectical Behavior Therapy

Rishab Gupta, M.B.B.S., M.D. Associate Psychiatrist/Instructor in Psychiatry, Department of Psychiatry, Brigham & Women's Faulkner Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
2.9 Neuropsychiatric Aspects of Other Infectious Diseases (Non-HIV)

Raquel E. Gur, M.D., Ph.D. Professor, Department of Psychiatry, University of Pennsylvania, Raymond and Ruth Perelman School of Medicine, Philadelphia, Pennsylvania
12.8 Functional Brain Imaging in Schizophrenia

Ruben C. Gur, Ph.D. Professor, Psychiatry, University of Pennsylvania, Raymond and Ruth Perelman School of Medicine, Philadelphia, Pennsylvania
12.8 Functional Brain Imaging in Schizophrenia

Oye Gureje, Ph.D., D.Sc., F.R.C.Psych., F.W.A.C.P. Professor of Psychiatry, WHO Collaborating Centre for Research and Training in Mental Health, Neuroscience and Substance Abuse, Department of Psychiatry, College of Medicine, University of Ibadan, Ibadan, Nigeria
33.19b Biologic Therapies in Africa

Debra A. Gusnard, M.D. Retired, Department of Radiology and Psychiatry, Washington University in St. Louis, School of Medicine, St. Louis, Missouri
1.29 Neuropsychiatry of Self

Robert W. Guynn, M.D. Professor Emeritus, Department of Psychiatry and Behavioral Sciences, University of Texas Health Science Center, McGovern Medical School, Houston, Texas
32.10 Interpersonal Psychotherapy

Kathleen Y. Haaland, Ph.D., A.B.P.P.-C.N. Professor, Department of Psychiatry and Behavioral Sciences, University of New Mexico, School of Medicine/Health Sciences Center, Albuquerque, New Mexico
7.4 Neuropsychological Assessment of Adults

Habibolah Khazaie, M.D. Clinical Professor of Psychiatry, Department of Psychiatry, Division of Sleep Disorders Research Center, Farabi Hospital/Kermanshah University of Medical Sciences, Kermanshah, Iran
23 Sleep Disorders

Saria El Haddad, M.D. Instructor, Department of Psychiatry, Harvard University, Harvard Medical school, Boston, Massachusetts
33.17c Pharmacologic Treatment of Sexual Dysfunction

Danella Hafeman, M.D., Ph.D. Assistant Professor, Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania
50.2 Early-Onset Bipolar Disorder

Nekisa Haghighat, M.D., M.P.H. Resident Physician, Department of Psychiatry and Neuroscience, University of California, School of Medicine, Riverside, California
33.7a Pimavanserin

Wayne D. Hall, Ph.D. Emeritus Professor, The National Centre for Youth Substance Use Research, The University of Queensland, Queensland, Australia
11.5 Cannabis-Related Disorders

Michael Hann, M.D., M.B.A., M.S., C.P.E., F.A.C.H.E., F.A.P.A. Chief Medical Officer, Department of Psychiatry, Brook Lane Health Services, Hagerstown, Maryland
55.3j Alcohol and Substance Abuse in Older Adults, 55.5 Financial Issues in the Delivery of Geriatric Psychiatric Care

Helena B. Hansen, M.D., Ph.D. Professor of Psychiatry, Department of Psychiatry, University of California, David Geffen School of Medicine, Los Angeles, California
7.10 Cultural Competency in the Organization and Delivery of Psychiatric Care

Jessica Harder, M.D. Neuropsychiatrist, Department of Psychiatry, Brigham and Women’s Hospital, Boston, Massachusetts
2.9 Neuropsychiatric Aspects of Other Infectious Diseases (Non-HIV)

Ebrahim Haroon, M.D. Associate Professor, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia
1.13 Immune–Brain Interactions in Psychiatry

Ansar M. Haroun, M.D. (Retired) Clinical Professor of Psychiatry, and Pediatrics and Adjunct Professor of Law, Departments of Psychiatry, Pediatrics, and Law, University of California, School of Medicine and University of San Diego School of Law; Supervising Psychiatrist, Department of Forensic Psychiatry, San Diego Superior Court, San Diego, California
55.6b Forensic Aspects, 55.6e The Legal and Ethical Analysis of Older Adult Abuse

Philip D. Harvey, Ph.D. Leonard M. Miller Professor of Psychiatry and Behavioral Sciences, Department of Behavioral Sciences, University of Miami Leonard M. Miller School of Medicine, Miami, Florida
12.10 Neurocognition in Schizophrenia, 33.2a Dopamine Antagonists

Lily T. Hechtman, M.D., F.R.C.P., A.B.P.N. Professor of Psychiatry and Pediatrics, Division of Child Psychiatry, Montreal Children's Hospital, McGill University Health Center, McGill University Medical School Montreal, Quebec, Canada
44.1 Attention-Deficit/Hyperactivity Disorder

Nicholas D. Henkel, Ph.D. Graduate Research Assistant, Department of Neurosciences, University of Toledo, College of Medicine and Life Sciences, Toledo, Ohio
12.6 Cellular and Molecular Neuropathology of Schizophrenia

Robert Hindman, Ph.D. Faculty, Beck Institute, Bala Cynwyd, Pennsylvania
32.9 Cognitive Behavior Therapy

Max Hirshkowitz, Ph.D. Retired Full Professor, Department of Medicine, Baylor University, College of Medicine, Houston, Texas
23 Sleep Disorders

J. Gregory Hobelmann, M.D., M.P.H. CO-CEO, Ashley Addiction Treatment, Havre De Grace, Maryland
27.15 Management of Chronic Pain

Gil D. Hoftman, M.D., Ph.D. Assistant Professor of Psychiatry, Division of Child and Adolescent Psychiatry, Resnick Neuropsychiatric Hospital; Assistant Professor of Psychiatry and Biobehavioral Sciences, Division of Child and Adolescent Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human

Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
36 Genetics in Child Psychiatry

Eric Hollander, M.D. Professor of Psychiatry and Behavioral Sciences, Department of Psychiatry, Albert Einstein College of Medicine, New York, New York
33.8a α_2*-Adrenergic Receptor Agonists: Clonidine, Clonidine Extended Release, Guanfacine, and Guanfacine Extended Release, 33.8b β-Adrenergic Receptor Antagonists*

Lindsay E. Holly, Ph.D. Assistant Professor, Department of Psychology, Marquette University, Milwaukee, Wisconsin
32.24 Evaluating Psychotherapy: Research Methods, Findings, and Future Directions

Paul E. Holtzheimer, M.D., M.S. Professor, Department of Psychiatry and Surgery, Dartmouth College, Geisel School of Medicine, Hanover, New Hampshire
33.5d Monoamine Oxidase Inhibitors

Vivian Hook, Ph.D. Distinguished Professor of Pharmaceutical Sciences, Pharmacology, and Neurosciences, Division of Pharmaceutical Sciences, University of California, School of Medicine, San Diego, California
55.4d Antipsychotic Drugs

Dina Hooshyar, M.D., M.P.H. Director, Veterans Health Administration National Center on Homelessness Among Veterans; Associate Professor, Department of Psychiatry, University of Texas Southwestern, Medical School, Dallas, Texas
12.15 Recovery in Schizophrenia

Rory P. Houghtalen, M.D. Clinical Professor of Psychiatry, Department of Psychiatry, University of Rochester, School of Medicine and Dentistry, Rochester, New York

7.1 Psychiatric Interview, History, and Mental Status Examination of the Adult Patient

Michael F. Hoyt, Ph.D. Independent Practice, Mill Valley, California
32.13 Single-Session Therapy

Jeffrey Hsu, M.D. Assistant Professor, Department of Psychiatry and Behavioral Sciences, The Johns Hopkins Hospital/Johns Hopkins University School of Medicine, Baltimore, Maryland
2.8 Neuropsychiatric Aspects of HIV Infection and AIDS

Jonathan D. Huppert, Ph.D. Professor of Psychology, Sam and Helen Beber Chair of Clinical Psychology, Department of Psychology, The Hebrew University of Jersualem, Jerusalem, Israel
14.6 Anxiety Disorders: Cognitive-Behavioral Therapy

Mustafa M. Husain, M.D. Professor of Psychiatry, Neurology and Bio-medical Engineering, Fellow, Peter O'Donnell Brain Institute (POBI), University of Texas Southwestern, Medical Center, Dallas, Texas; Adjunct Professor of Psychiatry and Neurology, Member, Duke Institute of Brain Sciences (DIBS), Duke University, School of Medicine, Durham, North Carolina
55.4f Electroconvulsive Therapy and Neurostimulation Treatments

Kumail Hussain, M.D. Child and Adult Psychiatrist, Clarity Clinic Arlington Heights, PLLC, Arlington Heights, Illinois
33.2c Dopamine Blockers and Neuromotor Adverse Effects

Philippe Hwang, M.B., B.Ch., B.A.O., M.Sc. Psychiatry Resident Physician, Department of Psychiatry, University of Ottawa, The Ottawa Hospital/Children's Hospital of Eastern Ontario/Royal Ottawa Mental Health Centre, Ottawa, Ontario, Canada
44.1 Attention-Deficit/Hyperactivity Disorder

Brian Iacoviello, Ph.D. Assistant Professor, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New

York
14.4 Anxiety Disorders: Neurobiology and Neuroscience

Rocco A. Iannucci, M.D. Instructor in Psychiatry, Division of Alcohol, Drugs and Addiction, McLean Hospital/Harvard University, Harvard Medical School, Belmont, Massachusetts
11.3 Stimulant-Related Disorders

Roya Ijadi-Maghsoodi, M.D., M.S.H.P.M. Assistant Professor-in-Residence, Department of Psychiatry and Biobehavioral Sciences, University of California, David Geffen School of Medicine, Los Angeles, California
54.8 School Consultation and Collaborating With the Educational System

Evan Iliakis, B.A., M.D./Ph.D. Candidate, Department of Neuroscience, University of Pennsylvania, Raymond and Ruth Perelman School of Medicine, Philadelphia, Pennsylvania
32.7 Treatment of Borderline Personality Disorder

Ali S. Imami, M.B.B.S., M.S. Graduate Research Assistant, Department of Neurosciences, University of Toledo, College of Medicine and Life Sciences, Toledo, Ohio
12.6 Cellular and Molecular Neuropathology of Schizophrenia

Brian Isakson, Ph.D. Professor, Department of Psychiatry and Behavioral Sciences, University of New Mexico School of Medicine, Albuquerque, New Mexico
11.8 Tobacco-Related Disorders

Anna Ivanenko, M.D., Ph.D. Professor of Clinical Psychiatry and Behavioral Sciences, Division of Child and Adolescent Psychiatry, Ann and Robert H. Lurie Children’s Hospital of Chicago/Northwestern University, Feinberg School of Medicine, Chicago, Illinois
54.13 Pediatric Sleep Disorders

Elena I. Ivleva, M.D., Ph.D. Associate Professor of Psychiatry, Director of Early Psychosis Clinical and Research Program, Department of Psychiatry, University of Texas Southwestern, Medical School, Dallas, Texas
12.16 Psychosis as a Defining Dimension in Schizophrenia

Julienne Jacobson, M.D. Clinical Associate Professor of Psychiatry and Behavioral Sciences and Pediatrics, Psychiatry and the Behavioral Sciences, Children's Hospital Los Angeles/University of Southern California, Keck School of Medicine, Los Angeles, California
54.3 Children's Reaction to Illness and Hospitalization

Robert J. Jaffe, M.D. Assistant Professor, Departments of Psychiatry and Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York
34.2 Child Development

Neeta Jain, M.D. Clinical Assistant Professor of Psychiatry, Division of Psychiatry and Psychology, Mayo Clinic of Arizona/Mayo Clinic, Alix School of Medicine, Scottsdale, Arizona
30.8 Physician and Medical Student Mental Health

Jimmy E. Jakobsson, Ph.D. Postdoctoral Research Associate in Radiology, Mallinckrodt Institute of Radiology, Washington University in St. Louis, School of Medicine, St. Louis, Missouri
12.9 Molecular Brain Imaging in Schizophrenia

Philip G. Janicak, M.D. Adjunct Professor of Psychiatry, Department of Psychiatry and Behavioral Sciences, Northwestern University, Feinberg School of Medicine, Chicago, Illinois
33.2c Dopamine Blockers and Neuromotor Adverse Effects

Oluwole O. Jegede, M.D., M.P.H. Assistant Professor of Psychiatry, Division of Addiction Psychiatry, Yale University School of Medicine, New Haven, Connecticut

33.13a.2 Opioid Receptor Antagonists: Naltrexone and Nalmefene

Dilip V. Jeste, M.D. Distinguished Professor, Department of Psychiatry and Neurosciences, University of California, School of Medicine, San Diego, California
55.1 Introduction to Geriatric Psychiatry, 55.6i Positive Psychiatry of Aging

Dylan J. Jester, Ph.D., M.P.H. Biostatistician, War Related Illness and Injury Study Center (WRIISC), VA Palo Alto, Palo Alto, California
55.1 Introduction to Geriatric Psychiatry

Jason K. Johannesen, Ph.D. Assistant Professor, Adjunct, Department of Psychiatry, Yale University, School of Medicine, New Haven, Connecticut
1.26 Electroencephalography in Psychiatry

Melissa B. Jones, M.D. Assistant Professor, Menninger Department of Psychiatry and Behavioral Sciences, Baylor University, College of Medicine, Houston, Texas
2.5 Neuropsychiatric Consequences of Traumatic Brain Injury

Ricardo E. Jorge, M.D. Professor, Beth K. and Stuart C. Chair in Brain Injury Medicine, Department of Psychiatry and Behavioral Science, Baylor University, College of Medicine, Houston, Texas
2.5 Neuropsychiatric Consequences of Traumatic Brain Injury

Laura M. Juliano, Ph.D. Professor, Department of Psychology, American University, Washington, DC
11.4 Caffeine-Related Disorders

Rahil R. Jummani, M.D. Clinical Associate Professor of Child and Adolescent Psychiatry, Department of Child and Adolescent Psychiatry, Child Study Center/Hassenfeld Children's Hospital at NYU Langone/New York University, Grossman School of Medicine, New York, New York

47 Tic Disorders

Alexandra Junewicz, M.D. Clinical Assistant Professor, Department of Child and Adolescent Psychiatry, New York University, Grossman School of Medicine, New York, New York
35.4 Psychiatric Emergencies in Children and Adolescents

Arie Kaffman, M.D., Ph.D. Assistant Clinical Professor of Psychiatry, Department of Psychiatry, Newington VA/Yale University, School of Medicine, New Haven, Connecticut
1.20 Epigenetics in Psychiatry: The Promise for New Biomarkers and Treatments

Sudhir Kakar, Ph.D. President, Delhi Institute of Psychoanalysis, Delhi, India
6.2 Erik H. Erikson

Amir Kalali, M.D. Chairman, CNS Summit, San Diego, California
33.1b Drug Development and Approval Process in the United States

Gen Kanayama, M.D., Ph.D. Research Associate in Psychiatry, Department of Psychiatry, McLean Hospital/Harvard University, Harvard Medical School, Belmont, Massachusetts
11.11 Anabolic-Androgenic Steroid-Related Disorders

John M. Kane, M.D. Professor of Psychiatry and Molecular Medicine, Department of Psychiatry, Zucker Hillside Hospital, Hofstra University/Northwell Health, Donald and Barbara Zucker School of Medicine, Hempstead, New York
12.12 Schizophrenia: Pharmacologic Treatment

Tuana Kant, M.Sc. Clinical Research Analyst, Molecular Brain Science, Neurogenetics Department, Centre for Addiction and Mental Health, University of Toronto, Toronto, Ontario, Canada
42.1 Language Disorder, 42.3 Speech Sound Disorder, 42.4 Stuttering

Adam I. Kaplin, M.D., Ph.D. Adjunct Faculty, Departments of Psychiatry and Neurology, Johns Hopkins University, School of Medicine, Baltimore, Maryland
1.9 Novel Neurotransmitters

Shweta Kapoor, M.D., Ph.D. Senior Associate Consultant and Assistant Professor of Psychiatry, Department of Psychiatry and Psychology, Mayo Clinic/Mayo Clinic, Alix School of Medicine, Scottsdale, Arizona
30.7 Psychiatry and Spirituality

Helmet T. Karim, Ph.D. Assistant Professor of Psychiatry and Bioengineering, Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania
55.3d Anxiety Disorders in Older Adults

Ethan F. Karsen, M.D. Staff Psychiatrist, Department of Psychiatry and Behavioral Sciences, NorthShore University Health System, Evanston, Illinois
33.5d Monoamine Oxidase Inhibitors

Mirelle Kass, B.A. Research Projects Coordinator, Center for the Developing Brain, Child Mind Institute, New York, New York
42.2 Social (Pragmatic) Communication Disorder

Sheryl Kataoka, M.D., M.S.H.S. Professor Emeritus, Department of Psychiatry, Division of Population Behavioral Health, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
54.8 School Consultation and Collaborating With the Educational System

Jeffrey W. Katzman, M.D. Professor, University of New Mexico School of Medicine, Department of Psychiatry, University of New Mexico Health Science Center, Albuquerque, New Mexico; Adjunct

Professor, Department of Psychiatry, Yale University, School of Medicine, Silver Hill Hospital, New Canaan, Connecticut
25 Adjustment Disorders

M. Sagman Kayatekin, M.D. Medical Director, Professionals Program, Menninger Department of Psychiatry and Behavioral Sciences, Baylor University, College of Medicine, The Menninger Clinic, Houston, Texas
57.2 Examining Professional Patients

David L. Kaye, M.D. Professor of Psychiatry, Department of Psychiatry, University at Buffalo, Jacobs School of Medicine and Biomedical Sciences, Buffalo, New York
53.1 Psychodynamic Psychotherapy (Child Psychiatry)

George Alan Keepers, M.D. Carruthers Professor and Chair of Psychiatry, Department of Psychiatry, Oregon Health & Science University, School of Medicine, Portland, Oregon
57.1 Graduate Psychiatric Education

Samuel J. Keith, M.D. Milton Rosenbaum Professor of Psychiatry, Emeritus, Department of Psychiatry, University of New Mexico School of Medicine, Albuquerque, New Mexico
12.2 Phenomenology of Schizophrenia

Kimberley M. Kendall, B.Sc. (Hons.), M.B.B.Ch., Ph.D., M.R.C.Psych. Wales Clinical Academic Track Psychiatrist, Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, United Kingdom
12.4 Genetics of Schizophrenia

Alvin Keng, M.D., F.R.C.P.C. Assistant Professor, Division of Geriatric Psychiatry, Department of Psychiatry, Baycrest Centre for Geriatric Care/University of Toronto, Toronto, Ontario, Canada
55.4k Group Therapy

James L. Kennedy, M.D., F.R.C.P.C., F.R.S.C. Professor, Department of Psychiatry; Department of Molecular Sciences, Centre for Addiction and Mental Health, University of Toronto, Toronto, Ontario, Canada

42.1 Language Disorder, 42.3 Speech Sound Disorder, 42.4 Stuttering

Sidney H. Kennedy, M.D., F.R.C.P.C., F.R.C.Psych. Professor, Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

13.6 Mood Disorders: Pharmacologic Treatment of Depression and Bipolar Disorders

Paul J. Kenny, Ph.D. Ward-Coleman Professor and Chair, Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York

1.22 Animal Models in Psychiatry

Otto F. Kernberg, M.D. Emeritus Professor of Psychiatry, Weill Cornell Medical College, Westchester Division, New York/Presbyterian Hospital, White Plains, New York

3.6 Neuroscience of Psychoanalytic Object Relations Theory

Amir A. Khan, M.D. Staff Psychiatrist/Behavioral Health Lead, Department of Behavioral Health, Corporal Michael J. Crescenz Department of Veterans Affairs Medical Center, Philadelphia, Pennsylvania

33.6b Nefazodone

Ravenna Khanna, M.D. Resident Physician, Department of Dermatology, Howard University College of Medicine, Washington, DC

27.12 Psychocutaneous Disorder

Nayla M. Khoury, M.D., M.P.H. Assistant Professor/Chief of Child and Adolescent Psychiatry Outpatient Clinic, Department of

Psychiatry, State University of New York, Upstate Medical University, Alan and Marlene Norton College of Medicine, Syracuse, New York
54.11 Impact of Terrorism and School Shootings on Children

Jihye Kim, M.D. Instructor, Department of Psychiatry, Weill Cornell Medical College and New York-Presbyterian/Weill Cornell Medical Center, New York, New York
1.10 Neurotrophic Factors

Soo-Jeong Kim, M.D. Associate Professor of Psychiatry and Behavioral Sciences, Division of Child and Adolescent Psychiatry, Seattle Children's Autism Center/University of Washington, School of Medicine, Seattle, Washington
39 Intellectual Disability

Young Shin Kim, M.D., M.S., M.P.H., Ph.D. Professor, Department of Psychiatry and Behavioral Sciences, University of California, School of Medicine, San Francisco, California
42.2 Social (Pragmatic) Communication Disorder

Ryan J. Kimmel, M.D. Professor, Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington
14.2 Clinical Features of the Anxiety Disorders

Eva R. Kimonis, Ph.D., M.S. Professor of Psychology, School of Psychology, University of New South Wales, Sydney, New South Wales, Australia
45 Disruptive Behavior Disorders in Children and Adolescents

Robert A. King, M.D. Professor of Child Psychiatry, Yale Child Study Center, Yale University School of Medicine, New Haven, Connecticut
35.1 Psychiatric Examination of the Infant, Child, and Adolescent

Laurence J. Kirmayer, M.D. James McGill Professor, Division of Social and Transcultural Psychiatry, McGill University, Montreal, Quebec, Canada
29.1 Cultural Concepts of Distress

Elizabeth R. Kitt, M.S., M.Phil. Graduate Student, Department of Psychology, Yale University, School of Medicine, New Haven, Connecticut
14.5 Neuroimaging and the Neuroanatomical Circuits Implicated in Anxiety Disorders

Cary Kogan, Ph.D., M.S.C.P. Professor of Clinical Psychology, School of Psychology, University of Ottawa, Ottawa, Ontario, Canada
33.5b $5HT_{1A}$ *Receptor Partial Agonists (Buspirone)*

Steve Koh, M.D., M.P.H., M.B.A. Chief, General Psychiatry Division; Clinical Professor, Department of Psychiatry, University of California, School of Medicine, San Diego, California
55.3j Alcohol and Substance Abuse in Older Adults, 55.5 Financial Issues in the Delivery of Geriatric Psychiatric Care

Maulika Kohli, M.S. Doctoral Candidate, Department of Psychiatry, San Diego State University/University of California, School of Medicine, Joint Doctoral Program in Clinical Psychology, San Diego, California
55.3a Assessment of Functioning

Robert Kohn, M.D. Professor, Practice of Epidemiology, Brown University, Providence, Rhode Island
4.2 Transcultural Psychiatry

Brandon A. Kohrt, M.D., Ph.D. Professor of Psychiatry, Global Health, and Anthropology, Center for Global Mental Health Equity, George Washington University School of Medicine and Health Sciences, Washington, DC

29.1 Cultural Concepts of Distress

Alexander Kolevzon, M.D. Professor of Psychiatry and Pediatrics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York
34.2 Child Development

Susan G. Kornstein, M.D. Professor of Psychiatry and Obstetrics and Gynecology; Executive Director, Institute for Women's Health, Virginia Commonwealth University School of Medicine, Richmond, Virginia
33.6a Trazodone, 33.6b Nefazodone

Anna Moukouri Kouoh, C.R.N.P. Nurse Practitioner, Psychiatry Department, Johns Hopkins Hospital/John Hopkins University, School of Medicine, Baltimore, Maryland
2.8 Neuropsychiatric Aspects of HIV Infection and AIDS

Emiko Koyama, Ph.D. Postdoctoral Research Fellow/Technical Staff I, Molecular Brain Science/Molecular Pathology of Psychiatric Disorders, Centre for Addiction and Mental Health, University of Toronto/Center for Brain Science, RIKEN, Toronto, Ontario, Canada; Wakō-shi, Saitama, Japan
42.1 Language Disorder, 42.3 Speech Sound Disorder, 42.4 Stuttering

Jennifer L. Kruse, M.D. Associate Professor of Psychiatry, Department of Psychiatry and Biobehavioral Sciences, University of California, David Geffen School of Medicine, Los Angeles, California
7.6 Medical Assessment and Laboratory Testing in Psychiatry

John H. Krystal, M.D. Robert L. McNeil, Jr., Professor of Translational Research, Professor of Psychiatry, Neuroscience, and Psychology; Chair, Department of Psychiatry; Co-Director, Yale Center for Clinical Investigation, Yale University, School of Medicine;

Chief of Psychiatry and Behavioral Health, Yale-New Haven Hospital, New Haven, Connecticut
1.1 The Neuroscience of Psychiatry, 1.25 Nuclear Magnetic Resonance Imaging and Spectroscopy: Basic Principles and Recent Findings in Neuropsychiatric Disorders, 1.27 Radiotracer Imaging with Positron Emission Tomography and Single Photon Emission Computed Tomography: Fundamental Principles, Methodology and Role in Neuropsychiatric Research

Marek Kubicki, M.D., Ph.D. Professor of Psychiatry and Radiology, Psychiatry and Radiology, Brigham and Women's Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
12.7 Structural Brain Imaging in Schizophrenia

Helen H. Kyomen, M.D., M.S. Clinical Assistant Professor, Department of Psychiatry, Boston University, Aram V Chobanian and Edward Avedisian School of Medicine; Adjunct Clinical Assistant Professor, Department of Psychiatry, Tufts University, School of Medicine; Lecturer, Part-Time, Department of Psychiatry, Harvard University, Harvard Medical School, Boston, Massachusetts
55.5 Financial Issues in the Delivery of Geriatric Psychiatric Care, 55.6d Gender Issues

Beth Ann Labardi, M.D. Clinical Assistant Professor of Psychiatry, Psychiatry and Behavioral Sciences, VA Puget Sound Health Care System/University of Washington School of Medicine, Seattle, Washington
55.3b Psychiatric Problems in the Medically Ill Geriatric Patient

Simon Lafrenière, M.D., M.Sc. Psychiatrist/Clinical instructor of psychiatry, Division of Psychiatry and Neurosciences, Institut universitaire en santé mentale de Québec /Université Laval, Quebec City, Quebec, Canada
33.5a.1 Serotonin Reuptake Inhibitors (SRIs)

James H. Lake, M.D. Adjunct Fellow, National Institute of Complementary Medicine, Health Research Institute, Western Sydney University, Sydney, Australia
30.4 Complementary, Alternative, and Integrative Approaches in Mental Health Care

H. Richard Lamb, M.D. Professor Emeritus of Psychiatry and the Behavioral Sciences, University of Southern California, Keck School of Medicine, Los Angeles, California
56.8 Criminalization of Persons With Serious Mental Illness

Eugene Laska, Ph.D. Research Professor, Department of Psychiatry and Division of Biostatistics, Department of Population Health, New York University Grossman School of Medicine, New York, New York
5.2 Statistics and Experimental Design

Tarannum Lateef, M.D., M.P.H. Associate Professor of Pediatrics and Neurology, Division of Neurology, George Washington University, School of Medicine and Health Sciences, Washington, DC; Pediatric Specialists of Virginia and Children's National Health System, Woodbridge, Virginia
2.11 Neuropsychiatric Aspects of Headache

Megan E. Lavery, Psy.D. Bariatric Psychologist, Department of Bariatric Surgery/Embedded Psychology, Christiana Care, Wilmington, Delaware
27.4 Psychosomatic Medicine: Obesity

Helen Lavretsky, M.D., M.S. Professor In-Residence of Psychiatry, Department of Psychiatry and Biobehavioral Sciences, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
55.2b Complementary and Integrative Geriatric Psychiatry

Lawrence W. Lazarus, M.D. Private Practice, Adult, Geriatric and Forensic Psychiatry, Santa Fe, New Mexico, San Francisco, California
55.4h Individual Psychotherapy

Ellen E. Lee, M.D. Assistant Professor and Division Chief of Geriatric Psychiatry, Psychiatry, University of California, School of Medicine, San Diego, California
55.1 Introduction to Geriatric Psychiatry

Francis S. Lee, M.D., Ph.D. Chair and Professor, Department of Psychiatry, Weill Cornell Medical College and New York-Presbyterian/Weill Cornell Medical Center, New York, New York
1.10 Neurotrophic Factors

Kelly C. Lee, Pharm.D., M.A.S., B.C.P.P. Professor of Clinical Pharmacy, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, School of Medicine, San Diego, California
55.4d Antipsychotic Drugs

Susan W. Lehmann, M.D. Professor of Psychiatry and Behavioral Sciences, Division of Geriatric Psychiatry and Neuropsychiatry, Johns Hopkins University School of Medicine, Baltimore, Maryland
55.3g Delirium

Sean R. LeNoue, M.D. Medical Director of Outpatient Services, Rogers Behavioral Health, Nashville, Tennessee
54.5 Adolescent Substance Use Disorders

Gregory B. Leong, M.D. Clinical Professor of Psychiatry, Department of Psychiatry and the Behavioral Sciences, University of Southern California, Keck School of Medicine, Los Angeles, California
58.1 Clinical-Legal Issues in Psychiatry

Bernard Lerer, M.D. Professor, Department of Psychiatry, Hadassah Medical Center, Hebrew University, Jerusalem, Israel
33.15b Thyroid Hormones

Alan J. Lerner, M.D. Director, Brain Health and Memory Center, Neurology Department, University Hospitals Cleveland Medical Center, Cleveland, Ohio
2.12 Neuropsychiatric Aspects of Neuromuscular Diseases

Nigel Lester, M.D. Director of Mental Health, Palm Health; Professor of Wellbeing, Anthropedia Foundation, St. Louis, Missouri
26 Personality Disorders

Janni Leung, Ph.D. Research Fellow of Epidemiology, National Centre for Youth Substance Use Research, The University of Queensland, Australia
11.5 Cannabis-Related Disorders

Bennett L. Leventhal, M.D. Irving B. Harris Professor of Child and Adolescent Psychiatry, Emeritus, University of Chicago, Division of Biological Sciences, Pritzker School of Medicine, Chicago, Illinois
42.2 Social (Pragmatic) Communication Disorder

Adam B. Lewin, Ph.D., A.B.P.P. Professor and Division Chief, Pediatric Neuropsychiatry, University of South Florida, Morsani College of Medicine, Tampa, Florida
51.1 Obsessive-Compulsive Disorder in Childhood

Bradley Lewis, M.D., Ph.D. Associate Professor of Health Humanities and Cultural Studies, Gallatin School of Individualized Study, New York University, Grossman School of Medicine, New York, New York
32.15 Narrative Psychiatry

David A. Lewis, M.D. Chair of Psychiatry and Distinguished Professor of Psychiatry and Neuroscience, Department of

Psychiatry, University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania
1.2 Functional Neuroanatomy

Dorothy Otnow Lewis, M.D., F.A.C.P. Retired Clinical Professor of Psychiatry, Yale Child Study Center, Yale University, School of Medicine, New Haven, Connecticut
28.2 Adult Antisocial Behavior, Criminality, and Violence

Stephen F. Lewis, B.S., M.D. Professor, Staff Psychiatrist, Department of Psychiatry and Behavioral Sciences, University of New Mexico School of Medicine, Albuquerque, New Mexico
12.2 Phenomenology of Schizophrenia

Roberto Lewis-Fernández, M.D., M.T.S. Professor of Clinical Psychiatry at Columbia University, Vagelos College of Physicians and Surgeons; Director of the New York State Center of Excellence for Cultural Competence; Director of the Hispanic Treatment Program; and Research Area Leader for Anxiety, Mood, Eating and Related Disorders at New York State Psychiatric Institute, Psychiatry (Columbia) and Anxiety, Mood, Eating and Related Disorders (NYS Psychiatric Institute), Columbia University, Vagelos College of Physicians and Surgeons and NYS Psychiatric Institute, New York, New York
20 Dissociative Disorders, 29.1 Cultural Concepts of Distress, 29.2 Cultural Assessments in Psychiatry

Julio Licinio, M.D., Ph.D., M.B.A., M.S. State University of New York, Distinguished Professor, Department of Psychiatry and Behavioral Sciences, State University of New York, Upstate Medical University, Alan and Marlene Norton College of Medicine, Syracuse, New York
1.21 Pharmacogenetics

Yeun J. Lim, M.D. Resident Physician, Department of Psychiatry, University of California, David Geffen School of Medicine, Olive View

Psychiatry Residency Program, Los Angeles, California
54.15 Social Media and Youth

Christine Lochner, Ph.D. Professor of Clinical Psychology, SAMRC Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry, Stellenbosch University, Cape Town, South Africa
15 Obsessive-Compulsive and Related Disorders

Richard J. Loewenstein, M.D. Adjunct Professor, Department of Psychiatry, University of Maryland School of Medicine, Baltimore, Maryland
20 Dissociative Disorders

John Lotfi, M.D., J.D. Resident Physician, Department of Psychiatry, Mather Hospital, Port Jefferson, New York
1.16 Pain Systems: Interface with Affective and Motivational Mechanisms

Carlos López-Jaramillo, M.D., Ph.D. Professor of Psychiatry, Department of Psychiatry, School of Medicine, Universidad de Antioquia, Medellin, Colombia
9.2 The Classification of Mental Disorders in the International Classification of Diseases (ICD-11)

Luminita Luca, M.D., Ph.D. Attending Psychiatrist, Department of Psychiatry and Behavioral Sciences, University of Miami Leonard M. Miller School of Medicine, Miami, Florida
33.11b Cholinesterase Inhibitors and Memantine

Loredana Lucarelli, Psy.D. Chair Professor in Dynamic and Developmental Psychopathology, Department of Pedagogy, Psychology, Philosophy, University of Cagliari, Cagliari, Italy
46 Feeding and Eating Disorders of Infancy and Early Childhood

John Luo, M.D. Health Sciences Clinical Professor of Psychiatry, Department of Psychiatry and Human Behavior, University of

California Irvine Medical Center/University of California Irvine Health/University of California, School of Medicine, Irvine, California
30.9 Computers in Psychiatry

Sean X. Luo, M.D., Ph.D. Assistant Professor of Clinical Psychiatry, Department of Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons, New York, New York
11.9 Opioid Use and Related Disorders: From Neuroscience to Treatment

Anne Bodmer Lutz, B.S.N., M.D. Executive Director of The Institute for Solution-Focused Therapy and Assistant Professor, Psychiatry: Child Psychiatry Division, Institute for Solution-Focused Therapy, University of Massachusetts, T.H. Chan School of Medicine, Worcester, Massachusetts
32.12 Solution-Focused Brief Therapy

Amanda E. Lyall, Ph.D. Assistant Professor, Department of Psychiatry, Brigham and Women's Hospital, Massachusetts General Hospital, Harvard University, Harvard Medical School, Boston, Massachusetts
12.7 Structural Brain Imaging in Schizophrenia

Stanley Lyndon, M.D. Associate Neuropsychiatrist, Center for Brain/Mind Medicine, Brigham and Women's Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
2.2 Neuropsychiatric Aspects of Cerebrovascular Disorders

Frank P. Macmaster, Ph.D. Vice President, Research and Innovation, IWK Health, Halifax, Nova Scotia, Canada
37 Neuroimaging in Psychiatric Disorders of Childhood

Vladimir Maletic, M.D., M.S. Clinical Professor of Psychiatry, Department of Psychiatry, University of South Carolina, School of Medicine, Columbia, South Carolina, and Greenville, South Carolina
13.8 Mood Disorders: Neurobiology

Daniel Mamah, M.D., M.P.E. Professor, Department of Psychiatry, Washington University in St. Louis, School of Medicine, St. Louis, Missouri
12.3 Global Burden of Schizophrenia

Alice R. Mao, M.D. Professor of Psychiatry and Behavioral Sciences, Child and Adolescent Psychiatry, Texas Children's Hospital, Baylor University, College of Medicine, Houston, Texas
54.12 Impact on Parents of Raising a Child with Psychiatric Illness and/or Developmental Disability

Stephen R. Marder, M.D. Daniel X. Freedman Professor of Psychiatry, Department of Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, and VA Greater Los Angeles, Los Angeles, California
33.2b Dopamine-Serotonin Antagonists, 33.3a Dopamine Partial Agonists

Russell L. Margolis, M.D. Professor of Psychiatry and Neurology, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
2.6 Neuropsychiatric Aspects of Movement Disorders

John J. Mariani, M.D. Associate Professor of Clinical Psychiatry, Division on Substance Use Disorders, New York State Psychiatric Institute/CUIMC, New York, New York
11.10 Sedative-, Hypnotic-, or Anxiolytic-Related Disorders

Humberto Marin, M.D. Professor of Psychiatry, Department of Psychiatry and UBHC, Rutgers University, Robert Wood Johnson Medical School, Piscataway, New Jersey
9.1 Present and Future of Classification Systems for Mental Disorders

John C. Markowitz, M.D. Professor of Clinical Psychiatry; Research Psychiatrist, Department of Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons; New York State Psychiatric Institute, New York, New York
13.4 Mood Disorders: Intrapsychic and Interpersonal Aspects

Charles R. Marmar, M.D. Littauer Professor and Chair, Department of Psychiatry, New York University, Grossman School of Medicine, New York, New York
17 Post-Traumatic Stress Disorder

María Marquine, Ph.D. Associate Professor, Department of Medicine, University of California, School of Medicine, San Diego, California
55.6g HIV and Aging

Laura Marsh, M.D. Professor, Menninger Department of Psychiatry and Department of Neurology, Baylor University, College of Medicine/Michael E. DeBakey Veterans Affairs Medical Center, Houston, Texas
2.6 Neuropsychiatric Aspects of Movement Disorders

Dori Marshall, M.D. Clinical Associate Professor of Psychiatry, Department of Psychiatry, University at Buffalo, Jacobs School of Medicine and Biomedical Sciences, Buffalo, New York
33.9a Barbiturates and Similarly Acting Substances, 33.9b Benzodiazepine Receptor Agonists and Antagonists, 33.10a Calcium Channel Inhibitors

Andrés Martin, M.D., Ph.D. Riva Ariella Ritvo Professor, Child Study Center, Yale University, School of Medicine, New Haven, Connecticut
53.6 Inpatient Psychiatric, Partial Hospitalization, and Residential Treatment for Children and Adolescents

Nancy Martin, M.D. Emergency Psychiatrist, Presbyterian Behavioral Health, Albuquerque, New Mexico
30.3 Physical Violence on Adults

Ardis C. Martin, M.D. Child, Adolescent, and Adult Psychiatrist, True InSight Psychiatric, Private Practice, Martinez, California
34.3 Adolescent Development

Caridad Ponce Martinez, M.D. Assistant Professor, Department of Psychiatry, University of Massachusetts, T.H. Chan School of Medicine, Worcester, Massachusetts
33.17c Pharmacologic Treatment of Sexual Dysfunction

Ruben G. Martinez, Ph.D. Postdoctoral Fellow, Department of Psychiatry, University of California, David Geffen School of Medicin, Los Angeles, California
51.3 Separation Anxiety, Generalized Anxiety, and Social Anxiety

Graeme F. Mason, Ph.D. Professor, Departments of Radiology and Biomedical Imaging, Psychiatry, and Biomedical Engineering, Yale University, School of Medicine, New Haven, Connecticut
1.25 Nuclear Magnetic Resonance Imaging and Spectroscopy: Basic Principles and Recent Findings in Neuropsychiatric Disorders

Christopher E. Mason, Ph.D. Professor of Genomics, Physiology, and Biophysics, Physiology and Biophysics, Weill Cornell Medical College, New York, New York
1.18 Genome, Transcriptome, and Proteome: The Molecular Genetics, Biochemistry, and Multi-Omics Underlying the Neurobiology of Mental Disorders

Chacku Mathai, A.A.S. Director, OnTrackNY Healthy Transitions SAMHSA Grant Department, Center for Practice Innovations at Columbia University, New York, New York
12.15 Recovery in Schizophrenia

Anu A. Matorin, M.D. Professor of Psychiatry and Behavioral Sciences, Department of Psychiatry and Behavioral Sciences, Ben Taub Hospital Harris Health/Baylor University, College of Medicine, Houston, Texas
8 Clinical Manifestations of Psychiatric Disorders

Sarah Mattson, Ph.D. Professor and Neuropsychologist, Department of Psychology, Center for Behavioral Teratology, San Diego State University, San Diego, California
2.13 Psychiatric Aspects of Child Neurology

Jeffrey E. Max, M.D., M.B.B.Ch. Professor of Psychiatry, Division of Child and Adolescent Psychiatry, University of California, School of Medicine, San Diego; Rady Children's Hospital, San Diego, California
2.13 Psychiatric Aspects of Child Neurology

Tina L. Mayes, Ph.D. Staff Psychologist, Geriatric Mental Health Clinic, San Diego VA Healthcare System/University of California, School of Medicine, San Diego, California
55.2d Psychological Changes With Normal Aging

Ruth McCann, M.D. Fellow in Neuropsychiatry and Behavioral Neurology, Department of Psychiatry, Brigham and Women's Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
2.9 Neuropsychiatric Aspects of Other Infectious Diseases (Non-HIV)

Kristine L. Mccaslin, Psy.D. Assistant Professor of Psychiatry, Division of Psychology, University of Texas Southwestern, Medical School, Dallas, Texas
6.3 Other Psychodynamic Schools

James T. McCracken, M.D. Professor of Psychiatry and Biobehavioral Sciences, Division of Child and Adolescent Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California

36 Genetics in Child PsychiatryRobert E. McCullumsmith, M.D., Ph.D. Chair of Department of Neurosciences, Department of Neurosciences, University of Toledo, College of Medicine and Life Sciences/ProMedica Neurosciences Institute, Toledo, Ohio

12.6 Cellular and Molecular Neuropathology of Schizophrenia

Courtney R. McDermott, B.A. Doctoral Candidate in Neuroscience, Neuroscience and Cell Biology, Rutgers University, Robert Wood Johnson Medical School, Piscataway, New Jersey

1.4 Neural Development and Neurogenesis

James J. McGough, M.D., M.S. Professor of Clinical Psychiatry and Biobehavioral Sciences, Division of Child and Adolescent Psychiatry, Resnick Neuropsychiatric Hospital/University of California, David Geffen School of Medicine, Los Angeles, California

44.2 Adult Manifestations of Attention-Deficit/Hyperactivity Disorder

Daniel McFarland, D.O., M.S. Associate Professor/Director of Psycho-Oncology, Departments of Psychiatry and Internal Medicine/Division of Hematology and Oncology, University of Rochester, School of Medicine and Dentistry/Wilmot Cancer Center, Rochester, New York

27.8 Psycho-Oncology

John S. McIntyre, M.D. Clinical Professor of Psychiatry, Department of Psychiatry, University of Rochester, School of Medicine and Dentistry, Rochester, New York

7.1 Psychiatric Interview, History, and Mental Status Examination of the Adult Patient, 7.3 Practice Guidelines in Psychiatry

Nicole C.R. McLaughlin, Ph.D. Associate Professor, Research Scholar Track, Neuropsychology, Department of Psychiatry and

Human Behavior, Butler Hospital/Brown University, Warren Alpert Medical School, Providence, Rhode Island
33.20c Neurosurgical Treatments

John R. McQuaid, Ph.D. Chief of Psychology, Psychology Service, VA Palo Alto Health Care System, Palo Alto, California; Vice Chair, Psychiatry and Behavioral Sciences, Weill Institute for Neurosciences, University of California, School of Medicine, San Francisco, California; Associate Chief of Staff for Health Equity, Office of the Chief of Staff, San Francisco VA Health Care System, San Francisco, California
13.7 Mood Disorders: Psychotherapy, 55.4i Cognitive-Behavioral Therapy

Paul D. Meesters, M.D. Geriatric psychiatrist, Education and Research, Mental Health Services, GGZ Friesland, Leeuwarden, The Netherlands
55.3h Schizophrenia and Other Psychotic Disorders

Darlene S. Melchitzky, M.S. Research Principal, Department of Psychiatry, University of Pittsburgh, Medical Center, Pittsburgh, Pennsylvania
1.2 Functional Neuroanatomy

Samantha Meltzer-Brody, M.D., M.P.H. Assad Meymandi Distinguished Professor and Chair, Department of Psychiatry, University of North Carolina, School of Medicine, Chapel Hill, North Carolina
33.17a Sex Differences in Psychopharmacology

Mario F. Mendez, M.D., Ph.D. Professor of Neurology and Psychiatry and Clinical Neurosciences, Department of Neurology, University of California, David Geffen School of Medicine and VA Greater Los Angeles Healthcare System, Los Angeles, California
2.4 Neuropsychiatric Aspects of Epilepsy and Psychogenic Nonepileptic Seizures

Steven Mennerick, Ph.D. John P. Feighner Professor of Neuropsychopharmacology, Department of Psychiatry, Washington University in St. Louis, School of Medicine, St. Louis, Missouri
1.5 Cellular and Synaptic Basis of Neural Signaling

James R. Merikangas, M.D. Clinical Professor of Psychiatry and Behavioral Sciences, Department of Psychiatry, George Washington University, School of Medicine and Health Sciences, Washington, DC
2.11 Neuropsychiatric Aspects of Headache

Kathleen Ries Merikangas, Ph.D. Distinguished Investigator, Intramural Research Program, National Institute of Mental Health, Bethesda, Maryland
2.11 Neuropsychiatric Aspects of Headache, 13.2 Mood Disorders: Epidemiology, 14.1 Anxiety Disorders: Introduction and Overview, 14.3 Epidemiology of Anxiety Disorders

Jonathan Michel Metzl, M.D., Ph.D. Professor, Chair, Department of Medicine, Health, and Society, Vanderbilt University, School of Medicine, Nashville, Tennessee
7.10 Cultural Competency in the Organization and Delivery of Psychiatric Care

Edwin Jens Mikkelsen, M.D. Associate Professor of Psychiatry, Harvard University, Harvard Medical School, Boston, Massachusetts
48 Elimination Disorders

Andrew H. Miller, M.D. Professor of Psychiatry and Behavioral Sciences, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia
1.13 Immune–Brain Interactions in Psychiatry

Barbara L. Milrod, M.D. Professor of Psychiatry, Albert Einstein College of Medicine, New York, New York
13.4 Mood Disorders: Intrapsychic and Interpersonal Aspects

Alireza Minagar, M.D., M.B.A. Retired Neurologist, University of Maryland Global Campus, Bioinformatics, Shreveport, Louisiana
2.10 Neuropsychiatric Aspects of Prion Disease

Daniel Minkin Levy*, M.D. Resident Physician, Women's Department, Jerusalem Mental Health Center, Jerusalem, Israel
33.1c Classification of Psychotropics: Neuroscience-Based Nomenclature (NbN), 33.19a Essential List of Medicines

Wendy G. Mitchell, M.D. Professor, Clinical Neurology, Department of Neurology, University of Southern California, Keck School of Medicine, Children's Hospital Los Angeles, Los Angeles, California
41 Developmental Coordination Disorder

Ramon Mocellin, M.B.B.S., M.Sc., F.R.A.N.Z.C.P. Consultant Neuropsychiatrist and Psychogeriatrician, Delmont Consulting Suites, Delmont Private Hospital, Glen Iris, Victoria, Australia
2.14 Neuropsychiatry of Neurometabolic and Neuroendocrine Disorders

F. Gerard Moeller, M.D. Professor, Department of Psychiatry, Virginia Commonwealth University, School of Medicine, Richmond, Virginia
24.1 Intermittent Explosive Disorder, 24.2 Pyromania, 24.3 Kleptomania

Janna K. Moen, B.A., Ph.D. Postdoctoral Research Fellow, Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York
1.22 Animal Models in Psychiatry

E. Mohandas, M.B.B.S., M.D. Chief Consultant in Psychiatry, Sun Medical and Research Centre, Kerala, India
33.19c Biologic Therapies in India

Ramin Mojtabai, M.D., Ph.D., M.P.H. Professor, Department of Mental Health, Bloomberg School of Public Health and Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, School of Medicine, Baltimore, Maryland
12.17 Other Psychotic Disorders

David J. Moore, Ph.D. Professor of Psychiatry, Department of Psychiatry, University of California, School of Medicine//HIV Neurobehavioral Research Program, San Diego, California
55.3a Assessment of Functioning

Raeanne C. Moore, Ph.D. Associate Professor, Psychiatry, University of California, School of Medicine, San Diego, California
55.6g HIV and Aging

Timothy H. Moran, Ph.D. Paul R. McHugh Professor of Motivated Behaviors, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
1.17 Basic Science of Appetite

Joshua C. Morganstein, M.D. Professor of Psychiatry, Psychiatry, Uniformed Services University of the Health Sciences, F. Edward Hébert School of Medicine, Bethesda, Maryland
30.6 Disaster Psychiatry: Disasters, Terrorism, and War

Stephen J. Morse, J.D., Ph.D. Ferdinand Wakeman Hubbell Professor of Law and Professor of Psychology and Law in Psychiatry, Law School and Department of Psychiatry, University of Pennsylvania, Raymond and Ruth Perelman School of Medicine, Philadelphia, Pennsylvania
4.1 Neurocentrism: Implications for Addiction and the Courtroom

Nidal Moukaddam, M.D., Ph.D. Associate Professor, Department of Psychiatry and Behavioral Sciences, Baylor University, College of Medicine, Houston, Texas
56.1 Public and Community Psychiatry, 56.2 Reforming Health Care

Khaled Moussawi, M.D., Ph.D. Assistant Professor of Psychiatry and Neurology, Department of Psychiatry, University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania
1.31 Neuroscience of Substance Use Disorders

Cynthia Mulder, L.C.S.W.S. Psychoanalyst, Clinical Assistant Professor, Menninger Department of Psychiatry and Behavioral Sciences, Baylor University, College of Medicine; Faculty, Center for Psychoanalytic Studies, Houston, Texas
6.1 Psychoanalytic Theories and Treatment: Perspective from North America and Great Britain

Grace Murray, B.A. Business Office Manager, Gunderson Personality Disorders Institute, McLean Hospital, Belmont, Massachusetts
32.7 Treatment of Borderline Personality Disorder

John D. Murray, Ph.D. Associate Professor of Psychiatry, Department of Psychiatry, Yale University, School of Medicine, New Haven, Connecticut
1.24 Computational Modeling Approaches to Psychiatry

James W. Murrough, M.D., Ph.D. Associate Professor, Department of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York
14.4 Anxiety Disorders: Neurobiology and Neuroscience

Drew Musa, C.P.S. Peer Specialist, Bronx, New York
12.15 Recovery in Schizophrenia

Philip R. Muskin, M.D., M.A., D.L.F.A.P.A., L.F.A.C.L.P. Professor of Psychiatry, Psychiatry/Consultation-Liaison Psychiatry, Columbia University Irving Medical Center/Columbia University, Vagelos College of Physicians and Surgeons, New York, New York
32.3 Psychoanalysis and Psychoanalytic Psychotherapy

Srinivas B. Muvvala, M.D., M.P.H. Associate Professor, Department of Psychiatry, Yale University, School of Medicine, New Haven, Connecticut
33.13a.2 Opioid Receptor Antagonists: Naltrexone and Nalmefene

Ashwini Nadkarni, M.D. Associate Vice Chair, Wellness and Assistant Medical Director, Brigham Psychiatric Specialties Group; Instructor, Harvard University, Harvard Medical School, Department of Psychiatry, Brigham and Women's Hospital, Boston, Massachusetts
27.3 Gastrointestinal Disorders

David Naimark, M.D. Forensic Psychiatrist, Adult Behavioral Health Services, County of San Diego Health & Human Services Agency, San Diego, California
55.6b Forensic Aspects, 55.6e The Legal and Ethical Analysis of Older Adult Abuse

William E. Narrow, M.D., M.P.H. Associate Professor, Department of Psychiatry and Behavioral Sciences, Johns Hopkins Bayview Medical Center/Johns Hopkins University School of Medicine, Baltimore, Maryland
5.1 Quantitative and Experimental Methods in Psychiatry: Epidemiology

J. Craig Nelson, M.D. Professor Emeritus, Department of Psychiatry and Behavioral Sciences, Weill Institute of Neurosciences/University of California, School of Medicine, San Francisco, California
33.5a.3 Tricyclics and Tetracyclics

Charles B. Nemeroff, M.D., Ph.D. Matthew P. Nemeroff Professor and Chair, Department of Psychiatry and Behavioral Sciences, University of Texas, Dell Medical School, Austin, Texas
1.8 Neuropeptides: Biology, Regulation and Role in Neuropsychiatric Disorders

Ariana Nesbit, M.D., M.B.E. Assistant Consulting Professor, Department of Psychiatry and Behavioral Sciences, Central Regional Hospital/Duke University, School of Medicine, Durham, North Carolina; Deputy Chief Medical Officer, Central Regional Hospital, Butner, North Carolina
55.6b Forensic Aspects, 55.6e The Legal and Ethical Analysis of Older Adult Abuse

Mark P. Nespeca, M.D. Clinical Professor of Neuroscience and Pediatrics, Departments of Neurosciences and Pediatrics, Rady Childrens Hospital San Diego, University of California, School of Medicine, San Diego, California
2.13 Psychiatric Aspects of Child Neurology

Randolph M. Nesse, M.D. Professor Emeritus of Psychiatry, Professor Emeritus of Psychology, Department of Psychiatry, University of Michigan, Medical School, Ann Arbor, Michigan
4.3 Evolutionary Foundations for Psychiatric Research and Practice

Paul S. Nestadt, M.D. Associate Professor of Psychiatry, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
27.10 Stress and Psychiatry

D. Jeffrey Newport, M.D., M.S., M.Div. Professor of Psychiatry & Behavioral Sciences and Women’s Health, Departments of Psychiatry and Behavioral Sciences and Women’s Health, University of Texas, Dell Medical School, Austin, Texas
33.15a Melatonin Receptor Agonists

Dorian S. Newton, Ph.D. Psychologist and Psychoanalyst, Oakland, California
6.2 Erik H. Erikson

Sarah A. Nguyen, M.D. Assistant Clinical Professor, Department of Psychiatry and Biobehavioral Sciences, University of California,

David Geffen School of Medicine, Los Angeles, California
55.2b Complementary and Integrative Geriatric Psychiatry

Sharon Nichols, Ph.D. Research Scientist, Department of Neurosciences, University of California, School of Medicine, San Diego, California
2.13 Psychiatric Aspects of Child Neurology

Frank John Ninivaggi, M.D. Assistant Clinical Professor of Child Psychiatry, Yale Child Study Center, Yale University School of Medicine, New Haven, Connecticut
28.1 Malingering, 28.3 Borderline Intellectual Functioning and Academic or Educational Problem

Wanjikũ F.M. Njoroge, M.D. Assistant Professor/Associate Chair of Diversity, Equity, Inclusion, Department of Psychiatry/Department of Child and Adolescent Psychiatry and Behavioral Sciences, University of Pennsylvania, Raymond and Ruth Perelman School of Medicine/Children's Hospital of Philadelphia, Philadelphia, Pennsylvania
54.17 Racism and Discrimination's Impact on the Health and Well-Being of Developing Children and Adolescents

Hannah O. Nordberg, M.A. Student, Department of Psychology, Southern Methodist University, Dallas, Texas
27.5 Respiratory Disorders

Erika L. Nurmi, M.D., Ph.D. Associate Professor of Psychiatry and Biobehavioral Sciences, Division of Child and Adolescent Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
36 Genetics in Child Psychiatry

David Nutt, F.R.C.P., F.R.C.Psych., F.B.Ph.S., F.Med.Sci., D.Laws. Edmond J. Safra Professor of Neuropsychopharmacology, Division

of Psychiatry, Imperial College London, London, United Kingdom
33.14b MDMA, Psilocybin, Ayahuasca, and Ibogaine

Jennifer A. O'Day, M.D. Psychopharmacology Consultant, Department of Psychiatry, California Department of State Hospitals, Los Angeles, California
33.7a Pimavanserin

Andres A. Avellaneda Ojeda, M.D. Associate professor of Psychiatry, Menninger Department of Psychiatry, Baylor University, College of Medicine, Houston, Texas
8 Clinical Manifestations of Psychiatric Disorders

Nora C. Oberfield, M.D. Assistant Professor of Psychiatry at Columbia University Medical Center, Child and Adolescent Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons, New York, New York
7.11 Medical Error

Taylor A. Ochalek, Ph.D. Postdoctoral Fellow, Department of Pharmacology and Toxicology, Virginia Commonwealth University, School of Medicine, Richmond, Virginia
24.1 Intermittent Explosive Disorder, 24.2 Pyromania, 24.3 Kleptomania

Michael Ogata, M.D. Teaching Faculty, Inpatient Psychiatry, Department of Psychiatry, Mission Community Hospital Psychiatry Residency Program, Panorama City, California
54.14 Internet Gaming Disorder in Youth

Ejine Okoroafor, M.D., F.A.P.A. Medical Director, Central Point Psychiatry, Mental Health OP Clinic, Las Vegas, Nevada
27.14 Psychiatric Care of the Burned Patient

Stephanie S. O'Malley, Ph.D. Elizabeth Mears and House Jameson Professor of Psychiatry and Deputy Chair for Clinical Research,

Department of Psychiatry, Yale University, School of Medicine, New Haven, Connecticut
33.13a.2 Opioid Receptor Antagonists: Naltrexone and Nalmefene

Lauren M. Osborne, M.D. Associate Professor and Vice Chair for Clinical Research; Associate Professor, Department of Obstetrics and Gynecology; Department of Psychiatry, Weill Cornell Medical College, New York, New York
30.1 Psychiatry and Reproductive Medicine

Hanadi Ajam Oughli, M.D. Health Sciences Assistant Clinical Professor, Department of Psychiatry and Biobehavioral Sciences, University of California, David Geffen School of Medicine, Los Angeles, California
55.2b Complementary and Integrative Geriatric Psychiatry

Fred Ovsiew, M.D. Clinical Professor of Psychiatry and Behavioral Sciences, Psychiatry and Behavioral Sciences, Northwestern Memorial Hospital/Northwestern University, Feinberg School of Medicine, Chicago, Illinois
2.1 The Neuropsychiatric Approach to the Patient

Kyle Owen, D.O. Lead Psychiatrist—Consult Liaison service, Department of Psychiatry, Presbyterian Health System, Albuquerque, New Mexico
30.3 Physical Violence on Adults

Michael J. Owen, B.Sc. (Hons.), M.B., Ch.B., Ph.D., F.R.C.P.sych. Professor of Psychological Medicine, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, United Kingdom
12.4 Genetics of Schizophrenia, 1.8 Neuropeptides: Biology, Regulation and Role in Neuropsychiatric Disorders

Vanessa Padilla, M.D. Assistant Professor, Consultation-Liaison Psychiatry, Department of Clinical Psychiatry and Behavioral

Sciences, University of Miami Leonard M. Miller School of Medicine, Miami, Florida
56.6 Psychiatric Rehabilitation

Barton W. Palmer, Ph.D. Professor in Residence, Department of Psychiatry, University of California, School of Medicine, San Diego, California; Staff Psychologist, Mental Illness Research, Education, and Clinical Center and Division of Psychology, VA San Diego Healthcare System, San Diego, California
55.2e Neuropsychological Evaluation, 55.6i Positive Psychiatry of Aging

Maryland Pao, M.D. Clinical and Deputy Scientific Director, National Institute of Mental Health, Office of the Clinical Director, National Institutes of Health, Bethesda, Maryland
54.3 Children’s Reaction to Illness and Hospitalization

Leila Parand, M.D. Neurologist, Neurobehavior Specialist, Department of Neurology, University of California Health/David Geffen School of Medicine, Los Angeles, California
2.4 Neuropsychiatric Aspects of Epilepsy and Psychogenic Nonepileptic Seizures

Lawrence T. Park, A.M., M.D. Medical Director, Experimental Therapeutics and Pathophysiology Branch, National Institute of Mental Health, Bethesda, Maryland
33.4a Dopamine Receptor Agonists, 33.14a Esketamine and Ketamine

Nansook Park, Ph.D. Professor of Psychology, Psychology, University of Michigan, Medical School, Ann Arbor, Michigan
32.16 Positive Psychology

Marcella Pascualy, M.D. Associate Professor of Psychiatry, Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington

55.3b Psychiatric Problems in the Medically Ill Geriatric Patient

Caroly Sue Pataki, M.D. Health Sciences Clinical Professor, Department of Psychiatry and Biobehavioral Sciences, University of California, David Geffen School of Medicine, Los Angeles, California
34.1 Introduction and Overview, 34.3 Adolescent Development, 35.2 Psychiatric Assessment of Preschool Children, 50.2 Early-Onset Bipolar Disorder, 54.14 Internet Gaming Disorder in Youth, 54.15 Social Media and Youth

Kartik Pattabiraman, M.D., Ph.D. Assistant Professor, Child Study Center, Yale New Haven Hospital, Yale University, School of Medicine, New Haven, Connecticut
1.3 Functional Genomics of Human Brain Development

Bindu D. Paul, Ph.D. Assistant Professor of Pharmacology, Psychiatry and Neuroscience, Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
1.9 Novel Neurotransmitters

Holly L. Peay, Ph.D., M.S. Senior Research Scientist, Genomics and Translational Research Center, RTI International, Research Triangle Park, Durham, North Carolina
30.2 Genetic Counseling for Psychiatric Conditions

Dale Aaron Peeples, M.D. Associate Professor of Psychiatry, Department of Psychiatry and Health Behavior, Child and Adolescent Psychiatry Division, Augusta University, Medical College of Georgia, Augusta, Georgia
54.1 Adoption and Foster Care

Kevin Pelphrey, Ph.D. Harrison-Wood Jefferson Scholars Foundation Professor, Department of Neurology, University of Virginia, School of Medicine, Charlottesville, Virginia
43 Autism Spectrum Disorders

Max Pemberton, M.B.B.S., B.Sc., M.R.C.Psych. Specialty Trainee in General Adult and Old Age Psychiatry, Department of Psychiatry, Camden and Islington NHS Foundation Trust, London, United Kingdom
60 World Aspects of Psychiatry

Tara S. Peris, Ph.D. Professor and Vice Chair for Research, Department of Psychiatry and Biobehavioral Sciences, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
51.3 Separation Anxiety, Generalized Anxiety, and Social Anxiety

Bernice A. Pescosolido, Ph.D. Distinguished Professor, Sociology Department, Indiana University, School of Medicine, Indianapolis, Indiana
56.7 A Sociocultural Framework for Mental Health and Substance Abuse Service Inequities

Ninoska D. Peterson, Ph.D. Adjunct Assistant Professor, Bariatric and Metabolic Institute, Cleveland Clinic Lerner College of Medicine, Cleveland, Ohio
27.4 Psychosomatic Medicine: Obesity

Joanna Petrides, M.B.S., Psy.D. Associate Professor, Department of Family Medicine, Rowan-Virtua School of Osteopathic Medicine, Stratford, New Jersey
32.17 Acceptance and Commitment Therapy

Megan L. Petrik, Ph.D. Assistant Professor, Department of Medicine, University of Minnesota Medical School, Minneapolis, Minnesota
32.24 Evaluating Psychotherapy: Research Methods, Findings, and Future Directions

John Piacentini, Ph.D., A.B.P.P. Professor of Psychiatry and Biobehavioral Sciences, Division of Child and Adolescent Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
51.1 Obsessive-Compulsive Disorder in Childhood

Ralph L. Piedmont, Ph.D. Managing Director, Center for Professional Studies, Timonium, Maryland
6.4 Approaches Derived from Philosophy and Psychology

Katherine S. Pier, M.D. Addiction Psychiatrist, Addiction Medicine and Recovery Services and Department of Psychiatry, The Permanente Medical Group, San Francisco, California
33.13a.2 Opioid Receptor Antagonists: Naltrexone and Nalmefene, 33.13b Disulfiram and Acamprosate

Sandra S. Pimentel, Ph.D. Associate Professor, Chief of Child and Adolescent Psychology, Department of Psychiatry and Behavioral Sciences, Montefiore Medical Center/Albert Einstein College of Medicine, New York, New York
53.2 Cognitive-Behavioral Psychotherapy for Children and Adolescents

Eric M. Plakun, M.D. Medical Director and Chief Executive Officer, The Austen Riggs Center, Stockbridge, Massachusetts
32.4 Psychoanalytic Treatment of Anxiety, Obsessive-Compulsive and Trauma-Related Disorders

Sirisha Pochareddy, Ph.D. Administrative Director of Genome Sciences Core, Penn State College of Medicine, Hershey, Pennsylvania
1.3 Functional Genomics of Human Brain Development

Carol Ann Podgorski, Ph.D., M.P.H., M.S. Professor of Psychiatry, Department of Psychiatry, University of Rochester, School of

Medicine and Dentistry, Rochester, New York
55.4j Family Intervention and Therapy With Older Adults

Bruce G. Pollock, M.D., Ph.D., F.R.C.P.C. Peter and Shelagh Godsoe Endowed Chair in Late-Life Mental Health, Centre for Addiction and Mental Health; Professor of Psychiatry and Pharmacology, Temerty Faculty of Medicine, University of Toronto, Temerty, Toronto, Ontario, Canada
33.18b Biologic Therapies: Geriatric Psychopharmacology, 55.4a Geriatric Psychopharmacology: General Principles

Norman K. Pollock, Ph.D. Senior Research Scientist, Department of Pediatrics, Boston Medical Center/Boston University, Aram V Chobanian and Edward Avedisian, School of Medicine, Boston, Massachusetts
30.5 Micronutrient Monitoring and Supplementation in Psychiatric Illness

Maurizio Pompili, M.D., Ph.D. Full Professor and Chair of Psychiatry, Department of Neurosciences, Mental Health and Sensory Organs, Sant'Andrea Hospital, Sapienza University of Rome, Rome, Italy
13.5 Mood Disorders: Suicidal Behavior

Harrison G. Pope, Jr., M.D. Professor of Psychiatry, McLean Hospital, Belmont, Massachusetts/Harvard University, Harvard Medical School, Boston, Massachusetts
11.11 Anabolic-Androgenic Steroid-Related Disorders

Robert M. Post, M.D. Clinical Professor of Psychiatry, Department of Psychiatry, George Washington University, School of Medicine and Health Sciences, Washington, DC
33.10c Carbamazepine, 33.10d Lamotrigine, 33.10e Valproate

Seth Powsner, M.D. Professor, Department of Psychiatry and Emergency Medicine, Yale University, School of Medicine, New

Haven, Connecticut
19 Factitious Disorder

Jenessa S. Price, Ph.D. Associate Professor, Department of Surgery, Division of Transplant Surgery; Department of Psychiatry and Behavioral Medicine, Medical College of Wisconsin, Milwaukee, Wisconsin
32.8 Behavior Therapy

Leslie S. Prichep, Ph.D. Chief Scientific Officer, BrainScope Company, Chevy Chase, Maryland; Professor of Psychiatry and Director of the Brain Research Laboratories (Retired), Department of Psychiatry, City University of New York, School of Medicine, New York, New York
7.7 Principles and Applications of Quantitative Electroencephalography in Psychiatry

Elizabeth J. Prince, D.O. Assistant Professor, Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
27.10 Stress and Psychiatry

Jessica L. Principe, M.D., M.S. Resident Physician in Psychiatry, Department of Psychiatry, Brigham and Women’s Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
30.5 Micronutrient Monitoring and Supplementation in Psychiatric Illness

Joan Prudic, M.D. Clinical Professor of Psychiatry, Department of Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons/Irving Medical Center, New York, New York
33.20a Electroconvulsive Therapy

Andrés J. Pumariega, M.D. Professor and Chief, Department of Psychiatry, Division of Child and Adolescent Psychiatry, University of Florida College of Medicine, Gainesville, Florida

53.7 Community-Based Treatment

Davin K. Quinn, M.D. Vice Chair, Adult Clinical Services, Department of Psychiatry and Behavioral Sciences, University of New Mexico School of Medicine, Albuquerque, New Mexico
19 Factitious Disorder

Avery Quynh, B.S. Staff Research Associate II, Department of Psychiatry, University of California, School of Medicine, San Diego, California
55.6h Technology for Seniors

Mark Ragins, M.D. Psychiatrist, Department of Student Health Services/Counseling and Psychological Services, California State University Long Beach, Long Beach, California
12.15 Recovery in Schizophrenia

Nusrat Rahman, Ph.D. Senior Researcher, Department of Research, American Psychiatric Association, Washington, DC
5.1 Quantitative and Experimental Methods in Psychiatry: Epidemiology

Charles L. Raison, M.D. Professor of Human Ecology and Psychiatry, Department of Psychiatry; Human Development and Family Studies, School of Medicine and Public Health; School of Human Ecology, University of Wisconsin, School of Medicine and Public Health, Madison, Wisconsin
1.13 Immune–Brain Interactions in Psychiatry, 13.8 Mood Disorders: Neurobiology

Tarek K. Rajji, M.D., F.R.C.P.C. Professor of Psychiatry, Geriatric Psychiatry Division, Centre for Addiction and Mental Health/Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
55.4f Electroconvulsive Therapy and Neurostimulation Treatments

Seethalakshmi Ramanathan, M.D., M.P.A. Clinical Associate Professor of Psychiatry, Department of Psychiatry and Behavioral Sciences, State University of New York, Upstate Medical University, Alan and Marlene Norton College of Medicine, Syracuse, New York
55.2f Genetics of Late-Life Neurodegenerative Disorders

Natalia Ramos, M.D., M.P.H. Assistant Clinical Professor, Psychiatry and Biobehavioral Sciences, University of California, David Geffen School of Medicine, Los Angeles, California
54.16 Gay, Bisexual, Transgender, Queer/Questioning, Intersex, and Asexual (LGBTQIA) Youth

Laura B. Ramsey, Ph.D. Associate Professor, Department of Pediatrics, Divisions of Clinical Pharmacology and Research in Patient Services, University of Cincinnati, College of Medicine and Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio
33.18a Pediatric Psychopharmacology, 53.5 Pediatric Psychopharmacology

Mark M. Rasenick, Ph.D. Distinguished University of Illinois Chicago, Professor of Physiology and Biophysics and Psychiatry, University of Illinois College of Medicine, Research Career Scientist, Jesse Brown VA Medical Center, President/CSO Pax Neuroscience, Chicago, Illinois
33.1a General Principles of Psychopharmacology

Natalie Rasgon, M.D., Ph.D. Professor of Psychiatry and Behavioral Sciences, Psychiatry and Behavioral Sciences, Stanford University, School of Medicine, Palo Alto, California
27.7 Endocrine and Metabolic Disorders

Wegdan Rashad, M.D. Knowledge Translation Coordinator, Department of Psychiatry, Centre for Depression and Suicide Studies, St. Michael's Hospital, Toronto, Ontario, Canada
13.6 Mood Disorders: Pharmacologic Treatment of Depression and Bipolar Disorders

Lakshmi N. Ravindran, M.D., F.R.C.P.C. Assistant Professor of Psychiatry, Department of Psychiatry, Centre for Addiction and Mental Health/University of Toronto, Toronto, Ontario, Canada
14.7 Somatic Therapies for Anxiety Disorders

Elliott Rees, B.Sc. (Hons.), Ph.D. Dr of Psychiatric Genetics, Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences/Cardiff University, Cardiff, United Kingdom
12.4 Genetics of Schizophrenia

Ulrich Reininghaus, Ph.D. Heisenberg Professor Head of Department, Department of Public Mental Health, Central Institute of Mental Health, Mannheim, Germany
12.5 The Clinical Epidemiology of Psychosis Spectrum Disorder

Maya A. Reiter, M.S., Ph.D. Visiting Scholar, Department of Psychology, University of California, School of Medicine, San Diego, California
55.2d Psychological Changes With Normal Aging

Judith Restrepo, M.D. Associate Director Transplant Psychiatry, Consultation Liaison Psychiatry Service, Massachusetts General Hospital, Boston, Massachusetts
27.2 Cardiovascular Disorders

David C. Rettew, M.D. Clinical Faculty, Oregon Health & Science University, School of Medicine, Department of Psychiatry, Portland, Oregon; Medical Director, Lane County Behavioral Health, Eugene, Oregon
38 Temperament: Risk and Protective Factors for Child Psychiatric Disorders

Victor I. Reus, M.D. Distinguished Professor Emeritus, Department of Psychiatry and Behavioral Sciences, University of California,

School of Medicine, UCSF Weill Institute for Neurosciences, San Francisco, California
1.12 Psychoneuroendocrinology

Paula D. Riggs, B.A., M.A., M.D. Professor and Vice Chair, Faculty Affairs and Director, Division of Addiction Science Prevention and Treatment, Department of Psychiatry, University of Colorado, School of Medicine, Aurora, Colorado
54.5 Adolescent Substance Use Disorders

Zoltán Rihmer, M.D., Ph.D., D.Sc. Clinical Professor of Psychiatry, Department of Psychiatry and Psychotherapy, Semmelweis University and National Institute of Mental Health, Neurology and Neurosurgery, Budapest, Hungary
13.2 Mood Disorders: Epidemiology, 13.5 Mood Disorders: Suicidal Behavior

Thomas Ritz, Ph.D. Professor, Department of Psychology, Southern Methodist University, Dallas, Texas
27.5 Respiratory Disorders

Laura Weiss Roberts, M.D., M.A. Professor and Chairman, Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California
58.2 Ethics in Psychiatry

Xiomara Rocha-Cadman, M.D. Assistant Clinical Professor of Psychiatry, Division of Psychiatry, City of Hope Comprehensive Cancer Center, Duarte, California
27.9 End-of-Life and Palliative Care

Joshua J. Rodgers, M.D. Assistant Clinical Professor, Department of Psychiatry, University of Colorado School of Medicine, Aurora, Colorado; Chief Medical Officer, Community Reach Center, Westminster, Colorado; Director, Behavioral Neurology and

Neuropsychiatry, Baylor University, College of Medicine, Houston, Texas
1.14 Chronobiology, Circadian Rhythm, and Psychiatry, 2.15 Catatonia, 2.7 Neuropsychiatric Aspects of Multiple Sclerosis and Other Demyelinating Disorders

Ana Rodriguez-Villa, M.D., M.B.A. Staff Psychiatrist, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts; McLean Hospital, Belmont, Massachusetts
32.7 Treatment of Borderline Personality Disorder

Christian Romanchek, M.D. Resident Physician, Department of Psychiatry, The Johns Hopkins Hospital, Baltimore, Maryland
27.11 Psychiatric Morbidity Following Critical Illness

Alasdair G. Rooney, MB.ChB., M.D., Ph.D. Consultant Neuropsychiatrist, Scottish Neurobehavioural Rehabilitation Service, Royal Edinburgh Hospital, NHS Lothian; Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, Scotland
2.3 The Neuropsychiatry of Brain Tumors

David R. Rosenberg, M.D. Professor and Chair of the Department of Psychiatry and Behavioral Neurosciences, Wayne State University, School of Medicine, Detroit, Michigan
37 Neuroimaging in Psychiatric Disorders of Childhood

M. Zachary Rosenthal, Ph.D. Associate Professor, Department of Psychiatry and Behavioral Science, Duke University, School of Medicine/Medical Center, Durham, North Carolina
32.11 Dialectical Behavior Therapy

Stefan B. Rowny, M.D., M.F.A. Attending Psychiatrist, Assistant Professor, Department of Psychiatry, New York State Psychiatric Institute/Columbia University, Vagelos College of Physicians and Surgeons/Irving Medical Center, New York, New York
33.20b Brain Stimulation Methods

Peter Roy-Byrne, M.D. Professor, Department of Psychiatry, University of Washington School of Medicine, Seattle, Washington
14.2 Clinical Features of the Anxiety Disorders

David R. Rubinow, M.D. Research Professor, Departments of Psychiatry and Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina
33.17b Reproductive Hormonal Therapy: Theory and Practice

Katrina A. Rufino, Ph.D. Associate Professor in Psychology, Department of Social Sciences, University of Houston– Downtown, Senior Research Scientist, Department of Psychiatry and Behavioral Sciences, The Menninger Clinic/Baylor University, College of Medicine, Houston, Texas
31.2 Suicide Treatment

Pedro Ruiz, M.D. Professor of Psychiatry, Department of Psychiatry, University of Texas Health Science Center, McGovern Medical School, Houston, Texas
8 Clinical Manifestations of Psychiatric Disorders, 60 World Aspects of Psychiatry

Samir A. Sabbag, M.D. Assistant Professor, Adult and Geriatric Psychiatry, Community Hospital of Monterey Peninsula, Monterey, California
33.11b Cholinesterase Inhibitors and Memantine

Ewurama Sackey, M.D. Child and Adolescent Psychiatry Fellow, Department of Psychiatry and Biobehavioral Sciences, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
54.17 Racism and Discrimination’s Impact on the Health and Well-Being of Developing Children and Adolescents

Joel I. Sadavoy, M.D., F.R.C.P.(C) Professor and Medical Director, Psychiatry (Reitman Centre), University of Toronto and Psychiatrist in Chief and inaugural Pencer Chair in Applied General Psychiatry at Mount Sinai Hospital, Toronto, Ontario, Canada
55.4g Psychosocial Factors in Psychotherapy of the Older Adults, 55.4h Individual Psychotherapy

Joseph R. Sadek, Ph.D., A.B.P.P.-C.N. Associate Professor, Psychiatry and Behavioral Sciences, University of New Mexico, School of Medicine and New Mexico VA Health Care System, Albuquerque, New Mexico
7.4 Neuropsychological Assessment of Adults

Virginia A. Sadock, M.D. Professor of Psychiatry, Department of Psychiatry, New York University, Grossman School of Medicine; Attending Physician, Department of Psychiatry, NYU Langone Hospital Center, New York, New York
21.1 Normal Human Sexuality and Sexual Dysfunctions

Benjamin J. Sadock, M.D. Menas S. Gregroy Professor of Psychiatry, Department of Psychiatry, New York University Grossman School of Medicine; Attending Physician, Department of Psychiatry, NYU Langone Hospital Center; Attending Physician, Bellevue Hospital, New York, New York
7.2 The Psychiatric Report and Outline for a Psychiatric Examination

Carl Salzman, M.D. Professor of Psychiatry, Department of Psychiatry, Beth Israel Deaconess Hospital and Massachusetts Mental Health Center, Harvard University, Harvard Medical School, Boston, Massachusetts
55.4b Antidepressants and Mood Stabilizers

Gerard Sanacora, M.D., Ph.D. George D. Gross and Esther S. Gross Professor of Psychiatry, Yale University School of Medicine, Director of Yale Depression Research, Department of Psychiatry, Co-Director, Yale-New Haven Hospital Interventional Psychiatry Services, New Haven, Connecticut

1.25 Nuclear Magnetic Resonance Imaging and Spectroscopy: Basic Principles and Recent Findings in Neuropsychiatric DisordersJohn Sargent, M.D. Director, Division of Child and Adolescent Psychiatry, Tufts Medical Center, Boston, Massachusetts

53.4 Family Therapy

Sally L. Satel, M.D. Lecturer, Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut

4.1 Neurocentrism: Implications for Addiction and the Courtroom

Stephen M. Saunders, Ph.D. Professor of Psychology, Psychology Department, Marquette University, Milwaukee, Wisconsin

32.24 Evaluating Psychotherapy: Research Methods, Findings, and Future Directions

Radu V. Saveanu, M.D. Professor and Executive Vice Chair, Department of Psychiatry and Behavioral Sciences, University of Miami Leonard M. Miller School of Medicine, Miami, Florida

33.6c Mirtazapine, 56.6 Psychiatric Rehabilitation

Gauri Savla, Ph.D. Clinical Psychologist, Private Practice, Gauri Savla, Encinitas, California

55.2e Neuropsychological Evaluation

Andrew J. Saxon, M.D. Professor, Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine/VA Puget Sound Health Care System, Seattle, Washington

33.13a.1 Opioid Maintenance Treatments: Methadone and Buprenorphine

Ayal Schaffer, M.D., F.R.C.P.C. Professor, Department of Psychiatry, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada

13.6 Mood Disorders: Pharmacologic Treatment of Depression and Bipolar Disorders

Alan F. Schatzberg, M.D. Kenneth T. Norris, Jr., Professor of Psychiatry and Behavioral Sciences, Department of Psychiatry and Behavioral Sciences, Stanford University, School of Medicine, Palo Alto, California
33.5a.4 Dopamine-Noradrenaline Reuptake Inhibitors (Bupropion), 33.16 Combination Pharmacotherapy

Steven C. Schlozman, M.D. Associate Professor, Department of Psychiatry, Dartmouth College, Geisel School of Medicine, Hanover, New Hampshire
53.8 The Treatment of Adolescents

Peter J. Schmidt, M.D. Chief, Section on Behavioral Endocrinology, National Institute of Mental Health, Intramural Research Program, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland
33.17b Reproductive Hormonal Therapy: Theory and Practice

Benjamin N. Schneider, M.D. Associate Professor of Psychiatry and Biobehavioral Sciences, Department of Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
51.3 Separation Anxiety, Generalized Anxiety, and Social Anxiety

Edward J. Schreiber, T.E.P. Psychodrama Director and Adjunct Professor, Zerka Foundation, Amherst, Massachusetts, and Lesley University, Cambridge, Massachusetts
32.22 Psychodrama, Sociodrama, Sociometry, and Sociatry

Matthew Alan Schreiber, M.D., Ph.D. Assistant Professor, Department of Psychiatry and Behavioral Sciences, University of Washington, School of Medicine, Seattle, Washington

55.3b Psychiatric Problems in the Medically Ill Geriatric Patient

Marc A. Schuckit, M.D. Distinguished Professor of Psychiatry, University of California, School of Medicine, San Diego, California
11.2 Alcohol-Related Disorders

Kenneth M. Schwartz, M.D., F.R.C.P.C. Assistant Professor of Medicine, Department of Psychiatry, Baycrest Center for Geriatric Care, Toronto, Ontario, Canada
55.4k Group Therapy

Nicholas P. Schweizer, Ed.D., L.C.A.D.C., L.C.P.C. Program Manager, Department of Psychiatry, Johns Hopkins Hospital, Johns Hopkins University, School of Medicine, AIDS Psychiatry Service, Baltimore, Maryland
2.8 Neuropsychiatric Aspects of HIV Infection and AIDS

Vanessa Scott, B.S. Staff Research Associate, Psychiatry, University of California, School of Medicine/San Diego Health, San Diego, California
55.6h Technology for Seniors

Thomas W. Sedlak, M.D., Ph.D. Assistant Professor, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, School of Medicine, Baltimore, Maryland
1.9 Novel Neurotransmitters

Mary Schwab Seligman, M.D., M.A.R., Ph.D. Professor of Psychiatry, Psychology, Epidemiology, Nursing, Biostatistics, and Clinical and Translational Science, Department of Psychiatry, University of Pittsburgh School of Medicine and Medical Center, Pittsburgh, Pennsylvania
35.1 Psychiatric Examination of the Infant, Child, and Adolescent

Nenad Sestan, M.D., Ph.D. Harvey and Kate Cushing Professor of Neuroscience, Department of Neuroscience, Yale University, School

of Medicine, New Haven, Connecticut
1.3 Functional Genomics of Human Brain Development

Daniel D. Sewell, M.D. Professor Emeritus, Department of Psychiatry, University of California, School of Medicine, San Diego, California
55.6f Geriatric Psychiatry: Sexuality and Aging

Sandra B. Sexson, M.D. Professor and Chief, Child, Adolescent and Family Psychiatry, Department of Psychiatry and Health Behavior, Augusta University, Medical College of Georgia, Augusta, Georgia
54.1 Adoption and Foster Care

Olivia Shadid, M.D. Child and Adolescent Psychiatry Fellow, Department of Psychiatry and Behavioral Sciences, University of New Mexico School of Medicine, Albuquerque, New Mexico
54.18 Forensic Evaluations of Child and Adolescent Migrants

Arya Shah, M.D. Resident, Department of Psychiatry, Brigham and Women's Hospital, Boston, Massachusetts
58.3 Correctional Psychiatry

Asim A. Shah, M.D. Professor & Executive Vice Chair, Menninger Department of Psychiatry, Baylor University, College of Medicine, Houston, Texas
8 Clinical Manifestations of Psychiatric Disorders, 56.1 Public and Community Psychiatry, 56.2 Reforming Health Care

Talya Shahal, M.D. Clinical Instructor in Psychiatry, Department of Psychiatry, Veterans Administration Boston Healthcare System/Harvard University, Harvard Medical School, Boston, Massachusetts
55.3e Geriatric Mood Disorders

Arieh Y. Shalev, M.D. Adjunct Professor of Psychiatry, Department of Psychiatry, New York University, Grossman School of Medicine, New York, New York
17 Post-Traumatic Stress Disorder

Gabrielle Shapiro, M.D., D.F.A.A.C.A.P., D.F.A.P.A. Clinical Professor of Psychiatry, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York
34.2 Child Development

Michael Shapiro, M.D. Clinical Associate Professor of Psychiatry, Division of Child and Adolescent Psychiatry, Department of Psychiatry/University of Florida College of Medicine, Gainesville, Florida
53.1 Psychodynamic Psychotherapy (Child Psychiatry)

Paul Shapshak, Ph.D. Professor, Internal Medicine, University of South Florida, Tamp, Florida
2.10 Neuropsychiatric Aspects of Prion Disease

Amir Sharafkhaneh, M.D., Ph.D. Professor of Medicine, Department of Medicine, Section of Pulmonary, Critical Care and Sleep Medicine, Michael E. DeBakey VA Medical Center, Baylor University, College of Medicine, Houston, Texas
23 Sleep Disorders

Neha Sharma, D.O. Associate Professor of Psychiatry, Child and Adolescent Psychiatry Fellowship Program Director, Department of Psychiatry, Division of Child and Adolescent Psychiatry, Tufts University, School of Medicine/Medicine—Tufts Medical, Boston, Massachusetts
53.4 Family Therapy

Shweta Sharma, Psy.D. Assistant Professor, Menninger Department of Psychiatry and Behavioral Sciences, Baylor University College of Medicine, Houston, Texas

6.1 Psychoanalytic Theories and Treatment: Perspective from North America and Great Britain

Jess P. Shatkin, M.D., M.P.H. Professor and Vice Chair for Education, Department of Child and Adolescent Psychiatry and Pediatrics, New York University Grossman School of Medicine, New York, New York
54.13 Pediatric Sleep Disorders

Martha E. Shenton, Ph.D. Professor, Psychiatry and Radiology, Brigham and Women's Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
12.7 Structural Brain Imaging in Schizophrenia

Edward Shorter, Ph.D., F.R.S.C. Jason A Hannah Professor of the History of Medicine/Professor of Psychiatry, Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
59 History of Psychiatry

Rajeet Shrestha, M.D. Assistant Professor, Psychiatry, Department of Psychiatry, University Hospitals Cleveland Medical Center/Case Western Reserve University, School of Medicine, Cleveland, Ohio
2.12 Neuropsychiatric Aspects of Neuromuscular Diseases

Shawn Singh Sidhu, M.D. Associate Professor of Psychiatry, Department of Psychiatry, Division of Child and Adolescent Psychiatry, University of California, School of Medicine, San Diego, California
54.18 Forensic Evaluations of Child and Adolescent Migrants

Carole E. Siegel, Ph.D. Research Professor of Psychiatry, Psychiatry and Population Health, New York University, Grossman School of Medicine, New York, New York
5.2 Statistics and Experimental Design, 7.10 Cultural Competency in the Organization and Delivery of Psychiatric Care

Daniel J. Siegel, M.D. Executive Director, Mindsight Institute, Santa Monica, California; Clinical Professor of Psychiatry, University of California, David Geffen School of Medicine, Founding Co-Director, Mindful Awareness Research Center at UCLA, Los Angeles, California
3.1 Contributions of the Psychological Sciences: Sensation, Perception, and Cognition

Linmarie Sikich, M.D., M.A. Associate Clinical Professor of Psychiatry and Behavioral Health, Division of Child and Family Mental Health and Community Psychiatry, Duke University School of Medicine, Durham, North Carolina
52 Early-Onset Psychotic Disorders

John C. Silbereis, Ph.D. Fellow, Department of Neuroscience, Yale University School of Medicine, New Haven, Connecticut
1.3 Functional Genomics of Human Brain Development

David A. Silbersweig, M.D. Stanley Cobb Professor of Psychiatry; Chairman, Department of Psychiatry, Brigham and Women's Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
1.29 Neuropsychiatry of Self

Samuel Justin Sinclair, Ph.D. Assistant Professor, Department of Psychiatry, Harvard University, Harvard Medical School; Director of the Psychological Assessment and Research Service (PARSe), Brigham and Women's Hospital, Boston, Massachusetts
7.8 Psychiatric Rating Scales

Manpreet Kaur Singh, M.D., M.S. Associate Professor of Psychiatry and Behavioral Sciences, Department of Psychiatry and Behavioral Sciences, Stanford University, School of Medicine, Palo Alto, California
33.4b Amphetamines, Methylphenidate, Atomoxetine, and Modafinil, 50.2 Early-Onset Bipolar Disorder

Patrick D. Skosnik, Ph.D. Professor, Northeastern University, Bouvé College of Health Sciences, Schools of Nursing and Pharmacy/Pharmaceutical Sciences, Boston, Masachusetts
1.26 Electroencephalography in Psychiatry

Mark Slifstein, Ph.D. Professor of Psychiatry, Director of Research PET, Department of Psychiatry and Behavioral Health, Stony Brook University, Renaissance School of Medicine, Stony Brook, New York
1.27 Radiotracer Imaging with Positron Emission Tomography and Single Photon Emission Computed Tomography: Fundamental Principles, Methodology and Role in Neuropsychiatric Research

Gary W. Small, M.D. Professor, Chair, Behavioral Health Physician-in-Chief, Department of Psychiatry and Behavioral Health, Hackensack Meridian School of Medicine, Hackensack University Medical Center, Hackensack Meridian Health, Hackensack, New Jersey
55.3f Alzheimer Disease and Other Neurocognitive Disorders

Tawny L. Smith, Pharm.D., B.C.P.P. Associate Professor, Department of Psychiatry and Behavioral Sciences, University of Texas, Dell Medical School, Austin, Texas
33.15a Melatonin Receptor Agonists

Yvette Smolin, M.D., M.Ed. Director of Consultation-Liaison Psychiatry, Department of Psychiatry, Westchester Medical Center/New York Medical College, Valhalla, New York
27.14 Psychiatric Care of the Burned Patient

Ingrid Söchting, Ph.D., R.Psych., C.G.P. Clinic Director and Clinical Professor, Psychology and Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada
55.4k Group Therapy

Takahiro Soda, M.D., Ph.D. Assistant Professor of Psychiatry, Department of Psychiatry, Division of Child and Adolescent

Psychiatry, University of Florida, College of Medicine/University of Florida Health, Gainesville, Florida
54.7 Ethical Issues in Child and Adolescent Psychiatry

LalithKumar Solai, M.D. Professor of Psychiatry/Vice-Chair, Clinical Services, Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, College of Medicine, Charleston, South Carolina
10.2 Delirium

Adrian N. Sondheimer, M.D. Clinical Assistant Professor, Psychiatry Department, University at Buffalo, Jacobs School of Medicine and Biomedical Sciences, Buffalo, New York
54.7 Ethical Is 1.3 Functional Genomics of Human Brain Developmenthiatry

Reneé M. Sorrentino, M.D. Clinical Assistant Professor, Part-time, Department of Psychiatry, Massachusetts General Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
21.3 Paraphilic Disorders

Traci J. Speed, M.D., Ph.D. Assistant Professor, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
27.10 Stress and Psychiatry

Henry I. Spitz, M.D., D.F.A.P.A., D.F.A.G.P.A. Clinical Professor of Psychiatry, Department of Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons, New York, New York
32.20 Family and Couples Therapy, 32.21 Group Psychotherapy

Susan Spitz, A.C.S.W., C.G.P. Clinical Instructor of Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons, New York, New York
32.20 Family and Couples Therapy

Stephen M. Stahl, M.D., Ph.D., D.Sc. (Hon.) Adjunct Professor of Psychiatry, Department of Psychiatry, University of California, School of Medicine, San Diego, California
33.7a Pimavanserin

Dan J. Stein, M.D., Ph.D. Professor and Chair, Department of Psychiatry and Neuroscience Institute, Groote Schuur Hospital/University of Cape Town, Cape Town, South Africa
15 Obsessive-Compulsive and Related Disorders

Murray B. Stein, M.D., M.P.H. Distinguished Professor of Psychiatry and Public Health, Department of Psychiatry, University of California, School of Medicine, San Diego, California
14.7 Somatic Therapies for Anxiety Disorders

Erika Steinbrenner, M.D. Psychiatrist, Imagine Healthcare, Chicago, Illinois
32.18 Hallucinogen-Assisted Psychotherapy

Kayla D. Stone, Ph.D. Postdoctoral Associate, Department of Psychiatry, Cumming School of Medicine, Alberta Children’s Hospital / University of Calgary, Calgary, Alberta, Canada
37 Neuroimaging in Psychiatric Disorders of Childhood

Cynthia M. Stonnington, M.D. Professor of Psychiatry, Psychiatry and Psychology, Mayo Clinic, Scottsdale, Arizona
30.8 Physician and Medical Student Mental Health

Martin T. Strassnig, M.D. Adjunct Associate Professor of Psychiatry, Psychiatry and Behavioral Sciences, University of Miami Leonard M. Miller School of Medicine, Miami, Florida
33.2a Dopamine Antagonists

Jeffrey R. Strawn, M.D. Professor of Psychiatry, Pediatrics and Clinical Pharmacology; Co-Director, Center for Clinical and Translational Science and Training, Department of Psychiatry &

Behavioral Neuroscience, University of Cincinnati College of Medicine, Cincinnati, Ohio
33.18a Pediatric Psychopharmacology, 53.5 Pediatric Psychopharmacology

Joan M. Striebel, M.D. Assistant Clinical Professor, Department of Psychiatry and Biobehavioral Sciences, University of California, David Geffen School of Medicine, Los Angeles, California
33.7a Pimavanserin

Shannon F. Stromberg, M.D. Medical Director and Chief of Psychiatry, Psychiatry and Behavioral Health Program, Presbyterian Healthcare System and Medical Group, Albuquerque, New Mexico
30.3 Physical Violence on Adults

Mandayam Sujata, M.D. Staff Psychiatrist, Department of Psychiatry, Western Psychiatric Hospital of UPMC, Pittsburgh, Pennsylvania
10.2 Delirium

Joji Suzuki, M.D. Director, Division of Addiction Psychiatry, Brigham and Women's Hospital, Boston, Massachusetts
11.1 Substance Use Disorders: Introduction, 11.6 Hallucinogen-Related Disorders

D.W. Svenson, M.A. Doctoral Candidate, Psychology/Behavioral, Cognitive Neuroscience, Wayne State University, School of Medicine/College of Liberal Arts and Sciences, Detroit, Michigan
11.7 Inhalant-Related Disorders

Dragan M. Svrakic, M.D., Ph.D. Professor, Department of Psychiatry, Washington University in St. Louis, School of Medicine; Staff Psychiatrist, VA Medical Center, St. Louis, Missouri
3.6 Neuroscience of Psychoanalytic Object Relations Theory

Rex M. Swanda, Ph.D., A.B.P.P.-C.N. Clinical Assistant Professor, Department of Psychiatry and Behavioral Sciences, University of New Mexico, School of Medicine, Albuquerque, New Mexico
7.4 Neuropsychological Assessment of Adults

Mary Margaret Sweeney, Ph.D. Assistant Professor, Department of Psychiatry and Behavioral Sciences, Behavioral Pharmacology Research Unit, Johns Hopkins University School of Medicine, Baltimore, Maryland
11.4 Caffeine-Related Disorders

Roxy Szeftel, M.D. Clinical Professor, Department of Psychiatry and Biobehavioral Sciences, University of California, David Geffen School of Medicine, Los Angeles, California
54.14 Internet Gaming Disorder in Youth, 54.15 Social Media and Youth

Zebulon Taintor, M.D. Adjunct Professor, Department of Psychiatry, New York University, Grossman School of Medicine, New York, New York
7.9 Electronic Media in Psychiatry

Carol A. Tamminga, M.D. Professor and Chair, Department of Psychiatry, University of Texas Southwestern, Medical School, Dallas, Texas
12.1 Introduction and Overview, 12.16 Psychosis as a Defining Dimension in Schizophrenia

Evan Taniguchi, M.D. Acting Assistant Professor, Department of Psychiatry and Behavioral Sciences/Department of Psychiatry and Behavioral Medicine, University of Washington, School of Medicine/Seattle Children's Hospital, Seattle, Washington
39 Intellectual Disability

Erica Taylor, M.D., M.P.H. Psychiatrist, Assistant Professor, Department of Psychiatry, MD Anderson Cancer Center, Houston,

Texas
56.3 The Role of the Hospital in the Care of Persons with Mental Illness

Michael E. Thase, M.D. Professor, Department of Psychiatry, University of Pennsylvania, Raymond and Ruth Perelman School of Medicine and Corporal Michael J. Crescenz Veterans Affairs Medical Center, Philadelphia, Pennsylvania
13.7 Mood Disorders: Psychotherapy, 13.8 Mood Disorders: Neurobiology, 33.5a.2 Serotonin-Norepinephrine Reuptake Inhibitors (SNRIs) 33.6c Mirtazapine

Mikiko Thelwell, M.D., M.P.H. Resident, Department of Psychiatry, University of Pennsylvania, Raymond and Ruth Perelman School of Medicine, Philadelphia, Pennsylvania
54.16 Gay, Bisexual, Transgender, Queer/Questioning, Intersex, and Asexual (LGBTQIA) Youth

Michelle M. Theodory, B.A. Post-baccalaureate IRTA, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland
14.3 Epidemiology of Anxiety Disorders

Margo L. Thienemann, M.D. Clinical Professor of Psychiatry, Department of Psychiatry and Behavioral Sciences, Stanford University, School of Medicine, Palo Alto, California
53.3 Group Psychotherapy

Giulio Tononi, M.D., Ph.D. Professor of Psychiatry, Department of Psychiatry, University of Wisconsin, School of Medicine and Public Health, Madison, Wisconsin
1.15 Basic Science of Sleep

Marika Toscano, M.D. Physician, Department of Obstetrics and Gynecology, University of Rochester Medical Center, Rochester, New York

30.1 Psychiatry and Reproductive Medicine

Karen E. Toth, Ph.D. Clinical Faculty in Psychology, Department of Psychology, University of Washington, School of Medicine, Seattle, Washington
39 Intellectual Disability

Glenn Jordan Treisman, M.D., Ph.D. Eugene Meyer III Professor of Psychiatry and Medicine, Departments of Psychiatry and Behavioral Sciences and Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland
2.8 Neuropsychiatric Aspects of HIV Infection and AIDS

Manuel Trujillo, M.D. Clinical Professor of Psychiatry, New York University, Grossman School of Medicine, New York, New York
32.5 Intensive Short-Term Psychodynamic Psychotherapies

Jacqueline Trumbull, B.A., Ph.D. Student, Department of Clinical Psychology, Duke University, School of Medicine, Durham, North Carolina
32.11 Dialectical Behavior Therapy

Thanh Thuy Truong, M.D. Assistant Professor of Psychiatry and Behavioral Sciences, Menninger Department of Psychiatry and Behavioral Sciences, Baylor University, College of Medicine/The Menninger Clinic, Houston, Texas
2.15 Catatonia

Susan Beckwitt Turkel, M.D. Emerita Associate Professor of Psychiatry, Department of Psychiatry, Division of Child-Adolescent Psychiatry, Children’s Hospital Los Angeles/University of Southern California Keck School of Medicine, Los Angeles, California
54.3 Children’s Reaction to Illness and Hospitalization

Catharyn A. Turner, II, M.D., M.Ed. Medical Director, Acute Inpatient Psychiatry, Department of Child and Adolescent Psychiatry

and Behavioral Sciences/Integrated Division, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania
50.1 Depressive Disorders and Suicide

Hiroyuki Uchida, M.D., Ph.D. Professor, Department of Neuropsychiatry, Keio University School of Medicine, Tokyo, Japan
33.10b Gabapentin, Topiramate, Tiagabine, Levetiracetam, Zonisamide, and Pregabalin

Juan Carlos Urizar, M.D. Medical Director of Geriatric Psychiatry/Instructor in Psychiatry, Division of Geriatric Psychiatry, Brigham and Women's Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
2.2 Neuropsychiatric Aspects of Cerebrovascular Disorders

Robert J. Ursano, M.D. Professor, Director, Center for the Study of Traumatic Stress, Uniformed Services University of the Health Sciences, F. Edward Hébert School of Medicine, Bethesda, Maryland
30.6 Disaster Psychiatry: Disasters, Terrorism, and War

David R. Vago, Ph.D. Visiting Faculty/Associate Professor, Contemplative Sciences Center, University of Virginia, School of Medicine, Charlottesville, Virginia
1.29 Neuropsychiatry of Self

Ipsit V. Vahia, M.D. Associate Chief, Department of Geriatrics, McLean Hospital/Harvard University, Harvard Medical School, Belmont, Massachusetts
55.6h Technology for Seniors

George E. Vaillant, M.D. Professor of Psychiarty, Harvard University, Harvard Medical School, Boston, Massachusetts
3.5 Normality and Mental Health

Jim van Os, M.D., Ph.D. Professor of Psychiatric Epidemiology and Public Mental Health, Division Neuroscience, Utrecht University

Medical Centre, Utrecht, The Netherlands
12.5 The Clinical Epidemiology of Psychosis Spectrum Disorder

Pieter Joost Van Wattum, M.D., M.A. Associate Clinical Professor of Child, Adolescent and Adult Psychiatry, Child Study Center, Yale University School of Medicine, New Haven, Connecticut
54.12 Impact on Parents of Raising a Child with Psychiatric Illness and/or Developmental DisabilityAaron Van Wright, M.D. Clinical Associate Professor of Psychiatry, Department of Psychiatry, University of Texas Southwestern, Medical School, Dallas, Texas

32.23 Other Methods of Psychotherapy

Daniel Varon, M.D. Assistant Professor of Psychiatry, Department of Psychiatry, Western Psychiatric Hospital, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania
10.4 Mild Cognitive Impairment and Amnestic Disorders

Jayson Varughese, M.D., M.S. Adult and Geriatric Psychiatrist, HRC Behavioral Health and Psychiatry, PA, Chapel Hill, North Carolina
55.3g Delirium

Yvonne E. Vaucher, M.D., M.P.H.J. Clinical Professor of Pediatrics, Division of Neonatology, University of California, School of Medicine, San Diego, California
2.13 Psychiatric Aspects of Child Neurology

Erin P. Vaughan, M.A. Graduate Student, Department of Psychology, Louisiana State University, Baton Rouge, Louisiana
45 Disruptive Behavior Disorders in Children and Adolescents

Dennis Velakoulis, M.B.B.S., M.P.M., Dip.Crim., D.Med.Sci., F.R.A.N.Z.C.P. Professor and Director, Neuropsychiatry, Royal Melbourne Hospital, Melbourne, Australia
2.14 Neuropsychiatry of Neurometabolic and Neuroendocrine Disorders

Antonio Ventriglio, M.D., Ph.D. Clinical Researcher, Department of Clinical and Experimental Medicine, University of Foggia, Foggia, Italy
60 World Aspects of Psychiatry

Marcy Verduin, M.D. Associate Dean for Students, Professor of Psychiatry, Departments of Medical Education and Clinical Sciences, University of Central Florida College of Medicine, Orlando, Florida
21.5 Sexual Addiction

John T. Vernon, M.D. Assistant Professor of Psychiatry, Virginia Commonwealth University, School of Medicine, Richmond, Virginia
33.6a Trazodone

Vijay Vishwanath, M.D., Ph.D. Clinical Assistant Professor, Pediatric Neurology, Children's Hospital Los Angeles/University of Southern California Keck School of Medicine, Los Angeles, California
41 Developmental Coordination Disorder

Shayne Vitemb, M.A., L.M.F.T., C.G.P. Clinical Psychology Licensed Marriage and Family Therapist, Private Practice, El Segundo, California
55.4k Group Therapy

Michael Vitulano, Ph.D. Clinical Instructor, Yale Child Study Center, Yale University School of Medicine, New Haven, Connecticut
35.1 Psychiatric Examination of the Infant, Child, and Adolescent

Giacomo Vivanti, Ph.D. Associate Professor, A.J. Drexel Autism Institute, Drexel University, College of Medicine, Philadelphia, Pennsylvania
43 Autism Spectrum Disorders

Fred R. Volkmar, M.D. Irving B. Harris Professor of Child Psychiatry, Pediatrics, and Psychiatry, Child Study Center, Yale New Haven

Hospital, Yale University School of Medicine, New Haven, Connecticut
43 Autism Spectrum Disorders

Joel L. Voss, Ph.D. Associate Professor, Department of Medical Social Sciences, Neurology, and Psychiatry, Northwestern University, Feinberg School of Medicine, Chicago, Illinois
3.4 Biology of Memory

Karen Dineen Wagner, M.D., Ph.D. Professor and Chair of Psychiatry, Department of Psychiatry and Behavioral Sciences, University of Texas Medical Branch, John Sealy School of Medicine, Galveston, Texas
50.1 Depressive Disorders and Suicide

Harold J. Wain, Ph.D. Professor in Psychiatry, Department of Psychiatry, (USUHS) Uniformed Services University of the Health Sciences, F. Edward Hébert School of Medicine, Bethesda, Maryland
32.19 Hypnosis

Milton Wainberg, M.D. Professor of Clinical Psychiatry, Department of Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons, New York, New York
56.4 Mental Health Services Research

B. Timothy Walsh, M.D. Ruane Professor of Psychiatry, Department of Psychiatry, Columbia University, Vagelos College of Physicians and Surgeons/Irving Medical Center, New York, New York
22 Feeding and Eating Disorders

Patricia D. Walshaw, Ph.D. Associate Clinical Professor of Psychiatry, Department of Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
35.3 Neuropsychological and Psychoeducational Assessment of Children

Mark Walterfang, M.B.B.S. (Hons), Ph.D., F.R.A.N.Z.C.P. Clinical Professor of Neuropsychiatry, Department of Neuropsychiatry, Royal Melbourne Hospital; Department of Psychiatry, University of Melbourne, Melbourne, Australia
2.14 Neuropsychiatry of Neurometabolic and Neuroendocrine Disorders

Danuta Wasserman, M.D., Ph.D. Professor of Psychiatry and Suicidology, National Centre for the Prevention of Suicide and Mental Ill-Helath (NASP), Karolinska Institutet, Stockholm, Sweden
31.1 Suicide: Overview of Epidemiology, Risk and Protective Factors, Treatment, and Prevention

Khadijah Booth Watkins, M.D., M.P.H. Associate Director, Child and Adolescent Psychiatry Residency Training, Massachusetts General Hospital/McLean Hospital; Associate Director, The Clay Center for Young Healthy Minds, Child and Adolescents Psychiatry Department, Massachusetts General Hospital/Harvard University, Harvard Medical School, Boston, Massachusetts
53.8 The Treatment of Adolescents

Clarence Watson, J.D., M.D. Clinical Associate Professor of Psychiatry, Department of Psychiatry, University of Pennsylvania, Raymond and Ruth Perelman School of Medicine, Philadelphia, Pennsylvania
58.1 Clinical-Legal Issues in Psychiatry

Kathleen Watson, Ph.D. Postdoctoral Scholar, Department of Psychiatry and Behavioral Health, Stanford University, School of Medicine, Palo Alto, California
27.7 Endocrine and Metabolic Disorders

Linda E. Weinberger, Ph.D. Professor Emerita of Clinical Psychiatry and the Behavioral Sciences, University of Southern California, Keck School of Medicine, Los Angeles, California
56.8 Criminalization of Persons With Serious Mental Illness

Margaret D. Weiss, M.D., Ph.D. Associate Professor of Psychiatry, Harvard University, Department of Psychiatry, Cambridge Health Alliance, Cambridge, Massachusetts
44.1 Attention-Deficit/Hyperactivity Disorder

Roger D. Weiss, M.D. Professor of Psychiatry; Chief, Division of Alcohol, Drugs, and Addiction, Division of Alcohol, Drugs, and Addiction, McLean Hospital/Harvard University, Harvard Medical School, Belmont, Massachusetts
11.3 Stimulant-Related Disorders

Abraham Weizman, M.D. Professor of Psychiatry, Department of Psychiatry, Geha Mental Health Center, Petah Tikva and Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
33.12a Histamine Blockers and Enhancers

Chelsey A. Werchan, Ph.D. Clinical Psychologist, Department of Behavioral Health, Parkland Health System, Dallas, Texas
27.5 Respiratory Disorders

James C. West, M.D. Associate Professor of Psychiatry, Department of Psychiatry, Uniformed Services University of the Health Sciences, F. Edward Hébert School of Medicine, Bethesda, Maryland
30.6 Disaster Psychiatry: Disasters, Terrorism, and War

Pariya L. Fazeli, Ph.D. Associate Professor, School of Nursing, University of Alabama, Marnix E. Heersink School of Medicine, Birmingham, Alabama
55.3a Assessment of Functioning

Heather Williams, M.D. Child Abuse Pediatrician, Department of Pediatrics, WakeMed Children’s Hospital/SAFEchild Advocacy Center, Raleigh, North Carolina
54.2 Child Maltreatment

Katherine E. Williams, M.D. Clinical Professor of Psychiatry, Director, Women's Wellness Clinic, Stanford University, School of Medicine, Palo Alto, California
27.7 Endocrine and Metabolic Disorders

Cynthia Wilson, M.D. Assistant Professor of Clinical Psychiatry, Psychiatry and the Child Study Center, Yale University, School of Medicine, New Haven, Connecticut
53.6 Inpatient Psychiatric, Partial Hospitalization, and Residential Treatment for Children and Adolescents

Gerald Scott Winder, M.D., M.Sc. Clinical Associate Professor of Psychiatry, Surgery, and Neurology Departments of Psychiatry, Surgery, and Neurology, University of Michigan, Medical School, Ann Arbor, Michigan
27.13 Organ Transplantation

Dean F. Wong, M.D., Ph.D. Professor, Department of Radiology, Psychiatry, Neurology, and Neuroscience, Washington University in St. Louis, School of Medicine, St. Louis, Missouri
12.9 Molecular Brain Imaging in Schizophrenia

Dora Wang, M.D., M.A. Psychiatrist, Palo Alto, California
19 Factitious Disorder

Ma-Li Wong, M.D., Ph.D. Professor, Department of Psychiatry, State University of New York, Upstate Medical University, Alan and Marlene Norton College of Medicine, Syracuse, New York
1.21 Pharmacogenetics

Harrell Woodson, Ph.D. Assistant Professor/Vice President of Clinical Services, Department of Psychiatry/Chief of Staff's Office, Baylor University, College of Medicine/Menninger Clinic, Houston, Texas
28.2 Adult Antisocial Behavior, Criminality, and Violence, 3.2 Piaget and Cognitive Development: History, Theory, Extensions, and

Modifications

Chester Wu, M.D. Psychiatrist, Department of Psychiatry and Sleep Medicine, Baylor College of Medicine, Houston, Texas
1.14 Chronobiology, Circadian Rhythm, and Psychiatry

Timothy P. Wuerz, D.O., M.Sc. Assistant Professor, Department of Psychiatry, Charles R. Drew University of Medicine and Science, Los Angeles County Department of Mental Health, Los Angeles, California.
2.12 Neuropsychiatric Aspects of Neuromuscular Diseases

Lakshmi N. Yatham, M.B.B.S., F.R.C.P.C.S., M.R.C.Psych(UK), M.B.A.(Exec) Professor and Head, Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada
13.6 Mood Disorders: Pharmacologic Treatment of Depression and Bipolar Disorders

Larry J. Young, Ph.D. William P. Timmie Professor, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia
1.8 Neuropeptides: Biology, Regulation and Role in Neuropsychiatric Disorders

Lamyaa Yousif, M.D., Ph.D., M.Sc. Senior DSM Operations Manager and Research Associate, Division of Research, American Psychiatric Association, Washington, DC
5.1 Quantitative and Experimental Methods in Psychiatry: Epidemiology

Ryan K. C. Yuen, Ph.D. Senior Scientist, Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada
43 Autism Spectrum Disorders

Carlos A. Zarate, Jr., M.D. NIH Distinguished Investigator, Chief, Experimental Therapeutics and Pathophysiology Branch, Intramural

Research Program, National Institute of Mental Health, Bethesda, Maryland
33.4a Dopamine Receptor Agonists, 33.14a Esketamine and Ketamine

Charles H. Zeanah, M.D. Professor of Psychiatry and Pediatrics, Section of Child and Adolescent Psychiatry, Tulane University School of Medicine, New Orleans, Louisiana
49.1 Reactive Attachment Disorder and Disinhibited Social Engagement Disorder

Sasson Zemach, M.D. Doctor of Psychiatry, Eitanim Campus, The Jerusalem Mental Health Center, Affiliated With the Hebrew University and the Hadassah School of Medicine, Jerusalem, Israel
33.12a Histamine Blockers and Enhancers, 33.19a Essential List of Medicines, 33.1c Classification of Psychotropics: Neuroscience-Based Nomenclature (NbN)

Douglas M. Ziedonis, M.D., M.P.H. Distinguished Professor of Psychiatry/Executive Vice President, Health Sciences and CEO, UNM Health System, Department of Psychiatry, UNM Health/UNM Health Sciences/University of New Mexico, School of Medicine, Albuquerque, New Mexico
11.8 Tobacco-Related Disorders

Bonnie T. Zima, M.D., M.P.H. Professor-in-Residence, Department of Psychiatry, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, David Geffen School of Medicine, Los Angeles, California
54.10 Child Mental Health Services Research

Joseph Zohar, M.D. Professor, Director, Post-Trauma Center, Sheba Medical Center, Ramat Gan, Israel
33.19a Essential List of Medicines, 33.1c Classification of Psychotropics: Neuroscience-Based Nomenclature (NbN)

Charles F. Zorumski, M.D. Samuel B. Guze Professor of Psychiatry, Professor of Neuroscience, Department of Psychiatry, Washington University in St. Louis, School of Medicine, St. Louis, Missouri
1.5 Cellular and Synaptic Basis of Neural Signaling, 3.6 Neuroscience of Psychoanalytic Object Relations Theory

Jon-Kar Zubieta, M.D., Ph.D. Professor, Department of Psychiatry, Northwell Health, Port Jefferson, New York
1.16 Pain Systems: Interface with Affective and Motivational Mechanisms

---

*Deceased.

Preface

The Comprehensive Textbook of Psychiatry was conceived in the 1960s when Harold Kaplan, MD, saw the need for a scholarly textbook in the style of the great medical textbooks. He published the first edition in 1967. Then, it was one volume with about 600 pages. However, as you see, it has grown considerably!

Much of the book's growth occurred under the remarkable careers of Benjamin and Virginia Sadock, both psychiatrists at the New York University School of Medicine, who coedited many of the volumes with Dr. Kaplan and then took on the book's leadership after Dr. Kaplan's death in 1998. They continued updating and revising the book through the 10th edition, published in 2017.

Over the book's course, they oversaw an expansion to two volumes and more than 5,000 pages, with more than 600 contributors. This growth reflects the fantastic expansion of the field of psychiatry, with advances in neuroscience, pharmacology, psychological theory and practice, cultural psychiatry, sociology, and many other affiliated areas. To accommodate this, the Sadocks organized the book into different subject area sections, each with an expert contributing editor.

With the Sadocks' retirement, the publishers sought a new set of editors who would respect the mission of these books. We are honored to have been chosen and appreciate the faith the Sadocks have put in us to continue their wonderful tradition. Long-time consumers of the Kaplan and Sadock series (one of us remembers buying the fifth edition during residency [RB]) and then the Sadock and Sadock series of books, we understand all too well the care and thought that has gone into both the organization and writing of these volumes. With that in mind, we have chosen an incremental approach. Thus, whereas some previous editions saw dramatic changes in subject matter or organization, this book will look similar to the last volume, at least superficially. Some chapters have been added and some removed, but the organization remains unchanged. The reader will find a few notable exceptions to this rule, where advances had outpaced the previous edition's approach. For example, readers will find significant changes in the organization and chapters in our biologic treatments section and the cultural and global psychiatry content.

The most significant change is the knowledge contained. Psychiatry is moving ever-increasingly, making this an exciting time to be in the field. We asked each author to update their content and review the latest research in their field. At times this has meant adding or replacing authors. Indeed, a sad part of our jobs was learning of the passing of several distinguished authors. This unfortunate experience was perhaps made more common during this ongoing pandemic. In many cases, the new authors retained much of the format and detail from the previous volumes and acknowledged their predecessors' work in their new chapters.

One practical update is in psychiatric classification. Both the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM) and the World Health Organization's International Classification of Diseases (ICD) have had updates since the last edition of this textbook. In the case of the DSM, the change is relatively modest, representing a textual revision of the fifth edition (DSM-5-TR). The ICD has undergone a more significant revision, updating much of the descriptions and nomenclature for many diseases to reflect advancing medical science. The WHO approved ICD-11 in 2019 and committed to transitioning to the newer system in 2022. For mental health professionals, the most significant change is in their efforts to align ICD with DSM, with more similar terminology and approaches to mental diseases than the previous versions. This revision will simplify things, particularly for hospital systems that rely on the ICD for classification and billing. However, the United States has classically been slow to adopt ICD changes, taking about 10 years for the last one. The United States hopes to transition to the newer nomenclature by 2025, although its full adoption may take longer. You can read more about these approaches and the results in this book's section on Classification in Psychiatry.

The book's many other revisions are too many to relate here, but they will be evident as you read it. Hopefully, readers will also appreciate the book's continued attempt, long championed by the Sadocks, to keep the book readable. As editors, we agree with the maxim, usually attributed to Einstein, that "If you can't explain it simply, you don't understand it well enough." However, we are also mindful of the other Einstein attribution, "Everything should be made as simple as possible, but no simpler." With that in mind, our contributors and we worked hard to make the book as readable as possible while maintaining its high scientific rigor. We wanted this book to be more than a compendium of facts, but rather something we would want to sit down and read—if not cover to cover, then at least through an entire chapter. Being the first to read the book as we edited it, we believe the many authors who contributed to this work have done an awe-inspiring job achieving that. We learned a great deal and enjoyed doing it. At least most of the time.

One aspect of the challenge to balance rigor and readability is the decision to omit within-text citations and limit the references. This approach has been a longstanding policy of this and many other standard textbooks; however, it made many of our authors uncomfortable. Their concern is that readers will see their chapter as unscientific "pop science" or just their opinion. Some even worried about ethical considerations: whether it would appear that they were presenting the work of others as if it was theirs. We trust that the readers will understand otherwise. We all intend each chapter to be a rigorous review of existing literature. Usually, a complete reflection of the many sources contributing to each chapter would require hundreds, if not thousands, of citations and references, making the result unwieldy. We trust that the readers will understand that each chapter represents the authors' synthesis of the many sources they reviewed to create a coherent work. Each chapter does list critical references at the end for further study. However, in most cases, many more references than we can include contribute to the ideas and facts in a chapter. We accept that this is the nature of works that attempt to review massive fields of study. Hopefully, readers will appreciate the reasons for and results of this approach. If you do not, please understand that this decision was ours alone.

The limitation in this approach, we realize, is somewhat mitigated by the connected world we live in, where interested readers can quickly and easily search for relevant papers online—accessibility likely to be broadened by the advances in artificial intelligence. Such intelligent systems are already producing reviews of complicated subjects, albeit with mixed results. Will that field advance so dramatically that we will not need further editions of this book? We doubt it. When we reflect on the most important influences in our education as psychiatrists, we do not recount the countless computer screens we have viewed. Instead, we remember the fantastic psychiatrists and other experts who taught us throughout our training and careers. We are delighted to say that we had the opportunity to include some of them in this book. Although the written word will never equal the power of a human relationship, we hope that the chance to sit down and read the thoughts of the many expert minds found here will be an experience we all can appreciate long into the future.

OUR MISSION STATEMENT

Our mission is to educate mental health and related professionals in contemporary psychiatry.

We emphasize the word "educate." Today, there is no end to information sources. What we seek here is not one more repository of facts but rather a system of education that is complete, authoritative, and accessible. This book represents a crucial part of that mission, being the most comprehensive of the books. However, it is part of a series of books devoted to psychiatric education. Also included are the *Synopsis of Psychiatry*, a distillation of this book into a single volume, and the *Concise Textbook of Psychiatry*, which further reduces the *Synopsis*, concentrating on the most clinically relevant sections of the book. Also related are books devoted to subspecialties in psychiatry, most notably the *Concise Textbook of Child and Adolescent Psychiatry*, edited by child psychiatrist Caroly Pataki, M.D. We are considering other specialty volumes as well. Finally, the *Study Guide and Self-Examination Review of Psychiatry* are written by Eric Williams, M.D., and Lindsay Moskowitz, M.D., both respected educators. That volume, used independently or alongside the *Synopsis*, represents a further example of our commitment to education. One of the best ways to learn new material is through self-assessment, and we produced the *Study Guide* with that purpose in mind.

Acknowledgments

Over time, thousands of psychiatric, mental health, and other experts have contributed to this book. Many are long gone, but they have put their mark on the current volume, and we appreciate them all. We also thank the hundreds of authors who often endured work pressures, the pandemic, and our nagging as they took on the challenge of summarizing their fields. They put in an incredible effort, the results are remarkable, and we gratefully thank them all.

We give special thanks to the section editors of this book. A book of this size is well beyond our abilities to review as thoroughly as we would wish, and these editors took on the lion’s share of the oversite in their sections. These editors include Iqbal “Ike” Ahmed, M.D., Gaston Baslet, M.D., Carlyle Chan, M.D., Dilip Jeste, M.D., John Krystal, M.D., Roberto Lewis-Fernández, M.D., Kathleen Merikangas, Ph.D., Christine Yu Moutier, M.D., Caroly Pataki, M.D., Joji Suzuki, M.D., Carol Tamminga, M.D., Sasson Zemach, M.D., and Joseph Zohar, M.D. It would be a wild understatement to say that we could not have done this without them, and we owe them all a lasting debt of gratitude.

We owe a similar debt to the publishing team at Wolters Kluwer. Particular thanks go to Chris Teja, our Acquisitions Editor; Anne Malcolm, the Associate Director for Content Development; Ariel S. Winter, our Development Editor; and most of all, Sean Hanrahan and Erin Hernandez, Editorial Coordinators, who had the dubious pleasure of being in almost daily communication with us.

We want to acknowledge our respective institutions and universities, The Menninger Clinic, The Baylor College of Medicine, and the University of Central Florida, for their enduring support in our careers.

Finally, we remain always thankful to Ben and Virginia Sadock, who are wonderful people and great scholars, for creating what we consider the finest textbooks in psychiatry and trusting us to continue the tradition.

Foreword

A BRIEF HISTORY

Harvey Cushing (1869–1939), the father of modern neurosurgery wrote a biography of Sir William Osler for which he won a Pulitzer Prize. In that book he described Osler's writing of his famous *Textbook of Medicine*. Osler devoted his life to that magnum opus which prompted Cushing to write the following: *The birth of a successful textbook, like that of a child, may hold its authors in unexpected bondage*. That saying resonated with us.

The first edition of the *Comprehensive Textbook of Psychiatry* (CTP) was published in 1967, over 55 years ago. Now in its 11th edition, it is the most widely used psychiatric textbook in this country and around the world, translated into several languages. The book has been well received not only by mental health professionals but also by reviewers. The *Journal of the American Medical Association* published many reviews of the book using terms like "well-organized, up to date, comprehensive" among other laudatory phrases. It has been called "a university without walls."

A foreword is rarely written by an author or editor of a previous edition of a book. So, we were surprised, pleased, and honored to be offered this opportunity by the editors. It gave us a chance to reflect on our work. To help us we decided to review the forewords written by others and the prefaces written by us. This would help jog our memories. We decided to take excerpts from some of the forewords to let the reader know how the editions evolved and how the issues facing the field changed over these many years.

The book began as the brainchild of Harold Kaplan, M.D. (1928–1998) who wanted to develop a textbook in psychiatry that would rival the great textbooks of medicine like *Nelson's Textbook of Pediatrics* or the *Cecil Textbook of Medicine*. He eventually achieved this objective but not before several rejections. One major medical publishing house rejected the project commenting that "trying to get a group of psychiatrists to collaborate would be like trying to herd cats." Years later, that same publisher said it was the biggest mistake they had ever made. One editor did accept the proposal, Chuck Reville at Williams & Wilkins (now Wolters Kluwer).

Unfortunately, Harold passed away in 1998 after the sixth edition was published, and we dedicated the seventh edition to him:

> In addition to his academic career, Dr. Kaplan was a clinical psychiatrist who maintained a large and active practice. He believed strongly in the art and science of psychotherapy, and although trained as a psychoanalyst, he followed no particular school. He was eclectic in the true sense of the word; he applied the best of each school in his work with patients. His vast clinical experience allowed him to create a textbook of value to basic behavioral scientists and more importantly, to clinicians.
>
> As an editor, Dr. Kaplan was a hard task master who demanded perfection. He encouraged each contributor, exhorted them when necessary, and expected and received their best work.

In Harold's last sixth edition, the following was written about the editor to be.

> In all our previous books, she (VS) served as an assistant to the editors and actively participated in every editorial decision. Her enthusiasm, sensitivity, comprehension, and depth of psychiatric knowledge were of immeasurable importance. She has ably represented not only the viewpoint of women in medicine but has also made many contributions to the content of this textbook.

The text could not have come into being or have continued without our contributors. We ended each preface thanking them for their cooperation and participation. As each edition appeared and the reputation of the text became more apparent, it was easier to recruit contributors. We sought experts in their respective fields who were often persons who developed or originated an area of study. Aaron Beck is one example. Having developed cognitive behavioral therapy, he wrote the chapter on CBT for the textbook. Arvid Carlsson, the Nobel laureate who discovered dopamine was another.

Our practice was to invite contributors and stay in touch with them by telephone. That contact with them was effective in getting them to complete their sections on time, a major concern for editors of a multi-authored textbook. And those phone calls enabled us to get to know our contributors personally, a rewarding and valuable fringe benefit of editing this book.

To date over 2,000 psychiatrists, psychologists and behavioral scientists have written sections for the book. Some are long gone but their work lives on in the various editions of the book that can be accessed via the internet. We consider many of their contributions to be classics.

As the book expanded we added a third editor, Pedro Ruiz, (1936–2023) a close personal and professional associate, to the ninth edition. At the time he was Professor of Psychiatry at the University of Texas Medical School at Houston and President of the World Psychiatric Association. His participation immeasurably enhanced and facilitated our work.

The first Foreword did not appear until the fourth edition. It was written by Robert Cancro, M.D., who was at the time Professor and Chairman of the Department of Psychiatry at New York University. He analyzed the prior editions. The first brought structure and discipline into the field; the second and third increased our understanding of clinical psychiatry by including basic science; and the fourth integrated the application of neuroscience as crucial to understanding the problems of patients. He went on.

> It is awkward for this writer to acknowledge the debt to Harold Kaplan and Benjamin Sadock, because, in addition to being colleagues they are personal friends and members of the same faculty. It is an irony of Western culture that one can acknowledge indebtedness to a stranger more easily than to those who are near and dear. The time and effort demanded of them to make this book possible is known to only a handful of people.

In the eighth edition, John Nemiah (1918–2009), at the time Professor of Psychiatry at Harvard Medical School and long-standing editor of the *American Journal of Psychiatry* commented on an issue that is still present: the conflict between a psychoanalytic and psychodynamic approach versus a biologic approach to understanding human behavior. He was prescient when he wrote the following.

> Psychiatry has been reassessing its position. It is fortunate that the current self-reflection coincides with recent major advances in knowledge of the biological side of psychiatric illness and human behavior. The understanding of brain function is beginning to approximate the sophistication of psychological observations and theories, inspiring the hope that psychiatrists are coming closer to substantial correlations between those two complementary approaches. Already, the contributions of psychopharmacology have made it possible for psychiatrists to alleviate the symptoms of major psychoses that were generally beyond the reach of their skills. Biological knowledge, in other words, is beginning to restore balance that has too long tipped steeply toward psychology. There is danger in this too - a danger that the balance dip too far in the other direction and that psychiatrists forget the important gains that have been made in the psychological and social understanding of human behavior.

For the ninth edition of the textbook, Robert Michels, M.D., then University Professor of Medicine and Psychiatry at Weill Cornell Medical College, wrote the foreword. He entitled his contribution *The Future of Psychiatry*. He was concerned about the social and economic structure of the health care system, as are we. He covered broad areas: medicine in general; science and research; education and many others. A brief excerpt of his many interests follows.

> The social and economic structure of the healthcare system has lagged behind the development of knowledge and is currently the limiting factor in the quality of care available to most patients. The future promises a continuing growth of our knowledge and, particularly, an increased rate of its translation to clinically relevant tools. However, developments in the healthcare system are more difficult to predict and more problematic. Psychiatry is increasingly recognized as a full participant in medicine and health care and is unlikely to return to its former marginal status, as represented in the past by the asylum, the stigma associated with mental illness, and the woeful underfunding of psychiatric services. However, we continue to grapple for a more rational and effective healthcare system in the United States, and although the urgency is increased, the outcome is uncertain. The magnitude

of the problem suggests that larger social and political forces will determine the course, and the psychiatric profession will have to struggle merely to participate in the dialogue.

Psychiatry for the future will evolve in the context of the future use of these–of medicine, of the other mental health professions, of the scientific basis of psychiatric practice, of the healthcare system, of education, and of the organization of the profession. Each will change, and as it is so often said, the future is, therefore, hard to predict.

For the last edition the foreword was written by Herbert Pardes, M.D. At the time he was the Professor of Psychiatry and Dean of the Columbia College of Physicians and Surgeons. His interests were in the recent history of psychiatry, especially legislation that made mental health “more visible, accessible and patient oriented.” He mentioned the benefits that accrued to the mentally ill from the Community Mental Health Act of 1963 that led to deinstitutionalization and community health centers. His foreword emphasized emerging research into the cause and treatment of mental illness, and he regretted the trend to defund research and treatment programs. He focused on problems that exist to this day where psychiatry has a vital role to play.

A multitude of factors, that is, veterans’ problems, increased suicide, increased cost of comorbid psychiatric and nonpsychiatric conditions, mass violence along with recognition of the overwhelming numbers of individuals who both need care yet are winding up in prisons and jails, has given mental health unprecedented visibility.

The many editions of CTP reviewed here are a living history of psychiatry in the last half-century. There have been so many changes during this period, in the delivery of care, in our understanding of what causes mental illness, in treatment methods and in methods of investigation. Neuroscience, which occupied no more than 200 pages in the first edition is now covered in almost 1,000 pages. The first edition was about 1,600 pages in length, the last edition had close to 5,000 pages. Such was and is, the expansion of knowledge. Despite this, psychiatry is a frontier science with large gaps still to be filled. The connectome which has as its goal the mapping of more than 86 billion neurons of the brain is in its infancy. It may hold the clue that finally explains psychiatric disorders. New editions of this textbook will record these events.

One of the insights we had in reviewing the editions was an awareness of how much we missed Harold. Thinking about him 25 years after his death brought back many memories. It was more than a work relationship—it was a friendship of long standing. We went on vacations with our families while working at the same time.

This has been a labor of love, but after more than 50 years of working on this textbook it was time to release ourselves from bondage. The search for new editors to carry on our work resulted in the choice of Robert Boland, M.D. and Marcia Verduin, M.D., distinguished academicians. We had the opportunity to give them some advice, but they produced the current edition on their own. We are pleased to see that they organized the book based upon the solid foundation of the previous edition.

We said goodbye to “our child” with, as in all partings, feelings of sadness. But we take comfort that the book is in very excellent hands. They have done a superb job and we wish them both many more editions of this classic textbook.

Benjamin Sadock, M.D. and Virginia Sadock, M.D.
New York University Grossman School of Medicine

1

Neural Sciences

1.1 The Neuroscience of Psychiatry

John H. Krystal, M.D.

This section of the *Textbook* contains a remarkable collection of 30 chapters that provide an overview of the landscape of the current understanding of the neuroscience foundations of psychiatry. The brain is the principal organ of psychiatry, so understanding brain science is the starting point for a comprehensive knowledge of the field. These chapters bear careful reading as they reflect the first indications of the emergence of explanatory science within psychiatry. This progress arises from the application of technologies that have transformed our understanding of the brain and brain-behavior relationships, such as optogenetics, DREADDs, and CRISPR. It also arises from the maturation of fields, such as cognitive neuroimaging and psychiatric genetics, transforming them from intellectual curiosities to strategies to answer important questions about human behavior and mental illness.

What are the signs of this progress? The last issue of *The Comprehensive Textbook in Psychiatry* appeared in 2009. Thus a quote attributed to Bill Gates, known as “Gates Law” applies, “most people overestimate what they can achieve in a year, but underestimate what they can achieve in ten years.” There are many inspiring stories. With respect to etiology, advances in genome sequencing have enabled the discovery of growing numbers of rare gene variants that contribute significantly to the genetic risk for autism and schizophrenia. We are now beginning to understand in what regions of the brain and in which cell types these genes are preferentially expressed, as well as when in development, these genes are critically engaged. This knowledge establishes a foundation for understanding when, where, and how the development of the brain goes awry to give rise to the pathophysiology of psychiatric disorders. With the emergence of spatial transcriptomics, we will soon resolve our molecular understanding of pathophysiology to cells embedded in specific microcircuits within particular brain regions. As we have seen for Mendelian disorders, such as spinal muscular atrophy, identification of the molecular etiology may set the stage for disease-modifying interventions, such as gene therapy. With respect to pathophysiology, optogenetic studies in animals, neuroimaging studies in humans, and a new generation of intracortical physiology and stimulation studies have provided deep new insights into how depression and other psychiatric disorders might be expressed at the level of neural circuits. These “raw data” provide the basis for computational neuroscience studies that rigorously describe for the first time the relationships between the activity of neurons and networks or networks and behavior. In turn, these insights are providing a basis for a growing array of neurostimulation treatments for mood disorders. At the same time, some transformative advances in treatments, such as the discovery of the rapid antidepressant effects of ketamine and the psychedelics, have stimulated basic science advances that shed new light on the neurobiology of stress. In fact, since the last edition of this textbook, the FDA has approved the first two mechanistically novel antidepressants in over 50 years, Esketamine (treatment-resistant depression and depression in the context of increased suicide risk) and brexanolone (postpartum depression).

Why do psychiatrists need to understand neuroscience to practice their art? Psychiatry is moving away from viewing individual psychiatric patients solely as representative of static diagnostic categories, as might be reflected in the *Diagnostic and Statistical Manual*. One might predict that as psychiatrists become increasingly comfortable incorporating an understanding of the neurobiology of psychiatric disorders into their clinical formulations of individual patients, it will enrich their understanding of their patients and suggest new opportunities to enhance treatment.

CASE STUDY

The author will illustrate this point by referring to a case conference in which he participated, which he first presented in a blog to the American College of Neuropsychopharmacology in January 2012. The Yale Psychiatry Residents invited the author to interview a patient and then present his formulation to their group. The patient was a young man who suffered from long-standing psychotic symptoms. He met diagnostic criteria for schizophrenia, exhibiting many domains of preserved intellectual capacity but noted problems with sustained attention and memory. He also exhibited low levels of affective blunting but impairments in many social and functional domains. Upon interview, he also vividly described a childhood pattern of accentuated attention to nonsocial environmental details, and he both described and displayed perseverative tendencies and mild irritability when challenged. He also mentioned that both the diagnoses of schizophrenia and bipolar disorder ran in his family. With regard to treatment, he felt that clozapine was the most effective medication prescribed for him. In addition, he was treated with anticonvulsants, lamotrigine, and valproate.

How could psychiatrists understand the appearance of features of autism spectrum disorder (ASD) and his family history of both schizophrenia and bipolar disorder?

This patient presented opportunities to reflect on challenges in our diagnostic schema. Bleuler suggested that autistic thinking was a core feature of schizophrenia, but this feature is no longer included in the diagnosis of schizophrenia. Yet, a growing number of rare gene variants with relatively large effects on the risk for schizophrenia also contribute to the risk for ASD. Overlap of schizophrenia and ASD symptoms might be an illustration of pleiotropy. Similarly, recent large-scale GWAS studies report regions of the genome that appear to convey relatively modest increases in the risk for both schizophrenia and bipolar disorder. Thus, genetics provides some clues that may help to guide thinking about the features of the presentation of individual patients that conflict with the diagnostic categories of DSM-IV.

Why would someone add the antipsychotics lamotrigine and valproate to enhance clozapine treatment of schizophrenia?

Currently, there is a small evidence base to guide the pharmacotherapy of clozapine-resistant schizophrenia symptoms. The author was reminded of the evidence in animals and humans that lamotrigine, by itself or in combination with clozapine, attenuates some of the effects of psychotogenic NMDA receptor antagonists. Subsequently, lamotrigine was not found generally effective as an antipsychotic augmentation strategy unless patients had symptoms that had been highly resistant to treatments, including clozapine. Further, anticonvulsants like lamotrigine might protect against the seizure risk associated with pharmacotherapy with higher clozapine doses. Thus, there was a rationale for adding lamotrigine to clozapine in this case. However, the same could not be said for valproate. Psychiatrists discussed the pharmacoepidemiologic data describing common coprescription of valproate and antipsychotic medications for schizophrenia despite the lack of compelling efficacy data for valproate in this adjunctive role.

The invitation from the residents to interview a patient in front of them and to present a “biopsychosocial” formulation was embedded in an unspoken question. What do we mean by the “bio” component of the biopsychosocial formulation? Historically, the biologic domain was represented by general knowledge of psychopharmacology, medicine, and clinical neurology. Basic neuroscience was not included. In fact, historically, neuroscience education was not required in the psychiatry residency. To put this question another way, does it make a difference to a practicing psychiatrist to be a student of genetics and translational neuroscience? In the case that I discussed, the answer turned out to be absolutely! In fact, it seemed that the author’s background in neuroscience informed nearly every aspect of the interview, including the questions that the author asked, the hypotheses that he generated about the patient, and his thoughts about the treatment plan. While it was not yet clear to the author how his knowledge of the genetics of schizophrenia would guide the treatment of this patient, it was not hard to see how keeping abreast of developments in neuroscience might someday make this possible. There is also a growing appreciation that neurobiologic data, along with other types of data, will inform the development of more systematized personalized medicine approaches in psychiatry to improve the matching of individual patients to specific treatments.

REVIEW OF THE NEUROSCIENCE CHAPTERS

To facilitate this type of interplay between neuroscience and clinical psychiatry, the authors structured this section of the *Textbook* to cover the waterfront of progress in neuroscience that was most relevant to psychiatry. Experts in the field were asked to provide chapters that would be rigorous but accessible to clinicians. The section builds from elemental questions, that is, genetics, the properties of neurons, the basics of chemical neurotransmission, cortical networks, and network functions, and the pathophysiology of psychiatric disorders. The topics covered are broad, but some topics were not included. Similarly, the reviews provide depth, but hopefully not at the expense of accessibility. The Neuroscience Section of the Textbook is organized into five thematic subsections that are described briefly below.

The first subsection begins with three chapters that describe the organization and development of the brain. The first chapter (Chapter 1.2) in this subsection introduces the reader to core themes related to the overall structural organization of the brain, while the second and third chapters (Chapters 1.3 and 1.4) introduce the reader to molecular and cellular mechanisms governing the development of what is arguably one of the most complex structures in the universe.

The next seven chapters provide an overview of the basic signaling mechanisms that underlie communication and plasticity within the brain. Chapter 1.5 introduces readers to core cellular and synaptic mechanisms underlying neural communication. The next five chapters review the principal classes of chemical neurotransmission in the brain, including amino acid neurotransmitters (Chapter 1.5), monoamine neurotransmitters (Chapter 1.7), neuropeptides (Chapter 1.6), neurotrophic factors (Chapter 1.7), and novel neurotransmitters (Chapter 1.8). Intraneuronal signaling, the mechanisms through which the effects of transmitters and drugs are transduced chemically within neurons, are reviewed in Chapters 1.9 and 1.10.

Homeostatic and appetitive mechanisms are covered in the five chapters in the third subsection. These chapters cover topics that are implicated in nearly every psychiatric condition, such as psychoneuroendocrinology (Chapter 1.12), immune-brain interactions (Chapter 1.13), chronobiology (Chapter 1.14), sleep and insomnia (Chapter 1.15), pain (Chapter 1.16), and appetite (Chapter 1.17).

The fourth subsection covers a wide range of topics that are sometimes referred to as “the ’omics.” Chapter 1.18 introduces the reader to the relationship between genetics and brain biology, that is, the genome (the sources of genetic variation), the transcriptome (the levels of gene expression), and the proteome (the levels of various proteins). The following chapters in this subsection address particular aspects of psychiatric risk and resilience, including genomics (Chapter 1.19) and epigenomics (Chapter 1.20). Epigenomics is a topic of growing interest with regard to psychiatry biomarker development as epigenomic modifications provide a path for various environmental exposures (stress, abused substances, rewarding stimuli) to produce lasting changes in gene expression without altering the DNA sequence. Chapter 1.21 reviews the challenging area of pharmacogenomics, that is, the effort to find genomic predictors for the safety and effectiveness of psychiatric treatments. Lastly, for this subsection, animal models for psychiatric disorders are reviewed in chapter 1.22. Animal models, also known as animal research platforms, provide unique opportunities to test causal hypotheses related to brain structure, chemistry, and function. As a result, they are a critical component of the translational neuroscience mission in psychiatry.

The last 10 chapters in Section 1 cover special topics in systems, cognitive, and behavioral neuroscience. Chapter 1.23 covers basic issues in systems neuroscience. Chapter 1.24 addresses the important area of computational modeling, and it provides a foundation for the emerging focus on computation in psychiatry, that is, “computational psychiatry.” Chapters 1.25 to 1.28 present the key techniques for studying human cortical circuit function and chemistry, including nuclear magnetic resonance (sMRI, fMRI, dwMRI/DTI, MRS), cortical electrophysiology (EEG, MEG), and radiotracer imaging (PET, SPECT). Neuroscience is also informing our understanding of our sense of self (Chapter 1.29). The last two chapters in this section address learning theory and its application to the study of addiction (Chapters 1.30 and 1.31).

CONCLUSIONS

The information conveyed in the following 30 chapters provides evidence that psychiatric neuroscience is progressing at a dizzying pace. Nonetheless, in the final analysis, psychiatrists must be humble with regard to our current understanding of the biology and treatment of psychiatric disorders. The brain is unimaginably complicated. To quote chapter 1.3, it has “approximately 86 billion neurons connected by hundreds of thousands of kilometers of myelinated axons and several hundred trillion to well over a quadrillion synapses, along with a roughly equal number of glial cells.” Fundamental aspects of the pathophysiology of neuropsychiatric disorders remain beyond our grasp. Although some parts of the psychiatric community are put off by the growing complexity of neuroscience, others are attracted to the increasing explanatory power of this work. Although cautious in applying neuroscience advances to patients, psychiatry probably overestimates our knowledge and underestimates the challenges faced in trying to understand the etiology, pathophysiology, treatment, and prevention of neuropsychiatric disorders. Nonetheless, psychiatrists have a clinical, scientific, and humanistic imperative to continue to advance our understanding of the neurobiology of psychiatric disorders and to see that this knowledge improves the clinical practice of psychiatry.

Before closing, the author wishes to acknowledge the outstanding quality of the chapters in this Section as well as the wonderful collaboration with Mr. Sean Hanrahan and Dr. Robert Boland in shaping this Section. The author had the pleasure of reading these chapters several times and, with each repetition, learned more. He thanks the authors of these chapters for helping to inform you about some of the most exciting advances in all of science and for making this process so enjoyable.

FURTHER READINGS

Anand A, Charney DS, Oren DA, et al. Attenuation of the neuropsychiatric effects of ketamine with lamotrigine: support for hyperglutamatergic effects of N-methyl-D-aspartate receptor antagonists. *Arch Gen Psychiatry*. 2000;57:270–276.

Association AP. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Press; 2013.

Brody SA, Geyer MA, Large CH. Lamotrigine prevents ketamine but not amphetamine-induced deficits in prepulse inhibition in mice. *Psychopharmacology (Berl)*. 2003;169:240–246.

Chen H, Deshpande AD, Jiang R, Martin BC. An epidemiological investigation of off-label anticonvulsant drug use in the Georgia Medicaid population. *Pharmacoepidemiol Drug Saf*. 2005;14:629–638.

Duman RS, Aghajanian GK, Sanacora G, Krystal JH. Synaptic plasticity and depression: new insights from stress and rapid-acting antidepressants. *Nat Med*. 2016;22:238–249.

Farber NB, Jiang XP, Heinkel C, Nemmers B. Antiepileptic drugs and agents that inhibit voltage-gated sodium channels prevent NMDA antagonist neurotoxicity. *Mol Psychiatry*. 2002;7:726–733.

Goff DC, Keefe R, Citrome L, et al. Lamotrigine as add-on therapy in schizophrenia: results of 2 placebo-controlled trials. *J Clin Psychopharmacol*. 2007;27:582–589.

Gulsuner S, Walsh T, Watts AC, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell*. 2013;154:518–529.

Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010;167:748–751.

Krystal JH, State MW. Psychiatric disorders: diagnosis to therapy. *Cell*. 2014;157:201–214.

Mendell JR, Al-Zaidy S, Shell R, et al. Single-dose gene-replacement therapy for spinal muscular atrophy. *N Engl J Med*. 2017;377(18):1713–1722.

Moskowitz A, Heim G. Eugen Bleuler's Dementia praecox or the group of schizophrenias (1911): a centenary appreciation and reconsideration. *Schizophr Bull*. 2011;37:471–479.

Ripke S, Sanders AR, Kendler KS, et al. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011;43:969–976.

Schwarz C, Volz A, Li C, Leucht S. Valproate for schizophrenia. *Cochrane Database Syst Rev*. 2008;(3):CD004028.

State MW, Levitt P. The conundrums of understanding genetic risks for autism spectrum disorders. *Nat Neurosci*. 2011;14(12):1499–1506.

Tiihonen J, Wahlbeck K, Kiviniemi V. The efficacy of lamotrigine in clozapine-resistant schizophrenia: a systematic review and meta-analysis. *Schizophr Res*. 2009;109:10–14.

Treadway MT, Pizzagalli DA. Imaging the pathophysiology of major depressive disorder—from localist models to circuit-based analysis. *Biol Mood Anxiety Disord*. 2014;4:5.

Tye KM, Mirzabekov JJ, Warden MR, et al. Dopamine neurons modulate neural encoding and expression of depression-related behaviour. *Nature*. 2013;493:537–541.

Wang XJ, Krystal JH. Computational psychiatry. *Neuron*. 2014;84:638–654.

Williams HJ, Zamzow CR, Robertson H, Dursun SM. Effects of clozapine plus lamotrigine on phencyclidine-induced hyperactivity. *Prog Neuropsychopharmacol Biol Psychiatry*. 2006;30(2):239–243.

Willsey AJ, Sanders SJ, Li M, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. 155:997–1007.

1.2 Functional Neuroanatomy

DARLENE S. MELCHITZKY, M.S., AND DAVID A. LEWIS, M.D.

The broad range of affective, cognitive, and behavioral characteristics of humans arise as a consequence of specific patterns of activation in networks of neurons that are distributed across the central nervous system (CNS). These patterns of activation are mediated by the connections among specific brain structures. Consequently, understanding the neurobiologic bases for the disturbances in affective, cognitive, and behavioral processes present in psychiatric disorders requires an appreciation of the major principles governing the functional organization of these structures and their connections in the human brain. This chapter reviews some of these anatomic principles and illustrates them in the functional circuitry of neural systems that are of particular relevance to psychiatric disorders.

PRINCIPLES OF BRAIN ORGANIZATION

Cells

The human brain contains approximately 10^{11} nerve cells or *neurons*. In general, neurons are composed of four morphologically identified regions (Fig. 1.2–1): (1) the cell body, or *soma*, which contains the nucleus and can be considered the metabolic center of the neuron; (2) the *dendrites*, processes that arise from the cell body, branch extensively, and serve as the major recipient zones of input from other neurons; (3) the *axon*, a single process that arises from a specialized portion of the cell body (the *axon hillock*) and conveys information to other neurons; and (4) the *axon terminals*, fine branches near the end of the axon that form contacts (*synapses*) generally with the dendrites or the cell bodies of other neurons, release neurotransmitters, and provide a mechanism for interneuronal communication.

Most neurons in the human brain are considered to be multipolar in that they give rise to a single axon and several dendritic processes. Although there are numerous classification schemes for neurons in different brain regions, almost all neurons can be considered to be either projection or local circuit neurons. *Projection neurons* have long axons and convey information from the periphery to the brain (sensory neurons), from one brain region to another, or from the brain to effector organs (motor neurons). In contrast, *local circuit neurons* or *interneurons* have short axons and process information within distinct regions of the brain.

Neurons can also be classified according to the neurotransmitters they contain; for example, the neurons of the substantia nigra are often classified by their dopamine content. However, some neurons may contain more than one substance used for neurotransmission. Identification of neurons by their neurotransmitter content in anatomic studies provides a means for correlating the structure of a neuron with certain aspects of its function. However, neurotransmitters have defined effects on the activity of neurons, whereas complex brain functions, such as those disturbed in psychiatric disorders, are mediated by the coordinated activity of ensembles of neurons. Thus, the effects of neurotransmitters (or of pharmacologic agents that mimic or antagonize the action of neurotransmitters) on behavioral, emotional, or cognitive states must be viewed within the context of the neural circuits that they influence.

In addition to neurons, the brain contains several types of *glial* cells (Table 1.2–1). *Astrocytes* are the most numerous class of glial cells and are classified as either protoplasmic or fibrous. Protoplasmic astrocytes are large cells located exclusively in the gray matter with many fine and elaborate processes. In contrast, fibrous astrocytes are smaller with less complex processes and reside exclusively in the white matter. Protoplasmic astrocytes have been demonstrated to serve a number of functions, including participation in the formation of the blood–brain barrier, removal of glutamate and γ-aminobutyric acid (GABA) from the synaptic cleft, and buffering of the extracellular potassium (K^+) concentration.

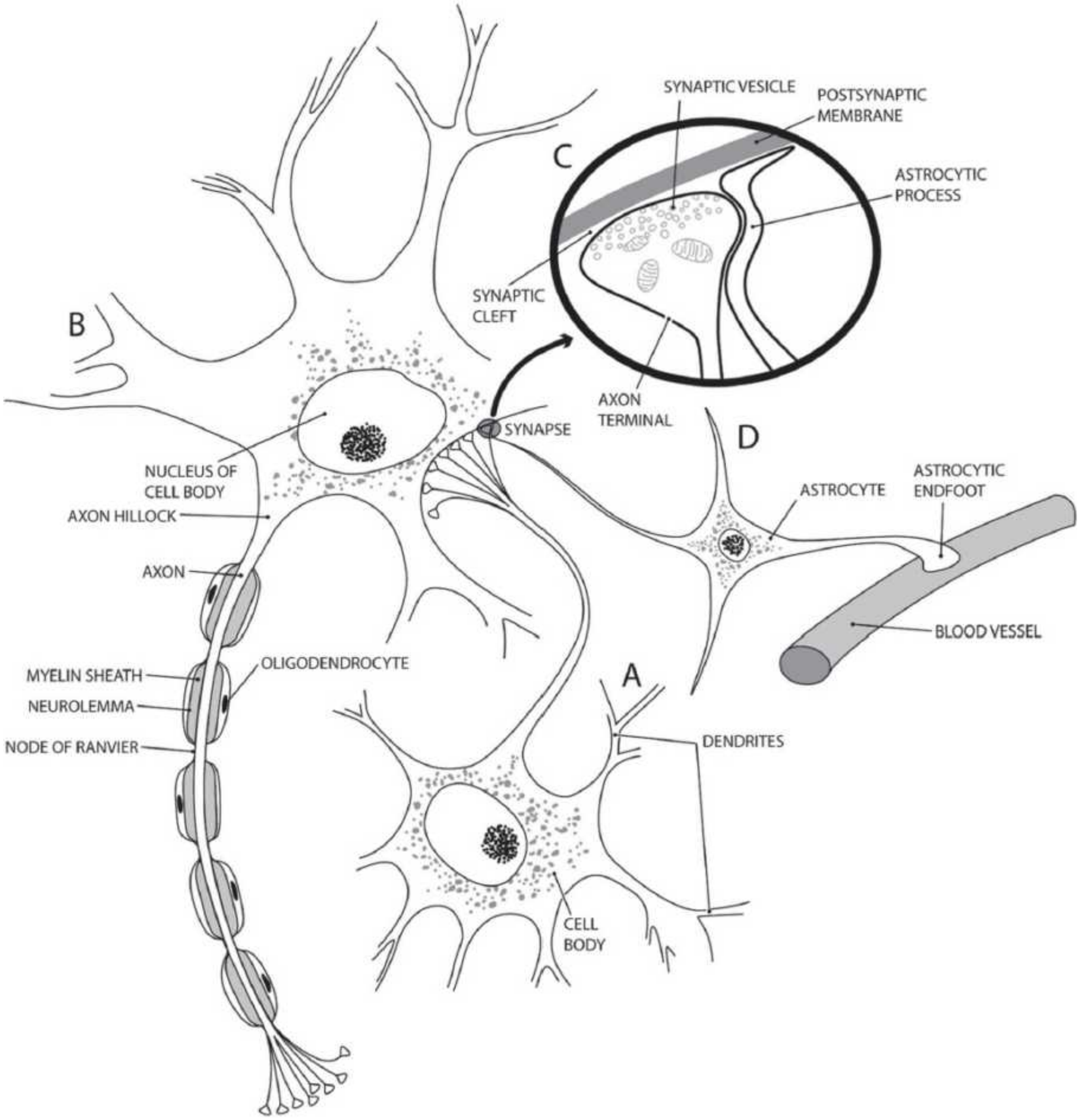
SYNAPTIC VESICLE
POSTSYNAPTIC MEMBRANE
C
ASTROCYTIC PROCESS
SYNAPTIC CLEFT
B
AXON TERMINAL
SYNAPSE
D
NUCLEUS OF CELL BODY
ASTROCYTE
ASTROCYTIC ENDFOOT
AXON HILLOCK
AXON
BLOOD VESSEL
OLIGODENDROCYTE
MYELIN SHEATH
NEUROLEMMA
A
NODE OF RANVIER
DENDRITES
CELL BODY

FIGURE 1.2–1. Drawing of the major features of a typical neuron, oligodendrocyte, and astrocyte. **A, B:** The main features of neurons include a cell body, dendrites, a principal axon, and axon terminals. The principal axon is myelinated by neighboring oligodendrocytes. **C:** Axon terminals branch off axons and form synapses with postsynaptic neurons via the release of neurotransmitters stored in synaptic vesicles. **D:** Fine astrocytic processes surround synapses and regulate neurotransmission, while astrocytic endfeet contact blood vessels and regulate blood flow. (Adapted with permission from Gilman S, Winans-Newman S. *Manter and Gatz's Essentials of Clinical Neuroanatomy and Neurophysiology*. 10th ed. FA Davis Co; 2003:2.)

The close contact of their processes with neurons and blood vessels suggests that protoplasmic astrocytes may help support the energy requirements of neurons. For example, distinctive processes of protoplasmic astrocytes called endfeet line the basal lamina of blood vessels. These endfeet appear to extract glucose from the blood and convert it to lactate which is then released for neurons to uptake and use as an energy substrate. The endfeet may also play a role in the regulation of cerebral blood flow leading protoplasmic astrocytes to be referred to as a neurovascular bridge. In addition to these regulatory functions, protoplasmic astrocytes are involved in synaptic neurotransmission in two distinct ways. First, perisynaptic astrocytic processes express a variety of neurotransmitter receptors, and these receptors are stimulated by the release of neurotransmitters from presynaptic axon terminals. The activated glial cell then releases gliotransmitters that can stimulate the postsynaptic neuron. Thus, the perisynaptic astrocyte is an active partner in synaptic transmission, creating what has been termed the tripartite synapse. This feature allows astrocytes to regulate both glutamatergic/excitatory as well as GABAergic/inhibitory neurotransmission. Second, neurotransmission is regulated by the structural network of astrocytes, which has been shown to comprise nonoverlapping domains of individual astrocytes. In the hippocampus and the cerebral cortex, individual astrocytes possess their own domain, with only their most distal processes interdigitating with processes of neighboring astrocytes. This pattern of organization creates distinct astrocytic domains, in which individual astrocytes modulate the activity of neurons and synapses.

Table 1.2–1. Glial Cells

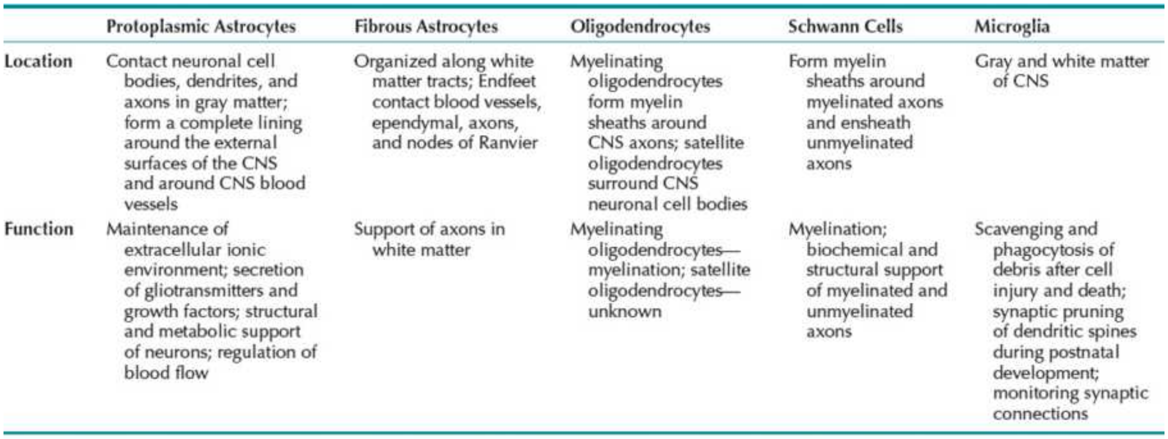

	Protoplasmic Astrocytes	Fibrous Astrocytes	Oligodendrocytes	Schwann Cells	Microglia
Location	Contact neuronal cell bodies, dendrites, and axons in gray matter; form a complete lining around the external surfaces of the CNS and around CNS blood vessels	Organized along white matter tracts; Endfeet contact blood vessels, ependymal, axons, and nodes of Ranvier	Myelinating oligodendrocytes form myelin sheaths around CNS axons; satellite oligodendrocytes surround CNS neuronal cell bodies	Form myelin sheaths around myelinated axons and ensheath unmyelinated axons	Gray and white matter of CNS
Function	Maintenance of extracellular ionic environment; secretion of gliotransmitters and growth factors; structural and metabolic support of neurons; regulation of blood flow	Support of axons in white matter	Myelinating oligodendrocytes—myelination; satellite oligodendrocytes—unknown	Myelination; biochemical and structural support of myelinated and unmyelinated axons	Scavenging and phagocytosis of debris after cell injury and death; synaptic pruning of dendritic spines during postnatal development; monitoring synaptic connections

Modified from Mihailoff GA, Haines DE. The cell biology of neurons and glia. In: Haines DE, Mihailoff GA. *Fundamental Neuroscience for Basic and Clinical Applications.* 5th ed. Elsevier; 2018:25.

The cell bodies of fibrous astrocytes are evenly distributed within the white matter, and unlike protoplasmic astrocytes, their processes overlap. Some processes form endfeet that contact blood vessels, ependymal cells, axons, and nodes of Ranvier. Fibrous astrocytes appear to play the traditional supportive role of neuroglia.

Oligodendrocytes and *Schwann cells*, found in the CNS and peripheral nervous system, respectively, are small cells that wrap their membranous processes around axons in a tight spiral. The resulting *myelin sheath* facilitates the conduction of action potentials along the axon. *Satellite oligodendrocytes* are found preferentially in gray matter, and their perineuronal location suggests that they may be involved in the regulation of the environment surrounding neurons.

The third class of glial cells, the *microglia*, is derived from macrophages and historically was thought to function solely as scavengers, eliminating the debris resulting from neuronal death and injury. However, studies over the last decade have demonstrated that microglia also regulate neurogenesis and synaptic function. For example, microglia are involved in the synaptic pruning of dendritic spines during postnatal development. In addition, microglia may have the ability to regulate the function of neuronal networks by constantly monitoring the status of synaptic connections and integrating new information when necessary.

Alterations in glial cells may contribute to the pathophysiology of psychiatric disorders. For example, stereologic studies of postmortem human brain tissue have revealed altered glial cell number in the dorsal prefrontal cortex of individuals with schizophrenia and in the subgenual medial prefrontal cortex in depressed individuals. Moreover, the decrease in cerebral blood flow in the prefrontal cortex of individuals with depression may be linked to a reduction in the amount of blood vessel coverage by astrocytic endfeet. Sleep disturbances can be comorbid with many psychiatric disorders, and thus clearance of neuronal and glial waste products through the glymphatic system during normal sleep may be impaired in these disorders. Although some putative susceptibility genes for schizophrenia are selectively expressed in glial cells, it is still unclear whether the alterations in glial cell number reflect the disease process or are a consequence of treating the disease.

Architecture

Neurons and their processes form groupings in many different ways, and these patterns of organization, or architecture, can be evaluated by several approaches. The pattern of distribution of nerve cell bodies, called *cytoarchitecture*, is revealed by aniline dyes called Nissl stains that stain ribonucleotides in the nuclei and the cytoplasm of neuronal cell bodies. The Nissl stains show the relative size and packing density of the neurons and, consequently, reveal the organization of the neurons into the different layers of the cerebral cortex. In certain pathologic states, such as Alzheimer disease (referred to as Major Neurocognitive Disorder due to Alzheimer type in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* [DSM-5]), neuronal degeneration and loss result in striking changes in the cytoarchitecture of some brain regions (Fig. 1.2–2).

Other types of histologic techniques, such as silver stains, selectively label the myelin coating of axons and consequently reveal the *myeloarchitecture* of the brain. For example, certain regions of the cerebral cortex—such as area MT, a portion of the temporal cortex involved in processing visual information—can be identified by a characteristic pattern of heavy myelination in the deep cortical layers. The progression of myelination is highly region specific, may not be complete for years after birth, and may be a useful anatomic indicator of the functional maturation of brain regions.

Immunohistochemical and other related techniques—which identify the location of neurotransmitters, their synthetic enzymes, or other molecules within neurons—can be used to determine the *chemoarchitecture* of the brain (Fig. 1.2–3B). In some cases, these techniques reveal striking regional differences in the chemoarchitecture of the brain that are difficult to detect in cytoarchitecture.

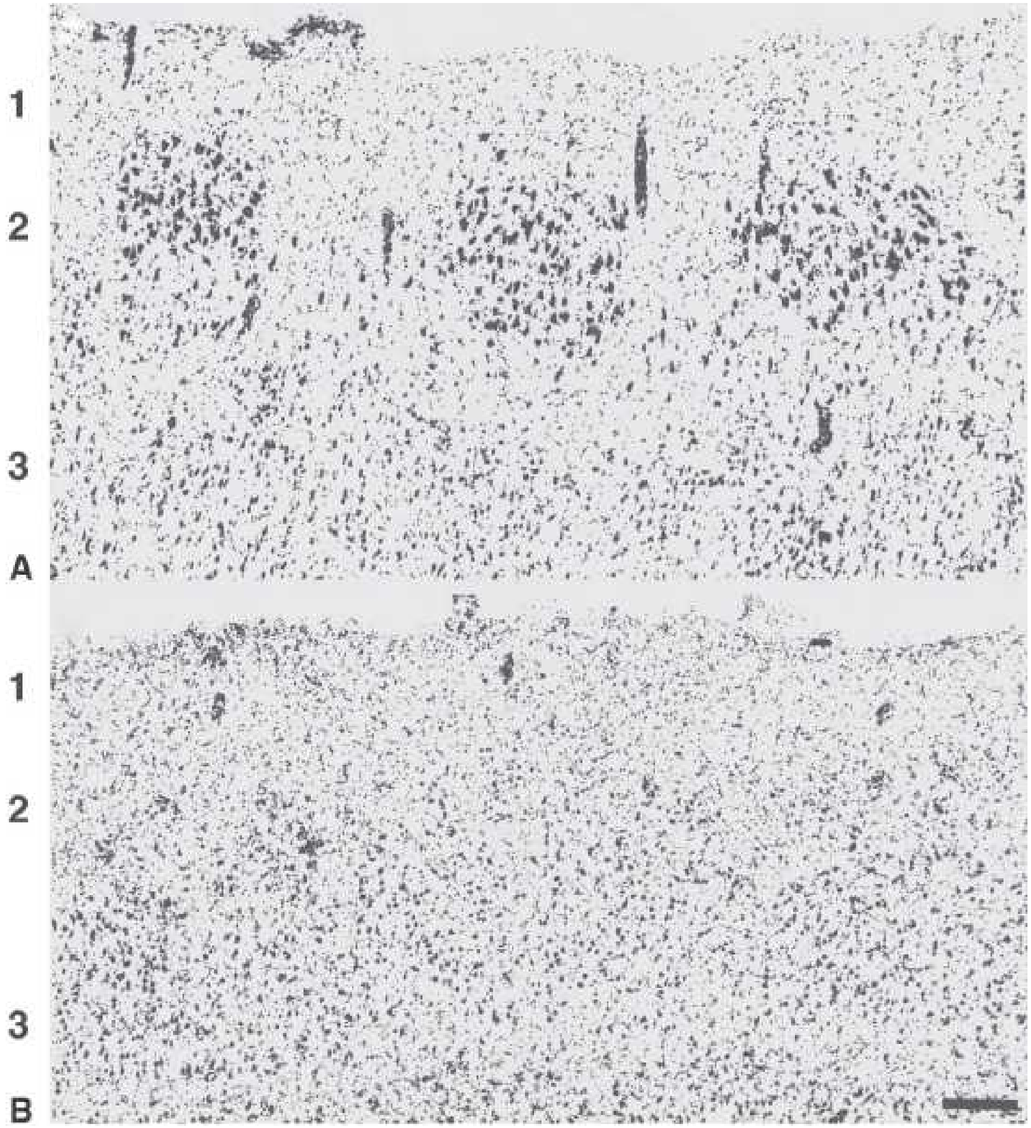

FIGURE 1.2–2. Nissl-stained sections of the superficial layers of the intermediate region of human entorhinal cortex. **A:** In the control brain, layer 2 contains clusters or islands of large, intensely stained neurons. **B:** In Alzheimer disease, these layer 2 neurons are particularly vulnerable to degeneration, and their loss produces a marked change in the cytoarchitecture of the region. Arabic numbers indicate the location of the cortical layers. Calibration bar (200 μm) applies to A and B.

Connections

Every function of the human brain is a consequence of the activity of specific neural circuits. The circuits form as a result of several developmental processes. First, each neuron extends an axon, either after it has migrated to its final location or, in some cases, before. The growth of an axon along distinct pathways is guided by molecular cues from its environment and eventually leads to the formation of synapses with specific target neurons. Although the projection of axons is quite precise, some axons initially produce an excessive number of axon branches, or *collaterals*, and contact a broader set of targets than are present in the adult brain. During later adolescence, the connections of particular neurons are focused by the pruning or elimination of axonal projections to inappropriate targets. The developmental timing of synaptic and axonal elimination seems to be highly specific across regions of the brain.

Within the adult brain, the connections among neurons or neural circuits follow several important principles of organization. First, many connections between brain regions are *reciprocal*. That is, each region tends to receive input from the regions to which it sends axonal projections. In some cases, the axons arising from one region may directly innervate the reciprocating projection neurons in another region; in other cases, local circuit interneurons are interposed between the incoming axons and the projection neurons that furnish the reciprocal connections. For some projections, the reciprocating connection is indirect, passing through one or more additional brain regions and synapses before innervating the initial brain region. In addition, connections within brain regions also display reciprocity. For example, in monkey prefrontal cortex, tract-tracing studies have shown that the axons and cell bodies of pyramidal neurons in layers 2 and 3 are arranged in a series of discrete stripes (Fig. 1.2–4). Reciprocity in this system is represented by the coregistration of anterogradely labeled axons and retrogradely labeled neurons within individual intrinsic and associational stripes. In addition, anterogradely labeled axon terminals form asymmetric synapses onto retrogradely labeled dendritic spines within individual stripes, providing further evidence of reciprocity in these connections.

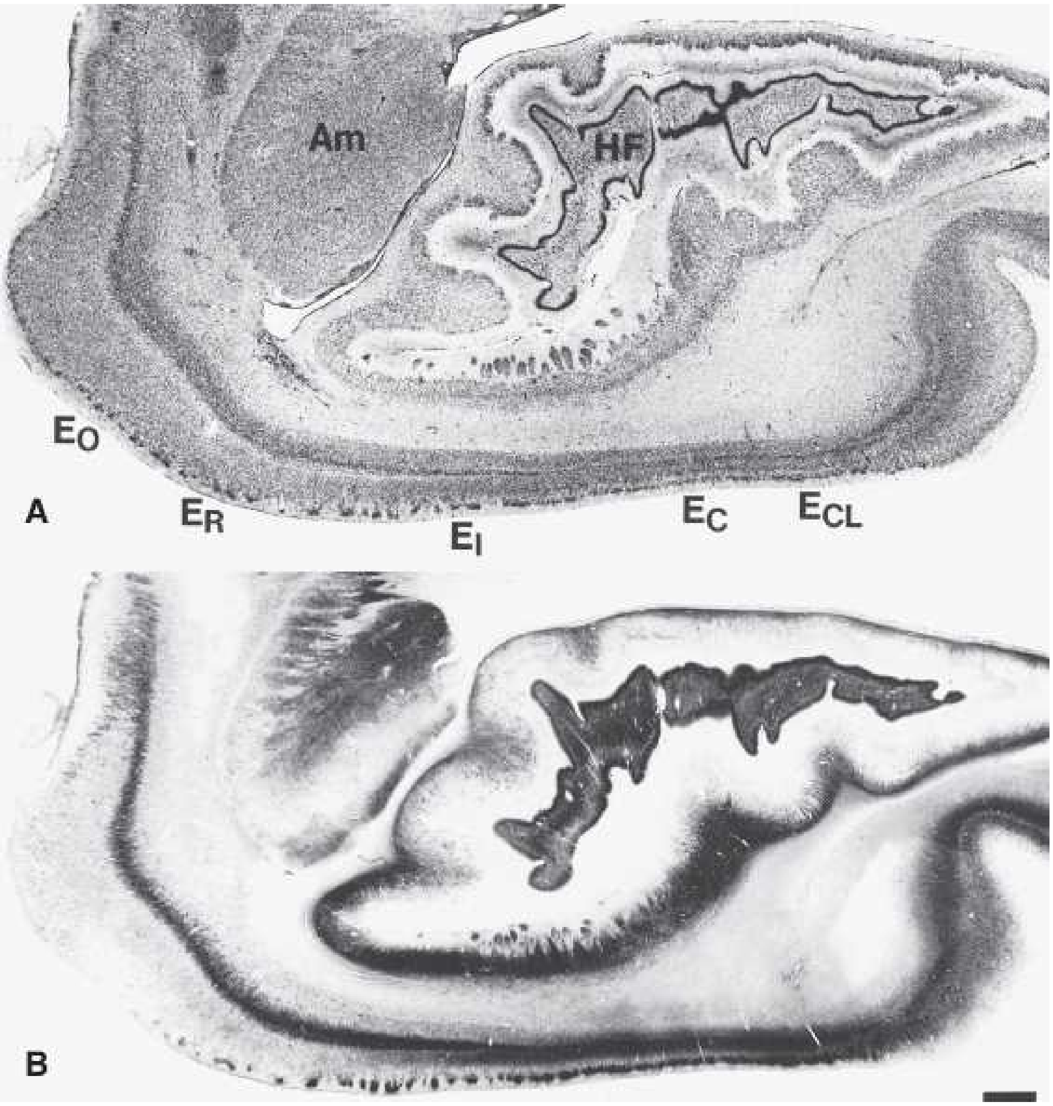

FIGURE 1.2–3. Adjacent sagittal sections through the medial temporal lobe of the human brain labeled to reveal the cytoarchitecture (**A**: Nissl stain) and chemoarchitecture (**B**: nonphosphorylated neurofilament protein immunoreactivity) of the entorhinal cortex. *Letters* indicate some of its subdivisions. Am, amygdala; HF, hippocampal formation. Calibration bar (2 mm) applies to both panels. (Reprinted with permission from Beall MJ, Lewis DA. Heterogeneity of layer II neurons in human entorhinal cortex. *J Comp Neurol*. 1992;321(2):241–266.)

Second, many neuronal connections are either divergent or convergent in nature. A *divergent* system involves the conduct of information from one neuron or a discrete group of neurons to a much larger number of neurons that may be located in diverse portions of the brain. The *locus coeruleus*, a small group of norepinephrine-containing neurons in the brainstem that sends axonal projections to the entire cerebral cortex and other brain regions, is an example of a highly divergent system. In contrast, the output of multiple brain regions may be directed toward a single area, forming a *convergent* system. Projection from multiple association areas of the cerebral cortex to the entorhinal region of the medial temporal lobe is an example of a convergent system. Connections within brain regions also display divergence and convergence (see Fig. 1.2–4). For example, in monkey prefrontal cortex, pyramidal neurons within an individual stripe have axons that project to several other stripes (divergence), and individual stripes receive input from more than one stripe (convergence). The divergence in this system may provide an anatomic substrate that would allow a spatially restricted input to recruit a group of neurons whose coordinated activation is necessary to generate a particular response. Convergence in this system could allow information from different modalities present in the array of stripes to be relayed to a single location, facilitating the integration of their information content.

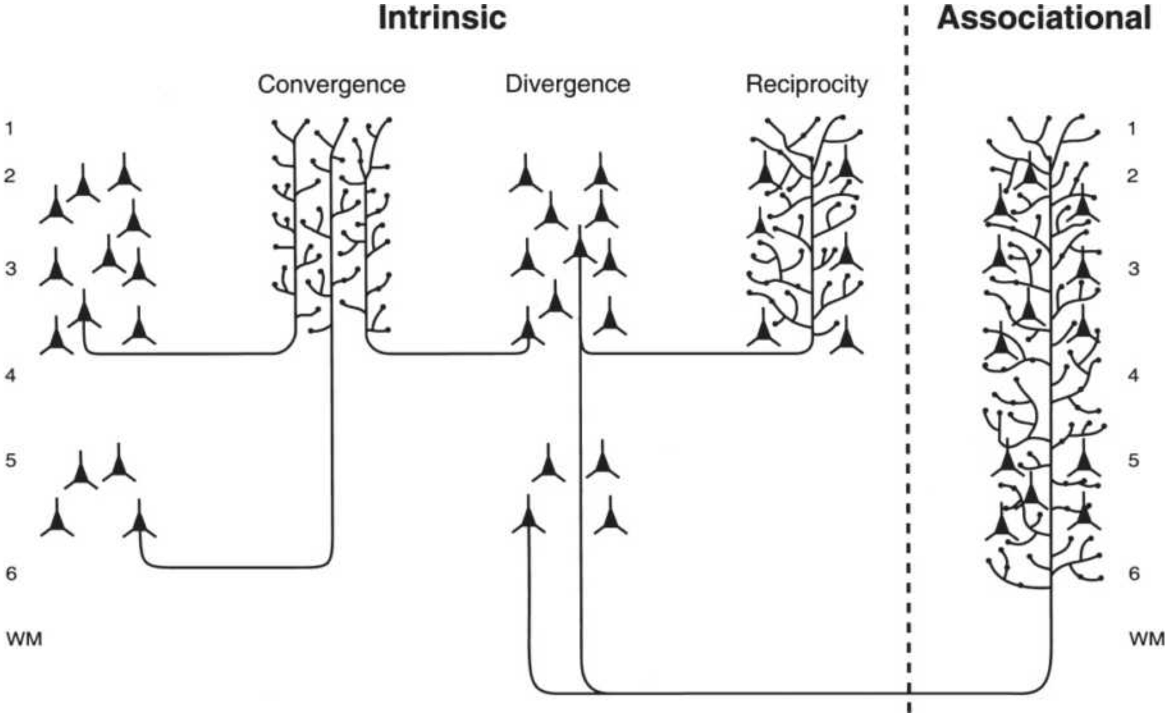

FIGURE 1.2–4. Schematic drawing comparing the spatial organization of intrinsic and associational connections and the reciprocity, convergence, and divergence in these connections in monkey prefrontal cortex. (Adapted with permission from Pucak ML, Levitt JB, Lund JS, Lewis DA. Patterns of intrinsic and associational circuitry in monkey prefrontal cortex. *J Comp Neurol*. 1996;376(4):614–630. Copyright © 1996 Wiley-Liss, Inc.)

Third, the connections among regions may be organized in a hierarchical or parallel fashion, or both. Visual input is conveyed in a serial or hierarchical fashion through several populations of neurons in the retina to the lateral geniculate nucleus, to the primary visual cortex, and then, progressively, to the multiple visual association areas of the cerebral cortex. Within the hierarchical scheme, different types of visual information (e.g., motion and form) may be processed in a parallel fashion through different portions of the visual system.

Finally, regions of the brain are specialized for different functions. Lesions of the left inferior frontal gyrus (Broca area) (Fig. 1.2–5) produce a characteristic impairment in speech production. Speech is a complex faculty. However, that depends not only on the integrity of Broca area but also on the distributed processing of information across numerous brain regions through divergent and convergent, serial, and parallel interconnections. Thus, the role of any particular brain region or group of neurons in the production of specific behaviors or in the pathophysiology of a given neuropsychiatric disorder cannot be viewed in isolation but must be considered within the context of the neural circuits connecting the neurons with other brain regions.

Distinctiveness of the Human Brain

Compared with the brains of other primate species, the human brain is substantially greater in size, with certain areas expanded disproportionately. The prefrontal cortex has been estimated to occupy only 3.5 percent of the total cortical volume in cats and 11.5 percent in monkeys but close to 30 percent of the much larger cortical volume of the human brain. Conversely, the relative representation of other regions is decreased in the human brain; for example, the primary visual cortex accounts for only 1.5 percent of the total area of the cerebral cortex in humans, but in monkeys, a much greater proportion (17 percent) of the cerebral cortex is devoted to this region. Thus, the distinctiveness of the human brain is attributable to its size and to the differential expansion of certain regions, particularly the areas of the cerebral cortex devoted to higher cognitive functions.

In addition, the expansion and differentiation of the human brain are associated with substantial differences in the organization of certain elements of neural circuitry. For example, compared with rodents, the dopaminergic innervation of the human cerebral cortex is much more widespread and regionally specific. The primary motor cortex and certain posterior parietal regions receive a dense dopamine innervation in monkeys and humans, but these areas receive little dopamine input in rats. The human brain is also distinctive at the cellular level, with greater relative percentages of certain neuronal and glial cells in the cerebral cortex. For example, the relative proportion of GABAergic interneurons is ∼15 percent in rodent cerebral cortex, but in humans, 20 to 25 percent of all neurons in most cortical areas are GABAergic. In addition, molecular characterization has revealed subtypes of neurons in the human cerebral cortex that are not found in rodent cerebral cortex. Intralaminar and polarized astrocytes are found exclusively in humans, and human protoplasmic astrocytes are larger and have more processes than their rodent counterparts. Furthermore, the astrocyte to neuron ratio is fivefold greater in humans compared to rodents, leading to greater astrocytic coverage of neurons and synapses and thus possibly enhanced regulation of neurotransmission in human cerebral cortex. These findings of a greater contribution of astrocytes in human brain function suggest that astrocytes are critical for higher brain functions and may be vulnerable in human psychiatric disorders. Thus, the human brain exhibits differences in gross anatomy, cellular constituents, and molecular composition, features that distinguish it from phylogenetically lower species.

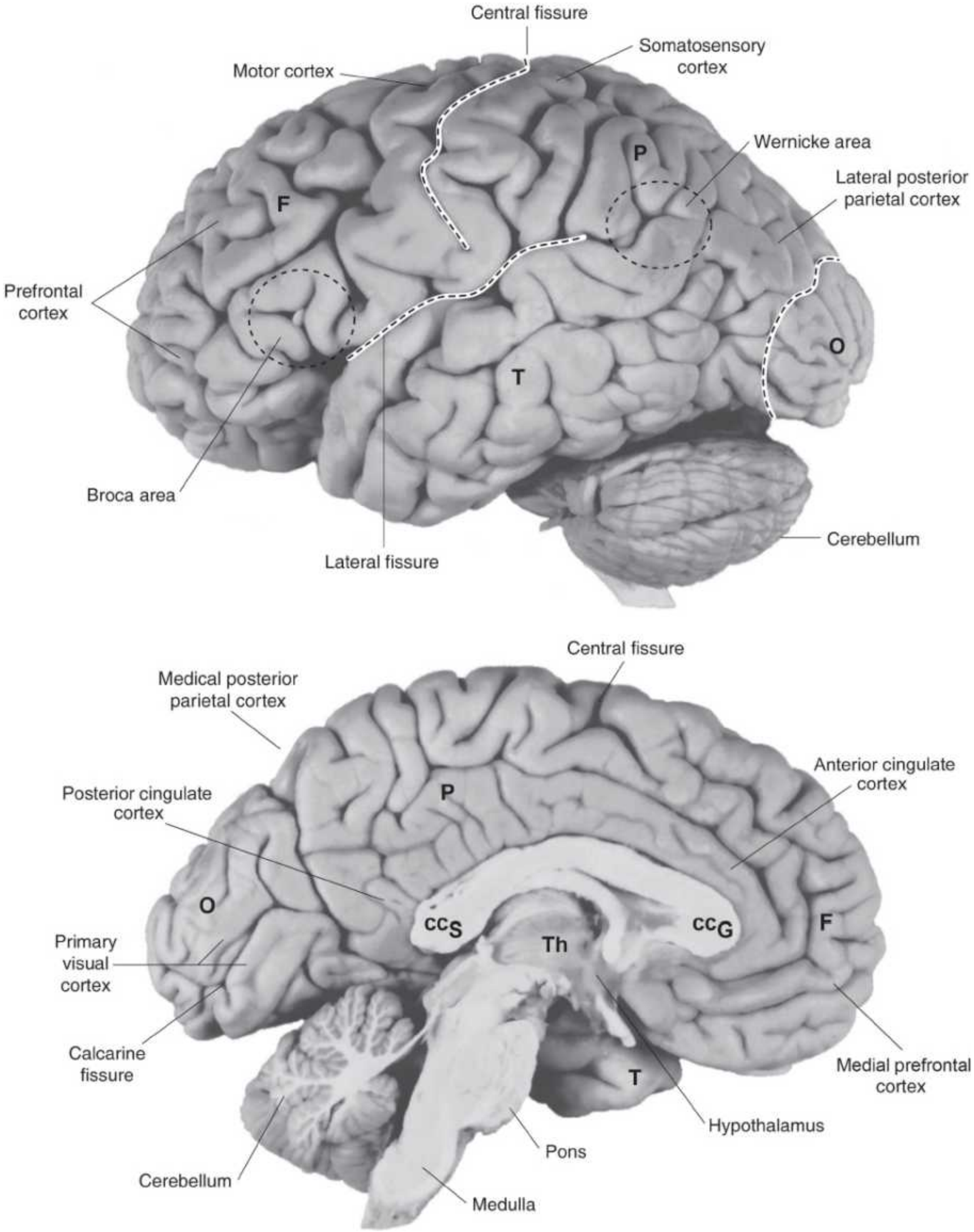
Central fissure
Motor cortex
Somatosensory cortex
P
Wernicke area
Lateral posterior parietal cortex
F
Prefrontal cortex
O
T
Broca area
Cerebellum
Lateral fissure
Central fissure
Medical posterior parietal cortex
Anterior cingulate cortex
Posterior cingulate cortex
P
O
cc$_S$
cc$_G$
F
Th
Primary visual cortex
Calcarine fissure
T
Medial prefrontal cortex
Hypothalamus
Pons
Cerebellum
Medulla

FIGURE 1.2–5. Photographs of the lateral (**top**) and medial (**bottom**) aspects of the left hemisphere of a human brain indicating the location of major surface landmarks. F, frontal lobe; O, occipital lobe; P, parietal lobe; T, temporal lobe; Th, thalamus; cc$_G$, genu of the corpus callosum; cc$_S$, splenium of the corpus callosum.

These types of species differences indicate that there are limits to the accuracy of generalizations made concerning human brain function when using studies in rodents or even nonhuman primates as the basis for the inference. Direct investigation of the organization of the human brain, however, is obviously limited and complicated by numerous factors.

As indicated earlier, the expansion of the human brain is associated with the appearance of additional regions of the cerebral cortex. For example, the entorhinal cortex of the medial temporal lobe in humans is sometimes considered to be a single cortical region, but the cytoarchitecture and chemoarchitecture of this cortex differ substantially along its rostral–caudal extent (see Fig. 1.2–3). It is tempting to identify these regions by their location relative to other structures, but sufficient interindividual variability exists in the human brain to make such a topologic definition unreliable. In the case of the entorhinal cortex, the location of its different subdivisions relative to adjacent structures, such as the amygdala and the hippocampus, varies across human brains. Therefore, in all studies, particularly studies using the human brain, areas of interest must be defined in a manner (e.g., using cyto-, chemo-, or myeloarchitectural features) that allows investigators to accurately identify the same region in all cases.

An additional limitation to the study of the human brain concerns the changes in morphology and biochemistry that can occur during the interval between the time of death and the freezing or fixation of brain specimens. In addition to the influence of the known postmortem interval, such changes may begin to occur during the agonal state preceding death. When comparing aspects of the organization of the human brain with that of other species, the researcher must try to account for changes that may have occurred in the human brain as a result of postmortem delay or agonal state. In the study of disease states, appropriate controls must be used because differences in neurotransmitter content or other characteristics among cases could be a result of factors other than the disease state, such as methods of tissue preparation. Studies of the human brain in vivo—using such imaging techniques as positron emission tomography (PET), magnetic resonance imaging (MRI), magnetic resonance spectroscopy (MRS), and diffusion tensor imaging (DTI)—circumvent many of these problems but are limited by insufficient resolution for the study of many aspects of human brain organization.

STRUCTURAL COMPONENTS

Major Brain Structures

In the early stages of human brain development, three primary vesicles can be identified in the neural tube: the *prosencephalon*, the *mesencephalon,* and the *rhombencephalon* (Fig. 1.2–6). Subsequently, the prosencephalon divides to become the *telencephalon* and the *diencephalon*. The telencephalon gives rise to the cerebral cortex, the hippocampal formation, the amygdala, and some components of the basal ganglia. The diencephalon becomes the thalamus, the hypothalamus, and several other related structures. The mesencephalon gives rise to the midbrain structures of the adult brain. The rhombencephalon divides into the *metencephalon* and the *myelencephalon*. The metencephalon gives rise to the pons and the cerebellum; the medulla is the derivative of the myelencephalon.

The cerebral cortex of each hemisphere is divided into four major regions: the *frontal, parietal, temporal*, and *occipital* lobes (see Fig. 1.2–5). The frontal lobe is located anterior to the central sulcus and consists of the primary motor, premotor, and prefrontal regions (Fig. 1.2–7). The prefrontal cortex can be divided into dorsolateral and ventrolateral regions, with each of these regions having different functional properties. For example, the dorsolateral prefrontal cortex (DLPFC) seems to be more involved in the manipulation of data during working memory tasks than does the ventrolateral prefrontal cortex, which seems to be more involved with pure maintenance of information during working memory. The primary somatosensory cortex is located in the anterior parietal lobe; in addition, other cortical regions related to complex visual and somatosensory functions are located in the posterior parietal lobe. The superior portion of the temporal lobe contains the primary auditory cortex and other auditory regions; the inferior portion contains regions devoted to complex visual functions. In addition, some regions within the superior temporal sulcus receive a convergence of input from the visual, somatosensory, and auditory sensory areas. The occipital lobe consists of the primary visual cortex and other visual association areas. The brains of humans and other anthropoid primates also contain a major area called the insula which is a cortical region located within the depths of the lateral sulcus (Figs. 1.2–8 and 1.2–9), covered dorsally by the frontal and parietal opercula and ventrally by the temporal operculum. The human insula is gyrencephalic (convoluted), unlike the insula in nonhuman primates, which is lissencephalic (smooth). The human insula has extensive connections with other cortical areas, the amygdala, and the thalamus.

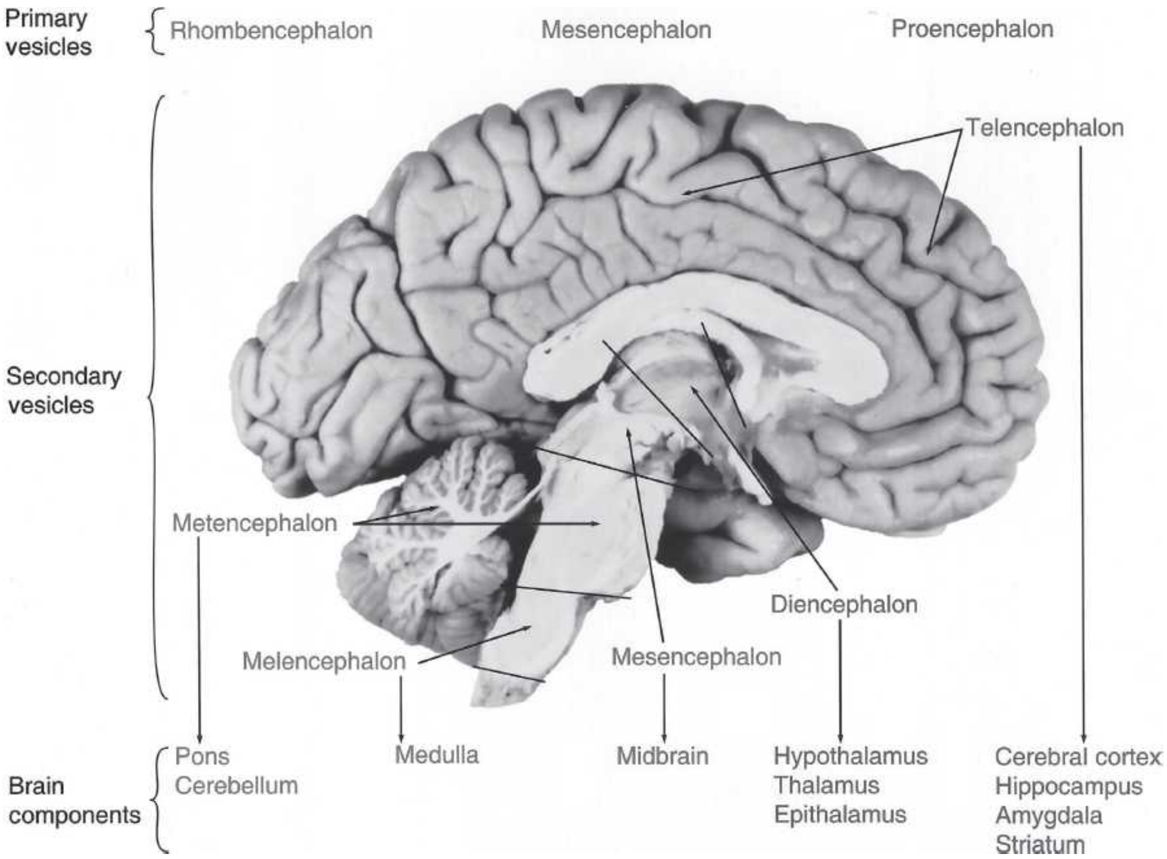
Primary vesicles
Rhombencephalon
Mesencephalon
Proencephalon
Telencephalon
Secondary vesicles
Metencephalon
Diencephalon
Melencephalon
Mesencephalon
Brain components
Pons
Cerebellum
Medulla
Midbrain
Hypothalamus
Thalamus
Epithalamus
Cerebral cortex
Hippocampus
Amygdala
Striatum

FIGURE 1.2–6. Medial view of the human brain with a schematic representation of the vesicles of the neural tube and their derivatives.

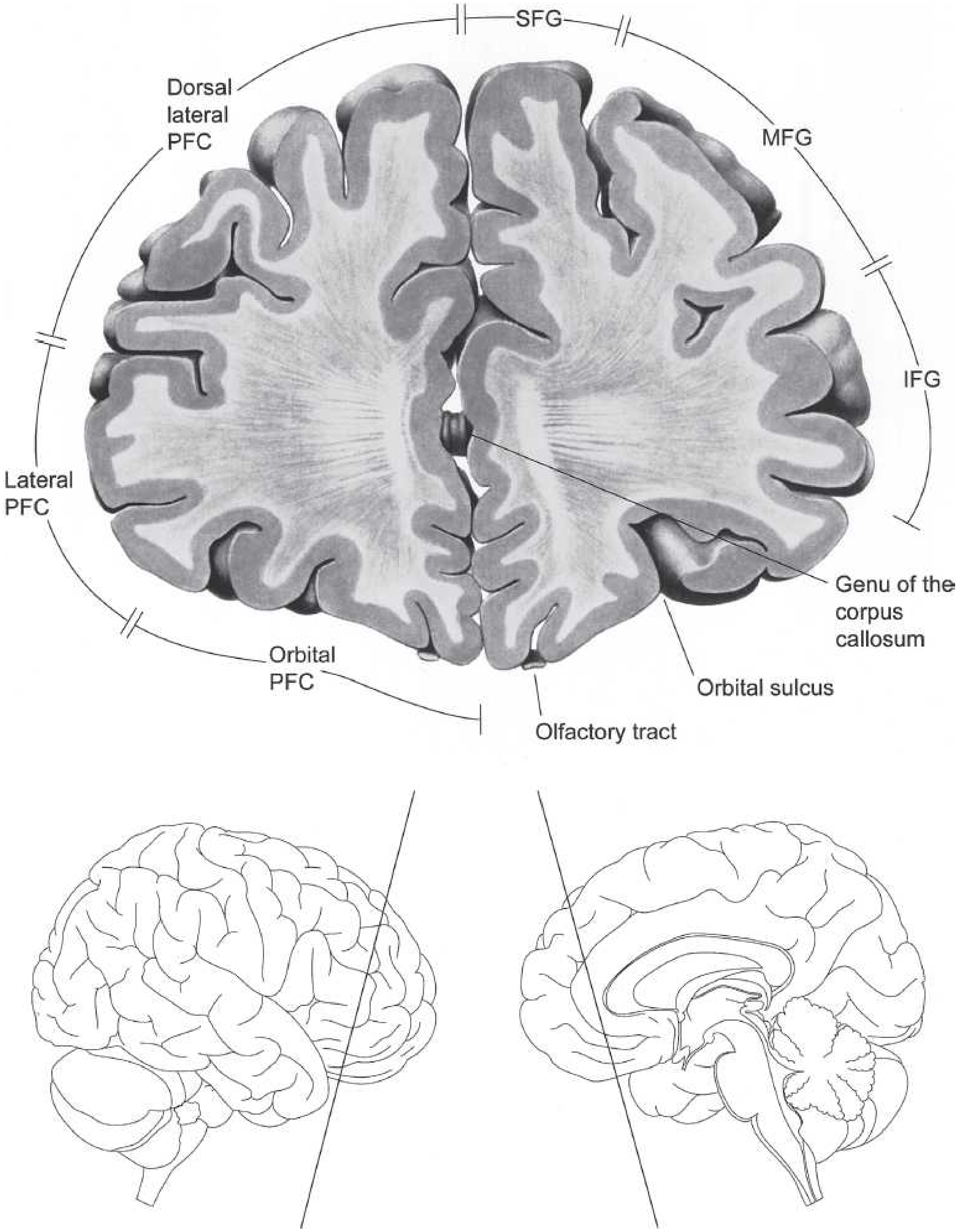
SFG
Dorsal lateral PFC
MFG
IFG
Lateral PFC
Genu of the corpus callosum
Orbital PFC
Orbital sulcus
Olfactory tract

FIGURE 1.2–7. Drawing of a coronal section just anterior to the genu of the corpus callosum of a human brain. The *inset* below indicates the level of the section. IFG, inferior frontal gyrus; MFG, middle frontal gyrus; PFC, prefrontal cortex; SFG, superior frontal gyrus. (Adapted with permission from Nieuwenhuys R, Voogd J, Huijzen C. Brain slices. In: *The Human Central Nervous System: A Synopsis and Atlas*. 4th ed. Springer Nature, 2008:141.)

Beneath the outer mantle of the cerebral cortex are many other major brain structures, such as the caudate nucleus, the putamen, and the globus pallidus (see Fig. 1.2–8). These structures are components of the basal ganglia, a system involved in the control of movement and certain cognitive processes. The hippocampus and the amygdala, components of the limbic system, are located deep in the medial temporal lobe (Figs. 1.2–9 to 1.2–11). In addition, the derivatives of the diencephalon, such as the thalamus and the hypothalamus, are prominent internal structures; the thalamus is a relatively large structure composed of numerous nuclei that have distinct patterns of connectivity with the cerebral cortex (see Figs. 1.2–9 to 1.2–11). In contrast, the hypothalamus is a much smaller structure involved in autonomic and endocrine functions.

White Matter Tracts

The cerebral hemispheres contain billions of myelinated axons or fibers, giving the white matter its characteristic color, which carry information to and from the cerebral cortex. These axons are bundled into white matter tracts that include projection, commissural, and associational fibers.

Projection Fibers. Two of the major projection fiber systems comprise fibers that originate in the cerebral cortex and project to subcortical targets (corticofugal) and fibers that originate outside of the telencephalon and project to the cerebral cortex (corticopetal). Examples of these are the corticothalamic and thalamocortical projections, respectively. These projection fibers travel through the *internal capsule*, a compact bundle of fibers that is structurally associated with the thalamus and lenticular nucleus (i.e., the putamen and globus pallidus considered as one structure). In each cerebral hemisphere, the internal capsule is bordered laterally by the lenticular nucleus and medially by the thalamus and head of the caudate (Fig. 1.2–12). Other fiber systems, such as the corticopontine, corticospinal, and corticobulbar tracts, descend from the cortex through the internal capsule and cerebral peduncle to reach their destinations in the pons, spinal cord, and brainstem. All the fibers traveling through the internal capsule form the *corona radiata* (see Fig. 1.2–8), a fan-like structure that sits just above the internal capsule.

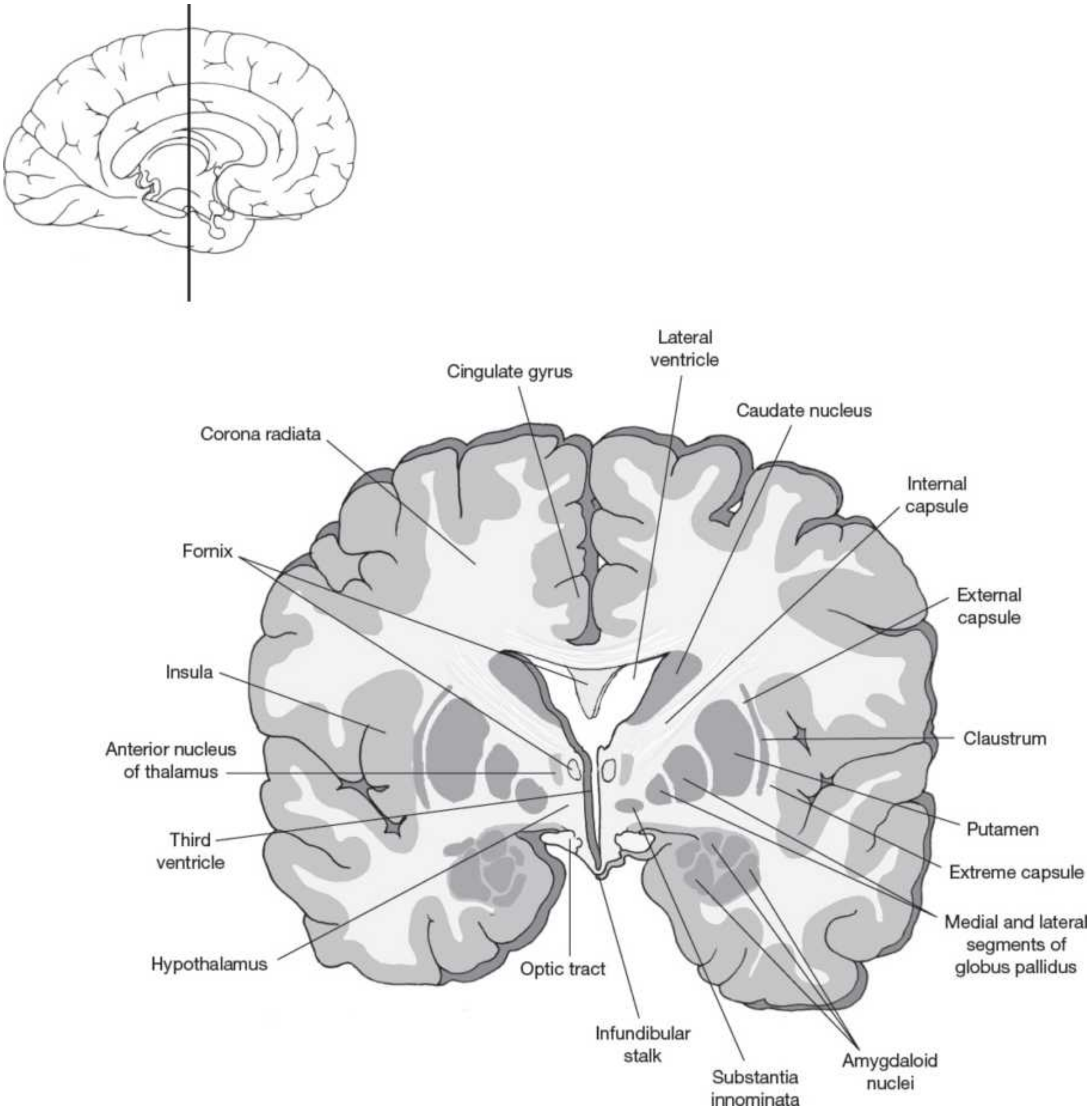
Lateral ventricle
Cingulate gyrus
Caudate nucleus
Corona radiata
Internal capsule
Fornix
External capsule
Insula
Claustrum
Anterior nucleus of thalamus
Putamen
Third ventricle
Extreme capsule
Medial and lateral segments of globus pallidus
Hypothalamus
Optic tract
Infundibular stalk
Substantia innominata
Amygdaloid nuclei

FIGURE 1.2–8. Drawing of a coronal section at the level of the anterior thalamus. The *inset* indicates the approximate level of the section. (Adapted with permission from Siegel A, Sapru HN. *Essential Neuroscience*. 4th ed. Wolters Kluwer; 2018.)

The internal capsule has been divided into five regions, with the name and location of each region based on its relationship with the lenticular nucleus (see Fig. 1.2–12). In addition, each region of the internal capsule contains different fiber systems. The *anterior limb* lies between the lenticular nucleus and the head of the caudate and carries frontopontine fibers and fibers interconnecting the thalamus and the frontal cortex. The *posterior limb*, the largest component, is located between the lenticular nucleus and the thalamus and conveys corticospinal fibers. The *genu* is the intersection of the anterior and posterior limbs and carries corticobulbar fibers. The *retrolenticular limb* lies posterior or behind the lenticular nucleus, and fibers in this region of the internal capsule form the bulk of the *optic radiation*, the large group of fibers projecting from the lateral geniculate thalamic nucleus to the primary visual cortex in the occipital lobe. The *sublenticular limb* lies inferior to the lenticular nucleus and contains fibers of the *auditory radiation*, a collection of fibers connecting the medial geniculate thalamic nucleus with the primary auditory cortex in the temporal lobe.

Commissural Fibers. Commissural fibers interconnect areas in the two cerebral hemispheres with each other. The two main commissural fiber systems are the corpus callosum and the anterior commissure. The *corpus callosum* is the largest fiber bundle in the brain, containing roughly 300 million axons. Most of these axons interconnect cortical regions in one lobe with homotopic (i.e., similarly placed) regions in the opposite lobe. However, heterotopic connections (i.e., those that link dissimilar cortical regions) also are carried in the corpus callosum. Almost all cortical regions are connected via the corpus callosum, with the notable exceptions of the hand area of the motor and somatosensory cortices and all of the primary visual cortex except the portion representing areas adjacent to the vertical midline.

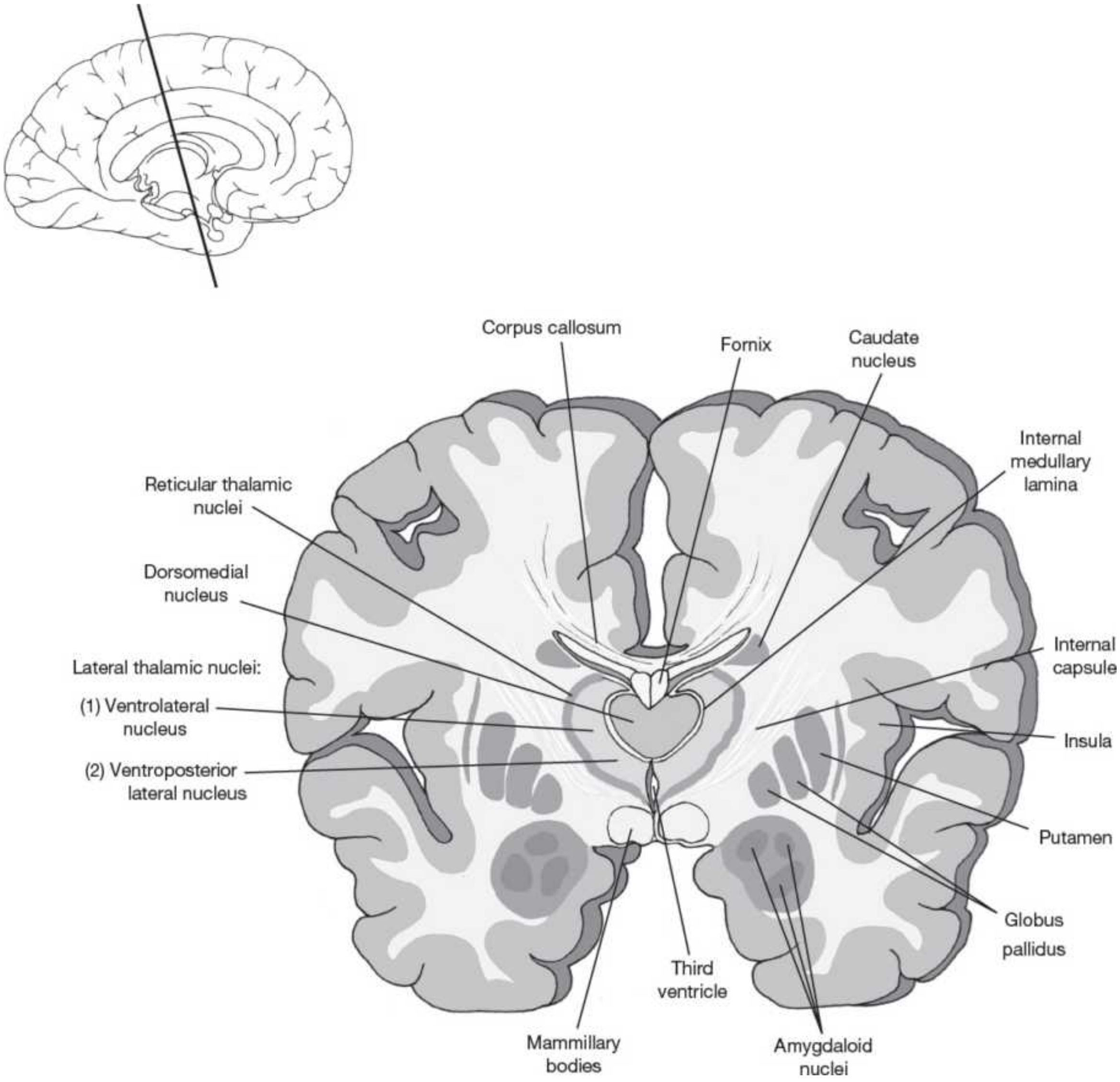
Corpus callosum
Fornix
Caudate nucleus
Internal medullary lamina
Reticular thalamic nuclei
Dorsomedial nucleus
Lateral thalamic nuclei:
(1) Ventrolateral nucleus
(2) Ventroposterior lateral nucleus
Internal capsule
Insula
Putamen
Globus pallidus
Third ventricle
Mammillary bodies
Amygdaloid nuclei

FIGURE 1.2–9. Drawing of a coronal section at the level of the middle thalamus. The *inset* indicates the approximate level of the section. (Adapted with permission from Siegel A, Sapru HN. *Essential Neuroscience*. 4th ed. Wolters Kluwer, 2018.)

The corpus callosum consists of four parts (Fig. 1.2–13). The corpus callosum starts at the *rostrum* and then curves anteriorly and dorsally to form the *genu*. The *body* of the corpus callosum is the largest part and gives way to the *splenium*, the enlarged, rounded posterior end. Sometimes the narrow portion of the corpus callosum between the body and splenium is referred to as the *isthmus*. Axons carrying higher-order cognitive and sensory information from the prefrontal, temporal, and parietal cortices primarily travel through the genu and splenium, whereas visual, auditory, and somatosensory information is carried predominantly in the body and isthmus of the corpus callosum.

The anterior commissure is a compact bundle of fibers that is caudal to the corpus callosum and crosses the midline in front of the fornix (see Fig. 1.2–8). The anterior commissure interconnects areas in the two temporal lobes and fibers from the anterior olfactory nucleus. Smaller commissural fiber tracts include the posterior commissure, which connects caudal portions of the diencephalon, and the hippocampal commissure, which interconnects the two hippocampal formations.

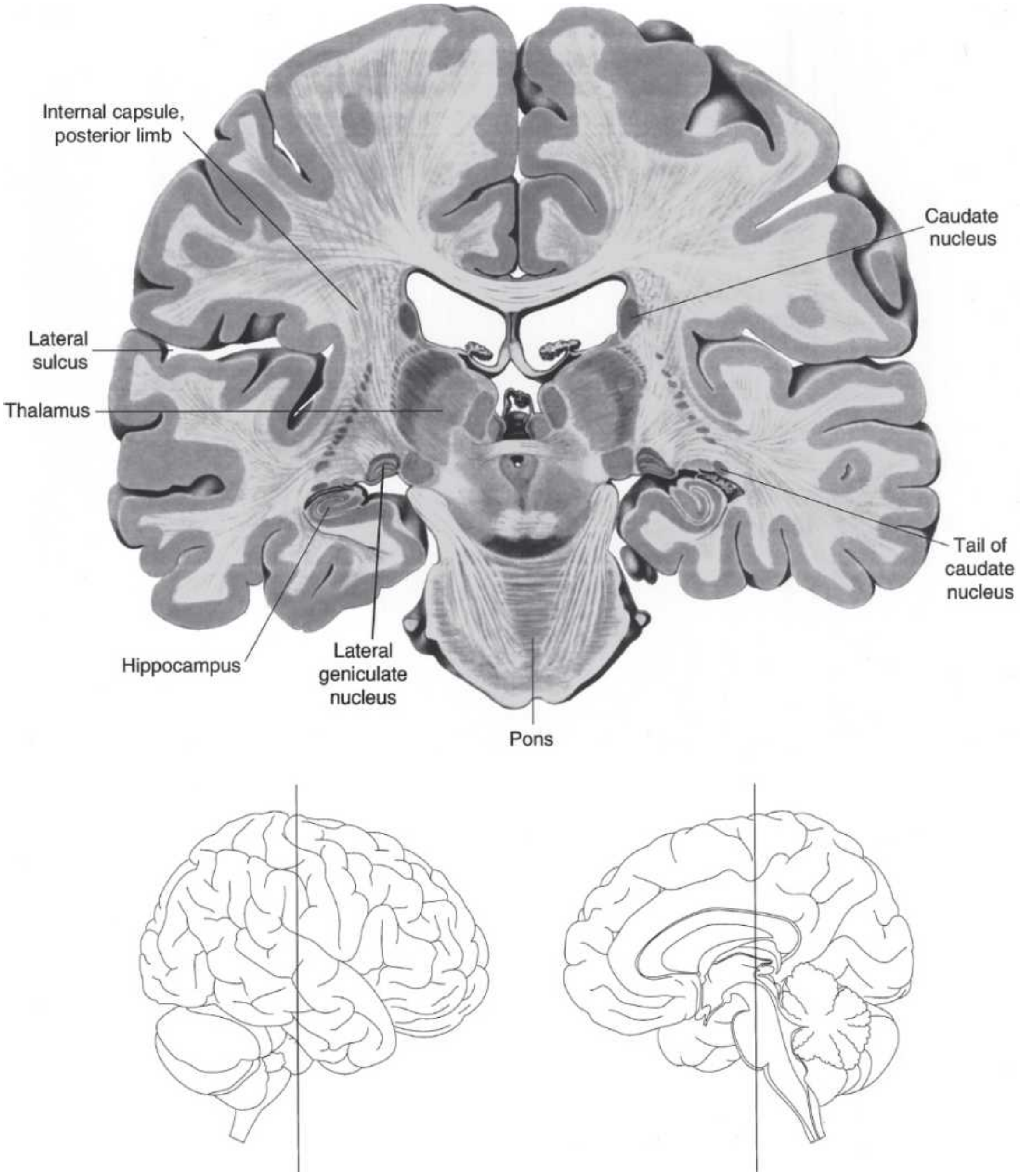
Internal capsule, posterior limb
Caudate nucleus
Lateral sulcus
Thalamus
Tail of caudate nucleus
Hippocampus
Lateral geniculate nucleus
Pons

FIGURE 1.2–10. Drawing of a coronal section through the posterior thalamus. The *inset* below indicates the level of the section. (Adapted with permission from Nieuwenhuys R, Voogd J, Huijzen C. Brain slices. In: *The Human Central Nervous System: A Synopsis and Atlas*. 4th ed. Springer Nature, 2008:147.)

Associational Fibers. Associational fibers connect cortical areas within a hemisphere and range in size from very short fibers that connect areas within the same lobe to longer fibers that connect areas within different lobes. *Short association fibers* connect adjacent gyri and are often called *U fibers* because they form a U connecting one gyrus to another gyrus (Fig. 1.2–14). There are five major tracts of *long association fibers* that connect distant cortical areas within the same hemisphere. The *superior longitudinal fasciculus* is located laterally within the hemisphere above the insula and connects frontal, parietal, and occipital cortices. The *arcuate fasciculus* interconnects the frontal and temporal lobes. The *uncinate fasciculus* is a curved fiber bundle that connects the orbital portion of the frontal lobe with the anterior region of the temporal lobe. As its name implies, the *inferior occipitofrontal fasciculus* connects the occipital and frontal lobe in a bundle of fibers that courses ventrally and laterally within the hemisphere. The *cingulum* lies within the white matter under the cingulate gyrus and connects this gyrus with the parahippocampal gyrus. The *inferior longitudinal fasciculus* connects the temporal and occipital lobes. These fiber bundles are not discrete, point-to-point pathways between cortical regions but are continuous pathways with fibers entering and leaving all along their course. Other associational fiber bundles include the *external capsule*, which is sandwiched between the claustrum and the putamen, and the *extreme capsule*, which lies between the claustrum and the insular cortex (see Fig. 1.2–8).

Disturbances in the connectivity within and between hemispheres have been implicated in the pathophysiology of schizophrenia. For example, MRI studies of individuals with schizophrenia have revealed lower white matter density in the corpus callosum, internal capsule, and anterior commissure. In addition, studies using DTI, which provides information on the organization and microstructure of tissue, have shown abnormalities in almost every major white matter tract, including the corpus callosum, internal capsule, cingulum bundle, uncinate fasciculus, and fornix, in patients with schizophrenia, as well as altered microstructure in the corpus callosum and uncinate fasciculus in patients with depression. Abnormalities in white matter tracts have also been reported in other neuropsychiatric disorders. For example, MRI studies have revealed a reduction in the cross-sectional area of the corpus callosum in individuals with Alzheimer disease as well as in children with autism.

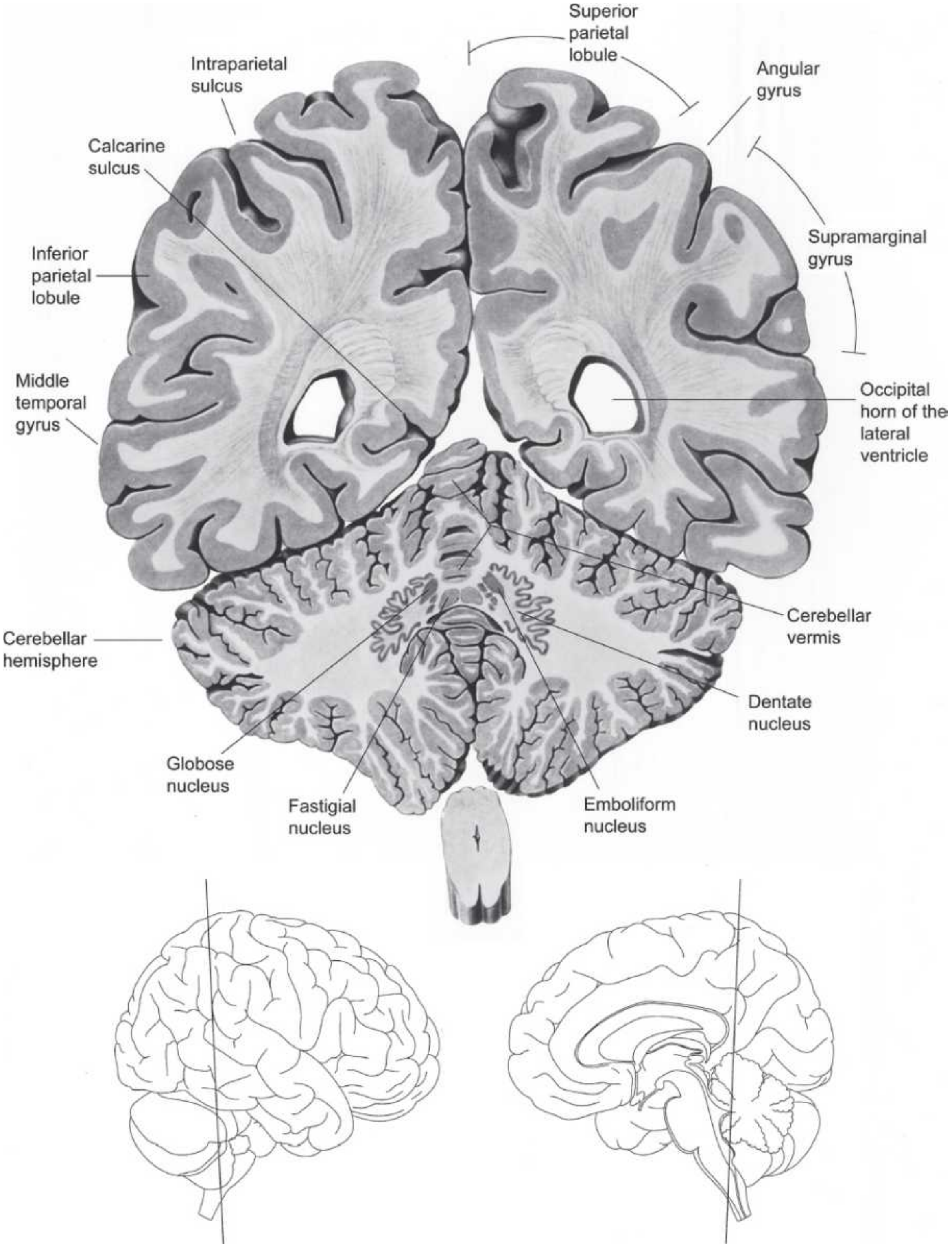
Superior parietal lobule
Intraparietal sulcus
Angular gyrus
Calcarine sulcus
Supramarginal gyrus
Inferior parietal lobule
Middle temporal gyrus
Occipital horn of the lateral ventricle
Cerebellar vermis
Cerebellar hemisphere
Dentate nucleus
Globose nucleus
Fastigial nucleus
Emboliform nucleus

FIGURE 1.2–11. Drawing of a coronal section through the cerebral hemispheres just posterior to the splenium of the corpus callosum and through the deep nuclei of the cerebellum. The *inset* below indicates the level of the section. (Adapted with permission from Nieuwenhuys R, Voogd J,

Huijzen C. Brain Slices. In: *The Human Central Nervous System: A Synopsis and Atlas*. 4th ed. Springer Nature; 2008:150.)

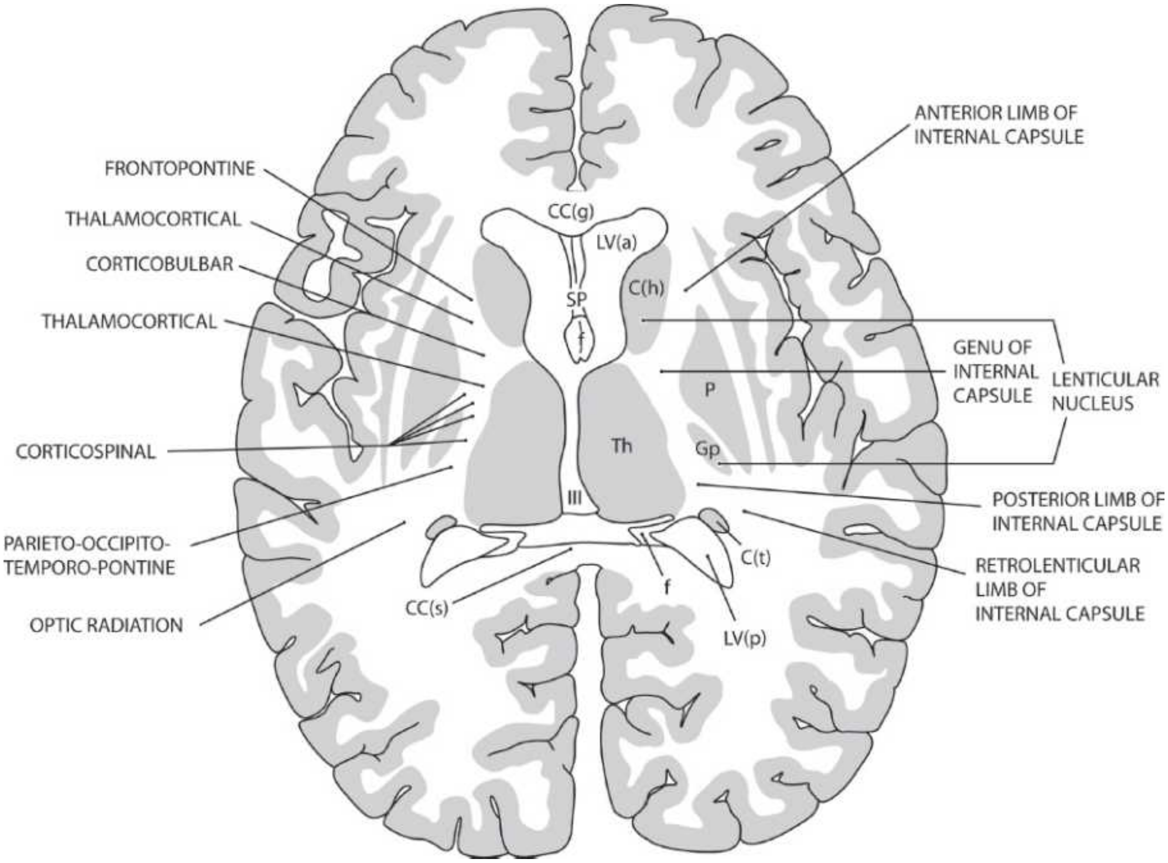
FRONTOPONTINE
THALAMOCORTICAL
CORTICOBULBAR
THALAMOCORTICAL
CORTICOSPINAL
PARIETO-OCCIPITO-TEMPORO-PONTINE
OPTIC RADIATION
CC(g)
LV(a)
SP
f
C(h)
P
Th
Gp
III
C(t)
f
CC(s)
LV(p)
ANTERIOR LIMB OF INTERNAL CAPSULE
GENU OF INTERNAL CAPSULE
LENTICULAR NUCLEUS
POSTERIOR LIMB OF INTERNAL CAPSULE
RETROLENTICULAR LIMB OF INTERNAL CAPSULE

FIGURE 1.2–12. A horizontal section through the cerebrum shows the location of the internal capsule fibers in relation to components of the basal ganglia and the corpus callosum. CC(g), corpus callosum, genu; CC(s), corpus callosum, splenium; C(t), caudate, tail; LV(a), lateral ventricle, anterior horn; LV(p), lateral ventricle, posterior horn; SP, septum pellucidum; Th, thalamus. (Adapted with permission from Gilman S, Winans Newman S. Manter and Gatz's *Essentials of Clinical Neuroanatomy and Neurophysiology*. 10th ed. F.A. Davis; 2003.)

Ventricular System

As the neural tube fuses during development, the cavity of the neural tube becomes the ventricular system of the brain. It is composed of two C-shaped *lateral ventricles* in the cerebral hemispheres that can be divided further into five parts: the anterior horn (which is located in the frontal lobe), the body of the ventricle, the inferior or temporal horn in the temporal lobe, the posterior or occipital horn in the occipital lobe, and the atrium (Fig. 1.2–15). The *foramina of Monro* (interventricular foramina) are the two apertures that connect the two lateral ventricles with the *third ventricle*, which is found on the midline of the diencephalon. The *cerebral aqueduct* connects the third ventricle with the *fourth ventricle* in the pons and the medulla.

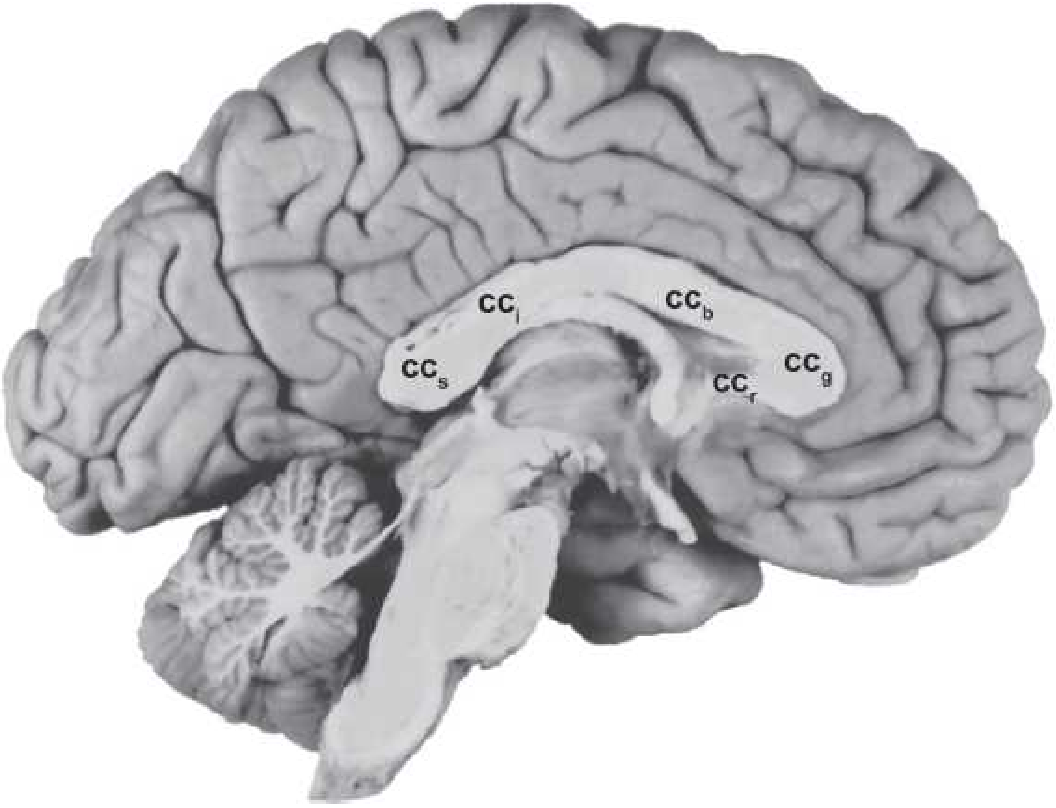

FIGURE 1.2–13. Photograph of the medial view of the right cerebral hemisphere of a human brain illustrating the different parts of the corpus callosum (cc). cc_r, rostrum; cc_g genu; cc_b, body; cc_i, isthmus; cc_s, splenium.

The ventricular system is filled with cerebrospinal fluid (CSF), a colorless liquid containing low concentrations of protein, glucose, and potassium and relatively high concentrations of sodium and chloride. Most (70 percent) of the CSF is produced at the choroid plexus located in the walls of the lateral ventricles and in the roof of the third and fourth ventricles. The *choroid plexus* is a complex of ependyma, pia, and capillaries that invaginate the ventricle. In contrast to other parts of the brain, the capillaries in the choroid plexus are fenestrated, which allows substances to pass out of the capillaries and through the pia mater. The ependymal or choroid epithelial cells, however, have tight junctions between cells to prevent the leakage of substances into the CSF; this provides what is sometimes referred to as the *blood–CSF barrier*. In other parts of the brain, the endothelial cells of the capillaries exhibit tight junctions that prevent the movement of substances from the blood to the brain; this is referred to as the *blood–brain barrier*.

The CSF is constantly produced and circulates through the lateral ventricles to the third ventricle and then to the fourth ventricle. The CSF then flows through the medial and lateral apertures to the cisterna magna and pontine cistern and, finally, travels over the cerebral hemispheres to be absorbed by the arachnoid villi and released into the superior sagittal sinus. Disruptions in the flow of the CSF usually cause some form of hydrocephalus; for example, if an intraventricular foramen is occluded, the associated lateral ventricle becomes enlarged, but the remaining components of the ventricular system remain normal.

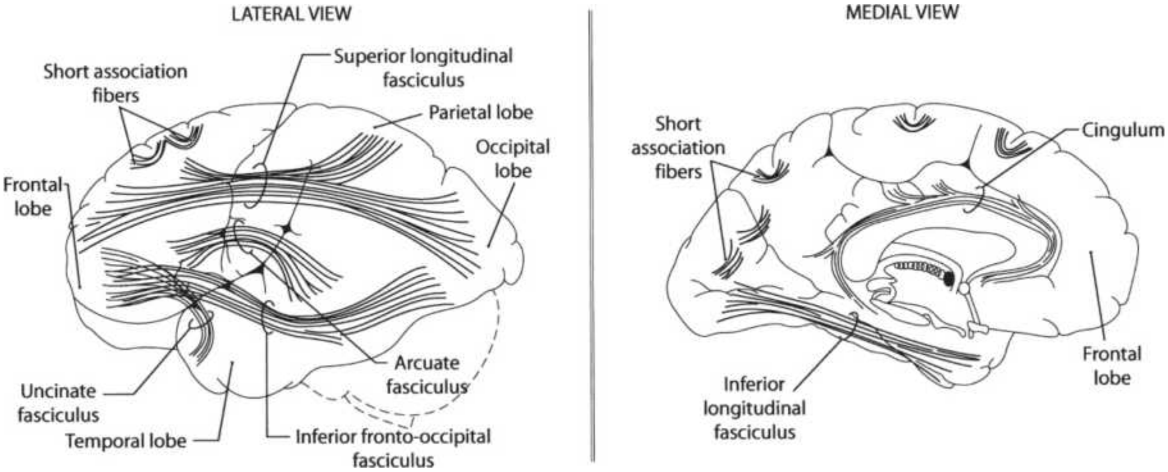
LATERAL VIEW
Short association fibers
Superior longitudinal fasciculus
Parietal lobe
Occipital lobe
Frontal lobe
Uncinate fasciculus
Temporal lobe
Arcuate fasciculus
Inferior fronto-occipital fasciculus
MEDIAL VIEW
Short association fibers
Cingulum
Frontal lobe
Inferior longitudinal fasciculus

FIGURE 1.2–14. Drawings illustrating the main associational fiber tracts as visualized from lateral **(left panel)** and medial **(right panel)** aspects of the left hemisphere. (Adapted from Haines DE. *Fundamental Neuroscience for Basic and Clinical Applications*. 3rd ed. Churchill Livingstone; 2006:253.)

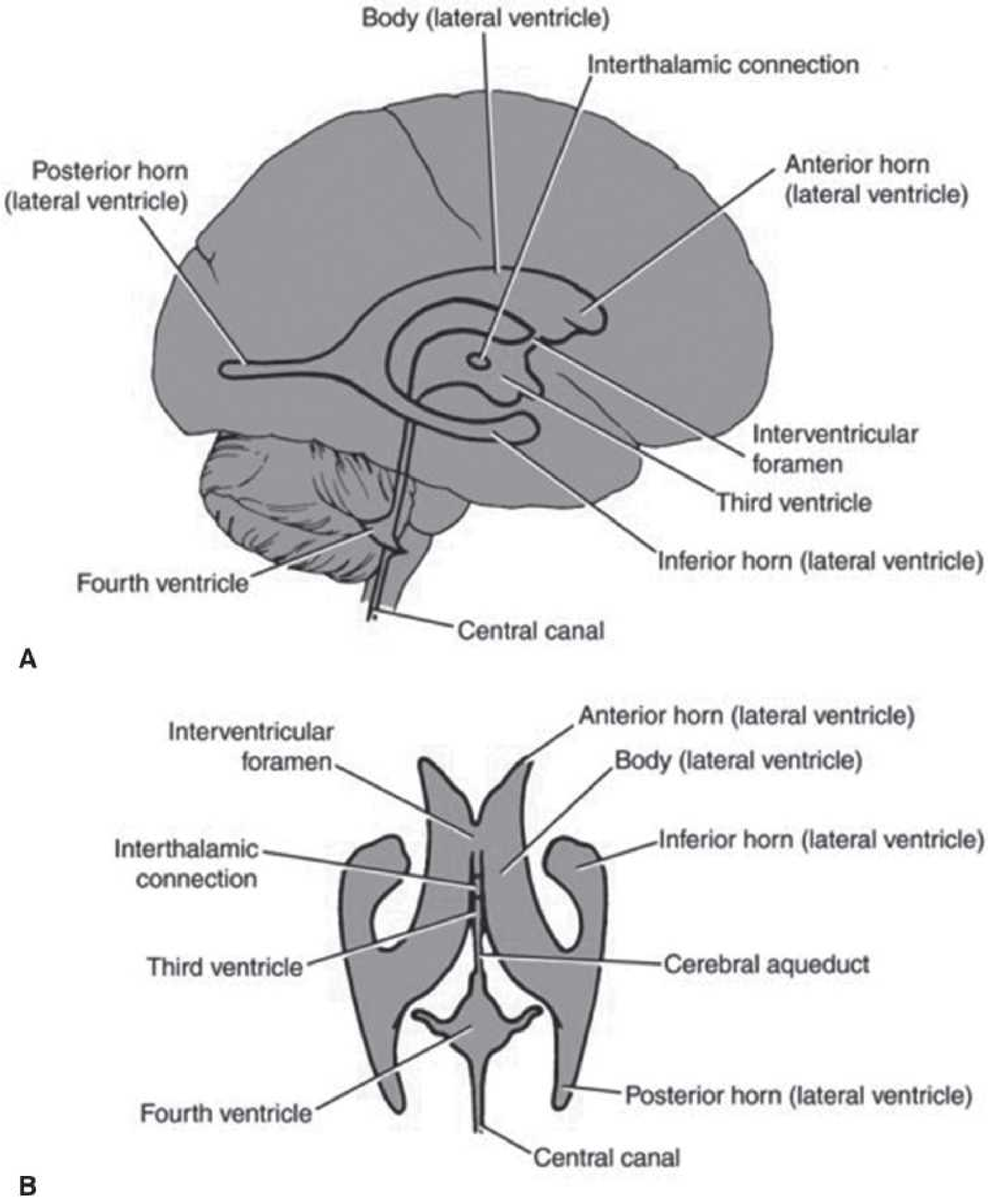
Body (lateral ventricle)
Interthalamic connection
Posterior horn (lateral ventricle)
Anterior horn (lateral ventricle)
Interventricular foramen
Third ventricle
Inferior horn (lateral ventricle)
Fourth ventricle
Central canal
A
Interventricular foramen
Anterior horn (lateral ventricle)
Body (lateral ventricle)
Interthalamic connection
Inferior horn (lateral ventricle)
Third ventricle
Cerebral aqueduct
Fourth ventricle
Posterior horn (lateral ventricle)
Central canal
B

FIGURE 1.2–15. A: Diagram of the ventricles of the brain and central canal of the spinal cord. **B:** A three-dimensional representation of the ventricles of the brain. (Reprinted with permission from Splittgerber R. Snell's *Clinical Neuroanatomy*. 8th ed. Wolters Kluwer; 2019.)

Several functions are attributed to the CSF: it serves to cushion the brain against trauma, maintain and control the extracellular environment, and spread endocrine hormones. Because the CSF bathes the brain and is in direct communication with extracellular fluid, it is possible to measure certain compounds in the CSF as a correlate of the amount of that substance in the brain. For example, CSF levels of amyloid beta$_{1\text{-}42}$ and total tau are thought to be markers of altered amyloid metabolism and axonal damage, respectively. The concentration of amyloid beta$_{1\text{-}42}$ and total tau in samples of the CSF taken in a lumbar puncture have been used in studies examining the progression from mild cognitive impairment to Alzheimer disease. In addition, CSF measures of amyloid beta$_{1\text{-}42}$ and total tau are well-established predictive biomarkers of Alzheimer dementia in patients with mild cognitive impairment.

FUNCTIONAL BRAIN SYSTEMS

The relationships between the organizational principles and the structural components of the human brain are illustrated in three functional systems: the thalamocortical, basal ganglia, and limbic systems.

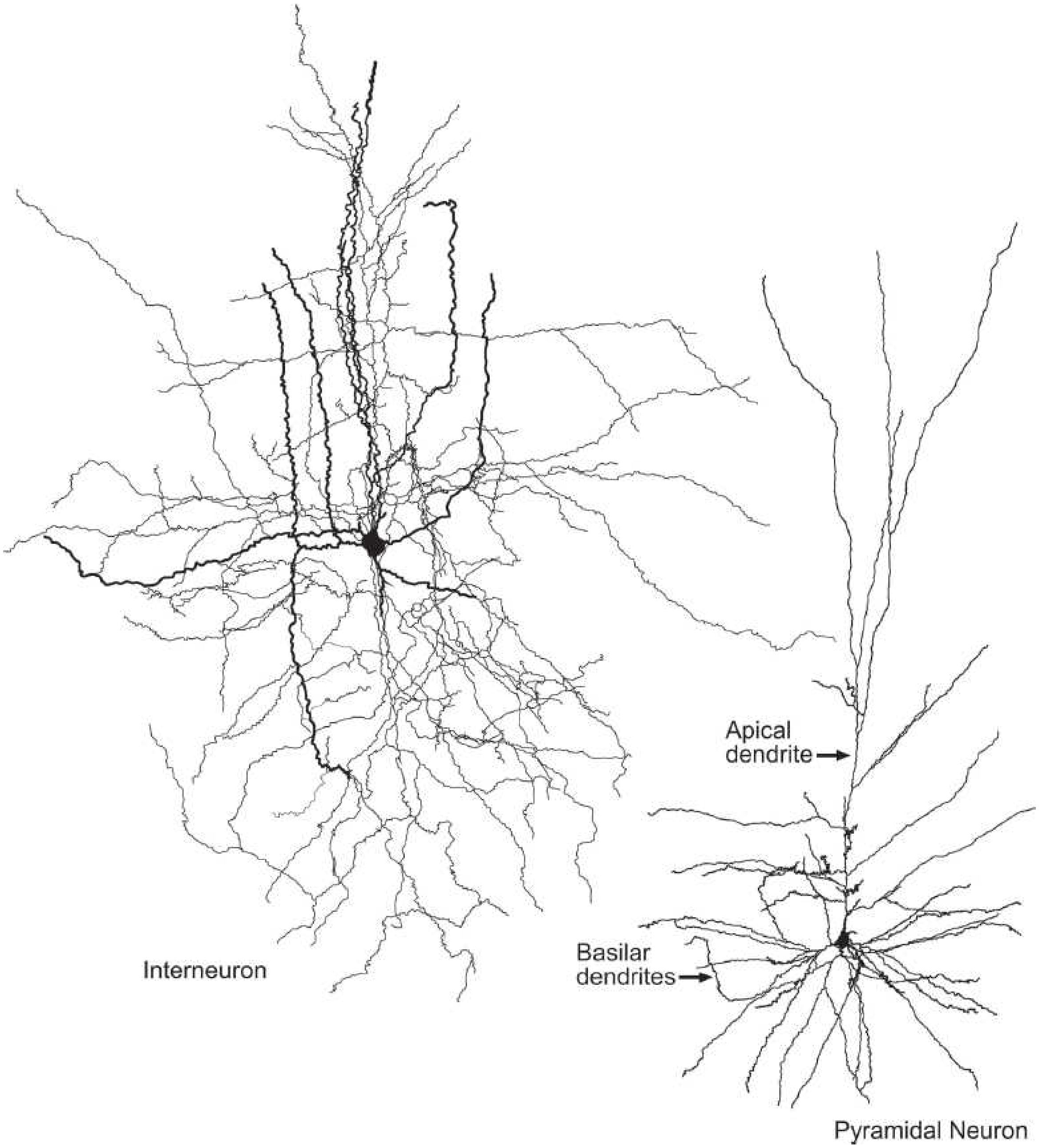

FIGURE 1.2–16. Drawings of an interneuron **(left)** and a pyramidal neuron **(right)**. Note the difference in the morphology of these two types of neurons. The soma of interneurons tends to be round or ovoid, whereas that of pyramidal neurons generally appears triangular from a two-dimensional perspective. Also, note the difference in the dendritic and axonal arbors between the two cells. The processes arising from the interneuron appear to branch in multiple directions, whereas the pyramidal neuron has prominent, well-defined apical and basilar dendrites. (Permission from: Neuromorpho.org; Ascoli GA, Donohue DE, Halavi M. NeuroMorpho.Org: a central resource for neuronal morphologies. *J Neurosci*. 2007;27(35):9247–9251(NM)-01861; Krimer LS, Zaitsev AV, Czarner G, et al. Cluster analysis-based physiological

classification and morphological properties of inhibitory neurons in layer 2–3 of monkey dorsolateral prefrontal cortex. *J Neurophysiol*. 2005;94(5): 3009–3022.)

Cerebral Cortex

The *cerebral cortex* is a laminated sheet of neurons, several millimeters thick, that covers the cerebral hemispheres. It consists of approximately 22 billion neurons communicating via approximately 165 trillion synapses. These neurons have approximately 12 million km of dendrites, and the cerebral cortex and subcortical regions are interconnected by approximately 100,000 km of axons. More than 90 percent of the total cortical area consists of the *neocortex*, which has a six-layered structure (at least at some point during development). The remainder of the cerebral cortex is referred to as the *allocortex* (cortical regions with other laminar structures) and consists of the *paleocortex* and the *archicortex*, regions that are restricted to the base of the telencephalon and the hippocampal formation, respectively.

Within the neocortex, the two major neuronal cell types are the pyramidal and stellate, or nonpyramidal, neurons (Fig. 1.2–16). *Pyramidal neurons*, which account for approximately 75 percent of all neocortical neurons, usually have a characteristically shaped cell body that gives rise to a single apical dendrite that ascends vertically toward the cortical surface. In addition, these neurons have an array of short dendrites that spread laterally from the base of the cell. The dendrites of pyramidal neurons are coated with short protrusions, called *dendritic spines*, which are the sites of most of the excitatory synapses to these neurons (Fig. 1.2–17). Most pyramidal cells are excitatory projection neurons that use glutamate as a neurotransmitter. Interestingly, in postmortem studies, subjects with schizophrenia appear to have fewer spines on the dendrites located at the base of pyramidal neurons in deep layer 3 of the DLPFC. This deficit in dendritic spines may be due to a genetic liability in the regulation of actin dynamics, leading to a loss of dendritic spines due to structural instability and/or excessive microglia-mediated pruning of spines. The presence of fewer spines, and presumably of the excitatory synapses they receive, results in less excitation of pyramidal neurons, eventually leading to less output of these pyramidal neurons and ultimately in lower indices of cortical activity.

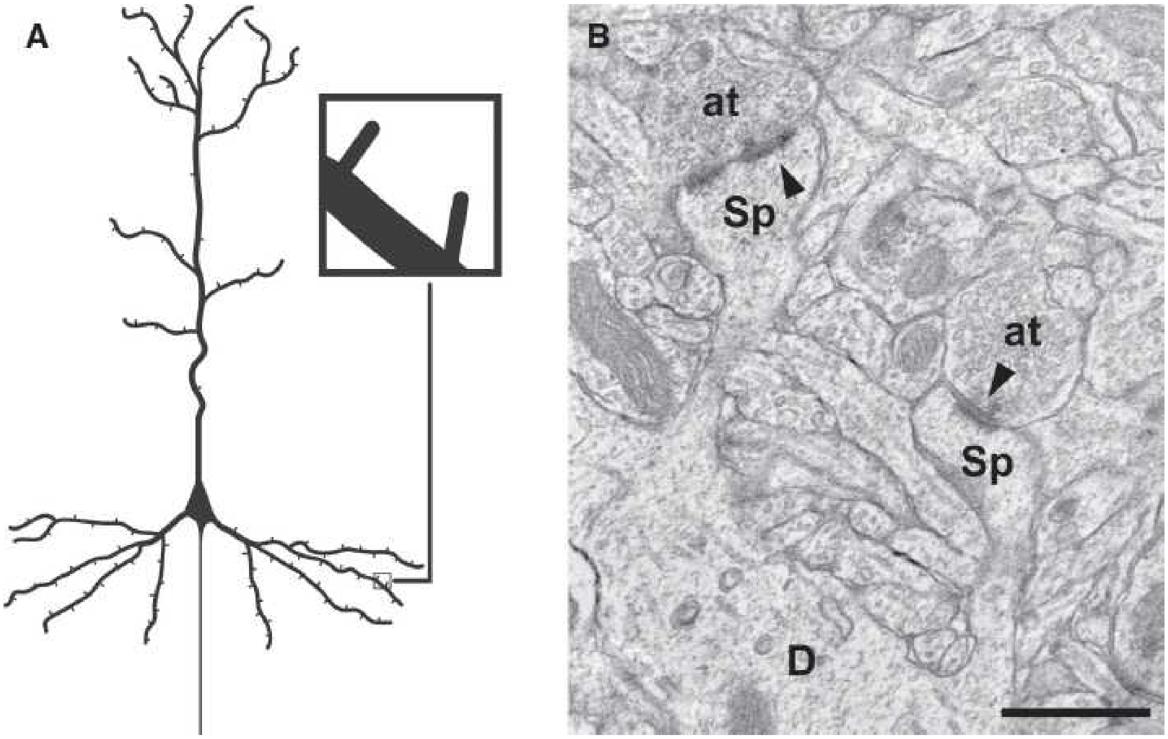

FIGURE 1.2–17. Dendritic spines in the cerebral cortex. **A:** Diagram of a cortical pyramidal neuron illustrating the location of dendritic spines. **B:** Electron micrograph from monkey prefrontal cortex showing two dendritic spines (*Sp*) emanating from a single dendrite (*D*), both of which receive an asymmetric synapse (*arrowhead*) from an axon terminal *(at)*. Calibration bar = 200 nm.

In contrast, *nonpyramidal cells* are generally small, local circuit neurons, many of which use the inhibitory neurotransmitter GABA (Figs. 1.2–16). Also known as *interneurons*, the axons of cortical GABA cells arborize within the gray matter, and most do not project out of the cortical region in which they reside. Multiple different subtypes of GABA neurons can be found in the cerebral cortex, and these can be distinguished biochemically, electrophysiologically, and morphologically. For example, subpopulations of GABA cells can be distinguished by the presence of certain neuropeptides or calcium-binding proteins. In addition, the organization of the axonal arbor and synaptic targets of the axon terminals differ greatly across these different subtypes. The chandelier class of GABA cells expresses the calcium-binding protein parvalbumin (PV) and has axon terminals that are arrayed as distinct vertical structures termed cartridges (Fig. 1.2–18). These axon terminals form inhibitory or symmetric synapses exclusively with the axon initial segments of pyramidal cells. PV-containing basket neurons form symmetric synapses onto the cell bodies and dendrites of pyramidal neurons. PV-containing neurons are predominantly located in layers 3 and 4. In contrast, the Martinotti class of GABA neurons contains the neuropeptide somatostatin and forms symmetric synapses onto the tuft dendrites of pyramidal neurons. Some double-bouquet GABA neurons have radially oriented axonal arbors, contain somatostatin and the calcium-binding protein calbindin, and form symmetric synapses onto the distal dendritic shafts and spines of pyramidal neurons. The neuropeptide cholecystokinin (CCK) is present in a separate population of basket cells which also express the brain cannabinoid receptor 1 (CB1). These CCK/CB1 basket cells predominantly synapse onto the dendritic shafts and cell bodies of pyramidal neurons. In contrast, calretinin-containing double-bouquet cells form symmetric synapses predominantly onto the dendritic shafts of other GABA neurons. Calretinin-containing Cajal–Retzius cells reside solely in layer 1 and target the tuft dendrites of pyramidal neurons.

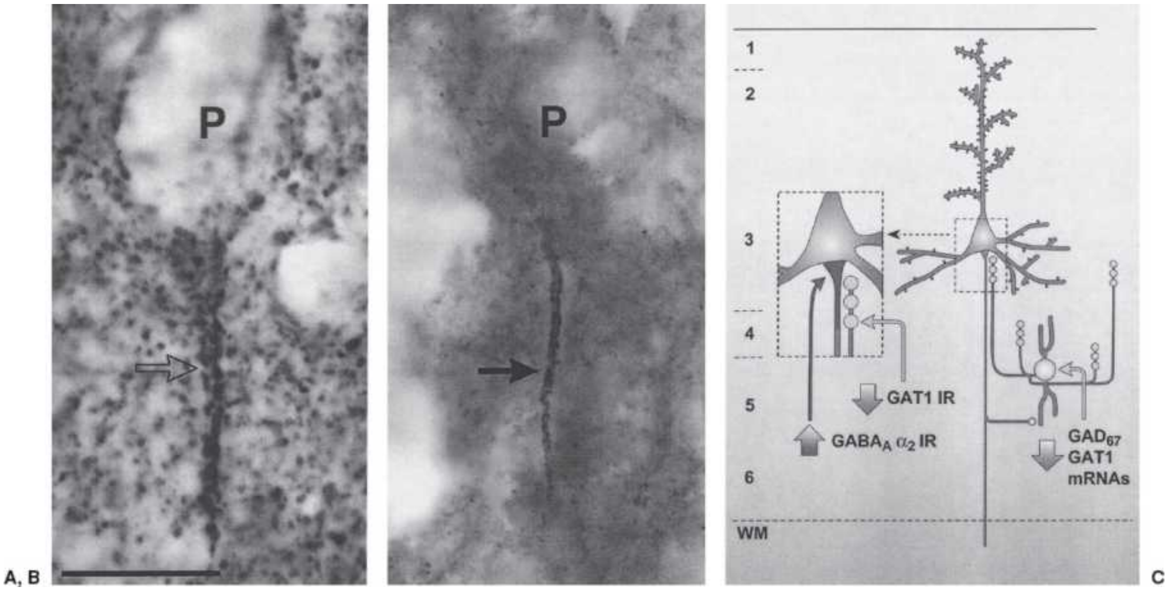

FIGURE 1.2–18. A: Brightfield photomicrograph of chandelier neuron axon terminal *(arrow)* immunostained for the GABA transporter-type 1 (GAT-1). **B:** Brightfield photomicrograph of a pyramidal neuron (*P*) axon initial segment (the site of action potential generation) immunostained for the α_2 subunit of the GABA$_A$ receptor. **C:** Schematic diagram of the synaptic relationship between chandelier and pyramidal neurons illustrating the pre- and postsynaptic changes in schizophrenia. (A and B, reprinted from and C, adapted from Volk DW, Pierri JN, Fritschy JN, Auh S, Sampson AR, Lewis DA. Reciprocal alterations in pre- and postsynaptic inhibitory markers at chandelier cell inputs to pyramidal neurons in schizophrenia. *Cereb Cortex*. 2002;12:1063, by permission of Oxford University Press.)

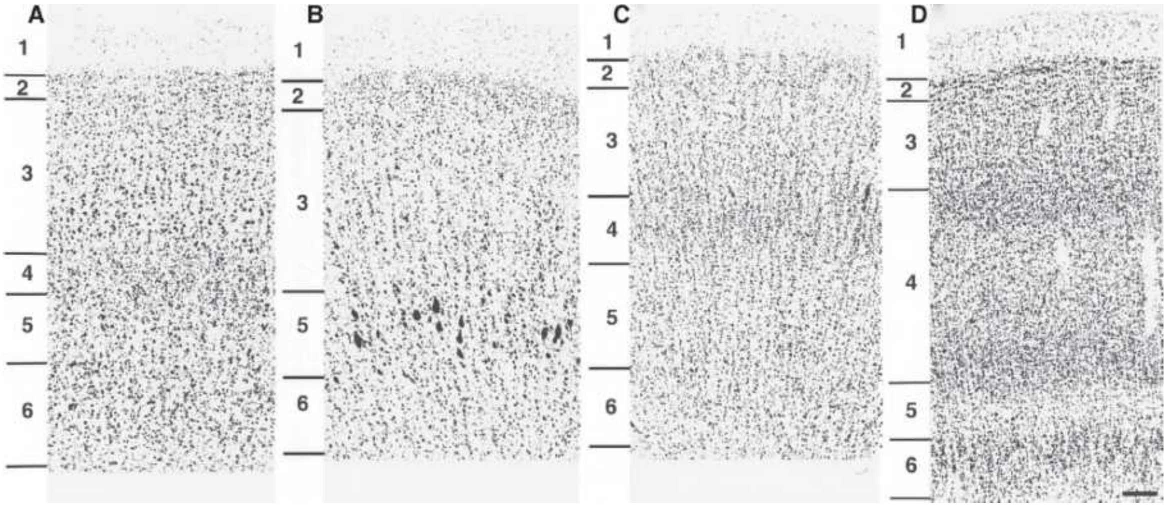

FIGURE 1.2–19. Nissl-stained sections of Brodmann area 46 (DLPFC) **(A)**, area 4 (primary motor cortex) **(B)**, area 41 (primary auditory cortex) **(C)**, and area 17 (primary visual cortex) **(D)** from a control human brain. Note the marked differences in the size and laminar organization of neurons across areas. The large neurons in panel B are Betz cells, which extend their axons to the spinal cord. Arabic numerals indicate the cortical layers. Calibration bar (200 μm) applies to **A–D**.

Many studies have revealed abnormalities in different populations of nonpyramidal neurons in a variety of neurologic and psychiatric disorders. For example, lower numbers of calretinin neurons and morphologic changes in the axonal arbors of PV neurons have been reported in epileptic tissue. In schizophrenia, basket cells express lower levels of PV and the 67 kilodalton form of the biosynthetic enzyme glutamic acid decarboxylase (GAD67). The axon terminals of chandelier neurons have reduced levels of the membrane GABA transporter-type 1 (GAT-1) and greater levels of $GABA_A$ α_2 receptors in postsynaptic pyramidal neurons (see Fig. 1.2–18C). In concert, these findings reflect dysfunctional prefrontal cortical circuitry, which may contribute to the cognitive symptoms observed in schizophrenia.

Neocortical neurons are distributed across six layers of the neocortex; these layers are distinguished by the relative size and packing density of their constituent neurons (Fig. 1.2–19). Each cortical layer tends to receive particular types of inputs and furnish characteristic projections. For example, afferents from thalamic relay nuclei terminate primarily in deep layer 3 and layer 4, and afferents from other cortical areas predominantly terminate in layers 1 to 3. In contrast, corticostriatal projections arise from layer 5 pyramidal cells, and corticothalamic projections originate mainly from layer 6 pyramidal neurons (Fig. 1.2–20). These laminar distinctions provide important clues for dissecting possible pathophysiologic mechanisms in psychiatric disorders. Reports of smaller somal size and lower spine density on deep layer 3 pyramidal neurons in the prefrontal cortex of individuals with schizophrenia suggest that these alterations may be associated with alterations in afferent projections from the medial dorsal thalamic nucleus, the primary thalamic projection nucleus to the prefrontal cortex.

In addition to the horizontal laminar structure, many aspects of cortical organization have a vertical or columnar characteristic. For example, the apical dendrites of pyramidal neurons and the axons of some local circuit neurons have a prominent vertical orientation, indicating that these neural elements may sample the input to or regulate the function of neurons in multiple layers respectively. Afferent inputs to the neocortex from other cortical regions also tend to be distributed across cortical layers in a columnar fashion. Finally, physiologic studies in the somatosensory and visual cortices have shown that neurons in a given column respond to stimuli with particular characteristics, whereas neurons in adjacent columns respond to stimuli with different features.

Although best studied in sensory cortices, this pattern of organization is also present in association cortices. More recent studies in monkeys using tract-tracing techniques have shown that clusters of prefrontal cortical neurons are organized into reciprocally connected, discrete modular stripes that appear to be the analog of columns identified in the visual cortex (see Fig. 1.2–4). It has been hypothesized that this organization may subserve prefrontal working memory and executive functions.

The neocortex can be divided into two general types of regions. Regions with a readily identifiable six-layer appearance are known as the *homotypical cortex* and are found in association regions of the frontal, temporal, and parietal lobes. In contrast, some regions of the neocortex do not have a six-layer appearance. These regions, referred to as the *heterotypical cortex*, include the primary motor cortex, which lacks a defined layer 4 (see Fig. 1.2–19B), and primary sensory regions, which exhibit an expanded layer 5 (see Fig. 1.2–19D). The neocortex can be divided further into discrete areas, each area having a distinctive architecture, a certain set of connections, and a role in particular brain functions. Most subdivisions of the human neocortex have been based on cytoarchitectural features; that is, subdivisions differ in size, packing density, and arrangement of neurons across layers (see Fig. 1.2–19). The most widely used system is that of Korbinian Brodmann (Fig. 1.2–21), who divided the cortex of each hemisphere into 44 numbered areas. Some of these numbered regions correspond closely to functionally distinct areas, such as area 4 (primary motor cortex in the precentral gyrus) and area 17 (primary visual cortex in the occipital lobe). In contrast, other Brodmann areas appear to encompass several cortical zones that differ in their functional attributes. Although Brodmann's brain map has been used extensively in postmortem studies of psychiatric disorders, many of the distinctions among regions are subtle, and the locations of the boundaries between regions vary across individuals.

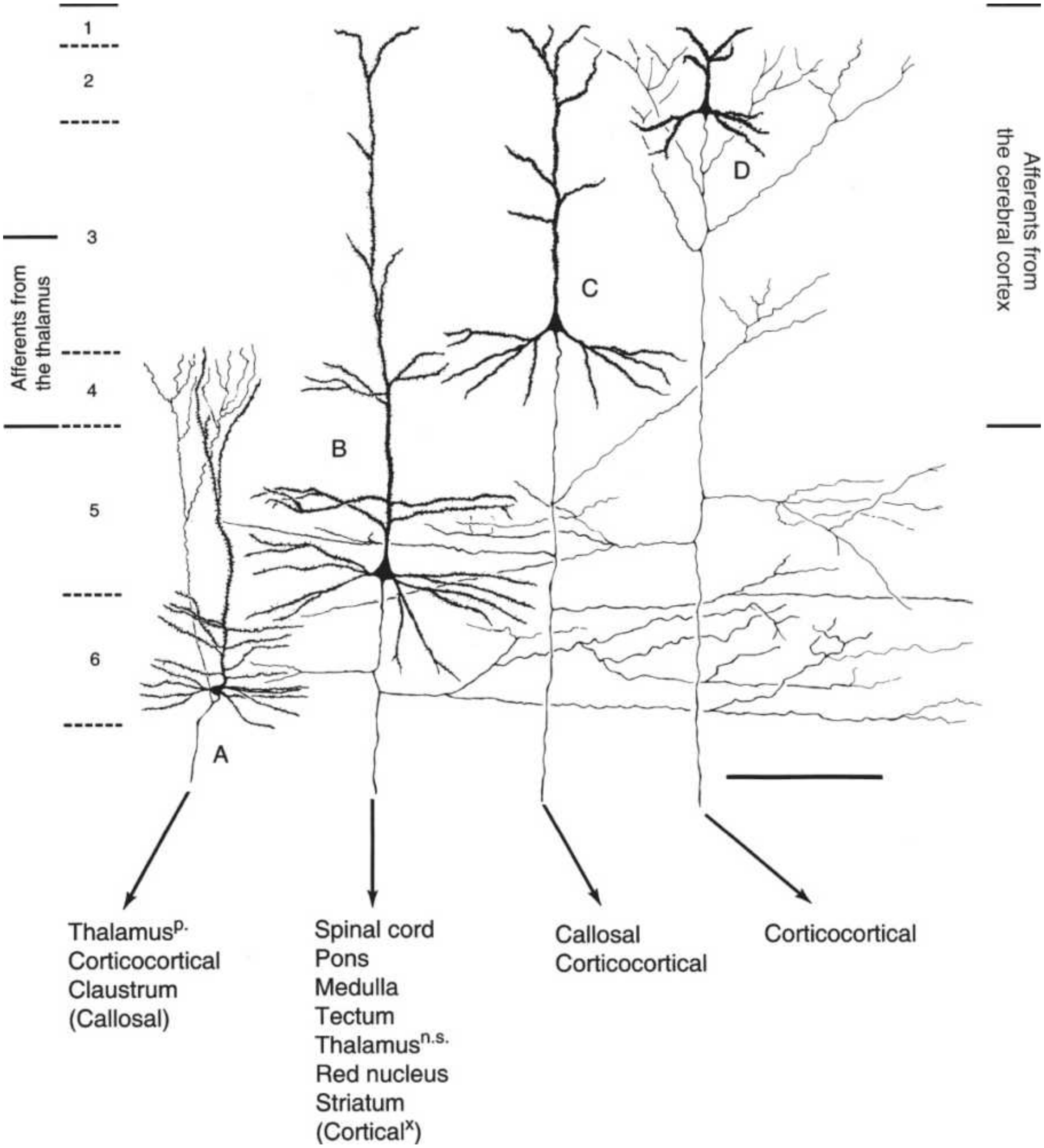
1
2
3
4
5
6
Afferents from the thalamus
Afferents from the cerebral cortex
A
B
C
D
Thalamus$^{p.}$
Corticocortical
Claustrum
(Callosal)
Spinal cord
Pons
Medulla
Tectum
Thalamus$^{n.s.}$
Red nucleus
Striatum
(Cortical^{x})
Callosal
Corticocortical
Corticocortical

FIGURE 1.2–20. Schematic diagram of the laminar origins of efferent projections from the cerebral cortex. These data are mainly derived from the study of monkeys via tract-tracing studies. Parentheses indicate projections that may not arise from the identified layer in all species or in all cortical areas. Note that afferents from the thalamus project mainly to the lower half of layers 3 and 4, whereas cortical afferents terminate in layers 1–3. (Adapted from Jones EG. Laminar distribution of cortical efferent cells. In: Peters A, Jones EG, eds. Cerebral cortex. *Volume 1: Cellular Components of the Cerebral Cortex*. Plenum Press; 1984:535, reproduced with permission of SNCSC.)

Although a given cortical area may receive other inputs, it is heavily innervated by projections from particular thalamic nuclei and from certain other cortical regions either in the same hemisphere (*associational fibers*) or the opposite hemisphere (*commissural fibers*). The patterns of connectivity make it possible to classify cortical regions into different types. *Primary sensory areas* are dominated by inputs from specific thalamic relay nuclei and are characterized by a topographic representation of visual space, the body surface, or the range of audible frequencies on the cortical surface of the primary visual, primary somatosensory, and primary auditory cortices, respectively. These regions project to nearby *unimodal association regions*, which are also devoted to processing information from a particular sensory modality. Output from these regions converges in *multimodal association areas*, such as the prefrontal cortex or the temporoparietal cortical regions. Neurons in these regions respond to complex stimuli and are thought to be mediators of higher cognitive functions. Finally, these regions influence the activity of the *motor areas* of the cerebral cortex that control behavioral responses.

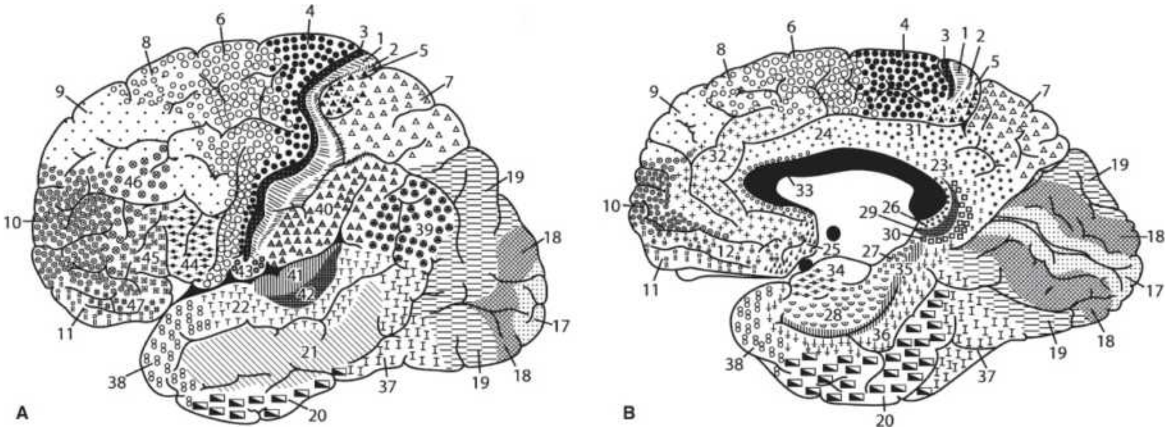

FIGURE 1.2–21. Drawing of the lateral view **(A)** and medial view **(B)** of the cytoarchitectonic subdivisions of the human brain as determined by Brodmann.

Although this classification scheme of cortical regions is accurate in many respects, it fails to account for some of the known complexities of cortical information processing. For example, somatosensory input from the thalamus projects to several distinct topographically organized maps in the cerebral cortex. In addition, information flow within the cortex is not confined to the serial processing route implied in the classification scheme but also involves parallel processing streams, such as sensory input from the thalamus to the primary and the association areas.

Although this discussion has not distinguished between the cerebral hemispheres, certain brain functions, such as language, are localized to one hemisphere. The structural bases for the lateralization of function have not been determined, but some anatomical differences between the cerebral hemispheres have been observed. For example, a portion of the superior temporal cortex, called the *planum temporale*, is generally larger in the left hemisphere than in the right hemisphere. That cortical area, which is located close to the primary auditory cortex and includes the region known as *Wernicke area* (see Fig. 1.2–5), seems to be involved in receptive language functions that are localized to the left hemisphere. In addition, Brodmann area 44 in the left inferior frontal cortex (Broca area) (see Fig. 1.2–5) contains larger pyramidal neurons than the homotopic region of the right hemisphere, a difference that may contribute to the specialization of Broca area for motor speech function. A lesion in Broca area causes broken speech, whereas a lesion in Wernicke area causes wordy speech that does not make sense.

Functional Circuitry

CORTICOCORTICAL CONNECTIONS. Corticocortical connections comprise the majority of connections within the cerebral cortex and can be divided into three basic categories: (1) intrinsic—connections within one cortical area, (2) associational—connections between cortical areas within the same hemisphere, and (3) callosal—connections between cortical areas in different hemispheres. Intrinsic connections are formed by the axon collaterals of pyramidal neurons, whereas associational and callosal connections are formed by the principal axons of pyramidal neurons.

Corticocortical connections are responsible for the coordinated activity of large groups of neurons distributed within and across cortical areas. The synchronization of this activity appears to be involved in the majority of perceptual and cognitive functions ascribed to the cerebral cortex.

The synchronous neural activity or oscillations of these connected groups of neurons generates characteristic patterns of electrical activity that can be recorded with an electroencephalogram (EEG). The greater the number of neurons oscillating together at a particular frequency, the greater the amplitude of the EEG signal above noise. These oscillations are grouped into different frequency bands such as theta (4 to 8 Hz), alpha (8 to 12 Hz), and gamma (30 to 80 Hz). Oscillatory activity in all frequency bands arises from networks of inhibitory, GABAergic neurons whose output modulates the activity of pyramidal neurons.

High-frequency oscillations have been suggested as a possible mechanism for the integration of distributed neuronal networks involved in a variety of mental processes such as perception, attention, and cognition. Gamma-band oscillations in the DLPFC are thought to be a key neural substrate for cognition. Moreover, the synaptic connections between the PV basket subclass of GABA neurons and pyramidal neurons within the supragranular layers of the DLPFC appear to generate this oscillatory activity. Interestingly, individuals with schizophrenia exhibit lower power of prefrontal gamma oscillations when performing cognitive tasks, perhaps a consequence of the abnormalities in the GABAergic system, including PV neurons, observed in postmortem DLPFC from individuals with schizophrenia.

NETWORKS OF CORTICOCORTICAL CIRCUITRY. The study of corticocortical connections had once only been possible in animals with the use of tract-tracing methods. Although improvements in these methods have refined and deepened our knowledge of the projections between different cortical areas, the connections revealed by this methodology can only be inferred to be the same in humans. Over the last 20 years, neuroimaging studies have vastly expanded our knowledge of corticocortical connectivity in humans. However, it is important to note that most of these neuroimaging studies have identified functional connectivity (i.e., brain areas that are coactivated during a specific task) and not physical connections between regions. Evidence of direct anatomical corticocortical connections in the human brain is sparse, although newer imaging techniques such as DTI do provide some structural detail. Large-scale cortical networks linking distributed areas that subserve particular functions, such as attention or memory, have long been suggested and are now more readily studied with the aforementioned neuroimaging techniques. Numerous networks have been described, but three particular networks have received the greatest focus.

Table 1.2–2.
Functional Networks in Cerebral Cortex

Network	Brain Areas Involved	Function
Central executive	Dorsolateral prefrontal cortex Lateral posterior parietal cortex	Attention Working memory Decision making
Salience	Anterior cingulate cortex Ventromedial prefrontal cortex Insula	Detection of relevant incoming stimuli
Default	Medial prefrontal cortex Posterior cingulate cortex Medial posterior parietal cortex	Internal cognition Theory of mind

The *central executive network* includes the DLPFC (see Fig. 1.2–7) and multiple areas of lateral posterior parietal cortex (see Fig. 1.2–5) and is involved in certain cognitive functions such as attention, working memory, and decision making (Table 1.2–2). The *salience network*, comprised of the anterior cingulate cortex (see Fig. 1.2–5), the ventromedial prefrontal cortex (see Fig. 1.2–7), and the insula (see Figs. 1.2–8 and 1.2–9), responds to the degree of salience of incoming cognitive, homeostatic, or emotional information (Table 1.2–2). The involvement of the insula in the salience network is understandable given its physical location as well as its function. The insula is primarily associated with interoception, the representation, and integration of the physiologic status of the organs and tissues of the body. However, recent studies indicate that the insula has more extensive functions, including bodily awareness, self-recognition, emotional awareness, and the integration of interoceptive information and emotions.

Perhaps the most intriguing network that has been described is the *default mode network*, which is formed by the medial prefrontal cortex (Fig. 1.2–5), the posterior cingulate cortex (Fig. 1.2–5), and the medial areas of the posterior parietal cortex (Fig. 1.2–5). The brain regions of the default mode network are active when a person is at rest and not engaged in an externally cued task, suggesting that this network is involved in internal modes of cognition, including recall of personal experiences, envisioning the future, and theory of mind (see Table 1.2–2).

Recent studies suggest that these three networks interact, with the salience network acting as a switch between the task-positive central executive network and the task-negative default mode network. The frontoinsular region of the salience network appears to play a key role in switching between the default mode and central executive networks. Interestingly, the human-specific von Economo class of neurons is found exclusively in the frontoinsular cortex, and it has been suggested that these neurons are responsible for the switching capability of the salience network.

Given that these networks are involved in the processing of cognitive and emotional information, it is not surprising that their dysfunction has been implicated in a variety of neuropsychiatric disorders. For example, individuals with depression have abnormally high functional connectivity within the default mode network and reduced connectivity among regions involved in the central executive network. In schizophrenia, the functional connectivity across the default mode, central executive, and salience networks is less organized. In particular, salience network dysfunction, perhaps driven by lower gray matter in the insula and/or anterior cingulate cortex, results in an imbalance between the default mode and central executive networks. This breakdown in network connectivity may lead to a disruption in the capability to distinguish between the external world and internal thoughts and feelings, manifested as psychotic symptoms such as delusions and hallucinations.

THE CLAUSTRUM. The *claustrum*, a telencephalic structure derived from the lateral pallium, consists of a thin strip of gray matter that is situated between the insular cortex and extreme capsule, the external capsule or the putamen (see Figs. 1.2–8 and 1.2–9). It extends caudally from the level of the very rostrum of the corpus callosum to the beginning of the pulvinar nucleus of the thalamus. Because of its small size and cloistered location, the claustrum has not been extensively studied. However, studies in nonhuman primates have shown that it has widespread connections across the brain. It is most extensively connected with the cerebral cortex, and these connections are topographic, with dedicated claustrum regions for specific cortical areas. In particular, sensory cortical regions are heavily interconnected with the claustrum, with regions of the claustrum identified by sensory modality (e.g., visual claustrum). This widespread connectivity has led some researchers to liken the claustrum to a salience detector, while others describe it as an area that binds together information from a variety of sources into a conscious experience. Interestingly, the claustrum contains the highest expression of serotonin 2A receptors in the brain. Activation of serotonin 2A receptors by psychedelic substances affects consciousness, cognition, and perception, providing a rationale for the current interest in the use of psychedelic substances as treatments for psychiatric disorders, although the location of the receptors that mediate the effects of psychedelic drugs remains an active research question.

Thalamus

The largest portion of the diencephalon consists of the *thalamus*, a group of nuclei located medial to the basal ganglia that serves as the major synaptic relay station for information reaching the cerebral cortex. Anatomically, the thalamic nuclei can be divided into six groups: anterior, medial, lateral, reticular, intralaminar, and midline nuclei (Fig. 1.2–22). A thin Y-shaped sheet of myelinated fibers, the *internal medullary lamina*, delimits the anterior, medial, and lateral groups of nuclei. In the human thalamus, the anterior and medial groups each contain a single large nucleus, the *anterior* and *medial dorsal nuclei*. The lateral group of nuclei can be subdivided further into dorsal and ventral tiers. The dorsal tier is composed of the *lateral dorsal*, the *lateral posterior*, and the *pulvinar nucleus*; the ventral tier consists of the *ventral anterior*, the *ventral lateral*, the *ventral posterior lateral*, and the *ventral posterior medial nuclei*. The lateral group of nuclei is covered by the *external medullary lamina*, another sheet of myelinated fibers. Interposed between these fibers and the internal capsule is a thin group of neurons forming the *reticular nucleus* of the thalamus. Also considered part of the lateral group, the lateral geniculate nucleus and the medial geniculate nucleus are located posterior to the ventral tier nuclei and inferior to the pulvinar nucleus, protruding posteriorly alongside the midbrain. The *intralaminar nuclei*, the largest of which is the *central median nucleus*, are located within the internal medullary lamina. The final group of thalamic nuclei, the *midline nuclei*, covers portions of the medial surface of the thalamus. The midline nuclei of each hemisphere may fuse to form the interthalamic adhesion, which is variably present.

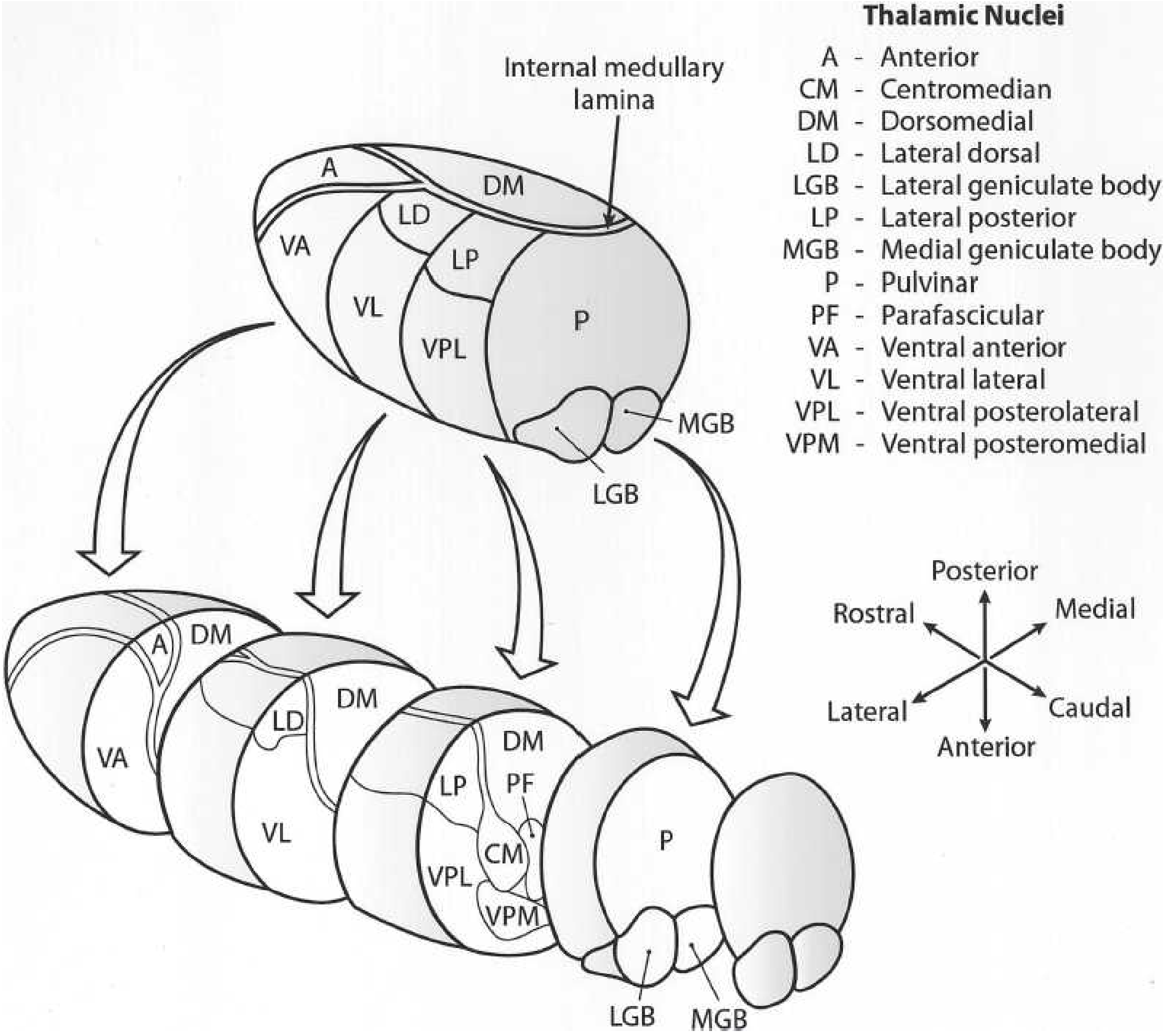
Thalamic Nuclei
A - Anterior
CM - Centromedian
DM - Dorsomedial
LD - Lateral dorsal
LGB - Lateral geniculate body
LP - Lateral posterior
MGB - Medial geniculate body
P - Pulvinar
PF - Parafascicular
VA - Ventral anterior
VL - Ventral lateral
VPL - Ventral posterolateral
VPM - Ventral posteromedial
Internal medullary lamina
A
DM
LD
VA
LP
VL
P
VPL
MGB
LGB
Posterior
Rostral
Medial
Lateral
Caudal
Anterior
A
DM
VA
LD
DM
VL
DM
LP
PF
CM
VPL
VPM
P
LGB
MGB

FIGURE 1.2–22. Expanded view of the dorsal thalamus illustrating the organization of thalamic nuclei. (Reprinted from Haines DE. *Fundamental Neuroscience for Basic and Clinical Applications*. 3rd ed. Churchill Livingstone; 2006:236.)

Thalamic nuclei also can be classified into several groups based on the pattern and information content of their connections (Table 1.2–3). *Relay nuclei* project to and receive input from specific regions of the cerebral cortex. These reciprocal connections apparently allow the cerebral cortex to modulate the thalamic input it receives. *Specific relay nuclei* process input either from a single sensory modality or from a distinct part of the motor system. For example, the lateral geniculate nucleus receives visual input from the optic tract and projects to the primary visual area of the occipital cortex. As summarized in Figure 1.2–23, neurons of the thalamic relay nuclei furnish topographically organized projections to specific regions of the cerebral cortex, although some cortical regions receive input from more than one nucleus.

In contrast, *association relay nuclei* receive highly processed input from more than one source and project to larger areas of the association cortex. For example, the medial dorsal thalamic nucleus receives input from the hypothalamus and the amygdala and is reciprocally interconnected with the prefrontal cortex and certain premotor and temporal cortical regions (Fig. 1.2–24). In contrast to relay nuclei, *diffuse-projection nuclei* receive input from diverse sources and project to widespread areas of the cerebral cortex and to the thalamus. The divergent nature of the cortical connections of these nuclei indicates that they may be involved in regulating the level of cortical excitability and arousal. Finally, the reticular nucleus is unique in that it contains inhibitory neurons that receive input from collaterals of the axons that reciprocally connect other thalamic nuclei and the cerebral cortex. Each portion of the reticular nucleus then projects to the thalamic nucleus from which it receives input. The pattern of connectivity indicates that the reticular nucleus samples cortical afferent and efferent activity and then uses that information to regulate thalamic function.

The classification of thalamic nuclei into different categories has been refined over the last 20 years to reflect our greater understanding of thalamic circuitry. Although classification into different nuclei is useful, the focus has shifted to the type of inputs to the various nuclei. For example, thalamic afferents can be defined as either drivers or modulators. Drivers convey information from the senses and periphery that are then relayed to the cerebral cortex. These inputs determine the nature of the thalamic relay. For example, the lateral geniculate nucleus is a visual relay because it transmits retinal input to the primary visual cortex. The driver inputs for association and diffuse-projection nuclei are more difficult to determine, but these inputs most likely arise from layer 5 pyramidal neurons in the cerebral cortex. Modulators include all other inputs to the thalamus, such as monoaminergic and cholinergic projections from brainstem nuclei and the feedback projections from layer 6 pyramidal neurons in the cerebral cortex. Thalamic circuits can then be distinguished based on the source of driver inputs. First-order relays have drivers that arise from ascending pathways and transmit information that has not yet been processed in the cortex. In contrast, higher-order relays have drivers that arise from layer 5 pyramidal neurons, allowing for the transmission of information from one cortical area to another via the thalamus. Thus, it is possible that a given thalamic nucleus contains both first-order and higher-order relays. Interestingly, over half of the circuits in the thalamic nuclei of primates are considered to be higher order, suggesting a complex role of the thalamus in corticothalamocortical function.

Functional Circuitry. The connections between the thalamus, the cortex, and certain related brain structures constitute three types of thalamocortical systems, each with different patterns of functional circuitry. These three systems—sensory, motor, and association systems—are described separately here but are heavily interconnected.

Table 1.2–3.
Connections of Thalamic Nuclei[a]

Specific Relay Nuclei

Nucleus	Afferents	Efferents	Function
Anterior	Mammillary body of hypothalamus Hippocampal formation	Cingulate cortex, parahippocampal gyrus	Expression of emotions Learning, memory
Ventral Anterior	Globus pallidus Substantia nigra	Premotor cortex Frontal eye fields	Control of limb movement Control of eye, face, and head movements
Ventral lateral	Dentate nucleus of cerebellum	Motor, premotor cortices	Control of head and limb movement
Ventral posterior lateral	Medial lemniscal and spinothalamic pathways	Somatosensory cortex	Processes pain, temperature, touch sensation, pressure and joint movement
Ventral posterior medial	Sensory nuclei of trigeminal nerve	Somatosensory cortex	Processes pain, temperature, touch sensation, pressure, and proprioception
Medial geniculate	Inferior colliculus	Auditory cortex	Hearing
Lateral geniculate	Optic tract	Visual cortex	Vision

Association Relay Nuclei

Nucleus	Afferents	Efferents	Function
Lateral dorsal	Pretectal area	Cingulate cortex, retrosplenial cortex, subiculum	Expression of emotions

Lateral posterior	Superior colliculus Pretectal area	Parietal cortex Precuneus	Sensory integration
Pulvinar	Superior colliculus Cerebellum	Temporal, parietal, occipital cortices	Integration of visual, auditory and somatosensory information
Medial dorsal	Amygdala, hypothalamus, brainstem	Prefrontal cortex	Expression of emotions, integration of sensory information
Diffuse-Projection Nuclei			
Nucleus	**Afferents**	**Efferents**	**Function**
Midline	Reticular formation, hypothalamus, striatum, hippocampus	Striatum, hippocampus, limbic cortex	Modulation of cortical excitability
Intralaminar	Reticular formation, spinothalamic tract, globus pallidus	Basal ganglia, cerebral cortex	Control of cortical activity
Reticular	Cerebral cortex Other thalamic nuclei	Other thalamic nuclei	Integration and control of thalamic activity

[a] This table does not include the cortical inputs to each thalamic nucleus.

Adapted with permission from Patestas MA, Gartner LP. *A Textbook of Neuroanatomy*. 2nd ed. Wiley 2016. Copyright © 2016 by John Wiley & Sons, Inc.

THALAMOCORTICAL SENSORY SYSTEMS. Several general principles govern the organization of the thalamocortical sensory systems. First, sensory receptors transduce certain stimuli in the external environment to neural impulses. The impulses ascend, often through intermediate nuclei in the spinal cord and the medulla, and ultimately synapse in specific relay nuclei of the thalamus.

Second, projections from peripheral sensory receptors to the thalamus and the cortex exhibit topography. That is, a particular portion of the external world is mapped onto a particular region of the brain. For example, in the somatosensory system, axons carrying information regarding a distinct part of the body synapse in a discrete part of the ventral posterior nucleus of the thalamus. Specifically, the ventral posterior medial nucleus receives inputs regarding the head, and the ventral posterior lateral nucleus receives inputs regarding the remainder of the body. The nuclei project topographically to the primary somatosensory cortex, where several representations of the contralateral half of the body can be found. These representations are distorted; regions heavily innervated by sensory receptors, such as the fingers, are disproportionately represented in the primary somatosensory cortex.

Third, in some cases, sensory inputs travel to the thalamus in a segregated manner according to the submodality of the information conveyed. The inputs are processed in a parallel fashion; particular pathways may be devoted exclusively to processing a submodality. For example, in the somatosensory system (Fig. 1.2–25), most fibers carrying tactile and proprioceptive information travel in the medial lemniscus, whereas fibers carrying pain and temperature information travel in the spinothalamic tract; both pathways terminate in the ventral posterior thalamic nuclei. Although some tactile information is carried in the spinothalamic tract, the submodalities of pain and temperature are largely segregated from tactile and proprioceptive inputs as they ascend to the thalamus.

Finally, sensory pathways exhibit convergence. That is, primary sensory areas process sensory information and project to unimodal association areas. Subsequently, the unimodal areas project to and converge in multimodal associational areas. Convergence in sensory pathways is illustrated in the somatosensory system. The primary somatosensory cortex, located in the anterior parietal lobe, has been divided into four regions on the basis of cytoarchitecture. Each of the cytoarchitectonic regions—numbered 1, 2, 3a, and 3b by Brodmann—contains a topographical representation of the body. The regions are heavily interconnected, and all project to the next level of somatosensory processing in area S-II (Brodmann areas 40 and 43). This type of projection, from one level of processing to a more advanced level, is termed a *feedforward* projection. The reciprocal connection, from the more advanced processing level back to the simpler level, is called a *feedback* projection. Both projections have distinct patterns of laminar termination: feedforward projections originate in cortical layers 2 and 3 and terminate in layer 4; feedback projections originate primarily in layers 5 and 6 and terminate outside layer 4. Further processing of somatosensory information occurs in higher-order somatosensory areas, such as area 7b of the posterior parietal cortex, which receives feedforward projections from S-II. Lesions of the posterior parietal cortex reflect the complexity of the information processed there; after a person has sustained a posterior parietal lesion, the ability to understand the significance of sensory stimuli is impaired, and extreme cases result in contralateral sensory neglect and inattention.

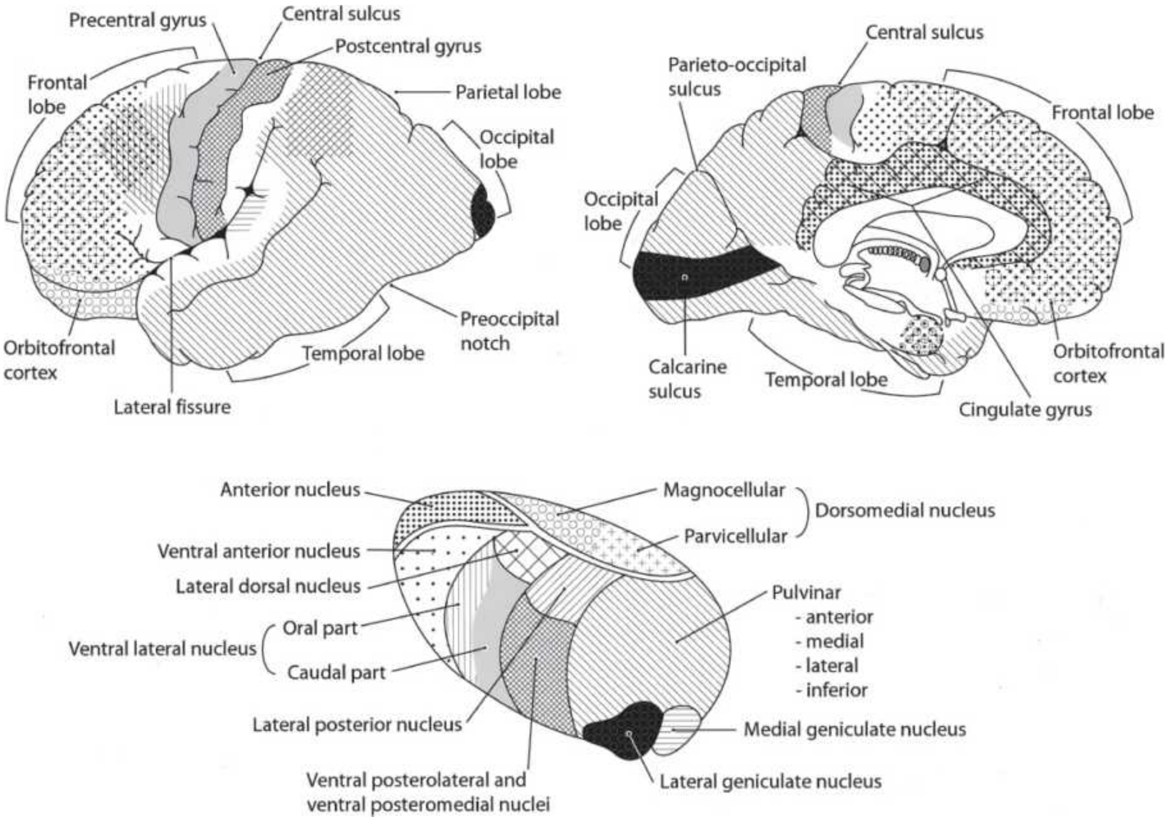
Precentral gyrus
Central sulcus
Postcentral gyrus
Frontal lobe
Parietal lobe
Occipital lobe
Orbitofrontal cortex
Temporal lobe
Preoccipital notch
Lateral fissure
Central sulcus
Parieto-occipital sulcus
Frontal lobe
Occipital lobe
Calcarine sulcus
Temporal lobe
Orbitofrontal cortex
Cingulate gyrus
Anterior nucleus
Magnocellular
Parvicellular
Dorsomedial nucleus
Ventral anterior nucleus
Lateral dorsal nucleus
Ventral lateral nucleus
Oral part
Caudal part
Pulvinar
- anterior
- medial
- lateral
- inferior
Lateral posterior nucleus
Medial geniculate nucleus
Ventral posterolateral and ventral posteromedial nuclei
Lateral geniculate nucleus

FIGURE 1.2–23. Schematic drawings of the lateral **(upper left)** and medial **(upper right)** surfaces of the right cerebral hemisphere and the right thalamus **(lower)**. Each thalamic nucleus is patterned coded to match its target area in the cerebral cortex. (Adapted from Haines DE. *Fundamental Neuroscience for Basic and Clinical Applications*. 3rd ed. Churchill Livingstone; 2006:237.)

THALAMOCORTICAL MOTOR SYSTEMS. The thalamocortical motor systems exhibit some unique organizational principles but also share many of the features present in the sensory systems. First, in contrast to sensory systems, which primarily ascend from sensory receptors to cortical association areas, motor systems descend from association and motor regions of the cortex to the brainstem and the spinal cord. The corticospinal tract originates in the large Betz cells of layer 5 in the premotor and primary motor cortices (see Fig. 1.2–19B) of the frontal lobe and terminates in the spinal cord to influence motor behavior.

Second, motor systems exhibit strong topography at the thalamic and cortical levels. The corticospinal tract is organized so that a topographic representation of the contralateral half of the body is evident in the primary motor and premotor cortices. The representation of the body is disproportionate, with large regions of the motor cortex devoted to areas of the body involved in fine movement, such as the face and the hands.

Finally, projections from several sensory association regions converge in the motor regions of the frontal cortex. The premotor cortex receives a convergence of afferents from higher-order somatosensory and visual areas of the posterior parietal cortex, whereas afferents from the primary somatosensory cortex converge on the primary motor cortex. In addition to cortical input, the primary motor cortex receives afferents from the ventral lateral nucleus of the thalamus; this nucleus receives afferents predominantly from the cerebellum. The premotor cortex receives input from the ventral anterior thalamic nucleus, which receives much of its input from the globus pallidus.

THALAMOCORTICAL ASSOCIATION SYSTEMS. The multimodal association areas of the cortex are organized according to several general principles. First, association regions receive a convergence of input from a variety of sources, including unimodal and multimodal association regions of the cortex, association nuclei of the thalamus, and other structures. The prefrontal cortex receives afferents from higher-order sensory cortices of the parietal and temporal lobes, the contralateral prefrontal cortex, the cingulate cortex of the limbic system, the medial dorsal thalamic nucleus (an association relay nucleus), and portions of the amygdala. The medial dorsal thalamic nucleus receives highly processed inputs from many sources, including some regions, such as the amygdala, hippocampus, and hypothalamus, that project directly to the prefrontal cortex. These redundant (direct and indirect) projections may serve to attach additional significance to certain inputs received by the prefrontal cortex. The significance of these inputs may also be influenced by their temporal and spatial coincidence with modulatory inputs from brainstem nuclei that use the monoamine neurotransmitters dopamine, norepinephrine, or serotonin. These monoamine systems project broadly to the cerebral cortex, although with substantial regional differences in density (Fig. 1.2–26).

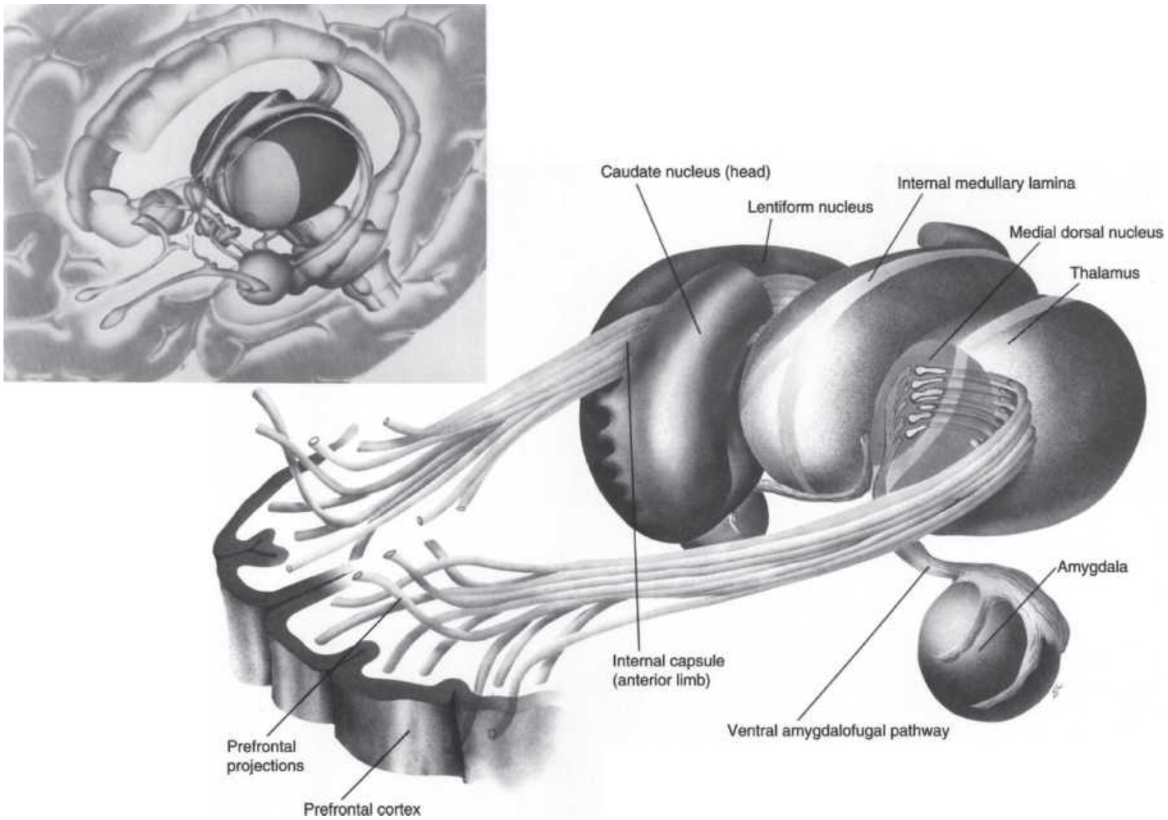
Caudate nucleus (head)
Internal medullary lamina
Lentiform nucleus
Medial dorsal nucleus
Thalamus
Amygdala
Internal capsule (anterior limb)
Ventral amygdalofugal pathway
Prefrontal projections
Prefrontal cortex

FIGURE 1.2–24. Drawing of the thalamus showing the pathway of projections from the mediodorsal thalamic nucleus through lateral thalamic nuclei to the prefrontal cortex. Also shown are afferents from the amygdala to the medial dorsal nucleus. The *inset* shows the thalamus embedded in the limbic system of which it is a key component. (Adapted from Hendelman WJ. *Student's Atlas of Neuroanatomy*. WB Saunders; 1994:199. Copyright © 1994 Elsevier.)

The second way that the projections are organized is according to topography. The projections that terminate in multimodal association regions exhibit a topographic organization. Different cytoarchitectonic regions of the medial dorsal thalamic nucleus project to discrete regions of the prefrontal cortex. In addition, some cortical afferents received by the prefrontal cortex are topographically organized; certain regions of the prefrontal cortex predominantly receive highly processed information from one sensory modality.

The patterns of connectivity are clearly related to some of the functional characteristics attributed to the prefrontal cortex. For example, in monkeys, lesions of the DLPFC consistently produce impairments in a monkey’s ability to perform spatial working memory tasks. These tasks require that monkeys maintain a spatial representation of the location of an object during a delay period in which the object is out of sight; the DLPFC appears to play a key role in maintaining the spatial representation of the object. This property requires the DLPFC to receive information regarding the location of objects in space, and in fact, such information is conveyed to the DLPFC via projections from association regions of the parietal cortex. Although the DLPFC is necessary for the performance of working memory tasks in monkeys, it is not sufficient for the performance of such tasks. For example, lesions of the medial dorsal thalamic nucleus in monkeys result in similar impairments in the performance of spatial working memory tasks.

Knowledge of the integration of afferent inputs into the neural circuitry of certain prefrontal regions may also be important for understanding the nature of prefrontal cortical dysfunction in schizophrenia. Individuals with schizophrenia perform poorly on tasks that are known to be mediated by the prefrontal cortex. These findings have been correlated with other measures to suggest, albeit indirectly, that the dopamine projections to the DLPFC are impaired in schizophrenia. Studies in nonhuman primates have shown that performance of delayed-response tasks that require working memory, the same type of behaviors that are impaired in subjects with schizophrenia, requires an appropriate level of dopamine input to the DLPFC.

During visuospatial delayed-response tasks, the DLPFC guides decisions about how to sample and process complex visual information, and it does this through circuitry involving the frontal eye fields, the mediodorsal thalamic nucleus, and the superior colliculus. The frontal eye fields are located in the frontal cortex, situated between the more anterior prefrontal cortex and the more caudal motor cortex, and control saccadic eye movements, which are an integral component of visuospatial delayed-response activity. Interestingly, the mediodorsal thalamic nucleus has recently been described as containing both first-order and higher-order relays, both of which appear to be involved in visuospatial delayed-response tasks. First-order brainstem afferents from the superior colliculus send eye movement–related information to the mediodorsal thalamic nucleus, which then transmits this information to the frontal eye fields. In addition to direct corticocortical connections between the DLPFC and the frontal eye fields, information from the frontal eye fields concerning the executed saccadic eye movement is transmitted to the prefrontal cortex via higher-order relays in the mediodorsal thalamic nucleus. This transthalamic circuitry suggests that the mediodorsal thalamic nucleus plays a significant role in contributing to the control of behavior executed by the DLPFC.

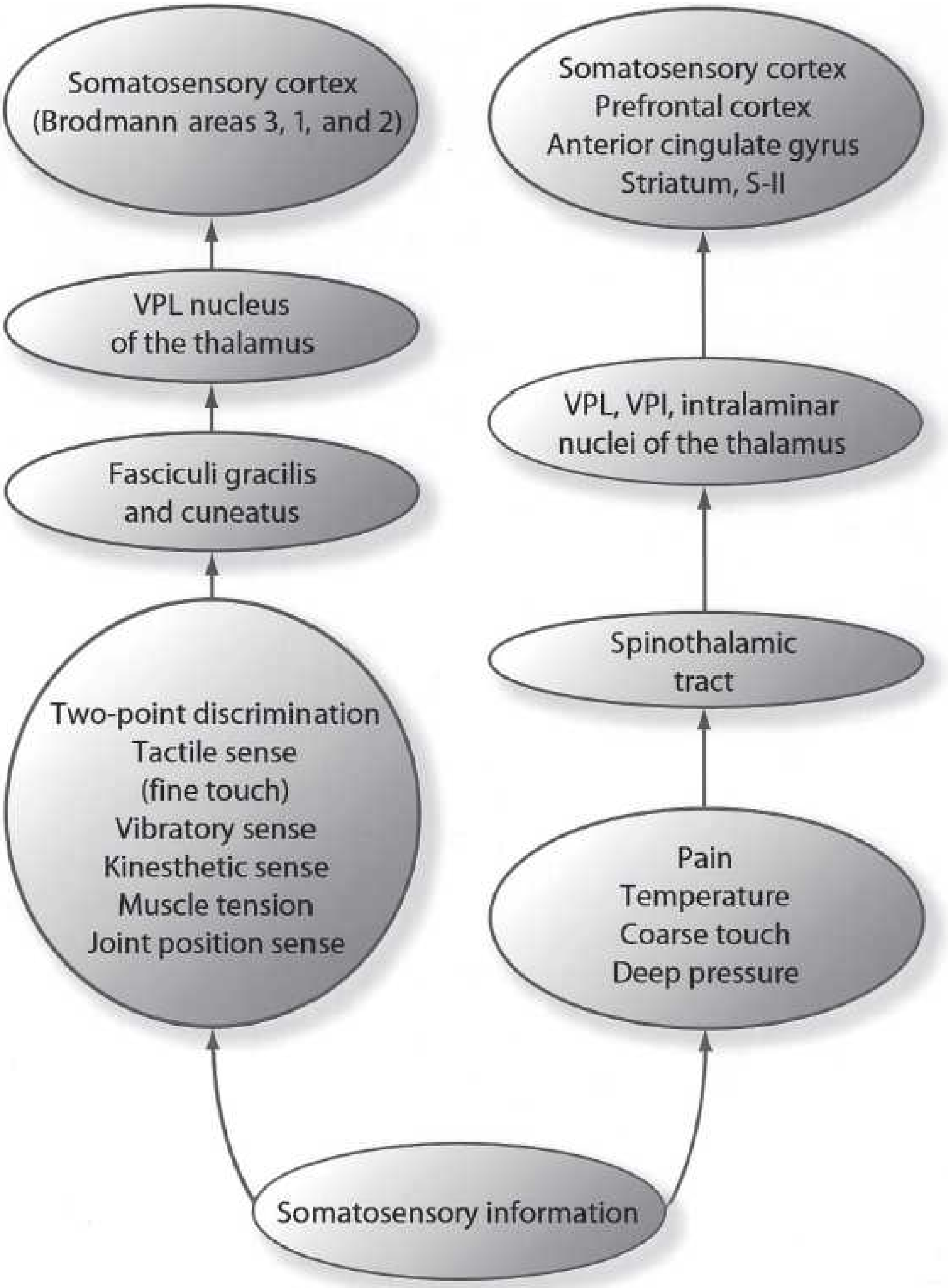
Somatosensory cortex
(Brodmann areas 3, 1, and 2)
Somatosensory cortex
Prefrontal cortex
Anterior cingulate gyrus
Striatum, S-II
VPL nucleus
of the thalamus
VPL, VPI, intralaminar
nuclei of the thalamus
Fasciculi gracilis
and cuneatus
Spinothalamic
tract
Two-point discrimination
Tactile sense
(fine touch)
Vibratory sense
Kinesthetic sense
Muscle tension
Joint position sense
Pain
Temperature
Coarse touch
Deep pressure
Somatosensory information

FIGURE 1.2–25. Pathway of somatosensory information processing. (Adapted from Patestas MA, Gartner LP. A *Textbook of Neuroanatomy*. 2nd ed. John Wiley & Sons; 2016:202.)

CEREBELLOTHALAMOCORTICAL SYSTEMS. The cerebellum traditionally has been considered to be involved solely with motor control, regulating posture, gait, and voluntary movements. More recent studies indicate, however, that the cerebellum may also play an important role in the mediation of certain cognitive abilities through inputs to portions of the thalamus that project to association regions of the cerebral cortex.

The cerebellum is located in the posterior cranial fossa, inferior to the occipital lobes (see Figs. 1.2–5 and 1.2–11). The external surface of the cerebellum, the cerebellar cortex, is composed of small folds, termed *folia*, separated by sulci. Viewed from the dorsal surface, the cerebellum contains a raised central portion, called the *vermis*, and lateral portions, called the *cerebellar hemispheres* (see Fig. 1.2–11). Located within the cerebellum are the deep cerebellar nuclei, which are arranged as follows: the fastigial nucleus is located next to the midline; the globose and emboliform nuclei are slightly more lateral; and the largest nucleus, the dentate, occupies the most lateral position (see Fig. 1.2–11). Generally, the cerebellar cortex processes the inputs to the cerebellum, and the deep nuclei process the outputs. Although many portions of the cerebellum are interconnected with brain regions that regulate motor actions, the circuitry of the cerebellum involved in cognitive functions may be of greatest interest from the standpoint of psychiatric illness. For example, the lateral cerebellar cortex and the dentate nucleus are markedly expanded in the primate brain. It has been suggested that these changes are associated with an increase in the size of cortical areas (especially the prefrontal regions) influenced by cerebellar output and an expanded role of the cerebellum in cognitive functions. More recent studies in nonhuman primates have shown that the DLPFC receives inputs from two ipsilateral thalamic nuclei (medial dorsal and ventral lateral), which receive inputs from the contralateral cerebellar dentate nucleus. The cells of the dentate nucleus involved in these connections are distinct from the cells that influence the motor and premotor regions of the cerebral cortex. Interestingly, functional imaging studies in individuals with schizophrenia have revealed abnormal patterns of activation in the cerebellum, thalamus, and prefrontal cortex, suggesting that dysfunction of this circuitry might be associated with the disturbances in cognitive processes exhibited by these individuals.

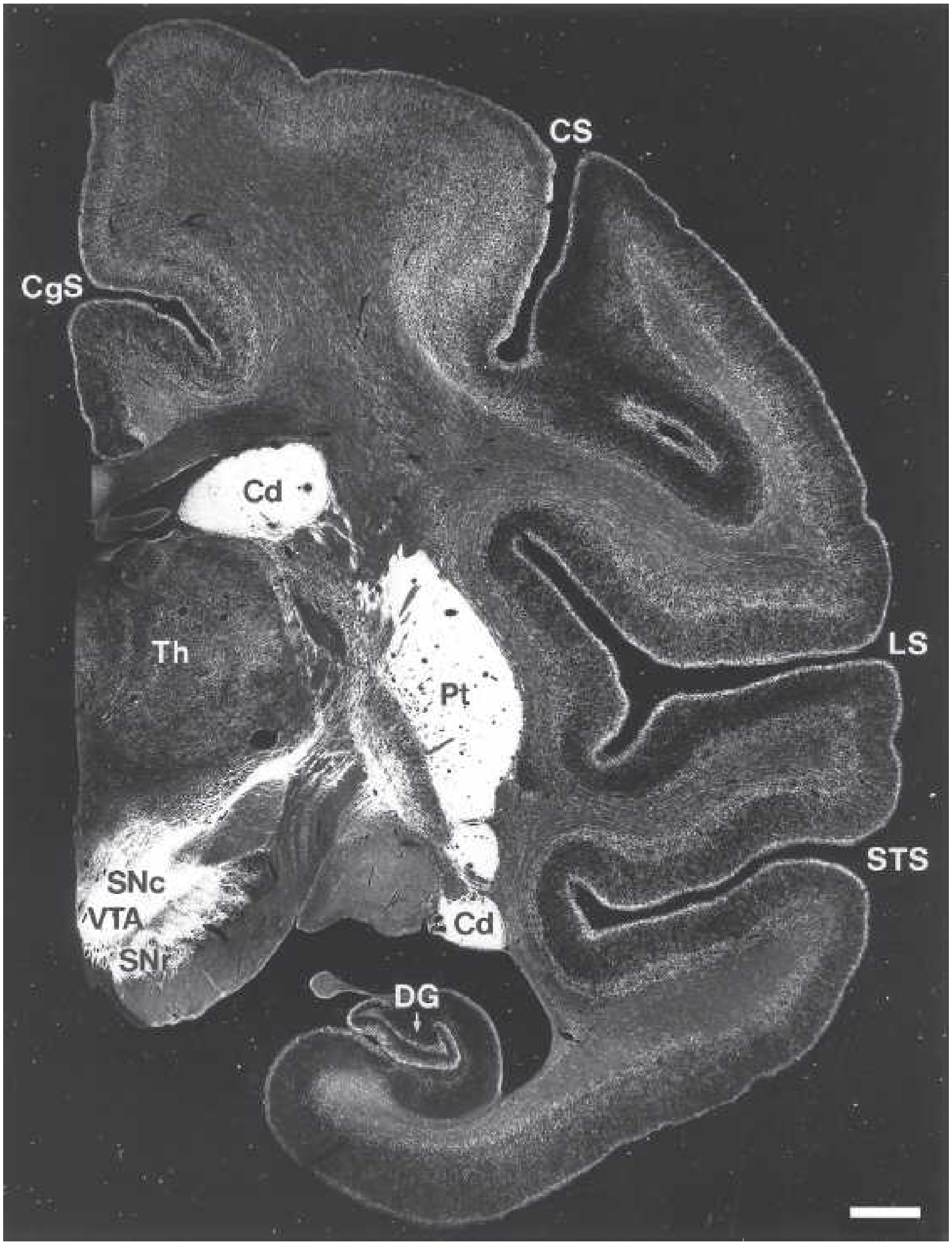

FIGURE 1.2–26. Darkfield photomicrograph of a coronal section through a hemisphere of a macaque monkey immunolabeled for the dopamine transporter. This image illustrates the differential distribution of dopamine-containing axons in different regions of the brain. The brighter the image, the greater the quantity of dopamine-containing structures. Dopamine-rich areas such as the caudate (*Cd*), putamen (*Pt*), ventral tegmental area (*VTA*), and the substantia nigra (*SNc* and *SNr*) appear white, whereas dopamine innervation of the cortex and thalamus, although clearly seen, is less dense and varies by the specific cortical and thalamic region. CgS, cingulate sulcus;

CS, central sulcus; DG, dentate gyrus; LS, lateral sulcus; STS, superior temporal sulcus; Th, thalamus. Calibration bar = 2 mm. (From Lewis DA, Melchitzky DS, Sesack SR, Whitehead RE, Auh S, Sampson A. Dopamine transporter immunoreactivity in monkey cerebral cortex: Regional, laminar and ultrastructural localization. *J Comp Neurol*. 2001;432(1):119. Used with permission.)

Basal Ganglia System

The basal ganglia are a collection of nuclei that have been grouped together on the basis of their interconnections. These nuclei play an important role in regulating movement and in certain disorders of movement (dyskinesias), which include jerky movements (chorea), writhing movements (athetosis), and rhythmic movements (tremors). In addition, more recent studies have shown that certain components of the basal ganglia play an important role in many cognitive functions.

Major Structures. The basal ganglia are generally considered to include the *caudate nucleus*, the *putamen*, the *globus pallidus* (referred to as the *pallidum*), the *subthalamic nucleus*, and the *substantia nigra* (Fig. 1.2–27). The term *striatum* refers to the caudate nucleus and the putamen together; the term *corpus striatum* refers to the caudate nucleus, the putamen, and the globus pallidus; and the term *lentiform nucleus* refers to the putamen and the globus pallidus together.

Although these nuclei are generally agreed to comprise the basal ganglia, some controversy exists concerning whether other nuclei should be included in the definition of the basal ganglia. Some investigators believe that additional regions of the brain have anatomic connections that are similar to other components of the basal ganglia and should, therefore, be included in the term. These additional regions are usually termed the *ventral striatum* and the *ventral pallidum*. The ventral striatum includes the nucleus accumbens (Fig. 1.2–28), which is the region where the putamen and the head of the caudate nucleus fuse, and the olfactory tubercle. The ventral pallidum is a region that receives afferents from the ventral striatum and includes, but is not limited to, a group of neurons termed the *substantia innominata* (see Fig. 1.2–8). This section focuses on the structures generally accepted as belonging to the basal ganglia but also discusses additional structures when relevant to the functional anatomy of the system.

CAUDATE NUCLEUS. The caudate nucleus is a C-shaped structure that is divided into three general regions. The anterior portion of the structure is referred to as the *head*, the posterior region is the *tail*, and the intervening region is the *body* (see Fig. 1.2–27). The caudate nucleus is associated with the contour of the lateral ventricles: the head lies against the frontal horn of the lateral ventricle, and the tail lies against the temporal horn (see Figs. 1.2–8 to 1.2–10). The head of the caudate nucleus is continuous with the putamen; the tail terminates in the amygdala of the temporal lobe.

PUTAMEN. The putamen lies in the brain medial to the insula and is bounded laterally by the fibers of the external capsule and medially by the globus pallidus (see Figs. 1.2–8 and 1.2–9). As noted earlier, the putamen is continuous with the head of the caudate nucleus (see Fig. 1.2–27). Although bridges of neurons between the caudate nucleus and the putamen show the continuity of the nuclei, the two structures are separated by fibers of the anterior limb of the internal capsule (see Fig. 1.2–28).

GLOBUS PALLIDUS. In contrast to the caudate nucleus and the putamen, which are telencephalic in origin, the globus pallidus is derived from the diencephalon. The globus pallidus constitutes the inner component of the lentiform nucleus (see Fig. 1.2–27, bottom panel); with the putamen, it forms a cone-like structure, with its tip directed medially (see Figs. 1.2–8 and 1.2–9). The posterior limb of the internal capsule bounds the globus pallidus medially and separates it from the thalamus; the putamen borders the globus pallidus laterally. In humans, the medial medullary lamina divides the globus pallidus into external (lateral) and internal (medial) segments (see Fig. 1.2–28).

SUBTHALAMIC NUCLEUS. The subthalamic nucleus (of Luys) is also derived from the diencephalon. The large-celled nucleus lies dorsomedial to the posterior limb of the internal capsule and dorsal to the substantia nigra (see Figs. 1.2–9 and 1.2–27). Discrete lesions of the subthalamic nucleus in humans lead to hemiballism, a syndrome characterized by violent, forceful choreiform movements that occur on the side contralateral to the lesion.

SUBSTANTIA NIGRA. The substantia nigra is present in the midbrain between the tegmentum and the basis pedunculi and is mesencephalic in origin (see Fig. 1.2–9). The substantia nigra consists of two components: a dorsal cell-rich portion referred to as the *pars compacta* and a ventral cell–sparse portion denoted the *pars reticulata*. Most of the neurons in the pars compacta of the substantia nigra in humans are pigmented because of the presence of neuromelanin; these cells contain the neurotransmitter dopamine (see Fig. 1.2–26). The dendrites of the pars compacta neurons frequently extend into the pars reticulata, where they receive synapses from the neurons of the pars reticulata that use the inhibitory neurotransmitter GABA.

In rodents, the dopamine-containing neurons of the substantia nigra (A9 region) have been distinguished from the neurons located in the ventral tegmental area (A10 region) (Fig. 1.2–29) and the retrorubral field (A8 region), but more recent studies in monkeys and humans suggest that dopamine neurons can be more meaningfully organized at a functional level into dorsal and ventral tiers (Fig. 1.2–29). The dorsal tier is formed by a medially–laterally oriented band of neurons that includes the dopamine-containing cells that are (1) located in the medial ventral mesencephalon, (2) scattered dorsal to the dense cell clusters in the substantia nigra, and (3) distributed lateral and caudal to the red nucleus. The ventral tier comprises the dopamine neurons that are densely packed in the substantia nigra and the cell columns that penetrate into the substantia nigra pars reticulata. Dorsal tier dopamine neurons express relatively low levels of mRNA for the dopamine transporter and the dopamine type 2 receptor (D_2), contain the calcium-binding protein calbindin, and send axonal projections to the regions of the striatum that are dominated by input from limbic-related structures and association regions of the cerebral cortex. In contrast, ventral-tier neurons contain high levels of mRNA for the dopamine transporter and the D_2 dopamine receptor, typically lack calbindin, and send axonal projections to the sensorimotor regions of the striatum. Each of these features may contribute to the greater vulnerability of ventral tier neurons to the pathology of Parkinson disease, whereas dorsal tier neurons may be more likely to be involved in the pathophysiology of schizophrenia.

Internal Organization. The caudate nucleus and the putamen are frequently referred to together as the striatum because of their common characteristics. In rodents, these nuclei are a continuous structure, and, in all mammals, they are composed of histologically identical cells. Most neurons in the striatum are medium-sized cells (10 to 20 μm in diameter) that possess spines on their dendrites; these so-called *medium spiny neurons* are known to send their axons out of the striatum. In addition to medium spiny neurons, medium-sized cells without spines (medium aspiny neurons) are present, as are large cells with and without spines (large spiny neurons and large aspiny neurons). With the exception of the medium and large spiny cells, most other striatal neurons are local circuit neurons.

Immunohistochemical and receptor-binding studies have shown a discontinuity in the distribution of certain neurotransmitter-related substances that form the functional circuitry of the basal ganglia. For example, in the striatum, zones that contain a low density of acetylcholinesterase (AChE) enzymatic activity are surrounded by regions rich in AChE activity. The AChE-rich regions are referred to as the *matrix*, and the AChE-poor zones are termed either *striosomes* in primates or *patches* in rodents. Several neuropeptide systems also follow this organization. For example, the distributions of enkephalin, substance P, and somatostatin immunoreactivity are organized in a similar manner as the AChE-rich and the AChE-poor areas in the striatum. In addition, in rodents, certain subtypes of dopamine receptors are present predominantly in one compartment. Furthermore, the distribution of some afferent systems terminating in the striatum follows the striosome-matrix organization. For example, afferents from the thalamus terminate preferentially in the matrix rather than in the striosome.

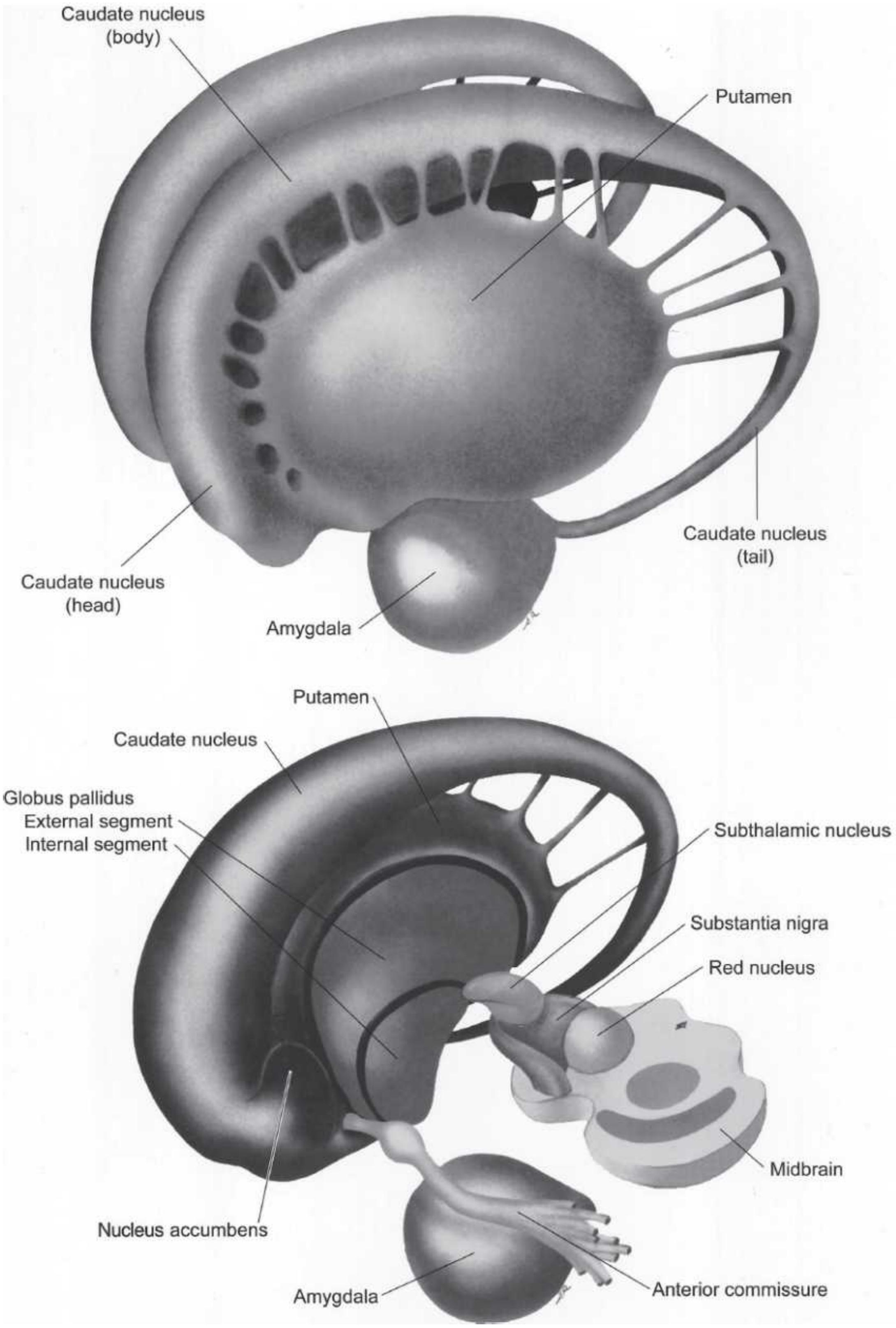
Caudate nucleus (body)
Putamen
Caudate nucleus (tail)
Caudate nucleus (head)
Amygdala
Putamen
Caudate nucleus
Globus pallidus
External segment
Internal segment
Subthalamic nucleus
Substantia nigra
Red nucleus
Midbrain
Nucleus accumbens
Amygdala
Anterior commissure

FIGURE 1.2–27. Schematic drawing of the isolated basal ganglia as seen from the dorsolateral perspective so that the caudate nucleus is apparent bilaterally. In the **bottom panel**, the basal ganglia from the left hemisphere

has been removed, exposing the medial surface of the right putamen and globus pallidus and the subthalamic nucleus and substantia nigra. (Adapted from Hendelman WJ. *Student's Atlas of Neuroanatomy*. WB Saunders; 1994:37, 41. Copyright © 1994 Elsevier.)

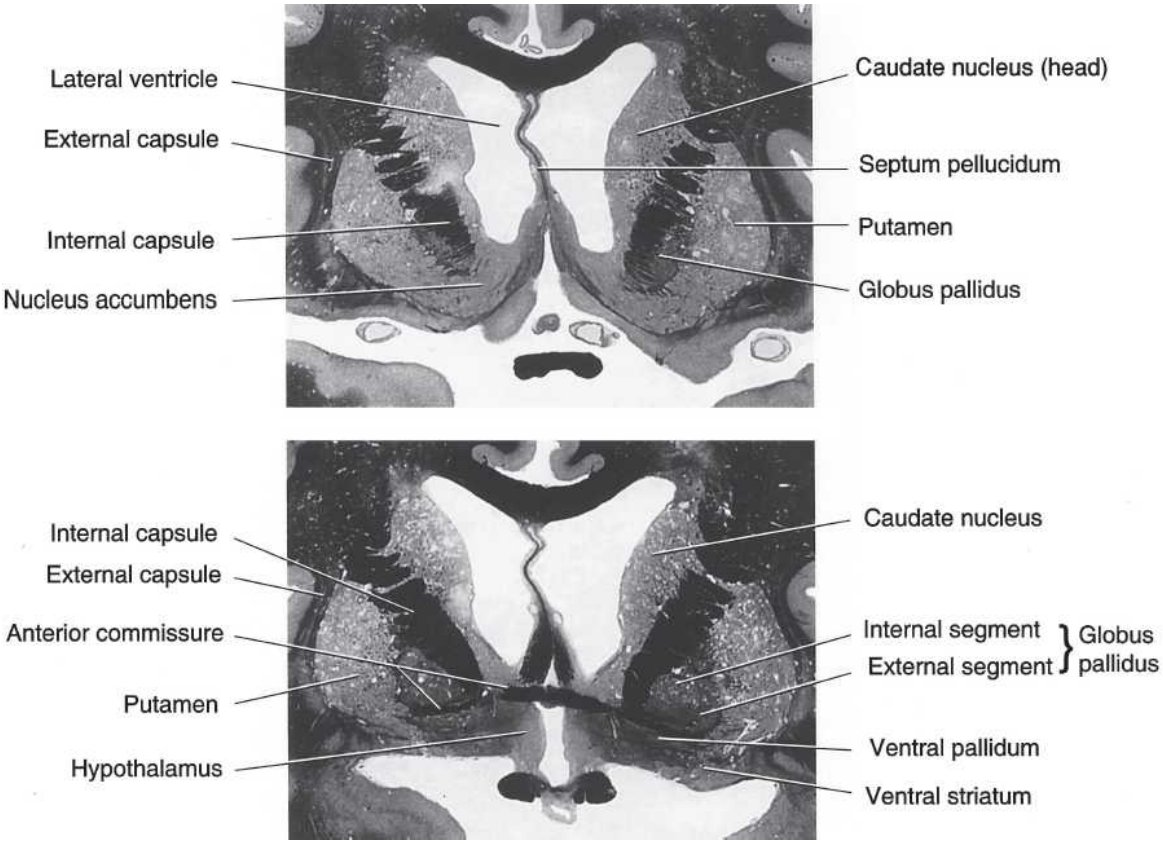
Lateral ventricle
External capsule
Internal capsule
Nucleus accumbens
Caudate nucleus (head)
Septum pellucidum
Putamen
Globus pallidus
Internal capsule
External capsule
Anterior commissure
Putamen
Hypothalamus
Caudate nucleus
Internal segment
External segment
Globus pallidus
Ventral pallidum
Ventral striatum

FIGURE 1.2–28. Photographs of cross-sections of human brain showing basal ganglia nuclei and related structures. (Reprinted from Haines DE. *Fundamental Neuroscience for Basic and Clinical Applications*. 3rd ed. Churchill Livingstone; 2006:416.)

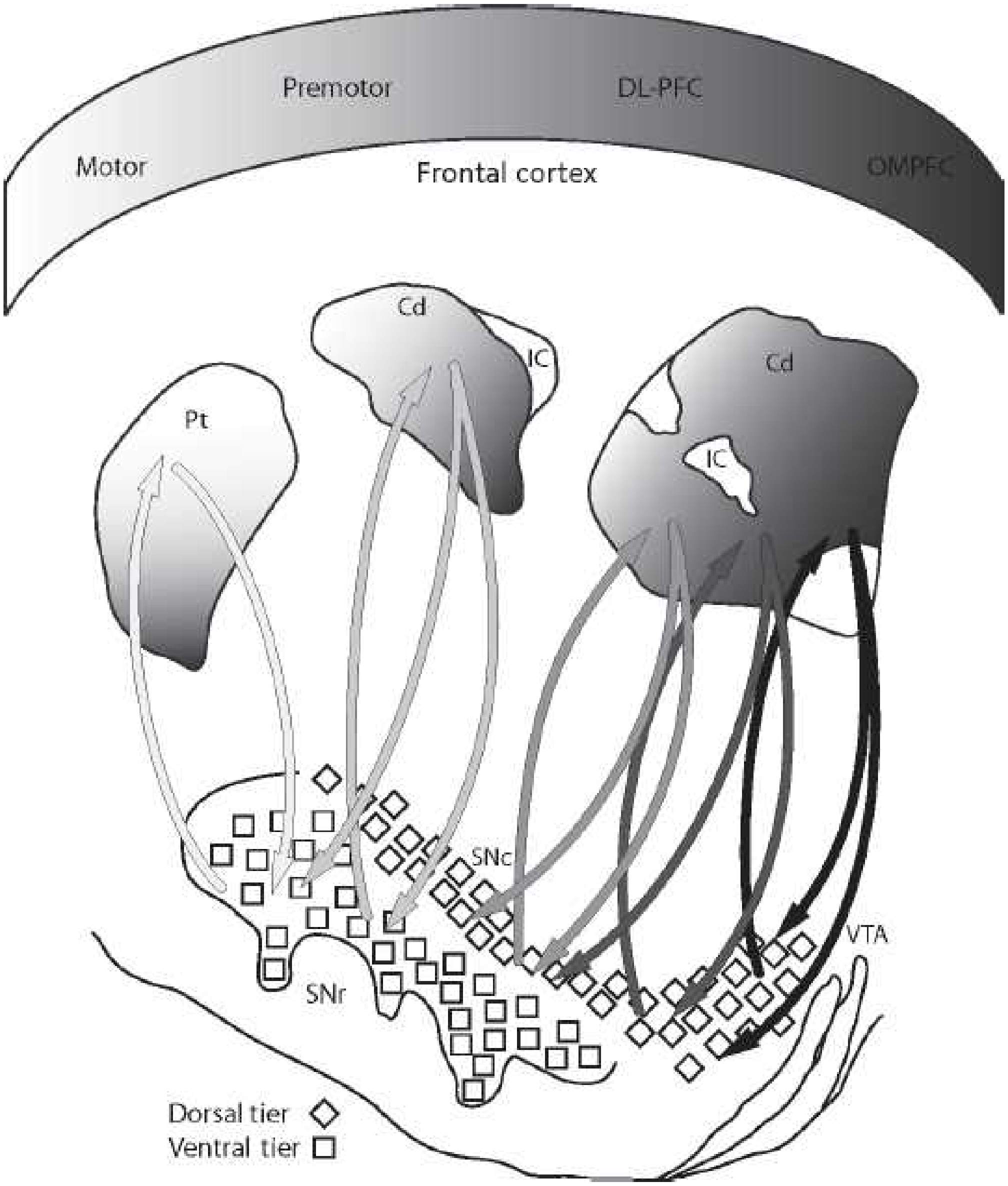
Premotor
DL-PFC
Motor
Frontal cortex
OMPFC
Cd
IC
Pt
Cd
IC
SNc
VTA
SNr
Dorsal tier ◇
Ventral tier □

FIGURE 1.2–29. Diagram of the organization of the striatonigrostriatal and corticostriatal projections in monkeys. DLPFC, dorsolateral prefrontal cortex; IC, internal capsule; OMPFC, orbital and medial prefrontal cortex; S, shell; SNC, substantia nigra pars compacta; SNr, substantia nigra pars reticulata; VTA, ventral tegmental area. (Adapted from Haber SN, Fudge JH, McFarland NR. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J Neurosci* 2000;20:2369. doi: 10.1523/JNEUROSCI.20-06-02369.2000. Copyright © 2000 Society for Neuroscience.)

Functional Circuitry. Projections into, within, and out of the basal ganglia are topographically organized and maintain this topography throughout the processing circuits of the basal ganglia. The existence of such patterns of connectivity has resulted in the hypothesis that parallel independent circuits in the basal ganglia process information from different regions of the brain and mediate separate complex functions. For example, and as illustrated in Figure 1.2–29, there is an inverse dorsal–ventral topographic organization to the projection from the dorsal and ventral dopamine neurons to the striatum. Dorsally and medially located dopamine neurons project to the ventral and medial parts of the striatum, whereas ventrally and laterally located dopamine neurons project to dorsal and lateral parts of the striatum. Another prominent input to the striatum comes from the cerebral cortex, and this projection has a topographic organization related to that of the striatonigrostriatal pathway. Orbital and medial prefrontal cortices project to the ventral striatum, the DLPFC projects to the central striatum, and premotor and motor cortices project to the dorsolateral striatum. These topographies create limbic, associative, and motor pathways within the corticostriatal and striatonigrostriatal projections.

INPUTS TO THE BASAL GANGLIA. The striatum is the major recipient of the inputs to the basal ganglia. Three major afferent systems are known to terminate in the striatum: the corticostriatal, nigrostriatal, and thalamostriatal afferents (Fig. 1.2–30). The corticostriatal projection originates from all regions of the neocortex, arising primarily from the pyramidal cells of layers 5 and 6, which use the excitatory neurotransmitter glutamate. A topography governing corticostriatal projections has been found in monkeys. Afferents from the sensorimotor cortex terminate predominantly in the putamen; association regions of the cortex terminate preferentially in the caudate nucleus. The prefrontal regions, in particular, provide a heavy input to the head of the caudate nucleus. In addition, afferents from the limbic cortical areas, the hippocampus, and the amygdala terminate in the ventral striatum. The second major class of afferents uses the neurotransmitter dopamine. In Figure 1.2–30, these projections are shown arising from the substantia nigra pars compacta, but, as noted earlier (see Fig. 1.2–29), different portions of the striatum receive input from the dorsal-tier or ventral-tier dopamine-containing neurons of the ventral mesencephalon. Electron microscopy studies have shown that many of the synapses formed by dopamine axon terminals on medium spiny neuron dendrites are immediately adjacent to the synapses provided by corticostriatal axons, suggesting that dopamine may play an important role in modulating the excitatory influence of cortical projections on striatal neurons. The third afferent system originates in the thalamus. The thalamic nuclei providing the projections are the intralaminar nuclei, particularly the central median nucleus.

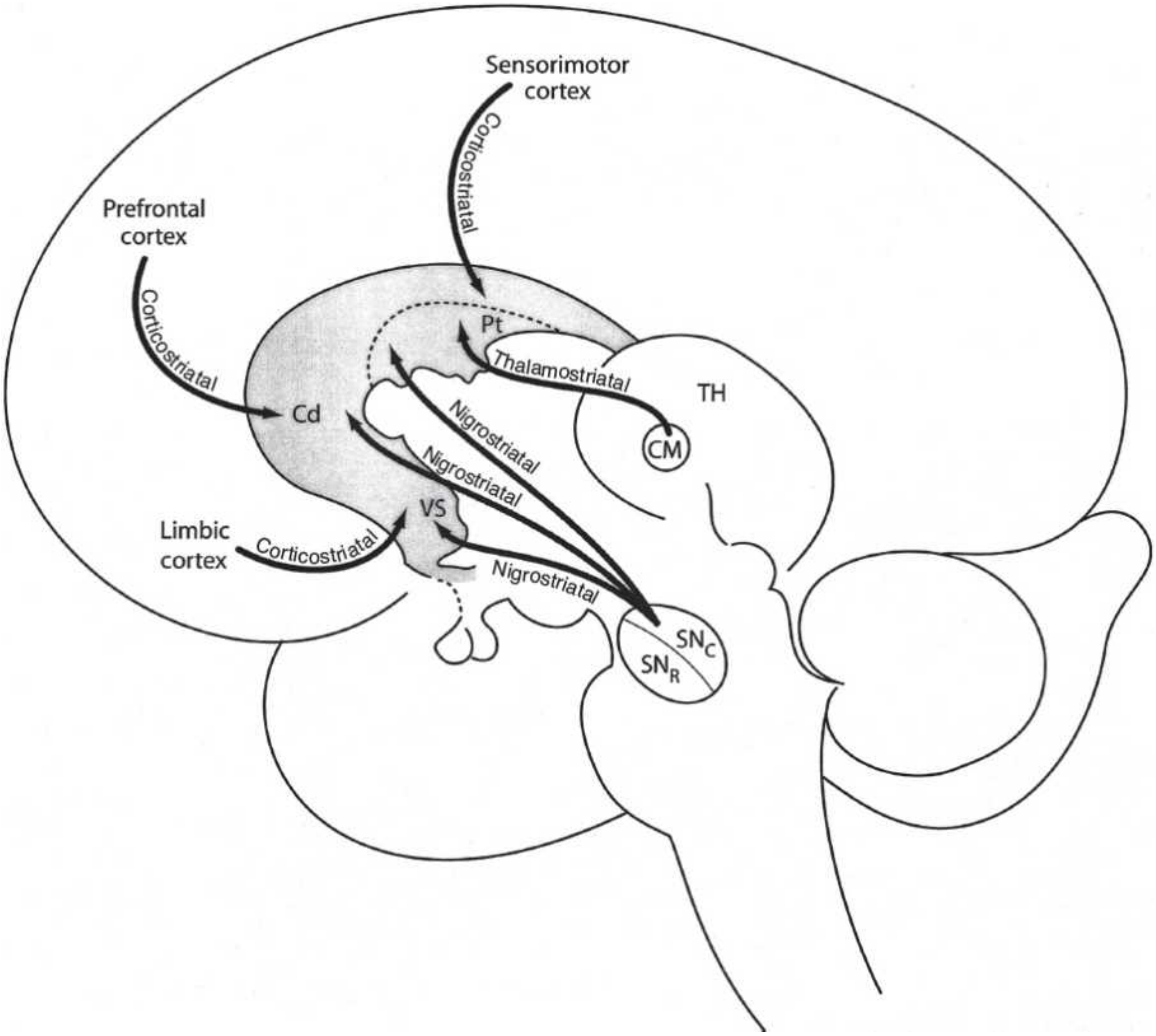
Sensorimotor cortex
Corticostriatal
Prefrontal cortex
Corticostriatal
Pt
Thalamostriatal
TH
Cd
Nigrostriatal
Nigrostriatal
CM
VS
Limbic cortex
Corticostriatal
Nigrostriatal
SN$_C$
SN$_R$

FIGURE 1.2–30. Diagram of the inputs to the basal ganglia system. Three major afferent systems have been identified: the corticostriatal, thalamostriatal, and nigrostriatal pathways.

Disruption of the input pathways of the basal ganglia has been associated with some movement disorders, such as Parkinson disease, which is characterized by muscular rigidity, fine tremor, shuffling gait, and bradykinesia. The most consistent neuropathologic feature of Parkinson disease is degeneration of the dopamine neurons in the substantia nigra pars compacta, accompanied by a loss of dopamine terminals in the striatum. The compound levodopa, a precursor in the biosynthesis of dopamine, is used as a treatment for Parkinson disease because of its ability to augment the release of dopamine from the remaining terminals. Conversely, the administration of typical antipsychotic agents in the treatment of schizophrenia is frequently associated with parkinsonian features and other motor system abnormalities; the fact that these agents are D_2 dopamine-receptor antagonists is thought to explain their movement-related side effects.

INTERNAL PROCESSING. The major processing pathways within the basal ganglia are summarized in Figure 1.2–31. As noted earlier, the striatum receives a major projection from the cerebral cortex. Within the striatum, the subclass of medium spiny neurons that contain the neuropeptide substance P sends inhibitory projections to the internal segment of the globus pallidus in what is termed the *direct pathway*. In contrast, the subpopulation of medium spiny neurons that contain the neuropeptide enkephalin provides inhibitory projections to the external segment of the globus pallidus, which sends inhibitory afferents to the internal segment of the globus pallidus in what is termed the *indirect pathway*. The globus pallidus external segment also projects to the pars reticulata of the substantia nigra. The topography found in the afferent projections to the striatum appears to be maintained in that processing pathway. For example, the sensorimotor territories of the striatum project most heavily to the ventral portion of the globus pallidus, whereas association territories project to the dorsal regions of the globus pallidus.

The external segment of the globus pallidus also gives rise to an inhibitory projection that terminates in the subthalamic nucleus. In contrast, neurons in the subthalamic nucleus provide excitatory projections that terminate in both segments of the globus pallidus and in the pars reticulata. Although most connections within the basal ganglia are unidirectional, a reciprocal projection is found between the external segment of the globus pallidus and the subthalamic nucleus. The intrinsic circuitry of the basal ganglia is disrupted by a severe loss of neurons in the striatum in Huntington disease. This autosomal dominant disorder is characterized by progressive chorea and dementia. The excessive number of trinucleotide repeats in the causal gene leads to the selective degeneration of striatal cells via the accumulation of the mutant huntingtin protein within the cytoplasm and nucleus of these neurons.

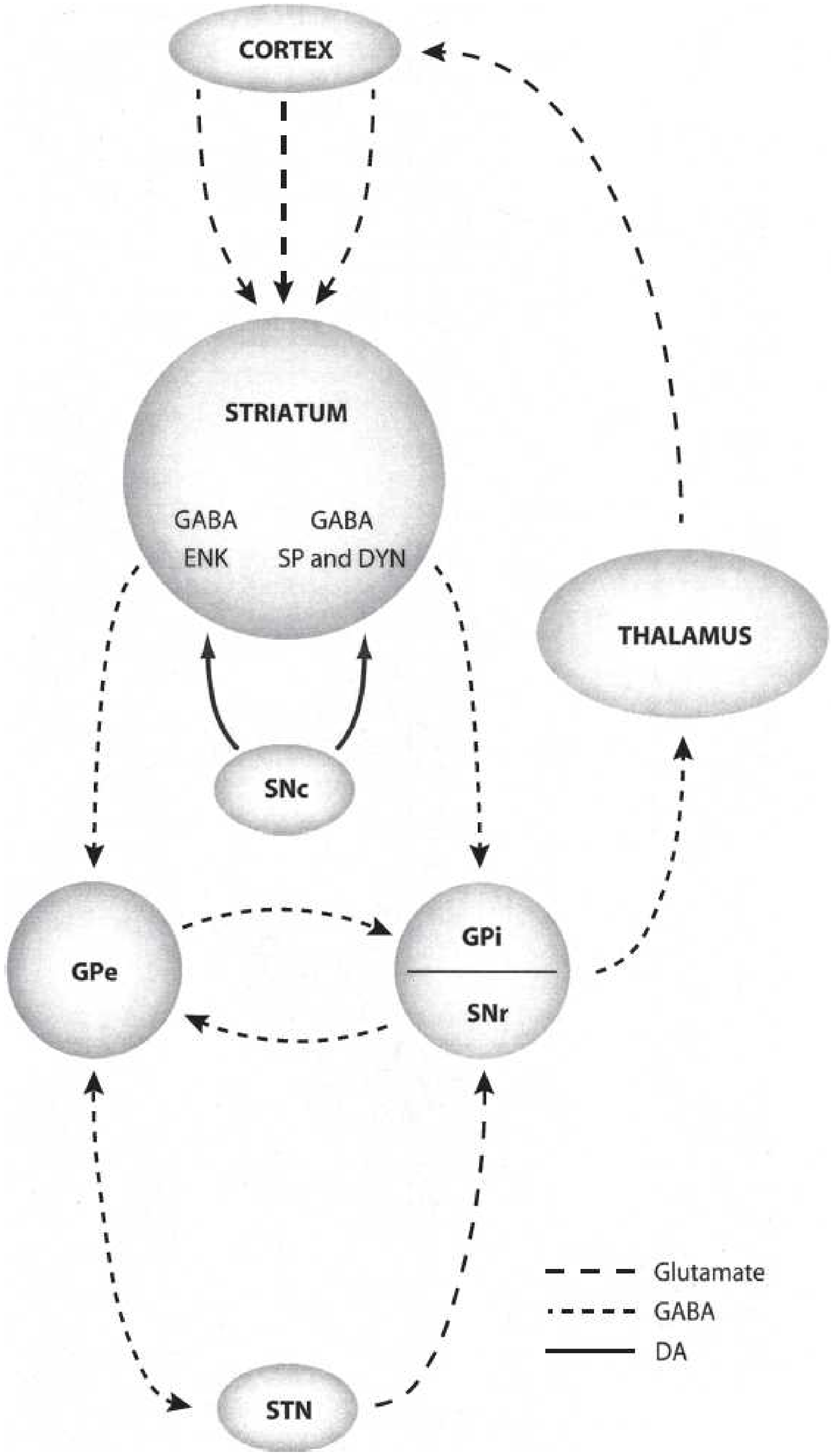
CORTEX
STRIATUM
GABA
ENK
GABA
SP and DYN
THALAMUS
SNc
GPe
GPi
SNr
STN
Glutamate
GABA
DA

FIGURE 1.2–31. Diagram of the intrinsic circuitry of the basal ganglia. Cortical neurons project to the striatum. Substance P (SP)–containing striatal neurons send an inhibitory projection directly to the internal segment of the globus

pallidus (GPi), whereas neurons containing enkephalin provide an inhibitory projection to GABA neurons in the external segment of the globus pallidus (GPe), which project to the GPi. The subthalamic nucleus (STN) receives a projection from the GPe and projects back to both segments. The STN and GPe then project to the substantia nigra pars reticulata (SNr), which projects to the thalamus. Finally, the thalamus projects back to the cerebral cortex, completing the loop.

OUTPUT OF THE BASAL GANGLIA. The internal segment of the globus pallidus is the source of much of the output of the basal ganglia (Fig. 1.2–32). This segment of the globus pallidus provides a projection to the ventral lateral and ventral anterior nuclei of the thalamus and to the intralaminar thalamic nuclei, particularly the central median nucleus. The pars reticulata of the substantia nigra also provides a projection to the ventral anterior and ventral lateral thalamic nuclei. These portions of the ventral lateral and ventral anterior thalamic nuclei project to the premotor and prefrontal cortices. As a result of the projections of the premotor and prefrontal cortices to the primary motor cortex, the basal ganglia are able to influence the output of the primary motor cortex indirectly. In addition, the cortical output of the basal ganglia exhibits marked convergence; although the striatum receives afferents from all regions of the neocortex, the eventual output of the globus pallidus and pars reticulata is largely conveyed through the thalamus to a much smaller portion of the neocortex—the premotor and prefrontal regions.

The functional consequences of the neural circuitry of the basal ganglia can also be considered in the context of some of the neurotransmitters used (see Figs. 1.2–31 and 1.2–32). Because the afferents from the cortex are thought to use glutamate, which is an excitatory neurotransmitter, cortical afferents presumably excite the structures of the basal ganglia in which they terminate. Many of the processing pathways within the basal ganglia use the inhibitory neurotransmitter GABA. Finally, the output pathways of the basal ganglia—the globus pallidus and the substantia nigra pars reticulata—use GABA as well. Excitation from cortical afferents eventually disinhibits the target structures of the basal ganglia because of the back-to-back inhibitory pathways of the basal ganglia.

Historically, motor systems have been divided into pyramidal (corticospinal), and extrapyramidal (basal ganglia) components; the names of these systems derive from the presence of the former in the pyramids, bilateral white matter structures which characterize the ventral surface of the medulla in the brainstem. This division into pyramidal and extrapyramidal is supported by clinical findings that lesions of each system result in distinct motor syndromes. For example, lesions of the extrapyramidal system result in involuntary movements, changes in muscle tone, and slowness of movement; lesions of the pyramidal system result in spasticity and paralysis. Because of these findings, the pyramidal and extrapyramidal systems were thought to control voluntary and involuntary movement independently. However, this division is no longer accurate for several reasons. First, other structures of the brain outside the traditional pyramidal and extrapyramidal systems, such as the cerebellum, are involved in the control of movement. Second, the pyramidal and extrapyramidal systems are not independent; the neural circuits of these systems are interconnected. For example, the basal ganglia influence motor behavior through certain regions of the cerebral cortex, which then directly (through the corticospinal tract) or indirectly (through specific brainstem nuclei) produce motor activity.

Finally, although the basal ganglia are important in the control of movement, this neural system also seems to be involved in other functions of the brain. More recent studies of the connections of the basal ganglia in nonhuman primates also support a role for these structures in cognitive functions. The DLPFC has been shown to receive inputs from portions of the thalamus that are the targets of projections from specific locations within the internal segment of the globus pallidus, providing evidence for a distinct pallidothalamocortical pathway. In addition to linking association regions of the cerebral cortex, such as the prefrontal and posterior parietal areas, with the control of motor activity in the primary motor cortex, some of the output of the basal ganglia seems to be directed back to regions of the prefrontal cortex. These findings suggest that “closed” loops are present between the prefrontal cortex and basal ganglia, which presumably have a cognitive rather than a motor function.

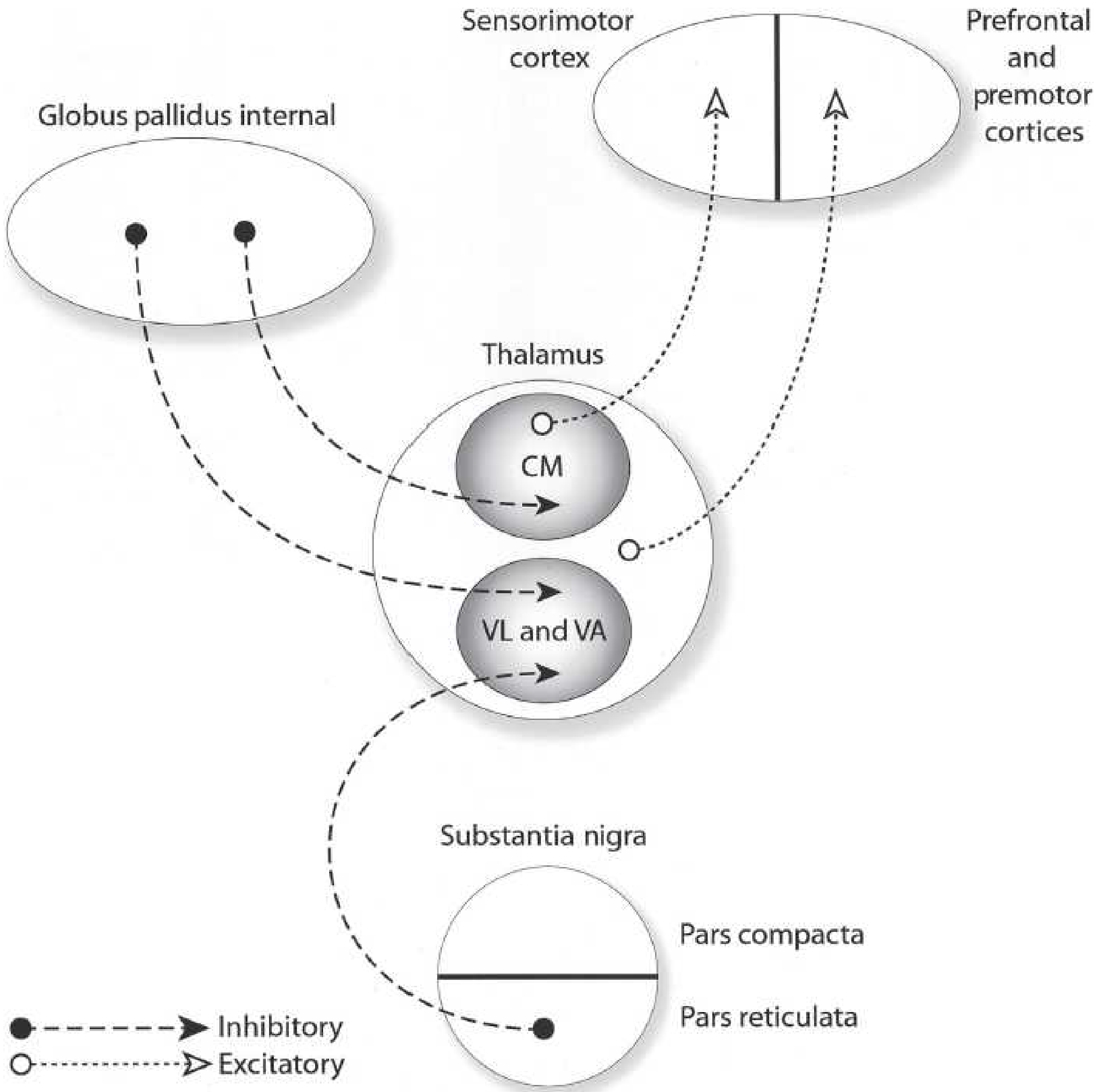
Sensorimotor cortex
Prefrontal and premotor cortices
Globus pallidus internal
Thalamus
CM
VL and VA
Substantia nigra
Pars compacta
Pars reticulata
Inhibitory
Excitatory

FIGURE 1.2–32. Diagram of the output of the basal ganglia system. The internal segment of the globus pallidus projects to the central median (*CM*), ventral lateral (*VL*), and ventral anterior (*VA*) nuclei of the thalamus. Those nuclei then project to sensorimotor, prefrontal, and premotor cortices. The substantia nigra pars reticulata also projects to the VL and VA nuclei.

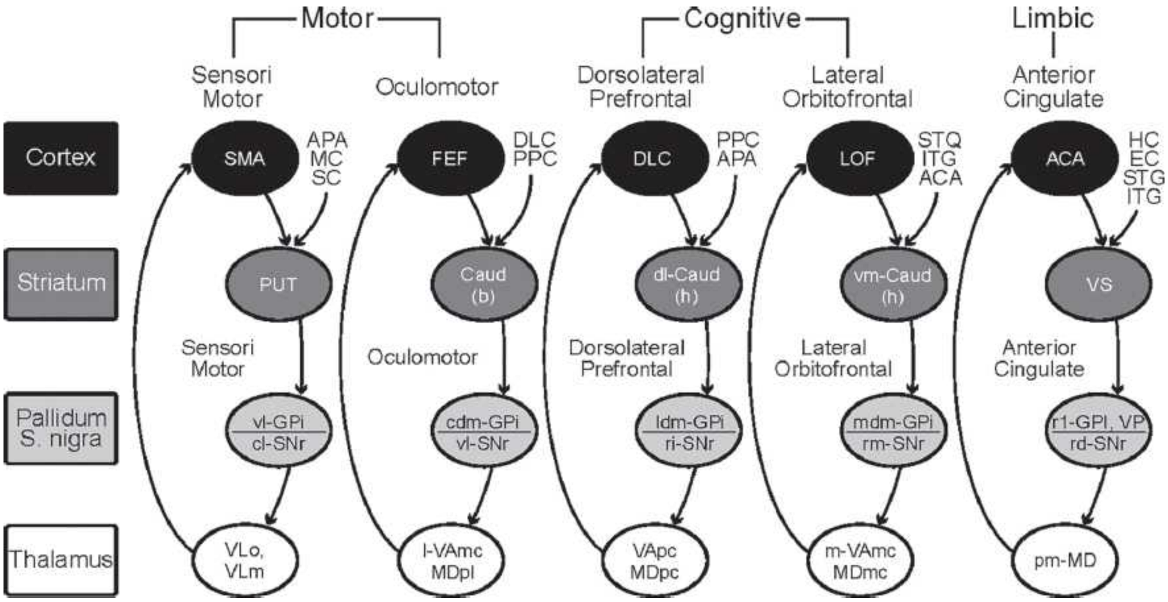
Motor
Cognitive
Limbic
Sensori Motor
Oculomotor
Dorsolateral Prefrontal
Lateral Orbitofrontal
Anterior Cingulate
Cortex
Striatum
Pallidum S. nigra
Thalamus
SMA
APA MC SC
FEF
DLC PPC
DLC
PPC APA
LOF
STQ ITG ACA
ACA
HC EC STG ITG
PUT
Caud (b)
dl-Caud (h)
vm-Caud (h)
VS
Sensori Motor
Oculomotor
Dorsolateral Prefrontal
Lateral Orbitofrontal
Anterior Cingulate
vl-GPi cl-SNr
cdm-GPi vl-SNr
ldm-GPi rl-SNr
mdm-GPi rm-SNr
r1-GPi, VP rd-SNr
VLo, VLm
l-VAmc MDpl
VApc MDpc
m-VAmc MDmc
pm-MD

FIGURE 1.2–33. Diagram illustrating various corticostriatalthalamocortical pathways originating in the frontal cortex. ACC, anterior cingulate cortex; Caud, caudate; b, body; h, head; DLPFC, dorsolateral prefrontal cortex; FEF, frontal eye fields; GPi, internal segment of globus pallidus; LOFC, lateral orbitofrontal cortex; MDpl, mediodorsal thalamus pars paralamellaris; MDmc, mediodorsal thalamus pars magnocellularis; MDpc, mediodorsal thalamus pars parvocellularis; PUT, putamen; SMA, supplementary motor area; SNr, substantia nigra pars reticulata; VAmc, ventral anterior thalamic nucleus pars magnocellularis; VApc, ventral anterior thalamic nucleus pars parvocellularis; VLm, ventral lateral thalamic nucleus pars medialis; VLo, ventral lateral thalamic nucleus pars oralis; VP, ventral pallidum; VS, ventral striatum; cl, caudolateral; cdm, caudal dorsomedial; dl, dorsolateral; l, lateral; ldm, lateral dorsomedial; m, medial; mdm, medial dorsomedial; pm, posteromedial; rd, rostrodorsal; rl; rostrolateral; rm, rostromedial; vm, ventro medial; vl, ventrolateral. (Adapted with permission from Annual Reviews. Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9(1):357–381. doi: 10.1146/annurev.ne.09.030186.002041.)

The basal ganglia are interconnected with the cerebral cortex and thalamus, forming five corticostriatalthalamocortical loops. These five parallel loops are depicted as closed, but recent studies suggest that they should be considered at least partially open due to some overlap among them. These loops can be categorized into Motor, Cognitive, or Limbic groups (Fig. 1.2–33). All five of these loops start in an area of the frontal cortex, make their way through specific input and output nuclei of the basal ganglia, and pass through a specific thalamic nucleus before returning to the cortical area of origin (Fig. 1.2–33). Each of these loops constitutes an integrated network mediating specific functions. The dorsolateral prefrontal loop contains structures known to be dysfunctional in schizophrenia, namely the DLPFC, the substantia nigra pars reticulata, and the mediodorsal thalamic nucleus. Given the role of the anterior cingulate cortex in affect regulation, the limbic loop is also likely affected in schizophrenia, as well as major depressive disorder. The ventral striatum's role in reward circuitry also implicates the limbic loop in the pathophysiology of major depressive disorder and addiction. Many regions within the lateral orbitofrontal loop appear to be dysfunctional in obsessive–compulsive disorder.

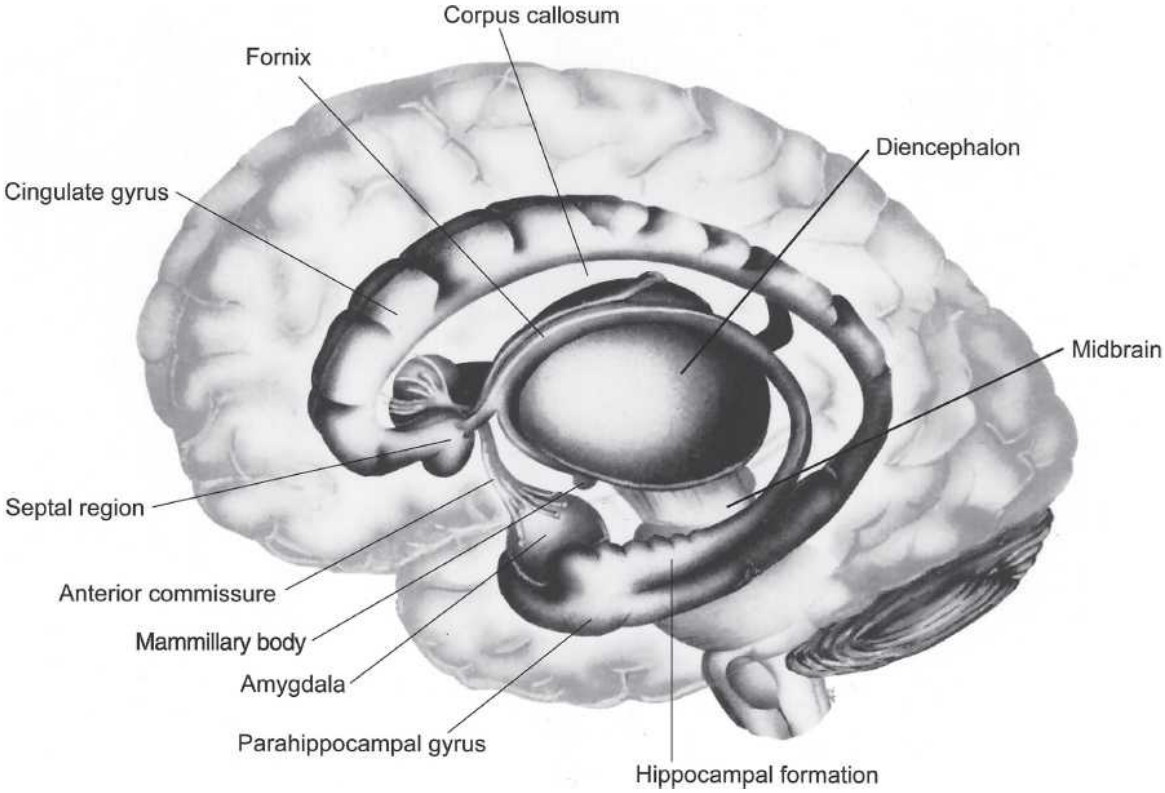
Corpus callosum
Fornix
Diencephalon
Cingulate gyrus
Midbrain
Septal region
Anterior commissure
Mammillary body
Amygdala
Parahippocampal gyrus
Hippocampal formation

FIGURE 1.2–34. Schematic drawing of the major anatomic structures of the limbic system. The cingulate and parahippocampal gyri form the "limbic lobe," a rim of tissue located along the junction of the diencephalon and the cerebral hemispheres. (Adapted from Hendelman WJ. *Student's Atlas of Neuroanatomy*. WB Saunders; 1994:179. Copyright © 1994 Elsevier.)

Limbic System

The concept of the limbic system as a neural substrate for emotional experience and expression has a rich but controversial history. More than 100 years ago, Paul Broca applied the term *limbic* (from the Latin *limbus*, meaning “border”) to the curved rim of the cortex, including the cingulate and the parahippocampal gyri, located at the junction of the diencephalon and the cerebral hemispheres (Fig. 1.2–34). In 1937, James Papez postulated, primarily on the basis of anatomic data, that these cortical regions were linked to the hippocampus, mammillary body, and anterior thalamus in a circuit that mediated emotional behavior (Fig. 1.2–35). This concept was supported by the work of Heinrich Klüver and Paul Bucy, who showed that temporal lobe lesions, which disrupt components of the circuit, alter affective responses in nonhuman primates. In 1952, Paul MacLean coined the term *limbic system* to describe Broca limbic lobe and related subcortical nuclei as the neural substrate for emotion. However, over the last 40 years, it has become clear that some limbic structures (e.g., the hippocampus) are also involved in other complex brain processes, such as memory. In addition, expanding knowledge of the connectivity of traditional limbic structures has made it increasingly difficult to define the boundaries of the limbic system. Despite these limitations, the concept of a limbic system may still be a useful way to describe the circuitry that relates certain telencephalic structures and their cognitive processes with the hypothalamus and its output pathways that control autonomic, somatic, and endocrine functions.

Major Structures. As noted above, no unanimity exists on the brain structures that constitute the limbic system. This section includes the brain regions that are most commonly listed as components of the limbic system: the cingulate and parahippocampal gyri (limbic cortex), the hippocampal formation, the amygdala, the septal area, the hypothalamus, and related thalamic and cortical areas.

LIMBIC CORTEX. The limbic cortex is composed of two general regions, the cingulate gyrus and the parahippocampal gyrus (see Fig. 1.2–34). The *cingulate gyrus*, a C-shaped stretch of cortex that lies dorsal to and follows the corpus callosum, includes several cortical regions that are defined by their relationship to the corpus callosum (Fig. 1.2–36). For example, *the subgenual anterior cingulate* lies ventral to the corpus callosum, extending from the anterior boundary of the genu to the rostrum (Fig. 1.2–36). The subgenual anterior cingulate is composed of Brodmann area 25 and subcallosal portions of Brodmann areas 32 and 24. The subgenual anterior cingulate is connected to other structures of the limbic system, such as the amygdala and hypothalamus, brain regions also often associated with depression. Recent imaging studies reveal that the subgenual anterior cingulate is overactive in people with depression and that a variety of treatments, such as antidepressants, electroconvulsive therapy, and repetitive transcranial magnetic stimulation, result in a decrease in activity in this area. Moreover, in preliminary studies of patients with treatment-resistant depression, deep brain stimulation (DBS) of the subgenual anterior cingulate and the underlying white matter has proven effective in the alleviation of symptoms. It’s important to note that the response documented after DBS stimulation likely results from not only direct activation of the subgenual anterior cingulate but also the activation of brain regions directly connected to it, including the amygdala, hypothalamus, and hippocampus. These areas are connected to the subgenual anterior cingulate via the callosum bundle, the unicate fasciculus, and the subcortical fascicle, and stimulation of these white matter tracts is necessary for the antidepressant response to DBS. Moreover, transsynaptic activation of areas not directly connected to the subgenual anterior cingulate, such as the DLPFC and the insula, might also play a role in the antidepressant effect of DBS.

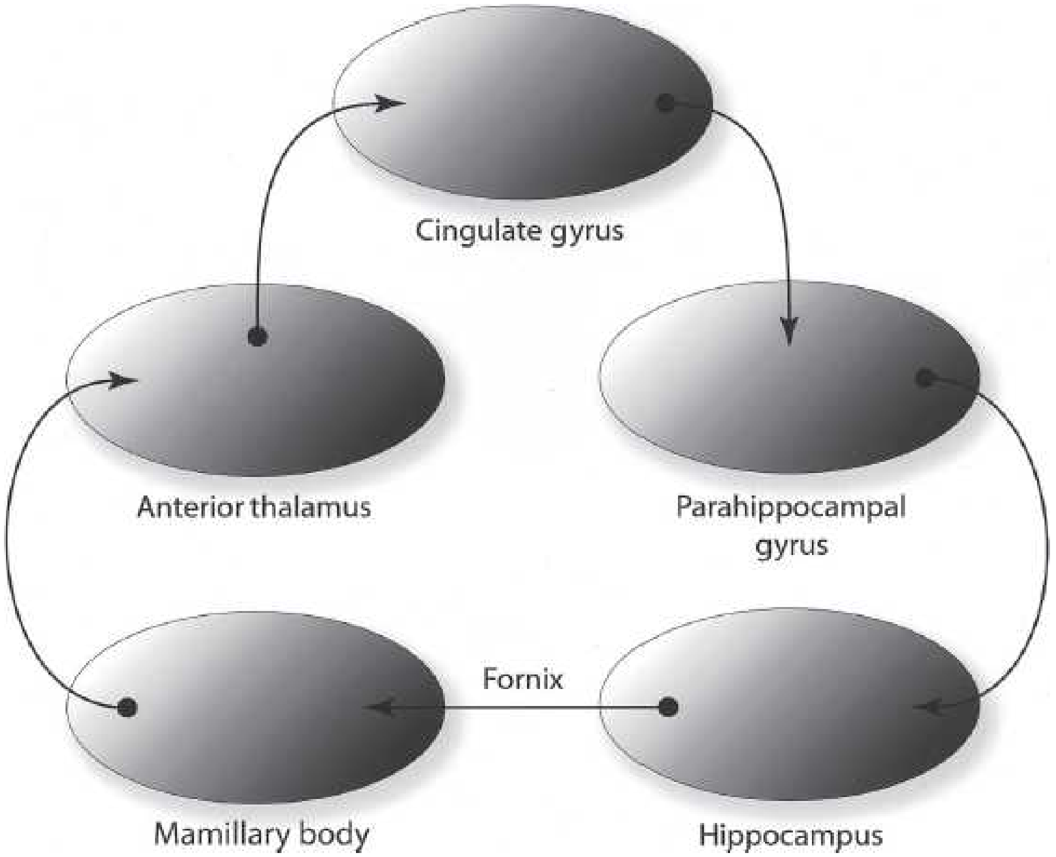
Cingulate gyrus
Anterior thalamus
Parahippocampal gyrus
Fornix
Mamillary body
Hippocampus

FIGURE 1.2–35. Diagram of the neural circuit for emotion as originally proposed by James Papez.

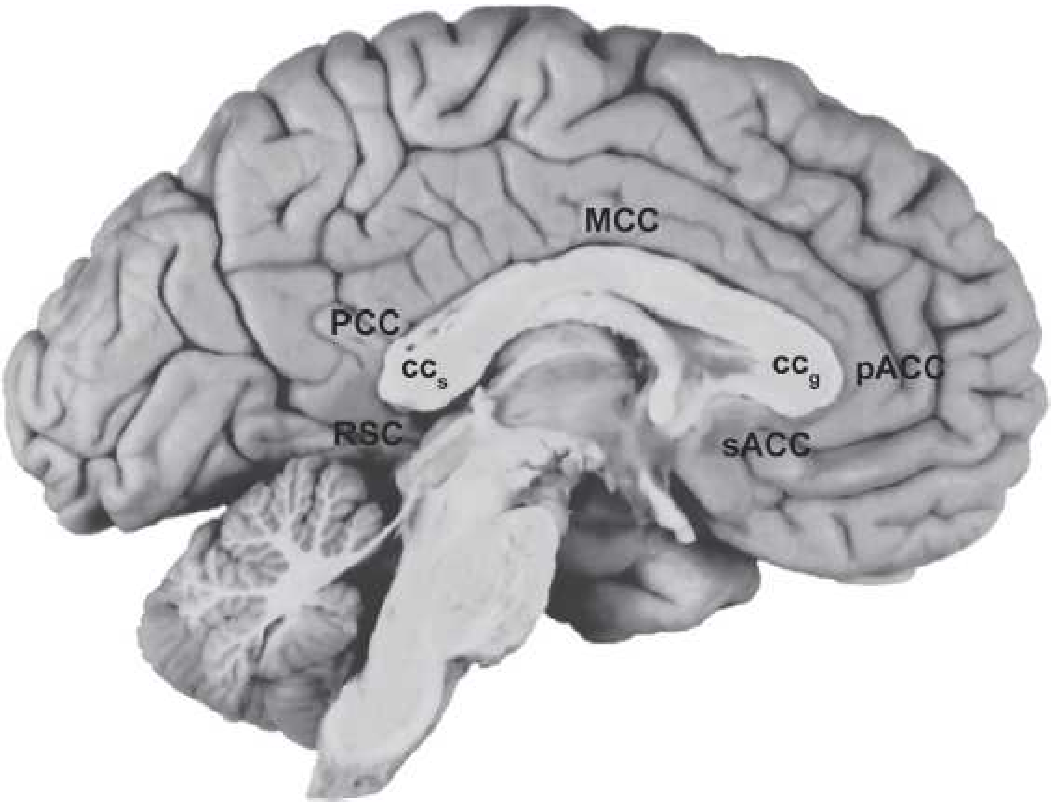

FIGURE 1.2–36. Photograph of the medial aspect of the left hemisphere of a human brain indicating the different subdivisions of the cingulate gyrus. sACC, subgenual anterior cingulate cortex; pACC, pregenual anterior cingulate cortex; MCC, midcingulate cortex; PCC, posterior cingulate cortex; RSC, retrosplenial cingulate cortex; cc_g, genu of the corpus callosum; ccs, splenium of the corpus callosum.

The habenula, although technically part of the epithalamus, has significant connections within the limbic system and is divided into medial and lateral subdivisions. The lateral habenula serves as a hub connecting limbic, basal ganglia, and brainstem circuits. These links, specifically those involving projections from the lateral habenula to serotonergic neurons of the raphe nuclei and dopaminergic cells of the ventral tegmental area, suggest that the lateral habenula may be involved in mood regulation and disorders, such as depression. Preliminary studies targeting the lateral habenula for DBS have been successful and opened the door to further study of this brain region in the pathophysiology and treatment of depression.

As illustrated in Figure 1.2–36, the cingulate gyrus is extensive, and as it travels posteriorly, it becomes continuous (via the cingulum bundle of fibers in the white matter) with the *parahippocampal gyrus*, located in the medial temporal lobe, which contains several distinct cytoarchitectonic regions. One of the most important of these regions is the *entorhinal cortex*, which not only funnels highly processed cortical information to the hippocampal formation but also is a major output pathway from the hippocampal formation.

HIPPOCAMPAL FORMATION. The hippocampal formation comprises three distinct zones—the dentate gyrus, the hippocampus, and the subicular complex—and is located in the floor of the temporal horn of the lateral ventricle (see Fig. 1.2–10). These zones are composed of adjacent strips of cortical tissue that run in a rostral–caudal direction but fold over each other mediolaterally in a spiral fashion, resulting in a C-shaped appearance. The *dentate gyrus* comprises three layers: an outer, acellular molecular layer, which faces the subarachnoid space of the hippocampal fissure; a middle layer composed of granule cells; and an inner polymorphic layer (Fig. 1.2–37). The granule cells extend their dendritic trees into the molecular layer and give rise to axons that form the mossy fiber projection to the hippocampus.

The *hippocampus* is also a trilaminate structure composed of molecular and polymorphic layers and a middle layer that contains pyramidal neurons. On the basis of differences in the cytoarchitecture and connectivity, the hippocampus can be divided into three distinct fields, which have been labeled CA3, CA2, and CA1 (see Fig. 1.2–37) (*CA* is derived from the term *cornu ammonis* after the Egyptian deity Ammon, who was depicted with ram’s horns, which some early investigators thought described the shape of the hippocampus). Some disagreement exists regarding the so-called CA4 region. This terminology has been applied to the portion of the hippocampus adjacent to CA3 and within the “C” created by the granule cell layer of the dentate gyrus. Connectional studies have revealed, however, that this area is more closely related to the dentate gyrus and should be referred to more appropriately as the hilar region or hilus of the dentate gyrus (see Fig. 1.2–37). The white matter adjacent to the polymorphic layer of the hippocampus is known as the *alveus*. The axons in this structure contribute to the fimbria, which, at the caudal end of the hippocampus, becomes the crus of the *fornix*. These bilateral structures converge to form the body of the fornix, which travels anteriorly and then turns inferiorly to form the columns of the fornix, which pass through the hypothalamus into the mammillary bodies (Fig. 1.2–38).

The *subicular complex* is generally considered to have five components—the presubiculum, the parasubiculum, the prosubiculum, the postsubiculum, and the subiculum—that together serve as transition regions between the hippocampus and the parahippocampal gyrus.

The components of the hippocampal formation have a distinct pattern of intrinsic connectivity that is largely unidirectional and provides for a specific flow of information. The major input to the hippocampal formation arises from neurons in layers 2 and 3 of the entorhinal cortex that project through the *perforant path* (that is, through the subiculum and the hippocampus) to the outer two-thirds of the molecular layer of the dentate gyrus where they synapse on the dendrites of granule cells. The mossy fiber axons of the granule cells provide a projection to the pyramidal neurons of the CA3 field of the hippocampus. Axon collaterals from CA3 pyramidal neurons project within CA3, and the primary axons of these CA3 neurons, the so-called Schäffer collaterals, project to the CA1 field of the hippocampus. This region projects to the subicular complex, which provides output to the entorhinal cortex, completing the circuit.

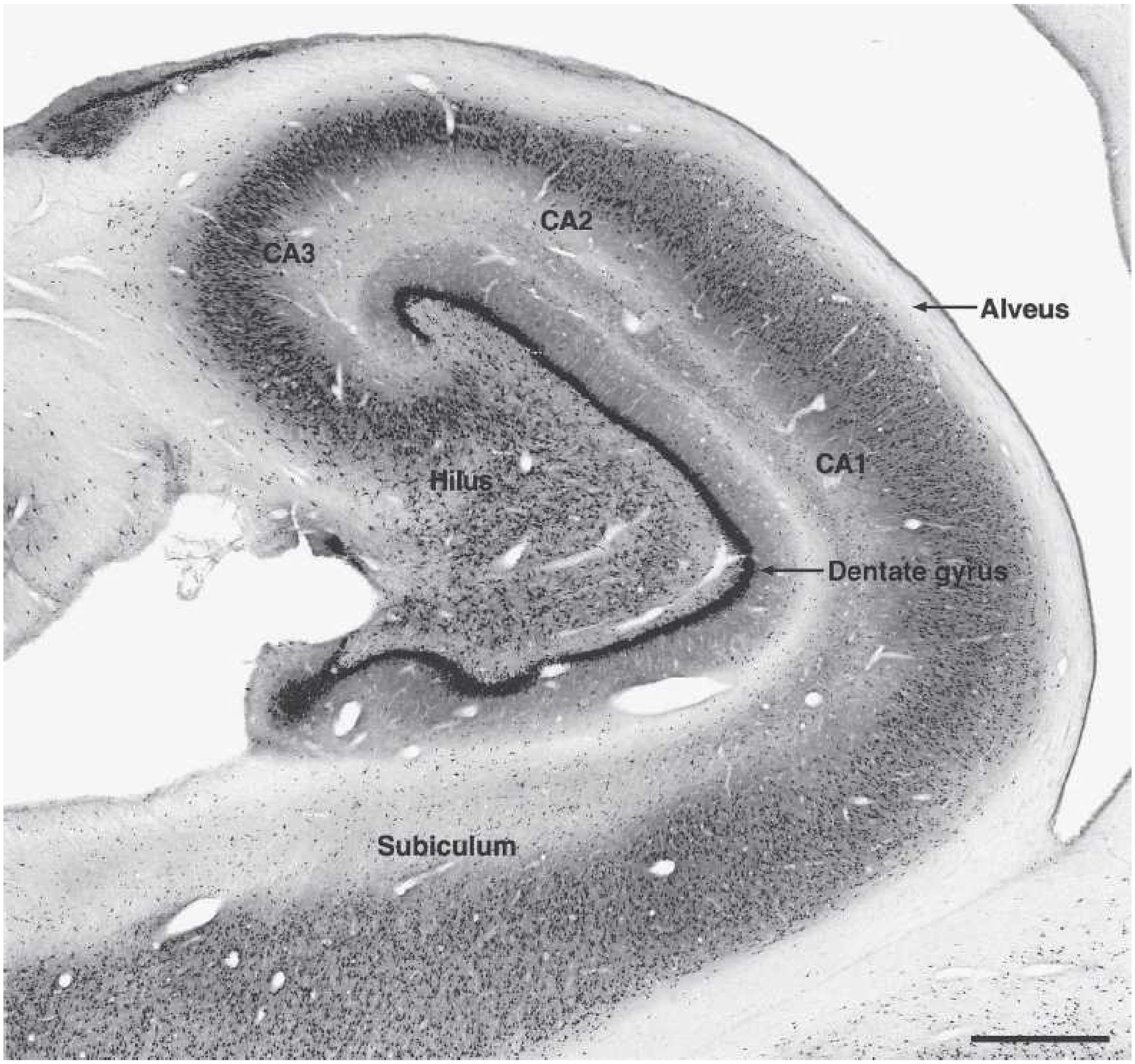

FIGURE 1.2–37. Photomicrograph of neurons immunoreactive for neuron-specific nuclear protein in the human hippocampal formation. The immunostaining illustrates the major components of the hippocampal formation. Scale bar = 1 mm.

AMYGDALA. Located in the medial temporal lobe just anterior to the hippocampal formation are a group of nuclei referred to as the *amygdala* (see Fig. 1.2–9). These nuclei form several distinct clusters: the basolateral complex, the centromedial amygdaloid group, and the olfactory group, which includes the cortical amygdaloid nuclei. These nuclei are usually delineated using cytoarchitectural features revealed by Nissl stains. However, the chemoarchitecture of cannabinoid CB_1 receptor–containing axons also clearly demarcates these nuclei (Fig. 1.2–39). CB_1 receptor immunoreactivity is found within the basolateral nuclei, the largest of the three groups, whereas the central and medial nuclei are devoid of CB_1 receptor labeling. The *basolateral complex* differs from the remaining amygdaloid nuclei in many respects. Although the basolateral complex is not a laminated structure, its connectivity and some other anatomic characteristics are more similar to cortical regions than to the remaining amygdaloid nuclei. For example, the basolateral nuclei are directly and reciprocally connected with the temporal, insular, and prefrontal cortices. In addition, similar to some cortical regions, the basolateral complex shares bidirectional connections with the medial dorsal thalamic nucleus and receives projections from the midline and intralaminar thalamic nuclei. Finally, neurons of the basolateral complex with a pyramidal-like morphology that use excitatory amino acids as neurotransmitters furnish projections to the striatum. On the basis of these anatomic characteristics, one may hypothesize that the basolateral complex actually functions similar to a multimodal cortical region.

In contrast, the *centromedial amygdala* appears to be part of a larger structure that is continuous through the sublenticular substantia innominata with the bed nucleus of the stria terminalis. This larger structure, which has been termed the *extended amygdala*, consists of two major subdivisions. The central subdivision of the extended amygdala includes the central amygdaloid nucleus and the lateral portion of the bed nucleus of the stria terminalis. This subdivision is reciprocally connected with brainstem viscerosensory and visceromotor regions and with the lateral hypothalamus. In addition, it receives afferents from cortical limbic regions and the basolateral amygdaloid complex. In contrast, the medial subdivision of the extended amygdala, composed of the medial amygdaloid nucleus and its extension into the medial part of the bed nucleus of the stria terminalis, is distinguished by reciprocal connections with the medial or endocrine portions of the hypothalamus.

SEPTAL AREA. The *septal area* is a gray matter structure located immediately above the anterior commissure (Fig. 1.2–40). The septal nuclei are reciprocally connected with the hippocampus, the amygdala, and the hypothalamus and project to numerous structures in the brainstem. The septum and the hippocampus have extensive reciprocal connections, forming the septohippocampal pathway, a neurobiologic circuit for the regulation of anxiety.

HYPOTHALAMUS. The hypothalamus, a small structure within the diencephalon, is a crucial component of the neural circuitry regulating not only emotions but also autonomic, endocrine, and some somatic functions. In addition to its relationships with other components of the limbic system, it is interconnected with various visceral and somatic nuclei of the brainstem and the spinal cord and provides an output that regulates the function of the pituitary gland. On its inferior surface, the hypothalamus is bounded rostrally by the optic chiasm and caudally by the posterior edge of the mammillary bodies. The area of the hypothalamus between these two structures, called the *tuber cinereum*, gives rise to the median eminence, which is continuous with the infundibular stalk and then the posterior lobe of the pituitary gland (Fig. 1.2–41). On the basis of these features, the hypothalamus is subdivided from anterior to posterior into three zones: the supraoptic region, the infundibular region, and the mammillary region. (In addition, the preoptic area, a telencephalic structure located immediately anterior to the supraoptic region, is usually considered part of the hypothalamus.) These three zones also are divided on each side into medial and lateral areas by the fornix as it travels through the body of the hypothalamus to the mammillary bodies. As shown in Table 1.2–4, the six parts of the hypothalamus contain different nuclei.

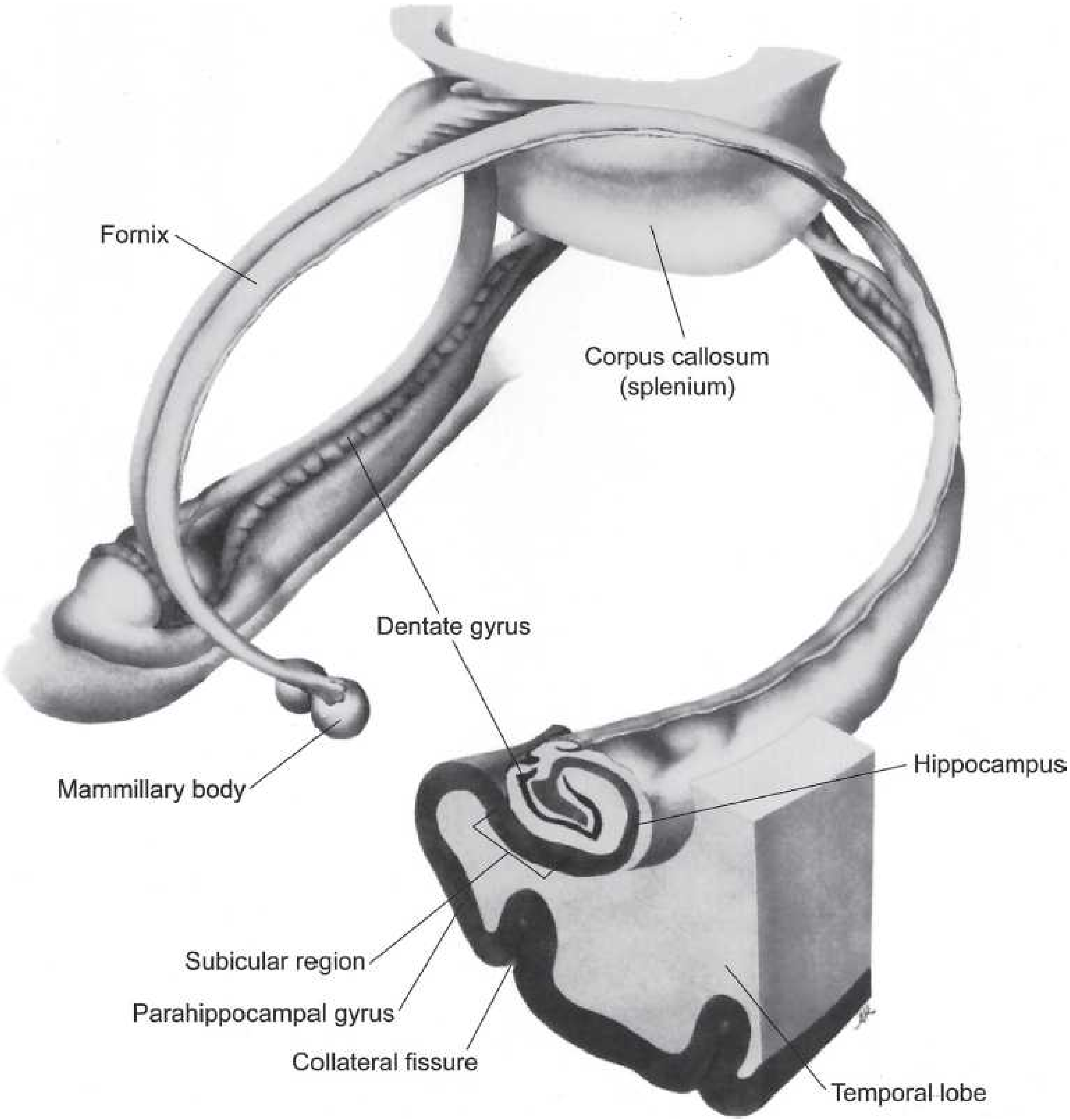
Fornix
Corpus callosum (splenium)
Dentate gyrus
Mammillary body
Hippocampus
Subicular region
Parahippocampal gyrus
Collateral fissure
Temporal lobe

FIGURE 1.2–38. Schematic drawing of a cross-sectional view of the hippocampal formation and the path of the fornix running between that structure and the mammillary bodies. (Adapted from Hendelman WJ. *Student's Atlas of Neuroanatomy*. WB Saunders; 1994:189. Copyright © 1994 Elsevier.)

These different nuclei mediate the diverse functions of the hypothalamus. The *suprachiasmatic nucleus* receives direct and indirect projections from the retina and is important in the regulation of diurnal rhythms. The *supraoptic* and *paraventricular nuclei* contain large cells (magnocellular neurons) that send oxytocin-containing and vasopressin-containing fibers to the posterior neural lobe of the pituitary. In addition, some neurons of the paraventricular nucleus project to the median eminence, where they release neuropeptides, such as corticotropin-releasing factor, into the portal blood system. These neuropeptides control the synthesis and release of anterior pituitary hormones. The paraventricular nucleus also gives rise to descending projections that regulate the sympathetic and parasympathetic autonomic areas of the medulla and the spinal cord.

Table 1.2–4.
Hypothalamic Nuclei

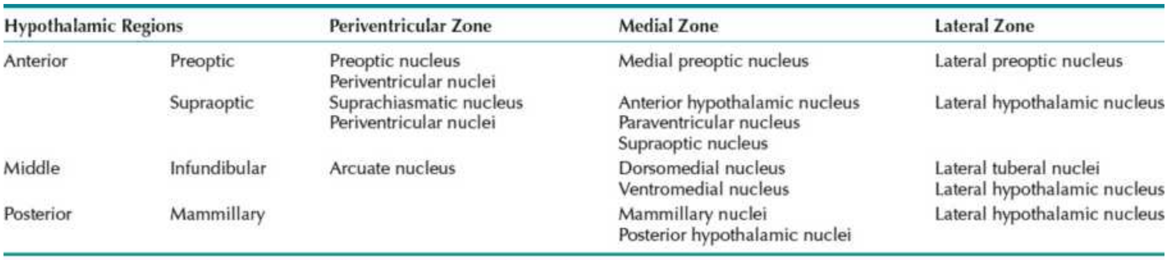

Hypothalamic Regions		Periventricular Zone	Medial Zone	Lateral Zone
Anterior	Preoptic	Preoptic nucleus Periventricular nuclei	Medial preoptic nucleus	Lateral preoptic nucleus
	Supraoptic	Suprachiasmatic nucleus Periventricular nuclei	Anterior hypothalamic nucleus Paraventricular nucleus Supraoptic nucleus	Lateral hypothalamic nucleus
Middle	Infundibular	Arcuate nucleus	Dorsomedial nucleus Ventromedial nucleus	Lateral tuberal nuclei Lateral hypothalamic nucleus
Posterior	Mammillary		Mammillary nuclei Posterior hypothalamic nuclei	Lateral hypothalamic nucleus

Modified from Patestas MA, Gartner LP. *A Textbook of Neuroanatomy*. 2nd ed. John Wiley & Sons; 2016:441.

Within the medial tuberal region of the hypothalamus, the *ventromedial* and *arcuate nuclei* also participate in the regulation of the anterior pituitary function. In addition, the ventromedial nucleus may play an important role in reproductive and ingestive behavior. The medial posterior section of the hypothalamus contains the *posterior nucleus* and the *mammillary bodies*. Within the mammillary bodies, the lateral and medial mammillary nuclei receive hippocampal input through the fornix (see Fig. 1.2–38) and project to the anterior nuclei of the thalamus. The posterior nucleus shares reciprocal connections with the extended amygdala. This nucleus appears to be relatively more developed in primates than in rodents, suggesting that it plays an important but still-to-be-clarified role in the human brain.

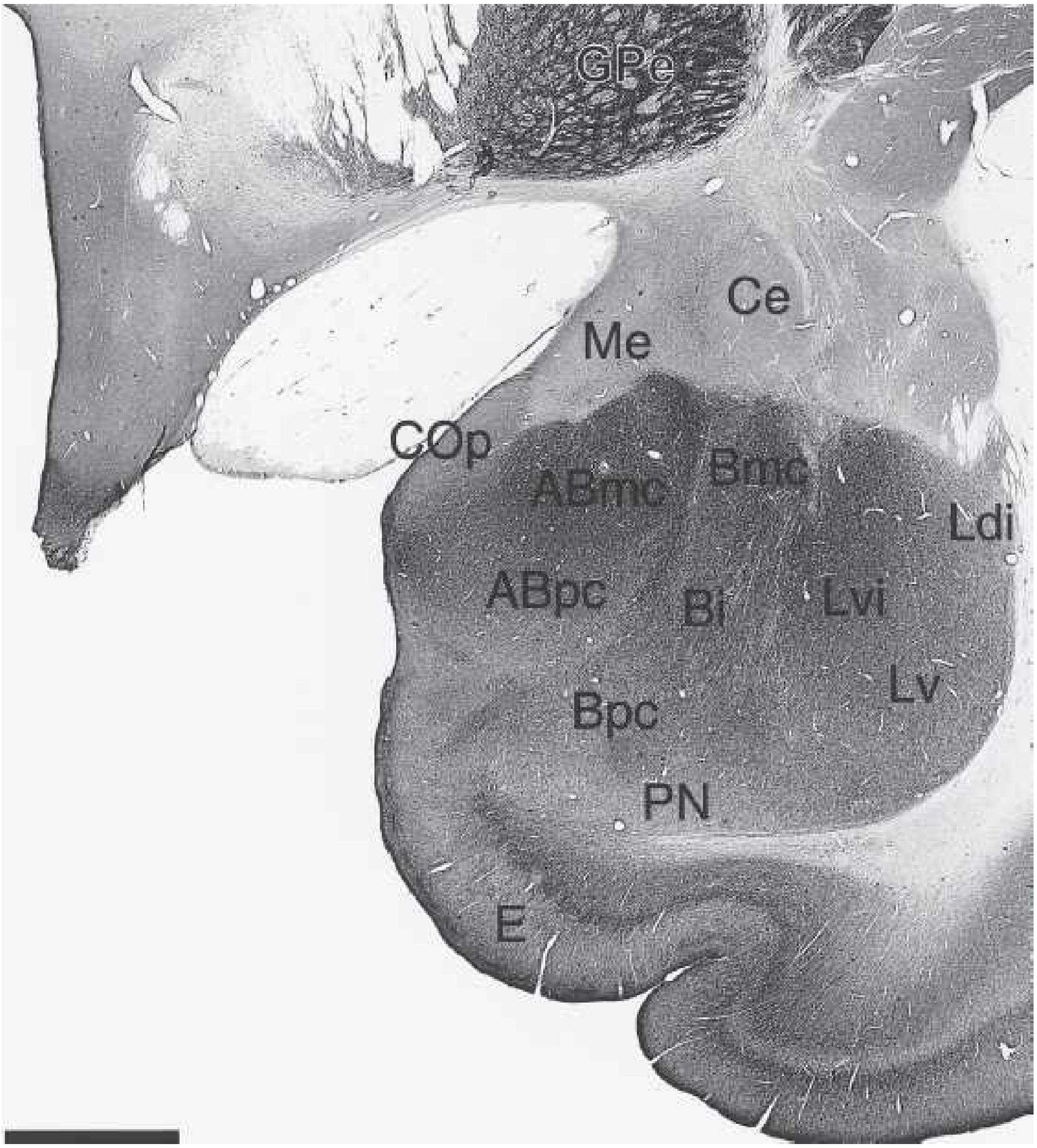

FIGURE 1.2–39. Photomicrograph of a coronal section through macaque monkey brain illustrating the distribution of the cannabinoid CB_1-immunoreactive axons in the amygdala. The density of labeled axons is high in the cortical-like basolateral nuclei (*ABmc, ABpc, Bi, Bmc, Bpc, Ldi, Lv, Lvi*), whereas the striatal-like central (*Ce*) and medial (*Me*) nuclei are devoid of CB_1-immunoreactive axons. ABmc, accessory basal nucleus, magnocellular division; ABpc, accessory basal nucleus, parvicellular division; Bi, basal nucleus, intermediate division; Bmc, basal nucleus, magnocellular division; Bpc, basal nucleus, parvicellular division; Ce, central amygdaloid nucleus; Cop, posterior cortical nucleus; E, entorhinal cortex; GPe, globus pallidus, external; Ldi, lateral nucleus, dorsal intermediate division; Lv, lateral nucleus, ventral division; Lvi, lateral nucleus, ventral intermediate division; Me, medial amygdaloid nucleus; PN, paralaminar nucleus. Calibration bar = 2 mm. (Adapted from Eggan SM, Lewis DA. Immunocytochemical distribution of the

cannabinoid CB1 receptor in the primate neocortex: A regional and laminar analysis. *Cereb Cortex*. 2007;17:175, by permission of Oxford University Press.)

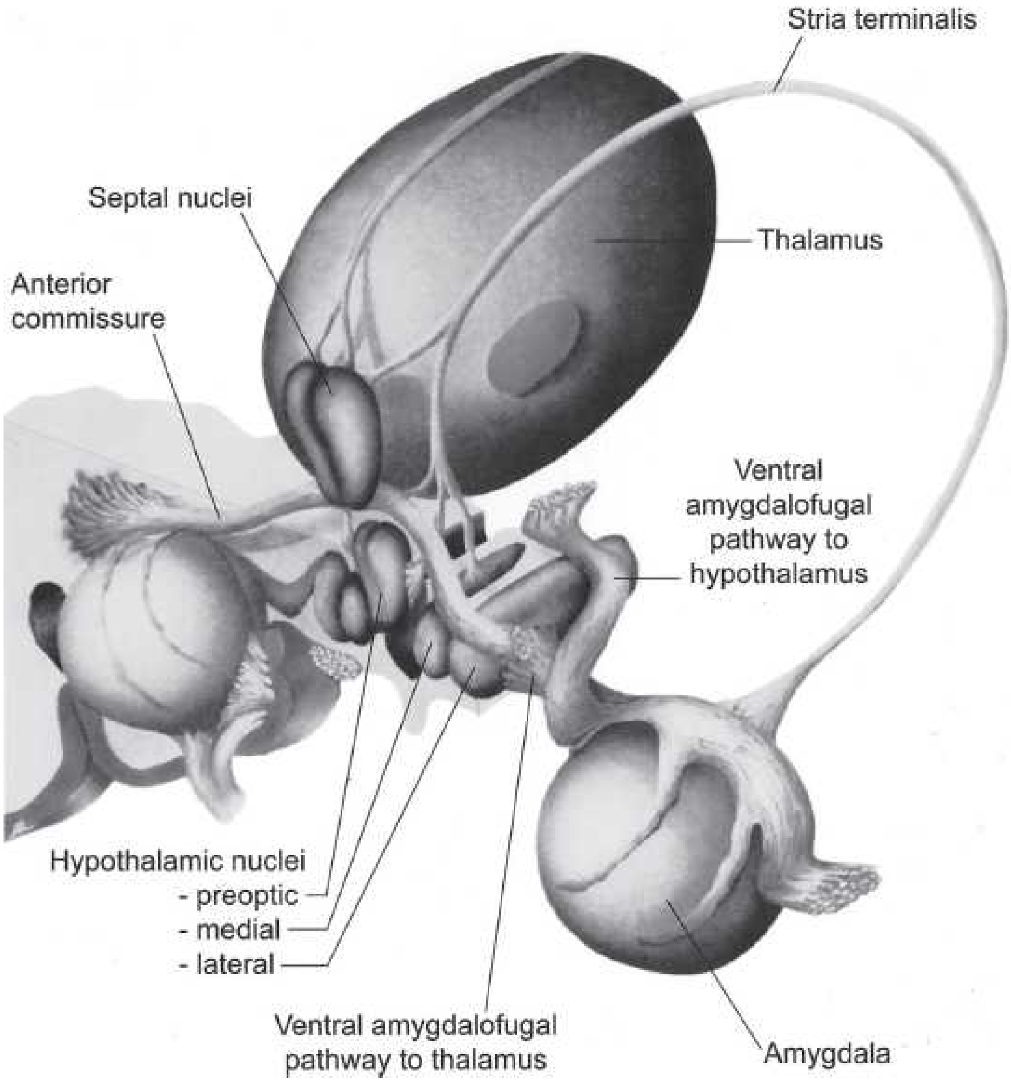
Stria terminalis
Septal nuclei
Thalamus
Anterior commissure
Ventral amygdalofugal pathway to hypothalamus
Hypothalamic nuclei
- preoptic
- medial
- lateral
Ventral amygdalofugal pathway to thalamus
Amygdala

FIGURE 1.2–40. Schematic drawing of some components of the limbic system showing the major output pathways of the amygdala, the stria terminalis, and the ventral amygdalofugal pathway. (Adapted from Hendelman WJ. *Student's Atlas of Neuroanatomy*. WB Saunders; 1994:183. Copyright © 1994 Elsevier.)

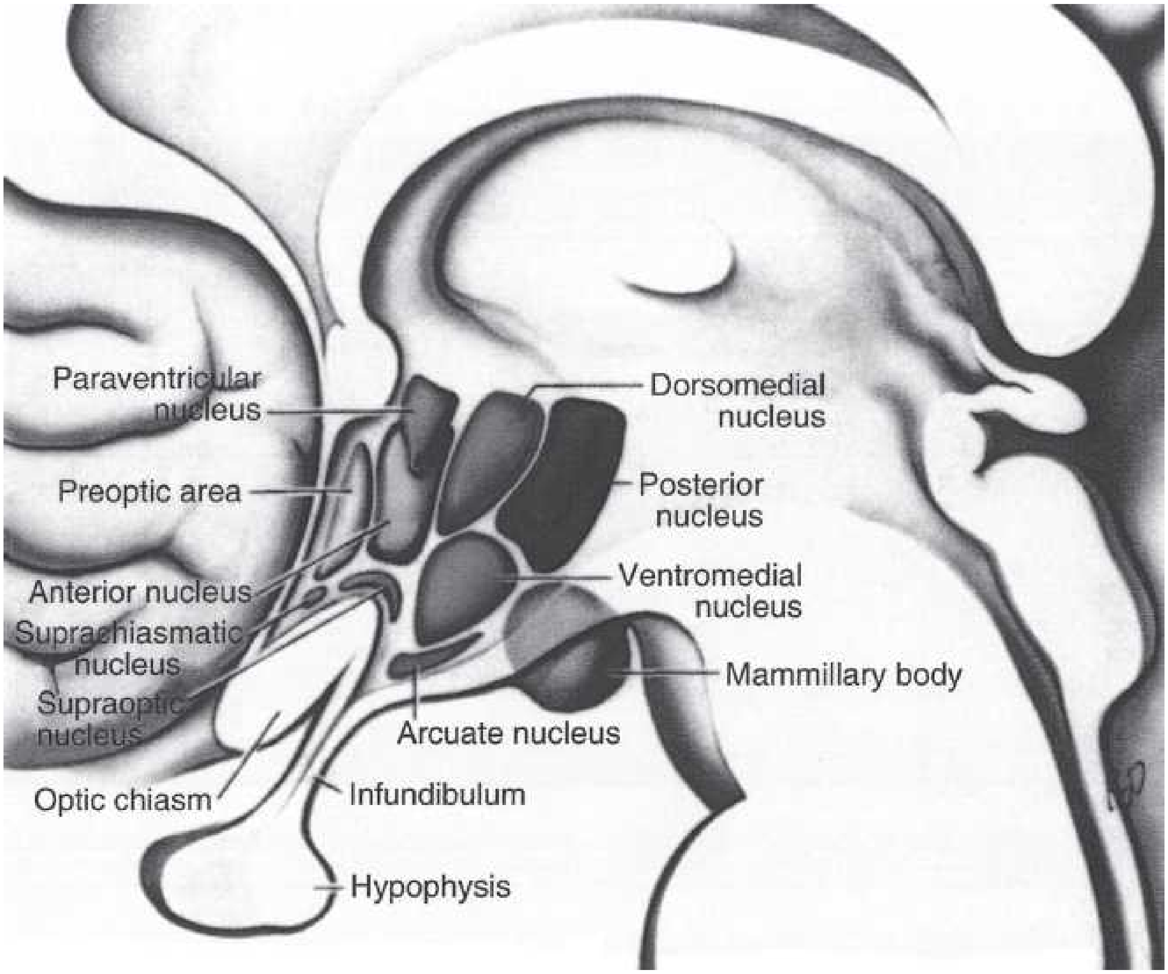
Paraventricular nucleus
Dorsomedial nucleus
Preoptic area
Posterior nucleus
Anterior nucleus
Ventromedial nucleus
Suprachiasmatic nucleus
Supraoptic nucleus
Mammillary body
Arcuate nucleus
Optic chiasm
Infundibulum
Hypophysis

FIGURE 1.2–41. Schematic drawing of the nuclei in the medial hypothalamus. (Modified from Parent A. *Carpenter's Human Neuroanatomy*. 9th ed. Williams & Wilkins; 1996:707.)

The lateral portions of the hypothalamus contain a low density of neurons scattered among longitudinally running fibers of the medial forebrain bundle. This region is interconnected with multiple regions of the forebrain, the brainstem, and the spinal cord. The lateral hypothalamic area also contains a population of neurons that express the orexin neuropeptides, orexin A and orexin B (also known as hypocretin A and hypocretin B), which seem to be involved in sleep and wakefulness. The approximately 70,000 orexin-producing neurons in the human brain project throughout the brain, with the exception of the cerebellum (Fig. 1.2–42). Orexin neurons project to most of the monoaminergic (i.e., substantia nigra, locus coeruleus, dorsal raphe) and cholinergic (i.e., medial septum, pedunculopontine, laterodorsal tegmental) nuclei. Orexin neurons also have widespread projections throughout the cerebral cortex.

Areas containing high densities of orexin axons include the paraventricular thalamic nucleus, the arcuate nucleus of the hypothalamus, the locus coeruleus, and the dorsal raphe nucleus. The projections of orexin neurons to neuronal systems involved in sleep and wakefulness (i.e., locus coeruleus, raphe nuclei, and laterodorsal/pedunculopontine tegmental nuclei) suggest that orexin neurons participate in these functions. Numerous studies in animals and humans show that an orexin deficiency is the main cause of narcolepsy. For example, mice that lack the orexin gene exhibit physiologic symptoms similar to human narcolepsy, and postmortem examination of the brains of narcolepsy patients have revealed an 85 to 95 percent reduction in the number of orexin-immunoreactive neurons.

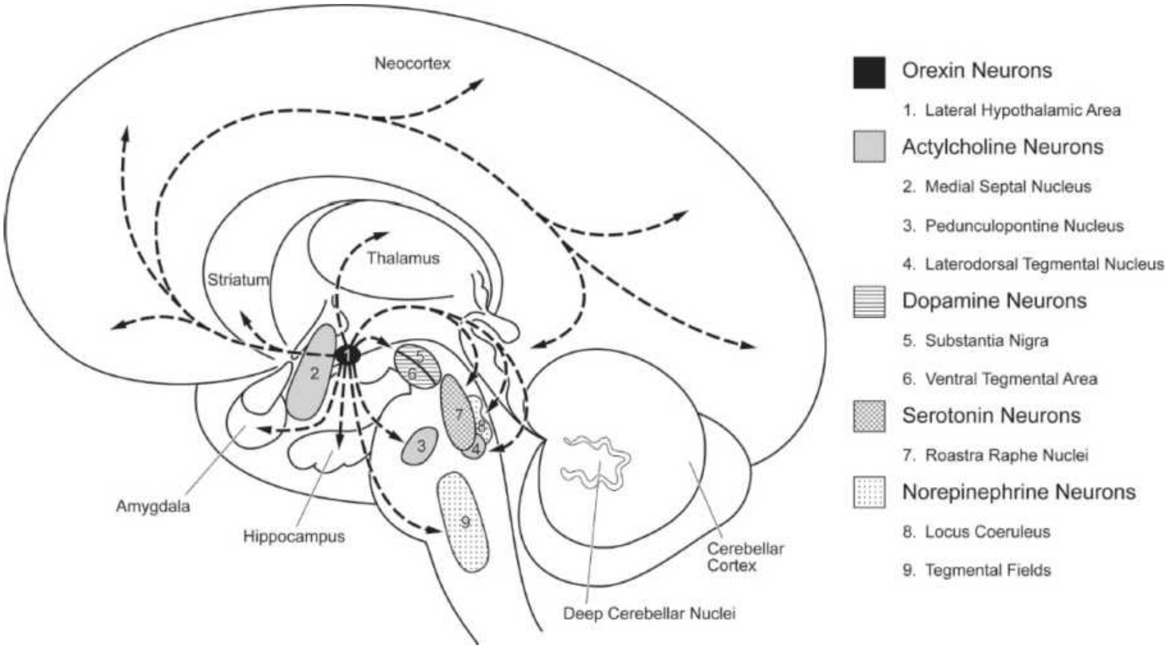
Neocortex
Thalamus
Striatum
Amygdala
Hippocampus
Cerebellar Cortex
Deep Cerebellar Nuclei
1
2
3
4
5
6
7
8
9
■ Orexin Neurons
1. Lateral Hypothalamic Area
▨ Actylcholine Neurons
2. Medial Septal Nucleus
3. Pedunculopontine Nucleus
4. Laterodorsal Tegmental Nucleus
▤ Dopamine Neurons
5. Substantia Nigra
6. Ventral Tegmental Area
▩ Serotonin Neurons
7. Roastra Raphe Nuclei
⬚ Norepinephrine Neurons
8. Locus Coeruleus
9. Tegmental Fields

FIGURE 1.2–42. Schematic drawing illustrating the circuitry of orexin neurons.

Functional Circuitry. The major structures of the limbic system are interconnected with each other and with other components of the nervous system in various ways. Several major output pathways of the limbic system are clearly defined. In one pathway (Fig. 1.2–43), highly processed sensory information from the cingulate, the orbital and temporal cortices, and the amygdala is transmitted to the entorhinal cortex of the parahippocampal gyrus and then to the hippocampal formation. After traversing the intrinsic circuitry of the hippocampal formation, information is projected through the fornix either to the anterior thalamus, which projects to the limbic cortex, or to the septal area and the hypothalamus. These latter two regions provide feedback to the hippocampal formation through the fornix. In addition, the mammillary bodies of the hypothalamus project to the anterior thalamus. Finally, the hypothalamus and the septal area project to the brainstem and the spinal cord.

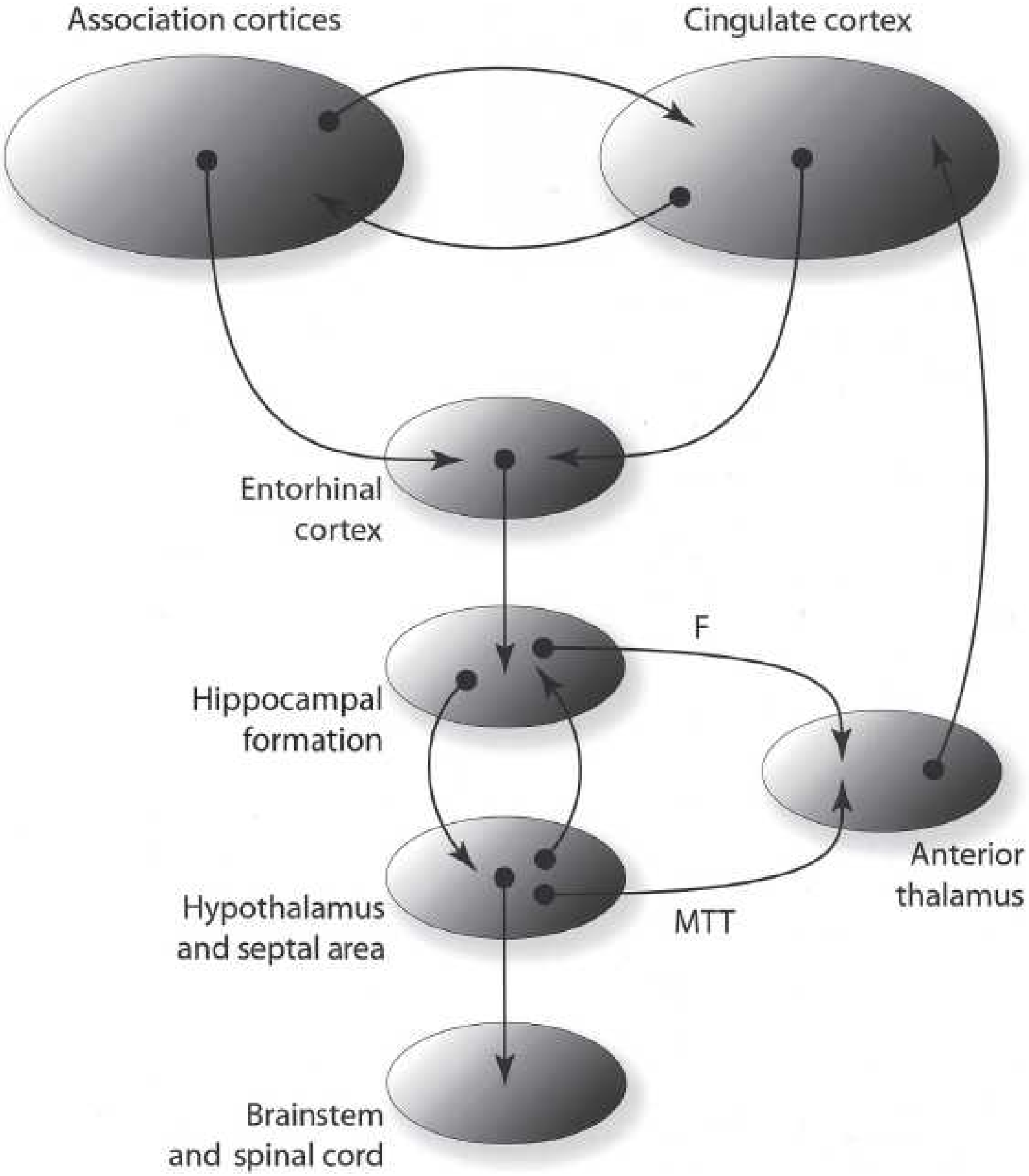
Association cortices
Cingulate cortex
Entorhinal cortex
F
Hippocampal formation
Anterior thalamus
Hypothalamus and septal area
MTT
Brainstem and spinal cord

FIGURE 1.2–43. Functional neural circuitry of the limbic system. This diagram illustrates the manner in which the hippocampal formation and the anterior thalamus provide a mechanism for the integration of information between the cerebral cortex and the hypothalamus. F, fornix; MTT, mammillothalamic tract. (Adapted from Nolte J. *The Human Brain: An Introduction to Its Functional Anatomy*. 3rd ed. Mosby;1993:399.)

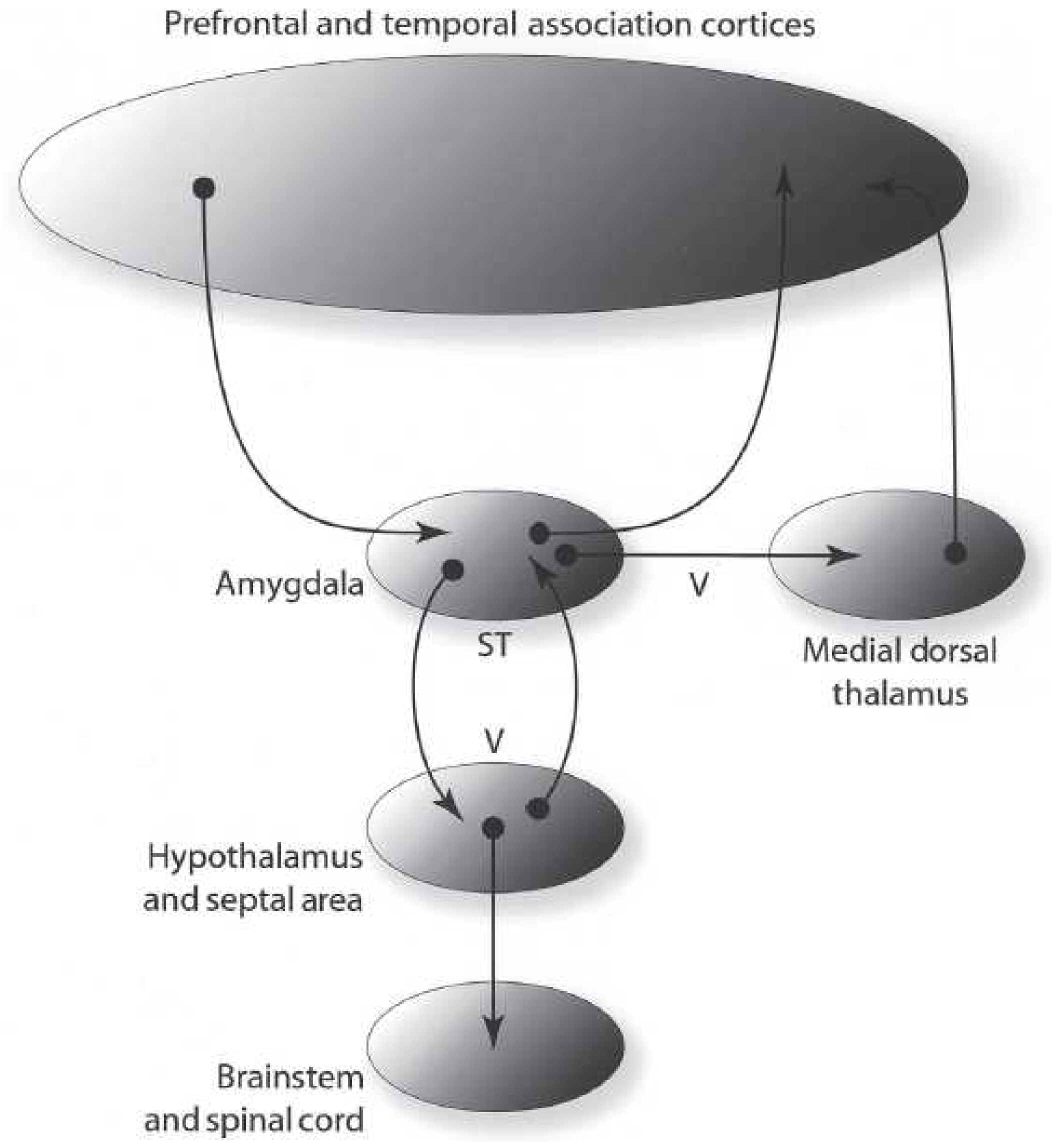
Prefrontal and temporal association cortices
Amygdala
ST
V
Medial dorsal thalamus
V
Hypothalamus and septal area
Brainstem and spinal cord

FIGURE 1.2–44. Functional neural circuitry of the limbic system. This diagram illustrates how the amygdala and the medial dorsal thalamic nucleus serve to integrate information processing between prefrontal and temporal association cortices and the hypothalamus. V, ventral amygdalofugal pathway; ST, stria terminalis. (Adapted from Nolte J. *The Human Brain: An Introduction to Its Functional Anatomy*. 3rd ed. Mosby; 1993:399.)

Another major pathway within the limbic system centers on output from the amygdala (Fig. 1.2–44). Highly processed sensory information, primarily from the association regions of the prefrontal and temporal cortices, projects to the amygdala. Output from the amygdala is conducted through two main pathways (see Fig. 1.2–40). A dorsal route, the *stria terminalis*, accompanies the caudate nucleus in an arch around the temporal lobe and contains axons that project primarily to the septal area and the hypothalamus. The second major output route, the *ventral amygdalofugal pathway*, passes below the lenticular nucleus and contains fibers that terminate in many regions, including the septal area, the hypothalamus, and the medial dorsal thalamic nucleus. As mentioned earlier, the medial dorsal thalamic nucleus projects heavily to prefrontal and some temporal cortical regions, suggesting that this pathway may be important for the integration of emotional and homeostatic information into cognitive circuits.

Both of these pathways reveal how the limbic system is able to integrate the highly processed sensory and cognitive information content of the cerebral cortical circuitry with the hypothalamic pathways that control autonomic and endocrine systems. In addition, the limbic system interacts with components of the basal ganglia system (Fig. 1.2–45). For example, the ventral amygdalofugal pathway also projects to the nucleus accumbens (ventral striatum), the area where the head of the caudate nucleus fuses with the putamen (see Figs. 1.2–27 and 1.2–28). This region sends efferents to the ventral pallidum, an extension of the globus pallidus, which, in turn, projects to the medial dorsal thalamic nucleus. This pathway illustrates that the functions of the basal ganglia extend beyond the regulation of motor activities and shows the necessity of considering the function or dysfunction of particular brain regions in the context of all aspects of their circuitry.

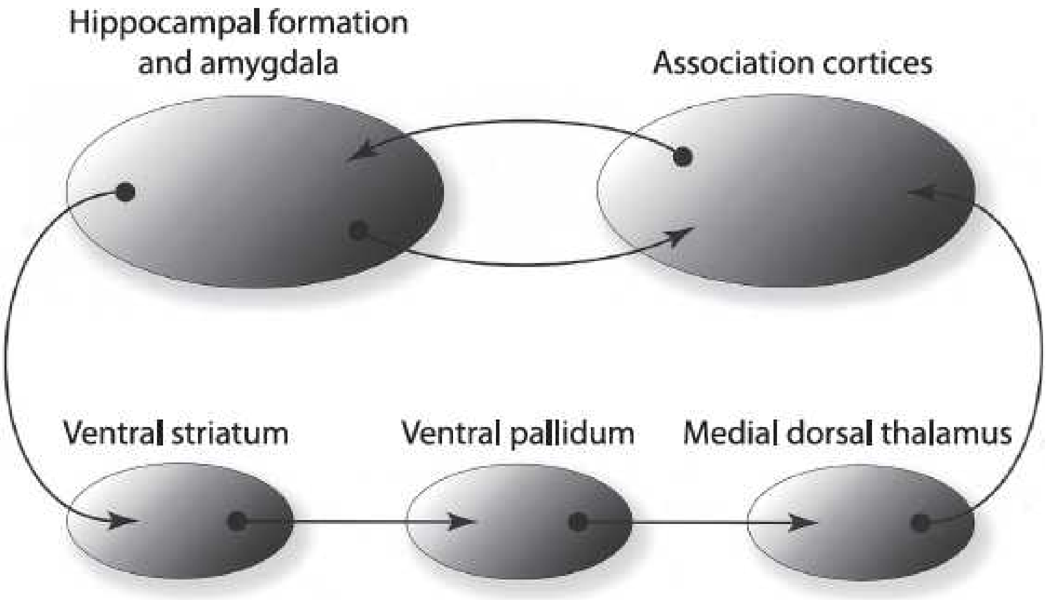

FIGURE 1.2–45. Functional neural circuitry of the limbic system. This drawing illustrates the interaction between the limbic system and certain components of the basal ganglia. (Adapted from Nolte J. *The Human Brain: An Introduction to Its Functional Anatomy*. 3rd ed. Mosby; 1993:412.)

THE BRAINSTEM. As noted earlier, the brainstem is divided into three subdivisions, ranging from rostral to caudal, the *midbrain*, *pons*, and *medulla* (Fig. 1.2–46). The brainstem is a multi-faceted structure and generally has been regarded as dealing with autonomic functions and sensory processing. The reticular nuclei of the brainstem together comprise the reticular formation, which includes all the neurons that are interspersed among the more compact and named nuclei of the brainstem. Numerous studies have revealed that several of the nuclei embedded within the brainstem play an important role in emotional and cognitive functions that are often disrupted in psychiatric illnesses.

Sensory nuclei are located at all three levels of the brainstem, and their associated pathways course from the spinal cord through the brainstem to higher levels, specifically the thalamus and cerebral cortex. The basic sensory representations in the brainstem share a common pathway to the insular cortex and adjacent frontal opercula areas, thus creating the interoceptive pathway. Interoception is an individual’s ability to detect, integrate, and interpret signals from the body, and it is tightly associated with the concept of self. Recent studies have reported reduced interoceptive accuracy in subjects with schizophrenia and increased interoceptive sensibility in subjects with obsessive–compulsive disorder.

The Midbrain. Within the midbrain, the substantia nigra and ventral tegmental area (see Figs. 1.2–26, 1.2–46, and 1.2–47) are of significant relevance to psychiatric disorders, as the neurons in these areas are the major producers of dopamine. The connections and their implications for psychiatric disorders are outlined in the previous section (Basal Ganglia/Major Structures/Substantia nigra).

Other notable nuclei within the midbrain include the *periaqueductal gray* (see Fig. 1.2–46), an area that has historically been associated with the regulation of emotions and experience of pain. The periaqueductal gray surrounds the cerebral aqueduct (see Fig. 1.2–47) and is comprised of four columns that are oriented rostral to caudal; the dorsomedial periaqueductal gray, the dorsolateral periaqueductal gray, the lateral periaqueductal gray and the ventrolateral periaqueductal gray (see Fig. 1.2–47). The two dorsal and two ventral columns are often combined to form dorsal periaqueductal gray and ventral periaqueductal gray, respectively. The dorsal periaqueductal gray is innervated by the amygdala, the ventromedial hypothalamus, the medial prefrontal cortex, and all sensory systems. The dorsal periaqueductal gray innervates the sympathetic nervous system and is thought to facilitate the fight/flight and other active coping responses associated with increased autonomic activation. The ventral periaqueductal gray is targeted by the anterior insula and dorsomedial prefrontal cortices and innervates the parasympathetic nervous system. The ventral periaqueductal gray is thought to mediate passive coping mechanisms.

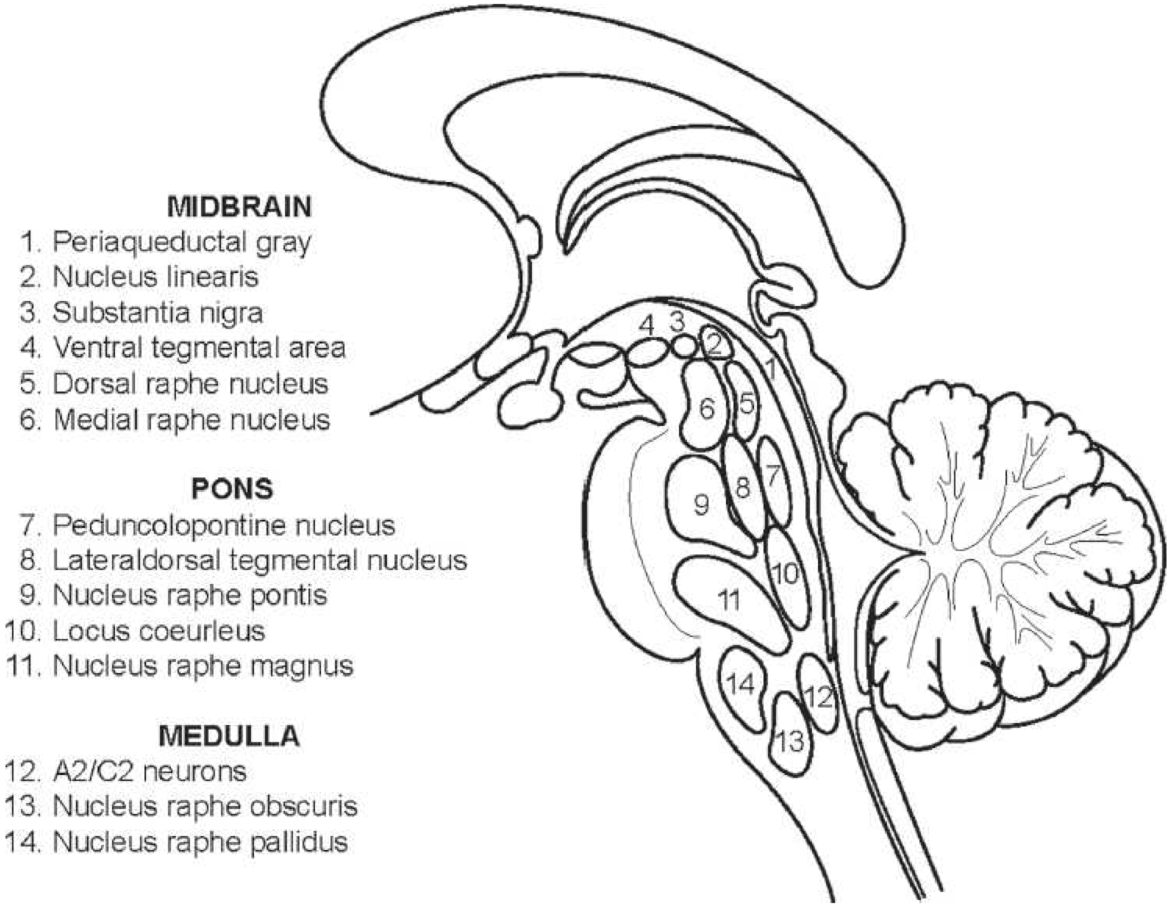
MIDBRAIN
1. Periaqueductal gray
2. Nucleus linearis
3. Substantia nigra
4. Ventral tegmental area
5. Dorsal raphe nucleus
6. Medial raphe nucleus
PONS
7. Peduncolopontine nucleus
8. Lateraldorsal tegmental nucleus
9. Nucleus raphe pontis
10. Locus coeurleus
11. Nucleus raphe magnus
MEDULLA
12. A2/C2 neurons
13. Nucleus raphe obscuris
14. Nucleus raphe pallidus
1
2
3
4
5
6
7
8
9
10
11
12
13
14

FIGURE 1.2–46. Schematic drawing of a midsagittal section of the brain stem indicating the positions of various nuclei implicated in psychiatric disorders. (Adapted from Carpenter MB, Sutin *J. Human Neuroanatomy*. 8th ed. Williams & Wilkins; 1983.)

The periaqueductal gray is often considered a gateway between the spinal cord and brainstem, transmitting sensory and pain sensations, and has been implicated in a variety of disorders, including panic disorder and chronic pain. Recent studies suggest that unchecked amygdala input to the periaqueductal gray (due to a reduction in prefrontal cortical input to the amygdala) results in inappropriate activation of the periaqueductal gray, possibly contributing to the symptoms of posttraumatic stress disorder.

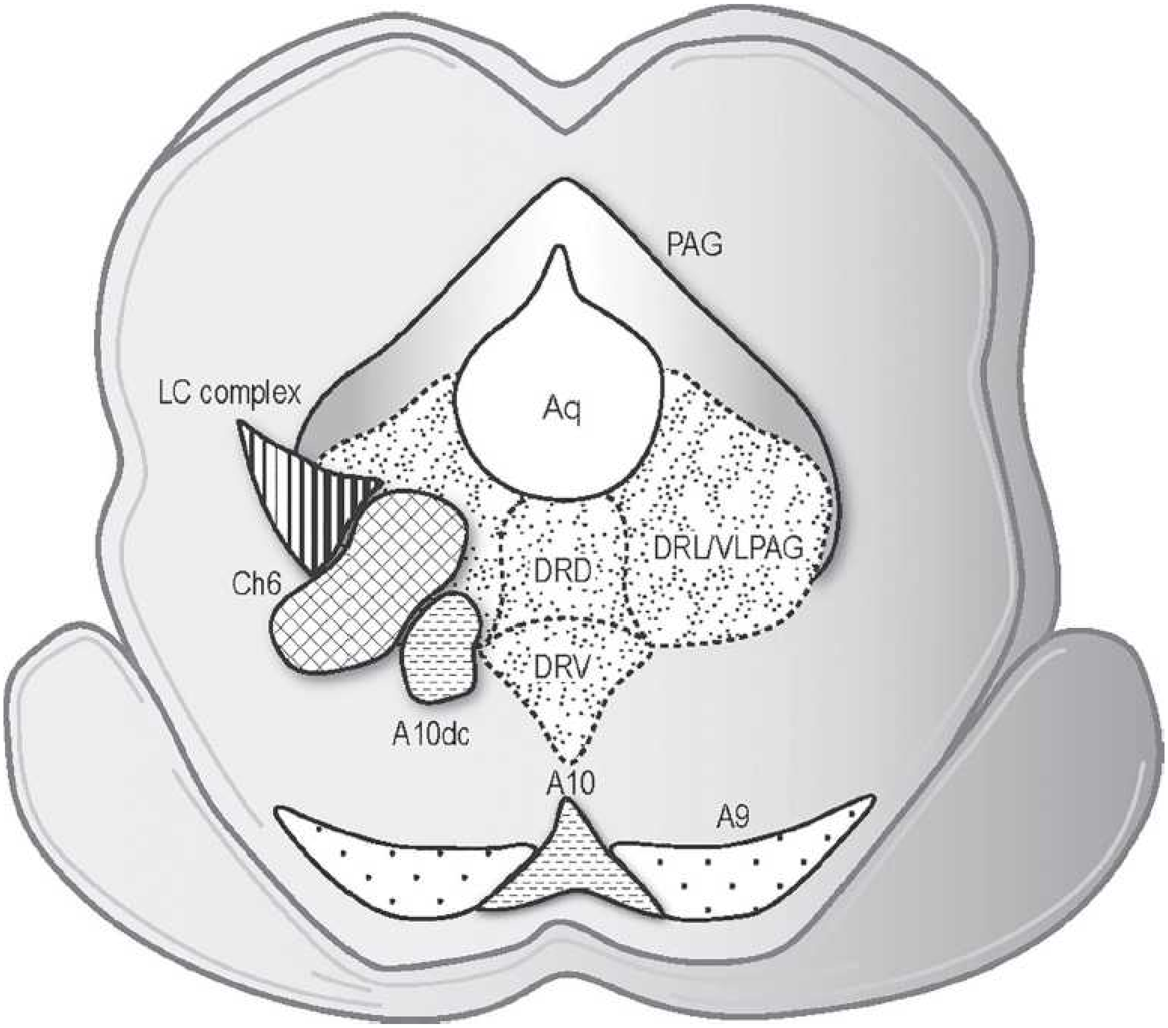

FIGURE 1.2–47. Schematic drawing of the monoaminergic and cholinergic nuclei in a coronal section through the brainstem. In order to include as many relevant nuclei, this section represents nuclei present in both the midbrain and pons. As described in the text, the dorsal raphe nucleus has several subdivisions, all of which can be related to the periaqueductal gray (PAG) and the cerebral aqueduct (Aq). DRD, dorsal raphe nucleus, dorsal subdivision; DRL, dorsal raphe nucleus, lateral subdivision; DRV, dorsal raphe nucleus, ventral subdivision; LDT, lateral dorsal tegmental nucleus; PPN, pedunculopontine nucleus; SNc, substantia nigra pars compacta; VLPAG, ventrolateral periaqueductal nucleus; VTA, ventral tegmental nucleus; VTAdc, ventral tegmental nucleus, dorsocellular subdivision. (Adapted from Ferrucci M, Limanaqui F, Ryskalin L, Biagioni F, Busceti CL, Fornai F. The effects of amphetamine and methamphetamine on the release of norepinephrine, dopamine, and acetylcholine from the brainstem reticular formation. *Front Neuroanat* 2019;13:48. License: https://creativecommons.org/licenses/by/4.0/)

The *raphe nuclei*, which span the length of the brainstem, from the midbrain to the medulla (see Fig. 1.2–46), are the primary source of serotonergic projections within the brain. Seven distinct raphe nuclei within the human brain can be distinguished based on their location within the brainstem. The nuclei can also be divided into a dorsal group and a caudal group, based primarily on their efferent projections. The dorsal group contains the nucleus linearis, the dorsal raphe nucleus, and the medial raphe nucleus.

The nucleus linearis is located along the midline between the two red nuclei, dorsal to the interpeduncular nucleus, and ending at the level of the decussation of the superior cerebellar peduncle. This nucleus is the most rostral raphe nucleus that synthesizes serotonin, although it also contains a variety of chemical-specific cells, including dopaminergic neurons.

At rostral levels, the dorsal raphe is located along the midline, ventral to the periaqueductal gray, and bordered laterally by the oculomotor nucleus and the medial longitudinal fasciculus, and caudally it remains ventral to the periaqueductal gray with the trochlear nucleus and medial longitudinal fasciculus serving as lateral borders. The dorsal raphe can be further subdivided into subnuclei, including the dorsal raphe, which contains the largest population of serotonergic neurons in the human brain, the lateral dorsal raphe, and the ventral dorsal raphe (see Fig. 1.2–47).

The median raphe is situated along the midline and ventral to the dorsal raphe. At some levels, the two nuclei are adjacent to each other. The median raphe extends from the decussation of the superior cerebellar peduncle to the level of the trigeminal motor nucleus and ends in the pons below the medial longitudinal fasciculus.

Numerous studies have revealed that serotonergic neurons within the dorsal raphe and median raphe send projections to the cerebral cortex and other forebrain structures, such as the thalamus, basal ganglia, amygdala, and hippocampus. The projections to the cerebral cortex of these two nuclei are parallel, dense projections and can be distinguished by the morphology of the fibers and their laminar arborization pattern. Briefly, axons from dorsal raphe neurons have small varicosities and are widely distributed in the cerebral cortex and hippocampus, whereas axons from the median raphe have large varicosities and are distributed mainly in the hippocampus and provide a more modest innervation of the cerebral cortex.

The dorsal and median raphe nuclei and their axonal projections are important targets for the treatment of many psychiatric conditions in which mood is affected, as indicated by the effectiveness of selective serotonin reuptake inhibitors.

The Pons. The predominant source of noradrenergic projections arises from the locus coeruleus nucleus, which is embedded within the tegmentum of the brainstem. The locus coeruleus nucleus is situated in the floor of the fourth ventricle and upper dorsolateral pons. Named for the blue pigmentation of the neurons within this nucleus (Latin, *caeruleus* for blue), this small cluster of neurons has widespread projections throughout the cerebral cortex. Although there are other noradrenergic nuclei within the brain, the locus coeruleus nucleus is the major source of projections to the cerebral cortex and thalamus.

Norepinephrine has a widespread and multifunctional impact on brain functioning and can be considered a hormone (both within the CNS and the body) as well as a neurotransmitter. Noradrenergic neurotransmission in the cerebral cortex has strongly been associated with anxiety disorders, including panic disorder and generalized anxiety disorders. However, given the widespread distribution of norepinephrine axons across all areas of the cerebral cortex, it is likely that the noradrenergic neurons of the locus coeruleus nucleus play a role in many psychiatric disorders.

The pedunculopontine and lateral dorsal tegmental nuclei are the primary sources of cholinergic projections to the thalamus as well as to other brainstem areas, including the ventral tegmental area. Most cholinergic projections to the cerebral cortex arise from the basal forebrain, and this circuitry has been the most studied in psychiatric disorders.

The Medulla. Anatomically, the medulla is characterized by the pyramids, white matter tracts that carry information from the precentral, postcentral, and premotor cortices to the spinal cord, forming the corticospinal tract. Medullar nuclei that may play a role in psychiatric disorders include the A2/C2 nuclei and two raphe nuclei; the nucleus raphe obscuris and pallidus. The A2/C2 nuclei are comprised of adrenergic and cholinergic neurons, which send their respective projections to other brainstem areas as well as the thalamus.

IMPLICATIONS FOR BIOLOGICALLY BASED DIAGNOSTIC SYSTEMS

The integrity of the neuroanatomic features described in this chapter can be assessed in individuals with psychiatric disorders at different levels of resolution. Disease-associated changes in neuron number, neuron size, or connections among neurons may be reflected in gross structural alterations detected by in vivo imaging techniques. It remains unclear, however, if the resolution afforded by these imaging techniques would be able to discriminate among different disease processes in a manner that could inform clinical diagnosis. For example, a disease-related difference in the volume of a given brain region could be due to fewer neurons, smaller neurons, or fewer neuronal connections; the same abnormality evident by structural imaging could represent very different underlying disease processes. Although postmortem studies provide the level of resolution needed to distinguish among such possibilities, the diagnostic value of findings from these investigations (as is the case for imaging studies) requires the capacity to distinguish among the following five "C's": (1) *cause*, an upstream factor related to the disease pathogenesis; (2) *consequence*, a deleterious effect of a cause; (3) *compensation*, the brain's response to either a cause or a consequence that helps restore homeostasis; (4) *comorbidity*, a product of factors frequently associated with, but not a part of, the disease process; or (5) *confound*, an artifact of the approach used to obtain the measure of interest. The future incorporation of anatomic data into a diagnostic schema for psychiatric disorders will depend on the ability to make measurements at the appropriate level of resolution and to determine which "C" category a given observation represents.

FURTHER READINGS

Beneyto M, Sibille E, Lewis DA. Human postmortem brain research in mental illness syndromes. In: Charney DS, Nestler EJ, *eds. Neurobiology of Mental Illness*. 3rd ed. Oxford University Press; 2009.

Chen AC, Oathes DS, Chang C, et al. Causal interactions between fronto-parietal central executive and default-mode networks in humans. *Proc Natl Acad Sci U S A*. 2013;110:19944–19949.

Chudasama Y, Robbins TW. Functions of frontostriatal systems in cognition: comparative neuropsychopharmacological studies in rats, monkeys and humans. *Biol Psychol*. 2006;73:19–38.

*Datta D, Arion D, Roman KM, Volk DW, Lewis DA. Altered expression of ARP2/3 complex signaling pathway genes in prefrontal layer 3 pyramidal cells in schizophrenia. *Am J Psych*. 2017;174:163–171.

DeFelipe J. The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Front Neuroanat*. 2011;5:29.

DeLong MR, Wichmann T. Circuits and circuit disorders of the basal ganglia. *Arch Neurol*. 2007;64:20–24.

Eggan SM, Lewis DA. Immunocytochemical distribution of the cannabinoid CB1 receptor in the primate neocortex: a regional and laminar analysis. *Cereb Cortex*. 2007;17:175–191.

*Fuster JM. *The Prefrontal Cortex*. 4th ed. Academic Press; 2008.

Glausier JR, Lewis DA. Dendritic spine pathology in schizophrenia. *Neurosci*. 2013;251:90–107.

Gonzalez-Burgos G, Cho RY, Lewis DA. Alterations in cortical network oscillations and parvalbumin neurons in schizophrenia. *Biol Psychiatry*. 2015;77:1031–1040.

Haines DE. *Fundamental Neuroscience for Basic and Clinical Applications*. 4th ed. Churchill Livingstone; 2006.

Hamami C, Mayberg H, Stone S, Laxton A, Haber S, Lozano AM. The subcallosal cingulate gyrus in the context of major depression. *Biol Psychiatry*. 2011;69:301–308.

Heimer L, Van Hoesen GW. The limbic lobe and its output channels: implications for emotional functions and adaptive behavior. *Neurosci Biobehav Rev*. 2006;30:126–147.

*Lewis DA, Curley AA, Glausier JR, Volk DW. Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. *Trends Neurosci*. 2012;35:57–67.

Lewis DA, Melchitzky DS, Gonzalez-Burgos G. Specificity in the functional architecture of primate prefrontal cortex. *J Neurocytol*. 2002;31:265276.

Mäki-Marttunen V, Andreassen OA, Espeseth T. The role of norepinephrine in the pathophysiology of schizophrenia. *Neurosci Biobehav Rev*. 2020;118:298–314.

*Nolte J, Sundsten JW. *The Human Brain: An Introduction to Its Functional Anatomy*. 6th ed. Mosby; 2008.

Patestas MA, Gartner LP. *A Textbook of Neuroanatomy*. Blackwell; 2006.

*Peters SK, Dunlop K, Downar J. Cortico-striatal-thalamic loop circuits of the salience network: a central pathway in psychiatric disease and treatment. *Front Syst Neurosci*. 2016;10:104.

Petrides M. Lateral prefrontal cortex: architectonic and functional organization. *Philos Trans R Soc Lond B Biol Sci*. 2005;360:781–95.

Ramnani N. The primate cortico-cerebellar system: anatomy and function. *Nat Rev Neurosci*. 2006;7:511–522.

Sillito AM, Cudeiro J, Jones HE. Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends Neurosci*. 2006;29:307–316.

*Sherman SM, Guillery RW. *Functional Connections of Cortical Areas: A New View from the Thalamus*. MIT Press; 2013.

Squire LR, Bloom FE, McConnell SK, et al. *Fundamental Neuroscience*. 4th ed. Academic Press; 2012.

Steriade M. Grouping of brain rhythms in corticothalamic systems. *Neuroscience*. 2007;137:1087–1106.

The Petilla Interneuron Nomenclature Group (PING). New insights into the class and nomenclature of cortical GABAergic interneurons. *Nat Rev Neurosci*. 2013;14:202.568.

Tsujino N, Sakuri T. Orexin/Hypocretin: a neuropeptide at the interface of sleep, energy homeostasis, and reward system. *Pharmacol Rev.* 2009;61:162–176.

*Venkatraman A, Edlow BL, Immordino-Yan MH. The brainstem in emotion: a review. *Front Neuroanat*. 2017;11:15.

Zhang F-F, Peng W, Sweeney JA, Jia Z-Y, Gong Q-Y. Brain structure alterations in depression: psychoradiological evidence. *CNS Neurosci Ther*. 2018;24:994–1003.

1.3 Functional Genomics of Human Brain Development

Kartik Pattabiraman, M.D., Ph.D., Sirisha Pochareddy, PH.D., John C. Silbereis, PH.D., John N. Constantino, M.D., and Nenad Sestan, M.D., PH.D.

The human central nervous system (CNS) comprises approximately 86 billion neurons connected by hundreds of thousands of kilometers of myelinated axons and several hundred trillion to well over a quadrillion synapses, along with a roughly equal number of glial cells. Reflecting this complexity, building the human CNS requires the precise regulation of molecular and cellular processes. Through this regulation, intrinsic processes encoded in the human genome and extrinsic modulators from the environment converge in development to shape the circuitry, anatomical features, synaptic interactions, and cellular processes that carry out neural function. The greater size and complexity of the human brain are, of course, foundational to the expanded intellectual, cognitive, and emotional capacities of humans. As a result of its size and complexity, the human brain as a whole, but most especially the association areas of the cerebral neocortex, takes relatively more time than the brains of other mammals to develop, and humans have a particularly long gestational time, as well as childhood and adolescence (Fig. 1.3–1). This prolonged developmental course and period of dependency allows, more so than in other species, environmental factors to shape the development of cognitive, emotional, and social capabilities and capacities. Indicative of the biologic challenge of precisely regulating diverse molecular and cellular processes over a protracted period of time and across myriad cell types and regions, the CNS exhibits regionally and temporally distinct patterns of vulnerability to various diseases and insults (see Fig. 1.3–1).

Thus, the many neurodevelopmental processes that enable human brain complexity seem to have required an evolutionary trade-off that rendered it especially susceptible to certain disorders.

Recent technologic advances and increased focus on human neurodevelopment have enabled a more comprehensive characterization of the human CNS and its development in both health and disease. In particular, "-omic"-level characterizations of the transcriptional (e.g., gene expression properties), gene regulatory, and epigenetic processes of the developing human CNS across developmental periods and brain regions combined with comparative studies in non-human primates and rodents have generated new insights into the molecular landscapes that define human CNS development and their dysfunction in disease. Furthermore, the genomic revolution has considerably advanced knowledge about the genetic variations that confer risk for various neurologic and psychiatric disorders, confirming that psychiatric disease is, by and large, genetically complex and etiologically diverse.

This chapter details the current understanding of the transcriptional and regulatory architecture of the developing human neocortex and other regions of the CNS. It further provides examples of how a broad dysfunction of these processes likely underlies neurodevelopmental and psychiatric disease. Emphasis will be given to studies of the neocortex because of its importance in higher-order cognition, emotional regulation, complex behaviors, and the pathobiology of many neurodevelopmental and psychiatric disorders. Another chapter of this book, titled "Neurodevelopment and Neurogenesis," describes the basic principles of the organization and development of the CNS, and it is recommended as companion reading.

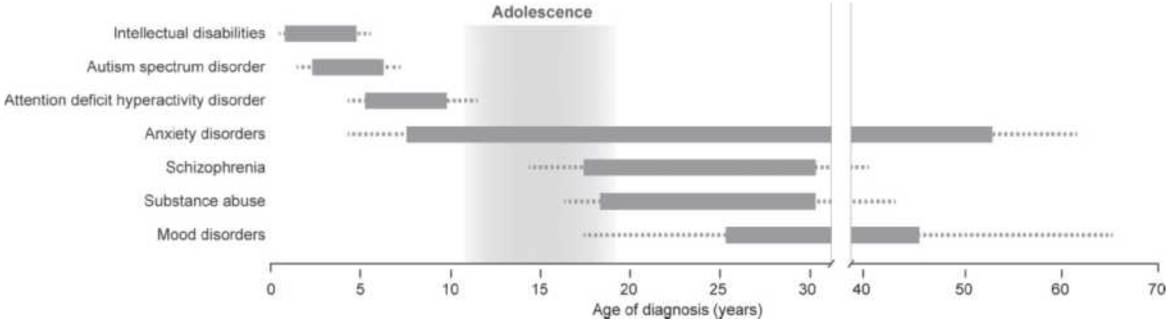

FIGURE 1.3–1. Psychiatric and neurodevelopmental disorders have discrete ages of onset. An illustration detailing the ages of onset of major psychiatric and neurodevelopmental disorders. The bars indicate the age range that each disorder commonly affects, with less frequent ages of diagnosis denoted as *dotted lines*. Note that the age of diagnosis is highly variable between brain disorders. Many psychiatric disorders emerge in adolescence and early adulthood, while neurodevelopmental disorders emerge earlier. This variability is indicative of dysregulation of tightly regulated developmental processes. (Based in part on data from Kessler RC, Amminger GP, Aguilar-Gaxiola S, et al. Age of onset of mental disorders: a review of recent literature. *Curr Opin Psychiatry*. 2007;20:359–364; Silbereis J, Pochareddy S, Zhu Y, Li M, Sestan N. The cellular and molecular landscapes of the developing human central nervous system. *Neuron*. 2016;89:248–268.)

THE TRANSCRIPTIONAL LANDSCAPE OF THE DEVELOPING HUMAN CNS

At the root of the cell and molecular processes governing human brain development is the immensely complex genome (see section “Organization of the Human Nuclear Genome”). Transcription refers to the conversion of genetic information encoded in DNA to mRNA; mRNA is then translated into the proteins that ultimately carry out the common and unique functions of specific cell types. Therefore, the selection of genes for transcription and the determination of gene expression levels are largely responsible for establishing the molecular and physiologic properties of distinct cell types and brain regions. Transcription is therefore tightly governed by a diverse array of regulatory mechanisms. The development of high-throughput technologies (see section “High-Throughput Techniques to Study Molecular Landscapes”) has greatly advanced the ability to explore the transcriptome in the developing human CNS in order to understand the diverse cell and molecular phenotypes that underlie CNS function and disease processes.

Organization of the Human Nuclear Genome

The human nuclear genome comprises approximately 6.2 (in males) or 6.4 (in females) billion base pairs distributed across 46 chromosomes and provides the blueprint for building the body and its function. Current GENCODE estimates (/www.gencodegenes.org/stats/current.html; version 39) predict 61,533 genes. Of these, 19,982 are protein-coding genes, 26,378 are noncoding RNA (long and short) genes and the remaining are pseudogenes. Furthermore, recent studies have found that greater than 90% of human protein-coding genes have splice variants, with the human brain showing the greatest splicing complexity. A significant part of the remaining genome is believed to be *cis*-regulatory elements (CREs), the DNA sequences that regulate transcription. According to some estimates, the human nuclear genome harbors approximately 400,000 putative enhancers and 70,000 promoters. Access to CREs by transcription factors and other components of the transcriptional machinery is regulated by epigenetic modification of both DNA and chromatin. Most notably, methylation of DNA generally represses gene expression by blocking DNA binding, while acetylation of histones promotes transcription by opening up the chromatin to make DNA more accessible. This highlights both the complexity of the human genome and the challenges in elucidating the molecular mechanisms underlying complex processes like neurodevelopment.

High-Throughput Techniques to Study Molecular Landscapes

Several major techniques have been applied over the past decade to study the human genome and its functions at a global scale. Many of these techniques rely on either hybridization or deep sequencing to identify and quantify the molecular species (RNA, DNA) tested. Some of the popular high-throughput techniques applied to investigate gene expression and its regulation at a genome-wide scale are discussed below:

Exon arrays—A microarray platform to quantify RNA expression at the exon level through hybridization to sequences of exons for known genes.

RNA-seq—Next-generation sequencing of RNA to identify and quantify the various RNA species, including coding and noncoding RNAs in a given sample. Allows unbiased detection of transcripts, thus leading to the identification of novel transcripts. This technique can be performed both at the tissue and single-cell levels. Single-cell level analysis has been especially important in studying the brain, given the numerous transcriptomically and functionally distinct cell types.

ATAC-seq—Assay for Transposase-Accessible Chromatin (ATAC) followed by next-generation sequencing identifies regions of genomic DNA that are defined as “open” chromatin, which are regions that can be bound by transcriptional machinery. Allows for unbiased detection of putative active CREs and gene promoters. This technique can be performed at the tissue and single-cell level.

ChIP-seq—Chromatin immunoprecipitation (ChIP) followed by next-generation sequencing identifies regions of genomic DNA at which a particular protein of interest is interacting. Typically, DNA interacting with the protein of interest (e.g., transcription factor or histone mark) is precipitated along with the protein by using an antibody specific to the protein of interest. The precipitated DNA is then sequenced to identify the genomic sequences. This technique has been applied to identify genomic regions enriched for certain histone modifications or bound by specific transcription factors.

Methylation arrays—Microarrays that quantify DNA methylation levels at an array of regulatory sites by selective hybridization of methylated DNA that utilizes a reaction with bisulfate that marks methylated but not unmethylated DNA by converting methylated cytosine to uracil.

WGBS—Whole genome bisulfite sequencing is a sequencing-based approach to determine DNA methylation levels on a genome-wide scale.

Spatial–Temporal Transcriptional Dynamics of Human Brain Development

A number of recent genome-wide profiling studies of the developmental transcriptome (see Silbereis et al. for a comprehensive review) have produced comprehensive datasets describing coding and noncoding RNAs across multiple time points and regions of the human brain. These analyses have shed new light on the transcriptional architecture of brain regions and neocortical areas, neurodevelopmental processes, and the underlying biology of complex neurodevelopmental disorders. This work indicates that most protein-coding genes (at least 86% according to Kang et al., 2011) and a growing number of noncoding genes (see section

"Organization of the Human Nuclear Genome") are involved in building the human CNS at some point in development. Of the protein-coding genes expressed in the brain, about 90% were either differentially expressed or spliced across brain regions and/or time. Such differences in the transcriptional architecture of brain regions and neocortical areas reflect their distinct ontologies, cellular compositions, biologic processes, and developmental timing (Fig. 1.3–2).

Another common finding of these studies is that transcriptomes differ more prominently across time and space than they do between sexes, ethnicities, or individuals, despite the underlying genetic differences. In particular, the majority of the temporal variance occurs during two phases: an early fetal transition (postconception weeks 6 to 10) and a late fetal transition between midfetal development and infancy. In contrast, expression changes over several decades of adulthood are relatively minute (Fig. 1.3–2A). When compared with dorsal pallial structures (amygdala, hippocampus, and neocortex), the cerebellum is the most transcriptionally distinctive region in the developing and adult brain, followed by the thalamus and striatum (Fig. 1.3–2B). Transcriptional differences in primordial areas of the developing neocortex are less robust than those between the neocortex and the other analyzed brain regions. A majority of these spatially enriched genes are also temporally dynamic; notably, some are transiently enriched during a specific time window, including protracted epochs of postnatal development such as late childhood and adolescence. Single-cell transcriptomic and epigenetic analyses in fetal tissues have similarly identified complex dynamic expression across individual neuronal and nonneuronal populations, including differential maturation rates. These findings reveal that regional transcriptomes are developmentally regulated and may reflect cellular and functional differences.

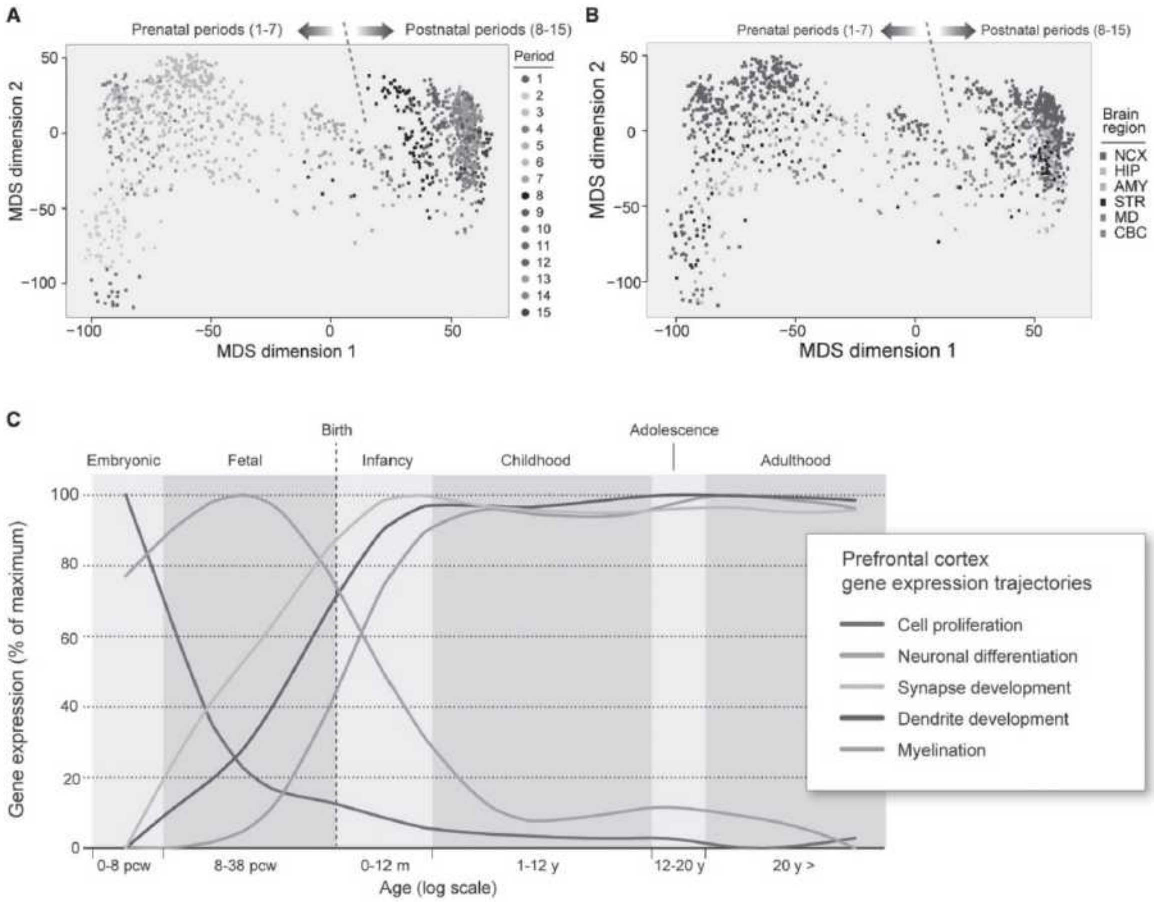

FIGURE 1.3–2. Global spatiotemporal dynamics of the human brain transcriptome. **A:** This panel shows a multidimensional scaling (MDS) plot that describes gene expression differences between developmental periods. MDS is statistical/graphical technique that reduces the complexity of data by plotting individual data points apart from one another at distances corresponding to the degree to which they are different. In the plot, each dot represents a sample and is colored according to period as defined by Kang et al., 2011; the *dotted line* indicates birth. Note that the most pronounced differences (approximately two-thirds) occur during prenatal development (periods 1 to 7). By contrast, over four decades of adulthood (periods 13 to 15), less than 1% of genes are differentially expressed. **B:** Here the same MDS plot as in (**A**) is shown except that the new data points are colored by brain region. Note that the most prominent transcriptional differences between regions are between cerebellar cortex (CBC) and forebrain regions. NCX, neocortex; HIP, hippocampus; AMY, amygdala; STR, striatum; MD, mediodorsal nucleus of thalamus. **C:** A line graph illustrating the developmental trajectory of expression of genes important for major neurodevelopmental processes demonstrating how gene expression profiles are linked to the cellular processes that take place at different periods of development. The y-axis represents the percentage of gene expression relative to maximum and the x-axis corresponds to age. (See eBook for color images.)

Transcriptional studies have also demonstrated correlations between gene expression dynamics and the morphologic and functional development of brain regions and consequently shed light on the timing of developmental processes and the onset of specific biologic functions (Fig. 1.3–2C). For example, gene coexpression network analyses have revealed that the developmental brain transcriptome is segregated into distinct modules or clusters of genes with highly correlated expression. The expression trajectories of such modules often exhibit dynamic patterns across development, indicating that they play roles in distinct and dynamic developmental processes. For example, in a study by Kang et al., 2011, the two largest modules were composed of genes that are coregulated across brain regions and time, indicating that many transcriptional processes are shared and coordinated across different regions as the brain develops. Many modules could be defined by functional ontologies, as they were highly enriched for genes important for specific biologic processes.

Looking broadly at the expression levels of genes implicated in major developmental processes by functional ontology analysis sheds light on the molecular dynamics of human brain maturation and the timing of key neurodevelopmental events across areas. For instance, in the prefrontal cortex, genes promoting cell proliferation are most highly expressed embryonically, and expression declines markedly during the fetal period; genes associated with neuronal differentiation peak in the fetal period and are downregulated in infancy; and genes important for myelination rise markedly during perinatal development, plateauing around 1 year of age (Fig. 1.3–2C). Taken together, transcriptomic analyses of human brain development across the spatial, temporal, and cellular dimensions have provided novel insights into normal neurodevelopmental processes and the interpretation of disease-associated mutations and underlying neuropathology, as discussed later in this chapter.

Transcriptional Organization of the Fetal Neocortex

These gene expression differences play essential roles in establishing the complex anatomical, circuit, and functional organization of the CNS, as can be exemplified in the development of the cerebral neocortex. At least two types of spatial information must be encoded in nascent projection neurons of the cortical plate (CP). First, their position in the radial direction, corresponding to their laminar position, with distinct cell types, synaptic connectivities, and functions within each lamina or layer; and second, their position in the tangential plane, corresponding to their particular cortical region/area identity. The human cerebral neocortex exhibits the basic organizing principle common to all mammals; however, there are human-enriched and defining features that are largely responsible for differences in cognition, emotion, behavior, and manifestations of disease.

Most of the neurons of the cortex are generated from the late embryonic period to the late fetal period, as is detailed in depth in chapter Neurodevelopment and Neurogenesis. During this period, the neocortical wall becomes segregated into laminar zones, each composed of characteristic cell compositions carrying out discrete neurodevelopmental processes. These lamina exhibits highly distinct gene expression patterns that define their biologic roles. For instance, in early and midfetal development genes associated with proliferation and stem cell identity are enriched in the ventricular and subventricular zones (VZ and SVZ), which are germinal lamina overlying the ventricles. In contrast, the CP, which is the primordium for the cerebral cortex, is characterized by high expression of genes promoting neuronal differentiation and synaptogenesis. Recent single-cell level studies are beginning to reveal greater heterogeneity in both progenitors and neurons in the VZ, SVZ, and CP, including the outer subventricular zone (OSVZ), which is greatly expanded in the human cerebral cortex.

The CP will ultimately form the mature neocortex, which is a six-layered structure with the upper layers (e.g., through 4) sending intracerebral projections and the lower layers 5 and 6 sending projections to the sub cerebrum, brain stem, spinal cord, and other cortical areas, as well as, in the case of layer 6, reciprocal connections to the thalamus. Each of these layers is also characterized by distinct gene expression patterns, which are important for conveying the unique projection identities, synaptic receptivities, and functional properties within them. Defining these layers requires the establishment of spatially restricted gene expression patterns through transcriptional regulation, which shall be described later in this chapter.

Transcriptional Patterns Define Neocortical Regions and Areas in Development

While the neocortex has a highly specialized laminar organization, its organization into functionally discrete regions and areas along the tangential dimension is arguably far more complex, especially in humans. Work in mice shows this is governed by graded expression of transcription factors during early embryonic development, followed by extrinsic signaling from thalamic axonal inputs. Studies of the human transcriptome have revealed substantial regional/areal gene expression differences in the developing human neocortex (Fig. 1.3–3). Indeed, each region/area exhibits discrete temporally specified transcriptional properties that are likely reflective of underlying biologic processes involved in the patterning and specialization of neural circuits encompassing particular regions and areas (Fig. 1.3–3A). Interestingly, these regional/areal transcriptional profiles were largely symmetric between hemispheres at the population level from fetal development through adulthood. This is surprising given the functionally distinct roles and different developmental trajectories of the left and right hemispheres, which arise in the late midfetal period and become more prominent during early postnatal periods. These findings, therefore, suggest that transcriptional neocortical left-right asymmetry may be too subtle for detection by tissue-level transcriptomic methods or that nontranscriptional mechanisms, such as neural activity, play more critical roles.

Interareal transcriptional differences were found to be most robust during midfetal development (postconception weeks 15 to 24), a crucial period for specification and the initial assembly of neocortical neural circuits (Fig. 1.3–3B,C). During these periods, there were prominent and specific transcriptional signatures associated with prospective prefrontal and perisylvian areas, which are involved in some of the most distinctly human aspects of cognition and behavior. In addition, prefrontal/frontal-enriched gradients of expression along the anteroposterior axis of the CP, as well as gradients with enrichment in temporal, occipital, occipitotemporal, perisylvian, and ventromedial areas, were observed. Recent transcriptomic analysis of the human midfetal cortex identified a novel enrichment of genes associated with retinoic acid signaling in the frontal lobe established by a prefrontal cortex-enriched rostrocaudal gradient of retinoic acid. Disruption of the gradient of retinoic acid in the developing mouse cortex led to disruption of prefrontal cortex specification and connectivity.

Strong early and midfetal interareal transcriptional differences and gradients diminish perinatally and increase again after adolescence (Fig. 1.3–3B,C). By contrast, in late fetal periods and early postnatal development, more general molecular programs governing glial development, synaptogenesis, synaptic plasticity, and dendritic maturation can predominate. The progressive increase in regional differences that emerge in adolescence could be reflective of variations in the maturational trajectories between certain neocortical areas (e.g., associative vs. primary areas) and may be important for late cognitive, emotional, and socio-behavioral development.

Taken together, these findings indicate that, similar to other species (Brain Initiative Cell Census Network, 2021), the generation and differentiation of neocortical cell types and their assembly into functional neural circuits in humans are achieved through precise regulation of spatiotemporal gene expression. Psychiatric disorders and neurodevelopmental disorders are associated with both selective dysfunctions of specific brain regions and neocortical areas and abnormal functional and anatomical connectivity between them. It is, therefore, likely that distinct etiologies of psychiatric disorders can be defined by abnormal gene expression at specific times and places in the developing brain.

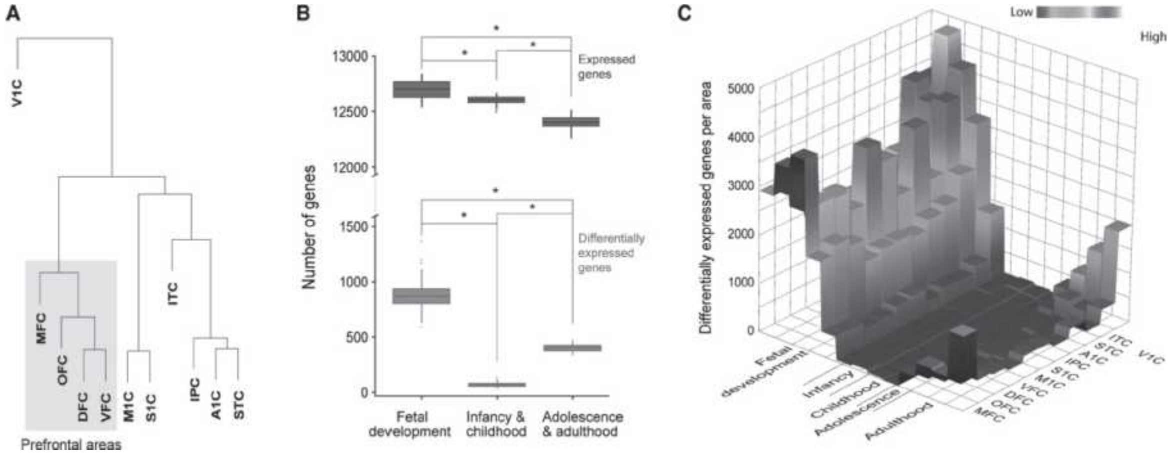

FIGURE 1.3–3. Transcriptional differences among human neocortical areas are temporally regulated. **A:** Illustration of the relative transcriptional differences between brain areas in an unsupervised hierarchical clustering chart that arranges the 11 neocortical regions/areas profiled by Pletikos et al., 2014, based on the similarity of the transcriptome of each area from early fetal development throughout adulthood. OFC, orbital prefrontal cortex; DFC, dorsal prefrontal cortex; VFC, ventral prefrontal cortex; MFC, medial prefrontal cortex; M1C, primary motor cortex; S1C, primary somatosensory cortex; IPC, posterior inferior parietal cortex; A1C, primary auditory cortex; STC, superior temporal cortex; ITC, inferior temporal cortex; V1C, primary visual cortex. **B:** Boxplots showing the number of expressed (*dark gray*, top panel) and differentially expressed (*light gray*, bottom panel) genes among neocortical areas across fetal development (periods 3 to 7), infancy (periods 8 and 9), childhood (periods 10 and 11), adolescence (period 12), and adulthood (periods 13 to 15). Note that the total number of genes expressed decreases over development. However, the number of differentially expressed genes observed over development exhibits a temporal hourglass pattern with the most substantial differential expression occurring in fetal development, a marked decline in infancy and childhood, and an increase in adolescence through adulthood. **C:** A three-dimensional Manhattan plot showing the number (post hoc Tukey test) of genes with differential expression between any two neocortical areas, demonstrating that the hourglass pattern of interareal differential expression persists in all neocortical areas, but is most prominent in MFC and V1C. Differential expression corresponds to the y-axis, the x-axis delineates developmental periods, and the z-axis delineates brain region. (Reprinted from Pletikos M, Sousa AM, Sedmak G, et al. Temporal specification and bilaterality of human neocortical topographic gene expression. *Neuron* 2014;81(2):321–332. doi: 10.1016/j.neuron.2013.11.018 Copyright © 2014, with permission from Elsevier.)

THE REGULATORY LANDSCAPE OF THE DEVELOPING HUMAN CNS

The spatiotemporal patterns of gene expression discussed above are established by elaborate regulatory mechanisms that govern each step of transcription. The current understanding of the general properties of transcriptional regulation of the developing human CNS is discussed below.

Regulation of Gene Transcription

A gene is transcribed when core transcriptional machinery, including RNA polymerase and other proteins, is recruited to the gene's proximal core promoter, a minimal region of contiguous DNA sequence that is sufficient to accurately initiate a basal level of gene expression. But the tissue specificity, timing, and level of a gene's expression are greatly controlled by distal DNA sequences like enhancers (sequences that increase expression), silencers (sequences that repress expression), and insulators (sequences that act as barriers), which along with core promoters are referred to as "cis-regulatory elements" (CREs) for their ability to mediate transcriptional activity in cis (i.e., proximal regions of the genome). Enhancers are believed to activate transcription by recruiting important accessory proteins to the promoter. They have binding sites for multiple transacting regulatory proteins, primarily transcription factors, and can function on their target promoters from locations as remote as another chromosome. Bound transcription factors recruit other transcription factors or chromatin-modifying enzymes and interact with mediator complexes. One model of activation proposes that enhancer-bound transcription factors, along with mediator complex proteins, loop the intervening DNA and bring enhancers and core promoters in close proximity, leading to the activation of gene expression. Like enhancers, silencers or repressors recruit proteins, but these proteins inhibit transcription. Silencer proteins act by inhibiting transcription initiation or elongation or by sequestering activator proteins or altering chromatin state. Insulators also bind proteins, and these DNA–protein complexes act as barriers in the genome, blocking the action of an enhancer/silencer on its noncognate promoter or preventing the spread of heterochromatin.

cis-Regulation of Transcription in Neocortical Development

Comparative genomic studies, which assess sequence conservation across different organisms, have shown that many of the regulatory networks involved in neurodevelopment are highly conserved across mammals. For instance, the regulation of *FEZF2*, a gene that is necessary for corticospinal (CS) tract development and that has remarkably similar neocortical expression patterns between mouse and human, are highly similar in these species. Studies of *Fezf2* regulation in mouse neocortical development exemplify key concepts that describe developmental regulatory networks. A conserved enhancer (E4), which lies downstream of the Fezf2 transcription start site, was identified to regulate *Fezf2* expression in the developing neocortical plate and is essential for the specification of the CS tract (see Fig. 1.3–4A). *Fezf2* expression is highly enriched in the layer projection neurons of the cerebral cortex that give rise to the CS tract. This pattern suggests that the expression has to be activated in L5 and inhibited in L6 and other layers of the cortex. The gene regulatory network that accomplishes this selective expression has recently been defined (see Fig. 1.3–4A) and provides a model of spatial control of gene expression. Two related transcription factors of the SOX family, SOX4 and SOX11, were shown to be essential for FEZF2 expression by binding and activating the E4 enhancer in the cortex. SOX5, another member of the SOX family of transcription factors, also binds E4 but represses FEZF2 expression (see Fig. 1.3–4A). It was observed that SOX4 and SOX11 functionally compete with SOX5 to determine if *Fezf2* transcription is activated or repressed. The spatiotemporal dynamics of neocortical FEZF2 expression are further controlled by TBR1, another transcription factor expressed in L6 neurons. TBR1 binds to a different enhancer in a highly conserved region downstream of the *Fezf2* transcription start site and represses the expression of FEZF2 in L6. This process is indicative of a common principle of gene regulatory networks in which competitive binding of transcriptional activators and repressors in key regulatory elements form gradients that determine the areas and levels of gene expression.

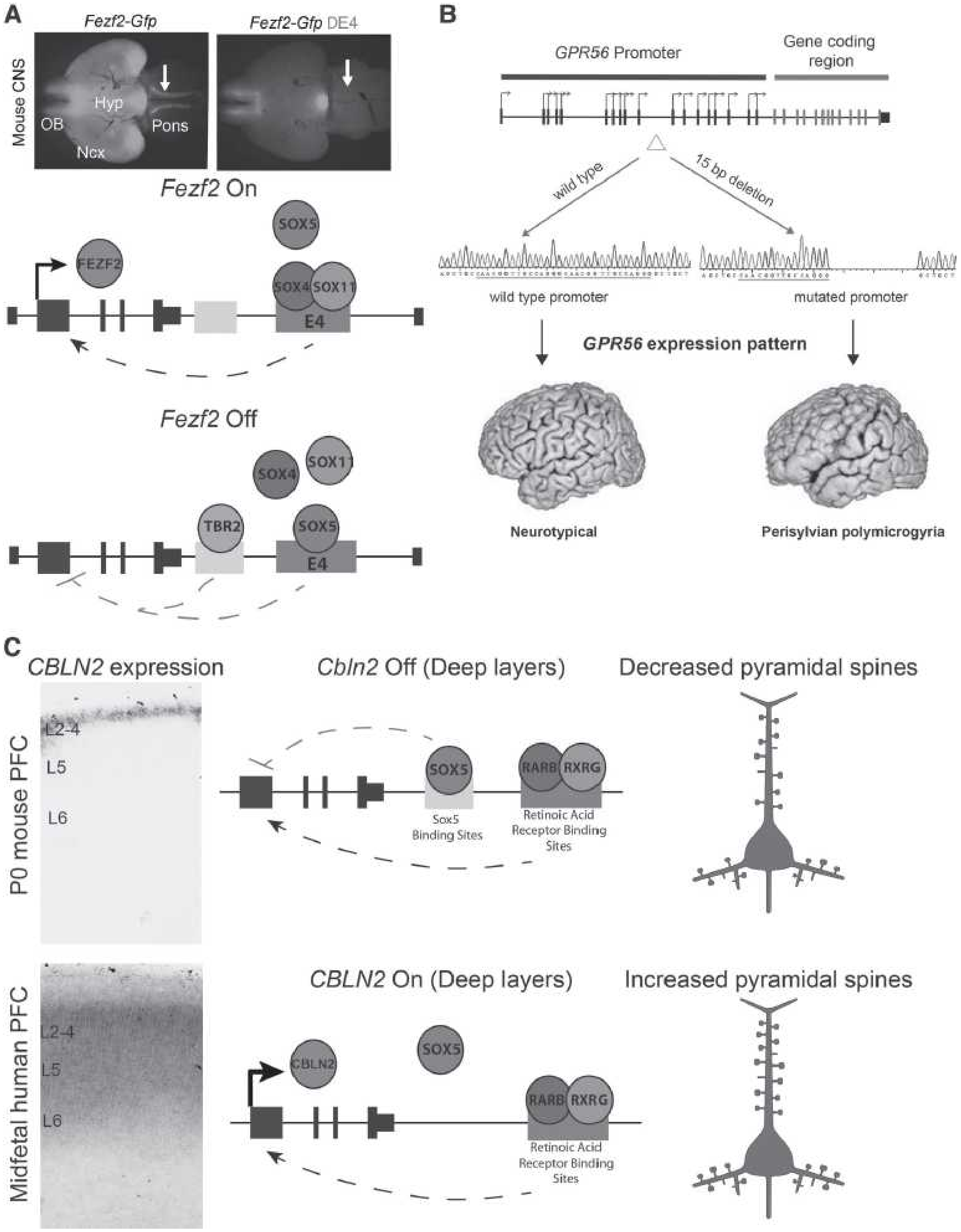
A
Fezf2-Gfp
Fezf2-Gfp DE4
Mouse CNS
OB
Hyp
Ncx
Pons
Fezf2 On
FEZF2
SOX5
SOX4
SOX11
E4
Fezf2 Off
SOX4
SOX11
TBR2
SOX5
E4
B
GPR56 Promoter
Gene coding region
wild type
15 bp deletion
wild type promoter
mutated promoter
GPR56 expression pattern
Neurotypical
Perisylvian polymicrogyria
C
CBLN2 expression
P0 mouse PFC
L2-4
L5
L6
Cbln2 Off (Deep layers)
SOX5
Sox5 Binding Sites
RARB
RXRG
Retinoic Acid Receptor Binding Sites
Decreased pyramidal spines
Midfetal human PFC
L2-4
L5
L6
CBLN2 On (Deep layers)
CBLN2
SOX5
RARB
RXRG
Retinoic Acid Receptor Binding Sites
Increased pyramidal spines

FIGURE 1.3–4. **A.** Deletion of the conserved E4 enhancer which regulates transcription factor, FEZF2, leads to loss of the corticospinal tract (*white arrow*) in mouse (top). Expression of FEZF2 is restricted to layer 6 of the cortex (Fezf2 On), where the corticospinal tract originates, through the complex interplay the transcription factors SOX4, SOX5, SOX11, and TBR1. **B.** A 15 base pair deletion in the promoter of the gene GPR56 was observed

in patients with perisylvian polymicrogyria. In these patients, this mutation likely led to malformations of the perisylvian neocortex important for language among other functions. (Adapted with permission from Bae BI, Tietjen I, Atabay KD, et al. Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. *Science* 2014; 343(6172):764–768. doi:10.1126/science.1244392.) **C.** CBLN2 expression is expanded into deep layers (L) 5 and 6 in the human prefrontal cortex (PFC) compared to mouse (right). This layer specific expansion CBLN2 is likely due to hominini-specific deletion, which contains repressive SOX5 bindings sites, in the otherwise conserved CBLN2 PFC enhancer (middle). Targeted replacement of the mouse Cbln2 enhancer with the human CBLN2 enhancer led to a stable increase in dendritic spines in pyramidal neurons in PFC, which is well-defined specialization of the human PFC. (Adapted from Shibata M, Pattabiraman K, Muchnik SK, Kaur N, Morozov YM, Cheng X, Waxman SG, Sestan N. Hominini-specific regulation of CBLN2 increases prefrontal synaptogenesis. *Nature* 2021;598:489–494. doi: 10.1038/s41586-021-03952-y.)

Recent studies have also identified human and primate-specific aspects of gene regulatory networks, which are likely essential for the unique properties of the human and primate cortex and possibly advanced human cognition. For example, the synaptic organizer, CBLN2, shows species-specific expression differences in the developing prefrontal cortex. In the midfetal human prefrontal cortex, *CBLN2* is expressed in all layers of the CP, while *Cbln2* expression is restricted to the upper layer of the mouse prefrontal cortex at developmental stages equivalent to human midfetal development. This difference in expression is due to a hominini (human, chimpanzee, and bonobo)-specific deletions in a prefrontal cortex enhancer of CBLN2 (Fig. 1.3–4C). These deleted regions contain binding sites for the repressive transcription factor, SOX5, leading to the de-repression of CBLN2 expression in the human midfetal prefrontal cortex. In situ genetic humanization of the mouse CBLN2 enhancer led to the human-like expression of CBLN2 and increased dendritic spines on PFC pyramidal neurons. Of note, there is a disproportionately higher number of spines on pyramidal neurons in the human PFC compared to the PFC of other analyzed primates.

Transcriptional Regulation of Human Neurodevelopment

Although studies of individual CREs in model organisms have provided great advances in the current understanding of neurodevelopment, it is also highly evident that a comprehensive understanding of gene regulation in human neurodevelopment requires a catalog of all CREs that are active in humans. Based on the conservation of genomic sequence during evolution, many putative CREs have been identified, a large number of which are proposed to regulate the expression of genes enriched in the human brain. Conversely, DNA sequences with dramatically increased substitution rates in the human lineage compared to other organisms have also been used to predict CREs in humans.

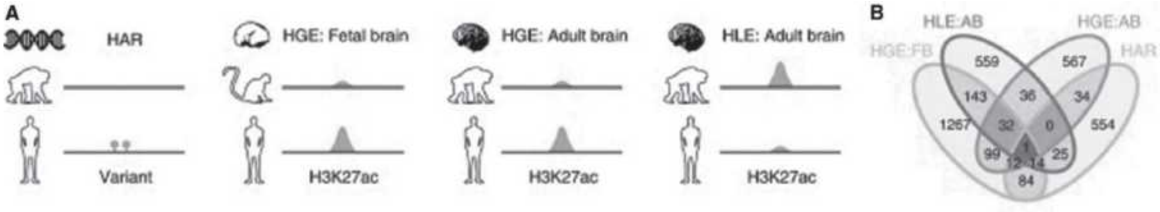

FIGURE 1.3–5. **A.** Schematic of human evolved elements. Human accelerated regions (HAR) are defined as regions with accelerated sequence change in human compared to other primates. Human hypermethylated regions are defined as regions with increased DNA methylation in human compared to other primates. Human-gained elements (HGEs) are defined as genomic regions with gained regulatory activity, defined by H3K27ac levels, in humans compared to other primates. Human lost elements (HLEs) are defined as regulatory regions lost in humans. **B.** Genes associated with human-evolved elements are enriched in outer radial glia and endothelial cells in the fetal brain (FB) and astrocytes in adult brain (AB). (Reproduced from Won H, Huang J, Opland CK, et al. Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. *Nat Commun* 2019;10(1):2396. doi: 10.1038/s41467-019-10248-3. License: https://creativecommons.org/licenses/by/4.0/)

Similar to studying the transcriptome, the development of high-throughput technologies (see section "High-Throughput Techniques to Study Molecular Landscapes") has greatly improved the ability to detect CREs in the human brain. Traditionally, chromatin accessibility and nucleosome depletion and, more recently, enrichment for specific posttranslational histone modifications have been considered to be among the defining characteristics of CREs. High-throughput assay for transposase accessible chromatin (ATAC) and ChIP assays followed by deep sequencing (ChIP-seq) has been applied in postmortem human brain tissue. Visel et al. determined the genome-wide occupancy of the enhancer-associated proteins p300/CBP in the human fetal neocortex to identify active enhancers. By mapping enrichment of two enhancer-enriched posttranslational modifications of histones, histone 3 lysine 4 monomethylation (H3K4me1) and histone 3 lysine 27 acetylation (H3K27ac), enhancers active in different brain regions and at different time points were deciphered. For example, Won et al. identified evolutionarily dynamic regions in humans, including human accelerated regions, human-gained enhancers, and human lost enhancers (Fig. 1.3–5A). They annotate those regions to potential target genes and found that human-evolved regions are enriched for genes expressed in adult cortical oligodendrocytes and fetal outer radial glial cells (Fig. 1.3–5B). Combined with human cortical region-specific ATAC-seq studies, these studies have identified enhancers that have been gained in humans and putative enhancers specific to cortical regions. In addition, coregulated enhancers that form cell-type and context-specific networks have been identified. Many of these enhancer networks are correlated with modules of coexpressed genes that function in neuronal proliferation, migration, and cortical-map organization. Thus, the regulatory networks that govern spatial-temporal gene expression in the developing human CNS are beginning to be elucidated; however, this information has yet to be fully placed into a functional context.

EPIGENETIC REGULATION OF HUMAN BRAIN DEVELOPMENT

Epigenetics refers to the chemical modification of DNA, chromatin structure, and RNA that can modulate gene expression without changing the DNA sequence. Epigenetic states are dynamic over neurodevelopment and influenced by both the genome and the environment. Therefore, epigenetic regulation is a mechanism by which susceptibility to psychiatric disorders can emerge. The two most commonly studied epigenetic mechanisms are DNA methylation, which generally represses gene expression by inhibiting transcription factors from binding DNA, and posttranslational histone modifications, which alter the chromatic structure to make it either more or less accessible to transcriptional machinery.

DNA methylation of cytosine, primarily at CpG nucleotides, which are enriched in gene promoters, plays a key role in neural development and function and can be profiled at hundreds of thousands to millions of CpGs in the human genome in an unbiased manner. A widespread methylome reconfiguration during fetal to young adult development has been observed in multiple studies. Global DNA methylation in the human prefrontal cortex was shown to be greatly reconfigured during fetal-to-postnatal development. These changes were more rapid during fetal development and slowed down after birth and with aging. Many of these methylation changes correlate significantly with the changing cellular composition of the human cortex during development, highlighting the important role of DNA methylation in human neurodevelopment.

In addition to methylation CpGs, accumulation of non-CpG methylation (mCH) during the early postnatal developmental period that coincides with the primary phase of synaptogenesis has recently been discovered in the human brain. This accumulation exhibits cell-type specificity, with abundant mCH in neurons and insignificant levels in glial cells. CH methylation has increased over the course of human brain evolution, contributing to human-specific downregulation of genes and coexpression modules. The effects of CH hypermethylation are particularly pronounced in neuronal subtypes and active in early human postnatal development; disruption of developmental processes mediated by CH methylation may therefore contribute to vulnerability to neuropsychiatric disorders, and understanding them may lend new insights into the derivation of targeted treatments for those conditions. For both CpG and mCH methylation, a strong correlation with the expression of adjacent genes has been observed.

Another key epigenetic mechanism for establishing cellular diversity and regulating the timing of neurodevelopment is the posttranslational modification of histones. Methylation (mono-, di-, and tri-) and acetylation are the most extensively studied modifications, although many more have been observed. Studies have generated chromatin state maps in a limited number of postmortem human brains by applying ChIP coupled with next-generation sequencing (ChIP-seq). Cheung et al. profiled histone 3 lysine 4 trimethylation (H3K4me3) across development in neuronal and nonneuronal cells of the human prefrontal cortex of 11 individuals ranging in age from 0.5 to 69 postnatal years. A key finding of their study was significant remodeling during postnatal development and aging of prefrontal cerebral cortical neurons. Both methylation and *demethylation* of DNA are required for neuronal differentiation and function in all mammalian species; it has been observed that high levels of 5-hydroxymethylcytosine (5hmC) accumulate in mammalian neurons after birth when the cells are no longer dividing. Our knowledge of the role of

5hmC in terminal differentiation or as an intermediate in active DNA demethylation is still evolving; recently, Stoyanova et al. demonstrated in an animal model system that neurons do not mature properly if 5hmC is not produced continuously following the first week of life. Their data highlighted the role of Tet proteins, which convert 5mC into 5hmC, in preparing methylation marks for removal and demonstrating that active removal of these marks is essential for neuronal differentiation. Taken together, these data confirm that epigenetic mechanisms play a key role in establishing the regional and temporal landscape of gene expression and contribute to human brain evolution.

TRANSCRIPTIONAL AND EPIGENETIC DYSREGULATION IN NEURODEVELOPMENTAL AND PSYCHIATRIC DISORDERS

Increasingly, it is becoming understood that many neuropsychiatric conditions are rooted in genetic and environmental perturbations of early neurodevelopment. Over the last few years, genetics research has made significant strides in identifying genetic risk factors for psychiatric disorders. Most psychiatric disorders are highly complex, with multiple genes contributing to disease risk. There is very strong evidence that the genetic factors involved in psychiatric disease comprise both common and rare variants (SNPs) and submicroscopic variations in chromosomal structure (i.e., CNVs). For example, in schizophrenia and bipolar disorder, gene discovery has centered on the linkage between these disorders with common SNPs, while in ASD, highly penetrant rare and de novo single nucleotide variants are capable of engendering severe syndromes, usually associated with intellectual disability. When a neuropsychiatric condition is manifest by a single causative gene, as in many neural developmental disorders (e.g., fragile X syndrome [FXS]), there is broad dysregulation of neurodevelopmental processes that ultimately lead to disease phenotypes independent of the mutated gene’s direct biologic roles. The polygenicity of common psychiatric disorders makes identifying the molecular mechanisms underlying the pathobiology of psychiatric diseases difficult. Further complicating the ability to understand the genetic bases of disease is that, in addition to the above-mentioned germline mutations, somatic mutations—de novo mutations occurring postfertilization and often in specific cell and tissue types—can contribute to psychiatric and neurodevelopmental disorders.

Genetics and complex polygenic risk architecture do not completely explain susceptibility to many psychiatric diseases. Indeed, for many disorders, the possibility of two identical twins developing the disease is lower than the genetic risk, indicating that environmental factors influence disease susceptibility. For example, environmental factors such as obstetric complications have been reported to increase the risk of developing schizophrenia. Aversive stimuli from the environment have broad effects on the cellular processes in the developing brain, ranging from altered gene expression levels to dysregulated neural activity, which then interacts with neural development to convey increased susceptibility to psychiatric problems. Moreover, adverse events in development can lead to epigenetic changes that might have permanent effects on gene expression; in neuropsychiatric disease states, *stochastic influences*—random events that minimally affect the course of typical development—may have amplified effects on severity and variation in outcome. This is a specific example of genetic variation accentuating sensitivity to variation in the environment in a manner that induces highly variable rather than highly predictable responses.

Despite this etiologic complexity, the neurodevelopmental disruptions underlying psychiatric diseases converge on distinct and presumably definable cell and molecular processes, brain circuits, and epochs of development. Resolving these processes conveys a better understanding of the underlying pathology and may lead to more targeted therapeutic strategies. As described in several examples below, increasingly comprehensive studies of the cell and molecular biology of neurotypical human neurodevelopment can be integrated with studies of psychiatric disease, genetics, and pathology to interrogate disease mechanisms and identify the time and place where disrupted molecular processes are likely to act.

Fragile X Syndrome: Single Gene Mutations Can Have Broad Downstream Consequences

FXS is the leading inherited cause of intellectual disability and is often accompanied by autistic-like features, aggression, attention deficits, and delays in speech and language development. FXS is caused by the loss of function of the FMR1 gene, which encodes an RNA-binding protein

(FMRP) involved in mRNA localization, stability, and translation. Therefore, loss of FMRP function has broad downstream consequences on gene expression and translation. For example, in a recent study, FMRP was shown to interact with the human neuronal nitric oxide synthase 1 (NOS1) *gene* through both conserved and species-dependent processes. *In the cerebral cortex,* NOS1 protein expression is restricted to layer 5 (L5) subcerebrally projecting pyramidal neurons and exhibits expression in alternating minicolumns in the frontoparietal operculum. In humans, FMRP interacts with sequences in the NOS1-coding region and leads to increased NOS1 protein expression. Concordantly, neocortical NOS1 protein levels are dramatically reduced in the developing FXS human neocortex, where NOS1 and FMRP are transiently coexpressed during synaptogenesis in subpopulations of pyramidal neurons in regions involved in speech, language, and complex social behaviors.

Down Syndrome and ASD: Polygenic Disruptions Converging on Specific Developmental Processes

Down syndrome (DS) provides an interesting neurodevelopmental disorder for studying the biologic consequences of polygenic, complex disorders, as the source of polygenic disruption (trisomy 21) is well established. Studies have indicated that trisomy 21 causes genome-wide disruption of gene expression. Among these, a characterization of the DS developmental brain transcriptome found that approximately 5% of genes on all chromosomes were dysregulated at some point in development, with a roughly equal number of up- and downregulated genes, with the exception of 18% of Chr21 genes, all of which were upregulated (Fig. 1.3–6). Gene coexpression analysis revealed gene networks that were implicated in processes previously associated with DS (e.g., synaptic function and neuronal specification) and aberrations in genes for particular cellular and molecular processes that had not been well established in DS, such as altered expression of genes associated with myelination and oligodendrocyte development. Subsequent analysis of postmortem human DS brain specimens revealed prominent hypomyelination, disorganization of myelinated fibers, and slower conductance through myelinated fiber tracts as features of DS neuropathology.

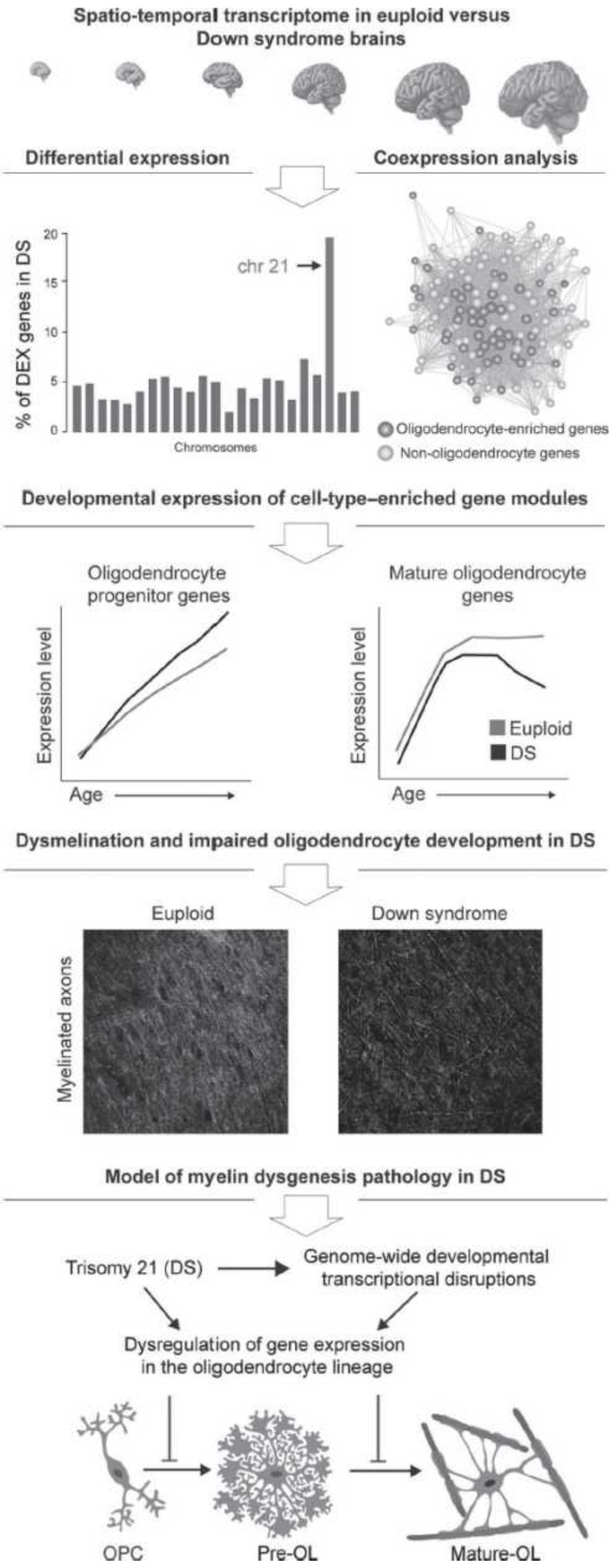
Spatio-temporal transcriptome in euploid versus Down syndrome brains
Differential expression
Coexpression analysis
% of DEX genes in DS
20
15
10
5
0
chr 21
Chromosomes
Oligodendrocyte-enriched genes
Non-oligodendrocyte genes
Developmental expression of cell-type–enriched gene modules
Oligodendrocyte progenitor genes
Mature oligodendrocyte genes
Expression level
Age
Euploid
DS
Dysmelination and impaired oligodendrocyte development in DS
Euploid
Down syndrome
Myelinated axons
Model of myelin dysgenesis pathology in DS
Trisomy 21 (DS)
Genome-wide developmental transcriptional disruptions
Dysregulation of gene expression in the oligodendrocyte lineage
OPC
Pre-OL
Mature-OL

FIGURE 1.3–6. Genome-wide dysregulation in Down syndrome (DS) leads to impaired oligodendrocyte development. Schematic of a study by Olmos-Serrano et al. showing how developmental transcriptome characterizations of

DS (trisomy 21, DS) have provided insight into how polygenic gene disruptions lead to far broader impacts on gene expression and ultimately cell and molecular function. In this study, global gene expression was characterized from midfetal development to adulthood in several brain regions of individuals with DS (first panel). It was determined that approximately 20% of genes of chromosome 21 genes were misexpressed at some point in development, while 3% to 8% of the genes on all other chromosomes were misexpressed. Thus, in total, far more nonchromosome 21 genes are misexpressed in DS (second panel). Network analysis of coexpressed genes revealed a module that was highly enriched in myelination genes. Subsequent analysis of gene networks enriched in genes highly expressed in either oligodendrocyte precursors or mature oligodendrocytes revealed substantial dysregulation of developmental gene expression dynamics in the oligodendrocyte lineage: progenitor genes were more highly expressed and mature genes were expressed at lower levels (third panel). Analysis of myelinated fiber tracts in the prefrontal cortex revealed hypomyelination and disorganization of myelinated axons in DS (fourth panel). This study suggests that trisomy of chromosome 21 leads to genome-wide disruptions in developmental gene regulation, which were particularly prominent in oligodendrocytes and resulted in impaired oligodendrocyte maturation and hypomyelination (fifth panel). (Modified from Olmos-Serrano JL, Kang HJ, Tyler WA, et al. Down syndrome developmental brain transcriptome reveals defective oligodendrocyte differentiation and myelination. *Neuron* 2016;89(6):1208–1222. Copyright © 2016, with permission from Elsevier. doi: 10.1016/j.neuron.2016.01.042.)

Similarly, to interrogate ASD and schizophrenia pathogenesis, recent studies have used computational approaches to integrate data on risk variants from GWAS and whole exome sequencing with spatiotemporal brain transcriptome data. For instance, by building coexpression networks of genes previously associated with ASD, researchers have identified modules of highly related genes and determined the time in development and cell types in the brain where these gene networks are enriched. Intriguingly, it was found that ASD-risk genes were highly enriched in cortical projection neurons in the frontal lobe during midfetal development (Fig. 1.3–7). Of note, genes associated with schizophrenia showed a similar enrichment. These data indicate that, in ASD and schizophrenia, development may be particularly disrupted in cortical projection neurons when they are in the process of terminal migration and beginning to elaborate dendrites and early axonal projections. Moreover, functional and anatomical imaging studies have revealed that brain circuits composed of these neurons are abnormal in both ASD and schizophrenia.

Noncoding Mutations in Neurodevelopmental Disorders

The majority of disease-associated sequence variants occur in noncoding regions of the genome. While the consequences of these mutations are difficult to study, many likely have developmental effects and represent risk factors for psychiatric disease. This is particularly the case for variants that lie within regulatory regions, such as enhancers and promoters. An illustrative example is a recently described study of perisylvian polymicrogyria that results from a deletion immediately upstream of one of 17 transcriptional start sites (TSSs) of *GPR56*, a G-coupled protein receptor encoding gene that is important for stem cell proliferation and cerebral cortical development. In five affected individuals, a 15-base pair deletion was detected in a region of the GPR56 core promoter that is essential for driving expression in perisylvian regions of the brain during prenatal development. This deletion likely impaired the expression of GPR56, specifically in the perisylvian cortex during neurogenesis, and led to reduced progenitor cell proliferation and microgyrification, specifically in the perisylvian cortex (Fig. 1.3–4B). While this example is extreme, it demonstrates how differences in noncoding sequences, particularly in regulatory elements, can convey different disease susceptibilities and neurodevelopmental phenotypes.

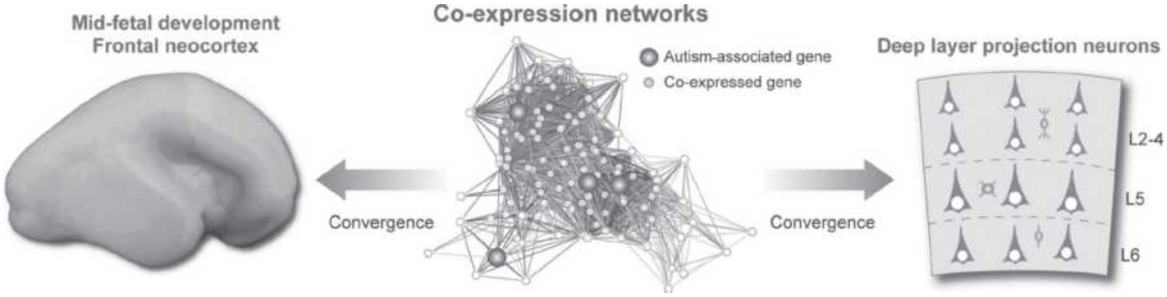

FIGURE 1.3–7. Dysregulation of autism spectrum disorder (ASD)-implicated genes converge on gene networks enriched in midfetal cortical projection neurons. This figure illustrates how a coexpression network analysis of ASD-associated genes with a spatiotemporal human transcriptome database inferred a convergence of autism-associated genes in frontal neocortex and deep layer projection neurons, suggesting that abnormalities in this cell type may play a role in ASD pathogenesis. ASD-associated genes identified in genome-wide sequencing studies were used as seed genes to build coexpression networks (middle panel). The networks of genes highly interconnected with ASD-associated genes were found to be enriched for genes highly expressed in the frontal cerebral cortex (left panel) and deep layer projection neurons (right panel). (Modified from Willsey AJ, Sanders SJ, Li M, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155(5):997–1007. Copyright © 2013, with permission from Elsevier. doi: 10.1016/j.cell.2013.10.020.)

Somatic Mutations: Regionally Discrete Gene Mutations and Patterns of Gene Misexpression

Although somatic mutations in the brain cannot be readily detected using standard genomic screening techniques, recent advances in deep next-generation sequencing and single-cell sequencing have enabled the detection of somatic mutations in specific cells and tissues and advanced the current understanding of their contribution to disease etiology. This work has broad implications, from uncovering new disease mechanisms to providing new insights into neurotypical development. For example, these studies have demonstrated that certain neurodevelopmental diseases may be caused by *de novo* somatic mutations in neural progenitors that then expand clonally to produce aberrations in specific neural cell types and brain regions that can lead to brain malformations and more subtle phenotypes, as recently suggested by postmortem studies of schizophrenia.

Epigenetic Dysregulation in Schizophrenia

Dysregulation of epigenetic mechanisms can lead to transcriptome changes observed in disease states. While understanding the role of epigenetic modifications in psychiatric diseases remains in its nascent phase, recent systematic reviews have identified a moderate association between changes in various epigenetic modifications, including DNA methylation, histone modifications, and noncoding miRNAs, and schizophrenia and other psychotic disorders.

Initial studies analyzed the role of DNA methylation changes in promoters of individual candidate genes. In schizophrenia, several lines of evidence suggest that reelin gene (*RELN*) expression, which encodes an extracellular matrix protein that plays critical roles in the developing and adult brain, is downregulated due to hypermethylation of its promoter. A similar observation linked the promoter of the glutamic acid decarboxylase 67 (GAD1) gene, a gene encoding an enzyme that is required for the synthesis of GABA, with schizophrenia. In addition, decreased enrichment of H3K4me3 and an excess of H3K27me3, a repressive mark, have also been observed in the GAD1 promoter. Highlighting the potential importance of GAD1 dysregulation in schizophrenia are the well-established findings that GAD1 and GABA are not only expressed at lower levels in schizophrenia but also that mutations in GAD1 have been implicated in childhood-onset schizophrenia and that GAD autoantibody syndrome typically presents as schizophrenia.

Additionally, mutations and structural variations in genes involved in histone modifications, known as chromatin modifiers, have been associated with various neurodevelopmental disorders and schizophrenia. For example, loss of functions variants of the gene, SETD1A, which is part of a protein complex that mediates histone methylation, confer a large increase in risk for schizophrenia. Mouse lines carrying a heterozygous loss-of-function mutation in SETD1A show deficits in cortical connectivity and working memory. Furthermore, genes putatively regulated by SETD1A in the mouse brain were enriched for neurodevelopmental genes enriched in the developing cortex. While these monogenic forms of schizophrenia are likely to be rare, integrated analyses of schizophrenia-associated genetic variations do suggest a greater contribution of chromatin regulators to the neurobiology of schizophrenia.

More recently, studies have explored possible roles of epigenetic mechanisms of psychiatric disease at a genome-wide scale. For example, upon integrating the DNA methylation differences observed during normal brain development and risk loci associated with schizophrenia in a genome-wide association study, Jaffe et al. observed that global CpG methylation is different between neurotypical prenatal and postnatal samples. CpG sites that were hypermethylated prenatally were enriched for schizophrenia risk loci. However, a significant enrichment for risk was not observed at the age of onset of illness (around 25 years). A direct comparison of the DNA methylation profiles between the control and schizophrenia adult prefrontal cortex identified approximately 2,100 CpG loci that were differentially methylated. Although these loci were not enriched in schizophrenia risk alleles, they again exhibited pre- and postnatal changes in the neurotypical cortex.

FUTURE DIRECTIONS

Neuropsychiatric disorders owe in part to broad dysregulation of the genome over the course of development. Transcriptome data suggest that, in complex psychiatric and neurodevelopmental disorders, disruptions in many genes can converge on particular cell types and biologic processes to produce disease phenotypes. Thus, selective dysfunction of spatially and temporally regulated gene expression may account for differences in the age of onset and affected neural circuits in neurologic and psychiatric disorders (see Fig. 1.3–1). However, a great deal of work must be undertaken to further elucidate the mechanisms of diseases, determine how genetic and somatic variations cause diseases and develop more rational therapeutic strategies.

It is important to clarify that the changes in the transcriptome may reflect the cause or the *consequence* of a disorder. This is most notable for postmortem brain samples from adult patients with ASD and schizophrenia, which are associated with *in utero* disruptions of brain development. One way to infer the direction of causation is to compare genetic correlations between disease pairings with their corresponding transcriptome overlap, and this approach has confirmed that major components of gene-expression patterns in neuropsychiatric disorders reflect biologic processes coupled with underlying genetic variation. Another approach is to focus on mutations in noncoding regions of genes, which comprise most disease-associated variants. Specifically, characterizing the spatiotemporal activity of relevant CREs and identifying their associated gene can directly link disease-associated variants to transcriptomic changes. Modern experimental approaches such as Massively Parallel Reporter Assay (MPRA) and computational approaches such as quantitative trait locus (QTL) mapping are currently being utilized to further characterize and associate CREs with specific genes.

Multidimensional genomic data, including more comprehensive transcriptomic, epigenomic, and regulatory maps of the developing human brain, will greatly advance the current understanding of the pathophysiology of complex diseases such as psychiatric disorders. These datasets may one day reveal genetic and molecular phenotypes of diseases and disease subtypes that might be useful for diagnosis and more targeted treatment paradigms. In addition to generating genomic and transcriptomic data, integrative analyses with pathologic, neuroimaging, behavioral, and clinical studies will be critical for a full understanding of psychiatric conditions.

FURTHER READINGS

BRAIN Initiative Cell Census Network (BICCN). A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature*. 2021;598(7879):86–102.

Castelbaum L, Sylvester CM, Zhang Y, Yu Q, Constantino JN. On the nature of monozygotic twin concordance and discordance for autistic trait severity: A quantitative analysis. *Behav Genet*. 2020;50(4):263–272.

Eze UC, Bhaduri A, Haeussler M, Nowakowski TJ, Kriegstein AR. Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat Neurosci.* 2021;24:584–594.

Gandal MJ, Haney JR, Parikshak NN, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science.* 2018;359(6376):693–697.

Hauberg ME, Creus-Muncunill J, Bendl J, et al. Common schizophrenia risk variants are enriched in open chromatin regions of human glutamatergic neurons. *Nat Commun.* 2020;11:1–16.

Jeong H, Mendizabal I, Berto S, et al. Evolution of DNA methylation in the human brain. *Nat Commun.* 2021;12:2021.

Johnson MB, Kawasawa YI, Mason CE, et al. Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron*. 2009;62:494–509.

Kang HJ, Kawasawa YI, Cheng F, et al. Spatio-temporal transcriptome of the human brain. *Nature*. 2011;478:483–489.

*Kelley KW, Pașca SP. Human brain organogenesis: Toward a cellular understanding of development and disease. *Cell.* 2022;185:42–61.

Kim MH, Kim IB, Lee J, et al. Low-level brain somatic mutations are implicated in schizophrenia. *Biol Psychiatry.* 2021;90(1):35–46.

Kwan KY, Lam MM, Johnson MB, et al. Species-dependent posttranscriptional regulation of NOS1 by FMRP in the developing cerebral cortex. *Cell*. 2012;149:899–911.

Li M, Santpere G, Imamura Kawasawa Y, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science*. 2018;362(6420):eaat7615.

Markenscoff-Papadimitriou E, Whalen S, Przytycki P, et al. A chromatin accessibility atlas of the developing human telencephalon. *Cell.* 2020;182:754–769.e18.

Miller JA, Ding SL, Sunkin SM, et al. Transcriptional landscape of the prenatal human brain. *Nature*. 2014;508:199–206.

Mukai J, Cannavò E, Crabtree GW, et al. Recapitulation and reversal of schizophrenia-related phenotypes in Setd1a-deficient mice. *Neuron.* 2019;104:471–487.e12.

Myers SM, Challman TD, Bernier R, et al. Insufficient evidence for “autism-specific” genes. *Am J Hum Genet*. 2020;106(5):587–595.

O’Leary DD, Chou SJ, Sahara S. Area patterning of the mammalian cortex. *Neuron*. 2007;56:252–269.

Olmos-Serrano JL, Kang HJ, Tyler WA, et al. Down syndrome developmental brain transcriptome reveals defective oligodendrocyte differentiation and myelination. *Neuron*. 2016;89(6):1208–1222.

Parikshak NN, Luo R, Zhang A, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*. 2013;155:1008–1021.

Shibata M, Pattabiraman K, Lorente-Galdos B, et al. Regulation of prefrontal patterning and connectivity by retinoic acid. *Nature*. 2021;598:483–488.

Shibata M, Pattabiraman K, Muchnik SK, et al. Hominini-specific regulation of CBLN2 increases prefrontal spinogenesis. *Nat.* 2021;598:489–494.

Shim S, Kwan KY, Li M, Lefebvre V, Sestan N. Cis-regulatory control of corticospinal system development and evolution. *Nature*. 2012;486:74–79.

*Silbereis John C, Pochareddy S, Zhu Y, Li M, Sestan N. The cellular and molecular landscapes of the developing human central nervous system. *Neuron*. 2016;89:248–268.

Stoyanova E, Riad M, Rao A, Heintz N. 5-Hydroxymethylcytosine-mediated active demethylation is required for mammalian neuronal differentiation and function. *Elife.* 2021;10:e66973.

Sullivan PF, Daly MJ, O’Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Gen*. 2012;13:537–551.

*Tebbenkamp AT, Willsey AJ, State MW, Sestan N. The developmental transcriptome of the human brain: implications for neurodevelopmental disorders. *Curr Opin Neurol*. 2014;27:149–156.

Werling DM, Pochareddy S, Choi J, et al. Whole-genome and RNA sequencing reveal variation and transcriptomic coordination in the developing human prefrontal cortex. *Cell Rep*. 2020;31:107489.

Willsey AJ, Sanders SJ, Li M, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. 2013;155:997–1007.

Won H, Huang J, Opland CK, Hartl CL, Geschwind DH. Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. *Nat Commun*. 2019;10(1):2396.

Ziffra RS, Kim CN, Ross JM, et al. Single-cell epigenomics reveals mechanisms of human cortical development. *Nature*. 2021;598:205–213.

1.4 Neural Development and Neurogenesis

EMANUEL DICICCO-BLOOM, M.D., AND COURTNEY R. MCDERMOTT, B.A.

NEURAL DEVELOPMENT AND NEUROGENESIS

An understanding of molecular and cellular mechanisms mediating nervous system development is critical in psychiatry as we know well that abnormalities of developmental processes contribute to many brain disorders. While a developmental basis may not be surprising in early childhood disorders, such as autism, fragile X mental retardation, and Rett syndrome, even mature diseases including schizophrenia and depression reflect ontogenetic factors. For example, there are reductions in forebrain region volumes, neuron and glial cell numbers that may be apparent at the time of diagnosis, although with maturation additional changes occur in neuronal cell density, neuropil, and myelination. Similarly, in autism, early brain growth may be increased, while abnormalities of cellular organization are observed that reflect fetal disturbances in developmental cell proliferation and migration. When there is abnormal regulation of early brain development, we lay down a foundation of altered neuron populations that may differ in cell types, numbers, and positions, or may elaborate abnormal connections, with consequences for interacting glial populations. With progressive postnatal development, the maturing brain systems call upon component neurons to achieve increasing levels of complex information processing, which may be deficient should initial conditions be disturbed. New neural properties emerge during maturation as neuron populations elaborate additional functional networks based upon and modified by ongoing experience. Given the brain's dynamic character, developmental abnormalities in neural populations and systems, caused by genetic as well as environmental factors, will likely manifest at diverse times in a person's life.

OVERVIEW OF NERVOUS SYSTEM MORPHOLOGIC DEVELOPMENT

In considering brain development, several overarching principles may serve to guide our understanding. First, different brain regions and neuron populations are generated at distinct times of development, which in turn provide distinct *windows of susceptibility* to genetic and environmental insults. Second, the sequence of cellular processes that comprise ontogeny predicts that abnormalities in early events necessarily lead to differences in subsequent stages, though such abnormalities may not be accessible to our clinical tools. For example, a deficit in the number of early-born neurons will likely lead to later reductions in axonal processes and ensheathing white matter, though only the latter may be detectable on brain imaging using MRI or DTI of the mature brain. Third, it is clear that specific molecular signals, such as extracellular growth factors and cognate receptors or transcription factors, regulate multiple developmental stages of the cell, and play roles in cell proliferation, survival promotion, neuron migration, process outgrowth, and the momentary synaptic modifications (plasticity) underlying learning and memory. Finally, higher functioning mammals possess a higher capacity for neurogenesis, including stem cell proliferation as well as prolonged maturation of neural precursors that may serve for continued growth, learning and memory, and disease vulnerability. This chapter considers these principles as it examines cellular and molecular systems regulating development, and discusses implications for psychiatric disease.

The Neural Plate and Neurulation

The nervous system of the human embryo first appears between 2 and 1/2 and 4 weeks of gestation. During development, emergence of new cell types, including neurons, results from interactions between neighboring layers of cells. On gestational day 13, the embryo consists of a sheet of cells. Following gastrulation (days 14 and 15), which forms a two-cell–layered embryo consisting of ectoderm and endoderm, the neural plate region is delineated on day 16 by the underlying mesoderm. The mesoderm forms by cells entering a midline cleft in the ectoderm called the primitive streak and it induces overlying ectoderm to become neural plate. Induction involves release of soluble growth factors from one group of cells, which in turn bind receptors on neighboring cells, eliciting changes in nuclear transcription factors which control downstream gene expression. In some cases, cell–cell contact-mediated mechanisms are involved. In the gene patterning section below, the important roles of soluble growth factors and transcription factor expression will be described.

The neural plate, whose induction is complete by 18 days, is a sheet of columnar epithelium, and is surrounded by ectodermal epithelium. After formation, the edges of the neural plate elevate, forming the neural ridges. Subsequently, changes in intracellular cytoskeleton and cell–extracellular matrix attachment cause the ridges to merge in the midline and fuse, a process termed neurulation, forming the neural tube, with a central cavity presaging the ventricular system (Fig. 1.4–1). Fusion begins in the cervical region at the hindbrain level (medulla and pons) and continues rostrally and caudally. Neurulation occurs at 3 to 4 weeks of gestation, and its failure results in anencephaly rostrally and spina bifida caudally. The risk of developing neurulation defects are increased following exposure to retinoic acid in dermatologic preparations and anticonvulsants, especially valproic acid (VPA), as well as diets deficient in folic acid.

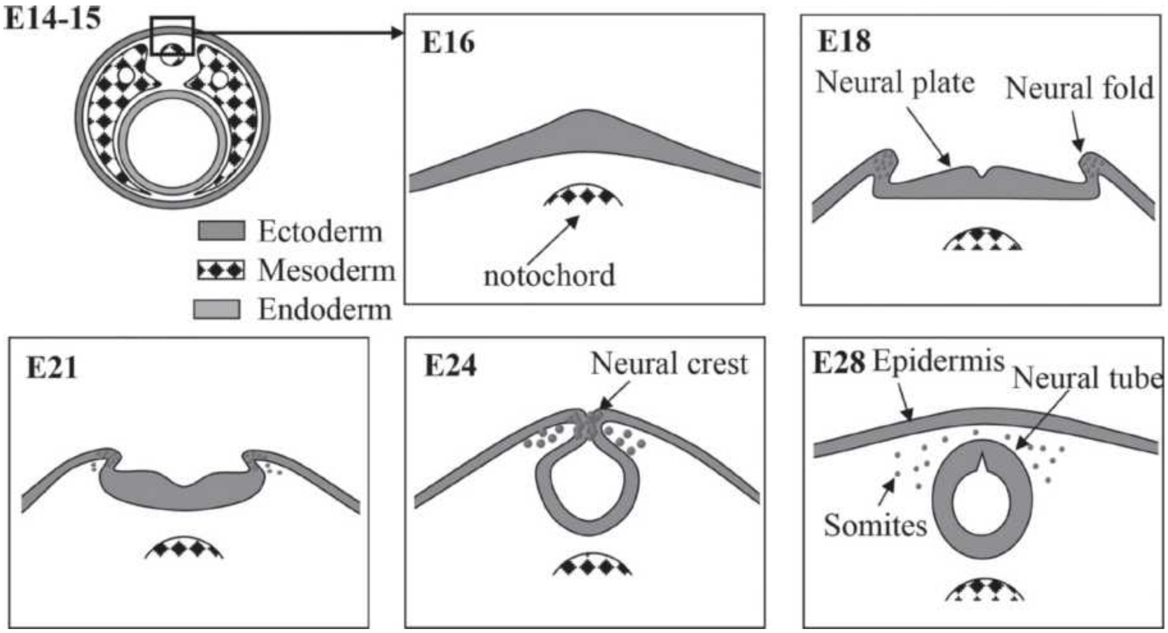

FIGURE 1.4–1. Mechanisms of neurulation. Neurulation begins with the formation of a neural plate in response to soluble growth factors released by the underlying notochord. The neural plate originates as a thickening of the ectoderm that results from cuboidal epithelial cells becoming columnar in shape. With further changes in cell shape and adhesion, the edges of the plate fold and rise, meeting in the midline to form a tube. Cells at the tips of the neural folds come to lie between the neural tube and overlying epidermis, forming the neural crest that gives rise to the peripheral nervous system and other structures.

Another product of neurulation is the neural crest, whose cells derive from the edges of the neural plate and dorsal neural tube. From this position, neural crest cells migrate dorsolaterally, under the skin to form melanocytes, and ventromedially to form dorsal root sensory ganglia and sympathetic chains of the peripheral nervous system and ganglia of the enteric nervous system. However, neural crest gives rise to diverse tissues including cells of neuroendocrine, cardiac, mesenchymal, and skeletal systems, forming the basis of many congenital syndromes involving brain and other organs. The neural crest origin at the border of neural and epidermal ectoderm and its generation of melanocytes forms the basis of the neurocutaneous disorders, including tuberous sclerosis and neurofibromatosis. Finally, another nonneuronal structure of mesodermal origin formed during neurulation is the notochord found on the ventral side of the neural tube. As shown below, the notochord plays a critical role during neural tube differentiation, since it is a signaling source of soluble growth factors, such as sonic hedgehog (Shh), which impact gene patterning and cell determination.

Regional Differentiation of the Embryonic Nervous System

After closure, the neural tube expands differentially to form major subdivisions that precede the major functional divisions of the brain. Different subdivisions are generated according to specific schedules of proliferation, migration, and differentiation. The neural tube can be described in three dimensions, including longitudinal, circumferential, and radial. The longitudinal dimension reflects rostrocaudal (anterior–posterior) organization, and most simply, the brain and spinal cord. Circumferential organization, tangential to the surface, represents two axes: in the dorsoventral axis, cell groups are uniquely positioned from top to bottom. In the medial to lateral axis, there is right–left mirror image symmetry. Finally, the radial dimension represents organization from the inner most cell layer adjacent to the ventricles to the outer most surface, and exhibits region-specific cell layering. At 4 weeks, the human brain is divided longitudinally into three “vesicles” the prosencephalon (forebrain), mesencephalon (midbrain), and rhombencephalon (hindbrain) (Fig. 1.4–2). By 5 weeks, these three “vesicles” divide further into five divisions with the prosencephalon forming the telencephalon (including cortex, hippocampus, and basal ganglia) and diencephalon (thalamus and hypothalamus), the mesencephalon, (midbrain), and the rhombencephalon yielding metencephalon (pons and cerebellum) and myelencephalon (medulla). These morphologic transformations depend on region-specific proliferation of precursor cells adjacent to the ventricles, the ventricular zones (VZs). Proliferation depends on soluble growth factors made by proliferating cells themselves or released from regional signaling centers. Growth factor production and cognate receptor expression also depend on region-specific patterning genes. VZ precursors, which appear morphologically homogeneous, express a checkerboard array of molecular genetic determinants that control the generation of specific types of neurons in each domain (Fig. 1.4–3).

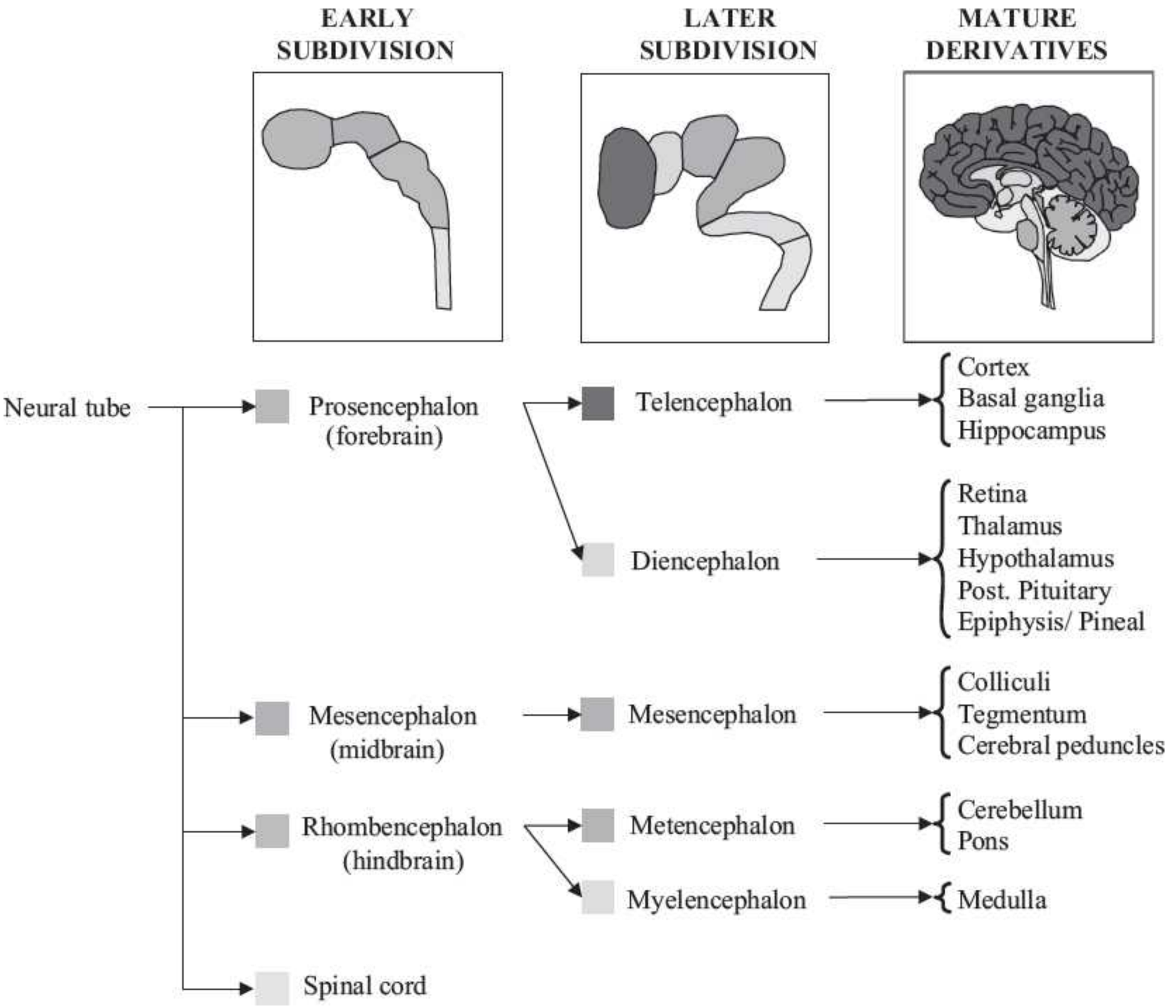
EARLY SUBDIVISION
LATER SUBDIVISION
MATURE DERIVATIVES
Neural tube
Prosencephalon (forebrain)
Telencephalon
Cortex
Basal ganglia
Hippocampus
Diencephalon
Retina
Thalamus
Hypothalamus
Post. Pituitary
Epiphysis/ Pineal
Mesencephalon (midbrain)
Mesencephalon
Colliculi
Tegmentum
Cerebral peduncles
Rhombencephalon (hindbrain)
Metencephalon
Cerebellum
Pons
Myelencephalon
Medulla
Spinal cord

FIGURE 1.4–2. Progression of brain regional differentiation. Early after neurulation, the neural tube differentiates into four regions (forebrain, midbrain, hindbrain, and spinal cord) that gives rise following later divisions and maturation to the different brain structures.

In the circumferential dimension, organization begins very early and extends over many rostrocaudal subdivisions. In spinal cord, the majority of tissue comprises the lateral plates, which later divide into dorsal or alar plates, composed of sensory interneurons, and motor or basal plates, consisting of ventral motor neurons. Two other diminutive plates, termed the roof plate and floor plate, play critical regulatory roles as growth factor signaling centers in the embryo. In response to Shh from the ventrally located notochord, the floor plate produces its own Shh, which induces neighboring cells in ventral spinal cord and brainstem to express region-specific transcription factors that specify cell phenotype and function. For example, floor plate Shh induces midbrain precursors to differentiate into dopamine secreting neurons of the substantia nigra. Similarly, the roof plate secretes bone morphogenetic proteins (BMPs), which induce dorsal sensory neuron cell fate in spinal cord, while in its absence, both cerebellum and midline hippocampus are missing. Finally, in the radial dimension, the organization of layers is subdivision-specific, produced by differential proliferation of VZ precursors and cell migration.

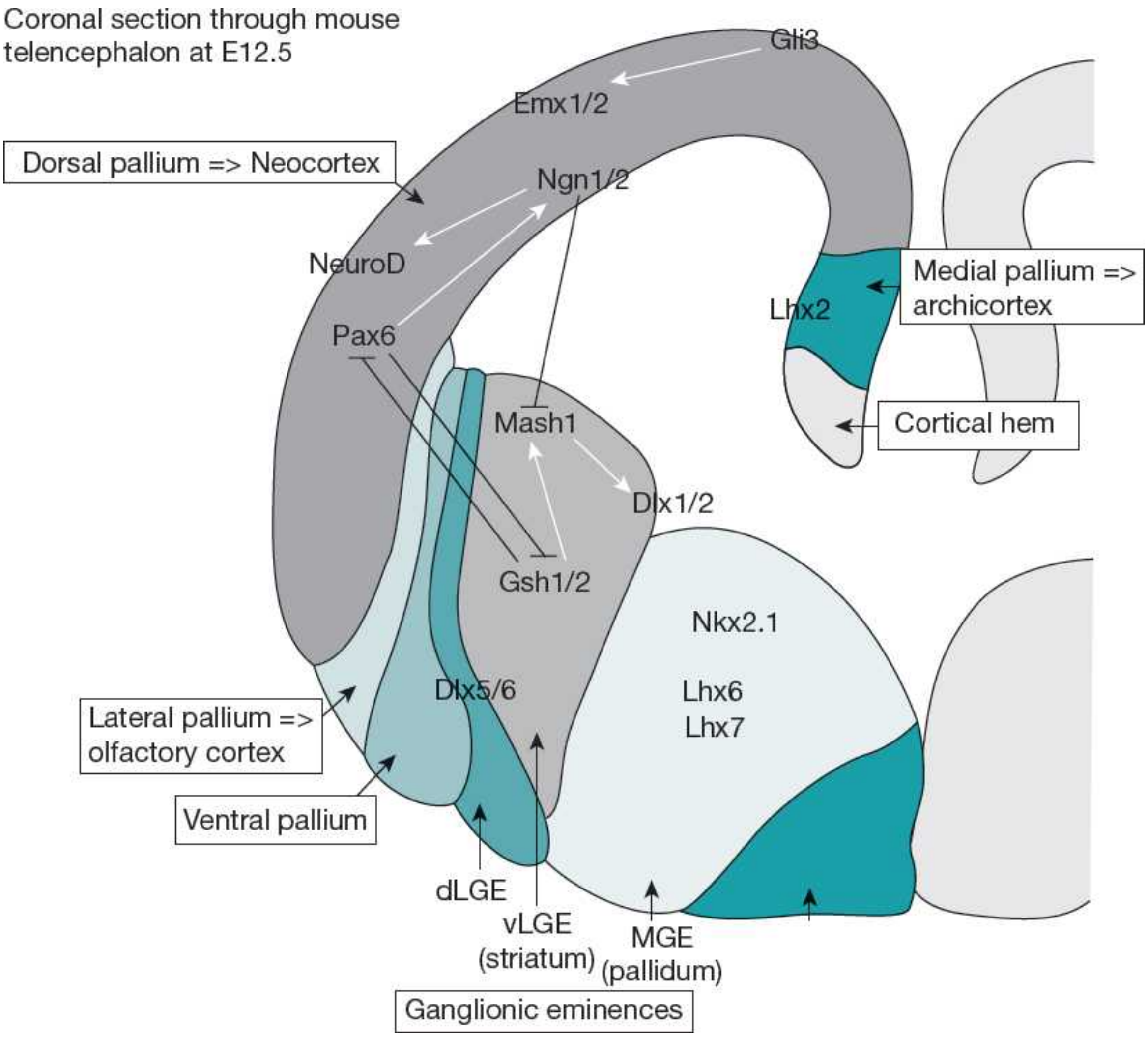
Coronal section through mouse telencephalon at E12.5
Gli3
Emx1/2
Dorsal pallium => Neocortex
Ngn1/2
NeuroD
Pax6
Mash1
Dlx1/2
Gsh1/2
Medial pallium => archicortex
Lhx2
Cortical hem
Nkx2.1
Lhx6
Lhx7
Dlx5/6
Lateral pallium => olfactory cortex
Ventral pallium
dLGE
vLGE (striatum)
MGE (pallidum)
Ganglionic eminences

FIGURE 1.4–3. Genetic interactions underlying regionalization of the mammalian telencephalon. A coronal section schematic of the E12.5 mouse telencephalon shows the dorsal and ventral subdomains as defined by their unique patterns of gene expression. Dorsal telencephalic progenitors express high levels of bHLH transcription factors Ngn1, Ngn2, and NeuroD and homeodomain proteins Emx1, Emx2, Lhx2, and Pax6. Ventral progenitors express the bHLH protein Mash1 and homeodomain factors Gsh1, Gsh2, Dlx1, Dlx2, Dlx5, and Dlx6. The LGE can be subdivided into dLGE and vLGE compartments, with higher levels of Pax6, Gsh2, Dlx2, and Mash1 in the dLGE, and Gsh1 expression only in the vLGE. In contrast, MGE progenitors alone express homeodomain transcription factors Lhx6, Lhx7, and Nkx2.1. Important cross-regulatory interactions between Ngn1/2 and Mash1 and Pax6 and Gsh2 participate in the maintenance of telencephalic progenitor identity. *Arrows* denote positive interactions; T-bars denote inhibitory interactions. (Adapted from Schuurmans C, Guillemot F. Molecular mechanisms underlying cell fate specification in the developing telencephalon. *Curr Opin Neurobiol* 2002;12(1):26–34. Copyright © 2002 Elsevier Science Ltd.)

Primary Proliferative Zones

The distinct patterns of precursor proliferation and migration in different regions generate the radial organization of the nervous system. In each longitudinal subdivision, the final size of the brain region population depends on the interplay of regulated neurogenesis with programmed cell death. Traditional concepts had suggested that excess cell production occurred everywhere, and that final cell number regulation was achieved primarily through selective cell death mediated by target-derived survival (trophic) factors. In fact, patterning genes play major roles in directing regional precursor proliferation that is coordinated with final structural requirements, and programmed cell death occurs at multiple stages. Consequently in diseases characterized by brain regions smaller than normal, such as schizophrenia, there may be a failure to generate neurons initially, as opposed to normal generation with subsequent cell loss.

The generation of specific cell types, the process of neurogenesis, involves proliferation of undifferentiated precursor cells (or progenitors), followed by cessation of proliferation (exit from the cell cycle), migration of cells to final destinations, and expression of specific phenotypic characters, such as neurofilaments and neurotransmitter systems. Precursor proliferation occurs primarily in two densely packed regions during development. The primary site is the VZ lining the walls of the entire ventricular system, which contributes to all brain regions in the rostrocaudal dimension. For select regions, however, including the cerebral cortex, hippocampus, and cerebellar cortex, precursors from the VZ migrate out to secondary zones where they generate a more restricted range of cell types.

In the early embryo, neural tube VZ progenitors are arranged as a one-cell layer thick, pseudostratified neuroepithelium. The bipolar VZ precursors have cytoplasmic processes that span from the ventricular to the pial surface. During the cell cycle, the VZ appears multi-layered, or stratified, because cell nuclei undergo movements, called interkinetic nuclear migration. Precursor cell division (M phase) occurs at the ventricular margin, producing two new cells (Fig. 1.4–4). The progeny then re-enter G1 as they move outward toward the pia. Under influence of extracellular signals these cells become committed to another round of division, marked by re-entry into S-phase, which occurs near the upper VZ margin. After replicating DNA, nuclei move back down during G2 to the ventricular surface where they undergo mitosis and divide. Different fields of study suggest several distinct roles for interkinetic migration: (1) Interkinetic nuclear migration may allow nuclei access to environmental cues produced by postmitotic cells that effect subsequent proliferation and gene expression; (2) Dividing cells at the VZ margin possess primary cilia that sense and respond to local regulatory signals including Shh, Wnt, and Notch; (3) Time lapse analysis suggests that because of the physically limited area of the VZ, and the high density of cells undergoing mitosis, the nuclei of nondividing cells move away from this region, but then return upon their next division. Human genetic mutations that interfere with interkinetic nuclear movement or the mitotic spindle organizer, the centrosome, impair cell migration, and produce heterotopic neurons and epilepsy syndromes.

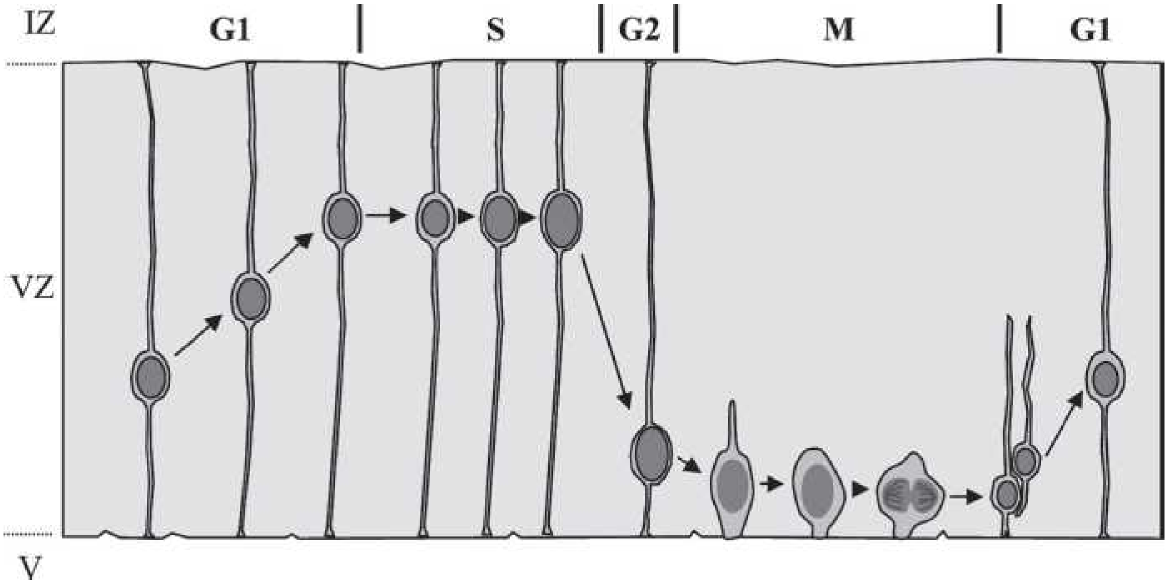

FIGURE 1.4–4. Interkinetic nuclear migration in the ventricular zone. During each cell cycle, nuclei move from the ventricular surface at G1 to the border of the ventricular zone where it enters S phase. Nuclei move down during G2, and reach the ventricular surface where they undergo mitosis. Asymmetric division leads to generation of a postmitotic cell that leaves the ventricular zone to produce a neocortical neuron, while the remaining stem cell continues to divide. IZ, intermediate zone; VZ, ventricular zone; V, ventricle. (Adapted from Nowakowski RS, Hayes NL. Cell proliferation in the developing mammalian brain. In: Rao MS, Jacobson M, eds. *Developmental Neurobiology*. 4th ed. Springer; 2005. doi: 10.1007/0-387-28117-7_2.)

At the earliest stages, VZ cells divide to increase the pool of progenitors. Then, during the period of neurogenesis, a precursor cell divides giving rise to both a postmitotic neuron and another dividing precursor. When neurogenesis ceases, precursor division produces two postmitotic neurons only, greatly increasing neuron production and depleting the precursor pool. Newly born neurons migrate to final destinations, such as cerebral cortical plate, traveling along the processes of radial glial cells (RGCs) (Fig. 1.4–5A). RGCs are bipolar VZ precursors that have one process associated with the ventricular surface and the other contacts the pial surface. The association between newborn neurons and radial glial processes allows cells generated within localized VZ domains that express distinct patterning genes, to migrate to specific cortical functional areas. This model has been termed the cortical "protomap" and was originally proposed by P. Rakic. Precursor cell phenotypic fate is already specified genetically before cells cease division and begin migration. To some degree cells born from the same VZ domain form an associated radial cell column (clonal population) that first exhibit electrical coupling and then preferentially develop chemical synaptic interconnections. This neuroanatomic relationship is one basis of the well-defined functional and neurophysiologic activities observed in cortical columns of rodents and all primates, including humans. Nonetheless the ingrowing thalamic afferents that synapse on cortical neurons also impact their final neuronal fate and function. In the human brain, neuron production generally occurs for the first 4 months of gestation, whereas from then on until birth, neurons undergo migration, whereas glial precursors proliferate, migrate, and begin producing myelin.

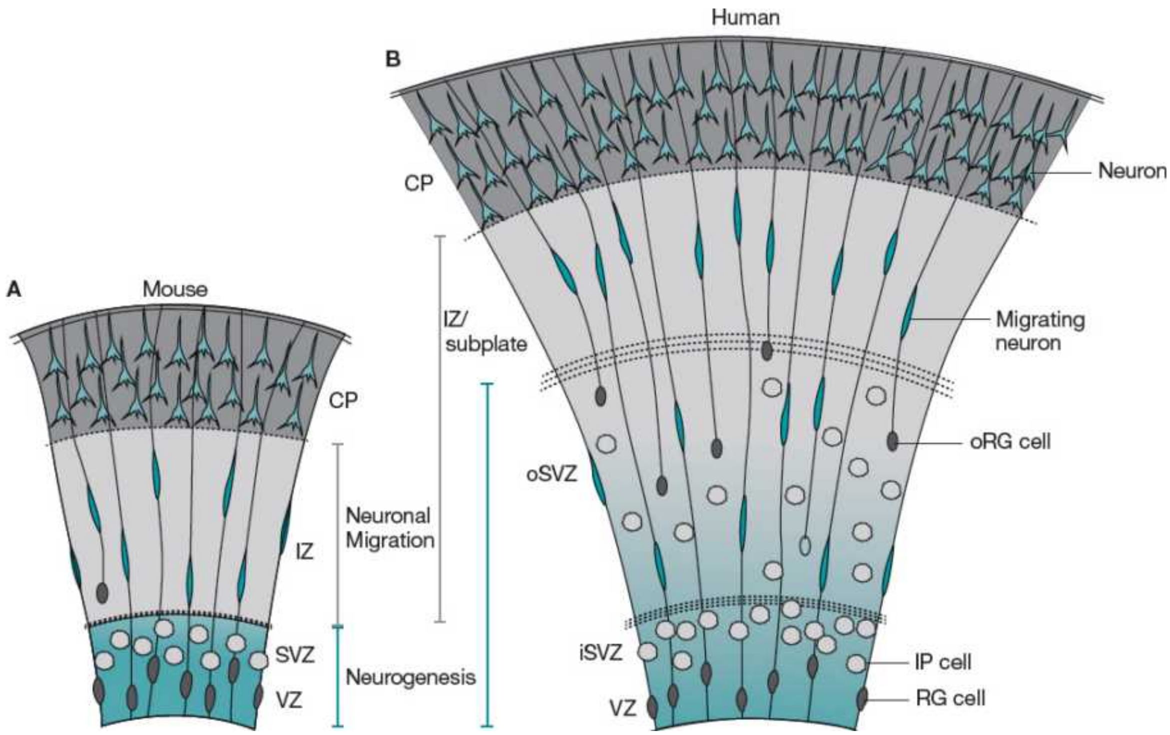

FIGURE 1.4–5. Comparison of the developing mouse and human cortex. **A:** Model of mouse neocortical development. The VZ and SVZ are the primary zones of neurogenesis (*dark gray*) where RG cells and IP cells reside, respectively. Neurons born in the VZ or SVZ migrate along the RG radial fiber scaffold found in the intermediate zone (IZ), the primary zone of neuronal migration (*light gray*) in the mouse, to reach the CP. oRG cells are infrequent and not located in a distinct progenitor zone. **B:** Model of human neocortical development. RG cells, IP cells, and oRG cells are found in neurogenic zones (*dark gray*), which are the VZ, the inner SVZ (iSVZ), and the outer SVZ (oSVZ). Neurons migrate through the OSVZ and IZ/subplate (zones of neuronal migration, *light gray*) to populate the CP. Neurons must navigate a larger distance than in the mouse, and a radial fiber scaffold of increased complexity, to reach the CP. (Reprinted from LaMonica BE, Lui JH, Wang X, Kriegstein AR. OSVZ progenitors in the human cortex: an updated perspective on neurodevelopmental disease. *Curr Opin Neurobiol.* 2012;22(5):747–753, Copyright © 2012, with permission from Elsevier.)

In addition to the general plan of VZ neurogenesis, secondary proliferative zones produce specific neuron populations in particular regions. For example, in cerebral cortex and thalamus, the subventricular zone (SVZ) produces intermediate precursor cells, that can generate oligodendrocytes, diverse astrocytes, and neurons. In hippocampus, the hilus and later the subgranular zone (SGZ) produce dentate gyrus (DG) granule neurons, a lifelong process of adult neurogenesis, though the duration of human adult neurogenesis is a subject of significant controversy (see below). Finally, in newborn cerebellum, the overlying external germinal layer (EGL) generates granule neurons for 7 to 24 months in humans, a population that is affected by medical treatments administered in the neonatal intensive care unit (NICU). In contrast to the VZ, secondary zone cells do not exhibit nuclear movements, suggesting distinct mechanisms of regulation. After neurogenesis is complete, the VZ differentiates into ciliated epithelial cells of the ependymal lining. Underlying the ependyma, undifferentiated cells of the SVZ, referred to as subependyma, have been identified as a neural stem cell (NSC) population, capable of proliferating and generating neurons and glia throughout life. In humans, the magnitude and final destinations of these newly generated neuroblasts are uncertain, but lateral ventricle-derived neuroblasts do migrate to the nearby striatum during adulthood to differentiate into interneurons. Furthermore, their depletion is observed during the advanced stages of Huntington disease.

Similar to the SVZ, humans and other primates have an additional proliferative region in the cortex called the outer subventricular zone (OSVZ), whose extensive proliferation accounts for the massive evolutionary increase in the cerebral cortex. This region first appears during the beginning of the second trimester and becomes the main site of neurogenesis by mid-trimester, surpassing the neurogenic capacity of the SVZ by almost fourfold. The OSVZ contains radial glia-like (oRG) cells that have only one long basal process that contacts the pia, unlike radial glia (RG) of the VZ, which have both apical (touching the ventricle) and basal processes (Fig. 1.4–5B). oRG cells have self-renewing capacity and undergo mitosis during somal translocation in the OSVZ, but do not translocate to the ventricle. They give rise to intermediate progenitor daughter cells that also have proliferative capacity and after migrating up the basal process, differentiate into excitatory neurons of upper cortical layers, II-IV. The OSVZ has a role in forming the gyrencephalic (highly folded) brain, where cortical folding allows the much larger cortical surface area found in higher vertebrates to fit into a limited cranial space. Therefore, the OSVZ appears to have evolved in higher functioning organisms to increase neurogenic capacity in the cortex, allowing for more complex functioning. Recent studies suggest that newly evolved primate genes, including NOTCH2NLA and its paralogs, underlie the expansive neurogenesis in humans by augmenting proliferation in the VZ and OSVZ to increase generation of cortical populations and gyrification.

Radial and Tangential Patterns of Neurogenesis and Migration

The different regions of the brain are generated through three well-recognized spatiotemporal patterns of neurogenesis. There are two radial patterns of precursor cell proliferation and cell migration away from the VZ that are called inside-to-outside and outside-to-inside patterns. The third involves nonradial or tangential migration of cells. Experimentally, one follows the localization of cells that were previously labeled in vivo by incorporating a thymidine analog into DNA that is synthetized prior to cell division. These analogs include tritiated (^{3}H)-thymidine or bromodeoxyuridine (BrdU). When looking at later developmental times, only cells that recently completed mitosis retain the label, because it is lost from cells that continue to divide. The last day a precursor cell incorporates the label is considered its birthday, after which it exits the cell cycle, differentiates, and migrates to its final position.

As a focus of psychiatry, the evolutionarily more recent cerebral cortex is the paradigmatic model of inside-to-outside neurogenesis. Derived from embryonic forebrain telencephalic vesicles, it displays the characteristic six-cell layer organization that represents a common cytoarchitectural and physiologic basis for neocortical function. Within each layer, neurons exhibit related axodendritic morphologies, use common neurotransmitters, and establish similar afferent and efferent connections. In general, pyramidal neurons in layer III establish synapses within and between cortical hemispheres whereas deeper layer V/VI neurons project primarily to subcortical nuclei, including thalamus, striatum, brainstem, and spinal cord (see Chapter 1.2). The majority of cortical neurons, especially glutamatergic excitatory neurons, originate from the forebrain VZ and OSVZ. At the earliest stages of cortex formation, the first postmitotic cells migrate outward from the VZ to establish a superficial layer termed the preplate, which comprises Cajal–Retzius cells that form outermost layer I, and subplate neurons, which lay beneath the future layer VI (Fig. 1.4–6). After its formation the prelate is divided in two by the arrival of the cortical VZ and OSVZ precursors that will form the mature cortical plate. These cortical precursors migrate into the prelate in an inside-to-outside fashion: first layer V/VI neurons arrive, and then later generated neurons migrate past them to form more superficial layers. Thus, the day on which a precursor exits the cell cycle predicts the kind and localization of the neurons generated.

The preplate Cajal–Retzius cells produce the extracellular glycoprotein reelin, an important signal for neuronal migration. When reelin is deleted in mice, cortical neuron migration is inverted so neurons accumulate in reverse pattern, in, outside-to-inside fashion. Significantly, abnormal levels of reelin protein and mRNA have been observed in many disorders, including major depression, bipolar disorder, schizophrenia, autism, and Alzheimer disease. Human reelin mutation is associated with lissencephaly (smooth brain), the absence of gyri and sulci, and abnormalities in cerebellum.

The other preplate component, the subplate neurons are critical for establishing cortical circuitry with distinct prenatal and postnatal roles. Neural circuits in the subplate are initially established by spontaneous neural activity that depends in part on electrical connections between cells mediated by gap junctions composed of connexin proteins. This activity is spontaneous, and is not dependent on nor influenced by external sensory inputs. As development proceeds, early afferent fibers from the thalamus that will later synapse on layer IV cortical neurons first make temporary synapses onto preplate neurons, awaiting arrival of their target cortical neurons. After the thalamic afferents leave the prelate cells, many undergo cell death. Clinically, subplate neurons are particularly vulnerable to perinatal insults during pregnancy and in the NICU, including hypoxia-ischemia, drugs, infection, or inflammation.

In contrast to inside-to-outside neurogenesis, the outside-to-inside pattern occurs in phylogenetically older structures such as hypothalamus, tectum, spinal cord, and hippocampal DG. First formed postmitotic neurons lie superficially and last generated cells localize toward the center. While this pattern might simply reflect passive cell displacement as more cells are generated, radial glia and specific migration signaling molecules are clearly involved. Furthermore, cells do not always lie in direct extension from their locus of VZ generation. Rather, some groups of cells migrate to specific locations, as observed for neurons of the inferior olivary nuclei. In the hippocampus the pyramidal cell layer, Ammon’s Horn CA 1 to 3 neurons, is generated in a typical outside-to-inside fashion in the dorsomedial forebrain for a discrete period, from 7 to 15 weeks of gestation, and exhibits complex migration patterns. In contrast, DG granule neurons start appearing at 18 weeks. Precursors from the forebrain VZ migrate along a subpial route into the DG hilus where they generate both neurons and glia. With time, the granule precursors move to a layer just under the dentate gyrus granule cell layer, termed the subgranular zone, SGZ, which produces neurons throughout life in adult rats, primates, and humans. In rodents, SGZ precursors proliferate in response to cerebral ischemia, tissue injury, seizures, neuropsychiatric drugs, diet, and exercise as well as growth factors. Clinically, the diminished hippocampal volumes reported in schizophrenia and depression raise the possibility that disordered neurogenesis plays a role in pathogenesis, either as a basis for dysfunction, or as a consequence of brain injuries and infections, or environmental and experiential stressors.

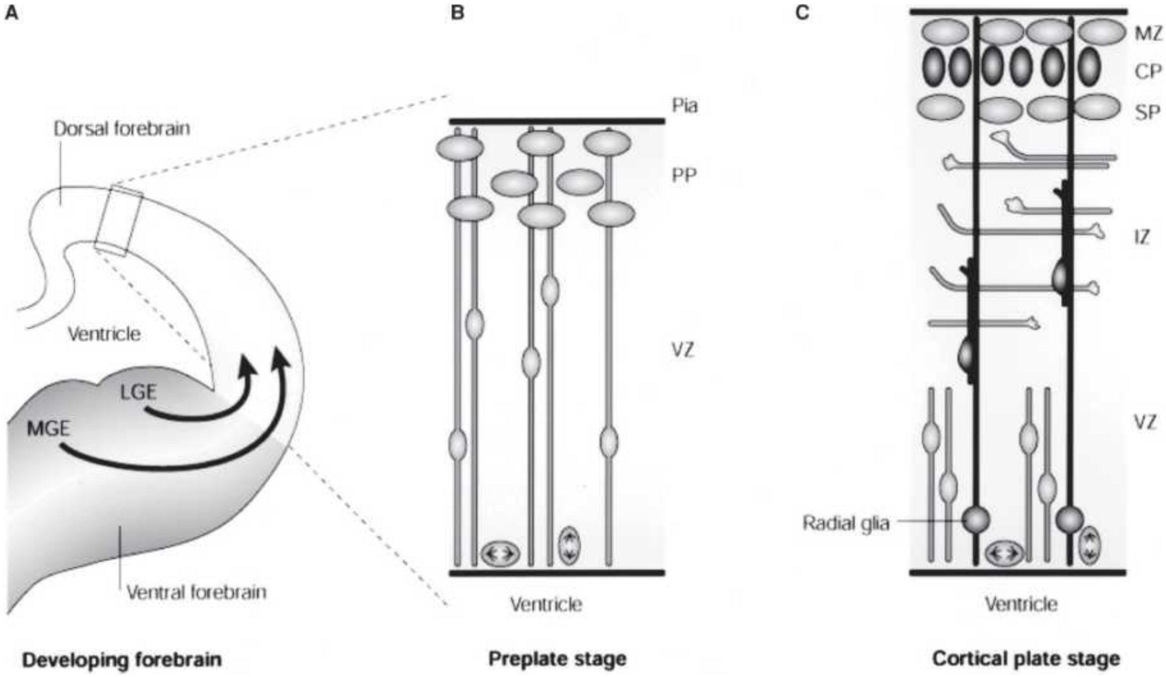

FIGURE 1.4–6. Schematic drawing of radial and tangential migration during cerebral cortex development. **A:** A coronal section of one-half of the developing rat forebrain. The dorsal forebrain gives rise to the cerebral cortex. Medial ganglionic eminences (MGEs) and lateral ganglionic eminences (LGEs) of the ventral forebrain generate neurons of the basal ganglia and the cortical interneurons. The *arrows* indicate the tangential migration route for γ-aminobutyric acid interneruons to the cortex. The boxed area (enlarged in **B** and **C**) shows the developing cortex at early and late stages. **B:** In the dorsal forebrain, the first cohort of postmitotic neurons migrate out from the ventricular zone (VZ) and create a preplate (PP) below the pial surface. **C:** Subsequent postmitotic neurons will migrate along radial glia through the intermediate zone (IZ) and take position in the middle of the preplate, creating a cortical plate (CP) between the outer marginal zone (MZ) and inner subplate (SP). Ultimately, the CP will be composed of six layers that are born sequentially, migrating in an inside-to-outside pattern. Horizontal processes in the IZ represent axon terminals of thalamic afferents. (Reprinted by permission from Springer Nature: Nadarajah B, Parnavelas J. Modes of neuronal migration in the developing cerebral cortex. *Nat Rev Neurosci* 2002;3:423–432, Copyright 2002. doi:10.1038/nrn845.)

A different combination of radial and nonradial migration is observed in cerebellum, which contains ~75 percent of all brain neurons, and that contributes to both motor and nonmotor circuits, with significance for ASDs. Cerebellar development begins at 4 weeks of gestation and continues until 2 years of age. The major GABAergic neurons, including Purkinje cells and deep nuclei, originate from the primary VZ surrounding the fourth ventricle, coincident with other brainstem neurons. This occurs in humans from 5 to 15 weeks of gestation. In contrast, glutaminergic neurons of the cerebellum including granule neurons and some interneurons are generated from precursors that originate in the rhombic lip, a dorsal structure distinct from the VZ that surrounds the roofplate of the fourth ventricle. Rhombic lip precursors migrate dorsally to cover the surface of the cerebellum at 8 to 9 weeks gestation, establishing a secondary proliferative zone, the EGL, which generates neurons with restricted neural fates, and persists until 2 years. When an EGL precursor stops proliferating, the cell body sinks below the surface, grows bilateral processes that extend transversely in the molecular layer, and then the soma migrates further down into the internal granule layer (IGL). Cells reach the IGL along specialized Bergmann glia, which serve guidance functions similar to the radial glia. Clinically, this postnatal population in infants makes cerebellar granule neurogenesis vulnerable to infectious insults of early childhood, and an undesirable target of therapeutic drugs, such as steroids, or cancer chemotherapy, which inhibit cell proliferation. Proliferation control of this stem cell population is lost in the common childhood brain tumor, medulloblastoma (Fig. 1.4–7).

Developmental Cell Death

During nervous system development, cell death is a reproducible, and spatially and temporally restricted death of cells whose elimination is required to coordinate the proportions of interacting neural cells. Three types of cell death have been described: (i) Phylogenetic cell death that removes structures in one species that served evolutionarily earlier ones, such as the tail; (ii) Morphogenetic cell death, which sculpts the fingers from the embryonic paddle and is required to form the optic vesicles as well as caudal neural tube; and (iii) Histogenetic cell death, a widespread process that allows removal of select cells in specific brain regions. Histogenetic cell death can impact between 20 and 80 percent of neurons depending on brain region. A major role for developmental cell death was proposed in the 1980s based on the paradigm of nerve growth factor (NGF). A competitive process was thought to ensure proper matching of new neuronal populations with the size of their final target field. After neurogenesis, survival of differentiating neurons was considered absolutely dependent on axons reaching targets to obtain limiting amounts of the NGF trophic (survival) factor. While demonstrably true for peripheral populations such as sympathetic and sensory neurons, a more complex process regulates cell death during neurodevelopment, and occurs at all stages that precede target innervation, including proliferation, migration, and differentiation of neural precursors. Programmed cell death is a necessary process during neurodevelopment, as genetic deletion of mediating enzymes, the caspases, in embryonic mice produces enlarged and disorganized brains with marked regional specificity.

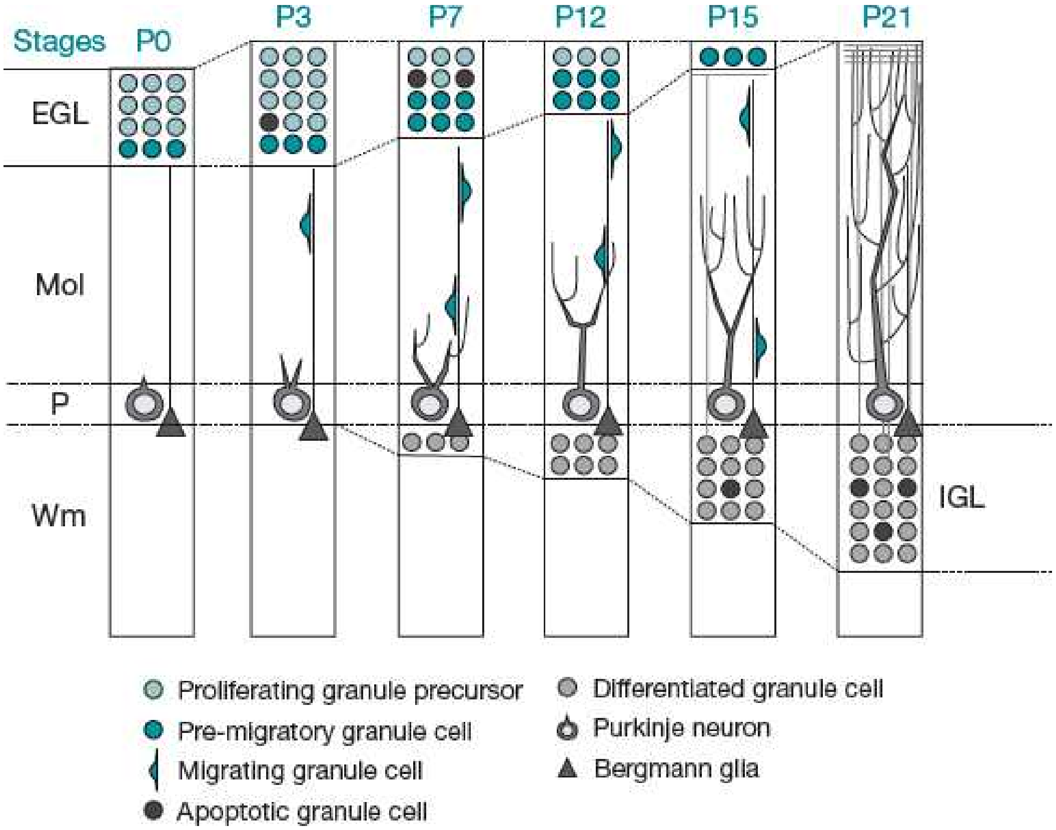
Stages
P0
P3
P7
P12
P15
P21
EGL
Mol
P
Wm
IGL
Proliferating granule precursor
Pre-migratory granule cell
Migrating granule cell
Apoptotic granule cell
Differentiated granule cell
Purkinje neuron
Bergmann glia

FIGURE 1.4–7. Neurogenesis, migration, and differentiation of granule cells during cerebellar development. Granule cell precursors proliferate in the external germinal layer. After exiting the cell cycle, they migrate through the molecular layer and past the Purkinje neurons to reach the internal granule layer where they differentiate and make synapses. Neurons that do not migrate properly or that do not establish proper synaptic connections undergo apoptosis. EGL, external germinal cell layer; Mol, molecular layer; P, Purkinje cell layer; IGL, internal granule cell layer; Wm, white matter.

Three types of programmed cell death based on morphologic criteria have been described. "Apoptotic cell death," is the most common, and is characterized by chromatin condensation and membrane blebbing, followed by nuclear fragmentation and cell shrinkage. "Autophagic degeneration” involves contiguous groups of degenerating cells and features large autophagic vacuoles and pyknotic nuclei. Much less common is “cytoplasmic-type cell death,” that exhibits similarities to necrosis.

As the major type of developmental cell death, apoptotic mechanisms have been extensively examined. Apoptosis involves activation of a series of protease enzymes, the cysteine-containing **asp**artate-specific prote**ases**, also called caspases, that elicits a sequence of intracellular events. A large number of signals (both pro- and antiapoptotic) converge to regulate common signaling pathways that consist of three phases. First, a regulatory phase, termed “initiation,” involves numerous extracellular and intracellular factors that promote either survival or cell death. Then, in the case of cell death ignition, the second phase called “execution” begins. The effector enzymes, caspases 3 and 7 are activated which cleave specific substrates leading to the last and irreversible step of apoptosis. During this final programmed degeneration, caspase-activated DNAse (CAD) cleaves exposed chromosomal DNA located between the nucleosomes, cytoskeletal components are disassembled, and plasma membranes swell into vesicles termed apoptotic bodies. The cell is then dismantled and phagocytosed without any release of its contents, which would otherwise induce an inflammatory response. Historically, two main pathways were described: (i) The “extrinsic pathway,” which mediates effects of death factors such as tumor necrosis factor α (TNF-α) and Fas ligand and involves caspase-8; (ii) The “intrinsic pathway” that involves release of mitochondrial factors, such as cytochrome C, and activation of caspase-9. It is now clear that in most cases cell death results from interactions between multiple factors, involving multiple pro- or antiapoptotic signaling molecules, exerting positive or negative regulation of one another.

Both developmental and pathologic cell death involve many of the same signaling cascades. A failure to inhibit apoptosis is involved in cancers and autoimmune diseases (multiple sclerosis), while excess stimulation of apoptosis is observed in neurodegenerative diseases during development (Huntington disease, lysosomal diseases, leukodystrophy) and aging (Alzheimer’s, Parkinson’s). Apoptosis is also observed during developmental brain injuries such as hypoxia-ischemia, fetal alcohol syndrome, or exposure to ionizing radiations, inhalant anesthetics, and neurotoxicants. Thus dysregulation of apoptotic cell death during development can lead to severe brain abnormalities, which may only become manifest later as mature functional impairments.

THE CONCEPT OF NEURAL PATTERNING

Ever since tools to reveal gene expression in tissues were developed 50 years ago, and specific gene knockout methods were perfected 20 years later, neurobiologists have searched for the genes responsible for determining different brain regions and associated functions. The following sections briefly review the bounty of molecular genetic studies performed primarily in mice, with relevance for human neurodevelopment. However, this focus on defining effects of single genes on brain development almost necessitates that deletions produce major disruptions that are readily and reproducibly observed. These studies delineate the primary interactions of genes expressed in neighboring regions and provide a foundation for regulatory principles. On the other hand, this reductionist approach fails to encompass the greater than 85 percent of protein-coding genes in the human genome (>19,000) that are expressed in the developing brain. Another chapter in this book titled “Functional Genomics of Human Brain Development” describes the gene transcriptional profile of the human brain and the regulatory mechanisms and it is recommended as companion reading.

Principles of Function

Nervous system development is controlled by interactions between extracellular factors and intrinsic genetic programs. Early on extracellular signals are primarily soluble growth factors secreted from regional signaling centers (i.e., notochord, floor or roof plates, surrounding mesenchymal tissues). The precursor’s ability to respond (competence) depends on cognate receptor expression, which is determined by patterning genes whose proteins regulate gene transcription. Patterning genes also control transcription of other genes whose products include other transcription factors or proteins involved in cellular processes (i.e., proliferation, migration, differentiation). Transcription factor proteins contain primarily two domains: a DNA-binding domain that binds promoter or enhancer regions of genes, and a trans-activating domain, which interacts with proteins that can activate or repress gene transcription. Transcription factors frequently form multimeric protein complexes which enable single transcription factors to play roles in multiple cell types and processes, as well as at distinct developmental stages. Consequently, altering expression of a single patterning gene will lead to a diversity of functional outcomes. Furthermore, protein interactions depend on protein–protein affinities. At low concentrations a transcription factor may associate primarily with one partner, but at a higher titer with yet another. The multimeric nature of regulatory complexes allows a single factor to stimulate one process while simultaneously inhibiting another. For instance, a patterning gene protein may promote one event by stimulating one gene promoter, while simultaneously sequestering another factor from a different promoter whose activity is required for an alternative phenotype. Finally, the factors frequently exhibit cross-regulatory activity, where one factor negatively regulates expression of another. This activity leads to the establishment of tissue boundaries, allowing the formation of regional subdivisions.

Patterning genes act in hierarchical fashion by exhibiting distinct temporal sequences of expression and function. To define such hierarchies, researchers use genetic manipulations (either deleting a gene [loss of function] or over-/ectopically expressing it [gain of function]) and observe the developmental consequences. In general, regionally restricted patterning genes participate in specifying the identity, and therefore function, of cells in which they are expressed. These genes regulate multiple developmental decisions including: (1) determining that ectoderm will give rise to nervous system (as opposed to skin), (2) defining the dimensional character of a region, such as positional identity in dorsoventral or rostrocaudal axes, (3) specifying cell class, such as neuron or glia, (4) defining when proliferation ceases and differentiation begins, (5) determining specific cell subtype, such as GABA interneuron, as well as projection pattern, and (6) defining laminar position, such as observed in cerebral cortex.

Transcription factors that pattern the nervous system belong to protein families highly conserved throughout evolution, with many discovered in Drosophila where they mediate body and organ segmentation and morphogenesis, or regulate neural development. In mammals, the Hox family delineates the anterior–posterior axis from tail to midbrain, defining segments of the hindbrain (rhombomeres) and its cranial nerves. The basic helix-loop-helix (bHLH) family, binding DNA and proteins through the basic and helix regions, respectively, regulates multiple stages from neural plate to neurogenesis. Other families bear names reflecting protein interaction domains and include LIM homeodomain (*Lhx*), zinc finger, paired domain (*Pax*), winged helix (*BF1 = Foxg1, Foxa2 = Hnf3ß*), and *Pou.*

Finally, patterning gene expression is regulated by environmental signals, especially growth factors released from regional signaling centers. Focal growth factor release establishes tissue gradients of these proteins that provide positional information (dorsal or ventral), impart cell specification, and/or control proliferative growth. Examples include the BMPs, the Wingless-int proteins (Wnts), Shh, fibroblast growth factors (FGFs), and epidermal growth factors.

Patterning the Embryonic Forebrain

In this section, we briefly review some of the transcription factors that determine brain region identities, also noting the roles of signaling center growth factors. Even at the third week of gestation, the neural plate stage, induction of the CNS begins when the underlying notochord and mesenchyme, and surrounding epidermal ectoderm produce signaling molecules that affect neighboring cell identity. Perhaps surprisingly, neural differentiation is the default state of early embryonic tissues unless it is inhibited. To prevent its own neural differentiation, the neighboring ectoderm produces BMPs that instead promote epidermal differentiation. To elicit neural differentiation, Hensen node, a signaling center at the rostral end of the primitive streak, secretes inhibitory proteins that block the BMP-mediated epidermis-inducing activity. Once the neural tube closes, the roof plate and floor plate become new signaling centers, organizing dorsal and ventral neural tube, respectively. As stated previously, the same ligand/receptor system is used sequentially for multiple functions during development. As a case in point are the BMPs: they prevent neural development at neural plate stage, while after neurulation, the factors are produced by the dorsal neural tube/roof plate itself to induce sensory neuron fates.

The dorsal forebrain structures include the hippocampus medially, cerebral cortex dorsolaterally, and entorhinal cortex ventrolaterally, whereas in basal forebrain, the globus pallidus lies medially and striatum laterally. In the mouse embryo, the forebrain is divided into a checkerboard-like pattern of transcription factors arranged in longitudinal columns and transverse segments that produce unique combinations within each subdivision (Fig. 1.4–3). Many of the genes, including*, Emx2, Pax6, FoxA2,* and *Dlx2,* are first expressed even before neurulation in the neural plate and are then maintained, comprising the "protomap" determinants of the VZ. Initial forebrain gene expression is influenced by an array of soluble factors derived from signaling centers, including (1) FGF8 and Shh from the anterior neural ridge (cranial mesenchyme), (2) BMPs and Wnts from midline structures, including cortical hem and commissural plate, and (3) FGF7, Wnt antagonists, and neuregulins and retinoic acid from the lateral junction of the dorsal and ventral palliums, the anti-hem. Shh also derives from the ventral midbrain region (Fig. 1.4–8). The influence of these localized sources diminishes as brain size increases.

As indicated, regional patterning genes including bHLH and homeodomain proteins frequently exhibit cross-regulatory activity, where one factor negatively regulates expression of its neighbors (Fig. 1.4–3). Thus loss of one factor leads to expansion in the boundaries of others. For example, the rostral/lateral cortex expresses high levels of homeodomain gene *Pax6,* whereas caudal/medial cortex exhibits *Emx2, Lhx2,* and *Lhx5.* A prediction would be that altering gene expression should cause a change in cortical areas, especially the proportions of motor to sensory cortex. Indeed, motor cortex markers are markedly diminished in mice mutant for *Pax6,* as is downstream bHLH transcription factor, *Ngn2*. Conversely, there is proportionate increase in caudal sensory cortex markers. Moreover, there are also changes in the dorsoventral dimension: Genes of the ventral striatum and pallidum, *Gsh* and *Dlx,* are now expressed ectopically in dorsal territory. A similar dorsal shift occurs with deletion of other dorsal transcription factors, *Ngn1/2* and *Gli3,* yielding loss of the cerebral cortex. Within the ventral forebrain, mutation of Nkx2.1 protein results in the acquisition of striatal-like properties by the presumptive pallidum, suggesting that Nkx2.1 both specifies the medial ganglionic (MGE; globus pallidus) fate and inhibits lateral ganglionic eminence (LGE) phenotypes. These observations indicate that patterning genes exert reciprocal inhibitory functions in several dimensions.

As expected, mutations in patterning genes and signaling molecule pathways lead to human brain malformations that are also observed in rodent models. *PAX6* deletion results in abnormalities of the eyes (cataracts, aniridia, anophthalmia) and the olfactory epithelium and bulbs.

Furthermore, the cerebral cortex is hypoplastic, exhibiting nodules of poorly differentiated cells adjacent to proliferative zones. Disruption of roof plate signaling causes holoprosencephaly (HPE), characterized by a single forebrain ventricle with a continuous cerebral cortex across the midline. HPE is linked to mutations in several components of the Shh inductive cascade, including *SIX3* and *ZIC2*. Both these transcription factors are upstream of Shh gene expression. HPE is also the consequence of mutations in the Shh Patched (*PTC*) receptor gene, and *TGIF,* a downstream component of the TGF-β family. Shh and *SIX3* are co-expressed in the anterior neural ridge and the ventral midline, whereas *Zic2* is expressed in the dorsal roof plate. *SIX3* and *SHH* are also mutated in some cases of schizencephaly. Finally, *FOXG1* (*BF1*) gene mutations cause microcephaly, corpus callosum dysgenesis and delayed myelination, associated with severe intellectual disability, motor delay, dyskinetic movements, and epilepsy.

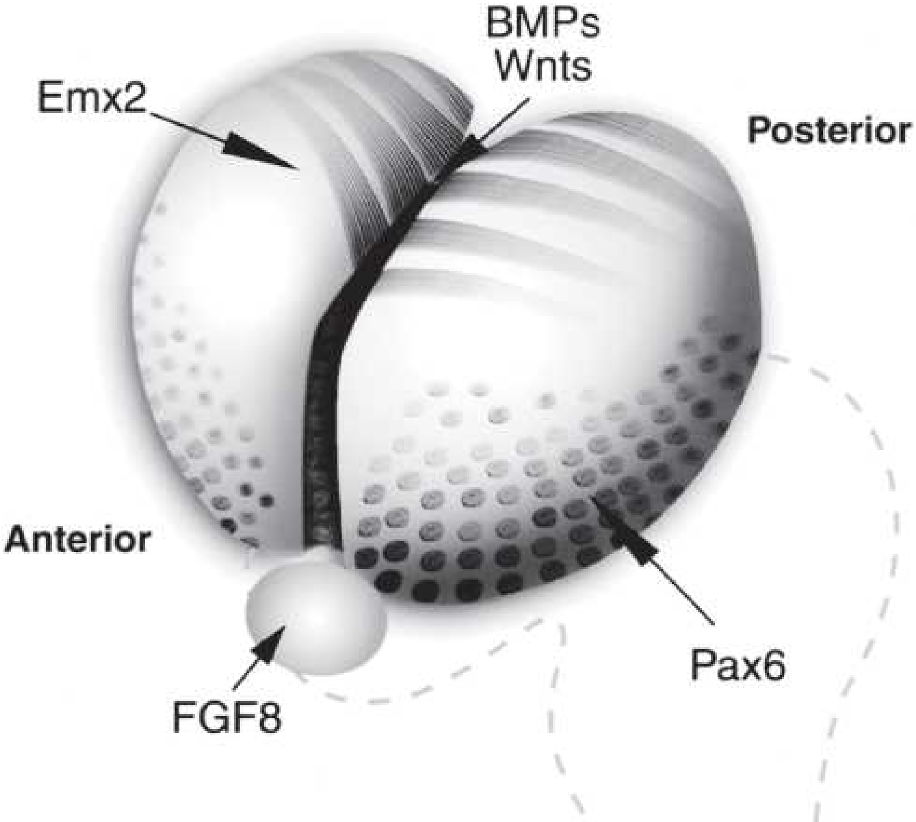

FIGURE 1.4–8. Patterning genes and signaling centers in the developing cerebral cortex. This schematic diagram shows a lateral-superior view of the two cerebral hemispheres of the embryonic mouse, sitting above the midbrain and hindbrain (*broken lines*). *Circles* indicate the anterior-lateral extent of Pax6 gene expression. The posterior-medial domain of Emx2 expression is indicated by stripes. The genes exhibit continuous gradients of expression that decrease as they extend to opposite poles. The signaling factor fibroblast growth factor 8 (FGF8) is produced by and released from mesenchymal tissue in the anterior neural ridge, which regulates Pax6 and Emx2 expression. In the midline, bone morphogenetic proteins (BMPs) and Wingless-Int proteins (Wnts) are secreted from other signaling centers, including the roof plate and the cortical hems. (Courtesy of E. DiCicco-Bloom and K. Forgash.)

Given the substantial role of molecular determinants (“protomap”) in regulating areal specification, do ingrowing thalamic afferents contribute to cortical development? In two different mouse mutants of ventral forebrain genes, *Gbx2* and *Mash1,* thalamocortical innervation is disrupted, yet the expression of cortical patterning genes is not altered, suggesting they have a limited prenatal role. However, after birth, thalamic axons interact with local regional cues and modify their outgrowth to establish proper connections that conform to cortical molecular identities. Furthermore, the developing cortex exhibits a remarkable level of flexibility: in the ferret, eliminating the thalamic lateral geniculate nucleus to which retinal ganglion axons project causes them to reroute to the auditory cortex, which then successfully mediates vision!

Layer-Specific Generation of Neurons during Cortical Development. Despite early determination of cortical areal identities by "protomap" patterning genes, the cerebral cortex exhibits a common six-layered cytoarchitecure that was classically described by histologic analysis, and is now characterized by "layer-specific" molecular markers. Many of the markers, which are highly conserved in mammals and primates including humans, are transcription factors that determine when neural precursors exit the cell cycle and or initiate neuron- and layer-specific cell differentiation. After neural precursors commit to a postmitotic fate, the diversity of cortical projection neurons suggests the critical importance of accurately timing the activation of intrinsic genetic programs. Glutamatergic projection neurons in the cortical layers target specific brain structures. Lower-layer (V–VI) neurons mainly project subcortically, with axons establishing synapses in the thalamus, brain stem, and spinal cord. Upper-layer (II–IV) neurons project intracortically, either within the ipsilateral hemisphere or to the contralateral hemisphere via the corpus callosum. During this period, upper-layer neurons express *Cux1/2*, *Satb2*, and *Brn1*, whereas lower-layer neurons exhibit *Ctip2*, *Foxp2*, *Tle4*, and *Fezf2* (Fig. 1.4–9). As with cortical areal determinants, genetic studies using gain- and loss-of-function approaches indicate their fate-specifying roles and their cross-regulatory actions. For example, in upper-layer neurons, when *Satb2* is deleted, axons do not extend to contralateral cortex via the corpus callosum, but instead project to subcortical targets, just like lower layer neurons. Indeed, these upper-layer neurons up-regulate levels of lower-layer factor, *Ctip2*. Conversely, reduction of *Ctip2* in lower-layer neurons prevents their normal projection to the spinal cord. Similarly, silencing of lower-layer *Fezf2* transcription factor also prevented corticospinal tract axons from reaching their targets, whereas ectopic overexpression of *Fezf2* in upper-layer neurons mis-directed axons to subcortical sites. Thus the cerebral cortex exhibits interacting as well as sequential arrays of transcription factors during development.

The Hippocampus

Genetic experiments have defined patterning genes localized to the cortical hem and hippocampal primordia, whose deletions result in a variety of morphogenetic defects. In mice lacking growth factor Wnt3a expressed in the cortical hem, the hippocampus is either completely missing or greatly reduced, while neighboring cerebral cortex is preserved. A similar phenotype results from deleting intracellular factor downstream of Wnt receptor activation, the *Lef1* gene, suggesting the Wnt3a-*Lef1* pathway is required for hippocampal cell specification and/or proliferation. When another cortical hem gene, *Lhx5* (also *Lhx2),* is deleted, mice lack both the hem and neighboring choroid plexus, both sources of growth factors. However, in this case, cortical hem cells proliferate in excess, and the hippocampus is disorganized due to abnormalities in proliferation, migration, and differentiation. Finally, a sequence of bHLH transcription factors plays roles in postnatal hippocampal neurogenesis: DG differentiation is defective in *NeuroD* and *Mash1* mutants. Significantly, all hippocampal-patterning genes are regulated by locally secreted factors including FGF8, Shh, BMPs, and Wnts.

The Basal Ganglia and the Origins of Inhibitory Interneurons

The genes patterning the ventral forebrain regions were described above, and featured in Figure 1.4–3. However, in addition to motor and cognitive functions, the basal ganglia are important because they are the embryonic origin of GABA interneurons that reach the neocortex through tangential migration (Fig. 1.4–6). As a simplified model in mice, the MGE expresses primarily *Nkx2.1* and gives rise to GABA interneurons of the cortex and hippocampus, whereas the LGE expresses *Gsh2* and generates GABA interneurons of the SVZ and olfactory bulb. When *Nkx2.1* is deleted, LGE transcription factor expression spreads ventrally into the MGE, and there is 50 percent reduction in neocortical and striatal GABA interneurons. In contrast, deletion of *Gsh2* leads to ventral expansion of the dorsal cortical molecular markers and concomitant decreases in olfactory interneurons. Finally, *Pax6* mutation causes both MGE and LGE to spread laterally and into dorsal cortical areas, yielding increased interneuron migration.

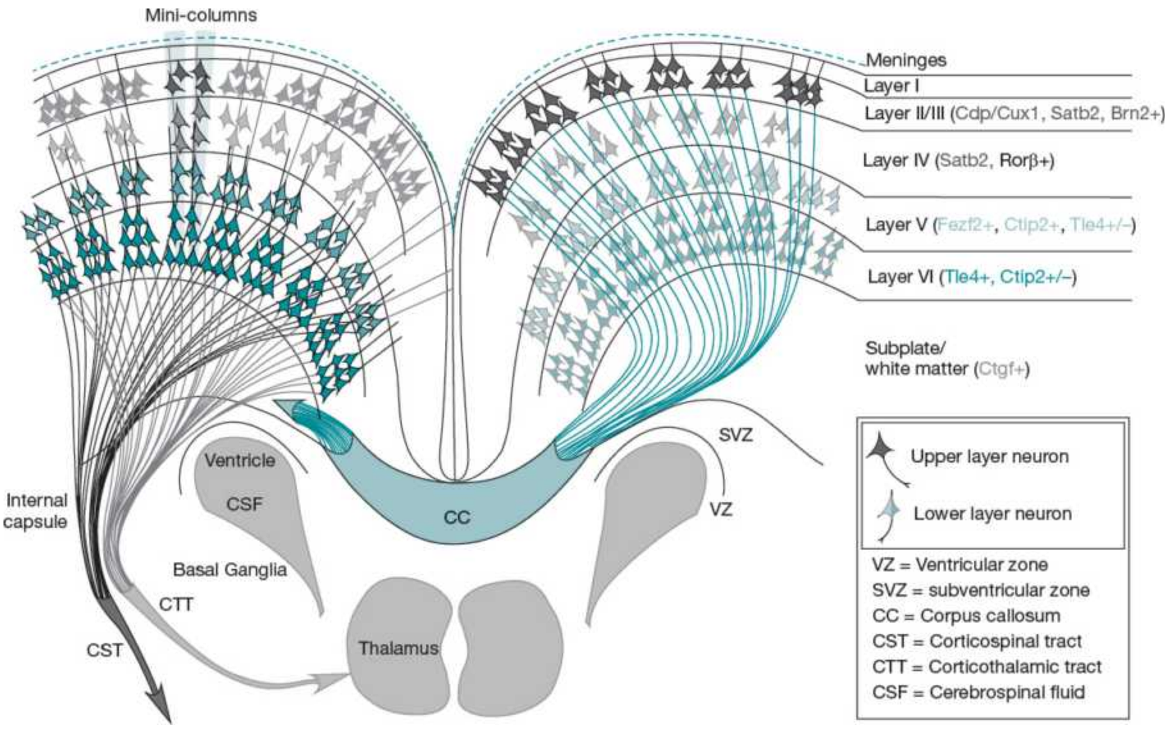
Mini-columns
Meninges
Layer I
Layer II/III (Cdp/Cux1, Satb2, Brn2+)
Layer IV (Satb2, Rorβ+)
Layer V (Fezf2+, Ctip2+, Tle4+/–)
Layer VI (Tle4+, Ctip2+/–)
Subplate/
white matter (Ctgf+)
SVZ
Ventricle
CSF
CC
VZ
Internal capsule
Basal Ganglia
CTT
CST
Thalamus
Upper layer neuron
Lower layer neuron
VZ = Ventricular zone
SVZ = subventricular zone
CC = Corpus callosum
CST = Corticospinal tract
CTT = Corticothalamic tract
CSF = Cerebrospinal fluid

FIGURE 1.4–9. A simplified schematic of the laminar organization of neocortical projection neurons. Glutamatergic projection neurons represent approximately 85 percent of all neocortical neurons and exhibit a six-layered format. Neuron subpopulations are characterized by specific molecular identities, dendritic morphologies, and terminal targets corresponding to each layer. Projection neurons that are born during later stages of prenatal neurogenesis localize to upper layers II–IV, express specific transcription factors such as CDP/Cux1, and project solely intracortically, forming the corpus callosum and ipsilateral intrahemispheric tracts. In contrast, earlier born projection neurons localize to lower layers V–VI, express TLE4 and FEZF2, and project subcortically to form the corticothalamic tract (CTT) or corticospinal tract (CST). In the subventricular zone (SVZ), adult progenitors are found giving rise to olfactory cortex neurons. (Reprinted from DeBoer EM, Kraushar ML, Hart RP, Rasin MR. Post-transcriptional regulatory elements and spatiotemporal specification of neocortical stem cells and projection neurons. *Neuroscience* 2013;248:499–528. Copyright © 2013 IBRO.)

Other transcription factors in the MGE and LGE (Fig. 1.4–3) regulate both the timing and type of interneurons generated. *Mash1* is expressed in early born cells, whereas *Dlx1/Dlx2* appears in later maturing neurons, having as targets other family members, *Dlx5/Dlx6.* In the *Dlx1/Dlx2* double knock out, there is 75 percent reduction in neocortical interneurons, and complete absence in the hippocampus, while olfactory neurons are preserved. A regulatory cascade has been suggested since *Mash1* can regulate *Dlx* expression, while *Dlx2* can induce expression of the GABA synthetic enzyme, glutamic acid decarboxylase. Indeed, *Mash1* deletion mutants exhibit reduced cortical GABA interneurons and striatal cholinergic interneurons. Similarly, *Nkx2.1* loss leads to complete absence of all cortical interneurons expressing NPY, somatostatin, and NOS. These studies suggest that transcription factors play roles at multiple stages in neuronal production including generic neuronal fate specification, as well as neuron subtype determination.

Neuronal Specification

Transcription factors regulate decisions of neural cell type, neuron or glial cell, as well as neuron subtypes. For example, *Mash1* can promote a neuronal over a glial fate, as well as induce the GABA interneuron phenotype. *Neurogenin2 (Ngn2)* can promote glutamatergic excitatory neuron fate. However, another bHLH factor, *Olig1/2,* can promote oligodendrocyte development in the forebrain, whereas it promotes motor neuron differentiation in the spinal cord, indicating that the variety of factors expressed in a specific cell leads to combinatorial effects and diverse outcomes for cell differentiation.

REGULATION OF NEURODEVELOPMENT BY EXTRACELLULAR FACTORS

Extracellular factors that regulate cell proliferation, migration, differentiation, and survival interact with intrinsic genetic determinants to control region-specific neurogenesis. This section focuses on precursor proliferation as one example. Patterning genes control the expression of growth factor receptors and the molecular machinery of the cell division cycle. Growth factors stimulate or inhibit proliferation of VZ precursors and are termed as autocrine, paracrine, and endocrine, meaning they originate from the cells themselves, neighboring cells/tissues, and from the general circulation, respectively. Although defined initially in cell culture, mitogenic growth factors are now well characterized in vivo. There are those that stimulate proliferation, such as bFGF, Wnt, IGF-I, Shh, as well as signals that inhibit cell division, including leukocyte inhibitory factor (LIF), pituitary adenylate cyclase–activating polypeptide (PACAP), GABA, glutamate, interferon-gamma, and glial-derived neurotrophic factor. A recent integrated transcriptomic and proteomic analysis of embryonic cortical precursors and neurons by Yuswa et al. (2016) has newly identified growth factors that elicit cell cycle exit while promoting neurogenic differentiation. In addition to acting as mitogens, extracellular signals also act as trophic factors, which enhance proliferation by promoting survival of the mitotic population. Stimulation of both pathways is necessary to produce maximal cell numbers. Several of the neurotrophins, especially BDNF and neurotrophin-3 (NT3), promote survival of mitotic precursors as well as the newly generated progeny.

Altering the expression and activity of mitogens and their receptors during development can elicit profound and permanent consequences. For example, growth factor administration to developing rodent embryos or pups induces changes in proliferation in prenatal cortical VZ, postnatal cerebellar EGL, and hippocampal DG that produce lifelong alterations in brain region size, cell composition, and function. Environmental factors (hypoxia, maternal/fetal infection, drug or toxicant exposure) can also dysregulate mitogenic pathways and produce major deficits. In humans, the previous practice of steroid treatments to promote lung maturation for mothers with preterm deliveries was discontinued when negative effects on cerebellar granule neurogenesis and cognition were clarified. Similarly, cancer regimens including chemotherapy and radiation have long been known to inhibit perinatal neurogenesis. Changes in regional neurogenesis may be relevant to structural differences observed in neuropsychiatric disorders: abnormalities in FGF ligand and receptor system expression have been found consistently in brain samples from those with major depressive disorders (MDDs), as well as many animal models of stress and depressive behavior. Finally, mitogens regulating cerebellar granule neurogenesis in animals include bFGF, Shh, and PACAP, whereas genetic mutations in both *SHH* and *WNT* determine the severity and prognosis of medulloblastoma in children.

It is well known that neurologic development may be delayed in children experiencing serious systemic illness that is associated with numerous inflammatory cytokines, and one may wonder whether this reflects interference with neurogenesis and concomitant processes, potentially producing long-term differences in cognitive and motor development. Similarly, maternal infection during pregnancy is a known risk factor for schizophrenia and possibly autism, and cytokines that cross the placental barrier may directly affect fetal brain cell proliferation and differentiation as well as cell migration, axon target selection, and synapse maturation as shown in animal models, eventually leading to multiple brain and behavioral abnormalities in the adult offspring.

Cell Migration

The movement of immature neurons from proliferative areas to their definitive positions in the brain is an essential process of neurodevelopment. The orientation and directionality of cell migration can be described as radial (cortex), tangential (inhibitory GABA interneurons from MGE/LGE), or through a migratory stream (granule interneurons of the olfactory bulb). Regardless of migration type, there are guidance cues that influence the direction, speed, and cell behavior of these motile cells. These cues are broadly categorized as: extracellular factors, adhesion molecules, and cytoskeletal-regulating molecules.

Much of our knowledge about these guidance cues originates from animal models. For example, the best characterized examples are mice with deficient levels of the extracellular factor reelin, a large, secreted glycoprotein produced embryonically by the earliest neurons in the cortical preplate, Cajal–Retzius cells. The mouse cortex exhibits dysregulated neuronal migration resulting in inverted cortical laminar structure, generated in outside-to-inside fashion. Further, there is ectopic hippocampal neurogenesis, resulting from aberrant migration into the hilus and molecular layer. Reelin binds and activates two receptors, the very–low-density lipoprotein receptor (VLDLR) and the apoprotein E receptor 2 (ApoER2), and the intracellular adapter protein, disabled 1 (Dab1). The roles of the VLDL and ApoE2 receptors are intriguing considering their well-known contributions to genetic risk for Alzheimer disease. Interestingly, postmortem brain protein levels have been reported as abnormal in schizophrenia and autism, though recent genomic studies do not reveal genome-wide significance for reelin gene variant in disease risk. Another extracellular factor important for neuronal migration is the polypeptide hormone, insulin-like growth factor 1 (IGF-1). Knock out mouse models have revealed a suspension of progenitors in a multipolar state, resulting in failed neuronal migration to upper cortical layers.

Adhesion molecules that participate in neuronal migration can be divided into cell–matrix and cell–cell interacting molecules. Cell–matrix adhesion molecules, such as integrin, participate in mediating the binding of cell surfaces and different ligands at the extracellular matrix. As for cell–cell adhesion molecules, there are the cadherins, which mediate

Ca^{2+}-dependent cell–cell adhesion. N-cadherin in particular is turned on during neurulation and its absence results in disorganized radial glia and intermediate progenitor cells across the cortex. Ca^{2+}-independent cell–cell adhesion molecules include nectins and nectin-like molecules, which are immunoglobulin-like cell adhesion proteins. These molecules are expressed in the hippocampus, cerebellum, and neocortex, and have been identified as components of the reelin-signaling pathway to instruct radial migration of Cajal–Retzius cells.

Many of the molecules described above regulate the cytoskeleton, which is critical for neurite outgrowth and migration. The neuronal cytoskeleton is comprised of actin, intermediate filaments, and microtubules (MTs). Actin filaments play a critical role in cell migration, with particular importance for growth cone exploration (discussed in next section). The dynamic nature of actin filaments is regulated by Rho GTPases, including Rac1, RhoA, Rnd2, and cdc42. Rnd2 inhibits RhoA to promote radial migration and neurite extension, whereas Rac1 and cdc42 play more of a role in axonal guidance and polarity. MTs are dynamic, polarized intracellular filaments that are highly abundant in neurons and generate the energy needed to propel the nucleus forward to aid in migration. Missense mutations in both murine and human MT subunit, *Tuba1a/TUBA1A*, results in an accumulation of neurons in proliferative regions, which fail to emigrate out, and are associated with several neuropsychiatric disorders. Further, MTs are regulated by microtubule-associated proteins (MAPs), which play a major role in ensuring that there is normal progression of cell migration and outgrowth. Disruptions in extracellular factors, adhesion molecules, and cytoskeletal regulators can lead to malformations in cortical development. For example, patients with lissencephaly often have brains devoid of sulci and gyri, originating from defects in neuronal migration. This disorder can be caused by single mutations in four different genes (*DCX, RELN, LIS1, ARX*) and a partial chromosomal deletion (Chr 17p—Miller–Dieker syndrome). Mutations in doublecortin (*DCX*), a well-characterized MAP, results in X-linked lissencephaly in males and bands of heterotopic neurons in white matter in females, appearing as a “double cortex” on imaging studies, and producing severe mental retardation and epilepsy. Studies utilizing human-induced pluripotent stem cell models revealed that loss of *LIS1* results in altered neuronal migration by dysregulating MT stability and their motor protein, dynein. Mutations in *DCX* and *LIS1* are suggested to account for about 75 percent of lissencephaly cases while *TUBA1A* accounts for approximately 4 percent of cases.

Differentiation and Neuronal Process Outgrowth

Once newly produced neurons and glial cells reach their final destinations, differentiation into a mature state can begin, peaking around 5 months of gestation. For neurons, this involves outgrowth of dendrites and extension of axonal processes, formation of synapses, and production of neurotransmitter systems, including receptors and selective neurotransmitter reuptake sites. This outgrowth begins with a break in polarity, where the distal tip of an immature neurite begins to protrude and respond to guidance cues. This structure is known as the growth cone, which was discovered and named by Santiago Ramón y Cajal.

The structure of the cone is subdivided into two domains: the central domain, which contains mitochondria and MTs, and the peripheral domain, containing veil-like lamellipodia and spike-like filopodia and populated primarily by actin (Fig. 1.4–10). Growth cone protrusion and motility is largely driven by the cyclical polymerization and depolymerization of actin filaments. When there is more polymerization or myosin activity, the growth cone extends, whereas depolymerization and retrograde movement results in retraction of the growth cone. Abnormalities in actin-modulating proteins (Arp2/3, Profilin, Cofilin) and related kinases (LIM, SRC) involved in rapid growth cone movements may contribute to language and cognitive disorders, such as *LIMK* kinase whose mutation causes Williams syndrome. During the first several years of life, many neuronal systems exhibit exuberant process growth and branching, which is later decreased by selective “pruning” of axons and synapses dependent on experience, while myelination continues for several years after birth and into adulthood.

Much of growth cone movement is influenced by a variety of attractive and repulsive extracellular guidance cues, including secreted extracellular matrix proteins, such as netrins, slits, and semaphorins, and cell–cell adhesion molecules, such as ephrins, cadherins, and nectins. For example, region-specific expression of guidance molecules, such as cadherins, is regulated by patterning genes *Pax6* and *Emx2.* These cues interact with filopodial receptors to cause growth cones to move forward, turn, or retract. Netrin proteins serve as bifunctional cues, where they can induce attractive and repulsive signals depending on their receptors.

Netrin-1 in particular is involved in the migration of GABAergic interneurons, while Netrin-4 influences the maturation of cortical neurons. Slit molecules act via binding to Robo receptors and act primarily as a repulsive axon guidance cue, enabling commissural axons to cross the midline only once, ensuring proper projection. Similar to Slits, Semaphorins also serve as repulsive cues for developing commissural axons. When their expression is altered, aberrant axonal projection occurs from failed pathfinding.

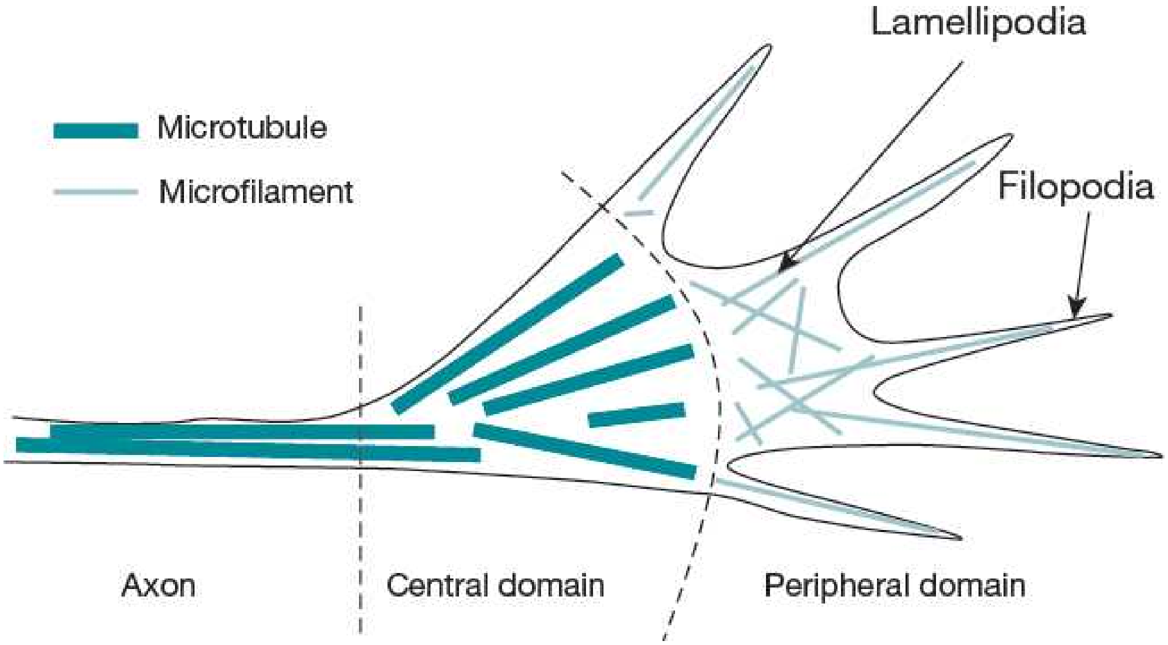

FIGURE 1.4–10. Structure of the growth cone. The cone is subdivided into two domains: the central domain, which contains mitochondria and microtubules, and the peripheral domain, containing veil-like lamellipodia and spike-like filopodia. In the lamellipodia, microfilaments consisting of actin form a meshwork, while in filopodia, they have the same orientation. Cell surface receptors on growth cone processes sense extracellular guidance cues to control navigation.

Ephrins frequently serve as chemorepellent cues, negatively regulating growth by preventing developing axons from entering incorrect target fields. For example, the optic tectum (superior colliculus) expresses ephrins A2 and A5 in decreasing gradients from posterior to anterior, whereas innervating retinal ganglion cells express a gradient of Eph receptors. Ganglion cell axons from posterior retina, which possess high Eph A3 receptor levels, preferentially innervate the anterior tectum because the low-level ephrin expression does not activate the Eph kinase that causes retraction. The role of cadherins and nectins in regulating axonal guidance and target recognition has been more deeply characterized in structures outside of the neocortex, including the spinal cord and motor or sensory nerves. However, a recent study demonstrated that absence of N-cadherin in thalamic and cortical co-cultures resulted in altered innervation patterns of thalamocortical axons, confirming the role of cadherins in the guiding axonal pathfinding.

POSTNATAL NEUROGENESIS IN MAMMALS: SUBVENTRICULAR ZONE AND HIPPOCAMPAL DENTATE GYRUS

Adult neurogenesis is the generation of new neurons from neural precursors in the brain. This process is restricted to two regions: the SVZ lining the lateral ventricles, and a narrow proliferative zone underlying the DG granule cell layer termed the subgranular zone, SGZ, in the hippocampus (Fig. 1.4–11). It is a highly plastic process that can be altered by various physiologic, pathologic, environmental, and pharmacologic stimuli. In mice, rodents, and monkeys, newly produced neurons migrate from the SVZ in an anterior direction toward the olfactory bulb to become GABA (and dopamine)-containing interneurons, and also into striatum under certain conditions.

Studies in rodents have demonstrated that adult neurogenesis is regulated by extracellular signals (such as EGF, bFGF, IGF-I, BDNF, LIF/CNTF) that control precursor proliferation and or cell survival. For example, peripheral administration of both bFGF and IGF-I stimulate neurogenesis, selectively increasing mitotic labeling in the SVZ and hippocampal SGZ, respectively. More classical studies were performed by application of growth factors via intracerebroventricular injections and demonstrated direct cellular actions. Interestingly, models of pregnancy indicate that both prolactin and oxytocin hormones enhance maternal neurogenesis which contributes to learning of their offspring identity. These studies typically utilize thymidine analogs BrdU and EdU, which can be injected into living animals and then detected by immunohistochemistry. Use of endogenous cell markers can also be used to label proliferating cells, including Ki-67 and proliferating cell nuclear antigen (PCNA).

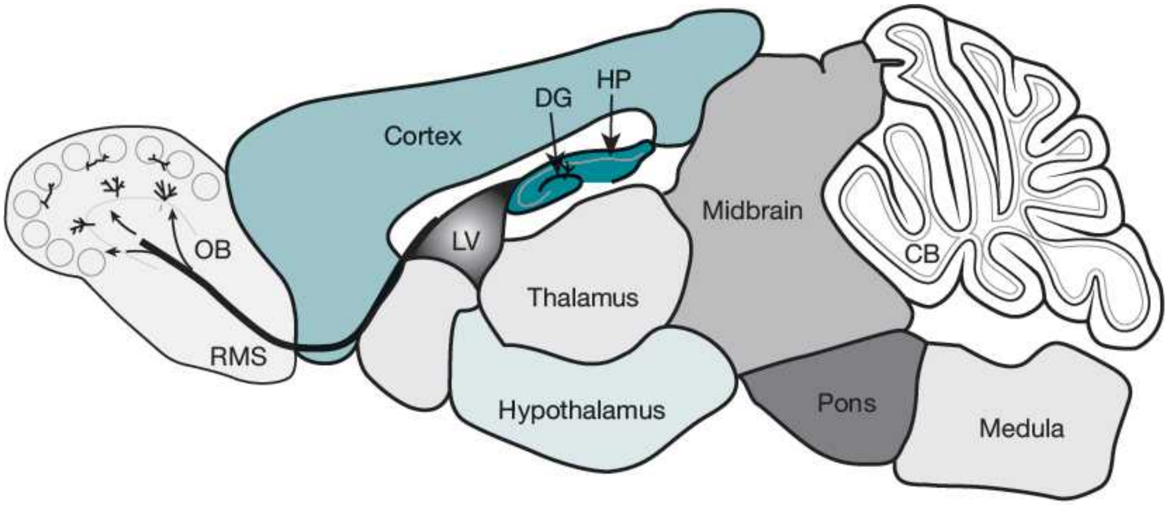

Factors that stimulate neurogenesis	Factors that inhibit neurogenesis
Antidepressants/Antipsychotics	Stress/Glucocorticoids
Exercise	Isolation
Enriched environments	Sleep deprivation
Oxytocin/Prolactin	Drugs of abuse

FIGURE 1.4–11. Adult neurogenesis in the hippocampal dentate gyrus and the subventricular zone. A sagittal section view of an adult rodent brain highlighting the two restricted regions that exhibit active adult neurogenesis: dentate gyrus (DG) in the hippocampal formation (HP), and the lateral ventricle (LV) subventricular zone to the rostral migratory stream (RMS) to the olfactory bulb (OB). (Adapted from Ming G, Song H. Adult neurogenesis in the mammalian brain: Significant answers and significant questions. *Neuron* 70(4):687–702. Copyright © 2011, with permission from Elsevier.)

As discussed in earlier sections, doublecortin (DCX) is a MAP expressed by neuronal precursor cells and immature neurons. This marker in particular has been extensively used to characterize the possibility of human adult hippocampal neurogenesis (AHN) in postmortem brain tissue. Improvements in methodology and resources have allowed researchers to decipher this complex question, which has led to an ongoing debate within the last few years. A handful of labs have concluded to find thousands of immature neurons (DCX+ cells) in variable stages of maturation within the DG of the hippocampus of healthy humans up to the ninth decade of life. Although these studies have used state-of-the-art technology and brain tissue from healthy humans, they have reported that detection of AHN markers in the human DG are reduced when brain tissue is in a fixative solution for a prolonged period of time, which is typically the case at brain banks worldwide. Studies that refute the hypothesis have used similar techniques and reported that numbers of young neurons decreased in density by age 13 years to undetectable levels in adulthood (18+). However, the analyzed brain tissue samples included those from humans with epilepsy, which often results in destruction of the neurogenic niche, perhaps contributing to the lack of AHN reported. Additional studies that take the history of the sample into account (neuropsychiatric disease, tissue fixation solution and time) and include rigorous methods (unbiased stereology, bioinformatics, non-invasive *in vivo* studies in humans) are needed to adequately support or refute the possibility of AHN.

The discovery of adult neurogenesis naturally leads to questions about whether new neurons can integrate into the complex cytoarchitecture of the mature brain, and the functional significance. Extensive analysis has shown that newly born neurons indeed integrate into hippocampal circuits and establish typical synaptic connections. From a functional perspective, the generation and/or survival of new neurons correlate strongly with multiple instances of behavioral learning and experience. For example, survival of newly generated neurons is markedly enhanced by hippocampal-dependent learning tasks and by an enriched, behaviorally complex environment. Of perhaps greater importance, a reduction in DG neurogenesis impairs the formation of trace memories, that is, when an animal must associate stimuli that are separated in time, a hippocampal-dependent task. From clinical and therapeutic perspectives, scientists have observed that changes in neurogenesis can contribute to disease or help the brain recover from injury. A neurogenetic response has been shown for multiple pathologic conditions in the adult, including brain trauma, stroke, and epilepsy. For instance, ischemic stroke in the striatum stimulates adjacent SVZ neurogenesis (Fig. 1.4–11), with neurons migrating to the injury site. These studies raise the possibility that newly produced neurons normally participate in recovery, and may be stimulated as a novel therapeutic strategy. However, neurogenesis can also play roles in pathogenesis. In a kindling mouse model of epilepsy, newly generated neurons migrated to incorrect positions and formed aberrant neuronal circuits, re-enforcing the epileptic state. Human samples from surgical resection of the DG in epilepsy reveal similar abnormalities. Conversely, reductions in neurogenesis can contribute to several conditions that implicate dysfunction or degeneration of the hippocampal formation. DG neurogenesis is inhibited by increased glucocorticoids levels observed in aged rats and is reversed by steroid antagonists and adrenalectomy. These observations are potentially relevant to the correlation of elevated human cortisol levels with reduced hippocampal volumes and the presence of memory deficits. Similarly, stress-induced increases in human glucocorticoids may contribute to decreased hippocampal volumes seen in schizophrenia, depression, and posttraumatic stress disorder (PTSD). Some studies have shown that antidepressants, antipsychotics, and/or physical exercise can reverse the decreased neurogenesis phenotype. For example, administration of the selective serotonin reuptake inhibitor (SSRI) antidepressant, fluoxetine, in mice led to an increase in neural progenitor cells, which eventually became new neurons in the hippocampus. Although successful antidepressant treatment response in humans has been associated with increased hippocampal volume on structural MRI analysis, evidence supporting a causative role of AHN in humans in clinical trials will be needed.

FURTHER READINGS

Alcamo EA, Chirivella L, Dautzenberg M, et al. Satb2 regulates callosal projection neuron identity in the developing cerebral cortex. *Neuron*. 2008;57(3):364–377.

Alvarez-Buylla A, Seri B, Doetsch F. Identification of neural stem cells in the adult vertebrate brain. *Brain Res Bull*. 2002;57:751–758.

Bishop KM, Goudreau G, O'Leary DD. Regulation of area identity in the mammalian neocortex by Emx2 and Pax6. *Science*. 2000;288:344–349.

Britanova O, de Juan Romero C, Cheung A, et al. Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. *Neuron*. 2008;57(3):378–392.

Cheng Y, Black IB, DiCicco-Bloom E. Hippocampal granule neuron production and population size are regulated by levels of bFGF. *Eur J Neurosci*. 2002;15:3–12.

de Agustín-Durán D, Mateos-White I, Fabra-Beser J, Gil-Sanz C. Stick around: cell-cell adhesion molecules during neocortical development. *Cells*. 2021;10(1):118.

DeBoer EM, Kraushar ML, Hart RP, Rasin MR. Post-transcriptional regulatory elements and spatiotemporal specification of neocortical stem cells and projection neurons. *Neuroscience.* 2013;248:499–528.

DiCicco-Bloom E, Lord C, Zwaigenbaum L, et al. The developmental neurobiology of autism spectrum disorder. *J Neurosci*. 2006;26:6897–6906.

Falluel-Morel A, Sokolowski K, Sisti HM, et al. Developmental mercury exposure elicits acute hippocampal cell death, reductions in neurogenesis, and severe learning deficits during puberty. *J Neurochem*. 2007;103:1968–1981.

Fiddes IT, Lodewijk GA, Mooring M, et al. Human-specific NOTCH2NL genes affect notch signaling and cortical neurogenesis. *Cell*. 2018;173(6):1356–1369.e22.

Garcia-Moreno F, Molnar Z. Variations of telencephalic development that paved the way for neocortical evolution. *Prog Neurobiol*. 2020;194:101865.

Guillemot F. Cell fate specification in the mammalian telencephalon. *Prog Neurobiol*. 2007;83:37–52.

Hansen DV, Lui JH, Parker PR, Kriegstein AR. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature*. 2010;464:554–561.

Hevner RF. Layer-specific markers as probes for neuron type identity in human neocortex and malformations of cortical development. *J Neuropathol Exp Neurol*. 2007;66:101–109.

Jin SC, Lewis SA, Bakhtiari S, et al. Mutations disrupting neuritogenesis genes confer risk for cerebral palsy. *Nat Genet*. 2020;52(10):1046–1056.

Kempermann G, Song H, Gage FH. Neurogenesis in the adult hippocampus. *Cold Spring Harb Perspect Biol*. 2015;7(9):a018812.

Mairet-Coello G, Tury A, DiCicco-Bloom E. Insulin-like growth factor-1 promotes G(1)/S cell cycle progression through bidirectional regulation of cyclins and cyclin-dependent kinase inhibitors via the phosphatidylinositol 3-kinase/Akt pathway in developing rat cerebral cortex. *J Neurosci*. 2009;29(3):775–788.

Molnar Z, Luhmann HJ, Kanold PO. Transient cortical circuits match spontaneous and sensory-driven activity during development. *Science*. 2020;370(6514):eabb2153.

Moreno-Jiménez EP, Terreros-Roncal J, Flor-García M, Rábano A, Llorens-Martín M. Evidences for adult hippocampal neurogenesis in humans. *J Neurosci*. 2021;41(12):2541–2553.

Nicot A, Lelièvre V, Tam J, Waschek JA, DiCicco-Bloom E. Pituitary adenylate cyclase-activating polypeptide and sonic hedgehog interact to control cerebellar granule precursor cell proliferation. *J Neurosci*. 2002;22:9244–9254.

Noctor SC, Flint AC, Weissman TA, Dammerman RS, Kriegstein AR. Neurons derived from radial glial cells establish radial units in neocortex. *Nature*. 2001;409:714–720.

O'Leary DDM, Borngasser D. Cortical ventricular zone progenitors and their progeny maintain spatial relationships and radial patterning during preplate development indicating an early protomap. *Cereb Cortex*. 2006;16(Suppl 1):i46–i56.

Prem S, Millonig JH, DiCicco-Bloom E. Dysregulation of neurite outgrowth and cell migration in autism and other neurodevelopmental disorders. *Adv Neurobiol*. 2020;25:109–153.

Rakic P. Decision by division: making cortical maps. *Trends Neurosci*. 2009;32(5):291–301.

Rakic P. Evolution of the neocortex: a perspective from developmental biology. *Nat Rev Neurosci*. 2009;10(10):724–735.

Rossman IT, Lin L, Morgan KM, et al. Engrailed2 modulates cerebellar granule neuron precursor proliferation, differentiation and insulin-like growth factor 1 signaling during postnatal development. *Mol Autism*. 2014;5(1):9.

Schuursmans C, Guillemot F. Molecular mechanisms underlying cell fate specification in the developing telencephalon. *Curr Opin Neurobiol*. 2002;12:26–34.

Shors TJ, Miesegaes G, Beylin A, Zhao M, Rydel T, Gould E. Neurogenesis in the adult is involved in the formation of trace memories. *Nature*. 2001;410:372–376.

Sorrells SF, Paredes MF, Zhang Z, et al. Positive controls in adults and children support that very few, if any, new neurons are born in the adult human hippocampus. *J*

Neurosci. 2021;41(12):2554–2565.

Yuzwa SA, Yang G, Borrett MJ, et al. Proneurogenic ligands defined by modeling developing cortex growth factor communication networks. *Neuron*. 2016;91(5):988–1004.

1.5 Cellular and Synaptic Basis of Neural Signaling

CHARLES F. ZORUMSKI, M.D., AND STEVEN MENNERICK, PH.D.

Neuropsychiatric disorders involve dysfunction in intercellular communication and neural networks. Although these defects often reflect synaptic communication between neurons, changes in the intrinsic excitability of neurons also contribute. Thus, pharmacologic and neuromodulation therapies aimed at neuronal excitability are becoming commonplace in the treatment of neurologic and psychiatric disorders. Furthermore, there is growing appreciation that psychotherapy alters specific aspects of neural communication to effect benefit.

Neurons use electrical signals to send and receive information. These electrical signals regulate the brain's local and network properties and involve the flow of ions across cell membranes through macromolecular ION CHANNELS. Neurons express two major classes of ion channels, gated and nongated. Nongated (or leakage) channels are open at rest and contribute to the cellular RESTING MEMBRANE POTENTIAL. The opening and closing of most ion channels are regulated (GATED) by changes in transmembrane voltage, extracellular chemicals, or intracellular messengers. VOLTAGE-GATED SODIUM CHANNELS open and close rapidly and provide the basis for communication within and between neurons. Rapid signaling via sodium channels results in "all or none" ACTION POTENTIALS that are generated near the neuronal cell body and transmitted with little decrement in amplitude along axons to nerve terminals. This high-fidelity signal propagation reflects action potentials regenerated by voltage-gated channels along the length of the axon. In myelinated axons, action potential propagation is enhanced by

SALTATORY CONDUCTION, the ability of electrical signals to “jump” between axonal nodes of Ranvier.

At nerve (synaptic) terminals, the wave of action potential–mediated depolarization opens VOLTAGE-GATED CALCIUM CHANNELS, and the influx of calcium promotes the release of a chemical neurotransmitter into the extracellular space, where the transmitter then influences a receptive cell. NEUROTRANSMITTERS bind two broad classes of RECEPTORS and alter neuronal excitability via actions on ion channels. LIGAND-GATED ION CHANNELS are directly opened by neurotransmitter binding, whereas G-PROTEIN–COUPLED RECEPTORS (GPCRs) influence ion channels indirectly via guanine nucleotide-binding proteins (G-proteins) or intracellular chemical messengers. This chapter outlines the basic principles of cellular electrophysiology and discusses the importance of electrical signals in brain function and neuropsychiatric illnesses.

PRINCIPLES OF CELLULAR ELECTROPHYSIOLOGY

Resting Membrane Potential

The concentration of potassium ion [K^+] is much higher inside cells than outside, reflecting the selective permeability of most cell membranes at rest to K^+. The expression of nongated (leak) K^+ channels in the cell membrane is the basis for this selective permeability. Potassium channels are a class of membrane-spanning proteins that have a hydrophilic pore that selectively conducts K^+ across the membrane through open pores of leak channels. Under baseline conditions, K^+ ions are attracted into the cell by the presence of large, impermeant intracellular anions (acids and proteins). As K^+ accumulates in the cell, the transmembrane electrical potential (electrical charge separation) of the cell becomes depolarized (less negative) and results in less K^+ entry because of electrorepulsion (like charges repel). Intracellular concentrations of K^+ reach levels of ~100 mM while extracellular [K^+] is between 2 and 6 mM; this unequal distribution establishes a chemical gradient, which, by itself, would result in net K^+ efflux from the cell. Thus, there are two gradients that act on K^+: intracellular electronegativity promoting K^+ influx and a chemical gradient driving K^+ efflux. At a specific membrane potential (around −90 mV), the electrical and chemical gradients for K^+ are equal and opposite. This membrane potential is known as the EQUILIBRIUM POTENTIAL (also called the NERNST POTENTIAL) for K^+. The equilibrium potential for an ion species is the transmembrane potential at which electrical and chemical gradients are balanced, and there is no net flux of the ion. In a cell whose membrane is exclusively permeable to K^+, the resting potential of the cell would be equal to the equilibrium potential for K^+.

The situation in most neurons is more complex because other ions, with different electrochemical gradients, are slightly permeant through ion channels that are open in the resting cell membrane. Each of these ions has its own characteristic equilibrium (Nernst) potential depending on ion concentrations inside and outside the cell. The cations Na^+ and Ca^{2+} are present at higher concentrations outside than inside the cell. Thus, at negative membrane potentials, both electrical and chemical gradients for these cations are inwardly directed, and the equilibrium potentials are positive to 0 mV. Chloride (Cl^-) concentrations are usually higher outside the cell, but because of this ion's negative charge, the equilibrium potential for chloride is near the resting potential (Fig. 1.5–1). The actual resting potential of a cell is determined by the average of the equilibrium potentials of all the permeant ions, weighted by the relative permeability of each ion. At rest, K^+ and Cl^- are much more permeant than other ions, so the resting potential is closest to the equilibrium potentials for these ions. Na^+ and Ca^{2+} are less permeant and contribute less to the resting potential, but the small permeability of nongated channels to these cations renders the actual resting potential more positive than the equilibrium potentials of K^+ or Cl^-. Typical values for neuronal resting potentials are between −55 and −70 mV. Astrocytes, by contrast, have a membrane more purely permeable to K^+ and, therefore, a more deeply negative resting membrane potential (∼−90 mV), very close to the K^+ equilibrium potential.

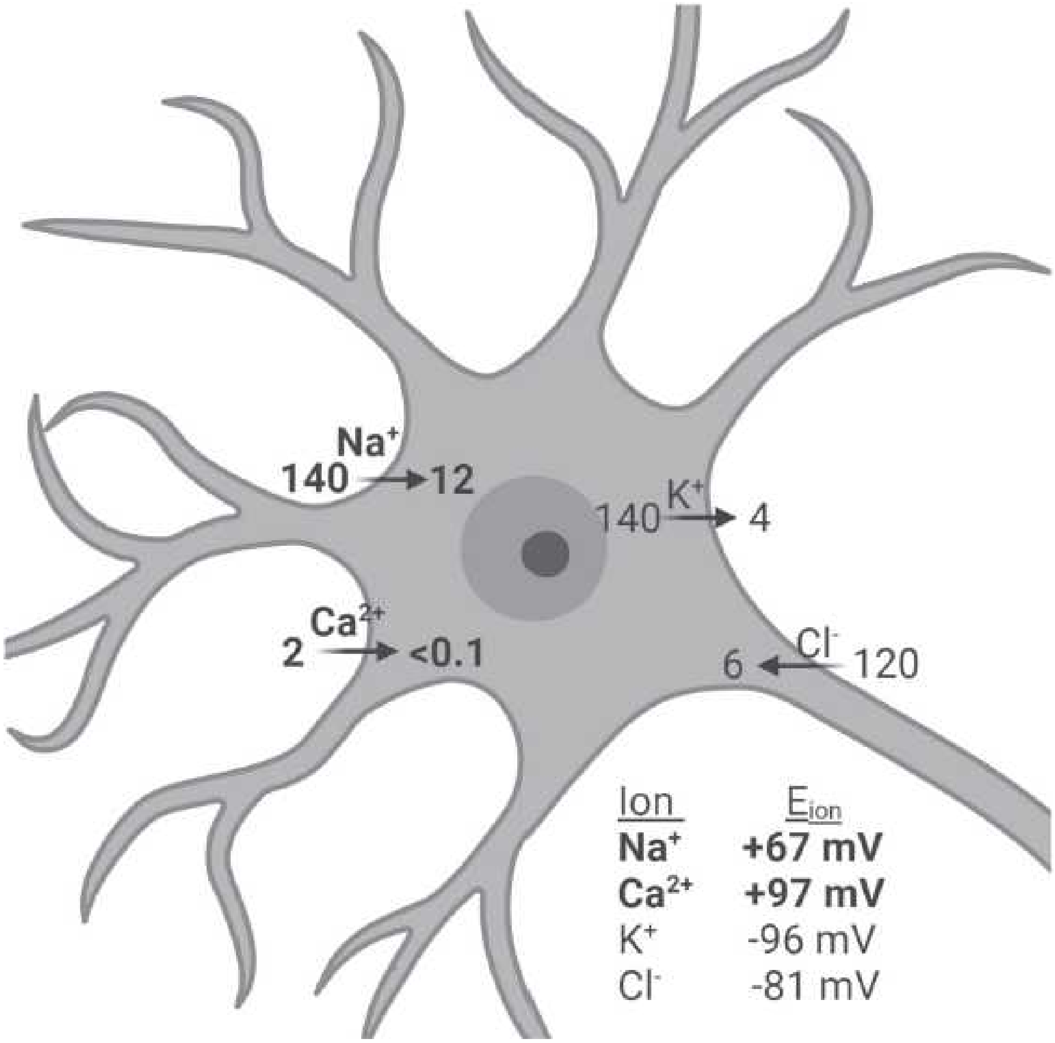

FIGURE 1.5–1. The figure shows the distribution of Na$^+$, K$^+$, Ca^{2+}, and Cl$^-$ across the membrane of a neuron; the *arrows* show the direction of the current flow down the electrochemical gradient. Depolarizing influences are shown on the **left**; hyperpolarizing influences are shown on the **right**. At the indicated ion concentrations, the equilibrium (Nernst) potentials at 37 °C are shown at the lower **left**. Excitatory influences are bolded. (Created with BioRender.com.)

The concepts of equilibrium potential and membrane potential outlined above are described quantitatively by the Nernst equation. The equilibrium potential for any ionic species is calculated based on the ion concentrations on either side of the membrane. For K$^+$, the equilibrium potential (designated E_K) is expressed as: $\mathbf{E_K = (RT/zF) \times \ln([K]_o/[K]_i)}$, where **R** is the ideal gas constant (8.31 joules/degree/mole), **T** is the temperature in degrees Kelvin, **z** is the valence of the ion, F is Faraday's constant (96,500 coulombs/mole, the charge on a mole of monovalent ions), and **$[K]_o$** and **$[K]_i$** are the concentrations of K^+ outside and inside the cell. At 37 °C, the equilibrium potential for K^+ is −96 mV, E_{Na} is +67 mV, E_{Cl} is −81 mV, and E_{Ca} is greater than +97 mV. These equilibrium potentials are important in determining how the membrane potential changes when an ion channel that is permeable to a specific ion species opens or closes because the opening of a specific ion channel drives the membrane potential toward the equilibrium potential for that ion. For example, when K^+-selective ion channels open, the neuronal membrane potential moves toward its equilibrium potential of −96 mV. This makes the inside of the cell more negative, an effect called HYPERPOLARIZATION. Na^+ and Ca^{2+} channel opening have the opposite effect, making the inside of the cell less negative (DEPOLARIZATION).

Because resting cell membranes are permeable to more than one ion, the true membrane potential is never exactly equal to the equilibrium potential for any one ion. The Goldmann/Hodgkin/Katz (GHK) equation describes the actual resting potential as the average of the various equilibrium potentials, weighted by the relative permeability of each ion. The equation is of the form:

$$E_m = \frac{RT}{F} \ln \frac{P_k[\mathrm{K}]_o + P_{Na}[\mathrm{Na}]_o + P_{Cl}[\mathrm{Cl}]_i}{P_k[\mathrm{K}]_i + P_{Na}[\mathrm{Na}]_i + P_{Cl}[\mathrm{Cl}]_o}$$

Most of the variables are familiar from the Nernst equation described above. E_m is the membrane potential, and P_{ion} is the permeability of the membrane to the ion. The resting membrane potential is thus a reversal potential (potential at which no net inward or outward current flows) for the various conductances open at rest.

The bulk ionic solutions on either side of the membrane are electrically neutral, with most intracellular negative charge contributed by large intracellular organic anions (acids and proteins). The differential distribution of ions across neuronal membranes is maintained by membrane pumps that use energy from hydrolysis of adenosine triphosphate (ATP) to move ions against their concentration gradients into or out of the cell. The best-characterized pump is the Na^+-K^+ ATPase (SODIUM PUMP) that transports 3 Na^+ out of and 2 K^+ into the cell during each cycle. Because an unequal amount of charge is moved during each cycle, the pump is ELECTROGENIC and contributes to intracellular negativity of the membrane potential. Na^+-K^+ ATPase activity is a major driver of brain energy utilization, with as much as 40% of brain oxygen consumption reflecting the pump activity required to re-establish ionic homeostasis following action potential firing and synaptic transmission. The cardiac glycosides, digoxin, and ouabain inhibit Na^+-K^+ ATPase in the heart and improve myocardial contractility by depolarizing cardiac myocytes and increasing intracellular Ca^{2+}.

The resting membrane potential is relatively static and represents the potential energy available for neuronal signaling. Negative resting potentials are not unique to excitable cells, but neurons and other excitable cells make unique use of the energy stored as the resting potential to generate transient membrane potential *changes*, the true currency of neuronal information exchange. A membrane potential change is typically initiated by a change in current flow across the membrane, usually resulting from the opening or closing of specific ion channels discussed below. The number of ions needed to change the membrane potential is very small relative to concentrations in bulk intracellular and extracellular solutions. For example, a potential change of 100 mV across a 1 cm^2 area of membrane requires the movement of only about 10^{-12} moles of a monovalent (singly charged) ion. By comparison, Na^+ and K^+ are present at about 10^{-1} M in the extra- and intracellular fluids, respectively.

Passive Membrane Properties

One way to conceptualize how electrochemical gradients, ion channels, and the unequal distribution of charges across the membrane are related is to view the cell membrane as an electrical circuit of resistors (conductors), batteries, and capacitors (Fig. 1.5–2). The use of the equivalent circuit allows the prediction of how the membrane potential will change in response to stimulation or channel opening. Because ions flow through ion channels and do not directly penetrate cell membranes, ion channels can be viewed as variable resistors. These ion channels are characterized by their SELECTIVITY (which ions flow through the channel) and their CONDUCTANCE (relative ease of passing ions). Conductance (g) is the inverse of RESISTANCE (R) in an electrical circuit ($g = 1/R$). The different types of ion channels are part of an electrical circuit in parallel with each other and with the membrane capacitance. The presence of a voltage across the membrane provides an electrical force driving ions through their channels and resulting in a transmembrane

CURRENT. The relationship among voltage (V), ionic current (I), and resistance (conductance) is given by a version of Ohm's law: $\mathbf{I_{ionic} = g(V_m - E_{rev})}$, where $\mathbf{V_m}$ is the membrane potential, $\mathbf{E_{rev}}$ is the equilibrium (Nernst) potential for the ion(s) flowing through the channel, and $\mathbf{(V_m - E_{rev})}$ is the DRIVING FORCE for ion flow.

CAPACITANCE is another important passive membrane electrical property. A capacitor is a device with two conductors separated by an insulating material that can store charges of opposite signs on the two conductors. In neurons, the conductors are the extra- and intracellular fluids, while the lipid membrane is the insulator. Whenever current flows through the membrane, some current must charge the membrane capacitance (C_m). The expression describing this capacitive current is: $\mathbf{I_{cap} = C_m(dV/dt)}$. Note that capacitive current flows only when the membrane potential is changing (i.e., there is some change in voltage [**dV**] over time [**dt**]). The total current flowing across a membrane at any given time is a sum of $\mathbf{I_{cap}}$ and $\mathbf{I_{ionic}}$. The membrane capacitance, along with the leak conductance of the membrane at rest, helps to define the low-pass filtering property of neurons. A major tool used to study ionic currents is a VOLTAGE CLAMP, often implemented using the PATCH CLAMP method. These techniques use specialized amplifiers to hold membrane potential constant and eliminate capacitive currents by preventing changes in voltage, making it possible to measure ionic currents, currents through ion channels, directly.

Active Membrane Properties: Action Potentials

Changes in membrane potential have important effects on excitability because some ion channels are regulated (gated) by voltage changes. When neurons depolarize, specific Na^+ channels open rapidly and drive the membrane potential toward the Na^+ equilibrium potential (~+66 mV). Because of leakage channels that are open at rest, there is initially a balance between leakage currents and the currents flowing through Na^+ channels opened by depolarization. At a specific depolarized membrane potential, the current flowing through Na^+ channels exceeds the current through the leak channels. The membrane potential at which Na^+ currents exceed leakage currents is called the THRESHOLD POTENTIAL, typically between −45 mV and −30 mV in neurons. At membrane potentials depolarized from the threshold, the entry of more Na^+ produces further depolarization and opens more Na^+ channels in a positive-feedback cycle. The membrane potential becomes more positive than 0 mV but never reaches the Na^+ equilibrium potential for three reasons. First, leak channels continue to play a role in determining the membrane potential during the action potential. Because of the relative K^+ selectivity of these channels, the membrane potential never reaches the Na^+ equilibrium potential. Second, during depolarization, Na^+ channels not only activate but also rapidly inactivate. INACTIVATION is a process by which voltage-gated channels enter a nonconducting state despite the continued presence of the activating stimulus (in this case, depolarization). Third, the depolarization produced by Na^+ entry also opens voltage-gated K^+ channels that drive the membrane potential toward the K^+ equilibrium potential (~ -90 mV). The net effect of activation and inactivation of Na^+ channels and delayed opening of voltage-gated K^+ channels is that the neuronal membrane potential rapidly changes to values more positive than 0 mV and then returns rapidly to the resting membrane potential. This rapid sequence occurs over several milliseconds and is called an ACTION POTENTIAL (or sodium SPIKE) (Fig. 1.5–3). The fact that the membrane potential transiently exceeds 0 mV is called an OVERSHOOT. To a first approximation, action potentials represent all-or-none increases in electrical excitability and are important contributors to information transfer within and between neurons. Because they propagate (see below), action potentials allow the neuronal cell body to communicate rapidly with its axon terminals. In axon terminals, the spike provides the depolarization that promotes Ca^{2+} channel opening and hence Ca^{2+}-dependent release of neurotransmitters.

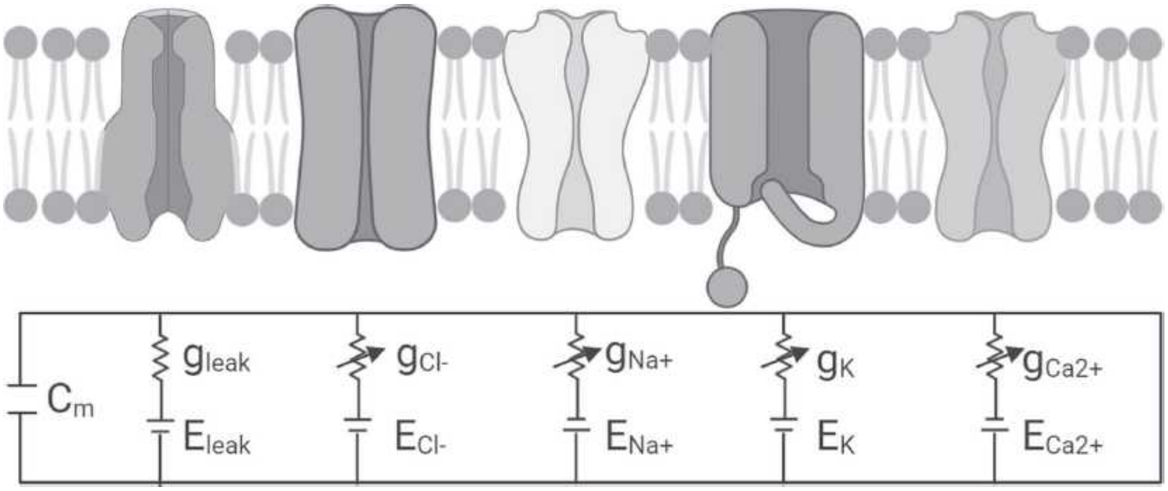

FIGURE 1.5–2. Ion channels are protein pores that span the lipid bilayer of cell membranes (**upper panel**). Because of the selective permeability of the membrane to potassium and because of membrane pumps, the intracellular surface of the membrane has a net negative charge with respect to the extracellular surface. The figure shows a membrane schematic that includes major leak and/or voltage-activated ion channels and the predominant direction of ion flux under physiologic conditions. As a result of the transmembrane potential and the presence of ion channels, the neuronal membrane can be described as an equivalent electrical circuit in which each ion channel is a variable resistor (conductor, g_x) in series with a battery (E_x) (**lower panel**). The leak conductance (g_{leak}) is represented as a constant resistance. Different ion channels are electrically in parallel with each other and in parallel with the membrane capacitance (C_m). (Created with BioRender.com.)

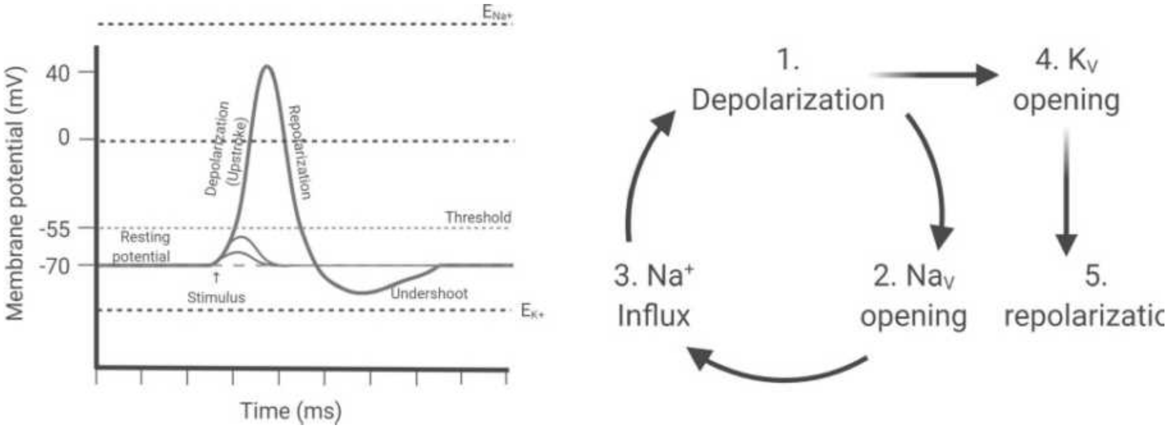
E_Na+
40
0
-55
-70
Membrane potential (mV)
Depolarization (Upstroke)
Repolarization
Threshold
Resting potential
Stimulus
Undershoot
E_K+
Time (ms)
1. Depolarization
4. K_V opening
2. Na_V opening
3. Na⁺ Influx
5. repolarizatic

FIGURE 1.5–3. The trace shows a neuronal action potential as recorded by an intracellular microelectrode. The portions of the action potential are described in the text. The right portion of the figure shows the sequence of events underlying an action potential. Na_V and K_V represent voltage-gated sodium channels and potassium channels, respectively. (Created with BioRender.com.)

In most neurons, the K^+ equilibrium potential is negative with respect to the resting membrane potential. Thus, action potentials are often followed by a transient UNDERSHOOT (or AFTERHYPERPOLARIZATION [AHP]) resulting from open voltage-gated K^+ channels that gradually close as the membrane hyperpolarizes (see Fig. 1.5–3). Following an action potential, there is a time during which depolarization either cannot elicit an action potential or during which it takes a very strong depolarization to evoke an action potential. These are called the ABSOLUTE and RELATIVE REFRACTORY PERIODS. The absolute refractory period results from nearly complete inactivation of Na^+ channels immediately following an action potential. The relative refractory period reflects the time it takes for voltage-gated K^+ channels to close following membrane repolarization.

Action Potential Conduction in Axons

Action potentials are typically generated in the initial segment of the axon (also called the AXON HILLOCK) within about 50 μm of the neuronal cell body. Because action potentials are initiated at a distance from nerve terminals where neurotransmitters are released, an important question concerns how action potentials propagate to synaptic terminals. In a strictly passive nerve fiber, leakage of current across the membrane results in DECREMENTAL CONDUCTION, with the signal (voltage change) fading over a distance that is determined by the longitudinal (axial) resistance of the fiber, the membrane capacitance, and the transmembrane resistance. This passive decremental conduction is more typical of how voltage changes spread along dendrites toward the neuronal cell body, although dendrites also express voltage-gated ion channels that support back-propagating action potentials and that play important roles in modifying synaptically generated voltage changes in dendrites.

Many, but not all, axons are encased in myelin sheaths that allow them to transmit action potentials more efficiently than bare (unmyelinated) axons. As a result of myelination, axons are electrically insulated except at NODES OF RANVIER, where there are dense collections of the voltage-gated Na^+ channels involved in action potential generation (Fig. 1.5–4). Myelin increases transmembrane resistance and decreases membrane capacitance. This diminishes leakage of current from the axon, making it easier for current to flow down the length of the axon and reach the nerve (axon) terminal, where neurotransmitter is stored. Action potentials propagate rapidly, and the wave of depolarization jumps from node to node in a form that transmits the signal with high fidelity to nerve terminals. This process of action potential spread through myelinated axons is called SALTATORY CONDUCTION (from the Latin word *saltare* meaning “to jump”) and is important because of the speed and fidelity with which electrical information is passed from a neuronal cell body to its terminals. Note that while the opening of voltage-gated channels is time dependent, reflecting changes in channel conformational state, the passive current flow between nodes is essentially instantaneous, despite the loss of fidelity. Thus, the passive spread of current longitudinally down the axon is important in determining the conduction speed. A typical value for conduction velocity in large myelinated axons is ∼100 m/s, while propagation in small unmyelinated fibers is only ∼0.3 m/s. In certain human illnesses, including multiple sclerosis and Guillain–Barre syndrome, demyelination of axons produces changes in axon conduction and specific neurologic defects.

ION CHANNELS

Structure and Function of Voltage-Gated Ion Channels

Voltage-gated ion channels allow the flow of ions in response to changes in transmembrane voltage and are key elements in regulating neuronal excitability. Although ion channels can usually pass more than a single ionic species, voltage-gated channels are named according to their predominant permeant ion, Na^+, K^+, Ca^{2+}, or Cl^-. Certain channels, including those opened by neurotransmitters such as glutamate and acetylcholine, are selective for positively charged ions (Na^+, K^+, and Ca^{2+}) but exclude Cl^-, and thus are NONSELECTIVE CATION CHANNELS. The family of voltage-gated ion channels has about 150 members encoded by genes in the human genome. Only the GPCRs and protein kinases have more members. The International Union of Basic and Clinical Pharmacology (IUPHAR) maintains a website with updated nomenclature on ion channels and receptors.

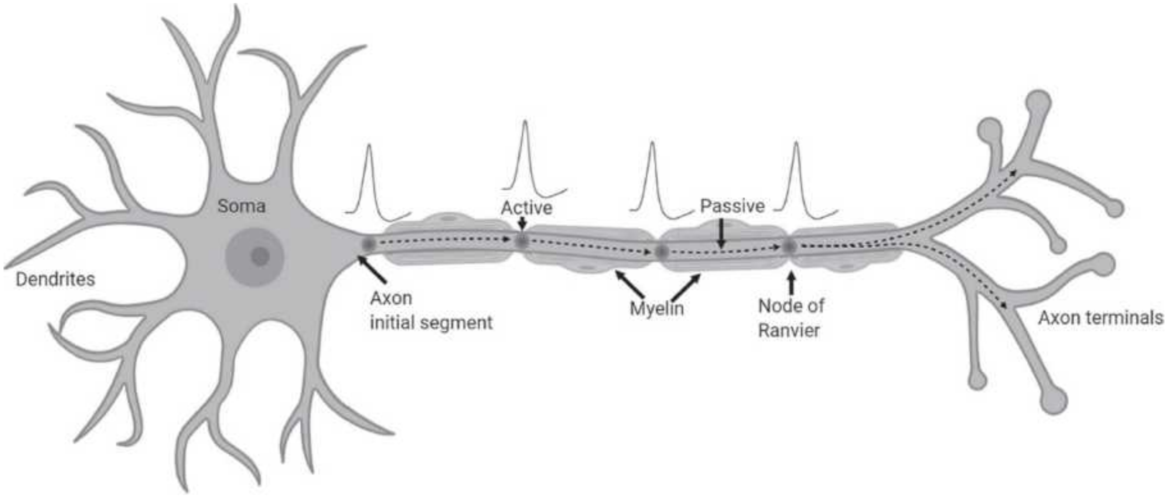

FIGURE 1.5–4. The diagram depicts saltatory conduction of an action potential in a neuron with a myelinated axon. The action potential is generated in the axon initial segment. As the waveform moves along the axon, current tends to leak from the cell diminishing the amplitude. However, myelin insulates the axon and markedly diminishes current leakage, thus enhancing axial current flow (*dashed arrows*) to the first node of Ranvier. At the nodes of Ranvier, Na^+ channels open in response to the wave of depolarization and reproduce (regenerate) the all-or-none action potential. The sequence is repeated at subsequent nodes of Ranvier until the action potential reaches the axon terminals to initiate transmitter release. (Created with BioRender.com.)

Sodium (Na^+) Channels

The opening of Na^+ channels in response to membrane depolarization underlies the fast upstroke of neuronal action potentials and thus are critical determinants of neuronal information processing. In some neurons, Na^+ channels also contribute to smaller depolarizations and pacemaker firing. Pacemaker activity refers to the ability of some neurons to depolarize spontaneously and to coordinate activity in a network of connected cells. As noted earlier, Na^+ channels ACTIVATE (open) rapidly in response to depolarization and most also INACTIVATE rapidly and nearly completely in response to prolonged depolarization.

Na^+ channels expressed in the brain typically have three protein subunits—a main (α) subunit with a molecular weight of 240 to 280 kilodaltons (kDa) and two minor (β) subunits with molecular weights of 30 to 40 kDa expressed in a 1:1:1 ratio. The α-subunit is a glycoprotein with four structurally similar domains that have six membrane-spanning (transmembrane) regions, S1 to S6 (Figs. 1.5–5 and 1.5–6). The α-subunit, when expressed alone, can form a functional channel. Unlike voltage-gated K^+ channels, which, as discussed below, are tetramers of distinct subunits with each subunit containing six transmembrane domains, functional sodium channels arise from a single α-subunit gene product. There are at least 10 mammalian genes that encode sodium channel α-subunits, Na_v 1.1 to 1.9 and Na_x. Na_v 1.1 to 1.3 and Na_v 1.6 to 1.9 are neuronal channels; Na_v 1.4 is expressed in muscle, and Na_v 1.5 is expressed in the heart. Some neuronal channels are expressed primarily in the central nervous system (CNS), and others in the peripheral nervous system (PNS). The properties of voltage dependence, ion permeation, pharmacologic sensitivity, activation, and inactivation are conferred by specific regions of these proteins.

Relationships between primary protein structure and ion channel function in Na^+ channels have been determined by site-directed mutagenesis of specific amino acid residues. Both the amino- and carboxy-terminals of α-subunits are located intracellularly. The fourth membrane-spanning region (S4) contains several electrically charged amino acid residues that sense transmembrane voltage changes and promote channel opening. Between the S5 and S6 membrane-spanning regions, a segment of hydrophobic amino acids does not completely traverse the membrane bilayer. This group of amino acids (called a "P-loop") lines the ion channel PORE (see Figs. 1.5–5 and 1.5–6). The P-loop is a feature shared by other voltage-gated ion channels and some nonselective cation channels (Figs. 1.5–7 and 1.5–8). A short intracellular loop between the third and fourth homologous domains plays a role in channel inactivation and physically blocks the ion pore during longer periods of depolarization. A high-resolution crystal structure of a bacterial

Na+ channel (NavAb) provides insights into structure–function relationships. This channel has a large external vestibule, a narrow ion selectivity filter, and a large, water-filled central cavity lined by S6 segments. The ion selectivity filter is formed by negatively charged side chains of glutamate residues that remove most, but not all, waters of Na+ hydration. An intracellular activation gate involves crossing parts of the S6 segments.

Sodium channel β−subunits (termed β1 to β4) are glycoproteins with a large extracellular amino terminus, a single transmembrane domain, and a short intracellular carboxy-terminus. Two β-subunits associate with a single α-subunit and regulate functional expression, channel kinetics, and gating. The large extracellular N-terminus is involved in cell adhesion via an immunoglobulin-like fold. A mutation in a cysteine residue in this extracellular fold is linked to a form of familial epilepsy. Studies using cryo-electron microscopy (Cryo-EM) have elucidated the structure of mammalian sodium channels complexed with a β-subunit.

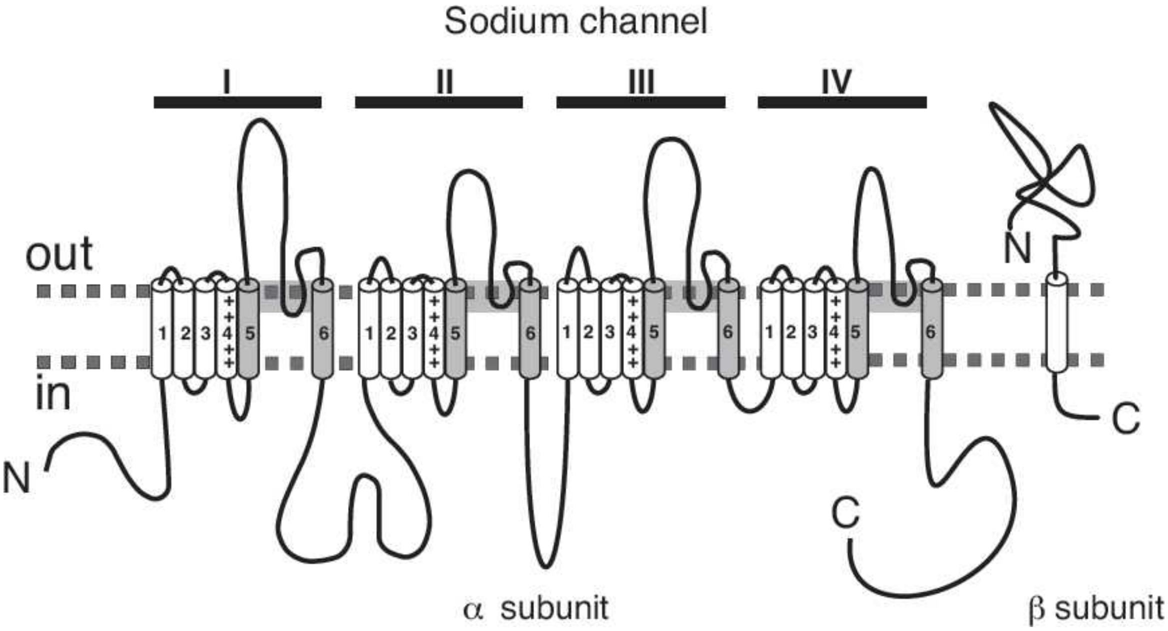

FIGURE 1.5–5. The diagram shows the schematized secondary structure of voltage-gated Na+ channels. The pore-forming α-subunit (**left**) consists of four homologous domains (I, II, III, and IV), each of which has six membrane-spanning regions (numbered 1 to 6). Both the amino (N) and carboxy (C) terminals are located intracellularly. A stretch of amino acids between S5 and S6, called the P-loop, forms two antiparallel β-sheets that line the ion channel pore. Positive charges in the fourth membrane-spanning (S4) region comprise the voltage sensor. An accessory (β) subunit is shown to the right.

Na^+ channels contain at least seven sites at which neurotoxins and clinical drugs act. Most, but not all, Na^+ channels contain an extracellular site at which tetrodotoxin (TTX) and saxitoxin (STX) block ion flow. TTX is a neurotoxin isolated from puffer fish that is commonly used in physiology experiments to block Na^+ channel function. At a site on Na^+ channels distinct from the TTX site, α-scorpion and sea anemone toxins act to modulate gating. α-scorpion toxins slow channel inactivation while α-scorpion toxins, acting at a distinct site, shift the voltage of activation to allow channels to open at voltages closer to the resting potential. The net effect of the scorpion toxins is to enhance excitation, contributing to the increased firing in pain fibers and paralysis associated with a scorpion sting. Mutations in α-subunits of skeletal muscle Na^+ channels cause the human disorder hyperkalemic periodic paralysis. Like anemone and α-scorpion toxins, these mutations slow channel inactivation.

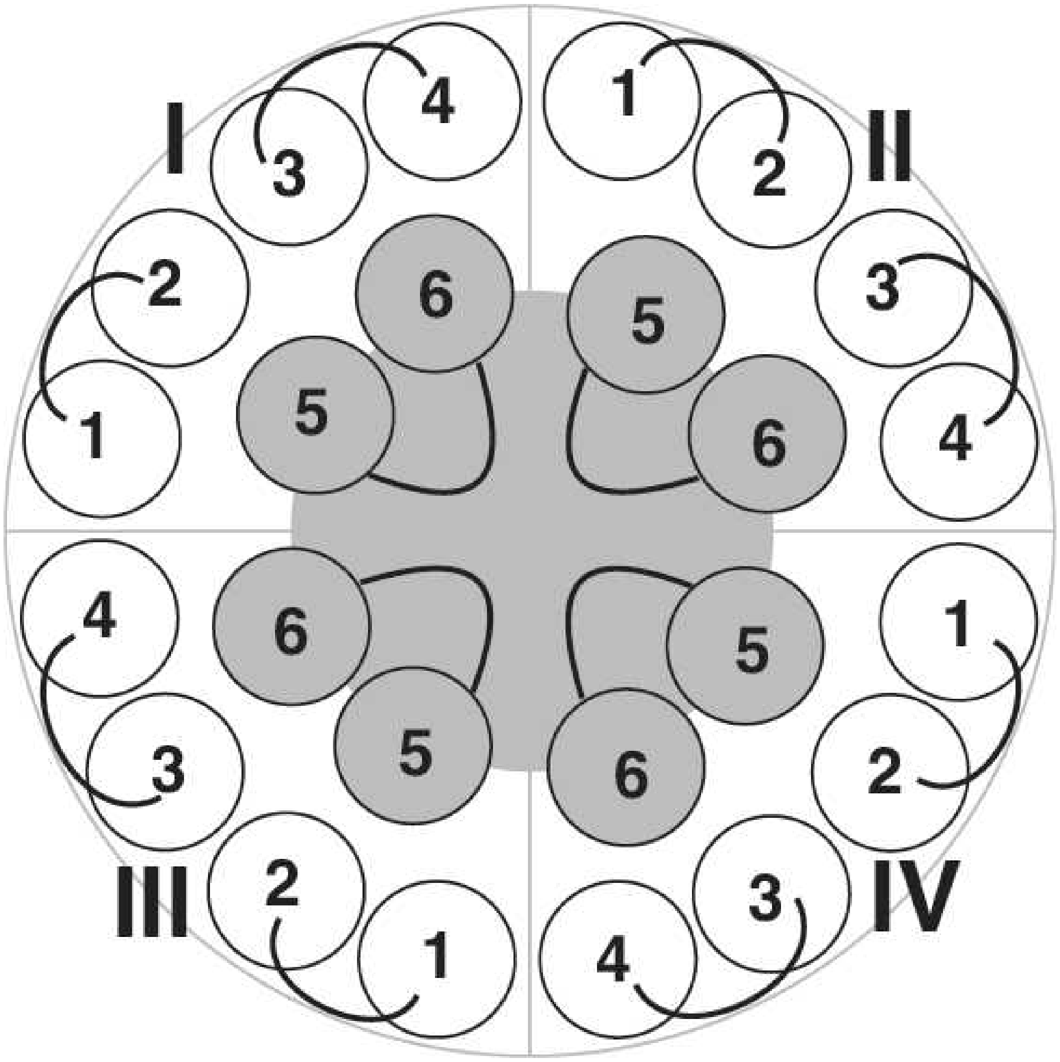

FIGURE 1.5–6. The diagram is a simplified depiction of the four homologous repeats (I to IV) in the cell membrane. The view is looking at the channel en face from the outside of the cell. Transmembrane helices are labeled 1 to 6. Extracellular linker regions are depicted by connecting lines. The P-loop is depicted as the linker between S5 and S6. Note the P-domains from each of the four homologous repeats contribute to lining the ion channel located at the center of the diagram. A more realistic depiction of the skeletal muscle form of the channel complexed with a β-subunit is available from cryo-electron microscopy work.

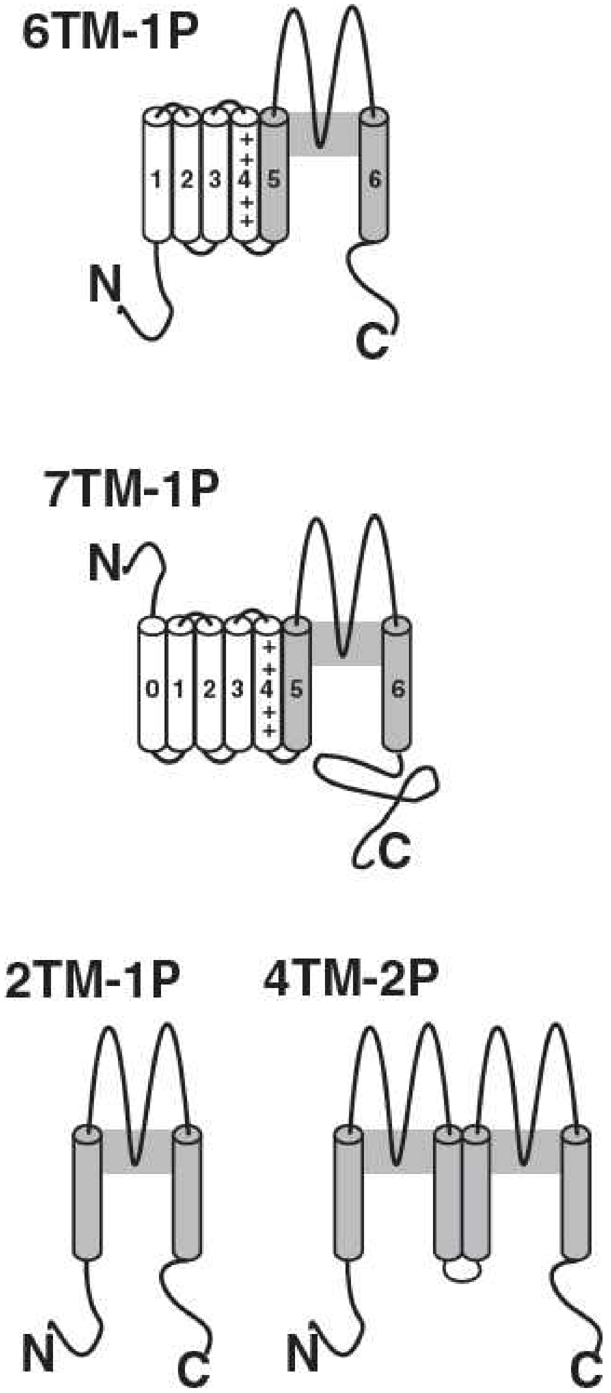

FIGURE 1.5–7. Potassium channels have diverse molecular structures. Four families of K$^+$ channels defined by the number of transmembrane (TM) domains, and the number of P-domains are shown. Of the channels described in the text, classical depolarization-gated K$^+$ channels, and KCNQ subunits are members of the six TM domain, one P-domain family. Some Ca^{2+}-dependent K$^+$ channels are also members of this family, but the BK Ca^{2+} and voltage-dependent channels are members of a separate family because of an extra TM domain and an extracellular N-terminus. Inward rectifiers, including

K_{ATP} and astrocyte leak channels, are members of the two TM domain, one P-domain family. The tandem pore (or TWIK) family of K^+ channels has TM regions and two P-domains.

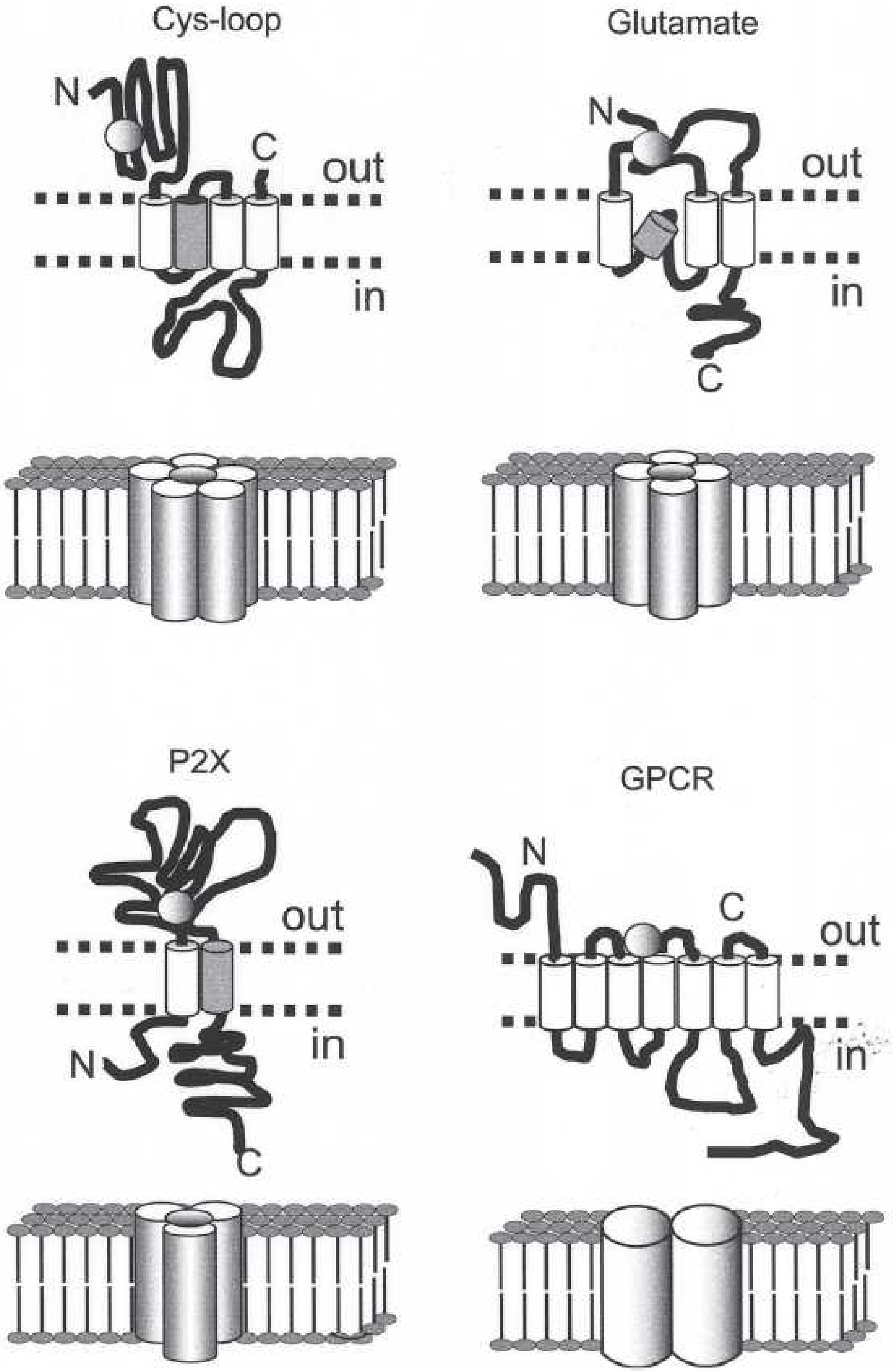
Cys-loop
N
C
out
in
Glutamate
N
out
in
C
P2X
out
N
in
C
GPCR
N
C
out
in

FIGURE 1.5–8. The diagrams show the schematized structure of receptor subunits for several neurotransmitters, including a $GABA_A$ receptor (member of the cys-loop family; heteropentameric functional receptor structure), an ion-channel–linked glutamate receptor (tetrameric structure), a channel gated by extracellular ATP (P2X; trimer) and a G-protein–coupled receptor (GPCR; often dimers). The ligand-binding domains of the subunits are depicted by the circles in the extracellular regions.

Other toxins from the buttercup family (aconitine), the lily family (veratridine), and frogs used for arrow poisons in South America (batrachotoxin) promote the direct opening of Na^+ channels and prolong the duration that channels stay open. The net effect is similar to the scorpion toxins. Certain local anesthetic drugs, including lidocaine and procaine, block Na^+ channels by binding to sites within the hydrophobic regions of the ion channel. The blockade of Na^+ channels contributes to local anesthetic effects and to antiarrhythmic effects in the heart. Certain clinically important anticonvulsants (carbamazepine, lamotrigine, phenytoin, riluzole) bind a site similar to that bound by procaine and have become important psychotherapeutic agents. All blockers described above, with the exceptions of TTX and STX, block Na^+ channels in a USE-DEPENDENT manner. That is, the drugs become more effective as cells become more depolarized. Mechanistically, this may be achieved by drugs stabilizing an inactivated state (conformation) of the channel, resulting in a clinically beneficial situation in which normal activity is relatively spared but abnormal excitation is blocked. Further clinical benefit from these drugs may result from their ability to dampen excitatory synaptic transmission selectively while sparing inhibitory transmission by mechanisms not fully understood. It is important to emphasize that not all Na^+ channels in neurons are sensitive to all of the above agents. For example, TTX-insensitive Na^+ channels exist in some excitable cells and play a prominent role in certain nociceptive neurons.

Potassium (K^+) Channels

K^+ channels are the most diverse family of ion channels and determine the resting and firing properties of neurons. At least 75 different K^+ channels are expressed in various cells. Among these are 40 mammalian genes for voltage-activated K^+ channels grouped in 12 families according to their major (α) subunit (termed K_v1 to K_v12). The various families have distinct properties and structural motifs (see Fig. 1.5–7). These include six transmembrane domain/1 P-domain channels (classical depolarization-gated channels), inwardly rectifying channels with two transmembrane domains, two-pore channels with four transmembrane domains, and two-pore channels with eight transmembrane domains (found in invertebrates). A fifth class of K^+ channels is regulated by calcium and depolarization and are known as BK (for their big single-channel conductance); these channels are similar to six transmembrane domain channels but also have a seventh membrane-spanning domain (S0 region) and an extracellular amino terminus. Adding to the diversity, there is evidence for both homo- and heteromeric K^+ channels. There are also several auxiliary subunits (β_1 to β_3) that associate with the α-subunits in an α4β4 stoichiometry. The diversity of potassium channels is daunting, and molecular and structural differences among these channels have broad functional implications, including the pattern of membrane potential changes in response to input signals.

K^+ channels with six transmembrane domains are particularly important in neurons, with depolarization-gated channels representing a major subgroup. Voltage-gated K^+ channels have several subfamilies, K_V1 (sometimes called *Shaker* for the *Drosophila* gene), K_V2 (*Shab*) K_V3 (*Shal*), and K_V4 (*Shaw*), each with constituent subunits (e.g., $K_V1.1$ to 1.4). A functional channel is a tetramer of four subunits from the same subfamily. Specific domains in the amino-terminal regions are responsible for forming tetramers and regulate gating, inactivation, and interactions with accessory proteins.

A common structural motif among voltage-gated K^+ channels is the presence of six membrane-spanning domains (S1 to S6) with a re-entrant loop (P-domain) between S4 and S5. The re-entrant loops of the four subunits line the ion channel pore. Carbonyl groups in the P-domain interact with K^+ and determine ion selectivity. These interactions mimic the hydration shell for K^+, and no water molecules intervene between K^+ and the carbonyl groups, differing from ion selectivity in Na^+ channels. The specificity of the interactions within the channel pore imparts selectivity for K^+ over other ions, and high-efficacy channel permeation involves electrostatic ion–ion repulsions of adjacent K^+. A re-entrant P-loop is common to all families of K^+ channels.

Voltage-gated K^+ channels help to define the spike shape and firing pattern of subtypes of neurons. The fast repolarization mediated by certain K^+ channels allows an increased rate of action potential firing that can then be used in frequency-dependent information coding (Fig. 1.5–9). Most neurons express multiple types of K^+ channels that differ in activation and inactivation kinetics, voltage dependence, and pharmacology. Because the equilibrium potential for K^+ is about −90 mV, the opening of K^+ channels results in K^+ efflux, membrane hyperpolarization, and a decrease in excitability.

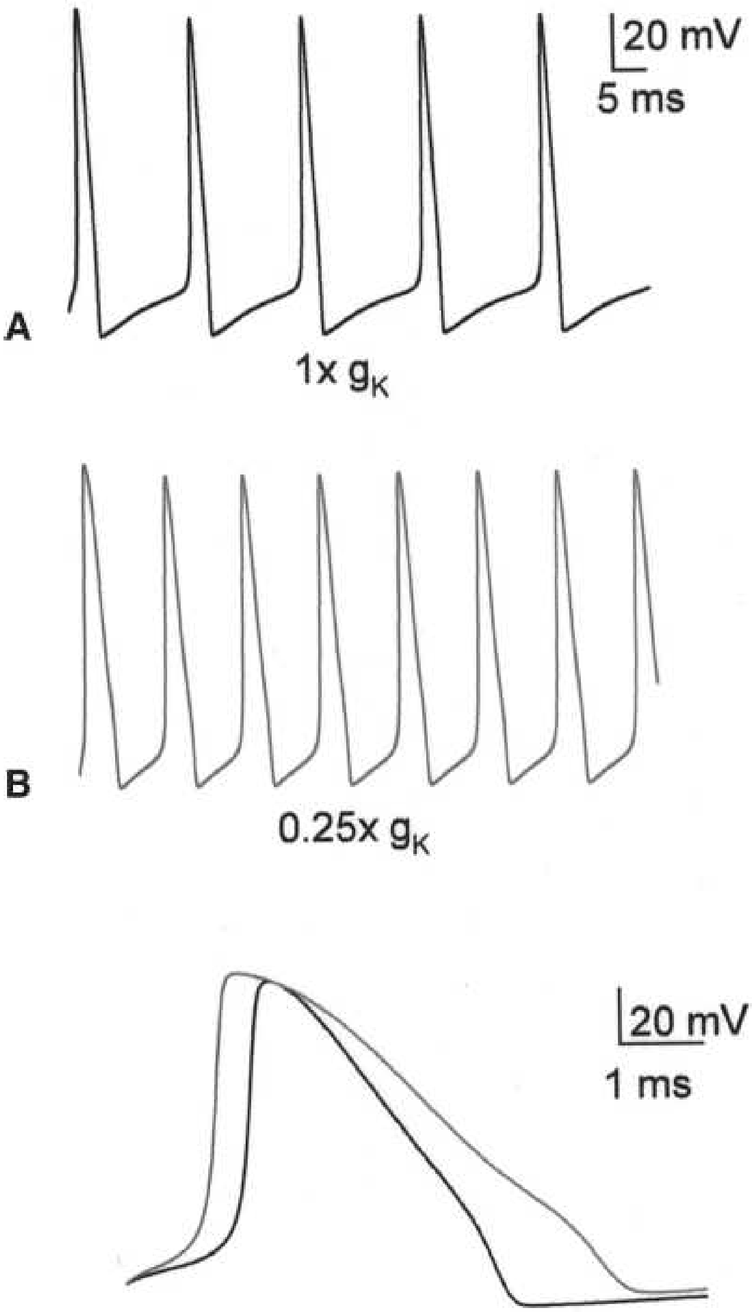

FIGURE 1.5–9. **A:** The traces show the effect of inhibiting K^+ channels involved in action potential repolarization. The traces were constructed using a simulation program that includes nongated leak channels and voltage-gated sodium and potassium channels similar to those described by Hodgkin and Huxley. The simulation includes a sustained depolarizing current injection of fixed amplitude to elicit spiking. The gray traces show the effect of reducing the delayed rectifier potassium conductance (g_K) to 25% of the initial baseline level (*black traces*). Note that with fewer voltage-gated potassium channels, the cell is hyperexcitable, exhibiting more action potentials to the same depolarizing input. **B:** The bottom panel shows the first 7 ms of the

simulations superimposed. Following K^+ channel block, individual action potentials are broadened and show a diminished undershoot.

The first K^+ channel identified physiologically was called a DELAYED RECTIFIER. These channels were predicted in the experiments of Hodgkin and Huxley on squid giant axons and were so named because their currents activate more slowly than the Na^+ channels that produce the upstroke of the action potential; channel opening is “delayed” relative to Na^+ channel opening. A “rectifier” (diode) is an electrical device that passes current better at certain voltages than at others; in other words, conductance varies with voltage. The K^+ current is a rectifier because the channel is more effective in allowing K^+ ions to exit than to enter the cell by virtue of depolarization opening the channels. Delayed rectifier channels show little inactivation during prolonged depolarization. These channels help to determine the frequency with which neurons fire action potentials in response to sustained depolarization. For instance, “fast-spiking” interneurons of the hippocampus and cortex possess a rapidly activating and deactivating delayed rectifier channel encoded by Kv3 family members. These channels are largely responsible for the narrow spike and brief refractory period of this interneuron class. Structurally, delayed rectifiers are members of the six transmembrane K^+ channel subfamily.

Experiments in squid axons indicated that delayed rectifiers were the primary K^+ channels involved in action potential repolarization. In mammalian neurons, the situation is more complex, with additional K^+ channels contributing significantly. These include two classes of CALCIUM-ACTIVATED K^+ CHANNELS that are opened by increases in intracellular Ca^{2+} (some are also opened by depolarization). These channels are important in action potential repolarization and in generating an AHP characteristic of some neurons. In cells that possess it, the AHP is responsible for the process of ACCOMMODATION (adaptation) that diminishes repetitive action potential firing during prolonged depolarization (see Fig. 1.5–9). The AHP has several temporal components that are mediated by big conductance (BK) and small conductance (SK) calcium-activated K^+ channels. BK channels mediate the fast component of the AHP and have a large single-channel conductance. BK channels belong to the SLO family of channels named for the *slowpoke* gene in *Drosophila* and include Ca^{2+}, Na^+, and H^+ sensitive K^+ channels. The large conductance of these channels results from a ring of negative charges in the inner and outer pore regions and a large inner pore. Some evidence suggests an association between a defect in the α-subunit of BK channels and autism with intellectual dysfunction. SK channels (SK1 to 3 in the CNS) are gated by intracellular Ca^{2+} but are voltage insensitive and contribute to the medium and slow components of the AHP. Some evidence has associated SK channel variants with schizophrenia and other psychotic disorders.

A-TYPE K^+ CHANNELS rapidly activate with depolarization to potentials positive to −60 mV and rapidly inactivate at depolarized potentials. A-channels help to regulate the interspike frequency with which neurons fire and contribute to action potential repolarization. They are important for frequency coding, the encoding of input strength in the firing frequency of a neuron. The *Shaker* A-channel from *Drosophila* was the first K^+ channel cloned.

M-CHANNELS are activated in a time- and voltage-dependent fashion and are inhibited by the neurotransmitter acetylcholine, acting at muscarinic receptors. These channels activate slowly and contribute little to action potential repolarization, but their activation at negative membrane potentials slows repetitive firing. In pyramidal neurons, M-channels are expressed in perisomatic regions where they regulate synaptic integration and adaptation to repetitive firing. M-currents arise from the heteromeric association of two members of the six transmembrane domain groups, *KCNQ2* and *KCNQ3* (members of the K_v7 subfamily), and possibly other KCNQ family members. Mutations in KCNQ subunits underlie benign familial neonatal convulsions (BFNC), deafness, and a form of cardiac long QT syndrome. The name of the subunits in this class of channels (KCNQ) derives from their role in long QT syndrome. Retigabine, an anticonvulsant, potently opens M-channels and slows their closing; studies in rodents suggest that agents of this type may have antidepressant activity.

In the sea snail *Aplysia californica*, K^+ channels that contribute to action potential repolarization are inhibited by the neurotransmitter serotonin and are called S-CHANNELS. S-channel activity is diminished during acute behavioral sensitization of the gill-withdrawal reflex in *Aplysia*, and studies of the role of these channels in synaptic function provided important insights into the cellular basis of certain forms of learning and memory.

Some K^+ channels (K_{ir} family) are opened by hyperpolarization instead of depolarization and are called INWARD RECTIFIERS (anomalous rectifiers) because they allow K^+ to more easily enter than exit the cell and strongly rectify near the equilibrium potential for K^+. The primary mechanism underlying this rectification is channel block by positively charged intracellular magnesium or polyamines. This channel block is relieved by K^+ influx through the channel, accounting for rectification that follows the K^+ equilibrium potential. Despite their inward rectification, the physiologic importance of these channels for neurons likely lies in passing small outward (hyperpolarizing) currents because neurons are rarely hyperpolarized beyond the K^+ equilibrium potential. Inward rectifiers are members of the two transmembrane domain groups of K^+ channels; this class has a single pore region and lacks a voltage sensor (see Fig. 1.5–7).

Brain astrocytes express inwardly rectifying K^+ channels of the $K_{ir}4$ and possibly $K_{ir}2$ subfamily that help to buffer extracellular increases in K^+ during neuronal activity. The expression of inward rectifiers in these cells helps to siphon K^+ away from axons and axon terminals where extracellular accumulation may occur. The unique properties of these channels favor the influx of K^+. Note that the influx of K^+ ions through these channels is caused by the change in equilibrium potential for K^+ rather than by channel opening. Transient conditions of elevated extracellular K^+ represent a rare situation in which the resting potential of surrounding astrocytes is more negative than the equilibrium potential for K^+.

Adding to the complexity of K^+ channels is the observation that certain neurotransmitters alter their function via GPCRs. For example, acetylcholine, acting via muscarinic receptors, inhibits several K^+ currents resulting in enhanced neuronal excitability. At different muscarinic receptors, acetylcholine activates inwardly, rectifying K^+ channels in multiple tissues, including the heart and brain. In the hippocampus and other brain regions, the neurotransmitters γ-aminobutyric acid (GABA), serotonin, and adenosine indirectly open inwardly, rectifying K^+ channels. These G-protein-regulated, inwardly rectifying channels (GIRKs) allow divergent synaptic inputs to exert regulatory influences over neuronal firing through a single class of ion channels.

In peripheral tissues and in some neurons, the $K_{ir}6$ subfamily of two transmembrane domain inward rectifiers is regulated by intracellular ATP (see Fig. 1.5–7). In the pancreas, K_{ATP} channels are a site of action of the hypoglycemic sulfonylurea drugs (tolbutamide and glibenclamide) that inhibit these channels and promote insulin release via membrane depolarization and calcium influx. Diazoxide, an antihypertensive drug that increases blood glucose levels, has opposite effects on pancreatic ATP-sensitive K^+ channels, opening the channels and diminishing insulin release. Sulfonylurea drugs do not interact directly with K_{ir} subunits of K_{ATP} channels but rather bind high-affinity sulfonylurea receptors (SURs) that combine with $K_{ir}6$ subunits in heteromeric complexes of four K_{ir} subunits and four SURs. K_{ATP} channels are also expressed in the CNS and regulate the release of certain neurotransmitters and perhaps determine neuronal responses to changes in intracellular energy levels. These channels are also expressed in mitochondrial membranes and may play a role in regulating apoptotic cell death.

HCN (hyperpolarization and cyclic nucleotide-gated) or H-CHANNELS are nonselective cationic channels structurally related to K^+ channels. These channels show high expression in certain neurons, including dopaminergic neurons of the substantia nigra and ventral tegmental area, cells implicated in motor behavior, arousal, attention, reward, motivation, and addiction. H-currents contribute to resting membrane potential and pacemaker firing and are expressed in dendrites of some pyramidal neurons, where they help to modulate synaptic inputs. HCN channels are similar to K_{ir} channels in that they activate at hyperpolarized voltages and close with depolarization. HCN channels differ in being permeable to both K^+ and Na^+, providing a depolarizing current at membrane potentials near rest. H-channels help stabilize neuronal membrane potentials, resisting changes that depolarize or hyperpolarize the cell. The anticonvulsant and mood stabilizer, lamotrigine, activates dendritic H-channels in pyramidal neurons as one of its potential mechanisms of action, and the dissociative anesthetic and antidepressant, ketamine, blocks these channels at concentrations relevant to some of its clinical effects. HCN channels differ structurally from K_{ir} channels but are similar to other voltage-activated channels with six membrane-spanning regions, an S4 voltage sensor, and a re-entrant P-loop. H-channels have an intracellular cyclic nucleotide-binding domain near their carboxy-terminus, and the binding of cyclic AMP shifts the voltage range for channel gating.

The K^+ channels described above have either two or six membrane-spanning regions and a single P-loop. Another class has four membrane-spanning regions and two P-domains (see Fig. 1.5–7). These TANDEM

PORE (K_{2P}) channels are widely expressed in the brain and periphery, and serve, at least in part, as leak conductances that help to establish the resting membrane potential. These subunits form functional dimers, and 18 members of this family have been identified. K_{2P} channels vary in electrical properties and sensitivity to activating stimuli (e.g., neurotransmitters, second messengers, arachidonic acid, acid, heat, stretch). The functional channels have a variety of names based on TWIK (tandem pore weak inward rectifying K^+) channels that were the first cloned. For example, TWIK-related arachidonic acid-sensitive K^+ channels are called TRAAKs, while acid-sensitive channels are called TASKs. Channels of this class are activated by certain volatile anesthetics and anticonvulsants (riluzole).

As noted, members of the TWIK family are activated by a host of novel stimuli, including acidic pH, heat, and mechanical activity. Another class of channels, called ASICs (acid-sensing ion channels), is also responsive to these stimuli. ASICs have two membrane-spanning domains with an apparent re-entrant loop between domains. At least nine mammalian family members falling into five subfamilies have been identified. Unlike TWIKs, ASICs are voltage-insensitive nonselective cationic channels that are more permeable to Na^+ than to K^+ and least permeable to Ca^{2+}. ASICs appear to form functional trimers and participate in peripheral sensory processing, including touch, heat, taste, and pain, and also contribute to certain forms of long-term synaptic plasticity in the hippocampus and amygdala, where they are involved in fear-associated learning.

Calcium (Ca^{2+}) Channels

The divalent cation, Ca^{2+}, is a critical contributor to cellular function, serving as both a regulator of intracellular chemistry (enzyme activity, metabolism, transmitter release) and an electrical signal for membrane depolarization. In some neurons, Ca^{2+} influx can drive action potentials (“calcium spikes”). However, excessive and prolonged increases in intracellular Ca^{2+} contribute to neurodegeneration. Under some conditions, prolonged deficiency of Ca^{2+} influx may also result in neuronal death, particularly in developing neurons. Thus, regulation of intracellular Ca^{2+} levels is vital to cellular function and survival.

Voltage-activated Ca^{2+} channels provide a major source of the Ca^{2+} that regulates cellular function and, in conjunction with Ca^{2+}-permeable ligand-gated ion channels (e.g., N-methyl-D-aspartate [NMDA]-type glutamate receptors and neuronal nicotinic acetylcholine receptors), represent major conduits for Ca^{2+} entry from the extracellular environment. Neurons express multiple classes of voltage-gated Ca^{2+} channels that are classified according to biophysical and pharmacologic properties. LOW-VOLTAGE–ACTIVATED (LVA, Ca_v3) Ca^{2+} CHANNELS are activated by relatively small depolarizations over the range of −80 to −50 mV. LVA channels inactivate rapidly and are largely insensitive to dihydropyridine Ca^{2+} channel blockers, such as nifedipine and nimodipine. LVA channels are also called T-TYPE Ca^{2+} CHANNELS because of their "transient" (inactivating) currents. Because LVA Ca^{2+} channels are activated at membrane potentials near rest, these channels contribute to burst firing and oscillatory neuronal activity. Oscillatory firing helps to drive coordinated movements and maintain complex behavioral states such as wakefulness. LVA channels are also expressed in neuronal dendrites and contribute to synaptic integration and spike-timing–dependent forms of synaptic plasticity. With a few exceptions, LVA channels do not participate in neurotransmitter release at axon terminals. LVA channels may be a target for the actions of some antipsychotic drugs, and the diphenylbutylpiperidines, pimozide, and penfluridol inhibit LVA channels at concentrations similar to those affecting D2 dopamine receptors. Other antipsychotics also inhibit LVA channels but do so at concentrations above those required at dopamine receptors.

HIGH-VOLTAGE–ACTIVATED Ca^{2+} CHANNELS (HVA, Ca_v1, and Ca_v2 family members) are activated by membrane depolarizations positive to −50 mV. In many neurons, even when voltage-gated Na^+ channels are blocked, HVA channels can produce regenerative spikes. These calcium spikes are slower in onset and longer in duration than Na^+ spikes, reflecting the slower kinetics of HVA channels.

HVA channels are heterogeneous with several major subtypes. L-TYPE Ca^{2+} CHANNELS (Ca_v1 family; named for their "long-lasting" responses) show slow inactivation during sustained depolarizations and are sensitive to block by dihydropyridines. L-type channels provide sufficient Ca^{2+} influx during action potentials to activate Ca^{2+}-dependent second messengers and gene expression. N-TYPE Ca^{2+} CHANNELS ($Ca_v2.2$ named because they were "neither" L- nor T-type) are HVA channels that regulate neurotransmitter release from some presynaptic terminals, particularly those of sensory neurons in the retina and inner ear. N-type channels are blocked irreversibly by ω-conotoxin GVIA, poison from the snail, *Conus geographicus*. Another conotoxin, ω-conotoxin MVIIA (Prialt), is a reversible N-channel blocker that is used to treat chronic pain. P/Q-TYPE CHANNELS ($Ca_v2.1$) are the third class of HVA channels and are named because of their presence in Purkinje cells of the cerebellum and pyramidal neurons of the hippocampus and cortex. P/Q-channels are insensitive to dihydropyridines and ω-conotoxin GVIA but are blocked by ω-Aga-IVA, a toxin from the funnel-web spider, *Agelenopsis aperta.* P/Q-channels, like N-channels, regulate the release of neurotransmitters in the CNS. N- and P/Q-type channels have an intracellular loop between channel domains II and III that binds certain presynaptic proteins, including syntaxin 1, Rim, and synaptotagmin 1. This site is referred to as a "synprint" region and plays a role in allowing synaptic proteins to modulate channel function.

R-TYPE CHANNELS ($Ca_v2.3$) also contribute to CNS function, but their actions are less well understood. R-type channels are resistant to the Ca^{2+} channel antagonists described above but are inhibited by SNX-482, a toxin derived from the African tarantula, *Hysterocrates gigas*. Although their presynaptic role is not as prominent as N- and P/Q-channels, R-channels participate in transmitter release at fast excitatory synapses in the CNS.

Skeletal muscle HVA Ca^{2+} channels were the first of these channels cloned and serve as a model for understanding the structure of other voltage-gated Ca^{2+} channels. These channels are involved in excitation–contraction coupling at neuromuscular junctions and consist of five distinct subunits, α1 (165 to 195 kDa), α2 (~150 kDa), δ (17 to 25 kDa), β (50 to 60 kDa), and γ (25 to 35 kDa) arranged in a 1:1:1:1:1 stoichiometry. The δ-subunit is produced by cleavage of α2/δ peptides while the other subunits are encoded by separate genes. α1-Subunits have about 30% sequence homology to voltage-gated Na^+ channels and form the ion channel pore. A recurring theme in α1-subunits is the existence of four homologous internal repeats that each contain six membrane-spanning regions and a pore-forming P-loop (see Figs. 1.5–5 and 1.5–6). The skeletal muscle HVA α1-subunit contains the dihydropyridine binding site. Point mutations in the α1-subunit of skeletal muscle T-tubule Ca^{2+} channels ($Ca_v1.1$) and the skeletal muscle Na^+ channel ($Na_v1.4$) cause the human disorder, hypokalemic periodic paralysis. The critical mutations occur in the S4 region of the channel involved in voltage sensing and result in a gain of function current that is open at rest and results in membrane depolarization and potential action failure; similar pore mutations may occur in other channelopathies. The functions of the β (β_{1-4}) and γ (γ_{1-8}) subunits are less certain but appear to regulate membrane expression and trafficking of α1-subunits. Loss of the γ2-subunit (*stargazin*) in the stargazer mouse markedly diminishes cell surface expression of certain glutamate receptors. The importance of auxiliary subunits in the function of Ca^{2+} channels is highlighted by evidence that analgesic effects of the anticonvulsants gabapentin and pregabalin involve binding to specific residues in α_2–δ_1-subunits.

Ten different α1 genes have been identified. These include four different α1 subtypes contributing to L-type channels (Ca_v1.1, 1.2, 1.3, 1.4), three α1 genes contributing to P/Q-, N-, and R-type channels (Ca_v2.1, 2.2, 2.3), and three variants of T-type Ca^{2+} channels (Ca_v3.1, 3.2, 3.3). Adding to this complexity, human genes typically go by other names. For example, the skeletal muscle L-type channel Ca_v 1.1 is referred to as α1S and CACNA1S; other human calcium channel genes are named accordingly. The importance of calcium channels to neuropsychiatric disorders is highlighted by the finding that mutations in the human CACNA1A gene encoding the α1-subunit of P/Q calcium channels are associated with familial hemiplegic migraine.

An important question concerns how channels establish selectivity for one ion over another. At one level, ionic charge and charged amino acid residues in ion channel proteins help select cations over anions. It is a more difficult problem for channels to select among different cations. In Ca^{2+} channels, this is particularly vexing because hydrated Ca^{2+} ions are significantly larger than Na^+ or K^+ ions. Thus, the size of the channel pore cannot determine selectivity for Ca^{2+}. Rather, selectivity in Ca^{2+} channels involves high-affinity binding of divalent cations within the channel. When Ca^{2+} is present, its binding within the pore excludes monovalent cations, rendering the channels selective for Ca^{2+}. As might be expected, when Ca^{2+} is not present, these channels will pass monovalent cations. This principle of ions binding to specific sites within a channel to regulate permeability and gating is a recurring theme that can sometimes be exploited for the development of drugs that modulate channel function.

In some regions of the CNS, particularly retinal photoreceptors and olfactory epithelial cells, intracellular cyclic nucleotides (e.g., cyclic AMP and cyclic GMP) gate specific classes of ion channels. These cyclic nucleotide-gated (CNG) channels have structural features similar to voltage-gated channels, including the presence of six transmembrane regions and a P-loop that lines the channel. CNG channels also have an S4-like voltage-sensing region, although these channels are not regulated by voltage. CNG channels are nonselectively permeable to cations. Like voltage-gated calcium channels, they bind divalent cations in the external pore region as a means of restricting the flow of monovalent cations, thereby selecting Ca^{2+} permeation over Na^+. CNG channels have structural similarities to HCN channels but have higher calcium permeability. Mutations in retinal photoreceptor CNG channels contribute to color blindness in humans.

The TRP SUPERFAMILY represents another class of cationic channels with six membrane-spanning regions and high calcium permeability. This family contains more than 25 members in at least seven subfamilies (TRPC, TRPV, TRPM, TRPML, TRPP, TRPA, and TRPN) that participate in multiple processes in the nervous system and in nonexcitable cells, ranging from sensory processing to vascular and cell cycle control. TRP channels are named for the first member to be identified, the *trp* (transient receptor potential) gene in *Drosophila*, and are linked to human disorders including polycystic kidney disease and mucolipidosis, a neurodegenerative illness. TRP channels are regulated by intracellular and extracellular signals, including changes in pH, temperature, capsaicin (the active ingredient in hot peppers), and anandamide (an endogenous ligand for cannabinoid receptors). Endogenous lipid mediators like the endocannabinoid anandamide are important regulators of the TRPC/VM subfamilies. TRPC1 and TRPC4 also contribute to “store-operated” channels that mediate extracellular calcium influx following calcium release from intracellular stores.

Chloride (Cl^-) Channels

In most neurons, Cl^- is present at higher concentrations outside than inside cells, and the equilibrium potential for Cl^- is near the resting membrane potential. Thus, the opening of Cl^- channels tends to maintain the membrane potential near rest and, in conjunction with K^+ channels, serves as a mechanism to dampen neuronal excitability. Some Cl^- channels also conduct bicarbonate, another physiologic anion. Cl^- channels contribute to the resting membrane potential in certain neurons and muscle cells. These channels spontaneously open at resting membrane potentials and exhibit weak voltage dependence. In muscle fibers, the background Cl^- conductance is the largest resting conductance and the transmembrane distribution of Cl^- is near equilibrium. There are three main families of Cl^- channels, voltage-gated Cl^- channels (ClCs), the cystic fibrosis (CF) transmembrane conductance regulator (CFTR), and ligand-gated Cl^- channels. The ClC family includes nine different channels that have 10 or 12 membrane-spanning regions and no defined S4 voltage-sensing region as in voltage-gated cation channels. Crystallographic studies indicate that ClC channels are double-barreled homodimers, with each subunit containing its own pore. In the human illness, myotonia congenita, an abnormality of a muscle Cl^- channel (ClC-1) results in abnormally low Cl^- conductance leading to increased muscular excitability and fatigue with exercise.

There are at least five activating stimuli for Cl^- channels. These include changes in membrane voltage (depolarization or hyperpolarization), increases in intracellular Ca^{2+}, ligand binding (usually GABA or glycine), cellular swelling, and phosphorylation by cyclic-AMP–dependent protein kinase (PKA). Ca^{2+}-activated Cl^- channels may help to determine the interspike firing frequency of neurons, while swelling-activated channels help to protect cells from damage during osmotic stress. These channels are pentamers with a single ion pore and a hydrophobic neck that forms the channel gate controlling ion flow. Ca^{2+} binds to a large cytosolic region of the protein to regulate channel opening. Cl^- channels also serve important functions in secretory cells, providing a major source of Cl^- in tears, sweat, and digestive juices. A defect in CFTR secretory Cl^- channels that renders the channels insensitive to activating stimuli is involved in CF. CFTR has a structure that differs from the ClC family and belongs to a larger family of ATP-binding cassette (ABC) proteins that require phosphorylation by protein kinase A and hydrolysis of ATP for activation. CFTR is the only member of the ABC family that also serves as a Cl^- channel. CFTR has two repeats with six membrane-spanning regions (12 transmembrane regions in total), a nucleotide-binding domain, and a regulatory domain. Drugs that combine agents that promote trafficking of CFTR to the cell surface with agents that enhance CFTR function have been approved for the treatment of patients with an F508 (phenylalanine) deletion that is the most common cause of CF. These agents are a major step forward for CF patients and a proof of concept for the ability to modulate defective ion channels for therapeutic purposes.

Mitochondrial membranes express unique voltage-dependent anion channels (VDACs) that pass negatively charged ions and have unusual gating properties, being open at potentials near 0 mV and closing with voltage changes in either direction. VDACs (or PORINS) participate in releasing metabolites from mitochondria and are components of the mitochondrial permeability transition pore (PTP) that regulates apoptotic cell death. PTP is a multiprotein complex that includes adenine nucleotide translocase, hexokinase, cyclophilin D, and a VDAC. PTP activity is regulated by mitochondrial benzodiazepine receptors (now called translocator protein 18 kDa, TSPO). VDACs may also be expressed in plasma membranes and may be complex with some neurotransmitter receptors, although the function of plasma membrane VDACs remains uncertain. Three VDACs (VDAC 1 to 3) have been identified and are β-sheet proteins that differ from the α-helical configuration of most ion channels.

NEUROTRANSMITTERS AND ION CHANNELS

Classes of Neurotransmitters

Most information transfer between neurons occurs via chemical synapses. These connections use a wide array of chemical messengers (NEUROTRANSMITTERS) that are released in a Ca^{2+}-dependent fashion from presynaptic terminals and act on specific membrane proteins (RECEPTORS) to produce biochemical and excitability changes. There are two primary groups of neurotransmitters—low-molecular-weight amines and neuroactive peptides. These agents act on two classes of receptors, ligand-gated ion channels, at which binding of the transmitter directly opens an ion channel, and GPCRs, where an activated G-protein modulates ion channels or alters biochemical SECOND MESSENGER SYSTEMS. Synaptic transmission can also be classified according to the speed of transmission (fast or slow) and according to the excitability change (excitatory or inhibitory). FAST transmission occurs on a time scale of up to several 100 milliseconds and is mediated primarily by amine neurotransmitters acting at ligand-gated ion channels. SLOW synaptic communication occurs on a scale of seconds to minutes or longer, usually through the actions of amine or peptide transmitters acting on GPCRs.

The specific ion channels involved and the electrochemical gradients of the permeant ions determine whether effects are EXCITATORY (depolarizing) or INHIBITORY (hyperpolarizing). Moreover, an excitatory synaptic input can exert a net inhibitory influence on regional excitability and vice versa. For example, the release of an excitatory neurotransmitter onto an inhibitory neuron can stimulate the inhibitory neuron and dampen the activity of a local network. Conversely, inhibition of inhibitory neurons (disinhibition) can enhance regional excitability.

There are at least nine low-molecular-weight amines that serve as neurotransmitters. These include GLUTAMATE, the major fast excitatory transmitter in the mammalian brain, ACETYLCHOLINE, the excitatory transmitter at the vertebrate neuromuscular junction, GABA, and GLYCINE, the major fast inhibitory transmitters in the brain and spinal cord, respectively, and the biogenic amines, DOPAMINE, NOREPINEPHRINE, EPINEPHRINE, SEROTONIN, and HISTAMINE. Purines, ADENOSINE, AND ATP can also act as transmitters. A large number of NEUROACTIVE PEPTIDES modulate neuronal excitability. It is uncertain, however, whether all of these substances function as bona fide neurotransmitters. Many peptides, including vasopressin and cholecystokinin, were first identified as hormones in the vasculature and gut. ATP and certain neuroactive peptides coexist with amine neurotransmitters in some nerve terminals and can be co-released. Thus interactions between classes of neurotransmitters may determine the ultimate effects of a presynaptic neuron on its postsynaptic target.

Conductance Mechanisms Underlying Neurotransmitter Actions

Neurotransmitter actions can also be described by effects on membrane conductances. Transmitters that act at ligand-gated ion channels increase the conductance of cell membranes (open channels) to specific ions. Excitatory transmitters, such as acetylcholine and glutamate, directly activate nonselective cation channels, increasing the flow of Na^+, K^+, and, in some cases, Ca^{2+}. Because of the mixed permeability of these channels, the reversal potential of currents generated by these channels is a weighted average of the equilibrium potentials of the individual permeant ions, typically near 0 mV. Thus, at membrane potentials negative to 0 mV, channel opening depolarizes (excites) the cell by allowing the inward flow of cations. If sufficient numbers of the channels open, they will drive the membrane potential toward 0 mV, exceeding the action potential threshold.

In contrast, the fast inhibitory transmitters, GABA and glycine, open ligand-gated channels permeable to Cl^-. Because Cl^- usually has an equilibrium potential near the resting membrane potential, the opening of these channels resists depolarization toward the spike threshold, decreasing (inhibiting) the likelihood of spike initiation and shunting excitatory influences. Note that the excitatory or inhibitory nature of transmitter actions is dictated by the electrochemical gradients of the permeant ions. Early in development, many neurons possess a sufficiently high intracellular Cl^- concentration that the Cl^- equilibrium potential is positive relative to the spike threshold, resulting in excitatory effects of GABA and glycine. With maturation and changes in the expression of specific Cl^- transporters, intracellular Cl^- levels diminish, and GABA and glycine become inhibitory transmitters. There are pathologic conditions, including behavioral stress, that alter the expression of Cl^- transporters. Under these conditions, GABA (and Cl^- channel opening in general) becomes excitatory. This is thought to contribute to the epileptogenesis associated with human brain tumors.

A second group of transmitters indirectly increases membrane conductance through G-proteins. For example, GABA, serotonin, and adenosine, by binding GPCRs, promote G-protein–mediated opening of inwardly rectifying K^+ channels (GIRKs). A third set of transmitter actions involves indirect effects on voltage-gated or leakage ion channels. These transmitters typically decrease membrane conductance by activating chemical second messenger systems via GPCRs. Certain voltage-gated K^+ and Ca^{2+} channels are specific targets of this inhibition, resulting in cellular excitation or inhibition, respectively. Most transmitters that act on GPCRs exert at least some of their effects by these decreased conductance mechanisms. The electrical principles underlying synaptic excitation or inhibition are identical to those described for other ion channels and are based on the relative permeabilities of the ion channels and the equilibrium potentials of the ions involved.

Several transmitters (e.g., GABA, glutamate, acetylcholine, and serotonin) act at both ligand-gated ion channels and GPCRs. Thus, receptors for these neurotransmitters and their effects are heterogeneous, with the nature of the effect depending on the specific receptor to which the transmitter binds. Receptors for most neurotransmitters are structurally varied, with multiple receptor subtypes being the rule rather than the exception. Further variability arises from RNA splicing differences and posttranslational modifications of receptor function. Thus, at the receptor level, there is tremendous flexibility in determining the effects of a neurotransmitter on a single neuron or on a set of neurons.

Structure of Neurotransmitter Receptors

Most transmitter-gated ion channels are multimeric proteins consisting of several (usually four or five) subunits that have multiple (two to five) membrane-spanning regions (see Fig. 1.5–8). Functional receptors typically have large amino-terminal regions that extend into the aqueous extracellular environment and participate in neurotransmitter binding. Intracellular regions of the receptor often contain sites at which phosphate groups can be attached. Phosphorylation is a mechanism by which second messenger systems modulate receptor and ion channel function. Many receptors are multiprotein complexes in which the receptor subunits that comprise the ion channel pore are in physical proximity with intracellular proteins (in some cases, 100 or more intracellular proteins). The intracellular proteins regulate receptor trafficking and expression as well as ion channel function and participation in intracellular signaling processes.

The neuromuscular nicotinic acetylcholine receptor was the first transmitter-gated channel to be purified and cloned, and five neuromuscular nicotinic receptor subunits have been identified. Each of these subunits has four membrane-spanning regions, and a pair of cysteine residues located 13 amino acids apart in the extracellular region of the protein (see Fig. 1.5–8). These cysteine residues form a disulfide bridge that contributes to transmitter binding. The nicotinic ion channel is a nonselective cation channel permeable to Na^+, K^+, and Ca^{2+}. Membrane-spanning regions form the ion channel, with the second transmembrane region lining the channel pore. Muscle nicotinic subunits assemble to form a pentamer with a stoichiometry of $\alpha 2$, β, δ, and γ or ε depending on the age of the animal. Nicotinic receptors are part of a superfamily that includes $GABA_A$, glycine, and serotonin type 3 (5HT-3) receptors. $GABA_A$ and glycine receptors are anion selective, passing primarily Cl^- in physiologic solutions, whereas nicotinic and 5HT-3 receptors are cation selective. Because all members of the family contain the characteristic extracellular cysteine residues, the family is referred to as the CYS-LOOP family. Differences in charges on amino acids at the entrance to the channel pore determine whether the channel passes cations or anions. A crystal structure of a human homomeric β3 $GABA_A$ receptor shows the GABA channel to have a cylindrical structure about 110 Å high and 60 to 80 Å wide, extending 65 Å into the extracellular space. A positively charged ring of amino acids halfway down the channel allows anion binding and determines ion selectivity. A closed gate at the base of the pore provides a mechanism for desensitization. Cryo-EM and crystallization studies continue to reveal nuances of structure–function relationships in these receptors.

Ligand-gated ion channels gated by extracellular ATP (P2X receptors) have structures somewhat similar to inwardly rectifying K^+ channels (see Figs. 1.5–7 and 1.5–8). ATP receptors have two membrane-spanning regions and a pore-forming region (P-loop) connected by a large loop of extracellular amino acids. A major difference between P2X receptors and inwardly rectifying K^+ channels is that P2X receptors are mostly extracellular, whereas the bulk of the K^+ channel is intracellular. P2X channels are cation selective and have high Ca^{2+} permeability. These receptors participate in fast excitatory synaptic transmission in certain regions of the CNS, including the thalamus. Native ATP receptors may consist of combinations of P2X subunits.

Ionotropic glutamate receptors differ from cys-loop receptors in having three membrane-spanning regions and a re-entrant P-loop between the first and second transmembrane regions that does not completely cross the membrane (see Fig. 1.5–8). The P-loops in glutamate-gated channels, like those in voltage-gated ion channels, line the ion channel. A difference from voltage-gated channels is that glutamate receptor P-loops enter the membrane from the cytoplasmic side. Crystallographic data on AMPA-type glutamate receptors indicate that these channels contain four subunits, each of which binds a glutamate molecule. Subunits form a dimer of dimers, and the receptor resembles a letter “Y” when viewed perpendicular to the membrane. The glutamate binding region has a hinged cleft resembling a venus flytrap. When agonists bind, the cleft closes to varying degrees depending on the ligand. This imparts structural changes necessary for ion channel opening. Competitive antagonists that block glutamate binding stabilize the open cleft configuration of the binding pocket, providing an explanation for how an agent can bind to the agonist recognition site but not produce the conformational changes that cause channel opening. A crystal structure for heterotetrameric GluN1-GluN2B NMDA receptors reveals differences from AMPA receptors with a more tightly packed amino-terminal domain and ligand-binding domain, possibly explaining why the amino terminus more effectively regulates ion channel activity in these receptors. In contrast to the Y-like shape of AMPA receptors, NMDA receptors resemble a hot-air balloon with an extracellular "balloon" and a transmembrane "basket." Further complicating the picture, it appears that the majority of NMDA receptors have a triheteromeric structure with GluN1 subunits in combination with two different GluN2 subunits. All glutamate receptors share a highly conserved motif that regulates channel gating despite major differences in how gating occurs in different receptor subtypes and modulation of gating by pharmacologic agents. There are also a variety of genetic mutations that alter the gating of glutamate receptors and result in complex neuropsychiatric illnesses.

GPCRs have structures that differ completely from ligand-gated ion channels. These receptors have seven membrane-spanning regions (see Fig. 1.5–8), and many GPCRs exist as homo- or heterodimers. Transmitter binding occurs in a pocket formed by the intramembranous domains of the receptor, except for the glutamate family of GPCRs, where binding is in a large extracellular amino-terminal region. The coupling of GPCRs to their G-protein occurs at intracellular loops of the receptor. Adding to the complexity of signaling, the binding of certain neurotransmitters to receptors that are ligand-gated ion channels can also trigger intracellular effects akin to GPCRs in the absence of ion flux through the channels. Examples of metabotropic effects via ligand-gated channels include glutamate acting at NMDA and kainate receptors.

CLINICAL ASPECTS OF ION CHANNELS

Rapid signal processing in the brain requires action potential firing and fast neurotransmission and hence depends critically upon the activity of ion channels. In this final section, we highlight some areas where understanding the function and mechanisms of ion channels is particularly important for psychiatry.

Oscillatory Neuronal Firing and Complex Behavioral States

Complex behavioral states, including wakefulness, attention, mood, and sleep, require a coherent activity within and between neural circuits. Neuronal networks show rhythmic activity (oscillations) over a wide range of frequencies from less than 1 to several 100 cycles/s (Hz). These oscillations provide energy-efficient mechanisms that determine optimal input and output frequencies of brain regions and provide coherence in functional networks. Activity in oscillating circuits involves the interplay of intrinsic electrical properties of neurons, the local balance of excitatory and inhibitory inputs (E/I balance), and sustained effects of more diffusely acting neuromodulator systems, including muscarinic and monoaminergic systems. Certain neurons express voltage-gated conductances that allow them to fire rhythmically and spontaneously, thus having properties of a PACEMAKER or OSCILLATOR. For example, cerebellar inferior olivary neurons fire action potentials spontaneously and sustain this firing for relatively long periods in the absence of external inputs. These neurons fire conventional fast Na^+ spikes that provide the depolarization needed to open HVA Ca^{2+} channels. In turn, Ca^{2+} influx activates a Ca^{2+}-dependent K^+ conductance that rapidly hyperpolarizes the membrane. When the membrane hyperpolarizes, LVA Ca^{2+} channels open and depolarize the membrane potential to the threshold for firing Na^+ spikes, which then activates another cycle. In inferior olivary neurons, it is the properties of the LVA Ca^{2+} channels that foster oscillatory firing. LVA channels are inactivated at the neuronal resting membrane potential but become activatable when the membrane is hyperpolarized to rest. Thus, hyperpolarization becomes a priming stimulus that promotes LVA channel opening. Pacemaker firing of inferior olivary neurons drives Purkinje neurons in the cerebellum to RESONATE at the inferior olivary neuron's preferred firing frequency. This resonating circuit contributes to the physiologic resting tremor that oscillates at about 10 cycles/s.

In thalamocortical neurons, network oscillations are driven by intrinsic pacemaker conductances in coordination with synaptic connections. This activity drives EEG changes during different stages of sleep and vigilance. Thalamocortical neurons exhibit two distinct activity states. During sleep, neurons show synchronized rhythms that resemble δ (1- to 3-Hz activity) and other slow waves on the EEG. During wakefulness and rapid eye movement (REM) sleep, these neurons exhibit tonic activity. LVA calcium channels are important contributors to thalamocortical network activity. The transition from sleep to wakefulness is mediated by the depolarization of thalamic reticular neurons and the inactivation of LVA channels. Specific abnormalities in thalamocortical neurons also contribute to the generation of 3 Hz spike and wave activity observed in childhood absence epilepsy. In spike-wave discharges, the interplay between LVA currents and H-currents appears to be critical in generating abnormal firing. Drugs targeting these channels (e.g., ethosuximide, an inhibitor of LVA channels, and lamotrigine, an activator of H-channels) are useful clinically. Mutations in Cav3.2 LVA channels have been associated with absence epilepsy.

In some brain regions, outputs of pacemaker cells are mediated by fast excitatory or inhibitory transmitters. However, some neurons can fire bursts of action potentials. BURSTS are periods of frequent spike firing followed by quiescent periods. This type of firing can drive activity in a local or distributed neural network and in the subiculum, contributing to synaptic plasticity via LVA Ca^{2+} channels. In addition, burst-like firing can provide sufficient intracellular Ca^{2+} to stimulate the release of peptide transmitters. In turn, the slow actions of certain neuropeptides can alter the frequency of oscillatory firing and bursting. An example of this is the repeated firing that occurs when spike frequency adaptation is inhibited by blocking Ca^{2+}-activated K^+ conductances. In this case, both the intrinsic electrical properties of neurons and the effect of modulatory transmitters conspire to determine a background level of activity (or tone) in specific neuronal systems.

There is considerable interest in understanding cellular, synaptic, and network mechanisms contributing to neuro-oscillations, and these mechanisms hold significant promise for translating observations in animals to human brain function and behavior. An example involves recent studies of behavioral dissociation induced by ketamine. Ketamine is an anesthetic, and ketamine-induced dissociation in rodents is associated with the generation of 1- to 3-Hz rhythms in layer 5 neurons of the retrosplenial cortex, an area important in circuits governing vigilance and conscious effort. These changes result from the effects of ketamine on HCN-1 pacemaker channels. Intriguingly, similar localized rhythms in the deep posteromedial cerebral cortex were observed in a patient who experienced dissociative symptoms prior to seizures, providing a proof of concept for extrapolation of the rodent studies to humans. Other studies have emphasized the increased power of γ-oscillations as a possible hallmark of ketamine’s antidepressant effects.

“E/I Balance” and Psychiatric Illnesses

While the pathophysiology of psychiatric disorders remains uncertain, there is increasing interest in determining whether shifts in relative regional excitation (E) and inhibition (I) (E/I BALANCE) contribute to psychiatric symptoms at a neural network level. According to this hypothesis, both synaptic and intrinsic excitability changes contribute to neurocircuitry abnormalities. In schizophrenia and autism, there appears to be augmented excitation relative to inhibition in certain brain regions, and this increased E/I ratio may account for abnormal hyperactivity in frontal and hippocampal circuits that underlie cognitive and social defects. The E/I balance hypothesis is consistent with functional neuroimaging findings and increases in high-frequency cortical oscillations recorded by electroencephalography (EEG) and magnetoencephalography (MEG). Changes in β- and γ-oscillations (activity above 20 Hz) are thought to contribute to defects in sensory processing and learning and inefficient information transfer within and between neural networks. Specific classes of inhibitory GABAergic interneurons play a critical role in regulating these high-frequency rhythms by influencing the firing frequency of excitatory glutamatergic (pyramidal) neurons.

There are few direct tests of the E/I balance hypothesis in humans, but studies in mouse models have provided tentative support. Studies using optogenetics employing cell type–specific expression of light-activated ion channels (e.g., channelrhodopsins) to depolarize specific neuronal types for prolonged periods have reproduced increases in high frequency (30- to 80-Hz) cortical γ-oscillations and some altered social behaviors akin to those observed in human illnesses. These electrical and behavioral changes are reversed by increasing the activity of specific GABAergic inhibitory interneurons that express the calcium-binding protein parvalbumin (PV). PV^+ interneurons are fast-spiking cells that serve as a model for understanding how the expression of specific classes of ion channels drives unique functions, allowing interneurons to regulate rhythmic activity within and across brain regions. Simulation studies provide further support for the E/I balance hypothesis and indicate that local feedback inhibition enhances information capacity and discrimination accuracy of more global and distributed networks. These studies suggest that E/I balance within specific brain regions helps to regulate regional spontaneous activity and information transfer across brain regions at large-scale levels.

While the E/I balance hypothesis is still under investigation, this concept has a significant influence on clinical neuroscience. Efforts to manipulate E/I balance and the plasticity of excitatory (glutamate) synapses are leading to novel ideas about the treatment of developmental neurocognitive disorders that have heretofore proven therapeutically intractable, including fragile X syndrome, Rett syndrome, and autism spectrum disorders, among others.

Ion Channels and Neuropsychiatric Disorders

There is increasing evidence that defects in ion channels (CHANNELOPATHIES) contribute to the pathophysiology of neuropsychiatric illnesses. Channelopathies involve the altered function of specific ion channels as a result of genetic mutations, transcriptional abnormalities, protein folding defects, or immune processes, and multiple ion channels have been implicated as causes of various syndromes.

Given the importance of ion channels in neuronal excitability, the observation that epilepsy, a group of disorders characterized by recurrent bouts of abnormal electrical activity, is associated with mutations in specific ion channels is not surprising. BFNC are associated with mutations in K^+ channel genes, *KCNQ2* and *KCNQ3*. These proteins (Kv7.2, 7.3) form M-channels that produce slowly activating and slowly inactivating K^+ currents that blunt neuronal firing. The expression of a single mutant allele of either gene—BFNC is a dominantly inherited disorder—decreases channel expression sufficiently to result in neuronal hyperexcitability. Generalized epilepsy with febrile seizures plus (GEFS+) is an autosomal dominant illness that results from mutations in Na^+ channel subunits. The role that ion channels play in these rare genetic syndromes suggests that idiopathic epilepsy might also reflect channelopathies and perhaps result from interactions between genetic defects in ion channels and adverse environmental exposures. An exome sequencing study examining more than 230 ion channel genes in sporadic idiopathic epilepsy, the most common neuronal excitability disorder, found rare missense variants in both epilepsy patients and unaffected individuals, including deleterious mutations in both groups. These results strongly suggest that ion channel genes confer illness risk, but illness expression depends on other factors. Identification of defective channel genes, nonetheless, offers an opportunity to improve clinical outcomes based on the potential for personalized treatments.

The calcium channel *CACNA1A* gene that encodes the α_1-subunit of HVA P/Q-type Ca^{2+} channels ($Ca_v2.1$) is associated with several rare genetic diseases. Familial hemiplegic migraine is an autosomal dominant illness with childhood onset and auras that include transient hemiparesis or hemiplegia. In some families, the disorder is associated with progressive ataxia. Familial hemiplegic migraine is genetically heterogeneous, with about 50% of cases involving mutations in *CACNA1A*; more than a dozen mutations are associated with illness risk. How the mutations cause, illness is not entirely clear because the most frequently identified mutation reduces Ca^{2+} currents, while another mutation increases Ca^{2+} flux. Episodic ataxia type 2 (EAT2) is another dominantly inherited disorder associated with mutations in *CACNA1A*. Patients with EAT2 experience episodes of nystagmus and ataxia; for some, the disorder is progressive and results in cerebellar atrophy, with about 50% of patients having migraine. Most mutations associated with EAT2 grossly disrupt protein expression, although several are point mutations that change only a single amino acid. Expression of the gene with one of the point mutations results in complete loss of function without a change in protein expression. Finally, a type of autosomal dominant spinocerebellar ataxia (SCA6) is linked to a mutation that causes expanded glutamine repeats in the carboxyl terminus of the $Ca_v2.1$ protein. The diverse syndromes associated with altered *CACNA1A* gene highlight the heterogeneity that is expected from the exploration of genes for psychiatric disorders.

Genetic studies indicate that certain pain syndromes are also channelopathies involving altered Na^+ channel function and expression. Mutations in the *SCN9A* gene on chromosome 2 that encodes $Na_v1.7$ are implicated in both paroxysmal extreme pain disorder (PEPD) and congenital inability to experience pain. In the case of PEPD, the causative mutations result in fast inactivation of the channel with a persisting tonic Na^+ current. In pain insensitivity disorder, the mutations result in loss of function. These syndromes are instructive not only from the perspective of ion channels and disease but also in guiding future work aimed at developing more specific pain therapies.

Some evidence indicates that mutations in ion channels also contribute to illnesses that present with psychiatric symptoms. For example, polymorphisms in SK calcium-activated K^+ channels are associated with certain forms of psychosis. A polymorphism in the *KCNH2* gene that encodes a K^+ channel is associated with schizophrenia and predicts lower intelligence and altered memory-related changes in functional neuroimaging. Other work suggests the involvement of BK channels in a form of autism and mental retardation, and mutations in T-type (LVA) calcium channels have been linked to autism. Similarly, a polymorphism in the α5 neuronal nicotinic acetylcholine receptor subunit is linked to nicotine dependence. This latter polymorphism is in a coding region of the protein in a position that influences ion channel kinetics. Large-scale studies have revealed genetic loci in schizophrenia that likely code several voltage-gated ion channels, including α1C- and β2-subunits of L-type (HVA) Ca^{2+} channels, Kv2.1 delayed rectifier K^+ channels, ClC-3 voltage-activated Cl^- channels and HCN1 channels, among others. These observations suggest that understanding how genetic and environmental factors influence membrane excitability will become increasingly important in psychiatry. Changes in ion channel function are likely to contribute to diagnosis, pathophysiology, and response to psychotropic medications, including predispositions to serious side effects, including cardiac arrhythmias.

Certain autoimmune disorders affecting ion channel expression and function can also result in complex neuropsychiatric syndromes, including new-onset psychosis, delirium, and seizures. Perhaps the best example is anti-NMDA receptor encephalitis (NMDARE), now thought to be the most common cause of autoimmune encephalopathy. In this disorder, patients develop antibodies against the NMDAR GluN1 subunit that result in receptor cross-linking and internalization, leading to NMDAR hypofunction and defects in synaptic plasticity. Defects in synaptic potentiation appear to contribute to defects in working memory observed in these patients. Most patients with NMDARE improve with immunotherapies, but some are left with residual and/or relapsing symptoms. Thus, there is interest in whether agents that enhance NMDAR ion channel function could be alternative treatment strategies; preclinical studies provide early proof of concept for the use of NMDAR-positive modulators. Autoantibodies have also been found against other receptors and ion channels, including AMPA-type glutamate receptors, certain GABA receptors, and voltage-activated ion channels.

Homeostatic Plasticity and Psychiatric Therapeutics

The brain is a remarkably plastic organ capable of learning, remembering, and adapting to complex environments. Much of this plasticity involves short- and longer-term, use-dependent changes in the function of synapses, although changes in neuronal excitability can also contribute. Learning is typically thought to involve experience-dependent synaptic changes that require coordinated pre- and postsynaptic activity (called "Hebbian" plasticity after Donald Hebb, who postulated the importance of these types of changes for learning). Hebbian plasticity includes long-term potentiation (LTP) and long-term depression (LTD), which result from coincidental activity of pre- and postsynaptic neurons and are typically triggered by activation of NMDA-type glutamate receptors. Defects in Hebbian plasticity are thought to contribute to defects in learning that occur in the context of stress and neuropsychiatric disorders.

Brain regions can also alter their function by non-Hebbian mechanisms in which neurons and networks adjust excitability in response to persistent increases or decreases in activity. These "homeostatic" changes reflect adjustments to help maintain E/I balance and optimal function of brain networks. These changes include alterations in synaptic function (called synaptic scaling) as well as changes in neuronal excitability (including changes in the function and expression of ion channels). Unlike Hebbian plasticity, homeostatic plasticity is unlikely to drive information storage but rather helps maintain longer-term function in response to local changes in excitation and inhibition. Some of these changes do not require input from other cells and are called "cell autonomous." In recent years, there has been interest in determining whether certain treatments in psychiatry modulate neural networks via homeostatic mechanisms. For example, the antidepressant ketamine rapidly upregulates excitatory synaptic responses in the hippocampus and cortex in a non-Hebbian fashion, while lithium dampens synaptic function in a similar way. There is also interest in determining whether other psychiatric treatments have similar homeostatic effects, including ECT and other brain stimulation methods (see below).

Another form of plasticity in neural circuits appears not to result directly in persisting changes in synaptic responses or excitability but rather modulates the ability to induce Hebbian plasticity. This latter form of neuronal modulation is called "metaplasticity" (or "plasticity of plasticity") and involves shifting the relative ease with which Hebbian changes can be induced. Metaplastic regulation shares molecular mechanisms with

Hebbian and homeostatic plasticity and appears to contribute to the dampening of Hebbian plasticity and learning observed under certain physiologic, pathologic, and pharmacologic conditions, including ongoing excitatory synaptic activity, stress, and exposure to certain drugs, including ethanol and ketamine. Taken together, the three general forms of plasticity outlined here offer potentially novel targets for therapeutics that could help to improve cognitive, emotional, and motivational dysfunction associated with psychiatric illnesses. Importantly, however, defects in these forms of plasticity may not be the primary causes of neuropsychiatric illnesses but rather provide strategies to improve neural network function for therapeutic purposes.

Neural Excitability and Brain Stimulation Methods in Neuropsychiatry

Brain stimulation (neuromodulation) methods are playing an increasing role in the treatment of psychiatric and neurologic disorders. These methods include electroconvulsive therapy (ECT), vagus nerve stimulation (VNS), repetitive transcranial magnetic stimulation (rTMS), deep brain stimulation (DBS), and transcortical direct current stimulation (tDCS). The selection of optimal stimulation parameters for these treatments requires knowledge about factors that regulate neuronal excitability and the effects of electrical stimulation on neuronal function. A major advance in ECT was the recognition that electrical stimuli play a key role in therapeutic and adverse effects. Because the goal of an ECT session is to cause a generalized brain seizure while minimizing cognitive side effects, understanding factors that determine seizure threshold and optimizing stimulus parameters becomes extremely important. To stimulate nerve cells, brief square-wave pulses of electrical current (0.1 to 2.0 ms) are much more effective than prolonged pulses or sine-wave stimuli. The rate of delivery of the current pulses is also important, with lower frequencies (30 to 40 Hz) being more efficient than very high-frequency trains (>100 Hz). These features reflect the fact that the entrainment of neurons in a seizure is more likely using parameters that mimic neuronal firing patterns. Long current pulses (particularly sine waves) are inefficient because of the relatively slow onset and offset of the stimulus allowing neuronal adaptation and because much of the stimulus is delivered during the absolute and relative refractory periods when neurons are less excitable. Similarly, very high frequencies of stimulation result in pulses being delivered during refractory periods. For reasons that are unclear, ultrabrief electrical pulses (~0.2-ms duration) are effective for unilateral but not bilateral ECT. The degree to which electrical doses exceed the seizure threshold is also important. For bilateral ECT, electrical doses that are ~1.5 times the threshold result in highly effective treatments that minimize cognitive impairment. For nondominant hemisphere (unilateral) ECT, these same electrical doses are ineffective, and stimuli that are five to six times the seizure threshold are required for clinical improvement. These observations indicate that electrical dosing and stimulation parameters have an important impact on ECT outcomes, and both bilateral and unilateral ECT can be made effective or ineffective depending on the electrical stimulus.

More recent additions to brain stimulation methods, particularly DBS, also require an understanding of the effects of electrical stimulation on neuronal activity and raise interest in the potential role of homeostatic plasticity as a therapeutic mechanism. The parameters used for DBS typically employ brief (0.06 ms) pulses administered continuously at a high frequency (>100 Hz). Modeling studies suggest that this stimulation has complex effects on network function with suppression of intrinsic firing at neuronal cell bodies in the stimulated region but enhanced axonal responses and increased efferent output to downstream targets. Advances in DBS are exploring a variety of targets for modulation of mood, memory, and other complex neuropsychiatric functions, including the use of stimulation frequencies and closed-loop techniques that more closely mimic ongoing brain rhythms to enhance neural physiology.

Progress is also being made in developing more effective ways to use other neuromodulation methods, including rTMS and VNS. rTMS involves the administration of electromagnetic pulses to specific brain regions (typically dorsolateral prefrontal cortex) that either excite (10 to 20 Hz to the left hemisphere) or inhibit (1 Hz to the right hemisphere) cortical function. These parameters mimic stimulus parameters used to induce LTP (high frequencies) or LTD (low frequencies) in preclinical studies in the hippocampus and cortex. Advances in rTMS are using stimuli that more closely mimic intrinsic brain rhythms (e.g., intermittent θ-pulse stimulation) and the use of neuroimaging-guided localization of the magnetic coil to direct rTMS pulses to specific cortical networks that contribute to attention and cognitive control; some evidence suggests that targeting different cortical networks based on clinical subtypes of depression may lead to more effective treatments.

VNS remains an intriguing but poorly understood form of neuromodulation that involves chronic, intermittent stimulation of afferent connections of the vagus nerve via an electrical pulse generator implanted in the chest. VNS has been used to treat refractory epilepsy for over 20 years and has been approved for use in treatment-resistant major depression. Again, optimal stimulation parameters are not completely defined at present, but VNS has several intriguing features that warrant study, including observations that full effects of VNS may take more than 6 months to a year of use to manifest but, once manifest, may persist for long periods of time (years) with continuing stimulation. Mechanisms underlying the long latency to effects and the persisting efficacy of VNS remain uncertain but could reflect homeostatic changes in the function of specific neurocircuits.

Although parameters for clinical benefit continue to be optimized, current neurostimulation/modulation techniques may ultimately prove insufficiently nuanced to manipulate circuits for maximum clinical benefit while limiting undesirable side effects. In this regard, techniques currently in use in animals to manipulate circuits with high selectivity may ultimately find their way to humans as next-generation neurostimulation treatments. For instance, optogenetics allows cell type-selective, local control over circuit behavior. Similarly, chemogenetics employs virally introduced designer receptors responsive to synthetic ligands, allowing experimenters to alter the electrical behavior of restricted, genetically defined populations of neurons. Since their introduction in the early 2000s, these techniques have already revolutionized studies of the role of neural circuits in animal behavior. Hurdles to human implementation include the requirement for gene therapy to introduce foreign channels and implantation of a fiber optic stimulator (optogenetics) or delivery of a foreign chemical (chemogenetics). However, optimizations to increase sensitivity and selectivity of the methods and to decrease invasiveness are published regularly and could reduce barriers to clinical implementation.

SUMMARY

Major advances are occurring in understanding the structure, function, and diversity of voltage-gated and ligand-gated ion channels. The properties of ion channels are being harnessed to yield an unprecedented understanding of the roles of neural circuits in behavior, cognition, and emotion. Efforts to modulate the electrical properties of neurons offer potential new vistas in neuropsychiatric therapeutics and are fundamental to modifying brain networks and to developing innovative treatments. Although electrical properties underlying neuronal excitability are relatively stereotyped, the numerous ion channels contributing to neuronal excitability offer huge flexibility in the control of cellular, regional, and network neural activity. Determining how alterations in ion channels contribute to neurodevelopment, behavior, cognition, and clinical syndromes remains a major goal of work in this field.

ACKNOWLEDGMENTS

The authors thank members of the Taylor Family Institute for Innovative Psychiatric Research for many valuable discussions. Work in the authors' labs is supported by the National Institute of Mental Health and the National Institute on Alcohol Abuse and Alcoholism.

FURTHER READINGS

Alexander SPH, Mathie A, Peters JA, et al; CGTP Collaborators. The concise guide to pharmacology 2019: ion channels. *Br J Pharmacol*. 2019;176:S142–S228.

Catterall WA. Sodium channels, inherited epilepsy, and antiepileptic drugs. *Annu Rev Pharmacol Toxicol*. 2014;54:317–338.

Deco B, Ponce-Alvarez A, Hagmann P, Romani GL, Mantini D, Corbetta M. How local excitation-inhibition ratio impacts the whole brain dynamics. *J Neurosci*. 2014;34:7886–7898.

Dib-Hajj SD, Waxman SG. Sodium channels in human pain disorders: genetics and pharmacogenomics. *Annu Rev Neurosci*. 2019;42:87–106.

Dolphin AC, Insel PA, Blaschke TF, Meyer UA. Introduction to the theme “Ion channels and neuropharmacology: from the past to the future.” *Annu Rev Pharmacol Toxicol*. 2020;60:1–6.

Duncan AL, Song W, Sansom MSP. Lipid-dependent regulation of ion channels and G protein-coupled receptors: insights from structures and simulations. *Annu Rev Pharmacol Toxicol*. 2020;60:31–50.

Gandal MJ, Geschwind DH. Polygenicity in psychiatry—like it or not, we have to understand it. *Biol Psychiatry*. 2021;89:2–4.

Hille B. *Ion Channels of Excitable Membranes*. Sinauer Associates Inc.; 2001.

Hu H, Gan J, Jonas P. Fast-spiking, parvalbumin+ GABAergic interneurons: from cellular design to microcircuit function. *Science*. 2014;345:1255263.

Isaacson JS, Scanziani M. How inhibition shapes cortical activity. *Neuron*. 2011;72:231–243.

Jan LY, Jan YN. Voltage-gated potassium channels and the diversity of electrical signaling. *J Physiol*. 2012;590:2591–2599.

Javitt DC, Siegel SJ, Spencer KM, et al. A roadmap for development of neuro-oscillations as translational biomarkers for treatment development in neuropsychopharmacology. *Neuropsychopharmacol*. 2020;45:1411–1422.

Kavalali ET, Monteggia LM. Targeting homeostatic synaptic plasticity for treatment of mood disorders. *Neuron*. 2020;106:715–726.

Kopfer DA, Song C, Gruene T, Sheldrick GM, Zachariae U, de Groot BL. Ion permeation in K^+ channels occurs by direct Coulomb knock-on. *Science*. 2014;346:352–355.

Lee C-H, Lü W, Michel JC, et al. NMDA receptor structures reveal subunit arrangement and pore architecture. *Nature*. 2014;511:191–197.

Mankin EA, Fried I. Modulation of human memory by deep brain stimulation of the entorhinal-hippocampal circuitry. *Neuron*. 2020;106:218–235.

Miller PS, Aricescu AR. Crystal structure of a human $GABA_A$ receptor. *Nature*. 2014;512:270.

Nanou E, Catterall WA. Calcium channels, synaptic plasticity and neuropsychiatric disease. *Neuron*. 2018;98:466–481.

Perszyk RE, Myers SJ, Yuan H, et al. Genetic and pharmacological control of glutamate receptor channels through a highly conserved gating motif. *J Physiology*. 2020;598:3071–3083.

Pollak TA, Lennox BR, Müller S, et al. Autoimmune psychosis: an international consensus on an approach to the diagnosis and management of psychosis of suspected autoimmune origin. *Lancet Psychiatry*. 2020;7:93–108.

Roux B. Ion channels and ion selectivity. *Essays Biochem*. 2017;61:201–209

Shah MM. Cortical HCN channels: function, trafficking and plasticity. *J Physiol*. 2014;592:2711–2719.

Shen Y, Campbell RE, Côté DC, Paquet ME. Challenges for therapeutic applications of opsin-based optogenetic tools in humans. *Front Neural Circuits*. 2020;14:41.

Spillane J, Kullmann DM, Hanna MG. Genetic neurological channelopathies: molecular genetics and clinical phenotypes. *J Neurol Neurosurg Psychiatry*. 2016;87:37–48.

Veit G, Roldan A, Hancock MA, et al. Allosteric folding correction of F508del and rare CFTR mutations by elexacaftor-tezacaftor-ivacaftor (Trikafta) combination. *JCI Insight*. 2020;5:e139983.

Vesuna S, Kauvar IV, Richman E, et al. Deep posteromedial cortical rhythm in dissociation. *Nature*. 2020;586:87–94.

Warikoo N, Brunwasser SJ, Benz A, et al. Positive allosteric modulation as a potential therapeutic strategy in anti-NMDA receptor encephalitis. *J Neurosci*. 2018;38:3218–3229.

Womelsdorf T, Valiante TK, Sahin NT, Miller KJ, Tiesinga P. Dynamic circuit motifs underlying rhythmic gain control, gating and integration. *Nature Neurosci*. 2014;17:1031–1039.

Wong HHW, Rannio S, Jones V, Thomazeau A, Sjöström PJ. NMDA receptors in axons: there's no coincidence. *J Physiology*. 2021;599:367–387.

Zorumski CF, Rubin EH. *Psychiatry and Clinical Neuroscience: A Primer*. Oxford University Press; 2011.

1.6 Amino Acid Neurotransmitters

Joseph T. Coyle, M.D.

For over 70 years, the biogenic amines have dominated thinking about the role of neurotransmitters in the pathophysiology of psychiatric disorders. However, over the last two decades, evidence has accumulated from postmortem, brain imaging, and genetic studies that the amino acid neurotransmitters, in particular glutamic acid and γ-aminobutyric acid (GABA), play an important, if not central, role in the pathophysiology of a broad range of psychiatric disorders including schizophrenia, bipolar disorder, major depression, autism spectrum disorder (ASD), Alzheimer disease, and anxiety disorders. The recent ascendance of sufficiently powered genome-wide association studies (GWAS) has had a tremendous impact in linking risk sites on the human genome to genes associated with amino acid neurotransmitters, especially glutamic acid. This is not entirely surprising given the fact that virtually every neuron in the CNS is innervated by GABAergic and glutamatergic neurons. Consistent with this, the brain concentrations of synaptic GABA and glutamate are in the millimolar range, whereas biogenic amine and peptide neurotransmitters are in the micromolar range or lower. The mM concentrations of the neurotransmitter amino acids make them uniquely accessible for measurement in vivo in the human brain by magnetic resonance spectroscopy (MRS), which does not involve exposure to radiation like computer-assisted tomographic (CAT) or positron emission tomographic (PET) scans. The purpose of this chapter is to review the current understanding of amino acid neurotransmitters and to address their potential involvement in psychiatric disorders.

GLUTAMIC ACID

Glutamate mediates fast excitatory neurotransmission in the brain and is the neurotransmitter for approximately 80 percent of brain synapses, particularly those innervating with dendritic spines. The repolarization of neuronal membranes that have been depolarized by glutamatergic neurotransmission accounts for as much as 80 percent of the energy expenditure in the brain. The concentration of glutamate in the brain is 10 mM, the highest of all amino acids, of which approximately 20 percent represents the neurotransmitter pool of glutamate.

The postsynaptic effects of glutamate are mediated by two families of receptors. The first are the glutamate-gated cation channels that are responsible for fast neurotransmission. The second type of glutamate receptor is the metabotropic glutamate receptor (mGluR), which are G-protein–coupled receptors (GPCRs) like α-adrenergic receptors and dopamine receptors. The mGluRs primarily modulate glutamatergic neurotransmission.

Synthesis of Glutamate

Given the excitatory (and potentially neurotoxic) effects of glutamate, it is not surprising that it is excluded from the brain by the blood–brain barrier (BBB). Thus, glutamate in the brain must be synthesized de novo from glucose through the tricarboxylic acid cycle, which generates α-ketoglutarate. The α-ketoglutarate receives an amino group via a transaminase reaction, converting it to glutamic acid. Glutamate is in equilibrium with α-ketoglutarate, and virtually all glucose entering the brain is cycled through glutamic acid. The portion of glutamate dedicated to neurotransmission (approximately 20 percent) is actively sequestered in storage vessels by the vesicular glutamate transporter (vGluT).

A second metabolic pathway is particularly important for replenishing synaptic glutamate. This pathway exploits the intimate relationship between the glutamatergic synapse (i.e., the synaptic bouton and the postsynaptic spine) and the astrocytic endfeet that envelop the synapse. It is the astrocyte and not the glutamatergic nerve terminal that expresses the excitatory amino acid transporters (EAAT1 and 2) that remove glutamate from the synapse, thereby terminating its action. Within the astrocyte, glutamine synthetase, a cytosolic adenosine triphosphate (ATP)-dependent enzyme, catalyzes the conversion of glutamate to generate glutamine. Glutamine synthetase is expressed in glia but not in neurons. Glutamine is then released by the astrocyte and taken up by the glutamatergic terminal, where it is converted back to glutamate by phosphate-activated glutaminase, a mitochondrial enzyme. This process is known as the “glutamine cycle” and accounts for approximately 40 percent of brain glutamate turnover.

Major Glutamatergic Pathways in the Brain

All primary sensory afferent systems use glutamate as their primary neurotransmitter, including retinal ganglion cells, cochlear cells, trigeminal nerve, and spinal afferents. The thalamocortical projections that distribute afferent information broadly to the cortex are glutamatergic. The pyramidal neurons of the corticolimbic regions, the major source of intrinsic, associational, and efferent excitatory projections from the cortex, are glutamatergic. A temporal lobe circuit that figures importantly in the development of new memories is a series of four glutamatergic synapses: The perforant path innervates the hippocampal granule cells that innervate CA3 pyramidal cells that innervate CA1 pyramidal cells. The climbing fibers innervating the cerebellar cortex are glutamatergic as well as the corticospinal tracks.

Ionotropic Glutamate Receptors

Three families of ionotropic glutamate receptors have been identified on the basis of selective activation by conformationally restricted or synthetic analogs of glutamate. These include α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA), kainic acid (KA), and N-methyl-ᴅ-aspartic acid (NMDA) receptors (Fig. 1.6–1). Subsequent cloning revealed 16 mammalian genes that encode structurally related proteins, which represent the subunits that assemble into functional receptors. Glutamate-gated ion channel receptors are tetramers, and the subunit composition affects both the pharmacologic and the biophysical features of the receptor.

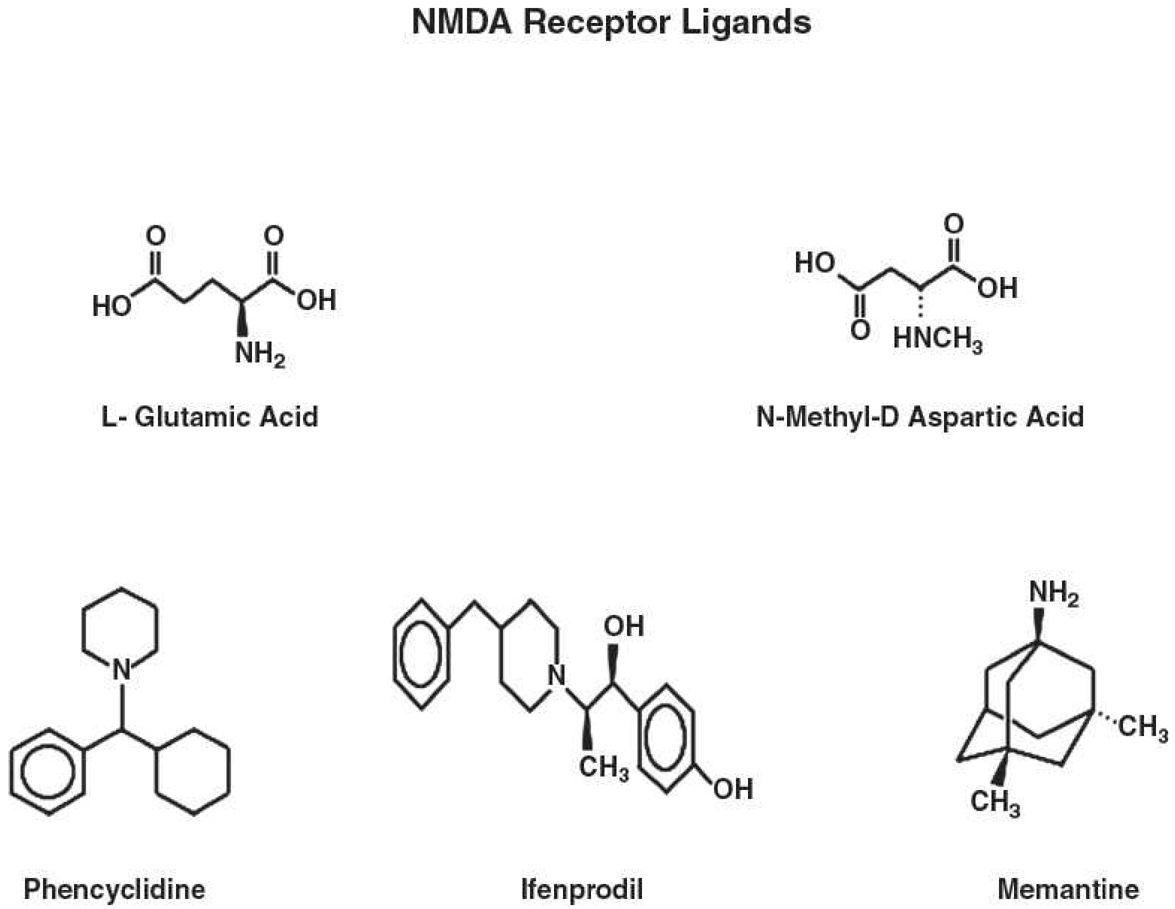
NMDA Receptor Ligands
O
O
HO
OH
NH2
L- Glutamic Acid
O
HO
OH
O
HNCH3
N-Methyl-D Aspartic Acid
N
Phencyclidine
OH
N
CH3
OH
Ifenprodil
NH2
CH3
CH3
Memantine

FIGURE 1.6–1. NMDA receptor ligands.

AMPA receptors (AMPARs) mediate the excitatory (depolarizing) postsynaptic currents (EPSCs) primarily responsible for excitatory neurotransmission and are broadly distributed in the CNS (Fig. 1.6–2). The AMPAR family consists of four subunits of GluR1 to GluR4. However, additional complexities affect AMPAR function. Alternative splicing of a 115 bp cassette for GluR1 to GluR4 messenger ribonucleic acid (mRNA) results in two forms (flip and flop) that give rise to receptors that differ in desensitization rate and regional distribution in the brain. In the second transmembrane domain, GluR1, 3, and 4 have a glutamine (Q) residue that results in high Ca^{2+} conductance, whereas GluR2 has an arginine (R) in this position that severely restricts Ca^{2+} passage and conducts only Na^+. mRNA editing by adenosine deaminase converts the codon for GluR2 from an arginine to a glutamine by a process known as Q/R editing. This radically increases the channel permeability to Ca^{2+} of AMPARs containing the edited GluR2 subunit. Similar mRNA editing mechanisms have been described for kainate receptor subunits GluR5 and 6.

The kainate receptor family consists of five subunits. GluR5 to GluR7 represent subunits that form glutamate-gated cation channels. KA1 and KA2 exhibit negligible channel activity but associate with GluR5 to GluR7 to form high-affinity kainate receptors. The role of kainate receptors is less clearly defined than that of AMPARs, but with their presynaptic localization on glutamatergic terminals, kainate receptors reduce glutamatergic neurotransmission when activated. Notably, kainate receptors can also function as GCPRs.

Seven genes encode subunits that comprise the NMDA receptor family. The NMDA receptor has several unique features (Fig. 1.6–3). First, the channel is blocked by magnesium (Mg^{2+}) at resting membrane potential. Thus, NMDA receptors are "silent" until activation of the AMPARs has sufficiently depolarized the neuronal membrane to relieve the Mg^{2+} blockade. The NMDA receptor requires the simultaneous binding of two ligands to separate recognition sites on two different subunits in order for the channel to open. On the NR1 subunit, which forms the channel, is a binding site termed the glycine modulatory site (GMS), for which glycine and D-serine are the endogenous ligands. The binding site for the neurotransmitter glutamate is located on the second nonchannel subunit (NR2A-D). In the forebrain, D-serine is the dominant coagonist and is synthesized in the synaptic spine of the postsynaptic neuron and released at the synapse where, as an autocrine agent, it primes the synaptic NMDARs to respond immediately to synaptic glutamate. The NMDAR has been described as a coincidence detector because three events must occur simultaneously for the channel to open. Thus, the NMDAR becomes functional only when a sufficient amount of presynaptic glutamate has been released to saturate the receptor, D-serine (or glycine) occupies the GMS, and the synaptic membrane is sufficiently depolarized AMPAR activation to remove the Mg^{2+} blockade. Typically, for this to occur, multiple, converging glutamatergic afferents must fire simultaneously.

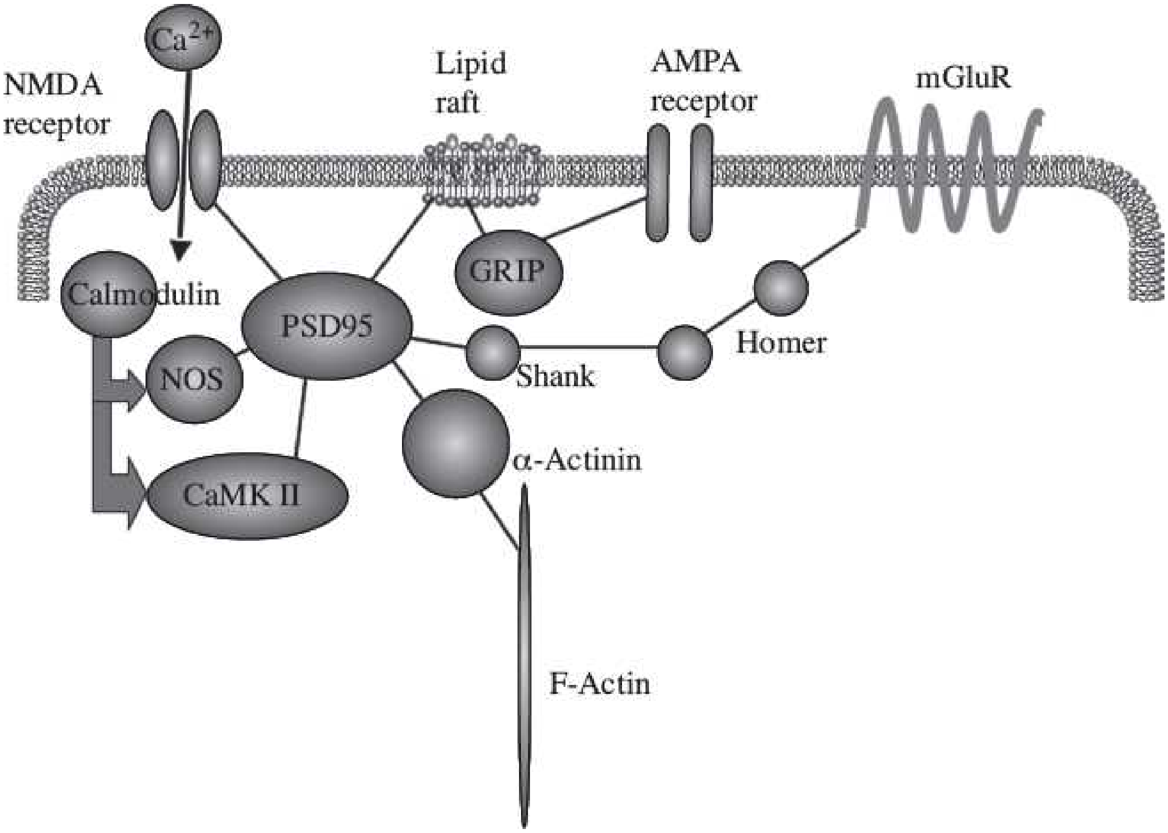

FIGURE 1.6–2. Schematic representation of the postsynaptic density (PSD) of the excitatory synapse. The NMDA receptor is bound to the principal organizing protein, PSD-95. Effector enzymes such as nitric oxide synthase (NOS) and calmodulin-activated protein kinase II (CaMKII) are also bound to PSD-95, keeping them in close proximity to the Ca^{2+} permeable NMDA receptor channel. PSD-95 is connected to the intracellular cytoskeletal protein, F-action by α-actinin. PSD-95 also is attached to neuronal membrane lipid rafts, which indirectly links it to the AMPA receptor through glutamate receptor–interacting protein (GRIP). Finally, postsynaptic group I mGluR receptors are linked to PSD-95 by the scaffolding proteins Shank and Homer. The gene-encoding Shank-3 is a risk gene for autism spectrum disorder.

The NMDAR channel is sufficiently large for Ca^{2+} to readily pass through it. Intracellularly, Ca^{2+} activates several kinases that ultimately affect gene expression in the neuron. For example, the immediate-early gene cFos is a sensitive surrogate for NMDAR activation. The type of NR2 subunit affects the pharmacology and biophysics of the NMDA receptor. For example, NR2B-containing receptors are much more Ca^{2+} permeable than NR2A-containing receptors. NR2A-containing receptors are expressed primarily in corticolimbic regions in the mature brain. NR2B is expressed at high levels in the immature cortex and decreases with maturation. NR2C is expressed primarily in the cerebellum, and NR2D is localized predominately to the cerebellum, the midbrain, and the brainstem.

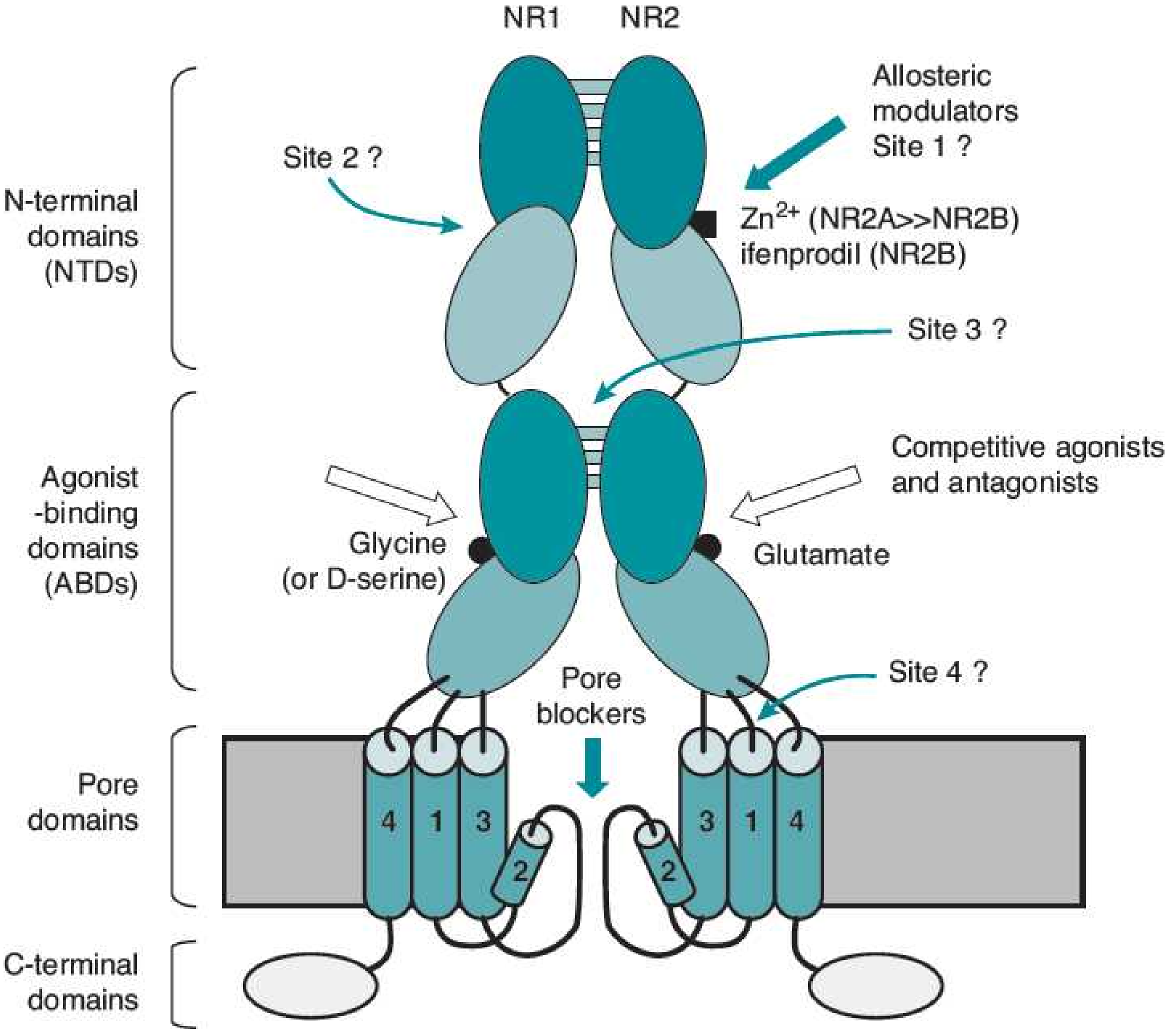
NR1
NR2
N-terminal domains (NTDs)
Site 2 ?
Allosteric modulators Site 1 ?
Zn2+ (NR2A>>NR2B)
ifenprodil (NR2B)
Site 3 ?
Agonist -binding domains (ABDs)
Competitive agonists and antagonists
Glycine (or D-serine)
Glutamate
Pore blockers
Site 4 ?
Pore domains
4
1
3
2
2
3
1
4
C-terminal domains

FIGURE 1.6–3. Schematic representation of the NMDA receptor. The NMDA receptor is a heterotetramer composed of the NR1 subunit that comprises the channel and the NR2 subunit, which contains the ligand binding site for the agonist glutamate. The channel is blocked at the resting membrane potential by Mg^{2+}. The glycine modulatory site on the NR1 subunit must be occupied by the endogenous agonists, D-serine or glycine, for the channel to open. The channel accommodates both Na^+ and Ca^{2+}. A polyamine site positively modulates the receptor. Within the channel is the binding site for dissociative anesthetics, which are use-dependent, noncompetitive inhibitors of the NMDA receptor.

Given their prominent role in learning and in excitotoxicity (see below), it is not surprising that NMDARs are among the most tightly regulated neurotransmitter receptors. As described above, two separate ligands must be bound to two distinct subunits on the NMDAR for it to function. In addition, there are binding sites for Zn^{2+} and H^+ that inhibit ion flux. A polyamine site enhances channel opening. Furthermore, the channel is sensitive to its redox state, which also affects ion currents. Subunit composition affects responses to modulators. For example, the NR2A subunit is much more sensitive to inhibition by Zn^{2+}, whereas the NR2B is differentially more sensitive to the polyamine site antagonist ifenprodil. The influx of Ca^{2+} via the NMDARs activates calmodulin, which then binds to the C-terminus of NR1 and reduces channel opening frequency and duration.

Metabotropic Glutamate Receptors

mGluRs are so designated because their effects are mediated by G-proteins and not by the influx of cations. All mGluRs are activated by glutamate, although their sensitivities vary remarkably. To date, eight mGluRs have been cloned. These genes encode seven membrane-spanning proteins that are members of the superfamily of GPCRs. They are subgrouped into three classes based upon amino acid sequence homology, agonist pharmacology, and signal transduction pathway utilized. Group I mGluRs, which includes mGluR1 and 5, activate PLC, presumably through G_Q. Group II includes mGluR2 and 3. Group III includes mGluR4, 6, 7, and 8. Group II and III mGluRs inhibit adenylyl cyclase through G_i-protein. The abundant neuropeptide N-acetylaspartylglutamate is a specific agonist at mGluR3.

mGluRs located postsynaptically modulate several channels and receptors. All three groups inhibit L-type voltage-dependent calcium channels, and groups I and II inhibit N-type calcium channels. In addition, mGluRs are reported to close voltage-dependent K^+ channels, thereby slowing depolarization and reducing excitability. Presynaptic mGluRs on both GABAergic and glutamatergic terminals inhibit neurotransmitter release, possibly by inhibiting the P/Q-type calcium channel.

Postsynaptic Density

Considerable advances have been made in characterizing the organization and dynamics of glutamate receptors at the postsynaptic density (PSD). The PSD is an osmophilic multiprotein complex that contains scaffolding proteins, cell adhesion molecules, and proteins for intercellular signaling pathways. A major scaffolding protein is PSD-95 (an acronym for *postsynaptic density* protein with a molecular weight of 95 kDa). PSD-95 contains several regions that bind other proteins. There are three PDZ domains (an acronym for PSD-95/disc large/zona occludens-1). The PDZ domains contain approximately 90 amino acids that bind the C-termini of proteins with complementary amino acid sequences. Neuroligin binds to the PDZ and extends into the synaptic cleft to bind to β-neurexin, which is anchored to the presynaptic component of the synapse. This arrangement stabilizes the synapse by connecting pre- and postsynaptic components. Two N-terminal cysteines of PSD-95 bind palmitic acid, which links the protein to lipid rafts in the plasma membrane. The NR2 subunit of the NMDAR binds to the PDZ domain. PSD-95 also binds α-actin, which is bound to filamentous actin, an important component of the cytoskeletal complex in the dendritic spine.

In contrast to NMDARs, AMPARs do not bind directly to PSD-95 but are associated with it indirectly by binding to intermediary proteins that bind to PSD-95. These include glutamate receptor–interacting protein (GRIP), protein interacting with C-kinase1 (PICK1), and synapse-associated protein of 97 kDa (SAP-97). In addition, there are the transmembrane AMPA receptor regulatory proteins (TARPs), which are involved in transporting AMPARs to and intercalating them within the PSD. Whereas the number of NMDARs in mature synapses tends to be relatively constant, the number of AMPARs in the PSDs varies tremendously. In fact, there are some PSDs that contain no AMPARs but do contain NMDARs. They are referred to as “silent synapses” because glutamate has no excitatory effects since the NMDARs are inactivated at the resting membrane potential with a few exceptions. Except for mGluR7, postsynaptic mGluRs are tethered to the periphery of the PSD by binding to two scaffolding proteins, homer and shank, the latter of which binds to PSD-95. Notably, allelic variants of gene-encoding proteins that comprise the PSD, including neurexin, neuroligin, and shank, have been implicated in autism.

The Role of Astrocytes

Specialized endfeet of the astrocyte surround glutamatergic synapses. The astrocyte expresses the two Na^+-dependent glutamate transporters that play the primary role in removing glutamate from the synapse, thereby terminating its action: EAAT1 and EAAT2 (excitatory amino acid transporter 1 and 2). The neuronal glutamate transporter, EAAT3, is expressed in upper motor neurons, whereas EAAT4 is expressed primarily in cerebellar Purkinje cells; and EAAT5 are restricted to the retina. Mice homozygous for null mutations of either EAAT1 or EAAT2 exhibit elevated extracellular glutamate and excitotoxic neurodegeneration. Notably, several studies have described the loss of EAAT2 protein and transport activity in the ventral horn in amyotrophic lateral sclerosis (ALS).

The astrocytes express AMPARs so that they can monitor synaptic glutamate release. GlyT1, which maintains subsaturating concentrations of glycine in the extracellular space, is expressed on the astrocyte plasma membrane. GlyT1 transports three Na^+ out for each molecule of glycine transported into the astrocyte. This stoichiometry results in a robust reversal of the direction of transport when glutamate released in the synapse activates the AMPARs on the astrocyte, thus depolarizing the astrocyte. Glycine release in the synapse by GlyT1 is coordinated with glutamatergic neurotransmission. D-Serine, which is found primarily in the forebrain, is synthesized by serine racemase (SR), which is expressed primarily in pyramidal neurons and, to a lesser extent, GABAergic neurons. D-Serine levels are also determined by D-amino acid oxidase (DAAO), with low D-serine levels in the cerebellum and brainstem where DAAO expression is high, and high D-serine levels are found in corticolimbic brain regions where DAAO expression is quite low. In contrast, the expression of GlyT1 is highest in the cerebellum and brainstem. This distribution suggests that D-serine is the primary coagonist of the NMDARs in the forebrain, whereas glycine is more prominent in the brainstem and cerebellum. Nevertheless, GlyT1 inhibitors have procognitive effects in experimental animals and are being tested clinically for the treatment of cognitive impairment and negative symptoms in schizophrenia.

Plasticity and Glutamatergic Neurotransmission

Hebb proposed that learning and memory involved use-dependent changes in synaptic efficacy. Neurophysiologic studies, originally exploiting the identifiable glutamatergic Schaffer collateral synapse on the hippocampal CA1 pyramidal cell, showed that a brief period of intense stimulation of the Schaffer collateral (100 Hz) resulted in a subsequent persistent increase in the efficacy of synaptic neurotransmission at these synapses. This phenomenon is known as long-term potentiation (LTP) and is quite widespread at glutamatergic synapses. In contrast, a period of stable low-frequency stimulation of the glutamatergic axons results in a persistently reduced efficacy of synaptic neurotransmission, a phenomenon known as long-term depression (LTD). LTP in these hippocampal studies requires the activation of NMDARs, as demonstrated by the fact that it is blocked by NMDAR antagonists such as the dissociative anesthetics ketamine and phencyclidine (PCP). Conditions resulting in blockade of LTP in the hippocampus are associated with impairments in the acquisition of new memories. The nature of these plastic changes has been the focus of intense research. The persistent changes in synaptic efficacy in LTP and LTD result from the insertion (LTP) or removal (LTD) of AMPARs from the PSDs of affected synapses. Thus, in contrast to the NMDARs, the AMPARs are quite dynamic, and their synaptic function is controlled through trafficking in and out of the synapse.

The extinction of conditioned fear has also been shown to be an active process mediated by the activation of NMDARs in the amygdala. Treatment of rats with NMDAR antagonists prevents the extinction of conditioned fear, whereas treatment with the GMS partial agonist D-cycloserine facilitates the extinction of conditioned fear. (D-Cycloserine is an antibiotic used to treat tuberculosis that is a partial agonist at the NMDAR with 50 percent efficacy of glycine). To determine whether the phenomenon generalizes to humans, patients with acrophobia were administered either placebo or a single dose of D-cycloserine along with cognitive-behavioral therapy (CBT). D-Cycloserine plus CBT resulted in a highly significant reduction in acrophobic symptoms that persisted for at least 3 months as compared to placebo plus CBT. Other placebo-controlled clinical trials support the notion that D-cycloserine is a robust enhancer of CBT, suggesting that pharmacologically augmenting neural plasticity may be used to bolster psychological interventions. D-Cycloserine may enhance cognitive remediation in patients with schizophrenia, although the evidence for this is limited to pilot data.

Glutamate-mediated synaptic plasticity is not only functional; it is also structural. Dendritic spines are dynamic appendages. Such dynamics have cleverly been demonstrated in real-time in vitro and in vivo in the brains of mice in which a gene from jellyfish that encodes a green fluorescent protein is inserted into the mouse genome in a manner so that only glutamatergic neurons become fluorescent. Using confocal microscopy, investigators have been able to demonstrate in real time that persistent activation of NMDARs results in spine maturation from long and skinny to fat and stubby and even the elaboration of new spines. This transformation is mediated in part by the influx of calcium through the activated NMDARs. Such rapid structural changes in the synaptic spine reflect the fact that protein synthesis (i.e., translation) has recently been demonstrated to occur within individual dendritic spines.

Fragile X mental retardation protein (FMRP), which is deficient in individuals with fragile X syndrome, is synthesized locally within the spine during times of NMDAR activation and also plays a role in transporting specific mRNAs to the dendritic spine for translation. Notably, mice in which the *FMRP* gene has been inactivated through a null mutation, as well as patients with fragile X syndrome, have fewer dendritic spines, the preponderance of which have an immature morphology. Loss of FMRP exaggerates responses of mGluR5, which stimulates dendritic protein synthesis. Treatment with an mGluR5 antagonist reverses the fragile-X–like phenotype in mice with the FMRP gene inactivated; however, this strategy failed in clinical trials.

Excitotoxicity

In the early 1970s, it was shown that the systemic administration of large amounts of monosodium glutamate to immature animals causes the degeneration of neurons in brain regions where the BBB was deficient such as the arcuate nucleus of the hypothalamus. Subsequent studies showed that the direct injection of ionotropic glutamate receptor agonists such as KA, ibotenic acid, and NMDA caused a pattern of neuronal degeneration that affected neurons with their cell bodies in proximity to the injection site but spared axons passing through the area arising from distant neurons as well as glia. Persistent and overwhelming activation of AMPA/kainate receptors and NMDA receptors causes a tremendous influx of Na^+ and Ca^{2+} and a secondary influx of H_2O. The resulting acute cellular edema causes *necrotic* cell death. At sites more distant from the injection, the persistent elevation of intracellular Ca^{2+} disrupts the mitochondria, which release cytochrome C and activate caspases, resulting in programmed cell death (*apoptosis*).

The neuronal degeneration occurring after ischemic stroke is the result of excitotoxicity. Local hypoxia due to ischemia results in a cessation of

ATP production, causing the collapse of the ATP-driven sodium gradient across the neuronal and astroglial membranes. Consequently, the direction of the sodium-dependent glutamate transporters is reversed, causing a massive release of glutamate. Although several drugs that block the ionotropic glutamate receptors or the downstream events caused by their overstimulation have proved effective in reducing the amount of neuronal damage in animal models of stroke, no such agent has yet proved effective in clinical trials.

Excitotoxicity has also been implicated as the proximate cause of neuronal degeneration in Alzheimer disease. Most evidence points to the toxic consequences of aggregates of β-amyloid, especially β-amyloid$_{1–42}$. The β-amyloid fibrils depolarize neurons, resulting in loss of the Mg^{2+} block and enhanced NMDA receptor sensitivity to glutamate. The fibrils also impair glutamate transport into astrocytes, thereby increasing the extracellular concentration of glutamate. β-Amyloid directly promotes oxidative stress through inflammation that further contributes to neuronal vulnerability to glutamate. Finally, the proliferation of inflammatory (“reactive”) astrocytes further promotes excitotoxicity by releasing the obligatory NMDAR coagonist, D-serine, rendering neurons vulnerable to elevated extracellular glutamate. Thus, several mechanisms contribute to neuronal vulnerability to NMDA-receptor–mediated excitotoxicity in Alzheimer disease. Memantine (Namenda), an FDA-approved treatment for mild to moderate Alzheimer disease, is a weak noncompetitive inhibitor of NMDA receptors. It reduces the tonic sensitivity of NMDA receptors to excitotoxicity but does not interfere with “phasic” neurotransmission, thereby attenuating neuronal degeneration in Alzheimer disease.

Inhibitory Amino Acids: GABA

GABA is the major inhibitory neurotransmitter in the brain, where it is broadly distributed and occurs in millimolar concentrations. In view of its physiologic effects and distributions, it is not surprising that the dysfunction of GABAergic neurotransmission has been implicated in a wide range of neuropsychiatric disorders, including anxiety disorders, schizophrenia, alcohol dependence, and seizure disorders. Chemically, GABA differs from glutamic acid, the major excitatory neurotransmitter, simply by the removal of a single carboxy group from the latter.

GABA is synthesized from glutamic acid by GAD, which catalyzes the removal of the α-carboxyl group. In the CNS, the expression of GAD appears to be restricted to GABAergic neurons, although in the periphery, it is expressed in pancreatic islet cells. Two distinct but related genes encode GAD. GAD65 is localized to nerve terminals, where it is responsible for synthesizing GABA that is concentrated in the synaptic vesicles. Consistent with its role in fast inhibitory neurotransmission, mice homozygous for a null mutation of GAD65 have an elevated risk for seizures. GAD67 appears to be the primary source of neuronal GABA because mice homozygous for a null mutation of GAD67 die at birth, have a cleft pallet, and exhibit major reductions in brain GABA.

GABA is catabolized by GABA transaminase (GABA-T) yielding succinic semialdehyde. Transamination generally occurs when the parent compound, α-ketoglutarate, is present to receive the amino group, thereby regenerating glutamic acid. Succinic semialdehyde is oxidized by succinic semialdehyde dehydrogenase (SSADH) into succinic acid, which reenters the Krebs cycle. GABA-T is a cell surface, membrane-bound enzyme that is oriented toward the extracellular compartment and is expressed by both neurons and glia. As would be anticipated, drugs that inhibit the catabolism of GABA have anticonvulsant properties. One of the mechanisms of action of valproic acid (VPA) is the competitive inhibition of GABA-T. γ-Vinyl-GABA is a suicide substrate inhibitor of GABA-T that is used as an anticonvulsant in Europe but can impair vision by potentiating retinal GABAergic neurotransmission (vigabatrin; Sabril).

The synaptic action of GABA is also terminated by high-affinity transport back into the presynaptic terminal as well as into astrocytes. Four genetically distinct GABA high-affinity transporters have been identified with differing kinetic and pharmacologic characteristics. They all share homology with the other neurotransmitter transporters with the characteristic of 12 membrane-spanning domains. The active transport is driven by the sodium gradient so that upon depolarization, transportation of GABA out of the neuron is favored. GABA transported into astrocytes is catabolized by GABA-T and ultimately converted to glutamic acid and then to glutamine, which is transported back into the presynaptic terminal for GABA synthesis. Tiagabine (Gabitril) is a potent GABA transport inhibitor that is used to treat epilepsy.

Anatomy of GABAergic Systems

In the corticolimbic regions of the brain, GABA is localized to the intrinsic (i.e., local circuit) neurons. In the columnar organization of the cerebral cortex, the GABAergic neurons provide the outer boundaries of the columns with inwardly directed axons. While the GABAergic interneurons comprise a minority of cortical neurons (15 to 25 percent), they exert a profound degree of inhibition of the activity of the glutamatergic pyramidal cells. The remarkable efficacy of inhibition reflects two neuroanatomical features of GABAergic synapses: they are concentrated on the shafts of synaptic spines to counteract glutamatergic depolarization and on the neuronal cell body and proximal axon to restrict the generation of action potentials. In the cerebral cortex, the GABAergic interneurons are the primary site of colocalization of neuropeptides. These include cholecystokinin (CCK), dynorphin, neuropeptide Y (NPY), somatostatin, substance P, and vasoactive intestinal peptide (VIP).

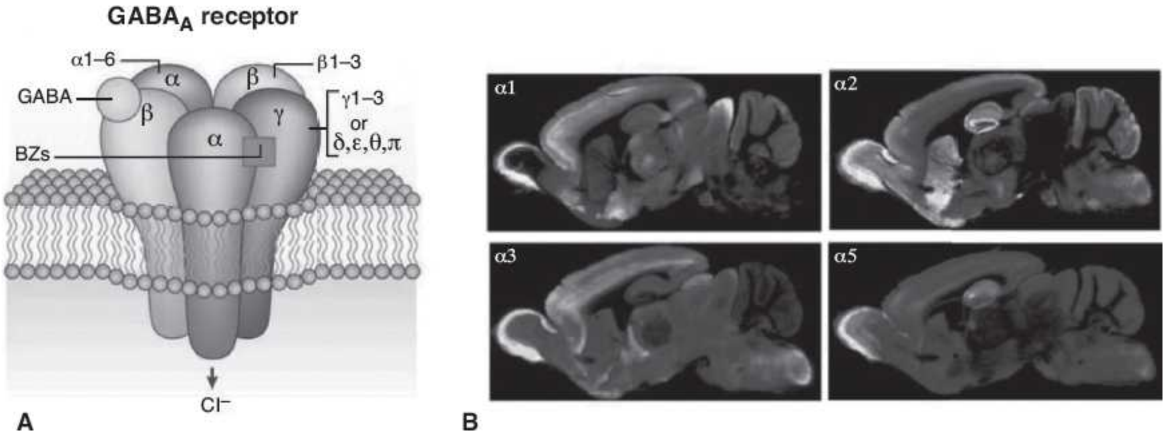

FIGURE 1.6–4. Schematic representation of the GABA$_A$ receptor. **A:** GABA$_A$ receptors are defined by a complex of five subunits, creating a pore through which Cl⁻ ions flow down their concentration gradient to mediate fast synaptic inhibition. The most common subunit assembly is depicted here with 2α, 2β, and 1γ subunits. The receptor is also defined by the presence of GABA and benzodiazepine binding sites, which are formed by α and β or α and γ subunits, respectively. (A, reprinted by permission from Springer Nature: Jacob TC, Moss SJ, Jurd R. GABAA receptor trafficking and its role in the dynamic modulation of neuronal inhibition. *Nat Rev Neurosci* 2008; 9:331–343. Copyright © 2008.) **B:** The four α subunits, which bind classical benzodiazepines (α1, α2, α3, and α5), have distinct expression patterns within the brain, as demonstrated by these autoradiographs of benzodiazepine binding in the rat brain. (B, reproduced by permission of ASPET from Mohler H, Fritschy JM, Rudolph U. A new benzodiazepine pharmacology. *J Pharm and Exp Ther*. 2002;300(1):2–8.)

In the striatum, GABAergic neurons project directly to the substantia nigra pars reticulata, which regulates dopaminergic neuronal activity. In addition, there are striatal GABAergic neurons that project to the globus pallidus to synapse on pallidal–subthalamic GABAergic neurons that regulate the excitatory output from the subthalamic nucleus as well as reciprocal GABAergic projections from the substantia nigra/ventral tegmental area to the striatum. In the cerebellum, the GABAergic Purkinje cells are its main efferent system.

GABA$_A$ Receptors

GABA$_A$ receptors are distributed throughout the brain. The GABA$_A$ receptor complex, when activated, mediates an increase in membrane conductance with an equilibrium potential near the resting membrane potential of −70 mV (Fig. 1.6–4). In the mature neuron, this typically results in an influx of Cl^-, causing membrane hyperpolarization. Hyperpolarization is inhibitory because it increases the threshold for generating an action potential. In immature neurons, which have unusually high levels of intracellular Cl^-, activating the GABA$_A$ receptor can counterintuitively cause depolarization. For this reason, anticonvulsants that act by enhancing GABA$_A$ receptor activity may exacerbate seizures in the neonatal period.

The GABA$_A$ receptor subunits exhibit sequence homology to a larger family of ligand-gated channels, including the nicotinic acetylcholine receptor and the glycine receptor. At least 19 distinct but closely related genes have been identified that encode GABA$_A$ receptor subunits. Each subunit contains four α-helical membrane-spanning domains, the sequences of which are highly conserved among the subunits. The receptor complex is a heteropentamer. The ligand-binding site is formed at the interface between the α- and the β-subunits. The subunit composition affects the biophysical and pharmacologic characteristics of the receptor. Different subunit-containing GABA$_A$ receptor complexes are expressed at different stages of development as well as in different regions of the brain.

The pharmacology of GABA$_A$ receptors is particularly rich. A component of the extracts of the psychoactive mushroom *Amanita muscaria* is muscimol, which is a direct agonist at the GABA$_A$ receptor. The prototypical GABA$_A$ antagonist is bicuculline, which acts by decreasing the frequency and duration of channel opening and has proconvulsant effects. Picrotoxin, another proconvulsant, acts by blocking the chloride channel.

Reminiscent of the NMDA receptor, the $GABA_A$ receptor complex is noteworthy for multiple allosteric modulatory interactions. These include benzodiazepines, barbiturates, general anesthetics, ethanol, and neurosteroids. Benzodiazepines bind to a distinct site in the $GABA_A$ receptor complex and allosterically increase the frequency of channel opening in response to GABA. Therefore, benzodiazepines do not directly activate the receptor, but they enhance the phasic responses to synaptically released GABA. This indirect mechanism of action and the localization of benzodiazepine-sensitive receptors account for the lower risk of respiratory suppression for benzodiazepines as compared to that of barbiturates. The benzodiazepine site is allosterically linked to the binding sites of other modulatory ligands, which can contribute to toxic interactions such as with ethanol. Notably, antagonists and inverse agonists for the benzodiazepine receptor have been developed that demonstrate anxiogenic and procognitive effects.

Barbiturates such as phenobarbital and pentobarbital are noted for their sedative and anticonvulsant activities. Barbiturates allosterically increase the affinities of the binding sites for GABA and benzodiazepines at concentrations that are pharmacologically relevant. Barbiturates also affect channel dynamics by markedly increasing the long open state and reducing the short open state, thereby increasing Cl^- conductance, and inhibition. This direct mechanism of action at the $GABA_A$ receptors accounts for fatal respiratory suppression when taken in excessive amounts.

Chemically modified analogs of progesterone and corticosterone have been shown in behavioral studies to have sedative and anxiolytic effects through their interaction with the $GABA_A$ receptor complex. They share features with barbiturates, although the hydrophobic neurosteroids accumulate in high concentrations within the cell membrane at their recognition site on the $GABA_A$ receptor. They allosterically enhance agonist binding to the receptor and increase the duration of Cl^- channel opening. A variety of behavioral effects associated with steroid administration or fluctuation of endogenous steroids and sex-specific effects of GABAergic drugs have been linked to the action of endogenous neurosteroids. Recently, the FDA approved brexanolone (allopregnanolone; Zulresso) as a treatment for women with postpartum depression.

Regarding $GABA_A$ receptor antagonists, picrotoxin, like the barbiturates, alters channel dynamics but in the opposite direction by reducing long open states and favoring the briefest open state. The proconvulsant pentylenetetrazol also acts by reducing Cl^- channel permeability. Penicillin, which at high concentrations is proconvulsant, binds to the positively charged residues in the channel, thereby occluding it. As a general class, anesthetics including barbiturates, steroids, and volatile anesthetics increase chloride conductance, thereby inhibiting neurotransmissions. Amino acids in the membrane-spanning domain of the GABA receptor subunits confer sensitivity to anesthetics. The precise mechanism whereby ethanol enhances $GABA_A$ receptor function remains unclear due to inconsistent results, suggesting that subunit composition may be important. However, recent studies suggest that ethanol increases the response of the tonic GABA-activated currents, which contain the δ-subunit and exhibit remarkably high affinity for GABA.

Recombinant DNA strategies exploiting site-directed mutagenesis have permitted the identification of sites on the specific subunits that mediate the pharmacologic action of drugs such as the benzodiazepines. Mutational removal of the binding ability for benzodiazepines has established that the α_1-subunit plays a major role in the sedative and amnestic effects of benzodiazepines, the α_2-subunit confers the anxiolytic effects, the α_3-subunit contributes to fear reduction and the α_5-subunit to muscle relaxation. Developing drugs that are subunit selective for benzodiazepines, such as the sedative zolpidem (Ambien), reduce off-target side effects associated with nonspecific benzodiazepines.

$GABA_B$ Receptors

The $GABA_B$ receptors are distinguished pharmacologically from $GABA_A$ receptors by the fact that they are insensitive to the canonical $GABA_A$ receptor antagonist bicuculline and that they are potently activated by baclofen (β-(4-chlorophenyl)-γ-aminobutyric acid; Lioresal), which is inactive at $GABA_A$ receptors. They are members of the G-protein–coupled superfamily of receptors but are highly unusual as they are made of a dimer of two seven transmembrane-spanning subunits. $GABA_B$ receptors are widely distributed throughout the nervous system and are localized both pre- and postsynaptically. The postsynaptic $GABA_B$ receptors cause a long-lasting hyperpolarization by activating potassium channels. The presynaptic $GABA_B$ receptors act as auto- and heteroreceptors to inhibit neurotransmitter release.

All $GABA_B$ receptors in the vertebrate brain are the sole products of the $GABA_B$(1) and $GABA_B$(2) genes (Fig. 1.6–5). Pharmacologic heterogeneity among $GABA_B$ receptors reflects different isoforms resulting from splice variants. The most common variants that are conserved across species are $GABA_B$(1A) and $GABA_B$(1B). They exhibit different regional distributions in the brain where, for example, $GABA_B$(1A) transcripts are confined to the granule cell layer, whereas $GABA_B$(1B) transcripts are expressed primarily in Purkinje cells. All $GABA_B$ agonists and competitive antagonists bind to the extracellular domain of the $GABA_B$(1) subunit. The $GABA_B$(2) subunit also has a large extracellular domain, which may be the site for allosteric modulation of the $GABA_B$ receptor. Interestingly, there is a binding site for Ca^{2+} in the ligand-binding pocket of the $GABA_B$(1) subunit that increases the affinity for GABA. This Ca^{2+} binding site is typically saturated at physiologic concentrations of Ca^{2+}.

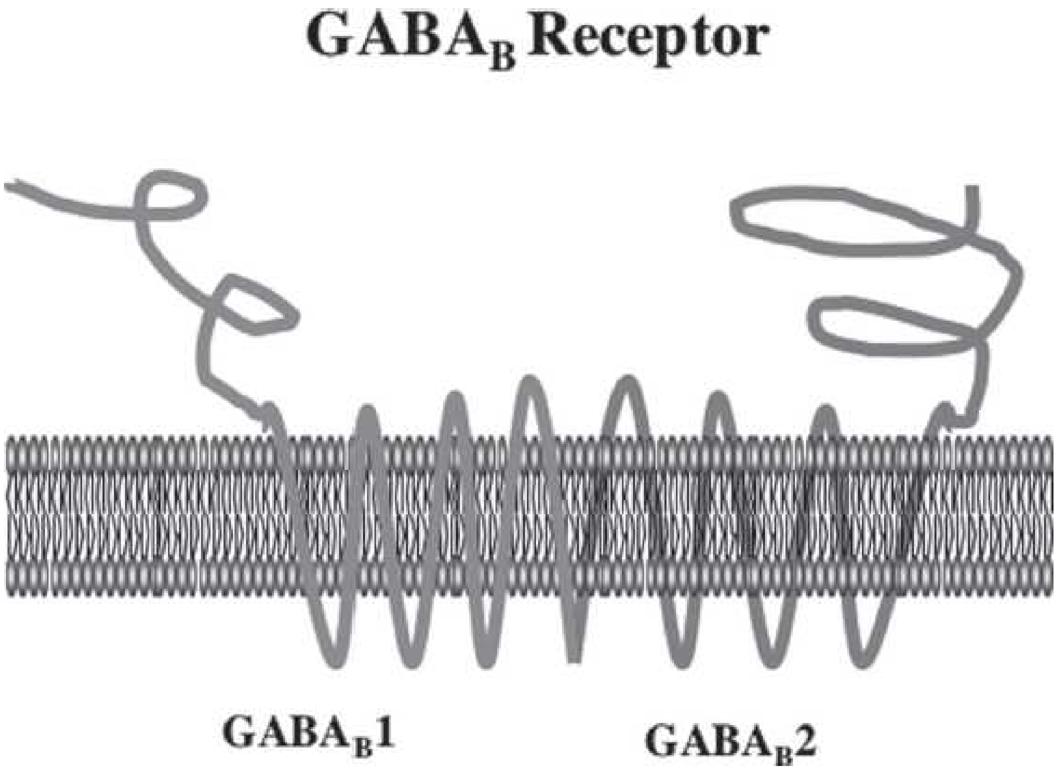

FIGURE 1.6–5. Schematic representation of the GABA$_B$ receptor. This G-protein–coupled receptor (GPCR) is a heterodimer of two seven-membrane-spanning GPCRs. The extensive extracellular domains have binding sites for GABA and allosteric modulators.

γ-Hydroxybutyrate (GHB), which is approved for the treatment of narcolepsy, has been misused as a “date rape” drug because it rapidly induces deep sleep. Although high-affinity binding sites for GHB have been described in the brain, the sedative and hypnotic effects of exogenous GHB can be prevented by GABA$_B$ receptor antagonists. Furthermore, administration of GHB to GABA$_B$−/− mice resulted in no behavioral effects of GHB, indicating that GABA$_B$ receptors are its likely site of action and not the so-called GHB receptors.

Glycine as a Neurotransmitter

Glycine is an inhibitory neurotransmitter primarily in the brainstem and spinal cord, although the expression of glycine receptor subunits in the thalamus, cortex, and hippocampus suggests a broader role. Glycine is a nonessential amino acid that is synthesized in the brain from L-serine by serine hydroxymethyltransferase. Glycine is concentrated within synaptic vesicles by H$^+$-dependent vesicular inhibitory amino acid transporter

(VIAAT or VGAT), which also transports GABA. Termination of the synaptic action of glycine is through reuptake into the presynaptic terminal by the glycine transporter II (GlyT2), which is quite distinct from the GlyT1 that is expressed in astrocytes and modulates NMDA receptor function.

The inhibitory effects of glycine are mediated by a ligand-gated chloride channel, which can also respond to β-alanine, taurine, L-alanine, L-serine, and proline but not to GABA. The canonical antagonist for the glycine receptor is the plant alkaloid strychnine. The receptor was first identified through the specific binding of [^{3}H]strychnine to synaptic membranes. [^{3}H]Glycine itself binds to two sites: One that is displaceable by strychnine and represents the glycine A receptor, and a second that is insensitive to strychnine and is designated the glycine B receptor, representing the GMS on the NMDA receptor.

The glycine A receptor is a macromolecular complex of approximately 250 kDa that is comprised of five subunits surrounding a central pore. There are two subunits with a high degree of homology: The 48-kDa α-subunit and the 58-kDa β-subunit. The subunits display a structural similarity to other members of this ion channel family with four hydrophobic domains that span the lipid bilayer in α-helices. The binding site for both glycine and strychnine is located on the α-subunit. There are four genes that encode the α-subunits, but only one gene encodes for the β-subunit. Interestingly, the β-subunit is expressed fairly abundantly in rostral brain regions that exhibit no [^{3}H]strychnine binding or α-subunit expression.

Hereditary hyperekplexia is a disorder due to mutations in genes encoding components of the glycinergic synapse. It is characterized by stiffness and an excessive startle reflex in infancy that subsides with maturation. Mutations causing hyperekplexia have been described in the α-subunit (GLRA1) and in the β-subunit (GLRB) of the glycine receptor but also in GlyT2 (SLC6A5). Treatment with oral clonazepam is usually an effective therapy.

Neuropsychiatric Implications of Amino Acid Transmitters

Schizophrenia Evidence accumulating from postmortem, pharmacologic, and genetic studies is shifting the focus of the pathophysiology of schizophrenia from dopamine to glutamate and GABA. Indeed, after the use of dopamine D_2 receptor antagonists of varying structures and side-effect profiles as the primary treatment of schizophrenia for the last 60 years, over two-thirds of the treated patients still remain substantially disabled. Postmortem studies conducted over 40 years ago indicated a reduction in the activity of GAD in the cerebral cortex in patients with schizophrenia as compared to suitable controls. With the advent of immunocytochemistry and gene expression techniques, it has been possible to more precisely define the GABAergic deficit in schizophrenia. The parvalbumin-positive GABAergic interneurons in the intermediate layers of the cortex bear the brunt of the pathology, which includes reduced expression of GAD67, parvalbumin, and the GABA transporter (GAT). The finding that $GABA_A$ receptors are upregulated, as measured by autoradiography or with antibodies, supports the theory that these changes reflect the hypofunction of the presynaptic GABAergic neurons. These particular GABAergic interneurons, which include the chandelier cells, play an important role in negative feedback inhibition to the pyramidal cells in the cortex.

The theory that hypofunction of NMDA receptors is an etiologic factor in schizophrenia initially arose from the observation that PCP and related dissociative anesthetics that block NMDA receptors produce a syndrome that can be indistinguishable from schizophrenia. Dissociative anesthetics are so named because they prevent the acquisition of new memories without diminishing the apparent level of consciousness. In fact, under laboratory conditions, infusion of low doses of ketamine, another dissociative anesthetic, can produce the positive symptoms, negative symptoms, and specific cognitive deficits associated with schizophrenia in clear consciousness. Subsequent studies indicated that low-dose ketamine can also cause enhanced amphetamine-induced subcortical dopamine release, as is observed in schizophrenia, as well as abnormal cortical event-related potentials (ERPs) and disruption of prepulse inhibition in experimental animals.

Early-gene studies in schizophrenia were limited by the biased selection of “candidate” genes and small numbers of subjects, resulting in contradictory results. In 2014, a large GWAS comparing nearly 40,000 subjects with schizophrenia to a hundred thousand controls identified 108 sites (single-nucleotide polymorphism [SNP]) on the genome, which attained the stringent level of genome-wide significance (5×10^{-8}) required by the multiple comparisons inherent to GWAS (Fig. 1.6–6). Recently, this was expanded to 70,000 schizophrenics and 236,000 controls. A clear pattern has emerged. Most of the risk-associated SNPs do not localize to coding regions of the genome but presumably act by altering gene expression or RNA splicing. Notably, none contributed more than 5 percent to the ascribable risk for schizophrenia, indicating that schizophrenia was a disorder of complex genetics like diabetes with multiple risk genes of modest effects interacting with the environment to produce the phenotype.

Three major categories of risk genes were identified: major histocompatibility complex (MHC) genes, genes related to neurotransmission, and genes involved in development. Nearly a dozen risk genes are within two degrees of separation from the NMDAR, including *SRR* (the gene-encoding serine racemase, SR), whereas a dozen more are involved in glutamate signaling such as voltage-dependent Ca^{2+} channels. Genetic inactivation of SR in mice results in a phenotype that shares many features with the neurochemical alterations, neuronal structural deficits such as loss of synaptic spines, electrophysiologic abnormalities such as impaired EEG ERPs and cognitive features of schizophrenia as well as increased striatal dopamine release consistent with psychosis. Although the GWAS in bipolar disorder are not as well powered as those in schizophrenia, risk genes shared between schizophrenia and bipolar disorder have been identified that implicate glutamatergic neurotransmission, including the voltage-gated calcium channel CACNA1C, and NAMDAR subunit, GRIN2A, in the pathophysiology of bipolar disorder.

Several clinical trials have been carried out with agonists at the GMS on the NMDA receptor in patients receiving concurrent treatment with antipsychotic medications. The hypothesis being tested was that enhancing NMDA receptor function would reduce negative symptoms and improve cognition, aspects of the disorder unaffected by antipsychotic drugs. A meta-analysis of the results of these placebo-controlled trials of agonists at the NMDA receptor GMS revealed highly significant reductions in depression, negative symptoms, cognitive deficits, and even positive symptoms. However, other meta-analyses that used different selection criteria have reported less compelling results, indicating that these weak agonists at the GMS are unlikely to be clinically relevant but rather contribute to “proof of principle.”

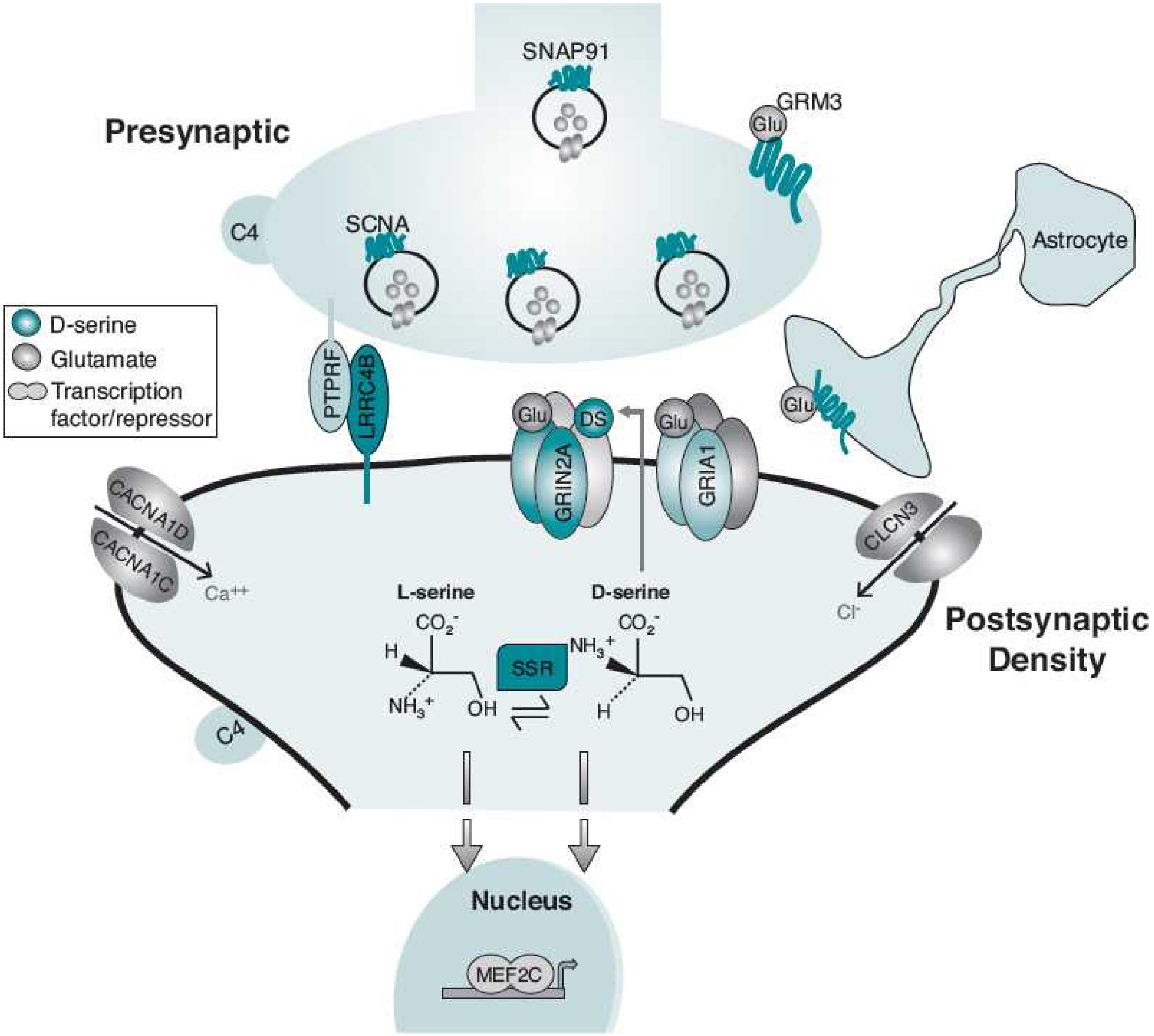
SNAP91
GRM3
Glu
Presynaptic
C4
SCNA
Astrocyte
D-serine
Glutamate
Transcription factor/repressor
PTPRF
LRRC4B
Glu
DS
Glu
Glu
GRIN2A
GRIA1
CACNA1D
CACNA1C
Ca++
CLCN3
L-serine
D-serine
CO2-
H
NH3+
OH
SSR
NH3+
CO2-
H
OH
Cl-
Postsynaptic Density
C4
Nucleus
MEF2C

FIGURE 1.6–6. Genetic risk for schizophrenia converges on the glutamatergic synapse. Genome-wide association studies (GWAS) of over nearly two hundred thousand subjects with schizophrenia and controls have identified over a hundred single nucleotide polymorphisms that exhibit genome-wide significant association (5×10^{-8}) with the risk for schizophrenia. Associated risk genes indicate a prominent role for the glutamatergic synapse in schizophrenia pathology. The selected schizophrenia risk genes are shown here demonstrate the relevance to all aspects of the glutamatergic synapse function to schizophrenia. At the presynaptic terminal, genes including *SNCA* (*synuclein alpha*) and *SNAP91* (*synaptosome associated protein 91*) regulate synaptic vesicle trafficking. Trans-synaptic cell adhesion molecules such as *PTPRF* (*Protein Tyrosine Phosphatase Receptor Type F*) and *LRRC4B* (*Leucine-Rich Repeat-Containing 4B*) regulate synaptic signaling, establishment, and remodeling. Multiple nonreceptor ion channel subunit genes are associated with schizophrenia, including voltage-dependent calcium channel subunits *CACNA1C*, *CACNA1D*, and *CACNA1G* and chloride ion channel subunit *CLCN3*. Similarly, subunits of glutamate receptors are themselves encoded by schizophrenia risk genes, including NMDA receptor

subunit *GRIN2A*, AMPA receptor subunits *GRIA1* and *GRIA3*, and metabotropic glutamate receptors *GRM1* and *GRM3*. The serine racemase enzyme encoded by *SRR* converts L-serine to D-serine, the primary coagonist of NMDA receptors. Complement component C4, encoded by the *C4A* and *C4B* genes within the major histocompatibility complex locus, contributes to excessive pruning of excitatory synapses in late adolescence and the decreased density of dendritic spines observed in schizophrenia. Finally, multiple nuclear factors, including the transcription factor *MEF2C*, regulate gene expression–modulated glutamatergic neurotransmission via the NMDA receptor. (Reprinted with permission from Coyle JT, Ruzicka WB, Balu DT. Fifty years of research on schizophrenia: The ascendance of the glutamatergic synapse. *Am J Psychiatry* 2020;177(12):1119–1128. Copyright © 2020. American Psychiatric Association. All Rights Reserved.)

Recent findings have provided a link between the GABAergic neuropathology and NMDAR hypofunction in schizophrenia. Chronic treatment of rats with NMDAR antagonists causes a downregulation of GAD67, parvalbumin, and GAT. The most sensitive subpopulation of GABAergic neurons is the rapidly firing interneurons that provide the perisomatic and axon hillock innervation of the pyramidal cells. Their NMDAR appear to be much more sensitive to antagonists than those less active GABAergic neurons and pyramidal cells because a critical feature of NMDAR function is membrane depolarization. Except for the genes encoding *GABBR1*, *GABBR2,* and *CALB2*, current GWAS results have not found other genes related to GABAergic function as significant risk genes, suggesting that cortical GABAergic deficits in schizophrenia may be a downstream consequence of some more proximal genetic defects affecting GABAergic neuronal function such as reduced NMDAR function.

The subtly reduced GABAergic inhibition results in a disinhibition of the glutamatergic pyramidal output. This degradation of the cortical inhibitory feedback likely accounts for the cognitive deficits and negative symptoms in schizophrenia, and the disinhibited pyramidal neuron output also results in elevated subcortical dopamine release causing psychosis. These findings that mimic the schizophrenia phenotype are replicated by silencing the risk gene for schizophrenia, *SRR*, which eliminates D-serine, the dominant NMDAR coagonist in the forebrain. Thus, psychosis appears to be a downstream consequence of the disruption in critical glutamatergic–GABAergic synaptic function in the cerebral cortex.

Anxiety and Depression

GABAergic dysfunction has been associated with anxiety disorders, especially panic disorder, as well as with major depressive disorder (MDD). Clinically, there is considerable comorbidity between anxiety and affective disorders. MRS reveals significant reductions in GABA in the prefrontal cortex, cingulate cortex, and occipital cortex in MDD. The reduced levels of GABA in the occipital cortex in episodes of MDD normalize with effective treatment with selective serotonin reuptake inhibitor (SSRI) or with electroconvulsive therapy (ECT). Postmortem studies have revealed upregulation of the $GABA_A$ receptor α_1- and β_3-subunits in the cerebral cortices of depressed patients who committed suicide, consistent with a reduction in GABAergic neurotransmission. Decreased levels of the $GABA_A$ receptor modulators, three α-reduced neuroactive steroids, have been found both in plasma and in CSF in MDD. Effective treatment with SSRI increases the neurosteroid levels. MRS has disclosed significant reductions in GABA levels in the anterior cingulate and in the basal ganglia of medicated patients with panic disorder. PET scanning reveals a highly selective reduction in benzodiazepine receptor sites bilaterally in the insular cortex in panic disorder. A genome-wide screen has shown significant linkage at 15q in a region containing $GABA_A$ receptor subunit genes and panic disorder.

Glutamatergic dysfunction has also been implicated in depression. NMDAR antagonists have antidepressant effects in several animal models of depression, including forced swim, tail suspension, and learned helplessness. A single injection of ketamine provides protection from the induction of behavioral despair in rats for up to 10 days. Chronic treatment with antidepressants alters the expression of NMDAR subunits and decreases GMS binding. Several placebo-controlled clinical trials have shown that a single dose of ketamine can produce a rapid, substantial, and persistent reduction in symptoms in patients with MDD, especially those who have not responded to SSRI/SNRI treatment. Furthermore, ketamine administration causes a rapid and persistent reduction in suicidal thoughts. In 2019, Jansen received FDA approval for the use of the S-isomer of ketamine (Esketamine) administered by nasal insufflation for treatment of depression and suicidal ideation. There is debate about the precise mechanism of rapid antidepressant action of ketamine, but most converge on its transient blockade of NMDARs resulting in a rapid increase in brain-derived neurotrophic factor (BDNF) release and strengthening of glutamatergic neurotransmission, especially in reward pathways.

Alcoholism

Ethanol at concentrations associated with intoxication has a dual action of enhancing GABAergic receptor function and attenuating NMDAR function. The GABA receptor effects may be associated with the anxiolytic effects of ethanol. Persistent abuse and dependency on ethanol result in a downregulation of $\mathrm{GABA_A}$ receptors and an upregulation of NMDARs such that acute discontinuation of ethanol results in a hyperexcitable state characterized by delirium tremens. Furthermore, supersensitive NMDARs in the context of thiamine deficiency may contribute to the excitotoxic neuron degeneration of Wernicke–Korsakoff syndrome.

Acamprosate (Campral) is a derivative of homotaurine that was developed as an agent to reduce alcohol consumption, craving, and relapse in alcohol use disorder (AUD). It exhibits moderate efficacy in clinical trials. Because of taurine’s resemblance to GABA, it was thought that acamprosate acted via $\mathrm{GABA_A}$ receptors, but electrophysiologic studies found little evidence to support this hypothesis. Subsequent studies demonstrated that it inhibited NMDAR responses in cortical slices and recombinant NMDARs. The precise mechanism whereby acamprosate alters NMDAR function, however, remains unclear. Recently, acamprosate has been combined with the opiate receptor antagonist, naltrexone (Vivitrol), to better prevent relapse in AUD.

Fetal alcohol syndrome (FAS) is the most common preventable cause of intellectual disability. Convincing evidence has been developed that the microcephaly associated with fetal alcohol exposure results from its inhibition of NMDAR function, resulting in widespread neuronal apoptosis in the immature cortex because NMDAR activation is essential for immature neuronal survival and differentiation.

CONCLUSION

Over the last 70 years, very few FDA-approved psychiatric drugs that deviated from the fundamental mechanisms of reducing dopamine function for psychosis, inhibiting biogenic amine uptake for depression and anxiety disorders, and enhancing $\mathrm{GABA_A}$ receptor function for anxiolysis. However, recently some novel effective treatments have been developed that work through altering amino acid neurotransmitter function, including acamprosate for AUD, esketamine, an NMDAR antagonist, for treatment-resistant depression, and brexanolone a $GABA_A$ receptor allosteric modulator, for postpartum depression.

The expansion of genetic studies to permit sufficiently powered GWAS are now disclosing risk genes that involve pathways for psychiatric disorders linked to glutamatergic and GABAergic neurotransmission. The expanding GWAS findings in serious mental illness will provide targets for rational drug development, deviating from the historical reliance on serendipity, an inefficient method for treatment development. These advances will present real challenges for students and psychiatric clinicians as more and more drugs that address symptom domains linked to risk gene pathways are introduced and erode the categorical approach to diagnosis and treatment. Given their preeminent role in brain neurotransmission, glutamate and GABA pathways will likely play a prominent role in these developments.

FURTHER READINGS

Anderson KM, Collins MA, Kong R, et al. Convergent molecular, cellular, and cortical neuroimaging signatures of major depressive disorder. *Proc Natl Acad Sci U S A*. 2020;117:25138–25149.

Benham RS, Engin E, Rudolph U. Diversity of neuronal inhibition: a path to novel treatments for neuropsychiatric disorders. *JAMA Psychiatry*. 2014;71:91–93.

Berg EL, Copping NA, Rivera JK, et al. Developmental social communication deficits in the Shank3 rat model of Phelan-McDermid syndrome and autism spectrum disorder. *Autism Res*. 2018;11:587–601.

Lüscher B, Möhler H. Brexanolone, a neurosteroid antidepressant, vindicates the GABAergic deficit hypothesis of depression and may foster resilience. *F1000 Res*. 2019;8:F1000Faculty Rev-751

Bode A, Lynch JW. The impact of human hyperekplexia mutations on glycine receptor structure and function. *Mol Brain*. 2014;7:2.

Cheng HY, McGuinness LA, Elbers RG, et al. Treatment interventions to maintain abstinence from alcohol in primary care: systematic review and network meta-analysis. *BMJ*. 2020;371:m3934.

Cooke SF, Bear MF. How the mechanisms of long-term synaptic potentiation and depression serve experience-dependent plasticity in primary visual cortex. *Philos Trans R Soc Lond B Biol Sci*. 2013;369:20130284.

Corssen G, Domino EF. Dissociative anesthesia: further pharmacologic studies and first clinical experience with the phencyclidine derivative CI-581. *Anesth Analg*. 1966;45(1):29–40.

Coyle JT, Ruzicka WB, Balu DT. Fifty years of research on schizophrenia: the ascendance of the glutamatergic synapse. *Am J Psychiatry*. 2020;177:1119–1128.

Dienel SJ, Lewis DA. Alterations in cortical interneurons and cognitive function in schizophrenia. *Neurobiol Dis*. 2019;131:104208.

Dutertre S, Becker CM, Betz H. Inhibitory glycine receptors: an update. *J Biol Chem*. 2012;287:40216–40223.

Gould TD, Zarate CA Jr, Thompson SM. Molecular pharmacology and neurobiology of rapid-acting antidepressants. *Annu Rev Pharmacol Toxicol*. 2019;59:213–236.

Granger AJ, Nicoll RA. Expression mechanisms underlying long-term potentiation: a postsynaptic view, 10 years on. *Philos Trans R Soc Lond B Biol Sci*. 2013;369:2013–2036.

Groc L, Choquet D. Linking glutamate receptor movements and synapse function. *Science*. 2020;368(6496):eaay4631.

Jensen AA, Fahlke C, Bjørn-Yoshimoto WE, Bunch L. Excitatory amino acid transporters: recent insights into molecular mechanisms, novel modes of modulation and new therapeutic possibilities. *Curr Opin Pharmacol*. 2015;20:116–123.

Lebois LAM, Seligowski AV, Wolff JD, Hill SB, Ressler KJ. Augmentation of extinction and inhibitory learning in anxiety and trauma-related disorders. *Annu Rev Clin Psychol*. 2019;15:257–284.

Lisman JE, Coyle JT, Green RW, et al. Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. *Trends Neurosci*. 2008;31:234–242.

Lynch JW, Zhang Y, Talwar S, Estrada-Mondragon A. Glycine receptor drug discovery. *Adv Pharmacol*. 2017;79:225–253.

Michalon A, Bruns A, Risterucci C, et al. Chronic metabotropic glutamate receptor 5 inhibition corrects local alterations of brain activity and improves cognitive performance in fragile X mice. *Biol Psychiatry*. 2014;75:189–197.

Moghaddam B, Krystal JH. Capturing the angel in “angel dust”: twenty years of translational neuroscience studies of NMDA receptor antagonists in animals and humans. *Schizophr Bull*. 2012;38:942–949.

Paoletti P, Neyton J. NMDA receptor subunits: function and pharmacology. *Curr Opin Pharmacol*. 2007;7:39–47.

Pinar C, Fontaine CJ, Triviño-Paredes J, et al. Revisiting the flip side: long-term depression of synaptic efficacy in the hippocampus. *Neurosci Biobehav Rev*. 2017;80:394–413.

Price RB, Shungu DC, Mao X, et al. Amino acid neurotransmitters assessed by proton magnetic resonance spectroscopy: relationship to treatment resistance in major depressive disorder. *Biol Psychiatry*. 2009;65:792–800.

Proctor WR, Diao L, Freund RK, Browning MD, Wu PH. Synaptic GABAergic and glutamatergic mechanisms underlying alcohol sensitivity in mouse hippocampal neurons. *J Physiol*. 2006;575:145–159.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–427.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604:502–508

Senter RK, Ghoshal A, Walker AG, Xiang Z, Niswender CM, Conn PJ. The role of mGlu receptors in hippocampal plasticity deficits in neurological and psychiatric disorders: implications for allosteric modulators as novel therapeutic strategies. *Curr Neuropharmacol*. 2016;14:455–473.

Singewald N, Schmuckermair C, Whittle N, et al. Pharmacology of cognitive enhancers for exposure-based therapy of fear, anxiety and trauma-related disorders. *Pharmacol Ther*. 2015;149:150—190.

Swick TJ. Swick TJ. Treatment paradigms for cataplexy in narcolepsy: past, present, and future. *Nat Sci Sleep*. 2015;7:159–169.

Yang Y, Paspalas CD, Jin LE, Picciotto MR, Arnsten AF, Wang M. Nicotinic α7 receptors enhance NMDA cognitive circuits in dorsolateral prefrontal cortex. *Proc Natl Acad Sci U S A.* 2013;110(29):12078–12083. doi:10.1073/pnas.1307849110. Epub 2013 Jul 1.

Zorumski CF, Paul SM, Covey DF, Mennerick S. Neurosteroids as novel antidepressants and anxiolytics: GABA-A receptors and beyond. *Neurobiol Stress*. 2019;11:100196.

1.7 Biogenic Amine Neurotransmitters

JOHN D. ELSWORTH, PH.D.

The biogenic amine neurotransmitters comprise the catecholamines (dopamine, norepinephrine, and epinephrine) and serotonin (5-hydroxytryptamine, 5HT), histamine, and acetylcholine. The importance of these messengers to central nervous system (CNS) function is reflected in the critical roles these systems play in cognition, memory, motivation, motor function, and arousal, and their relevance to psychiatry is demonstrated by their involvement in the pathophysiology and pharmacologic treatment of multiple disorders.

These neurotransmitters have an unusual and functionally significant organization in the CNS. Their cell bodies are restricted to a relatively small number of nuclei in the midbrain, brain stem, basal forebrain, and hypothalamus, yet their axons project extensively, with some biogenic amine transmitter systems displaying a more complex topographical arrangement than others. This widely disbursed influence permits each of these neurotransmitters to modulate activity in diverse circuits throughout the CNS, sometimes in a coordinated fashion. This anatomical organization poses a major challenge to deciphering the precise molecular, cellular, and systems-level contributions these neurotransmitters exert in neuropsychiatric disorders. Advances in genomics have revealed insights into the complexity of these neurotransmitter systems, exemplified by the identification of a large number of genes that regulate monoaminergic neurotransmission, such as the enzymes, receptors, and transporters that mediate the synthesis, cellular actions, and cellular reuptake of these neurotransmitters, respectively. Human genetics studies have provided evidence of tantalizing links between allelic variants in specific monoamine-related genes and psychiatric disorders and personality traits, while the ability to modify gene function and cellular activity in experimental animals has clarified the roles of specific genes and neural pathways in mediating behavioral processes.

In this chapter, each of the biogenic amine transmitters is considered separately with the details of its “life cycle,” specifically the details of enzymes and transporters devoted to the synthesis, degradation, and transporters of the transmitter. These descriptions also incorporate transmitter-specific receptors since, ultimately, the effects of biogenic amine transmitters on CNS function and behavior depend on their interactions with receptor molecules. The binding of biogenic amines to these plasma membrane proteins initiates a series of intracellular events that modulate neuronal excitability. Unlike transporters, multiple receptor subtypes exist for each biogenic amine transmitter. Neurotransmitter receptors produce intracellular effects by one of two basic mechanisms: (1) by interactions with G proteins that couple receptors (GPCRs) to intracellular effector systems and (2) by providing channels through which ions flow when transmitters bind (ligand-gated ion channels). The neurotransmitter receptors covered in this chapter, with the exception of the serotonin 5-HT$_3$ receptor subtype and the cholinergic nicotinic receptors, all belong to the superfamily of GPCRs. However, within each neurotransmitter receptor group, the subtypes are heterogeneous with regard to the G proteins with which they interact and to the second messenger effects that they induce. Receptors are also diverse in their regional patterns of expression within the brain, their neurotransmitter binding affinities, and their synaptic localization. Whereas many receptor subtypes are located exclusively in postsynaptic membranes, others are located presynaptically, and some are both pre- and postsynaptically. Populations of receptors on the presynaptic terminal that respond to a transmitter released by that neuron are known as “*autoreceptors*,” which often act to inhibit neurotransmitter release in order to regulate the concentration of transmitter in the synapse (“local feedback inhibition”). When a receptor for a particular transmitter is located on terminals of a neuron expressing a different transmitter, it is termed a “*heteroreceptor,*” and its activation can either increase or decrease transmitter release.

Much effort has been focused on determining the location and function of individual neurotransmitter receptor subtypes. Limited availability of selective agonist and antagonist drugs hampers this effort, but recent advances in gene manipulation and targeted gene editing are advancing the fields of neuroscience and pharmacology considerably. Such studies in animals provide clues to receptor function and to the contributions of each receptor to the actions of nonspecific drugs. It is anticipated that the near future will witness the development of receptor subtype–selective compounds that achieve therapeutic effects by altering biogenic amine transmission in a more refined manner.

DOPAMINE AND NOREPINEPHRINE

Dopamine Anatomy and Functions

Several distinct populations of dopaminergic neurons have been identified in the brain and are classified as groups A8–A16. The majority of dopaminergic innervation in the brain is provided by long projection neurons arising from two clusters of cell bodies in the midbrain (mesencephalon), the substantia nigra (A9), and the ventral tegmental area (VTA, A10). Neurons from the substantia nigra provide a dense input to the dorsal striatum (caudate nucleus and putamen), while projections from the VTA principally target limbic regions (e.g., nucleus accumbens and septum) in addition to cortical regions (e.g., prefrontal cortex and cingulate cortex). It should also be noted, though, that it is not possible to completely segregate the motor, motivational, and cognitive behaviors that are often attributed to the striatal, limbic, and cortical dopamine systems, respectively.

Dysfunction in dorsal striatal dopamine transmission is strongly associated with movement disorders such as Parkinson disease or tardive dyskinesia. The hallmark signs of Parkinson disease (hypokinesia, rigidity, and tremor) originate from a loss of striatal dopaminergic innervation. Tardive dyskinesia (repetitive, involuntary movements) can occur following high-dose or long-term exposure to certain antipsychotic drugs that block dopamine receptors. Nigrostriatal dopamine neurons do not merely allow motor behavior to occur. They also play an important role in the selection and initiation of actions and in establishing motor skills and habits.

A unique feature of nigrostriatal dopamine neurons is the complex arborization of their terminals. This is illustrated by the estimate that over 1 million striatal synapses are associated with each human nigrostriatal dopamine neuron, which at least two orders of magnitude greater than for other types of neuron. This large architecture, together with the unmyelinated nature of its axons, puts the nigrostriatal dopamine neuron under a phenomenal energy demand and probably is a critical factor in their particular susceptibility to a variety of biochemical, environmental, and pharmacologic insults.

Dopamine neurons arising in the VTA and terminating in the frontal cortex play an important role in executive function, a set of mental functions such as directing attention, working memory, maintaining task goals, decision making, flexible thinking, and self-control. Dopamine neurons innervating limbic regions such as the nucleus accumbens are strongly implicated in goal-oriented (motivated) behaviors, in addition to reward, attention, and pharmacologically induced locomotion. Enhancement of dopamine transmission in the mesolimbic system has been linked with the reinforcing and sensitizing effects of repeated exposure to psychostimulant drugs of abuse and with stress. However, blunting of dopamine release in ventral striatal dopamine has been associated with established addictions and may contribute to neuroadaptive changes that provoke craving during abstinence. Tourette syndrome and attention deficit hyperactivity disorder (ADHD) have been linked to a dysfunction of dopaminergic neurotransmission in limbic and cortical regions, based mainly on the success of dopaminergic drugs in the treatment of their symptoms.

The oldest credible neurochemical hypothesis of schizophrenia posited that the disease is related to excessive dopamine activity in striatal and limbic regions. Subsequently, it was shown that psychosis symptoms of schizophrenia were associated with increased dorsal/associative striatum dopamine release, while negative symptoms and cognitive impairments were associated with reductions in cortical and ventral/limbic striatal dopamine release. There is solid evidence that dopamine is involved in the pathophysiology of schizophrenia, but not necessarily its root cause. Rather, dopamine systems probably act as the common final pathway for a wide variety of predisposing factors, either environmental, genetic, or both, that lead to symptoms of the disease.

Anterior to the mesencephalic substantia nigra and VTA are several other relatively small groups of dopaminergic cell bodies. The most prominent of these are the tuberoinfundibular dopamine neurons, which are located in the arcuate nucleus of the hypothalamus and project to the median eminence. Dopamine released from these neurons enters the circulation and is transported to the anterior pituitary, where it inhibits the release of prolactin from the anterior lobe of the pituitary. Antipsychotic drugs that block dopamine receptors in the pituitary induce an increase in blood prolactin levels.

Electrophysiologic studies have shown that dopamine cells fire in two different modes, either single spiking or burst firing, presumably as a result of a change in the balance between excitatory and inhibitory inputs. As cells are able to switch between these levels, the transition in activity may be a mechanism for altering the impact of dopamine neurotransmission on receptive cells. Bursting activity elicits more terminal dopamine release per action potential than pacemaker activity, and interestingly stimulation of mesocortical neurons in a burst pattern produces a greater release than tonic stimulation of both dopamine and a colocalized peptide, neurotensin. The different rates of firing of dopamine neurons probably code different responses in the brain. At the fastest scale, with burst firing, dopamine neurons can signal a reward or indicate to what extent a reward occurs differently than predicted. At the slowest time scale, the pacemaker activity of dopamine neurons has been linked to a tonic-enabling function on a variety of motor, cognitive, and motivational processes, which are deficient in Parkinson disease and which can be restored by dopamine replacement therapy. Recent data indicate that phasic and tonic modes of dopamine neurotransmission can operate as a continuum, with the postburst elevation of extracellular dopamine being sustained to engage tonic functions of dopamine; such an explanation is consistent with the need for both types of dopamine signaling in the motivational and cognitive functions.

In addition to the dopamine systems in the brain and pituitary, dopamine has restricted yet critical roles in the periphery. The best-defined action for peripheral dopamine neurons is in renal regulation of sodium homeostasis, and impairment of this control may underlie the development of hypertension in some cases. Another important function of dopamine outside the brain is in the retina, where dopamine-containing amacrine cells modulate high acuity, light-adapted vision, and photoreceptor coupling.

Norepinephrine Anatomy and Functions

The different populations of noradrenergic neurons in the brain have been designated A1–A7. The widespread noradrenergic innervation in the brain mainly derives from a small cluster of neurons located in the locus coeruleus (LC, part of the pons, A6) in addition to a more diffuse grouping in the lateral tegmental area (part of the midbrain). These two populations provide input to virtually all brain regions, with LC neurons primarily targeting the forebrain, thalamus, cerebellum, and spinal cord, and with lateral tegmental neurons principally sending fibers to the hypothalamus and basal forebrain. An unusual feature of noradrenergic innervation is that a single neuron typically innervates several brain regions. LC noradrenergic neurons exhibit different activity modes: tonic firing alternating with short periods of burst firing. Increased tonic activity has been associated with behavioral flexibility, whereas the phasic mode is linked with arousing stimuli. There is good evidence for an important role of LC noradrenergic neurons in arousal, vigilance, attention, working memory, and regulation of the sympathetic nervous system.

Epinephrine Anatomy and Functions

In the sympathetic nervous system and the adrenal medulla, epinephrine shares with norepinephrine the role of the final neurotransmitter, and the proportion of this overlap is species dependent and hormonally modified. The relatively few epinephrine-containing cells in the brain exist mostly in the medulla. These neurons predominate in the C1, C2, and C3 regions of the medulla and send axons to a variety of brainstem and forebrain structures, including the LC, dorsal raphe, nucleus accumbens, and provide innervation to subregions of the hypothalamus.

Life Cycle of Dopamine, Norepinephrine, and Epinephrine

Synthesis. These three catecholamine neurotransmitters share a common synthetic pathway (Fig. 1.7–1). Bloodborne tyrosine, derived from dietary proteins and from phenylalanine metabolism, enters the brain by the large neutral amino acid (LNAA) transport system. Tyrosine in brain extracellular fluid is taken up into catecholamine neurons by a high-affinity amino acid transporter. In the first reaction in the cytosol, tyrosine is converted to L-dopa by the enzyme tyrosine hydroxylase. Because this step is rate limiting, it sets the pace for the conversion of tyrosine to dopamine and makes this reaction the most susceptible to physiologic regulation and pharmacologic manipulation. Tyrosine hydroxylase requires tetrahydrobiopterin as a cofactor, and the synthesis of tetrahydrobiopterin is itself dependent on the activity of another enzyme GTP-cyclohydrolase-1. The next reaction is in catecholamine synthesis, the decarboxylation of L-dopa by aromatic amino acid decarboxylase

(AADC) in the cytosol to form dopamine. AADC decarboxylates L-dopa so avidly that the level of this amino acid in the brain is very low under normal conditions. L-dopa is the principal drug used to treat Parkinson disease as it restores the deficient striatal dopamine level that characterizes the disorder. Supplying exogenous L-dopa to the neuron ensures efficient conversion to dopamine because it relies only on AADC and bypasses tyrosine hydroxylase, the rate-limiting step in dopamine biosynthesis. An inhibitor of AADC that does not cross the blood–brain barrier (BBB) (e.g., carbidopa or benserazide) is routinely administered with levodopa to patients with Parkinson disease to inhibit the peripheral decarboxylation of L-dopa to dopamine, which reduces peripheral side effects resulting from abnormally high circulating levels of dopamine. One prominent side effect of excessive circulating dopamine is emesis resulting from the activation of dopamine receptors in the area postrema, an area outside the BBB. In fact, D_2 antagonists (e.g., promethazine [Phenergan]) are a widely used class of antiemetic agents. However, because dopamine antagonists are prone to induce parkinsonian-like side effects, alternatives such as $5HT_3$ antagonists (e.g., ondansetron [Zofran]) and neurokinin-1 receptor antagonists (e.g., aprepitant [Emend]) are increasingly used for controlling the emergence of nausea and vomiting that can occur after chemotherapy and surgery.

In norepinephrine neurons, as well as in chromaffin cells of the adrenal medulla, an additional enzyme, dopamine-β-hydroxylase (DBH), is present that enables the conversion of dopamine to norepinephrine. This enzyme is concentrated inside the storage vesicles, so unlike other reactions, this step of norepinephrine synthesis does not occur in the cytoplasm. Disulfiram (Antabuse), a drug used to treat alcohol addiction, is an inhibitor of DBH. However, the principal relevant property of disulfiram in treating alcoholism is to block acetaldehyde dehydrogenase, the enzyme that metabolizes acetaldehyde formed from ethanol, and this results in elevated acetaldehyde levels that induces moderate to severe adverse effects after consumption of ethanol.

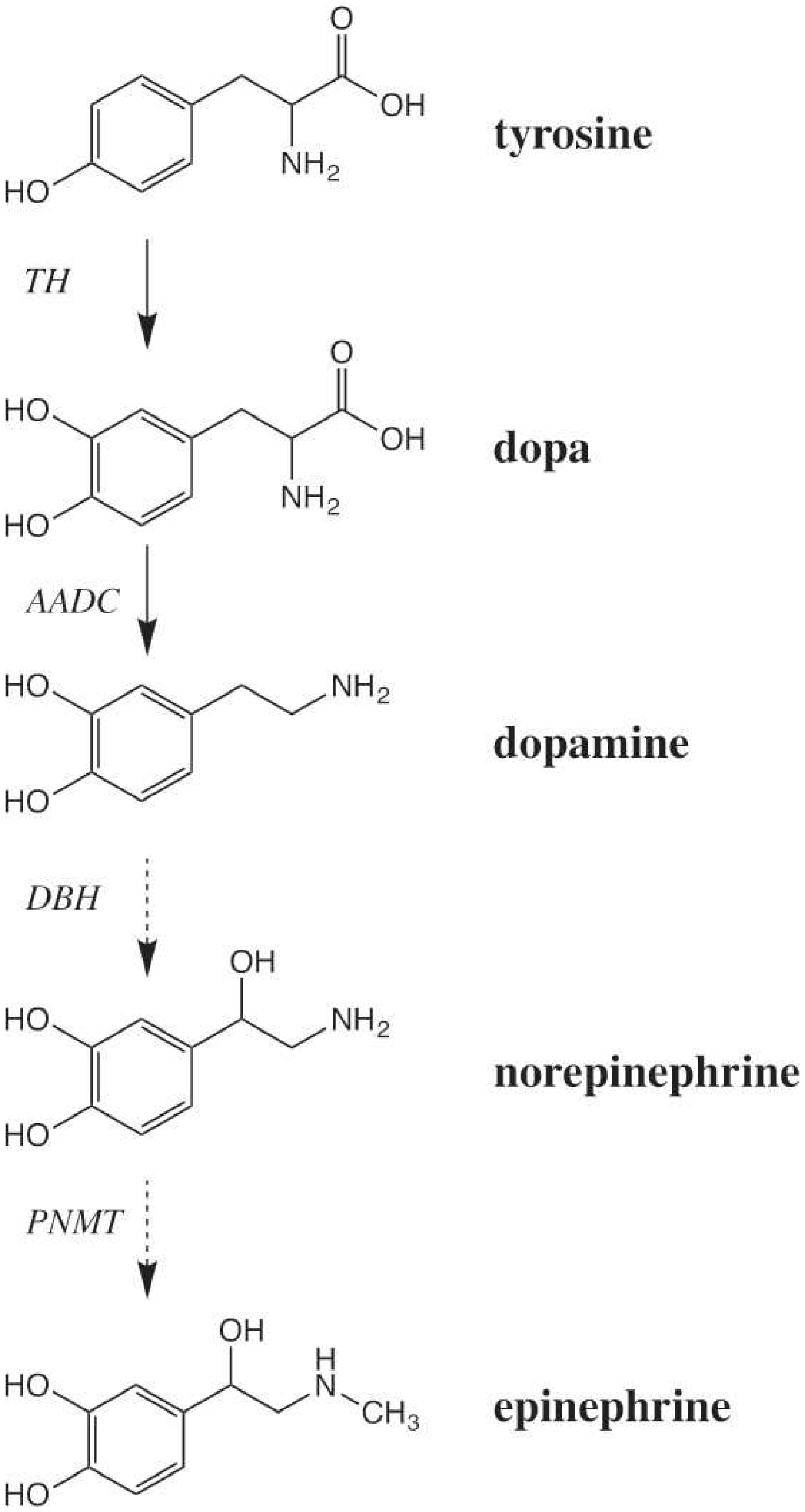
tyrosine
TH
dopa
AADC
dopamine
DBH
norepinephrine
PNMT
epinephrine

FIGURE 1.7–1. Biosynthesis of dopamine, norepinephrine, and epinephrine. The catecholamine neurotransmitters are produced by a multistep synthesis that starts with the dietary amino acid, tyrosine. Dopamine neurons contain

the enzymes TH and AADC, whereas norepinephrine neurons also express DBH. PNMT is an additional enzyme that is present only in epinephrine neurons. AADC, aromatic amino acid decarboxylase; DBH, dopamine-β-hydroxylase; PNMT, phenol-N-methyltransferase.

In central epinephrine neurons and those in the adrenal medulla and heart, norepinephrine is N-methylated by the enzyme phenylethanolamine-N-methyltransferase (PNMT) to form epinephrine. As PNMT is located in the cytoplasm, norepinephrine has to exit its storage site in vesicles to be converted to epinephrine before being transported back into vesicles. This reaction with PNMT presumably takes place during spontaneous leakage of norepinephrine from vesicles in the cytoplasm or after norepinephrine has been released into the synapse and taken back up into the cytoplasm.

Phenylketonuria (PKU) is a rare genetic disease caused by a mutation in the phenylalanine hydroxylase gene, resulting in the production of a nonfunctional enzyme, hampering the metabolism of circulating phenylalanine, which can lead to severe mental retardation unless steps are taken to limit dietary intake of phenylalanine.

Vesicular Transporter. Vesicles that are present in the presynaptic terminals are specialized for the uptake and storage of catecholamines, thus protecting the transmitter from degradation by the enzyme monoamine oxidase (MAO) found in mitochondria. The accumulation and concentration of catecholamine transmitters inside the vesicles is accomplished by the vesicular monoamine transporter (VMAT), and this sequestration is disrupted by amphetamine-like drugs, leading to the mobilization of the transmitter from vesicular storage into the cytoplasm, which increases the concentration available for release to the extracellular compartment. There are two forms of VMAT, with VMAT-2 being primarily located in neuronal cells of the CNS, peripheral nervous system, and enteric nervous system and VMAT-1 being expressed in extraneuronal tissues, such as chromaffin cells of the adrenal medulla.

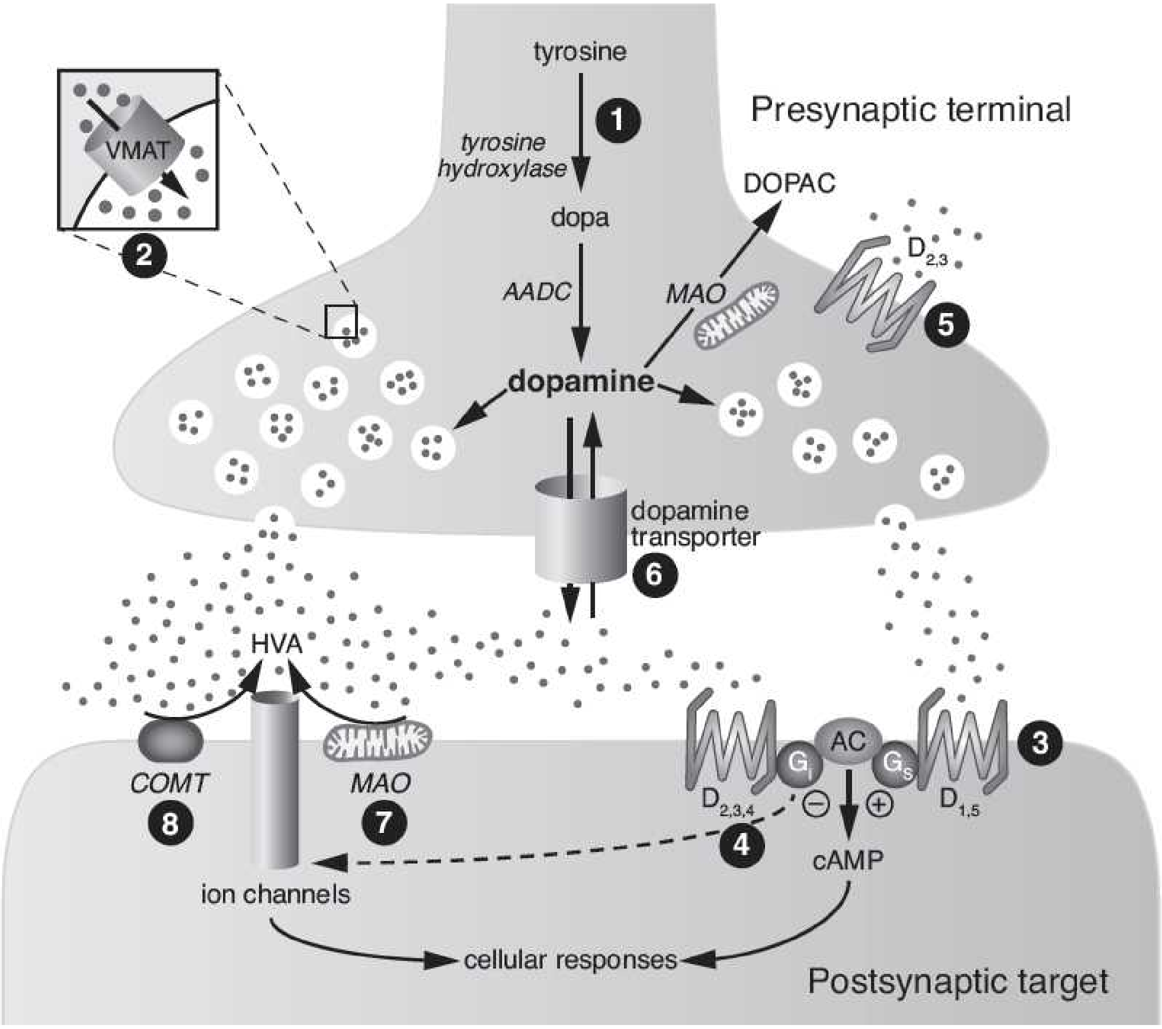
tyrosine
1
tyrosine hydroxylase
Presynaptic terminal
VMAT
2
dopa
DOPAC
AADC
MAO
$D_{2,3}$
5
dopamine
dopamine transporter
6
HVA
COMT
8
MAO
7
AC
G_i
G_s
3
$D_{2,3,4}$
$D_{1,5}$
4
cAMP
ion channels
cellular responses
Postsynaptic target

FIGURE 1.7–2. Schematic model of a dopaminergic synapse with sites of drug action. **Site 1:** Tyrosine hydroxylase reaction blocked by the competitive inhibitor, α-methyltyrosine. Administration of L-dopa augments dopamine levels and is the mainstay treatment of Parkinson disease. Carbidopa is an inhibitor of AADC that does not cross the blood–brain barrier, and is used in conjunction with L-dopa in the treatment of Parkinson disease to reduce the peripheral side effects of L-dopa. **Site 2:** VMAT transports cytoplasmic dopamine into storage vesicles, decreasing its cytoplasmic concentration and preventing metabolism by monoamine oxidase. Reserpine inhibits VMAT and depletes dopamine stores. Tetrabenazine also interferes with the uptake–storage mechanism and has been used for treatment of hyperkinetic movement disorders. **Sites 3 and 4:** Once released, dopamine can interact with postsynaptic D1-like receptors (site 3) and D2-like receptors (site 4). Bromocriptine (Parlodel) and apomorphine (Apokyn) are D2 agonists used in the treatment of Parkinson disease. Pramipexole (Mirapex) is a D3 agonist used in the treatment of Parkinson disease and restless legs syndrome. Aripiprazole (Abilify) is a partial D2 agonist used as an antidepressant. All antipsychotic drugs block D2 receptors to some degree. **Site 5:** Interaction of

dopamine with autoreceptors regulates synthesis and release of dopamine. Low doses of apomorphine have been used experimentally as an autoreceptor agonist. **Site 6:** The plasma membrane transporter recaptures synaptic dopamine. The drug of abuse, cocaine, inhibits the dopamine transporter, and amphetamine increases dopamine release via the transporter. **Site 7:** MAO present outside the presynaptic terminal can metabolize dopamine that has been released. Selegiline (Eldepryl) is an MAO-B inhibitor used in the treatment of Parkinson disease. **Site 8:** Dopamine is a substrate for COMT, an enzyme that is inhibited by tolcapone and entacapone, which are used in treatment of Parkinson disease. AADC, aromatic amino acid decarboxylase; AC, adenylate cyclase; cAMP, cyclic AMP; COMT, catechol-O-methyltransferase; D1–5, dopamine receptor subtypes; DOPAC, dihydroxyphenylacetic acid; G, G protein; HVA, homovanillic acid; MAO, monoamine oxidase; VMAT, vesicular monoamine transporter. (Modified from Nestler E, Hyman S, Holtzman D, Malenka R. *Molecular Neuropharmacology*. McGraw-Hill Companies, Inc; 2015.)

Dopamine Plasma Membrane Transporter. A high-affinity dopamine transporter (DAT) serves to recapture released dopamine and limits the concentration and actions of dopamine in the synapse (Fig. 1.7–2). Importantly, dopamine neurons in the striatum and nucleus accumbens are equipped with a high density of DATs and so are especially susceptible to drugs that target this site and inhibit its actions. For example, blockade of dopamine reuptake by cocaine in the nucleus accumbens prolongs and intensifies dopamine's actions in this region and is key to the euphoric effects of the drug. Similarly, amphetamine and methamphetamine amplify dopamine signaling in the nucleus accumbens by blocking dopamine reuptake and increasing the release of dopamine. The VTA-nucleus accumbens pathway is part of the brain's reward pathway that is activated by all types of reinforcing stimuli such as food, sex, and many drugs of abuse. While drugs that rapidly increase synaptic dopamine concentration can be addictive, some of these agents (e.g., methylphenidate, amphetamine) are used at lower doses to treat ADHD and may achieve their beneficial effects by actions on the dopamine and norepinephrine transporters (NETs) and subsequent interaction of the transmitter with their autoreceptors, which are generally sensitive to a lower concentration of agonists than is necessary to activate postsynaptic receptors. Lisdexamfetamine (Vyvanse) is a long-acting oral prodrug that is hydrolyzed in blood to active D-amphetamine, and it has been approved for treating ADHD. The DAT also provides access to dopamine neurons for parkinsonian neurotoxins, 6-hydroxydopamine, and 1-methyl-4-phenyl-

1,2,3,6-tetrahydropyridine (MPTP) because they are substrates for the transporter.

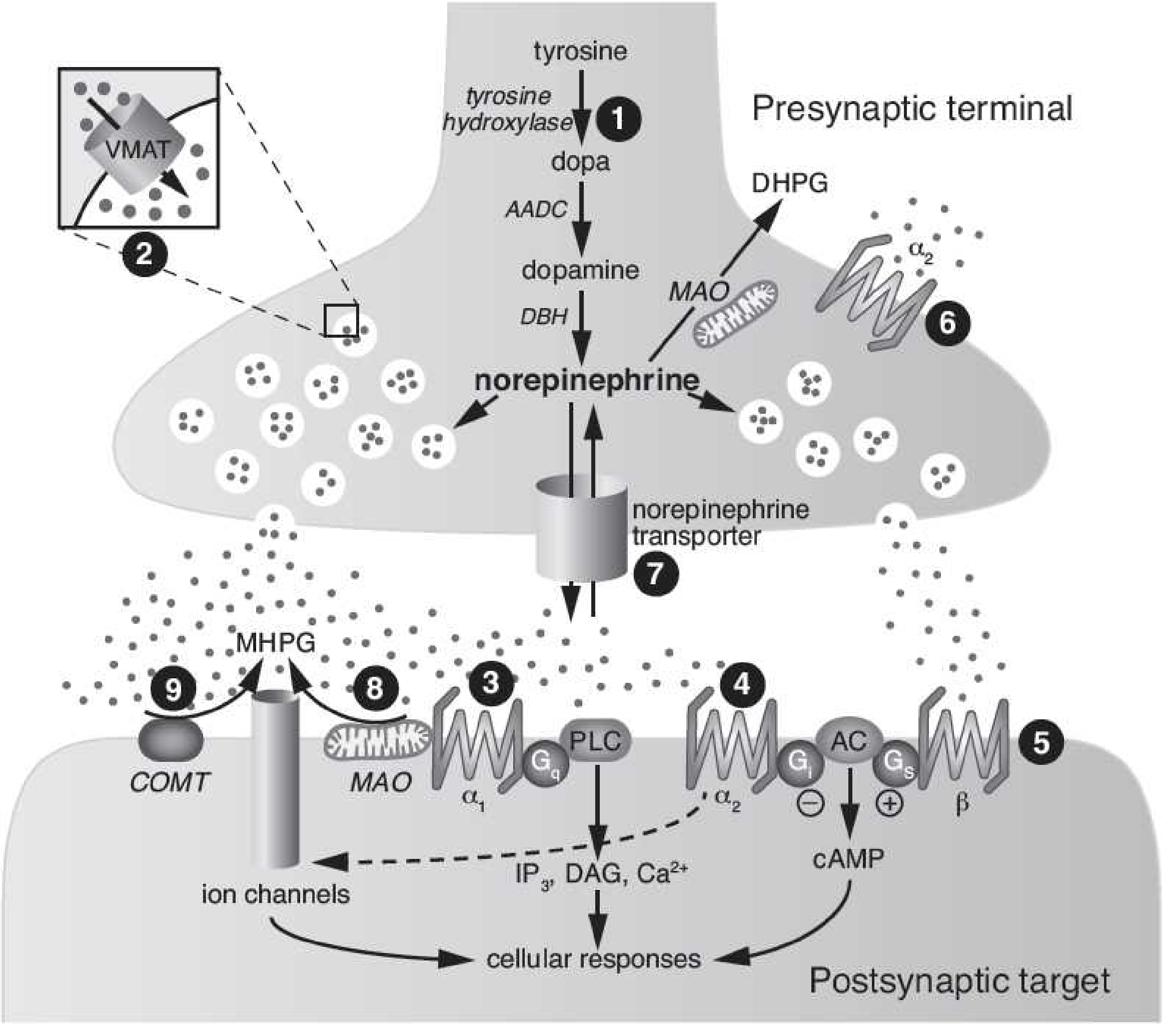
tyrosine
tyrosine hydroxylase
1
dopa
AADC
dopamine
DBH
norepinephrine
Presynaptic terminal
DHPG
MAO
α_2
6
VMAT
2
norepinephrine transporter
7
MHPG
9
8
3
4
5
COMT
MAO
α_1
G_q
PLC
α_2
G_i
AC
G_s
⊖
⊕
β
ion channels
IP_3, DAG, Ca^{2+}
cAMP
cellular responses
Postsynaptic target

FIGURE 1.7–3. Schematic model of a central norepinephrine synapse with sites of drug action. **Site 1:** Tyrosine is transported into the presynaptic terminal by an active uptake mechanism and converted to NE by a series of enzymatic steps. The competitive inhibitor, α-methyltyrosine blocks the tyrosine hydroxylase reaction. DBH reaction is prevented by disulfiram. **Site 2:** Norepinephrine is taken up from the cytoplasm and transported into storage vesicles by the VMAT-2. Reserpine inhibits VMAT and depletes norepinephrine stores; the drug was once widely prescribed for its antihypertensive actions. Tetrabenazine also interferes with the uptake–storage mechanism and has been used for treatment of hyperkinetic movement disorders. **Sites 3, 4, and 5:** After release, norepinephrine can interact with two categories of G-protein–coupled adrenergic receptors, α1 and α2 (**sites 3 and 4**) and β (**site 5**). The majority of adrenergic receptors are localized postsynaptically where they mediate the cellular responses of the postsynaptic neuron. Propranolol (Inderal) is an antagonist of β-adrenergic receptor, and is used to treat performance anxiety and hypertension. α1-Receptor (**site 3**) is inhibited prazosin (Minipress), used as an antihypertensive, but which also has anxiolytic effects used in the treatment

PTSD. **Site 6:** α2-Autoreceptors localized in the nerve terminals modulate the synthesis and release of norepinephrine, while cell body and dendritic α2-autoreceptors modulate impulse flow. Low doses of clonidine (Catapres) and guanfacine (Intuniv) are agonists at these sites. α2-Agonists are used to treat opioid withdrawal, Tourette syndrome, and hypertension. Yohimbine is an antagonist at this receptor site. **Site 7:** The synaptic action of norepinephrine is terminated by reuptake into the presynaptic terminal by norepinephrine plasma membrane transporter. The tricyclic antidepressants, such as desipramine (Norpramin), are inhibitors of this uptake mechanism. Some newer antidepressants are more selective for NET, such as atomoxetine (Strattera), and by SNRIs such as duloxetine (Cymbalta). **Site 8:** Norepinephrine is metabolized by MAO and it is a preferential substrate for MAO-A. Inhibitors of MAO-A block metabolism of norepinephrine, and this action may contribute to their antidepressant property. Examples of irreversible nonselective inhibitors are isocarboxazid, phenelzine, and tranylcypromine. Moclobemide inhibits MAO-A reversibly. **Site 9:** Norepinephrine is also metabolized by COMT, which is inhibited by tolcapone (Tasmar) and entacapone (Comtan), drugs that are used in treatment of Parkinson disease. The combined actions of MAO and COMT produces MHPG, a major brain norepinephrine metabolite. α and β*,* subtypes of norepinephrine receptor; AADC, aromatic amino acid decarboxylase; AC, adenylate cyclase; cAMP, cyclic AMP; COMT, catechol-O-methyltransferase; DAG, diacylglycerol; DBH, dopamine β-hydroxylase; DHPG, 3,4-dihydroxyphenylglycol; G, G protein; IP3, inositol triphosphate; MAO, monoamine oxidase; MHPG, 3-methoxy-4-hydroxyphenylglycol; SNRI, serotonin–norepinephrine reuptake inhibitor; VMAT, vesicular monoamine transporter. (Modified from Nestler E, Hyman S, Holtzman D, Malenka R. *Molecular Neuropharmacology*. McGraw-Hill Companies, Inc; 2015.)

Norepinephrine Plasma Membrane Transporter. The NET is located on all noradrenergic neurons and efficiently recaptures most of the transmitter that is released into the synapse (Fig. 1.7–3). In addition to the impact of the psychostimulants (e.g., cocaine and amphetamines) on DAT, these drugs also interact with NET and intensify the synaptic actions of NE. The mechanism of certain antidepressant drugs includes blockade of NET, and these include the tricyclic antidepressants (TCAs) and the more specific serotonin–norepinephrine reuptake inhibitors (SNRIs) such as duloxetine (Cymbalta) and venlafaxine (Effexor), norepinephrine–dopamine reuptake inhibitors (NDRIs) such as bupropion (Wellbutrin), and the selective norepinephrine reuptake inhibitors (NRIs), such as maprotiline and atomoxetine (Strattera), which is also used in the treatment of ADHD.

Some of the original studies on dopamine uptake revealed that dopamine is a good substrate for the NET. However, not until relatively recently has the importance of this been appreciated. In regions, such as the striatum and nucleus accumbens, where dopamine neurons express a high density of DAT, there is a relatively sparse norepinephrine innervation, and dopamine uptake depends primarily on DAT. In contrast, in regions such as the prefrontal cortex, where dopamine neurons express low levels of DAT, and there is a rich norepinephrine innervation, dopamine uptake depends strongly on NET. In fact, it has been suggested that dopamine can be co-released with norepinephrine by noradrenergic neurons as a result of nonspecific uptake of dopamine by the NET and/or because dopamine is a precursor in the biosynthesis of norepinephrine.

Epinephrine Plasma Membrane Transporter. There is no evidence for a specific transporter for epinephrine, and no drug that selectively blocks epinephrine reuptake has been described. All of the pharmacologic agents that block the NET selectively are also effective inhibitors of epinephrine uptake. Thus, it appears unlikely that a selective transporter for epinephrine exists.

Metabolism. The major mammalian enzymes of importance in the metabolic degradation of catecholamines are MAO and catechol O-methyltransferase (COMT), and either of these enzymes can catalyze the first step in catecholamine metabolism. MAO converts its substrates to their corresponding aldehydes, and these aldehyde intermediates are rapidly metabolized either by the aldehyde dehydrogenase or aldehyde reductase. Thus, multiple metabolites of catecholamines are formed. In the primate brain, the major end metabolites for dopamine are homovanillic acid (HVA), whereas, in the rodent brain, 3,4-dihydroxyphenylacetic acid (DOPAC) and DOPAC sulfate predominate. For norepinephrine, 3-methoxy-4-hydroxyphenylglycol (MHPG) is the main end metabolite in the primate brain, but in the rodent brain, both 3,4-dihydroxyphenylglycol (DHPG) and MHPG exist in high concentrations, and both metabolites are mainly sulfate conjugated. In the periphery, the major metabolite of dopamine is HVA, and for norepinephrine, it is vanillylmandelic acid (VMA), with a significant proportion of both compounds being conjugated to either sulfate or to glucuronide.

It is well documented in animals that short-term fluctuations in the levels of dopamine metabolites, DOPAC, and HVA, in the striatum, nucleus accumbens, and prefrontal cortex provide a useful index of alterations in impulse flow in the nigrostriatal, mesolimbic, and mesocortical dopamine pathways, respectively. Changes in dopamine metabolite concentration in CSF have also been associated with alterations in central dopaminergic activity or integrity of central dopamine systems. Similarly, shifts in the concentration of MPHG or DHPG have been correlated with changes in norepinephrine function.

MAO. This enzyme is located on the outer membrane of mitochondria. Separate genes encode two isoforms of MAO (types A and B), which can be distinguished by substrate specificity and sensitivity to selective inhibitors. In the brain, MAO-A is preferentially located in dopaminergic and noradrenergic neurons, while MAO-B is the major form present in serotonergic neurons and glia. Dopamine and norepinephrine are effective substrates for both forms, whereas serotonin is a preferential substrate for MAO-A. The predominance of MAO-B in serotoninergic neurons appears counterintuitive as this transmitter is a better substrate for MAO-A; it is possible that MAO-B in serotoninergic neurons serves to regulate phenylethylamine levels a “trace amine” that is a selective MAO-B substrate and can affect serotonin uptake and efflux. The prototypical inhibitors of the two forms are clorgyline for MAO-A and selegiline (formerly known as deprenyl) for MAO-B. Selegiline (Eldepryl) or rasagiline (Azilect) are irreversible MAO-B inhibitors, and Safinamide (Xadago) is a reversible MAO-B inhibitor that is used in the treatment of Parkinson disease, usually in combination with L-dopa to prolong or enhance the effect of dopamine formed from L-dopa.

COMT. There are two isoforms of COMT, a membrane-bound form, and a soluble form. Membrane-bound COMT is the major form found in the CNS, has a higher affinity for catecholamines, and is located principally in neurons. The soluble form has a lower affinity for catecholamines and is the major form expressed in the periphery but is present also in CNS glia. Two COMT inhibitors, tolcapone and entacapone, are used as adjuncts to L-dopa in the treatment of Parkinson disease to slow L-dopa metabolism and thereby prolong its effects and reduce L-dopa requirements.

Accumulating evidence from clinical and preclinical studies shows that COMT plays a more significant role in dopamine metabolism in the prefrontal cortex than in other regions such as the striatum, likely because of the paucity of DAT in the prefrontal cortex. Another interesting facet of the role of COMT in the frontal cortex is the finding that the COMT gene contains a polymorphism (Val158Met) that affects the in vivo activity of the enzyme. Met158 homozygotes have approximately one-third less COMT enzyme activity in the prefrontal cortex than Val158 homozygotes, and this aligns with an altered rate of dopamine metabolism by the alleles. COMT Val158Met polymorphism has been associated with different phenotypes, such as reward responsiveness and reward-seeking, cognitive abilities, and stress processing. However, it should be noted that an inverted "U"–shaped curve describes the relationship between dopaminergic transmission in the prefrontal cortex and behavior-like cognition, with both sub- and supraoptimal dopamine activity impairing performance.

Dopamine Receptors

The physiologic actions of dopamine are mediated by five distinct but closely related GPCRs. On the basis of their structure, pharmacology, and primary effector mechanisms, they are grouped as D_1-like (D_1 and D_5) and D_2-like (D_2, D_3, D_4) dopamine receptors (Fig. 1.7–2). The D_1-like dopamine receptors activate the Gs family of G proteins, increase adenylate cyclase activity, and are found postsynaptically on dopamine-receptive cells. The D_2-like dopamine receptors (D_2, D_3, and D_4) couple to the Gi family, inhibit adenylate cyclase activity and are expressed both postsynaptically on dopamine-receptive cells and presynaptically on dopaminergic neurons.

In the brain, D_1 dopamine receptors are expressed in particularly high density in the nigrostriatal, mesolimbic, and mesocortical dopamine systems, such as the striatum, nucleus accumbens, and frontal cortex. In contrast, D_2 receptors are more evenly distributed in the brain, with the highest levels in the striatum and nucleus accumbens. Dopamine transmission through D_1 and D_2 receptors in the striatum and prefrontal cortex plays an essential role in many aspects of motor control, reward systems, and cognition. The alternative splicing of the D_2 gene results in the generation of two major variants, termed D_{2S} (D_2-short) and D_{2L} (D_2-long). There is some evidence that D_{2S} is mostly expressed presynaptically and is involved in autoreceptor functions, with D_{2L} being predominantly a postsynaptic isoform. The D_3 dopamine receptor has a more limited pattern of distribution, with the highest level of expression being observed in the limbic areas, such as in the shell subdivision of the nucleus accumbens. Some data suggest that D_3 receptors can operate as autoreceptors to complement the D_2 autoreceptor role in regulating the neuronal firing rate, synthesis, and release of dopamine. Among the dopamine receptors, the D_4 subtype has the lowest level of expression in the brain, with distribution in the frontal cortex, amygdala, hippocampus, and hypothalamus. D_5 dopamine receptors are expressed at low levels in multiple brain regions, including the frontal cortex, hypothalamus, and hippocampus. D_1, D_2, and D_4 dopamine receptors are also present in the retina. The important role of dopamine in regulating the hypothalamic–pituitary axis accounts for the high expression of D_2 dopamine receptors in the pituitary gland. In the periphery, all subtypes of dopamine receptors have been observed in varying proportions in the kidney, adrenal glands, sympathetic ganglia, gastrointestinal (GI) tract, blood vessels, and heart.

Some of the older, so-called “*typical*” antipsychotic drugs, such as chlorpromazine, block both D_1 and D_2 receptors, while others, like haloperidol, are selective for D_2 receptors. Typical antipsychotic drugs are prone to induce extrapyramidal side effects, which include Parkinsonism, dystonia, and choreiform movements (tardive dyskinesia). “A*typical*” antipsychotics were originally defined by their diminished risk of extrapyramidal side effects compared with typicals, and mechanistically this has been traditionally linked with blockade of D_2 dopamine receptors and $5HT_{2A}$ serotonin receptors with little or no impact on D_1 receptors. Atypical antipsychotics are thought to be more effective in reducing cognitive and negative symptoms of schizophrenia than the typical antipsychotics, although it is not clear to what extent the interaction with serotonin receptors improves the antipsychotic effect or reduces some of the unwanted effects of dopamine receptor blockade. As the full beneficial effects of antipsychotic drugs can take weeks to develop, it is apparent that the therapeutic mechanism involves more than mere blockade of the D_2 receptor site. An additional side effect of antipsychotic drugs is elevated serum prolactin, resulting from a disturbance of dopamine-mediated regulation of prolactin secretion. The consequences of hyperprolactinemia include menstrual disturbances, galactorrhea, sexual dysfunction, gynecomastia, infertility, decreased bone mineral density, and possibly breast cancer. Atypical antipsychotics with an especially high affinity for D_2 receptors, such as risperidone (Risperdal), are more prone to induce hyperprolactinemia, whereas this side effect is rare with those drugs that bind to D_2 less avidly, such as aripiprazole (Abilify), clozapine (Clozaril), olanzapine (Zyprexa), quetiapine (Seroquel), and ziprasidone (Geodon). Although clozapine was discovered in 1959 and first used as an antipsychotic in the 1970s, it still has an important role in combatting treatment-resistant schizophrenia with a low likelihood of extrapyramidal side effects. Current thinking is that clozapine's combined actions at several receptors may be responsible for its therapeutic properties. However, even with careful monitoring, approximately 0.4% of patients treated with clozapine develop serious agranulocytosis. Of the newer antipsychotic drugs, olanzapine has the closest receptor profile to clozapine. Weight gain and induction of metabolic syndrome are central challenges to the use of several atypical drugs, especially with olanzapine, and coadministration of medications such as metformin or samidorphan have been utilized to combat these side effects

D_2-like agonists have an important role in the symptomatic treatment of Parkinson disease. Such drugs include apomorphine, a D_2 agonist, and also pramipexole (Mirapex) and ropinirole (Requip), which are D_2-like agonists with the greatest affinity for the D_3 receptor. Nonselective dopamine agonists (e.g., rotigotine) have also been used to treat symptoms of Parkinson disease, in addition to restless legs syndrome.

Historically, GPCRs were believed to function as monomeric units; however, it is now apparent in some instances, specific biogenic amine receptors may directly interact with other members of the same receptor family or even with structurally divergent families of receptors to form heteromers (e.g., D1-D2, D2-5HT2, D2-cannabinoid receptor-1) which have unique functional properties. Pharmacologic targeting of heteromers offers a novel strategy for the therapeutic manipulation of cell signaling.

Adrenergic Receptors

For many years, receptors for norepinephrine and epinephrine (adrenergic receptors) were divided into just three distinct classes, α_1, α_2, and β. A postsynaptic location is typical for α_1- and β-receptors, whereas α_2-receptors exist both pre- and postsynaptically. Currently, three α_1

subtypes (α_{1A}, α_{1B}, α_{1D}), four α_2-receptor subtypes ($\alpha_{2A–D}$), and three β-receptor subtypes ($\beta_{1–3}$) are recognized (Fig. 1.7–3). All β-receptors are Gs-coupled, whereas most α_1-receptors are Gq-coupled, and α_2-receptors are typically Gi-coupled. As norepinephrine is the transmitter used by most postganglionic neurons of the autonomic nervous system (ANS), agonists and antagonists at various adrenergic receptors exert profound effects on the functioning of many peripheral organs by mimicking or blocking the sympathetic innervation respectively. Agonists are referred to as "*sympathomimetics*" and antagonists as "*sympatholytics.*" In the brain, the LC noradrenergic system exerts a widespread influence on neuronal circuits that are essential substrates of arousal and cognition, and several affective disorders have been posited to involve a dysregulation of noradrenergic neurotransmission. In addition, by sending direct projections to the sympathetic preganglionic neurons in the spinal cord and parasympathetic preganglionic neurons in the brainstem and spinal cord, the noradrenergic LC plays a central role in regulating autonomic activity.

The predominant peripheral effect of α_1-adrenergic agonists is the constriction of arterial smooth muscle, which in turn causes an increase in blood pressure, and drugs such as phenylephrine and methoxamine are used occasionally in the clinic to increase blood pressure during a hypotensive crisis. α_1-Adrenergic agonists are also widely used as nasal decongestants, and their decongestant activity is mediated by constriction of blood flow to the nasal mucosa and by a reduction in airway secretions. Antagonists of α_1-adrenergic receptors inhibit smooth muscle contraction and reduce arteriolar resistance by dilating arteries. Their major uses are as an adjuvant treatment for hypertension and for treating symptomatic benign prostatic hypertrophy. α_1-Antagonists such as prazosin (Minipress) have also been shown to be useful in treating some of the symptoms of post-traumatic stress disorder (PTSD).

α_2-Adrenergic agonists, like clonidine (Catapres) and guanfacine (Intuniv), have been used to combat hypertension and the mechanism appears to be a reduction in sympathetic tone elicited by activation of α_2-receptors in the anterior hypothalamus. α_2-Adrenergic agonists also have a role in the treatment of certain CNS disorders, especially those involving prefrontal cortex dysfunction. Damage to the catecholaminergic innervation of the prefrontal cortex impairs performance in tests of executive function, and although dopamine has a prominent role in mediating the effects of these lesions, substantial evidence indicates that norepinephrine also exerts a potent modulatory influence on prefrontal cortex functions. Early pharmacologic studies on the role of α_2-receptors in prefrontal cortex function focused on clonidine, although this drug produces mixed effects, presumably due to competing pre- versus postsynaptic effects and limitations from the sedative and hypotensive side effects. A superior profile has been observed with the more selective α_{2A}-agonist, guanfacine, which can improve working memory in animals and in some human populations, possibly by binding to postsynaptic α_{2A}-receptors and increasing noradrenergic tone. Guanfacine is now being used to treat a variety of disorders with prominent prefrontal cortex dysfunction, including Tourette syndrome and ADHD. Clonidine is also employed to manage the acute phase of opioid withdrawal, an effect achieved by normalizing the hyperactivity of noradrenergic LC neurons.

β-Adrenergic agonists have profound effects on the cardiac and pulmonary systems and are one of the primary agents used to treat asthma and other chronic obstructive pulmonary diseases. Antagonists of β-adrenergic receptors are mainstays in the treatment of ischemic heart disease and hypertension. β-Adrenergic antagonists, such as propranolol, are also effective in reducing signs of increased sympathetic activity in individuals with social anxiety (stage fright), such as hand tremor, sweating, and rapid heartbeat. Preclinical studies have demonstrated that propranolol reduces the reconsolidation of fear memories, although its effectiveness in the treatment of PTSD has not been resolved.

SEROTONIN

Anatomy and Functions

Mammals employ serotonin (5HT) as a neurotransmitter within both the CNS and peripheral nervous systems. Serotonin is also produced in many cells that are not neurons, such as mast cells and enterochromaffin cells, and bloodborne serotonin is taken up and concentrated in platelets. More than 80% of the serotonin in the body occurs in the GI system, where it modulates motility and digestive functions. In the brain, only one in a million neurons produces serotonin, yet this evolutionary ancient signaling molecule influences virtually all aspects of CNS function due to the extensive innervation pattern of the neurons. Due to this widespread distribution in the brain, serotonin has been implicated in a vast array of physiologic and pathologic processes.

The populations of serotonergic cell bodies, classified as groups B1–B9, are clustered in multiple midline raphe nuclei of the brainstem. The caudal raphe nuclei send descending projections into the medulla, cerebellum, and spinal cord, which are involved in motor activity, pain control, and regulation of autonomic processes. The more rostrally located dorsal and median raphe nuclei send ascending axonal projections to the rest of the brain. The median raphe provides most of the serotonergic fibers that innervate the limbic system, while the dorsal raphe nucleus provides most of the serotonergic fibers that innervate the cerebral cortex, thalamus, and striatum, in addition to the dopaminergic neurons residing in the substantia nigra and VTA. These rostral serotonergic systems have a modulatory role in many functions, including mood, anxiety, aggression, cognition, feeding, the sleep–wake cycle, and sexual behavior. The link between serotonin and depression is particularly compelling, and the clinical efficacy of antidepressant drugs is strongly associated with an enhancement of serotonergic function.

Interestingly, the projections from the dorsal and median raphe can be distinguished morphologically. Dorsal raphe serotonergic fibers are very fine, with small vesicle-coated swellings called varicosities that are fusiform or granular in appearance, while median raphe fibers are thicker with large spherical or beaded varicosities. The two systems coexist in most areas of the brain, although the ratio varies considerably. The striatum, thalamus, and dopaminergic cell bodies in the midbrain receive innervation predominantly by the neurons originating in the dorsal raphe, while the hippocampus, septum, and hypothalamus receive a particularly high proportion of fibers arising in the median raphe. In the cerebral cortex, there is extensive overlap between dorsal and median raphe projections. The two types of fibers exhibit differential sensitivity to the neurotoxic effects of the amphetamine analog 3,4-methylenedioxy-methamphetamine (MDMA, “ecstasy”), which causes long-lasting or permanent damage to the fine axons of the dorsal raphe while sparing the thick beaded axons of the median raphe.

Life Cycle

Synthesis. In the brain, the first step in serotonin synthesis is the active uptake of the dietary amino acid tryptophan across the BBB and then into neurons (Fig. 1.7–4). Brain tryptophan is utilized in all brain cells for protein synthesis, in addition to serving as a substrate for transmitter synthesis in serotonin neurons. The carrier processes that transport tryptophan from plasma into the brain and from the extracellular fluid into neurons are open to competition from LNAAs, including other aromatic amino acids (e.g., tyrosine and phenylalanine), branched-chain amino acids (e.g., leucine, isoleucine, and valine), and others (e.g., methionine and histidine). The competitive nature of the LNAA carrier means that brain levels of tryptophan will be determined not only by the plasma concentration of tryptophan but also by the plasma concentration of competing neutral amino acids. The rate-limiting enzyme for the synthesis of serotonin in the brain, tryptophan hydroxylase, is not saturated under normal physiologic conditions. Therefore, any decrease or increase in brain tryptophan levels results in a reduction or stimulation of serotonin synthesis, respectively. Thus, dietary protein and carbohydrate content can specifically influence brain tryptophan and serotonin levels by effects on plasma amino acid patterns. Because plasma tryptophan has a daily rhythmic variation in its concentration, this concentration variation can also influence the rate and synthesis of brain serotonin. However, it is unclear to what extent pre- and postsynaptic homeostatic mechanisms prevent tryptophan-induced alterations in serotonin synthesis from being translated into changes in neurotransmission.

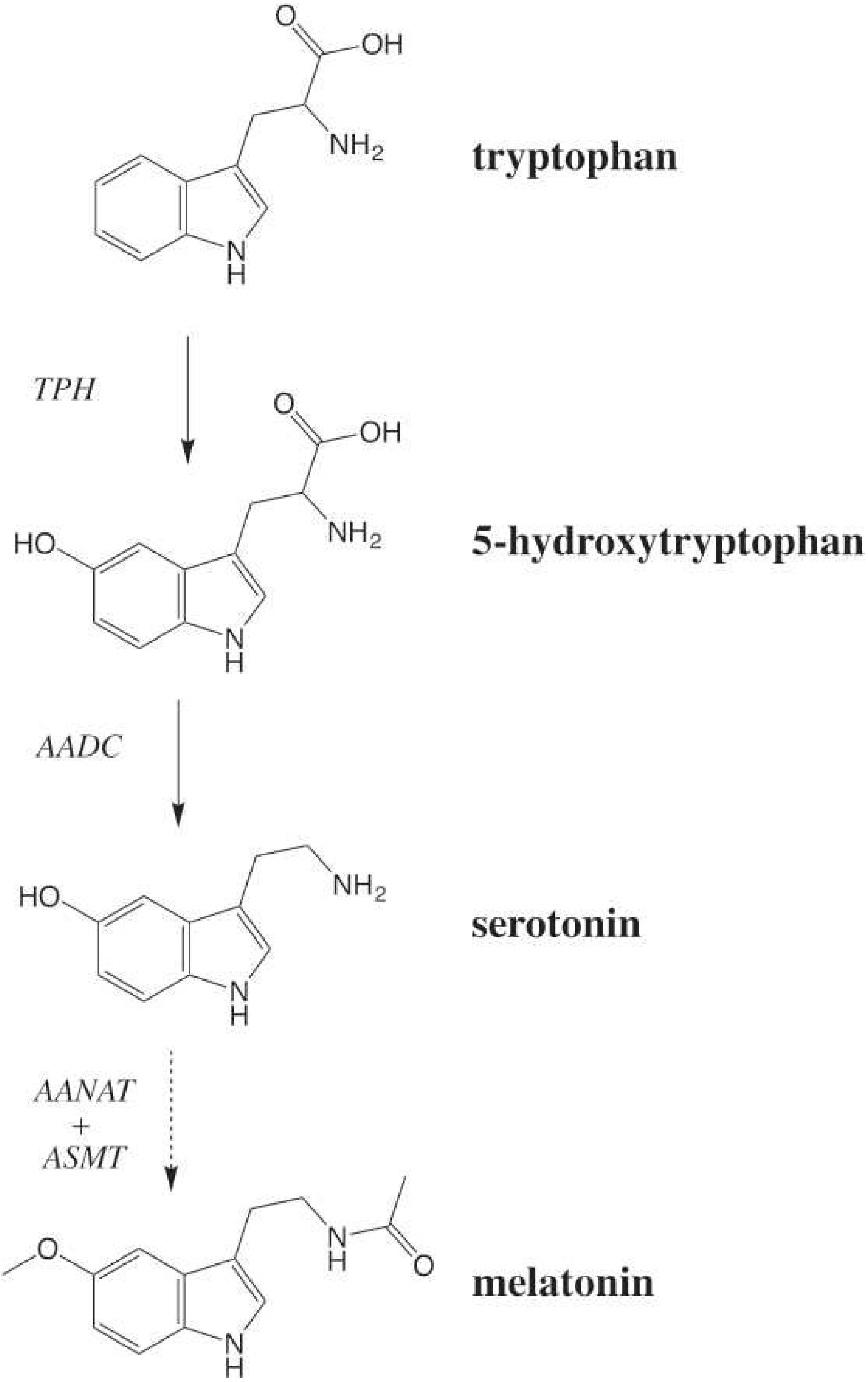
O
OH
NH2
NH
tryptophan
TPH
O
OH
HO
NH2
NH
5-hydroxytryptophan
AADC
HO
NH2
NH
serotonin
AANAT
+
ASMT
O
NH
O
NH
melatonin

FIGURE 1.7–4. Biosynthesis of serotonin and melatonin. Serotonin (5-hydroxytryyptamine, 5HT) and melatonin are produced by a multistep synthesis that starts with the dietary amino acid, tryptophan. Cells in the

pineal gland are the major site of melatonin synthesis, and they contain the enzymes AANAT and ASMT that are necessary to convert serotonin to melatonin. AADC, aromatic amino acid decarboxylase; AANAT, aralkylamine *N*-acetyltransferase (also known as serotonin acetylase); ASMT, acetylserotonin O-methyltransferase (also known as hydroxyindole O-methyltransferase); TPH, tryptophan hydroxylase.

Tryptophan hydroxylase catalyzes the formation of 5-hydroxytryptophan from tryptophan, and this step in the synthesis of serotonin can be specifically blocked by p-chlorophenylalanine, which competes directly with tryptophan and binds irreversibly to the enzyme. Once synthesized from tryptophan, 5-hydroxytryptophan is almost immediately decarboxylated to yield serotonin. The enzyme responsible for this conversion is AADC and is identical to the enzyme that decarboxylates L-dopa in dopamine and norepinephrine neurons. AADC is widespread in distribution; it is found in the peripheral and CNS associated with catecholamine- and serotonin-containing neurons and in the adrenal and pineal glands. It is also found in the kidney, liver, and various other tissues in which little or no monoamine transmitter is normally produced. It is possible to increase serotonin formation experimentally by administering 5-hydroxytryptophan and bypassing the rate-limiting tryptophan hydroxylase step. Thus, unlike tryptophan administration, which can result in a selective increase in serotonin in serotonin-containing neurons, 5-hydroxytryptophan administration will result in the nonspecific formation of serotonin at all sites containing AADC, including the catecholamine-containing neurons.

In the pineal gland, serotonin can be further metabolized by acetylation and O-methylation to form melatonin. Tryptophan is a precursor to serotonin and melatonin, yet ingestion of large quantities of tryptophan will not amplify the production of melatonin in the pineal because of the restricted availability and activity of enzymes necessary to complete this conversion. Melatonin has a major role in the regulation of cyclic bodily activities. The suprachiasmatic nucleus of the anterior hypothalamus is the master clock controlling circadian rhythms in mammals, and it controls melatonin secretion from the pineal by an indirect pathway in response to the environmental light/dark cycle. Peak melatonin secretion occurs during darkness. Melatonin signals both the time of the day (“clock” function) and the season of the year (“calendar” function) to many tissues of the body. Melatonin is available for over-the-counter use as a hypnotic agent to induce sleep but has limited effectiveness in healthy subjects at night when endogenous production of melatonin is high. Administration of melatonin may be of use to treat jet lag or to help shift workers adjust to new work hours. Melatonin agonists have been developed and recently marketed, such as ramelteon (Rozerem).

Transporters. After the synthesis of serotonin in the cytoplasm, it is actively sequestered and concentrated in synaptic vesicles to await release (Fig. 1.7–5). The uptake of serotonin into vesicles is accomplished by VMAT-2 in the brain and is inhibited by drugs such as reserpine and tetrabenazine. Once released into the synapse, serotonin can interact with pre- and postsynaptic serotonin receptors. The synaptic effects of serotonin are limited by reuptake into the presynaptic terminal and by enzymatic conversion of serotonin to inactive metabolites. Reuptake serves as the major mechanism for the termination of the action of synaptic serotonin. The primary targets for the TCAs such as amitriptyline (Elavil), clomipramine (Anafranil), desipramine (Norpramin), imipramine (Tofranil), and nortriptyline (Pamelor) is blockade of serotonin and NETs, which prolongs the action of the transmitters in the synapse, although this class of drug possesses several other pharmacologic actions. Newer selective serotonin reuptake inhibitors (SSRIs) are now the most widely prescribed antidepressants and antianxiety medications, such as citalopram (Celexa), escitalopram (Lexapro), fluoxetine (Prozac), fluvoxamine (Luvox), sertraline (Zoloft), and paroxetine (Paxil). More recently, SNRIs, such as duloxetine (Cymbalta), venlafaxine (Effexor), and desvenlafaxine (Pristiq), have been introduced for the treatment of depressive disorders. SSRIs are also acknowledged as the first-line pharmacologic treatment of choice for obsessive-compulsive disorder (OCD). A polymorphism occurs in the promoter region of the serotonin transporter gene that affects the function of the protein, and this finding has provoked considerable controversy over the possible consequences of a particular variant (short allele of 5-HTTLPR). It now appears that there is a small but significant effect of this polymorphism on susceptibility for bipolar depression (BPD) and a reduced remission rate following antidepressant treatment.

Psychomotor stimulants like amphetamine enter monoamine nerve terminals via DAT, NET and the serotonin transporter (SERT) and stimulate the release of all three monoamines into the synapse by rapid reverse transport. Halogenated amphetamine derivatives (e.g., fenfluramine) are selective for SERT compared with DAT and NET. There is some evidence that the appetite-suppressing action of fenfluramine is mediated by the release of serotonin and activation of $5HT_{2C}$ receptors in the medial hypothalamus. Fenfluramine alone or in combination with phentermine (called “fen-phen”) was formerly prescribed as an appetite suppressant. However, fenfluramine was withdrawn from the market when it was found to be associated with a high incidence of pulmonary hypertension and valvular heart disease. This toxic effect is believed to be related either to the peripheral release of serotonin from platelets or, more likely, as a result of activation of cardiac $5HT_{2B}$ receptors by norfenfluramine, the main metabolite of fenfluramine.

Metabolism. Serotonin is enzymatically degraded by MAO, and the product of this reaction, 5-hydroxyindoleacetaldehyde, is oxidized to 5-hydroxyindoleacetic acid (5-HIAA). Some research has shown that people with a particular loss-of-function variant of the MAO-A gene are predisposed to violence and other antisocial behaviors, but this research is quite controversial. Inhibitors of MAO were the first drugs to be marketed as antidepressants, and these monoamine oxidase inhibitors (MAOIs) were not selective for the A or B isoforms and inhibited the enzyme irreversibly, such as tranylcypromine (Parnate) and phenelzine (Nardil). However, the propensity of tyramine-rich foods (such as cheese) to provoke dangerous hypertensive reactions in patients taking these MAOIs was soon recognized and forced their use to be severely curtailed. As the intestinal mucosa contains MAO-A predominantly, it is an important site for metabolizing ingested MAO substrates like tyramine. When intestinal MAO-A is inhibited, dangerous amounts of tyramine can enter the circulation and release norepinephrine from sympathetic neurons and epinephrine from the adrenal medulla, where the metabolism of these transmitters is blunted because of MAO-A inhibition, leading to a pressor response. Selective MAO-B inhibitors were shown to be safer with regard to this “cheese reaction” but unfortunately lacked antidepressant efficacy. Nevertheless, the selective MAO-B inhibitor, selegiline, is marketed as an antidepressant but has to be used at high enough doses to inhibit MAO-A. The delivery of selegiline transdermally (Emsam) is designed to reduce the risk of tyramine-provoked reactions by minimizing inhibition of GI and hepatic MAO-A activity while allowing inhibition of central MAO-A and

MAO-B. Reversible inhibitors of MAO-A (RIMA) (e.g., moclobemide) are efficacious in depressive disorders and appear relatively safe with respect to tyramine sensitivity.

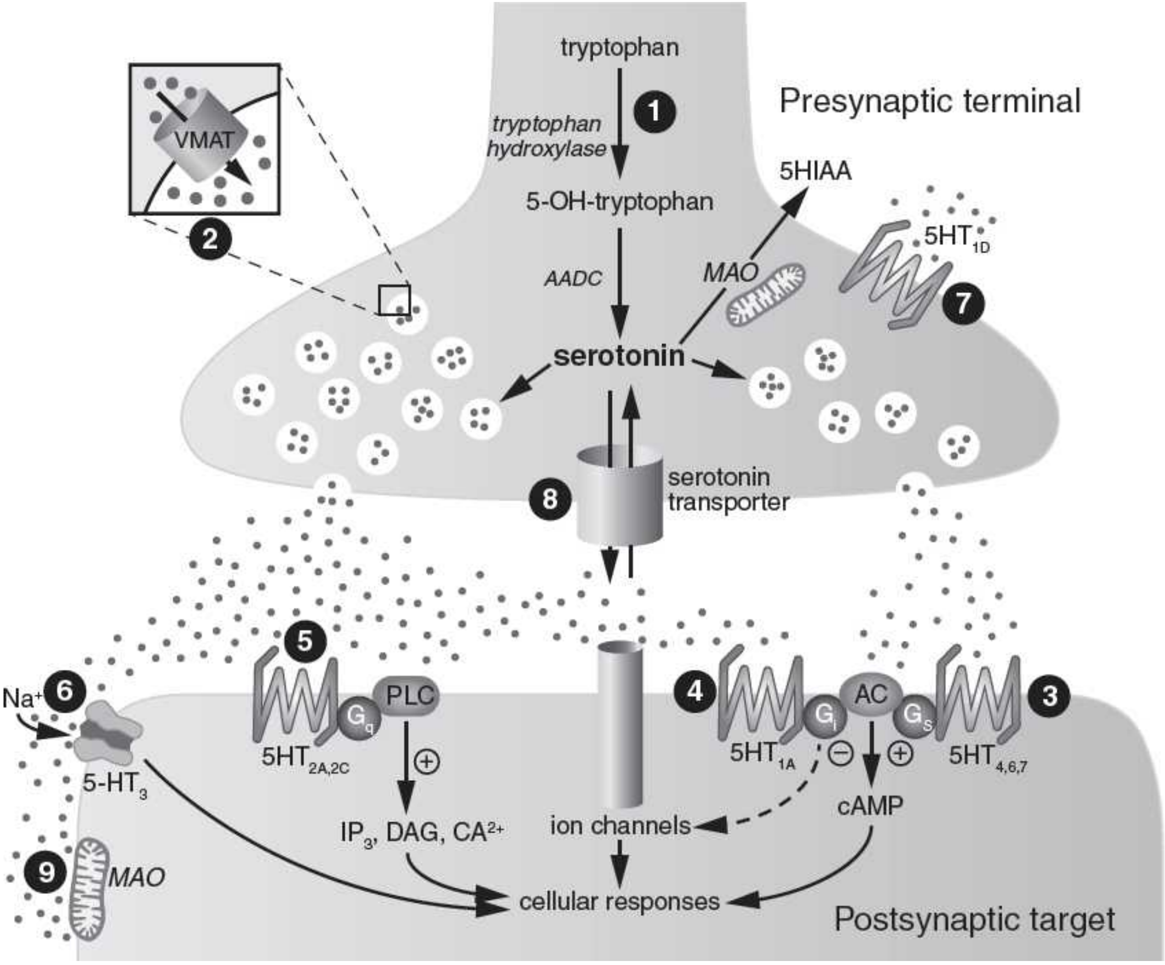
tryptophan
1
tryptophan hydroxylase
Presynaptic terminal
VMAT
2
5-OH-tryptophan
5HIAA
5HT1D
AADC
MAO
7
serotonin
8
serotonin transporter
5
PLC
Gq
5HT2A,2C
Na+
6
5-HT3
4
AC
Gi
Gs
3
5HT1A
5HT4,6,7
IP3, DAG, CA2+
ion channels
cAMP
9
MAO
cellular responses
Postsynaptic target

FIGURE 1.7–5. Schematic model of a serotonin synapse with sites of drug action. **Site 1:** Tryptophan is taken up into the neuron by an active transport mechanism and converted to 5-OH-tryptophan by the rate-limiting enzyme tryptophan hydroxylase; this enzyme can be effectively inhibited by p-chlorophenylalanine. AADC converts 5-OH-tryptophan to serotonin. **Site 2:** Serotonin is taken up and stored in vesicles by VMAT-2. Reserpine and tetrabenazine interfere with the uptake–storage mechanism of the amine granules, causing a marked depletion of serotonin. **Sites 3, 4, and 5:** Once released, serotonin can interact with as many as 14 different receptors. A number of G-protein–coupled serotonin receptors are located postsynaptically. Lurasidone (Latuda), a novel antipsychotic drug with possible procognitive properties, has high affinity for the 5HT7 receptor (**site 3**). The anxiolytics, buspirone (Buspar) and gepirone (Ariza), are 5HT1A receptors (**site 4**). Many antipsychotic drugs inhibit 5HT2A receptors (**site 5**). The hallucinogenic effects of drugs such as LSD, mescaline, and psilocybin are mediated by agonist interactions at the 5HT2A receptor. **Site 6:** In contrast to the other serotonin receptors, the 5HT3 receptor is a ligand-gated ion channel. Ondansetron is a 5HT3 antagonist that is an effective antiemetic and lacks parkinsonian side effects inherent in D2 antagonists like metoclopramide (Reglan) or prochlorperazine (Compazine). **Site 7:** Serotonin autoreceptors

(5-HT1B rodent and 5-HT1D human) modulate the stimulus-induced release of serotonin and are located on the presynaptic terminals. 5HT1A autoreceptors are located on the serotonin cell bodies and dendrites where they modulate impulse flow. Antimigraine triptan drugs, such as sumatriptan (Imitrex), are 5HT1D/1B agonists. **Site 8:** Termination of the synaptic actions of serotonin is accomplished by its reuptake into the neuron by the serotonin transporter. SSRIs and SNRIs are classes of antidepressant that inhibit the serotonin transporter and increase the synaptic actions of serotonin. The halogenated amphetamine, fenfluramine, causes release of serotonin by rapid reverse transport. **Site 9:** Synaptic actions of serotonin are also attenuated by its metabolism. Inhibitors of MAO-A block metabolism of serotonin and are used as antidepressants. Examples of irreversible inhibitors are isocarboxazid (Marplan), phenelzine (Nardil), and tranylcypromine (Parnate), whereas moclobemide inhibits MAO-A reversibly. 5HIAA, 5-hydroxyindoleacetic acid; 5HT1–7, subtypes of serotonin receptor; 5-OH-tryptophan, 5-hydroxytryptophan; AADC, aromatic amino acid decarboxylase; AC, adenylate cyclase; cAMP, cyclic AMP; DAG, diacylglycerol; G, G protein; IP3, inositol triphosphate; MAO, monoamine oxidase; SSRI, selective serotonin reuptake inhibitor; VMAT, vesicular monoamine transporter. (Modified from Nestler E, Hyman S, Holtzman D, Malenka R. *Molecular Neuropharmacology*. McGraw-Hill Companies, Inc; 2015.)

Receptors

Seven distinct families of 5HT receptors have been identified (5HT$_1$ to 5HT$_7$), and subpopulations have been described for several of these (Fig. 1.7–5). All of them except the ionotropic 5HT$_3$ receptor are G protein coupled. For serotonin GPCRs, three main types of primary coupling to G proteins have been described. The 5HT$_1$ and 5HT$_5$ receptors activate Gi/Go proteins, the 5HT$_2$ receptors activate Gq/11, and the 5HT$_4$, 5HT$_6$, and 5HT$_7$ activate Gs receptors.

5HT$_1$ Receptors. 5HT$_1$ family comprises five different gene products: 5HT$_{1A}$, 5HT$_{1B}$, 5HT$_{1D}$, 5HT$_{1E}$, and 5HT$_{1F}$ receptors. The 5HT$_{1A}$ receptor is the most widely distributed, and they are expressed on the soma and dendrites of all serotonin neurons, where they serve as autoreceptors to regulate cell excitability and the synthesis and release of 5HT. 5HT$_{1A}$ receptors are also located postsynaptically and on astrocytes and other glia. Partial 5HT$_{1A}$ agonists, such as buspirone (Buspar), are clinically useful anxiolytic drugs, and they augment the efficacy of antidepressants in the treatment of depression.

$5HT_{1B}$ and $5HT_{1D}$ receptors resemble each other in structure and distribution. They are predominantly located on the presynaptic terminals of serotonin neurons, where they function as autoreceptors to control local synthesis and release of 5HT. $5HT_{1B}$ receptors are also present on axon terminals of nonserotonergic neurons, where they act as heteroreceptors to control the presynaptic function of other transmitters. The serotonergic system has long been linked to migraine, and $5HT_{1B/D}$ agonists, such as sumatriptan (Imitrex) and other triptan drugs, are effective treatments for these types of headaches. The mechanism is not clear, however, and may be due either to vasoconstriction induced by activation of $5HT_{1B}$ and $5HT_{1D}$ receptors on cranial blood vessels or to modulation of activity in trigeminal afferents that carry pain information from the meninges. With the development of PET ligands for $5HT_{1B}$, these receptors have been implicated in the biology of psychiatric disorders, including PTSD and major depression. The antagonist of $5HT_{1B}$ and β-adrenergic receptors, pindolol, has been reported to augment the efficacy of antidepressant medications by promoting 5HT release, but this is not yet firmly established.

$5HT_{1E}$ is a highly expressed receptor in the human brain and is particularly enriched in the hippocampus and frontal cortex, making it a possible drug target for memory disorders, such as Alzheimer disease or temporal lobe epilepsy. Interestingly, this receptor subtype is absent in rats and mice. Research into the functions of the $5HT_{1E}$ receptor has been hampered by the lack of selective pharmacologic tools and specific antibodies.

Expression of $5HT_{1F}$ has been reported in various brain regions, in addition to the trigeminal ganglia and cerebral blood vessels. Triptans, which are used to treat migraine, have an appreciable affinity for $5HT_{1F}$ receptors, and the development of $5HT_{1F}$ agonists may provide promising new treatments for migraine.

$5HT_2$ Receptors. The $5HT_2$ family has three members, $5HT_{2A}$, $5HT_{2B}$, and $5HT_{2C}$, which share similarities in amino acid sequence and signaling properties. They all are coupled to phospholipase C (PLC) by Gq/11.

$5HT_{2A}$ is expressed widely throughout the brain, with especially high abundance in the cortex, where it is present in both interneurons and pyramidal neurons. The disturbance in sensory perception induced by hallucinogenic (psychedelic) drugs (e.g., D-lysergic acid diethylamide [LSD], mescaline, psilocybin, and 2,5-dimethoxy-4-methylamphetamine) is mediated by full or partial agonism at the $5HT_{2A}$ receptor. Second-generation antipsychotic drugs are characterized by their relatively higher potency as antagonists at the $5HT_{2A}$ site compared to D_2 receptors. There is some evidence that downregulation of $5HT_{2A}$ receptors and/or $5HT_{2C}$ receptors may be an adaptive mechanism contributing to the antidepressant effect of SSRIs, and in fact, blockade of $5HT_{2A}$ or $5HT_{2C}$ receptors enhances antidepressant-like effects of SSRIs in animal models. In agreement with this hypothesis is evidence of the effectiveness of atypical antipsychotic drugs given with SSRIs in treatment-resistant depression. Pimavanserin (Nuplazid) exhibits inverse agonist and antagonist activity at the $5HT_{2A}$ receptor and has been approved for the treatment of hallucinations and delusions associated with psychosis in patients suffering from Parkinson disease. Preliminary clinical data suggest that pimavanserin may also be beneficial for the short-term treatment of dementia-related psychosis. One useful effect of $5HT_{2A}$ receptor activation is a reduction in intraocular pressure, and this has led to the development of compounds that reduce pressure inside the eyes but without crossing the BBB and producing hallucinogenic side effects.

$5HT_{2B}$ receptors have a restricted expression in the CNS, which contrasts with their relative abundance in the periphery. Research on $5HT_{2B}$ has been limited due to the cardiotoxicity of $5HT_{2B}$ agonists, yet the addition of the atypical antipsychotic, aripiprazole (Abilify) which has a high affinity for $5HT_{2B}$ receptors, to antidepressants such as SSRIs have been reported to produce beneficial effects in treatment-resistant depressed patients.

The $5HT_{2C}$ receptor has a widespread and heterogeneous distribution in the brain. $5HT_{2C}$ receptors have been implicated in the mechanism underlying weight gain resulting from chronic administration of antipsychotic drugs, as those with the potential to produce weight gain, such as clozapine (Clozaril) and olanzapine (Zyprexa), have a greater affinity for the $5HT_{2C}$, compared with $5HT_{2A}$ receptor. Antipsychotic drugs that are more selective antagonists at the $5HT_{2A}$ subtype have less liability for weight gain. Agonists at the $5HT_{2C}$ receptor are being investigated as antiobesity agents. Interestingly, $5HT_{2C}$ in humans is located at the X chromosome, whereas all other 5HT receptors are autosomal. Thus, it is possible that polymorphisms at this receptor would affect the two sexes differently. $5HT_{2C}$ polymorphism appears to influence levels of circulating leptin, providing a potential mechanism underlying the genetic association of the $5HT_{2C}$ receptor with weight gain. $5HT_{2C}$ receptors are closely linked with the midbrain dopamine neurons, being able to regulate dopamine release in its terminal regions. For example, the firing rate of VTA dopamine neurons is inhibited or increased by $5HT_{2C}$ agonists or antagonists, respectively, such that $5HT_{2C}$ receptors regulate the release of dopamine in the nucleus accumbens. The basis for this interaction is the location of $5HT_{2C}$ receptors in the VTA on both dopamine and GABA neurons that innervate the nucleus accumbens. Because of the involvement of $5HT_{2C}$ receptors with the VTA-nucleus accumbens pathway, specific ligands for this receptor subtype are under development for the treatment of schizophrenia and substance abuse disorders.

$5HT_3$ Receptors. As mentioned above, the $5HT_3$ receptor is a ligand-gated ion channel, so it differs markedly in structure and mechanism from the other 5HT receptor subtypes. $5HT_3$ receptors are located in many brain areas, with the highest densities in the medulla and spinal cord. Some $5HT_3$ receptors can be located postsynaptically, but in many regions, they also are frequently expressed presynaptically, where their activation modulates the release of a variety of neurotransmitters. $5HT_3$ antagonists, such as ondansetron (Zofran), granisetron (Kytril), dolasetron (Anzemet), and palonosetron (Aloxi) are used clinically in the treatment of emesis associated with cancer chemotherapy and radiotherapy and to combat symptoms of irritable bowel syndrome. Owing to the unfavorable effects of $5HT_3$ agonists (e.g., nausea and anxiety), no clinical use of these is likely in the near future. In the periphery, $5HT_3$ receptors are prominently expressed in a variety of peripheral ganglia, and they have been proposed to modulate nociception and pain responses, as well as enteric and cardiovascular reflexes.

$5HT_4$ Receptors. These receptors are present in the brain and periphery. The $5HT_4$ agonist cisapride (Propulsid) is in clinical use as a gastroprokinetic agent to increase motility in the upper GI tract.

$5HT_5$ and $5HT_6$ Receptors. Although $5HT_5$ receptors have been detected in the brain, so far, no drug with selectivity for this subtype is used clinically. However, animal studies suggest that $5HT_{5A}$ receptor antagonists have the potential to treat cognitive impairments associated with schizophrenia. The $5HT_6$ receptor is expressed almost exclusively in the brain, and its abundance in limbic and cortical regions has stimulated considerable research interest, as it is blocked by some antipsychotic medications. Several selective $5HT_6$ ligands reached clinical trials to treat cognitive deficits in Alzheimer disease, but none achieved significant efficacy in the selected populations.

$5HT_7$ Receptors. The $5HT_7$ receptor has been identified in many brain regions, in addition to its presence in the periphery. There is evidence for the existence of both presynaptic and postsynaptic $5HT_7$ receptors. The function of this receptor has been most firmly linked with the control of circadian rhythm and thermoregulation. However, a major recent interest in $5HT_7$ receptors derives from their possible relevant role in cognitive processes, and the presence of $5HT_7$ receptors in the hippocampus and prefrontal cortex supports this. Furthermore, it is relevant that several antipsychotics and antidepressants have a high affinity for the $5HT_7$ receptor, which may contribute to their beneficial effects. $5HT_7$ has also been implicated in the modulation of structural plasticity in adolescent and mature brain circuits. Lurasidone (Latuda) is an atypical antipsychotic drug that is a particularly potent $5HT_7$ antagonist and is used to treat schizophrenia and depressive episodes associated with bipolar 1 disorder. Studies have even indicated that lurasidone can have procognitive properties in normal monkeys.

ACETYLCHOLINE

Anatomy and Function

In the brain, the axonal processes of cholinergic neurons either project to distant targets or contact cells locally within the same region (interneurons). Widely projecting cholinergic neurons originate from several clusters of cells within the basal forebrain complex and mesopontine complex in the upper brainstem. The basal forebrain population of cholinergic neurons comprises locations in the medial septal nucleus, the diagonal band, and the nucleus basalis of Meynert, and these neurons provide the vast majority of the cholinergic innervation to the telencephalon. Within the mesopontine complex, cholinergic neurons are located in the pedunculopontine nucleus and in the laterodorsal tegmental nucleus and innervate the thalamus, and midbrain regions (including the substantia nigra and VTA), in addition to descending innervations to regions such as the LC and dorsal raphe. While long projecting cholinergic neurons do innervate the striatum, the principal source of striatal acetylcholine is the cholinergic interneuron, which comprises only about 1% to 2% of all striatal cells, yet sends dense arbors of projections throughout the striatum. As would be predicted from their widespread distribution in the brain, central cholinergic neurons are involved in a wide variety of processes such as cognition, learning and memory, arousal, motor control, and CBF. In Alzheimer disease, there is significant degeneration of neurons in the nucleus basalis, leading to a substantial reduction in cortical cholinergic innervation. The extent of neuronal loss correlates with the degree of dementia, and the cholinergic deficit may contribute to the cognitive decline in this disorder, consistent with the beneficial effects of drugs that promote acetylcholine signaling.

In the ANS, acetylcholine is responsible for preganglionic transmission for both parasympathetic and sympathetic neurons and also at postganglionic synapses for the parasympathetic system. However, there are exceptions to this organization, such as the sympathetic innervation of adrenal glands and sweat glands that are provided by acetylcholine.

Life Cycle

The intraneuronal synthesis of acetylcholine involves only a single-step reaction (Fig. 1.7–6). The substrate choline is transported into cholinergic neurons by both high- and low-affinity mechanisms, and acetylcholine is formed when the enzyme choline acetyltransferase catalyzes the transfer of an acetyl group from cytoplasmic acetyl coenzyme A to choline. The rate-limiting step in acetylcholine synthesis is the high-affinity uptake of choline, and while it is not clear how this is regulated, it is known that increased neuronal activity enhances choline entry into neurons, and high-affinity uptake is blocked by hemicholinium. Once synthesized, acetylcholine is taken up into vesicles by the vesicular cholinergic transporter, a process that can be inhibited by vesamicol.

Several potent toxins affect the release of acetylcholine from vesicles into the synapse. Botulinum toxin A and tetanus toxin are derived from bacteria and prevent the release of acetylcholine, resulting in paralysis. However, if it is injected in small amounts, botulinum toxin can weaken a muscle locally, and this effect has been utilized in the treatment of spasms and dystonias and as a cosmetic application in smoothing facial wrinkles (Botox). Once released from the presynaptic neuron into the synapse, acetylcholine is not recaptured by an uptake mechanism, as occurs with the other biogenic amines. Instead, the termination of cholinergic transmission is primarily dependent on the breakdown of acetylcholine to choline and acetate, a reaction catalyzed by AChE, an enzyme that exists both in soluble and membrane-bound forms. Almost half the choline released by the metabolism of acetylcholine is taken back up into neurons for reuse. Inhibition of AChE activity by drugs enhances the stimulation of cholinergic receptors by acetylcholine. Neostigmine (Prostigmin) is a reversible inhibitor that does not cross the BBB and is used to improve muscle tone in cases of myasthenia gravis. As AChE inhibitors decrease fluid pressure in the eye, they are used to treat certain types of glaucoma. Centrally acting reversible inhibitors of AChE that penetrate the brain well after oral administration, such as donepezil (Aricept) and rivastigmine (Exelon), are used to improve the symptoms of Alzheimer disease and dementia associated with Parkinson disease. Irreversible inhibitors of AChE permanently inactivate the enzyme, such that restoration of activity requires synthesis of new protein, and such compounds are employed as insecticides and have been used as nerve gases (e.g., sarin, soman) in chemical warfare.

Receptors

Acetylcholine receptors fall into two major classes, muscarinic and nicotinic receptors, named for their respective naturally occurring agonists, muscarine and nicotine (Fig. 1.7–6).

Muscarinic Receptors. Five subtypes of muscarinic receptors (M_1 to M_5) have been cloned, and all are present in the brain. They are coupled to G proteins and either act directly on ion channels or are linked to a variety of second-messenger systems. M_1, M_3, and M_5 are Gq/11 coupled, whereas M_2 and M_4 are Gi/o coupled. Owing to the lack of specific activators or inhibitors of the subtypes, the roles of each subtype in the brain have not been well delineated. M_1 and M_3 are primarily located in the cerebral cortex and hippocampus and mediate some of the effects of acetylcholine on learning and memory. High densities of M_1 and M_4 are located in the striatum and presumably mediate cholinergic signaling in extrapyramidal motor circuits and response to rewards. M_2 receptors are concentrated in the basal forebrain, where they function as autoreceptors to control the synthesis and release of acetylcholine. The M_5 receptor is expressed at low levels in the brain and is present in substantia nigra and VTA. A recent short-term trial of the M1- and M4-selective muscarinic receptor agonist, xanomeline, in combination with a peripherally acting cholinergic antagonist to limit side effects, revealed a significant improvement in positive and negative symptoms in schizophrenia, which is an interesting development as xanomeline has no appreciable affinity for dopamine receptors. Nonspecific muscarinic agonists include some naturally occurring alkaloids such as pilocarpine and arecoline and synthetic compounds such as carbachol, oxotremorine, and bethanechol. Such drugs are not used to treat CNS disorders but are employed to reduce symptoms of glaucoma and nonobstructive urinary retention and to treat the dry mouth that occurs in Sjogren syndrome. Arecoline is the active ingredient in the areca nut and betel leaf that are chewed in many Asian cultures for the stimulating effect they induce. The most prominent adverse effects of muscarinic cholinergic agents (bradykinesia, GI distress, salivation, and sweating) are mediated by activation of peripheral M_1, M_2, and M_3 receptors.

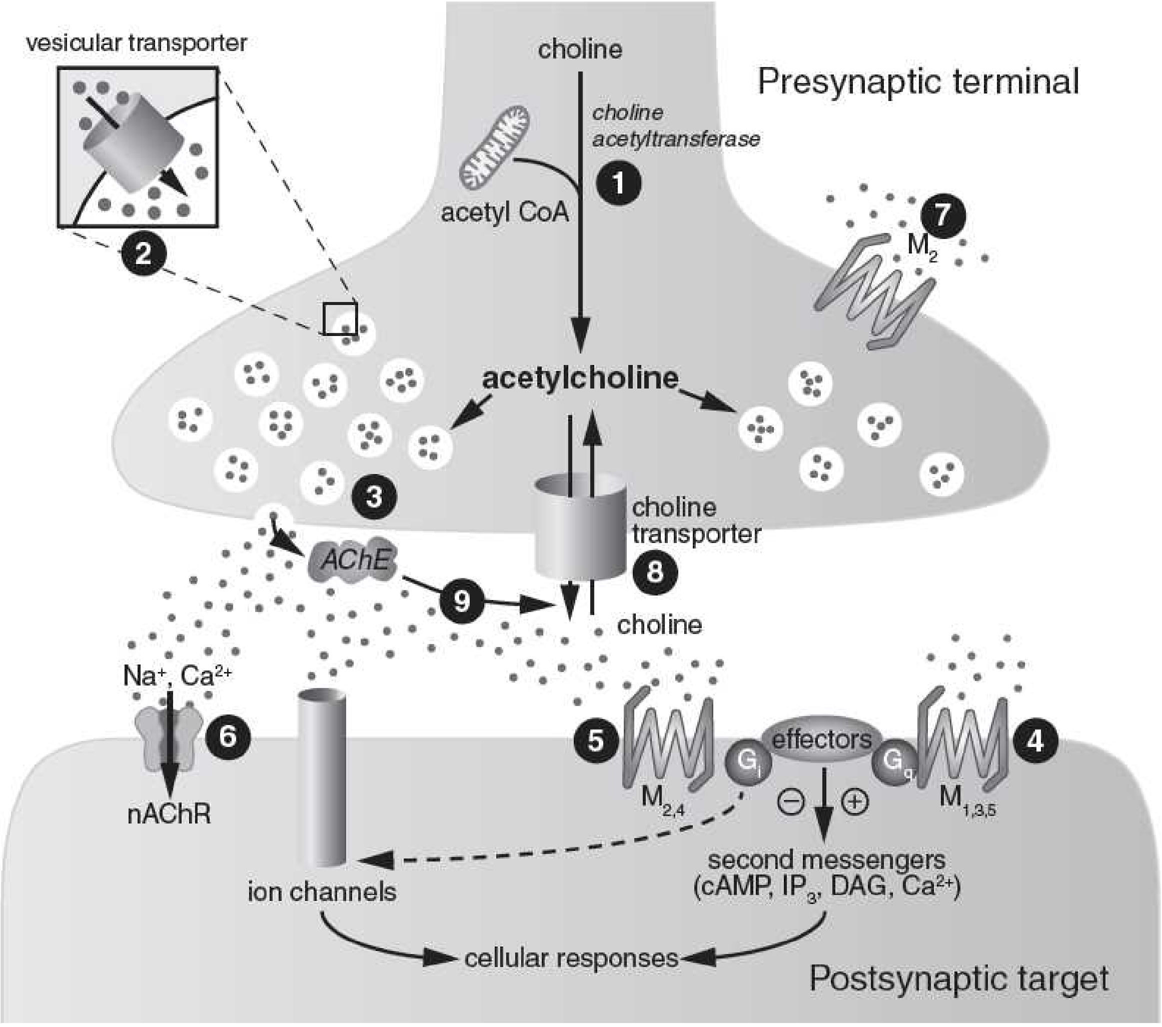
vesicular transporter
choline
Presynaptic terminal
choline acetyltransferase
1
acetyl CoA
7
M_2
2
acetylcholine
3
choline transporter
8
AChE
9
choline
Na^+, Ca^{2+}
6
nAChR
5
effectors
4
G_i
G_q
$M_{2,4}$
⊖
⊕
$M_{1,3,5}$
second messengers
(cAMP, IP_3, DAG, Ca^{2+})
ion channels
cellular responses
Postsynaptic target

FIGURE 1.7–6. Schematic model of an acetylcholine synapse with sites of drug action. **Site 1:** Choline is transported into the presynaptic terminal by an active uptake mechanism and converted to acetylcholine by a single enzymatic step. Acetyl coenzyme A required for acetylcholine synthesis is provided by presynaptic mitochondria. **Site 2:** Acetylcholine is taken up and stored inside vesicles. The vesicular transporter in cholinergic neurons is distinct from VMAT-2 present in other biogenic amine neurons and is blocked by vesamicol. **Site 3:** Release is promoted by black widow spider venom and blocked by botulinum toxin (peripheral synapse) and tetanus toxin (central synapse). **Sites 4 and 5:** After release, acetylcholine may interact with both G-protein–coupled muscarinic receptors (**sites 4 and 5**) on the postsynaptic cell. There are no muscarinic receptor subtype selective drugs, although some agents with partial selectivity are used experimentally. A clinically relevant nonselective muscarinic agonist is carbachol (Miostat), which is used in ophthalmologic procedures. A typical antagonist is atropine, which is also in ophthalmology and to reduce salivary secretions during surgery. A subtype of nicotinic receptors is activated by varenicline, which is used to treat nicotine dependence. Nicotinic receptors are blocked by the pharmacologic tools, α-bungarotoxin and hexamethonium. **Site 6:** Acetylcholine released from presynaptic terminals may also bind to nicotinic ligand-gated ion channels. A subtype of nicotinic receptors is activated by varenicline (Chantix), which is used to treat nicotine dependence. Nicotinic receptors are blocked by the pharmacologic tools, α-bungarotoxin, and hexamethonium. **Site 7:** Acetylcholine interacts with presynaptic muscarinic receptors to modulate subsequent release of transmitter. **Site 8:** In contrast to monoamine synapses, the plasma membrane transporter at a cholinergic synapse does not return the neurotransmitter to the presynaptic terminal but rather recycles precursor choline liberated after metabolism of acetylcholine. Hemicholinium-3 is a drug used to block choline transport experimentally. **Site 9:** Acetylcholine is metabolized by AChE, which present on both pre- and postsynaptic membranes, and serves to terminate its action. Examples of AChE inhibitors are physostigmine, which is used in eye drops to treat glaucoma, and donepezil (Aricept), which improves symptoms of Alzheimer disease. AChE, acetylcholinesterase; cAMP, cyclic AMP; DAG, diacylglycerol; G, G protein; IP3, inositol triphosphate; M1–5, subtypes of muscarinic receptor; MAO, monoamine oxidase; nAChR, nicotinic acetylcholine receptor; VMAT, vesicular monoamine transporter. (Modified from Nestler E, Hyman S, Holtzman D, Malenka R. *Molecular Neuropharmacology*. McGraw-Hill Companies, Inc; 2015.)

The prototypical muscarinic antagonists are atropine and scopolamine, which occur naturally in the plant deadly nightshade (Atropa belladonna). The translation of “belladonna” from Italian is “beautiful woman,” and this name was given because some women in Renaissance Italy applied belladonna berry juice to their eyes in order to dilate the pupils, giving them an alluring appearance, albeit with impaired vision. Ingestion of muscarinic antagonists leads to other effects resulting from inhibition of the parasympathetic nervous system, such as dry mouth, constipation, and urinary retention, in addition to central effects such as cognitive impairment and, in extreme cases, delirium. However, muscarinic antagonists have some therapeutic uses. For example, scopolamine has antiemetic activity and is used to treat nausea and vomiting caused by motion sickness. In addition, muscarinic antagonists such as benztropine (Cogentin) and trihexyphenidyl (Artane) are employed to treat the motor symptoms of Parkinson disease and are also sometimes helpful in ameliorating tardive dystonia, a possible side effect induced by long-term use of “typical” antipsychotic drugs. Many psychotropic drugs possess unwanted muscarinic antagonistic properties, including some TCAs and antipsychotic drugs. Some first-generation H_1 antagonist antihistamines are sold in over-the-counter hypnotics and antiallergy medications, and these drugs also have significant muscarinic antagonist actions that are responsible for side effects such as dry mouth, constipation, urinary retention, and even delirium to which geriatric patients are more susceptible.

Nicotinic Receptors. These are ligand-gated ion channels located at the neuromuscular junction, autonomic ganglia, adrenal medulla, brain, and spinal cord. Nicotinic receptors that are expressed in both neurons and muscle mediate rapid, excitatory signaling but differ in their subunit composition. Every nicotinic receptor consists of five subunit molecules organized around a central channel that opens to admit ions when the receptor is activated; they are arranged either in various homomeric (all one type of subunit) or heteromeric (at least one α and one β) groupings. The most common subtypes in the brain are the heteromeric α4β2-subtype and homomeric α7 receptors. The subunit combination affects the pharmacologic and functional properties of an individual complex, such as cation permeability and the rate of desensitization. These various nicotinic acetylcholine receptor subunits are categorized into functional classes: (1) neuronal (α_2–α_8 and β_2–β_4) and (2) skeletal (α_1, β_1, δ, γ, and ε). Neuronal nicotinic acetylcholine receptors are widely distributed in the brain, but they do not function only in the classical postsynaptic, directly excitatory manner of their muscle counterparts. Indeed, a large proportion of nicotinic binding sites are found on pre-, peri- and extrasynaptic sites in addition to the classic postsynaptic location. Nicotinic receptors have been implicated in several disorders and pathologic conditions. The most obvious role of neuronal nicotinic acetylcholine receptors is the mediation of reinforcing and addicting properties of nicotine in chronic tobacco and e-cigarette users, in addition to the symptoms of withdrawal experienced on cessation of use. The reinforcing effects of nicotine depend on its interaction with nicotinic receptors in the brain’s reward system, notably in releasing dopamine from neurons that originate in the VTA and innervate the nucleus accumbens. Mutations in one of the nicotinic receptor subunit genes contribute to a specific form of epilepsy, autosomal dominant frontal lobe epilepsy (ADNFLE). In addition, abnormal expression of α4β2 nAChRs has been implicated in several disorders, including autism spectrum disorders, nicotine addiction, and Parkinson disease. Alterations of α7 nAChR expression have been associated with various brain disorders, such as dementia and schizophrenia spectrum disorders. An interesting, distinctive property of ionotropic α7 receptors is their ability to activate downstream metabolic cascades, much like the metabotropic G-protein–coupled receptors. α_7-Nicotinic receptors have emerged as a potential therapeutic target for the treatment of neurocognitive dysfunctions in schizophrenia, Alzheimer disease, and Parkinson disease. Because of the evidence for their involvement in a number of CNS disorders, there is great interest in the future development of selective nicotinic receptor agonists as therapeutics. Although drugs such as hexamethonium and mecamylamine are useful pharmacologic tools used to block nicotinic receptors, there are relatively few nicotinic agonists or antagonists that are currently used therapeutically. One example is succinylcholine, a weak partial agonist at nicotinic receptors, which is used clinically during anesthesia as a paralytic agent. Varenicline (Chantix) is a partial agonist with selectivity at the α_4/β_2 nicotinic receptor, which is now widely marketed as medication to treat nicotine dependence and as an aid in the cessation of smoking. Galantamine (Razadyne) is an allosteric potentiator of several subtypes of neuronal nicotinic receptors and a weak competitive reversible inhibitor of AChE, which is also used in the treatment of Alzheimer disease.

HISTAMINE

Anatomy and Functions

Despite fulfilling a remarkably wide range of functions throughout the body, the role of histamine in the brain has not received as much attention as the other biogenic amines. Histamine-containing neuronal cell bodies are restricted to the tuberomammillary complex of the posterior hypothalamus. Like other biogenic amine neurons, the histaminergic system comprises long, highly divergent, mostly unmyelinated, varicose fibers that provide diffuse projections to many telencephalic, mesencephalic, and cerebellar structures. Individual neurons can innervate widely divergent areas but make few typical synaptic contacts. Functions closely associated with histaminergic projections are the regulation of sleep and arousal and the control of feeding and appetite. In the periphery, histamine released from mast cells plays an important role in allergic reactions, and histamine released from enterochromaffin-like cells and mast cells stimulates acid secretion in the stomach.

Life Cycle

Histamine is formed in the brain primarily by decarboxylation of histidine by the enzyme histidine decarboxylase (Fig. 1.7–7), although histidine is also a substrate for AADC, which is also involved in the synthesis of the catecholamines and serotonin. A mutation in the histidine decarboxylase gene has been implicated as a rare genetic cause of Tourette syndrome, and research with a mouse model bearing a knockout of this gene has shown it to recapitulate the core phenomenology of the disorder and has bolstered evidence for histamine-dopamine interactions in the basal ganglia to be a locus of pathology. The transporter protein responsible for concentrating histamine in synaptic vesicles has not been definitively established, although both reserpine and tetrabenazine, which interfere with VMAT-2 function, induce the release of brain histamine. In the brain, histamine is principally metabolized by histamine methyltransferase to the n-methylhistamine (tele-methylhistamine). In the periphery, histamine is also metabolized by diamine oxidase to imidazole acetaldehyde, an intermediate that is rapidly oxidized to imidazoleacetic acid. Unlike the other biogenic amines, there is no evidence for a specific high-affinity plasma membrane transporter to recapture and terminate the effects of released histamine in the brain. However, histamine is a substrate for some of the low-affinity transporters in the organic cation transporter (OCT) family and the low-affinity plasma membrane transporter.

Receptors

Four subtypes of histamine receptors have been identified (H_1 to H_4), and all belong to the G-protein–coupled superfamily of receptors (Fig. 1.7–7). In the brain, both H_1 and H_2 receptors are expressed in neurons postsynaptic to histamine terminals and also in glia. Blockade of H_1 receptors results in sedation, a side effect inherent in some antipsychotic drugs, such as clozapine (Clozaril), antidepressant drugs, such as doxepin (Deptran), and allergy medications, such as diphenhydramine (Benadryl) that inhibit H_1 receptors. In fact, the sedative property of doxepin is used therapeutically to alleviate the symptoms of primary insomnia. To overcome the sedating effect of H_1 antagonists in treating allergic reactions, newer nonsedating antihistamines, such as loratadine (Claritin), have been designed that do not cross the BBB. H_1 antagonists, such as meclizine (Antivert) and dimenhydrinate (Dramamine), are also employed in the treatment of motion sickness and vestibular disturbances such as Meniere disease.

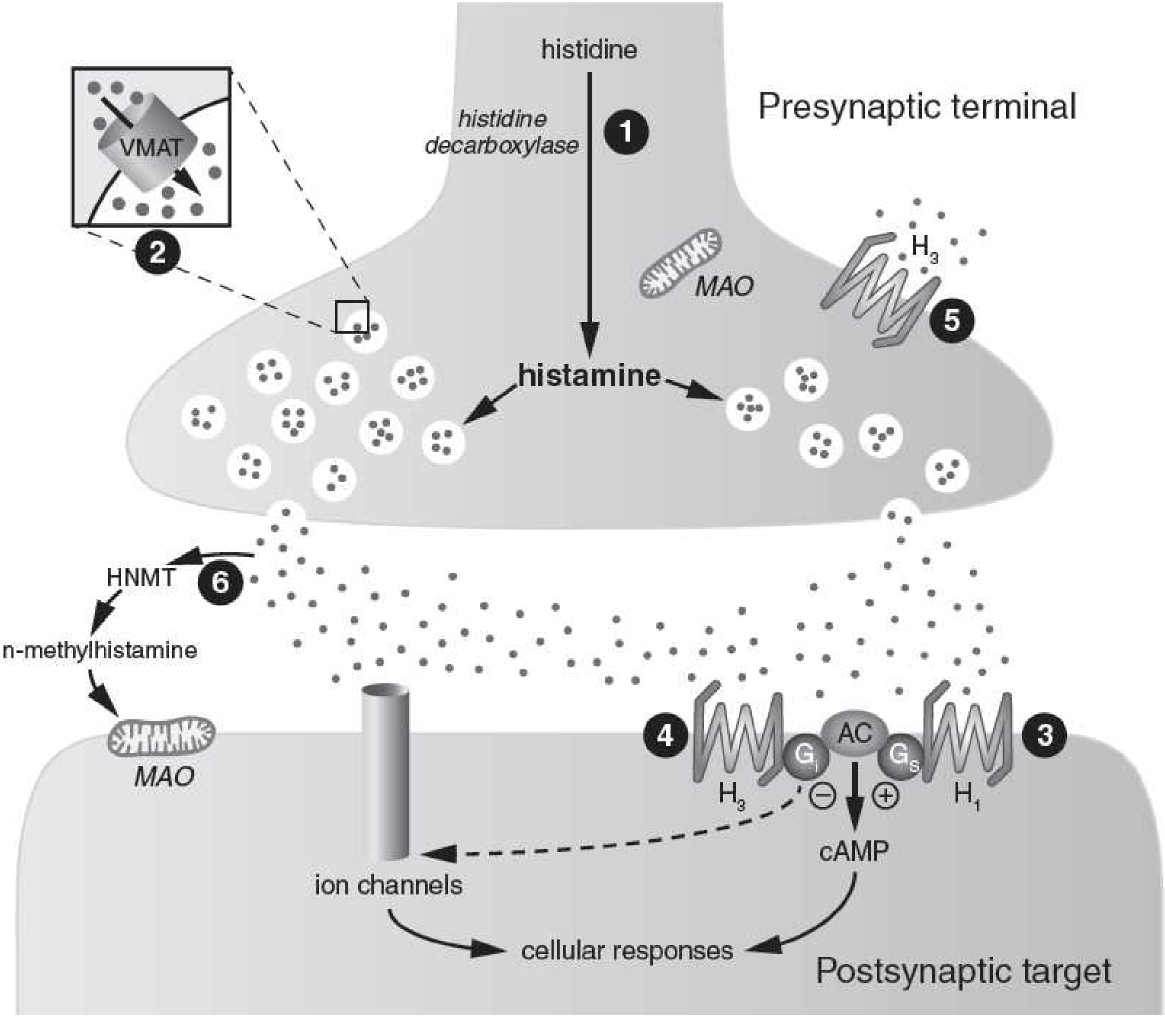
histidine
Presynaptic terminal
histidine decarboxylase
1
VMAT
2
MAO
H3
5
histamine
HNMT
6
n-methylhistamine
MAO
4
AC
3
Gi
Gs
H3
H1
cAMP
ion channels
cellular responses
Postsynaptic target

FIGURE 1.7–7. Schematic model of a central histaminergic synapse with site of drug action. **Site 1:** Histidine is taken up into histamine-containing neurons and converted to histamine by histidine decarboxylase. This enzyme is distinct from AADC that is used in the synthesis of other biogenic amines. No inhibitors of the histidine decarboxylase are used clinically. **Site 2:** Histamine is accumulated in storage vesicles by the action of VMAT. Reserpine interferes with vesicular storage of histamine. **Sites 3 and 4:** After release histamine interacts with postsynaptic receptors. Chlorpheniramine and diphenhydramine possess H1 antagonist properties and are used as antiallergy medications. H1 antagonists, such as meclizine (Bonine) and dimenhydrinate (Dramamine) have proved useful in the treatment of motion sickness and vestibular disturbances. **Site 5:** Activation of autoreceptors serves to regulate subsequent release of histamine. Pitolisant is a selective inverse agonist at the H3 receptor that increases histamine release, and is used to treat excessive diurnal sleepiness. There are no therapeutic products that act as selective antagonists for H3 receptors. **Site 6:** There is no recapture of synaptic histamine, and its actions are terminated by metabolism by histamine n-methyltransferase. No selective drugs are used therapeutically to inhibit this enzyme. The metabolite, n-methylhistamine is a substrate for MAO. AADC,

aromatic amino acid decarboxylase; AC, adenylate cyclase; cAMP, cyclic AMP; H1–3, subtypes of histamine receptor; HNMT, histamine N-methyltransferase; G, G protein; MAO, monoamine oxidase; VMAT, vesicular monoamine transporter. (Modified from Nestler E, Hyman S, Holtzman D, Malenka R. *Molecular Neuropharmacology*. McGraw-Hill Companies, Inc; 2015.)

H_2 receptors are expressed in neurons postsynaptic to histamine terminals. Compared with H_1 receptors, much less is known of the function of H_2 receptors, partly due to the paucity of specific agonists and antagonists that penetrate the BBB, but their involvement has been implicated in memory retention and nociception. H_2 antagonists that do not cross the BBB, such as cimetidine (Tagamet) and famotidine (Zantac 360, Pepcid), are used to lower the secretion of gastric acid and so reduce the pain and hasten the healing of gastric ulcers.

H_3 receptors are located presynaptically on histamine neurons, where they serve as autoreceptors to inhibit the release of histamine. In addition, H_3 receptors are present as heteroreceptors on terminals of other neurotransmitters to regulate their function. Recently, a selective inverse agonist at the H_3 receptor, pitolisant (Wakix), has been shown to be effective at counteracting excessive diurnal sleepiness in narcolepsy, Parkinson disease, and obstructive sleep apnea. Antagonists and inverse agonists for the H_3 receptor have potential as treatments for cognitive disorders.

The presence of H_4 receptors in the brain is not well established, although they are expressed in the periphery at several sites, including the spleen, bone marrow, and leucocytes, and are linked with the regulation of the immune system. Increasing attention has been directed toward the therapeutic modification of H_4 receptor signaling in disorders such as allergy, asthma, chronic pruritus, and autoimmune diseases.

TRACE AMINES

Several biologically active compounds with structural similarity to the classical biogenic amines are termed trace amines because of their low concentration in vertebrate brains and other tissues. They include phenylethylamine, tyramine, and tryptamine which are formed by decarboxylation of the amino acid precursors for the catecholamines and serotonin, phenylalanine/tyrosine, and tryptophan, respectively. However, trace amines are also abundant in many common foods and are also produced by microbiota. They diffuse across membranes easily, and no specific transport proteins or storage compartments have been identified for them. They are substrates for several oxidases, including MAO. Early research identified indirect sympathomimetic, amphetamine-like effects of phenylethylamine, albeit at higher concentrations than occurs endogenously. The identification of specific receptors for trace amine receptors galvanized interest in the function of trace amines and their possible role in the etiology and treatment of psychiatric conditions, especially as some psychotropic agents, such as amphetamines and MDMA, exhibit high affinity for the receptors. Humans possess six functional isoforms (TAAR1, TAAR2, TAAR5, TAAR6, TAAR8, and TAAR9). All except TAAR1 are expressed in olfactory epithelium neurons. However, TAAR1 is most thoroughly studied, and both central and peripheral roles have been linked with it. In the brain, TAAR1 has relatively high expression in VTA and dorsal raphe nuclei, regions where dopamine and serotonin cell bodies, respectively, are clustered. Generation of mice that overexpress TAAR1 or have the receptor knocked out have helped delineate possible contributions of trace amines to neuronal homeostasis. Such studies indicate that activation of the TAAR1 receptor inhibits dopaminergic and serotonin neurons, exemplified by their increased firing rate in TAAR1 knockout mice. There is evidence to suggest that the mechanism for TAAR1 modulation may involve the formation of a heterodimer between TAAR1 and presynaptic dopamine or serotonin autoreceptors that negatively regulate transmission. Although further research is needed to clarify the roles of trace amines in the brain, TAAR1 has been identified as a novel therapeutic target for disorders such as schizophrenia and the prevention of psychostimulant addiction relapse. Indeed, TAAR1 agonists are currently undergoing clinical trials for the treatment of schizophrenia, with available evidence indicating no abuse potential of TAAR1-directed agents.

FURTHER READINGS

Azizi SA. Monoamines: dopamine, norepinephrine, and serotonin, beyond modulation, “switches” that alter the state of target networks. *Neuroscientist*. 2022;28(2):121–143.

Baldan LC, Williams KA, Gallezot JD, et al. Histidine decarboxylase deficiency causes Tourette syndrome: parallel findings in humans and mice. *Neuron*. 2014;81:77–90.

Bernstein AI, Stout KA, Miller GW. The vesicular monoamine transporter 2: an underexplored pharmacological target. *Neurochem Int*. 2014;73:89–97.

Borroto-Escuela DO, Ambrogini P, Chruscicka B, et al. The role of central serotonin neurons and 5-HT heteroreceptor complexes in the pathophysiology of depression: a historical perspective and future prospects. *Int J Mol Sci*. 2021;22(4):1927.

Chestnykh DA, Amato D, Kornhuber J, Müller CP. Pharmacotherapy of schizophrenia: mechanisms of antipsychotic accumulation, therapeutic action and failure. *Behav Brain Res*. 2021;403:113144.

Cools R, Arnsten AFT. Neuromodulation of prefrontal cortex cognitive function in primates: the powerful roles of monoamines and acetylcholine. *Neuropsychopharmacology*. 2022;47(1):309–328.

Delcourte S, Etievant A, Haddjeri N. Role of central serotonin and noradrenaline interactions in the antidepressants' action: electrophysiological and neurochemical evidence. *Prog Brain Res*. 2021;259:7–81.

Fernstrom JD. Large neutral amino acids: dietary effects on brain neurochemistry and function. *Amino Acids*. 2013;45(3):419–430.

Gainetdinov RR, Hoener MC, Berry MD. Trace amines and their receptors. *Pharmacol Rev*. 2018;70:549–620.

Gershman SJ, Uchida N. Believing in dopamine. *Nat Rev Neurosci*. 2019;20(11):703–714.

Gomes I, Ayoub MA Fujita W, Jaeger WC, Pfleger KDG, Devi LA. G protein-coupled receptor heteromers. *Annu Rev Pharmacol Toxicol*. 2016;56:403–425.

Grace AA, Gomes FV. The circuitry of dopamine system regulation and its disruption in schizophrenia: insights into treatment and prevention. *Schizophr Bull*. 2019;45(1):148–157.

Liu C, Kaeser PS. Mechanisms and regulation of dopamine release. *Curr Opin Neurobiol*. 2019;57:46–53.

Liu Y, McNally GP. Dopamine and relapse to drug seeking. *J Neurochem*. 2021;157(5):1572–1584.

Lohani S, Martig AK, Underhill SM, et al. Burst activation of dopamine neurons produces prolonged post-burst availability of actively released dopamine. *Neuropsychopharmacology*. 2018;43(10):2083–2092.

Martinez D, Gil R, Slifstein M, et al. Alcohol dependence is associated with blunted dopamine transmission in the ventral striatum. *Biol Psychiatry*. 2005;58:779–786.

McCutcheon RA, Abi-Dargham A, Howes OD. Schizophrenia, dopamine and the striatum: from biology to symptoms. *Trends Neurosci*. 2019;42(3):205–220.

Meltzer HY. New trends in the treatment of schizophrenia. *CNS Neurol Disord Drug Targets*. 2017;16(8):900–906.

Moran SP, Maksymetz J, Conn PJ. Targeting muscarinic acetylcholine receptors for the treatment of psychiatric and neurological disorders. *Trends Pharmacol Sci*. 2019;40(12):1006–1020.

Nolan SO, Zachry JE, Johnson AR, Brady LJ, Siciliano CA, Calipari ES. Direct dopamine terminal regulation by local striatal microcircuitry. *J Neurochem*. 2020;155(5):475–493.

Ott T, Nieder A. Dopamine and cognitive control in prefrontal cortex. *Trends Cogn Sci*. 2019;23(3):213–234.

Pani L, Porcella A, Gessa GL. The role of stress in the pathophysiology of the dopaminergic system. *Mol Psychiatry*. 2000;5(1):14–21.

Panula P. Histamine receptors, agonists, and antagonists in health and disease. *Handb Clin Neurol*. 2021;180:377–387.

Papke RL, Horenstein NA. Therapeutic targeting of α 7 nicotinic acetylcholine receptors. *Pharmacol Rev*. 2021;73(3):1118–1149.

Peters KZ, Cheer JF, Tonini R. Modulating the neuromodulators: dopamine, serotonin, and the endocannabinoid system. *Trends Neurosci*. 2021;44(6):464–477.

Picciotto MR, Higley MJ, Mineur YS. Acetylcholine as a neuromodulator: cholinergic signaling shapes nervous system function and behavior. *Neuron*. 2012;76(1):116–129.

Scammell TE, Jackson AC, Franks NP, Wisden W, Dauvilliers Y. Histamine: neural circuits and new medications. *Sleep*. 2019;42(1):zsy183.

Stone EA, Lin Y, Sarfraz Y, Quartermain D. The role of the central noradrenergic system in behavioral inhibition. *Brain Res Rev*. 2011;67(1-2):193–208.

Suchting R, Tirumalajaru V, Gareeb R, et al. Revisiting monoamine oxidase inhibitors for the treatment of depressive disorders: a systematic review and network meta-analysis. *J Affect Disord*. 2021;282:1153–1160.

Tunbridge EM, Narajos M, Harrison CH, Beresford C, Cipriani A, Harrison PJ. Which dopamine polymorphisms are functional? Systematic review and meta-analysis of COMT, DAT, DBH, DDC, DRD1-5, MAOA, MAOB, TH, VMAT1, and VMAT2. *Biol Psychiatry*. 2019;86(8):608–620.

Waterhouse BD, Navarra RL. The locus coeruleus-norepinephrine system and sensory signal processing: a historical review and current perspectives. *Brain Res*. 2019;1709:1–15.

Wise RA, Robble MA. Dopamine and addiction. *Annu Rev Psychol*. 2020;71:79–106.

Wong DL, Tai TC, Wong-Faull DC, et al. Epinephrine: a short- and long-term regulator of stress and development of illness: a potential new role for epinephrine in stress. *Cell Mol Neurobiol*. 2012;32(5):737–748.

Xing B, Li YC, Gao WJ. Norepinephrine versus dopamine and their interaction in modulating synaptic function in the prefrontal cortex. *Brain Res*. 2016;1641(Pt B):217–233.

Zoli M, Pucci S, Vilella A, Gotti C. Neuronal and extraneuronal nicotinic acetylcholine receptors. *Curr Neuropharmacol*. 2018;16(4):338–349.

1.8 Neuropeptides: Biology, Regulation and Role in Neuropsychiatric Disorders

Michael J. Owen, B.Sc. (Hons.), M.B., Ch.B., Ph.D., F.R.C.Psych., Shaun P. Brothers, Ph.D., M.B.A., Larry J. Young, Ph.D., and Charles B.

NEMEROFF, M.D., PH.D.

Neuropeptides represent the most diverse class of signaling molecules in the central nervous system (CNS). Initially discovered for their role in the hypothalamic regulation of pituitary hormone secretion, the complex role of peptides in brain function has emerged over the last 40 years. Neuropeptides and their receptors are widely distributed throughout the CNS, where they exert an extraordinary array of direct or neuromodulatory effects, ranging from modulating neurotransmitter release and neuronal firing patterns to the regulation of emotionality and complex behaviors. Over 100 unique biologically active neuropeptides have been identified in the brain, a subset of which is presented in Table 1.8–1. Adding to the complexity of neuropeptide systems in the CNS, the actions of many peptides are mediated via multiple receptor subtypes localized in different brain regions. In fact, the discovery of new peptides and receptor subtypes has outpaced our understanding of the roles of these peptides in normal or aberrant CNS function. Pharmacologic, molecular, and genetic approaches are now leading the way in our understanding of the contribution of neuropeptide systems to psychiatric disorders. The first half of this chapter provides a general overview of neuropeptide biology, including their biosynthesis, distribution, and signaling. The second half of the chapter provides specific, detailed information for a select group of neuropeptides with special relevance to behavior and psychiatric disorders.

Table 1.8–1.
Selected Neuropeptide Transmitters

Adrenocorticotropin hormone (ACTH)
Angiotensin
Apelin
Atrial natriuretic peptide
Bombesin
Calcitonin
Calcitonin gene–related peptide (CGRP)
Cocaine- and amphetamine-regulated transcript (CART)
Cholecystokinin (CCK)
Corticotropin-releasing factor (CRF; corticotropin-releasing hormone; CRH)
Dynorphin
β-Endorphin
leu-enkephalin
met-enkephalin
Galanin
Gastrin
Ghrelin
Gonadotropin-releasing hormone (GnRH)
Growth hormone
Growth hormone–releasing hormone (GHRH; GRF)
Insulin
Kisspeptin
Motilin
Neurokinin A (formerly Substance K)
Neuropeptide S
Neuropeptide Y (NPY)
Neurotensin
Neuromedin-N
Nociceptin
Orexin (also known as hypocretin)
Oxytocin
Pancreatic polypeptide
Prolactin
Secretin

Somatostatin (SS; SRIF)
Substance P
Thyrotropin-releasing hormone (TRH)
Urocortin (1,2, and 3)
Vasoactive intestinal polypeptide (VIP)
Vasopressin (AVP; ADH)

By definition, a neuropeptide is a chain of two or more amino acids linked by peptide bonds and differs from other proteins only in the length of the amino acid chain. Neuropeptides range in length from 2 (e.g., carnosine and anserine) to over 40 amino acids (e.g., corticotropin-releasing factor [CRF] and urocortin). By convention, peptides greater than 90 amino acids in length (about 10,000 molecular weight) are considered proteins. The neuropeptides highlighted in detail in this chapter include thyrotropin-releasing hormone (TRH), CRF, oxytocin (OT), arginine vasopressin (AVP), neurotensin (NT), and neuropeptide Y (NPY) with brief summaries of relevant data for cholecystokinin (CCK), substance P (SP), galanin, and orexins. The structures of these highlighted neuropeptides are illustrated in Table 1.8–2 and are written using the single-letter amino acid code by convention from the amino terminus (NH_2^-) beginning on the left to the carboxyl terminus (-COOH) on the right. A detailed discussion of all neuropeptide systems of potential relevance to psychiatry is beyond the scope of this chapter, and other signaling peptides, including the opioid peptides, growth factors, and cytokines, are considered elsewhere in this volume. TRH and CRF are hypothalamic hypophysiotropic hormones, which stimulate the release of thyroid-stimulating hormone (TSH) and adrenocorticotropic hormone (ACTH), respectively, from the anterior pituitary (adenohypophysis) into the bloodstream. OT and AVP are neurohypophysial peptides that are released directly into the bloodstream from the posterior pituitary (neurohypophysis) under specific physiologic conditions. However, all of these above-mentioned peptides, including NT, also function in the CNS as neurotransmitters, neuromodulators, or neurohormones in ways that are often quite distinct and independent from their effects on peripheral endocrine systems. Indeed, many neuropeptides appear to have evolved two distinct activities, one in the CNS and a sometimes highly different one in the periphery. OT and AVP, for example, are thought to contribute to emotionality in the CNS and stimulate uterine contraction peripherally during pregnancy. Perhaps more confounding to being able to achieve clear research outcomes is that these seemingly disparate systems may even have some overlap, such as contributing to postpartum nurturant behavior in mothers. Another example is with μ-opioid receptors in the CNS that regulate nociception and reward and peripherally regulate respiration and gut motility. For this reason, opiate drugs have the side effects of respiratory depression and constipation.

Table 1.8–2.
Selected Neuropeptide Structures

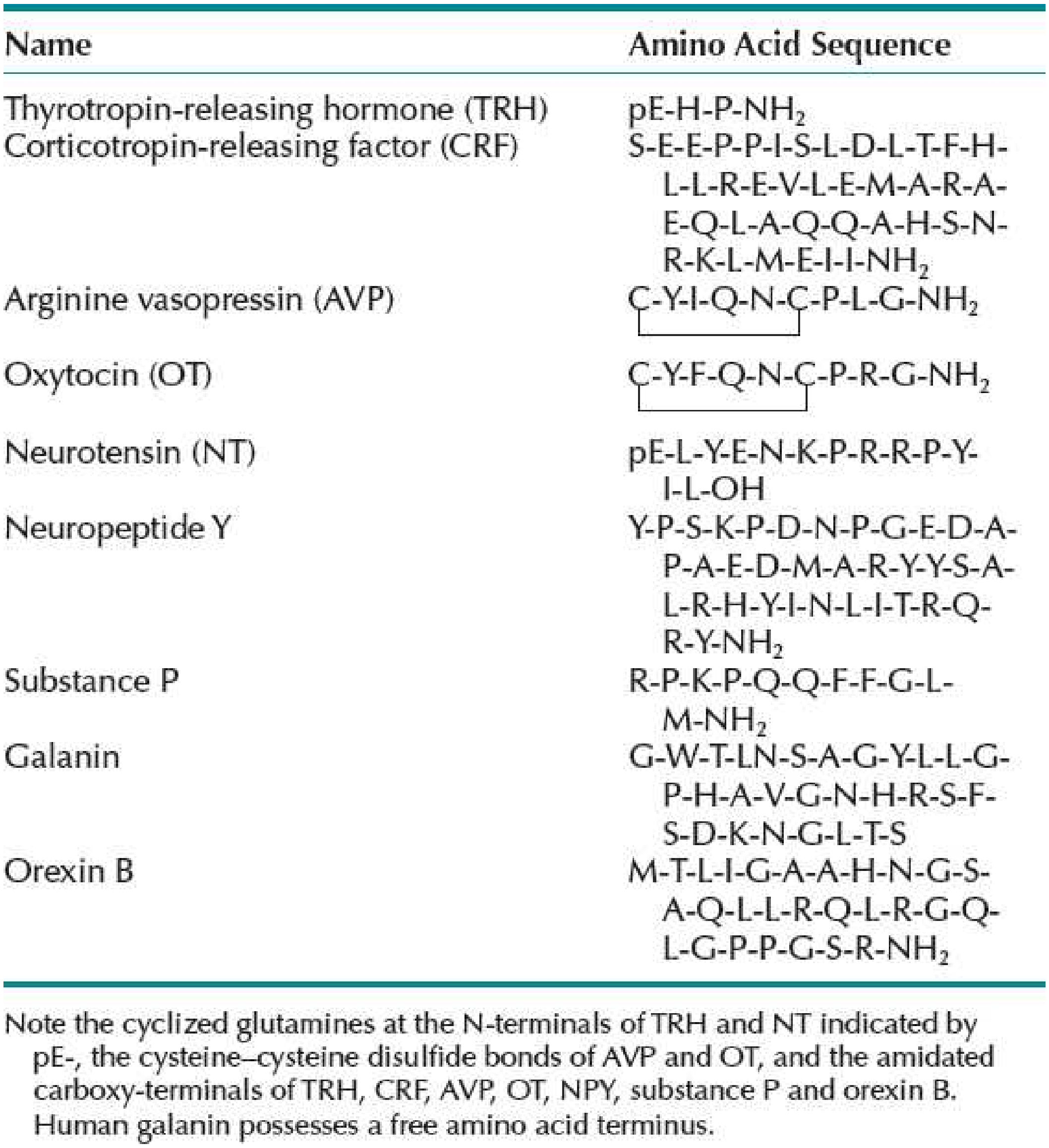

Name	Amino Acid Sequence
Thyrotropin-releasing hormone (TRH)	pE-H-P-NH_2
Corticotropin-releasing factor (CRF)	S-E-E-P-P-I-S-L-D-L-T-F-H-L-L-R-E-V-L-E-M-A-R-A-E-Q-L-A-Q-Q-A-H-S-N-R-K-L-M-E-I-I-NH_2
Arginine vasopressin (AVP)	C-Y-I-Q-N-C-P-L-G-NH_2
Oxytocin (OT)	C-Y-F-Q-N-C-P-R-G-NH_2
Neurotensin (NT)	pE-L-Y-E-N-K-P-R-R-P-Y-I-L-OH
Neuropeptide Y	Y-P-S-K-P-D-N-P-G-E-D-A-P-A-E-D-M-A-R-Y-Y-S-A-L-R-H-Y-I-N-L-I-T-R-Q-R-Y-NH_2
Substance P	R-P-K-P-Q-Q-F-F-G-L-M-NH_2
Galanin	G-W-T-LN-S-A-G-Y-L-L-G-P-H-A-V-G-N-H-R-S-F-S-D-K-N-G-L-T-S
Orexin B	M-T-L-I-G-A-A-H-N-G-S-A-Q-L-L-R-Q-L-R-G-Q-L-G-P-P-G-S-R-NH_2

Note the cyclized glutamines at the N-terminals of TRH and NT indicated by pE-, the cysteine–cysteine disulfide bonds of AVP and OT, and the amidated carboxy-terminals of TRH, CRF, AVP, OT, NPY, substance P and orexin B. Human galanin possesses a free amino acid terminus.

Neuropeptides have been implicated in the regulation of a variety of behavioral and physiologic processes, including thermoregulation, food and water consumption, sex, sleep, locomotion, learning and memory, responses to stress and pain, emotion, and social cognition, to name a few. Involvement in such behavioral processes suggests the likelihood that neuropeptidergic systems contribute to the symptoms and behaviors exhibited in the major psychiatric disorders, including autism, schizophrenia, mood disorders, and anxiety.

INVESTIGATING NEUROPEPTIDE FUNCTION

The roles of neuropeptides in CNS function and behavior have been examined using a multitude of experimental techniques. The levels of analysis include molecular structure and biosynthesis of the peptide and its receptor(s), the neuroanatomical localization of the peptide and its receptor(s), the regulation of the expression and release of the peptide, and finally, the behavioral effects of the peptide. Much of the information on neuropeptide biology has been derived from laboratory animal studies; however, there is an ever-increasing database on the localization, activity, and potential psychiatric relevance of several neuropeptides systems in humans.

Most neuropeptide structures have been identified based on chemical analysis of purified biologically active peptides leading ultimately to the cloning and characterization of the genes encoding them. Characterization of the gene structure of peptides and their receptors has allowed for an understanding of the molecular regulation of these systems, and their chromosomal localization is useful in genetic studies examining the potential roles of these genes in the pathogenesis of psychiatric disorders. Structural characterization permits the production of immunologic and molecular probes that are useful in determining peptide distribution and regulation in the brain. Quantitative enzyme-linked immunosorbent assay (ELISA) on microdissected brain regions or immunocytochemistry on brain sections are typically used to characterize the regional distribution of peptides throughout the brain. Both techniques use specific antibodies generated against the neuropeptide to detect its presence. Immunocytochemistry allows researchers to visualize the precise cellular localization of peptide-synthesizing cells as well as their projections throughout the brain, though the technique is generally not quantitative. Using molecular probes homologous to the mRNA encoding the peptides or receptor, *in situ* hybridization can be used to localize and quantify gene expression in brain sections. This is a powerful technique for examining the molecular regulation of neuropeptide synthesis with a precise neuroanatomical resolution, which is impossible for other classes of nonpeptide neurotransmitters that are not derived directly from the translation of messenger RNAs, such as dopamine, serotonin, and norepinephrine. In addition to immunocytochemistry and *in situ* hybridization, receptor autoradiography on brain sections (Fig. 1.8–1) or “grind and bind” receptor binding assays on microdissected brain tissue are frequently used to localize and quantify neuropeptide receptors in specific regions of the brain. Receptor autoradiography involves allowing a radiolabeled ligand to bind the receptor on a thin slice of tissue and then detecting the bound by visualizing it on x-ray film or other means. Other molecular techniques, such as quantitative real-time polymerase chain reaction (RT-PCR) and RNA sequencing (transcriptomics), are also commonly used to measure neuropeptide and receptor expression and regulation by quantifying the mRNAs encoding the peptide or receptor. However, quantification of neuropeptide gene expression or immunoreactivity within a cell or tissue homogenate does not provide information on neuropeptide release. *In vivo* microdialysis, in which peptides concentrated in the extracellular fluid and presumably originating from neuronal release, are collected at sequential time intervals using dialysis probes implanted into specific brain regions, may be used to quantify neuropeptide release under defined physiologic or behavioral circumstances.

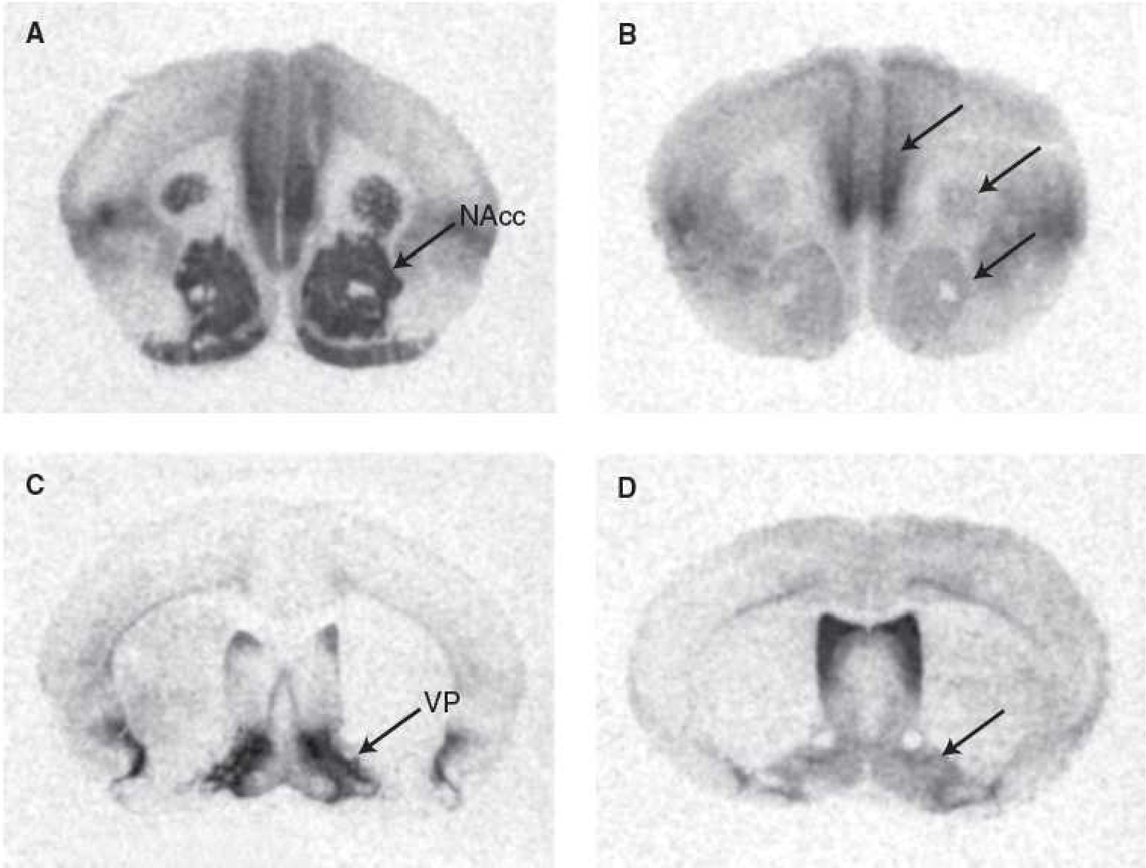

FIGURE 1.8–1. Oxytocin and vasopressin receptor distribution patterns in the brain are associated with social behavior. The **upper panels** depict receptor autoradiograms illustrating the localization of oxytocin receptor binding in the highly social and monogamous prairie vole (**A**) and the asocial montane vole (**B**). The **lower panels** illustrate vasopressin V1a binding in the monogamous prairie vole (**C**) and nonmonogamous montane vole (**D**). Note the high density of oxytocin receptor in the nucleus accumbens (NAcc) and V1a receptor binding in the ventral pallidum (VP) of the prairie vole but not the montane vole. These receptor populations are critical for social attachment in monogamous rodents. (Adapted by permission from Young LJ, Wang ZX. The neurobiology of the social bond. *Nat Neurosci* 2004;7:1048–1054. Copyright © 2004. Nature Publishing Group.)

In general, the behavioral effects of neuropeptides are initially investigated by infusions of the peptide directly into the brain. Unlike many nonpeptide neurotransmitters, most neuropeptides do not penetrate the blood–brain barrier in amounts sufficient enough to produce CNS effects. Furthermore, serum and tissue enzymes tend to degrade the peptides before they reach their target sites. The degradation is usually the result of the cleavage of specific amino acid sequences targeted by a specific peptidase designed for that purpose. Thus intracerebroventricular (ICV) or site-specific infusions of peptide in animal models are generally required to probe for behavioral effects of peptides. However, there are some examples of the delivery of neuropeptides via intranasal infusions in human subjects, which in some cases appears to permit access of the peptide to the brain (e.g., OT), though this remains controversial.

In many cases, the interpretation of neuropeptide infusion studies is complicated because of the considerable cross-talk between specific neuropeptides and several heterologous receptors. For example, OT and vasopressin differ by only 2 of 9 amino acids and both peptides cross-react to some degree with both receptor types. In some cases, highly selective peptide receptor synthetic agonists or antagonists have been developed that allow researchers to examine the role of specific neuropeptide receptors in the regulation of behavior or physiologic processes. In addition, transgenic and knockout mouse approaches are commonly used to investigate neuropeptide function. For example, mutant mouse strains with null mutations (knockouts) in either the peptide gene or the corresponding receptor have been developed and have proven quite useful for exploring the role of neuropeptides in behavioral processes. Small interfering RNA (siRNA) or short hairpin RNA (shRNA) techniques, which lead to the selective degradation of the targeted mRNA in specific brain regions, have been used to examine the function of specific neuropeptide-producing neuronal populations. The CRISPR/Cas system that has been used for gene editing holds the potential to begin to separate spatiotemporal effects of up-/downregulating neuropeptides and genes. Similar efforts with viruses to selectively overexpress or knock down neuropeptide or receptor RNA in specific microinjected brain regions have been pursued as well. Transgenic and viral vector approaches are increasingly being used to interrogate the activity and function of neuropeptide or receptor-expressing neurons using optogenetics, chemogenetics, and fiber photometry. These approaches are made possible by genetically engineered rodents, typically mice, that express Cre recombinase in tandem with the peptide or peptide receptor transcript. When combined with Cre-dependent viral vectors that express light-sensitive opsins, novel “designer” receptors, or intracellular calcium indicators (e.g., GCaMP), researchers can exquisitely control or monitor activity in specific populations of peptide or receptor neurons. This approach must be combined with pharmacologic manipulations to dissect the functions of the neurons from neuropeptide signaling.

As noted above, one of the greatest impediments to exploring the role and potential therapeutic value of neuropeptides is the inability of the peptides or their peptidergic agonists/antagonists to penetrate the blood–brain barrier. Thus, the behavioral effects of most peptides in humans are largely uninvestigated, with the exception of a few studies utilizing intranasal delivery. However, in some instances, small-molecule, nonpeptide agonists/antagonists have been developed that can be administered peripherally and permeate the blood–brain barrier in sufficient quantities to affect receptor activation. Other approaches include pharmacologically evoking endogenous neuropeptide release using a brain-penetrant drug.

Humans are less-than-ideal subjects for neuropeptide research for several reasons. First, although blood samples to determine plasma hormone concentrations are relatively easy to obtain, the independent regulation of peripheral and CNS peptide release, the high concentration of plasma peptidases, and the blood–brain barrier make it virtually impossible to infer CNS peptide physiology from plasma hormone concentrations. Also, the use of biopsy to directly assess tissue peptide concentrations is not ideal because it is not routinely repeatable, is limited to superficial structures, and suffers from potential morbidity. In contrast, however, cerebrospinal fluid (CSF) has been shown to reflect extracellular fluid concentrations of several transmitter substances, is in direct contact with the CNS, is screened from peripheral serum sources by the blood–brain barrier, and may be sampled across time. The limitations of human CSF studies include a lack of information about the regional CNS source of any changes in peptide concentration detected, the use of lumbar CSF, which is somewhat removed from higher forebrain CNS sources of peptides and subject to spinal cord peptide contributions, the potentially confounding effects of previous drug treatments or disease episodes, and the reluctance of research subjects to consent to this procedure. Postmortem tissue studies of neuropeptide concentration changes in psychiatric disease have been informative in many cases, but interpretation must include consideration of postmortem delay, previous drug treatment, and coexisting illnesses. Most of the data on alterations in CSF or tissue concentrations of neurotransmitters have been derived from comparisons between diagnostically defined psychiatric groups and control groups. However, the controls may be so-called “neurologic or psychiatric controls,” not healthy volunteers, and the accuracy and consistency of the diagnoses may be less than optimal. In addition, the etiology of a syndromal diagnosis may differ among subjects in the same diagnostic group. Even after matching for age, gender, or other demographic variables, heterogeneity among human research populations results in individual variations of absolute peptide values that are often quite wide. Such variances severely reduce the power of group comparisons to detect alterations in peptide concentrations. The use of pre- and posttreatment CSF samples, or of samples obtained during the active disease state versus when the patient is in remission, addresses some of the serious limitations in study design. For such progressive diseases as schizophrenia or Alzheimer disease (AD), serial CSF samples may be a valuable indicator of disease progression or response to treatment. Even with these constraints, significant progress has been made in describing the effects of various psychiatric disease states on neuropeptide systems in the CNS.

BIOSYNTHESIS

Unlike other neurotransmitters, the biosynthesis of a neuropeptide involves transcription of an mRNA from a specific gene, translation of a polypeptide preprohormone encoded by that mRNA, and then posttranslational processing involving proteolytic cleavage of the preprohormone to yield the active neuropeptide. Over the past 40 years, the gene structures and biosynthetic pathways of most neuropeptides have been elucidated. The gene structure of selected neuropeptides is illustrated in Figure 1.8–2. Neuropeptide genes are generally composed of multiple exons that encode a protein preprohormone. The N-terminus of the preprohormone contains a signal peptide (SP) sequence, which guides the growing polypeptide to the rough endoplasmic reticulum (RER) membrane. The single preprohormone molecule often contains the sequences of multiple peptides that are subsequently separated by proteolytic cleavage by specific enzymes. For example, translation of the gene encoding NT yields a preprohormone, which upon enzymatic cleavage produces both NT and neuromedin-N. Other neuropeptide genes, such as the TRH gene, encode multiple copies of the peptide sequence or, as in the case of OT and vasopressin, also encode other proteins essential in the posttranslational processing and transport of the neuropeptide. The neuroanatomical localization and abundance of neuropeptides are determined primarily by the region-specific expression and regulation of their gene. Each neuropeptide gene is expressed in well-defined populations of neurons within the brain. The precise neuroanatomical pattern of peptide hormone gene expression is determined by regulatory DNA sequences surrounding the gene. This has been elegantly demonstrated for the OT gene. Viral vector constructs containing the 5′ regulatory region of the OT gene reliably direct transgene expression in OT neurons, permitting opto- or chemogenetic manipulation of the activity of these neurons and OT release in behavioral contexts. Transcription factor binding sites located in the promoter of the gene are also involved in the physiologic regulation of peptide gene expression. Analysis of promoter sequences of peptide genes has provided insights into the molecular regulation of peptide biosynthesis.

The mRNA encoding the preprohormone is translated by ribosomes associated with the RER, and the growing polypeptide is translated with the SP anchored in the RER membrane within the cisternae. Once translated, the SP of the *pre*prohormone is cleaved by a signal endopeptidase, freeing the *pro*hormone polypeptide. The prohormone is then shuttled to the Golgi apparatus, where packaging into granules or vesicles occurs. Proteolytic cleavage of the prohormone into the biologically active neuropeptide begins in the Golgi and continues in the granules. Production of biologically active neuropeptides from prohormones begins with cleavage at specific sites adjacent to the neuropeptide sequence by specific endopeptidases known as prohormone convertases. Prohormone convertases generally cleave at pairs of basic amino acids (e.g., Lys-Arg, Lys-Lys, and Arg-Arg), flanking the neuropeptide sequence. There are at least seven prohormone convertases, each with unique properties, including substrate specificity and neuroendocrine distribution. Prohormone convertases are co-packaged with the prohormones in the granules at the Golgi apparatus. The substrate specificity and differential distribution of the prohormone convertases provide a mechanism by which different neuropeptides encoded by a single prohormone can be differentially produced in an active form. Following endopeptidase cleavage, the peptide fragments are subjected to exoproteolysis by carboxypeptidases and/or aminopeptidases in order to remove the residual basic residues on the carboxy- or amino-terminus of the peptide fragments. The synthesis and processing of neuropeptides are illustrated in Figure 1.8–3.

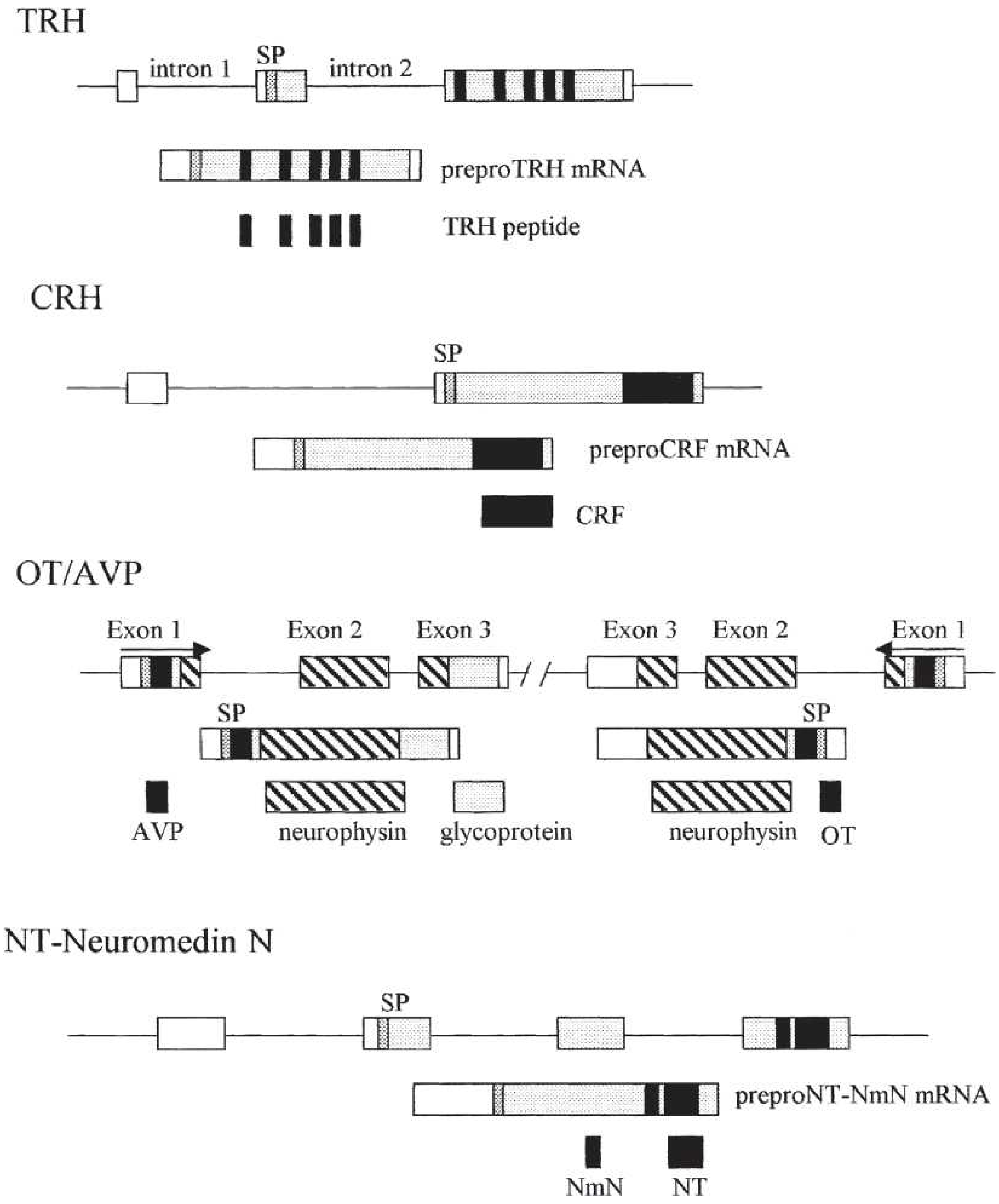
TRH
intron 1
SP
intron 2
preproTRH mRNA
TRH peptide
CRH
SP
preproCRF mRNA
CRF
OT/AVP
Exon 1
Exon 2
Exon 3
Exon 3
Exon 2
Exon 1
SP
SP
AVP
neurophysin
glycoprotein
neurophysin
OT
NT-Neuromedin N
SP
preproNT-NmN mRNA
NmN
NT

FIGURE 1.8–2. Schematics illustrating the gene structure, preprohormone mRNA, and processed neuropeptides of TRH, CRF, OT, AVP, and NT. Boxed regions indicate the locations of the exons in the respective genes. Shaded or hatched regions indicate coding regions. Each preprohormone begins with a signal peptide sequence (SP). *Black boxes* indicate the location of the sequences encoding the neuropeptide.

Although many known peptides are complete and biologically active when cleaved from the prohormone, many others are subjected to additional posttranslational processing. Certain peptides have a metabolically blocked carboxy-terminus that is often amidated. A glycine residue in the prohormone sequence often acts as the amide donor and, in the case of TRH, is cleaved by a monooxygenase that is contained in secretory granules. TRH is further processed on the N-terminus, where glutamine is cyclized by a glutamyl cyclase to yield a pyroglutamyl moiety. These alterations are usually effective in reducing susceptibility to degradation and are often required for biologic activity, as is the case for TRH, which is rendered inactive when the C-terminal amide is removed by proline endopeptidase to generate the free-acid structure. Other posttranslational processing events for active peptides include glycosylation, phosphorylation, and the formation of disulfide bonds, which are often required for either biologic activity or transport. Several neuropeptides, including OT, vasopressin, and orexin A, contain one or more cysteine–cysteine disulfide bonds, resulting in cyclic peptide structures (see Table 1.8–2). Finally, neuropeptide diversity can occur through alternative splicing of the primary RNA to generate different mRNAs and, ultimately, peptides with markedly differing actions (e.g., calcitonin and calcitonin gene–related peptide [CGRP]).

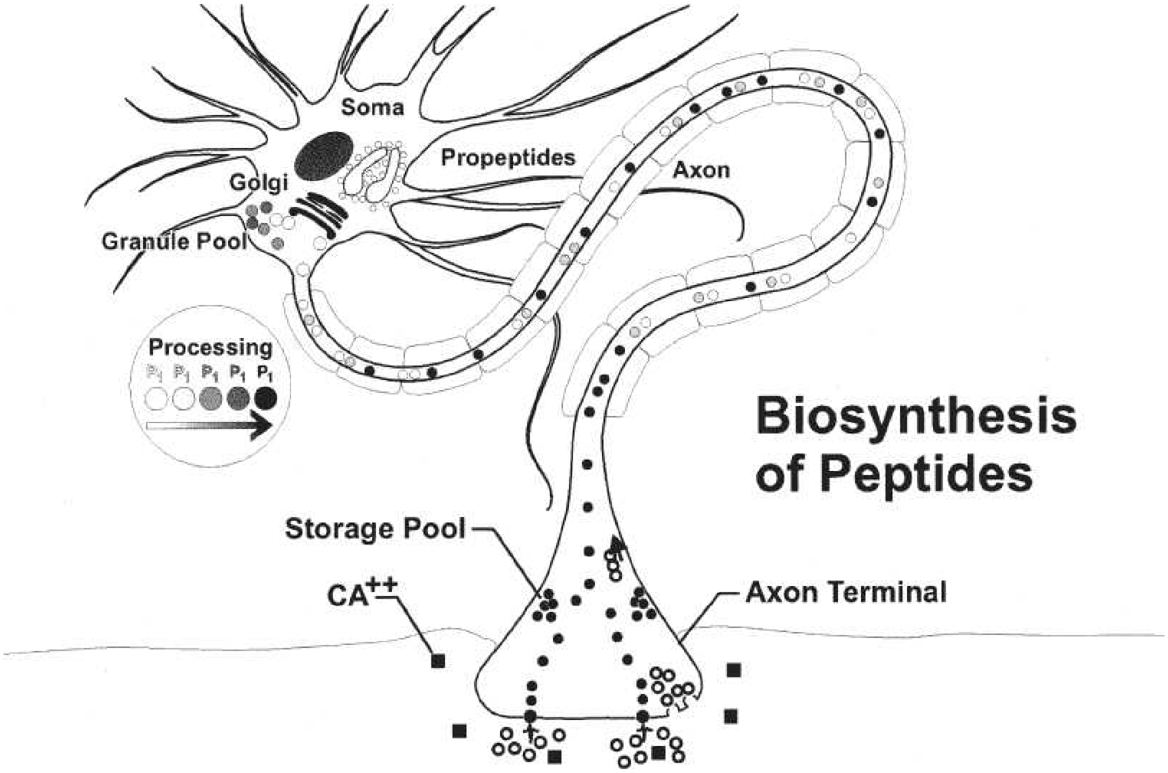

FIGURE 1.8–3. The peptidergic neuron. The figure shows the main steps in the chain of events, from the information stored in the DNA molecule to the peripherally detected peptide fragments. The DNA sequence in the nucleus is transcribed to the mRNA molecule for further transport to the endoplasmic reticulum, where a translation takes place to form a large precursor protein (preproprotein). That protein is prepared for axonal transport by packaging into neurosecretory vesicles or granules within the Golgi complex. During transport, the precursor protein is processed by specific cleavage enzymes into active and inactive peptide fragments. After release, the peptides are further degraded into smaller peptide fragments or constituent amino acids.

DISTRIBUTION AND REGULATION

Although many neuropeptides were originally isolated from the pituitary and peripheral tissues, the majority of neuropeptides have been found to be widely distributed throughout the brain. Those peptides involved in regulating pituitary secretion are concentrated in the hypothalamus. Hypothalamic releasing and release-inhibiting factors are produced in neurosecretory neurons adjacent to the third ventricle that send projections to the median eminence, where they release peptides into the hypothalamohypophysial portal circulatory system. Peptides produced in these neurons are often subject to feedback regulation by the peripheral hormones that they regulate. For example, TRH regulates the activity of the hypothalamic–pituitary–thyroid (HPT) axis, and thyroid hormones negatively feedback TRH gene expression. However, neuropeptide-expressing neurons and their projections are found in many other brain regions, including limbic structures, cerebral cortex, midbrain, hindbrain, and spinal cord. Neuropeptides are usually, if not exclusively, colocalized and released with other neuropeptides or classic neurotransmitters. The colocalization of neuropeptides within classical neurotransmitter circuits, first described for somatostatin in 1977, suggests an interaction between these systems and the modulation of monoamine neurotransmitter (e.g., norepinephrine) function by neuropeptides (e.g., NPY and galanin) is common and well established. Indeed, neuropeptides are co-transmitted with nonpeptide neurotransmitters and serve to modulate neurotransmission in both the pre- and postsynaptic cells.

NEUROPEPTIDE SIGNALING

Neuropeptides may act as neurotransmitters, neuromodulators, or neurohormones. Neuropeptide neurotransmitters are typically released from axonal terminals into a synapse where binding to neuropeptide receptors changes the postsynaptic membrane potential, either depolarizing or hyperpolarizing the cell. For classical neurotransmitters, this often involves direct modulation of voltage-gated ion channels. In contrast, neuropeptide neuromodulators and neurohormones do not directly affect the firing of the target cell itself but may alter the response of the cell to other neurotransmitters through the modulation of second messenger pathways. Neuropeptide release is not restricted to synapses or axon terminals but may occur at multiple sites along the axon or even from dendrites. Neuropeptides may also diffuse quite a distance from the release site to the target cell, which possesses the neuropeptide receptor, where it acts as a neurohormone and has been labeled volume neurotransmission. Such paracrine neuropeptide release can even occur from nonneuronal cell populations in the brain, such as glial cells. In fact, there are numerous examples of a mismatch between neuropeptide and neuropeptide receptor distribution in the brain. Neuropeptides are released by exocytosis of the granules in response to electrical or hormonal stimulation of the neuron containing the neuropeptide. Typically, neuropeptide release occurs when neurons fire at a high frequency or in burst patterns. Stimulation results in an increase in intracellular calcium concentrations, which leads to the fusion of the peptidergic granules to the plasma membrane and expulsion of the peptide into the extracellular space.

The cellular signaling of neuropeptides is mediated by specific neuropeptide receptors. Thus, understanding neuropeptide receptor function is essential for understanding neuropeptide biology. Neuropeptide receptors have undergone the same process of discovery and characterization that receptors for other neurotransmitters have enjoyed. The vast majority of neuropeptide receptors are G-protein–coupled, seven-transmembrane domain receptors (GPCR) belonging to the same family of proteins as the monoamine receptors. Each neuropeptide receptor is usually, but not always exclusively, coupled to one type of G protein (e.g., G_s, G_i, G_q). Receptors can also homo- and heterodimerize, allowing multiple modes of action via a single receptor population, though this process, its regulation, and the physiologic implications are not currently well understood. Depending on the subtype of G protein with which the receptor interacts, receptor activation may result in the stimulation or inhibition of specific second messenger pathways. The most common types of receptor signaling pathways involve the activated G protein modulating the activity of either adenylate cyclase (e.g., Gs or Gi) or phospholipase C (e.g., Gq). Stimulation of adenylate cyclase results in increased cAMP production in the cell, while stimulation of phospholipase C results in increased diacylglycerol (DAG) and inositol triphosphate (IP_3). These so-called “second messengers” then effect increases in intracellular calcium concentrations, activation of protein kinases, and ultimately a host of cellular responses, including altered gene expression, altered membrane potential, and/or cellular degranulation.

Many neuropeptides exert their effects through multiple different receptor subtypes, which have different affinities for the peptides and activate different second messenger pathways. These receptor subtypes are typically differentially distributed throughout the brain. Furthermore, many receptors may be modulated by more than one neuropeptide. For example, there are three subtypes of the vasopressin receptor, the V1a, V1b, and V2 subtypes, with the V1a and V1b predominating in the brain, while the V2 is localized in the kidney. Each of these receptor subtypes exhibits a unique tissue distribution, interacts with different G proteins, and activates different second messenger systems. Within the brain, there are remarkable species differences as well as individual variation in OT and V1a receptor distribution, which has been associated with intra- and interspecies variation in social behaviors. In addition, OT may stimulate vasopressin receptor subtypes, and vasopressin may stimulate the OT receptor (OTR). Likewise, the main two CRF receptors are differentially localized within the brain, and both receptors can be modulated by both CRF and urocortin I, making it difficult to ascertain the relative role of each receptor in CRF functioning.

Many, if not perhaps all, neurotransmitters appear to be co-released with at least one other neurotransmitter at the synapse. ATP, for example, can apparently be co-released from all neuronal types, though the action of the molecule will vary according to the biology of the postjunctional cell. This co-transmission certainly also occurs for neuropeptides. One of the first to be delineated, NPY, is co-released with norepinephrine from the same vesicles within cells to act simultaneously on the target cell. NPY can also act *via* feedback inhibition on the releasing cell to reduce signaling tone. In this instance, NPY appears to play a neuromodulatory role at the synapse to control signal strength and act as a shutoff switch. Galanin also plays a similar role in norepinephrine-releasing neurons.

Other types of co-transmission can occur, including the differential release of transmitters and cotransmitters (e.g., *via* frequency-modulated release), post- and/or prejunctional targeting co-transmitters, synergistic versus inhibitory co-transmitters, and others. It is important to appreciate neuropeptides as co-transmitters in such systems because co-transmission has the potential to mediate or modulate the underlying disease pathophysiology. Using NPY as an example again, NPY is also co-transmitted with NE in the adrenals, where cortisol is released in response to stress. NPY appears to play an overall suppressive role in NE signaling within the adrenals, thus suppressing the stress response. It has been widely theorized that activation of NPY receptors can ameliorate some of the trauma associated with chronic stress disorders and drug abuse/alcoholism, where there is prolonged activation of the stress response and where stress is viewed as playing a major role in relapse and drug-seeking behaviors.

Most, if not all, genes for neuropeptide receptors and ligands have been cloned and characterized, though some receptors called “orphan receptors” have, as of yet, unknown endogenous ligands. Similarly, some ligands, for example, “cocaine- and amphetamine-regulated transcript”

(CART), are thought to signal via GPCRs, which remain to be identified. It is likely, however, that all known GPCRs and ligands will be “deorphanized” within the next decade, as numerous efforts are underway to do so. Once the cDNA of the receptor has been isolated, it can be used to produce purified receptor protein for structural and functional studies. By mutating specific amino acids in the receptor structure and determining relative binding affinities of peptides with various amino acid substitutions, it is possible to elucidate the nature of the ligand–receptor interaction and to inform crystal structures of the receptors. This information should facilitate the development of compounds that specifically modulate receptor function, including nonpeptide drugs, leading to the ability to manipulate peptide systems in ways that are currently enjoyed by the more classic neurotransmitters. The availability of cDNAs encoding the receptor also permits the neuroanatomical mapping of the receptor-producing cells in the brain by use of probes that target the mRNA of the receptor specifically, which is critical for understanding the neural circuits modulated by the peptide. Finally, with the cloned receptor in hand, it is possible to use transgenic techniques, such as targeted gene overexpression or gene knockouts, to further elucidate receptor function. Viral-mediated RNAi and CRISPR techniques now allow the targeted synthesis disruption of specific receptor populations, allowing researchers to examine the role of these receptor populations on physiology and behavior. Peptide receptor Cre transgenic animals in combination with Cre-dependent vectors enable investigators to identify the role of receptor-expressing neuronal populations in behavior using opto- and chemogenetics.

The three factors that determine the biologic role of a neuropeptide hormone are (i) the temporal–anatomical release of the peptide, (ii) coupling of the neuropeptide receptor to intracellular signaling pathways, and (iii) the subtype of the receptor expressed and their distribution in different cell populations and circuits. Genetic studies have demonstrated that regulatory sequences flanking the receptor coding region determine the expression pattern of the receptor and thus the physiologic and behavioral response to the neuropeptide. For example, mice and voles differ in the localization of AVP receptors in the brain, and they also differ in their behavioral response to AVP. However, when transgenic mice were created carrying the vole AVP receptor gene with the flanking regulatory sequences, the *mice* expressed the receptor in a pattern similar to that of the *vole*, and then displayed behavioral responses to AVP similar to that of voles. This study suggests that polymorphisms in the regulatory region of a neuropeptide receptor gene could result in significant differences in neuropeptide function and interindividual differences in response to neuronal stimulation and, ultimately, behavior relevant to psychiatric disorders. Many receptor genes are being examined in genetic studies for associations with psychiatric disorders.

Historically, the inability to block specific neuropeptide signals pharmacologically has severely hindered research into the roles of endogenous peptides in various behaviors and physiologic effects. However, for many neuropeptide receptors, selective agonists and antagonists are now available, which have been extremely informative in preclinical studies to examine receptor function. As mentioned above, most of these compounds are derivatives of the peptide itself and therefore do not pass the blood–brain barrier. More recently, a number of pharmaceutical companies, and even academic labs, have synthesized nonpeptidergic, lipophilic compounds that can penetrate the blood–brain barrier and may act as neuropeptide agonists or antagonists. The development of these types of compounds is essential for understanding the role of neuropeptide receptor function in human behavior and may also be useful in the development of positron emission tomography (PET) radioligands to study receptor distribution in living human subjects. These compounds also hold promise as therapeutic agents in the treatment of certain psychiatric disorders.

Peptidases

Unlike monoamine neurotransmitters, peptides are not inactivated by being actively taken up by presynaptic nerve terminals. Rather, released peptides are degraded into smaller fragments and eventually into single amino acids by specific enzymes termed peptidases. The enzymes may be found bound to post- or presynaptic neural membranes or in solution in the cytoplasm and extracellular fluid, and they are distributed widely in peripheral organs and serum, as well as in the CNS. As a result, neuropeptides generally have half-lives on the order of minutes once released. There are several general classes of peptidases, with several distinct enzymes in each class. Those classes include the serine endopeptidases containing such enzymes as trypsin and chymotrypsin; the thiol peptidases, such as pyroglutamate aminopeptidase and cathepsin B and C; the acid proteases, such as pepsin and renin; the metalloendopeptidases, such as neural endopeptidase and angiotensin-converting enzymes (ACE); and the metalloexopeptidases, such as the aminopeptidases and the carboxypeptidases such as enkephalin convertase and carboxypeptidase A and B. These degradative enzymes are often the same as those used in processing but have different subcellular locations. An example is carboxypeptidase B, which cleaves the dibasic amino acid residues flanking the active peptide sequence in the prohormone during processing or reduces activity at the receptor if the peptide contains dibasic amino acids in the active sequence, such as NT. Peptidases have pH and temperature optimums for activity and can be inhibited by various chemicals or chelators or by amino acid substitution at vulnerable points in the peptide chain. Alterations in peptidase activity or concentration can contribute to alterations in the synaptic availability of a peptide, and the regulation of peptidase levels may be as exquisitely controlled as receptor number and peptide synthesis and release. Cleavage of the actively released form of the peptide usually ends or significantly reduces biologic activity, but examples abound of partial or complete receptor activation by partially metabolized peptides or their fragment.

Peptidases offer yet another potential opportunity for the integration and regulation of neuropeptide transmitter actions and synaptic availability. Because current peptidase inhibitors are relatively nonspecific in their ability to inhibit various peptidases, there have been few attempts to influence peptide concentrations by pharmacologic blockade of their associated peptidases. The ACE inhibitors such as captopril and lisinopril are one exception to that generality. It is expected that newer generations of peptidase inhibitors, with discrete peptidase and possibly regional specificity, will be developed that eventually may allow the truly elegant manipulation of endogenous neuropeptide concentrations.

NEUROPEPTIDE SYSTEMS AS DRUG TARGETS

Because of the many regulatory roles neuropeptide receptors play in the CNS, it is no surprise that they also make attractive drug targets. There have been numerous attempts to modulate neuropeptide receptors for the treatment of psychiatric diseases, with some success and also some notable failures. The first and most robust class of therapeutics, particularly for peripheral neuropeptide receptors, is in administering the neuropeptide itself, with perhaps some small modifications to make it more drug like. The issue in psychiatric diseases is, however, that the vast majority of the peptides are not brain penetrant. While there have been numerous attempts to either create brain-penetrant peptides or else centrally administer them, these efforts have largely yielded little of clinical value. The field remains focused on this approach, however, and developments in gene editing and other biologic drug approaches hold some potential to yield useful results.

It is because of these and other reasons that most approaches to drugging CNS neuropeptide receptor systems remain focused on the identification and development of low molecular weight (ideally less than 500 Da) brain-penetrant molecules. A plethora of small molecules has been developed that selectively and potently target different neuropeptide receptors. These small molecules are highly versatile and can be designed to have highly specific activities. Typical approaches are to activate the receptor (agonist), inhibit receptor signaling (antagonist), or reduce intrinsic receptor activity (inverse agonist). More specialized approaches have been developed in recent years. A positive allosteric modulator (PAM) small molecule, for example, will potentiate the signaling of a neuropeptide by binding to the receptor at a site on the protein (known as the allosteric site) that is not the site where the neuropeptide binds (known as the orthosteric site) and cause the receptor to initiate more signaling with the same given amount of neuropeptide. The PAM approach is elegant and has the potential benefits of selectively activating only those receptors that would normally be activated, as compared to a typical orthosteric agonist, which may activate receptor populations throughout the body. A similar negative allosteric modulator (NAM) approach can be used to reduce signaling through receptors with the same given amount of neuropeptide present. Finally, a great deal of emphasis has been recently placed on creating “biased” signaling drugs. A given receptor will usually have more than one intracellular signaling pathway that it activates within normal physiology. A biased compound is aimed at selectively activating or inhibiting one pathway over another. A famous example is with the μ-opioid receptor, where it is thought that an opiate drug that does not activate the β-arrestin signaling system will have less tolerance and, therefore, addiction potential. These highly specialized approaches to modulating receptors remain mostly experimental at this point, but there are many efforts aimed at developing such drugs. A major unanswered question is whether a relatively minor modulation of receptor signaling systems that are generally only modulators of more robust neuronal firing and network systems can be powerful enough to result in positive clinical outcomes.

This issue that neuropeptide systems appear to be largely highly modulatory may also be the reason why more traditional drug discovery for neuropeptide systems has not lived up to expectations. There have been several notable failures in developing neuropeptide receptor drugs. Particularly, the “big three” failures all around the same time in the mid-2000s seemed to cause the field to reconsider whether neuropeptide small-molecule drugs can be useful for psychiatry. These were phase III clinical trial failures for CRF receptor antagonists for anxiety and depression, NK1 (SP) receptor antagonists for pain, and NPY Y5 receptor antagonists for obesity. While the question of utility potential remains largely unanswered, the same questions and answers have appeared in other neuroscience fields with notable failures, such as in failures of AD drugs. As with AD, it is critical that the exact approach, dosing, and mechanism of action needs to be better defined and evaluated before developing a drug. Moreover, the disease and clinical outcomes deserve much more thorough consideration before initiating a trial. A clear example is CRF receptor antagonists, which failed to have meaningful outcomes for generalized anxiety disorder (GAD). GAD is, like many DSM5 disorders, highly heterogeneous. CRF receptor antagonist approaches may be quite effective in a much more defined, enriched subpopulation of patients, for example, post-traumatic stress disorder (PTSD) patients with a genetic and childhood maltreatment history promoting increased CRF activity. It appears that for all neuropsychiatric disorders, clinical population, clinical trials design, trials outcomes, and other factors must receive considerably more attention than was given in the past if there is any chance for success. As a clear example and expanded upon later in this chapter, the ongoing hypothesis that CRF hypersecretion may represent a pathophysiologic basis for mood and anxiety disorders, as well as alcohol use disorder (AUD), is supported by significant preclinical and clinical research. However, only a subset of individuals, perhaps 20% to 30%, with these disorders are expected to be CRF hypersecreters. Clinical trial designs based on broad diagnostic categories lose tremendous statistical power in assessing the value of these medications as well as potentially causing untoward effects in subjects with normal CRF function. The field appears to largely be working through these issues, and a resurgence in the development of neuropeptide receptor small-molecule clinical trials is expected.

SPECIFIC NEUROPEPTIDES AS EXEMPLARS OF NEUROPEPTIDE BIOLOGY

Thyrotropin-Releasing Hormone

In 1969, TRH, a pyroglutamyl-histidyl-prolylamide tripeptide (see Table 1.8–2), became the first of the hypothalamic-releasing hormones to be isolated and characterized, work that, in part, resulted in the Nobel prize for Drs Andrew Schally and Roger Guillemin in 1977. The discovery of the structure of this hormone led to the conclusive demonstration that peptide hormones secreted from the hypothalamus regulate the secretion of hormones from the anterior pituitary. The gene for TRH in humans resides on chromosome 3q13.3-q21. In the rat, it consists of three exons (coding regions) separated by two introns (noncoding regions) (see Fig. 1.8–2). The first exon contains the 5′ untranslated region of the mRNA encoding the TRH preprohormone, the second exon contains the SP sequence and much of the remaining amino-terminal end of the precursor peptide, and the third contains the remainder of the sequence, including five copies of the TRH precursor sequence, the carboxy-terminal region, and the 3′ untranslated region. The 5′ flanking region of the gene, or promoter, contains sequences homologous to the glucocorticoid receptor and the thyroid hormone receptor DNA binding sites, providing a mechanism for the regulation of this gene by cortisol and negative feedback by thyroid hormone. Enzymatic processing of TRH begins with excision of the progenitor peptides by carboxypeptidases, amidation of the C-terminal proline, and cyclization of the N-terminal glutamine to yield five TRH molecules per prohormone molecule. TRH is widely distributed in the CNS, with TRH immunoreactive neurons being located in the olfactory bulbs, entorhinal cortices, hippocampus, extended amygdala, hypothalamus, and midbrain structures. As is the case for most neuropeptides, the TRH receptor is also a member of the seven-transmembrane domain, G-protein–coupled receptor family.

Hypothalamic TRH neurons project nerve terminals to the median eminence, where they release TRH into the hypothalamohypophyseal portal system, where it is transported to the adenohypophysis causing the release of TSH into the systemic circulation. TSH subsequently stimulates the release of the thyroid hormones triiodothyronine (T_3) and thyroxine (T_4) from the thyroid gland. TRH neurons in the paraventricular nucleus contain thyroid hormone receptors and respond to increases in thyroid hormone secretion with a decrease in TRH gene expression and synthesis. This negative feedback of thyroid hormones onto the TRH synthesizing neurons was first demonstrated by a decrease in TRH content in the median eminence (due to increased release in trying to drive thyroid function) but not in the paraventricular nucleus of the hypothalamus after thyroidectomy. This effect can be reversed with exogenous thyroid hormone treatment. The treatment of normal rats with exogenous thyroid hormone decreases TRH concentration in the paraventricular nucleus and the posterior nucleus of the hypothalamus. Using a probe against the TRH preprohormone mRNA, *in situ* hybridization studies have demonstrated that TRH mRNA is increased in the paraventricular nucleus 14 days after thyroidectomy. The ability of thyroid hormones to regulate TRH mRNA can be superseded by other stimuli that activate the HPT axis. In that regard, repeated exposure to cold (which releases TRH from the median eminence) induces increases in the levels of TRH mRNA in the paraventricular nucleus despite concomitantly elevated concentrations of thyroid hormones. Further evidence of the different levels of communication of the HPT axis is seen in the ability of TRH to regulate the production of mRNA for the pituitary TRH receptor and for TRH concentrations to regulate the mRNA coding for both the α- and β-subunits of the thyrotropin (TSH) molecule. In addition, TRH-containing synaptic boutons have been observed in contact with TRH-containing cell bodies in the medial and periventricular subdivisions of the paraventricular nucleus, thus providing anatomical evidence for ultrashort feedback regulation of TRH release. Negative feedback by thyroid hormones may be limited to the hypothalamic TRH neurons because negative feedback on TRH synthesis by thyroid hormones has not been found in extrahypothalamic TRH neurons.

The early availability of adequate tools to assess HPT axis function (i.e., radioimmunoassays, synthetic peptides), coupled with observations that primary hypothyroidism is associated with depressive symptomatology, ensured extensive investigation of the involvement of this axis in affective disorders. Early studies established the hypothalamic and extrahypothalamic distribution of TRH. This extrahypothalamic presence of TRH quickly led to speculation that TRH might function as a neurotransmitter or neuromodulator. Indeed, a large body of evidence supports such a role for TRH. Within the CNS, TRH is known to modulate several different neurotransmitters, including dopamine, serotonin, acetylcholine, and the opioids. TRH has been shown to arouse hibernating animals and counteracts the behavioral response and hypothermia produced by a variety of CNS depressants, including barbiturates and ethanol.

Interest in putative CNS actions of TRH was stimulated by studies of the HPT axis and depression by Prange and colleagues. Nearly four decades ago, it was hypothesized that thyroid function was integral to the pathogenesis of and recovery from affective disorders due to the numerous interactions among thyroid hormones, catecholamines, and adrenergic receptors in the CNS. Overall, these studies suggested a role for thyroid dysfunction in refractory depression and are consonant with clinical studies suggesting the existence of an increased rate of hypothyroidism and autoimmune thyroiditis among patients with refractory depression.

The use of TRH as a provocative agent for the assessment of HPT axis function evolved rapidly after its isolation and synthesis. Clinical use of a standardized TRH stimulation test, which measures negative feedback responses, revealed blunting of the TSH response in approximately 25% of euthyroid patients with major depression. These data have been widely confirmed. The observed TSH blunting in depressed patients does not appear to be the result of excessive negative feedback due to hyperthyroidism because thyroid measures such as basal plasma concentrations of TSH and thyroid hormones are generally in the normal range in these patients. It is possible that TSH blunting is a reflection of pituitary TRH receptor downregulation as a result of median eminence hypersecretion of endogenous TRH. Indeed, the observation that CSF TRH concentrations are elevated in depressed patients as compared to controls supports the hypothesis of TRH hypersecretion but does not elucidate the regional CNS origin of this tripeptide. In fact, TRH mRNA expression in the paraventricular nucleus of the hypothalamus is decreased in patients with major depression. However, it is not clear whether the altered HPT axis represents a causal mechanism underlying the symptoms of depression or simply a secondary effect of depression-associated alterations in other neural systems. More recently, Cohen and colleagues have reported that even mild increases in plasma TSH concentrations (>3.5 mIU/mL), though technically in the upper end of the so-called “normal” range, are associated with a poor response to antidepressant treatment.

Corticotropin-Releasing Factor and Urocortins

In the 1950s, it was observed that pituitary extracts contained a factor, referred to as CRF, that could stimulate the release of ACTH from anterior pituitary cells *in vivo*. After a search spanning nearly three decades, Vale and colleagues isolated and characterized CRF as a 41 amino acid peptide in 1981. The gene for CRF in humans is located on chromosome 8q13 and is composed of two exons, with the CRF preprohormone being encoded entirely on exon 2 (see Fig. 1.8–2). More recently, the related neuropeptides, urocortin 1, urocortin 2, and urocortin 3, have been identified and share similar gene structures. CRF is the primary hypothalamic ACTH secretagogue in most species, and it also functions as an extrahypothalamic neurotransmitter/neuromodulator in a CNS network which, along with the urocortins, globally coordinates responses to stressors. There is convincing evidence to support the hypothesis that CRF and the urocortins play a complex role in integrating the endocrine, autonomic, immunologic, and behavioral responses of an organism to stress.

Although it was originally isolated because of its functions in regulating the hypothalamic—pituitary–adrenal (HPA) axis, CRF is widely distributed throughout the brain. The paraventricular nucleus of the hypothalamus (PVN) is the major site of CRF-containing cell bodies that influence anterior pituitary ACTH secretion. These neurons originate in the parvocellular region of the PVN and send axon terminals to the median eminence, where CRF is released into the hypothalamic–hypophyseal portal system in response to stressful stimuli. A small group of PVN neurons also project to the brainstem and spinal cord, where they regulate autonomic aspects of the stress response. CRF-containing neurons are also found in other hypothalamic nuclei, the neocortex, the extended amygdala, brainstem, and spinal cord. Central CRF infusion into laboratory animals produces physiologic changes and behavioral effects similar to those observed following stress, including increased locomotor activity, increased responsiveness to an acoustic startle, and decreased exploratory behavior in an open field.

In a manner similar to that described for TRH and thyroid hormones, CRF gene expression and content in the PVN are negatively influenced by glucocorticoids (cortisol) and positively regulated by a wide variety of stressors. Adrenalectomy results in an increase in CRF mRNA expression in the PVN, and glucocorticoid replacement decreases CRF mRNA expression in a dose-dependent manner. In contrast to their effects on the PVN, glucocorticoids increase CRF mRNA expression in the amygdala rather than decreasing it.

CRF is also found in the raphe nuclei and the locus coeruleus (LC), the origin of the major serotonergic and noradrenergic projections to the forebrain, respectively, circuits long postulated to play a role in the pathophysiology of depression and anxiety. Increased anxiety observed after direct CNS administration of CRF has been hypothesized to be mediated, in part, by increased noradrenergic activity. Stress has been shown to produce an increase in CRF content in the locus ceruleus and a decrease in CRF concentrations in the median eminence (consistent with increased release). Other studies have shown that CRF-containing nerve terminals impinge upon noradrenergic neurons of the locus ceruleus and that exogenous CRF applied to those neurons alters their firing rate. Some of the noradrenergic locus ceruleus neurons, in turn, project to the hypothalamic PVN where their input increases CRF synthesis and release. Because CRF injection into the locus ceruleus elicits fearful or anxious behavior, one could postulate that stress activates the CRF neurons terminating on the locus ceruleus noradrenergic neurons, which then may, acting along with other inputs to the PVN, stimulate the stress-induced increased release of CRF from the median eminence. Another hypothesis that has been pursued in recent years is severe stress causes nucleus accumbens dopamine neurons to become insensitive to CRF, resulting in reduced dopamine release and increased anhedonia. Interestingly, adult animals exposed to maternal separation early in life, an animal model for early adverse childhood experiences and a very severe form of stress, exhibit elevated CRF concentrations in the LC, increased CRF mRNA expression, and exaggerated HPA axis responses to stress. Relatedly, increasing evidence strongly supports a role for CRF in stress-induced relapse in substance abuse disorders.

The physiologic and behavioral roles of the urocortins are less understood, but several studies suggest that urocortin 2 and 3 are anxiolytic and may dampen the stress response. This has led to the hypothesis that CRF and the urocortins act in opposition, but this is likely an oversimplification. Urocortin 1 is primarily synthesized in the Edinger–Westphal nucleus, lateral olivary nucleus, and supraoptic hypothalamic nucleus. Urocortin 2 is synthesized primarily in the hypothalamus, while urocortin 3 cell bodies are found more broadly in the extended amygdala, perifornical area, and preoptic area.

The CRF system is further complicated by the fact that the effects of CRF and the urocortins are mediated by at least two receptor subtypes, CRF_1- and CRF_2-receptor (Fig. 1.8–4). The CRF_1 receptor is abundantly expressed in the cerebral cortex, cerebellum, medial septum, and anterior pituitary, whereas the CRF_2 receptor is predominantly found in the lateral septum ventromedial hypothalamus and choroid plexus of rodents but has considerable expression in the human cortex. The CRF_1 receptor appears to be the predominant receptor mediating the effects of CRF in the stress response. The CRF_1 receptor has a 4- to 10-fold higher affinity for CRF than for urocortin 1, with very low affinity for the other urocortins. In contrast, the CRF_2 receptor has a 40-fold higher affinity for the urocortins relative to CRF. Thus, the urocortins have been proposed to be the endogenous ligands for the CRF_2 receptor, but little is known regarding their physiologic role. As expected, CRF_1 receptor knockout mice display decreased anxiety-like behavior, have an impaired stress response, and exhibit elevated CRF mRNA expression in the PVN due to a lack of glucocorticoid negative feedback. In contrast, CRF_2 receptor knockout mice display increased anxiety-like behavior and are hypersensitive to stress.

Hyperactivity of the HPA axis in major depression remains one of the most consistent findings in biologic psychiatry. The reported HPA axis alterations in major depression include hypercortisolemia, resistance to dexamethasone suppression of cortisol secretion (a measure of negative feedback), blunted ACTH responses to intravenous CRF challenge, increased cortisol responses in the combined dexamethasone/CRF test, and elevated CSF CRF concentrations. The exact pathologic mechanism(s) underlying HPA axis dysregulation in major depression and other affective disorders remains to be elucidated.

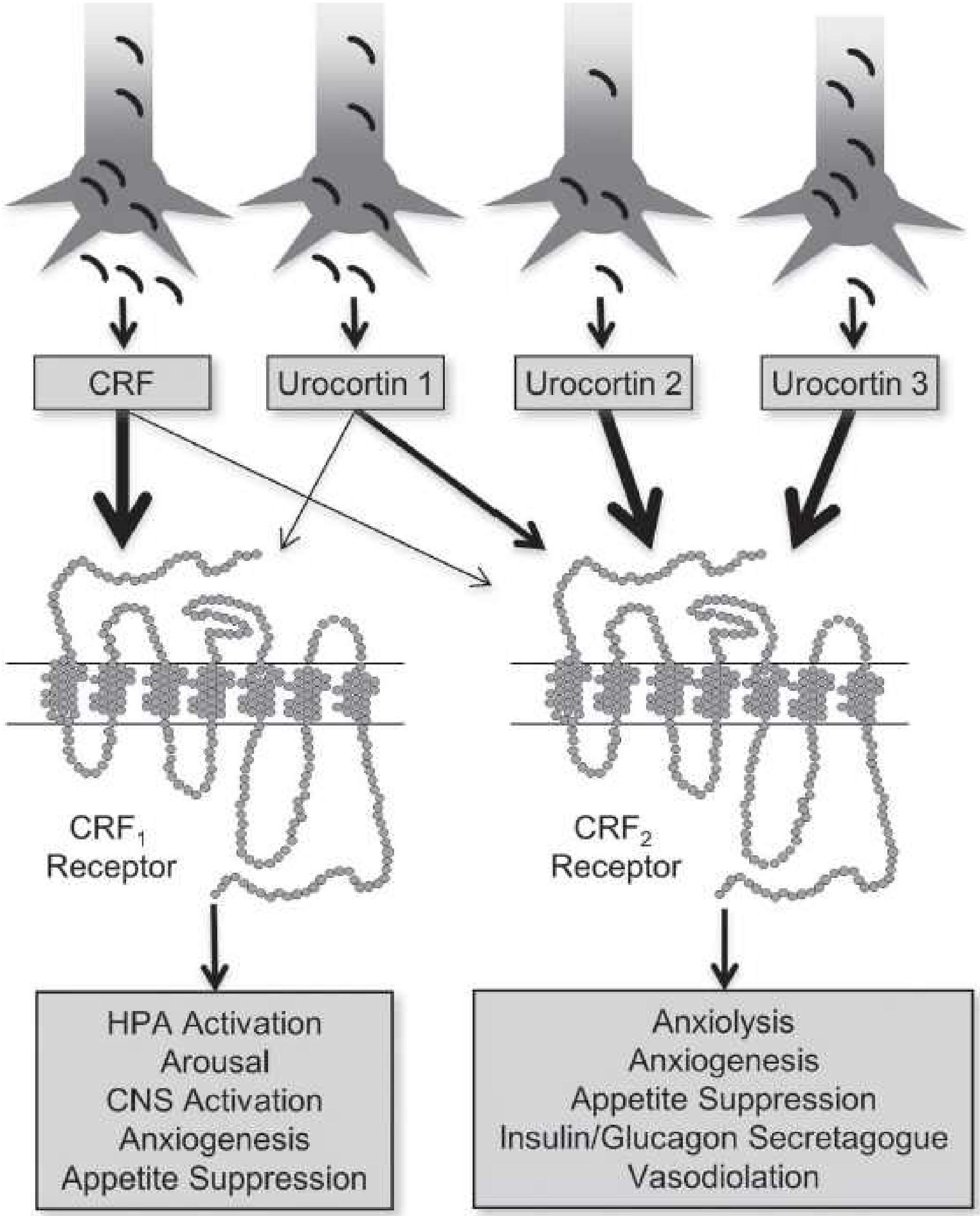

CRF
Urocortin 1
Urocortin 2
Urocortin 3
CRF_1 Receptor
CRF_2 Receptor
HPA Activation
Arousal
CNS Activation
Anxiogenesis
Appetite Suppression
Anxiolysis
Anxiogenesis
Appetite Suppression
Insulin/Glucagon Secretagogue
Vasodiolation

FIGURE 1.8–4. Ligands and receptors of the corticotropin-releasing factor (CRF) signaling network and their putative roles. The figure illustrates the complexity of the CRF system with four different ligands modulating two different receptors, each of which regulates divergent physiologic processes. The thickness of the *arrows* represents the relative affinity of each ligand for the respective receptors. (Adapted by permission from Nemeroff CB, Vale WW. The neurobiology of depression: Inroads to treatment and new drug discovery. *J Clin Psychiatry* 2005;66(S7):5–13. Copyright 2005, Physicians Postgraduate Press.)

Once the phenomenon of HPA axis hyperactivity in patients with major depression was established, many research groups utilized various provocative neuroendocrine challenge tests as a “window into the brain” in attempts to elucidate pathophysiologic mechanisms. In normal subjects, the CRF stimulation test, using either rat/human or ovine CRF, yields robust ACTH, β-endorphin, β-lipotropin, and cortisol responses following intravenous or subcutaneous administration. However, in patients with major depression, blunting of ACTH or β-endorphin secretion with a normal cortisol response has been repeatedly reported. Patients with PTSD, 50% of whom also fulfill the criteria for major depression, also show blunted ACTH secretion in response to a CRF challenge. Importantly, researchers have reported normalization of the ACTH response to CRF following clinical recovery from depression, suggesting that the blunted ACTH response, like dexamethasone nonsuppression, may be a state marker for depression. Early-life stress apparently sensitizes the HPA axis and leads to a greater risk of developing depression later in life. Depressed women who were victims of childhood abuse exhibit an exaggerated ACTH and cortisol response to a psychosocial stressor, presumably due to CRF hypersecretion. Depressed men with a history of childhood abuse exhibit marked HPA axis hyperactivity in the combined dexamethasone/CRF test.

Mechanistically, two hypotheses have been advanced to account for the ACTH blunting following exogenous CRF administration. The first hypothesis suggests that pituitary CRF receptor downregulation occurs as a result of hypothalamic CRF hypersecretion. The second hypothesis postulates altered sensitivity of the pituitary to glucocorticoid negative feedback. Substantial support has accumulated, favoring the first hypothesis. However, neuroendocrine studies represent a secondary measure of CNS activity; the pituitary ACTH responses principally reflect the activity of hypothalamic CRF rather than that of the corticolimbic CRF circuits. The latter of the two is more likely to be involved in the pathophysiology of depression.

A potentially more direct method for the evaluation of extrahypothalamic CRF tone may be obtained from measurements of CRF concentrations in the CSF. A marked dissociation between CSF and plasma neuropeptide concentrations has been described, thus indicating that neuropeptides are secreted directly into CSF from brain tissue as opposed to being derived from plasma-to-CSF transfer. Evidence that CSF CRF concentrations originate from extrahypothalamic CRF neurons has been obtained from studies in which CSF CRF concentrations were repeatedly measured over the course of the day. CSF CRF concentrations in rhesus monkeys do not appear to be entrained with pituitary–adrenal activity. The proximity of corticolimbic, brainstem, and spinal CRF neurons to the ventricular system of the brain suggests that these areas make substantial contributions to the CSF CRF pool.

In a series of studies, significant elevations of CSF CRF concentrations in drug-free patients with major depression or following suicide have been demonstrated. In addition, the severity of depression appears to correlate significantly with CSF CRF concentrations in patients with anorexia nervosa, multiple sclerosis, and Huntington disease. The elevation of CSF CRF concentrations in patients with anorexia nervosa reverts to the normal range as these patients approach normal body weight. No alterations of CSF CRF concentrations have been reported in other psychiatric disorders, including mania, panic disorder, and somatization disorders, as compared to controls. It is now clear that patients with early-life trauma in the form of child abuse or neglect exhibit increased CSF CRF concentrations, as has now been demonstrated in patients with major depression, PTSD, and antisocial personality disorder.

Of particular interest is the demonstration that the elevated CSF CRF concentrations in drug-free depressed patients are significantly decreased after successful treatment with ECT treatment, indicating that CSF CRF concentrations, like hypercortisolemia, represent a state, rather than a trait, marker. Other studies have confirmed this normalization of CSF CRF concentrations following successful treatment with fluoxetine. One group demonstrated a significant reduction of elevated CSF CRF concentrations in 15 female patients with major depression who remained depression-free for at least 6 months following antidepressant treatment as compared to little significant treatment effect on CSF CRF concentrations in 9 patients who relapsed in this 6-month period. This suggests that elevated or increasing CSF CRF concentrations during antidepressant treatment may be the harbinger of a poor response in major depression despite early symptomatic improvement. Interestingly, treatment of normal subjects with desipramine or, as noted above, of depressives with fluoxetine is associated with a reduction in CSF CRF concentrations.

In preclinical studies, CRF hypersecretion is associated with CRF-receptor downregulation. Depression is a major determinant of suicide, with more than 50% of completed suicides accomplished by patients with major depression. If CRF hypersecretion is a characteristic of depression, evidence of related CRF-receptor downregulation should be evident in the CNS of depressed suicide victims. Indeed, in two studies, a marked decrease in the density of CRF_1 receptors and CRFR1 mRNA expression in the frontal cortex of suicide victims as compared to matched control samples was observed.

If CRF hypersecretion is a factor in the pathophysiology of depression, then reducing or interfering with CRF neurotransmission might be an effective strategy to alleviate depressive symptoms. Over the past several years, a number of pharmaceutical companies have committed considerable effort to the development of small-molecule CRF_1 receptor antagonists, also termed CRHR1 antagonists, that can effectively penetrate the blood–brain barrier. Several compounds have been produced with reportedly promising drug-like characteristics.

More than 20 clinical trials have been conducted on the effects of CRF receptor antagonists on mood and anxiety disorders. These clinical trials have largely been disappointing because of the lack of efficacy for CRF_1 receptor antagonists in patients with major depression or GAD. It is important to note that these studies have not used narrowly defined depressive patient subtypes such as psychotic, anxious, or atypical depression but instead have utilized a much broader DSM inclusion criterion. Considering the fact that there is overwhelming evidence that CRF is hypersecreted in depressed patients with a history of childhood abuse and neglect, a clinical trial of CRF receptor antagonists in this discrete subset of patients with major depression would be of considerable interest. Moreover, two studies have documented elevated CSF CRF concentrations in PTSD. Here also, clinical trials have been disappointing, with little emerging that suggests a potential for disease modification.

In spite of the overwhelming preclinical and clinical pathophysiologic evidence for the role of CRF circuit hyperactivity in the pathogenesis of depression, PTSD, and several anxiety disorders, as well as AUD, the results of randomized controlled clinical trials in these populations with one or another CRF_1 antagonist have been universally disappointing. Recently, our group conducted a trial of GSK561679, a CRF_1 antagonist in women with PTSD. In the 96 subjects who completed the study, using the primary endpoint, the Clinician-Administered PTSD Scale (CAPS) score, there was no beneficial effect of the CRF_1 antagonist; a secondary analysis revealed that in subjects with a history of child abuse/neglect, who were homozygous for the GG genotype of the CRF_1 receptor, exhibited significant improvement after treatment with the CRF_1 antagonist. In a subsequent study, our group reported that DNA methylation of the PTSD-relevant genes NR3C1 and FKBP5, previously shown to predict treatment outcome of PTSD patients after psychotherapy, interacted with child abuse to predict PTSD symptom change following CRF_1 antagonist treatment.

Oxytocin and Vasopressin

The vasopressor effects of posterior pituitary extracts were first described in 1895, and the potent extracts were named vasopressin. In 1953, OT became the first peptide hormone to have its structure elucidated and the first to be chemically synthesized, leading to the Nobel Prize in Chemistry being awarded to Vincent du Vingeaud in 1955. The human OT and AVP genes are situated tandemly oriented in opposite transcriptional directions on chromosome 20p13, separated by a several kilobase intergenic sequence (see Fig. 1.8–2). Both peptides are cyclical nonapeptides containing a cysteine–cysteine disulfide bond and differ at only two amino acid residues (see Table 1.8–2). Like the sequence homology of the peptides themselves, the genes for OT and AVP share a common structure, suggesting that the two hormones are derived from a single ancestral hormone as a result of a gene duplication event early in vertebrate evolution. The two genes are organized in opposite orientations, and the OT and AVP mRNA are transcribed from opposite DNA strands toward each other. Each gene consists of three exons, with the first exon encoding the 5′ untranslated region and the translation initiation codon followed by the SP sequence and the peptide hormone portion of the preprohormone. Exons 2 and 3 encode the neurophysin portion of the prohormone molecule. The AVP prohormone also contains a glycoprotein, copeptin, whose function is unclear but that has been suggested as a biomarker in a number of diseases. Neurophysin is thought to play a role in the posttranslational processing and transport of the peptides, and copeptin is thought to do the same.

OT and vasopressin mRNAs are among the most abundant messages in the hypothalamus, being heavily concentrated in the magnocellular neurons of the paraventricular and the supraoptic nuclei of the hypothalamus, which send axonal projections to the neurohypophysis. These neurons produce all of the OT and AVP that are released into the bloodstream, where these peptides act as a hormone on peripheral targets. OT and AVP are generally synthesized in separate neurons within the hypothalamus. OT released from the pituitary is most often associated with female reproductive functions such as regulating uterine contractions during parturition and the milk ejection reflex during lactation. AVP, also known as antidiuretic hormone, regulates water retention in the kidney and vasoconstriction through interactions with vasopressin V2 and V1a receptor subtypes, respectively. AVP is released into the bloodstream from the neurohypophysis following a variety of stimuli, including plasma osmolality, hypovolemia, hypertension, and hypoglycemia. The actions of OT are mediated largely via a single receptor subtype (OTR), which is distributed in the periphery and within the limbic CNS. In contrast to the OTR, there are three AVP receptor subtypes, V1a, V1b, and V2 receptors, each of which is a G-protein–coupled, seven-transmembrane domain receptor. The V2 receptor is localized in the kidney and is not found in the brain. The V1a receptor is distributed widely in the CNS and is thought to mediate most of the behavioral effects of AVP. The V1b receptor is concentrated in the anterior pituitary and in the hippocampus. Although its function in the brain is largely unknown, V1b knockout mice display decreased aggression and altered social behaviors.

Some parvocellular neurons in the paraventricular nucleus of the hypothalamus project to the median eminence, where AVP is released into the portal system and delivered to the anterior pituitary. Through interactions with V1b receptors located on corticotrophs in the adenohypophysis, AVP acts to potentiate the effects of CRF on ACTH secretion. AVP is colocalized with CRF in the parvocellular neurons of the paraventricular nucleus. Given the link between HPA axis dysregulation and depression, recent attention has been given to the possible relationship between AVP secretion and psychiatric disorders. Although alterations in CSF AVP concentrations have been reported in patients with major depression, bipolar disorder, schizophrenia, autism, anorexia, and AD, the findings are not as consistent as for CRF, and many discrepant reports have appeared. In a postmortem study, an increase in the number of paraventricular AVP neurons colocalized with CRF cells has been reported in depressed patients compared to controls. Selective, nonpeptide V1b receptor antagonists have been developed as putative antidepressants with mixed results. Microdialysis experiments have demonstrated that AVP is released within the CNS in response to stressful stimuli.

In addition to the hypophyseal OT and AVP systems, magno- and parvocellular hypothalamic and extrahypothalamic neurons produce OT and AVP and send projections to the forebrain and brainstem. The release of peptides from these neurons is generally independent of neurohypophysial release, and it should be noted that OT and AVP released into the bloodstream do not efficiently re-enter the brain due to their inability to traverse the blood–brain barrier. OT and AVP projections from the paraventricular nucleus to the brainstem regulate a host of autonomic functions. However, in the forebrain, these peptides are now known to regulate a number of processes, ranging from anxiety, learning, and memory to complex social behaviors.

Central OT produces clear anxiolytic effects in animal models. This is particularly evident during lactation in rats when OT results in a blunted behavioral and ACTH response to a stressful acoustic stimulus. In contrast, central AVP appears to exert anxiogenic effects. In animal models, OT has been most intensively studied for its role in facilitating specific, complex social behaviors. OT has been reported to facilitate female sexual behavior, increase social interest, and facilitate the onset of maternal behavior. For example, in parturient rats, the onset of maternal behavior is blocked by OT antagonists, whereas maternal behavior can be observed in virgin females after infusion of OT directly into the brain. Stimulating OT release enhances mouse maternal behavior in part by modulating excitatory/inhibitory balance in the auditory cortex, elevating the salience of pup vocalizations. Likewise, in sheep, mother–infant bonding is facilitated by OT infusions. Studies with OT knockout mice suggest that this peptide plays a specific role in the processing of socially salient stimuli. For example, OT knockout mice have normal nonsocial cognitive abilities but have a specific deficit in the ability to recognize previously encountered individuals, even though olfactory processing is intact. Studies in highly social, monogamous rodents suggest that OT is also involved in the formation of selective social attachments between mates as well as empathic responses toward distressed partners.

Furthermore, species differences in OTR expression patterns appear to correlate with species differences in social behavior in rodents. For example, monogamous prairie voles have high densities of OTRs in the striatum, while nonmonogamous species do not (see Fig. 1.8–1). Behavioral pharmacologic studies demonstrate that these ventral striatal receptors are critical for social bond formation. Ventral striatal OTRs mediate social reward in mice. Studies in humans suggest that OT plays a similar role in regulating human social cognition. Polymorphisms in the human OTR gene predict face recognition abilities, and intranasal OT delivery causes men to rate their partner as more attractive and enhances ventral striatal activity in response to viewing the partner's face. OT has also been demonstrated to modulate human–canine interactions. All of these findings have led to the hypothesis that OT is involved in the regulation of the social brain, suggesting that dysregulation of this peptide could potentially explain social deficits in certain psychiatric disorders such as autism. Several studies using intranasal delivery of OT, which may enter the brain via the compromised blood–brain barrier of the nasal epithelium, now confirm that this neuropeptide modulates brain function and cognition in humans. For example, intranasal OT enhances eye gaze and the ability to infer the internal states of others from subtle affective facial expressions. Imaging studies reveal that intranasal OT reduces amygdala activation and reduces the coupling of the amygdala to brainstem regions implicated in autonomic and behavioral manifestations of fear in response to fear-inducing visual stimuli. There is evidence that early-life experience also alters the OT system because women with a history of childhood abuse or neglect exhibit reduced CSF OT concentrations, and polymorphisms of the OTR have recently been shown to mediate vulnerability to depression and anxiety. In voles, individual variation in OTR signaling confers susceptibility or resilience to early-life neglect in relation to adult social attachment behaviors. OT system dysfunction has also been implicated in autism spectrum disorders (ASDs). Several studies have suggested that intranasal OT can improve several aspects of social cognition, including emotion recognition and social reciprocity, as well as enhance activation of the reward system in autistic subjects. The OT system is currently viewed as the most promising target for improving social cognition in psychiatric disorders. However, the efficacy of activating the OT system through intranasal OT administration is limited due to poor brain penetration of the OT peptide, and there are no useful small-molecule brain-penetrant OT agonists. There is potential, however, for enhancing OT signaling by pharmacologically evoking endogenous OT release, such as occurs following melanocortin agonist or MDMA administration.

Studies in rodents suggest that OT enhances the signal-to-noise neurotransmission of social signals through neuromodulation. This effect is thought to amplify the neural response to social signals (olfactory, auditory, and visual) and enhance the flow of social information across brain networks involved in valence, reward, and memory. Hence, OT signaling is thought to enhance the salience and reinforcing value of social stimuli. However, the effects of intranasal administration are transient due to peptidase activity. Thus, optimal therapeutic use of drugs targeting the OT system will likely be as adjunctive therapy with behavioral interventions.

Extrahypothalamic AVP-producing neurons in the extended amygdala are sexually dimorphic, with males having many more AVP-expressing neurons than females. These neurons project through the ventral forebrain to the lateral septum, where they form a dense plexus of AVP-containing fibers in males, much more so than in females. Castration diminishes this sex difference, and androgen treatment re-establishes the sexually dimorphic pattern. Thus, extrahypothalamic AVP is predicted to be involved in the regulation of sex-specific behaviors in males. Vasopressin has been reported to modulate a variety of behaviors in males, including anxiety, aggression, affiliation, and social attachment in several animal models. For example, infusion of AVP into the hamster brain stimulates territorial and aggressive behaviors within minutes of administration. Extending this observation to humans, one study reported that individuals with a history of violent tendencies have elevated levels of AVP in the CSF compared to nonviolent controls.

One of the most intriguing features of the AVP system is the species specificity in the behavioral effects of AVP. Consistent with this observation, the neuroanatomical localization of V1a AVP receptors is highly species specific, often with little overlap between even closely related species. In fact, the specific behavioral role of AVP seems to be correlated with the localization of V1a receptors in specific brain regions. For example, AVP facilitates affiliation and social attachment in monogamous mammals. In the prairie vole, a monogamous rodent, AVP has been identified as the neurochemical trigger that stimulates pair bonding between the male and its mate. Comparisons among monogamous rodents and closely related nonmonogamous species have revealed that species differences in social organization are associated with species differences in receptor distribution within the brain. In several monogamous species, including the prairie vole, the vasopressin V1a receptor subtype is abundant in the mesolimbic dopamine reward pathway. In contrast, this region has few V1a receptors in the nonmonogamous, asocial montane vole (see Fig. 1.8–1). Infusion of a V1a receptor antagonist directly into the ventral pallidum of the prairie vole completely blocks pair bonding. Thus, AVP released during mating facilitates social bonding by modulating the mesolimbic dopamine pathway in prairie voles but cannot do so in nonmonogamous species because of the lack of receptors in that region. Molecular analysis of the V1a receptor genes of these different species has revealed DNA sequences in the promoter of the gene that may be responsible for the differential distribution of the receptor in the brain and thus the differences in behavioral patterns. This variability in distribution across species, along with the association between expression patterns and behavior, has led to the hypothesis that individual differences in receptor expression due to individual variation in gene promoter elements could potentially have important behavioral consequences in humans. In fact, three separate genetic association studies now have reported associations between polymorphisms in the V1aR promoter and ASD. Thus, dysregulation of AVP and/or its receptor may represent a risk factor that contributes to the social cognition deficits in ASD.

There has been increasing attention to the AVP system as a potential biomarker or drug target for psychiatric disorders with impaired social function, such as ASD. Nonhuman primates with low sociality are reported to have decreased CSF AVP compared to more social individuals. In addition, a small study suggested that children diagnosed with ASD, or infants who later developed ASD symptoms, have lower CSF AVP concentrations compared to non-ASD controls. Finally, intranasal AVP administration enhanced social abilities in children with ASD. While this work needs replication, and the importance of context has not been explored, these results suggest that the AVP system may be an important target for improving social functioning in psychiatric disorders.

Neurotensin

NT was isolated, based on its hypotensive properties, from the bovine hypothalamus in 1973. The NT-neuromedin N gene was originally cloned from canine ileal mucosa, and complementary deoxyribonucleic acid (cDNA) probes constructed against this form were used to clone the rat gene. The rat gene contains four exon sequences separated by three introns and spans approximately 10.2 kilobases (see Fig. 1.8–2). In the rat, the NT-neuromedin N sequence is contained in the fourth exon, and the single copy of each peptide sequence is bounded and separated by Lys-Arg basic amino pairs. The human NT gene has been localized to chromosome 12 (12q21). In pheochromocytoma (PC-12) neurons in culture, the NT-neuromedin N gene is regulated by lithium, nerve growth factor, cyclic adenosine monophosphate (AMP) activators, and dexamethasone through their effects on a 5′ flanking promoter region. The distribution of the NT-neuromedin N mRNA is generally the same as described for NT-containing neuronal cell bodies, except in the hippocampus and subiculum, where few neurons stain immunohistochemically for NT, and yet an abundance of the NT-neuromedin N mRNA is found. NT-producing cells are found in the midbrain (ventral tegmental area [VTA] and, to a lesser extent, the substantia nigra [SN]), ventral striatum, extended amygdala, lateral septum, and arcuate nucleus. The actions of NT are mediated by three receptors, the NT_1, NT_2, and NT_3 receptor subtypes. The NT_1 and NT_2 receptors are seven-transmembrane, G-protein–coupled receptors, while NT_3 is a Type I amino acid receptor with a single transmembrane domain and is located intracellularly.

Although NT is found in a number of brain regions, it has been most thoroughly investigated in terms of its association with other neurotransmitter systems, particularly the mesolimbic dopamine system, and has gained interest in research on the pathophysiology of schizophrenia. There are several lines of evidence suggesting that NT and its receptors should be considered potential targets for pharmacologic intervention in this disorder. First, the NT system is positioned anatomically to modulate the neural circuits implicated in schizophrenia. Second, peripheral administration of antipsychotic drugs has been shown to consistently modulate NT systems. Third, there is evidence that central NT systems are altered in schizophrenic patients.

Although likely other neurotransmitter systems are involved, one prevalent model of the pathophysiology of schizophrenia is overactivity of the mesolimbic dopamine system. Within the midbrain, NT-producing neurons are found in the VTA and the SN. Within the VTA, NT is found in dense-core vesicles only in tyrosine hydroxylase–positive staining cell bodies, indicating colocalization with dopamine. These NT-dopamine cells project to the prefrontal cortex, striatum, amygdala, and lateral septum. A subset of those NT-dopamine cells projecting from the VTA to the prefrontal cortex also produces CCK. In contrast to the VTA, the NT-producing cells in the SN are tyrosine hydroxylase negative. In addition to the NT-producing cells, dense fibers in the VTA staining positive for NT and originating from projections from the forebrain do not contain tyrosine hydroxylase. The midbrain also expresses NT receptors, with the vast majority of NT receptor–containing neurons in the VTA being dopamine-positive neurons. NT-producing cells, fibers, and NT receptors are also located in the ventral striatum. Thus, NT is colocalized with DA in the mesolimbic dopamine system, and this system is, in turn, sensitive to NT modulation due to the presence of the NT receptors.

NT was first shown to interact with dopamine systems while undergoing characterization of its potent hypothermic and sedative potentiating activity. Subsequent work indicated that NT possessed many properties that were also shared by antipsychotic drugs, including the ability to inhibit avoidance, but not escape, responding in a conditioned active avoidance task; the ability to block the effects of indirect dopamine agonists or endogenous dopamine in the production of locomotor behavior; and the ability to elicit increases in dopamine release and turnover. Perhaps most importantly, both antipsychotic drugs and NT neurotransmission enhance sensorimotor gating. Sensorimotor gating is the ability to screen or filter relevant sensory input, deficits that may lead to involuntary flooding of indifferent sensory data. Increasing evidence suggests that deficits in sensorimotor gating are a cardinal feature of schizophrenia. Both dopamine agonists and NT antagonists disrupt performance on tasks designed to gauge sensorimotor gating. Unlike antipsychotic drugs, NT is not able to displace dopamine from its receptor. As noted above, NT is colocalized in certain subsets of dopamine neurons and is co-released with dopamine in the mesolimbic and medial prefrontal cortical dopamine terminal regions that are implicated as the site of dopamine dysregulation in schizophrenia. Antipsychotic drugs that act at dopamine D_2 and D_4 receptors increase the synthesis, concentration, and release of NT in those dopamine terminal regions but not in others. That effect of antipsychotic drugs in increasing NT concentrations persists after months of treatment and is accompanied by the expected increase in NT mRNA concentrations, as well as expression of the "immediate early gene" c-fos within hours of initial drug treatment. The altered regulation of NT expression by antipsychotic drugs apparently extends to the peptidases that degrade the peptide because reports have revealed decreased NT metabolism in rat brain slices 24 hours after the acute administration of haloperidol.

While NT is certainly important in schizophrenia, there are numerous other neuropeptides implicated as well. For example, cortical neuropeptides that localize to parvalbumin-positive cells, such as somatostatin and CCK, may also play a role. These cells express parvalbumin calcium-binding protein and coordinate cortical γ-aminobutyric acid (GABA) release. Coordinated GABA release is central to the generation of γ-wave rhythmicity. In schizophrenic patients, γ-oscillations appear to be impaired owing to dysfunctional inhibitory GABA circuits. Neuropeptides such as somatostatin and CCK are also released from GABA-ergic Martinotti and basket cells, respectively, and contribute to the γ-oscillation frequency. Both peptides are also reduced in schizophrenia, and likely contribute to the disease state and represent attractive targets for potential therapies for this disease

With regard to schizophrenia, decreased CSF NT concentrations have been reported in several populations of patients when compared to controls or other psychiatric disorders. Although treatment with antipsychotic drugs has been observed to increase NT concentrations in the CSF, it is not known whether this increase is causal or merely accompanies the decrease in psychotic symptoms seen with successful treatment. Postmortem studies have shown an increase in NT concentrations in the dopamine-rich Brodmann area 32 of the frontal cortex, but that result may have been confounded by premortem antipsychotic treatment. Other researchers have found no postmortem alterations in NT concentrations in a wide sampling of subcortical regions. Decreases in NT receptor densities in the entorhinal cortex have been reported in postmortem samples of schizophrenic patients. A critical test of the hypothesis that NT acts as an endogenous antipsychotic-like substance awaits the development of an NT receptor agonist that can penetrate the blood–brain barrier. Recently, small-molecule full, partial, and inverse NT receptor agonists have been reported, but no psychopharmacology trials have yet been reported.

Other Neuropeptides

A number of other neuropeptides have been implicated in the pathophysiology of psychiatric disorders. These include CCK, SP, galanin, and NPY, though numerous others are implicated, far too many to cover in this chapter. A brief overview of the potential involvement of several of these neuropeptides in therapeutics or psychiatric disorders is provided as exemplars.

CCK, originally discovered in the gastrointestinal (GI) tract, and its receptors (CCK_1 primarily in the GI tract and CCK_2 primarily in the CNS) are found in areas of the brain associated with emotion, motivation, and sensory processing (e.g., cortex, striatum, hypothalamus, hippocampus, and amygdala). CCK is often colocalized with dopamine in the VTA neurons that comprise the mesolimbic and mesocortical dopamine circuits. Like NT, CCK decreases dopamine release. Intravenous infusions of a CCK fragment have been reported to induce panic in healthy individuals, and patients with panic disorder exhibit increased sensitivity to the CCK fragment compared to normal controls. Pentagastrin, a synthetic CCK agonist, dose-dependently produced increased blood pressure, pulse, HPA activation, and physical symptoms of panic. Recently, a CCK receptor gene polymorphism has been associated with panic disorder.

The undecapeptide SP is localized in the amygdala, hypothalamus, periaqueductal gray, LC, and parabrachial nucleus and is colocalized with norepinephrine and serotonin. SP serves as a pain neurotransmitter, and administration to animals elicits behavioral and cardiovascular effects resembling the stress response. More recent data suggest a role of SP in major depression and PTSD. Both depressed and PTSD patients exhibit elevated CSF SP concentrations. Furthermore, in PTSD patients, marked increases in CSF SP were detected following the precipitation of PTSD symptoms. One study has indicated that an SP receptor (termed the neurokinin 1 or NK1 receptor) antagonist capable of passing the blood–brain barrier is more effective than placebo and as effective as paroxetine in patients with major depression with moderate-to-severe symptom severity, though subsequent studies failed to confirm these findings, including in patients with PTSD and/or comorbid alcohol use. SP is present in sensory neurons and the spinal dorsal horn, and antagonism of

SP at the NK1 receptor was examined as a new type of analgesic. It failed in this realm but was serendipitously found to benefit in the treatment of chemotherapy-induced nausea, and several NK1 antagonists, with aprepitant being the first, are FDA approved for this indication.

NPY is a 36 amino acid peptide found in the hypothalamus, brain stem, spinal cord, and several limbic structures and is involved in the regulation of appetite, reward, anxiety, and energy balance. NPY is colocalized with serotonergic and noradrenergic neurons and is thought to facilitate the containment of negative effects following exposure to stress. Suicide victims with a diagnosis of major depression have been reported to exhibit a pronounced reduction in NPY levels in the frontal cortex and caudate nucleus. Furthermore, CSF NPY levels are decreased in depressed patients. Chronic administration of antidepressant drugs increases NPY concentrations in the neocortex and hippocampus in rats. Plasma NPY levels were found to be elevated in soldiers subjected to the “uncontrollable stress” of interrogation, and NPY levels were correlated with the feeling of dominance and confidence during the stress. In addition, low NPY response to stress has been associated with increased vulnerability to depression and PTSD. Remarkably, only a few scattered efforts have focused on developing NPY receptor modulators for psychiatric purposes. Although NPY itself is an anxiolytic-like substance, a synthetic compound that blocks the Y2 receptor (Y2R) will likely also be an anxiolytic because the NPY Y2R is located presynaptically and negatively regulates NPY release in the brain. Consistent with this hypothesis, Y2R null mice have reduced anxiety-like behavior. There are currently only two small-molecule NPY Y2R antagonists, and both are in preclinical development. Thus, this hypothesis remains to be tested in humans.

Galanin, first isolated from porcine intestine, is a 30 amino acid–containing peptide in humans interacting with three receptor subtypes, GalR1, GalR2, and GalR3. Preclinical studies in rodents show that galanin co-released with norepinephrine by noradrenergic cell bodies with the locus ceruleus, the main source of noradrenergic neurons within the brain, acts on GalR1somatodendritic autoreceptors to hyperpolarize and inhibit further noradrenergic cell firing. It is hypothesized that this helps prevent locus ceruleus overexcitation and may be a source of resilience to stressors. Unlike rodents, the GalR3 receptor is most prominent in humans and is hypothesized to play a similar role in both the locus ceruleus and dorsal raphe (the main source of serotonergic cell bodies). The GalR2 receptor is thought to act as a presynaptic autoreceptor in noradrenergic terminal fields, again decreasing norepinephrine release. Galanin also appears to potently enhance norepinephrine α_2 autoreceptor inhibition of LC firing. If inhibition of the noradrenergic or serotonergic system is too strong, depression may be the result. A postmortem study in human brain observed upregulation of galanin and GalR3 mRNA in the locus ceruleus of male and female subjects that completed suicide. Genetic variants for galanin and its receptors have been reported to confer an increased risk of anxiety and depression in individuals with adverse early life experiences or recent severe stressors. A small-molecule GalR3 antagonist, SNAP37889, proposed to disinhibit both noradrenergic and serotonergic pathways, was found to have significant *in vitro* toxicity, and clinical trials to test this therapeutic hypothesis have not proceeded.

Orexin A (hypocretin 1) and orexin B (hypocretin 2) are 33 and 28 amino acid–containing peptides, respectively. They possess approximately 50% sequence identity and are cleaved from the same preprohormone, and interact at the OX_1 and OX_2 G-protein–coupled receptors. Orexins are synthesized in no more than 15,000 to 20,000 neurons in the lateral and posterior hypothalamus but project widely throughout the brain and promote wakefulness. Preclinical studies originally observed that orexins stimulated feeding by delaying normal satiety mechanisms. A loss of function mutation in the orexin receptor was found to be the cause of canine narcolepsy, and this was recapitulated in a knockout mouse model. Subsequently, many humans with narcolepsy have been found to possess markedly low CSF concentrations of orexins. This is presumably due to autoimmune selective destruction of orexin neurons, although the initiating events are not clear. Therapeutic development of small-molecule orexin receptor ligands for sleep, alertness, obesity, and addiction (preclinical) has progressed, and two high-affinity, nonselective orexin receptor antagonists, suvorexant and lemborexant, are FDA approved for the treatment of insomnia.

FUTURE DIRECTIONS

Our current understanding of the role of neuropeptide systems in psychiatric disorders is derived primarily from correlational studies in human samples (e.g., CSF peptide concentrations or postmortem analyses), which preclude inferences of causality, or from animal models, which may or may not accurately reflect psychopathology. The inability to directly modulate CNS neuropeptide receptor activity in human subjects is a major impediment to the direct examination of the role of neuropeptide systems in human psychopathology. Considerable effort is being devoted to the development of small-molecule nonpeptide drugs that readily pass the blood–brain barrier and selectively modulate CNS peptide receptor activity. Small-molecule agonists or antagonists for several neuropeptide systems, including CRF, OT, AVP, orexin, and SP, are the subject of preclinical and clinical investigations. Over the next decade, these new pharmacologic tools will likely contribute significantly to our understanding of the roles of these peptides in both normal human behavior and various psychopathologies. Small-molecule drugs targeting neuropeptide receptors will undoubtedly lead to novel pharmacotherapy approaches for the treatment of psychiatric disorders such as anxiety disorders, depression, and ASDs. Small-molecule agonists or antagonists will also likely lead to the development of novel PET ligands, allowing visualization of peptide receptors in the CNS of human subjects, a great unmet need.

In addition to drug development and novel brain imaging tools, advances in psychiatric genetics are likely to reveal novel relationships between neuropeptide systems and psychopathology over the next few years. Polymorphisms in several neuropeptide receptor systems have already been implicated as risk factors in psychiatric disorders. Combining brain imaging techniques with genetic analyses will aid in understanding how these polymorphisms affect brain functioning. The plethora of recent GWAS studies in patients with mood and anxiety disorders and schizophrenia will likely identify novel neuropeptides associated with disease vulnerability. Finally, psychopharmacogenomics, which examines how genotype influences clinical responses to drugs, may lead to individualized therapies targeting peptide systems based on the patient's genotype. This burgeoning field has been helped along with technology to be able to routinely fully sequence a patient genome but still needs to be broadly informed by examining patient cohorts in order to identify relevant genotypes.

Clearly, we are (still) just beginning to understand the complexity of the brain's rich neuropeptide systems and their contribution to mental health. The unprecedented ability to manipulate neuropeptide circuits with remarkable anatomical and temporal precision is revealing the circuit-based neuropeptide-behavior relationships at a phenomenal pace. The significance of therapeutic contexts of pharmacologic manipulations is only beginning to be considered. This area of research will continue to provide novel insights into the biologic basis of psychopathology over the next few decades and will likely produce the next generation of pharmacologic interventions for psychiatric disorders.

FURTHER READINGS

Andari E, Hurlemann R, Young LJ. A precision medicine approach to oxytocin trials. *Curr Top Behav Neurosci*. 2018;35:559–590.

Binder EB, Nemeroff CB. The CRF system, stress, depression and anxiety-insights from human genetic studies. *Mol Psychiatry*. 2010;15(6):574–588.

Boules MM, Fredrickson P, Muehlmann AM, Richelson E. Elucidating the role of neurotensin in the pathophysiology and management of major mental disorders. *Behav Sci (Basel)*. 2014;4(2):125–153.

Brothers SP, Saldanha SA, Spicer TP, et al. Selective and brain penetrant neuropeptide y y2 receptor antagonists discovered by whole-cell high-throughput screening. *Mol Pharmacol*. 2010;77(1):46–57.

Brothers SP, Wahlestedt C. Therapeutic potential of neuropeptide Y (NPY) receptor ligands. *EMBO Mol Med*. 2010;2(11):429–439.

Bunevicius R. Thyroid disorders in mental patients. *Curr Opin Psychiatry*. 2009;22(4):391–395.

Burkett JP, Andari E, Johnson ZV, Curry DC, de Waal FBM, Young LJ. Oxytocin-dependent consolation behavior in rodents. *Science*. 2016;351:375–378.

Cáceda R, Kinkead B, Nemeroff CB. Involvement of neuropeptide systems in schizophrenia: human studies. *Int Rev Neurobiol*. 2007;78:327–376.

Coccaro EF, Lee R, Owens MJ, Kinkead B, Nemeroff CB. Cerebrospinal fluid substance P-like immunoreactivity correlates with aggression in personality disordered subjects. *Biol Psychiatry*. 2012;72(3):238–243.

Cohen BM, Sommer BR, Vuckovic A. Antidepressant-resistant depression in patients with comorbid subclinical hypothyroidism or high normal TSH levels. *Am J Psychiatry*. 2018;175:598–604.

DeMayo MM, Young LJ, Hickie IB, Song YJC, Guastella AJ. Circuits for social learning: a unified model and application to autism spectrum disorder. *Neurosci Biobehav Rev*. 2019;107:388–398.

Dunlop BW, Binder EB, Iosifescu D, et al. Corticotropin-releasing factor receptor 1 antagonism is ineffective for women with posttraumatic stress disorder. *Biol Psychiatry*. 2017;82:866–874.

Fliers E, Alkemade A, Wiersinga WM, Swaab DF. Hypothalamic thyroid hormone feedback in health and disease. *Prog Brain Res*. 2006;153:189–207.

Ford CL, Young LJ. Translational opportunities for circuit-based social neuroscience: advancing 21st century psychiatry. *Curr Opin Neurobiol*. 2021;68:1–8.

Froemke RC, Young LJ. Oxytocin, neural plasticity, and social behavior. *Annu Rev Neurosci*. 2021;44:359–381.

Gordon I, Vander Wyk BC, Bennett RH, et al. Oxytocin enhances brain function in children with autism. *Proc Natl Acad Sci U S A*. 2013;110(52):20953–20958.

Green JJ, Hollander E. Autism and oxytocin: new developments in translational approaches to therapeutics. *Neurotherapeutics*. 2010;7(3):250–257.

Griebel G, Holsboer F. Neuropeptide receptor ligands as drugs for psychiatric diseases: the end of the beginning? *Nat Rev Drug Discov*. 2012;11(6):462–478.

Hökfelt T, Barde S, Xu ZQD, et al. Neuropeptide and small transmitter coexistence: fundamental studies and relevance to mental illness. *Front Neural Circuits*. 2018;12:106.

Johnson ZV, Young LJ. Oxytocin and vasopressin neural networks: implications for social behavioral diversity and translational neuroscience. *Neurosci Biobehav Rev*. 2017;76:87–98.

Kamath J, Yarbrough GG, Prange AJ Jr, Winokur A. The thyrotropin-releasing hormone (TRH)-immune system homeostatic hypothesis. *Pharmacol Ther*. 2009;121(1):20–28.

Lemos JC, Wanat MJ, Smith JS, et al. Severe stress switches CRF action in the nucleus accumbens from appetitive to aversive. *Nature*. 2012;490(7420):402–406.

Myers AJ, Williams L, Gatt JM, et al. Variation in the oxytocin receptor gene is associated with increased risk for anxiety, stress and depression in individuals with a history of exposure to early life stress. *J Psychiatr Res*. 2014;59:93–100.

Nemeroff CB, Vale WW. The neurobiology of depression: inroads to treatment and new drug discovery. *J Clin Psychiatry*. 2005;66(Suppl 7):5–13.

Neumann ID, Landgraf R. Balance of brain oxytocin and vasopressin: implications for anxiety, depression, and social behaviors. *Trends Neurosci*. 2012;35(11):649–659.

Oztan O, Garner JP, Constantino JN, Parker KJ. Neonatal CSF vasopressin concentration predicts later medical record diagnoses of autism spectrum disorder. *Proc Natl Acad Sci U S A*. 2020;117:10609–10613.

Oztan O, Garner JP, Partap S, et al. Cerebrospinal fluid vasopressin and symptom severity in children with autism. *Ann Neurol*. 2018;84:611–615.

Pape JC, Carrillo-Roa T, Rothbaum BO, et al. DNA methylation levels are associated with CRF1 receptor antagonist treatment outcome in women with post-traumatic stress disorder. *Clin Epigenetics*. 2018;10:136.

Parker KJ, Oztan O, Libove RA, et al. A randomized placebo-controlled pilot trial shows that intranasal vasopressin improves social deficits in children with autism. *Sci Transl Med*. 2019;11(491):eaau7356.

Rakofsky JJ, Holtzheimer PE, Nemeroff CB. Emerging targets for antidepressant therapies. *Curr Opin Chem Biol*. 2009;13(3):291–302.

Strand FL. Neuropeptides: general characteristics and neuropharmaceutical potential in treating CNS disorders. *Prog Drug Res*. 2003;61:1–37.

Walum H, Young LJ. The neural mechanisms and circuitry of the pair bond. *Nat Rev Neurosci*. 2018;19:643–654.

Young LJ, Barrett CE. Neuroscience. Can oxytocin treat autism? *Science*. 2015;347:825–826.

1.9 Novel Neurotransmitters

Thomas W. Sedlak, M.D., Ph.D., Bindu D. Paul, Ph.D., and Adam I. Kaplin, M.D., Ph.D.

Neurotransmitters are chemicals that amplify or inhibit the depolarization signal from one neuron to that of an adjacent neuron. A neurotransmitter is typically released from a *presynaptic neuron* and travels across a small space, the *synaptic cleft* or *synapse*, to act upon the *postsynaptic neuron*. Neurotransmitters may be excitatory (e.g., glutamate) in that they promote depolarization of the target cell or inhibitory by promoting hyperpolarization (e.g., most GABA). Neurotransmitter receptors that are ion channels typically elicit depolarization via sodium or calcium, and hyperpolarization via chloride. Ion channel receptors differ from metabotropic receptors, the latter producing second messengers that have complex effects upon target cells, such as altering gene expression or modulating the propensity to depolarize.

An *action potential* is a wave of depolarization that travels down a neuronal axon to the *presynaptic terminal*, a specialized appendage where neurotransmitters are stored in specialized *vesicles*. The action potential opens voltage-sensitive calcium channels in the membrane, allowing for an increase in cellular calcium that results in vesicles releasing their contents into the synaptic cleft and acting upon receptors on the postsynaptic neuron membrane. When binding to their targets, neurotransmitters are typically full or partial agonists, and therapeutic drugs range from partial to full agonists to antagonists or inverse agonists (Fig. 1.9–1).

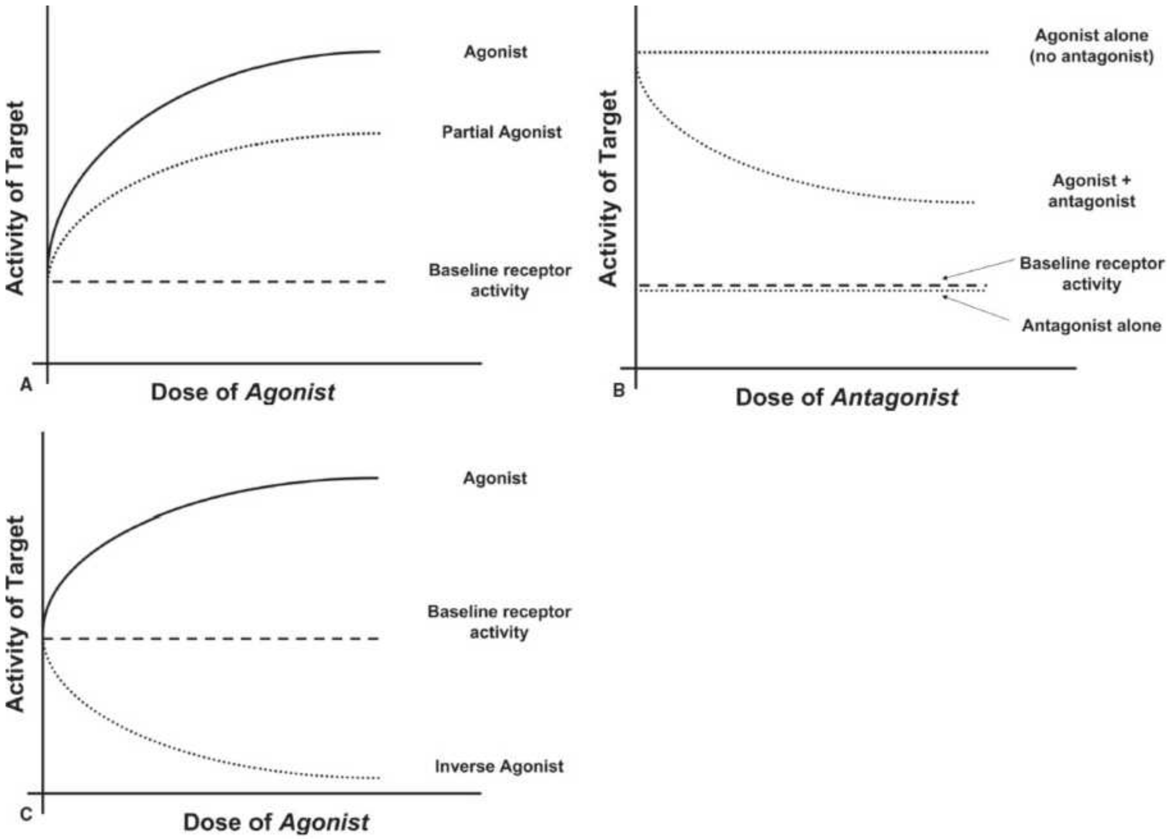

FIGURE 1.9–1. Agonists, antagonists, partial agonists, and inverse agonists. **A:** *Agonist* drugs bind to a target, such as a neurotransmitter receptor, and increase its activity until the receptors saturate. A *partial agonist* increases the activity of its target, but to a level below that of its maximum. In some cases, a partial agonist leads to diminished overall activity of a neurotransmitter receptor because it competes with the full agonist neurotransmitter. Occasionally a drug considered to be an agonist becomes reclassified as a partial agonist once a stronger agonist is found. THC was once considered a full agonist for the CB1 receptor, but was found to be a partial agonist after the discovery of the more potent synthetic cannabinoids, CP55,940 and WIN55,212-2. **B:** An *antagonist* has no intrinsic activity for activating or inhibiting a receptor. It acts to inhibit the activity of an agonist, in many cases by competing with an agonist for the binding site. **C:** An *inverse agonist* inhibits the activity of its target to a level below that of its baseline level with no drug present. Rimonabant is a CB1 receptor inverse agonist. It can block the baseline activity of the receptor even in the absence of cannabinoid agonists.

The definition of what a neurotransmitter is and is not has changed over the decades. The neurotransmitters initially discovered were small chemicals, first acetylcholine and later the biogenic amines such as serotonin, dopamine, norepinephrine, epinephrine, and histamine. Later it was found that amino acids and peptides could act as neurotransmitters, such as enkephalin acting upon the opiate receptor. By the 1990s it became apparent that the neurotransmitter acting on cannabinoid receptors was derived from cellular lipids. Furthermore, even a gas, nitric oxide, could be a neurotransmitter, bypassing membrane receptors and acting directly within postsynaptic neurons.

GASES AS NEUROTRANSMITTERS

Nitric Oxide

The discovery that gases could function as neurotransmitters revealed that highly atypical modes of signaling existed between neurons. In the early 1990s, nitric oxide was the first gas to be ascribed a neurotransmitter function, and proved to be an atypical neurotransmitter for several reasons. First, it was not stored in or released from synaptic vesciles. It was a small gas it could freely diffuse across cell membranes and into the target neuron. Second, its target was not a specific receptor on the surface of a target neuron, but intracellular proteins whose activity could directly be modulated. Nitric oxide also lacks a reuptake mechanism to remove it from the synapse. Although enzymatic inactivation of it is postulated to exist, nitric oxide is limited by a short half-life of a few seconds.

Nitric oxide was initially discovered as a bactericidal compound released from macrophages and as an endothelial cell-derived relaxation factor producing dilation of blood vessels. A role for nitric oxide in the brain followed, revealing a role for the gas in neurotransmission, learning and memory processes, neurogenesis, and neurodegenerative disease.

Synthesis of Nitric Oxide. Nitric oxide is chemically designated NO, with the dot representing that the molecule is a free radical, imparting a highly reactive nature. Nitric oxide is occasionally confused with nitrous oxide (N_2O), the gaseous anesthetic, and nitrogen dioxide (NO_2), a pollutant found in exhaust fumes, although these are not synthesized endogenously in mammals. A specific enzyme exists to generate nitric oxide within cells, nitric oxide synthase (Fig. 1.9–2). This enzyme generates nitric oxide by abstracting nitrogen from the amino acid arginine and reacting it with an oxygen atom. The enzyme utilizes nicotinamide adenine dinucleotide phosphate (NADPH) and generates citrulline as a byproduct.

Three distinct enzymatic forms of nitric oxide synthase exist, each with different locations and activation patterns within the body. *Neuronal nitric oxide synthase* (nNOS) was the first form discovered by Bredt and Snyder and is the predominant form in the brain. Neuronal NOS is expressed in neurons, especially those of the cortex, the dentate gyrus of the hippocampus, the corpus striatum, and the cerebellum. Although nNOS-containing neurons are only 1% of cortical neurons, their neuronal processes are so extensively distributed that almost all neurons make contact with an nNOS-containing nerve terminus. nNOS enzyme activity is markedly augmented by calcium levels via the accessory protein calmodulin. Thus, nitric oxide may be synthesized following neuronal depolarization, in which calcium levels transiently increase.

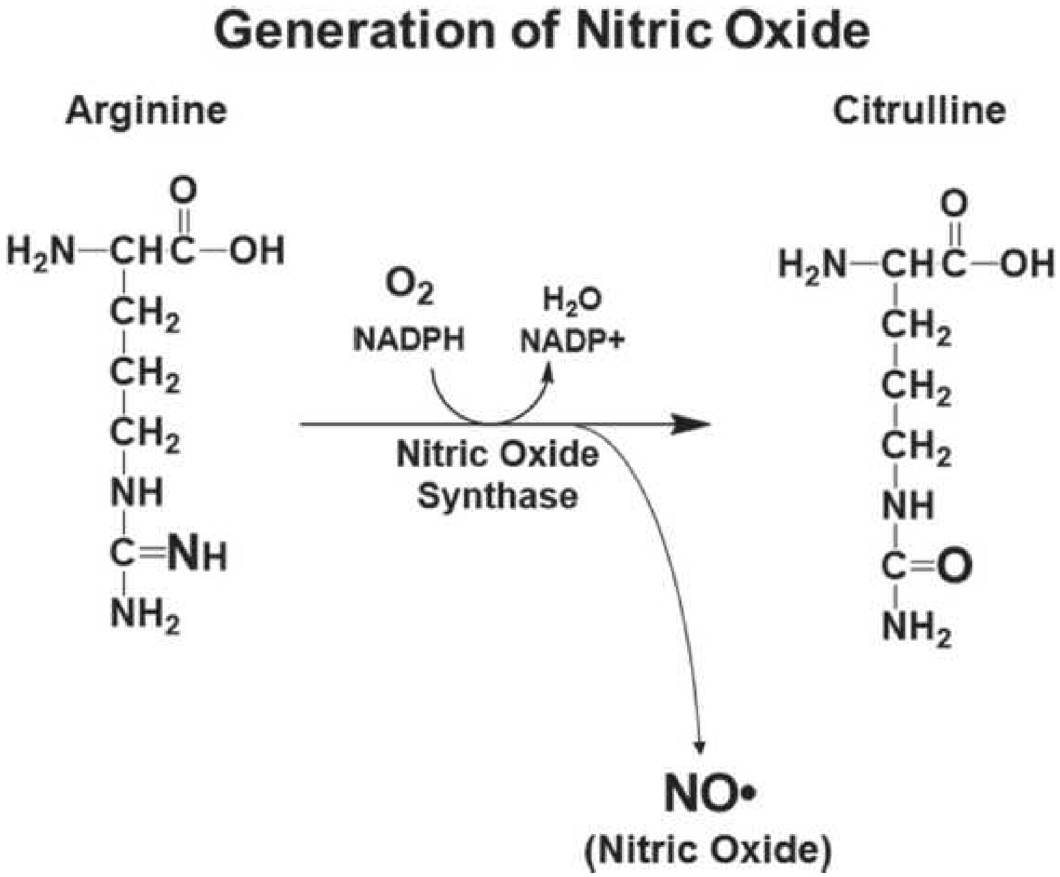
Generation of Nitric Oxide
Arginine
$H_2N-CH(CH_2CH_2CH_2NHC(=NH)NH_2)-C(=O)-OH$
O_2
NADPH
H_2O
NADP+
Nitric Oxide Synthase
Citrulline
$H_2N-CH(CH_2CH_2CH_2NHC(=O)NH_2)-C(=O)-OH$
NO•
(Nitric Oxide)

FIGURE 1.9–2. Enzymatic generation of nitric oxide. The gaseous neurotransmitter nitric oxide is generated by the enzyme nitric oxide synthase. The amino acid arginine is converted to citrulline and nitric oxide, employing oxygen and the reductant NADPH. Of note, nitric oxide is not preformed and stored in synaptic vesicles but is synthesized on demand.

Endothelial NOS (eNOS) is predominantly found in blood vessels, where it plays a crucial role in allowing for the relaxation and dilation of blood vessels. Nitroglycerine and sodium nitroprusside exert their vasodilatory effects via conversion to nitric oxide. eNOS activity is augmented by phosphorylation and increases in intracellular calcium. The third NOS, *inducible NOS* (iNOS), exists in many tissues in minuscule amounts. However, its levels are strongly increased by a great variety of cell stressors, especially inflammation. In the brain, it is largely induced in glial cells but also neurons.

Mechanism of Action of Nitric Oxide: Cyclic GMP Pathway. Long-term changes in brain function, such as learning and memory, involve a great variety of cellular processes, including changes in the patterns of gene and protein expression and the physical remodeling of neuronal architecture, such as dendrites. *Signal transduction* is the process by which extracellular signals, such as neuronal depolarization and receptor activation, lead to modified cellular function. Cyclic guanosine monophosphate (cGMP) is a prototypic intracellular messenger whose synthesis by guanylyl cyclase is stimulated after a membrane receptor is activated. cGMP then activates protein kinases that phosphorylate proteins as part of the cascade that elicits the biologic effect. As is the case with other neurotransmitters, such as serotonin, nitric oxide also activates cGMP production in neurons (Fig. 1.9–3). While a typical neurotransmitter activates guanylyl cyclase via a G-protein coupled to a membrane receptor, nitric oxide directly activates soluble guanylyl cyclase, the enzymatic form found in the cytoplasm. The active site of guanylyl cyclase contains a heme-group cofactor whose iron atom is bound by nitric oxide, leading to a protein conformation change and production of cGMP. Nitric oxide can also interact with the heme groups of other proteins, including hemoglobin/myoglobin, ferritin, and cytochrome P450.

Mechanism of Action of Nitric Oxide: S-Nitrosylation Pathway. A second method by which nitric oxide exerts its effects upon cells is the process of S-nitrosylation. In this signaling mechanism, nitric oxide directly modifies the sulfur atom of a protein cysteine residue, forming an S-nitrosothiol group (Fig. 1.9–4) that modifies the activity of select proteins. The generation of this modification requires no enzyme, though there is evidence that some proteins may carry the modification around to different areas of the cell and transfer it to a destination protein that elicits the target effect. Proteins that have been nitrosylated vary in their response to this modification; some are activated, and others are inactivated.

The number of protein targets of S-nitrosylation is rapidly expanding and includes molecules involved in signal transduction, programmed cell death, transcription factors, cytoskeletal proteins, ion pumps, and ion channels. In many cases modification of a single cysteine residue in a target protein is sufficient for nitric oxide to regulate its activity. Specific targets that are *activated* by S-nitrosylation include L-type calcium channels, calcium-activated potassium channels, and GABA-A receptors. Proteins *inhibited* by nitrosylation include several types of sodium channel, the N-methyl-d-aspartate (NMDA) subtype of the glutamate neurotransmitter receptor and several metabolic enzymes. S-Nitrosylation as a means of signal transduction is somewhat analogous to protein phosphorylation, as both are reversible covalent modifications that regulate protein function to change cell activity. S-Nitrosylation might participate in memory, learning, and behavior as many brain proteins are nitrosylated through the activity of neuronal nitric oxide generation.

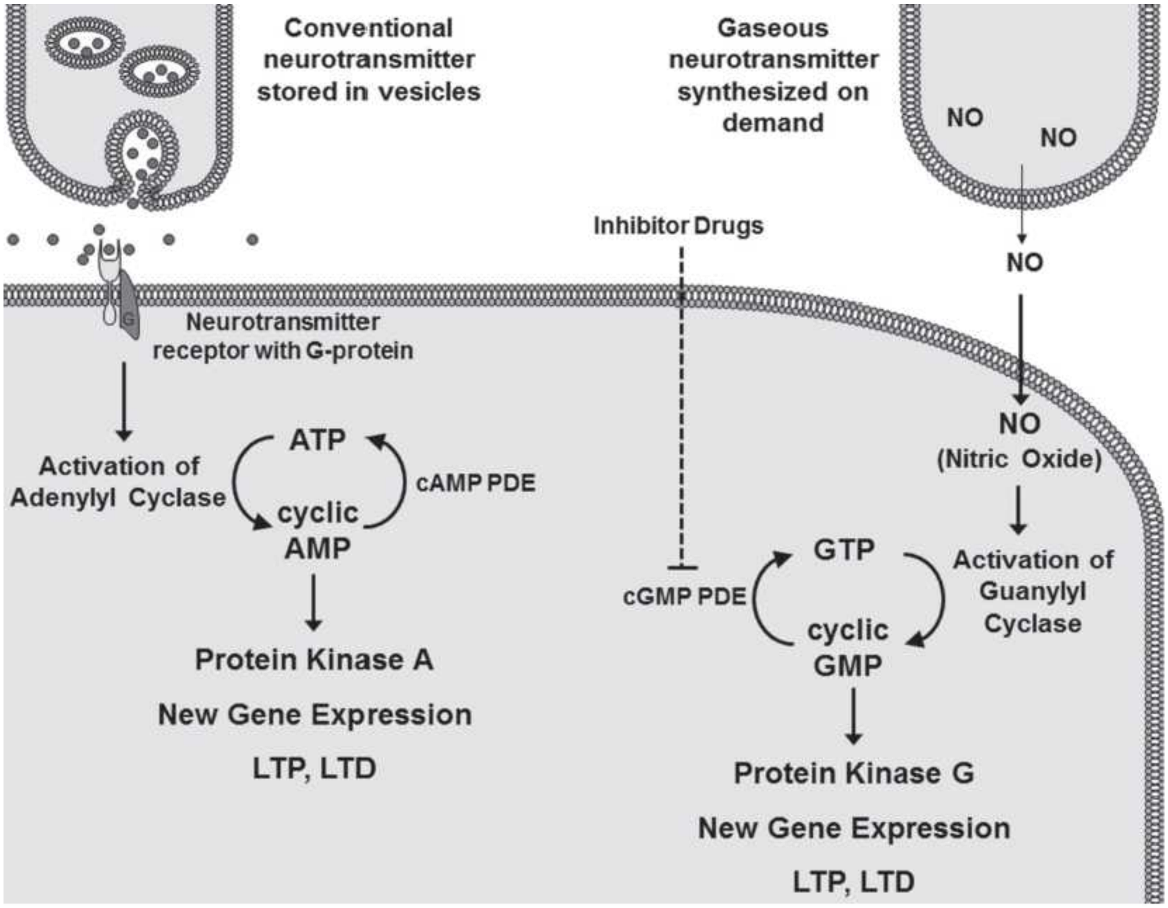
Conventional neurotransmitter stored in vesicles
Gaseous neurotransmitter synthesized on demand
NO
NO
Inhibitor Drugs
NO
G
Neurotransmitter receptor with G-protein
NO (Nitric Oxide)
Activation of Adenylyl Cyclase
ATP
cAMP PDE
cyclic AMP
Protein Kinase A
New Gene Expression
LTP, LTD
cGMP PDE
GTP
Activation of Guanylyl Cyclase
cyclic GMP
Protein Kinase G
New Gene Expression
LTP, LTD

FIGURE 1.9–3. Neurotransmitter and signaling functions of nitric oxide (NO) via production of cyclic GMP. Gaseous nitric oxide is generated enzymatically and freely diffuses into an adjacent neuron (**upper right**). In comparison to traditional neurotransmitters (**upper left**), NO does not act via a specific neurotransmitter receptor on the surface membrane of a neuron. In contrast, NO freely diffuses across the neuronal membrane and activates the enzyme guanylyl cyclase, which converts GTP into the second messenger, cyclic GMP. Nitric oxide effects are mediated, in part, by cyclic GMP activation of neuronal protein kinases, new gene expression, and effects on neuronal long-term potentiation (LTP) and long-term depression (LTD). ATP, adenosine triphosphate. PDE, phosphodiesterase.

Nitric Oxide and Neurotransmission. Long-term potentiation (LTP) is the process by which repetitive stimulation of a presynaptic neuron leads to stronger firing of a postsynaptic neuron, a process that underlies changes in learning and behavior. Induction of LTP depends on activation of postsynaptic NMDA receptors, while LTP maintenance relies on presynaptic mechanisms. Neurotransmission through the NMDA receptor facilitates LTP, in part, through the activity of nitric oxide. Activation of the

NMDA receptor leads to a cellular calcium increase, promoting nitric oxide synthesis and cGMP formation in the postsynaptic cell (Fig. 1.9–5).

Pharmacologic inhibitors of NOS have revealed deficits in LTP in rodent, bird, and honeybee models, and nitric oxide has been implicated in both short-term and long-term memory acquisition. Genetically modified mice lacking either nNOS or eNOS demonstrate no changes in LTP in the hippocampus, however, mice deficient in *both* nNOS and eNOS show decreased LTP in the CA1 region of the hippocampus. One form of NOS may compensate for the absence of the other, or the two may function cooperatively in LTP. Studies of the enteric nervous system have also revealed roles for nitric oxide in relaxation of the pyloric sphincter, and mice deficient in nNOS reveal a marked hypertrophy of the pylorus. Nitric oxide may also regulate monoaminergic neurotransmission. Inhibition of NOS in rats enhances the effects of cocaine and amphetamine, while the reverse is observed by increasing nitric oxide.

Nitric Oxide and Behavior. Nitric oxide neurotransmission may play a role in behavior as nNOS-deficient male mice display exaggerated aggressive tendencies and increased sexual activity. In female mice, the contrary is true, as they have reduced aggression.

In the periphery, nNOS localizes to neurons that innervate the blood vessels of the penis, including the corpus cavernosum. Stimulation of these nerves releases nitric oxide, leading to cGMP formation, blood vessel wall relaxation and vasodilatation, penile engorgement and initiation of erection. The sustained phase of erection also depends on nitric oxide; turbulent blood flow leads to phosphorylation of eNOS and sustained nitric oxide production. Drugs used in the treatment of erectile dysfunction, sildenafil (Viagra), tadalafil (Cialis), and vardenafil (Levitra), act to inhibit type 5 phosphodiesterase, an enzyme that degrades cGMP in the penis (see Fig. 1.9–3), thereby augmenting nitric oxide neurotransmission and penile erection.

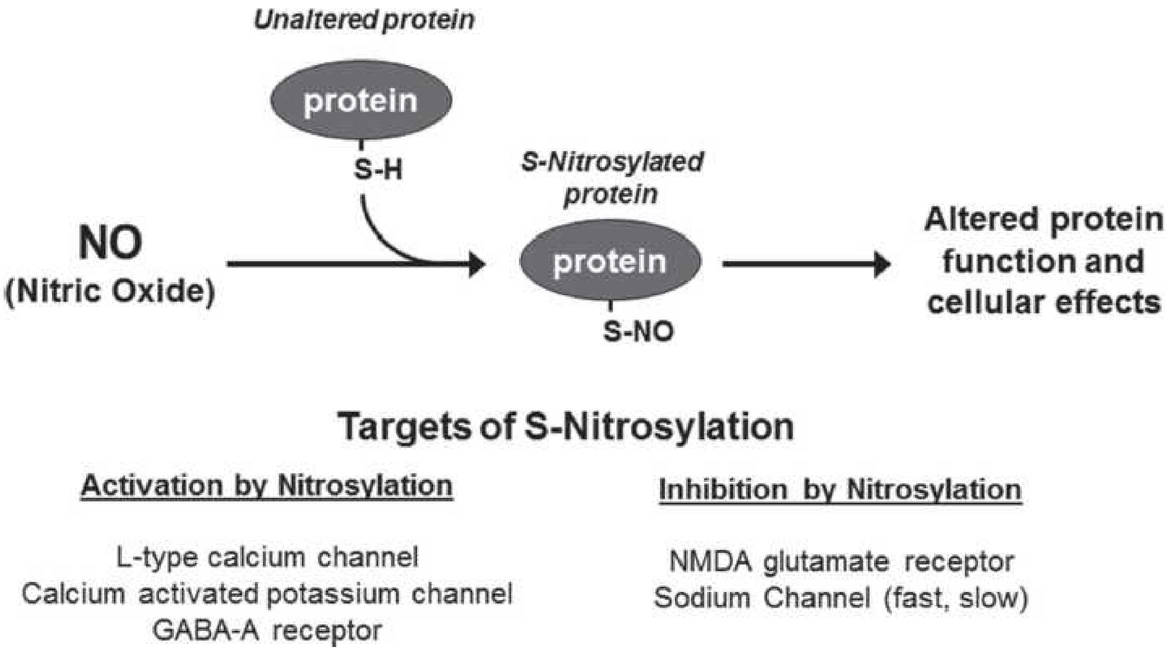
Unaltered protein
protein
S-H
NO
(Nitric Oxide)
S-Nitrosylated protein
protein
S-NO
Altered protein function and cellular effects
Targets of S-Nitrosylation
Activation by Nitrosylation
L-type calcium channel
Calcium activated potassium channel
GABA-A receptor
Inhibition by Nitrosylation
NMDA glutamate receptor
Sodium Channel (fast, slow)

FIGURE 1.9–4. Nitric oxide (NO) signaling via S-nitrosylation. In addition to NO activation of guanylyl cyclase, NO may also directly alter protein function via the process of S-nitrosylation. In this process, which does not require enzymatic catalysis, NO reacts with –SH groups of protein cysteine residues, resulting in an –SNO modification and altered protein function. Some proteins demonstrate robust activation following S-nitrosylation, whereas others are inhibited by the process.

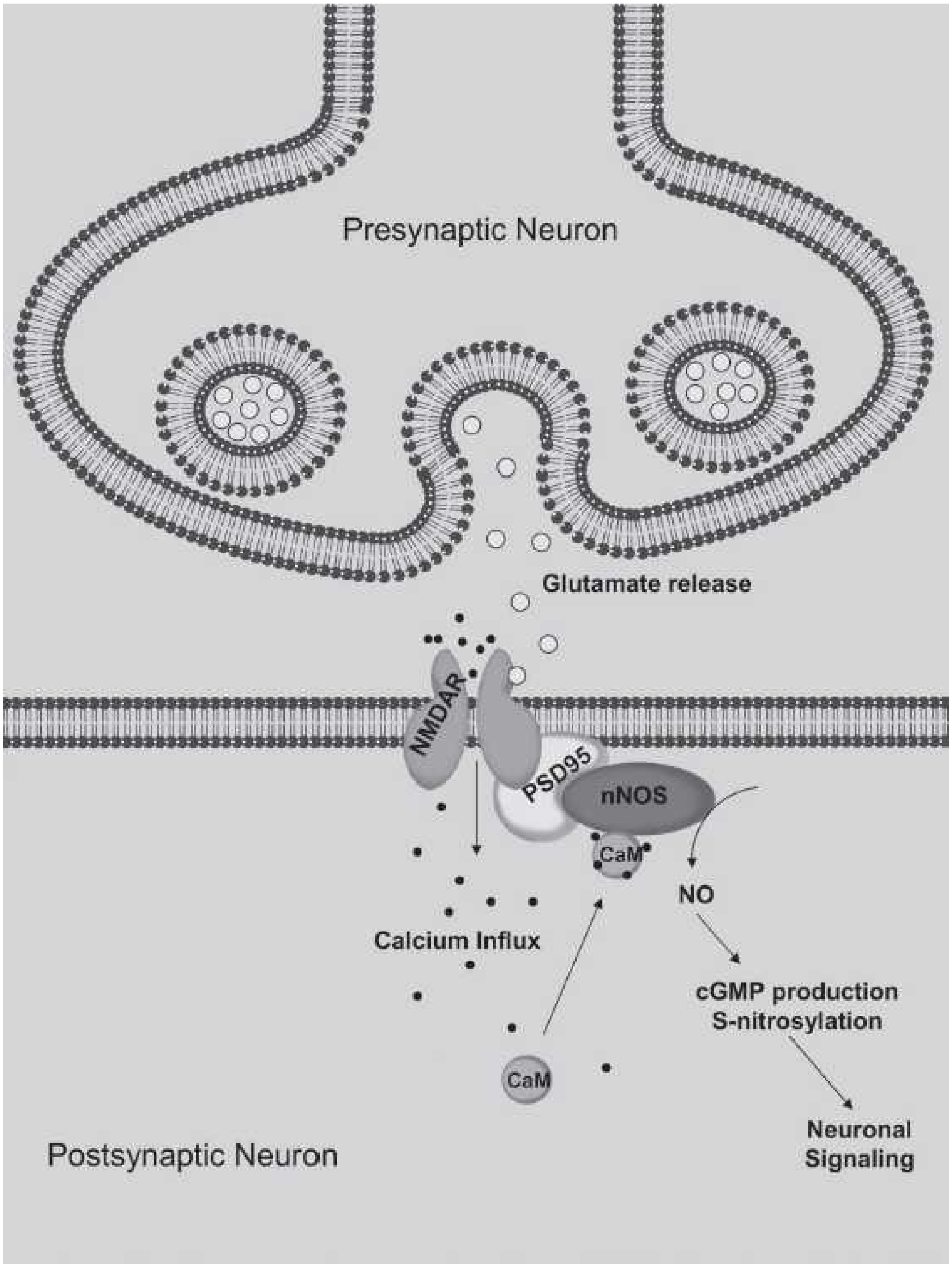
Presynaptic Neuron
Glutamate release
NMDAR
PSD95
nNOS
CaM
NO
Calcium Influx
cGMP production
S-nitrosylation
CaM
Neuronal
Signaling
Postsynaptic Neuron

FIGURE 1.9–5. Nitric oxide (NO) generation following NMDA receptor activation. The presynaptic neuron (**top**) releases glutamate (not shown), activating the NMDA glutamate receptor and allowing for calcium entry into the postsynaptic neuron. Calcium binds to the protein calmodulin (CaM), which in turn activates neuronal nitric oxide synthase (nNOS) to synthesize NO. A freely diffusible gas, NO exerts effects upon the target neuron via formation of cyclic GMP and S-nitrosylation (Figs. 1.9–3 and 1.9–4).

Numerous lines of evidence have suggested a role for nitric oxide in the regulation of sleep–wake cycles. nNOS expressing neurons occur in several areas that initiate REM sleep, including the pons, dorsal raphe nucleus, laterodorsal tegmentum, and pedunculopontine tegmentum. In animal models, microinjection of compounds that release nitric oxide results in decreased wakefulness and increased slow-wave sleep. Consistent with this, NOS inhibitors show a trend toward decreasing slow wave and rapid eye movement (REM) sleep. Studies of NOS-deficient mice suggest that nitric oxide may serve a more complex role than merely promoting sleep. nNOS-deficient animals also show reduced REM sleep; however, iNOS-deficient mice demonstrate the reverse, suggesting a complex interplay between NOS enzymatic isoforms.

Nitric Oxide and Mood Disorders. NOS-expressing neurons are well represented in brain areas implicated in depression, including the dorsal raphe nucleus and prefrontal cortex. A role for nitric oxide has been suggested in antidepressant response as selective serotonin reuptake inhibitor (SSRI) antidepressants can directly inhibit NOS activity. Moreover, in animal studies such as the forced swim test, NOS and soluble guanylyl cyclase inhibitors can achieve antidepressant-like effects. Plasma nitric oxide levels were elevated in patients with bipolar disorder compared to healthy control subjects. However, in depressed subjects, studies have found decreased nitric oxide levels and increased plasma nitrite, a byproduct of nitric oxide. Reduced NOS has also been described in the paraventricular nucleus of patients with schizophrenia and depression compared to controls.

Neurogenesis, the process by which new neurons are generated in the adult brain, is increasingly appreciated to participate in both mood disorder pathophysiology and antidepressant response. Increased hippocampal neurogenesis is associated with antidepressant response, and smaller hippocampal volume may be a risk factor for mood and anxiety disorders. Serotonin, itself, appears to promote neurogenesis in the hippocampus, while nitric oxide has been found to inhibit neurogenesis. Pharmacologic inhibitors of NOS result in increased serotonin and neurogenesis in the dentate gyrus of the hippocampus, a paramount site of this process. These NOS inhibitors also lead to an increase in serotonin in the dentate gyrus. Unsurprisingly, nNOS-deficient animals also manifest increased proliferation in the dentate gyrus. As steroids appear to induce NOS expression, nitric oxide may contribute to effects upon mood and anxiety often observed in those treated with these agents.

Nitric oxide has been interrogated for its ability to regulate neurotransmission at serotonin, norepinephrine and dopamine nerve termini. Nitric oxide appears to possess the capability of either increasing or decreasing activity at these neurons depending upon the specific brain region and timing of its activation.

Nitric Oxide and Schizophrenia. Nitric oxide has been investigated as a candidate molecule contributing to symptoms of schizophrenia. Genetic studies have identified schizophrenia-associated single nucleotide polymorphisms (SNPs) in CAPON, a protein that associates with nNOS. SNPs in nNOS have been inconsistently associated with schizophrenia. Changes in NOS levels have been reported in postmortem brain samples of individuals with schizophrenia. Abnormalities have been noted in the cortex, cerebellum, hypothalamus and brainstem, although no specific trend can be discerned. Elevated NOS activity has been noted in platelets from drug-naïve and drug-treated individuals with schizophrenia, but results have been inconsistent. Some investigators find increased nitric oxide activity and others the reverse. In autopsy samples, schizophrenic patients were found to have abnormally localized NOS expressing neurons in the prefrontal cortex, hippocampus and lateral temporal lobe, consistent with abnormal migration of these neuronal types during development. In a rat model, prenatal stress led to reduced NOS expressing neurons in the fascia dentata and hippocampus.

Neuropathologic Roles of Nitric Oxide. Abundant evidence exists that nitric oxide is a direct participant in a variety of neuropathic events. Superoxide, a byproduct of cellular metabolism, can react with nitric oxide to form peroxynitrite (chemical formula, $ONOO^-$). This labile and toxic compound forms chemical adducts with protein tyrosine residues, a process termed protein nitration, and DNA, leading to cellular dysfunction. Cell loss resulting from ischemic stroke is mediated in part by overstimulation of the glutamate NMDA receptor, a process termed *excitotoxicity*. Nitric oxide produced by NMDA activation appears to mediate a significant portion of this excitotoxic neuronal death, and stroke damage is reduced in mice with a genetic deletion of nNOS.

S-Nitrosylation has also been implicated in pathologic processes in the brain. Mutations in the Parkin protein are associated with early-onset Parkinson disease. Parkin is an E3 ubiquitin ligase, adding ubiquitin molecules to proteins and targeting them for destruction in the cell proteasome. In sporadic Parkinson disease, that is, without the early-onset mutation, nitric oxide can nitrosylate the Parkin protein and inhibit its protective E3 ubiquitin ligase function. An overabundance of nitric oxide signaling may thus predispose to the dysfunction and cell death of dopaminergic neurons in Parkinson disease by interfering with proteins essential for cell functioning. In Alzheimer disease, excess oxidation of brain protein, lipids, and carbohydrates has long been appreciated, but nitrosative stress from excess nitric oxide also appears to participate. Protein disulfide isomerase (PDI) is a cellular protective protein that may help combat the accumulation of misfolded proteins such as the amyloid fibrils occurring in the disease. In both Alzheimer and Parkinson disease brains, S-nitrosylation of PDI impedes its cellular protective function.

The discovery that nitric oxide participates in neurodegenerative processes raises the possibility for improved diagnostic processes, such as detecting damage to cellular components produced by nitric oxide prior to the onset of full-blown symptoms. In addition, drugs may be designed to attenuate the damage to crucial neuronal proteins that protect against disease onset. However, completely and nonspecifically inhibiting or stimulating nitric oxide synthesis is likely to produce significant side effects because of its wide-ranging activities throughout the body.

Nitric oxide is one of the most intensively studied compounds in the body, and possesses pleiotropic activities in different organs. While nitric oxide has physiologic roles in the brain, vasculature and immune system, it is a complex and incompletely understood molecule that also participates in disease. However, oxygen, which is a highly reactive molecule like nitric oxide, also possesses the capacity to contribute to disease pathogenesis, such as in oxidative damage, while still being essential for life.

Carbon Monoxide

Although carbon monoxide (chemical formula, CO) is most well known as an air pollutant derived from combustion reactions, it is produced physiologically in a great variety of organisms ranging from human to bacterium. While once thought to be a toxic byproduct of metabolic reactions, carbon monoxide is increasingly recognized to regulate a variety of physiologic processes in the brain and other organs. These varied effects include regulation of olfactory neurotransmission, blood vessel relaxation, smooth muscle cell proliferation, and platelet aggregation.

Carbon monoxide is far more well known for its toxic effects than its activities at physiologic concentrations. It binds tightly to heme molecules within hemoglobin, forming carboxyhemoglobin, which can no longer transport oxygen to tissues. One to two pack per day smokers typically have 3% to 8% of their hemoglobin as carboxyhemoglobin, with nonsmokers have less than 2%. Following acute carbon monoxide poisoning, 5% to 10% carboxyhemoglobin is associated with impaired alertness and cognition, and 30% to 50% carboxyhemoglobin leads to significant drops in oxygen transport to tissues.

Enzymatic Generation of Carbon Monoxide. Carbon monoxide is produced during the metabolism of heme by the action of *heme oxygenase* (HO). This enzyme utilizes oxygen, the reducing equivalents of NADPH and the electron donor cytochrome P450 reductase to break open the carbon ring of heme and release a one-carbon fragment as carbon monoxide (Fig. 1.9–6). The reaction also produces the green pigment, biliverdin, and free iron. Biliverdin is converted to the yellow pigment, bilirubin, which like carbon monoxide is no longer solely considered a toxic byproduct. At physiologic concentrations bilirubin is an enormously potent antioxidant which can be partly recycled back to its precursor, biliverdin.

Three forms of HO exist. HO1 is similar to iNOS in that it typically exists at very low levels, but its expression may be potently induced by a great variety of stimuli, ranging from oxidative stress, inflammation, dopamine, steroids, and growth factors. Indeed, HO1 is one of the most easily induced proteins known. HO2 is not an inducible protein and is predominantly expressed in the brain and testis. HO2 is expressed in discrete neuronal populations throughout the brain, including cortical and hippocampal pyramidal cells, dentate gyrus granule cells, the olfactory bulb, thalamus, hypothalamus, brainstem, and cerebellum. HO3 is an isoform whose significance is poorly understood.

Molecular Actions of Carbon Monoxide. Similar to nitric oxide, gaseous carbon monoxide can freely diffuse across membranes and directly activate soluble guanylyl cyclase, though it is approximately 30-fold less potent than nitric oxide. With the ensuing increase in cGMP, protein kinases are activated leading to some of the manifold effects of carbon monoxide upon cells. The expression of HO2 closely mirrors the expression pattern of guanylyl cyclase, implicating the two as part of a common pathway of neuronal signaling, and inhibitors of HO2 block the generation of cGMP. Similar to the case of nNOS, HO2 can be activated by calcium/calmodulin and phosphorylation, fulfilling the important criterion that a neurotransmitter to be rapidly released in response to neuronal depolarization. Carbon monoxide can also activate p38 MAP kinase, via a mechanism that does not require cGMP. This important kinase promotes a variety of cellular effects including inhibition of inflammation, cell proliferation, and programmed cell death (apoptosis).

Carbon Monoxide and Neurotransmission. Carbon monoxide appears to participate in the neurotransmission of odorant perception. Odorants lead to carbon monoxide production and subsequent cGMP synthesis that promotes long-term adaptation to odor stimuli. Carbon monoxide has the potential to regulate a variety of perceptual and cognitive processes that are yet untested. Similarly, in the rat retina, long periods of light exposure lead to increased HO1 expression, carbon monoxide production, and cGMP signaling. Carbon monoxide may also participate in adaptation to chronic pain. HO2-deficient animals manifest reduced hyperalgesia and allodynia after exposure to chronic pain stimuli. Carbon monoxide may thus tune the threshold for pain perception, although it is unclear whether the effect occurs in the central or peripheral nervous system (PNS). Aside from its role in promoting cGMP production, carbon monoxide may also directly bind to and open the BK_{Ca} channel.

In the gastrointestinal nervous system, carbon monoxide serves as a neurotransmitter to relax the internal anal sphincter in response to nonadrenergic noncholinergic (NANC) nerve stimulation and vasoactive intestinal peptide (VIP). Mice rendered genetically deficient of HO2 demonstrate a 50% reduction in NANC neurotransmission, as do nNOS-deficient animals. Mice that are bred to have loss of both HO2 and nNOS have their NANC neurotransmission abolished, establishing a physiologic process mediated entirely by gaseous neurotransmitters.

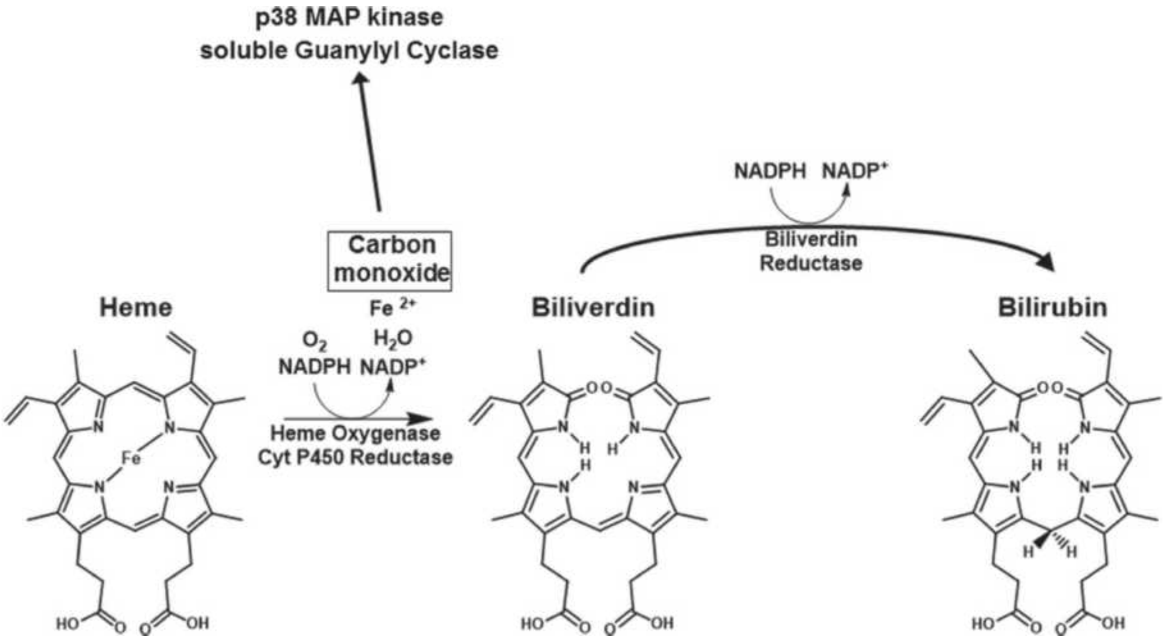

FIGURE 1.9–6. Synthesis of carbon monoxide, an atypical neurotransmitter. Gaseous carbon monoxide (CO) is enzymatically synthesized in neurons via the enzyme, heme oxygenase, also converting heme into the molecule biliverdin, and liberating free iron (Fe). Similar to nitric oxide, CO is not stored in neuronal vesicles and can freely diffuse across neuronal membranes. CO also similarly activates soluble guanylyl cyclase, and leads to activation of multiple intracellular signaling molecules such as p38 MAP kinase. CO exerts its neurotransmitter and signaling functions at concentrations far below that at which classical CO toxicity occurs. The significance of this pathway in neurons is underlined by the existence of two distinct heme oxygenase enzymes, one of which is predominantly expressed in the brain. Biliverdin is converted to bilirubin via the enzyme biliverdin reductase. Similar to CO, bilirubin is no longer relegated to the status of toxic byproduct and may be an important antioxidant.

Carbon monoxide has been implicated in the development of hippocampal LTP, although lines of evidence are contradictory. Carbon monoxide and tetanic stimulation of nerves leads to increased excitatory postsynaptic potentials (EPSPs). HO inhibitors that block carbon monoxide production lead to impaired induction of LTP and reduced calcium-dependent release of glutamate neurotransmitter. However, HO2-deficient animals fail to manifest any differences in LTP. These disparate findings may be explained by a role for HO1 in LTP, or an ability of HO inhibitors to nonspecifically block some other processes important to LTP induction.

Animals with loss of HO2 demonstrate a reduced fear of falling from a suspended wire and greater exploratory behavior in open-field testing. These animal model correlates may suggest a role for carbon monoxide signaling in anxiety states. Subtle abnormalities on memory have been noted in these HO2 knockout mice, although findings have been mixed. As HO2 is abundantly expressed in male testis and penile autonomic ganglia, it is unsurprising that carbon monoxide and nitric oxide both appear to regulate the male reproductive autonomic nervous system. HO2-deficient mice manifest abnormal neurotransmission of the myenteric plexus and diminished bulbospongiosus muscle reflex, leading to ejaculatory abnormalities.

Other Signaling Roles of Carbon Monoxide. The expression of HO2 in the hypothalamus led investigators to test whether carbon monoxide could regulate the release of peptide hormones. In animal models carbon monoxide was found to block the secretion of both oxytocin and vasopressin from the hypothalamus. Cell culture systems have also suggested that carbon monoxide inhibits corticotrophin-releasing factor (CRF) release from hypothalamic cells, but near toxic levels do the reverse, stimulating CRF release.

Heme metabolism and carbon monoxide production appear to be regulated in a circadian fashion, consistent with a role in the regulation of sleep–wake cycles. The mammalian transcription factor NPAS2 binds BMAL1 to form a complex (NPAS2/BMAL1) that, along with Clock/BMAL1, activates transcription of period and cryptochrome proteins. Period and cryptochrome have dual functions, inhibiting the circadian rhythm machinery, but also blocking NPAS2/BMAL1 and Clock/BMAL1, forming a periodic circuit. NPAS2 contains two heme moieties that can bind carbon monoxide, leading to an inhibition of the NPAS2/BMAL1 complex being able to bind to DNA and regulate transcription.

At toxic levels, carbon monoxide is well known to impair oxygen transport by binding to hemoglobin with higher affinity than oxygen. Amazingly, carbon monoxide itself plays a physiologic role in the mechanism by which the carotid body senses oxygen. HO, expressed in glomus cells of the carotid body, uses oxygen as a substrate in the production of carbon monoxide (see Fig. 1.9–6). When oxygen levels drop, so does carbon monoxide production, leading to a resetting of the threshold in which the carotid body senses oxygen. The molecular mechanism may occur via carbon monoxide regulation of the carotid body BK ion channel.

Neuroprotective Roles of the Heme Oxygenase Pathway. Mice rendered genetically deficient in HO2 manifest increased susceptibility to traumatic brain injury and stroke damage. The neuroprotective function of HO may be impaired in Alzheimer disease as HO is found in amyloid plaques. The amyloid precursor protein (APP), a source for toxic amyloid beta fragments, can bind to and inhibit HO neuroprotective function, and APP mutants associated with early-onset Alzheimer disease are the most potent at blocking HO function.

Hydrogen Sulfide: The Newest Gaseous Messenger Molecule

As carbon monoxide and bilirubin had reputations of toxicity prior to the appreciation of their physiologic functions, a similar tale is unfolding for the gas, hydrogen sulfide. While still abundantly recognized as a foul smelling and toxic emission of bacteria and sewage treatment plants, hydrogen sulfide (chemical formula, H_2S) may yet prove to be a significant neuromodulator and atypical neurotransmitter.

At least two different enzymes can generate hydrogen sulfide, cystathionine beta synthase (CBS) and cystathionine gamma lyase (CSE). Each catalyzes the same reaction converting cysteine and water to hydrogen sulfide, pyruvate, and ammonia. Intriguingly, each enzyme also catalyzes an independent reaction. CBS converts homocysteine and serine to cystathionine, which CSE uses to produce cysteine, ammonia, and 2-oxobutyric acid. Similar to nitric oxide, hydrogen sulfide also regulates blood pressure by dilation of arterioles. Mice deficient in CSE are hypertensive, and hydrogen sulfide treatments lower blood pressure.

In the brain, hydrogen sulfide exists at concentrations as high as 160 micromolar, consistent with a role in regulating brain function. CBS is abundantly expressed in brain, while CSE is undetectable. Similar to the enzymes that generate the other gaseous neurotransmitters, CBS is activated by calcium/calmodulin. CBS-deficient mice have altered hippocampal LTP. Hydrogen sulfide also potentiates NMDA receptor currents and has neuroprotective effects at low levels.

Environmental hydrogen sulfide will elicit varied dose-dependent effects. Humans detect the odor at 0.1 to 1 parts per million (ppm), and the olfactory nerve is paralyzed at 100 to 150 ppm. Pulmonary edema occurs at 320 to 530 ppm and 530 to 1,000 ppm produces strong central nervous system (CNS) activation, respiratory arrest, coma, and death. Extremely high levels occur in industrial chemical synthesis and petroleum refineries, with more modest levels in sewage-processing plants, geysers, hot springs, and geothermal sites.

ENDOCANNABINOIDS: FROM MARIJUANA TO NEUROTRANSMISSION

Whether known as cannabis, hemp, hashish, ma-fen, or a variety of slang terms, marijuana has been cultivated and utilized by human populations for thousands of years. Despite long debate as to whether its risks and benefits are evenly matched, is has only been in recent decades that some of the mystery has been revealed by which marijuana exerts its effects on the brain. The “high” users experience, euphoria and tranquility, relates to cannabis acting upon a neural pathway involving cannabinoids endogenous to the human brain, or endocannabinoids.

The first described medicinal use of cannabis dates to approximately 2700 BC in the pharmacopeia of Chinese Emperor, Shen Nung, who recommended its use for a variety of ailments. At this time, adverse properties were also apparent, and large amounts of the fruits of hemp could lead to “seeing devils” or a user might “communicate with spirits and lightens one’s body.” For centuries, cannabis was employed in India as an appetite stimulant, and habitual marijuana users remain well acquainted with “the munchies.”

For many years the mechanisms by which the active components of marijuana, *cannabinoids*, exerted their psychoactive effects remained a mystery. Chemists sought to isolate the psychoactive components of cannabis from the many components of the plant oil. First elucidated in 1940 was cannabinol, a compound now appreciated to be an oxidation product of other cannabinoids, and not a psychoactive compound (Fig. 1.9–7). However, with a structure in hand, chemists were now able to synthesize synthetic cannabinoids that did possess psychoactive properties. Within a few years it was apparent that tetrahydrocannabinols might be the active components of cannabis. Following advances in chemical separation techniques, Mechoulam and Gaoni in 1964 identified delta-9-tetrahydrocannabinol (THC), a compound that accounts for nearly all of the psychoactive effects of cannabis. THC-acid is the predominant form of plant THC, and this is readily converted to THC upon heating, such as when cannabis is smoked.

Discovery of the Brain Endocannabinoid System

Estimates suggest that 20 to 80 micrograms of THC reach the brain after one smokes a marijuana cigarette (i.e., “joint”). Consider that there are 100 to 200 micrograms of norepinephrine neurotransmitter present in the entire human brain. In the 1960s, there were at least two schools of thought on how THC exerted its psychoactive effects. One held that THC worked in a manner similar to the inhaled volatile anesthetics, that is, no specific receptor existed and it might have a generalized effect on neuronal membranes or widespread actions upon neurotransmitter receptors. A competing school hypothesized that specific receptors for cannabinoids existed in the brain, but they were difficult to identify due to the lipophilic nature of these chemicals. Novel cannabinoids were synthesized that were more water soluble, and in the late 1980s, this allowed for the discovery of a specific cannabinoid receptor, named CB1.

Now that a cannabinoid receptor was known to exist in the brain, the research community felt that it was unlikely that such receptors would have evolved solely to allow for action of plant cannabinoids. Indeed, after Pert and Snyder discovered opiate receptors in the 1970s, it was soon apparent that these did not evolve for the purpose of morphine drugs, but as the targets of the endogenous enkephalin neurotransmitters, discovered soon after. A hunt for the endogenous brain ligand for the CB1 receptor was underway, and the existence of such a substance was hypothesized to be an endogenous cannabinoid. This proved true in 1992 when Mechoulam and colleagues reported the discovery of anandamide, a lipid produced endogenously in the brain that could activate cannabinoid receptors and function as a neurotransmitter (Fig. 1.9–8). The name of this substance was derived from the Sanskrit word, *ananda*, which translates as bliss. Anandamide could also mimic THC in a variety of animal behavioral tests that generally predict whether a substance will have psychoactive properties in humans, including inhibition of spontaneous movement, promotion of freezing spells, reducing pain sensitivity, and decreasing body temperature.

Several additional endocannabinoids were soon discovered, 2-arachidonylglycerol (2-AG), N-arachidonyldopamine (NADA), 2-arachidonoylglycerol ether (noladin ether), and virodhamine (see Fig. 1.9–

8). The reason for having several different endocannabinoids may lie with their differing affinities for the cannabinoid receptors, CB1 and CB2. Anandamide appears to have the greatest selectivity for the CB1 receptor, followed by NADA and noladin ether. In contrast, virodhamine prefers CB2 receptors and has only partial agonist activity at CB1. 2-AG appears not to discriminate between CB1 and CB2.

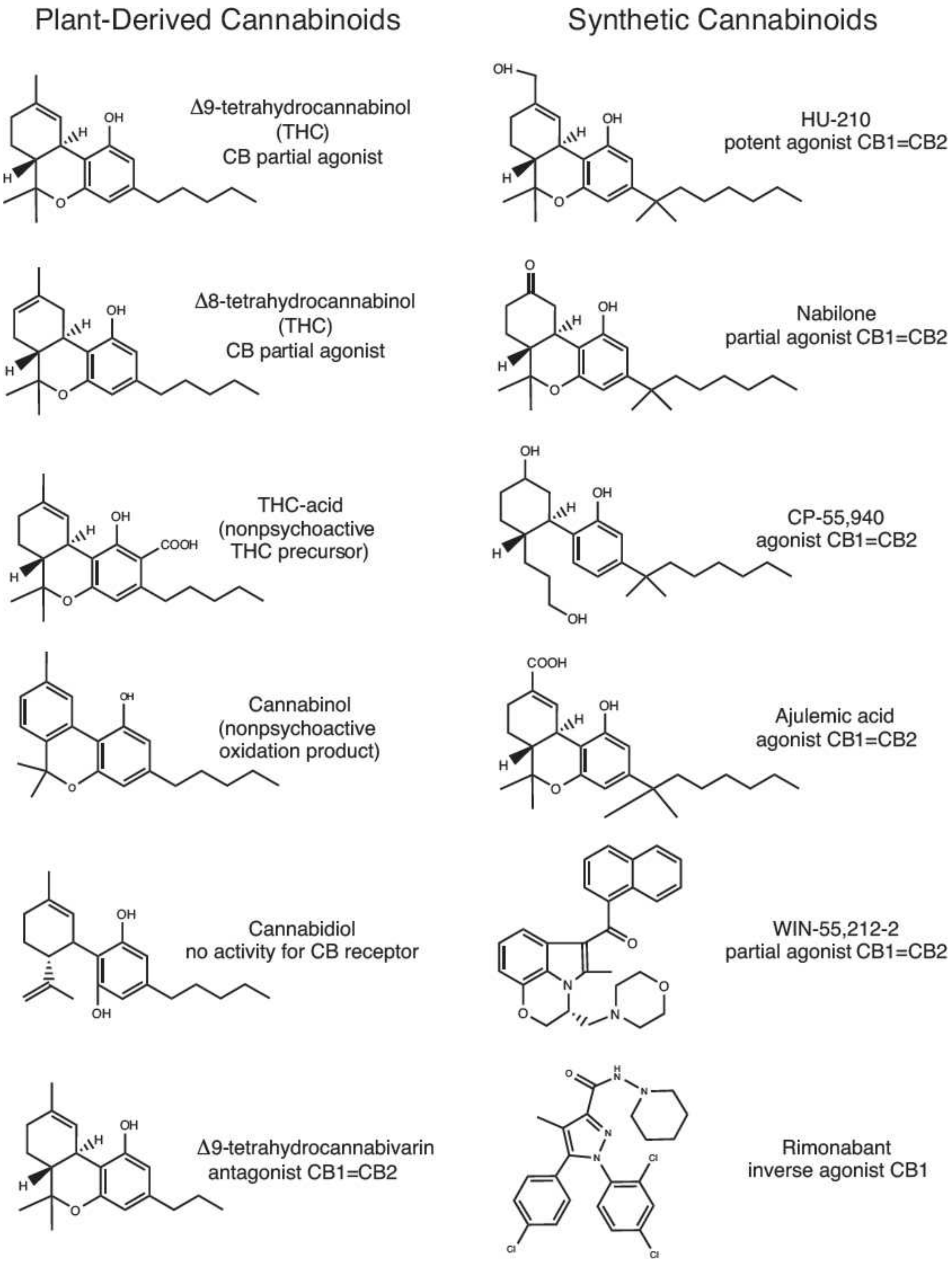
Plant-Derived Cannabinoids
Synthetic Cannabinoids
Δ9-tetrahydrocannabinol (THC) CB partial agonist
HU-210 potent agonist CB1=CB2
Δ8-tetrahydrocannabinol (THC) CB partial agonist
Nabilone partial agonist CB1=CB2
THC-acid (nonpsychoactive THC precursor)
CP-55,940 agonist CB1=CB2
Cannabinol (nonpsychoactive oxidation product)
Ajulemic acid agonist CB1=CB2
Cannabidiol no activity for CB receptor
WIN-55,212-2 partial agonist CB1=CB2
Δ9-tetrahydrocannabivarin antagonist CB1=CB2
Rimonabant inverse agonist CB1

FIGURE 1.9–7. Select plant-derived and synthetic cannabinoids. THC is the main psychoactive component of cannabis. The drug rimonabant (Acomplia) is a potent inverse agonist for cannabinoid CB1 receptors. FAAH, fatty acid amide hydrolase.

Biosynthesis of Endocannabinoids

Arachidonic acid is utilized as a building block for biosynthesis of endocannabinoids, prostaglandins, and leukotrienes, and is found within cellular phospholipids of the plasma membrane and other intracellular membranes. Synthesis of anandamide requires the sequential action of two enzymes (Fig. 1.9–9). In the first reaction, the enzyme NAT transfers an arachidonic acid side chain from a phospholipid to phosphatidylethanolamine (PE), generating NAPE. In the second reaction the enzyme NAPD-PLD converts NAPE to anandamide. As NAPE is already a natural component of mammalian membranes, it is the second step which generates anandamide that is most crucial to neurotransmission.

Biosynthesis of 2-AG also requires two enzymes. In the first step, a phospholipid containing arachidonic acid at the middle position is converted to sn-1-Acyl-2-arachidonyl glycerol (DAG) via the action of the enzyme phospholipase C. The second reaction generates 2-AG via either of two specific diacylglycerol lipases (DAGL). The enzymes for biosynthesis of the other endocannabinoids are undefined.

Endocannabinoids are not stored in synaptic vesicles for later use but are synthesized on demand as is done for the gaseous neurotransmitters. An important criterion for a signaling molecule to be considered a neurotransmitter is that neuronal depolarization should lead to its release. Depolarization leads to increases in cellular calcium, which in turn promotes synthesis of the endocannabinoids and their release. The mechanism is explained in part by calcium activation of NAPE-PLD and DAGL, leading to augmented biosynthesis of anandamide and 2-AG, respectively.

Endocannabinoids generated in a neuron must cross the synaptic cleft to act upon cannabinoid receptors. Like THC, endocannabinoids are highly lipophilic, and thus poorly soluble in cerebrospinal fluid. It is hypothesized that a specific endocannabinoid transporter exists to allow endocannabinoids to cross the synaptic cleft and allow entry into the target neuron.

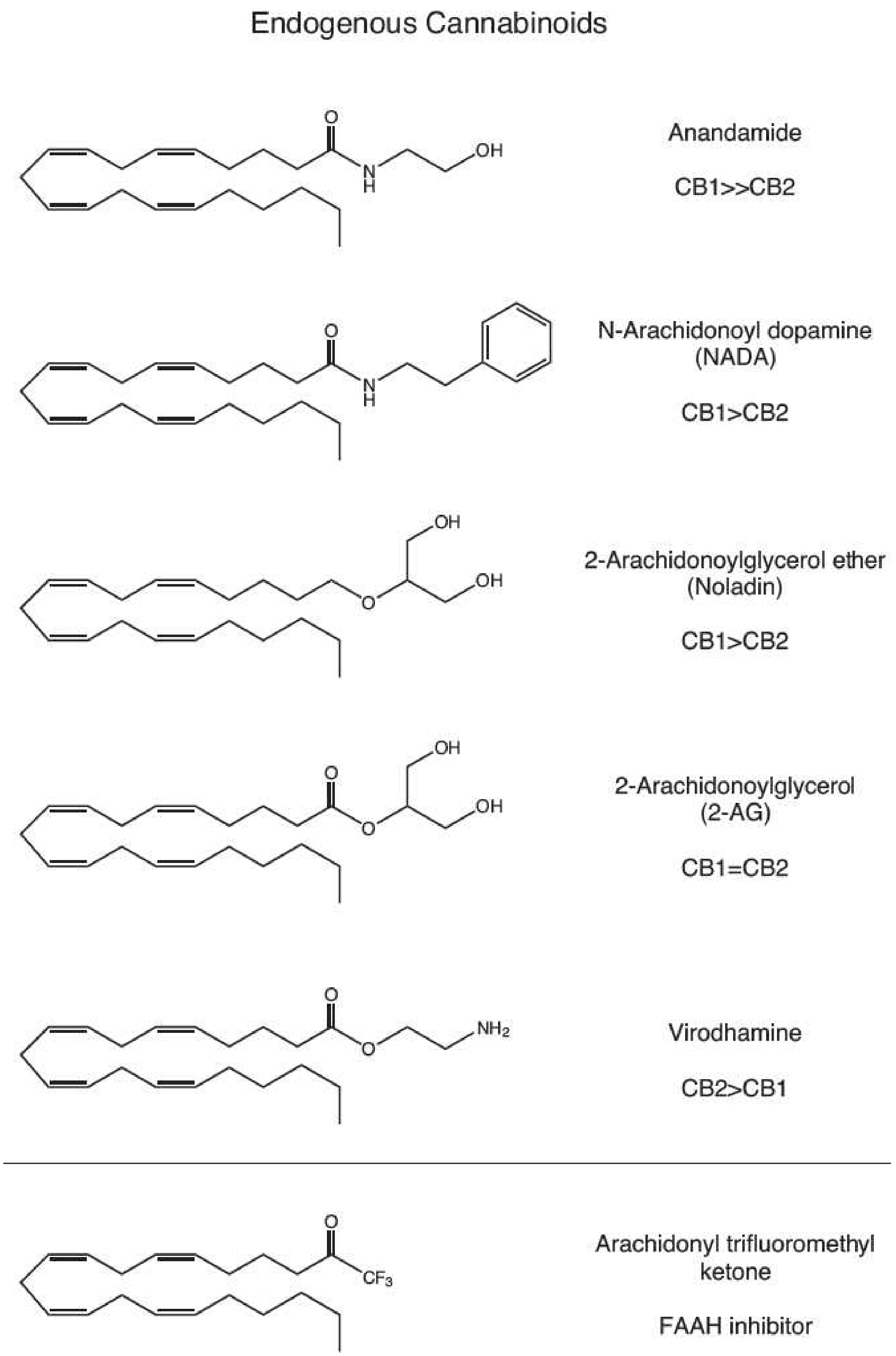
Endogenous Cannabinoids
O
N
H
OH
Anandamide
CB1>>CB2
O
N
H
N-Arachidonoyl dopamine
(NADA)
CB1>CB2
OH
OH
O
2-Arachidonoylglycerol ether
(Noladin)
CB1>CB2
OH
O
OH
O
2-Arachidonoylglycerol
(2-AG)
CB1=CB2
O
O
NH2
Virodhamine
CB2>CB1
O
CF3
Arachidonyl trifluoromethyl
ketone
FAAH inhibitor

FIGURE 1.9–8. Endogenous cannabinoids. At least five endocannabinoids exist in the mammalian brain, each differing in affinity for CB1 and CB2 cannabinoid receptors. All are derived from the essential omega-6 fatty acid,

arachidonic acid, which is also a precursor in the formation of prostaglandins and leukotrienes.

Inactivation of Endocannabinoids

Neurotransmitters are typically inactivated either by reuptake from the neurons that release them or degradation by highly specific enzymes, such as acetylcholine being hydrolyzed by acetylcholinesterase. At least two enzymes exist to target the destruction of endocannabinoids and attenuate their neurotransmission. Fatty acid amide hydrolase (FAAH) converts anandamide to arachidonic acid and ethanolamine (see Fig. 1.9–9). FAAH is found in regions of the brain where CB1 receptors are predominant and localizes to postsynaptic neurons where anandamide is made. Rapid degradation of anandamide in part explains its relatively low potency compared to THC. Confirming the role of FAAH in anandamide inactivation, knockout mice without FAAH exhibit a 15-fold increase of anandamide, but not 2-AG. These mice have greater behavioral responses to exogenous anandamide, owing to its decreased degradation. The endocannabinoid 2-AG is inactivated by FAAH, but also by a monoacylglycerol lipase (MAGL) located in presynaptic neurons.

Pharmacologic inhibitors of FAAH have analgesic effects and reduce anxiety in animal models, but do not have the undesirable effects of THC such as immobility, lowered body temperature or greater appetite. Such a pharmacologic strategy would be analogous to monoamine oxidase inhibitors (MAOI) and catechol o-methyltransferase inhibitors (COMTI). MAOI, used in depression, inhibits breakdown of serotonin and other monoamines, thereby increasing serotonin, while COMTI serve an analogous role in blocking destruction of dopamine and other catecholamines.

Cannabinoid Receptors

Underscoring their importance in neural functions, CB1 receptors are possibly the most abundant G-protein–coupled receptors in the brain. They occur at highest density in the basal ganglia, cerebellum, hippocampus, hypothalamus, anterior cingulate cortex, and cerebral cortex, particularly the frontal cortex. Humans or animals that receive large doses of THC develop catalepsy, a reduction of spontaneous movement and freezing into bizarre and unnatural postures. The action of cannabinoids in the basal ganglia and cerebellum may be associated with these behaviors, which may prove relevant in understanding catatonic symptoms in schizophrenia.

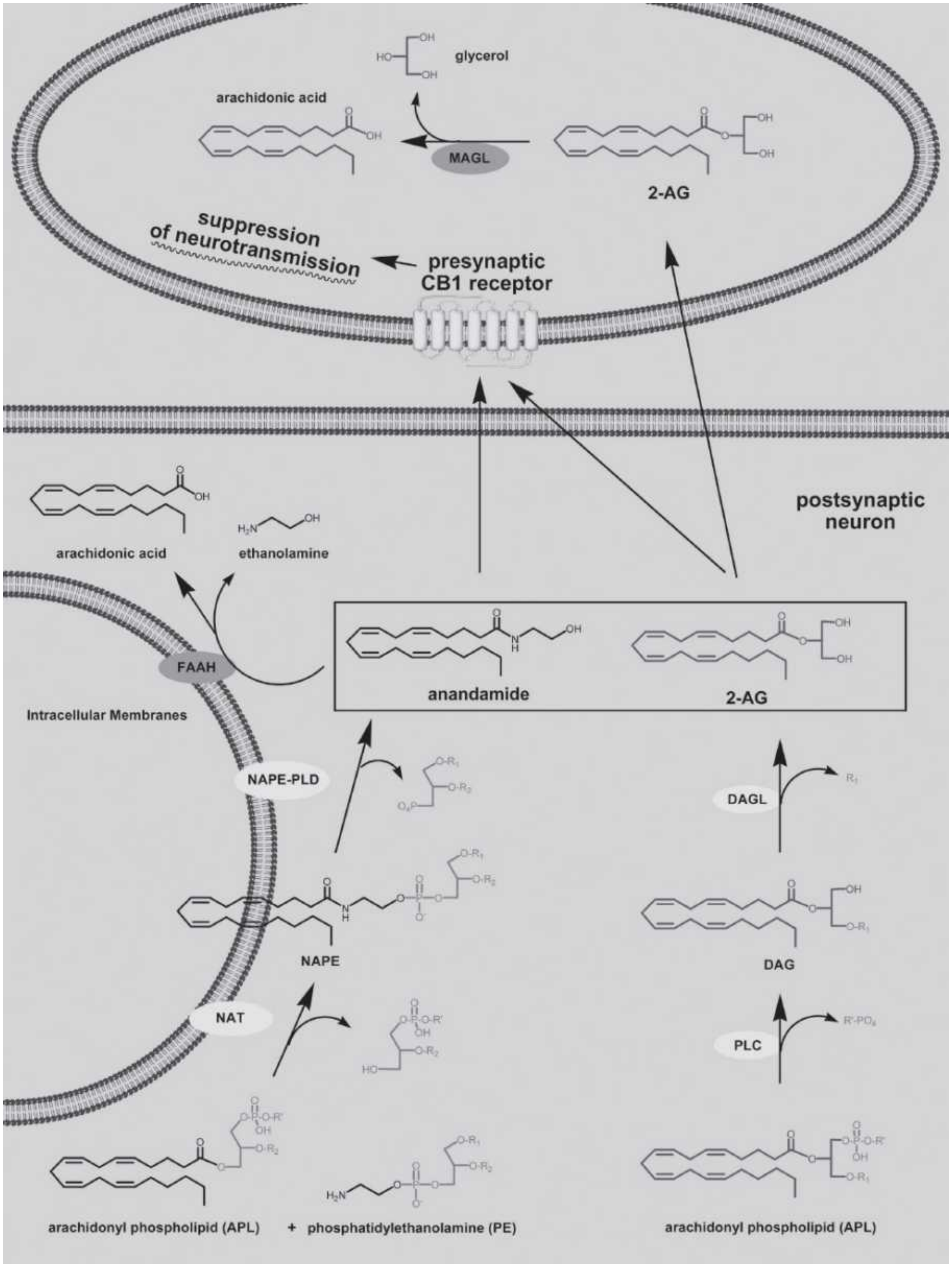
glycerol
arachidonic acid
MAGL
2-AG
suppression of neurotransmission
presynaptic CB1 receptor
postsynaptic neuron
arachidonic acid
ethanolamine
FAAH
Intracellular Membranes
anandamide
2-AG
NAPE-PLD
DAGL
NAPE
DAG
NAT
PLC
arachidonyl phospholipid (APL)
+ phosphatidylethanolamine (PE)
arachidonyl phospholipid (APL)

FIGURE 1.9–9. Retrograde neurotransmission of the endocannabinoids, anandamide, and 2-arachidonoylglycerol (2-AG). Anandamide is synthesized on demand for neurotransmission via a two-step process. The enzyme NAT transfers the arachidonic acid chain from a phospholipid (APL) to phosphatidyl

ethanolamine (PE), producing NAPE. A second enzyme, NAPE-PLD, generates anandamide. 2-AG is similarly synthesized in two steps by the enzymes PLC and DAGL. The endocannabinoids made in a postsynaptic neuron cross the synapse and activate presynaptic CB1 receptors, suppressing neurotransmission of the presynaptic neuron (although activation of the presynaptic neuron occurs in some cases). 2-AG is predominantly inactivated in the presynaptic neuron by MAGL, whereas anandamide is destroyed in the postsynaptic neuron by FAAH. PE, phosphatidylethanolamine; APL, arachidonyl phospholipids; NAT, N-acyltransferase; NAPE, N-arachidonoyl-phosphatidylethanolamine; NAPE-PLD, N-arachidonoyl-phosphatidylethanolamine phospholipase D; FAAH, fatty acid amide hydrolase; MAGL, monoacylglycerol lipase; PLC, phospholipase C; PDE, phosphodiesterase; DAG, diacylglycerol; DAGL, diacylglycerol lipase; R_1-R_3, various acyl or alkyl side chains of phospholipids; R’, side chain of phospholipid head group.

CB1 receptors are predominantly found on axons and nerve termini, with little present on neuronal dendrites and the cell body. CB1 receptors tend to be localized to the presynaptic rather than postsynaptic side of the neuronal cleft, suggesting a role in regulation of neurotransmission. A second cannabinoid receptor, CB2, is predominantly expressed on the surface of white blood cells of the immune system, but small amounts appear to be present in the brain.

Effects on Neurotransmission. The cannabinoid CB1 receptor is associated with G proteins that mediate its intracellular signaling, in part, through inhibition of adenylyl cyclase. This leads to a decrease in levels of the important second messenger, cyclic adenosine monophosphate. Activation of the CB1 receptor also leads to activation of potassium channels and inhibition of N-type calcium channels. As calcium is integral to neurotransmitter release, cannabinoids can block neurotransmission through this mechanism. Cannabinoid receptors also activate mitogen-activated protein kinases.

Via the use of cell culture models and slices of brain, cannabinoids have been shown to block the release of a variety of neurotransmitters, including GABA, norepinephrine, and acetylcholine. Norepinephrine and acetylcholine tend to be excitatory neurotransmitters, and cannabinoid inhibition of their release would be expected to have an overall inhibitory effect. However, GABA is an inhibitory neurotransmitter, and cannabinoid inhibition of it would lead to overall excitatory effects, demonstrating that cannabinoids can have complex effects on neurotransmission depending upon the context. Cannabinoids also appear to increase the release of brain endorphin neurotransmitters and increase dopamine release in the nucleus accumbens, a “reward center” relevant to addiction and learning. The endocannabinoids have been implicated in a variety of forms of synaptic plasticity, including LTP and long-term depression (LTD).

Retrograde Transmission Regulated by Endocannabinoids

It has long been apparent to neuroscientists that a postsynaptic neuron could regulate the activity of a presynaptic neuron, for instance, inhibiting further release of neurotransmitter by the presynaptic neuron. Endocannabinoids may be the best candidate to date as the *retrograde messenger* that diffuses from a postsynaptic neuron to act upon a presynaptic neuron. During development the enzymes responsible for cannabinoid synthesis are expressed in both pre- and postsynaptic neurons. However, in adult brain synthesis of endocannabinoids appears to be predominantly in postsynaptic neurons. This suggested that they might work *backward* and regulate neurotransmission of a presynaptic neuron.

A typical presynaptic neuron containing dopamine or glutamate releases its neurotransmitter, leading to depolarization of the postsynaptic neuron. This second neuron can release endocannabinoid that diffuses across the synaptic cleft and inhibits further neurotransmitter release from the presynaptic neuron. Such a mechanism has been identified in the rodent hippocampus, in which CB1 receptor antagonists block suppression of the presynaptic neuron. Similarly, mice rendered genetically deficient in CB1 receptors lose hippocampal suppression of the presynaptic GABA neurons (a process also known as depolarization-induced suppression of inhibition, or DSI). The reverse has also been demonstrated, as inhibitors of endocannabinoid destruction enhance retrograde neurotransmission in the hippocampus. Endocannabinoids have also been implicated in regulation of other neurotransmission processes, such as LTD and depolarization induced suppression of excitation (DSE). The ability of endocannabinoids to inhibit neurotransmitter release may be an important general mechanism by which neurotransmission is regulated.

Endocannabinoids in Anxiety and Mood

Endocannabinoid neurotransmission may be an important regulator of anxiety, and cannabis users regularly describe the tranquilizing effect of THC. Loss of signaling by the endocannabinoid system appears to promote anxiety-like states in animal studies. CB1 receptor–deficient animals exhibit more pronounced anxiety behavior when exposed to stress or new environs.

Marsicano and colleagues suggested a role for cannabinoid signaling in forgetting painful memories. A mouse model was employed in which a tone was paired with an electric shock. Typically, when exposed to the tone, mice “freeze” in anticipation of the shock. Once the shock is no longer paired with the tone, animals demonstrate “extinction,” that is, they no longer freeze in response to just the tone by itself. Remarkably, animals deficient in CB1 receptors failed to demonstrate this normal extinction. Thus, endocannabinoid neurotransmission may mediate the ability to “forget” the anxiety associated with a painful memory. CB1 antagonist drugs, given just before the tone, revealed a similar effect. The amygdala participates in many anxiety responses, and endocannabinoids may act upon this brain region to attenuate anxiety. In support of this, levels of anandamide and 2-AG were found to increase in the amygdala immediately following exposure of mice to the tone.

FAAH knockout mice lack the enzyme that degrades endocannabinoids, and exhibit both increased anandamide levels and reduced anxiety in behavioral tests. Enhancing levels of endocannabinoids may represent a therapeutic target for anxiety. Whereas an agonist might overactivate cannabinoid receptors where little neurotransmission normally occurs, blocking the breakdown of endocannabinoids would be expected to facilitate activity in areas already utilizing these molecules, and thereby having fewer side effects. Novel FAAH inhibitors reduce the breakdown of anandamide and reduce anxiety-like behaviors in animals. Though the “forced swim test” and “tail suspension test” are far from perfect models of depression in the mouse, FAAH inhibitors improved the ability of the animals to cope with these stresses, a benefit also observed by treatment with antidepressant drugs.

The endocannabinoid pathway may represent an attractive target in understanding post-traumatic stress responses and phobias. Although one cannot yet routinely measure brain endocannabinoid levels in human subjects, this model is supported by clinical trials of the cannabinoid receptor blocker, rimonabant, which may offer promise as a strategy for weight loss (see below). A frequent adverse reaction to the drug was increased anxiety and depression. In a 2007 meta-analysis, Christensen et al. reported that those receiving rimonabant had a 2.5 times greater risk of stopping treatment because of depression and a threefold greater risk of stopping due to anxiety. These psychiatric side effects occurred despite the studies having excluded those who had a history of recurrent depressive or anxiety disorders, underlining an important role of this system in the regulation of anxiety and mood.

Endocannabinoids may play a role in mood disorders, as cannabis use is associated with a tranquilizing effect on mood, while some users experience paradoxical anxiety. A postmortem study of depressed individuals who died by suicide found increased CB1 receptors in the prefrontal cortex, also observed in a study of suicide associated with alcohol dependency. Genetic association studies have sought links between the CB1 receptor and psychiatric illness. While results have been mixed, some have suggested links to depression in Parkinson disease and an association with attention-deficit disorder in alcoholic patients.

FAAH inhibitors have yet to meet benchmarks in human trials. SSR-411298 had inadequate improvement in the treatment of depression, and PF-04457845 failed to improve the pain of knee arthritis. In phase 1 clinical trial of BIA, 10–2474, five of 90 participants experienced hemorrhagic and necrotic brain lesions, and one died, highlighting the difficulty of human pharmacologic intervention of this pathway.

Addiction. The endocannabinoid system may be an attractive target for the understanding of addiction. Mice deficient in CB1 receptors are unsurprisingly resistant to the behavioral effects of cannabinoids. However, they also appear to have reduced addiction to and withdrawal from opiates. Further interaction has also been found between the opioid and cannabinoid systems, as cannabinoids appear to increase the release of dopamine in the nucleus accumbens, a key reward area of the brain implicated in addiction. This dopamine release appears to require mu-opioid receptors, as pharmacologically inhibiting these receptors blocks the ability of cannabinoids to increase dopamine release. Rats with a preference for alcohol have decreased FAAH activity, suggestive of greater cannabinoid signaling. CB1 receptor antagonists dampen their alcohol consumption while inhibiting FAAH increases their alcohol consumption. Furthermore, CB1-deficient animals also appear to have reduced alcohol intake. A single amino acid mutation in human FAAH has been found to be associated with drug abuse, and this abnormal enzyme appears to be less stable than its wild-type counterpart.

Endocannabinoids in Psychosis

Heavy use of cannabis can produce psychotic symptoms in individuals with no prior history of psychiatric disorder, though it is unclear whether this is solely due to the drug or an underlying vulnerability to psychosis. Cannabis use often worsens psychosis in schizophrenia, and heavy use has been associated with developing schizophrenia, although some suggest that this association is an accelerated development of symptoms in those who would eventually manifest schizophrenia. Nonetheless, the endocannabinoid system has implications for the pathophysiology of schizophrenia, as cannabinoid signaling appears to increase the release of dopamine. Medications that act as antagonists of dopamine D2 receptors will likely remain a component of schizophrenia treatment for some time.

Elevated levels of anandamide have been noted in the cerebrospinal fluid of individuals with schizophrenia, including in follow-up studies of medication naïve patients. Elevated anandamide levels have also been noted in the blood of those with schizophrenia, and such elevations normalized with clinical improvement.

Elevated CB1 receptor levels have been reported in postmortem brain samples from those with schizophrenia, particularly the dorsolateral prefrontal cortex and cingulate cortex. While some have identified polymorphisms in the CB1 receptors associated with schizophrenia, others found no association. Furthermore, the implications of such polymorphisms on CB1 function are unclear.

Feeding. Following drug ingestion, THC users develop an increased appetite (“the munchies”), and cannabis has been utilized as an appetite stimulant for centuries. This effect may depend on CB1 receptors present in the hypothalamus. Endocannabinoid levels increase in the hypothalamus and limbic system when animals are deprived of food. Mice genetically deficient in CB1 receptors become resistant to developing obesity after being given a high-fat diet. Similarly, the CB1 receptor antagonist, rimonabant, appears to facilitate weight loss by blocking cannabinoid signaling. In a clinical trial of over 3,000 obese patients, those treated with 20 mg/day of rimonabant lost 6.3 kg at 1 year, compared to 1.6 kg in the placebo group. Nausea was the most common side effect reported. A 2007 meta-analysis of clinical trials reported an overall 4.7-kg weight loss with rimonabant treatment, besting the weight-loss drugs orlistat (2.9 kg) and sibutramine (4.2 kg). Unfortunately the severe side effects of anxiety and depression prevented its clinical use.

Effects on Brain Injury and Pain

In mouse models of traumatic brain injury, 2-AG appears neuroprotective, reducing brain edema, infarct size, and cell death while improving functional outcomes. Anandamide also protected against brain injury in a model of multiple sclerosis, and human patients with the disease have increased production of anandamide. A study of cannabinoid agonist, HU-211, led to more rapid clinical improvement following head trauma. FAAH inhibitors improved motor symptoms in a mouse model of Parkinson disease, likely via cannabinoids increasing dopamine neurotransmission.

Neurotransmission via the endocannabinoid pathway is increasingly appreciated to regulate pain perception. THC and cannabinoid agonists have proven effective in animal models of acute and chronic pain, ranging from burn injury to nerve damage and inflammation. The CB1 receptor plays an important role in these effects as the analgesic effects of cannabinoid drugs are lost when the CB1 antagonist rimonabant is given. Similarly, the analgesic effect of THC is lost in mice genetically deficient in the CB1 receptor.

Stress has long been associated with diminished pain perception, such as in cases of injured military personnel who demonstrate remarkable pain tolerance, a phenomenon known as stress-induced analgesia. The endocannabinoid system may mediate these effects. Animal models reveal anandamide and 2-AG production after stress, and stress-induced analgesia is blocked by the CB1 blocker, rimonabant, in these animals. In a placebo-controlled, randomized study of over 600 individuals with multiple sclerosis, Zajicek and colleagues found that oral THC administration led to improvement in mobility and pain. However, THC offered little benefit in postoperative pain following a hysterectomy.

Endocannabinoid regulation of pain perception appears to be distinct from that of the endogenous opiate system, but the two pathways may share overlapping neural pathways. Evidence for this has been provided using CB1 blocker, rimonabant, and naloxone, which blocks opiate receptors. Rimonabant attenuates analgesia provided by THC and cannabinoids but only partly blocks the response to morphine. However, the opposite is true for opiates: naloxone blocks morphine-induced analgesia but also partially blocks the analgesia of THC and cannabinoid drugs. Combinations of cannabinoid and opiate drugs evince synergistic analgesic effects in animal models.

Although it was initially assumed that cannabinoids exert their analgesic effects via the CNS, in animal models, it has been shown that localized administration of cannabinoids may also be effective, including drugs selective for the CB2 receptor, whose expression is minimal in the CNS.

Endocannabinoids may also influence pain sensitivity by mechanisms that do not involve the CB1 and CB2 receptors. Both anandamide and NADA can also activate a calcium channel known as the vanilloid receptor (also known as TRPV-1) found on sensory nerves. This same receptor is notable for being activated by capsaicin, which causes the hot sensation of chili peppers. Thus, endocannabinoids can exert opposing functions: promoting analgesia through the CB1 and CB2 receptors but potentially increasing pain via TRP channels. While CB2 receptors are largely expressed in the periphery, postmortem analyses reveal upregulation in the brain of those with Alzheimer disease.

The rapid development of novel cannabinoid drugs may allow for the targeting of specific symptoms rather than eliciting all of the typical effects of THC. For instance, ajulemic acid demonstrates analgesic and anti-inflammatory properties but may offer the benefit of limited psychoactive side effects. In a randomized clinical trial of this compound, ajulemic acid proved efficacious in reducing chronic neuropathic pain.

Effects in the Periphery

Cannabinoids lead to the direct relaxation of vascular smooth muscle by local CB1 receptors. This vasodilatation extends to the conjunctiva of the eyes, leading to a “bloodshot” appearance in some cannabis users. Relaxation of ocular arteries by cannabinoids may offer utility as a treatment for glaucoma, a condition of high intraocular pressure, and activation of CB1 receptors in the kidney can improve renal blood flow. A role in generalized blood pressure regulation is unproven, and blood pressure is unaltered in persons treated with rimonabant or animals deficient in CB1 receptors. Cannabinoid signaling may also be relevant to ectopic pregnancy, as CB1-deficient mice retain many embryos in the oviduct.

Synthetic Cannabinoids

Hundreds of chemically distinct synthetic cannabinoids have been generated since the 1980s (see Fig. 1.9–7). Initially generated for research purposes, they have grown in popularity for users to seek euphoric effects. Governments have sometimes been quick to ban such uses, only to have new and different chemicals appear. These and many others are variably mixed together in recreational drugs referred to as “Spice” and “K2” that are typically smoked or ingested by mouth and not detected by standard drug tests for THC. Delta-8-tetrahydrocannabinol (delta-8-THC) is a close structural analog of THC. It is present in small quantities in cannabis plants though it is commonly made by chemists using cannabidiol as a precursor. Its psychoactive effects are less potent than THC, in part relating to its binding to CB1 receptors about sixfold less efficiently than THC. Dronabinol (Marinol) is a synthetic THC with FDA approval to treat chemotherapy-induced nausea and vomiting. Nabilone is a partial agonist of CB1 and CB2, once approved for similar applications but subsequently withdrawn from the market.

Nonpsychoactive Cannabinoids

Although THC is the principal psychoactive component of cannabis, the many nonpsychoactive cannabinoids also have intriguing properties and may regulate neurotransmission. Cannabidiol may offer potential therapeutic effects and appears to stimulate TRP-V1 receptors and influence endocannabinoid degradation. Cannabidiol also demonstrated a protective effect in a mouse model of inflammatory arthritis and has been purported to have anticonvulsant properties. Although results have been mixed, purified cannabidiol may also exert antipsychotic activity, although the net effect of plant cannabis use typically exacerbates schizophrenia symptoms owing to the THC component. Tetrahydrocannabivarin is a plant cannabinoid that antagonizes CB1 receptors. It is a candidate marker to distinguish whether a patient has been using plant-derived cannabis or prescription THC, which contains no tetrahydrocannabivarin.

EICOSANOIDS

Overview

Clinical findings suggest that the dietary supplements Omega-3 fatty acids, eicosapentaenoic acid (EPA), its ester ethyl-eicosapentaenoic (E-EPA), and docosahexaenoic acid (DHA), help relieve symptoms of depression, bipolar illness, schizophrenia, and cognitive impairment. DHA and EPA may help reduce behavioral outbursts and improve attention in children.

Chemistry

Essential fatty acids are a group of polyunsaturated fats that contain a carbon–carbon double bond in the third position from the methyl end group in the fatty acid chain. They are essential because, unlike monosaturated and saturated fatty acids, polyunsaturated fatty acids cannot be synthesized de novo and can only be acquired through diet from natural fats and oils. Linoleic acid (LA) is the parent compound of omega-6 fatty acids, and α-linolenic acid (ALA) is the parent compound of omega-3 fatty acids. Both omega-3 and omega-6 groups use the same enzymes for desaturation and chain elongation. Omega-3 fatty acids are synthesized by algae and plankton. Fish such as herring, salmon, mackerel, and anchovy feed on these aquatic species and become a rich dietary source of omega-3. EPA and DHA are highly unsaturated omega-3 fatty acids that contain 6 and 5 double bonds on their long structural chain, respectively. They are positioned in the cell membrane by phospholipids and play a crucial role in cell membrane signaling.

Effects on Specific Organs and Systems

The strongest scientific evidence for treatment with fatty acid supplements comes from the cardiovascular literature. Several human trials have demonstrated that omega-3 fatty acids lower blood pressure, reduce the rate of recurrent myocardial infarction, and lowers triglyceride levels. In the nervous system, fatty acids are essential components of neurons, immune cells, and glial phospholipid membrane structures. They increase cerebral blood flow, decrease platelet aggregation, and delay the progression of atherosclerosis in the cardiovascular system. Omega-6 fatty acids appear to reduce inflammation and neuronal apoptosis and decrease phosphatidylinositol second messenger activity. Omega-3 fatty acids have been suggested to alter gene expression.

In the CNS, fatty acids are selectively concentrated in neuronal membranes and involved in cell membrane structure. Omega-6 arachidonic acid has been shown to enhance glutamate neurotransmission, stimulate stress hormone secretion and trigger glial cell activation in the setting of oxidative toxicity and neurodegeneration. The omega-3 fatty acids DHA and EPA appear to protect neurons from inflammatory and oxidative toxicities. Increases in serotonin, enhancement of dopamine, and regulation of corticotrophin-releasing factor have been demonstrated in cell culture models.

In rodent models of depression, chronic EPA treatment normalized behavior in an open-field test. Serotonin and norepinephrine were also increased in the limbic regions. Mice fed Omega-3 poor diets had reduced memory, altered learning patterns, and more behavioral problems.

Therapeutic Indications

Clinical research on the use of fish oil for mood disorders began with epidemiology studies where there appears to be a negative correlation between fish consumption and depressive symptoms. Countries with lower per capita fish consumption had up to 60 times increased rates of major depression, bipolar disorder, and postpartum depression. Observational studies concluded that the lower incidence of seasonal affective disorder in Iceland and Japan than latitude predicted is related to the amount of fatty acid these populations consume in their diet. A study in Norway showed that the use of cod-liver oil decreased depressive symptoms. Depression after a myocardial infarction shows higher arachidonic acid to EPA ratio. Postmortem studies in the brains of patients diagnosed with major depressive disorder show reduced DHA in the orbitofrontal cortex.

The first randomized, controlled pilot study of omega-3 fatty acids focused on adjunctive treatment in both bipolar and unipolar patients with depression in addition to their standard lithium or valproic acid treatment. The omega-3 fatty acid group had significant improvement on the Hamilton Depression Rating Scale and a longer period of remission than the placebo group. A subsequent larger study supported the benefit of treatment with E-EPA for bipolar illness. However, a study group of patients with either bipolar disorder or rapid cycling treated with E-EPA showed no significant difference in any outcome measure between the

EPA and placebo groups. Bleeding time was also increased in the treatment group.

In 2002, two landmark double-blind, placebo-controlled studies showed that supplementation with E-EPA or DHA in addition to standard treatment for unipolar depression led to a significant reduction in Hamilton Depression Rating Scales. A later study with a longer duration and a larger number of subjects failed to see an effect of EPA and DHA supplementation on depressed mood or on cognitive function but also found no adverse effects from the supplements. Although meta-analysis shows the antidepressant efficacy of omega-3 fatty acids, larger trials may identify favorable target subjects, optimal dosing, and the composition of omega-3 in treating depression.

Limited consumption of seafood during pregnancy is recommended in American guidelines, but Japanese studies suggest that low DHA but not EPA increases the risk for postpartum depression. A randomized trial of omega-3 fatty acids as monotherapy for major depressive disorder during pregnancy showed a reduction in Hamilton Depression rating scores and compliance with treatment compared to the control group.

The most convincing evidence comes from early brain development and learning studies. Pregnant mothers who consumed foods rich in DHA gave birth to infants that had improved problem-solving skills but not necessarily improved memory. Visual acuity and eye development are also associated with DHA supplementation during pregnancy.

The Oxford-Durham study of dietary supplementation with omega-3 fatty acids in children with developmental coordination disorder suggests a potentially controversial role in learning disabilities, attention-deficit hyperactivity disorder, and autism. The authors saw significant reductions in inattention, hyperactivity, and impulsivity. The crossover placebo group improved after switching to fish oil supplements, and a multivitamin showed no additional benefits for ADHD symptoms. This led to a plan by education officials in the Durham County Council in England to spend £1 million on omega-3 fish oils to help 5,000 children as they approach their school placement exams in order to help to improve their performance.

In behavioral studies, prisoners in England that consumed higher amounts of seafood containing mega-3 fatty acids saw a decrease in assault rates. A Finnish study of violent criminals identified lower levels of omega-3 fatty acids in their system compared to nonviolent offenders.

The negative and psychotic symptoms of schizophrenia may be improved with supplementation with omega-3 fatty acids. Antipsychotic medications like haloperidol appear to have fewer extrapyramidal side effects when combined with antioxidants and omega-3 fatty acids.

EPA and DHA have been associated with decreased dementia incidence. After reviewing the Rotterdam study of a longitudinal cohort of over 5,300 patients, fish consumption appeared to be inversely related to the development of new cases of dementia. Later analysis of the study after 6 years demonstrated that low intake of omega-3 fatty acids was not associated with an increased risk of dementia. In contrast, the Zutphen study, also in the Netherlands, concluded that high fish consumption was inversely related to cognitive decline at both 3 years and after 5 years. Well-designed clinical trials are needed before omega-3 fatty acids can be recommended for the prevention of cognitive impairment.

Precautions and Adverse Reactions

The most adverse complication is an increased risk of bleeding. Dietary sources can contain heavy metals, and there is no standard preparation for capsule formulations. Treatment studies have yielded a variety of different doses.

NEUROSTEROIDS

Background

While steroids are critical for the maintenance of body homeostasis, neurosteroids are synthesized from cholesterol in the brain and independent of peripheral formation in the adrenal gland and gonads. Neurosteroids are produced by a sequence of enzymatic processes governed by P450 and non-P450 enzymes either within or outside the mitochondria of several types of CNS and PNS cells (Fig. 1.9–10).

Neurosteroids can operate through a nongenomic pathway to regulate neuronal excitability via their effects on neurotransmitter-gated ion channels. Receptors are generally located in the nucleus, membrane, or microtubules of the CNS and PNS. Although steroids and neurosteroids may act on the same nuclear receptors, neurosteroids differ from steroids in their topologic distribution and regional synthesis. The most well-known effect of neurosteroids is on the GABA receptor, particularly the GABAA receptor alpha subunit. These agents do not directly activate the receptor but are positive allosteric modulators that increase the ability of GABA to open the GABAA receptor channel. Neurosteroids acting primarily at this site include allopregnanolone (3α5α tetrahydroprogesterone), pregnenolone (PREG), and tetrahydrodeoxycorticosterone (THDOC). DHEA-S (dehydroepiandrosterone sulfate), the most prevalent neurosteroid, acts as a noncompetitive modulator of GABA, and its precursor DHEA (dehydroepiandrosterone) has also been shown to exert inhibitory effects at the GABA receptor. Some neurosteroids may also act at the N-methyl-D-aspartate receptor (NMDA-R), alpha-amino-3-hydroxy-5-methyl-4-isoxazole-propanoic acid receptor (AMPA-R), kainate, glycine, serotonin, sigma type-1, and nicotinic acetylcholine receptors. Progesterone is also considered a neurosteroid and can regulate gene expression at progesterone receptors. Questions remain as to how the neurosteroids differ in spectrum from benzodiazepines, which are also positive allosteric modulators of the GABAA receptor alpha subunits. One clue may be that of the six different types of alpha subunits of GABAA, benzodiazepines do not bind to alpha-4 or alpha-6, whereas the neurosteroids do, among other subunits.

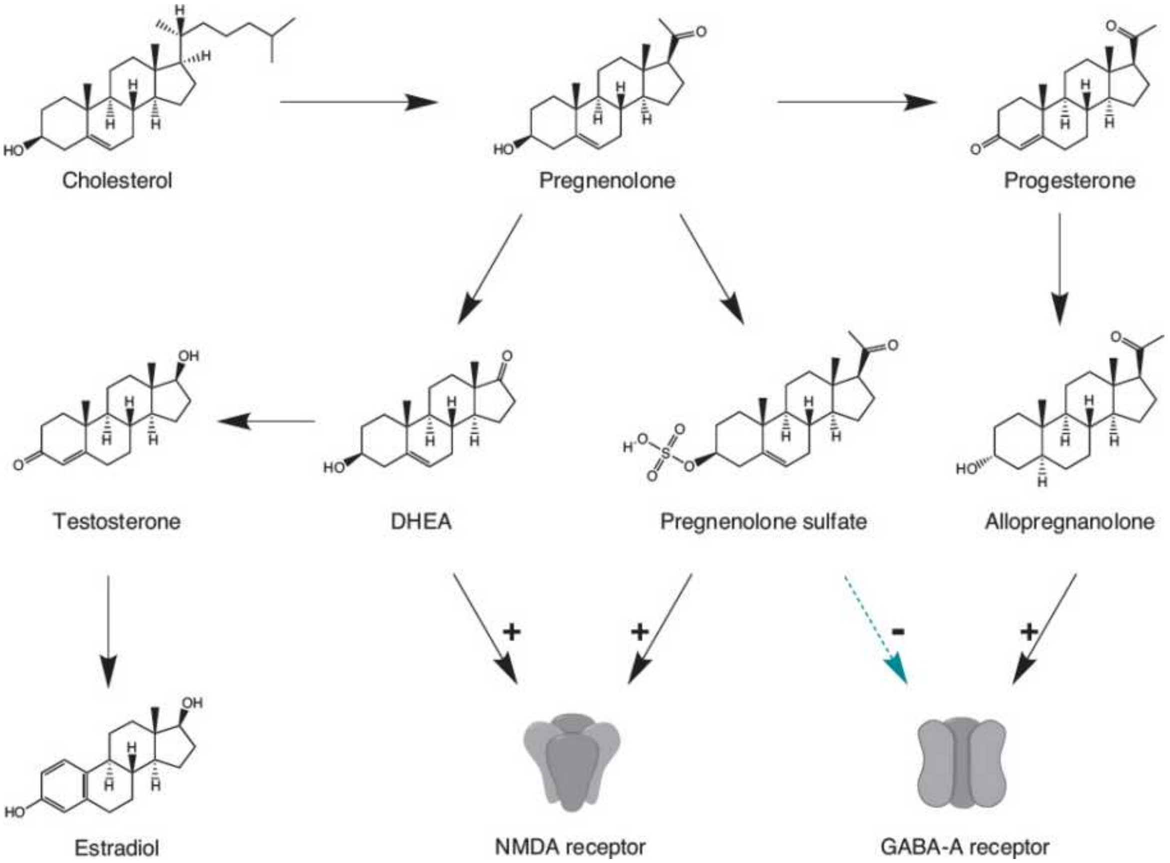
Cholesterol
Pregnenolone
Progesterone
Testosterone
DHEA
Pregnenolone sulfate
Allopregnanolone
Estradiol
NMDA receptor
GABA-A receptor
+
+
-
+

FIGURE 1.9–10. Biosynthetic pathway of select neurosteroids. + indicates positive allosteric effect upon GABA-A and NMDA receptors. (-) negative allosteric effect.

Neurosteroids in Neurodevelopment and Neuroprotection

Neurosteroids have been shown to stimulate axonal growth and promote synaptic transmission. Specific neuroprotective effects are unique to each neurosteroid. DHEA acts to regulate brain serotonin and dopamine levels, suppress cortisol, increase hippocampal primed burst potentiation and cholinergic function, decrease amyloid β-protein, inhibit the production of proinflammatory cytokines, and prevent free radical scavenging. DHEA and DHEA-S have both been shown to have a role in glial development and neuronal growth and to promote their survival in animals; their injection into the brains of mice promoted long-term memory while reversing amnesia. Progesterone is linked to myelinating processes like aiding in the repair of damaged neural myelination. Allopregnanolone contributes to the reduction of contacts during axonal regression.

Role of Neurosteroids in Mental Illness

Neurosteroids have distinct implications for the maintenance of normal neurologic function and contribute to neuropsychiatric illness. Neurosteroids are differentially regulated in males and females and may impact the manifestation of psychological disorders in these two populations. Specifically, they play a distinct role in mood and anxiety disorders. Allopregnanolone (brexanolone, Zulresso) has a salutary benefit for postpartum depression and is an FDA-approved intravenous treatment. Zuranolone (SAGE-217) is an orally available GABAA-positive modulator that has shown benefit in a phase 2 study of major depressive disorder.

Depression. When compared with nondepressed controls, studies show that depressed patients have lower plasma and CSF concentrations of allopregnanolone; additionally, this research has elucidated an inverse relationship between allopregnanolone concentrations and the severity of depressive illness. Antidepressant drugs, specifically fluoxetine, have been shown in multiple studies to increase the levels of certain neurosteroids. Nonetheless, there is debate over the therapeutic properties of neurosteroids, prompting the investigation of neurosteroid concentrations in patients undergoing nonpharmacologic therapies. Preliminary results indicate that the lack of modifications in neurosteroid levels during nonpharmacologic treatments supports the validity of the pharmacologic properties of antidepressants, not their therapeutic action, in the elevation of neurosteroid levels in medicated populations. In a 2006 clinical study with mirtazapine, allopregnanolone concentrations increased in patients with major depressive disorder regardless of the therapeutic benefit.

Anxiety Disorders. In patients with anxiety disorders, the major mechanism of action is on the GABA receptor. Homeostasis, characterized by normal GABAergic activity, is restored after panic attacks as neurosteroids are released in response to stress. Allopregnanolone stimulates GABAergic activity with 20 times the strength of benzodiazepines and 200 times the potency of barbiturates. Both positive and negative regulation of the GABAA receptor is correlated with anxiolytic and anxiogenic action, respectively.

Psychotic Disorders. In addition to their primary relevance to the pharmacologic treatment of mood and anxiety disorders, neurosteroids contribute to psychotic, childhood, substance abuse, eating, and postpartum disorders. The effect of neurosteroids on psychotic disorders like schizophrenia is mediated by DHEA and DHEA-S. DHEA has been dispensed to decrease anxiety in schizophrenics, as DHEA and DHEA-S suppress GABA inhibition and heighten the neuronal response at the NMDA-R and sigma receptors. DHEA and DHEA-S levels are typically elevated first episode psychosis.

Attention-Deficit Disorder. In children, the clinical symptoms of ADHD are inversely correlated with DHEA and pregnenolone levels.

Substance Abuse. Alcohol is theorized to regulate the GABA receptor and induce de novo steroid synthesis in the brain; specifically, pregnenolone, allopregnanolone, and allotetrahydrodeoxycorticosterone levels are increased in the brain and periphery in response to increases in peripheral alcohol levels. It is hypothesized that sharp increases in ethanol concentration may mimic the acute stress response and elevate neurosteroid concentrations by the hypothalamic–pituitary–adrenal axis. To prevent ethanol dependence, researchers are investigating fluctuations in neurosteroid levels and in vivo neurosteroid responsiveness. Neurosteroids (increased allopregnanolone levels in particular) are associated with drug abuse. However, DHEA-S may check the acquisition of morphine tolerance. Past research has shown DHEA-S levels were also increased in patients who abstained from cocaine use as per a treatment program, and as patients relapsed, DHEA-S concentrations decreased accordingly.

Eating Disorders. DHEA has been shown to diminish food intake, temper obesity, moderate insulin resistance, and lower lipids in rodent models of youth-onset, hyperphagic, genetic obesity. By regulating the serotonergic system, DHEA is hypothesized to promote a reduced caloric load. Although hypothetical, low levels of DHEA and DHEA-S are recorded in young women with anorexia nervosa, and 3 months of oral DHEA supplementation increased bone density and tempered the emotional problems associated with the disorder.

Postpartum and Gynecologic Disorders. As estrogen and progesterone levels fluctuate during pregnancy and drop markedly after delivery, neurosteroids are thought to contribute to postpartum disorders.

Low postpartum DHEA concentrations have been linked to mood instability. In addition, allopregnanolone levels correlated with mood disorders during pregnancy and premenstrual syndrome (PMS). It has been noted that women with premenstrual dysphoric disorder have higher allopregnanolone/progesterone ratios than normal controls; women treated for this disorder reported improvement as allopregnanolone levels decreased.

Neurosteroids, Memory Disorders, and Aging. Neurosteroid levels may be irregular in neurodegenerative disorders and aging conditions such as Alzheimer's and Parkinson's. DHEA levels at age 70 are only about 20% of their maximum value recorded in the late 20s, and some researchers believe DHEA supplementation can prevent or slow the cognitive declines associated with the aging process. However, conflicting studies have indicated that DHEA administration does not improve cognitive measures in patients. Additionally, in those patients with Alzheimer disease, the DHEA concentrations have been found to be markedly decreased.

Conclusion

Neurosteroids may play a role in a variety of psychiatric pathologies. Breakthroughs with animal models do not always correlate with advances in understanding the role of neurosteroids in humans, complicating research. Moreover, further research is exploring neurosteroid levels over the treatment course of a diverse array of psychiatric illness to get a more complete picture of disease management. New medications approved for human use are expanding the repertoire of treatments.

NOVEL NEUROTRANSMITTERS: BEYOND THE CLASSICAL DEFINITION OF NEUROTRANSMITTER

The classical criteria for a chemical to be considered a neurotransmitter were (a) synthesis in a presynaptic neuron, (b) storage and release from a presynaptic neuron, (c) binding to a receptor on a postsynaptic membrane, and (d) removal from the synaptic cleft by reuptake or degradation. Within the past few decades, the discovery of novel neurotransmitters has led to a reformulation of these strict criteria. Messengers such as the gases, cannabinoids, and eicosanoids are not stored in vesicles in presynaptic neurons but appear to be generated and released "on demand." The endocannabinoids appear to have an important role as *retrograde messengers*, being released from a conventional postsynaptic neuron and acting upon a presynaptic neuron (see Fig. 1.9–9). Finally, the gases do not act upon a receptor on the extracellular membrane of a postsynaptic neuron but diffuse into the cell and act directly upon multiple cellular proteins, bypassing membrane receptors entirely. Some scientists feel the classical definition of a neurotransmitter should not be altered, and the aforementioned might generically be termed atypical neurotransmitters or novel neural messengers. Additional yet undiscovered chemical messengers may further refine the concept of what a neurotransmitter is.

FURTHER READINGS

Appleton KM, Voyias PD, Sallis HM, et al. Omega-3 fatty acids for depression in adults. *Cochrane Database Syst Rev*. 2015;11(11):CD004692.

Cutajar MC, Edwards TM. Evidence for the role of endogenous carbon monoxide in memory processing. *J Cogn Neurosci*. 2007;19:557–562.

Deng J, Lei C, Chen Y, et al. Neuroprotective gases–fantasy or reality for clinical use? *Prog Neurobiol*. 2014;115:210–245.

Di Marzo V, Stella N, Zimmer A. Endocannabinoid signalling and the deteriorating brain. *Nat Rev Neurosci*. 2015;16:30–42.

Grosso G, Galvano F, Marventano S, et al. Omega-3 fatty acids and depression: Scientific evidence and biological mechanisms. *Oxid Med Cell Longev*. 2014;2014:313570.

Gunduz-Bruce H, Silber C, Kaul I, et al. Trial of SAGE-217 in patients with major depressive disorder. *N Engl J Med*. 2019;381:903–911.

*Iversen LL. *The Science of Marijuana*. Oxford University Press; 2008.

Joy CB, Mumby-Croft R, Joy LA. Polyunsaturated fatty acid supplementation for schizophrenia. *Cochrane Database Syst Rev*. 2006;2006(3):CD001257.

Keck PE Jr, Mintz J, McElroy S L, et al. Double-blind, randomized, placebo-controlled trials of ethyl-eicosapentanoate in the treatment of bipolar depression and rapid cycling bipolar disorder. *Biol Psychiatry*. 2006;60:1020–1022.

Kidd PM. Omega-3 DHA and EPA for cognition, behavior, and mood: clinical findings and structural-functional synergies with cell membrane phospholipids. *Altern Med Rev*. 2007;12:207–227.

Kreitzer AC, Malenka RC. Endocannabinoid-mediated rescue of striatal LTD and motor deficits in Parkinson's disease models. *Nature*. 2007;445:643–647.

*Legare CA, Raup-Konsavage WM, Vrana KE. Therapeutic potential of cannabis, cannabidiol, and cannabinoid-based pharmaceuticals. *Pharmacology*. 2022;107:131–149.

Lim E, Mbowe O, Lee AS, Davis J. Effect of environmental exposure to hydrogen sulfide on central nervous system and respiratory function: A systematic review of human studies. *Int J Occup Environ Health*. 2016;22:80–90.

Lynch AM, Loane DJ, Minogue AM, et al. Eicosapentaenoic acid confers neuroprotection in the amyloid-beta challenged aged hippocampus. *Neurobiol Aging*. 2007;28:845–855.

*Maccarrone M, Guzman M, Mackie K, Doherty P, Harkany T. Programming of neural cells by (endo)cannabinoids: From physiological rules to emerging therapies. *Nat Rev Neurosci*. 2014;15:786–801.

Mazier W, Saucisse N, Gatta-Cherifi B, Cota D. The endocannabinoid system: Pivotal orchestrator of obesity and metabolic disease. *Trends Endocrinol Metab*. 2015;26:524–537.

*McEwen BS, Bowles NP, Gray JD, et al. Mechanisms of stress in the brain. *Nat Neurosci*. 2015;18:1353–1363.

Mifflin K, Benson C, Kerr B, et al. Involvement of neuroactive steroids in pain. *Mod Trends Pharmacopsychiatry*. 2015;30:94–102.

*Mischoulon D, Freeman MP. Omega-3 fatty acids in psychiatry. *Psychiatr Clin North Am*. 2013;36:15–23.

Moncada S, Bolanos JP. Nitric oxide, cell bioenergetics and neurodegeneration. *J Neurochem*. 2006;97:1676–1689.

Nemets H, Nemets B, Apter A, Bracha Z, Belmaker RH. Omega-3 treatment of childhood depression: A controlled, double-blind pilot study. *Am J Psychiatry*. 2006;163:1098–1100.

*Paul BD, Snyder SH. Gasotransmitter hydrogen sulfide signaling in neuronal health and disease. *Biochem Pharmacol*. 2018;149:101–109.

Pi-Sunyer FX, Aronne LJ, Heshmati HM, Devin J, Rosenstock J; RIO-North America Study Group. Effect of rimonabant, a cannabinoid-1 receptor blocker, on weight and cardiometabolic risk factors in overweight or obese patients: RIO-North America: A randomized controlled trial. *JAMA*. 2006;295:761–775.

Porter J, Van Vrancken M, Corll C, Thompson H, Svec F. The influence of dehydroepiandrosterone and 8-OH-DPAT on the caloric intake and hypothalamic neurotransmitters of lean and obese Zucker rats. *Am J Physiol Regul Integr Comp Physiol*. 2005;288:R928–R935.

Ratner MH, Kumaresan V, Farb DH. Neurosteroid actions in memory and neurologic/neuropsychiatric disorders. *Front Endocrinol*. 2019;10:169.

Renard J, Krebs M-O, Le-Pen G, Jay TM. Long-term consequences of adolescent cannabinoid exposure in adult psychopathology. *Front Neurosci*. 2014;8:361.

Richardson AJ, Montgomery P. The Oxford-Durham study: A randomized, controlled trial of dietary supplementation with fatty acids in children with developmental coordination disorder. *Pediatrics*. 2005;115:1360–1366.

Rogers PJ, Appleton KM, Kessler D, et al. No effect of n-3 long-chain polyunsaturated fatty acid (EPA and DHA) supplementation on depressed mood and cognitive function: A randomised controlled trial. *Br J Nutr*. 2008;99:421–431.

Rubino T, Parolaro D. The impact of exposure to cannabinoids in adolescence: Insights from animal models. *Biol Psychiatry*. 2016;79:578–585.

Rudolph U, Crestani F, Benke D, et al. Benzodiazepine actions mediated by specific γ-aminobutyric acid A receptor subtypes. *Nature*. 1999;401:796–800.

Schule C, Romeo E, Uzunov DP, et al. Influence of mirtazapine on plasma concentrations of neuroactive steroids in major depression and on 3alpha-hydroxysteroid dehydrogenase activity. *Mol Psychiatry*. 2006;11:261–272.

Sedlak TW, Snyder SH. Messenger molecules and cell death: Therapeutic implications. *JAMA*. 2006;295:81–89.

Song C, Shieh CH, Wu YS, Kalueff A, Gaikwad S, Su KP. The role of omega-3 polyunsaturated fatty acids eicosapentaenoic and docosahexaenoic acids in the treatment of major depression and Alzheimer's disease: Acting separately or synergistically? *Prog Lipid Res*. 2016;62:41–54.

Su KP, Huang SY, Chiu TH, et al. Omega-3 fatty acids for major depressive disorder during pregnancy: Results from a randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry*. 2008;69:644–651.

Sulaieva O, Wallace JL. Gaseous mediator-based anti-inflammatory drugs. *Curr Opin Pharmacol*. 2015;25:1–6.

Tagen M, Klumpers LE. Review of delta-8-tetrahydrocannabinol (Δ8-THC): comparative pharmacology with Δ9-THC. *Br J Pharmacol*. 2022;179:3915–3933.

Vincent SR. Nitric oxide neurons and neurotransmission. *Prog Neurobiol*. 2010;90:246–255.

Volk DW, Lewis DA. Insights into the pathophysiology of endocannabinoid signaling in schizophrenia. *JAMA Psychiatry*. 2019;76:887–888.

Zorumski CF, Paul SM, Covey DF, Mennerick S. Neurosteroids as novel antidepressants and anxiolytics: GABA-A receptors and beyond. *Neurobiol Stress*. 2019;11:100196.

1.10 Neurotrophic Factors

JIHYE KIM, M.D., FRANCIS S. LEE, M.D., PH.D., AND MOSES V. CHAO, PH.D.

Neurotrophins are a unique family of polypeptide growth factors that influence the proliferation, differentiation, survival, and death of neuronal and nonneuronal cells. The effects of neurotrophins depend on their levels of availability, their affinity of binding to transmembrane receptors, and the downstream-signaling cascades that are stimulated after receptor activation. While the biologic roles for neurotrophins were initially characterized during development of the nervous system, it is now clear that they have multiple roles in the adult nervous system, such as regulating synaptic connections, synapse structure, neurotransmitter release, long-term potentiation (LTP), mechanosensation, and pain.

Alterations in neurotrophin levels have been implicated in neurodegenerative disorders, such as Alzheimer disease (AD), as well as in psychiatric disorders, such as anxiety disorders, depression, and substance abuse. A significant shift has occurred in conceptualizing the role of these growth factors for psychiatric disorders toward their key roles in mediating antidepressant and antianxiety treatment response. These recent insights have informed the role they play in the etiology and treatment of psychiatric disorders.

THE NEUROTROPHIN FAMILY

Many polypeptide factors affect the survival, growth, and differentiation of the nervous system. The neurotrophins, comprising of nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), and neurotrophin-4 (NT-4), are the best understood and the most widely expressed in the nervous system. The neurotrophins are initially synthesized as precursors or proneurotrophins that are cleaved to release the mature, active proteins. The mature proteins, approximately 12 to 14 kDa in size, form stable, noncovalent dimers and are normally expressed at very low levels during development. Proneurotrophin cleavage and processing can occur at multiple sites, including the Golgi apparatus by furin, the dense core vesicle by proconvertases, and extracellularly by plasmin or selective matrix metalloproteinases.

NGF was the first identified neurotrophic factor, and it has a restricted distribution. In the peripheral nervous system (PNS), it acts on sympathetic neurons and sensory neurons involved in nociception and temperature sensation. In the central nervous system (CNS), NGF promotes the survival and functioning of cholinergic neurons in the basal forebrain. These neurons project to the hippocampus and are believed to be important for memory processes, which are specifically affected in AD.

The other neurotrophins are more widely expressed in the CNS. BDNF and NT-3 are highly expressed in cortical and hippocampal structures, and they have been linked to the survival and functioning of multiple neuronal populations. BDNF is present in much higher concentrations after stress, injury, neural stimulation, and exercise. In addition to neuronal survival and differentiation, BDNF is highly responsive to activity-dependent plasticity (see subsection “Neurotrophins and Synaptic Plasticity”). As is the case for other neurotrophins, any pain, injury, or inflammation can result in significant increases in transcription of NGF, BDNF, and other proinflammatory polypeptides.

NEUROTROPHIN RECEPTORS

Neurotrophins are unique in exerting their cellular effects through the actions of two different receptors: the tropomyosin-related kinase (Trk), a receptor tyrosine kinase receptor, and p75 neurotrophin receptor ($p75^{NTR}$), a tumor necrosis factor (TNF) receptor superfamily member. Trk receptors consist of an extracellular ligand-binding region, a single transmembrane domain, and a highly conserved intracellular tyrosine kinase domain. $p75^{NTR}$ consists of an extracellular ligand-binding region, a single transmembrane domain, and an intracellular portion containing a protein association region termed the death domain.

Neurotrophins bind as dimers to Trk family members, leading to receptor dimerization and activation of the catalytic tyrosine protein kinase domains. The dimerized Trk receptors autophosphorylate several key intracellular tyrosine residues, which rapidly initiate intracellular-signaling cascades. This is accomplished by the phosphorylated tyrosines on the receptor acting as recognition sites for the binding of specific adaptor proteins that contain phosphotyrosine-binding motifs, such as Src homology domain 2 (SH2). In particular, the Shc adaptor protein links the activated Trk receptor to two separate intracellular-signaling pathways that mediate most of the biologic effects of neurotrophins.

The primary survival pathway involves Shc, linking the Trk receptor activation to increases in phosphotidylinositol-3-kinase activity. This in turn activates another protein kinase, Akt (protein kinase B), which has multiple effects on the cell's apoptotic pathways. Also, Shc phosphorylation by Trk receptor activation leads to increases in Ras and mitogen-activated protein kinase (MAPK) activities. These events influence transcriptional and translational events, such as the induction of the cyclic adenosine monophosphate (cAMP) response element–binding protein (CREB) transcription factor and mammalian target of rapamycin complex (mTORC). CREB produces a multitude of effects on the cell cycle, neurite outgrowth, and synaptic plasticity. mTORC has been shown to regulate learning and memory by promoting localized activity-dependent translation, which leads to remodeling of spines that accompanies LTP. These mechanisms have been implicated in the antidepressant actions of various drugs, including ketamine (see related subsection "Neurotrophic Hypothesis for Mood Disorders: Therapeutic Potential of Neurotrophins"). In addition, phospholipase-C-γ (PLC-γ) binds to activated Trk receptors and initiates an intracellular-signaling cascade release of inositol phosphates and activation of protein kinase C. Trk receptor activation leads to a multitude of downstream signaling events, leading to changes in transcriptional and translational programs.

NGF binds most specifically to TrkA, BDNF and NT-4 to TrkB, and NT-3 to TrkC receptors (Fig. 1.10–1). $p75^{NTR}$ can bind to each neurotrophin but has the additional capability of regulating a Trk's affinity for its cognate ligand. Trk and $p75^{NTR}$ have been referred to as high- and low-affinity receptors, respectively. However, this is not correct because TrkA and TrkB bind mature neurotrophins with an affinity of 10^{-9} to 10^{-10} M, which is lower than the high-affinity site ($K_d = 10^{-11}$ M). Also, the precursor form of NGF displays high-affinity binding to $p75^{NTR}$. Trk-mediated responsiveness to low concentrations of NGF is dependent on the relative levels of $p75^{NTR}$ and TrkA receptors and their combined ability to form high-affinity sites. This is important since the ratio of receptors can determine responsiveness and ultimately neuronal cell numbers.

Although $p75^{NTR}$ and Trk receptors do not bind to each other directly, there is evidence that complexes form between the two receptors. As a result of these interactions, increased ligand selectivity can be conferred onto Trk receptors by $p75^{NTR}$. One way of generating specificity is by imparting greater discrimination of ligands for the Trk receptors. For example, BDNF, NT-3, and NT-4/5 can each bind to the TrkB receptor; however, in the presence of $p75^{NTR}$, only BDNF provides a functional response. Likewise, NGF and NT-3 both can bind to TrkA, but $p75^{NTR}$ restricts the signaling of TrkA to NGF and not to NT-3. Hence, $p75^{NTR}$ and Trk receptors interact to provide greater discrimination among different neurotrophins.

Neurotrophins acting via Trk receptors promote cell survival and differentiation. Paradoxically, they can also induce cell death. $p75^{NTR}$ serves as a proapoptotic receptor during developmental cell death and after injury to the nervous system. Increase in $p75^{NTR}$ expression is responsible for apoptosis in embryonic retinas and sympathetic neurons during the period of naturally occurring neuronal death. Whereas BDNF binding to $p75^{NTR}$ in sympathetic neurons causes rapid cell death, NGF binding to the TrkA receptor on the same neurons provides a survival signal. In the context of neurotrophin processing, proneurotrophins are more effective than mature NGF in inducing p75-dependent apoptosis. These results suggest that the biologic action of the neurotrophins can be regulated by proteolytic cleavage, with proforms preferentially activating $p75^{NTR}$ to mediate apoptosis and mature forms selectively activating Trk receptors to promote survival.

NEUROTROPHIC FACTORS AND DEVELOPMENT

With their initial discovery, neurotrophins were observed to be highly expressed in the PNS during early development. They are essential for survival of selective populations of neurons during different developmental periods. These findings led to the conceptualization of the neurotrophic hypothesis for development, which provides a functional explanation for the role of neurotrophic factors in the development of the PNS. During development, neurons approaching the same final target, such as in the skin or muscle, compete for limited amounts of target-derived neurotrophic factors. In this way, the nervous system molds itself to maintain only the most competitive and appropriate connections. Competition among neurons for limiting amounts of neurotrophin molecules produced by target cells accounts for selective cell survival. Two predictions emanate from this hypothesis. First, the efficacy of neuronal survival will depend on the amounts of trophic factors produced during development. Second, specific receptor expression in responsive cell populations will dictate neuronal responsiveness.

On one level, neurotrophins fit well with the neurotrophic hypothesis, as many peripheral neuronal subpopulations depend on a specific neurotrophin during the period of naturally occurring cell death. In addition, specificity of the biologic effects of neurotrophins in the PNS during development can be modulated by the intracellular location of the neurotrophin ligand receptor complex. During development, neurotrophins are produced and released from the target tissues. They are internalized into vesicles and transported to the cell body. Interestingly, the biologic effects of neurotrophins require that signals are conveyed over long distances from the nerve terminal to the cell body. Therefore, a central theme of the neurotrophic hypothesis is that neuronal survival and differentiation depend on the retrograde signaling of neurotrophic factors produced at the target tissue.

In contrast to the PNS, the role of neurotrophins in the CNS during development is significantly different, possibly due to predominant expression of the Trk receptors and limited expression of p75^{NTR}. In the CNS, the overlapping expression of multiple Trk neurotrophin receptors and their cognate ligands allows for more diverse connectivity, which extends well into adulthood. Of note, unlike in the PNS, the primary neurotrophins expressed in the CNS are BDNF and NT-3. From genetic knockout studies, it has become clear that their primary functions are not with regard to neuronal survival but rather neuronal differentiation and synaptic plasticity. To date, much of the evidence base for neurotrophin function in the CNS has been focused on BDNF.

One of the first examples to illustrate the roles of neurotrophins in plasticity was identified with regard to regulating plasticity related to critical periods in the visual system. Developmental regulation of synaptic plasticity in the visual system is illustrated by the formation of ocular dominance columns in layer 4 of the cortex. This process is strongly influenced by BDNF, regulating the maturation of inhibitory gamma-aminobutyric acid-producing (GABA) interneurons and thereby, regulating the critical period for visual plasticity. Mice overexpressing BDNF have been shown to have accelerated development of visual acuity and early termination of the critical period. Also, these effects on the visual system plasticity can be blocked by antibodies for the neurotrophins as well as neurotrophin antagonists (e.g., TrkB–IgG fusion proteins that bind neurotrophins), indicating that an alteration in the levels of endogenous neurotrophins has dramatic consequences on the formation of critical periods.

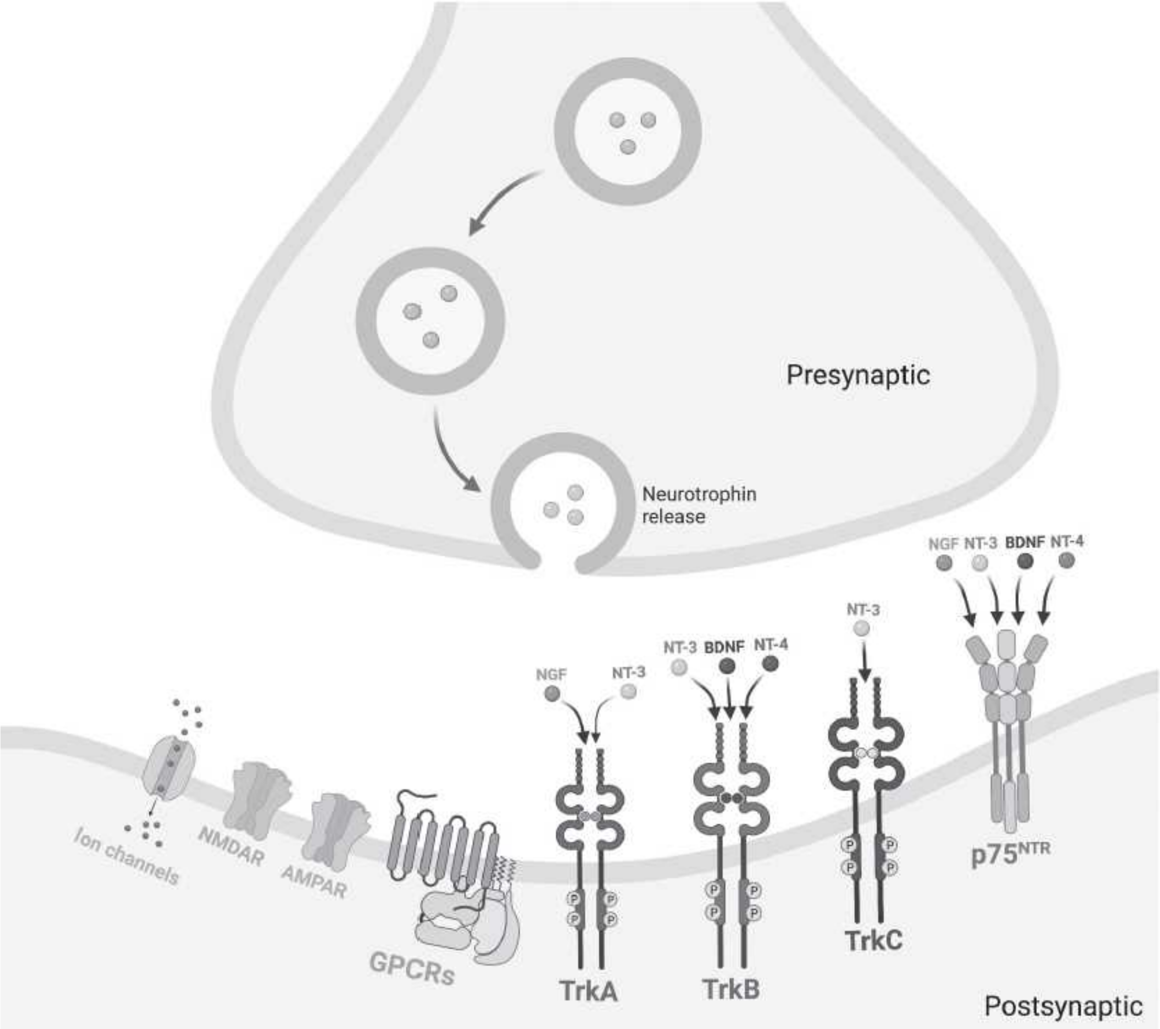
Presynaptic
Neurotrophin release
NGF
NT-3
BDNF
NT-4
Ion channels
NMDAR
AMPAR
GPCRs
TrkA
TrkB
TrkC
p75NTR
Postsynaptic

FIGURE 1.10–1. Neurotrophins at the synapse. Neurotrophins are released at the presynaptic side and bind selectively to specific tropomyosin-related kinase (**Trk**) receptors. All neurotrophins bind to p75 neurotrophin receptor (**p75^{NTR}**). Neurotrophin-3 (**NT-3**) and neurotrophin-4 (**NT-4**) bind to multiple Trk receptors. Nerve growth factor (**NGF**) binds to TrkA and brain-derived neurotrophic factor (**BDNF**) binds specifically to TrkB. The boldness of *arrow* correlates with strength of binding affinity of a specific neurotrophin to a specific receptor. (Created with BioRender.com.)

BDNF has also been shown to be involved in regulating the critical period for fear regulation. A single nucleotide polymorphism in the human BDNF gene (rs62565), resulting in a valine (Val) to methionine (Met) substitution at position 66 in the BDNF prodomain region has been shown to lead to decreased BDNF secretion and impaired periadolescent development of the prefrontal cortex–hippocampus neural circuit in mice. The lack of proper development during this critical window of periadolescence leads to a decreased ability to extinguish a cued fear response in adulthood. Both human BDNF Met allele carriers and mice with the human BDNF Met allele have shown prolonged responses to conditioned aversive cues during the acute extinction phase, suggesting less efficient initial cued fear extinction learning process. Consequently, BDNF Met carriers would likely benefit from more intensively tailored psychiatric treatment, such as a longer or more intensive course of exposure therapy to extinguish the conditioned aversive cue. Relatedly, both Met allele–carrying humans and BDNF Met mice exhibit increased anxiety phenotypes when placed under various environmental stressors. In this context, a recent meta-analysis of all the available human BDNF Met allele studies examining gene–environment interactions between the BDNF Met allele and life stressors showed that the Met allele significantly moderated the relationship between stressful life events and childhood adversity and depression. In sum, the BDNF Met allele likely confers a genetic vulnerability for depressive and anxiety-based symptoms that may require personalized psychiatric treatments for full responses.

NEUROTROPHINS AND SYNAPTIC PLASTICITY

BDNF's role in modulating synaptic plasticity in the adult brain has also been established in the various brain regions such as the hippocampus, prefrontal cortex, and striatum. Many studies have shown that BDNF–TrkB signaling plays a key role in synaptic plasticity in the adult brain. Based on application of exogenous BDNF, genetic manipulations, or blocking pharmacologic reagents, it has been established that BDNF–TrkB signaling promotes, but is not required for, the induction of high-frequency stimulation-induced LTP in the hippocampus. Neurotrophins have also been shown to evoke a chemical form of LTP. Brief application of exogenous BDNF produces long-lasting synaptic enhancement in hippocampus. Interestingly, despite strong evidence that BDNF release and BDNF-dependent spine regrowth and synaptic plasticity are required for learning and memory, this chemical "BDNF-LTP" remains only partially understood. What is currently understood about BDNF-LTP is that intracellular calcium signaling is involved, as well as some key downstream pathways involving extracellular signal-regulated kinases (ERKs). Thus, in the postnatal CNS, BDNF can modulate synaptic strengthening and neurotransmission as well as promote axonal and dendritic growth and differentiation. Conversely, the precursor of BDNF, proBDNF, has been shown to facilitate LTD in hippocampal slices by activation of $p75^{NTR}$. Exogenous proBDNF application induces synaptic depression at the neuromuscular synapse, suggesting that proBDNF opposes the actions of mature BDNF on synaptic plasticity. In some contexts, such as in developing neuromuscular synapses, proBDNF and mature BDNF have been shown to be potential "punishment" and "reward" signals, respectively, for inactive and active terminals. Specifically, activity-dependent activation of extracellular proteases can lead to the conversion of proBDNF to mature BDNF in active terminals, leading to stabilization of the synapse. Conversely, $p75^{NTR}$-expressing terminals will bind the proBDNF and are given the "punishment" signal to retract. These counterbalancing effects of proBDNF and mature BDNF are likely most pronounced in the early stages of development, when $p75^{NTR}$ levels are the highest. It has also been shown in rodent models of chronic stress and depression, that there are increased levels of proBDNF which normalizes with fluoxetine treatment. However, human studies comparing peripheral proBDNF and mature BDNF levels in patients with major depressive disorder (MDD) or other related disorders have been largely mixed and inconclusive. Therefore, it is unlikely that levels of proBDNF and mature BDNF in patients would serve as reliable biomarkers for psychiatric disease or response to treatment.

NEUROTROPHINS AND METABOLISM

BDNF and TrkB have also been established to play key roles in the central control of satiety and energy balance regulation. Both processes are regulated by a combination of CNS circuits predominantly in hypothalamic regions, as well as hormones and neurotransmitters produced centrally and by peripheral tissues. Analyses of loss of function genetic knockout mice with reduced expression of either BDNF or TrkB led to consistent phenotypes of increased food intake, as well as body weight. Recent studies have identified the paraventricular hypothalamus (PVH), a heterogenous cluster of nuclei containing separate neuronal populations expressing BDNF and TrkB, as a key brain region for neurotrophin-dependent suppression of appetite. Selective deletion of BDNF in the anterior PVH leads to significant hyperphagia and weight gain, while injection of recombinant BDNF in the PVH of control animals suppresses food intake. Deletion of TrkB in a separate set of PVH neurons also leads to significant hyperphagia and weight gain, which could be mimicked by decreasing neuronal activity in these neurons expressing TrkB.

BDNF–TrkB has also been implicated in the regulation of energy expenditure, which involves basal metabolism, physical activity, and thermoregulation. Deletion of BDNF in the same anterior PVH subregion led to decreased energy expenditure and locomotor activity. This is likely through projections of these neurons to the spinal cord sympathetic neuronal ganglia network-controlling thermogenesis. Conversely, injection of recombinant BDNF in the PVH of control animals led to increased energy expenditure. As BDNF–TrkB has been shown in other brain circuits to mediate enhanced synaptic plasticity, it is likely that BDNF regulation of appetite suppression and energy balance modulates synaptic plasticity within these hypothalamic circuits.

In parallel, human genetic studies have directly linked mutations in BDNF and TrkB with obesity and increased body mass index. A rare de novo missense mutation in TrkB’s kinase domain as well as de novo mutations leading to BDNF haploinsufficiency have been identified in subjects with hyperphagia and severe obesity. Genome-wide association studies of measures of obesity have identified several SNPs in or near the BDNF gene that have met criteria for genome-wide significance including the BDNF Val66Met allele. Finally, a randomized controlled study of human BDNF Met allele carriers taking antipsychotics showed that the BDNF Met allele is a risk factor for antipsychotic-associated obesity and insulin resistance. Together, these findings in rodents and in humans identify novel roles of BDNF and TrkB in the central control of energy expenditure and feeding behavior. While the brain circuitry involved in these processes diverges from their established roles in learning and memory and mood regulation, the underlying facilitation of synaptic transmission and plasticity are likely common themes for their divergent roles involved in this form of regulation of metabolism.

CORRELATES TO NEURODEGENERATIVE DISORDERS

Neurotrophins have been implicated in the pathophysiology of a wide variety of neurodegenerative and neuronal injury disorders. The brain’s response to acute and chronic environmental changes leads to alterations in neuronal function. The hypothesis underlying these clinical correlations, as well as the development of therapeutic strategies using neurotrophic factors, assumes that these disease states result in the following: decreased availability of neurotrophins for the affected neurons; fewer neurotrophin receptors on the affected neurons; and/or decreased neuronal survival. These deficits can be ameliorated by the addition of neurotrophic factors.

An example of how decreased availability of neurotrophins can account for a neuropsychiatric clinical problem is Huntington disease (HD), which is characterized by progressive death of medium spiny neurons in the striatum. One mechanism to explain the pathogenic effects of mutant huntingtin protein is through a decrease in BDNF levels brought on by lower anterograde transport from the cortex to the striatum. Similar logic applies to AD, one of the most common neurocognitive disorders. The addition of NGF or BDNF to vulnerable regions of the brain, such as the entorhinal cortex or hippocampus, increases spatial learning and memory retention. In these disease states, the assumption has been that exogenous neurotrophic factors would provide symptomatic treatment for the disease state rather than a cure for the core pathophysiology of these nervous system disorders. A more complete cell expression profile of the neurotrophins and their receptors in well-defined circuits will provide strategies to target degenerative disorders of different origins.

CORRELATES TO PSYCHIATRIC DISORDERS

The profound effects of neurotrophic factors on synaptic connections, synaptic plasticity, and neurotransmission have formed the basis for their association with psychiatric disorders, such as anxiety disorders, depression, and substance abuse. Neurotrophins became attractive molecular intermediates that may be involved in the pathophysiology of psychiatric disorders. Environmental inputs to neurotrophins can presumably lead to alterations in neuronal circuitry and ultimately, behavior. It has become clear that neurotrophins can produce long-term changes by regulating transcriptional programs on the functioning of adult neurons. This could explain the long delay in therapeutic action of many psychiatric treatments. Again, the clinical correlation assumes that there is a deficit in access or responsiveness to neurotrophic factors contributing to the phenotype of the disease state.

Major Depressive Disorder

The strongest evidence for a role of neurotrophins has come from the pathophysiology of depression, especially those associated with stress. For depression, it is believed that there is a fundamental dysregulation of synaptic plasticity and neuronal survival in regions of the brain such as the hippocampus. There are several lines of evidence suggesting a role of neurotrophins in depression. First, in animal models, restraint stress leads to decreased expression of BDNF in the hippocampus. In addition, chronic physical or psychosocial stress leads to atrophy and death of hippocampal neurons especially in the CA3 region in rodents and primates. Also, magnetic resonance imaging (MRI) studies have shown that patients with depression or post-traumatic stress disorder (PTSD) exhibit a small decrease in hippocampal volume. It is unclear though whether the atrophy and/or death of these neurons is directly related to the decreased availability of BDNF. In addition, not all forms of depression are associated with stress. However, if structural remodeling and synaptic plasticity are involved in the cellular pathophysiology of depression, BDNF is an attractive candidate molecule to mediate these alterations.

Second, exogenously administered BDNF in the hippocampus had antidepressant effects in two animal models of depression (i.e., the forced swim and learned helplessness paradigms) comparable to those of chronic treatment with pharmacologic antidepressants. In addition, BDNF has also been shown to have trophic effects on serotonergic and noradrenergic neurons in vitro and in vivo. Mutant mice with decreased levels of BDNF have been shown to have a selective decrement in serotonergic neuron function and corresponding behavioral dysfunction consistent with serotonergic abnormalities.

Third, serotonin and norepinephrine reuptake inhibitor antidepressants upregulate BDNF and CREB, a cAMP-dependent transcription factor, in a time course that corresponds to therapeutic action (10 to 20 days). The CREB transcription factor is involved in the induction of BDNF gene expression in neurons. This effect on the cAMP pathway provides a link between monoamine antidepressants and neurotrophins. Antidepressant treatments also lead to increases in expression of TrkB receptors in the hippocampus in a time course that also parallels the extended time course of therapeutic action of these treatments.

Notably, although the effects of prolonged serotonin and norepinephrine reuptake inhibitor treatment involve enhancing neurotrophin signaling, exogenously administered BDNF in the mesolimbic dopamine system appears to have an opposite effect—increasing depression-like behavior. In addition, removal of BDNF in this dopamine circuit appears to have antidepressant effects on a social defeat paradigm. These findings emphasize the complexity of BDNF's role in mediating aspects of behavior related to depression. Together, these studies provide a framework to examine further the neurotrophin system as a potential therapeutic target for the treatment of depression.

NEUROTROPHIC HYPOTHESIS FOR MOOD DISORDERS: THERAPEUTIC POTENTIAL OF NEUROTROPHINS

The recent clinical trials have provided limits in designing therapeutic strategies to use neurotrophic factors for neurodegenerative and psychiatric disorders. First, it has become clear that the physical delivery of sufficient quantities to target neurons is a major obstacle. Development of small molecules that readily cross the blood–brain barrier to activate neurotrophin receptors or potentiate the actions of neurotrophins is an approach that is in its infancy. Second, because neurotrophins have multiple effects on neuronal activity, indiscriminate "flooding" of the CNS with neurotrophic factors will likely lead to untoward side effects, such as epileptic activity. It had been noted in the clinical trials that there is a downregulation of the TrkB receptors after unregulated application of BDNF, likely contributing to the minimal therapeutic effects. New strategies are being studied that include more local and regulated application of neurotrophins through stereotactic injection of viral vectors or engineered progenitor cells. This approach is currently being applied to AD, in which there is a defined neuronal population in the basal forebrain cholinergic neurons that undergoes degeneration and is dependent on a specific neurotrophin (NGF).

The activation of the neurotrophin system through other receptor signaling systems offers an alternative strategy. For example, antidepressant agents acting via monoamine G-protein–coupled receptors

(GPCRs) can lead to increased expression of both neurotrophins and neurotrophin receptors. Importantly, only the neurons that express the monoamine GPCRs will have enhanced production of the neurotrophin or Trk receptor. It has also been shown that other GPCRs, the purine adenosine 2A (A2A) receptor, and pituitary adenylate-cyclase–activating peptide (PACAP) neuropeptide receptor can transactivate Trk neurotrophin receptors in the absence of neurotrophins in hippocampal neurons in vitro. Therefore, small molecules can activate Trk receptors in the absence of neurotrophins. These results raise the possibility that small molecules may be used to elicit neurotrophic effects for the treatment of neurodegenerative diseases by selective targeting of neurons that express specific GPCRs and Trk receptors.

Another gold-standard treatment for treatment-resistant and late-life depression that has been shown to depend on BDNF is electroconvulsive therapy (ECT). In rodents, long-term ECT has been shown to increase BDNF transcription and elicit the sprouting of hippocampal neurons. This effect was attenuated in mutant mice that express lower levels of BDNF. Despite the promising rodent studies, clinical studies have been more mixed in terms of the effects of ECT on increasing BDNF levels. Some show increases in serum and cerebrospinal fluid BDNF levels with ECT treatments while others do not. Studies that have attempted to look at BDNF as a biomarker for treatment response have concluded that BDNF levels were a poor predictor of ECT treatment outcome. This is likely because BDNF is affected by wide variety of genetic, biologic, and environmental factors. More rigorously powered studies are necessary to determine whether the clinically significant therapeutic effects of ECT are BDNF dependent.

Other potent ways of increasing the level of neurotrophins involve using specific classes of antidepressant drugs. Various serotonin and serotonin–norepinephrine reuptake-inhibiting antidepressants are clinically effective only after several weeks of administration. Newer approaches have examined the effects of ketamine, a fast-acting, noncompetitive N-methyl-D-aspartate (NMDA) receptor antagonist. Ketamine has been shown to mediate a rapid antidepressant response through a BDNF-dependent mechanism. NMDA receptor blockade via ketamine leads to the deactivation of eukaryotic elongation factor-2 kinase (eEF2 kinase). This alleviates the block on BDNF translation, resulting in rapid BDNF protein synthesis (within 30 minutes). This rapid upregulation of BDNF then triggers spine remodeling, synaptogenesis, and induction of synaptic plasticity. Alternatively, the blockade of NMDA receptors by ketamine can lead to a rapid, transient glutamate burst that subsequently stimulates AMPA receptors. The stimulation of AMPA receptors leads to local release of neurotrophins, including BDNF and vascular epithelial growth factor (VEGF). Both BDNF and VEGF then stimulate their respective receptors TrkB and Flk-1 to signal downstream to activate Akt and mTORC1. This leads to translation of various synaptic proteins that promote synaptogenesis and synapse maturation. These findings suggest that the rapid effects of glutamatergic modulators, such as ketamine, may differ from conventional antidepressants in that the pools of neurotrophins that are being upregulated and released in an activity-dependent way lead to enhancement of cellular signaling at separate synaptic sites.

It should be emphasized that all these strategies that utilize neurotrophic factors assume that pathologic symptoms arise from impaired neurons and neuronal circuits. This impairment implies not only cell survival, but also proper physiologic functioning and connectivity of these neurons. With greater understanding of the signal transduction pathways that are activated by neurotrophins, alternate strategies can be devised to manipulate these pathways through new drug development. A completely different approach is to block neurotrophins, such as NGF, which elicits hyperalgesia after tissue damage or inflammation. Development of neutralizing antibodies against NGF and their administration to reduce pain thresholds has proven to be efficacious in many clinical pain conditions. For example, NGF has been shown to activate the nociceptive neurons that send chronic pain signals of osteoarthritis (OA) from the PNS to CNS. Treatment with anti-NGF antibodies has been shown to inhibit joint pain and improve functions in those with moderate to severe OA. Further understanding of the core pathophysiologic mechanism for neurodegenerative and psychiatric disorders will benefit the development of rational therapies that involve engaging the neurotrophin system.

ACKNOWLEDGMENTS

Funding: Burroughs Wellcome Fund (JK), R01NS052819 (FSL), R01MH123154 (FSL).

FURTHER READINGS

An JJ, Liao G-Y, Kinney CE, Sahibzada N, Xu B. Discrete BDNF neurons in the paraventricular hypothalamus control feeding and energy expenditure. *Cell Metab*. 2015;22:175–188.

Bai Y-Y, Ruan C-S, Yang C-R, et al. ProBDNF signaling regulates depression-like behaviors in rodents under chronic stress. *Neuropsychopharmacology*. 2016;41:2882–2892.

Berton O, McClung CA, Dileone RJ, et al. Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. *Science*. 2006;311:864–868.

Cabelli RJ, Hohn A, Shatz CJ. Inhibition of ocular dominance column formation by infusion of NT4/5 or BDNF. *Science*. 1995;267:1662–1666.

Casey BJ, Glatt CE, Lee FS. Treating the developing versus developed brain: translating preclinical mouse and human studies. *Neuron*. 2015;86:1358–1368.

Castrén E, Monteggia LM. Brain-derived neurotrophic factor signaling in depression and antidepressant action. *Biol Psychiatry*. 2021;90:128–136.

Chao MV. Neurotrophins and their receptors: a convergence point for many signaling pathways. *Nat Rev Neurosci*. 2003;4:299–309.

Chen ZY, Jing DQ, Bath KG, et al. Genetic variant BDNF (Val66Met) polymorphism alters anxiety-related behavior. *Science*. 2006;314:140–143.

Deyama S, Bang E, Wohleb ES, et al. Role of neuronal VEGF signaling in the prefrontal cortex in the rapid antidepressant effects of ketamine. *Am J Psychiatry*. 2019;176:388–400.

Duman RS, Heninger GR, Nestler EJ. A molecular and cellular theory of depression. *Arch Gen Psychiatry*. 1997;54:597–606.

Felmingham KL, Dobson-Stone C, Schofield PR, Quirk GJ, Bryant RA. The brain-derived neurotrophic factor Val66Met polymorphism predicts response to exposure therapy in posttraumatic stress disorder. *Biol Psychiatry*. 2013;73(11):1059–1063.

Giza JI, Kim J, Meyer HC, et al. The BDNF Val66Met polymorphism disassembles dendritic spines, altering fear extinction circuitry and behavior. *Neuron*. 99(6):1356.

Je HS, Yang F, Ji Y, Nagappan G, Hempstead BL, Lu B. Role of pro-brain-derived neurotrophic factor (proBDNF) to mature BDNF conversion in activity-dependent competition at developing neuromuscular synapses. *Proc Natl Acad Sci U S A*. 2012;109:15924–15029.

Lee FS, Chao MV. Activation of Trk neurotrophin receptors in the absence of neurotrophins. *Proc Natl Acad Sci U S A*. 2001;98:3555–3560.

Lee FS, Kim AH, Khursigara G, Chao MV. The uniqueness of being a neurotrophin receptor. *Curr Opin Neurobiol*. 2001;11:281–286.

Lee R, Kermani P, Teng KK, Hempstead BL. Regulation of cell survival by secreted proneurotrophins. *Science*. 2001;294:1945–1948.

Li A, Jing D, Dellarco DV, et al. Role of BDNF in the development of an OFC-amygdala circuit regulating sociability in mouse and human. *Mol Psychiatry*. 2021;26:955–973.

Lyons WE, Mamounas LA, Ricaurte GA, et al. Brain-derived neurotrophic factor-deficient mice develop aggressiveness and hyperphagia in conjunction with brain serotonergic abnormalities. *Proc Natl Acad Sci U S A*. 1999;96:15239–15244.

Mariga A, Mitre M, Chao MV. Consequences of brain-derived neurotrophic factor withdrawal in CNS neurons and implications in disease. *Neurobiol Dis*. 2017;97:73–79.

Minichiello L, Calella AM, Medina DL, Bonhoeffer T, Klein R, Korte M. Mechanism of TrkB-mediated hippocampal long-term potentiation. *Neuron*. 2002;36:121–137.

Park H, Poo MM. Neurotrophin regulation of neural circuit development and function. *Nat Rev Neurosci*. 2013;14:7–23.

Petryshen TL, Sabeti PC, Aldinger KA, et al. Population genetic study of the brain-derived neurotrophic factor (BDNF) gene. *Mol Psychiatry*. 2010;15(8):810–815.

Riccio A, Ahn S, Davenport CM, Blendy JA, Ginty DD. Mediation by a CREB family transcription factor of NGF-dependent survival of sympathetic neurons. *Science*. 1999;286:2358–2361.

Shirayama Y, Chen ACH, Nakagawa S, Russell DS, Duman RS. Brain-derived neurotrophic factor produces antidepressant effects in behavioral models of depression. *J Neurosci*. 2002;22:3251–3261.

Song M, Martinowich K, Lee FS. BDNF at the synapse: why location matters. *Mol Psychiatry*. 2017;22:1370–1375.

Strand AD, Baquet ZC, Aragaki AK, et al. Expression of profiling of Huntington’s disease models suggests BDNF depletion plays a major role in striatal degeneration. *J Neurosci*. 2007;27:11758–11768.

Thoenen H. Neurotrophins and neuronal plasticity. *Science*. 1995;270:593–598.

Van Zutphen EM, Rhebergen D, van Exel E, et al. Brain-derived neurotrophic factor as a possible predictor of electroconvulsive therapy outcome. *Transl Psychiatry*. 2019;9:155.

Wise BL, Seidel MF, Lane NE. The evolution of nerve growth factor inhibition in clinical medicine. *Nat Rev Rheumatol*. 2021;17:34–46.

Zhao M, Chen L, Yang J, et al. BDNF Val66Met polymorphism, life stress and depression: a meta-analysis of gene-environment interaction. *J Affect Disord*. 2018;227:226–235.

1.11 Intraneuronal Signaling

John A. Gray, M.D., Ph.D.

OVERVIEW OF NEURONAL SIGNAL TRANSDUCTION

Signal transduction is the process by which a cell converts extracellular signals to intracellular signals and the subsequent cascade of events that leads to alterations in cellular function. The initial step in signal transduction usually involves the binding of an extracellular signal, such as a neurotransmitter, to a designated plasma membrane receptor.

Binding of this molecule, or ligand, to its receptor stabilizes a unique conformation of the receptor, resulting in the transmission of the signal across the plasma membrane. In its simplest form, the binding of a neurotransmitter to an ion channel stabilizes a conformation of the channel protein that allows the channel to open and ions to flow into or out of the cell, setting off a cascade of intracellular events. In more complex examples of signal transduction, the stabilized conformation of the receptor allows for the binding of other proteins to the intracellular portions of the receptor. These intracellular proteins subsequently become "activated" and go on to initiate various downstream events.

Before the details of each signaling pathway are discussed, it is useful to understand the common themes in the flow of information from receptor binding to the final alterations in neuronal function. In general, these systems are organized into several layers (Fig. 1.11–1). First, extracellular signals are detected by receptors and transmitted across the plasma membrane to adaptor proteins. These adaptor proteins then activate one or more intracellular signaling pathways, which, in turn, alter the function of effector proteins, either directly or via intermediates, such as protein kinases. A protein kinase is an enzyme that adds a phosphate group to a protein. Protein phosphorylation is a primary mechanism in signal transduction as phosphorylation changes a protein's conformation, which can alter its enzymatic activity or its ability to bind with other proteins. Typically, protein phosphorylation leads to the activation of a protein. The eventual outcome of these signaling pathways is the alteration of neuronal activity and changes in the expression of various genes.

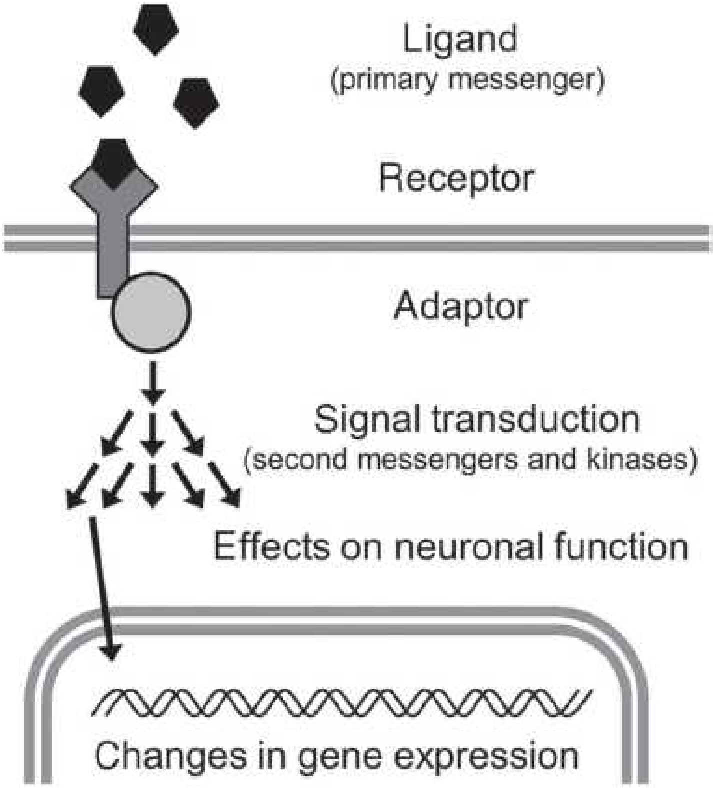
Ligand
(primary messenger)
Receptor
Adaptor
Signal transduction
(second messengers and kinases)
Effects on neuronal function
Changes in gene expression

FIGURE 1.11–1. General outline of signal transduction. Transmission of an extracellular signal to an intracellular signal usually involves the binding of a ligand, such as a neurotransmitter, to a receptor. Conformational changes in the receptor protein transmit the signal across the plasma membrane to adaptor proteins which then activate one or more intracellular signaling pathways. These signaling pathways often involve the generation of small-molecule second messengers or the activation of kinase cascades leading to changes in neuronal function and gene expression.

So why do neurons and other cells have these complex intracellular signaling pathways? First, in addition to transmitting an extracellular signal across the plasma membrane, these signal transduction pathways amplify the signal, allowing for cells to have large responses to very minute quantities of extracellular stimuli. Furthermore, the multiplicity of intracellular signaling pathways allows signals to be directed in a specific manner, thus enabling cells to maintain separate channels of information that can be integrated only when appropriate. For example, these separate channels of information allow neurons to detect when different stimuli are presented concurrently, thus enabling them to alter their response accordingly. In addition, each signaling pathway has distinctive spatial and temporal characteristics that allow for the optimal handling of different types of information. For example, in some instances, it may be advantageous for a neuron to have extremely high sensitivity to an uncommon stimulus but to ignore repetitive inputs. Thus, these complex signaling pathways determine a neuron's sensitivity and responsiveness to its environment in the context of its current circumstances and its past experiences.

Overall, a detailed understanding of the complex biochemical processes operating inside neurons is critical to appreciate how the brain not only responds to individual stimuli but also how it can continuously adapt to endless environmental changes. In addition, advances in the understanding of the molecular processes occurring within neurons will lead to improved insight into the basis of behavior and psychotropic drug action and will likely guide the development of improved psychiatric treatments and diagnostic tools.

Major Neuronal Signaling Pathways

There are three main schemes of signal transduction in neurons (Fig. 1.11–2). The first involves ligand-gated ion channels, which are the primary mechanism of signal transduction for amino acid neurotransmitters such as glutamate and γ-aminobutyric acid (GABA). In addition, a number of other neurotransmitters, including acetylcholine and serotonin, have a subset of receptors that are ligand-gated ion channels. Upon binding of a neurotransmitter, these ion channels are stabilized in a conformation that alters the conductance of the channel to particular ions, usually Na^+, K^+, Ca^{2+}, or Cl^-. At synapses, these receptors can very quickly convert an extracellular signal into a postsynaptic electrical signal constituting so-called “fast synaptic transmission.” A prototypical example of a ligand-gated ion channel is the nicotinic acetylcholine receptor, which consists of five protein subunits with two binding sites for acetylcholine enclosing a central aqueous pore. When both acetylcholine-binding sites are filled, the internal pore of the channel opens, allowing Na^+ ions to flow into the cell, down their electrochemical gradient. Channels that allow Ca^{2+} to enter the cell can have effects on the electrical properties of the cell and can stimulate additional calcium-mediated intracellular signaling cascades, as discussed below.

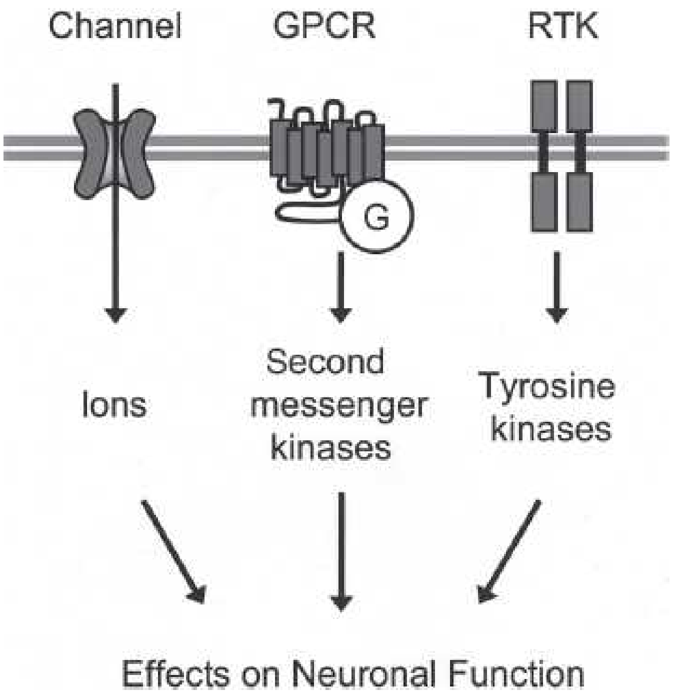

FIGURE 1.11–2. Outline of the three major receptor types mediating intraneuronal signal transduction in neurons: ligand-gated ion channels, G-protein–coupled receptors (**GPCR**s), and receptor tyrosine kinases (**RTK**s). Each receptor type, when activated by its extracellular signal, induces intracellular signaling pathways that result in alterations of neuronal function. G, heterotrimeric G protein.

A second primary scheme of neuronal signal transduction involves the binding of a neurotransmitter to seven-transmembrane domain receptors. These receptors are also known as GPCRs because they activate heterotrimeric guanine nucleotide–binding proteins (G proteins). GPCRs represent the single largest family of receptors (with more than 700 members) in the genome and are the primary form of receptor for many neurotransmitters, including serotonin and dopamine. GPCRs primarily signal by activating G proteins that subsequently activate effector enzymes that generate small molecules termed "second messengers" (the "first messenger" being the extracellular signal itself). Second messengers, in turn, mediate many of the downstream intracellular signaling cascades, largely involving protein kinases. Because of the additional step of creating small-molecule second messengers, signaling through GPCRs generally requires more time to develop than the opening of a ligand-gated ion channel, and thus this scheme accounts for the majority of "slow synaptic transmission" in neurons.

The third common scheme of signal transduction in the brain involves the activation of a distinct class of protein kinases that phosphorylate proteins on tyrosine residues. Activation of these protein tyrosine kinases is the primary signaling pathway for most neurotrophic factors, such as NGF and BDNF, as well as various cytokines and chemokines. The binding of a neurotrophic factor or chemokine to its respective plasma membrane receptor leads to the dimerization of the receptor, with another copy of the receptor transmitting the signal across the plasma membrane. This results in the activation of a cascade of protein kinases. In some cases, the intracellular portion of the receptor itself contains the first tyrosine kinase in the cascade, and in other cases, the dimerized receptors recruit cytoplasmic tyrosine kinases that then become activated. While the cascades of kinases can be complicated, the ultimate purpose is to amplify the initial signal and affect numerous changes in neuronal function.

G-PROTEIN–COUPLED RECEPTOR SIGNALING

Accounting for at least 2% of the genes in the genome, GPCRs comprise a very large family of proteins that represent targets for a wide array of molecules ranging from hormones and neurotransmitters to odorants and even light. The binding of an agonist to its cognate GPCR stabilizes an active conformation of the receptor, inducing the dissociation and activation of a receptor-specific heterotrimeric G protein into its α- and βγ-subunits (Fig. 1.11–3). G proteins were discovered by Alfred Gilman, Martin Rodbell, and colleagues and are a family of proteins named because they use the exchange of guanosine diphosphate (GDP) and guanosine triphosphate (GTP) as a molecular “switch” to regulate cell processes. In their baseline state, heterotrimeric G proteins consist of α-, β-, and γ-subunits, with a GDP molecule bound to the α-subunit. When a GPCR is activated by an agonist, the G-protein α-subunit becomes associated with the receptor, causing conformational changes that release the GDP molecule. This allows a GTP molecule to bind, thus activating the α-subunit. GTP binding to the α-subunit also causes the dissociation of the α-subunit from the βγ-subunits and from the receptor. These dissociated subunits are now biologically active and activate or inhibit a number of downstream effectors, including nucleotide cyclases, phospholipases, kinases, and ion channels, resulting in a variety of downstream cellular effects. The system is returned to its basal state when the α-subunit, which has intrinsic GTPase activity, hydrolyzes the GTP back to GDP. This hydrolysis of GTP to GDP within the α-subunit also leads to the reassociation of the α-subunit with the βγ-subunits and thus the restoration of the inactive heterotrimer. If the receptor remains bound to its agonist, then the GDP can again dissociate from the α-subunit, and another G-protein cycle begins; however, if the receptor becomes inactive, the hydrolysis of GTP to GDP halts intracellular signaling.

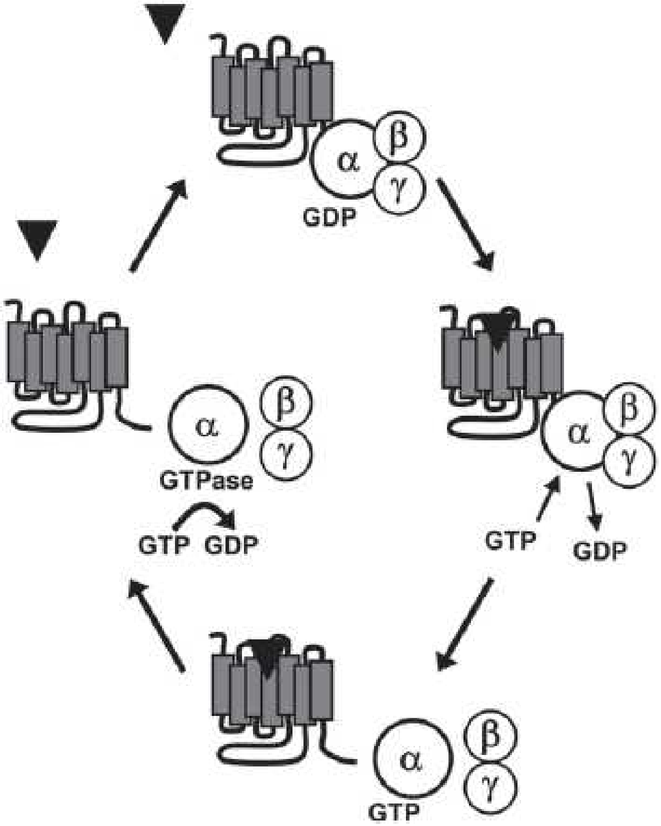

FIGURE 1.11–3. Outline of G-protein function. Under basal conditions, G proteins exist as heterotrimers consisting of single α-, β-, and γ-subunits, with the inactive α-subunits bound to GDP. After the G-protein–coupled receptor is activated by its ligand, it causes the associated G protein to release GDP, allowing GTP to bind. GTP binding to the α-subunit causes the dissociation of the α-subunit from the βγ-subunits and from the receptor. The free G-protein subunits are functionally active and can activate and regulate a variety of downstream effector proteins. The α-subunit has intrinsic GTPase activity that hydrolyzes the GTP back to GDP, causing the reassociation of the α- and βγ-subunit, restoring the system to its basal state.

The primary effector systems mediated by G proteins share a common overall form, whereby the activated G protein activates an effector enzyme that generates small-molecule second messengers that then initiate specific protein kinase cascades. The particular class of G proteins associated with each individual receptor dictates which specific second messenger system is activated. The four major classes of G proteins that are involved in neurotransmitter-mediated signaling are called G_s, G_i, G_q, and $G_{12/13}$ and are discussed below with the details of each second messenger system. Briefly, both the G_s and G_i classes of G proteins regulate the cyclic adenosine monophosphate (cAMP) system, while the G_q class regulates the phosphoinositide (PI) signaling system. The $G_{12/13}$ class modulates Rho/Rac signaling cascades but will not be discussed. Importantly, a single receptor can activate multiple G proteins, which can then activate multiple effector enzymes that synthesize many second messenger molecules, leading to exponential amplification of the initial signal.

Cyclic Adenosine Monophosphate Pathway

The discovery of cAMP in the 1950s by Earl Sutherland and Theodore Rall established the concept that small intracellular molecules can act as second messengers that convey information from cell-surface receptors to their targets within the cell. Since that seminal discovery, decades of intensive research on the cAMP system have elucidated its operating principles, making it the prototypical second messenger system.

GPCRs that activate the cAMP signaling pathway are coupled to the G_s class of G proteins. When G_s becomes activated by a receptor, it dissociates from the receptor and stimulates a membrane-bound effector enzyme called adenylate cyclase that converts ATP to cAMP (Fig. 1.11–4). Conversely, other receptors are coupled to the G_i class of G proteins that inhibit adenylate cyclase and thus decrease cAMP production. The net level of cAMP production by a given neurotransmitter is thus determined by the specific G_s- and G_i-coupled receptor subtypes expressed on a given neuron or synapse. For example, norepinephrine stimulates adenylate cyclase via its interaction with β-adrenergic receptors and inhibits adenylate cyclase via stimulation of α_2-adrenergic receptors.

The major target of action of cAMP in most cells is the cAMP-dependent protein kinase, also known as protein kinase A (PKA), which mediates many of the actions cAMP has on neuronal function. PKA is a multisubunit serine/threonine kinase consisting of two regulatory subunits and two catalytic subunits. In the absence of cAMP, the regulatory subunits are bound to the catalytic subunits, thus keeping the kinase inactive. However, when cAMP is present, it binds to the regulatory subunits, thus causing a conformation change that dissociates the regulatory subunits from the catalytic subunits. The release of the regulatory subunits activates the catalytic subunits, which are then free to phosphorylate various cellular proteins on serine and threonine residues. This kinase has a broad range of substrate proteins involved in regulating virtually every aspect of neuronal function. In particular, several important neuronal targets for PKA have been identified, including various ion channels, synaptic vesicle machinery, neurotransmitter synthetic enzymes, and proteins involved in regulating gene transcription. Thus, alterations in cAMP levels are able to affect neuronal function over a broad range of time scales. For example, rapid effects are achieved by targeting ion channel gating and neurotransmitter release machinery, while slower effects occur with the targeting of neurotransmitter synthesis and cellular energy metabolism. Furthermore, cAMP elicits longer-lasting changes in neuronal function by controlling the expression of specific target genes.

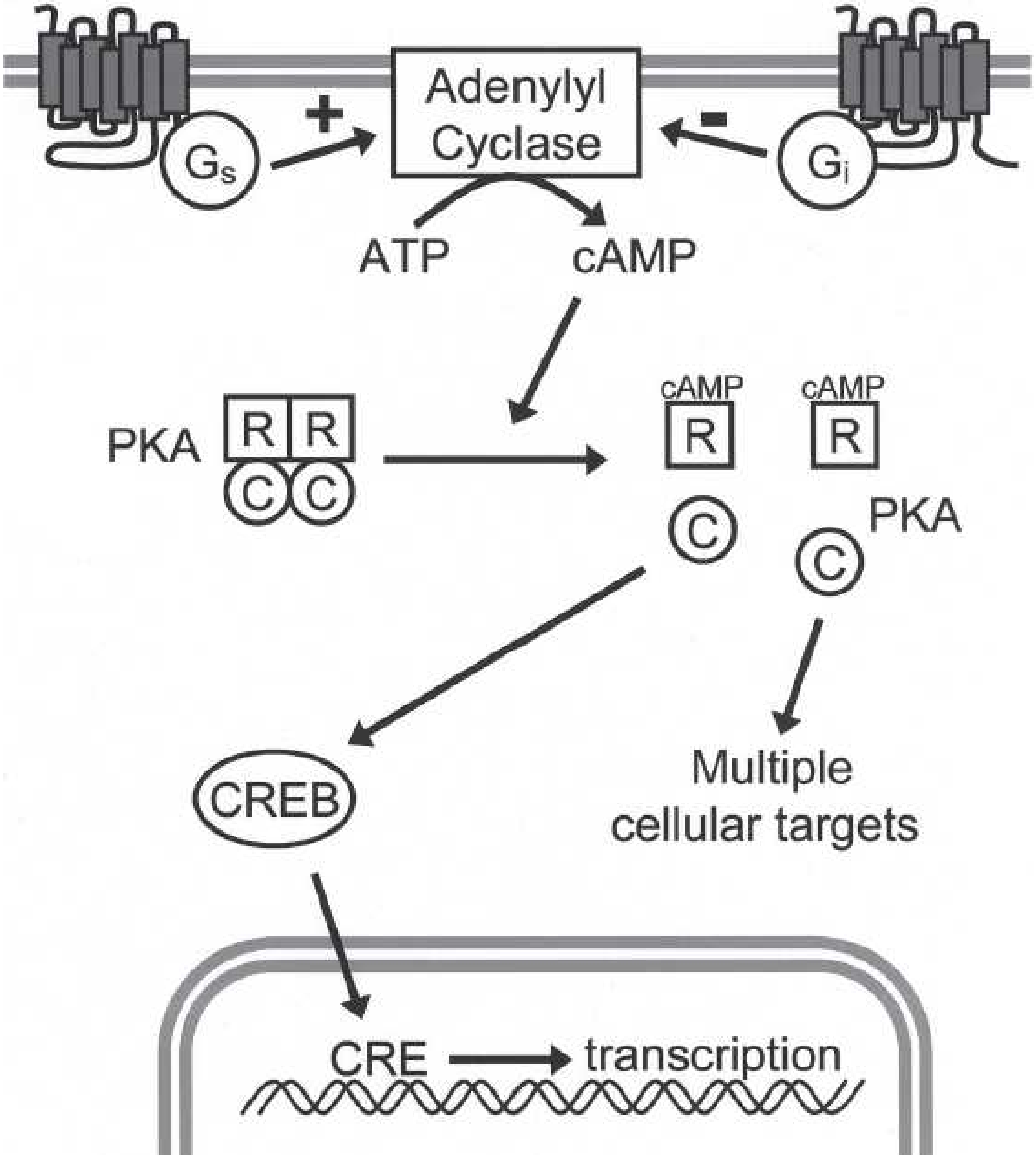
Adenylyl Cyclase
+
-
Gs
Gi
ATP
cAMP
PKA
R
R
C
C
cAMP
cAMP
R
R
C
C
PKA
CREB
Multiple cellular targets
CRE
transcription

FIGURE 1.11–4. Basic organization of the cyclic adenosine monophosphate (cAMP) signaling pathway. Production of cAMP from ATP by adenylate cyclase can either be stimulated or inhibited by G-protein–coupled receptors. Receptors coupled to the G-protein Gs stimulate cAMP synthesis whereas those coupled to Gi inhibit adenylate cyclase. Many of the cellular actions mediated by cAMP occur through protein kinase A (PKA), which exists at baseline as a tetramer of two regulatory subunits (R) that tonically inhibit two catalytic subunits (C). When the R subunits become bound by cAMP, they dissociate from the C subunits, which become activated and can phosphorylate multiple cellular proteins. One such target is the transcription factor cAMP response element–binding (CREB) protein, which, when activated by PKA, can bind to DNA sequences called cAMP response elements (CREs) and promote gene transcription.

One important substrate of PKA is a transcription factor that enables elevations in cAMP to regulate gene expression. This transcription factor, called CREB protein, regulates the expression of various genes by binding to short deoxyribonucleic acid (DNA) sequences called cAMP response elements (CREs). Phosphorylation of CREB by PKA activates it, thus allowing it to bind to CRE sequences in the regulatory regions of target genes, where it increases or decreases the transcription of certain genes. Proteins whose expression is regulated by cAMP through CREB are thought to be involved in various neuronal processes, including neuronal development and survival and the formation of long-term memories.

Termination of cAMP signaling is mediated by the actions of phosphodiesterases (PDEs), enzymes that cleave cAMP to AMP. There are multiple isoforms of PDEs expressed throughout the brain that are differentially regulated, adding a level of complexity to the precise cellular control of cAMP signaling. At high concentrations, caffeine nonselectively inhibits PDEs, possibly contributing to some of its pharmacologic effects. In addition, considerable efforts have been made to develop PDE inhibitors that are selective for individual isoforms. For example, rolipram, an inhibitor selective for type IV PDEs, showed promise as an antidepressant but was limited due to side effects. In addition, inhibitors of PDE10A, an isoform expressed at high levels in the brain, have antipsychotic potential.

Phosphatidylinositol Pathway

After the discovery of the cAMP system, it became apparent that there were many neurotransmitter receptors that did not act via cAMP, suggesting the possible existence of other second messenger systems. Beginning in the 1950s, there were hints that PIs may be involved in various cellular pathways, though a coherent view of this second messenger system did not emerge until the early 1980s. The phosphatidylinositol signaling pathway parallels many basic aspects of the cAMP system (Fig. 1.11–5), though it also includes several unique features. The phosphatidylinositol signaling pathway is initiated by receptors that are coupled to the G_q class of G proteins, which activate the effector enzyme phospholipase C (PLC). This enzyme cleaves phosphatidylinositol bisphosphate (PIP_2), an inositol-containing phospholipid located in the cytoplasmic leaf of the plasma membrane, into two second messengers, diacylglycerol (DAG) and inositol trisphosphate (IP_3). These two second messengers can then go on to affect distinct cellular pathways.

DAG, which is hydrophobic, remains in the plasma membrane where it activates various isoforms of PKC, a serine/threonine protein kinase. In its inactive state, PKC is found in the cytoplasm, but when DAG is generated, PKC relocates to the plasma membrane, becomes activated, and phosphorylates multiple cellular substrates. The water-soluble IP_3 second messenger is released from the plasma membrane and diffuses to the endoplasmic reticulum, where it binds to the inositol trisphosphate receptor. This receptor is a ligand-gated ion channel that, when bound by IP_3, rapidly releases the large stores of Ca^{2+} from the endoplasmic reticulum. This released Ca^{2+} also functions as a second messenger, regulating various cellular functions. In addition to DAG, some isoforms of PKC also require Ca^{2+} to become activated.

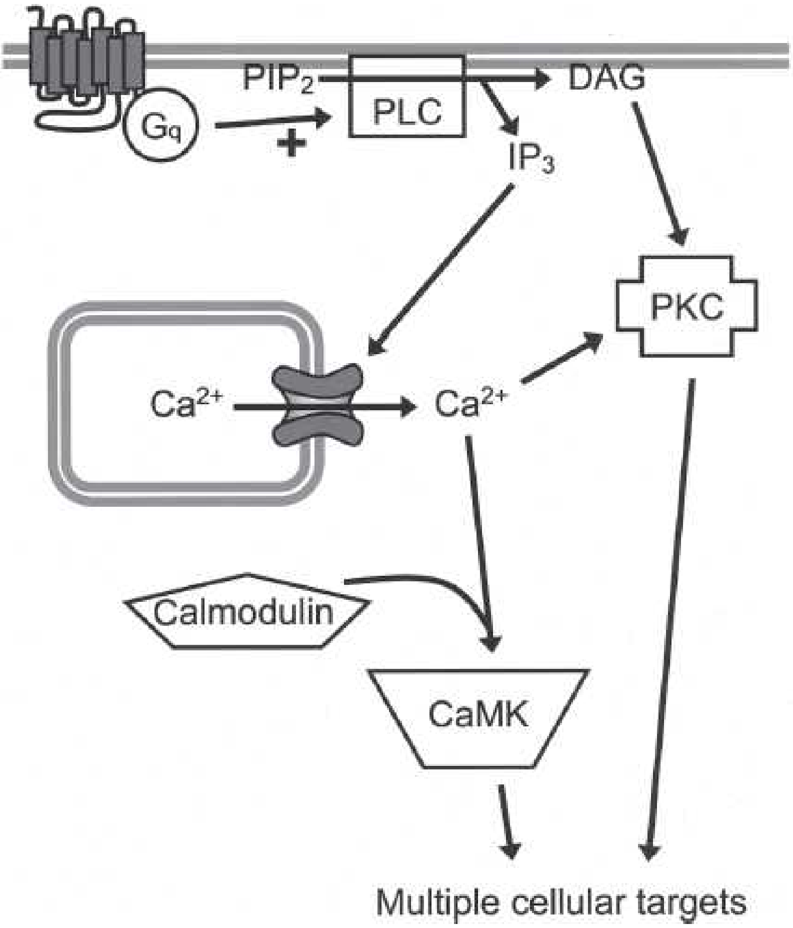

FIGURE 1.11–5. Basic organization of the phosphatidylinositol signaling pathway. The Gq class of G proteins activates phospholipase C (PLC), which cleaves the membrane phospholipid phosphatidylinositol bisphosphate (PIP2) into two second messengers, diacylglycerol (DAG) and inositol trisphosphate (IP3). IP3 diffuses to the endoplasmic reticulum where it binds to the inositol trisphosphate receptor, rapidly releasing large stores of Ca^{2+}, which also functions as a second messenger. DAG activates protein kinase C (PKC), though some isoforms also require Ca^{2+} to become activated. The released Ca^{2+} mediates many of its functions by binding to calmodulin, which can then activate multiple cellular targets, including calcium/calmodulin-dependent kinase (CaMK). PKC and CaMK can phosphorylate multiple cellular targets, including various transcription factors.

The rapid rise in intracellular Ca^{2+}, mediated by IP$_3$-induced release of cellular stores, has both immediate and more delayed effects on neuronal functioning. Immediate effects are triggered by the direct binding of Ca^{2+} itself to various effector proteins and include the release of synaptic vesicles and the opening of calcium-activated ion channels in the plasma membrane. Delayed effects of Ca^{2+} signaling are similar to those of cAMP, such as effects on cellular energy metabolism and gene expression. Many of these slower effects of calcium signaling are mediated by the association of calcium with calmodulin, a small, ubiquitous, calcium-binding protein. Calmodulin is activated when the intracellular concentration of Ca^{2+} is high enough for four ions to bind to the calmodulin protein. Activated calmodulin has multiple cellular targets, including activating various kinases, such as the calcium/calmodulin-dependent (CaM) kinases. It is important to note that calcium-based signaling pathways can also be activated by an influx of Ca^{2+} from the cell surface by various voltage- and ligand-gated ion channels, independent of G-protein signaling.

Termination of the phosphatidylinositol pathway involves multiple steps. DAG is degraded by lipases into glycerol and fatty acids or recycled into membrane phospholipids. Ca^{2+} is rapidly cleared from the cytoplasm by Ca^{2+}–ATPase pumps on the plasma membrane and endoplasmic reticulum, the action of which is enhanced by Ca^{2+} itself through the interaction of activated calmodulin with the transport pump. IP_3 is sequentially dephosphorylated by inositol phosphatases to inositol, which can then be reintegrated into membrane phospholipids. Interestingly, lithium is an inhibitor of these inositol phosphatases and leads to an accumulation of IP_3 and other inositol phosphates within cells. This leads to a depletion of the free cellular inositol needed to replenish membrane PIP_2 for further signaling, prompting the hypothesis that the rundown of the phosphatidylinositol cycle may underlie lithium's therapeutic action, though this remains controversial. Indeed, lithium is also known to inhibit several adenylate cyclases and protein kinases.

Other Second Messenger Systems

In addition to cAMP, another cyclic nucleotide, cGMP, is a second messenger that is regulated by neurotransmitter receptor stimulation. However, there are significant differences between the two systems. Guanylate cyclases are primarily cytoplasmic enzymes that are not directly activated by G proteins but are activated by the gas nitric oxide. Nitric oxide is synthesized in cells by NOS, which is activated by calmodulin and is thus mediated by increases in intracellular Ca^{2+} levels (Fig. 1.11–6). This demonstration that a gas can act as a second messenger blurs the distinction between extracellular and intracellular messengers, as nitric oxide is capable of diffusing across cell membranes and at synapses may act as a retrograde signal to the presynaptic neuron. Synthesis of cGMP also leads to various downstream effects, many through its activation of protein kinase G. Like cAMP, cGMP is degraded by various PDEs. Indeed, drugs for erectile dysfunction, such as sildenafil (Viagra), act by selectively inhibiting a PDE isoform that is highly enriched in the vascular smooth muscle of the penis, highlighting the utility of developing drugs that target intracellular signaling pathways.

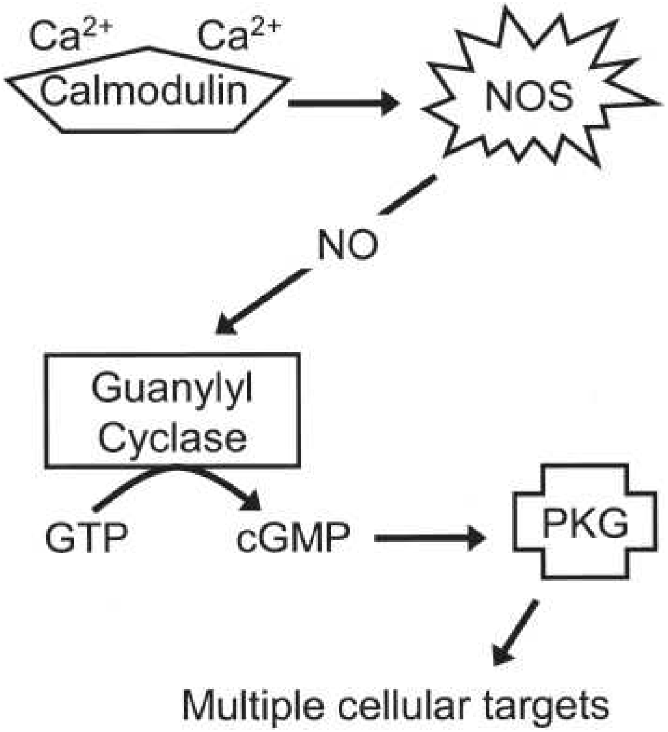

FIGURE 1.11–6. The cyclic guanosine monophosphate (cGMP) signaling pathway. In contrast to the cyclic adenosine monophosphate (cAMP) system, cGMP synthesis via guanylate cyclase is not regulated by G proteins. Instead, guanylate cyclase is activated by nitric oxide, which is synthesized by nitric oxide synthase (NOS) after it is activated by a calcium/calmodulin complex. Like cAMP, cGMP affects neuronal function by stimulating its cognate kinase, protein kinase G (PKG).

Another intracellular signaling system that appears to play an important role in neuronal function involves metabolites of the fatty acid arachidonic acid. Various receptors activate an enzyme called phospholipase A_2, possibly through an unidentified G protein or elevations in cytoplasmic calcium levels. Phospholipase A_2 cleaves membrane phospholipids, typically PIP_2, releasing free arachidonic acid, which is rapidly converted to a number of active metabolites (Fig. 1.11–7). For example, arachidonic acid may be cleaved by cyclooxygenase to yield, after multiple enzymatic steps, several types of prostaglandins and thromboxanes. Alternatively, arachidonic acid may be cleaved by lipoxygenases to yield the leukotrienes. These active metabolites then regulate many intracellular functions, including ion channels and protein kinases, and are important for modulating signaling through other pathways by regulating adenylate and guanylate cyclases. In addition, as these compounds are lipophilic, they can diffuse out of the neuron and act as ligands for their own GPCRs on other neurons. Interestingly, cyclooxygenase inhibitors have been hypothesized to improve schizophrenia and depression, possibly by reducing stress-mediated inflammatory processes in the brain.

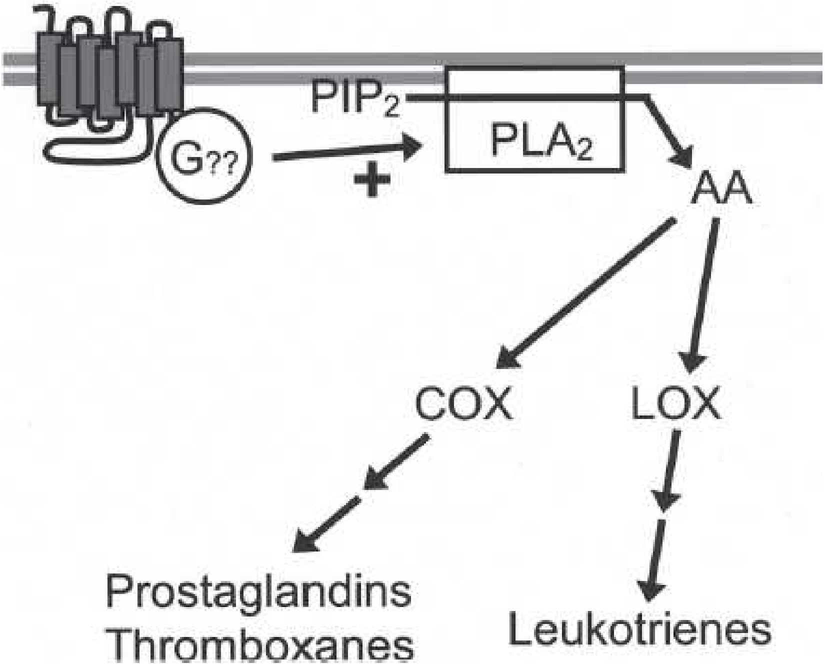

FIGURE 1.11–7. Organization of the arachidonic acid signaling pathway. Various G-protein–coupled receptors activate an enzyme called phospholipase A2 (PLA2), possibly through an unidentified G protein (G??). PLA2 primarily cleaves the membrane phospholipid phosphatidylinositol bisphosphate (PIP2), releasing free arachidonic acid (AA), which is rapidly converted to a number of active metabolites. For example, arachidonic acid may be cleaved by cyclooxygenase (COX) to yield prostaglandins and thromboxanes and by lipoxygenases (LOXs) to yield leukotrienes. These AA metabolites can regulate many intracellular functions and can diffuse out of the neuron and act as ligands for their own G-protein–coupled receptors on other neurons.

Regulation of Ion Channels by GPCRs

As discussed, the primary function of G proteins is to initiate second messenger signaling cascades that go on to affect a multitude of cellular functions, including modulating ion channel gating through phosphorylation by kinases such as PKA, PKC, and CaM kinases. In addition, second messengers such as cAMP and cGMP can regulate ion channels directly. However, it is now clear that the G proteins themselves are also able to directly bind to and regulate ion channels independent of second messenger cascades. This process is best established for receptors that couple to G_i, such as muscarinic acetylcholinergic, α_2-adrenergic, D_2-dopaminergic, and 5-HT_{1A} serotonergic receptors. As before, receptor activation causes the dissociation of the G-protein α- and βγ-subunits. While the α-subunit goes on to inhibit adenylate cyclase, the βγ-subunits can bind directly to the cytoplasmic regions of different ion channels. In some cells, βγ-subunits bind to and directly open specific K^+ channels known as G-protein–regulated inwardly rectifying K^+ (GIRK) channels. These channels are called inwardly rectifying because, if under no electrochemical gradient, they more readily pass current inward; however, under normal physiologic circumstances, K^+ flow through GIRKs is primarily outward, leading to neuronal hyperpolarization. In other cells, βγ-subunits directly inhibit voltage-gated Ca^{2+} channels and can inhibit presynaptic vesicle release. Interestingly, recent evidence suggests that GIRKs may interact directly with GPCRs, likely to promote the near-instantaneous opening of the ion channel following G-protein activation. Another mechanism for the regulation of neuronal ion channels by GPCRs may be through changes in membrane PIP_2 levels. Many ion channels either require PIP_2 to function or are strongly modulated by PIP_2 and other phospholipids. Thus, the activity of these channels may be altered by G_q-mediated PLC activation and local depletion of membrane PIP_2, though the physiologic relevance of this mode of regulation is still under investigation.

Regulation of GPCR Signaling

Because GPCRs play such a key role in cellular signaling, it is not surprising that their activity is tightly regulated. Indeed, regulation occurs at almost every point along the signaling pathways. Above, the termination of second messenger signaling was discussed, such as with the degradation of cAMP and cGMP by PDEs. In this section, regulation of the G proteins and GPCRs will be briefly discussed.

The termination of G-protein signaling is mediated by the intrinsic GTPase activity of the α-subunit, which hydrolyzes GTP to GDP and thus inactivates the G protein. A separate class of proteins, called regulators of G-protein signaling (RGS) proteins, can accelerate the GTPase activity of the α-subunits, thereby shortening the duration of G-protein signaling. There are more than 20 subtypes of RGS proteins that are differentially expressed throughout the brain and are involved in the regulation of all G-protein α-subunits, except for the G_s family. Specific subtypes of RGS proteins have been shown to regulate important neuronal functions, including behavior. For example, mice with a targeted deletion of the gene encoding RGS2, which regulates G_q, show increased anxiety and decreased male aggression, highlighting the important role of RGS proteins in modulating behavior. Interestingly, expression levels of RGS9, which inhibits G_i-mediated dopamine signaling in the striatum, were found to be reduced in the postmortem brains of patients with schizophrenia, consistent with the hypothesis of excessive dopamine signaling resulting in psychosis. Overall, the diversity and heterogeneous distributions of the RGS proteins make them an attractive target for drug development, as drugs affecting individual isoforms may exert highly selective effects.

In addition to initiating intracellular signal transduction cascades, agonist activation of GPCRs also triggers cellular and molecular mechanisms that lead to the attenuation (desensitization) of receptor signaling. Desensitization is an adaptive mechanism to attenuate receptor responsiveness to repetitive environmental stimuli. At the level of the whole organism, the mechanisms underlying receptor desensitization are likely responsible for the development of tolerance to psychopharmacologic agents such as opiates, as well as the delayed therapeutic response to antidepressants and antipsychotics.

Receptor desensitization is typically mediated by feedback phosphorylation of the receptor by a class of kinases called G-protein–coupled receptor kinases (GRKs) (Fig. 1.11–8). The GRKs phosphorylate the intracellular domains of receptors only when the receptors are bound by an agonist. Receptor phosphorylation by GRKs enables a protein called arrestin to bind to the receptor, preventing the G protein from recoupling, thus rendering the receptor inactive. Arrestin binding to a GPCR also causes the receptor to be internalized into endocytic vesicles.

This process is mediated by an interaction of the arrestin molecule with proteins in clathrin-coated pits and membrane invaginations that are pinched off during endocytosis. After internalization, receptors may be recycled back to the cell surface or degraded. While the precise role of receptor endocytosis is not entirely clear, targeting receptors for degradation may be involved in the downregulation of brain receptor levels following chronic drug administration. In addition, while GPCRs have traditionally been thought to signal only from the cell surface, recent studies have shown receptor signaling from intracellular membrane compartments, including endosomes. GPCR signaling from intracellular compartments may be functionally distinct from plasma membrane receptors, and drugs will have differential accessibility to these intracellular receptors. Thus, the regulation of GPCRs by endocytic membrane trafficking may allow for the development of novel pharmacologic agents.

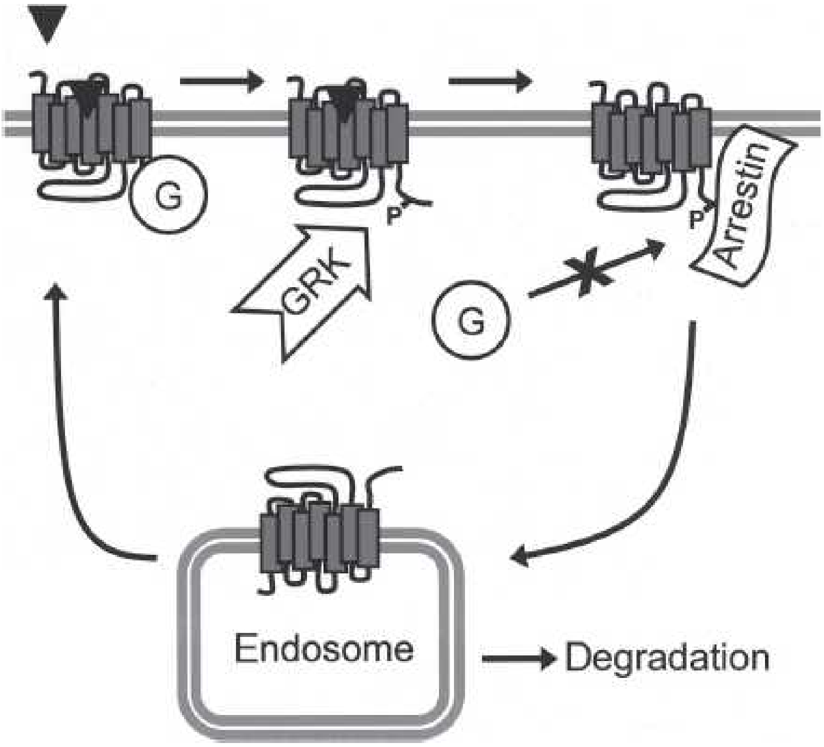

FIGURE 1.11–8. Regulation of G-protein–coupled receptors (GPCRs) by desensitization and internalization. Desensitization is an adaptive mechanism that attenuates receptor responsiveness to repetitive stimuli. GPCR desensitization is typically mediated by feedback phosphorylation (P) by a specific G-protein–coupled receptor kinase (GRK). Receptor phosphorylation by a GRK causes a protein called arrestin to bind to the receptor, effectively preventing G proteins (G) from recoupling to the receptor. Arrestin binding also causes the receptor to be internalized into endocytic vesicles, which may then recycle the receptors back to the cell surface or target them for degradation.

Role of Phosphatases

Because phosphorylation plays such a central role in intracellular signaling pathways, it is not surprising that protein phosphatases, which reverse the effect of protein kinases, also have a major impact on these signaling pathways. There are four major protein phosphatases that are differentially distributed in the brain that dephosphorylate targets of the second messenger kinases, named protein phosphatases 1, 2A, 2B, and 2C. For example, protein phosphatase 2B, also called calcineurin, is activated by the binding Ca^{2+}/calmodulin. Thus, neurotransmitters coupled to G_q proteins as well as Ca^{2+} channels can activate calcineurin and influence the phosphorylation of various cellular proteins. Indeed, phosphatases can be targets for pharmacologic agents, as demonstrated by tacrolimus, an immunosuppressant agent used to prevent organ transplant rejection, which is a selective inhibitor of calcineurin that interferes with T lymphocyte signaling.

Another key mechanism for regulating protein phosphatases involves a separate class of proteins called protein phosphatase inhibitors. These proteins, such as phosphatase inhibitors 1 and 2, are highly potent inhibitors of protein phosphatase 1, a major neuronal phosphatase, and their inhibitory activity is greatly enhanced when they are phosphorylated by PKA and other second messenger kinases. Thus, neurotransmitters that signal through cAMP can influence the phosphorylation of target proteins, both by PKA activation and through PKA-induced indirect inhibition of protein phosphatase 1.

Another protein phosphatase inhibitor, called dopamine- and cAMP-regulated phosphoprotein of 32 kDa (DARPP-32), is of particular interest because it is highly concentrated in regions of the brain that receive dopaminergic input. Similar to other kinase events, phosphorylation of DARPP-32 by PKA greatly enhances its ability to inhibit protein phosphatase 1. Interestingly, DARPP-32 is dephosphorylated by calcineurin, resulting in an increase in protein phosphatase 1 activity. Because calcineurin is activated by increases in intracellular Ca^{2+} concentrations, DARPP-32 may be involved in the integration of current signals from different pathways. Indeed, DARPP-32 is a key mediator of dopaminergic signaling and appears to play an important role in the effects of drugs of abuse. A share of the 2000 Nobel Prize for Physiology or Medicine was awarded to Paul Greengard for elucidating this role of DARPP-32.

TYROSINE KINASE PATHWAYS

While the vast majority of the protein phosphorylation that occurs in cells is on serine and threonine residues, the phosphorylation of tyrosine residues has an extremely important role in a distinct set of intracellular signaling pathways. In particular, tyrosine-phosphorylation–based signaling is mediated by receptors for neurotrophic factors such as NGF and BNDF. Functionally, neurotrophic factors modulate a wide variety of cellular events, such as cell growth and differentiation, synaptic plasticity, metabolism, and cell survival, and thus have been classically studied for their role in neurodevelopment. However, these factors have been shown to be expressed throughout the entire life span, and exciting new research is delineating the roles that neurotrophic factors play in regulating behavior and responses to stress.

Protein tyrosine kinases represent a diverse superfamily of proteins. These include the receptor tyrosine kinases, which are transmembrane receptors with a tyrosine kinase built into the intracellular domains, and nonreceptor tyrosine kinases, which are soluble cytoplasmic enzymes that are often recruited to membrane receptors to become activated. Neurotrophins such as NGF and BDNF bind to the Trk family of receptor tyrosine kinases that will be the primary focus of this section. Neurotrophins bind to two individual Trk receptors, resulting in the dimerization of the two receptors activating the protein tyrosine kinases that reside in the cytoplasmic domain of each receptor (Fig. 1.11–9). The activated Trk receptors subsequently phosphorylate the opposite dimer on tyrosine residues, a process called autophosphorylation. These phosphorylation events produce new binding sites for various other intracellular signaling proteins. For example, an adaptor protein called growth-factor-receptor–bound protein 2 (Grb2) contains an Src homology 2 (SH2) domain that binds to specific phosphorylated tyrosine residues on Trk and leads off a complex signaling cascade.

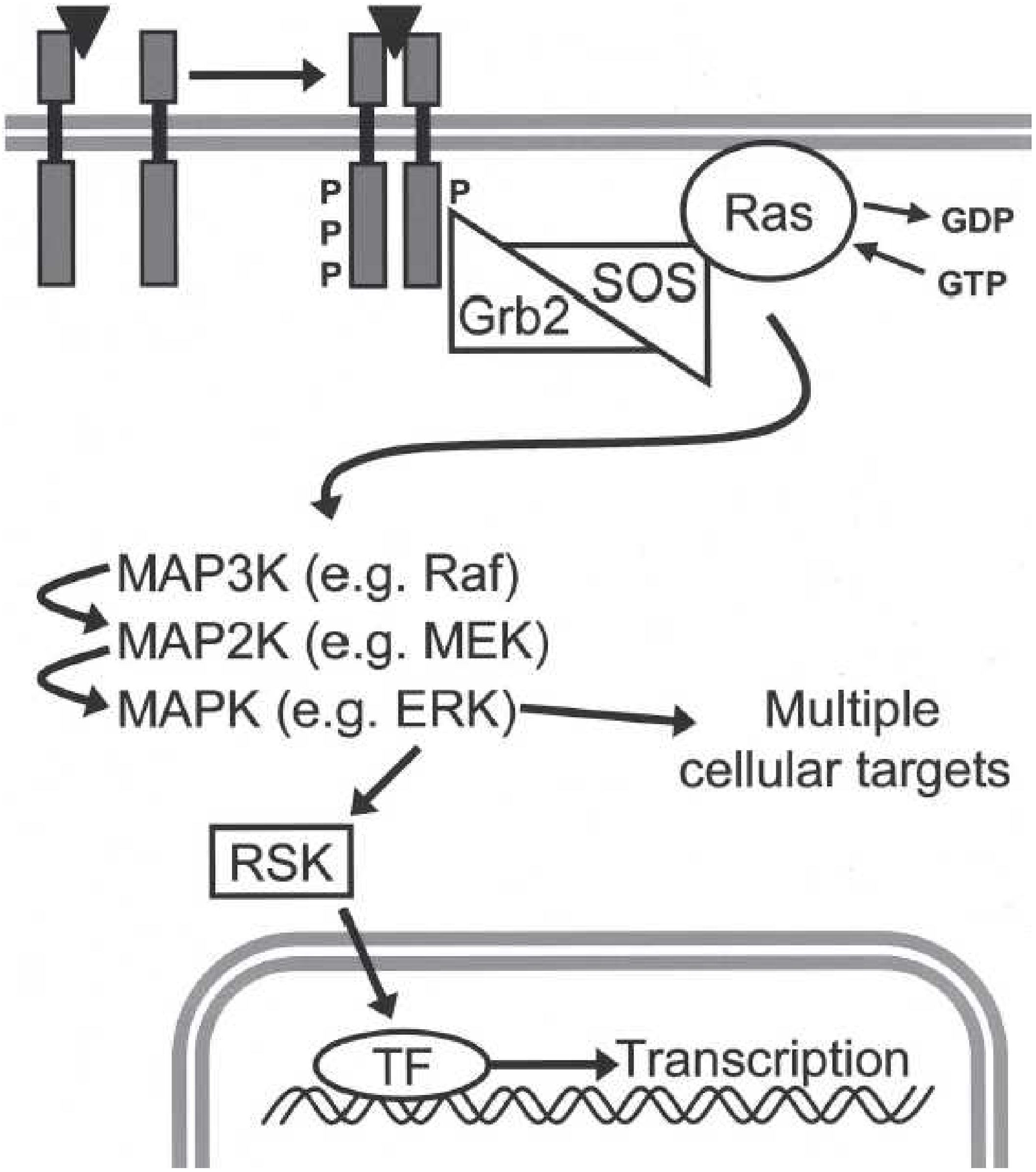

FIGURE 1.11–9. General organization of neurotrophic factor signaling through receptor tyrosine kinases. Receptor tyrosine kinases are transmembrane receptors with a tyrosine kinase built into the intracellular domains. Neurotrophic factor binding induces the dimerization of two receptors and the activation and autophosphorylation of their intrinsic tyrosine kinase domains. These phosphorylated (P) tyrosines become the binding sites for adaptor proteins such as growth-factor-receptor–bound protein 2 (Grb2), which can then attract a protein called Son of Sevenless (SOS) that activates the small G protein Ras by enhancing the exchange of GTP for GDP. In its active GTP-bound form, Ras activates multiple downstream effector pathways, including mitogen-activated protein kinase (MAPK) cascades. A MAPK such as extracellular signal-regulated kinase (ERK) is activated by a MAPK kinase (MAP2K) such as the MAPK/ERK kinase (MEK), which is activated by a MAPK kinase kinase (MAP3K) such as Raf. After the cascade, ERK can

activate various cellular targets including p90 ribosomal S6 kinases (RSK), which translocates to the nucleus and activates various transcription factors (TFs) and regulates gene expression.

As with neurotransmitter receptors, G proteins play a major role in the signal transduction from activated receptor tyrosine kinases. In this case, however, the G proteins are members of the Ras, Rho, and Ral families, collectively referred to as small G proteins. Like the heterotrimeric G proteins described previously, small G proteins are bound to GDP in their inactive state and become active when GTP is bound. However, unlike the heterotrimeric G proteins, the small G proteins are not directly activated by the receptor but by distinct proteins called guanine nucleotide exchange factors (GEFs). For Trk, binding of Grb2 to the phosphorylated tyrosine residues recruits a GEF protein called SOS (for Son of Sevenless), which then activates the small G protein Ras by exchanging GTP for GDP. Once recruited to the receptor complex, SOS can activate many molecules of Ras, thus amplifying the initial signal. In its active GTP-bound form, Ras is able to activate the multiple downstream effector pathways described below. Ras is inactivated by hydrolysis of GTP to GDP, which can be accelerated by a specific GTPase-activating protein (GAP) called RasGAP, which is analogous to the action of RGS proteins on heterotrimeric G proteins.

Mitogen-Activated Protein Kinase Cascades

In contrast to the GPCR-mediated second messenger signaling pathways, the activation of small G proteins such as Ras by receptor tyrosine kinases does not lead to the production of small-molecule intermediates. Instead, the small G proteins stimulate a signaling pathway that is organized as a kinase cascade in which a series of three or more kinases sequentially phosphorylate another kinase. Several parallel kinase cascades can be activated by various receptors and are collectively referred to as MAPK pathways. MAPKs are serine/threonine kinases making up three main classes characterized in mammals: The extracellular signal-regulated kinases (ERKs), the c-Jun N-terminal kinases (JNKs), and isoforms of p38. The ERK pathway is the classical MAPK pathway that is preferentially activated by the neurotrophins and other growth factors, while the JNK and p38 pathways are activated by other receptors, signaling pathways, and various forms of cellular stress.

The kinase cascades that lead to MAPK activation follow an organization that is evolutionally well conserved from yeast to mammals: A MAPK kinase kinase kinase (MAP4K) phosphorylates a MAPK kinase kinase (MAP3K) that phosphorylates a MAPK kinase (MAP2K) that then phosphorylates the MAPK (see Fig. 1.11–9). The cascade of events leading to the activation of ERKs by neurotrophins through the Trk receptor begins with the activation of the small G protein Ras. When Ras is active, it recruits a MAP3K called Raf to the cell surface, where it is phosphorylated by a MAP4K that is not yet well described. Raf then phosphorylates and activates a MAP2K called MEK (for MAP kinase/ERK kinase) that then phosphorylates and activates ERK.

The ERK pathway is the subject of a considerable amount of current biomedical research, as it is involved in regulating a wide variety of cytoplasmic proteins as well as multiple transcription factors. For example, ERK phosphorylates and activates protein kinases such as the p90 ribosomal S6 kinases (RSK), which in turn phosphorylates an array of transcription factors, including c-myc and CREB. Interestingly, stimulation of the ERK signaling pathway has also been linked to neurotransmitter receptors through PKC. Thus, ERK activation plays a key role in modulating long-term neuronal function and may represent a critical node for the interplay between other signaling pathways.

Phosphoinositide 3-Kinase Pathway

The elucidation of the signaling pathways downstream of Ras has identified another major kinase cascade that mediates many of the powerful effects of neurotrophins on neuronal differentiation and survival. This cascade involves the phosphoinositide 3-kinase (PI3K) pathway (Fig. 1.11–10). In this pathway, PIP_2, the same membrane phospholipid that is cleaved to DAG and IP_3 by PLC, is phosphorylated by PI3K, a lipid kinase, to yield PIP_3, which then acts to recruit various proteins to the membrane. One of the proteins that PIP_3 recruits to the membrane is Akt, a serine/threonine kinase that, upon translocation, becomes activated, dissociates from the membrane, and phosphorylates several substrate proteins important for controlling cell survival. For example, Akt activates the “rapid-acting” transcription factor nuclear factor-κB (NF-κB), resulting in the transcription of prosurvival genes. NF-κB is present in cells in an inactive state bound to a protein called inhibitor of κB (IκB). Akt activates a kinase called IκB kinase that phosphorylates IκB, thus tagging it for degradation. This degradation releases NF-κB, which can then migrate to the nucleus and regulate gene expression. Akt also inhibits glycogen synthase kinase 3 (GSK-3), a metabolic regulatory protein that may be a cellular target for lithium (see below). Indeed, a major challenge of current research involves determining which of the many effects of neurotrophins are mediated by these various signaling cascades. Interestingly, some G_i-coupled neurotransmitter receptors can also trigger the activation of PI3K and Akt, suggesting that agonists of these receptors may represent novel strategies for enhancing neuronal survival.

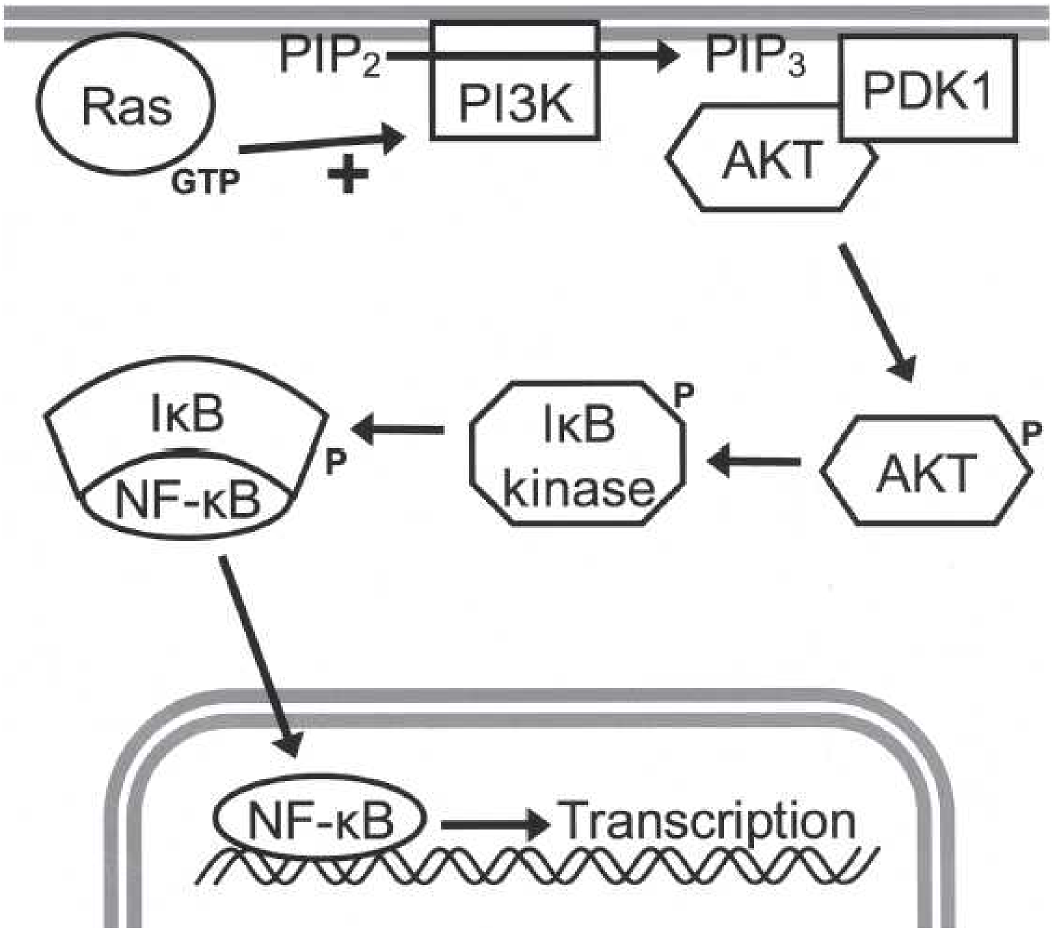

FIGURE 1.11–10. The phosphoinositide 3-kinase (PI3K) pathway. In addition to the mitogen-activated protein kinases (MAPKs), receptor tyrosine kinase activation of Ras can activate the PI3K pathway. PI3K adds another phosphate to the membrane phospholipid phosphatidylinositol bisphosphate (PIP2) to yield phosphatidylinositol trisphosphate (PIP3). PIP3 recruits various proteins to the membrane including 3-phosphoinositide–dependent protein kinase 1 (PDK1) and a kinase called Akt. PDK1 phosphorylates Akt, which then dissociates from the membrane and can phosphorylate multiple cellular proteins important for controlling cell survival. For example, Akt can lead to the activation of a transcription factor called nuclear factor-κB (NF-κB), which is normally found in an inactive state bound to "inhibitor of κB" (IκB). Akt activates a kinase called IκB kinase, which phosphorylates IκB, tagging it for degradation, which releases NF-κB that can then translocate to the nucleus and regulate gene expression.

Mammalian Target of Rapamycin

A downstream target of the PI3K and a central node of integration of multiple extracellular signals is an atypical kinase called mammalian target of rapamycin (mTOR). mTOR is a central controller of local protein synthesis that is required for the formation of new synaptic connections. mTOR function is influenced by the activity of each of the major classes of receptors, including ligand-gated ion channels, GPCRs, and neurotrophin receptors, suggesting that mTOR may act as a signaling “node” at the convergence of several signaling pathways. Indeed, mTOR activity is regulated by multiple signaling pathways, notably the PI3K-Akt pathways.

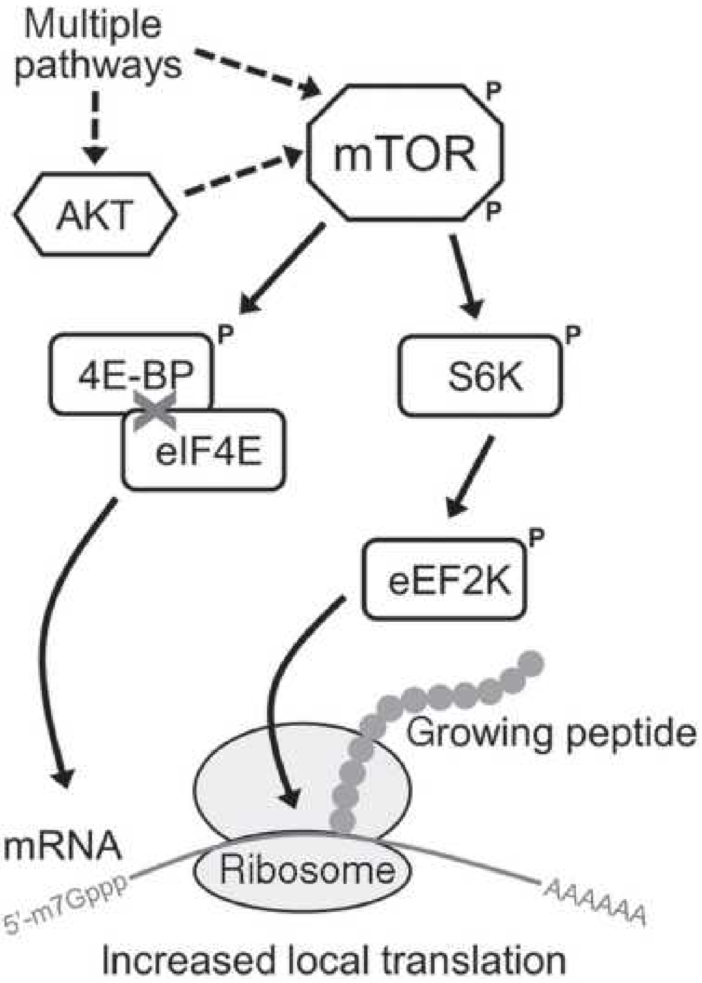

FIGURE 1.11–11. Regulation of protein translation by the mammalian target of rapamycin (mTOR). Research on mTOR regulation has suggested that mTOR is a key point of convergence of multiple signaling pathways, notably via Akt. mTOR is a central controller of local protein synthesis by regulating proteins involved in protein translation initiation and elongation at ribosomes. The activated mTOR complex phosphorylates 4E-BP (eukaryotic initiation factor 4E-binding protein) which dissociates it with eIF4E (eukaryotic initiation factor 4E). eIF4E can then facilitate translation initiation by binding mRNA at the 7-methyl-guanosine cap (5’-m7Gppp). mTOR also facilitates the elongation of nascent polypeptides by phosphorylating p70 ribosomal S6 kinases (S6K) which can then activate the eukaryotic elongation factor 2 kinase (eEF2K).

Protein synthesis is a highly regulated process that can be separated into three general phases: initiation, elongation, and termination. mTOR is a master controller of protein synthesis by regulating both the initiation and elongation steps of protein synthesis (Fig. 1.11–11). One of the first mTOR targets to be discovered was 4E-BP (eukaryotic initiation factor 4E-binding protein) that, when phosphorylated by mTOR, releases eIF4E (eukaryotic initiation factor 4E) which can then initiate local protein translation by binding to the 7-methyl-guanosine cap of mRNAs. In addition, mTOR activates the p70 ribosomal S6 kinases (S6K) which then activates eukaryotic elongation factor 2 kinase (eEF2K), facilitating protein synthesis.

Recently, mTOR signaling has been linked to long-lasting synaptic plasticity and various neuropsychiatric disorders. For example, the rapid antidepressant activity of NMDA receptor antagonists such as ketamine may be mediated by the activation of the mTOR leading to increased translation of synaptic proteins and generation of new spine, particularly in the prefrontal cortex. In addition, genetic mutations that lead to pathologically sustained mTOR activity may underlie several neuropsychiatric disorders, including fragile X syndrome, autism, and tuberous sclerosis.

WNT SIGNALING

Another signaling pathway gaining interest in psychiatry and neurobiology is the Wnt signaling pathway. Wnts are a family of secreted glycoproteins known to play a critical role in embryogenesis. However, components of the Wnt signaling pathway are expressed in the adult brain, and Wnt signaling is important in adult behavior and possibly the pathophysiology of psychiatric and neurologic disorders. The primary, or canonical, Wnt signaling pathway begins with the binding of secreted Wnt proteins to cell-surface receptors of the Frizzled family. Frizzled receptors are seven-transmembrane domain receptors similar to the GPCRs, though it remains unclear whether Frizzled signals through heterotrimeric G proteins. It is clear, however, that Frizzled receptors activate a cytoplasmic protein called Dishevelled, that ultimately leads to the regulation of gene expression through an increase in a transcriptional coactivator called β-catenin.

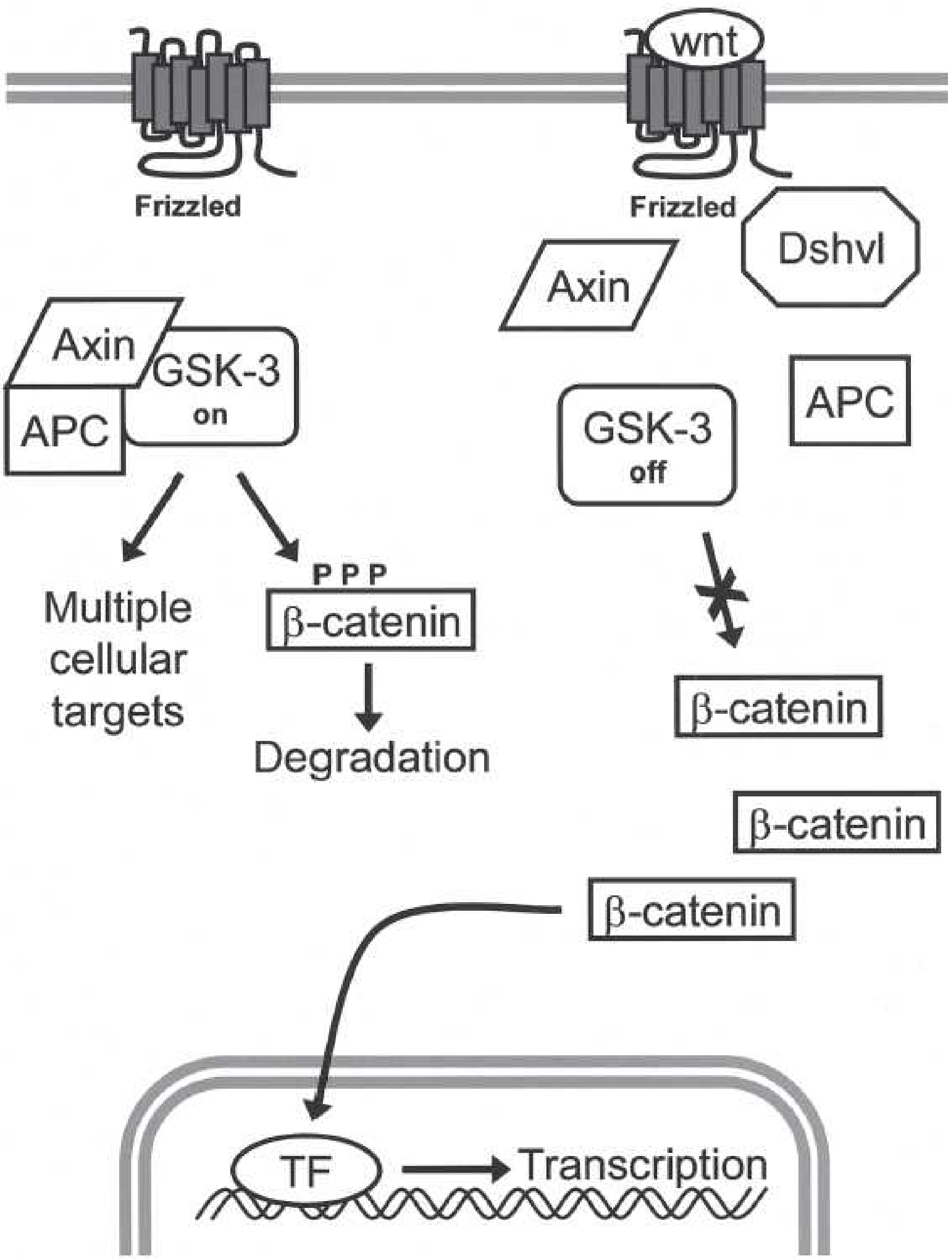
wnt
Frizzled
Frizzled
Dshvl
Axin
Axin
GSK-3
on
APC
APC
GSK-3
off
Multiple
cellular
targets
P P P
β-catenin
Degradation
β-catenin
β-catenin
β-catenin
TF
Transcription

FIGURE 1.11–12. Basic organization of the canonical Wnt signaling pathway. The primary, or canonical, Wnt signaling pathway begins with the binding of secreted Wnt proteins to cell-surface receptors of the Frizzled family. In the absence of Wnt signaling, a complex of proteins that includes axin, adenomatosis polyposis coli (APC) protein, and glycogen synthase kinase 3 (GSK-3) maintains an active form of GSK-3 (on), which phosphorylates multiple cellular proteins, including the transcriptional coactivator β-catenin. When β-catenin is phosphorylated, it is targeted for degradation. When Wnt

signaling is initiated, Frizzled activates a protein called Dishevelled (Dshvl), which causes the dissociation of the axin/APC/GSK-3 complex, resulting in an inhibition of GSK-3 activity. Decreased GSK-3 activity reduces the degradation of β-catenin, which then translocates to the nucleus, interacts with transcription factors (TFs), and promotes specific gene expression changes.

In the absence of Wnt signaling, a complex of proteins that includes axin, GSK-3, and the protein adenomatous polyposis coli (APC) regulates the intracellular levels of β-catenin (Fig. 1.11–12). Through phosphorylation by GSK-3, this protein complex promotes the proteolytic degradation of β-catenin. However, when Wnt signaling is initiated, the activation of Dishevelled causes this protein complex to dissociate and other proteins to associate and inhibit GSK-3 activity, preventing the degradation of β-catenin. Thus, the cytoplasmic levels of β-catenin increase and β-catenin translocate to the nucleus, interacting with transcription factors and promoting specific gene expression changes. In addition to the canonical signaling pathway, Wnt signaling has also been shown to follow other pathways, including increasing intracellular calcium and activating the MAPK JNK. Thus, while the details of the Wnt signaling pathway are not fully delineated, there is likely some intersection and cross-regulation with other signaling pathways, and this is an area of active research.

Glycogen Synthase Kinase 3

Though initially discovered as a kinase involved in the regulation of glucose metabolism, GSK-3 is emerging as a promising target for the development of psychiatric and neurologic medications. In 1996, it was discovered that lithium inhibited GSK-3, raising the possibility that GSK-3 inhibition might play a role in the treatment of bipolar disorder. Recently, there has been an emergence of research supporting the hypothesis that the inhibition of GSK-3 represents a therapeutically relevant target for mood stabilization.

GSK-3 is a ubiquitous kinase found in both neurons and glia and has two isoforms that are highly homologous but may have slightly different biologic effects. It is generally considered to be constitutively active, meaning that it phosphorylates target proteins until a signal regulates it to stop. For example, as described above, constitutive phosphorylation of β-catenin by GSK-3 leads to its proteolytic degradation but signaling through the Wnt pathway turns off GSK-3, releasing β-catenin to affect gene expression. GSK-3 was initially characterized in 1980 as an enzyme that phosphorylated and deactivated glycogen synthase, leading to studies of its role in insulin signaling and diabetes mellitus. Indeed, insulin binding to the insulin receptor, a receptor tyrosine kinase, leads to the activation of Akt, which also phosphorylates and deactivates GSK-3. Because GSK-3 normally phosphorylates and inactivates glycogen synthase, insulin's ability to turn off GSK-3 allows cells to utilize the elevated plasma glucose levels to make glycogen.

As discussed above, Akt is also activated by neurotrophic factors such as BDNF, low levels of which are implicated in depression and other neuropsychiatric disorders. Other kinases that regulate GSK-3 include PKA, PKC, and RSK, demonstrating that the mechanisms of regulation and biologic targets of GSK-3 are quite diverse. This convergence of diverse signaling pathways onto GSK-3 is a characteristic that has led to the labeling of GSK-3 as a crucial signaling "node." The precise mechanisms that regulate the cross-talk among these distinct pathways are not well established, and this is an area of active research. However, it is likely that the compartmentalization of GSK-3 to distinct regions of the cell minimizes much of the potential cross-talk among pathways.

SYNAPTIC PLASTICITY

Changes in the strength and efficiency of synaptic signaling, termed synaptic plasticity, underlie one of the most important neurochemical foundations of learning and memory. Because these processes play a prominent role in a variety of psychiatric disorders and psychotherapies, there has been intense interest in defining the cellular and molecular events mediating these processes. Because synaptic plasticity is activity dependent, various intraneuronal signaling pathways are important for coordinating these changes. Indeed, there are several mechanisms that cooperatively affect synaptic plasticity, including changes in the release of presynaptic neurotransmitters and changes in how effectively the postsynaptic neuron responds to those neurotransmitters. A postsynaptic mechanism that is widely considered to be a major mediator for enhancing synaptic efficacy is called long-term potentiation (LTP).

LTP is roughly defined as an increase in the strength of a synapse that lasts from hours to several days and is widely considered one of the major mechanisms by which memories are formed and stored in the brain.

Given the diversity of neuronal cell types in the brain, there are many variations in the processes involved in LTP; however, the prototypical model of LTP has been derived from the glutamatergic synapses in the CA1 region of the hippocampus. At these synapses, there are both early and late stages of LTP that are initiated by the actions of two glutamate-gated ion channels α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) and NMDA receptors. Early LTP is mediated by modulating proteins already at or near the synapse, while late LTP requires new protein synthesis.

The initiation of early LTP begins with glutamate binding to AMPA receptors, which allows Na^+ to enter the synapse, depolarizing the membrane (Fig. 1.11–13). When the postsynaptic membrane is sufficiently depolarized, NMDA receptors open, leading to a rapid increase in intracellular Ca^{2+} concentrations. The rise in Ca^{2+} levels leads to the activation of CaM kinase II (CaMKII) and PKC, which rapidly phosphorylate AMPA receptors, which increases their conductance for Na^+, enhancing the strength of the postsynaptic response. In addition, CaMKII activation facilitates the insertion of additional AMPA receptors into the postsynaptic membrane. By increasing the number of AMPA receptors at the synapse, future signaling stimuli are able to generate larger postsynaptic responses. This process can additionally be regulated by the concomitant activation of other signaling pathways through various GPCRs.

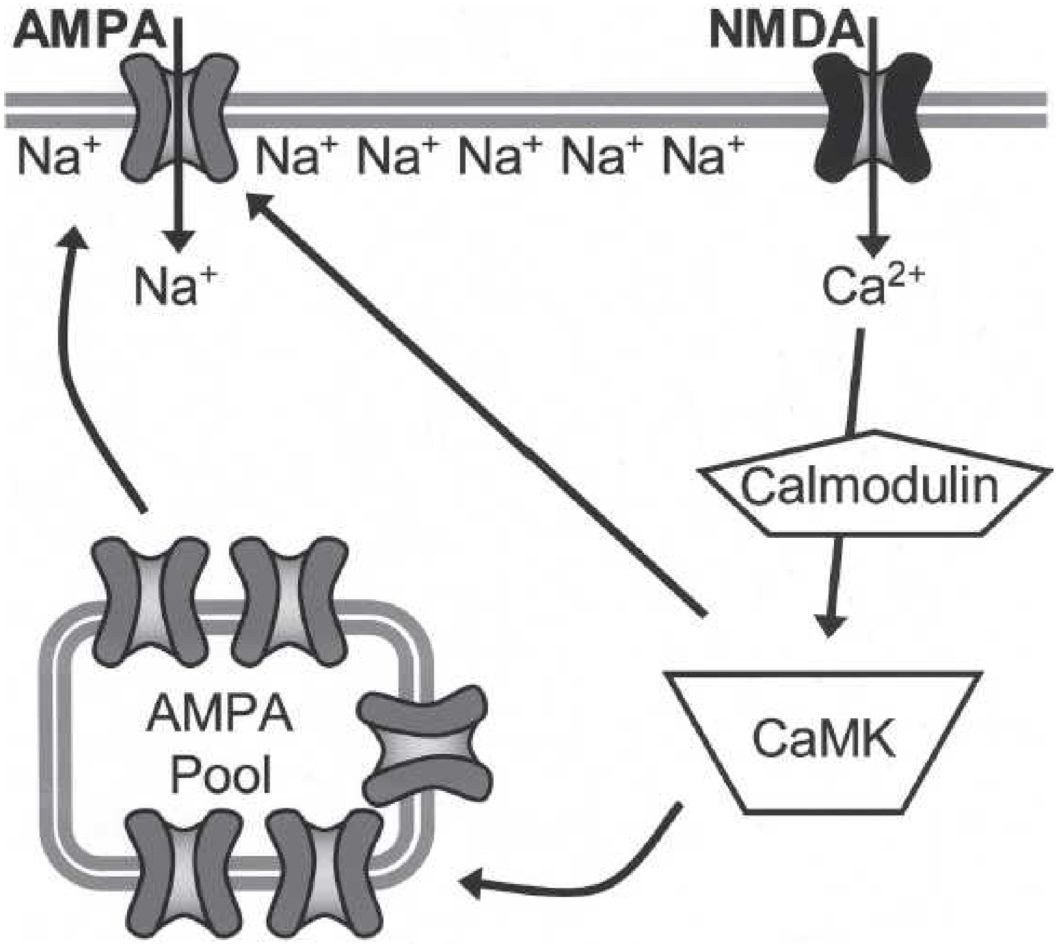

FIGURE 1.11–13. An example schematic of a postsynaptic mechanism involved in long-term potentiation (LTP) and synaptic plasticity. Initiation of LTP begins with glutamate binding to α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptors gating the influx of Na^+ ions into the synapse, depolarizing the membrane. When the postsynaptic membrane is sufficiently depolarized, N-methyl-d-aspartic acid (NMDA) receptors open, leading to a rapid increase in intracellular Ca^{2+} levels, which, through binding to calmodulin, activate Ca^{2+}/calmodulindependent kinases (CaMKs) and protein kinase C (PKC). These activated protein kinases can then both phosphorylate AMPA receptors and regulate the insertion of additional AMPA receptors into the postsynaptic membrane intracellular pool. Together, these processes increase the efficiency of future synaptic transmission.

Although these phosphorylation events underlie the rapid changes in synaptic efficacy during early LTP, enduring changes characteristic of “late LTP” depend on the targeting of newly synthesized proteins to the synapses. These newly targeted proteins, which can include additional receptors and scaffolding proteins, induce a remodeling of the synapse and can profoundly strengthen postsynaptic responses to stimuli. At least some of this new protein synthesis occurs locally in the dendrites, allowing direct targeting of the new proteins to the activated synapse, a process largely regulated by mTOR activity.

Overall, LTP and the other cellular mechanisms involved in synaptic plasticity are highly active areas of neuroscience research. The current understanding of these processes highlights the complexity and importance of tight temporal and spatial regulation of synaptic signaling. Continued elucidation of these biochemical mechanisms underlying synaptic plasticity will enhance the understanding of learning and behavior, provide insights into psychiatric diseases, and may allow the development of pharmacologic agents that can improve cognition.

SIGNALING COMPLEXES AND SCAFFOLDS

Several additional types of proteins are central to the organization of signaling pathways. These include scaffolding and anchoring proteins, which provide mechanisms to ensure that information being signaled in cells is transferred to the appropriate targets in a timely and efficient manner. They do this by mediating the localized assembly of multiprotein complexes that contain, for example, receptors, second messenger–generating enzymes, kinases, phosphatases, and substrates. By keeping many of the components of a signaling cascade in close proximity, these complexes minimize the need for activated proteins to diffuse through a dense cytoplasm to find their targets, thus greatly enhancing signaling efficiency. In addition, these systems both maintain the separation, or compartmentalization, of distinct signals when simultaneous signaling events are occurring and are crucially involved in moderating and integrating this information. The characterization of the multitude of ways that signaling pathways interact via these complexes is an active area of research.

Many of these scaffolding and adaptor proteins use specific protein–protein interactions to mediate the transport and localization of signaling proteins and form these specialized multiprotein complexes. Protein interactions are often formed by distinct domains within adaptor proteins that are responsible for recognizing and binding to specific regions of other proteins. An example mentioned above is the Src homology (SH) domain involved heavily in neurotrophin signaling. SH2 domains, found in adaptor proteins such as Grb2, are roughly 100 amino acids long and specifically bind to short amino acid sequences that contain a phosphorylated tyrosine residue. Multiple proteins can bind to these phosphotyrosine sequences via their SH2 domains, some of which are subsequently phosphorylated and activated, and others act as adaptors that recruit other substrates to the kinase. Indeed, activated receptor tyrosine kinases, such as Trk, serve as scaffolds for an array of activated signaling molecules. Another important example is the localization of PKA to distinct sites within cells by a family of scaffolding proteins called A-kinase–anchoring proteins (AKAPs). These AKAPs bind to the regulatory subunits of PKA, holding it near particular substrates while awaiting activation by cAMP. Some AKAPs, for example, are thought to localize PKA next to synaptic ion channels, greatly enhancing the rate of substrate phosphorylation by eliminating delays due to protein diffusion.

Scaffolding also organizes signaling complexes involving the classic neurotransmitter receptors and ion channels at synapses. A common protein domain involved in protein scaffolding in these systems is the PDZ domain. PDZ domains are found in more than 400 proteins in humans and bind tightly to the extreme C-terminal segment of proteins in which the last three amino acids are commonly S/TXV (i.e., serine [S] or threonine [T], followed by any amino acid [X], followed by valine [V] or another hydrophobic amino acid), though other variations exist. Given the large number of proteins containing PDZ domains and PDZ recognition elements, scaffolds containing these proteins can assemble into very large molecular complexes. The best-known example of these large scaffolds is the postsynaptic density (PSD) of excitatory neurons, which organizes glutamate receptors and their associated signaling proteins at the postsynaptic membrane and helps to determine the size and strength of synapses. Indeed, scores of proteins have been identified in the PSD: ion channels, GPCRs, kinases, phosphatases, and cytoskeletal proteins are all targeted and maintained in the PSD by various adaptor and scaffolding proteins. Thus, scaffolding proteins are major players in the organization, efficiency, and specificity of intraneuronal signaling machinery.

Functional Selectivity

Even though most cell-surface receptors have classically been described as activating a single primary intracellular signaling cascade, as discussed above, most receptors also activate one or more additional pathways.

Some GPCRs, for example, have been shown to signal through the phosphatidylinositol pathway, the arachidonic acid pathway, and the MAPK/ERK pathway, among others. In addition, activation of these receptors stimulates the biochemical mechanisms involved in their desensitization and internalization. A classical concept of receptor pharmacology is that a receptor ligand can be either classified as a full agonist, partial agonist, or antagonist at that receptor and that this classification will be consistent for all of the signaling pathways for that receptor. In other words, if a receptor-ligand fully activates the phosphatidylinositol pathway, then it is expected to fully activate all of the other signaling and regulatory pathways linked to that receptor. However, there is now evidence that some ligands may inherently be able to produce different levels of signaling among the various pathways (Fig. 1.11–14). This phenomenon is most often referred to as "functional selectivity" (also called biased agonism).

A well-recognized form of functional selectivity is the differential ability of ligands to activate the classical G-protein–mediated signaling pathway compared to the ability to recruit arrestins for receptor desensitization and endocytosis. Arrestins also act as a scaffold for additional signaling pathways, particularly MAPK/ERK signaling. An example of this form of biased agonism lies with the μ-opioid receptor (MOR), which is a target for endogenous enkephalin peptides as well as opioid analgesics such as morphine. Interestingly, mice with a targeted deletion of a particular arrestin isoform develop less tolerance to the analgesic effects of morphine. Thus, significant efforts are underway to develop new opioids with a strong bias toward the analgesia-mediating G-protein signaling over arrestin signaling. With reduced tolerance, there may be less risk of respiratory depression and overdose due to increased doses. Importantly, however, these arrestin knockout mice still develop morphine dependence. Thus, the overall benefits of biased opioid agonists remain unknown. Indeed, truly balanced opioids similar to endogenous enkephalins may be more beneficial in limiting opioid dependence.

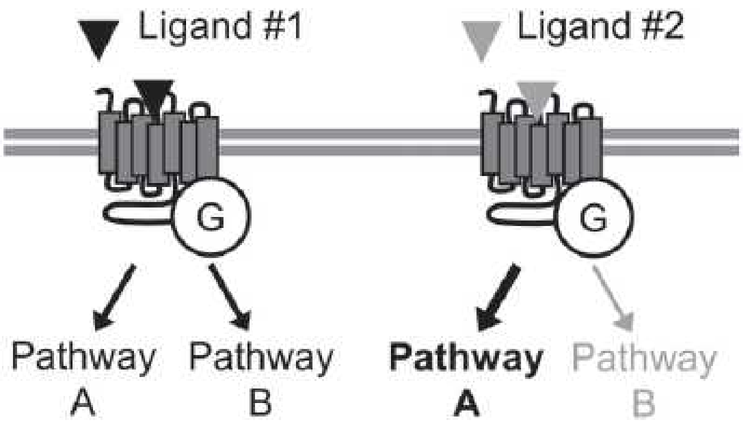

FIGURE 1.11–14. Functional selectivity. In addition to the canonical G-protein–mediated signaling pathways, most GPCRs can activate other signaling pathways. While classical receptor pharmacology theory posits that a full agonist at a receptor will activate all signaling pathways equally, it is now clear that some ligands may inherently be able to produce different levels of signaling from the various pathways. This phenomenon is referred to as "functional selectivity."

While the example above focuses on opioid receptors, functional selectivity has been demonstrated in many GPCRs and may be mediated by a variety of mechanisms. First, different ligands sample and stabilize unique conformational changes in the receptor protein, resulting in differential activation of the various signaling pathways. Second, functional selectivity may also arise through the diversity of G proteins and other signaling and scaffolding proteins or the observed ability of GPCRs to dimerize and oligomerize, the function of which remains poorly understood. Third, as discussed above, drugs can differentially interact with receptors on intracellular membranes suggesting that intracellular bioavailability can lead to a location-biased form of functional selectivity. Thus, not only is the concept of functional selectivity an interesting one, but it is likely to have an important impact on future psychiatric drug development.

REMOTE CONTROL OF INTRANEURONAL SIGNALING

In recent years, emerging synthetic biology technologies have allowed for the targeted activation of intracellular signaling pathways with spatial and temporal precision. These techniques, collectively termed chemogenetics, have facilitated the study of complex circuits and behaviors and are revolutionizing the understanding of brain function. In one approach, chimeric receptors were engineered that replaced the intracellular loops of the light-activated rhodopsin with those of specific adrenergic or serotonergic receptors. These artificial receptors, called “opto-XRs,” allow the use of light to activate the signaling pathways normally recruited by the receptor used for the intracellular loops. Because of the speed of optical manipulation, this approach allows for rapid stimulation of intracellular signaling and examination of the effects of tonic versus phasic activation, though it requires the invasive insertion of optical fibers into the brain. Another approach uses specifically engineered GPCRs, or chemogenetic actuators, that are only activated by synthetic ligands that do not activate endogenous receptors. Termed DREADDs (designer receptors exclusively activated by designer drugs) or RASSLs (receptors activated solely by synthetic ligands), these technologies similarly allow for precise spatiotemporal control of intracellular signaling in vivo. Various versions of these chemogenetic actuators activate the major GPCR signaling pathways (G_s, G_i, and G_q). Additional chemogenetic actuators are being developed that could control intraneuronal signaling cascades, including engineered enzymes and drug-induced dimerization or destabilization domains. Since they are activated by drugs, chemogenetic actuators allow the noninvasive modulation of intracellular signaling and thus have the potential to be used for transformative therapeutic approaches for a host of neuropsychiatric disorders.

FUTURE DIRECTIONS

Translating the advances in molecular neurobiology into improved diagnostic and therapeutic capabilities represents the greatest opportunity and challenge facing modern psychiatry. The current armamentarium of medications used in treating psychiatric diseases has facilitated decades of progress in understanding intercellular signaling mediated by cell-surface receptors. However, dramatic and ongoing advances in the understanding of the intraneuronal signaling pathways activated by these receptors will likely lead to novel, innovative, and improved pharmacologic agents for psychiatric diseases, as has been achieved in other branches of medicine. Thus, future efforts in drug discovery should move beyond the current strategies of solely targeting synaptic neurotransmission at the receptor level to the development of agents acting on components of intracellular signaling pathways. By nature, signaling pathways have significant redundancy and interactions. Thus, identifying and targeting critical points within these networks may lead to improved molecular diagnostic tests and treatments.

FURTHER READINGS

Airan RD, Thompson KR, Fenno LE, Bernstein H, Deisseroth K. Temporally precise in vivo control of intracellular signalling. *Nature*. 2009;458:1025–1029.

Alberts B, Heald R, Johnson AD, Morgan D, Raff M, Roberts K, Walter P, Wilson J, Hunt T. *Molecular Biology of the Cell*. 7th ed. W.W. Norton & Company; 2022.

*Blitzer RD, Iyengar R, Landau EM. Postsynaptic signaling networks: cellular cogwheels underlying long-term plasticity. *Biol Psychiatry*. 2005;57:113–119.

Boeckers TM. The postsynaptic density. *Cell Tissue Res*. 2006;326:409–422.

Bohn LM, Gainetdinov RR, Lin FT, Lefkowitz RJ, Caron MG. μ-Opioid receptor desensitization by β-arrestin-2 determines morphine tolerance but not dependence. *Nature*. 2000;408:720–723.

Cheyette BN, Moon RT. Wnt protein family. In: Henry HL, Norman AW, eds. *Encyclopedia of Hormones*. Academic Press; 2003:655–674.

Doupnik CA. GPCR-Kir channel signaling complexes: defining rules of engagement. *J Recept Signal Transduct Res*. 2008;28:83–91.

Duman RS, Voleti B. Signaling pathways underlying the pathophysiology and treatment of depression: novel mechanisms for rapid-acting agents. *Trends Neurosci*. 2012;35:47–56.

Duncan AL, Song W, Sansom MSP. Lipid-dependent regulation of ion channels and G protein-coupled receptors: insights from structures and simulations. *Annu Rev Pharmacol Toxicol*. 2020;60:31–50.

Gainetdinov RR, Premont RT, Bohn LM, Lefkowitz RJ, Caron MG. Desensitization of G protein-coupled receptors and neuronal functions. *Annu Rev Neurosci*. 2004;27:107–144.

Gillis A, Kliewer A, Kelly E, et al. Critical assessment of G protein-biased agonism at the μ-opioid receptor. *Trends Pharmacol Sci*. 2020;41:947–959.

Hanyaloglu AC, von Zastrow M. Regulation of GPCRs by endocytic membrane trafficking and its potential implications. *Annu Rev Pharmacol Toxicol*. 2008;48:537–568.

Herring BE, Nicoll RA. Long-term potentiation: from CaMKII to AMPA receptor trafficking. *Annu Rev Physiol*. 2016;78:351–365.

Irannejad R, Pessino V, Mika D, et al. Functional selectivity of GPCR-directed drug action through location bias. *Nat Chem Biol*. 2017;13:799–806.

Kroeze WK, Sheffler DJ, Roth BL. G-protein-coupled receptors at a glance. *J Cell Sci*. 2003;116:4867–4869.

Le Novère N, Li L, Girault JA. DARPP-32: molecular integration of phosphorylation potential. *Cell Mol Life Sci*. 2008;65:2125–2127.

Müller N, Schwarz MJ. COX-2 inhibition in schizophrenia and major depression. *Curr Pharm Des*. 2008;14:1452–1465.

*Nestler EJ, Hyman SE, Malenka RC. *Molecular Neuropharmacology: A Foundation for Clinical Neuroscience*. 4th ed. McGraw-Hill Education; 2020.

*Patapoutian A, Reichardt LF. Trk receptors: mediators of neurotrophin action. *Curr Opin Neurobiol*. 2001;11:272–280.

Peineau S, Bradley C, Taghibiglou C, et al. The role of GSK-3 in synaptic plasticity. *Br J Pharmacol*. 2008;153(Suppl 1):S428–S437.

*Sternson SM, Roth BL. Chemogenetic tools to interrogate brain functions. *Annu Rev Neurosci*. 2014;37:387–407.

*Urban JD, Clarke WP, von Zastrow M, et al. Functional selectivity and classical concepts of quantitative pharmacology. *J Pharmacol Exp Ther*. 2007;320:1–13.

Wild AR, Dell'Acqua ML. Potential for therapeutic targeting of AKAP signaling complexes in nervous system disorders. *Pharmacol Ther*. 2018;185:99–121.

Willars GB. Mammalian RGS proteins: multifunctional regulators of cellular signalling. *Semin Cell Dev Biol*. 2006;17:363–376.

Wootten D, Christopoulos A, Marti-Solano M, Babu MM, Sexton PM. Mechanisms of signalling and biased agonism in G protein-coupled receptors. *Nat Rev Mol Cell Biol*. 2018;19:638–653.

Yaffe M, Cantley LC. Grabbing phosphoproteins. *Nature*. 1999;402:30–31.

1.12 Psychoneuroendocrinology

VICTOR I. REUS, M.D.

Endocrine disorders are frequently associated with secondary psychiatric symptoms, such as depressed mood and disturbances in thought. In addition, certain psychiatric syndromes are associated with distinct patterns of endocrine dysfunction. The term *psychoneuroendocrinology* encompasses the inextricable structural and functional relationships between hormonal systems and the CNS and behaviors that modulate and are derived from both. Classically, *hormones* have been defined as the products of endocrine glands transported by the blood to exert their action at sites distant from their release. Advances in neuroscience have shown, however, that in the CNS, the brain not only serves as a target site for regulatory control of hormonal release but also has secretory functions of its own and serves as an end organ for some hormonal actions. These complex interrelationships make classic distinctions between the origin, structure, and function of neurons and those of endocrine cells dependent on physiologic context. Multiple interactions between the endocrine and immune systems have pointed to a parallel regulatory complexity.

HORMONE EVOLUTION

Over the course of evolution, as organisms have increased in complexity, hormones that first appeared in unicellular organisms have been recruited to serve a multiplicity of functions, a quality that is referred to as *pleiotropy.* A single hormone may act at multiple sites, including binding to receptors on the membrane, cytoplasm, or nucleus, each with different effects, and subtle differences in molecular structure or metabolic processing can have profound physiologic consequences. Hormones are thus ideally suited to regulate complex behavioral activities and to play a role in the *plasticity* of the organism, allowing it to adapt to the changing demands of the environment, as, for example, in the alteration of sexual phenotype in certain amphibians and reptiles in response to changing environmental conditions. Exogenous exposure to endocrine-disrupting chemicals in the environment, such as bisphenol A and phthalates, are additional confounding factors.

HORMONE CLASSIFICATION

Hormones are divided into two general classes by structure—(1) proteins, polypeptides, and glycoproteins; and (2) steroids and steroid-like compounds—and into three classes by location of function (Tables 1.12–1 and 1.12–2). In addition to classical action on a target tissue, hormones may also act as *neuromodulators,* regulating the effects of neurotransmitters and, in some cases, meeting criteria for neurotransmitter function independently.

Table 1.12–1.
Classification of Hormones by Structure

Structure	Examples	Storage	Lipid Soluble
Proteins, polypeptides, and glycoproteins	Adrenocorticotropic hormone, β-endorphin, thyrotropin-releasing hormone, luteinizing hormone, follicle-stimulating hormone	Vesicles	No
Steroids and steroid-like compounds	Cortisol, estrogens, testosterone, progesterone, dehydroepiandrosterone	Diffusion after synthesis	Yes

Table 1.12–2.
Classification of Hormones by Location of Function

Hormone Classification	Function
Autocrine	Self-regulatory effects
Paracrine	Local or adjacent cellular action
Endocrine	Distant target site

HORMONE SECRETION

Hormone secretion is stimulated by the action of a neuronal secretory product of neuroendocrine transducer cells of the hypothalamus. Examples of hormone regulators (Table 1.12–3) include corticotropin-releasing hormone (CRH), which stimulates adrenocorticotropin (ACTH); thyrotropin-releasing hormone (TRH), which stimulates the release of TSH; gonadotropin-releasing hormone (GnRH), which stimulates the release of luteinizing hormone (LH) and follicle-stimulating hormone (FSH); and somatostatin (somatotropin release–inhibiting factor [SRIF]) and growth hormone–releasing hormone (GHRH), which influence growth hormone (GH) release. Chemical signals cause the release of these neurohormones from the median eminence of the hypothalamus into the portal hypophyseal bloodstream and subsequent transport to the pituitary to regulate the release of target hormones. Pituitary hormones, in turn, act directly on target cells (e.g., ACTH on the adrenal gland) or stimulate the release of other hormones from peripheral endocrine organs. In addition, these hormones have feedback actions that regulate secretion and exert neuromodulatory effects in the CNS.

HORMONE SYNTHESIS AND STRUCTURE

Peptide hormones represent subsections of larger amino acid chains or polypeptides called *prohormones.* Production of a peptide hormone occurs by the cleavage of its prohormone chain at a given site on the chain by the appropriate enzyme. Proopiomelanocortin (POMC) is an example of a prohormone that contains the sequences for ACTH, β-endorphin, β-lipotropin, and melanocyte-stimulating hormone (MSH). Some hormones, called *dimers,* contain two or more peptide chains (e.g., FSH, LH, and TSH). Further cleavage of these hormone peptide chains in the course of metabolism may create additional biologically active peptides that have different effects from those of the parent peptide, and even minor modifications of structure can drastically change binding properties and metabolic processing.

Tropic hormones, such as ACTH and gonadotropins, in turn, induce steroidogenesis in two distinct ways. Acute regulation occurs through activation and rapid synthesis (over minutes) of steroidogenic acute regulatory (StAR) protein, which regulates the rate-limiting step of steroid hormone synthesis, the transport of cholesterol from the outer to the inner mitochondrial membrane. In contrast, chronic stimulation induces transcription and increases P450scc protein and steroidogenesis over hours to days.

Table 1.12–3.
Examples of Regulating Hormones

Regulating Hormone	Hormone Stimulated (or Inhibited)
Corticotropin-releasing hormone	Adrenocorticotropic hormone
Thyrotropin-releasing hormone	Thyroid-stimulating hormone
Luteinizing hormone–releasing hormone	Luteinizing hormone
Gonadotropin-releasing hormone	Follicle-stimulating hormone
Somatostatin	Growth hormone (inhibited)
Growth hormone–releasing hormone	Growth hormone
Progesterone, oxytocin	Prolactin
Arginine vasopressin	Adrenocorticotropic hormone

CELLULAR MODE OF ACTION

Genomic

The first known mode of action of steroid hormones (glucocorticoids, estrogen, and testosterone) and thyroid hormones (triiodothyronine [T_3] and thyroxine [T_4]) is by binding to intracellular receptors in the cytoplasm. The hormone-receptor complex, in turn, binds to common response elements on chromosomal DNA and alters transcription through a conformational change that unmasks the binding site. The hormone complex can also interact with transcription factors such as those produced by c-*fos,* c-*jun,* or activator protein-1 (AP-1) to amplify or inhibit gene expression and, by these mechanisms, regulate the induction of such gene products as enzymes and other cell proteins that affect metabolic change.

Nongenomic

Alternatively, as in the case of estrogen-stimulated prolactin release, certain hormones may exert a physiologic effect within seconds to minutes, a time course precluding a genomic mechanism. Nongenomic action is hypothesized to involve membrane hormone receptors. Some nongenomic effects appear to be mediated through distinct, nonclassical membrane receptors in that they are not blocked by classical receptor inhibitors and do not require gene transcription, protein synthesis, or a coagonist. Hormones also may act through ion-gated neurotransmitter receptors as coagonists or antagonists, as in the modulation of gamma aminobutyric acid type A ($GABA_A$) receptors by neurosteroids or by altering the fluidity and microenvironment of membrane receptors through the intercalation of the steroid in the phospholipid bilayers. Genomic and nongenomic mechanisms may be active simultaneously, with cross-talk a likely occurrence. Rapid actions of glucocorticoids include facilitating the surface diffusion of glutamate receptors and estradiol and the membrane trafficking of neurotensin and GABA-A receptors.

Combined Action

An example of a hormone-induced behavioral response with both genomic and nongenomic mechanisms is corticosteroid stimulation of aggressive behavior in rats. A rapid surge of glucocorticoids precedes the aggressive behavior, but early and later stimulation of aggressive behavior is mediated differently. The initial phase is promoted by a nongenomic mechanism needed for rapid response, which can play a decisive role in the outcome of the encounter. (In the case of rats, an early aggressive response may scare off the intruder and avoid a fight.) The effects of this mechanism rapidly subside, and aggressive behavior in later phases is stimulated by genomic mechanisms. This later, genomic corticosteroid-stimulated behavior is blocked by a protein synthesis inhibitor, but earlier nongenomic stimulation of aggression is not. Besides meeting the immediate need for rapid action, the nongenomic mechanism may serve the purpose of preparing for the genomic response by activating changes needed for an aggressive response, such as changes in energy metabolism.

Tissue Specificity

Many hormones, such as estrogen, act through a multistep process involving a hormone-receptor complex. This complex binds to specific

DNA sites called hormone response elements. The formation of a cluster with coactivators, corepressors, and transcriptional factors stimulates transcription and protein synthesis, with the response specific to the type of cell. Tissue specificity may arise from several mechanisms. Selective estrogen receptor modulators (SERMs) are synthetic hormones developed to target certain actions while avoiding other unwanted responses through coactivators or repressors, or other elements of the cluster that do not recognize the SERM in certain tissues. Alternatively, many hormones, such as estrogen or DHEA, may act directly or indirectly as a prohormone to be converted into other hormones in specific tissues. For example, tibolone is a synthetic hormone structurally related to 19-nortestosterone derivatives that has weak estrogen, progestogen, and androgen effects in specific tissues, in part because it is differentially metabolized. Accordingly, it has been used to reduce hot flashes and sweating, improve mood and libido and sexual functioning, decrease bone loss, stimulate semantic memory, and lessen vaginal atrophy without stimulating the growth of endometrium or breast tissue as do estrogens.

CHARACTERISTICS OF ENDOCRINE ACTIVITY

In general, hormonal compounds often exert their effect in a tonic rather than phasic fashion, being diffused in a less precise manner than a neurotransmitter and over a longer time period. Theoretically, such a characterization would allow hormones to be more closely linked to integrated behavioral responses. The release of many hormones is pulsatile, and the pattern of these pulses (i.e., duration, interpulse interval, the slope of increase or decrease in rate, and amplitude) is crucial to their effects. Other factors that can influence the regulation and effect of a hormone in a given individual include genetic differences, a history of exposure to the hormone during critical developmental encoding periods, the frequency and chronicity of past exposure, the time since last exposure, and the status of other influences on the target system. A decrease in the amplitude of response after repeated exposure is referred to as *habituation,* and an enhancement is termed *sensitization.* Facilitation of a previously habituated response after exposure to a novel stimulus or more severe stressor is called *dishabituation* (i.e., this allows enhancement of the previously habituated response to cope more effectively with the stimulus). In the case of the hypothalamic–pituitary–

adrenal (HPA) system, the release of cortisol by the adrenal gland is dependent on an integration of three separate control systems. These include an underlying circadian rhythm regulated by the suprachiasmatic nucleus, a stress-responsive circuit involving inputs to the hypothalamus from the brainstem, limbic system, and cerebral cortex, and a feedback control system exerted through two classes of corticosteroid receptors.

PHARMACOGENETICS AND PHARMACOGENOMICS

The study of pharmacogenetics and pharmacogenomics is leading to a better understanding of the mechanisms of psychopathology and its response to treatment. Pharmacogenetics refers to the effects of single genes. Pharmacogenomics is the genome-wide approach to discovering individual differences in the outcome of drug therapy. Recent studies of treatment response to antidepressants suggest that polymorphisms in some genes, such as for glucocorticoid receptors (GRs), CRH receptor 1, and a GR-regulating cochaperone, can predict treatment outcome or rapidity of response, but to date, studies have been of insufficient size, and the findings are difficult to replicate. Larger scale genome wide association studies are needed. Polymorphisms in genes that regulate hormonal response or in genes for other chemical messengers that modulate hormonal activity may also explain treatment responsiveness through more indirect mechanisms. Examples include genes for neurotrophic factors; excitatory cytokines, such as tumor necrosis factor; α-adrenergic receptors; GABA receptors; and metabolizing enzymes, such as monoamine oxidase inhibitor (MAO). Treatment interventions, including nonpharmacologic ones such as exercise, can produce changes in the expression of certain genes that may be more beneficial to those with particular genotypes. The use of genetically engineered mice has allowed the targeting of specific genes for the study of the dysfunction of hormonal systems in relation to anxiety and depression, including peptides modulating hormone release, receptors, binding proteins, and proteins controlling the access of hormones into the CNS. This research is still in its early stages, but the clinical implications of discovering genetic differences as biomarkers for subtyping the disorder and selection of treatment agents are exciting.

DEVELOPMENTAL PSYCHONEUROENDOCRINOLOGY AND EPIGENETIC TRANSMISSION

Although a review of the effect of hormones on brain development is beyond the scope of this chapter, it is important to note that hormones can have organizational as well as activational effects. Exposure to gonadal hormones during critical stages of neural development directs changes in brain morphology and function (e.g., sex-specific behavior in adulthood) and differentiation of dopaminergic neurons. Similarly, thyroid hormones are essential for the normal development of the CNS, and thyroid deficiency during critical stages of postnatal life severely impairs the growth and development of the brain, resulting in behavioral disturbances that may be permanent if replacement therapy is not instituted. Prenatal exposure in animals to endogenous or exogenous glucocorticoids or to stressful circumstances reduces offspring birth weight and may result in long-lasting changes in immune response, hypertension, hyperglycemia, hyperinsulinemia, cardiovascular function, and neuroendocrine responses and behavior, including attentional deficits, increased anxiety, and disturbed social behavior. Maternal deprivation in strains of rats with increased glucocorticoid response to stress has similarly been shown to lead to increased startle responses, anxiety-like behavior, increased alcohol preference, and difficulties with spatial learning in adulthood.

Epigenetic transmission occurs through changes in chromatin and DNA structure that do not involve changes in the sequence of DNA but which alter gene expression and phenotype. Maternal behavior can cause epigenetic alterations to steroid receptor genes and produce long-term changes influencing postpartum behavior. Cross-fostered rats show maternal care behaviors similar to those of their “adoptive” mother rather than those of their biologic parent. This change in hormone-influenced behavior is thought to involve estrogen and oxytocin receptor changes.

PSYCHONEUROENDOCRINOLOGY METHODOLOGY IN HUMANS

Human studies are often limited to examining the relationship between hormone concentrations or changes in concentration and psychiatric disease states, symptoms, neurotransmitter function, or response to treatment. Concentrations may be measured in plasma, urine, saliva, cerebrospinal fluid (CSF), or postmortem tissue and are sometimes used as indicators of regulatory neurotransmitter function in response to a given stimulus. For example, cortisol or prolactin response to d-fenfluramine has been used to assess serotonin activity and GH response to clonidine (Catapres) to assess dopaminergic function.

One example of a provocative psychoendocrine test used to assess the HPA system is the combined dexamethasone–CRH test, which assesses response to two hormonal stimuli, one inhibiting (dexamethasone [Decadron]) and the other stimulating (CRH), on cortisol secretion. Typically, 1.5 mg of dexamethasone is given in the evening, and plasma cortisol concentration is measured 16 hours later on the following day. An infusion of 100 μg of CRH is then given, and cortisol level and ACTH are measured again several times over the next 75 minutes. Abnormalities in this test are found in a variety of psychiatric illnesses, including bipolar disorder, major depression, schizophrenia, and PTSD, although the sensitivity or specificity for these disorders does not make the test a reliable diagnostic tool. Several studies or case reports have suggested an association between changes in suppression and outcome for these disorders.

Another method of studying the relationship of hormones to psychiatric disorders is the administration of hormones or other secretagogues (substances that cause other substances to be secreted) to experimentally correct an abnormal hormone concentration and examine the effects. Some secretagogues can be given orally to replace a parenterally administered hormone and be given multiple times during a day to mimic circadian rhythms or other variations in concentration when studying circadian effects.

Functional brain imaging studies can help localize the areas of changed activity produced by hormonal action affecting a variety of cognitive and behavioral activities.

HYPOTHALAMIC–PITUITARY–ADRENAL AXIS

Since the earliest conceptions of the stress response by Hans Selye and others, investigation of HPA function has occupied a central position in psychoendocrine research. CRH, ACTH, and cortisol are all elevated in response to a variety of physical and psychological stresses and serve as prime factors in the maintenance of homeostasis and the development of adaptive responses to novel or challenging stimuli. Allostatic load, an index of chronic stress-induced physiologic dysregulation, has been correlated with a number of adverse behavioral trajectories and outcomes. Glucocorticoids regulate glucose metabolism, blood pressure, immune response, lipid metabolism, glycogen deposition, and energy homeostasis for fight-or-flight response and are essential for embryonic development and neonatal survival. A normal glucocorticoid stress response helps to recover after the challenge and helps to store the experience for coping with future encounters, although sustained levels may impair some forms of memory. The hormonal response is dependent not only on the characteristics of the stressor itself but also on how the individual assesses and is able to cope with it. In primates, social status can influence adrenocortical profiles and, in turn, can be affected by exogenously induced changes in hormone concentration. Aside from generalized effects on arousal, distinct effects on sensory processing, stimulus habituation and sensitization, pain, sleep, and memory storage and retrieval also have been documented with CRH, ACTH, and cortisol (or corticosterone).

Exposure to chronic stress produces increased concentrations of CRH and arginine vasopressin (AVP) in the paraventricular nucleus (PVN) of the hypothalamus and, over time, leads to a reduction in CRH receptor number in the anterior pituitary. The release of CRH results in simultaneous activation of the locus coeruleus noradrenergic circuit, which functionally increases arousal and selective attention and decreases vegetative functions, such as appetite and sex drive. ACTH concentrations are increased in acute stress but diminish over time in chronic stress.

GRs are ubiquitously distributed throughout the body. At least two intracellular receptor subtypes bind corticosteroids: The mineralocorticoid receptor (MR) (or type I receptor) and the GR (or type II receptor). The human GR has an α and a β form, the α form showing high affinity for dexamethasone, a modest affinity for cortisol, and low affinity for aldosterone, deoxycortisol, and the sex steroids; and the β form acting as a negative regulator. MRs have high affinity but low capacity, whereas GRs have low affinity but high capacity. Owing to the difference in affinity, low corticosteroid levels generally result in a predominant MR occupation, with higher steroid levels shifting the balance in favor of the GR.

"Permissive actions" of MR activation before the onset of stress are tonically involved in the mediation of the initial stress response. When more GRs become occupied, local excitability may decrease or, in some cases, may increase on a short-term basis to inhibit or enhance the initial effects of stress-responsive hormones. Over time, continued stress produces increasing "allostatic load," the sustained effects of hypercortisolemia giving rise to hyperglycemia, increased visceral fat, elevated blood pressure, decreased bone density, hyperlipidemia, and changes in electrolytes and immune response. Interactions between MR and GR in the hippocampus may be relevant to understanding the regulation of stress response in depression and the efficacy of antidepressants. Studies of GR function have pointed to a relevant decrement in response to agonists in depressed patients, but MR function is generally preserved.

Three types of inhibitory feedback of glucocorticoids on CRH and ACTH have been characterized. Fast, rate-sensitive feedback occurs while plasma concentrations of the glucocorticoid are rising and regulate release rather than synthesis of CRH and ACTH. Intermediate, delayed feedback occurs from 1 to 2 hours after steroid administration, is dose sensitive and duration sensitive rather than rate sensitive, and inhibits the release of CRH and ACTH as well as the synthesis of CRH. Slow feedback is similar to intermediate feedback but occurs over a longer period of time (hours) and is distinguished by decreased synthesis of CRH and ACTH (and other POMC derivatives).

Glucocorticoid release is amplified, at least acutely, by serotonergic and cholinergic input and is inhibited by GABA and opioids. Catecholamines play a role in response to stress and interact with the limbic–hypothalamic–pituitary–adrenal axis. The acute addition of glucocorticoids can increase dopaminergic activity in certain areas of the brain, but chronic hypercortisolemia may decrease dopamine activity (depending on the region involved). Glucocorticoids may exert their influence on dopamine activities by functioning as a transcriptional regulator, by acting on promoter regions of dopamine receptors, and by modulating catecholamine biosynthesis.

Pathologic alterations in HPA function have been associated with a number of psychiatric disorders, including mood disorders, PTSD, dementias, and substance use disorders (SUDs). Disturbances of mood are found in more than 50% of patients with Cushing syndrome (characterized by elevated cortisol concentrations), with psychosis or suicidal thought apparent in more than 10% of cases studied. Cognitive impairments similar to those seen in major depressive disorder (MDD) are common and relate to the degree of hypercortisolemia present and a possible reduction in hippocampal size. In general, therapeutically induced reductions in cortisol levels result in a normalization of mood and mental status. Mifepristone (RU486) has been reported to ameliorate psychosis and depression in Cushing patients, and several studies have reported that it also alleviated psychosis or depression in psychotic depression not associated with Cushing syndrome. In Addison disease (characterized by adrenal insufficiency and diminished glucocorticoid output), apathy, social withdrawal, impaired sleep, and decreased concentration frequently accompany prominent fatigue. Replacement of glucocorticoid results in resolution of behavioral symptomatology, although correction of the associated electrolyte disturbances by itself does not.

Exogenous administration of synthetic corticosteroids is commonly associated with mild activation, but a higher dose and more sustained treatment may produce depression, mood lability, memory, attentional impairment, and sometimes psychosis. Alterations in HPA function associated with depression include elevated cortisol concentrations, failure to suppress cortisol in response to dexamethasone, increased adrenal size and sensitivity to ACTH, a blunted ACTH response to CRH, and elevated CRH concentrations in the brain. In addition to altered slow feedback, several groups have demonstrated decreased sensitivity to glucocorticoid fast feedback as well, and increased cortisol near sleep onset and on first awakening have been shown to be predictive of increased risk of future depression. The pattern of these abnormalities has not, thus far, led to a definitive theory of mechanism, and other elements, such as AVP, are important to understanding the change in homeostasis. The finding that corticosteroids have multiple regulatory effects on serotonergic function, particularly on the serotonin (5-hydroxytryptamine type 1A [5-HT$_{1A}$]) receptor, may also be relevant, as may be the state-dependent stimulant-like effects that glucocorticoids can exert on mesencephalic dopamine transmission.

Excessive glucocorticoid activity may contribute to the symptoms of psychotic mood disorders. Some mood stabilizers, such as lithium, carbamazepine (Tegretol), and valproate, inhibit the transcription activity of GRs and may exert some of their therapeutic effects in this way, and the development of CRH-1 antagonists as potential pharmacotherapies for depression and anxiety disorders, alone or in combination with current antidepressant medications, is currently underway. Cortisol response to CRH or ACTH is abnormal in some psychiatric illnesses, although studies in this area have been highly variable and confusing. For example, in one study, patients with PTSD exhibited low cortisol levels despite high CRH activity, while in another, cortisol response to ACTH was increased.

A number of studies have linked abnormalities in HPA activity, hippocampal volume, and certain memory functions, but it is not clear which is causal. One imaging study of identical twins discordant for PTSD showed similarly diminished hippocampal volume, but the treatment of the symptoms of PTSD with SSRIs is followed by an increase in hippocampal volume and improvement in memory. Genomic and nongenomic actions of cortisol in hippocampal cells promote enzymatic processes that can lead to cell death, but as the hippocampus mediates negative feedback of cortisol release, hippocampal damage may impair cortisol negative feedback. Increasing glucocorticoid levels by administration or other methods has decreased hippocampal volume in several (but not all) studies in non-human primates and other species. PTSD symptoms and cortisol at baseline predicted hippocampal size reduction later in one study, and a decrease in cortisol following the treatment of Cushing disease has been shown to reverse hippocampal atrophy.

In schizophrenia, difficulty suppressing cortisol after dexamethasone is associated with negative symptoms and cognitive impairment. Clozapine (Clozaril) improves cognitive functioning, possibly in part through preventing or reversing cortisol-induced hippocampal damage, and reverses stress-induced impairment of long term potentiation (LTP), a measure of synaptic plasticity important for the storage of information and memory, in the frontal cortex. The GR antagonist mifepristone also blocks cortisol-induced impairment of LTP.

Stress is often reported as a reason for relapse in substance-dependent individuals, and animal studies suggest that glucocorticoid administration is involved in drug self-administration. Glucocorticoid acute enhancement of dopamine activity likely contributes to the motivational changes involved with drug use. Alcohol usage and withdrawal produce profound changes in HPA regulation, pseudocushingoid features are a phenotypic feature of chronic alcohol intake, and HPA adaptation to alcohol withdrawal varies by family history of alcoholism. There is also suggestive evidence that alteration in HPA response to acute alcohol challenge may represent an endophenotype of genetic risk of dependence. CRH is involved not only in the HPA axis but also in extrahypothalamic systems that play a role in the relapse to alcohol and other drug use after stress.

Another disorder associated with a disturbance in glucocorticoid negative feedback is polydipsia. Polydipsic schizophrenic patients, particularly those with hyponatremia, show marked impairment in cortisol suppression of ACTH.

METABOLIC SYNDROME

Metabolic syndrome is a cluster of multiple metabolic risk factors, including elevated insulin levels and resistance, hyperglycemia, visceral obesity, hyperlipidemia, and hypertension. Glucocorticoids interfere with glucose transport and utilization. The insulin resistance and increased insulin concentrations that develop decrease lipid deployment and increase lipid accumulation. GRs are found in high concentrations in intra-abdominal fat tissue, accounting for the truncal obesity resembling that in Cushing disease. Elevated intracellular glucocorticoid tone is thought to be an etiology of metabolic syndrome. Selective inhibitors of 11β-hydroxysteroid dehydrogenase 1, which decrease cortisol production, are thus being tested as a therapeutic intervention. Stress can produce these changes, along with increased vascular resistance through effects on the somatic nervous system. Many patients with schizophrenia who exhibit increased cortisol and epinephrine have more central obesity, higher plasma cortisol, and an increased risk of diabetes, even when medication free. Many of the atypical antipsychotics and some of the other psychotropic medications also can cause hyperglycemia, hyperlipidemia, and visceral obesity, limiting their use. Reports of changes in HPA axis hormones with treatment by typical and atypical antipsychotics have been inconsistent. However, atypical antipsychotics may cause more glucose elevation after glucose challenge compared to typicals, putting patients at greater risk of developing diabetes.

PROOPIOMELANOCORTIN, MELANOCORTINS, MELANOCYTE-STIMULATING HORMONE, AND

MELATONIN

POMC is a prohormone from which several hormones called melanocortins are derived. These include ACTH (described above), melanocortins, and MSH.

Studies of rodent sexual behavior suggest that the activation of melanocortin receptors may be an avenue to treat women with hypoactive sexual desire, and the development of melanocortin medications for sexual disorders is underway. Blockade of the melanocortin-4 receptor decreases the reinforcing effects of cocaine, and melanocortins may be neuroprotective for ischemic injury, reducing hippocampal cell death, an area important in learning. Inactivating the melanocortin-5 receptor in rats also reduces pheromone-induced aggression.

MSH, an anterior pituitary peptide, controls the secretion of melatonin and melanin. Melanin is a pigment not only in hair, skin, and eyes but also in neurons in the substantia nigra and the locus coeruleus. Melatonin (described below) is a hormone that plays a major role in regulating the circadian rhythm. Phenothiazines increase pituitary MSH secretion and pigmentation in some patients. In a double-blind, cross-over trial in humans, an infusion of α-MSH resulted in a significant improvement in verbal memory but little change in mood. A dose-related biphasic effect on mood has been reported for MSH release–inhibiting factor. Recent data indicate that MSH interacts with leptin to counteract neuropeptide Y (NPY), decrease food intake, and increase energy expenditure. It may also antagonize the antidepressant effects and anxiolytic effects of NPY while also antagonizing the anxiogenic effects of the proinflammatory cytokine interleukin-1β. MSH inhibitors have been found to decrease dyskinetic movements in animals.

Melatonin is a pineal hormone that is derived from the serotonin molecule and controls photoperiodically mediated endocrine events (particularly those of the hypothalamic–pituitary–gonadal [HPG] axis). It also modulates immune function, mood, and reproductive performance; it is a potent antioxidant and free-radical scavenger and may have oncostatic effects. Melatonin has a depressive effect on CNS excitability and exerts neuroprotective effects against excitotoxicity. Melatonin has analgesic effects through its actions on opiate receptors and has regulatory effects on serotonin metabolism. Altered secretory patterns and levels of melatonin have been found in various psychiatric disorders, such as unipolar and bipolar depression, seasonal affective disorder (SAD), bulimia, anorexia, schizophrenia, panic disorder, and obsessive-compulsive disorder. Although suppression of melatonin is not necessary for the efficacy of light therapy in SAD, melatonin can be a useful therapeutic agent in the treatment of circadian phase disorders, such as jet lag, and intake of melatonin increases the speed of falling asleep as well as its duration and quality. A number of synthetic melatonin-like drugs have recently been or are being developed as hypnotic agents.

ENDOGENOUS OPIOIDS

Since the discovery of endogenous opioid receptors and their endogenous ligands in the early 1970s, research into the possible behavioral roles of such compounds has grown at a rapid pace. At least three different receptor systems for these ligands have been identified (μ, δ, and κ), and each of these has subtypes. Opioid receptors μ, δ, and κ are activated by the endogenous ligands β-endorphin, enkephalins, and dynorphin, respectively, among others. σ-Receptors, originally included because some opioids that suppressed coughing acted on them, are no longer considered opioid receptors. β-Endorphin is the principal opioid peptide prototype and, like ACTH, MSH, and β-lipotropin, is derived from POMC. Methionine enkephalin (met-enkephalin) and leucine enkephalin (leu-enkephalin) are two small pentapeptides that also possess direct opioid activity, met-enkephalin being contained in POMC and another precursor, proenkephalin, and leu-enkephalin being contained in the prohormones proenkephalin and prodynorphin. The best-documented function of endogenous opiates is analgesia and alteration of pain perception, but effects on stress, appetite regulation, learning and memory, motor activity, and immune function also appear to be of physiologic importance. CRF and endogenous opioids also interact to coregulate the locus coeruleus, a role important in early adaptation to stress.

Early enthusiasm for the idea that a dysfunctional opiate system was etiologically related to schizophrenia has waned in the face of contradictory findings. Increases in various endorphin compounds have been reported in plasma as well as in postmortem brain tissue of patients with schizophrenia, but studies of short- and long-term treatment with opioid antagonists show no consistent or reproducible effects on psychopathology. However, hypersecretion of opioids in the CNS of patients with PTSD has been postulated to be an adaptive response to a traumatic experience, and CSF β-endorphin concentration is inversely related to intrusive and avoidant symptoms of PTSD. Naltrexone (Vivitrol; Revia), an opioid receptor antagonist, has decreased symptoms in autistic children and can improve functioning, with decreases in social withdrawal, stereotypy, and abnormal speech being directly related to decreases in β-endorphin levels.

In animal models, a number of stressors, including those that are purely psychological, induce opiate-mediated effects such as analgesia and hypomotility that are reversed by the opiate antagonist naloxone (Narcan). Several studies have found that concentrations of plasma β-endorphin in humans are correlated with measures of stress elicited by surgery, exercise, parachuting, or pain. Short-term administration of opioid agonists also increases eating, whereas antagonists reduce food intake by as much as 30%, diminish the intake of fats and highly palatable foods, and increase caloric expenditure. Thus far, however, their long-term use in obesity and eating disorders has not proven clinically useful. Some studies of opioid antagonist treatment have found certain binge parameters (e.g., duration) to be reduced in bulimia, but no studies have demonstrated weight loss in obese subjects. Naltrexone is helpful as an adjunct in the treatment of alcohol as well as opioid dependence, reducing drinking, craving, the high derived from drinking alcohol, and the likelihood that sampling alcohol would precipitate a relapse. In addition to the μ agonist methadone, buprenorphine, a partial μ agonist, has been helpful for opioid dependence, probably both because of its alleviation of withdrawal and because of its blockade of opioid-induced euphoria.

It is well known that exogenous opioids (e.g., heroin and morphine) can induce a euphoric mood state and that exercise increases the release of endogenous opioids and is associated with mood enhancement; these observations, together with findings that exercise-induced mood enhancement is blocked by naloxone, suggest that endogenous opioids are also involved in the mediation of mood. Such conclusions must be moderated, however, by the recognition that additional specific and nonspecific effects on other neurochemical systems are possible contributors to exercise-related mood effects.

HYPOTHALAMIC–PITUITARY–GONADAL AXIS

GnRH is a decapeptide that was sequenced and synthesized by Andrew Schally and colleagues in 1971. GnRH administration results in the rapid release of LH and FSH from the pituitary in healthy subjects and, in some pathologic states, such as acromegaly, an abnormal release of GH or prolactin. The cell bodies of GnRH are located principally over the optic chiasm in the arcuate area, with projections to the median eminence and in the lamina terminalis. GnRH release is stimulated by norepinephrine and is inhibited through negative feedback of gonadal steroids. Administration of GnRH can result in a depressive-like state, characterized by hot flashes, anxiety, insomnia, decreased libido, and fatigue in euthymic subjects, but it is not known whether this is a direct effect of the agent or is caused by the hypoestrogenic state that is produced when GnRH is given continuously. A GnRH analog has been found to have some efficacy in the treatment of paraphilia by decreasing testosterone.

The gonadal hormones (progesterone, androstenedione, testosterone, E_2, and others) are steroids that are secreted principally by the ovary and testis, but significant amounts of androgens arise from the adrenal cortex as well. The prostate gland and adipose tissue are also involved in the synthesis and storage of dihydrotestosterone and contribute to individual variance in sexual function and behavior.

The timing and presence of gonadal hormones play a critical role in the development of sexual dimorphisms in the brain. Developmentally, these hormones direct the organization of many sexually dimorphic CNS structures and functions, such as the size of the hypothalamic nuclei (INAH3) and corpus callosum, the neuronal density in the temporal cortex, the organization of language ability, and the responsiveness in Broca area. Women with congenital adrenal hyperplasia, a deficiency of the enzyme 21-hydroxylase leading to high exposure to adrenal androgens in pre- and postnatal life, have been found to be more aggressive and less interested in “traditional female roles” than female control subjects. Sexual dimorphisms may also reflect acute and reversible actions of relative steroid concentrations (e.g., higher estrogen levels transiently increase CNS sensitivity to serotonin). The importance of timing in the exposure to sex steroids is highlighted by studies of administration in gender identity disordered adults. These individuals perform on cognitive tests according to their physical gender and not to their perceived gender. Treatment with sex steroids produces substantial cross-sex changes but no changes in cognitive performance.

Testosterone

Testosterone is the primary androgenic steroid, having androgenic (i.e., facilitating male gonadal development) and anabolic (i.e., facilitating linear body growth and somatic growth) functions. Testosterone is important for sexual desire in men and women. In males, muscle mass and strength, sexual activity, desire, thoughts, and intensity of sexual feelings are dependent on normal testosterone levels, but these functions are not clearly augmented by supplementing testosterone in those with normal androgen levels. The addition of small amounts of testosterone to normal hormonal replacement in postmenopausal women has, however, proven to be as beneficial as its use in hypogonadal men. Testosterone has both genomic and nongenomic actions, including acting directly at the cell membrane or modulating the activity of other membrane receptors or second messengers, and regulating the actions of a wide range of neurotransmitters.

Testosterone administration has been shown to result in increased violence and aggression in animals, and testosterone level tends to be correlated with aggression in humans, but anecdotal reports of increased aggression with testosterone treatment have not been uniformly substantiated in human scientific investigations. Reports may be confounded by factors such as past history and social factors, which are particularly important determinants of the effects of hormones in primates and in humans. For instance, in the cynomolgus monkey, administration of testosterone increases dominant behavior in dominant monkeys and submissive behavior in submissive monkeys; in hypogonadal men, it improves mood and decreases irritability. Psychological tests such as the Point Subtraction Aggression Paradigm have been developed to measure aggression in a human laboratory setting and support the clinically observed association between increased testosterone and aggression, although personality characteristics are confounding factors. Free testosterone in the CSF has been associated with measures of aggression, sensation seeking and monotony avoidance, suspiciousness, and reduced socialization, and testosterone-induced aggression in mice has been associated with a decrease in brain allopregnanolone. Abnormal testosterone levels have been inconsistently reported in a variety of disorders, including schizophrenia, PTSD, depression, and anorexia. Studies of testosterone treatment in depression have generally been inconclusive, although two randomized, placebo-controlled trials found intramuscular (IM) therapy effective in eugonadal men with late-life depression and testosterone gel supplementation efficacious in refractory depression. Testosterone may play a role in pre-menstrual syndrome, as concentrations are higher in those with the disorder. Testosterone has also been reported to improve mood and fatigue in HIV-positive men. A recent trial in women with treatment-resistant depression, however, found no benefit.

Anabolic steroids are synthetic derivatives of testosterone modified to enhance their anabolic actions, such as muscle growth, rather than their androgenic actions. Varying effects of anabolic-androgenic steroids on a wide variety of moods have been noted anecdotally. In one prospective placebo-controlled study of anabolic-androgenic steroid administration in normal subjects, positive mood symptoms, including euphoria, increased energy, and sexual arousal were reported, in addition to increases in negative mood symptoms of irritability, mood swings, violent feelings, anger, and hostility. In addition to a variety of adverse physical effects, anabolic steroids can cause a variety of adverse psychiatric effects, including aggressive behavior, mood and psychotic disturbances, and psychological dependence, particularly when taken in high doses.

Dehydroepiandrosterone

Dehydroepiandrosterone (DHEA) and DHEA-S are adrenal androgens secreted in response to ACTH and represent the most abundant circulating steroids. DHEA is also a neurosteroid that is synthesized in situ in the brain. DHEA has many physiologic effects, including a reduction in neuronal damage from glucocorticoid excess and oxidative stress. Behavioral interest has centered on its possible involvement in memory, mood, and a number of psychiatric disorders. *Adrenarche* is the prepubertal onset of adrenal production of DHEA-S and may play a role in human maturation by increasing the activity of the amygdala and hippocampus and promoting synaptogenesis in the cerebral cortex. DHEA has been shown to act as an excitatory neurosteroid and to enhance memory retention in mice, but studies of DHEA administration in humans have not consistently shown any improvement in cognition. Several trials of DHEA administration point to an improvement in well-being, mood, energy, libido, and functional status in depressed individuals. Administration of DHEA to women with adrenal insufficiency (e.g., Addison disease) has repeatedly been demonstrated to enhance mood, energy, and sexual function; effects in men remain to be assessed. Mood, fatigue, and libido improved in HIV-positive patients treated with DHEA in one study, and DHEA and DHEA-S have been found to be inversely correlated with the severity of attention deficit-hyperactivity disorder (ADHD). Women diagnosed with fibromyalgia have significantly decreased DHEA-S levels, but supplementation does not improve outcomes. Several cases of possible DHEA-induced mania have been reported, and DHEA has been reported to be inversely related to extrapyramidal symptoms (EPS) in schizophrenics treated with antipsychotics. DHEA administration in these cases improves EPS.

Double-blind treatment studies have shown antidepressant effects of DHEA in patients with major depression, midlife-onset dysthymia, and schizophrenia, although beneficial effects on memory have not been reliably demonstrated. A small, double-blind trial of DHEA treatment for Alzheimer disease failed to reveal significant benefit, although a near-significant improvement in cognitive function was seen after 3 months of treatment.

Animal studies suggest that DHEA may be involved in eating behavior, aggressiveness, and anxiety as well, with its effects resulting from its transformation into estrogen, testosterone, or androsterone from its antiglucocorticoid activity or from direct effects on $GABA_A$, n-methyl-D-aspartate (NMDA), and σ-receptors. Because of the putative antiglucocorticoid effects, the ratio of cortisol to DHEA levels may be particularly important in understanding adaptive responses to stress. Both cortisol and DHEA appear to be involved in fear conditioning, with the cortisol/DHEA ratio hypothesized to be an index of the degree to which an individual is buffered against the negative effects of stress. This ratio has been found to be related to some measures of psychopathology and response to treatment, predicting the persistence of the first episode of major depression and being related to the degree of depression, anxiety, and hostility in schizophrenics and response to antipsychotic treatment. PTSD patients have higher DHEA levels and lower cortisol/DHEA ratios related to symptom severity, suggesting a role in PTSD recovery. Fear-potentiated startle is larger in individuals with high as compared to low cortisol/DHEA-S ratios and is positively associated with cortisol and negatively with DHEA-S. Greater DHEA response to ACTH is related to lower PTSD ratings and the cortisol/DHEA ratio to negative mood symptoms. A genetic variation in an ACTH receptor promoter has been found to influence DHEA secretion in response to dexamethasone and may underlie some individual differences in stress response.

Estrogen and Progesterone

Sex differences in the prevalence or expression of psychopathology or response to treatment may be caused by differences in the concentrations of hormones or gender-related CNS differences in brain morphology and function. Mood and some other psychiatric disturbances are particularly likely to occur during times of sex hormone changes in women, such as the postpartum period, premenstrually, and perimenopause. The primary estrogens are estradiol (E_2), estrone (E_1), and estriol (E_3), with E_2 being the major secretory product of the ovaries. Two different estrogen receptors have been identified (α and β), each with different anatomical distributions and physiologic effects. Estrogens can influence neural activity in the hypothalamus and limbic system directly through the modulation of neuronal excitability and have complex multiphasic effects on nigrostriatal dopamine receptor sensitivity. Estrogens also enhance dopamine synthesis and release, modify basal firing rates, and can lead to stereotypical behavior in rodents. Accordingly, there is evidence that the antipsychotic effects of psychiatric drugs may change over the menstrual cycle and that the risk of tardive dyskinesia is partly dependent on estrogen concentrations. However, studies of symptom changes in schizophrenic women show significant differences for anxiety–depression and withdrawal–retardation subscales but not for psychotic subscales across the menstrual cycle. Nevertheless, lower levels of estrogen are associated with episodes of acute psychosis in both women and men and with more severe negative symptomatology as well as poorer cognitive function. Estrogen pretreatment attenuates anticholinergic drug-induced problems with attention.

Several studies have suggested that gonadal steroids modulate spatial cognition and verbal memory and are involved in impeding age-related neuronal degeneration. There is also some evidence that estrogen administration may decrease the risk and delay the onset of dementia of the Alzheimer type in postmenopausal women, but acute treatment in dementia has been ineffective in reducing symptoms.

Estrogen has mood-enhancing properties and can also increase sensitivity to serotonin, possibly by inhibiting MAO; in animal studies, long-term estrogen treatment results in a decrease in 5-HT_1 and an increase in 5-HT_2 receptors. In oophorectomized women, significant reductions in tritiated imipramine binding sites (which modulate presynaptic serotonin uptake) were restored with estrogen treatment. Severe postpartum depression has been successfully treated with sublingual 17-β-estradiol, as have depressive disorders in perimenopausal women in a large, randomized, double-blind trial. Prophylactic estrogen administration has been reported to prevent the recurrence of postpartum depression.

Associations between response to antidepressant treatment and age group in women have been found. Low-dose estrogen augmentation of antidepressant medication for perimenopausal women is reported to improve mood, and poorer response to selective serotonin releasing inhibitor (SSRI) treatment in older women can be eliminated by hormone replacement therapy. Inadequate response to SSRIs may be due to a chronic hypoestrogenic state, as serotonin receptor binding appears to be estrogen dependent. Tamoxifen (Nolvadex), a SERM used to treat breast cancer, showed beneficial effects on mania in one study. The greater sensitivity of women to certain kinds of stress may also be attributed in part to differences in tissue sensitivity. Brain norepinephrine system activation, for example, may be stronger in women because of the differential postsynaptic sensitivity of locus coeruleus neurons to CRH. Many of estrogen's psychiatric effects also could be mediated through its stimulation of brain-derived neurotrophic factor (BDNF) or NPY through the estrogen–BDNF–NPY cascade (see Neuropeptide Y).

Progesterone, the primary progestin, is produced by the corpus luteum of the ovary. Although progesterone itself may be anxiogenic, metabolites of progesterone (allopregnanolone and pregnenolone) appear to have anxiolytic and hypnotic properties via GABA_A agonistic activity. Progesterone is colocalized with serotonin in cells of the median raphe and causes increased serotonin uptake and turnover in the brain in several species. Progesterone, which has antiestrogen effects, such as downregulating estrogen receptors and increasing MAO activity, is often associated with dysphoric mood. The ratio of progestin to estrogen in oral contraceptives has been associated with negative mood change, although this effect varies with depression history, and meta-analyses have not shown a clear increase in depressive symptoms with contraceptive use overall. Independently, progesterone interactions with dopamine have been hypothesized as serving a possible etiologic role in schizophrenia.

The association of these hormones with serotonin is hypothetically relevant to mood change in premenstrual and postpartum mood disturbances. Women with a history of depression have higher FSH and LH and lower E_2 levels and are at risk of beginning perimenopause at a younger age. Similarly, the 2:1 female-to-male ratio in depression prevalence may speculatively be related to the rapid changes in hormone levels in menarche. Premenstrual dysphoric disorder is a disorder in which a constellation of symptoms resembling MDD (in some respects) occurs in most menstrual cycles, appearing in the luteal phase and disappearing within a few days of the onset of menses. No definitive abnormalities in estrogen or progesterone levels have been demonstrated in women with premenstrual dysphoric disorder, but decreased serotonin uptake with premenstrual decreases in steroid levels have been correlated with the severity of symptoms in some studies. Progesterone downregulates the estrogen receptor, and it has been suggested that, despite high circulating concentrations of estrogen, the luteal phase is a period of functional estrogen withdrawal, with concomitant effects on the serotonergic system. Recent evidence indicates that the abrupt decline of progesterone and allopregnanolone in the luteal phase results in increased production of the α_4 subunit of $GABA_A$ and changes in receptor sensitivity that could account for the typical behavioral symptoms noted. This effect is correlated with the insensitivity of the GABA receptor to modulation by the benzodiazepine class of tranquilizers (and hence is anxiogenic). SSRIs, particularly fluoxetine (Prozac), have demonstrated efficacy, and as many as 50% of women may respond to fluoxetine administered only in the second half of each cycle. Alprazolam (Xanax), a $GABA_A$ agonist, has been found to be more effective than placebo in several studies for the treatment of premenstrual dysphoric disorder. In women with severe symptoms that are not responsive to these treatments, the long-term use of a GnRH agonist to abolish menstrual cycling with added estrogen–progestin may be therapeutic.

The menstrual phase also has been associated with aspects of substance abuse. Although reports vary, craving for cigarettes and tobacco withdrawal appears to vary with the menstrual phase (worse in the luteal phase). Women show greater heart rate and pleasurable drug effects after cocaine administration during the follicular phase but report that cocaine improves dysphoric mood during the luteal phase.

The bulk of the psychological symptoms associated with menopause is actually reported during perimenopause rather than after the complete cessation of menses. Reported symptoms include worry, fatigue, crying spells, mood swings, diminished ability to cope, and diminished libido or intensity of orgasm. Estrogen replacement alone may be beneficial, but combination androgen–estrogen replacement may be superior for reinstating energy, a sense of well-being, and libido. In women with an intact uterus, the addition of progestin is necessary to protect against endometrial hyperplasia but can attenuate the beneficial effects of estrogen on mood.

The postpartum period appears to be a particularly risky time for the emergence or relapse of psychiatric illnesses. Women with bipolar disorder appear to be particularly sensitive to alterations in gonadal steroid levels. A high risk of relapse during the postpartum period has been observed and appears to be maintained over subsequent pregnancies; evidence for a familial preponderance of the puerperal trigger also exists, suggesting a genetic contribution. During pregnancy, a number of mechanisms play a role in dampening the response to stress in order to promote maternal care and the survival of the offspring. Hormones such as oxytocin and prolactin produce inhibitory effects on the HPA axis, decreased excitatory activity, and a more positive mood state. The reversal of many of these effects may contribute to the vulnerability of women to postpartum psychiatric illnesses. Lactation has the potential to dampen some postpartum changes and has been associated with decreased maternal postpartum depression.

PREGNENOLONE AND ALLOPREGNANOLONE

Neuroactive steroids, such as pregnenolone and allopregnanolone (allo), modulate activity at GABA_A, NMDA, σ-1, 5-HT_3, nicotinic, kainate, oxytocin, and glycine receptors, among others. Pregnenolone is a neurosteroid synthesized from cholesterol in the brain and is partially metabolized to all subsequent steroids. It, and especially its sulfate, pregnenolone sulfate, appear to have memory-enhancing properties in animal studies. Pregnenolone increases the rate and extent of the formation of microtubules, which are important in neuronal plasticity and function. Progesterone counteracts this effect. Pregnenolone sulfate is an excitatory neurosteroid and has GABA inhibitory effects.

Allo is a neurosteroid derived from progesterone with high concentrations in the CNS. It acts as a $GABA_A$ receptor agonist, decreases CRH concentrations in the hypothalamus, and reduces the anxiety evoked by CRH in rats. Testosterone-induced aggression in mice is thought to be mediated by downregulation in allo, and normalization of allo with progesterone and estrogen prevents aggression. SSRIs increase brain allo levels in rodents. Social isolation also downregulates allo and increases aggression. Fluoxetine reduces both allo downregulation and aggression. CSF levels of allo in humans have been correlated with clinical improvement over the first 8 to 10 weeks of SSRI treatment, an effect that appears to be independent of the serotonin reuptake inhibition. Allo increases in response to ethanol administration and may play a role in ethanol withdrawal through modulation of $GABA_A$. Administration of allo also has been shown to reduce anxiety and hyperlocomotion associated with benzodiazepine withdrawal in animals. Allo is currently unavailable for clinical use, but studies have already commenced on clinical treatment with allo or with related compounds. Brexanolone is an FDA-approved neurosteroid analog that has been shown to have efficacy in the treatment of moderate to severe postpartum depression, although utilization has been limited by its expense and the need to be administered through infusion.

PROLACTIN

Since its identification in 1970, the anterior pituitary hormone prolactin has been examined as a potential index of dopamine activity, dopamine receptor sensitivity, and antipsychotic drug concentration in studies of CNS function in psychiatric patients and as a correlate of stress responsivity. Prolactin signaling is mediated by the transmembrane prolactin receptor, PRL-R, which is expressed in a wide variety of tissues. The secretion of prolactin is under direct inhibitory regulation by dopamine neurons located in the tuberoinfundibular section of the hypothalamus and is, therefore, increased by classical antipsychotic medications. Prolactin also inhibits its own secretion by means of a short-loop feedback circuit to the hypothalamus. In addition, a great number of prolactin-releasing or -modifying factors have been identified, including estrogen, serotonin (particularly through the 5-HT_2 and 5-HT_3 receptors), norepinephrine, opioids, TRH, T_4, histamine, glutamate, cortisol, CRH, and oxytocin, with interaction effects possible. For example, estrogen may promote the serotonin-stimulated release of prolactin.

Prolactin is primarily involved in reproductive functions but also in weight and metabolism control. During maturation, prolactin secretion participates in gonadal development, whereas, in adults, prolactin contributes to the regulation of the behavioral aspects of reproduction and infant care, including estrogen-dependent sexual receptivity and breastfeeding. In female rats, prolactin secretion is strongly stimulated with exposure to pups. In women, basal prolactin levels are elevated in the postpartum period before weaning, and prolactin release is stimulated by suckling. Hyperprolactinemia is associated with low testosterone in men and reduced libido in men and women. In rodents, prolactin is increased along with corticosterone in response to such stressful stimuli as immobilization, hypoglycemia, surgery, and cold exposure and may be specifically associated with the use of passive coping in the face of a stressor. Prolactin promotes various stress-related behaviors in rats, depending on the condition, such as increasing object-directed exploration while decreasing other exploration.

Hyperprolactinemic patients often complain of depression, decreased libido, stress intolerance, anxiety, and increased irritability. These behavioral symptoms usually resolve in parallel with decrements in serum prolactin when surgical or pharmacologic treatments are used. In psychotic patients, prolactin concentrations and prolactin-related sexual disturbances have been positively correlated with the severity of tardive dyskinesia. Prolactin levels are also positively correlated with negative symptoms.

HYPOTHALAMIC–PITUITARY–THYROID AXIS

Thyroid hormones are involved in the regulation of nearly every organ system, particularly those integral to the metabolism of food and the regulation of temperature, and are responsible for the optimal development and function of all body tissues. Moreover, rates of secretion and metabolism of all other major hormones (cortisol, gonadal hormones, and insulin) and catecholamines depend on thyroid status. The thyroid gland secretes two thyroid hormones: T_3 and T_4. T_3 is the more potent of the two, and most of the T_3 circulating in the blood is created by the peripheral metabolism of T_4. The brain relies on its own conversion of T_4 to T_3 rather than on circulating T_3. The hypothalamus secretes TRH into the capillaries of the pituitary portal venous system, and the pituitary responds with the synthesis and secretion of TSH, which stimulates thyroid cells. Negative feedback regulation occurs when T_3 and T_4 act in the pituitary and hypothalamus to inhibit TSH and TRH, respectively. Finally, a corticotropin-release–inhibiting factor (CRIF) has been identified in the rat that inhibits the synthesis and secretion of ACTH. This peptide, prepro-TRH 178–199, is derived from the prohormone TRH and may play a role in integrating the regulation of the HPA and HPT axes. The expression of genes that encode for thyroid hormone receptors, such as TRα and TRβ*,* plays a major role in the regulation of neuronal differentiation and the action of immediate early genes. As in the case of steroids, thyroid hormones regulate the transcription of a variety of genes through binding to thyroid response elements (TREs) in regulatory sequence elements and may also be critical to understanding sex-dependent neuronal signaling.

There is general agreement that central noradrenergic systems are primarily stimulatory to TSH secretion and that central dopamine neurons inhibit TSH release. Thyroid hormones, in turn, are important regulators of central adrenoceptor function, generally decreasing presynaptic noradrenaline release and increasing postsynaptic β-adrenergic receptor number. Hypothyroidism is conversely associated with a decreased β-receptor number. Changes in serotonin function are also apparent, with T_3 increasing 5-HT in the frontal cortex and inducing downregulation of 5-HT_{1A} autoreceptors. These changes in neurotransmitter release and receptors in response to thyroid hormones parallel the alteration in α- and β-receptor sensitivity associated with pharmacologic and electroconvulsive antidepressant treatments and may explain the therapeutic efficacy of supplemental thyroid hormone in treatment-resistant depression. Alternatively, the therapeutic benefit may be secondary to the alteration of gene expression and remodeling of synaptic connectivity. In addition to its prime endocrine function, TRH has direct effects on neuronal excitability, behavior, and neurotransmitter regulation, particularly on central cholinergic systems located in the septohippocampal band and on mesolimbic and nigrostriatal dopamine systems. In lower animals, TRH possesses mild stimulant properties. Initial reports of its mood-elevating effects in healthy human subjects led to a number of projects investigating its short- and long-term antidepressant effects in clinical populations. Despite some initial enthusiasm, the degree of mood alteration does not seem to be great, nor is its occurrence reliable.

Given these observations, it is not surprising that alterations in behavioral function have been observed in patients with primary thyroid gland dysfunction, beginning with the earliest reports in the medical literature. It has been noted that thyroid disorders may induce virtually any psychiatric symptom or syndrome, although regular associations of specific syndromes and thyroid conditions are not consistently found. Hyperthyroidism is commonly associated with fatigue, irritability, insomnia, anxiety, restlessness, weight loss, and emotional lability; marked impairment in concentration and memory may also be evident. Such states can progress into delirium or mania, or they can be episodic in nature. On occasion, a true psychosis develops, with paranoia being a particularly common presenting feature. In some cases, psychomotor retardation, apathy, and withdrawal rather than agitation and anxiety are the presenting features. Symptoms of mania also have been reported after rapid normalization of thyroid status in hypothyroid individuals and may covary with thyroid levels in individuals with episodic endocrine dysfunction. In general, behavioral abnormalities resolve with normalization of thyroid function and are responsive symptomatically to traditional psychopharmacologic regimens. Long-term residual complaints of fatigue, cognitive impairment, and emotional distress have been reported in some individuals even after remission of the precipitating thyroid dysfunction, and there is evidence for an increased risk of dementia over time. Caution should be exerted, however, regarding the use of MAOIs or tricyclic antidepressant (TCA) medications in hyperthyroid states because of possible synergistic cardiotoxicity. In several case reports, haloperidol (Haldol) has been linked to increasing thyrotoxicity, and hyperthyroidism has been associated with an enhancement of the neurotoxic effects of antipsychotic medications.

The psychiatric symptoms of chronic hypothyroidism are generally well recognized. Most classically, fatigue, decreased libido, memory impairment, and irritability are noted, but a true secondary psychotic disorder or dementia-like state also can develop. In milder, subclinical states of hypothyroidism, the absence of gross signs accompanying endocrine dysfunction may result in its being overlooked as a possible cause of a mental disorder. Accordingly, the evaluation of the basal TSH concentration or the TSH response to TRH infusion is necessary to arrive at the proper diagnosis. Hypothyroidism impairs neurogenesis in the hippocampus, and this impairment is associated with rat models of depression. Thyroid hormone administration corrects the impairment and reverses the depression.

A blunted response of TSH to TRH infusion has been found in a significant percentage of patients with a variety of disorders, including eating disorders, panic disorder, alcoholism, schizophrenia, and, most commonly, MDD, and probably reflects a transient hyperthyroxinemia. No evidence of TRH hypersecretion has been shown. Large-scale studies suggest that such subjects are, in fact, euthyroid, and the predictive sensitivity of the test is low. Thyroid autoimmunity may play a role in some psychiatric illnesses. Antithyroid antibodies are found more frequently in women with depression than in control subjects and in patients with bipolar disorder and may contribute to relative treatment resistance as well as to postpartum behavioral disturbance. The frequency of thyroid antibodies may also be higher in certain nonpsychiatric diseases with prominent psychiatric symptoms, such as fibromyalgia and rheumatoid arthritis. Patients with major depression also have been found to have low levels of CSF transthyretin, a protein involved in thyroid transport. Basal T_3 level has been inversely related to time to episode recurrence, and basal TSH positively correlated with the overall severity of depressed mood in an unselected inpatient population. Depressed patients who show an improvement in mood after one night of sleep deprivation also appear to have lower T_3 uptake at baseline and greater nocturnal TSH release. Thyroid receptor alterations can produce symptoms of ADHD in animals, and genetic resistance to thyroid hormone has been associated with ADHD symptoms in humans. Thyroid receptor coactivators also may play a role in increased vulnerability to a number of psychiatric disorders, including depression and psychosis.

Most antidepressant therapies have some influence on thyroid concentrations at baseline; T_4 and T_3 concentrations have been correlated with antidepressant response, as have antidepressant-induced changes in thyroid hormones as well as changes induced by electroconvulsive therapy (ECT). Lithium increases antithyroid antibodies and inhibits iodine uptake into the thyroid, iodination of tyrosine, the release of T_3 and T_4 from the thyroid, and peripheral breakdown of thyroid hormones. It also regulates *TR* gene expression, blocks the thyroid-stimulating effects of TSH through interference with adenylate cyclase, and may, in certain circumstances, precipitate rebound thyrotoxicosis. Approximately 30% of patients receiving lithium have an elevated TSH level during treatment, and approximately one-sixth of these patients go on to develop frank hypothyroidism. Attention to subtle alteration in thyroid status induced by lithium treatment is important in the clinical evaluation of symptomatic complaints, such as fatigue, memory impairment, and anhedonia; a specific association between lower serum T_4 and mood instability during lithium maintenance suggests even subclinical changes may be clinically relevant. Carbamazepine, an anticonvulsant shown to have antimanic properties akin to lithium, also decreases peripheral thyroid hormone concentrations while increasing TSH. Administration of T_3 accelerates clinical response to TCAs and is sometimes helpful in patients with treatment-resistant depression, whereas adjunctive T_4 contributes to decreasing cycling in patients with rapid-cycling bipolar I disorder. Supraphysiologic dosing is sometimes required, with 200 to 500 μg per day being given as an adjunct in treatment-resistant depression and intractable bipolar disorder. Administration of mirtazapine (Remeron), a mixed-action antidepressant, increases free T_3 levels and decreases free T_4. Higher T_3 concentrations predicted improvements in depression in one study.

PARATHYROID HORMONE

Parathyroid hormone (PTH) was originally isolated as an endocrine factor having effects on bone, gut, and kidney and contributing to calcium and phosphorus homeostasis. However, the frequent and often profound neuropsychiatric changes that can result from altered parathyroid gland function are consistent with other central actions of PTH that have been described in recent years. Hyperparathyroidism can cause lethargy, stupor, coma, depression, delirium, psychosis, primarily visual hallucinations, or anxiety. Hypoparathyroidism can cause cognitive impairment, psychosis, depression, or anxiety by alterations in calcium and magnesium levels. PTH administration can impair the active uptake and release of norepinephrine and dopamine and result in adrenergic-like effects (not blocked by β-adrenergic antagonists), learning and memory problems, and a state of hyperalgesia.

Lithium treatment can raise the concentrations of serum calcium and may increase PTH over a period of months to years by direct stimulation of PTH secretion and through a shift in the set point for inhibition of PTH secretion by calcium. When such effects are associated with somatic or behavioral changes, discontinuation of lithium should result in rapid symptomatic improvement. When this does not happen, a parathyroid adenoma is sometimes discovered fortuitously. Primary hyperparathyroidism most commonly occurs secondary to a single parathyroid adenoma, the removal of which almost invariably results in lysis of behavioral symptoms, regardless of severity or chronicity. Animal studies suggest that long-term lithium administration can stimulate the development of extant parathyroid tumors but does not induce tumors in normal parathyroid tissue. Thus, reinstitution of lithium treatment after surgical removal of the tumor should be possible.

GROWTH HORMONE

Somatotropin or GH, a hormone required for normal growth, is synthesized and released by the anterior pituitary gland. Dopamine, serotonin acting at the 5-HT$_{1D}$ receptor, and norepinephrine acting at the α_2-adrenergic receptor appear to play a role in its release. GH acutely stimulates lipolysis and ketogenesis, which are important in the adaptation to stress, and prevents hypoglycemia. Most psychiatric studies of the regulation of GH have used strategies similar to those described for prolactin. Accordingly, studies of GH response to various provocative stimuli, such as GHRH or psychotherapeutic drugs, have been seen as a means to evaluate central neurotransmitter function. Augmentation of GH secretion in response to GHRH, LH-releasing hormone (LHRH), or TRH in patients with schizophrenia or dementia of the Alzheimer type has been interpreted as reflecting an alteration in catecholamine and, possibly, prostaglandin regulation, which facilitate the secretion of human GH. In general, however, there is a large variation in GH response to GHRH; a blunted response has been variably linked to the length of illness, presence of negative symptoms, and platelet MAO activity, but the validity of the conclusions drawn from this test is controversial.

The stress responsiveness of somatotrophs is well established but species dependent, with increases in circulating GH noted in humans and inhibition of secretion noted in rodents. In humans, the direction of the GH stress response may depend on the persistence of the stressor. GH appears to be relatively more responsive to exercise and hypoglycemic stress than to psychological stress. However, GH has been reported to increase in response to psychological stress in anxious subjects, perhaps due to hyperactivity of the noradrenergic system. Case reports have documented reversible GH deficiencies and marked growth retardation and delay of puberty secondary to a stressful experience. Administration of GH to individuals with GH deficiency has a beneficial effect on cognitive function in addition to its more obvious somatic effects. A significant proportion of adult-onset patients with GH deficiency are depressed, and GH therapy significantly improves their depression scores. Some prepubertal as well as adult patients with diagnoses of MDD show hyposecretion of GHRH during an insulin tolerance test, a deficit that has been interpreted as reflecting alterations in cholinergic and serotonergic mechanisms. Blunted response to 5-HT_{1D} agonists also has been found. Panic disorder patients may have a blunted GH response to clonidine (Catapres), an α_2-adrenergic agonist, that does not normalize with antidepressant treatment. A number of GH abnormalities also have been noted in patients with anorexia nervosa, but secondary factors, such as weight loss, may be responsible for alterations in endocrine release in depression and eating disorders. At least one study has reported that GHRH stimulates food consumption in patients with anorexia nervosa and attenuates elevated food consumption in patients with bulimia. Administration of GH to elderly men results in an increase in lean body mass, but controlled trials have been unable to replicate anecdotal reports of improved mental clarity, muscle strength, or vigor. Recent evidence indicates that a novel GH secretagogue (GHS), ghrelin, may represent an important alternative regulatory influence over food intake and sleep patterns.

Many GHS can be administered orally, in contrast to GH, which must be injected on a daily basis. In addition, they may be used to fine-tune body concentrations in a way that daily injections cannot. GHB, a potent GHS, has been used to mimic the physiologic secretory pattern of GH and as a way to increase SWS in patients with fibromyalgia and, secondarily, to a reduction in pain and fatigue. The use of the GHS GHB also has been used by bodybuilders as a way of increasing muscle mass, but respiratory depression and sedation, and dependence with severe withdrawal effects may result.

SOMATOSTATIN

Somatostatin (SRIF) is a hypothalamic tetradecapeptide that is located principally in the nerve endings of the median eminence and in neurosecretory neurons located in the PVN. SRIF inhibits the anterior pituitary secretion of ACTH, thyrotropin, GH, and prolactin, alters the release of catecholamine neurotransmitters, and stimulates serotonin release. A number of receptor subtypes have been cloned, and receptor-specific ligands have been developed. SRIF was so named because of its action in inhibiting the release of immunoreactive GH, a function that is subserved by SRIF-2 receptors.

In rats, SRIF delays the extinction of active avoidance behavior and antagonizes amnesia induced by electric shock. Alterations in the concentration of SRIF have been associated with a number of conditions in which cognitive dysfunction is present, including Huntington disease, Parkinson disease, multiple sclerosis, and Alzheimer disease. Decreases in SRIF are highly correlated with decreases in acetylcholinesterase, suggesting a close relationship between the cholinergic and somatostatinergic systems. Decreased concentrations of SRIF in the CSF are inconsistently found in patients with depression, and central injection of SRIF in rats causes decreased slow-wave and rapid-eye movement sleep, altered appetite, and locomotor activity, impaired cognition, and decreased sensitivity to pain. Early stressful experiences also have been related to sustained elevations of CRH and somatostatin in the CSF of adult primates. Altered somatostatin concentrations have been reported in a number of other illnesses and with treatment, but the physiologic relevance of these changes is still unclear. Somatostatin receptors 2 and 4 have been specifically implicated in SUDs.

ARGININE VASOPRESSIN

AVP (or antidiuretic hormone [ADH]) is a posterior pituitary hormone that maintains plasma osmolarity through the regulation of renal water excretion and stimulates hepatic glycogenolysis. AVP release is triggered by pain, emotional stress, dehydration, increased plasma osmolarity, or decreases in blood volume and acts synergistically with CRH to control ACTH release. AVP potentiates the stimulatory effect of CRF. An AVP receptor antagonist blocks ACTH release, norepinephrine release, and hyperthermic response to stress and attenuates some stress-related behaviors. AVP receptor blockade does not impair motor or cognitive processes or produce tolerance, as do many other anxiolytics.

Animal and normal human studies of AVP administration (or longer-acting synthetic analog compounds) have indicated that the hormone may enhance the consolidation and retrieval of memory, particularly that associated with aversive learning. AVP has been shown to prevent the loss of tolerance to the incoordinating, sedative–hypnotic, and hypothermic effects of alcohol after cessation of ingestion and to delay the loss of sexual behavior after castration.

Altered AVP function has been reported in depression and in eating disorders. Anorexic and bulimic patients show hypersecretion of centrally directed AVP, and patients with bulimia nervosa or depression may have an attenuated AVP response to hypertonic saline. Vasopressin delays the extinction of behaviors acquired during aversive conditioning and may be related to an obsessional preoccupation with the aversive consequences of eating and weight gain. An inverse relationship between AVP concentration and motor activity in depression and an increased number of vasopressin and oxytocin neurons also have been reported in the hypothalamus of depressed patients. Although dexamethasone suppression of ACTH and cortisol release is attenuated in depressed patients, suppression of ACTH and cortisol release in response to vasopressin is not.

Profound alterations in fluid ingestion and excretion have been observed in psychiatric patients. Polydipsia occurs in 10% to 15% of hospitalized psychiatric patients and is unrelated to diagnosis; in many cases, the syndrome is secondary to inappropriate secretion of AVP, which occurs as a feature of the altered behavioral state itself and resolves with treatment or, conversely, is precipitated by a variety of antidepressant or antipsychotic agents. In contrast, fluoxetine treatment for depression decreases the CSF vasopressin levels. Alprazolam, an inhibitor of CRH secretion, inhibits the vasopressin-stimulated release of ACTH and cortisol.

Several subtypes of AVP receptors and been discovered, including $V_{1a}R$, $V_{1b}R$, and V_2R. $V_{1a}R$ knockout (KO) mice show impairments in spatial learning, and both $V_{1a}R$ and $V_{1b}R$ KO mice show impairments in prepulse inhibition and social behavior. AVP $V_{1a}R$ KO mice have reduced social recognition. With transgenic technology, this gene, when introduced into the nonmonogamous mouse, increased a behavior associated with monogamy. AVP antagonists prevent bonding in the monogamous prairie vole. Animal studies show that centrally released AVP produces anxiogenic and depression-like actions, including generating passive coping strategies in stressful situations. That AVP modulates the stress response is shown by the ability of the $V_{1a/b}$ receptor antagonist to normalize the usually abnormal dexamethasone–CRH test in rats bred for high anxiety (HAB). Serotonin increases the release of AVP, but SSRIs reduce AVP secretion or normalize AVP overactivity in HAB rats, resulting in the normalization of the dexamethasone–CRH test and less depression-like behavior in response to stress. Animal studies of the role of AVP in social behavior suggest that receptor variations may be involved in such disorders as schizophrenia and autism. In fact, several studies have found an association between an AVP receptor gene and autism and social behavior in a nonclinical population.

OXYTOCIN

Oxytocin is a posterior pituitary hormone that is involved in osmoregulation, the milk ejection reflex, food intake, and female maternal and sexual behaviors and has many effects reciprocal to those of vasopressin. Convergent evidence, using a range of methodologies, indicates that oxytocin inhibits food and sodium intake. Oxytocin binding in the hypothalamus is increased by estrogen and glucocorticoids and in estrogen-primed women. Oxytocin also can act as a neuromodulator of limbic dopamine concentrations and thus may be involved in the adaptation to substances of abuse, and it can act as a mediator of the effect of CRH on ACTH. Oxytocin has anxiolytic activity.

Many of oxytocin's behavioral effects are affiliative, and it promotes a variety of reproductive (grooming, arousal, lordosis, orgasm, nesting, and birthing) and maternal behaviors (breastfeeding and mother–infant bonding). Infusion of oxytocin in female subjects of monogamous species facilitates pair bonding in the absence of mating, and administration of an oxytocin antagonist prevents pair bonding. It has been called the *amnestic neuropeptide* owing to its ability to attenuate memory consolidation and retrieval. Patients with autism and anorexia have been reported to have reduced levels of oxytocin, but depressed patients do not differ from healthy controls in basal levels. Autistic children also do not show the expected increase in oxytocin with age. In adults with autism, oxytocin infusion reduces repetitive behaviors. In one animal study, chronic administration of phencyclidine (PCP), which lowers hypothalamic oxytocin, also decreased social interaction. An association between an oxytocin receptor gene and autism has recently been identified. Oxytocin interacts with the mesolimbic dopamine system and is believed to facilitate the acquisition of drug use disorders, particularly of some of the "party drugs," such as 3,4 methylenedioxymethamphetamine and gamma hydroxybutyrate (GHB), which are often used to promote social behavior. Variants in the oxytocin receptor gene, particularly rs53576, have been shown to correlate with measures of empathy and exogenous administration of oxytocin to increase affiliative behavior in humans.

NEUROPEPTIDE Y

NPY is closely linked with stress response and with the action of a number of steroid hormones. NPY is widely distributed throughout the CNS and is one of the most conserved peptides in evolution, suggesting an important role in the regulation of basic physiologic functions, including learning and memory. NPY and NPY-related peptides bind to at least five receptors, which are widely distributed but relatively concentrated in the hypothalamus, the hippocampus, and several other limbic regions. NPY is synthesized in the arcuate nucleus of the hypothalamus. Immunoreactive NPY is found in the serotonin-containing raphe nucleus, and it has been implicated in the modulation of emotional processing. Marijuana use appears to elevate the expression of NPY-1 receptor mRNA levels, perhaps explaining some of the drug's effects. NPY has been found to increase feeding, particularly carbohydrate ingestion, and to counteract leptin effects in a variety of animal models, possibly through its effects on inhibition of fear and anxiety circuits. It has a mutually inhibitory relationship with insulin, and its release is stimulated by stress and corticosteroids and associated with norepinephrine release.

NPY has been studied for its potential anxiolytic, antinociceptive, antihypertensive, and memory-enhancing effects and for a possible role in seizure disorders, schizophrenia, and depression. Treatments for depression, such as some antidepressants, lithium (Eskalith), and ECT, increase NPY concentrations in a number of brain areas in rats, while significantly low levels of NPY have been found in the temporal cortices of patients with schizophrenia. Greater NPY release during stress is associated with less psychological distress in humans and the modulation of the activity of GABAergic neurons. Neuroactive steroids, in turn, regulate NPY transmission. NPY appears to influence alcohol consumption and alcohol and morphine withdrawal effects. Reduced levels of NPY are found in discrete brain regions of alcohol-preferring rats, and polymorphisms of the NPY gene are associated with alcohol dependence and with alcohol withdrawal seizures in humans. Polymorphisms in the genes for the NPY peptide and for a promoter region also have been found to be associated with depression in some individuals.

GALANIN

Galanin is a 29 amino acid inhibitory peptide that is stimulated in a coordinated fashion with gonadal steroid release. Its documented actions include increased release of GH and inhibition of insulin release, locus coeruleus noradrenergic firing, and acetylcholine release, as well as impairment of memory, control of food intake, nociception, energy expenditure, and the mediation of some emotional responses. Studies are underway using galanin receptor agonists and antagonists and galanin receptor KO mice to explore its role in mediating anxiety- and depression-related behavior and in decreasing opiate withdrawal and drug dependence overall. Inhibition of GALR1 and GALR3 and activation of GALR2 appear to attenuate depressive symptoms, suggesting galanin as a possible novel antidepressant drug target. Spexin and kisspeptin share an ancestral origin with galanin and may be similarly involved in emotion and pain regulation, as well as in energy homeostasis.

INSULIN

Insulin is a protein hormone secreted by the β cells of the pancreas in response to elevations of glucose and amino acids; insulin receptors occur in high density in the hippocampus and are believed to help neurons to metabolize glucose by controlling the transport across cell membranes. Some atypical antipsychotics impair response to insulin and raise blood glucose, increasing the risk of developing diabetes (see Metabolic Syndrome). Psychotic stress itself may impair insulin sensitivity. Those antidepressants that predominantly increase catecholamine activity, such as many of the TCAs, also reduce sensitivity to insulin. Those that increase serotonergic function, such as the SSRIs, increase sensitivity to insulin and are preferred in the treatment of depression comorbid with diabetes. However, insulin may play a more active role in psychiatric symptoms as increasing evidence indicates that insulin may be integrally involved in learning, memory, and mood. Depression is frequent in patients with diabetes, as are indices of impaired hormonal response to stress, but it is not known whether these findings represent direct effects of the disease or are secondary, nonspecific effects. Higher fasting insulin levels also have been associated with better psychopathology scores in schizophrenic patients.

LEPTIN/GHRELIN/GLUCAGON-LIKE PEPTIDE

Leptin is a protein hormone synthesized and secreted in a pulsatile fashion by adipose tissue and involved in the regulation of food intake. Obesity is associated with leptin resistance, principally its metabolic actions, because sympathetic effects are preserved. Leptin also affects the HPG axis, inhibits insulin-induced steroidogenesis and human chorionic gonadotropin-induced testosterone secretion, and may play a role in menstruation, pregnancy, lactation, puberty, and amenorrhea due to weight loss in anorexia nervosa. Leptin stimulates hematopoiesis, T-cell activation, phagocytosis, and cytokine production and decreases susceptibility to infection. Mediators of leptin action include orexigenic neuropeptides, such as NPY, galanin, and galanin-like peptide, melanin-concentrating hormone (MCH), and anorexigenic neuropeptides, such as CRH and α-MSH hormone, as well as gut microbiota. Weight gain produced by some atypical antipsychotics may be mediated in part through increases in leptin. Patients with major depression have lower leptin levels, and these levels are inversely correlated with depression severity. However, higher leptin levels have been found in subjects exposed to trauma with hyperarousal, and these levels are related to hypervigilance. Because these relationships were correlations, the causal relationship is not known. Increased leptin may be an attempt at adaptation, as postulated for changes in opioids and neuroactive steroids in PTSD.

Ghrelin is an orexigenic peptide released in the stomach that is involved in energy regulation. Studies have shown that its release is sensitive to a number of physical and psychological stresses and, as such, may modulate coping mechanisms and central reward pathways through brain-mediated receptor signaling. Increased blood levels have been reported in both restricting and binge-eating/purging subtypes of anorexia nervosa.

Glucagon-like peptide (GLP-1) is another gut peptide involved in appetite suppression that has also shown some effects on central control of stress response and emotional regulation.

CHOLECYSTOKININ

Cholecystokinin (CCK) is a peptide neurotransmitter originally isolated from the gut. In addition to its presence in the pancreas and the gastrointestinal (GI) tract, CCK is found in high concentrations in the cerebral cortex, limbic system, and hypothalamus. CCK is involved in the regulation of such behavioral functions as inhibition of intake of solid and liquid food, production of satiety, pain relief (probably from modulation of the endogenous opioid system), cardiovascular and respiratory function, neurotoxicity, and seizures, sexual and reproductive behaviors, and memory. Of the two identified receptor subtypes, CCK type A (CCK-A) is found primarily in the periphery and in some discrete brain areas, whereas CCK type B (CCK-B) is plentiful in the brain. The primary form of CCK, CCK-8S (a sulfated octapeptide), coexists with dopamine in the VTA and substantia nigra, and its interactions with dopamine are context and location specific. CCK stimulates the synthesis of NGF and plays a neuroprotective role. CCK also modulates HPA activity.

Of particular interest to psychiatry is the colocalization of CCK with dopamine in mesolimbic and mesocortical, but not in nigrostriatal systems. CCK contributes to the modulation of dopamine-mediated behavior and might be dysregulated in psychiatric syndromes thought to involve altered dopamine transmission. CCK-A receptor antagonists have been proposed for the treatment of schizophrenia, and initial evidence suggests that CCK agonists may be useful for decreasing the severity of parkinsonian symptoms. Cerulein, a mixed CCK-A and CCK-B agonist has weak, neuroleptic-like effects on prepulse inhibition in schizophrenic patients. Evidence is stronger for a role of CCK-B receptor antagonists in the treatment of anxiety. CCK has anxiogenic effects in some animal models, and several small-scale human studies have demonstrated that the administration of CCK-4 or pentagastrin can induce panic attacks and can increase neurosteroid release in a significant percentage of healthy volunteers as well as in anxiety disorder patients, even in the absence of arousal or environmental stress. Panic attacks in subjects with preexisting panic disorder can be elicited at doses of CCK-4 that do not reliably induce panic in healthy subjects, indicating enhanced sensitivity. Not only can selective CCK-B antagonists completely abolish the anxiogenic effects of CCK-4, but in an animal model of anxiety used to evaluate the efficacy of benzodiazepines, CCK-B antagonists also demonstrated independent anxiolytic properties. However, CCK overactivity may be involved in submissive behavior in animal models, and CCK receptor expression is higher in suicide victims.

Modulation of mesolimbic dopamine-related behavior, including exploratory and rewarded behaviors, may underlie CCK’s effects on substance use. Central CCK activity has been linked with a preference for drugs of abuse, such as cocaine or alcohol, and polymorphisms of the CCK gene may be one of the risk factors for smoking. This suggests a potential role for CCK receptor antagonists in the treatment of drug dependence.

GASTRIN AND GASTRIN-RELEASING PEPTIDE

Gastrin is a peptide hormone closely related to CCK that stimulates the secretion of gastric acid by the stomach. Pentagastrin is a synthetic polypeptide with effects similar to those of gastrin. It is a CCK agonist and produces anxiety and panic in patients with anxiety disorders and to a lesser extent in those without anxiety disorders. It increases ACTH and cortisol release. Gastrin-releasing peptide (GRP), as its name implies, stimulates gastrin release but has a number of other actions, including interacting with GABA, dopamine, and GRs. GRP appears to enhance memory storage, and a GRP receptor antagonist impairs emotionally motivated memory tasks in rats. GRP-receptor-deficient mice show increased locomotor activity and changes in social behavior.

NEUROTENSIN

Neurotensin is a tridecapeptide that appears to play a role in neuroendocrine regulation and coordination as a signaling molecule. Gonadal and adrenal steroids and thyroid hormones alter neurotensin levels in the hypothalamus, preoptic area, and arcuate nucleus. Neurotensin has a close neuroanatomical relation with serotonin and dopaminergic pathways and is involved in the control of anterior pituitary activity, stimulating the release of prolactin and TSH, as well as in the regulation of a subpopulation of serotonergic neurons in the dorsal raphe and frontal cortex and GABAergic and glutamatergic neurons. Stimulation of serotonin neurons may be responsible for its analgesic effects and reduction of the stress response, whereas the effects on dopamine suggest a possible antipsychotic role. Subgroups of drug-free schizophrenic patients have low neurotensin CSF concentrations and altered neurotensin receptor binding in the entorhinal cortex. Psychotogenic drugs (e.g., methamphetamine) inhibit the release of striatal neurotensin via an inhibitory effect of the dopamine type 1 (D_1) receptor. Most antipsychotic drugs increase neurotensin concentrations in the nucleus accumbens and caudate nucleus; schizophrenic patients with decreased CSF neurotensin show an increase compared to baseline values after antipsychotic drug treatment and clinical improvement. Because of neurotensin’s association with the nigrostriatal dopamine and the serotonin systems, it is suspected of playing a role in movement disorders caused by antipsychotic drugs. Central administration of neurotensin in rats produces motor effects seen in animal models of parkinsonian and dystonic reactions (catalepsy) and tardive dyskinesia. Neurotensin may exert an antipsychotic action through intramembrane receptor interactions that reduce the affinity of the dopamine type 2 (D_2) agonist binding site. Overexpression of neurotensin 1 receptors in rats results in decreased activation of the mesolimbic dopamine system, similar to that produced by atypical antipsychotics but without changing prepulse inhibition or locomotor behavior, as atypical antipsychotic drugs do. This suggests that neurotensin receptor agonists may be candidates for the treatment of psychosis and attenuate dopamine-induced motor behaviors.

Involvement in the development of drug dependence also has been hypothesized. Blocking the effects of neurotensin with antiserum or a receptor antagonist enhances dopamine release in the nucleus accumbens, and neurotensin itself blocks stimulant-induced motor activity. However, doses that block hyperlocomotion do not attenuate the self-administration of cocaine and even enhance conditioned place preference, an animal model of rewarding effects. Neurotensin's modulating effects on dopamine activity may depend on stimulus intensity, enhancing the rewarding properties of a subthreshold stimulus from intracranial self-stimulation (ICSS), as do psychostimulants, but decreasing maximal stimulation rate, as do antipsychotic drugs. Acute administration of stimulants increases neurotensin in the nucleus accumbens, but with chronic administration, levels return to normal. Ibogaine, a hallucinogen used in indigenous religious ceremonies, has been shown to interrupt cocaine and methamphetamine abuse in patients and may act similarly to ICSS, increasing neurotensin concentrations in the nucleus accumbens when given alone but attenuating cocaine-induced increases in neurotensin. Neurotensin also mimics many of the effects of alcohol, chronic alcohol downregulates neurotensin receptors, and lower concentrations of neurotensin have been found in the frontal cortices of alcohol-preferring rats.

FUTURE DIRECTIONS: ENDOCRINE VARIABLES IN THE DIAGNOSIS AND TREATMENT OF PSYCHIATRIC DISORDERS

Although it is clear that alterations in endocrine regulation are involved in the pathophysiology and treatment responses of many psychiatric disorders, incorporating these findings into clinical diagnostic assessment, and decision making remains problematic. Questions about state/trait differences and the role of confounding variables plague nearly all observations, and most of the findings to date are based on small numbers of subjects studied under experimental conditions. Large-scale longitudinal or cost-effectiveness studies are rare, despite indications that baseline alterations in glucocorticoid regulation and thyroid status (two of the best-studied abnormalities) may actually be useful in subtyping psychiatric disorders and in the prediction of outcome. Although the dexamethasone suppression test (DST) was perhaps prematurely put forward as a diagnostic aid, the fact remains that alterations in HPA/stress regulation underlie a number of psychiatric diagnoses and may serve as complementary independent variables in defining treatment response and course of illness to the classical behavioral categories that have thus far defined current practice. Studying genetic polymorphisms in factors regulating hormonal response may help us better understand the influence of hormonal variability on the illness and also possible underlying differences in the nature of the illness reflected in these genetic subtypes. The incorporation of endocrine variables into the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) would have been premature, but the routine assessment of endocrine status in clinical trials and epidemiologic surveys is long overdue. Incorporation of biologic variables into psychiatric diagnosis and treatment decision making requires prospective proof of concept studies that unequivocally demonstrate the clinical superiority of such assessments over that of behavioral variables alone. Unfortunately, no such data exist at the present time.

FURTHER READINGS

Barabas K, Godó S, Lengyel F, Ernszt D, Pál J, Ábrahám IM. Rapid non-classical effects of steroids on the membrane receptor dynamics and downstream signaling in neurons. *Horm Behav*. 2018;104:183–191.

Barchi-Ferreira AM, Osório FL. Associations between oxytocin and empathy in humans: a systematic literature review. *Psychoneuroendocrinology*. 2021;129:105268.

Baumbach JL, Zovkic IB. Hormone-epigenome interactions in behavioral regulation. *Horm Behav*. 2020;118:104680.

Bernard V, Young J, Binart N. Prolactin-a pleiotropic factor in health and disease. *Nat Rev Endocrinol*. 2019;15(6):356–365.

Bodnar RJ. Endogenous opiates and behavior: 2019. *Peptides*. 2021;141:170547.

Bridges RS. The behavioral neuroendocrinology of maternal behavior: past accomplishments and future directions. *Horm Behav*. 2020;120:104662.

Cipolla-Neto J, Amaral FGD. Melatonin as a hormone: new physiological and clinical insights. *Endocr Rev*. 2018;39(6):990–1028.

Comeras LB, Herzog H, Tasan RO. Neuropeptides at the crossroad of fear and hunger: a special focus on neuropeptide Y. *Ann N Y Acad Sci*. 2019;1455(1):59–80.

D’Amico D, Amestoy ME, Fiocco AJ. The association between allostatic load and cognitive function: a systematic and meta-analytic review. *Psychoneuroendocrinology*. 2020;121:104849.

Del Rio JP, Alliende MI, Molina N, Serrano FG, Molina S, Vigil P. Steroid hormones and their action in women's brains: the importance of hormonal balance. *Front Public Health*. 2018;6:141.

Dwyer JB, Aftab A, Radhakrishnan R, et al; APA Council of Research Task Force on Novel Biomarkers and Treatments. Hormonal treatments for major depressive disorder: state of the art. *Am J Psychiatry*. 2020;177(8):686–705.

Eck SR, Bangasser DA. The effects of early life stress on motivated behaviors: a role for gonadal hormones. *Neurosci Biobehav Rev*. 2020;119:86–100.

Eslami-Amirabadi M, Sajjadi SA. The relation between thyroid regulation and impaired cognition/behavior: an integrative review. *J Neuroendocrinol*. 2021;33(3):e12948.

Fetissov SO, Hökfelt T. On the origin of eating disorders: altered signaling between gut microbiota, adaptive immunity and the brain melanocortin system regulating feeding behavior. *Curr Opin Pharmacol*. 2019;48:82–91.

Fritz EM, Singewald N, De Bundel D. The good, the bad and the unknown aspects of ghrelin in stress coping and stress-related psychiatric disorders. *Front Synaptic Neurosci*. 2020;12:594484.

Genders SG, Scheller KJ, Djouma E. Neuropeptide modulation of addiction: focus on galanin. *Neurosci Biobehav Rev*. 2020;110:133–149.

Geniole SN, Bird BM, McVittie JS, Purcell RB, Archer J, Carré JM. Is testosterone linked to human aggression? A meta-analytic examination of the relationship between baseline, dynamic, and manipulated testosterone on human aggression. *Horm Behav*. 2020;123:104644.

Gold PW. Endocrine factors in key structural and intracellular changes in depression. *Trends Endocrinol Metab*. 2021;32(4):212–223.

Juruena MF, Cleare AJ, Young AH. Neuroendocrine stress system in bipolar disorder. *Curr Top Behav Neurosci.* 2021;48:149–171.

McEwen BS. From serendipity to clinical relevance: how clinical psychology and neuroscience converged to illuminate psychoneuroendocrinology. *Psychoneuroendocrinology.* 2019;105:36–43.

McEwen BS. Hormones and behavior and the integration of brain-body science. *Horm Behav*. 2020;119:104619.

Mourtzi N, Sertedaki A, Charmandari E. Glucocorticoid signaling and epigenetic alterations in stress-related disorders. *Int J Mol Sci.* 2021;22(11):5964.

Robinson KJ, Bosch OJ, Levkowitz G, Busch KE, Jarman AP, Ludwig M. Social creatures: model animal systems for studying the neuroendocrine mechanisms of social behavior. *J Neuroendocrinol*. 2019;31(12):e12807.

Robinson SL, Thiele TE. A role for the neuropeptide somatostatin in the neurobiology of behaviors associated with substance abuse and affective disorders. *Neuropharmacology*. 2020;167:107983.

Soria V, González-Rodríguez A, Huerta-Ramos E, et al; PNECAT Group. Targeting hypothalamic-pituitary-adrenal axis hormones and sex steroids for improving cognition in mood disorders and schizophrenia: a systematic review and narrative synthesis. *Psychoneuroendocrinology*. 2018;93:8–19.

Sparapani S, Millet-Boureima C, Oliver J, et al. The biology of vasopressin. *Biomedicines*. 2021;9(1):89.

Spencer-Segal JL, Akil H. Glucocorticoids and resilience. *Horm Behav.* 2019;111:131–134.

Stone LA, Harmatz ES, Goosens KA. Ghrelin as a stress hormone: implications for psychiatric illness. *Biol Psychiatry*. 2020;88(7):531–540.

Turner AI, Smyth N, Hall SJ, et al. Psychological stress reactivity and future health and disease outcomes: a systematic review of prospective evidence. *Psychoneuroendocrinology*. 2020;114:104599.

Wingenfeld K, Otte C. Mineralocorticoid receptor function and cognition in health and disease. *Psychoneuroendocrinology*. 2019;105:25–35.

▲ 1.13 Immune–Brain Interactions in Psychiatry

EBRAHIM HAROON, M.D., ANDREW H. MILLER, M.D., AND CHARLES L. RAISON, M.D.

The immunologic capacity is essential to all multicellular organisms. The relationship between the immune and central nervous systems (CNS) may underlie successful adaptations to environmental challenges. Increasingly, immune brain communication can be a cause of many psychiatric and medical disorders and is the result of multiple lifestyles and individual risk factors. Figure 1.13–1 provides an overview of these risk factors. However, for a long time, the brain was considered an immune-privileged organ as it existed outside the reaches of the body's immune system. Indeed, the blood–brain (BBB) and blood–cerebrospinal fluid (CSF) barriers separated the brain from the rest of the body, allowing only a few essential molecules or cells to pass through their tight walls. Consequently, immune molecules or cells have limited access to the CNS.

Instead, the brain employs its own immune system with lymphatics, lymphocytes in the meninges and choroid plexus, and various subsets of macrophage-like cells found in border regions and the brain parenchyma. However, despite the protection provided by the tight BBB, the brain immune system does not function autonomously from the peripheral immune system. Upheaval in the body's immune system impacts the brain and modifies its functioning. The brain, in turn, shapes and prepares the body's immune system for current and anticipated dangers. The brain has co-opted cellular and molecular elements of its immune system to regulate its development and modulate synaptic plasticity. This chapter will discuss the basic elements of this immune–brain communication and how this newly acquired knowledge can help foster newer therapeutic advances in psychiatry.

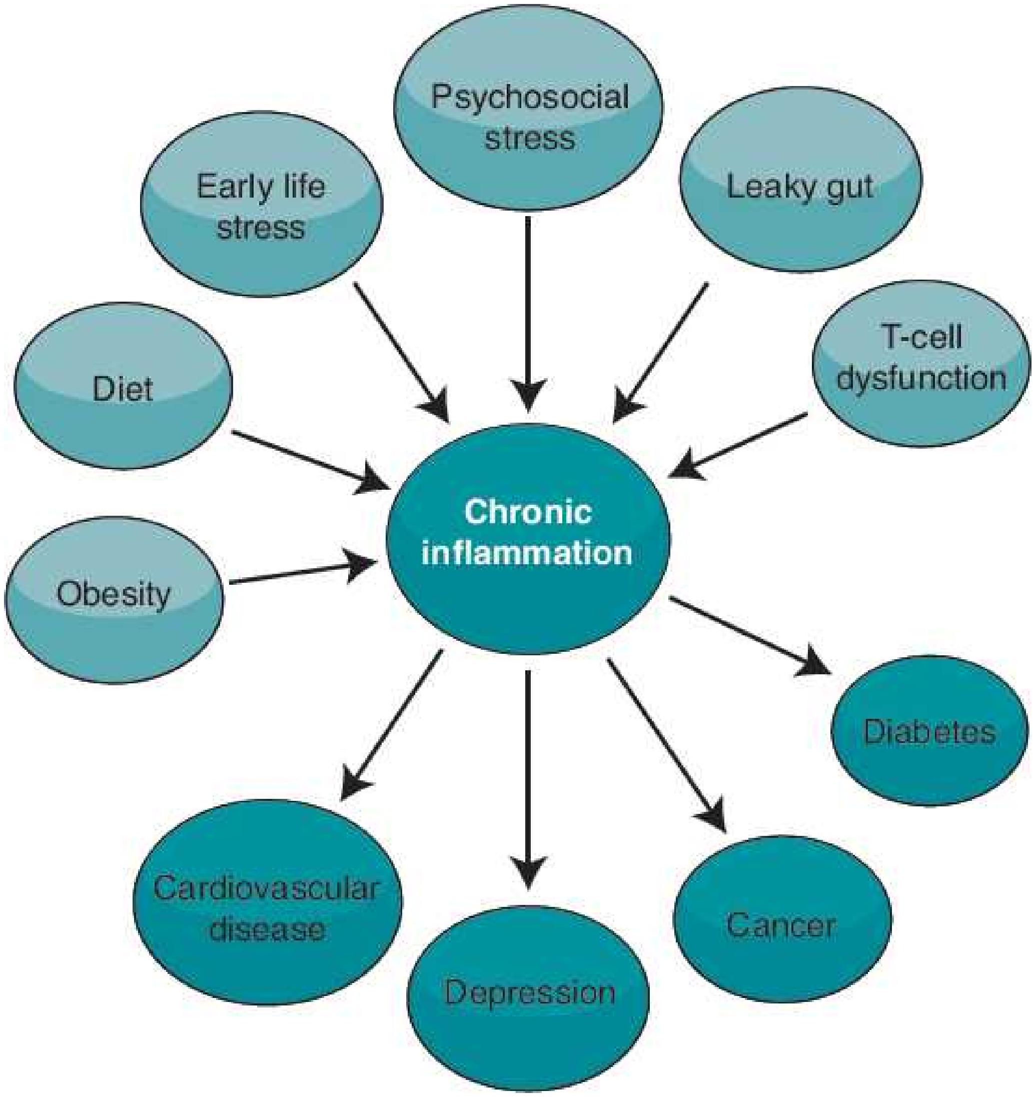
Psychosocial stress
Early life stress
Leaky gut
Diet
T-cell dysfunction
Chronic inflammation
Obesity
Diabetes
Cardiovascular disease
Depression
Cancer

FIGURE 1.13–1. The myriad of health consequences resulting from chronic immune activation is presented in purple-colored ellipses and the etiologic factors leading to increase in inflammation is depicted on blue-colored ellipses. (See eBook for color images.) (Reprinted from Haroon E, Raison C, Miller A. Psychoneuroimmunology meets neuropsychopharmacology: Translational implications of the impact of inflammation on behavior. *Neuropsychopharmacology* 2012;37:137–162. doi:10.1038/npp.2011.205.)

COMPONENTS OF THE IMMUNE SYSTEM

The immune system performs myriad functions, including:

1. protecting the body from the invasion of foreign pathogens, such as viruses, bacteria, fungi, and parasites,
2. detecting and eliminating cells that have become neoplastically transformed,
3. surveying, controlling, and maintaining the various symbiotic and commensal microbiota, notably in the gut,
4. sculpting of brain circuitry, and
5. regulating neurodevelopment and neuronal communication.

These functions are accomplished through specialized sensors (receptors) on immune and glial cell surfaces that recognize pathogenic molecules from invading organisms or neoplastic cells. The cells communicate via a rich intercellular communication network that involves direct cell-to-cell interactions and signaling between cells of the immune system by soluble factors called cytokines. Intriguingly, immune signaling molecules also communicate with brain cells via cytokine receptors on glial cells and some neurons, orchestrating a complex neurobehavioral response to any threat (physical or psychosocial).

Table 1.13–1.
Divisions of the Immune System: Innate versus Acquired

Lines of Defense	Innate	Acquired
Physiochemical barriers	Skin, mucous membranes	Cutaneous and mucosal immune systems
Cells	Phagocytes (macrophages, neutrophils, and natural killer cells)	Lymphocytes (B and T cells)
Soluble mediators that affect other cells	Macrophage-derived cytokines, i.e., IL-1, IL-6, TNF, IFN-a	Lymphocyte-derived cytokines, i.e., IL-2, IL-4, IL-10, IFN-g
Memory	Trained	Yes
Circulating molecules	Complement, acute-phase reactants	Antibodies

IL, interleukin; IFN-a, interferon-alpha; IFN-g, interferon-gamma; TNF, tumor necrosis factor.

Adapted from Abbas AK, Lichtman AH, Pillai, S. *Cellular and Molecular Immunology*. Elsevier; 2022, with permission.

Arms of the Immune System—Innate and Adaptive Immunity

Table 1.13–1 delineates the functional differentiation of the two arms of the immune system, that is, innate or natural immunity and adaptive or acquired immunity. Immune cells originate from stem cells in the bone marrow but differentiate into myeloid and lymphoid stem cell lines. The myeloid family gives rise to erythrocytes, platelets, and mononuclear phagocytic cells, and the lymphoid family yields T and B lymphocytes and natural killer (NK) cells.

Key differences and overlaps exist between innate and adaptive immune systems. The components of innate immunity act rapidly but are relatively nonspecific against pathogens or infected cells. On the other hand, the adaptive immune system is highly specific, carrying an “adaptive immune memory” that often lasts a lifetime. However, recent data have indicated that the innate immune system may also carry the recent memory of an encountered pathogen—a phenomenon called “trained innate immune memory.” Trained immunity involves creating a short-term “memory” of a pathogen by modifying innate immune cells’ epigenome, genome, and, ultimately, function.

Thus, trained immunity lasts up to several months, in contrast to the adaptive immune memory, which may last a lifetime. Furthermore, trained immunity does not create antibodies in preparation for a second encounter, unlike its adaptive counterpart. In other words, trained immunity is akin to recent memory, while adaptive immunity is akin to long-term memory. Therefore, innate immunity (trained or acute) is considered evolutionarily more primitive and less specialized than the T and B lymphocyte–mediated adaptive/acquired immunity. Intriguingly, lymphocytic cells, such as NK cells, act as innate immune cells. Operationally, however, all the above nodes of immunity interact and cooperate. Most notably, the immune system exhibits a high level of biologic redundancy, with several molecules converging on the same functional target, ostensibly as a fail-safe mechanism in the face of threat or danger.

Innate Immune System

Innate Immune Surveillance. The innate immune system is usually the first point of contact with internal or external threats. The cells mediating innate immunity are maintained at a constant background level of activation, partly by exposure to microbial components from the gut. These cells do not require prior activation and recognize invading pathogens without delay. Thus, they provide an important first line of defense against infectious agents during the early stages of an immune response. Mononuclear phagocytic cells and NK cells are examples of innate immune cells. NK cells are a specialized but important component of innate immunity. These cells originate from the lymphoblastic heritage-like adaptive immune cells but act as innate cells. They can destroy virally infected cells by binding to them and releasing cytolytic factors, including perforin. NK cells also can recognize and destroy neoplastically transformed host cells, especially those of hematopoietic origin, thus protecting against some cancers. Figure 1.13–2 depicts the hematopoietic origins of different immune cells.

Mononuclear phagocytic cells include the following:

1. Monocytes (circulating in the blood)

2. Macrophages (monocytic cells entering tissues from circulation)
3. Microglia (the macrophage equivalent in the brain)
4. Dendritic cells (immune cells residing in tissues)
5. Reticular cells (immune-capable fibroblasts residing in reticuloendothelial organs [spleen, lymph nodes, and lymphoid organs])

Of note, the reticuloendothelial organs may represent the crossing point at which immune information is exchanged between the innate and adaptive systems. These cells originate from the lymphoblastic heritage like adaptive immune cells but act as innate cells. They can destroy virally infected cells by binding to them and releasing cytolytic factors, including perforin. NK cells can also recognize and destroy neoplastically transformed host cells, especially those of hematopoietic origin, thus protecting against some cancers.

Threat Detection and Immune Signaling. Figure 1.13–3 summarizes the key events that take place during innate immune activation—of relevance to psychiatry. Once a threat is recognized, the innate system sends alarm signals to all other components of the immune and nervous systems. Immune molecules and proteins are keys to coordinating the system-wide responses to immune stimulation. The most important factors will be discussed below.

Surface Pattern Recognition Receptors. Innate immune cells are part of the immune surveillance system that survey circulating antigens and mobilize rapid immune responses upon discovering pathologic products. Innate cells recognize extracellular pathogens (e.g., bacteria and parasites) through relatively simple molecular pattern recognition receptors (e.g., toll-like receptors, or TLRs). TLRs recognize two types of pathogenic molecular patterns. The first involves evolutionarily conserved molecular patterns of infectious agents, called pathogen-associated molecular patterns (PAMPs). The second refers to the recognition of deoxyribonucleic acid (DNA) and other debris from dying or neoplastic cells, known as danger-associated molecular patterns (DAMPs). Thus, the immune response induced by DAMPs is referred to as sterile inflammation and can be provoked by biochemicals leaking from dying or distressed cells, including adenosine triphosphate (ATP), DNA, heat shock proteins

(HSPs), and high-mobility group box 1 (HMGB1). Then, these cells rapidly neutralize the threat by engulfing and degrading pathogens or dead cells.

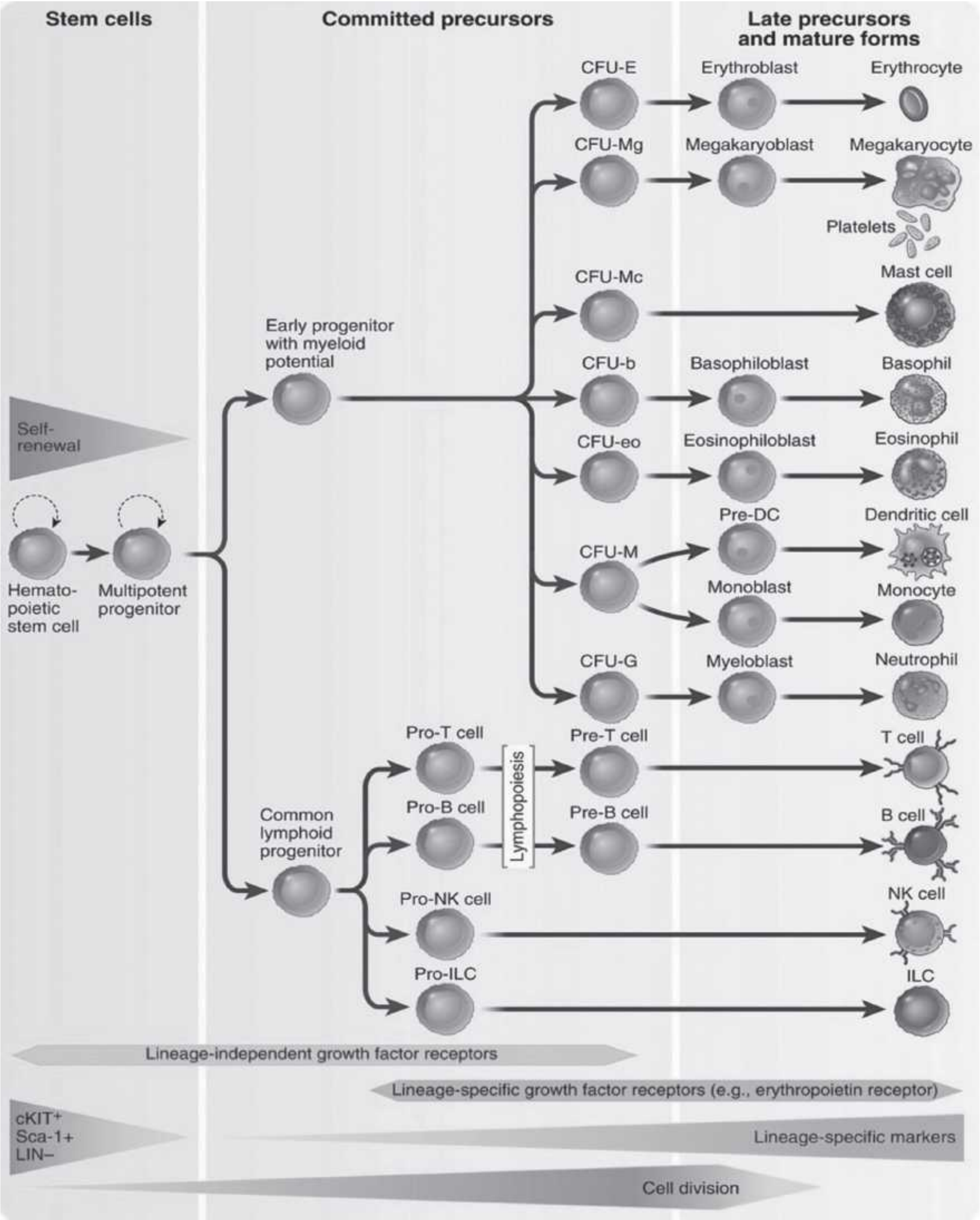

Stem cells
Committed precursors
Late precursors and mature forms
Self-renewal
Hemato-poietic stem cell
Multipotent progenitor
Early progenitor with myeloid potential
Common lymphoid progenitor
CFU-E
Erythroblast
Erythrocyte
CFU-Mg
Megakaryoblast
Megakaryocyte
Platelets
CFU-Mc
Mast cell
CFU-b
Basophiloblast
Basophil
CFU-eo
Eosinophiloblast
Eosinophil
Pre-DC
Dendritic cell
CFU-M
Monoblast
Monocyte
CFU-G
Myeloblast
Neutrophil
Pro-T cell
Pre-T cell
T cell
Pro-B cell
Pre-B cell
B cell
Lymphopoiesis
Pro-NK cell
NK cell
Pro-ILC
ILC
Lineage-independent growth factor receptors
Lineage-specific growth factor receptors (e.g., erythropoietin receptor)
cKIT+
Sca-1+
LIN–
Lineage-specific markers
Cell division

FIGURE 1.13–2. The development of the major lineages of blood cells is depicted in this hematopoietic tree. The principal cytokines that drive the maturation of different lineages are described in the source references provided below. *CFU,* Colony-forming unit; *CFU-Mc,* CFU mast cell; *CFU-b,* CFU B cell; *CFU-eo,* CFU eosinophil; *CFU-G,* CFU granulocytes; *CFU-M,* CFU macrophages; *DC,* dendritic cell; *ILCs,* innate lymphoid cells; *NK,* natural

killer. (Reprinted from Abbas AK, Lichtman AH, Pillai, S. *Cellular and Molecular Immunology*. Elsevier; 2022, with permission.)

Intracellular Signaling Cascades. Activation of TLRs by pathogens (and by some cytokines) triggers intracellular signaling cascades—akin to second messenger systems in neurons. These intracellular signaling cascades include nuclear factor-kappa-B (NF-kB) and a distinct subtype of mitogen-activated protein kinase (MAPK) known as p38 MAPK. Both NF-kB and MAPK play a sentinel role in initiating the innate immune response.

Inflammasomes. Sterile inflammatory responses to noninfective but toxic stimuli (dead, neoplastic cells, or products of stress and distress) are mediated by inflammasomes, which are cytosolic protein complexes in the myeloid cells that respond to microbial and sterile inflammatory stimuli. Just-in-time assembly of inflammasome machinery leads to activation of the enzyme caspase 1, which cleaves the inactive precursors of cytokines such as interleukin-1 (IL-1) beta and interleukin (IL)-18 into active forms that can drive the inflammatory response. Inflammasomes act in concert with TLRs but focus on responding to debris from dying cells or distressed cells (sterile inflammation). Indeed, studies suggest that the release of DAMPs such as adenosine from ATP of stressed cells may explain how psychological stress activates inflammation—in the absence of an infectious stimulus. These studies have revealed that adenosine and glutamate released by stressed neurons trigger activation of the inflammasome. Obesity is a well-known cause of chronic inflammation. DAMPs released from the formation and death of adipocytes lead to chronic innate immune activation in obesity. While beyond the scope of this chapter, inflammasome activation in monocytes and macrophages by SARS-CoV-2 viral particles results in much of the tissue damage in COVID-19–associated lung dysfunction.

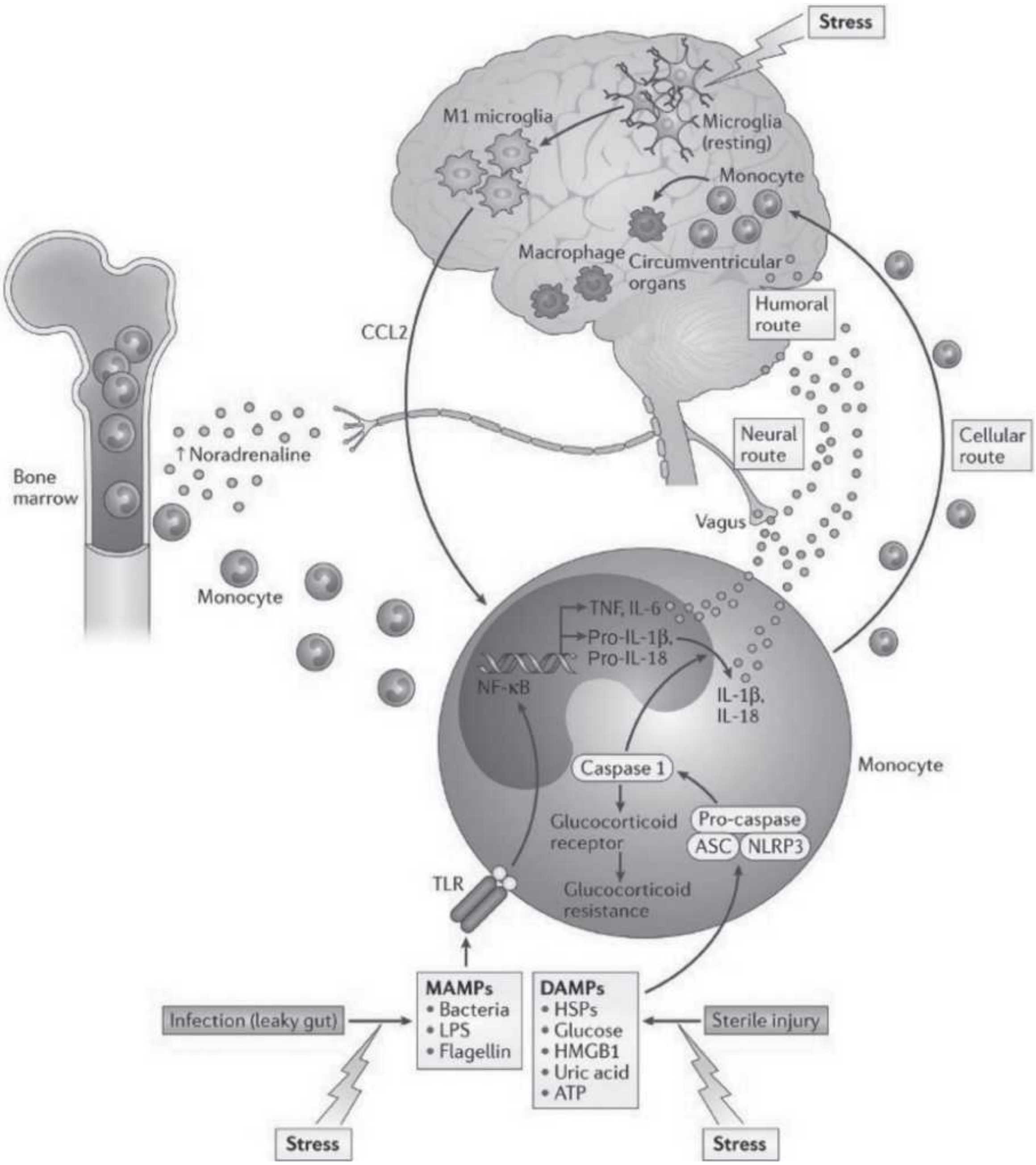
Stress
M1 microglia
Microglia (resting)
Monocyte
Macrophage
Circumventricular organs
Humoral route
CCL2
↑ Noradrenaline
Bone marrow
Neural route
Cellular route
Vagus
Monocyte
TNF, IL-6
Pro-IL-1β, Pro-IL-18
NF-κB
IL-1β, IL-18
Caspase 1
Monocyte
Glucocorticoid receptor
Pro-caspase
ASC
NLRP3
Glucocorticoid resistance
TLR
MAMPs
• Bacteria
• LPS
• Flagellin
DAMPs
• HSPs
• Glucose
• HMGB1
• Uric acid
• ATP
Infection (leaky gut)
Sterile injury
Stress
Stress

FIGURE 1.13–3. Locally cytokines act on endothelial and tissue stromal cells. Tissue stromal cells produce chemotactic factors, recruiting other immune cells to the site of injury. The endothelial cells produce adhesion molecules, enhancing immune cell margination and diapedesis. In the brain, proinflammatory cytokines—including interleukin (IL)-1, IL-6, and tumor necrosis factor (TNF)—activate the hypothalamic–pituitary–adrenal (HPA) axis and induce behavioral changes that subserve the metabolic demands of fever and inflammation. The proinflammatory cytokines also induce the liver to produce acute phase proteins. CAM, cellular adhesion molecule; IFN, interferon; NF-κB, nuclear factor-kappa-B; PGE, prostaglandin E; TLR, toll-like receptor. (Reprinted with permission from Cowles MK, Miller AH. Stress,

cytokines and depressive illness. In: Squire LR, ed. *The New Encyclopedia of Neuroscience*. Elsevier: Academic Press; 2009, 519–527.)

Cytokines. Cytokines are large molecular weight proteins that serve as the primary long-distance signaling molecules of the immune system. Some of their myriad functions are illustrated in Table 1.13–2. Activated mononuclear cells release proinflammatory cytokines, including tumor necrosis factor (TNF), IL-1, and IL-6. TNF and IL-1 are the principal mediators of the response to gram-negative bacteria and are among the earliest cytokines released in the inflammatory cascade, followed by IL-6. TNF and IL-1 are endogenous pyrogens capable of inducing fever by increasing the synthesis of prostaglandins by cytokine-stimulated hypothalamic cells. TNF also leads to cachexia, characterized by the wasting of muscle cells and appetite suppression. Activated mononuclear cells also release type I interferons (e.g., interferon [IFN]-alpha) with direct antiviral properties.

Acute Phase Response Proteins. Acute Phase Response Proteins (APP) synthesized primarily in the liver, limit tissue damage, isolate, and destroy invading pathogens, and set repair functions in motion. APPs are divided into positive and negative acute-phase proteins. Negative acute-phase proteins are those molecules whose production is decreased to make way for the synthesis of positive APPs. Negative acute-phase proteins include albumin, transferrin, and retinol-binding protein and are often used to measure the extent of chronic inflammation (e.g., following cachexia). Positive APPs aid in inflammatory activity and include the following:

Table 1.13–2.
Cytokine Mediators of Innate Immunity

Cytokines and Subunits	Principal Cell Source	Principal Cellular Targets and Biologic Effects
Interleukin-6 (IL-6)	Macrophages, DCs, endothelial cells, T cells	*Liver:* Synthesis of acute-phase proteins, *B cells:* Proliferation of antibody-producing cells, *T cells:* Th17 differentiation
Interleukin-2 (IL-2)	T cells	T cells: Proliferation and differentiation into effector and memory cells; promotes regulatory T cell development, survival, and function, *NK cells:* Proliferation, activation
Interleukin-4 (IL-4)	CD4 $^+$ T cells (Th2), mast cells	*B cells:* Isotype switching to IgE, IgG4 (in humans; IgG1 in mice), *T cells:* Th2 differentiation, proliferation, *Macrophages:* Alternative activation and inhibition of IFN-g-mediated classical activation
Interleukin-27 (IL-27)	Macrophages, DCs	*T cells:* Enhancement of Th1 differentiation; inhibition of Th17 differentiation
Granulocyte–monocyte CSF (GM-CSF)	T cells, macrophages, endothelial cells, fibroblasts	*Immature and committed progenitors, mature macrophages:* Induced maturation of granulocytes and monocytes, macrophage activation
Granulocyte-CSF (G-CSF, CSF3)	Macrophages, fibroblasts, endothelial cells	*Committed hematopoietic progenitors:* Maturation of granulocytes
IFN-a (type I IFN; multiple proteins)	Plasmacytoid DCs, macrophages	*All cells:* Antiviral state, increased class I MHC expression, *NK cells:* Activation
IFN-g (type II IFN)	T cells (Th1,	*Macrophages:* Classical activation

	CD8 $^{+}$T cells), NK cells, ILC1s	(increased microbicidal functions), *B cells:* Isotype switching to opsonizing and complement-fixing IgG subclasses (in mice, not humans), *T cells:* Th1 differentiation, *Various cells:* Increased expression of class I and class II MHC molecules, increased antigen processing and presentation to T cells
Interleukin-10 (IL-10)	Macrophages, T cells (mainly regulatory T cells)	*Macrophages:* DCs: Inhibition of expression of IL-12, co-stimulators, and class II MHC
Tumor necrosis factor (TNF, also called TNF-alpha; TNFSF2)	Macrophages, NK cells, T cells	*Endothelial cells:* Activation (inflammation, coagulation), *Neutrophils:* Activation, *Hypothalamus:* Fever, *Muscle, fat:* Catabolism (cachexia)
Interleukin-1 beta (IL-1β)	Macrophages, DCs, fibroblasts, endothelial cells, keratinocyte; major type of biologically active IL-1	*Endothelial cells:* Activation (inflammation, coagulation), *Hypothalamus:* Fever, *Liver:* Synthesis of acute-phase proteins, *T cells:* Th17 differentiation
IL-1 receptor antagonist (IL-1RA)	Macrophages	*Various cells:* Competitive antagonist of IL-1
Interleukin-17A (IL-17A), Interleukin-17F (IL-17F)	CD4 $^{+}$T cells (Th17), ILC3s	*Epithelial cells, macrophages, and other cell types:* Increased chemokine and cytokine production; GM-CSF and G-CSF production

Transforming growth factor-beta (TGF-β)	T cells (mainly Tregs), macrophages, other cell types	*T cells:* Inhibition of proliferation and effector functions; differentiation of Th17 and Treg, *B cells:* Inhibition of proliferation; IgA production, *Macrophages:* Inhibition of activation; stimulation of angiogenic factors, *Fibroblasts:* Increased collagen synthesis
MCP (Monocyte chemoattractant protein)-1, (CCL2)	Many cells	*Mixed leukocyte:* Recruitment
IL-8, (CXCL8)	Many cells	*Neutrophil:* Recruitment
Fractalkine, (CX3CL1)	Many cells	*T cell, NK cell, and monocyte:* Recruitment

CCL2, CXCL8, CXC3CL1, chemokine nomenclature for MCP-1, IL-8 and Fractalkine, respectively; DC, dendritic cells; IFN, interferon; IFN-a, interferon-alpha; IFN-g, interferon-gamma; Ig, immunoglobulin; MHC, major histocompatibility complex; NK, natural killer cells; Th, T helper cells.

Adapted with permission from Abbas AK, Lichtman AH, Pillai, S. *Cellular and Molecular Immunology*. 10th ed. Elsevier; 2022.

1. Protease inhibitors: for example, alpha 1 antitrypsin, alpha 1 antichymotrypsin
2. Coagulation proteins: for example, fibrinogen, prothrombin
3. Complement proteins: for example, C2, C3, C4, C5, etc.
4. Transport proteins: for example, haptoglobin
5. Inflammatory proteins: for example, C-reactive protein (CRP), serum amyloid A (SAA)

The combination of liver-derived plasma proteins induced by TNF and IL-1 with those induced by IL-6 constitutes the acute-phase response (see Fig. 1.13–3). Notably, the synthesis of CRP in the liver is increased by the action of IL-6 on hepatocytes. Thus, plasma CRP is a well-regarded measure of overall inflammatory activation. It is standardized across clinical laboratories and is well known to predict the risk of acute (pneumonitis) and chronic inflammation (cardiovascular disease, autoimmune disorders).

Complement Factors. Complement factor proteins, also produced by the liver, represent another important component of the innate immune response. These functionally linked proteins interact with each other in an orderly manner to subserve innate immune functions, including cell lysis, opsonization, attracting inflammatory cells (chemotaxis), stimulating immune cells to release chemical mediators of inflammation, and neutralization of antigen–antibody complexes to prevent tissue damage. Of note, complement proteins tag cells for identification and destruction by phagocytic cells. For instance, complement proteins such as the C4A-complement shape neuronal pruning of synapses by microglial phagocytosis during development. Aberrant phagocytic signals leading to excess pruning are believed to play an etiologic role in some cases of autism and schizophrenia. Nascent evidence indicates that activation of immunologic mechanisms via maternal or fetal infection may worsen pathology associated with aberrant complement activation.

Adaptive Immunity

T and B lymphocytes are the crowning achievements of the evolution of immune cell specialization and account for the diversity, specificity, and adaptive functionality within the immune system. The adaptive immune response includes three phases, that is, a recognition phase, an activation phase, and an effector phase of antigen elimination.

Recognition Phase. During the recognition phase, the presence of an infectious agent, antigen, or neoplastically transformed cell is sensed using specialized receptors for antigens on the surfaces of B and T lymphocytes. The receptors for an antigen on a particular B or T cell are identical and unique to their descendants (clones). A family of lymphocytes with identical antigen receptor specificity is called a clonal line. Thus, diversity in antigen recognition is derived from the vast number of different B- and T-cell clonal lines present in each individual. When an antigen is detected, the corresponding clonal line is selected and activated. Of note, the surface receptors on lymphocytes are highly specialized and customized to each pathogen or dying cell and are typed using their surface protein expression—referred to as a cluster of differentiation (CD). Intriguingly, the reticuloendothelial organs are conceptualized as librarians holding vast libraries of clonal lines, ready for activation. Several antigen-presenting innate cells traffic into the sinusoids of these organs via lymphatic channels and come into proximity with the lymphocytes. Thus, antigen recognition by lymphocytes may occur in a paracrine manner in the reticuloendothelial organs such as the spleen, thymus, and lymph nodes. Of note, the miseducation of lymphocytes at this interface may also trigger some leukemias.

Activation Phase. The activation phase includes the proliferation and mobilization of the immune cells relevant to eradicating the infectious agent. The binding of foreign antigens alone is usually insufficient to produce B- and T-cell activation—an accessory signal is required. Important accessory signals are generated by the IL group of cytokines secreted by T helper (Th) lymphocytic cells acting in concert with antigen-presenting cells (APCs), such as macrophages and dendritic cells. The Th cells and APCs stimulate each other in a reverberating manner to mount and amplify the immune response. APCs secrete IL-1 and other cytokines that stimulate Th cells to secrete a host of cytokines, including IFN-gamma, which further increase the activity of APCs, ultimately improving their antigen-presenting capacity. Thus, cytokines involved in the acquired immune response often simultaneously serve multiple functions. Binding to antigen in the presence of stimulatory cytokines results in activating intracellular systems regulating cell growth, division, and proliferation of T and B lymphocytes. Activation also results in the clonal expansion of immune cells with the identical high-affinity specificity for the foreign antigen. In addition, some progenies undergo further differentiation during clonal expansion into mature effector cells, such as antibody-secreting plasma B cells and cytotoxic CD8+ T lymphocytes (CTLs) that participate in cell-mediated immunity.

In contrast, some descendants of activated B or T cells become memory cells primed for activation upon future stimulation by the same antigen. Thus, reexposure to the same antigen elicits a more rapid and robust secondary immune response (thus the name acquired immunity) than the first or primary immune response to that antigen. Notably, memory cells may live for many years providing long-lasting acquired immunity, as demonstrated by individuals who have received a vaccine or had their first contact with a specific infectious agent during infancy. Similarly, exaggerated immune memory may play a part in hypersensitivity reactions.

Effector Phase. During the effector phase, the pathogen is neutralized and eliminated. Acquired immune responses employ humoral (antibody-mediated) and cell-mediated (CTLs) processes to neutralize a threat. Antibodies secreted from B cells are principal effectors of humoral immunity. Humoral immunity is especially effective in combating extracellular pathogens, such as bacteria and parasites. Cell-mediated immunity using CTLs is used to kill target cells directly. In contrast to humoral immunity, cell-mediated (CTL-based) immunity effectively protects against viral and intracellular bacterial infection and, as with NK cells, may provide some protection against tumor cells. The T cells also release macrophage-activating cytokines such as IFN-gamma that enhance the antimicrobial functions of macrophages, including microglia. Regulation of the acquired immune response is another important component of the effector phase. It includes CD4+CD25+ regulatory T cells (also known as CD4 cells), which are immunosuppressive and restrict immune responses to self-antigens. In addition, the inhibitory costimulatory molecule called programmed death-1 (PD-1) and its ligands have been shown to play an important role in regulating/suppressing T-cell activation. Indeed, loss of PD-1 has been associated with an autoimmune diathesis in laboratory animals. PD-1 depletion has also been proposed as a mechanism leading to “cytokine storm” in conditions such as SARS-CoV-2 infections.

Integration with the Innate Immune System. B and T cells also recruit, enhance, and direct innate immune mechanisms to neutralize the threat posed by specific antigens. For example, circulating antibodies can neutralize pathogens by binding to and coating the pathogens (opsonization). Pathogens that are opsonized are made susceptible to lysis by complement factors and phagocytosis. NK cells and phagocytic cells (such as neutrophils and macrophages) have receptors for the Fc fragment (fragment crystallizable region) of antibodies. Furthermore, complement proteins bind to and are activated by the Fc fragments of some types of antibodies. Thus, antibodies can link effector cells and cytolytic proteins of innate immunity with pathogens, lending a level of specificity that is not inherent in the effector arms of the innate immune processes themselves.

Cytokines Synchronize Innate and Acquired Immunity. Cytokines are the primary signaling molecules that convey immune information between different components of the immune system. Cytokines can be broadly divided into categories based on initiating, regulating, and maintaining the immune response. Hence, some cytokines mediate innate immunity and inflammation, others regulate acquired immunity, and the remaining control proliferation and differentiation of immature immune cells (see Table 1.13–2). Although structurally distinct, these cytokines overlap in function and act together to govern the dynamic events of immunity.

Regulatory T Cells. Within the acquired immune response, significant interest has been paid to two varieties of Th cells (CD 4+ T cells) known as Th1 and Th2. Th1 cells produce cytokines such as IL-2 and IFN-gamma that promote T-cell proliferation and activity and promote inflammatory responses that are especially relevant for protection against intracellular pathogens. Th2 cells produce cytokines such as IL-4 that promote antibody production and provide protection against parasites. In addition, Th2-type T-regulatory cells dampen immune activation both via direct cell-to-cell effects and via the production of anti-inflammatory cytokines, especially IL-10. Exaggerated Th2 responses have been closely associated with allergic and hypersensitivity reactions and asthma. A third subset of Th17 cells, developmentally distinct from Th1 or Th2 cells, also has been identified. These cells produce the cytokine IL-17, which has recently been linked to autoimmune (psoriasis) and allergic conditions once thought to be more closely involved with Th1 or Th2 dysregulation. More recently, two additional subsets of effector cells have been described including, Th9 cells, which secrete IL-9 and seem to be predominantly proinflammatory, and Th22 cells, which secrete IL-22 and can have either pro- or anti-inflammatory effects.

IMMUNE–BRAIN COMMUNICATION

The CNS and immune system interact and modulate each other in health and disease. This complex interplay of immune–brain communication is illustrated in Figure 1.13–4. While both systems defend against threats, each offers unique evolutionary advantages. For example, the immune system offers a specialized network to neutralize threats, but its operational domain is at the closest proximity and immediacy of the threat. On the other hand, as a superior prediction machine, the brain can anticipate and prepare the immune system for a future threat. Similarly, the immune system can control behavior in a manner that promotes healing and recovery. In the CNS, resident immune cells (microglia) are found mostly in the parenchyma, while trafficking immune cells are stationed in specialized niches inside the brain—including meninges and choroid plexus, ready to be called into action when indicated. Astrocytes participate in immune responses only after priming by microglial cells, as happens during aging or brain infections. The BBB separates the cerebral blood vessels from brain tissue and tightly regulates the flow of peripheral immune molecules into the CNS. Under usual circumstances, direct between the blood and the CNS can occur only at specific locations. The blood–brain interchange typically occurs in fenestrated capillaries found in circumventricular organs (CVOs). The recently discovered brain lymphatic (glymphatic) system facilitates the movement of fluid and small molecules between the CSF and interstitial fluid. The glymphatic system is one possible route for cytokines from the meningeal compartment to the brain. Thus, the CSF, carrying cytokines (from the peripheral circulation) in the subarachnoid space, can enter the Virchow–Robin spaces, dive deep along with the perivascular spaces, and then diffuse into the brain parenchyma. Subsequently, the cytokines may interact with cytokine receptor–expressing neurons, glia, and stroma in the brain parenchyma. However, how cytokines cross the arachnoid mater to reach the subarachnoid space during homeostasis remains unanswered.

Neuronal activity can alter the immunologic response within the brain. Microglia, the brain's resident immune cells, respond to cues from ongoing neuronal activity. Neural activity is accompanied by the spillover of neurotransmitters that escape from the synapse. Microglia sense these spillover neurotransmitters using surface receptors. The activation of these receptors modifies the function and phenotype of microglial cells. In addition, brain activity can also stimulate epithelial cells in the CVOs by increasing the release of chemotactic molecules, changing the permeability of the BBB, and promoting the recruitment of immune cells from peripheral circulation into the brain.

Comparisons of immune–brain communication pathology in stress-induced psychiatric disorders such as depression and post-traumatic stress disorder (PTSD) reveal a more complex picture that defies simplistic explanations. There is sufficient evidence linking peripheral and central immune changes in unipolar and bipolar affective disorders (see later sections in this chapter). Emerging data indicates that this process may be less connected in PTSD. This will be discussed in the section on PTSD.

Immune Receptors on Neurons

Both immune and neurologic systems detect and neutralize threats and maintain homeostasis. Both systems perceive internal and external environments and use specific receptors to monitor ambient stimuli. For example, both sensory afferent fibers of the autonomic nervous system (vagal afferents) and peripheral sensory neurons in the dorsal root ganglia display immune receptors (TLRs and cytokine receptors) and transmit immunologic signals to the brain. Peripheral neurons become more sensitive to pain and itch stimuli when these immune receptors are stimulated, resulting in behavioral modifications. For example, avoidance behavior occurs when an organism protects itself from injury, danger, or infection by avoiding a substance or circumstance deemed harmful by immune-associated sensory cues.

Neurochemical Receptors on Immune Cells

In addition to signaling neural pathways, peripheral immune cells express receptors for a wide range of chemicals regulated by or produced from several body tissues, as shown in Table 1.13–3. The adrenergic receptor expressed on T and B cells was one of the first receptors to be identified in lymphocytes. Following that, receptors for various small molecule neurotransmitters such as noradrenaline (NE), dopamine (DA), acetylcholine (ACh), serotonin (5HT), and glutamate were recognized on immune cells including microglia. Neurotransmitter receptors on immune cells, like those in the CNS, are found in the cell surface and, in most cases, are linked to G proteins and second-messenger pathways. The impact of CNS chemicals and hormones on the immune system varies depending on which step of the immune response is involved (recognition, activation, or effector). For example, norepinephrine has been shown to boost immune activity during recognition but suppress response during the effector phase. As a result, while persistent stress is linked to increased baseline inflammation, it also reduces the body’s ability to fight infections and develop a strong immunologic response to immunizations. These data suggest that the time of exposure as well as the dose are critical.

Lymphoid Tissue Nerve Innervation

One of the first indications of immune–brain communication was the discovery of the sympathetic nervous system (SNS) innervation of immunologic tissues. SNS fibers have been found in reticuloendothelial organs (bone marrow, thymus, spleen, and lymph nodes) involved in lymphocyte growth, education (e.g., self–nonself-discrimination), and function. The superior cervical and stellate ganglia, for example, supply sympathetic fibers to the thymus, whereas the mesenteric and celiac plexuses supply sympathetic input to the spleen. Furthermore, the sympathetic supply to the bone marrow is derived from the spinal cord level corresponding to the bone’s position. Efferent input from the vagal or recurrent laryngeal nerves normally provides parasympathetic counter-regulation. Sympathetic nerve fibers usually enter lymphoid tissues in tandem with vascular supply. These fibers are widely known for modulating vascular tone by controlling the contractility of the smooth muscles that line the arteries. Nevertheless, penetrating nerve fibers also terminate near developing lymphocytes and macrophages. By altering blood flow into lymphoid organs and locally releasing neurotransmitters like NE and neuropeptides, SNS fibers can modulate immune cell functions.

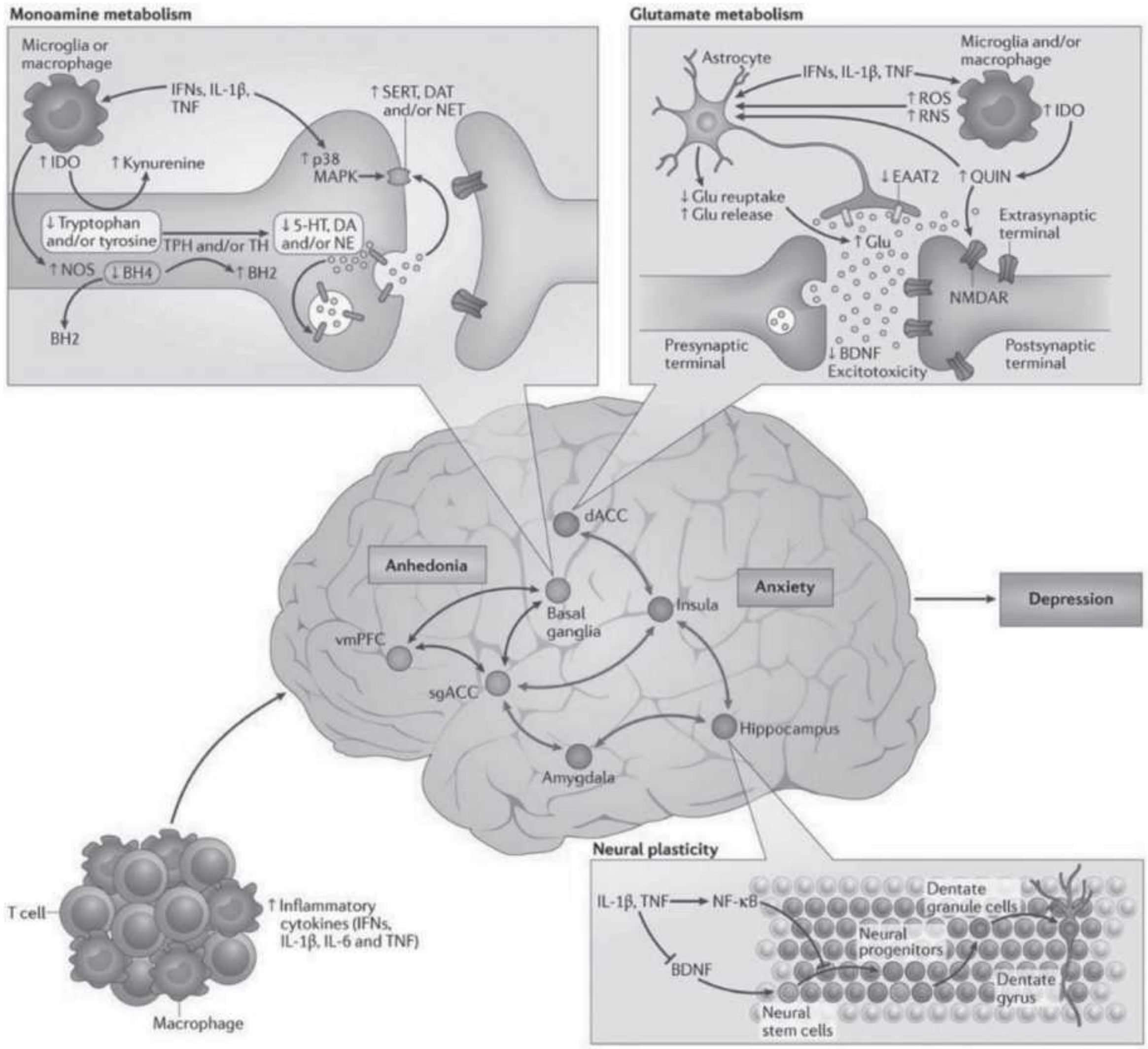
Monoamine metabolism
Microglia or macrophage
IFNs, IL-1β, TNF
↑ SERT, DAT and/or NET
↑ p38 MAPK
↑ IDO
↑ Kynurenine
↓ Tryptophan and/or tyrosine
TPH and/or TH
↓ 5-HT, DA and/or NE
↑ NOS
↓ BH4
↑ BH2
BH2
Glutamate metabolism
Astrocyte
IFNs, IL-1β, TNF
Microglia and/or macrophage
↑ ROS
↑ RNS
↑ IDO
↓ EAAT2
↑ QUIN
↓ Glu reuptake
↑ Glu release
↑ Glu
Extrasynaptic terminal
NMDAR
Presynaptic terminal
↓ BDNF
Excitotoxicity
Postsynaptic terminal
dACC
Anhedonia
Basal ganglia
Insula
Anxiety
Depression
vmPFC
sgACC
Hippocampus
Amygdala
Neural plasticity
IL-1β, TNF
NF-κB
Dentate granule cells
Neural progenitors
BDNF
Dentate gyrus
Neural stem cells
T cell
↑ Inflammatory cytokines (IFNs, IL-1β, IL-6 and TNF)
Macrophage

FIGURE 1.13–4. In the context of psychosocial stress, catecholamines (such as noradrenaline) released by activated sympathetic nervous system fibers stimulate bone marrow production and the release of myeloid cells (for example, monocytes) that enter the periphery where they encounter stress-induced damage-associated molecular patterns (DAMPs), bacteria and bacterial products such as microbial-associated molecular patterns (MAMPs) leaked from the gut. These DAMPs and MAMPs subsequently activate inflammatory signaling pathways such as nuclear factor-kappa-B (NF-κB) and the NOD-, LRR- and pyrin domain-containing protein 3 (NLRP3) inflammasome. Stimulation of NLRP3 in turn activates caspase 1, which leads to the production of mature interleukin-1-beta (IL-1β) and IL-18 while also cleaving the glucocorticoid receptor contributing to glucocorticoid resistance. Activation of NF-κB stimulates the release of other proinflammatory cytokines including tumor necrosis factor (TNF) and IL-6, which together with IL-1β and IL-18 can access the brain through humoral and neural routes. Psychosocial stress can also lead to the activation of microglia to an M1 proinflammatory phenotype, which releases CC-chemokine ligand 2 (CCL2) that in turn attracts

activated myeloid cells to the brain via a cellular route. Once in the brain, activated macrophages can perpetuate central inflammatory responses. ASC, apoptosis-associated speck-like protein containing a CARD; HMGB1, high mobility group box 1; HSP, heat shock protein; LPS, lipopolysaccharide; TLR, Toll-like receptor. (Reprinted from Miller A, Raison C. The role of inflammation in depression: from evolutionary imperative to modern treatment target. *Nat Rev Immunol* 2016;16:22–34. doi: 10.1038/nri.2015.5)

Table 1.13–3.
Neurotransmitters, Hormones, and Peptides— Targeting Immune cells

Neurotransmitters	Hormones	Peptides
Dopamine	Corticosteroids–glucocorticoids	ACTH
Glutamate, receptors, and transporters	Mineralocorticoids	Alpha-MSH
Norepinephrine	Estrogen	AVP
Serotonin	Progesterone	Calcitonin
Histamine	Testosterone	CRH
Acetylcholine	GHRH	IGF-1
GABA	GnRH	Melatonin
	Growth hormone	NPY
	Opioids (endorphins, enkephalins)	PTH
	Thyroid hormone	Somatostatin
	Prolactin	Substance P
		TRH
		TSH VIP
		CGRP

Receptors for the above molecules have been identified on the immune cells, indicating their impact on the immune system.

ACTH, adrenocorticotropin; alpha-MSH, alpha-melanocyte-stimulating hormone; AVP, arginine vasopressin; CGRP, calcitonin gene–related peptide; CRH, corticotrophin-releasing hormone; GABA, gamma amino butyrate; GHRH, growth hormone–releasing hormone; GnRH, gonadotropin-releasing hormone; IGF-1, insulin-like growth factor-1; NPY, neuropeptide Y; PTH, parathyroid hormone; TRH, thyrotropin-releasing hormone; TSH, thyroid-stimulating hormone; VIP, vasoactive intestinal peptide.

Microscale Correlates of Inflammatory Activation on the Brain

Table 1.13–4 illustrates a selected list of key neurotransmitters of relevance to psychiatry. We will provide an abbreviated discussion of selected contents of the table below.

Table 1.13–4.
Immunologic Effects of Representative Neurotransmitters

Neurotransmitters	Immune Markers	Preclinical	Human
Serotonin	TNF and IL-1	Increased 5HTT (number and activity)	
	TNF		Increased brain stem 5HTT
SSRI (paroxetine)	IFN-alpha		Attenuate severity of depression
Dopamine			
Dopamine	IFN-alpha	*Primates, microdialysis:* Decrease in DA secretion, rescued by DA supplementation	
Dopamine	IL-6	*Rodents, microdialysis:* Decreased DA	
	IFN-alpha		Decreased presynaptic DA function on PET, Decreased basal ganglia activity on fMRI
	IL-6		CSF BH4 negatively correlated
DA promoting medications	CRP >1 mg/L		Improvement in Depressive Symptoms with

			bupropion, stimulants
Glutamate			
Glutamate	IFN-alpha		Increases in MRS glutamate signals
Glutamate	CRP >3 mg/L		Increases in MRS glutamate signals
Glutamate	KYNA/QUIN		Negatively Correlated with MRS glutamate in depression
Ketamine	LPS	*Rodents:* Pre-administration of Ketamine blocked sickness symptoms	
Ketamine	IL-6, CRP, kynurenine		Predictors of clinical response
GABA			
GABA	IL-18		Predicted depression severity
Acetyl choline			
Acetyl choline	Anti-inflammatory		Predicted increase in heart rate variability
	KYNA	*Rodents:*	Increased KYNA

Increased KYNA cognitive deficits by decreasing a7nAChR	related to experimental stress in schizophrenic subjects

a7nAChR: alpha-7-nicotinic acetyl choline receptors; DA: dopamine; HTT: serotonin reuptake transporters; KYNA: kynurenic acid; LPS: lipopolysaccharide bacterial endotoxin; MRS: magnetic resonance spectroscopy; and QUIN; quinolinic acid.

Norepinephrine. The effects of immune molecules on NE have been extensively studied. NE binding to beta-adrenergic receptors is a key mechanism by which stress activates immune cells. Beta-adrenoceptor blockade therapy may be useful in controlling the aberrant behavior of immune cells in medical disorders. Thus, for example, administration of beta-blockers before cancer therapy may be helpful to mitigate the risk of metastasis.

Serotonin. Innate immune cytokines, including TNF and IL-1, have also been shown to increase the expression and function of synaptic reuptake pumps for serotonin (and NE) via stimulation of p38 MAPK-linked pathways. The relevance of this mechanism to depression is highlighted by a recent human study showing that increased circulating levels of TNF were associated with increased brainstem serotonin reuptake availability. Conversely, etanercept, an antagonist to TNF, lowered brainstem serotonin transporter availability and improved symptoms of depression. Finally, treatment with the serotonin reuptake inhibitor, paroxetine, has been shown to attenuate the behavioral consequences of IFN-alpha administration, further highlighting the importance of effects on serotonin metabolism in cytokine-induced mood disturbances.

Dopamine. In addition to the effects of chronic cytokine exposure on noradrenergic and serotonergic transmission, it has been reported that chronic administration of IL-2 or IFN-alpha significantly alters DA metabolism while also leading to depression. For example, induction of inflammation in nonhuman primates (using IFN-alpha) led to decreases in DA metabolism in the basal ganglia, which was associated with depression-like behavior and responded to local L-Dopa administration via reverse microdialysis. These data parallel data from rodent studies, which demonstrated decreased DA in the nucleus accumbens following intraperitoneal injections of IL-6. Furthermore, chronic (i.e., 4 to 6 weeks) IFN-alpha treatment led to anhedonia, depression, and fatigue that correlated with reduced basal ganglia activation on functional magnetic resonance imaging (MRI) in response to motivational signals and decreased presynaptic striatal DA function. Several recent studies have investigated the mechanisms that underlie DA deficits. The pteridine derivative tetrahydrobiopterin (BH4) is an essential cofactor in synthesizing key monoamines relevant to psychopharmacology, including DA, NE, and serotonin. By diverting scarce BH4 to catalyze the generation of reactive oxygen and nitrogen species, inflammatory cytokines (such as IFN-gamma, TNF, and IL-6) decrease the production of key monoamines such as DA. For example, CSF concentrations of BH4 correlate negatively with CSF IL-6 levels among patients treated with IFN-alpha. Notably, plasma phenylalanine to tyrosine ratio (an indirect measure of BH4 activity) correlates with CSF concentrations of DA and severity of depressive symptoms following IFN-alpha–treatment. Targeting the DA system in inflammation using bupropion or stimulants decreases inflammation-induced anhedonic symptoms in human and preclinical models.

Glutamate. Although initial interest in the impact of inflammation focused on the monoamines (i.e., serotonin, NE, DA), studies also implicate changes in glutamate transmission as being relevant to the behavioral effects of cytokines. For example, increased levels of glutamate in the basal ganglia and dorsal anterior cingulate cortex (dACC) have been observed in patients receiving IFN-alpha, and higher levels of glutamate in these brain regions correlated with an increase in depressive symptoms. Glutamate exerts a paradoxical effect on neural plasticity and survival by promoting both neuronal health and toxicity. It may do so via its location-specific effects on glutamate receptors. Stimulation of synaptic N-methyl D-aspartate (NMDA) receptors enhances synaptic growth and survival, while stimulation of extrasynaptic NMDA receptors can promote cell death. The underlying mechanisms of toxicity include both receptor overstimulation and decreased synthesis of brain-derived neurotrophic factor (BDNF). Finally, increased extrasynaptic glutamate may stimulate presynaptic metabotropic glutamate receptors (mGluR) that inhibit the release of glutamate in the synapse, further impacting synaptic dysfunction. In this context, activated immune cells, including microglia and trafficking macrophages, are plentiful sources of extracellular/extrasynaptic glutamate in the brain. Activated immune cells have high antioxidant requirements due to accelerated mitochondrial metabolism and increased synthesis of reactive oxygen species. The release of glutamate by immune cells may be increased via system xC-, which extrudes glutamate molecules in exchange for cysteine to synthesize the antioxidant glutathione. In addition, the above cells release glutamate into the extrasynaptic space due to inflammation- and quinolinic acid (QA)-induced glial toxicity (discussed in the following sections). Moreover, immune-associated dysfunction of astrocytic glutamate reuptake pumps leads to decreased clearance and inefficient containment of glutamate, resulting in its “spillover” into the extracellular space. Increases in brain glutamate measures have been associated with increased neuropathology in neuroimmune disorders such as MS and HIV, depression-like behavioral changes following IFN-alpha, and among depressed individuals with high inflammation. Thus, there is significant evidence supporting immune-associated dysregulation of glutamate in depression and other neuropsychiatric disorders. In addition, glutamate dysregulation may also be bound to the activation of the kynurenine (KYN) pathway.

Kynurenine Pathway Molecules with Glutamate-Like Effects. KYNs are inflammatory molecules that bridge peripheral inflammation with neurobehavioral outcomes. When induced by cytokines like IFN-gamma or TNF, the enzyme indoleamine 2,3-dioxygenase (IDO) produces KYN from tryptophan in immune cells. KYN then undergoes a series of biotransformations resulting in glutamate-modifying metabolites like QA and kynurenic acid (KYNA). QA affects glutamate toxicity by decreasing the expression of astrocyte glutamate reuptake pumps, stimulating astrocytic glutamate release, and activating extrasynaptic NMDA receptors. KYNA, on the other hand, decreases the activation of NMDA receptors by allosterically modulating their glycine-binding sites. Antagonistic actions of KYNA on nicotinic acetylcholine receptors type-7 (nAChR7), which are linked to cognitive deficits in schizophrenia, also reduce presynaptic glutamate release. QA levels rise during inflammatory activation because inflammatory molecules selectively activate the QA pathway. Exaggerated NMDA activity in preclinical models and increased severity of depressive symptoms and suicidal behavior in depressed humans have been linked to concurrent increases in QA and decreases in KYNA. In the brain, QA is produced by microglia, while astrocytes and oligodendrocytes produce KYNA. This compartmentalization of their synthesis could play a role in the QA/KYNA imbalance.

During inflammation, more than 95% of free tryptophan is metabolized via the KYN pathway, with the remainder being utilized to produce serotonin and melatonin. Based on this finding, some researchers proposed that diverting tryptophan into the KYN pathway could result in tryptophan depletion, leading to depression. However, the ability of cells to retain tryptophan availability even in circumstances of significant tryptophan depletion and the preservation of high plasma levels of tryptophan even after continuous IDO activation has argued against this idea. IFN alpha and gamma and TNF stimulate IDO, resulting in large reductions in plasma tryptophan levels and increases in KYN metabolites in the blood. The only study examining the link between IFN-alpha-induced alterations in CSF metabolites and the onset of depression revealed that the cytokine did not affect central tryptophan levels while dramatically increasing KYN and QA levels in the CSF. The capacity of KYNs to easily pass the BBB via specific amino acid transporters is their most prominent property. As a result, KYN molecules directly link inflammation to excitatory neurotransmission and brain alterations that might cause neuropsychiatric symptoms.

Notably, changes in the KYN pathway have been targeted for neuropsychiatric treatment. Inhibitors of the KYN pathway are being developed to treat medical conditions, including cancer, that could be repurposed to treat psychiatric problems. Interestingly, blocking glutamate receptors with ketamine or inhibiting IDO activity protects mice from lipopolysaccharides (LPS)- or stress-induced depressive-like behavior but leaves the inflammatory response intact. These results indicate that increased activation of glutamate receptors by glutamate and/or QA may be a common pathway through which inflammation causes depressive-like behavior.

Macroscale Neural Correlates of Inflammatory Activation on the Brain

Studies utilizing positron emission tomography (PET) and MRI prove that peripheral cytokine activity can induce centrally mediated behavioral changes. These and other imaging modalities provide a means by which various behavioral alterations can be associated with changes in the functioning of specific brain regions. Studies on IFN-alpha, an innate immune cytokine used to treat infectious disorders and cancer, have extensively contributed to our understanding of the brain response to peripheral immune stimulation. IFN-alpha induces substantial behavioral abnormalities, including major depression in up to 50% of patients, depending on the amount and duration of therapy. Thus, IFN-alpha is a potent and reliable inducer of peripheral inflammation and its behavioral correlates (depression, anhedonia, fatigue, and cognitive impairments). During a functional MRI task of visuospatial attention, in comparison to controls, patients administered IFN-alpha exhibited significantly greater activation of the dACC that was correlated with the number of errors during task performance. Interestingly, increased dACC activity during cognitive tasks has also been demonstrated in patients vulnerable to mood disorders, such as those with high trait anxiety, neuroticism, or OCD. In addition, IFN-alpha leads to basal ganglia hypermetabolism measured using PET-uptake of 18-fluorodeoxyglucose), which was positively correlated with the severity of fatigue. In addition, impairments in DA metabolism mediated the development of fatigue following IFN-alpha. Of note, recent studies have provided further support to the selective targeting of reward circuitry by increased inflammation among depressed subjects. Using magnetic resonance spectroscopy (MRS), increased peripheral inflammation was also associated with increased glutamate signals in the dACC and basal ganglia regions among two independent cohorts (depressed patients and subjects treated with IFN-alpha) that was associated with severity of depression, anhedonia, and psychomotor slowing. Subsequently, we demonstrated that these increases in MRS glutamate signals were associated with greater incoherence in spontaneous brain oxygen level–dependent fluctuations on resting-state functional MRI and impaired functional connectivity between the basal ganglia and medial prefrontal regions. Similarly, studies from other groups have demonstrated similar brain and behavioral targets of inflammation.

SOCIAL ADVERSITY, ENVIRONMENTAL THREATS, AND THE IMMUNE SYSTEM

While stress was once thought to be linked to immune suppression, evidence over the last several decades suggests that this conclusion oversimplifies the complexities of the mammalian immune response to environmental perturbation. It is widely accepted that stress can activate certain aspects of the immune system, particularly the innate inflammatory response (see Fig. 1.13–4). Most factors that increase the risk of depression in today's environment also raise levels of peripheral inflammatory biomarkers, and many factors that improve emotional well-being have been found to reduce inflammation. A succession of animal and human studies demonstrating that stressful stimuli can impact the development of immune-related disorders such as infectious diseases, cancer, and autoimmune disorders sparked interest in the effects of the broader environment on the immune system. These findings demonstrate that peripheral T- and monocytic cell functioning—specifically the homing of these cells to the meninges of the brain in response to psychosocial stress—is required for learning in rodents in a process that drives hippocampal BDNF production via an IL-4–dependent mechanism.

Stress (Acute/Mild)

Data from laboratory animals suggest that brief or mild stressors might help animals adjust to new situations. In mouse models, brief/mild stress preceding antigen challenge resulted in an increased immunologic response. Acute stress boosts leukocyte activity by boosting leukocyte recruitment from the bone marrow into the bloodstream, infiltration into the pathology site, and proliferation at the infection site. Similar variations in NK cells and CD8 cytotoxic cell activity have been observed during stress. Acute stress appears to decrease regulatory and other suppressor T cells, according to new research. A rise in activated immune cells in the blood has been observed following acute stress such as parachute jumping, arithmetic problem solving, and confrontational human interactions. However, immunologic activation caused by acute stress is promptly decreased following the resolution of the stress. Notably, increased sympathetic activity is responsible for the acute effects of stress on the immune system, but increased parasympathetic activity is responsible for its suppression following stress resolution.

Severe/Chronic Stress

Chronic stress has been linked to immunosuppressive effects, despite its ability to increase proinflammatory responses. Chronic or severe stress is linked to increased health risks, frequently accompanied by proinflammatory alterations. In mouse models, chronic unexpected stress (predator stress, restraint stress) promotes inflammatory cell mobilization and functional activity of macrophages. These stressors work by increasing the release of DAMPs such as ATP, which activate inflammasomes leading to the release of IL-1.

Health Consequences of Chronic Stress

Allergies: Chronic stress, unlike acute stress, shifts the balance between the proinflammatory Th1 and the anti-inflammatory Th2 phenotypes, favoring the latter at the expense of the former. Thus, acute/mild stress in animals and humans may increase the in vitro production of the Th1 cytokine IFN-gamma without increasing the Th2 cytokine IL-4. In contrast, chronic/severe stressors tend to suppress IFN-gamma and IL-2 by Th1 cells, while not affecting or increasing the Th2 cytokines IL-4 and IL-10. This shift from Th1 to Th2 may also increase the susceptibility to allergic or hypersensitivity reactions, where Th2 responses play an important role. Chronic stressors in humans lead to a decrease in NK cell activity and lymphocyte proliferation in response to mitogen, and a decrease in monocyte sensitivity to endogenous cortisol. Furthermore, stress from a caregiver or an examination hinders the body’s ability to function or mount effective immunity against pathogens or immunization.

Th17 cells orchestrate tissue inflammation in autoimmune disorders (multiple sclerosis) by inducing proinflammatory cytokines (IL-6, IL-1, and TNF) and chemokines which recruit Th1 cells to the target tissue (monocyte chemoattractant protein [MCP]-1). Although functional T-regulatory cells accumulate in the target tissue, they are not efficacious during the inflammatory phase due to the overwhelming amounts of proinflammatory cytokines, such as IL-6, which protect effector and cytotoxic T cells from the suppressive effects of Treg cells. Therefore, boosting Treg cells and controlling tissue inflammation may provide the most effective strategy for controlling tissue inflammation and treating autoimmune diseases.

RELEVANCE OF IMMUNE–CNS INTERACTIONS TO PATHOLOGY AND TREATMENTS IN PSYCHIATRY

Major Depression

For many years, major depression was considered a quintessential example of how stress-related disorders may decrease immunocompetence. More recently, however, it has become evident that stress also activates inflammatory pathways, even while suppressing measures of acquired immunity. Thus, chronic stress-induced inflammatory activity may lead to depression and suppress *in vitro* functional assays, such as lymphocyte proliferation and NK cell activity. Nevertheless, despite potential negative impacts on health in the modern world, depressive behavior (with its attendant reduction in exploratory behavior) may represent an adaptive redeployment of limited metabolic energy reserves to innate immunity at the expense of adaptive immunity to promote survival across environmental challenges faced by hominids during evolution. Figure 1.13–5 depicts the key neural targets of inflammatory molecules in the etiopathogenesis of depressive disorders.

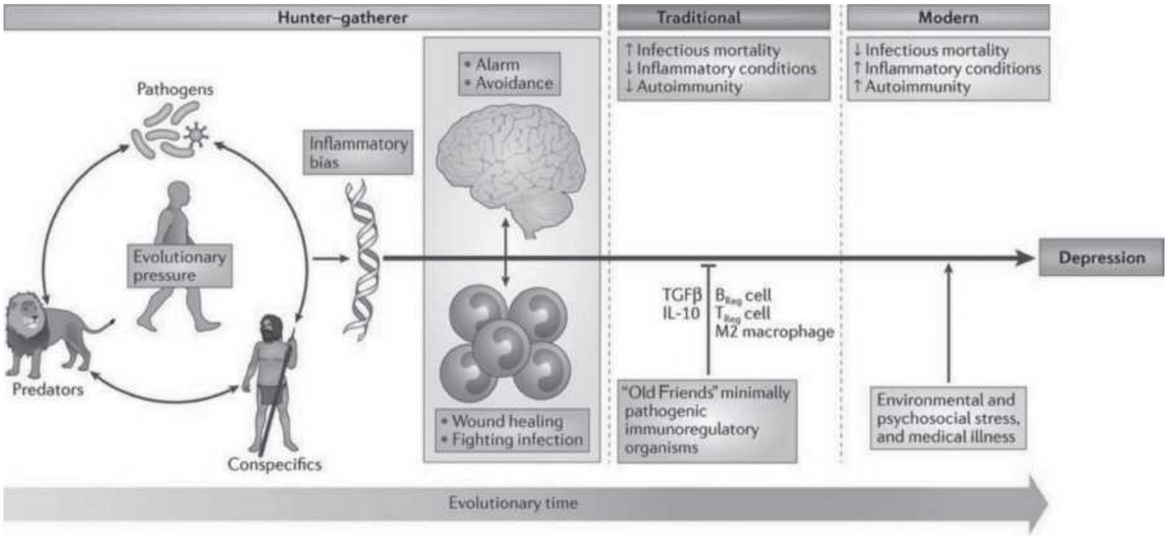
Hunter–gatherer
Traditional
Modern
Pathogens
Evolutionary pressure
Predators
Conspecifics
Inflammatory bias
• Alarm
• Avoidance
• Wound healing
• Fighting infection
↑ Infectious mortality
↓ Inflammatory conditions
↓ Autoimmunity
↓ Infectious mortality
↑ Inflammatory conditions
↑ Autoimmunity
Depression
TGFβ
IL-10
B_{Reg} cell
T_{Reg} cell
M2 macrophage
"Old Friends" minimally pathogenic immunoregulatory organisms
Environmental and psychosocial stress, and medical illness
Evolutionary time

FIGURE 1.13–5. Once in the brain, the inflammatory response can affect metabolic and molecular pathways influencing neurotransmitter systems that can ultimately affect neurocircuits that regulate behavior, especially behaviors relevant to decreased motivation (anhedonia), avoidance, and alarm (anxiety), which characterize several neuropsychiatric disorders including depression. On a molecular level, proinflammatory cytokines including type I and II interferons (IFNs), interleukin-1 beta (IL-1β), and tumor necrosis factor (TNF) can reduce the availability of monoamines—serotonin (5-HT), dopamine (DA) and noradrenaline (NE)—by increasing the expression and function of the presynaptic reuptake pumps (transporters) for 5-HT, DA, and NE through activation of mitogen-activated protein kinase (MAPK) pathways and by reducing monoamine synthesis through decreasing enzymatic co-factors such as tetrahydrobiopterin (BH4), which is highly sensitive to cytokine-induced oxidative stress and is involved in the production of nitric oxide (NO) by NO synthase (NOS). Many cytokines, including IFN-gamma, IL-1 beta, and TNF, can also decrease relevant monoamine precursors by activating the enzyme indoleamine 2,3-dioxygenase (IDO), which breaks down tryptophan, the primary precursor for serotonin, into kynurenine. Activated microglia can convert kynurenine into quinolinic acid (QUIN), which binds to the *N*-methyl-D-aspartate receptor (NMDAR), a glutamate (Glu) receptor, and together with cytokine-induced reduction in astrocytic Glu reuptake and stimulation of astrocyte Glu release, in part by induction of reactive oxygen species (ROS) and reactive nitrogen species (RNS), can lead to excessive Glu, an excitatory amino acid neurotransmitter. Excessive Glu, especially when binding to extrasynaptic NMDARs, can in turn lead to decreased brain-derived neurotrophic factor (BDNF) and excitotoxicity. Inflammation effects on growth factors such as BDNF in the dentate gyrus of the hippocampus can also affect fundamental aspects of neuronal integrity including neurogenesis, long-term potentiation, and dendritic sprouting, ultimately affecting learning and memory. Cytokine effects on neurotransmitter systems, especially DA, can inhibit several aspects of reward motivation and anhedonia in corticostriatal circuits

involving the basal ganglia, ventromedial prefrontal cortex (vmPFC), and subgenual and dorsal anterior cingulate cortex (sgACC and dACC, respectively), while also activating circuits regulating anxiety, arousal, alarm and fear including the amygdala, hippocampus, dACC, and insula. BH2, dihydrobiopterin; DAT, dopamine transporter; EAAT2, excitatory amino acid transporter 2; NET, noradrenaline transporter; NF-κB, nuclear factor-kappa-B; SERT, serotonin transporter; TH, tyrosine hydroxylase; TPH, tryptophan hydroxylase. (Reprinted from Miller A, Raison C. The role of inflammation in depression: from evolutionary imperative to modern treatment target. *Nat Rev Immunol* 2016;16:22–34. doi: 10.1038/nri.2015.5)

Inflammation and Depression—Cross-Sectional Associations. As evidenced by multiple meta-analyses, individuals with depressive disorders demonstrate elevated levels of several pro- and anti-inflammatory cytokines, including IL-2, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12, IL-13, IFN-gamma, CRP, and TNF. From this diverse set of inflammatory cytokines, the most robust and consistent associations between depression and inflammatory biomarkers have been observed for CRP, IL-6, IL-1, and TNF. Of note, the elevation of anti-inflammatory cytokines may occur to counteract the proinflammatory bias.

Inflammation and Depression—Longitudinal Associations. There is support for the conclusion that chronic inflammatory activation increases the risk of depression. Medical disorders associated with chronic inflammation, such as autoimmune disorders (rheumatoid arthritis, multiple sclerosis, lupus), infectious disorders (HIV, hepatitis, cytomegalovirus), vascular disorders (stroke, ischemic disease), and metabolic disorders (diabetes, obesity) are all associated with substantial increases in the frequency of depression. These findings were validated and supported by results from a large-scale population-based cohort study in Denmark, using survival analysis adjusted for age, birth year, and gender. The risk of mood disorders increased by 45% following diagnosis of autoimmune disorders and 62% following hospitalization for infectious diseases. Immune-stimulating treatments such as IFN-alpha are well-known inducers of depression. Multiple studies, including a recent meta-analysis, have indicated that antidepressant pretreatment may reduce the risk of developing depression in IFN-alpha–treated subjects. Finally, increased CRP and IL-6 also predicted the development of depression prospectively, with some variability across studies.

Caveats in Interpreting Inflammation–Depression Associations. Despite these associations, it is difficult to conclude whether inflammation is universally associated with depression or if depression is a neuroinflammatory disorder like multiple sclerosis. It appears that not all subjects with depression have increased levels of inflammatory cytokines, and not all subjects with elevated plasma cytokines develop depression. Inflammation is more likely to be linked to pathophysiology in a subset of depressed subjects with increased individual vulnerabilities or biologic predispositions to inflammation-induced depression. Another factor impacting this association involves confounding factors such as obesity, elevated blood pressure, and diabetes, which are common among depressed subjects. It is unclear if such associations between inflammation and depression exist after controlling for these confounding factors. For example, one study that controlled for waist circumference and blood pressure could not replicate the previously reported associations between depression and increased inflammatory markers in the plasma. Intriguingly, some studies have reported high correlations between plasma and CSF levels of inflammatory markers such as CRP, TNF, and IL-6, only among untreated depressed individuals after controlling for confounding factors, including obesity as reflected by body mass index (BMI). Furthermore, a meta-analysis of 69 studies also reported an increase in IL-6 and TNF concentrations in the CSF of depressed patients.

Association Between Inflammation and Domains of Depressive Symptoms. The extreme heterogeneity in the diagnosis of depression has severely hampered the search for biomarkers related to its development, progression, and treatment response. One recent approach has been to narrow the diagnostic entity by breaking it down into subsyndromal symptom domains. Some key domains of interest relevant to inflammation's effects on the brain and behavior include anhedonia, anxiety, suicidality, cognitive symptoms, and sleep—although these symptom domains may extend into diagnostic categories other than depression in a "transdiagnostic manner." Indeed, as noted above, increased inflammation has been reliably associated with effects on neurotransmitter systems and brain regions and circuits that regulate motivation and motor activity as well as anxiety, arousal, and alarm. The specificity of these effects suggests a narrower influence of inflammation on the wider universe of symptoms exhibited by patients with mood and anxiety disorders. Indeed, in a recent paper examining the relationship between inflammation and specific symptoms of depression in 56,351 subjects from 15 population-based cohort studies, anhedonia, fatigue, and sleep problems were reliably associated with inflammation. These data are very consistent with what we know about how inflammation affects the brain. Nevertheless, a detailed review of this approach is beyond the scope of this chapter, but the readers are referred to exhaustive reviews on these topics.

Inflammatory Biomarkers as Predictors of Responses to Conventional Antidepressant Treatments. Identifying biomarkers that can predict treatment response is crucial in clinical practice if relevant lab testing is low cost and not time intensive. Ideally, these biomarkers would help categorize patients into subgroups (depressed subjects with varying biomarker profiles) and suggest the best treatment options for subjects who are unresponsive to traditional antidepressant drugs. In summary, current evidence suggests a negative relationship between increased inflammation and clinical response to classical antidepressant therapy, notwithstanding encouraging data using ketamine and ECT (to be discussed later). Some commonly studied markers are examined below.

TUMOR NECROSIS FACTOR. The lack of response to antidepressant treatment among TNF-gene knockout mice suggested that antidepressants require some TNF to be effective. This preclinical finding has not been fully supported in human studies. Responders to antidepressant treatment with escitalopram had lower baseline TNF levels than nonresponders to healthy controls. SSRI-resistant depressive subjects demonstrated significantly higher levels of IL-6 and TNF in the plasma. In a recent study using 98 depressed subjects who had failed several antidepressant trials, we reported that subjects who failed three or more treatment trials demonstrated significantly higher TNF, sTNFR2, and IL-6 than subjects who failed fewer treatment trials. In addition, plasma levels of the three markers predicted treatment trial failures with significant accuracy (AUC >0.7). Increased TNF activation has generally formed the basis for targeting the immune system in depression using TNF-blocking monoclonal antibodies such as infliximab.

INTERLEUKIN-6. IL-6 has been consistently linked to treatment resistance in depression since early 2000. Compared to healthy controls, depressed patients who did not respond to SSRI therapy had higher baseline levels of IL-6 and TNF. Surprisingly, inflammatory markers such as IL-6 predict a good response to ketamine. An earlier report indicated that ketamine responders exhibited higher plasma IL-6 levels than nonresponders. Intriguingly, preclinical treatment-resistant depression models tended to indicate comparable findings. Similarly, a more recent larger analysis with more structured control of confounding factors appeared to support this conclusion from previous studies. In this study, baseline IL-6 predicted a greater reduction in depressive symptoms at the end of the study. More recently, another study has demonstrated that elevated KYN metabolites also predicted a positive response to ketamine, indicating that ketamine benefits may be more pronounced among patients with inflammatory activation. Along similar lines, higher baseline IL-6 levels predicted greater response to a course of ECT, an effect that was moderated by female sex. Intriguingly, early treatment increases in IL-6 and CRP during treatment fell to normal levels at the end of the study. In summary, circulating levels of IL-6 predict treatment response depending upon the type of treatment, with elevated IL-6 being associated with the nonresponse to SSRIs or SNRIs and positive response to ketamine and ECT.

INTERLEUKIN-8. IL-8 is a chemokine and proinflammatory cytokine that has demonstrated association with the severity of anxiety and depression scores, especially among females. Recent data indicate that baseline levels of IL-8 predicted treatment response to an index course of ECT, specifically in females. For instance, female responders to a course of ECT treatments demonstrated lower IL-8 levels at baseline. However, unlike IL-6, decreases in IL-8 from baseline to end-of-treatment were negatively correlated with the percentage of symptom improvement only among females. Thus, IL-8 may be a sex-specific marker of treatment response to ECT in major depression. Another study reported that IL-8 predicted response to ketamine among female depressives.

C-REACTIVE PROTEIN. A secondary analysis from the GENDEP study (a multicenter, partially randomized open-label pharmacogenetics study) examined the impact of baseline CRP in predicting treatment response among depressed patients randomized to escitalopram or nortriptyline for

12 weeks. Baseline CRP levels predicted treatment outcomes for antidepressants differently. Patients with CRP > 1 mg/L responded more favorably to nortriptyline treatment than escitalopram. Similar findings were reported from the COMED trial, where a CRP > 1 predicted a poor response to escitalopram and a good response to escitalopram plus bupropion. Another more recent study indicated that baseline peripheral CRP levels were higher among depressed subjects who were treatment resistant than those who responded.

INTERLEUKIN-1. Two genotypes of the single-nucleotide polymorphisms (GG genotype) of the IL-1 gene were associated with nonremission after 6 weeks of antidepressant treatment. In addition, an increase in the number of G-alleles in both SNPs (G-, GG) was associated with decreases in the responsiveness of the amygdala and ACC to emotional stimulation.

KYNURENINE PATHWAY. A postmortem study reported that microglia in the anterior cingulate cortex of acutely depressed, suicidal patients expressed greater QA levels. These increases in QA and its downstream effects on NMDA receptor modulation have provided a molecular basis to examine ketamine (a noncompetitive NMDA receptor antagonist) as an antidepressant therapy among depressed subjects with elevated inflammation. Intriguingly, initial data have also linked KYNs to the outcome of ketamine therapy for depression. A recent analysis has indicated that KYN pathway toxicity may underlie some neurocognitive deficits following Covid-19 infections.

In conclusion, increases in peripheral inflammatory markers may likely be useful biomarkers to identify subgroups of patients with differential responses to treatments. However, the choice of immunologic biomarkers will be limited by the nonavailability of standardized assay techniques for all immune biomarkers other than CRP. In addition, the high variability in the concentrations of these markers over time and redundancy in functional activity may impact the interpretation of immune measures. Nevertheless, it is possible to use CRP as a baseline screening tool to stratify inflammatory risk among depressed subjects—as routinely done in other disciplines such as cardiology.

Anti-Inflammatory Therapies for Depressive Disorders

NONSTEROIDAL ANTI-INFLAMMATORY AGENTS. Inhibition of enzymes such as cyclooxygenase-1 and 2 (COX-1 and COX-2, respectively) using nonsteroidal anti-inflammatory agents (NSAIDs) have been evaluated as potential treatments for depression. Both COX enzymes lead to the synthesis of prostaglandins that promote inflammation while protecting gastrointestinal and vascular smooth muscles. NSAIDs are classified into nonselective agents, including aspirin that acts on COX-1 and COX-2 and celecoxib that are selective COX-2 inhibitors.

NSAID Monotherapy. Despite initial enthusiasm, monotherapy with NSAIDs has so far yielded mixed results. The earliest study in this area reported that Australian men who had used aspirin in the past 5 years but not at the time of the study had 40% higher odds of experiencing depressive symptoms. Thus, the authors concluded that aspirin failed to protect against developing depression. However, aspirin was likely prescribed to target chronic medical conditions associated with increased inflammation and depression. A later study indicated that a combined intake of statins and aspirin reduced the risk of depression. However, statins, via their anti-inflammatory effects, may have confounded the immunoprotective effect of aspirin. Another study indicated a relationship between aspirin intake and reduced frequency of depression among older males with high levels of homocysteine in the plasma—signifying increased vascular inflammation. A study of depressed older individuals from an Alzheimer disease (AD) prevention trial comparing celecoxib 200 mg twice a day or naproxen sodium 220 mg twice a day with a placebo indicated that neither celecoxib nor naproxen group experienced significant remission of depressive symptoms. Data from individuals receiving ibuprofen, naproxen, and celecoxib for the treatment of osteoarthritis experienced a significant reduction in the intensity of depressive symptoms. However, the limitation of disability and pain from osteoarthritis may likely have contributed to improvements in depression scores. More importantly, the benefits of NSAIDs are often outweighed by the risks of renal, gastrointestinal, and cardiovascular side effects. Notably, cardiovascular toxicity following treatment with rofecoxib (Vioxx) was noted as early as 2 months following the initiation of treatment. In this context, a very large study published recently found no protective effect of aspirin in depression.

NSAID Augmentation Therapy. The efficacy of NSAIDs as augmentation agents to antidepressant therapy has also been mired in mixed results. An early study used a double-blind, placebo-controlled randomized design to compare celecoxib plus reboxetine (antidepressant), reboxetine alone, and placebo. While both the reboxetine alone and reboxetine plus celecoxib groups achieved a reduction in symptom severity, the substantial number of subjects dropping out of the celecoxib group implied intolerance to the side effects of this medication. Subsequently, two studies indicated positive benefits with 6-week celecoxib augmentation of fluoxetine. However, another study using the same universe of subjects indicated that the augmentation effect of celecoxib peaked at 4 weeks and waned off at 8 weeks. The authors concluded that celecoxib may accelerate response to SSRIs. Using baseline inflammatory markers such as PET labeling of translocator protein density on the glial cells predicts better response to celecoxib in depressed individuals. However, a recent, well-conducted study indicated that augmentation with celecoxib was no better than a placebo in augmenting the antidepressant efficacy of vortioxetine. Thus, the overall efficacy of NSAID monotherapy or augmentation is not well supported by current data. Several factors, including the selection of patients without increased inflammation, global outcome variables such as changes in depression rating scale scores, and using drugs with off-target effects, were implicated in trial failures. A large population-based study from Korea reported that a combination of NSAIDs with antidepressants increased the risk of intracranial hemorrhage with a hazard ratio of 1.6 (95% confidence interval = 1.3 to 1.85), raising serious concerns about NSAID augmentation.

CYTOKINE INHIBITION. Given the link between some depressions and inflammatory cytokines, recent research has focused on examining the antidepressant efficacy of cytokine-blocking agents. TNF-antagonist such as infliximab were among the earliest to arrive on the scene. In 2013, 60 medically stable patients with treatment-resistant major depression were treated with three infusions of the TNF antagonist infliximab (5 mg/kg) or placebo (n = 30 in each group) at baseline, weeks 2, and 6 of a 12-week trial. While the overall change in depression severity scores using Hamilton Depression Rating Scale did not differ between treatment groups, there was a treatment by time by group interaction, whereby subjects with a high baseline CRP level >5 mg/L exhibited a clinically meaningful response in favor of infliximab. In addition, baseline TNF and its receptors were significantly higher in infliximab responders versus nonresponders. Intriguingly, the extent of adverse events and dropouts was very few in contrast to studies with NSAIDs. This article pointed out several factors of relevance. While providing proof-of-concept evidence linking TNF-antagonism and antidepressant activity, the authors stressed significant caution in interpreting the results. Some of the limitations included the small sample available for exploratory analysis, limited generalizability of CRP >5 mg/L to most depressed patients in clinical practice, and the global measure of depression outcome that was used. Of note, the use of TNF-antagonists (infliximab, adalimumab, etanercept) to treat medical disorders such as psoriasis has also resulted in substantial improvement of comorbid depressive symptoms. These findings are subject to cautionary interpretation as the improvement in depression may have resulted from improvement in disability and suffering from disease activity. Treatment with TNF antagonists increases the risk of opportunistic infections such as tuberculosis, histoplasmosis, and bacterial sepsis. Recently, siltuximab (chimeric monoclonal antibody against IL-6) for multicentric Castleman disease and sirukumab (human monoclonal antibody against IL-6) for the treatment of rheumatoid arthritis were evaluated for their antidepressant effects in these medically ill cohorts. Treatment with the above two IL-6 antagonists reduced the prevalence of depression and anhedonia in this analysis, though limited by the confounding effects of primary medical conditions and their treatment. Finally, a recent trial of infliximab for bipolar depressed patients (see below) found similar effects on anhedonia but without a clear relationship with baseline inflammation. Nevertheless, a history of childhood maltreatment, which has been reliably associated with inflammation, was found to be predictive of response.

KETAMINE. A detailed discussion of ketamine is available in other chapters of this textbook. Its relevance to inflammatory activation is a source of interest for this chapter. Earlier sections outlined changes in immune markers with ketamine therapy. Despite long-standing knowledge that ketamine is an antagonist at the NMDA receptor site, it is far from clear if its mechanisms are limited to NMDA blockade. Notably, ketamine reduces IDO activation and decreases the KYN/tryptophan ratio in the

CNS. Ketamine also stimulates the mammalian target of rapamycin (mTOR), which plays a major role in maintaining synaptogenesis. Intriguingly, ketamine has also been shown to reduce IL-6, TNF, and IL-1 levels in preclinical depression models. Furthermore, ketamine has been shown to reduce levels of NF-kB, the transcription factor that promotes inflammation in depression.

MINOCYCLINE. Minocycline is a microglial inhibitor and exerts anti-inflammatory effects in the brain by inhibiting the production of proinflammatory cytokines and reactive oxygen species released by microglia in the brain. In addition, minocycline is well known to decrease IDO activity and reduce the synthesis of neurotoxic KYN metabolites by microglial cells. An early study reported that oral intake of minocycline 100 mg twice a day significantly reduced depression severity in a group of subjects with HIV infection. Subsequently, minocycline 150 mg/day in combination with antidepressants was reported as effective and well tolerated in the treatment of unipolar psychotic depression. However, despite evidence of the efficacy of minocycline in depressed patients with a CRP >3 mg/L, oral minocycline at similar doses was not found to reduce TSPO labeling in major depressive disorder. Several recent studies have evaluated the efficacy of minocycline in treating depression and other mental disorders with disappointing results (to be discussed under bipolar disorders). Similarly, recent large-scale efforts to examine the effect of minocycline in treating bipolar depression have not yielded meaningful results.

Antidepressant Effect on Cytokines

PRECLINICAL. *In vitro* analyses of human blood samples have yielded variable findings that are difficult to interpret, depending on the antidepressant used. Venlafaxine decreased IL-6 and sertraline increased IFN-alpha and IL-6, both associated with decreases in NF-kB activity. Both imipramine and fluoxetine decreased the release of TNF and IL-1 by rat glial cells subjected to immune stimulation using in vitro application of LPS mediated by decreases in NF-kB activity. Amitriptyline led to decreases in serum levels of IL-1 and TNF in preclinical peritonitis models by decreasing nitric oxide production. Paradoxically, citalopram increased the production of IL-1, IL-6, TNF, and IL-22; mirtazapine increased IL-1, TNF, and IL-22; and escitalopram decreased IL-17 production by immune-stimulated T and B lymphocytes from depressed individuals. Reviews on this area attributed the high variability to the type of drug used and the method of immune stimulation used. However, it does appear that antidepressants as a class may converge on the intracellular immune signal transduction pathways mediated by NF-kB.

CLINICAL STUDIES. Measures of plasma inflammatory markers from clinical samples of treated depressed subjects have also added to our understanding. An early meta-analysis failed to support a universal relationship between antidepressants and immune markers. Instead, there was a class-specific effect with SSRIs impacting IL-6 and both SSRIs and TCAs impacting IL-1. A recent meta-analysis of 35 longitudinal clinical studies using plasma measures of cytokines reported net decreases in IL-6 levels resulting from antidepressant treatment.

Schizophrenia

Innate Immune Changes in Schizophrenia. Increased inflammation is generally not associated with the development of positive psychotic symptoms such as those encountered in schizophrenia. Studies using circulating immune marker assays do not help distinguish schizophrenia from affective disorders. A meta-analysis of over 85,000 people found modest but reliable increases in CRP among schizophrenic subjects. Furthermore, CRP levels were moderately increased in persons with SZ independent of antipsychotics or the course of the illness (first episode vs. progressive illness), and an increase in peripheral CRP levels paralleled the increase in severity of positive symptoms but not negative symptoms. However, a more inclusive meta-analysis reported that elevated plasma levels of two cytokines (IL-6, TNF), one soluble cytokine receptor (sIL-2R), and one cytokine receptor antagonist (IL-1RA) were increased in acute schizophrenia, bipolar mania, and major depression. Finally, a recent meta-analysis confirmed the relationship between increases in inflammatory markers (IL-6, IL-8, and IGg/albumin ratio) in the CSF of patients with schizophrenia and affective disorders. These authors pointed to the difficulty in arriving at a firm conclusion due to bias and poor control of confounding factors. However, another meta-analysis controlling for variability and confounders indicated moderate effects for IL-6, IFN-gamma, and IL-17 (Hedges g 0.3 to 0.6 for mean differences). Finally, a meta-analysis of studies examining translocator protein (TSPO)-PET binding in schizophrenia reported moderate elevations in TSPO tracer binding in grey matter relative to other brain tissue in schizophrenia when using binding potential as an outcome measure, but no difference when the volume of distribution is the outcome measure. Thus, variability in marker estimations, bias in sample selection, and poor selection of confounding factors cloud the picture. However, studies of first-episode psychosis have yielded more consistent findings. A recent meta-analysis using 3,453 drug-naive first episode versus 4,152 control subjects reported that elevated IL-6, IL-8, and TNF were seen among first-episode psychotic subjects. Given the rarity of reports on first-episode subjects, this approach of pooling across multiple studies may be the most pragmatic approach.

Adaptive Immune Changes in Schizophrenia. A detailed discussion of the adaptive immune cell change in schizophrenia is beyond the scope of this chapter. The findings are often variable and contradictory. However, all these studies underline that alterations in T and B lymphocytes and other adaptive immune cells may be integral to the pathology of some patients with schizophrenia. An increasing cytotoxic, proinflammatory bias (as opposed to regulatory T-cell activity) is noted in the peripheral blood of susceptible individuals. Intriguingly, the pathology of B lymphocytes may be related to the expression of autoimmune encephalitis linked to psychotic symptoms. In addition, transcriptional profiling of postmortem brain specimens from individuals with schizophrenia indicated that genes associated with innate immune (cytokine, chemokines), postsynaptic signaling, and glutamate pathways were enriched. Other studies have reported upregulated expression of microglial activation, cytotoxic lymphocyte infiltration, and IgG deposition.

Maternal Immune Activation. While human studies have demonstrated the statistical link between exposure to maternal inflammation in schizophrenia, animal studies have revealed that exposure to chronic maternal inflammation is associated with persisting brain morphology and functional changes in the offspring. For example, various immune stimuli (activating both antiviral and antibacterial immunity) administered to pregnant lab animals result in aberrant developmental trajectories comparable to schizophrenia in the offspring. These behavioral deficits include deficient social interaction, repetitive and stereotyped behaviors, poor working memory, impaired executive functioning, and impaired sensorimotor gating as assessed via prepulse inhibition. In addition, some pathogenic viruses may act by triggering cell-death cascades in dopaminergic neurons, resulting in fluctuating DA levels and behavioral symptoms of schizophrenia. Multiple epidemiologic studies have linked *in utero* exposure to infectious pathologies with the development of schizophrenia later in life. Prenatal infections that have been associated with schizophrenia include rubella, influenza, and toxoplasmosis. Similarly, obstetric complications, gestational diabetes, and severe physical or psychosocial stressors (e.g., famine, war, psychological trauma, maternal depression) can increase maternal inflammation, which in turn predicts the development of schizophrenia. Maternal cytokines, including IL-8, are also significantly increased in pregnancies leading to the development of schizophrenia. For example, fetal exposure to increases in IL-8 was associated with significant increases in ventricular CSF, significant decreases in left entorhinal cortex volumes, and significant decreases in right posterior cingulate volumes among subjects who developed schizophrenia later in life. In addition, elevations in maternal TNF have also been linked to the later development of schizophrenia. Intriguingly, strategies to block inflammatory cytokines such as IL-1, IL-6, and TNF reduced the risk of behavioral and cortical deficits associated with schizophrenia-like behavior in laboratory animal models. Similar results have been obtained by strategies that increase the synthesis and release of the anti-inflammatory cytokine IL-10 in animal models using poly IC. Thus, schizophrenic disorders in some subjects may represent the outcome of immune activation and inflammatory pathologies in the developing brain.

Autoimmune Psychosis. A study using the Danish National Registry reported that previous history of autoimmune disorders increased the risk of later schizophrenia by 29% and a history of hospitalization for infection increased the risk to 60%. Furthermore, when autoimmunity was combined with infection, the risk of subsequent schizophrenia increased by 125%, even after controlling for confounding factors. In this context, the role played by autoimmunity in promoting risk for psychosis has been emphasized. More specifically, pathologic activation of neurotransmitter receptors by adaptive immune response has offered a possible avenue to study immune dysregulation in psychosis. The most dramatic presentation involves the clinical and laboratory features suggestive of autoimmune

NMDA receptor encephalitis. Autoimmune encephalitis is best diagnosed in the presence of clear neurologic features, including seizures, cognitive dysfunction, or movement disorders. In addition, these patients have pathogenic antibodies that target surface epitopes on synaptic and related proteins—principally the NMDA receptor system, voltage-gated potassium channel (VGKC)-system, leucine-rich-glioma inactivated (LGI)-1, and contactin-associated protein-like (CASPR)-2 systems. A detailed description of these syndromes and pathologies is provided elsewhere. Identification of receptor-specific antibodies on activated lymphocytes and characteristic protein bands in the CSF may be of diagnostic value. MRI findings, including medial temporal hyperintensities on T2-FLAIR images and increased or decreased uptake of FDG-PET in temporal and parietal regions, have also been reported. Finally, EEG changes, including diffuse slowing, intermittent rhythmic delta or theta activity, or overt epileptogenic activity, have been used in the diagnostic workup of these disorders.

Even though the study of neuroimaging and EEG markers indicates a specific vulnerability of medial temporal regions in the disease process, clinical evaluation remains the cornerstone in the diagnosis of these disorders. If the antibody activity is associated with an occult tumor (such as a thymoma), its removal may help alleviate the symptoms. In addition, immunotherapies such as plasma exchange, immunoadsorption, or intravenous immunoglobulins have all been tried with varying levels of success. Immunosuppressive therapies such as high-dose steroids, methotrexate, or azathioprine have also been used in acute treatment. Anti-CD20 agents that reduce B lymphocyte activity, such as rituximab, have also met with some success. Finally, there may be a role for the modulation of the IL-27 system by altering the function of Th27 cells.

Immune Therapies for Schizophrenia. The reader is referred to extensive meta-analyses and reviews on this area. This section will provide a summary of this burgeoning area of research. In general, most studies have found modest benefits of add-on anti-inflammatory agents in treating psychotic disorders. However, sophisticated meta-regression analyses indicated that smaller studies with limited samples exerted substantially larger effect sizes, diminishing the confidence of pooled estimates

NONSTEROIDAL ANTI-INFLAMMATORY DRUGS. Two initial studies that used high doses of aspirin (1,000 mg/day) for 3 to 4 months demonstrated modest efficacy with an effect size of ~0.30. However, this is an area that raises significant risk versus benefit concerns. Furthermore, aspirin may not readily cross the BBB and is well known to have off-target effects in reducing hypothalamic–pituitary–adrenal (HPA) axis activity. The studies that used celecoxib 400 mg/day for 5 to 11 weeks yielded low effect sizes (0.15) in the treatment of schizophrenia. Thus, the use of NSAIDs is not well supported in the literature.

ESSENTIAL FATTY ACIDS. In addition to neuroprotective effects, omega-3 fatty acids such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) have moderate anti-inflammatory effects on IL-1, TNF, and IFN-gamma. Daily doses of fatty acid with the EPA doses ranging from 0.5 to 4 g combined with DHA doses of 0.4 to 2.2 g have been studied for durations of 8 weeks up to 2 years. Most studies point to mild benefits with an effect size of 0.23. Nevertheless, some subjects may likely benefit from the addition of essential fatty acids.

ESTROGENS. It is well known that estrogens modulate immune activation through a variety of mechanisms. Some proposed mechanisms include increasing antioxidant capacity, improving glucose homeostasis, and possibly decreasing DA dysregulation. Estrogen doses ranging from 0.05 mg per day to 2 mg per day have been tried with some benefit. Pooled effect sizes have ranged in the 0.5 range among both female-only and bisexual groups. However, other risks from chronic estrogen therapy, such as increases in cancer risk, may need to be balanced against the benefits.

MINOCYCLINE. Minocycline has been reported to have benefit in some studies. However, a large study reported substantially negative results, limiting the confidence in prescribing minocycline. In addition, given its antibiotic characteristics, adverse effects on the microbiome have led to concerns about its widespread use. Therefore, much more evidence is necessary before recommending the use of minocycline.

N-ACETYLCYSTEINE. N-acetylcysteine (NAC) is a well-studied antioxidant molecule that may have a modulatory effect on inflammation, oxidative stress, and glutamate metabolism and can potentially cross the BBB.

Doses in the range of 600 to 3,600 mg/day have been used for periods ranging from a few weeks to a year. Its efficacy in chronic schizophrenia is similar to other immunomodulatory interventions with effect sizes in the ~0.4 range. A greater utility may be seen in first-episode schizophrenia. However, we note that there are few studies of first-episode subjects, and more data will need to be acquired before confidently prescribing this agent.

DRUGS TO TREAT METABOLIC SYNDROME. Metabolic syndrome and diabetes are commonly associated with increases in proinflammatory activity. Targeting metabolic dysregulation using regulators of glucose metabolism has been proposed as a viable strategy to reduce immune activation in some psychotic patients. Pioglitazone is a peroxisome proliferator-activated receptor (PPAR)-gamma inhibitor with anti-inflammatory activity and glucose-regulating effects. Pioglitazone has been shown to demonstrate beneficial effects on the course of schizophrenia in one study. Similar experiences with modulators of insulin sensitivity, such as metformin, have also been reported. Finally, modulators of lipid metabolism, such as statins, have also targeted inflammatory activation in some cases. Limited trial data indicate that statins may confer modest benefits in targeting schizophrenic symptoms. One should note that metabolic syndrome, diabetes, and lipid dysregulation are all common among schizophrenic subjects where the mainstay treatment, atypical antipsychotics, are well known for their contribution to these pathologies. Thus, it is highly likely that subjects will be exposed to these medications at some point during their lifetimes.

Bipolar Disorder

The link between immune activation and bipolar disorder has generally followed research on the links between inflammation and major depression. Several studies on major depression have included both unipolar and bipolar depressions. Accordingly, the increases in circulating inflammatory markers, activation of the KYN pathway, and inflammation-associated dysregulation in glutamate and DA systems have been all reported in bipolar disorders. However, this notion may have been oversimplified, and differences may exist in the pathophysiology of these two disorders.

Bipolar versus Unipolar Differences. For example, both unipolar and bipolar subjects may experience greater increases in inflammatory Th1 with concurrent decreases in anti-inflammatory Th2 and Th17 cells. However, while the T-cell imbalance persisted in MDD, they may be temporarily restored in BD until the next relapse. Indeed, gene expression studies indicate that increased expression of inflammatory genes precedes the onset of clinical symptoms but subsides following symptomatic remission in BD (as opposed to unipolar depression). The precise mechanism that underlies these differences is not clear. However, the common risk in proinflammatory gene activation shared by both MDD and BD may be related to environmental factors such as childhood adversity, age of onset, duration of illness, and frequency of episodes.

Staging or Neuroprogressive Models. Each recurrence or relapse of bipolar disease increases the risk of subsequent episodes, each with greater severity, a poorer likelihood of remission, and greater cognitive/neurologic dysfunction—a process referred to as kindling. This empirical model is referred to as the “staging or neuroprogressive model.” Immune dysfunction has been invoked as a possible etiologic factor in promoting kindling. Intriguingly, much of the data supporting the neuroprogressive model has been focused on innate immune factors. Two recent meta-analyses examined the association between plasma CRP and bipolar disorder. The first study reported that plasma CRP concentrations were substantially elevated during acute manic episodes and moderately during euthymia but not depressed periods. The second study reported increases in plasma CRP to be elevated during all three phases of illness—mania, depression, and euthymia. Increased circulating levels of inflammatory markers, including TNF, IL-6, soluble receptor antibodies for IL-1 (IL-1RA), and sIL-2R, have been associated with acute bipolar symptoms in some but not all studies. One recent study reported that circulating TNF markers may partially mediate the association between prior severe mood episodes and executive dysfunction in BD.

Allostasis and Neuroprogression. It has been proposed that chronic immune activation following repeated episodes of BD may result from increasing “allostatic load.” Briefly, allostatic load represents a buildup of the adaptive costs of chronic, recurrent stress. More work remains to be done to understand how recurrent episodes increase the allostatic load, contribute to brain pathology, and promote cognitive dysfunction.

Immune System and Bipolar Psychopharmacology. Several immunotherapies have been examined in clinical trials. However, the data should be interpreted very cautiously due to the small number of studies, small sample, and diversity of agents used. A summary is presented below.

N-ACETYLCYSTEINE. A major pathway to neuroprogression may involve the activation of oxidative stress by immune molecules. As a co-adjuvant agent, NAC has shown positive results in all phases of bipolar disorder, including acute treatments. An earlier study demonstrated that NAC treatment resulted in ~8 point reduction in the Montgomery–Asberg rating scale for depression. Subsequent studies have reported conflicting results. A recent study comparing NAC, NAC plus aspirin, aspirin, and placebo reported a better response rate for NAC plus aspirin than other groups. This area opens a promising new therapeutic target for the treatment of bipolar disorders.

CYTOKINE ANTAGONISTS. A major recent study examined the effect of adjunctive intravenous infliximab for treating bipolar depression with high inflammation. The elevated inflammatory load was defined using plasma CRP (>5 mg/L) or “phenotypic” inflammation (history of obesity, diabetes type I/2, inflammatory bowel disease, rheumatologic disorder, cigarette smoking, or migraine headaches). This study reported a negative result for infliximab compared with placebo. However, an association between infliximab responders and early life stress was reported demonstrating subgroup effects within the larger cohort. This study reinforced the value of narrowly defining inflammation based on immune markers. Of note, a recent meta-analysis that summarized this nascent area of research indicated that cytokine antagonists as a group were moderately beneficial through targeted group selection.

OMEGA-3 FATTY ACIDS (DHA/EPA). Despite initial enthusiasm, several controlled studies investigating the use of omega-3 fatty acids in BD have trended negatively. While meta-analyses have indicated modest effect sizes, follow-up meta-regression studies indicated that studies with small samples strongly influenced the pooled effect size. Notably, studies using small samples or open-label design have demonstrated significant effects. To date, most of the studies have examined the efficacy of omega-3 fatty acids in the depressed phase of bipolar disorder. Thus, these findings require replication in larger samples.

MINOCYCLINE. An earlier meta-analysis reported a significant pooled effect size for the use of minocycline to treat bipolar depression. Subsequently, a large-scale study was conducted to examine the effects of anti-inflammatory agents in bipolar depression. The study compared varying combinations of aspirin 81 mg twice a day, minocycline 200 mg twice a day, and placebo. The aspirin plus minocycline group demonstrated significant improvement in clinical endpoints along with a significant three-way interaction between the two drugs and plasma IL-6. While these effects indicate possible beneficial effects of minocycline, these authors clarified that the effect of remission might have been primarily driven by aspirin, as both arms with aspirin benefited most. However, a large-scale study described earlier did not indicate the benefit of either celecoxib or minocycline or their combination. Thus, more investigation is required to enable clinicians to determine risk-benefit considerations.

PIOGLITAZONE. Proof-of-concept studies have indicated an antidepressant efficacy of pioglitazone in the acute treatment of bipolar depression. However, larger clinical trials may be necessary to establish the risk benefit of pioglitazone, which has a substantial side effect burden.

Anti-Inflammatory Effects of BD Treatments. Antimanic medications such as lithium are known to exert a variety of anti-inflammatory effects. Some of the anti-inflammatory effects attributed to lithium include attenuation of COX-2, inhibition in the synthesis of IL-1, and inhibition of NF k-B. The latter effect is believed to be mediated via the effects of lithium on glycogen synthase kinase (GSK)-3 beta. Furthermore, lithium may also decrease microglial activation and expression of TLR-4. In addition, valproate may also inhibit the same site—TLR-4.

Conclusion. So far, the psychopharmacologic trials of anti-inflammatory agents to treat bipolar disorder have progressed in tandem with the advances in our understanding of the staging model. In general, moderate effect sizes ~0.4 have been reported suggesting moderate antidepressant effects across a wide range of anti-inflammatory medications. However, most studies did not stratify patients based on inflammation. Thus, more studies are required to address the utility of immune-modifying agents on bipolar mania and upon the course of illness specifically in the context of increased inflammation.

Autism

Immune dysregulation in autism is believed to result from a confluence of prenatal and postnatal risk factors. Maternal immune activation (MIA) resulting from maternal infections and presence of maternally derived antibrain autoantibodies found in ~20% of mothers constitute some of the prenatal risk. The postnatal environment on the other hand is characterized by somewhat distinct profiles of immune dysregulation, inflammation, and endogenous autoantibodies.

Autism has been associated with several immune abnormalities, including unbalanced Th1/Th2 cytokine production, reduced NK and T-cell activation, and increased TNF and IL-12, especially in patients with comorbid gastrointestinal symptoms. In addition, patients with autism exhibit increased autoimmune-based genes, including HLA-DR-Bβ1*04 and the complement C4B null allele. More recent meta-analytic studies in autism spectrum disorders report increased circulating concentrations of IL-1, IL-6, IL-8, IFN-gamma, and MCP-1 and reduced anti-inflammatory cytokine TGF-beta. Still, while a convincing case can be made for a significant immune component to autism, the relationship of immune abnormalities to the neurobehavioral symptoms of the disease remains controversial. In this context, recent epidemiologic studies have not substantiated the claim that childhood vaccines trigger autism, and immune-based therapies for autism have not been reliably effective. However, a large epidemiologic study from Denmark found that *in utero* exposure to maternal influenza infection was associated with a twofold increased risk of autism, while prolonged episodes of maternal fever increased the risk by threefold. Similarly, maternal autoimmune diseases have been shown to increase the risk of autism in children by 34%.

Animal studies also support the possibility that inflammatory activation either in utero or early in life may increase the risk for autism. For example, exposure of pregnant macaques to low-dose LPS resulted in significantly increased brain volume in the affected offspring, a finding repeatedly observed in human children with autism. As stated earlier, it is increasingly recognized that immune processes play central roles in brain development. Perhaps the most striking example of the power of immune processes to guide and disrupt neural development in ways relevant to autism comes from a mouse model of MIA that produces various behaviors in offspring relevant to autism in humans. Prenatal immune stimulation reduced the number of circulating anti-inflammatory Th2 type T-regulatory cells while increasing IL-6 and IL-17 production by Th1 type cells. Transplantation of immunologically normal bone marrow cells into MIA offspring whose native immune cells had been ablated using irradiation reversed autism-like behavioral changes.

Other Disorders

Far less attention has been given to the role of immunity in general, particularly inflammation in the pathogenesis of neuropsychiatric conditions other than major depression, bipolar disorder, schizophrenia, and autism. Nonetheless, increasing evidence suggests that increased inflammation is a hallmark of nearly all types of psychiatric disturbance. Indeed, it is intriguing that no currently defined psychiatric condition has been reliably associated with reduced levels of inflammatory biomarkers.

Post-Traumatic Stress Disorder. Given the well-characterized relationship between stress and the immune system, it is easy to extrapolate that a similar relationship would exist between PTSD and inflammatory activation. Compared with healthy controls, PTSD patients had significantly higher IL-1, IL-2, IL-6, IFN-gamma, and TNF-alpha with moderate to strong pooled effect sizes. Less strong associations were noted between PTSD and CRP and WBC counts. Intriguingly, these studies noted no relationship between PTSD and adaptive immune system or anti-inflammatory cytokines. Individuals exposed to childhood trauma had significantly elevated baseline peripheral levels of CRP, IL-6, and TNF-alpha. Subgroup analyses for specific types of traumas (sexual, physical, or emotional abuse) revealed that the type of trauma impacts individual inflammatory markers differentially. Age, BMI, and gender had no moderating effects. A large-scale study reported that plasma CRP might be prospectively associated with PTSD symptom emergence, suggesting that inflammation may predispose to PTSD. A more recent study highlighted a bidirectional genetic association between PTSD and CRP, also suggesting a potential role of socioeconomic status in the interplay between childhood support and inflammatory processes with respect to PTSD risk. Despite considerable heterogeneity, these recent studies have reported an association between PTSD and peripheral inflammation cross-sectionally and longitudinally. In addition, immune stimulation using endotoxin is associated with a more robust inflammatory response in subjects with PTSD, mediated via increased expression of NF-kB. The role of immune factors in fear learning and fear extinction mechanisms has been of recent interest. It has been demonstrated that inflammation-associated impairment in synaptic plasticity of fear extinction circuitry may underlie the behavioral dysfunction in PTSD. Much more work remains to be done to clarify these hypotheses. In addition, increased inflammation has been associated with decreased connectivity between reward-related brain regions and anhedonia in trauma-exposed women with PTSD symptoms, further emphasizing the consistency of effects of inflammation on the brain across disorders and related to motivation.

Comparing stress-related disorders such as depression and PTSD indicate that stress-induced immune pathophysiology is complex, heterogeneous, and defies simplistic explanations. Data from the two well-known stress-related disorders, that is, depression and PTSD, has shown differential associations between peripheral and central neuroimmune activation profiles. Despite well-documented increases in peripheral inflammation in PTSD, decreased binding of TSPO in prefrontal–limbic regions and decreased in microglial activation markers in postmortem brain specimens in transcriptomic analysis point to a delinking the central–peripheral immune responses. Thus, it is likely that the microglial response varies with the underlying psychopathologic syndromes.

OCD. Pediatric autoimmune neuropsychiatric syndromes associated with streptococcal infections (PANDAS) such as Sydenham chorea result from antineuronal antibodies induced by group A beta-hemolytic streptococcal infection. Children with Sydenham chorea frequently exhibit obsessive-compulsive symptoms, and sudden onset of OCD, tics, emotional lability, hyperactivity, and ADHD have been described among children following infection with group A beta-hemolytic streptococcus. Moreover, children with repeated streptococcal infections exhibit higher distal choreiform movements and behaviors consistent with ADHD. Interestingly, recent data suggest that OCD more generally may also be associated with increased levels of several cytokines, including TNF, IL-6, and IL-1.

Neuro-Covid

NEUROLOGIC SYMPTOMS

Pathology. Severe pathologies induced by SARS-CoV-2 are the result of an exaggerated host immune response than the extent of the infection. The host immune response can include an exaggerated activity of both innate and adaptive immune systems, although increased innate activity has been decisively linked to acute mortality from COVID-19. Specific neurologic syndromes consistently reported include anosmia, encephalopathy, and stroke. In addition, recent data implicate activation of intracellular inflammasome pathways and KYN system activation in perpetuating short- and long-term SARS-CoV-2 brain pathologies.

Anosmia/Ageusia (Loss of Taste). SARS-CoV-2 infection does not directly infect olfactory neurons in the nasal mucosa but instead causes significant infection of the olfactory neural supporting cells known as sustentacular cells. Specific vulnerability of sustentacular cells can result from the expression of angiotensin-converting enzyme (ACE) -2 receptors and surface expression of transmembrane serine protease two that facilitates viral entry. SARS-CoV-2 has not yet been demonstrated to infect the brain directly; few viral proteins were detected in the brain parenchyma. However, the hypothesis that anosmia may result from impairment of olfactory bulb function has been effectively replicated in animal models. Neuroimaging studies have indicated increased hyperintensity and swelling of the olfactory bulb consistent with acute inflammation, which resolved during subsequent follow-up. Anosmia and loss of taste are common following mild infections but persist for longer durations.

Delirium and Altered Mental Status. Encephalopathic changes underlie the development of delirium among hospitalized SARS-CoV-2 infections. Elevated plasma and CSF levels of cytokine and neurofilament light chain proteins (generated from the breakdown of fibrils in large-caliber nerve fibers) correlate with pathologic and behavioral changes among acutely confused patients. In addition, overactive astroglia was seen in widely distributed cortical regions among the very ill. Region-specific increases in microglial and cytotoxic T-cell activity have also been related to SARS-

CoV-2 infections. Inflammasome and KYN pathway activation may underlie some, if not all, of the SARS-CoV-2.

Stroke. Profound inflammatory response involving the endothelium of cerebral blood vessels resulting in strokes, hemorrhages, and disseminated encephalopathies are frequent during acute phases of the infections. This is attributed to the enrichment of ACE receptors on cerebral endothelial cells, which are the primary portals of entry for SARS-CoV-2. Stroke remains the most damaging and life-threatening of neurologic COVID-19 pathologies. Large vessel cryptogenic strokes complicated by hemorrhagic transformation have been reported. In addition, progressive microvascular changes associated with chronic white matter pathologies associated with cognitive and psychiatric symptoms have also been noted.

Behavioral Syndromes. The data on psychiatric manifestations of SARS-CoV-2 infections are emerging. Hence, only a preliminary overview of the encountered syndromes will be presented below. Current information on the neuropsychiatric sequelae of SARS-CoV-19 infection is based on cohorts of seriously ill hospitalized patients. However, the reader must remember that many infected individuals experienced mild to moderate infections that did not lead to hospitalization.

Cognitive Difficulties. Subjective cognitive difficulties described as “mental fog” are among the commonest self-reported symptoms following SARS-CoV-2 infections, especially among nonhospitalized subjects. The symptoms are characterized by decreased attention, concentration, and possibly impaired memory symptoms. Complaints of subjective cognitive symptoms are more frequent during the acute phase of the illness, and it is not yet clear if these symptoms will persist over a longer follow-up period.

Fatigue. Fatigue is common after several viral infections such as Epstein–Barr virus infections and SARS-CoV-2 infections resemble many of them. Pathologic overlaps between myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) are an area of active investigation.

NEUROPSYCHIATRIC SYMPTOMS. Neuropsychiatric symptoms have been examined in recent meta-analytic reviews. During the acute phase included depressed mood (33%), anxiety (36%), impaired memory (34%), impaired attention concentration (38%), insomnia (42%), and confusion (28%). The cognitive symptoms appeared to be independent of the age of the samples. Upon follow-up, the point prevalence of anxiety disorder diagnoses was 14.8% (mean follow up of 11.6 months), depression was 14.9% (mean follow-up of 22.6 months), PTSD was 32.2% (mean follow-up of 33.6 months). Unfortunately, reports of period prevalence have not yet been published. However, symptom severity scores indicated mild to moderate anxiety and depression scores during follow-up. In addition, some papers noted greater severity of PTSD symptoms, especially in terms of intrusion, avoidance, and hyperarousal symptoms. Several subjects also reported poor subjective quality of life, including mental well-being. On the contrary, several subjects also reported positive emotions, including gratitude to support and appreciation for the value of life and relationships.

SARS-COV-2 INFECTION AMONG MENTALLY ILL. The likelihood of a positive test was 22.3% for the schizophrenia spectrum group, 25.4% for the mood disorder group, 24.1% for the anxiety disorder group, and 28.2% for the reference group. However, 22% to 25% of subjects diagnosed with schizophrenia, mood, and anxiety disorders were positive for SARS-CoV-2 using a medical record–based study. In addition, a premorbid diagnosis of a schizophrenia spectrum disorder was significantly associated with 2.67 times increased mortality following SARS-CoV-2.

PSYCHOTROPICS AND ANTI-COVID TREATMENTS. Given the high risk of SARS-CoV-2 exposure among the mentally ill, it would be helpful to review some risks of combining psychotropics with outpatient anti-COVID medications. The most notable development in the treatment is the temporary approval of antiviral combination medications (Paxlovid, Pfizer Inc.) for acute SARS-CoV-19. The drug is composed of two agents, nirmatrelvir and ritonavir, the latter being a powerful inhibitor of the P450 (CYP) 3A4 system. Medications such as carbamazepine and herbals such as St. John’s wort can decrease the efficacy of Paxlovid. Risk of serious toxicity are likely with clonazepam, diazepam, clobazam, estazolam, flurazepam, clozapine, lumateprone, lurasidone, quetiapine, and pimozide. It is recommended that these medicines be discontinued before initiation and not resumed till 3 days after the conclusion of Paxlovid therapy. Drugs with potential interaction include trazodone, aripiprazole, haloperidol, iloperidone, risperidone, alprazolam, buspirone, and flunitrazepam. It has been recommended that Paxlovid should not be initiated unless these agents can be safely discontinued if necessary. Finally, agents with possible risks include ziprasidone and zolpidem. The recommendations include fully counseling patients to discontinue these agents if they feel unwell.

FACTORS ASSOCIATED WITH IMMUNE ACTIVATION IN PSYCHIATRY

Multiple individual variables impact immune–brain communication in a variety of psychopathologic states. Several of these factors, such as social adversity, stress, and health status, have been discussed under appropriate headings. However, some factors transcend disease categories and are reviewed below.

Aging

Aging is paradoxically associated with a deteriorated immune defense (immunosenescence) and increased basal levels of tissue inflammation (inflammaging). Inflammaging is associated with the exaggerated inflammatory response of innate cells. This change results in an increased or exaggerated inflammatory activity at baseline and in responding to immune stimulation. Paradoxically, this exaggerated immune response is less efficient in fighting pathogens or mounting immune responses to immunizations. This exaggeration in inflammation often results from phenotypic changes in immune cells such as macrophages and monocytes, resulting in increased secretion of cytokines such as TNF, IL-6, and IL-1, and chemokines such as MCP-1, leading to increased synthesis of CRP. Notably, microglial cells in the brain also develop a similar exaggeration in the immune-response profile. In addition, recent studies indicate that astroglial cells also transform into a hypersecretory phenotype that can amplify and exaggerate brain immunity. The precise meaning of these increases in age-associated inflammatory cell changes is far from clear.

One possible explanation is that immune cells undergo “priming” by the age-associated accumulation of spontaneously mutating neoplastic cells or misfolded proteins such as beta-amyloid or alpha-synuclein. In addition, other brain insults incurred during the long human lifespan may also play a contributory role. Some of these include chronic cerebrovascular disease, diabetes, obesity, and intermittent episodes of traumatic brain injury. The exaggerated immune response of primed cells may help partially neutralize threats but may lead to other unintended consequences via “collateral” tissue damage. For example, it is well known that chronic immune activation in healthy and chronically depressed individuals may be associated with a wide array of age-associated chronic medical disorders.

There are two divergent views on understanding innate immune activation in age-associated brain disorders. One perspective emphasizes that priming and inflammatory bias of glial cells is conducive to more efficient removal of cellular debris and misfolded proteins and may serve a valuable purpose. The contrarian view holds that exaggerated immune cell activity in the brain may lead to greater neuronal loss and neurodegeneration. While there may be truth in both views, much more research will be necessary to understand the biologic meaning of these changes. Of note, most if not all inflammatory changes in major depression are exaggerated among older depressed patients.

The adaptive immune response is decreased with aging. The precise nature of adaptive immune changes with aging is even less well understood. However, recent studies indicate that the decline in adaptive immune responses may result from increased “checkpoint” regulation, thus suppressing adaptive immune activity, beneficial to clearance of amyloid and other debris. Thus, the recent development and discovery of checkpoint inhibitors may one day improve brain health and prevent disorders such as AD.

Nutrition, Obesity, and Activity

Obesity, defined as BMI greater than 30 kg/m^2, represents the commonly encountered threat to neuroimmune homeostasis. A two-way association between the immune system and glucose/lipid management systems has been consistently reported. Obesity/high BMI is reliably linked with chronic low-grade increases in inflammatory activity measured using proinflammatory markers such as CRP and IL-6. This universal association is common to all population categories, especially the mentally ill. It is well known that adipocytes in abdominal fat pads recruit inflammatory proteins and cells that generate chronic inflammation.

On the other hand, immune molecules also may cause weight gain. Recent data have indicated that inflammatory molecules reprogram cellular glucose metabolism and promote insulin resistance, hyperglycemia, and hyperlipidemia, thus promoting metabolic syndrome. Reinstatement of TLRs in genetically null mice (TLR-null mice) leads to rapid weight gain following a high-fat diet. In addition, chronic sympathetic activation resulting from stress leads to gene expression changes among tissue macrophages in brown adipose tissue, leading to aberrant thermogenesis and weight loss.

Lifestyle modifications geared toward weight reduction generally result in a reduction of chronic inflammatory activity. For instance, reduction in CRP and related cytokines are associated with "healthy" Mediterranean style diets and increases in CRP and sTNFR2 have been noted following intake of high transfat/high-caloric diet. Along similar lines, fitness-promoting activities and physical exercise are well documented to lower the proinflammatory molecules and their activity. Conversely, a sedentary lifestyle has been linked with significant increases in proinflammatory activity in several population groups, including the mentally ill. Furthermore, obesity tends to combine with other inflammation-promoting factors such as advancing age, stress, smoking/alcohol use, and chronic medical disorders to accentuate chronic inflammatory activity.

As stated earlier, increased inflammation may characterize a distinct subgroup of depressed subjects. Most recent formulations have concluded that the presence of obesity and early-life adversity may phenotypically identify a subgroup of depressed subjects with high inflammation. Improved response to L-methyl folate augmentation of antidepressant therapy was seen among subjects with high BMI and high inflammation. Similarly, several authors, including the previous edition of this textbook, had noted that the link between early life adversity and BMI is likely to be strong in and of itself. Thus, it may be very difficult to dissociate the effects of stress and BMI upon chronic inflammation among psychiatrically ill populations. Most studies of bipolar disorder indicate that high BMI, high-calorie diet, sedentary lifestyle, and lack of exercise are high-risk factors for poor treatment outcomes. It is most likely that inflammation is an underlying common link among these factors. Intriguingly, obesity and high inflammation are also common among first-episode psychotic patients. Though the relationship between psychosis, obesity, and inflammation is unclear, inflammation may be related to the disease process itself rather than obesity. In addition, links between PTSD and inflammation have been reported even after controlling for obesity.

These data indicate that obesity plays a substantial role in the etiology of chronic inflammation among mentally ill subjects. It may play a direct role in driving up the inflammatory status among patients. In addition, it may exacerbate the inflammatory potential of other disease mechanisms in disorders such as schizophrenia and PTSD and add to the proinflammatory potential of lifestyle factors such as inactivity, tobacco/alcohol use, and stress.

Microbiome

A detailed description of microbiota is beyond the scope of this chapter, and the readers are referred to reviews on this area. The collection of microorganisms associated within a certain location is called microbiota, and its collective genetic material is called the microbiome. The human body has a large amount of microbiota, mostly located in the gastrointestinal system. It has been stated that most of the genetic material found in the human body is made up of microbial genes. While some microbial invaders lead to profound immune stimulation, human microbiota is predominantly symbiotic and often does not stimulate host inflammatory responses. It is likely that the human gut hosts microbes from more than 60 genera. Recent advances in the analysis of microbial DNA have led to a better understanding of the influence of microbiota on various aspects of the host's brain health. Multiple modes of interaction between the brain, including synthesis of metabolites and neurotransmitters, activation of the vagus nerve, and activation of the anti-inflammatory arc of the immune system, have all been reported.

Colonization of the human gut microbiome begins at birth mostly by exposure to the maternal birth canal in natural delivery and the maternal skin in C-section. Other factors such as prematurity, breastfeeding, presence of pets, caloric intake, and obesity may all impact the developing microbiome. Dietary intake is believed to be an important factor in the regulation of the gut microbiome. For instance, a fiber-rich Mediterranean diet is associated with Prevotella species, while a high-calorie Western diet is associated with Bacteroides. There is a complex two-way relationship between the gut and brain. Significant preclinical data appear to support the notion that gut microbiota influence brain development and function. Intriguingly, there is emerging evidence that brain functions may also disrupt the microbiome. Stress alters the microbiota composition in various species, and probiotics have been shown to reverse the effects of stress in animal models.

Reduced prevalence of Lactobacillus and Bifidobacterium species were noted among subjects with the first-episode psychosis. Importantly, differences in microbiota predicted severity of negative symptoms and quality of remission during follow-up. Differences in the oropharyngeal microbiome were also reported from schizophrenic subjects. Probiotic treatment yielded no perceptible clinical benefits, but it improves severe bowel problems in a group of subjects with schizophrenia. Differences in two separate genera of Firmicutes and a positive response to probiotic supplementation were noted in bipolar disorder.

The most compelling evidence implicating microbiome in psychiatric disorders comes from the study of depression. Preclinical data also indicate that probiotic and prebiotic therapies may reduce depressive behavior in rodent models. Multiple systematic reviews indicate that probiotics may improve mood in depressed humans. Notably, transplantation of the gut microbiome from depressed humans induced greater depressive and anxiety-like behavior in rodents. Several studies have noted altered microbial composition in depressed subjects with a reduction in the relative abundance of *Feacalbacterium* species that has been negatively correlated with the severity of depression.

In summary, the development of mental illnesses may be related at least in part due to the loss of immunomodulatory effects of microbiota. The precise mechanisms by which microbiota trigger or alter inflammatory pathologies are still unclear. However, it has been proposed that the balance between microbiota with pro- and anti-inflammatory biases may be disrupted during times of stress leading to an exaggeration of inflammation-promoting organisms. This change may result in possible activation of gastrointestinal mucosa leading to the release of inflammatory cells and microbial products directly into the systemic circulation. This “leaky gut” hypothesis has attracted considerable attention in the etiology of chronic inflammation in stress-related disorders.

Human-Microbial Symbiosis—“Old Friends” Hypothesis. Figure 1.13–6 depicts how evolutionary pressures and adaptation to our current living conditions may have resulted in heightened inflammatory bias. Studies have argued that genes promoting depression may have evolved and persisted in the human genome, at least partly because many physiologic and behavioral aspects of depression in ancestral environments may have enhanced host defense against pathogens. In addition to microbiota, two more classes of commensal organisms (pseudocommensals and parasites) are collectively referred to as “old friends,” that is, organisms living within humans that promote healthy immune function while suppressing proinflammatory activity. Some recent authors have hypothesized that modern environmental practices have disrupted human relationships with “old friends,” leading to an increased prevalence of immune-based pathologies common in industrialized societies than in more primitive environments.

Significant in vitro and in vivo preclinical data indicate that all three classes of “old friend” organisms (microbiota, pseudocommensals, and parasites) have the potential to prevent or ameliorate pathology in animal models for autoimmune, allergic, and inflammatory bowel diseases, as well as neoplasms and certain infections, and do so via reductions in inflammatory activity. Human–microbial and human–parasitic interactions in disease pathogenesis lead to the possibility that organism-based interventions may show promise if delivered early to prevent the development of schizophrenia, autism, and mood disorders.

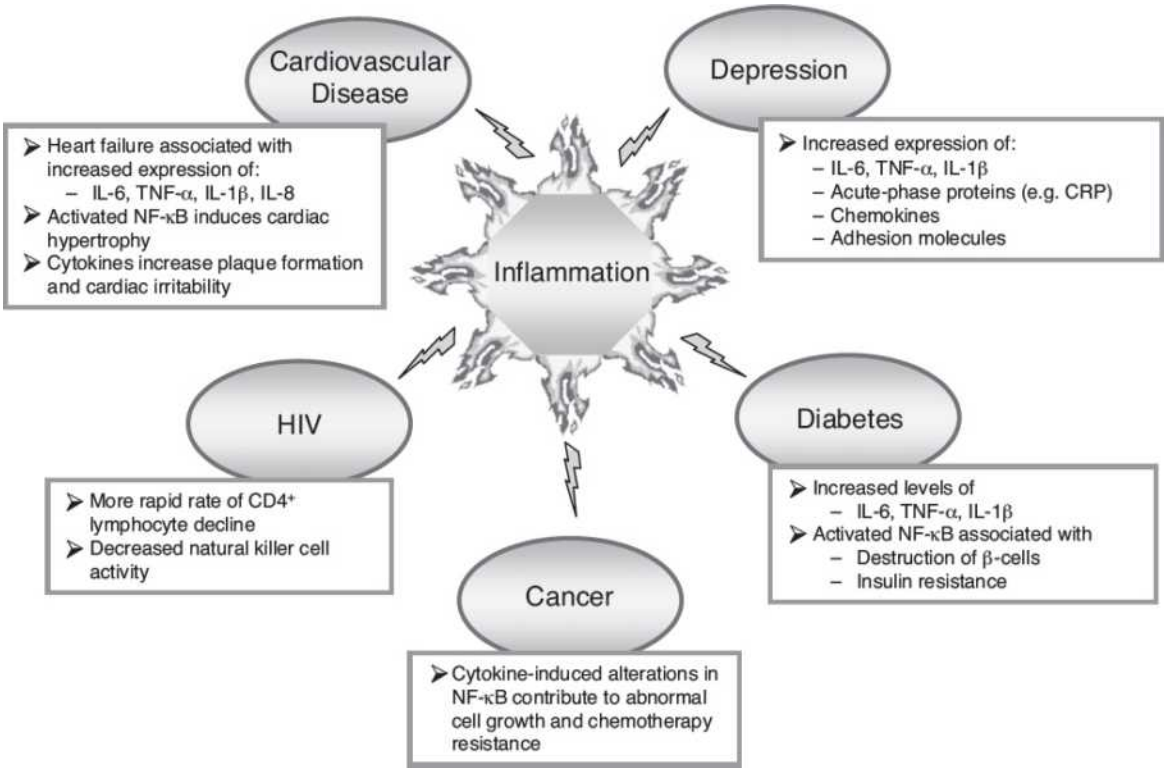
Cardiovascular Disease
Heart failure associated with increased expression of:
– IL-6, TNF-α, IL-1β, IL-8
Activated NF-κB induces cardiac hypertrophy
Cytokines increase plaque formation and cardiac irritability
Depression
Increased expression of:
– IL-6, TNF-α, IL-1β
– Acute-phase proteins (e.g. CRP)
– Chemokines
– Adhesion molecules
Inflammation
HIV
More rapid rate of CD4+ lymphocyte decline
Decreased natural killer cell activity
Diabetes
Increased levels of
– IL-6, TNF-α, IL-1β
Activated NF-κB associated with
– Destruction of β-cells
– Insulin resistance
Cancer
Cytokine-induced alterations in NF-κB contribute to abnormal cell growth and chemotherapy resistance

FIGURE 1.13–6. Early evolutionary pressures derived from human interactions with pathogens, predators, and human conspecifics (such as rivals) resulted in an inflammatory bias that included an integrated suite of immunologic and behavioral responses that conserved energy for fighting infection and healing wounds, while maintaining vigilance against attack. This inflammatory bias is believed to have been held in check during much of human evolution by exposure to minimally pathogenic, tolerogenic organisms in traditional (i.e., rural) environments that engendered immunologic responses characterized by the induction of regulatory T (TReg) cells, regulatory B (BReg) cells, and immunoregulatory M2 macrophages as well as the production of the anti-inflammatory cytokines interleukin-10 (IL-10) and transforming growth factor-beta (TGFβ). In modern times, sanitized urban environments of more developed societies are rife with psychological challenges but generally lacking in the types of infectious challenges that were primary sources of morbidity and mortality across most of human evolution. In the absence of traditional immunologic checks and balances, the psychological challenges of the modern world instigate ancestral immunologic and behavioral repertoires that represent a decided liability, such as high rates of various inflammation-related disorders including depression. (Reprinted from Miller A, Raison C. The role of inflammation in depression: from evolutionary imperative to modern treatment target. *Nat Rev Immunol* 2016;16:22–34. doi: 10.1038/nri.2015.5)

Sleep

Disturbances in sleep rhythm and cycle are ubiquitous in the field of psychiatry. Most recent authors consider sleep as a transdiagnostic marker that transcends conventional psychiatric diagnostic boundaries. Given the large volume of data linking sleep and inflammation, the readers are referred to source papers on this area. Naturally occurring sleep regulates the immune system by controlling the production of inflammatory cytokines and the distribution of inflammatory cells. It is well known that sleep regulates the functioning of stress response systems, including the HPA axis and SNS. Methodologically rigorous studies that repeatedly sampled immune markers during natural sleep and states of sleep deprivation have yielded intriguing results. Systemic levels of IL-6 tend to peak between 7PM and 5AM, and nocturnal sleep may be the single deciding factor in the nocturnal increases of IL-6 and decreases in TNF during the night. Experimental sleep deprivation leads to decreased nocturnal IL-6 levels, delays the nocturnal peak in IL-6 secretion, and possibly attenuates TNF production by monocytes during the night. Subjects with sleep disturbances, including decreased duration of slow-wave sleep and increased duration of rapid eye movement (REM) sleep, demonstrate increased daytime IL-6 coupled with its decreased nighttime production. Finally, nocturnal sleep appears to increase concentrations of the soluble IL-6 receptor (sIL6-R) during nighttime, increasing its ability to influence neuronal cells that are not normally responsive to the effects of IL-6. In a large-scale meta-analysis, sleep disruption was associated with higher CRP and IL-6 in the plasma.

More critically, psychosocial stressors and adversity may impact inflammation by disrupting sleep efficiency and quality. The characteristic changes in sleep architecture in depression are classic examples of such a shift. It has been hypothesized that this shift in the sleep cycle may induce an inflammatory transcriptional state in the immune cells such as monocytes leading to altered patterns of release of cytokines such as IL-6 and TNF.

In contrast, increases in peripheral inflammation can also disrupt sleep. For example, viral and bacterial infections are well known to alter the balance between non-REM and REM sleep—an effect mediated by inflammatory cytokines. Neural sensors of inflammation located in the vagus nerve are believed to convey immune signals to the sleep centers in the brain. Thus, resection of the vagus nerve appears to abolish the effect of immune molecules on sleep rhythm. A key area of research involves the relationship between sleep disturbance and inflammatory aging (inflammaging). Notably, one night of partial sleep deprivation–induced genomic and transcriptional changes suggestive of accelerated aging and increased cellular senescence in immune cells. The early or premanifest stage of AD is associated with increases in neuroinflammation and sleep disruptions. It has been proposed that both these processes combine to cause disease progression in AD. The neuroinflammation is believed to result from primed microglia (microglia stuck in a state of persistent activation) exposed to amyloid and tau proteins. Thus, neuroinflammation and sleep deprivation worsen each other in a manner that may contribute to illness progression.

Of note, sleep disturbance also has a role in the risk of medical illnesses (such as hypertension and diabetes) and lifestyle factors (sedentary lifestyle and social isolation), and sleep disorders (sleep apnea, REM-sleep disorder), all of which may increase the risk of dementia. Finally, the relationship between sleep disturbances in depression has been well studied in the literature. Persistent insomnia remains a strong predictor of depression recurrence. Notably, recent studies indicate that depressive response to inflammatory stimuli such as endotoxin challenge may be amplified by sleep deprivation. Alternatively, the onset of sleep disturbance in subjects with pre-existing inflammatory disorders can often precipitate the onset of depression. Thus, the use of behavioral sleep therapies that appears to improve natural sleep (as opposed to sleep induced by hypnotic drugs) may offer a promising avenue to reduce the impact of sleep disturbance on the onset of depression.

CONCLUSIONS TO THE CHAPTER

Disrupted immune–brain communication in psychiatric disorders may represent the cumulative effects of a complex causal chain. However, despite their complexity, a good grasp of immune–brain communication pathologies can help us better target a host of potentially modifiable risk factors that can improve the lives of our patients.

FURTHER READINGS

Bastiaanssen TFS, Cryan JF. The microbiota-gut-brain axis in mental health and medication response: parsing directionality and causality. *Int J*

Neuropsychopharmacol. 2021;24(3):216–220.

Bhatt S, Hillmer AT, Girgenti MJ, et al. PTSD is associated with neuroimmune suppression: evidence from PET imaging and postmortem transcriptomic studies. *Nat Commun*. 2020;11(1):2360.

Brown AS. Prenatal infection as a risk factor for schizophrenia. *Schizophr Bull*. 2006;32(2):200–202.

Cho M, Lee TY, Kwak YB, Yoon YB, Kim M, Kwon JS. Adjunctive use of anti-inflammatory drugs for schizophrenia: a meta-analytic investigation of randomized controlled trials. *Aust N Z J Psychiatry*. 2019;53(8):742–759.

Dantzer R, O'Connor JC, Freund GG, Johnson RW, Kelley KW. From inflammation to sickness and depression: when the immune system subjugates the brain. *Nat Rev Neurosci*. 2008;9(1):46–56.

Dargél AA, Godin O, Kapczinski F, Kupfer DJ, Leboyer M. C-reactive protein alterations in bipolar disorder: a meta-analysis. *J Clin Psychiatry*. 2015;76(2):142–150.

Eraly SA, Nievergelt CM, Maihofer AX, et al; Marine Resiliency Study Team. Assessment of plasma C-reactive protein as a biomarker of posttraumatic stress disorder risk. *JAMA Psychiatry*. 2014;71(4):423–431.

Fernandes BS, Steiner J, Molendijk ML, Dodd S, Nardin P, Goncalves CA, et al. C-reactive protein concentrations across the mood spectrum in bipolar disorder: a systematic review and meta-analysis. *Lancet Psychiatry*. 2016;3(12):1147–1156.

Franceschi C, Santoro A, Capri M. The complex relationship between immunosenescence and inflammaging: special issue on the New Biomedical Perspectives. *Semin Immunopathol*. 2020;42(5):517–520.

Frank P, Jokela M, Batty GD, Cadar D, Steptoe A, Kivimäki M. Association between systemic inflammation and individual symptoms of depression: a pooled analysis of 15 population-based cohort studies. *Am J Psychiatry*. 2021;178(12):1107–1118.

Haroon E, Daguanno AW, Woolwine BJ, Goldsmith DR, Baer WM, Wommack EC, et al. Antidepressant treatment resistance is associated with increased inflammatory markers in patients with major depressive disorder. *Psychoneuroendocrinology*. 2018;95:43–49.

Haroon E, Miller AH, Sanacora G. Inflammation, glutamate, and glia: a trio of trouble in mood disorders. *Neuropsychopharmacology*. 2017;42(1):193–215.

Haroon E, Raison CL, Miller AH. Psychoneuroimmunology meets neuropsychopharmacology: translational implications of the impact of inflammation on behavior. *Neuropsychopharmacology*. 2012;37(1):137–162.

Irwin MR. Sleep and inflammation: partners in sickness and in health. *Nat Rev Immunol*. 2019;19(11):702–715.

Kiraly DD, Horn SR, Van Dam NT, Costi S, Schwartz J, Kim-Schulze S, et al. Altered peripheral immune profiles in treatment-resistant depression: response to ketamine and prediction of treatment outcome. *Transl Psychiatry*. 2017;7(3):e1065.

Lanquillon S, Krieg JC, Bening-Abu-Shach U, Vedder H. Cytokine production and treatment response in major depressive disorder. *Neuropsychopharmacology*. 2000;22(4):370–379.

Lucido MJ, Bekhbat M, Goldsmith DR, Treadway MT, Haroon E, Felger JC, et al. Aiding and abetting anhedonia: impact of inflammation on the brain and pharmacological implications. *Pharmacol Rev*. 2021;73(3):1084–1117.

Marazziti D, Mucci F, Fontenelle LF. Immune system and obsessive-compulsive disorder. *Psychoneuroendocrinology*. 2018;93:39–44.

Marzolini C, Kuritzkes DR, Marra F, Boyle A, Gibbons S, Flexner C, et al. Recommendations for the management of drug-drug interactions between the COVID-19 antiviral nirmatrelvir/ritonavir (Paxlovid) and comedications. *Clin Pharmacol Ther*. 2022;112(6):1191–1200.

Meltzer A, Van de Water J. The role of the immune system in autism spectrum disorder. *Neuropsychopharmacology*. 2017;42(1):284–298.

Meyer JH, Cervenka S, Kim MJ, Kreisl WC, Henter ID, Innis RB. Neuroinflammation in psychiatric disorders: PET imaging and promising new targets. *Lancet Psychiatry*. 2020;7(12):1064–1074.

Miller AH, Raison CL. The role of inflammation in depression: from evolutionary imperative to modern treatment target. *Nat Rev Immunol*. 2016;16(1):22–34.

Miller BJ, Goldsmith DR. Towards an immunophenotype of schizophrenia: progress, potential mechanisms, and future directions. *Neuropsychopharmacology*. 2017;42(1):299–317.

Pollak TA, Lennox BR, Müller S, et al. Autoimmune psychosis: an international consensus on an approach to the diagnosis and management of psychosis of suspected autoimmune origin. *Lancet Psychiatry*. 2020;7(1):93–108.

Rogers JP, Chesney E, Oliver D, et al. Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: a systematic review and meta-analysis with comparison to the COVID-19 pandemic. *Lancet Psychiatry*. 2020;7(7):611–627.

Roman M, Irwin MR. Novel neuroimmunologic therapeutics in depression: a clinical perspective on what we know so far. *Brain Behav Immun*. 2020;83:7–21.

Rosenblat JD, McIntyre RS. Efficacy and tolerability of minocycline for depression: a systematic review and meta-analysis of clinical trials. *J Affect Disord*. 2018;227:219-25.

Rosenblat JD. Targeting the immune system in the treatment of bipolar disorder. *Psychopharmacology (Berl)*. 2019;236(10):2909–2921.

Salvador AF, de Lima KA, Kipnis J. Neuromodulation by the immune system: a focus on cytokines. *Nat Rev Immunol*. 2021;21(8):526–541.

Schou TM, Joca S, Wegener G, Bay-Richter C. Psychiatric and neuropsychiatric sequelae of COVID-19—a systematic review. *Brain Behav Immun*. 2021;97:328–348.

Schwarcz R, Stone TW. The kynurenine pathway and the brain: challenges, controversies and promises. *Neuropharmacology*. 2017;112(Pt B):237–247.

Shelton RC, Miller AH. Eating ourselves to death (and despair): the contribution of adiposity and inflammation to depression. *Prog Neurobiol*. 2010;91(4):275–299.

Solomon T. Neurological infection with SARS-CoV-2—the story so far. *Nat Rev Neurol*. 2021;17(2):65–66.

Strawbridge R, Arnone D, Danese A, Papadopoulos A, Herane Vives A, Cleare AJ. Inflammation and clinical response to treatment in depression: a meta-analysis. *Eur Neuropsychopharmacol*. 2015;25(10):1532–1543.

Vasupanrajit A, Jirakran K, Tunvirachaisakul C, Maes M. Suicide attempts are associated with activated immune-inflammatory, nitro-oxidative, and neurotoxic pathways: a systematic review and meta-analysis. *J Affect Disord*. 2021;295:80–92.

1.14 Chronobiology, Circadian Rhythm, and Psychiatry

CHESTER WU, M.D., AND JOSHUA J. RODGERS, M.D.

Chronobiology is the study of biologic rhythms. Of the various periodicities that fall within the realm of chronobiology, *circadian* (Latin: *circa,* about; *dies,* day) rhythms that have a period of about 1 day and serve to synchronize cellular processes and behavior with the environmental day–night cycle are among the most extensively studied and best understood biologic rhythms. Circadian rhythms are evolutionarily conserved across nearly all known living organisms, a fact that suggests these rhythms are fundamental to adaptive functioning on this planet.

The site of the primary circadian oscillator in mammals, including humans, is the suprachiasmatic nucleus (SCN), located in the anterior hypothalamus just dorsal to the optic chiasm from which the afferent retinohypothalamic tract (RHT) separates (Fig. 1.14–1). Via local genetic feedback loops of relatively fixed timing, the SCN generates a circadian period of approximately 24.18 hours. Therefore, the master circadian clock must constantly adjust to being effective at maintaining the proper phase relationships of behavioral and physiologic processes within the context of the 24-hour day. The characterization of persons as owls or larks ("night person" or "morning person") comes largely from variations in genetically determined longer or shorter circadian period lengths, respectively.

Organisms have evolved to use time cues or *zeitgeber* (German: *zeit,* time; *geber,* giver) to adjust or resynchronize the endogenous circadian clock. One of the most critical zeitgebers is the daily variation in light levels. Changes in illuminance are sensed by a photoreceptive apparatus that resides in the retina in humans and is communicated to the SCN via the RHT. Other nonphotic zeitgebers include food/feeding, social interactions, temperature, substances (such as amphetamines), and exercise. Signals from these stimuli are relayed to the SCN from the thalamus and midbrain raphe.

The circadian clock modulates almost all human physiology, from the molecular level with transcription and translation to larger order metabolic and cognitive functioning. It does this by driving rhythms, including behavior, core body temperature, sleep, feeding, drinking, and hormonal levels.

Taken together, the circadian axis of mammals can be divided into three distinct functional components: (1) a master pacemaker situated in the SCN, (2) a photoreceptive input to the SCN that originates in the eye, and (3) the myriad of rhythmic outputs that provide insight into the clockwork of the circadian pacemaker.

CIRCADIAN PACEMAKERS

Anatomy

The mammalian circadian system is organized as a hierarchy of pacemakers. The SCN is the master oscillator that orchestrates a multitude of oscillators outside of the SCN. These oscillators are found in a wide range of peripheral tissues, including the kidney, liver, lung, and other sites in the brain. The SCN consists of two paired nuclei in the hypothalamus situated immediately dorsal to the optic chiasm (see Fig. 1.14–1).

The neurons of the SCN are among the smallest neurons in the entire brain. They are densely packed and possess short dendrites that are not extensively branched. Subdivisions of the SCN have been defined according to immunohistochemical and neural tract tracing criteria. These subdivisions include the *core* of the SCN, defined by the presence of calbindin-positive neurons. The remainder of the SCN that surrounds the core is the *shell.* Afferent and efferent projections to and from these subdivisions are beginning to provide insights into their putative functions.

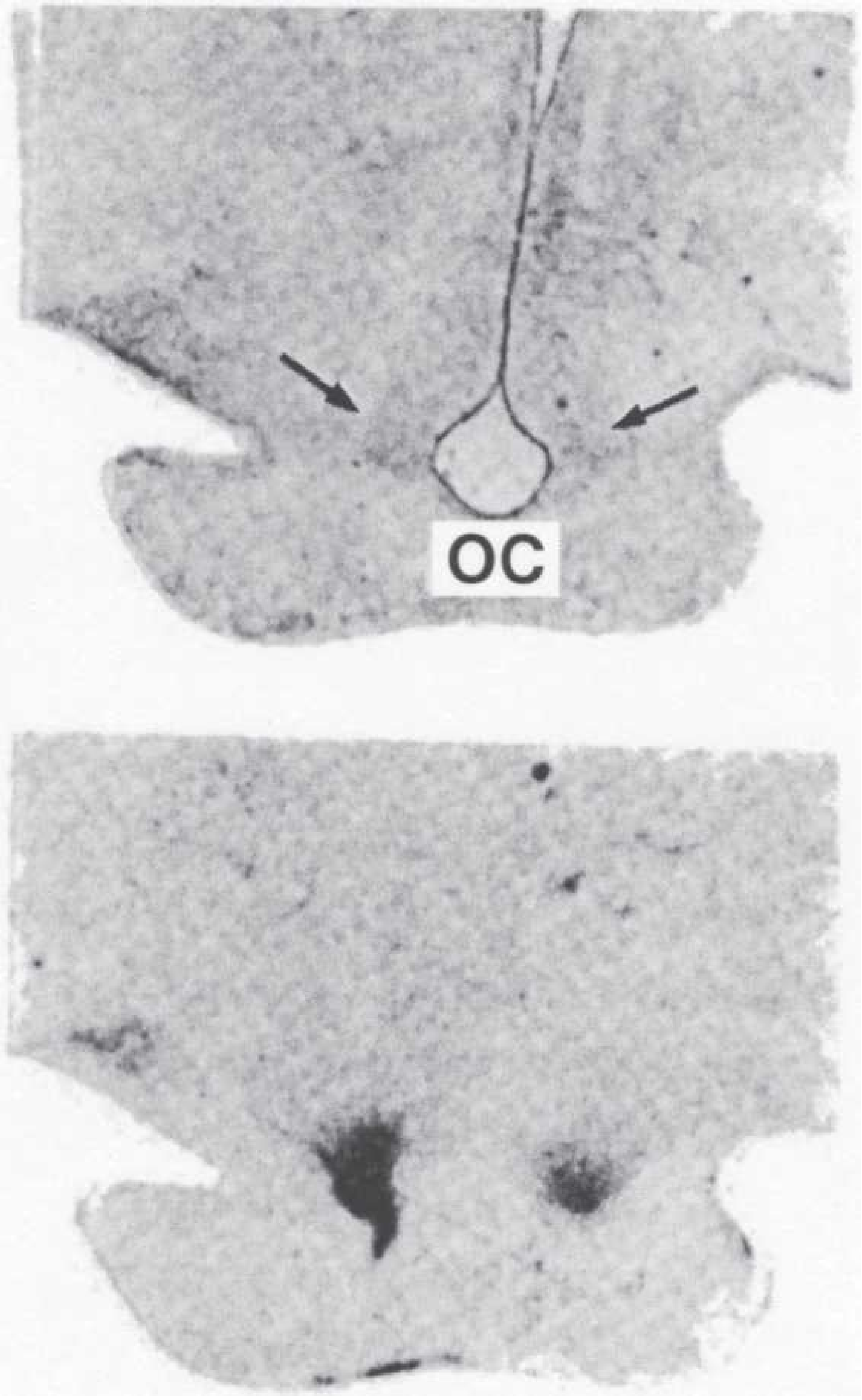

FIGURE 1.14–1. The human suprachiasmatic nucleus. **Top:** Nissl stain of section through the human hypothalamus. The suprachiasmatic nuclei are indicated by *arrows* and are located dorsal to the optic chiasm (OC). **Bottom:**

Autoradiograph of same section. Specific binding of a radioiodinated analog of melatonin is indicated by the darkening of the suprachiasmatic nuclei. (From Reppert SM, Weaver DR, Rivkees SA, Stopa EG. Putative melatonin receptors in a human biological clock. *Science* 1988;242(4875):78. Reprinted with permission from AAAS. doi: 0.1126/science.2845576.)

Afferent Projections

The RHT is the primary afferent input to the SCN. It originates in the retina from a small subset of retinal ganglion cells, which are particularly sensitive to blue light wavelengths, and innervates the entire volume of the SCN. The excitatory neurotransmitter glutamate is the primary neurotransmitter of the RHT. Thus, glutamate receptor antagonists can block the effects of light on the circadian axis. Likewise, blue wavelength filters for computer screen output or blue wavelength blocking eyeglass lenses can also mute these signals.

Photic projections are also received from the geniculohypothalamic tract from the ipsilateral intergeniculate leaflet (IGL), providing a secondary pathway from the retina to the SCN. Neuropeptide Y (NPY) is the predominant transmitter of the IGL-to-SCN projection.

Additional afferent inputs include from median raphe nucleus and the pedunculopontine, parabigeminal, and laterodorsal tegmentum nuclei. Most prominent among these is a distinct serotonergic projection from the midbrain raphe. It is hypothesized serotonin modulates the effect of light on SCN function, potentiating glutamate input during the day and inhibiting it at night.

Efferent Projections

Most SCN efferent projections remain within the limits of the hypothalamus.

The best-studied projection originating in the SCN is the multisynaptic projection to the pineal gland, which is involved in the regulation of melatonin release. Axons of inhibitory GABAergic SCN neurons traverse the hypothalamus dorsally to the paraventricular nucleus (PVN). These signals are propagated to the superior cervical ganglion, which projects to the pineal gland. At the synapse to the pineal gland, norepinephrine release stimulates beta-1– and alpha-1–adrenergic receptors on pinealocytes to trigger the production of melatonin. The ultimate outcome of this negative-feedback efferent pathway is that melatonin synthesis increases when SCN activity decreases. Thus, melatonin is a hormone signal for “biologic night” (not just sleep) for the brain and body. In normal circumstances, natural melatonin plasma concentration begins to rise (the so-called dim light melatonin onset, DLMO) 2 to 4 hours before the onset of sleep.

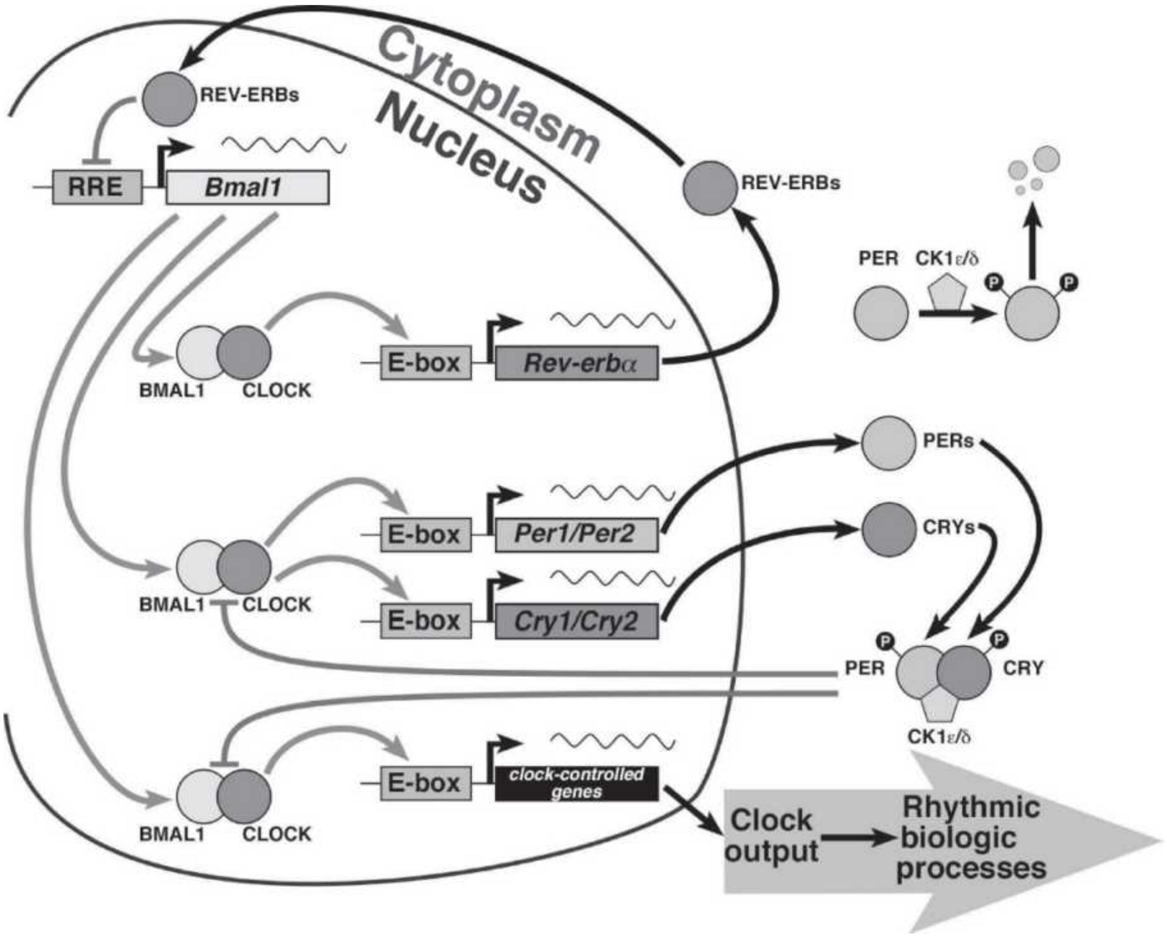
Cytoplasm
Nucleus
REV-ERBs
RRE
Bmal1
REV-ERBs
PER
CK1ε/δ
P
P
BMAL1
CLOCK
E-box
Rev-erbα
PERs
CRYs
E-box
Per1/Per2
E-box
Cry1/Cry2
BMAL1
CLOCK
P
P
PER
CRY
CK1ε/δ
E-box
clock-controlled genes
BMAL1
CLOCK
Clock output
Rhythmic biologic processes

FIGURE 1.14–2. Transcriptional–translational feedback loops that constitute the molecular clockworks of the mammalian circadian clock. (Adapted from Ko CH, Takahashi JS. Molecular components of the mammalian circadian clock. *Human Mol Genet* 2006;15(Rev 2):R271, by permission of Oxford University Press. doi:10.1093/hmg/ddl207.)

The SCN also coordinates peripheral circadian clocks contained in organs such as the lung, liver, kidney, and heart. A second, less understood efferent pathway from the SCN plays an important role in the control of cortisol which has a strong circadian component, being highest in the early morning in humans. Peak cortisol levels occur at approximately 6 AM to 8 AM, just as melatonin levels approach baseline.

The SCN projection to the hypothalamic PVN synapses on neurons that contain corticotrophin releasing hormone (CRH), eventually stimulating the release of adrenocorticotrophic hormone (ACTH), which in turn acts on the adrenal gland to release cortisol.

Molecular Regulation

Isolated SCN neurons can generate circadian rhythms. The mammalian circadian clockwork consists of interacting positive and negative transcriptional and translational feedback loops (Fig. 1.14–2). The expression of the period gene (*PER*) and cryptochrome genes (*CRY*) is positively regulated by the binding of Clock-Bmal1 (Bmal1) heterodimers to enhancers in the promoters of these genes. The products of the *Per* and *Cry* genes translocate back into the nucleus and repress their own transcription. This series of events constitutes the negative feedback limb of the core oscillation.

In addition to activating the transcription of the *PER* and *CRY* genes, the Clock-Bmal1 complex also activates the expression of the orphan nuclear receptor gene *Rev-Erba.* The gene product of *Rev-Erba,* in turn, translocates into the nucleus and represses transcription of the *Bmal1* gene through *Rev-Erb*/*Ror* response elements in the *Bmal1* gene promoter. Bmal1 subsequently heterodimerizes with Clock and again activates expression of the *PER, CRY,* and *Rev-Erba* genes. This derepression (or activation) of the *Bmal1* gene, subsequent heterodimerization with Clock, and activation of the *Per, Cry,* and *Rev-Erba* genes constitutes the positive feedback limb of the core oscillator.

The interactions of clock proteins and the translocations of these proteins between cellular compartments are tightly regulated by posttranslational modifications. For example, phosphorylation of some of the *Per* proteins by casein kinase Iε or δ is important for their translocation into the nucleus or their degradation in the cytoplasm.

SLEEP AND CIRCADIAN RHYTHMS

Sleep–Wake Regulation

Restful consolidated sleep is most appreciated when sleep disturbances are experienced. Sleep is the integrated product of two dynamic processes. The first process, frequently referred to as the *sleep* *homeostat,* or *process S* arises from the accumulation and dissipation of sleep debt. The biologic substrates encoding sleep debt are not fully known, although adenosine accumulation in the VLPO and other areas during wakefulness and its recycling during sleep is emerging as a primary candidate neuromodulator of the sleep homeostat. The adenosine receptor antagonist caffeine is thought to promote wakefulness by interfering with the homeostatic drive for sleep. The second oscillatory process, *process C*, is governed by the circadian clock and controls a daily rhythm in sleep propensity or, conversely, arousal.

The circadian promotion of arousal (wakefulness) steadily increases throughout the day, counteracting the homeostatic drive for sleep and reaching a maximum in the evening hours before beginning to wane, triggering the increase in plasma melatonin (Fig. 1.14–3). Arousal promotion subsequently decreases to coincide with the circadian trough in core body temperature. Experiments imposing forced sleep schedules throughout the circadian day have shown that an uninterrupted 8-hour bout of sleep is best obtained if sleep is initiated approximately 6 hours before the temperature nadir, which typically occurs around 5:00 AM to 6:00 AM. In healthy individuals, initiating sleep between 11:00 PM and 12:00 AM affords the highest probability of getting 8 solid hours of sleep.

Diurnal preference varies among individuals as a function of age, endogenous circadian periods, and other factors. Individuals who are labeled as morning types or *morning larks* tend to awaken earlier and experience the core body temperature minimum at an earlier clock time relative to those who are labeled as evening types or *night owls.* Sleep deprivation studies have shown that the homeostatic component of sleep is remarkably similar among individuals of similar age. Therefore, chronotype is dictated almost exclusively by the circadian process of sleep regulation. It should be noted that there is a well-established age-dependent decline in sleep need from infancy to adulthood. However, the more pronounced decline in total sleep time and sleep fragmentation experienced by many older adults is due to medical and psychosocial comorbidities, as well as, in some, degeneration of VLPO neurons and can result in consequential sleep deprivation.

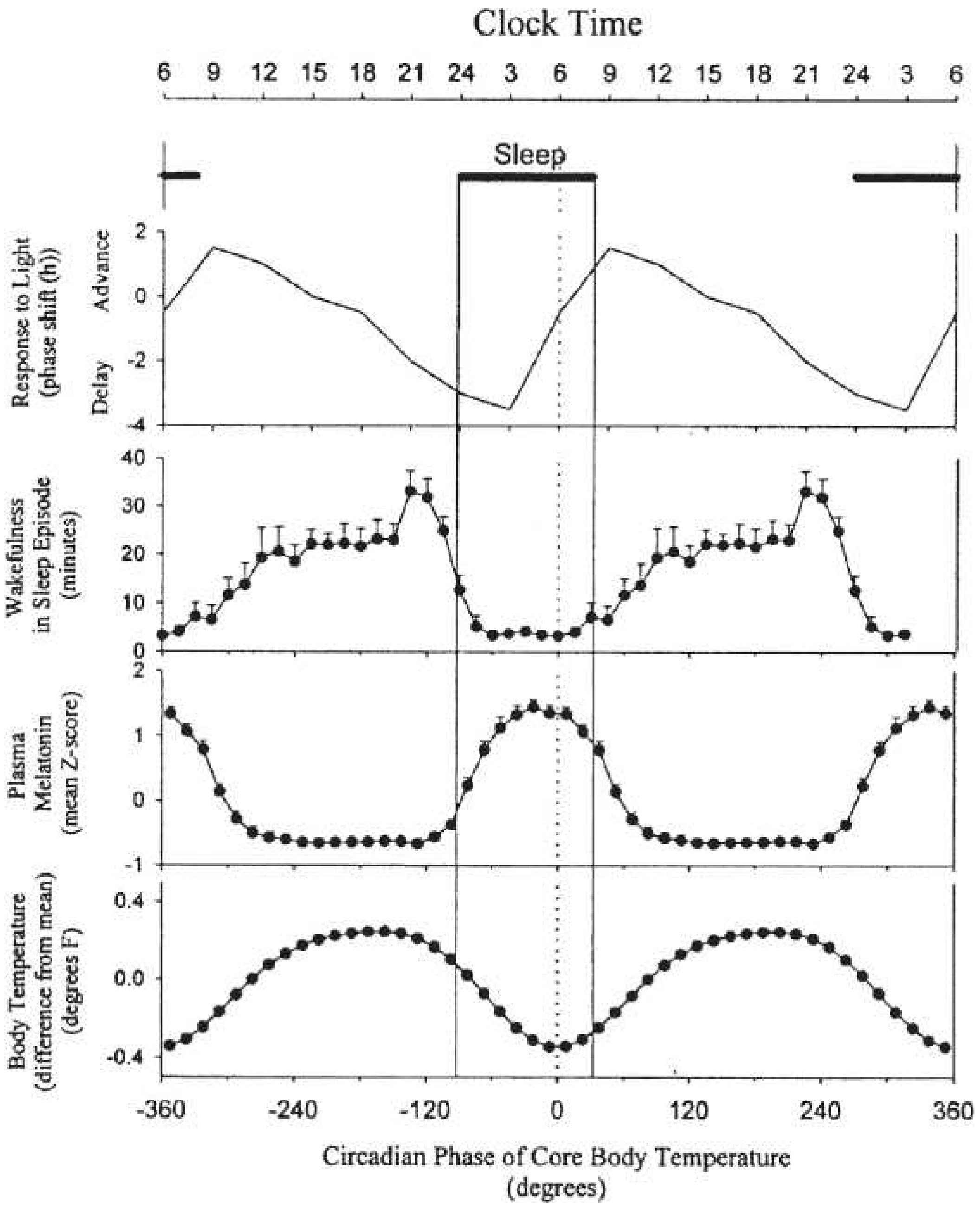

FIGURE 1.14–3. Relative phase relationship of sleep in young adults to other circadian phase markers. (From Dijk DJ, Lockley SW. Invited review: Integration of human sleep-wake regulation and circadian rhythmicity. *J Appl Physiol* 2002;92(2):852, with permission. doi:10.1152/japplphysiol.00924.2001.)

SEASONALITY

As described previously, melatonin levels are elevated during the night, not simply when sleeping. A long night, such as that experienced during the winter, results in an elevated melatonin profile of relatively long duration. A short summer night, by contrast, results in a short duration of elevated melatonin.

Seasonal variations in human behaviors include seasonal peaks in birth rate, sexual activity, conception, pain thresholds, suicide, and even criminal behavior. Seasonal variations are seen in biologic processes, including growth rate, cortisol, and testosterone levels. The general population experiences a winter worsening of mood, heightened appetite, lengthened sleep duration, lower energy, and less social behavior. These seasonal variations in mood and behavior are potentially maladaptive.

CIRCADIAN RHYTHMS IN PSYCHIATRIC DISORDERS

Seasonal Affective Disorder

Seasonal affective disorder (SAD) is the most overt manifestation of seasonality. It is characterized by recurrent major depressive episodes followed by periods of remission or change to mania or hypomania that recur on a seasonal basis. SAD is diagnosed as major depressive disorder (MDD), bipolar I, or bipolar II disorder, with the specifier "with seasonal pattern" in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5). The seasonal pattern specifier criteria include:

A. Regular temporal relationship between the onset of major depressive episodes and a particular time of the year (unrelated to obvious season-related psychosocial stressors).
B. Full remissions (or a change to mania or hypomania) also occur at a characteristic time of the year.
C. Two major depressive episodes meeting criteria A and B have occurred in the last 2 years, and no nonseasonal episodes have occurred in the same period.
D. Seasonal major depressive episodes substantially outnumber the nonseasonal episodes over the individual's lifetime.

(Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. text rev. American Psychiatric Association; 2022:214.)

The most prevalent form of SAD has an onset in the late fall and early winter and remits in the late spring and early summer. This condition is frequently referred to as *winter SAD,* or *winter depression,* but is a clinically significant level of depression, unlike the “winter blues.” Surveys indicate the prevalence of winter SAD among the general population to be between 4% and 9%. Women are up to four times as likely as men to be affected, and as much as 20% of the population may have subsyndromal symptoms. Rates of SAD are higher among relatives of those with a confirmed diagnosis of SAD, and twin studies identified a heritability of 29% to 43% for seasonality and/or SAD.

The gold standard treatment for winter SAD is light therapy. About half of individuals with SAD experience remission in response to light therapy, representing one of the highest treatment response rates in the field. However, the mechanism behind light therapy is not yet clear. A typical prescription for light therapy involves 45 to 90 minutes of daily exposure to a broad spectrum, ultraviolet-filtered, white light source of relatively high irradiance (5,000 to 10,000 lux). Studies have suggested that cognitive behavioral therapy (CBT), which includes education on sleep hygiene, may be equally efficacious compared to light therapy for acute treatment and may offer prophylactic benefits against recurrences in subsequent years. Multiple antidepressant medications have also been used successfully to treat winter SAD.

The antidepressant effect of phototherapy in winter SAD patients has given rise to several hypotheses regarding the etiology of the disorder. One hypothesis proposes that SAD patients experience the consequences of a chronically phase delayed circadian clock, suggesting that the aberrant phase angle between the clock and the environment is causative of winter SAD. Although most individuals with SAD report being evening chronotypes, nearly half do not have significantly delayed rhythms. Some studies indicate that SAD patients have a longer duration of elevated melatonin that increases at the same clock time as that of controls but stays elevated longer, thus impinging on the morning hours. Although these data are not consistent with the phase delay hypothesis, they may explain the effectiveness of morning bright light therapy that would acutely suppress the extended melatonin profile of winter SAD patients. It should be noted that the sculpting of the melatonin profile by morning light exposure cannot entirely explain the proven success of phototherapy. In some winter SAD patients, bright light administered during the evening is also antidepressant. In fact, some phototherapy treatment paradigms prescribe morning and evening light exposures, although data suggest that morning light therapy is most effective for the average patient.

The success of light therapy administered at various clock times suggests that winter SAD may not have a circadian-based etiology. An alternate hypothesis proposes that patients experiencing winter SAD are generally less sensitive to light than their healthy counterparts. Such photic insensitivity would become apparent during the decreased light levels of late fall and early winter, depriving these individuals of the threshold of light required to stave off depression. Accordingly, daily supplementation of light through bright phototherapy would be expected to exceed this theoretical threshold.

In general, it cannot be denied that phototherapy has proven to be an effective treatment for winter SAD. The mechanisms by which bright light ameliorates the symptoms of this disease remain unknown.

NONSEASONAL UNIPOLAR DEPRESSION AND CIRCADIAN RHYTHMS

Abnormal circadian rhythms have been documented in those with unipolar, nonseasonal depression. For instance, individuals with depression are more likely to be evening chronotypes and often show delayed rhythms and reduced peak amplitude of melatonin and core body temperature relative to healthy controls. Daily patterns of worsening mood symptoms in the morning or afternoon are evident in nonseasonal depression and coincide with brain glucose metabolism in affect regulation areas of the brain. Several variants in clock genes have been found to be associated with a depression diagnosis, including *Clock* and *Cry,* although reports are inconsistent.

Sleep disturbances are characteristic of nonseasonal depression, particularly insomnia. Additionally, REM latency, or the time from sleep onset to the first REM episode, is shortened, and REM density is increased in depression. As improvement in sleep is an important marker of lower recurrence rates, circadian control of sleep is a potential avenue to improve long-term outcomes in depression. A curious phenomenon related to depression and sleep is that total sleep deprivation can provide a transient antidepressant effect in a majority (approximately 60%) of depressed patients. This “wake therapy” involves sleep restriction for one night or for the latter half of one night but cannot be used as monotherapy because depressive symptoms will return after a full night of sleep, and the treatment effect must be maintained with another type of treatment. Other chronotherapies beyond this type of sleep deprivation include advancing sleep phase, bright light therapy, and melatonin administration. Some studies indicate that bright light therapy can lead to improvement for nonseasonal depression, suggesting that light therapy can be an option for adjunctive treatment in addition to antidepressant medication.

Given that wake therapy can result in fast antidepressant effects, the practice of combining wake therapy with antidepressant medications and/or light therapy shows some promise for achieving a rapid response followed by sufficient maintenance of the antidepressant response. In addition, it appears that nearly all effective antidepressant medications have therapeutic effects on sleep processes, such as inhibiting excessive REM sleep or lengthening REM latency, possibly reflecting underlying circadian resynchronization.

When negative life events occur in depression, this may lead to a disruption of social rhythms (e.g., timing of first daily social encounter after the death of a spouse). Limited evidence suggests that such disruption or irregularity of social rhythms leads to similar disruptions in biologic rhythms and that treatments to restore the regularity of social rhythms may be effective (see the next section for a discussion of Interpersonal and Social Rhythm Therapy). Circadian regulation of hormonal, temperature, and mood rhythms differ, may be differentially influenced by social zeitgebers as well as genetic factors, and are likely to interact, underlining the complexity of circadian influences on mood disorders.

Bipolar Disorder

As in nonseasonal depression, circadian abnormalities have been suggested as a potential important etiologic mechanism in the development of Bipolar disorder. Bipolar disorder has been linked to variants in the core clock genes in humans and in animal models. For example, mice with a mutation in the *Clock* gene (*Clock*Δ19) show manic-like behaviors, including hyperactivity and decreased sleep. These *Clock*Δ19 mice have offered a helpful model to better understand the contribution of circadian disruption in bipolar disorder. *Clock*Δ19 mice have increased dopamine synthesis in the VTA, suggesting the importance of the *Clock* gene in dopaminergic activity in the ventral tegmental area (VTA). Further, expression of CLOCK protein in the VTA rescues most of the *Clock* mutant mice's manic-like behavior, including hyperactivity and anxiety. Together, *Clock*Δ19 mice models suggest the importance of the circadian modulation of dopaminergic activity within the VTA in bipolar disorder.

Human studies in bipolar disorder also point to abnormal circadian regulation as a potential etiologic mechanism. Circadian abnormalities seem to vary based on which state an individual is experiencing. For example, depression is associated with a phase delay, whereas mania is associated with a phase advance. Therefore, stabilization of these rhythms may help prevent or reduce the occurrence of manic or depressive episodes in bipolar individuals. Indeed, interpersonal social rhythm therapy (IPSRT), in which zeitgebers are routinized to effectively leverage circadian forces, was developed in response to the observation that irregular daily routines cause circadian dysregulation and precipitate mood episodes. IPSRT has been shown to be as effective as mood stabilizers for bipolar II disorder. Importantly, light therapy has been shown to unmask manic tendencies in those with a predisposition for developing bipolar disorder, perhaps due to the relatively abrupt circadian changes.

Circadian Sleep Disorders

In the DSM-5, circadian rhythm sleep–wake disorders include multiple disorders with circadian symptoms and presumed circadian etiology. Among these, advanced sleep phase syndrome (ASPS) is a pathologic extreme of the morning lark phenotype. An autosomal-dominant familial form of ASPS (FASPS) has been genetically characterized. Afflicted family members exhibit a striking 4-hour advance of the daily sleep–wake rhythm. They typically fall asleep at approximately 7:30 PM and spontaneously awaken at approximately 4:30 AM. Affected individuals have an SNP in the *PER2* gene leading to a mutant protein that is normally an integral component of the circadian molecular clockwork. Similarly, delayed sleep phase syndrome (DSPS) is likely influenced by genetics. A length polymorphism in a repeat region of the *PER3* gene appears to be associated with morning chronotype in DSPS patients, and the shorter allele is associated with evening chronotype.

Although sleep disorders are discussed in the next chapter such as DSPS, ASPS, hypersomnia, sleep maintenance insomnia in the older adults, jet lag and shift work disorders, and non–24-hour and irregular sleep–wake-type circadian rhythm disorders, all show improvement in response to bright light therapy. Light therapy is also helpful in normalizing sleep during withdrawal from alcohol and could therefore be considered as an adjunctive option for many disorders in which sleep and mood are impaired. Chronotherapeutics, treatments administered either according to biologic rhythms or to normalize biologic rhythms, include bright light therapy, melatonin, and its agonists, and wake and dark therapy. Such treatments are often the first line of treatment in circadian rhythm sleep–wake disorders. In chronotherapeutics, the appropriate timing of these interventions in relation to the environment or the desired schedule (e.g., for use in accommodating shift work) is critical. For example, as previously mentioned, dim light melatonin onset begins to rise 2 to 4 hours before sleep onset. Taking high-dose (5+ mg) exogenous melatonin at bedtime may be soporific (help induce sleep) but could also induce a phase delay in the circadian rhythm relative to the environment, an effect commonly referred to as “jet lag” as one is most likely to experience it when quickly changing time zones by flying.

Memory, Aging, and Neurocognitive Disorders

In general, as humans age and normative age-related memory problems begin, individuals experience a shortening of the circadian period, decreased amplitudes of most circadian rhythms, and phase advances resulting in earlier waking times and bedtimes. As mentioned previously, degeneration of VLPO neurons and other medical and psychosocial comorbidities may exacerbate these changes. Further, aging individuals may be less tolerant of dramatic phase shifts such as those caused by jet lag.

Age-related circadian changes, like decreases in amplitude or misalignment, are exacerbated in neurocognitive disorders, including neurodegenerative disorders (e.g., phase delay in Alzheimer disease [AD], phase advance in frontotemporal dementia [FTD], and sleep–wake reversal in Parkinson disease [PD] and progressive supranuclear palsy [PSP]). For example, those with moderate dementia exhibit fragmented rhythms and decreased amplitude in activity rhythms relative to mild dementia and healthy control counterparts. The amount of disruption in activity rhythms has also been associated with the severity of neuropathology found in postmortem brains, such that those with the most aberrant activity rhythms around their time of death also exhibited the most neuropathology, often with disruption of cholinergic functioning. Dysregulated circadian rhythms have been found to predate the occurrence of neurocognitive disorders (dementia) and/or clinical diagnosis of AD and significantly predict the occurrence of AD or mild neurocognitive disorder. Together, these pieces of evidence would suggest that circadian disruption may play a central role in the development of the neuropathology and neurocognitive deficits associated with AD. Light therapy aimed at normalizing an individual's circadian rhythms has been shown to improve AD patients' sleep and cognitive scores. Likewise, the disruption of cholinergic tone associated with circadian misalignment in these disorders is associated with the phenomena of "sundowning." These disturbances aggravate other symptoms, may be a major source of discouragement, increase the caretaker burden, and lead to earlier nursing home placement.

CIRCADIAN RHYTHMS AND PHARMACOTHERAPY

Circadian rhythmicity can be affected by drugs, and conversely, the circadian clock can modulate the efficacy of drugs throughout the course of the day. A better understanding of these interactions will lead to more effective pharmacotherapies. Some of the best-studied interactions between medications and the circadian clock have included the circadian effects of antidepressants. Elevated nocturnal body temperature is a common feature among depressed patients. This effect may be due to the reduced amplitude of the master circadian oscillator in the hypothalamus that drives body temperature. TCAs and SSRIs reduce elevated nocturnal body temperature while simultaneously enhancing circadian amplitude, and part of their antidepressant effect is thought to be in counteracting the dampened amplitude in daily activity rhythms depressed patients exhibit.

Short-acting benzodiazepines (e.g., triazolam [Halcion] and brotizolam [Lendormin]) also exert chronobiologic effects. In hamsters, triazolam or brotizolam administered during the middle of the subjective day induces circadian phase advances. Brotizolam has been shown to reduce the light-induced expression of clock genes *Per1* and *Per2* in the SCN.

Recreational drugs of abuse also impact the circadian system. 3,4-Methylenedioxymethamphetamine (MDMA) or "ecstasy" can act as a serotonin neurotoxin. Animal models have demonstrated that MDMA reduces medication-induced phase shifts, diminishes the ability to re-entrain rhythms, and reduces serotonergic axonal terminals in the SCN. Chronic administration of methamphetamine disorganizes rodent activity rhythms, although some studies indicate that it may help re-establish rhythmicity in rodents rendered arrhythmic through ablation of the SCN.

The efficacy and toxicity of many pharmacotherapeutics vary as a function of the circadian phase. Daily variations in fixed-dose lethal toxicity have been appreciated in rodents for years. Many anticancer drugs, ranging in mechanism from antimetabolites to DNA intercalators to mitotic inhibitors, have been shown to have 2- to 10-fold changes in tolerability in rodents over the course of the day. Much of this difference is attributed to circadian variations in the body’s ability to absorb, distribute, metabolize, and eliminate toxic compounds. These four processes are affected by underlying circadian rhythms in physiologic processes such as daily variations in gastric pH, gastrointestinal mobility, glomerular filtration rate, and membrane viscosity. The rhythmic intake of food during traditionally timed meals also influences how therapeutic drugs are handled by the body. It is becoming clear that to maximize efficacy and minimize the toxicity of drugs, the circadian phase of administration must be considered. The emergence of the field of chronotherapy is a reflection of the increased understanding of the impact of the circadian system on the effectiveness of pharmacologic treatments.

FUTURE CONSIDERATIONS

A growing body of research indicates that the influence of the circadian system is broad and foundational to normal functioning. Desynchrony between the SCN and peripheral oscillators is likely to be involved in the pathogenesis of multiple maladies, including psychiatric disorders. As our appreciation of the relationship between circadian disturbances and the pathophysiology of psychiatric illness improves, hopefully, treatments designed to manipulate the light input, central clock, or clock output mechanisms will show increasing promise in treating psychiatric disorders.

FURTHER READINGS

Arendt J, Skene DJ. Melatonin as a chronobiotic. *Sleep Med Rev*. 2005;9(1):25–39.

Berger M, Vollmann J, Hohagen F, et al. Sleep deprivation combined with consecutive sleep phase advance as a fast-acting therapy in depression: an open pilot trial in medicated and unmedicated patients. *Am J Psychiatry*. 1997;154:870–872.

Bloom HG, Ahmed I, Alessi CA, et al. Evidence-based recommendations for the assessment and management of sleep disorders in older persons. *J Am Geriatr Soc*. 2009;57(5):761–789.

Bohle P, Di Milla L, Fletcher A, Rajaratnam S. Introduction: aging and the multifaceted influence on adaptation to working time. *Chronobiol Int*. 2008;25:155–164.

Carter DS, Goldman BD. Antigonadal effects of timed melatonin infusion in pinealectomized male Djungarian hamsters (*Phodopus sungorus sungorus*): duration is the critical parameter. *Endocrinology*. 1983;113:1261–1267.

Chen-Goodspeed M, Lee CC. Tumor suppression and circadian function. *J Biol Rhythms*. 2007;22:291–298.

Czeisler CA, Duffy JF, Shanahan TL, et al. Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science*. 1999;284:2177–2181.

Davidson AJ, Sellix MT, Daniel J, Yamazaki S, Menaker M, Block GD. Chronic jet-lag increases mortality in aged mice. *Curr Biol*. 2006;16:R914–R916.

Dijk D-J, Lockley SW. Integration of human sleep-wake regulation and circadian rhythmicity. *J Appl Physiol*. 2002;92:852–862.

Germain A, Kupfer DJ. Circadian rhythm disturbances in depression. *Hum Psychopharmacol*. 2008;23:571–585.

Harvey AG. Sleep and circadian rhythms in bipolar disorder: seeking synchrony, harmony, and regulation. *Am J Psychiatry*. 2008;165:820–829.

Herzog ED, Schwartz WJ. A neural clockwork for encoding circadian time. *J Appl Physiol*. 2002;92:401–408.

Klein DC, Moore RY, Reppert SM, eds. *Suprachiasmatic Nucleus: The Mind's Clock*. Oxford University Press; 1991.

Klerman EB, Shanahan TL, Brotman DJ, et al. Photic resetting of the human circadian pacemaker in the absence of conscious vision. *J Biol Rhythms*. 2002;17:548–555.

Ko CH, Takahashi JS. Molecular components of the mammalian circadian clock. *Hum Mol Genet*. 2006;15:R271–R277.

Kondratova AA, Kondratov RV. The circadian clock and pathology of the ageing brain. *Nat Rev Neurosci*. 2012;13(5):325–335.

Levi F, Schibler U. Circadian rhythms: mechanisms and therapeutic implications. *Annu Rev Pharmacol Toxicol*. 2007;47:593–628.

McClung CA. How might circadian rhythms control mood? Let me count the ways. *Biol Psychiatry*. 2013;74(4):242–249.

Miller JD, Morin LP, Schwartz WJ, Moore RY. New insights into the mammalian circadian clock. *Sleep*. 1996;19:641–667.

Moore RY. Circadian rhythms: basic neurobiology and clinical applications. *Annu Rev Med*. 1997;48:253–266.

Morin LP. The circadian visual system. *Brain Res Brain Res Rev*. 1994;19:102–127.

Nelson DE, Takahashi JS. Sensitivity and integration in a visual pathway for circadian entrainment in the hamster (*Mesocricetus auratus*). *J Physiol*. 1991;439:115–145.

Quintero JE, Kuhlman SJ, McMahon DG. The biological clock nucleus: a multiphasic oscillator network regulated by light. *J Neurosci*. 2003;23:8070–8076.

Ralph MR, Foster RG, Davis FC, Menaker M. Transplanted suprachiasmatic nucleus determines circadian period. *Science*. 1990;247:975–978.

Roecklein KA, Wong PM, Miller MA, Donofry SD, Kamarck ML, Brainard GC. Melanopsin, photosensitive ganglion cells, and seasonal affective disorder. *Neurosci Biobehav Rev*. 2013;37:229–239.

Roybal K, Theobold D, Graham A, et al. Mania-like behavior induced by disruption of CLOCK. *P Natl Acad Sci USA*. 2007;104(15):6406–6411.

Ruby NF, Brennan TJ, Xie X, et al. Role of melanopsin in circadian responses to light. *Science*. 2002;298:2211–2213.

Smith BN, Sollars PJ, Dudek FE, Pickard GE. Serotonergic modulation of retinal input to the mouse suprachiasmatic nucleus mediated by 5-HT1B and 5-HT7 receptors. *J Biol Rhythms*. 2001;16:25–38.

Takahashi JS, Turek FW, Moore RY, Takahashi JS, Turek FW, eds. *Handbook of Behavioral Neurobiology: Circadian Clocks*. Vol. 12. Kluwer Academic Publishers; 2001.

Toh KL. Basic science review on circadian rhythm biology and circadian sleep disorders. *Ann Acad Med Singap*. 2008;37(8):662–668.

Toh KL, Jones CR, He Y, et al. An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science*. 2001;291:1040–1043.

Weekes NY, Lewis RS, Goto G, Garrison-Jakel J, Patel F, Lupien S. The effect of an environmental stressor on gender differences on the awakening cortisol response. *Psychoneuroendocrinology*. 2008;33:766–772.

Yamazaki S, Goto M, Menaker M. No evidence for extraocular photoreceptors in the circadian system of the Syrian hamster. *J Biol Rhythms*. 1999;14:197–201.

1.15 Basic Science of Sleep

Ruth M. Benca, M.D., Ph.D., Chiara Cirelli, M.D., Ph.D., and Giulio Tononi, M.D., Ph.D.

Sleep is a fundamental behavior of all animal species, although its specific functions are not yet fully understood. Sleep occupies approximately one-third of the human lifespan, and loss of sleep can lead to cognitive, emotional, and physical impairment. Systems involved in the regulation of sleep and wakefulness appear to overlap or interact with systems involved in the regulation of emotion and other behaviors. Therefore, it is not surprising that sleep abnormalities are commonly found in patients with psychiatric disorders, part of the diagnostic criteria for several disorders, and have predictive value for the future development of psychiatric disorders. Furthermore, a majority of commonly used psychiatric medications have effects on sleep.

NORMAL HUMAN SLEEP

Definition of Sleep

From a behavioral standpoint, sleep is a state of decreased awareness of environmental stimuli that is distinguished from states such as coma or hibernation by its relatively rapid reversibility. Sleeping individuals move little and tend to adopt stereotypical postures. Although sleep is characterized by a partial disconnection from the external world and a general lack of memory of the state, unlike in comatose states, people generally recognize when they feel sleepy and are aware that they have been asleep at the termination of an episode.

For clinical and research purposes, sleep is generally characterized by combining behavioral observation with electrophysiologic recording. Like most other mammals, humans express two types of sleep: rapid eye movement (REM) and nonrapid eye movement (NREM) sleep. These states have distinctive neurophysiologic and psychophysiological characteristics. REM sleep derives its name from the frequent bursts of eye movement activity that occur. It is also referred to as *paradoxical sleep* because the electroencephalogram (EEG) during REM sleep is similar to that of waking. In infants, the equivalent of REM sleep is called *active sleep* because of prominent phasic muscle twitches. NREM sleep, or *orthodox sleep,* is characterized by decreased activation of the EEG; in infants, it is called *quiet sleep* because of the relative lack of motor activity.

Stages of Sleep

Within REM and NREM sleep, there are further classifications called *stages* (Table 1.15–1 and Fig. 1.15–1). For clinical and research applications, sleep is typically scored in epochs of 30 seconds, with stages of sleep defined by the visual scoring of three parameters: EEG, electrooculogram (EOG), and electromyogram (EMG) recorded beneath the chin. Most of the criteria defined by Rechtschaffen and Kales in 1968 are still accepted in clinical practice and for research around the world, although new rules that modify the older criteria were adopted in 2007 by the American Academy of Sleep Medicine (AASM) in *The AASM Manual for the Scoring of Sleep and Associated Events*, which was most recently updated in 2020. During wakefulness, the EEG shows a low-voltage fast activity or activated pattern. The EMG has a high tonic activity with additional transient muscle activity related to voluntary movements. Voluntary eye movements and eye blinks can also be observed during wakefulness. When the eyes are closed in preparation for sleep, alpha activity (8 to 13 Hz) becomes prominent, particularly in the occipital regions.

Table 1.15–1.
Stages of Sleep—Electrophysiologic Criteria

	EEG	EOG	EMG
Wakefulness			
Stage W	Low-voltage, mixed frequency Alpha (8–13 Hz) with eyes closed, vertex sharp waves	Eye movements and eye blinks	High tonic activity and voluntary movements
Nonrapid Eye Movement			
Stage N1	Low-voltage, mixed frequency Theta (4–7 Hz) and vertex sharp waves may be present	Slow eye movements	Tonic activity slightly decreased from wakefulness
Stage N2	Low-voltage, mixed-frequency background with sleep spindles (12–14-Hz bursts) and K-complexes (negative sharp wave followed by positive slow wave)	None	Low tonic activity
Stage N3	High-amplitude (≥75 μV) slow waves (≤2 Hz) occupying at least 20% of epoch	None	Low tonic activity
Rapid eye movement	Low-voltage, mixed frequency Saw-tooth waves, theta activity, slow alpha activity may be present	Rapid eye movements	Tonic atonia with phasic twitches

Criteria from Iber C, Ancoli-Israel S, Chesson AL, Quan SF. *The AASM Scoring Manual for the Scoring of Sleep and Associated Events*. American Academy of Sleep; 2007.

NREM sleep, which usually precedes REM sleep, is subdivided into three (N1 to N3) stages (Table 1.15–1 and Fig. 1.15–1). Sleep usually begins with a transitional state, stage N1 (formerly stage 1 sleep), characterized by the loss of alpha activity and the appearance of a low-voltage, mixed-frequency EEG pattern with prominent theta activity (4 to 7 Hz) and occasional vertex sharp waves (V waves) over the central regions. Eye movements become slow and rolling, and skeletal muscle tone relaxes. Subjectively, stage N1 may not be perceived as sleep, although there is a decreased awareness of sensory stimuli, particularly visual, and mental activity becomes more dreamlike. Motor activity may persist for a number of seconds during stage N1. Occasionally individuals experience sudden muscle contractions, sometimes accompanied by a sense of falling and/or dreamlike imagery; these hypnic jerks (or sleep starts) are generally benign and may be exacerbated by sleep deprivation. Typically, sleep-deprived individuals will enter periods of “microsleep” consisting of brief (5 to 10 seconds) bouts of stage N1 sleep; these episodes are unavoidable in sleepy individuals and can have serious consequences in situations that demand constant attention, such as driving a motor vehicle.

After a few minutes of stage N1, sleep usually progresses to stage N2 (formerly stage 2), which is heralded by the appearance of sleep spindles (11 to 16 Hz, lasting ≥0.5 seconds) and K-complexes (high-amplitude, negative sharp waves followed by positive slow waves, lasting ≥0.5 seconds) in the EEG. Stage N2 and subsequent stages of NREM and REM sleep are all subjectively perceived as sleep. Particularly at the beginning of the night, stage N2 is generally followed by N3 (formerly stages 3 and 4), a period when 20% or more of each sleep epoch consists of slow waves, that is, waves of 0.5- to 2-Hz frequencies with peak-to-peak amplitudes of >75 μV over frontal regions. N3 is also defined as *slow-wave sleep (SWS)*, *delta sleep,* or *deep sleep* because the arousal threshold increases incrementally from stages N1 to N3. Previously, SWS was subdivided according to the proportion of slow waves in the epoch (stage 3, 20% to 50%; stage 4, >50%), and older studies used this nomenclature, but the validity and biologic significance of this subdivision was called into question. Eye movements typically cease during stages N2 and N3, and EMG activity decreases further.

REM sleep, or stage R, is not subdivided into stages but is rather described in terms of tonic (persistent) and phasic (episodic) components.

Tonic aspects of REM sleep include the activated EEG similar to that of stage N1, which may exhibit increased activity in the theta band and a generalized decrease of the tone of skeletal muscles except for the extraocular muscles and the diaphragm. Sawtooth waves, trains of triangular, serrated 2- to 6-Hz waves may also be present. Phasic features of REM include irregular bursts of REMs and muscle twitches. An example of stage R showing tonic and phasic components is shown in Figure 1.15–2.

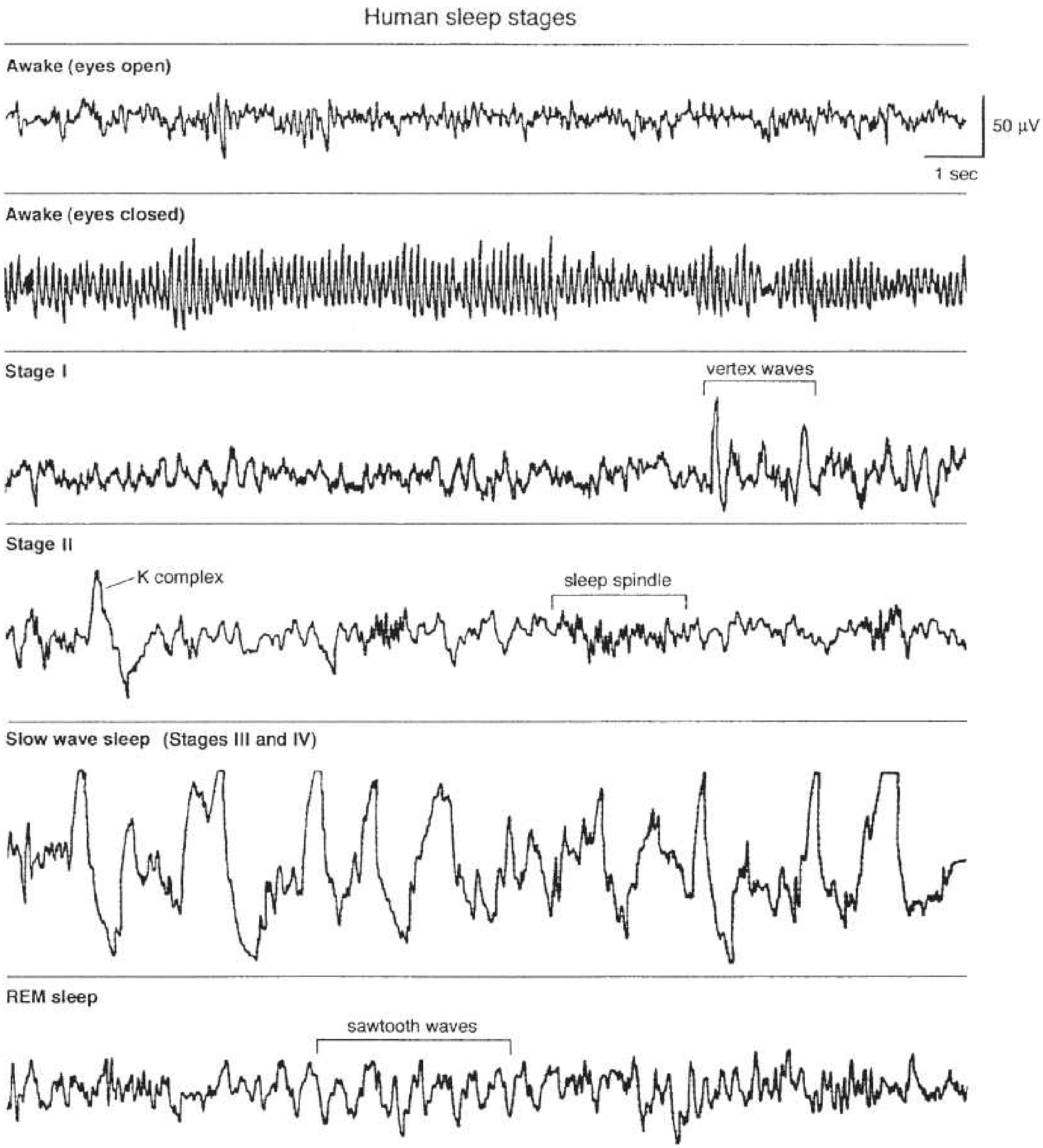

FIGURE 1.15–1. Electroencephalogram patterns for stages of sleep and wakefulness. REM, rapid eye movement. (From Butkov N. *Atlas of Clinical Polysomnography*. Synapse Media; 2010, with permission.)

Organization of Sleep

The amount of sleep obtained during the night varies among individuals and with age; the National Sleep Foundation convened an expert panel to review the scientific literature to make recommendations for sleep duration across the lifespan. They concluded that most adults need about 7 to 9 hours of sleep per night to function optimally, although those over

65 might only require 7 to 8 hours. Although these are population-based recommendations, there also exist short sleepers who appear to function adequately with less than 6 hours per night and long sleepers who may need 12 or more hours per night. In addition to age and genetic factors that influence daily sleep needs, medical or psychiatric disorders also strongly influence sleep patterns. Regardless of the number of hours needed, the proportion of time spent in each stage and the pattern of stages across the night is fairly consistent in normal adults (Fig. 1.15–3). A healthy young adult will typically spend about 5% of the sleep period in stage N1 sleep, about 50% in stage N2, and 20% to 25% in each of stages N3 and R. Sleep occurs in cycles of NREM–REM sleep, each lasting approximately 90 to 110 minutes. SWS (stage N3) is most prominent early in the night, especially during the first NREM period, and diminishes as the night progresses. As SWS wanes, periods of REM sleep lengthen while showing greater phasic activity and generally more intense dreaming later in the night. Older adults have decreased amounts of SWS.

The ascending arousal system sends projections from the brainstem and posterior hypothalamus throughout the forebrain. Neurons of the laterodorsal tegmental (LDT) nuclei and pedunculopontine tegmental (PPT) nuclei send cholinergic fibers (Ach) to many forebrain targets, including the thalamus, which then regulate cortical activity. Aminergic nuclei diffusely project throughout much of the forebrain, regulating the activity of cortical and hypothalamic targets directly. Neurons of the tuberomammillary nucleus (TMN) contain histamine (HIST), neurons of the raphe nuclei contain serotonin (5-HT), and neurons of the locus coeruleus (LC) contain noradrenaline (NA). Sleep-promoting neurons of the ventrolateral preoptic (VLPO) nucleus and other regions of the anterior hypothalamus and basal forebrain contain GABA and galanin (Gal).

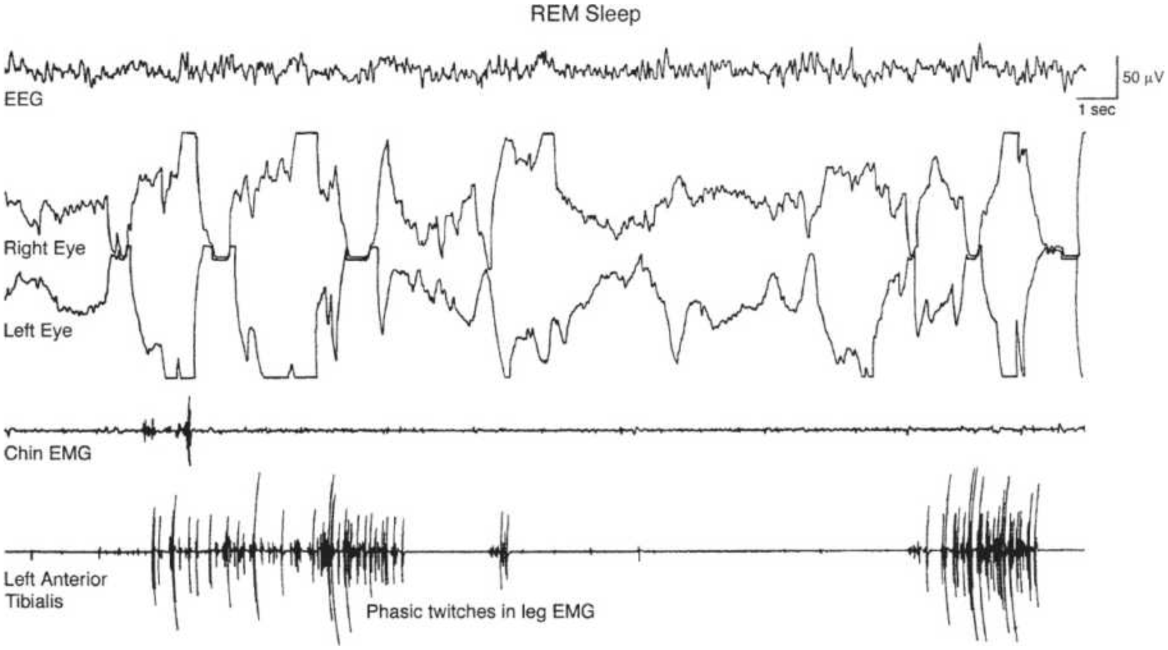

FIGURE 1.15–2. Example of human rapid eye movement (REM) sleep. EEG, electroencephalogram; EMG, electromyogram. (From Butkov N. *Atlas of Clinical Polysomnography*. Synapse Media; 2010, with permission.)

EFFECTS OF AGE ON SLEEP

Sleep patterns change markedly across the lifespan, with the most rapid changes occurring during the first years of life. Development of EEG patterns of sleep and wakefulness begins at about 24 weeks of gestational age, and differentiation into active (REM) and quiet (NREM) sleep occurs during the last trimester. Newborn infants spend about 14 to 17 hours per day sleeping, and premature infants may sleep even more. Infants do not show evidence of strong diurnal sleep patterns for the first several months of life; they exhibit short sleep–wake cycles of about 3 to 4 hours and a reduced length of active–quiet sleep cycles (about 50 minutes). At birth, active sleep occupies about half of their sleep time, and they tend to enter sleep through active rather than quiet sleep. At approximately 3 to 4 months of age, several important developmental changes occur: Babies shift to the adultlike pattern of initiating sleep with NREM, sleep starts to become consolidated during the night, and the sleep EEG shows more mature waveforms characteristic of NREM and REM sleep. Sleep duration also drops slightly, to 12 to 15 hours per day.

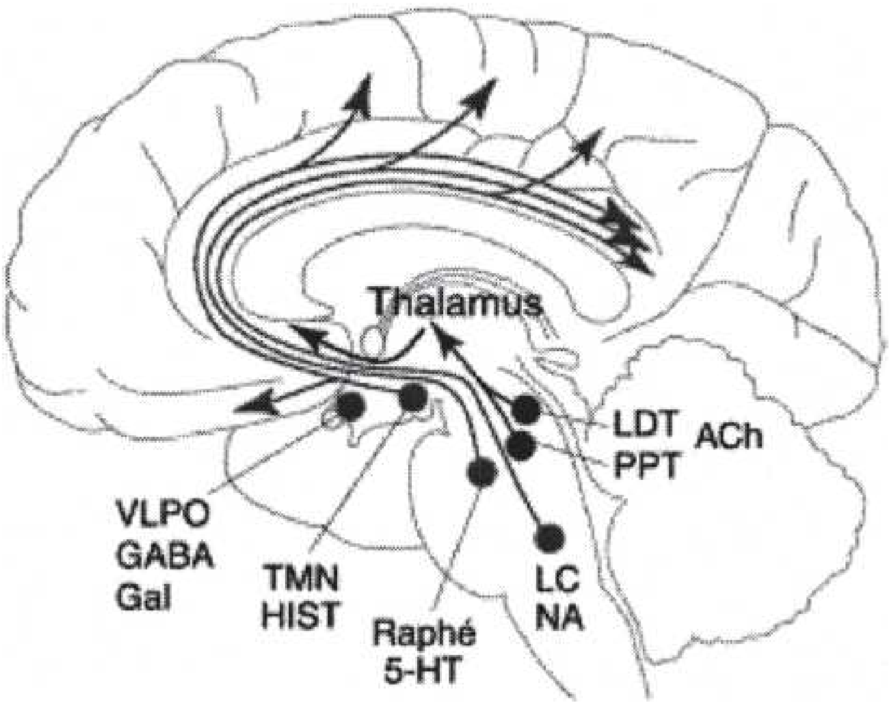

FIGURE 1.15–3. The ascending arousal system sends projections from the brainstem and posterior hypothalamus throughout the forebrain. Neurons of the laterodorsal tegmental (LDT) nuclei and pedunculopontine tegmental (PPT) nuclei send cholinergic fibers (Ach) to many forebrain targets, including the thalamus, which then regulate cortical activity. Aminergic nuclei diffusely project throughout much of the forebrain, regulating the activity of cortical hypothalamic targets directly. Neurons of the tuberomammillary nucleus (TMN) contain histamine (HIST), neurons of the raphe nuclei contain serotonin (5-HT), and neurons of the locus coeruleus (LC) contain noradrenaline (NA). Sleep-promoting neurons of the ventrolateral preoptic (VLPO) nucleus contain γ-aminobutyric acid (GABA) and galanin (Gal). (Reprinted from Saper CB, Chou TC, Scammell TE. The sleep switch: hypothalamic control sleep and wakefulness. *Trends Neurosci.* 2001;24(12):726–731. Copyright © 2001, with permission from Elsevier.)

During early childhood, total sleep time continues to decrease, such that recommended sleep amounts for preadolescent school-aged children range from 9 to 11 hours per night, and REM sleep proportion drops to adult levels (20% to 25%). Napping normally continues during the preschool years but is often abandoned once children begin school full-time. Young children show the highest percentages of SWS, with particularly high arousal thresholds; this accounts for both the difficulty in arousing them at the beginning of the sleep period as well as the high incidence of bedwetting and SWS-related parasomnias such as sleepwalking and night terrors.

SWS duration and the amplitude of slow waves diminish significantly during adolescence, possibly related to cortical synaptic pruning. Irwin Feinberg has suggested that abnormalities in synaptic elimination may account for the seemingly coincidental timing of the maturation of sleep patterns and the increasing incidence of schizophrenia in late adolescence and early adulthood. In addition to showing decreases in SWS duration, adolescents often decrease their total sleep time significantly, although this is probably due to behavioral changes rather than representing a true decrease in sleep need; their recommended sleep amount is estimated at 8 to 10 hours per night. They also show a tendency to become “night owls,” preferring to stay up late rather than wake up early. This shift to “eveningness” may be related to an increase in the intrinsic period of the circadian clock.

In addition to the changes in amounts of SWS, there are dramatic changes in the topography of scalp slow-wave activity (SWA) that mirror the pattern of cortical brain maturation. Occipital regions that develop first show the most SWA in very young children. As the cortex matures in a posteroanterior direction during development, maximal SWA also shifts forward, showing the adult pattern of dominance over the frontal cortex by late adolescence/early adulthood.

Older adults show an increased incidence of primary sleep disorders (e.g., sleep apnea and periodic limb movements [PLMs]), medical illnesses, and psychiatric disorders that may interfere with sleep; it is, therefore, difficult to determine which sleep changes represent normal versus pathologic aging. SWS declines across adulthood and may disappear entirely by age 60; SWA declines globally over the scalp. Sleep also becomes more fragmented, with prolonged latency to sleep onset, increased numbers of arousals, more time spent awake during the sleep period, and increased daytime napping. Older adults thus may spend more time in bed but obtain less sleep; 5 to 6 hours may be a sufficient amount of sleep for some individuals. The older adults also show a trend to wake up earlier and feel more alert in the morning than in the evening. The percentage of REM sleep only shows a small decline with age in normal individuals. Patients with Alzheimer disease and other degenerative disorders of the CNS, however, show a loss of REM sleep and deterioration in diurnal patterns of sleep–wakefulness; this deterioration may become so severe that nursing home patients may not spend a single hour of the day where they are consistently either asleep or awake.

MONITORING HUMAN SLEEP

Sleep–wakefulness-related alterations in the human EEG were first reported by Hans Berger in 1929. Since this discovery, the EEG and other electrophysiologic parameters have been used to investigate normal human sleep. Starting in the 1970s, researchers and clinicians began to use similar monitoring techniques to characterize and diagnose sleep disorders. In addition to the EEG, the EOG and EMG are recorded to measure eye movements and muscle activity, respectively, both requisites to distinguish wakefulness and the stages of sleep. The EEG is recorded from electrodes affixed to the scalp overlying specific regions of the brain according to the International 10–20 system of electrode placement. Because the EEG features that define wakefulness and the stages of sleep are most readily recorded bilaterally from different regions of the brain, a minimum of three strategically chosen regions (frontal, central, and occipital) are required. K-complexes and slow waves are optimally recorded with the frontal pair of electrodes. Sleep spindles (Fig. 1.15–1) are instead best recorded from electrodes placed over the central region. The third pair of electrodes is placed over the occipital lobe to optimize the detection of alpha activity, correlated to a relaxed waking state with closed eyes. Additional electrodes are applied to obtain a higher spatial resolution of EEG activity under certain research or clinical circumstances (e.g., diagnosis of sleep-related seizure disorders). The EOG recording is used to detect REMs associated with wakefulness and REM sleep, as well as the slow rolling eye movements that occur during stage N1 sleep. Because the retina is electrically negative compared to the cornea, eye movements generate small electrical fields that can be detected from electrodes attached to the skin near the eyes. The EMG recording is used to detect tonic and phasic changes in muscle activity that correlate with changes in behavioral state. In particular, skeletal muscle tone reaches the lowest tonic levels during REM sleep, reflecting the general paralysis associated with this stage of sleep. Typically, the EMG is recorded from electrodes attached to the chin. In clinical settings, additional EMG electrodes are placed over the anterior tibialis and intercostal muscles to detect leg movements and respiratory effort, respectively. Depending on the presenting symptoms, clinical monitoring may also include additional monitoring such as respiratory effort and flow sensors, ECG, oxyhemoglobin saturation, end-tidal and/or transcutaneous CO_2 levels, and limb EMG recording.

Standard Sleep Stage Scoring

Standardized visual scoring rules were first instituted by Rechtschaffen and Kales in 1968 and are still used, with more recent modifications, to quantify time spent in wakefulness and each sleep stage as well as the temporal distribution, or architecture, of sleep across the recording period. In addition, mathematical techniques, such as power spectral analysis based on the *Fast Fourier Transform*, are frequently employed to quantify the relative contributions of various brain wave frequencies to the overall EEG recording. For example, power spectral analysis has shown that SWA is greatest early in the sleep period and progressively declines across successive periods of SWS, a finding that provided the basis for the two-process model of sleep regulation.

Sleep stage scoring yields several measures of sleep quantity and quality, including clinically relevant markers of sleep and psychiatric disorders. Specific clinical or research circumstances may call for slightly different definitions for each sleep measure. *Sleep latency* is the time elapsed from the start of the recording to the onset of any stage of sleep. *REM sleep latency* is defined as the time elapsed from sleep onset to the first epoch of stage R sleep. *Total sleep time* is the cumulative time spent in all sleep stages, and *sleep efficiency* is how much of the total recording time was spent in any stage of sleep. Other measures of interest include the proportion of sleep spent in each sleep stage.

Other clinical events scored in clinical sleep studies include respiratory events and limb movements. Respiratory events include apneas (drops in respiratory sensor signal of at least 90% for at least 10 seconds) and hypopneas (drops in respiratory sensor signal of at least 30% for at least 10 seconds and with an oxygen saturation drop of at least 3%). Respiratory events are classified as obstructive if respiratory effort continues and central when there is no inspiratory effort. Sleep apnea is considered mild with an average of 5 or more respiratory events per hour of sleep, and moderate to severe with 15 or more events per hour. Limb movements are recorded from electrodes placed over the anterior tibialis muscles, and a leg movement is scored when there is an increase of at least 8 µv that lasts for at least 0.5 seconds. When leg movements occur repetitively during sleep (i.e., at least four leg movements separated by 5 to 90 seconds), they are considered PLMs; 15 or more per hour is considered abnormal.

Measuring Daytime Sleepiness

In addition to characterizing sleep patterns across the major sleep period, researchers and clinicians are sometimes interested in quantifying daytime sleepiness. Daytime sleepiness manifests as an increased propensity to fall asleep and/or a decreased ability to maintain wakefulness. Sleepiness can be measured using subjective questionnaires or objective electrophysiologic monitoring. Sleep questionnaires such as the Stanford Sleepiness Scale (SSS) or Epworth Sleepiness Scale (ESS) are easy to use in the laboratory or physician's office. The SSS asks the individual to rate their current level of sleepiness, whereas the ESS asks the individual to rate their probability of falling asleep under various circumstances. Both questionnaires may be influenced by the individual's ability to assess their level of alertness or propensity to fall asleep, as well as their motivation to obtain treatment.

Objective measures of sleepiness generally require time-consuming, laboratory-based testing but yield a more reliable estimate of sleepiness. Two standardized objective tests are typically employed to measure daytime sleepiness; the multiple sleep latency test (MSLT) measures the propensity to fall asleep, whereas the maintenance of wakefulness test (MWT) measures the ability to maintain wakefulness. Both tests employ the electrophysiologic sleep-monitoring techniques described above. The MSLT consists of four or five sleep latency tests spaced 2 hours apart. At the start of each test, individuals are asked to lie quietly in a darkened room and allow themselves to fall asleep. The latency to sleep onset is recorded, and if sleep does not occur within 20 minutes, the test is terminated. In the clinical setting, patients who fall asleep are usually allowed to sleep for 15 minutes to determine whether they tend to enter REM sleep in an abnormally short period of time. The mean sleep latency is calculated as well as the number of tests in which REM sleep was detected. A mean sleep latency of <8 minutes reflects excessive sleepiness and may indicate a CNS hypersomnia such as narcolepsy or idiopathic hypersomnia, whereas a mean sleep latency ≥15 minutes is considered normal. Idiopathic hypersomnia may also be diagnosed in individuals who sleep >660 minutes on a 24-hour polysomnographic recording in the absence of chronic sleep deprivation. In conjunction with other symptoms, two or more tests with REM sleep are suggestive of narcolepsy. During the MWT, individuals are placed under similar conditions to that described for the MSLT, but rather than being asked to fall asleep, they are asked to stay awake. Although 20- to 40-minute tests can be used, the 40-minute protocol has the advantage of minimizing ceiling effects. A mean sleep latency >35 minutes on the 40-minute MWT is considered normal. Although both the MSLT and the MWT are objective tests of sleepiness, the results can be influenced by various factors, such as an individual's prior sleep history. In the clinical setting, the goal is to assess whether a patient exhibits a pathologic degree of sleepiness despite obtaining normal amounts of sleep at night; short sleep latencies on the MSLT are of limited clinical utility if the patient is clearly sleep deprived. Consequently, it is critical that the patient's prior sleep history be evaluated by monitoring in the sleep laboratory during the preceding night and questionnaires logging sleep habits for the previous week. Under certain circumstances, patients may wear small devices on their wrist that detect and log motion, a correlate of wakefulness, for several days or weeks before undergoing laboratory testing. Such activity monitoring (or actigraphy) is also used to obtain objective measures of probable sleep at home in patients complaining of insomnia. Other objective measures of sleepiness based on power spectral analysis of the EEG, pupillary activity, and performance on cognitive tests are currently under development.

PHYSIOLOGY IN SLEEP

An appreciation of the physiologic changes that occur during sleep helps understand the effects of normal and disturbed sleep on medical disorders. Various neuroendocrine and autonomic nervous system (ANS) physiology parameters show changes related to circadian rhythms and/or sleep itself (Table 1.15–1).

Autonomic Nervous System

During NREM sleep and tonic REM sleep, there is a relative increase in parasympathetic activity relative to sympathetic activity. The ANS reaches its most stable state during SWS in comparison to wakefulness; for example, blood pressure, heart rate, and respiratory rate are at their lowest mean values and least variable during SWS. However, during phasic REM sleep, there are brief surges in both sympathetic and parasympathetic activity, resulting in a high degree of autonomic instability.

Cardiovascular System

Blood pressure, heart rate, and cardiac output decrease during NREM sleep, reaching their lowest average values and least variability in SWS. Although on average, these parameters remain somewhat reduced during REM sleep in comparison to waking, they attain their peak values during REM sleep. Arrhythmias are also more prevalent during REM sleep, which may contribute to the increased rate of cardiovascular mortality in the early morning, the time of greatest REM sleep propensity. It is also possible that increased mortality rates related to cardiovascular causes seen in major depression might be related to the tendency of depressed patients to have increased amounts of REM sleep with greater phasic activity.

Pulmonary System

Temporary breathing instability and/or periodic breathing may occur at the onset of sleep and are related to the loss of waking-related respiratory drive as well as a decreased sensitivity of central chemoreceptors to pCO_2; sensitivity to pCO_2 declines further during REM sleep, along with a decrease in the ventilatory response to reduced pO_2. Respiratory rate and minute ventilation decrease during sleep and upper airway resistance increases as a result of muscle relaxation, most significantly during REM sleep. These changes contribute to exacerbations of underlying pulmonary disease as well as sleep-related breathing disorders such as sleep apnea.

Thermoregulation. In addition to the circadian-related nocturnal decrease in body temperature, sleep directly affects thermoregulation. Brain and body temperature are downregulated during NREM sleep, particularly SWS, due to both a decreased hypothalamic temperature set point and active heat loss. People commonly experience this phenomenon when they go to sleep feeling somewhat cold and wake up several hours later to throw off their extra covers because they feel too warm and may have been sweating. There is a decreased ability to regulate body temperature during REM sleep through sweating and shivering.

Neuroendocrine Changes. Most hormones regulated by the circadian system also show significant interactions with sleep–wakefulness patterns. Growth hormone (GH) is released primarily during the early part of the night, and SWS enhances its secretion. Sleep also stimulates prolactin secretion, although prolactin peaks after GH, usually during the middle portion of the night. However, pulses of GH and prolactin can occur after sleep onset, regardless of the timing. Both GH and prolactin may have feedback effects on sleep as well; GH seems to enhance SWS, whereas prolactin may increase REM sleep. In contrast, *TSH* reaches its peak level in the evening just before sleep onset; its secretion is inhibited by sleep and stimulated by sleep deprivation. The *HPA axis* is usually at its most inactive state at nocturnal sleep onset. Sleep onset inhibits cortisol release, whereas *ACTH* and *cortisol* levels rise at the end of the sleep period, shortly before awakening, and likely contribute to morning arousal. Severe sleep disruption or sleep deprivation may have significant clinical effects on the endocrine system; for example, patients with obstructive sleep apnea show decreased levels of GH and prolactin, and sleep deprivation produces evidence of HPA axis activation in the evening of the day following deprivation.

Melatonin secretion is mediated by a combination of circadian control and the effects of the light–dark cycle. It can only be released at night if it is dark; darkness does not stimulate melatonin secretion during the day. Thus melatonin can transduce the duration of the photoperiod; it has a significant role in timing reproductive function in some mammals, although its function in humans is less clear. Melatonin can feedback on the circadian clock and may serve to maintain entrainment, which is why it is sometimes recommended for the treatment of jet lag or sleep schedule disorders. It may also have a modest hypnotic effect in humans but can produce arousal in nocturnal animals, suggesting that it acts as a modulator of nocturnal behaviors.

Sexual Function

One of the characteristics of REM sleep in men is the occurrence of penile erections, beginning in infancy and persisting into old age. Nocturnal penile tumescence studies are therefore helpful in determining whether cases of impotence are related to organic or psychogenic etiologies. In women, REM sleep produces increased vaginal blood flow and clitoral erection. These changes are not necessarily linked to sexual content in associated dreams.

NEUROBIOLOGY OF SLEEP AND WAKEFULNESS

Sleep and wakefulness are governed by separate yet interacting systems. Although the specific mechanisms are not fully understood, it is clear that multiple structures and systems in the medulla, brainstem, hypothalamus, and basal forebrain are involved in the orchestration of wakefulness, NREM sleep, and REM sleep. No single brain lesion has been able to produce persistent insomnia or sleep. Given the importance of these behaviors to survival, it is not surprising that there is a certain amount of apparent redundancy in their mechanisms.

Wakefulness

As mentioned above, the waking EEG is characterized by an activated pattern with low-voltage fast activity. Correspondingly, PET studies show that blood flow and metabolic activity are higher during resting wakefulness than in NREM sleep. As indicated by increased regional cerebral blood flow (rCBF), the most active brain areas include the prefrontal cortex, anterior cingulate, parietal cortex, and precuneus, which are areas known to be involved in attention, cognition, and memory.

Maintenance of wakefulness is dependent on the *ascending reticular activating system* (ARAS), which is comprised of inputs from the oral pontine and midbrain tegmentum as well as posterior hypothalamus (Fig. 1.15–4). In animals, electrical stimulation of these regions produces an activated EEG pattern and behavioral arousal. Lesions may produce a comalike state, although recovery of cortical activity can sometimes occur given sufficient time, suggesting that structures outside the brainstem reticular formation are also involved in maintaining wakefulness. Clinically, lesions in the midbrain, diencephalon, or posterior hypothalamus can produce somnolence, stupor, or coma.

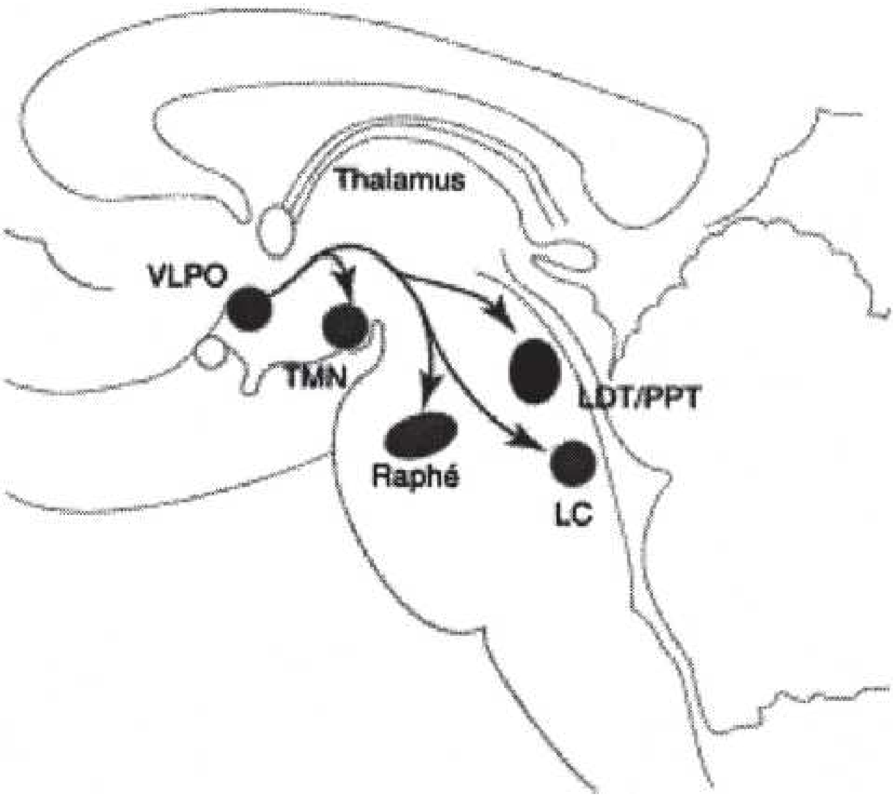

FIGURE 1.15–4. The projections from the ventrolateral preoptic (VLPO) nucleus to the main components of the ascending arousal system. LC, locus coeruleus; LDT, laterodorsal tegmental nuclei; PPT, pedunculopontine tegmental nuclei; TMN, tuberomammillary nucleus. (Reprinted from Saper CB, Chou TC, Scammell TE. The sleep switch: hypothalamic control sleep and wakefulness. *Trends Neurosci.* 2001;24(12):726–731. Copyright © 2001, with permission from Elsevier.)

It is now known that several distinct structures and neurochemical systems with diffuse projections are involved in wakefulness, including noradrenergic cells in the LC, cholinergic cells in the *PPT* and *LDT* nuclei, histaminergic cells in the *TMN* of the posterior hypothalamus, orexin/hypocretin cells in the lateral hypothalamus, and glutamatergic and GABAergic neurons in various structures in the CNS. The ARAS produces cortical activation via input to the thalamus as well as through an extrathalamic pathway with projections to the hypothalamus and basal forebrain.

Noradrenergic cells from the LC project directly throughout the forebrain and cerebral cortex and show their highest discharge rates during wakefulness. These cells decrease their firing during NREM sleep and cease firing altogether during REM sleep. Converging evidence suggests that wakefulness and sleep differ not only in terms of behavior, metabolism, and neuronal activity but also in terms of gene expression. LC cells are responsible for at least some of the changes in gene expression that occur in the brain between wakefulness and sleep, and specifically for the upregulation of plasticity-related genes involved in synaptic potentiation during wakefulness. The LC has long been considered a homogeneous cell group with broad and uniform action on the cortex, but recent experiments found little overlap among the LC projections to three subregions of the rat prefrontal cortex, including the medial prefrontal cortex (mPFC), or between the LC afferents to each of these prefrontal areas and the primary motor cortex, M1. Moreover, relative to the neurons targeting M1, LC cells projecting to mPFC have a spontaneous firing rate three times as high, and are more excitable, suggesting that decision-making prefrontal areas are under a stronger modulatory influence by the LC than movement-generating cortical areas.

Cholinergic cells from the oral pontine region fire at high rates when the EEG is activated (i.e., during wakefulness and REM sleep) but reduce firing during NREM sleep. They promote cortical activation through inputs to the thalamus, hypothalamus, and basal forebrain. In addition, cholinergic cell bodies in the basal forebrain, including the nucleus basalis, substantia innominata, diagonal band of Broca, and septum, receive input from the ARAS and in turn provide excitatory input to the entire cortex. Stimulation of the basal forebrain results in cortical release of acetylcholine, EEG activation, and depolarization of cortical neurons. Moreover, the cholinergic cells of the basal forebrain also fire maximally during wakefulness and REM sleep and minimally during NREM sleep. Optogenetic experiments in mice suggest that cholinergic cells in the pons promote REM sleep more than wake, while cholinergic cells in the basal forebrain may mainly promote EEG activation. Loss of cholinergic cells in Alzheimer patients is associated with slowing of the cortical EEG. Drugs with anticholinergic activity, including tricyclic antidepressants and atropine, can cause sedation, suppress REM sleep, and increase SWA. However, cholinergic agonists (e.g., nicotine) or anticholinesterase inhibitors (e.g., neostigmine) enhance arousal.

Histaminergic neurons in the TMN of the posterior hypothalamus also appear to have an important wakefulness-promoting function, in part inferred from the fact that antihistaminergic drugs typically produce sedation. The significance of this region for waking was first identified by

Constantin von Economo in the early part of the 20th century following an outbreak of viral encephalitis; encephalitis lethargica, as it was called, produced lesions of the posterior hypothalamus and profound somnolence. Histaminergic TMN neurons project throughout the cortex and, like noradrenergic cells, fire at the highest rates during wakefulness and are inhibited during sleep. However, cholinergic and noradrenergic cells in the pons increase firing hundreds of milliseconds or even a few seconds before the EEG switches from the synchronized NREM sleep pattern to the activated wake pattern suggesting a causal role in arousal, while histaminergic neurons increase firing only at wake onset. Histamine infusion into the CNS causes arousal, whereas experimental lesions of the TMN decrease waking and increase SWS and REM sleep. The wakefulness-promoting effect of histamine is mediated by H_1 receptors. Histamine promotes wakefulness in the thalamus, cortex, basal forebrain, and pontine tegmentum by enhancing glutamatergic and cholinergic transmission.

The dopaminergic system also appears to modulate arousal. Dopamine-containing neurons in the substantia nigra and VTA innervate the frontal cortex, basal forebrain, and limbic structures. The mean firing rate of these cells does not change across behavioral states. However, their bursting activity, which is known to induce synaptic dopamine release, increases during the consumption of palatable food and during REM sleep relative to NREM sleep. Lesions of areas containing dopaminergic cell bodies in the ventral midbrain or their ascending pathways can lead to the loss of behavioral arousal while maintaining cortical activation. Psychostimulants such as amphetamines and cocaine that block the reuptake of monoamines, including norepinephrine, dopamine, and serotonin, promote prolonged wakefulness and increase both cortical activation and behavioral arousal.

The importance of the peptide hypocretin (orexin) in the maintenance of wakefulness was defined through discovering its role in the disorder narcolepsy. Hypocretin is produced by cells in the lateral hypothalamus that provide excitatory input to all components of the ARAS, including the LC, PPT and LDT, VTA, basal forebrain, and TMN. These cells, too, are most active during waking, especially in relation to motor activity and exploratory behavior, almost completely stop firing during both NREM and REM sleep, and start firing again before the EEG activates upon arousal. Optogenetic stimulation of hypocretin cells promotes arousal from sleep, although not as quickly as the stimulation of noradrenergic cells. Narcolepsy in animal models is related to deficits in the hypocretin system; canine narcolepsy is caused by a mutation in the hypocretin type 2 receptor gene, whereas narcoleptic symptoms (sleep attacks and sleep-onset REM periods) occur in hypocretin knock-out mice. Mouse experiments also show that the slow and progressive elimination of hypocretin cells mimics many aspects of human narcolepsy: sleep fragmentation is the first symptom, and cataplexy develops only after ~95% of all hypocretin cells are lost. Human narcoleptics show the loss of hypocretin cells and/or protein in the CSF. Hypocretin cell loss is also present in Parkinson disease, which is often associated with sleep disturbances similar to narcolepsy.

Serotonergic cells from the *dorsal raphe* nucleus also project widely throughout the cortex. Serotonergic neurons, like noradrenergic neurons, fire at higher levels during waking, lower levels in NREM sleep, and fall silent during REM sleep. However, in sharp contrast to noradrenergic neurons, serotoninergic neurons are inactivated during orientation to salient stimuli and are activated instead during repetitive motor activity such as locomotion, grooming, or feeding. Selective Serotonin Reuptake Inhibitors (SSRIs) tend to decrease sleep time and increase arousal during sleep. However, the role of serotonin in sleep is not straightforward since there is also evidence that serotonin may be involved in sleep induction (see below).

Several other neurotransmitters and neuromodulators appear to have wakefulness-promoting effects. These include substance P, neurotensin, epinephrine, and hypothalamic peptides such as corticotrophin-releasing factor, VIP, and thyrotropin-releasing factor, all of which can increase arousal levels. Cortisol also promotes wakefulness. It is thus possible that sleep disturbance in depression, including early morning awakening, could be related in part to the associated hyperactivity of the HPA axis.

NREM Sleep

The EEG of NREM sleep is dramatically different from that of waking and is characterized by oscillatory waveforms such as sleep spindles, K-complexes, slow waves (0.5 to 2 Hz), and slow oscillations (mainly 0.7 to 1 Hz). Brain activation generally decreases in NREM sleep, particularly SWS, which is characterized by an overall decrease in CBF. PET imaging studies show the deactivation of many structures, including the brainstem, thalamus, anterior hypothalamus, basal forebrain, basal ganglia, cerebellum, and frontal, parietal, and mesiotemporal cortical areas.

The control of NREM sleep, like wakefulness, involves multiple structures ranging from the lower brainstem through the thalamus, hypothalamus, and forebrain. Electrophysiologic studies in animals have clarified the production of various thalamocortical oscillations that produce the characteristic waveforms in NREM sleep. The generation of sleep oscillations requires the interplay between intrinsic cellular properties and synaptic activity mediated by corticocortical, corticothalamocortical, and thalamoreticular loops. Work in animals has shown that, shortly before the transition from waking to sleep, changes in the activity of cholinergic, noradrenergic, histaminergic, hypocretinergic, and glutamatergic neuromodulatory systems with diffuse projections belonging to the ARAS bring about a change in the firing mode of thalamic and cortical neurons. Thalamocortical cells are hyperpolarized, whereas reticulothalamic cells are facilitated and further inhibit thalamocortical cells, with the consequence that sensory stimuli are gated at the thalamic level and often fail to reach the cortex. Rebound firing due to the activation of intrinsic currents in thalamocortical cells leads to the emergence of oscillations in the spindle frequency range within local thalamoreticular circuits. Local thalamic spindle sequences are globally synchronized and grouped with other rhythmic activities by the slow oscillations that originate in the cortex. Intracellular recordings have shown that the slow oscillation is the result of a brief hyperpolarization of cortical neurons (lasting a few hundred milliseconds), which is seen in the surface EEG as a high-amplitude negative wave. The hyperpolarization phase, also known as the down state, is followed by a slightly longer depolarization phase, known as the up state, during which the firing of cortical neurons entrains and synchronizes spindle sequences in thalamic neurons, resulting in EEG-detectable spindles. K-complexes are made up of the cortical depolarization phase followed by its triggered spindle. The slow oscillation also organizes the slow (also called) delta waves (0.5 to 4.5 Hz), which can be generated both within the thalamus and in the cortex. The hyperpolarization phase of the slow oscillation is associated with the virtual absence of synaptic activity within cortical networks. By contrast, during the depolarized phase, cortical cells fire at rates that can be higher than those seen in waking. Moreover, during the depolarized phase, thalamocortical neurons often fire rhythmically in the gamma range (around 40 Hz), which was previously thought to be exclusively associated with activated states such as wakefulness and REM sleep. The slow oscillation is found in virtually every cortical neuron and is synchronized across the cortical mantle by corticocortical connections. Intrinsic currents are also involved in initiating and terminating the oscillation.

The importance of hypothalamic structures for sleep induction was recognized in early studies in which electrical stimulation of the anterior hypothalamus resulted in increased SWA in the cortex. Conversely, some cases of encephalitis lethargica, in which lesions occurred in the anterior rather than posterior hypothalamus, were characterized by severe insomnia. A few years ago, attention became focused on a small portion of the anterior hypothalamus, the VLPO area, as a possible sleep switch. However, it is now clear that many other neurons scattered through the anterior hypothalamus (for instance, in the median preoptic nucleus, MnPO) and the basal forebrain also play a major role in initiating and maintaining sleep. These neurons tend to fire during sleep and stop firing during wakefulness. MnPO neurons can also track sleep pressure, because their firing is higher during the first sleep cycle when sleep pressure is high and declines in later cycles. When they are active, many release GABA and the peptide galanin and inhibit most wakefulness-promoting areas, including cholinergic, noradrenergic, histaminergic, hypocretinergic, and serotonergic cells. In turn, the latter groups of cells inhibit several sleep-promoting neuronal groups. This reciprocal inhibition provides state stability in that each state reinforces itself as well as inhibits the opposing state. A subset of cortical GABAergic cells is also active during sleep and may track sleep pressure. Finally, sleep-promoting GABAergic cells were recently identified in the medulla, in the parafacial zone. Stimulation of this cell group promotes NREM sleep by inhibiting the wake-promoting neurons of the parabrachial nucleus, located at the junction between the midbrain and pons, which send glutamatergic projections to the basal forebrain.

In terms of NREM sleep neurochemistry, many substances modulate sleep, but no unique sleep factor has been identified. GABA, the major inhibitory neurotransmitter in the CNS, appears to be involved in thalamocortical oscillations and inhibition of waking centers by sleep-active cells. Most hypnotics, including barbiturates, benzodiazepines, and several of the newer nonbenzodiazepine hypnotics, act by enhancing

GABA transmission. However, many GABAergic neurons in the brainstem and basal forebrain are active during waking.

Adenosine has been increasingly recognized as having a role in sleep. Caffeine probably exerts its stimulant effects by blocking adenosine receptors. Adenosine, a degradation product of adenosine triphosphate (ATP), is a byproduct of brain metabolism whose levels increase as a result of increased glutamatergic activity. It accumulates in the basal forebrain and cerebral cortex during prolonged wakefulness and decreases during sleep, suggesting that it may serve to transmit the homeostatic signal for sleep. Adenosine inhibits cholinergic neurons in the pons and basal forebrain, as well as hypocretin neurons and cortical glutamatergic neurons. By contrast, adenosine excites VLPO neurons, and its infusion promotes NREM sleep.

Early studies raised the possibility that serotonin might also be involved in SWS because lesions of serotonergic nerve cells in the dorsal raphe led to insomnia. However, more recent evidence suggests that some of the early results may have been confounded by lesion-induced loss of thermoregulation, which caused increased locomotion (and thus decreased sleep) as a compensatory response to maintain body heat. Serotonergic neurons decrease firing rates in NREM sleep and are completely inhibited during REM. However, they inhibit cholinergic neurons and produce behavioral inhibition, raising the possibility that they help facilitate sleep onset, although they may not be directly involved in inducing or maintaining sleep. It remains controversial as to how much serotonin contributes to sleep versus arousal.

Many other substances, including peptides and neuromodulators, have been attributed with having sleep-promoting properties. These include a variety of hormones, such as melatonin, αMSH, GH-releasing factor, insulin, CCK, and bombesin; cytokines such as interleukin-1, interleukin-6, and TNF; muramyl peptides produced from gut bacteria; and dozens of other substances. Most of these factors have mild hypnotic and/or circadian effects consistent with the usual timing of their release (e.g., melatonin or GH-releasing factor, which is normally released at night) or the physiologic state of the organism (e.g., cytokines produced during infectious illness promote sleep).

REM Sleep

REM sleep is characterized by an activated EEG and increased neuronal activity and CBF. Recent studies using functional imaging techniques have shown that, during REM sleep, there are some brain regions showing increased activation as well as others with decreased activation in comparison to wakefulness. Areas involved in REM sleep generation in the mesopontine tegmentum, thalamus, posterior cortical areas, and limbic areas, particularly the amygdala, are highly activated during REM sleep. In contrast, frontal and parietal cortices are relatively deactivated.

REM sleep is somewhat unique among the three states because a specific brain region—the *pons* and *caudal midbrain*—is both necessary and sufficient to generate the main features of REM sleep and represent the final common pathway for the induction of REM sleep. A series of transection studies have demonstrated that REM sleep is preserved only in the portion of the CNS containing these structures. Bilateral lesions within the pons and caudal midbrain can completely eliminate REM sleep. Although these findings have created a tendency to focus on brainstem mechanisms of REM sleep, more rostral brain regions, including the preoptic area, are also important in the homeostatic regulation of REM sleep and in organizing REM episodes through the night.

As in wakefulness, cholinergic neurons produce EEG activation and a hippocampal theta rhythm during REM sleep. LDT/PPT neurons provide input to the thalamus and cholinergic basal forebrain neurons that in turn activate the limbic system and cortex. Hobson and McCarley proposed the reciprocal interaction hypothesis to explain NREM–REM cycles based on interactions between cholinergic and aminergic neurons in the mesopontine junction. Cholinoceptive and/or cholinergic REM-on cells in the PPT and LDT regions become activated during REM sleep, whereas noradrenergic or serotonergic REM-off cells are inhibitory of the REM-on cells. The aminergic cell groups are most active during waking; they decrease activity somewhat during NREM sleep, and meanwhile, cholinergic activity increases to turn on REM sleep. REM sleep episodes are terminated because REM-on cells are self-inhibitory and provide excitatory input to the REM-off cells. Recent experiments show that GABAergic and glutamatergic neurons in the mesopontine tegmentum are also very important in controlling REM sleep.

The role of cholinergic–monoaminergic interactions in regulating REM sleep is supported by a variety of experimental data. Local infusion of cholinergic agonists such as carbachol in the region of the LDT/PPT of cats or systemic administration of physostigmine, arecoline, pilocarpine, or other cholinergic agonists in humans produces prolonged REM sleep episodes and reduced latency to REM sleep. Cholinergic induction of REM sleep appears to be related primarily to the activation of M_2 muscarinic receptors in the pontine reticular formation. Facilitation of REM sleep also results from the depletion of brainstem monoaminergic activity, for example, in human subjects with an acute depletion of serotonin after being fed a tryptophan-deficient diet. Most antidepressants cause significant reductions in REM sleep, particularly those that increase synaptic availability of norepinephrine and/or serotonin; anticholinergic effects seen in TCAs and MAOIs may also contribute to REM sleep suppression.

Muscle atonia in REM sleep can be eliminated by small lesions in the pontine reticular formation lateral to the LC or in the medial medulla that eliminate inhibitory input from this area to the spinal motoneurons. Spinal motoneurons are inhibited by both GABA and glycine and excited by NA, serotonin, and glutamate. The brainstem circuit responsible for REM sleep atonia includes two major pathways: (1) a glutamatergic projection from the subcoeruleus area to the medullary reticular formation, which then sends GABAergic and glycinergic inputs to motoneurons; (2) a glutamatergic projection from the subcoeruleus area to GABAergic interneurons, which then inhibit motoneurons. Disfacilitation of spinal motoneurons resulting from decreased noradrenergic and serotonergic activity also contributes to the suppression of muscle tone during REM sleep. Several clinical conditions illustrate how atonia may be separated from the state of REM sleep. In narcolepsy, muscle atonia can occur during wakefulness, either as cataplexy, a sudden loss of muscle tone usually brought on by emotional stimuli, or sleep paralysis, in which atonia persists briefly after waking out of REM sleep. In contrast, patients with REM sleep behavior disorder do not develop atonia during REM sleep and act out their dreams, sometimes with such violence that they may injure themselves or their bed partners. REM-sleep–suppressing antidepressants such as TCAs, MAOIs, and SSRIs may induce REM sleep behavior disorder in some individuals.

In animals, ponto-geniculo-occipital (PGO) waves that occur sequentially in the pons, lateral geniculate nucleus of the thalamus, and occipital cortex appear just prior to the initiation of REM periods and in conjunction with phasic activity in REM sleep, including eye movements and muscle twitches. They originate from cholinergic burst cells in the peribrachial region and are suppressed by serotonergic cells in the raphe nuclei. Eye movements during REM sleep are tightly linked to PGO waves and are mediated by inputs to vestibular neurons that in turn activate oculomotor cells. PGO waves are thought to be the internal representation of orienting responses because they can also be observed during a startle response during waking. In humans, PGO-like potentials have been recorded in the pons before and during REM sleep and in the subthalamic nucleus.

In addition to imaging data showing dramatically increased activity in forebrain structures during REM sleep, previous anatomical work has also suggested forebrain involvement in REM sleep regulation. Transection studies that separate the forebrain from pons disrupt NREM–REM cycling caudal to the transection. The amygdala has reciprocal connections with REM-generating brainstem regions, and electrical stimulation or infusion of carbachol into the amygdala can increase REM sleep. Abnormal activation of limbic structures occurs in depression and may contribute to associated changes in REM sleep.

CIRCADIAN AND HOMEOSTATIC REGULATION OF SLEEP

The regulation of sleep—both NREM and REM—involves at least two key components—a circadian and a homeostatic component. The circadian component is responsible for the change in sleep propensity tied to the time of day, with obvious adaptive advantages. The homeostatic component refers to the fact that the longer one stays awake, the greater the propensity to sleep, and it represents the essential aspect of sleep whose function remains mysterious.

The Two-Process Model

Several models of sleep regulation positing a circadian and a homeostatic process have been proposed and validated on the basis of a large amount of data. One of the most influential is the *two-process model* developed by Alexander A. Borbély and colleagues, which predicts sleep propensity based on the interaction between the homeostatic *process S* and the circadian *process C.* Process S builds up across the day in response to the increase in sleep pressure caused by wakefulness and decreases during sleep. The circadian process C for sleep propensity, however, reaches its peak during the latter half of the night. Thus nocturnal sleep onset is primarily driven by process S, whereas process C maintains sleep through the latter part of the night. It is common for humans to have a brief period of arousal in the middle of the night, possibly related to a reduction in overall sleep drive from the fall in process S before process C has reached its maximal values. Similarly, the tendency for afternoon napping may be caused by the increase in process S across the day before process C has reached its lowest values in the late afternoon/early evening. Although napping during the day has become relatively uncommon in industrialized societies, most people experience a period of increased sleepiness in the afternoon. This afternoon “dip” can produce significant daytime sleepiness in individuals who are already somewhat sleep-deprived, are taking sedating medications, and/or have sleep disorders causing excessive daytime sleepiness such as sleep apnea.

Circadian Rhythms

In humans and other mammals, the primary pacemaker for generating circadian rhythms lies in the *SCN* of the hypothalamus. The SCN regulates several neuroendocrine and behavioral parameters, including sleep propensity as measured by process C, to coordinate the state of the organism with the 24-hour light–dark cycle. Circadian sleep regulation is strongly linked to the endogenous temperature rhythm; subjective sleepiness, sleep propensity, and REM sleep propensity are all maximal at the minimum (nadir) of core body temperature, usually in the very early morning, several hours prior to waking up. Sleep tendency is greater during the falling phase of the temperature curve during the night. When core body temperature begins its rising phase in the morning hours, people tend to wake up; arousal levels, performance, and cognitive function are maximal in association with the rise of body temperature across the day.

In animals with lesions of the SCN, sleep is no longer concentrated in one main episode but is dispersed across the entire 24-hour cycle. Still, sleep propensity increases in these animals as a function of previous waking. Thus, although process C and process S normally work together and interact significantly, they can, to a large extent, be separated. Several approaches have been used to investigate specifically the circadian regulation of sleep in humans. The *constant routine* protocol was designed to minimize the influences of factors other than the circadian clock on behavioral state. In this protocol, subjects are kept awake, usually for more than 24 hours, while sitting in bed in a dimly lit room. This technique has been used successfully to demonstrate the persistence of various circadian outputs in the absence of external cues, but the need to enforce sleep deprivation limits the duration of the experiment. With *temporal isolation,* subjects are placed in an environment without time cues for periods of weeks to months. Although circadian rhythms of sleep persist, the "day length" as defined by the period between successive bedtimes is close to 25 hours. These results led to the conclusion that the endogenous period of the human circadian clock is about 25 hours and that the light–dark cycle serves to entrain it to the 24-hour day.

To accurately determine the endogenous period of the circadian pacemaker, it is necessary to disentangle the circadian output from the effects of sleep or prolonged sleeplessness; however, neither of the two protocols discussed above can satisfactorily achieve this. The *forced desynchrony* protocol, originally developed by Nathaniel Kleitman in 1938, consists of having subjects live on a 28-hour "day" while living in temporal isolation. Since it is not possible to entrain the human circadian clock to a period of 28 hours, after a sufficient period of time, sleep periods will have occurred at all phases of the circadian cycle. Thus, circadian and homeostatic influences on human sleep organization can be teased apart. Data from forced desynchrony studies suggest that the endogenous human circadian period is, in fact, close to 24 hours (24.1 to 24.2 hours). Moreover, they show that SWS is primarily regulated by the homeostatic sleep drive, whereas REM sleep is primarily regulated by the circadian clock. However, REM sleep is also homeostatically regulated, as indicated by increased attempts to initiate REM sleep if that stage of sleep is prevented.

More recent studies have demonstrated that circadian timekeeping is more complex; in addition to the central pacemaker in the SCN, there are peripheral clocks in many cells and organs. These peripheral clocks, with their local circadian oscillations, are likely important in the regulation of secretion of multiple hormones, and both misalignment of the central SCN oscillator with the light–dark cycle as well as internal desynchronization between the SCN and peripheral oscillators may be contributory to metabolic and endocrine disorders, as well as to a number of psychiatric disorders.

HOMEOSTATIC REGULATION OF SLEEP AND THE EFFECTS OF SLEEP DEPRIVATION

In humans, as well as in virtually all animal species in which sleep has been carefully studied, sleep deprivation produces sleepiness and increased sleep pressure that soon becomes overwhelming. Sleep deprivation is followed by a “sleep rebound,” that is, a compensatory increase in the duration and/or the intensity of sleep. After sleep deprivation, sleep latency is decreased, and sleep efficiency is increased; that is, sleep is less fragmented. The amount of NREM sleep (especially stage N3 in humans) increases, together with markers of NREM sleep intensity such as SWA, which measures the predominance of slow waves in the cortical EEG. REM sleep amount also increases, but it is unclear whether this is also true for REM sleep “intensity.” Such exquisite homeostatic regulation is one of the most important indications that there must be a distinct physiologic, biochemical, or molecular process that builds up beyond its usual level in the brain if sleep initiation is postponed.

The most prominent effect of total sleep deprivation in humans is cognitive impairment, with striking practical consequences. It is estimated that 35% to 40% of Americans suffer from chronic sleep disorders resulting in difficulty falling asleep or daytime sleepiness. Each year, the costs attributed to sleep problems, including the direct costs of treatment as well as the indirect costs of absenteeism, decreased productivity, accidents, and increased morbidity and mortality, range into the hundreds of billions of dollars in the United States alone. Moreover, the National Highway Traffic Safety Administration reported that in 2017 drowsy driving was responsible for at least 91,000 automobile crashes, resulting in 50,000 injuries and 800 fatalities, according to police reports, but that the true number of yearly fatalities due to drowsy driving is likely about 6,000 according to analyses reported by the Centers for Disease Control. A sleep-deprived person tends to take longer to respond to stimuli, particularly when tasks are monotonous and low in cognitive demands. However, sleep deprivation produces more than just decreased alertness. Tasks requiring higher cognitive functions, such as logical reasoning, encoding, decoding, and parsing complex sentences; complex subtraction tasks; and tasks requiring divergent thinking, such as those involving the ability to focus on a large number of goals simultaneously; are all significantly affected even after one single night of sleep deprivation. Tasks requiring sustained attention, such as those including goal-directed activities, can also be impaired by even a few hours of sleep loss. Thus, sleep loss causes attention deficits, decreases in short-term memory, speech impairments, perseveration, and inflexible thinking. These deficits can explain why sleep-deprived subjects underestimate the severity of their cognitive impairment, often with tragic consequences. Another reason people may underestimate their impairment due to sleep loss because the lack of sleep does not completely eliminate the capacity to perform but rather makes the performance inconsistent and unreliable. Thus, a sleepy driver will either normally respond to an emergency or not at all due to rapid changes in vigilance state and the sudden intrusion of microsleeps during waking. Similarly, subjects may still be able to transiently perform at baseline levels in short tests even after 3 to 4 days of sleep deprivation. However, the same subjects will perform very poorly when engaged in tasks requiring sustained attention. New evidence suggests that not just a few hours of sleep but several days of normal sleep–wake patterns are required to normalize cognitive performance after sleep deprivation.

Cognitive impairment is not only the consequence of total sleep deprivation. Cognitive performance is also affected by sleep restriction (6 hours per night or less) if it continues for several days and by chronic sleep discontinuity such as that occurring in patients with chronic pain, sleep apnea, or other sleep disorders. According to the 2020 “Sleep in America” poll conducted by the National Sleep Foundation in United States residents of at least 18 years of age, 44% reported feeling sleepy 2 to 4 days per week, and 28% felt sleepy 5 to 7 days per week. Women reported feeling sleepy an average of 3.4 days per week, whereas men felt sleepy 2.7 days per week on average. Respondents attributed their sleepiness either to not sleeping well enough (55%) or having insufficient time to sleep (44%). Those who reported feeling sleepy were far more likely to report an impact on their mood, attention, and work performance. A recent analysis of sleep duration obtained from the National Health Interview Survey found that sleep duration decreased from 2013 to 2017, more noticeably among Hispanics and non-Hispanic blacks; overall, 32.9% of adults reported <6 hours of sleep per night. Whether chronic sleep restriction, due to inadequate opportunity or poor quality of sleep, is sufficient to affect objective measures of cognitive performance is not known, but it is certainly concerning given the data from sleep restriction studies that point to impairment and decreased performance. It is becoming apparent that almost all sleep measures, such as the duration of total sleep or SWS and the amount of SWA in the cortical EEG during NREM sleep, show great variability across different subjects but high consistency within the same individual from one night to another. The extent of the cognitive impairment after sleep deprivation also varies significantly across subjects. Converging evidence shows that individual differences in the cognitive impairment caused by sleep deprivation are stable over time and traitlike, suggesting a genetic basis.

Brain and peripheral tissues respond differently to sleep loss. Like in sleep-deprived animals, the peripheral metabolic rate is increased in sleep-deprived human subjects and in normal sleepers on nights of poor sleep relative to baseline nights; this increased rate is also present generally in people with insomnia relative to normal sleepers. In both animals and humans, glucose metabolism is higher in many brain regions in waking than that in NREM sleep. After one day of sleep deprivation, selective brain areas can still be activated metabolically when the subject is engaged in specific tasks. However, the global cerebral metabolic rate does not increase and actually decreases relative to normal waking values in areas such as the thalamus and the midbrain. Thus, while peripheral metabolic rate is persistently increased during sleep deprivation, brain metabolic rate is not. This may be an indication that the brain cannot sustain high-energy metabolism for too long.

In addition to causing cognitive impairment, sleep deprivation in humans may also affect various physiologic systems with impacts on overall health. It has been suggested that sleep loss can affect host defense systems; for example, sleep-deprived rats show increased rates of bacteremia. Sleep deprivation has also been shown to lead to decreased glucose tolerance, increased sympathetic nervous system (SNS) activation, and elevated cortisol levels, suggesting that it may contribute to disorders such as diabetes, hypertension, and obesity. Patients with insomnia have increased rates of health problems, including cardiac disease, further suggesting a possible causal relationship between reduced sleep amounts and health outcomes. Some studies in both humans and animals have linked long-term disturbances in sleep with reduced longevity. Others have suggested that, in fact, short sleep and insomnia are associated with little risk distinct from comorbidities. Whether there is a U-shaped association between sleep duration and mortality risk (i.e., both short sleep and long sleep are associated with increased mortality) remains unclear, and meta-analyses have come to conflicting conclusions. Recent meta-analyses, published in 2019 and 2020, concluded that long sleep duration was associated with increased risk for all-cause mortality in men and women, but short duration of sleep was significantly associated with all-cause mortality only in men and in only one of the analyses. However, it is important to note that insomnia is not equivalent to sleep deprivation and does not necessarily indicate reduced sleep amount.

The most extreme case of prolonged sleep loss in humans is observed in patients affected by fatal familial insomnia (FFI), a prion disorder characterized by near-complete loss of sleep and associated with various neurologic symptoms and spongiform degeneration in selected brain regions. The disease is invariably fatal after a few months to 2 to 3 years. In FFI with a short clinical course (death in <1 year), insomnia is almost complete from the onset, and spongiform degeneration is mainly restricted to the thalamus and inferior olive but does not extend to the cerebral cortex. In FFI cases with a longer clinical course (death in 2 to 3 years), insomnia develops more gradually, and spongiform degeneration is found in most cortical regions. While several clinical aspects of FFI can be attributed to diffuse spongiform degeneration, sleep loss per se likely plays an important role in determining the evolution of the disease, especially in short-course FFI. Indeed, the severity of the clinical course correlates with the severity of insomnia rather than with the accumulation of prion protein.

In humans, sleep deprivation is never enforced for more than 3 to 4 days (the record is 11 days but in only one subject), making it impossible to determine whether there are other more severe effects of very prolonged sleep loss. Prolonged sleep deprivation is possible in animals, where several techniques have been used, from forced locomotion to pharmacologic stimulation with amphetamines. Irrespective of the method used, an important and consistent finding is that even drastic manipulations such as brain electrical stimulation are unable to enforce complete and uninterrupted wakefulness after the first 24 hours. Even prolonged sleep deprivation attempts cannot enforce complete and uninterrupted wakefulness for more than one day. Evidently, sleep pressure overcomes whatever method is used to maintain wakefulness in animals as well as in humans. This pressure can manifest itself in different ways. The most obvious way is “microsleeps,” short lapses lasting a few seconds during which the waking EEG clearly switches to a sleep pattern. Microsleeps can cause obvious impairment in performance, for instance, losing control of the car while driving. Less obvious changes consist in an overall “slowing of the EEG,” with an increase in low range frequencies (delta/theta activity). This change is detectable by quantitative analysis of the EEG but not by simple visual observation, although it may still affect performance in cognitive tasks. Finally, an even more subtle change is the recently identified phenomenon of “local sleep” during wake. This phenomenon, first described in rodents using intracortical recordings, refers to the fact that during prolonged wakefulness, rats can move around and show a normal wake activated EEG, while some cortical neurons may go “offline” and stop firing, showing down states of membrane hyperpolarization and neuronal silence similar to those normally present only during NREM sleep. Because these down states are local and short-lasting, they can only be detected using invasive intracortical recordings. However, in humans, a useful EEG correlate of this phenomenon is the occurrence of local delta/theta waves (2 to 6 Hz), which can be identified using high-density EEG (hd-EEG, 256 channels). When human subjects stay awake for 24 hours and engage in intensive learning of cognitive tasks, hd-EEG can detect a local increase in delta/theta waves mainly in the cortical areas mostly involved in these tasks. During recovery sleep after sleep deprivation, the same areas will show a local increase in SWA.

The most comprehensive series of sleep deprivation studies in animals have been performed using the *disk-over-water apparatus* (DOW), which can prevent sleep for days and even weeks. This method uses minimal stimulation to enforce chronic sleep deprivation in the sleep-deprived rat, while it simultaneously applies the same stimulation to the control rat but without severely limiting its sleep. Sleep deprivation in rats with the DOW produces dramatic physiologic changes that invariably culminate in death after 2 to 3 weeks of sleep loss. Within the first 1 to 2 days, rats develop a syndrome characterized by increased food intake, energy expenditure, and heart rate, followed by a decrease in body weight and a decline in body and brain temperature. The sleep deprivation syndrome and its lethal consequences have also been observed after selective REM sleep deprivation, although the pathology associated with the loss of sleep takes longer to appear, the survival time is longer (4 to 5 weeks rather than 2 to 3 weeks), and body and brain temperature are not significantly decreased. Despite a long series of studies, the DOW sleep deprivation syndrome has not been fully explained, nor is it clear why the animals die of sleep deprivation. It is evident, however, that the syndrome produced by the DOW is not unique, and other methods of chronic sleep deprivation used in different animal species have produced similar effects. Also, in dogs, rabbits, and to a lesser extent, cats, sleep deprivation for several days causes an increase in food intake and heart rate, weight loss, and eventually death. Even fruit flies, if prevented from sleeping for several days, die of sleep deprivation. Again, these findings suggest that there must be at least one potentially vital function for sleep and that this function is conserved across different animal species.

DREAMING

For a long time, sleep has been regarded as the annihilation of consciousness, save for the occasional dream that is remembered when one wakes up. Just as the old notion that the brain is silent during sleep has been disproved, so has the myth of cognitive death during sleep. Most subjects, if awakened and asked to report whatever may be going through their minds, report some kind of mental activity most of the time. Often such mental activity constitutes a dream, which can be defined as a complex, temporally unfolding, hallucinatory episode that occurs during sleep.

The Relationship between Dreaming and Normal Waking Consciousness

Dream hallucinations are typically more vivid than waking images. Images are predominantly visual, although all modalities can be represented, and impossible motor activities, such as flying, may occur. Hearing speech or conversation is extremely frequent. Dreaming is generally delusional—events and characters in the dream are taken for real—and confabulatory—a dream involves making up a story. There is often disorientation, that is, uncertainty about space (where one is in the dream), time (when the dream is taking place in personal history), and confusion about the gender, age, and identity of dream characters. However, dreams appear to run in real time, as there are good correlations between the subjective duration of a dream, the length of the dream report, and the time it would take to reenact a dream. Emotions are prominent in many dreams, especially fear and anxiety. While the self is almost always at the center of the dream, there is some reduction of reflective consciousness. Dreams have been described as single-minded, that is, missing the alternation of primary and reflective consciousness that characterize wakefulness. Self-monitoring, directed thinking, and volition are reduced, leading to an inability to analyze situations, to question assumptions, and to make appropriate decisions. The ability to form new memories is drastically impaired (dream amnesia), in part due to low noradrenergic activity. Not surprisingly, psychiatrists have pointed out similarities between some aspects of dreaming cognition and certain psychotic symptoms, as well as delirium.

Despite these psychopathologic traits, perhaps the most important fact about dreaming consciousness is how remarkably similar it can be to waking consciousness. The sleeping brain, disconnected from the “real” world, is capable of generating an imagined world, a “virtual reality,” which is fairly similar to the real one and is indeed experienced as equally real. Considering dreams from the perspective of the waking state, people are often intrigued by the bizarre quality of some dreams, especially morning dreams. Bizarre dreams may be more memorable; however, if collected systematically throughout the night, dreams are much more mundane—they are a faithful replica of waking life. As a rule, one can only dream what one can imagine, although dreams can be much more vivid, probably because of the reduced competition from external signals.

Conditions that affect people’s brains during wakefulness extend to their dreams. For instance, if blind people can still construct visual images, then they have visual dreams; otherwise not. If a stroke abolishes the ability to perceive color during waking, then visual images, as well as dreams, become achromatopsic. Like the ability to form mental images, dreaming seems to depend most strongly on the integrity of cortical areas higher than the primary visual cortex. Similarly, somatosensory/motor and audioverbal/motor imagery are normal in the dreams of hemiplegic and aphasic patients. Another example of the close connection between waking and dreaming cognitive competence comes from the study of children’s dreams. Dream reports are rare and extremely short until 4 to 5 years of age—in agreement with the limited ability of children below that age to imagine and narrate complex stories with high emotional content. Finally, there is a good correlation between waking and dreaming mood, imaginativeness, and personality. Studies of dream content in psychiatric populations, especially in depression and PTSD, have generally been unremarkable. However, some intriguing findings have been reported, for example, that changes in dream content may anticipate changes in the course of the disorder.

Dreaming and Sleep Stages

It was initially suggested that full-fledged dreams could be elicited almost exclusively during REM sleep. However, later studies have shown conclusively that dreamlike mental activity can also be elicited from NREM sleep, especially at sleep onset and during the last part of the night—the times when NREM sleep is less deep. Up to 15% of subjects, however, never recall any mental content when awakened from NREM sleep. Moreover, typical REM dream reports can easily be distinguished from NREM ones by being, on average, much longer (up to seven times). When dream reports are longer, this suggests a longer duration of the actual dream, but they may also be due to an increased density or bizarreness of the experienced scenes. Whether REM dreams are not only longer but also qualitatively different—more bizarre, hallucinatory, delusional, narrative, and emotional than NREM dreaming—remains controversial. So is the suggestion that NREM dreams may be “covert” REM dreams due to an intrusion of REM characteristics in NREM sleep. There is no doubt, however, that typical dreams, as well as nightmares, can be experienced in certain phases of NREM sleep. Dreamlike mental activity having a hallucinatory–delusional character can also occur during quiet wakefulness, especially under conditions of reduced sensory input (daydreaming). Conversely, some waking-like mental activity—nonhallucinatory and nondelusional—is occasionally reported at sleep onset. A good example of a mixed state is lucid dreaming, where dreamers are aware that they are dreaming and, to some extent, can control the course of their dreams.

The initial equating of the cognitive state of dreaming with the physiologic state of REM sleep was encouraged by the remarkable similarity between the EEG of REM sleep with that of conscious waking and by their equivalent differences from the NREM sleep EEG. It seemed natural to infer that the activated (low-voltage, high-frequency) EEG of waking and REM sleep would support vivid conscious experience, while the deactivated (high-voltage, low-frequency) EEG of NREM sleep would not. While frequent dream reports at sleep onset could still be reconciled with the mixed EEG of stage N1 sleep, the presence of dreamlike experiences, although shorter, during NREM stages characterized by EEG slow waves seemed paradoxical. This paradox may be resolved by considering the time course of neural excitability during the slow oscillation of NREM sleep. It is now known that, during the depolarized phase of the slow oscillation, neural activity is as intense as in waking or REM sleep, and neurons are highly excitable. However, during sleep, the depolarized phase is interrupted by a hyperpolarized phase during which neural activity ceases throughout the cerebral cortex, and neurons are much less excitable. The inevitable occurrence of hyperpolarization (down state) after a period of depolarization (up state) occurs exclusively during NREM sleep and reflects a peculiar bistability of cortical circuits. This bistability is especially strong during SWS early in the night—the depolarized phase may not last longer than a second or so before it is followed by the hyperpolarized phase—and becomes less marked during stage N2 sleep late in the night when longer periods of depolarization are possible. Thus, if dreamlike experiences occur during the depolarized phase, then they will be short and poor during SWS early in the night and become progressively longer and richer later in the night. Altogether, it would seem that the likelihood of conscious experience during sleep is related to moment-to-moment cortical activation, where activation should be understood as the readiness of cortical neurons to respond in a rich and differentiated manner to incoming signals.

Neural Correlates of Dreaming Consciousness

Recently, lesion and imaging studies have provided new insights into brain correlates of the characteristic differences between dreaming and waking consciousness. The most obvious difference is that dreaming consciousness is only marginally influenced by external stimuli, very few of which are incorporated into the dream narrative. During NREM sleep, partial functional disconnection is mediated by thalamic inhibition. During REM sleep, external stimuli more easily pass the thalamic gate and reach primary cortical areas, but they do not seem to influence higher cortical areas, as if the brain were not paying attention to them. The hallucinatory character of dreams is obviously facilitated by such sensory disconnection. Visual hallucinosis of dreams is indeed associated with increased activity in higher visual areas, while the primary visual cortex is less active, as indicated by PET studies. Whether the intense visual experience and scene changes characteristic of dreams are triggered primarily by phasic signals from the brainstem, areas involved in visual imagery, or both is an unresolved issue.

Another relevant difference between dreaming and waking consciousness concerns the ability to reflect on oneself and one's experience, especially the ability to judge the verisimilitude of dreaming experience. Imaging studies indicate that the dorsolateral prefrontal cortex (DLPFC), a brain region implicated in volitional control and self-monitoring, is less active in sleep compared to waking. Reduced activation of the DLPFC may also contribute to the disorientation and reduction of directed thinking and working memory observed in dreams and contribute to dreaming amnesia. Recent episodic memories are conspicuously absent in dreams, and memory for dreams is strikingly labile unless one wakes up and rehearses the dream.

Dreams are also characterized by a high degree of emotional involvement, especially fear and anxiety. Correspondingly, imaging studies have revealed a marked activation of limbic and paralimbic structures such as the amygdala and anterior cingulate, insular, and medial orbitofrontal cortices during REM sleep. Altogether, cognitive activity during sleep provides a powerful indication of the extent to which fluctuating levels of several neuromodulators, whose dysfunction is implicated in several psychiatric disorders, can affect mental function in healthy subjects. Indeed, Hobson has suggested that most aspects of dreaming cognition can be explained by considering the level of three processes—brain activation, input source (external or internal), and neuromodulation—across the sleep–wakefulness continuum.

Recently, hd-EEG has also been used to characterize the neural correlates of conscious experience during sleep. These experiments have used a no-task, within-state paradigm in which sleeping subjects recorded with hd-EEG reported retrospectively whether they had been conscious or not. It was found that the neural correlates of consciousness in both NREM and REM sleep map to two localized hotspots: (1) a parieto-occipital hotspot characterized by reduced low-frequency EEG activity and (2) a frontal hotspot characterized by increased high-frequency EEG activity. It was also found that the frontal hotspot is involved in highly thought-like experiences, while the parieto-occipital hotspot is implicated in highly perceptual experiences, suggesting that the neural correlate of consciousness during sleep may be restricted to a single hotspot and change in location depending on the nature of the subject's dream. These studies have also shown that specific contents of a subject's dreams—such as faces and spatial setting—are associated with increased high-frequency EEG activity in specific cortical areas that closely match those engaged during waking perception of the same contents. These results provide strong evidence that dream reports reflect conscious experience that occurs during sleep and are not simply confabulations that are produced upon awakening.

Theories of Dreaming

In what he considered his most important work—*The Interpretation of Dreams*—Freud suggested that dreams provide disguised wish fulfillment and, if properly interpreted, would provide essential clues to the most profound determinants of psychic life, "the royal road to the unconscious." The systematic investigation of dreams has not provided much support for this notion.

A radically different suggestion was made by Hobson and McCarley based on their neurophysiologic studies of REM–sleep-generating mechanisms in the brainstem. According to their activation–synthesis hypothesis, dreams were the forebrain's attempt to make sense of the random activation of thalamocortical networks by the upper brainstem, like the music produced by unmusical fingers wandering over the keys of a piano. Another suggestion, based on a comprehensive evaluation of dream reports, was made by Foulkes and others. According to this view, dreams reveal not so much the psychodynamic unconscious but instead the cognitive development, competence, and style of the dreamer, just as in waking cognition. This view carefully eschews the one-to-one conflation of dreams with REM sleep.

Along these lines, Solms has recently reviewed neuropsychologic evidence and shown that the ability to dream depends not on the upper brainstem but on forebrain regions. In more than 100 cases of cessation of dreaming, the responsible lesion was either the parieto-temporo-occipital junction (uni- or bilaterally) or white matter near the orbitomesial prefrontal cortex (bilaterally). Despite the cessation of dreaming, REM sleep was almost always preserved. The parieto-temporo-occipital junction is important for mental imagery, spatial cognition (on the right side), and symbolic cognition (on the left side), all central features of dreaming. The white matter underlying the ventromedial prefrontal cortex that is necessary for dreaming is the same brain area that is targeted by modified prefrontal leucotomy. Its lesion reduces the positive symptoms of schizophrenia but produces adynamia. Chemical activation of these circuits by stimulants such as amphetamines, as well as by 3,4-dihydroxy-L-phenylalanine (L-DOPA), can produce hallucinations and delusions, suggesting that dreams may be facilitated by the activation of the mesolimbic and mesocortical dopaminergic systems.

THE FUNCTIONS OF SLEEP

The function of sleep is still a mystery. Sleep may have evolved from the circadian rest–activity cycle and may thus represent a default state. However, it is likely that sleep serves some more fundamental function. Otherwise, why should animals engage in prolonged periods of quiescence with increased arousal thresholds during which they cannot monitor potential dangers in the environment? Sleep seems to be universal. All animal species studied so far sleep, from invertebrates such as fruit flies and bees to birds and mammals, although mammals and birds generally display more elaborate sleep cycles that include an alternation of NREM and REM sleep. Ridding or even reducing the body’s need for sleep does not seem to be easy. Some marine mammals who may need continuous vigilance while flying or swimming, such as certain dolphins and porpoises, have developed alternating unihemispheric sleep rather than eliminating sleep altogether. Sleep deprivation in all species studied leads to an increase in sleep pressure, manifested as sleepiness, which rapidly becomes irresistible. Sleep pressure may lead to microsleeps or piecemeal sleep that result in cognitive impairment. Sleep deprivation is followed by longer and more intense sleep, suggesting a regulated need for sleep. If sleep is prevented for several weeks, then the consequences are fatal. Rats deprived of sleep for more than 2 to 3 weeks invariably die. Humans affected by a rare prion disorder called FFI also die, although it is not certain whether death is due to sleep deprivation per se.

Many hypotheses have been formulated about the functions of sleep. A good hypothesis should account for the following facts: (1) Sleep involves partial disconnection from the environment, which is potentially dangerous; sleep must therefore provide something not provided by quiet waking. (2) Sleep is accompanied by intense neural activity in most brain regions; an explanation is needed for the brain's activity in the absence of overt behavior. (3) Sleep seems to constitute a universal requirement, but its amount varies greatly across different species; conservation of sleep across species must be accounted for while accommodating key differences in sleep patterns. (4) In most species, sleep is prevalent early in life; sleep should perform important functions both during development and in the adult. (5) Sleep is often made up of both NREM and REM stages; it must be explained whether the sequence of the two stages is important and whether both stages serve a similar function. These various functions and characteristics of sleep have proved difficult to accommodate within one hypothesis; no single hypothesis has been proposed that can account for all these facts.

Sleep and Brain Restitution

Most hypotheses currently under investigation are concerned with a role of sleep in restoring some metabolic function or in serving neural plasticity. It is likely that sleep may preserve energy by enforcing body rest in animals with high metabolic rates. Indeed, animals and humans eat more during sleep deprivation. However, in humans, the metabolic efficiency of sleep is only marginally better than that of quiet waking. Most bodily organs can obtain rest through quiet wakefulness, except for the brain; thus, sleep may be especially important for the brain. Some molecular pathway or chemical in the brain may be depleted during waking and restored during sleep. For example, it has been suggested that sleep may favor the replenishment of glycogen in glial stores, although recent evidence shows that this may only be true in a few brain regions. Alternatively, sleep could counteract synaptic fatigue by favoring the replenishment of calcium in presynaptic stores, the replenishment of glutamate vesicles, the resting of mitochondria, recycling of membranes, or transfer of proteins along axons and dendrites. While recent studies have revealed that molecular changes do occur between sleep and waking and after sleep deprivation, the significance of such changes is still unknown. If sleep restores something lost during waking, it is not known what it is.

Sleep, Learning, and Memory

A connection between sleep and memory was noted a long time ago. For example, after struggling to learn a new piece of music for much of the day, one will often play it better after a night of sleep. Recently, the importance of sleep and even naps occurring after certain types of declarative and nondeclarative learning has been documented in well-controlled experiments. Moreover, recent evidence suggests that sleep deprivation impairs not only the ability to consolidate previously learned tasks but also the capacity to acquire new information. Sleep could indeed offer a favorable context for certain aspects of learning and memory. The sensory disconnection associated with sleep reduces interference between ongoing activities and the consolidation of previously acquired memories. Moreover, sleep has been suggested to permit the repeated reactivation, in an off-line mode, of the neural circuits originally activated during a memorable experience. Studies using multielectrode recordings in animals and PET in humans have shown that brain areas or cells activated during waking are preferentially reactivated during subsequent sleep, although this reactivation also occurs during periods of quiet wakefulness following the learning experience. A further advantage of sleep could be that the relevant neural circuits can be reactivated in a spaced and interleaved fashion. This would favor the integration of new with old memories and would avoid catastrophic interference. The intense, high-frequency bursts of spontaneous neural activity that occur during sleep may be particularly important for both triggering molecular mechanisms of synaptic consolidation and enlarging the network of associations.

Many unknowns remain, however, as to whether sleep may favor the consolidation of newly established memories or the maintenance of older ones. The molecular correlates of such processes are still not known. Molecular markers of memory acquisition are turned off during sleep, which may be advantageous given that the intense neural activity of sleep occurs while the animal is disconnected from the environment. Much of the early literature connecting sleep and memory was concerned with REM sleep. However, prolonged inhibition of REM sleep in humans through MAOIs does not seem to disrupt memory, nor does the complete disappearance of REM sleep after certain brainstem lesions. Perhaps the most convincing evidence concerns the role of sleep in developmental plasticity. Recent experiments have shown that REM sleep deprivation in newborn rats and NREM sleep deprivation in kittens influence the activity-dependent development of visual system circuits. If sleep affects synaptic maturation and plasticity, then one can expect that sleep disturbances in early life may affect psychopathologic development.

SLEEP AND PSYCHIATRY

Historically, psychiatrists have been interested in sleep and dreaming and their relationship with mental illness. From Sigmund Freud and Carl Jung in the latter part of the 19th century, the interpretation of dreams became an important tool in psychoanalytical psychotherapy. After the discovery of REM sleep in 1953, psychiatrists began to explore whether specific sleep abnormalities could be correlated with psychiatric disorders. One of the first questions addressed was whether schizophrenia might represent a disorder of REM sleep, given the similarities between dreaming and psychosis. Although psychosis could not be explained as a REM sleep disorder, both REM sleep and schizophrenia were later found to be associated with decreased activity in the DLPFC. Narcolepsy, however, turned out to be caused by the abnormal intrusion of REM sleep phenomena into wakefulness.

Research has focused more on sleep in depression than any other psychiatric aspects of sleep by far. Even before the discovery of REM sleep, people with depression were known to have disrupted sleep. Sleep EEG recordings in the 1950s and 1960s also showed that they had a relative loss of SWS in comparison to age-matched control subjects as well as specific changes in REM sleep, including reduced latency to REM sleep, greater proportion of REM sleep during the first third of the night, increased frequency of REMs during REM sleep (i.e., increased REM density), and increased percentage of sleep time spent in REM sleep. Although not every patient with depression shows changes in REM sleep and not every patient with short REM latency has depression, REM sleep abnormalities are one of the more robust biologic markers for depression. Both reduced REM sleep latency and loss of SWS appear to be trait markers for depression in that they persist even during clinical remission and are found at higher rates in first-degree family members of people with depression. Although the mechanisms for sleep changes in depression are not fully understood, there is a convergence of data suggesting that sleep and mood are regulated by common systems. For example, the cholinergic–monoaminergic imbalance hypothesis of depression is consistent with the observed increase in REM sleep and reduction in SWS that would be caused by increased cholinergic activity. Individuals with depression show a heightened sensitivity to REM sleep induction by cholinergic drugs in comparison to nondepressed control subjects as well.

Sleep deprivation studies are even more suggestive of a functional relationship between sleep and mood. Total deprivation of a single night of sleep or even partial deprivation of sleep in the latter half of the night can have an immediate antidepressant response in many moderately to severely depressed individuals. Sleep loss can induce or perpetuate mania in bipolar patients, who may go for periods of several days with little or no sleep. In contrast, even a short bout of sleep can reverse the antidepressant effect of sleep deprivation, and prolonged sleep can induce depression in some individuals. Functional imaging studies have shown that sleep deprivation, like antidepressant drug therapy, normalizes the increased metabolic activity seen in the anterior cingulate gyrus in individuals with depression. Selective REM sleep deprivation has also been shown to have antidepressant effects, and it has been suggested that REM-sleep–suppressing antidepressants may act in part through their effects on sleep. REM sleep suppression, however, is not a requirement for antidepressant efficacy because some agents such as bupropion and nefazodone appear to cause no significant reduction of REM sleep. More recent evidence suggests that both sleep deprivation and antidepressant medications may act by similar mechanisms, namely, by selectively upregulating genes involved in neural plasticity and synaptic potentiation.

Although therapeutic sleep deprivation may be beneficial for some cases of depression, the preponderance of evidence suggests that disturbed sleep is a risk factor for the development of psychiatric illness. People with insomnia as well as hypersomnia have higher rates of psychiatric disorders, particularly mood and anxiety disorders; in primary care settings, insomnia is more strongly associated with depression than with any other medical disorder. Individuals who have insomnia or even difficulty sleeping during times of stress are significantly more likely to develop depression in the future. A causal link between sleep and depression has not yet been established, but the fact that both sleep deprivation and depression are becoming more prevalent in society suggests that the relationship between the two must be clarified. In patients with psychiatric disorders as well as in those without known psychiatric illness, insomnia is an independent risk factor for suicidal behaviors.

From an epidemiologic perspective, although sleep disturbance is most strongly associated with psychiatric disorders, there are a number of other significant health-related correlates. People with insomnia have higher rates of other medical illnesses, use more health care services, have higher rates of absenteeism, accidents, and disability, and have poorer outcomes with some medical disorders, including cardiac disease. Whether treatment of sleep problems can prevent the development of any of these comorbidities remains to be seen.

Sleep abnormalities are also seen in virtually all other psychiatric disorders, particularly disturbances in sleep continuity, including prolonged latency to sleep onset, diminished efficiency of sleep, and decreased amounts of total sleep. During NREM sleep, individuals with insomnia show increased fast beta activity (16 to 32 Hz) globally, as well as increased alpha activity (8 to 12 Hz) more centrally, suggestive of hyperarousal during sleep. REM sleep abnormalities similar to depression have been described in some studies of patients with schizophrenia, alcoholism, eating disorders, and borderline personality disorder. Patients with panic disorder may have panic attacks arising from sleep, usually at the transition into SWS, which further emphasizes the biologic etiology of this disorder. Recent studies using hd-EEG have demonstrated a loss of spindle activity and decrements in SWA in schizophrenic subjects, which may be related to a dysfunction in the thalamocortical network and associated with clinical features of sleep disruption and cognitive impairments. Furthermore, spindle deficits have been reported both in schizophrenic patients early in the course of illness as well as in unaffected first-degree relatives, suggesting that they may represent biomarkers and/or risk factors for schizophrenia.

An appreciation of sleep neurobiology is essential for clinicians, because psychiatric patients often have sleep problems associated with their illnesses and most psychiatric drugs have significant effects on sleep; these range from sedation caused by many antipsychotics, benzodiazepines, TCAs, and antiparkinsonian agents to sleep disturbance caused by MAOIs, SSRIs, and psychostimulants. Antidepressants may precipitate some sleep disorders, such as periodic movements in sleep and REM sleep behavior disorder, and abrupt withdrawal of REM-suppressing antidepressants including tricyclics, MAOIs, and SSRIs can lead to REM sleep rebound, characterized by increased duration and intensity of REM sleep and sleep disruption. Sleep apnea may be exacerbated by drugs that produce muscle relaxation (e.g., benzodiazepines or barbiturates) or weight gain (e.g., antipsychotics, antidepressants, and mood stabilizers).

Sleep is a revealing, though not yet transparent, window into the functional state of the human brain. The study of sleep has shed light on many aspects of consciousness and the workings of the human brain. Clarifying the functions and mechanisms of sleep will undoubtedly provide us with a greater understanding of psychiatric disorders and their treatments.

FURTHER READINGS

Anafi RC, Kayser MS, Raizen DM. Exploring phylogeny to find the function of sleep. *Nat Rev Neurosci*. 2019;20(2):109–116.

Benca RM, Obermeyer WH, Thisted RA, Gillin JC. Sleep and psychiatric disorders: a meta-analysis. *Arch Gen Psychiatry*. 1992;49:651–668.

Borbely AA, Achermann P. Sleep homeostasis and models of sleep regulation. *J Biol Rhythms*. 1999;14:557–568.

Brown RE, Basheer R, McKenna JT, Strecker RE, McCarley RW. Control of sleep and wakefulness. *Physiol Rev*. 2012;92:1087–1187.

Cirelli C. Cellular consequences of sleep deprivation in the brain. *Sleep Med Rev*. 2006;10:307–321.

Cirelli C, Tononi G. Is sleep essential? *PLoS Biol*. 2008;6(8):e216.

Cirelli C, Tononi G. Cortical development, EEG rhythms, and the sleep/wake cycle. *Biol Psychiatry*. 2015;77(12):1071–1078.

Colten HR, Altevogt BM, eds. *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. The National Academies Press; 2006.

Consensus Conference Panel; Watson NF, Badr MS, et al. Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society on the Recommended Amount of Sleep for a Healthy Adult: Methodology and Discussion. *J Clin Sleep Med.* 2015;11(8):931–952.

Dang-Vu TT, Schabus M, Desseilles M, Sterpenich V, Bonjean M, Maquet P. Functional neuroimaging insights into the physiology of human sleep. *Sleep*. 2010:33(12):1589–1603.

Ferrarelli F. Endophenotypes and biological markers of schizophrenia: from biological signs of illness to novel treatment targets. *Curr Pharm Des*. 2013;19(36):6462–6479.

Foster RG, Peirson SN, Wulff K, Winnebock E, Vetter C, Roenneberg T. Sleep and circadian rhythm disruption in social jetlag and mental illness. *Prog Mol Biol Transl Sci*. 2013;119:325–346.

Gamble KL, Berry R, Frank SJ, Young ME. Circadian clock control of endocrine factors. *Nat Rev Endocrinol*. 2014;10:466–475.

García-Perdomo HA, Zapata-Copete J, Rojas-Cerón CA. Sleep duration and risk of all-cause mortality: a systematic review and meta-analysis. *Epidemiol Psychiatr Sci*. 2018;28:578–588.

Goel N, Basner M, Rao H, Dinges DF. Circadian rhythms, sleep deprivation, and human performance. *Prog Mol Biol Transl Sci*. 2013;119:155–190.

He M, Deng X, Zhu Y, Huan L, Niu W. The relationship between sleep duration and all-cause mortality in the older people: an updatd and dose-response meta-analysis. *BMC Public Health*. 2020;20:1179.

Hirshkowitz M, Whiton K, Albert SM, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary, *Sleep Health*. 2015;1:40–43.

Iber C, Ancoli-Israel S, Chesson AL, Quan SF. *The AASM Scoring Manual for the Scoring of Sleep and Associated Events*. American Academy of Sleep Medicine; 2007.

Klinzing JG, Niethard N, Born J. Mechanisms of systems memory consolidation during sleep. *Nat Neurosci*. 2019;22(10):1598–1610.

Kryger MH, Roth T, Dement WC. *Principles and Practices of Sleep Medicine*. WB Saunders; 2017.

Kurina LM, McClintock MK, Chen JH, Waite LJ, Thisted RA, Lauderdale DS. Sleep duration and all-cause mortality: a critical review of measurement and associations. *Ann Epidemiol*. 2013;23:361–370.

Kurth S, Olini N, Huber R, LeBourgeois M. Sleep and early cortical development. *Curr Sleep Med Rep*. 2015;1(1):64–73.

Neske GT. The slow oscillation in cortical and thalamic networks: mechanisms and functions. *Front Neural Circuits*. 2016;9:88.

Palagini L, Baglioni C, Ciapparelli A, Gemignani A, Riemann D. REM sleep dysregulation in depression: state of the art. *Sleep Med Rev*. 2013;17(5):377–390.

Rasch B, Born J. About sleep's role in memory. *Physiol Rev*. 2013;93(2):681–766.

Scammell TE, Arrigoni E, Lipton JO. Neural circuitry of wakefulness and sleep. *Neuron.* 2017;93(4):747–765.

Sheehan CM, Frochen SE, Walsemann KM, Ailshire JA. Are U.S. adults reporting less sleep?: findings from sleep duration trends in the National Health Interview Survey, 2004–2017. *Sleep*. 2019;42(2):zsy221.

Siclari F, Baird B, Perogamvros L, et al. The neural correlates of dreaming. *Nat Neurosci*. 2017;20(6):872–878.

Silber MH, Ancoli-Israel S, Bonnet MH, et al. The visual scoring of sleep in adults. *J Clin Sleep Med*. 2007;3:121–131.

Tononi G, Cirelli C. Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration. *Neuron*. 2014;81:12–34.

Tononi G, Koch C. The neural correlates of consciousness: an update. *Ann N Y Acad Sci*. 2008;1124:239–261.

1.16 Pain Systems: Interface with Affective and Motivational Mechanisms

John Lotfi, M.D., J.D., and Jon-Kar Zubieta, M.D., Ph.D.

Pain is a cardinal symptom of many somatic diseases. Acutely, it serves to warn of danger and shapes behavior to avoid the environment eliciting the stimulus. It is also a subjective experience that incorporates sensory (e.g., intensity, location, quality), affective, and cognitive dimensions. Its chronification is associated with adaptive and maladaptive neurobiologic changes serving to perpetuate the experience, leading to an increase in its aversive and motivational significance. In individuals for whom acute pain transitions into a persistent condition, its experience no longer serves a protective role and is frequently associated with the development of depression and anxiety comorbid conditions in approximately 50% of cases, with bidirectional exacerbation.

While the modulation of acute pain can be achieved pharmacologically by the introduction of agents that reduce nociceptive input (e.g., opioid compounds, pharmacologic conduction blockade and modulation, monoamine reuptake inhibitors), the relief of persistent pain is not adequately achieved in many conditions, particularly those without obvious or identifiable pathophysiology (e.g., fibromyalgia, irritable bowel syndrome, nonneuropathic back pain). The negative effects of chronic opioid use have received much attention because of the development of tolerance to these compounds, withdrawal symptoms, and their potential for misuse. In this context, opioid use disorders have grown into a public health crisis in the United States. Currently, annual drug-related deaths exceed the number of casualties linked to either automobile or handgun deaths. In 2015, according to the Centers for Disease Control and Prevention (CDC), more than 33,000 people died accidentally from an opioid overdose. Three years later, in 2018, the death toll from an opioid overdose had risen even higher, to 48,802, and the combination of the opioid epidemic with the COVID-19 pandemic has further strained fragile health care infrastructure, particularly for more complex, comorbid conditions.

The main nonopioid strategies to reduce pain severity and comorbidity include potentiation of monoaminergic neurotransmission through reuptake inhibition or similar mechanisms. However, agents that reduce inflammatory responses and neuronal excitability are increasingly studied to reverse neuroplastic changes that take place during the transition from acute to chronic pain.

PAIN SYSTEMS

Pain perception is typically initiated by the activation of peripheral nociceptors, sensory end organs in the skin, muscle, joints, and viscera that selectively respond to noxious or potentially damaging stimuli. An important property of nociceptors is their sensitization as a consequence of tissue injury and inflammation, reducing their threshold of response and increasing the magnitude of their responses to noxious stimulation, this being a potential new target for pharmacologic interventions. Nociceptor endings in tissue are unencapsulated ("free") and are commonly associated with either (1) medium diameter myelinated Aδ axons, which transmit acute, well-localized "first pain" signals, or (2) small-diameter, unmyelinated slowly conducting C fibers that convey poorly localized, "second pain," "burning," or slow pain. Aδ and C nociceptors are further subdivided into multimodal or more specialized types, responding to mechanical, chemical, or heat stimuli with varying degrees of sensitivity. Larger, rapidly conducting myelinated Aβ fibers respond to innocuous mechanical stimulation such as touch.

The cell bodies of nociceptors for the body are located in the dorsal root ganglia (DRG) of the spinal cord, whereas the trigeminal ganglion houses the cell bodies of nociceptors for the face, which transmit information through the trigeminal nucleus caudalis. Axonal projections from the DRG neurons transmit information through excitatory glutamatergic and peptidergic neurotransmission to second-order neurons of the dorsal horn of the spinal cord. Second-order neuron axons then cross contralaterally and form ascending pathways that synapse in the brainstem (spinoreticulothalamic tract) and thalamic nuclei (spinothalamic tract), thought to respectively subserve less localized and more sensory-discriminative (e.g., intensity and localization) elements of pain.

Spinoreticulothalamic projections to the rostroventral medulla (RVM) and periaqueductal gray (PAG) engage descending pain inhibitory pathways in feedback loops that regulate dorsal horn neuronal excitability while thalamic nuclei project to cortical structures. Other spinoreticulothalamic projections reach the parabrachial nucleus of the pons, which is directly connected to the amygdala. The amygdala is a region implicated in the representation and modulation of aversive, affectively negative elements of the pain experience and which provides further input to the insular cortex.

The regulatory functions of the PAG and RVM have received much attention since they are engaged as a modulatory mechanism of the nociceptive signal, which is heavily influenced by neurotransmitters such as the enkephalinergic endogenous opioid, serotonin, and norepinephrine systems. These are the targets of commonly used pain control agents, such as monoamine reuptake inhibitor antidepressants and opioid medications. Pain signals to the PAG activate excitatory and peptidergic transmission, which projects to the RVM. In the RVM, “on” cells become activated in response to noxious painful stimuli and are involved in the formation of protective “nocifensive” behaviors, such as withdrawal reflexes. “Off” cells, on the other hand, respond acutely to endogenous opioids, GABA, and serotonin, to induce analgesic responses through descending connections to the dorsal horn of the spinal cord. Norepinephrine, through alpha 1 and alpha 2 receptors, respectively, can have both pro- and antinociceptive effects by modulation of “on” and “off” cells of the RVM (Fig. 1.16–1).

Thalamic input reaching the somatosensory cortex is implicated in sensory-discriminatory functions, while projections to regions such as the anterior cingulate and insular cortex are thought to integrate sensory and emotional aspects of the experience. Often referred to as “lateral” (sensory-discriminatory) and “medial” (affective-cognitive) projections, they form a large, distributed brain network “pain matrix” that reflects and processes the complex, multifaceted experience of pain. During acute pain, the primary and secondary somatosensory cortices, insular cortex, anterior cingulate, and prefrontal cortices become consistently activated. The basal ganglia, amygdala, and hippocampus are also involved in these processes.

It should be noted that many of these “medial projection” integrative areas overlap with those implicated in the representation of other nonpainful, salient, or stressful stimuli and are thought to be dysregulated when pain becomes chronic, reflecting alterations in reward and motivational mechanisms observed in persistent pain conditions. Notably among these, the anterior insular, anterior cingulate, and prefrontal cortices, as well as the nucleus accumbens and amygdala, are thought to be implicated in the representation and regulation of emotional, motivational, and evaluative aspects of the pain experience as discussed in further detail below. Conversely, the ventrolateral thalamus, somatosensory cortex, dorsal posterior insula, and some cellular groups of the anterior cingulate and anterior insular cortex are thought to be more pain-specific, responding to variations in pain intensity and being differentially activated by pain stimuli as opposed to nonpain-specific sensory stimuli or other components of the experience of pain. In addition to the modulatory role of the PAG and RVM on nociceptive suppression, telencephalic regions that include the dorsolateral prefrontal cortex, anterior cingulate and insular cortices, thalamus, nucleus accumbens, and amygdala are implicated in regulating the pain experience across sensory and affective (e.g., unpleasantness) components through the local activation of endogenous opioid and other neurotransmitter systems such as the noradrenergic, GABA, endocannabinoid, and glycinergic, as well as through descending projections to the PAG and RVM (Fig. 1.16–2).

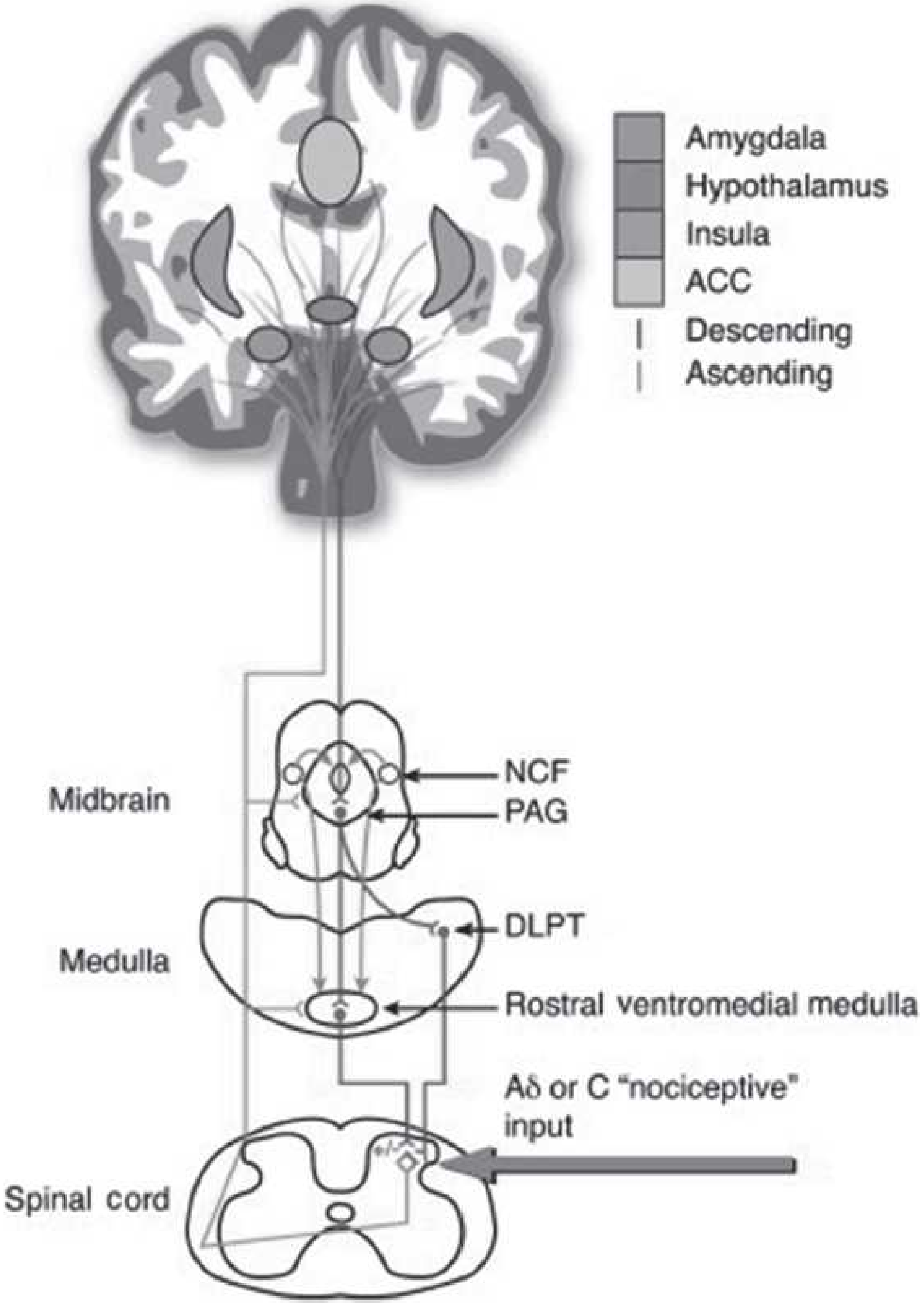
Amygdala
Hypothalamus
Insula
ACC
Descending
Ascending
Midbrain
NCF
PAG
Medulla
DLPT
Rostral ventromedial medulla
Aδ or C "nociceptive" input
+/−
Spinal cord

FIGURE 1.16–1. Descending pain modulatory system. NCF, nucleus cuneiformis; PAG, periaqueductal gray; DLPT, dorsolateral pontine tegmentum; ACC, anterior cingulate cortex; +/− indicates both pro- and antinociceptive influences, respectively. (Reprinted from Tracey I, Mantyh PW.

The cerebral signature for pain perception and its modulation. *Neuron.* 2007;55(3):377–391. Copyright © 2007, with permission from Elsevier.)

PAIN AND MOTIVATIONAL MECHANISMS

From an evolutionary perspective, pain is a strong motivational signal with negative affective connotations, while pain relief can be conceptualized as a form of relief and reward. Therefore, pain represents a potent factor directing behavior. The basal ganglia and associated limbic cortex are thought to integrate internal and external signals to encode the value of the sensory signals received. They are engaged during anticipation of pain or expectations of pain relief and, together with prefrontal cortical structures, are implicated in the integration of sensory information but also the modulation of pain through endogenous opioid modulatory systems.

During acute pain, brain functional activity of the saliency/reward/motivation network, which includes the orbitofrontal cortex, nucleus accumbens, and ventral tegmental area, is increased and associated with the capacity to express endogenous opioid analgesia in humans. Interindividual variations in the capacity of the nucleus accumbens to respond to reward value have been further linked to the capacity to develop placebo-induced analgesic responses, suggesting that reward and saliency evaluation mechanisms are also engaged to induce appropriate analgesic responses when required to maintain homeostasis.

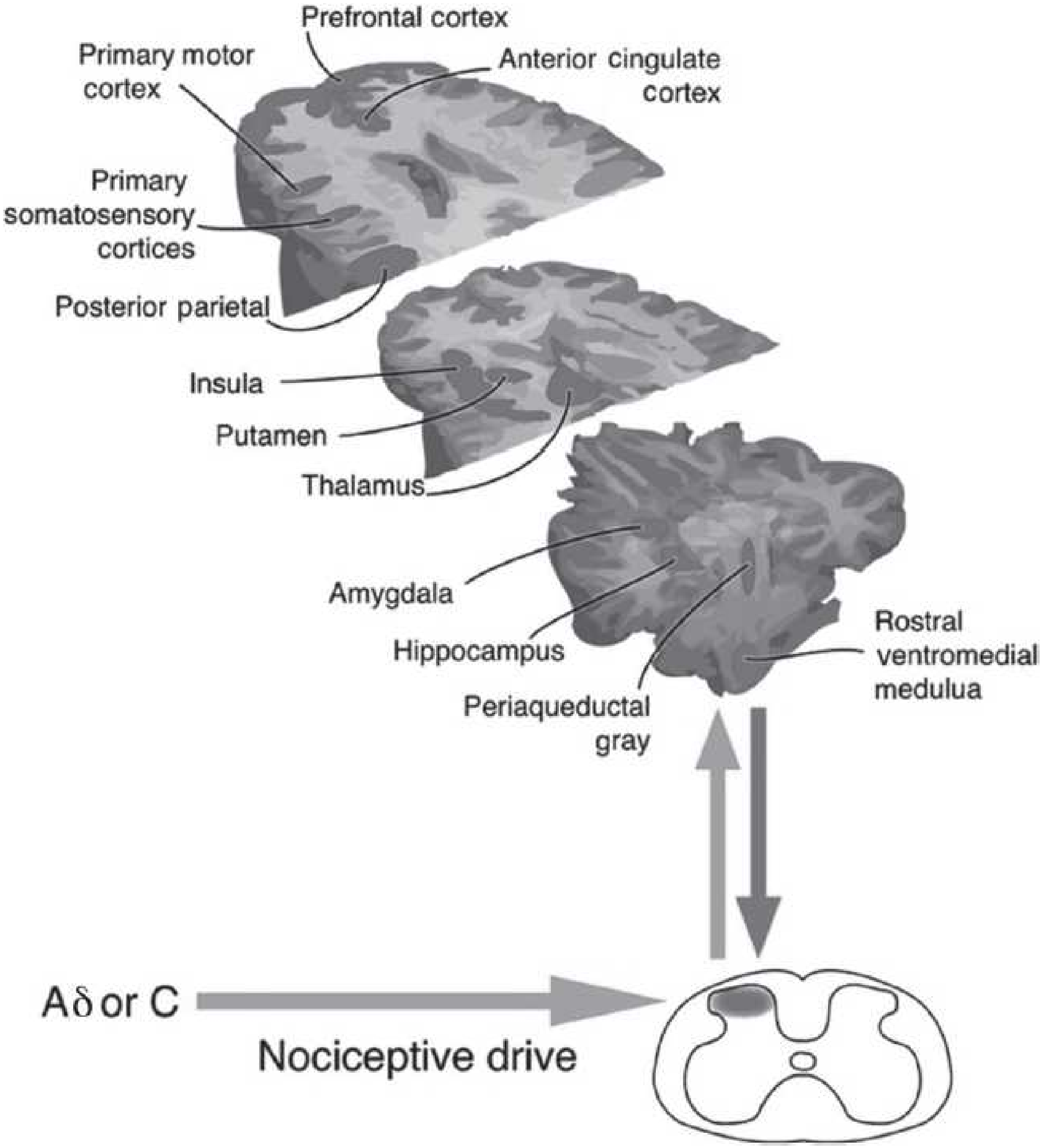
Prefrontal cortex
Primary motor cortex
Anterior cingulate cortex
Primary somatosensory cortices
Posterior parietal
Insula
Putamen
Thalamus
Amygdala
Hippocampus
Periaqueductal gray
Rostral ventromedial medulua
Aδ or C
Nociceptive drive

FIGURE 1.16–2. Neuroanatomy of pain processing. Brain regions that activate during a painful experience, with bilateral representation (contralateral to the noxious stimulus in orange). (Reprinted from Tracey I, Mantyh PW. The cerebral signature for pain perception and its modulation. *Neuron.* 2007;55(3):377–391. Copyright © 2007, with permission from Elsevier.)

The nucleus accumbens receives pain information through regions involved in pain and emotional processing, such as the prefrontal cortex, anterior cingulate, and the amygdala, and direct projections from spinal cord neurons to the nucleus accumbens have been described. During experimental pain in humans, dopamine neurotransmission is activated in the basal ganglia. In dorsal regions receiving input from the somatosensory cortex, dopamine is released in a manner proportional to the intensity and sensory elements of the pain signal, while in the nucleus accumbens, it was related to negative affective responses in humans. This effect appears to be localized in the nucleus accumbens core based on animal model data, while in the shell area of this region, it is released after pain stimulus termination. This is consistent with the concept that pain relief represents a form of positive reinforcement upon its termination as shown using neuroimaging data in humans diagnosed with persistent pain. The transition from injury and acute pain to persistent low back pain in humans has been further associated with alterations in the functional connectivity of the basal ganglia to prefrontal cortical structures, suggesting that persistent pain-induced alterations in motivational systems form a critical part in the perpetuation of the pain experience in humans.

In patients diagnosed with persistent pain conditions, such as fibromyalgia and chronic back pain, dopamine receptors and dopaminergic neurotransmission in the nucleus accumbens are downregulated compared to those of healthy individuals. These reductions appear adaptive in nature, hampering the salience of the chronic noxious stimulus but also enhancing the functional capacity of downstream antinociceptive endogenous opioid mechanisms. However, dysregulation of nucleus accumbens function in the context of chronic pain has also been linked to reductions in motivation, such as the capacity to perform effortful tasks to gain rewards in animal models. Therefore, nucleus accumbens function and dopaminergic neurotransmission in this area are involved in the signaling of the negative value and salience of noxious input and have been implicated in the transition from acute to persistent pain. Dopaminergic mechanisms also seem to become adaptively downregulated in the context of persistent pain, however, with potentially negative behavioral consequences such as the diminished capacity to engage motivational mechanisms and the effortful pursuit of rewards, core features of anhedonia and depression.

PAIN AND AFFECTIVE REGULATION AND DYSREGULATION

Persistent pain becomes a form of physical and emotional stressor that induces negative mood states, part of maladaptive responses to chronic noxious stimuli. Clinically significant symptoms of depression and anxiety appear in approximately half of the patients diagnosed with persistent pain conditions and form part of the comorbid pathology associated with chronic pain. Psychopharmacologic therapies for comorbid depression and anxiety that may also modulate descending neuropathic pain pathways are discussed below.

Brain regions such as the anterior cingulate, which receive nociceptive signals through the spinothalamic pathway, are implicated in the encoding of pain unpleasantness. This unpleasantness can be modulated by cognitive processes such as hypnotic suggestions, whereby endogenous opioid mechanisms are engaged in the context of suggestions of analgesia during placebo administration. Opioid receptors mediating endogenous opioid analgesia are found downregulated in this and other regions, such as the thalamus, nucleus accumbens, and amygdala, in chronic pain states (such as inflammatory, central, and neuropathic pain as well as in fibromyalgia) with accompanying enhancements in negative affective and pain unpleasantness ratings. Similarly, in chronic back pain patients, deficits in endogenous opioid neurotransmission have been encountered in the amygdala, a region implicated in antinociception as well as responses and regulation of affectively valenced stimuli, and have been associated with both clinical pain ratings and greater negative affect. These findings parallel those encountered in patients diagnosed with major depression and are an example of alterations in biologic processes that overlap between pathologies.

The anterior insular cortex, one of the regions that mediate responses to emotional stimuli and is implicated in the pathophysiology of depression, has also been shown to have an important role in the encoding of interoceptive information in the context of pain. It has been linked to the formation of attentional biases towards physical physiologic cues, including pain and implicated in fear learning and anxiety responses to painful stimuli. Clinical interventions such as meditation-based therapies geared towards the reduction of interoceptive sensitivity and anxiety in patients diagnosed with persistent pain have been shown to regulate the anterior cingulate and anterior insular cortex in a manner proportional to clinical benefit.

Psychopharmacologic therapies for comorbid depression and anxiety that may also modulate descending neuropathic pain pathways in chronic neuropathic pain include serotonin and norepinephrine reuptake inhibitors (SNRIs), tricyclic antidepressants (TCAs), gabapentinoids, and carbamazepine. SNRIs have antidepressant and anxiolytic effects stemming from their inhibition of serotonin and norepinephrine reuptake which increases the relative availability of these neurotransmitters. This increased availability of serotonin and norepinephrine may be the mechanism behind an SNRI's ability to inhibit descending pain pathways, although opioid mechanisms may also be indirectly implicated. TCAs similarly inhibit serotonin and norepinephrine reuptake for their antidepressant and anxiolytic effects; however, TCAs also act as antagonists at muscarinic acetylcholine, histamine, and adrenergic receptors, which result in a greater side-effect profile than SNRIs. Gabapentinoids bind with high affinity to the alpha-2-delta subunit of voltage-gated calcium channels which inhibits excitatory neurotransmitter release, although the exact mechanism of its anxiolytic and descending pain pathway modulation remains unclear. Carbamazepine acts as a use-dependent voltage-sensitive sodium channel blocker which inhibits the release of glutamate, acting to modulate pain pathways as well as manage acute treatment-refractory depression. Oxcarbazepine, which is structurally related to carbamazepine, may have similar effects but bears further study. Ketamine, an NMDA receptor agonist, also appears to be effective in reducing persistent pain in recent metanalysis, in a manner parallel to that described for major depression.

As noted above, substantial overlap exists among brain regions and neurobiologic mechanisms implicated in the processing of emotionally salient stimuli and those found to represent and regulate pain signaling. This is consistent with the concept of persistent pain as an allostatic process that promotes negative affective responses and enhances the individual propensity toward developing depression and anxiety symptomatology. Different psychopharmacologic therapies can target comorbid depression and anxiety while also modulating pain pathways.

COGNITIVE PROCESSING AND PAIN

The perception of pain is sensitive to mental processes such as attentional biases towards the noxious stimulus. Heightened attention and expectations of threat toward painful stimuli have been shown to induce catastrophic thinking and higher pain sensitivity. Conversely, a distraction from pain and reframing its meaning have been associated with reductions in pain-associated responses of the somatosensory cortex, thalamus, insula, and anterior cingulate cortex, with accompanying changes in sensory, cognitive, and affective appraisals of the painful stimulus. Therefore, both enhancements and reductions in pain reports can be achieved by the influence of higher-order cognitive processes effects that at least in part take place through the top-down engagement of the PAG and descending pain modulatory circuits impacting spinal nociceptive transmission. Executive cortical regions such as the dorsolateral and ventrolateral prefrontal cortex are thought to provide input to the anterior cingulate, insular, and thalamic circuitry, which in turn project to the PAG, a region that potently influences the intensity of ascending signals from the spinal cord.

Another example of how cognitive processes can affect pain processing is the so-called placebo analgesic effect. Besides the perspective that placebo effects confound controlled trials of potentially therapeutic agents, the information so far acquired points to neurobiologic systems that, when activated by positive expectations and maintained through reward learning, are capable of inducing physiologic changes that lead to the experience of analgesia and improvements in an emotional state. This line of research has described neural networks and neurotransmitter systems implicated in the formation of placebo analgesic responses. These include the dorsolateral prefrontal cortex, which is implicated in the processing of environmental cues and expectations, as well as the ventrolateral prefrontal cortex, anterior cingulate, insular cortex, nucleus accumbens, amygdala, and PAG. Salience assessments mediated by dopamine neurotransmission in the nucleus accumbens appear to engage pain regulatory networks across these regions, inducing the adaptive activation of pain and affect regulatory neurotransmitters and neural networks (or maladaptive in the case of nocebo responses, adverse effects in response to therapeutic cues). Endogenous opioids are centrally involved in these processes as well as the endocannabinoid and cholecystokinin systems. The formation of biologic placebo effects, therefore, appears to implicate the engagement of resiliency mechanisms with interindividual variations partially determined by genetic factors and uncovered by the cognitive-emotional integration of the expectations created by the therapeutic environment. Of note, similar mechanisms appear to take place in the formation of placebo antidepressant effects with partial regional overlap between those engaged during the development of placebo-induced analgesia.

FUTURE DIRECTIONS

The majority of the work carried out in the study of pain perception and regulation has been concerned with the transmission of nociceptive input and its regulation at the level of peripheral, spinal, and brainstem structures, providing exquisite detail on these processes. Increasing information is being acquired as to the contribution of inflammatory mechanisms in both glia and neurons as well as neuroplastic changes in excitatory pathways in the transition from acute to persistent pain and its further chronification. New insights have also been acquired as to telencephalic mechanisms that are engaged during acute, sustained, and persistent pain, particularly in humans, and which undoubtedly contribute to pain chronicity and comorbidity with depression, anxiety, and alterations in reward and motivation. Emerging literature demonstrates substantial overlaps between neural systems engaged during pain and those that respond to other stressors, the dysregulation of which would explain the complex presentation of patients diagnosed with persistent pain conditions. Future advancements in the area will require particular attention to those interfaces to exert actual change in pain chronicity and associated neuropsychiatric symptomatology, including the frequently comorbid substance use disorders, and to provide new insights into novel treatment targets for these conditions.

From a public health perspective, and given the complex relationships among pain, mental health, substance abuse, and disparities in economic and social status, the fragmentation of health care into silos of care is highly unproductive. Comorbidities are known to deeply impact health outcomes, with 80% of Medicare spending being directed toward patients with four or more chronic conditions. However, disease or organ-based practice continues to be the prevailing approach to care, reducing the capacity of health practitioners to personalize care for those with co-occurring conditions.

FURTHER READINGS

Apkarian AV, Bushnell MC, Treede R-D, Zubieta J-K. Human brain mechanisms of pain perception and regulation in health and disease. *Eur J Pain*. 2005;9:463–484.

Baliki MN, Geha PY, Fields HL, Apkarian AV. Predicting value of pain and analgesia: nucleus accumbens response to noxious stimuli changes in the presence of chronic pain. *Neuron*. 2010;66:149–160.

Baliki MN, Petre B, Torbey S, et al. Corticostriatal functional connectivity predicts transition to chronic back pain. *Nat Neurosci*. 2012;15:1117–1119.

Basbaum AI, Bautista DM, Scherrer G, Julius D. Cellular and molecular mechanisms of pain. *Cell*. 2009;139:267–284.

Bushnell MC, Ceko M, Low LA. Cognitive and emotional control of pain and its disruption in chronic pain. *Nat Rev Neurosci*. 2013;14:502–511.

Fields H. State-dependent opioid control of pain. *Nat Rev Neurosci*. 2004;5:565–575.

Gold MS, Gebhart GF. Nociceptor sensitization in pain pathogenesis. *Nat Med*. 2010;16:1248–1257.

Harris RE, Clauw DJ, Scott DJ, McLean SA, Gracely RH, Zubieta JK. Decreased central mu-opioid receptor availability in fibromyalgia. *J Neurosci*. 2007;27:10000–10006.

Hsu DT, Sanford BJ, Meyers KK, et al. It still hurts: altered endogenous opioid activity in the brain during social rejection and acceptance in major depressive disorder. *Mol Psychiatry*. 2015;20:193–200.

Lutz A, McFarlin DR, Perlman DM, Salomons TV, Davidson RJ. Altered anterior insula activation during anticipation and experience of painful stimuli in expert meditators. *Neuroimage*. 2013;64:538–546.

Martikainen IK, Nuechterlein EB, Peciña M, et al. Chronic back pain is associated with alterations in dopamine neurotransmission in the ventral striatum. *J Neurosci*. 2015;35:9957–9965.

Martikainen IK, Peciña M, Love TM, et al. Alterations in endogenous opioid functional measures in chronic back pain. *J Neurosci*. 2013;33:14729–14737.

Navratilova E, Porreca F. Reward and motivation in pain and pain relief. *Nat Neurosci*. 2014;17:1304–1312.

Peciña M, Bohnert AS, Sikora M, et al. Association between placebo-activated neural systems and antidepressant responses: neurochemistry of placebo effects in major depression. *JAMA Psychiatry*. 2015;72:1087–1094.

Peciña M, Zubieta JK. Molecular mechanisms of placebo responses in humans. *Mol Psychiatry*. 2015;20:416–423.

Pertovaara A. The noradrenergic pain regulation system: a potential target for pain therapy. *Eur J Pharmacol*. 2013;716:2–7.

Phan KL, Wager T, Taylor SF, Liberzon I. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*. 2002;16:331–348.

Rainville P, Carrier B, Hofbauer RK, Bushnell MC, Duncan GH. Dissociation of sensory and affective dimensions of pain using hypnotic modulation. *Pain*. 1999;82:159–171.

Ru-Rong J, Chamessian A, Yu-Qiu Zhang Y. Pain regulation by non-neuronal cells and inflammation. *Science*. 2016;354(6312):572–577.

Schwartz N, Temkin P, Jurado S, et al. Chronic pain. Decreased motivation during chronic pain requires long-term depression in the nucleus accumbens. *Science*. 2014;345:535–542.

Scott DJ, Heitzeg MM, Koeppe RA, Stohler CS, Zubieta JK. Variations in the human pain stress experience mediated by ventral and dorsal basal ganglia dopamine activity. *J Neurosci*. 2006;26:10789–10795.

Scott DJ, Stohler CS, Egnatuk CM, Wang H, Koeppe RA, Zubieta JK. Individual differences in reward responding explain placebo-induced expectations and effects. *Neuron*. 2007;55:325–336.

Scott DJ, Stohler CS, Egnatuk CM, Wang H, Koeppe RA, Zubieta J-K. Placebo and nocebo effects are defined by opposite opioid and dopaminergic responses. *Arch Gen Psychiatry*. 2008;65:220–231.

Segerdahl AR, Mezue M, Okell TW, Farrar JT, Tracey I. The dorsal posterior insula subserves a fundamental role in human pain. *Nat Neurosci*. 2015;18:499–500.

Simons LE, Elman I, Borsook D. Psychological processing in chronic pain: a neural systems approach. *Neurosci Biobehav Rev*. 2014;39:61–78.

Skolnick P, Volkow ND. Re-energizing the development of pain therapeutics in light of the opioid epidemic. *Neuron*. 2016;92(2):294–297.

Smith TH, Sim-Selley LJ, Selley DE. Cannabinoid CB1 receptor-interacting proteins: novel targets for central nervous system drug discovery? *Br J Pharmacol*. 2010;160:454–466.

Tracey I, Mantyh PW. The cerebral signature for pain perception and its modulation. *Neuron*. 2007;55:377–391.

Vandenberg RJ, Ryan RM, Carland JE, Imlach WL, Christie MJ. Glycine transport inhibitors for the treatment of pain. *Trends Pharmacol Sci*. 2014;35:423–430.

Wager TD, Atlas LY, Lindquist MA, Roy M, Woo CW, Kross E. An fMRI-based neurologic signature of physical pain. *N Engl J Med*. 2013;368:1388–1397.

Wanigasekera V, Lee MC, Rogers R, et al. Baseline reward circuitry activity and trait reward responsiveness predict expression of opioid analgesia in healthy subjects. *Proc Nat Acad Sci USA*. 2012;109:17705–17710.

Wiech K, Ploner M, Tracey I. Neurocognitive aspects of pain perception. *Trends Cogn Sci*. 2008;12:306–313.

Wood PB, Schweinhardt P, Jaeger E, et al. Fibromyalgia patients show an abnormal dopamine response to pain. *Eur J Neurosci*. 2007;25:3576–3582.

Zubieta JK, Smith YR, Bueller JA, et al. mu-opioid receptor-mediated antinociceptive responses differ in men and women. *J Neurosci*. 2002;22:5100–5107.

Zubieta JK, Smith YR, Bueller J M, et al. Regional mu opioid receptor regulation of sensory and affective dimensions of pain. *Science*. 2001;293:311–315.

1.17 Basic Science of Appetite

NORI GEARY, PH.D., AND TIMOTHY H. MORAN, PH.D.

Appetite comprises the complexly determined and highly individuated behavioral and subjective phenomena related to eating. This chapter, (1) summarizes progress in understanding the core physiology of appetite, including eating behavior, the sensory signals that control it, the brain mechanisms that integrate these signals, and some important modulators of appetite, and (2) provides examples of basic research contributions to the understanding of disordered eating. Although much has been learned, both normal and disordered eating remains poorly understood. Thus, further research could facilitate the development of more effective treatment strategies than are presently available for disordered eating, whether presenting in psychiatric eating disorders or simple obesity.

Appetite Is a Network Function

Eating is controlled by a widely distributed neural network in the brain that integrates oral, nasal, and pharyngeal food stimuli, gastrointestinal (GI) signals, metabolic signals (including signals related to body adiposity), numerous other internal stimuli (e.g., circadian signals, stress responses, immune responses, reproductive-axis function), and external stimuli with learned meanings. Furthermore, the network displays structural and functional plasticity throughout the life span. This perspective emphasizes the multiple factors affecting eating, their varied effects on information processing in the brain, and their complex synergistic and antagonistic interactions. It contrasts with the traditional view that only one or a few signals, for example, blood-glucose levels, acting on a few circumscribed hypothalamic sites produce a unitary control of appetite.

The view that appetite is a property of a neural network rather than a single brain region has several important implications. (1) There may be no brain locus where all neural signals controlling appetite converge. (2) It is unlikely that a neural basis for a unitary motivational process for appetite will be found. (3) Controls of eating that are elaborated in different parts of the network may not always operate in a coordinated manner. For example, eating might simultaneously be inhibited by some controls (e.g., homeostatic signals related to metabolic fuel availability) and stimulated by others (e.g., flavor hedonics). For this reason, patterns of within meal-eating rates or patterns of spontaneous meal sizes and frequencies may change in the absence of changes in the total amount eaten. (4) The existence of such partially autonomous controls may contribute to the modest effects of pharmacologic therapies for disordered eating that are based on the manipulation of single signaling molecules.

The Meal as the Unit of Analysis

The meal is the most appropriate unit of analysis for both functional and physiologic analyses of appetite. The meal is the functional unit of eating. (1) The timing, size, and content of meals provide a complete description of the organism's response to all the influences affecting appetite described above. (2) Meal onset, the rate of eating during meals, meal termination, meal size, and intermeal intervals provide concrete metrics for the study of appetite. For example, accounts of hunger cannot succeed if they do not explain the timing of meal onset. (3) Disordered meals are the behavioral phenomenology of disordered eating. Abnormally large meals are the defining behavioral change in bulimia nervosa and in binge-eating disorder, and abnormally small meals are the defining behavioral change in anorexia nervosa.

The meal is also the biologic unit of eating. Humans and most other vertebrates organize eating into discrete bouts or meals, separated by intervals of noneating, with species-specific variations related to circadian, seasonal, and other influences. Meal-ending satiation is an active process: meals are followed by a period of sedation or sleep, which is clearest in animals and infants. These characteristics are undoubtedly products of both natural selection and cultural evolution, shaped by such influences as the digestive consequences of cooked rather than raw food and the logistics of food acquisition and preparation.

Behavioral Neuroscience of Eating

The behavioral neuroscience of eating follows three general approaches. (1) *Input approaches* are aimed at understanding the environmental and experiential contingencies and the exteroceptive stimuli and interoceptive sensory signals that influence eating. Exteroceptive stimuli include stimuli whose meaning is learned as well as a few innately meaningful stimuli such as the hedonic significance of sweet taste. Important interoceptive stimuli include circulating levels of hormones and metabolites as well as neural signaling from peripheral organs to the brain. (2) *Network approaches* focus on the central neural networks that integrate inputs and produce the motivational processes that drive eating. This is now done with increasingly sophisticated molecular-genetic approaches in animals, and functional brain imaging approaches in humans. An important advantage of imaging is the ability to associate subjective and brain responses. These approaches are most meaningful when meals are studied, rather than only summary measures such as calories ingested per day or only subjective measures. (3) *Motor approaches* include efforts to understand the neural control of the movements of eating, often called the microstructure of eating. Food is licked, sucked, bitten, or masticated prior to being moved by lingual and palatal movements to the oropharynx and swallowed. Analysis of these movements has the potential to track their neural control through the lower motor neurons of cranial nerves V, VII, IX, X, and XII into (i) the local circuits in the motor nuclei of these nerves, (ii) the central pattern generators in the hindbrain that project onto these nuclei, (iii) upstream integratory interneuronal networks, and (iv) ultimately to sensory and brain-autonomous inputs, such as memories, circadian rhythms, and so forth.

SIGNALS FOR MEAL INITIATION

Hunger Processes and Learning

The patterns of neural activity leading to food-seeking and meal initiation are conceptualized as hunger motivation. Three distinct motivational processes are thought to be involved: (1) a negative-reinforcement process, (2) a positive-reinforcement process related to flavor hedonics, and (3) a positive-reinforcement process related to postingestive consequences of eating (Fig. 1.17–1). Each of these suffices to stimulate eating, and eating leads to negative feedbacks that reduce hunger. However, only negative-reinforcement hunger involves an aversive state that one wishes to escape. In contrast, the expectation of satiation or of a positive hedonic state is usually pleasant. Most people rarely experience the aversive state of strong deprivation-induced hunger (among those who do are the poor, strict dieters, and people with anorexia nervosa). Although the upper GI tract is relatively empty after overnight fasts, people generally do not report intense hunger in the morning and breakfast is usually the smallest meal of the day. Thus, normal adult hunger probably represents a lifetime of positive-reinforcement learning with innate “homeostatic” mechanisms not usually playing important roles.

Studies in animals have suggested that partially independent brain processes mediate a motivational aspect of hedonics, known as incentive motivation or wanting, and a subjective aspect, known as liking. This distinction may be linked to the separable influences of flavor hedonics and flavor-nutrient learning on food selection discussed below.

Circadian Rhythms

Circadian rhythms importantly affect eating and may explain low morning hunger (Fig. 1.17–2). Such circadian rhythms presumably originate in the circadian pacemaker in the suprachiasmatic nucleus of the hypothalamus (SCN). This area has extensive neural connections with the arcuate nucleus of the hypothalamus (ARC), the dorsomedial nucleus of the hypothalamus (DMN), the lateral hypothalamic area (LHA), and other areas through which it could modulate the neural and endocrine signals controlling eating, as reviewed below.

Metabolic Signals

Stimuli resulting from decreased glucose or fatty-acid utilization can elicit hunger and meal initiation. These effects are clear in extreme conditions, such as during extended nutrient depletion, abrupt shifts in the macronutrient composition of the diet, and biochemical hypoglycemia. It seems unlikely that they operate under usual physiologic conditions.

Small transient declines in plasma glucose sometimes occur prior to spontaneous meals in rats and humans and may be a normally operating hunger stimulus. These are too small (∼10%) to influence cellular glucose availability. Rather, the temporal dynamics of the decline may signal hunger.

Endocrine Signals

Two GI hormones are candidate endocrine signals for hunger, ghrelin and motilin. Ghrelin has received more attention to date. Ghrelin is a peptide hormone synthesized and released primarily by enteroendocrine cells in the stomach and proximal small intestine. It is secreted in an inactive form and metabolized to the biologically active, acylated form, octanoyl- or acyl-ghrelin, by the enzyme ghrelin O-acyltransferase. Plasma levels are high in the morning, decrease after meals, and increase during intermeal intervals. The mechanism that stimulates ghrelin secretion is unknown. The gradual reduction in GI contents in the hours after meals apparently releases a tonic stimulatory control of ghrelin secretion from inhibition.

Inhibition is reinstated when nutrients reach the small intestine during and after meals. Carbohydrates and proteins inhibit ghrelin secretion more than isoenergetic lipid loads.

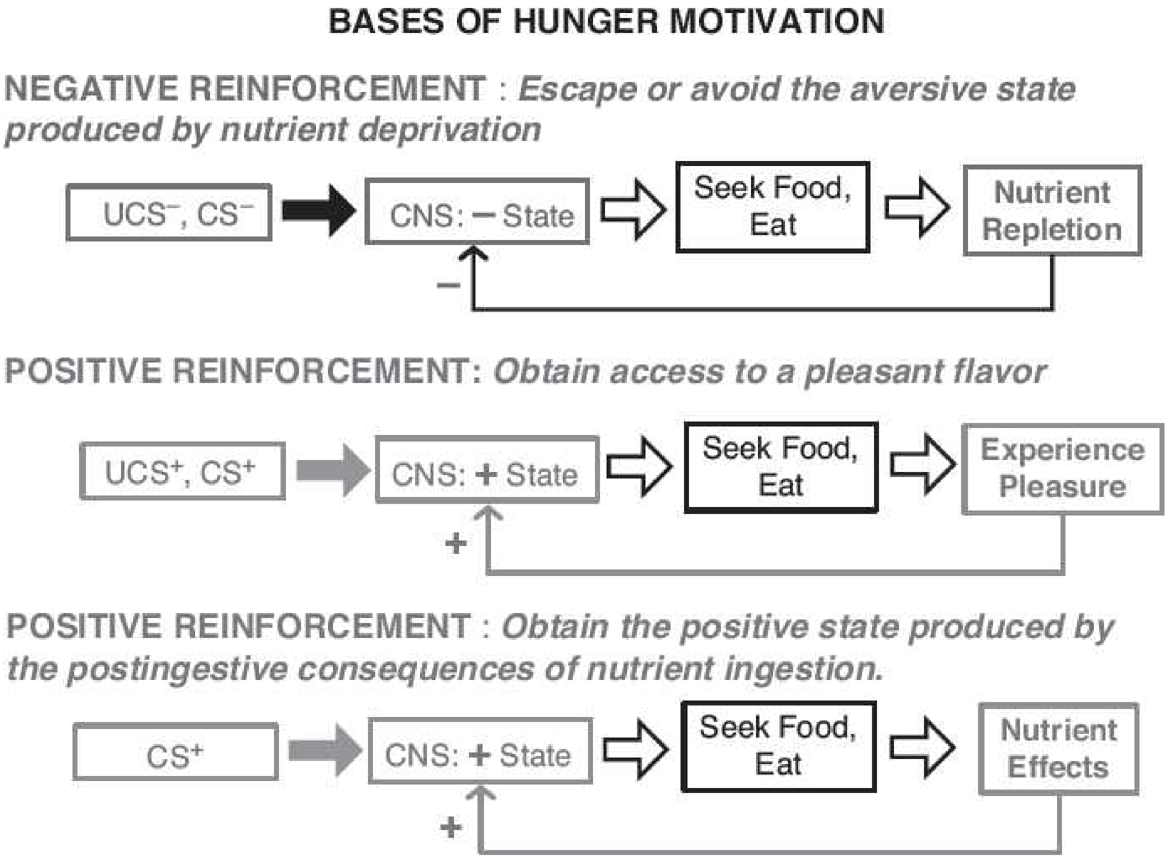
BASES OF HUNGER MOTIVATION
NEGATIVE REINFORCEMENT : Escape or avoid the aversive state produced by nutrient deprivation
UCS⁻, CS⁻
CNS: − State
Seek Food, Eat
Nutrient Repletion
−
POSITIVE REINFORCEMENT: Obtain access to a pleasant flavor
UCS⁺, CS⁺
CNS: + State
Seek Food, Eat
Experience Pleasure
+
POSITIVE REINFORCEMENT : Obtain the positive state produced by the postingestive consequences of nutrient ingestion.
CS⁺
CNS: + State
Seek Food, Eat
Nutrient Effects
+

FIGURE 1.17–1. Hunger-motivation constructs developed to explain food-seeking, meal initiation, and eating. **Top:** Negative-reinforcement hunger is the classical construct of hunger generated by nutrient depletion or homeostatic need. Depletion of metabolic energy supplies or of certain micronutrients such as sodium lead to aversive CNS states (UCS−) that are escaped by ingestion of the appropriate commodity. Such aversive states act as unconditioned negative reinforcers, such that with experience, initially arbitrary stimuli can elicit a similar state; these CS− permit avoidance as well as escape. **Middle:** Positive-reinforcement hunger generated by flavor hedonics. Some flavors such as sweet flavor, lead to positive CNS states (UCS$^+$) that elicit approach and ingestion. Other flavor stimuli (CS$^+$) come to elicit similar states via a variety of learning mechanisms, often not directly associated with UCS; even initially, aversive flavors such as chili pepper can become such CS$^+$. Such CS$^+$ are likely to be the most important hunger motivation mechanism in affluent humans. **Bottom:** Positive-reinforcement hunger generated by stimuli associated with postingestive GI or metabolic effects produced by ingested food. These effects generate positive CNS states that act as unconditioned positive reinforcers, which, with experience, can lead to the development of conditioned satiating stimuli (CS$^+$) that can elicit approach and ingestion. This process has been referred to as *appetition* because it is not clear that the GI and metabolic stimuli that generate it are identical to those that generate satiation. Note (i) that negative-reinforcement hunger is a negative feedback process (*solid black arrows*), whereas the

positive-reinforcement hungers are positive-feedback processes (*gray arrows*); and (ii) that the conditioned stimuli in all three types of hunger motivation can come to motivate eating without deprivation, which is presumably a cause of overeating.

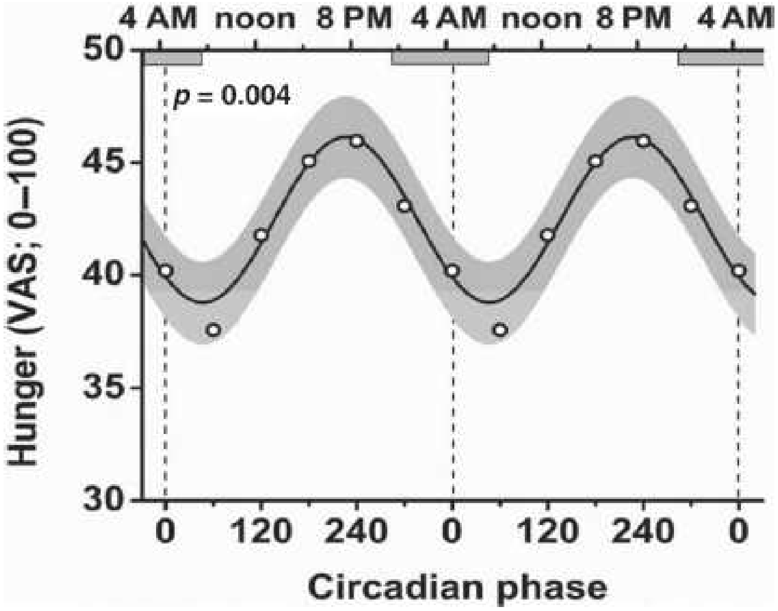

FIGURE 1.17–2. The circadian rhythm of hunger. Hunger was rated using visual analog scales (VAS; 0 = not at all hungry, 100 = extremely hungry) by healthy, normal-weight men who underwent a 13-day "forced desynchronization" protocol that balanced all behaviors, including meals and sleep periods, evenly across the endogenous circadian cycle under constant dim light. Data are double plotted; points are mean ratings in 4-hour bins; the black line is a cosine function ± 95% confidence interval (*gray area*) fit to the rating data, and horizontal gray rectangles indicate average habitual scheduled sleep episodes. Note that hunger minima occurred in the biologic morning and hunger maxima in the evening and that there was a 17% maximum–minimum difference. (From Shear FA, Morris CJ, Shea SA. The internal circadian clock increases hunger and appetite in the evening, independent of food intake and other behaviors. *Obesity (Silver Spring)*. 2013;21:421–423, with permission.)

Ghrelin became a candidate endocrine hunger signal when it was discovered that ghrelin infusion decreased the latency to eat and increased meal size in rats and mice; indeed, repeated ghrelin administration induced obesity. Animal experiments suggest that ghrelin may activate both negative- and positive-reinforcement hunger: (1) systemic ghrelin administration stimulated rats' ingestion of palatable foods more than normal chow and led to the formation of conditioned place preferences, suggesting that ghrelin stimulates positive-reinforcement hunger, and (2) ghrelin injected into the third cerebral ventricle failed to affect rats' initial rate of eating during the meal, which is an index of hedonic hunger, but did increase the rate of responding in a progressive-ratio test, an index of motivation, which together suggest that ghrelin elicited negative-reinforcement hunger. These results also suggest that ghrelin acts in multiple brain loci to stimulate eating. Among the sites of ghrelin, receptors suggested to mediate hunger are the ARC, the nucleus of the solitary tract (NTS), the ventral tegmental area (VTA), and the nucleus accumbens (NAc); these areas are discussed further below.

Evaluating Physiologic Endocrine Signaling. Endocrinology and neuroscience have evolved criteria for identifying normal physiologic functions of chemical signals. Table 1.17–1 presents a current set. Criteria 1 and 2 address the plausibility of the candidate hormonal signal to control a particular function; criteria 3 to 5 concern its sufficiency; and criterion 6, its necessity. For candidate endocrine signals such as ghrelin, plasma measurements and intravenous infusions suffice to test these criteria. As described below, however, some GI hormones also act locally. Modes of local action of GI hormones include paracrine effects and synaptic connections between processes of enteroendocrine cells called neuropods and vagal afferents. Assessing the physiologic relevance of such local actions requires measurements and administration in the local environment, that is, in the intestinal lamina propria or, like other neurotransmitters or neuromodulators, in the synaptic cleft.

Table 1.17–1.
Empirical Criteria to Assess Physiologic Status of Chemical Signals[a] in Meal-Related Functions

1. Concentrations of the molecule change at the site of action in a pattern consistent with the effect.[b]
2. Cognate receptors for the molecule are expressed at its site(s) of action.
3. Exogenous administration of the molecule in amounts duplicating the meal-related changes in endogenous patterns at the site of action produces the effect.
4. Administration of secretagogues for the molecule produces effects similar to the effect of the hormone.
5. The molecule's effect occurs in the absence of abnormal behavioral, physiologic, or subjective effects.
6. Administration of selective agonists and antagonists of the molecule's receptors produce effects that are consistent with their receptor pharmacologies.[c,d]

[a] Signaling molecules may be hormones, cytokines, neurotransmitters, and so forth.

[b] At a minimum, the change in concentrations of the proposed signal should precede the effect.

[c] For example, the administration of specific and potent receptor antagonists should delay or reduce eating in the case of a hunger signal or increase eating in the case of a satiation signal.

[d] We do not include phenotypic evaluation of global transgenic or spontaneous genetic loss-of-function models in this criterion. These are valuable research tools, but complications due to developmental compensatory effects, pleiotropic actions, and species differences preclude their use as a "necessity" criterion for physiologic function. Rapidly inducible, tissue-specific reductions in gene function, however, may complement the use of receptor antagonists in establishing necessity.

At present, ghrelin fulfills some but not all of these criteria. It is a plausible hunger signal (criteria 1 and 2) because plasma ghrelin levels increase before meals, increase during food deprivation, and decrease after meals in animals and humans (Fig. 1.17–3) and because the infusion of ghrelin to brain areas that contain ghrelin receptors simulates eating in animals. Such effects can be obtained without side effects (criterion 5). Two cardinal criteria, however, have not yet been met in humans: it is not known whether ghrelin infusions that mimic the changes in ghrelin produced by premeal increases in endogenous ghrelin are sufficient to stimulate eating (criterion 3) or whether antagonism of endogenous ghrelin signaling before meals is sufficient to delay meal onset (criterion 6), although this has been shown in animals. The recent discovery that liver-enriched antimicrobial peptide 2 (LEAP2), an endogenous GHS antagonist, changes levels in the opposite direction to changes in ghrelin levels with eating and body mass suggests that the critical stimulus for initiating eating could be the ghrelin/LEAP2 ratio.

MEAL MAINTENANCE

Flavor Hedonics

The maintenance of eating during the meal is attributed to the ongoing influence of the hunger processes reviewed above, together with the eating-stimulatory effect of positive flavor hedonics. Flavor stimuli elicited during eating arise from olfactory (cranial nerve I), gustatory (VII, IX, X), and somatosensory (V) receptors in the oronasopharynx. Flavor stimuli contribute to (1) detection and selection of foods; (2) stimulation or inhibition of eating, mainly through positive or negative hedonic processes, presumably via mechanisms overlapping with those producing hedonic hunger discussed above; and (3) associative learning processes, both as conditioned stimuli (CS) and, in the case of sweet and fat flavors, unconditioned or reinforcing stimuli. An important example is learned incentive salience. In this process, the perceptual salience and valence of stimuli associated with preferred flavors are increased with experience, leading to increased identification, selection, and intake of foods with those flavors.

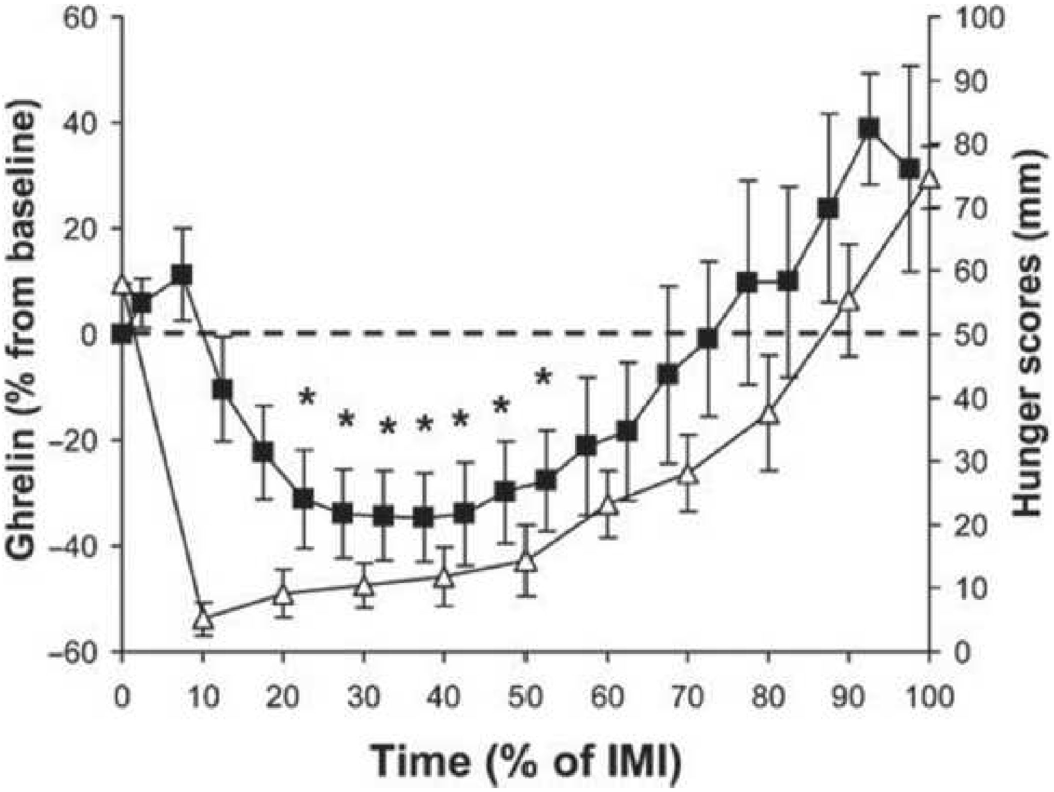

FIGURE 1.17–3. Close association of hunger scores (*open triangles,* rated on a 100-mm visual analog scale) and plasma ghrelin concentration (*filled squares,* total ghrelin-like immunoreactivity) during the interval between lunch (begun at time 0) and a freely requested dinner (time 100%). Subjects were six healthy, normal-weight males. The mean lunch size was 800 kcal, the mean lunch duration was 20 minutes and the mean intermeal interval was 359 minutes. Data are mean ± standard error of the mean. Baselines are values at lunch onset. *Significantly different from time 0. (Reprinted with permission from Cummings DE, Fayo RS, Marmonier C, Aubert R, Chapelot D. Plasma ghrelin levels and hunger scores in humans initiating meals voluntarily without time- and food-related cues. *Am J Physiol.* 2004;287(2):E297. doi:10.1152/ajpendo.00582.2003.)

Psychophysics. Subjective aspects of flavor hedonics are often measured with visual analog scales (VAS). One common scheme is to instruct subjects to indicate how pleasant a flavor is by marking a 100-mm line labeled at the ends as “not at all” and “extremely.” Most subjects produce reliable and coherent VAS ratings, but patients with anorexia nervosa often do not. In addition, it is now clear that VAS ratings are relatively insensitive to individual differences and not appropriate for comparisons between different classes of subjects who may use the scale differently, such as men versus women or individuals with and without obesity. This problem can be avoided with the use of generalized scales whose extreme values are labeled with the most intense imaginable affective experience of any kind.

Sham and Brief-Access Eating. Sham eating and brief-access eating tests are designed to isolate controls of eating by flavor stimuli in the absence of postingestive consequences. An elegant method for brief-access tests in rats involves measurement of licks of different fluids during repeated 10-second presentation periods, with presentation order randomized so that although the fluids enter the stomach normally, their mixture prevents the association of particular flavors with postingestive effects. Sham eating eliminates postingestive food stimuli altogether. It is done in rats with a surgically implanted gastric cannula, which when opened, causes liquid food to drain out rather than accumulate in the stomach or enter the intestines in appreciable amounts. Rats' brief-access and sham intakes of sweet solutions or of oil emulsions increase linearly over a wide range of concentrations, indicating the potent effect of flavor hedonics on eating (Fig. 1.17–4). Sham eating is done in humans simply by having subjects take food into the mouth and expel it without swallowing. This method has been used successfully with women with anorexia nervosa or bulimia nervosa. Interestingly, it revealed a greater propensity to sham eat in women with bulimia nervosa than in women without a history of eating disorders.

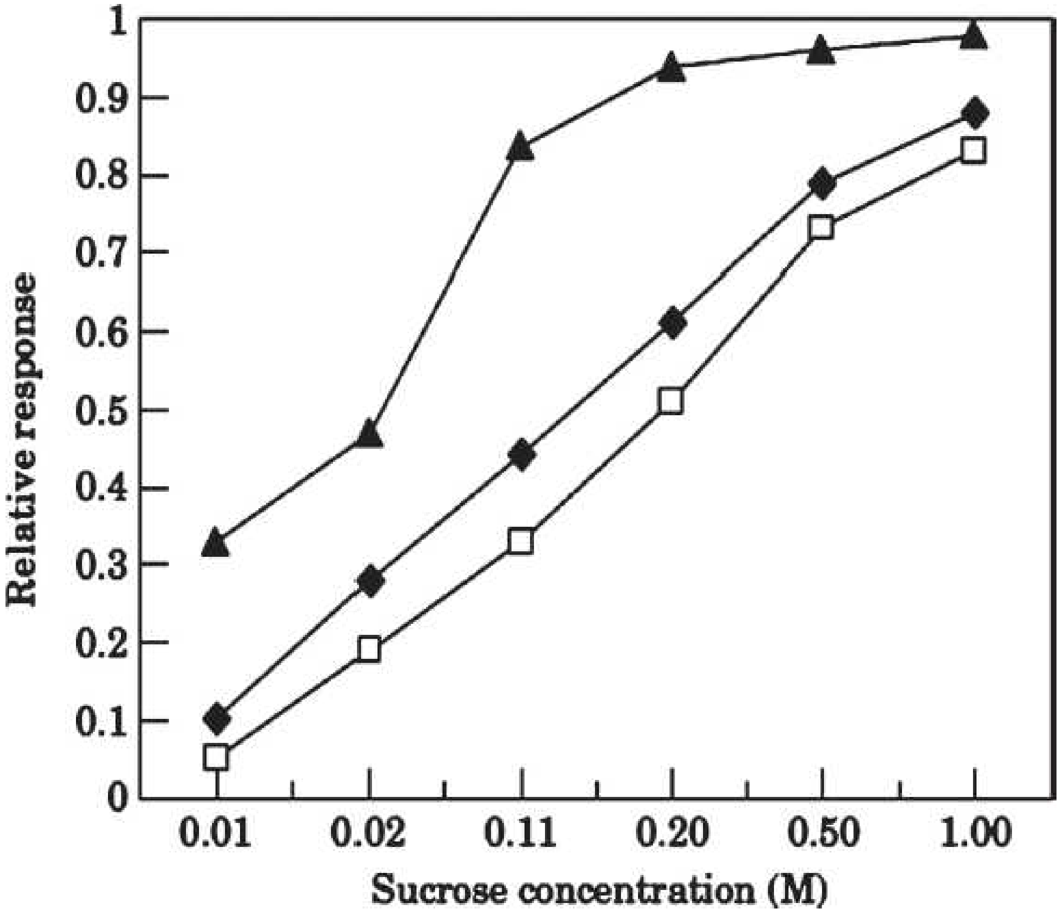

FIGURE 1.17–4. Flavor hedonics stimulates ingestion in rats. Rats' responses (expressed relative to maximal) to varying sucrose concentration in 30-minute sham-feeding tests (*open squares*), in 30-second brief-access licking tests (*filled triangles*), and from electrophysiologic recordings from the greater superficial petrosal nerve (a branch of cranial nerve VII, *filled diamonds*). (Reprinted from Smith JC. The history of the "Davis Rig." *Appetite*. 2001;36(1):93–98. Copyright © 2001, with permission from Elsevier.)

Eating in the Absence of Hunger. Eating in the absence of hunger (EAH) denotes eating driven by hedonics after satiation is achieved with less preferred foods (Table 1.17–2). Pediatric EAH, assessed by offering snack foods after a standard meal is strongly associated with overweight and obesity. Note that EAH refers only to the presumed absence of negative-reinforcement hunger; thus, "hedonic hunger" would be less ambiguous.

Innate and Learned Flavor Hedonics

Although some preferences (sweet) and aversions (bitter or sour) for basic gustatory stimuli appear to be innate, preferences and aversions for the vast majority of flavors are predominately learned. Foods' GI and metabolic consequences can reinforce such learning. This occurs in conditioned satiations, conditioned aversions (including the marked aversions for flavors associated with acute upper GI illness), and "specific hungers" (preferences for flavors associated with foods containing some, but by no means all, vitamins or minerals that can be learned during states of nutritional deficiency). Note that in these situations, with the exception of sodium appetite during sodium depletion, it is the discriminative rather than the hedonic aspects of flavor stimuli that are important, although increased or decreased hedonic responses are part of what is learned. Human flavor preferences also arise from the emotional, cognitive, and cultural associations attached to various foods, independent of their nutritional or physiologic properties. Most individuals endorse especially strong preferences, or *cravings,* for particular foods. Because they dramatically affect an individual's tonic level of food intake and their adherence to therapeutic dietary regimens, the origins and plasticity of human food preferences are important research challenges.

Table 1.17–2.
Two Demonstrations of "Eating in the Absence of Hunger"

	Meal	**EAH**
Standard meal (kcal)	861 ± 30	365 ± 20
Large-array meal (kcal)	1,309 ± 55*	295 ± 18

Adolescent subjects were first offered a midday meal consisting of either a large array of lunch foods in a buffet setting or a single standardized meal consisting of a subset of foods from the large-array meal (chicken nuggets, grapes, baby carrots, tortilla chips, cookies, 2% fat milk, etc.) served in excess. Subjects were instructed to eat until they were no longer hungry and to open the room when they were done. Note the effect of variety: subjects ate significantly larger meals in the large-array condition (*). To measure "eating in the absence of hunger" (EAK) subjects were offered unexpectedly an array of novel, highly palatable snack-food items 30 min after their meals. Note that they ate 300 to 400 kcal more regardless of the type of first course served, although there was a significant difference in the conditions ($^{+}$). Data are mean ± standard error of the mean. (Data from Shomaker LB, Tanofsky-Kraff M, Zocca JM, et al. Eating in the absence of hunger in adolescents: intake after a large-array meal compared with that after a standardized meal. *Am J Clin Nutr.* 2010;92:697.)

Variety. Variety is a strong hedonic control of eating. Offering a variety of nutritionally similar foods with different, preferred flavors leads to larger meals than does offering only one of the alternatives, even the single most preferred one. Table 1.17–2 shows an example. The effect of variety on intake is referred to as sensory-specific satiety, reflecting the decline in the preference for a consumed food in relation to preferences for foods that have not been recently consumed. Flavor variety can tonically increase rats' intake and increase body weight.

Familiarity. Familiarity is sufficient to condition flavor preferences. This phenomenon likely explains much of the marked cultural variety of preferred foods, the social contexts or times of day when they are eaten, and so forth. For example, children repeatedly faced with foods that they find unpalatable and with the sight of adults enjoying them learn to enjoy and eat them as well. This learning has to overcome the innate tendency to avoid novel flavors (neophobia), which is especially strong in young children.

Social Facilitation. Eating with acquaintances markedly increases intake in comparison with eating alone, whereas no facilitation is evident when eating with strangers. Given how often people eat with others, the characteristics and mechanisms of this effect deserve analysis.

Individual Differences. Heterogeneity in human gustatory capacity is a source of individual differences in hedonics. For example, the density of fungiform papillae on the anterior tongue is usually greater in women than in men, and this is associated with both individual and sex differences in bitter and sweet perception and hedonics. Furthermore, individual differences in sweet or bitter taste capacity are associated with the hedonic valence of a wide variety of flavors (Fig. 1.17–5). Such hedonic differences, in turn, are associated with body mass index (BMI*;* weight in kg/height in m^2), indicating that they tonically increase meal size and caloric intake. Recent work has indicated that the degree of sweet preference and the change following surgery can serve as a predictor of weight loss with Roux-en-Y bariatric surgery, suggesting that change in preference can play a role in modulating caloric intake.

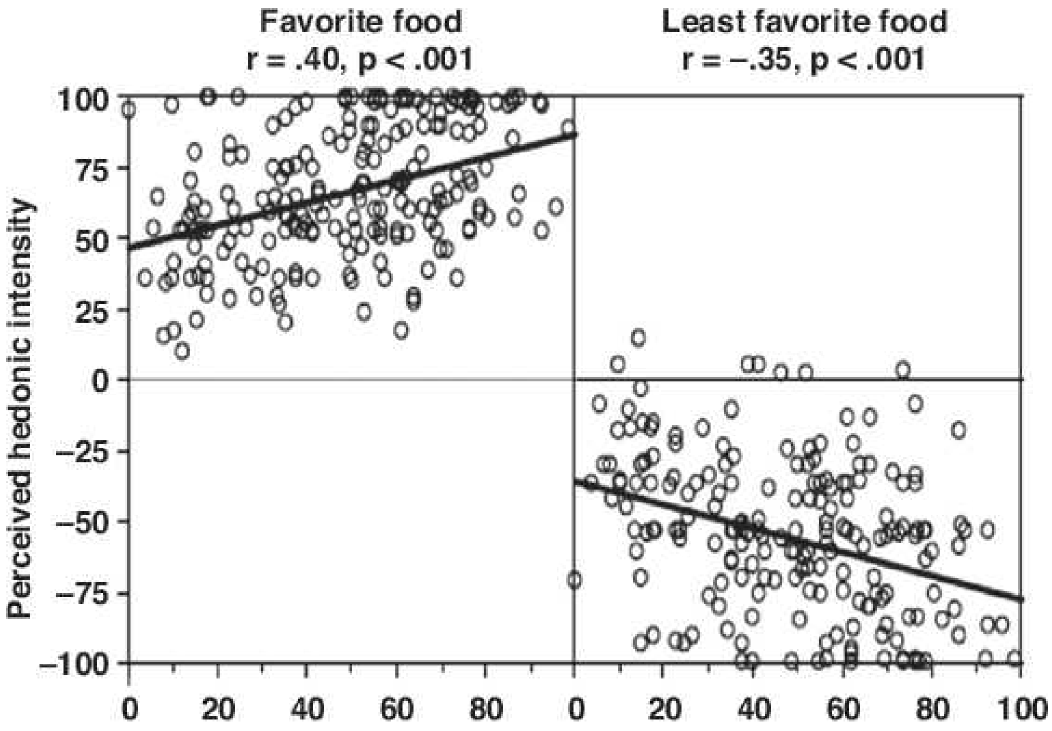

FIGURE 1.17–5. Individual differences in flavor hedonics. Hedonic intensities were rated with a general labeled magnitude, the gLMS scale, for favorite and least favorite foods as described in the text. Ratings are plotted against subjects' taste capacity, assessed by the average sensory intensity rating of four standard taste solutions (1 M NaCl, 1 M sucrose, 0.032 M citric acid, and 0.001 M quinine hydrochloride). Note the significant correlations between taste capacity and both positive (favorite food) and negative (least favorite food) flavor hedonics. (From Kalva JJ, Sims CA, Puentes LA, Snyder DJ, Bartoshuk LM. Comparison of the hedonic general labeled magnitude scale with the hedonic 9-point scale. *J Food Sci.* 2014;79(2):S238–S245.)

SATIATION AND POSTPRANDIAL SATIETY

Eating-inhibitory processes that occur during meals and contribute to meal termination are known as satiation processes. These are differentiated from longer-lasting processes that inhibit eating during the subsequent intermeal interval, known as postprandial or across-meal satiety processes. Because some signals appear to contribute to both, these processes are discussed together.

Gastric Mechanoreception

Although experimental dissection of gastric and postgastric signals in the control of eating is challenging, a variety of evidence indicates that gastric volume contributes directly to satiation via mechanoreceptors that are tuned to both tension and stretch or length. For example, (1) imaging studies have revealed significant correlations between fullness and gastric volume, (2) increasing a meal's volume by mixing in air decreased meal size, and (3) incorporating water into foods maintained gastric volume at higher levels, increased fullness, and decreased subsequent intake more than ingestion of the same amounts of food and water served separately. Gastric mechanoreceptors are linked to the brain mainly by vagal afferents.

Intestinal Nutrient Sensing

Food stimuli in the small intestine activate mechano-, osmo-, and chemoreceptors that initiate a number of potent satiation signals. Most important are the chemoreceptors tuned to detection of glucose, fatty acids, amino acids, and other digestive products. Negative-feedback signals generated by these nutrient sensors are relayed to the brain via endocrine and neural pathways, some of which are described below.

Oral preloads can be used to investigate the contributions of a food's metabolic energy content, nutrient composition, colligative and osmotic properties, and so forth to its satiating potency. When mixed nutrient preloads are used, decreases in eating are often proportional to the sizes of the preloads, that is, to their metabolizable energy contents. When the macronutrient compositions of the preloads vary, however, protein is typically more satiating on a calorie-for-calorie basis than carbohydrate or fat. Comparisons of intraduodenal and intravenous infusions indicate that intestinal signals inhibit eating more potently than postabsorptive signals. For example, in one comparison of intraduodenal and intravenous glucose infusions that produced identical increases in systemic glucose levels, intraduodenal glucose but not intravenous glucose decreased premeal hunger ratings.

The next sections describe just three of the many intestinal hormones thought to contribute to satiation and postprandial satiety. The probable existence of numerous signaling mechanisms for satiation raises the question of their interactions and their relative contributions in various normal and abnormal contexts. As yet, little is known about these questions. Finally, the peripheral origin and, in most cases, peripheral sites of action on the eating of GI hormones make them attractive targets for the development of pharmacologic therapies.

Cholecystokinin. Cholecystokinin (CCK) is secreted by enteroendocrine cells in the small intestines. CCK secretion is stimulated by glucose and other sugars, amino acids, and long-chain fatty acids. Secretion begins soon after meal onset and usually returns to basal levels during the intermeal interval.

CCK's role as an endocrine satiation signal is uniquely well established. Intravenous CCK infusions mimicking postprandial plasma levels reduced meal size without adverse physical or subjective effects in men and women (Table 1.17–1, criterion 3), intraduodenal lipid infusions stimulated satiation, which was reversed by a CCKA receptor antagonist (criteria 4 and 6), and infusions of the antagonist alone increased meal size (criterion 6) (Fig. 1.17–6). No comparable antagonist effects have been reported in humans for any other hypothesized endocrine control of eating. Finally, occasional individuals bearing null mutations of the CCKA (or CCK1) receptor are extremely obese, and more common CCKA receptor polymorphisms are associated with increased meal size, increased food intake, and obesity. CCK also fulfills the criteria for a normal physiologic satiation signal in animals. Indeed, CCK research in rats and mice has been paradigmatic for the study of the control of eating in animal models. CCK elicits highly specific, dose-related decreases in meal size. The specificity is indicated by: (1) CCK fails to inhibit water intake in water-deprived rats under conditions where it inhibits eating liquid food; (2) CCK elicits behavioral signs of satiation, including postprandial grooming and sleep, in rats that are sham feeding with open gastric cannulas, which otherwise sham feed essentially indefinitely; and (3) CCK can decrease food intake without signs of illness, such as pica or formation of conditioned taste aversions (of course, large doses can induce illness). In summary, a wide variety of data indicates CCK is an important part of the natural process of satiation. As well these data indicate that the physiologic system controlling satiation, however complicated, is not completely redundant.

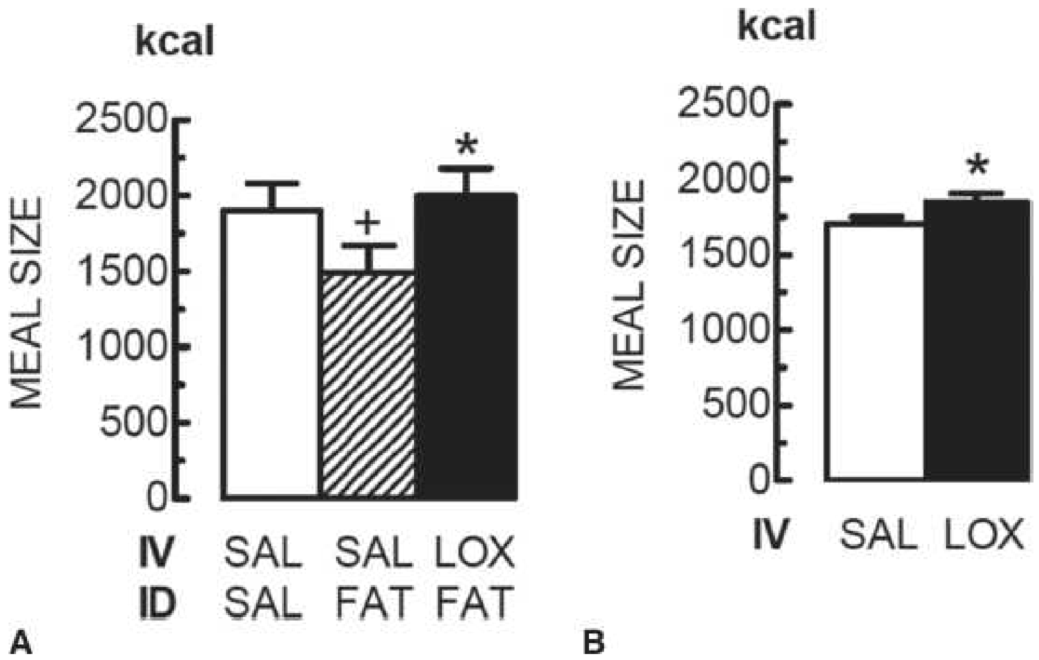

FIGURE 1.17–6. Evidence from CCKA receptor-antagonist studies that endogenous CCK is a physiologic satiation signal. **A:** The CCKA receptor antagonist loxiglumide (LOX) antagonized the satiating action of endogenous CCK stimulated by intraduodenal (ID) infusion of a fat emulsion. Healthy-weight adult males began a midday lunch buffet 4 hours after a standard breakfast, 90 minutes after the onset of an IV infusion of 10 $\mu mol \cdot kg^{-1} \cdot h^{-1}$ LOX or SAL, and 60 minutes after an ID infusion of 0.4 $mL \cdot min^{-1}$ corn oil (FAT) or SAL, and 20 minutes after an oral preload of 400 mL of a low-fat banana milkshake (bananas were used because they do not elicit CCK secretion). Infusions were continued throughout the meal. ID fat infusion significantly reduced the size of the lunch meal (+), and this was reversed by LOX (*); no physical or subjective side effects occurred in any condition. (Data from Matzinger D, Gutzwiller JP, Drewe J, et al. Inhibition of food intake in response to intestinal lipid is mediated by cholecystokinin in humans. *Am J Physiol.* 1999;277:R1718.) **B:** Antagonism of CCK signaling with the CCKA receptor antagonist LOX stimulated eating. Healthy-weight adult males began a midday lunch buffet 4 hours after a standard breakfast and 60 minutes after beginning an IV infusion of 22 $\mu mol \cdot kg^{-1} \cdot h^{-1}$ LOX or SAL. Infusions were continued throughout the meal. LOX significantly increased meal size (*) without physical or subjective side effects. (Data from Beglinger C, Degen L, Matzinger D, D'Amato M, Drewe J. Loxiglumide, a CCK-A receptor antagonist, stimulates calorie intake and hunger feedings in humans. *Am J Physiol.* 2001;280(4):R1149.)

CCK appears to act on CCKA receptors in the pyloric area, liver, and brain to elicit satiation. At least in rats, CCK also acts locally to elicit satiation, that is, in the intestinal lamina propria before absorption. For example, intravenous infusion of a small-molecule CCKA receptor antagonist, which presumably diffused from the capillaries into the lamina propria, increased food intake, but infusions of a CCK antibody, which would not escape the vasculature, did not. Activation of abdominal CCKA receptors stimulates vagal afferent neurons, which relay the feedback information to the NTS.

Glucagon-Like Peptide-1. Glucagon-like peptide-1 (GLP-1) is produced by enteroendocrine in both the small and large intestine and by a small population of neurons in the NTS. All three macronutrients lead to intestinal GLP-1 secretion. Secreted GLP-1 is rapidly inactivated by dipeptidyl peptidase-IV (DPP-IV), a proline/alanine-specific peptidase found on the luminal surface of capillary endothelial cells, in the liver, and in the blood. Despite this, and in contrast to CCK, GLP-1 secretion and plasma levels are elevated for hours after meals and return to basal only overnight. Thus, GLP-1 may function either as a satiation signal, a postprandial satiety signal (i.e., delaying onset of the next meal), or an across-meal satiation signal (i.e., decreasing the size of the next meal). GLP-1 receptors on abdominal vagal afferents and possibly in the brain mediate GLP-1 satiation.

Whether GLP-1 signals satiation in humans has not been fully established. Consistent with a satiating effect, intravenous infusion of physiologic doses of GLP-1 reduced meal size in the absence of adverse effects in normal-weight men (it has not been tested in women). In two tests, however, GLP-1 receptor antagonism failed to increase meal size, although the same antagonist had clear effects on insulin secretion and glycemia in other tests. GLP-1's status as a normal satiation signal deserves further study for several reasons: (1) GLP-1 antagonism stimulated secretion of glucagon and PYY(3–36, see below), which may have interfered with its desatiating effect; (2) the same antagonist increased eating in several tests in rats and mice; (3) the antagonist may not have reached GLP-1 receptors in the lamina propria that may be important for satiation; (4) the possibility that GLP-1 may function as an across-meal control of eating has not been explicitly tested; and (5) long-lasting GLP-1 agonists such as liraglutide (Victoza or Saxenda, Novo Nordisk) and semaglutide (Ozempic, Novo Nordisk) reduce eating and body weight in patients with obesity likely through both peripheral and brain sites of action.

Peptide Tyrosine Tyrosine. Intestinal enteroendocrine cells that secrete peptide tyrosine tyrosine (PYY) are located mainly more distal than CCK and GLP-1 cells, so plasma levels increase more slowly after meals. All three macronutrients stimulate PYY secretion, with long-chain fatty acids the most potent. DPP-IV acts on secreted PYY (i.e., PYY[1–36]) to produce the active form, PYY(3–36). Interestingly, in rats, peripheral administration of PYY(3–36) selectively stimulates Y2 receptors and inhibits eating, whereas central administration of PYY(1–36) stimulates other Y-receptor types and stimulates eating.

PYY(3–36)'s role in postprandial satiety has been tested using intermeal infusions that began after standard meals and ended before test meals. Although these tests led to reductions in eating without signals of illness, it was unclear whether the doses were physiologic. Infusions of supraphysiologic doses of PYY(3–36) during meals also reduced intake, but in this case, the subjects usually became ill. Thus, whether PYY(3–36) has a normal role in the control of eating is uncertain. It will be important for future research on this issue to focus on the possibility that PYY signals locally via neuropod signaling.

Vagal Afferent Signaling. A variety of peripheral eating-inhibitory feedbacks are signaled to the brain via vagal afferents. Vagal afferent nerve endings densely innervate the lamina propria in most of the GI tract and, less densely, the hepatic paraganglia, the portal vein, and the bile duct. Spinal visceral afferents of the splanchnic nerves have a similar innervation pattern, but their role in eating is less well understood. Abdominal vagal afferents are stimulated by glucose and other digestive products, by CCK, GLP-1, and other hormones, and by serotonin (5HT) and other signaling molecules that are released by food stimuli. Their importance in satiation is reflected in the increase in meal size that follows a selective surgical section of sensory abdominal vagal fibers in rats (Table 1.17–3). Total vagotomy may not result in increased meal size because of the impairment of GI motor function caused by loss of vagal efferents.

Vagal afferents integrate neural signals as well as relay them to the brain. For example, in rats, (1) some individual vagal afferent fibers are stimulated by both gastric fill and CCK infusion, and combinations of the two greatly amplify the vagal response (Fig. 1.17–7), and (2) other fibers are sensitive to both ghrelin and CCK, and the two functionally antagonist stimuli sum to determine the net vagal electrophysiologic response. Vagal afferent cell bodies in the nodose ganglion also demonstrate integrative properties in that patterns of mRNA expression of orexigenic and anorexigenic peptides and receptors vary depending on the nutritional state. Vagal (and spinal visceral) afferents project to the NTS, as described in the CNS section. Recent mouse work has suggested laterality in the function of vagal afferent fibers, with fibers arising from the left nodose mediating satiating feedback and fibers from the right nodose mediating reward or appetition aspect of ingestion.

Table 1.17–3.
Disconnection of Subdiaphragmatic Vagal Afferents Leads to Chronic Increases in Spontaneous Meal Size in Rats

	Control	SDA
Meal size (licks)	1,657 ± 109	2,309 ± 102*
Intermeal interval (min)	78 ± 6	92 ± 5*

Subdiaphragmatic vagal deafferentation (SDA) combines unilateral dorsal vagal rhizotomy at the brainstem and section of the ipsilateral abdominal vagal trunk, which arises from the contralateral side of the brainstem. Rats were maintained on a liquid diet (Ensure, Ross Laboratories). Control rats received sham surgery. SDA increased spontaneous meal size by ~1/3 (*) and increased intermeal interval by ~15%, likely in response to the larger meals. Data are averages over 3 day in the 3rd week after SDA surgery, when average postsurgical body weights and rates of body weight gain had reached their presurgical values, mean ± standard error of the mean. (Adapted from Schwartz GJ, Salorio CF, Skoglund C, Moran TH. Gut vagal afferent lesions increase meal size but do not block gastric preload-induced feeding suppression. *Am J Physiol.* 1999;276:R1623, with permission.)

Energy Homeostasis

Adiposity and Energy Homeostasis. The body contains both online and stored substrates for energy metabolism. Both are homeostatically regulated. Online energy substrates are metabolites circulating in the blood that can be taken up by cells and used to generate ATP. The most prominent is blood glucose, whose levels are actively regulated. Stored energy substrates include glycogen in the muscle and liver, which amounts to a few thousand calories and triacylglyceride stored in the adipocytes, which amounts to more than 70,000 kcal in people at healthy weights. The relative constancy of body weight displayed by most people through adulthood, during which one eats about 10 times as many calories per year as are stored in the adipose tissue, strongly suggests that adipose-tissue mass is physiologically regulated. The influences of energy stores on eating are mediated primarily via the control of meal size. Recent work indicates that the control of eating by energy stores can be accurately modeled by simple negative-feedback control systems. At the same time, the epidemic rates of type-2 diabetes mellitus and obesity in many societies indicate that lifestyles in these societies stress the capacities for glycemic and adipose tissue mass regulation beyond their capacities to handle surfeits. The feedback signals used by the brain to detect increases in the mass of adipose tissue have not been identified.

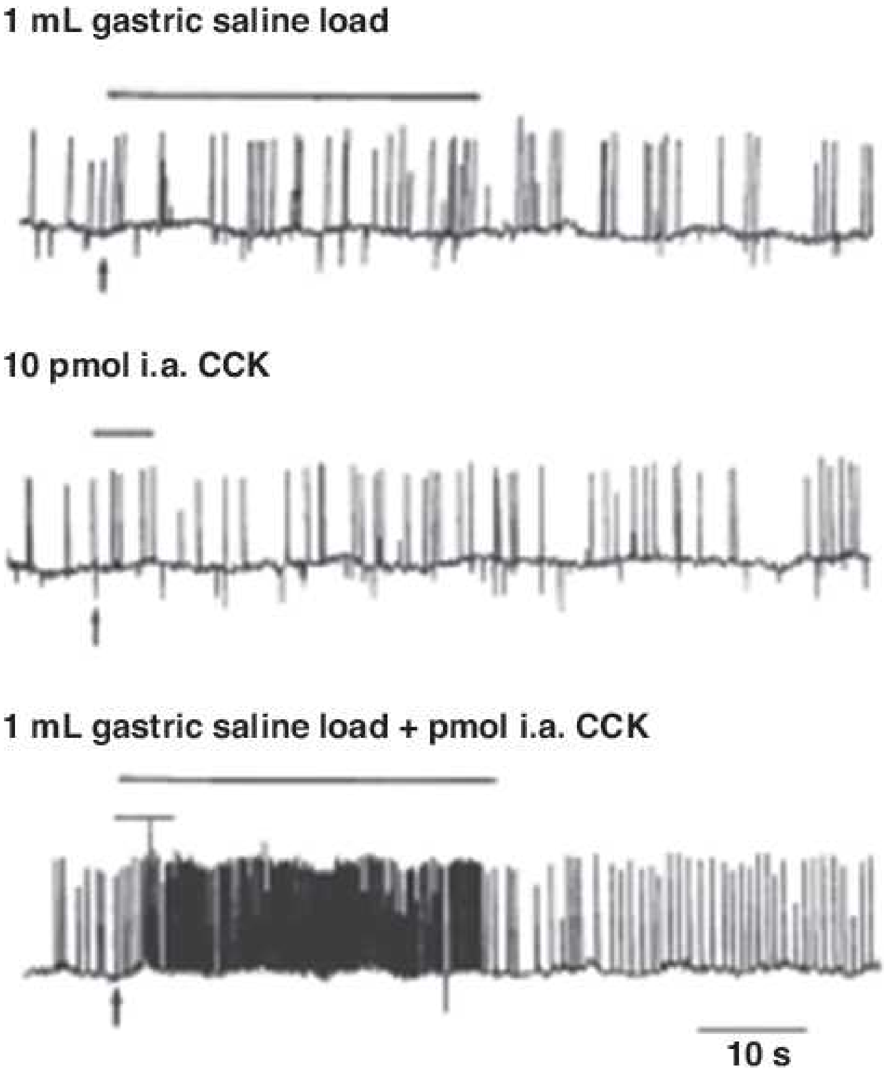

FIGURE 1.17–7. Synergistic effect of gastric mechanoreception and CCK on vagal afferent signaling in anesthetized rats. Data are extracellular recordings of action potentials in single vagal afferents in response to 1 mL intragastric saline loads and intraarterial (i.a.) infusions of 10 pMol CCK. (From Schwartz GJ, McHugh PR, Moran TH. Gastric loads and cholecystokinin synergistically stimulate rat gastric vagal afferents. *Am J Physiol*. 1993;265:R872–R872.)

Leptin. Leptin is a hormone secreted by adipocytes in proportion to their volume. Leptin's role in energy homeostasis is to defend against underweight, not to defend against overweight. That is, decreases in leptin levels below the usual level are sufficient to increase eating and decrease energy expenditure, but increases in leptin above the usual level have little effect. Note that "usual level" means the level to which the organism is adapted. Thus, long-term obese persons will increase their eating and decrease energy expenditure when leptin is reduced to levels that are still much higher than those in never-obese persons. How long a stable weight level must be maintained for such adaptation is unclear. Leptin physiology makes leptin monotherapy ineffective against obesity. Leptin has, however, shown promise in preclinical studies when administered in combination with other hormone-based therapies, such as long-acting forms of amylin, GLP-1, or CCK. With sustained weight loss, leptin levels fall below what would be predicted for the degree of adiposity. Restoring leptin levels with exogenous administration increase energy expenditure and decreases eating, suggesting that there could be a role for leptin therapy in maintaining a weight loss. Leptin is actively transported through the blood–brain barrier (BBB) and acts in a variety of brain loci to affect, among others, eating, energy expenditure, and reproductive-axis function.

BRAIN MECHANISMS OF APPETITE

Eating is mediated by a complex, anatomically diffuse neural network that extends from the caudal brainstem to the frontal lobes of the cerebral cortex. Figure 1.17–8 schematizes its general organization. A large number of interneuronal signaling molecules involved in this network have been identified and the application of increasingly powerful molecular-genetic and sophisticated behavioral methods has led to the construction of an at least rudimentary “wiring diagram.” It is no criticism of this demanding work to say that it approaches but does not quite attain the truly physiologic. That is, what normally leads to the excitation and inhibition of these neurons and under what condition the signaling molecules involved meet the criteria of Table 1.17–1 for endogenous neurocrine function are not yet clear.

Brainstem

The brainstem has vital sensory, motor, and integrative functions related to eating. As mentioned above, the NTS receives vagal and spinal visceral afferents. It also is the target of primary gustatory afferents. In addition, it expresses a variety of metabolite receptors and hormone receptors (e.g., ghrelin, GLP-1, and leptin) that directly sense eating-relevant information. The NTS is also the target of neural efferents from the area postrema, which, like other circumventricular structures, lacks a normal BBB and has receptors for a variety of circulating molecules, including: (1) receptors for the hormone amylin, which is a candidate satiation signal; (2) receptors for several toxins, which mediate conditioned taste aversion learning and elicit the emesis reflex; and (3) osmoreceptors. The brainstem also contains the lower motor neurons that produce eating and swallowing movements, upper motor neurons that drive lower motor neurons, and central pattern generators that coordinate these movements. Accordingly, difficulty eating and swallowing are signs of motor neuron diseases, such as bulbar palsy, that affect these motor networks. Finally, the NTS integrates a wide variety of information contributing to the control of eating. In addition to the sensory inputs described above, it is the target of projections from a variety of brainstem and hypothalamic areas, some of which are described below.

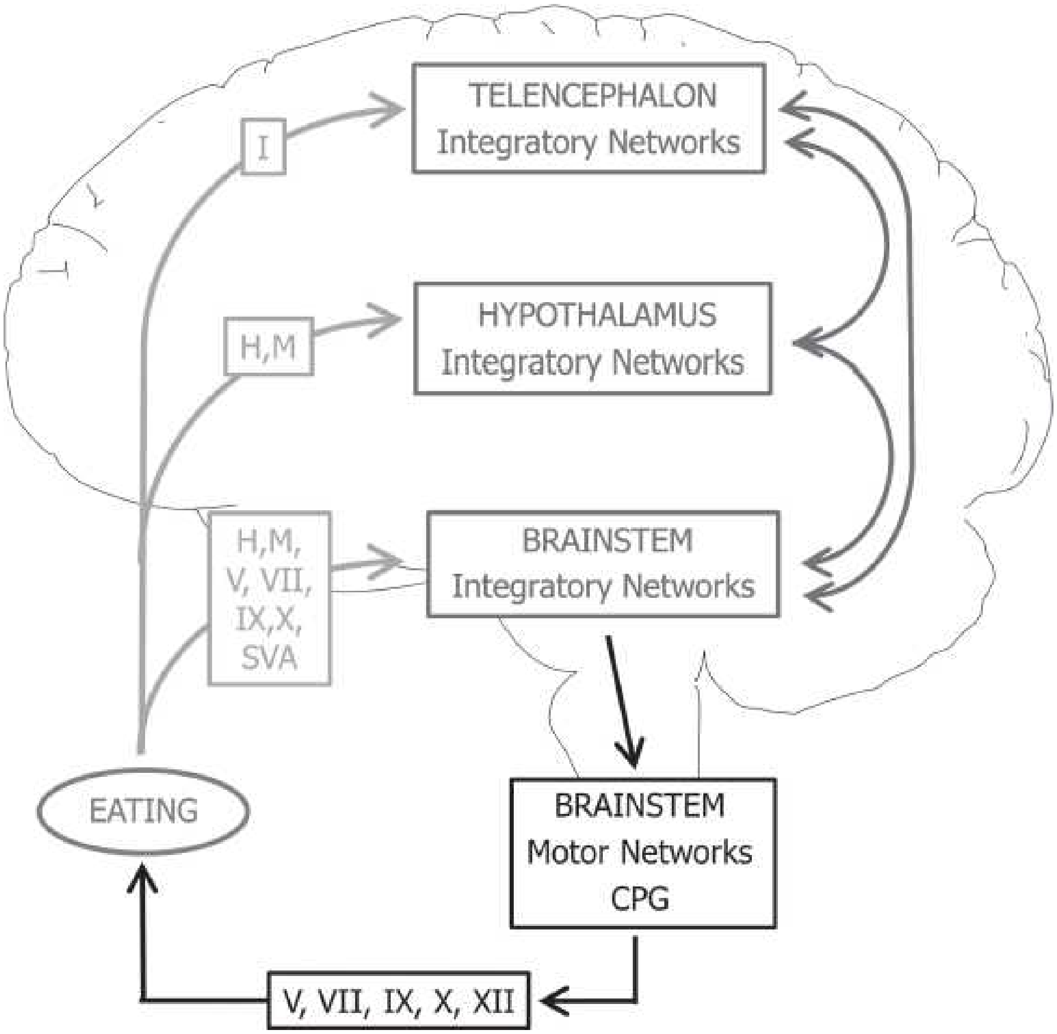
TELENCEPHALON
Integratory Networks
I
HYPOTHALAMUS
Integratory Networks
H,M
BRAINSTEM
Integratory Networks
H,M,
V, VII,
IX,X,
SVA
EATING
BRAINSTEM
Motor Networks
CPG
V, VII, IX, X, XII

FIGURE 1.17–8. Schematic of the major brain areas involved in the control of eating as well as their principal afferent connections (*light gray arrows*), interconnections (*dark gray arrows*), and motor outputs (*black arrows*). Roman numerals indicate cranial nerves; H, hormones; M, metabolites; SVA, spinal visceral afferents; CPG, central pattern generator.

Tests of decerebrate rats, that is, rats bearing transections of the brain at the level of the midbrain, reveal the impressive integratory capacity of the caudal brainstem. When liquid food is infused into their mouths via oral catheters, they (1) ingest food in a meal-like pattern (which they end by food drain from the mouth) and, after meals, groom and become quiescent, similar to intact rats (2) eat progressively more of sweeter food (Fig. 1.17–9), (3) display the orofacial movements associated with preferred and nonpreferred foods in intact rats, (4) eat less after CCK injection, and (5) eat more after pharmacologic blockade of glucose metabolism. The caudal brainstem does not, however, have the capacity to initiate eating spontaneously or to increase eating in response to food deprivation; thus, these functions require the forebrain.

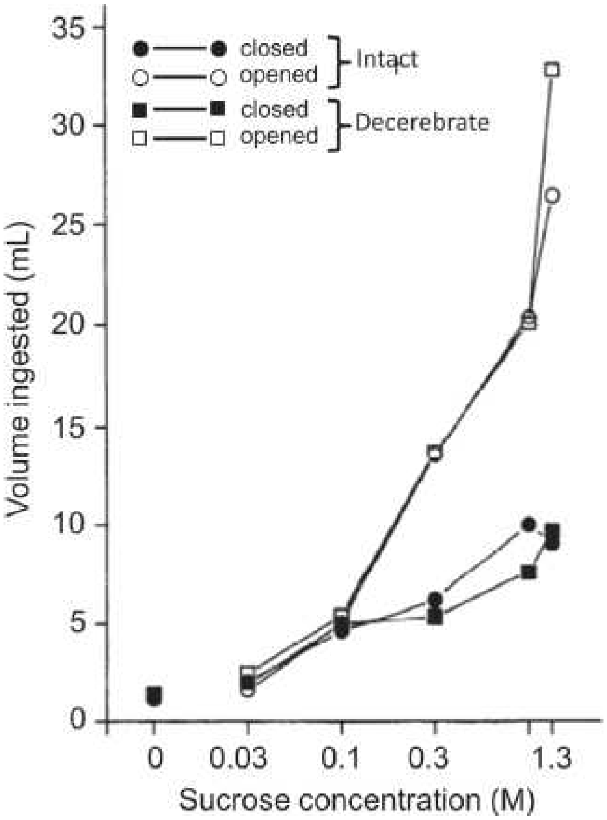

FIGURE 1.17–9. Eating behavior in intact and chronic decerebrate rats. Various concentrations of sucrose were delivered by intraoral catheters so that rats could either ingest it or passively reject it by allowing it to drip from the mouth. Intake was an increasing function of sucrose concentration in decerebrate rats that fed normally (closed condition). The gain of this function was dramatically increased in sham-feeding rats in which postingestive controls of meal size were minimized by opening a gastric cannula (opened condition). Note the close correspondence, in intact and decerebrate rats, of the stimulatory effect on eating of increasing sucrose concentration and the interaction of this stimulatory effect with the inhibitory effect of normal postingestive food stimuli. (Reprinted with permission from Grill HJ, Kaplan JM. Sham feeding in intact and chronic decerebrate rats. *Am J Physiol.* 1992;262(6):R1070. doi:10.1152/ajpregu.1992.262.6.R1070.)

Ascending Projections. Several ascending neural projections originate in the brainstem. Each contributes to a range of arousal, affective, and motivational functions related to eating.

A2 NORADRENERGIC NEURONS. A2 noradrenergic neurons have cell bodies in a region of the caudal medulla containing the NTS. They receive multiple inputs and project directly to the hypothalamus, cerebral cortex, and many other sites throughout the brain. They contribute to functions including vagovagal reflexes, satiation (including CCK satiation), and several aspects of affect and learning, especially related to stress or illness. Although noradrenergic neurons originating in the locus coeruleus (A4 and A6 cell groups) receive more attention in the study of stress, stress-related changes in eating are more likely to involve A2 noradrenergic neurons because these provide greater inputs to the paraventricular hypothalamic nucleus (PVN) of the hypothalamus, central nucleus of the amygdala (CeA) and NAc, and other sites mediating eating.

DOPAMINE. The VTA contains dopaminergic neurons that project via the mesocorticolimbic pathways to numerous hypothalamic and telencephalic targets as well as to the NTS and other caudal sites. Each of these areas is also reciprocally connected to the VTA. This network is important in food reward, as discussed further below, as well as having numerous other affective and cognitive functions. Importantly, dopamine in different brain areas can have very different influences on eating. For example, dopamine neurotransmission in the NAc stimulates eating, whereas dopamine in the perifornical hypothalamus inhibits eating (a phenomenon potentially explaining the clinical finding that neuroleptics that antagonize dopamine increase body weight, apparently by increasing food intake).

SEROTONIN. Serotonin (5-hydroxytryptamine, 5HT) neurons in the midbrain raphe nuclei give rise to another diffuse ascending projection system that is involved in affect and motivation, including the control of eating. For example, 5HT signaling via the 2C receptor has been implicated in the eating-inhibitory effects of CCK, GLP-1, leptin, and the bacterial toxin LPS. Some of these effects may involve melanocortin signaling in the ARC (discussed below).

Hypothalamus

The importance of the hypothalamus for the neural control of eating has been recognized since the classic descriptions by John Brobeck and his colleagues in the mid-20th C of the syndromes of hyperphagia and aphagia that result from lesions of the mediobasal (VMH and ARC) and lateral hypothalamus (LHA), respectively. Originally understood as functional centers, these and a number of other hypothalamic areas are now conceptualized as nodes in a distributed neural network. Figure 1.17–10 shows a network element based on two distinct neuronal populations in the ARC. One population expresses a precursor for α-melanocyte-stimulating hormone (αMSH). The second population expresses NPY, agouti-related peptide (AgRP), and GABA. Both sets of neurons express receptors to ghrelin and leptin (as well as receiving a variety of additional inputs), with the hormones' electrophysiologic effects opposite in sign within and between the two sets of ARC neurons. Studies in animals indicate that stimulation of AgRP neurons stimulates negative-reinforcement hunger, consistent with classic negative-feedback models of the homeostatic regulation of body weight. More recent studies, however, indicate that AgRP neuronal activity is reduced when mice view food before they consume it, suggestive of anticipatory, positive-reinforcement hunger. Current research generating gene-expression profiles of individual neurons so as to identify distinct subsets of neurons in these areas, will enable further insights into the organization and functions of the ARC and other hypothalamic and extrahypothalamic areas.

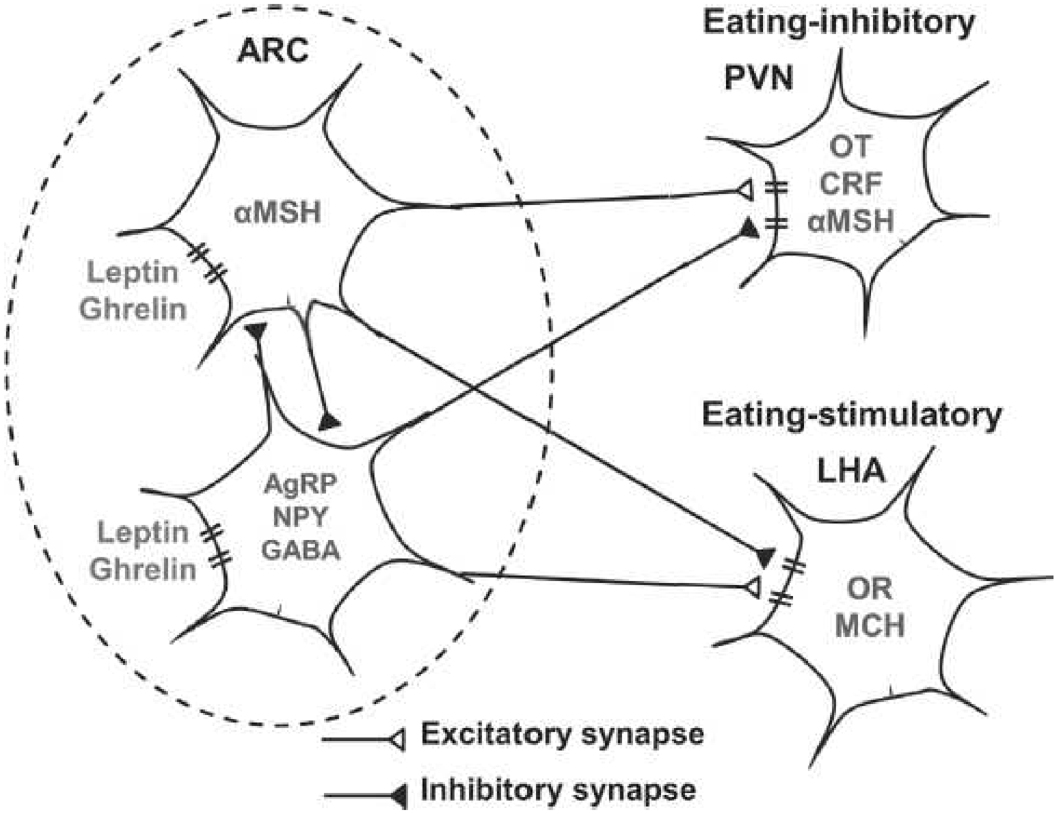

FIGURE 1.17–10. Schematic diagram of part of the hypothalamic component of the neural network controlling eating. Two populations of neurons in the hypothalamic arcuate nucleus (ARC) receive inputs from the hormones leptin and ghrelin as well as other humoral signals and neural afferents (not shown). One of these populations expresses α-melanocyte-stimulating hormone (αMSH) as a neurotransmitter; these neurons are stimulated by leptin and inhibited by ghrelin. The other population expresses agouti-related peptide (AgRP), neuropeptide Y (NPY), and γ-aminobutyric acid (GABA) as neurotransmitters; these neurons are inhibited by leptin and stimulated by ghrelin. The two populations of ARC neurons have mutually inhibitory connections. Among their principal projection targets are neurons in the paraventricular nucleus of the hypothalamus (PVN) that express oxytocin (OT), corticotropin-releasing factor (CRF), and αMSH as neurotransmitters and neurons in the lateral hypothalamic area (LHA) that express orexin (OR) and melanin-concentrating hormone (MCH) as neurotransmitters. αMSH neurons stimulate PVN neurons, and AgRP/NPY/GABA neurons inhibit them; in contrast, αMSH neurons inhibit LHA neurons, and AgRP/NPY/GABA neurons stimulate them. If the net balance of activity in this network is to stimulate the LHA neurons, eating is stimulated (thus, AgRP, NPY, OR, and MCH are often called orexigenic peptides; GABA is too widely expressed to have such a simplistic designation). If the net balance of activity in this network is to stimulate the PVN neurons, eating is inhibited (thus, αMSH, OT, and CRF are called anorexigenic peptides). Both these LHA and PVN neurons project to many sites in the brain to control eating.

Both αMSH and AgRP/NPY/GABA ARC neurons project to the PVN and the LHA, with electrophysiologically opposite effects in these sites. This suggests the model that if the net result of neural activity in this circuit is to stimulate PVN neurons and inhibit LHA neurons, then eating is inhibited; conversely, if the net effect is to stimulate LHA neurons and inhibit PVN neurons, then eating is stimulated. Interestingly, the effects of αMSH and AgRP are primarily mediated through their opposite effects on a single receptor, the melanocortin-4 receptor (MC4R). αMSH stimulates the receptor, and AgRP is an inverse agonist, that is, a functional antagonist. Loss-of-function mutations in *MC4R* or *POMC,* the gene expressing αMSH, are the commonest form of monogenic human obesity, occurring in 1% to 2% of patients with obesity. A melanocortin agonist, setmelanotide (Imcivree, Rhythm Pharmaceuticals), was recently approved for the treatment of obesity due to loss-of-function mutations in *POMC*, the leptin-receptor gene *LEPR,* or *PCSK1*, which encodes a serine endoprotease involved in neuropeptide synthesis.

Descending Projections. Forebrain eating-control signals are relayed to the integratory networks in the caudal brainstem rather than directly to the motor networks for eating. For example, (1) leptin administration in intact rats increased the satiating potency of both gastric loads and CCK; (2) leptin increased the neuronal activation produced by gastric loads and CCK in the NTS; (3) rats bearing mutations of the leptin receptor had reduced sensitivity to exogenous CCK; and (4) transgenic restoration of the leptin receptor function to these rats' ARC restored CCK's potency to inhibit eating and activate the NTS. Some of the neural pathways mediating these effects have been identified. As shown in Figure 1.17–11, ARC leptin-sensitive αMSH neurons project to oxytocin neurons, the PVN, which then project to NTS neurons that also mediate CCK satiation.

Telencephalon

The best-understood telencephalic contribution to appetite is the *telencephalic reward network,* which comprises a number of subcortical areas (e.g., NAc, ventral pallidum [VP]) and cortical areas (e.g., limbic, orbitofrontal, anterior cingulate, opercular, and insular cortex). Most of these structures receive dopaminergic, noradrenergic, and serotonergic inputs. In addition, opioid, cannabinoid, acetylcholine, orexin, benzodiazepine, glutamate, and GABA neurotransmission has been implicated in flavor reward. In addition to its roles in hedonics and reinforcement related to eating, this network is involved in other natural rewards, such as sex and in drug rewards.

Subcortical Reward Nodes. The NAc and VP have been extensively analyzed in rodents. Each has connections with the other and has numerous connections with more rostrally and caudally located structures. Dopamine is released in the NAc in a dose-dependent fashion as rats sham feed sucrose or liquid fat (Fig. 1.17–12), and local administration of dopamine-receptor antagonists into the NAc reduces sham feeding of sucrose solutions, indicating that NAc dopamine neurotransmission plays an important role in flavor reward. This is further supported by anatomical studies indicating that all the sensory modalities involved in flavor, that is, gustatory, olfactory, and somatosensory converge in the NAc. Furthermore, these projections are relatively direct; that is, they involve a minimal number of synaptic connections.

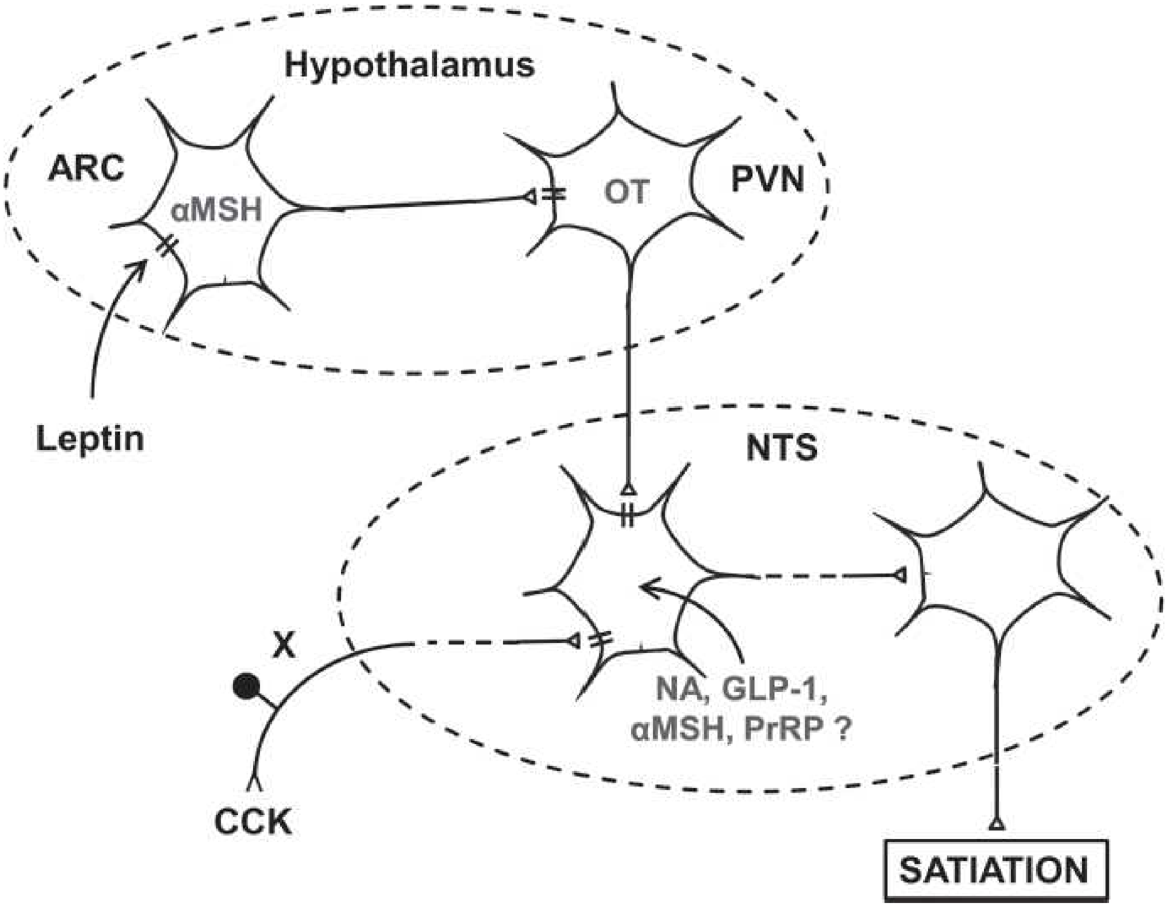

FIGURE 1.17–11. Schematic depiction of the neural network through which plasma leptin levels modulate the satiating potency of cholecystokinin (CCK). Increases in plasma leptin levels stimulate neurons in the hypothalamic arcuate nucleus (ARC) that express α-melanocyte-stimulating hormone (αMSH), and decreases in plasma leptin levels inhibit them (this appears to be the more physiologically relevant effect). ARC αMSH neurons project to neurons in the paraventricular nucleus of the hypothalamus (PVN) that express oxytocin (OT). These, in turn, project to neurons in the nucleus of the solitary tract (NTS) in the brainstem that also receive direct or indirect inputs from vagal afferent neurons (X) that are stimulated by CCK. Increases or decreases in the activity of this network lead to increases or decreases in satiation and meal size. The NTS network mediating CCK satiation appears to involve noradrenergic (NA), glucagon-like peptide-1 (GLP-1), αMSH, and prolactin-releasing peptide (PrRP) neurotransmission. Note: *broken lines* on the schematic axons indicate polysynaptic connections.

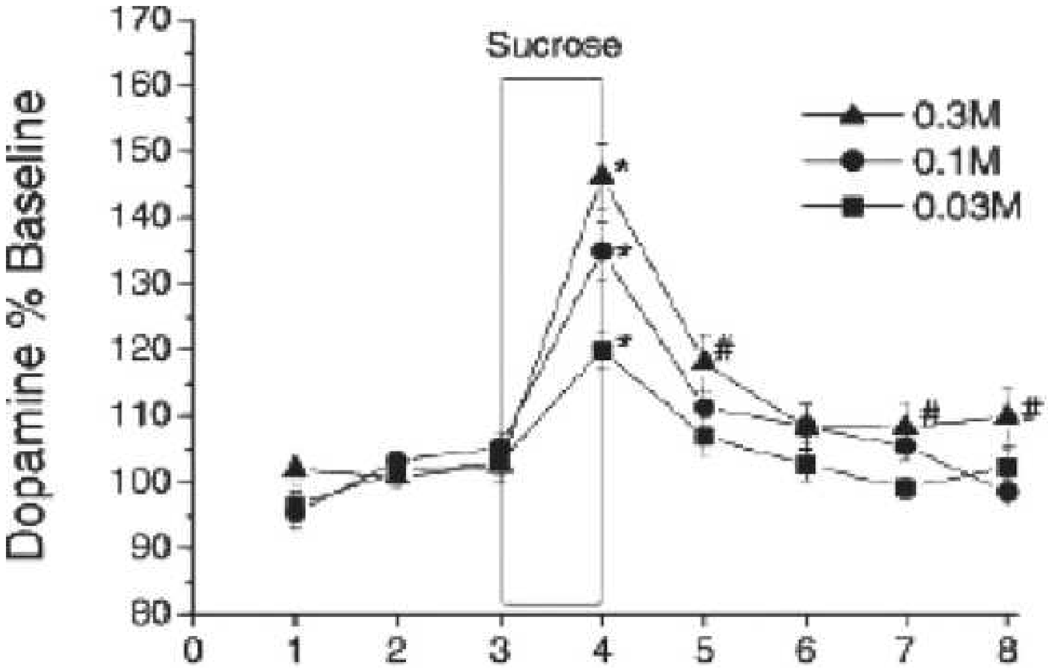

FIGURE 1.17–12. Evidence for the role of dopaminergic neurotransmission in the nucleus accumbens (NAc) in sweet-taste reward. Different molar sucrose concentrations were presented for 20 minutes per day to *ad libitum*-fed rats, which sham fed with open gastric cannulas to minimize postingestive negative feedbacks. Rats ingested increasing amounts of more concentrated sucrose (means of 18, 31, and 43 mL of 0.03, 0.1, and 0.1 M sucrose, respectively), and simultaneous extracellular microdialysis revealed a graded increase in the release of dopamine in the NAc during ingestion. Control tests (not shown) revealed that similar graded dopamine release occurred when the volume of sucrose ingested was held constant. Data are shown for 20-minute periods as mean ± standard error of the mean of baseline. *$P < .01$; #$P < .05$ versus baseline. (Reprinted with permission from Hajnal A, Smith GP, Norgren R. Oral sucrose stimulation increases accumbens dopamine in the rat. *Am J Physiol.* 2004;286(1):R31. doi:10.1152/ajpregu.00282.2003.)

A dense projection from the NAc targets the VP, a subdivision of the substantia innominata. Opioid neurotransmission via μ-opioid receptors appears to mediate reward in this area. In a fascinating rat study, lesion of the caudal VP caused rats to display aversive facial reactions to sucrose rather than positive reactions, suggesting that neural processing in this area is necessary for the generation of positive flavor hedonics.

Cortical Reward Nodes. Several areas of the frontal and temporal lobes of the cortex have been implicated in flavor reward. These include both evolutionarily older areas (e.g., the hippocampus and cingulate cortex) and evolutionarily newer or neocortical areas (e.g., orbitofrontal cortex and prefrontal cortex). Functional imaging studies in human subjects as well as single-neuron recording studies in monkeys indicate that flavor hedonics is mirrored by the activity of the midanterior orbitofrontal cortex. For example, orbitofrontal cortical activity is much higher if palatable foods are sampled when hungry than when satiated. In contrast, the medial orbitofrontal, ventromedial prefrontal, and middle anterior insula cortices may be less involved in generating the experience of flavor hedonics than in monitoring and predicting flavor rewards.

PHYSIOLOGIC MODULATORS OF APPETITE

Learning

Meal initiation, food selection, satiation, and meal size are all readily conditionable in animals and humans. Both classical (or Pavlovian) and instrumental (or operant) learning are involved. Environmental context and flavor usually provide the CS. For example, when a sound/light CS was presented to rats before each of six scheduled meals for several days and then tested during “extinction,” that is, when the rats had free access to the same diet, the CS elicited initiation of a very large meal on each daily presentation for 3 weeks. Thus, cues that predict food availability during food deprivation can provoke the initiation of a large meal in the absence of deprivation. The unconditioned stimuli (UCS) for such conditioned hunger have not been identified. Satiation and food selection are also conditionable, as shown by the extinction of conditioned controls of meal size during sham feeding (Fig. 1.17–13). In this case, the UCS is the postingestive consequences of eating. For example, rats that drink flavored water while nutrients are infused into the stomach, duodenum, jejunum, or liver subsequently eat less food with that flavor (flavor-nutrient conditioning). This is satiation, not aversion because when a choice of foods is offered, the rats choose the flavor associated with nutrient infusions. Studies in humans and rats indicate that food selection and amount eaten are often driven more by flavor-nutrient conditioning than by flavor hedonics. Recent work indicates that in the case of infused fats, vagal afferents relay the UCS signals centrally, leading to dopamine release in the striatum.

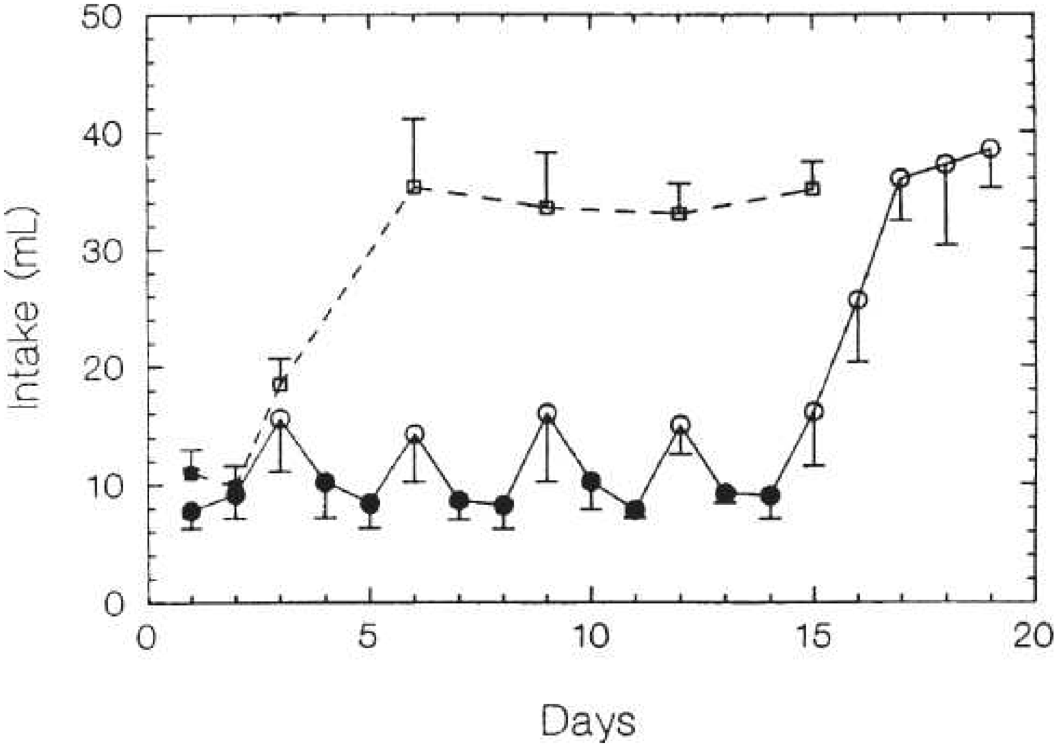

FIGURE 1.17–13. The role of conditioned satiation in eating in the rat. Data are 30-minute intakes of 0.8 M sucrose in two groups of rats in successive tests. Rats were prepared with steel gastric cannulas that were either closed for real-feeding tests (closed symbols) or open for sham-feeding tests (open symbols), during which sucrose did not accumulate in the stomach or enter the intestines. One group (*squares*) real-fed sucrose once and then sham fed. The other group (*circles*) alternated real- and sham-feeding tests. The progressive increase in sucrose intake during consecutive sham-feeding tests was prevented by interspersing two normal feeding tests between each sham-feeding test. When this was done, sham intake was significantly larger than real intake but significantly smaller than the asymptotic sham intake. These data indicate that normal feeding of sucrose produces a learned association between the taste of the sucrose and its postingestive effects, that this association limits intake, and that it extinguishes during consecutive sham-feeding tests. (Reprinted with permission from Davis JD, Smith GP. Learning to sham feed: Behavioral adjustments to loss of physiological postingestional stimuli. *Am J Physiol.* 1990;259(6):R1228. doi:10.1152/ajpregu.1990.259.6.R1228.)

Higher-order conditioning is likely to play important roles in human eating. For example, cultural socialization may result in the apparent positive hedonic response to and selection of the flavor of capsaicin (chili), which all infants avoid. Indeed, with the exception of a few unconditioned gustatory preferences and aversions (such as for sweet and bitter tastes), *all* food selection appears to be learned. Some forms of learned controls of eating are special forms of conditioning. Flavor aversions conditioned by upper GI food poisoning, for example, can be learned after only one CS–UCS pairing and, despite extraordinarily long CS–UCS delays (hours), are unusually long lasting.

Basic and clinical research on learned food preferences should be a high priority. For example, behavior therapy programs, because they are relatively highly structured and can present specific food stimuli, should provide excellent opportunities for the clinical use of conditioned physiologic controls of eating.

Interactions

This review emphasizes that different functional processes controlling eating and their physiologic bases interact in ways indicating that boundaries among them are merely heuristic. Fundamental learned interactions between flavor stimuli and GI or metabolic controls of eating were described in the previous section. There is also an unconditioned change in flavor hedonics related to physiologic state, such that flavor stimuli are rated as more pleasant by hungry than by full subjects ("alliesthesia"). There are also interactions between GI and adiposity signals. As described above, CCK satiation was reduced in rats with loss-of-function mutations in leptin receptors and was rescued by transgenic restoration of leptin receptors in the ARC. Similar interactions have been described between adiposity signaling and GLP-1 or other controllers of meal size.

Energy Expenditure

Resting Metabolic Rate (RMR). RMR, which is determined mainly by fat-free mass, is linearly correlated with daily energy intake, accounting for ~50% of the variance, and with meal size. Fat mass is not strongly correlated with daily energy intake. These data suggest that RMR tonically stimulates appetite. Teleologically, it seems sensible that organisms ensure sufficient energy to maintain the function of the brain, liver, kidneys, etc. The mechanisms involved are unknown. Note that RMR is unlikely to be a regulated variable; that is, although RMR drives appetite, there is no feedback loop maintaining RMR at a particular level.

Exercise. Energy intake also is related to habitual levels of exercise or physical-activity energy expenditure, but whether acute bouts of exercise affect eating is not clear. Recent research has delineated interesting relationships between exercise and food intake in an animal model of obesity, Otsuka Long Evans Tokushima Fatty (OTELF) rats, which lack CCK1 receptors and are hyperphagic and obese. When given access to running wheels, they are equally active to the control strain but they reduce their food intake within a day down to levels of the control rats and maintain this lower intake for as long as running wheel access is maintained. This results in the normalization of body weight. When they no longer have access to running wheels, food intake increases. Importantly, the magnitude and extent of the increase depend upon the developmental stage during which exercise is experienced. In mature rats, intake and body weight return to pre-exercise levels. However, when access to exercise is provided to young OLETF rats, the effects appear to be permanent. Food intake temporarily increases but not to levels in OLETF rats that did not have running wheel access and body weights are maintained at levels significantly below those of rats that did not have access to running wheels. Similar effects of exercise have been demonstrated for rats bred to be susceptible to dietary obesity and in mice lacking melanocortin-4 receptors. The mechanisms underlying such actions have yet to be determined.

Other recent work in rats has also demonstrated the effects of exercise on diet selection. Access to a running wheel reduced the selection of a previously preferred high-fat diet and can produce avoidance of a novel, normally preferred palatable diet.

Reproductive Axis

Ovarian cycling and other reproductive states affect eating. Adult women and female animals of many mammalian species decrease eating during the follicular and periovulatory phases of the cycle. In women, this decrease, which is not consciously appreciated, amounts to a few 100 kilocalories per day, which over the long term can affect body weight (Fig. 1.17–14). Studies of oophorectomy and estrogen treatment in rats and mice indicate that estrogens acting on estrogen receptor-1 (or α) in the medial NTS mediate these effects. Estrogen levels are also high in the midluteal phase. It is thought that progestins counteract the eating-inhibitory effects of estrogens and maintain higher levels of food intake during the midluteal phase of the cycle.

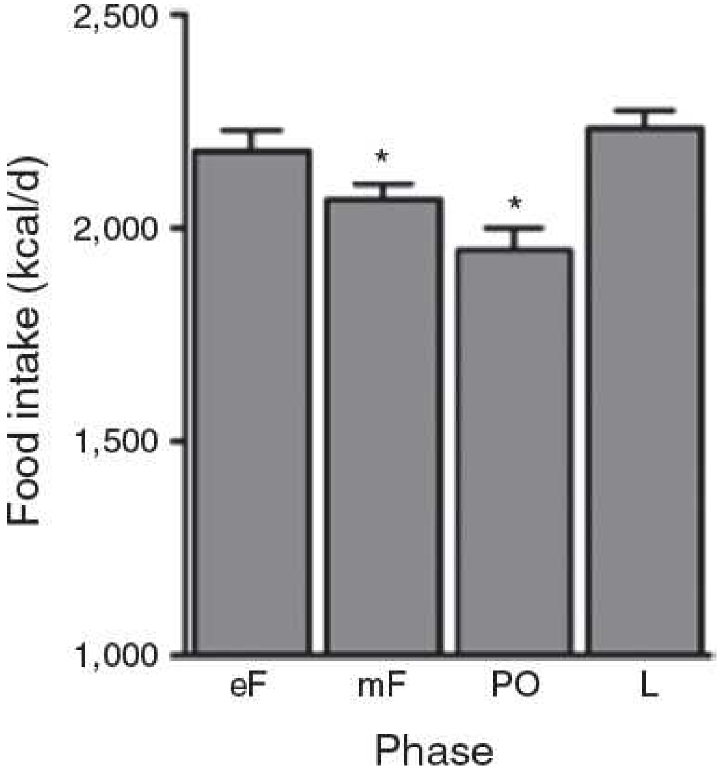

FIGURE 1.17–14. Daily food intake during the ovarian cycle in women averaged across the early-follicular (eF; 4 days), mid-follicular (mF; about 9 days), periovulatory (PO; 4 days), and luteal (L; about 11 days) phases. Note the progressive decrease in food intake during the follicular phase. Old-world monkeys, whose cycle is apparently identical to the human cycle, show a similar follicular decrease in daily food intake and also a high, constant level of food intake during most of the luteal phase which has not been verified in women. *Significantly different from luteal phase. (From Asarian L, Geary N. Sex differences in the physiology of eating. *Am J Physiol Regul Integr Comp Physiol*. 2013;305:R1215–R1267.)

In rats and mice, the periovulatory decrease and the postoophorectomy increase in food intake are caused by selective changes in meal size; meal frequency does not contribute. Indeed, when the hyperphagia of oophorectomized rats abates after they have increased body weight ~20%, it is because meal frequency decreases; the oophorectomy-induced increase in meal size is permanent. This is a good example of the importance of meal pattern analysis in understanding the controls of eating. Estrogens decrease meal size during the periovulatory phase in rats at least in part by selectively increasing the satiating potency of CCK.

Both human and animal research indicates that there is a sex difference in the avidity for sweets, with females preferring them more than men. One reflection of this is that women more frequently endorse sweets as their most preferred or craved foods, whereas men more frequently endorse savory foods. The mechanisms underlying this sex difference are unknown.

Reproductive axis function also seems to affect binge eating, as described below.

Illness Anorexia

Anorexia of varying intensity and duration is a common element of both innate and acquired immune responses to infection, trauma, neoplasm, and other challenges. The transient anorexia of the acute phase response of the innate immune system, like fever, is an unpleasant but adaptive response that can facilitate recovery. More chronic illness anorexia, however, is a maladaptive response that can increase illness severity and interfere with therapy.

A number of cytokines, including interleukin-1 and TNF-α, as well as prostaglandin-E_2 and other immune signaling molecules are involved in illness anorexia. Peripheral immune signals apparently affect the brain both directly and indirectly via endocrine and peripheral neural responses and converge on the same neural networks that mediate normal eating. For example, serotonergic neurons in the dorsal raphe nuclei (DRN), which project into the hypothalamus and telencephalon, have been implicated in both the normal control of eating and illness anorexia. In animals, illness can lead to reduced meal size, reduced meal number, or both. Illness anorexia in humans remains poorly understood.

PSYCHIATRIC EATING DISORDERS

Laboratory Studies in Patients with Eating Disorders

Subjective Appetite in Patients with Anorexia Nervosa. The study of disordered subjective appetite around meals in patients with eating disorders provides important insights into disordered eating. Patients with anorexia nervosa display marked individual differences in subjective appetite. It seems that physiologically rooted perceptions of appetite are often outweighed by psychological influences related to body image, fear of losing control of eating, conditioned food preferences, etc. Figure 1.17–15 shows an example. Disentangling these disparate factors remains challenging.

Eating in Patients with Bulimia. Binge eating is accessible to laboratory study. In a residential-laboratory setting in which subjects had free access to a variety of foods, patients with bulimia nervosa and control subjects usually took similar numbers of meals and, usually, similarly-sized meals (40 to 980 kcal). But about a quarter of the meals taken by the patients were much larger (1,058 to 6,728 kcal). These meals occurred mainly late in the day and consisted largely of sweet and snack foods. Remarkably, experimentally manipulated cognitive stimuli, that is, an invitation to binge if possible, are sufficient to induce binges in the laboratory in patients with bulimia that exhibit spontaneous binges. This opens the possibility of within-individual studies of the behavioral and neural differences between normal and binge meals.

Postingestive negative-feedback satiation signals are less potent in patients with bulimia. (1) Equivalent preloads of food decrease intake less in patients with bulimia nervosa than in controls, particularly when the patients are eating large meals. (2) Patients with bulimia must eat larger amounts of food to produce equivalent self-reports of fullness during a meal. (3) Volume distention of the stomach produces a decreased perceptual and mechanical response in patients with bulimia, presumably because their stomachs are larger than normal as the result of accommodation to the frequent ingestion of large meals. (4) Meal-stimulated CCK release is less in bulimic patients than that in controls (Fig. 1.17–16). These abnormalities appear to resolve as binging decreases, suggesting that they are not the initial cause of bulimia. Nevertheless, it is possible that they facilitate the development of the disorder once it has begun and impede recovery from it.

Binge eating is also associated with ovarian cycling. Binge frequency is maximal during the mid- and lateluteal phases of the cycle in women with bulimia nervosa and least during the follicular phase. Binge frequency and emotional eating (the tendency to overeat or binge eat in response to negative emotions) show similar cyclic effects in community samples with greater tendencies for disordered eating associated with high levels of both estradiol and progesterone and lessor tendencies associated with high levels of estradiol and minimal levels of progesterone.

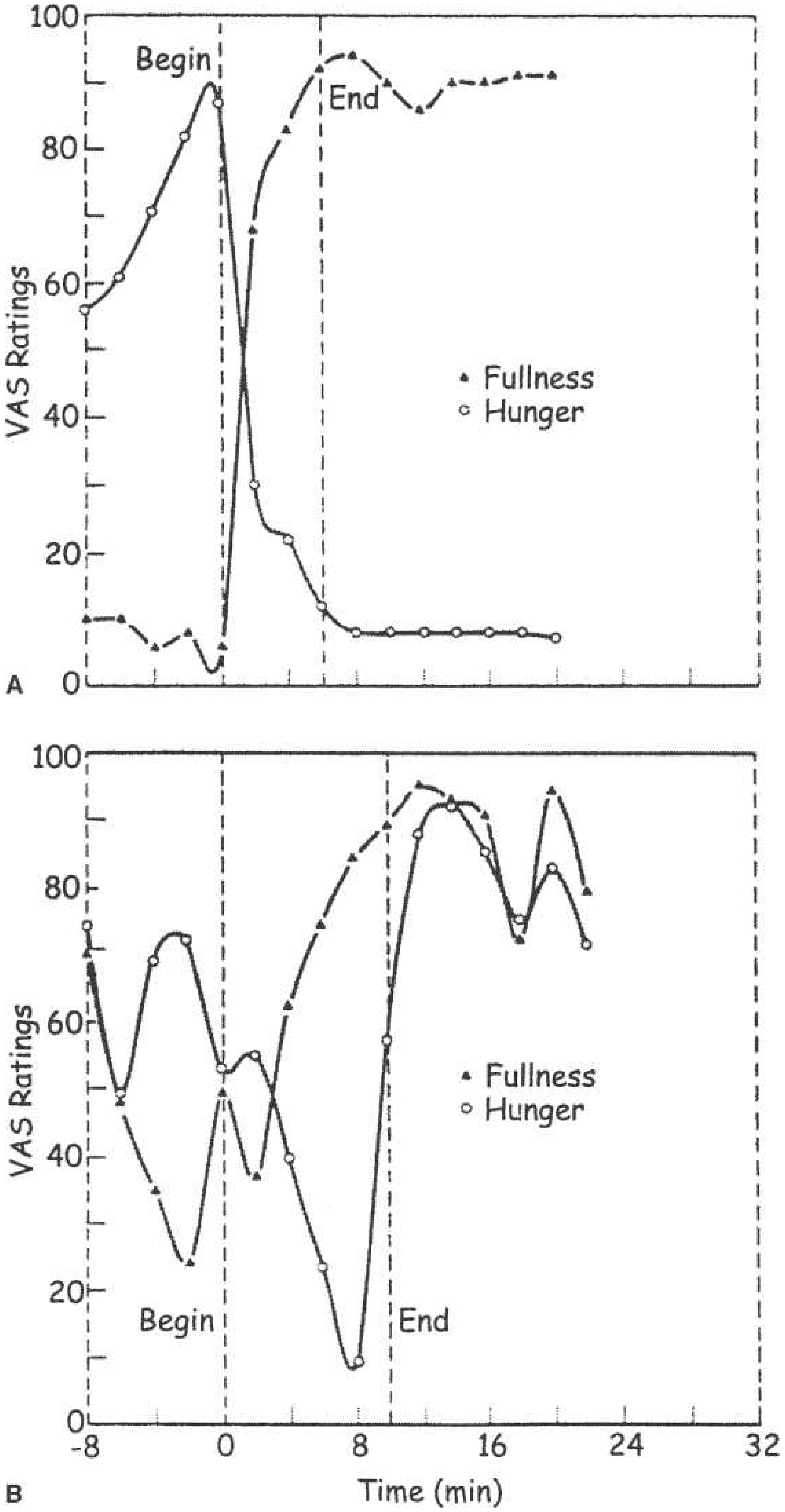

FIGURE 1.17–15. Disturbed subjective appetite in a woman with anorexia nervosa. Momentary intensities of hunger and fullness around a meal were rated by (**A**) a healthy woman and (**B**) a woman with anorexia nervosa using

visual analog scales. In the healthy woman, hunger is high before the meal, fullness is low before the meal, and the two percepts change in a reciprocal fashion before, during, and after eating; this is typical of most healthy normal subjects. In contrast, in this patient with anorexia nervosa, hunger and fullness change irregularly before, during, and after the meal and are often not reciprocally related; other patients with anorexia nervosa showed a variety of responses, some more and some less like the usual responses of healthy women than these. Note that the figure does not show other experiences associated with meals, such as pleasure and tranquilization. The challenge for this research is to disentangle the various psychological and physiologic influences producing these percepts. (Reprinted from Owen WP, Halmi KA, Gibbs J, Smith GP. Satiety responses in eating disorders. *J Psychiatr Res.* 1985;19(2–3):279–284. Copyright © 1985, with permission from Elsevier.)

Serotonin and Eating Disorders. Functional brain imaging, often involving radiolabeling methods to identify neurotransmitter function, is a major research focus in eating disorders. The most progress to date involves the study of serotonergic mechanisms. (1) Women with bulimia nervosa displayed reductions in $5HT_{2A}$ receptor binding in the orbital frontal cortex following recovery, suggesting that this might reflect a trait difference underlying vulnerability for this disorder. (2) Women who recovered from anorexia nervosa also displayed reduced $5HT_{2A}$ receptor binding, albeit in different brain areas.

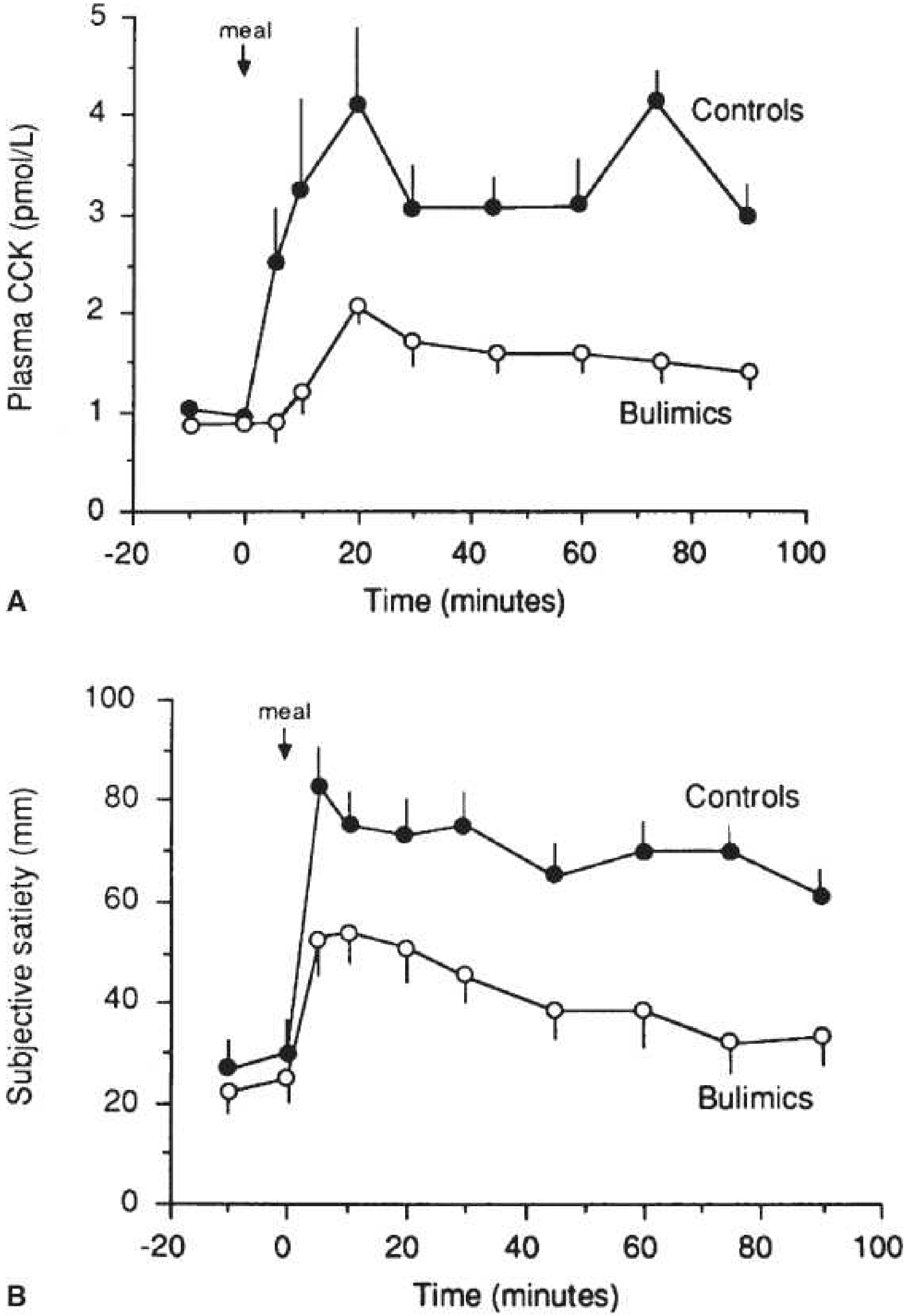

FIGURE 1.17–16. Prandial concentrations of cholecystokinin (CCK) (**A**) and the subjective experience of postprandial satiety (**B**) are reduced in patients with bulimia nervosa. Fourteen patients and 10 control women matched for age and weight were offered a 400-mL liquid meal after an overnight fast (*arrows*) and ate it in 1 to 2 minutes. Plasma CCK was measured with a

selective bioassay. Satiety was measured by a 100-mm visual analog scale (0 = empty; 100 = full). Both the peak CCK concentration and the integrated CCK response (area under the curve) were significantly reduced in patients with bulimia nervosa, and this correlated with reports of significantly less satiety beginning 5 minutes after meal onset. (From Geracioti TD, Jr., Liddle RA. Impaired cholecystokinin secretion in bulimia nervosa. *N Engl J Med*. 1988;319:683–688, with permission.)

A key challenge in research of this type is to identify the causal factors contributing to vulnerability. For example, serotonergic changes might be related to anxiety disorders, which occur in the majority of patients with anorexia nervosa and bulimia nervosa, rather than being specific to the disordered control of eating. Unfortunately, relatively few brain-imaging studies of eating disorders in patients occur in the context of eating. More work of this type might reveal state or trait differences related to the physiology of appetite between eating disorders and suggest new targets for improved treatment.

The acute tryptophan depletion (ATD) test has been used to investigate the role of serotonin in eating disorder behaviors. The test involves ingestion of a tryptophan-free protein load, which diverts plasma tryptophan into peripheral protein synthesis, leading to a decrease in central serotonin synthesis. One study indicated that ATD decreased mood in both patients with bulimia nervosa and healthy controls and increased the urge to binge in patients who were medicated with serotonin-reuptake inhibitors (SSRI) at the time of testing but not in patients who were not so medicated and not in controls. These data are consistent with a specific role of serotonin in acute binge urges. Further research is warranted to determine why ATD affected binging only in the medicated group (perhaps that group had a more severe disorder or perhaps ATD reversed the beneficial effects of SSRI treatment but that serotonin is not involved in the binge disorder itself), where in the brain the critical effect on serotoninergic neurotransmission occurred, and what functional system it influenced.

Animal Models of Psychiatric Eating Disorders

Anorexia Nervosa. Activity-based anorexia (ABA) is a widely used model of anorexia nervosa. In this model, rats or mice are given access to running wheels for most of the day and access to food for only 1 to 2 hours per day. Over the course of several days, animals run progressively more and eat progressively less, to the point of self-starvation. In some, but not all studies, signs of anxiety and anhedonia occur. ABA also elicits many endocrine and neurotransmitter changes similar to those associated with anorexia nervosa. Additionally, as in anorexia nervosa, adolescence is a period of increased vulnerability. Individual differences also play a role in ABA. For example, in a recent study of adult female C57BL/6N mice, only about half the animals displayed progressive weight loss, while the others maintained their initial weights. When mice with transgenic knockdown of the dopamine transporter and increased dopamine levels were tested, however, more mice were vulnerable due to exaggerated increases in wheel-running activity.

Binge Eating. Binge-like eating can be produced in rats and mice by a variety of methods. Restricted access to a highly preferred food (usually 1 to 2 h/day for 3 day/wk) increases intake in comparison with 1 to 2 hours of access every day. Intake of preferred foods is also increased if the food is offered 1 to 2 h/day following electric shock or other physical stressors. Psychogenic stress, produced by placing the preferred food just outside the cage, also suffices to elicit binge-like eating.

Several studies have addressed the influences of the reproductive axis on binge-like eating. Estradiol treatment reduced binge-like eating in oophorectomized rats, similar to the effect found in women, but, unlike in women, the effect did not vary cyclically and was not reversed by progesterone. Female rats are markedly more vulnerable to displaying binge-like eating than are male rats (Table 1.17–4). The development of this “binge-eating proneness” trait was pursued to follow up on a report that girls with lower levels of estrogen during puberty have increased heritability of binge eating. Rats that were prepubertally oophorectomized failed to show normal pubertal development and displayed an increased risk of binge-like eating as adults (Table 1.17–4). These data, consistent with the human data, suggest that normal pubertal estrogens levels are protective against the development of binge eating (note that this apparent organizational effect of estrogens contrasts with the activational or acute effect of estrogens to reduce binge eating in women and binge-like eating in animals). Finally, another recent study indicated that estrogens act on serotonergic neurons in the DRN to inhibit binge-like eating in adult female mice.

Table 1.17–4.
Influences of Sex and Pubertal Development on the Propensity for Binge-Like Eating in Rats

A. Sex	**BER**	**BEP**
Female (%)	3*	47*
Male (%)	37	3
B. Prepubertal oophorectomy	**BER**	**BEP**
Likelihood of BEP (odds ratio)	0.6	2.4*
	(0.3–1.3)	(1.0–5.7)

A. Male and female rats were offered standard chow *ad libitum* and a palatable commercial cake frosting for 4 h/day, 3 day/wk for 2 weeks, a procedure that leads to increased palatable food intake in comparison to *ad libitum* access to the same food. Palatable-food intakes in six tests were ranked across all rats, and propensity for binge-like eating was scored as binge-eating–prone phenotype (BEP; scored in the highest tertile of palatable-food intake in three or more of the six tests and never scored in the lowest tertile) or binge-eating-resistant phenotype (BER: scored in the lowest tertile of palatable food intake in three or more of the six tests and never scored in the lowest tertile). Females were significantly more likely to develop the BEP phenotype than males (*) and significantly less likely to develop the BER phenotype than males (*). Data are percent of rats of each phenotype (rats with intermediate scores are not shown). (Adapted from Klump KL, Racine S, Hildebrandt B, Sisk CL. Sex differences in binge eating patterns in male and female adult rats. *Int J Eat Dis.* 2013;46(7):729, with permission.) **B.** Prepubertal female rats were oophorectomized and phenotyped as described above as adults. Oophorectomy prevented normal pubertal development, as assessed by vaginal opening, and significantly increased the odds of developing the BEP phenotype (*). Data are odds ratios (95% confidence intervals) of the likelihood of developing BER or BEP phenotype following oophorectomy relative to non-oophorectomized rats. (From Klump KL, Sinclair EB, Hildebrandt BA, et al. The disruptive effects of estrogen removal before puberty on risk for binge eating in female rats. *Clin Psychol Sci*. 2020;8:839–856, with permission.)

SUMMARY

The behavioral and subjective aspects of eating are generated by a widely distributed neural network extending from the caudal brainstem to the frontal lobes. The fundamental output of this network, and therefore the focus of behavioral neuroscience of appetite is the meal. Physiologic stimuli for meal initiation include circadian rhythms, metabolic signals, and endocrine signals, of which ghrelin has received the most attention.

Conditioned signals, however, are usually more important. After eating begins, flavor hedonics increase or decrease eating. Other than the preference for sweet taste, these are mainly conditioned with a variety of mechanisms ranging from postingestive nutrient sensing to social and cultural factors, contributing. Meal-ending satiation is controlled by physiologic signals involving gastric mechanoreception and gastrointestinal hormones, such as CCK and GLP-1, as well as by conditioned influences. Some of these mechanisms maintain satiety after meals. A great deal of progress has been made in delineating nodes in the underlying neural network in the hypothalamus, mainly in the context of energy homeostasis and the ventral forebrain, mainly in the context of food reward. AgRP, NPY, GABA and αMSH are important mediators of the hypothalamic contributions to energy homeostasis. Reduced leptin in underweight appears to activate this system to stimulate appetite, but the extent to which overweight engages a regulatory compensatory influence remains in doubt. A number of subcortical (NAc, VP, etc.) and cortical sites (hippocampus and cingulate, insular, orbitofrontal, and prefrontal cortices) contribute importantly to hunger, satiety and, especially, food reward. Ascending projections from the brainstem and midbrain α-adrenergic, serotoninergic, and dopaminergic neurons provide important inputs to most of these areas. A general principle is that the functional divisions of appetite are increasingly understood to result from interactions among all of the physiologic mechanisms mentioned. Subclinical disordered eating and psychiatric eating disorders are accessible to direct experimental study as well as to study in animal models.

FURTHER READINGS

Asarian L, Geary N. Sex differences in the physiology of eating. *Am J Physiol Regul Integr Comp Physiol*. 2013;305:R1215–R1267.

Asarian L, Gloy V, Geary N. Homeostasis. In: Ramachandran VS, ed. *The Encyclopedia of Human Behavior*, Vol. 2. Academic Press; 2012.

Asarian L, Langhans W. A new look on brain mechanisms of acute illness anorexia. *Physiol Behav*. 2010;100:464–471.

Beeler JA, Mourra D, Zanca RM, et al. Vulnerable and resilient phenotypes in a mouse model of anorexia nervosa. *Biol Psychiatry*. 2020;90:829–842.

Blundell J, de Graaf C, Hulshof T, et al. Appetite control: methodological aspects of the evaluation of foods. *Obes Rev*. 2010;11:251–270.

Blundell JE, Gibbons C, Beaulieu K, et al. The drive to eat in homo sapiens: energy expenditure drives energy intake. *Physiol Behav*. 2020;219:112846.

Brobeck JR. Remembrance of experiments almost forgotten. *Appetite*. 1993;21:225–231.

Bruce KR, Steiger H, Young SN, Kin NM, Israël M, Lévesque M. Impact of acute tryptophan depletion on mood and eating-related urges in bulimic and nonbulimic women. *J Psychiat Neurosci*. 2009;34:376–382.

Cao X, Xu P, Oyola MG, et al. Estrogens stimulate serotonin neurons to inhibit binge-like eating in mice. *J Clin Invest*. 2014;124:4351–4362.

de Araujo IE, Schatzker M, Small DM. Rethinking food reward. *Annu Rev Psychol.* 2020;71:139–164.

Fowler N, Klump KL. A virtual issue highlighting animal studies of eating disorders as valuable tools for examining neurobiological underpinnings and treatment of eating disorders. *Int J Eat Disord*. 2020;53:1569–1578.

French SA, Epstein, LH, Jeffrey, RW, Blundell JE, Wardle J. Eating behavior dimensions. Associations with energy intake and obesity. A review. *Appetite*. 2012;59:541–549.

Geary N. Control-theory models of body-weight regulation and body-weight-regulatory appetite. *Appetite*. 2020;144:104440.

Han W, Tellez LA, Perkins MH, et al. A neural circuit for gut-induced reward. *Cell.* 2018;175:665–678.e23.

Kaelberer MM, Buchanan KL, Klein ME, et al. A gut-brain neural circuit for nutrient sensory transduction. *Science*. 2018;361:eaat5236.

Kaye WH, Weltzin TE, McKee M, McConaha C, Hansen D, Hsu LK. Laboratory assessment of feeding behavior in bulimia nervosa and healthy women: methods for developing a human-feeding laboratory. *Am J Clin Nutr*. 1992;55:372–380.

Levin BE, Magnan C, Dunn-Meynell A, Le Foll C. Metabolic sensing and the brain: who, what, where and how? *Endocrinology*. 2011;152:2552–2557.

Lowell BB. New neuroscience of homeostasis and drives for food, water, and salt. *N Engl J Med*. 2019;380:459–471.

Ma R, Mikhail ME, Culbert KM, Johnson AW, Sisk CL, Klump KL. Ovarian hormones and reward processes in palatable food intake and binge eating. *Physiology (Bethesda)*. 2020;35:69–78.

Mani BK, Puzziferri N, He Z, et al. LEAP2 changes with body mass and food intake in humans and mice. *J Clin Invest*. 2019;129:3909–3923.

Moody L, Liang J, Choi PP, Moran TH, Liang NC. Wheel running decreases palatable diet preference in Sprague–Dawley rats. *Physiol Behav*. 2015;150:53–63.

Morton GJ, Blevins JE, Williams DL, et al. Leptin action in the forebrain regulates the hindbrain response to satiety signals. *J Clin Invest*. 2005;115;703–710.

Polidori D, Sanghvi A, Seeley RJ, Hall KD. How strongly does appetite counter weight loss? Quantification of the feedback control of human energy intake. *Obesity.* 2016;24:2289–2295.

Ravussin Y, Leibel RL, Ferrante AW Jr. A missing link in body weight homeostasis: the catabolic signal of the overfed state. *Cell Metab*. 2014;20:565–572.

Ruddock HK, Brunstrom JM, Vartanian LR, Higgs S. A systematic review and meta-analysis of the social facilitation of eating. *Am J Clin Nutr*. 2019;110:842–861.

Schalla MA, Stengel A. Activity based anorexia as an animal model for anorexia nervosa—a systematic review. *Front Nutr*. 2019;6:69.

Schwartz GJ, Moran TH. Sub-diaphragmatic vagal afferent integration of meal-related gastrointestinal signals. *Neurosci Biobehav Rev*. 1996;20:47–56.

Smith KR, Papantoni A, Veldhuizen MG, et al. Taste-related reward is associated with weight loss following bariatric surgery. *J Clin Invest*. 2020;130:4370–4381.

Steinert R, Feinle-Bisset C, Asarian L, Horowitz M, Beglinger C, Geary N. Ghrelin, CCK, GLP-1 and PYY(3–36): secretory controls and physiological roles in eating and glycemia in health, obesity, and after RYGB. *Physiol Rev*. 2017;97:411–463.

Treesukosol Y, Moran TH. Analyses of meal patterns across dietary shifts. *Appetite*. 2014;75:21–29.

Walsh BT, Kissileff HR, Cassidy SM, Dantzic S. Eating behavior of women with bulimia. *Arch Gen Psychiat*. 1989;46:54–58.

1.18 Genome, Transcriptome, and Proteome: The Molecular Genetics, Biochemistry, and Multi-Omics Underlying the Neurobiology of Mental Disorders

EVAN E. AFSHIN, M.D., AND CHRISTOPHER E. MASON, PH.D.

This is an era in which the global study of the DNA, ribonucleic acid (RNA), and protein-building blocks of cells has become feasible, accessible, and increasingly routine. Genome-wide association studies (GWAS) of a million or more common genetic markers from large populations are now relatively inexpensive and conducted routinely. The same can be said for studies that sequence the exome, that is, the DNA of expressed genes as well as the entire genome. The growing understanding of genome variation provided by the International HapMap Consortium and continued major advances in genotyping technology together have made it possible to conduct high-throughput, cost-effective GWAS in large numbers of individuals with detailed information on phenotypic traits and environmental exposures. The resulting data will be used to identify genetic variants potentially related to mental disorders, assess the prevalence of these variants in large and diverse samples, and examine possible modifiers of gene–disease relationships. Soon, the 1,000 genome project followed, organizing a comprehensive database of SNPs, copy number variations (CNVs), insertions and deletions (indels), as well as large structural rearrangements. ENCODE project (described later in this chapter) expanded these studies further into functional genomics. Functional genomics is already becoming routine in brain research and is being applied to the study of postmortem human brains from individuals with mental illness and to animal models of relevance to clinical neuroscience.

These exciting molecular genetic approaches permit the study of biologic information from a global perspective. This information is contained in the **human genome**, the three billion base sequence comprising the full complement of DNA in human cells. The genome of each individual is contained within every cell in the body that carries a nucleus and represents the raw material necessary to allow for normal development. This genetic material also plays a causal or contributory role in much of human disease, including mental illness. **Genomics** is the study of the full complement of genetic material of an organism. The ability to consider humans in this light is a fairly recent development, one that has been energized by the sequencing of the human genome and the subsequent large-scale efforts to identify and characterize human genetic variation and all functional and regulatory elements. The **transcriptome** is a term that can be applied to the sum total of RNAs transcribed from genes in an organism, a process often referred to as gene “expression.” Studies of gene expression usually focus on **messenger RNA** (mRNA), the intermediate between genes and proteins. They also vary depending on the experimental design and may include only mature, polyadenylated mRNA or total RNA (mRNA plus ncRNAs) in the cell. Global studies of gene expression compose the field of **functional genomics** and may use such tools as **next-generation sequencing** (NGS), a methodology that relies on the power of massive parallel sequencing of short reads derived from genomic DNA or RNA. A simplified diagram of NGS workflow is shown in Figure 1.18–1.

The **proteome** denotes the full panoply of proteins expressed in an organism and studied globally by proteomics. As the general understanding of these biologic domains in humans and many other organisms has grown, so too has the appreciation of its complexity. The deeper we delve into its structure, function, and complex interactions, the more we are pressed to reconsider basic concepts. For instance, the amount of genetic variation in apparently normal healthy people is underestimated. While it has long been appreciated that the sequence of the DNA varies slightly between individuals, it has recently been found that as much as 20% of the human genome can possess large-scale structural variation as well. Hundreds of thousands to millions of nucleotides of the genetic code may be missing, duplicated, or have multiple copies within a single individual without obvious adverse effects. These variations, which are referred to as **copy number polymorphisms**

(CNPs) or CNVs are just beginning to be cataloged and studied in depth. In addition, there are extremely large numbers of common differences between individuals on a single nucleotide level, known as SNPs. Understanding the importance of these many forms of genetic variation promises to revolutionize the current thinking about the role of genetic structural variation in normal development, shed light on the heritability and biology of traits associated with psychiatric disorders, and perhaps ultimately reshape psychiatric diagnostic schema.

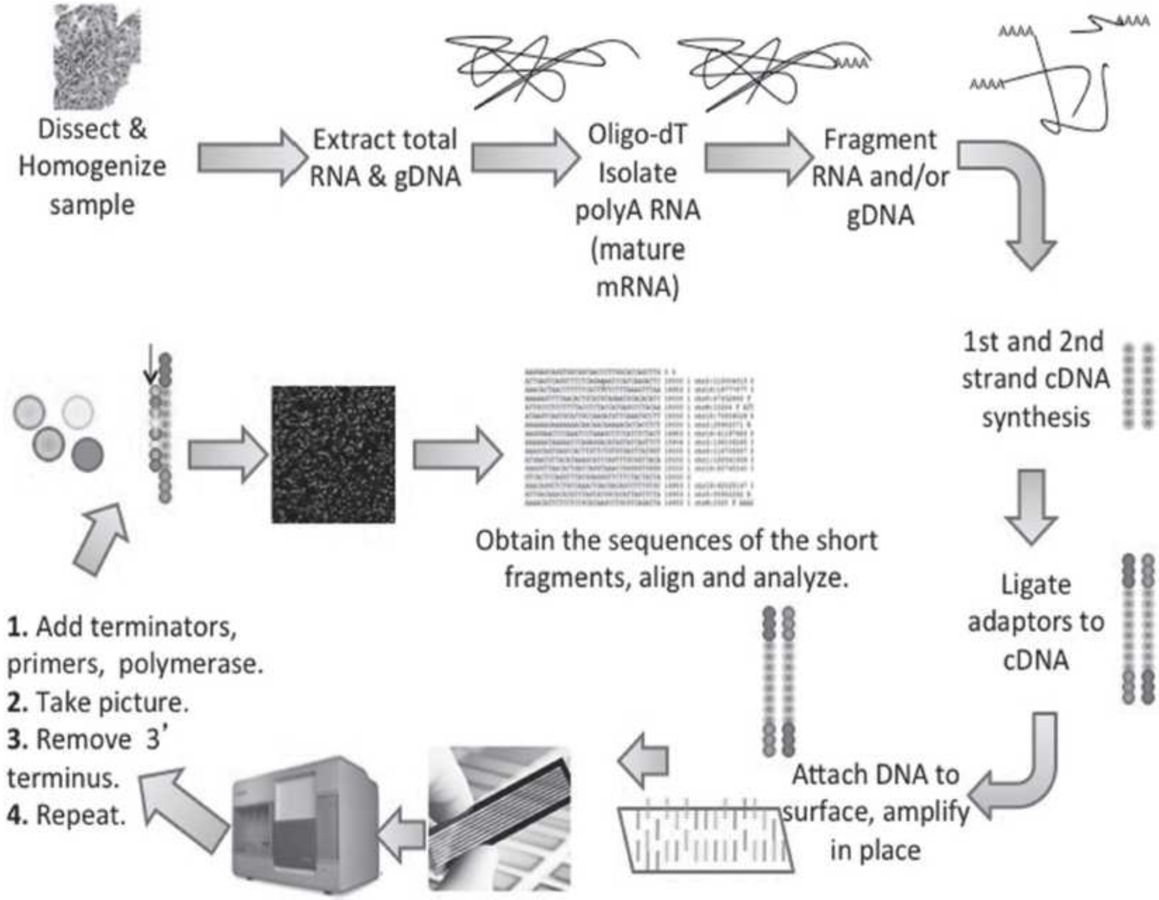
Dissect & Homogenize sample
Extract total RNA & gDNA
Oligo-dT Isolate polyA RNA (mature mRNA)
Fragment RNA and/or gDNA
1st and 2nd strand cDNA synthesis
Ligate adaptors to cDNA
Attach DNA to surface, amplify in place
1. Add terminators, primers, polymerase.
2. Take picture.
3. Remove 3' terminus.
4. Repeat.
Obtain the sequences of the short fragments, align and analyze.

FIGURE 1.18–1. Overview of the NGS workflow. Tissue is being dissected and homogenized, and genomic DNA and/or RNA are extracted. In case of studying mature coding mRNAs, oligo-dT purification is performed. Next, DNA and/or RNA is fragmented into short pieces (~200 nt) and cDNA library is derived. Next adaptors are ligated, who help anchor the cDNA to a hard surface, and sequencing by synthesis is performed. The final results are the sequences of the short fragments of DNA or RNA.

The notion of a “gene” is evolving as well. It was long thought that a single gene led to the production of a single protein. However, by 1977, it was discovered learned that one gene may code for multiple versions of a protein that differed functionally. Recent data challenge the conventional definition of the boundaries of an individual gene, findings which will likely lead to a major reappraisal of genomic function and regulation. For example, most genes have multiple 3′ and 5′ termini and various splicing sites and harbor various regulatory elements within their gene bodies. In addition, elements called enhancers can control gene expression long-range, and transcription factor binding site’s (TFBSs) can regulate the ability of genes to be activated, attenuated, or silenced.

In addition to forcing a reconsideration of fundamental genetic concepts, the sequencing of the entire human genome also set the stage for a far deeper understanding of how genetic variation alters the regulation of gene expression and how variation either in functional or regulatory domains of genes affects brain structure and function, and how this variation contributes to the risk for mental disorders.

This chapter will review fundamental concepts, including the nature of DNA, chromosomes, RNA, proteins, and the concept of coding and noncoding genes. This chapter will then review what is currently known about the sequence characteristics of the human genome on a global scale and consider how this sequence varies between individuals and among populations. Subsequently, this chapter will consider how the genome is regulated and the manner in which the study of proteomics and systems biology promises to leverage these recent advances in functional genomics to enhance the general understanding of normal and pathologic brain development and functioning.

THE ORGANIZATION AND STRUCTURE OF THE HUMAN GENOME

A **genome** is defined as the total complement of DNA replicated in a living organism. The sequencing of genomes of free-living organisms began in 1995 with bacteria and progressed to larger and more complicated organisms, such as yeast (1996), worms (1998), and fruit flies (2000). A critical milestone was reached in 2001 with the completion of the first draft of the human genome. Currently, there are thousands of genomes sequenced. In 2011, the Genome in a Bottle Consortium was established with the goal of providing an entire technical infrastructure that will enable data from genome sequencing to be used in translational medicine.

As these data became available, four curious patterns emerged: (1) the number of genes is lower than expected. Estimates prior to the completion of the draft human sequence ran as high as 160,000. As of the last issue of GENCODE annotation, the official annotation of the ENCODE project, there are 60,155 genes, with 19,881 of them being protein-coding genes. It is now evident that there are, in total, about 20,000 protein-coding genes in *Homo sapiens;* (2) gene number is not a predictor of organismal complexity. For example, amoebas are simple, single-celled organisms but are predicted to have many more genes than humans (Table 1.18–1); mice and humans have been found to have approximately the same number of protein-coding regions; (3) genome size is not a predictor of organismal complexity. An amoeba (*Amoeba dubia)* has the largest known genome (670 gigabases)—more than 200 times larger than the human genome; humans and chimpanzee genomes are essentially the same size but they hold obvious and important differences with respect to the development of the cerebral cortex; and (4) a small percentage of eukaryotic genomes codes for proteins. Only 2% of the human genome encodes proteins. The vast majority is comprised of intronic and repetitive sequences, some of which are clearly functional and much of which remains a mystery. Apart from protein-coding genes, there are a substantial number of actively transcribed genes that don't code any proteins (ncRNAs, to be discussed below), sometimes called the "dark matter" of the genome. Some of this "dark matter" is known to be functional, although the functional significance of much of ncRNAs is yet to be resolved.

Table 1.18–1.
Number of Genes and Genome Sizes across Different Organisms

Organism	Gene Count	Genome Size (base pairs)	Number of Chromosomes
Amoeba	50,000	670,000,000,000	13
Plant (Fern)	37,500	100,000,000,000	90
Human	25,000	3,000,000,000	46
Chimpanzee	25,000	3,000,000,000	48
Mouse	25,000	2,500,000,000	40
Honey bee	15,000	300,000,000	32
Fruit fly	14,000	130,000,000	10
Worm	19,000	97,000,000	12
Fungus	6,000	13,000,000	32
Bacterium	3,000	5,000,000	1
Mycoplasma genitalium	500	580,000	1
DNA virus	450	50,000	0
RNA virus	20	10,000	0
Viroid	1	500	0

Sorted by Genome Size. All numbers are approximate.

Cumulatively, these observations pose a central question for psychiatry: if the genome of humans is not very distinctive from an evolutionary perspective, how has it produced the unique qualities of the human CNS? Recent developments have begun to shed light on this critical question.

DNA and Chromosomes

DNA is made of four **nucleic acids**, also known as **nucleotides** or **bases**, adenosine (A), cytosine (C), guanine (G), and thymine (T). In total, human genomic DNA is comprised of approximately 3 billion nucleotides, and this full complement is found in every cell in the body that contains a nucleus. Within the nucleus, the genome is organized into 46 strands of DNA that complex with multiple proteins to form **chromosomes** (23 inherited from mother and 23 inherited from father). In the nuclei of cells, the strands of DNA are combined with **histone proteins** that enable DNA to be wound tightly into histone-DNA complexes called **nucleosomes** and then organized into a super structure called **chromatin**.

From the first observations of chromosomes under the light microscope, chromatin has been divided into two types corresponding to the familiar light and dark banding patterns: **euchromatin** (lighter, less dense material) and **heterochromatin** (darker, more dense material). In one sense, chromatin acts simply as a spool around which DNA is wound, ensuring that the entire genome fits within the nucleus. This property enabled cells to remain small, a property that influenced the evolution of multicellular organisms. However, it also plays a key role in coordinating the function of the genome. The histone protein cores around which the DNA is organized may be altered by chemical reactions, including **acetylation, phosphorylation**, and **methylation**. The addition or subtraction of these and other chemical modifiers are able to dictate the conformation of regions of DNA within the nucleus, a process that helps coordinate the regulation of gene expression. A field of study, epigenetics has emerged to study the way that gene expression is altered by modifying chromatin or DNA methylation but not altering the DNA sequence (see the section on Epigenetics in this chapter). Finally, the organizational structure of chromatin creates scaffolding that helps guide the movement of DNA necessary for chromosomal replication and during cellular division (**mitosis**).

The darker staining of heterochromatin is due to the very tight packing of nucleosomes, making the DNA largely inaccessible to cellular machinery, whereas the light color of the euchromatin reflects a more open, unfolded, and usually active state. Euchromatic regions were found to be distinct in other ways as well: they hold many more genes and contain less repetitive sequences. However, heterochromatic regions of the genome can become active and some heterochromatic regions are modified back and forth, becoming active only when needed. These processes will be discussed in more detail subsequently when epigenetic mechanisms are considered at the conclusion of this chapter.

Genes, RNAs, and Proteins

Gregory Mendel first discovered genes in 1866 when he identified hereditary units he called “factors” and elaborated the principles of their inheritance through his experiments with pea plants. However, he had no physical or mechanistic understanding of what constituted a gene. The first medical application of genes came in 1902 when Archibald Garrod found that alkaptonuria, a rare disorder characterized in part by the urine turning black when exposed to air, followed Mendel laws of inheritance. He identified that the transmission of this condition followed an autosomal recessive pattern. However, while it was clear that genes were passed from one generation to the next and carried disease liability, it was still unclear precisely how or in what substance.

In 1910, Thomas Morgan demonstrated that genes were discrete units on chromosomes through his work with fruit flies, and he first proposed the idea of genetic linkage due to chromosomes exchanging material in a process called **crossing over** (Fig. 1.18–2).

In 1941, experiments in bread molds (*Neurospora crassa*) by Edward Tatum and George Beadle showed further that certain enzymes were functionless if the genes were mutated. This led to the widely accepted notion that specific genes make specific proteins, commonly referred to as the “**one gene, one enzyme**” hypothesis.

Despite these remarkable discoveries, the physical substrate for the transmission of genetic information was still a matter of debate. Early in the 20th century, proteins were favored, as they were known to be comprised of 20 different amino acids. As DNA was the product of only four different nucleotides, it was thought that the former was most likely to allow for the diversity required to make the complex instructions necessary for sustaining life. However, in 1944, genes were shown to be composed of DNA by Oswald Avery, through his demonstration that DNA could create a heritable transformation in bacteria, but proteins could not.

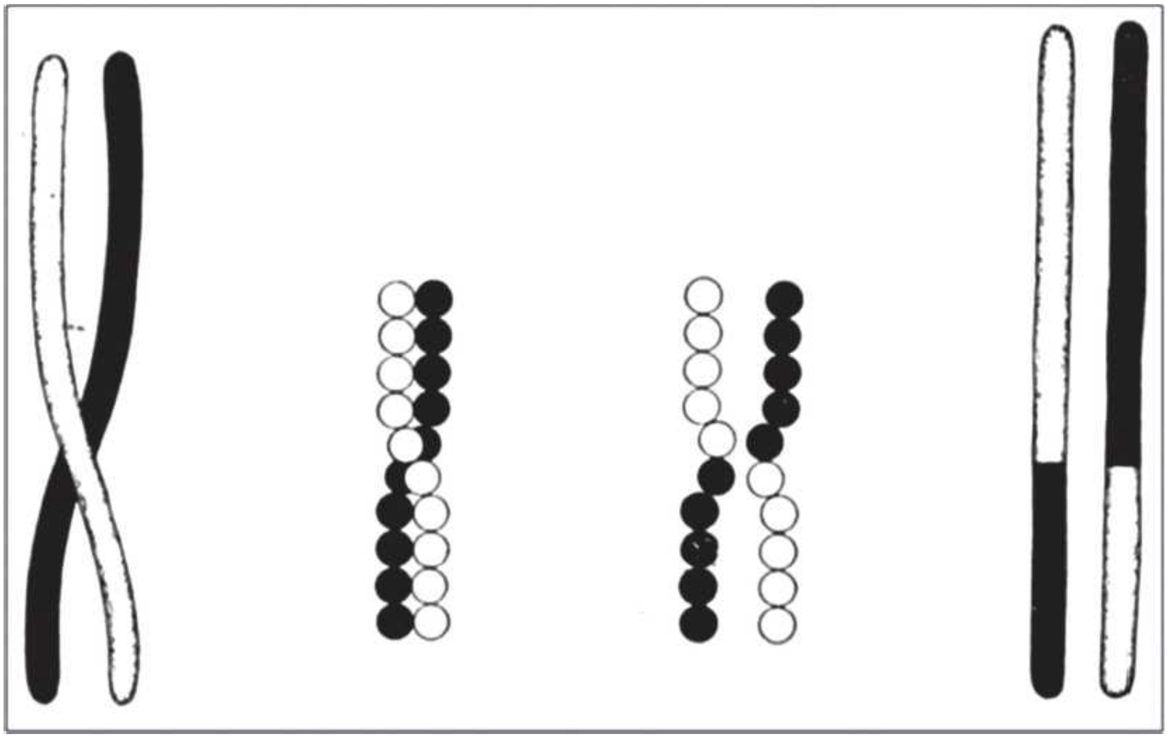

FIGURE 1.18–2. Crossing Over. A schematic of genes positioned along chromosomes (*circles*) and exchanging information (crossing over) with a different series of genes (*black circles suddenly with white circles*) during meiosis. The farther apart genes are on a chromosome, the more often they cross over—a phenomenon known as recombination.

Finally, in 1953, Watson and Crick published the chemical structure of DNA, showing that it was a **double helix** formed by two sugar-phosphate backbones supporting the four nucleotides. These, in turn, were recognized to form specific pairs between them, such that an “A” on one strand paired with “T” on the other and “C” on one paired with “G” on the other.

Coding Genes

The last several decades of progress have now led to an appreciation that there are, in a broad sense, two classes of genes: those that lead directly to the production of proteins (**coding genes**) and those that do not (**noncoding genes**). Coding genes are stretches of DNA that are **transcribed** by an enzyme (RNA polymerase) into a temporary mRNA, which is then **translated** into a protein (made of peptides), following what is known as the central dogma of molecular biology: DNA → RNA → Protein (Fig. 1.18–3).

In the 1960s, it was discovered that coding DNA is “read” in three-letter segments called **codons** (see Fig. 1.18–3), each of which becomes an amino acid in a peptide or protein. However, unlike simpler prokaryotic organisms, human coding genes are transcribed initially as sequences that require editing. Coding genes have segments that are read as codons (called **exons**) and also noncoding sequences that are transcribed but not translated into a protein (called **introns**). Noncoding sequences can be present in three areas of a gene: in front (**upstream** with respect to the sequence of transcription) of the gene, inside the gene, and behind (**downstream**) the gene. Those sequences upstream of the gene are called the **5′ UTR**. The noncoding region downstream of the coding interval is the **3′ UTR**. Mature mRNA contains a polyadenylation signal (i.e., AAUAA) that directs transcriptional machinery to insert a long string of adenosines, known as a poly-A tail, at the 3′ terminus of the mRNA. A single gene may contain several polyadenylation signals in series in its 3′ UTR, but the mechanism of determining which one is going to be used has yet to be established. In turn, multiple polyadenylation signals allow for transcripts with variable 3′ UTR length to be produced from a single gene. Transcripts with longer 3′ UTRs are relatively prevalent in the brain, giving rise to higher numbers of miRNA binding sites that enable fine-tuning of RNA regulation. In addition, longer 3′ UTRs are known to harbor protein-binding motifs that shuttle proteins to a specific subcellular location. The BDNF gene is known to possess a form with long 3′ UTR with a binding motif for BDNF that enables it to be transported to dendrites. Thus, the nascent mRNA transcript produced from a coding gene must be first processed to lead to the coherent set of instructions necessary to produce a protein product at the appropriate subcellular location. The removal of introns, through a process called **gene splicing**, is one critical step in determining a gene’s function, and it will be addressed in more detail when considering the function of each part of the genome.

Since gene transcripts are read in three-base codons, and DNA is made up of four nucleotides, there are 64 possible codons (4 × 4 × 4). In most organisms, mRNA that are translated into proteins possess a three-base sequence, called a **start codon** (AUG), delineating where the reading or translation of the codons should begin. Three **stop codons** (UAA, UAG, UGA) signal the last amino acid in the growing polypeptide chain. Start and stop codons are the punctuation needed for reading a gene, while the other 60 codons determine which amino acids are included in the peptide or protein. Since the cell uses only 20 different amino acids to make proteins, there exists some redundancy in the genetic code such that different codons can encode for the same amino acid (Fig. 1.18–4). This redundancy in codon usage may have a specific function. There is a so-called codon usage bias, where translation favors some codons over others. Differences in codon translation efficiency are related to the concentration of tRNAs that correspond to those codons, thereby implying that there is a complex interplay in-between them. Such complex evolutionary changes may fine-tune the translation process, especially in tissues with an extremely broad repertoire of expressed genes, such as neurons.

The processed RNAs are shuttled through the cell to organelles called **ribosomes** to be turned into a protein chain, which then folds into increasingly complex levels of organization (Table 1.18–2).

The total of all the proteins made by the genome is called the proteome. Many proteins serve as subunits for larger protein complexes (including the large enzyme that performs transcription, RNA polymerase II), and the study of many proteins at once is known as proteomics, a topic that will also be addressed in more detail later in this chapter.

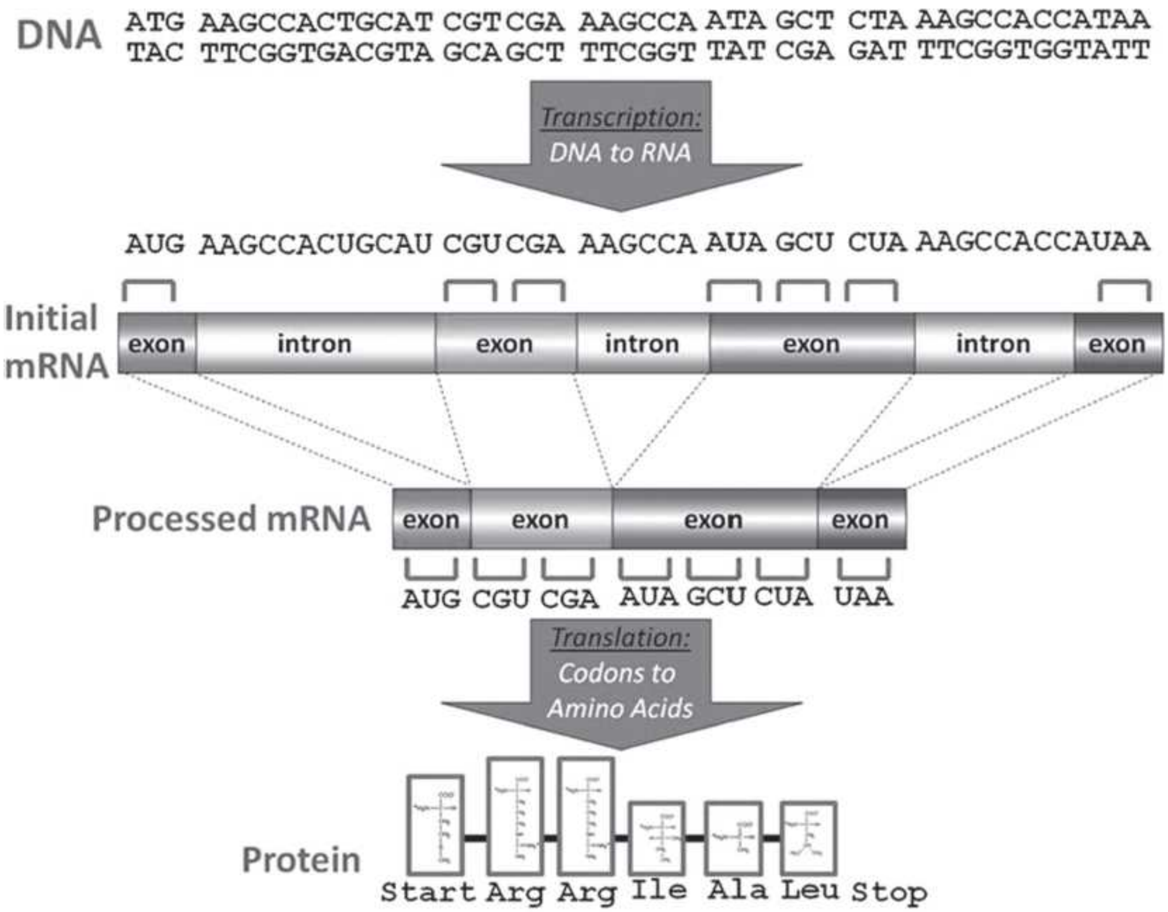
DNA
ATG AAGCCACTGCAT CGT CGA AAGCCA ATA GCT CTA AAGCCACCATAA
TAC TTCGGTGACGTA GCA GCT TTCGGT TAT CGA GAT TTCGGTGGTATT
Transcription: DNA to RNA
AUG AAGCCACUGCAU CGU CGA AAGCCA AUA GCU CUA AAGCCACCAUAA
Initial mRNA
exon
intron
exon
intron
exon
intron
exon
Processed mRNA
exon
exon
exon
exon
AUG CGU CGA AUA GCU CUA UAA
Translation: Codons to Amino Acids
Protein
Start Arg Arg Ile Ala Leu Stop

FIGURE 1.18–3. Central Dogma of Molecular Biology. During transcription, the thymine (T) in DNA is represented with a different nucleic acid (U, for Uracil), so that the codon can be understood by the cell's machinery. Introns are removed from the final messenger RNA (mRNA), leaving a product that is ready to become a protein through translation of the codons.

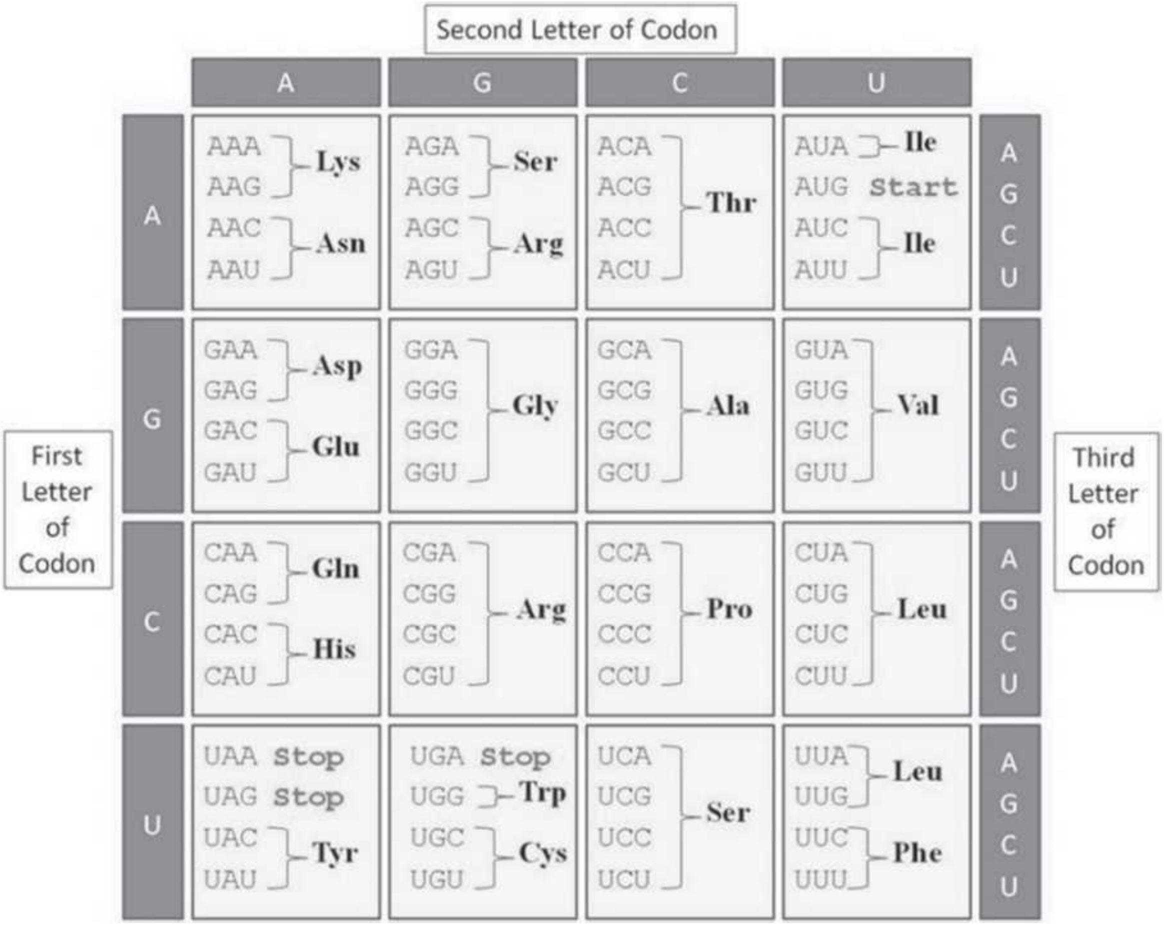

First Letter of Codon	Second Letter of Codon				Third Letter of Codon
	A	G	C	U	
A	AAA, AAG – Lys AAC, AAU – Asn	AGA, AGG – Ser AGC, AGU – Arg	ACA, ACG, ACC, ACU – Thr	AUA – Ile AUG Start AUC, AUU – Ile	A G C U
G	GAA, GAG – Asp GAC, GAU – Glu	GGA, GGG, GGC, GGU – Gly	GCA, GCG, GCC, GCU – Ala	GUA, GUG, GUC, GUU – Val	A G C U
C	CAA, CAG – Gln CAC, CAU – His	CGA, CGG, CGC, CGU – Arg	CCA, CCG, CCC, CCU – Pro	CUA, CUG, CUC, CUU – Leu	A G C U
U	UAA Stop UAG Stop UAC, UAU – Tyr	UGA Stop UGG – Trp UGC, UGU – Cys	UCA, UCG, UCC, UCU – Ser	UUA, UUG – Leu UUC, UUU – Phe	A G C U

FIGURE 1.18–4. **The Genetic Code**. Genes are read in three-letter nucleotide segments (codons) and are then read by the cellular machinery to correspond to particular amino acids.

Table 1.18–2.
Protein Structure Definitions

Name	Definition
Primary	A chain of amino acid (peptides), making a polypeptide
Secondary	The amino acids are bound within a chain by hydrogen bonds
Tertiary	The polypeptide forms entire sheets and helices for a 3D structure
Quaternary	Multiple polypeptide chains bond together for a macromolecular structure

Noncoding Genes (or Noncoding RNAs)

General statistics of the GENCODE annotation for the past 4 years show a very interesting trend: the total number of annotated genes is rising; however, the number of protein-coding genes decreases at the expense of a rising number of ncRNAs, especially pseudogenes and long intergenic noncoding RNAs (LINCs), as shown in Table 1.18–3.

While coding genes conform most closely to early notions of how hereditary information is stored and processed, over time, it has become clear that these represent only about 2% of the genome. Moreover, an alternative set of instructions contained within the DNA governs transcription that is not followed by the production of a protein. Over 30,000 noncoding genes have so far been cataloged in the human genome. These are comprised of RNA (and consequently are also called **noncoding RNAs**, or **ncRNAs**). Table 1.18–4 lists several types of RNAs. Some are involved in regulating normal cellular processes of gene transcription and protein translation, whereas others have only recently been discovered and their functions are not as well understood. Moreover, some ncRNAs can function all by themselves, catalyzing reactions and acting autonomously in the cell.

Some of the best-characterized ncRNAs are called **siRNAs**. These are transcripts coded for in the DNA that are complementary (or antisense) to another transcribed sequence from the genome. Once a siRNA binds to the complementary mRNA sequence, the cell degrades the transcript, thus functionally silencing the gene that produced the mRNA without altering the DNA or mRNA sequence. Similarly, a different subset of noncoding genes, called micro-RNAs (miRNAs), are short (18–22 bp) sequences that bind most often to the 3′ UTR of a transcript and most commonly lead to a decrease in the level of the resulting protein. Both siRNAs and miRNAs operate post-transcriptionally to regulate gene expression. Thus, the multiple extremely important ways along the process from DNA to protein in which gene expression may be altered is of major importance to scientists seeking to understand the function of the genome. This topic will be addressed in the second section of this chapter. The recognition that ncRNAs are widespread and play a key role in the regulation of gene expression strongly suggests that they may also confer risk for mental disorders and other complex diseases. Moreover, it was quickly realized that siRNAs could be useful to assess the consequences of loss of expression of a particular gene at the level of the cell or the model organism; as a result, siRNAs have quickly become an indispensable research tool.

Table 1.18–3.
Distributions of Annotated Genes by Their Coding Capacity and Biotype for GENCODE 2010–2020

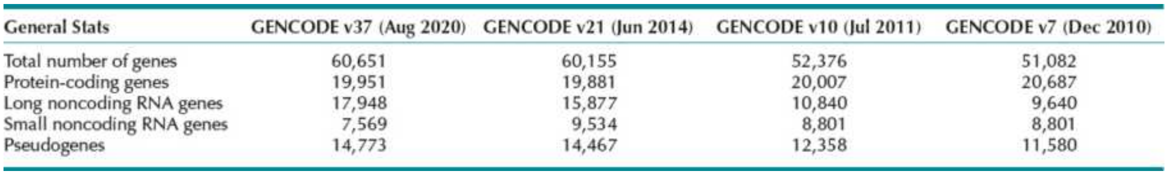

General Stats	GENCODE v37 (Aug 2020)	GENCODE v21 (Jun 2014)	GENCODE v10 (Jul 2011)	GENCODE v7 (Dec 2010)
Total number of genes	60,651	60,155	52,376	51,082
Protein-coding genes	19,951	19,881	20,007	20,687
Long noncoding RNA genes	17,948	15,877	10,840	9,640
Small noncoding RNA genes	7,569	9,534	8,801	8,801
Pseudogenes	14,773	14,467	12,358	11,580

Two other important classes of ncRNAs are pseudogenes and LINCs. Pseudogenes are ubiquitous in the brain transcriptome and encompass several types of transcripts, transcriptionally active genes that lost their coding potential—unitary pseudogenes (GULO-vitamin C–producing enzyme), duplicated genes that lost their coding potential, and reverse transcribed genes (processed pseudogenes). The reason their number keeps growing in each subsequent annotation is because they share a substantial portion of their sequence with the functional genes, which makes them extremely hard to identify. Bioinformaticians have developed highly sophisticated algorithms to be able to distinguish pseudogenes from functional genes. The "sponge" hypothesis suggests that pseudogenes may compete with functional genes for a limited pool of miRNAs. If so, then pseudogenes may influence overall gene regulation without possessing intrinsic function.

LINCs also appear to influence neural function. Most LINCs alter transcription (H19, XIST, AIR, HOTAIR) in various ways (i.e., direct effect on transcription, changing the epigenetic signature). Others exert their function on the translational level (GAS5).

While a significant amount of ncDNA is transcribed to functional ncRNAs (see Table 1.18–4), it has not been clear how much of the genome was actively processed nor how much of this transcription might contribute directly to biologic function. To answer this question, the National Human Genome Research Institute (NHGRI) launched a project in 2003 to identify all the functional elements in the genome called ENCODE (the Encyclopedia Of DNA Elements). By the second phase of the ENCODE project, it appeared that 75% of the human genome was transcriptionally active. However, 70% of the genome showed extremely low expression levels, raising questions regarding functional significance. Also, 11% of the genome was associated with motifs in transcription factor-bound regions or high-resolution DNase footprints in one or more cell types. Promoter-associated histone modifications or enhancers marked ~20% of the genome, while as much as one-third of the genome was marked by modifications associated with transcriptional elongation. Taken together, it appears that the vast majority of the genome is either transcribed or harbors regulatory elements (Fig. 1.18–5).

These observations strained the ability to classify genetic material. While there is a clear distinction between coding and noncoding genes, a large amount of transcriptionally active DNA produces RNAs that do not fit into current functional categories. For clarity, the term ncRNA will be used to describe noncoding genes such as siRNAs and miRNAs, whose general structure is well characterized. The additional transcribed material is referred to as **transcripts of unknown function (TUFs)** or **transcriptionally active regions (TARs).**

TUFs and TARs may represent (1) new protein-coding genes that were missed by previous experiments and gene-finding algorithms; (2) new noncoding genes; (3) novel antisense transcriptional units (siRNAs and miRNAs) that may regulate other genes; (4) alternative isoforms of known genes that include intronic regions in the final mRNA product; (5) misannotated genes that should be longer or shorter, divided into two genes, or merged into one larger gene; (6) biologic artifacts representing aberrant transcription or accidental transcriptional read-through; or (7) experimental artifacts due to small amounts of genomic DNA contamination.

Table 1.18–4.
Types of RNA

Name	Abbrev	Year	Function
Messenger	mRNA	1956	Carries the transcribed message from the DNA
Ribosomal	rRNA	1958	Components of the protein-making machinery, the ribosome
Transfer	tRNA	1962	Function in translation to bring amino acids to the ribosome
Small nuclear	snRNA	1977	Help the splicing of immature mRNAs to reach their final form
Small nucleolar	snoRNA	1986	Direct chemical modifications of rRNAs and other RNAs
Micro	miRNA	1993	Single-stranded RNA that regulates gene expression
Small interfering	siRNA	1999	Small, double-stranded RNAs that interfere with transcription
Piwi-acting	piRNA	2006	Germ-line acting RNA that stops parasitic genetic elements
Virus-derived, small interfering	viRNAs	2011	Virus-derived, small interfering RNAs that mediate trans-generational inheritance of viral silencing

Future experiments will be needed to clarify the role of these TARs, but some conclusions are already possible. For instance, the majority (93%) of the euchromatic genome is functional in some capacity (either transcribed or regulating another sequence). Also, gene regulation is symmetrical along the genome, with no bias for upstream or downstream placement of regulatory structures. This contrasts with previous notions that regions 5′ or upstream to the starting codon of a gene were most likely to modulate the regulation of that transcript. Finally, surprisingly, many of these regulatory or active regions of the genome are not highly conserved across species. Taken together, these data show that any search for genes and genetic mechanisms contributing to brain development and mental disorders must extend far beyond the horizon of coding sequences.

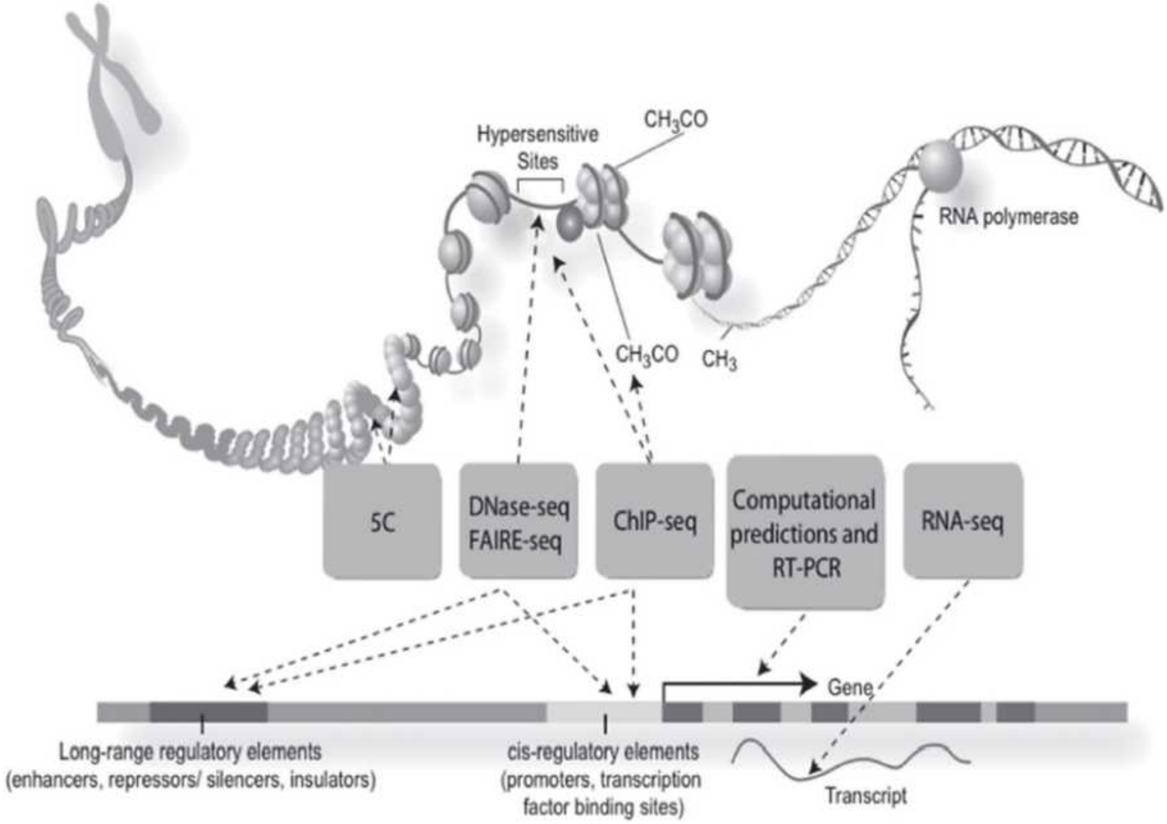
Hypersensitive Sites
CH_3CO
RNA polymerase
CH_3CO
CH_3
5C
DNase-seq
FAIRE-seq
ChIP-seq
Computational predictions and RT-PCR
RNA-seq
Gene
Long-range regulatory elements
(enhancers, repressors/ silencers, insulators)
cis-regulatory elements
(promoters, transcription factor binding sites)
Transcript

FIGURE 1.18–5. Summary of the ENCODE Project. The discovery and annotation of TARs was performed by sequencing RNA from a diverse range of cell lines, followed by comparative genomics, integrative bioinformatic methods, and human curation. Investigation of the regulatory elements involved DNA hypersensitivity assays, assays of DNA methylation, and chromatin immunoprecipitation (ChIP) of proteins that interact with DNA, including modified histones and transcription factors, followed by sequencing (ChIP-Seq).

Repetitive DNA

With the sequencing of the human genome and the recognition that only a small proportion is present in the form of coding DNA or ncRNAs, interest in the remainder of the genome has increased, and much of this material is present in the form of repetitive elements.

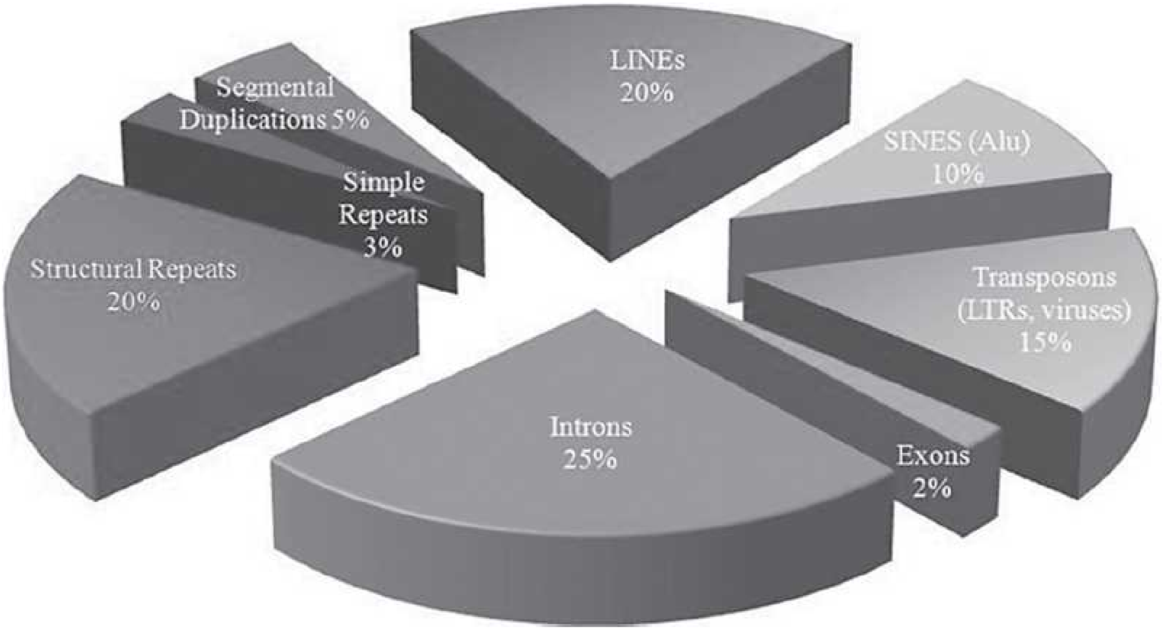

FIGURE 1.18–6. Structural Organization of the Human Genome. Most of the human genome is repetitive DNA sequences and transposable elements (LINES, SINEs, LTRs, and viruses). Very little of the genome is coding sequence (exons) but there is great room for gene flexibility and change with many of the gene's long intron sequences (introns).

The structural organization of the human genome is shown in Figure 1.18–6. Nearly one-third of the genome consists of repeated elements. Some repeated sections of the human genome serve structural purposes at the end or middle of chromosomes. However, many types of repeats are of unknown function, including **simple sequence repeats** (SSRs) and **segmental duplications** (SegDups). SSRs are small sequences (2 to 6 bp) that are tandemly repeated in the genome. A short trinucleotide repeat (CAG) in the coding part of the HGG gene responsible for coding the Huntingtin protein is known to be the cause of Huntington disease, a progressive neurodegenerative disorder that is ultimately fatal. While healthy individuals have 10 to 35 repeats in the HGG gene, affected individuals can have as many as 120. The onset and severity of the disease correlate very strongly with the number of repeats. This is one of the few neuropsychiatric diseases with a clearly defined genetic etiology that can be diagnosed prior to the onset of symptoms with a simple test. Children of a parent with Huntington disease have a 50% risk of developing this disorder. Thus, the capacity to diagnose an illness in presymptomatic individuals raises important ethical issues related to the role of genetics in neuropsychiatric clinical practice. SegDups are sequences of 1,000 bp or greater that appear at least twice throughout the genome. Interestingly, some of these SegDups contain extra copies of entire genes.

The most ubiquitous repeated genetic elements in the genome are transposable elements (TEs or **transposons**), which can jump from one place to another in the genome. This process may lead to new forms of genes that may be useful to humans, but it also poses a danger of disrupting essential genes. TEs are divided into several subgroups: **long interspersed elements** (LINEs), **short interspersed elements** (SINEs), and the small 300 bp *Alu* element (considered a SINE). The *Alu* element is present in 75% of introns and accounts for 10% of the entire genome, whereas the 6,000 bp (6 kb) LINEs account for 20% of the human genome and can be found in thousands of genes (in 5′ UTRs, exons, 3′ UTRs, and introns). Almost all genes in the human genome have at least one TE.

Alu elements are both TEs and retrotransposons, which alter DNA through reverse transcription in the manner of retroviruses, such as the 7SL RNA-derived primate-specific repeat family. Their distribution throughout the evolution of primate lineage has resulted in as many as 1.1 million copies in the human genome. Insertions of the Alu elements occurred in a few major waves, deriving the old *AluJ* subfamily (including *AluJo* and *AluJb*; about 81 million years ago), the middle-aged *AluS* subfamily (including *AluS*, *AluSx*, *AluSg*, *AluSq*, *AluSp*, *AluSc*, and *AluSb*; about 48 to 19 million years ago), and the young and still active *AluY* wave (including *AluY*, *AluYa5*, *AluYa8*, and *AluYb8*; starting about 6 million years ago and continuing to the present) (M7). Albeit ubiquitous in the genome, Alu repeats (as any other repeat or low-complexity sequence) are extremely hard to map accurately using the current short-read sequencing. This fact, along with the limited RNA sequence data derived from primates, explains the lack of proper mapping of these elements. The body of the *Alu* element is approximately 300 bp bases in length, formed from two diverged dimers, separated by a short A-rich region. Such structure enables them to fold perfectly into double-stranded structures, thereby being substrates for a process called RNA editing (discussed below).

Novel findings demonstrate that *Alu* elements are actively transposing in neurons in culture as well as in neurons in vivo. These changes in the neuronal genomes might generate neurons with unique genetic, transcriptomic, and proteomic makeup, increasing the complexity of the brain.

Since the human’s divergence from the common ancestor with the chimpanzee, approximately 98,000 viruses have invaded the human genome. These viruses, called **endogenous retroviruses**, total 8% of the human genome and are made up of **long terminal repeats** (LTRs) that reverse transcribe themselves from RNA to DNA, DNA TEs, and some viruses that lay dormant and can no longer replicate (about 4% of the human genome). Some of these sequences are essentially dead genomes that have incorporated into human DNA and are now replicated along with the genome, whereas others are still actively infecting humans by retrotransposition.

The large volume of repetitive noncoding DNA present in the human genome totals 98% of the human genome. These regions were thought to be nonfunctional, earning the misnomer “junk” DNA. However, in light of the many functions ascribed to noncoding DNA, it is generally thought that the functions of this “dark matter” in the genome remain to be discovered. Even if a function for a specific sequence has not been identified, it may still be critical to the functioning of other genes, either nearby or surprisingly remote. Even if a sequence is a repetitive TE that is present in thousands of places in the genome, the location of a single TE may be needed for the correct activity of that specific gene that now possesses the element. In a complete reversal from only a decade ago, every sequence of the genome is now considered putatively functional, opening the door for a better understanding of the genomic biology of psychiatry and revealing a much larger genome that requires critical examination.

The Mitochondrial Genome

All eukaryotes, including humans, possess two fully functional living genomes, that is, the DNA particular to that species and the genomes of mitochondria (in animals) or chloroplasts (in plants). The theory of endosymbiosis suggests that in the early days of evolving life on Earth (1.5 to 2 billion years ago), multicellular life forms merged with respiratory bacteria, which were able to produce cellular energy in a more accessible form (ATP). This relationship proved to be mutually beneficial for both the protoeukaryotes and the respiratory bacteria since the bacteria could evade prey, and the protoeukaryotes had a new source of energy. This theory is supported by the fact that mitochondria have their own genome and separate replication cycles and that the closest living relatives of mitochondria (rickettsia bacteria) are parasites of eukaryotic cells.

The human genome and the mitochondrial genome now have a completely intertwined coexistence. Nearly 300 genes from the mitochondria are present on the 23 human chromosomes. These are referred to as **nuclear DNA of mitochondrial origin** (NUMTs), and they are produced by the nuclear genome to support the function of mitochondria. Further, 27 of these NUMTs do not appear in the chimpanzee or other genomes and have therefore been incorporated into the human genome within the last 4 to 6 million years. Most of them (23/27) are present within known or predicted human genes, indicating that the symbiosis between these two genomes is significantly gene-centered, perhaps ensuring stronger protection against transposable elements and/or protect mitochondrial DNA (mtDNA) from damage inflicted by **reactive oxygen species** (ROS). Moreover, some neuromuscular diseases and metabolic defects arise from mutations or errors specific to mtDNA or NUMTs (Table 1.18–5), underscoring the importance of the mitochondrial genome in influencing the risk of mental disorders. Finally, recent studies have shown how integral the mitochondria are to human physiology, particularly the body's response to stress. A multi-omics and systems biology approach to profiling astronauts revealed significant enrichment for mitochondrial processes as a consistent phenotype of spaceflight.

Exosomes

Exosomes are a type of extracellular vesicle that contains DNA, RNA, and protein of the cells that secrete them. The physiologic purpose of exosomes remains largely unknown and needs further investigation. Some theorize exosomes play a role in maintaining cellular homeostasis by removing excess or unnecessary molecules, while others believe exosomes are integral to regulating complex intercellular and intertissue communication. Indeed, studies have shown exosomes are associated with cardiovascular diseases, viral pathogenicity, cancer progression, inflammatory diseases, and CNS-related diseases, including neurodegeneration. Findings have suggested that the molecular mechanisms and intrinsic properties of exosomes may have potential utility in the therapeutic control or diagnostic evaluation of these diseases. Also, exosomes can be engineered or programmed to deliver therapeutic agents and immune modulators to specific targets. Moreover, they have been reported in all biologic fluids and found to be readily accessible via liquid biopsies, making them a potentially useful diagnostic and prognostic marker for clinicians. Exosomes have been shown to both promote and limit the aggregation of misfolded proteins in the brain, making them a potential player in both the pathogenesis and treatment of neurodegeneration and other neurologic disorders. Their unique intrinsic property of being able to efficiently pass through the **blood–brain-barrier**, a vascular network functioning as a selective filter to keep drugs or toxins from reaching the brain, make exosomes a key therapeutic agent of interest for researchers and clinicians.

Table 1.18–5.
Human Diseases Arising from Mitochondrial-Encoded Genes

Disorder	Affected Gene(s)
NARP (Neuropathy, Ataxia, Retinitis, Pigmentosa)	ATPase6
MELAS (Mitochondrial Encephalomyopathy, Lactic Acidosis, Stroke)	tRNA (Leu), Cytochrome C
MERRF (Myoclonic Epilepsy; Ragged Red Fibers)	tRNA (Lys), tRNA (Ser)
LHON (Leber hereditary optic neuropathy)	MTND4
Friedrich ataxia	FRDA
Leigh syndrome	ATPase6, MTND5

HUMAN GENETIC VARIATION

The whole-genome sequencing of a reference human in 2001 set the stage for intensive studies of genetic variation between individuals and among populations. Generally, between any two unrelated persons, differences are present at approximately 1 in every 1,000 base pairs, that is, 0.1% of their genomes. It also appears that not everyone has two copies of every part of the genome. These larger regions of variability (>1 kb) occur in heterochromatic and euchromatic regions and are estimated to cover 5% to 20% of the total human genomic sequence. The prevalence of these variants has dramatically altered thinking about the total burden of variation within the human species. The search for disease susceptibility genes in psychiatry may be more precisely thought of as the hunt for that small proportion of genetic variation that is either causal or contributing to psychopathology.

Sequence Variation

DNA sequences of individuals vary in terms of the genetic code at the nucleotide level. Such changes may involve the substitution of one nucleotide for another or, in some cases, the insertion or deletion of one or a small number of bases or **indels.** In addition, short repetitive sequences in the DNA may differ in terms of the number of nucleotides present in these regions, variations that are known as **short tandem repeats** (or STRs) and **small sequence repeats** (SSRs).

There are varying thresholds for classifying a variant as common. Typically, if a sequence variation is found in less than 1% of the population, it is considered rare, though some authors use a threshold of 5%. Specific genetic variants found above this frequency in a given population are typically referred to as **polymorphisms**, but this term may also be applied to any change in the genome regardless of its frequency and regardless of whether or not it is deleterious to the function of the RNA or protein that it encodes or regulates. Similarly, any permanent change in the DNA nucleotide sequence can be called a **mutation,** although this term is most commonly applied to disease-related genetic variation.

Common single-base substitutions with a population frequency greater than 1% are often called **SNPs**. These have been intensively studied and proven critical to gene discovery efforts. SNPs are thought to have occurred spontaneously in human history, to have only happened at one point in that history, and to have distributed subsequently throughout a population over time. SNPs are generally stable from generation to generation, and a very large number have accumulated in the genome as the human species has evolved. Currently, 30 million known SNPs have been identified. These individual SNPs, or the arrangement of SNPs on a single chromosome (called **haplotypes**), allow scientists to trace inheritance within families or to identify variations that may carry risks for disease within a population. In addition, the distribution of SNPs and haplotypes among various populations has provided critical insight into human evolution. For instance, this type of data along with other genetic and archeologic evidence has provided strong evidence for the theory that humans originated from African hominids.

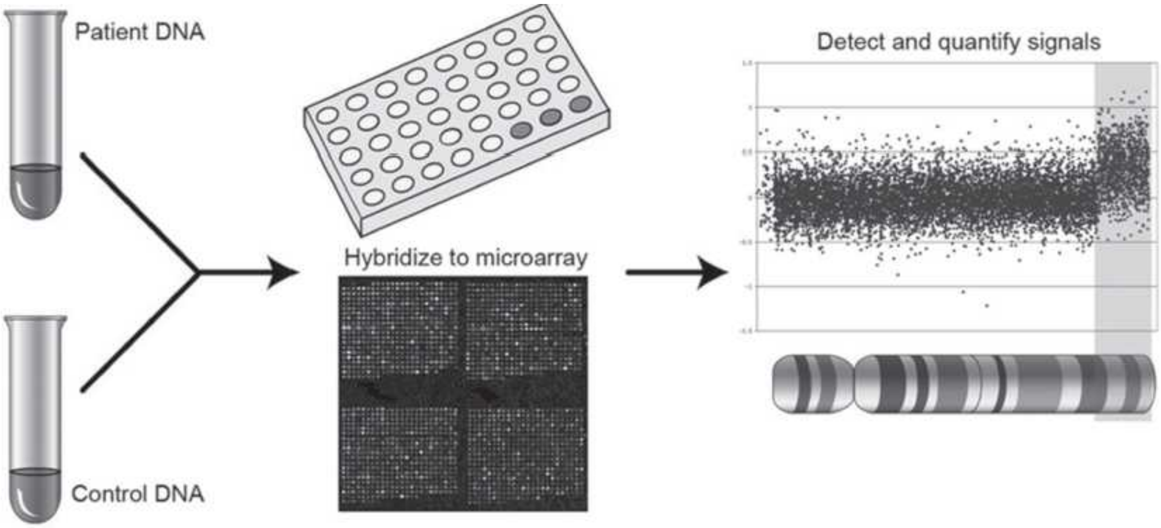

FIGURE 1.18–7. Detecting Structural Variation in the Human Genome. Using microarrays, which hold thousands of spots of DNA from the human genome, a comparison between amount of patient DNA and control DNA along the genome can be made. If a region of a chromosome is altered (upward shift in intensity, in red along the chromosome), it is then defined as a CNV.

Structural Variation

While large microscopic variations including balanced and unbalanced translocations, large deletions or inversions, and extra chromosomes, have been studied for decades, only recently has it become apparent that there is also a tremendous amount of submicroscopic structural variation in the genome.

In the late 1990s, several labs developed tools to identify chromosomal aberrations that were below the resolution of the light microscope. Given the known relationship between chromosomal abnormalities and various human disorders including cancer and mental retardation, it was anticipated that the ability to identify formerly unseen chromosomal anomalies would lead to the identification of additional disease susceptibility genes. The surprising result of these efforts was the identification of differences in the number of copies of chromosomal segments found among the control samples used to validate the technology. Given the clear evidence that these changes were widespread among individuals without any obvious phenotype, these changes became known as **CNPs** or **CNVs.**

Subsequent studies have widely replicated these initial results and the development of higher resolution technologies has demonstrated that deviations from the expected two copies of each genetic locus are common throughout the genome. These variants are not replacements or changes in sequence (like mutations or SNPs) but are more comparable to indels, where regions of the genome are missing a copy (**heterozygous deletions**), missing both copies (**homozygous deletions**), or have extra copies (**amplification**). Healthy human beings possess from several dozen to several hundred CNVs and may carry as many as a thousand total CNVs. When the sum total of all the identified variants is considered (SNPs, CNVs, and larger DNA deletions and/or amplifications), as much as 20% of the genome has been shown to vary between individuals (Fig. 1.18–7).

Gene (and large genomic segments) duplications have resulted in many evolutionary changes during the evolution of hominid lineage. Trichromatic vision was developed in the primate lineage around 40 million years ago as a result of gene duplication. Such rearrangement was facilitated by the Alu repeats, which flank the opsin genes, allowing the duplication to occur. The genes coding for cytochrome P450 enzymes frequently contain CNVs in humans. They are key enzymes in drug metabolism and, as such, extremely important in the novel field of pharmacogenomics (discussed below).

The discovery of CNVs has had an enormous impact. Approximately 10% of the genes in the human genome, including genes influencing development, overlap common CNVs, suggesting that the human genome is able to tolerate haploinsufficiency (the loss of one of two haploid copies) to a far greater extent than previously anticipated. Several mechanisms lead to CNV formation. Some CNVs appear to have developed once in human history with subsequent distribution among populations. Other CNVs appear to be more dynamic. In this case, a region of the genome may be primed to lead to CNVs, but this process may occur randomly among different individuals and lead to multiple different CNVs in a single chromosomal region.

The relationship between CNVs and human disease is a topic of intense study. It is now clear that the relationship between genotype and phenotype with regard to structural variation is quite varied. Previously, if a deletion was identified in a patient being evaluated for mental retardation, it was presumed that the deletion was the likely cause of the clinical problem. Presently, the identification of a loss of a gene in such a patient, even one known to be involved in brain development, is no longer *prima face* evidence of causality. At present, it is generally presumed that: (1) some CNVs have no phenotypic consequences; (2) as with other forms of variation, some common CNVs will be found to contribute incrementally to complex, multigenic disease; and (3) some copy number changes, particularly those that are rare and or *de novo* (newly created in a single person), may carry large risks for some disorders. Given that the characterization of these types of variation in the human genome is in the early stages, distinguishing among these alternatives in a given patient or population remains a formidable challenge.

Common versus Rare Genetic Variation

Irrespective of whether a genetic variant is sequence-based or structural, once it is introduced into the human genome, one expects that it will be subject to natural selection. That is, changes that do not alter reproductive fitness may be readily passed from generation to generation and, over time, have the potential to become common. Alternatively, changes that result in reduced fitness are likely to be subject to purifying selection, leading the frequency of that variant to decline over time in the population.

The impact on fitness is only one of several forces that influence the frequencies of genetic variants. For instance, a new mutation is, by definition, rare until it has had time to disseminate into a population and be subject to selective pressures. Moreover, the history of ethnic populations and migration patterns can dramatically influence the dynamics of genetic variants over time. However, it is important to note that as social constructs, race, and ethnicity can complicate genetic studies involving common variance, oftentimes leading to spurious findings and conclusions.

One would presume that rare variants contribute to disorders that are lethal early in development and those that reduce fertility or otherwise lead to decreased reproductive fitness. One would conversely presume that common genetic variants contribute to disorders that appear later in life or for which there is not a negative impact on fitness. Common variants are also presumed to be implicated in conditions (e.g., sickle cell trait) where a positive effect (malaria resistance in this case) of a given genetic variation might counterbalance negative consequences. The competing impacts of positive and negative effects of a gene variant are called **balancing selection.**

The field of human genetics has a record of tremendous accomplishment in those disorders in which a single rare genetic change causes or dramatically increases the risk for a disease or syndrome. However, the task of clarifying disorders in which multiple gene variants contribute to the manifestation of disease has remained a daunting challenge. Two alternative, but not mutually exclusive, paradigms have emerged to account for the genetics of common complex disease: a common variant: common disease model and a rare variant: common disease alternative.

The former has been largely favored in mental disorders; for example, it is widely held that schizophrenia, major depression, bipolar disorder, and autism result from the combined effect of multiple common genetic variants, each with modest effect and interacting with environmental factors to exceed a biologic threshold. For these and other complex disorders, both modeling and early experimental evidence have essentially ruled out the contribution of a single gene of major effect, assuming that there is a single phenotype (with relatively little variation) for a particular illness. As most of the variation within a population is carried in common variants, it is thought that common disorders will likely reflect this underlying genetic architecture.

An alternative model posits that many individually rare variants with relatively large effects contribute alone or in combination to the risk for common disorders. This hypothesis seems intuitive for disorders with early-onset and those that alter reproductive fitness. In the absence of balancing selection, one would expect that alleles with large effects, even if these are contributory and not causal, would be likely to be rare. In addition, one would expect a significant burden of rare mutations for disorders in which so-called sporadic or *de novo* variation played an important role. Indeed, recent studies show that *de novo* mutations (mostly *de novo* missense and *de novo* likely gene-disrupting mutations) in children affected by autistic spectrum disorder (ASD) may account for more than 10% of the diagnosed cases.

These two possibilities are not mutually exclusive. As more research is completed, it is likely that both common and rare variants will be found to contribute to the variable manifestation of many common mental disorders. However, the distinction is of tremendous importance with respect to current genetic studies because the methodologies employed to show a causal or contributory role for genes differ tremendously in terms of their ability to detect the contribution of rare versus common genetic variation. In addition, diseases with complex and extremely variable phenotypes, such as schizophrenia or ASDs, may be caused by various combinations of rare and common variations, which would, at least in part, explain the large spectra of pathologies seen in the aforementioned conditions.

THE FUNCTION OF THE GENOME

Now that the sequence of the human genome has been elaborated, research now focuses on how its 3 billion nucleotides work in a coordinated fashion. Given that the human genome appears largely unremarkable compared to other species in terms of size, gene number, or percentage of the genome devoted to protein-coding, one must presume that the capabilities of the human brain must be powerfully influenced by variation in a relatively limited number of genes or influenced by factors other than raw DNA sequence. The characterization of brain-specific functional components of the genome and the consequences of genetic variation in these elements will undoubtedly become a prime concern for psychiatric genetics in the coming era. More broadly, such an understanding will identify those unique molecular changes that define *Homo sapiens* as a species.

As recently as the early 1970s, the importance of gene regulation for shaping the development of the human brain remained in question. Several hypotheses were entertained regarding the origin of the differences between humans and other closely related species. It was thought that speciation might be the result of small-scale sequence changes (DNA or protein sequence chances), regulatory changes (from controller genes or other similar elements), or large-scale sequence changes (gene and genome duplication). In 1975, Mary-Claire King and Allan Wilson published a paper in which they sequenced and compared 44 human and chimpanzee genes. To the surprise of many, the vast majority of these sequences were found to be nearly identical among chimp and human and virtually all of the small-scale changes identified were synonymous. Moreover, the gene and genomic organization of the transcripts was essentially unchanged between the two species. These data demonstrated for the first time that neither coding sequence nor genomic organization was likely to account for the differences between the two species.

King and Wilson's work supported the alternative conclusion, that is, the sequence surrounding genes that regulated gene expression accounted for key evolutionary changes that include dramatic differences in cortical architecture. This work also implied that sequencing of the human genome and that of closely related species would not be sufficient for understanding key differences in their biology. These conclusions largely have been borne out: it is clear that in order to understand brain function, the question of whether a gene sequence is present is often not as important as determining *when* and *how* the transcript is active in that organism. In cases of highly divergent neurodevelopmental disorders such as autism, it has been shown that there are distinctive spatiotemporal expression patterns of the ASD-associated genes in early developmental stages of the brain. Aberrant spatiotemporal expression of these genes could easily contribute to developmental impairments, resulting in autism.

Yet another confirmation of King and Wilson's work came up recently with obtaining the genomes of the Old World Apes. Namely, several genomic regions that underwent rapid evolution in humans were identified —human accelerated regions (HARs). While HAR1 is a gene expressed during the development of the neocortex, and HAR2 (also known as HACNS1) is a human-specific developmental enhancer, another 721 HARs were identified by comparing conservation scores of the sequences. Surprisingly, 92% of them are noncoding genes, emphasizing the likely importance of regulatory sequences in recent human genome evolution.

REGULATION OF TRANSCRIPTION AND THE TRANSCRIPTOME

Perhaps the largest contribution to the complexity of eukaryotes comes from **RNA splicing**, the process by which the exons of a gene can be combined in multiple ways (Fig. 1.18–8).

The discovery that genes were transcribed as both exons and introns came from studying the adenovirus gene *hexon* in 1977. The implications of this discovery for human evolution were quickly recognized: a single region of DNA could lead to a variety of transcripts and multiple versions of a protein. These, in turn, could become targets of evolutionary selection that might generate new functions without having to sacrifice the utility of the original transcript. Thus, the “universe” of the transcriptome suddenly became much larger.

During RNA splicing, exons can either be retained in the mature message or targeted for removal in different combinations to create a diverse array of mRNAs from a single pre-mRNA, a process referred to as **alternative RNA** splicing. By 1994, estimates of the number of human genes exhibiting alternative splicing ranged from 1% to 5%. However, these estimates were known to be imprecise due to the sparse availability of sequence data. Interest in resolving this issue led to a portion of the HGP being devoted to studying the relationship between gene structure and gene expression, and two methods were utilized: **complementary DNAs** (cDNAs) and **expressed sequence tags** (ESTs). Both methods depend on the mRNA being processed first, which removes all the introns and then adds a long polyadenosine (poly-A) tail to the initial transcript. The presence of these poly-A tails provides a means for transcripts to be bound and isolated in the laboratory by running the total extracted RNA through columns packed with complementary poly-T sequences. Once isolated, the single-stranded mRNAs serve as a template in a reaction to create the complementary sequence, using an enzyme called reverse transcriptase. This creates the cDNAs, which can then be sequenced. ESTs are small fragments of cDNAs and they are slightly different because they use cloning techniques to digest the cDNA and merge the sequence with a bacterial fragment in order to discern the sequence. Once created and analyzed, these EST and cDNA libraries confirmed a massive amount of alternative splicing in the genome and showed that within genes, some exons are always used (constitutively spliced) while others are very rare and specific to a certain time or place in the body (alternatively spliced).

As these sequence libraries were being created, other technologies simultaneously emerged to allow researchers a clearer view of gene expression and splicing. In 1995, Pat Brown and colleagues first measured expression for 45 genes in parallel in the plant *Arabidopsis* by creating a microarray. They placed complementary sequences to the 45 genes in small spots on a glass microscope slide and then isolated mRNAs from two different *Arabidopsis* samples. The relative expression of each of the 45 genes could be determined between the two samples based on the fluorescence intensity determined at each spot on the slide.

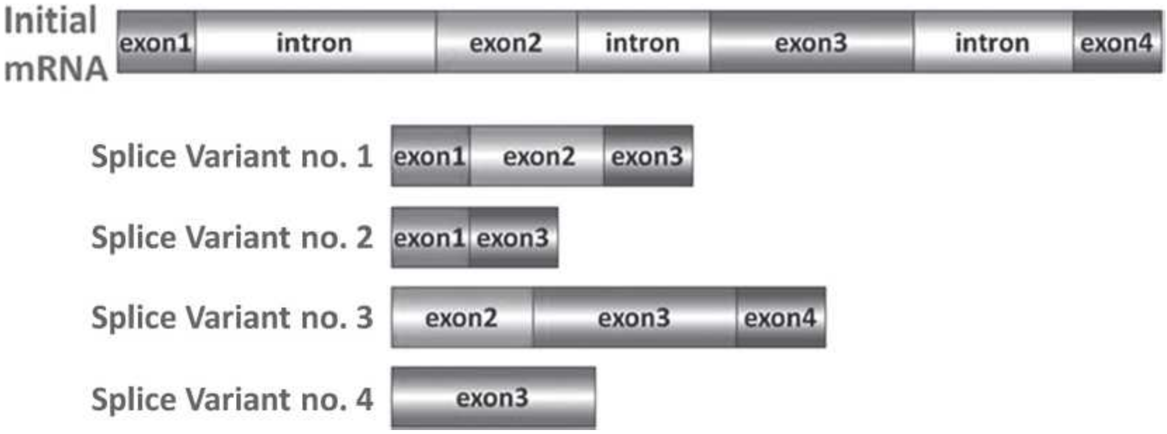

FIGURE 1.18–8. **Alternative Splicing in Eukaryotic Genes**. Different variations of certain multi-exon genes can be generated by the cell, depending on environmental conditions or developmental stage. These different splice variants will then create a slightly modified version of the gene's protein.

Since this time, a series of technical and methodologic developments have led to an astonishingly rapid increase in the density and specificity of the probes (spots) that may be placed on microarrays. Current technologies allow for millions of sequences to be arrayed on a single slide, providing the basis for comprehensive and global studies of gene expression within and among organisms. With hundreds of thousands of cDNAs and ESTs isolated, the current estimate is that at least 80% of human genes undergo alternative splicing.

While the size and gene number found in an organism do not generally correlate with the biologic sophistication of an organism, such is not the case for alternative splicing. At the most fundamental level, this distinction is made clear by the contrast between prokaryotes (which have no alternative splicing) and eukaryotes (which do). There is also some evidence that within higher organisms increases in splicing rates correspond to greater complexity. For instance, in yeast, most genes (~96%) have no introns and very few splicing events have been discovered. Moreover, recent evidence shows that invertebrates have fewer genes undergoing alternative splicing (50% to 60% in the fruit fly) than do vertebrates (80% in humans).

While the total amount of alternative splicing in the human genome remains a question, it is already clear that the brain shows the largest amount of any tissue in the human body. This observation suggests that alternative splicing and its regulation may be quite important for understanding what makes the human brain so different from other closely related species and in understanding the neurobiology of mental disorders.

Indeed, the process of alternative splicing bestows an opportunity for genes to undergo accelerated evolutionary change. Most genes are under strong negative selection pressure, meaning that it is difficult for them to change, especially if they are critical to the early development of an organism. Alternative splicing leading to the formation of a gene's less common products can be an evolutionary shortcut that relaxes this negative selection pressure. Since enough of the major (or more common) form is still present, the new, minor splice form would not be subject to the same degree of negative selection and may even create a "neutral space" for the novel gene product to acquire a new function for the organism.

Ample evidence of this process can be found within the DNA sequence. The exonic constituents of the major versus minor forms of a given gene are largely conserved between the mouse and human genomes. The major-use exons typically show higher degrees of sequence conservation than minor-use exons, suggesting stronger negative selection on the former. Cross-species comparisons demonstrate that minor-use exons are a more recent addition in evolutionary time, supporting the hypothesis that the basic biologic functions of a gene can be augmented through the addition of new splice variants. There is strong evidence that when a minor-use exon subserves a new important function, its level of conservation quickly approaches that of more ancient exons. These phenomena take place even if the new splice variant is highly restricted temporally or spatially, and this is amply demonstrated in the human brain.

It has often been the case that researchers have focused on changes that affect the most highly conserved regions of a gene or the most highly conserved amino acids in a protein sequence based on the notion that these underpin the most vital functions of the transcript. Recent cross-species comparisons and the enhanced knowledge of splice variation suggest that it is also possible that changes in less-conserved sequences or those that alter only minor splice variants that are spatially or temporally restricted could be quite relevant.

An appreciation of the importance of splice variations points to the critical nature of its regulation. As with many other aspects of the genome, the mechanisms of the process are incompletely understood. It has long been appreciated that there are canonical splice sites that define the boundaries of introns: the 5′ splice site typically has the sequence GT and the 3′ site most often has the sequence AG. However, these are not absolute as they are absent in approximately 2% of known introns. It has been appreciated that regulation of gene splicing is also influenced by sequences nearby or within the gene. These include **exonic splicing enhancers** (ESEs) and **exonic splicing silencers** (ESSs), which are specific sequences contained within exons that regulate splicing proliferation or attenuation. Changes in these sequences can have a profound impact on gene function without altering the amino acid composition of a protein. Synonymous sequence mutations, that is, those that alter the nucleotide sequence but do not change the amino acid sequence, involving these motifs have been shown in some cases to lead to serious developmental disorders.

In addition to the accumulating knowledge on gene splicing, several other aspects of regulating the transcriptome have been well characterized. Individual genes are regulated by transcription factors, which determine the timing and degree of a gene's activity. Transcription factors bind to specific sequences upstream of genes called TFBSs or **response elements** (REs) (Fig. 1.18–9). Most transcription factors consist of a DNA-binding domain that confers some target specificity and an activation domain that confers its regulatory function.

RNA Editing

Another way of regulating and diversifying the transcriptome is RNA editing. RNA editing is one of the most unusual post-transcriptional RNA modifications since it can modify the information content of an organism without changing the underlying DNA in the genome. It is mediated by the ADAR type of enzymes, deaminating adenosine (A) and turning it into inosine (I), which is then recognized as guanosine (G) by both the spliceosome and the ribosome (Fig. 1.18–10). This type of transcriptome diversification contributed to evolutionary changes in many cases instead of gene duplication/mutation or alternative splicing, possibly due to its efficiency and low energy consumption. RNA editing may produce changes in the coding sequence, miRNA binding sites, and may create novel splicing sites. Until recently, only a small number of edited human genes were discovered and most of them have important roles in the CNS. Brain shows the most versatile repertoire of RNA editing events, including AMPA, HTR2C receptors, sodium channels, and calcium channels. Humans have the highest amount of A to I editing as compared to rodents, rhesus macaque, and chimp, which may contribute to human traits.

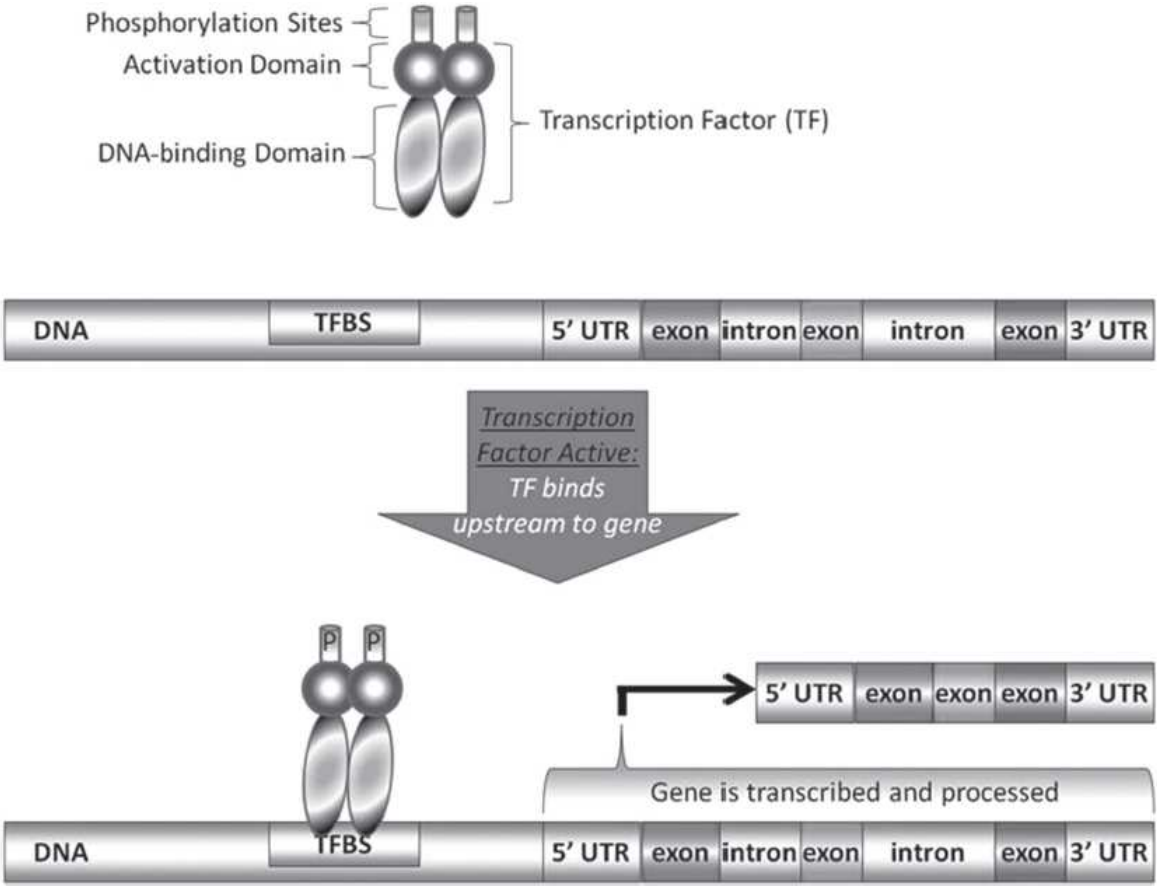
Phosphorylation Sites
Activation Domain
DNA-binding Domain
Transcription Factor (TF)
DNA
TFBS
5' UTR
exon
intron
exon
intron
exon
3' UTR
Transcription Factor Active:
TF binds upstream to gene
P
P
5' UTR
exon
exon
exon
3' UTR
Gene is transcribed and processed
DNA
TFBS
5' UTR
exon
intron
exon
intron
exon
3' UTR

FIGURE 1.18–9. **Gene Regulation by Transcription Factors**. Once a transcription factor is activated in its activation domain, usually by phosphorylation (the P sites on the TF), the DNA binding domain is then activated and finds the binding site in front of the appropriate gene (TFBS). Once bound, the gene is then activated.

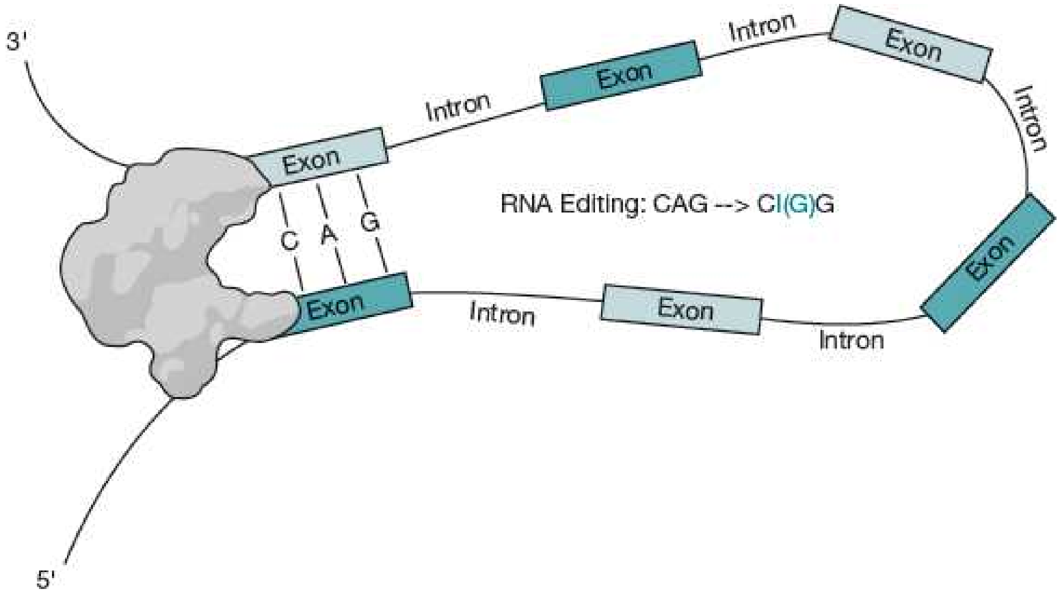

FIGURE 1.18–10. ADAR-Mediated RNA Editing. Double-stranded secondary structure of the RNA (*black line*) enables ADAR binding and deamination of the adenosine in the exon. Deamination derives inosine, which is then acts like guanosine in each interaction that requires Watson-Crick base pairing. Nascent CAG codon showed on the scheme, has been changed to CI(G)G. Created with BioRender.com.

RNA editing occurs almost exclusively on RNAs that form double-stranded secondary structures (dsRNA). Alu elements, due to their specific sequence, are a perfect target for RNA editing. Due to the fact that so many genes involved in brain functions undergo editing (calcium, sodium, potassium channels; GABA, serotonin, glutamate receptors), it is thought that RNA editing is a “fine tuner” of the brain transcriptome.

One of the best known eukaryotic regulatory elements is the TATA box, so named because the sequence TATA is present upstream of a gene. This motif binds a protein complex that allows for the initiation of transcription. The core promoter of individual genes need not contain all elements. Many promoters lack a TATA box and use instead of the functionally analogous **initiator element** (INR). These sequences are related to the number of protein-coding genes in the genome. Determining the number of coding genes in an organism is simply a matter of counting the number of TATA boxes and INRs in promoters that are nearby coding regions. Sequences like the TATA box, which are immediately upstream of a gene, are called **cis-regulatory** (meaning “on the same side”) due to their presence nearby the gene and on the same strand of DNA. Regulatory elements, TFBSs, or gene products that act on a gene from the other strand of DNA or from a distant region of the genome are called **trans-regulatory elements.**

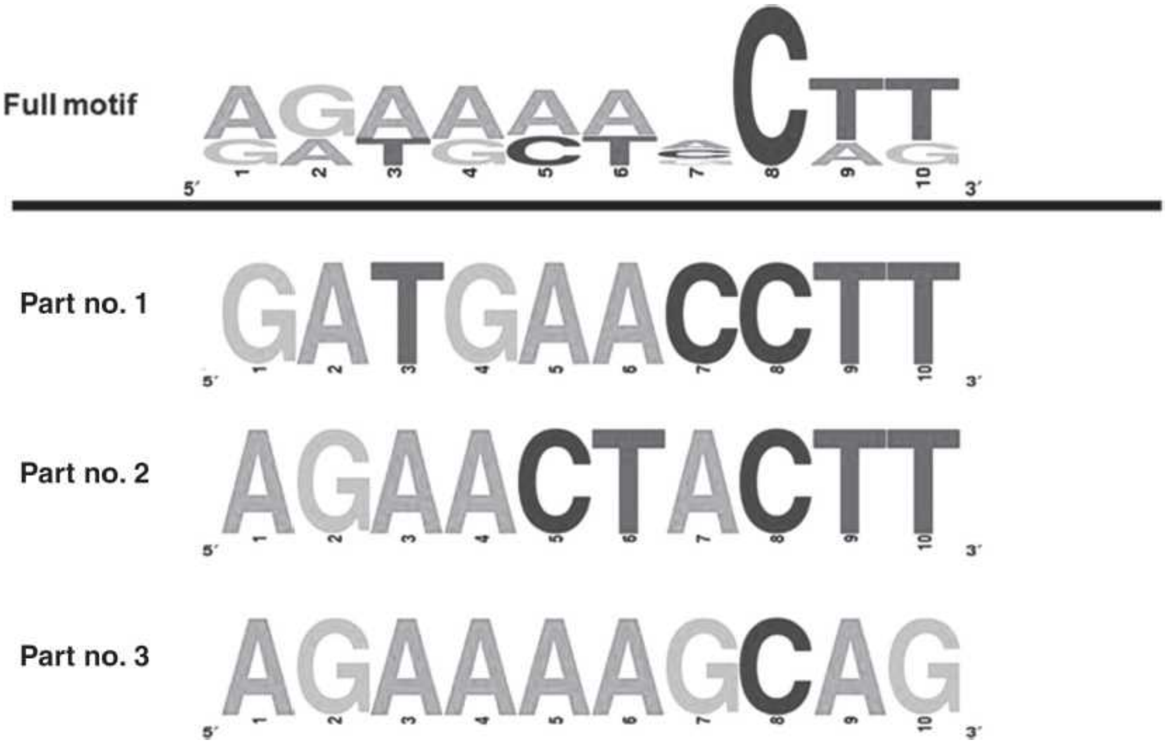

FIGURE 1.18–11. Building a Consensus Sequence for a Transcription Factor Binding Site (TFBS). For any TFBS upstream of many genes, there are often several different functional motifs (**Part #1, #2,** and **#3**) that allow for binding of that specific TF. When merged, these three 10bp sequences create a consensus motif that is degenerate across all nucleotides except for position 8, which is always a cytosine (C).

After the human genome was sequenced, this simple dichotomy of the regulation (cis- and trans) has become more complicated. The most well-understood promoter elements (TATA box and INR) are only present in the promoter regions of 54% of human genes, where promoters are defined within 100 bp in either direction of the **transcriptional start site** (TSS). This suggests that unknown promoter sequences exist to regulate the expression and function of these genes, or much of gene regulation occurs in trans-regulatory elements. A great deal of bioinformatics research is now focused on this very question.

Over 10 other core promoter elements are now predicted, but there is no longer an expectation that these elements will be universally present in all or even most genes. Rather, genes are activated in large simultaneous groups or come in multiple waves of activation, and these coexpressed genes can have a specific set of regulatory elements and modules that restrict expression to a certain time or place. Once a large set of coexpressed genes is found, an examination of the upstream regions of those genes can be performed to ascertain the sequences of the regulatory elements.

For a set of coexpressed genes, it is not typical to identify precisely the same regulatory sequence upstream of all of these transcripts. Due to the flexibility in the TF's DNA binding domain, there is often some sequence variability that is tolerated in gene regulation. Studies of many TFBSs have demonstrated that certain nucleotides in the binding site are critical, while others may vary. Aligning these sites across many genes results in the identification of a consensus sequence or consensus motif for a TFBS (Fig. 1.18–11).

Hundreds of transcription factors have now been studied with a technique called ChIP (Chromatin Immunoprecipitation). The dynamic process of TF binding to regulatory elements is frozen in a moment in time by fixing bound TFs to their cognate regulatory site via a chemical process known as **crosslinking**. Genomic DNA, some of which is now bound by TFs, is then fragmented in a manner that preserves this crosslinking. Using antibodies specific to the known TFs, the bound genomic sequences can be identified, separated from the rest of the genome (the unbound fragments), and then characterized. These bound fragments can be sequenced directly to identify where precisely a specific TF is functioning or may be labeled with different dyes and studied on a microarray. This latter method has become known as **ChIP-chip** (chromatin immunoprecipitation on a chip).

Table 1.18–6.
Effects on Gene Expression of Mono-, Di-, Tri-Methylation, As Well As Acetylation of Lysine for the Most Prominent Histone Markers

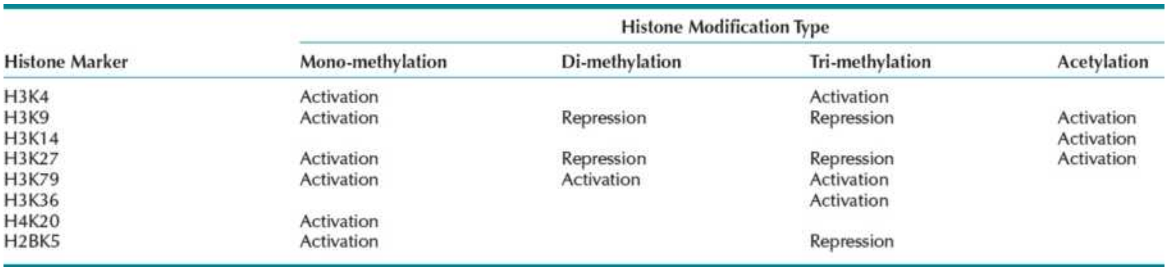

	Histone Modification Type			
Histone Marker	Mono-methylation	Di-methylation	Tri-methylation	Acetylation
H3K4	Activation		Activation	
H3K9	Activation	Repression	Repression	Activation
H3K14				Activation
H3K27	Activation	Repression	Repression	Activation
H3K79	Activation	Activation	Activation	
H3K36			Activation	
H4K20	Activation			
H2BK5	Activation		Repression	

These experiments identified a large amount of regulatory complexity in the human genome. The position of the TSS, where the beginning of transcription occurs for a gene, can sometimes be found hundreds of thousands of bases (>100 kb) away from the first exon. Also, many of these regulatory elements, TSSs, and TFBSs for genes are not found beside their respective gene but within introns or within different genes altogether. Sometimes these regulatory elements allow for genes to behave as discrete elements; however, at times, these regulatory motifs allow for a group of genes to merge together. There have been observations of trans-splicing, whereby exons from one gene can combine with exons from a completely different gene, creating chimeric mRNAs and chimera proteins that contain some or all of the elements from the two independent proteins.

Important points in understanding gene regulation include the following: any regulatory element in the genome should not be presumed to be regulating the gene to which it is adjacent; multiple genes may be activated in concert by a single, distant regulatory element; and a single gene is not always a single gene, and can (if needed) merge with some or all of the parts of another gene, even at relatively great distances.

Epigenetics

One additional realm of genetic regulation rests outside of the genomic sequence in the **epigenome**. Epigenetic regulation is any heritable change in the genome that does not change the nucleotide sequence. These changes can occur by either a chemical modification to the DNA known as methylation or by chemical changes to the protein complexes of histones. Histone modifications are far more numerous, with phosphorylation, acetylation, methylation, and ubiquitination modulating the activity of a gene or group of genes by altering DNA-protein complex conformation. This, in turn, has an impact on DNA availability to transcription factors and the recruitment of other molecules that inhibit or enhance gene expression.

Epigenetic marks on DNA consist of methylation of cytosine by specific methyltransferases, thereby deriving 5-methylcytosine (5 mC). 5 mC presence is predominant at regions of the genome that contain neighboring CG tandems—CpG sites. Cytosine methylation of CpG sites located at the gene promoters is a hallmark of transcriptionally silenced genes. DNA methylation is a reversible process and it is thought that the oxidative product of 5 mC, 5-hydroxymethylcytosine (5hmC), is an intermediate in the conversion of 5mC back into cytosine. The presence of 5hmC in germ cells is in favor of this model, as demethylation (resetting of the epigenetic code) is needed for the production of gametes. In addition, 5hmC shows high concentrations in CNS, with novel data suggesting roles of 5hmC in regulating splicing and gene expression in the brain.

Important histone modifications and their effect on the gene expression are presented in Table 1.18–6.

Probably the most reliable signs of gene repression are the presence of H3K27Me3 at gene promoters as well as H3K9Me2/3, which is a strong mark for the presence of heterochromatin at the specific location. In contrast, H3K4Me3 presence at the promoters is a sign of open, actively transcribed chromatin regions, while H3K36Me3 marks during the length of the gene body are also an indication of an actively transcribed region.

Epigenetic programming allows for tissue-specific regulation of gene function and structural and functional specialization throughout the body. Epigenetics also clearly helps explain differences between identical twins who share the same DNA sequence. This nonsequence-based heritable change can also lead to parent-specific effects during fertilization, known as **epigenetic imprinting**. These epigenetic markers are established or “imprinted” in the germline (sperm or egg cells) of the parents and are maintained through mitotic cell divisions in the somatic cells of an organism. Each person, when making eggs or sperm, imprints the epigenetic characteristics of their own sex since they must reprogram the epigenetic signature that was given by their own parents. Prader–Willi and Angelman syndromes are examples of diseases attributable to abnormal epigenetic imprinting.

Epigenetic programming not only regulates single genes but also can alter the behavior of entire chromosomes. X-chromosome inactivation is a normal process that occurs in every human female, in which one X chromosome is genetically silenced in every cell. This allows for dosage compensation, in which the amount of expression of any gene mapped to the X chromosome is essentially the same for males and females (even though males carry only a single X). The choice of which chromosome is active (X_a) and inactive (X_i) within a given cell is typically a random process. However, in some cases, a particular chromosome is more likely to be inactivated, contributing to aspects of Rett syndrome and other X-linked disorders.

Parent of origin effects, defined as the transmission of a risk for disease based on whether an autosome has been transmitted by the mother or father, has been hypothesized to play a role in several common mental disorders. The strongest evidence that this may be the case is the observation that duplications of chromosome 15q inherited from the mother, but not the father, accounts for a small but significant proportion of cases of autism.

Epitranscriptome

Another layer of complexity arises from modifications to the bases in RNA—the epitranscriptome. The current database list 112 forms of RNA modification. One type, methyl-6-Adenosine (m^6A), is found in all kingdoms of organisms. m^6A is enriched within 3′ UTRs and long exons in both mice and humans. Sequences enriched with m^6A are evolutionary conserved. Preliminary analysis showed that m^6A may hinder RNA editing. Also, as m^6A can be demethylated back into adenosine, it may constitute a mechanism for regulating RNA editing. m^6A is implicated in the modulation of the circadian rhythm, perhaps by influencing the trafficking of circadian cycle regulatory genes through the nuclear envelope. In addition, it may play a role in the regulation of translational dynamics for transcripts enriched with m^6A. These functional relationships suggest that post-transcriptional modifications of RNA may modify a transcript's properties and function, generating the so-called epitranscriptome.

PROTEIN REGULATION AND THE PROTEOME

Just as gene expression is regulated at multiple steps, so too can the protein products specified by coding genes (Table 1.18–7). Over 20 known modifications may occur to proteins. These modifications include chemical modification of specific amino acids, exchanging one amino acid for another, attaching other biochemical functional groups, attaching larger molecules (e.g., lipids or carbohydrates), or making structural changes to a protein. Even after a protein is fully formed and biochemically active, it can still be selectively targeted for degradation.

These types of protein modifications play an important role in neurotransmission since most neurotransmitters are shuttled across synaptic membranes and clefts by proteins that have been specially created to accomplish this task. For example, the protein that transports dopamine across synaptic membranes, the DAT, undergoes a range of posttranslational changes that are specific to its needed function in each cell.

Table 1.18–7.
Regulatory Mechanisms across the Genome, Transcriptome, and Proteome

	Regulatory Mechanism
Epigenetic	Allelic exclusion (imprinting one allele of a gene, so only the other is expressed)
	X chromosomal inactivation
	Chromatin re-modeling and chemical alteration (methyl, acetyl, phosphate)
	Short-range cell-to-cell signaling
Pretranscriptional	Binding of tissue-specific transcription factors
	Binding of competing, inhibitory factors
	Hormones binding to response elements in inducible genes
	Use of different transcription start sites, stop sites, and promoters
Post-Transcriptional	Alternative splicing within the same gene
	Tissue-specific RNA editing
	Translational control mechanisms
	Interference and degradation by siRNAs
	Trans-splicing between different genes
Post-Translational	Proteolytic cleavage
	Protein merging into a larger protein complex
	Glycosylation
	Targeted for degradation (ubiquitination)
	Chemical alteration (methyl, acyl, phosphate, or sulfate groups)
	Addition of heavy metals

To measure and understand these changes for each protein, a new discipline, **proteomics**, has emerged that examines each individual protein's various chemical modifications and how those proteins interact with each other to create larger cellular complexes and machinery. Components of the proteome have been the focus of intensive study, for example, all the active and interacting proteins in a synapse are dubbed the **synapse proteome,** and those in the brain are termed the **brain proteome**. All of these protein characterizations and interactions require extensive experimental support, large datasets and overlapping experimental methods.

Mass spectrometry is useful for rapid cataloging of all proteins present within a sample, such as the ~1,000 proteins in the synapse proteome (~1,000 proteins). High-pressure liquid chromatography (HPLC) coupled with mass spectrometry has the capability of resolving the full proteome of the brain. Two-dimensional (2D) electrophoresis is used to study protein interactions. In this approach, proteins are pushed across each other's path in a gel with a change in the migration pattern suggesting a direct interaction. Yeast two-hybrid experiments are another approach to characterizing protein binding partners. In this case, two proteins of interest are cloned into a yeast cell. Each protein is a hybrid, one containing the first protein and the DNA-binding domain of a specific transcription factor, while the other contains the second protein and the activation domain of the same transcription factor. If the two proteins interact, the transcription factor will have both necessary components to activate its target reporter gene, which is typically designed to be easy to assay. More recently, **protein chips**, which are similar to expression microarrays, have been created. These chips place thousands of proteins on a microscope slide and then they are washed over by a single sample, thus detecting every interaction between a protein and the rest of the queried proteome.

All these methods have enabled the discovery and characterization of the proteins made by the ~25,000 genes in the human genome. Current estimates suggest that from this relatively small number of coding genes, as many as 2 million distinct proteins may result. Several groups have emerged to tackle the task of accurately cataloging all of these proteins, including the Human Proteome Initiative (HPI) and the Human Proteome Organization (HUPO).

In summary, the complex machinery that regulates the genome can be thought of as encompassing four levels: pretranscriptional regulation (exemplified by the functioning of transcription factors), post-transcriptional regulation (exemplified by miRNAs and siRNAs), post-translational regulation (exemplified by chemical modification and processing of proteins), and epigenetic regulation.

If the notion of understanding how these processes contribute to the complex functioning of the human brain and how they might contribute to mental illness was not sufficiently daunting, one need only consider the astronomical potential for diversity inherent in the transcriptome and proteome: Though the human genome is only 3 billion letters long, for every gene with more than one exon, the number of possible transcripts increases at the rate of $2^n - 1$. Consequently, a two-exon gene can be spliced into a product containing each one of the two exons, or both combined, totaling three possible transcripts. Similarly, a three-exon gene can be spliced into one of the three exons, or four other combinations (1–2, 2–3, 1–3, 1–2–3), totaling seven, and so on. When this formula is applied to the gene with the most exons in the human genome (*titin,* with 234 exons), one can see that this one coding unit could theoretically combine into 2.76×10^{70} different transcripts. In addition, given about 2 million proteins in the human proteome, there exist 4 trillion possible protein–protein interactions. In total, this becomes about 7.0×10^{82} possible gene products and subsequent interactions to examine. As a comparison, the number of atoms in the universe is estimated to be around 4.0×10^{80}. Though these numbers are theoretical approximations, they illustrate the vast potential complexity of the human genome and the need for sophisticated methods to understand it.

PHARMACOGENOMICS IN NEUROPSYCHIATRIC DISORDERS

Pharmacogenomics is the effort to characterize gene variants that influence the response to medications with the long-term goal of optimizing the treatment of individual patients. Pharmacogenomic tests are already established in several different fields of clinical practice including oncology, infectious disease, anesthesiology, gastroenterology, cardiology, gynecology, rheumatology, neurology, and psychiatry. The Federal Drug Administration (FDA) maintains a table of pharmacogenomic biomarkers in drug labeling that covers over 350 drug-gene combinations across over 75 genes.

In psychiatry, pharmacogenomics has had its greatest impact through characterizing gene variants coding for P450 microsomal enzymes that influence the rate of metabolism of psychotropic medications. These tests have been applied clinically as an alternative to measuring drug–blood levels in order to identify whether a lack of clinical response could be attributed to rapid drug clearance, necessitating treatment with higher medication doses. They might identify slow metabolizers, who might be more suspectable to medication side effects. To date, pharmacogenomics studies related to the pharmacodynamics actions of psychotropic medications do not yet have sufficient validity or explanatory power for routine clinical use but may become more common in the future.

THE BRAIN AS A SYSTEM

The field of **systems biology** focuses on the emergent properties of biologic systems. There is increasing interest in developing sophisticated computational approaches that enable scientists to evaluate the interplay of genome, transcriptome, proteome, and other emergent properties of the brain, such as behavior.

Systems biology is a particularly powerful approach to extraordinarily large datasets generated by large-scale international collaborations aiming for a complete functional annotation of the human genome, transcriptome, and proteome. These efforts have been made possible by technologic advances that have allowed for the accumulation of very large datasets that encompass all of the types discussed in this chapter: regulatory sequences, the transcription factors that bind them, the expressed genes (coding and noncoding), their subsequent modifications by other enzymes, the final protein product(s), and the interaction of that protein with any other protein (**protein–protein interactions** [PPIs]). These data are derived from three sources: (1) computational predictions of transcriptional activity, splicing potential, or protein interactions; (2) experimentally based observations extracted from published scientific reports and databases; and (3) the use of algorithms that scan all of the world's published scientific literature and report any interaction with a specific gene or protein of interest (**text mining**). In addition, large databases of protein sequences have been compiled and curated, such as UniProt.

Over the past several years, major advances in bioinformatics and computational biology have allowed an entire biochemical or neurologic pathway to be modeled and a change in one segment of the pathway (from a mutated exon to a drop in synaptic plasticity) to be traced to every other possibly interacting biologic entity. This technique—known as **pathway analysis**—creates predictions for expected changes to other genes or proteins in the same pathway. Already, several such tools are publicly available online, including Ingenuity, DAVID, KEGG, and Interactome, an effort at Cold Spring Harbor Laboratory to build a working model for every interaction (gene–gene, gene–protein, protein–protein) that exists in the human body.

Table 1.18–8.
Online Scientific Resources

Description	Address
National Center for Biotechnology Information	https://www.ncbi.nlm.nih.gov/
Wellcome Trust Genome Research Center	https://www.sanger.ac.uk/
The HapMap Consortium	https://www.genome.gov/10001688/international-hapmap-project
National Human Genome Research Institute	https://www.genome.gov/
Protein Data Bank	https://www.rcsb.org/
The ENCODE Project	https://www.encodeproject.org/
UCSC Genome Bioinformatics Center	https://genome.ucsc.edu/
Personal Genome Project	https://www.personalgenomes.org/
Allen Brain Atlas from Mouse Brains	https://portal.brain-map.org/
Functional Annotation of Genes	http://geneontology.org/
The Human Protein Atlas	https://www.proteinatlas.org/
Database and Research Tools for Proteins	https://proteinlounge.com/index.php
Modeling of Biomolecular Interactions	http://interactome.org/

An extraordinary public resource has recently been created that catalogs region-specific expression analyses in the mouse brain, known as Allen Brain Atlas. This database currently contains a three-dimensional rendering of almost every gene's expression pattern. The mouse model serves as a good proxy for understanding region-specific brain expression in humans, and as a result of this comprehensive resource, essentially any gene of interest can be analyzed with all the others, for instance, to investigate the spatial and temporal correlation among groups of genes or across the entire genome. These and other resources are publicly available on the World Wide Web (Table 1.18–8).

FUTURE PROSPECTS

The brain is the most complex organ in the human body and perhaps the most complex structure known to humanity. Scientists have never before had so many tools and technologies to bring to bear to study the brain. These approaches create an unparalleled ability to gather and merge information about a biologic system, but the exact methods by which clinicians and researchers can merge patient data and abstract predictions is a work in progress.

Though sequencing, microarrays, and cloning have been indispensable techniques, they are already being eclipsed by newer, faster, and more high-throughput technologies. All of the methods currently used for analyzing and understanding an individual patient's genome are restricted to specific regions, but this is only because of the limitations of the current technology. Once each patient's genome can be sequenced cheaply and quickly, another revolution in the postgenomics era will begin. Instead of searching for mutations or structural abnormalities only in genes or places where a change is suspected, every small- and large-scale genetic variation will be detected for every location in the genome.

Clearly, the availability of a complete picture of each patient's DNA will vastly empower efforts to identify that proportion of genetic variation that contributes to mental disorders, influences therapeutic response, and mediates the results of environmental stressors. There is no doubt that recent discoveries have already heralded the first steps toward the development of personalized genomic medicine. The <$1,000 genome is a major step toward realizing this potential. The tremendous payoff from molecular genetics will greatly enhance the understanding of DNA sequence and structural changes; these important first steps will be critical in dissecting the molecular neurobiology of mental disorders and in developing novel therapeutics.

FURTHER READINGS

An JJ, Gharami K, Liao GY, et al. Distinct role of long 3′ UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell*. 2008;134(1):175–187.

Bass BL. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem*. 2002;71:817–846.

Capra JA, Erwin GD, McKinsey G, Rubenstein JL, Pollard KS. Many human accelerated regions are developmental enhancers. *Philos Trans R Soc Lond B Biol Sci*. 2013;368(1632):20130025.

da Silveira WA, Fazelinia H, Rosenthal SB, et al. Comprehensive multi-omics analysis reveals mitochondrial stress as a central biological hub for spaceflight impact. *Cell.* 2020;183(5):1185–1201.e20.

Drozda K, Müller DJ, Bishop JR. Pharmacogenomic testing for neuropsychiatric drugs: current status of drug labeling, guidelines for using genetic information, and test options. *Pharmacotherapy*. 2014;34(2):166–184.

Dulai KS, von Dornum M, Mollon JD, Hunt DM. The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates. *Genome Res*. 1999;9(7):629–638.

ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447(7146):799–716.

Frankish A, Harrow J. GENCODE Pseudogenes. *Pseudogenes*. Springer; 2014:129–155.

Gerstein MB, Bruce C, Rozowsky JS, et al. What is a gene, post-ENCODE? History and updated definition. *Genome Res*. 2007;17(6):669–681.

Jens M, Nikolaus R. Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat Rev Genet*. 2015;16(2):113–126.

Kalluri R, LeBleu VS. The biology, function, and biomedical applications of exosomes. *Science*. 2020;367(6478):eaau6977.

Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet*. 2007;8(6):413–423.

Kellis M, Wold B, Snyder MP, et al. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci USA*. 2014;111(17):6131–6138.

Khaja R, Zhang J, MacDonald JR, et al. Genome assembly comparison identifies structural variants in the human genome. *Nat Genet*. 2006;38(12):1413–1418.

King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. *Science*. 1975;188:107–116.

Lander ES, Linton LM, Birren B, et al.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921.

Levy S, Sutton G, Ng PC, et al. The diploid genome sequence of an individual human. *PLoS Biol*. 2007;5(10):e254.

Liu GE, Alkan C, Jiang L, Zhao S, Eichler EE. Comparative analysis of *Alu* repeats in primate genomes. *Genome Res*. 2009;19:876–885.

Mendel JG. Versuche über pflanzenhybriden. In: *Verhandlungen des naturforschenden Vereines in Brünn 4*. Abhandlungen; 1866:3–47.

Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA Methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*. 2012;149(7):1635–1646.

Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci U S A*. 2006;103(47):17973–17978.

Paz-Yaacov N, Levanon EY, Nevo E, et al. Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. *Proc Natl Acad Sci USA*. 2010;107(27):12174–12179.

Pipes L, Li S, Bozinoski M, et al. The non-human primate reference transcriptome resource (NHPRTR) for comparative functional genomics. *Nucleic Acids Res*. 2013;41(D1):D906–D914.

Plotkin JB, Grzegorz K. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genetics*. 2011;12(1):32–42.

Pollard KS, Salama SR, King B, et al. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet*. 2006;2(10):e168.

Pollard KS, Salama SR, Lambert N, et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*. 2006;443(7108):167–172.

Stolc V, Gauhar Z, Mason C, et al. A gene expression map for the euchromatic genome of Drosophila melanogaster. *Science*. 2004;306(5696):655–660.

Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001;291(5507):1304–1351.

Wang HY, Chien HC, Osada N, et al. Rate of evolution in brain-expressed genes in humans and other primates. *PLoS Biol*. 2007;5(2):e13.

Wilson FH, Hariri A, Farhi A, et al. A cluster of metabolic defects caused by mutation in a mitochondrial tRNA. *Science*. 2004;306(5699):1190–1194.

Zhang ZD, Paccanaro A, Fu Y, et al. Statistical analysis of the genomic distribution and correlation of regulatory elements in the ENCODE regions. *Genome Res*. 2007;17(6):787–797.

1.19 Gene Mapping Investigations of Psychiatric Disorders

SCOTT C. FEARS, M.D., PH.D.

The field of genetics essentially started from the rediscovery of Gregor Mendel's basic concepts at the turn of the 20th century and has matured into an essential cornerstone not only of the biologic sciences but also of all of medicine. The discovery of the basic structure and properties of DNA in the middle of the century led to an exponential acceleration in the current understanding of all aspects of the life sciences, including deciphering the complete sequence of the human genome and those of myriad other species. Massive databases of such sequences now provide 21st-century biologists with the task of decoding the functional significance of all this information. In particular, attention has turned to determining how sequence variations contribute to the phenotypic variation between species and between individuals within a species; in humans, it is hoped that discoveries about the relationship between genotypes and phenotypes will revolutionize the understanding of why and how some individuals but not others develop common diseases. This hope is particularly strong for psychiatry, as current knowledge of the pathogenic mechanisms of psychiatric disease remains sparse.

With a few notable exceptions, the field of psychiatric genetics struggled for many decades to identify and confirm genetic linkages and associations for mental health disorders. Recently, advances in high-throughput sequencing and organization of well-powered sample sizes through large-scale collaborations have started to deliver long-awaited results. Prior to 2010, no confirmed genetic variants influencing risk for schizophrenia had been discovered, but since then, over 250 risk variants have been identified and replicated in rigorously designed gene-mapping investigations. Similar success is beginning to emerge for other mental health disorders, including Alzheimer disease, autism, bipolar disorder, and depression. The growing list of genes associated with psychiatric disorders promises to redefine the current understanding of the biologic basis of mental disorders. The emerging catalog of gene-mapping findings does not support the popular theories of the last 50 years that have focused on alterations of monoaminergic systems as fundamental to the development of mood and psychotic disorders. Rather, genetic findings indicate that the etiologies of psychiatric disorders are linked to a complex network of partially overlapping cellular processes, including neurodevelopment, synapse functioning, ion regulation, immunity/inflammation, and intracellular signaling. Further, genetic analyses are confirming what has long been suspected; that there is substantial heterogeneity within categorically defined diagnostic groups and significant genetic overlap between different categories of disorders. In contrast to traditional genetic approaches that aimed to identify one or a few rare genetic variants with a large impact on the development of disease (i.e., Mendelian genes), much of the recent success has been gained by focusing on the identification of common genetic variants, each of which contributes only a small increase in disease risk. There are still major issues to be resolved, for example, deciphering the relative contributions of common versus rare genetic mutations; however, as a whole, the field is finally able to begin building an empirically rigorous foundation upon which many of the deep unknowns regarding the pathophysiology of mental disorders can begin to take shape.

Genetic mapping studies aim to identify the specific genes and sequence variations implicated in heritable diseases based on their chromosomal location. These studies are carried out by investigating affected individuals and their families through two approaches: linkage and association (Figs. 1.19–1 and 1.19–2). The two approaches have distinct advantages and are discussed in more detail below. Briefly, linkage methods are implemented within samples composed of related family members and are well suited for mapping genes with large effects on disease traits (high penetrance), whereas association methods are usually implemented in outbred populations (nonfamily members) and are designed to identify genes with small effects on the disease (low penetrance). It is now straightforward to use linkage methods to genetically map large-effect Mendelian traits, which are traits for which a specific genotype at one particular locus is both necessary and sufficient to cause the trait. Psychiatric diseases, however, do not follow simple Mendelian inheritance patterns but rather are examples of etiologically complex traits. Etiologic complexity may be due to many factors, including the following:

1. *Incomplete penetrance* in which the phenotype is expressed in only some of the individuals carrying the disease-related genotype.
2. *The presence of phenocopies*, which are forms of the disease that are not caused by genetic factors.
3. *Locus heterogeneity,* which occurs when different genes are associated with the same disease in different families or populations.

4. *Polygenic inheritance* in which risk for disease increases only if susceptibility variants at multiple genes act in concert.

Mapping a complex disorder involves several component steps, including the definition of the phenotype to be studied, epidemiologic studies to determine the evidence for genetic transmission of that phenotype, choice of an informative study population, and determination of the appropriate experimental and statistical approaches. This chapter will start with an overview of the epidemiologic methods used to identify genetically influenced traits and disorders and then describe the various approaches that are used to map the specific genetic variants associated with complex traits. The final part of this chapter will review the progress to date in mapping genes for several key psychiatric disorders.

EPIDEMIOLOGY

Genetic Epidemiologic Approaches

Prior to mapping specific genes, genetic epidemiologic investigations provide quantitative evidence regarding the degree to which a given trait aggregates in families and, furthermore, can suggest to what degree such aggregation reflects a genetic contribution to the etiology of the trait. Family studies compare the aggregation of disease among the relatives of affected individuals compared to control samples. Because these studies do not differentiate between genetic and environmental contributions to such familial aggregation, they provide only indirect evidence regarding the heritability of a trait. Often these studies measure the familial relative risk (λ), defined as the rate of occurrence of a disease among specified categories of relatives of an affected individual compared to the rate of occurrence of the disease for the general population. A relative risk of >1 suggests a genetic etiology and the magnitude of the measure gives an estimate of the genetic contribution to the disease. Relative risks can be calculated for sibling pairs, parent–offspring pairs, and various other types of family relationships. Likely modes of transmission can be assessed by comparing the degree of relative risk for each type of relationship. Multiple family studies have been carried out for many of the major psychiatric disorders, including major depression, bipolar disorder, schizophrenia, and obsessive compulsive disorder (OCD). While these studies have consistently reported familial aggregation for all of these disorders, the degree of such aggregation has varied substantially across studies, largely reflecting differences in phenotype definition and how study samples were ascertained and assessed (see "The Phenotype" section below for more details).

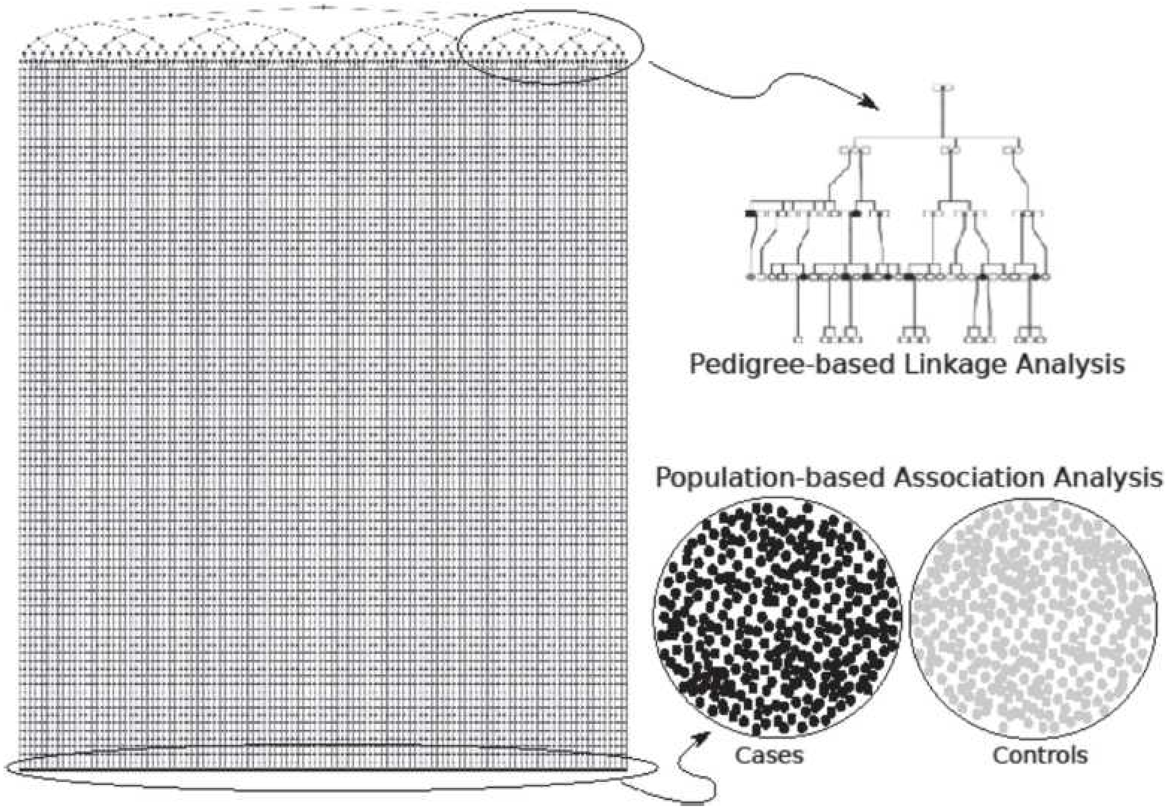
Pedigree-based Linkage Analysis
Population-based Association Analysis
Cases
Controls

FIGURE 1.19–1. There are two basic approaches to mapping genes: linkage (**top right**) and associations (**bottom right**). All gene-mapping approaches are based on identifying genetic mutations that are inherited identical by descent (IBD). The left side of the figure shows a schematic representation of a founder at the top center of the figure and all of their descendants inheriting the risk allele IBD. Linkage analysis (**top right**) samples multiple generations of individuals within a family and aims to identify cosegregation of genetic markers of known location with the disease trait in affected individuals. Linkage analysis is optimal for detecting genes of large effect in family members relatively close to the origination of the genetic mutation (before natural selection results in the removal of the mutation). In contrast, association analysis (**bottom right**) samples generations very distant from the founder and attempts to identify enrichment of specific genetic variants within affected individuals compared to controls. Genetic mutations that have survived for many generations tend to confer a relatively small increase in disease risk (which is generally expected to lower fitness), which is reflected in the small effect sizes of variants identified by association studies.

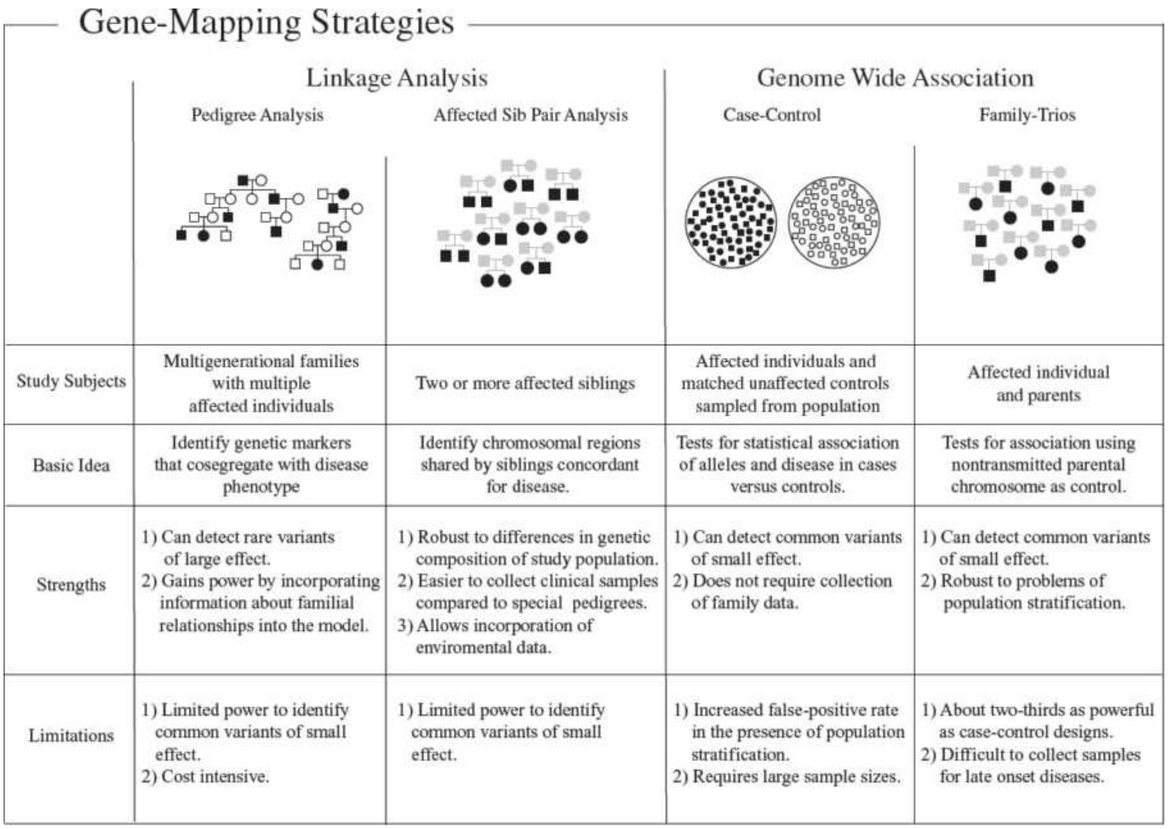

Gene-Mapping Strategies

	Linkage Analysis		Genome Wide Association	
	Pedigree Analysis	Affected Sib Pair Analysis	Case-Control	Family-Trios
Study Subjects	Multigenerational families with multiple affected individuals	Two or more affected siblings	Affected individuals and matched unaffected controls sampled from population	Affected individual and parents
Basic Idea	Identify genetic markers that cosegregate with disease phenotype	Identify chromosomal regions shared by siblings concordant for disease.	Tests for statistical association of alleles and disease in cases versus controls.	Tests for association using nontransmitted parental chromosome as control.
Strengths	1) Can detect rare variants of large effect. 2) Gains power by incorporating information about familial relationships into the model.	1) Robust to differences in genetic composition of study population. 2) Easier to collect clinical samples compared to special pedigrees. 3) Allows incorporation of enviromental data.	1) Can detect common variants of small effect. 2) Does not require collection of family data.	1) Can detect common variants of small effect. 2) Robust to problems of population stratification.
Limitations	1) Limited power to identify common variants of small effect. 2) Cost intensive.	1) Limited power to identify common variants of small effect.	1) Increased false-positive rate in the presence of population stratification. 2) Requires large sample sizes.	1) About two-thirds as powerful as case-control designs. 2) Difficult to collect samples for late onset diseases.

FIGURE 1.19–2. Genetic mapping approaches can be divided into those that rely on linkage analysis and those that rely on association analysis. Linkage studies can be further categorized as either focused on investigation of pedigrees or focused on investigation of sib pairs. Association studies can be categorized as either case control or family based. Some of the key features as well as advantages and disadvantages of these different approaches are shown in the figure.

Twin studies examine the concordance rates of a particular disorder (the percentage of twin pairs where both twins have the disorder) in monozygotic (MZ) and dizygotic (DZ) twins. For a disorder that is strictly determined by genetic factors, the concordance rate should be 100% in MZ twin pairs (who share 100% of their genetic material) and 25% or 50% in DZ twin pairs (who are no more closely related than any siblings), depending on whether the disease is recessive or dominant, respectively. For a disorder where genetic factors play a role in disease causation but are not the exclusive cause of disease, the concordance rates should be greater for MZ twins than those for DZ twins. The higher the degree of concordance of MZ twins, the higher the trait heritability or the evidence for a genetic contribution to disease risk. When genetic factors do not play a role, the concordance rates should not differ between the twin pairs, under the simplifying assumption that the environment for MZ twin pairs is no more similar than that for DZ twin pairs. The twin studies that have been conducted for traits such as autism, bipolar disorder, and schizophrenia have consistently suggested high heritability and have therefore spurred efforts to genetically map loci for each of these conditions. Different twin studies may, however, generate varying estimates for the heritability of any given disorder. When evaluating the results of twin studies, it is therefore important to scrutinize how the phenotype was ascertained because, as with family studies, the different heritability estimates are likely due to differences in the mode of assessing and defining phenotypes. For example, early twin studies of psychiatric disorders often relied for their phenotypes on unstructured interviews by a single clinician. In contrast, modern studies generally utilize standardized assessments and review of diagnostic material by a panel of expert clinicians. Similarly, part of the apparent variation in heritability between different twin studies can be attributed to the fact that some studies employ narrow definitions of affectedness for a given phenotype, while other studies employ broader phenotype definitions (e.g., considering a twin with major depressive disorder (MDD) to be phenotypically concordant with a cotwin diagnosed with bipolar disorder). Because of such differences in approach across studies, it is usually prudent to view such investigations as providing a rough estimate of the genetic contribution to trait variability. Nevertheless, even such estimates are useful in deciding which traits are likely to be mappable.

BASIC CONCEPTS OF GENE MAPPING

Introduction

Once genetic epidemiologic studies of particular phenotypes have suggested that these phenotypes are heritable, genetic mapping methods are implemented to identify the specific genetic variants that contribute to the risk of the disorder. The fundamental essence of gene mapping is to identify the order of genetic loci along a chromosome and determine the distance between them. Prior to the discovery of DNA as the molecular basis of genes, mapping studies typically sought to identify patterns of transmission in families between disease phenotypes and enzymatic or physiologic phenotypes like the ABO blood group to establish linkage. Current mapping studies start with an extensive set of DNA polymorphisms of known location and methodically traverse through the genome to identify the location of an unknown disease locus. There are two basic approaches to mapping genes: linkage and association (Figs. 1.19–1 and 1.19–2). Linkage analysis uses information from family members to identify loci that are transmitted together across generations more often than would be expected by chance. In contrast, association designs do not make explicit use of family information and aim to identify nonrandom co-occurrence of genetic variants of known location with disease traits. Linkage analysis is ideally suited to identify rare genetic variants that have a relatively large effect on the disease trait, whereas population-based association methods are designed to identify common genetic variants that have modest effects on the disease trait. The following sections will provide an overview of the basic concepts of gene mapping, starting with the phenomena of recombination and linkage upon which all mapping methods are based.

Recombination and Linkage

All genetic mapping methods aim to identify disease-associated variants based on their chromosomal position and the principle of genetic linkage. All cells contain two copies of each chromosome (called homologs): one inherited from the mother and one inherited from the father. During meiosis, the parental homologs cross over or recombine, creating unique new chromosomes that are then passed on to the progeny. Genes that are physically close to one another on a contiguous strand of DNA within a single chromosome are genetically linked (tend to be inherited together), and those that are farther apart or are on different chromosomes are genetically unlinked. Genes that are unlinked will recombine at random (i.e., there is a 50% chance of recombination with each meiosis). Genetic loci that are linked will recombine less frequently than expected by random segregation, with the degree of recombination proportional to the physical distance between them (i.e., the closer the physical distance, the more frequently loci will be transmitted together). The principle of linkage underlies the use of genetic markers, segments of DNA of known chromosomal location that contain variations or polymorphisms (described in more detail below). Strategies to map disease genes are based on identifying genetic marker alleles that are shared—to a greater extent than expected by chance—by affected individuals. It is presumed that such sharing reflects linkage between a disease locus and a marker locus, that is, the alleles at both loci are inherited “identical by descent” (IBD) from a common ancestor, and, furthermore, that this linkage pinpoints the chromosomal site of the disease locus.

The evidence for linkage between two loci depends on the recombination frequency between them. Recombination frequency is measured by the recombination fraction (Θ), which can be empirically determined by observing how often two traits or genetic loci are transmitted together and can be used to estimate the genetic distance between the two loci (1% recombination equals 1 centimorgan [cM] in genetic distance and, on average, covers a physical distance of about 1 megabase [mB] of DNA). A recombination fraction of 0.5% or 50% indicates that two loci are not linked but rather that they are segregating independently, whereas lower recombination fractions indicate that loci are sited near enough to each other that there is a tendency to be transmitted together: the lower the recombination fraction, the smaller the distance between the loci. In practice, a logarithm of the odds ratio (LOD) score is estimated from empirically derived observations to determine the likelihood that two loci are linked at any particular genetic distance. The central statistic for a LOD score is an odds ratio which is calculated by dividing the likelihood of acquiring the observed data if the loci are linked at a given recombination fraction by the likelihood of acquiring the data if the loci are unlinked ($\Theta = 0.5$). The log (base 10) of this odds ratio is the LOD score. A LOD score of 3.3 translates to an approximately 1,000-to-1 odds that two loci are linked rather than segregating independently. A LOD score can be obtained for various values of the recombination fraction, from $\Theta = 0$ (completely linked) to $\Theta = 0.5$ (unlinked). The value of Θ that gives the largest LOD score is considered to be the best estimate of the recombination fraction between the disease locus and the marker locus. This recombination fraction can then be converted into a genetic map distance between the two loci. In practice, large families and/or samples including a large number of smaller families are needed to estimate accurate LOD scores.

Linkage Disequilibrium

Linkage disequilibrium (LD) is a phenomenon that is analogous to family-based measures of linkage but is used to evaluate the genetic distance between loci in populations rather than in families. When alleles at two loci occur together in the population more often than would be expected, given the allele frequencies at the two loci, those alleles are said to be in LD. When strong LD is observed between two loci, it usually indicates that the two loci are sited in very close physical proximity to one another on a given chromosome and is useful in mapping disease susceptibility loci because one locus can be used to predict the presence of another locus. This predictability is important because current gene-mapping strategies are able to sample only a subset of the estimated 10 million common human polymorphisms. Because of the existence of LD, one can use data from a subset of genotyped polymorphisms to infer genotypes at nearby loci. Clusters of alleles that are in LD and inherited as a single unit are termed haplotypes. Thus, LD mapping "consolidates" genomic information by identifying haplotypes in populations that can then be used to infer IBD sharing among unrelated individuals (a central tenet of gene mapping using the association methods described below).

There are several methods to measure the extent of LD. One of the most commonly used measures of LD is r^2, a measure of the difference between observed and expected haplotype probabilities. A large r^2 value indicates that the observed frequency of association between two alleles is greater than that expected by chance; that is, the alleles are in LD. LD studies have traditionally been used to complement traditional pedigree analyses, for example, to hone in on a locus that has been mapped by linkage analysis. However, LD-based association analysis has become the method of choice for whole-genome screens, particularly for diseases where traditional linkage studies have been unsuccessful. These studies have one great advantage over a traditional family analysis; because affected individuals are chosen from an entire population rather than from one or a few pedigrees, the number of potential subjects is limited only by the size of the population and the frequency of the disease. Maximizing the potential number of affected individuals that can be included in the analysis is extremely important for disorders where genetic heterogeneity or incomplete penetrance is likely to be factors.

Genetic Markers

Mapping studies, regardless of their type, depend on the availability of genetic markers. The most widely used markers are single nucleotide polymorphisms (SNPs) and microsatellite markers (also called simple tandem repeats [STRs], or simple sequence length polymorphisms [SSLPs]). SSLPs are stretches of variable numbers of repeated nucleotides two to four base pairs in length. These markers are highly polymorphic, as the number of repeat units at any given STR locus varies substantially between individuals. SNPs, as the name implies, are single base pair changes at a specific nucleotide; they are the most common form of sequence variation in the genome. SNPs are widely used for genetic mapping studies because they are distributed so widely across the genome and because they can be assessed in a high-throughput, automated fashion. Other forms of genetic variation that have been investigated for use as genetic markers include small insertion or deletion polymorphisms, termed indels, that generally range between 1 and 30 base pairs, and copy number variations (CNVs), which can refer to either deletions or duplications. Recent genome-wide surveys have revealed that CNVs are common and can range in length from several base pairs to several million base pairs. CNVs may contribute to chromosomal recombination and rearrangements, thereby playing an important role in generating genetic diversity, and also, as many of these variants are sizable, it is hypothesized that they may significantly influence the expression of genes that encompass or are adjacent to the variant.

MAPPING STRATEGIES

The genetic variants that contribute to disease susceptibility can be roughly categorized into those that are highly penetrant and those that are of low penetrance. High-penetrance variants, by definition, have a large effect on phenotype and therefore identifying these variants usually provides fundamental insights into the biologic pathways contributing to the disease. Because individuals carrying high-penetrance variants have a high probability of expressing a disease phenotype, such variants tend to be rare and segregate in families and are generally most powerfully mapped using pedigree-based linkage mapping approaches (Figs. 1.19–1 and 1.19–2). In contrast, low-penetrance variants have a relatively weak effect on phenotype, and therefore, identifying individual low-penetrance variants may, at least initially, provide relatively little new biologic knowledge. However, because of their small effects, such variants are typically common in the population, and therefore identifying them may add to the understanding of disease risk in the population as a whole. It is not expected that these variants will segregate strongly with the disease phenotype in pedigrees; efforts to identify them focus on population samples and require large sample sizes because of their relatively small effect on the phenotype. In the following sections, two common family-based linkage strategies (pedigree and sib pair analyses) are described, followed by an overview of population and family-based association analyses.

Pedigree Analysis

The primary goal of pedigree analysis is to determine if two or more genetic loci (i.e., a genetic marker of known location and the unknown disease loci) are cosegregating within a pedigree. A typical pedigree analysis, which is conducted in multigenerational families, consists of scanning the genome or a portion of the genome with a series of markers in one or more affected pedigrees, calculating a LOD score between the marker and the disease phenotype at each marker position, and thereby identifying the chromosomal regions linked to the disease trait.

Following the successful application of pedigree analysis to map Mendelian disorders such as Huntington disease, many investigators adopted this strategy for mapping psychiatric disease genes with, at best, mixed success. In the late 1980s and mid-1990s, several pedigree-based studies reported the mapping of susceptibility loci for Alzheimer disease, bipolar disorder, and schizophrenia. Although the linkage findings for three Alzheimer disease loci were relatively quickly replicated, the findings reported for bipolar disorder and schizophrenia were ultimately determined to have been false positives. While a number of different explanations have been proposed for the failure of pedigree-based approaches to map psychiatric loci, most investigators now recognize that these studies were generally drastically underpowered considering the apparent etiologic complexity of psychiatric disorders.

Pedigree analysis in psychiatry has increasingly turned toward an application for which it is more appropriately powered, namely, the mapping of quantitative trait loci (QTLs). QTLs are defined as genetic loci that contribute to the variation in continuously varying traits as opposed to categorical traits such as disease diagnoses. QTLs are typically loci of small effects that only contribute to a portion of the observed variance of a trait in the population. It is now generally accepted that using analytical methods developed in the late 1990s, it may be possible to use pedigree studies to map a wide range of quantitative traits that are relevant for understanding psychiatric disorders, including brain and behavioral phenotypes. Several such studies are now being undertaken, typically with multiple phenotypes being assessed in each individual in the pedigree. A more detailed discussion of quantitative phenotypes is provided after the overview of population-based mapping strategies.

Sib Pair Analysis

Affected sib pair (ASP) analysis, first proposed in 1935, became widely used during the 1990s for the genetic mapping of complex traits, including many psychiatric disorders. Sib pair analysis examines the frequency with which sibling pairs concordant for a trait share a particular region of the genome compared with the frequency that is expected under random segregation. Sib pair analysis is based on the fact that siblings share approximately 50% of their genomes IBD. Therefore, if a set of unrelated sib pairs affected with a given trait shares a particular area of the genome at a frequency significantly greater than 50% (the proportion of sharing expected under conditions of random segregation), then that area of the genome is likely to be linked to the trait in question. In this method, siblings are genotyped and population frequencies and parental genotypes are used to estimate the proportion of genes shared IBD at each site for each sib pair. The linkage analysis then compares those pairs concordant and discordant for each locus.

Like pedigree studies, ASP studies have more power to locate genes of large effect than genes of small effect. This limitation can be partially addressed by a two-tiered design that incorporates additional markers or family members after an initial linkage study in affected siblings or by increased sample size. It generally requires less effort to identify and assess even large sets of affected sibs than to identify and assess all members of extended pedigrees, particularly when investigators can take advantage of data repositories that include samples and phenotype data from sib pairs ascertained from multiple sites. For example, the U.S. National Institutes of Mental Health (NIMH) maintains such repositories for sizable collections of sib pairs affected with schizophrenia, bipolar disorder, autism, and Alzheimer disease. An additional benefit of the ASP design is that it allows for the incorporation of epidemiologic information, permitting the simultaneous examination of environmental and gene–environment interactions.

Association Studies

In the past few decades, there has been increasing acceptance of the notion that association studies are more powerful than linkage approaches for mapping the loci of relatively small effect that are thought to underlie much of the risk for complex disorders. While linkage studies attempt to find cosegregation of a genetic marker and a disease locus within families, association studies examine whether a particular allele occurs more frequently than expected in affected individuals within a population. As noted previously in this chapter, mapping genes using association studies is based on the idea that certain alleles at markers closely surrounding a disease gene will be in LD with the gene; that is, these alleles will be carried in affected individuals more often than expected by random segregation, because they are inherited IBD.

There are two common approaches to association studies (Fig. 1.19–2), case-control designs and family-based designs, which typically investigate trios (mother, father, and an affected offspring). In a case-control study, allele frequencies are compared between a group of unrelated affected individuals and a matched control sample. This design is generally more powerful than a family-based design, as large samples of cases and controls are easier to collect than trios and are less expensive as they require the genotyping of fewer individuals. Case-control samples may be the only practical design for traits with a late age of onset (such as Alzheimer disease) for which parents of affected individuals are typically unavailable. The main drawback of the case-control approach is the potential problem of population stratification; if the cases and controls are not carefully matched, then they may display substantial differences in allele frequency that reflect population differences rather than associations to the disease.

Family-based association studies are designed to ameliorate the problem of population stratification. In this design, the nontransmitted chromosomes (the copy of each chromosome that is not passed from parent to child) are used as control chromosomes, and differences between allele frequencies in the transmitted and nontransmitted chromosomes are examined, eliminating the problem of stratification as the comparison group is by definition genetically similar to the case group. Although more robust to population stratification than a case-control study, family-based studies are only about two-thirds as powerful using the same number of affected individuals.

In the past, it was not practical to conduct association studies on a genome-wide basis, as relatively few SNPs were available. Therefore, association studies focused on testing one or a few markers in candidate genes chosen on the basis of their hypothesized function in relation to a given disease. However, as a result of international efforts that have identified millions of SNPs distributed relatively evenly across the genome and the development of technology for genotyping them relatively inexpensively, genome-wide association (GWA) studies are now a common reality, and over the past decade, such studies are starting to generate an extensive catalog of common variants contributing to common diseases. Early GWA studies in psychiatric disorders failed to identify any risk variants; however, recent studies have made use of very large samples, providing further support for the hypothesis that underpowered study designs bear much of the responsibility for the disappointing results of previous psychiatric genetic investigations.

Statistical Considerations

Scientists in other biomedical research fields are often surprised by the apparently high level of statistical evidence that geneticists require to consider a linkage or association result to be significant. Most simply, this requirement can be thought of in terms of the very low expectation that any two loci selected from the genome are either linked or associated with one another. The likelihood that any two given loci are linked (i.e., the prior probability of linkage) is expected to be approximately 1:50, based on the genetic length of the genome. To compensate for this low prior probability of linkage and bring the posterior (or overall) probability of linkage to about 1:20, which corresponds to the commonly accepted significance level of $P = .05$, a conditional probability of 1,000:1 odds in favor of linkage is required, corresponding to the traditionally accepted LOD score threshold of 3. This generally provides an acceptable false-positive rate (Fig. 1.19–3), but some false-positive findings have exceeded even this threshold.

It is important to note that the accepted significance threshold for linkage studies is relatively low compared to the much more stringent thresholds required for current GWA studies. The reason for more stringent thresholds is that the expectation that any two loci in the genome are associated with one another is even lower than that of their being in linkage, and typically a P value of less than about 10^{-7} is considered to indicate “genome-wide significance.” Another very important consequence of the expectation that any two loci are unlikely to be associated with each other is that candidate gene studies should be held to the same stringent standards used in genome-wide studies. A rigorous candidate gene study would use the agnostic assumption that all loci have the same low prior probability of association and would discount the high prior probability that some investigators assign to variants in candidate genes chosen on the basis of their hypothesized functional relevance to a given disorder or trait. GWA studies are now replicating associations with very low *P* values for a wide range of complex traits, while the vast majority of candidate gene associations (which usually report significantly much higher *P* values) remain unreplicated. It is therefore increasingly apparent that genome-wide levels of significance are appropriately applied to all initial association studies for a given trait, even when tests are limited to a small number of genetic variants within a defined set of candidate genes. As more rigorously confirmed biologic data is obtained for associations between genetic loci and diseases or behavioral traits, future association studies will be able to relax the stringent agnostic standards. For example, if a genome-wide mapping study identifies an association between a locus and a trait like neuroticism using a stringent threshold, future studies of the same locus and similar traits like a diagnosis of an anxiety disorder would be considered meaningful with a less stringent threshold.

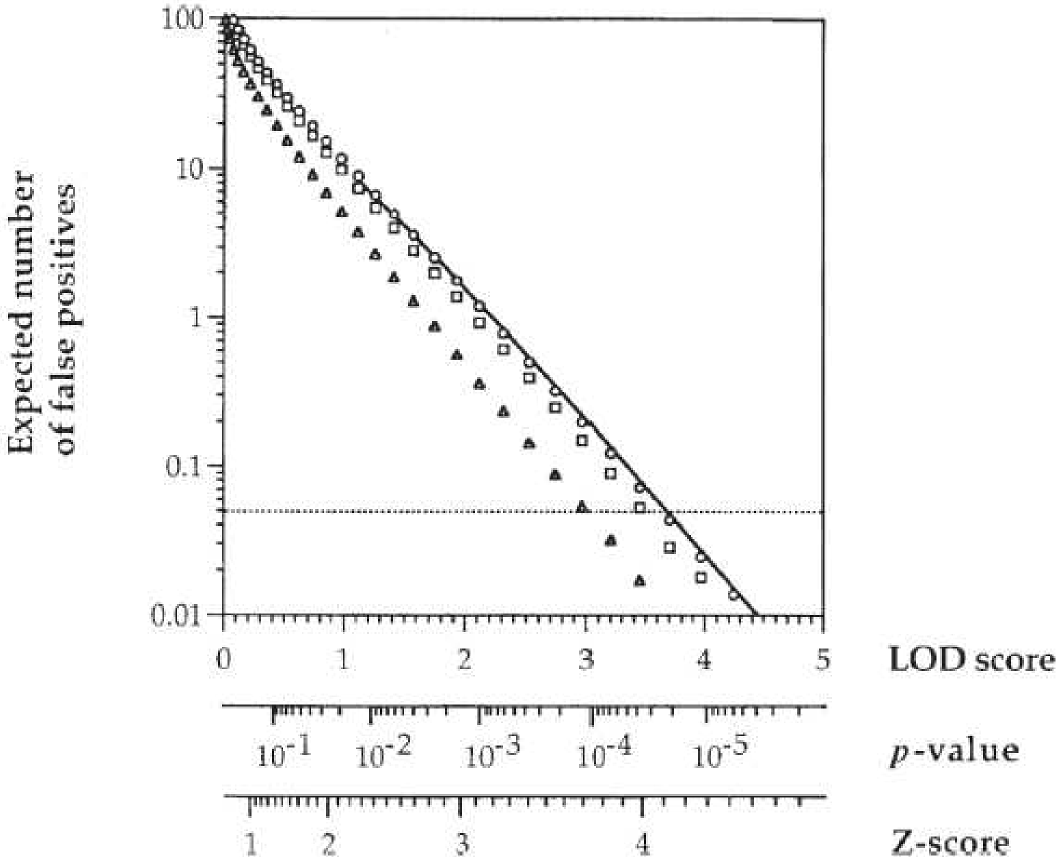

FIGURE 1.19–3. Number of false positives expected in a whole genome scan for a given threshold of logarithm of odds (LOD) score. *Solid line* represents the expectation for a perfect genetic map. Symbols represent the results for 100 sib pairs using genetic maps with markers spaced every .1 cM (*circles*), every 1 cM (*squares*), and every 10 cM (*triangles*). The *dotted line* indicates the 5% genome-wide significance level. (Reprinted from Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting linkage results. *Nat Genet*. 1995;11(3):241–247, with permission from Springer Nature.)

THE PHENOTYPE

The generally disappointing results of psychiatric genetic mapping studies have focused increasing attention on the problem of defining and assessing phenotypes for such studies. Most psychiatric mapping studies to date have relied on categorical disease diagnoses, as exemplified by the *Diagnostic and Statistical Manual* (DSM) classification scheme. Criticisms of this approach rest on two arguments. First, diagnosis of psychiatric disease depends on subjective clinical evaluation, a fact that underscores the difficulty in ascertaining individuals who can be considered definitely affected by a given disease. Second, even when a psychiatric diagnosis can be established unambiguously, the menu-based system used for psychiatric classification provides the possibility that any two individuals affected with a given disorder may display largely nonoverlapping sets of symptoms, likely reflecting distinct etiologies. Concern that the diagnosis-based approach to phenotyping may represent one of the chief obstacles to the genetic mapping of psychiatric phenotypes has generated considerable interest in mapping heritable traits known to demonstrate continuous variation in the population. Continuous measures that are hypothesized to be related to psychiatric disorders include biochemical measures (e.g., serum or CSF levels of neurotransmitter metabolites or hormones), cognitive measures, personality assessments, structural or functional brain images, biophysical markers such as responses to evoked potentials, or molecular assays such as gene expression profiles. Key features of categorical and continuous phenotyping strategies are shown in Figures 1.19–4 and 1.19–5, and each is discussed in more detail below.

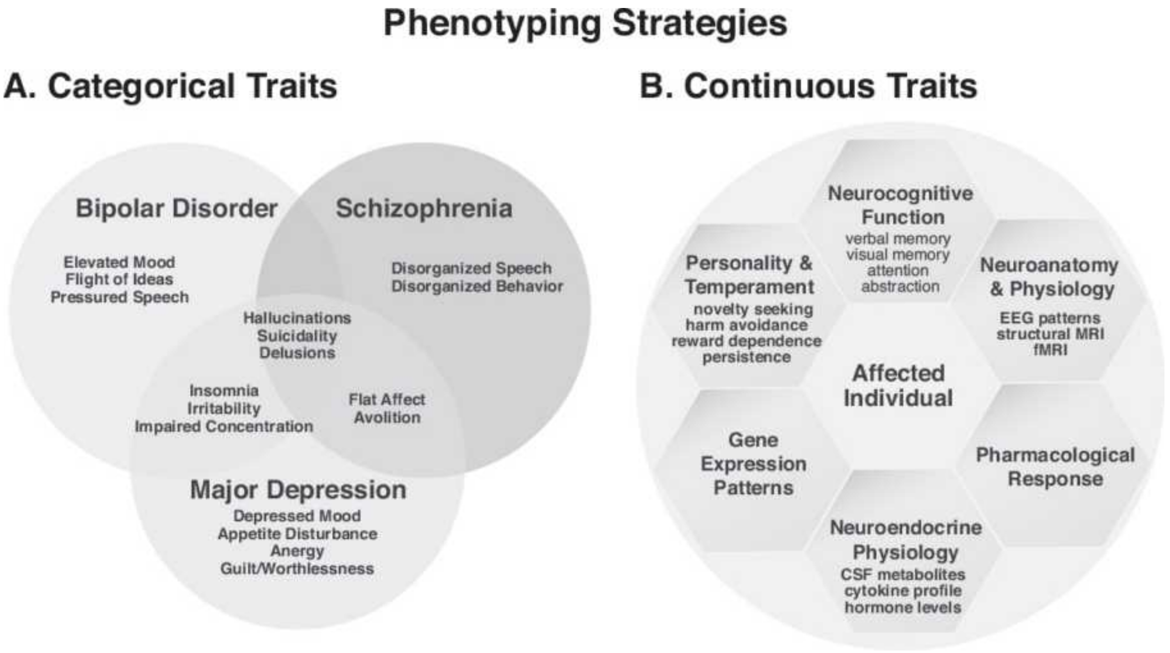
Phenotyping Strategies
A. Categorical Traits
B. Continuous Traits
Bipolar Disorder
Elevated Mood
Flight of Ideas
Pressured Speech
Schizophrenia
Disorganized Speech
Disorganized Behavior
Hallucinations
Suicidality
Delusions
Insomnia
Irritability
Impaired Concentration
Flat Affect
Avolition
Major Depression
Depressed Mood
Appetite Disturbance
Anergy
Guilt/Worthlessness
Neurocognitive Function
verbal memory
visual memory
attention
abstraction
Personality & Temperament
novelty seeking
harm avoidance
reward dependence
persistence
Neuroanatomy & Physiology
EEG patterns
structural MRI
fMRI
Affected Individual
Gene Expression Patterns
Pharmacological Response
Neuroendocrine Physiology
CSF metabolites
cytokine profile
hormone levels

FIGURE 1.19–4. Two alternate schemes for conceptualizing psychiatric phenotypes. **A:** Categorical traits as conceptualized by the *Diagnostic and Statistical Manual* (DSM) represent a “menu-based” approach to psychiatric disorders. Individuals are assessed for a checklist of signs and symptoms that are then used to categorize the individual as “affected” according to a specific diagnosis. Not all symptoms are present in samples of individuals who carry a particular DSM diagnosis, and many of these symptoms occur across diagnostic boundaries, as illustrated in this Venn diagram. DSM phenotypes therefore probably represent etiologically heterogeneous categories. **B:** Alternatively, in the continuous traits model, “affectedness” can be conceptualized in terms of an expectation that an individual will demonstrate extreme values on a set of continuous measures that correlate with psychopathology and thus are hypothesized to underlie the disorder (as illustrated by examples of six different types of measures shown in the hexagon). Such measures may also be associated with particular components of categorical phenotypes, such as those depicted in the Venn diagram. The justification for using continuous measures as the phenotypes for genetic mapping studies is that they are considered etiologically simpler and more reliably assessed compared to categorical phenotypes. In addition, mapping such traits combines information from all members of the study population (affected and unaffected individuals alike), which adds considerably to power.

Categorical Phenotypes

The most commonly used categorical phenotypes in psychiatry are DSM diagnoses. Some studies focus on a single DSM diagnosis, while other studies include individuals with a range of different diagnoses. The latter approach is typically used for disorders that are hypothesized to represent a single disease spectrum, such as mood disorders. Using the categorical approach, it is important to be able to classify subjects as unambiguously as possible. Several strategies are used to accomplish this goal. The first strategy involves deciding on the appropriate diagnostic criteria for the study in question and deciding how these criteria will be applied to individuals in the study. One way of standardizing the procedures used to identify and assess potential study subjects is to use only experienced clinicians in the diagnostic process and to train them in the administration of the instruments and the diagnostic criteria to be employed. In addition, a “best estimate” procedure and/or a consensus diagnosis are frequently used. The best estimate process involves making use of every piece of available information, including medical records, interviews, and videotapes, to arrive at a diagnosis. For a consensus diagnosis, two or more diagnosticians independently review the material and make a diagnosis for each individual. The diagnoses are then compared and individuals for whom an agreement in diagnosis cannot be reached are not entered as “affected” into the study.

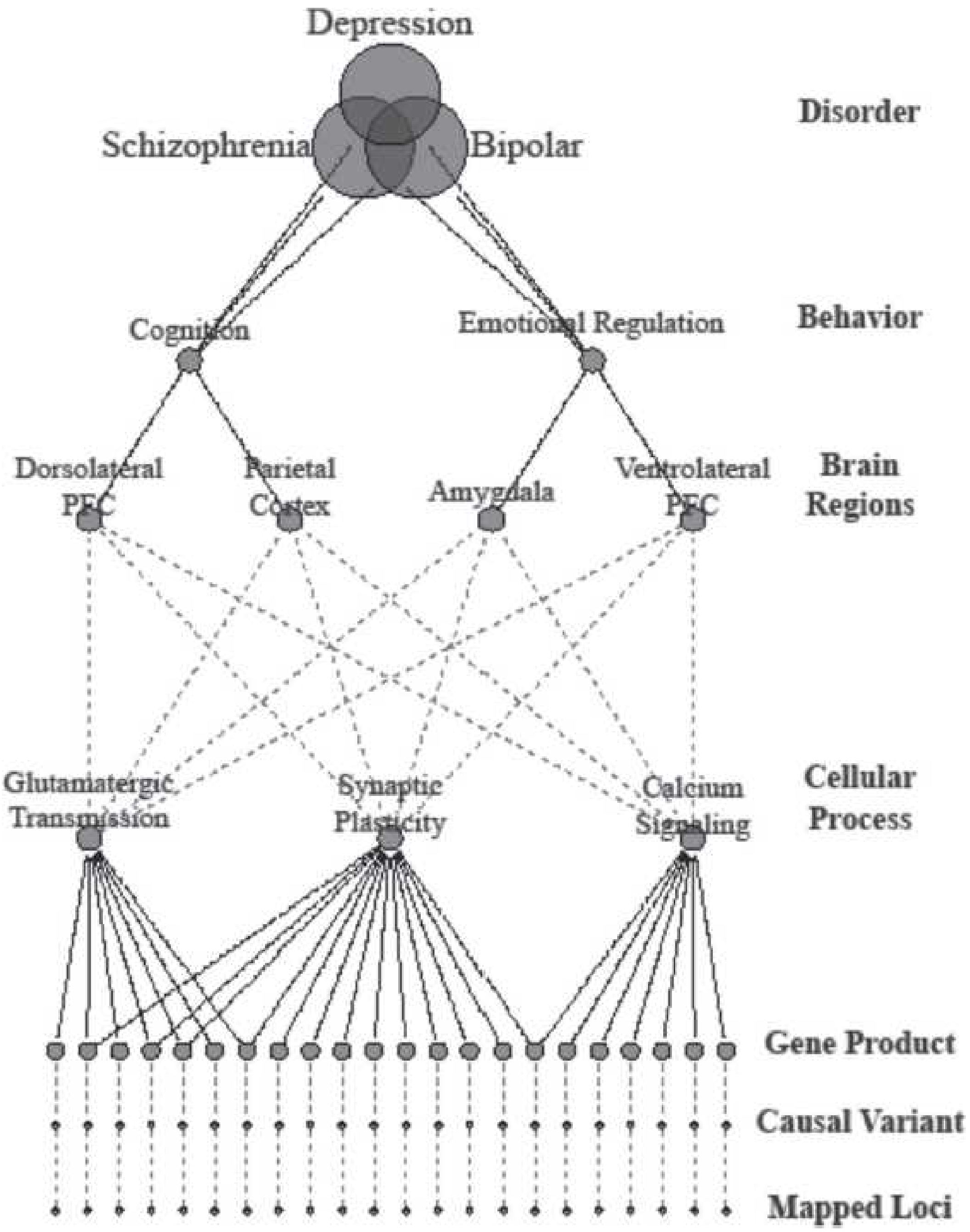
Depression
Schizophrenia
Bipolar
Disorder
Cognition
Emotional Regulation
Behavior
Dorsolateral PFC
Parietal Cortex
Amygdala
Ventrolateral PFC
Brain Regions
Glutamatergic Transmission
Synaptic Plasticity
Calcium Signaling
Cellular Process
Gene Product
Causal Variant
Mapped Loci

FIGURE 1.19–5. The endophenotype (also known as intermediate phenotype) strategy aims to map genes associated with quantitative measures acquired from multiple levels of physiology in order to characterize the biologic pathways that cause psychiatric disorders. In contrast to conventional studies that map genes associated with dichotomous disease categories, the endophenotype method can be implemented across diagnostic categories to identify the differences/commonalities in biologic systems that contribute to psychiatric disorders. Because intermediate phenotypes are "closer to genes" compared to diagnostic categories, they are expected to have more power in gene-mapping studies.

A well-designed study makes use of all available information about the genetic epidemiology of the disorder to choose a sample of affected individuals to study. It is often the case that a subset of families carries the disorder in what appears to be a simple Mendelian pattern, while the inheritance pattern is less clear for other families or groups. In a disorder where there are likely to be multiple genes contributing to the phenotype, it makes sense to begin with a study sample where there may be major loci. Redefining the disease phenotype can often simplify the mapping process by identifying such groups or families. For example, in the search for a genetic defect for Alzheimer disease, the process was advanced enormously by limiting the study population to those individuals who had an early age of onset (before age 65); the early onset trait segregated in an autosomal dominant fashion. Other ways of redefining the phenotype include focusing on factors such as ethnic background, treatment response, symptom severity, or the presence of comorbid disorders.

Narrowing the phenotype using the approaches discussed above may increase the chances of finding a genetic defect in complex diseases, but it can also greatly reduce the power of the study by limiting the number of available affected individuals. For this reason, it has been argued that for some disorders, broadening the phenotype is an appropriate strategy. The suggestion is that for some complex diseases, the phenotype of interest may represent the extreme end of a spectrum and that to have enough power to map genes; other phenotypes within the spectrum must also be included. For example, mapping studies of bipolar disorder might include affected individuals with MDD as well as those individuals diagnosed with bipolar disorder.

Although the two approaches of narrowing the disease phenotype and broadening the disease phenotype may seem to be mutually exclusive, many groups studying complex disorders have incorporated both approaches into their study designs. One way to do this is to create stratified diagnostic categories ranging from a narrow diagnostic category to a broad diagnostic category, and test for genetic linkage under each of these schemas. Some investigators argue that for complex diseases that are part of a spectrum, this strategy decreases the rate of false negatives, that is, of missing an existing linkage because of misspecification. Others argue that using several models and picking the one that gives the highest scores greatly increase the rates of false positives, that is, of identifying an area of linkage where none exists. One problem that clearly exists with the use of multiple diagnostic categories is that as more models are used (and therefore, more statistical tests are performed), increasingly stringent levels of evidence are required to consider a result significant.

While categorical phenotypes remain the mainstay of psychiatric genetic studies, the limitations of DSM nosology as the basis of phenotyping for genetic studies are becoming clear. Genetic investigations are focusing increasingly on traits that may be components of one or more DSM diagnostic categories. For example, there is growing evidence that genetic susceptibility to psychosis, broadly defined, contributes to both severe bipolar disorder and schizophrenia, and a number of investigative approaches are being employed to attempt to identify genes that underlie such susceptibility and even to explore possible etiologic relationships between psychiatric and nonpsychiatric disorders. For example, bioinformatics models have been employed to investigate medical records databases and have uncovered extensive pairwise correlations among a diverse list of psychiatric disorders, neurologic disorders, autoimmune disorders, and infectious diseases. Eventually, the results of such model-fitting experiments may provide a framework to design more powerful linkage and association studies that can search for alleles that contribute to susceptibility to multiple disorders.

Continuous Phenotypes

Because of the difficulties experienced in genetic mapping of categorical diagnoses, neurobehavioral geneticists are increasingly focused on investigating quantitative traits that are hypothesized to underlie a particular psychiatric diagnosis and that may be simpler to genetically map. The rationale for efforts to map such alternative phenotypes, termed endophenotypes or intermediate phenotypes, is that the genes identified through such efforts may provide clues regarding the biologic pathways that are relevant to understanding a particular disorder or multiple disorders that may have overlapping etiologic pathways (Fig. 1.19–5). Several features characterize useful endophenotypes. First, they should be state independent; that is, they should not fluctuate as a function of the disease course or medication treatment and should show adequate test–retest stability. Second, they should be heritable; that is, there should be evidence that genetic factors are responsible for a substantial proportion of the variability of the trait within the population. Third, the endophenotype should be correlated with the disease under investigation;

that is, different values of the trait measure are observed in patients compared to unrelated control subjects. Measures of brain structure and function are now under investigation as endophenotypes for psychiatric disorders. For example, several features of brain morphometry (as assessed by magnetic resonance imaging [MRI]) are highly heritable (in the range of 60% to 95%), including total brain volume, cerebellar volume, gray and white matter density, amygdala and hippocampal volume, and regional cortical volume. Several studies show that brain structural features that are correlated in clinical samples with disorders such as schizophrenia or bipolar disorder are also abnormal in relatives of affected individuals. Physiologic measures of brain activity that have been employed as candidate endophenotypes for psychiatric disorders include electroencephalography (EEG) patterns. In addition to brain measures, behavioral traits relevant to psychiatric disorders are an active field of gene-mapping studies, including neurocognition, impulsivity, emotional regulation, and temperament (i.e., neuroticism).

Animal Models

In contrast to categorical phenotypes, endophenotypes can be more straightforwardly related to phenotypes that can be assessed in animal models. Studies of genetic variations that affect circadian rhythms provide a good example. Variations in circadian rhythms have long been recognized as important features of mood disorders, and quantitative assessments of activity patterns have been proposed as endophenotypes for such disorders. Numerous studies in animal models have demonstrated that genetically controlled biologic clocks determine circadian activity and that variations in clock genes are associated with variations in such activity from bacteria to humans. Genetic mapping efforts in fruit flies starting in the early 1970s resulted in the identification of at least seven “clock genes,” beginning with *period.* Subsequent studies showed that the homologs of several of these genes play essential roles in regulating mammalian circadian rhythms. Genetic mapping studies in mice also have identified previously unknown circadian rhythm genes, beginning with the discovery and characterization in the early 1990s of *clock.* These genetic discoveries have not only explicated the cellular networks and neurophysiologic circuits responsible for the control of mammalian circadian rhythms but have also generated animal models that may shed light on the pathobiology of psychiatric syndromes such as bipolar disorder. For example, mice carrying a targeted mutation in *clock* demonstrate abnormal activity patterns, such as hyperactivity and decreased sleep, which are apparently modified by the administration of lithium.

PROGRESS IN THE GENETIC DISSECTION OF SPECIFIC DISORDERS

Taken as a whole, the progress in identifying susceptibility genes for psychiatric disorders has been slower compared to that observed for nonpsychiatric disorders. The final sections of this chapter will review the progress that has been made in identifying the genetic underpinnings of several specific psychiatric disorders. Alzheimer disease represents the most successful application of gene-mapping strategies to complex neurobehavioral disorders, and the section on this disease provides an example of how genetic linkage studies add to understanding the pathogenesis of a complex trait. Gene-mapping studies of autism provide examples of genetic investigations of syndromes that have relatively simple inheritance patterns and demonstrate how these studies have provided starting points for investigations of more complex autism spectrum disorders (ASDs). Finally, recent progress identifying risk variants for schizophrenia, bipolar disorder, and major depression are used to illustrate the exciting directions that are beginning to revolutionize the current understanding of the etiology of these complex disorders.

ALZHEIMER DISEASE

Alzheimer disease provides an excellent example of the power of genetics to elucidate the complex biology of a neuropsychiatric disorder. Alzheimer disease is a well-defined form of dementia characterized by progressive impairment of memory and intellectual functioning. The clinical signs and symptoms, although characteristic, are not limited to Alzheimer disease but are also found in several other types of dementia. Alzheimer disease has several distinguishing histopathologic features; the presence of senile plaques (made up of a core of β-amyloid fibrils surrounded by dystrophic neurites), tau-rich neurofibrillary tangles, and congophilic angiopathy in the brain parenchyma and associated blood vessels are pathognomonic for Alzheimer disease.

A variable age of onset has been noted for Alzheimer disease, ranging from as early as age 35 to as late as age 95. Family history is the second greatest risk factor after age for the disease. The concordance rate for Alzheimer disease in MZ twin pairs is about 50%, indicating a moderately strong genetic contribution to disease risk. It is now evident from a wide range of genetic studies that Alzheimer disease can be divided into two broad categories: familial forms, which account for about 5% of Alzheimer disease cases and are characterized by early onset and autosomal dominant inheritance with high penetrance; and sporadic forms, in which the genetic contribution is hypothesized to be similar to that characterizing other common neuropsychiatric diseases.

The search for the genetic basis of familial Alzheimer disease began with traditional linkage studies. First, an investigation of a candidate locus on chromosome 21 in humans identified mutations in the *APP* gene in a small number of families in which significant linkage had previously been observed to markers from this region. The *APP* encodes the amyloid precursor protein, a cell surface and transmembrane precursor protein that is cleaved to form multiple peptides of various function. Some of the peptides are secreted and function to promote transcriptional activation, whereas other peptide products form the basis of the β-amyloid fibrils. Transgenic mice with different *APP* mutations were created and have been shown to produce β-amyloid deposits and senile plaques as well as to show synapse loss, astrocytosis, and microgliosis, all part of the pathology of Alzheimer disease. Mutations in the genes that encode β-APP all lead to an increase in the extracellular concentration of longer fragments of β-amyloid (Aβ42). Most of the strains of transgenic mice with mutations in APP exhibit increased rates of behavioral changes and impairment in several memory tasks, indicating dysfunction in object-recognition memory and working memory among others. These findings represent striking evidence that mutations in the β-amyloid gene are indeed responsible for at least some of the histopathologic elements of Alzheimer disease.

Even as the above findings were being reported, it was clear that mutations in the β-amyloid gene could not completely explain the etiology and pathology of Alzheimer disease, not least because it was shown that linkage to chromosome 21 was excluded in most early-onset Alzheimer disease families. In addition, no neurofibrillary tangles are observed in most of the different β-amyloid transgenic mice. The subsequent search for the genetic underpinnings of Alzheimer disease using genome-wide linkage analysis of early-onset Alzheimer disease families resulted in the identification of two additional Alzheimer disease susceptibility genes: *presenilin-1* (*PS-1*) on chromosome 14q24.3 and *presenilin-2* (*PS-2*) on chromosome 1q. PS-1 and PS-2 are integral transmembrane proteins with at least seven transmembrane domains. The *presenilins* are aspartyl proteases that carry out γ-secretase cleavage of *APP* to generate β-amyloid. Similar to mutations in *APP,* genetic variation in *presenilins* causes an increase in the expression of longer fragments of β-amyloid.

These findings emphasize one of the strengths of using family-based linkage analysis. Pedigree-based studies are especially suited to identify highly penetrant disease genes that serve critical roles in important biologic processes. Mutations in *APP* and *presenilin* are rare, but different mutations in these genes have occurred in different founders; to date, 185 mutations have been identified in *PS-1,* 13 mutations for *PS-2,* and 24 mutations for *APP.* Virtually all of these mutations are transmitted in an autosomal dominant and fully penetrant manner. The convergence of different mutations within the same genes clearly confirms the involvement of these genes in the disorder. Although the mutations in *APP* and *presenilin* are rare, research into the biology of the expressed proteins has provided key insights into the pathophysiology of dementia. Because these highly penetrant mutations elucidate important biologic functions, they also provide a firm ground to design therapeutic interventions. For example, β-amyloid "vaccines" designed to induce an immunogenic response to pathogenic amyloid are now in advanced clinical trials. In addition, γ-secretase modulators have been developed to target the dysfunction of the *presenilin* gene products. Unlike the current psychopharmacologic treatments for Alzheimer disease that nonspecifically target cholinergic and glutamatergic systems, the β-amyloid vaccines and γ-secretase modulators specifically treat the causes of Alzheimer disease by generating an immune response or altering the processing of *APP* and may actually prevent or reverse the deposition of senile plaques.

Sporadic and Late-Onset Alzheimer Disease

Mutations in *APP, PS-1,* or *PS-2* are present in a majority of familial cases of early-onset Alzheimer disease but do not account for sporadic or familial late-onset Alzheimer disease. For this reason, investigators turned to other approaches to search for the evidence of linkage in a large number of small families with late-onset Alzheimer disease. In 1991, the results of a linkage study using 36 markers in late-onset Alzheimer disease families provided evidence for a susceptibility gene on the long arm of chromosome 19. In 1993, association studies revealed that the e4 allele of the *apolipoprotein E* gene was strongly associated with late-onset Alzheimer disease and that this association almost certainly was responsible for the previously observed linkage signal on chromosome 19. There are three known alleles of this gene—e2, e3, and e4. These three alleles are distinguished by combinations of amino acids at residues 112 and 158; e2: Cys_{112}/Cys_{158}, e3: Cys_{112}/Arg_{158}, and e4: Arg_{112}/Arg_{158}. In most populations, the e3 allele is the most common. However, in familial late-onset Alzheimer disease, the incidence of e4 is approximately 50%, and in sporadic late-onset Alzheimer disease, it is 40%, compared with about 16% in normal controls. Inheritance of a single copy of e4 increases the risk for the disease by 4-fold, and inheritance of two copies of the allele increases the risk by 10-fold. Epidemiologic studies suggest that between 30% and 60% of late-onset Alzheimer disease cases have at least one *apoE-e4* allele. The e4 genotype appears to be a more important risk factor for Alzheimer disease in populations of European and Asian origin when compared with populations of African origin. The *apoE* gene has a well-characterized role in the metabolism and transport of lipids and is believed to play a role in the clearance of β-amyloid from the brain. Overall, the association of *apoE-e4* with Alzheimer disease remains probably the strongest association yet identified for a common human disease.

The establishment of *apoE-e4* as a susceptibility allele for late-onset Alzheimer disease has led to the search for additional alleles that might interact with *apoE-e4* to modify disease risk. In 2007, investigators used GWA strategies in histologically confirmed cases and controls to identify *GAB2* (GRB-associated binding protein 2) as an additional risk allele in *apoE-e4* carriers (but not in Alzheimer disease patients who were not e4 carriers). Initial studies suggest that carriers of both *apoE-e4* and *GAB2* risk alleles have an almost 25-fold greater risk for Alzheimer disease than individuals who do not carry either risk allele. *GAB2* appears to modulate β-amyloid production through its interaction with GRB2 (growth factor receptor-bound protein 2), which physically interacts with APP and the presenilins.

Since the initial GWA studies that identified *GAB2,* there have been more than a dozen additional GWA studies and, together, have identified more than 40 confirmed Alzheimer risk genes. Unlike *APP* and the *presenilins,* these genes have small effect sizes (odds ratio <1.2) and are relatively common in the general population. Their role in the pathophysiology is less clear compared to the current understanding of *APP* and the *presenilins* biology, though analysis of the known function of genes identified in GWA studies point to the involvement of immunity, lipid metabolism, tau binding proteins, and amyloid precursor protein metabolism. Notably, the population-based GWA studies have identified genes associated with β-amyloid processing other than *APP* and *presenilins*, showing a convergence on a common pathway in both familial and sporadic forms of the disease. For example, the *ATXN1* gene product modulates γ-secretase levels and impacts the cleavage of APP and the murine knockout model of *ATXN1* shows cognitive impairment by age 6 months. Other genes identified by GWA studies appear to influence the risk of Alzheimer disease by impacting immune and inflammation pathways, including factors that regulate cellular interactions required for immune reactions (e.g., CD33) and inflammatory processes (e.g., *CR1,* the receptor for the complement molecule C3b). These immune/inflammatory processes have been postulated to mediate the neurotoxic response to β-amyloid.

As the results of GWA studies of Alzheimer disease continue, genes associated with protection and resilience have emerged alongside genes associated with risk. Resilience to Alzheimer disease, also known as asymptomatic Alzheimer disease, is of great interest because of the possibility of identifying biologic pathways that confer protection from the disease. As many as 70% of cognitively normal older adults show clear evidence of Alzheimer histopathologic changes, and up to 30% of cognitively normal older adults meet full autopsy-confirmed neuropathologic criteria for the disease. Characterization of the genetic and environmental factors that confer protection despite clear evidence of brain pathology holds promise for prevention or treatment strategies to mitigate the impact of the disease processes. For example, a recently identified missense mutation within the APP gene near the β-secretase cleavage site was shown to be protective against the disease and slow the cognitive decline. In vitro studies showed that the mutation resulted in suboptimal β-secretase cleavage with subsequent reduced aggregation of pathologic Aβ by 50% compared to wild-type cells. Such studies provide potential pathways to intervene with the goal of preventing or slowing the processes that eventually result in the full manifestation of the disease.

Summary

Progress in the field of Alzheimer research has achieved significant momentum, and there are now more than 40 genes implicated in the pathogenesis of this disorder. Linkage studies of rare familial forms of Alzheimer disease led to the discovery of high-penetrance variants that have had a profound impact on the understanding of Alzheimer disease pathogenesis and on the basic understanding of a wide range of cellular processes within the CNS. Association studies have unequivocally identified lower-penetrance variants that together explain much of the genetic contribution to disease risk at the population level. Taken together, the risk genes converge on a few basic biologic pathways: processing of β-amyloid, lipid metabolism, immunity, and inflammation. In addition to the identification of risk alleles, gene-mapping studies have also identified genetic variants that confer protection from the disease. The identification of risk and protective variants has provided the promise of two types of medical breakthroughs. New therapies are in development that target the molecular pathways identified through these studies. In addition, the emerging picture of genetic risk for common forms of Alzheimer disease suggests that it may soon be possible to focus prevention and early intervention strategies on individuals who are at high risk.

AUTISM

Autism is a severe neurodevelopmental disorder that was originally described as “enclosure in one’s self” by Leo Kanner in 1943 and is characterized by three primary features: impaired language and communication, abnormal or impaired social interaction, and restricted, repetitive, and stereotyped patterns of behavior. Understanding of the etiology of autism has proceeded slowly, but there is now convincing evidence that alterations in specific cellular and molecular neurodevelopmental pathways are important in its etiology. There is particularly strong evidence for a genetic contribution to the risk of autism and ASDs. The sibling recurrence risk for autism and/or ASD is between 2% and 6%. Given a population prevalence of about 1 in 2,000 (0.04%), this means that the siblings of autistic individuals are approximately 50 to 100 times more likely to develop autism than a person in the general population. Twin studies of autism not only show extraordinarily high heritability (as demonstrated by MZ twin concordance of 80% to 92%) but also demonstrate the genetic complexity of these disorders, with the DZ twin concordance rate of 1% to 10% suggesting a highly multigenic mode of inheritance.

Increasing interest is focused on the possibility that large-scale chromosomal aberrations (CNVs) contribute an important role in autism. Early studies found an increased frequency of CNVs in autistic individuals at a rate of 5% to 10% compared to 1% to 3% in controls. Subsequent studies failed to replicate the initially identified high rate of CNVs compared to controls; however, recent work has suggested that the specific location and functional consequence of CNVs are more critical than the relative number of abnormalities. Two large studies identified a heterogeneous collection of CNVs that were not randomly distributed within the genome but rather concentrated within a subset of loci. Functional analysis of these loci showed that they are involved in synapse development, neuronal migration, and axon targeting. These functional findings converge nicely with investigations of well-characterized non–CNV-related abnormalities (discussed in detail below) and provide confidence that sporadic or inherited CNVs represent an important etiologic factor for some cases of the disorder. Collectively, the large population studies estimated that CNVs account for 5% to 8% of syndromic cases of autism.

Even prior to the demonstration of high rates of de novo mutations in autism, epidemiologic studies had strongly suggested that the genetic basis of this disorder is likely complex. For example, although the risk of autism in first-degree relatives of autistic probands is high, there is a substantial falloff for second- and third-degree relatives of such probands, suggesting that multiple genetic variants must interact to increase susceptibility to this syndrome. Segregation analyses of autism also support the hypothesis that it is a heterogeneous disorder that reflects the actions of multiple genetic variants of small effect. Genetic studies of autism have included whole-genome screens, candidate gene studies, chromosome rearrangement studies, mutation analyses, and comparative genomic hybridization studies. Taken together, these studies have contributed to an emerging picture of autism susceptibility that includes genes involved in three major systems: those involving synapse formation and maintenance, morphogenesis, and calcium regulation and signaling.

Synapse Formation and Maintenance

Perhaps the biggest breakthroughs in identifying susceptibility genes for autism have come from studies of disorders that display clinical features associated with autism or ASDs but with simpler inheritance patterns, including fragile X syndrome, tuberous sclerosis, and Rett syndrome. In general, the genetic defects associated with these disorders affect synapse formation and maintenance. Fragile X, which accounts for 3% to 4% of autism cases, is caused by an unstable trinucleotide repeat in the 5′ region of the *FMR1* gene at Xq27.3. This repeat expands as it is transmitted to succeeding generations, resulting in abnormal methylation and inhibition of expression of *FMR1. FMR1* produces an RNA-binding protein that acts as a chaperone for the transport of RNA from the nucleus to the cytoplasm and is involved in mRNA translation at the synapse. Abnormalities in dendritic spine density (increased over normal) and morphology (longer and thinner than normal) have been reported in individuals with fragile X as well as in mouse models of this disorder. Tuberous sclerosis, which accounts for perhaps 2% to 10% of autism cases (the rate of tuberous sclerosis is higher among autistic individuals with seizure disorders), results from mutations in one of two tumor suppressor genes, *TSC1* on 9q34, and *TSC2* on 16p13, both of which are involved in guanosine triphosphatase (GTPase) inactivation. Loss of a single copy of *TSC1* in mice has been shown to disrupt cytoskeletal dynamics and dendritic spine structure.

The *neuroligin* (*NLGN*), *SHANK,* and *neurexin* gene families are additional risk loci that appear to play a role in synapse formation and may be affected by mutations observed in some individuals affected with autism. The *NLGN* genes, situated on the X chromosome, produce cell adhesion molecules that are located on postsynaptic glutamatergic neurons. When mutated in rodents, these genes show defective trafficking and synapse induction. In nonmutated form, their expression induces the formation of normal, presynaptic terminals in axons. In mutants, there is defective transport of vesicles to the synaptic membrane and reduced synaptic induction. Mice carrying a human mutation of *NLGN3* show an interesting phenotype characterized by impaired social novelty preference and reduced vocalizations, two of the central features of the disorder in humans. The *SHANK* gene family encodes scaffolding proteins that are located in the postsynaptic membrane and serve as master organizers of the PSD. *SHANK3* is one of three members of the *SHANK* gene family and is located in the minimal region of the 22q13 deletion syndrome, known as Phelan-McDermid syndrome, which includes prominent autistic features. Larger studies have found that *SHANK3* mutations are present in about 0.85% of all autistic subjects. A third family of synapse-related genes associated with autism are the neurexins. Neurexins interact with neuroligins and are critical to the formation of presynaptic boutons. Neuroexins also link calcium channels to synaptic vessels and mediate neurotransmitter release through exocytosis. Like neuroligins and SHANK3, mutations in neurexins (*NRXN1*) occur rarely but have emerged multiple times in independent familial lines of individuals with autism.

Morphogenesis

Many individuals with syndromic forms of autism display facial dysmorphisms and/or abnormal body growth rates. About 20% of autistic children show macrocephaly with increased trajectories of head growth during the first few years of life, although a smaller proportion of patients show microsomia. Although the etiology of the abnormal growth patterns is not yet clear for most cases, several syndromic forms of autism have been associated with rare mutations of genes involved in embryogenesis and organogenesis. Mutations in *HOXA1,* a gene necessary for normal development of the brain stem, cerebellum, cranial nerves, and portions of the calvarium, have been identified in several families with heterogeneous phenotypes that include autistic features. Missense mutations have been identified in the tumor suppressor gene *PTEN* on chromosome 10q23. *PTEN* mutations have been associated with a broad cluster of syndromes, including Cowden syndrome, Bannayan–Riley–Ruvalcaba syndrome, Proteus syndrome, and Lhermitte–Duclos disease. *PTEN* has been shown to interact in the ERK/P13K/mTOR pathway, which is critical for incorporating information regarding nutrient availability and growth factor signals to regulate cell proliferation and body growth. *PTEN* mice knockout models show macrocephaly, CNS overgrowth, and excessive dendritic and axonal growth. Syndromes involving *PTEN* mutations have heterogeneous features, including nonpsychiatric phenotypes like cancer, highlighting that the final expression of specific genetic mutations depends on a complex interplay of genetic background (other mutations) and environmental factors.

Another syndromic form of autism is Rett syndrome, an X-linked pervasive developmental disorder (the first with a known genetic etiology) that occurs only in girls and is associated with normal early development followed by loss of skills, particularly social engagement and purposeful hand skills by age 4. The syndrome also points to abnormalities in synapse formation and maintenance in ASD and ASD-like disorders. Rett syndrome is caused by mutations in *MeCP2,* which makes a methylated-DNA–binding protein that regulates gene expression and chromatin structure and is required for normal brain development. Loss of *MeCP2* is associated with delayed neuronal maturation and abnormal synaptogenesis. Loss-of-function mutations in *MeCP2* are lethal in males and cause severe symptoms in about 70% of females; however, some mutations result in relatively mild phenotypes with some intellectual disability and/or verbal deficits. The heterogeneous expression of these mutations on the syndromic phenotype again highlights the key role that genetic background plays in determining the final manifestation of symptoms.

Calcium Regulation and Signaling

Another class of genes that has been linked to at least some cases of autism includes ion channels, receptors, and Ca^{2+}-regulated signaling molecules. Gain-of-function mutations in several L-type voltage-gated Ca^{2+} channels that express constitutively active channels and lead to excessive Ca^{2+} influx have been associated with autism. Timothy syndrome is caused by gain-of-function mutations in *CACNA1C,* an L-type voltage-gated Ca^{2+} channel and is characterized by intellectual impairment and autism. Interestingly, *CACNA1C* has been associated with schizophrenia and bipolar disorder, providing an example of a genetic risk factor that can cross diagnostic categories. Gain-of-function mutations in *CACNA1F,* another L-type voltage-gated Ca^{2+} channel, are the cause of congenital night blindness that can be accompanied by autistic features. Several other genetic mutations have been identified in other genes that are linked to calcium regulation. In one autism case, a mutation was identified in *SCN2A,* which expresses a sodium channel that is regulated by calmodulin. The mutation disrupts the binding affinity between calmodulin and the channel, leading to sustained channel activity during depolarization. In another case, a chromosomal translocation was identified that disrupted *KCNMA1* that codes for a potassium channel that is activated by depolarization or intracellular Ca^{2+} and leads to constitutively depolarized resting membrane potential and anomalous neuronal excitability.

The Role of Common Variants in Autism

The genetic mutations described in the preceding sections are rare and collectively explain less than 10% of autism cases. In large samples of genotyped individuals, it has been estimated that common variants explain between 40% and 60% of the liability for the disorder, and in 2019, after many underpowered gene-mapping investigations failed to identify any significant findings, an international consortium assembled a sample of almost 20,000 ASD cases and 30,000 controls to identify the first robust GWA study findings: five replicated risk alleles for the disorder. Analysis of transcriptome data from fetal and adult brains showed that the ASD risk loci were associated with molecular pathways impacted by rare and/or de novo variants found in nonfamilial cases of Autism, including *PTBP2*, *CADPS*, and *KMT2E*, demonstrating the shared cellular biology between familial and nonfamilial cases of the disorder.

The same ASD GWA study described above integrated the ASD sample with a large independent sample that included phenotypes on educational attainment and found that the ASD sample shared genetic factors associated with *higher* educational achievement and intelligence. The finding of shared genetic influence between intelligence and ASD diagnosis was only found among higher functioning individuals with ASD (e.g., individuals classified with Asperger syndrome), showing that the spectrum of higher functioning autism disorders can be distinguished from other pervasive developmental disorders based on their genetic architecture. In fact, combined with results from other genetic studies, the emerging picture of autism and other developmental disorders demonstrates that many highly heritable psychiatric disorders are positively correlated with intelligence and have extensive polymorphic contributions from small effect common genetic factors. These are etiologically distinct from the heterogenous cluster of sporadic pervasive developmental disorders that have lower heritability (i.e., less influenced by inherited alleles) and a higher prevalence of de novo mutations that have a larger impact on cognitive functioning.

Summary

Genetic investigation of autism has progressed considerably in the last several years. Successes in autism genetics may be attributed largely to three factors: (1) the high heritability of many of the forms of this disorder, (2) international collaborations that have made large samples of autism families readily available to the scientific community for a wide range of investigations, and (3) the relatively consistent findings obtained using complementary approaches, including linkage studies, investigation of known chromosomal abnormalities, comparative genomic hybridization, mutation analyses, and investigation of animal models that display phenotypes relevant to autism. Compared to other psychiatric disorders like schizophrenia, bipolar disorder, and major depression, there are now relatively elaborated models of the molecular biology and cellular pathways that are involved in the etiology of ASDs, which appear to have significant overlap between familial and nonfamilial versions of the disorder. The progress is largely attributable to the investigation of syndromic cases of autism which are caused by rare mutations of large effect, but emerging evidence is beginning to uncover the genetic architecture of familial cases. Interestingly, well-powered GWA studies of familial autism show that there is a common set of cellular pathways between syndrome and familial cases but also show that the genetic basis of familial cases diverges from syndromic cases for genetic factors associated with cognitive functioning and intelligence. Familial cases of ASD are associated with genes that contribute to higher intellectual functioning, whereas nonfamilial cases are associated with higher rates of de novo mutations that predispose to lower intellectual functioning. Recent progress has been enlightening and has much more to discover; the current genetic loci associated with ASD explain only a small proportion of the total heritability, indicating that there are many loci remaining to be discovered.

BIPOLAR DISORDER, SCHIZOPHRENIA, MAJOR DEPRESSION, AND ANXIETY

Bipolar Disorder

The search for the genetic basis of bipolar affective disorder has been fraught with missteps and partial answers. The history of genetic mapping attempts for bipolar disorder illustrates not only the extreme complexity of psychiatric disorders but also the evolution of genetic approaches to such diseases. Numerous genetic epidemiologic investigations conducted over several decades have strongly supported a genetic contribution to the risk for bipolar disorder. Current estimates of concordance for bipolar disorder range between 65% and 100% in MZ twins and between 10% and 30% in DZ twins, indicating that the disorder is highly heritable (between about 60% and 80%). Several studies have shown that bipolar disorder is substantially more heritable than unipolar major depression, which has an estimated heritability between 30% and 40%. Early family studies suggested that bipolar disorder segregation patterns were compatible with single-gene inheritance of a locus of major effect. Although it is possible that some bipolar disorder pedigrees segregate such a locus, mounting evidence indicates that if such pedigrees exist, they must be quite rare. Furthermore, the fact that genetic linkage studies have failed to uncover such a locus with unequivocal evidence in any pedigrees argues against this possibility. The observed rapid decrease in recurrence risk for bipolar disorder from MZ cotwins to first-degree relatives is also not consistent with single gene inheritance models but rather suggests models of multiple interacting genes.

Tremendous excitement followed the first reports of linkage to bipolar disorder on chromosomes X and 11 in 1987. Investigators noted that in several families, bipolar disorder and other affective disorders appeared to be inherited in an X-linked fashion. Likewise, these disorders appeared to cosegregate in several Israeli families with color blindness and G6PD deficiency, which map to the X chromosome. Linkage studies in these pedigrees using color blindness or G6PD deficiency as marker loci, gave LOD scores between 4 and 9. Early studies of chromosome 11 were similar to those for chromosome X in that they reported significant linkage after testing only a few markers in a single region, in this case, in an extended Old Order Amish pedigree heavily loaded for bipolar disorder.

Not surprisingly, these findings generated a great deal of interest. Both studies showed high LOD scores and seemed to provide clear evidence for linkage. However, replication studies in other populations failed to produce positive results for either the X chromosome or chromosome 11, and evidence for linkage essentially disappeared in both chromosomal regions in the samples in which linkage was originally reported when the pedigrees were extended to include additional affected individuals and when additional markers were typed in the putative linkage regions. The most likely explanation in each case is that the original linkage results were false-positive findings and may have reflected an overoptimistic interpretation of evidence that, in retrospect, was relatively scant.

The early linkage studies of bipolar disorder evaluated only a few markers because they were all that were available. With the construction of genetic linkage maps of the genome in the 1990s, linkage studies of most complex traits, including bipolar disorder, began to search genome wide. The advantage of genome-wide mapping studies is that they do not require a priori knowledge of the biologic underpinnings of a particular phenotype. Complete genome screens provide an opportunity to evaluate the evidence of linkage at all points in the genome without bias. While genome-wide studies clearly had greater power to detect true linkage than studies focused on only a few markers in arbitrary locations or around a few candidate genes, these investigations have also generally had disappointing results. The challenge of achieving replicated significant linkage results for bipolar disorder and other complex traits is apparent when one reviews the many gene-mapping studies that have suggested—but not demonstrated unequivocally—bipolar disorder susceptibility loci on chromosome 18.

The first report of linkage came from a partial genome screen that examined 11 markers on chromosome 18 and identified suggestive linkage near the centromere. Because the inheritance patterns for bipolar disorder are unknown, the results were analyzed using both recessive and dominant models. Some of the markers were positive under a recessive model in some families, some were positive under a dominant model in other families, and some markers gave positive LOD scores in a subset of families under both models. Attempts to replicate this finding in other populations have been mixed. So far, at least two groups have found no evidence for linkage to the pericentromeric region of chromosome 18 in their samples, although one other group has found evidence to support linkage to this region. Other studies have found suggestive evidence for linkage on chromosome 18, including a complete genome screen in two large Costa Rican pedigrees that gave evidence for linkage on chromosome 18q22–23 as well as in an area on 18p. The combined evidence of these several studies, although somewhat contradictory and confusing, points to at least two different susceptibility loci on chromosome 18: one on 18p and another on 18q.

The complexities outlined above for chromosome 18 highlight the potential difficulties encountered in gene-mapping studies. Replications of significant findings in small studies often fail in samples of different families. One approach to overcome the low power of small linkage studies is to use meta-analyses to combine numerical results of several individual studies to estimate the average effect of interest. The largest meta-analysis to date included almost 1,000 families, but even for this large sample, it is still debatable whether any unequivocal linkage signals were identified. Various explanations have been put forward to interpret the lack of findings for these meta-analyses; at the top of the list is the possibility that they have been statistically underpowered. Another consideration is that multiple genes of small effect are interacting to cause the disorder, in which case these genes will be difficult to detect with linkage approaches. Another possibility is that each family is segregating a unique set of genetic variants, and linkage analysis has little power to identify risk loci in these situations. Ultimately, it may turn out that the genetic architecture of bipolar disorder includes a mix of common variants segregating in the general population and rare alleles that are specific to recent ancestors and only segregating within a relatively small set of families. Much of the recent gene-mapping effort in psychiatric genetics has focused on identifying common risk loci of small effect using GWA methods. Compared to several decades of frustrating linkage studies, recent GWA investigations are appreciating considerable success (described below). However, individual common variants identified in GWA studies account for only a small proportion of disease risk (~1% to 2%). Therefore, family-based linkage methods have a potentially important role in the full elaboration of the genetic architecture of psychiatric disorders by identifying rare risk variants that occur in the background of common risk variants and determining which individuals develop the disorder.

In contrast to the disappointing results from linkage analysis described above, GWA studies of bipolar have begun to yield statistically rigorous significant findings. In 2007, the Wellcome Trust Case Control Consortium (WTCCC) conducted the first GWA study of bipolar disorder in 2,000 cases and 3,000 controls genotyped at ~500,000 SNPs but did not identify any genome-wide significant risk variants. The negative finding was notable given that six other common disorders were included in the WTCCC, all of which generated some significant findings. In 2008, a similarly sized investigation was implemented in samples from the Systematic Treatment Enhancement Program for Bipolar Disorder and University College of London (STEP and UCL) and also failed to identify any genome-wide significant risk loci. The WTCCC, STEP, and UCL then combined their data to obtain a sample size of 4,387 cases and 6,209 controls and finally reported the first genome-wide significant association for the disorder for an SNP residing in the *ANK3* (ankyrin G) gene on chromosome 10q21.2. The *ANK3* finding has subsequently been replicated multiple times in additional studies, which have also extended the list of risk loci such that by the end of 2020, about 30 loci had been established as confirmed risk factors for bipolar disorder. As expected for population-based GWA findings, the effect sizes of the risk alleles are low (odds ratios between 1.1 and 1.3), indicating that individually these common variants only contribute to a small increase in disease risk.

Although a definitive list of genes and biologic pathways impacted by risk loci requires more work, there is emerging confidence in several findings. The *ANK3* gene has been identified in multiple GWA studies and encodes ankyrin B, which is highly expressed in brain tissue and is involved in the organization of the neuronal axon. Several of the SNPs associated with bipolar disorder are also associated with expression levels of the ankyrin B. A conditional knockout of the gene in pyramidal neurons of the mouse forebrain results in disruption to the voltage-gated sodium and potassium channel architecture of the axon. Interestingly, the knockout mice showed disrupted behavior resembling mania, which was modified with lithium. A second gene with strong evidence linked to bipolar disorder is the *CACNA1C* gene, which is also associated with autism (see above), depression, and schizophrenia (see below). *CACNA1C* encodes an L-type voltage-gated channel involved in neuronal development and synaptic functions. A third gene strongly associated with bipolar disorder is *TRANK1*, which is also associated with schizophrenia and may be involved in blood–brain barrier physiology. Most of the risk loci associated with bipolar disorder have not been firmly linked to biologic functions. Pathway analysis leverages the growing body of evidence in gene ontology databases to group genes into specific biologic pathways and has provided evidence for the involvement of a range of cellular processes in bipolar disorder, including ion channelopathies, endocrine regulation, and posttranscriptional modification. The current understanding of the genetic architecture of bipolar disorder, which was murky just 5 years ago, has started to come into focus and in combination with family-based studies and intermediate phenotype approaches, promises to translate into clear mechanistic disease models.

Schizophrenia

Similar to bipolar disorder, the mapping of genetic variants associated with schizophrenia began with strong epidemiologic evidence of heritability followed by pedigree-based linkage studies that produced equivocal results, then followed by a period of rapid discovery of common, small effect variants through GWA studies. As of 2021, there are over 200 replicated risk variants with plans to continue increasing sample sizes through international consortia. Prior to successful gene mapping, family studies had clearly established that schizophrenia was strongly influenced by genetic factors, with heritability estimates between 60% and 80%. In the early 1990s, family-based linkage studies began to identify chromosomal regions linked to the disorder. Chromosome 6p24–22 was among the first regions to be implicated by a complete genome screen for schizophrenia, in this case, from a study of Irish families heavily loaded for schizophrenia. Six additional linkage studies have shown positive results over approximately the same region, but at least three studies have found no linkage to the region. Fine-scale mapping of this region using association analysis in the original Irish kindreds led to the proposal of *Dysbindin* (*DTNB1*) as a candidate gene for schizophrenia. Additional association studies of *Dysbindin* have been equivocal. Separate linkage studies subsequently pointed to a region on chromosome 1 containing the candidate genes *DISC1* and *DISC2* (*disrupted in schizophrenia 1* and *2*) located on chromosomes 1q21–22 and 1q32–42. These genes were initially identified in a large Scottish pedigree in the early 1990s. A balanced translocation between chromosomes 1 and 11 was segregated in this pedigree and was possibly associated with serious mental illness. *DISC1* and *2* were identified in the original Scottish family because of their location near the chromosomal translocation breakpoint. As with *Dysbindin,* follow-up studies of *DISC1* and *2* have been equivocal.

Around the same time as the linkage studies mentioned above, a genome-wide screen focused on extended Icelandic kindreds identified a schizophrenia candidate region on chromosome 8p21–22. Fine mapping of the region narrowed the search and eventually led to the proposal of *Neuregulin 1* (*NRG1*) as a schizophrenia candidate gene. Association studies again provided equivocal and difficult-to-interpret results. Meta-analysis of 14 separate studies using the SNP marker that demonstrated an association in the original study showed significant heterogeneity between the follow-up studies. It also showed that there is no consistent association between the specific risk allele “tagged” by the marker SNP and schizophrenia in different populations. However, after taking into account the statistical power of each association study, the meta-analysis showed a positive association between *NRG1* at the level of the gene (as opposed to the SNP or haplotype level).

Despite the equivocal genetic studies, significant resources have been channeled into molecular and neurophysiologic investigations of the functional products of *Dysbindin, DISC1*, and *Neuregulin.* Mutant mice for each of the three genes are now available and have been used to demonstrate interesting biologic findings. For example, *Dysbindin* is expressed in the hippocampus and dorsolateral prefrontal cortex. The dysbindin protein binds to β-dystrobrevin and has been implicated in synaptic structure and signaling. *DISC1* has been shown to influence neurite formation in cellular studies, and mutant mice for *DISC1* show impairments in a wide variety of tests, including learning, memory, and sociability. Neuregulin belongs to a family of growth factors that mediate numerous functions, including synapse formation, neuronal migration, and neurotransmission. Targeted disruption of *erbB4,* the postsynaptic target of neuregulin, leads to synaptic glutamatergic hypofunction. Despite the interesting biology uncovered, it remains unclear whether and to what extent any of these genes contribute to the etiology of schizophrenia in humans and many geneticists have been cautious in their endorsement of the legitimacy of the mutant mice generated from the current list of candidate genes as models of psychiatric disorders.

Like other psychiatric disorders, GWA studies of schizophrenia produced disappointing results until 2009, when the International Schizophrenia Consortium, Molecular Genetics of Schizophrenia, and the Schizophrenia Working Group of the Psychiatric Genomics Consortium combined samples and identified an SNP in the major histocompatibility complex (MHC) on chromosome 6p22 as a risk locus for the disorder. The mapping of a schizophrenia risk locus to the MHC region was an exciting discovery, although the full significance of the finding is difficult to interpret. It has long been known that exposure to maternal infection is a risk factor for schizophrenia. Therefore, specific MHC configurations may lead to anomalous immune responses that impact brain development and/or processes that predispose to the disorder. In addition, neurons express MHC molecules where they are involved in several aspects of brain development, including synaptogenesis, neurite outgrowth, and long-term plasticity. Therefore, genetic variations in the MHC may directly impact brain development outside immunomodulatory mechanisms.

Shortly after the discovery of the MHC as a risk locus, GWA studies identified more than a dozen additional risk loci, and as sample sizes increased, the number of novel and replicated risk loci accumulated. In 2014, the Schizophrenia Working Group of the Psychiatric Genomics Consortium, which by then had reached a sample size of 37,000 cases and 113,000 controls, identified 108 distinct loci contributing to risk for schizophrenia.

It is important to note that the identification of a genetic association at a specific SNP does not imply that the genotyped SNP is the causal allele. In fact, it is unlikely that the genotyped SNP on most current genotyping platforms has functional significance. The genotyped SNPs are more likely in LD in the immediate vicinity (∼100 kb) of the causal variant. In addition, given the extensive sequence variation even within relatively small regions, establishing a specific variant or set of variants as causal factors in the development of a complex disorder is not trivial. Complicating the definitive identification of variants is the expectation that many of the risk loci identified by GWA studies will harbor causal variants that impact more subtle aspects of gene functioning compared to Mendelian genes of large effect. While Mendelian mutations often lead to major disruptions in the coding sequencing of genes, common risk variants of small effect have been hypothesized to impact gene functioning at the level of expression or regulation. In contrast to mutations that impact protein-coding regions, it is more difficult to identify mutations that lie outside the intron/exon structure of genes that may be affecting regulatory regions that control expression or chromatin structure. Ongoing initiatives like the ENCODE project aim to characterize biochemically active regions of the genome, including regulatory protein binding sites and open chromatin regions, thereby providing a valuable resource for guiding the interpretation of mutations in noncoding regions.

Although the elaboration of the specific biologic pathways involved in the pathogenesis of schizophrenia will require definitive identification of causal mutations and their impact on specific genes, the recent accumulation of risk loci provides suggestive directions. The risk loci do not appear to be randomly distributed among genes involved in all cellular processes, but rather appear to be enriched near genes involved in a smaller subset of processes important for brain functioning, including neurodevelopment (FXR1 on 3q26.33, SATB2 on 2q33.1, PODXL on 7q32-q33, BCL11B on 14q32.2, TLE1 on 9q21.32, TLE3 on 15q22, and BRINP2 on 1q24), glutamatergic neurotransmission (GRIN2 A on 16p13.2, GRIA1 on 5q33.2, SRR on 17p13.3, CLCN3 on 4q33, and GRM3 on 7q21.12), neuronal calcium signaling (CACNA1C on 12p13.3, CACNB2 on 10p12, CAMKK2 on 12q24.2, NRGN on 11q24, ATP2A2 on 12q24.11, CACNA1I on 22q13.1, and RIMS1 on 6q12–13), and synaptic function and plasticity (KCTD13 on 16p11.2, NLGN4X on Xp21.33–32, IGSF9B on 11q25, CNTN4 on 3p26.3, MEF2C on 5q14.3, PTN on 7q33, CNKSR2 on Xp22.12, PAK6 on 15q14, and SNAP91 on 6q14.2).

Although GWA studies have appreciated substantial success over the past several years, it has become clear that common variants segregating in the general population are not able to account for all of the estimated genetic contribution to psychiatric disorders; this fact has been labeled as the “missing heritability” problem. For example, based on twin and family studies, the heritability for schizophrenia is around 70%, but the total amount of variance explained by common genetic polymorphisms is estimated to be about 30%, such that there is a “gap” of unexplained heritability. Various explanations have been put forward to account for the missing heritability problem. One factor that is likely to explain at least some of the missing heritability is the possibility that rare genes of medium-to-large effect that are not detectable with GWA methods are present in affected individuals. As with all other psychiatric disorders, the future directions of genetic mapping for schizophrenia will include continued efforts to (1) a mass ever-larger sample sets to increase power for population GWA studies, (2) identify rare variants of medium/large effect that are present in single individuals (i.e., de novo mutations) or segregating in extended families, and (3) using intermediate phenotype approaches to characterize the genetic architecture of the biologic pathways involved in the pathogenesis of the disorder.

Major Depression

Major depression is the most common psychiatric disorder and is one of the leading causes of morbidity worldwide. Lifetime prevalence varies across countries but generally ranges between 8% and 12%. Evidence for a genetic contribution to major depression comes from family studies that show an increased risk of the disorder in first-degree relatives of probands that is two to three times greater than the general population. The heritability of major depression is generally lower than other mood and psychotic disorders and is estimated to be between 30% and 40%. Large population studies have consistently shown that there is a higher prevalence of the disorder in women, and twin studies have shown that the disorder is, in fact, more genetically influenced in women, with an estimated heritability of ~40% compared to ~30% for men. These same studies have estimated that although a substantial proportion of genetic causes of major depression are shared between men and women (genetic correlation between the sexes is estimated between 0.6 and 0.9), a significant proportion of genetic factors are sex specific. In addition to genetic differences between the sexes, there also appear to be genetic differences between subtypes of depression; for example, recurrent depression appears to have a higher genetic influence compared to single-episode (heritability of 40% vs. 30%), and postpartum depression has higher heritability compared to nonpostpartum depression (50% vs. 30%). The differing estimates of genetic influence across sexes and disorder subtypes highlight the heterogeneity of the disorder and the fact that many complex interacting biologic and environmental factors influence the emergence or resilience of the disorder in each individual.

Prior to implementing GWA studies, genetic research was largely focused on testing candidate genes for influence on depression. Starting four decades ago, researchers began testing specific genes or small clusters of genes related to neurotransmitters and associated metabolic pathways that were hypothesized to be relevant to the biology underlying depression. More than 1,500 investigations of over 200 purported candidate genes have been published over the past decades with conflicting and confusing results. Recent GWA studies have adequate power to identify significant associations reported in candidate gene studies but have consistently failed to do so. A focused analysis in very well–powered datasets of 18 of the most commonly studied candidate genes found evidence for the *DRD2* (dopamine receptor D2) gene but failed to find any evidence for any of the other genes indicating that the vast majority of “significant” candidate gene findings reported to date are false positives.

Like other neuropsychiatric disorders, the genetic architecture of major depression is expected to include a mix of rare or novel variants of medium-to-large effect and many common loci of small effect. To date, family-based linkage studies have failed to identify any strong findings for rare variants contributing to the disorder. Recent whole-exome and whole-genome sequencing studies have provided suggestive evidence that individuals with major depression have higher burdens of deleterious single exonic variants, but larger sample sizes will be needed to confirm these findings. There is also suggestive evidence that short deletion CNVs mapped to enhancer regions are enriched in individuals with major depression. Despite the suggestive evidence for CNVs and exonic variants, the characterization of novel and rare variants will require larger datasets and more refined analysis methods (e.g., analyses of disorder subtypes and intermediate phenotypes).

Unlike the unsuccessful searches for rare variants, GWA studies of major depression have started to bear fruit. However, compared to other psychiatric disorders like schizophrenia, the lower heritability and higher heterogeneity of major depression have led to the need for much larger discovery samples. For example, GWA studies of schizophrenia began appreciating rigorous success with sample sizes of around 15,000 and are identifying new loci, on average, with each new addition of around 250 cases. In contrast, the first rigorous genetic findings for major depression required 75,000 cases and need, on average, about 1,500 new cases for the discovery of new risk loci. The first successful GWA was achieved in 2016 by the direct-to-consumer company 23andMe using self-reported depression diagnoses. In subsequent years, GWA studies by the Psychiatric Genomics Consortium and the Million Veterans Program added to the list of confirmed risk loci, and currently, over 100 risk loci have been identified. Collectively, the overall variance explained by recent GWA studies explains only about 3% of all the estimated genetic contribution to the disorder, again highlighting the highly polygenic nature of major depression and the need for increasing sample sizes and more refined analyses of disorder subtypes and intermediate phenotypes.

Gene-level and pathway analyses have been used to aggregate associations across the genome and explore the biologic pathways that contribute to the disorder. These analyses have implicated synaptic machinery, voltage-gated calcium channels, glutamate receptors, genes involved in neuronal growth and differentiation, and genes involved in immunity and inflammation. The most robustly supported individual genes are *NEGR1* (neuronal growth regulator 1) and *OLFM4* (olfactomedin 4). Each of these loci has a small effect size with an estimated odds ratio of 1.04. *NEGR1* is a cell adhesion molecule involved in neuronal outgrowth and synapse formation. *OLFM4* is secreted glycoprotein involved in inflammation and immunity. Other genes within the set of confirmed loci also participate in these biologic processes confirming the emerging picture that common variants of small effect accumulate within a subset of biologic pathways, including synapse function, neuroreceptors, and immunity/inflammation to collectively bend an individual toward major depression contingent on the individual's exposure to environmental risk factors.

Anxiety

Pathologic forms of the otherwise normal emotions of fear and anxiety are among the most common mental health conditions, with a lifetime prevalence of over 20%. Prior to DSM-V, OCD and post-traumatic stress disorder were included in the category of anxiety disorders but have been removed from the category due to their distinct characteristics and biologic mechanisms. The anxiety disorders now include the diagnoses of generalized anxiety disorder, panic disorder, and phobias. Genetic mapping studies of anxiety disorders have followed the same pattern as other psychiatric disorders; first, epidemiologic investigations showed evidence for familial aggregation, followed by pedigree-based linkage studies that identified largely nonreplicated loci and numerous candidate gene studies that produced inconsistent findings, and finally, GWA studies that have started to produce rigorous findings of common variants in large population samples.

Epidemiology investigations of anxiety disorders in families have provided consistent evidence for familial aggregation, with first-degree relatives of patients having a four- to six-fold increased risk for anxiety disorders. Twin studies estimate the heritability of the various anxiety disorders to be between 30% and 50%. Risk within families includes the development of a range of anxiety disorders, suggesting that there is substantial shared genetic overlap among the different diagnoses. Statistically, modeling methods confirm that a relatively uniform set of common genetic factors contribute to the core set of anxiety disorders, with environmental factors influencing the final form of the disorder.

Pedigree-based linkage studies of anxiety disorders have largely focused on families with panic disorder, but the strongest mapping signals have been obtained for broad phenotypes that include panic, phobia, generalized anxiety symptoms, and even medical conditions like headaches. During the late 1990s and early 2000s, a series of investigations identified linkage signals on chromosomes 1q, 2q, 7p, 9q, 13q, 15q, and 22q. Compared to similar studies on Alzheimer disease and bipolar disorder, there have been significantly fewer efforts to map anxiety disorder genes in families and none of the findings have been replicated to date. Around the same time that linkage studies were being conducted, many groups were focused on candidate gene association studies based on hypothesized biologic mechanisms. Like candidate gene studies for other psychiatric disorders, candidate gene studies of anxiety phenotypes produced inconsistent results, and the field now recognizes that the sample sizes were very underpowered to identify statistically significant findings until recent GWA studies were able to accumulate large population samples.

Like GWA studies of other psychiatric disorders, the initial foray into anxiety disorders had underpowered sample sizes and did not generate any replicable findings. In 2016, a meta-analysis of seven previous studies assembled a sample size of over 18,000 cases and identified two genome-wide significant loci, including an SNP in the *CAMKMT* gene on chromosome 3. Shortly after, several groups assembled sample sizes exceeding 20,000 and identified another half-dozen risk loci that included a region on chromosome 9 that had been previously associated with neuroticism. One of the risk genes with the strongest biologic support is *PDE4B*, a phosphodiesterase family gene involved in signal transduction. Altered activity of the protein has been characterized in bipolar disorder, and schizophrenia and mice models deficient in the protein exhibit anxiety-like behaviors. The largest GWA study to date was implemented in the Million Veterans Program with a sample size of almost 200,000 individuals. The investigators used screener questions from the generalized anxiety disorder 2-item instrument (GAD-2) as the phenotype for gene mapping. The GAD-2 does not provide a specific diagnosis but uses self-report on a Likert scale to assign a continuous measure of anxiety based on two questions that query the frequency of anxious feelings and the inability to stop worrying. Using this phenotype, the study identified five significant genome-wide loci, including one on chromosome near the *SATB1* gene, a global regulator of gene expression, and one on chromosome 6 near the estrogen receptor gene *ESR1* which encodes an estrogen receptor. A third loci on chromosome near the *MAD1L1* gene was previously identified in GWA studies of schizophrenia and bipolar disorder. The overlapping findings between anxiety risk loci and other psychiatric disorders highlight the genetic overlap among the disorders and are the focus of the next section.

GENETIC OVERLAP AMONG PSYCHIATRIC DISORDERS

The preceding sections have emphasized that categorically defined psychiatric disorders display substantial genetic heterogeneity, indicating that the symptom clusters that specify diagnostic categories can emerge from mutations in different sets of genes. The converse is also true; mutations in specific genes can cross diagnostic boundaries. Early family studies showed that relatives of an individual diagnosed with a specific disorder (i.e., schizophrenia) are at higher risk of developing not only schizophrenia but also major depression or bipolar disorder. Gene-mapping studies have confirmed that genetic risk loci are shared by different disorders. For example, mutations in the calcium channel gene, *CACNA1C,* have been linked to autism, bipolar disorder, and schizophrenia.

The availability of genotypes for large samples of individuals allows for the estimation of genetic correlations between psychiatric disorders and nondiagnostic traits. The genetic correlation (r_g) between two traits can be determined from family data or from population-based genotypes and provides an estimate of the shared genetic factors that are influencing both traits, also known as pleiotropy. The Psychiatric Genomics Consortium investigated the overlap of genome-wide common variants in a large sample (~30,000 cases and ~30,000 controls) that included patients from five categorically defined psychiatric disorders: autism, attention deficit hyperactivity disorder (ADHD), bipolar disorder, major depression, and schizophrenia. The investigation found a remarkable degree of overlap between schizophrenia and bipolar disorder; about 68% of the polygenic risk attributable to common variants was shared between these two disorders. Major depression also showed a substantial overlap with other disorders, where about 45% of the genetic risk attributable to common variants was shared with both schizophrenia and bipolar disorder, and about 30% was shared with ADHD. Schizophrenia showed a low (~15%) but significant overlap with ASDs. Autism did not show significant genetic overlap with bipolar disorder, major depression, or ADHD. Similarly, aside from the genetic overlap with major depression, ADHD did not show significant overlap with any of the other disorders. These findings seem to parallel clinical scenarios, where, for example, schizophrenia and bipolar disorder can sometimes be difficult to disentangle, but bipolar disorder and autism are usually easy to distinguish. The findings also provide hints into the emerging pattern of cellular pathways underlying psychiatric disorders, alterations in some molecular pathways (e.g., calcium signaling) can perturb neural systems nonspecifically such that the final psychiatric presentation is determined by other genetic and/or environmental factors. This scenario may be particularly true for schizophrenia, major depression, and bipolar disorder, which share a substantial proportion of common genetic risk factors. In contrast, disorders like ADHD do not share as many common genetic risk factors with other disorders and may emerge through alterations of cellular pathways distinct from those leading to psychotic and mood spectrum disorders.

Genetic correlations can also be used to help understand the genetic relationship between psychiatric disorders and nonpsychiatric traits. For example, the GWA study of GAD-2 scores described in the section on anxiety is highly correlated with depression ($r_g = 0.81$) and the personality trait neuroticism ($r_g = 0.75$). Similarly, major depression shares significant genetic factors with neuroticism ($r_g = 0.70$) and subjective well-being ($r_g = -0.75$). The negative estimate for the correlation with subjective well-being indicates that genetic factors that confer increased risk for depression influence well-being in the opposite direction, that is, genetic loci that increase the risk for depression tend to reduce measures of subjective well-being. Major depression also has a negative genetic correlation with educational attainment and cognitive performance, albeit much lower than with the personality traits at about $r_g = -0.15$. The genetic correlations with nonpsychiatric traits can help to start dissecting subtypes of the disorder. For example, there appears to be shared genetic factors between atypical depression and body mass index (BMI) in patients that experience weight gain but not in patients who do not experience changes in appetite or weight. As genetic correlations between psychiatric disorders and a wide range of medical and nonmedical traits become available, we will develop a clearer understanding of the complex heterogeneity that exists in mental health. This may be especially important for disorders like depression, which reflects a common endpoint for many diverse and independent processes that manifest in a relatively common cluster of symptoms.

FUTURE DIRECTIONS

The field of psychiatric genetics is in a state of transition. New technologies have made the genetic exploration of complex behaviors a real possibility. While 35 years ago, human gene-mapping studies involved the use of a few markers in a small number of individuals, it is now routine to genotype millions of markers in samples of tens of thousands of individuals. The development of exciting new technologies for phenotyping the human nervous system—such as high-resolution neuroimaging—and the implementation of many such assays in genetic mapping studies have created extraordinary opportunities for the large-scale association of phenotypes and genotypes in psychiatric research. While the analysis of such vast and complex datasets remains a substantial challenge, the development of methodologies for this purpose is now a central focus of research in statistical genetics and bioinformatics.

As the field of psychiatric genetics progresses, several major questions that have emerged may soon be answered, at least partially, through the analysis of large-scale datasets. One major question is whether genetic risk for psychiatric disorders derives mainly from a few rare variants of large effect or many common variants of small effect (or some combination of rare and common variants). While well-powered GWA studies should identify common variants, they have very little power to identify rare variants. The failure thus far of linkage studies to identify rare causative variants for psychiatric disorders does not preclude the possibility that rare variants of somewhat lesser effect could play an important role in susceptibility to these diseases. Whole-genome sequencing of large samples is now feasible and is being implemented with the goal of identifying rare variants segregating in families with heavy genetic loading for psychiatric disorders.

The second question concerns the degree to which psychopathology will be dividable into discrete disorders, as in the current DSM nosology, or if the heterogeneity and overlap among syndromes are so great as to require a thorough overhaul of the current classification systems. Estimates from population-based investigations of common genetic variations show that the mood and psychotic disorders share a substantial overlap of genetic risk factors, whereas disorders like autism and ADHD may be more genetically distinct from other psychiatric disorders. To elaborate on this issue more completely, it will be necessary to begin examining how genetic risk factors interact with other risk genes and the environment to predict the symptomatology in specific individuals. In addition, the increasing focus of the field on the genetic investigation of brain, behavioral, and other endophenotypes for major psychiatric disorders may be a step in the direction of such a revolution.

A third major question is characterizing the role of interactions between genes and environment in the etiology of disorders. Gene-environment interactions occur when specific phenotypes emerge or are influenced due to the interplay of an individual genetic background with specific features of the environment. Identifying gene-environment interactions requires even more statistical power than identifying first-order correlations between genes and phenotypes, and although many such studies have been published to date, they have been underpowered to make firm conclusions. However, as sample sizes with adequate assessments of demographic and historical data continue to grow, the rigorous identification of gene-environment interactions will become possible. For example, a recent investigation of over 500,000 individuals from the U.K. Biobank with genetic and socioeconomic data as measured by the Townsend Deprivation Index showed: (1) socioeconomic scores are associated with anxiety, bipolar disorder, depression, self-harm, and overall health, and (2) suggestive associations from a genome-wide scan of gene-by-environment analysis between specific genetic loci that contributed to disease phenotypes depending on the socioeconomic scores. The interpretation is that certain genetic loci will contribute to disease risk only in the presence of the environmental factor. As big data accumulates with methodically acquired data related to socioeconomics, trauma, exposure to substances, and other known or suspected influences on mental health, gene-environment analyses will provide powerful tools to characterize the complex pathways that contribute to risk and resilience and will provide a guide to understanding both the etiology and the treatment of such disorders.

The final question concerns the impact of anticipated psychiatric genetic discoveries on health outcomes—in terms of either prevention or improved treatments. For Alzheimer disease, concrete advances have been made, and answers to this question may be obtained soon. For the other disorders discussed in this chapter, rapid advances have been made in the last several years, but the emerging picture is still fluid, and it may still be several years before enough is understood to frame this question in answerable terms.

FURTHER READINGS

Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet.* 2006;7(10):781–791.

Bienvenu OJ, Davydow DS, Kendler KS. Psychiatric 'diseases' versus behavioral disorders and degree of genetic influence. *Psychol Med*. 2011;41(1):33–40.

Cardno AG, Rijsdijk FV, West RM, et al. A twin study of schizoaffective-mania, schizoaffective-depression, and other psychotic syndromes. *Am J Med Gen B Neuropsychiatr Genet*. 2012;159B(2):172–182.

Chanock SJ, Manolio T, Boehnke M, et al; NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype-phenotype associations. *Nature*. 2007;447(7145):655–660.

*Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits; *Nat Rev Genet.* 2014;15(1):34–48.

Craddock N, Sklar P. Genetics of bipolar disorder. *Lancet*. 2013;381(9878):1654–1662.

*Flint J, Kendler KS. The genetics of major depression. *Neuron*. 2014;81(3):484–503.

Flint J, Munafò M. Schizophrenia: genesis of a complex disease. *Nature*. 2014;511(7510):412–413.

Dennison CA, Legge SE, Pardiñas AF, Walters JTR. Genome-wide association studies in schizophrenia: recent advances, challenges and future perspective. *Schizophr Res.* 2020;217:4–12.

Glahn DC, Knowles EEM, McKay DR, et al. Arguments for the sake of endophenotypes: examining common misconceptions about the use of endophenotypes in psychiatric genetics. *Am J Med Genet B Neuropsychiatr Genet*. 2014;165B(2):122–130.

Grove J, Ripke S, Als TD, et al; Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium, BUPGEN, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, 23andMe Research Team. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431–44.

Kellis M, Wold B, Snyder MP, et al. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A*. 2014;111(17):6131–6138.

*Kendall KM, Van Assche E, Andlauer TFM, et al. The genetic basis of major depression. *Psychol Med.* 2021;51(13):2217–2230.

*Lee SH, Ripke S, Neale BM, et al; Cross-Disorder Group of the Psychiatric Genomics Consortium, International Inflammatory Bowel Disease Genetics Consortium (IIBDGC).

Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013;45(9):984–994.

Meier SM, Deckert J. Genetics of anxiety disorders. *Curr Psychiatry Rep*. 2019;21:16.

Munafò MR, Thiselton DL, Clark TG, Flint J. Association of the NRG1 gene and schizophrenia: a meta-analysis. *Mol Psychiatry*. 2006;11(6):539–546.

Mutsuddi M, Morris DW, Waggoner SG, Daly MJ, Scolnick EM, Sklar P. Analysis of high-resolution HapMap of DTNBP1 (Dysbindin) suggests no consistency between reported common variant associations and schizophrenia. *Am J Hum Genet*. 2006;79(5):903–909.

Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet*. 2015;16:275–284.

O'Tuathaigh CMP, Babovic D, O'Meara G, Clifford JJ, Croke DT, Waddington JL. Susceptibility genes for schizophrenia: characterisation of mutant mouse models at the level of phenotypic behaviour. *Neurosci Biobehav Rev*. 2007;31(1):60–78.

Pinto D, Delaby E, Merico D, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet*. 2014;94(5):677–694.

Reiman EM, Webster JA, Myers AJ, et al. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron*. 2007;54(5):713–720.

Roybal K, Theobold D, Graham A, et al. Mania-like behavior induced by disruption of CLOCK. *Proc Natl Acad Sci U S A*. 2007;104(15):6406–6411.

Rzhetsky A, Wajngurt D, Park N, Zheng T. Probing genetic overlap among complex human phenotypes. *Proc Natl Acad Sci U S A*. 2007;104(28):11694–11699.

*Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421–427.

Shih RA, Belmonte PL, Zandi PP. A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *Int Rev Psychiatry.* 2004;16(4):260–283.

Sklar P, Pato MT, Kirby A, et al. Genome-wide scan in Portuguese Island families identifies 5q31–5q35 as a susceptibility locus for schizophrenia and psychosis. *Mol Psychiatry*. 2004;9(2):213–218.

Stahl EA, Breen G, Forstner AJ, et al; eQTLGen Consortium, BIOS Consortium, Bipolar Disorder Working Group of the Psychiatric Genomics Consortium. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet*. 2019;51(5):793–803.

Stanewsky R. Genetic analysis of the circadian system in Drosophila melanogaster and mammals. *J Neurobiol*. 2003;54(1):111–147.

Tanzi RE. The genetics of Alzheimer disease. *Cold Spring Harb Perspect Med*. 2012;2(10):a006296.

Thomas DC. *Statistical Methods in Genetic Epidemiology*. Oxford University Press; 2004.

Visscher PM, Hill WG, Wray NR. Heritability in the genomics era–concepts and misconceptions. *Nat Rev Genet*. 2008;9(4):255–266.

Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. *Nat Rev Genet*. 2014;15(11):765–776.

1.20 Epigenetics in Psychiatry: The Promise for New Biomarkers and Treatments

ARIE KAFFMAN, M.D., PH.D.

In 1942 Conrad Waddington coined the term “*epigenetics*” by fusing the terms “*genetics*” and “*epigenesis*”—an old term used to describe the developmental program by which the zygote gives rise to different cell types in a multicellular organism. Epigenetics, as defined by Waddington, represented the cooperation between genetic traits and environmental conditions that together established a stable cellular phenotype. The notion that nongenetic factors play an important role in development was further formalized by defining epigenetics as “a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence.” More recently, the term epigenetics has been further broadened to emphasize the contribution that environmental factors make to the formation of a stable cellular phenotype that may or may not be heritable. In this review, the latter definition was adopted, and the term epigenetics is used to describe “*changes in gene expression that lead to predictable, stable, and possibly heritable alterations in cell function that are not due to changes in DNA sequence*.” This definition is particularly relevant in psychiatry, where environmental factors, such as stress or substance abuse, alter the expression of genes in postmitotic neurons that are not heritable but contribute to psychiatric pathologies such as major depression, anxiety, and addiction.

Table 1.20–1.
Differences and Similarities between the Genetic and the Epigenetic Codes

Variable	Genetic Code	Epigenetic Code
Components	Long strand of DNA composed of four nucleotides (A, G, C, T)	Histones, histone-modifying enzymes, structural proteins, DNA modifying enzymes, ATP-dependent chromatin-remodeling machinery
Structure	One-dimensional linear strand	Dynamic three-dimensional structure
Cellular content	Identical in all cells	Unique for each cell type Dynamic, may change in response to environmental cues and cell cycle progression
Environmental impact	No change in response to environmental changes	Rapid changes in response to environmental changes, could also be long lasting
Can modify gene expression	Yes	Yes
Heritable	Yes	Sometimes

Table 1.20–1 summarizes some of the key similarities and differences between the “*genetic*” and the “*epigenetic*” codes. Structurally, the genetic code is made of a one-dimensional sequence of nucleotides that are joined together to form a long DNA strand. The epigenetic code, on the other hand, is composed of a host of proteins and noncoding RNAs that coat the DNA and transform it into a three-dimensional structure that can be compressed, decompressed, and twisted into loops. The genetic code is identical in all the cells in the body, while the epigenetic code is highly variable between different cell types and may also change during different phases of the cell cycle within the same cell. Epigenetics thus comprises the mechanisms by which the same genetic blueprint is used to generate different cell types. This concept is illustrated in Figure 1.20–1, where the same genetic code, represented as a linear sequence of letters, is selectively masked by two different epigenetic programs to read STAND (top) or SIT (bottom), depending on what letters within the underlying code are made visible. The genetic code is static, whereas the epigenetic code is highly dynamic and can change rapidly in response to environmental cues. This dynamic property allows postmitotic neurons to modify the expression of genes that drive changes in synaptic plasticity, learning, and behavior. Finally, the genetic

and the epigenetic codes share some similarities. These include the ability to modify gene expression in a manner that can be transmitted across generations.

Epigenetics research is one of the most exciting and rapidly growing frontiers in psychiatric research. This is due to its potential to advance the field in six specific areas (summarized in Table 1.20–2). First, epigenetics provides a molecular mechanism to explain how environmental factors shape vulnerability to psychiatric conditions. In some cases, such as substance abuse or post-traumatic stress disorder (PTSD), the contribution of the environment is quite obvious. In other examples, such as schizophrenia or bipolar disorder, a strong environmental influence can be surmised from the observation that the concordance between monozygotic twins is slightly less than 50%. One of the goals of this chapter is to provide specific preclinical and clinical examples in which epigenetic changes explain how specific environmental conditions modify the risk for psychopathology. It is important to recognize that environmental manipulations can also induce epigenetic changes that alleviate psychiatric symptoms. In fact, many of the known interventions in psychiatry, such as antidepressant medications, electric convulsive shock (ECS), diet, and exercise, have been shown to cause epigenetic changes that modify vulnerability to stress, anhedonia, and drug-seeking behaviors in animals. A better understanding of how epigenetics mediate environmental risk and protective factors is likely to generate new and perhaps more effective strategies to treat mental illness.

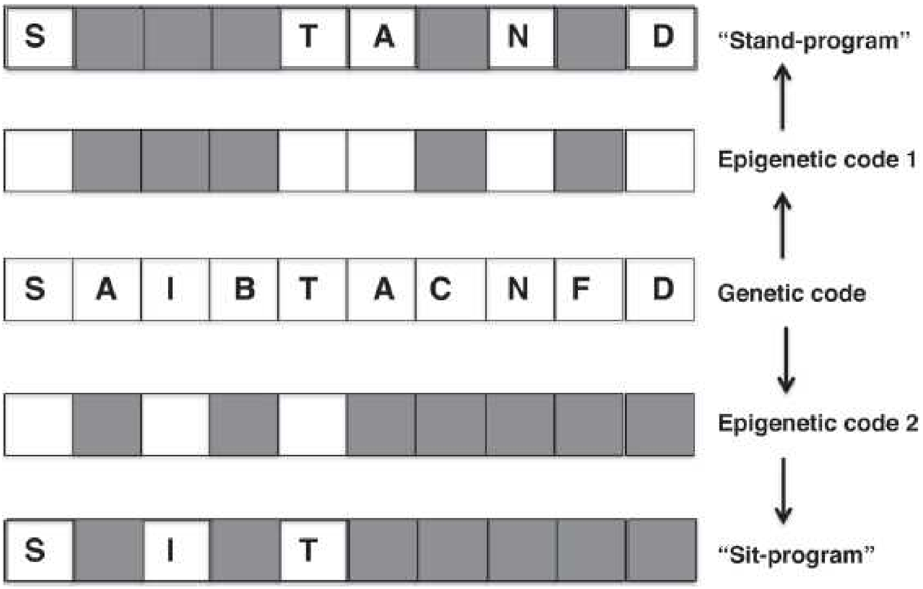

FIGURE 1.20–1. The epigenetic code modifies the genetic instructions in response to environmental demands to generate either a "STAND" (**top**) or "SIT" program (**bottom**).

Second, epigenetics play a central role in guiding neurodevelopment and therefore provides a possible mechanism to explain how abnormal brain development contributes to the vulnerability to many psychiatric conditions. This observation is evidenced by the large number of neurodevelopmental disorders that are caused by epigenetic dysregulation (Table 1.20–3). A better understanding of this process is likely to lead to the development of new treatments (see The Epigenetic Etiology of Fragile X Syndrome section below). Moreover, a growing body of work, some of which is summarized in this chapter, demonstrates that epigenetics play a critical role in modifying brain development and the vulnerability to several psychopathologies in children exposed to abuse or neglect (discussed in detail below).

Third, epigenetics research is responsible for recent advances in generating inducible pluripotent stem cells (iPSCs). This technology allows researchers to harvest somatic cells (e.g., skin cells) from individuals with specific psychiatric illnesses (e.g., schizophrenia) and reprogram their epigenetic code in a way that transforms them into neurons, glial cells, and even entire organoids. These cells can then be grown and studied in culture. This remarkable process solves one of the key hurdles in psychiatric research, which is how to gain access to and manipulate functional brain tissue from humans.

Fourth, despite significant progress over the past decade to map genetic causes of highly heritable psychiatric conditions such as schizophrenia (81%) and bipolar disorder (85%), genetic mutations account for only a small fraction of the cases. There is a growing consensus that epigenetics may also contribute to the heritability of these disorders because heritability is calculated as the relative rates of concordance between monozygotic and dizygotic twins, and monozygotic twins share many more epigenetic similarities than dizygotic twins. In addition, the ability of some epigenetic markers to be inherited may explain how exposure to a specific stressor in the parental generation may modify the risk for psychopathology in the offspring. For example, prolonged parental stress in rodents modifies noncoding small RNA content in the sperm that is responsible for alterations in neurodevelopment, physiology, and behavior in the offspring. These preclinical findings provide a novel conceptual framework for evaluating similar changes in noncoding small RNA sperm content seen in stressed men.

Fifth, the dynamic interactions between the underlying DNA sequence and its packaging chromatin provide a molecular mechanism to explain how genes modify the ability of certain environmental conditions to induce psychopathology. Such gene-by-environment interactions have been elegantly demonstrated between different alleles of the serotonin transporter and the ability of childhood maltreatment to modify the risk for adult depression and are illustrated in this chapter by examining how genetic and epigenetic changes coordinate the expression of FKBP5 and thus influence the risk for developing PTSD.

Table 1.20–2.
Epigenetics Promise in Psychiatry

Contribution	Practical Implications
1. Provides molecular mechanisms to explain how environmental change modifies behavior and psychopathology	Novel therapies
2. Clarifies how abnormal neurodevelopment leads to psychopathology	More effective early interventions
3. Allows for the generation of inducible pluripotent stem cells	Gaining access to “brain tissue” from live patients with mental illness
4. Provides additional ways to explain how psychopathology is inherited	Novel diagnosis and interventions
5. Elucidates the mechanisms by which gene and environmental factors interact to modify vulnerability to mental illness	More specific biomarkers and interventions
6. Provides reliable biomarkers to diagnose psychiatric vulnerabilities	The availability of objective and quantifiable markers to diagnose and assess treatment response

Sixth, many epigenetic markers, such as DNA methylation, are stable biomarkers that can be accurately quantified using a very small number of cells. These properties make epigenetic changes a potentially valuable biomarker for diagnosing and assessing treatment response. Biomarkers are desperately needed in psychiatry, where no objective disease markers are currently available to diagnose or assess treatment response. Recent findings suggesting that the methylation rate of a stress-related gene, SKA2, can be used to predict suicidal behavior are presented in the last section of this chapter. This kind of work exemplifies the utility of epigenetics in identifying useful biomarkers of psychiatric conditions.

This chapter is divided into three sections. The first section provides a comprehensive review of the cellular machinery involved in epigenetic regulation. The second section describes several preclinical examples in which epigenetic changes have been shown to modify psychiatrically relevant phenotypes in animals and how targeting these changes can potentially generate novel psychiatric treatments. The third and final section provides several clinical examples to illustrate how epigenetics may modify clinical understanding and practice in psychiatry.

THE EPIGENETICS MACHINERY

Evolutionary studies in archaebacteria suggest that chromatin originally evolved as a mechanism to rapidly regulate gene expression in response to environmental demands. This more complex multilayered form of transcriptional regulation was later adapted to support cellular differentiation, in which the same genetic code is used to generate different cell types. With the expansion of the genome, chromatin was essential to compact the lengthy DNA so it could fit into a small nucleus. Finally, chromatin serves many other functions that are essential for maintaining the integrity of the genetic code. These include DNA repair and the formation of telomeres and centromeres.

Table 1.20–3.
Chromatin-Mediated Neurodevelopmental Disorders

Disease	Protein/Gene	Chromatin Modifications	Clinical Features
Alpha-thalassemia/mental retardation syndrome, X-linked (ATR-X)	ATRX/XH2, member of SWI/SNF family of proteins (chromatin-remodeling enzyme)	Defective chromatin remodeling that is implicated in down-regulating the α-globin locus	X-linked inheritance; mental retardation, hemolytic anemia, splenomegaly, facial, skeletal, and genital anomalies
Prader–Willi syndrome	The parental 15q11–13 chromosome is maintained in an open chromatin state, allowing for the expression of genes necessary for normal development	Deletion of parental 15q11–13 locus	Initial neonatal hypotonia with a failure followed by hyperphagia, childhood obesity, hypogonadism and mild intellectual disability
Angelman syndrome	The maternal chromatin at the 15q11–13 locus is highly compacted, preventing the expression of UBE-3A antisense RNA and allowing the expression of the UBE3A gene	Deletion of maternal 15q11–13 locus or mutations in the maternal UBE3A gene.	Cortical atrophy, cerebellar abnormal myelination, seizures, mental retardation
Coffin–Lowry syndrome	RSK2 (Kinase)	Abnormal histone acetylation and phosphorylation	X-linked inheritance; psychomotor retardation, skeletal and craniofacial deformities
Fragile X syndrome	FMR1 and FMR2	DNA hypermethylation of expanded trinucleotide repeats at the FMR1 and FMR2 promoters	Most common inherited form of intellectual disability, signs of autistic behavior, long and narrow face with large

			ears and macroorchidism
Immunodeficiency–centromeric instability–facial anomalies syndrome (ICF)	Dnmt3B (De novo DNA methyltransferase)	Hypomethylation at centromeric regions of chromosomes 1, 9, and 16	Mild mental retardation, immunodeficiency, facial anomalies
Rubinstein–Taybi syndrome	CBP (HAT)	Reduced histone acetylation, Autosomal dominant inheritance	Mental retardation, abnormal facial features, blunted growth
Rett syndrome	MeCP2 (MBP)	Abnormal recruitment of MeCP2 to methylated DNA	X-linked, pervasive developmental disorder, predominantly affecting girls with arrested brain development, cognitive decline, and autistic-like behavior
The 9q34 subtelomeric deletion syndrome	GLP (HMT)	Abnormal H3K9 methylation	Severe mental retardation, hypotonia, facial anomalies

The basic folding unit of chromatin is the *nucleosome*. The nucleosome consists of a histone octamer core wrapped by a DNA loop (Fig. 1.20–2A). The octamer core is assembled from four highly conserved, positively charged histone proteins (H2A, H2B, H3, and H4). H3 and H4 form a tetramer that is sandwiched between two H2A/H2B dimers forming the octamer core. The octamer core is wrapped by a 146bp DNA loop, a structure that is stabilized by large numbers of ionic contacts made between the negatively charged DNA and positively charged amino acids found in histones. Each histone protein has a long unstructured N-terminal tail that protrudes and probes the surrounding space. This tail plays a critical role in the ability of the chromatin to condense or open up.

In the open chromatin conformation, also known as "*beads on a string*," neighboring nucleosomes are separated by a short (10 to 80 bp) linker DNA (Fig. 1.20–2B). The "beads on a string" chromatin structure alters the properties of the underlying DNA in two important ways. First, it compacts the DNA by about 5- to 10-fold. Second, significant portions of the wrapped DNA are not accessible for the binding of transcriptional activators. As a result, even in its most accessible form, chromatin represents a barrier to basal transcription. Moreover, the long N-terminal tails on histones make contacts between neighboring nucleosomes that tend to compact the chromatin. These higher-order filaments can be stabilized by the binding of a fifth histone, histone H1, forming higher-order 30-nm filaments that further compact the DNA, making it even less accessible. This natural tendency of chromatin to coil into less accessible, more compact structures is regulated by several mechanisms that include posttranslational modifications of the histone N-terminal tails, recruitment of chromatin-remodeling machinery, and the use of invariant histones, described in some detail below. In summary, the nucleosome unit regulates accessibility to the DNA sequence, which is critical for gene expression without altering the DNA chemistry or genetic code.

Histones Modifications

The N-terminal tails of histones are decorated by a large number of posttranslational modifications such as acetylation, methylation, phosphorylation, ubiquitylation, and sumoylation as some of the main examples. These modifications are reversible and are carried out by large multisubunit protein complexes that coordinate their activities in a way that can either condense or open the chromatin in response to environmental cues. Histone modifications affect chromatin structure via three distinct mechanisms. First, some modifications disrupt ionic contacts between positively charged residues on histone tails and the negatively charged DNA backbone, loosening DNA-histone contacts. Second, nucleosomes contact one another via their N-terminal tails. These inter-nucleosomal contacts can be strengthened or disrupted by posttranslational modifications. Third, alterations of the histone tails provide binding sites that recruit chromatin-remodeling machinery. These proteins can use ATP hydrolysis to slide or remove nucleosomes from chromatin (Fig. 1.20–3).

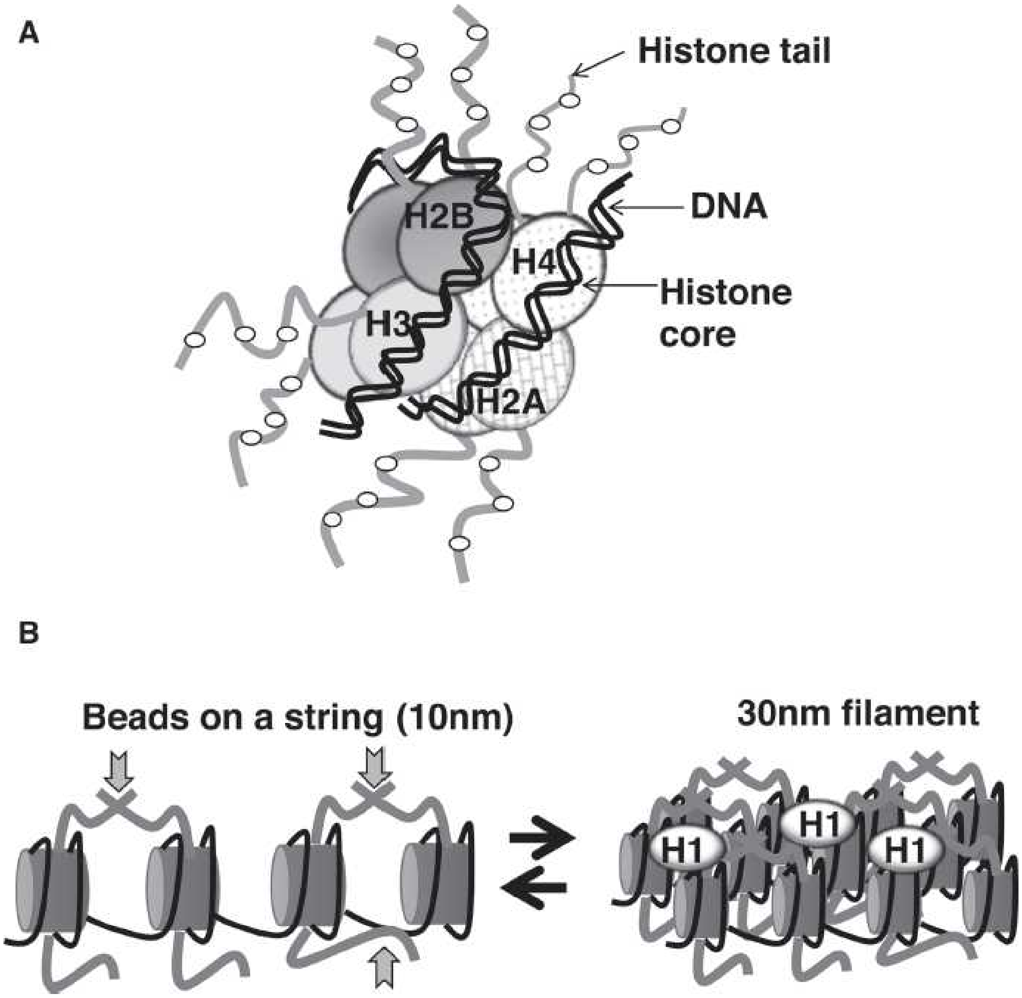

FIGURE 1.20–2. **A:** The nucleosome as a basic folding unit of chromatin. **B:** Chromatin can transition from a "beads on a string" 10-nm filament to a more compact 30-nm filament structure that is stabilized by the presence of histone H1 (*light ovals*). Histone tails (in *gray*) make contacts between neighboring nucleosomes known as internucleosomal contacts (*top arrows*) and with the underlying DNA (*bottom arrow*).

Histone Acetylation. Acetylation neutralizes lysine residues, rendering the N-terminal tails on histones less positively charged. This decrease in positive charge is thought to disrupt contacts with the negatively charged DNA and with neighboring nucleosomes. Histone acetylation is associated with transcriptional activation and other cellular processes that require access to the underlying DNA, such as DNA replication and repair (Table 1.20–4). Histone acetyltransferases (HATs) acetylate lysine residues on the tails of histones H3 and H4, while histone deacetylases (HDACs) remove these acetylations (Fig. 1.20–4). Acetylation modifies lysines (K) that are systematically spaced on Histone 3 (e.g., 9, 14,18, and 23) and Histone 4 (e.g., 5, 8, 12, and 16). Different HATs and HDACs are responsible for acetylating and deacetylating specific residues. On histone H3, for example, K14 and K18 are acetylated by cAMP response element-binding protein (CREB)-binding protein (CBP/P300), while K9, K14, and K18 are acetylated by a different HAT, PCAF/GCN5.

This enzyme specificity provides an opportunity to develop more selective and targeted pharmacotherapies to act on this system. Recent work using reconstituted purified histones found that a single acetylation at lysine 14 in histone 4 (H4K14-Ac) was sufficient to maintain chromatin in the “beads on a string” state by preventing interactions between neighboring nucleosomes. Interestingly, the H4K14-Ac modification inhibited the ability of the chromatin-remodeling enzyme ACF to slide nucleosomes, suggesting that this modification “locks” chromatin into an open state.

Consistent with histone acetylation opening chromatin, deacetylation of histones leads to chromatin compaction and transcriptional repression (Fig. 1.20–4). HDACs fall into three classes, Class I, Class II, and the NAD-dependent Class III of the SIR family. These classes are based on structural similarities and are therefore inhibited by different pharmacologic agents. In general, HDACs seem to have less specificity than HATs with regard to the lysine resides they deacetylate. Different HDACs are targeted and bind to different compartments within the nucleus, allowing them to conduct somewhat distinct functions. For example, HDAC1 seems to play a general role in chromatin condensation and repression, while HDAC3 is tethered to the nuclear periphery, where it forms highly repressive null chromatin (see Chromatin Domains section below).

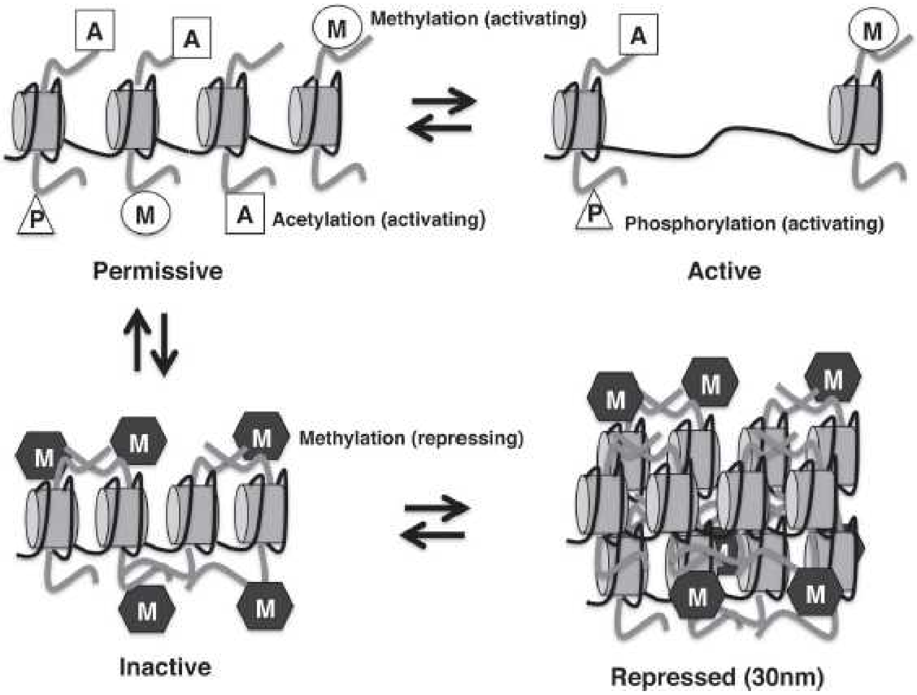

FIGURE 1.20–3. "Permissive chromatin" (**top left**) is acetylated (*gray squares*), methylated (*open ovals*) and phosphorylated (*gray triangle*). These histone modifications open up the chromatin by disrupting internucleosomal contacts and DNA–nucleosome contacts and recruit ATP-dependent chromatin-remodeling enzymes (not shown) that further open up the chromatin to form "active chromatin" (**top right**). Permissive chromatin can be modified to form "inactive chromatin" (**bottom left**) by removing activating histone modifications and the addition of inhibitory histone methylations (windowpane-patterned hexagon). These repressive modifications compact chromatin by promoting internucleosome and DNA–nucleosome contacts. Inactive chromatin can be further condensed to form "repressed chromatin" (**bottom right**) by recruiting ATP-dependent chromatin-remodeling enzymes and other structural proteins (not shown).

Table 1.20–4.
Enzymatic Activities That Modify Chromatin

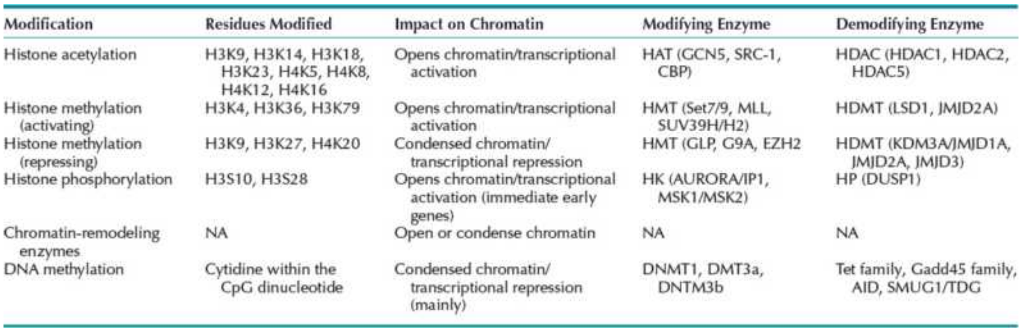

Modification	Residues Modified	Impact on Chromatin	Modifying Enzyme	Demodifying Enzyme
Histone acetylation	H3K9, H3K14, H3K18, H3K23, H4K5, H4K8, H4K12, H4K16	Opens chromatin/transcriptional activation	HAT (GCN5, SRC-1, CBP)	HDAC (HDAC1, HDAC2, HDAC5)
Histone methylation (activating)	H3K4, H3K36, H3K79	Opens chromatin/transcriptional activation	HMT (Set7/9, MLL, SUV39H/H2)	HDMT (LSD1, JMJD2A)
Histone methylation (repressing)	H3K9, H3K27, H4K20	Condensed chromatin/ transcriptional repression	HMT (GLP, G9A, EZH2	HDMT (KDM3A/JMJD1A, JMJD2A, JMJD3)
Histone phosphorylation	H3S10, H3S28	Opens chromatin/transcriptional activation (immediate early genes)	HK (AURORA/IP1, MSK1/MSK2)	HP (DUSP1)
Chromatin-remodeling enzymes	NA	Open or condense chromatin	NA	NA
DNA methylation	Cytidine within the CpG dinucleotide	Condensed chromatin/ transcriptional repression (mainly)	DNMT1, DMT3a, DNTM3b	Tet family, Gadd45 family, AID, SMUG1/TDG

Proteins that have a bromodomain-folding motif are able to bind acetylated histones in order to maintain open chromatin. Mutations in the HAT domain of CBP cause Rubinstein–Taybi syndrome, which is associated with mental retardation and cardiac abnormalities (Table 1.20–3). Interestingly, a mouse model of this condition showed that the cognitive deficits associated with low levels of CBP could be reversed by administrating HDAC2 inhibitors to adult mice. These findings suggest that the relative activities of HATs and HDACs in postmitotic neurons play an important role in neuroplasticity with broad implications for neurodevelopmental disorders, aging, and drug addiction (see Epigenetics in Preclinical Work). Indeed, increased HDAC2 has been shown to silence genes that are necessary for learning and memory, and HDAC2 activity was found to be elevated in Alzheimer's disease. Clinical work is currently underway to test whether specific HDAC inhibitors can be used to improve cognition in conditions such as early Alzheimer's disease, schizophrenia, and traumatic brain injury.

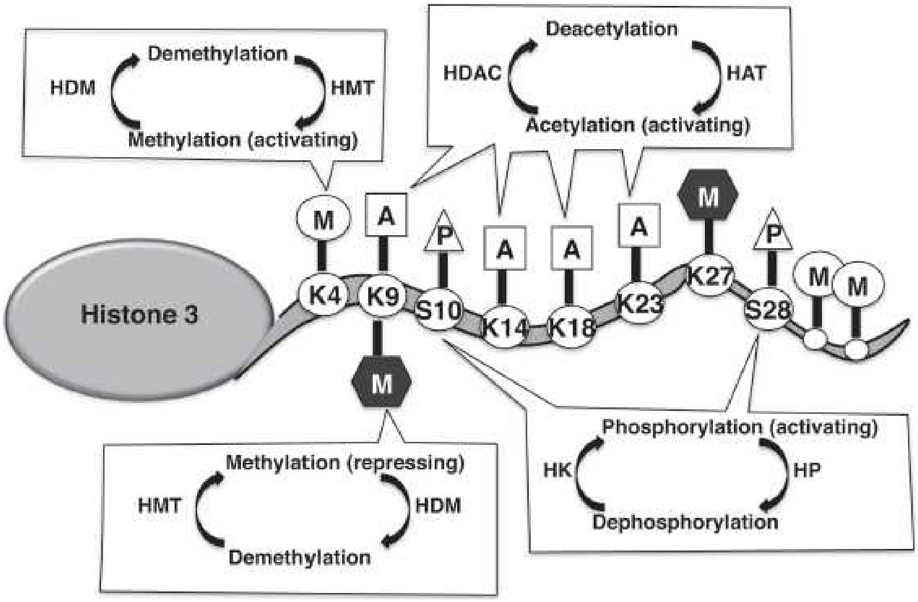
Demethylation
HDM
HMT
Methylation (activating)
Deacetylation
HDAC
HAT
Acetylation (activating)
Histone 3
M
A
P
A
A
A
M
P
M
M
K4
K9
S10
K14
K18
K23
K27
S28
M
Methylation (repressing)
HMT
HDM
Demethylation
Phosphorylation (activating)
HK
HP
Dephosphorylation

FIGURE 1.20–4. Histone tails are reversibly modified at specific residues. HAT-Histone acetyltransferases, HDAC-histone deacetylase, HMT-histone methyltransferase, HDMT-histone demethyltransferase, HK-histone kinase, HP-histone phosphatase.

Histone Methylation. Histone methylation is more complex than acetylation because both lysines (K) and arginines (R) can be methylated. Moreover, each lysine can be methylated once, twice, or three times, providing another layer of complexity. Lysine methyltransferases are more selective compared to acetylating enzymes, targeting specific lysine residues within the histone tail. Methylation on three sites is associated with transcriptional activation (H3K4, H3K36, H3K79), whereas methylation on 3 other lysine resides is associated with repression (H3K9, H3K27, H4K20), see Table 1.20–3 for a summary. In addition to transcriptional regulation, histone methylation also plays a role in DNA repair.

As with other chromatin modifications, methylated lysines can be demethylated by histone demethylases, allowing for dynamic and reversible chromatin changes (Fig. 1.20–4). Proteins that contain chromo-like domains (e.g., chromo, tudor, MBT) or the structurally unrelated PHD motif are able to bind methylated histones. Such proteins form large complexes that contain chromatin-remodeling enzymes and other histone modifiers. These complexes are recruited to methylated histones and, depending on the exact composition of the complex, are able to efficiently open or close the chromatin. For example, appropriate activation of the HOXC8 gene is essential for normal axis development in all vertebrates. Activation of the HOXC8 gene requires histone methylation during a specific period in development. This methylation step is required to recruit the nucleosome-remodeling factor (NURF) via the bromodomain PHD finger transcription factor (BPTF). The NURF complex contains a chromatin-remodeling enzyme (SNF2L) that removes nucleosomes from the HOXC8 promoter allowing for transcriptional activation. Mutations that interfere with the ability of BPTF to bind the methylated promoter cause abnormal expression of HOXC8 and abnormal axis formation in tadpoles. This sequence of events illustrates the multilayered coordination of chromatin regulation in which histone modification recruits additional machinery making the chromatin more accessible and allowing for the expression of genes necessary for normal axis development (Fig. 1.20–5).

Methylation at lysine 9 in histone 3 (H3K9-me2/ H3K9-me3) plays a critical role in maintaining a specialized form of chromatin known as *pericentric/constitutive heterochromatin*. One of the main functions of the pericentric chromatin is to support the formation of centromeres that are used to align and segregate chromosomes during cell division (see also Table 1.20–4). H3K9 is methylated by a specific histone methyltransferase (Suv39h1/h2), which in turn triggers the binding of the repressor protein HP1 via its chromodomain. HP1 then associates with several other proteins to form a complex that extends this specialized repressed chromatin structure through large sections of the underlying DNA. This is accomplished by the ability of HP1 to recruit additional Suv39h1/h2, allowing further histone methylation and HP1 recruitment to propagate down the DNA (Fig. 1.20–6). HP1 is also associated with DNA methyltransferases, DNMT1 and DNMT3a, allowing for DNA methylation and further chromatin compaction. Finally, during replication, HP1 is associated with chromatin factor 1 (CAF1) protein, which deposits histones at the replication fork. The presence of a complex containing HP1, Suv39h1/h2, DNMT1, and CAF1 at the replication fork provides a possible mechanism by which pericentric chromatin is inherited in daughter cells and a good example of how DNA methylation and histone modification activities are coordinated (see DNA Methylation section below).

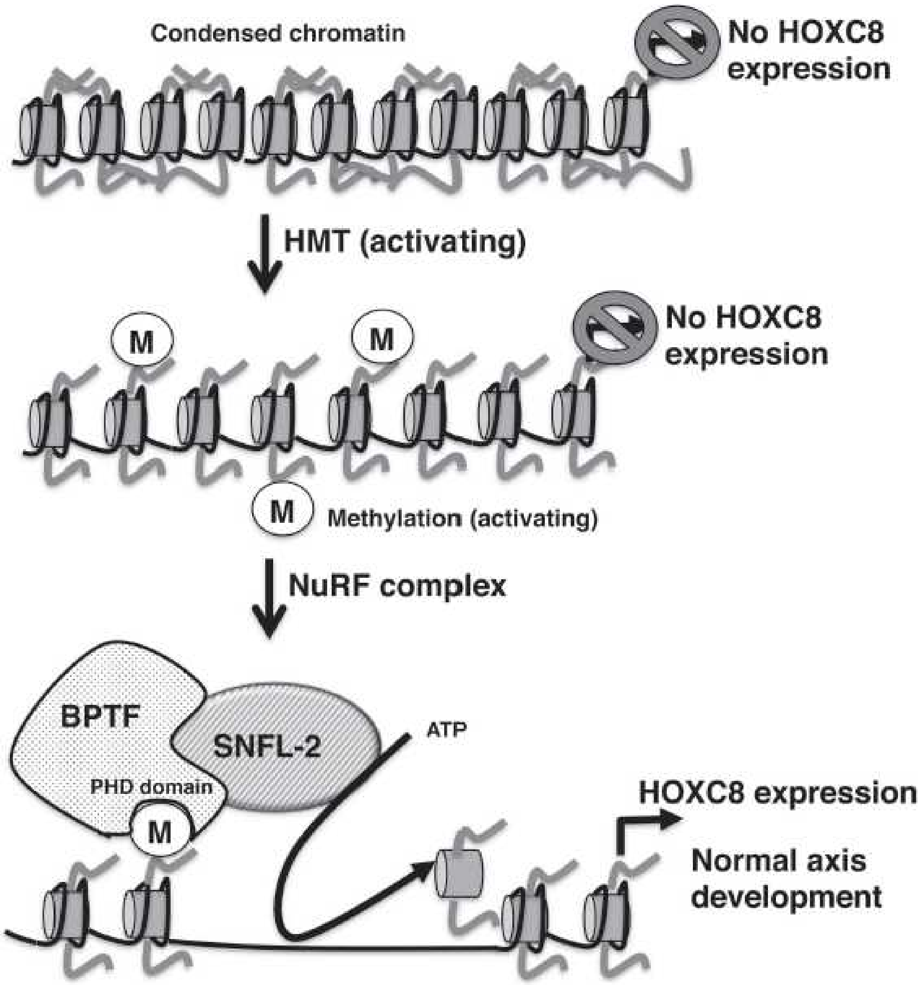
Condensed chromatin
No HOXC8 expression
HMT (activating)
M
M
No HOXC8 expression
M
Methylation (activating)
NuRF complex
BPTF
SNFL-2
PHD domain
M
ATP
HOXC8 expression
Normal axis development

FIGURE 1.20–5. Histone methylation recruits BPTF via its PHD domain bringing with it the chromatin-remodeling enzyme SNFL2 as part of the NuRF complex. Binding of SNFL2 allows for ATP-dependent removal of nucleosomes from the promoter, HOXC8 expression, and normal axis development in vertebrates.

Histone methylation is also critical for the creation of repressive polycomb chromatin structures, which are instrumental in maintaining the pluripotency of embryonic stem cells, regulating cell cycle progression, cell differentiation, and chromosome-X inactivation. In this process, histone methylation at H3K27 is catalyzed by the Polycomb Repressive Complex (PRC2), a highly evolutionarily conserved protein complex. PRC2 is recruited to specific regions of DNA via interactions with other DNA-binding proteins and noncoding RNA. Once recruited, PRC2 methylates H3K27 to form the di, and the tri methylated states (H3K27-me3), which then serve as tethering sites for the entire complex. H3K27-me3 also recruits the chromatin-remodeling machinery (PRC1) that further condenses chromatin into the highly repressive polycomb conformation (see Fig. 1.20–7 and Table 1.20–4). Interestingly, the PRC2 complex maintains its binding to DNA during DNA synthesis, providing a mechanism by which H3K27-me3 and polycomb chromatin repression are maintained in daughter cells. Mutations in the histone methylase activity of PRC2 prevent the stable propagation of polycomb repression into daughter cells, suggesting an essential role for methylation in this epigenetic form of inheritance, with critical implications for many developmental programs, including X-inactivation and the transition from cellular pluripotency to differentiation.

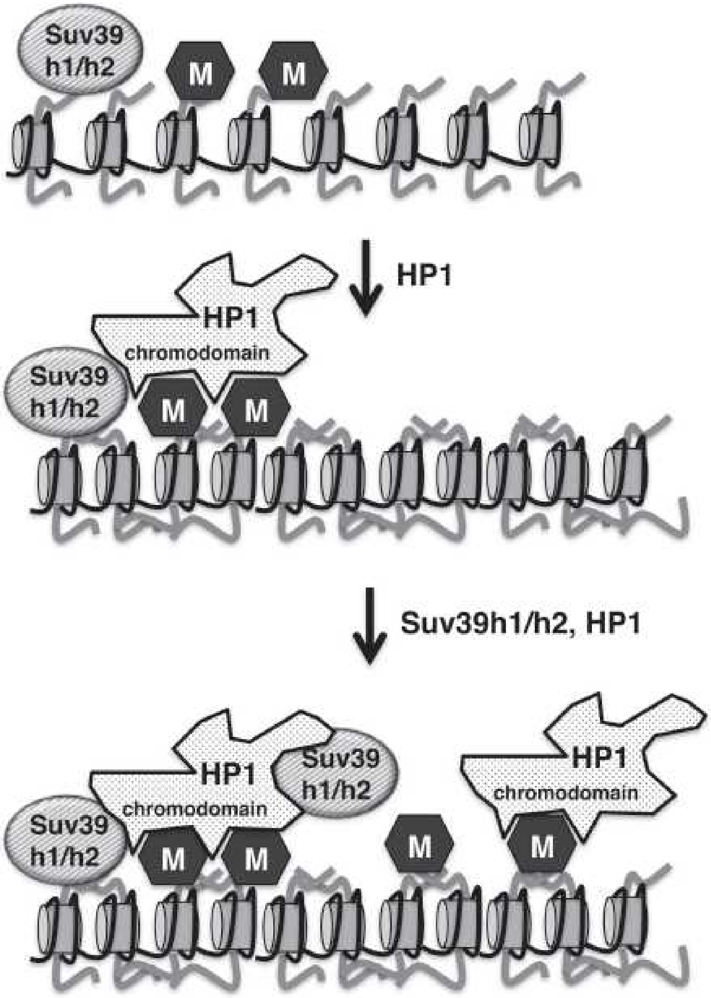

FIGURE 1.20–6. The HMT Suv39h1/h2 methylates H3K9 (**top**), recruiting HP1 via its chromodomain (**middle**). HP1 binds Suv39h1/h2 allowing for the propagation of this repressive pericentric/constitutive heterochromatin (**bottom**).

Histone Phosphorylation. Phosphorylation of histones plays a role in chromosome segregation during mitosis, induction of early response genes, and DNA repair. Histone kinases such as Aurora/Ip1 and the mitogen and stress-activated protein kinases 1 and 2 (MSK1/MSK2) phosphorylate serines (S) 10 and 28 on histone 3. The proximity of these serine residues to lysines 9 and 27 influences the decision as to whether these lysines will be acetylated or methylated. This influence is critical because acetylation at lysines 9 and 27 tends to activate transcription, whereas methylation at these same sites represses transcription (Fig. 1.20–4). Several environmental stimuli, such as stress, cocaine, changes in neurotransmission, and UV light, induce H3S10 phosphorylation by activating MSK1 and MSK2. Activated MSK1 is recruited to the promoters of immediate early genes, such as c-fos and Jun, in a complex containing 14-3-3 proteins, chromatin-remodeling machinery, and a HAT. This complex efficiently activates the expression of immediate early genes that provide appropriate cellular responses depending on the type of cells and the environmental context. In the dentate gyrus, this response facilitates helplessness behavior, while in the striatum, this response modifies drug-seeking behavior (see Epigenetics in Preclinical Work section below).

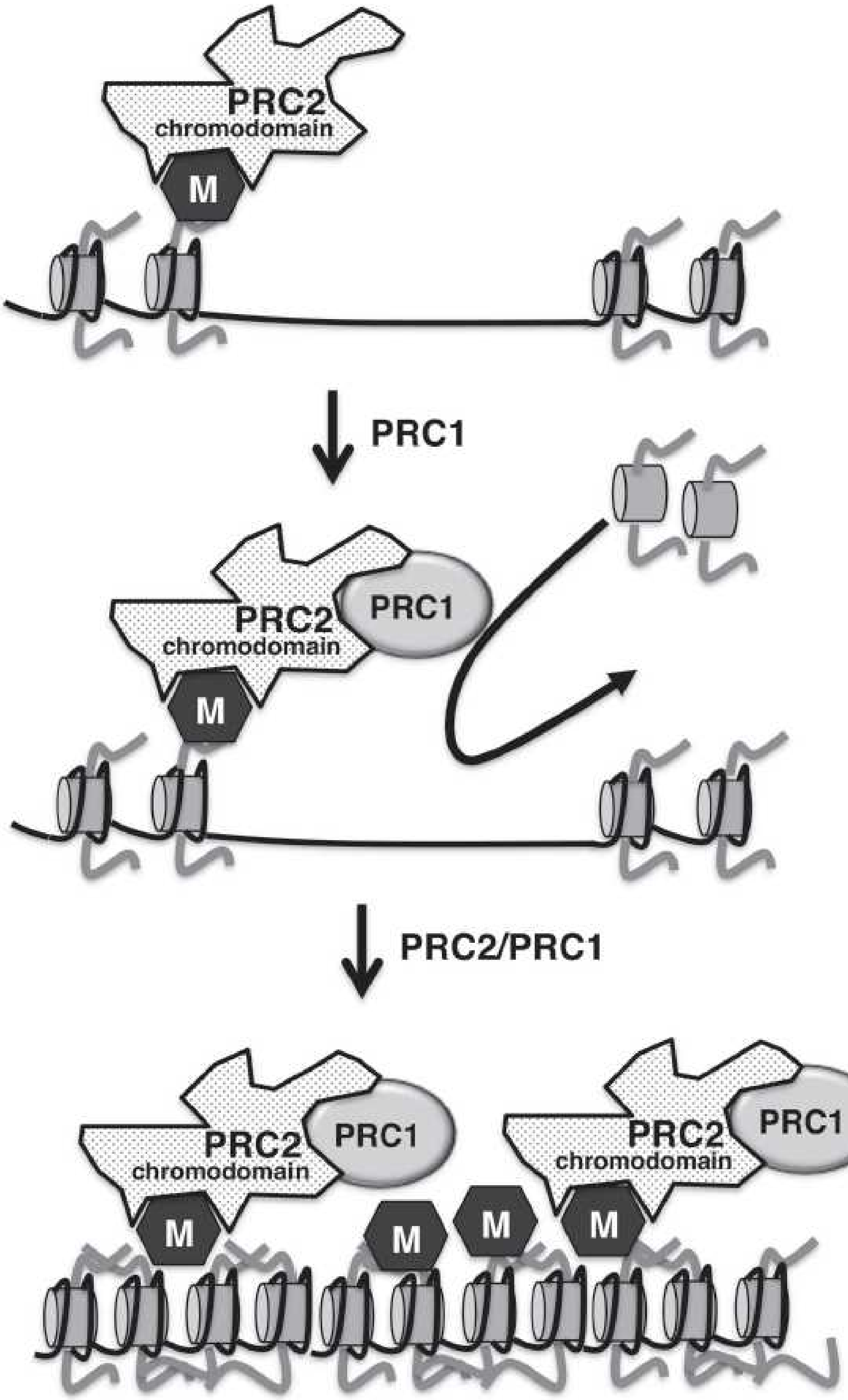
PRC2
chromodomain
M
PRC1
PRC2
chromodomain
PRC1
M
PRC2/PRC1
PRC2
chromodomain
PRC1
M
M
M
PRC2
chromodomain
PRC1
M

FIGURE 1.20–7. The PRC2 complex has an HMT activity that generates H3K27-me3, which further stabilizes PRC2 binding to chromatin (**top**). PRC2 recruits the PRC1 complex that further compacts chromatin by increasing nucleosome density (**middle**). These newly recruited nucleosomes are methylated by PRC2, allowing for the propagation of this polycomb repressive chromatin (**bottom**).

ATP-Dependent Chromatin Remodeling

Nucleosomes are distributed in a stereotypical pattern along actively transcribed housekeeping genes. This pattern is highly conserved from yeasts to mammals and is characterized by a 250 bp nucleosome-free region (NFR) that is flanked by two positioned nucleosomes in the promoter of housekeeping genes (Fig. 1.20–8). Similar NFRs are also found at the 3' end of the gene and in enhancer regions. This pattern allows the transcriptional machinery to bind and constitutively express housekeeping genes. Nucleosome positioning is accomplished in vivo by two mechanisms. First, specific DNA sequences within the promoter region can guide the nucleosome's position, providing an important example of how the genetic code influences the epigenetic code. However, this arrangement is static and does not allow for the rapid reorganization of chromatin in response to environmental demands. The second and perhaps more important mechanism is mediated by chromatin-remodeling enzymes. These enzymes are equipped with chromatin binding motifs and use ATP to remove, add, or slide nucleosomes along the DNA strand. In some cases, chromatin-remodeling enzymes maintain NFR by removing nucleosomes from specific promoter sequences. This, in turn, allows transcription factors to bind DNA and activate transcription (Fig. 1.20–8). Alternatively, nucleosomes can be pushed by chromatin-remodeling enzymes into NFRs, where they block transcription and form higher-order chromatin structures (Fig. 1.20–8). Recruitment of chromatin-remodeling enzymes has been implicated in the development of addictive-like behaviors after chronic cocaine exposure in mice via modification of the promoters of late-onset addiction-related genes, such as cyclin-dependent kinase 5 (CDK5) (see Epigenetic Mechanisms of Cocaine Addiction section).

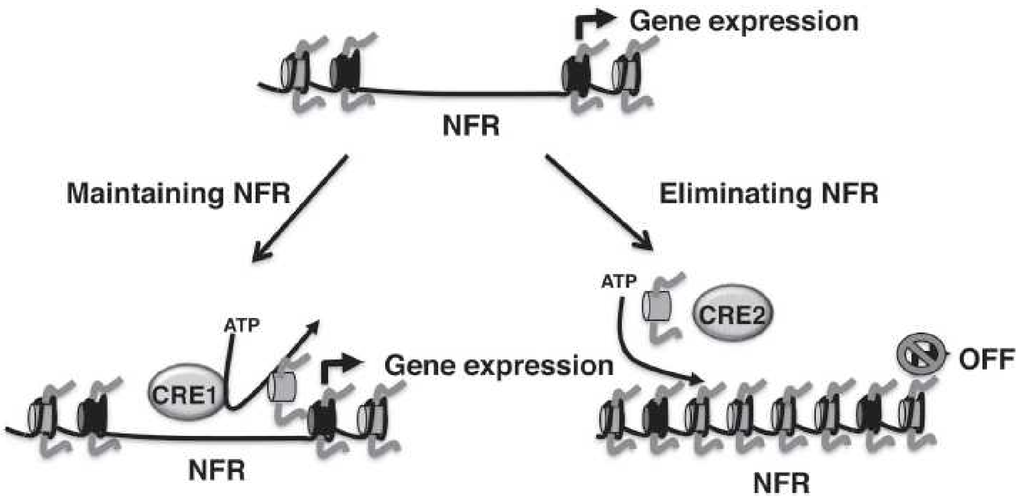

FIGURE 1.20–8. Top: Nucleosome-free region (NFR) can be passively maintained by DNA sequences that position nucleosomes on both sides of the NFR (shown in *dark gray*). This, in turn, allows transcription factors to bind and activate gene expression. In many cases, NFR is actively maintained by ATP-dependent removal of nucleosomes (depicted as CRE1 on the **lower-left panel**). ATP-remodeling activities (CRE2, on the **right lower panel**) use ATP to deposit nucleosomes making the NFR inaccessible and therefore turning transcription off.

DNA Methylation

Unlike histone modifications that do not alter the chemical structure of the genetic code, DNA methylation directly changes the molecular structure of cytosines. This is accomplished by transferring a methyl group from a donor molecule to the 5′ position on the cytosine ring to generate 5-methylcytidine (5mc). In vertebrates, DNA methyltransferase (DNMT) enzymes methylate cytosines in a dinucleotide (CpG) palindromic sequence. Cytosine methylation does not affect its ability to form correct Watson-Crick pairing with a guanosine on the antiparallel DNA strand and thus does not affect the underlying genetic information. Instead, DNA methylation marks the DNA at specific locations in a manner that can be inherited and allows the recruitment of histone and nucleosome-remodeling machinery to do the heavy lifting of opening or compressing chromatin.

Within the mammalian genome, CpG dinucleotides are either clustered in a high-density region known as CpG islands (CGIs) or found sporadically across the genome. CGIs usually mark promoters of active housekeeping genes and are thus generally unmethylated. In rare cases, when CGIs are methylated, they form a highly repressive chromatin structure that inhibits gene expression and, in many cases, is inherited. Some well-known examples include X-inactivation and imprinted genes. The observation that abnormal regulation of imprinted genes leads to several forms of intellectual disability demonstrates the importance of correct methylation of these CGIs for normal neurodevelopment (see Table 1.20–3 and Rescuing Prader–Willi Syndrome section below).

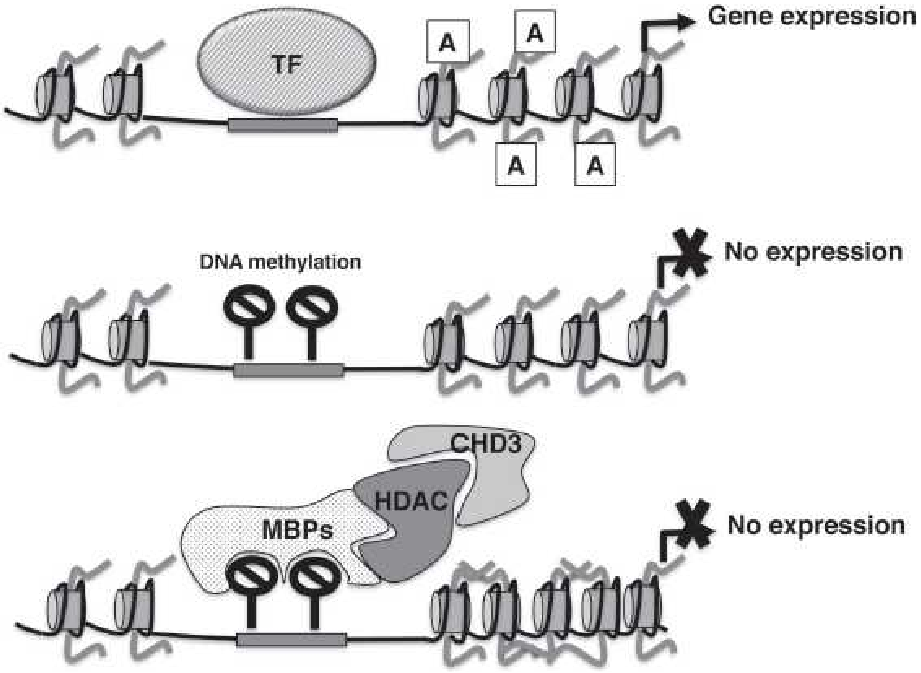

FIGURE 1.20–9. **Top:** In the absence of DNA methylation, a generic transcription factor (TF) binds a specific DNA sequence within the promoter and activates gene expression. **Middle:** DNA methylation (*black* no entry signs) prevents the TF from binding its binding site and therefore prevents gene expression. **Bottom:** The NuRD complex is recruited to methylated DNA via methyl-binding proteins (MBPs). These MBPs are complexed with enzymatic activities such as HDACs and the chromatin-remodeling protein CHD3 to compact DNA and terminate gene expression.

Methylation of CpG in sporadic regions usually down-regulates gene expression. This downregulation is achieved in one of two ways. First, DNA methylation can prevent the binding of a transcription factor to its binding site (Fig. 1.20–9). A second mechanism involves the direct binding of a methyl-binding protein (MBP) to methylated cytosines. Docking of MBP to methylated DNA recruits enzymes that further compress chromatin and inhibit gene expression. A good example of this repressive cascade is the nucleosome-remodeling deacetylases (NuRD) repressive complex. NuRD is a highly conserved multisubunit complex that synergistically represses chromatin. The NuRD complex includes two HDACs (HDAC1 and HDAC2), a chromatin-remodeling subunit (CHD3 or CHD4), and several MBPs (MBD2 and MBD3). The MBPs recruit the NuRD complex to methylated cytosines where CHD3/CHD4-mediated ATP hydrolysis compacts nucleosomes and stimulates deacetylation by HDACs. This synergism between ATP hydrolysis and histone deacetylation is thought to be due to the ability of chromatin-remodeling enzymes to make histone tails more accessible for deacetylation. The result is a highly repressive chromatin structure around methylated sites (Fig. 1.20–9). DNA methylation can therefore be thought of as a "molecular bridge" between the genetic code (i.e., CpG within specific binding sites) and the other chromatin-modifying machinery (e.g., Figs. 1.20–9, 1.20–13, and 1.20–14). The clinical section of this chapter provides an example of this concept, showing how the genetic code and changes in DNA methylation induced by early life stress interact to modify the expression of the FKBP5 gene and subsequent vulnerability to PTSD.

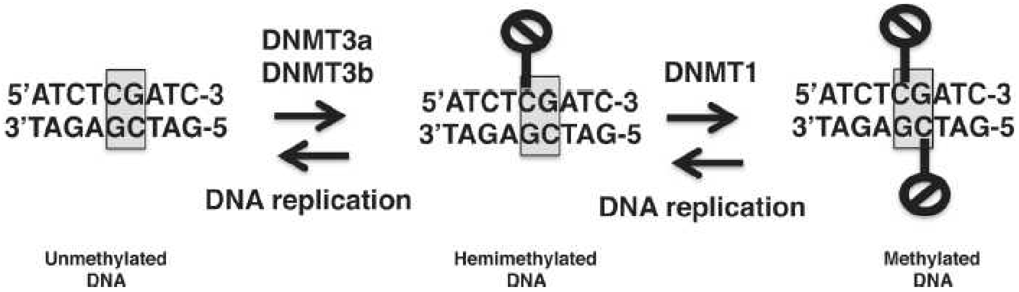

FIGURE 1.20–10. Unmethylated DNA is methylated by the de novo DNMTs, DNMT3a and DNMT3b, to form hemimethylated DNA. Hemimethylated DNA is methylated further by the maintenance DNMT, DNMT1. DNA methylation is lost after each round of DNA replication. The CpG dinucleotide is shown in gray, and the DNA methylation is shown as *black* “no entry” signs.

Two de novo DNMTs enzymes (DNMT3a and DNT3b) are able to methylate CpG sequences on naked or unmethylated DNA, whereas the maintenance DNMT (DNMT1) methylates only hemimethylated sequences (Fig. 1.20–10). De novo DNMTs are responsible for remethylating chromatin after two global demethylation events that occur during normal development. The first global demethylation occurs in the embryonic gametes in order to reprogram these cells to become sperm or an oocyte. The second global demethylation occurs immediately after fertilization and is necessary to establish a pluripotent state in the zygote. Deletion of either DNMT3a or DNMT3b is embryonic lethal, underscoring the importance of de novo methylation for normal development. De novo DNMTs are also expressed in postmitotic neurons, where they methylate promoter elements that regulate the expression of genes implicated in synaptic plasticity, learning, and the response to cocaine.

In contrast to the two de novo DNMTs (e.g., DNMT3a, DNMT3b), DNMT1 does not efficiently methylate naked DNA (i.e., unmethylated DNA). Instead, DNMT1 efficiently methylates DNA in which only one strand is methylated (Fig. 1.20–10). This hemimethylated DNA is commonly formed after DNA replication of a fully methylated CpG site. The presence of DNMT1 at the replication fork allows it to faithfully copy the DNA methylation pattern present on the maternal strand onto the newly formed daughter strand. This mechanism is responsible for the heritability of DNA methylation across mitosis. Deletion of DNMT1 in neural stem cells (NSCs) causes premature induction of gliogenesis (i.e., the production of glial cells) during embryogenesis and multiple defects in neuronal differentiation and synaptogenesis. In addition, exposure of rodents to early life stress increases the expression of DNMT1 in NSCs, which is associated with increased methylation in promoter regions of the retinoic acid receptor. This promoter methylation results in decreased expression of the retinoic acid receptor, which leads to defects in the ability of NCSs to differentiate into neurons. In summary, appropriate DNMT1 activity is required for normal differentiation of NSCs.

The prevailing dogma for many years has been that DNA methylation is a stable epigenetic marker that can be removed only by DNA replication coupled with the inactivation of DNMT1 (see Fig. 1.20–10), a process known as passive demethylation. Therefore, the presence of DNA methylation in postmitotic neurons was thought to be a stable marker that caused long-term suppression of gene expression. This assertion was based on the observations that the carbon-carbon covalent bond formed between the methyl group, and the cytosine ring is highly stable, and no enzymatic activities have been conclusively shown to cleave this bond *in vivo*. However, a growing body of work has consistently shown that demethylation of some CpG sites can occur rapidly in postmitotic neurons. For example, from a total of 219,000 CpG sites examined in postmitotic neurons in the dentate gyrus of the hippocampus, roughly 3000 sites showed a change in DNA methylation four hours after the administration of a single ECS. About two-thirds of these sites showed an increase in DNA methylation, while the remaining third showed a decrease in DNA methylation, indicating rapid and active demethylation. DNA methylation and demethylation modified the expression of genes involved in synaptogenesis, providing a possible mechanism to explain how ECS modifies neuroplasticity in the hippocampus. Interestingly, similar changes in DNA methylation were observed after 3 days of voluntary running, suggesting that ECS and exercise share similar epigenetic mechanisms for modifying synaptogenesis in the hippocampus.

The observation that DNA can be rapidly demethylated in neurons has led to an intense search for the responsible mechanisms. Different models have been proposed, and additional work is needed to further clarify this issue, especially given that several pathways are likely to be involved. Nevertheless, compelling evidence has emerged to support the role of the ten-eleven translocation (TET) family of proteins, especially Tet1, in this process. This family of Fe(II)/2-oxoglutarate-dependent dioxygenases catalyzes the hydroxylation of methylated cytosine (5mC) to 5-hydroxymethyl-cytosine (5hmc). 5mhc is poorly recognized by DNMT1, providing a mechanism by which passive DNA demethylation can occur in mitotic cells (see also above). In neurons, 5mhc is recognized by deaminases, which catalyze the formation of 5-hydroxyuracil (5mhU). This alteration is recognized as a U:G mismatch that is repaired by the base excision repair (BER) machinery (Fig. 1.20–11). The BER complex consists of a DNA glycosylase that hydrolyzes the N-glycosidic bond creating an apyrimidinic/apurinic (AP) site that is removed by an AP endonuclease, and the resulting gap is filled by DNA polymerases. The growth arrest and DNA damage 45 protein family (Gadd45a, Gadd45b, Gadd45c) facilitate this process by recruiting deaminases or glycosylating enzymes to 5mhc sites (Fig. 1.20–11). Interestingly, the C to T transition in a CpG dinucleotide is the most common mutation in human disease, suggesting that this convoluted DNA demethylation/repair mechanism is associated with a significant cost for the organism. This observation further highlights the intimate relationship between the epigenetic and the genetic code.

Chromatin Domains

Chromatin has been traditionally divided into euchromatin (active chromatin) and heterochromatin (repressed chromatin). However, recent advances in chromatin precipitation and high throughput sequencing allow for more precise chromatin mapping. This work has shown that in the mammalian cell, chromatin is organized in topologically associating domains (TAD) that divide each chromosome into roughly 1Mb segments. TADs can be grouped into four major categories of chromatin: active chromatin (equivalent to euchromatin) and three distinct subtypes of heterochromatin known as polycomb-repressed chromatin, null chromatin, and constitutive heterochromatin (see Table 1.20–5). The boundaries between different TADs are fairly well delineated by CGIs of actively transcribed housekeeping genes and binding sites for insulator proteins. However, some flexibility and shift within these boundaries can occur in response to environmental cues. The positions and boundaries of TADs are programmed by the genetic code and are therefore similar in all cells in the body. However, the exact chromatin content within each TAD is different in different cell types (Fig. 1.20–12). These changes in TAD content are induced by histone modifications, chromatin-remodeling enzymes, and DNA methylation allowing cells to alter gene expression in response to different environmental conditions and to use the same genetic code to form different cell types (Fig. 1.20–1). This segmented organization further illustrates how genetic information and the epigenetic machinery interact with each other.

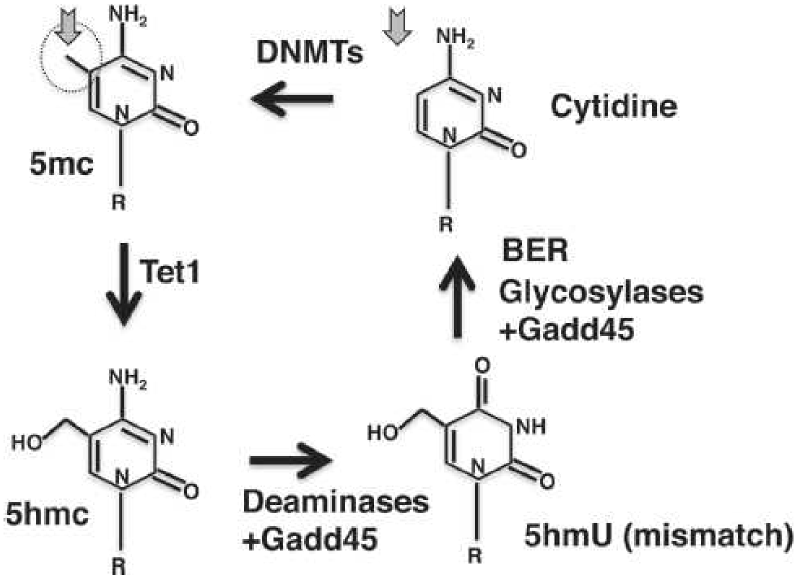

FIGURE 1.20–11. Pathways for DNA demethylation (methylation site is highlighted with an *arrow*).

Active Chromatin. Active chromatin is characterized by the presence of specific histone acetylations (H3K9-ac) and methylations (H3meK4) that are associated with open chromatin and active transcription (see Table 1.20–4). Genes within an active TAD are usually co-expressed and are functionally related. Active TADs are mainly found in the interior of the nucleus and, in some cases, are aggregated together in specific sub-regions of the nucleus to form efficient “assembly lines” (Fig. 1.20–12). Perhaps the most well-characterized example is the nucleolus, where ribosomal RNAs (rRNAs) are synthesized and assembled with ribosomal proteins to form ribozymes. The rRNA transcriptional unit is arranged in a tandem series of about 100 such units clustered on different chromosomes in regions known as nucleolus organizer regions (NORs). Not all NORs are active, and the activation process is at least partly dependent on the ability of the UBF transcription factor to recruit NORs into the nucleolus. This recruitment process helps maintain the rRNA chromatin in an active state, creating a more efficient assembly of ribosomes. This ability of chromatin to loop enhancer elements, some of which are hundreds of kbs away from each other, in and out of contact with each other, and the transcriptional machinery represents an important mechanism by which chromatin regulates gene expression in response to environmental cues (Fig. 1.20–12). The effect of such looping on the risk of developing PTSD is discussed in the clinical section.

Table 1.20–5.
Chromatin Can Be Categorized into Four Different Topologically Associated Domains

TAD Chromatin	Histone Modifications and Associated Proteins	Nuclear/Chromosomal Position	Chromatin Structure and Transcription Activation
Active	H3K9-ac, H3meK4	Interior of the nucleus	Open chromatin, active transcription
Polycomb	H3K27-me3, PRC1, PRC2	Peripheral in differentiated cells	Compact chromatin, transcriptional repression
Null	H3K9me2, nuclear laminin, H1 histones	Nuclear envelope	Compact chromatin, transcriptional repression
Constitutive/Pericentric	H4K20me3, H3K9me2/me3, HP1, H1 histones	Chromocenters, centromeres, telomeres, tandem repeats	Compacted chromatin with intermediate transcriptional repression

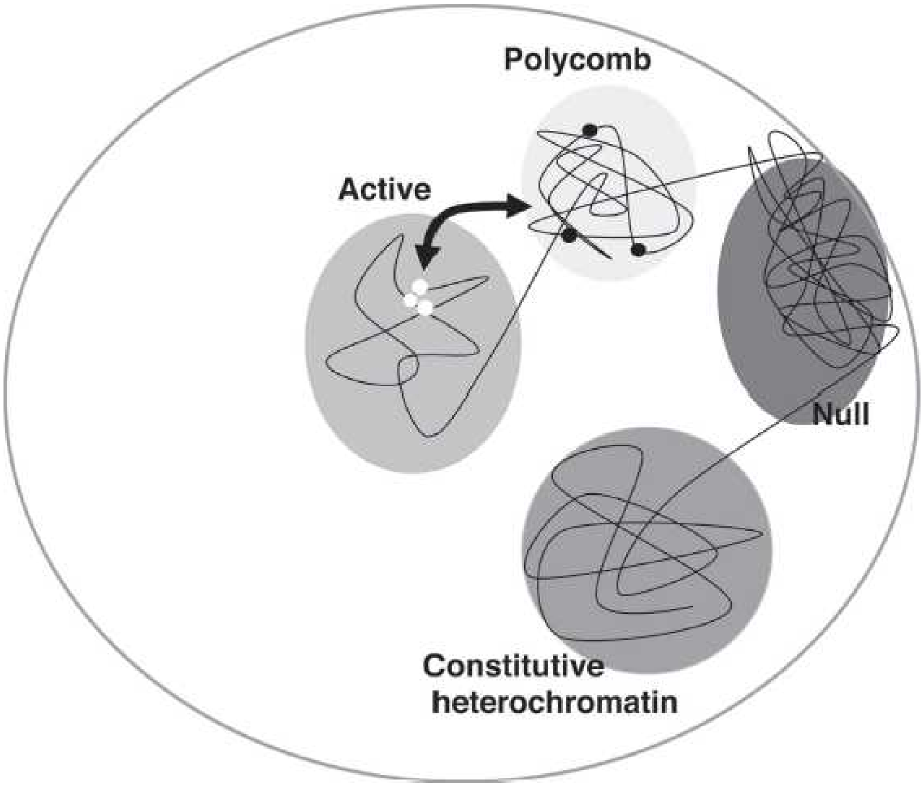

FIGURE 1.20–12. Chromatin in the nucleus is organized into TADs that are separated by house-keeping genes or insulator proteins (not shown). TADs can be categorized as active, polycomb, null, or constitutive/heterochromatin. The chromatin identity of each TAD may be different between different cell types or across the cell cycle within the same cell. Long-distance loops of chromatin bring DNA regions that are far apart in the genetic code into close proximity in the epigenetic code, allowing the formation of assembly lines (*black dots*). These chromatin loops are dynamic and can change in response to environmental cues (*black arrow*).

Polycomb-repressed Chromatin. Polycomb TADs are characterized by polycomb-group response elements (PREs). These sequences are bound by specific transcription factors that recruit the PRC1 and PRC2 complexes to these TADs (Fig. 1.20–7). Polycomb TADs are therefore characterized by the presence of H3K27-me3, polycomb proteins, and repressed chromatin. Polycomb TADs regulate many important cellular processes such as X-inactivation, imprinting, neural stem cell differentiation and cortical development.

Null Chromatin. In mammals, null chromatin is characterized by H3K9me2 histone methylation. This modification is required to tether TAD null chromatin to the nuclear envelope via laminin proteins (Fig. 1.20–12). Null TADs can be significantly larger than other TADs extending up to 10 Mb in humans. They consist of highly repressive, late replicating chromatin that is characterized by high levels of H1 histones and laminin. HDAC1 is required for the general silencing of null chromatin, whereas HDAC3 seems to play an important role in maintaining null chromatin tethered to the nuclear envelope. Mutations in laminin cause the movement of null chromatin from the nuclear envelope to the central part of the nucleus and are associated with chromatin decompaction, gene activation, and severe developmental defects such as Hutchinson–Gilford progeria syndrome.

Constitutive Heterochromatin. Constitutive heterochromatin is compacted during the entire cell cycle and is usually located around centromeres, telomeres, and regions that are enriched with tandem repeats. About 30% of the human chromatin is constitutive chromatin and is characterized by the presence of methylated histones, such as H4K20me3 and H3K9me2/me3, specialized histone variants, and high density of the HP1 protein (see also Fig. 1.20–6). Constitutive heterochromatin provides a more repressive chromatin than active chromatin but is less repressive compared to polycomb or null chromatins (Fig. 1.20–12).

EPIGENETICS IN PRECLINICAL WORK

Several lines of animal research elegantly demonstrate how epigenetic changes cause stable alteration of neuronal circuitry and psychiatrically relevant phenotypes such as stress reactivity, drug-seeking behaviors, and helplessness behavior. Four examples are discussed to illustrate how epigenetic changes modify complex behaviors in animals and the potential therapeutic implications of these findings.

Epigenetic Transmission of Stress Reactivity and Maternal Behavior

The importance of early life experience for the development of the organism has long been appreciated from both psychodynamic and biologic perspectives. Cases of early life abuse and neglect often fit into a transgenerational pattern in which individuals that were maltreatment as children will frequently display maladaptive parenting styles later in life. Exposure to early life stress (ELS) is linked with enhanced endocrine and autonomic responses to stress that persist into adulthood. This enhanced stress reactivity is closely related to an increased risk of developing several psychopathologies, including affective disorders, PTSD, and substance abuse (see Epigenetics Mechanisms in Psychiatric Illness). Animal models taking advantage of the normal heterogeneity in “parenting styles” in rats have allowed investigation of the biologic mechanisms by which parental neglect increases stress reactivity in a manner that persists into adulthood and can be transmitted across generations.

The frequency by which rats lick and groom (LG) their pups during the first week of life differs significantly between individuals, with some mothers LG their pups almost three times compared to others. Despite the fact that differences in the frequency of LG are seen only during the first postpartum week, they have profound effects on many behavioral phenotypes in the offspring (summarized in Table 1.20–6). For example, offspring raised by high LG mothers perform better in several memory tests, are less fearful, and show decreased corticosterone secretion in response to stress compared to “neglected” offspring raised by low LG mothers. Several observations suggest that these behavioral changes are due to the frequency of LG during the first week of life and not to genetic differences between the high and low LG mothers. First, cross-fostering studies show that offspring born to low LG mothers but raised by high LG dams are behaviorally indistinguishable from pups born and raised by high LG mothers. Similarly, offspring born to high LG dams but raised by low LG mothers resemble the biologic offspring of low LG mothers. Second, mimicking maternal behavior by providing rat pups with different amounts of paintbrush strokes during the first weeks of life can induce similar behavioral changes. These distinct phenotypes persist into adulthood, indicating that exposure to different levels of tactile stimulation during the first week of life in some way programs adult behavior and physiology. Importantly, female offspring raised by high LG dams become high LG mothers themselves, while those born to low LG mothers become low LG mothers (Table 1.20–6). As a result, the low and high LG phenotypes are propagated transgenerationally by maintaining specific levels of tactile stimulation within a lineage early in life.

Table 1.20–6.
Differences between Offspring Exposed to High and Low LG Mothers

Behavioral Phenotype in Adulthood	High LG Offspring	Low LG Offspring
Hypothalamic pituitary axis (HPA) reactivity (GR resistance)	Low	High
Anxiety	High	Low
Hippocampal dependent learning	Intact	Impaired
Maternal behavior (LG)	High	Low

These observational studies beg the question of how seven days of early life experience determine stress reactivity and maternal behavior in adulthood. The emerging answer is that the frequency of LG during the first week of life causes stable changes in DNA methylation in promoter elements that regulate the expression of genes in neuronal circuits that control stress- reactivity and maternal behavior. For example, the promoter of the glucocorticoid receptor (GR) is heavily methylated at birth, leading to low GR levels in the hippocampus of newborn pups. Exposure to high LG during the first week of life induces several transient changes in the hippocampus that ultimately lead to extensive demethylation of the GR promoter. Demethylation of the GR promoter allows the transcription activator nerve growth factor inducible protein A (NGFI-A) to bind to the GR promoter. The binding of NGFI-A recruits the HAT, and CBP, to acetylate nearby histones, which further opens chromatin and enhances GR expression in the hippocampus of high LG offspring (Fig. 1.20–13). High levels of GR persist in the hippocampus of adult high LG offspring and exert an inhibitory influence on neurons in the paraventricular nucleus of the hypothalamus (PVN) that secrete corticotropin-releasing hormone (CRH) in response to stress. The promoter of the GR in offspring of low LG mothers is kept hypermethylated. This hypermethylation prevents NGFI-A from binding to the GR promoter, leading to reduced GR levels in the hippocampus of low LG offspring (Fig. 1.20–13). In sum, exposure to high LG during the first week of life decreases stress reactivity in adulthood by demethylating the GR promoter during a critical period of development.

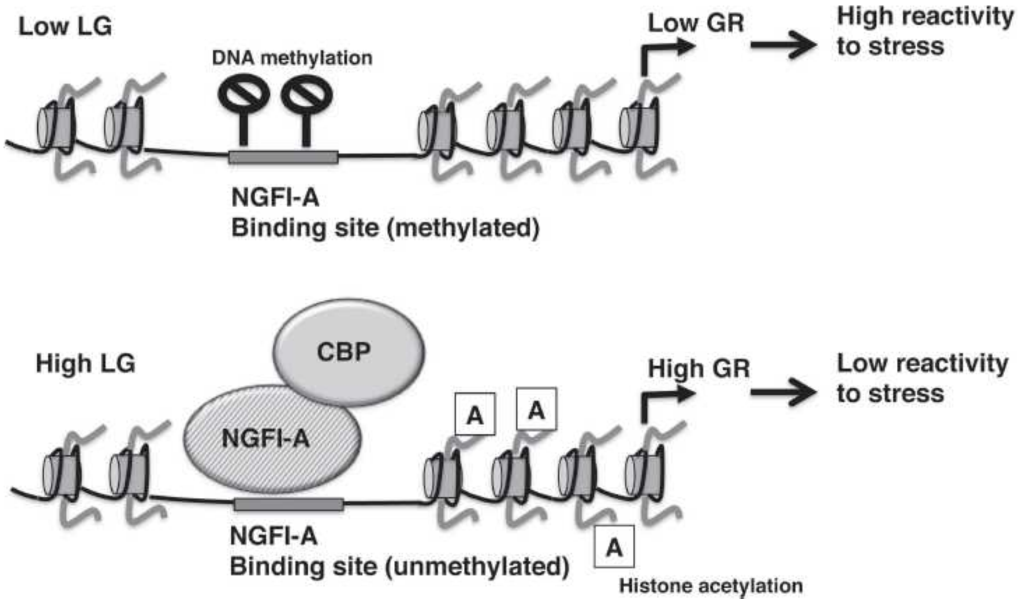

FIGURE 1.20–13. **Top:** The promoter of the GR is maintained in a methylated state in the offspring of low LG mothers, leading to low levels of GR expression in the hippocampus and high-stress reactivity. **Bottom:** High LG during the first week of life leads to demethylation of the GR promoter, allowing the NGFI-A transcription factor to bind and activate GR expression by recruiting the HAT, CBP. Elevated levels of GR in the hippocampus decrease stress reactivity in the offspring of high LG mothers.

Exposure to high LG behavior also decreases DNA methylation at the estrogen receptor alpha (Erα) promoter in the medial preoptic area (MPOA) of the hypothalamus. Demethylation of this promoter recruits the transcription factor Stat5 to the Erα promoter allowing for increased Erα levels in the MPOA of high LG female offspring. Offspring raised by low LG mothers maintain low levels of Erα in the MPOA due to the persistence of DNA methylation at the Erα promoter. Differences in Erα levels in the MPOA persist into adulthood, regulate oxytocin expression, and are thought to program subsequent LG behavior in adult female offspring.

An important clinical question regarding the capacity of LG to cause stable alterations in DNA methylation, gene expression, and behavioral phenotype is whether these changes can be reversed pharmacologically in adulthood. The answer to this question seems to be yes. For example, intraventricular infusion of an HDAC inhibitor decreases DNA methylation at the GR promoter, increases GR levels in the hippocampus, and decreases stress reactivity in adult low LG offspring. While HDAC inhibitors do not change DNA methylation directly, they increase chromatin accessibility, a process that may facilitate the demethylation of DNA. Conversely, intraventricular infusion of methionine (a donor for methylation reactions) increases DNA methylation at the GR promoter, reduces hippocampal GR expression, and increases stress reactivity in adult high LG offspring. These findings suggest that the epigenetic landscape at the GR promoter can be modified pharmacologically in adulthood by tilting the balance between activities that promote open chromatin (histone acetylation, DNA demethylation) and activities that compact chromatin (methylating agents). It is currently unclear whether the developmental phenotypes induced by LG at the GR promoter will be reestablished after a prolonged washout of the HDAC inhibitors or methionine. Addressing this question will have important implications regarding the true reversibility of these developmentally sensitive epigenetic changes.

The relevance of these findings to human psychopathology has been examined in postmortem studies by assessing the brains of suicide victims with documented histories of childhood abuse or neglect compared to suicide victims with no childhood trauma. Similar to the preclinical work in rats, childhood maltreatment was associated with increased DNA methylation at the GR promoter and reduced GR levels in the hippocampus. While the challenges of assessing epigenetic signatures in living patients remain (see Epigenetics Mechanisms of Psychiatric Illness section below), these studies suggest that preclinical work in rodents can provide important insights into the mechanisms by which epigenetics contribute to psychiatric conditions in humans.

Rescuing Prader–Willi Syndrome in the Mouse

Prader–Willi syndrome (PWS) is an imprinting disorder affecting roughly 1/15,000 to 30,000 individuals. The core features of this variable syndrome include neonatal hypotonia and a failure to thrive, followed by hyperphagia, childhood obesity, hypogonadism and mild intellectual disability (Table 1.20–3). PWS is caused by a lack of expression of a group of genes located on the paternal 15 chromosome (e.g., 15q11-q13). This region contains several protein-coding genes such as Necdin (NDN), Small Nuclear Ribonucleoprotein Polypeptide N (SNPRN), Ubiquitin Protein Ligase E3A (UBE3A) and noncoding RNAs such as SNORD116 and UBE3A-ATS whose expression is regulated by the PWS imprinting center (*PWS-IC*). In males, the DNA at the PWS-IC is unmethylated and is maintained in an open chromatin state, allowing for the expression of PWS genes (Fig. 1.20–14, top). The expression of UBE3A-ATS in males causes local suppression of its neighboring UBE3A gene (Fig. 1.20–14-, top). On the maternal chromosome, the PWS-IC DNA is highly methylated, leading to the recruitment of the histone methylase G9A. G9A methylates histone 3 on lysine 9 (H3K9-Me2 and H3K9-Me3), resulting in the expansion of compacted chromatin over the entire PWS locus and the blockade of gene expression from the maternal chromosome (Fig. 1.20–2, middle). The absence of UBE3A-ATS expression on the maternal chromosome allows for the transcription of the UBE3A gene. This is important because conditions that impair the expression of functional UBE3A from the maternal chromosome cause Angelman syndrome (Table 1.20–3).

Deletion of the paternal 15q11-q13 locus eliminates the only genomic source for these genes and is the most common cause of the PWS. The exact contribution of each of these genes to the PWS is yet to be clarified, but a growing body of work has implicated the noncoding, highly conserved RNA SNORD116 as an important contributor to the pathology. The presence of an intact yet highly compacted chromatin over the maternal chromosome raised the possibility that “forcing” the maternal chromatin to open may eliminate or at least mitigate the clinical features of PWS. To accomplish this, embryonic fibroblasts were harvested from transgenic mice that express the green fluorescence protein (GFP) from the maternal PWS-IC. Since the maternal PWS-IC is inactive (Fig. 1.20–14, middle), no GFP expression is seen in this cell line. However, when exposed to a large library of over 9000 small compounds, two compounds (UNC0638 and UNC0642) were able to induce the expression of GFP in this cell line. Both compounds were known inhibitors of the HMT G9A, suggesting that blocking histone methylation might be the key to opening the compacted chromatin structure on the maternal chromosome. Further analysis confirmed that UNC0638 and UNC0642 opened up the chromatin structure at the maternal locus and restored expression of PWS genes in mouse embryonic fibroblasts, with similar results seen when fibroblasts obtained from PWS patients were exposed to UNC0638 and UNC0642. What is particularly promising about this approach is that, unlike agents that block DNA methylation (such as 5-Aza-dc), UNC0638 and UNC0642 increased the expression of most of the genes found in the PWS locus, including SNORD116. Further, treatment of a mouse model of PWS showed no evidence of toxicity and led to a significant reduction in mortality caused by a failure to thrive in this model. In vivo and in vitro treatment with UNC0638 and UNC0642 did not activate the expression of more downstream genes such as UBE3A-ATS which would block the expression of UBE3A and lead to Angelman syndrome (Fig. 1.20–14, bottom). UNC0638 and UNC0642 blocked histone methylation but did not reduce DNA methylation or interfere with G9A binding at the maternal PWS-IC region, suggesting that blocking histone methylation is able to restore some gene expression and reduce lethality even in the presence of high levels of DNA methylation. Expression of genes from the maternal chromosome reached only 30-50% of levels expressed from the parental locus, indicating that DNA methylation and histone methylation are both necessary for full repression and that reversing one layer of regulation within this system is likely to provide only a partial rescue of the PWS. Nevertheless, the use of a reporter-mediated cell line coupled with a high throughput screen of small molecules to induce localized changes in chromatin structure provides an exciting new frontier in epigenetic drug discovery.

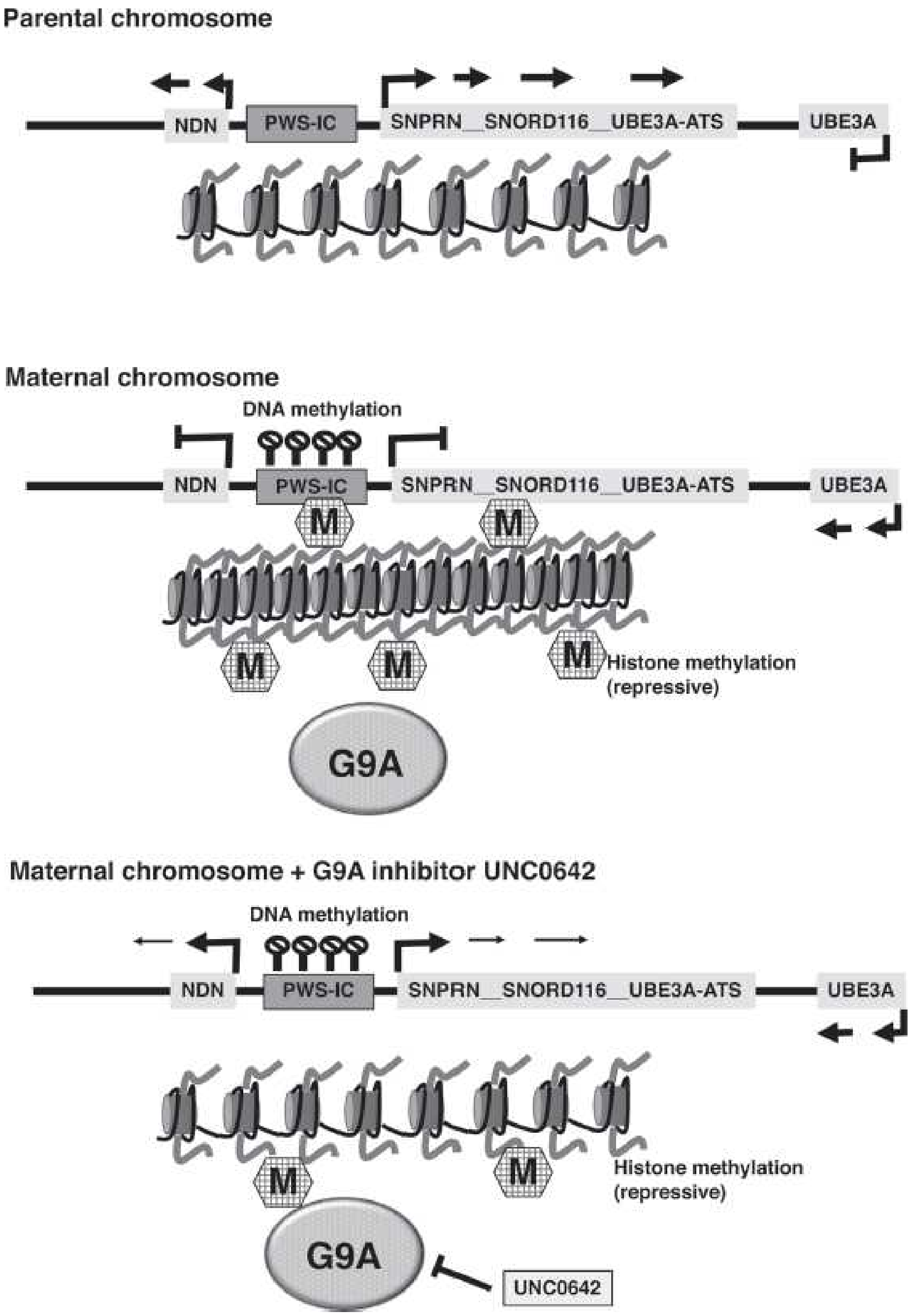
Parental chromosome
NDN
PWS-IC
SNPRN__SNORD116__UBE3A-ATS
UBE3A
Maternal chromosome
DNA methylation
NDN
PWS-IC
SNPRN__SNORD116__UBE3A-ATS
UBE3A
M
M
M
M
M
Histone methylation (repressive)
G9A
Maternal chromosome + G9A inhibitor UNC0642
DNA methylation
NDN
PWS-IC
SNPRN__SNORD116__UBE3A-ATS
UBE3A
M
M
Histone methylation (repressive)
G9A
UNC0642

FIGURE 1.20–14. Top: The paternal DNA at the Prader Willis imprinting center (PWS-IC, shown in *dark gray*) is unmethylated, allowing the formation of open chromatin and the expression of genes such as NDN, SNPRN, SNORD116 that are necessary for normal neurodevelopment. Expression of UBE3A-ATS prevents UBE3A transcription from the parental chromosome. **Middle:** The maternal DNA at the PWS-IC is methylated, recruiting the HMT G9A and leading to histone methylation and chromatin compaction across the entire PWS locus. Compacted chromatin prevents gene expression, including UBE3-ATS, while restoring transcription of UBE3A from the maternal chromosome. **Bottom:** UNC0642 inhibits G9A HMT activity allowing for partial expression of genes such as NDN, SNPRN, SNORD116 (but not UBE3-ATS) from the maternal PWS locus and reducing pup mortality in a mouse model of PWS.

The Epigenetic Mechanisms of Depressive Behavior and Antidepressant Action

Rodent models of depression and PTSD typically employ some form of chronic, unpredictable, or severe stress that leads to despair, avoidance, or anhedonic-like behaviors. One such paradigm involves chronic social defeat in which a mouse is briefly exposed to an aggressive intruder over the course of 10 days. Defeated mice show a decrease in sucrose consumption, suggesting an anhedonic-like state and increased helplessness in the forced-swim and the tail-suspension tests. Like human depression, many of these behavioral changes are reversed by chronic but not acute antidepressant treatment.

Several observations suggest that the behavioral consequences of social defeat and the rescue of depressive phenotypes by antidepressant medications are due to epigenetic changes in the expression of genes implicated in neuroplasticity. One of the best examples of epigenetic regulation of neuroplasticity genes is the opposing effects of stress and antidepressants on the expression of brain-derived neurotrophic factor (BDNF) in limbic brain areas. BDNF expression in the hippocampus is downregulated in defeated mice in a manner that persists for many weeks after terminating the stress, an effect that can be reversed with chronic treatment with the tricyclic antidepressant, imipramine. The ability of social defeat to decrease hippocampal BDNF levels appears to be mediated by increasing H3K27 methylation at the BDNF promoter. These histone modifications lead to a repressive chromatin state that inhibits BDNF expression (see Fig. 1.20–15). These findings are consistent with human postmortem studies showing that depressed individuals have increased H3K27 dimethylation (H3K27-me2) at the BDNF promoter and reduced BDNF levels. Treatment with imipramine rescued the depressive phenotype and increased BDNF levels in the hippocampus of defeated mice but did not alter H3K27-me2 levels. Instead, imipramine treatment increased acetylation at H3K9 residues at the BDNF promoter, a modification that is associated with open chromatin and transcriptional activation (Fig. 1.20–15). These results underscore the combinatorial nature of the histone code, in which acetylation can override the repressive nature of H3K27-me2. Treatment with imipramine decreased HDAC5 expression in defeated but not control mice, raising the possibility that reduced HDAC5 in defeated mice might promote H3K9 acetylation and therefore mediate the therapeutic effects of imipramine (Fig. 1.20–15). This assertion is supported by data showing that overexpression of HDAC5 in the hippocampus blocks imipramine's antidepressant effects in defeated animals. These preclinical findings have suggested that treatment with HDAC inhibitors may provide a novel therapeutic strategy to increase BDNF and treat depression. However, despite significant effort, no such treatment is currently available, highlighting the complexity and the challenges of using epigenetic-modifying agents to treat psychiatric conditions (see more on this issue below)

Epigenetic Mechanisms of Cocaine Addiction

While the prior sections have addressed how negative life experiences (fear conditioning, chronic social defeat, or maternal neglect) cause long-lasting changes in chromatin structure, intensely rewarding experiences can also be coded epigenetically. A classic example is drug addiction, in which the natural reward system is hijacked by highly rewarding exogenous substances, resulting in maladaptive, compulsive drug-seeking behavior.

The transition from the initial experimentation phase of drug consumption, driven by the rewarding experience, to the addictive phase, characterized by compulsive maladaptive behavior, is a gradual process that is seen across diverse species, from humans to rodents. In rodents, the transition to addictive behavior is mediated by epigenetic changes in the expression of genes that modify synaptic plasticity in neurons located within the reward system. This section provides a brief summary of some of the main lessons learned about this process by focusing on the role that histone acetylation in the nucleus accumbens (NAc) plays in the development of drug-seeking behaviors.

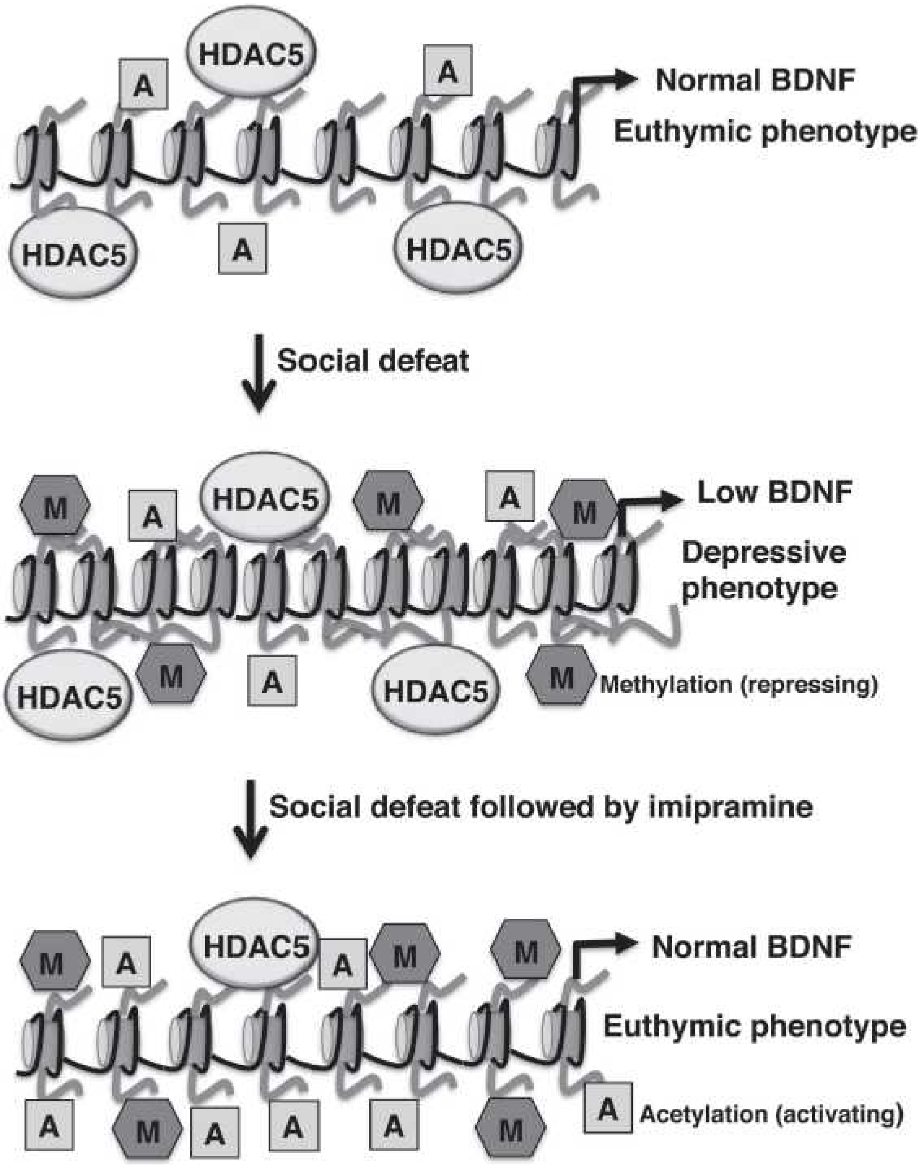
HDAC5
A
Normal BDNF
Euthymic phenotype
HDAC5
A
HDAC5
Social defeat
M
A
HDAC5
M
A
M
Low BDNF
Depressive phenotype
HDAC5
M
A
HDAC5
M
Methylation (repressing)
Social defeat followed by imipramine
M
A
HDAC5
A
M
M
Normal BDNF
Euthymic phenotype
A
M
A
A
A
M
A
Acetylation (activating)

FIGURE 1.20–15. Top: Control mice show normal BDNF levels and euthymic-like phenotype. Middle. Exposure to repeated social defeat leads to an increase in repressive HK27-me2 methylation (windowpane-patterned hexagons), leading to condensed chromatin, decreased BDNF levels, and a depressive-like phenotype. **Bottom:** Treatment of socially defeated mice with imipramine does not alter HK27-me2 levels but instead decreases HDAC5 levels. Reduced HDAC5 levels contribute to the increase in histone acetylation (*gray squares*), chromatin decondensation, increased BDNF levels, and restoration of euthymic phenotype.

The immediate rewarding effects of cocaine are closely associated with an increase in dopamine levels in several brain circuits, with the Nac being one of the most important and well-studied targets. The initial rewarding effects of cocaine are experienced within minutes and lead to the transient expression of immediate early genes such as c-fos. Chronic, but not acute, administration of cocaine is necessary to increase the expression of an additional group of late-onset genes such as FosB/

ΔfosB, CDK5, and BDNF. The expression of many of these late-onset genes can persist for many weeks even in the absence of cocaine and is necessary to elicit robust drug-seeking behaviors in rodents. This effect has been extensively studied for the highly stable transcription factor ΔfosB. Repeated activation of c-fos leads to the accumulation of ΔfosB in a subset of Nac neurons. High levels of ΔfosB are necessary to express synaptic proteins that enhance synaptic plasticity in Nac neurons, which results in the stable changes that are necessary for drug-seeking behavior. Accumulation of ΔfosB is also seen in other compulsively consumptive behaviors such as exercise and feeding, consistent with the notion that cocaine is able to hijack endogenous reward pathways.

Histone acetylation plays a critical role in coordinating the transition from inducing immediate early genes to the stable expression of late-onset genes such as ΔfosB. Acute cocaine administration induces c-fos expression by increasing H4 acetylation and H3 phosphorylation. With repeated cocaine use, these chromatin changes are no longer present at the c-fos promoter leading to a decline in c-fos levels. Instead, repeated cocaine administration leads to an increase in H3 acetylation at promoters of late-onset genes (e.g., FosB, ΔfosB, Cdk5, and BDNF). Manipulations that inhibit histone acetylation, such as viral knockdown of the HAT CBP or the overexpression of different HDACs in the Nac, block the ability of chronic cocaine to induce late-onset genes and reduce compulsive drug-seeking behaviors. These findings suggest a model in which the transition to drug addiction is mediated by an epigenetic switch from an initial rapid H4 acetylation to a slower wave of H3 acetylation that is necessary to induce late-onset genes.

It is important to note, however, that this is a highly simplified model and that the epigenetic effects of chronic cocaine are far more complex. For example, prolonged inhibition of HDAC1 in the Nac reduces cocaine locomotor sensitization, which is inconsistent with the "expected" increase in histone acetylation and the promotion of drug-seeking behavior. This unexpected result is likely due to the ability of chronic HDAC1 inhibition to maintain the repressive H3K9-me2 marker at the promoters of GABA receptor genes even after chronic cocaine exposure. Thus, the retention of the H3K9-me2 repressive marker prevents the expression of these genes even in the presence of high levels of acetylation. This example underscores the complexity by which different histone modifications interact with one another to modify chromatin structure and gene expression (see also Fig. 1.20–15, above).

Chronic cocaine also down-regulates the expression of DNMT3a in the Nac, and manipulations that decrease DNMT3a enhance addictive phenotypes. These findings suggest that some of the behavioral consequences of cocaine addiction are mediated by the demethylation of DNA in the Nac. Finally, some of the epigenetic changes seen after cocaine administration appear to be cell-type specific. For example, knockout of the HMT (G9A) in D1 neurons decreases the behavioral consequences of cocaine, but the opposite phenotype is seen when the same enzyme is knocked out in D2 neurons. This cell-type specificity is not surprising, given that different cell types possess very different chromatin structures and functions (see the introduction and Fig. 1.20–1). Nevertheless, this cell-type specificity represents a challenge for pharmacologic interventions that target specific chromatin activities (e.g., the HMT G9A) and for interpreting chromatin changes seen in heterogeneous brain regions such as the Nac.

EPIGENETIC MECHANISMS IN PSYCHIATRIC ILLNESS

While epigenetics in psychiatric research holds significant promise for addressing several fundamental unanswered questions in psychiatric research (Table 1.20–2), significant challenges in its clinical implementation have also emerged in recent years. This final section provides three clinical examples that illustrate the promise and challenges that epigenetic research holds for revolutionizing psychiatric care.

The Epigenetic Etiology of Fragile X Syndrome

Fragile X syndrome (FXS) is the leading cause of inherited intellectual disability in the world, with an estimated prevalence of 1:2,500 males. It is an X-linked monogenetic neurodevelopmental psychiatric disorder characterized by moderate intellectual disability, social deficits, a high incidence of ADHD, and multiple physical abnormalities (Table 1.20–3). Fragile X syndrome is caused by the loss of function of the fragile X mental retardation gene 1 (FRM1). The FMR1 gene codes for an RNA-binding protein that regulates levels of synaptic and chromatin-modifying genes that are critical for normal synaptic development and cognition. While the term “monogenetic” implies mutated genetic code, the actual disease-causing mechanism involves epigenetic changes that silence the promoter of this gene.

The number of CGG repeats in FXS can be categorized into three groups: normal, premutated, and fully mutated alleles. The normal allele consists of 5-55 CGG repeats allowing for normal expression of the gene. The premutated allele has 56-200 repeats and maintains normal expression of the FRM1 protein. However, the premutated allele is associated with increased risk for further expansion, pushing it into the fully mutated range, defined as greater than 200 repeats. Since each CGG repeat contains a CpG dinucleotide, this expanded repeat is recognized as an unstable genomic region that becomes heavily methylated. As in the case of other heavily methylated CpG islands (see The Epigenetic Machinery, DNA Methylation section), this methylated region recruits additional remodeling enzymes that form a highly repressive chromatin structure that prevents the expression of the FRM1 gene (Fig. 1.20–16).

In a subset of rare cases, individuals with fully mutated alleles (i.e., >200 repeats) do not develop clinical symptoms of FXS. In these rare cases, the expanded repeats somehow escape DNA methylation providing compelling evidence that DNA methylation is necessary to induce the clinical syndrome in individuals with the full mutation. Therefore, therapies that prevent DNA methylation could potentially rescue the neurodevelopmental sequelae of individuals with the fully mutated allele. Indeed, treatment with DNA demethylating agents such as 5-azacytidine of cell lines and iPSCs obtained from individuals with FXS were able to partially demethylate and express FMR1protein. However, the clinical utility of these compounds is limited due to their high degree of cytotoxicity and risk for mutagenesis, and there is a need to develop nontoxic, stable, and highly specific epigenetic interventions. Indeed, a recent proof of concept study has successfully used crisper technology to meticulously excise the CGG repeat from iPSCs derived from an FXS patient leading to full activation of the FMR1 gene and restoring normal electrophysiologic properties in neurons generated from these iPSCs. Although far in the future in terms of clinical utility, this work demonstrates the utility of iPSCs in testing outcomes of novel epigenetic interventions in humans.

Gene-by-Environment Interaction Modulates FKBP5 Expression and the Risk for PTSD

In many cases, the effects of stress on complex behaviors are mediated by genetic vulnerability. Several examples of this type of gene by environment interaction have been described in the psychiatric literature. However, the molecular details by which these interactions take place have only just started to emerge. This was recently illustrated in a series of studies showing how childhood trauma preferentially interacts with a specific genetic predisposition to increase the expression of the FK506-binding protein 5 (FKBP5) and the vulnerability to PTSD and depression.

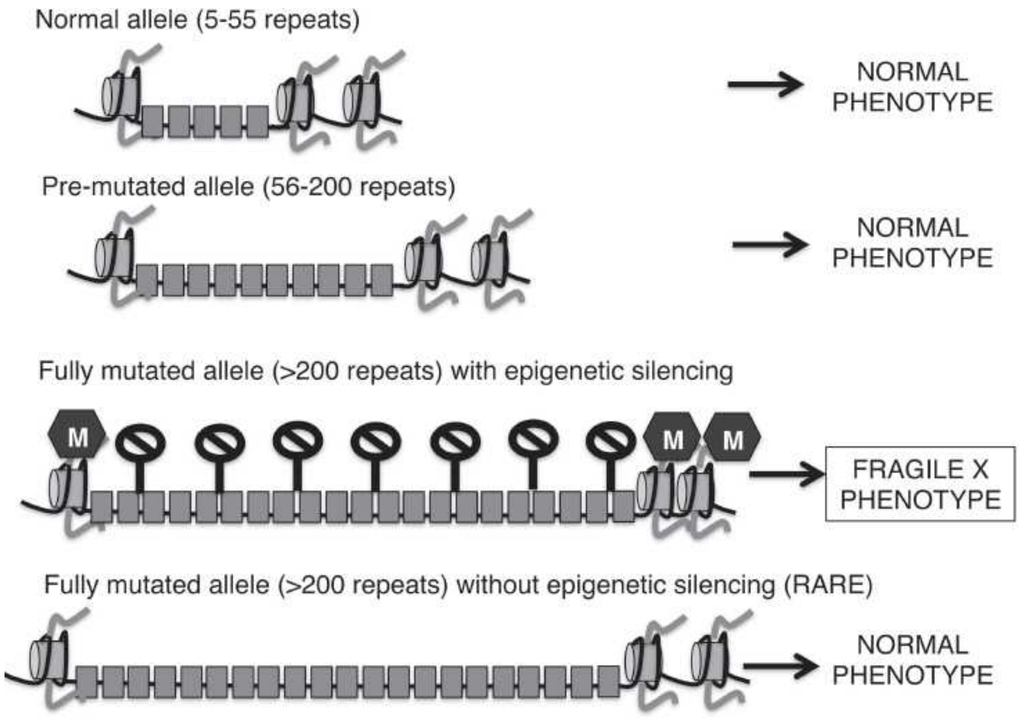
Normal allele (5-55 repeats)
NORMAL PHENOTYPE
Pre-mutated allele (56-200 repeats)
NORMAL PHENOTYPE
Fully mutated allele (>200 repeats) with epigenetic silencing
M
M
M
FRAGILE X PHENOTYPE
Fully mutated allele (>200 repeats) without epigenetic silencing (RARE)
NORMAL PHENOTYPE

FIGURE 1.20–16. In normal individuals, the FRM1 gene contains 5–55 CGG repeats (each gray box represents 10 CGG repeats). Individuals with the premutation have 56–200 repeats and are phenotypically normal. However, they are at increased risk for repeat expansion during meiosis resulting in a fully mutated phenotype in their offspring. The DNA promoter of fully mutated individuals (>200 CGG repeats) becomes highly methylated (black no entry signs), leading to abnormally reduced FMR1 expression and the Fragile-X phenotype. Rarely individuals have been identified who possess the full mutation yet lack epigenetic silencing. These children are phenotypically normal, suggesting that DNA methylation is necessary to induce the clinical syndrome.

FKBP5 is a stress-related gene that exerts its effects via interaction with the GR. Under basal conditions, the GR is maintained in the cytoplasm by the heat shock 90 (Hsp90) and FKBP5 proteins, which bind GR and mask its nuclear localization signal. High levels of cortisol induced by stress dissociate Hsp90 and FKBP5 from GR, allowing it to enter the nucleus and induce two transcriptional programs. The first transcriptional program is essential for terminating further release of cortisol from the adrenal gland (see Epigenetic Transmission of Stress Reactivity and Maternal Behavior section). The second task involves the induction of FKBP5, a negative feedback loop that prevents further entry of GR into the nucleus. Retaining GR in the cytoplasm is thought to reset the stress response so it can be reactivated in response to additional threats (Fig. 1.20–17).

Conditions in which FKBP5 is overexpressed prevent normal activation of the GR, leading to inefficient termination of the stress response. This inability to terminate HPA activity leads to prolonged exposure to high levels of cortisol and is referred to as “*GR resistance*.” GR resistance is associated with an increased risk for depression, PTSD and several other medical complications (Fig. 1.20–17). This disease model is similar to that described for insulin resistance, in which ineffective insulin signaling causes end-organ toxicity by maintaining high levels of glucose in the circulation.

The rs1360780 haplotype of the FKBP5 gene has an A/T and a C/G allele. At baseline, the A/T allele is associated with higher levels of GR-induced FKBP5 expression, leading to more robust HPA activation compared to the protective C/G allele (Fig. 1.20–18). While the A/T allele alone is not associated with increased risk for stress-related psychiatric disorders, the risk allele interacts with early life trauma in a dose-dependent manner to increase the risk for PTSD. In other words, individuals with the A/T allele that were exposed to more severe childhood maltreatment are more likely to develop PTSD and depression. This is not the case for individuals with the protective C/G allele in which exposure to childhood trauma is not associated with an increased risk for PTSD or depression.

The mechanistic explanation for these clinical observations points to a synergy between genetic and epigenetic regulation of FKBP5 levels and GR resistance. At the genetic level, A/T susceptible individuals show a more efficient GR-mediated FKBP5, producing higher levels of FKBP5 than C/G individuals. The FKBP5 gene has two glucocorticoid response elements (GREs) that are occupied by GR in response to stress. One GRE is found in intron 2 (GRE-2) is roughly 50 kb from the transcriptional start site (TSS), while a second GRE is found in intron 7 (GRE-7), about 100 kb from the TSS of FKBP5. Despite these long distances, both GREs are able to contact the TSS via long-distance loops. These chromatin-mediated loops recruit these GREs to the vicinity of the TSS, driving FKBP5 expression. When GR is activated, the GRE-7 site is recruited to the TSS regardless of the genotype. In contrast, GRE-2 is recruited to the TSS only in AT individuals. The presence of GRE-2 in the vicinity of the TSS results in a more robust expression of FKBP5 in response to stress and hence greater HPA activation in individuals with the A/T allele (Fig. 1.20–18). It is currently unclear how these genetic changes contribute to these differential looping configurations, but the presence of these loops is an important function of chromatin (see Active Chromatin section).

As a result of these genetic differences, A/T individuals exposed to childhood abuse are predicted to secrete higher levels of cortisol compared to C/G individuals. Greater cortisol levels are thought to decrease DNA methylation at the GRE-7–binding site in individuals with the risk allele. Decreased DNA methylation at GRE-7 leads to more efficient recruitment of GR, further augmenting FKBP5 expression and GR resistance in A/T individuals that were exposed to childhood trauma (Fig. 1.20–18). GRE-7 continues to be methylated in G/C individuals that were abused or neglected, demonstrating specific gene-by-environment interactions for this epigenetic marker. *Ex vivo* experiments with patient-derived monocytes show a strong negative correlation between GRE-7 methylation and glucocorticoid potency, supporting a functional role for this epigenetic biomarker in mediating GR resistance.

The ability of trauma to induce these epigenetic changes appears to be confined to a sensitive period, as exposure to trauma in adulthood does not lead to a decrease in the GRE-7 methylation rate. These clinical findings are supported by cell culture experiments in which dexamethasone decreases GRE-7 methylation in proliferating

NSCs, but not in fully differentiated postmitotic neurons. When exposed during proliferation, this decrease in methylation is maintained after the NSCs differentiate into neurons, even once the dexamethasone stimulus is removed, suggesting stable propagation of demethylated GRE-7 into differentiated neurons. These observations suggest that exposure to stress during childhood, when high levels of neurogenesis are present, is responsible for this gene by early life stress interaction. Therefore, interventions that are aimed at decreasing stress early in life are likely to be more effective in reducing vulnerability to PTSD compared to interventions that eliminate stress in adulthood.

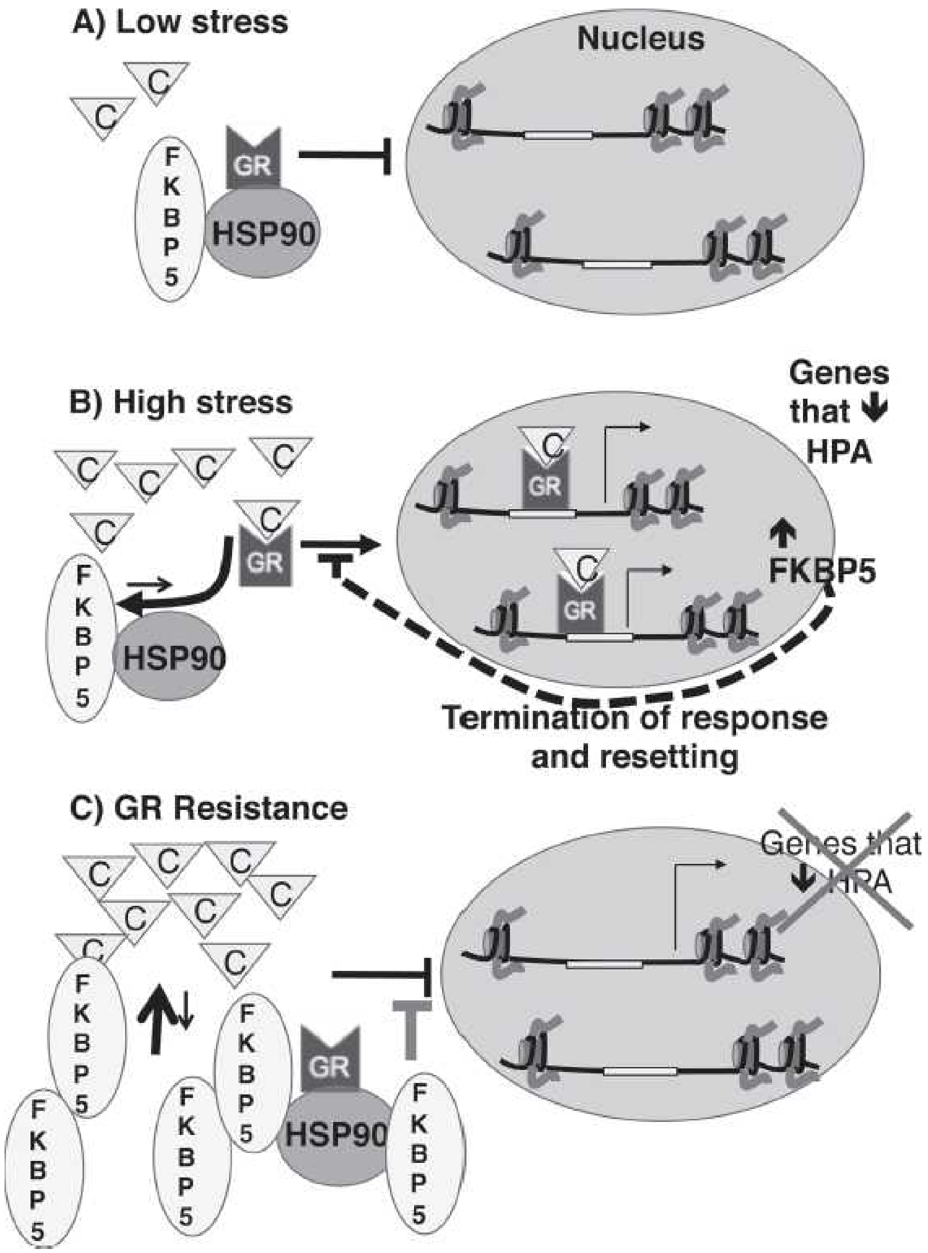
A) Low stress
Nucleus
C
C
F K B P 5
GR
HSP90
B) High stress
Genes that ↓ HPA
C
GR
F K B P 5
HSP90
↑ FKBP5
Termination of response and resetting
C) GR Resistance
Genes that ↓ HPA
C
F K B P 5
GR
HSP90

FIGURE 1.20–17. **A:** In the absence of stress, when cortisol levels ("C" triangles) are low, the glucocorticoid receptor (GR) is bound to the chaperone proteins FKBP5 and HSP90. In this bound state, the GR is prevented from entering the nucleus and activating transcription. **B:** Exposure to stress increases cortisol levels allowing GR to enter the nucleus by displacing the FKBP5/HSP90 complex. Inside the nucleus, GR activates transcription of genes that terminate further cortisol release and activates transcription of FKBP5. Increased FKBP5 levels prevent further GR entry into the nucleus, resetting the system so it can respond to subsequent waves of stress. **C:** Overexpression of FKBP5 leads to GR Resistance by preventing high levels of cortisol from driving GR into the nucleus. As a result, additional levels of cortisol are being released, leading to high levels of circulating cortisol.

Epigenetic Suicide Signatures: Advance Toward a Biomarker

Suicide is the 11th most common cause of death in the United States. The importance of accurate risk assessment is highlighted by the observation that roughly half of the individuals who commit suicide are seen by a mental health worker prior to their death. In the absence of reliable biomarkers to predict suicide, psychiatric evaluation relies on a long list of risk factors to assess the likelihood of lethal self-harm. These population-risk factors have poor predictive value, as most people who have these risks do not commit suicide, and some individuals who ultimately take their own life may not share these risks. Therefore, the recent findings that certain peripheral epigenetic markers can predict suicide with more than 80% accuracy provide a possible breakthrough for preventing these tragedies.

To search for epigenetic markers of suicide, DNA methylation patterns from the prefrontal cortex of depressed individuals who committed suicide were compared to depressed individuals who died of other causes. DNA methylation is amenable to study in postmortem tissue because of its relative stability and its ability to predict gene expression (see DNA methylation). Only one gene, coding for the spindle and kinetochore associated complex subunit 2 (SKA2), was found to be differentially methylated in neurons of suicide completers compared to the control group. SKA2 is a scaffolding protein that chaperones the GR into the nucleus and therefore helps terminate HPA activation in response to stress. Increased DNA methylation at the 3′ untranslated region (3′-UTR) of SKA2 was replicated in a separate postmortem cohort, suggesting that it represents a reliable retrospective biomarker of suicide. Methylation of SKA2 was associated with a decrease in SKA2 protein levels, suggesting that this epigenetic modification regulates SKA2 activity in suicide victims. Interestingly, the CpG methylation site at the 3′-UTR of SKA2 contains an alternative single-nucleotide polymorphism allele in which the cytidine is replaced with thymidine. This T allele cannot be methylated and is therefore considered to be the protective allele.

While postmortem replication is an important first step, the more important question is whether the presence of this epigenetic biomarker in peripheral tissue can predict future suicidal behavior in living patients. This was assessed by first comparing gene expression in prefrontal cortex neurons and glial cells obtained from the postmortem analysis and peripheral immune cells obtained from a different cohort of individuals with depression. This analysis showed a high correlation between gene expression in neurons, but not glia, and peripheral immune cells, suggesting that the two cell populations share similar epigenetic signatures, which is a critical feature in establishing a peripheral biomarker. Next, the relationship between SKA2 methylation state in peripheral immune cells, genetic makeup (T vs. C alleles), and suicidal ideation were assessed prospectively in three separate cohorts. The remarkable findings were a significant interaction between genetic disposition (T vs. C alleles) and epigenetic status (low vs. high methylation) in predicting suicidal ideation. In addition, levels of SKA2 methylation in immune cells were highly correlated with morning (waking) cortisol levels suggesting a possible functional link between SKA2 methylation and stress levels. In some cases, high levels of methylation at SKA2 were detected months before the reported onset of suicidal ideation, raising the possibility that this biomarker was present in these individuals before they developed suicidal thoughts.

The inclusion of additional variables such as cortisol levels and anxiety further improved the predictive value of the model. This revised model was tested in another prospective cohort using genotype (T vs. C alleles), methylation rate of SKA2, and anxiety scores. This model was able to predict suicidal behavior with 71% accuracy, which was increased to 82% when levels of salivary cortisol were added. Further refinement of the stringency of categorizing suicidal ideation improved the predictive value to 97%, successfully predicting all 4 eventual suicide attempts in this particular cohort.

These provocative findings have now been replicated by other groups providing one of the most tangible examples linking epigenetic markers with complex and devastating human behaviors such as suicide. The challenges of relating peripheral biomarkers to human psychopathology and behavior were elegantly addressed in this study by showing that the transcriptional and the epigenetic programs in the peripheral cells resemble those seen in postmortem neuronal tissue and by relating the methylation rate of SKA2 in the periphery with morning cortisol levels and suicidal ideation. It will be interesting to see if the peripheral methylation rate of SKA2 could be modified by treatment with ketamine, which has recently emerged as a rapid and promising treatment for suicidal ideation.

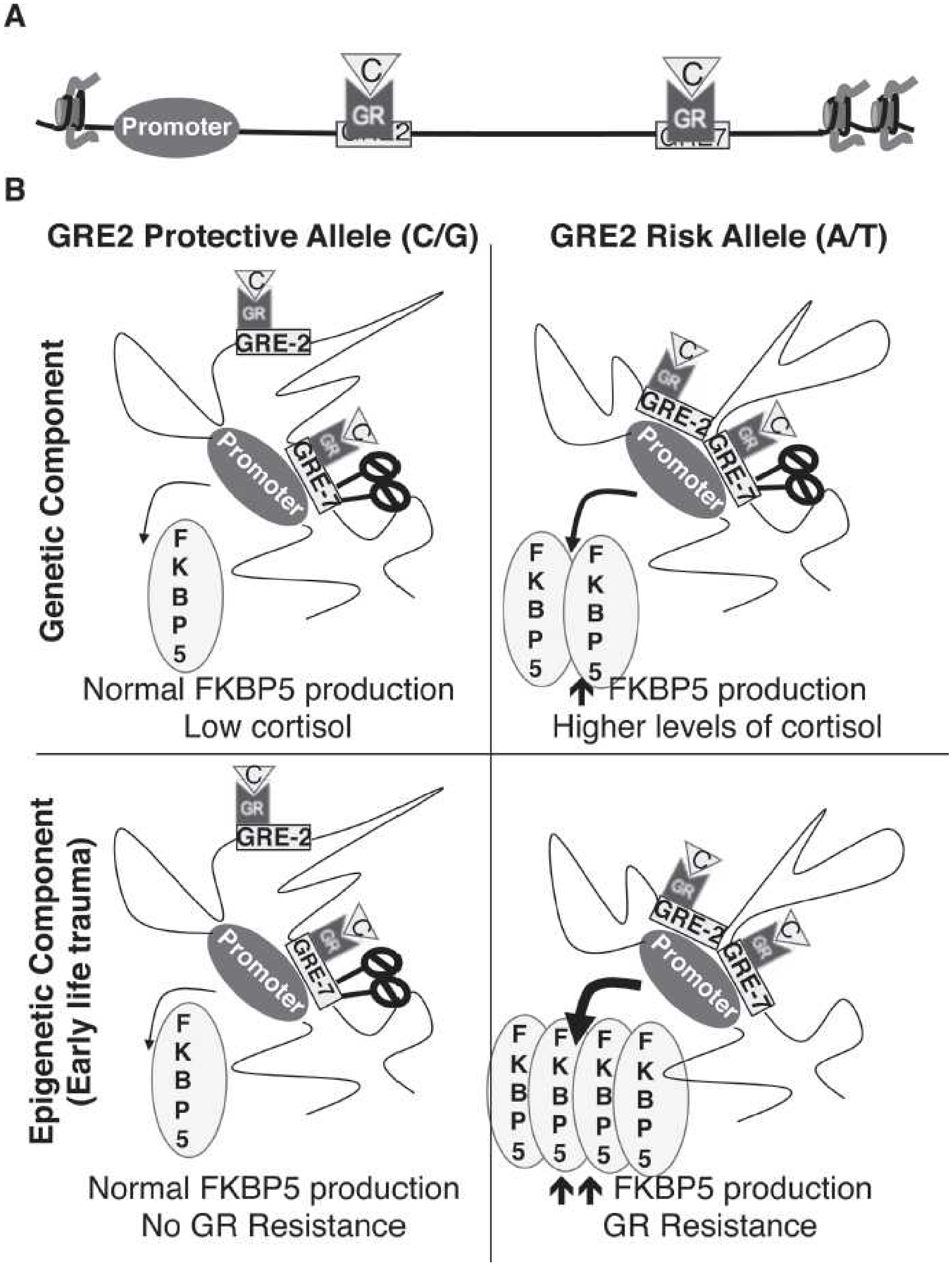
A
Promoter
C
GR
GRE-2
C
GR
GRE-7
B
GRE2 Protective Allele (C/G)
GRE2 Risk Allele (A/T)
Genetic Component
C
GR
GRE-2
Promoter
GRE-7
FKBP5
Normal FKBP5 production
Low cortisol
GRE-2
Promoter
GRE-7
FKBP5
FKBP5
↑ FKBP5 production
Higher levels of cortisol
Epigenetic Component (Early life trauma)
GRE-2
Promoter
GRE-7
FKBP5
Normal FKBP5 production
No GR Resistance
GRE-2
Promoter
GRE-7
FKBP5
FKBP5
FKBP5
FKBP5
↑↑ FKBP5 production
GR Resistance

FIGURE 1.20–18. **A:** The FKBP5 gene has two GREs located in introns 2 (GRE-2) and intron 7 (GRE-7). **B:** The top two panels demonstrate the impact of the *genetic risk* on FKPB5 expression and GR resistance. On the left, GRE-2 containing the protective C/G allele is not recruited to the promoter/GRE-7 complex producing relatively low levels of FKBP5 and low cortisol **levels** in response to mild stress. The presence of the GRE-2 at the promoter/GRE-7 complex in the A/T allele (**top right**) produces higher levels of FKBP5 and hence higher levels of cortisol in response to mild stress. The bottom panels demonstrate the *epigenetic* contribution of early life stress to FKPB5 expression. Early life stress results in relatively low levels of cortisol and normal methylation at the GRE-7 in individuals with the C/G protective allele (**lower left panel**). In contrast, early life stress leads to higher cortisol levels and a decrease in the DNA methylation at the GRE-7 of

individuals with the A/T risk allele. Demethylation at GRE-7 leads to more efficient GR binding to this site and, therefore, higher FKBP5 levels and GR resistance (**lower right panel**).

FURTHER READINGS

Cheung S, Woo J, Maes MS, Zai CC. Suicide epigenetics, a review of recent progress. *J Affect Disord*. 2020;265:423–438.

Ciabrelli F, Cavalli G. Chromatin-driven behavior of topologically associating domains. *J Mol Biol*. 2015;427:608–625.

Duffy KA, Bale TL, Epperson CN. Germ cell drivers: transmission of preconception stress across generations. *Front Hum Neurosci*. 2021;15:642762.

Guintivano J, Brown T, Newcomer A, et al. Identification and replication of a combined epigenetic and genetic biomarker predicting suicide and suicidal behaviors. *Am J Psychiatry*. 2014;171:1287–1296.

Guo JU, Ma DK, Mo H, et al. Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat Neurosci*. 2011a;14:1345–1351.

Guo JU, Su Y, Zhong C, Ming GL, Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*. 2011b;145:423–434.

Kaffman A, Meaney MJ. Neurodevelopmental sequelae of postnatal maternal care in rodents: clinical and research implications of molecular insights. *J Child Psychol Psychiatry*. 2007;48:224–244.

Kaminsky ZA, Tang T, Wang SC, et al. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet*. 2009;41:240–245.

Kim Y, Lee HM, Xiong Y, et al. Targeting the histone methyltransferase G9a activates imprinted genes and improves survival of a mouse model of Prader-Willi syndrome. *Nat Med*. 2017;23:213–222.

Klengel T, Mehta D, Anacker C, et al. Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat Neurosci*. 2013;16:33–41.

Kouzarides T. Chromatin modifications and their function. *Cell*. 2007;128:693–705.

Kraan CM, Godler DE, Amor DJ. Epigenetics of fragile X syndrome and fragile X-related disorders. *Dev Med Child Neurol*. 2019;61:121–127.

Kumar A, Choi KH, Renthal W, et al. Chromatin remodeling is a key mechanism underlying cocaine-induced plasticity in striatum. *Neuron*. 2005;48:303–314.

Liu XS, Wu H, Krzisch M, et al. Rescue of fragile X syndrome neurons by DNA methylation editing of the FMR1 gene. *Cell*. 2018;172:979–992 e976.

O'Donnell KJ, Meaney MJ. Epigenetics, development, and psychopathology. *Annu Rev Clini Psychol*. 2020;16:327–350.

Paris J. Predicting and preventing suicide: do we know enough to do either? *Harv Rev Psychiatry*. 2006;14:233–240.

Pena CJ, Bagot RC, Labonte B, Nestler EJ. Epigenetic signaling in psychiatric disorders. *J Mol Biol*. 2014;426:3389–3412.

Peter CJ, Akbarian S. Balancing histone methylation activities in psychiatric disorders. *Trends Mol Med*. 2011;17:372–379.

Piatti P, Zeilner A, Lusser A. ATP-dependent chromatin remodeling factors and their roles in affecting nucleosome fiber composition. *Int J Mol Sci*. 2011;12:6544–6565.

Reul JM, Chandramohan Y. Epigenetic mechanisms in stress-related memory formation. *Psychoneuroendocrinology*. 2007;32 (Suppl 1):S21–25.

Rogge GA, Wood MA. The role of histone acetylation in cocaine-induced neural plasticity and behavior. *Neuropsychopharmacology*. 2013;38:94–110.

Schafer A. Gadd45 proteins: key players of repair-mediated DNA demethylation. *Adv Exp Med Biol*. 2013;793:35–50.

Swygert SG, Peterson CL. Chromatin dynamics: interplay between remodeling enzymes and histone modifications. *Biochim Biophys Acta*. 2014;1839:728–736.

Tsankova NM, Berton O, Renthal W, Kumar A, Neve RL, Nestler EJ. Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. *Nat Neurosci*. 2006;9:519–525.

Vizan P, Beringer M, Ballare C, Di Croce L. Role of PRC2-associated factors in stem cells and disease. *FEBS J*. 2014.

1.21 Pharmacogenetics

JULIO LICINIO, M.D., PH.D., M.B.A., M.S., AND MA-LI WONG, M.D., PH.D.

Why do some individuals respond so well to drugs–pharmacologic treatment and others do not? This fundamental question underlies the heterogeneity of treatment responses, which is one of the biggest challenges in the pharmacologic treatment of psychiatric disorders. While some respond exceptionally well to psychopharmacologic interventions, others do not. Likewise, some, but not others, have clinically meaningful side effects. Treatment outcomes consist of multiple combinations of positive and negative treatment responses. Those combinations can be broadly grouped into six categories as follows:

1. Fully positive clinical response, that is, remission or cure with minimal or occasional side effects.
2. Fully positive clinical response, but with clinically impactful side effects.
3. Partial clinical response with minimal side effects.
4. Partial clinical response with significant side effects.
5. Limited clinical response with minimal side effects.
6. Limited clinical response with substantial side effects.

The pharmacotherapy of psychiatric disorders is a process of shared decision making with patients, as both parties must balance risks and benefits. Choices may involve compromising efficacy to improve the tolerability of treatment. The better that physicians can predict risks and benefits for patients, the more likely treatment can be optimized to achieve the shared objectives of both physician and patient. A significant challenge for this process is the limited ability of physicians to predict clinical efficacy or medication tolerability before initiating treatment.

THE SEARCH FOR INDIVIDUALIZED PSYCHOPHARMACOLOGY

For decades, clinicians have sought markers that identify the likelihood of favorable outcomes before pharmacologic treatment is administered. These markers would move clinical practice into a situation where treatments would be personalized (personalized medicine) or given with precision (precision medicine). In current routine clinical practice, when considering medication targets, the choice focuses more on the issue of adverse drug reactions (ADRs). For example, as a first line of treatment, one would avoid giving a medication with potential cardiac side effects to someone with impaired cardiac function and heart disease.

The cost of our limited ability to predict clinical response is that many patients drop out of treatment before achieving a meaningful clinical response. This point was illustrated by the National Institute of Mental Health (NIMH)-funded Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, the largest and longest study ever conducted to evaluate depression treatment. The overall goal of the STAR*D trial was to assess the effectiveness of depression treatments in patients diagnosed with major depressive disorder (MDD) in primary and specialty care settings. The study had four levels, each testing a different medication or combination. The primary goal of each level was to determine if the treatment used during that level could adequately treat participants' MDD. Those who did not remit (i.e., no symptom more than mild) could proceed to the next level of treatment. In level 1, about one-third of the participants reached remission, and about 10% to 15% more responded but did not reach remission. It is noteworthy and clinically crucial that almost 70% is remitted. However, the rate at which participants withdrew from the trial was meaningful and rose with each level—21% withdrew after level 1, 30% withdrew after level 2, and 42% withdrew after level 3.

A paradigm shift that would transform psychopharmacology from the mostly imprecise effort that it is today into a precision effort would be truly transformational for clinical psychiatry.

Pharmacogenetics' tremendous appeal comes from the promise of real-life clinical pharmacologic treatment guided by science and given to patients with precision based on their genetic makeup. Patients with similar clinical presentations treated with the same drug and dosage may have different outcomes. The underlying central hypothesis of pharmacogenetics is that one's genetic and epigenetic makeup is a significant determinant of treatment response. Thus, if we identify the genetic sequences that correlate with treatment outcomes with a specific drug, these sequences can serve as biomarkers to guide matching particular patients to specific treatments. Conversely, genetic or epigenetic biomarkers that predict adverse outcomes to particular medications could be used to guide patients to alternative medications and to prevent these bad outcomes.

PHARMACOGENETICS AND PHARMACOGENOMICS

There are two historical phases in pharmacogenetics: Before and after 2000, when the rough draft of the entire human genome was finished. Before 2000, genetic sequencing was expensive and limited as we needed the human genome blueprint. Until then, we could test for particular genetic variants at a high cost. In contrast, after 2000, technology grew at a vertiginous pace, and we can now sequence a person's entire genome at a moderate cost. As a reference, the total cost of the Human Genome Project is reported to have been USD 3 billion. Today, whole genome sequencing is offered at a direct-to-consumer price range of USD 300 to 800, depending on the level of DNA sequencing depth.

As we look at genome-wide data in some circles, a distinction was created between pharmacogenetics, based on testing specific sequences, and pharmacogenomics, based on whole genome sequence data. However, in many settings, including this chapter, the terms pharmacogenetics and pharmacogenomics are used interchangeably.

PHARMACOKINETICS AND PHARMACODYNAMICS

As we seek to optimize treatment response, it is essential to understand the two critical aspects of pharmacology with direct relevance to pharmacogenetics: pharmacokinetics and pharmacodynamics. Different proteins and their encoding genes mediate each of them. Therefore, some pharmacogenetic targets will map to pharmacokinetic pathways, while others will be related to pharmacodynamics (Fig. 1.21–1).

PHARMACOGENETICS OF PHARMACOKINETICS

Pharmacokinetics refers to all the factors that regulate drug availability. Drugs enter the body and are metabolized. Their bioavailability reflects the balance between drug administration, metabolism, and excretion. Drug metabolism occurs primarily in the liver, particularly as drugs enter the liver from the portal venous blood before reaching the systemic bloodstream in the “hepatic first pass.” Drug excretion happens principally through the kidney.

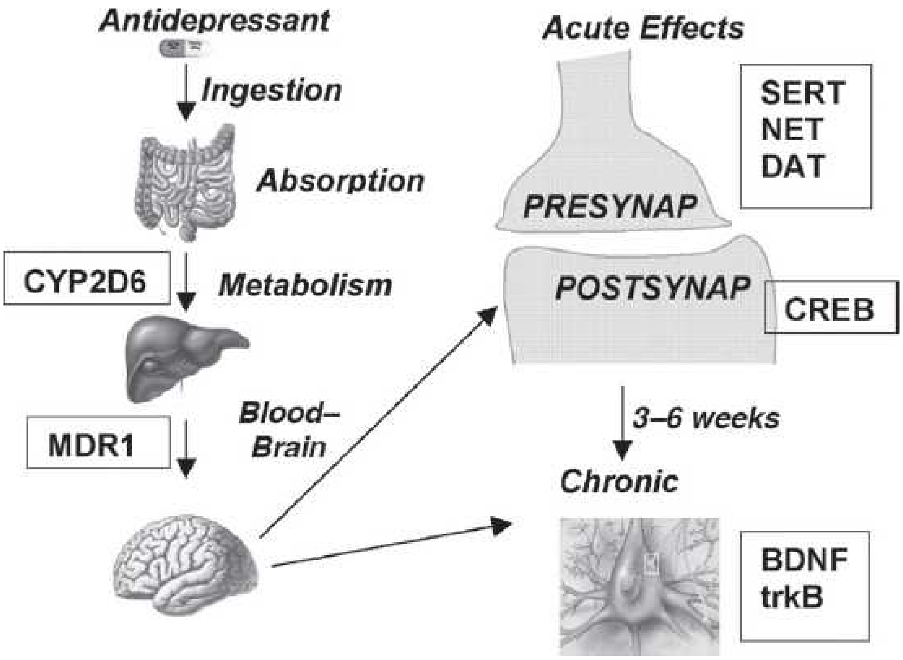

FIGURE 1.21–1. The trajectory of an antidepressant drug from ingestion through pharmacokinetics (*left-hand column*) to pharmacodynamics (*right-hand column*). Boxes show representative pharmacogenetic targets. CYP2D6, cytochrome P450 family 2 subfamily D member 6; MDR1, multidrug resistance protein 1, also known as adenosine triphosphate (ATP)-binding cassette subfamily B member 1, ABCB1; SERT, serotonin transporter, also known as solute carrier family 6 member 4, SLC6A4; NET, norepinephrine transporter, also known as solute carrier family 6 member 2, SLC6A2; DAT, dopamine active transporter, also known as solute carrier family 6 member 3, SLC6A3; CREB, cyclic adenosine monophosphate (cAMP) response element-binding protein; BNDF, brain-derived neurotrophic factor, and trkB, tyrosine receptor kinase B, the receptor for BDNF. (See eBook for color images.)

The drug-metabolizing enzymes belong to the cytochrome P450 (CYP) superfamily. Data from the Human Genome Project show 57 human genes coding for the various CYP enzymes. Of those, a small but highly significant subset is particularly relevant for psychiatry because they metabolize many of the drugs we use in our field. Furthermore, they have multiple genetic variants of functional relevance and are prime targets for treatment optimization based on pharmacogenetics. Specifically, six CYPs metabolize 90% of drugs; the two most significant enzymes are CYP3A4 and CYP2D6.

Note that there are complex relationships between multiple drugs and CYP enzymes. On the one hand, a drug can be metabolized by more than one CYP. On the other hand, when different drugs are taken together, the effect of one drug on the function CYP450 enzyme can affect the metabolism of another drug by that same enzyme. Specifically, some drugs are solely substrates for CYP enzymes, but other drugs or herbal products can either induce or inhibit the activity of a CYPP450 enzyme. Therefore, taking one drug that either induces (e.g., haloperidol) or inhibits (e.g., bupropion, methylphenidate, sertraline, or escitalopram) the activity of CYP2D6 will affect the metabolism of drugs that are substrates for CYP2D6, such as imipramine, fluoxetine, venlafaxine, duloxetine, mirtazapine, or oxycodone. It is a relatively common situation clinically that someone will have attention deficit hyperactivity disorder (ADHD) and depression and may be treated with a combination of the CYP2D6 inhibitor methylphenidate and an antidepressant that is a CYP2D6 substrate, such as fluoxetine, venlafaxine, duloxetine, or mirtazapine. In these cases, patients may better respond to lower doses of the (CYP2D6 substrate) antidepressant.

CYP3A4

This enzyme metabolizes about 50% of all drugs on the market, including tricyclic antidepressants (TCAs), selective serotonin reuptake inhibitors (SSRIs), some other antidepressants, buspirone, multiple antipsychotics, opioids (mainly analgesics), methadone, benzodiazepines, anticonvulsants, mood stabilizers, and barbiturates. Over 38 single-nucleotide polymorphisms (SNPs) have been identified in the *CYP3A4* gene, which is part of a cluster of CYP450 genes on chromosome 7q22.1, some resulting in decreased and others increased catalytic activity with certain ligand drugs, compared to wild-type CYP3A4 metabolism. Of note, CYP3A4 can be induced on exposure to substrates. CYP3A4 activity levels have also been linked to diets and environmental factors, such as duration of exposure to xenobiotic substances and nutritional status, given the enzyme's extensive presence in the intestinal mucosa. Therefore, directly translating pharmacogenetics into significant interindividual variability *in vivo* is not straightforward.

CYP2C19

CYP2C19 is a member of the CYP2C subfamily of the mixed-function oxidase system, which includes enzymes responsible for metabolizing xenobiotics such as proton pump inhibitors and antiepileptic drugs. In humans, the *CYP2C19* gene, located on chromosome 10q24, encodes the CYP2C19 protein, which is primarily found in the liver and responsible for metabolizing at least 10% of drugs currently used in clinical practice. These include clopidogrel (Plavix), which is used as an antiplatelet treatment, pain medications such as omeprazole, antiseizure drugs like mephenytoin, and diazepam, as well as some barbiturates. CYP2C19 is also involved in epoxygenase activity and plays a crucial role in the formation of signaling agents through the attack of double bonds in long-chain polyunsaturated fatty acids.

Also known as loss-of-function alleles, *CYP2C19**2* and *3* have been linked to a reduction in enzyme activity, while *CYP2C19**17*, a gain-of-function allele, leads to increased activity. These three variant alleles are recommended to be included in the minimal clinical pharmacogenomic testing panel.

CYP2C19 is also involved in the metabolism of several TCAs such as imipramine, amitriptyline, trimipramine, and clomipramine, as well as the secondary amine TCA nortriptyline. In CYP2C19 poor metabolizers (PMs), these TCAs have been found to lead to greater plasma concentrations and active metabolite levels, resulting in a higher risk of adverse effects. CYP2C19 is also involved in the metabolism of SSRIs, including citalopram, sertraline, fluoxetine, and venlafaxine, as well as the reversible monoamine oxidase inhibitor moclobemide. CYP2C19 genotype has been found to affect citalopram serum concentration, while patients with one or two loss-of-function alleles have higher dose-adjusted serum concentrations of sertraline than normal metabolizers, which can predict outcomes.

The clinical relevance of CYP2C19 ultrarapid metabolizers (UMs) is still under investigation, and *CYP2C19*17* has been found to be linked to lower serum concentrations of several antidepressants in comparison to normal metabolizers. However, further research is required to ascertain the clinical importance of UM genotypes in antidepressant therapy.

CYP2D6

CYP2D6 constitutes up to 2% of hepatic CYP content and it is responsible for the metabolism and elimination of approximately 25% of clinically used drugs by adding or removing specific functional groups, such as hydroxylation, demethylation, and dealkylation. CYP2D6 is encoded by genes on human chromosome 22. *CYP2D6* has multiple functional alleles, some of them providing standard effects, while others have either low or high activity. *The CYP2D6*1* allele is commonly used as a reference allele and is assumed when no variant is detected by a genotyping test. Individuals with a normal enzyme activity and no variants detected are often associated with the *CYP2D6*1* allele, which is known as the extensive metabolizer phenotype. Some groups use this term to refer to individuals with this phenotype. The *CYP2D6* gene may also be duplicated or multiduplicated in some individuals who then have very high CYP2D6 enzymatic activity. Please refer to Figure 1.21–2 for further details. Those drugs are among the most widely used in medicine and psychiatry. Please refer to Figure 1.21–3 for a diagram showing recognized pharmacogenes according to their association with drugs used in different therapeutic areas.

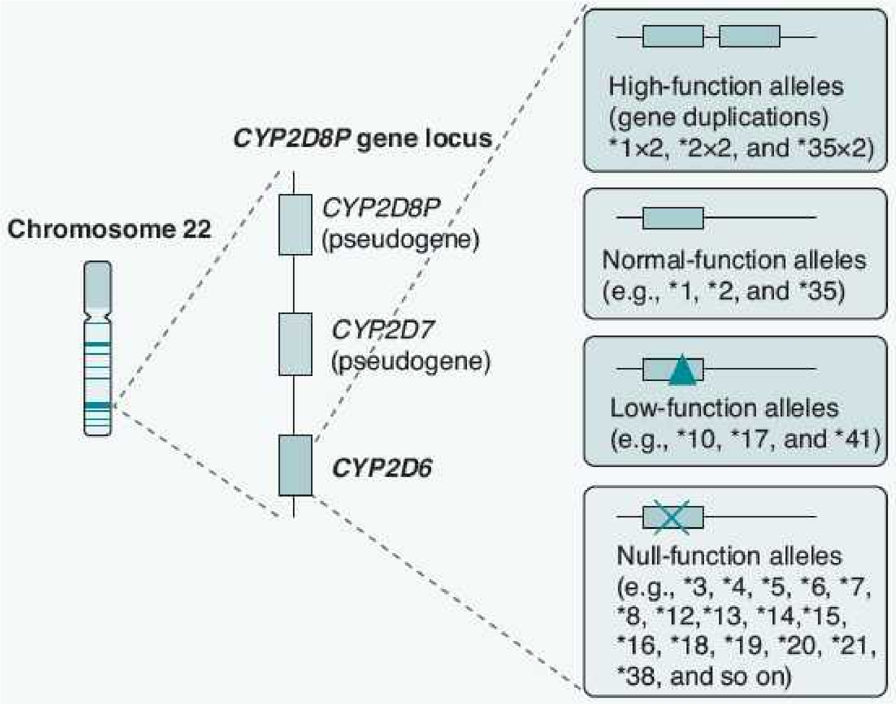
CYP2D8P gene locus
Chromosome 22
CYP2D8P
(pseudogene)
CYP2D7
(pseudogene)
CYP2D6
High-function alleles
(gene duplications)
*1×2, *2×2, and *35×2)
Normal-function alleles
(e.g., *1, *2, and *35)
Low-function alleles
(e.g., *10, *17, and *41)
Null-function alleles
(e.g., *3, *4, *5, *6, *7,
*8, *12,*13, *14,*15,
*16, *18, *19, *20, *21,
*38, and so on)

FIGURE 1.21–2. Selected CYP2D6 genetic polymorphisms, with functional implications. Genetic variability is the primary cause of interindividual phenotypic diversity in CYP2D6. The genetic spectrum of CYP2D6 comprises several hundred well-defined allelic variants, such as null alleles, functional variants, splice variants, and structural variants including complete gene loss, gene amplifications, and hybrid genes. When compared to extensive metabolizers (EMs), who have at least one fully functional allele, individuals classified as poor metabolizers (PMs) carry two nonfunctional alleles and have a significantly reduced ability to eliminate CYP2D6 drug substrates. The intermediate metabolizer (IM) phenotype is another genetically determined trait that results from the presence of reduced-function alleles, while the ultrarapid metabolizer (UM) phenotype is caused by CYP2D6 gene duplications of functional alleles. (See eBook for color images.) (With permission from Zanger UM. Genetic variability of CYP2D6: basic and clinical aspects. In: Baumann P, ed. *CYP2D6: Genetics, Pharmacology and Clinical Relevance*. Future Medicine Ltd, Future Science Group; 2014:28–41. https://doi.org/10.2217/fmeb2013.14.16.)

Even though CYP3A4 metabolizes more drugs, CYP2D6 may be clinically more critical as it metabolizes more of the most widely used drugs. CYP2D6 is particularly relevant to psychiatry, as it metabolizes all TCAs, most selective serotonin receptor reuptake inhibitors—such as fluoxetine, paroxetine, fluvoxamine—as well as venlafaxine, duloxetine, mianserin, mirtazapine, minaprine, opioids, codeine, oxycodone, hydrocodone, antipsychotics—such as haloperidol, risperidone, perphenazine, thioridazine, aripiprazole, chlorpromazine, levomepromazine, and remoxipride—cocaine, amphetamine, methoxyamphetamine, dextromethamphetamine, lisdexamfetamine, and atomoxetine. CYP2D6 also metabolizes widely used drugs, such as beta-blockers, antiarrhythmics, calcium channel blockers, H3 receptor antagonists, antihistamines, and multiple other therapeutically essential drugs, such as tamoxifen.

Based on one's combination of *CYP2D6* genetic alleles, individuals can be functionally classified as PMs (little or no CYP2D6 function), intermediate metabolizers (IMs) (metabolize drugs at a rate somewhere between the poor and extensive metabolizers), extensive metabolizers (standard CYP2D6 function), and UMs, who have multiple copies of the *CYP2D6* gene, resulting in greater than normal function. CYP allele frequencies and metabolizer status vary across populations worldwide. Therefore, normative data from one population may not be applicable across the board.

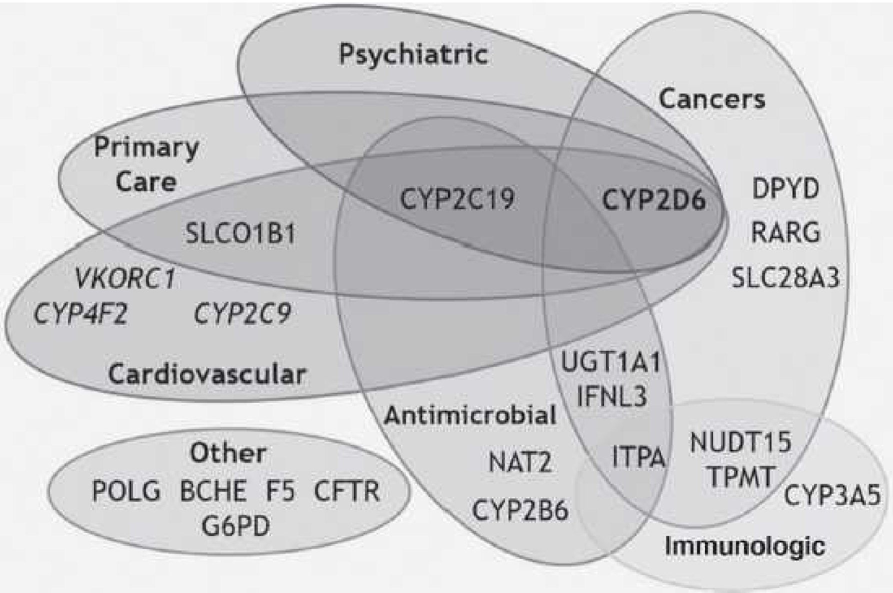

FIGURE 1.21–3. Venn Diagram showing recognized (non-HLA) pharmacogenes according to their association with drugs used in different therapeutic areas. CYP2D6 overlaps with four distinct therapeutic areas. BCHE, butyrylcholinesterase; CFTR, cystic fibrosis transmembrane conductance regulator; CYP2B6, CYP2C9, CYP2C19, CYP3A5, CYP4F2, members of cytochrome P450 superfamily; DPYD, dihydropyrimidine dehydrogenase; F5, factor 5; G6PD, glucose-6-phosphate dehydrogenase; IFNL3, Interferon lambda 3; ITPA, inosine triphosphate pyrophosphohydrolase; NAT2, N-acetyltransferase 2; NUDT15, nudix hydrolase 15; POLG, DNA polymerase subunit gamma; RARG, retinoic acid receptor gamma; SLCO1B1, solute carrier organic anion transporter family member 1B1; SLC28A3, solute carrier family 28 member 3; TPMT, thiopurine S-methyltransferase; UGT1A1, UDP-glucuronosyltransferases; VKORC1, vitamin K epoxide reductase complex. (See eBook for color images.) (With permission from Taylor C, Crosby I, Yip V, Maguire P, Pirmohamed M, Turner RM. A review of the important role of CYP2D6 in pharmacogenomics. *Genes*. 2020;11:1295. doi:10.3390/genes11111295. License: https://creativecommons.org/licenses/by/4.0/)

The more active CYP2D6 is, the faster the drugs metabolized by this enzyme will be cleared. Therefore, standard doses may be insufficient for UM, and higher doses are required. CYP2D6 UMs may not respond to a psychiatric drug simply because their standard dose is too low. They are then inappropriately labeled as nonresponders when they have yet to receive an adequate dosage of a drug that might be excellent for them. Conversely, poor and IMs will break down prescribed drugs more slowly than those with standard CYP2D6 function. At least 15 allelic variants of *CYP2D6* can cause PM phenotype, but 75% of PMs are *CYP2D6*4* homozygotes. Therefore, by taking standard dosages, they accumulate bioactive drugs that may rapidly result in serious side effects. The patients are then incorrectly labeled as intolerant to a drug that was actually given to them in doses that far exceeded their metabolic capacity. In this case, the doses should be lowered to offset the slow metabolism, and individuals may respond well to the individually tailored dosage.

Multiple reports have shown that individuals who appear to be treatment refractory were either UM or PM. The UM did not receive enough medication dosage, and the PM received doses that exceeded their capacity to metabolize the drug. Therefore, they involuntarily and rapidly overdosed and experienced ADRs that led them to drop out of treatment before a clinical effect could be achieved.

CYP2D6 Phenocopying and Autophenocopying. The use of CYP2D6 substrates in therapy can be complicated, not only due to genetic variability but also due to drug–drug interactions. Several drugs, such as statins, are CYP2D6 inhibitors, and co-administration with a CYP2D6 substrate can alter the apparent phenotype of a patient. This occurrence is referred to as phenocopying. In such situations, an individual with a standard metabolizer phenotype may appear to be an IM or a PM because most of the available enzyme is being inhibited by the confounding drug. Another related phenotype that can arise with chronic administration of a CYP2D6 drug is called autophenocopying, in which the substrate inhibits its own metabolism as the drug concentration reaches a steady state. Consequently, the pharmacokinetic profile of a single dose and of repeated dosing for drugs that exhibit phenocopying can vary significantly.

PHARMACOKINETICS BEYOND DRUG LEVELS OF THE PARENT DRUG

The effects of P450 genotypes on drug bioavailability go beyond the simple pharmacologic prediction of the levels of the parent compound. In many cases, drug development takes drug metabolism into action, and the drug that is prescribed and ingested is a prodrug that requires metabolism for bioactivity. In these cases, we have the reverse situation of what was described above. PM may have insufficient active drugs as they do not metabolize the parent drug well, and UM may then have excessive bursts of the active drug as they metabolize the parent compound very rapidly.

Drugs may be formulated as prodrugs for a variety of reasons. These include improved stability during manufacture and storage, enhanced absorption by the body, or superior pharmacokinetics. All of these advantages of prodrugs make them attractive for use in drug formulations. Since its inception in 1958, the concept of prodrugs, proposed by Albert, has been used to alter the physical and chemical properties of drugs to render them pharmacologically inactive until metabolized in the body.

Lisdexamphetamine

Lisdexamphetamine (LDX) is a unique type of long-acting stimulant prodrug that was the first of its kind for treating ADHD and binge-eating disorder. It also has similar off-label uses as other pharmaceutical amphetamines. LDX is inactive and requires hydrolysis to convert it into the potent central nervous system (CNS) stimulant dextroamphetamine. The chemical structure of LDX consists of the amino acid L-lysine attached to dextroamphetamine. Although drug efficacy may not be directly related to pharmacokinetic characteristics, LDX's significant impact on symptoms could be linked to the predictable bioavailability of the active ingredient, d-amphetamine, upon hydrolysis from the inactive prodrug. The United States approved LDX for medical use in 2007, and in 2020, it was the 85th most commonly prescribed medication in the country, with over 8 million prescriptions. LDX is classified as a Class B controlled substance in the United Kingdom and a Schedule II controlled substance in the United States.

Codeine

Codeine is a prodrug that is converted to morphine, a potent mu-opioid receptor agonist, by the CYP2D6 enzyme, as illustrated in Figure 1.21–4. Ultrarapid metabolizers (UM), who produces higher levels of active metabolites such as morphine and morphine-6-glucuronides, are at an increased risk for opioid-related side effects, including respiratory depression. Several case reports have documented severe and even life-threatening opioid intoxication in patients with UM due to functional duplication of the CYP2D6 gene. In addition, a clinical study showed that breastfeeding newborns whose mothers had an ultrarapid metabolism genotype were at a higher risk of respiratory depression. As a result, the Clinical Pharmacogenetics Implementation Consortium has published guidelines recommending genotyping before prescribing codeine. Moreover, an observational study on postpartum analgesia revealed that patients with poor metabolism genotypes received no analgesic effect from codeine, whereas those with ultrarapid metabolism genotypes experienced sedation.

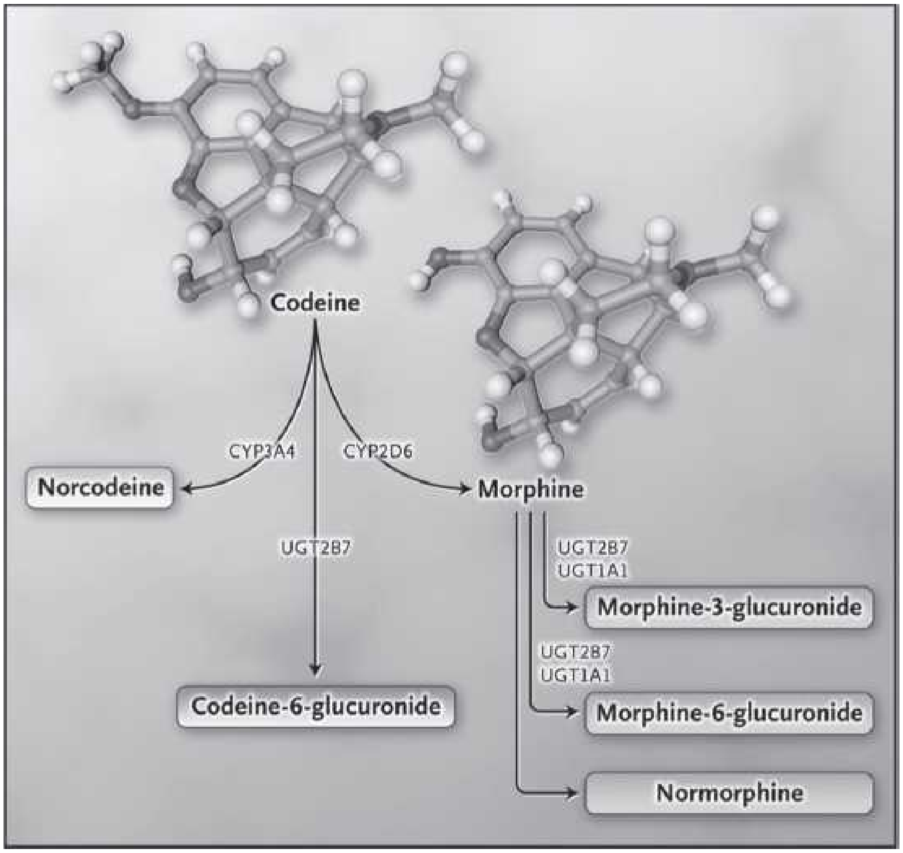
Codeine
CYP3A4
CYP2D6
Norcodeine
Morphine
UGT2B7
UGT2B7
UGT1A1
Morphine-3-glucuronide
UGT2B7
UGT1A1
Codeine-6-glucuronide
Morphine-6-glucuronide
Normorphine

FIGURE 1.21–4. Codeine Metabolism Pathway. The activity of codeine depends on its conversion to morphine by CYP2D6. Morphine is subsequently metabolized to the active morphine-6-glucuronide by means of UDP-glucuronosyltransferase 2B7. The gene encoding CYP2D6 has many genetic variations that affect the amount of codeine that is converted to an active form and that results in the drug's variable effect. (From Racoosin JA, Roberson DW, Pacanowski MA, Nielsen DR. New evidence about an old drug–risk with codeine after adenotonsillectomy. *N Engl J Med*. 368;23:2155–2157.)

Tramadol

Similarly, tramadol, another prodrug, is metabolized by CYP2D6 to O-desmethyltramadol, also known as (+)-M1, which has mu-opioid receptor activity. Patients with UM genotypes have been found to have higher levels of tramadol metabolites and a reduced experience of discomfort from experimental pain. In contrast, those with PM genotypes showed minimal clinical effects with low levels of detectable metabolites. Case reports have also highlighted the potential risk of respiratory depression in patients with UM genotypes.

THE PHARMACOKINETICS OF ONE DRUG MAY GENERATE MULTIPLE ACTIVE COMPOUNDS

In other, more complex cases, both the parent drug and its metabolite(s) are bioactive. There are several examples where the pharmacokinetics of administered drugs results in active metabolites with properties that may differ from that of the parent compound. The pharmacogenetics of CYP enzymes may determine final therapeutic effects based on the different combinations of parent compounds versus metabolites that are generated due to the bioactivity of CYP enzyme isoforms. There are several such cases in psychopharmacology. We provide here some illustrative examples.

Clozapine and Desmethylclozapine

Clozapine and its active metabolite, N-desmethylclozapine (NDMC), are metabolized by several CYP enzymes, including CYP1A2, CYP2D6, CYP3A4, and CYP2C19. The activity of CYP1A2 is particularly relevant in determining the required dose of clozapine. While clozapine is an atypical antipsychotic that acts as an antagonist at the D2/D3 receptors, NDMC has intrinsic activity at these receptors and acts as a weak partial agonist, similar to aripiprazole and bifeprunox. Moreover, NDMC has been shown to be a potent agonist at the M1 and delta-opioid receptors, unlike clozapine. These unique actions of NDMC led to the hypothesis that it might be responsible for the clinical superiority of clozapine over other antipsychotics. However, clinical trials have demonstrated that NDMC alone is ineffective in treating schizophrenia, specifically in managing positive symptoms. Nevertheless, it remains possible that NDMC could contribute to the efficacy of clozapine in treating cognitive and negative symptoms of the disease. Targeting M1 receptor agonism could be a promising avenue for developing drugs that can address cognitive deficits in schizophrenia and other neuropsychiatric disorders.

Clomipramine and Desmethyl Clomipramine

Clomipramine, a TCA, undergoes significant biotransformation before excretion. One of the primary routes of hepatic first-pass metabolism is demethylation into desmethyl clomipramine by several CYP enzymes, including CYP2C19, CYP3A4, and CYP1A2. In-vivo studies have shown that CYP2C19 is involved in this process, while in-vitro studies have demonstrated the involvement of CYP3A4 and CYP1A2. Notably, smokers exhibit lower plasma concentrations of clomipramine than nonsmokers, suggesting a role for smoking-induced CYP1A2. Desmethyl clomipramine is an active metabolite that is a more potent inhibitor of norepinephrine reuptake and a weaker serotonin reuptake inhibitor than the parent drug. Both the parent drug and the active metabolite undergo hydroxylation, catalyzed by CYP2D6, to form several metabolites, including 8-hydroxy desmethyl clomipramine, 8-hydroxy clomipramine, 2-hydroxy clomipramine, and 2-hydroxy desmethyl clomipramine. The 2- and 8-hydroxylated metabolites are primarily excreted through the kidneys as glucuronide conjugates. Desmethyl clomipramine can also be further demethylated to produce didesmethyl clomipramine. The elimination half-life is approximately 24 hours for clomipramine and 96 hours for desmethyl clomipramine. Typically, it takes around 3 weeks to reach steady-state concentration.

Dextromethorphan and Dextrorphan

Dextromethorphan (DM) and its active metabolite, dextrorphan (DXO), share pharmacodynamic similarities with the antidepressant ketamine and have been suggested to possess rapid-acting, conventional, and treatment-resistant antidepressant properties. DM is primarily metabolized to DXO by CYP2D6. DXO is an N-methyl-D-aspartate (NMDA) antagonist and contributes to the psychoactive effects of DM, producing phencyclidine-like behavioral effects in animals. In addition, DXO exhibits anticonvulsant and neuroprotective properties in various experimental models.

Diazepam and Oxazepam

Diazepam is metabolized to either nordiazepam by CYP 2C19 or temazepam by CYP3A4. Nordiazepam and temazepam are hydroxylated and demethylated to oxazepam. Oxazepam is then glucuronidated and excreted in the urine. Therefore, due to its pharmacokinetics, those who take diazepam have a combination of multiple active metabolites, including nordiazepam, temazepam, and oxazepam. Diazepam treats anxiety, but it is more likely to cause dependence. Oxazepam treats short-term anxiety and insomnia and is also used as a first-line treatment to prevent alcohol withdrawal symptoms.

PHARMACOKINETICS AND ALCOHOLISM

Drug metabolism (pharmacokinetics) can determine the susceptibility to alcohol use disorder (AUD). The effects of alcohol on various organs, including the brain, are influenced by the amount of ethanol and the duration of exposure. These factors are impacted by how quickly the ethanol is absorbed into the bloodstream and tissues and how efficiently it is metabolized. The liver is the primary site of ethanol metabolism, and the main process for ethanol metabolism involves converting it into acetaldehyde through alcohol dehydrogenases (ADHs) enzymes and then into acetate by aldehyde dehydrogenase (ALDH) enzymes. Other enzymes, such as CYP2E1, metabolize a small portion of ingested ethanol.

Multiple genes encode ADH and ALDH enzymes, and the variants of these enzymes differ in their rates of ethanol and acetaldehyde metabolism and production levels. These variations can impact an individual’s drinking patterns and the likelihood of developing alcohol dependence. Certain ADH and ALDH alleles have been shown to reduce the risk of becoming alcohol dependent, and these gene associations are widely recognized and significant. However, the frequency of ADH and ALDH variants varies among ethnic groups.

The mechanism by which ADH and ALDH variants influence the risk of alcoholism involves an increase in acetaldehyde levels. This may be due to either more rapid oxidation of ethanol (in cases with more active ADH variants) or slower oxidation of acetaldehyde (in cases with less active ALDH variants).

Acetaldehyde is a toxic substance that can produce an unpleasant reaction, including facial flushing, nausea, and tachycardia. The ADH1B2 allele, associated with fast ethanol oxidation, protects against alcohol dependence in various populations.

In East Asians, carrying one ADH1B2 allele reduces the odds ratio of developing alcoholism to 0.19, and carrying two ADH1B2 alleles reduces it to 0.12, compared to a person carrying two ADH1B1 alleles. The ADH1B2 allele is less prevalent in European and African populations but still offers protection against alcoholism. Among people of Jewish ancestry, the ADH1B2 allele occurs with moderate frequency and reduces binge drinking and the risk of alcoholism. However, the protective effect of ADH1B2 is weaker in European populations than in Asian populations, possibly due to varying social and environmental factors. It could also be co-inherited with other unidentified ADH variations that affect the risk of alcoholism, which could differ among Europeans and Asians.

PHARMACOKINETICS, PHARMACOGENETICS, AND DRUG ENTRY INTO THE BRAIN

In addition to liver enzymes, the pharmacokinetics of drugs acting in the brain, such as those used to treat psychiatric disorders, depend on proteins that regulate drug entry into the CNS. As an example, adenosine triphosphate (ATP)-binding cassette (ABC) subfamily B member 1 (*ABCB1*), also known as multidrug resistance protein 1 (*MDR1*), is 1 of 49 putative members in the superfamily of human ABC transporters, encoding transporter and channel proteins that function as efflux pumps. *ABCB1/MDR1* appears to modulate the brain concentrations of several antidepressants.

GEOGRAPHIC AND ETHNIC PHARMACOKINETIC VARIABILITY

The CYP enzymes are highly polymorphic, with immense genetic variability. That variability is deposited and freely available in two federally funded and connected databases, the Pharmacogene Variation Consortium or PharmVar (https://www.pharmvar.org) and the Pharmacogenomics Knowledgebase or PharmGKB (https://www.pharmgkb.org). As an example, PharmVar shows 171 CYP2D6 alleles, most with multiple variants, demonstrating the existence of several hundred CYP2D6 allele variants. For CYP2C19, PharmVar shows 39 alleles, some of them with several variants, while for CYP2C19, PharmVar shows 38 alleles, some of them also having several variants. Those allele frequencies vary substantially worldwide, across different regions and ethnic groups. Therefore, therapeutic ranges for specific drugs can vary widely depending on the genetic makeup of diverse populations worldwide.

Using CYP2D6 UMs as an example, considerable variability in allele frequency is found in UM. This is because those individuals tend to be undertreated, as standard doses of medications metabolized by CYD2D6 are rapidly cleared from their bodies. It is, therefore, not surprising to see that UMs have higher hospital admission rates, presumably, because their underlying diseases are undertreated. To show the very drastic variations in rates of UM across populations, a large study conducted in Kentucky, United States, showed that the overall prevalence of CYP2D6 UMs was 1.5% in Caucasians and 2.0% in African Americans. This is consistent with other studies. In contrast, other work by Aklillu et al. (2006) has shown that 29% of the population in Ethiopia carried alleles with duplicated or multiduplicated CYP2D6 genes. In Saudi Arabia, 21% of subjects studied had genetic profiles indicative of CYP2D6 gene duplication, as documented by McLellan RA et al. (1997).

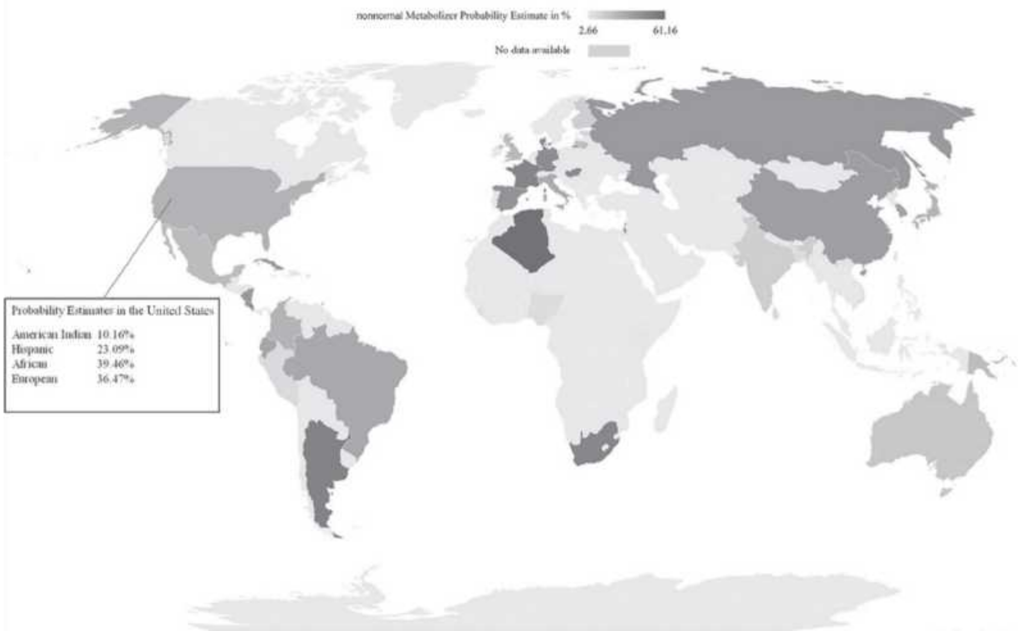

FIGURE 1.21–5. Worldwide CYP2D6 nonnormal metabolizer probability estimates. (From Koopmans AB, Braakman MH, Vinkers DJ, Hoek HW, van Harten PN. Meta-analysis of probability estimates of worldwide variation of CYP2D6 and CYP2C19. *Transl Psychiatry.* 2021;11:141. doi:10.1038/s41398-020-01129-1. License: https://creativecommons.org/licenses/by/4.0/)

Koopmans et al. conducted a meta-analysis of probability estimates of worldwide variation of CYP2D6 and CYP2C19. Due to the high frequency of allele duplications, high percentages of CYP2D6 UM were found in the Mozabite people, a Berber ethnic group in the Sahara, North Africa (39.5%), in non-Austronesian Melanesians (21.5%), and the ethnoreligious Druze from the Middle East (21.4%). On the other hand, high percentages of CYP2D6 PM were found in Europeans, for example, in the British (12.1%), the Danish (10.6%), and Basque (French) people (9.7%) due to the high frequency of *CYP2D6*4*. Frequencies of CYP2D6 predicted phenotypes by ethnicity worldwide are summarized in Figure 1.21–5.

Such evidence covers UM as an illustration. Substantial cross-population variability also exists for CYP2D6 PMs. Bertilsson et al. described that among Swedish Caucasians, 6.3% were PM. This incidence is very similar to that in other European Caucasian populations (7% to 10%). In collaboration with Lou and associates in Beijing, they showed that the incidence of PM among the Chinese was only 1%. A similar low incidence of PM has been shown in Japanese and Koreans. As they assessed PM in various communities, they discussed the genetic heterogeneity of different black populations of Africa. Looking at CYP2D6 phenotype from genotype across world populations, Gaedigk and colleagues showed that allele frequencies vary considerably across the major ethnic groups predicting PM status (activity score, AS = 0) between 0.4% and 5.4% across world populations. The prevalence of genotypic intermediate (AS = 0.5) and standard (AS = 1, 1.5, or 2) metabolizers ranges between 0.4% and 11% and between 67% and 90%, respectively. Finally, 1% to 21% of subjects (AS >2) are predicted to have UM status. The ranges of 0.4% to 5.4% and 1% to 21% are extensive and demonstrate that standard doses are not standard across populations worldwide.

CULTURE OR ETHNICITY-RELATED PHARMACOGENETICS?

Another point to consider is that both cultures and allele frequencies vary across populations, and sometimes pharmacogenetic variants may be mistaken for cultural factors. For example, in Hispanic women, depressive symptoms and ADRs frequently present as somatic symptoms. These symptoms may include fatigue, headaches, backaches, insomnia, and loss of appetite, which in Hispanics may be misattributed to cultural factors when they could be due to genetic background. In our pharmacogenetic studies, we reported on Hispanic patients who were homozygous for *CYP2D6*4* and were consequently PM. While receiving low doses of antidepressants, the blood levels of the treatment drug were substantially above the therapeutic range, with concomitant ADRs, namely worsened anxiety and physical symptoms including palpitations, hyperventilation, and myalgias. They were appropriately managed because they were in a prospective antidepressant pharmacogenetic study. Their genetically determined somatic symptoms might have been easily attributed to cultural factors in other settings.

PHARMACOGENETICS OF PHARMACODYNAMICS

Pharmacodynamics refers to the interactions of drugs with their targets, in our case, the CNS, as well as all downstream effects relevant to a drug's effects and mechanisms of action. Drugs can fail because they must achieve sufficient concentrations at the targets in the context of pharmacokinetic factors. On the other hand, even at sufficient or even high levels at the target cells, some individuals may not respond sufficiently to a drug because of genetic variants with functional relevance at the level of drug targets, such as receptors and transporters, or at the level of downstream second messenger pathways. While clinical pharmacology has evolved substantially to identify the enzymes involved in drug metabolism and their genetically based functional variability, we still need to fully identify the ultimate mechanisms of action and therapeutic pathways for most psychotropic drugs. With few exceptions, such as drugs used to treat anxiety and ADHD, most drugs used in psychiatry take a long time to exert their therapeutic effects, despite rapidly reaching immediate targets. Those targets include the human serotonin transporter encoded by the *SLC6A4* gene, the target for SSRIs, and the dopamine receptor D_2, also known as D2R, encoded by the *DRD2* gene, the target for antipsychotics. Clinical effects require many days to weeks of continuous drug treatment; therefore, drug action mechanisms are beyond the initial targets, and the final therapeutic aims have not yet been identified. It is challenging to develop pharmacogenetic biomarkers of therapeutic response at the pharmacodynamic level without fully elucidating the mechanisms of action for most drugs used in psychiatry. As that knowledge basis grows, the number of potential pharmacodynamic pharmacogenetic targets increases exponentially. For example, in 2011, we published a paper covering 29 pharmacodynamic targets in the pharmacogenetics of antidepressants (Licinio & Wong, 2011). In 2023, when searching PharmGKB.org for antidepressant pharmacogenomics, we find 1,300 variant annotations and 334 clinical annotations.

On a positive note, as new pharmacodynamic pathways of treatment responses are identified, not only will the genes related to those pathways be natural aims for future pharmacogenetic testing, but they will also have the potential to become therapeutic targets for conceptually novel treatments.

PHARMACOGENETICS OF SPECIFIC PSYCHIATRIC DISORDERS

This evolving field has been expertly reviewed elsewhere. We have edited one of the first books on this topic (Licinio & Wong, 2009). We would refer readers to relevant recent papers on each disorder that are listed in our reference section, including the pharmacogenetics of major depression and bipolar disorder, exciting work on the pharmacogenetics of lithium treatment response, as well as advances on the pharmacogenetics of ketamine. In schizophrenia, a particular area of interest has been the use of pharmacogenetics to identify those most likely to gain substantial weight and develop metabolic complications following antipsychotic use, particularly, but exclusively restricted clozapine. Other areas of emerging pharmacogenetic interest include, but are not restricted to, ADHD, obsessive-compulsive disorder (OCD), and anorexia nervosa.

Using antidepressants as an example, Figure 1.21–1 depicts the steps from drug intake to action, with representative pharmacogenetic targets along the way. With antidepressants and antipsychotics as case examples, Figures 1.21–6 and 1.21–7 depict CYP2D6-mediated quantitative influences on the pharmacokinetics of antidepressant (see Fig. 1.21–6) and antipsychotic (see Fig. 1.21–7) drugs.

PHARMACOGENETIC TESTING: A BRIEF OVERVIEW

Optimizing drug dosage based on genetic profiles related to CYP enzymes is the constituent element of pharmacogenetics that is most available for immediate translation to clinical settings. As an example of emerging translation, there are now various Medicare-funded pharmacogenetic tests that inform the optimal prescribing of antidepressants by taking CYP enzyme genotypes and their predicted functional capacity into consideration to adjust drugs and dosages.

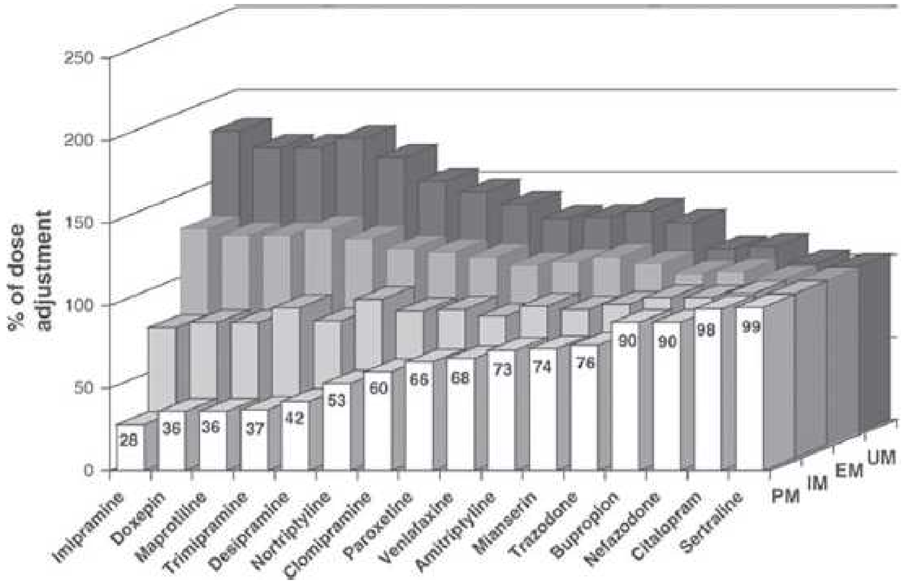

FIGURE 1.21–6. CYP2D6-mediated quantitative influences on the pharmacokinetics of antidepressant drugs expressed as percent dose adjustments: CYP2D6 PM (*white*), IM (*gray*), EM (*dark gray*), UM (*black*). If data on active drug moiety (consisting of active principle metabolite plus parent drug of active enantiomers of a racemic drug) were given, dose recommendations were based on these data only (other studies not providing such detailed information were not incorporated). If more than one study was integrated, the weighted mean for the dose adjustment was taken according to the number of poor metabolizers in each study. Data on mirtazapine, moclobemide, fluoxetine, and maprotiline were not shown in the figure since no dose adjustment based on CYP2D6 can be recommended at present. (With permission from Kirchheiner J, Nickchen K, Bauer M, et al. Pharmacogenetics of antidepressants and antipsychotics: the contribution of allelic variations to the phenotype of drug response. *Mol Psychiatry.* 2004;9:442–473. doi: 10.1038/sj.mp.4001494.)

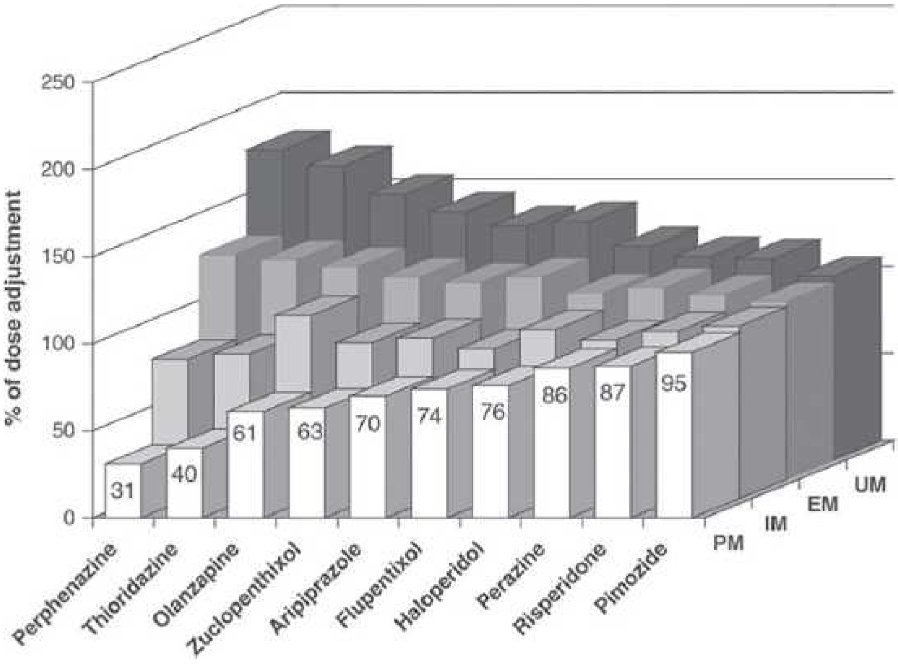

FIGURE 1.21–7. CYP2D6-mediated quantitative influences on the pharmacokinetics of antipsychotic drugs expressed as percent dose adaptations: CYP2D6 PM (*white*), IM (*gray*), EM (*dark gray*), UM (*black*). (With permission from Kirchheiner J, Nickchen K, Bauer M, et al. Pharmacogenetics of antidepressants and antipsychotics: the contribution of allelic variations to the phenotype of drug response. *Mol Psychiatry.* 2004;9:442–473. doi: 10.1038/sj.mp.4001494.)

Despite the potential ease of translational pharmacogenetics related to pharmacokinetics, it must be highlighted that this pharmacology component primarily addresses bioavailability. Accordingly, we only ensure that adequate drug amounts reach their therapeutic targets by optimizing pharmacokinetics through pharmacogenetics. However, that does not ensure therapeutic outcomes that will be dependent on pharmacodynamics.

Limitations of our knowledge of pharmacodynamic targets hamper the full translation of pharmacogenetics to clinical settings. As illustrative examples based on information currently provided at their websites, Medicare-funded Castle's IDgenetix Pharmacogenetic Testing supports psychiatric pharmacology practice by covering 15 genes, as follows: on the pharmacokinetic side: *ABCB1*, *CYP1A2*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A4*, and *CYP3A5*. In contrast, on the pharmacodynamic side, the genes tested are *ADRA2A*, *COMT*, *HTR2A*, *HTR2C*, *MTHFR*, *NAT2*, *OPRM1*, and *SLC6A4*. Myriad's GeneSight covers variants across 8 genes (pharmacokinetics: *CYP1A2*, *CYP2C9*, *CYP2C19*, *CYP3A4*, *CYP2B6*, and *CYP2D6*, and pharmacodynamics: *HTR2A* and *SLC6A4*). The Invitae Mental Health Pharmacogenomics Panel tests for variants in the following 15 genes, seven of them encoding for CYP enzymes, namely *CYP1A2, CYP2B6, CYP2C19, CYP2C9, CYP2D6, CYP3A4, CYP3A5, COMT, GRIK4, HLA-A, HLA-B, HTR2A, HTR2C, MTHFR,* and *UGT2B15*.

Commercially Available Pharmacogenetic Testing: Significant Translational Limitations

Before commercially available tests can predict actual treatment responses, much more knowledge is needed on the mechanisms of action and related pharmacogenetics of drugs commonly used to treat psychiatric disorders. As described by Hicks et al. (2020), in 2018, the Food and Drug Administration (FDA) cautioned that there might be a lack of clinical evidence supporting the utility of clinical pharmacogenetic testing. Moreover, the FDA also issued a warning letter to a laboratory regarding the marketing and gene content of their clinical pharmacogenetic tests. However, in a press release of February 20, 2020, entitled "FDA Announces Collaborative Review of Scientific Evidence to Support Associations Between Genetic Information," the FDA recognizes that "genetic testing can enhance patient management by improving the selection of medications or the dosage of a medication, tailoring the therapy to the patient—but only when it is based on sound science. That is why [they introduced] a collaboration between the FDA's Center for Devices and Radiological Health and Center for Drug Evaluation and Research intended to provide the agency's view of the state of the current science in pharmacogenetics." This "new web-based resource includes a table that describes some of the gene–drug interactions for which we believe there is sufficient scientific evidence to support the described associations between certain genetic variants, or genetic variant-inferred phenotypes and altered drug metabolism, and in certain cases, differential therapeutic effects, including differences in risks of adverse events." As promised in that statement, the FDA now maintains and constantly updates a Table of Pharmacogenetic Associations—for reference, see Rubinstein and Pacanowski, 2021. In this chapter, Table 1.21–1 summarizes the large FDA table focused only on drugs relevant to psychiatry and neurology. It can be seen in Table 1.21–1 that all genes highlighted are involved only in pharmacokinetics or in ADRs. No pharmacodynamic targets are identified in that table.

It is self-evident that the mechanisms of action of antidepressants, for example, will, at the pharmacodynamic level, involve many more genes and their variants than those tested by current commercially available pharmacogenetic tests. Therefore, that type of testing may be helpful for dose adjustment (pharmacokinetics); however, a high percentage of patients will not respond well to a given drug (pharmacodynamics), despite the dosage.

In a 2020 editorial Smith and Nemeroff stated that "the field of PGx [pharmacogenetics] is advancing quickly, but evidence of its utility is meager. Clearly, insufficient evidence supports the universal use of PGx testing to guide antidepressant treatment decisions. Of course, gene structure is only one piece of the puzzle that contributes to the inter-individuality of antidepressant response. Personal or environmental variables, such as concomitant medications, ethnicity, gender, smoking status, and comorbidities, are also important."

A fundamental limitation for the clinical applicability of pharmacogenetics in psychiatry is the insufficient availability of rigorously peer-reviewed, prospective, multicenter studies on the effectiveness of psychiatric treatment that are federally or foundation funded, and therefore without pharmaceutical or testing company bias. For example, over a 20-year stretch, there were no new drugs for treating acute lymphoblastic leukemia (ALL) in children. However, clinical outcomes continued to improve substantially during that period due to the better use of existing drugs, guided by rigorous multicenter studies. Those types of extensive, rigorous, prospective, multicenter, peer-reviewed studies, possibly with pharmacogenetics as an integral part of the study design, have been carried out in various therapeutic areas such as cancer, heart disease, diabetes, and neurologic diseases. However, they have not been conducted as needed for psychiatric disorders. While STAR*D, cited above, was an important step forward, it was a contract by the National Institutes of Health (NIH), not a peer-reviewed multicenter R01 or similar research grant study. A significant limitation of STAR*D was that pharmacogenetics was not part of the study design and was included later as an afterthought. The study was already underway when it was added, and blood was collected for that purpose. Participants who dropped out early on for multiple reasons, including pharmacogenetics related to ADRs, were lost, and their DNA was not sequenced. There are significant limitations to the growth of evidence-based pharmacogenetic testing in psychiatry in the absence of extensive (and costly), national, peer-reviewed, multicenter, rigorously controlled clinical research studies to assess the prediction of clinical outcomes prospectively.

Table 1.21–1.

Representative Gene–Drug Combinations Relevant to Psychiatry and Neurology Extracted from the FDA Table of Pharmacogenetic Associations in February 2023

Section 1: Pharmacogenetic Associations for Which the Data Support Therapeutic Management Recommendations

Drug	Gene	Affected Subgroups+	Description of Gene–Drug Interaction
Amphetamine	CYP2D6	Poor metabolizers	May affect systemic concentrations and adverse reaction risk. Consider lower starting dosage or use alternative agent.
Aripiprazole	CYP2D6	Poor metabolizers	Results in higher systemic concentrations and higher adverse reaction risk. Dosage adjustment is recommended. Refer to FDA labeling for specific dosing recommendations.
Aripiprazole lauroxil	CYP2D6	Poor metabolizers	Results in higher systemic concentrations. Dosage adjustment is recommended. Refer to FDA labeling for specific dosing recommendations.
Atomoxetine	CYP2D6	Poor metabolizers	Results in higher systemic concentrations and higher adverse reaction risk. Adjust titration interval and increase dosage if tolerated. Refer to FDA labeling for specific dosing recommendations.
Carbamazepine	HLA-B	*15:02 allele positive	Results in higher adverse reaction risk (severe skin reactions). Avoid use unless potential benefits outweigh risks and consider risks of alternative therapies. Patients positive for HLA-B*15:02 may be at increased risk of severe skin reactions with other drugs that are associated with a risk of Stevens–Johnson Syndrome/Toxic Epidermal necrolysis (SJS/TEN). Genotyping is not a substitute for clinical vigilance.
Citalopram	CYP2C19	Poor metabolizers	Results in higher systemic concentrations and adverse reaction risk (QT prolongation). The maximum recommended dose is 20 mg.
Phenytoin	CYP2C9	Intermediate or poor	May result in higher systemic concentrations and higher adverse

		metabolizers	reaction risk (central nervous system toxicity). Refer to FDA labeling for specific dosing recommendations. Carriers of CYP2C9*3 alleles may be at increased risk of severe cutaneous adverse reactions. Consider avoiding phenytoin as an alternative to carbamazepine in patients who are CYP2C9*3 carriers. Genotyping is not a substitute for clinical vigilance and patient management.
Phenytoin	HLA-B	*15:02 allele positive	May result in higher adverse reaction risk (severe cutaneous reactions). Patients positive for HLA-B*15:02 may be at increased risk of Stevens-Johnson Syndrome/Toxic Epidermal necrolysis (SJS/TEN). Consider avoiding phenytoin as an alternative to carbamazepine in patients who are positive for HLA-B*15:02. Genotyping is not a substitute for clinical vigilance and patient management.
Thioridazine	CYP2D6	Poor metabolizers	Results in higher systemic concentrations and higher adverse reaction risk (QT prolongation). Predicted effect based on experience with CYP2D6 inhibitors. Contraindicated in poor metabolizers.
Venlafaxine	CYP2D6	Poor metabolizers	Alters systemic parent drug and metabolite concentrations. Consider dosage reductions.
Vortioxetine	CYP2D6	Poor metabolizers	Results in higher systemic concentrations. The maximum recommended dose is 10 mg.

Section 2: Pharmacogenetic Associations for which the Data Indicate a Potential Impact on Safety or Response

Drug	Gene	Affected Subgroups+	Description of Gene–Drug Interaction
Carbamazepine	HLA-A	*31:01 allele positive	Results in higher adverse reaction risk (severe skin reactions). Consider risk and benefit of carbamazepine use in

			patients positive for HLA-A*31:01. Genotyping is not a substitute for clinical vigilance.
Oxcarbazepine	HLA-B	*15:02 allele positive	Results in higher adverse reaction risk (severe skin reactions). Patients positive for HLA-B*15:02 may be at increased risk of severe skin reactions with other drugs that are associated with a risk of Stevens–Johnson Syndrome/Toxic Epidermal necrolysis (SJS/TEN). Genotyping is not a substitute for clinical vigilance.
Perphenazine	CYP2D6	Poor metabolizers	Results in higher systemic concentrations and higher adverse reaction risk.

Section 3: Pharmacogenetic Associations for which the Data Demonstrate a Potential Impact on Pharmacokinetic Properties Only

Drug	Gene	Affected Subgroups+	Description of Gene–Drug Interaction
Amitriptyline	CYP2D6	Ultrarapid, intermediate, or poor metabolizers	May alter systemic concentrations.
Amoxapine	CYP2D6	Ultrarapid, intermediate, or poor metabolizers	May alter systemic concentrations.
Clomipramine	CYP2D6	Ultrarapid, intermediate, or poor metabolizers	May alter systemic concentrations.
Desipramine	CYP2D6	Ultrarapid, intermediate, or poor metabolizers	May alter systemic concentrations.
Diazepam	CYP2C19	Poor metabolizers	May affect systemic concentrations.
Donepezil	CYP2D6	Ultrarapid or poor metabolizers	Alters systemic concentrations.

Doxepin	CYP2C19	Intermediate or poor metabolizers	Results in higher systemic concentrations.
Doxepin	CYP2D6	Ultrarapid, intermediate, or poor metabolizers	May alter systemic concentrations.
Escitalopram	CYP2C19	Ultrarapid, intermediate, or poor metabolizers	May alter systemic concentrations.
Fluvoxamine	CYP2D6	Poor metabolizers	Results in higher systemic concentrations. Use with caution.
Galantamine	CYP2D6	Poor metabolizers	Results in higher systemic concentrations. Titrate dosage based on tolerability.
Imipramine	CYP2D6	Ultrarapid, intermediate, or poor metabolizers	May alter systemic concentrations.
Nortriptyline	CYP2D6	Ultrarapid, intermediate, or poor metabolizers	May alter systemic concentrations.
Paroxetine	CYP2D6	Ultrarapid, intermediate, or poor metabolizers	May alter systemic concentrations.
Propranolol	CYP2D6	Poor metabolizers	May affect systemic concentrations.
Protriptyline	CYP2D6	Poor metabolizers	Results in higher systemic concentrations.
Risperidone	CYP2D6	Poor metabolizers	Alters systemic parent drug and metabolite concentrations.
Trimipramine	CYP2D6	Ultrarapid, intermediate, or poor metabolizers	May alter systemic concentrations.
Viloxazine	CYP2D6	Poor metabolizers	May result in higher systemic concentrations.

Note: This table includes select examples and does not represent all gene–drug combinations of potential interest to family physicians and may not include all interactions within the drug class. For a complete list that is periodically updated, go to https://www.fda.gov/medical-devices/precision-medicine/table-pharmacogenetic-associations. Table appears as published by the FDA.

FDA, U.S. Food and Drug Administration; INR, international normalized ratio; NAT, arylamine N-acetyltransferases.

[+] The table describes gene–drug interactions and indicates specific affected subgroup(s) to which the interaction applies. The affected subgroup(s) may be carriers of a specific genetic variant (e.g., *HLA-B**15:02) or a genotype-inferred phenotype, ultrarapid, normal, intermediate, or poor metabolizers/function transporters of a drug-metabolizing enzyme/drug transporter. Normal metabolizers or normal transporters do not have genetic variants that are expected to impact metabolism or transport function. In general, ultrarapid metabolizers have two or more copies of a genetic variant that increases metabolic function; intermediate metabolizers or reduced function transporters are individuals who have one or two copies of a genetic variant that reduces the ability to metabolize or transport a drug; and poor metabolizers or poor function transporters are individuals who generally have two copies of a genetic variant that results in little to no ability to metabolize or transport a drug.

From Food and Drug Administration. Table of pharmacogenetic associations. 2022. https://www.fda.gov/medical-devices/precision-medicine/table-pharmacogenetic-associations.

While the promise of pharmacogenetics-driven personalized and precision medicine is exciting and appealing to patients and physicians, our current scientific knowledge is still insufficient to predict, based on pharmacogenetic testing alone, who will respond particularly well to specific psychotropic medications, despite the marketing claims of various testing companies.

CONCLUSIONS

Those who practice medicine know that patients respond differently to identical treatments. In general, the best predictor of drug response is the patient’s personal or familial treatment response history, which suggests an inherited, hence genetic, basis for treatment response. However, that type of information is only sometimes available. The clinician often observes that certain patients following the same dose of the same drugs will show complete symptoms remission. In contrast, others may report only severe side effects or no response. Physicians need to balance a patient’s health with adverse drug interactions when prescribing various antidepressants; however, this may fail to predict drug response.

Work by our group and others on the translational and clinical pharmacogenomics of psychiatric disorders has been focused on identifying genetic markers as predictors of positive and negative treatment outcomes. Rigorous clinical studies must continue to be performed to closely examine the relation between genotype and the drug response phenotype. In well-conducted clinical trials, positive clinical responses and adverse events can be related to specific genetic polymorphisms. Such research is designed to identify markers significantly linked with treatment responses. This line of investigation offers enormous promise for the individualization of drug therapy.

It would be highly beneficial to public health if the likelihood of favorable or adverse responses were known before administering a drug. Despite the potential for enormous returns, a myriad of problems emerges in conducting such work. Problematic clinical factors include the placebo response, sample size issues, genetic background, ethnic stratification, continual versus categorical outcome measures, drug choice and treatment strategy, treatment regime adherence, and the impact of the environment on treatment outcome. The dilemmas raised by this line of work's genetic components are also substantial and discussed elsewhere. A significant difficulty is choosing the appropriate genetic polymorphisms associated with drug response. A statistical conundrum is created by the need to assess multiple variables that are partially related to one another (without a priori knowledge of the exact nature of such interactions) and that contribute to minor effects in sample sizes that require standardized clinical research treatment and that therefore are highly costly and often cannot be as large as desired. Despite these limitations, pharmacogenetic testing has been translated from the bench to the bedside, with utility for tailoring specific treatments, such as those for major depression.

Neither the biology of psychiatric disorders nor the mechanisms of action of psychotropic drugs are entirely known to date, and knowledge continues to accumulate progressively. The pharmacogenetic perspective contributes considerably to developing customized treatments and presents a good starting point for understanding the pathophysiologic mechanisms underlying psychiatric disorders. All pharmacogenetic studies are based on the neurobiologic research theories for these disorders.

Pharmacogenomics is a new area of medical science based on novel databases that have been emerging since the completion of the Human Genome Project. High throughput technologies have been developed to enable researchers to harness the overwhelming amount of genomic data to improve therapeutics. In addition, genomics now gives us the tools to search for new biomarkers and therapeutic targets by providing new avenues of treatment optimization and research. Finally, applying advanced molecular biology techniques and a complete understanding of human genomic sequences relevant to treatment responses will facilitate searching for conceptually novel therapeutic target sites for psychiatric disorders.

FURTHER READINGS

Aklillu E, Persson I, Bertilsson L, Johansson I, Rodrigues F, Ingelman-Sundberg M. Frequent distribution of ultrarapid metabolizers of debrisoquine in an ethiopian population carrying duplicated and multiduplicated functional CYP2D6 alleles. *J Pharmacol Exp Ther*. 1996;278:441–446.

Bertilsson L, Dahl ML, Dalén P, Al-Shurbaji A. Molecular genetics of CYP2D6: clinical relevance with focus on psychotropic drugs. *Br J Clin Pharmacol*. 2002;53:111–122.

Bunka M, Wong G, Kim D, et al. Evaluating treatment outcomes in pharmacogenomic-guided care for major depression: a rapid review and meta-analysis. *Psychiatry Res*. 2023;321:115102.

Campos AI, Byrne EM, Mitchell BL, et al. Impact of CYP2C19 metaboliser status on SSRI response: a retrospective study of 9500 participants of the Australian Genetics of Depression Study. *Pharmacogenomics J*. 2022;22:130–135.

Cuéllar-Barboza AB, McElroy SL, Veldic M, et al. Potential pharmacogenomic targets in bipolar disorder: considerations for current testing and the development of decision support tools to individualize treatment selection. *Int J Bipolar Disord*. 2020;8:23.

Edenberg HJ. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res Health*. 2007;30:5–13.

Flores DL, Alvarado I, Wong ML, Licinio J, Flockhart D. Clinical implications of genetic polymorphism of CYP2D6 in Mexican Americans. *Ann Intern Med*. 2004;140:W71.

Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Klein T, Leeder JS. Prediction of CYP2D6 phenotype from genotype across world populations. *Genet Med*. 2017;19:69–76.

Goodman DW. Lisdexamfetamine dimesylate (vyvanse), a prodrug stimulant for attention-deficit/hyperactivity disorder. *P T*. 2010;35:273–287.

Gray K, Adhikary SD, Janicki P. Pharmacogenomics of analgesics in anesthesia practice: a current update of literature. *J Anaesthesiol Clin Pharmacol*. 2018;34:155–160.

Howland RH. Sequenced treatment alternatives to relieve depression (STAR*D). Part 2: study outcomes. *J Psychosoc Nurs Ment Health Serv*. 2008;46:21–24.

Kirchheiner J, Nickchen K, Bauer M, et al. Pharmacogenetics of antidepressants and antipsychotics: the contribution of allelic variations to the phenotype of drug response. *Mol Psychiatry*. 2004;9:442–473.

Koopmans AB, Braakman MH, Vinkers DJ, Hoek HW, van Harten PN. Meta-analysis of probability estimates of worldwide variation of CYP2D6 and CYP2C19. *Transl Psychiatry*. 2021;11:141.

Krishnan S, Zhang Y. Relative bioavailability of lisdexamfetamine 70-mg capsules in fasted and fed healthy adult volunteers and in solution: a single-dose, crossover pharmacokinetic study. *J Clin Pharmacol*. 2008;48:293–302.

Lett TAP, Wallace TJM, Chowdhury NI, Tiwari AK, Kennedy JL, Müller DJ. Pharmacogenetics of antipsychotic-induced weight gain: review and clinical implications. *Mol Psychiatry*. 2012;17:242–266.

Li Z, Huang M, Ichikawa J, Dai J, Meltzer HY. N-desmethylclozapine, a major metabolite of clozapine, increases cortical acetylcholine and dopamine release in vivo via stimulation of M1 muscarinic receptors. *Neuropsychopharmacology*. 2005;30:1986–1995.

Licinio J, Wong ML. *Pharmacogenomics: The Search for Individualized Therapies*. John Wiley & Sons; 2009.

Licinio J, Wong ML. Pharmacogenomics of antidepressant treatment effects. *Dialogues Clin Neurosci*. 2011;13:63–71.

Llerena A, Naranjo MEG, Rodrigues-Soares F, Penas-LLedó EM, Fariñas H, Tarazona-Santos E. Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations. *Expert Opin Drug Metab Toxicol*. 2014;10:1569–1583.

Lynch T, Price A. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. *Am Fam Physician*. 2007;76:391–396.

McLellan RA, Oscarson M, Seidegård J, Evans DA, Ingelman-Sundberg M. Frequent occurrence of CYP2D6 gene duplication in Saudi Arabians. *Pharmacogenetics*. 1997;7:187–191.

Meshkat S, Rodrigues NB, Di Vincenzo JD, et al. Pharmacogenomics of ketamine: a systematic review. *J Psychiatr Res*. 2021;145:27–34.

Milosavljevic F, Bukvic N, Pavlovic Z, et al. Association of CYP2C19 and CYP2D6 poor and intermediate metabolizer status with antidepressant and antipsychotic exposure: a systematic review and meta-analysis. *JAMA psychiatry*. 2021;78:270–280.

Patrick KS, Radke JL, Raymond JR, et al. Drug regimen individualization for attention-deficit/hyperactivity disorder: guidance for methylphenidate and dexmethylphenidate formulations. *Pharmacotherapy*. 2019;39:677–688.

Racoosin JA, Roberson DW, Pacanowski MA, Nielsen DR. New Evidence about an old drug–risk with codeine after adenotonsillectomy. *N Engl J Med*. 368;23:2155–2157.

Rubinstein WS, Pacanowski M. Pharmacogenetic gene-drug associations: FDA perspective on what physicians need to know. *Am Fam Physician*. 2021;104:16–19.

Smith S, Woodside B. Is there a role for pharmacogenetics in the treatment of anorexia nervosa? *Pharmacogenomics*. 2016;17:1381–1383.

Smith TL, Nemeroff CB. Pharmacogenomic testing and antidepressant response: problems and promises. *Braz J Psychiatry*. 2020;42:116–117.

Takahashi PY, Ryu E, Pathak J, et al. Increased risk of hospitalization for ultrarapid metabolizers of cytochrome P450 2D6. *Pharmgenomics Pers Med*. 2017;10:39–47.

Taylor C, Crosby I, Yip V, Maguire P, Pirmohamed M, Turner RM. A review of the important role of CYP2D6 in pharmacogenomics. *Genes (Basel)*. 2020;11.

Vecera CM, Fries GR, Shahani LR, Soares JC, Machado-Vieira R. Pharmacogenomics of lithium response in bipolar disorder. *Pharmaceuticals (Basel)*. 2021;14.

Wong ML, Arcos-Burgos M, Liu S, et al. Rare functional variants associated with antidepressant remission in Mexican-Americans: short title: antidepressant remission and pharmacogenetics in Mexican-Americans. *J Affect Disord*. 2021;279:491–500.

Wong ML, Licinio J. Research and treatment approaches to depression. *Nat Rev Neurosci*. 2001;2:343–351.

Zai G. Pharmacogenetics of obsessive-compulsive disorder: an evidence-update. *Curr Top Behav Neurosci*. 2021;49:385–398.

Zanger UM. Genetic variability of CYP2D6: basic and clinical aspects. In: Baumann P, ed. *CYP2D6: Genetics, Pharmacology and Clinical Relevance*. Future Medicine Ltd; 2014:28–41. https://www.futuremedicine.com/doi/abs/10.2217/fmeb2013.13.270

1.22 Animal Models in Psychiatry

JANNA K. MOEN, B.A., PH.D., AND PAUL J. KENNY, PH.D.

THE USE OF ANIMAL MODELS IN PSYCHIATRIC RESEARCH

Animal research allows experimenters to manipulate and measure behavioral and biologic processes that are relevant to psychiatry. Animal research in psychiatry falls into two broad categories. First, basic research aims to improve our understanding of psychological and neural processes in their species-specific context using established and novel technical approaches. Second, in applied research, animal experiments are used to model specific aspects of human psychiatric conditions, either to help identify novel targets for therapeutic drug development, or to provide a preclinical test of the effectiveness of a compound on behavior considered relevant to a particular disorder. As the usefulness of the latter type of research depends on the degree to which findings can be translated between species, it is constrained by how effectively aspects of psychiatric conditions can be modeled in animals.

There are clearly substantial differences between humans and nonhuman primates (NHPs) and rodents, which are the species most often used in animal research. Abstract thought in humans allows for self-regulation while self-reflection can lead to rumination, which may play a role in human susceptibility to depression or anxiety. Language allows humans to engage in uniquely complex communication, and insights gained during therapy may be instrumental in coping with psychiatric disorders. Biologic differences probably underlie these differences between humans and other species, including a more developed cortex and prefrontal cortex compared to rodents and even NHPs. In addition, although the vast majority of human brain regions and genes have homologous versions in NHPs and rodents, there are substantial differences at the level of cell structure, connectivity, and gene regulation. These differences notwithstanding, there are many similarities between humans and experimental animals in both functional and biologic terms. Thus, basic psychological processes such as stress, fear, anxiety, motivation, attention, and reward are present in species-specific manners in rodents and NHPs. In addition, subcortical structures are quite well conserved, with a common brain design and substantial homologies for most neural structures.

Table 1.22–1.
Implication of Class of Symptoms in Major Psychiatric Conditions

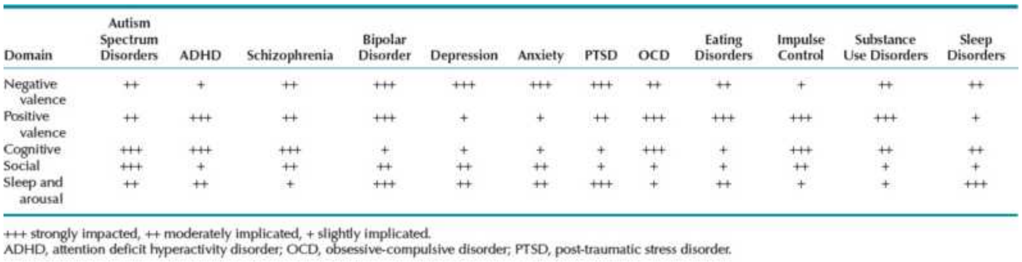

Domain	Autism Spectrum Disorders	ADHD	Schizophrenia	Bipolar Disorder	Depression	Anxiety	PTSD	OCD	Eating Disorders	Impulse Control	Substance Use Disorders	Sleep Disorders
Negative valence	++	+	++	+++	+++	+++	+++	++	++	+	++	++
Positive valence	++	+++	++	+++	+	+	++	+++	+++	+++	+++	+
Cognitive	+++	+++	+++	+	+	+	+	+++	+	+++	++	++
Social	+++	+	++	++	++	++	+	+	+	++	+	+
Sleep and arousal	++	++	+	+++	++	++	+++	+	++	+	+	+++

+++ strongly impacted, ++ moderately implicated, + slightly implicated.
ADHD, attention deficit hyperactivity disorder; OCD, obsessive-compulsive disorder; PTSD, post-traumatic stress disorder.

Based on these considerations, and on the limited success in using animal models to develop new therapeutics for psychiatric disorders, it has been suggested that animals should be used to model particular symptoms instead of entire human psychiatric disorders. In particular, productive animal research in psychiatry should focus on well-defined phenotypes that have a known biologic basis which can be measured to facilitate experimentation, and preferentially focus on brain substrates that have reasonable homologies between humans and the animal model of choice. In addition, when genetic models are used in animals, it is useful to focus on highly penetrant genetic variants to ensure relevance to human disease. In a general sense, the usefulness of animal models in modeling aspects of human psychiatric disorders can be assessed by three criteria:

- *Face validity* refers to the extent to which a behavior that is studied in animals resembles the behavior in humans it is intended to model.
- *Predictive validity* refers to the extent to which manipulations in animal models predict clinical effectiveness in psychiatric patients, and an ideal model would have low false-positive and false-negative rates.
- *Construct validity* refers to the extent to which an animal model reproduces the etiology of the human psychiatric disorder.

Following recent guidelines from the National Institute of Mental Health, this chapter is organized by a description of available animal models in five broad classes of symptoms, which may be observed across psychiatric disorders (Table 1.22–1).

NEGATIVE VALENCE

In psychology, “valence” refers to the intrinsic value of an internal feeling or emotional state. In psychiatric disorders, “negative valence” typically indicates a level of aversiveness or discomfort. Behaviors associated with negative valence that can be studied in animal models include the following:

Defensive Behaviors

Ethologic studies of defensive behavior in rodents have shown that threatening situations evoke defensive behaviors that depend on the proximity of the threat. When a threat is very near and acute, panic responses consisting of flight or defensive attack are elicited, while fear responses consisting of concealment and freezing occur when a threat is present but more distal. Finally, potential threats elicit lower-intensity anxiety-like behavior, with risk assessment and reduction of exploratory behaviors for extended periods of time. In general, animal models of defensive behavior are focused on acute normal defensive responses, while only a few models involve chronic stress manipulations.

Panic-Like Behavior. Panic is often defined as a sudden and sometimes uncontrollable episode of intense fear or anxiety provoked by a real or perceived threat. Panic responses can be studied in rats and mice in its ethologic context in defensive tests that use a predator (human, rat, or in some recent studies, robotic) closing in on subjects in a circular runway. In this test, called the defense test battery, the predator approaches the subject on a long oval runway, and flight responses can be quantified. If the runway is blocked, the imminent approach of the predator elicits flight or defensive threat and attack responses. In addition to good face validity, pharmacologic agents that are clinically effective at reducing panic attacks, such as benzodiazepines, have been shown to reduce flight in this model, while anxiolytic agents that are less effective in reducing panic attacks do not have a strong effect on flight, suggesting good predictive validity. However, the main drawback of this model is its relative complexity, and it has not been widely adopted to date. Panic responses can also be evoked artificially in rats by electrical stimulation of the dorsal periaqueductal grey matter (dPAG). As the dPAG harbors a hypoxia-sensitive alarm system, this model is most relevant to the putative respiratory subtype of panic disorder.

Fear-Like Behavior. Fear can be broadly defined as a noxious behavioral state provoked by exposure to potentially threatening stimuli. Fear responses are typically studied in the context of an aversive stimuli being presented at random (unconditioned) or following association with a predictive cue (conditioned). Unconditioned fear responses to predators can also be measured in the defense battery tests for rats or mice. In both cases, freezing behavior is observed in the blocked runway when the predator approaches up to a certain point. Further approach is followed by active defensive attack. Conditioned fear responses have traditionally been studied to elucidate the neural basis of learning and memory but are increasingly used to study normal and pathologic defensive behaviors. In the basic fear conditioning setup, experimenters consistently present a conditioned stimulus (typically a tone or a light) before an aversive unconditioned stimulus (typically a mild foot shock). After repeated pairings, the subject will be able to predict presentation of the aversive stimulus when the conditioned stimulus is presented. Conditioned fear can be measured in the fear-potentiated startle paradigm, where presentation of the conditioned stimulus potentiates the locomotor response to a loud startle tone. This procedure can be used in both rodents and humans with minor adaptations. Alternatively, conditioned fear is often measured in rodents in enclosed environments as a freezing response. Although both tests can be confounded by manipulations that affect locomotion, when a pharmacologic manipulation reduces fear-potentiated startle and increases locomotion by reducing freezing, the results can be interpreted as fear reducing with confidence.

Generalization of conditioned fear is a learning phenomenon that refers to the transfer of conditioned fear responses from the proper conditioned stimulus that predicts an aversive outcome to stimuli that are related, but distinct. Theoretically, putative abnormal generalization of conditioned fear to related stimuli is an attractive explanation for the pervasiveness of defensive responses in panic disorder and post-traumatic stress disorder. Operationally, rodents, NHPs, or humans can be tested for generalization of conditioned fear by measuring fear responses following the presentation of stimuli similar to but distinct from the original conditioned stimulus.

Extinction of conditioned fear is another learning phenomenon that has clear relevance to clinical psychiatry. In this procedure, after fear conditioning in rodents, NHPs, or humans, the conditioned stimulus is presented repeatedly in the absence of the unconditioned stimulus. Over time, subjects will gradually reduce, or extinguish, their conditioned fear responses until the stimuli no longer elicits an aversive response. This procedure has a clear parallel in exposure therapy that is used to successfully treat generalized anxiety disorder, specific phobias, and post-traumatic stress disorder. Pharmacologic treatments that expedite development of extinction in animals can enhance the beneficial effect of exposure therapy in human patients. The drug D-cycloserine has been shown to improve outcomes of exposure-based therapy in the treatment of anxiety disorders and phobias, but may not be an effective option for treating post-traumatic stress disorder.

While fear is often discussed as a different emotional state from anxiety distinguished by the presence or absence of an active threat, respectively, there is considerable debate among behavioral neuroscientists regarding whether fear and anxiety involve different neurobiologic mechanisms or instead encompass different aspects of a shared internal state. For example, stress-induced fear learning in rodents elicits anxiety-like responses in the elevated plus maze in addition to conditioned fear responses (i.e., freezing) following extensive pairing of foot shock with conditioned stimuli, indicating that the anxiety response in this model serves to increase fear behaviors. Further study is necessary to distinguish the difference between fear and anxiety.

Anxiety-Like Behavior. Anxiety is typically considered to be a more generalized emotional state than fear, with anxiety corresponding to intense uneasiness or apprehension regarding something with an uncertain outcome in the absence of an imminent threat. The most common way to measure anxiety-like behavior in rodents is based on approach-avoidance conflicts. Rats and mice are naturally cautious in open, elevated, and brightly lit environments because of the risk of predation, and the avoidance of such places can be a measure of anxiety-like behavior. Typically, animals are allowed to explore an apparatus with two parts: one enclosed or dark compartment, and a second more open or brightly lit or elevated compartment. The elevated plus maze, light–dark box, and open field tests are the most popular tests in this category. In addition, the elevated plus maze has been adapted to measure stretch-attend postures, reflecting risk assessment. There is also a human version of the open field test, called the human behavioral pattern monitor, which combines free exploration of a novel environment with presentation of novel objects, thus allowing for translational studies of anxiety-like behavior. Tests of approach-avoidance behavior have been criticized as a test of anxiety-like behavior, as drugs may affect the drive to explore independently of anxiety, which could be misinterpreted as an anxiolytic effect. In addition, these behavioral tests were developed primarily to measure the anxiolytic effect of benzodiazepine drugs, and might therefore not be very effective at detecting novel anxiolytics that work through alternative mechanisms. Indeed, selective serotonin reuptake inhibitor (SSRIs) that are clinically effective at reducing anxiety in human patients generally do not have strong anxiolytic effects in approach-avoidance tests.

A second approach to study anxiety-like behavior is through conflict tests, in which food- or water-restricted animals are punished with a mild electric shock for lever pressing for food (Geller–Seifter conflict test) or licking a spout for water (Vogel conflict test). The resulting suppression of appetitive behavior is interpreted as anxiety-like behavior and is sensitive to anxiolytic drugs.

A third approach for studying anxiety-like behavior involves measuring responses to novel objects. A classic example is the marble-burying test, in which rodents are placed in a cage with plentiful bedding and unfamiliar marbles. Rodents will typically bury these novel objects in bedding, which can be interpreted as a defensive anxiety-like behavior. Importantly, the marble-burying test is controversial in the field of behavioral neuropharmacology. It is difficult to ascribe a rodent's anxiety toward an object like a novel marble, and drugs that are effective at reducing marble burying may instead serve to increase freezing or object avoidance, which can also be interpreted as anxiety-like responses. A model with more face validity is the shock probe burying test for rats, in which animals are confronted with an electrified probe that delivers aversive shocks on contact. Defensive burying of the probe with bedding material serves as the measure of anxiety-like effects.

Finally, there are tests for anxiety-like responses that do not measure anxiety-like responses as behavioral inhibition, as general sedative effects might be mistaken for anxiolytic effects in the tests described above. In the stress-induced hyperthermia test, animals are stressed with rectal placement of a thermometer, and the gradual increase in body temperature is measured as an index of the effect of stress. In addition, light-enhanced startle is a test that uses a continuous bright light in the startle chamber, which gradually potentiates the startle response. As all of the acute models described in this section have both benefits and pitfalls, the general consensus when testing potential anxiolytic drugs is to perform a battery of behavioral tests before drawing conclusions.

Social Anxiety-Like Behavior. Several tests of social anxiety have been developed in rodents and NHPs. In the social interaction test for rats and mice, two animals are either allowed to interact freely, or through a wire bar enclosure. Time spent interacting with this conspecific (animal of the same species) is quantified, and treatments that increase interaction time are considered anxiolytic. In addition, ultrasonic distress vocalizations, particularly in rats, are social signals that serve to indicate distress. Such signals in pups separated from mothers, or in mature animals in distressing conditions, can be taken as a measure of anxiety.

Although the most common animal models of anxiety are based on rat or mouse studies, the human intruder test for primates is well validated and offers particularly good face and construct validity as a model of human social anxiety. In this test, confined macaque or marmoset primates are confronted with the presence of a human who does not look at them directly. The mere presence of the human intruder causes behavioral inhibition, aggression, and displacement stress responses such as scratching and autogrooming. The behavior of animals in this test has been shown to be sensitive to clinically effective anxiolytic drugs and correlates well with fear responses in other paradigms, but can vary significantly between facilities and between different human intruders. A more recent primate model of anxiety and fearful temperament has been developed in which NHPs are rewarded for touching visual stimuli on a touch screen device. During test trials, the face of an unknown male conspecific with direct gaze is presented on the touch screen, and behavioral inhibition can be observed in subjects who hesitate or fail to touch the screen for a juice reward. This touch screen paradigm may be less sensitive to facility and experimenter variability, and offers a more ethologic approach compared with the human intruder test.

Depressive Behaviors

Reduced Effort to Escape Aversive Stimuli. Hopelessness is critical symptom of depression, and the rodent forced swim test is one of the most widely used acute behavioral screens for antidepressant activity. In this test, rodents are placed in a cylinder filled with water. While the animals will initially try to escape the cylinder with vigorous escape movements, eventually the animals will show immobility, which is interpreted as learned helplessness. Antidepressant effects are inferred by an increase in the time spent in escape movements. Importantly, immobility time in the forced swim test may be influenced by other factors, such as adaptive floating for energy conservation. As such, these acute models are often performed as part of a battery of behavioral tests to further characterize a putative antidepressant-like response. A related task for mice is the tail-suspension test, in which mice are hung by the end of their tails and a similar transition from active escape movements to immobility occurs. Various classes of clinically effective antidepressants produce acute reductions in immobility in these tests, suggesting good predictive validity. In addition, lower doses of antidepressants show an effect after chronic but not acute treatment, in line with the delayed effectiveness of antidepressants for humans. In the learned helplessness paradigm, uncontrollable stress can lead to reductions in escape attempts in subsequent situations where the stressor can actually be evaded. Such escape deficits are also reversible by antidepressant drugs. Notably, these tests have been developed explicitly to detect the effects of monoamine-targeted antidepressants, although recent studies have shown other drugs like psilocybin and ketamine can produce similar antidepressant-like behavioral phenotypes in a rat model.

Reduced Effort to Obtain Rewards. Anhedonia, or the loss of interest in otherwise pleasurable activities, is also a key symptom of depression which can be modeled in rodents. Preference for palatable sucrose over water, or the maximum number of lever presses an animal is willing to expand to obtain food in the progressive ratio test, can both provide measures of anhedonia. In addition, chronic implantation of intracranial self-stimulating electrodes in reward centers of the brain can allow for repeated measures of the threshold current at which animals are willing to self-stimulate, thus allowing for a more robust measure of reward function.

Chronic Stress Models. Finally, antidepressant activity can also be measured in rodent tests that use chronic stress manipulations. In the chronic mild stress paradigm, experimenters administer various mild unpredictable stressors for prolonged periods. Common stressors include restraint, damp bedding material, and disrupted light–dark cycles. The effects of chronic stress can be measured as reduced sucrose preference and reduced grooming behavior in mice. Although the chronic mild stress model has good face and predictive validity, it has received criticism for its high level of variability.

In the chronic social defeat model, subject mice are introduced in the home environment of a large aggressive mouse each day for 10 days. The sequential defeat experience leads to a relative avoidance of a novel nonaggressive mouse in the social interaction test, as well as reduced sucrose preference. The study of susceptible versus resilient animals in the chronic social defeat stress model has yielded interesting insights into the molecular biology underlying this behavior. An important drawback of the traditional chronic social defeat model is that the test can only be performed using male mice, as female mice typically do not display or elicit the same aggression when confronted with a novel conspecific. As women have higher rates of depression and distinct transcriptomic profiles when depressed compared with men, it is necessary to adapt this model for use in female animal subjects. In one adaptation, male odorants are applied to female subjects who are then confronted with a male aggressor. Another model involves housing intact female subjects with castrated males. These females will readily attack another female intruder, presumably as an adaptive defensive mechanism to protect potential offspring. These models allow for the study of both biologic sexes in chronic social defeat paradigms, and have resulted in the discovery of important sex differences in the development and manifestation of chronic stress-induced adaptations.

POSITIVE VALENCE

"Positive valence" refers to an emotional state that is intrinsically attractive, rewarding, or valuable. In psychiatric disorders, symptoms that fall under the category of positive valence include, for example, excess energy and impulsive behavior. Behaviors associated with positive valence that can be studied in animal models include the following:

Mania

Sleep disturbance often triggers manic episodes in bipolar patients, and sleep deprivation paradigms in rodents can produce short-term effects that resemble some aspects of mania such as hyperactivity, stereotypy, insomnia, aggressive behavior, and hypersexuality. Typically, rats or mice are placed in a small platform surrounded by water that does not allow for sleep without falling into the water. Seventy-two hours of sleep deprivation is sufficient to produce a brief manic-like state evidenced by increased frequency of the behavioral patterns mentioned above. Interestingly, these behaviors are sensitive to chronic treatment of mood-stabilizing drugs such as lithium. In addition, acute amphetamine injections in rats have been used to induce hyperactivity.

Impulsivity

Impulsivity can broadly be divided into actions that are initiated prematurely or without appropriate cognitive control, and actions that favor options with short-term payoff without taking longer-term consequences into consideration.

Impulsive Action. Behavioral inhibition can be measured in rodents, NHPs, and humans in the stop signal reaction time task. In the rat version, animals are trained to respond to a go stimulus, unless a no-go stimulus is concurrently presented. The no-go stimulus is presented at various time delays after the go stimulus, enabling a quantification of the maximal delay time at which a prepotent response can still be stopped. Clinically effective anti-impulsive drugs such as methylphenidate and atomoxetine decrease reaction time in both rodents and humans. Another test for impulsive action is embedded in the 5-choice serial reaction time task for rats. In this visuospatial attention task, after trial initiation, animals are faced with a standard intertrial interval before a visual stimulus prompts the correct response. When this intertrial interval is lengthened, rats tend to show premature responses, indicative of motor impulsivity. Clinically effective anti-impulsive medications tend to reduce premature responses in this task. Thus, both measures of motor impulsivity have good face and predictive validity. However, their operation requires substantial training, particularly for rodents and NHPs.

Impulsive Choice. Impulsive choices can be measured in delayed discounting paradigms in rodents, NHPs, and humans. Typically, subjects can choose between one option that results in an immediate small reward, and another option that results in a delayed larger reward. The relative discounting of future rewards can be deduced from choice behavior and can serve as an index of impulsivity. In animals, delayed reinforcement can only be delivered through experience and prior training. In contrast, human studies can use either experience of actual delayed rewards inside the timeframe of the experiment, or description of future rewards (outside of the timeframe of the lab visit). As such, animal experiments can best be compared to human studies using experience of delayed outcomes. Clinically effective anti-impulsivity drugs such as atomoxetine and methylphenidate have been shown to reduce choice of smaller immediate rewards.

Compulsive Behaviors

Repetitive invariant behavior with no apparent function or goal can easily be measured in its natural context. For example, excessive grooming in mice, circling behaviors in dogs, acral licking in dogs, or rocking, scratching, and head banging in NHPs occur naturally in stressful situations. Experimentally induced forms of compulsive behaviors include burying of unfamiliar marble objects, and the exaggerated drinking behavior that is observed in rats that respond for food under a fixed time schedule of reinforcement (schedule-induced polydipsia). Both types of compulsive behaviors respond well to clinically effective SSRI treatment for obsessive-compulsive disorder (OCD) patients. However, outcomes in the marble burying task also respond to antianxiety drugs like benzodiazepines, which are not clinically effective for reducing OCD symptoms in humans. In addition, other drugs that can effectively decrease marble burying have not been effective in the clinic for treating either anxiety or OCD. As such, results from the marble burying task alone are not considered sufficient evidence of reduced compulsivity.

In addition, more complex cognitive measure of perseverative responding can be obtained in operant tasks for rodents and primates. In a reversal learning paradigm, the ability of the animal to suppress lever pressing following previously reinforced stimuli is measured, in which continued responding despite the loss of the reinforcement reward can be interpreted as perseverative or compulsive responding. An analogous measure in the 5-choice serial reaction time task is based on repeated responding to a previously rewarded stimuli. Signal attenuation is a similar but more complex task that involves lever pressing rats that are presented with a sensory signal to mark reward delivery. When that cue is then presented without reward delivery (cue extinction), subsequent lever pressing under extinction conditions is not always accompanied by magazine entries to check for food rewards. The resulting lever pressing is considered to be perseverative. Of these three cognitive tasks, signal attenuation is most responsive to clinically effective SSRIs.

Abnormal Eating

The combination of undereating, compulsive or excessive exercise, and severe weight loss that is often seen in anorexia nervosa patients can be modeled in rodents. In the activity-based anorexia model, rats or mice are allowed to exercise on a running wheel, while food is only available for a limited amount of time per day. Paradoxically, access to the running wheel decreases food intake, which combines to produce severe weight loss that can lead to self-starvation. To date, no pharmacotherapies have been approved for the treatment of anorexia nervosa, so it is not clear what the predictive validity of this model is.

Excessive calorie intake can be measured by feeding rats or mice high-fat and carbohydrate food in the so-called cafeteria diet, which induces gradual weight gain and various metabolic changes. Binge-eating models restrict access to highly palatable food, such as chocolate- or fruit-flavored pellets, to short periods of time. Rats and mice show considerable increases in intake of the restricted highly palatable food over time, while intake of normal chow food decreases. These paradigms are useful for modeling aspects of binge eating disorder in humans, and recent studies have shown that such binge eating behaviors may be correlated with a dysregulated reward system similar to what is seen in addictive disorders. The etiology of binge eating disorder is still poorly understood with no clinically approved pharmacologic interventions.

Substance Use Disorders

Substance use disorders (SUDs) are chronic, relapsing conditions characterized by compulsive drug intake despite negative consequences. Patients with SUD frequently present with a genuine desire to decrease or eliminate their drug intake but typically experience great difficulty in achieving and maintaining abstinence. The prevailing framework considers SUD to be an acquired brain disease revolving around a dysregulated reward system, in which prolonged use of addictive drugs results in the overvaluation of drug reward at the expense of naturally rewarding stimuli. As these maladaptive changes make drug abstinence extremely challenging, preclinical research in addictive disorders typically focuses on methods to reduce drug craving/reward and withdrawal symptoms in order to assist a patient’s attempt to quit.

Drug Reward. In its simplest form, drug reward can be measured by administering a compound to rodents and measuring the behavioral consequences. In the conditioned place preference test, drugs are repeatedly injected when the animal is confined to one of two compartments, while control injections are paired with the alternative compartment. Subsequently, in a free exploration test, animals tend to spend more time in the drug-paired environment, which is taken as an index of drug reward. The conditioned preference test can also be carried out in humans, in which human subjects are exposed to two rooms, one of which contains free palatable food. After repeated exposure, subjects in this test will prefer to spend time in the food-paired room even when the reward is absent. Although this test is relatively simple to administer it has limited construct validity due to the lack of animal-driven behavior, whereas patients with SUD actively seek out and control drug consumption. Another way to quantify the rewarding effects of drugs of abuse is to combine drug administration with threshold intracranial self-stimulation testing, where a rewarding effect is measured as a decrease in intracranial self-stimulation threshold.

Drug Self-Administration

Both rodents and NHPs can be trained to self-administer drugs of abuse, with drug delivery occurring either orally or through intravenous infusion via indwelling catheters. Animals will self-administer almost all drugs with abuse potential, with the exception of hallucinogens like lysergic acid diethylamide and psilocybin. In the simplest form, stable self-administration can be established, and the effect of environmental, pharmacologic, or genetic manipulations on self-administration can be studied. In these experiments, a dose–response function for earned rewards can be established to assess changes in sensitivity to drug dosing (leftward or rightward shift in preferred dose) or the reinforcing effect of the drug (upward or downward shift in the number of rewards earned). There are numerous more specific behavioral manipulations that are based on self-administration. The progressive ratio is a related test in which the number of operant responses required to earn a drug reward increases over training, allowing for measurement of the maximum effort (i.e., maximum lever presses) animals are willing to perform to obtain a fixed dose of the drug. Although intake of drugs of abuse tends to be well controlled and stable in short daily sessions, extended access to many drugs of abuse leads to a progressive escalation of drug intake over days. Finally, after prolonged drug self-administration, animals can be challenged with an aversive punishment such as foot shock alongside drug administration. The willingness of some animals to continue earning drug rewards despite an aversive punishment is thought to model compulsive drug seeking, much in line with human patients who continue to seek out drugs despite negative consequences.

Patients with SUDs typically attempt to quit drug consumption but often fail to do so, and many models of addiction involve an abstinence phase after chronic self-administration. Immediately after the cessation of self-administration, animals may undergo psychological or physical withdrawal symptoms. After a period of abstinence or extinction of the drug-seeking response, reinstatement of drug-seeking can be precipitated by the drug-seeking context, response-contingent presentation of the drug-paired cue, stressors like mild foot shock, or noncontingent administration of the drug by the experimenter. Incubation of craving refers to the finding that the magnitude of reinstatement tends to increase with the duration of abstinence. Finally, although most drug self-administration tests are based on free operant procedures, discrete trial procedures allow for a discrete choice between drugs of abuse and natural rewards.

There are few effective and approved medications for various SUDs, and it remains difficult to evaluate the predictive validity of drug self-administration procedures for screening potential pharmacotherapies to treat SUDs. However, chronic drug self-administration procedures have many commonalities with SUD presentation in humans, including prolonged and self-driven drug consumption, lending these models excellent face and construct validity compared with noncontingent drug exposure paradigms.

COGNITION

Preattentive Processing

Prior to conscious processing, sensory information is selected and processed in the brain to facilitate further analysis. One measure of preattentive processing in rodents, NHPs, and humans is prepulse inhibition. In this test, a loud startle tone is preceded by a quieter pretone with a specific latency. The presentation of the quieter pretone reduces or inhibits the startle response to the loud tone, with the reduced loud tone response representing the degree of prepulse inhibition. Prepulse inhibition is disrupted in schizophrenia patients with psychosis, but also in many other psychiatric patients. The antipsychotic activity of therapeutic agents can be measured in combination with the administration of dopaminergic agents that impair prepulse inhibition, such as apomorphine. The P50 response is a similar measure of preattentive processing in rodents and humans. In this test, two identical auditory stimuli are presented approximately 50 ms apart, and the second stimulus normally produces a less strong event-related potential than the first.

Attention

Attention, or the process of selectively focusing on specific information, is typically tested by training subjects to report the presence of discrete signals over prolonged periods of time. There are several sustained attention tasks for rodents based on analogous continuous performance tasks for humans. These tasks differ somewhat in that humans typically do not undergo training before testing. In the aforementioned 5-choice serial reaction time task, rats are trained to report the presentation of a stimulus light in one of five locations. In the behavioral vigilance task, rats are required to report the presence or absence of a visual stimulus by different responses, and background noise is used to increase the difficulty of the task.

Working Memory

The capacity to transiently hold and process new and already stored information can be tested in rodents, NHPs, and humans. Behavioral models of working memory depend on a delay between information that informs the correct response and the opportunity to execute the response. In delayed alternation tasks, rats are trained to select alternating arms in a T-maze, either based on sequential free choice or with alternating forced choices and free choice trials. In the radial arm maze, eight arms are baited with a food reward, and the frequency of reentry in already visited arms serves as an index of working memory function. Operant delayed nonmatching to place is another rodent task that has a direct primate analog. In this task, rats or mice are presented with a stimulus lever (right or left) that is subsequently retracted, and after a variable delay, the animals have to press the alternative lever in order to earn a reward.

Rodent models can provide valuable information on the neural basis of working memory and can serve as a first screen for potential pharmacotherapies to treat working memory deficits in humans. However, evolutionary differences in cognitive capacity and prefrontal cortex architecture suggest that NHP models of working memory provide a closer approximation of human working memory. Primate tests of working memory focus on spatial or object working memory, which depends on dissociable parts of the prefrontal cortex. In the spatial task, a reward is presented in one of several locations, after which a screen is lowered for various delays. When the screen is lifted up, the animal can find a reward if it looks in the same location. In the delayed match to sample object task, primates are presented with multiple objects, and they have to select the matching object after a delay in order to receive a reward. Critically, the objects have to be reused frequently for the same animals so that they cannot solve the task based on recognition memory. An analogous system for humans is the Cambridge Neuropsychological Test Automated Battery (CANTAB) cognitive assessment system.

Declarative Memory

Animal models of declarative memories have been developed for rodents and NHPs. In the Morris water maze, rodents are introduced to a large swimming pool with cool opaque water and a hidden submerged platform. Over multiple trials, rodents will learn the location of the platform in the pool and will quickly swim to the platform location regardless of their starting position. Spatial memory can then be tested by removing the platform and measuring the amount of time the rodent spends swimming in the location where the platform was previously located, with more time spent in the platform's previous location being interpreted as good spatial memory. The Morris water maze has been criticized by some researchers because of a rodent's natural tendency to spend time at the edges of a container, which may influence outcome measures. In addition, other factors, such as stress induced by swimming or sensitivity to water temperature, can alter the responses in this assay independent of memory formation. Another similar model is the Barnes maze, in which rodents are placed in the center of a circular open field with up to 20 circular holes around the edges of the field, with 1 hole leading to an escape tunnel. Over time, the animal will learn to head directly to the escape hole with few errors or entries into incorrect holes. This behavioral paradigm takes advantage of the rodent's natural aversion to open and brightly lit spaces, offering an ethologic advantage over the Morris water maze. Both the Barnes and Morris water maze have strong spatial memory components that are heavily reliant on hippocampal function.

The spontaneous object recognition task for rats is based on a sampling phase with two identical and novel objects at two ends of a y-maze. After a variable retention delay, one of the objects is replaced by a new object, and rats with intact memory tend to spend more time exploring the novel object. The most popular NHP model of object recognition is the delayed nonmatch to sample task. The basic setup of the task is similar to the object working memory task described above, except that the objects are new to the animal at the time of testing, so the task can be solved based on recognition memory.

SOCIAL FUNCTIONING

There are many rodent tests of social functioning. In addition to the social anxiety-like behavior described above, there are specific models of social attachment, dominance, and aggression.

Social Attachment

Mother–infant attachment is often studied in rodents in the homing test, where pups at postnatal days 8 to 10 are separated from their mothers and allowed to choose between bedding from their own nest and bedding from another age-matched litter. The pups will prefer bedding from their own nest, which is interpreted as normal social function. Similarly, isolated rat pups will emit specific ultrasonic vocalizations upon separation from the mother. Further in development, the social functioning of mice can be assessed in the social preference test, in which time spent near a confined conspecific versus a nonsocial stimulus (social preference) or familiar conspecific versus novel conspecific (novelty preference) can be measured. In addition, social interactions can also be studied in freely interacting conspecifics. Social play in young animals can be quantified in rats by placing two unfamiliar young rats together and observing the frequency of species-specific playful behavior. In addition, the rewarding qualities of social play can be quantified in an adaptation of the place preference procedure, where play behavior is confined to one chamber of a two-chambered apparatus. Finally, the formation of monogamous pair bonds between partners is often studied in another rodent species, the prairie vole. Unlike most rodent species, prairie voles are monogamous and form strong lifelong partner bonds with their mate, even to the extent that males will not seek a new mate after their female partner dies.

Social Hierarchy and Aggression

Both mice and rats tend to form stable hierarchies and engage in aggressive behaviors under particular circumstances. Aggression in males tends to occur between cagemates in order to establish dominance hierarchies and toward intruders in order to protect reproductive success. Female aggression in rodents tends to occur specifically postpartum to protect offspring, although sexually naive female cage mates still form stable social hierarchies similar to males. Dominance and aggression can be observed in natural and seminatural settings using a deep, clear structure such as the visible burrow system. Dominance hierarchies can also be studied in mice in the tube test, in which pairs of mice are placed opposite of each other at the ends of a tube, and the mouse that does not retreat is considered the dominant mouse. Further, aggression is often provoked by the resident intruder test, sometimes after social isolation, which enhances resident aggression.

SLEEP AND AROUSAL

Arousal

Arousal and hyperactivity can be measured in rats and mice, typically through automated analysis of locomotor behavior in the home cage or in testing equipment like the open field test. Such tests can quantify overall movement and the relative occurrence of rest and activity.

Sleep

Natural Sleep. Natural sleeping behaviors are most often studied in rats and mice. Rodent models of sleep have been informative for human sleep regulation. Similarly to humans, sleep parameters can be quantified on the basis of behavioral observation or electroencephalographic (EEG) measures. More detailed analyses focus on the amount of sleep per day and the temporal distribution of sleep. However, unlike humans, rodents are nocturnal and typically sleep during the light phase. In addition, rodents show highly fragmented sleep episodes with several activity bouts during their normal sleeping hours. As such, results from rodent experiments on sleep disturbances should be interpreted with some caution.

Insomnia. The inability to sleep has been modeled in rats and mice through the effects of stress manipulations on sleep architecture. Stressors may include immobilization, social stress, or fear conditioning. These manipulations generally prolong the latency to fall asleep, reduce sleep duration, and increase the fragmentation of sleep. Most of these changes are reversible by medications used to treat insomnia and other sleep disorders.

Narcolepsy. Drowsiness during normal waking hours and sudden attacks of sleep have been modeled successfully in dogs and rodents. There are several strains of dogs that have naturally occurring narcolepsy, and the investigation of the genetic determinants of narcolepsy in dogs led to the identification of the orexin 2 receptor and its natural peptide ligand orexin as critical for normal arousal and wakefulness. The sudden arrest of movement, or cataplexy, is a hallmark of narcolepsy and can be measured in dogs with a loss of function mutation in the hypocretin/orexin gene in the food-elicited cataplexy test, in which bits of food are placed along a semicircle, and sudden arrests in the movement are taken as partial or complete cataplexy. Analogous mouse models of cataplexy have been developed to investigate the effect of genetic manipulations on arousal and sleep systems. In susceptible mice, cataplexy can be elicited by various manipulations, including anticipation of food, wheel running, and social interactions. Rodent cataplexy can be measured as brief episodes of behavioral immobility, in particular atonia of muscles in the back of the neck, and a sleep-like EEG signature.

Sexual Function

Normal Sexual Behavior. Normal and dysfunctional sexual behavior have most often been studied in rodents. Although sexual interaction in rodents has important species-specific components, such as a strong role for olfaction and a series of approaches and withdrawals, many aspects of sexual behavior are preserved between rodents and humans. Normal sexual function can be characterized in rodents by observation studies, and male ultrasonic vocalizations toward females can be measured. Mice and rats both undergo a dynamic copulatory sequence in which female rodents will display solicitation behaviors such as darting and ear wiggling to entice a male partner. These behaviors, alongside the female rodent's sexual posture (lordosis), mounting, intromission, and ejaculation, can be measured as a readout of sexual function. Another rodent species used to study sexual behavior is the Syrian hamster. This species does not engage in solicitation behaviors or locomotor activity during the copulation sequence, and as such, these animals can be used for experiments in which a relatively fixed posture is ideal, such as measuring neurotransmitter release or cellular activity.

Sexual Dysfunction. Although not often studied, stress can have detrimental effects on male sexual behaviors in rats, leading to longer latencies to mounting and intromission but shorter latencies to ejaculation. In addition, there are many models of male erectile dysfunction that are based on physiologic manipulations, such as cavernous nerve damage.

CONCLUDING REMARKS AND FUTURE DIRECTIONS

Preclinical studies using animal models, including those described above, have undoubtedly provided critical information on both the etiology and treatment of psychiatric disorders. Medications to treat psychiatric conditions are typically required to undergo extensive preclinical testing using rodents and NHPs before proceeding to human clinical trials, with the majority of effective pharmacotherapies having relied on preclinical testing during their development. In this way, preclinical animal models have led to an improved quality of life in countless patients that benefit from current psychiatric medications. In the past decade, rodents, in particular, have become critical model organisms to study psychiatric conditions due to the wide availability of genetically modified mice that allow for targeted manipulations of specific cell types or brain areas, shedding light on the biologic mechanisms underlying complex behavioral phenotypes. The recent discovery and implementation of CRISPR/Cas9 gene editing tools has rapidly accelerated the ability to generate these specialized mouse lines, while technologies like RNA sequencing and proteomic screens can provide a wealth of information about the molecular and cellular adaptations present in abnormal behaviors. Many would argue that we are closer than ever to both understanding and treating psychiatric conditions.

Even with the advent of advanced methods to study psychiatric conditions in animal models, significant barriers remain that will require careful consideration by researchers. Most of the models described in this chapter were developed with the intention of screening potential therapeutic agents for treating specific behavioral abnormalities, with construct validity being tied to the predictive validity of pharmacologic agents being effective in both humans and animal models. However, the vast majority of drugs that show promise in preclinical studies have no or only modest success in clinical trials. In addition, most pharmacotherapies that are approved for treating psychiatric conditions in humans show success in only a subset of patients, with prominent examples including SSRI antidepressants and smoking cessation medications. As psychiatric disorders are still classified and diagnosed almost exclusively through behavioral symptoms, it is likely that common behavioral symptoms occur through a variety of distinct biologic mechanisms. Animal studies, by nature, tend to have high levels of variability, and the careful consideration of individual differences in the preclinical setting may help address these barriers. In addition, the majority of classical studies that inform the development and utilization of animal models utilized only male animals, a trend that unfortunately persists in neuroscience research despite growing evidence that sex as a biologic variable influences psychiatric conditions and treatment outcomes. Moving forward, a more complete understanding of psychiatric conditions at the preclinical level will require the integration of advanced scientific technology with the investigation of individual, population-based, and species-specific differences in behavioral models.

FURTHER READINGS

Ågmo A. Animal models of female sexual dysfunction: basic considerations on drugs, arousal, motivation and behavior. *Pharmacol Biochem Behav*. 2014;121:3.

Belin D, Jonkman S, Dickinson A, Robbins TW, Everitt BJ. Parallel and interactive processes within the basal ganglia: relevance for the understanding of addiction. *Behav Brain Res*. 2009;199:89.

Bethell EJ, Cassidy LC, Brockhausen RR, Pfefferle D. Toward a standardized test of fearful temperament in primates: a sensitive alternative to the human intruder task for laboratory-housed rhesus macaques (macaca mulatta). *Front Psychol*. 2019;10.

Blanchard RJ, Flannelly KJ, Blanchard CD. Life-span studies of dominance and aggression in established colonies of laboratory rats. *Physiol Behav*. 1988;43:1.

Blanchard CD, Griebel G, Blanchard RJ. The mouse defense test battery: pharmacological and behavioral assays for anxiety and panic. *European J Pharm*. 2003;463:97.

Burkett JP, Young LJ. The behavioral, anatomical and pharmacological parallels between social attachment, love and addiction. *Psychopharmacology (Berl)*. 2012;224:1.

Castner SA, Goldman-Rakic PS, Williams GV. Animal models of working memory: insights for targeting cognitive dysfunction in schizophrenia. *Psychopharmacology (Berl)*. 2004;174:111.

Chung E, De Young L, Brock GB. Investigative models in erectile dysfunction: a state-of-the-art review of current animal models. *J Sex Med*. 2011;8:3291.

Coleman K, Pierre PJ. Assessing anxiety in nonhuman primates. *ILAR J*. 2014;55:333.

Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med*. 2013;11:126.

Cryan JF, Holmes A. The ascent of mouse: advances in modelling human depression and anxiety. *Nat Rev Drug Discovery*. 2005;4:775.

Cryan JF, Markou A, Lucki I. Assessing antidepressant activity in rodents: recent developments and future needs. *Trends Pharmacol Sci*. 2002;23:238.

D'Angelo LC, Eagle DM, Grant JE, Fineberg NA, Robbins TW, Chamberlain SR. Animal models of obsessive-compulsive spectrum disorders. *CNS Spectr*. 2014;19:28.

* Eichenbaum H. Exploring declarative memory using animal models. In: Eichenbaum H, ed. *The Cognitive Neuroscience Of Memory: An Introduction*. Oxford University Press; 2011.

* Evenden JL. Varieties of impulsivity. *Psychopharmacology (Berl)*. 1999;146:348.

Garcia LSB, Comim CM, Valvassori SS, et al. Acute administration of ketamine induces antidepressant-like effects in the forced swimming test and increases BDNF levels in the rat hippocampus. *Prog Neuropsychopharmacol Biol Psychiatry*. 2008;32.

Gestel van MA, Kostrzewa E, Adan RAH, Janhunen SK. Pharmacological manipulations in animal models of anorexia and binge eating in relation to humans. *Br J Pharmacol*. 2014;171:4767.

Griebel G, Holmes A. 50 years of hurdles and hope in anxiolytic drug discovery. *Nat Rev Drug Discovery*. 2013;12:667.

Hibicke M, Landry AN, Kramer HM, Talman ZK, Nichols CD. Psychedelics, but not ketamine, produce persistent antidepressant-like effects in a rodent experimental system for the study of depression. *ACS Chem Neurosci*. 2020;18:11.

Hyman SE. Revitalizing psychiatric therapeutics. *Neuropsychopharmacology*. 2014;39:220.

Hyman SE, Malenka RC, Nestler EJ. Neural mechanisms of addiction: the role for reward-related learning and memory. *Annu Rev Neurosci*. 2006;29:565.

Labonté B, Engmann O, Purushothaman, I. et al. Sex-specific transcriptional signatures in human depression. *Nat Med*. 2017;23.

Litz BT, Salters-Pedneault K, Steenkamp MM, et al. A randomized placebo-controlled trial of D-cycloserine and exposure therapy for posttraumatic stress disorder. *J Psychiatr Res*. 2012;46(9).

Markou A, Chiamulera C, Geyer MA, Tricklebank M, Steckler T. Removing obstacles in neuroscience drug discovery: the future path for animal models. *Neuropsychopharmacology*. 2009;34:74.

Miczek KA, Maxson Sc, Fish EW, Faccidomo S. Aggressive behavioral phenotypes in mice. *Behav Brain Res*. 2001;125:167.

Oddi D, Crusio WE, D'Amato FR, Pietropaolo S. Monogenic mouse models of social dysfunction: implications for autism. *Behav Brain Res*. 2013;251:75.

Peleg-Raibstein D, Feldom J, Meyer U. Behavioral models of antipsychotic drug actions. In: Gross G, Geyer MA, eds. *Current Antipsychotics, Handbook of Experimental Pharmacology*. Springer-Verlag; 2012.

Perusini JN, Fanselow MS. Neurobehavioral perspectives on the distinction between fear and anxiety. *Learn Mem*. 2015;22(9).

Rothbaum BO, Price M, Jovanovic T, et al. A randomized, double-blind evaluation of D-cycloserine or alprazolam combined with virtual reality exposure therapy for posttraumatic stress disorder in Iraq and Afghanistan War veterans. *Am J Psychiatry*. 2014;171(6).

Sarter M, Givens B, Bruno JP. The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Res Brain Res Rev*. 2001;35:146.

Schenberg LC, Schimitel FG, de Souza Armini R, et al. Translational approach to studying panic disorder in rats: hits and misses. *Neurosci Biobehav Rev*. 2014;46:472.

Shaham Y, Shalev U, Lu L, de Wit H, Stewart J. The reinstatement model of drug relapse: history, methodology and major findings. *Psychopharmacology (Berl)*. 2003;168:3.

Toth LA, Bhargava P. Animal models of sleep disorders. *Comp Med*. 2013;63:91.

Wang F, Zhu J, Zhu H, Zhang Q, Lin Z, Hu H. Bidirectional control of social hierarchy by synaptic efficacy in medial prefrontal cortex. *Science*. 2011;334:693.

1.23 Basic Systems Neuroscience

ANDRÉ A. FENTON, PH.D., AND DINO DVORAK, PH.D.

Systems neuroscientists study how neural circuits represent, analyze, and compute information by generating internal neural representations of variables in the external world to create knowledge, make decisions, and execute actions. Systems neuroscience aims to understand the causal relationships between the molecular and cellular levels of understanding brain structure and function, including the high-level mental functions that systems of neural circuits implement. For the basic systems neuroscientist, schizophrenia is at once a fascinating and frustrating subject to contemplate. Schizophrenia is fascinating because the core symptoms are of such substantial consequence. As such, schizophrenic symptomology is a profound reminder that thoughts and neural representations are neurobiologically mechanistic constructions that should be within our abilities to understand and decode, despite the apparently immense complexity. However, the absence of a mechanistic understanding of schizophrenia is a fundamental challenge to the effort to develop model organisms and model systems that could greatly accelerate scientific progress toward more effective treatments. The pendulum is decidedly on an optimistic swing as we enjoy the present era of advanced research tools that embrace the fact that the problem of schizophrenia, like that of understanding any mental phenomena, is fundamentally a systems problem that engages and demands integration across all the levels of biologic function and understanding from genes and molecules to cells and circuits and brain-regional specializations to individual and social behaviors (Fig. 1.23–1).

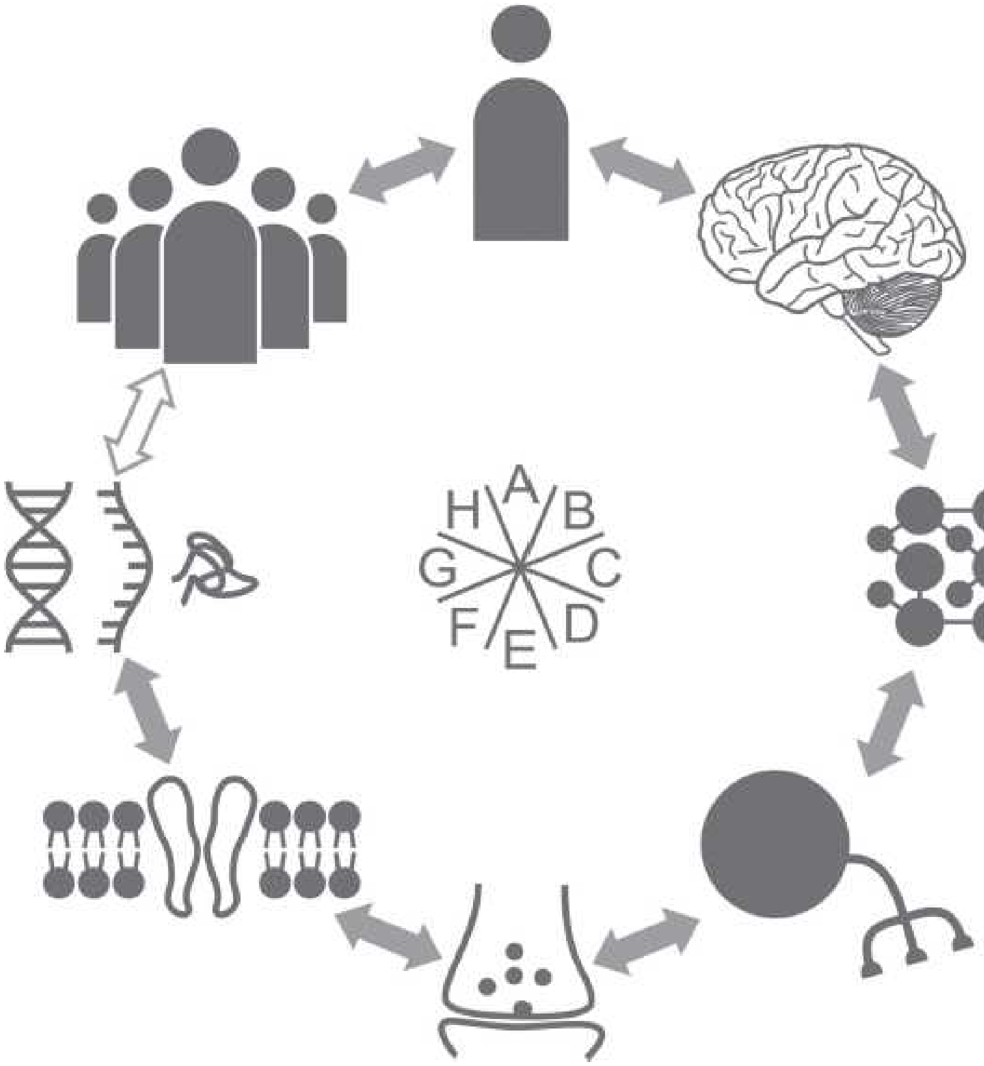

FIGURE 1.23–1. Circle of Being. This cartoon illustrates distinct levels of biologic organization, each of which has a direct mechanistic influence on the adjacent levels of biologic function and thus, an indirect influence on any other level. If we start arbitrarily at the level of an individual's behavior (**A**), the schematic illustrates that individual behavior both influences and is influenced by social behavior (**H**) as well as brain activity (**B**). Brain activity influences and is influenced by neural circuit operations (**C**), and that activity influences and is influenced by the activity and properties of single cells (**D**). Single-cell activity influences and is influenced by synaptic function (**E**). Synaptic properties and function are determined by membrane and channel properties (**F**), and those in turn influence and are influenced by genetic function, depicted here as DNA, RNA, and peptides (**G**). While genetic function is not known to directly influence social behavior, the effects are indirect via the other levels of biologic function depicted here. Nonetheless, much of what is known about the genetics of mental illness derives from studies that seek to correlate gene function with social and individual behavior, and this extends to attempts to use this genetic knowledge for therapeutic interventions.

This chapter will review the fundamentals of the basic systems neuroscience of cognitive circuit function in rodent models as may be relevant to concepts in schizophrenia. We will focus on the hippocampus, not merely because it is specifically abnormal in schizophrenia, but because the hippocampus circuit is a particularly well-characterized model system within which the systems neuroscience perspective on neural computation has been developed. This point invites the opportunity to clarify what can be productively expected of an animal model of schizophrenia in particular and any neuropsychiatric illness in general, at least from the systems perspective.

THE IMPORTANCE AND UTILITY OF ANIMAL MODELS

The importance of animal research for translating basic science to understanding mental illnesses like schizophrenia is highlighted by the fact that knowledge of basic neural and pharmacologic mechanisms has grown enormously while the expansion of treatment options has been much more modest. A central difficulty is the wide separation of biologic levels between the behavioral and mental spheres of the clinical manifestations and the biochemical domain of therapies and potential interventions. This separation is known as the “problem of the missing middle”; whereas research tools are optimized for studies at the microscopic levels of biologic function (genes, proteins, and cells), as well as the large systems level like the functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) methodologies; the technologies and concepts for investigating the middle levels of biologic function are far less developed or possible to implement in people. These missing levels include the temporally organized action potential discharge within ensembles, that is, subpopulations of single neurons and the temporally activated subpopulations of activated synapses and the consequences of the patterns of intracellular and extracellular current flow. Without this middle level, it has proven rather difficult to connect the nitty-gritty of receptors and transmitters to the clinical observables that drive treatment decisions. Though crucial, this middle level of investigation is only practical in animal research and is especially well developed for rodent preparations where it is feasible to record the extracellular electrical discharge from several tens to thousands of single neurons, concurrent with the population-level summation of synaptic currents that manifest as oscillations in the local field potential (LFP).

Since much of this chapter will be concerned with animal research that is conducted in the effort to understand schizophrenia-related issues, it is necessary to clarify what is and is not meant by the term “animal model.” Although the term is commonly used to indicate that the animal is being used to mimic some aspect of the clinical disorder, this is unlikely to be a valid or productive conceptualization. The mental phenomena that are the foundation of a schizophrenia diagnosis are hardly applicable to animals. There is great uncertainty about whether animals have the corresponding mental capacities, and if they do, it is highly unlikely that they will manifest as they do in people. There has been considerable discussion over the validity of the current DSM-based diagnostic criteria and whether there should be greater emphasis on objective biologic criteria as proposed by the National Institute of Mental Health’s Research Domain Criteria (RDoC) project. A consequence from the vantage of animal research is that uncertainty in defining the disease translates into making it impossible to mimic something that is poorly defined. But even basing diagnosis on objective criteria like measurements of functions like attention, working memory, and long-term memory, one has to be aware and careful that such functions can manifest in species-specific ways. Consider, for example, the Stroop test as a tool to assess cognitive control. Surely rodents also use cognitive control, but effectively assessing cognitive control in the rodent certainly does not resemble a Stroop test. There is substantial overlap in the genetic abnormalities that predispose to the diverse conditions that are classified as autism spectrum disorder, attention deficit hyperactivity disorder (ADHD), bipolar disorder, major depressive disorder, and schizophrenia. Which disorder would be modeled by the genetic abnormality in a mouse? Even though it is clearly practical and advantageous to mimic genetic alterations associated with schizophrenia in mutant mice, one must also be wary that pleiotropy has species-specific outcomes. A gene expressed in one species may confer considerably different phenotypes in different species, making it challenging to even interpret how a human genetic alteration might be mimicked in a model organism. Finally, the penetrance of genetic variants is typically a few percent, which from the perspective of mimicry, should mean that there is a low probability of detecting that a specific genetic modification leads to an abnormal phenotype. There are a multitude of reasons not to seriously consider that animal models of neuropsychiatric disorders mimic the disorder. Yet, beyond the problem of the missing middle, animal models are among our most powerful tools to test hypotheses about schizophrenia and other neuropsychiatric disorders. We do not consider any of the rodent models reviewed here to be models of schizophrenia itself. The animals are themselves tools, somewhat akin to a reagent, which can be used to effectively evaluate the key hypotheses that drive understanding of schizophrenia.

NEURAL ELECTRICAL ACTIVITY AND NEURAL CODES

A major general challenge in neuroscience that may, in particular, provide an opportunity for progress toward a systems-level understanding of schizophrenia is to determine how information is represented in the electrical activity of neurons. The standard textbook treatment of this fundamental question is the "dedicated coding" hypothesis in which the action potential discharge of single neurons is interpreted to signal the presence (or absence) of a complex stimulus like an object or the occurrence of a complex event like a movement or being in a particular location of space, as is the case for hippocampus place cells (see Fig. 1.23–2). There is ample cardinal cell evidence for the dedicated coding hypothesis. These are neurons dedicated to signaling high-order stimuli and concepts, like "face" or "grandmother." Especially dramatic, well-studied examples are the place cells, head direction cells, border cells, and grid cells in the rodent hippocampus and related structures that signal location, direction, environmental boundaries, and distance in external space. Another dramatic example recorded from the same brain regions of human epilepsy-surgery candidates is the cells with firing that are tuned to well-known individuals like the actresses Jennifer Aniston and Halle Berry. A problem with dedicated coding is that the capacity of a dedicated code is limited to the number of functional cell types, but a major attraction of the hypothesis is that it is conceptually straightforward to conceive of how it is possible for neurons to read and transform the information in the firing of such cardinal cells by, for example, higher-order representations like of a place emerging from the integration of converging connections from neurons representing lower-order concepts like the direction and distance to borders.

Much of the contemporary research into the neural activity correlates of information representation in animal models that are relevant to schizophrenia has been conducted and interpreted within the framework of the alternative “ensemble-coding” hypothesis. The hypothesis asserts that information is encoded by the pattern of activity across many neurons. An ensemble code is analogous to a stadium jumbotron display that uses many light bulbs to represent a message. There is ample evidence of ensemble coding, even from the same spatially tuned cells in the hippocampal and related neocortical areas, which are interpreted as evidence of dedicated coding (Figs. 1.23–2 and 1.23–3). Analogous to the jumbotron bulbs, the activity of a given neuron contributes to multiple representations, yet, no particular neuron is essential to a representation; many more representations can be encoded than there are neurons, and crucially, the same neuron can only represent one message at a single moment. Like the jumbotron, these properties require that the activity within the system of neurons (or light bulbs) is temporally coordinated so that the proper subsets of cells are coactive at the proper time intervals to represent and transmit information, perhaps at multiple time scales. Hebb’s idea of the cell assembly, a set of cells that fire together to represent information by the pattern of active and inactive cells, is an example of an ensemble code. While there is substantial evidence of ensemble codes operating in the mammalian and invertebrate brains, there is a major issue with the hypothesis. Unlike a dedicated code, it is unclear how neurons might detect, read, and transform information from an ensemble code. The difficulty arises from the fact that (a) the information in an ensemble code is distributed across time and space (i.e., diverse locations of cells) and (b) single neurons tend not to possess the necessary connectivity to collect an overview of all the cellular activity that defines the representational patterns carrying information. How might we fare, either individually or collectively, to read messages on a jumbotron display if each person could only see a subset of the light bulbs?

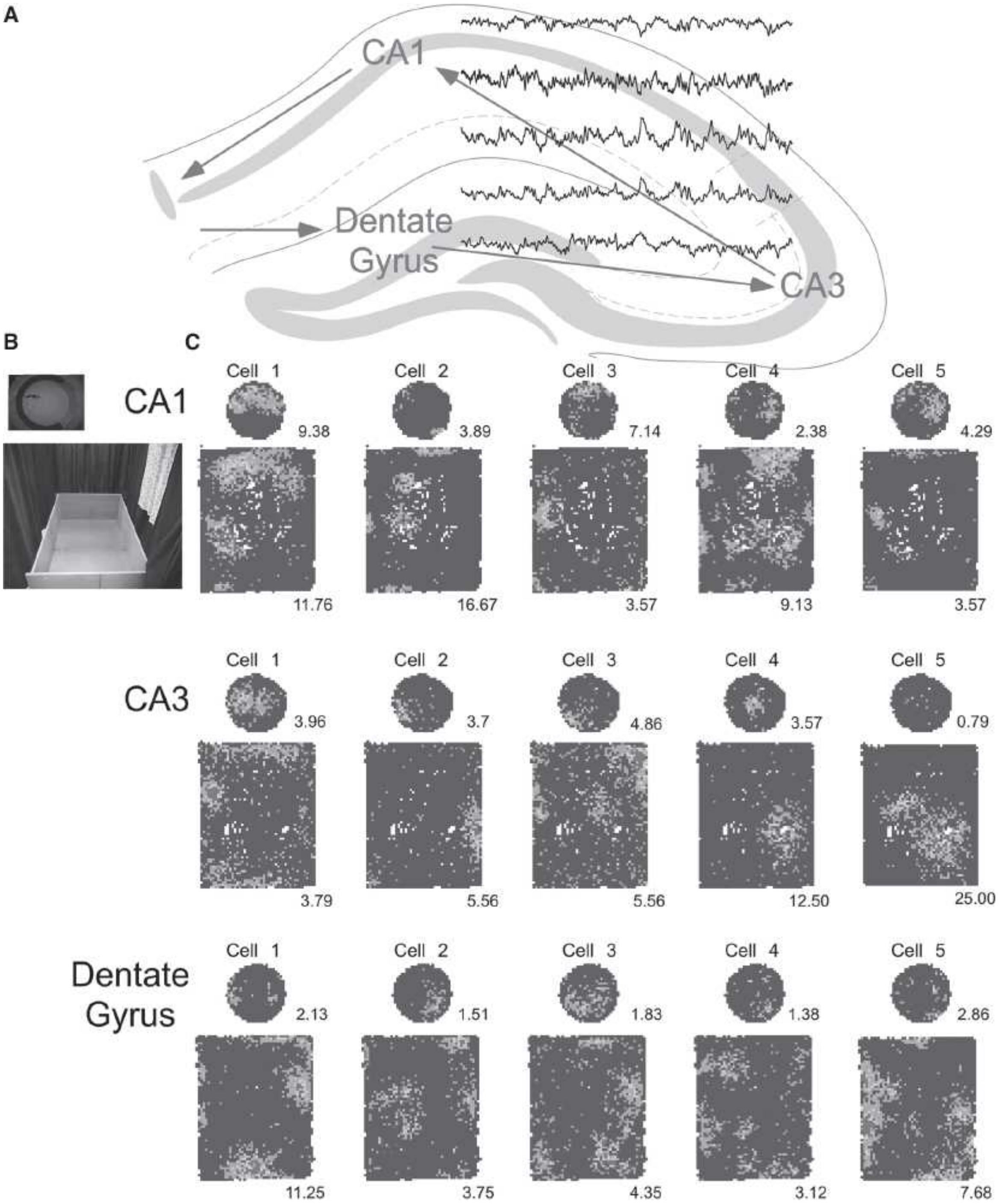

FIGURE 1.23–2. **Hippocampus place cells generate an ensemble place code. A:** Schematic of the dorsal hippocampus with the major subdivisions labeled and arrows indicating the feed-forward flow of information through the trisynaptic circuit. A 2-s example of the location-specific LFP has been superimposed such that the start of the LFP corresponds to the location where it was recorded. **B:** Photos of the two familiar environments in which place cells were recorded, a 76-cm diameter cylinder and a 1.8 × 1.4 m box. **C:** Blue-to-red color-coded average firing rate maps document the spatial

tuning of place cells recorded as a rat foraged for scattered food in the cylinder and box. Each set of five cells was recorded simultaneously from the same anatomical subdivision of the hippocampus circuit. Notice that in the cylinder, most place cells only discharge robustly in a single location, called the cell's place field, which suggests that the discharge provides a dedicated neural code for location. However, the fundamental nature of the place code is actually an ensemble code because, in the larger environment of the box, the same cells typically discharge in multiple place fields. If each cell's discharge in the box is interpreted as a dedicated code, the activity of each cell provides an ambiguous indication of location that cannot be reliably disambiguated within the dedicated coding framework. If, however, the discharge is interpreted as an ensemble code, the pattern of activity across the set of cells provides an accurate indication of location because the across-cell pattern of activity is unique at each location. The minimum firing rate (measured as spikes/s) in the *red* color (highest firing rate) category is given below each firing rate map. The *blue* color indicates a firing rate of 0 spikes/s. (See eBook for color images.) (B&C, reprinted from Park E, Dvorak D, Fenton AA. Ensemble place codes in hippocampus: CA1, CA3, and dentate gyrus place cells have multiple place fields in large environments. *PLoS ONE.* 2011;6(7):e22349. doi: 10.1371/journal.pone.0022349.)

The ensemble coding hypothesis offers a natural systems-level account for the cognitive and perceptual abnormalities that are observed in schizophrenia. Like the jumbotron, a neural network operating to represent information in an ensemble code must avoid the total loss of information resulting from simultaneously activating multiple representations with the same cells. A jumbotron can display two messages with the same lights, but only one message after the other as a temporally coordinated series of patterns, not simultaneously. Similarly, two cell assemblies that share neurons and are each defined as the subset of coactive neurons can represent multiple items so long as the two-coactivation patterns are separated in time. But there is a catastrophic loss of information if the two cell assembly patterns coactivate, because the coactivity merges all the active cells into a distinct, coactivity-defined cell assembly that represents something other than what is present or intended. Like any system that uses an ensemble code, a neural system can only operate effectively if mechanisms are in place to coordinate that cells within a single assembly coactivate and that multiple coactivity-defined assemblies do not coactivate. We will presently examine neural oscillations as one class of these coordinating mechanisms, but let us first consider what kind of conditions present a challenge to the ability to coordinate multiple neuronal ensemble representations of information.

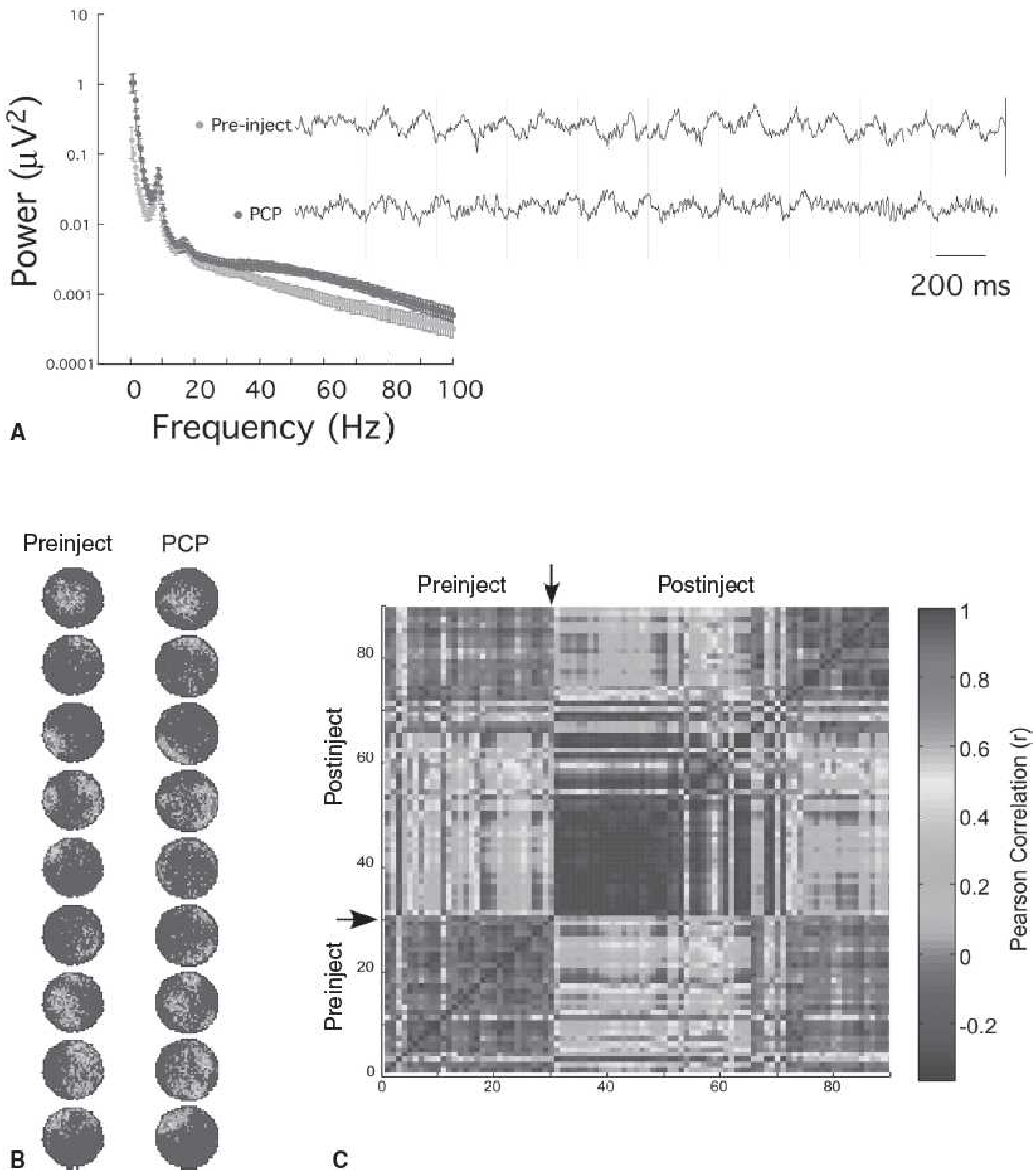

FIGURE 1.23–3. The psychotomimetic phencyclidine (PCP) discoordinates neural discharge with minimal effects on the spatial tuning of individual place cells. **A:** The blue-to-red color-coded firing rate maps of an ensemble of 9 CA1 place cells that were recorded as a rat foraged for scattered food in a familiar cylinder before and after systemic injection of 5 mg/kg of PCP. The PCP had no discernable effect on the spatial firing of the individual place cells. **B:** In contrast, when the same data are investigated as a time-varying across-cell pattern of discharge, it is obvious that the PCP radically changed the ensemble firing of the cells (PCP was administered at 30 minutes). To generate the correlation matrix that is shown, every minute, the number of

spikes that each cell discharged was counted and organized as an activity vector that described the temporal pattern of discharge across the population. Each activity vector was correlated with every other activity vector, and the correlation is represented using the color code. The correlation at any point in the matrix represents the activity at time x correlated with the activity at time y. The correlations along the x = y line are exactly 1 because it is the activity vector correlated with itself. The activity vectors under PCP immediately changed relative to the preinjection vectors (independent or negative correlations), but the PCP activity formed a stable pattern indicated by the correlations of approximately 1. Thirty to 40 minutes after PCP, the place cell ensemble activity switched back and forth between the preinjection and the PCP patterns until it settled back to the preinjection pattern. Notice that at 3 and 11 minutes during the preinjection period, the PCP pattern was expressed (in the absence of PCP). Not only do these data demonstrate discoordination of neural activity without affecting the information tuning of individual neurons, but also they indicate that ensemble activity is multistable. The system prefers to exist in one effectively stable activity pattern, and PCP radically changes the probability of observing the distinct patterns. (See eBook for color images.) (Data from Kao HY, Dvorak D, Park E, Kenney J, Kelemen E, Fenton AA. Phencyclidine discoordinates hippocampal network activity but not place fields. *J Neurosci.* 2017;37(49):12031–12049.)

A hypothesis for the information processing, perceptual, and cognitive impairments in schizophrenia falls out of the foregoing discussion of ensemble coding. Loss or aberrant coordination of neural activity will generate information processing failures in a network of neurons, and such failures are more likely when multiple sources of conflicting information compete for the same neural resources. Such ideas have been put forth in a number of ways. In essence, the notion asserts that discoordinated patterns of electrical neural activity are the core deficit in schizophrenia from which other deficits derive. This “discoordination” hypothesis is a systems-level attempt to account for the cognitive deficits in schizophrenia. It asserts that, by and large, cognitive and perceptual deficits of the disease are caused by the aberrant coordination of the many individually “normal” neurons that comprise a neural network or circuit. This discoordination is asserted to occur both within and between neural networks. Note that the coordination and discoordination of neural activity can only really be investigated and assessed directly by measuring the timing relationships of neural discharge among two or more neurons or other measures of neural activity, such as the EEG that reflects spatiotemporal patterns of synaptic currents.

An instructive example of the neural discoordination that is hypothesized to underlie the cognitive deficits in schizophrenia is demonstrated by the effect of the psychotomimetic phencyclidine (PCP) on rat hippocampus place cells. The drug selectively increases the power of gamma oscillations (30 to 100 Hz) in the hippocampal LFP (Fig. 1.23–3A), but it does not cause a detectible change in the spatial tuning of hippocampal place cells (Fig. 1.23–3B). Interpreting the effect on place cells from the dedicated coding perspective, one must conclude that PCP has a minimal effect on the discharge of these cognition-related cells. Remarkably, the conclusion is rather different if the same data are interpreted from the ensemble-coding perspective and evaluated as an across-cell pattern of discharge (Fig. 1.23–3C). The correlation matrix shows that PCP has a large and immediate effect on the temporally organized pattern of discharge across the sample of hippocampal place cells. PCP causes the familiar, stable pattern of discharge to change to a new, different stable pattern of discharge that does not correspond to the familiar representation of this familiar environment. As the PCP effect begins to wear off, about 30 minutes after administration, the neural activity pattern alternates between the preinjection and the PCP patterns before it settles into the familiar preinjection pattern. It is notable that the PCP discharge pattern had been transiently expressed at minutes 3 and 11 of the recording before PCP was administered. These data indicate that hippocampal neural activity is multistable and reveals that PCP changed the probabilities of the system being in the familiar preinjection and the aberrant PCP states, where the network state is defined by the coordinated temporally defined ensemble pattern of discharge and not on the basis of a measure of the discharge properties of the individual cells. Ensemble disruptions have also been measured in neocortical ensembles in genetic and chronic ketamine mouse models that are relevant to schizophrenia.

The discoordination hypothesis conceptualizes the coordinating processes as distinct or, at least, dissociable from the more unitary processes that characterize neural activity like the firing rate and the tuning curves of individual neurons, and the frequency of oscillations in the field potentials that can be recorded as EEG from the scalp or as an LFP when it can be localized to a specific location of a neural circuit. However, since both the unitary physiologic properties of cells and their coordinating processes result from the deployment and operation of the cellular machinery, including protein channels, at some sufficiently detailed level, alterations of the coordinating activity also influence these unitary properties. As currently conceived, the discoordination hypothesis predicts that basic properties and neural network functions like the tuning of neural responses to stimuli, the behavioral responses to simple, unitary stimuli, and the expression and use of memory will be relatively normal when conflicting information is minimal. Whereas, functional and discoordination abnormalities will be prominent with increasing likelihood under complex conditions that require ignoring and selectively using the accessible information to manage and meet competing demands. A canonical example of such a situation has been formalized as the Stroop test. The basic form of the test presents stimuli to the subject that provide two distinct forms of information, such as the words of colors, for example, “blue” and “red” written in either a congruent (“blue” written in blue) or incongruent color (“blue” written in red). Because the subject is instructed to report the color and thus ignore the meaning of the word, to be effective, the subject’s relevant brain circuits must selectively represent the color and not the meaning. Using tasks that present a cognitive control challenge like this, one can observe the discharge across an ensemble of neurons in rodents coalesce into coactivity-defined groups or assemblies of cells that represent one class of information, but the activity toggles between distinct patterns of coactive cells, each representing one or the other class of information. At any moment, the network of cells prefers to represent the information with the greater behavioral relevance.

Origins of Local Field Potentials and Neural Oscillations

Whether it is a single-neuron action potential or an ensemble of synaptic potentials, neural electrical activity derives from changes in the transmembrane potentials of excitable cells like neurons. The extracellular currents emanate through the conductive extracellular medium and are recorded as field potentials, the origins of which can be localized to the vicinity of an intracranial electrode and are hence called an LFP, or in the case of the EEG recorded with a scalp electrode. The spatial localization of the neural origins of the EEG potentials is much reduced. There are, however, developments in the ability to do source localization computations that have allowed fMRI-like imaging of functional brain activity from the EEG collected at sufficiently dense scalp electrode sites.

Extracellular fields are generated by the flow of positively or negatively charged ions through the semipermeable membranes of neurons. When positively charged ions (e.g., Na^+ or Ca^{2+}) leave the extracellular space in the cell, they cause local electrical sinks, which appear as a negative deflection in the LFP, measured at the proximity of the active membrane. Passive current in the opposite direction balances the current influx and is termed the return current. Excitatory neurons, such as those that release the neurotransmitter glutamate, cause a current influx in postsynaptic cells and, therefore, a depolarization of the postsynaptic membrane from its resting potential and an increased probability of a postsynaptic action potential. Inhibitory cells typically release gamma-aminobutyric acid (GABA), which in the postsynaptic cell causes hyperpolarizing and/or shunting inhibition where negative chloride ions can enter or leave the cell through chloride ion channels associated with the GABA receptors. This hyperpolarizes or stabilizes at rest the membrane potential of the postsynaptic cell, both of which make postsynaptic action potentials less likely. Currents through the membrane are extremely small, and, therefore, any measurable changes in the LFP are a result of many spatially uniform and temporally synchronized events. The largest contribution to the LFP comes from slow and synchronized currents activated by inhibitory neurons. While fast Na^+ and K^+ conductance-based action potentials generate the largest extracellular currents (an extracellular action potential or spike), they only have a limited contribution to the LFP spectrum below 100 Hz because of the short duration of the action potential events (<1 ms).

Neuronal oscillations are periodic fluctuations in neural activity in the central nervous system. Oscillations can be observed across multiple spatial scales. At the level of a single neuron, they either have a form of rhythmic, subthreshold changes of the membrane potential or rhythmic changes in action potential activity. The synchronized activity of large numbers of neurons gives rise to macroscopic oscillatory electric fields that can be observed in the LFP, invasively, using extracellular electrodes, or in the EEG, noninvasively, using scalp electrodes. Oscillations differ in several aspects, such as the typical frequency, anatomical localization, generating mechanism, relative power, and relationship to the ongoing brain state. Here, we will stress the importance of theta (5 to 12 Hz) and gamma (30 to 100 Hz) oscillations because their biophysical origins are well understood, and their interactions provide a functional coordinating infrastructure within which information-carrying action potentials are organized and controlled.

Theta oscillations can be detected in brain structures such as the hippocampus, entorhinal cortex, amygdala, sensory cortex, and prefrontal cortex. The theta rhythm has been associated with encoding and retrieval of memory, synaptic plasticity, long-range synchronization, and top-down control of cognitive functions such as attention and cognitive control. The hippocampal theta rhythm is the most prominent oscillation associated with the online state of the hippocampus; it is most notable during the intake of sensory information such as active locomotion, whisking, and sniffing, although it is also consistently present in the rapid eye movement (REM) stage of sleep. Theta rhythmicity is generated by pacemaker cells located in the medial septum-diagonal band of Broca, which provides GABAergic and cholinergic drive to hippocampal CA1 interneurons. On the other hand, excitatory CA3 hippocampal subfield and entorhinal cortical inputs to the dendrites of the CA1 hippocampal subfield may also produce stronger and higher frequency theta oscillations. The presence of multiple, weakly coupled theta oscillators might allow the necessary flexibility in the selection of appropriate cell assemblies for various cognitive tasks. One of the most important features of the theta rhythm is its ability to modulate the discharge of interneurons that generate local gamma oscillations by the postsynaptic inhibitory effects of their release of GABA. This mechanism was proposed as a general mechanism for chunking information throughout the brain. The result of such bias or so-called “nesting of oscillations” reflects the observation that the phase of the slower (theta) rhythm modulates the amplitude of the faster (gamma) rhythm (Fig. 1.23–4).

Gamma oscillations have the most ubiquitous rhythm; they have been localized in all brain structures where fast GABA-mediated inhibition is provided by soma-targeting interneurons. It is not surprising that something so widespread has been associated with perception, attention, memory, synaptic plasticity, and consciousness. In contrast to the theta rhythm, gamma oscillations are more locally generated, faster oscillations that cover a wide range of frequencies from 30 Hz to approximately 150 Hz. The upper limit of the gamma band has been a subject of intensive discussions. While the gamma rhythm was originally defined as the 35- to 85-Hz band, in more recent work, authors have subdivided the hippocampal gamma band into slow (<50 Hz), midfrequency (60 to 90

Hz), and “high” or “fast” (>100 Hz) components. However, caution is necessary for oscillatory activity above 100 Hz because of their possible contamination by the wideband spectral content of action potentials and associated after potentials. Gamma oscillations are generated by the interplay of excitatory and inhibitory neurons. GABAergic interneurons, especially interneurons that express the Ca^{2+}-binding protein parvalbumin, have been shown to be a particularly important part of the pacemaker system that generates gamma frequency oscillations. These interneurons receive glutamate receptor-mediated excitatory inputs. Transient excitation of the network of parvalbumin-expressing interneurons, which are mutually interconnected through electrical gap junctions and GABAergic synapses, generates rhythmic inhibitory postsynaptic potentials which can synchronize the firing of large populations of pyramidal neurons (Fig. 1.23–5).

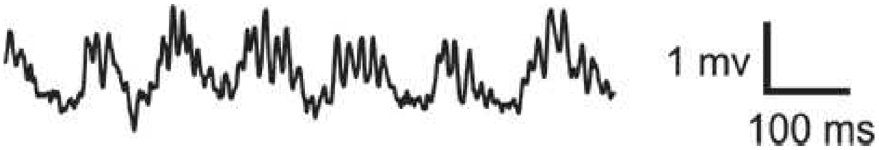

FIGURE 1.23–4. Phase-amplitude coupling of distinct neural oscillations. This LFP example illustrates the “nesting” of fast (60 to 100 Hz) gamma oscillations within theta (~8 Hz) oscillations. There are approximately 6.5 theta cycles depicted here. Notice that the largest gamma oscillations tend to occur preferentially following the peak of the concurrent theta oscillation.

Canonical pattern generating networks are shown in Figure 1.23–6. Feed-forward inhibition (Fig. 1.23–6 left) provides a mechanism for attenuation of the afferent excitation, where a distant excitatory cell activates a local interneuron or a group of local interneurons connected via gap junctions, which in turn inhibits multiple principal cells in the local network. This mechanism increases the temporal precision of firing by narrowing temporal windows of discharge probability of the principal neurons. Because the periodic repetition of the “window of opportunity” of the principal neuron discharge is defined by the time constant of the inhibitory neuron, this pairing of excitatory and inhibitory neurons is proposed as one of the generating mechanisms of gamma and other types of oscillations. In feedback inhibition (Fig. 1.23–6 middle), the interneuron receives excitatory glutamatergic input from the local principal cell and, in turn, provides inhibition to a specific compartment of the principal cells in the same network. Feedback inhibition provides a mechanism for reducing the excitability of principal cells and a mechanism for inhibition of a specific dendritic segment of the principal cell. Lateral inhibition (Fig. 1.23–6 right) provides an effective “winner-take-all” mechanism for selective activation of distinct ensemble patterns by increased activation of selected principal neurons (left population), which in turn inhibits other networks of neurons, therefore reducing their chance to fire (right population).

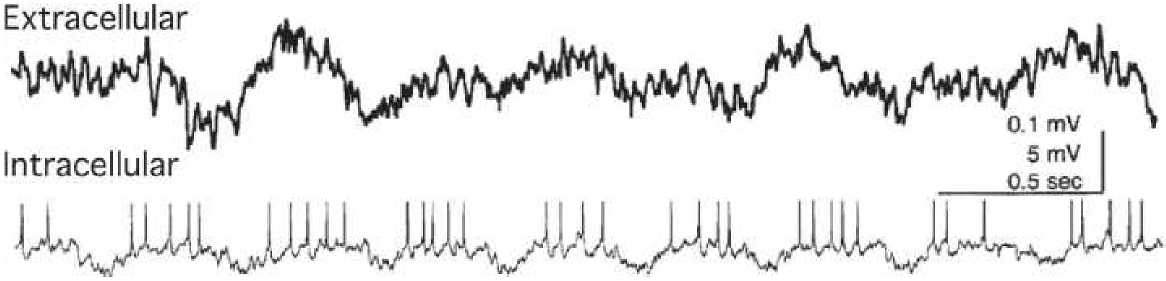

FIGURE 1.23–5. Theta-paced gamma frequency discharge of a parvalbumin-expressing basket cell type interneuron. A potential source of theta and gamma oscillations in the field potential is illustrated in this pair of simultaneous records from an intracellular recording of a parvalbumin-expressing basket cell and an extracellular recording of the LFP. (Adapted from Ylinen A, Soltesz I, Bragin A, Penttonen M, Sik A, Buzsaki G. Intracellular correlates of hippocampal theta rhythm in identified pyramidal cells, granule cells and basket cells. *Hippocampus*. 1995;5(1):78–90. doi: 10.1002/hipo.450050110.)

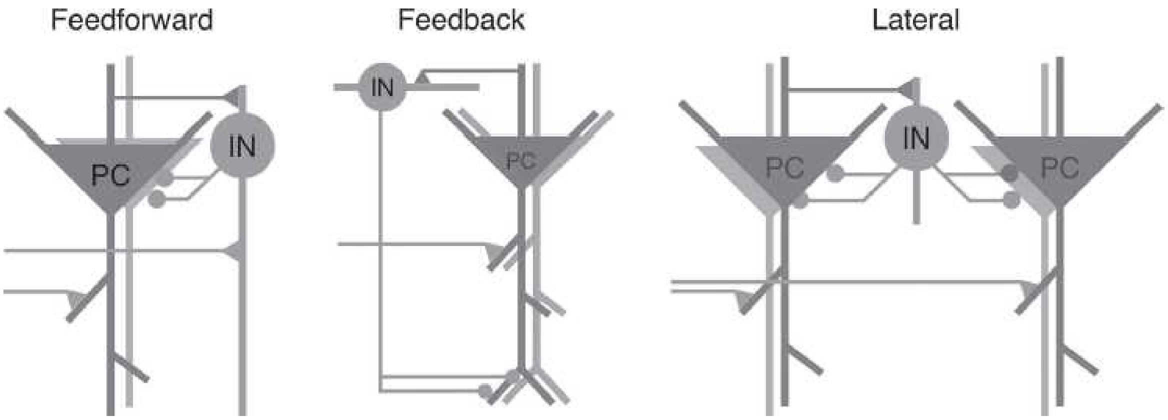

FIGURE 1.23–6. Basic mechanisms of network excitation-inhibition coordination. In these schemes, each cell type that is depicted, principal cell (PC) and inhibitory interneuron (IN), represents a population of the cell type, not an individual neuron. Circuitry for feed-forward inhibition (**left**) provides a mechanism for attenuation of the afferent excitation. Feedback inhibition circuitry (**middle**) provides a mechanism for reducing the excitability of principal cells and a mechanism for inhibition of a specific dendritic segment of the principal cell. Lateral inhibition circuitry (**right**) provides an effective "winner-take-all" competitive network mechanism for selective activation of distinct ensemble patterns according to relative effective input strength. The stronger of the momentary excitatory inputs will dominate and suppress the cells that are experiencing the weaker inputs, such that the network activity will exaggerate and sustain differences in input. Note that each of these functionally distinct schemes can be realized in the same circuitry based on the coordinated timing and levels of activity in the relevant neural elements.

Gamma activity is not actively propagated; instead, its role in neural coordination is to synchronize the spiking of neurons. From the perspective of the dedicated-coding hypothesis, it may not seem intuitive that inhibition would be synchronizing. However, from the perspective of the ensemble-coding hypothesis, inhibition synchronizes by defining narrow windows of opportunity for principal cells to discharge. The coupling between spiking activity and gamma oscillations could arise because local inhibitory neurons (crucial for generating gamma) fire preferably at the trough of the local gamma cycle. This mechanism makes the spiking of excitatory projection neurons more likely at the offset phase when inhibition is weaker. This feature of coupling between gamma and spiking has been hypothesized as a central aspect of cortical processing. Several hypotheses have also been proposed for the role of the gamma rhythm. According to one hypothesis, the phase of firing relative to gamma encodes the stimulus strength because stronger afferent excitation will permit firing to occur earlier in the window of opportunity that is set by ongoing gamma oscillations. Another idea is that gamma influences the communication between neuronal populations so that communication can entrain and increase when there is gamma phase-coherent spiking in the two populations.

Neural Coordination – Measuring the Varieties of Synchrony in Neural Oscillations. Various forms of neural synchrony, the coordinated, synchronized activation of same-function cells, and the active desynchronization of different-function cells have been proposed as a fundamental mechanism of neural computation. Neural synchrony appears across multiple scales. On the cellular (microscopic) scale, neural synchrony mediated by action potential and synaptic activity manifests as a transient coupling between functionally distinct neuronal subpopulations. This coupling of neural subpopulations creates a form of synchrony that manifests as oscillations in the LFP and the synchrony we observe on a mesoscopic scale of neural networks. The synchronized activity of a large number of neural networks, in turn, gives rise to the signals we observe at the macroscopic level, such as the scalp EEG.

An important aspect of neural synchrony is its cross-scale nature. For example, neural oscillations such as gamma synchronize action potentials within and between networks of neurons. Furthermore, slow oscillations, such as theta, synchronize transient bursts of fast oscillations such as gamma (Figs. 1.23–5 and 1.23–6). Coordinated neural interactions, therefore, appear across scales and across frequency bands. Although it is an extremely important and well-studied area, because of the difficulty of recording spiking data in clinical settings, we will not discuss spike-field synchrony further and instead focus on synchrony within and among the oscillations in field potentials.

Oscillations can be observed at multiple temporal and spatial scales. Cortical oscillations can be recorded noninvasively by conventional EEG. Circular cup electrodes of approximately 10 mm in diameter are placed on the scalp with a small amount of conductive gel between the scalp and the electrode. To facilitate fixing the electrode to a specific location of the head, electrodes are sometimes embedded in a fabric cap, which stabilizes their location during the recording. The location of electrodes has been standardized, as in the international 10 to 20 systems. Signals from electrodes are usually recorded using differential amplifiers where the common reference can be one of the electrodes (typically the *CZ* electrode at the center of the 10 to 20 system) or linked mastoid electrodes. Scalp EEG has several important limitations. The tissue between the cortex and the electrode is formed by the cerebrospinal fluid, pia mater, arachnoid, dura mater, skull, periosteum, and scalp. The tissue is highly heterogeneous, most importantly in terms of its conductance. The presence of the low-conductance bone causes signal smearing that decreases spatial resolution and source localization. Another problem is related to the distance between the electrode and the recorded neural tissue. Because the intensity of the electric field falls with the square of the distance, the measurable voltage change on the electrode (several tens of microvolts) corresponds to the synchronized activity of approximately 6 cm^2 of cortical gyri tissue corresponding to 60 million neurons forming the electrical dipole layer. The last important limitation of the scalp EEG is its frequency range. Because the spectral magnitudes of brain oscillations are inversely related to their frequency, according to a power law, it is progressively more difficult to record higher frequencies on the scalp because of the progressively worse signal/noise ratio. This fact limits the upper bandwidth of the EEG to approximately 70 Hz, although several techniques have been used to circumvent this limitation, such as trial-based averaging (event-related potentials, ERPs).

Higher spatial and frequency resolution than the EEG can be achieved by electrocorticography (ECoG) by using subdural electrodes which come into direct contact with brain tissue. In humans, this technique is limited to presurgery monitoring of epilepsy patients, where epileptogenic source localization provides the justification for this highly invasive technique. Compared to standard EEG electrodes, subdural electrodes are usually implemented in the form of grids or strips with typical electrode diameters of 2 to 3 mm and spacing of 10 mm between the electrodes. The close contact between the electrode and the neural tissue allows recording of LFPs up to 500 Hz.

The highest spatial and frequency resolution can be obtained by using penetrating high-impedance microelectrodes placed directly into the neuronal dipole sheet to record the electric signal of small groups of neurons. Microelectrodes come in several different forms, such as high-density microelectrode arrays used primarily in nonhuman primate research or micromachined multicontact silicon probes used primarily with rodents. When a microelectrode is placed in the CA1 region of the hippocampus, it can record the summed activity of up to 1,000 neurons within a radius of 140 μm.

The remaining parts of the recording system are principally the same, no matter if scalp EEG or LFPs are acquired. Conventionally, the signal is first routed into a signal follower, an operational amplifier with unity gain, although new low-power digital electronics can skip this step. The goal is to reduce the signal impedance to minimize additional electric noise caused by the signal passing through the cable that connects the subject to the main amplifier and digitization stages, and this can be avoided if the digitization occurs at the electrode site. After this stage, the signal is amplified using one or multiple operational amplifiers, so the signal's voltage is increased from tens to hundreds of microvolts to several volts. At this point, the signal is high-pass filtered to avoid saturation of the analog-to-digital converter (ADC) and then low-pass filtered to avoid aliasing effects in the ADC stage and finally routed to the ADC for digital conversion. The output of the ADC is a 16- or 24-bit binary number corresponding to the voltage between the active electrode and the reference electrode. The binary number is stored in computer memory for further processing. Recent advances in the use of metal-oxide-semiconductor silicon fabrication technology have made it possible to design dense arrays of electrodes with the electrode contacts integrated with the amplifier and digitization circuits. Instead of placing a limited number of electrodes at specific targets, this technology allows a large, excess number of electrodes to be placed such that only the subset that are in regions of interest will be recorded. For example, Neuropixels 2.0, that were designed with this technology, have 5,120 15-μm square recording sites distributed over four shanks such that using computer-controlled electronic switches, a user can select to record the electrical activity from any 384 of the sites. The 384 digital signals are multiplexed into a single digital data stream and transmitted on a single wire, allowing for the electrical activity from thousands of cells to be recorded with a recording system that weighs 1.1 g. This electronics technology advance is especially important because, although it has been possible to record the electrical activity of thousands of neurons by transducing the calcium transients that are associated with action potentials into a fluorescent signal using genetically encoded calcium indicators (GECIs), the time resolution (∼10's ms) of even the latest GCaMP8 GECI is too slow to resolve individual action potentials or relate them to ongoing oscillations in the LFP. Nontoxic genetically encoded voltage sensors that are now being developed promise another revolution in recording technology. Such voltage sensors will enable the recording of the subthreshold transmembrane fluctuations as well as the action potentials from many tens to hundreds of individual cells in awake-behaving brains, all with high-temporal fidelity and μm spatial resolution. These unprecedented possibilities are sure to bring novel insights and permit novel questions and specific biophysical hypotheses to be evaluated, but in the meantime, analysis of neural oscillations is the state of the art for patient recordings.

Analysis of Oscillations. The most basic approach for the analysis of oscillations is based on Fourier analysis, where the time representation of the voltage signal is transformed into the frequency domain using the fast Fourier transform (FFT). The result of this transformation is a power spectrum, which defines the original voltage time series as a function of oscillations of different frequencies. Specifically, the power spectrum is the amount (power) of each frequency-defined sine function from which the original voltage time series signal is composed. Importantly, computing the FFT requires a sufficiently long time series of data, and it is assumed that the frequency content of the signal did not vary in any meaningful way in time. Consequently, appropriate analysis using the power spectra assumes stationarity of the signal and, as such, cannot provide any information about the dynamics of oscillations, should the stationarity assumption be invalid. Given that neural signals are time varying, especially those that might contribute to cognitive processes, it is important to recognize that several methods that have been developed for digital signal processing engineering applications are now available and are growing in their use and utility for systems neuroscience applications. One such development is called the short-time Fourier transform (STFT), an extension of the Fourier transform. The STFT adds the time dimension to the power spectra, which allows analysis of the dynamical changes of the power at individual frequencies over time. Limitations of the time-frequency resolution of the STFT had been partially overcome by wavelet analysis, which enables time-frequency analysis at various levels of detail or so-called scales. When a complex wavelet is used, one can also estimate the instantaneous phase and amplitude information from the signal. This concept is similar to the application of a pass-band filter followed by the Hilbert transform, which will also yield a time series of the phase and amplitude of a frequency-defined signal. The ability to measure the time-varying spectral content of EEG and LFP signals has made it possible to study the interactions of distinct oscillations, like that of the theta and gamma oscillations that can be seen in the LFP fragment in Figures 1.23–3 and 1.23–4. Interactions of oscillations can be studied by methods of bivariate data analysis. In general, these methods analyze the relationships between various aspects of signals, such as their amplitudes in time (cross-correlation), their phase relationship (coherence, phase locking), or the relationship between amplitude and phase (cross-frequency coupling [CFC]). The most advanced methods, such as phase locking and CFC, are based on a similar scheme. First, both signals under analysis are band-pass filtered. After that, the Hilbert transform separates the phase and amplitude information of both signals. In the next step, the phase and/or the amplitude information is compared using indicators such as circular variance or mutual entropy. The last important step is testing the statistical significance of any phase and/or amplitude relationships, for example, by using surrogate data techniques.

There are several types of synchrony that can be readily measured. Amplitude synchrony (Fig. 1.23–7 left) is maximal when the amplitudes of two signals covary and coincide in time. This form of synchrony can be detected by linear cross-correlation. Phase synchrony (Fig. 1.23–7, middle) is maximal when the difference between the instantaneous phases of two signals is constant. Phase synchrony can be detected by measures such as the phase-locking value (PLV) or coherence, although amplitude synchrony also contributes to coherence. Phase-amplitude synchrony (Fig. 1.23–7, right) is maximal when the amplitude of a fast signal like gamma is modulated by the phase of a slow signal, like theta. That coordination is called phase-amplitude coupling (PAC), but because such interactions also happen between different frequency bands, it is also known as CFC. The term “nesting” has also been used to describe PAC, and it emphasizes the fact that the phase of the slow oscillation modulates the amplitude of the fast oscillation within the cycle of the slow oscillation. While all forms of neural synchrony have a role in neural circuit and brain function, the present goal is to be illustrative rather than comprehensive, so we will expand the discussion of PAC to illustrate key principles of how neural oscillations participate in neural circuit computation.

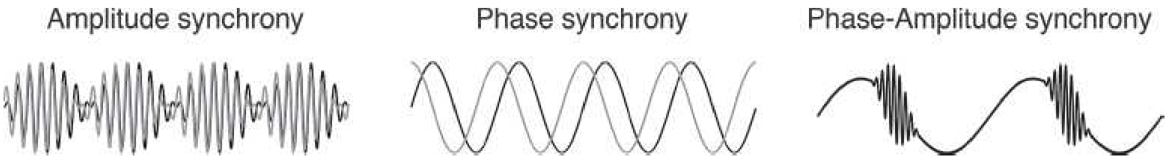

FIGURE 1.23–7. Types of synchrony. In amplitude synchrony (**left**), the amplitudes of two signals (depicted as *blue* and *black*) follow similar temporal patterns. In phase synchrony (**middle**), the phases of two signals maintain a constant phase offset over time. In phase-amplitude synchrony (**right**), the amplitude of the fast oscillation appears at a specific phase of the slower oscillation.

Phase-Amplitude Coupling. PAC has been observed across a variety of brain regions as well as across different animal species. We will not consider here how PAC is computed but refer the reader to a review of the methods and their limitations. The power of fast gamma oscillations (80 to 150 Hz) is phase-locked to ~8 Hz theta oscillations in the human neocortex, and different behavioral tasks evoke distinct patterns of theta–gamma coupling across the cortex. The properties of theta-slow gamma (25 to 50 Hz) coupling in the human neocortex reflect the number of items stored in working memory. Increasing memory load is associated with a reduction in the frequency of the slow oscillation, consistent with the idea that longer oscillation cycles are required to accommodate an increased number of information representations, each of which is paced by a gamma cycle.

PAC is thought to coordinate and route information flow into computationally different compartments of a neural circuit. In the rat CA1 region of the hippocampus, the theta phase of local oscillations is demonstrated to modulate the amplitude of two somewhat overlapping gamma frequency bands, slow and fast gamma. A role of the slow delta (1 to 4 Hz) rhythm in the sensory selection of the task-relevant input stream has also been demonstrated. Although the physiologic mechanisms of PAC are not well understood, it is hypothesized that perisomatic inhibitory basket cells in the hippocampus play a role in the spatiotemporal organization of cell assemblies by firing theta rhythmic trains of action potentials at gamma frequency, and therefore, this activity contributes to both rhythms and their phase–amplitude and phase–phase relationships (Fig. 1.23–5).

Several conceptual frameworks were also proposed for the computational role of PAC. The concept of a cross-scale organization of neural activity offers a possible neural mechanism for integrating information between several functionally distinct networks to accomplish perceptual binding, selective attention, cognitive control, and the recruitment of specific computational cell assemblies. Neural activity in macroscopic (slow oscillations), mesoscopic (high-frequency oscillations), and microscopic (single-neuron activity) scales are braided together such that a progressively faster activity occurs within a short time window of slower activity. A particularly clear example is seen in the hippocampus LFP of the freely behaving rat. The phase of the hippocampal theta rhythm modulates the amplitude of hippocampal gamma oscillations (see Fig. 1.23–4). In turn, gamma oscillations provide millisecond-scale "windows of opportunity" when the discharge probability of hippocampal pyramidal cells is the greatest. It has been suggested that because of conduction delays, slow oscillations such as theta are particularly suitable for the synchronization of functional networks over longer distances. On the contrary, fast oscillations such as gamma might transiently synchronize neuronal action potential discharge into same-function assemblies over relatively short spatial scales that can be as long as a few hundred microns. The interplay between the two oscillations might then provide a mechanism for spatial integration from a local to a global scale. In this scheme, the faster gamma oscillations that emerge at the specific phase of the slower theta oscillations might recruit subsets of cells that signal the same information into a cell assembly of coactive cells for neuronal information processing. Theta might then serve as an integration mechanism between several functional networks, each associated with a particular gamma episode across longer periods of time. Another conceptual framework for the functional significance of PAC is based on the idea of phase coding, where PAC is a mechanism for encoding different information at different phases of the theta cycle. In this scenario, faster gamma oscillations appearing at specific phases of slower theta oscillations are hypothesized to activate distinct neural ensembles, which encode distinct representations by the across-cell pattern of discharge as illustrated in Figure 1.23–4.

Discoordination of Neural Discharge and Oscillations

Disturbances of neural synchrony are proposed to underlie the cognitive deficits that associate with diverse brain disorders that include schizophrenia, autism, Alzheimer disease, Parkinson disease, and epilepsy. It is not a simple matter that dysfunction emerges from too little synchrony. Rather, the idea is that neural synchrony is optimized for cognitive function and either too much or too little manifests as abnormal cognition. Gestalt perception of Mooney faces (black and white contour pictures of human faces) is impaired in schizophrenia patients, and the impairment is associated with reduced phase synchrony in the gamma band. It has also been reported there is a selective alteration of theta-frequency oscillations, suggesting deficits in the large-scale synchronization while local synchrony in the gamma band is largely intact. Several studies have found that the positive symptoms, specifically visual and auditory hallucinations and delusions, are correlated with enhanced amplitude and phase synchronization of evoked and induced beta- and gamma-band activity. On the other hand, negative symptoms have been associated with both enhanced and reduced high-frequency oscillations. In fact, there is substantial evidence that resting, baseline (i.e., unprovoked) gamma power is elevated in schizophrenia. This might, in part, explain why stimulus-evoked gamma elevations appear to be attenuated in schizophrenia patients. While exact mechanisms are not known to explain these diverse findings of neural discoordination in schizophrenia, it is hypothesized that reduced parvalbumin expression in the cortical GABAergic interneuron population indicates GABAergic dysfunction, and this results in impaired gamma-band synchronization, which is crucial for the local synchronized firing of large populations of pyramidal neurons and may lead to deficits in the temporal coordination of distributed neural activity.

Neural Discoordination in Neurodevelopmental Models of Mental Disorders. An important development in understanding mental illness in general and schizophrenia, in particular, is the recognition that these are neurodevelopmental disorders. The effect of disturbing neurodevelopment has been investigated using animal models, which provide some of the most compelling support for the assertion that disorders like schizophrenia have a neurodevelopmental origin. In the MAM-E17 rat model, the mitotoxin methylazoxymethanol acetate (MAM) is administered to pregnant dams on embryonic day 17 to preferentially disrupt the ongoing embryogenesis of the hippocampus and prefrontal cortex. Adult MAM rats exhibit deficits in cognitive flexibility and hyperresponsiveness to dopaminergic agonists, hippocampal glutamatergic dysfunction, and reduced parvalbumin-positive interneuron expression. These abnormalities accompany reduced cortical up- and down-state transitions that underlie state-selective neural responses and slow oscillations in sleep. The coupling of hippocampal (150 to 250 Hz) ripple oscillations and sleep thalamocortical spindles in the prefrontal cortex is also disrupted in adult MAM rats. Thus, the MAM neurodevelopmental insult causes a wide variety of abnormalities with general relevance to schizophrenia and clear evidence of discoordination of neural network function.

The neonatal ventral hippocampus lesion (NVHL) rat is one of the most studied neurodevelopmental models in schizophrenia research. It was developed on the hypothesis that an early-life brain insult causes aberrant development of neural circuits. The excitotoxic lesion of the ventral hippocampus occurs on postnatal day 7, corresponding to the third trimester of pregnancy in human brain development. NVHL studies provide some of the strongest evidence that early brain insults can disturb neural circuits and cognitive functions relevant to schizophrenia. Adult NVHL rats demonstrate abnormalities that are relevant to mental illness, especially schizophrenia. These include abnormal cortical dendritic morphology, reduced GABAergic metabolism, excessive dopaminergic signaling, hyperresponsive behavior to dopaminergic agonists, and various behavioral abnormalities that emerge in adulthood. There is an increased prevalence of seizures in schizophrenia, and the incidence of epileptiform activity is increased in adult NVHL rats. Making the lesion in adulthood does not produce the same NVHL abnormalities, consistent with a developmental component to the consequences.

Single neuron responses in the prefrontal cortex and nucleus accumbens are aberrant in adult NVHL rats. Prefrontal firing normally decreases in response to dopamine from stimulation of the ventral tegmental area, but prefrontal activity increases in NVHL adults. Such disruptions of excitation–inhibition coordination occur, in part, because of impaired interneuron activation by dopamine. Disinhibition of the adult NVHL neocortex is associated with cognitive impairments, such as in decision tasks. In control rats, the adolescent prefrontal network is also disinhibited and less coordinated than in adults, suggesting that NVHL information processing networks are functionally immature.

NVHL cortical network activity patterns are discoordinated within and between the hippocampus and prefrontal cortices during cognitively challenging behaviors. During an odor discrimination task, rats generate behavioral sampling event-related increases in low-frequency (1 to 30 Hz) prefrontal oscillations. These increases are weak in NVHL rats. Consistent with reduced oscillations being an indicator of reduced excitation—inhibition organization, prefrontal neurons in NVHL rats are hyperresponsive when the animals sample odors and this was exacerbated on large reward trials when dopamine levels are expected to increase. These abnormalities were not evident during baseline behavior when differences in oscillatory power are hard to detect, indicating that network activity patterns are sensitive to what the subject is doing and presumably also thinking. It follows that group and treatment differences tend to emerge during cognitive effort, but it is also the case that physiologic differences may be irrelevant to cognitive performance, and this is difficult to know unless the behavior and physiology are assessed concurrently.

The interdependence of cognitive behavior and network activity is evident in studies of NVHL adults performing an active place avoidance task on a slowly rotating arena. The task requires intact hippocampus function and maintained hippocampal long-term potentiation (LTP) as well as cognitive control to use stationary spatial cues and ignore irrelevant rotating olfactory cues. NVHL adults are impaired at place avoidance, but with extended training, they can approach the performance of controls. NVHL rats are not impaired in adolescence, nor are they impaired if the cognitive control demand is reduced, indicating that an increased incidence of epileptiform activity is not sufficient to cause basic performance learning and memory deficits. Hippocampal and prefrontal LFPs are reduced across a wide band of frequencies during walking, but this is insufficient to explain the cognitive deficits which depend on experience and the availability of cues. Compared to during the task, interhippocampal and interprefrontal phase and amplitude synchrony of oscillations are reduced during home-cage behaviors such as quiet rest, grooming, and eating, except in the gamma range, which is similar in NVHL and controls in all conditions. Interhemispheric phase synchrony increases during place avoidance compared to home cage behaviors, but prefrontal synchrony is not different between NVHL and control groups during place avoidance. Theta-phase synchrony between the hippocampi is weaker in NVHL adults, even when the NVHL rats are performing as well as controls. These differences in adulthood were prevented by cognitive training in adolescence, presumably because of persistent changes in neural network function. The neural discoordination and cognitive abnormalities are also partly attenuated by reducing excitability with the antiepileptic ethosuximide, consistent with the notion that excitation–inhibition disbalance causes discoordination in NVHL adults and the possibility that this discoordination may be sufficient to explain the cognitive abnormalities.

Substantial advances in identifying genetic deletions and syndromes that confer risk for schizophrenia have allowed these genetic alterations to be modeled in mice. One basic observation is that such genetic deletions tend to confer risk for multiple disorders, including major depression, bipolar, ADHD, and autism, in addition to schizophrenia. This apparent pleiotropy highlights the value of understanding the mechanistic consequences of the genetic alterations and the fact that changes in cell biologic fundamentals like synaptic proteins and receptors can have diverse consequences depending on them when they are expressed and the specific cell types in which they are expressed. The *Df(16)A(+/−)* mutant mouse is emerging as an important genetic model because it models the 22q11.2 microdeletion syndrome with increased risk for a broad spectrum of neuropsychiatric diseases, including schizophrenia. These mutant mice express synaptic, cellular, neurocircuit, and behavioral abnormalities that include discoordinated neural activity in both the neocortex and hippocampus, including their interactions that predict cognitive dysfunction. Perhaps the most popular genetic model is the *Fmr1*-null mouse, which lacks the gene to make the RNA-binding protein named fragile X mental retardation protein (FMRP). The *Fmr1*-null mouse models the genetic defect in fragile X syndrome that results from silencing of the FMR1 gene due to an excessive (>200) CGG nucleotide repeat expansion. Lacking FMRP, *Fmr1*-null mice express a variety of synaptic, cellular, and neurophysiologic abnormalities, many of which can be restored by restoring FMRP or the molecular signaling for translation of mRNA to protein or even countering levels of dysregulated cAMP metabolism and protein expression resulting from FMRP loss. Consistent with the possibility of attenuating cognitive deficits and the discoordination hypothesis, despite the diverse cellular abnormalities, *Fmr1*-null mice can express normal hippocampus place cells and single-cell electrophysiologic characteristics, but discoordinated action potential timing among these cells and their relationship to ongoing oscillations in the LFP. These findings suggest fundamental excitation–inhibition discoordination in *Fmr1*-null as well as the *Cntnap2(−/−)*, *16p11.2(del/+)*, *Tsc2(+/−)* genetic models of defects observed in autism spectrum disorders. Such dissection and elucidation of the complex multilevel consequences of even a single-gene defect illustrate the power and promise of animal models for identifying novel therapeutic options. While these and other findings in *Df(16)A(+/−)*, *Fmr1*-null and other mutant mice with relevance for neuropsychiatric disorders are consistent with the discoordination hypothesis, understanding how these deficits of neural dysfunction emerge and how they can be normalized remains a daunting challenge. Consequently, the field is focusing on powerful new neurobiologic technologies that utilize animal models to identify elusive mechanisms.

SUMMARY

The systems neuroscience perspective embraces the diverse levels of biologic organization, and when this is practiced in an integrative framework, the goal is to make measurements at the different biologic levels from the same subject to look for cross-scale relationships and principles. This systems approach can be especially powerful when it is applied to testing hypotheses using animal models that are associated with mental illness, such as models that are relevant to schizophrenia, because it is extremely difficult, if not impossible, to perform the requisite multiscale systems-level analyses in a clinical population. As a closing example, let's consider the loss of GABAergic tone that is associated with schizophrenia and manifests as a reduction of parvalbumin immunoreactivity that is associated with GABAergic interneurons like the one depicted in Figure 1.23–5. Such compromised GABAergic function is consistent with the systems-level discoordination hypothesis that asserts that cognitive dysfunction is a result of multiscale disturbances that impact the different levels of biologic organization (Fig. 1.23–1). According to this viewpoint, genetic abnormalities and altered developmental programs cause, at the level of synapses and cells, impaired molecular interactions due to inefficient calcium binding, which is the normal function of parvalbumin. This can have diverse consequences, including abnormal synaptic signaling and the dysregulation of synthesis of synaptic proteins, all of which would contribute to disbalancing excitation and inhibition at the level of neural circuits. Because these circuits are dynamic and nonlinear, at least in part because of the network mechanisms depicted in

Figure 1.23–6, the consequence of reduced parvalbumin expression in GABAergic neurons is a time-varying excitation–inhibition discoordination. The excitation–inhibition discoordination results in discoordinated neural activity that manifests as aberrant temporal patterns of discharge in neural ensembles as depicted in Figure 1.23–3, disruptions of the kind of spike-field synchrony that is illustrated in Figure 1.23–4, and the discoordination of the various forms of synchrony between oscillations in the field potentials that reflect population synaptic activity. According to the hypothesis, these diverse forms of discoordination cause cognitive impairments exemplified by failures of cognitive control and the disorganization of purposeful goal-directed behavior.

From the systems perspective, it is not obvious that standard pharmacologic interventions could be effective for improving cognitive deficits, at least because the systems view is often that interventions are needed at specific brain locations at specific times, depending on the particular ongoing cognitive effort. More optimistically, as exemplified by the research on the NVHL rat neurodevelopmental model, the systems neuroscience perspective suggests that there is an opportunity to harness experience itself to aid in the location- and time-specific correction and adjustment of dysfunctional neural circuit function. The hope and promise are that as we develop and elaborate knowledge of the detailed recruitment and input-specific activation and suppression of neural circuit operations, it will be possible to design and implement specific experience-based therapies, like cognitive behavioral therapy and computer-based cognitive training programs that harness the mechanisms of synaptic plasticity to tune and adjust dysfunctional neural circuits and in so doing, compensate and perhaps even correct impaired cognitive behavior. Although it is still a hypothesis, there is substantial optimism that if pharmaceutical or brain simulation-based interventions to regulate neuroplasticity can be coupled with the cognition-driven exercise of specific neural circuits, then this combination can be a powerful strategy to use the self-organizing neural consequences of experience to improve or normalize abnormal brain operations. There is much work to do but also reason for considerable optimism because basic systems neuroscience has never had such a powerful and vast toolbox to deploy.

FURTHER READINGS

Antoine MW, Langberg T, Schnepel P, Feldman DE. Increased excitation-inhibition ratio stabilizes synapse and circuit excitability in four autism mouse models. *Neuron*. 2019;101:648–661.e4.

Berry-Kravis EM, Harnett MD, Reines SA, et al. Inhibition of phosphodiesterase-4D in adults with fragile X syndrome: a randomized, placebo-controlled, phase 2 clinical trial. *Nat Med*. 2021;27:862–870.

Buzsáki G. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*. 2010;68:362–385.

Buzsáki G, Anastassiou CA, Koch C. The origin of extracellular fields and currents–EEG, ECoG, LFP and spikes. *Nat Rev Neurosci*. 2012;13:407–420.

Cuthbert BN, Insel TR. Toward new approaches to psychotic disorders: the NIMH Research Domain Criteria project. *Schizophr Bull*. 2010;36:1061–1062.

Dvorak D, Fenton AA. Toward a proper estimation of phase-amplitude coupling in neural oscillations. *J Neurosci Methods*. 2014;225:42–56.

Fenton AA. Excitation-inhibition discoordination in rodent models of mental disorders. *Biol Psychiatry*. 2015;77:1079–1088.

Fernández-Ruiz A, Herreras O. Identifying the synaptic origin of ongoing neuronal oscillations through spatial discrimination of electric fields. *Front Comput Neurosci*. 2013;7:5.

Gandal MJ, Edgar JC, Klook K, Siegel SJ. Gamma synchrony: towards a translational biomarker for the treatment-resistant symptoms of schizophrenia. *Neuropharmacology*. 2012;62:1504–1518.

Gurney ME, Cogram P, Deacon RM, Rex C, Tranfaglia M. Multiple behavior phenotypes of the fragile-X syndrome mouse model respond to chronic inhibition of phosphodiesterase-4D (PDE4D). *Sci Rep*. 2017;7:14653.

Hamm JP, Peterka DS, Gogos JA, Yuste R. Altered cortical ensembles in mouse models of schizophrenia. *Neuron*. 2017;94:153–167.e8.

Insel TR. Rethinking schizophrenia. *Nature*. 2010;468:187–193.

Kao HY, Dvorˇák D, Park E, Kenney J., Kelemen E, Fenton AA. Phencyclidine discoordinates hippocampal network activity but not place fields. *J Neurosci*. 2017;37:12031–12049.

Karayiorgou M, Morris MA, Morrow B, et al. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc Natl Acad Sci U S A*. 1995;92:7612–7616.

Kelemen E, Fenton AA. Dynamic grouping of hippocampal neural activity during cognitive control of two spatial frames. *PLoS Biol*. 2010;8:e1000403.

Laughlin RB, Pines D, Schmalian J, Stojkovic BP, Wolynes P. The middle way. *Proc Natl Acad Sci U S A*. 2000;97:32–37.

Le Van Quyen M, Bragin A. Analysis of dynamic brain oscillations: methodological advances. *Trends Neurosci*. 2007;30:365–373.

Lee H, Dvorak D, Kao HY, Duffy AM, Scharfman HE, Fenton AA. Early cognitive experience prevents adult deficits in a neurodevelopmental schizophrenia model. *Neuron*. 2012;75:714–724.

Lee SH, Ripke S, Neale BM, et al; Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013;45:984–994.

Lieberman JA, Stroup TS, McEvoy JP, et al; Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med*. 2005;353:1209–1223.

Liu Q, Ganzetti M, Wenderoth N, Mantini D. Detecting large-scale brain networks using EEG: impact of electrode density, head modeling and source localization. *Front Neuroinform*. 2018;12:4.

Mitchell KJ, O'Donnell P, Durstewitz D, et al. A framework for the use of models in schizophrenia. In: Silverstein SM, Moghaddam B, Wykes T, eds. *Schizophrenia: Evolution and Synthesis*. MIT Press; 2013:212–226.

Nunez PL, Srinivasan R. *Electric Fields of the Brain: the Neurophysics of EEG*. Oxford University Press; 2006.

Park E, Dvorak D, Fenton AA. Ensemble place codes in hippocampus: CA1, CA3, and dentate gyrus place cells have multiple place fields in large environments. *PLoS One*. 2011;6:e22349.

Phillips WA, Silverstein SM. Convergence of biological and psychological perspectives on cognitive coordination in schizophrenia. *Behav Brain Sci*. 2003;26:65–82;discussion 82–137.

Piatkevich KD, Bensussen S, Tseng H-A, et al. Population imaging of neural activity in awake behaving mice. *Nature*. 2019;574:413–417.

Sigurdsson T, Stark KL, Karayiorgou M, Gogos JA, Gordon JA. Impaired hippocampal-prefrontal synchrony in a genetic mouse model of schizophrenia. *Nature*. 2010;464:763–767.

Steinmetz NA, Aydin C, Lebedeva A, et al. Neuropixels 2.0: a miniaturized high-density probe for stable, long-term brain recordings. *Science*. 2021;372:eabf4588.

Subramaniam K, Luks TL, Fisher M, Simpson GV, Nagarajan S, Vinogradov S. Computerized cognitive training restores neural activity within the reality monitoring network in schizophrenia. *Neuron*. 2012;73:842–853.

Talbot ZN, Sparks FT, Dvorak D, Curran BM, Alarcon JM, Fenton AA. Normal CA1 place fields but discoordinated network discharge in a Fmr1-null mouse model of fragile X syndrome. *Neuron*. 2018;97:684–697.e4.

Thomazeau A, Bosch M, Essayan-Perez S, Barnes SA, De Jesus-Cortes H, Bear MF. Dissociation of functional and structural plasticity of dendritic spines during NMDAR and mGluR-dependent long-term synaptic depression in wild-type and fragile X model mice. *Mol Psychiatry*. 2021;26:4652–4669.

Thomson SR, Seo SS, Barnes SA, et al. Cell-type-specific translation profiling reveals a novel strategy for treating fragile X syndrome. *Neuron*. 2017;95:550–563.e5.

Weinberger DR. On the plausibility of "the neurodevelopmental hypothesis" of schizophrenia. *Neuropsychopharmacology*. 1996;14:1S–11S.

Ylinen A, Soltész I, Bragin A, Penttonen M, Sik A, Buzsáki G. Intracellular correlates of hippocampal theta rhythm in identified pyramidal cells, granule cells, and basket cells. *Hippocampus*. 1995;5:78–90.

Zaremba JD, Diamantopoulou A, Danielson NB, et al. Impaired hippocampal place cell dynamics in a mouse model of the 22q11.2 deletion. *Nat Neurosci*. 2017;20:1612–1623.

1.24 Computational Modeling Approaches to Psychiatry

JOHN D. MURRAY, PH.D., AND ALAN ANTICEVIC, PH.D.

A primary goal of psychiatry is to understand the neural bases and mechanisms underlying mental disorders. Mechanistic understanding is integral to the development of new diagnostic and therapeutic approaches. A great challenge faced by the field of psychiatry is bridging the vast explanatory gaps across levels of analysis: from neurons to circuits to complex human behavior. At present, there is a lack of unifying understanding across this computational hierarchy for any psychiatric symptom. This profoundly limits the field's capacity to develop novel and rationally guided therapies that are grounded in well-established neural mechanisms.

It is no longer controversial to conceptualize severe mental illness as brain-based disease. The explosion of research in neuroscience has established that the brain is essentially a massively complex, nonlinear computational device that governs the expression of an organism's behavior. This perspective is also captured by the ongoing paradigm shift within the field of psychiatry, from categorical behavioral psychiatric diagnoses toward an understanding of dimensional disruptions in complex behaviors that map onto well-defined neural substrates. This perspective is not intended to argue for a reductionist model of complex mental illness. Critically, this view does not posit that psychiatric disorders cannot emerge as a result of external causal factors or perturbations (e.g., long-term stress or extreme trauma) that affect the brain. Nevertheless, such perturbations ultimately impact a set of neural processes and computations that govern a behavioral response (e.g., fear expression). Put simply, there is an organ that psychiatrists are attempting to treat—namely, the central nervous system.

Given the complexity of the brain and the vast number of possible disruptions across distinct levels of analysis, the field of psychiatry should attempt to harness all available tools to mechanistically understand abnormal behavior, regardless of whether the upstream cause lies in genes or external stressors. One promising way to link these levels is to combine empirical findings from clinical neuroscience with approaches from computational and theoretical neuroscience in an emerging field often called “computational psychiatry.” The key logic here is to use mathematical principles and formalism to generate consistent, rigorous, and testable hypotheses that can lead to a better understanding of mechanisms across levels of analyses.

In that sense, computational psychiatry offers the opportunity to interface with experimental and treatment studies at the level of neurons, neural systems, and, ultimately, abnormal behavior (Fig. 1.24–1). Put differently, disrupted mechanisms that contribute to a particular psychiatric disease can occur at the level of neurons and synapses, whereas psychiatric symptoms express at the level of behavior, which involves large-scale brain networks. Linking these disparate levels is vital for gaining mechanistic insight into mental illness and for the rational development of pharmacologic treatments, which act at the synaptic level. The primary objective of computational psychiatry is to help bridge this existing gap and to mathematically formalize computations across noted levels of analysis.

To date, mathematical models of neural systems have contributed greatly to our understanding of how neural dynamics and behavior arise from underlying mechanisms. Building on this mechanistic understanding, disruptions in the underlying elements of the model can then be linked to neural and behavioral dysfunction. In particular, researchers can evaluate which models or model parameters better describe empirical findings in a disease state compared to controls. By examining these disruptions within the models, we can describe key differences in underlying processes in a mathematically precise way. This approach can also test hypotheses about lower-level phenomena (e.g., specific synaptic disruptions) with experimentally accessible noninvasive measures at higher levels in humans (e.g., functional magnetic resonance imaging [fMRI] or psychophysical behavior).

Indeed, in the past several decades, the field of psychiatry has profited tremendously from the ongoing development of noninvasive neuroimaging technology, more sophisticated human cognitive neuroscience, as well as advances in basic neuroscience. The combination of these approaches has directly impacted the broad evidence base in psychiatry, providing support for neural alterations across psychiatric conditions. Such evidence for neural alterations in psychiatric disease has emerged from a wide range of experimental modalities. Analysis of postmortem brain tissue reveals differences in neuronal and synaptic structure across psychiatric disorders. Genetic analyses implicate specific neuronal and synaptic alterations. Neurochemical alterations can be assessed *in vivo* using noninvasive methods such as positron emission tomography (PET) and magnetic resonance spectroscopy (MRS). The spatiotemporal properties of brain activity during various tasks and at rest can be measured with noninvasive neuroimaging, such as fMRI, electroencephalography (EEG), and magnetoencephalography (MEG). Neuroimaging can also reveal structural differences in grey matter distribution and long-range white-matter pathways. As noted, these neural measurements are complemented by rapid and parallel advances in cognitive neuroscience that allow us to quantitatively characterize, at the behavioral level, cognitive, affective, and perceptual impairments in patients.

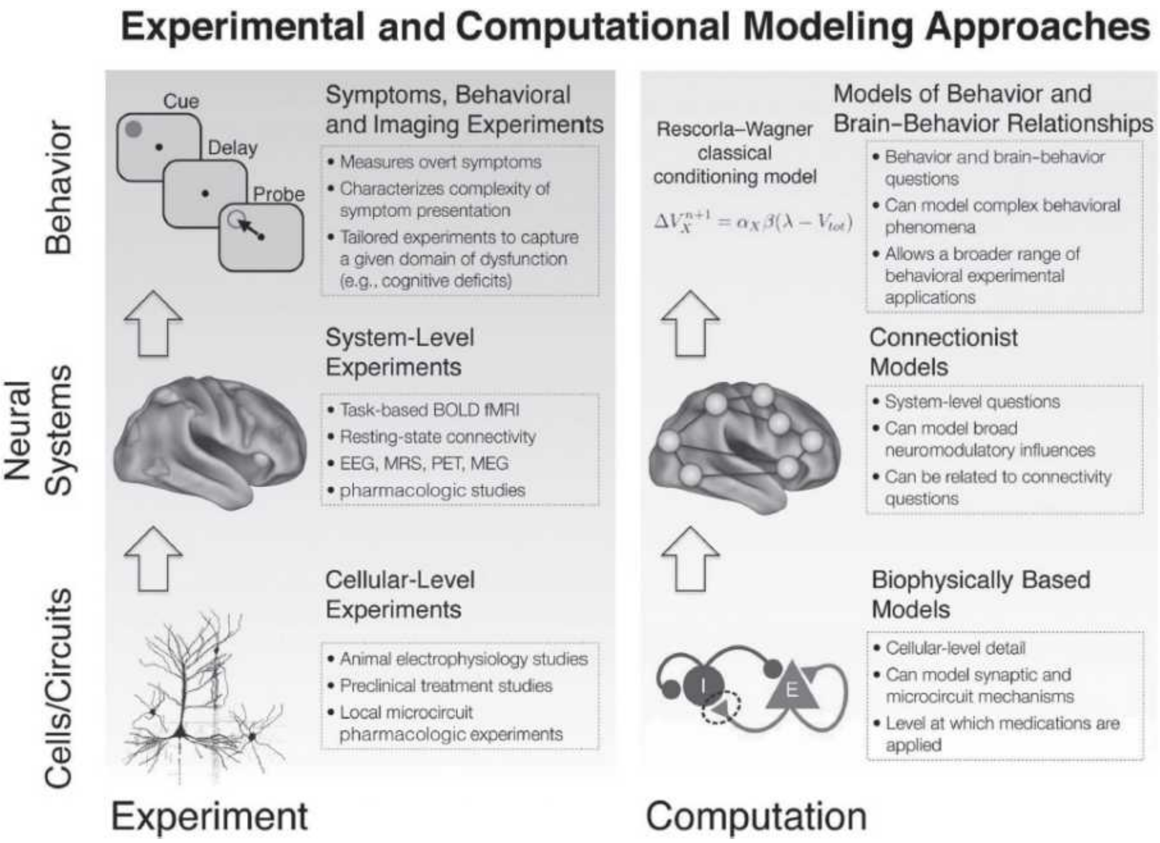
Experimental and Computational Modeling Approaches
Behavior
Neural Systems
Cells/Circuits
Cue
Delay
Probe
Symptoms, Behavioral and Imaging Experiments
• Measures overt symptoms
• Characterizes complexity of symptom presentation
• Tailored experiments to capture a given domain of dysfunction (e.g., cognitive deficits)
System-Level Experiments
• Task-based BOLD fMRI
• Resting-state connectivity
• EEG, MRS, PET, MEG
• pharmacologic studies
Cellular-Level Experiments
• Animal electrophysiology studies
• Preclinical treatment studies
• Local microcircuit pharmacologic experiments
Rescorla–Wagner classical conditioning model
$\Delta V_X^{n+1} = \alpha_X \beta(\lambda - V_{tot})$
Models of Behavior and Brain–Behavior Relationships
• Behavior and brain–behavior questions
• Can model complex behavioral phenomena
• Allows a broader range of behavioral experimental applications
Connectionist Models
• System-level questions
• Can model broad neuromodulatory influences
• Can be related to connectivity questions
Biophysically Based Models
• Cellular-level detail
• Can model synaptic and microcircuit mechanisms
• Level at which medications are applied
I
E
Experiment
Computation

FIGURE 1.24–1. **Conceptual illustration of the computational modeling and experimental interplay across levels of analysis**. The utility of computational modeling, particularly in the study of schizophrenia, is its ability to inform a given level of experimental study. Because we study abnormalities in schizophrenia from the cellular level at the neural system level, and ultimately at the level of behavior, we have to utilize our modeling approaches to best fit the experimental framework. For instance, cellular-level experiments use techniques and produce measurements that are best captured using models that contain the necessary level of biophysical realism (**bottom panels**). Such models can, for instance, inform synaptic processes that may govern the microcircuit phenomena under study such as neural oscillations. In turn, a number of neuroimaging studies have focused on characterizing system-level disturbances in schizophrenia both using task-based paradigms and resting-state functional connectivity approaches. To best inform such system-level cognitive neuroscience experiments, models should capture the relevant detail and complexity of larger-scale neural systems (**middle panel**). Such models can perhaps better inform the role of systemic pharmacologic manipulations on BOLD fMRI or can be used to predict results of functional connectivity studies in schizophrenia. Finally, schizophrenia produces complex and devastating behavioral symptoms, which can be measured via increasingly sophisticated behavioral paradigms. Here, the use of models that formalize complex behavior can provide a powerful tool to quantitatively examine a given behavioral process in patients (e.g., reinforcement learning).

Despite this explosive technologic and experimental progress, the field of psychiatry has yet to produce systematic and mechanistic disease models that are amenable to the rational development of therapies. As noted, to achieve such a multilevel understanding of psychiatric diseases, the field of psychiatry must ultimately link these levels of analysis: from synapses to cells, to neural circuits, to large-scale systems, and to behavioral disturbances. The emerging field of computational psychiatry is in a unique position to help provide the theoretical foundation needed to close these gaps. computational psychiatry benefits from the diverse range of complementary modeling approaches within the broader field of computational neuroscience. Different types of models are best suited to address different questions and different types of experimental data. In that sense, the level of computational modeling is matched to the appropriate levels of experimental inquiry, enabling the interplay between experiment and theory. A multilevel understanding of complex abnormalities in behavior (i.e., psychiatric symptoms) directly benefits from these complementary modeling approaches (Fig. 1.24–2).

This chapter describes three broadly defined levels of computational modeling. (1) Biophysically based models, which are constrained by realistic physiologic properties of synapses, neurons, and circuits. These models are well suited for testing synapse-level hypotheses and allowing comparison to physiologic and pharmacologic studies in psychiatry. This type of modeling is also well suited to study neurophysiologic biomarkers of disease states. (2) Connectionist models use more abstract neural elements but incorporate a systems-level neural architecture. These models can be applied to more complex behaviors and are also well suited to interface with task-based neuroimaging and behavior—namely cognitive neuroscience approaches. (3) Mathematical models of behavior formalize the algorithms underlying cognitive computations, which can, in turn, formalize behavioral symptoms and cognitive deficits. These models are well suited to quantitatively fit behavior and symptoms. Collectively, this chapter provides a conceptual overview of the current state of the art in computational psychiatry, spanning from local neuronal circuits to complex behavior such as reinforcement learning (RL). Finally, ongoing challenges and future directions in computational psychiatry are discussed.

BIOPHYSICALLY BASED CIRCUIT MODELS

Biophysically based circuit modeling refers to a computational framework for modeling neural systems that are grounded in neurobiologic mechanisms. These models can simulate neural circuit activity and computation at the levels of neurons and synapses, constrained by detailed physiologic characterization of neuronal and synaptic dynamics. These models can be further constrained with systems-level experimental measurements of circuit architectures, neural activity, and behavior in tasks. In these models, dynamic activity is typically simulated through systems of differential equations governing the biophysical properties of neurons, synapses, and ion channels. Patterns of activity across neurons are shaped by connectivity within the circuit along with external inputs, which are also grounded in basic anatomy principles. This allows such models to be tested with empirical measures of neural activity. In certain models, behavioral responses can be read out from the patterns of activity, allowing comparison with behavioral data from various task paradigms.

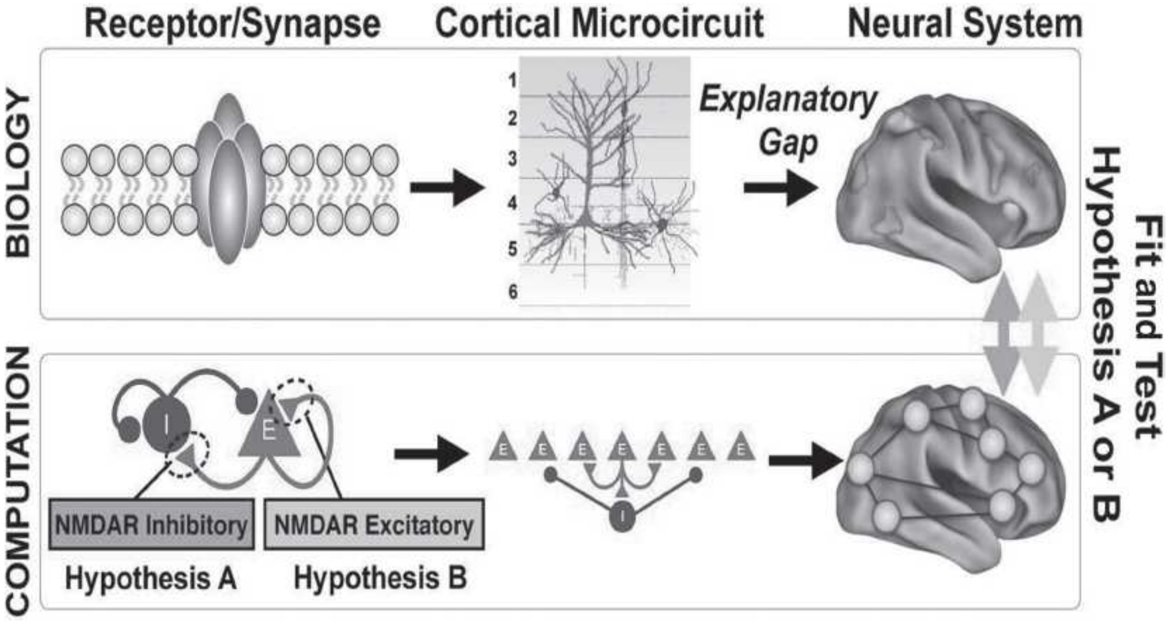
Receptor/Synapse
Cortical Microcircuit
Neural System
BIOLOGY
1
2
3
4
5
6
Explanatory Gap
COMPUTATION
I
E
NMDAR Inhibitory
NMDAR Excitatory
Hypothesis A
Hypothesis B
Fit and Test Hypothesis A or B

FIGURE 1.24–2. Multilevel analysis of hypothesized disruption informed by computational psychiatry. While current noninvasive neuroimaging methods offer major insights into neural system dysfunction in neuropsychiatric conditions, they cannot reveal synaptic or neuronal-level mechanisms. This leaves important explanatory gaps between neuroimaging systems-level observations and cellular-level hypotheses, such as those proposing cellular alterations on specific receptor subtypes. Computational psychiatry aims to help close these gaps by generating testable and translatable hypotheses across distinct levels of inquiry.

In computational psychiatry, the biophysical basis of circuit models provides opportunities to probe how synapse-level disruptions produce aberrant neural activity and deficits in cognition and behavior. For instance, one can perturb specific synaptic parameters in the model that are physiologically interpretable and implicated in hypothesized disease mechanisms. This allows the direct implementation of hypothesized perturbations related to disease mechanisms and pharmacologic manipulations. One can then characterize the impact of these perturbations on emergent neural activity and behavior. Therefore, the model results can be related to human neuroimaging and psychophysics in healthy and clinical populations, as well as in animal models of disease states. Biophysically based circuit models have the specific potential to inform our understanding and design of pharmacologic therapies because they incorporate details of synaptic and neuronal properties that are altered by psychiatric medications.

Despite these inherent advantages, due to the level of detail in biophysically based modeling, it remains a challenge to relate circuit models to specific complex cognitive functions. For some core cognitive computations, we have knowledge of the neural circuit basis underlying these processes, which typically involve contributions from preclinical animal studies. For these cases, detailed circuit models can be developed rigorously to provide the link from synaptic disruptions to behavior (e.g., cognitive deficits discussed below). However, due to the large number of physiologic parameters, these models are more difficult to fit quantitatively to behavioral performance. In other cases, psychiatric symptoms relate to complex functions for which we lack understanding of the underlying neural circuit processes. At present, these circuit models are limited and cannot be applied to complex behavioral tasks, for which we lack understanding of neural circuit correlates.

Biophysically based circuit models can also provide insight into the mechanisms underlying neurophysiologic biomarkers of disease states (see section below). These biomarkers can include electrophysiologic measurements of aberrant neural oscillatory dynamics, and neuroimaging characterization of large-scale dysconnectivity in resting-state activity. Put differently, some neural measures are directly translatable from cells, to neural systems and from animals to humans (e.g., electrophysiologic signatures measured via single electrodes, electrode arrays or via electroencephalography in humans). In such cases, biophysically based models can provide testable predictions that can be readily translated from animal to human experiments.

Linking Synapses to Cognition: Working Memory

An important application of biophysically based neural circuit models to cognitive processing has been in the topic of working memory. Working memory refers to the ability to encode and transiently maintain or manipulate information internally, in the absence of sensory stimulation (e.g., remembering a phone number for a few seconds). This neural computation is fundamental for normal expression of behavior and relies on the delicate balance of local microcircuit neurons. Impairment in working memory is a prominent cognitive deficit in a number of psychiatric and neurologic disorders, including schizophrenia and attention deficit hyperactivity disorder.

Working memory function is also a promising candidate in clinical neuroscience as an endophenotype, a quantitatively measurable core trait that is intermediate between genetic risk factors and a psychiatric disorder. Such transdiagnostic endophenotypes may be more tractable for research into mechanistic underpinnings of symptom dimensions. In that sense, working memory is a useful test bed for applying biophysically based modeling to cognitive function, because there are experimental paradigms that provide multilevel data that can constrain and test models.

Neural Circuit Basis of Working Memory. Neuroscientific studies in animals and humans have revealed important neural circuits and systems involved in working memory. Working memory engages a distributed network of cortical and subcortical areas, with prefrontal cortex as a key node in this network. Single-neuron recording experiments in monkeys have found that a neural correlate of working memory in the prefrontal cortex is stimulus-selective persistent activity, in which neurons show elevated firing rates during the mnemonic delay. For instance, in one well-studied experimental paradigm, the oculomotor delayed response task, the subject must maintain in working memory the spatial location of a visual cue across a delay period to guide a saccadic eye movement toward that location. During the mnemonic delay, a subset of prefrontal neurons exhibits tuned persistent activity patterns, with single neurons firing at elevated rates for a preferred spatial location. This paradigm has been successfully extended to human studies and extensively tested via functional neuroimaging, confirming the neural correlates of working memory in humans. In that sense, this complex cognitive behavior is an ideal example where the synaptic, neuronal, system-level, and behavioral mechanisms are well linked in both animal and human studies. We focus here on biophysically based models of working memory that can offer specific insights into the circuit mechanisms causing disruptions in this complex behavior.

Attractor Network Models. A class of circuit models called *attractor networks* provides a leading framework for understanding the neural circuit basis of persistent activity supporting working memory. The term *attractor* comes from the mathematical formalism of dynamical systems: an attractor state is an activity pattern that is stable in time so that after a small perturbation to the state of the system, the system converges back to the attractor state. In attractor network models, these dynamics are typically formalized by a set of differential equations that capture the neuronal dynamics over time. An attractor network model of working memory typically possesses multiple attractor states: a low-activity baseline state and multiple stimulus-selective memory states in which a subset of neurons are persistently active. Because the memory state is a stable attractor state, it is self-reinforcing and resistant to noise or perturbation, allowing the memory to be maintained over time. This critical property of such models has been shown to rely on the slow currents mediated via the N-methyl-D-aspartate receptors (NMDARs) on excitatory pyramidal cells. It is important to highlight here that it is precisely this level of detail in biophysically based models (i.e., NMDAR conductance) that allows for perturbing and testing of specific synaptic contributions to the neural circuit or, ultimately, behavior (i.e., working memory performance).

In a typical attractor network, subpopulations of neurons are selective to different stimuli. Neurons of similar selectivity show a Hebbian pattern of strong recurrent excitatory connections between them. When the strength of recurrent excitatory connections is strong enough, they can support a persistent-activity memory state. Strong, structured recurrent excitation sustains the persistent activity. Inhibitory GABAergic interneurons mediate strong feedback inhibition that serves two roles. First, it stabilizes the low-activity baseline state. Second, during persistent activity, they provide lateral inhibition that enforces selectivity of the working memory representation, preventing the spread of excitation to the entire population of neurons in the network. The stability of attractor dynamics, which depends on excitation and inhibition, provides the network with a natural resistance against noise and distractors that intervene during the delay (i.e., a stable attractor state). This robustness of working memory against distraction is a key cognitive property that supports cognitive processing.

Detailed biophysically based models have revealed important constraints on circuit function and made specific predictions that are experimentally testable. Strongly recurrent networks are prone to oscillations, which can destabilize persistent activity patterns and destroy working memory function. As noted, a decade of computational modeling work has predicted that persistent activity can be stabilized if recurrent excitation is mediated primarily by slow NMDAR rather than by fast alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors to provide temporal stability of the delay activity. A number of model predictions have been tested through studies that record the activity of single neurons while focally applying pharmacologic agents localized to the cortical tissue surrounding that neuron. In line with model predictions, it was experimentally established that locally blocking NMDA receptors can abolish delay-period persistent activity and that locally blocking gamma aminobutyric acid receptors leads to decreased stimulus selectivity during the delay through elevated firing. These experimental tests support the critical roles of balanced excitation and inhibition in the prefrontal cortex in supporting working memory function, a neural property that is hypothesized to be affected by severe psychiatric conditions (e.g., schizophrenia).

Dysfunction in Attractor Models. A number of computational modeling studies have analyzed how attractor network function is impacted by alterations of synaptic and neuronal properties. Optimal cortical function depends on the proper balance of excitation and inhibition (E/I balance). Disruption of E/I balance in the cortex is implicated in the neuropathology of a number of psychiatric disorders, including schizophrenia, autism, and attention deficit hyperactivity disorder. As described above, attractor models require strong recurrent interactions between pyramidal neurons and interneurons and are, therefore, sensitive to perturbations of E/I balance. Because these models also implement the core cognitive process of working memory, which is known to be disrupted in psychiatric disorders, attractor models provide a useful test bed to link disrupted E/I balance to cognitive deficits and generate critical experimental predictions across levels of analysis (see Fig. 1.24–2).

These models have been most often applied to circuit disruptions implicated in schizophrenia, for which a number of cortical alterations are proposed. Although neuronal microcircuit alterations in schizophrenia are complex, many studies have converged on interneuron dysfunction as a key component of its pathophysiology. Impaired interneuron function can result in weakened recruitment of feedback inhibition, causing disinhibition of pyramidal neurons and an elevated E/I balance in the circuit. Computational models have explored the neural and cognitive effects of disinhibition in working memory circuits. A mild disinhibition can induce a broadening of working memory representations, which can be understood from the role of inhibition in maintaining the selectivity of persistent activity patterns. A strong disinhibition can destabilize the baseline state, causing working memory patterns to spontaneously appear (e.g., false memory representations). Importantly, these distinct regimes make dissociable predictions for performance in working memory tasks, which allows testing of these hypotheses with behavior. In addition to inhibitory deficits, there is also strong evidence that recurrent excitation is compromised in the prefrontal cortex in schizophrenia. In the attractor models, reductions in recurrent excitation can abolish persistent activity, preventing the circuit from implementing working memory. This regime again makes specific testable predictions for behavior and patterns of neural activity in the context of working memory tasks.

Attractor models have also been applied to study the effects of neuromodulation on working memory. Neuromodulators are known to alter the effective conductances of a number of synaptic receptor and neuronal channels. Dopamine (DA) has diverse effects on prefrontal cortical circuits, modulating multiple cellular and synaptic sites on both pyramidal cells and interneurons. DA also modulates working memory function, and DA signaling is found to be dysregulated in a number of psychiatric disorders. Numerous lines of evidence support the hypothesis that in schizophrenia, there is impairment of DA signaling, via the D1 receptor, in the prefrontal cortex. One circuit modeling study found that D1 stimulation increases the stability of both baseline and memory states in the network, which would be functionally beneficial during working memory maintenance. This, in turn, predicts that a hypodopaminergic state in the prefrontal cortex impairs working memory by weakening persistent activity and thereby rendering it vulnerable to disruption by noise or distraction. In contrast, an aberrantly hyperdopaminergic state could cause attractor states to be too strong, rendering the network inflexible to signals that should normally induce transitions between internal states.

In turn, biophysically based models can also be scaled to the level of neural systems. Extending biophysically based models in this way can generate experimental predictions for neuroimaging experiments that can be tested via causal pharmacologic manipulations in humans or during cognitive tasks in psychiatric populations. More specifically, one of the most established findings coming out of human neuroimaging experiments implicates an “anticorrelated” neural architecture. There are brain regions specifically involved in orienting to an externally demanding cognitive task (i.e., task-positive areas) or areas that are more active when engaged in internal mental states but suppressed during cognitive demand (i.e., task-negative areas or sometimes referred to as the default mode network, DMN). Studies have found that during demanding cognitive tasks, patients with schizophrenia typically cannot suppress task-negative areas and concomitantly cannot activate task-positive regions—a process also shown to occur following NMDAR antagonist administration in healthy individuals during working memory. Biophysically based microcircuit computational models have been expanded to capture the putative reciprocal antagonism between task-positive and task-negative neural systems in a biologically plausible fashion by incorporating mutual synaptic inhibition. In this way, such models can generate system-level predictions for task-based neuroimaging experiments in humans while still capturing the relevant synaptic detail. Nevertheless, as noted previously, these models are still limited in the repertoire of complex behaviors they are able to capture due to the inherent constraint placed by biologic realism in such models. Therefore, this branch of computational psychiatry is not yet positioned to address the range of complex cognitive and affective computations that can be represented in humans. This is, at present, much better accomplished by another type of modeling—namely connections models—which we turn to next.

COGNITIVE PSYCHIATRIC NEUROSCIENCE AND CONNECTIONIST MODELS

The sections above introduce the interface between biophysically based computational modeling and clinical neuroscience. As noted, this branch of modeling is distinguished by building on the physical properties of real neurons and synapses, producing realistic cellular-level dynamics. This level of detail is ultimately vital to generate mechanistic predictions at the level of treatments or to elucidate synaptic mechanisms governing complex behaviors in psychiatric patients. Nevertheless, a parallel line of computational psychiatry research has emerged in lockstep with the cognitive neuroscience revolution, building on the rapid and productive advances in noninvasive neuroimaging technology. Specifically, this branch of modeling can be broadly referred to as *connectionist models* that are designed to capture the function of large-scale neural networks and generate behavioral predictions in conjunction with human cognitive neuroscience experiments.

Emergence of Cognitive Neuroscience in Psychiatry

Before detailing the contributions that connectionist modeling has made for psychiatry, it is important to set the stage for the type of research questions these models are best suited for—namely cognitive neuropsychiatry. The past two decades have witnessed an explosive growth of knowledge regarding the neural correlates of various cognitive and affective processes in healthy individuals. This burgeoning field of cognitive neuroscience has generated an increasingly robust platform for interpreting clinical neuropsychiatric phenomena. This is primarily accomplished by garnering an increased understanding of neural systems known to be involved in various cognitive operations in healthy individuals. This understanding, in turn, constrains our search space for what aspects of brain circuitry may be abnormal in clinical populations exhibiting deficits in these same cognitive operations. Therefore, using a cognitive neuroscience framework with a specific application toward understanding computational mechanisms offers a promising tool for elucidating and ultimately treating psychiatric symptoms by delineating abnormalities in neural circuits whose functions are increasingly understood in healthy populations.

This cognitive neuroscience framework has emerged in parallel with the development of computational modeling in psychiatry. Historically, the types of computational psychiatry models that first appeared were most directly aligned with the neural systems and behavioral levels of analyses.

Therefore, there exists a productive and ongoing interplay between cognitive neuroscience in psychiatry and connectionist modeling approaches.

Connectionist and Parallel Distributed Processing Models

There is a long history of using models to understand central nervous system function. Such a review is beyond the scope of this chapter, and we refer the reader to one of the references at the end of the chapter. However, briefly, it is worth considering the long-standing analogy of the brain as a “computer.” The growing sophistication of algorithms and principles in basic computer science has cross-fertilized the emerging application of such approaches in neuroscience. One particular framework that has emerged out of this effort is the modeling of neural networks. More formally, models of this class are called parallel distributed processing (PDP) or connectionist models. This broad class of models has emerged in an effort to elucidate the neural correlates of complex psychological functions, including cognitive control, language, and fear processing.

Importantly, these models are distinct in their flavor, architecture, and algorithmic implementation from the biophysically based (or behavioral) models reviewed in this chapter. The key distinction is represented in their ability to incorporate some level of abstraction beyond strict biologic realism. A major objective of PDP models is to develop a deeper understanding of the mechanisms driving cognition from a psychological perspective, but using principles thought to capture aspects of computations as they might be carried out in the brain. Put differently, such models do not necessarily build in explicit neurobiologic constraints or assumptions.

A typical PDP model architecture is conceptualized as a series of “units” that are interconnected in some fashion and that represent some computational process. Each unit implements a simple transfer function that summates its inputs from other model units and transforms that into an output activity level. The state of the network is described by the graded activity patterns over populations of these simple units, and information processing occurs through the transformation of activity patterns from one set, or layer, of units to another. To achieve the desired information processing, the connection strengths between units are modified through a learning algorithm, such as *back propagation*. This training of connection weights is meant to capture some elements of the type of Hebbian learning known to govern changes in connection strength between neurons.

In the back-propagation learning algorithm, the network is presented with an input pattern, and the output layer is allowed to generate its response. At each layer of the network, an error signal is computed as the difference between the response generated by the network and the desired target response. These error signals are propagated back through the network toward the input layer, used to modify the connection weights between units. The implementation of back-propagation learning provides an example of how connectionist models, while capturing complex neural phenomena, may not map onto specific neurobiologic mechanisms (in contrast to biophysically based models described above). Nevertheless, this approach captures the idea that learning occurs through an incremental minimization of error by small changes in the connections between units (a phenomenon referred to as *gradient descent*).

In typical PDP models, the units are abstracted, without a direct correspondence to individual neurons or explicit neurobiologic assumptions. Units instead represent processing accomplished by neuronal assemblies and capture the general principles of neural information processing. However, some modeling studies have incorporated neuroanatomical constraints and specific neurotransmitter modifications, at varying levels of resolution.

Connectionist models help to build bridges between understanding of low-level properties of neural systems, and their participation in higher-level (systems) behavior. They are able to capture a wide range of complex behaviors and neural systems interactions, which cannot at present be modeled as effectively by biophysically based models. Therefore, PDP models continue to be successfully applied in computational psychiatry to address questions and scenarios in which the field lacks firm empirical evidence to inform strong biologic assumptions.

Applications of Connectionist Models in Psychiatry

This section presents three example areas of research where PDP models have provided novel insights and established a vital interplay between basic neuroscience and psychiatric questions. First, the focus is placed on how PDP models have been used to understand the formations of hallucinations and delusions in schizophrenia. In turn, we review the role of PDP models in understanding cognitive control deficits, which cut across many psychiatric conditions. Third, we discuss the use of connectionist models in the context of fear conditioning, specifically from a systems neuroscience perspective, with possible implications for understanding posttraumatic stress disorder (PTSD).

Connectionist Modeling of Hallucinations in Schizophrenia. Some of the seminal models using the PDP framework were developed to examine the impact of abnormal synaptic pruning in cortex and its possible consequences on symptoms observed in schizophrenia, specifically formation of hallucinations, which are abnormal internal sensory experiences in the absence of external sensory input. This mechanism is based on the neurobiologic conceptualization of schizophrenia as a neurodevelopmental illness associated with abnormalities of synaptic pruning over the course of development, which in turn impacts the functioning of large-scale neural systems.

PDP models implemented this mechanism by eliminating the connections between units thought to represent synaptic connections. This line of research showed that if a sufficient number of connections were removed, then the model would repeatedly return to an abnormal state referred to as “parasitic foci.” That is, researchers observed patterns of activity across units in the model that did not correspond to any learned memory, but to the “state” to which the model kept returning—perhaps capturing an abnormal internally generated state that could occur during hallucinations. Researchers in this area hypothesized that these parasitic foci may behaviorally map onto delusions and hallucinations. A concrete example of how PDP models may help explain hallucinations came from applications to understanding language processing. Using the PDP-style architecture, the research found that language could be represented within an “Elman” network—namely, a sequential recurrent network that consists of an input and an output layer connected by a “hidden layer” that computed internal processing. This architecture included an additional layer connected to the hidden layer that could temporarily store information, thought to support working memory functions (temporary storage of information) during speech perception. In this architecture, it was shown that the elimination of connections between units (i.e., “synapses”) in the working memory subnetwork of the model initially improved speech perception. However, as the severity of the elimination increased, the model started to fail to detect words correctly and started to generate “hallucinations.” In this PDP model, this was computationally captured as patterns of activity across the output layer that corresponded to words even when there was no structured input into the model. In contrast to synapse elimination, a small amount of “cell” removal (e.g., elimination of units) also initially improved speech perception, but greater cell loss did not generate hallucinations. Critically, these early PDP models provided an important “proof of concept” for how various neurobiologic changes might affect connectionist networks.

In turn, such PDP models can also incorporate some aspects of broad neuromodulation. For instance, researchers that extended the PDP models of hallucinations in schizophrenia also captured the potential influence of increased DA in the model. This is concretely accomplished as a gain function within the model by abstracting a neurobiologically plausible broad ”neuromodulatory” influence on the entire PDP network. In schizophrenia, the role of DA is hypothesized to play a key role in cortical “tuning.” Based on this hypothesis, PDP models incorporated increased DA by changing the bias onto the units in the working memory module, effectively requiring them to accumulate greater information to maintain information. This manipulation only relates to one aspect of the influence of DA on neuronal information processing and by itself, it did not fully capture clinical observations. However, a combination of increased DA and synaptic pruning more successfully captured the performance of those patients who experienced hallucinations.

These types of models incorporate aspects of brain-style computational principles, but did not make strong commitments to key aspects of neurobiology during model development. Nonetheless, they provided critical steps and a mathematical formalism to understand how neurobiologic or neural network deficits might contribute to the formation of core symptoms of psychosis, such as hallucinations. Although the models themselves did not produce behaviors with the same format and content as shown by human patients, the “in principle” examples provided conceptual leverage to bridge the gap between biology and behavior and even extend to treatment recommendations (e.g., work in the area of transcranial magnetic stimulation that explicitly leveraged insights from these models). Therefore, despite not explicitly committing to biologic realism, PDP models can provide useful predictions for treatment development.

Connectionist Modeling of Cognitive Control Deficits. A parallel line of research leveraged PDP models to understand cognitive control deficits that emerge across psychiatric conditions but are particularly prominent in schizophrenia. These models developed a more neurobiologically plausible architecture by postulating that "cognitive control" in the central nervous system arises from interactions between the DA neurotransmitter system and the PFC. These models suggested that goal-related information, or context information, was maintained in PFC and used to bias stimulus–response mappings represented in the posterior cortex, serving as a source of top-down support for controlling behavior. In these models, the "context" module was associated with the functions of the dorsolateral prefrontal cortex (DLPFC). Active maintenance in the absence of external inputs was assumed to occur via recurrent excitation, as suggested by neurobiologic data that was reviewed in the sections above. This branch of models assumed that feedback projections from the context/DLPFC module biased processing in posterior systems via direct excitation but that competition between representations within a processing stream occurred via lateral inhibition. Further, the effects of DA were assumed to be modulatory rather than purely inhibitory, such that changes in DA activity could either increase or decrease activity depending on the nature of inputs that DA was modulating. Put differently, increases in DA were assumed to increase the signal-to-noise ratio of a unit's activation value in relation to its input.

Concretely, these models have been successfully used to generate predictions about a specific pattern of complex behavioral errors that are observed during cognitive paradigms evaluating context processing deficits in schizophrenia. Specifically, the models were used to test the hypothesis that reductions in DA input into lateral PFC lead to impairments in the ability to represent and maintain context information in lateral PFC, and that this impairment in the representation of context leads to deficits in a range of cognitive tasks among individuals with schizophrenia. Behaviorally, the model predicted that individuals with cognitive deficits should exhibit a specific pattern of behavioral errors when context information needed to be maintained. This pattern of errors was largely confirmed across a series of studies in schizophrenia patients. In turn, this type of PDP model was also successfully leveraged to generate predictions about patterns of lateral PFC signals found during task-based cognitive activation neuroimaging paradigms. The specific prediction from these models suggested that individuals with schizophrenia should show reduced lateral PFC activity in conditions with a strong demand for context maintenance. Again, a number of empirical reports confirmed this prediction, showing that patients with schizophrenia exhibited deficits in recruitment of lateral PFC specifically in line with model predictions. Subsequent work in this area suggested that deficits in DA signaling may contribute to abnormal context maintenance and information gating in lateral PFC, which in part contributes to cognitive processing deficits seen in this patient population.

Collectively, this line of research illustrates how PDP models can be used to make both behavioral and neural predictions of a given behavioral symptom (i.e., cognitive deficits) which can in turn be explicitly tested via cognitive neuroscience paradigms in patients. This line of research has also been successfully leveraged to understand the neural underpinnings of specific affective deficits in psychiatric populations—namely fear conditioning.

Connectionist Modeling of Fear Conditioning. It is well established that severe deficits in emotion processing cut across different diagnostic categories, including schizophrenia, bipolar illness, and anxiety disorders to name a few. Here, we specifically focus on computational modeling studies in a specific area of affect processing—namely fear expression. Abnormal fear expression provides a powerful translational target that can be understood computationally across mammalian species given the well-preserved neurobiologic mechanisms underlying basic fear expression. At the basic preclinical level, the role of the amygdala in fear generation is perhaps one of the best-studied neural computational mechanisms, having been translated across levels of analysis, including human studies. Clinically, abnormal fear expression can be found across a host of anxiety disorders, including generalized anxiety, social anxiety, specific phobias, and PTSD. Therefore, computationally formalizing abnormal fear circuits represents a particularly broad and important challenge for computational psychiatry.

Computational studies in this area have used the amygdala as a central starting point. The amygdala is a small structure comprised of several nuclei located anterior to the hippocampal formation, and is critical for the normal expression of fear and learning of fear-related associations (i.e., classical fear conditioning). In that sense, PDP models in this area build on both well-understood biophysical principles of this neural structure as well as its well-established large-scale anatomical connectivity. This is a good example of how PDP-style models can at times incorporate more explicit commitments to biologically plausible circuits and instantiate biophysical realism while maintaining a system-level architecture with useful behavioral predictions (e.g., fear learning). Specifically, such anatomically constrained PDP models of fear circuitry have been used to generate predictions for how multiple interacting neural structures could work jointly to accomplish a pairing between the unconditioned stimulus (US, nociceptive stimulus) and the conditioned stimulus (CS, auditory input) (see Fig. 1.24–3). Using this model architecture, it was found that fear learning can be established via the joint pattern of activation across specific unit pairs that map the CS and UC. Furthermore, the model provided a powerful way to generate predictions for how focal lesions may affect fear learning—a specific advantage of PDP models that are developed based and constrained via neuroanatomically plausible principles. This line of work in particular can help generate useful predictions for abnormalities in fear expressions that may result from neuropathology to one or more key components of the fear circuit.

Collectively, this work demonstrates how connectionist models can be used creatively and flexibly to generate behavioral and neuroimaging predictions across a wide range of neural systems and behaviors. As noted, this flexible repertoire of PDP-style models is possible because their architecture is not always explicitly developed or constrained based on neurophysiology or neuroanatomy. That said, these models still make a number of strong commitments to represent neural computations. For this reason, PDP models are not always capable of formalizing even more complex behaviors that may occur in psychiatric patients (e.g., formation of complex and abnormal belief patterns such as delusions, or deficits in motivation and reward learning). In such cases where neither PDP models nor biophysically based models can provide the final links to behavior, we turn to mathematical models of behavior that are explicitly developed to formalize the patterns of complex psychiatric symptoms.

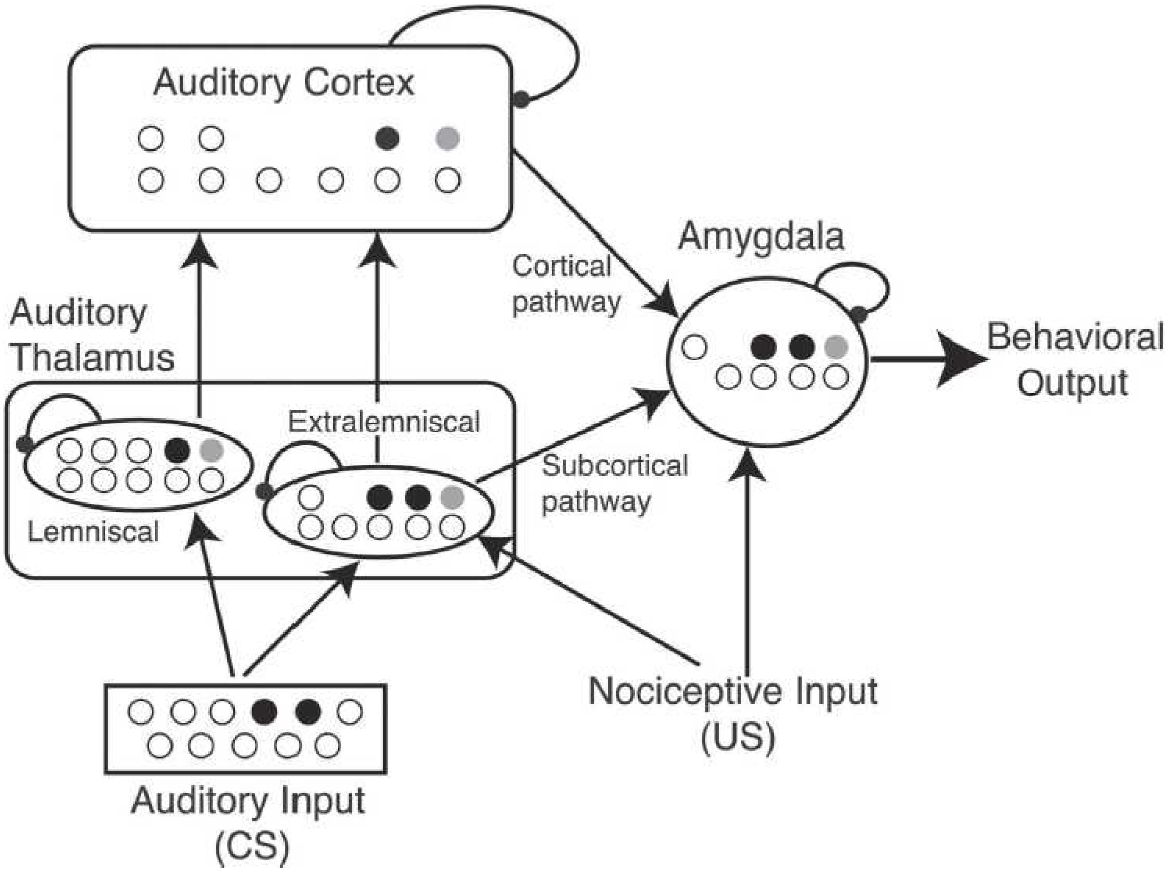
Auditory Cortex
Amygdala
Cortical pathway
Auditory Thalamus
Behavioral Output
Extralemniscal
Subcortical pathway
Lemniscal
Nociceptive Input (US)
Auditory Input (CS)

FIGURE 1.24–3. Illustrative example of a connectionist PDP-style model in the context of fear conditioning. In the case of fear learning, PDP models have successfully merged both connectionist-level architecture across distinct "modules" that represent specific neuroanatomical structures while remaining committed to a certain level of biophysical realism (i.e., by modeling the amygdala based on animal physiology experiments). Here, we show an adapted figure from X. In this exemplar case, each larger module is comprised of individual nonlinear units (black units, maximum activation; white units, zero activation; gray units, intermediate level of activation), as is typically done in PDP models. *Large arrows* denote connections across modules, which are modeled as feedforward and excitatory, corresponding to pathways in the actual rat brain (although they are simplified here). Connection strengths across the modules are modulated as a function of learning by incorporating an extended Hebbian rule. (Reprinted from Armony JL, Servan-Schreiber D, Romanski LM, Cohen JD, LeDoux JE. Stimulus generalization of fear responses: effects of auditory cortex lesions in a computational model and in rats. *Cereb Cortex*. 1997;7(2):157–165. doi:10.1093/cercor/7.2.157)

MATHEMATICAL MODELS OF BEHAVIOR

The computational psychiatry approaches discussed above, biophysically based circuit modeling and connectionist modeling, aim to develop models with elemental components that instantiate how neural systems transform their inputs into outputs. These approaches are often limited in the range of cognitive behaviors to which they can be applied, due to our limited knowledge of the relevant neurophysiologic mechanisms. An alternative approach to modeling cognition is to develop mathematical models that provide an account of the psychological computations underlying a given cognitive function. These models are often grounded in so-called normative accounts that aim to define the algorithms that optimize performance at a task and to characterize constraints that limit such task performance.

Although developed to explain behavior, these models can potentially make important links to neural circuits. For example, experiments can reveal that activity in specific brain areas represents the internal variables of the model, providing support for this computational account and providing insight into the division of labor among brain systems. In turn, such neuroscientific findings can inform the algorithms and modules that compose the computational architecture of the model.

Impairments in decision making are common cognitive symptoms across psychiatric and neurologic disorders. Mathematical models of behavior can formalize these symptoms in rigorous and quantitative terms, potentially isolating specific components of decision-making computations that are altered in a given psychiatric condition. These models typically have fewer parameters than the neural systems models described above, which enables them to be more easily fit to behavioral data or to explore the full range of a given parameter. In that sense, behavioral models can provide a deeper intuition of the full range of a possible behavioral repertoire and help rigorously define the behavioral states that may be abnormal or maladaptive.

Reinforcement Learning

One form of dysfunction in decision making is a disruption in the process through which one learns the values of available actions. Mathematical modeling and experiments have made great progress in studying this process in the context of RL. In RL, an agent learns the values of stimuli and actions through successive trials of feedback (reward and punishment) and selects actions based on these values to maximize one's overall outcomes.

A class of *temporal-difference RL* (TDRL) models has been widely applied to explain the process of learning the mapping from stimuli and states to actions. In these models, actions are selected on the basis of the total expected value for that action. The value of each action is learned through sequential updating based on a *reward prediction error* (RPE). In its simplest form, the RPE is the difference between the reward received after that action and the expected value of the action. Thus, if the reward is greater than expected, a positive RPE signals that the value should be increased. If the reward is equal to that expected, the RPE should be zero. The formulation of the RPE can be extended to incorporate additional factors, such as temporal discounting of future outcomes. TDRL models are well supported by behavioral studies, as well as neuroscientific studies that have found brain areas that signal action values and RPEs. In particular, DA neurons in the subcortical ventral tegmental area (VTA) quantitatively reflect internally computed RPEs through bursts of action potentials.

Many key aspects of addiction can be modeled as dysfunction in a TDRL model. It is thought that drugs of abuse access the natural RL systems in the brain, leading an agent to over-select drug-seeking actions. Specifically, drugs of abuse can intervene in the pathways that activate the DA neurons that are involved in signaling RPEs. For instance, cocaine can induce phasic firing of DA neurons in the VTA. In TDRL, this can interfere with RPE signaling, strongly increasing the learned values for drug-seeking actions so that they come to dominate behavior. Modeling studies have found that implementing this drug-induced component to RPE signaling in a TDRL model can parsimoniously explain a number of important aspects of addiction-related behavior.

The TDRL model described above, which learns the values of actions for a given state, applies to habit-related learning. More elaborate RL models have also been developed to explore the roles of multiple learning systems in the brain. In addition to this habit-related form of learning (also called model-free RL), there is another type called model-based RL. Model-based RL calculates the values of actions from the values of future states, using an internal representation of the causal structure of the environment (i.e., knowing which states are reached through the actions). Model-based RL is more computationally demanding for the agent (but also more flexible) than model-free RL and is associated with cognitive processing in the PFC.

By comparing behavior in RL tasks to model performance, researchers can probe the relative reliance on model-free versus model-based RL. Cognitive deficits may impair the function of model-based RL, leading to an increased reliance on model-free RL. Researchers have designed behavioral paradigms to measure the relative contributions of model-free versus model-based RL in decision-making. Studies have found, both in addiction and in obsessive-compulsive disorders, that there is a favoring of the habitual model-free RL compared to healthy controls. These findings support the view that these disorders exhibit habitual, repetitive behaviors.

Decision Making via Sequential Sampling

Another topic of mathematical modeling of decision making is the process of categorically selecting among available options. This has been most extensively studied in the context of two-alternative forced-choice (2AFC) tasks. In 2AFC tasks, the decision-making process can be formulated as the sequential accumulation over time of evidence in favor of one option or the other. The drift diffusion model (DDM) is a phenomenologic but mechanistic model that is widely applied to 2AFC tasks and is likely the optimal solution in certain contexts.

In the DDM, a decision variable evolves in time, accumulating noisy evidence. Because evidence sampling is inherently noisy, the decision variable undergoes a biased random walk, with a mean drift rate in one direction determined by the net evidence in favor of one option over the other. Computationally, the two choices correspond to a lower and an upper bound; when the decision variable crosses a bound, the corresponding choice is made. Because the noise is different on each decision-making trial, the choice and reaction time vary across trials. One strength of the DDM is that it generates not just a proportion of choices but a full distribution of reaction times that can be used to fit empirical choice data. At the neural level, the DDM has received support from electrophysiologic recordings from a number of cortical areas that show slow ramping of neural activity during decision making.

The DDM can also explain aspects of impaired decision making by describing behavior in terms of the values of DDM parameters. This has the potential to more precisely isolate and quantify differences in decision-making behavior. Consider, for instance, a clinical group that shows a propensity toward selecting one option over the other. In the DDM framework, this could be due to one decision boundary being closer than the other. Alternatively, the mean drift rate may include a component of the drift rate in favor of that option. Critically, these two potential mechanisms have dissociable effects on the patterns of errors and reaction times. Therefore, fitting subjects' behavioral data can isolate which alteration better describes behavior in a group of subjects.

Bayesian Approaches to Inference

Another fruitful area of research is the mathematical formalism of belief formation. This is most clinically relevant in cases where strong but abnormal beliefs can result in internal states that are clinically characterized as delusions. Here, the biggest impact has been achieved by a class of mathematical models of *Bayesian inference*. Bayesian inference describes how the estimated probability of event A should be updated by observation of event B, according to Bayes' theorem:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

where $P(A \mid B)$ is the probability of A given that B is true, $P(A \mid B)$ is the probability of B given that A is true, $P(A)$ is the probability of A, and $P(B)$ is the probability of B. Therefore, the *posterior* $P(A \mid B)$ is proportional to the *likelihood* $P(B \mid A)$ times the *prior* $P(A)$. The prior probability, the expectation for event A to occur, plays a key role in the inference of the posterior estimate from the observation of event B.

In the context of delusion formation, Bayesian models hypothesize that psychosis is a neural and psychological state that is marked by abnormal priors about the world. Abnormal priors may result from a state of aberrant salience, that is, events in the environment are hypothesized to be more salient than usual, perhaps due to elevated DA signaling in the striatum. Bayesian models can, in turn, be applied to a range of behavioral experimental contexts where expectations are formed and then confirmed or violated, with the reported posterior estimates providing a window into the abnormal priors. Studies in this area support the view that abnormal priors (or posterior probability) lead to the formation of abnormal inferences, which could lead to a stable internal belief that is not supported by current evidence (i.e., a delusion). Put simply, Bayesian models of behavior, when applied to psychiatry, have postulated that delusions may be an extreme state of abnormal inference. This framework has been subsequently linked to abnormal prediction error coding and extended to a neuroimaging context, providing a putative neural marker of abnormal belief formation.

MODELING NEUROPHYSIOLOGIC BIOMARKERS

A major goal of clinical neuroscience is the discovery of robust neurophysiologic biomarkers for psychiatric disorders. Understanding the mechanistic basis underlying these biomarkers could aid in the development of treatments, by mapping specific circuit perturbations to neurophysiologic measurements. Biophysically based circuit modeling is well suited for linking these levels, as the models link synapse-level parameters to network-level dynamics. Models can make predictions for how physiologic biomarkers are affected by putative synaptic disruptions. In turn, model predictions provide the basis for inference in the other direction, to test hypotheses regarding the site of a synaptic perturbation with measured changes in physiologic biomarkers. Therefore, biophysically based circuit modeling provides a means to interpret specific biomarkers and generate hypotheses that can be tested further with synapse-level and cellular-level measurements. This understanding allows for interpretation of aberrant neural dynamics measured in patients in terms of circuit perturbations, which generates specific and testable hypotheses for experimental studies in animal models and clinical populations. Two topics studied in this way are cortical oscillations and large-scale resting-state brain networks.

Oscillations

Cortical circuits generate a wide range of oscillatory dynamics that arise from strong recurrent network interactions. Abnormal oscillations have been observed in a number of psychiatric and neurologic disorders, typically measured through EEG or MEG. Such oscillatory deficits are observed both in local dynamics, and in long-range coupling between brain areas. In schizophrenia, oscillatory abnormalities in the gamma range (30 to 80 Hz) are of particular interest. Cortical circuit models, comprised of pyramidal cells and interneurons, have been used to explore the disruption of gamma oscillations in response to perturbations implicated in psychiatric disorders such as schizophrenia.

Computational and experimental studies support a model in which neocortical gamma oscillations arise from a feedback loop of excitatory pyramidal cells reciprocally connected with perisomatic-targeting, fast-spiking (and parvalbumin-expressing) inhibitory interneurons. In these circuit models, gamma oscillations can arise through a patterned cycle of activation of pyramidal cells and interneurons. Excitatory input causes pyramidal cells to fire. This activity then recruits inhibitory interneurons to fire short delay. This feedback inhibition suppresses firing in the circuit. After this synaptic inhibition decays, the cycle begins again. Models show that such oscillatory behavior arises naturally in systems with strong recurrent excitation and inhibition if the feedback inhibition is slow or delayed with respect to the excitation. As in the case of working memory function described above, gamma oscillations depend on strong recurrent interactions between excitatory and inhibitory cells, and oscillatory dynamics are sensitive to E/I disruptions.

A number of computational studies have developed models of cortical microcircuits generating gamma oscillations to study how gamma-range synchronization is affected by synaptic perturbations implicated in disease states, with a specific focus on schizophrenia. These studies have characterized how the schizophrenia-associated disruptions of parvalbumin-expressing interneurons can impair the generation of gamma oscillations. Modeling studies have also characterized how gamma oscillations may be affected by dopaminergic modulation and antagonism of NMDA receptors on specific cell types. In each of these cases, the models provide testable hypotheses for the physiologic effects of synaptic manipulations.

These models have proven useful in understanding the circuit basis of aberrant cortical dynamics in disease states. However, the roles of aberrant oscillations in cognitive deficits and psychiatric symptoms remain unclear. There is still controversy regarding whether oscillatory synchronization is fundamentally necessary for the computations performed by neocortical circuits during cognition. Theoretical models have shown that oscillatory synchronization can subserve specific computational functions, such as efficient information transmission between neural populations. Further computational studies are needed that posit specific roles for oscillations in cognitive processes and make specific predictions for how behavior could be impaired in disease states.

Large-Scale Brain Dynamics

An emerging topic in computational neuroscience is the dynamics of large-scale brain networks. Progress in this area has the potential to provide insight into the growing body of experimental literature characterizing psychiatric disorders with whole-brain functional neuroimaging via fMRI. These data may provide important biomarkers of disease states. One common approach in fMRI is to analyze low-frequency fluctuations present in the blood-oxygenation-level-dependent (BOLD) signal during rest. In particular, resting-state fMRI can provide a pattern of functional connectivity, that is, the variability of the BOLD signal across brain parcels. These patterns of functional connectivity can be related to the underlying structural connectivity between brain areas, which can be measured noninvasively in humans through diffusion tractography,

Recent computational models have been developed to capture how the global pattern of resting-state functional connectivity emerges through long-range interactions among cortical areas. These large-scale network models simulate the dynamics of multiple nodes, each representing a cortical parcel. The nodes interact through an anatomical coupling matrix reflecting long-range connections, typically derived from tractography studies. Ongoing temporal dynamics in the nodes produce patterns of correlated fluctuations in activity across the network. The functional connectivity of the model can then be calculated and compared to the functional connectivity observed in experiments.

In one class of these resting-state models, each node is simulated with reciprocally connected pools of excitatory pyramidal cells and inhibitory interneurons, with diffusion tractography in humans used to define the structural connectivity pattern of excitatory long-range projections between nodes. Each node receives fluctuating background input to drive dynamics. As in the microcircuit models, the biophysical basis of these large-scale models makes them directly applicable to address the dynamical consequences of anatomical and physiologic changes induced by disease processes or pharmacologic manipulation. Changes in the strengths of local and long-range connections may induce differential effects in the patterns of functional connectivity. The models reveal that resting-state functional connectivity is sensitive to both the structure of underlying anatomical connectivity and the physiologic parameters that scale local and long-range connections. This can be used to test synapse-level hypotheses for a disease state by their predictions for resting-state alterations observed via fMRI. For instance, the hypotheses of elevated versus reduced E/I balance in local cortical circuits make dissociable predictions for the patterns of variability and covariability in the cortical BOLD signals, which can be tested in patients with schizophrenia.

FUTURE CHALLENGES

Integration Across Levels of Analysis

This chapter describes how computational psychiatry has offered a method to bridge levels of analysis, from local neuronal circuits to large-scale neural systems and, ultimately, psychiatric symptoms. The challenge that still faces this rapidly growing field is how to provide a formal bridge between these levels of inquiry for most psychiatric symptoms. We highlighted example cases where such progress has occurred—namely, in the context of working memory and fear conditioning. In such cases, computational psychiatry has begun to establish a formalism for how to link local circuits to large-scale neuronal dynamics and expression of behavior. However, the current models in this area of computational psychiatry are still simplistic in their repertoire of behavior and are designed to capture a very limited range of experimental contexts.

To accurately capture more complex cognitive and affective behaviors, models must incorporate multiple interacting brain areas possessing various distributed and modular computations. In this area, biophysically based modeling will be informed by connectionist-style models that are more abstract and removed from biophysical detail but that can be readily applied to modeling psychological processes and interactions among distributed neural systems. Similarly, a highly related challenge facing computational psychiatry is to provide a formal extension of biophysically based modeling to large-scale distributed brain networks beyond the microcircuit level, as this level of analysis offers the potential for the development of disease-related biomarkers. For instance, a key extension of such models would be to capture, in a biophysically realistic way, the large-scale thalamocortical functional loops that may play a critical role across a range of psychiatric disorders. The next generation of computational psychiatry studies will be vital to formalize these bridges across levels of analysis.

Computational Phenotyping

Psychiatric diagnoses are made using inclusionary and exclusionary criteria based entirely on co-occurring sets of symptoms. The Diagnostic and Statistical Manual of Mental Disorders (DSM) has been iteratively revised to improve diagnostic reliability and provide clear guidance for clinical decision making. However, consequently, a key objective of psychiatric research has been to identify biomarkers of disease under DSM categories, often via case-controlled studies that search for mean differences between groups of asymptomatic controls versus patients meeting diagnostic criteria for a single disorder. The challenge with using such a system is that patients with the same diagnosis may exhibit highly distinct behavioral and neural phenotypes, suggesting that considerable heterogeneity exists within diagnostic categories. It is appreciated by both researchers and clinicians that individuals’ psychiatric symptoms present with substantial heterogeneity, which cuts both within and across diagnostic categories. This conceptual problem is captured in the case of existing tension between category-based DSM-style diagnostic rubrics and the dimensional Research Domain Criteria (RDoC) framework. Instead of making an a priori decision regarding which framework may be better for a given psychiatric symptom or syndrome, we posit that perhaps there should be a parallel movement toward “precision psychiatry.” Simply put, it may be possible to harness advances in computational psychiatry to generate individual-level inferences regarding the type of underlying computational mechanisms that govern the observed patterns of neuroimaging effects or behavioral symptoms.

The current state of the art in this area of research focuses on group effects. That is, the majority of work that has emerged in computational psychiatry has focused on explaining group-level effects across a given diagnosis. We argue that one highly promising future application of computational psychiatry may be to provide individual-level computational phenotyping based on mechanistic models. Such a level of precision may possibly provide a method to manage the massive clinical heterogeneity that exists both within and across our current clinical rubrics in psychiatry. Ultimately, mapping such clinical heterogeneity onto underlying computational mechanisms will be vital to develop better mechanistically informed treatments.

A necessary and important step toward developing more effective and targeted treatments for psychiatric symptoms is a reliable and specific mapping onto underlying neural features. The limitations of categorical frameworks have been reflected in initiatives such as the RDoC, which advocated for studying a dimensional framework for mapping functional domains onto neural circuits. Such efforts have motivated large-scale, multisite, transdiagnostic studies for mapping symptom and neural variation across psychiatric disorders. These larger datasets have made it viable to explore multivariate brain–behavioral relationships across psychiatric spectra, paving the way toward quantitatively grounded individualized treatment and moving beyond traditional clinical scales that may not optimally parse neurobehavioral heterogeneity. For instance, despite many data-driven factor analytic studies of the psychiatric symptom space in the literature, a common approach in psychiatric neuroimaging studies remains "to relate" composite scores (such as total "positive" or "negative" psychosis symptoms) to neural measures. Critically, these analyses do not test the possibility that neural alterations in psychiatric disorders may be better captured by a weighted combination of symptoms, which would not be reflected in summed composite scores.

While such weighted combinations can be tested using multivariate approaches such as canonical correlation analysis (CCA), these studies themselves are prone to overfitting, leading to a failure to replicate results. This is often because of the intrinsically high dimensionality of the neural and behavioral data evaluated against a relatively low number of subjects. Indeed, recent work has demonstrated that in most cases for CCA, the sample size must be at least an order of magnitude greater than the number of neural/behavioral features used to detect a true effect (e.g., assuming a canonical correlation of $r = 0.3$). Even with prior dimensionality reduction of the neurobehavioral feature space (a common practice that typically reduces the number of features from hundreds of thousands down to hundreds), this necessitates a sample size of thousands of subjects. Of note, most current state-of-the-art clinical neuroimaging studies have target enrollment totals of 200 to 600 subjects, although several population-based and prodromal studies have achieved enrollment numbers in the thousands. Thus, while multivariate methods remain a promising and powerful approach to simultaneously computing maximal covariation between neural and behavioral features, the reproducibility and reliability of the results must be assessed.

In computational phenotyping, the parameters of a model are fit to best capture the empirical data for a single subject. Computational phenotyping could take advantage of a range of empirical data and modeling approaches. For instance, an RL model could be fit to a subject's decision-making behavioral data. Similarly, a biophysically based network model could be fit for a subject's resting-state neuroimaging data. The promise of this approach is that these model parameters may have diagnostic value, potentially leading to more predictive biomarkers. These model parameters may allow effective classification according to diagnostic criteria and may reveal clinically relevant clusters or dimensions within a diagnostic category or be predictive of a patient's response to treatment. Longitudinal studies will be needed to validate the diagnostic potential of computational phenotyping.

Modeling Treatments

Finally, of utmost relevance to psychiatry, computational modeling has the potential to provide a method for simulating possible effects of treatments (or compensations) that can occur across the described levels of analysis. For instance, some modeling efforts have already begun to generate predictions for how compensations onto NMDAR or GABAR synapses in local circuits can impact cortical disinhibition that adversely affects working memory computations. However, this area of research is at present underdeveloped and has future potential for guiding intuition in cases where complex neural dynamics are involved.

In order for any computational features to be clinically actionable for guiding treatment design, it is necessary to relate the feature(s) to targetable neural mechanisms. BOLD neuroimaging measures do not inherently provide information about the underlying mechanisms nor how these mechanisms may be disrupted to produce psychiatric symptoms. It is, therefore, important to validate neuroimaging measures using mechanistically informed neurobiologic features (genetic, cellular, or circuit-level measures) that can act as "molecular benchmarks." One such example is pharmacologic neuroimaging maps, which contain information about characterized mechanisms of action and can be used as mechanistic references for symptom-neural maps. For example, by quantifying a patient's similarity to one molecular target map versus another (e.g., NMDA antagonism via ketamine versus serotonin receptor agonism via serotonin), it is possible to evaluate a single patient's differential similarity to one potential treatment mechanism versus another. Another approach is to use gene expression data for genes implicated in psychiatric pathophysiology, such as transcriptomic maps derived from the Allen Human Brain Atlas. Quantifying the relationships between the transcriptomic patterns of certain genes (e.g., genes encoding neurotransmitter receptor subunits or interneuron markers) and neuroimaging phenotype maps can inform how specific mechanisms and neural systems are implicated in observed brain–behavioral features. These transcriptomic maps may thus provide an independent benchmark for treatment targets that may be associated with a specific receptor.

Here, further development and refinement of biophysically based models have particular advantages. Given the relevant cellular detail, such models can actually generate both biomarker and behavioral predictions of disruptions on specific synaptic sites. Such a level of detail may be out of reach in cases that are not informed by animal models. Nevertheless, this is precisely the point of the cutting-edge interplay between future research in computational psychiatry and clinical neuroscience. We posit that future studies in this area hold the exciting promise of rationally guiding treatment development in psychiatry, grounded in basic neuroscience.

FURTHER READINGS

*Abbott LF, Dayan P. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press; 2001.

Adams, R. A., Brown, H. R., & Friston, K. J. (2014). Bayesian inference, predictive coding and delusions. *AVANT. J. Philos. Int. Vanguard*, 5, 51–88.

Anticevic A, Cole MW, Repovs G, et al. Connectivity, pharmacology, and computation: toward a mechanistic understanding of neural system dysfunction in schizophrenia. *Front Psychiatry*. 2013;4:169.

*Anticevic A, Murray JD, Barch DM. Bridging levels of understanding in schizophrenia through computational modeling. *Clin Psychol Sci*. 2015;3(3):433–459.

Armony JL, Servan-Schreiber D, Cohen JD, Ledoux JE. Computational modeling of emotion: explorations through the anatomy and physiology of fear conditioning. *Trends Cogn Sci*. 1997;1(1):28–34.

Cohen JD, Braver TS, O'Reilly RC. A computational approach to prefrontal cortex, cognitive control and schizophrenia: recent developments and current challenges. *Philos Trans R Soc Lond B Biol Sci*. 1996;351(1346):1515–1527.

Corlett PR, Taylor JR, Wang XJ, Fletcher PC, Krystal JH. Toward a neurobiology of delusions. *Prog Neurobiol*. 2010;92(3):345–369.

Deco G, Kringelbach Morten L. Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. *Neuron*. 2014;84(5):892–905.

Huys, Q. J., Guitart-Masip, M., Dolan, R. J., & Dayan, P. (2015). Decision-theoretic psychiatry. *Clinical Psychological Science*, 3(3), 400–421.

Insel TR. The NIMH research domain criteria (RDOC) project: precision medicine for psychiatry. *Am J Psychiatry*. 2014;171(4):395–397.

Kishida KT, King-Casas B, Montague PR. Neuroeconomic approaches to mental disorders. *Neuron*. 2010;67(4):543–554.

Lee D. Decision making: from neuroscience to psychiatry. *Neuron*. 2013;78(2):233–248.

Lisman J. Excitation, inhibition, local oscillations, or large-scale loops: what causes the symptoms of schizophrenia? *Curr Opin Neurobiol*. 2012;22(3):537–544.

*Maia TV, Frank MJ. From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci*. 2011;14(2):154–162.

*Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. *Trends Cogn Sci*. 2012;16(1):72–80.

Redish AD, Jensen S, Johnson A. A unified framework for addiction: vulnerabilities in the decision process. *Behav Brain Sci*. 2008;31(4):415–437; discussion 437.

Stein DJ, Ludik J. *Neural Networks and Psychopathology: Connectionist Models in Practice and Research*. Cambridge University Press; 2007.

Stephan KE, Mathys C. Computational approaches to psychiatry. *Curr Opin Neurobiol*. 2014;25:85–92.

Thomas MS, McClelland JL. Connectionist models of cognition. In: Sun R, ed. *The Cambridge Handbook of Computational Psychology*. Cambridge University Press; 2008.

Uhlhaas Peter J, Singer W. Neuronal dynamics and neuropsychiatric disorders: toward a translational paradigm for dysfunctional large-scale networks. *Neuron*. 2012;75(6):963–980.

Wang XJ. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol Rev*. 2010;90(3):1195–1268.

*Wang XJ, Krystal John H. Computational psychiatry. *Neuron*. 2014;84(3):638–654.

1.25 Nuclear Magnetic Resonance Imaging and Spectroscopy: Basic Principles and Recent Findings in Neuropsychiatric Disorders

GRAEME F. MASON, PH.D., JOHN H. KRYSTAL, M.D., AND GERARD SANACORA, M.D., PH.D.

Historically the role of MRI in psychiatry has been limited to the use of structural MRI in ruling out known organic causes of emotional, cognitive, and behavioral symptoms. At present, there is a growing interest in the use of MR technology in psychiatry, largely owing to its noninvasive nature and its increasing ability to provide information on the structure, function, and physiology of the brain. While the potential usefulness of MR in the diagnosis of psychiatric disorders is immense, the current clinical applications remain extremely limited. However, the last decade has seen an explosion of MR research applications related to the study of the pathophysiology of psychiatric disorders and the mechanisms of action related to various forms of treatment. Imaging measures are also now included explicitly in the National Institute of Health's Research Domain Criteria (RDoC). In the context of the RDoC, the imaging results are specifically for research purposes, attempting to characterize pathophysiologic processes across a spectrum of disorders with various forms of measurements rather than limiting imaging to its standard diagnostic purposes.

HISTORY

MRI and other MR techniques were originally called nuclear magnetic resonance (NMR). The "nuclear" part of the name derives from the basis of the technique in energy transfer to and from the nuclei of atoms. The "magnetic" in the name refers to the use of magnetic fields in the experimental environment. The “resonance” component of the name comes from the use of particular frequencies of energy used for the measurements. Since its introduction, NMR has been used for basic investigations of chemistry and physics, but with its introduction to medical applications, the word “nuclear” has generally been dropped because patients may associate it erroneously with ionizing radiation, nuclear power, and nuclear weapons. Technical advances continue, not only in data acquisition but also in analysis, including artificial intelligence. A history and survey of many MR techniques has been published recently (Kose, 2021).

In 1946, Edward Purcell (1912–1997) and Felix Bloch (1905–1983) independently demonstrated NMR at Harvard and Stanford Universities, and for this discovery, they shared the 1952 Nobel Prize in Physics. Early NMR measurements were performed with a *swept field,* in which a narrow-banded radio frequency application was performed as the sample was exposed to a time-varying magnetic field. To enhance the efficiency of the measurement, pulse-and-Fourier-transform NMR was developed and successfully demonstrated at Varian Instruments in Palo Alto, California, in 1964 by Richard Ernst (born 1933), who proceeded to extend the concept to MRI and received the Nobel Prize in Chemistry in 1991. Raymond Damadian, a New York physician, observed in 1969 that tumors had different MR-related characteristics that provide the potential to differentiate cancerous tissue from healthy tissue, thereby demonstrating a potential clinical application of NMR. Dr. Damadian applied for the first patent for an MRI machine in 1972, in which he described the purpose of the device. In 1973, Zenuemon Abe and colleagues published a method that produced images by a slow process of acquiring signals sequentially as the position of the object was shifted through a series of positions, and in 1976 Dr. Damadian published a similar method applied to a mouse. A dramatic breakthrough for MRI occurred when Paul Lauterbur (1929–2007) of SUNY Stony Brook and Sir Peter Mansfield (born 1933) of Nottingham, England, independently developed the acquisition of MRI by means of magnetic field gradients, which will be discussed in this chapter, and that achievement led them to share the 2003 Nobel Prize in Physiology or Medicine.

BASIC PHYSICS OF NMR, MRS, AND MRI

For human use, the techniques are typically called MRI and MRS. Although MR methods are based on a quantum mechanical property called *spin,* they make use of phenomena that, for most purposes, are most appropriately described using classical mechanics.

Spin, the Larmor Frequency, and Net Magnetization

The primary points of this section are (1) only certain species of nuclei can be detected, (2) the frequency of detection depends on the species of nucleus observed and the strength of the magnetic field of the scanner, and (3) only a very small fraction of the nuclei can actually be detected.

The fundamental basis of MRI and MRS lies in quantum mechanics because the atomic nuclei that produce the signals are quantum particles. The nuclei carry a positive electrical charge, and observable nuclei have a quantum mechanical characteristic called *spin.* A spin of ½ is one of the basic properties of protons and neutrons. That is, a proton or a neutron can have a spin of either +½ or −½, and when neutrons and protons combine to form a nucleus, their spin values combine to yield a nucleus with the net of their spins. The most abundant form of hydrogen has a nucleus that is simply a proton, so the spin of the nucleus of ^{1}H is ±½. The most common form of carbon has a nucleus that is composed of six protons and six neutrons, half with spin +½ and half with spin −½, such that in combination, they yield a nucleus for ^{12}C that has no spin; thus, ^{12}C is not observable with MR techniques. However, 1.1% of naturally occurring carbon is ^{13}C, and the additional neutron that resides within its nucleus confers upon ^{13}C a net spin of ½; therefore, ^{13}C can be observed with MR techniques. Table 1.25–1 lists several MR-observable nuclei that have been applied to study the human brain.

Nuclei that have nonzero spin can be detected with NMR by applying energy and observing their responses, and so investigators who develop MR technology often discuss the behavior of spins in a magnetic field and the manipulation of spins during a measurement. When a charged particle like the atomic nucleus has spin, it produces a *magnetic dipole,* which can be imagined as a small bar magnet with North and South ends.

Table 1.25–1.
MR-Detectible Nuclei and Some of Their Applications to Brain Research

Nucleus	γ (MHz/Tesla)	Sensitivity Per Atom (% of ^{1}H)	Application
^{1}H	42.58	100.0	Amino acid neurotransmitters and neuronal and glial compounds; imaging of water and lipids
^{2}H	6.54	1.89×10^{-8}	Metabolism of fats, glucose, acetate: Deuterium Metabolic Imaging (DMI)
^{31}P	17.25	.0666	High-energy phosphate levels, intracellular pH, phospholipid head groups
^{13}C	10.71	.0174	Metabolism of amino acid neurotransmitters, glucose, lipids, acetate
^{15}N	−4.32	-3.83×10^{-4}	Metabolism of ammonia, amino acids
^{17}O	−5.77	1.89×10^{-7}	Cerebral oxygen consumption, production of H_2O
^{19}F	40.08	83.5	Cerebral pharmacokinetics of fluorinated drugs like fluoxetine and fluvoxamine
^{23}Na	11.26	1.60×10^{-3}	Effects of hypoxia, pharmacologic challenges to sodium pumps
^{7}Li	16.55	4.62×10^{-3}	Cerebral pharmacokinetics of lithium

When the nuclei with spin are placed in a strong magnetic field of magnitude B_0, they *precess* around the axis of that field (Fig. 1.25–1) with a frequency called the *Larmor frequency* or the *resonance frequency*, with γ and B_0 defined in the following paragraph.

$$v = \gamma B_0$$

The Larmor frequency is the frequency that must be transmitted to affect the sample, and it is also the frequency that is received by the MR scanner. Equation 1 shows that the Larmor frequency depends on the strength of the static magnetic field, as well as on γ. The frequency of detection of a nucleus in a clinical magnet operating at a magnetic field strength B_0 of 3 Tesla (abbreviated 3 T) is twice the frequency of detection of the same nucleus at 1.5 T. The Larmor frequency of the most abundant isotope of hydrogen, ^{1}H, is 63 MHz at 1.5 T and 126 MHz at 3 T, the most common clinical field strengths. These frequencies lie in the FM band of the radio, and accordingly, MR scanners are usually in rooms that are electrically sealed to prevent ambient signals from contributing what looks like noise in the data. The other variable that affects the frequency is γ, the *gyromagnetic ratio,* which differs for each species of the nucleus. So, because the value of γ for the common hydrogen nucleus is approximately four times as large as that for ^{13}C, the Larmor frequency of hydrogen is about four times as large as that of ^{13}C.

The reason for the large static magnetic field is to create a condition called *net magnetization.* In the absence of an external magnetic field, the spins of the nuclei in the magnetic field are oriented randomly in space, and if you were to draw an arrow through the spin orientation of each nucleus, then collect all those arrows into a common starting point, you would have a ball in which every arrow was paired with an equal and opposite arrow on the other side of the ball (Fig. 1.25–2A.1) so that the directions of the arrows averaged to zero. That is, zero net magnetization in the absence of an external magnetic field (Fig. 1.25–2B.1). To create the potential for a signal, some nonrandomness must be introduced. In the presence of a strong magnetic field, the spins become slightly biased in one direction (Fig. 1.25–2A.2), such as one might imagine from a dandelion seed head under a slight breeze, and creates the net magnetization (Fig. 1.25–2B.2). It is this small bias in the alignment of the nuclei that MR scanners can detect, and the greater the net alignment, the larger will be the signal that can be detected from the sample. The net magnetization expressed as a fraction of the total spins is given by the *Boltzmann distribution*

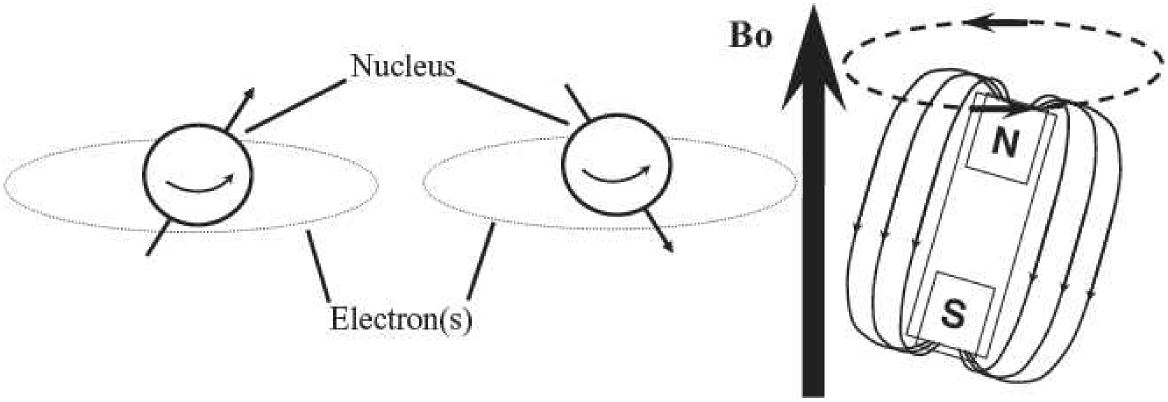

FIGURE 1.25–1. Individually, nuclei are particles that have the quantum mechanical property *spin*. By virtue of their charge and the spin, they behave as small magnets that can be aligned either with or against the large magnetic field of the MR scanner (B_0), rotating at the *Larmor frequency.*

$$N_{high}/N_{low} = 1 - (\mathrm{v}hB_0/kT)$$

where k is the Boltzmann constant, h is Planck's constant, and T is the absolute temperature. Most people have little occasion to use Equation 2, but it illustrates several important features of NMR measurements. First, a larger net magnetization can be obtained if a nucleus with a larger gyromagnetic ratio is used. Hydrogen has the largest ratio and has the sensitivity to be used for routine anatomic and other imaging. Sodium has a low ratio and is less sensitive and more difficult to use for imaging. Second, a larger B_0 yields more net magnetization, so a stronger magnet can yield more signal. This is why Earth's field MR has not been used for studies of the brain: a 3-T magnet is 67,000 times stronger than the Earth's average magnetic field, and so the signal available at Earth's field is thousands of times lower. In practical terms, measurements on 3-T machines have twice the signal available from a 1.5-T machine, while 7-T machines have over 4½ times the signal available. Third, lower temperature adds signal, but for measurements in humans, T is usually fixed at about 37 °C (310 K). Finally, if one plugs in typical temperatures and magnetic field strengths used for human clinical scans, the equation shows that for hydrogen, the net magnetization consists of only 0.0004% to 0.0008% of the nuclei in the sample.

Detection of the Net Magnetization

The techniques of MRS and MRI in vivo detect the quantities of net magnetization in various locations and elucidate the properties of the cellular environment. For MRS, there is the particular purpose of discovering in what chemicals the magnetization resides, and for MRI, the specific purpose is to measure the spatial distribution of the magnetization. Having the nuclei precess at the Larmor frequency and follow a net distribution oriented along with the scanner's magnetic field sets the stage for MR detection, but one more component is necessary to obtain a signal.

Within seconds of placing a subject in the scanner, the nuclei in the brain are in a fully *relaxed* state (Fig. 1.25–2A.2 and B.2). In the jargon of MR, a fully relaxed state is one in which the nuclei have a very small net orientation with the magnetic field, according to their Boltzmann distribution. Each spin precesses about the axis of the magnet, and as each one precesses, it causes a minute magnetic field to oscillate at the Larmor frequency. Those oscillating fields are how a signal can be detected, but, at any instant, for every spin oriented in one direction about the axis, there is another spin oriented in the opposite direction, and their oscillating magnetic fields cancel one another. In this condition, no signal can be detected in the sample. To solve this problem and detect the precessions, one must get the spins into a coordinated group precession. A magnetic field oscillating at the Larmor frequency (called radiofrequency, or *RF*) is applied, generally lasting a few hundred microseconds to a few milliseconds, and the nuclei absorb energy from the transmission, which causes the whole assembly of spins, including its net magnetization, to rotate out of alignment with the large, static magnetic field of the scanner (Fig. 1.25–2A.3 and B.3). As the energy rotates the net magnetization toward the perpendicular to the scanner's magnetic field, the spins acquire *phase coherence.* That is, the net magnetization is put in a position that has the net spins precessing together instead of having a random distribution about the *Z*-axis. In that position, their tiny oscillating magnetic fields occur coherently, which the detector can pick up as a signal, which also has the Larmor frequency. The energy is detected with a *probe,* sometimes also called a *coil,* which is placed around or against the part of the body that is being studied. In many cases, the same coil is used for transmission and detection. After the energy is applied, the signal gradually dies away as the spins that comprise the net magnetization disperse their orientation and return to alignment with Z.

Regarding terminology, as the net magnetization is rotated away from the magnetic field orientation, any magnetization that is perpendicular to the scanner's magnetic field is called *transverse magnetization* (Fig. 1.25–2A.3). In the terminology of classical mechanics, MR can only detect the component of the magnetization vector that lies in the *transverse* plane (Fig. 1.25–2B.3). If less energy is used, then a lower flip angle is achieved, and less signal can be detected. It is possible to rotate the magnetization through any angle, including 180 degrees, which is a special rotation called an *inversion,* and it is used frequently for MRI and MRS.

Quantum and Classical Mechanics

Many explanations of MRI and MRS are based principally on quantum mechanics. The explanations are often daunting and are generally unnecessary. A phenomenon is called "quantum mechanical" if classical mechanics are unable to explain the observed behavior. For example, a rotating bicycle wheel can be described with great effort using quantum mechanical calculations, but the wheel's behavior is explained with great precision and considerably greater ease with classical mechanics. Therefore, a rotating bicycle wheel is considered a classical phenomenon, not a quantum mechanical phenomenon. In the case of MR, observations of a single spin can be explained only with quantum mechanics, but when many spins are observed, as is the case for an MRI study of the brain, the nuclei behave as a set of spins that have a continuum of orientations, though slightly biased along the magnetic field. The behavior is explained completely in the realm of classical mechanics. This point is described in detail by Hanson (2008).

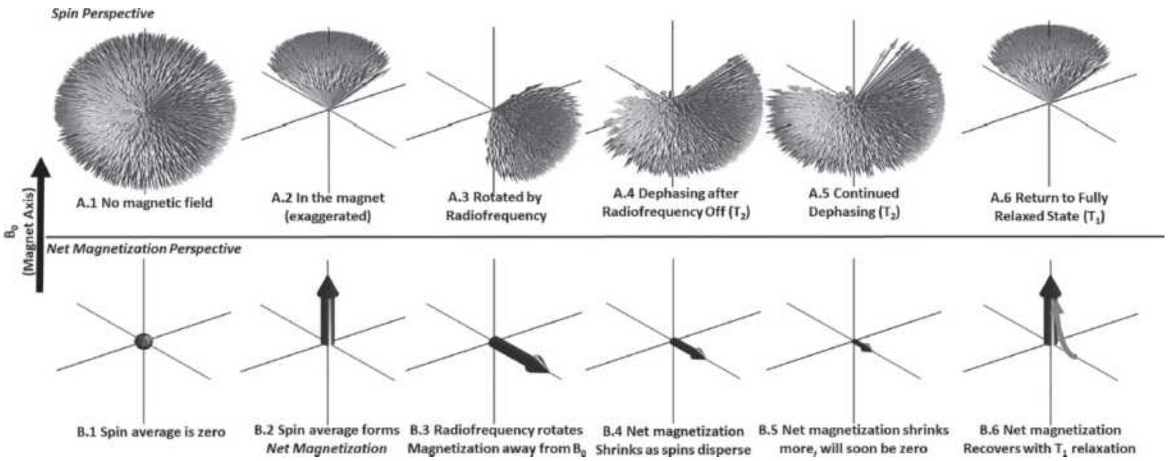
Spin Perspective
A.1 No magnetic field
A.2 In the magnet (exaggerated)
A.3 Rotated by Radiofrequency
A.4 Dephasing after Radiofrequency Off (T_2)
A.5 Continued Dephasing (T_2)
A.6 Return to Fully Relaxed State (T_1)
B_0 (Magnet Axis)
Net Magnetization Perspective
B.1 Spin average is zero
B.2 Spin average forms Net Magnetization
B.3 Radiofrequency rotates Magnetization away from B_0
B.4 Net magnetization Shrinks as spins disperse
B.5 Net magnetization shrinks more, will soon be zero
B.6 Net magnetization Recovers with T_1 relaxation

FIGURE 1.25–2. Magnetization in an MR experiment shown with perspectives of quantum and classical mechanics. **A:** Quantum mechanical: **A.1** and **A.2:** In the presence of an external magnetic field, the individual spins are aligned with or against the field, and a slight majority, on the order of just 0.0004 percent, is aligned with the field. The very small size of this ratio is the reason that MR is considered to be insensitive compared to other forms of imaging. **A.3:** When a radiofrequency burst is applied at the Larmor frequency of the spins to apply what is called a 90-degree pulse, the net distribution is randomized so that the spins are evenly distributed with and against the field. At the same time that the spins are randomized, they acquire *phase coherence,* coming together in one orientation with respect to the *XY* plane. **A.4** and **A.5:** As time passes, interactions of the individual spins with the environment changes their orientation in the *XY* plane according to T_2 relaxation, so that they gradually disperse and lose that phase coherence. **A.6:** Eventually the spins not only have lost any net *XY* orientation, and they also have retuned to their original state along Z, so the small portion of net magnetization now looks like **A.2**. **B:** Classic mechanical: **B.1:** In the absence of a magnetic field, the spins of **A.1** average to zero, just a dot, no net magnetization, and hence no available signal. **B.2:** In the presence of the external magnetic field, there exists a magnetization vector whose size is proportional to the sum of the small majority of spins that have net alignment with the field. **B.3:** When a 90-degree pulse is applied, the magnetization vector rotates to the *XY* plane. This is the average orientation of the vectors shown in **A.3. B.4** and **B.5:** After the excitation, the magnetization vector shrinks over time as the individual spins in **A.4** and **A.5** gradually point in various directions in the *XY* plane. **B.6:** The net magnetization eventually loses all its *XY* orientation and returns to the *Z* axis according to T_1 relaxation.

T_1, T_2, and T_2* Relaxation

Once the magnetization vector has been rotated to the transverse plane, phenomena called *relaxation* to eliminate the signal and return the vector to the *Z* axis through exponential processes called T_1 and T_2 relaxation, or *longitudinal* and *transverse* relaxation, respectively (Fig. 1.25–2A.5,6, and B.5,6). Longitudinal relaxation returns the magnetization to alignment with the scanner magnetic field at a rate governed by the exponential time constant T_1. As will be explained later, this relationship can be used to create contrast among materials with different values of T_1, such as gray matter and white matter.

Repeating scans too rapidly with too large an excitation angle leads to a reduction in signal amplitude through a process called *saturation.* Transverse magnetization can only be generated from *Z* magnetization, which is created during the T_1 relaxation process. Therefore, if too little time is allowed for recovery of *Z* magnetization, then the signal will be reduced on repeated scans. For full relaxation after a 90-degree pulse, typically, a delay of several seconds is required. Less saturation occurs when a lower angle of excitation is used, but that also results in the generation of less transverse magnetization. In practice, the most efficient acquisitions incorporate a compromise, in which an excitation angle of something less than 90 degrees is used, and an interscan delay of less than three T_1s is allowed between scans.

In vivo, T_1 relaxation depends on the static magnetic field strength, the rate of tumbling of the molecules in which the spins reside, and the distance between the relaxing nucleus and other charged spins in the sample. As a consequence, as magnetic field strength rises, the value of T_1 increases. As will be shown later, the dependence of T_1 on field strength can be exploited to make more effective experiments but can also place some limitations on measurements at different magnetic field strengths.

For measurements in vivo, long before the magnetization vector has returned to the *Z* axis, the signal decays to nothing because of T_2 relaxation. T_2 relaxation destroys the phase coherence of the component of the magnetization in the *XY* plane, such that the MR signal decays in an exponential process (Fig. 1.25–2A.4,5 and B.4,5). The signal gradually dies away, yielding a *free induction decay,* or FID. T_2 relaxation is primarily due to interactions of the nucleus with other spins in the sample, and in the brain, in vivo, its value is largely independent of magnetic field strength. As phase coherence is lost, the signal from each spin in the sample becomes increasingly canceled by another spin in the sample so that gradually the average signal coming from the sample approaches zero. The value of T_2 depends largely on the mobility of the molecule in which the nucleus resides: The greater the restrictions on motion, the shorter the T2 will be. For example, the white matter has an abundant, rigid structure that prevents water from diffusing rapidly and rotating freely, and the white matter has a shorter T_1 than cerebral spinal fluid (CSF), in which the water is free to rotate. The differences between white matter and gray matter provide an effective basis for high contrast for diseases in which fluid must be emphasized in MRI. The T_2 of bone is very short and is therefore not typically detected in MRI or MRS.

A parameter related to T_2 is T_2*, which results from the inhomogeneity of the magnetic field. Such inhomogeneity can derive from macroscopic causes, such as an interface between tissue and bone or air. The sinuses, ear canals, and temporal bones cause large inhomogeneities that are often troublesome for MRI and MRS studies. Inhomogeneities also result from microscopic influences, such as the presence of deoxyhemoglobin. T_2* is the time constant for an exponential loss of phase coherence, but the process also occurs with the dephasing of magnetization in different parts of the sample due to tissue inhomogeneities or shape. The sample volume could be an image pixel or region selected for an MRS measurement. If one corner of an image pixel experiences a slightly higher magnetic field strength, then the Larmor frequency in that corner will be a little higher. Perhaps the center of the image pixel has a lower frequency, so the Larmor frequency in the middle of the pixel will be lower. As time progresses, the magnetization in different parts of the sample will eventually randomize their phase, such that any signal coming from one part of the sample will be canceled by signals with the opposite phase from another part of the sample. In certain kinds of MRI and MRS studies, such effects can dramatically reduce a signal, causing regional darkening of images in MRI and blurring and loss of signal in MRS.

Influences on Frequency and Identification of Molecules in MRS Measurements

The main points of this section are that the major factors that influence a resonance frequency are (1) magnetic field strength and (2) gyromagnetic ratio, which affect primarily what transmitter frequency will be used. Smaller influences are the electron cloud conformation and the interactions with other nuclei in the molecule, which together imbue MRS with its ability to distinguish neurochemicals from one another.

The primary actor in the differentiation of chemicals is *shielding,* a phenomenon in which the electron cloud in a molecule partially blocks the static magnetic field that penetrates the nucleus. Consider the cases of glutamate and glutamine (Fig. 1.25–3). The fourth carbon in the chain (C4) of glutamate resides next to a carboxyl group whose oxygen atoms are highly electrophilic. The nearby oxygen atoms of the carboxyl group draw the electrons at that end of the molecule to themselves so that the C4 carbon nucleus of glutamate is more exposed to the large magnetic field and therefore has a larger Larmor frequency. Stated differently, the carboxyl group reduces the shielding of the carbon nucleus, slightly modifying the value of B_0 and thereby causing the carbon to resonate at a slightly higher frequency. Now consider glutamine, which has only one oxygen near the C4 carbon position, in addition to the nonelectrophilic amide group, and so the electrons more fully shield the glutamine C4 carbon nucleus. The C4 of glutamine resonates at a lower frequency than the C4 carbon of glutamate.

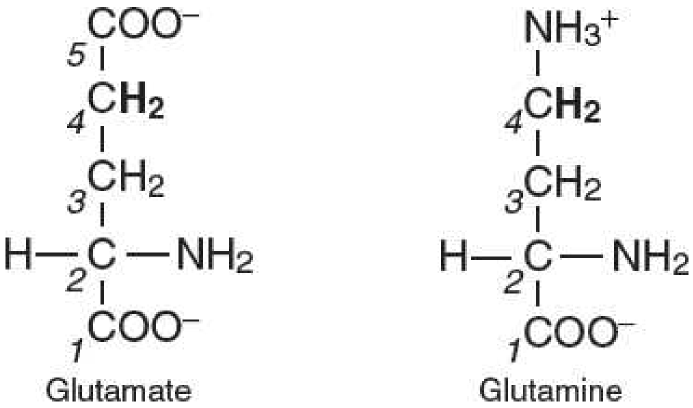

FIGURE 1.25–3. Neurochemicals differ slightly in the Larmor frequencies of their atoms according to chemical identity and location within the molecule. The frequencies differ according to their magnetic environment, and an electron cloud provides shielding for the nuclei. In the case of the $–CH_2–$ of glutamate, whose carbon is numbered 4, adjacent to the carboxyl group, the electrophilic oxygen atoms on the carboxyl group draw the electron cloud toward themselves, leaving the hydrogen nuclei exposed to the large static magnetic field. The nuclei of the corresponding hydrogen atoms in glutamine do not have the adjacent electrophilic oxygen atoms to unshield them and therefore experience a slightly lower magnetic field strength. Therefore, the resonance frequency of glutamate at 1.5 T is about 7 Hz higher than that of glutamine.

Glutamate and glutamine are also separated in the ^{1}H spectrum. At a magnetic field strength of 1.5 T, the resonances of the two hydrogen nuclei attached to the C4 of glutamate and glutamine are centered 7.7 Hz apart for glutamate relative to glutamine. At 3 T, the same hydrogen nuclei resonate 15.4 Hz apart. At 7 T, they resonate 36 Hz apart, which separates the signals more effectively at the higher magnetic field strength. Because MRS studies are performed across a wide range of magnetic field strengths, some common set of units are used to ease the identification of chemicals based on their resonance frequency. The frequency of resonance for nuclei at a particular position in a given molecule is called the *chemical shift,* and the units are parts per million or *ppm.* The notation of ppm is obtained by dividing the frequency difference in hertz by the Larmor frequency in megahertz (i.e., 10^6 Hz). The result has units of hertz per million hertz, or ppm. In these units, those protons of glutamate and glutamine are said to be centered .11 ppm apart, no matter what the static magnetic field strength.

Chemical shifts must be seen as relative differences in frequency location. By convention, for ^{1}H and ^{13}C MRS, the field has agreed that the signal from trimethylsilylpropionate (TSP) be designated as the reference compound with a shift of 0 ppm. All other compounds, by convention, are described by their frequencies relative to that of TSP. For example, lactate has protons that resonate 1.33 and 4.15 ppm away from TSP, respectively (Fig. 1.25–4). For reasons based on the original techniques used to obtain MRS measurements, called *swept field NMR,* lower chemical shifts are termed *upfield,* and the higher chemical shifts are called *downfield.*

The final influence on the frequency of a nucleus is *J-coupling,* which is the influence of neighboring nuclei with spin upon one another. The state of the spin of one nucleus exerts a very slight effect on the *transition energy* of the neighboring spin. To describe J-coupling, a quantum mechanical explanation is necessary. The transition energy is the amount of energy that a spin can absorb and change its energy state, and it is a defined, fixed amount and must be delivered at precisely the Larmor frequency. If one spin is coupled to another, usually through a covalent chemical bond, one spin will change the other's transition energy to a slightly higher or a slightly lower value. Because the transition energy is ultimately what defines the resonant frequency of a nucleus, half the spins have a slightly higher frequency and half a lower frequency. A nucleus that is affected by one nearby spin, such as the $–CH_3$ group at 1.33 ppm in lactic acid, has a *doublet* or two resonances instead of one, each representing half of the total resonance for –CH_3, as is the case for the methyl group of lactate, in which three identical hydrogen nuclei are affected identically by a single adjacent hydrogen (Fig. 1.25–4A). If nuclei are adjacent to two spins, then their resonance is divided into a *triplet,* as demonstrated by the spectrum of ethanol (Fig. 1.25–4B). Nuclei adjacent to three spins, such as seen in the hydrogen adjacent to the methyl group of lactic acid, yield a quartet that arises from the effect of the three methyl protons of lactic acid on the single ^1H at 4.11 ppm (Fig. 1.25–4A) or the –CH_2 protons of ethanol (Fig. 1.25–4B). The spectral distance between each resonance in the multipeak structure is J Hz and is typically reported in hertz, not ppm because the splitting due to J-coupling is independent of magnetic field strength. For the ^1H–^1H splitting of Figure 1.25–4, the value of J usually lies between 3 and 15 Hz.

SYSTEM HARDWARE

Before moving from the basic physics of MR to methods of MRI and MRS, it is helpful to examine the MR machine itself. MR machines are composed of three or four main components: A magnet, a *transmit path,* and a *receive path,* and almost all MR scanners also have *gradient coils.* The magnet is the source of B_0, which is what creates the net magnetization. The transmit path generates the transverse magnetization. The receive path detects the signal from the sample and carries it back to the computer for storage.

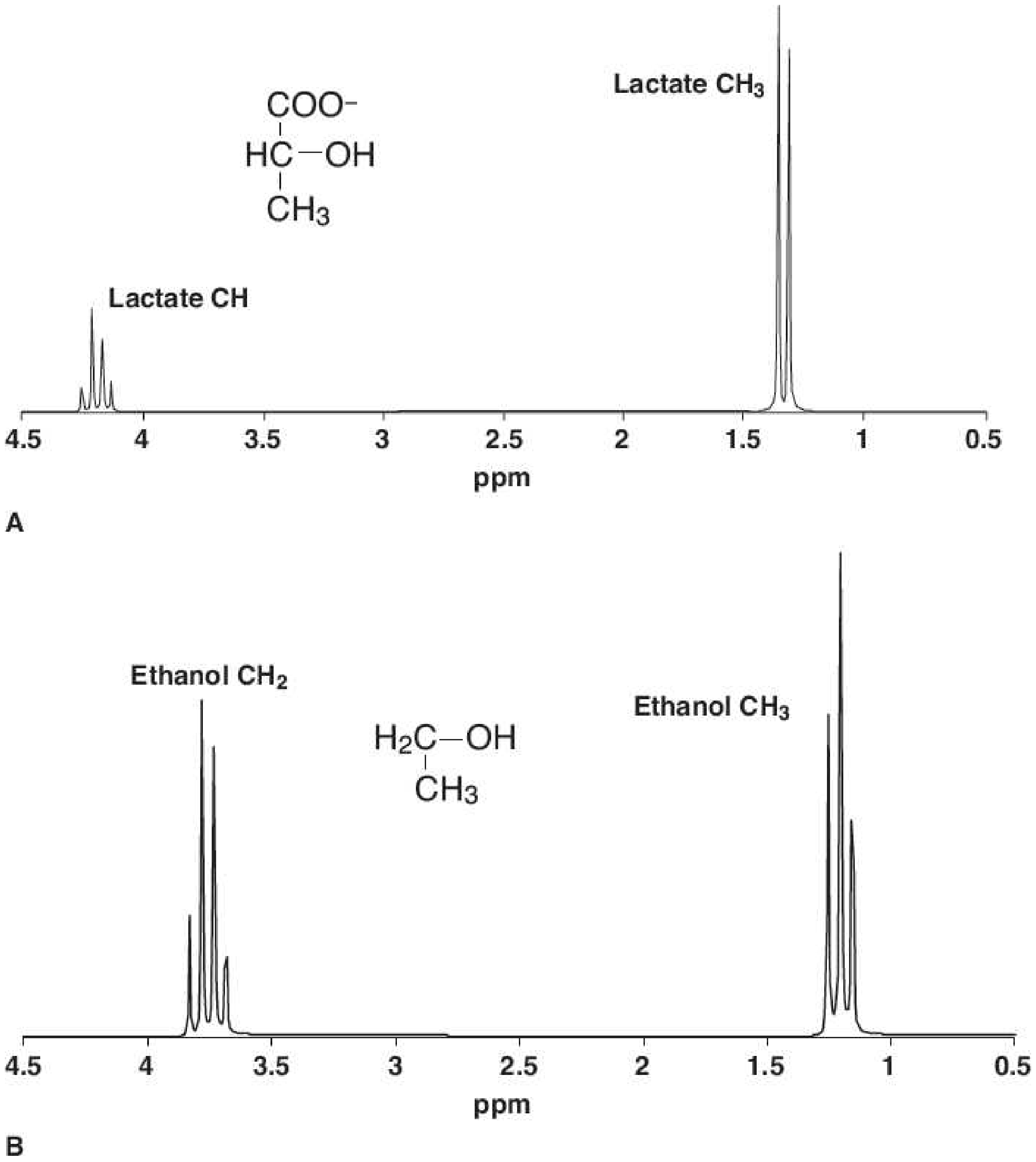

FIGURE 1.25–4. Effects of chemical shift and *J*-coupling. Chemicals in MRS studies are identified primarily on the basis of their chemical shifts, but significant information also arises from the *J*-coupling interaction of spins. **A:** Lactate has resonances at 4.15 and 1.33 ppm from TSP. The methyl protons are *equivalent* and are all centered at 1.33 ppm. However, they interact with the adjacent single proton at 4.15 ppm and split their resonance into a doublet centered at 1.33 ppm. The single proton at 4.15 ppm interacts with all three methyl protons and is split into a quartet. **B:** Ethanol has a methyl group whose resonance is centered at 1.2 ppm and interacts with an ethylene group whose resonance is centered at 3.7 ppm. The two protons split the methyl

resonance into a triplet, while the three protons of the methyl group divide the ethylene resonance into a quartet.

The magnet (Fig. 1.25–5) consists of hundreds of miles of superconducting wire wound in precise formations. When a magnet is installed, a strong vacuum is applied to remove all ambient air and water from around the wire. Then the evacuated container is filled with liquid helium to maintain its superconductivity at a temperature of 4 K, or −453 °F. The helium bath is insulated by a vacuum chamber to reduce boil-off. Liquid helium is expensive, and a recent worldwide shortage is expected to continue for several years, so in many magnets, the helium structure lies within a bath of liquid nitrogen or a refrigeration unit. The helium in the magnet must be maintained at high purity because humidity or air that enters the helium chamber (called a *dewar*) will solidify and may create a blockage that can result in an explosion. Another sort of failure is a *quench,* which is a sudden loss of charge in the electrical windings of the magnet, simultaneously with the almost instantaneous expulsion of helium from the magnet. Such events can be caused by a loss of vacuum around the helium container, a failure to maintain the helium at an adequate level, or the introduction of a ferromagnetic object in the vicinity of the magnet.

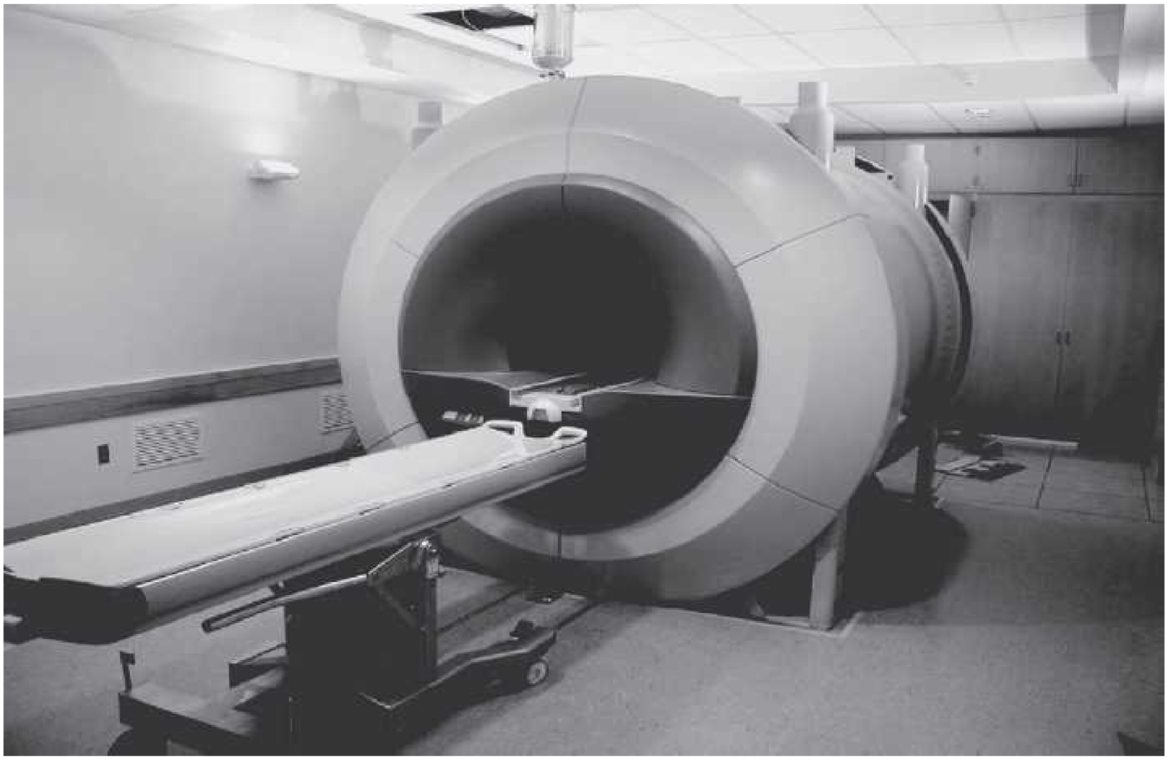

FIGURE 1.25–5. A whole-body MR scanner. Magnets for MRI are usually cylindrical, with a hole through the middle. The hole accommodates a set of cylindrical gradient coils that are used for shimming and experiments that require spatial information, such as imaging. The gradient coil cylinder has a bore large enough to accommodate a human torso or, in some new models, just a person's head. A patient bed slides in and out of the scanner. On top of the magnet are vents to allow for the boil-off of helium and, in many cases, nitrogen gas. There is also an entry point to allow charging of the superconducting windings when the magnet is installed and to allow controlled shut down if the magnet needs to be de-energized after some years of operation.

The magnet is constructed with additional circuits called *shims,* which are used to fine-tune the homogeneity of the magnetic field. The quality of MRI and MRS improves dramatically with the homogeneity of B_0. The magnets are designed and built to achieve a homogeneous magnetic field within a sphere in the center of the magnet, but when a magnet is placed at the site where it will operate, the homogeneity is disrupted. Initial adjustments to B_0 homogeneity are made using small pieces of metal, called *shims,* placed strategically at different places in the bore of the magnet in a painstaking process, and then engineers adjust the magnetic field homogeneity further with the shim circuits. Likewise, when a person is placed in the scanner, the homogeneity is altered, and it is necessary to adjust the shim circuits to compensate. For many cases, it is only necessary to apply three linear adjustments to the magnetic field, but for some applications, including MRS, second- or third-order changes must often be applied.

The transmit path begins with a frequency source whose output is modified to obtain a desired amplitude, phase, and shape. The modulated signal is sent to a radio frequency amplifier to boost the transmission, often into the range of kilovolts. The transmission is directed to the coil that is around the head, where it generates an oscillating magnetic field, which by convention is called B_1.

Part of the transmit path is the radio frequency coil. A typical coil for brain imaging and spectroscopy is shown in Figure 1.25–6. The coil surrounds the brain so that relatively uniform coverage is achieved. Another type of detector, a *surface coil,* is used for cases in which localized coverage with high sensitivity is required. All coils must be tuned to the Larmor frequency, and to do that, adjustable capacitors form part of the circuit. The coil not only transmits to the brain but also receives the signal from the brain and therefore forms the beginning of the receive path. The signal in the receive coil has an amplitude of microvolts at the Larmor frequency. If it traveled unaltered through many feet of wires to the computers for processing and storage, then the noise that accumulates along its path would overwhelm the signal. Therefore, as close as possible to the coil, magnet-compatible *preamplifiers* are used to boost the received signal to a level of millivolts. The signal is modified to shift the frequency into the auditory range, covering kilohertz, and stored on a computer. Typical causes of coil failure include a preamplifier damaged by leakage of radio frequency transmission into the receive path or damage to capacitors from power-induced cracking or sparking. Such failures become more likely in a head coil by transmitting high power levels into coils, especially when a sample whose size and conductivity differ significantly from a human head.

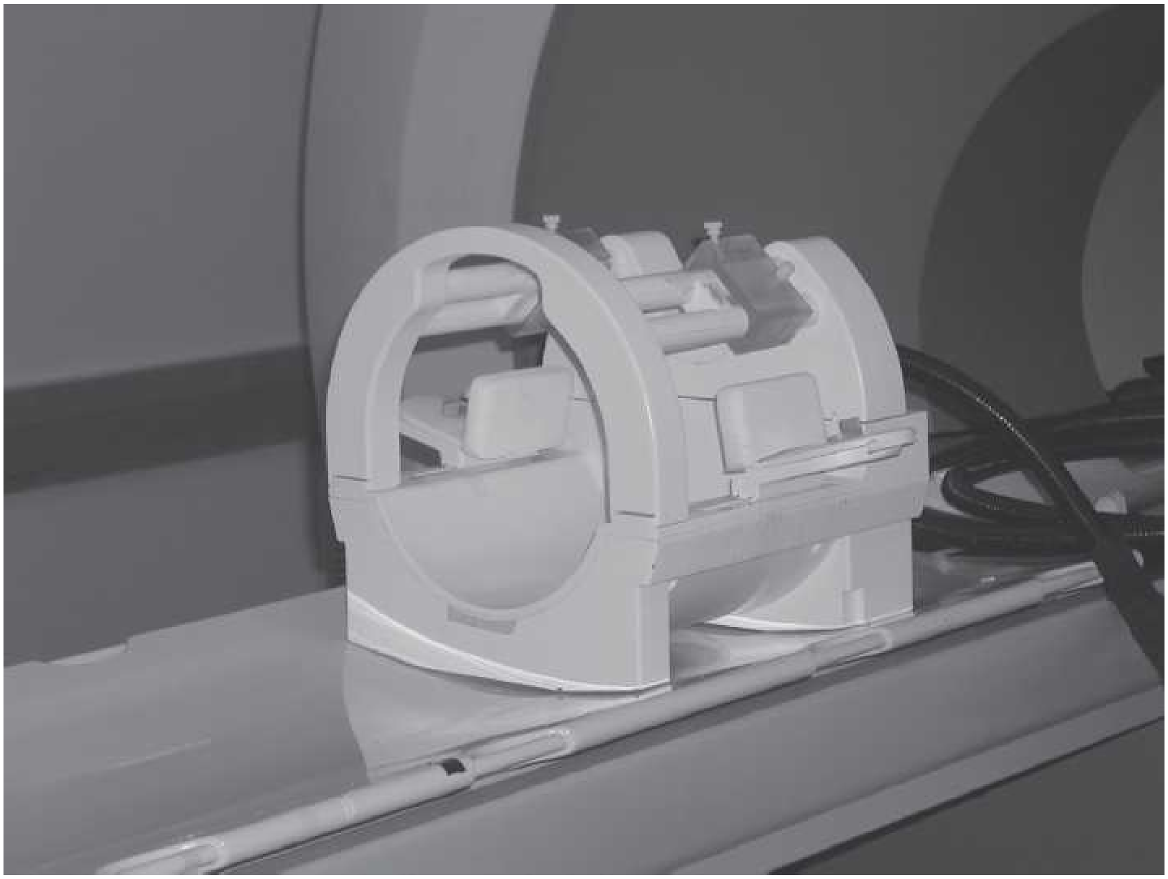

FIGURE 1.25–6. A typical coil used for MRI and MRS of the brain. Built for whole-brain coverage, this coil surrounds the head. The ideal would be a long, empty conducting tube, but to accommodate a head, it must be short, and to allow for patient comfort and construction constraints, it must have discrete conducting elements. Four of the elements are housed within the bars that can be seen along the upper part of the coil. The other elements are housed in the bottom section of the coil, beneath where the head rests during a scan.

Gradient coils form an almost universal fourth component of hardware in an MR scanner. In the absence of gradient coils, MR signals are obtained with little or no information about what part of the sample they represent. For example, if a standard brain imaging protocol is run without gradient coils, then the result would be an MR spectrum of a large water signal and a small fat signal. The *X, Y,* and *Z* gradient coils provide the positional information. When current is directed through the *X* gradient coil, a spatially linear ramp is added to B_0, such that $B_{0\text{effective}} = \mathrm{B}_0 + G_x x$, where G_x is the gradient strength in units of millitesla/centimeter, and *x* is the distance from the magnet's center. Similar relationships exist for the *Y* and *Z* gradients and the *y* and *z* directions in the magnet. When power is applied to the gradients, the Larmor frequency depends linearly on the position within the magnet, and by this encoding, images can be obtained. Patients commonly ask the source of the loud noises of an MR scanner, and the answer is the gradient coils. When the current is switched on and off rapidly, it generates a magnetic field that opposes the primary field of the magnet and generates large forces that manifest as loud noises. Patients may also sense vibrations inside the scanner from the same cause.

FOURIER TRANSFORM

Having examined the basic physics and the MR machine, it is helpful to turn to one of the most basic analytical tools that are used for MR studies. The purpose of this section is to explain and illustrate the relationships between the observed signals and the processed images and spectra of MRI and MRS. The signals are observed over periods of milliseconds, and those time-dependent signals must be transformed into spatial pictures in the case of MRI and spectra in the case of MRS.

MRS and MRI both make use of a mathematical operation called a *Fourier transform,* which is used in many fields to switch between time and temporal frequency (e.g., seconds and hertz) or between space and *k-space* (centimeters and reciprocal centimeters). Every time-dependent signal can be represented as a set of frequencies, measured in hertz, and every object can be expressed as a set of signals in *k*-space. The idea behind Fourier transforms is that any shape can be created by adding the right frequencies and amplitudes of sinusoidal signals together. Figure 1.25–7 shows the decomposition and recreation of a profile of the Taj Mahal in India. The Fourier transformation at the upper right is plotted as a histogram of weightings associated with the range of frequencies that make up the profile, with low frequencies in the center of the graph and higher frequencies on either side. The panels at the center-left show that the contributions of the lowest frequencies to the profile have to do with large features. From the low frequencies, one can see that the profile has a large object in the middle and something large on either side, but there is little detail. The next panels to the right show the effects of adding progressively higher frequencies, which gradually produce finer features. That is because the higher-frequency components of the Fourier transform contain the details of a signal. When all of the frequencies are included, the original image is obtained. The Taj Mahal is built from sine waves.

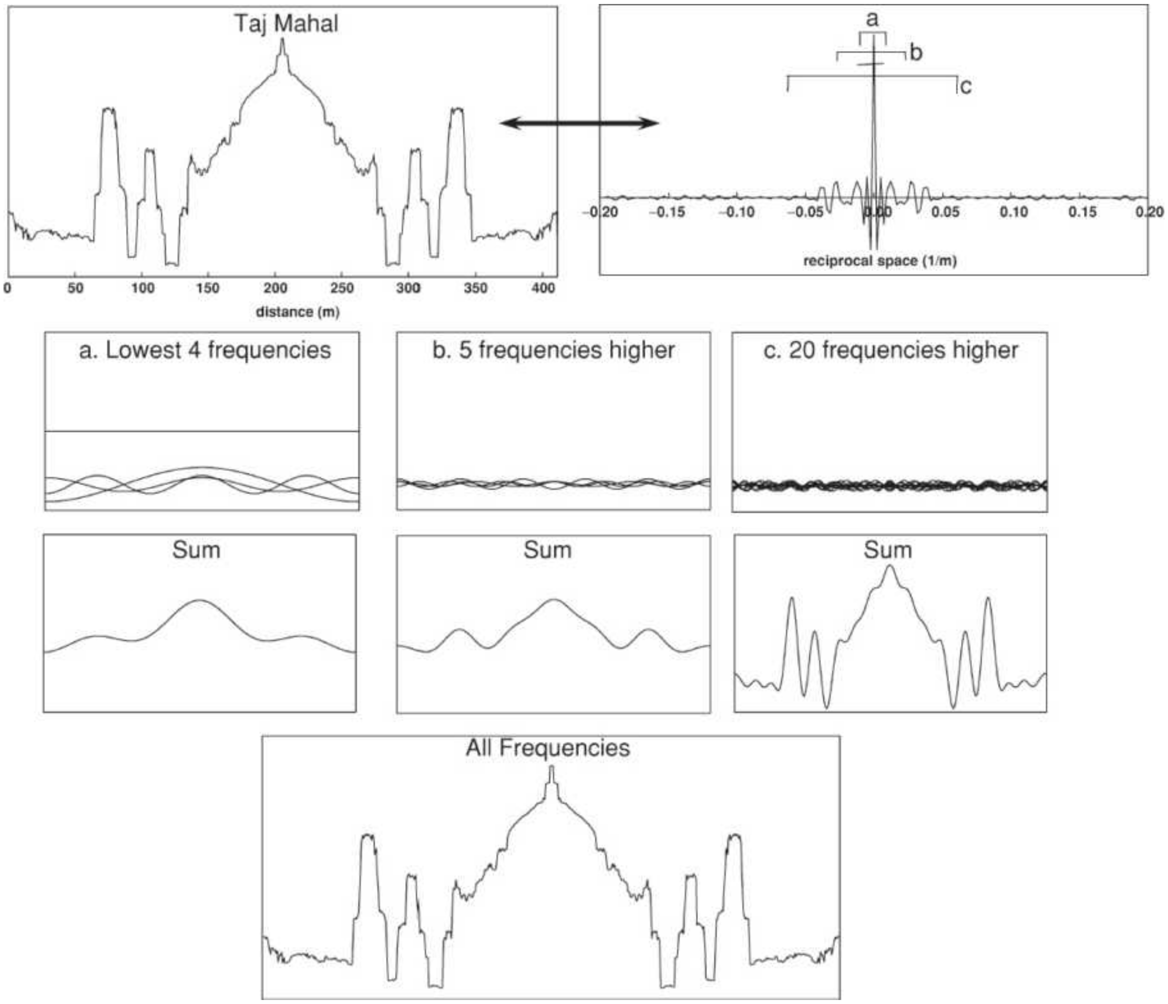

FIGURE 1.25–7. The Fourier transform is based on the premise that the shape of any object can be created with the right combination of sinusoidal waves. In this example, a profile of the Taj Mahal is shown on the **upper left**, and its Fourier transform is shown on the **right**. If the center group of frequencies, labeled **a**, are plotted and summed (**two center-left panels**), then a general shape of the profile is achieved. If the next group of frequencies, labeled **b**, are plotted and added (**two middle panels**), then the shape of the center dome begins to take shape. The next band of frequencies, labeled **c**, can be included (**right center panels**) to show the presence of the pairs of towers on each side of the central dome. When all frequencies are included, the original picture is seen (**bottom**). Two principles are shown by this example: An object can be expressed as a series of sinusoids, and the details of an object lie primarily with the frequencies that lie farthest from zero.

Aside from its appearance in dreams, the Taj Mahal may have little to do with psychiatric brain imaging, but it illustrates the fundamentals of the Fourier transform. Signals in MRS are acquired as a function of time, just as one can record sounds as time-dependent vibrations (Fig. 1.25–8A). The Fourier transform of the sounds reveals the frequencies and volumes of sound that comprise a note, just as the Fourier transform of an MRS signal reveals the frequencies and amplitudes of signals that uniquely identify neurochemicals that make up the FID (Fig. 1.25–8B). The MR signals in brain imaging are acquired in the *k*-space domain and undergo Fourier transformation to generate the standard anatomic pictures. Low-frequency data are those that reside near the center of *k*-space, and just as for the picture of the Taj Mahal, the data that resides near the center of *k*-space lack details but indicate primarily how large the head is and that the shape is roughly circular. The information about sulci and other fine details is found farther from the center of the *k*-space (Fig. 1.25–9). Schemes exist to acquire only the center or the outer edges of *k*-space to save time in cases that do not require all of the information about the image, but the great majority of brain imaging is done with standard coverage of *k*-space.

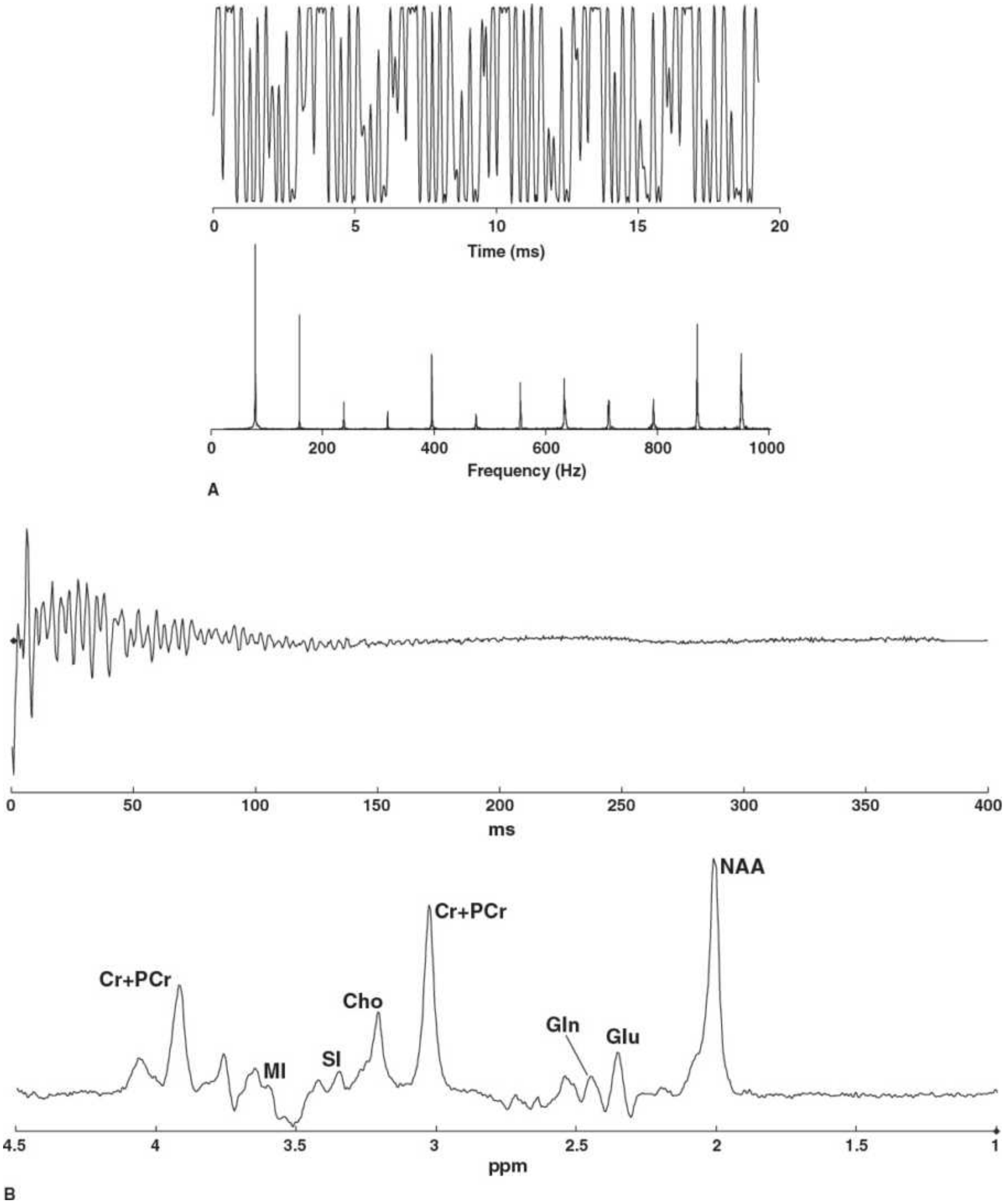

FIGURE 1.25–8. The Fourier transform allows time-dependent signals to be expressed in terms of frequency, and in this respect, MRI is indistinguishable from sounds. **A:** The upper graph shows the sound of a note played on a saxophone and cymbal by the author's daughter and son. Fourier transformation reveals some of the frequencies that are generated from the instruments. The richness of the sound of a saxophone derives largely from the combination of frequencies and various amplitudes of tones that it emits, and the cymbal's characteristic sound derives from the multitude of

frequencies in addition to its primary pitch at 540 Hz. **B:** The MRS signal, viewed relative to the Larmor frequency, is indistinguishable from sound; the MRS signal also appears as a time-dependent oscillation. Some early MR scanners were connected to audio speakers to allow the operators to make quick evaluations of signal quality. In this case, the Fourier transform converts a complicated time-domain measurement into a spectrum with easily identifiable features that are typical for a sample of healthy brain tissue. The volume contains *N*-acetylaspartate (NAA), creatine + phosphocreatine (Cr + PCr), myoinositol (MI), choline, scylloinositol (SI), glutamate (Glu), and glutamine (Gln).

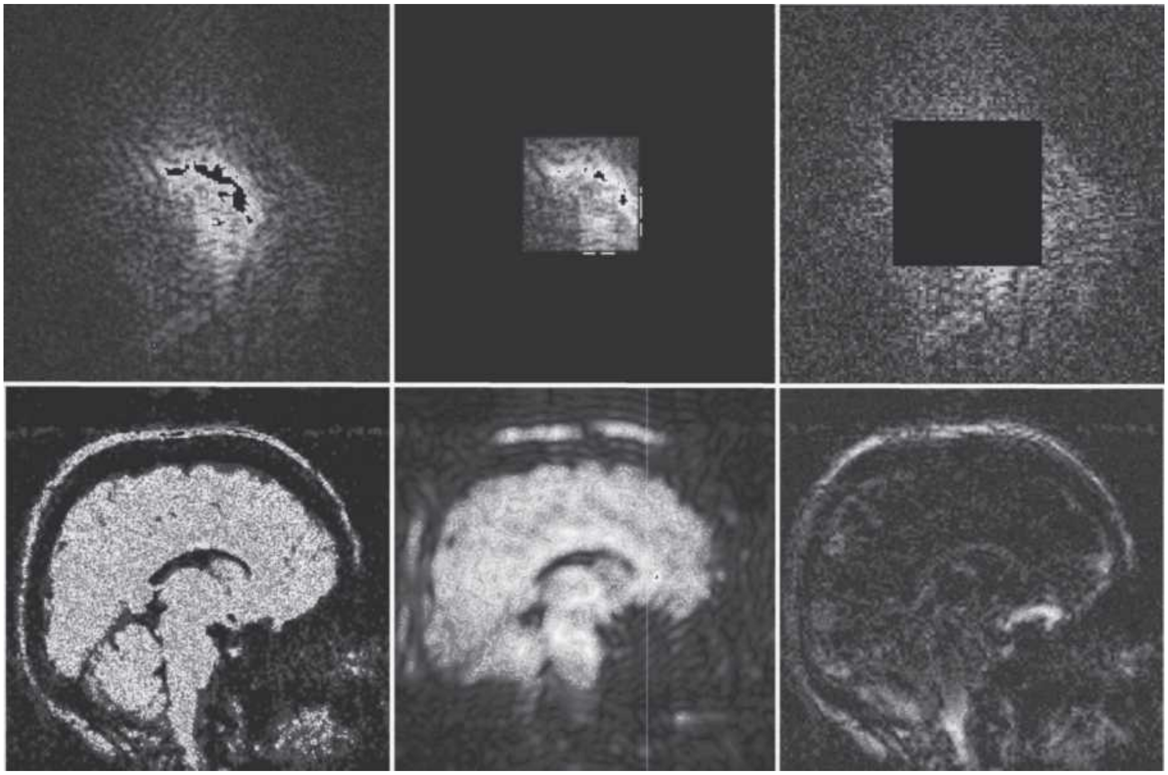

FIGURE 1.25–9. *K*-space data (**top**) paired with associated images (**bottom**). Typical *k*-space coverage is uniform and yields a standard brain image (**left pair**). If only the center of *k*-space is acquired (**center pair**), then the image loses much of its detail, which is analogous to the example of the reconstruction of the profile of the Taj Mahal in the earlier example. If only the outer regions of *k*-space are acquired (**right pair**), then only the fine details remain.

MRI METHODOLOGY

MRI methodology is about seeing boundaries and differences in tissue composition, which requires the generation of *image contrast*. The word "contrast" is often used clinically to mean an injectable *contrast agent,* but the general meaning of contrast in MRI is anything that differentiates the appearance of one material from another. To that end, the contrast in brain imaging is obtained by exploitation of differences in T_1, T_2, blood oxygenation, diffusion coefficient, and concentration of water or fat. The most basic neuropsychiatric application of MRI is the mapping of brain anatomy, and this is generally done on the basis of the natural contrast from T_1, T_2, and T_2* of brain water. On the side of clinical applications, the purpose is often to check for particular causes of symptoms, such as tumors or multiple sclerosis. All image sequences consist of one or more region-specific radio frequency excitations followed by applications of gradients to scan *k*-space. The resulting *k*-space distributions undergo Fourier transformation to create images. By convention, MRI results are displayed such that the bottom of sagittal or coronal images points toward the feet and the right side of axial or coronal images points toward the patient's left.

The most basic MRI sequences are called *gradient-echo* and *spin-echo.* The gradient-echo sequence is simply an excitation pulse in conjunction with gradients to localize the sources of the signals (Fig. 1.25–10). For each slice, the measurement consists of a slice-selective excitation pulse applied while a slice-selection gradient is applied. The longer the pulse and the stronger the gradient, the thinner the slice will be; scanners equipped with stronger gradients are generally able to provide thinner slices. A *phase-encoding gradient* is then applied perpendicular to the slice selection in one of the in-plane dimensions to shift the acquisition to a fixed distance in *k*-space; larger phase-encoding gradients shift the acquisition farther in *k*-space. After the phase-encoding gradient, a preparatory gradient is applied in the second in-plane dimension to shift the acquisition in the second dimension of *k*-space. Finally, a *read gradient* is applied in that second dimension of *k*-space. The process is repeated, using a slightly different phase-encoding gradient strength each time to provide appropriate coverage of two dimensions of *k*-space, allowing the generation of a two-dimensional image. With a gradient-echo image, signal loss occurs due to distortions in the magnetic field, such as seen near the sinuses and other tissue/air/bone interfaces, which can be avoided with a spin-echo sequence. The spin-echo sequence contains all of the elements found in a gradient-echo sequence, but with the addition of a 180-degree pulse called a *refocusing* pulse applied halfway between the excitation pulse and the acquisition, one effect is to refocus magnetization lost due to magnetic field inhomogeneities and thereby recover the signal, such as near the sinuses. A spin-echo sequence also provides a way to obtain tissue contrast and visualize certain pathologies, as will be discussed shortly.

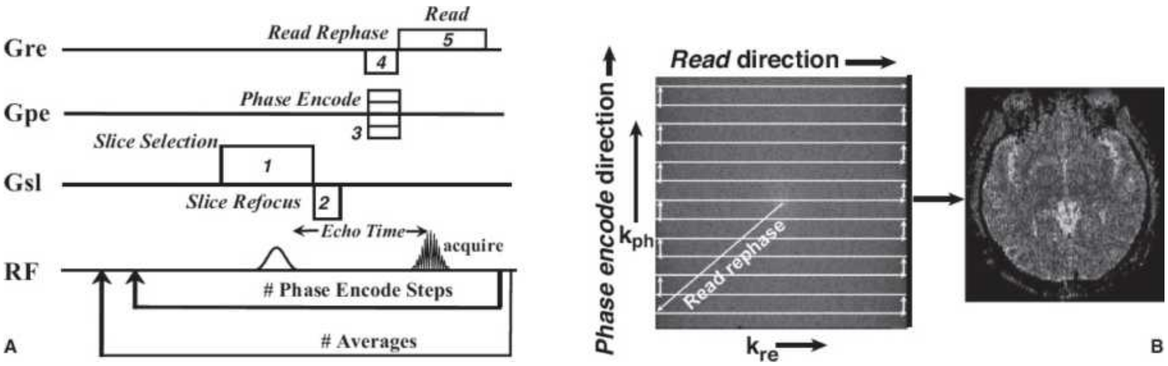

FIGURE 1.25–10. Coverage of *k*-space for brain imaging. **A:** The gradient-echo sequence is a pulse-acquire measurement that includes gradients to localize the signal. (1) A slice-selection gradient is applied with a frequency-selective pulse to excite the magnetization in a slice. (2) The slice selection gradient disperses the magnetization, so a slice refocus gradient is applied to refocus it before the acquisition. (3) Phase encode gradients move the magnetization in the phase encode direction. (4) A read rephase gradient opposite to the read gradient is applied to move to one edge of *k*-space. (5) A read gradient moves the signal through *k*-space in the read direction. **B:** The *k*-space data are shown at the left, with indications of the path that was covered by the read and phase-encoding gradients. Transformation of the *k*-space data results in the image shown at the right.

T_1 Weighting, Injected Contrast Agents, and T_2 Weighting

T_1-based contrast exploits the different rates of realignment of spins with the static magnetic field. T_1 weighting is exploited in different ways, including injections of exogenous agents, but a common noninvasive approach is *inversion recovery* (IR), with timing set to null, or blacken, regions of the image with the long T_1 of CSF. IR imaging works on the principle that the detectible *XY* magnetization can only be created from *Z* magnetization, which means that the elimination of *Z* magnetization in the CSF will eliminate the source of the signal in areas that contain CSF. When a 180-degree pulse is applied to the magnetization, the net magnetization is inverted, and it immediately begins to return to the alignment with the magnetic field. The T_1 of white matter is the smallest, so the magnetization in white matter recovers first. The T_1 of CSF is the longest, so its magnetization recovers most slowly, and if the images are acquired near the point at which its Z magnetization reaches 0, the CSF will appear black in the images. The gray matter has an intermediate T_1, so its value will be recovered enough to detect but not as much as that of white matter: Gray matter will appear significantly darker than white matter. Figure 1.25–9 shows such an image. T_1-weighted images are commonly used for structural analysis, including volumetric research.

Another way to obtain T_1-based contrast is with the use of an injectable agent, most of which contain gadolinium. The contrast agents pass through the blood but do not cross the healthy blood–brain barrier (BBB). The gadolinium that is trapped in the molecule decreases the T_1 of the hydrogen nuclei in the water around it and brightens the image intensity wherever it appears. In cases in which the BBB has been disrupted, as is the case with some tumors, the gadolinium agent enters the abnormal tissue easily, and the region appears bright in the images (Fig. 1.25–11).

Losses of signal due to the T_2 decay of the water signal can be used to highlight CSF relative to white matter and gray matter, which lose signal significantly more quickly during echo delays. A T_2-weighted image typically uses a spin-echo type of image, which combines an excitation pulse to create XY magnetization and then one or more 180-degree pulses to track the T_2 decay that progresses with time after the excitation pulse. In the case of multiple sclerosis (Fig. 1.25–12), the fluid in pockets within the ventricles, sulci, and lesions have longer T_2 values than the other tissue. If a long echo delay is used, then the signals from gray and white matter decay to 50% or less of their maximum values, while the signal from the fluid remains strong.

One disadvantage of T_2-weighted images is the poor contrast between gray and white matter. However, an advantage is that the types of image sequences used for T_2 weighting typically eliminate inhomogeneity-related signal loss, called T_2*-loss, which typically darkens images near the ear canals and sinuses. The sequences employed for T_2 weighting avoid this behavior.

Most MR images are acquired to evaluate structure. A treating physician may be interested in checking for tumors or multiple sclerosis, as already demonstrated in Figures 1.25–11 and 1.25–12. Other common clinical purposes are the detection of atrophy or cerebral malformations.

An area of structural MRI that may move beyond its current status as a research tool is volumetric analysis (Fig. 1.25–13). A variety of approaches exist, both manual and automated, but their purpose is to measure the volumes of brain regions, often with separate quantification of white matter and cortical tissue. The specific goal of the research is usually to characterize particular populations of patients with respect to the volumes of various brain structures, commonly the frontal lobe, the hippocampus, or the corpus callosum.

Diffusion MRI

The diffusivity of water can provide a mechanism of contrast for MRI. The most common form, *diffusion-weighted imaging* (DWI), reduces the signal in each image pixel based on the *diffusivity* of water. The process by which the signal is reduced is similar to that of T_2* decay: Gradients are used as an artificial source of T_2* with a strong spatial dependence. A spin-echo–type sequence is applied with the diffusion gradients on each side of the refocusing pulse. Nuclei that diffuse very little between the gradients experience the same magnetic field strength and are refocused by the spin-echo, but nuclei that diffuse appreciably *along the direction of the gradient* experience two different magnetic fields and are incompletely refocused, yielding less signal than is seen for the stationary nuclei. DWI can reveal abnormalities in otherwise normal-appearing tissue, as in regions of edema that surround a cerebral infarct. Regions with edema show greater signal reductions because water can diffuse more readily when the tissue structure is disrupted and filled with more fluid.

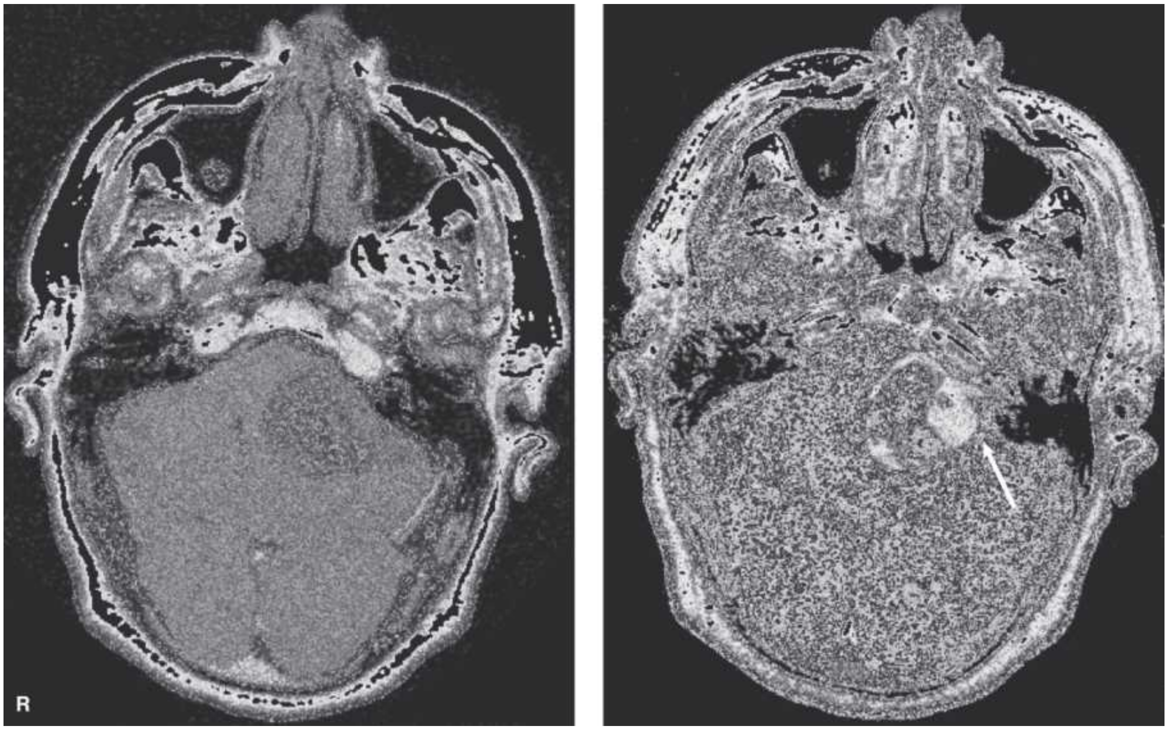

FIGURE 1.25–11. Acoustic schwannoma in the left cerebellar–pontine angle. The T_1-weighted image before the injection of the contrast agent (**left**) demonstrates a mass in the left cerebellar–pontine angle. After administration of the gadolinium contrast agent (**right**), blood vessels are highlighted, and because the tumor is not protected by the blood–brain barrier, the mass is more clearly characterized as being partially cystic. There is enhancement in the internal auditory canal (*arrow*), which is consistent with an acoustic schwannoma.

Water does not diffuse equally in all directions in the brain because the microenvironment restricts the mobility of water. For example, in white matter, water diffuses much more easily along the length of the fiber tracts than across them. That means that fiber tracts that run in the same direction as the diffusion gradient will show a greater signal reduction than fibers that run perpendicular to the orientation of the diffusion gradient. Application of diffusion gradients in six or more different orientations allows the directionality of the diffusion of the water, the *diffusion tensor,* to be measured for every image pixel, generating *diffusion tensor images* (DTIs). Measurement of the orientation of water diffusion permits a measurement of *fractional anisotropy,* which provides a measure of the randomness of tissue structure. For example, if tissue is highly ordered, then the diffusion of water will be restricted and more oriented than is the case for tissue that is disordered. A more calculation-intensive use of DTIs comes from following the diffusion tensors through chains of pixels to track fibers. Fiber tracking allows detailed analyses of anatomic connectivity (Fig. 1.25–14), which holds great potential clinically but, for now, remains a research tool.

Imaging Blood Flow

MRI is used clinically to image blood flow in the head. A physician may need to differentiate between a degenerative disorder and a compromised blood supply for a region of the brain. For that purpose, *MR angiography* (MRA) and perfusion MRI can provide helpful information. The example of Figure 1.25–15 shows *time-of-flight* angiography, which uses a gradient-echo measurement to make images in which inflowing water has more magnetization than stationary spins and thereby appears brighter than the surrounding tissue. Another approach to image perfusion is to use an injection of an intravascular contrast agent that enhances the T_1 relaxation and increases the signal on a T_1-weighted MRI. A T_1-weighted MRI is obtained without the agent and then with the agent, and the image pixels that are associated with flowing blood will be brighter with the contrast agent.

Another approach is *arterial spin labeling* (ASL), which is based on the manipulation of nuclei in water outside an image slice and the detection of changes in image intensity as those nuclei arrive in the slice of interest. One approach is to acquire two images, one a baseline image and the other identical, but for the application of an inversion pulse (*spin tagging*) to generate negative magnetization in the water in a slab of the neck, including any water that happens to be flowing up the arteries when the inversion pulse is applied. If images are acquired through slices of the brain, then any of the water that flows with the blood up from the neck will have a negative signal. Upon subtraction of the pair of images from one another, pixels that represent static brain water will approach zero, but pixels that lie in the path of blood flow will appear up to twice as bright as those in the baseline image. In this way, blood flow can be mapped without the injection of a contrast agent.

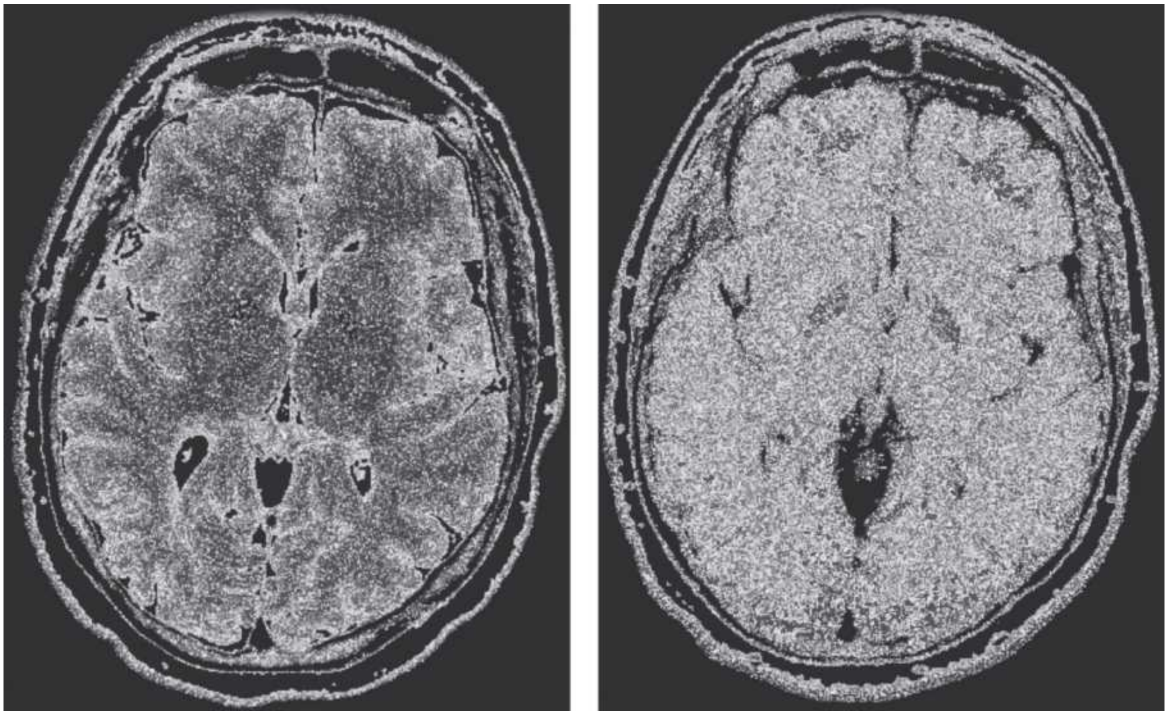

FIGURE 1.25–12. Multiple sclerosis (MS) in periventricular regions. **Left:** A fast spin-echo image enhances materials with large values of T_2, including cerebral spinal fluid (CSF) in the sulci and the ventricles. The MS lesions are enhanced, but it is difficult to differentiate them from the normal CSF. The periventricular MS lesions have T_1 values that are significantly lower than the value for CSF, so T_1 weighting can be used to differentiate the lesions from fluid. **Right:** The lesions are more conspicuous with the fast spin-echo, fluid-attenuated inversion recovery image (FLAIR) image. The FLAIR sequence incorporates an inversion-recovery pulse that uses T_1 relaxation to suppress signals from the CSF, making it easier to detect lesions that are adjacent to CSF spaces.

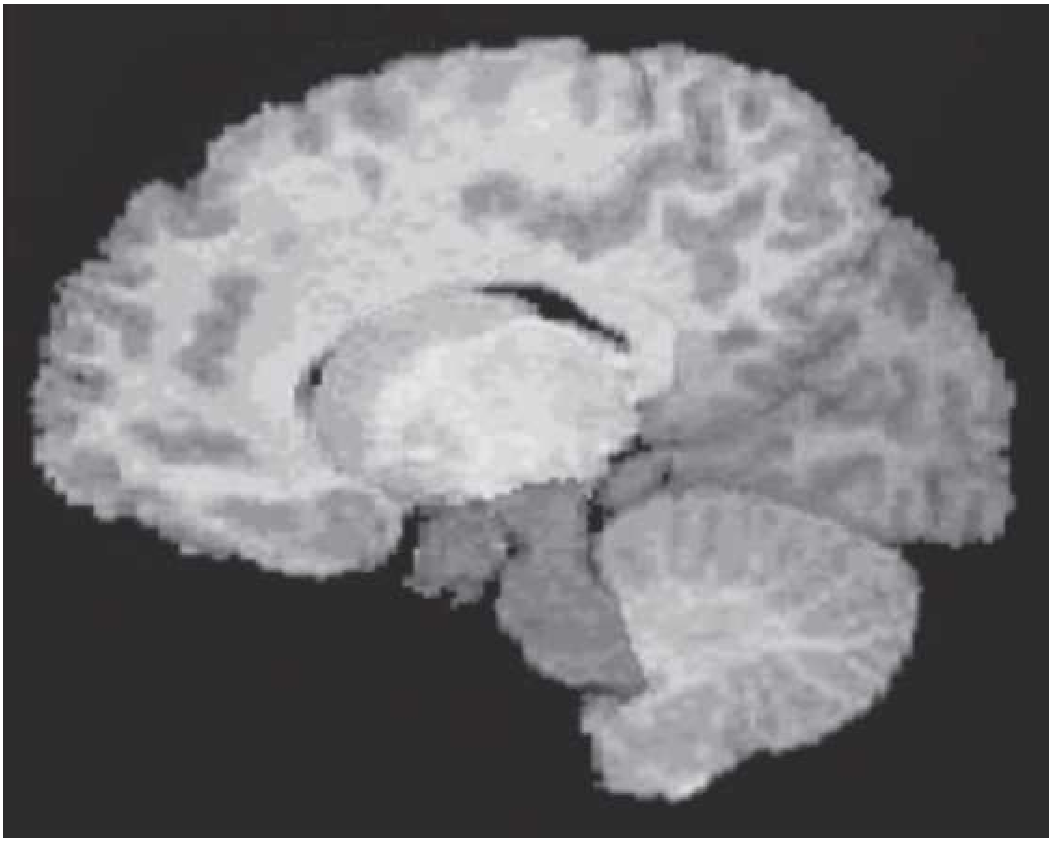

FIGURE 1.25–13. Structural T_1-weighted image that has been *segmented* by brain region for volumetric analysis in a group of patients with posttraumatic stress disorder. In this study, sets of contiguous slices were acquired to cover the entire brain, and then anatomic landmarks were used to demarcate brain regions. (Courtesy of Andrea Jackowski, Ph.D., Universidade Federal de São Paulo, Brazil, and Joan Kaufman, Ph.D., Johns Hopkins School of Medicine.)

APPLICATIONS OF STRUCTURAL IMAGING TO PSYCHIATRY

To date, the clinical utility of structural imaging in psychiatry has been limited to ruling out organic brain diseases. However, the methodology has provided information that has helped to shape many of the current theories related to the pathophysiologic mechanisms associated with several psychiatric disorders, suggesting that both neurodevelopmental and neurodegenerative abnormalities are associated with the various disorders. The use of imaging to contribute to the understanding of disease forms a key part of the RDoC criteria of NIH. Recent advances in the methodology now provide promise that the technique may soon find clinical utility in the diagnosis and prognosis of some psychiatric disorders.

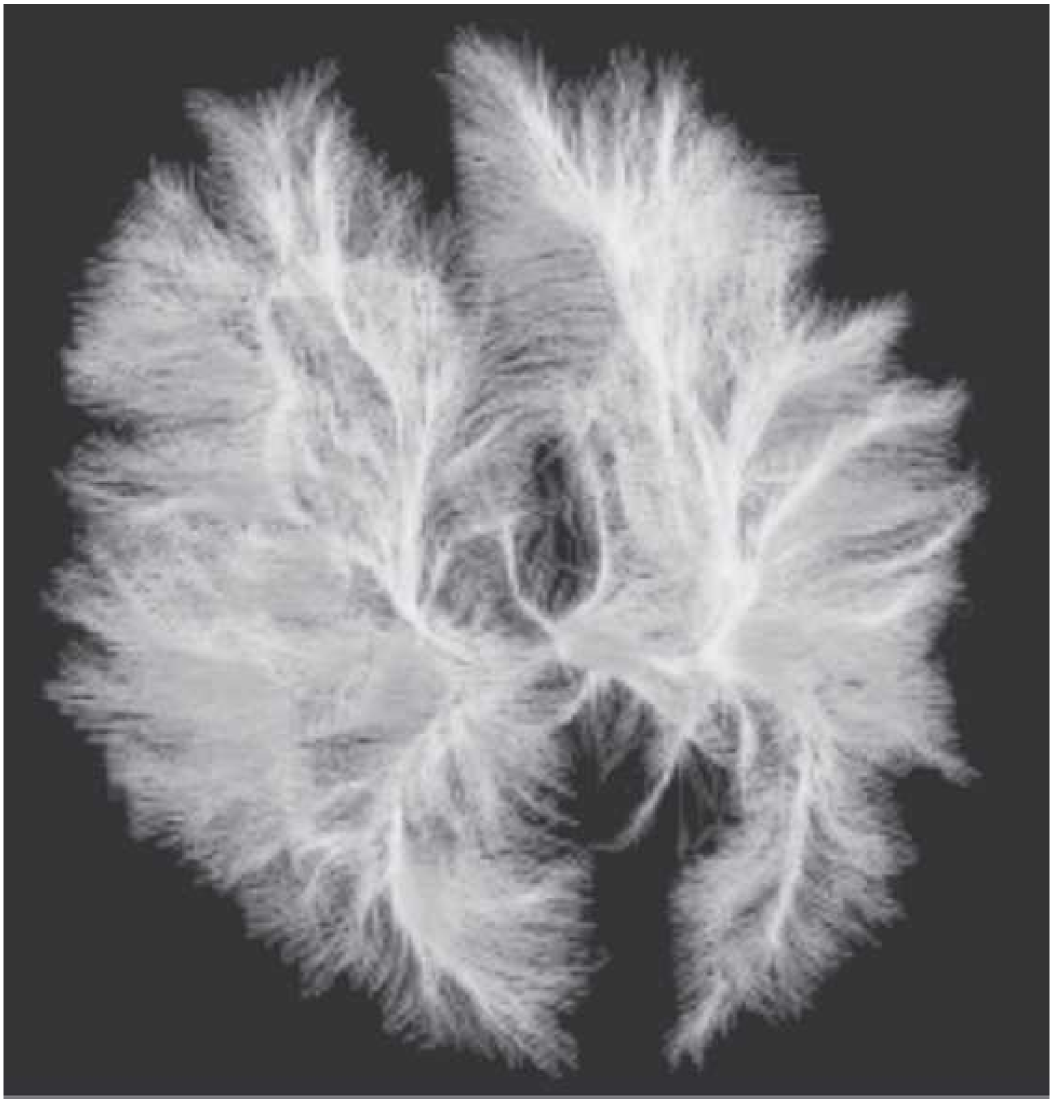

FIGURE 1.25–14. Fiber tracks determined from DTI are presented in an axial perspective. The diffusion tensor was mapped by the acquisition of images with diffusion weighting in multiple directions. A numerical integration technique followed the principal direction of diffusivity from pixel to pixel and slice to slice to track fibers through the brain. (Courtesy of Marcel Jackowski, Ph.D., Universidade de São Paulo, Brazil, and Todd Constable, Ph.D., Yale University.)

MRI Applications to Dementia

The diagnosis of dementia comprises several illnesses that are characterized by memory loss and other cognitive disturbances. The most common illness associated with dementia is Alzheimer disease, which is characterized by the presence of amyloid plaques and neurofibrillary tangles in several cortical and subcortical limbic regions. The disease is associated with neuronal and glial pathology, as well as global and regional brain atrophy and loss of gray matter. Structural imaging is now beginning to emerge as a clinically relevant component of Alzheimer disease diagnosis. Although the current diagnostic criteria for Alzheimer disease used by the National Institute of Neurological Disorders and Stroke—Alzheimer Disease and Related Disorders and *Diagnostic and Statistical Manual of Mental Disorders,* 4th edition, text revision (DSM IV-TR) do not consider MRI data in the diagnostic process, a recent position paper published on the basis of an international conference held to reevaluate the National Institute of Neurological Disorders and Stroke—Alzheimer Disease and Related Disorders diagnostic criteria suggested that MRI be used diagnostically as a supportive feature in the diagnosis of Alzheimer disease.

Several MRI changes, including increased numbers of subcortical hyperintensities, generalized atrophy, and ventricular enlargement, are associated with normal aging. However, it is well established that some changes seem more specific to the diagnosis of Alzheimer disease and may be clinically useful in formulating the diagnosis and prognosis of the disorder. MRI evidence of medial temporal lobe (MTL) atrophy appears to be most closely associated with the disorder. MTL atrophy is extremely common in individuals with Alzheimer disease, with rates reported to be between 71 and 96 percent depending on disease severity. MTL atrophy is often observed in the early stages of mild cognitive impairment (MCI), making it potentially useful as a prognostic tool. However, the sensitivity and specificity associated with the current MRI methodology limit its clinical usefulness in prognostic imaging. One approach that may help to improve the clinical utility of MRI in the diagnosis and prognosis of Alzheimer disease and other forms of dementia is to follow the rate of change in brain structures over time. Longitudinal follow-up studies have shown the rates of volume loss to be significantly greater in subjects with prodromal Alzheimer disease (up to 5 percent brain volume/year) compared to those experiencing normal age-related reductions (0.1 percent brain volume/year). Other methodologic advances are also in development that may help to improve the clinical utility of MRI in the diagnosis of Alzheimer disease. Methods providing an improved ability to rapidly and reliably assess the volumes of specific subregions of the MTL, such as the hippocampus and the entorhinal cortex, provide some promise. Several recent studies suggest hippocampal volume loss determined by MRI volumetric analyses in subjects with MCI is capable of predicting conversion to Alzheimer disease in the subsequent years. Further advances in the ability to assess the specific subfields of these regions could lead to dramatic improvements in the specificity and sensitivity of the studies. Examples of these methods include voxel-based morphometry, an image analysis technique that can simultaneously identify multiple areas of cortical and subcortical degeneration, and surface mapping techniques that use specific cortical features as constraints to guide the matching of anatomy across subjects to provide information on cortical gray matter thickness for specific brain regions. All of these techniques are aided in part by the development of what are today considered high-field (3 to 4 T) and ultrahigh-field magnets (7 T and above). The possibility of increasing accuracy by the addition of other measures like flurodeoxyglucose positron emission tomography (FDG-PET) is showing promise. Although these techniques were originally developed for neuroscience research, their potential clinical relevance is becoming appreciated in providing increased structural resolution and even the potential application of imaging microscopy.

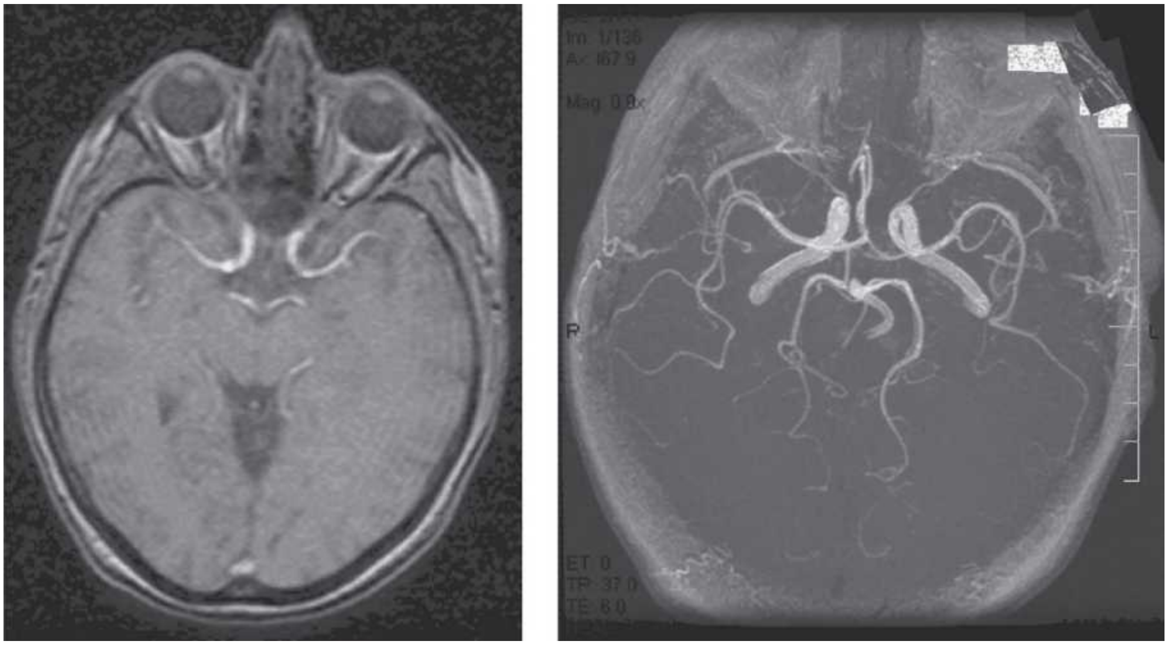

FIGURE 1.25–15. Imaging vasculature by *time-of-flight* (TOF) angiography. **Left:** Images are acquired with rapid repetitions that saturate the signal from nonvascular nuclei, thereby reducing the signals from tissues other than flowing blood. **Right:** After being processed to select image regions that show enhanced signal intensity, flowing blood is identified.

DTI studies have identified significant frontal, temporal, and parietal white matter diffusion tensor changes in MCI and Alzheimer disease patients that appear to correlate with cognitive functioning. Other recent studies have also found relevant microstructural changes in the normal-appearing white matter of patients with Alzheimer disease and MCI, although fractional anisotropy may not be particularly sensitive to the disorder. The changes are consistent with a hypothesized loss of axonal processes in the frontal and temporal regions in Alzheimer disease.

MRI Applications to Schizophrenia

Structural neuroimaging studies have provided some of the most important evidence of abnormal cortical development in patients with schizophrenia. Initial studies reported consistent cerebroventriculomegaly, reductions in total cortical volumes, and regional volumetric deficits, particularly in frontal and temporal cortical regions. One consistent exception to this rule was the striatum, where treatment with typical antipsychotics produces increased volume in patients compared to healthy subjects or never-medicated patients. With the application of segmentation techniques to separate cortical gray and white matter within MR images, it became clear that both tissue compartments showed regional deficiencies in patients. The application of more recent techniques has enabled investigators to describe abnormal patterns of cortical gyrification, curvature, and reduced cortical thickness in patients, although to date, abnormalities are insufficient to discriminate patients from healthy comparison subjects on an individual basis.

Structural MRI also has been useful for characterizing features of the heritable risk for schizophrenia. Schizophrenia is associated with a heritability estimate of 70% to 90%. Consistent with this view, family members of individuals with schizophrenia show a pattern of reductions in cortical gray and white matter volume that resembles but is milder than that associated with schizophrenia. Both the degree of polygenic risk for schizophrenia based on common gene variants that individually have a very small effect and rare mutations that have a significant impact on schizophrenia risk are associated with brain structural changes as measured by MRI.

Structural MRI also has provided some insights into the progressive course of this disorder and the impact of antipsychotic treatment. The most striking findings have been in childhood-onset schizophrenia, where a series of neuroimaging studies using a variety of MRI approaches, including specified regions of interest, automated voxel-based morphologic analysis, and deformation analysis (an extremely sensitive approach for measuring within-subject change), have clearly shown that the development and evolution of schizophrenia are associated with a progressive deviation of regional cortical volumes from those of healthy comparison populations. In adolescents and young adults who manifest symptoms of the schizophrenia prodrome, "ultra high risk" state, or who are in their first episode of schizophrenia, many of the changes associated with chronic schizophrenia are already present. Despite earlier data to the contrary from computed tomography (CT) imaging, there are now a growing number of studies in adults with chronic schizophrenia that indicate that even at this stage of the disorder, there is subtle but progressive cortical gray and white matter volume loss in a number of cortical regions, relative to time-dependent volumetric reductions in age-matched healthy subjects. In an era where nearly all patients with schizophrenia receive some degree of pharmacotherapy, studies face the challenge of distinguishing the long-term trajectory of cortical volume changes associated with the illness from the long-term effects of antipsychotics that might emerge over decades of treatment. In this regard, it is interesting to note that some data suggest the possibility that treatment with some "second generation" or "atypical" antipsychotic medications may reduce the short-term, that is, 1 to 2 years, progression of cortical volume reductions or the typical antipsychotic treatment-related increase in basal ganglia volume in patients with schizophrenia. There is hope that these types of studies might reveal neuroprotective effects of future pharmacotherapies for schizophrenia.

DTI has been employed in a growing number of studies that have attempted to characterize disturbances in the organization and integrity of many major white matter pathways associated with schizophrenia. Because DTI measures the movement of water molecules, it is particularly sensitive to the constraining effects of neural projections and their associated glial sheaths upon the normally random diffusion of water, as well as the impact of disturbances in neural and glial integrity. DTI findings in schizophrenia include (1) abnormalities in the degree of anisotropy (lack of organization in the direction that water molecules move within a given voxel), presumably reflecting altered integrity or degree of branching of cortical white matter pathways, and (2) reduced total diffusivity (total water movement within a given voxel), a dependent measure related to anisotropy and related to the integrity of white matter pathways, and (3) reduced size, integrity, or branchedness of specific white matter pathways identified using DTI tractography. Using standard techniques, one cannot differentiate the contributions of local dendritic fields, local axonal pathways, and converging inputs from distant neurons.

MR-based perfusion studies have begun to complement other techniques, principally PET and single-photon emission computed tomography (SPECT), in describing cortical perfusion deficits associated with schizophrenia. An early generation of PET studies of perfusion and metabolism deficits associated with schizophrenia yielded the impression that schizophrenia was associated with resting “hypofrontality,” that is, a relative deficit in the metabolism or perfusion of the frontal cortical structures. However, a careful review of this first generation of studies did not yield a firm conclusion on the nature and or validity of resting hypofrontality associated with schizophrenia. More recently, PET, SPECT, and MR-based perfusion studies have been used to characterize task-related alterations in regional cortical perfusion associated with schizophrenia. The novel opportunity provided by MR-based perfusion studies, largely unrealized, is to take advantage of the potential for absolute perfusion quantification in order to reduce interpretation errors when comparing across subjects or across test days where there might be baseline differences that cannot be appreciated using semiquantitative imaging techniques, such as Echo-Planar Imaging (EPI) with Blood-Oxygen Level-Dependent (BOLD) contrast. Studies of cortical perfusion in schizophrenia have been useful to help to evaluate the potential direct vascular effects of substances that might also be of interest with respect to their modulation of neural and glial function, as in the case of nicotine.

MRI Applications to Mood Disorders

Older MRI studies evaluating larger brain regions have yielded few replicable findings related to mood disorders beyond vague reports of lateral ventriculomegaly. However, more recent findings examining specific subregions of the brain have yielded interesting and replicable findings of smaller volumes in several brain regions, including the basal ganglia, thalamus, hippocampus, frontal lobe, orbitofrontal cortex, and gyrus rectus. Perhaps MRI’s greatest contribution to the field of mood disorders research comes from the studies demonstrating reduced hippocampal volume in individuals suffering from major depressive disorder (MDD). Hippocampal volumetrics have been studied in several neuropsychiatric disorders. Reduced hippocampal volumes have been reported in epilepsy, Parkinson disease, Huntington disease, Cushing disease, Alzheimer disease, traumatic brain disorder as well as alcoholism, obsessive-compulsive disorder (OCD), and schizophrenia. Despite the confounds introduced by study heterogeneity (age, gender, duration of illness, and responsiveness to treatment) that characterizes the studies in MDD, a 2004 meta-analysis of 12 MRI studies revealed a significant reduction in hippocampal volume in both hemispheres of individuals with MDD. Average reductions of 8% on the left and 10% on the right were reported. Other evidence suggests that the total number of episodes, duration of illness, and exposure to childhood abuse may further contribute to the degree of volume reduction associated with MDD. This data has been taken as strong support for a model of stress-induced hippocampal toxicity involving glucocorticoid receptor activation, hypoxia, and elevated levels of glutamate neurotransmission. However, recent studies also suggest that pediatric patients with familial depression also have reduced hippocampal volumes, suggesting that reduced hippocampal volume may be suggestive of a risk factor for developing MDD.

The anterior cingulate gyrus is another region that has gained much attention due to findings of significant abnormalities associated with mood disorders. Structural MRI has, in one study, demonstrated reductions of 39% and 48% in the mean gray matter volume in the prefrontal cortex ventral to the genu of the corpus callosum in both familial bipolar depressives and familial unipolar depressives. Although this finding has not been replicated consistently in later structural imaging studies, functional imaging studies have consistently found abnormalities in this brain region, and histologic studies have further identified significant glial cell reductions in this area.

Patterns of structural abnormalities in neural regions, including the hippocampus and anterior cingulated gyrus, supporting emotion processing were recently found to differentiate MDD and bipolar disorder (BPD), suggesting a multivariate pattern classification techniques may hold promise as future diagnostic aids. Similar to the case in dementia, future studies using higher field strengths and other methods to allow for improved measures of hippocampal subfields are likely to provide marked advantages over the currently used methods.

Findings remain inconclusive with regard to the amygdala, for which several studies have reported no relationship between amygdala volume and mood disorders, while other studies have found either an increase or a decrease in amygdala volumes in mood disorder subjects compared with those of healthy control subjects. Other studies have attempted to define morphometric abnormalities in the basal ganglia related to mood disorders. While there is some evidence suggesting a possible decrease in basal ganglia volume associated with MDD and a potential increase associated with BPD, in general, the data remain inconclusive. Recent advances with high-field magnets such as 7 T may clarify the disparate results, possibly even providing a way to differentiate bipolar psychosis from schizophrenia.

In the area of late-life depression, structural MRI findings have had a large impact on guiding the direction of research. White matter hyperintensities (WMHs) are areas of increased signal intensity that are particularly apparent on T_2-weighted images. As mentioned above, they are commonly associated with aging and are thought to reflect diverse neuropathologic underpinnings related to ischemia, hypertension, myelin pallor, inflammatory disease, amyloidosis, and other causes. WMHs are considered important markers in the study of the consequences of vascular risk factors on the brain and are commonly interpreted as a hallmark of chronic cerebral ischemia. The consistent findings of significantly increased numbers of WMHs in individuals with late-life depression, especially late-onset depression, suggest an etiologically different subtype of MDD. These findings have fueled the cerebrovascular hypothesis of late-life depression that posits that white matter lesions resulting from cerebrovascular disease disrupt fiber tracks within the frontostriatal circuitry.

Recent studies suggest that white matter changes predate the development of late-onset depression and that the severity of the changes predicted subsequent depressive symptoms in the following year even after controlling for age, baseline cognitive function, and depressive symptoms. This adds support to the claim of a causative relationship, but it remains to be determined how important the location and lateralization of the actual WMHs are to the pathogenesis of late-life depression or whether they are merely a marker of progressing vascular disease. Future standardized and automated grading systems of WMH, currently in development, should greatly facilitate the use of this modality in clinical practice.

MRI Applications to Alcohol Dependence

MRI studies have been the principal tool to describe in vivo the many sources of neurotoxicity associated with alcoholism, including (1) the direct neurotoxic and gliotoxic effects of ethanol, (2) the neurotoxic effects of poor nutrition that often accompany the abuse of alcohol, (3) the excitotoxicity associated with the ethanol withdrawal state, and (4) the possible disruption in adult-neurogenesis–associated ethanol intoxication and withdrawal. These studies documented a striking age dependence on the overall neurotoxicity associated with alcoholism. Alcohol dependence in adolescents has been associated with pronounced and lasting reductions in cortical volume, suggesting a disruption of a neurodevelopmental process associated with the familial risk for alcoholism or exposure to alcohol. In healthy young adults, careful studies suggest that there are relatively subtle neurotoxic effects of ethanol that may not be present in most heavy drinkers. However, with the onset of “middle age” and progressively thereafter, coincident with increasing cumulative exposure to alcohol and bouts of alcohol withdrawal, there appears to be a progressive sensitivity to alcohol-related neurotoxicity. This toxicity is reflected as loss of both gray and white matter, and these losses appear to be related to cognitive impairments. Gray matter microstructure has been reported to change first, and the damage is more severe in those who drink early, possibly because early heavy drinkers often proceed to heavier drinking than people who begin drinking at a later age. DTI studies provide evidence of disturbance in the integrity of white matter tracks associated with alcoholism. The volumetric changes are particularly severe in patients with Wernicke–Korsakoff syndrome, arising from nutritional deficiency. Generally, there is a gradual and partial recovery of volume loss with extended sobriety, that is, a more rapid phase of volume restoration over several months with some continued recovery over several years. White matter shows evidence of earlier recovery than gray matter, although relapse leads to accelerated white matter damage. Structural MRI research has disproven an early hypothesis that the volumetric brain changes associated with alcohol dependence and the initiation of sobriety reflect the dehydration and rehydration of the brain, respectively.

FUNCTIONAL MRI METHODOLOGY

Brain function consumes energy, which is supplied by blood flow to carry oxygen and nutrients to the tissue. Changes in brain function lead to alterations in hemodynamics and blood oxygenation, and such changes affect the magnetization that is detected by MRI. The effects are mediated through the magnetic susceptibilities of oxygenated and deoxygenated hemoglobin, working through local T_2* effects. Deoxygenated hemoglobin perturbs the magnetic field homogeneity in its vicinity, while oxygenated hemoglobin does not. When the level of deoxyhemoglobin increases, the signal in the vicinity of the blood decreases and lower deoxyhemoglobin leads to larger signals around blood vessels. This leads to the phenomenon of BOLD contrast. Sets of images are acquired under two or more conditions, using different functional or pharmacologic challenges, in which case the technique is sometimes called *ph*MRI, and changes in brain function are detected as increased or decreased image intensity (Fig. 1.25–16).

Because brain function is linked to hemodynamics and the extraction of oxygen from the blood, the level of oxygenated hemoglobin depends on brain function. Paradoxically, when brain activity increases, the fraction of deoxygenated hemoglobin *falls* because blood flow increases so dramatically that it overcomes the extraction, and the MR signal in the vicinity of the blood vessel rises. Therefore, if MRI data are acquired in one condition, and then while the patient is performing some cognitive task that increases the oxygen extraction in Area 9, for example, the signal in Area 9 will be higher, with an increase typically on the order of 1 or less. If a cognitive task is selected that instead decreases the utilization of Area 9, then the second image will be darker by a small percentage in Area 9.

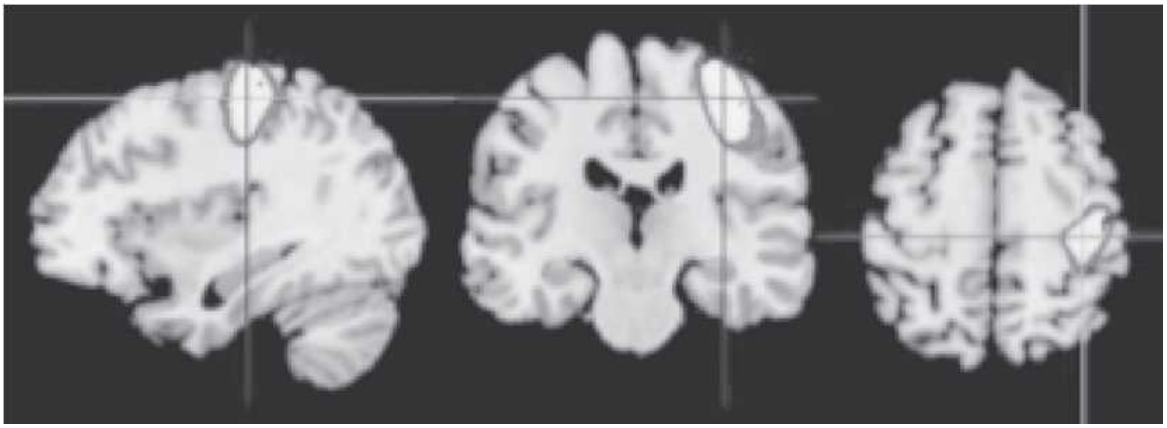

FIGURE 1.25–16. Functional MRI at 3 T during right-hand movement, with sagittal (**left**), coronal (**center**), and axial (**right**) views. The T_1-weighted image forms the base of the display for anatomic identification. A series of images were acquired with the hand alternating between stationary and moving conditions, and then the groups of images were compared to identify regions that showed statistically significant changes in signal during hand motion. The left motor cortex showed activation, with regions of most intense change highlighted.

Although functional magnetic resonance imaging (fMRI) is often said simply to reflect blood oxygenation and/or blood flow, the contrast in fMRI depends on a variety of sources that vary in their degrees of specificity. The mechanisms depend on the diameter of the blood vessels, the blood volume, the rate of blood flow, the amount of deoxygenated hemoglobin in the blood, the strength of the magnet used for the scans, and the method of MRI acquisition. For blood vessels that are larger than 20 μm, it is the *static dephasing* that dominates the signal reduction from deoxygenated hemoglobin. The static dephasing results from the proximity of deoxyhemoglobin in water in the tissue around a blood vessel, and it can be eliminated with a refocusing pulse in a spin-echo–type MRI sequence. For capillary beds and other small blood vessels, the dominant mechanism is *intravascular dynamic dephasing* and *extravascular dynamic dephasing,* which occur when water molecules wander randomly through the area of a distorted magnetic field that surrounds the deoxyhemoglobin. Although intravascular dynamic dephasing provides information about activation, it yields signals that are not always located at the site of activation. The intravascular dynamic dephasing dominates the BOLD response at 1.5 T but exerts progressively smaller effects as field strengths rise. At high magnetic fields such as 7 T, there is a negligible contribution from intravascular dynamic dephasing. The degree of signal change from these mechanisms is also influenced by the orientation of each blood vessel with respect to the static magnetic field B_0. The random orientations in capillary beds yield random shifts in B_0 that, once averaged, reduce the signal in a particular image pixel. Larger vessels exert a net change in phase that will also reduce the signal but will do so in an orientation-dependent manner. A change in oxygenation in a large vessel that flows straight across the thickness of a slice will yield an fMRI change that differs from that of a vessel of the same size that flows along the plane of the slice: A different slice orientation can yield a different degree of fMRI signal change even with the same functional task.

Another mechanism of contrast in fMRI is inflow. During an fMRI study, the nuclei in each image slice are repeatedly excited, and the repeated excitations lead to *saturation.* That is, the observed signal from a series of rapid scans is smaller because too little time is allowed for full T_1 relaxation to the maximum Z magnetization. The saturation only occurs for nuclei that have been excited, which are those that are in the slice and remain in the slice. Nuclei that are carried into the slice by blood flow between the scans have experienced no saturation and therefore have greater Z magnetization. They contribute more Z magnetization per unit volume than their resident counterparts. The more saturated the magnetization within the slice, the greater the impact of inflow contrast will be. Inflow contrast is greater when faster scan repetitions and larger slice excitation angles are used. If the blood flow increases, then the spins that arrive will be less influenced by the slice excitation, so increased blood flow contributes a greater signal to the fMRI scan. Any influence that increases blood flow will increase the signal intensity from blood vessels, which may or may not lie in a region of activated tissue, so the inflow contrast contributes nonspecific changes to the fMRI maps.

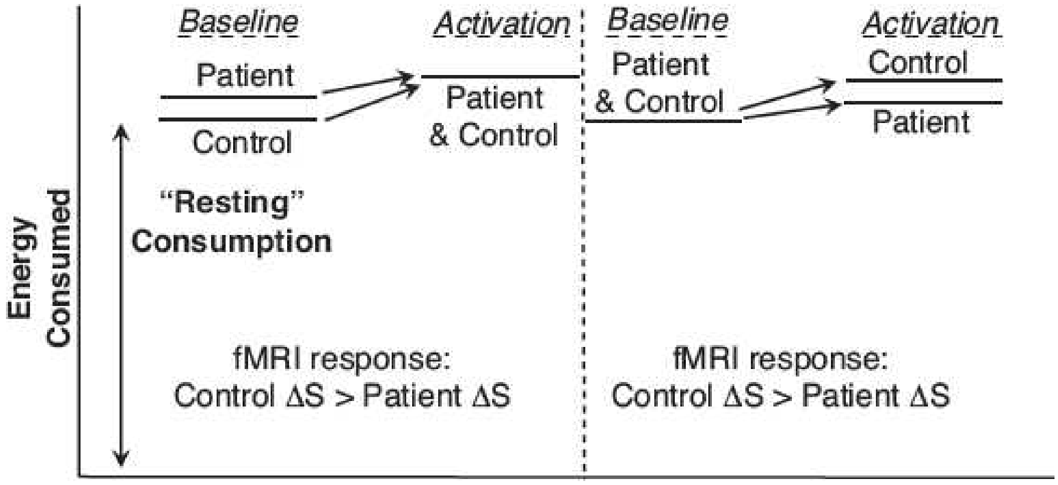

FIGURE 1.25–17. Relative changes of brain energy utilization when changing functional state and their impact on fMRI. The baseline state in the awake human corresponds to a high level of neuronal activity, blood flow, and metabolism. Changes in brain function are typically represented by small changes in brain energy utilization and, as a result, small changes in fMRI image amplitudes. Because fMRI reflects conditions relative to a large baseline, the interpretation of fMRI responses should consider baseline conditions. **Left:** The patient has a higher baseline activity than the control subject, and when the functional task shifts both to the same energetic condition, the patient shows a smaller fMRI response than the control subject. **Right:** The patient and the control subject have the same baseline energy utilization, but the control subject increases energy consumption more than the patient, such that, in this case also, the patient has a smaller fMRI response than the control subject. Other factors that may lead to ambiguities are differences in cerebral vascular coupling to energy demand due to disease or pharmaceutical interventions.

Once the fMRI signal changes have been evaluated, an additional factor must be considered: The meaning of the baseline image. A healthy adult might show an increase of 1% of signal amplitude in the dorsolateral prefrontal cortex (DLPFC) while performing a memory task, while a patient shows a 2% rise in the same region for the same task. In both cases, the baseline activity of the neurons is significant with respect to a variety of pertinent parameters, including electroencephalography (EEG), perfusion, and metabolic measurements (Fig. 1.25–17). One interpretation is that the patient's neural circuitry functions less efficiently and requires more energy to do the same work. An equally viable explanation is that the patient had less baseline activity than the healthy individual but that when the task began, the same region in the patient's DLPFC behaved normally and used a similar amount of energy. In both cases, an abnormality has been identified, but the understanding of the underlying pathology differs with the interpretation. When the challenge is pharmacologic, an additional layer of complexity should be considered: The potential influences of the drug, like acute or chronic tobacco use or heavy drinking on the cerebral vasculature, are independent of the functional demands of the tissue.

Taken together, the numerous mechanisms of contrast show that fMRI maps must be considered with respect to their method of acquisition, slice orientation, magnetic field strength, and pharmacologic influence on the cerebral vascular system. The various sources of contrast must be considered when evaluating the meaning and quantity of change that is observed.

APPLICATIONS OF FMRI IN PSYCHIATRY

fMRI is a noninvasive neuroimaging technique that can be used to study the neural correlates of complex cognitive, emotional, and perceptive processes. fMRI research holds tremendous promise for advancing the current understanding of the pathophysiology associated with psychiatric disorders where gross brain structure is preserved, but disruptions of brain function exist. By revealing the neural correlates of symptoms, cognitive biases, and emotional responses, fMRI studies have already been used to elucidate the pathophysiology of dementia, mood, anxiety, psychotic, and addictive disorders and the mechanisms of psychopharmacologic treatments for these conditions. To date, the clinical capabilities of fMRI for informing diagnostic or treatment decisions have not been established. The abnormalities identified by fMRI have thus far lacked sufficient sensitivity and specificity to discriminate individual patients from healthy subjects or from subjects with other illnesses. However, with future advances in the field, such as quantitative fMRI, it is hoped that the technique may ultimately lead the way to a pathophysiology-based classification of psychiatric phenotypes that can be used to improve both research and clinical practice. Examples of such research are provided below.

fMRI of Dementia

fMRI methods provide information that can potentially be used in the study, diagnosis, and prognosis of Alzheimer disease and other forms of dementia, as well as providing insights into normal age-related changes in cognitive processing. Evidence that aging is associated with weaker and more diffuse activations, as well as decreased hemispheric lateralization, suggests either compensation for lost regional intensity or dedifferentiation of processing. The weaker activations, especially prefrontally, suggest potential encoding-stage dysfunctions associated with aging. The alterations in these processes that occur with neurodegenerative disease appear to be superimposed on the course of normal aging. fMRI studies have consistently demonstrated that patients with Alzheimer disease have decreased fMRI activation in the hippocampus and related structures within the MTL during the encoding of new memories compared to cognitively intact older subjects, including alterations in what is called the default mode network. More recently, combined PET/structural and fMRI revealed that decreased connectivity was associated with more white matter lesions and greater amyloid burden. Further studies are needed to validate the use of fMRI studies in these populations, particularly longitudinal studies, to investigate the pattern of alterations in functional activity over the course of prodromal Alzheimer disease and the relationship to Alzheimer disease pathology. A major factor to consider is how age- and disease-related changes in vasculature affect the representation of neuronal function by fMRI.

fMRI of Schizophrenia

fMRI has emerged as the primary approach for probing disturbances in the activity of particular brain regions and specified circuits associated with the risk for developing schizophrenia, the symptoms and cognitive impairments associated with schizophrenia, and the impact of antipsychotic treatments. Given the large number of published studies (over 8,700 related papers since 1987 according to PubMed), the range of MR-based approaches included within the rubric of fMRI, and the enormous range of cognitive tests that have been used to stimulate cortical activity, a comprehensive review of fMRI studies of schizophrenia is beyond the scope of this review.

Studies of Low-Frequency Cortical Connectivity

Studies of functional connectivity of brain regions as assessed by fMRI complement the structural evidence emerging from DTI studies that the organization and integrity of white matter tracks are compromised in patients with schizophrenia. In these studies, functional connectivity is evaluated by quantifying the covariance of cortical activity across brain regions at rest over extended periods of time.

fMRI studies of perception and cognition alterations associated with schizophrenia support the central hypothesis emerging from MRI, DTI, and functional connectivity studies, that is, that there are widespread alterations in regional brain functional activation that disturb the normal pattern of circuit activation associated with perception and higher cognitive functions including attention, learning and memory, and judgment. Here the neuroimaging issues related to this research are discussed. A common assumption of this research is that these alterations in regional brain activation constitute the basis of the impairments in perceptual and cognitive processes associated with schizophrenia. While this assumption probably applies in a general way to these studies, EPI-BOLD imaging is actually measuring blood flow and volume changes rather than directly measuring neural activation and, as reviewed earlier, current clinical research imaging, typically at 3 T, has modest temporal and spatial resolution with respect to the processes that it attempts to measure. Therefore, it is more accurate to assume that these studies provide an indirect measurement, a biomarker, of neural and glial processes that predominantly occur with spatial and temporal characteristics that exceed the resolution of fMRI. An additional complexity is that the relationship between the demands of a task and its ability to stimulate circuit activation is often nonlinear. That is, it follows an inverted-U pattern. This issue emerged in the study of working memory deficits in schizophrenia, where it was first shown that patients exhibited reduced prefrontal cortex activation when performing tasks that put demands on working memory. Later, using tasks of graded difficulty, it was shown that patients actually showed prefrontal hyperactivation relative to healthy subjects in response to easy tasks but began to show functional activation deficits in response to moderately difficult tasks when healthy individuals were showing further increases in prefrontal cortex activation.

There are a number of additional challenges associated with the interpretation of fMRI studies of schizophrenia. First, because of widespread disturbances in brain structure, it is difficult to ascribe the underlying abnormality to the structure that performs abnormally, as opposed to its inputs. This limitation has profound implications for cognitive functions that are organized in a hierarchical or “bottom-up” fashion. That is, judgment depends on memory, which depends on attention and perception. But there are also executive or “top-down” controls of these processes. That is, we are more effective in perceiving and remembering stimuli that we expect or seek. From this perspective, the effort to determine the “central” cognitive impairments of schizophrenia and to ascribe particular cognitive impairments to particular brain regions is a challenging task. Second, because of their circuitry dysfunctions, individuals with schizophrenia may perform more poorly on tasks, or they may successfully perform particular tasks differently than individuals without schizophrenia, analogous to the way that individuals with hearing deficits learn to read lips. With regard to performance failure, studies typically select tasks that can be performed successfully by individuals with and without schizophrenia. However, there are many questions related to the performance deficits themselves. In this case, studies often use tasks with graded levels of difficulty so that the relationships between task difficulty, cortical activation, and performance may be determined. With regard to the second issue, it can be very difficult to know whether divergence between people with and without schizophrenia arises from the primary circuitry dysfunction or from the different strategies that these groups tend to use to solve the same types of problems. This highlights a critical issue for functional neuroimaging, which is the need to define the structures of what people do and how they do it. Since one cannot obtain perfect control of these issues, it is useful and important to evaluate performance and, if possible, strategy. The effort to map circuitry dysfunction associated with schizophrenia, already challenging for the reasons outlined above, is made more difficult because the engagement of brain circuits in cognitive processes is highly state dependent and therefore varies with the level of emotional arousal, the impact of other state-dependent processes (such as the symptoms of schizophrenia), the presence of other neurobiologic modulators (such as the impact of therapeutic medications, exposure to substances of abuse, and other psychiatric and medical comorbidities).

fMRI of Mood Disorders

Over 1,200 fMRI studies exploring various aspects of circuitry related to mood and mood disorders have been reported. Studies investigating the neural circuitry mediating adaptive emotion regulation and cognition have been instrumental in advancing the current understanding of the pathophysiology of mood and anxiety disorders. Functional imaging studies using PET, SPECT, and fMRI spectroscopy performed over the last decade have provided us with a much greater understanding of the structure and functional correlates of emotional and cognitive processing as well as mood regulation. Since individuals showing vulnerability to or suffering from episodes of depression, mania, and anxiety appear to exhibit impaired cognitive processing and a decreased ability to effectively regulate mood and affect. These findings are considered highly informative in identifying regions of specific interest to the neurobiology of the disorders. Functional brain imaging studies have identified critical neural circuits that modulate emotional behavior. Several frontolimbic networks including (1) the limbic–thalamic–cortical (LTC) circuit, comprising the amygdala, mediodorsal thalamus, and orbital and medial prefrontal cortices, and (2) the limbic–cortical–striatal–pallidal–thalamic (LCSPT) circuit, which includes components of the LTC circuit as well as areas of the striatum and pallidum, have emerged. Reciprocal connections between the various regions within the circuits are believed to modulate emotional responses. The use of fMRI to study the contributions of these networks to the pathophysiology of mood and anxiety disorders provides several unique advantages. The modality offers excellent spatial and temporal resolution relative to other methodologies, allowing the integration of cognitive and emotional tasks that serve as probes to activate the specific networks and further elucidate the interconnectedness of the individual structures. However, it is important to understand that the measures are always made relative to other measures of brain activation and cannot be considered absolute measures of brain activity.

Although, as would be expected based on the heterogeneity of the illnesses and the multitude of study designs employed in the exploration, the method has yielded many inconsistent findings related to the function of these networks in individuals with mood and anxiety disorders, there is a growing consensus that disruptions in the function of these circuits do contribute to the pathophysiology of these disorders. For example, several groups have found activation of the amygdala, a region involved in the identification and signaling of emotionally significant stimuli, to be increased and abnormally sustained during emotional task performance in subjects diagnosed with mood and anxiety disorders. Interestingly, abnormal amygdalar activation has also been observed in healthy individuals with the short allele of the functional promoter polymorphism for the serotonin transporter gene, suggesting a possible link to the increased stress-related susceptibility to depression associated with the genotype.

The cingulate cortex, another region commonly found to exhibit abnormal patterns of cerebral blood flow (CBF) and metabolism in mood and anxiety disorder, patients receive inputs from the thalamus, as well as several cortical regions, and projects to the amygdala and entorhinal cortex via the cingulum. It functions as an integral component of the limbic system, being specifically involved in emotional processing, learning, memory, and the suppression of inappropriate unconscious priming. Activation of the subgenual anterior cingulate cortex (ACC) has repeatedly been measured to be decreased in unipolar and bipolar depression relative to control samples. However, after volume reductions in this region identified by MRI studies have been accounted for, it is possible that the remaining tissue in this area actually exhibits relative hypermetabolism. Depressed patients have been found to have greater activation of this region during the processing of sad stimuli.

Abnormal activation patterns of two other prefrontal cortical regions have also consistently been found to be associated with mood and anxiety disorders. Several studies have reported abnormally elevated levels of CBF and metabolism in the orbital cortex as well as in the ventrolateral prefrontal cortex (VLPFC) (a region believed to play a role in the learning of new affective stimulus associations and the reversal of previously learned associations) in individuals with mood and anxiety disorders. Dysfunction of these regions is postulated to predispose individuals to the perseverative cognitions and emotional responses characteristic of depression. fMRI studies have demonstrated an inverse relationship between the amygdala and ventromedial prefrontal cortex (VMPFC) activation when healthy individuals reappraise the affective meaning of negative pictures. Individuals showing increased VMPFC and decreased amygdala activation during reappraisal show steeper, more adaptive diurnal patterns of daily cortisol, suggesting that the capacity to successfully modulate limbic circuitry has implications for well-being in daily life. A recent study demonstrating a significantly different correlation between VMPFC and amygdala activation in control versus depressed individuals during an effortful affective reappraisal task further supports this notion.

The DLPFC, a region specifically activated by cognitive tasks related to working memory and attention, has been consistently shown to exhibit abnormal reductions in CBF and metabolism associated with depression. The reduced activity of this region in response to cognitive task performance has been related to the deficits in cognitive performance characteristically observed in conjunction with mood disorders. Several fMRI studies demonstrate decreased DLPFC activity in response to cognitive tasks. Similar to the VLPFC, recent studies also find that depressed individuals display decreased relationships between the amygdala and DLPFC activity, potentially signifying decreased functional relationships among these structures.

The thalamus and the basal ganglia have extensive connections with the amygdala as well as the orbital cortex, VLPFC, and cingulate cortex. In addition, the ventral striatum receives key projections from the ventral tegmental area (VTA), comprising part of a critical reward pathway of the brain. Unipolar depressed and bipolar depressed subjects both manifest abnormal increases in metabolism and CBF in the left mediodorsal nucleus of the thalamus. Functional imaging studies of these regions in mood disorder subjects have shown them to have a decreased striatal response to happy stimuli during episodes of depression and increased striatal activity in a manic state.

In summary, functional imaging studies afford a new window into the neural circuitry of emotional and cognitive processing, thus providing novel information that can be used in the generation and testing of new hypotheses. In conjunction with work derived from other branches of cognitive neuroscience, these studies have led to the general hypothesis that mood and anxiety disorders result from impaired function of the LCSPT circuits responsible for regulating emotional and cognitive processing. Specifically, a “bottom-up” model has been presented in which overactivation of limbic structures such as the amygdala drives enhanced emotional responsiveness to stimuli and further decreases the ability to effectively cognitively process the incoming stimuli due to reciprocal inhibition of the DLPFC. A second “top-down” model postulates that the ability of higher-order cortical structures to modulate limbic activation is impaired in individuals with mood and anxiety disorders. Several studies have used fMRI protocols to identify specific activation patterns reflecting the top-down or bottom-up models that seem to be associated with treatment response. Additional studies are now underway to further explore the potential clinical relevance of these models and to determine if the patterns of activation could be used to predict differential treatment responses to the various treatment interventions available.

fMRI of Alcohol Dependence

fMRI studies have provided insights into the functional consequences of alcoholism-related neurotoxicity. Studies suggest that recovering alcohol-dependent patients show abnormal activation patterns in the frontal cortex, thalamus, striatum, cerebellum, and hippocampus related to impairments in attention, learning and memory, motor coordination, and inhibitory control of behavior. Studies have begun to explore pharmacologic modulation of resting circuit activity to probe mechanisms underlying circuit dysfunction in alcoholism, illustrated by blunted responses to benzodiazepines.

Studies also have provided insights into the vulnerability to alcoholism. Adolescents and adults who abuse alcohol or who are at risk for abusing alcohol show circuit-related dysfunction associated with reward anticipation, response inhibition, control of attention, and other dimensions of cognition. In addition, these individuals show increased limbic and orbitofrontal cortex activation when exposed to alcohol-related cues that elicit alcohol craving, associated with structural abnormalities of the brain. Studies are now attempting to utilize these craving-related changes in fMRI for testing putative pharmacotherapies for alcoholism.

^{1}H MRS

In 1983, a spectra of small metabolites, including glutamate and GABA, were observed in the living rat brain by Kevin Behar in the laboratory of Robert Shulman at Yale University. Since that time, MRS studies of the brain have expanded enormously in their variety and in their applications to the study of the human brain. Today, MRS is applied throughout much of the world for diagnosis and basic research on the brain in health and disease.

MRS employs the differentiation of molecules, primarily on the basis of their frequencies. The most commonly used nuclei have been ^{1}H, ^{31}P, and ^{13}C. Figure 1.25–9B shows an example of a ^{1}H MR spectrum obtained in a human brain. The primary components of the MR spectrum are methyl groups of creatine and phosphocreatine, choline, and *N*-acetylaspartate

(NAA). The two creatine compounds have resonances that overlap in vivo, so they appear as a single peak. Although creatine and phosphocreatine play key roles in cellular energetics, their combined level remains constant in the face of even severe acute challenges such as ischemia. Some chronic effects such as aging have been reported to increase, decrease, or have no effect on levels of tissue creatine, and this is one of the areas that are currently under active investigation in the MRS community. The choline resonance, comprised primarily of phosphocholine and glycerophosphocholine, is believed to reflect membrane degradation and synthesis, with a negligible contribution from acetylcholine. NAA is found primarily in glutamatergic neurons, including neuronal processes, and it is synthesized by *N*-acetyltransferase, a mitochondrial enzyme. Glia have transporters to import NAA and also contain the enzyme that degrades NAA, raising the possibility of a shuttle of NAA from neurons to glia. Indeed, there is evidence that the acetyl group of NAA is incorporated into myelin. Until the early 1990s, decreases in NAA were believed to signify neuronal death, but since that time, studies have found conditions in which NAA decreases and then recovers. NAA is now believed to represent neuronal health, viability, or mitochondrial integrity.

Much of MRS research has been based on these three metabolites that dominate the ^{1}H MR spectrum. Clinical applications have been proposed, including the diagnosis and evaluation of Alzheimer’s disease (AD), hepatic encephalopathy, and other disorders. Other metabolites of neurochemical interest that can be detected using ^{1}H MRS are myoinositol, glutamate, glutamine, GABA, and glutathione (GSH), and these have all been scrutinized from a research perspective in a variety of neuropsychiatric disorders. Glutamate, glutamine, and particularly GABA have required advanced MRS methods for detection in the human brain because they overlap with one another and are modulated into complex patterns under the influence of *J*-coupling. Methods classed as *spectral editing* are used to isolate the signals of GABA, glutamate, GSH, and other compounds from larger, overlapping resonances. Research is currently under way to assess if and how total tissue levels of glutamate and glutamine with respect to neurotransmitter activity and whether any of the MRS-detected chemicals can indicate neuroinflammation.

Magnetic resonance spectroscopic imaging (MRSI), also called *chemical shift imaging* (CSI), combines the power of chemical discrimination with the ability to map their spatial distribution, and it is most commonly implemented for ^{1}H MRS. Because the signals from most detected neurochemicals are typically five to six orders of magnitude lower than that of water, the signal-to-noise ratio is much lower for MRSI measurements, significantly longer acquisition times are required to obtain adequate sensitivity, and the resolution is lower than what is normally used for MRI. However, the ability to view large areas of the brain simultaneously has led to the beginnings of clinical applications of this approach to conditions in which abnormalities must be localized, as in epilepsy, and conditions of dispersed changes in the brain, such as multiple sclerosis and alcoholism.

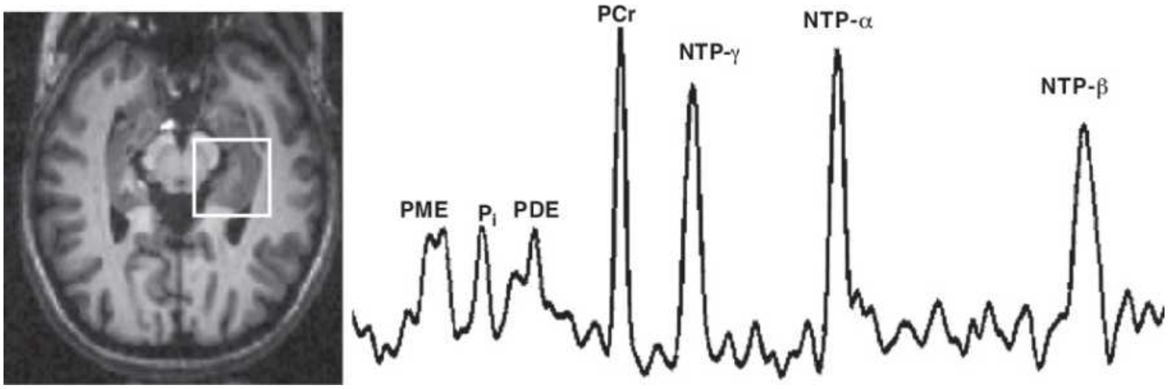

FIGURE 1.25–18. ^{31}P MRS in a healthy volunteer, using a 4-T magnet. **Left:** T_1-weighted image obtained with inversion recovery to null the CSF and create contrast between gray and white matter. **Right:** Spectrum that shows, from left to right, phosphomonoesters (PMEs), inorganic phosphate (P_i), phosphodiesters (PDEs), phosphocreatine (PCr), and the γ, α, and β phosphate residues of the nucleotide phosphates. (Courtesy of Jullie Pan, M.D., Ph.D. University of Missouri Columbia.)

^{31}P MRS

In 1978, Britton Chance and colleagues published MR spectra of the phosphorylated metabolites in the brain of a living mouse. ^{31}P MRS yields measurements of high-energy compounds, including phosphocreatine and adenosine triphosphate (ATP), and it can provide measurements of phosphomonoesters (PMEs) and phosphodiesters (PDEs) (Fig. 1.25–18), with applications to study schizophrenia, depression, and other disorders. The ^{31}P isotope nucleus is much less sensitive than that of ^{1}H on a per-atom basis, and further sensitivity is lost when one considers that many of the ^{1}H resonances represent methyl or ethylene groups that provide three or two nuclei per resonance instead of just one. Favoring sensitivity is the fact that 100% of naturally occurring phosphorus is the isotope ^{31}P. The unique information that it provides on high-energy phosphates and other metabolites without destroying the tissue make it a powerful tool for brain research. Some laboratories are applying MRSI of ^{31}P to map these neurochemicals throughout the brain.

The dominant resonance in the ^{31}P spectrum of the brain is that of phosphocreatine, which by convention is assigned a chemical shift of 0 ppm. To the right are three resonances that represent the γ, α, and β resonances of nucleotide triphosphates (NTPs). To the left are the PME, PDE, and inorganic phosphate. The PME and PDE resonances are believed to reflect mobile components of lipids and membrane metabolites. Inorganic phosphate is free phosphate that has been released from phosphocreatine and ATP during energy consumption. In cases of extreme insults, such as ischemia, the phosphocreatine resonance falls, the NTP resonances decrease, and the inorganic phosphate peak rises as the high-energy phosphate pool is exhausted and the phosphate groups are released. In psychiatric disorders, the changes are generally more subtle, with some reports of altered PME or PDE, although the directionality and extent of the changes are uncertain at this time. Recent advances have moved kinetic studies of ^{31}P metabolism to application in the brain, from muscle, where they have been used for many years. A technique called *saturation transfer* reduces the signal of ^{31}P in one chemical, like phosphocreatine, and then tracks the signal reduction in ATP when the phosphorus is transferred from phosphocreatine to ATP. This approach has allowed the measurement of phosphocreatine and ATP synthesis in the brain.

^{13}C MRS

Less common is ^{13}C MRS, used to detect the nonradioactive isotope carbon-13, which naturally occurs as 1.1% of the world's carbon. Like most ^{1}H MRS and all ^{31}P MRS, ^{13}C MRS remains a research tool, and in that role, it has proven useful for the measurement of the kinetics of neuroenergetic processes and neurotransmission. Its primary application in the brain has been the measurement of carbon oxidation and glutamate–glutamine neurotransmitter cycling.

^{13}C carries the powerful advantage of a broad chemical shift dispersion that allows the resolution of many metabolites that cannot be resolved in the ^{1}H MR spectrum. For example, glutamate and glutamine C4 and C3 resonances are completely resolved even at a magnetic field strength of 1.5 T, with the resolution improving further at higher field strengths. The great disadvantage of ^{13}C is that its sensitivity is even less than that of ^{31}P on a per-atom basis, and the sensitivity is further reduced by its low natural abundance. However, that low natural abundance can be turned into an advantage if viewed from the perspective that it contributes very little background signal (Fig. 1.25–19). That low level of background means that if ^{13}C is introduced into metabolic pathways and is incorporated into neurochemicals, then almost everything that is detected is known to result from the added carbon. This was first done in cell suspensions and then rabbits and rats and is now applied by several laboratories to studies of metabolism in the human brain. The most commonly used method today is the administration of ^{13}C-glucose or ^{13}C-acetate, which leads to the labeling of glutamate, glutamine, and GABA through glycolysis, oxidation, and the release and trafficking of neurotransmitters among neurons and astrocytes. The incorporation of ^{13}C is tracked over time simultaneously in multiple metabolites, and the more rapid the metabolic processes in the relevant pathways, the faster will be the appearance of the ^{13}C in the MR spectra. The glucose generally provides more information about neurons, while acetate yields greater sensitivity to glial metabolism.

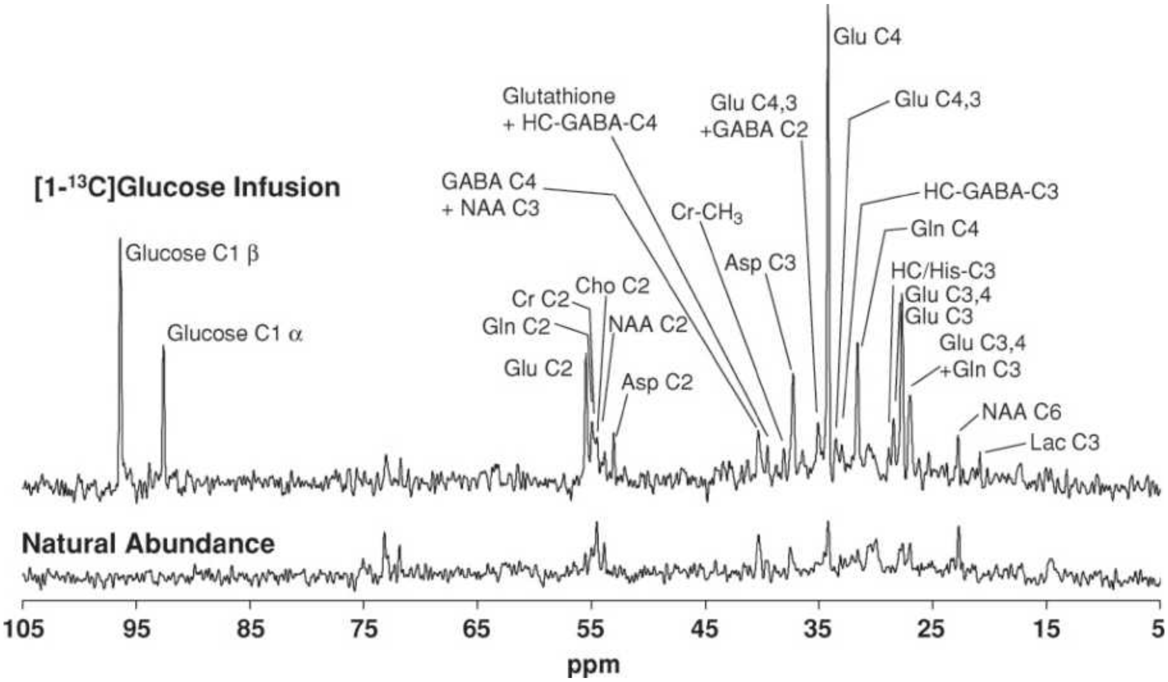

FIGURE 1.25–19. ^{13}C MR spectra from a human brain before (**bottom**) and during (**top**) a 2-hour infusion of [1–^{13}C]glucose. Before the infusion, only the 1.1% natural abundance signals can be seen in the spectrum. When exogenous ^{13}C is provided in glucose, through metabolism, the isotope appears in numerous products in the brain and can be detected with ^{13}C MRS. On the left can be seen the glucose itself. The largest single resonance is that of the C4 of glutamate (Glu C4). C3 and C2 of glutamate are also present, as are C4, C3, and C2 of glutamine (Gln). GABA is labeled at the C2, C3, and C4 positions, and aspartate (Asp) is labeled at C2. Lactate (Lac) is labeled at the C3 position. NAA and creatine (Cr) appear similar in the natural abundance and labeled spectra because their synthesis is too slow for significant labeling to occur in 2 hours.

Different methods have been developed to detect ^{13}C, with the appropriate choices based on the applications. The simplest approach is *direct detection,* which requires excitation and detection only of the ^{13}C nuclei. The sensitivity can be increased up to threefold by preapplying radio frequency at the ^{1}H frequency, in what is called the nuclear Overhauser effect (NOE). *Polarization transfer,* a more complex approach, can increase the sensitivity by up to a factor of four by transferring some of the larger polarization of coupled ^{1}H nuclei to the ^{13}C nuclei through covalent bonds, and for the acquisition, the ^{13}C nuclei are observed. *Proton-observed/carbon-edited* (POCE) spectroscopy provides a greater improvement in sensitivity. POCE detects ^{1}H that are bonded to ^{13}C, and because the ^{1}H nuclei are observed, the detection has a much greater sensitivity of ^{1}H MRS. The greater sensitivity means that volumes of a few milliliters can be studied instead of 75 mL or more for the methods that acquire data in the ^{13}C domain. The primary disadvantage is that the detection in the ^{1}H domain eliminates the benefits of the high resolution available from the ^{13}C chemical shift dispersion. *Heteronuclear single quantum coherence spectroscopy* (HSQC) provides a way to achieve the resolution available in the ^{13}C spectrum with the sensitivity of the ^{1}H acquisition by yielding data that simultaneously contain the information that would be seen by acquiring both nuclei, and this holds promise for achieving higher spatial resolution.

APPLICATIONS OF MRS IN PSYCHIATRY

As MRI provides information about brain structure and fMRI provides information about regional brain function, MRS provides information related to brain chemistry and metabolism. In many ways, it is the youngest and least developed of the MR methodologies, and until more recently, its application to psychiatry has also been limited by its relatively poor spatial and temporal resolution and need for technical expertise. However, the information provided by MRS studies in vivo over the past decade has proven valuable in identifying neurochemical and metabolic abnormalities associated with several psychiatric disorders that have helped to reshape the current thinking related to the pathophysiologic models of these disorders. With the advent of improved MRS imaging techniques and dissemination of expertise, the modality is gaining greater application to psychiatric research that is allowing the field to examine novel hypotheses related to pathophysiology and pathogenesis of psychiatric disorders.

MRS in Dementia

^{1}H MRS presents the opportunity to obtain noninvasive measures of several neurochemicals related to neurotransmission, energy metabolism, and cellular function. Studies using ^{1}H MRS have shown a trend for a general reduction in NAA measures with increasing age in medial temporal and frontal cortical brain regions, but with greater decreases in dementia and NAA levels related to plaque density. NAA is perhaps the most reliably measured metabolite in the brain due to it being present between 7 and 13 mmol/kg and the physical characteristics of the molecule (presence of identical three hydrogen atoms on a methyl group that provide triple the sensitivity of what would be provided by a single proton). As mentioned above, the significance of brain NAA content is not completely understood. It is commonly believed to serve as a marker of neuronal health, but it is also known to reflect aspects of mitochondrial energy metabolism and myelin maintenance. The studies on MCI and AD are in general agreement, reporting patients with these disorders to have decreased levels of NAA and increased levels of myoinositol (a form of inositol normally found in the brain that contributes to osmotic regulation) compared to those of age-matched comparison subjects. One interpretation is that decreased NAA and increased myoinositol are related to the loss of neurons and an increase in gliosis. However, the loss of NAA may reflect decreases in neural function. This hypothesis is supported by the fact that the changes in NAA can be reversed in some conditions. Several groups have also reported a correlation between the reductions in NAA content and clinical neuropsychological scores. Interestingly, this correlation has also recently been reported in healthy volunteer subjects.

Overall, studies of other metabolites have been less convincing in relationship to AD and other dementias, but a few studies suggest decreased levels of glutamate and elevated levels of glutamine can be found in subjects with MCI impairment and AD. Together these findings are believed to reflect a combination of ongoing metabolic dysfunction and increased gliosis. At present, there is no clinical use of MRS in the diagnosis or treatment of dementia.

MRS in Schizophrenia

^{1}H MRS has been applied widely in studies of cortical chemistry in schizophrenia. These studies documented reductions in NAA levels in many cortical and limbic brain regions in schizophrenic individuals and smaller reductions in family members of people diagnosed with schizophrenia. NAA level reductions have been described in some brain regions in medication-naïve patients, but there also do appear to be progressive reductions with advancing illness and continued antipsychotic treatment. Because NAA is localized to neurons, reductions in NAA levels may reflect the postmortem findings of reduced neuronal size and reduced dendritic arborization as well in vivo evidence from DTI studies of disturbances in the integrity of long fiber pathways. However, NAA is synthesized by a mitochondrial enzyme. As a result, reduced NAA levels may reflect metabolic disturbances that may be related to mitochondrial dysfunction suggested in postmortem studies or regional metabolic alterations associated with the pathophysiology of schizophrenia or the impact of treatment. This possibility is consistent with the emerging differences between the impact of typical and atypical antipsychotic treatment on NAA levels in patients. Other metabolites have been measured in ^{1}H MRS studies of schizophrenic patients. The most interesting findings to date may be the description of normal or low levels of glutamate and GABA and increased levels of glutamine in medication-free patients with schizophrenia, with evidence accumulating that medication may be responsible for some of the abnormal metabolite levels reported in schizophrenia. Nonetheless, metabolite alterations are observed in high-risk patients prior to the onset of illness and in the early course of schizophrenia. Brain glutamine is synthesized primarily by glia from glutamate which emerges as a by-product of energy metabolism, as well as from glutamate that is taken up by glia following its release by neurons and glia in the context of excitatory neurotransmission. Thus, it is possible that the alterations of glutamine and glutamate in individuals diagnosed with schizophrenia could reflect changes in glutamatergic neurotransmission.

A number of other metabolites have been examined using ^{1}H MRS and ^{31}P MRS in patients with schizophrenia, including glutathione, choline, PME, and PDE. These metabolites are of potential interest because studies of these metabolites suggest that schizophrenia is associated with inflammation and abnormalities in membrane integrity and might be reflective of disturbances in the processing of myelin suggested by postmortem and clinical molecular genetic studies. However, the relationships of these metabolites to specific membrane disturbances associated with schizophrenia remain unclear. Also, in the ^{31}P MRS studies, inconsistencies in the findings made by various groups have called these findings into question, generally. However, they may be consistent with recent findings using magnetization transfer that oxidative phosphorylation is slower in schizophrenia, using a volume of tissue that included a large fraction of white matter indirectly suggest that myelin may be compromised in patients with schizophrenia.

MRS in Mood Disorders

There has been a rapid increase in the use of MRS in the study of mood and anxiety disorders over the past decade. While the methodology does not yet enter into the clinical practice related to the diagnosis and treatment of mood disorders, it has in part led to an evolution in the concepts of the pathophysiology of the disorders. The majority of studies have used either ^{1}H MRS or ^{31}P MRS to provide information about the abnormalities in neurochemistry and energy metabolism associated with the illnesses. A few other studies have used ^{7}Li and ^{19}F MRS to study the pharmacokinetics and brain concentrations of drugs used in the treatment of mood and anxiety disorders. Recent new work suggests that the use of ^{13}C MRS and ^{31}P will be useful in further defining the pathophysiology of the disorders.

To date, ^{1}H MRS has been the most widely used MRS methodology in the study of mood disorders. While the technology continues to evolve, several of the measures are becoming more routine and reliable, thus allowing for investigators to more confidently draw conclusions from data obtained at different sites using different protocols and spectrometers.

There have been over 200 published studies examining NAA levels in relation to mood and anxiety disorders. However, due to the heterogeneity of the illnesses studied and the variety of regions examined, it remains difficult to draw firm conclusions regarding NAA’s contributions to the neurobiology of mood disorders. In general, there have not been consistent findings showing NAA changes in association with major depression. Yet, there have been specific studies suggesting that NAA may be reduced in the hippocampus of depressed and anxious patients. There is also some evidence suggesting that NAA levels may be reduced in the frontal lobe of bipolar patients, potentially serving as a diagnostic marker. The creatine peak is believed to remain relatively constant in most disorders, leading to the use of creatine as an internal standard in many studies. Studies examining creatine in relation to mood disorders appear to be in general agreement with this view, but exceptions do exist. Choline is an essential precursor to membrane lipids and the neurotransmitter acetylcholine. A series of studies seem to demonstrate elevated choline levels in the basal ganglia of mood disorder subjects compared to those of healthy comparisons. These studies have been suggested as evidence of altered membrane turnover and impaired signal transduction mechanisms within the basal ganglia of patients.

As the major inhibitory and excitatory neurotransmitters in the brain, it is not surprising that the amino acid neurotransmitters GABA and glutamate have been linked to the pathophysiology of several neuropsychiatric disorders, thus making them extremely interesting targets for MRS investigation. Due to overlapping resonances, these compounds are more difficult to measure using standard ^{1}H MRS methods. This has delayed the use of MRS to study the amino acid neurotransmitters and led many to use the term Glx to refer to the combined measure of the GABA, glutamate, and glutamine. With the development of several new editing techniques, several groups are now capable of isolating the individual peaks. The relatively recent application of MRS to investigate amino acid neurotransmitter systems has proven highly productive in the area of mood disorders research, and there is mounting evidence that suggests markedly abnormal concentrations of GABA and glutamate in several brain regions of mood and anxiety disorder patients. Furthermore, recent studies suggest that antidepressant treatments may be capable of reversing some of these abnormalities. These data helped to increase the awareness of the potential role of the amino acid neurotransmitters in the neurobiology of mood disorders and the mechanism of antidepressant action. Since it is now recognized that glial cell function is critical to amino acid neurotransmitter metabolism, these data, in conjunction with several postmortem findings, have also provided support for a pathophysiologic model of glial cell impairment in relation to MDD. Recent observations of altered brain glutathione levels in mood disorders may also be related to alterations in glial/neuronal interactions.

^{31}P MRS provides in vivo measures of brain membrane phospholipid metabolism, high-energy phosphate metabolism, and intracellular pH. Initial ^{31}P MRS studies in bipolar disorder set out to test the hypothesis that the phosphoinositide pathway is enhanced in bipolar disorder and to explore the contributions of membrane phospholipids and membrane defects in bipolar disorder by measuring levels of PME and PDE. A meta-analysis of these studies found significant diagnosis- and mood state–dependent abnormalities in PME content in patients with bipolar disorder, which suggests dysregulation of brain-signal transduction systems and membrane metabolism may have relevance in bipolar disorder. Further, they suggest that lithium treatment increases PME levels, providing a possible clue into the mechanism of action. Beyond concentration levels, in patients with first-episode bipolar disorder, kinetic ^{31}P MRS has also shown slower transfer of phosphate between ATP and phosphocreatine, an aspect that may have profound consequences, given that phosphocreatine functions as a high-energy reserve to maintain levels of ATP.

Other studies have attempted to evaluate whether neuroenergetic defects are related to mood disorders by measuring the concentrations of nucleotide phosphates (the three phosphate residues from NTP) and phosphocreatine. In general, these studies have found lower levels of nucleoside phosphates in the basal ganglia of subjects in depressive episodes, suggesting that high-energy phosphate metabolism is altered in the basal ganglia of subjects with depression and drawing attention to the possible role that the mitochondria may play in mood disorders.

Reduced pH_i has been observed in pathologic states arising from ischemic insult to the brain. Abnormalities in pH_i concentration have also been reported in bipolar subjects. Several studies have found significantly reduced pH_i in both the basal ganglia and whole brain of bipolar subjects in the euthymic state, while other studies have found increased pH_i levels to be associated with the bipolar and depressed states. While the specific pathophysiology underlying these effects remains unknown, the findings have helped to focus attention on the potential contributions of neuroenergetics to mood disorders.

As described earlier in this chapter, the unique physical properties of ^{13}C make it an excellent tool to measure rates of glucose metabolism as well as amino acid neurotransmitter cycling and being a stable isotope, it does not present the risks and challenges associated with radiochemical research. Studies using glucose labeled with ^{13}C have been used to measure the rates of glucose oxidation, glutamate/glutamine cycling, and GABA synthesis. Studies using ^{13}C-labeled acetate have recently yielded measures of glial cell metabolism, a potentially valuable tool considering the recent findings suggesting that glial cell pathology may play a prominent role in the neurobiology of mood disorders. Although the application of ^{13}C MRS in vivo is relatively new to psychiatry, recent data show significantly lower rates of oxidative metabolism in unmedicated patients with major depression. A conceptually new direction is the use of ^{13}C MRS kinetic measures to measure synaptic strength as a test of long-term potentiation/long-term depression using Vtca/Vcycle. Future studies using this modality are likely to provide new insights into the association relationships between mood disorders, neuroenergetics, amino acid neurotransmitter functioning, and glial cell function.

MRS in Alcohol Dependence

^{1}H MRS studies evaluating NAA and choline have provided neurochemical evidence that complements the MRI findings related to the emergence and recovery from alcohol-related neurotoxicity. ^{1}H MRS studies of GABA have provided insights into alterations in cortical inhibitory neurotransmissions associated with the recovery from alcohol dependence. During acute withdrawal, cortical GABA levels appear to be normal, consistent with animal studies. With recovery from alcohol dependence, cortical GABA levels appear to decline and may be significantly below the level seen in healthy subjects with extended sobriety. These time-dependent changes in cortical GABA levels during sobriety have been hypothesized to reflect a recovery from the adaptations in $GABA_A$ receptor populations produced by chronic ethanol exposure. ^{13}C measures may reveal factors that promote energetic factors, which promote heavy drinking and relapse, such as the enhanced ability of the brain to oxidize alcohol and its metabolic product, acetate. To the extent that the brain becomes dependent on acetate as a fuel, withdrawal from alcohol may be compounded by the near-absence of circulating acetate once patients stop drinking. Further, when alcohol is oxidized, it creates acetaldehyde, which has been shown to have rewarding properties inside the brain, even though systemically acetaldehyde is aversive.

^{7}Li and ^{19}F MRS

A few studies have attempted to use ^{7}Li MRS to study the pharmacokinetics of lithium as it relates to the treatment of mood disorders. These studies have demonstrated that lithium relatively rapidly clears the BBB and has approximately a 28-hour half-life in the brain. A potentially powerful clinical application of the methodology is the ability to measure brain concentrations of Li relative to serum levels and to assess intracellular versus extracellular lithium. While limited in number, these studies seem to suggest only a modest correlation between serum lithium levels that are commonly used to titrate treatment dosage and brain lithium concentration. Similarly, studies have employed ^{19}F MRS to determine the pharmacokinetics and the relationship between serum and brain levels of fluorine-containing drugs such as fluoxetine and fluvoxamine. While such studies are of limited clinical usefulness at present, they may be useful in determining dose-response relationships for other compounds in the future.

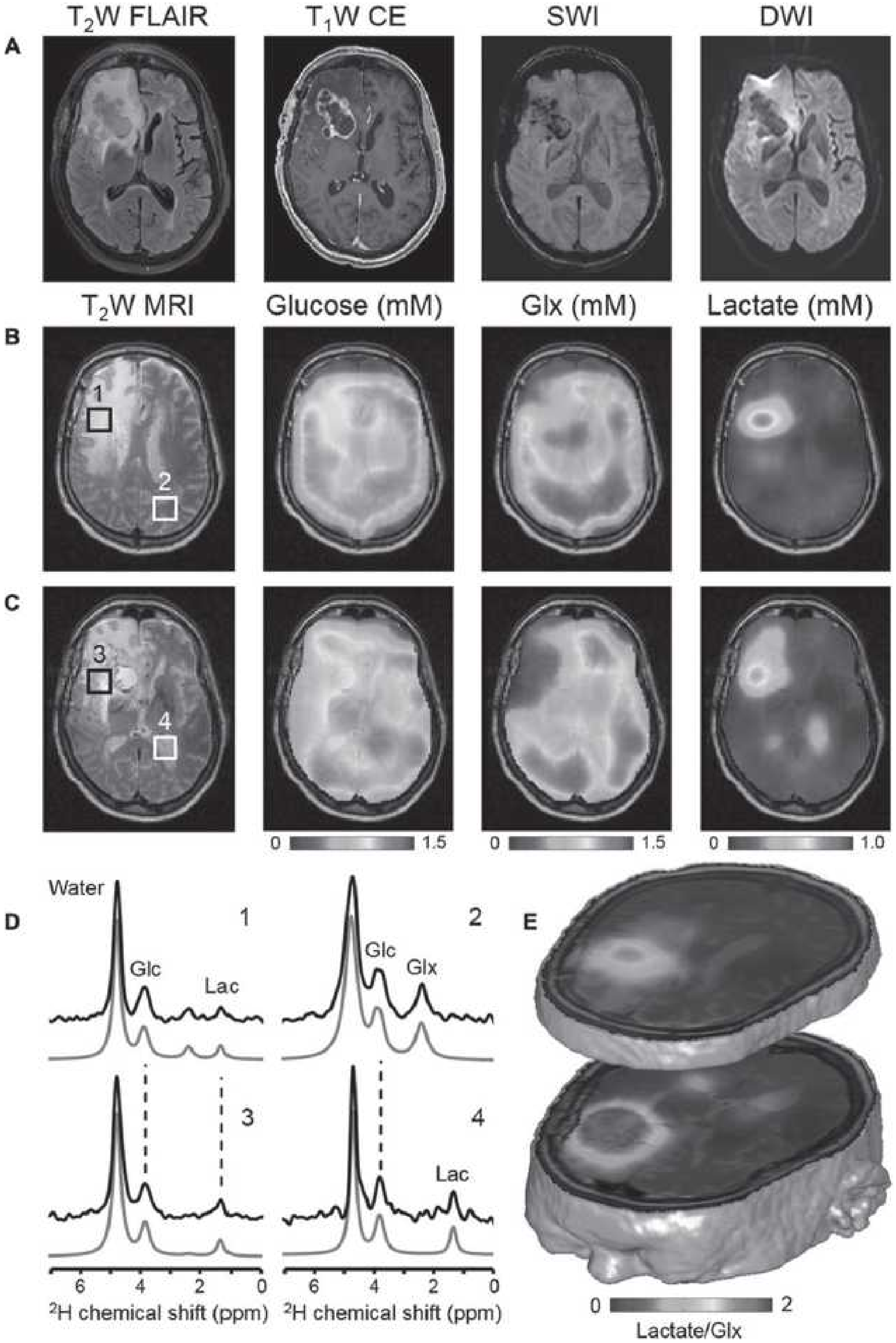
T2W FLAIR
T1W CE
SWI
DWI
A
T2W MRI
Glucose (mM)
Glx (mM)
Lactate (mM)
B
1
2
C
3
4
0
1.5
0
1.5
0
1.0
D
Water
Glc
Lac
1
Glc
Glx
2
3
4
Lac
6
4
2
0
²H chemical shift (ppm)
6
4
2
0
²H chemical shift (ppm)
E
0
2
Lactate/Glx

FIGURE 1.25–20. Deuterium metabolic imaging (DMI) shows the aberrant glucose metabolism known as the Warburg effect in a patient with an aggressive brain tumor. DMI detects deuterium-labeled substrates and their

downstream metabolites in vivo using a standard MRI scanner. In the case of glucose, the administration is as simple as drinking the water-dissolved ^{2}H-glucose. The method can be applied in different organs, such as the brain and the liver, and using different deuterium-labeled substrates to measure different metabolic features. This simple and robust MRI method can be easily integrated into existing scan protocols and can become a widespread metabolic imaging modality available in the clinic. (Reprinted with permission of AAAS from De Feyter HM, Behar KL, Corbin ZA, et al. Deuterium metabolic imaging (DMI) for MRI-based 3D mapping of metabolism in vivo. Sci Adv 2018;4(8). doi: 10.1126/sciadv.aat7314. © The Authors, some rights reserved; exclusive licensee AAAS. Distributed under a CC BY-NC 4.0 License (http://creativecommons.org/licenses/by-nc/4.0/).)

DEUTERIUM METABOLIC IMAGING (DMI)

A recent development that is likely to find practical, widespread use is the MRS of the ^{2}H isotope of hydrogen, commonly known as deuterium, to measure metabolism. Although the isotope has much lower sensitivity than ^{1}H, it has several advantageous qualities. ^{2}H has short T_1, which allows rapid repetitions to accumulate sufficient signal to generate 3D images that currently have resolution similar to that seen with PET. ^{2}H also has a very low natural abundance, a small fraction of 1%, so there is essentially no background signal if one wants to administer ^{2}H as a metabolic tracer. ^{2}H has a very short T_2, which means that distortions of the magnetic field homogeneity in regions of importance to psychiatry (e.g., the orbitofrontal cortex) are negligible by comparison, as well as robustness due to the low resonance frequency, and so metabolism can be accessed throughout the brain (Fig. 1.25–20), without the regional limitations that affect MRS of ^{1}H, ^{13}C, ^{31}P, and some other prominently used isotopes. While ^{2}H in large quantities can cause nausea, DMI uses doses that are far lower, with subjects experiencing no ill effects. Deuterium ultimately clears from the body in the form of water, and measurements can be repeated safely. For some applications, the deuterium can be administered orally in the form of deuterated water or as a glucose solution. A very important aspect of the method is the simplicity: the electronics and software to acquire the data are among the simplest possible for NMR.

PRACTICAL CONSIDERATIONS

Although the technology invokes some complexities, there are practical considerations of safety and comfort that are of great importance when considering a possible MRI or MRS scan for a patient.

Safety

The primary consideration is the safety of the patient. MRI and MRS are considered to be among the safest ways to examine the human body. Because MRI and MRS use magnetism and radio waves, and not X-rays, they are not known to be mutagenic or associated with cancer risk–like methods that use ionizing radiation. Radio waves can cause heating of tissue, including the brain, so the U.S. FDA has set guidelines for magnet strength and exposure to radio waves, and the manufacturers and operators of MR equipment carefully observe those guidelines. For exposure to the radio frequencies, powers, and durations that are used in human MR scans, no harmful effects have been seen.

MRI and MRS pose some risks for certain people. People who have a pacemaker or certain metal objects inside their bodies are generally not candidates for MR scans because the strong magnets in the MR scanner might cause the implants to malfunction or move. Aneurysm clips pose a particularly major risk. As current technologies increase magnetic field strength, one must take care with metallic products that are designated to be MR-safe. As far as can be achieved, companies attempt to maintain the necessary qualities of their products while removing as much of the ferromagnetism from the material as possible. However, trace levels of ferromagnetism may remain, posing a greater risk as magnetic field strengths increase. The claims of MR compatibility are generally true for the magnetic field strength of 1.5 T or lower, and in some cases, the devices are safe at 3 T. However, many devices have not been tested above 1.5 T, and fewer still above 3 T, so for now, many metallic implants pose an unknown and unacceptable risk above 3 T and should be checked individually for compatibility at a given magnetic field strength.

Another risk is that of a metallic object flying through the air toward the magnet and hitting the patient. To reduce this risk, MR facilities usually require all personnel, including the patient, to remove all metal from their clothing and all metal objects from their pockets. In many cases, patients will be required to wear hospital garb. Nothing metal can be brought into a magnet room except for equipment or supplies that have been specially screened by the MR operators. The doors to the magnet room should be kept closed except for brief entry and exit by the operators and patient or by personnel directly approved by the operators. Failure to safeguard against this risk had resulted in catastrophes, such as the death of a child when a magnetic oxygen cylinder crushed his skull at speeds in excess of 50 miles per hour. A less dramatic but more common risk is that of injury to a patient's eye from a flying ballpoint pen or a paperclip.

A common question is about the safety of pregnancy. At magnetic field strengths of 1.5 T, in use since the early 1980s, and 3 T, in use since the early 1990s, no adverse effects of the magnetic field have been associated with pregnancy or reproductive health. Above magnet field strengths of 3 T, far fewer data exist, although no ill effects have been reported at this time. At 7 T and higher, some individuals experience nausea or dizziness when entering or leaving the magnetic field due to magnetic field effects on the inner ear while moving through the region at which the magnetic field changes near the bore of the magnet.

The use of injectable contrast agents poses an additional but small source of risk. There are risks associated with the placement of an intravenous line, including bruising and infection, although the lines should be placed with care under sterile conditions that minimize those risks. The FDA has approved contrast agents that contain the metal gadolinium for use in human subjects and does not recognize any major risks with its use. Less than 3% of patients may experience mild nausea, headache, or dizziness after the injection, and those side effects usually resolve themselves without treatment. Less than 1% of patients experience an anaphylactic reaction, including hives, itching, or difficulty breathing. Subjects should be asked if they have a history of allergic reactions to MR or other injected contrast agents or if they have a history of kidney disease, asthma, allergic respiratory disorders, anemia or other diseases that affect red blood cells. Women who are pregnant are generally not given gadolinium contrast agents in the United States, although in Europe, it has been judged to be safe and is permitted. Women who are breastfeeding are instructed to express milk for 48 hours after the scan before returning to breastfeeding.

A magnet quench was described earlier in this chapter as the sudden de-energization of the magnet, accompanied by a rapid boil-off of liquid helium and possibly nitrogen. Nitrogen and helium are inert and nontoxic, but in confined spaces, they displace oxygen and can cause suffocation. Therefore, MR scanners are equipped with ventilation systems to remove those gases in the event of a quench. Furthermore, on the very rare occasion that a quench occurs, the operators are trained to remove a patient immediately from the magnet room. Patients are unlikely to be present at the times that a quench is most likely to occur, these being during magnet energization that occurs at installation or after major equipment upgrades, or when helium is being added to the magnet to replace material that has slowly been lost from the system. Quenches may also occur if a large metal object strikes the magnet. To date, no patients have been injured or killed by quenches.

What to Tell the Patient before a Scan

A patient should be told that the study requires lying on a bed, usually with a detector that has some rods that are placed around the head. The bed is rolled into the scanner, which typically looks like a tube so that the individual's head is approximately centered along the length of the tube, and the feet, legs, and sometimes even the waist lie outside the tube. Patients should be told that the machine does not move, although they will hear knocking and other sounds and possibly feel some vibrations while the scanner is acquiring the data. It is crucial to stress the need to hold still. Movement can lead to images that are blurry or otherwise uninterpretable. If there is an injection of a contrast agent, the individual may feel some discomfort associated with the intravenous line, just as any intravenous catheter may cause some bruising or other discomforts. Otherwise, MR has no painful parts.

Some people feel uncomfortable or anxious because of having to lie still in a tube that is generally less than three feet in diameter and several feet long. If a subject is feeling anxious, then a small dose of an anxiolytic drug such as a benzodiazepine is sometimes administered, as long as it does not adversely affect the measurements. A subject should be told that the operators will monitor them constantly by sight and sound, so a patient can ask the operator to be brought out of the scanner. It is important to warn a patient that some people feel dizzy, develop some stomach upset, experience a metallic taste, or find tingling sensations or muscle twitches. These sensations usually go away quickly, but it is important to tell the research staff if they occur.

The patient should be presented with an MR safety questionnaire to be answered very carefully, and it is important to verify the answers verbally with each patient. If parents accompany a child into the MRI room, the parents should be screened likewise. The questionnaire will ask about the surgical history and the presence of a variety of potentially MR-incompatible items, including pacemakers, intrauterine contraceptive devices, aneurysm clips, and injuries from shrapnel and metal work, especially to the eye. Patients commonly ask questions about implanted metal, including dental work and screws used to repair broken bones. Fillings are safe. Most permanently installed dental work is safe. Partial plates that can be removed should be removed before the scan, partly for safety but primarily because ferromagnetic material in some of them can cause T_2* effects that compromise the quality of the MRI or MRS. Bone screws and other surgically implanted materials should be examined in conjunction with the MR facility's safety officer. It is also important to know the manufacturer and model of any implanted devices, even if a patient insists that the device is MRI-safe. Some devices may be safe at 1.5 T, but not 3 T, and others safe at 3 T may pose risks at magnetic fields above 3 T. An X-ray or review of a patient's surgical records may be necessary to evaluate the presence and MRI compatibility of internal materials. The patient should be encouraged more than once to read the safety form carefully and ask questions about any uncertainties, even if the patient thinks that it is probably safe. It is best for a patient to ask about something that poses no risk than to keep quiet out of embarrassment and die.

Jewelry is a major source of questions and sometimes of resistance. Although patients often refuse, they should not enter the magnet room without first removing piercing jewelry, such as tongue rings, nose rings, belly button rings, and earrings. They may claim that a favorite necklace is pure gold, which is nonmagnetic, only to have the MR machine demonstrate the opposite. In some cases, an inadequately anchored piece, such as the backing of an earring, can come loose and pose a flying hazard or become lodged in a joint somewhere in magnet cowell or gradient coil assembly. The more common problem is that the small quantities of ferromagnetic materials distort or, in some cases, generate huge holes to appear in the MRI, even when the patient clearly has an intact head. Safety is enhanced if a blanket ban on jewelry is followed.

A significant number of subjects in the general population are likely to show abnormalities. When these are unexpected, they are called *incidentalomas,* and the proper course of action upon finding an incidentaloma has been and continues to be a subject of debate.

Observations that are of potential clinical significance should be referred to a radiologist for evaluation and decisions about further courses of action. For example, aneurysms pose a cause for further evaluation. Some conditions, such as an asymptomatic meningioma in an 85-year-old woman, might be best monitored but untreated unless symptoms later emerge. Each case must be evaluated on its own circumstances and according to the policies of each institution.

FUTURE DIRECTIONS

Today, established MR technologies provide a limited but valuable clinical service to psychiatry through their ability to rule out gross structural abnormalities. Presently, most of the methodologies described in this chapter are used to acquire data in academic research settings in efforts to help psychiatrists understand the pathophysiology, pathogenesis, and treatment of psychiatric disorders.

The primary limitations that face MR today are spatial resolution, sensitivity, and cost. Regarding spatial resolution, MRI data in humans are rarely acquired with grid sizes finer than 0.5 to 0.75 μL and, more typically, 1.5 to 2 μL. The sensitivity of MRS is typically far less than that of MRI because the concentrations of the metabolites are far less than those of the water and fat that are usually detected for MRI, with a general limitation of 0.5 to 1 mmol/kg as the detection limit in the human brain in vivo, with volumes in some cases as small as 0.25 mL but more typically 8 mL for ^{1}H MRS and larger volumes for ^{31}P and ^{13}C. Because of the need to detect lower concentrations, MRSI fares more poorly than MRI with regard to resolution and/or sensitivity, but the ability to map chemicals through the brain provides a unique power to investigate mental illnesses. Although whole-brain ^{1}H MRS can be implemented, much of the brain is poorly shimmed and yields no useful data. Methods that greatly improve problematic shimming are also under development and may resolve this problem in the coming years. The DMI method discussed in this chapter is far less affected by such problems and holds great promise for mapping metabolism over the entire brain, similar to FDG-PET emission tomography but without radioactivity. fMRI receives great attention in the psychiatric research community today, and a major limitation that must be kept in mind when interpreting the imaging results is that the measurement is relative, and any function-related changes must be evaluated in the context of possible baseline conditions. Cost poses an impediment to increases in diagnostic approaches with new MR methodologies. With scans now costing $500 to $4,000, depending on the procedures ordered and the institution where the scan is done, additional methods will only gain widespread acceptance if they are clearly shown to add clinically significant information.

The future of MR can be viewed from several perspectives. The noninvasive, nonradioactive nature of MR provides a strong impetus for continued development in the diagnostic evaluation of patients and use in basic science research. By virtue of its safety profile, MR is likely to support research on childhood manifestations of psychiatric disorders. Magnetic field strengths for research scanners are increasing, having reached 11.7 T for human use; higher magnetic field strengths for most brain studies represent greater sensitivity. As manufacturers gain experience with higher-field research machines, the technology has in the past lowered the cost of medium-field magnets and led to improvements in reliability and stability. A second potential advance lies with the elimination of helium as the coolant for the superconducting wire in the magnet. Helium currently poses a significant cost of installation and operation of MR scanners. The cost is rising with increasing limitations on worldwide helium supplies, and some industry sources predict exhaustion of the world supply by the year 2030. Researchers around the world are working to develop superconductors with the necessary physical flexibility and ability to carry large electrical currents at liquid nitrogen temperatures and also more powerful permanent magnets (not produced by electric currents); such breakthroughs should lower MR operating costs and eliminate the need for the limited helium supply. Another area of intense investigation is hyperpolarization of ^{13}C, increasing its sensitivity by several orders of magnitude. Researchers are actively pursuing the use of hyperpolarized ^{13}C for use as nontoxic contrast agents for angiography and metabolism. A very different approach, *molecular MRI,* increases the ability to monitor processes such as the migration of nerve cells, gene expression, and targeting of particular cell types. Such molecular imaging abilities remain in the realm of research on animals, but as methods continue to develop, they appear likely to find their way to applications to human diseases. Another major increasing area is multimodal imaging: linking MRI, fMRI, MRS, or other measures, as well as binding those measures with PET. For example, there are now efforts to link metabolite level changes to alterations in synaptic density using the PET tracer SV2A.

In the future, MR is likely to find increasing utility as a research tool because it provides a quantitative window on neuroanatomy, function, and neurochemistry. Developers are also working to turn the research methodologies into robust clinical techniques for the testing and use of biomarkers to monitor treatment efficacies. MR studies are becoming increasingly multimodal, with various types of MRI and MRS acquired in single scan sessions and on individual subjects, as well as in complementary measurements with methods like PET and EEG recordings. As outlined above, MR has already begun to change the current thinking with respect to the pathophysiology of psychiatric disorders, neuroenergetics, amino acid neurotransmission, and plasticity. From most perspectives, MR appears likely to continue to do so in the future.

FURTHER READINGS

Alloza C, Blesa-Cábez M, Bastin ME, et al. Psychotic-like experiences, polygenic risk scores for schizophrenia, and structural properties of the salience, default mode, and central-executive networks in healthy participants from UK Biobank. *Transl Psychiatry*. 2020;10:122.

Attwell D, Laughlin SB. An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab*. 2001;21:1133–1145.

Averill LA, Jiang L, Purohit P, Coppoli A, Averill CL, Roscoe J, Kelmendi B, De Feyter HM, de Graaf RA, Gueorguieva R, Sanacora G, Krystal JH, Rothman DL, Mason GF, Abdallah CG. Prefrontal Glutamate Neurotransmission in PTSD: A Novel Approach to Estimate Synaptic Strength in Vivo in Humans. Chronic Stress (Thousand Oaks). 2022 Apr 11;6:24705470221092734. doi:10.1177/24705470221092734. PMID: 35434443; PMCID: PMC9008809.

Carlson PJ, Singh JB, Zarate CA Jr, Drevets WC, Manji HK. Neural circuitry and neuroplasticity in mood disorders: insights for novel therapeutic targets. *NeuroRx*. 2006;3:22.

De Feyter HM, Behar KL, Corbin ZA, et al. Deuterium metabolic imaging (DMI) for MRI-based 3D mapping of metabolism in vivo. *Sci Adv*. 2018;4:eaat7314.

Drevets WC, Price JL, Simpson JR, et al. Subgenual prefrontal cortex abnormalities in mood disorder. *Nature*. 1997;386:824–827.

Du F, Cooper AJ, Thida T, et al. In vivo evidence for cerebral bioenergetic abnormalities in schizophrenia measured using ^{31}P magnetization transfer spectroscopy. *JAMA Psychiatry*. 2014;71:19–27.

Du F, Yuksel C, Chouinard VA, et al. Abnormalities in high-energy phosphate metabolism in first-episode bipolar disorder measured using ^{31}P-magnetic resonance spectroscopy. *Biol Psychiatry*. 2018;84:797–802.

Dukart J, Mueller K, Barthel H, et al. Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. *Psychiatry Res*. 2013;212:230–236.

Feldstein Ewing SW, Sakhardande A, Blakemore S-J. The effect of alcohol consumption on the adolescent brain: a systematic review of MRI and fMRI studies of alcohol-using youth. *Neuroimage: Clinical*. 2014;5:420–437.

Goghari VM, Rehm K, Carter CS, MacDonald AW III. Regionally specific cortical thinning and gray matter abnormalities in the healthy relatives of schizophrenia patients. *Cereb Cortex*. 2007;17:415–424.

Gore JC. Artificial intelligence in medical imaging. *Magn Reson Imag*. 2020;68:A1–A4.

Hanson LG. Is Quantum Mechanics Necessary for Understanding Magnetic Resonance? *Concepts Magn Reson*. 2008;32A:329–340.

Kempton MJ, Salvador Z, Munafò MR, et al. Structural neuroimaging studies in major depressive disorder. Meta-analysis and comparison with bipolar disorder. *Arch Gen Psychiatry*. 2011;68:675–690.

Kose K. Physical and technical aspects of human magnetic resonance imaging: present status and 50 years historical review. *ADV PHYS-X*. 2021;6:1885310.

Kroenke CD, Rohlfing T, Park B, Sullivan EV, Pfefferbaum A, Grant KA. Monkeys that voluntarily and chronically drink alcohol damage their brains: a longitudinal MRI study. *Neuropsychopharmacology*. 2014;39:823–830.

Kumar V, Vajawat B, Rao NP. Frontal GABA in schizophrenia: a meta-analysis of ^{1}H-MRS studies. *World J Biol Psychiatry*. 2021;22:1–13.

Lv J, Di Biase M, Cash RF, et al. Individual deviations from normative models of brain structure in a large cross-sectional schizophrenia cohort. *Mol Psychiatry*. 2021;26:3512–3523.

Mahon PB, Lee DS, Trinh H, et al. Morphometry of the amygdala in schizophrenia and psychotic bipolar disorder. *Schizophr Res*. 2015;164(1–3):199–202.

Merritt K, McGuire PK, Egerton A, et al. Association of age, antipsychotic medication, and symptom severity in schizophrenia with proton magnetic resonance spectroscopy brain glutamate level: a Mega-analysis of individual participant-Level data. *JAMA Psychiatry*. 2021;78:667–681.

Minati L, Grisoli M, Bruzzone MG. MR spectroscopy, functional MRI, and diffusion-tensor imaging in the aging brain: a conceptual review. *J Geriatr Psychiatry Neurol*. 2007;20:3–21.

Modenato C, Kumar K, Moreau C, et al. Effects of eight neuropsychiatric copy number variants on human brain structure. *Transl Psychiatry*. 2021;11:399–408.

Ogawa S, Menon RS, Tank DW, et al. Functional brain mapping by blood oxygenation level-dependent contrast magnetic-resonance imaging—A comparison of signal characteristics with a biophysical model. *Biophys J*. 1993;64:803–812.

Onwordi EC, Whitehurst T, Mansur A, et al. The relationship between synaptic density marker SV2A, glutamate and N-acetyl aspartate levels in healthy volunteers and schizophrenia: a multimodal PET and magnetic resonance spectroscopy brain imaging study. *Transl Psychiatry*. 2021;11:393.

Redlich R, Almeida JJR, Grotegerd D, et al. Brain morphometric biomarkers distinguishing unipolar and bipolar depression. A voxel-based morphometry-pattern classification approach. *JAMA Psychiatry*. 2014;71:1222–1230.

Smucny J, Carter CS, Maddock RJ. Medial prefrontal cortex glutamate is reduced in schizophrenia and moderated by measurement quality: a meta-analysis of 1H-MRS studies. *Biol Psychiatry*. 2021;90(9):643–651.

Wang L, Hermens DF, Hickie IB, Lagopoulos J. A systematic review of resting-state functional-MRI studies in major depression. *J Affect Disord*. 2012;142:6–12.

Whitehurst TS, Osugo M, Townsend L, et al. Proton magnetic resonance spectroscopy of N-acetyl aspartate in chronic schizophrenia, first episode of psychosis and high-risk of psychosis: a systematic review and meta-analysis. *Neurosci Biobehav Rev*. 2020;119:255–267.

Zahr NM. Structural and microstructural imaging of the brain in alcohol use disorders. In: Sullivan EV, Pfefferbaum A, eds. *Handbook of Clinical Neurology: Alcohol and the Nervous System*. 2014;125:257–290.

Zhang B, Ferman TJ, Boeve BF, et al. MRS in mild cognitive impairment: early differentiation of dementia with Lewy bodies and Alzheimer’s disease. *J Neuroimaging*. 2015;25:269–274.

1.26 Electroencephalography in Psychiatry

JASON K. JOHANNESEN, PH.D., AND PATRICK D. SKOSNIK, PH.D.

The *electroencephalogram* (EEG), considered the first functional brain imagining technique, has contributed substantially to the study of pathophysiology across the field of psychiatry. EEG provides a noninvasive measurement of brain activity that is well suited for a range of clinical and scientific applications, with low cost, flexibility in testing environment and equipment, and millisecond-level temporal resolution (considered some key advantages). Approximating "real-time" measurement, EEG is well suited for addressing questions concerning how brain activity changes in response to different conditions, such as states of arousal transitioning from wakefulness to sleep, to basic sensory stimulation (including stages of early sensory registration and perceptual processing), and during more complex mental operations engaging attention, memory, and learning processes. Despite the continued development of high-density digital EEG recoding and associated methods of interpolating sources of underlying neural activations, the strength of EEG remains to be nested in the ascertainment of events as they occur in time rather than the location of activation to precise brain structures. This distinction is commonly discussed in terms of the tradeoff between "temporal" and "spatial" resolution, which applies to all methods of brain imaging and, thereby, should guide the selection of methods most appropriate to investigate questions concerning the timing and sequencing of evolving neural processes versus the sources from which they emanate.

A renascence of EEG methods has been promoted by the accessibility of these instruments to cognitive and clinical scientists who seek cost-efficient and noninvasive means of applying measures of brain function to new experimental, diagnostic, and computational questions. EEG methodology can be exquisitely applied to examine hypotheses on the pathogenesis of psychiatric illness, no longer conceptualized as the result of a dysfunction of select brain regions but as a consequence of the failure to integrate neural activity through networks engaged across multiple brain regions. Advances in EEG measurement of brain function have been aided by technologic advances at levels of bioamplification, electrode sensors, and computational algorithms for signal processing and quantification. EEG methodology and its interface with experimental neuroscience continue to hold an important position at the forefront of psychiatry research, contributing both to areas of illness characterization and treatment development. In time, and with necessary large-scale replication, the ultimate goal will be to translate research findings to clinical protocols that can be used to inform the early prediction of illness, differential diagnosis, sample enrichment, and treatments developed for common psychiatric disorders.

Electroencephalography, the practice of recording and interpreting EEG, remains a growing discipline within the field of medicine broadly, encompassing a range of neuroevaluative applications used in the diagnosis of seizure and epilepsy, sleep disorders, head trauma, tumors, infections, encephalopathy, anoxia, and cases where acute toxicity may be suspected. There are a number of examples from clinical research, such as the case of traumatic brain injury secondary to blast pressure waves, where EEG has demonstrated superior sensitivity to pathophysiology than structural brain scans. Diagnostic techniques that rely on EEG measurement include evoked potentials (EP) recorded in response to sensory stimulation. EPs result from the phase-synchronization of the ongoing EEG that is time locked to stimulation, producing characteristic waveforms that are measured in amplitude (voltage change) and latency, and interpreted as a representation of change in brain activity “evoked” by the external sensory event. EPs reflect the earliest stages of sensory processing and help to determine whether the transmission of sensory information is continuous or interrupted along known pathways. Tests common to neurologic evaluation include the visual-evoked response (VER) test, detecting problems with the optic nerve that affects sight, the brainstem auditory-evoked response (BAER) test used to test hearing ability (as used in infants), and the detection of brainstem tumors, and the somatosensory-evoked response (SSER) test. The SSER test makes exquisite use of the EEG’s temporal properties, aiding in the examination of nerve transmission along the spinal cord by measuring the amount of time it takes stimulation (usually electrical current) applied to an affected extremity to reach the brain.

In the practice of clinical psychiatry, EEG remains to be used primarily as a screening tool for ruling out organic factors that may contribute to the presentation of psychiatric symptoms, such as a change in perceptual experience (e.g., hallucinations), personality, and behavior. The differential diagnosis of a psychiatric disorder from organic pathology is further informed by medical history and clinical interview. At this time, there are no specific profiles of EEG activity or related diagnostic procedures that, taken alone, would be considered confirmatory of a psychiatric diagnosis. That said, a range of EEG abnormalities are described in the psychiatric literature, with considerable interest in the use of such features as *biomarker*s of underlying pathogenesis disrupting brain function. As an objective, embodied measure of brain function, EEG may have utility in many of the same ways psychiatric symptoms are used, such as the classification of illness subtypes, as measures of disease progression, and as signals that respond to effective pharmacotherapy.

HISTORY

Experimental evidence of brain-derived electrical potentials was first established in 1874 with Richard Caton's (1842 to 1926) discovery of spontaneous electrical activity recorded from the exposed cortices of cats, rabbits, and monkeys. However, the lengthy transition from laboratory experiment to the eventual acceptance of EEG was plagued by intense controversy. Despite the continuing accumulation of evidence, the notion of electrical potentials emanating from the brain was rejected for nearly 50 years by leading authorities. Caton's work was replicated by Vasili Danilevsky's 1877 report that electrical oscillations recorded from the animal brain could be altered by strong sensory stimuli. Three further historical highlights of note include (1) the 1891 demonstration by Adolph Beck that the dog visual cortex produced large electrical potentials when the eyes were rhythmically illuminated (thus laying the experimental foundation for EEG photic driving), (2) Beck and Napoleon Cybulski's 1892 report that local injury to the cortex could alter the characteristics of recorded spontaneous electrical activity, and (3) Cybulski's 1914 report that brain wave seizure discharges could be induced in the cortex by applying electrical stimulation to the cortex (thus presaging the use of

EEG in epilepsy). Despite this stream of successful experimental work, the EEG phenomenon continued to be scrutinized by the scientific community.

Caton’s seminal work was finally brought to light when cited in 1929 by Hans Berger (1873 to 1941), reporting the discovery of the alpha waves (~8 to 12 Hz) at the emergence of the field of *electroencephalography.* The focused perseverance of Hans Berger, a biologically oriented Professor of Psychiatry and Director of the Psychiatric Clinic in Jena, Germany, finally brought EEG to a position of acceptance and clinical utility. After years of unsuccessful attempts to record brain waves from humans (he was able to obtain recordings from animals), he finally succeeded in recording the human EEG in 1924 and, in 1929, published the first in his classic series of 23 papers describing many aspects of the human EEG. Among his vast achievements, Berger demonstrated that brain electrical activity came from neurons and not blood vessels or connective tissue, that recordings from patients with brain tumors contained high-voltage slow waves (his recording technique did not permit localization), that waking *alpha* waves were blocked by eye-opening, and that the characteristics of EEG activity change with age, sensory stimulation, state of consciousness, and physiochemical state of the body. However, acceptance was still temporarily delayed when Lord Adrian, a Nobel laureate neurophysiologist, claimed that Berger’s findings were “impossible.” Later, in 1934, Lord Adrian publicly confirmed acceptance of Berger’s work, and the field of electroencephalography was born.

BASIC PRINCIPLES OF ELECTROENCEPHALOGRAPHY

The term *electroencephalogram,* coined by Hans Berger, refers to the graphic record of brain electrical activity or brain waves. The related term, *electroencephalograph,* refers to specialized instrumentation used to record electrical potentials at the scalp surface. The abbreviated term, *EEG,* may be used interchangeably to refer to either the record or the recording instrument. As rendered by EEG, the brain wave reflects the transient difference in electrical potential (greatly amplified) between any two points on the scalp or between an electrode placed on the scalp and one located at a neutral off-scalp location (e.g., ear lobe, mastoid, nose), used as a reference. In a simplistic sense, the EEG is an extremely sensitive voltmeter, with the unit of measurement being the microvolt (μV), representing dynamics occurring on a scale of millionths of a volt. Typical EEG signals range from approximately 30 to 80 μV but can be as low as 1 μV (for early EPs) or as high as 150 or 200 μV in some high-voltage "spike" discharges. The difference in electrical potential measured between any two EEG electrodes fluctuates or "oscillates" rapidly, usually many times per second. It is this oscillation that produces the characteristic waveforms that even many lay persons now recognize as the appearance of "brain waves."

The earliest EEG recordings involved only one pair of electrodes or one *channel* of recording, and although this could detect certain normal and abnormal features, effective clinical application remained for the future. Soon, the ability to record two channels of EEG was accomplished, making it possible to record activity simultaneously from homologous locations in each hemisphere. Before long, the rapid advances in recording technology allowed four- and eight-channel recordings to be made, and EEG became a viable clinical tool. Eventually, 10-, 12-, and 16-channel recording machines became the standard workhorses of clinical and research EEG laboratories around the world. EEG equipment capable of simultaneous recording from 32 to 64 channels is now commonplace in most EEG laboratories, with 128- and 256-channel systems also used in research where "high-density" recording may lend further accuracy to the estimation of current sources. The ability to simultaneously record brain waves from many scalp locations is important because it allows the detection of asymmetries across and within cortical regions, permits recording arrays to locate focal or regional abnormal features more clearly, and increases the ability to detect various *artifacts* (i.e., sources of signal of noncerebral origin that contaminate the EEG).

Brain wave activity captured on the scalp by EEG is a result of the summed extracellular electric potential fields generated by large numbers of neurons. EEG cannot detect the electrical activity generated by a single neuron or even by several neurons close to the scalp. Rather, the scalp-recorded EEG signals are the result of summated field potentials generated by excitatory and inhibitory postsynaptic potentials (i.e., EPSPs and IPSPs) in vertically oriented pyramidal cells of the cortex. An EPSP in a dendrite produces electrical negativity in the immediately surrounding area, but the electrical field becomes positive with increasing distance from the source. The reverse occurs with an IPSP, generating an electrical positivity nearby and a negative field at a distance. The summation of EPSPs and IPSPs is enhanced because the neurons are tightly packed together and oriented vertically in parallel. In order to be detectable by surface EEG electrodes, neural activity has to be strong enough to cross other tissue and fluid (i.e., CSF, dura, skull, scalp). It is the transduction of current across tissue and fluid of different densities, with different electrical resistance, that contributes to the wide diffusion of current as it is recorded over the scalp. This is one important aspect of EEG recording that must be considered when making assumptions about the underlying source and the direction of the current projection. The transduction of electrical activity from cortex to scalp relies on the ability of the brain to synchronize large populations of cells, which can occur either spontaneously or in response to a stimulating event. Neuronal activity that is unsynchronized cancels out due to destructive interference and thus does not contribute substantially to the electric field potentials recorded as scalp EEG. Because of the manner in which the dominant intrinsic brain waves are generated, EEG is maximally sensitive to cortical neuronal activity and relatively insensitive to electrical potentials generated at subcortical structures. However, there are exceptions, as subcortical neuronal events may influence cortical neuronal firing via afferent transmissions along subcortical–cortical tracts.

Probably the first observation about brain waves, going back to the time of Caton, was that many recorded potentials oscillate and repeat in a rhythmic fashion. Indeed, the term *intrinsic rhythms* is often used to refer to normal background activity, and the term *dysrhythmic* is used for activity with irregularities in oscillatory patterns that may be suggestive of pathology. Within reasonable limits, the repetitive rhythmic nature of the EEG is stable across individuals and within individuals over time, barring the introduction of pathophysiologic events. Indeed, the quantified EEG signal has been shown to have high test–retest reliability in a number of recording situations, with coefficients of $r = 0.90$ reported in healthy adults and in psychiatric samples. Examination of test–retest reliability in task-related EEG, recorded under conditions of mental challenge, also tends to produce high-reliability coefficients. However, the reliability of these measurements is also known to be influenced by stimuli, task conditions, and trial count. The stability of EEG spectral characteristics may also have a genetic basis, with an average biometric heritability estimate of 0.78 reported for a large-scale ($N = 4{,}026$) community sample of adolescent twins and their parents. Although the EEG does reflect state-dependent changes, sensitive to states of wakefulness and relaxation, for instance, its reliability within individuals over time has been interpreted by some authors as evidence of an EEG set point, a natural, spontaneous rhythm that is individually characteristic and stable over time. The concept of a set point suggests that repeated testing with averaging across test sessions would help to eliminate measurement error, thus maximizing the chances of detecting the onset or progression of neurodegenerative processes. Given further evidence that EEG characteristics are to some extent heritable, EEG measures may also be useful for monitoring the evolution of neuropathology in individuals regarded to be at elevated risk for developing a psychiatric disorder due to family history.

Despite the more than 130 years that have passed since the original recording of EEG by Richard Caton, over which a variety of abnormal findings have been reported in many psychiatric conditions, EEG methods have yet to find a prominent role in the training and clinical practice of psychiatry. This is in contrast to EEG's well-established position in clinical neurology as an approved subspecialty, sanctioned by the American Board of Medical Specialties (ABMS) and the American Board of Psychiatry and Neurology (ABPN). De-emphasis of EEG training in psychiatry is likely a consequence of factors unique to the psychiatric diagnosis itself. Prominent among these factors are the complex and heterogeneous nature of psychiatric disorders in general, for which there are currently no formally accepted diagnostic techniques beyond the clinical interview. The lack of familiarity of psychiatrists with EEG techniques, and heightened standards for evidence-based practices, are additional factors to contend with when considering the feasible application of EEG in routine psychiatry practice.

Basics of EEG Interpretation

The interpretation of the EEG record is essentially a problem of learning to recognize all of the myriad intrinsic waveforms, their expected distributions over cortical regions, and their range of variation in amplitude and degree of symmetry. Although there is little doubt that 64, 32, or even 16 channels of oscillating waves appear quite confusing to the beginner, there is an inherent regularity to the brain's electrical output that an experienced electroencephalographer comes to recognize. Because of this inherent regularity, the intrinsic EEG rhythms and their range of variation that characterizes the wake, drowsy, and various sleep states become easily identified through practice and experience. The various electrophysiologic phenomenon of nonencephalic origin, which need to be distinguished from EEG signals, also produce localized or regional waveforms of stereotypical appearance that one can learn to identify. The primary job of the interpreter is to identify those EEG waveforms that appear to fall outside of what one might call the *normal range of variation* of the intrinsic background activity and therefore are (1) frankly abnormal with known pathophysiologic correlates or (2) controversial abnormal waveforms of potential clinical relevance, pending the results of further continuing clinical investigation.

Interpreter Reliability

Because the EEG record is obtained by precisely measuring the microvolt fluctuations of electrical potentials over the scalp, a misconception often emerges that EEG interpretation is purely objective in the sense that measurements of temperature, weight, length, and volume are objective. In truth, there is a large, subjective element of judgment in EEG interpretation, and the achievement of skill in this area only follows a period of thorough training and the guiding hand of clinical experience. Accepted EEG abnormalities do not always appear in the EEG tracing in clear-cut textbook form, and there are always ambiguities in areas where EEG activity that is clearly normal becomes blurred and shades off into waveforms that are unequivocally abnormal.

Although the important topic of reliability of clinical EEG interpretation has not received extensive study, it has generally been assumed that high levels of interpretation reliability can be obtained through training and clinical experience. Two formal methods of assessing reliability involve comparing the independent readings of two or more EEG interpreters (*interrater reliability*) and asking one electroencephalographer to blindly reinterpret a sample of EEGs after an elapsed time (*intrarater reliability*). When these methods have been put to the test in studies of board-certified clinical neurophysiologists asked to classify EEGs into one or more diagnostic categories, findings have suggested that clinical interpretation of EEG should receive much greater attention as a matter of training and practice (Grant et al., 2014). It was reported that while trained readers tended to provide high confidence ratings of their interpretation (median score of ≥99% accuracy), intrarater reliability coefficients ranged from 0.33 to 0.73 with an aggregate value of 0.59, while interrater coefficients ranged from 0.29 to 0.62, with an aggregate value of 0.44. The take-home point of this study is that while experienced epileptologists have very high confidence in their EEG interpretations, this confidence may not be founded in actual performance and that further training to improve EEG accuracy and reliability is needed. One suggestion for accomplishing this goal is through continuing medical education modules that measure and report EEG reliability scores to individual users.

EEG RECORDING AND RELATED METHODOLOGY

Much has been written about the complexities of EEG recording, interpretation, and the corresponding level of skill needed to obtain an adequate EEG. What may be insufficiently recognized is the fact that there are also clinical situations in which a greatly simplified EEG obtained by a properly trained registered nurse or resident can have substantial diagnostic usefulness. There are some important EEG findings of particular relevance to emergency room settings and, possibly, even to some acute psychiatric admission or triage units that can be obtained in only 10 minutes by clinicians with a minimum level of technical skill. Clinical cases presenting moderate to marked confusion and agitation, delirium, or possible nonconvulsive status epilepticus (NCSE) may have diffuse EEG abnormalities that are more or less continuous in the tracing, detectable almost as soon as the recording is turned on. Such findings do not require sophisticated localization studies, and their presence, or absence, is diagnostically relevant. Prompt access to an EEG laboratory may not always be possible, especially on weekends or evenings. Thus, on-site screening may be most helpful.

Beyond routine screening EEGs described above, recording the EEG does, in fact, require a considerable amount of skill and experience. It is not merely a technical act performed by a technician. The unfolding clinical EEG record is a constantly moving and shifting display of complex waveforms recorded simultaneously from numerous scalp locations, and the EEG patterns differ dramatically during wakefulness, drowsiness, and various sleep levels. The appearance of the EEG also changes from one recording montage to another, along with differences in the reference selected, while a host of normal and abnormal EEG waveforms and contaminating artifacts must be identified in their obvious and subtle forms. Identifying the source of artifacts and minimizing their intrusion on the EEG is most critical during the recording procedure, as the application of offline removal and correction procedures requires an advanced skill set and appreciation of the intricacies of how such transforms may inadvertently change the appearance of features retained in the EEG. In addition to having the skills to correctly apply electrodes and operate amplifiers and recording equipment, better technologists are also capable of sophisticated EEG interpretation. It may not be obvious how important this is. EEG abnormalities do not always emerge in clear-cut, textbook form but instead may be distorted and, hence, ambiguous. Interpretative ability is necessary to recognize probable abnormalities and then arrange recording montages and monitor states of patient alertness (wake or sleep) in ways that can enhance the patterns and allow a more definite interpretation by the electroencephalographer. A minimum of 1 year of full-time training, including didactic instruction and supervised, hands-on recording and interpretation experience, is necessary for an EEG technologist to achieve competence. Formal training schools for EEG technologists exist in many places, and the graduates can become registered EEG technologists by taking and passing a two-part written and practical examination.

Working in a psychiatry environment has its own challenges that technicians must be comfortable handling. First and foremost is being able to competently manage the needs of a disturbed individual. It should be appreciated that clinical EEG is typically administered under conditions proximal to an acute change in mental state; therefore, psychiatric symptoms may be exacerbated. An EEG technologist who practices in a psychiatric setting must know when to terminate a procedure, call for help, and try to de-escalate a situation. Even the most cooperative psychiatric patients can also present challenges for EEG recording due to increased eye and involuntary body movements associated with tardive dystonia and akathisia. Muscle activity, particularly of the eyes, mouth, and jaw, produces changes in electrical activity on a much larger scale than brain activity and can contaminate an EEG record to the point where no reliable interpretation can be made. Given that good awake recording is necessary, sedation is not an acceptable procedure. On the other hand, the need for sleep in order to more completely assess *paroxysmal* activity, such as spike-wave discharges associated with seizures, may require that other steps be taken to help the patient to fall asleep. This may require decreasing the lighting and allowing the patient to relax. Room temperature in EEG recording environments is ideally kept between 70 °F and 75 °F (21 °C to 24 °C), a comfortable temperature at which an individual would not be expected to perspire (which can cause skin potential artifacts). It is also important to minimize changes in body temperature that could contribute to perspiration at the scalp, as sweat interacts with electrolytic gel applied to maintain conductance and reduce electrode impedance, produces slow drift and other artifacts affecting EEG signal, and can contribute to "salt bridging" of adjacent electrodes (essentially connecting electrodes together so no differentiation can be made between signal recorded at the bridged sites). Perspiration and associated changes in skin conductance are also associated with increased sympathetic arousal, which should be recognized in agitated or highly anxious patients. For these reasons, patients should be encouraged to remove heavy clothing or excessive layers if tested in street attire, and, as stated previously, every effort should be made to keep the patient comfortable and relaxed. Explaining the purpose of the procedure, the EEG apparatus, and the signals obtained by EEG recording (e.g., an analogy to an ECG of cardiac activity but with electrodes placed on the head to record brain activity) during electrode application can go a long way in helping the patient feel informed and at ease with the process. Careful procedures will prolong the recording time in psychiatric patients. However, in laboratories where psychiatric patients are not routinely evaluated, technicians may be inclined to hurry the procedure when it should be slowed down.

EEG Apparatus

The basic components of the EEG recording apparatus are the (1) electrodes, (2) amplifier, and (3) recording device, including means of visualizing, acquiring, and storing the EEG record. Each will be discussed in the following sections.

Electrodes. As EEG emerged into the clinical arena, electrodes simply were placed on the scalp symmetrically by eye, using salient landmarks on the head as reference points, and not all laboratories used the same placement system. Eventually, in 1947, it was decided at an International EEG Congress held in London that some effort should be made to standardize the system of electrode placement so that clinical and research findings would be more directly comparable across different laboratories. The challenge was taken up by Herbert Jasper, who developed the *10–20 International System* of electrode placement, which has become standard worldwide since 1958. The system is based on two anatomical landmarks for positioning; the *nasion,* identified as the depressed area between the eyes above the bridge of the nose, and the *inion,* which is the highest point of the external occipital protuberance, located by a prominent bump at the back of the head near the lowest point of the skull. Briefly, measurement is taken from the total nasion-to-inion distance, and the midpoint is marked at a distance of 50% between these points (and 50% of the transverse distance between the preauricular points on the ears). The 10–20 nomenclature refers to the distance between electrodes, positioned at either 10% or 20% of the actual anterior–posterior or transverse distance of the skull (Fig. 1.26–1A,B). Each electrode location is associated with an underlying area of the cerebral cortex, indicated by the letters F, T, C, P, and O, reflecting frontal, temporal, central, parietal, and occipital lobes, respectively, followed by either the lowercase letter “z” for midline electrodes or a number indicating homologous left- (odd numbers) or right- (even numbers) sided locations. Thus, electrodes O1 and O2, for example, are placed over the left and right sides of the occipital region, respectively. Electrode Cz also referred to as the vertex, is the most centrally located electrode and is used as the midpoint for alignment to the intersection of nasion-to-inion and transverse distance measures. The standard 10–20 system is based on a total of 21 electrodes, with 19 scalp sites, including two electrodes at frontal polar sites (Fp1, Fp2) and two electrodes affixed at earlobes (A1, A2) which have traditionally served as a common reference for scalp electrodes.

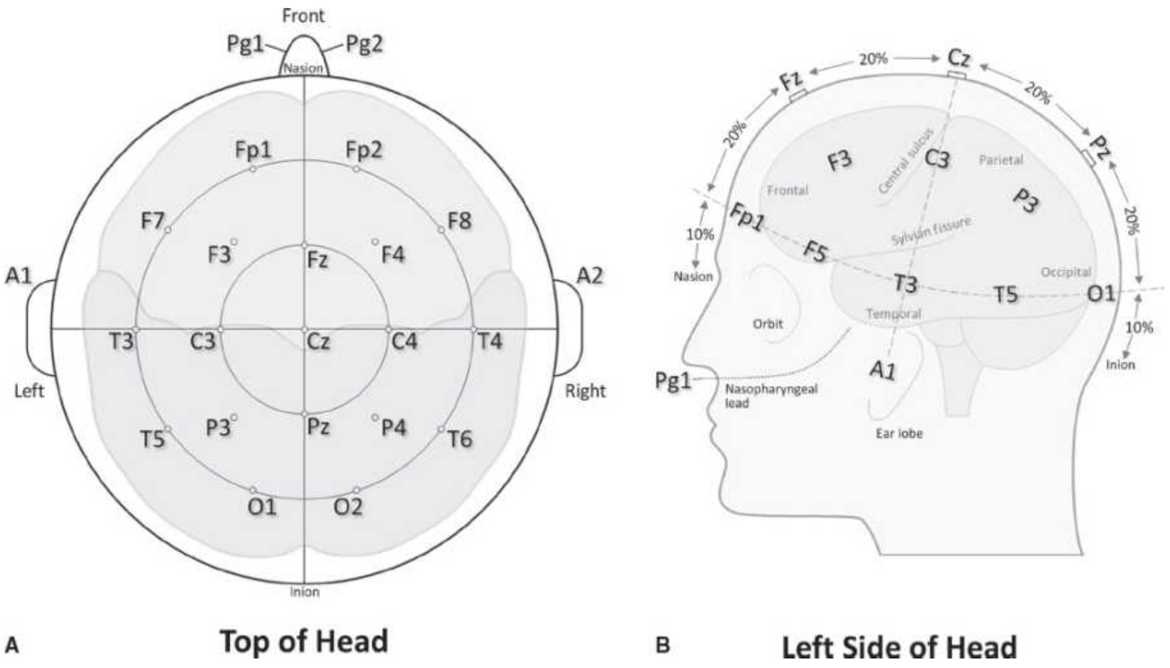
Front
Pg1
Pg2
Nasion
Fp1
Fp2
F7
F8
F3
Fz
F4
A1
A2
T3
C3
Cz
C4
T4
Left
Right
T5
P3
Pz
P4
T6
O1
O2
Inion
A
Top of Head
Fz
20%
Cz
20%
20%
Pz
F3
C3
Central sulcus
Parietal
P3
Frontal
Fp1
F5
Sylvian fissure
10%
Nasion
20%
Occipital
T3
T5
O1
Orbit
Temporal
10%
Inion
Pg1
Nasopharyngeal lead
A1
Ear lobe
B
Left Side of Head

FIGURE 1.26–1. **A:** International 10–20 Electrode Placement System, as viewed from the top of the head. **B:** A left-lateral diagram of the head showing the locations of the routine 10–20 electrodes. Note the relationships between electrodes placed on the scalp and underlying brain structures.

Although precise localization with respect to underlying functional anatomy should not be assumed, some basic generalizations have been made concerning electrode location and the nature of brain activity detected in their region. For instance, the frontocentral electrode Fz is located in proximity to frontal lobe structures important to intention and motivational processes, while central sites near the vertex (C3, Cz, C4) are sensitive to sensory (particularly auditory registration) and motor processes, and occipital locations O1 and O2 to early-stage visual processes. In addition, temporal sites (T3, T4, T5, T6) have shown sensitivity to memory and emotional processes, with temporoparietal electrode sites (P7, P8) typically selected for experimental studies on the neural activity involved in structural facial encoding. Again, the mention of these findings is not intended to insinuate that EEG is an effective tool for localization of function to a known source but rather to provide examples that illustrate the importance of multichannel recording and standardized electrode placement for characterizing unique and spatially distributed brain signals. Of note, the functional associations described above would not be inferred from routine clinical EEG but rather from experimental stimulus-evoked and task-related procedures, to be discussed later in this chapter.

The 10–20 system enables expansion of electrodes to provide greater coverage, with additional electrodes typically placed at 10% of the distance between the original 10–20 locations. The American EEG Society (currently the American Clinical Neurophysiology Society [ACNS]) has developed a nomenclature for the designation of as many as 75 defined electrode locations. However, it must be stressed that extremely large numbers of scalp electrodes are unnecessary for currently established clinical EEG applications. For currently accepted clinical indications, optimally useful EEG recordings can be achieved with as few as 16 (minimum recommended standard) or, at the most, 32 scalp electrodes. While preparation and setup time must always be considered as a constraint on the number of electrodes that can feasibly be applied in a given recording situation, arrays of 128 or even 256 scalp leads are not uncommon in specialized research applications where the aims of data collection include current source analyses and three-dimensional modeling (benefitting from additional electrode placements below scalp). An example of an extended 64-channel electrode array can be seen in Figure 1.26–2.

The 10–20 placement system has several advantages. Because the placement system is based on rigorous measurements, electrode placement error, particularly the asymmetrical placement of electrodes for homologous electrode pairs, is greatly minimized. The system also renders recordings entirely comparable between laboratories, as well as across serial recordings obtained from a single subject. Because percentages of distances between landmarks on the head are used for placement locations, scalp electrodes overlie the same cortical regions despite differences in head size. Furthermore, the relationships between electrodes placed on the scalp and underlying brain structures have been well established (Fig. 1.26–1) by using placements on cadavers (with holes drilled at the electrode sites to later identify the cortical area under the electrode), as well as studies using CT scanning.

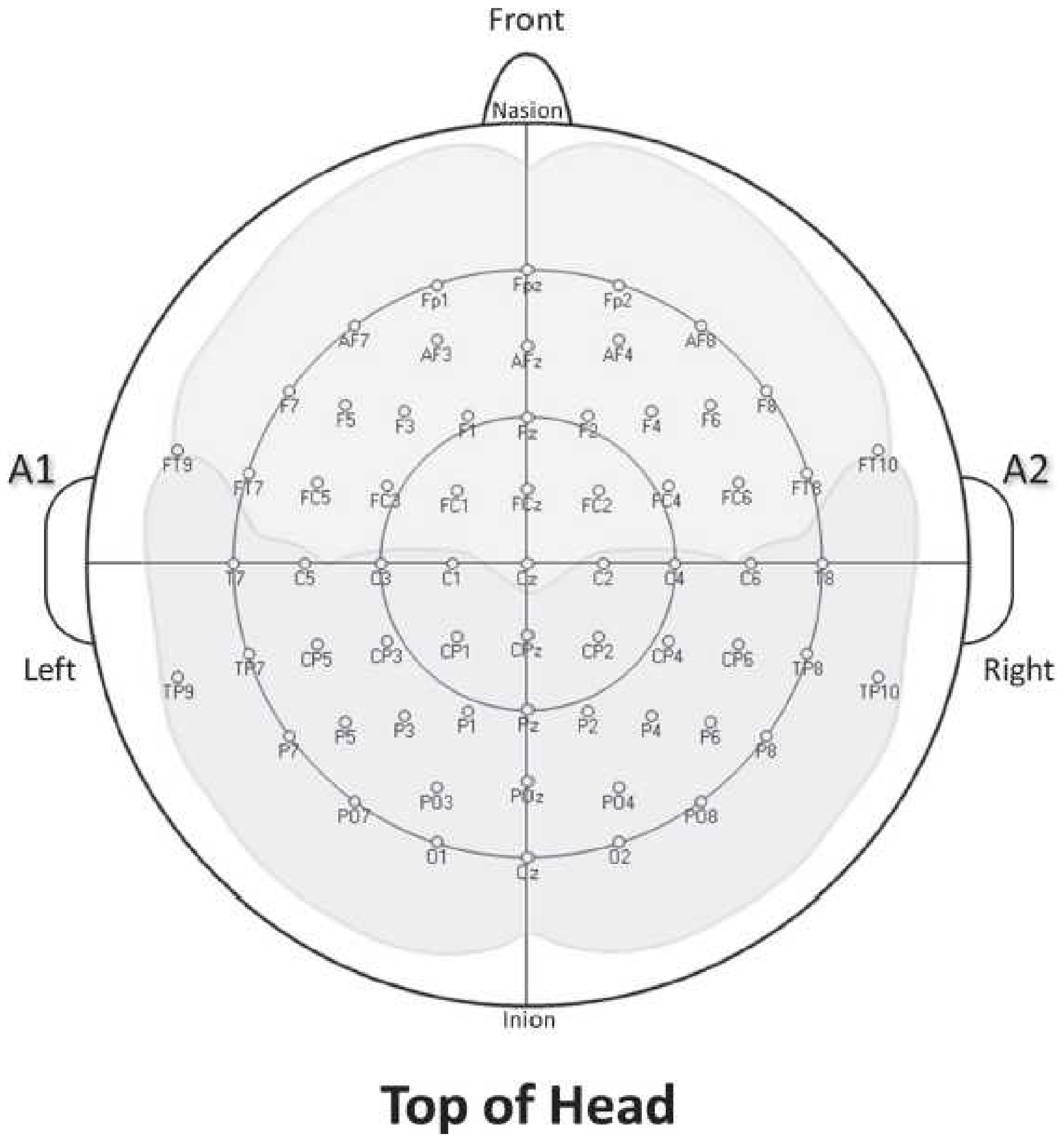
Front
Nasion
Fp1
Fpz
Fp2
AF7
AF3
AFz
AF4
AF8
F7
F5
F3
F1
Fz
F2
F4
F6
F8
A1
FT9
FT7
FC5
FC3
FC1
FCz
FC2
FC4
FC6
FT8
FT10
A2
T7
C5
C3
C1
Cz
C2
C4
C6
T8
Left
TP7
CP5
CP3
CP1
CPz
CP2
CP4
CP6
TP8
Right
TP9
TP10
P7
P5
P3
P1
Pz
P2
P4
P6
P8
PO7
PO3
POz
PO4
PO8
O1
Oz
O2
Inion
Top of Head

FIGURE 1.26–2. An expanded 64-electrode array. Such extensive placement systems are primarily used for special research studies and are rarely used for clinical recordings.

Scalp electrodes must be applied carefully. The skin under the electrode must be clean and completely free of oil or grease. A common practice is to rub the area with a slightly abrasive cleansing electrolyte material that also removes some of the superficial epidermis. When this is done, a metal disc electrode can be applied to the scalp by using a conducting electrode gel. Electrode impedance should be maintained at equal to or less than 10,000 ohms (Ω) (10 KΩ) and closely matched across electrodes, ideally within 1 KΩ of each other. The electrode application procedure should not be uncomfortable for the subject.

However, scalp abrasion has typically been used to perforate the most superficial layer of skin as a method to reduce scalp impedance. Modern EEG systems designed for research enable high-impedance recording through the use of *active electrodes* (of note, active systems have not yet been approved for clinical use). Active electrode systems do not require skin abrasion, and rather, acceptable signal-to-noise quality is achieved through preamplification at the electrode stage. Placing the first stage of amplification nearer to the scalp reduces the propagation of electrical interference through the shielded electrode cable to the amplifier, at which point line noise is amplified at the same scale as the EEG signal. Active electrodes do not overcome the necessity to assure good conductance at the scalp site through careful setup and application of electrolyte gel but may reduce overall preparation time by foregoing the need for scalp abrasion common to passive electrode systems.

Beyond the standard montage of scalp electrodes, certain applications may require additional electrodes specially designed to record from deep brain structures. Nasopharyngeal (NP) electrodes are inserted into the nostrils to achieve EEG recording with closer proximity to the basal temporal lobe that can be obtained by scalp electrodes (Fig. 1.26–1B); these leads are designated Pg1 and Pg2 in the 10–20 placement system. No actual penetration of tissue occurs. The NP lead is a long (as long as 15 cm for adults), curved S- or Z-shaped insulated wire with a silver ball (the electrode) on the tip, which is rotated laterally after insertion in order to place the ball in contact with the roof of the nasopharynx. With a cooperative patient and a skilled technologist, the procedure can be well tolerated. Although this lead is presumed to be better positioned to detect activity from the orbitofrontal cortex, temporal pole, and hippocampus, it has numerous disadvantages. Chief among these is a high propensity to produce pulse and respiration artifacts and the fact that NP leads cannot be used when a deviated septum or nasal inflammatory process is present. They also are contraindicated, with many psychiatric patients displaying behaviors such as confusion, agitation, or belligerence who could pull the leads out, possibly lacerating the nasal passage. Their use also may interfere with obtaining a sleep-activated EEG and, not infrequently, otherwise cooperative patients simply refuse the procedure.

MONTAGE SELECTION. A common misconception is that the EEG records the voltage detected at each electrode site. Instead, each channel on the

EEG chart represents the shifting or oscillating difference in electrical potential *between* two electrodes. Thus, in a multichannel recording, the activity from each channel represents the shifting difference in voltage between two selected electrodes. When 16, 21, or more electrodes are placed on the head, the number of possible electrode pairs becomes large, and how these pairs are arranged among the recording channels can become complex. In EEG parlance, the way electrode pairs are arranged for a recording is called a *montage,* and although many montages are possible, only a limited number have become popular and useful. Prior to the advance of digital EEG equipment, selection of the montage was done prior to recording and would determine how the EEG was visualized in the paper record. When digital EEG equipment is used, montages are preprogrammed. A recording is performed using one montage, but the recorded (digitized) signal can be viewed offline in any configuration desired.

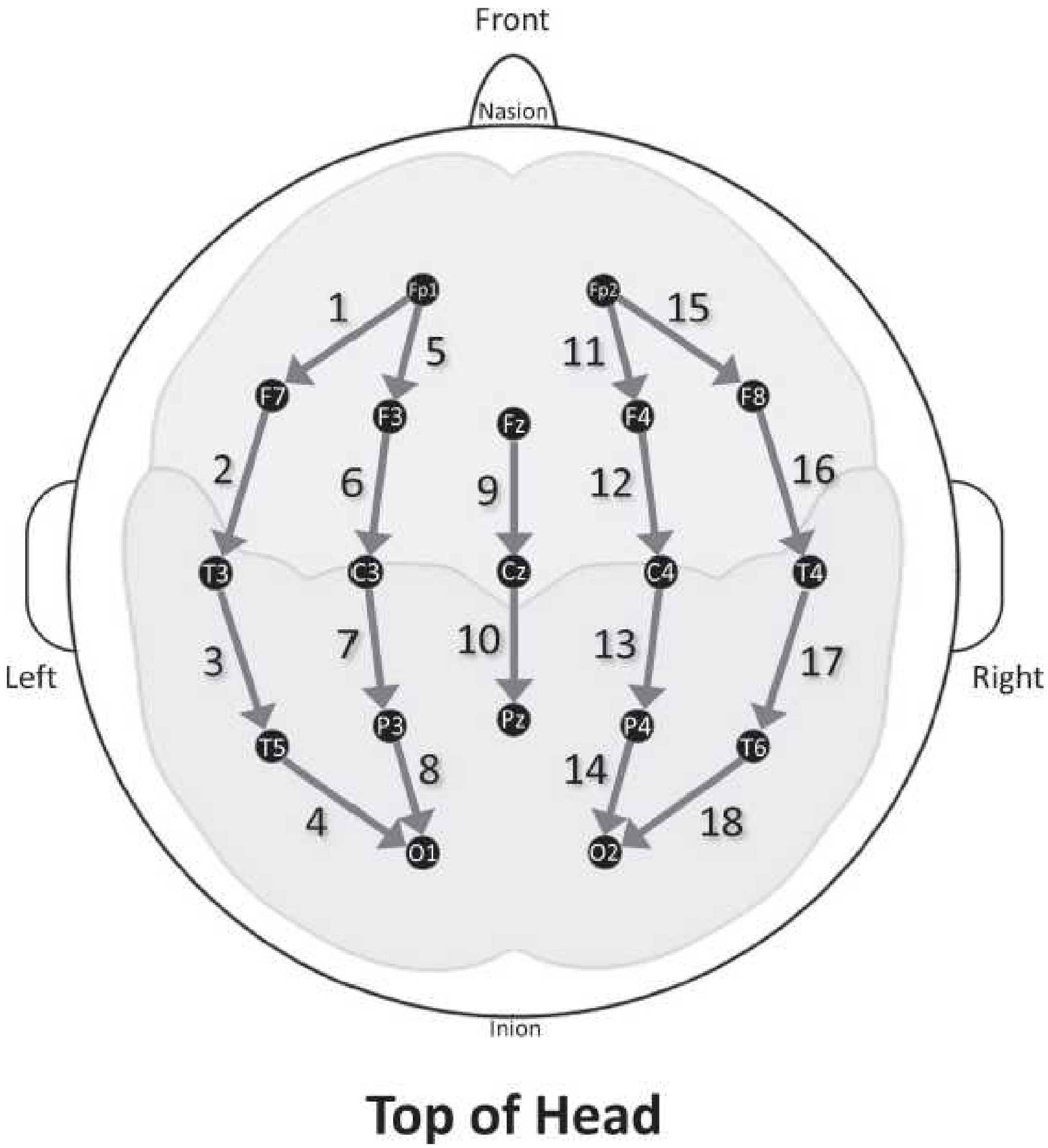
Front
Nasion
Fp1
Fp2
1
5
11
15
F7
F3
Fz
F4
F8
2
6
9
12
16
T3
C3
Cz
C4
T4
3
7
10
13
17
Left
Right
T5
P3
Pz
P4
T6
8
14
4
18
O1
O2
Inion
Top of Head

FIGURE 1.26–3. Example of an 18-channel bipolar montage with anterior to posterior linkages. The numbers between electrode locations designate recording channels. Thus, the number 6 means channel 6, which measures the difference in electrical potential between F3 and C3 electrodes.

Montages are designed to facilitate the detection of EEG abnormalities in different brain regions and to facilitate comparisons between left and right hemisphere activity. There are general guidelines for how montages are to be set up. The most important rule is simplicity for the sake of interpretation. Additional rules include the stipulation that odd numbers refer to the left-side, whereas even numbers refer to right-side electrodes (Fig. 1.26–1A). Furthermore, left-side electrodes are routinely displayed at the top of or before right-side electrodes on the record. Similarly, anterior electrodes are displayed on top of or before more posteriorly placed electrodes.

There are two main types of montages: referential and bipolar. With referential montages, all electrodes are referenced to a single common reference point that traditionally consisted of linked ears (the mastoid prominence can be used in place of the ear lobe), with variations being left or right ear reference alone, ipsilateral ear reference in which all electrodes in one hemisphere are referenced to the ear on that side, or a contralateral ear reference in which all electrodes in one hemisphere are referenced to the opposite-side ear. Given the risk of unintended distortion of EEG (causing artifactual asymmetry between hemispheres), that can occur when using references based on bilateral sites, particularly if the impedance is not carefully balanced between the two electrodes, single, midline, sites including the nose or vertex (Cz) have been recommended in some cases. An in-depth discussion of arguments in favor and against the various reference options is beyond the scope of this chapter, but suffice it to say that no one option will be ideal (or collectively accepted) for all EEG recording applications. A host of “reference-free” recording approaches have also been employed in digital EEG recordings to circumvent a choice of reference altogether. Reference-free montages are based on common average or other mathematical derivations of signal from all electrodes, such as *current source density* (CSD), defined by the second spatial derivative of the local field potential (LFP) (for a detailed discussion of field potentials in EEG, the reader is referred to the text of Nunez and Srinivasan, 2006). As per some suggested guidelines, the selection of a reference electrode should entail a location that is electrically neutral, uncontaminated by obvious physiologic (e.g., mastoid reference can be sensitive to cardiac signal if not properly located) or motion (e.g., nose reference can be sensitive to head movement if not properly affixed) artifact, and relatively distal to electrodes in the region of interest (e.g., EEG activity most proximal to reference is most extensively subtracted). A referential montage would most likely be chosen for applications in which it is important to ascertain the magnitude of an abnormality in terms of how large a sharp wave or slow wave is in microvolts in one region compared to another.

Bipolar montages, on the other hand, are useful (and, indeed, are much more widely used in neurology practice) for pinpointing the area of maximal abnormality or the exact source of abnormal activity. In bipolar montages, electrodes are referenced from one scalp location to an adjacent scalp location in chains of electrodes going across the head from front to back (Fig. 1.26–3) or from left to right (Fig. 1.26–4). Thus, one can think of a given channel of EEG activity as being derived from two inputs. The majority of abnormal cerebral activities tend to appear at the surface as negative potentials.

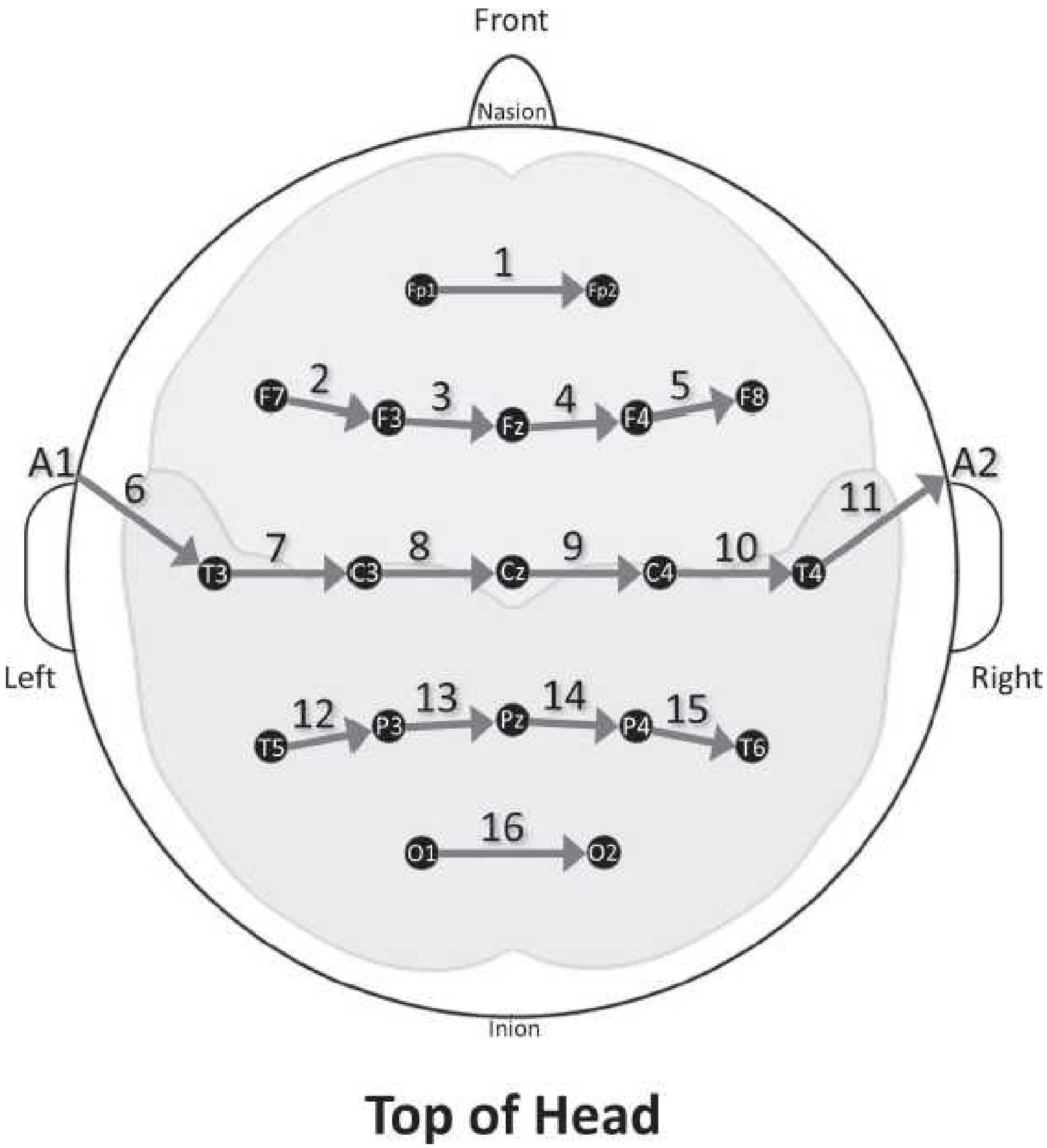
Front
Nasion
1
Fp1
Fp2
2
3
4
5
F7
F3
Fz
F4
F8
A1
A2
6
7
8
9
10
11
T3
C3
Cz
C4
T4
Left
Right
12
13
14
15
T5
P3
Pz
P4
T6
16
O1
O2
Inion
Top of Head

FIGURE 1.26–4. Example of a 16-channel transverse bipolar montage.

By convention, the first electrode of a pair constitutes input 1, whereas the second electrode provides input 2. Thus, in the electrode pair C3-P3, for example, the first electrode, C3, constitutes input 1. The direction of deflection in the EEG waveform is based on whether input 1 is, relatively speaking, “more negative” or “less negative” (i.e., relatively more “positive”) than input 2. If the electrode designated as input 1 is closer to the source of a negative field and, hence, more “negative” than input 2, an upward deflection (i.e., negative to positive/up) will be observed as an indication of input 1’s closer proximity to the source of an event. Conversely, if input 1 is more distant from the source of the field than input 2 and, hence, less negative, then the deflection of the bipolar pair is downward (i.e., positive to negative/down). As bipolar pairs of electrodes move in a longitudinal or transverse direction from one side of a strong and highly localized negative field to the other side, the deflection changes direction across bipolar pairs as input 1 shifts from being relatively more negative to relatively more positive than input 2. This change of deflection is called a *phase reversal* and is a powerful method for the localization of sharply focal abnormalities.

One additional montage configuration deserves special mention because it may be particularly useful in psychiatric EEG. This is a montage that combines referential and bipolar electrode arrangements. Following four bipolar connections from the frontal regions through the temporal areas and ending in the occipital region, a referential placement connects each posterior temporal region (T5 and T6) to the opposite ear. This arrangement allows the activity of low amplitude to be highlighted by the referential electrodes for further examination via the bipolar electrode pairs. This montage is commonly referred to as the *Queen Square montage* (Fig. 1.26–5).

The appearance of EEG activity varies greatly from one recording montage to another. Large interelectrode distances often yield higher voltages, whereas a close spacing between electrodes in bipolar pairings tends to reduce voltage since both electrodes will overlie approximately the same portion of an electrical field, and the signal common to both cancels out. Furthermore, specific EEG patterns visible in one montage may be distorted or even completely canceled out in another montage. Although some montages may permit differentiation of activity between two or more brain regions, other montage choices may not do so. For example, EEG sleep patterns are well visualized and differentiated in central and occipital regions when a common (monopolar) reference recording is made. However, differentiation between central and occipital sleep activity is no longer possible when bipolar anterior–posterior linkages are used (C3-O1 and C4-O2), and, with transverse bipolar links between homologous electrodes, the sleep patterns may not be visible at all. The issue is not merely academic. Discharges of interest to the electroencephalographer that are detectable in some recording montages may be completely or nearly undetectable, despite their neurophysiologic presence, when a different montage is being used. As previously mentioned, while traditional analog EEG required any change in a montage to be done before recording, modern digital EEG is much more flexible in allowing changes of reference (i.e., rereferencing) or montage to be done as an *offline* procedure (i.e., after data is recorded). In the case of digital EEG, a change in the reference does not alter the EEG record, only the appearance of the data for the purpose of visualization.

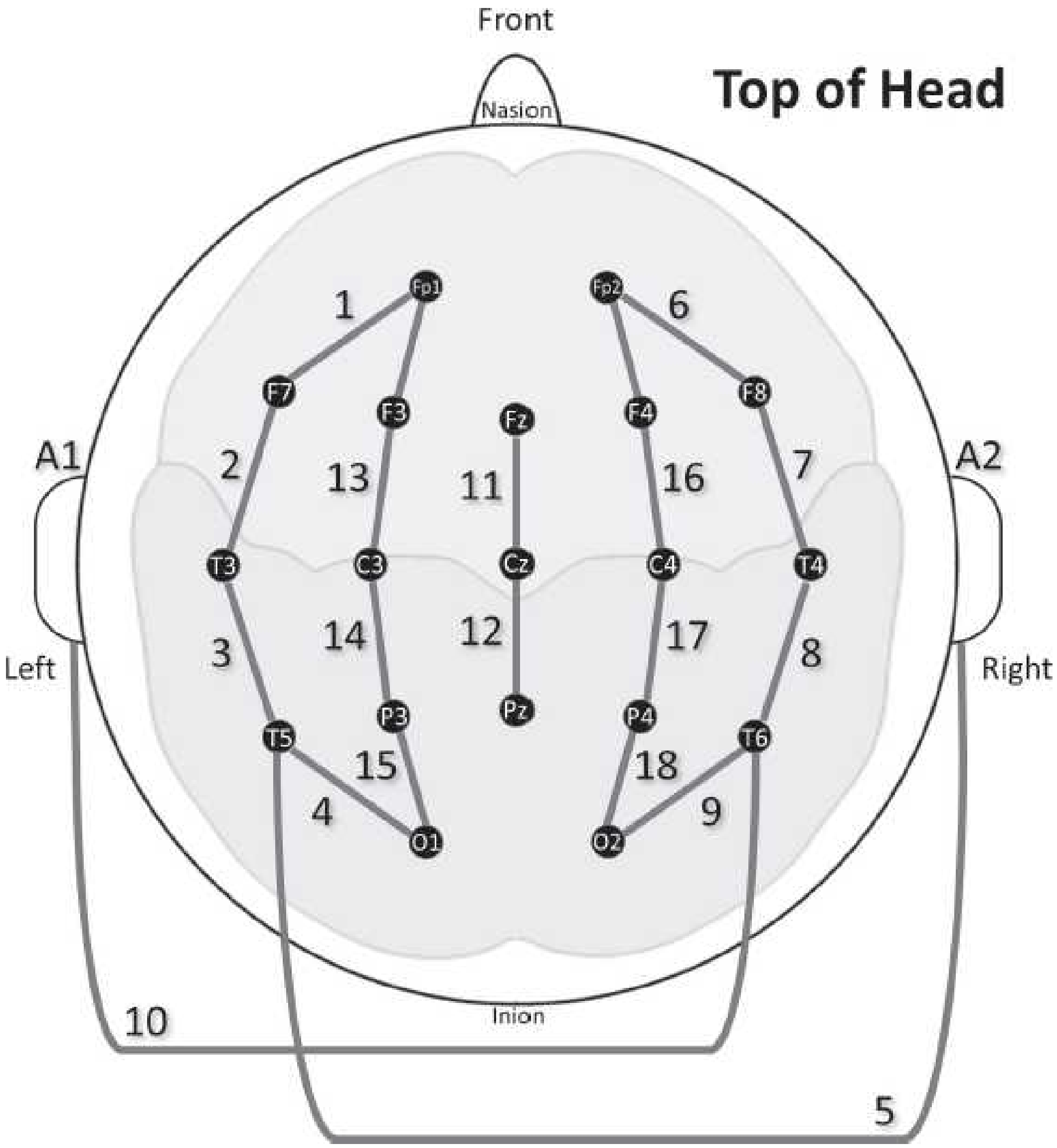
Front
Top of Head
Nasion
A1
A2
Left
Right
Inion
Fp1
Fp2
F7
F3
Fz
F4
F8
T3
C3
Cz
C4
T4
T5
P3
Pz
P4
T6
O1
O2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

FIGURE 1.26–5. Diagram of the Queen's Square montage. This is an 18-channel montage modified to include two referential leads to highlight temporal lobe activity.

EEG Amplifiers. Biopotentials recorded at the scalp are very small and require amplification (typically with a voltage gain of 1,000 to 100,000 times or 60 to 100 dB) in order to be visualized and, subsequently, analyzed. Indeed, scalp potentials are typically smaller in scale than many sources of interference, including nonencephalographic biopotentials (e.g., myographic, ocular, cardiac), electric line noise, and shifts in voltage that occur with poor electrode conductance at the recording surface. Therefore, a critical requirement for an EEG amplifier is the ability to boost brain signals while minimizing the amplification of these contaminating signals. This is achieved through the use of a *differential amplifier* using two input sources, one from the electrode and the other from a common system reference. The term “differential” refers to the manner in which a signal is amplified, wherein the difference between the two input voltages is passed through to the amplification stage while any common voltage shared between inputs (regarded as noise) is rejected. The proportion of common voltage rejection relative to the amplified difference signal can be quantified as the *common-mode rejection ratio* (CMRR), expressed in decibels (dB) as a unit of relative gain. As previously mentioned, the quality of EEG recording depends highly on the ability to achieve and maintain low impedance at scalp and reference sites. With regard to CMRR, it is especially important to ensure that impedance is balanced across electrodes, in that unequal impedance reduces the differential amplifier’s ability to detect and reject common-mode signals. Early analog EEG amplifiers required adjustment of *sensitivity* (i.e., the ratio of input voltage to pen deflection) to find the range of amplification that would boost low-voltage signals high enough to be visualized but prevent high-voltage signals from pushing recording pens above their deflection limits where “squaring off” of waveform peak would occur. The necessity to adjust sensitivity, or gain, in this manner is eliminated in modern EEG amplifiers that employ analog-to-digital (A/D) conversion, replacing pen and paper with software for control of EEG acquisition, visualization, and data storage. Most software allows input gain can be adjusted as needed for visualization at any point during EEG acquisition, with no consequence to the actual recording of the signal.

A feature found both in traditional analog and modern digital EEG amplifiers is an online filter, placed before the sampling stage, where it functions to limit the bandwidth of incoming signals in accord with the sampling theorem. This filter is set to the amplifier’s *Nyquist frequency,* defining the highest frequency of a discrete signal that can be accurately characterized at a specific sampling rate. High signal frequency relative to sampling rate can produce *aliasing* in the recording, a distorted rendering of a sinusoid that results from undersampling. *Nyquist theorem* dictates that the sampling frequency (F_S) must be at least twice the frequency of the highest frequency contained in the signal (F_C). Accordingly, Nyquist frequency is constrained by the amplifier sampling rate, as expressed mathematically in the equation $F_{\text{Nyquist}} = F_C < F_S/2$, with both signal frequency and sampling rate measured in cycles/samples per second using the common unit Hertz (Hz). Thus, the first stage of filtering may be regarded as an "*antialiasing filter*" used to attenuate EEG frequencies at and above the limit that can be accurately sampled. Aliasing has become less of a concern over time as amplifiers used in clinical applications sample minimally at rates of 256 to 512 Hz, which is more than sufficient to characterize EEG within the range of most human brain functions (e.g., ∼1 to 100 Hz). Indeed, modern digital amplifiers are capable of sampling well beyond 1,000 Hz, and usually up to 20 kHz, though doing so comes at the cost of much larger data files, with higher computer processing speed and storage space requirements, and little useful information gained. One notable exception to this generalization is testing of the auditory brainstem response (ABR), which can include events sampled at rates of 7 kHz to 50 kHz using specialized amplifiers.

Digital Recording. Over the last 30 years, the analog EEG traces of the old EEG machines have been largely replaced by computer systems providing an A/D conversion of the recorded EEG signals, as well as amplification, digital filtering, storing, and quantitative analysis of multichannel EEG. Digital systems are now commonplace in the laboratory and clinical settings, are relatively inexpensive, and offer several practical advantages. They can help reduce the space problem of storing paper records over many years and also allow review of an EEG record using multiple montages, filters, and vertical (gain or sensitivity) and horizontal scaling (e.g., paper speed) selected after the original recording. Moreover, digital recording systems allow many new and exciting possibilities for EEG signal processing and quantification that rely not only on time-series analysis but also spectral–frequency, time–frequency, related methods of coherence and phase synchrony, and spatial–temporal source localization and modeling (to be discussed in subsequent sections).

FREQUENCY FILTER SETTINGS. Nearly all of the EEG activity that is analyzed for clinical or research purposes falls within the frequency range of 0.5 to 50.0 Hz (and occasionally up to 100 Hz). Filters are used essentially to limit the range of frequencies to a bandwidth containing features of interest. It has become conventional in clinical EEG to use a *low-pass filter* setting of 70 Hz, which means that EEG frequencies above this filter setting become progressively attenuated while those below the filter setting as allowed to “pass” through and are retained for further analysis. At the other end of the spectrum, most laboratories also employ a *high-pass filter* between 0.5 and 1.0 Hz to reduce the registration of frequencies below this level. Because the EEG is represented as a two-dimensional image of amplitude over time, with the visualized waveform representing a summation of all frequencies, a *band-pass filter* (combines a high- and low-pass filter) may be used to isolate discrete frequency bands (e.g., alpha band, 8 to 12 Hz) of interest from the broadband signal.

Unfortunately, scalp electrodes pick up a variety of electrical potentials of noncerebral origin, and many of these have frequencies within or close to the EEG frequency spectrum. Frequency filters may, to some degree, mitigate against the distorting effects of frequencies generated by artifacts. However, filters must be used judiciously and with caution because they also can filter out real brain waves that one wishes to observe. For example, slow waves generated by changes in skin conductance during sweating (galvanic skin response) can be of a sufficiently high magnitude to completely obliterate genuine EEG activity in the affected recording channels (bilateral frontal-anterior temporal areas around the forehead are particularly prone). Raising the high-pass filter setting to 5 Hz effectively eliminates this source of contamination in the recording but does so at the expense of attenuating any real generalized or focal slow activity that also may be present. It is also common practice to adjust the low-pass filter downward from 70 to 35 or even 15 Hz to eliminate or reduce unwanted electromyographic (EMG) activity from the recording. Again, the choice to do this involves a compromise because attenuating high-frequency activity may obscure the accurate detection of certain fast spike discharges. Of note, adjusting filters in this way is permissible for clinical applications, with the goal of identifying significant features at the individual level. Naturally, the selection of filter settings in research applications must be kept standard across data samples so as not to systematically introduce differences that would invalidate the measure and would be based primarily on the resolution of the frequencies important to the experiment or neurophysiologic processes being evaluated.

EEG CHARACTERISTIC FEATURES

While the appearance of the raw EEG may vary somewhat between individuals, the range of frequencies, voltages, and waveforms that characterize the normal EEG during wake and sleep has been well established. The normal EEG tracing is composed of a complex mixture of many different frequencies, and based upon modern spectral analysis; consistent, recognizable patterns can be observed in most healthy individuals. Some frequency bands are expressed more strongly over some cortical regions than others (e.g., alpha over occipital cortex). Moreover, the frequency profile varies considerably as the subject being recorded transitions from alert wakefulness into sleep. Following the lead of Berger, discrete frequency bands within the broad EEG frequency spectrum are designated with Greek letters. However, it should be noted that the sequence of Greek letters is not entirely logical and, in fact, derives from historical precedents. For example, the term alpha rhythm, discovered by Berger, was chosen because it was the first rhythm detected and described. By contrast, the designation theta derives from the fact that activity in this frequency range was initially thought to originate in the thalamus.

Normal Intrinsic Frequencies

A brief discussion of the most established and recognizable *intrinsic rhythms* follows. Examples of each of these rhythms can be found in Figure 1.26–6, which provides visual representations of raw waveforms, the spectral (EEG power) bands they occupy, and general functional association. The recognition of these intrinsic rhythms enables the electroencephalographer to interpret EEG records in a clinical context. An in-depth review is beyond the scope of this chapter, and the interested reader is directed to the 2020 comprehensive text of Greenfield, Geyer, and Carney.

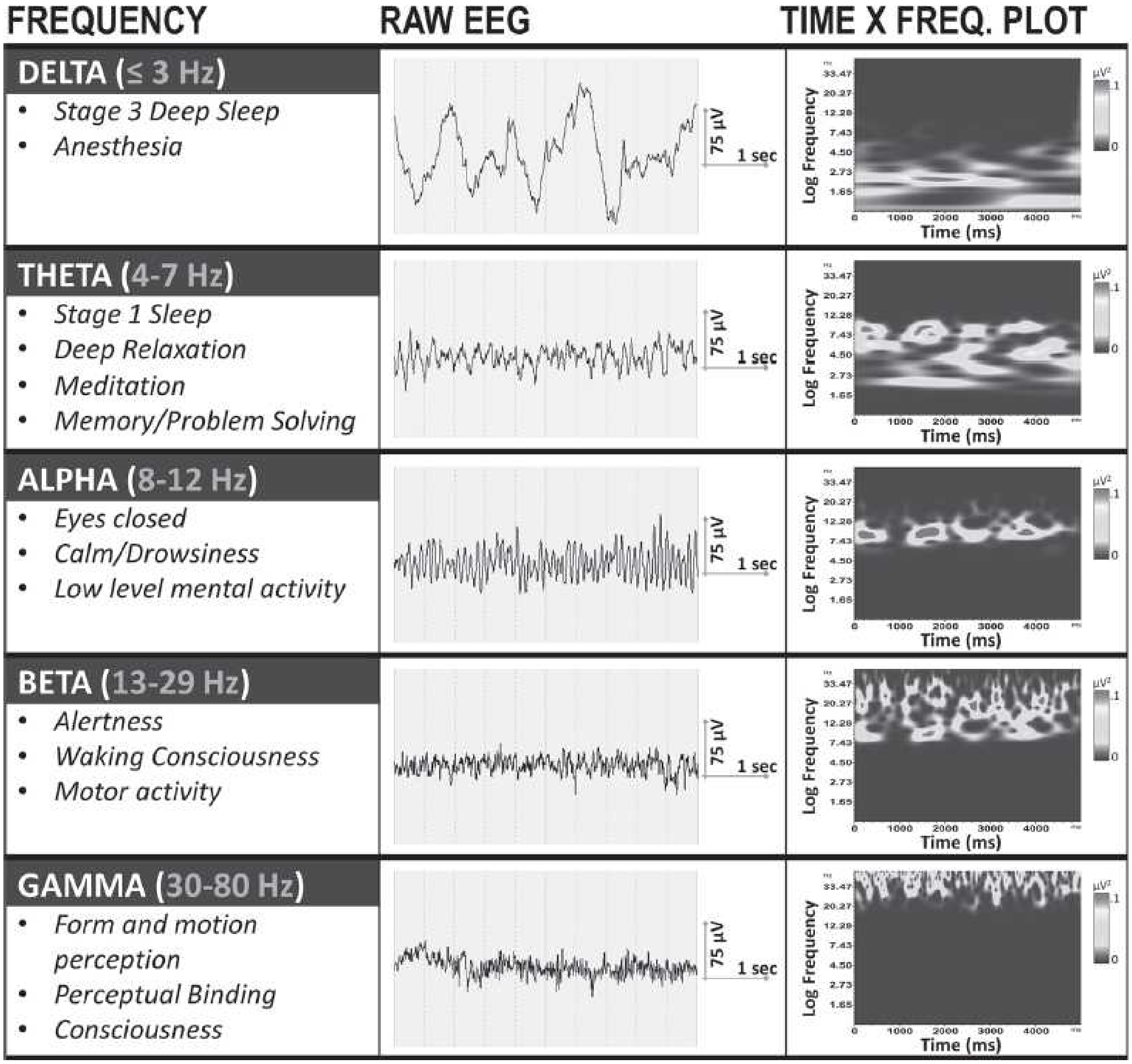

FIGURE 1.26–6. Example of the typical intrinsic brain rhythms observed in the normal EEG. The **middle panel** illustrates a prototypal example of each rhythm in the raw EEG. The **far-right panel** demonstrates how each rhythm would appear after a time x frequency transform.

Delta. Delta activity (≤3 Hz) is not present in the normal waking EEG but is a prominent feature of deeper stages of sleep, particularly stage 3. A slower delta rhythm (<1 Hz) is also observed when an individual is under anesthesia. The delta rhythm has been shown to be primarily generated by thalamocortical circuits. The presence of a significant generalized or focal delta in the waking EEG is strongly indicative of a pathophysiologic process.

Theta. EEG with a frequency of 4 to 7 Hz is referred to as theta activity. A small amount of sporadic, arrhythmic, and isolated theta activity can be seen in many normal waking EEGs, particularly in frontal-temporal regions. While ongoing theta activity is somewhat limited in the waking EEG, it can be a feature of drowsiness during sleep tracing and is highly prominent in stage 1 sleep. Brief periods of theta activity have also been shown to be associated with mental activities such as problem solving and processes related to memory. The fact that theta appears to be generated in the hippocampus supports the role of theta in memory. Theta rhythms are also generated in the neocortex, particularly during working memory tasks. Excessive theta during waking states, generalized or focal in nature, suggests a focal pathologic process such as CBF reduction or metabolic encephalopathies.

Alpha. Highly rhythmic EEG with a frequency range from 8 to 12 Hz is termed alpha activity and was the first brain rhythm to be described by Berger in 1929 (i.e., the Berger rhythm). Alpha constitutes the dominant brain wave frequency of the normal closed-eye waking EEG and is also common in states of relaxation or during periods of minimal mental activity. Hence, alpha has routinely been interpreted as constituting an EEG correlate of an “idling brain.” In middle-aged adults, the vast majority of individuals exhibit an alpha frequency at, or close to, 10 Hz, whereas, with normal geriatric populations, a slower alpha frequency of 8 to 9 Hz is not uncommon. Alpha activity is most prominent over the posterior cortex, particularly the parietal, posterior temporal, and occipital cortex, with the occipital region being best suited to detect this activity. In fact, the neural generators of alpha activity appear to be the visual cortex, with some influence from the visual thalamus. The detection of alpha activity diminishes as one records from more anterior locations, and this frequency is rare at prefrontal electrode sites. Alpha activity is abolished, or at least severely attenuated, by eye-opening. Further, alpha activity also disappears as one enters sleep (indeed, one of the criteria for stage 1 sleep is <50% alpha activity per epoch). Alpha activity can be highly responsive to cognitive activity, such as focused attention or concentration. For example, under closed-eye recording conditions, alpha can be blocked or attenuated by engaging in visual imagery, numeric calculation, or almost anything requiring significant concentration. Alpha frequency can be increased or decreased by a wide variety of pharmacologic, metabolic, or endocrine variables.

Beta. Frequencies in the 13 to 29 Hz range are termed beta waves, and they are one of the dominant frequencies in normal adult waking EEGs, particularly over frontocentral regions. Beta rhythms are typically associated with alertness, and it is not uncommon for beta to appear as runs of rhythmic activity as opposed to sporadic isolated waves. Mixed frequency EEG, including beta, is also observed in REM sleep. In addition, research has shown that beta frequencies are correlated with peripheral EMG activity, suggesting that rhythmic beta in the motor cortex plays a key role in the generation of motor commands. The voltage of beta activity is also almost always of a smaller amplitude than that of the lower frequencies described above. Hence, along with gamma (see below), beta rhythms are not as easily visualized in the raw EEG compared to delta, theta, and alpha (Fig. 1.26–6).

Gamma. One of the most studied EEG frequencies during the past several years is the gamma rhythm, which is typically considered to be activity in the 30 to 80 Hz range (note that some investigators extend gamma up to 100 Hz). Gamma rhythms were largely ignored throughout most of the history of EEG research, which was due to in part to (1) recording methodologies, and (2) the notion that such higher frequencies were nonneural in origin (i.e., noise). However, seminal work in the 1990s by researchers such as Rodolfo Llinás, Andreas Engel, Wolf Singer, and others demonstrated that transient gamma rhythms play a key role in the integration or “binding” together of neural networks in order to generate coherent cortical representations. Since that time, the role of gamma rhythms in human sensation, perception, and cognition has gained wide support. Abnormalities in gamma rhythms are now thought to be a core feature of several psychiatric disorders, most notably schizophrenia.

EEG Artifacts

EEG artifacts, as previously alluded to, refer to any source of recorded activity that is not of cerebral origin and, therefore, contaminates accurate characterization of the desired signal. The various sources of artifacts are unfortunately abundant but can generally be reduced to two sources: *physiologic* and *extraphysiologic.* Physiologic artifacts are real biosignals recorded from the patient but generated by sources other than the brain. Extraphysiologic artifacts are those caused by external sources, including EEG or other medical equipment, and typically electronic devices located on the person or in the recording environment. Familiarity with the appearance of each artifact in the EEG, and its associated source, is critical for the EEG technician as most of these contaminating signals can be mitigated during recording but tend to be much more difficult to remove from the recorded data.

Physiologic Artifacts. Important physiologic features to be aware of during EEG recording and interpretation are ocular artifact, muscle artifact, perspiration, and cardiac artifact, though others (e.g., lingual motion) also exist. Eye blinks and vertical shifts in eye movement can be prominently observed in the EEG record, identified by rapid changes of voltage over brief intervals over frontal electrode sites, particularly the frontal poles (Fig. 1.26–7). Blinks and *vertical electrooculogram* (VEOG) responses can be further captured in EEG recoding by affixing electrodes above and below the orbit of one eye as a bipolar channel. Similarly, horizontal eye movement, or “saccades,” produces a characteristic artifact that appears as a change in voltage that tends to flatten over a brief period of time, with phase reversal observed between left- and right-sided frontal electrode channels (the side of positive deflection indicates the direction of eye movement). Like VEOG, *horizontal electrooculogram* (HEOG) responses are detected in scalp EEG channels (particularly F7 and F8), but optimal characterization benefits from a bipolar channel of electrodes placed near the outer canthi of the eyes. Fortunately, several “blink correction” algorithms have been developed and can be employed in digital EEG data to remove VEOG/HEOG artifacts from the EEG record (Fig. 1.26–8).

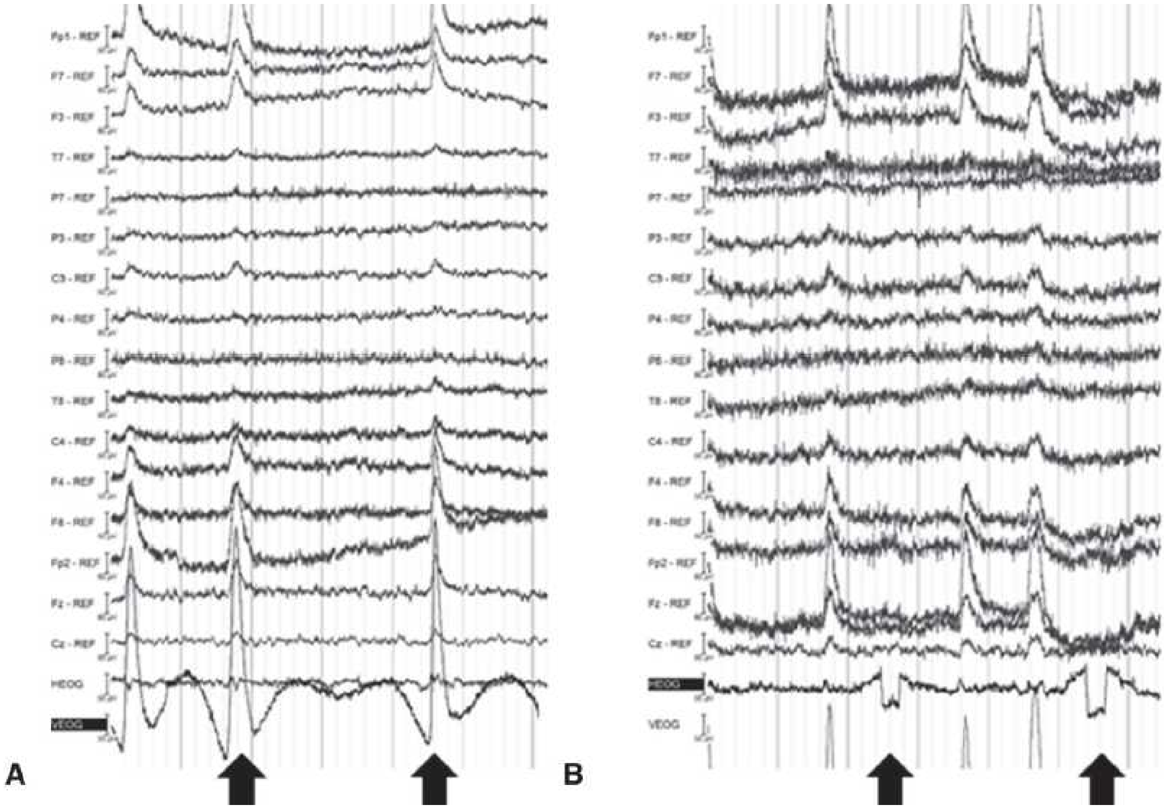

FIGURE 1.26–7. Electrooculographic (EOG) artifact. **Panel A:** Vertical EOG (VEOG) activity caused by eye blinks. VEOG recorded from electrodes affixed above and below the orbit of the right eye, also evident in frontal poles (Fp1, Fp2) and adjacent frontal sites with attenuated signal distributed posteriorly. **Panel B:** Horizontal EOG (HEOG) caused by lateral shifts in gaze, recorded at electrodes affixed bilaterally at outer canthi of eyes.

EMG activity of the temporalis muscle, indicating jaw clenching or chewing, and of the frontalis muscle, indicating tension and contraction of the forehead, appear as a high-frequency artifact that can spread across a large portion of scalp electrodes (Fig. 1.26–9). It is more difficult to correct EEG for this type of EMG artifact offline and could result in rejection of large segments of data if prominent. As mentioned previously, a low-pass filter may be used to attenuate EMG artifacts, but the use of this option may equally affect frequencies of interest. Therefore, it is best to reduce this form of artifact during EEG recording by training the patient to relax facial muscles beforehand and stopping recording, as necessary, to remind them. Obviously, anything that contributes to the use of these muscles, such as squinting to read in a dark room, chewing gum, or talking, should be avoided during EEG recording.

Perspiration artifacts (galvanic skin response) can be detected by a slow, undulating drift in signal over long durations (e.g., 2 seconds), thus being easy to differentiate from cerebrally generated activity. However, it also resembles slow roving eye motion. Perspiration artifact is caused by sweat (sodium chloride) interacting with electrolytic gel and electrode metals and tends to be most prominent over frontal electrode sites (Fig. 1.26–10). Properly temperature-regulated recording environments and instructions to patients to dress in layers that can be removed as needed are the best solutions for avoiding perspiration, which can be difficult to reduce once it begins. Less frequently seen are artifacts from an electrocardiogram (ECG), though certain physical characteristics of the patient, such as a short neck or heavy, “barrel-chested” build, tend to increase their presence. ECG artifacts reflect both electrical and mechanical actions of the heart and can be detected by a consistent, rhythmic pattern with a period that aligns closely to pulse rate (thus, near 1 Hz) and are observed most commonly in temporal electrode sites and references placed at ears or mastoid (when electrodes placed too low and nearer to the carotid artery). Although ECG artifact is not likely to directly affect many scalp EEG channels, the intrusion of ECG into reference channels could give this appearance during data acquisition using a reference montage or after data has been rereferenced offline. These problems are then best avoided by moving reference channels, when possible, to a location where the ECG signal is not picked up.

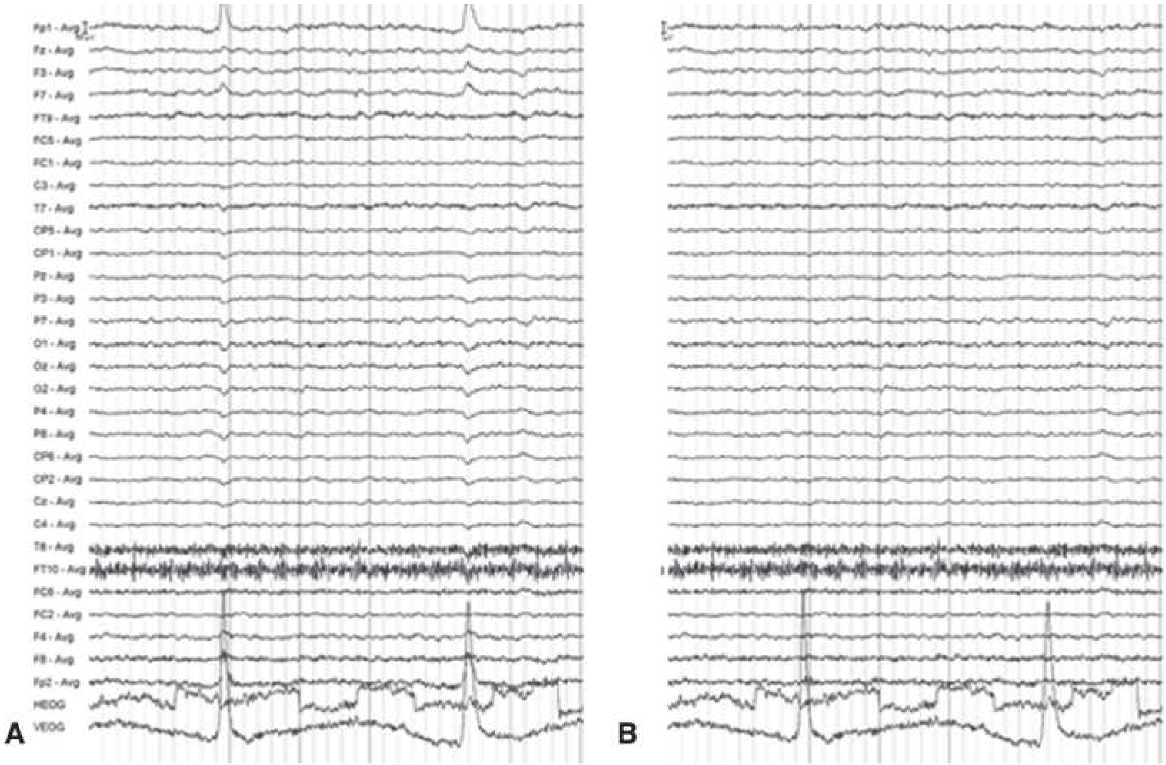
Fp1 - Avg
Fz - Avg
F3 - Avg
F7 - Avg
FT9 - Avg
FC5 - Avg
FC1 - Avg
C3 - Avg
T7 - Avg
CP5 - Avg
CP1 - Avg
Pz - Avg
P3 - Avg
P7 - Avg
O1 - Avg
Oz - Avg
O2 - Avg
P4 - Avg
P8 - Avg
CP6 - Avg
CP2 - Avg
Cz - Avg
C4 - Avg
T8 - Avg
FT10 - Avg
FC6 - Avg
FC2 - Avg
F4 - Avg
F8 - Avg
Fp2 - Avg
HEOG
VEOG
A
B

FIGURE 1.26–8. Ocular correction. **Panel A:** EEG shows ocular motion artifact on VEOG and HEOG channels with propagation to scalp electrodes. **Panel B** shows the same data following offline ocular correction by Gratton & Coles method. Note that correction is applied only to scalp electrodes in this case, leaving the original artifact present in VEOG and HEOG.

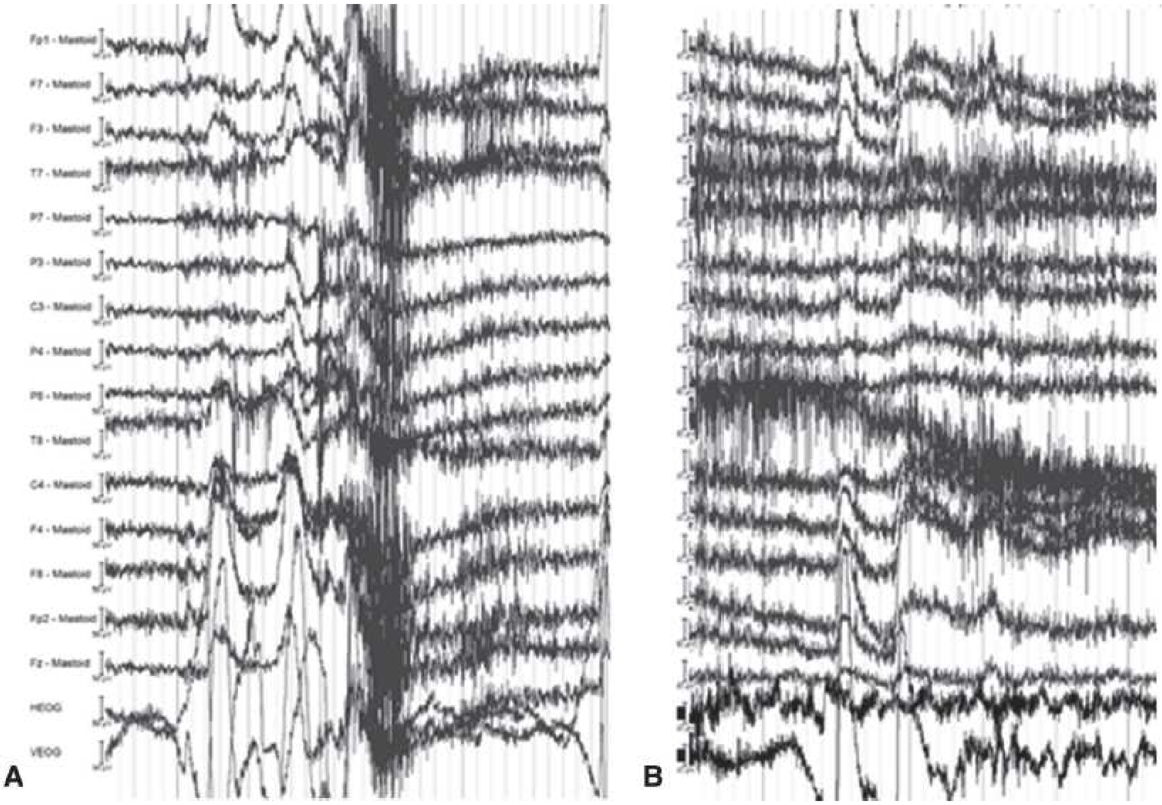
Fp1 - Mastoid
F7 - Mastoid
F3 - Mastoid
T7 - Mastoid
P7 - Mastoid
P3 - Mastoid
C3 - Mastoid
P4 - Mastoid
P8 - Mastoid
T8 - Mastoid
C4 - Mastoid
F4 - Mastoid
F8 - Mastoid
Fp2 - Mastoid
Fz - Mastoid
HEOG
VEOG
A
B

FIGURE 1.26–9. Electromyographic artifact. **Panel A:** Jaw (temporalis) muscle tension. **Panel B:** Forehead (frontalis) muscle tension. These artifacts can typically be ameliorated by asking the subject to relax facial muscles when observed.

Extraphysiologic Artifacts. The main sources of extraphysiologic artifacts have to do with equipment and environmental issues that can be minimized through careful attention to proper electrode setup, cleaning and maintenance of equipment between uses, and mitigation of ambient electromagnetic noise sources in the recording room. Electrical devices such as computer monitors, lights, and poorly shielded power cables emit a 60 Hz signal, the line frequency of AC delivery in the United States (50 Hz in most other parts of the world). Electrodes with high or unstable impedance due to poor contact with gel/scalp or break in the lead will be particularly susceptible to picking up ambient noise in the recording environment, which is identified by persistent 60 Hz activity that appears as a fast sine wave imposed over background EEG (Fig. 1.26–11). Some EEG acquisition software packages enable viewing EEG in spectral frequency mode, which can quickly identify the characteristic peak at 60 Hz if there is any doubt.

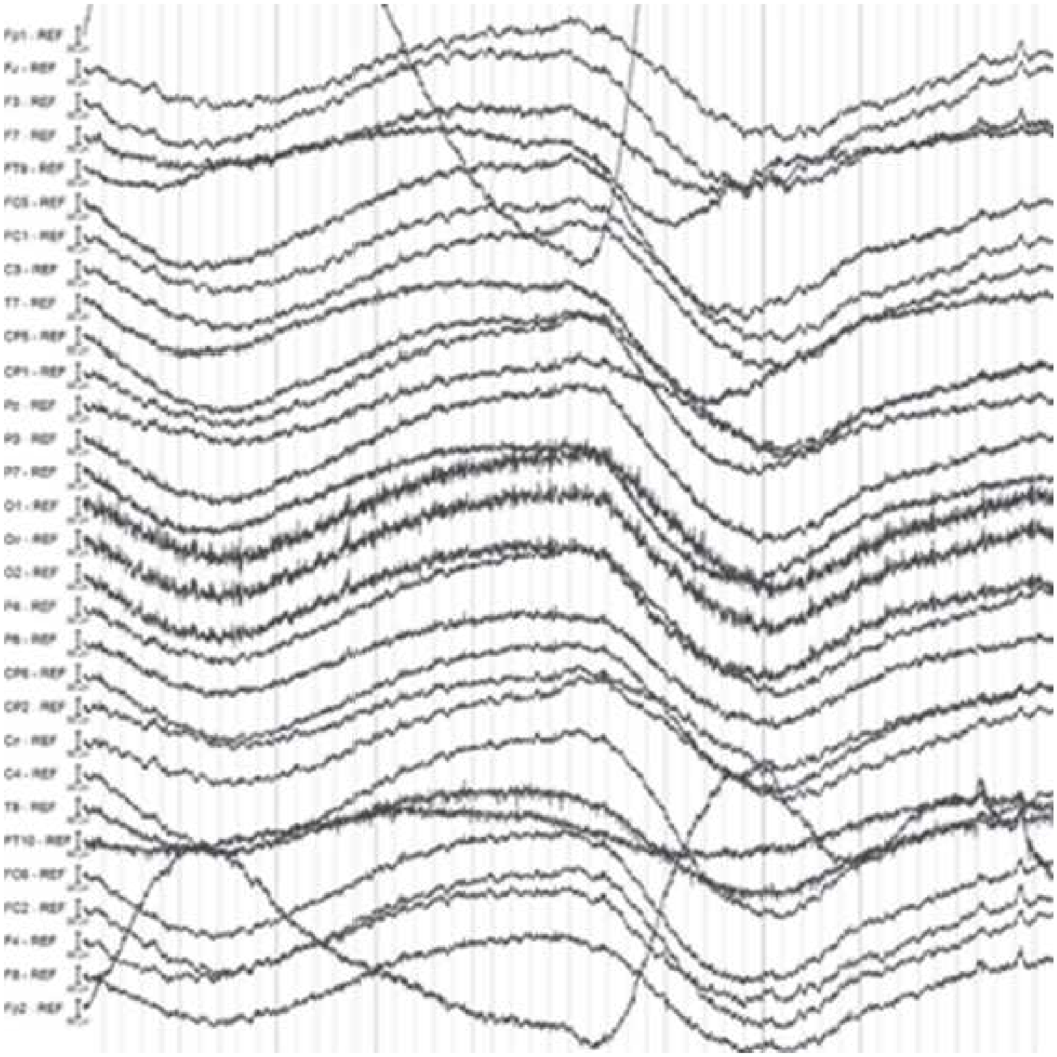
Fp1 - REF
Fz - REF
F3 - REF
F7 - REF
FT9 - REF
FC5 - REF
FC1 - REF
C3 - REF
T7 - REF
CP5 - REF
CP1 - REF
Pz - REF
P3 - REF
P7 - REF
O1 - REF
Oz - REF
O2 - REF
P4 - REF
P8 - REF
CP6 - REF
CP2 - REF
Cz - REF
C4 - REF
T8 - REF
FT10 - REF
FC6 - REF
FC2 - REF
F4 - REF
F8 - REF
Fp2 - REF

FIGURE 1.26–10. Sweat drift artifact in EEG record referenced to the mastoid. Drift at frontal poles (electrode Fp1, Fp2) suggests sweat on the forehead. Drift common across remaining scalp electrodes suggests sweat at mastoid reference electrodes.

Maintaining proper electrode contact over long EEG recordings may require reapplication of gel, which can always be a first troubleshooting step for addressing noisy channels. Detection of a connection problem, which may also be caused by electrode movement, is not difficult as the resulting effect on the recorded signal in no way resembles true EEG (Fig. 1.26–12). In these cases, there tend to be very large changes of voltage, which may “rail” out of range at gain settings used for regular EEG monitoring, with no regularity in the signal.

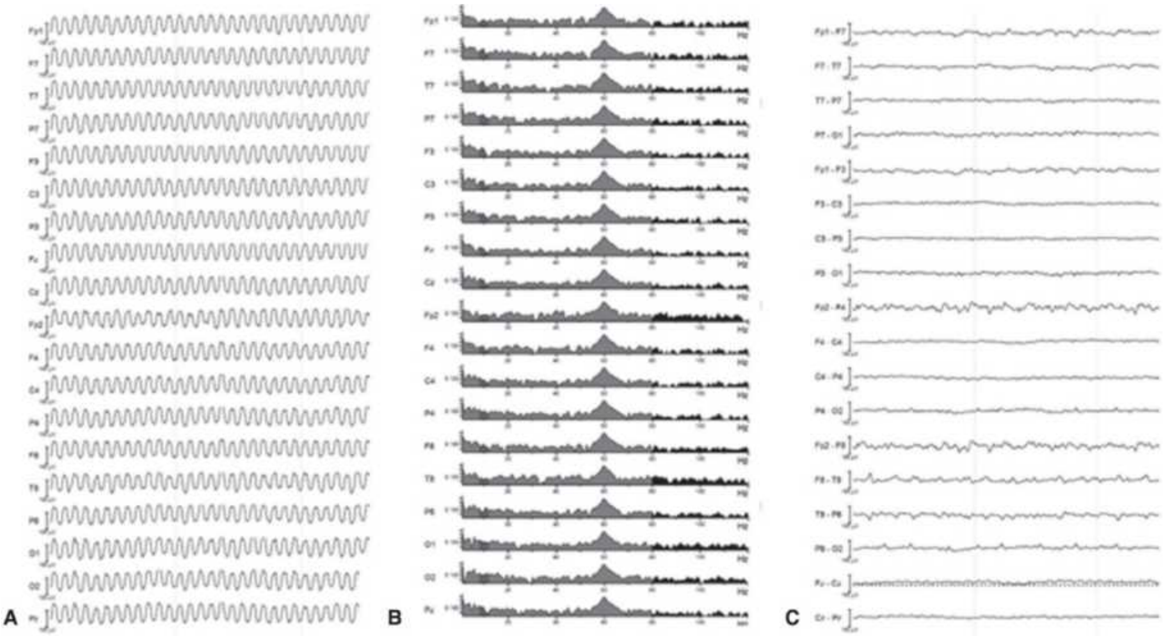

FIGURE 1.26–11. Panel A: 60 Hz ambient noise in unreferenced EEG viewed at a range of 100 μV. **Panel B:** Power spectral analysis by Fast Fourier Transform (FFT) depicting prominent 60 Hz activity in **Panel A** segment. **Panel C: Panel A** EEG segment viewed in bipolar reference montage at 100 μV. Note that 60 Hz activity remains present in EEG but cancels out as a common signal when referenced.

One common artifact that is more difficult to prevent in the electrode is a "pop," appearing as a steep and rapid deflection in voltage, usually isolated to a single electrode (unless a reference or a channel is included in a bipolar montage) and occurrence (or may reoccur a few times sporadically), that is caused by the discharge of electrical potential between the scalp and electrode where electrical charges can store (as if in a capacitor). The pop is the only extraphysiologic artifact mentioned here (though this list is not exhaustive) that may resemble abnormal EEG activity, with some characteristics similar to epileptiform activity. Importantly, pops can be distinguished from true neurophysiologic paroxysmal activity by the fact that they tend to occur in only one electrode at a time, thus do not give the appearance of voltage change from an electrical field source, and appear with an asymmetric morphology characterized by relatively steep rise and slow recovery. These characteristics should differentiate pops from interictal spikes to the trained eye.

EEG Quantification and Feature Detection

EEG recording, combined with stimulus-evoked paradigms and experimental measures of mental task performance, provides a powerful methodology for studying neural dynamics of human brain function and pathology. EEG data is dimensional and complex, based on a time series of events sampled with high temporal resolution (i.e., millisecond level) and distributed spatially across multiple scalp locations (e.g., montages of 32 to 256 channels). If one considers a sampling rate of 256 samples per second, across 32 scalp channels, for 10 minutes of recording, each parameter representing a lower bound of what is possible for digital EEG recording and typical for a recording session (which may last hours), the resulting data matrix could conceivably include 4,915,200 discrete data points. While examples of clinical EEG discussed thus far would not entail a detailed analysis of each sample point by the electrode (this would never be done under any circumstance), with objectives that are routinely achieved by visual inspection of data over time scales of several seconds, such abundant data no doubt contains important features that cannot be ascertained (or even seen) through visual inspection and clinical judgment alone. The advent of digital EEG recording technology, among other advantages over analog paper EEG, makes highly detailed EEG analysis possible through computerized signal processing. Basic processes such as *segmentation,* dividing the time series record into shorter periods or epochs based on experimental events, and *averaging* over multiple segments (consistencies in signal average together while noise averages out) provide efficient means of isolating important features in data, upon which a number of additional signal extraction methods can be applied. Digital EEG analysis is broadly referred to as "*Quantitative EEG*," abbreviated to *Q-EEG* or *qEEG,* emphasizing measurement of the magnitude and differentiation of features contained within the EEG rather than qualitative decisions over whether or not an observed feature indicates a pathophysiologic process. However, even while the heavy computational burden is now passed on to the machine, Q-EEG still requires the same human judgments concerning data quality and interpretation. When the issue of artifact contamination of the Q-EEG is not properly addressed, one has a situation described a generation ago by computer programmers, namely, "garbage in, garbage out," leading to erroneous conclusions about the processed data. Software used for Q-EEG analysis includes automatic or semiautomatic functions for removal of artifact, based on thresholds for acceptable amplitude or change in amplitude over time, which can be used to mark and remove some sources of EEG contamination. However, these *artifact-rejection* procedures cannot typically mitigate all problematic conditions, such as the low-frequency activity of a drowsy patient, and visual data inspection should always be undertaken as the first step of the analysis.

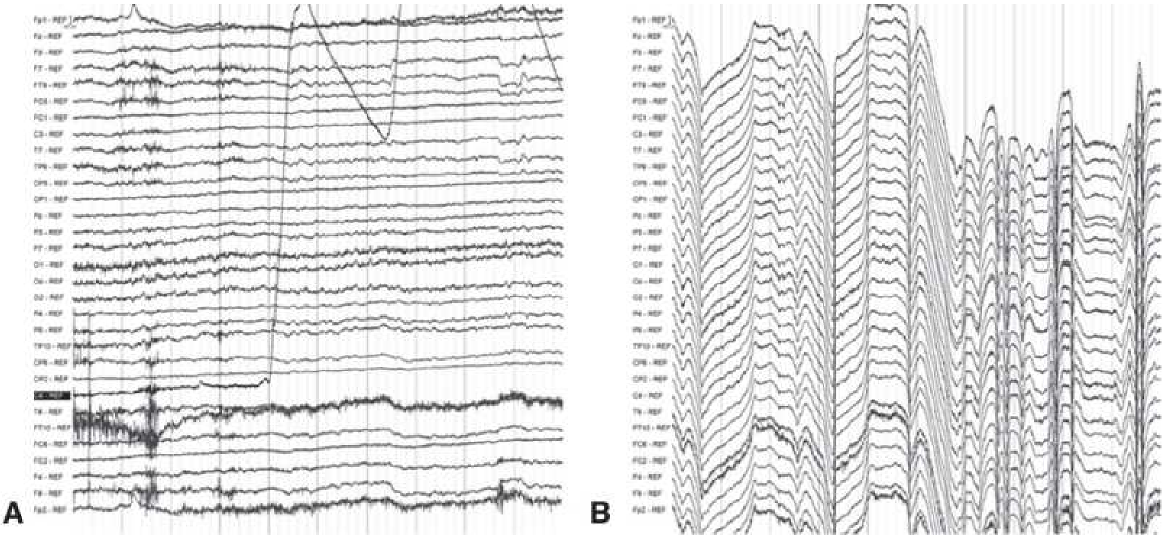

FIGURE 1.26–12. Physical electrode movement in a referential montage. **Panel A** shows the effect at a single electrode (C4), suggesting the electrode is being physically shifted in position and unseated from the point of scalp contact. **Panel B** shows a consistent pattern of shift across all electrodes, indicating movement at the reference electrode site.

Beyond the initial data cleaning, the technician or experimenter is faced with many more technical decisions regarding the selection of time points and signal extraction methods used to best characterize the neurophysiologic phenomena under examination. It is sensible, if not required, for some processing steps to be taken before others, but each involves a number of decisions over parameters for which there is no common agreement or standardization. Furthermore, there is usually more than one way to implement a given function, which may differ between software packages, and the user (typically lacking extensive training in computer science and engineering) may be unaware of the intricacies and assumptions involved in each. There are also a number of free, open-source toolkits available for EEG analysis that enable the inclined user to customize various analysis routines. Therefore, while digital recording and analysis greatly increase the capacity to explore and extract features from EEG, the application of such methods remains to be more of an exquisite art than a standardized practice.

The most commonly used method for EEG quantification is spectral analysis, principally based on the *Fourier transform.* The Fourier transform decomposes a signal represented as a function of time into its constituent frequency components. Given the analogy of a chord played in music, playing the chord as an arpeggio (e.g., the A major chord played as individual notes A-C#-E) enables the listener to appreciate the frequency and loudness of each note as a component of the chord's overall harmonic quality. In a similar way, the Fourier transform represents a signal as a measure of *spectral power* (square of the signal amplitude) according to the *total power* of the broadband (e.g., chord) and *band power* of individual frequencies (e.g., notes). In EEG signal processing, the Fourier transform is most commonly implemented as the *Fast Fourier Transform* (FFT), which applies an efficient algorithm for computing the Fourier transform achieved by reducing the total number of samples submitted to computation by their base 2 logarithms. Applying FFT, therefore, assumes that the data represents a fixed-length waveform with a total number of samples that can be expressed as a factor of 2-to-the-nth power (e.g., 64, 128, 256, 512). The frequency resolution of the FFT is also dependent on the length of the data sampled, with longer segments of data needed to resolve lower frequencies. For example, the lowest frequency that could be characterized accurately using a segment length of 1 second is 1 Hz, while resolving a 0.1 Hz signal would require a segment length of 10 seconds. When using FFT, it is also important to keep in mind that the analysis treats all samples within the segment equally without regard to time; therefore, it is insensitive to time-varying functions and most appropriate for signals assumed to be stationary. Quantification of changes in spectral power that occur over a period of time is made possible through *time–frequency decomposition.* A variation on the FFT, called the *short-time Fourier transform* (STFT), has been developed for this purpose, but EEG analyses typically employ more sophisticated algorithms based on the *wavelet* transform. The theory and mathematical computation involved in the wavelet transform are beyond the scope of this chapter, but an eloquent description written in the context of EEG analysis can be found in the methods handbook of Cong, Ristaniemi, and Lyytinen (2015). The basic point to keep in mind in differentiating FFT, or time to frequency transformation, and wavelet, or time–frequency transformation, is that frequency extraction by FFT applies to the entire time window of analysis (signal is not localized in time) while wavelet analysis provides localization of frequency extraction to each sample of the time window. The latter is advantageous when examining change in frequency component structure as signals evolve over time and so tends to be the more appropriate analysis for perceptual–cognitive experiments. That said, there are still cases where FFT may be preferred, given that time resolution does come at a cost to frequency resolution.

In both the clinical and research literature, several types of time–frequency outcome measures are often discussed, which vary depending on the paradigms used to elicit them and the signal processing techniques utilized. Most commonly, oscillatory activity measured by time–frequency decomposition is differentiated into three categories: transient evoked, steady-state evoked, and induced (Fig. 1.26–13).

Q-EEG using frequency-based analyses has contributed largely to the appreciation of EEG as an ensemble of neurophysiology-generated oscillatory components with unique functional properties. For instance, it has been suggested that the delta rhythm, traditionally regarded as a pathologic finding in the awake EEG, has been related to internal concentration (i.e., conditions in which subjects have to disregard external input). Further examples include the association of frontally generated theta activity with encoding and working memory functions, the slow alpha band (7.5 to 9.5 Hz) with attentional processes, the fast alpha band (9.5 to 13.5 Hz) with memory retrieval, and beta activity with the integration of sensory information. Furthermore, functions associated with frequencies above 30 Hz, the so-called gamma band, have been more recently discovered and are currently a major focus in psychiatric research. The demonstration of a stimulus-induced gamma oscillatory activity in the mammalian brain was the first empirical evidence that synchronization of neural groups (individually coding for different aspects of the sensory input) serves a basic “binding” process in the central nervous system (CNS), enabling information communicated over sensory afferents to combine in forming a unified percept. While high-frequency gamma synchronization is thought to support communication of short-range local networks, slower oscillatory rhythms have been implicated in cognitive processes requiring integration over large cortical distances (e.g., among areas of different sensory modalities) that extend over longer periods of time (e.g., memory encoding).

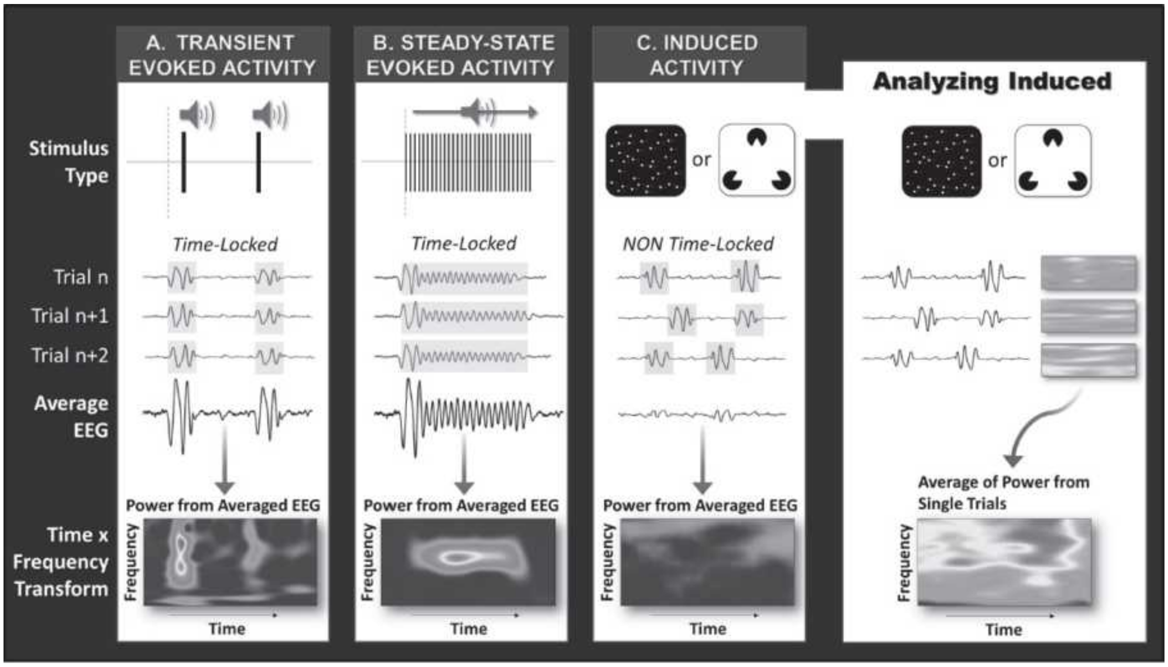
A. TRANSIENT EVOKED ACTIVITY
B. STEADY-STATE EVOKED ACTIVITY
C. INDUCED ACTIVITY
Analyzing Induced
Stimulus Type
or
or
Time-Locked
Time-Locked
NON Time-Locked
Trial n
Trial n+1
Trial n+2
Average EEG
Power from Averaged EEG
Power from Averaged EEG
Power from Averaged EEG
Average of Power from Single Trials
Time x Frequency Transform
Frequency
Time

FIGURE 1.26–13. Transient and steady-stated–evoked activity are phase- and time-locked oscillations (**A** and **B**) elicited by discrete or steady-state sensory stimuli. For both of these types of activity, averaging across trials has the effect of canceling out nonphase-locked oscillations (due to destructive interference), leaving the evoked response intact and magnified. Hence, time–frequency analysis can be performed on the averaged EEG (i.e., after averaging all the individual trials). By contrast, induced activity is typically elicited during higher perceptual processing, for example, during the perception of coherent motion or of illusory shapes (see **Panel C**). These responses are often nontime locked (jittered across time from trial to trial). Hence, averaging the trials cancels out the response of interest due to destructive interference and is not visible in the time–frequency transform (bottom of **Panel C**, left). However, these responses can be detected if the time–frequency analysis is performed on the single trials, and the time–frequency transforms of individual trials are then averaged (**Panel C**, right). (Adapted from Skosnik PD, Cortes-Briones JA. Electroencephalography and cannabis: from event-related potentials to oscillations. In: Preedy VR, ed. *The Neuropathology of Drug Addictions and Substance Misuse*. Academic Press; 2016:851–862, with permission from Elsevier.) (See color panel.)

Time- and frequency-based analyses mentioned thus far (Figs. 1.26–6 and 1.26–13) have paved the wave for more precise characterization of neural activity and function. However, they may fall short in providing a fully accurate representation of neural communication in that the information extracted by these methods remains spatially localized (e.g., analyzed based on one electrode or field of electrodes). Furthermore, while these methods effectively quantify the magnitude of neural activity as an average representation of time, the degree of consistency in activation across time segments is lost. Relationships between frequency ensembles located at different spatial locations or occurring at different time points, determined as a measure of synchronization, can be ascertained using *coherence* measures. Coherence analysis follows frequency (e.g., FFT) and time–frequency (e.g., wavelet) decomposition of the signal and is applied to a complex form of these data as a measure of the strength of association in the linear dependence of two signals. Coherence analysis can be applied to stationery and time-series data as a measure of the relationship of spectral power between two regions (i.e., *spatial coherence*), the dependence of one frequency on another (i.e., *cross-spectrum coherence*), and of consistency in the phasic cycle of oscillatory activity over repeated trials (i.e., *intertrial coherence*). Taken together, the combination of spectral power and coherence measures greatly enhances the ascertainment of neural dynamics comprising the human EEG and increases the capability to understand the functional interactions of brain areas on a large spatial scale.

Q-EEG Topographic Analysis and Source Analysis

Computerized EEG topography (CET) or Q-EEG mapping was developed for multichannel Q-EEG data. The technique enables the construction of a two- or three-dimensional matrix for a topographic representation of Q-EEG parameters, such as instant amplitude or band power. Matrix points that do not correspond to recording electrodes are calculated by interpolating the values of the nearest three or four recording electrodes. As for other imaging techniques, statistical probability maps can also be implemented for comparisons between different subject populations or experimental conditions. Color-coded maps make the multichannel information more immediately intelligible; however, the topographic information provided by the technique is critically dependent on the reference electrode used for EEG recording. To address this issue, several transformations have been proposed that provide reference-independent data, including average-reference (i.e., subtraction of the mean of all electrodes from each electrode value) and surface Laplacian or CSD approaches previously mentioned. However, it is imperative to realize that the topography at the scalp, even when resulting from reference-independent calculations, does NOT allow physiologic inferences on the underlying brain generators due to the so-called inverse solution problem. Essentially, this problem appreciates the imprecision of localizing sources of field potentials as recorded at the level of the scalp since numerous neural activation patterns can generate the same scalp topographies.

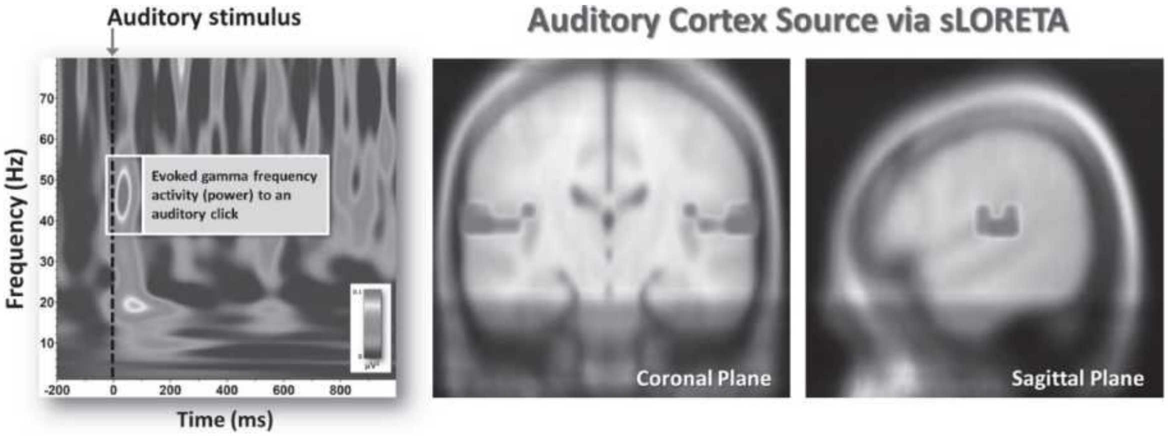
Auditory stimulus
Auditory Cortex Source via sLORETA
Frequency (Hz)
70
60
50
40
30
20
10
-200
0
200
400
600
800
Time (ms)
Evoked gamma frequency activity (power) to an auditory click
μV²
Coronal Plane
Sagittal Plane

FIGURE 1.26–14. Example of EEG source analysis using sLORETA. The time–frequency transform on the left indicates a burst of transient gamma activity to an auditory click stimulus, which was localized to the primary auditory cortex (right).

Thus, in 1989 Frank Duffy, Dietrich Lehman, Fernando Lopes da Silva, and other eminent scientists formed the International Society for Brain Electromagnetic Topography (ISBET), established to promote the application and further development of techniques investigating brain electromagnetic activity, topography, and source analysis. In recent years, different algorithms have been proposed to solve the inverse problem. They can be divided into *equivalent current dipole models* and *current distributed source models.* According to recent comparative studies, the dipole models are suitable only when a single source is expected. Among the distributed source methods, *low-resolution brain electromagnetic tomography* (LORETA), or more recently, *standardized low-resolution brain electromagnetic tomography* (sLORETA), have proven greater effectiveness in terms of localization error, particularly when multiple sources and noise are present. Briefly, the sLORETA method is a properly standardized discrete, three-dimensional distributed, linear, minimum norm inverse solution. The particular form of standardization used in sLORETA endows the tomography with the property of exact localization to test point sources, yielding images of standardized current density with exact localization, albeit with a low spatial resolution (i.e., neighboring neuronal sources will be highly correlated). sLORETA is currently a method of choice since it has no localization bias, even in the presence of measurement and biologic noise. In the current implementation of sLORETA, computations are made in a realistic head model using the MNI152 template with the three-dimensional solution space restricted to cortical gray matter, as determined by the probabilistic Talairach atlas. Thus, sLORETA images represent the standardized electric activity at each voxel in neuroanatomic Montreal Neurological Institute (MNI) space as the exact magnitude of the estimated current density (Fig. 1.26–14). Several studies have reported consistency between sLORETA and other neuroimaging methods such as fMRI. The LORETA/sLORETA software is based on the pioneering work of Roberto Pascual-Marqui and his group, and relevant references can be freely downloaded (http://www.uzh.ch/keyinst/loreta.htm). It should be noted, however, that methods/models of source localization, analysis steps, and relevant software packages continue to evolve.

Status of Q-EEG in Applied Clinical Practice

Despite advances in EEG signal extraction enabled by digital technology, there are few specific conditions for which Q-EEG has gained acceptance as a clinical diagnostic tool, and in these cases, it is used only to enhance the interpretation of conventional clinical EEG. According to the AETNA Health Insurance list of approved indications (last reviewed on April 15, 2021), Q-EEG is medically necessary only as an adjunct to traditional EEG for the following conditions (from http://www.aetna.com/cpb/medical/data/200_299/0221.html, accessed November 17, 2021):

1. For the ambulatory recording of EEG to facilitate subsequent expert visual EEG interpretation; *or*
2. For continuous EEG monitoring by frequency trending to detect early, acute intracranial complications in the operating room or ICU; *or*
3. For evaluation of certain members with symptoms of cerebrovascular disease whose neuroimaging and routine EEG studies are not conclusive; *or*

4. For evaluation of dementia and encephalopathy when the diagnosis remains unresolved after initial clinical evaluation; *or*
5. For screening for possible epileptic seizures in high-risk ICU members; *or*
6. For screening for possible epileptic spikes or seizures in long-term EEG monitoring; *or*
7. For topographic voltage and dipole analysis in presurgical evaluations for intractable epilepsy.

The list of indications described is no more inclusive than those recommended nearly 20 years prior by the ACNS and American Academy of Neurology (AAN) Executive Board's publication, appearing in the journal *Neurology* in 1997. Also consistent with the conclusions of the ACNS/AAN report, AETNA considers the use of Q-EEG "experimental and investigational" for all other indications due to the lack of adequate scientific evidence to support its clinical utility. Psychiatric conditions listed explicitly among unsupported indications include alcohol and substance use disorders, anxiety, Asperger syndrome and other autism spectrum disorders, attention deficit disorders, bipolar disorder and major depression, panic disorder, PTSD, and schizophrenia, in addition to cases of acquired injury involving mild or moderate head injury, sports concussion, and tinnitus, among others. Furthermore, despite favorable reports in the research literature, Q-EEG has yet to receive approval for use in predicting the efficacy of psychotropic medication.

EEG IN APPLIED CLINICAL PRACTICE

Despite the fact that EEG originated in psychiatry, the strongest initial impetus for its use came from neurology, particularly the study of epilepsy. The years from 1934 to 1940 saw a marked proliferation of EEG studies focused on structural brain lesions and a variety of seizure disorders. In 1934, the team led by Fred Gibbs discovered the classic three-per-second spike-and-wave complex, which proved to be specific to petit mal absence attacks. Before the decade ended, they had described EEG patterns associated with grand mal and myoclonic seizures and a diffuse spike-and-wave pattern that was slower in frequency than petit mal (and given the confusing name of petit mal variant) and that was associated with grand mal seizures and a high incidence of mental retardation. They also introduced the term *psychomotor seizures* (now *complex partial seizures*

[CPSs]) and described the EEG manifestations characterizing an ictal psychomotor attack. Later, in 1947 and 1948, they described the anterior temporal spike focus that became the interictal EEG correlate of this disorder. On the other side of the neurologic coin, the study of structural brain lesions was advanced through landmark EEG discoveries of the 1930s to 1950s. In 1935, Otfried Foerster and Helmut Altenberger reported that focal slow waves in the EEG recording often appeared near brain tumors, and 1 year later, Grey Walter made a major advance by demonstrating a technique for EEG localization of brain tumors. Denis Williams at Oxford subsequently began using EEG recordings to study and localize traumatic intracranial injuries received during World War II. Finally, through the collaboration of neurosurgeon Wilder Penfield and psychologist (turned neurophysiology researcher) Herbert Jasper, who brought EEG to the Montreal Neurologic Institute in 1937, a technique for direct intracranial EEG recording termed *electrocorticography* (ECoG) was developed. ECoG was applied to the detection of epileptogenic zones for surgical resection as part of the groundbreaking *Montreal procedure* in the 1950s.

Normal versus Abnormal: General Considerations

One of the factors that may have limited the perception of EEG as a useful assessment tool in psychiatry is the simplistic attempt to conceptualize EEG findings as a normal-versus-abnormal dichotomy. Psychiatric practitioners commonly ask if their patient’s EEG is abnormal and seem content with an affirmative or negative answer. The expectation was that if the EEG was abnormal, then this meant that the patient belonged to neurology, and if normal, that would confirm the psychiatric status of the patient. Furthermore, much of the published research literature reports the percentage of abnormal EEG findings in a given study population without stating the exact nature of the abnormalities. In actuality, the simple normal–abnormal dichotomy is inconsistent with the dynamic nature of EEG, particularly when interpreted by conventional visual inspection without the use of more objective means of quantification that could equate EEG to more common and reliably discernable measures of the physiologic function used in medicine (e.g., blood pressure). Rather than being considered as a dichotomy, the range of EEG findings exists on a broad continuum anchored on one end by unequivocally normal EEGs and extending in the other direction to features of possible and, eventually, more definite clinical significance. Some EEG patterns that are correctly classified as abnormal are on the low end of the continuum of clinical expressivity and may not always contribute strongly to diagnostic decisions. The psychiatrist receiving a report of a minor finding of limited clinical relevance may understandably begin to question the value of EEG when no current or past history of organic signs are detected. The bottom line is that one should not ask simply if an EEG is abnormal, as if it were one side of a dichotomy, but instead should ask what specific features were present and what, if any, known clinical correlates may be associated with that feature. Even when questions remain about the clinical significance of an abnormal feature, this finding should not simply be disregarded and, at least, could indicate a need for follow-up EEG, additional brain imaging, neuropsychological assessment, review of medications or illicit drugs used, or other laboratory measures of cardiac and metabolic function.

In addition to the confusion caused by the reliance on the normal–abnormal dichotomy, literature reports stating that 10% to 20% of the normal population have abnormal EEG findings could lend uncertainty regarding the value of the EEG in psychiatric practices. Because such statements are almost never properly clarified, they tend to be terribly misleading. Clearly, there can be an understandable reluctance to follow up on an abnormal EEG report if one has read in a presumably authoritative source that 20% of the normal population have abnormal EEG findings, decreasing the impression of their diagnostic specificity and, logically, clinical utility. More importantly, there also may be an equally understandable reluctance to refer a patient for an EEG study for the same reason. To place the issue in proper perspective, the reports of substantial percentages of abnormal EEGs within the normal population are heavily skewed by the inclusion of minor findings that often have low levels of clinical or diagnostic relevance. A comprehensive review of the literature on EEG findings in normal control populations by Nash Boutros and colleagues highlights the fact that most control studies have been seriously compromised by failure to screen subjects for a variety of medical and psychiatric variables that can impact EEG results. This group had previously reanalyzed data from numerous published EEG control population studies. When borderline EEG findings were removed from the data, the incidence of EEG abnormality dropped to 3.2% of 6,182 healthy adults and 3.5% of 1,450 healthy children, with the prevalence of accepted paroxysmal EEG abnormalities reduced to only 1.14% of a sample of 11,560 mixed-age healthy subjects.

Variants of Normal EEG

Sleep Patterns. The EEG patterns that characterize drowsy and sleep states are different from the patterns seen during wake. A detailed accounting of the nuances of sleep EEG is beyond the scope of this chapter. Briefly, the rhythmic posterior alpha activity of the closed-eye waking state subsides during progressive drowsiness and is replaced by irregular low-voltage theta activity and sporadic vertex sharp waves, which indicate entry into stage 1 sleep. The progression into stage 2 sleep is marked by the appearance of 12- to 14-Hz sleep spindles (also called sigma waves) and k complexes. These patterns eventually give way to high-voltage (>75 µV) delta waves, which indicate stage 3 deep sleep. Lastly, REM sleep is characterized by low-voltage, mixed frequencies that resemble waking EEG. Some individuals may also exhibit low-amplitude sawtooth waves in REM sleep.

Changes with Age. The appearance of the EEG changes dramatically from birth to advanced age. From a preponderance of irregular medium- to high-voltage delta activity in the tracing of the infant, EEG activity gradually increases in frequency and becomes more rhythmic with increasing age. Rhythmic activity in the upper theta–lower alpha range (7 to 8 Hz) can be seen in posterior areas by early childhood, and, by the time mid-adolescence is reached, the EEG essentially has the appearance of an adult tracing. The interpretation of EEGs obtained from children demands a solid grounding in the age-related changes during this period and is best performed by one specializing in pediatric EEG. For the sake of the current chapter, it is sufficient to raise awareness that the developmental period (among other factors to be discussed) must be considered in making absolute determinations regarding whether or not the presence of a given feature (e.g., presence of waking delta wave activity) is clinically significant. This topic of age-related change in EEG and trajectories of pathology receive detailed treatment in the text *Developmental Psychophysiology,* edited by Louis Schmidt and Sidney Segalowitz.

Controversial Sharp Waves or Spike Patterns

Early in the emergence of clinical EEG, a considerable amount of heated controversy developed around the clinical significance of certain sharp wave or spike patterns that were easily identified in the EEG tracing. These so-called controversial patterns deserve a specific mention here because they have been observed consistently in numerous studies to be more prevalent in psychiatric populations than healthy or general medical patient populations. Despite the increased prevalence in psychiatric patients, defining the clinical and neurobiologic correlates of these patterns has proved to be an elusive goal. The task is made difficult because none of these patterns is specific to a diagnostic category, and none predictably correlates with seizure disorders. Because the physiologic or pathologic correlates of these patterns have not been well studied using currently available methodologies and assessment techniques, the assertion that they are irrelevant to neuropsychiatric conditions is not fully supportable. At a minimum, the appearance of such patterns may give support to the possible organic nature of presenting symptoms. The number of accepted and still controversial EEG abnormalities is quite large and well beyond the scope of this chapter.

Paradigms for Eliciting Pathophysiologic Features

Over the years, electroencephalographers have recognized that certain *activating procedures* tend to increase the probability that abnormal discharges, particularly spike or spike-wave seizure discharges, will occur. Some activating techniques remain standard in many laboratories, others are used only rarely for specific purposes, and still, others introduced in the past largely have been abandoned because they were not easy to use or involved risk. It has long been recognized that seizure manifestations or aberrant behaviors with a pathophysiologic basis involving seizures can be triggered by specific stimuli. In this regard, cases of audiogenic, musicogenic, photogenic, and reading epilepsy readily come to mind, even though such cases are rare, and most practitioners have never encountered them. Seizure phenomena related to other sensory system inputs (e.g., somatosensory and gustatory) are even rarer. Sometimes, it may be possible for laboratory personnel to duplicate or approximate sensory triggers in various modalities to determine if they activate EEG seizure discharges combined with overt symptoms. Some of the more conventional activating techniques are described below.

Photic Stimulation. In the earliest days of EEG, it was known that the frequency of normal EEG activity recorded from posterior scalp regions could be made (within narrow limits) to follow the frequency of a flickering light that was flashed slightly faster or slower than the intrinsic brain wave frequencies, a phenomenon that came to be referred to as *photic-driving.* An illustration of this phenomenon is presented in Figure 1.26–15, depicting the instantiation of a steady-state response of approximately 9 Hz over the visual cortex following stimulation at a 9-Hz flicker rate. When it also became known that photic-driving would sometimes cause paroxysmal discharges to occur in the EEG, *intermittent photic stimulation* (IPS) emerged as a technique for eliciting EEG abnormalities referred to, under conditions of stimulation, as *paroxysmal photosensitive responses* (PRRs). Although there is some variation between laboratories, IPS generally involves placing an intense strobe light (at least 0.70 J) approximately 30 cm in from the nasion and flashing at frequencies that can range from 1 to 50 Hz, depending on how the procedure is carried out. It is recommended that a round stimulator be used to best ensure uniform stimulation of the entire retina. In addition to flashing light, flickering grid stimuli may be used to further assess pattern-specific photosensitive responses, which may be indicated in patients who tend to have seizures induced by television watching or video gaming. Some laboratories sample independent flash frequencies separately and randomly, without necessarily sampling all frequencies between 1 and 50 Hz, while others use a *zoom technique* in which the flashes start at a low frequency, such as 1 Hz, and are then gradually and continuously increased to higher frequencies. In some individuals, IPS produces facial, and eye muscle jerks called a *photomyoclonic response* (PMR). More than 40 years ago, Henri Gastaut observed that PMR occurs in 0.3% of normal subjects, 3% of epileptic patients, and 17% of patients with psychiatric disorders. The PMR also is enhanced in the early stages of alcohol withdrawal in chronic alcoholics and after a sudden withdrawal from barbiturates and other sedatives. The clinical relevance of PMR in psychiatric patients is yet to be fully explored in systematic research.

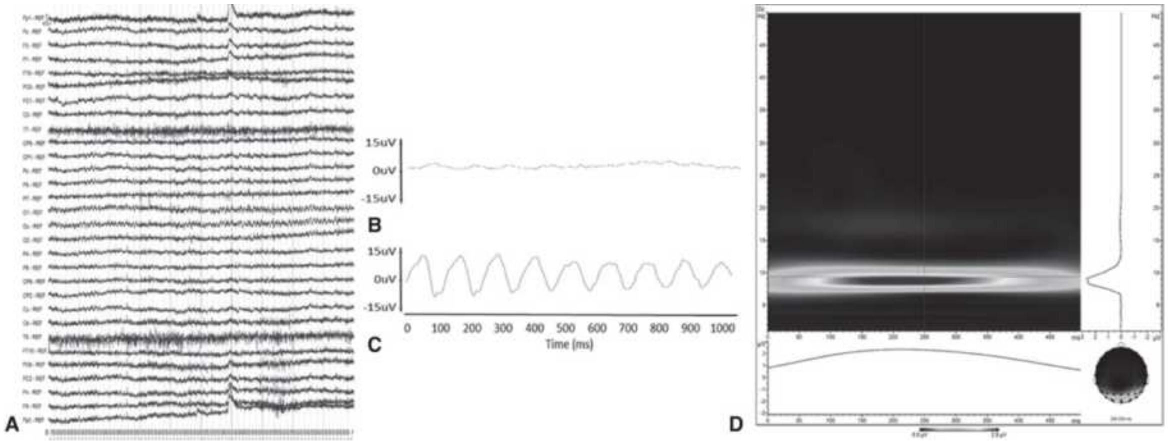

FIGURE 1.26–15. Photic driving using checkerboard pattern presented at 9 Hz flicker rate. **Panel A:** 10-second display of EEG at the beginning of photic stimulation. The onset of the visual checkerboard display, indicated by “s” at the bottom of the display. **Panel B:** 1-second segment based on average EEG for 10-second period prior to stimulation recorded at electrode Oz. **Panel C:** Visual-evoked potential (VEP) based on an average of 10 seconds recorded during stimulation at electrode Oz. The VEP represents an oscillatory pattern of 9 cycles over a 1-second segment. **Panel D:** VEP represented as time–frequency decomposition by wavelet transform of frequencies of 1 to 50 Hz at electrode Oz, with peak activity at approximately 9 Hz. The scalp plot (bottom right) indicates activity is generated primarily over occipital electrode sites.

In terms of activating EEG abnormalities, one looks for a PRR that consists of bilaterally synchronous, usually diffuse, spike-and-wave discharges of various frequencies or diffuse, multiple spike-wave complexes. When the resting EEG is normal, and a seizure disorder or behavior that is suspected to be a manifestation of a paroxysmal EEG dysrhythmia is suspected, IPS can be a valuable activation to use. Its primary limitations, especially as a routine procedure, are the “false positive” incidence of PRR in individuals with no history of seizure disorder or current symptoms suggestive of such and similarly low sensitivity in cases of verified seizure disorder. The persistence of the spike-wave discharges after the cessation of the IPS may also be indicative of a PRR. Moreover, a pathologic process might be indicated by the spread of the spike-wave activity to lead outside the occipital region.

Polysomnography—Sleep Study. Largely because of the pioneering studies and perseverance of Fred and Erna Gibbs, EEG recording during sleep, natural or sedated, now is widely accepted as an essential technique for eliciting a variety of paroxysmal discharges, when the wake tracing is normal or for increasing the number of abnormal discharges to permit a more definitive interpretation to be made. A variety of focal and diffuse spike and spike-wave discharges, as well as several minor or controversial paroxysmal patterns, occur much more often during drowsiness and light sleep than during the wake recording, and some of them are seen almost exclusively during the sleep recording. Paroxysmal patterns differ substantially among themselves in the degree to which their appearance in the tracing is sleep-activation dependent. Although most clinical EEGs should contain drowsy- and light-sleep tracings to be complete, stage 3 deep sleep with generalized high-voltage delta slowing have almost no activating property and is not clinically useful.

Sleep Deprivation. It has been shown that the stress produced by 24 hours of sleep deprivation alone can lead to the activation of paroxysmal EEG discharges in some cases. This effect is presumably independent of the known activating properties of natural or sedated sleep itself. Sleep deprivation is without risk for the healthy patient but may be contraindicated for patients medically or physically compromised. The primary disadvantage is the tendency for sleep-deprived individuals to enter into a deep sleep (stage 3) immediately at the start of the recording, thus reducing the chances of detecting spike activity. The optimal method is to ensure that the subject stays awake until the recording begins and remains so for the initial recording period, after which a gradual transition into drowsy sleep can be observed.

Drug Activation and Suppression Tests. Psychiatrists evaluating a patient with an atypical behavioral reaction to a drug (e.g., an unusually severe response to a small amount of alcohol or other abused drug or some other highly idiosyncratic response) may consider a drug-activated EEG to determine if ingestion of the drug activates a seizure or other paroxysmal abnormality of clinical significance. In the past, drug activation using hypoglycemic agents (e.g., tolbutamide) and convulsants (e.g., pentylene tetrazol) was used to activate epileptiform discharges. However, this type of activation has been essentially eliminated from clinical work in recent years due to the high incidence of discharges produced by these drugs in healthy subjects and the low value of positive findings in individuals with epilepsy. Today, drug modulation of EEG remains to be used as an investigative procedure for ascertaining drug effects on presumed neurophysiologic mechanisms, both with relevance to enhancing functions in the case of drug development and in modeling pathologic conditions in healthy subjects using the pharmacologic challenge design. While drug activation is no longer used in routine clinical practice, drug suppression paradigms are still regarded as useful for the identification of focal epileptiform activity. Seizure foci can be difficult to detect on EEG when in the case of generalized discharges and accompanying spread of activation across electrodes. In a drug suppression test, generalized discharges can be reduced by intravenous injection of diazepam or barbiturate methohexital (e.g., methohexital suppression test; MHXT). The rationale behind this approach is that by using drugs that block excitatory neurotransmission and/or enhance inhibitory transmission, much of the background EEG is suppressed while autonomous epileptogenic lesions, which are resistant to synaptic blockade, continue to fire. This approach can therefore improve the localization of epileptiform lesions against background EEG and generalized activity that is dependent on synaptic input.

Hyperventilation. Hyperventilation is one of the oldest and still one of the most frequently used activation procedures in clinical laboratories. While remaining reclined with the eyes closed, the patient is asked to overbreathe through their open mouth for 1 to 5 minutes (not useful beyond 5 minutes), emphasizing deep breathing rather than rapid breathing. The purpose is to empty the lungs to the extent of producing hypocarbia, and to reduce oxygen and glucose supply to the brain through cerebral vasoconstriction. The normal EEG response to hyperventilation (referred to in EEG parlance as a *buildup*) consists of an increase in generalized medium- to high-voltage synchronous slow waves in the delta range, which then quickly subside when overbreathing stops (within 1 to 2 minutes normally). Not everyone has a build-up response to hyperventilation, and children are far more likely to respond with diffuse EEG slowing than adults. Indeed, this procedure is not effective in the older adults, who show minimal reactivity of the cerebral vasculature to hypocarbia, and is contraindicated in those aged 65 and older. In terms of activating EEG abnormality, hyperventilation is especially effective in eliciting the classic diffuse three-per-second spike-and-wave complex of petit mal seizures when the pattern does not first appear in the standard waking EEG. Hyperventilation may, to a lesser degree, activate other synchronous diffuse spike-wave patterns, and activation of focal seizure activity has been infrequently reported. An interesting observation is that significantly low blood glucose levels have been associated with large, synchronous delta wave hyperventilation buildups that may persist beyond 2 minutes after overbreathing is stopped. Accordingly, a large delta wave buildup in an adult may signal the existence of covert, unsuspected pathologic hypoglycemia. If this suspicion presented itself during a recording, a good idea would be to give a sugar drink to the patient and then repeat hyperventilation later. If the glucose ingestion reduces or abolishes the large hyperventilation buildup, then the suspicion of hypoglycemia is likely confirmed. In general, hyperventilation is one of the safest EEG activating procedures and, for the majority of the population, will present no physical risk. However, it may pose a risk for patients with cardiopulmonary disease or risk factors for cerebral vascular pathophysiology, thus would also be contraindicated in these cases.

Brief Synopsis of EEG Findings in Organic Pathophysiology

The EEG has a high degree of sensitivity to a wide variety of medical conditions and events that impinge on CNS function. For this reason, EEGs sometimes have contributed to the detection of unsuspected organic pathophysiologies influencing a psychiatric presentation. The range of EEG findings in medical disease is enormous and far beyond the scope of this chapter, but we touch on four of the most common organic pathologies informed by EEG in the context of a diagnostic workup.

Seizure Disorder. The hallmark EEG finding for a seizure disorder is the generalized, hemispheric, or focal spike or spike-wave discharge, or both. However, this statement constitutes a large oversimplification because many types of EEG abnormalities have been associated with seizures at one time or another, and some patients with a bona fide seizure disorder have been known to have normal interictal EEG tracings. Indeed, while epileptiform activity detected on EEG is a reasonably specific indicator of epilepsy, sensitivity can be quite low, ranging from 25% to 56%, with only about 50% of patients diagnosed with epilepsy evidencing *interictal epileptiform discharge* (IED) of their first EEG examination (of note, this rate can increase to 80% in sleep EEG). Furthermore, seizure disorders have many possible etiologies, whether genetically based or a consequence of head trauma, cerebrovascular pathophysiology, metabolic disorders, structural brain lesions, infectious or toxic encephalopathies, certain drug-abuse withdrawal states, or iatrogenic causes to name only a few. The diagnosis of seizure disorder based on EEG findings should consider family and personal medical history while also keeping in mind that seizures can be idiopathic.

The classification of seizure types is also broad, and the majority of specific seizure manifestations may be of little concern to the practicing psychiatrist. One of the exceptions may be *absence* (i.e., petit mal) seizures, during which the patient presents with grossly impaired consciousness marked by pronounced confusion, greatly slowed mental processes, and stuporous or somnolent behavior. Although these seizures tend to be brief, normally lasting only 15 to 30 seconds, absence episodes can reoccur as many as 100 times per day. The psychiatrist is most likely to encounter this clinical presentation in the emergency room, where it is often confused with other functional or organic syndromes or intoxicated states. If suspected, then the status can be confirmed quickly by an EEG demonstration of continuous, diffuse spike-wave activity. Typical absence seizures also may be of relevance to the child psychiatrist as this type of epilepsy has a peak incidence from childhood to early adolescence and, in some cases, may be mistaken for inattention or other functional behavioral reactions. Spike foci in the anterior temporal region associated with CPSs (also known as psychomotor or temporal lobe seizures) should also be of interest to the psychiatrist due to the wide range of possible seizure manifestations, consisting of almost any combination of "automatic" motor movements, sensory disturbances, psychiatric symptoms, or autonomic signs. The symptom cluster almost always remains the same from one attack to another, and the "automatic" behavior during the ictal event may be quite unusual, with the patient undressing, picking at clothing and lip-smacking, walking in circles and yelling, trying to climb on a table while making guttural noises, experiencing hallucinatory symptoms, or almost any other action. When continuous or closely spaced CPS attacks (complex partial status) lasting hours or longer than a day occur, they may mimic hysterical dissociative states.

Spike activity of various kinds and locations is sometimes found in the EEGs of psychiatric patients with no obvious seizure manifestations. In such cases, one might consider possible relationships to episodic aberrant behavior (particularly explosive aggressiveness), if present, as well as a trial of anticonvulsant medication. Sometimes, such spike discharges in the nonseizure psychiatric patient may indicate an elevated risk for iatrogenic seizures after treatment with neuroleptic medication.

Structural Lesions. Brain tumors are the leading cause of new-onset seizures in adults. EEG is rarely used for the diagnosis of structural and space-occupying lesions, which are typically assessed with CT or MRI. Nonetheless, it should be noted that 90% of cortical brain tumors can be detected and localized on the scalp with routine EEG.

Closed Head Injury. Generalized or focal slowing is the main characteristic of EEG following head trauma, appearing over the site of the trauma or over a *contrecoup* (e.g., the side opposite the side of impact) location. Attenuation of posterior alpha is another hallmark feature associated with a head injury. The basic premise guiding the interpretation of these observations is that the frequency and power of scalp-recorded EEG depends on the numbers of activated neurons and the integrity of their connections. Therefore, in the case of injury, reductions in neurons and disruption of normal connectivity result in attenuation of signal power and a shift to slower frequency. The EEG change may be relatively transient with resolution over days or weeks, or it may persist for extended periods of time. Focal spiking may be taken as an early indication of a posttraumatic seizure disorder. Subdural hematomas after a head injury may present with focal delta slowing or more widespread slowing and, on occasion, amplitude asymmetry of waking EEG with the voltage decreased over the site of the hematoma. That said, conventional EEG is generally not regarded as clinically useful in cases of mild traumatic brain injury (mTBI), particularly beyond the first several hours following injury and when loss of consciousness (LOC) with injury is 2 minutes or less. EEG findings may appear normal early in the postinjury period following mTBI despite the presence of postconcussive symptoms, such as headache, dizziness, fatigue, irritability, impaired memory and concentration, insomnia, and sound and light sensitivity. Clinically significant EEG changes in association with mTBI, if at all present, may not be recognized as deviant until compared with follow-up EEG taken later in the postinjury period. Of note, literature reviews suggest that as many as 20% of cases with noncomplicated mTBI recover from clinical symptoms within the first 24 hours, and as many as 85% within 1 week following injury. Similarly, abnormalities detectable by conventional clinical EEG tend to resolve within months of injury.

While routine clinical EEG may not be regarded as a useful tool for routine screening of mTBI, at least not beyond the detection of posttraumatic seizure activity, there is evidence that Q-EEG may hold some promise. Numerous studies by Robert Thatcher and colleagues, as well as others, have reported Q-EEG findings associated with the postconcussion syndromes following mild to moderate closed head trauma. Reported Q-EEG changes involve a combination of increased theta power, decreased alpha power, increased theta–alpha ratios, and increased coherence between frontal and temporal areas in the resting EEG. A number of studies have applied discriminative analysis to Q-EEG collected in mTBI versus control sample, yielding high classification accuracy, and have demonstrated the utility of Q-EEG in predicting functional disability. Q-EEG and related topographic analysis are likely to outperform routine clinical EEG in these cases because the quantification process aims to identify sharply focal or regional alterations in activity according to statistically significant deviations from normative parameters, which could remain below the threshold for accurate visual detection. The use of Q-EEG in the assessment of mTBI nevertheless remains a heavily debated topic, with supporting evidence criticized for the composition of study samples and other methodologic choices.

Metabolic and Endocrine Disorders. EEG is not a tool that one would use in assessing endocrine or metabolic disorders. Yet the majority of these conditions alter the EEG, and unexpected abnormal findings during routine EEG screening have occasionally been the first positive laboratory results leading to the subsequent identification of mild endocrine or metabolic disturbance. Because metabolic and endocrine disorders, as well as certain vitamin deficiencies and toxic exposures, affect the brain in a global fashion, they tend to produce diffuse, generalized slowing of wake frequencies, with the degree of abnormality correlated with the severity of the disorder. As would be expected, focal slowing is almost never seen, and paroxysmal spike activity is quite rare, although spike activity has been observed in cases of hypocalcemia, acute intermittent porphyria, and exposures to some toxic substances (e.g., lead, carbon monoxide). Early hepatic disease is commonly associated with generalized slowing in the theta range, which may increase in severity with elevated blood ammonia levels. If the disorder progresses to hepatic encephalopathy, a distinctive and diagnostically relevant EEG pattern called *triphasic waves* (or, more informally, *liver waves*) can emerge, characterized by frontally dominant or diffuse 1.5- to 3.0-per-second high-voltage slow waves initiated by blunt or rounded spike-like transients. For a thorough review of the effect of various metabolic disorders on EEG, the reader is directed to the referenced chapter by Pang and colleagues.

EEG IN PSYCHIATRY

Psychiatry Overview

Starting in approximately 1938, a flurry of continuing EEG investigations began to reveal an increase in the overall prevalence of minor abnormalities in almost all psychiatric populations as compared to healthy or nonpsychiatric individuals, which continue to be investigated to this day. On the other hand, two major factors led to the rapid disillusionment of the field of psychiatry with EEG. The first was the lack of specificity of EEG abnormalities to known psychiatric syndromes. The second factor, alluded to previously, was the continuing discovery of EEG abnormalities correlating with epilepsy, tumors, encephalopathies, stroke syndromes, and coma. With discoveries of significant EEG changes accompanying neurologic problems on the rise and EEG abnormalities associated with psychiatric symptomatology remaining to be relatively obscure and noncontributory to the diagnostic process, the field of *clinical EEG* (and later *clinical neurophysiology*) became more firmly established as a subspecialty of neurology, with a de-emphasis of EEG among psychiatrists.

The recent surge of interest in the neurobiology of psychiatric disorders, the emergence of the clinical field of *neuropsychiatry,* and the unprecedented advances in computerized analyses of EEG and other neurophysiologic signals have resulted in a strong rekindling of interest in electrophysiology among psychiatrists. In 1995, John Hughes undertook the mammoth task of compiling a comprehensive outline of the broad area of EEG and psychiatry with 181 significant references selected for citation from before 1950 until 1994. When such compilations are inspected, the findings reveal that more than one-half of the EEG-psychiatry references appear after 1980, and one-third were written within the 5 years preceding Hughes' 1995 report. Continued inspection of the literature indicates that this trend has not abated. A *PubMed* search using the terms "EEG + Psychiatry" identified an additional 12,054 articles published in this field since 1995 (a nearly twofold increase since 2015), including both clinical and basic science research. Furthermore, two scientific organizations have been established for the primary purpose of accelerating the translation of electrophysiologic research findings to clinically utilizable laboratory tests. In 1991, the American Psychiatric Electrophysiology Association (APEA) was founded. In 1999, the APEA merged with the American Medical EEG Association (AMEEGA), and the EEG and Clinical Neuroscience Society (ECNS) was formed (http://ecnsweb.org).

In spite of the enormous research potential recognized in psychiatry, electrophysiologic methods are still of limited impact in clinical practice, where their primary applications are to aid in differential diagnostic purposes (i.e., the exclusion of “organic” brain pathology) and the investigation of sleep disorders. There is no accepted indication of EEG methods for the diagnosis of DSM-5 psychiatric disorders or for drug treatment choice and monitoring. This may seem surprising when considering that (1) robust results of Q-EEG abnormalities have been reported and independently confirmed for major psychiatric disorders, especially for schizophrenia; (2) evidence has been provided that abnormalities of event-related potentials (ERPs) and of eye movements are related to genetic risk factors, symptom dimensions, prognosis, and diagnostic subtypes of schizophrenia and depression; and (3) EEG and related methodologies already provide translational paradigms and have demonstrated promise in predicting the likelihood that novel therapies and compounds will exhibit clinical efficacy in preclinical development as well as in clinical trials. However, it should also be acknowledged that despite acceptance of psychiatric disorders as “brain disorders,” there are currently no tests from any discipline of brain imaging or behavioral assessment that have been formally integrated into the clinical diagnostic formulation. Perhaps this disconnect reflects the fact that it is not brain function, *per se,* but disordered behavior (e.g., reduction in normal routine behavior, elevation in abnormal behavior causing distress to self/others) that signifies the need for clinical attention and, thereby, serves as the basis for diagnosis and treatment decisions. In this respect, it is understandable that even the strongest evidence linking an electrophysiologic feature to a symptom or disorder may not necessarily indicate the use of that feature as a diagnostic measure when clinical symptoms are acutely present.

Where then could electrophysiology find a place in applied clinical psychiatry? One promising area is in the early detection of pathophysiologic processes during preclinical periods that could eventuate in psychiatric illness. For example, the prodromal period of psychosis (i.e., DSM-5 Attenuated Psychosis Syndrome) is now well recognized as a stage of disease progression during which at-risk individuals (e.g., presenting genetic liability, subclinical symptoms) begin to express functional impairment, which may correspond to a combination of factors affecting critical periods of brain maturation. Early detection has been a topic of great interest to psychiatric research as well as to the *National Institute of Mental Health* (NIMH), which, under the directorship of Thomas Insel, coalesced eight independent research laboratories in forming the *North American Prodrome Longitudinal Study* (NAPLS) consortium. Early *p*sychosis *p*rediction and *p*revention (i.e., EP3) was subsequently identified as a NIMH priority issue in 2013, with plans to focus funding efforts on this area for the following 5 years. These developments are consistent with aims to reduce the disability of severe psychiatric illness through early detection and intervention, reducing long periods of untreated illness that may precede a first clinical encounter. EEG methodology has figured prominently in the conceptualization and data collection of NAPLS investigations, with paradigms demonstrated to be sensitive to the pathophysiology of schizophrenia used as predictors of conversion to formal psychotic disorder during the prodromal stage. EEG has provided a highly valuable assessment tool for this work, having the advantages of being low-risk, noninvasive, and well-tolerated by this study population. Importantly, EEG has established sensitivity to normal neurodevelopmental processes that are hypothesized to be altered in the schizophrenia disease process. Finally, it should not be overlooked that EEG systems can be purchased at a relatively low cost, enabling standardization of recording apparatus and methodology across centers participating in multisite research. Innovations of the NAPLS study have included traveling subjects and additional calibration procedures to assess the reliability of instrumentation across sites and has firmly established the feasibility of large-scale data collection efforts in psychiatry.

Nonspecificity of Results

Two particular problems plague and hinder research in the field of EEG abnormalities in psychiatric populations. First is the belief that EEG abnormalities can be found in up to 20% of otherwise healthy individuals. The second major problem is another pervasive belief that interictal epileptiform activities, in the absence of clinically diagnosable seizures, are incidental (or nonspecific) findings without clinical relevance. Well-designed studies addressing these two important issues are needed to allow further research in this field to be more meaningful.

Moreover, there are only a limited number of ways in which brain electrical activity can respond to normal or pathologic influences. Brain waves change simply by becoming faster or slower in frequency or lower or higher in voltage, or perhaps some combination of these two responses. Thus, very similar patterns of EEG abnormality can emerge from different etiologic causes. For example, a neoplasm, subdural hematoma, brain abscess, cerebral vascular accident, or closed head injury may result in similar, although not always identical, focal EEG slowing. Generalized slowing is a common abnormal finding for which the etiologic causes include cortical atrophy, drug-induced encephalopathy, ECT, encephalitis, certain endocrine disorders (hypothyroidism and hypopituitarism), porphyria, head trauma, lead exposure, hypocalcemia, and Wernicke encephalopathy, to name but a few. The nonspecificity of results or the failure to specifically denote etiology is a genuine limitation, but one that is not as bleak as one may think.

Concerns raised regarding the sensitivity and specificity of EEG abnormalities do pertain primarily to conventional clinical EEG, which realistically only permits examination of the few dimensions of data that can be ascertained by visual inspection. The same concerns, though still appropriate, may, in time, be alleviated to an acceptable level given advances in the field of Q-EEG. Q-EEG profiles, as well as deviations on evoked responses, have produced high sensitivity and specificity in classification analysis comparing subjects with psychiatric disorders to healthy comparison and other diagnostic groups. Although, by and large, Q-EEG remains a research tool, its potential to contribute to diagnostic decision-making has been demonstrated in small-scale proof of concept studies. Clinical development to date has been hampered, in part, by the availability of sufficient normative samples and of statistical tools suited for the task of conducting classification in large, intercorrelated data sets. The application of modern machine learning (ML) algorithms shows promise as a solution for uncovering the bounds of disease classes using “unsupervised” models that alleviate dependence on known diagnostic categories and can be optimized to make use of the complex dimensionality of EEG represented in spatial, temporal, and spectral domains.

Applied EEG in Psychiatry

As stated previously, the role of EEG in applied clinical psychiatry remains fairly limited to the differential diagnosis of organic etiologies. In this regard, the EEG can be most useful when the presenting symptoms, age of first symptom onset, or response to treatment is atypical, suggesting that a more structural cerebral pathology may be contributing to the syndrome being evaluated. In such cases, the presence of paroxysmal EEG activity or focal or diffuse slowing should be ruled out. If paroxysmal activity is identified, then it is important to be clear that such an EEG finding is not sufficient grounds to diagnose epilepsy unless, retrospectively, the symptoms can be seen as being of an epileptic nature. The presence of a paroxysmal EEG abnormality in a patient with an atypical presentation, particularly in the absence of a family history of the diagnosis that is being considered, should lead the clinician to be wary of and less confident in the diagnosis. It may be a good practice at the current state of knowledge to take advantage of the *not otherwise specified* category. Whether anticonvulsants are indicated in these situations remains a question open to clinical judgment. Early studies in this area suggested that anticonvulsants can sometimes block EEG epileptiform discharges and also can lead to dramatic clinical improvement in individuals exhibiting aggressive behavior. Furthermore, carbamazepine (Tegretol) was thought to be clinically useful in individuals with schizophrenia who exhibited temporal lobe EEG findings and aggressive tendencies without a history of seizure disorder. Based on current practice, it would seem as though prescribers are comfortable using adjunctive anticonvulsant therapy for many indications, irrespective of EEG findings, as would be suggested by common prescription of this medication class in anxiety disorders, dementia, schizophrenia, depression, bipolar disorders, and in cases of drug and alcohol withdrawal syndromes. Prescription in these cases, perhaps with the exception of drug withdrawal, is not based on the question of seizures but rather on the efficacy of anticonvulsants for treating psychiatric symptoms, including aggression, mood swings, mania, and panic symptoms.

Similar comments can be made regarding the presence of a focal slow-wave abnormality in the EEG. A minimal focal slow-wave abnormality is unlikely to reflect a serious space-occupying lesion in a patient with a normal neurologic examination. However, if no logical explanation for the minor focal slowing is apparent (e.g., old head injury) and the finding did not appear on earlier EEGs, then a repeat EEG may be valuable to determine whether the finding is static or progressive. Clinicians should use their clinical judgment regarding when imaging (and what imaging) should be performed. If a space-occupying lesion is suspected, then a CT scan is usually sufficient. An MRI is more likely to yield abnormalities if the status of the lesion is less acute (e.g., old head injury). The presence of a focal slow abnormality, similar to the presence of paroxysmal activity, should cause the clinician to be wary of the diagnosis. Consideration also should be given to obtaining a neuropsychological evaluation to assess whether the focal cerebral abnormality identified by the EEG has any clinically correlating deficits.

The situation is different when an EEG is obtained in the course of managing an unstable or nonresponsive patient. The most important data to be gained from an EEG performed on an acutely agitated patient are the presence of diffuse slowing (indicating a delirious encephalopathic state) or ongoing epileptic activity NCSE. In both conditions, the EEG abnormality is fairly obvious and can be detected in a short period of time by temporarily retraining the patient, if needed, and using an electrode cap to minimize setup time. The EEG can be extremely useful in ruling out delirium secondary to medication toxicity, and, in this respect, it should be noted that diffuse EEG slowing (suggestive of a diffuse encephalopathic process) may be seen despite serum plasma levels within the therapeutic range.

When dealing with geriatric populations, sensitivity to organic pathology should be particularly heightened. In addition to the evaluation for dementia and delirium, EEG could be invaluable when there is a suspicion of seizures, and, in this respect, it is noted that older adult patients may account for as many as one-fourth of the cases of new-onset epilepsy. One also must keep in mind that a normal EEG in a patient diagnosed with dementias, once past a mild stage, should raise the suspicion that a depressive disorder, or other reversible cause, may be contributing to the clinical picture.

EEG Findings in Psychiatric Disorders

In addition to the general considerations regarding the interface between EEG and psychiatry, a number of psychiatric disorders deserve specific mention. Findings deemed promising for clinical applications will be briefly reviewed.

Schizophrenia. EEG abnormalities have been extensively examined in schizophrenia patients. A recent literature review entailing 133 published studies revealed that common EEG abnormalities observed in schizophrenia are also detected at the time of the first psychotic episode, as well as in high-risk clinical individuals, including increases in delta and gamma band. Moreover, some of these features appeared predictive of transition to psychosis in at-risk samples and associated with poorer functional status, suggesting utility as markers of disease progression and prognosis. The EEG literature in chronic schizophrenia is extensive, including at least 30 meta-analytic studies covering specific EEG and ERP measures investigated as putative psychosis-related biomarkers, including mismatch negativity (MMN), P50 sensory gating, the P300, and the gamma-band auditory steady-state response (ASSR) (see ERP sections below).

A study published in 2015 by the *Bipolar and Schizophrenia Network on Intermediate Phenotypes* (BSNIP) multicenter consortium examined the polygenic architecture of resting-state EEG frequency activity in samples of schizophrenia, psychotic bipolar disorder, and healthy individuals using 1 million single-nucleotide polymorphisms (SNPs) derived from DNA. Multivariate data clustering and component extraction methods identified five frequency and nine SNP components, with theta and delta activity found to correlate with gene components involved in brain development, neurogenesis, and synaptogenesis. Increased anterior delta and posterior theta abnormality were present in both schizophrenia and bipolar samples, while theta differed between the two disorders. Theta abnormalities were also mediated by unique gene clusters involved in the glutamic acid pathway and synaptic cell adhesion processes. These results replicated prior findings pertaining to excess resting slow-wave activity in schizophrenia and further differentiated the component structure of this abnormality from psychotic bipolar disorder patients, thus elevating the specificity of this feature to a diagnostic differential that is difficult to make using conventional clinical assessment. These findings further point to plausible genetic networks, consisting of genes previously implicated in schizophrenia as well as novel mechanisms that could underlie abnormal features observed in resting EEG.

Catatonia. Although the routine EEGs of catatonic patients with known functional etiology tend to be normal, catatonia could be a symptom of a number of serious medical conditions, including epilepsy and encephalopathy secondary to a variety of general medical conditions. Neuroleptic malignant syndrome has presented, at times, as catatonia, and, in such cases, the EEG is likely to show a picture of diffuse slowing indicative of an encephalopathic process. Although a well-known patient whose usual presentation includes catatonic symptoms may not need an EEG, it is advisable to obtain an EEG on every new patient who presents with catatonic symptoms.

Panic Disorder. Panic attack symptoms carry a significant resemblance to symptoms induced by temporolimbic epileptic activity, particularly those originating from the Sylvian fissure. Fear, derealization, tachycardia, diaphoresis, and abdominal discomfort are characteristic symptoms of simple partial seizures with psychiatric and autonomic symptomatology. Studies comparing symptomatology of patients with panic disorder to patients with CPSs have reported many similarities, suggesting that there may be a common neurophysiologic substrate linking the disorders. Notable conclusions of this literature are that individuals suffering from panic disorder show a higher rate of unspecific EEG abnormalities that include a unique increase in beta power, thought to align with the common clinical observation of hyperarousal in these patients. Although some evidence may suggest that IEDs occur at higher than normative baseline rates in panic disorder, relationships between panic symptoms and seizure activity remain far from clear, with EEG findings ranging from paroxysmal epileptiform discharges to asymmetrical increases in slow-wave activity and focal slow-wave abnormalities (reportedly observed in up to 25% of patients with panic disorder). Along these lines, it has been suggested that the use of anticonvulsant drugs, including valproic acid (Depakene), be considered in cases of panic disorder if conventional anxiolytic treatment is inefficacious.

On the other side of the coin, panic attacks also have been demonstrated to occur more frequently in epileptic patients. In some cases, these attacks could lead to overmedication for seizures if their nature is not precisely defined. Again, EEG monitoring during multiple attacks is indispensable in rendering such an evaluation. Other reports have concluded that the most common psychiatric disorder that must be differentiated from temporal lobe epilepsy is panic disorder. Epileptic seizures are commonly briefer and more stereotyped than panic attacks. Additionally, aphasia and dysmnesia often accompany seizure activity.

Obsessive-Compulsive Disorder. As recently reviewed by Zaboski et al. (2021), utilizing EEG in the study of the neural underpinnings of obsessive-compulsive disorder (OCD) is relatively scarce and has to date, primarily revealed disruptions in the P300 and in the spectral power of resting EEG. For example, EEG studies in patients with OCD reported a widespread increase of slow waves, with an additional report of epileptiform activities over the left temporal lobe. The frequency of these abnormalities varies among studies.

Q-EEG studies have more often involved the anterior regions of the scalp, supporting the hypothesis of frontal dysfunction in the pathogenesis of OCD. The cingulate gyrus, in particular, has been implicated as a locus of dysfunction by Q-EEG topographic brain mapping. Data concerning the involvement of a specific frequency band have been heterogeneous, probably due to differences in methodologies, small study sample sizes, and a broadly heterogeneous patient population in terms of the presentation of OCD symptoms. Although not given the coverage that might be deserving of the topic in the current chapter, EEG biofeedback or, more precisely, *neurofeedback* has gained interest as a nonpharmacologic treatment option for individuals suffering from OCD. Basically, neurofeedback involves a graphic representation of ongoing EEG to the patient, who may learn to implicitly modulate activity upward or downward toward what would be regarded as normalization from a pathophysiologic baseline. Again, although varied in terms of the specific EEG frequencies and locations involved in OCD therapy, neurofeedback has been approved by the FDA as a treatment for stress management and, like stress, may benefit from improved self-regulation that can be achieved as a common therapeutic mechanism of neurofeedback.

Mood Disorders. Depression may result from a variety of brain and other somatic diseases (e.g., tumors, dementia, and vascular accidents) that EEG can help to identify. With this exception, the diagnosis of mood disorders is typically not informed by EEG. Nevertheless, research findings of EEG abnormalities in depression, starting with the first publication by Fredrick Lemere in 1936, have provided many intuitive interpretations relating the clinical features of depression to brain mechanisms. Of note, Lemere's first paper, in which he proposed *"… The ability to produce 'good' alpha waves seems to be a neurophysiological characteristic which is related in some way to the affective capacity of the individual …*," followed Hans Berger's original report on the human EEG by only a few years and, indeed, referred to 10 Hz alpha waves as "Berger rhythms." Observations involving an abnormality in the alpha frequency have continued to be the major story pertaining to EEG research and conceptualization of depression up through current times. Since the early 1980s, work credited to Richard Davidson and colleagues has highlighted frontal alpha asymmetry and the associated *Approach-Withdrawal* model (aka the dispositional model) as an explanatory synthesis of the neural basis of depression. Briefly, an asymmetry of frontal alpha involving hyperactivation of the right hemisphere relative to the left has been interpreted as the interplay of *approach* (facilitating appetitive behaviors in association with positive affect) and *withdrawal* (facilitating avoidance of aversive stimuli in association with negative affect), with a reduction in left-sided alpha, thought to reflect deficient approach system activation in depression. The model further stipulates that relatively higher right-sided withdrawal system activation contributes to anxiety and other forms of psychopathology. EEG research has continued to support this seminal finding of frontal alpha asymmetry and, to some extent, this formulation of depression, with studies by other groups suggesting that alpha asymmetry more likely reflected a state rather than a trait feature in major depression. New models based on alpha asymmetry have since been proposed, including the capability model, which postulates that brain activity elicited under emotional challenge will be a stronger indicator of psychopathology than when measured at rest, and therefore entails testing with emotionally provocative stimuli.

Few large-scale EEG studies have been carried out in acute, drug-free subjects during a manic episode, probably due in part to the difficulty of obtaining such data. However, Joyce Small and colleagues published several papers in the late 1990s on a sample of approximately 200 patients admitted to the Larue D. Carter Memorial Hospital of Indianapolis in acute manic states, many of whom received follow-up EEGs when medicated, after medication washout, and when rehospitalized during subsequent episodes. Q-EEG was also performed in a subsample of about 40 patients who were cooperative enough to test. Remarkably, the majority of EEG data collected was interpreted to be normal. A slight increase in abnormal features (16% of the sample) was associated with a positive family history of mood disorder, and spike-wave activity was observed in 17% of the sample. Q-EEG differentiated patients on the basis of prescribed pharmacotherapy, with higher theta activity at baseline predicative of poorer response to psychotropic drugs. Abnormal hemispheric laterality was detected in a smaller portion of patients (9%), which implicated left hemisphere involvement.

Recent evidence suggests the usefulness of baseline Q-EEG measurements in the prediction of positive as well as negative outcomes of treatments with antidepressant drugs. Andrew Leuchter and his team introduced cordance, a Q-EEG measure combining relative and absolute power, to investigate correlates of response to fluoxetine in unipolar depression with a double-blind, placebo-controlled paradigm. Significantly more depressed subjects with high cordance in the theta band responded to fluoxetine in comparison to those with low cordance. The results were interpreted as suggesting that only depressed subjects with low baseline dysregulation respond to treatment. More recent studies conducted by Leuchter's group have also found a change in prefrontal cordance measures to be predictive of a positive response (e.g., reductions in Hamilton Depression Scale scores) to placebo conditions. However, the direction of change in concordance indicating improvement has not been entirely consistent across studies.

Attention-Deficit/Hyperactivity Disorder. Attention-deficit/hyperactivity disorder (ADHD) is a prevalent disorder among children and adolescents, as well as adults. The prevalence of abnormalities in the conventional EEG is higher in ADHD than that in age- and gender-matched healthy control subjects, with approximately one-third of ADHD children having abnormal routine EEGs and a significant proportion exhibiting spike or spike-wave discharges. In fact, some studies even report abnormal EEG incidence to be as high as 60% for this disorder.

Q-EEG has identified increased theta activity, particularly in the frontal lobe regions, as a consistent finding in ADHD. Even when learning disability is removed as a confounding variable, increased theta remains as one of the most prevalent EEG findings in this disorder, replicated by a number of groups from different parts of the world. In addition to being a consistent finding, the theta increase seen in ADHD is resistant to age effects, whereas abnormalities of beta EEG frequency tend to decrease with age. It also should be noted that early work suggests that increased theta abnormality frontally may be a strong predictor of response to methylphenidate and other psychostimulants and that favorable clinical responses may be associated with a normalization of the EEG abnormality. Because decreased beta activity has sometimes been reported in ADHD, some investigators have tried to examine the theta-to-beta ratio as a stronger measure of ADHD-related EEG abnormality with a higher detection power. There are, however, additional Q-EEG reports of an increase (instead of decrease) in beta activity in similar patients, raising questions concerning whether or not different subtypes of ADHD differentially affect beta activity. Finally, discriminant function analyses of Q-EEG data have been shown to be capable of significantly discriminating between healthy attention-deficit disorder and ADHD and learning disability subpopulations.

Dementia and Delirium. Because patients with advanced dementia rarely have normal EEG, a normal test result can play an important role in diagnosing cases of pseudodementia (symptoms of dementia secondary to depression or psychosis). When dementia and depression coexist, it becomes important to have some idea about the relative contribution of each disorder to the overall clinical presentation.

The routine EEG is also useful in following the progression of Alzheimer disease and may be useful in differentiating the Alzheimer etiology from that of frontotemporal dementia (FTD), which has been associated with either normal routine EEGs or mild frontal slowing despite progressive behavioral deterioration. Relatively normal EEG in FTD should therefore not be taken as an indication that behavioral features in these patients result from primary psychiatric causes.

A number of publications strongly support the utility of Q-EEG as diagnostic and prognostic tests for early Alzheimer disease. Collaborative work between centers in Stockholm and New York City, under the leadership of Christoph Lehmann, provided diagnostic efficiency statistics for EEG-based classification of mild Alzheimer disease from healthy controls of 85% sensitivity and 78% specificity. On the basis of available literature, EEG measures are likely to rapidly emerge as a cornerstone of the diagnostic workup for suspected Alzheimer disease.

The differential diagnosis of acutely disturbed and disorganized patients often includes delirium. In acutely agitated delirious patients, the EEG is often helpful in indicating whether the alteration in consciousness is due to (1) a diffuse encephalopathic process, (2) a focal brain lesion, or (3) continued epileptic activity without motor manifestations. Most often, patients with delirium have toxic metabolic encephalopathy. In general, with the progression of the encephalopathy, there is diffuse slowing of the background rhythms from alpha (8 to 12 Hz) to theta (4.0 to 7.5 Hz) activity. Delta activity (<3.5 Hz) usually does not become prominent until the patient approaches nonresponsiveness. The major exception to this rule is seen during withdrawal from alcohol and during delirium tremens. Instead of diffuse slowing, as described previously, excessive fast activity dominates the EEG in patients with alcohol withdrawal delirium. Patients in alcohol withdrawal who are not delirious could have a normal EEG. The low-voltage fast activity also can be induced by benzodiazepines, and this should be borne in mind when interpreting the EEG. The EEG thus can be helpful in differentiating between alcohol withdrawal delirium tremens and encephalopathies from other causes, including iatrogenic encephalopathy. The typical changes observed with delirium are slowing or even loss of the posterior dominant rhythm, slow-wave increase, reduced reactivity to eye opening, and intermittent, especially frontally located, rhythmic delta activity. These changes are thought to relate to the dopaminergic–cholinergic dysbalance as the final common pathway of delirium with a lack of cholinergic stimulation. In contrast to delirium cases revealing a cholinergic deficit as a common pathophysiologic feature, alcohol or benzodiazepine abstinence delirium results in decreased EEG voltage and an increase in fast activities, while nonconvulsive seizure state may be indicated by rhythmic activity and typical epileptiform discharges. Serial EEG assessments are considered the gold standard of delirium diagnostics.

Q-EEG is capable of detecting subtle changes and is especially useful for the evaluation of frail older adults with a high prevalence of dementia, accumulated organic brain disease, and age-related EEG changes. Using relative power measures, alpha decrease and delta increase, as well as the reduction in peak or mean frequency and diminished fast-to-slow-wave ratios, have been found to indicate delirium. In the presence of dementia, Q-EEG greatly improves delirium detection and monitoring, especially when supplemented with simple activation procedures.

Autism Spectrum Disorders. Several types of EEG abnormalities have been found in autism spectrum disorders (ASDs), including epileptiform activity, background slowing, and abnormal EEG hemispheric asymmetry. Altered mu rhythms (transient, nonoccipital alpha rhythms that occur over motor cortex) have also been reported, which some have postulated to indicate altered “mirror neuron” function, referring to neurons that fire both when an individual performs an action and when observing the same action performed by another. Of the various EEG abnormalities reported, epileptiform activity has been the most consistent finding in ASDs, detected at rates of 18.9% up to 60.7% of individuals with autism. Interestingly, approximately one-third of children with ASDs go on to develop epilepsy. That being said, several reviews have concluded that there is no justification for the use of anticonvulsant medications or surgery in children with pervasive developmental disorders without seizures; that is, there is no evidence that treatments to eliminate EEG spikes will have a therapeutic benefit on the behavioral abnormalities in autism.

EEG ALTERATIONS FROM MEDICATIONS AND DRUGS

A great many medications, as well as substances consumed for therapeutic or recreational purposes, can produce some degree of alteration in the EEG. This section attempts to highlight those compounds most relevant to clinical and research psychiatry.

Pharmaco-EEG

During the 1970s, several laboratories consistently obtained a classification of psychotropic drugs on the basis of EEG changes induced by a single dose routine in healthy subjects. Pharmaco-EEG methods are currently commonplace in preclinical studies to identify, at early stages of drug development, the therapeutic indications of new drugs, to test drug bioavailability or “target engagement” at the CNS level, to evaluate pharmacokinetic profiles based on correlation with drug plasma concentration, to predict therapeutically useful dosages, and compare the bioavailability of different formulation of drugs. Pharmaco-EEG studies identified the antidepressant activities of mianserin and doxepin, which had been classified as antiallergic and anxiolytic, respectively, by preclinical studies, as well as the sedative activity of fenfluramine, classified as a psychostimulant by animal studies. The discovery of the antidepressant activity of mianserin was instrumental in the development of new animal tests, which, in turn, promoted the discovery of new antidepressants, such as fluvoxamine (Luvox) and fluoxetine (Prozac). For fluvoxamine and sertraline (Zoloft), pharmaco-EEG studies enabled the identification of therapeutic doses in the early stages of drug development.

Few attempts have been made to translate pharmaco-EEG methods to psychiatric clinical settings, where this approach could inform the expectation for therapeutic benefit in response to a given psychotropic drug or dose level. Pharmaco-EEG methods for response prediction were applied in research, mainly involving the examination of Q-EEG changes following either a single administration of a drug (the so-called test-dose procedure) or at a point early in the course of treatment. Such methods, if proven capable of providing a reliable prediction of the therapeutic response to treatment, could have an enormous clinical impact when considering that failure to respond to initial treatment tends to increase attrition from care, the severity of prognosis, and ultimately the costs associated with psychiatric illness.

Regarding early prediction of treatment response, several QEEG studies of first-generation antipsychotics (FGA) in schizophrenia produced consistent findings but with ultimately no clinical impact. For instance, administration of high-potency FGA medication was found to produce an increase of alpha activity, particularly in the slow alpha range (7.5 to 9.5 Hz), that several independent groups found predictive of a favorable clinical response.

The test-dose procedure seldom has been used in pharmaco-EEG studies of response prediction in depressed patients. In one example, Q-EEG cordance (an index combining relative and absolute power) was measured during pretreatment and following 1 and 4 weeks of antidepressant therapy in a sample of treatment-resistant inpatients. Responders showed a decrease, while nonresponders showed an increase in prefrontal theta cordance after the first week of treatment.

These findings, in line with other brain imaging results, suggest that early changes in prefrontal activity mediate the clinical response to antidepressants.

In conclusion, Q-EEG indices could be a valuable complement to a patient's clinical assessment, providing further guidance to inform the clinician's choice of the appropriate drug treatment. Toward this end, guidelines have been published by the International Pharmaco-EEG Society (IPEG), with the goal of standardizing the acquisition and processing of data obtained in pharmaco-EEG studies. The introduction of these methods in clinical routine will require the replication of findings in large patient populations as well as the investigation of new drugs recently introduced in the clinical practice.

Psychotropic Medications—General Considerations

It is well known that psychotropic agents can affect the EEG. For the routine EEG, with the exception of the benzodiazepines and some compounds with a propensity to induce paroxysmal EEG discharges, there is little, if any, clinically relevant effect when the medication is not causing any toxicity. For example, benzodiazepines, even in small doses, generate a significant amount of beta activity that is seen diffusely and can last substantially longer (up to 2 weeks) than might be expected given the rapid response and peak occupancy (typically 1 to 4 hours) of these drugs. This response is so universal that it has been suggested that if a particular brain region fails to exhibit the expected benzodiazepine-induced beta activity, the area may be in some way damaged.

Psychotropic Medications—Toxic Effects. It has long been accepted that the EEG is sensitive to the neurotoxic effects of psychotropic medications, and clinical vignettes illustrating the value of EEG in detecting a neurotoxic reaction to medications commensurate with clinical deterioration have been reported. It is often stressed that such a scenario could occur at any time during the course of treatment because many factors impact the patient, and the symptoms could be subtle. An EEG investigation may be useful when patients undergoing long-term therapy present with clinical deterioration, particularly if the patient is known to be taking the medication and the serum plasma levels are within the therapeutic range. Such clinical situations also may highlight the need for having baseline EEGs available for comparison when a patient presents with clinical exacerbation. Normative EEG databases are available but are based on cross-sectional evaluations and do not take into account the dynamic nature of psychiatric disorders or changes due to medication status. Thus, for the purpose of detecting toxicity, routine clinical EEG remains of immediate relevance to the everyday management of patients prescribed psychotropic medication. The appearance of significant diffuse EEG slowing in a patient who is receiving psychotropic medications and whose clinical condition is not stable (particularly the older adults) should prompt the clinician to consider medication toxicity, in addition to other causes of encephalopathy (e.g., electrolyte imbalance and thyroid problems, to name only two).

Psychotropic Drug-Induced EEG Abnormality (Nonparoxysmal and Paroxysmal). Almost from the time psychotropic medications were introduced, it was known that some of these compounds could precipitate EEG abnormalities, including paroxysmal EEG discharges (spike and spike-wave activity) in some individuals. Usually, medication-induced paroxysmal EEG activity remains behaviorally silent and is not accompanied by iatrogenic overt seizure manifestations.

GENERAL ANESTHETICS. General anesthetics primarily induce their effects by binding to affecting neuronal ion channels. In terms of EEG signatures, anesthetics initially evoke a frontal fast activity followed by a disruption of alpha rhythms. As the anesthesia deepens, EEG activity becomes slower and of low voltage (not unlike deep sleep), and burst-suppression patterns can be observed. At the deepest levels, EEG activity ceases and can even resemble patterns seen in comatose states.

BARBITURATES. Barbiturates are now minimally used for medical or recreational purposes. The mechanism of action of this class of drugs involves binding to a specific site on the $GABA_A$ receptor, thus potentiating the effects of the inhibitory neurotransmitter GABA. When barbiturates are taken for medical or recreational purposes, beta activity is introduced into the EEG, particularly over frontal regions. Higher doses are associated with lower frequencies and dissolution of the alpha rhythm (as drowsiness and sleep ensue). Abrupt withdrawal from barbiturates after long-term exposure can produce EEGs containing generalized paroxysmal activity and spike discharges, which often are not associated with overt motor seizure manifestations. These effects may be seen for days up to 3 to 4 weeks after the last drug exposure.

TRADITIONAL NEUROLEPTICS. The primary mechanism of action of the typical antipsychotics/neuroleptics (e.g., chlorpromazine, haloperidol) is the antagonism of the D2 dopamine receptor. EEG changes with the neuroleptics include an increase in alpha activity, a shift to lower frequencies, and a decrease in beta activity. At high doses, seizures and spikes may occur.

ATYPICAL ANTIPSYCHOTICS. Atypical antipsychotic agents such as clozapine and risperidone have different mechanisms of action compared to traditional neuroleptics and have an affinity for 5HT2a, alpha-adrenergic, and D4 receptors, among others. EEG frequency changes during the administration of clozapine include increases in delta, theta, and beta, which are prominent over frontal, central, and parietal areas. Some paroxysmal slowing and spikes can occur, along with rare instances of seizure. Little is known regarding the EEG correlates of other atypical antipsychotic drugs.

TRICYCLIC ANTIDEPRESSANTS. Older generation tricyclic antidepressants (TCAs) (e.g., imipramine, doxepin) typically act as both serotonin and norepinephrine reuptake inhibitors. They can increase the amount of slow and fast rhythms, decrease the frequency of alpha, and can induce paroxysmal slow waves and spikes. At higher doses, they can induce seizures, which can be particularly concerning given that tricyclics have a somewhat narrow therapeutic range.

SELECTIVE SEROTONIN REUPTAKE INHIBITORS. By definition, selective serotonin reuptake inhibitors (SSRIs) increase synaptic levels of serotonin by binding to the serotonin transporter. These agents tend to decrease alpha activity while increasing theta frequencies. Further, the rates of seizures are less in SSRIs compared to tricyclics.

LITHIUM. Lithium is still commonly used in the treatment of bipolar disorder and can induce marked changes in the EEG. Lithium salts, particularly at intoxicating levels, are associated with diffuse EEG slowing and paroxysmal spikes, and triphasic waves, which can remain after serum levels have normalized.

BENZODIAZEPINES. Like the barbiturates, benzodiazepines potentiate the action of GABA, although the mechanism of action entails binding to a separate benzodiazepine site on the $GABA_A$ receptor. As such, benzodiazepines are commonly used for their anxiolytic and anticonvulsant properties. Benzodiazepines at therapeutic doses are potent enhancers of beta rhythms and, in addition, slightly increase theta power. Decreases in alpha frequencies are also noted. Benzodiazepines are well known to reduce epileptiform activity and, indeed, are the first drugs of choice for status epilepticus.

OTHER THERAPEUTIC AGENTS. While beyond the scope of the current chapter, a host of other nonpsychiatric therapeutic agents have been shown to induce robust effects on the EEG. Most notably, EEG changes in the context of antiepileptic drug treatment are well documented. Further, antibiotics, immunosuppressants, and antineoplastic agents, among others, have all been shown to affect EEG signatures. For a thorough treatment of these medication types, see referenced chapter by Bauer and Bauer.

Drugs of Abuse

The recreational use of addictive drugs remains a significant public health issue. In fact, in terms of the cost of illness, it has been estimated that drug addiction ranks second only to cardiovascular disease. Hence, drug dependence is becoming of increasing concern for those involved in the assessment and treatment of psychiatric disorders. Nearly all drugs of abuse are capable of altering the frequency spectrum of the EEG, and the degree of alteration varies with recreational versus heavy use and whether the EEG was obtained during acute intoxication, during intervening periods of abstinence, or during clinical withdrawal in the addicted individual.

With only infrequent exceptions, the use of an abused drug does not introduce frank clinical abnormalities into the visually analyzed EEG tracing. This is especially true for recreational drug use and is even the case for dependent and addictive use as well. For this reason, drug abuse alone is not a sufficient reason for EEG referral. Although the alterations of EEG frequency and voltage in the visually analyzed EEG produced by many abused drugs are often unimpressive, topographic Q-EEG analyses can reveal marked alterations of the EEG spectrum that constitute significant deviations from population norms (even though the clinical implications may not be clear). Nonetheless, it is important to appreciate the various effects of different drug classes on the EEG.

Alcohol. Ethyl alcohol, or ethanol, exhibits many of the same pharmacodynamics as benzodiazepines (i.e., acting as a positive allosteric modulator of the $GABA_A$ receptor). Hence, many of its behavioral and EEG effects are similar. There is considerable consensus that an increase in the amount of alpha activity and a slight slowing of alpha frequency typically accompany alcohol consumption and that higher blood alcohol levels increase the amount of theta. Some reports have indicated that chronic alcohol consumption may be associated with a lower voltage and slightly faster resting EEG, although the clinical relevance of this remains obscure.

Alcohol withdrawal can lead to epileptic seizures as a result of perturbations of GABA and glutamate neurotransmitter/receptor function. Higher-frequency beta activity may be substantially increased in the addicted alcoholic undergoing withdrawal, and if delirium tremens complicates the clinical picture, then excessive fast activity may dominate the EEG. EEG signatures during acute alcohol withdrawal can be desynchronized with low voltage, and generalized spikes may be observed. Alcohol-related seizures need to be differentiated from true epileptiform activity with comorbid alcohol dependence. Routine EEG can be extremely helpful in this regard. A normal EEG during periods of sobriety, particularly if associated with an abnormal EEG during early withdrawal, strongly suggests that the seizures are withdrawal induced. It should be noted that the EEG tends to normalize faster (with the exception of a generalized decrease in amplitude) during withdrawal from alcohol as compared to that from barbiturates.

Opiates. The effects of opiates such as morphine and heroin on the EEG are similar to those of alcohol and involve slight reductions in alpha frequency. They also may increase the voltage of the EEG, particularly the power of theta and delta activity. Opioid anesthesia has also been shown to induce spike discharges resembling benign epileptiform transients of sleep. However, when an opiate overdose produces a comatose clinical presentation, the EEG usually consists of clinically abnormal diffuse slowing.

Cannabis. Marijuana, or cannabis, remains one of the most prevalent drugs of abuse. Moreover, rates of its consumption are on the rise due to the trend of decriminalization and the acceptance of the use of “medical marijuana.” Cannabis, or its primary psychoactive constituent delta-9-tetrahydrocannabinol (THC), affects the brain via central cannabinoid receptors (CB1Rs). CB1Rs act primarily by affecting the release of the neurotransmitters GABA and glutamate. As reviewed by Skosnik and Cortes (2016), the effects of cannabis/THC on EEG can be broken down into effects on resting EEG (intrinsic rhythms), ERPs, and evoked/induced oscillations.

Acute cannabis smoking typically produces minimal visible effects on traditional EEG. In terms of contemporary studies utilizing Q-EEG, chronic cannabis users typically show reductions in power in the alpha-, theta-, and beta-frequency bands while in the nonintoxicated state. Acute THC has been shown to reduce resting theta power, which has been correlated with altered memory performance. Both chronic and acute cannabis/THC has been shown to disrupt the P300 ERP, which is congruent with cannabis’ effect on attention and memory. Finally, both chronic and acute cannabis/THC has been shown to disrupt sensory-evoked and induced gamma rhythms, which may provide a mechanism for the perceptual-altering effects of cannabinoids.

Finally, it should also be noted that one of the other nonpsychoactive constituents in cannabis (cannabidiol) has been shown to have antiepileptic properties. Hence, this may represent a potential future direction in the “medicinal” use of cannabinoids.

Psychostimulants. Psychostimulants such as cocaine, amphetamine, and methylphenidate act primarily to increase central dopamine function. They typically increase the amounts of beta and alpha activity and generally reduce EEG voltage. Overdose with psychostimulants produces typical coma patterns in the EEG. The effect of psychostimulants on ERPs has yielded mixed results, although they may exhibit some effects similar to those observed in disorders of dopamine function such as schizophrenia.

Inhalants. Inhalation abuse of volatile substances (e.g., airplane glue, cleaning fluid, paint thinner, and gasoline) can produce a nearly instantaneous sensation of euphoria, and, in the early period of use, there may be no obvious residuals after the acute response subsides. However, with continued inhalant abuse, serious neurologic and neurocognitive deficits can emerge that may not be completely reversible with abstinence. The immediate effects of inhalation of volatiles on the human EEG appear not to have been well studied. Where persistent neurologic or neurocognitive sequelae follow chronic inhalant abuse, clinically abnormal diffuse EEG slowing in the lower theta to upper delta range may be seen.

Hallucinogens. Drugs such as lysergic acid diethylamide (LSD) and mescaline appear to have only minor effects on the visualized EEG and do not produce clinically relevant changes. However, recent studies of the drug ecstasy (MDMA) have shown abnormalities in several ERPs, which may be related to the purported memory-impairing effects of long-term MDMA use.

Tobacco. Tobacco (nicotine) does not appear to produce dramatic alterations in the analog EEG. However, topographic Q-EEG analyses reveal striking EEG changes with acute exposure to, as well as withdrawal from, tobacco. The immediate effects of smoking include a decrease in slower frequencies (especially theta), increased power of frequencies in the upper one-half of the alpha-frequency band, and beta activity. Twenty-four hours of tobacco deprivation produce a marked decrease in alpha frequency, with a corresponding marked increase in the relative power (amount) of theta activity. The effects of acute smoking and abstinence are essentially opposite to one another. Further, nicotine has been associated with alterations in “sensory gating,” as indexed by studies of P50 ERP suppression. However, these effects can vary depending on whether they are examined during acute nicotine exposure use or during withdrawal from nicotine.

Caffeine. The use of coffee and products containing caffeine are quite prevalent. The use of caffeine is of little concern to the clinician interpreting the visually analyzed EEG. Withdrawal from caffeine in the caffeine-dependent individual, however, produces a markedly significant increase in the amplitude, or voltage, of theta activity—an effect that is reversible within 15 minutes of consuming one cup of coffee.

Other Drugs of Abuse. In recent years, several novel synthetic drugs of abuse have emerged, most notably “K2/spice,” which are synthetic cannabinoids, and “bath salts,” which are synthetic psychostimulants/hallucinogens typically in the cathinone family. However, little research has been performed on these substances, particularly in the context of EEG.

EEG AND EXPERIMENTAL RESEARCH PARADIGMS

The majority of EEG conducted in the field of psychiatry may indeed be for research purposes, including studies both on the intrinsic rhythms ascertained under normal resting conditions as well as a number of experimental paradigms associated with different sensory–perceptual and cognitive tasks. The high temporal resolution of EEG essentially provides “real-time” measurement of changes in brain activity in response to experimental conditions. EEG allows the experimenter to study changes in brain activity related to overt processes, such as differences in the auditory response to frequently versus infrequently presented stimuli, as well as more covert processes ranging from the anticipation of stimulus delivery (e.g., contingent negative variation; CNV) to presupplementary motor area activation (e.g., lateralized “readiness” potential; LRP) in preparation of a button press or consolidation–retrieval processes associated with working memory following presentation of a stimulus set. As illustrated by these examples, EEG is an ideal tool for assessing differences or changes in neural activity generated as a sequence of mental operations over a time course. Regardless of whether or not such functions can be localized to discrete cortical or subcortical regions, which is never the first aim of EEG research, the ability to precisely record changes in neural activity in response to known events (e.g., under controlled test conditions) enables the researcher to ask, and potentially answer, a limitless set of questions concerning human information processing and abilities affected by psychiatric illness.

Evoked and Event-Related Potentials

The genesis of experimental EEG research no doubt has roots in *EP* methods, which continue to see routine clinical use in neurology practice as a test of CNS function. Briefly, EPs are electrophysiologic responses to external stimuli, usually administered over several repetitions with changes in stimulus properties (e.g., frequency, intensity) across presentations. EPs can be generated in any sensory modality but in most clinical applications involve *visual-evoked potentials* (VEPs), and *brainstem auditory-evoked potentials* (BAEPs) measured by EEG, or *somatosensory-evoked potentials* (SSEPs) measured electromyographically (i.e., EMG). An important point to keep in mind regarding the EP is that the response is time locked to the delivery of a stimulus and, therefore, can be strictly regarded as “evoked.” This point distinguishes the EP from the conceptually and semantically similar *ERP*, which also represents an electrophysiologic response to stimulation. However, the subtle difference that distinguishes these measures can be appreciated in the examples of so-called overt versus covert brain processes provided in the preceding paragraph; whereas the overt auditory-evoked response can be regarded as an EP, examples of ERPs including the CNV and LRP occur prior to stimulus delivery and, therefore, reflect brain activity engaged in the process of performing mental operations that is not necessarily time locked, or “evoked” in response, to a stimulus. The term “event-related” has therefore replaced “evoked” over time in the research arena, where EEG recording is used to study both “bottom-up” sensory-evoked responses as well as changes in neural activity associated with higher-order cognitive processes that may be represented by induced brain signals.

Another basic principle regarding the EP or ERP is that these potentials are usually of small magnitude (as small as a few microvolts) compared to the ongoing EEG activity (reaching 50 μV or more). This is due to the fact that EPs/ERPs represent only the components of the EEG that are consistently present across repeated trials of stimulus delivery, which may consist of hundreds of trials. The EP/ERP is extracted from the digital recording by averaging, resulting in a signal that reflects a change in neural activity in response to stimulus events with ongoing background EEG activity (which is not time locked to the stimuli) canceled out by the averaging process. The resulting average is then graphed as a change in voltage over time plots and can be quantified by measures of peak amplitude and latency.

By convention, BAEPs are labeled with the Roman numerals I, II, III, IV, and V, reflecting the order of the component, with components I to III generated by the auditory branch of cranial nerve VIII and components IV and V by the upper brainstem. ERPs (as well as SSEPs and VEPs) are labeled according to both their polarity and latency. The letters “P” or “N” indicates the positive or negative direction of deflection relative to the prestimulus baseline, though whether the polarity is depicted graphically in an up or down direction seems to be a matter of personal choice and, in many cases, polarity does reverse depending on electrode location due to the dipolar distribution of associated electrical fields. Labeling nomenclature representing ERP peak latency is done more simply, with exceptions. While the ERP associated with the label “P300” is a positive deflection peaking at around 300 ms following stimulus delivery, the shorthand “P3” can also be used, signifying the third positive deflection of the ERP, which follows P1 or P100 at around 100 ms and P2 or P200 at around 200 ms. Time courses and conventional labeling of ERPs most commonly studied in psychiatry are presented in Figure 1.26–16.

When a stimulus is presented to the brain, it travels through the receiving sensory organ and the brainstem (except for visual and olfactory stimuli) on its way to the primary cortical regions representing the particular sensory input. EPs recorded over primary sensory cortices indicate how the stimulus was received and registered by the brain, thereby providing a test of the integrity of the underlying neural architecture. For instance, the BAEP is measured in EEG waves arising within the first 10 ms following delivery of an auditory stimulus, thought to reflect the intactness of the auditory pathway, and most are useful for the diagnosis of hearing loss. It cannot be stated with certainty at this time whether basic EPs are broadly or specifically affected by psychiatric disorders, as ERPs have been the focus of most contemporary neuropsychiatric research. However, disturbances in early sensory processes are thought to be fundamental to schizophrenia pathophysiology, for example, and may conceivably involve impairment detected at the earliest stages of sensory processing.

Regarding the stage of information processing, a basic delineation has been made between *early-latency potentials,* which would include the BAEP and other components occurring before 10 ms poststimulus, *mid-latency potentials* occurring up to 200 ms, and *long-latency potentials* occurring after 200 ms. As a general rule of thumb, early-latency potentials are thought to mainly reflect processing dependent on *exogenous* factors determined outside of the CNS, such as stimulus intensity or rate. Exogenous potentials can be measured regardless of a subject’s level of consciousness, are thought to be fundamentally unaffected by cognitive processes, and represent basic sensory registration. Once received by the brain, a stimulus undergoes psychological processing and evaluation, and these operations influence later ERP waves that are broadly distributed outside of primary sensory cortices. While ERPs also contain exogenous components (though typically recorded at sampling rates too slow to accurately measure early-latency potentials), mid-latency and long-latency potentials tend to be of most interest to experimental research. In later stages of information processing, ERPs increasingly reflect activity relating to cognitive processes, such as anticipation, encoding, and recognition, thus can be influenced by the level of alertness and attentional resources devoted to the experimental task. Accordingly, mid-latency and late ERPs mainly reflect *endogenous potentials,* with measurements of amplitude or other characteristics reflecting the interaction of stimulus properties and experimental task demands with individual differences in information processing. Because ERPs are sensitive to experimental manipulations, various paradigms have been developed to study brain function at the interface of sensory and cognitive processing. These paradigms have demonstrated many abnormalities during mid-latency and later processes across the psychiatry research literature and thus are of considerable interest for their potential as clinical tools.

Mid-Latency Components (P50 and N100)

Three of the most examined ERP components in the auditory modality are designated P50 (a positive component occurring between 35 and 80 ms after stimulus onset), N100 (a negative component occurring between 80 and 150 ms after stimulus onset), and P200 (a positive component occurring between 150 and 250 ms after stimulus onset). Figure 1.26–16 shows a typical example of these mid-latency ERPs. These components share the characteristic that their amplitude decreases with repetition (habituation or sensory gating). The P50 and N100 components have been studied extensively in psychiatric patients.

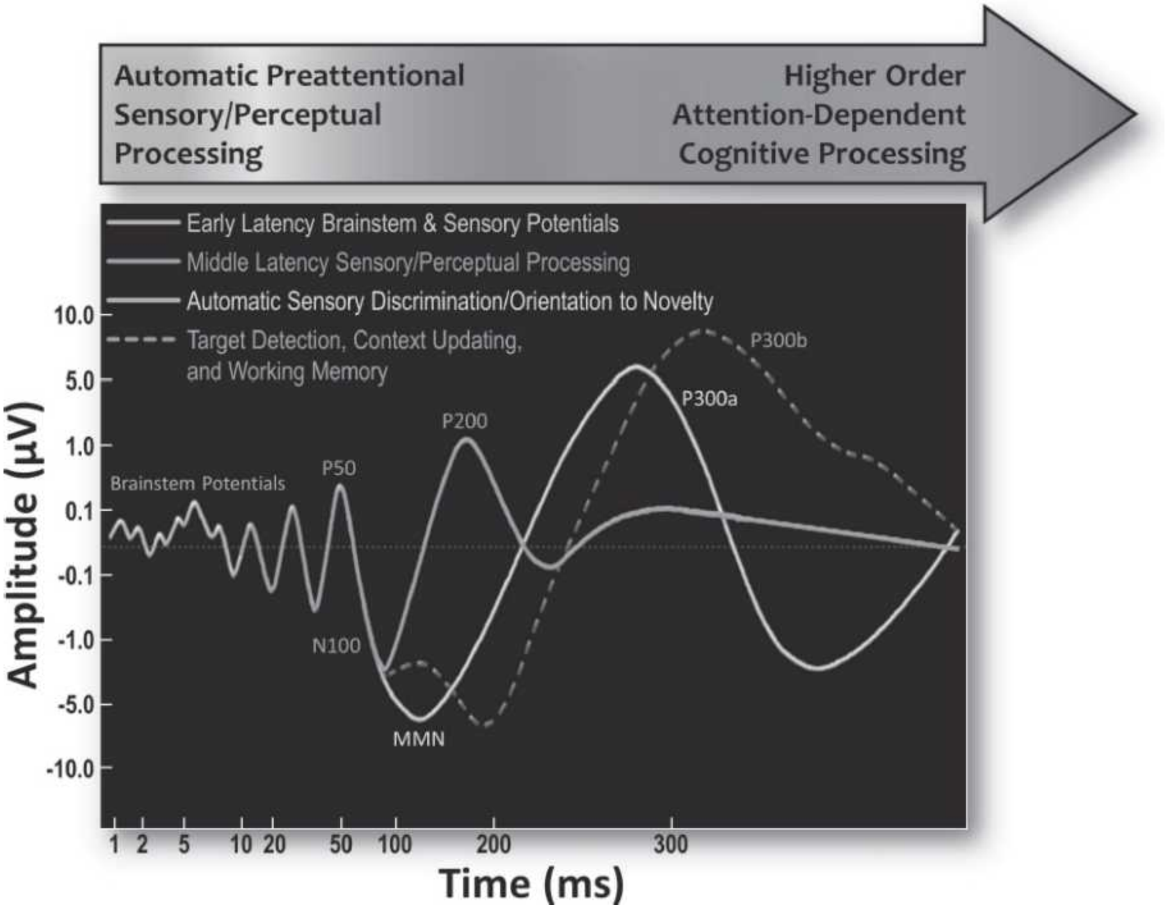

FIGURE 1.26–16. Schematic illustrating prototypical ERPs. Time is shown using a logarithmic scale. Stages of information processing from automatic preattentional to higher-order attentional processes are shown from left to right, respectively. (Adapted from Skosnik PD, Cortes-Briones JA. Electroencephalography (EEG) and cannabis: from event-related potentials (ERPs) to oscillations. In: Preedy VR, ed. *The Neuropathology of Drug Addictions and Substance Misuse*. Academic Press; 2016, pp. 851–862, with permission from Elsevier.) (See color panel.)

The P50 is a positive voltage, mid-latency ERP component peaking at around 50 ms poststimulus, best elicited by discrete auditory stimuli (e.g., brief white noise clicks). In the standard P50 gating or “dual-click” paradigm, the amplitude of the P50 to a second paired click (S2) is attenuated relative to the P50 amplitude to the first click (S1). A 500-ms interstimulus interval is typically utilized and is optimum to elicit gating. While the P50 to S1 is thought to be related to the capacity of the brain to register salient stimuli, the reduction of the P50 amplitude to S2 (compared to S1) reflects the automatic suppression of inhibitory gating of redundant and irrelevant stimuli. It is thought to be mediated by the hippocampus, temporoparietal region, auditory cortex, and prefrontal cortex. Deficient P50 suppression is commonly associated with schizophrenia and has been proposed as a candidate endophenotype for psychosis risk (representing a physiologically detectable feature linked to the underlying genetic cause of illness), supported in part by evidence for linkage to several psychosis risk genes and by a similar reduction in P50 suppression observed in first-degree family members of schizophrenia probands. Evidence for an inhibitory abnormality (i.e., sensory gating deficit) in schizophrenia patients has been accumulating over the past 30 years. In an early meta-analysis of proposed endophenotypes of schizophrenia, sensory gating of the P50 response was found to represent a magnitude of deficit equal to or greater than the most robust findings reported across neuroimaging and behavioral studies.

The N100 component is a large exogenous ERP that exists regardless of task demands (although it can be modulated by attention). The N100 is generally associated with early perceptual processing with origins in the primary sensory cortices and is elicited in response to various forms of auditory and visual stimuli, regardless of the experimental context. Of note, while the visual N100 is regionally localized over occipital and posterior parietal electrode sites, the N100 (like all auditory ERPs) tends to be maximal over the central or frontocentral sites, representing the confluence of potentials generated in bilateral auditory cortices. The N100 component has also been shown to exhibit suppression similar to the P50 ERP when assessed using the dual-click paradigm and is similarly affected in schizophrenia patients. Work by Johannesen and colleagues showed that reductions in N100 amplitude in schizophrenia, whether elicited by paired clicks or by pure tones presented at shorter interstimulus intervals (∼1,200 ms apart), appear to have equally large effect sizes when contrasted against healthy comparison subjects, thus appear more dependent on internal neural processing than on the stimulus features used to elicit the auditory response. In the same study, while P50 and P300 measures failed to differentiate schizophrenia from bipolar disorder, the N100 provided unique variance in classifying these psychiatric groups with reasonable (72%) accuracy. Of note, the classification of schizophrenia from bipolar participants based on N100 was only slightly less accurate than the classification of schizophrenia against matched healthy participants (79% accuracy) using an optimal multivariate model consisting of several ERP and spectral-frequency measures. Therefore, auditory processes indexed by the N100 appear to represent a consistent (e.g., obtained in different experimental contexts), robust, and perhaps diagnostically relevant neurophysiologic deficit among individuals with schizophrenia.

P300

The P300 is a late positive, postattentional ERP component thought to be related to directed attention, contextual updating of working memory, and the attribution of salience to deviant or novel stimuli. This response is typically elicited utilizing standard “oddball” paradigms in which low-probability target stimuli (~10% to 20%) are embedded within a repeating sequence of high-probability standard stimuli (~80% to 90%) differing in some physical dimension (e.g., frequency, duration, amplitude). It this thought that the P300, particularly the P300b component (elicited in response to designated targets), reflects activity from a distributed neural network including areas such as the thalamus, hippocampus, inferior parietal lobe, superior temporal gyrus, and frontal cortex. Cognitive and affective neuroscientists have been conducting basic research on P300 since it was discovered over 40 years ago, providing a rich knowledge base that has yet to be fully exploited by clinical scientists.

Mismatch Negativity

The MMN is an automatic, preattentive, negative voltage ERP component that is generated primarily in the superior temporal and prefrontal cortex. The MMN is hypothesized to index the automatic, preattentive processing of the “mismatch” between standard and periodic deviant stimuli, indexed by the amplitude of a negative component over latencies of 100 to 200 ms measured in waveforms created by the subtraction of ERPs produced by the deviant and standard stimuli. The MMN experiment is similar to the P300 oddball paradigm, where the serial presentation of a standard stimulus (e.g., 10-ms presentation of 1,000 Hz tone) is occasionally interrupted by deviants that differ in subtle ways (e.g., 10-ms presentation of 1,500 Hz tone, 50-ms presentation of 1,000 Hz tone). However, unlike the P300, attention to the stimuli is not required and is normally redirected from the primary auditory discrimination task by a concurrent, noninterfering task (e.g., reading, visual target detection). Also, unlike the P300, which is sensitive to both auditory and visual stimuli, MMN has been mostly studied as it relates to acoustic information processing.

Because it is virtually sensitive to any feature of an acoustic stimulus, including frequency, duration, and learned features such as rhythmic patterns and timing of stimulus delivery, the MMN has been especially useful in the evaluation of problems associated with specific aspects of auditory processing. Deviations in MMN are reported across several psychiatry disorders, including depression, alcoholism, and dementia, but have received the most attention in schizophrenia.

In a longitudinal investigation of first-hospitalized schizophrenia patients, Salisbury and colleagues (2007) showed that reduction in the MMN over a 1.5-year period following the first hospitalization was associated with progressive reductions in Heschl gyrus gray matter volumes in schizophrenia but not in psychotic bipolar or healthy comparison subjects, who evidenced no significant change in or correlation between MMN and gray matter measures. This important finding links an experimental electrophysiologic process to the brain structure and highlights the sensitivity of MMN to neurodegenerative processes that may be uniquely involved in the etiology of schizophrenia.

Error-Related Negativity

The error-related negativity (ERN), typically studied in a choice reaction time task, occurs approximately 60 to 150 ms after an incorrect choice has been made. The ERN has been posited to reflect sensitivity to error detection, the remedial actions that are taken to compensate for the error, and the conflict generated by the process. Time–frequency analysis indicates that the ERN is mainly comprised of anterior theta activity that appears to originate from the ACC, part of a neural system involved in behavioral regulation and self-monitoring. Because it appears to involve both affective and cognitive information processing and evaluation, it is not surprising that the ERN is also sensitive to clinical features of various psychiatric disorders, including OCD and other anxiety disorders, major depression, and externalizing psychopathology related to behavioral disinhibition. The findings of these studies suggest that anxious individuals because they generate larger ERNs, are more sensitive to making errors, while depressed individuals and those with disorders associated with behavioral disinhibition are less so. These results are generally consistent with expectations given the clinical characterization of these disorders.

SUMMARY AND FUTURE DIRECTIONS: EEG IN THE SYSTEMATIC STUDY OF PSYCHOPATHOLOGY

It is now firmly established that EEG anomalies, including ERPs, are associated with a wide variety of psychiatric conditions, including schizophrenia, mood disorders, anxiety, ADHD, and substance abuse disorders, which a voluminous research literature associated with each.

Because ERPs are derived from the EEG, it is possible to use time–frequency analysis to identify the constituent EEG spectra, which, because they are time locked to the stimulus presentation, compose the ERP waveform. This type of analysis, in effect, renders a possible determination of the amount of each EEG frequency present at each point in time, elapsing from the onset of a stimulus to the resolution of the associated ERP. These transient-evoked oscillations (Fig. 1.26–13) can be partitioned into the same frequency bands identified for spontaneous/resting EEG but represent stimulus-elicited rhythms that are likely to serve roles that are functionally different from those associated with their appearance in resting EEG. For example, while higher resting low frequency (e.g., delta, theta) is associated with pathology in a number of clinical conditions, higher activation of low-frequency activity under task-related conditions (e.g., working memory) is often interpreted to indicate better brain function. With respect to the P300 ERP, higher amplitude (better function) is mainly dependent on the activation of delta and theta frequencies, accompanied by a reduction in alpha frequency.

Using one example, transient-evoked oscillations are beginning to provide insights regarding brain dynamics and psychopathology, especially regarding the development of alcohol dependence. Reduced delta- and theta-evoked oscillations have been associated with risk for the development of alcoholism over and above the risk associated with reduced P300 amplitude, suggesting that the decomposition of the P300 ERP into constituent oscillatory components may improve the sensitivity of this measure over the band-passed signals (typically filtered at 0.5 to 12 Hz or as high as 24 Hz) characterized by ERPs. These evoked oscillations have also been linked to gene variants involved in cholinergic and GABAergic neurotransmitter systems, which have been implicated in the neurochemical pathways involved in many psychiatric conditions. This line of ERP or evoked oscillation research conducted in alcoholism exemplifies an evolving scientific effort to use experimental electrophysiologic assays of human information processing as fundamentally biologic, and perhaps genetically related, indicators of risk for the development of psychiatric disorders.

TRANSLATION OF RESEARCH FINDINGS TO CLINICAL APPLICATIONS

In the absence of a guiding system or accepted approach for evaluating and integrating new measurement approaches to clinical decision making, EEG findings that carry significant promise are likely to continue to remain interesting research findings discoverable only by investigators who intend to embark on similar research. Without translating such findings into diagnostically robust, clinically meaningful, and psychometrically validated procedures, the evolution of EEG in psychiatry appreciated by the academic community is unlikely to translate to clinical applications. We review and comment on the guidance provided in 2007 by Nash Boutros and Cynthia Arfken as illustrative of steps taken to psychometrically develop a diagnostic tool:

Step 1: A biologic variable is observed to be deviant from healthy controls in a particular patient population. The demonstration of test–retest reliability of the finding using blinding procedures is an essential component of this early step. The goal is to refine a measure and the procedures used to obtain it to a standard, assuring the measure is repeatable and generalizable across intended use scenarios. Such an endeavor cannot feasibly be undertaken by a single laboratory or even a small collection of research sites and requires a cooperative effort among large groups of investigators committed to harmonizing measurement methodology and who are able to generate the financial resources required to support such an effort. In practice, the intention would be to obtain samples diverse enough to represent demographic, regional, and perhaps international differences required to estimate meaningful population parameters. The NAPLS consortium is one example of such an effort, involving eight study cites across the United States and one in Canada, who collectively applied for grant funding under a common mechanism, and coordinated measurement efforts led by experts of each site who informed the standardization of test protocols across sites.

Step 2: Demonstration of the potential clinical usefulness of the specific finding. The two most important objectives at this step are a demonstration of the difference between the target patient population and appropriate control groups (these should be groups of patients with diagnoses that commonly appear on the differential diagnostic lists of the target disorder). Estimation of the effect size of the finding could be a reasonable guide to which findings should be considered good candidates for Step 3 studies. Of note, guidance on clinical meaningfulness may take several forms, depending on the criteria by which "meaningfulness" is demonstrated. While separation of diagnostic entities fulfills basic criteria concerning the typicality of a diagnostic feature as representative of a broader class, the "meaningfulness" of biomarkers alone remains a highly contested issue between drug developers and federal regulatory authorities, particularly in cases of disorders of the CNS. Accordingly, additional consideration should be given to precedent provided by current medical standards. For instance, if a new technology (e.g., remote ambulatory EEG monitoring) is demonstrated to perform as well as a clinical standard for diagnostic surveillance (e.g., clinic-based sleep study), the path to development is more straightforward. In examples where EEG is considered useful as a prognostic indicator, such as early detection of illness or as a measure of early response to treatment, criteria for clinical meaningfulness would be predicated not only on the presence of the marker in a known disease state but on its predictive power to inform future outcomes.

During Step 3: The performance characteristics of the test should be established. Specifically, the sensitivity, specificity, and positive and negative predictive values of the biologic marker should be examined. These criteria consider not only the measures' ability to accurately identify the intended index cases (e.g., diagnostic group, treatment responder, those who transition from "at-risk" stage) with high sensitivity (positive marker indicates condition) but also to accurately discriminate these cases from those who belong to other classes (negative marker indicates the absence of condition), reflecting a measure's specificity to a classification outcome. Depending on the intended use and risk–benefit of the decision the classifier is used to inform, preference may be given to measures that optimize sensitivity over specificity or vice versa.

Step 4 defines the clinical application of the test and helps standardize the technique used in large and multicenter clinical trials. Multicenter trials should guide methodology for standardization of laboratory procedures used to conduct the test as well as obtaining data regarding cost-effectiveness and impact on both short-term and long-term clinical outcomes. It is strongly recommended that studies describing efforts toward developing diagnostic testing should adhere to the guidelines proposed by the Standards for Reporting of Diagnostic Accuracy (STARD) initiative (see https://www.equator-network.org/reporting-guidelines/stard/).

Given this rubric for the development of clinical diagnostic and/or endpoint measures, which is considered appropriate for the translation of EEG research methods to clinical tools, it should also be appreciated that when undertaken in isolation of broader stakeholders informing health care policy, these efforts may fall short in scope or support required for acceptance and integration with current standards of care. In this way, effective translation will also benefit from the input of insurance payers, patient advocacy groups, private industries (e.g., EEG manufacturers, pharmaceutical companies), and federal regulatory authorities such as the Food and Drug Administration (FDA), whose remit is to protect public health by ensuring the safety and efficacy of drugs as well as medical devices. The interested reader is directed to publications provided by the FDA Critical Path Initiative. Launched in 2004, the CPI is FDA's national strategy for transforming the way FDA-regulated medical products are developed, evaluated, and manufactured. At this time, few initiates involving EEG are under consideration. That said, the FDA recently provided 510(k) clearance for a remote EEG wearable device that enables physicians to monitor the EEG of individuals from the home as they continue their daily activities. FDA clearance pertains to examples where a new device manufacturer is able to demonstrate that the product is "substantially equivalent" to a similar market device that has already been approved and requires the device manufacturers to notify the FDA of their intent to market a medical device within a specified time frame. The path to approval of a new device or measure with no market equivalent (i.e., predicate) is more complex by comparison, requiring that, through clinical testing results, the agency can conclude that the benefits of the test outweigh the known risks. A current effort to establish EEG measures as drug development tools, involving a collaboration of pharmaceutical industry sponsors invested in CNS drug development, academic leaders, and an EEG device manufacturer, is provided here as another example of what, through careful science, may be on the horizon for EEG in psychiatry: ERP Biomarker Qualification Consortium (https://erpbiomarkers.org/).

FURTHER READINGS

Bauer G, Bauer R. EEG, drug effects, and central nervous system poisoning. In: Schomer DL, Lopes da Silva FH, eds. *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins; 2011.

Boutros NN, Mirolo HA, Struve F. Normative data for the unquantified EEG: examination of adequacy for neuropsychiatric research. *J Neuropsychiatry Clin Neurosci*. 2005;17:84–90.

Boutros NN, Arfken CL. A four-step approach to developing diagnostic testing in psychiatry. *Clin EEG Neurosci*. 2007;38:62.

Chang BS, Schomer DL, Niedermeyer E. Epilepsy in adults and the elderly. In: Schomer DL, Lopes da Silva FH, eds. *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins; 2011.

Cong F, Ristaniemi Y, Lyytinen H. Advanced Signal Processing on Brain Event-Related Potentials: Filtering ERPs in Time, Frequency and Space Domains Sequentially and Simultaneously. World Scientific; 2015.

Grant AC, Abdel-Baki SG, Weedon J, et al. EEG interpretation reliability and interpreter confidence: a large single-center study. *Epilepsy Behav*. 2014;32:102–107.

Greenfield LJ, Geyer JD, Carney PR. *Reading EEGs: A Practical Approach*. 2nd Edition. Lippincott Williams & Wilkins; 2020.

Johannesen JK, O'Donnell BF, Shekhar A, McGrew JH, Hetrick WP. Diagnostic specificity of neurophysiological endophenotypes in schizophrenia and bipolar disorder. *Schizophr Bull*. 2013;39:1219–1229

Michel CM, Brunet D. EEG Source Imaging: A Practical Review of the Analysis Steps. *Front Neurol*. 2019;10:325.

Narayanan B, Soh P, Calhoun VD, et al. Multivariate genetic determinants of EEG oscillations in schizophrenia and psychotic bipolar disorder from the BSNIP study. *Transl Psychiatry*. 2015;5(6):e588.

Nunez PL, Srinivasan R. *Electric Fields of the Brain: The Neurophysics of EEG*. 2nd ed. Oxford University Press; 2006.

Pang T Selvitelli M, Schomer DI, Niedermeyer E. Metabolic disorders in EEG. In: Schomer DL, Lopes da Silva FH, eds. *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins; 2011.

Perrottelli A, Giordano GM, Brando F, Giuliani L, Mucci A. EEG-based measures in at-risk mental state and early stages of schizophrenia: a systematic review. *Front Psychiatry*. 2021;12:653642.

Roach BJ, Carrión RE, Hamilton HK, et al. Reliability of mismatch negativity event-related potentials in a multisite, traveling subjects study. *Clin Neurophysiol*. 2020;131(12): 2899–2909.

Salisbury DF, Kuroki N, Kasai K, Shenton ME, McCarley RW. Progressive and interrelated functional and structural evidence of post-onset brain reduction in schizophrenia. *Arch Gen Psychiatry*. 2007;64:521–529.

Schmidt LA, Segalowitz SJ. *Developmental Psychophysiology: Theory, Systems, and Methods*. Cambridge University Press; 2007.

Skosnik PD, Cortes-Briones JA. Electroencephalography (EEG) and cannabis: from event-related potentials (ERPs) to oscillations. In: Preedy VR, ed. *The Neuropathology of Drug Addictions and Substance Misuse*. Elsevier Science; 2016.

Thatcher RW. Validity and reliability of quantitative electroencephalography. *J Neurother*. 2010;14:122–152.

Zaboski BA, Stern EF, Skosnik PD, Pittenger C. Electroencephalographic correlates and predictors of treatment outcome in OCD: a brief narrative review. *Front Psychiatry*. 2021;12:703398.

1.27 Radiotracer Imaging with Positron Emission Tomography and Single Photon Emission Computed Tomography: Fundamental Principles, Methodology and Role in Neuropsychiatric Research

MARK SLIFSTEIN, PH.D., JOHN H. KRYSTAL, M.D., AND ANISSA ABI-DARGHAM, M.D.

OVERVIEW

Radiotracer imaging is a methodology with the unique ability to quantify spatially localized neurochemical information *in vivo* in the brains of humans and preclinical species. Radiotracers are chemical probes designed to interact with neurochemical systems, either by binding to target molecules such as neuroreceptors, reuptake transporters, or intracellular enzymes or by acting as substrates for the metabolic pathways of endogenous substances in parallel with them. “Tracer” refers to an experimental approach in which a chemical substance containing a quantifiable reporter probe is introduced into a biologic system, and its interaction with the system is “traced” by the probe. It is implicit that the tracer will be introduced in a small enough concentration so that the normal function of the system is not noticeably perturbed by the presence of the tracer itself. Tracer methods were in use for decades before the advent of modern imaging methods and technology. George De Hevesy pioneered the use of radioactive isotopes incorporated in tracers in the mid-20th century to study a wide range of chemical phenomena, including biologic systems. Other investigators developed pharmacokinetic models based on tracer methods to quantify processes such as the perfusion of blood in biologic tissues, which form the foundation of the pharmacokinetic models used in modern radiotracer imaging, but the acquisition of experimental data frequently involved invasive measurements in animals. Approaches that were noninvasive enough to use routinely in living humans did not become available until the late 1970s and 1980s, when technologic advances in imaging systems and computational power made such methods feasible. In radiotracers, the reporter probe is a radioactive atomic isotope incorporated in a molecule that is designed to interact with a (neuro)biologic system. Radioactive isotopes have nuclear configurations that are unstable; individual radioactive atoms decay to a more stable state at unpredictably random times, but when a large number of radioactive atoms are present, the rate of decay at any time is a very well–characterized fraction of the initial rate. There are many different schemes by which radioactive isotopes decay, but all involve the release of energy. When they decay, radioactive isotopes used in imaging release high-energy photons (also called γ-rays) that readily pass through biologic tissue and are then detected in imaging systems external to the body. Photon energy is usually described in units of electron volts (eV, 1 eV = $1.6*10^{-19}$ Joules, the energy change that an electron would experience when passed through a 1-volt potential difference. 1 keV = 10^3 eV, 1 MeV = 10^6 eV), and photons released by isotopes used in imaging range from ∼150 keV for some isotopes to 511 keV for others; most materials other than very dense ones, such as lead or the inorganic scintillating crystals used as detectors in imaging systems, are transparent to photons in this energy range. Radioactive atoms are incorporated into a very small portion of the tracer molecules; these are termed labeled molecules and are assumed to be well mixed within the tracer sample so that at any given time, the rate of decay measured in a location within the tissue will have a known proportionality to the number of unlabeled tracer molecules there. By counting the number of photons that originate in a local volume within the brain over an interval of time, the concentration of the radiotracer in that volume can be inferred. The ability to spatially localize the source of radioactivity requires the use of computational methods known as tomographic reconstruction. The mathematical theory behind tomography was already well established in the early part of the 20th century but was only practically implemented in the final decades of the century with the advent of fast, powerful computers in combination with the development of electronics and detector materials that could provide the necessary spatial, temporal, and energy resolution. There are two imaging modalities that utilize similar experimental methods but differ in the way photons are counted and the mode of decay of the radioactive isotopes used. These are positron emission tomography (PET) and single photon emission tomography (SPECT). The term emission refers to the fact that the radioactive isotope is part of a molecule that is injected intravenously (or, less frequently, inhaled or ingested) so that the detected photons are emitted from within the body. This is in contrast to transmission imaging such as x-ray and computed tomography (CT) scans, in which a beam of photons is directed at the body from an external source, and the portion that passes through without interacting with the intervening tissues is counted on the opposite side. Psychiatry researchers have used PET and SPECT imaging to characterize many aspects of neurochemistry in sample cohorts of subjects with many different neuropsychiatric conditions. While the number of radioactive isotopes used in imaging is small, the number of radiotracers that these can be incorporated into, and biochemical systems that can be studied, are virtually limitless, and many different tracers have been used in research studies with human subjects. Currently, in psychiatry, these imaging methods are used almost exclusively in research. They do not yet have the sensitivity or specificity to be used diagnostically. Most radiotracer studies in psychiatric populations have based their conclusions on statistical inference from moderately large sample cohorts in which there is considerable overlap in the measured quantities between patients and comparison subjects. But these studies provide important clues to other investigators working in basic neuroscience or drug development about processes that may be pathologic in psychiatric conditions and can be examined more closely using *in vitro* methods or in preclinical species, or that might suggest molecular targets for pharmacotherapy. In this sense, radiotracer imaging is an important translational tool for understanding brain neurochemistry in psychiatric patient populations. In the remainder of this chapter, the physics and technology of imaging, pharmacokinetic models, and experimental designs used in psychiatric imaging research will be described, followed by a survey of the radiotracers, molecular targets, and psychiatric populations that have been studied and some well-replicated results.

BASIC PHYSICAL PRINCIPLES

Radioactivity, Radionuclides, and Photon Emission

Half-Life. As stated above, the time at which any individual radioactive atom will undergo decay is random and unpredictable. However, radiotracer doses contain a very large number of labeled molecules. For example, a typical PET dose contains on the order of 10^{10}-labeled molecules, each of which has the same probability of decay per unit time, independently of all the others. In this setting, the rate, or number of decays per unit time, follows an exponential formula, $A(t) = A(0)e^{-\lambda t,}$ where A(t), the activity, is the number of decays per unit time at time t, A(0) is the initial activity, and λ is a very well–characterized rate constant that is different for each radioactive isotope. The

Système International (SI) unit of radioactivity is the Becquerel (Bq), equal to 1 decay per second. An older, though still frequently used unit is the Curie, equal to $3.7*10^{10}$ decays per second. For the range of activity used in imaging, it is more convenient to use units of MBq (10^6 Bq) and mCi (10^{-3} Ci). It is also convenient to factor λ into $\ln(2)/T_{1/2}$, where $T_{1/2}$, the half-life, is the time it takes for A to decrease to half its initial value. The choice of time and the total activity when the initial measurement A(0) is made does not affect the half-life; following any measurement, the activity will decrease by 50% after an additional $T_{1/2}$ has passed. Half-lives and other parameters for the isotopes most commonly used in PET and SPECT are given in Table 1.27–1. The isotopes used in radiotracer imaging are also referred to as radionuclides.

Decay by Positron Emission and PET Radionuclides. In decay by positron emission, a proton within the atomic nucleus is converted into a neutron and a positron (and an extremely low mass, uncharged particle called a neutrino). A positron is a particle with the same rest mass as an electron but opposite electric charge. When the positron is ejected from the nucleus, it will have initial kinetic energy that will be reduced as the particle interacts with the surrounding matter. The sequence of collisions with nearby matter will cause the positron to follow an unpredictable path with many changes of direction, and it will usually come to rest within nanoseconds of its release and within 1 to 2 millimeters of its source atom for the most commonly used PET radionuclides (Table 1.27–1). At that point, the positron and a local electron will be attracted to each other by their opposite charges and undergo an annihilation reaction, in which their masses are converted to energy. Two photons will be produced, each having the energy of the rest mass, 511 keV, of an electron, as the energy from the rest masses of the positron and electron is conserved. The photons travel away from each other in directions that are very nearly opposite to each other (the total momentum is conserved: the momenta of the positron and the electron are nearly zero at the time of the annihilation, and the momenta of the photons are opposite and sum to zero, although the angle is usually slightly different than 180 degrees due to a small amount of residual momentum in the particles at the time of the annihilation event) (Fig. 1.27–1). PET technology exploits the fact that two nearly collinear photons are emitted simultaneously from the location of the annihilation event by utilizing the principle of coincidence detection (see below). The two radionuclides most commonly used in PET brain imaging are ^{11}Carbon (^{11}C, $T_{1/2} \cong 20$ minutes) and ^{18}Fluorine (^{18}F, $T_{1/2} \cong 110$ minutes). ^{15}Oxygen (^{15}O, $T_{1/2} \cong 2$ minutes), incorporated in ^{15}O-labeled water, can be used to assess cerebral blood flow, but its use for this purpose has been mostly superseded by MRI techniques. Also, ^{13}Nitrogen (^{13}N, $T_{1/2} \cong 10$ minutes), incorporated in ^{13}N-labeled ammonia, is used frequently in other areas such as cardiology. Other positron emitters such as ^{64}Copper (^{64}Cu, $T_{1/2} \cong 12.7$ hours) have been used as radionuclides in settings where longer half-lives are desirable, but to date, ^{64}Cu has not been applied in neuroimaging. The radionuclides used in PET are not naturally occurring, and their short half-lives place strict constraints on how long before the study and, therefore, how far away from the study site they can be produced and incorporated into radiotracers. Production of PET radionuclides is performed in a particle accelerator called a cyclotron, an instrument that is produced by several medical device manufacturers but is large, expensive, requires a high level of skill to operate and maintain, and is subject to regulatory scrutiny because it produces radioactivity. Rapid incorporation of the radionuclide in a radiotracer requires elaborate chemistry and quality control instruments and highly skilled chemists to perform these tasks. An approximate rule of thumb is that a radiotracer must be able to be synthesized and undergo tests for sterility and chemical purity in an hour or less to be suitable as an ^{11}C-labeled radiotracer. The use of ^{11}C is therefore restricted to sites that can produce it locally at the time of the PET scan, and PET studies using ^{11}C radiotracers are performed almost exclusively at large academic medical centers that can support the required staffing and equipment. While the requirements for producing ^{11}C-labeled radiotracers are clearly demanding, there are some major advantages to ^{11}C that makes its use very desirable. First, the ubiquity of carbon in virtually all molecules of interest as tracers means that labeling can potentially be performed by substituting ^{11}C for ^{12}C in an organic molecule without altering its biochemical properties. Second, the 20-minute half-life allows for multiple scans on the same day so that scanning can be performed under different experimental conditions on a subject in a single study session. Academic medical centers also routinely produce ^{18}F-labeled radiotracers. There is clearly less time pressure in the production of ^{18}F-labeled tracers than with ^{11}C. In addition, the 110-minute half-life of ^{18}F allows enough of a time window for ^{18}F tracers to be produced by specialized manufacturers with facilities located within several hours of traveling time from the study site. Furthermore, ^{18}F labeling is preferable for cases where the pharmacokinetics of the radiotracer require scans of several hours duration, as in this case, ^{18}F will continue to provide a reliable signal for the entire scan duration, whereas ^{11}C activity would become very low, leading to unreliable data late in the scan. However, multiple ^{18}F scans on the same day in a single subject are challenging to analyze accurately due to activity from the first scan overlapping with activity from the second scan.

Table 1.27–1. Characteristics of Commonly Used PET and SPECT Radionuclides

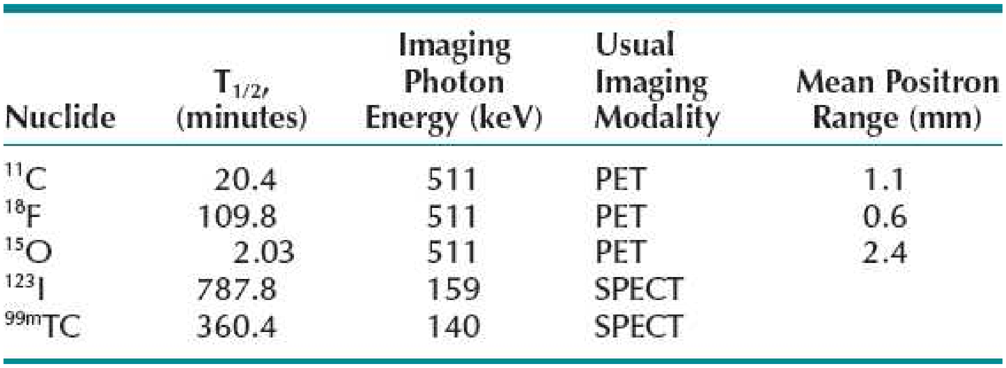

Nuclide	$T_{1/2}$, (minutes)	Imaging Photon Energy (keV)	Usual Imaging Modality	Mean Positron Range (mm)
^{11}C	20.4	511	PET	1.1
^{18}F	109.8	511	PET	0.6
^{15}O	2.03	511	PET	2.4
^{123}I	787.8	159	SPECT	
^{99m}TC	360.4	140	SPECT	

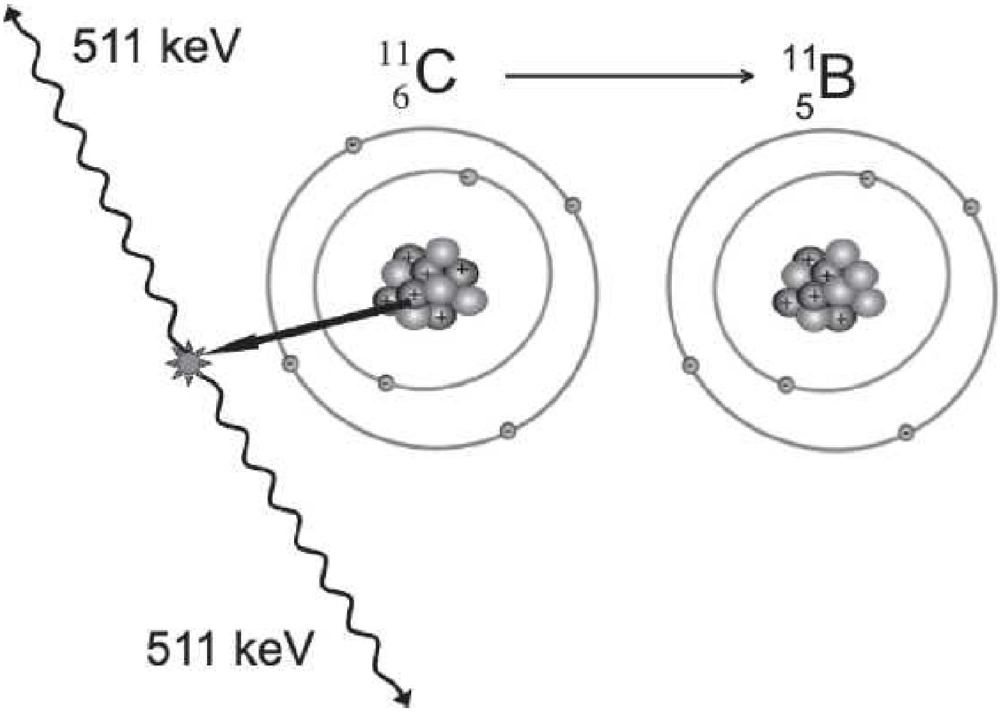

FIGURE 1.27–1. Schematic representation of the decay of unstable carbon 11 to stable boron. A positron is ejected from the carbon nucleus, eventually colliding with an electron. The pair annihilate, and two 511 keV photons are released in opposite directions. Distances are not drawn to scale.

SPECT Radionuclides. As the name implies, SPECT scanners measure single emitted photons, unlike PET scanners that utilize coincidence detection. In theory, PET radionuclides could be used in SPECT systems, but other radionuclides that also emit photons in a usable energy range and decay with longer half-lives are more practical. A SPECT radionuclide that has proven to be particularly useful for brain imaging is ^{123}Iodine (^{123}I, $T_{1/2} \cong$

13.1 hours). ^{123}I decays by the process of electron capture, in which an orbital electron combines with a proton to form a neutron (plus a neutrino), and a characteristic x-ray photon is emitted when another, outer orbital electron, fills the orbital shell vacated by the transformed electron. There are multiple electron capture decay schemes that ^{123}I can undergo, leading to the release of photons of different energies, but approximately 83% of decays result in the emission of 159 keV photons, and these are the ones detected by SPECT scanners. ^{123}I can be produced in a cyclotron, but the long half-life allows for ^{123}I- or ^{123}I-labeled radiotracers to be distributed within a several thousand-mile radius of the manufacturing site. Across all SPECT applications, the most commonly used radionuclide is ^{99m}Technicium (^{99m}TC, $T_{1/2} \cong 6$ hours). ^{99m}TC is a "metastable" product of the decay of ^{99}Molybdenum (^{99}Mo, $T_{1/2} \cong 66$ hours), meaning that its nucleus is in a moderately excited state that will decay to the more stable ^{99}TC form without transmuting to a different element. ^{99m}TC can decay to stable ^{99}TC by several processes, but approximately 88% of decays result in 140 keV photon release, and these photons are detected by SPECT scanners. ^{99m}TC is most commonly distributed to imaging sites in ^{99}Mo generators, portable devices containing ^{99}Mo produced in large nuclear reactors. The ^{99}Mo decays to ^{99m}Tc in the generator, and the ^{99m}Tc can be separated from ^{99}Mo by chromatographic methods. ^{99m}TC, however, does not readily cross the blood–brain barrier (BBB), whereas ^{123}I does so more easily due to the more lipophilic nature of iodine. As a result, the majority of SPECT radiotracers for brain imaging are ^{123}I labeled.

Physical Principles of Emission Tomography

γ-Ray Detection. The same tissue-penetrating property that makes the 511 keV photons emitted in PET and the 140 to 159 keV photons emitted in SPECT useful for imaging leads to the necessity for special materials to detect and transduce their energy into a form more amenable to processing by the electronic systems that count them. Detectors in PET and SPECT are dense inorganic scintillating crystals that interact with, that is, stop the high-energy photons resulting from radioactive decay and emit a portion of the absorbed energy in the form of lower energy photons in the visible to UV range. The most commonly used scintillator in SPECT is thallium-doped sodium iodide (NaI(TI)). Many scintillating detectors have been used in PET; some of the most frequently used ones are bismuth germinate (BGO), lutetium oxyorthosilicate (LSO), and lutetium yttrium orthosilicate (LYSO). Each has different properties that are more optimal for some applications than others.

After the scintillator releases its lower energy photons, the next stage in signal transduction for both PET and SPECT is to convert that signal into an electrical current that represents the detected event in the electronic circuits that perform the remainder of the activity-counting process. This component is called the photosensor. Up to and through the first decade of the 2000s, the photosensor in most commercially produced scanners was the photomultiplier tube (PMT). PMTs are evacuated glass tubes with a light-sensitive material on the end that interfaces with the scintillator (the cathode) and a sequence of metal plates (dynodes) along the length of the tube that are maintained at successively higher voltages. Light striking the cathode will induce electrons to leave the surface and strike the nearest dynode, inducing more electrons to break free from that plate and strike the next one. The cascade continues, with enough amplification of the electron stream by the last stage (the anode) to form a current in the range that is suitable for the next stage of the electronics (Fig. 1.27–2). Common configurations for detectors and PMTs in PET are rectangular blocks of detector material attached to a small number of PMTs. The blocks are assembled into polygonal rings that surround the imaging field of view (Fig. 1.27–3), and a number of rings are assembled concentrically to form a cylindrical imaging gantry. Similar designs are feasible in SPECT, but more commonly, SPECT scanners contain several large slabs of detector materials attached to large arrays of PMTs that rotate around the subject to obtain data from all angles (Fig. 1.27–3). Recently, there has been a great deal of progress in developing solid-state photosensors to replace PMTs, driven largely by the development of hybrid PET-MR scanners (see below). PMTs interact with and do not operate properly in the strong magnetic fields generated by MRI scanners, whereas solid-state photosensors can operate in that environment. Two designs that have been studied extensively are called avalanche photodiodes (APMs) and APMs operating in Geiger mode (referred to as silicon photomultipliers, SiPMs). APMs and SiPMs have become the standard photosensors in many more recently produced scanners, not just PET-MRs.

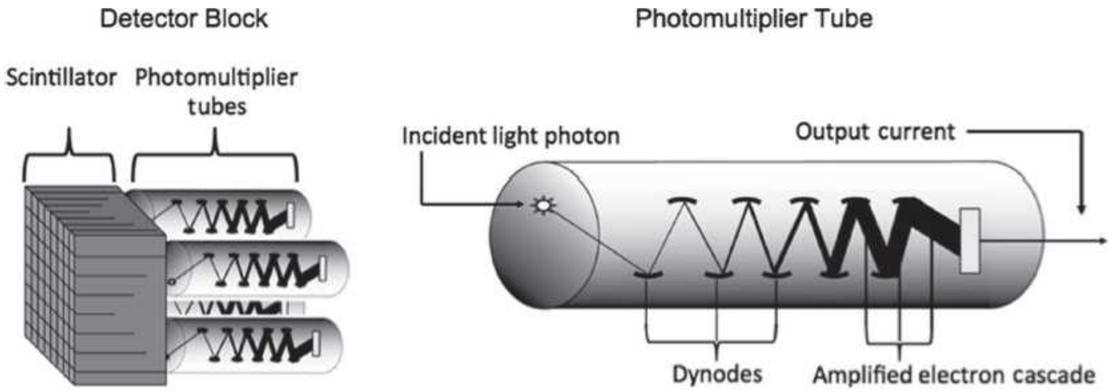

FIGURE 1.27–2. Representation of a detector block with scintillator crystal attached to an array of photomultiplier tubes (**left**) and a photomultiplier tube (**right**) with a schematic display of the amplification of the electron stream.

Coincidence Detection in PET. In PET, radioactive decay events are recorded when two 511 keV photons strike two detectors simultaneously. The path between the two detectors is referred to as a line of response (LOR). An electronic circuit detects if two-photon arrival events occur within a very narrow time interval called a coincidence timing window. The coincidence timing windows on modern scanners are in the range of 3 to 6 nanoseconds. Each detected coincidence event causes a computer to increment the count total for that LOR by one unit (Fig. 1.27–3).

Photon Detection in SPECT by Collimation. In SPECT, decay events are localized to an LOR by the use of lead collimators. These are usually formed from blocks of lead that are attached to the detectors and have long, narrow channels that only allow photons from a narrow set of angles with respect to the detector to pass through. The simplest design utilizes parallel cylindrical channels, all of which are perpendicular to the flat face of the detector, but many different configurations are in use with different scanner designs and different image reconstruction algorithms. The necessity for collimation in SPECT causes its scanners to utilize a smaller portion of the photons that are emitted from radioactive decay than PET scanners. The overall efficiency with which scanners convert radioactivity in the field of view to detected counts is referred to as sensitivity. Because of collimation, SPECT systems have lower sensitivity than PET systems.

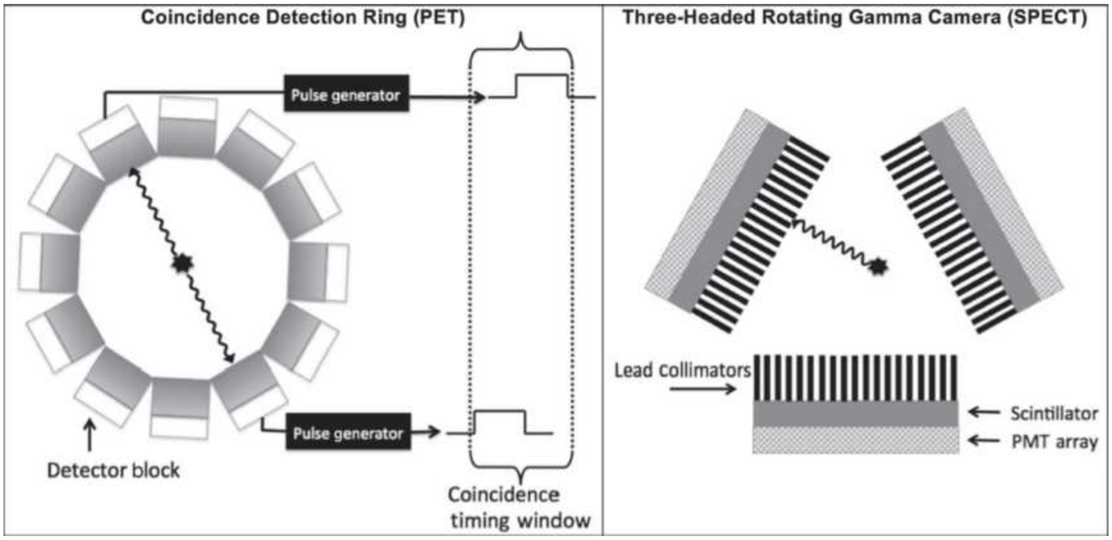
Coincidence Detection Ring (PET)
Pulse generator
Pulse generator
Detector block
Coincidence timing window
Three-Headed Rotating Gamma Camera (SPECT)
Lead collimators
Scintillator
PMT array

FIGURE 1.27–3. A PET ring detector (**left**) with a representation of a coincidence timing circuit and a 3-headed γ-camera for SPECT (**right**). In PET, an electronic pulse generator sends a signal to a coincidence detection circuit. If two pulses are detected within a coincidence timing window from two detectors, an event is counted on the LOR between them. In SPECT, photon detections are restricted to LORs defined by the collimators. The detector arrays rotate around the imaged object.

Image Reconstruction

The data obtained with both PET and SPECT are called projection data. For each LOR, a single number is recorded that contains counts from all loci along the LOR. Conversely, as the direction in which photons are released following a decay event is random, each location in the imaging field of view that contains activity will contribute counts to several, possibly many, LORs. The LOR data contain enough information to reconstruct a facsimile of the original distribution of decay events, but this requires the application of mathematical models of the projection data and numerical implementations of the models on computers. There are two general approaches to reconstruction, filtered backprojection (FBP), and iterative reconstruction. Both involve some degree of mathematical complexity, and there are many variants of each. Below, only the general concepts of both methods are described, not specific implementations. The projection data are stored as a numerical array called a sinogram, which is organized by angle and radial distance from the origin (the center of the field of view). For each angle θ and distance d, the value recorded in the sinogram at the point (θ, d) is the number of detected events (the counts) in the LOR that is perpendicular to the line through the origin at angle θ and passes through that line at distance d from the origin (Fig. 1.27–4). The name sinogram derives from the fact that an object with radioactivity only at a single spatial point would be recorded as a sine wave in the (θ, d) representation. The reconstructed image is usually represented with Cartesian coordinates in a rectangular grid of pixels (in two dimensions) or voxels (for “volume element” in three dimensions).

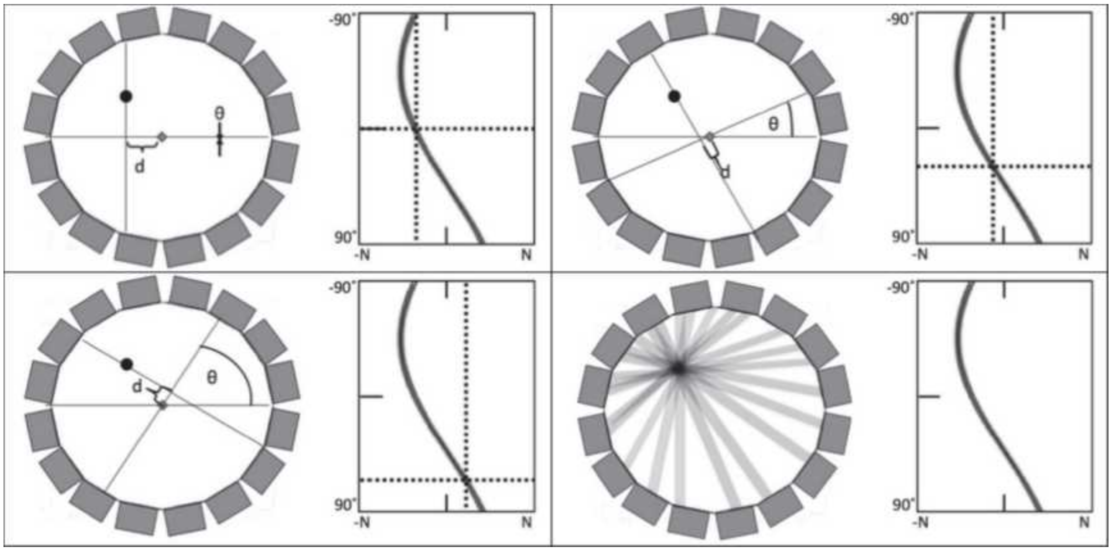

FIGURE 1.27–4. Depiction of back-projection. A point source of activity is placed in the upper left quadrant of the field of view. The **top row** and **bottom left panel** depict three different LORs and their locations in the sinogram of the point source. The **bottom right panel** shows the result of back projecting all LORs; the only ones that are non-zero are those that pass through the location of the point source.

Filtered Backprojection. FBP, sometimes called analytic reconstruction, was the most common approach to image reconstruction from the inception of CT up until the early 2000s, partly because the iterative methods, which appeared in several published papers in the 1980s, were too computationally intensive, and therefore too slow, for the computers of that time period. Iterative reconstruction programs, which lead to visually clearer reconstructed images, are now routinely provided as part of the software package with commercially produced scanners, but many researchers still prefer FBP reconstruction due to statistical issues described below. The basic equations of FBP are derived from mathematical formulas called the Radon transform and the inverse Radon transform, theoretical constructs first described in the early part of the 20th century explaining the relationship between a mathematical function defined in two-dimensional space, that is, f(x,y) which defines a single numerical value for each (x,y) pair of Cartesian coordinates, and its projections. FBP is a discretized version of the inverse Radon transform; that is, a version based on a finite number of LORs rather than a full continuum of angles and distances from the origin. The first step can be thought of as overlaying the projections from all the LORs and summing (Fig. 1.27–4). This step is called back projection. Because the loci that contain radioactivity contribute to all the LORs that pass through them, the back-projected data will have elevated intensities in these positions. However, pure back projection would also attribute activity along each entire LOR containing counts, so an adjustment called spatial filtering is applied prior to back projection to remove the starburst-like artifact that would otherwise result (Fig. 1.27–5). This is a conceptual description of FBP; real implementations use more sophisticated computer algorithms, and mathematical procedures such as Fourier transforms. The main advantage of FBP is that it is fast. However, there are disadvantages. The filtering step and the fact that there are a finite number of LORs combine to reduce spatial resolution and introduce the so-called streak artifacts (Fig. 1.27–6).

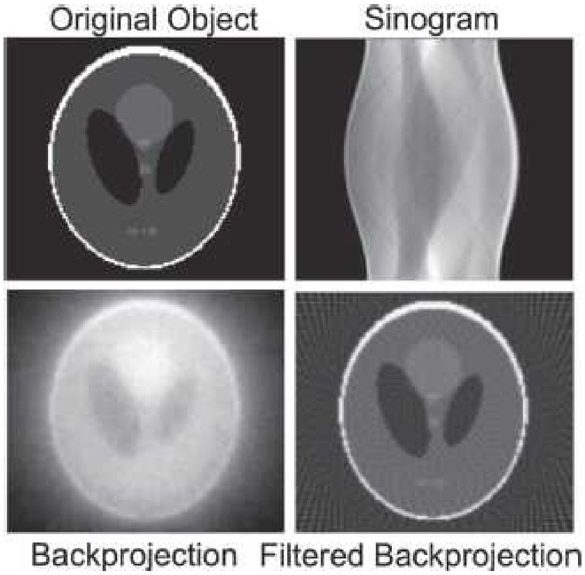

FIGURE 1.27–5. Filtered back-projection. The object is a Shepp-Logan phantom, a set of superimposed ellipses of different intensities used as a standard for comparing reconstruction algorithms. There are several different filters that can be used. The one used here is called a ramp filter.

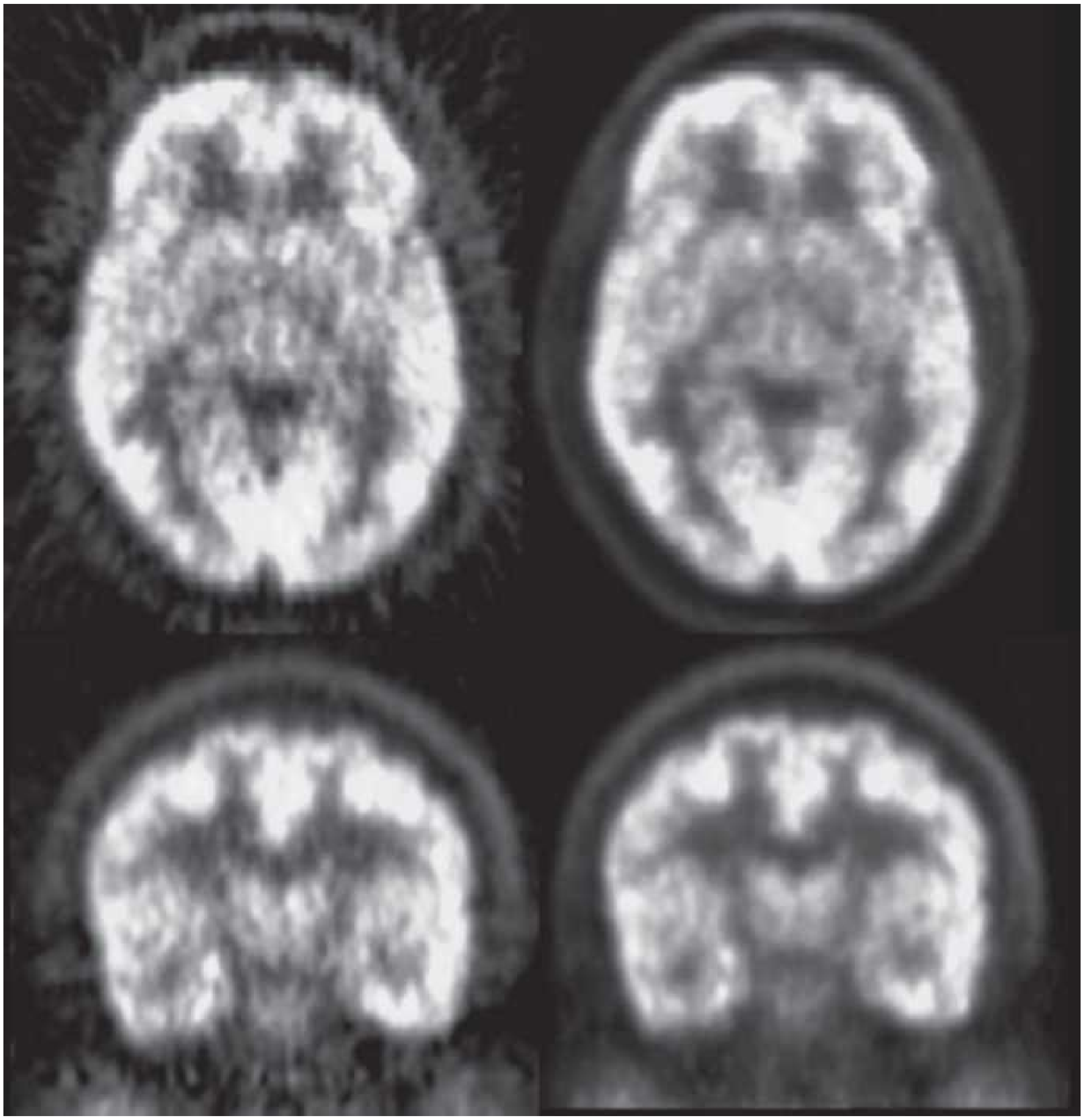

FIGURE 1.27–6. Comparison between FBP (**left**) and OSEM (**right**). The image is summed data from a 120-minute [^{11}C]MDL100907 scan, shown in the transverse (**top**) and coronal (**bottom**) views. The radiotracer binds to the serotonin 5-HT$_{2A}$ receptor. The summed image represents a combination of distribution of the tracer throughout the brain during the early part of the scan and specific binding to receptors in the later part. Streak artifacts are visible in the FBP, especially in the space between the scalp and brain. The FBP reconstruction also has a more mottled appearance than the OSEM reconstruction.

Iterative Reconstruction. A number of algorithms have been proposed that treat the relationship between projection data and voxel values probabilistically by framing reconstruction as a statistical estimation problem. These treat the voxel values as a sample drawn from a probability distribution and seek to estimate the most probable distribution of counts in voxels, given the measured projection data and the known physical properties of the imaging system, such as the detector geometry. They are called iterative because they start with an arbitrary estimate of the voxel values and then successively improve the estimate by updating a formula that takes the current estimate as input and returns an improved estimate as output. Each cycle through the procedure is called an iteration. The iterative algorithm that is available in the software packages provided by most commercial scanner manufacturers is called Ordered subset expectation maximization (OSEM), a modified version of the expectation maximization (EM) algorithm. EM and OSEM are based on the maximum likelihood (ML) principle. ML estimation procedures are statistical methods that establish formulas to estimate the parameters in a parametric probability distribution from observed sample data. For example, the conventional formula for estimating the mean μ of a Gaussian random variable as the sample average can be derived from the ML principle. The Likelihood is the formula for the probability of the observed sample data, expressed as a function of these data and the parameters of the distribution. As the name suggests, ML methods seek the value of the parameter(s) that maximizes the likelihood of the observed data. Voxel data are usually modeled as Poisson random variables; the Poisson distribution can take on only nonnegative integer values and has a single parameter λ which is the mean of the distribution. In the EM and OSEM approaches, counts emanating from each voxel are treated as statistically independent of every other voxel. The mean or expected value of the counts at voxel j is assumed to be λ_j, the parameter of the Poisson distribution for that voxel. These, however, are hidden variables because the only observed data are the projection data on LORs. The EM algorithm is designed to obtain the ML parameter estimate for the distribution of a hidden variable when only incomplete data, such as the projection data, are available (the projection data along with the unobserved voxel means are called the complete data). For the particular case of projection data and the assumed Poisson distribution of voxel values, the procedure leads to a simple and elegant formula that is guaranteed to converge, over the course of successive iterations, to the set $\{\lambda_j\}$ of ML estimates of voxel means. However, the full Poisson EM algorithm requires a huge number of computations, given the large number of LORs and voxels in PET and SPECT image data, and was therefore not practical to apply on the computers available at the time the EM algorithm was proposed. The OSEM algorithm is a modified version introduced in the 1990s that made iterative reconstruction feasible. By cleverly ordering the full set of LOR data into a sequence of smaller subsets and updating the voxel estimates based on computations performed on each successive subset rather than only once after each complete iteration through all the LORs, fewer equations need to be solved, and the necessary number of iterations, and therefore the computational load, is reduced significantly. Unlike full EM, OSEM is not guaranteed to converge to the ML estimate of the voxel values, but practical experience has shown that it leads to very satisfactory image reconstructions after a fixed, small number of iterations (Fig. 1.27–6). The main advantage of OSEM reconstruction is that it leads to clearer images without streak artifacts or the level of random noise that is seen in FBP reconstruction (Fig. 1.27–6). A disadvantage is that following necessary preliminary corrections to the sinogram data prior to reconstruction for several noise sources that occur, especially in PET, the Poisson distribution assumptions are not completely satisfied, and this can lead to bias in which the true proportionality between brain regions with high and low activity is not preserved, although the bias is relatively small. For this reason, many neuroimaging researchers, who are more interested in quantification than visual presentation, have historically preferred FBP to OSEM. But continued refinements to EM-based algorithms along with increasingly powerful computer hardware have led to faster and less biased iterative methods, making recent versions more competitive for quantitative imaging.

Another recent development is the practical application of time of flight (TOF) information in iterative reconstruction algorithms. If the location of the annihilation event is not identically midway between opposing detectors, the photons will arrive at the detectors at slightly different times. Theoretically, if this timing difference were known exactly, the location of the coincidence event along the LOR could be identified as well. In reality, most scanners can only resolve the timing differences down to a precision of a nanosecond (10^{-9} s) or more, leading to too much uncertainty in the estimate of the location of the event for this information to be useful in brain imaging. Timing resolution has been improved considerably, though, in more recently manufactured scanners, with tested values as low as 210 picoseconds ($2.1*10^{-10}$ s) in some systems. This allows for localization down to several centimeters. While still not precise enough to adequately identify the location of the annihilation event by itself, this information can be incorporated into the iterative reconstruction algorithm, significantly improving the signal-to-noise ratio of the reconstructed image.

Factors Affecting Image Quantification and Image Quality

The sections above describing the release and subsequent detection of photons are idealized in the sense that it is assumed that 100% of photons pass through biologic tissue without any interaction and that 100% of photons or coincident pairs striking the detectors are perfectly converted into detected counts. In reality, there is some interaction between photons and tissue, LORs are not really lines but three-dimensional columns in which the probability of detection is greater at the center than toward the edges, and electronics have finite limits to their temporal and energy resolution. These factors and others degrade the overall image quality and spatial resolution, some severely enough to require corrective procedures to be applied prior to the reconstruction phase.

Photon Attenuation. While many of the photons from radioactive decay do pass through tissue unimpeded, a fraction of them do not. For γ-rays in the energy range of those released by PET and SPECT isotopes and matter like biologic tissue, the majority of interactions are due to Compton scattering, in which the photon collides with an atom and is deflected from its original path, with some of its energy absorbed by an orbital electron from the atom. Less frequently, the photon is completely absorbed in the collision, and the atom ejects an electron, a process called the photoelectric effect. The reduction of detected photons due to these interactions is called attenuation. The probability of attenuation decreases with increasing γ-ray energy and increases with the density and thickness of the intervening matter between the photon source and the detector. For a given kind of matter, such as soft tissue, the probability that a photon passes through without being attenuated is an exponential function of the depth of the tissue where the photon is released; the probability is proportional to $e^{-\mu d}$ where d is the depth in centimeters and μ is the linear attenuation coefficient of the material in units of cm^{-1}. The attenuation in air is nearly zero, so only the thickness of the tissue is a factor in attenuation, not the total distance to the detectors. The linear coefficient μ is a function of both the energy of the photon and the nature of the material through which the photon is passing; the larger μ is, the greater the attenuation. Because PET photons have higher energy than SPECT photons, they will undergo less attenuation. For example, μ in water for 511 keV photons is $0.097\ \mathrm{cm}^{-1}$, whereas for 150 keV photons, it is approximately $0.15\ \mathrm{cm}^{-1}$, and μ in soft tissue is similar to μ in water. Without correction for attenuation, voxels located deep in tissue would have lower detected counts than voxels closer to the surface having the same concentration of radiotracer. Therefore, attenuation correction is a necessary step in quantitative imaging. The fact that PET utilizes coincidence detection allows the implementation of a very accurate form of attenuation correction. This is because the probability of either photon from annihilation not being attenuated is independent of whether or not the other is attenuated, and therefore the probability of both photons not being attenuated is the product of their probabilities; it is proportional to $e^{-\mu d1}e^{-\mu d2} = e^{-\mu(d1+d2)}$ which depends only on the total length of tissue, d1 + d2, traversed by the LOR, and does not depend on the location within the LOR where the annihilation event occurs (Fig. 1.27–7). Attenuation correction in PET is usually applied to the sinogram prior to reconstruction by multiplying the value of each LOR by a correction factor measured in a separate scan called a transmission scan. In dedicated PET systems, this is usually performed by rotating a point source or a rod containing multiple point sources of a positron-emitting material of known activity around the measured object (the subject's head) and recording the fraction of activity detected on the opposite side. Different approaches are required for hybrid PET-CT and PET-MR scanners (see below). In single-photon detection, as in SPECT, other strategies must be employed. While some sophisticated algorithms for measuring attenuation using a transmission source have been devised for application in SPECT, uniform attenuation correction is most frequently applied. In this approach, an image of an elliptical object with voxel values equal to the correction factors for water is aligned to the reconstructed SPECT image, and the values in the two images are multiplied on a voxel-by-voxel basis.

Photon Scatter. Compton scattered photons are not absorbed in the tissue completely but rather change their trajectory and continue at reduced energy. The direction of change is random, and the reduction in energy increases as the angle of deflection increases. In PET imaging, the scattered photon and its partner will arrive at two detectors within a coincidence timing window, and therefore the annihilation event will be misattributed to an LOR that does not pass through the true location. In SPECT imaging, the photon will be misread as a photon emanating along an LOR from the location of the scattering event in the scattered direction. In both modalities, detector circuitry can discriminate photon energy and reject low-energy photons to some degree, but there is enough overlap between the range of accepted energies and the energies of scattered photons such that many scattered photons are counted as true decay events. Scatter adds random noise unrelated to the true radiotracer distribution and would degrade the image quality considerably if not corrected. There have been many scatter correction methods proposed and implemented. Modern scanners incorporate sophisticated software algorithms that utilize information from the attenuation map generated from the transmission image to make probabilistic models of the scatter distribution, which are subtracted from sinogram data prior to reconstruction.

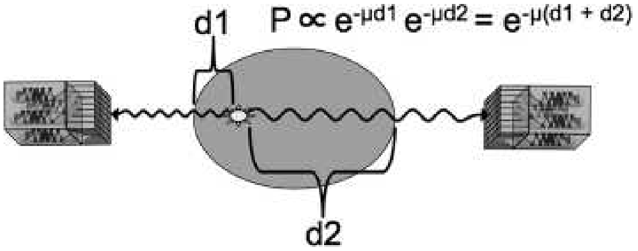

FIGURE 1.27–7. Attenuation. The ellipse in the center represents a hypothetical object (i.e., a head) in the field of view. In SPECT, the photon traveling to the left would have the probability proportional to $e^{-\mu d1}$ of not being attenuated, and the one traveling to the right would have a probability proportional to $e^{-\mu d2}$ of not being attenuated. In PET, the coincidence event along this LOR would have a probability proportional to $e^{-\mu(d1+d2)}$ regardless of where along the LOR it originated.

Random Coincidences in PET. An additional source of noise specific to PET is accidental, or random, coincidences. These arise when two photons from distinct annihilation events arrive at two detectors within a coincidence timing window and are counted as an annihilation event along the LOR connecting the two detectors. Random coincidences are uniformly distributed across the field of view and cause an artifact of elevated background activity level, roughly in proportion to the square of the singles rate (the rate at which individual photons are detected). As with scatter, random coincidences would considerably degrade the image quality if not addressed. Most scanners have a delayed coincidence timing channel, an electronic circuit that allows estimation of random coincidences by purposely counting pairs of detected photons as coincidence events that arrive several nanoseconds further apart than a coincidence timing window and are therefore known not to have originated from the same annihilation event. The estimated random coincidences are subtracted from sinogram data prior to any other processing. In regions of the field of view where there is no, or very low, true activity, this step can lead to apparent negative counts.

Spatial Resolution. All imaging systems have characteristic spatial resolutions and indices of the smallest object dimensions that they can distinguish. Resolution in PET and SPECT systems are limited by the physical characteristics of the scanner, such as the finite width of LORs (as opposed to infinitesimally thin lines) or the fact that projection data are measured at a finite number of angles and distances from the origin at each angle, rather than the full continuum of angle and length values. Resolution limiting factors specific to PET are the range of positrons and the fact that the annihilation events are only localized within a sphere of 1 to 2 mm radius from the sites of the decay, and that the angle between released photons is not identically 180 degrees. Resolution in SPECT is limited by collimator design and the distance of the imaged object from the collimators. The resolution of imaging systems is measured empirically by imaging radioactive sources of standard shapes and characterizing the accuracy of the image. If a small point source or a thin line source of activity is placed in the field of view, the reconstructed image spreads in a bell-shaped curve around the point or line called the point spread function (PSF) or the line spread function (LSF). Resolution is usually reported as the full width at half maximum (FWHM) of the PSF. This is the length, in millimeters, between the two points at which the value of the PSF is half of its maximum central value (Fig. 1.27–8). The PSF and FWHM are usually not constant across the field of view; they vary slightly with position but are typically reported as one average number. They are also dependent on the method of image reconstruction applied; performance evaluation reports are usually conducted using standardized methods to facilitate comparisons across scanners and reports. Most modern production PET scanners have FWHM in the range of 4 to 5 mm. Some specialized PET scanners have FWHM in the range of 2 to 3 mm. Resolution in SPECT is dependent on the collimator system in use. Clinical, rotating head SPECT scanners have FWHM in the range of 10 to 12 mm. Specialized small animal SPECT scanners that can exploit the close proximity of the animal to the collimator and use special collimator designs can have submillimeter FWHM. Images are usually reconstructed with voxel dimensions that are smaller than the FWHM to avoid having a checkerboard-like or pixelated appearance. As a result, each voxel will contain overlapping information from the voxels that are adjacent to it.

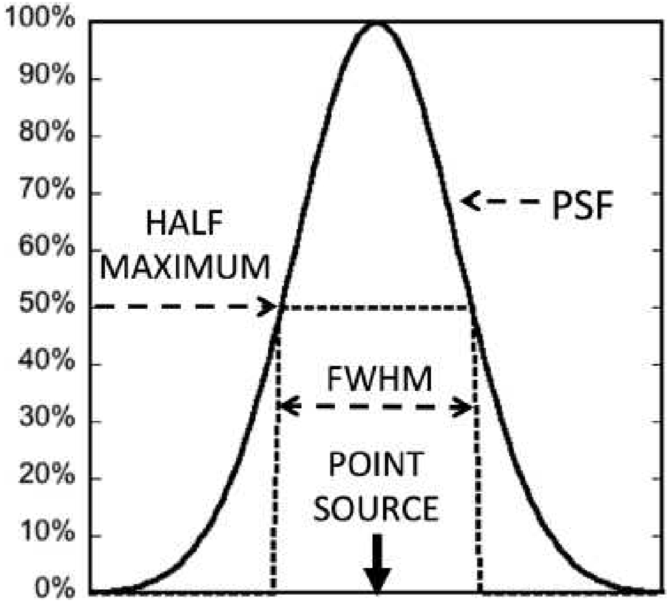

FIGURE 1.27–8. Schematic representation of a PSF from a point source of activity at the center of the horizontal axis.

Hybrid Imaging Systems

In the years since 2000, scanner manufacturers have begun producing multimodal scanners having a PET or SPECT component combined with an additional imaging system such as CT or magnetic resonance (MR). In fact, dedicated PET systems for imaging humans have become increasingly rare, with the majority of recently produced systems containing a CT component. Hybrid PET-MR scanners are also becoming more commonplace, with all major manufacturers now producing systems for clinical use. The adaptations necessary to use these systems and their potential to expand the information obtained in the context of neuroimaging are briefly described.

PET-CT and SPECT-CT. A CT system is mounted along the same axis as the PET or SPECT system. The CT and radionuclide imaging are performed in sequence, not simultaneously, as the locations of the imaging detectors for the two systems are different. Very accurate coregistration of the images is possible because the distance between the patient positions for the two modalities is known precisely and can be used for image fusion in which the neurochemical information from the radiotracer image is superimposed on the more anatomically precise, high-resolution CT image. This has proven to be very valuable for imaging the chest or abdomen, as organs are easily identifiable on CT images but not on PET or SPECT images. It is of less value for brain imaging because CT images do not distinguish between soft tissue types. MRI images, on the other hand, do distinguish between gray matter, white matter, and CSF, and MRI is, therefore, a preferred modality for anatomical coregistration and structure identification. Frequently in neuroimaging studies, MRI images will be obtained in a separate session, and coregistration will be performed by software methods. PET-CT and SPECT-CT systems utilize the CT scan as a transmission scan for attenuation correction. An advantage of this use is the high signal-to-noise ratio obtainable in the CT images. The photons from CT scanners are x-rays spread over a relatively broad energy spectrum from ~40 to 140 keV. These have different (higher) attenuation than the higher energy photons from PET or SPECT and require the use of a lookup table method to transform their μ values into appropriate values for the other modalities.

PET-MR. Unlike PET-CT, the PET and MRI components are concentric, and images are acquired simultaneously. For whole-body imaging, the main advantages are the ability to obtain anatomical images without the additional radiation burden of CT and reduced overall imaging time. For neuroimaging, there is the potential, in addition to the anatomical information from the MR, to simultaneously utilize other MR techniques such as perfusion imaging, magnetic resonance spectrometry, or BOLD fMRI, greatly expanding the experimental information that can be obtained in a single imaging session. An ongoing technical challenge is attenuation correction. The concentric scanner ring design and the current size of the components from both modalities preclude the inclusion of rod or point sources of radioactivity used for conventional transmission scans. On the other hand, the MRI images, while providing very precise anatomical information about brain tissue, provide no direct information about the attenuation of high-energy photons. In addition, skull bone and other materials in the PET field of view, such as the patient bed or the MRI RF coil (which is outside the MR field of view) that do attenuate photons are not detected with MRI (skull can be detected using special fast echo RF pulse sequences and manufacturers have started to include these in their software). Current approaches for brain imaging involve coregistering acquired MRI images to standard atlases and using standardized μ values from lookup tables. Other, simpler methods, such as fat/water tissue segmentation algorithms in which each voxel is assigned to a tissue type with a standardized μ value, have proven adequate for whole-body scanning. While much progress has been made over the past decade, attenuation correction in PET-MR remains an active area of research; improvements and refinements can be expected as the technology matures and interest grows.

Safety

Radiation Exposure and Internal Dosimetry. Radioactive decay leads to the deposition of energy in tissue when positrons, electrons, or photons interact with molecules in tissues. While the type and amounts of radiation that PET and SPECT research subjects are exposed to are very unlikely to lead to acute effects, the long-term impact and the possibility of genetic damage are less well understood. Risk/benefit considerations have led governments and international advisory groups to propose guidelines and regulations for exposure limits to research subjects. These are based on the linear nonthreshold model that assumes observations made at much higher exposures with qualitatively different radiation can be extrapolated to the levels that research subjects are exposed to in PET and SPECT studies. There is not currently a consensus among radiation biologists as to the validity of this assumption, but in the absence of definitive data, a conservative approach has been adapted. In the United States, there are FDA regulations limiting maximum radiation exposure to research subjects. Other countries are guided by the recommendations of the International Commission on Radiological Protection (ICRP). The amount of energy deposited in any organ or part of the body is a complicated function of the tissue type, the form of the organ, the isotope, and the distribution of the radiotracer throughout the body. The estimated radiation exposure associated with a radiotracer is most often determined through a combination of mathematical models, computer simulations, and test scans known collectively as internal dosimetry studies. The distribution of the radiolabel throughout the body will differ between radiotracers according to how the tracer is metabolized and cleared, and this is determined empirically by acquiring whole-body scans, either in a small cohort of human subjects or in animals with the application of scaling factors to account for size differences between species. Organ radiation doses are calculated as weighted sums of the energy due to decays occurring in the organ itself and the energy deposited in it from decay events originating in other parts of the body. The unit of absorbed dose is the Gray (Gy, joules per kg tissue). An older unit is the radiation absorbed dose (rad, 1 rad = 0.01 Gy). Exposure is usually reported as an absorbed dose per injected unit of radioactivity, such as mGy/MBq (1 mGy = 10^{-3} Gy) or rad/mCi. In the ICRP guidelines, absorbed doses are multiplied by an additional quality factor that accounts for the type of radiation, leading to different units (the guidelines also apply to workers with occupational exposures to more damaging types of radioactivity than used in PET and SPECT). After multiplication by the quality factors, Gy are transformed into units called Sieverts (Sv), and rads become Roentgen-Equivalent Man (rem). However, the quality factors are all equal to 1 for the types of radiation from PET and SPECT isotopes so that Sv are numerically equivalent to Gy and rems are numerically equivalent to rads. The weighting coefficients used to calculate the radiation dose to each organ as a sum of contributions of activity originating from all organs are derived from tables generated through detailed computer simulations that account for organ shape, size, composition, and distance from other organs in mathematical models of humans of different sizes, ages, and sexes. The FDA regulations are based on the concept of the critical organ, in which the total allowable radiation per study is limited by the absorbed radiation dose to the single organ with the highest exposure. The ICRP guidelines are based on weighted averages of doses to a predetermined set of organs. There have been multiple revisions of these guidelines and the names and definitions of the weighted sums. The names have been somewhat confusingly similar (dose equivalent, effective dose, equivalent dose, effective dose equivalent, etc.), but the standard as of ICRP Publication 103 (2007) is the effective dose (ED, Sv) equal to a weighted sum of absorbed doses from 28 specified organs and tissues. While the precise dose limits differ from tracer to tracer, several scans per subject per study can be performed with most radiotracers while remaining within the allowable exposure range.

Pharmacologic Toxicity. Compounds that are administered to human research subjects must undergo extensive toxicology screening prior to approval. The low mass dose usually administered to research subjects, in the range of several μg or less, generally keeps the dose many orders of magnitude lower than the level at which physiologic effects are noticed. This is especially true of radiotracers that are neuroreceptor antagonists. At tracer doses, these maximally occupy several percent of the receptors, well below the amount necessary for pharmacologic effects on neurotransmission. On the other hand, some agonist radiotracers have been known to elicit a pharmacologic response even at very low mass doses. Two examples are the μ-opioid agonist [^{11}C]carfentanil, which can cause sedation and hypoventilation at μg doses, and the dopamine D2/D3 receptor agonist [^{11}C]-(+)-PHNO, which can cause transient nausea several minutes after a bolus injection.

PHARMACOKINETIC MODELS AND METHODS, EXPERIMENTAL DESIGNS

General Framework and Compartment Models

Above, some of the physical and image processing theories, technology, and algorithms used to create a PET or SPECT image were described. From the experimenter's perspective, all these components lead to one key characteristic of the imaging data that makes quantitative neuroimaging feasible: in each spatial location within the image, the measured amount of activity is proportional to the concentration of the tracer at that location. These measured amounts can be extracted as the average activity over a well-defined brain region of interest which can be as large as thousands of voxels or as small as a single voxel. Imaging data can then be statistically fitted to mathematical models of the observed biologic processes, and parameter estimates can be derived to quantitatively characterize the processes. The imaging data can range from single, static snap shots of the tracer distribution obtained from 10 to 20 minutes of scanning to sequential dynamic images that record the kinetic profile of the tracer over the course of hours. The mathematical models can range from simple measurements of the ratio of radiotracer concentrations in different tissues to detailed differential equation models that relate the pharmacokinetic profile of the brain concentration of the radiotracer over hours to the concentration in arterial blood plasma, collected in parallel with the brain images over the same time period. The radiotracer enters the brain from the bloodstream, and the most comprehensive models combine an equation for transport across the BBB with equations for the interaction of the radiotracer with biologic processes of interest in brain tissue; simpler models can sometimes be derived from these if appropriate conditions are met. The transport equation is usually based on Fick's principle, in which the net rate of influx or efflux across the BBB is proportional to the gradient in radiotracer concentrations between brain tissue and blood. Models of biologic processes can be broadly categorized according to whether the radiotracer is a ligand that binds reversibly to a target molecule or is a substrate for a metabolic pathway and becomes irreversibly trapped at some stage. A general, flexible framework that encompasses the reversible and irreversible cases is the compartment model (Fig. 1.27–9). A compartment model describes radiotracer kinetics with a simple set of rules, in which states of the radiotracer are represented as compartments, drawn as boxes, and changes of state are represented as flow between compartments, denoted by arrows showing allowable directions of flow between compartments, at rates that are proportional to the tracer concentration in the source compartment. Corresponding to the compartment diagram is a system of differential equations describing the rates. Outcome measures containing relevant biologic or pharmacokinetic information are formed as functions of the rate constants. The rate constants are unknown parameters; they are estimated by using a computer-based data fitting procedure that provides the best fit of the model to the measured PET or SPECT data. The most common criterion for best fit is least-squares minimization, in which the program searches through the entire plausible range of rate constants and finds the ones that minimize the sum of the squared differences between the measured data, recorded as discrete “frames” at a sequence of time points over the course of the scan, and the model predictions at these time points, although many data fitting and parameter extraction procedures are in use in addition to the conventional least-squares approach.

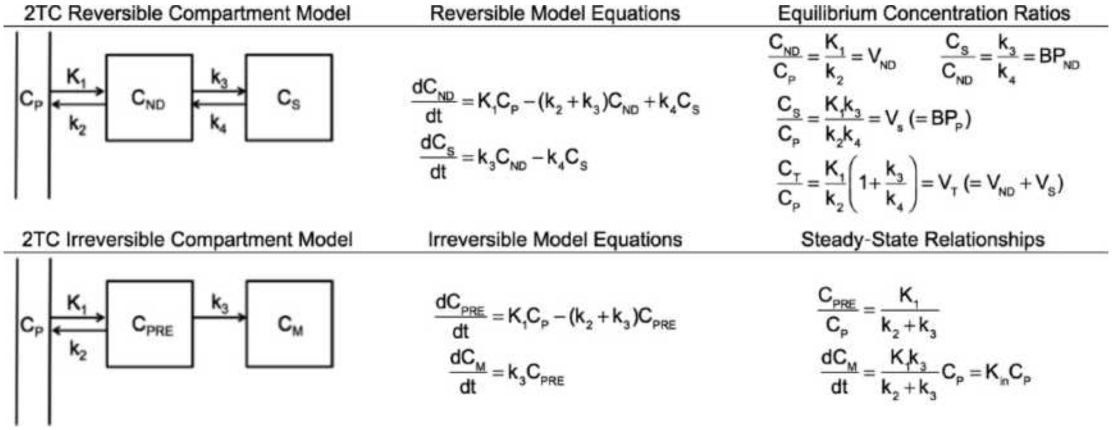
2TC Reversible Compartment Model
Reversible Model Equations
Equilibrium Concentration Ratios
C_P
K_1
k_2
C_{ND}
k_3
k_4
C_S
$\frac{dC_{ND}}{dt} = K_1C_P - (k_2 + k_3)C_{ND} + k_4C_S$
$\frac{dC_S}{dt} = k_3C_{ND} - k_4C_S$
$\frac{C_{ND}}{C_P} = \frac{K_1}{k_2} = V_{ND}$
$\frac{C_S}{C_{ND}} = \frac{k_3}{k_4} = BP_{ND}$
$\frac{C_S}{C_P} = \frac{K_1k_3}{k_2k_4} = V_s\ (= BP_P)$
$\frac{C_T}{C_P} = \frac{K_1}{k_2}\left(1 + \frac{k_3}{k_4}\right) = V_T\ (= V_{ND} + V_S)$
2TC Irreversible Compartment Model
Irreversible Model Equations
Steady-State Relationships
C_P
K_1
k_2
C_{PRE}
k_3
C_M
$\frac{dC_{PRE}}{dt} = K_1C_P - (k_2 + k_3)C_{PRE}$
$\frac{dC_M}{dt} = k_3C_{PRE}$
$\frac{C_{PRE}}{C_P} = \frac{K_1}{k_2 + k_3}$
$\frac{dC_M}{dt} = \frac{K_1k_3}{k_2 + k_3}C_P = K_{in}C_P$

FIGURE 1.27–9. Some of the mathematics of compartment models. The box representations (**left**) for reversible and irreversible 2TC models. The center column shows the associated differential equations. These are derived by the simple rules that the overall rate of change of a compartment is a weighted sum of all source compartment concentrations (ones with *arrows* pointing into it) minus a weighted sum of the compartment concentration itself if the compartment is reversible (*arrows* pointing out of it). The weights are the rate constants labeled K_1, k_2, k_3, and k_4. The **right** column shows outcome measures. For the reversible tracers, all outcome measures are equilibrium ratios of concentrations. Because they are equilibrium ratios, the rates of change of each compartment (the **left** side of the equations in the center column) are 0, and the concentration ratios can be derived as functions of the rate constants. For the irreversible model, the steady-state representation of C_{PRE} can be derived in a similar way, and then substitution into the differential equation for the rate of accumulation of C_M leads to a rate constant expression for K_{in}.

Reversible Models. Reversibly binding tracers are used most frequently to quantify receptor, transporter, enzyme, or other protein density in the brain. These tracers are also referred to as radioligands. The modeled processes for reversible radioligand binding are bimolecular association between the radioligand and target protein and unimolecular dissociation of the target-ligand complex. The differential equation for this process is the mass action law. The canonic compartment model applied to reversible radioligands is the two-tissue compartment model (2TC, Fig. 1.27–9). The 2TC model requires that the concentration of radiotracer in arterial plasma be measured during the scan as an input to the model. This is invasive, as it requires arterial catheterization, and technically demanding, as it requires the application of chromatography to estimate the fraction of radioactivity in the blood that is due to unmetabolized radioligand rather than radiolabeled metabolites that form in the periphery and accumulate in the blood over the course of the scan. For some radiotracers, there are methods for estimating relevant parameters that do not require an arterial plasma input function, but even these are often conceptually derived from the 2TC, highlighting its central role in models of reversible tracers. There are two main simplifying assumptions in the 2TC. When the radioligand crosses the BBB, it is initially in a free, unbound state. From the free state, it can associate specifically with the target protein it was designed to bind to, or it can loosely bind to membranes nonspecifically. Specific binding is characterized by certain properties: it is saturable (it cannot exceed the concentration of the available target protein regardless of how high the ligand concentration is), and it is displaceable by other ligands that bind to the same site. Conversely, nonspecific binding is not saturable and not displaceable. Also, compared to the rate of the specific binding process, free and nonspecifically bound radiotracer equilibrate rapidly. The first simplifying assumption is that this equilibration occurs instantaneously so that free and nonspecifically bound radiotracer are combined into a single compartment called the nondisplaceable compartment (C_{ND}). Only the free fraction of C_{ND}, denoted f_{ND}, is able to bind specifically to the target molecule. The bound state is the Specific binding Compartment (C_S). The second simplification is called the tracer dose approximation. The bimolecular association term of the mass action law contains the product of the unbound target protein concentration and the free radiotracer concentration. The tracer dose approximation is that the concentration of radioligand is so low that, at most several percent of the target is bound, and therefore the target concentration can be treated as if it is constant. To satisfy the tracer dose approximation condition, very low mass, usually several micrograms or less, of radioligand are injected. This also limits the possibility that the radiotracer will pharmacologically perturb the system under study. In theory, models could be formed that do not include these simplifications—free and nonspecifically bound tracer could be treated as separate compartments, and the equations of the full mass action law without the tracer dose assumption could be implemented, but experience has proven the 2TC model to be the most comprehensive model that is also robust and stable enough for reliable parameter estimation. The two compartments, C_{ND} and $C_{S,}$ represent states that are occurring simultaneously in the same voxels and cannot be separated in the image. Only their sum, $C_T = C_{ND} + C_S$ (T for Total concentration), can be observed; the relationship between C_{ND} and C_S must be inferred through the model. The concentration of the radiotracer in arterial plasma is denoted C_P. There are two types of related outcome measures derivable from the 2TC model. These are distribution volumes and binding potentials. These represent equilibrium ratios of radiotracer concentrations under tracer dose conditions (Fig. 1.27–9), but they are rarely estimated under equilibrium conditions. Rather, they are derived from dynamic scan data, usually acquired following a bolus injection of the radiotracer, as functions of the estimated rate constants. Some of the mathematics leading to the derived relationships between distribution volumes and rate constants are provided in greater detail in the legend to Figure 1.27–9. Distribution volumes represent equilibrium ratios between tracer concentration in compartments and tracer concentration in arterial plasma. For the 2TC model, there are three relevant distribution volumes: $V_{ND} = C_{ND}/C_P$, $V_S = C_S/C_P$ and $V_T = V_{ND} + V_S$. Binding potentials are parameters that are proportional to the ratio of specifically bound to free radiotracer at equilibrium. There are several forms of binding potential with different constants of proportionality, resulting from different ways of approximating the free concentration, but the version that is frequently reported in PET and SPECT studies is BP_{ND}, the ratio C_S/C_{ND} at equilibrium. From the distribution volume definitions, BP_{ND} can be seen to equal V_S/V_{ND}. Physiologically, BP_{ND} is equivalent to $f_{ND}B_{MAX}/K_D$, where B_{MAX} (nM) is the concentration of the target protein. Receptor studies report this as B_{AVAIL} to emphasize the fact that some portion of the total concentration B_{MAX} may not be available to bind to the radiotracer due to phenomena such as baseline receptor occupancy by the endogenous neurotransmitter or trafficking of receptors between the cell surface and cytoplasm. K_D (nM) is the equilibrium dissociation constant (reciprocal of the affinity). From the equilibrium definition C_S/C_{ND}, it is apparent that BP_{ND} is a ratio of specific binding, the quantity of most interest, to nondisplaceable concentration, which incorporates nonspecific binding, essentially a nuisance variable. For this reason, BP_{ND} is sometimes thought of as a signal-to-background ratio. In studies where C_P is measured, V_T can be estimated in all brain regions. BP_{ND} can be estimated through the use of the 2TC model provided there is a reference tissue, a brain region with negligible target protein concentration (Fig. 1.27–10). In a reference tissue, $V_T = V_{ND}$. An additional assumption applied in all compartment models that is well supported by data within experimental error is that V_{ND} is the same in all brain tissues so that V_T in the reference tissue provides an estimate of V_{ND} throughout the brain (Fig. 1.27–10). In theory, BP_{ND} is equivalent to the ratio of rate constants k_3/k_4 in the 2TC model. But in reality, this is not a reliable method to determine BP_{ND} when applying a statistical curve fitting to real data. The individual rate constants and their ratios suffer from poor identifiability—many different combinations of rate constants can fit the same data well. V_T, on the other hand, is identifiable, and a more robust way to estimate BP_{ND} in all brain regions is through distribution volumes as $[V_T(\text{region})—V_T(\text{reference tissue})]/V_T(\text{reference tissue})$. Also, if there is a reference tissue, there are mathematical models that eliminate the arterial plasma concentration C_P from the equations and estimate BP_{ND} directly from the relationship between the target-rich brain regions and the reference tissue; these allow BP_{ND} estimation without the requirement for arterial blood samples.

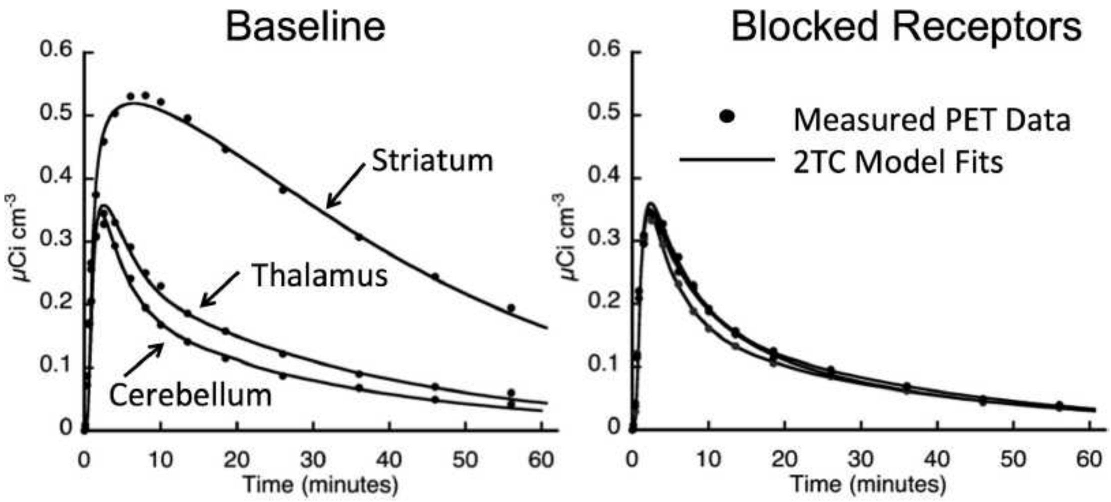

FIGURE 1.27–10. Time activity curves and 2TC fit at baseline and after receptor blockade. The data are from [^{11}C]raclopride scans, a dopamine D2-like receptor tracer, in a nonhuman primate at baseline (**left**) and following D2 receptor blockade with haloperidol (**right**). The figure illustrates several aspects of 2TC models. Distribution volumes (V_T) at baseline are 3.9 in the striatum, 1.4 in the thalamus, and 1.0 in the cerebellum. Using cerebellum V_T as an estimate of V_{ND} leads to BP_{ND} = 2.9 in the striatum and 0.4 in the thalamus. Following receptor blockade, all regions appear similar to the cerebellum, and all V_T are close to 1, illustrating the concept that V_T in the reference tissue, the cerebellum, provides a good estimate for V_{ND} in regions containing receptors.

Experimental Designs for Reversible Models. The most direct application of reversible models is the comparison of target protein density across groups or conditions. If the assumption is made that K_D and f_{ND} do not differ, on average, between scans being compared, then BP_{ND} can be taken to represent differences in B_{MAX} or B_{AVAIL}. Another flexible paradigm that can model many different phenomena is to compare the relative change in BP_{ND} across conditions within subjects, sometimes denoted as ΔBP_{ND}: $\Delta BP_{ND} = [BP_{ND}(\text{condition 2}) – BP_{ND}(\text{condition 1})]/BP_{ND}(\text{condition 1})$. The most common applications are those in which condition 1 is a baseline measurement, and condition 2 is a scan following a pharmacologic challenge that perturbs the system. Stimulant challenges induce an endogenous ligand for the binding site, such as a neurotransmitter in the case of receptor binding, to be released and compete with the radioligand for binding sites. Another example is a depletion paradigm, in which the challenge prevents or reduces the synthesis of the endogenous substance that normally binds to the site under baseline conditions, leading to the unmasking of the underlying B_{MAX}. ΔBP_{ND} also provides an estimate of the target fraction, or percent occupancy, attained by an exogenously administered drug. In the simplest case, when there is only a single competing ligand occupying some portion of the target molecules in condition 2 and no competitors in condition 1, BP_{ND} will be reduced by the factor $1/(1 + [L]/K_{DL})$ where [L] is the concentration of the competitor, and K_{DL} is the equilibrium dissociation constant of L for the target. The factor $1/(1 + [L]/K_{DL})$ equals 1 – occupancy, where occupancy is the fraction of the target concentration occupied at equilibrium by L at concentration [L]. Therefore, ΔBP_{ND} provides an estimate of the target occupancy attained by competitor L, and PET imaging has become an important tool in drug development and characterization of existing drugs. In vivo occupancy studies with PET have become nearly de rigueur for demonstrating target engagement by candidate drugs, and the need for such studies has motivated the development of tracers for new target molecules under investigation as sites for drug therapy. Occupancy study designs have included both simple measurement of receptor occupancy in stably dosed patients and more comprehensive designs utilizing experimental ranges of doses to form models of the relationship between doses or drug plasma levels and target occupancy, with estimates of related parameters such as ED_{50} or EC_{50} (dose and blood plasma concentration at which 50% of the target is occupied). Studies where ΔBP_{ND} is used to measure changes other than target occupancy by an exogenously administered drug, especially ones that utilize pharmacologic challenges to induce changes in neurochemical systems outside of the normal physiologic range, likely involve underlying pharmacokinetics or biophysical processes that are more complicated than the simple competition model implies, but ΔBP_{ND} is still an informative parameter to quantify the magnitude of the challenge effect.

Irreversible Models. The most commonly used irreversible compartment model also contains two tissues but has no route of return from the terminal compartment to the intermediate compartment as in the reversible model; the radiotracer or radiolabeled metabolic byproduct is effectively trapped in the final state. This model is most often applied in cases where the radiotracer is a substrate for several, but not all, of the steps in the metabolic pathway of an endogenous substance, with the radiolabel retained on the metabolic byproduct of each of these steps. If the process by which clearance from the last shared step is slow enough such that virtually all of the accumulated radiolabeled metabolite remains in that state for the duration of the scan, it can be treated as a trapped or irreversible state. Unlike the reversible models, irreversible models do not have a broadly agreed-upon nomenclature, but here the terminology C_{PRE} (for precursor, subsuming all reversible states in the brain prior to the last, trapped state) and C_M (for metabolite, the trapped state) are adapted. While there is no equilibrium state in an irreversible model, a hypothetical steady state can be envisioned in which the efflux and influx of C_{PRE} to and from the arterial plasma are just balanced by the forward transformation of C_{PRE} to $C_{M,}$ and the concentration C_{PRE} remains constant. From the compartment model equations, this steady-state would satisfy the condition $C_{PRE} = [K_1/(k_2 + k_3)]C_P$. The overall rate of trapping, equal to k_3C_{PRE}, would then equal $[K_1k_3/(k_2 + k_3)]C_P$. The expression $K_1k_3/(k_2 + k_3)$ is called K_{in}, the steady-state uptake rate constant of the tracer, and is the usual outcome measure estimated in irreversible models. In a similar fashion to reversible models, if a reference tissue for the irreversible process exists, a modified version of K_{in} can be estimated without the need for collecting arterial plasma.

Examples of Irreversible Models. The assumptions in irreversible models are more specific to the individual applications than with reversible models. A general sense of the way they work is most readily described through the examples of [^{18}F] DOPA and ^{18}FDG, two of the most widely used metabolism tracers in brain PET.

[^{18}F] DOPA is a substrate for amino acid decarboxylase (AADC), the enzyme that catalyzes L-Dihydroxyphenylalanine (DOPA) into dopamine within the dopaminergic terminals. [^{18}F] DOPA readily passes the BBB and cell membranes and is metabolized by AADC into 6-fluorodopamine (6-FDA), effectively ^{18}F-labeled dopamine, in dopaminergic terminals. 6-FDA, like endogenous dopamine, does not cross the BBB and so is "trapped" in the proximity of the terminals and the cycle of vesicular loading, release through exocytosis and reuptake into the terminal, in parallel with endogenous dopamine in the striatum, where reuptake through dopamine transporters is the dominant mode of clearance from the synapse. K_{in}, estimated with the irreversible trapping model, represents a marker of presynaptic dopaminergic condition. The trapping is only approximate because 6-FDA is a substrate for monoamine oxidase (MAO) and for catechol-O-methyltransferase (COMT), and the radiolabeled metabolic by-products from 6-FDA metabolism can readily diffuse across the BBB. Several approaches have been proposed to incorporate the slow efflux from the brain, sometimes called k_{loss}, into the modeling and design of [^{18}F] DOPA experiments, but many studies quantify K_{in} with the more basic irreversible model. [^{18}F] DOPA data sets can comprise anywhere between one and several hours of dynamic imaging. Several brain regions, including the cerebellum, can function as a reference tissue for [^{18}F] DOPA imaging.

^{18}F-labeled 2-fluoro-2-deoxy-D-glucose (^{18}FDG) is by far the most widely used PET radiotracer. Originally developed for quantifying brain glucose metabolism, it was observed early on that ^{18}FDG is taken up more rapidly by some types of tumor cells than by the surrounding tissue, and ^{18}FDG imaging became a widely used diagnostic and staging tool in oncology and other medical fields. It is still frequently used in research for the quantification of brain glucose metabolism as well. ^{18}FDG is a substrate for the same carrier protein that transports glucose into brain tissue. It is also a substrate for hexokinase, the enzyme that metabolizes glucose, and is phosphorylated into ^{18}FDG-6-PO4. ^{18}FDG-6-PO4 does not follow further steps of glucose metabolism. It does dephosphorylate, but at a rate that is very slow compared to the forward process, and ^{18}FDG-6-PO4 is treated as a trapped state. Here, K_{in} represents the steady-state phosphorylation rate of ^{18}FDG. This is proportional, not identical, to the rate of these steps in brain glucose metabolism, owing to the fact that the transport and phosphorylation rates are different for the two compounds. When multiplied by a conversion factor accounting for this difference (1/LC, the "lumped constant" because it lumps two conversion factors together), K_{in}/LC times the plasma glucose concentration is taken as an estimate of brain glucose metabolism. As glucose is utilized throughout the brain, there are no reference tissues. The most comprehensive approach to estimating K_{in} with ^{18}FDG is to obtain arterial samples and fit the PET data to the 2TC irreversible model with arterial input or a modified version that accounts for the slow dephosphorylation, but this is rarely done in clinical research studies, where simplifications of both the blood input and the PET data acquisition are frequently employed. Simplification of blood input data can entail utilizing venous blood that has been “arterialized” by warming the site of the venous catheter to produce vasodilation and more rapid equilibration between arterial and venous ^{18}FDG concentrations, or even simpler, collecting only a small number of samples and using these to scale a standardized population input. PET data acquisition can be reduced by approximating K_{in} as the ^{18}FDG concentration in the brain at steady state divided by the area under the curve of the arterial plasma input; the image of the average ^{18}FDG concentration over a relatively short period of time, typically 30 minutes or less of acquisition starting between 45 to 60 minutes after a bolus injection, is used in this approach. Finally, the steady-state approach can be simplified even further by estimating a parameter called the standard uptake value (SUV), in which the area under the curve of the ^{18}FDG in blood is replaced by the injected radioactivity divided by the subject’s body weight. SUV is a widely used outcome measure in clinical scans but has come under criticism as a research tool because, without any direct assessment of the concentration of ^{18}FDG in blood, the method cannot account for differences across subjects or conditions in the rate of peripheral clearance of ^{18}FDG, and therefore can only provide information about relative metabolism across the brain within a scan.

RADIOTRACERS

Table 1.27–2 lists PET and SPECT radiotracers that have been successfully used in human subjects in recent years. Tracers for new targets and improved tracers for established targets are constantly under development; the table is intended to convey a sense of the resources available in 2021. Many more tracers have been proposed but have either failed to be adapted or have not yet progressed beyond testing in vitro or in preclinical species. The pharmacokinetic properties that lead to useful reversibly binding radiotracers are delicate and somewhat different than those that might make a compound suitable for *in vitro* binding experiments or as a therapeutic drug. In general, balance is required between a number of different properties in order for quantification to be reliable. For example, a prerequisite is that a tracer can be adequately supplied to brain tissues in order to provide a detectable signal. Several factors can affect this. The tracer will bind nonspecifically to plasma proteins in the blood. The protein-bound fraction (the unbound fraction is called f_P; $1 - f_P$ is the bound fraction) can be very high; for some radiotracers $1 - f_P > 99\%$. In some cases, this may prevent an adequate concentration of radiotracer from entering the brain for reliable image quantification. BBB penetrability is an issue that must be addressed in the *in vivo* setting that does not arise in cell cultures or *in vitro* assays utilizing brain tissue. A property related to this is lipophilicity. High lipophilicity generally improves the potential of a compound for BBB penetration, but it also leads to high nonspecific binding. If the nonspecific binding is too high relative to specific binding, the specific binding signal will be difficult to separate from the background, leading effectively to low BP_{ND} and unreliable quantification. Thus lipophilicity must have an intermediate value (log P in the range of 1.5 to 4) to balance between the need for brain penetrance and the undesirable effect of high nonspecific binding. Also related to BBB penetrability is the fact that some radiotracer candidates are substrates or weak substrates for P-glycoprotein (P-gp) or other efflux mechanisms that actively prevent brain entry; this can be hard to predict prior to *in vivo* testing. Another property that needs to be considered is the relationship between the affinity of the tracer for its target and the target density. As $BP_{ND} = f_{ND}B_{MAX}/K_D$, it can be seen that for a given target availability and f_{ND}, BP_{ND} varies directly with affinity. If BP_{ND} is too low, specific binding cannot be reliably measured relative to nonspecific binding (Fig. 1.27–11). Conversely, if the target density is very high, a high-affinity tracer may not reach a washout phase within a reasonable time frame, and parameter estimates will not be stable—small perturbations of the data late in the scan will lead to large changes in outcome measures (Fig. 1.27–11). Targets that are distributed very heterogeneously in the brain may require multiple radiotracers, with moderate affinity tracers for high-B_{MAX} regions and high-affinity tracers for low-B_{MAX} regions. Some tracers that appear promising in development may turn out to be too far to one end or the other of this spectrum to be useful in humans: either BP_{ND} too low everywhere or washout too slow everywhere. Another phenomenon that can limit the usefulness of a radiotracer is the discovery that *in vivo*, its selectivity for the primary target molecule compared to a secondary target is less than predicted *in vitro*. In general, tracers are screened *in vitro* for binding affinity to many brain proteins, but there have been cases where the *in vitro* affinity for a primary target is several hundredfold greater than for another protein, but *in vivo,* the affinities are found to differ by much less and binding to the secondary target presents a confound for quantification. Finally, a tracer may appear to have high potential according to all of the above criteria and then turn out to have a radiolabeled metabolite in the blood that crosses the BBB and makes quantification difficult or impossible.

Chemists working in radiotracer development can control some of these parameters, but much of the process is trial and error. Development might entail starting with an initial pharmacophore or a molecule known to have specificity for the target and then iterating through the sequence of devising a radiolabeling procedure, testing the result *in vitro*, testing in preclinical species, altering the molecule to improve properties based on test results, and repeating until a suitable candidate has strong enough potential for human testing. This process can be long and arduous, often taking years to arrive at a suitable radiotracer for a given target. Some recent work has focused on ways to streamline the process or on the development of a rational approach to predicting the quantitative potential of new tracers early in their development.

NEUROCHEMICAL IMAGING IN PSYCHIATRIC RESEARCH

As Table 1.27–2 demonstrates, radiotracers for imaging many different neurochemical systems have progressed as far as human imaging. Some of these are recently developed or have only been used in one or several studies, whereas dozens or even hundreds of studies of neuropsychiatric illnesses and other conditions related to brain function have been published using some of the more established tracers. Some of the most studied areas are schizophrenia and related conditions, major depression and other mood and affect disorders, substance and alcohol abuse and addiction, and more recently, Alzheimer disease and other dementias, but many more have been investigated. Some of these extensively studied areas and some well-replicated results are described here. This section is not intended to be a comprehensive review of all neuroimaging results in psychiatry but rather a selective sample to convey the nature of the research, some of the more well-established or informative results, and cases illustrating the capabilities as well as limitations of radiotracer imaging.

Table 1.27–2.

PET and SPECT Radiotracers in Published Studies with Human Subjects

Target System	Specific Target	PET Radioligand	SPECT Radioligand
Dopamine	D2/D3 receptor	[^{11}C]raclopride	[^{123}I]IBZM
		[^{18}F]fallypride	[^{123}I]epidepride
		[^{11}C]fallypride	
		[^{11}C]FLB457	
		[^{11}C]-(+)-PHNO	
		[^{11}C]NPA	
		[^{11}C]MNPA	
	D1/D5 receptor	[^{11}C]NNC112	
		[^{11}C]SCH23390	
		[^{18}F]MNI-968	
	Reuptake transporter	[^{11}C]PE2I	[^{123}I]PE2I
		[^{11}C]altropane	[^{99m}Tc]TRODAT-1
		[^{18}F]CFT	[^{123}I]β-CIT
		[^{11}C]CFT	[^{123}I]FP-CIT
	DA terminals	[^{18}F]DOPA	
		[^{18}F]FMT	
	VMAT2	[^{11}C]DTBZ	
		[^{18}F]AV133	
Serotonin	SERT	[^{11}C]McN5652	[^{123}I]β-CIT
		[^{11}C]DASB	[^{123}I]ADAM
		[^{11}C]AFM	
		[^{11}C]HOMADAM	
	5-HT$_{1A}$ receptor	[*carbonyl*-^{11}C]WAY100635	
		[^{18}F]MPPF	
		[^{18}F]FCWAY	

		[^{18}F]MEFWAY	
		[^{11}C]CUMI101	
	5-HT$_{2A}$ receptor	[^{11}C]MDL100907	[^{123}I]5-I-R91150
		[^{11}C]CIMBI-36	
		[^{18}F]Altanserin	
	5-HT$_{1B}$ receptor	[^{11}C]P943	
		[^{11}C]AZ10419369	
	5-HT$_{6}$ receptor	[^{11}C]GSK215083	
Norepinepherine	NET	[^{18}F]MeNER	
		[^{11}C]MRB	
Opioid	μ Opioid receptor	[^{11}C]carfentanil	
	Opioid receptor	[^{11}C]diprenorphine	
	κ Opioid receptor	[^{11}C]GR103545	
		[^{18}F]LY2459989	
		[^{11}C]EKAP	
Acetylcholine	α4β2* nicotinic receptor	2-[^{18}F]-FA-85380	[^{123}I]5-1A-85380
		[^{18}F]AZAN	
		[^{18}F]-(-)-Flubatine	
		[^{18}F]-XTRA	
		[^{18}F] Nifene	
	α7 nicotinic receptors	[^{18}F]ASEM	
	Muscarinic M1 receptors	[^{11}C]LSN3172176	
	Muscarinic M2 receptors	[^{18}F]FP-TZTP	
	Muscarinic receptors (nonselective)	[^{11}C]NMBP	
		[^{11}C]-(+)-MPB	

	Vesicular acetylcholine Transporter	[^{18}F]FEOBV	[^{123}I]IBVM
		[^{18}F]VAT	
	Acetylcholinesterase	[^{11}C]MP4A	
		[^{11}C]PMP	
Glycine	GlyT1	[^{11}C]GSK931145	
		[^{11}C]RO5013853	
GABA	GABA$_A$ receptor	[^{11}C]flumazenil	[^{123}I]iomazenil
		[^{11}C]Ro15-4513	
Adenosine	Adenosine 2A receptor	[^{18}F]CPFPX	[^{123}I]MNI420
Glutamate	mGluR5	[^{11}C]ABP688	
		[^{18}F]FPEB	
Synaptic vesicle Protein	SV2A	[^{11}C]UCB-J	
Substance P	NK1 receptor	[^{18}F]SPA-RQ	
Enzymes	MAO B	[^{11}C]deprenyl	
	MAO A	[^{11}C]clorgyline	
		[^{11}C]harmine	
	PDE4	[^{11}C]rolipram	
	PDE10	[^{11}C]IMA107	
		[^{18}F]MNI659	
Neuroinflammation	TSPO	[^{11}C]PK11195	[^{123}I]PK11195
		[^{11}C]PBR28	[^{123}I]CLINDE
		[^{18}F]PBR06	
		[^{18}F]PBR111	
		[^{18}F]FEPPA	
		[^{11}C]ER176	
	cyclooxygenase-1	[^{11}C[PS13	
	cyclooxygenase-2	[^{11}C]MC1	

Cannabinoid	CB1 receptor	[^{11}C]OMAR	[^{123}I]AM281
		[^{18}F]MK9470	
	FAAH	[^{11}C]CURB	
Misfolded proteins	β-Amyloid	[^{11}C]PiB	
		[^{11}C]AZD2184	
		[^{11}C]AZD4694	
		[^{18}F]GE067	
		[^{11}C]SB13	
		[^{18}F]BAY94-9172	
		[^{18}F]AV45	
		[^{18}F]FDDNP†	
	Tau	[^{18}F]T807	
		[^{18}F]-THK105	
		[^{18}F]-THK523	
		[^{11}C]-PBB3	
		[^{18}F]MK6240	

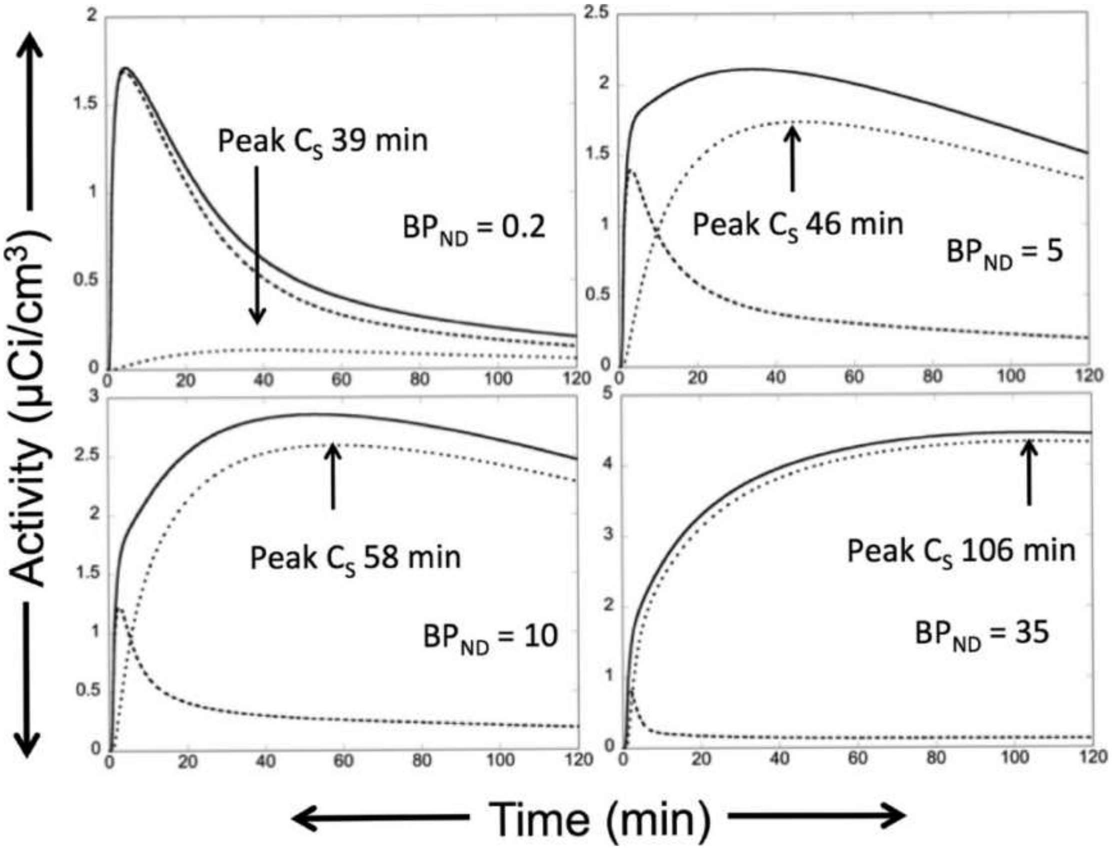

FIGURE 1.27–11. Time to peak-specific binding in simulated 2TC time–activity curves. These simulated curves were all created with the same input function and the same K_1, k_2, and k_4 values; they differ only in k_3, which is proportional to B_{AVAIL}. They illustrate the fact that the time of the peak-specific binding increases with increasing B_{AVAIL}. In the **upper left panel** ($BP_{N\,D}$ = 0.2), the majority of the activity is attributable to the C_{ND} curve, so that it will be difficult to distinguish C_T from $C_{ND,}$ and BP_{ND} will be difficult to estimate accurately; the signal-to-background ratio is too low. In the **lower right panel** (BP_{ND} = 35), the peak-specific binding occurs so late in the scan that small perturbations of the last several data points could lead to large differences in estimated V_T or BP_{ND}. The **other two panels** (BP_{ND} = 5, 10) illustrate cases where parameter estimation would be robust, both because C_S comprises a large part of C_T, and peak-specific binding and a washout phase are reached during the course of the scan. *Dashed curves* are C_{ND}, *dotted curves* are C_S, and *solid curves* are C_T.

Dopamine

The dopaminergic system is one of the most extensively studied neurotransmission systems with both PET and SPECT imaging. Studies have examined receptor and transporter availability as well as transmitter synthesis, storage, and release capacity in cohorts from a broad range of patient populations. Several tracers have also been developed for intracellular enzymes involved in the signaling pathways of dopamine receptors, such as phosphodiesterase 10A (PDE10A). Because all antipsychotic drugs block dopamine D2 receptor transmission to some degree, and because of the role of dopamine transmission in functions such as attribution of salience to environmental stimuli or fine-tuning of working memory, dopaminergic dysfunction is considerably implicated in schizophrenia, and there have been a particularly large number of imaging studies characterizing different aspects of the dopaminergic system in patients with schizophrenia. Another area that has produced many imaging studies of the dopaminergic system is substance dependence. Many drugs of abuse induce the synaptic release of dopamine. This is especially true of drugs such as cocaine and amphetamines that act directly on dopamine terminals but is also true for other substances such as alcohol or nicotine that act indirectly. Because of this, and because of dopamine's roles in motivation, identification of rewarding stimuli, and reinforcement of rewarding behaviors, dopaminergic dysfunction has been implicated in drug abuse and addiction, and this has been an area of focus for imaging.

D1 Receptor Availability. Dopamine D1-like receptors are distributed throughout most of the brain, with a high concentration in the striatum and low to moderate concentration in the frontal cortex. D1 concentration in the cerebellum is below the detectable level with currently available tracers, and the cerebellum is treated as a reference tissue. Because of the involvement of cortical dopaminergic function in cognition and working memory, there has been great interest in assessing cortical D1 in conditions where these are affected, especially schizophrenia. There are two PET radiotracers that have been most frequently used to assess D1 receptor availability. These are [^{11}C]NNC112 and [^{11}C]SCH23390. Neither tracer can distinguish between D1 and D5 receptors. The signal-to-noise ratio, embodied in BP_{ND} values, is approximately threefold higher for [^{11}C]NNC112 than for [^{11}C]SCH23390. Also, the rapid peripheral metabolism of [^{11}C]SCH23390 has made it difficult to measure arterial plasma concentration reliably, so [^{11}C]SCH23390 is used exclusively with reference tissue methods, whereas [^{11}C]NNC112 can be used with both arterial input and reference tissue methods. Several studies have been performed to examine D1 availability in patients with schizophrenia. The studies utilizing [^{11}C]NNC112 have found increased cortical D1 in antipsychotic drug-free patients and especially patients without prior history of antipsychotic drug treatment compared to demographically matched control subjects, whereas the studies utilizing [^{11}C]SCH23390 have found decreased D1. Initially, it was thought this might reflect differences in tracer properties, but subsequent studies from both sets of investigators, comparing subsamples of the original cohorts who had been administered both tracers, have yielded results in the same directions as originally reported, suggesting this may be more related to population heterogeneity than to the tracers. During the first decade of the 2000s, several investigators demonstrated that the selectivity of both [^{11}C]NNC112 and [^{11}C]SCH23390 for D1 compared to serotonin 5-HT$_{2A}$ receptors is approximately 10-fold rather than the 100-fold difference measured in vitro and that 25% to 35% of cortical BP$_{ND}$ is attributable to 5-HT$_{2A}$ binding for both tracers. Investigators presented some evidence that 5-HT$_{2A}$ binding probably did not contribute to group differences in the schizophrenia studies, but unless more D1 selective radiotracers are developed, future studies will need to account for this factor when interpreting their results.

D2/D3 Receptor Availability. Dopamine D2-like receptors are heterogeneously distributed, with a high concentration in the striatum, and low concentration in the thalamus, midbrain (substantia nigra and ventral tegmental area, SN/VTA), and medial temporal cortical regions (amygdala, hippocampus) and very low concentration in the frontal cortex. These receptors are among the most studied with imaging, and many tracers have been developed. Widely used tracers are [^{11}C]raclopride for PET and [^{123}I]IBZM for SPECT. Both are suitable for striatal imaging but do not have high enough BP$_{ND}$ for imaging outside the striatum. The cerebellum is frequently used as a reference tissue for each, and neither tracer distinguishes between D2 and D3 receptors. Several higher affinity tracers have been developed for extrastriatal imaging, including [^{18}F]fallypride and [^{11}C]FLB457 for PET and [^{123}I]epidepride for SPECT. [^{11}C]FLB457 can only be used quantitatively in the cortex and limbic regions as striatal binding does not reach equilibrium within a suitable time frame for ^{11}C imaging. Also, unlike most other tracers for D2-like receptors, specific binding of [^{11}C]FLB457 is detectable in the cerebellum. [^{18}F]fallypride, with its more long-lived radioisotope, can be used to simultaneously image the striatum and extrastriatal regions, though cortical BP$_{ND}$ is still very low, and striatal imaging requires lengthy scanning of at least 3 to 4 hours. Specific binding in the cerebellum has also been detected with [^{123}I]epidepride, which is chemically similar to [^{11}C]FLB457. The long half-life of ^{123}I allows simultaneous striatal and extrastriatal imaging, but again, with very long scan times in excess of 4 hours. In addition, some investigators have detected radiolabeled metabolites of [^{123}I]epidepride in the brain. None of these high-affinity tracers can be used to distinguish between D2 and D3 receptors. [^{11}C]-(+)-PHNO is an agonist radiotracer that is strongly D3-preferring. Estimates of its in vivo D3:D2 selectivity range from 20- to 50-fold. D3 receptors are much less widely distributed and in lower concentrations than D2 receptors, but the D3-preferring nature of [^{11}C]-(+)-PHNO has allowed noninvasive in vivo characterization of D3 receptors for the first time. Binding is quantifiable in the striatum, the globus pallidus, the thalamus, the SN/VTA, and the hypothalamus. Dorsal striatum BP_{ND} is attributable mostly to D2 binding, SN/VTA and hypothalamus BP_{ND} are attributable mostly to D3 binding, and globus pallidus, thalamus, and ventral striatum BP_{ND} represent a mixture of both receptor subtypes. The cerebellum is used as a reference tissue.

D2 Receptor Availability in Schizophrenia

Studies of striatal D2 receptor availability in schizophrenia have been mixed, with meta-analyses suggesting that striatal D2-like receptor availability may be slightly elevated, if at all. Most of these studies were performed with [^{11}C]raclopride or [^{123}I]IBZM. One recent study also failed to detect a difference in striatal [^{11}C]-(+)-PHNO BP_{ND} between patients with schizophrenia and matched control subjects. Studies have also been published assessing extrastriatal D2 availability in schizophrenia using [^{18}F]fallypride, [^{11}C]FLB457, and [^{123}I]epidepride imaging. Individual studies have reported differences between patients and controls in several brain regions such as the temporal cortex, anterior cingulate, or thalamus, but these have not been seen consistently across studies or even go in opposite directions, suggesting that observed changes were either state dependent or due to population heterogeneity. In summary, the imaging literature to date suggests that dopaminergic dysfunction in schizophrenia is not likely mediated through receptor expression levels but rather through either pre- or postsynaptic phenomena.

D2 Receptor Availability in Substance Dependence

In substance dependence, a very different picture has emerged in which studies have consistently shown decreased striatal D2 availability in subjects who are dependent on a variety of substances, including cocaine, methamphetamine, alcohol, or heroin. An exception is cannabis use, in which such differences have not been detected in several studies using [^{11}C]raclopride in moderate users or in one study that imaged heavy users with [^{11}C]-(+)-PHNO.

Dopamine Release, Synthesis, and Storage Capacity. Many studies have used imaging methods to indirectly assess aspects of dopamine transmission. These include studies measuring ΔBP_{ND} of a receptor-binding tracer to infer levels, or changes in levels, of synaptic dopamine following pharmacologic challenges and metabolism studies using mainly [^{18}F]DOPA to assess presynaptic dopamine tone. In theory, any radiotracer that binds to receptors should be sensitive to changes in the synaptic concentration of the endogenous neurotransmitter, but for reasons that are not yet well understood, radiotracers for D2-like receptors have turned out to be the most sensitive. Thus, studies examining dopamine transmission have mainly used [^{11}C]raclopride or [^{123}I]IBZM. More recently, several studies have been published using [^{18}F]fallypride or [^{11}C]FLB457 to examine extrastriatal dopamine transmission, or [^{11}C]-(+)-PHNO, which provides another perspective due to its agonist and D3-preferring properties. Two of the most widely used pharmacologic compounds for increasing synaptic dopamine are dextroamphetamine (d-AMPH) and methylphenidate (MPH). d-AMPH causes dopamine to be extruded from intracellular synaptic vesicles and to be released in synapses by reversing the direction of action of the dopamine reuptake transporter (DAT). ΔBP_{ND} in the striatum between a baseline scan and a post–d-AMPH scan shows decreased binding ($\Delta BP_{ND} <0$) consistent with competition at the receptors between the radiotracer and dopamine. The d-AMPH paradigm has been carefully validated; preclinical testing has demonstrated that measured ΔBP_{ND} is strongly correlated with released dopamine levels measured by microdialysis, that the measured levels are d-AMPH dose dependent, and that the d-AMPH effect is negated ($\Delta BP_{ND} \sim 0$) when dopamine is depleted prior to administration of d-AMPH, that is, the effect requires the presence of dopamine. There is some question as to whether, in addition to competition between the radiotracer and dopamine, the d-AMPH effect is mediated by other processes such as trafficking of receptors between the cell surface and cytoplasm because the decreased BP_{ND} is sustained for at least several hours longer than the dopamine surge following administration, but nonetheless ΔBP_{ND} following d-AMPH has been accepted as a reliable indicator of the capacity for dopamine synthesis, storage, and release. Microdialysis results from nonhuman primates show increased extracellular concentration of dopamine following d-AMPH of 1,000% or more and suggest that in the dose range administered to human subjects, dopamine release is high, but the effects on BP_{ND} are still small, with striatal ΔBP_{ND} in the range of 10% to 20% for [^{11}C]raclopride, [^{123}I]IBZM, and [^{18}F]fallypride. There is some evidence that dopamine may compete more effectively with agonist tracers than antagonists. Studies examining d-AMPH–induced dopamine release with [^{11}C]-(+)-PHNO or with [^{11}C]NPA, another agonist radiotracer that is not D3 preferring, found a larger magnitude ΔBP_{ND} than when the same subjects were scanned with [^{11}C]raclopride. This effect may be due to the agonist and dopamine competing for the same subset of receptors that are G-protein coupled and therefore in a high-affinity state for agonist binding, whereas the antagonist tracer binds with equal affinity to coupled and uncoupled receptors. This interpretation, however, is somewhat speculative and has yet to be rigorously demonstrated as the cause of the difference in tracer responses to dopamine release. MPH blocks reuptake through the transporter, so that released dopamine accumulates in the synapse and extracellular space. The magnitude of accumulated dopamine concentration following MPH is less than the amount released following d-AMPH, but the effect is still reliably detected with imaging and MPH challenge has also been widely used as a measure of presynaptic dopamine tone. Because radiotracers for D2-like receptors are sensitive to binding competition from endogenous dopamine, and there is a sustained, low baseline level of dopamine binding to receptors, there is some question as to the extent to which apparent baseline BP_{ND} is reduced by baseline dopamine. Depletion studies, in which preadministration of a compound that pharmacologically prevents dopamine synthesis or vesicular storage, are designed to unmask the contribution of baseline dopamine to apparent BP_{ND}. The tyrosine hydroxylase inhibitor α-methyl-para-tyrosine (AMPT) prevents dopamine synthesis and causes depletion of dopamine stores when administered over a period of 48 hours or more; AMPT has been used in human subjects to unmask baseline occupancy of D2 receptors by dopamine.

DOPAMINE TRANSMISSION IN SCHIZOPHRENIA. Studies using the d-AMPH challenge have consistently shown greater dopamine release (larger magnitude of ΔBP_{ND}) in patients with schizophrenia compared to matched control subjects. These have been observed at the level of the whole striatum, using [^{123}I]IBZM or [^{11}C]raclopride imaging. Studies using the AMPT depletion paradigm have found elevated baseline dopamine at the level of the whole striatum using [^{123}I]IBZM and, more recently, specifically in the associative striatum (caudate and precommissural putamen), using [^{11}C]raclopride imaging. These studies suggest that in humans, baseline occupancy of D2 receptors by dopamine is in the range of 10% to 20%. One study in which a small cohort of patients was imaged with both d-AMPH and AMPT challenges showed a high correlation between these, that is, greater AMPH-induced dopamine release was associated with greater baseline dopamine binding as inferred from AMPT unmasking. Thus imaging measures of dopamine transmission have consistently found elevated presynaptic tone in the striatum in schizophrenia. This is also in accord with [^{18}F] DOPA imaging in patients with schizophrenia and prodromal subjects at high risk for schizophrenia which have shown elevated K_{in} compared to matched controls in patients and intermediate elevation in prodromal subjects, with an association between the magnitude of [^{18}F] DOPA K_{in} and the likelihood of progression to psychosis in the prodrome. Conversely, one study used the d-AMPH paradigm in conjunction with [^{11}C]FLB457 imaging and found blunted dopamine release in patients with schizophrenia in the cortex and midbrain, in line with theoretical models of cortical hypodopaminergic transmission in schizophrenia. It has yet to be demonstrated that the two phenotypes, increased release in the striatum and decreased release in the cortex, occur in the same patients.

DOPAMINE TRANSMISSION IN SUBSTANCE DEPENDENCE. In contrast to patients with schizophrenia, in substance-dependent subjects, reduced dopamine release relative to matched controls has been most frequently observed following either d-AMPH or MPH administration. This has been observed in cocaine users, heroin users, heavy cannabis users, and alcohol-dependent subjects. One study also used the AMPT depletion model to show that baseline dopamine levels were reduced in cocaine users. Combined with observations of reduced D2-like receptor availability, the imaging literature suggests an overall reduced function or blunting of the dopaminergic system in substance dependence. These studies cannot address the question of whether low dopamine function is a predisposing factor or consequence of drug use, but one recent study using a nonhuman primate model of cocaine dependence found that presynaptic vesicle availability in dopamine terminals, as measured with the VMAT2 radiotracer [^{11}C]DTBZ, is lowered following chronic cocaine administration, suggesting a direct role for cocaine itself in the observed reduction of dopamine tone in users.

Dopaminergic Target Engagement. As far back as the late 1980s and early 1990s, PET studies were published using ΔBP_{ND} to infer receptor occupancy by commonly used antipsychotic drugs. It was shown, using [^{11}C]raclopride, that haloperidol (Haldol) and sulpiride (Dogmatil, Dolmatil, Eglonyl, Espiride, Modal, Sulpor) administered at therapeutic doses occupied between 75% and 90% of D2-like receptors and, using [^{11}C]SCH23390, that binding of both drugs to D1-like receptors was below detectable levels. The atypical antipsychotic clozapine (Clozaril) was shown to occupy ~50% of D2-like receptors and ~40% of D1-like receptors. These papers also were the first to demonstrate a direct link between extrapyramidal symptoms (EPS) and the level of D2 receptor occupancy, showing a threshold for the appearance of EPS at approximately 80% occupancy. Since that time, receptor occupancy has been characterized for many antipsychotic drugs as well as drugs from many other classes. Studies using [^{11}C]raclopride have repeatedly shown some widely used atypical antipsychotics such as risperidone (Risperdal) or olanzapine (Zyprexa) occupy 70% to 80% of striatal D2-like receptors in patients on stable doses and have also replicated earlier results that D2 occupancy by therapeutic levels of clozapine is lower than for other antipsychotics. In addition, risperidone, olanzapine, and clozapine have all been shown to occupy a greater percentage of serotonin 5-HT$_{2A}$ receptors than dopamine D2 receptors using radiotracers for 5-HT$_{2A}$. Studies measuring striatal and extrastriatal occupancy of D2-like receptors by the partial agonist antipsychotic aripiprazole (Abilify) using [^{18}F]fallypride have demonstrated very high occupancy, ~90% at stable doses, without EPS, consistent with the expectation that partial agonists should support higher levels of D2 receptor blockade without EPS than antagonists.

Serotonin

The serotonergic (5-HT) system has also been widely studied with neuroimaging. The involvement of 5-HT in major depression and other mood disorders has led to an extensive body of imaging studies measuring the availability of receptors and reuptake transporters in those conditions, but serotonergic indices have been characterized across a wide spectrum of neuropsychiatric and related conditions.

5-HT$_{1A}$ Receptor Availability. 5-HT$_{1A}$ receptors are distributed throughout the brain, with the highest concentration in the insula and medial temporal cortex, high concentration in the cingulate, and moderate concentration in the prefrontal cortex and posterior cortex (parietal and occipital cortices). The receptors are also found in the dorsal raphe nuclei (DRN), where they act as inhibitory autoreceptors. DRN concentration is low but still of particular interest because of the role of 5-HT$_{1A}$ autoreceptors in the regulation of serotonin transmission, as evidence from preclinical studies and other sources suggests that the mechanism of selective serotonin reuptake inhibitors (SSRIs) may involve desensitization of 5-HT$_{1A}$ autoreceptors in DRN over the first several weeks of treatment, leading eventually to increased serotonin transmission. [*carbonyl*-^{11}C]WAY100635, usually abbreviated as [^{11}C]WAY100635, a selective 5-HT$_{1A}$ receptor antagonist, has been the most widely used radiotracer for this target. The cerebellum is frequently used as a reference tissue with [^{11}C]WAY100635, but there is detectable specific binding in the cerebellar vermis, so the reference tissue must be defined to exclude that part of the cerebellum. [^{11}C]WAY100635 is difficult to produce due to a long and complicated synthesis procedure, and quantification is also difficult due to rapid peripheral metabolism and very low signal in the cerebellum, so there have been numerous attempts to develop modified versions and alternatives. Many of these use ^{18}F as the radiolabel and have run into problems with radiolabeled metabolites in the brain or uptake of [^{18}F]fluoride ion in the skull bone. Most recently, a new ^{11}C radiotracer, [^{11}C]CUMI101, has been introduced but has yet to be widely adopted. Thus despite the difficulties associated with its use, most published studies to date in patient populations have been performed with [^{11}C]WAY100635.

5-HT1A RECEPTOR AVAILABILITY IN DEPRESSION. There have been a large number of studies examining 5-HT$_{1A}$ availability with [^{11}C]WAY100635 in major depression disorder (MDD) and related mood conditions. The majority of these have reported decreased BP$_{ND}$ in the DRN and many other brain regions in MDD patients compared to matched controls; a few have reported no differences between groups. A small number of investigators, however, report increased binding in DRN and other regions when using the outcome measure BP$_{F}$, derived as V_S/f_P. BP$_{F}$ is a form of the binding potential that uses nonprotein bound tracer in arterial plasma as the reference value rather than the nondisplaceable concentration in the brain used with BP$_{ND}$. There has been debate as to whether these exceptional findings are due to methodologic differences, differences in patient samples, or a combination of these. Studies using BP$_{ND}$ with cerebellum as reference tissue may be biased by group differences in cerebellum binding of the tracer or by group differences in f_{ND}. Conversely, arterial plasma and f_P are challenging to measure and may contribute variability to data leading to discrepant results. Some evidence suggests that [^{11}C]WAY100635 is a weak substrate for P-gp. Possible group differences in P-gp activity have not been investigated in MDD to date, but if they exist, they might bias plasma-based outcome measures more than BP$_{ND}$. Thus, despite the importance of understanding the role of 5-HT$_{1A}$ availability in MDD, there are currently controversies in the field that may only ultimately be resolvable with the development and use of more quantitatively reliable radiotracers for this target.

Serotonin Reuptake Transporter Availability. Serotonin reuptake transporter (SERT) is heterogeneously distributed in the brain, with the highest concentration midbrain, high concentration in the striatum, amygdala, and thalamus, moderate concentration in the cingulate and hippocampus, and low concentration in the frontal cortex. The cerebellum is treated as a reference tissue for SERT tracers. The first radiotracer used to image SERT was [^{123}I]β-CIT. This tracer has an affinity for both SERT and the DAT and was thus not suitable for imaging SERT in the striatum where DAT concentration is high. Effectively, [^{123}I]β-CIT could only be used for SERT measurement in the midbrain. Subsequently, several PET radiotracers were developed for SERT imaging. The first of these, developed in the early 1990s, was [^{11}C]McN5652, which has a relatively low signal-to-noise ratio but is suitable for imaging in the midbrain and striatum. [^{11}C]DASB was developed in the late 1990s and has BP_{ND} values that are nearly double those of [^{11}C]McN5652 but are still very low in the prefrontal cortex. [^{11}C]DASB has become the most widely used radiotracer for SERT imaging. More recently, higher affinity tracers such as [^{11}C]AFM and [^{11}C]HOMADAM have been developed that may provide reliable quantification in the cortex, but studies using these have not yet been widely reported. SERT availability, measured with imaging, has been interpreted as a marker of 5-HT innervation and the integrity of the serotonergic system.

SERT AVAILABILITY IN MDMA USERS. 3,4-methylenedioxymethamphetamine (MDMA, ecstasy) releases 5-HT and increases its synaptic concentration by binding to and reversing the direction of action of SERT. MDMA has been used as a recreational drug due to its euphoria- and empathy-inducing qualities. Animal studies performed in the late 1980s indicated MDMA produced neural toxicity and reduced SERT at doses comparable to those of recreational users, raising the possibility that MDMA use could lead to reversible or even permanent neural damage. Initial studies performed in the late 1990s of SERT availability using [^{11}C]McN5652 found dramatically lowered BP_{ND} in MDMA users compared to matched control subjects. Several subsequent studies using [^{11}C]McN5652 did not replicate the dramatic result observed in the first study but did show moderately decreased BP_{ND}, in the range of 20% to 30%, in current MDMA users that appeared to be reversible both in cross-sectional studies that included former users, as well as in correlations, with the duration of abstinence or amount of use, in another study. Since that time, at least eight studies have been performed using [^{11}C]DASB in current and former MDMA users. These have consistently shown moderate decreases in cortical SERT availability in recently abstinent current MDMA users. One study examining former users found no difference compared to matched controls, suggesting the cautious interpretation of reversibility of MDMA-related neurotoxicity. An interesting aspect of this set of findings is that reported [^{11}C]DASB BP_{ND} values in the prefrontal cortex range from approximately 0.05 to 0.15 in healthy controls; these are usually considered well below the limit of quantitative reliability, but nonetheless, the result that there is reduced cortical SERT in recently abstinent MDMA users has been repeatedly replicated.

Serotonin Release, Synthesis, and Storage Capacity. Attempts to indirectly measure 5-HT release through competition with radiotracers following stimulant challenges have not been as successful as they have been for dopamine. The options for experimental design are more limited because strong 5-HT—releasing agents such as fenfluramine or MDMA are not suitable for human administration, and studies using human subjects have had to rely on SSRI administration to increase 5-HT levels, which likely leads to more modest effects than compounds that release 5-HT directly. But studies in rodents and nonhuman primates using fenfluramine to release 5-HT failed to consistently demonstrate the sensitivity of most 5-HT_{1A} and 5-HT_{2A} tracers available in the first decade of the 2000s to 5-HT, despite microdialysis evidence that extracellular 5-HT levels rise dramatically in response to the pharmacologic challenge. Several more recently developed tracers demonstrated sensitivity to changes in 5-HT levels in nonhuman primate models. These include two radiotracers for the 5-HT_{1B} receptor, [^{11}C]AZ10419369 and [^{11}C]P943, and [^{11}C]CIMBI-36, an agonist tracer for 5-HT_{2A} receptors. All three have proven sensitive to 5-HT release by fenfluramine with robust ΔBP_{ND} following fenfluramine administration. Results from the small number of studies performed in humans to date have been more mixed, with minimal response to SSRIs. One study did show a robust decrease of [^{11}C]CIMBI-36 binding in response to dAMPH.

The only currently available approach for reducing 5-HT levels to unmask possible baseline binding to receptors is rapid tryptophan depletion (RTD). Tryptophan is the precursor of 5-HT. The only source of tryptophan is dietary, and by having subjects ingest preparations that have high concentrations of other amino acids but not tryptophan, plasma levels decrease rapidly, leading, in turn, to reduced 5-HT in the brain. Behavioral studies have shown that RTD induces symptoms of depression in remitted depression patients. Glucose metabolism as measured with PET and FDG, cerebral blood flow measured with PET and ^{15}O water and neural activity measured with fMRI all suggest a detectable effect of RTD on brain function. Neurochemical imaging with PET has been more equivocal. Studies using [^{11}C]WAY100635 or [^{11}C]DASB failed to detect effects at 5-HT$_{1A}$ receptors or SERT. Two studies using the 5-HT$_{2A}$ tracers [^{11}C]MDL100907 or [^{18}F]Setoperone detected paradoxical effects, in which tracer binding was slightly reduced, rather than increased, in several brain regions following RTD. More recent studies with [^{11}C]P943 and [^{11}C]CIMBI-36 have also failed to definitively detect the effects of RTD on tracer binding.

Serotonergic Target Engagement. PET imaging has been used to establish SERT occupancy necessary for SSRI efficacy. These studies have been remarkably consistent across patient groups, drugs, and radiotracers in demonstrating that therapeutically effective SSRI doses attain 80% or greater occupancy of SERT. Studies have included patients with MDD, panic disorder, obsessive-compulsive disorder, and other conditions and have used both [^{11}C]McN5652 and [^{11}C]DASB as radiotracers. Tested drugs have included citalopram (Celexa), fluoxetine (Prozac), paroxetine (Paxil), sertraline (Zoloft), venlafaxine (Effexor), and others. These findings provide a quantitative dimension to the understanding of SERT blockade as an antidepressive mechanism and a benchmark for drug development and clinical trial design. An interesting related set of studies provide an instructive example of the application of imaging studies in drug development. The long latency to therapeutic effects of SSRIs is believed to be related to a period of reduced 5-HT release following an initial transient increase due to inhibition of serotonergic cell firing in the DRN caused by stimulation of 5-HT$_{1A}$ autoreceptors. The inhibition was thought to subside over the course of several weeks due to the desensitization of 5-HT$_{1A}$ autoreceptors. Thus a strategy for reducing the latency would be to block 5-HT$_{1A}$ receptors in the DRN. But while this might reduce autoinhibition in the DRN, systemic administration of a 5-HT$_{1A}$ antagonist would also block postsynaptic 5-HT$_{1A}$ receptors in the terminal fields, possibly interfering with antidepressant activity. Pindolol (Visken) is a β-adrenergic receptor blocker that is also a weak partial agonist at 5-HT$_{1A}$ receptors, and in the 1990s, some preclinical and postmortem data indicated that it might bind preferentially to DRN autoreceptors, suggesting that augmentation of SSRI treatment with pindolol might decrease the latency and increase the efficacy of SSRIs. Double-blind, placebo-controlled clinical trials showed mixed results, with some reporting beneficial effects and others not. In the early 2000s, several PET studies were performed to measure occupancy of pre- and postsynaptic 5-HT$_{1A}$ receptors by pindolol using [^{11}C]WAY100635 imaging. These studies showed that at the doses of pindolol administered as augmentation therapy in the clinical trials, probably less than 30% of DRN 5-HT_{1A} receptors were occupied, calling into question whether or not the tested doses provided enough blockade of autoreceptors to be effective. In the particular case of pindolol, higher doses could not be tested due to adrenergic effects, but these studies demonstrated the key role that imaging can play in quantifying target engagement by therapeutic agents. Among the several PET studies measuring pindolol occupancy, the most comprehensive one did observe higher occupancy in DRN than in the cortex. Theoretical models have been suggested, but the mechanism for this difference has never been carefully characterized.

NEUROINFLAMMATION IMAGING IN PSYCHIATRY

Imaging of neuroinflammatory markers, initially studied in neurologic and neurodegenerative diseases, has more recently been widely applied in psychiatry as well, motivated by various lines of evidence such as increased expression of inflammatory cytokines in schizophrenia and depression. Microglia are the resident immune cells in the brain, and 18 kDa translocator protein (TSPO) is a mitochondrial protein highly expressed in activated microglia and reactive astrocytes. Many tracers have been developed for TSPO imaging (Table 1.27–2), and this is the neuroinflammatory target that has been most extensively imaged. The protein is not exclusive to activated microglia, but elevated TSPO has still been interpreted as indicative of increased activated microglia. Support for this interpretation has come from studies in humans and nonhuman primates in which administration of the inflammatory agent lipopolysaccharide caused increased TSPO radiotracer binding. The first PET tracer for TSPO was [^{11}C]PK11195, and many early studies used this radioligand. [^{11}C]PK11195, however, has a low signal-to-background ratio, and this motivated the search for better tracers for this target. Many new tracers were developed, all of which were discovered to be sensitive, to varying degrees, to a single nucleotide polymorphism (rs6971). Research participants have been found to have different binding affinities of these second-generation tracers for TSPO according to whether they have two copies of the major allele (high-affinity binders, HABs), are heterozygous for the major and minor alleles (mixed affinity binders, MABs), or homozygous for the minor allele (low-affinity binders, LABs). For some second-generation tracers, like [^{11}C]PBR28, specific binding is so low in LABs that it is virtually undetectable. The strategy adopted by most researchers has been to perform genetic prescreening, exclude LABs from their studies, and treat HAB or MAB group membership as a factor in statistical models. As there is no obvious reference tissue, most investigators report V_T as the outcome measure. Some studies, especially earlier [^{11}C]PK11195 studies, have also reported BP_P or BP_{ND}, derived directly from ratios of the estimated rate constants from the 2TC fits. As described above in the section on kinetic models, these should be interpreted cautiously.

TSPO in schizophrenia and psychosis. Some of the earliest studies using [^{11}C]PK11195 reported higher BP_P or BP_{ND} throughout the cortex in patients with schizophrenia than in matched controls. Many more recent studies with second-generation tracers did not replicate those results. Meta-analyses of studies utilizing second-generation tracers in schizophrenia and other psychotic disorders have found either globally lowered V_T in patients or no difference between patients and controls. Given the better imaging properties of the second-generation tracers and the more reliable outcome measure V_T, the available data do not support the hypothesis of elevated TSPO in psychotic disorders.

TSPO in major depression. Studies in MDD have established a stronger link between TSPO binding and the illness than in schizophrenia. To date, seven studies comparing patients with MDD to matched controls have been performed. Six of these, mostly using second-generation tracers, have found elevated binding in the patients throughout the cortex. Additional findings among the individual studies included associations of TSPO binding with peripheral proinflammatory markers or duration of untreated illness and predictive value of TSPO binding for treatment response to anti-inflammatory drugs.

TSPO in post-traumatic stress disorder (PTSD). Imaging markers of neuroinflammation have not been extensively studied in PTSD yet, but an intriguing result has been reported. Contrary to hypotheses that stress-related disorders are associated with brain neuroimmune activation, patients with PTSD were found to have lower [^{11}C]PBR28 V_T than matched controls in prefrontal-limbic brain regions, with V_T negatively associated with both symptom severity and plasma C-reactive protein levels. These results were also consistent with postmortem data showing lower TSPO expression in females with PTSD than in controls. The study requires replication, but the differences in TSPO imaging results among PTSD, MDD, and psychotic disorders suggest very different involvement of the neuroimmune system in different psychiatric disorders.

Other targets for neuroinflammation imaging. The limitations of TSPO imaging have spurred the quest to develop radiotracers for other targets associated with neuroinflammation. Among these, the ones that have progressed the farthest are cyclooxygenase 1 and 2 (COX-1, COX-2), both of which have roles in inflammatory processes. COX-2 has low basal levels but rapidly increases in response to acute inflammatory conditions and is expressed both in microglia and neurons. The tracer [^{11}C]MC1 is specific for COX-2, but the low basal levels are difficult to detect and quantify. COX-1 is constitutively expressed in all tissues, with predominant brain expression in microglia. The ability to quantify baseline binding may make COX-1 a more desirable target for detecting chronic inflammation in psychiatric populations. [^{11}C]PS13 is a COX-1 tracer that has a heterogeneous binding distribution in the brains of healthy human volunteers, with the highest V_T in the hippocampus, occipital cortex, and sensorimotor cortex, consistent with known mRNA expression from postmortem brains. While still in the early exploratory phase, COX-1 imaging with [^{11}C]PS13 is likely to be tested as a paradigm for probing neuroinflammation in psychiatric populations in the near future.

FUTURE DIRECTIONS

The field of radiotracer imaging with PET and SPECT continues to grow and expand, with the number of peer-reviewed publications using these methods increasing annually. Progress in certain areas, both in the technology and the biologic science, may lead to expanded utility and adaptation of imaging. The manufacture of scanners or cyclotrons that are less expensive or more portable would facilitate even wider use. Some manufacturers have already developed lines of lower-cost PET instruments, for example, by utilizing BGO detector crystals, which have some less optimal properties but are also less expensive than LSO and other state-of-the-art scintillators. At the other extreme, manufacturers are continuing to make major technologic advances in their high-end scanners, leading to finer spatial resolution, higher sensitivity, and better signal-to-noise properties in reconstructed images. It can be expected that these improvements in image quality will translate into more precise results in research imaging as these instruments come into wider use. The most immediate way to expand the scientific questions that can be addressed with PET or SPECT is the development of tracers for more currently inaccessible target molecules or processes. Despite numerous attempts, progress has been slow in developing tracers for the glutamatergic and GABAergic transmitter systems, but clearly, direct observation of the major excitatory and inhibitory neurochemical systems would benefit all brain research. The library of tracers for human imaging of other transmitter systems such as the cholinergic and opioid systems has broadened considerably in recent years, and it can be expected that these will lead to studies providing new insights into psychiatric illnesses. Recent developments in alternative therapeutic targets in schizophrenia and depression will require new imaging paradigms to confirm mechanisms of action. Finally, it is possible that eventually, neurochemical information currently obtained by PET and SPECT radiotracer imaging will be replaced by less invasive methods, for example, by the development of neurochemical probes that can be detected with MRI. But for the foreseeable future, PET and SPECT neuroimaging methods remain essential translational tools for understanding the brain in health and disease and for informing the drug development process.

FURTHER READINGS

Abi-Dargham A, Gil R, Krystal J, et al. Increased striatal dopamine transmission in schizophrenia: confirmation in a second cohort. *Am J Psychiatry*. 1998;155:761–767.

Abi-Dargham A, Mawlawi O, Lombardo I, et al. Prefrontal dopamine D1 receptors and working memory in schizophrenia. *J Neurosci*. 2002;22:3708–3719.

Abi-Dargham A, Xu X, Thompson JL, et al. Increased prefrontal cortical D1 receptors in drug naive patients with schizophrenia: a PET study with [11C]NNC112. *J Psychopharmacol*. 2012;26:794–805.

Baldassarri SR, Park E, Finnema SJ, et al. Inverse changes in raphe and cortical 5-HT1B receptor availability after acute tryptophan depletion in healthy human subjects. *Synapse*. 2020;74(10):e22159.

Bhatt S, Hillmer AT, Girgenti MJ, et al. PTSD is associated with neuroimmune suppression: evidence from PET imaging and postmortem transcriptomic studies. *Nat Commun*. 2020;11(1):2360.

Buchert R, Thiele F, Thomasius R, et al. Ecstasy-induced reduction of the availability of the brain serotonin transporter as revealed by [11C](+)McN5652-PET and the multi-linear reference tissue model: loss of transporters or artifact of tracer kinetic modelling? *J Psychopharmacol*. 2007;21:628–634.

Cherry SR, Sorenson JA, Phelps ME. *Physics in Nuclear Medicine*, 4th ed. Place, Published, Saunders; 2012.

Erritzoe D, Ashok AH, Searle GE, et al. Serotonin release measured in the human brain: a PET study with [(11)C]CIMBI-36 and d-amphetamine challenge. *Neuropsychopharmacology*. 2020;45(5):804–810.

Farde L, Wiesel FA, Halldin C, Sedvall G. Central D2-dopamine receptor occupancy in schizophrenic patients treated with antipsychotic drugs. *Arch Gen Psychiatry*. 1988;45(1):71–76.

Farde L, Nordstrom AL, Wiesel FA, Pauli S, Halldin C, Sedvall G. Positron Emission Tomographic Analysis of Central D1-Dopamine and D2-Dopamine Receptor Occupancy in Patients Treated with Classical Neuroleptics and Clozapine—Relation to Extrapyramidal Side-Effects. *Arch Gen Psychiatry*. 1992;49:538–544.

Gunn RN, Slifstein M, Searle GE, Price JC. Quantitative imaging of protein targets in the human brain with PET. *Phys Med Biol*. 2015;60(22):R363–R411.

Howes OD, Kambeitz J, Kim E, et al. The nature of dopamine dysfunction in schizophrenia and what this means for treatment. *Arch Gen Psychiatry*. 2012;69:776–786.

*Innis RB, Cunningham VJ, Delforge J, et al. Consensus nomenclature for in vivo imaging of reversibly binding radioligands. *J. Cereb. Blood Flow Metab*. 2007;27:1533–1539.

Kegeles L, Abi-Dargham A, Frankle WG, et al. Increased synaptic dopamine in associative regions of the striatum in schizophrenia. *Arch. Gen. Psychiatry*. 2010;67:231–239.

Laruelle M. Imaging dopamine transmission in schizophrenia. A review and meta-analysis. *Q J Nucl Med*. 1998;42:211–221.

Laruelle M. Imaging synaptic neurotransmission with in vivo binding competition techniques: a critical review. *J Cereb Blood Flow Metab*. 2000;20:423–451.

Laruelle M, Slifstein M, Huang Y. Relationships between radiotracer properties and image quality in molecular imaging of the brain with positron emission tomography. *Mol Imaging Biol*. 2003;5:363–375.

Martinez D, Narendran R, Foltin RW, et al. Amphetamine-induced dopamine release: markedly blunted in cocaine dependence and predictive of the choice to self-administer cocaine. *Am J Psychiatry*. 2007;164:622–629.

Martinez D, Slifstein M, Broft A, et al. Imaging human mesolimbic dopamine transmission with positron emission tomography. Part II: amphetamine-induced dopamine release in the functional subdivisions of the striatum. *J. Cereb. Blood Flow Metab*. 2003;23:285–300.

Meyer JH, Cervenka S, Kim MJ, Kreisl WC, Henter ID, Innis RB. Neuroinflammation in psychiatric disorders: PET imaging and promising new targets. *Lancet Psychiatry*. 2020;7(12):1064–1074.

Narendran R, Mason NS, Laymon CM, et al. A comparative evaluation of the dopamine D(2/3) agonist radiotracer [11C](-)-N-propyl-norapomorphine and antagonist [11C]raclopride to measure amphetamine-induced dopamine release in the human striatum. *J Pharmacol Exp Ther*. 2010;333(2):533–539.

Okubo Y, Suhara T, Suzuki K, et al. Decreased prefrontal dopamine D1 receptors in schizophrenia revealed by PET. *Nature*. 1997;385:634–636.

Plavén-Sigray P, Matheson GJ, Collste K, et al. Positron emission tomography studies of the glial cell marker translocator protein in patients with psychosis: a meta-analysis using individual participant data. *Biol Psychiatry*. 2018;84(6):433–442.

Shotbolt P, Tziortzi AC, Searle GE, et al. Within-subject comparison of [C-11]-(+)-PHNO and [C-11]raclopride sensitivity to acute amphetamine challenge in healthy humans. *J Cereb Blood Flow Metab*. 2012;32(1):127–36.

Shrestha S, Hirvonen J, Hines CS, et al. Innis: Serotonin-1A receptors in major depression quantified using PET: controversies, confounds, and recommendations. *Neuroimage*. 2012;59:3243–3251.

Slifstein M, Laruelle M. Models and methods for derivation of *in vivo* neuroreceptor parameters with PET and SPECT reversible radiotracers. *Nucl Med Bio*. 2001;28:595–608.

Slifstein M, van de Giessen E, Van Snellenberg J, et al. Deficits in prefrontal cortical and extrastriatal dopamine release in schizophrenia: a positron emission tomographic functional magnetic resonance imaging study. *JAMA Psychiatry*. 2015;72(4):316–324.

Urban NB, Girgis RR, Talbot PS, et al. Sustained recreational use of ecstasy is associated with altered pre and postsynaptic markers of serotonin transmission in neocortical areas: a PET study with [(1)(1)C]DASB and [(1)(1)C]MDL 100907. *Neuropsychopharmacology*. 2012;37:1465–1473.

van de Giessen E, Weinstein JJ, Cassidy CM, et al. Deficits in striatal dopamine release in cannabis dependence. *Mol Psychiatry*. 2017;22(1):68–75.

Volkow ND, Wang GJ, Fowler JS, et al. Decreased striatal dopaminergic responsiveness in detoxified cocaine-dependent subjects. *Nature*. 1997;386:830–833.

1.28 Functional Brain Connectivity and Psychopathology

DEANNA M. BARCH, PH.D.

Since the advent of modern "biological psychiatry" in the 1950s, much work on understanding the etiology of mental illness has focused on identifying the neural mechanisms that contribute to the development and maintenance of varying forms of psychopathology. This work has been aided by advances in novel neuroimaging technologies, such as the use of SPECT, PET, and MRI for the assessment of structural and functional characteristics of the human brain. The focus of this chapter will be on the use of measures of human brain functional connectivity and the ways in which understanding brain connectivity can inform the general understanding of the mechanisms that lead to the development and maintenance of various forms of psychopathology. The chapter begins with a brief history of the development of methods to measure and understand human functional brain connectivity. Next, the chapter provides a brief overview of the currently available methods for studying functional brain connectivity, with some discussion of their relative pros and cons. An overview of the current knowledge about core "intrinsic" functional brain connectivity networks in humans is also provided. Finally, the chapter selectively describes some of the work using functional brain connectivity to understand the neurobiology of mental illness, with a specific focus on psychosis and depression, as these are two facets of psychopathology with perhaps the largest existing literature on functional connectivity impairments.

In terms of brain function, historically, much of the functional neuroimaging work in psychopathology has focused on the use of either SPECT/PET or fMRI. These approaches provide indirect estimates of neural activity through the measurement of either glucose or oxygen metabolism (PET), blood flow (SPECT), or the measurement of blood oxygen level–dependent (BOLD) activity. The initial studies addressing questions about the neural mechanisms of psychopathology primarily used task-based PET or fMRI to assess differential activation of brain regions while participants performed various cognitive, affective, or social tasks. These studies have been essential for the development and testing of hypotheses about which brain regions may be affected in different manifestations of mental illness, under the inference that abnormal BOLD activity during performance of cognitive or affective tasks indicates abnormal underlying neuronal processes. As some examples, this research has led to the development of theories about the role of the dorsolateral prefrontal cortex (DLPFC) in cognitive dysfunction in psychosis, the role of the amygdala in depression, the role of the amygdala and insula in at least some anxiety disorders, and the role of striatal mechanisms in addictive disorders.

Both the basic and translational research on task-related brain activation has been hugely informative in terms of the current understanding of the neural substrates of particular cognitive and effective processes and how these may go awry in psychopathology. However, over time, it has become clear that rarely does any particular task activate only a single brain region, and rarely is any particular domain of psychopathology associated with abnormal activation in only a single brain region. Further, basic neuroscience research has long made it clear that activity in any individual brain region (or any individual neuron!) is the result of inputs from and outputs to different areas of the brain. Such realizations led to a shift in focus to questions regarding neural circuits rather than a focus only on specific brain regions. More specifically, this shift has been to questions about the relationship *between* and *among* different brain regions in producing successful cognitive and affective function in health and the ways in which abnormalities at the level of circuits contribute to the development and maintenance of particular dimensions of psychopathology. Questions regarding the role of the DLPFC in cognitive dysfunction in schizophrenia provide an excellent case study in this regard. Much of the early work on DLPFC dysfunction associated with cognitive impairment in psychosis focused almost exclusively on the DLPFC itself. However, what if, for example, the abnormal activity observed in the DLPFC during a cognitive control task was actually the result of faulty inputs from another brain region(s)? In other words, what if the DLPFC itself functioned normally in schizophrenia, but the information it received was leading to the observed abnormalities? Alternatively, what if the problem was not in the activity within any specific brain region but rather in the connections among or the ability to coordinate activity among brain regions? The growing work on the role of brain oscillations in coordinating activity among and between neural networks is consistent with such hypotheses that localize impairment in psychopathology at the circuit level of function rather than within specific individual brain regions. Further, in many ways, it is easier to connect circuit-level hypotheses with the larger literature on impairments at other levels of analysis in psychopathology (e.g., neurotransmitter availability, receptors, E-I balance). Such impairments are frequently not regionally specific and involve processes and mechanisms that would reasonably be expected to alter inter-regional communication, either within or between cortical regions or between cortical and subcortical regions.

THE ADVENT OF STUDYING FUNCTIONAL CONNECTIVITY

The ability to examine these types of questions has been propelled by the development of functional connectivity methods for use in humans. Originally, the concept of functional connectivity was applied to simultaneous recordings of neuronal spike trains, which are thought to contribute to the functional connectivity observed in humans using noninvasive neuroimaging methods. A main inference of functional connectivity is that if two regions have highly correlated neuronal activity (i.e., have high functional connectivity), then they are more likely to engage in a common set of processing mechanisms. If so, then functional connectivity provides a tool for understanding which brain regions may be communicating during the completion of cognitive or affective demands and, therefore, what brain circuits support performance and ability in different domains of cognition, emotion, and/or social processing.

RESTING-STATE FUNCTIONAL CONNECTIVITY

A major breakthrough in the development and application of functional connectivity methods for humans came in 1995 when Biswal and colleagues reported that spontaneous activity from regions in the right and left motor cortices were highly correlated even while an individual was resting, suggesting functional connectivity between these brain regions even when the individual was not performing a task. Interestingly, this correlated activity was observed in the spontaneous low-frequency fluctuations of the BOLD signal (0.01 to 0.10 Hz), a frequency band that has often been discarded as noise in task-based studies, though such correlations can be seen at other frequencies as well. This work spurred a major field of exploration of what is now referred to as resting-state functional connectivity. Importantly, this resting brain state activity is thought to consume a major portion of the body's energy (~20%), despite the brain only being 2% of the body's total mass. Further, changes in metabolism due to engagement in a specific task are typically less than 5%, which suggests that ongoing resting-state activity may provide as rich a source, if not richer, of disease-related variability than changes observed due to task performance. In addition, there is also robust work demonstrating that a large portion of the trial-to-trial variability in task-related activity reflects coherent and organized spontaneous fluctuations in brain activity, providing another piece of evidence that this is a meaningful source of variation in human brain function.

RESTING-STATE FUNCTIONAL CONNECTIVITY NETWORKS

One of the key aspects of resting-state functional connectivity that has spurred interest in this aspect of brain function is the realization that *resting-state functional connectivity* reveals intrinsically organized networks of brain regions that are consistently functionally connected, even in the absence of task-induced perturbations in ongoing brain activity. This has been supported by numerous “network” mapping studies that have identified consistent, robust, and reproducible networks of brain regions that show coordinated activity at rest. Among others, these include the default mode network (DMN), the frontal–parietal network, the cingulo-opercular network, and the dorsal attention network, to name but a few. One recent example of this is shown in Figure 1.28–1, where 356 regions (“parcels”) identified as having highly homogenous patterns of resting-state functional connectivity were grouped into 12 resting-state networks that map nicely to networks identified in previous resting-state studies. Such networks are robustly identified across varying methods for assessing functional connectivity, are robustly identified across individuals, though they show important individual differences in the size and location of particular regions, and emerge at least in part early in the course of life as early as birth. Importantly, many of these networks map closely or at least partially to networks that have been identified in task-based studies, providing validation to the functional meaningfulness of this coordinated activity.

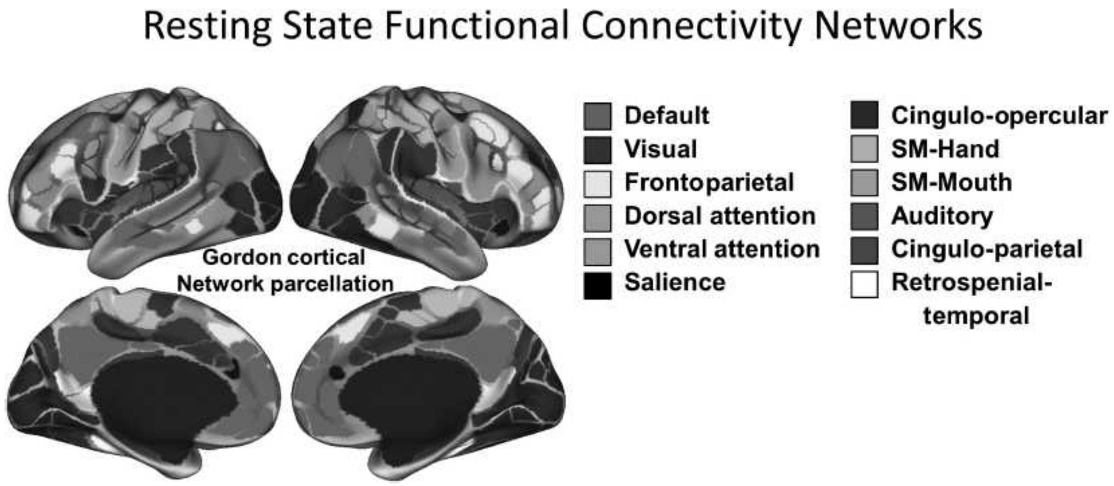
Resting State Functional Connectivity Networks
Gordon cortical
Network parcellation
Default
Visual
Frontoparietal
Dorsal attention
Ventral attention
Salience
Cingulo-opercular
SM-Hand
SM-Mouth
Auditory
Cingulo-parietal
Retrospenial-temporal

FIGURE 1.28–1. Functional connectivity networks derived from resting state data. The **top row** is a lateral view of the brain, and the **bottom row** is a medial view of the brain. The colored "parcels" were derived from a boundary identification approach using a "watershed by flooding" algorithm to identify boundaries in the gradients of change in whole-brain maps of functional connectivity. The community structure of the parcels derived from this boundary mapping was generated using an InfoMap algorithm. (See eBook for color images.) (From Gordon EM, Laumann TO, Adeyemo B, Huckins JF, et al. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb Cortex.* 2016;26(1):288–303, by permission of Oxford University. doi: 0.1093/cercor/bhu239.)

Such functional connectivity networks have their basis in part in structural connectivity but are not isomorphic with regions that show direct structural connectivity. As such, the mapping of resting-state networks to known task networks has led to the hypothesis that resting-state networks reflect, in part, organized interactions that arise partly from a history of coactivation over development, though, as noted above, key components of this organization are present as early as birth. Such networks are now frequently referred to as "intrinsic" connectivity networks, given that they are not dependent on the performance of a particular task for identification, are present even at rest, and are relatively consistent across task and environmental states. Recent major projects to study the human functional and structural connectome across multiple scales, modalities, and ages—the Human Connectome Projects Across the Lifespan—have corroborated the presence of these core networks in the human brain. Moreover, the results of the Human Connectome Project have confirmed and extended current knowledge about the critical relationship between individual variation in functional brain connectivity and individual variation in a range of behaviors relevant to mental health and both negative and positive life outcomes across the lifespan.

Although in principle, any human brain network is relevant to the understanding of psychopathology, several of the most commonly identified networks are particularly relevant to the functions and processes often found to be impaired in psychopathology. For example, both the *frontal–parietal* and the *cingulo-opercular* networks are ones that have been consistently associated with a variety of cognitive control functions. The frontal–parietal network includes dorsal regions of both the lateral prefrontal cortex and parietal cortex, while the cingulo-opercular network includes the dorsal anterior cingulate (dACC), bilateral dorsal anterior insula, and sometimes both thalamic regions and anterior prefrontal regions. As will be discussed in more detail below, these networks have been found to be abnormal in populations with known impairments in cognitive control, such as schizophrenia, or in emotion regulation, such as major depression. The *dorsal* and *ventral attention* networks have also been associated with cognitive function, though more so top-down and/or stimulus-driven attention functions. The dorsal attention network is closely aligned with the frontal–parietal network in that it includes not only both dorsal, frontal, and parietal regions but also more dorsal supplementary motor and eye field areas. The ventral attention system includes the temporal–parietal junction and the ventral lateral prefrontal cortex (VLPFC), and has been associated with attention to salient events in the environment, often activated when events in the world that disrupt ongoing processing occur. Altered connectivity has been associated with anxiety disorders under the hypothesis that it could contribute to altered attention to threat-relevant stimuli in anxiety disorders.

The *DMN* has been associated with a variety of functions, but in particular, it may be associated with attention to internal emotional states and the ability to distinguish or shift between internal and external modes of attention. Further, there is a large body of literature showing that the DMN decreases activity during engaged task states and some hints that a failure to "shut down" the DMN may impair cognitive function. The DMN (see Fig. 1.28–1) includes a wide variety of regions, but most notably medial prefrontal, medial posterior cingulate, and precuneus regions. Further, a number of studies have suggested that there may be multiple DMN subnetworks that may subserve somewhat different functions). Increased connectivity among DMN regions has been associated with rumination in major depression. The *salience* network is one that has been identified somewhat more recently than some of the other networks and includes more rostral regions of the anterior cingulate and insula than typically associated with the cingulo-opercular network (see Fig. 1.28–1), though both have strong connectivity with limbic and subcortical regions. Connectivity of the salience network has been associated with anxiety and arousal and is hypothesized to be a network of regions that serve to process and coordinate reactions to salient events in the environment.

MODALITIES FOR ASSESSING RESTING-STATE FUNCTIONAL CONNECTIVITY

Resting-state functional connectivity in humans can be assessed using a variety of methods. By far, the most common is to use fMRI with a focus on spontaneous fluctuations in BOLD-related activity. This approach is noninvasive and provides good-to-excellent spatial resolution, depending on the pulse sequences used, but it is inherently limited by the fact that BOLD-related signals are indirect measures of neural activity (i.e., reflecting a compensatory blood flow response in reaction to neural activity that alters local magnetic homogeneity) and by the fact that BOLD signals are slow and temporally extended over time. Resting-state functional connectivity can also be measured using EEG or magnetoencephalography (MEG), which can provide much finer-grained temporal resolution and the ability to resolve connectivity at different frequencies, though these methods can be limited in their ability to localize signals (i.e., EEG) or to address deep sources of activity in regions such as the striatum or thalamus (i.e., MEG). Functional connectivity can also be assessed using more invasive techniques, such as electrocorticography (ECoG), which involves the placement of grids of electrodes subdurally. Such an approach has the obvious advantage of direct mapping to the electrical activity generated by neuronal activity along with good spatial localization but is limited to use in populations in need of intracranial recordings for medical reasons.

Further, although much of the focus on functional connectivity has been on data acquired while participants are resting quietly with their eyes closed or open, numerous studies have now demonstrated that very similar results can be derived from data acquired during the performance of specific tasks or even watching movies, though with some caveats. One such approach is to use data acquired during the performance of a task but to try to remove the deterministic influence of the task structure on BOLD activity through regression methods and then base the functional connectivity analyses on the residuals from the regression. Results from such studies can be very similar to results from traditional resting-state studies. However, the degree to which this is true may depend on the number of different task states over which the data is integrated. Doing this for only a single task state is essentially functionally equivalent to the data generated using psychophysiologic interaction analysis approaches (PPI), which focus on comparing connectivity differences in one task state versus another task state. Further, there is some evidence in the literature that prior “states” or performance of specific types of tasks can influence subsequent assessments of resting-state connectivity. As such, integrating over several task states may be more effective at eliminating deterministic or residual effects of any one task state and thus increase generalizability and similar to resting-state connectivity. Another approach that has been increasing in popularity has been to examine functional connectivity while people watch movies, with evidence that this can perform as well as rest data in predicting behavior. Not surprisingly, such an approach is particularly attractive for special populations who may have difficulty resting quietly, such as children.

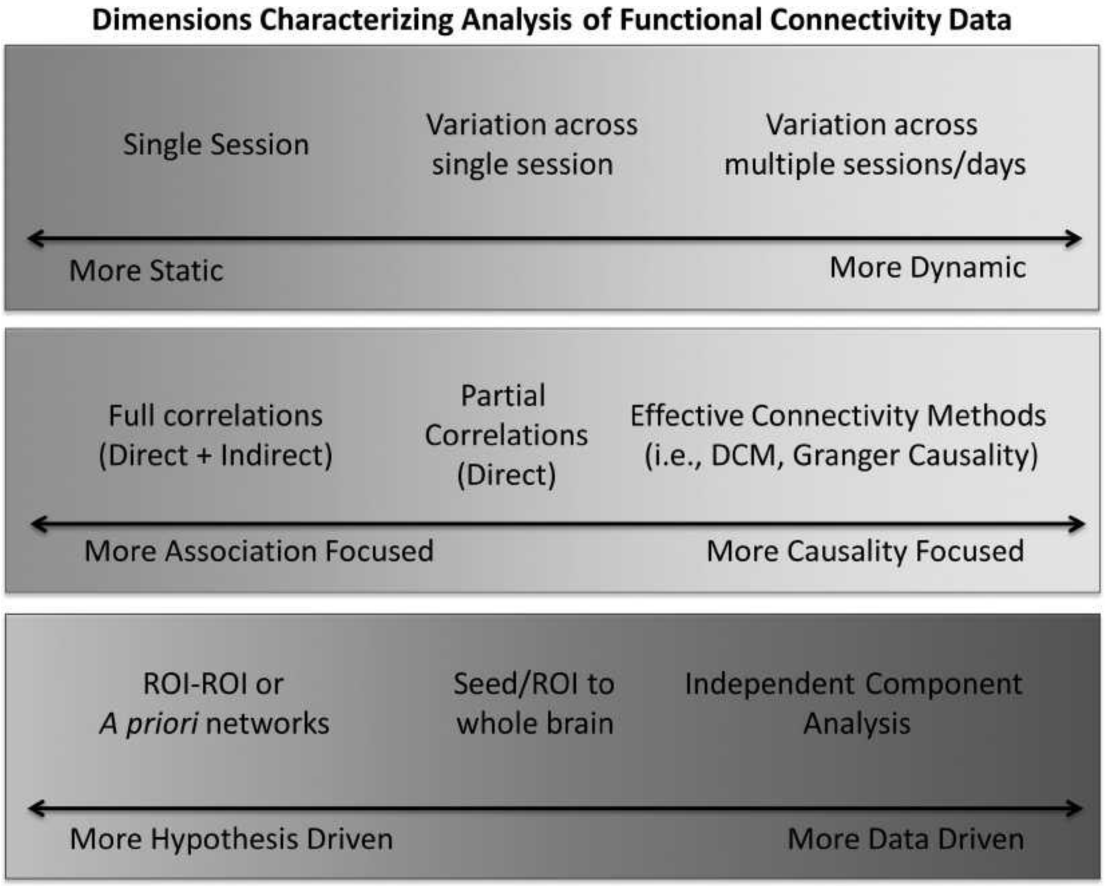
Dimensions Characterizing Analysis of Functional Connectivity Data
Single Session
Variation across single session
Variation across multiple sessions/days
More Static
More Dynamic
Full correlations (Direct + Indirect)
Partial Correlations (Direct)
Effective Connectivity Methods (i.e., DCM, Granger Causality)
More Association Focused
More Causality Focused
ROI-ROI or *A priori* networks
Seed/ROI to whole brain
Independent Component Analysis
More Hypothesis Driven
More Data Driven

FIGURE 1.28–2. Dimensions characterizing functional connectivity analysis. The **bottom row** illustrates the dimension from more hypotheses driven to more data-driven approaches to the analysis of functional connectivity data. The **middle row** illustrates the dimension of analysis approaches more focused on characterizing association versus those attempting to determine causality or directional relationships. The **top row** illustrates the dimension of analysis approaches focused on static estimates of connectivity in a single session versus approaches more focused on the dynamics of change within a session or across sessions.

METHODS FOR ANALYZING RESTING-STATE FUNCTIONAL CONNECTIVITY

As is true for most neuroimaging methods, the use of resting-state functional connectivity data introduces many choice points for the researcher, primarily in terms of how functional connectivity should be calculated, how to identify different brain regions, and how noise should be handled. Each of these choices will influence the structure and quality of the data, but there is no clear consensus about a "best" method for all resting-state studies, only trends of commonly used approaches in the literature. A review of all possible variations or choices relevant to these issues is beyond the scope of this chapter and has been reviewed elsewhere. The focus of this chapter now turns to some of the choices most relevant to understanding the current literature on psychopathology. For most studies of psychopathology, the main goal of resting-state functional connectivity analyses is to assess variations in the relationships between different brain regions, either due to an interest in a specific, a priori region and its connectivity or to better understand disruptions in functional networks. As such, one dimension (see Fig. 1.28–2) along which characteristics analysis approaches is from relatively more a priori and hypothesis driven to relatively more data driven. Along this dimension, the three most common analysis techniques are a region of interest (ROI) or a priori network-based analyses, seed/ROI to whole-brain connectivity analysis, and independent component analysis (ICA) or other clustering methods. All three of these methods have pros and cons, and it is up to the researcher to decide which analysis will allow them to most accurately test their research question at hand.

ROI–ROI or Network-Based Connectivity

ROI to ROI or network-based analyses typically take as their starting point a specific set of regions or a specific network of regions already identified and characterized in the literature and focus on extracting the time series from these regions and analyzing the connections among those specific ROIs or among the regions in that specific network. As an example, one might wish to test the hypothesis that a particular dimension of psychopathology is associated with abnormal connectivity between two specific regions (e.g., medial prefrontal and precuneus regions from the DMN) or with abnormal connectivity among all or some of the regions associated with the DMN. For example, prior work has examined whether connectivity among or between regions in the frontal–parietal network, the cingulo-opercular network, or the default-mode network is altered in schizophrenia and whether it relates to variation in cognitive function. As another example, Perino and colleagues examined whether connectivity among and between the dorsal and ventral attention systems predicts anxiety in youth. Of note, recent research suggests that there may be important individual differences in the size and location of regions within these networks and that it is possible with sufficient data to reliably assess person-specific regions and networks, an approach that may enhance connections to behavior and psychopathology.

ROI/Seed-Based Connectivity

Another common approach is to start with a particular ROI or "seed" and examine connectivity between the seed and the rest of the brain, or at least a subset of the brain. Seed-based connectivity is a flexible method that can be performed, if so desired, but is more commonly used to understand the functional connectivity maps for specific regions of interest. There are numerous examples of this in the literature, such as work looking at the connectivity of the amygdala, or subregions of the amygdala, in relation to depression. Seed regions can also be used to identify functional networks, assuming that the seed region is a robust member of a specific network. However, just because many regions are strongly connected with, say, the DLPFC does not necessarily indicate that those regions form a unitary network. In the case of seed-derived functional networks, it is important to verify that most other nodes are functionally connected to one another as well as to the seed. Taken together, one of the biggest advantages of ROI–ROI, network, or seed-based connectivity methods is that they allow for a direct test of specific research questions, making interpretation of findings in such studies relatively straightforward, modulo all of the potential interpretive caveats that arise in any functional connectivity study.

Importantly, however, this advantage to ROI or seed-focused studies may also contribute to some of the drawbacks of such studies. ROI/seed-based studies allow researchers to make specific predictions about relationships between certain regions or networks of the brain and to test those questions directly. However, by preselecting an anatomical region as the seed for all other correlations, researchers can introduce a bias in their data. There is no perfect way to provide the most accurate and reliable time series for the functional area of interest in all subjects. Consider the possibility that functional areas from two different networks could sit side by side in the brain, such that the inclusion of several voxels from the wrong region could result in connectivity maps that reflect two different functional networks. Add on top of that the consideration that at least some clinical populations may also have anatomical abnormalities that could make the identification of regions challenging. However, as noted above, increasing evidence suggests that it is possible to reliably identify individual-specific brain regions based on their patterns of functional connectivity, which helps to address some of these issues. Further, seed or ROI-focused studies may also suffer from the “streetlamp” problem, such that a researcher is biased to find significant results in a priori regions simply because those are the ones examined, and the researcher does not know whether even stronger or more consistent results might have been generated through the use of a different set of regions.

Another limitation of seed-based connectivity approaches is that they provide a univariate analysis of what is, in reality, a multivariate system. In such ROI or seed-based connectivity analyses, the time series of one or more regions is correlated with the time series of one or more other regions or voxels in an effort to understand pairwise relationships. What is being ignored in this process are the complex interactions that lead to the observed pairwise relationships, such as modulations of the connectivity of that pair of regions or voxels by other connections. Therefore, seed-based connectivity analysis not only restricts the anatomical definition of functional networks, as described above, but interpretations of these functional networks do not take into account the full complexity of the system in which the network is embedded.

Independent Components Analysis

A common alternative to ROI or seed-based connectivity analysis is ICA, which is a multivariate data-driven approach to identifying functional networks of the brain. ICA is an algorithm that takes the entire set of BOLD data and decomposes it into spatially segregated, maximally statistically independent components. In doing so, ICA addresses several of the limitations associated with ROI/seed-based connectivity that were discussed above. Namely, ICA does not require the definition of a priori regions, resolving the potential bias of selecting nonoptimal voxel clusters or of looking only at certain connections. ICA also takes into account the complexity of multivariate data since the components identified by ICA are determined using the entire BOLD data set, which includes information about the time series from all voxels.

Not surprisingly, ICA also introduces several limitations that, similar to seed-based connectivity, have the potential to bias the data. Most notably, ICA identifies components of the data, but it is up to the researcher to determine what each component represents. For example, researchers must decide which components are noise and which components represent functional networks, though strategies exist for helping investigators to accomplish this, such as automated identification of previously validated noise components.

Finally, the interpretation of ICA data is not always as straightforward as the interpretation of seed-based results. Specifically, the metric of ICA is the magnitude of the component, not the magnitude of a specific pairwise relationship. The magnitude of a component can be reduced for a number of different reasons, including variations in spatial distribution, reduced variability in signal, or reduced correlations among time series across regions, and it is not always clear which of these is driving the results.

Association versus Directionality

Another dimension (see Fig. 1.28–2) on which the analysis of resting-state data can vary is the degree to which the methods focus on simply characterizing association versus attempting to identify causal or directional associations. By far, the most common approach to the analysis of resting-state data is to simply use full correlations or ICA, which do not presume any directionality to the associations, and which capture covariance due both to direct connections between regions and to shared variance with other regions. At the opposite end of the spectrum are approaches such as dynamic causal modeling, Granger causality, and lagged time series analyses, which attempt to establish temporal or causative directionality from the patterns of covariance, though there are challenges to the use of this method with resting-state fMRI data. Somewhat in between these two extremes is the use of partial correlations, which attempt to isolate the direct associations of connectivity between regions or voxels, partially out the covariance due to shared connections with other regions.

Static versus Dynamic Assessments

Yet another dimension along which analysis of resting-state data can vary is a focus on a static estimate of a pattern of functional connectivity in an individual at one point in time versus dynamic variations in connectivity “states” over time. A number of studies have now begun to examine how patterns of connectivity among brain regions may vary over time and be classified into a limited set of connectivity “states” or patterns of connectivity. These state patterns may also vary across clinical populations, either in terms of the number of states achieved, the amount of time spent in different states, or the speed with which transitions between states occur.

Network Approaches to Functional Connectivity

As a complement to seed/ROI-based and data-driven approaches, researchers are increasingly applying tools from network science to understand brain organization, primarily through the use of graph theory. When adopting a graph theory approach, investigators have examined variations in several graph properties, including global and local efficiency, small-worldness, and participation coefficients. Global efficiency measures a brain network's functional integration by indexing how easy it is for information to be communicated across brain regions in a network (e.g., across regions in the DMN). Local efficiency measures how well a brain region's neighbors continue to communicate with each other following the removal of that region ("node" in graph theory terminology) and therefore represents something like the fault tolerance of a network. Participation coefficients indicate the number of different networks with which a particular brain region communicates, providing information about how important that region is in overall brain communication. Regions with high participant coefficients are often referred to as "hubs" and are thought to be particularly important for information transfer. The hubs that have been identified in the human brain include the DLPFC, the precuneus, and the insula, among others, and those that are particularly highly connected to each other are referred to as the "rich club." Graph theory approaches have been used to examine brain organization in a number of forms of psychopathology, but perhaps most frequently in schizophrenia, as will be discussed in more detail below.

BENEFITS AND CHALLENGES IN APPLYING RESTING-STATE FUNCTIONAL CONNECTIVITY ANALYSES TO THE STUDY OF PSYCHOPATHOLOGY

Studies of resting-state functional connectivity have not only shaped the way that human brain activity and organizations are understood but also have had important implications for research in clinical populations. One potential advantage of examining resting-state data over task-based data in individuals with psychiatric impairments is that it may not be as complicated by task performance as task-based imaging data, which has proved to be a difficult confound in studies comparing healthy individuals and various patient populations. Primarily, a concern of task-based studies is often whether the differences in observed BOLD activity between controls and patients might be due to differences in the ability to complete the task rather than due to inherent problems using those brain regions since modulation of task difficulty often influences BOLD activity within subjects. This has led many researchers into a challenging situation of designing tasks that are difficult enough for healthy controls to yield variability in BOLD signal and performance but easy enough for patients to learn and perform reliably. In addition, this task performance confound has shaped the recruitment strategies of researchers who tend to focus on higher-functioning patients who can tolerate a cognitive or effective task within the scanner. Because of this, the majority of task-based fMRI studies in schizophrenia may be biased by the use of populations of higher-functioning patients and, therefore, may not always generalize to more symptomatic individuals. Therefore, resting-state functional connectivity potentially represents an opportunity to assess neurobiologic abnormalities in a wider range of individuals with schizophrenia.

However, it is also important to keep in mind that resting-state studies come with their own particular set of challenges when using such data to understand the neural mechanisms of psychopathology. The first and foremost challenge is the pernicious confound introduced by participant movement, which is often greater among clinical populations. Movement can introduce all sorts of artifacts and confounds into resting-state data and, if not properly addressed during processing, can lead to spurious conclusions about the nature and source of connectivity changes in clinical populations. For example, one of the effects of movement is to increase the magnitude of connections among regions that are close together but to decrease the strength of connections among more distant regions, a confound referred to as *distance-dependent movement artifact*. The existence of such artifacts challenges the interpretation of studies claiming to find global reductions in long-range connections coupled with increased local connectivity in clinical populations when movement confounds have not been sufficiently addressed in the data. Another potential challenge is variation in factors such as arousal/sleepiness, respiratory rate, and eyes open versus closed state, all of which can also modulate patterns of resting-state functional connectivity to some extent. Further, there are additional challenges if one wants to use such measures in longitudinal studies, for example, work trying to predict conversion to illness, treatment response, or longitudinal course or change over time. One needs to be able to distinguish actual change from the unreliability of measurement. This latter concern is relevant to any neuroimaging measure, though, not just functional connectivity.

Insights from Resting-State Functional Connectivity Studies in Psychopathology

To date, there are now thousands of resting-state functional connectivity studies that have been conducted in various populations with different aspects of psychopathology. As such, it is well beyond the scope of this chapter to review all of these studies. However, to illustrate some of the insights gained from using resting-state functional connectivity to study psychopathology, this chapter will selectively review studies focused on the two domains with arguably the largest resting-state literature—schizophrenia and major depression. The specific focus will be to identify areas where there is starting to be a convergence of results that have been replicated across studies and which point to clear pathways or directions for future research.

Functional Connectivity Abnormalities in Schizophrenia Spectrum Disorders

To date, there are over 500 empirical studies examining various aspects of functional connectivity in schizophrenia spectrum disorders, with the vast major conducted in individuals with *Diagnostic and Statistical Manual of Mental Disorders–fourth edition* (DSM-IV) or fifth edition (DSM-5)-diagnosed schizophrenia. These studies have used a range of methods and approaches and thus have generated quite a diversity of findings. However, there are areas with increasingly consistent results that point to clear and compelling aspects of altered functional connectivity in individuals with schizophrenia.

Altered Thalamic Connectivity. A strong focus on functional connectivity in schizophrenia is consistent with long-standing hypotheses about schizophrenia as a “disconnection” syndrome, where impairments in cognition and behavior occur because of a failure of coordinated action across multiple brain regions. As many researchers have noted, versions of this hypothesis were put forth as early as the work of Wernicke and Bleuler. Further, theories about abnormalities in connections among brain regions have also played a central role in more modern theories of the pathophysiology of schizophrenia. One such prominent theory, put forth by Andreasen and colleagues, suggested that schizophrenia involves a disruption in the integration of cortical–subcortical–cerebellar circuits, a hypothesis termed “cognitive dysmetria.” Alterations in the structure and function of the thalamus are prominent in schizophrenia literature.

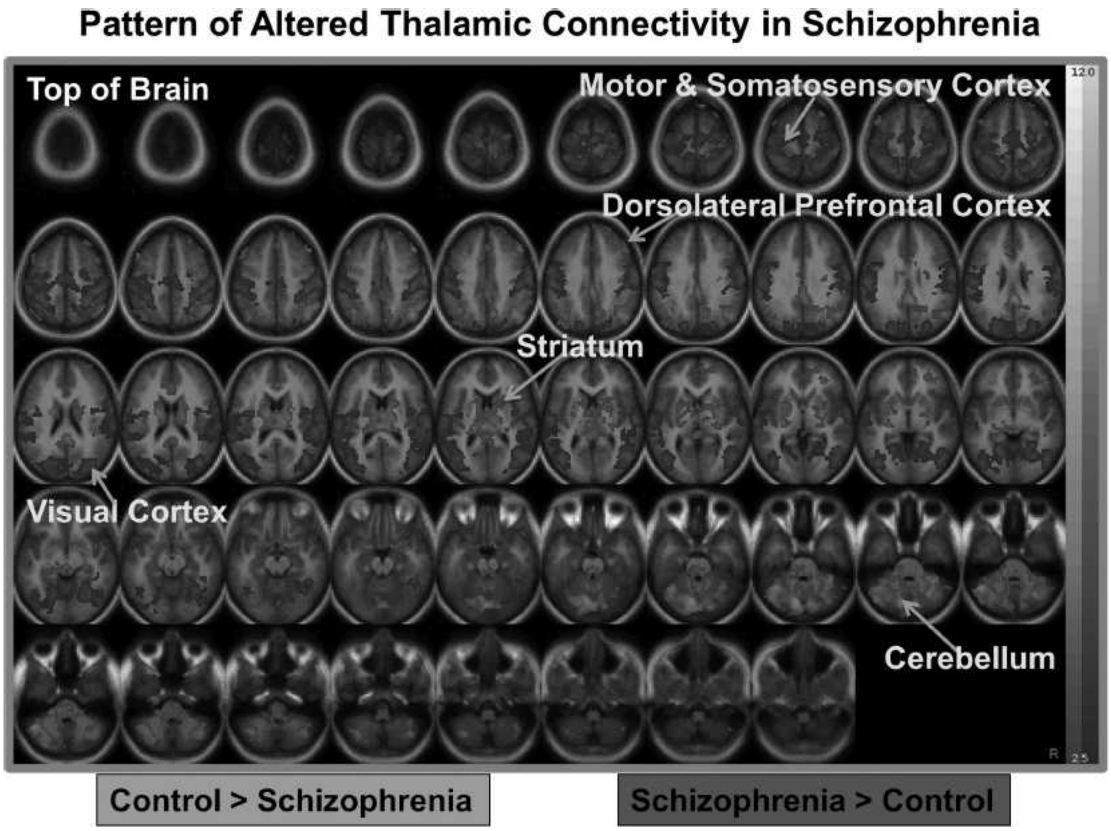
Pattern of Altered Thalamic Connectivity in Schizophrenia
Top of Brain
Motor & Somatosensory Cortex
Dorsolateral Prefrontal Cortex
Striatum
Visual Cortex
Cerebellum
12.0
2.5
R
Control > Schizophrenia
Schizophrenia > Control

FIGURE 1.28–3. Patterns of altered thalamic connectivity in schizophrenia. Unpublished data showing regions (in *red*) where healthy individuals show greater thalamic connectivity than individuals with schizophrenia and regions (in *blue*) where individuals with schizophrenia show greater thalamic connectivity than healthy individuals. The regions showing decreased thalamic connectivity in schizophrenia include the cerebellum, striatum, dorsolateral prefrontal cortex, and the dorsal anterior cingulate. The regions showing increased thalamic connectivity in individuals with schizophrenia include motor and somatosensory cortices, visual cortex, and anterior temporal poles. (See eBook for color images.)

As shown in anatomical studies of primates, the thalamus is topographically organized into parallel pathways connecting specific thalamic nuclei to different regions of cortex brain regions, helping to form parallel loops for various types of information processing between subcortical and cortical regions. In particular, the medial dorsal and anterior nuclei of the thalamus project to the DLPFC. A large and consistent body of evidence has now shown connectivity from bilateral thalamic regions, medial dorsal and anterior nuclei in particular, to bilateral DLPFC, dACC, parts of the striatum, and bilateral cerebellum. In a number of studies, these patterns of decreased connectivity have also been coupled with increased connectivity between the thalamus. These patterns of altered connectivity are illustrated in Figure 1.28–3, which shows regions showing reduced thalamic connectivity in schizophrenia in blue and those showing increased connectivity in red. Like the published literature, individuals with schizophrenia in this sample show reduced thalamic connectivity with bilateral prefrontal and parietal cortices, the ACC, and bilateral striatum and cerebellum. In contrast, they show increased connectivity bilaterally with motor, somatosensory, and visual cortices.

Similar patterns of altered connectivity are seen in individuals at both clinical and genetic high risk for schizophrenia and even to some extent in other disorders characterized by psychosis, such as bipolar disorder. Further, among individuals at high clinical risk, the patterns of altered thalamic connectivity are most pronounced among those who subsequently converted to a full-blown psychosis diagnosis. In addition, studies of structural connectivity are also starting to support altered thalamic (especially medial dorsal) to prefrontal connectivity, including reduced fractional anisotropy of the anterior limb of the internal capsule. Importantly, reduced connectivity between the thalamus and the DLPFC predicts cognitive impairment. These robust and consistent findings across studies of altered thalamic connectivity are fully consistent with the idea that coordination among these brain regions is disrupted in individuals with schizophrenia spectrum disorders. Further, the relationship of this impaired thalamic connectivity with cognitive function is in line with the idea that thalamocortical networks are critical for organizing brain oscillations important for effective cognitive function and that disruptions in such circuits due to impairments at other levels of analysis (e.g., E-I balance) may significantly impair large-scale brain function.

A variety of aspects of the results regarding thalamic connectivity suggest that it needs to be given major weight as a potential biomarker for the risk of psychosis. The first is the fact that such results have been found even among individuals who have never taken medications. The second is the fact that it is present across the course of the disorder and even in those who share the risk of psychosis via either clinical symptom manifestation of genetic risk. As such, the next important step will be to relate thalamic connectivity to other early indicators or risk markers for the development of psychosis. Such markers would include the evidence for increased presynaptic dopamine availability in the striatum, as well as altered glutamate levels in the thalamus.

Altered Default Mode Connectivity. Many researchers have been interested in the idea that individuals with schizophrenia might have alterations in the connectivity of the DMN. As described above, one of the major functions ascribed to the DMN is attention to internal emotional states. A number of researchers have hypothesized that this attention to internal states must be dampened when individuals need to engage in specific tasks or attend to the external world and that this leads to the reduced activity of the DMN typically seen when individuals engage in a focused cognitive task. Thus, the interest in the DMN from the perspective of schizophrenia is the idea that some of the altered perceptual experiences that individuals with schizophrenia might have, or even some of the delusional material, might reflect *enhanced* attention to internal thoughts, feelings, and experiences, due to altered connectivity as well as activity of the DMN. Further, it has been hypothesized that a failure to suppress DMN activity during cognitive task performance may be one of the mechanisms leading to disrupted cognitive function in schizophrenia.

Although such hypotheses are intriguing, the empirical literature on abnormalities in default mode connectivity in schizophrenia as a broad disorder is actually quite mixed, with relatively little consistent evidence for clear disruptions in default mode connectivity. However, one of the challenges in this literature may be that many of the studies have simply focused on comparing individuals with schizophrenia, regardless of symptom profile, to healthy controls. Given the fact that symptom profiles and severity can vary widely across individuals with the same diagnosis and across time in the same individuals, it is critical to also examine default mode connectivity in relation to variations in the severity of specific symptoms. For example, a growing body of work links auditory hallucinations to altered connectivity of the temporal–parietal junction and cortical midline structures that are part of the DMN (anterior and posterior cingulate, precuneus), as well as to disrupted connectivity between default mode and language network regions. The direction of the relationship is such that the presence of auditory verbal hallucinations tends to be associated with reduced connectivity among these default mode structures. This is perhaps not the direction one would have predicted under the hypotheses that auditory hallucinations reflect abnormal attention to internal states due to hyperconnectivity or hyperactivity of default mode regions, but are consistent with the idea that disruptions in networks thought to be important for monitoring internal states are present among individuals with hallucinations.

Graph Theory Analyses of Functional Brain Connectivity in Schizophrenia. Earlier, this chapter described the application of graph theory approaches to understanding functional brain connectivity as a means of summarizing major properties of brain organization. Such approaches have been increasingly applied to study functional (and structural) brain connectivity in schizophrenia. Although the findings have been somewhat mixed and complex, an emerging theme is that individuals with schizophrenia show reductions in both global and local efficiency, particularly when examining specific networks. The networks most frequently associated with reduced global and/or local efficiency tend to involve the anterior cingulate and/or insula (i.e., the cingulo-opercular network) or frontal regions (i.e., the frontal–parietal network). Such findings are consistent with a large body of other work using ROI-based or ICA-based methods that have also identified reduced functional connectivity in or between frontal–parietal and cingulo-opercular networks in schizophrenia. Interestingly, reductions in both the local and global connectivity of the frontal–parietal and cingulo-opercular networks have also been associated with impaired cognition in schizophrenia, suggesting the functional relevance of these network alterations.

There is also intriguing evidence that individuals with schizophrenia show impaired function and connectivity of "hub" regions in the brain, particularly those that are thought to be part of the "rich club" of highly interconnected hubs. More specifically, individuals with schizophrenia show significant reductions in the connections among regions characterized as hubs, including the insula, the superior frontal cortex, the precuneus, and the superior parietal lobule. However, they do not show reductions in the connections between these hub brain regions and nonhub regions (referred to as "feeder" connections) or in the connections among nonhub regions. Importantly, for some regions, such as the anterior insula, individual differences in the degree of hubness, as indexed by participation coefficients, also predict the level of cognitive impairment in schizophrenia. As a piece of converging evidence, numerous studies have also shown reduced gray matter volume in at least some of these hub regions, including the anterior insula.

Summary of Schizophrenia Functional Connectivity Literature. As noted above, there is now a very large literature on functional connectivity in schizophrenia, and some of the findings are mixed, including the results regarding default mode connectivity. However, several findings stand out as consistently replicable. By far, the most robust set of findings is of reduced thalamic connectivity with frontal, cingulate, striatal, and cerebellar regions, a result that is strongly consistent with long-standing theories about schizophrenia as a disconnection syndrome. In addition, there is a growing literature suggesting reduced global and local efficiency in schizophrenia, particularly in frontal–parietal and cingulo-opercular networks, as well as evidence for reduced connectivity of hub regions in this illness, both of which predict cognitive function. It is highly likely that these two sets of findings are related, as the large-scale thalamic disconnectivity seen in schizophrenia may contribute to reductions in the efficiency with which networks can process and transfer information and may also interfere with the ability of hub regions to communicate information across networks. In many cases, these impairments are present at the start of the illness or even before the start and are more prominent among those individuals who go on to develop full-blown psychosis. Thus, these findings converge to suggest that abnormalities in functional connectivity may be a key component of the pathophysiology of psychosis and a promising avenue for future biomarker and intervention work.

Functional Connectivity Abnormalities in Major Depression

Like the literature on schizophrenia, there are now a burgeoning number of empirical studies examining various aspects of functional connectivity in major depression, using a variety of different analytic approaches. Again, like the schizophrenia literature, there is diversity in the findings, which may relate to methodology, stage of illness, and medication factors. However, there are several emerging trends that provide important clues about the ways in which altered functional brain connectivity may be contributing to the development and/or maintenance of mood disorders.

Amygdala Connectivity and Emotion Regulation. The amygdala is a subcortical structure that has received massive research attention because of its role in fear conditioning, threat sensitivity, and generally responding to salient events in the environment. In addition, numerous studies have shown that poverty, stress, and trauma influence the development, function, and structure of the amygdala. Further, a number of investigators have suggested that such alterations in amygdala structure/function are a mediating mechanism linking early adversity to later negative mental health outcomes. Importantly, these findings in humans are consistent with a large animal literature also showing detrimental effects of stress and maternal care on amygdala function and structure, as well as behavior.

The amygdala shows a normative pattern of both positive and negative connections with other regions of the brain. Figure 1.28–4 illustrates the typical pattern of amygdala connectivity found in both children and adults, using a sample of school-aged children and adolescents. As can be seen, the amygdala shows positive resting-state functional connectivity with other regions in the medial and anterior temporal lobes, including the hippocampus, as well as the ventral medial prefrontal cortex (at age 10+ for this region). These regions showing positive connectivity with the amygdala are also important for emotion and stress responsivity. In some studies, the amygdala has been found to couple with other regions in the salience network, including ventral regions of the anterior cingulate (ventral medial prefrontal cortex, as in Fig. 1.28–4) and ventral regions of the insula and anterior temporal poles. In contrast, as shown in Figure 1.28–4, the amygdala also shows negative functional connectivity—anticorrelation—with bilateral dorsal prefrontal and parietal regions, as well as the dorsal anterior cingulate. These regions showing negative correlations with the hippocampus and amygdala are the same regions that are activated in studies of emotion regulation and overlap in large part with regions that are part of the frontal–parietal and cingulo-opercular networks. Importantly, these "anticorrelations" are thought to indicate top-down regulation of emotion and stress responsivity occurring in amygdala regions, suggesting a critical role for the integrity of these connections between the amygdala and frontal–parietal and cingulo-opercular networks in mood and effective function.

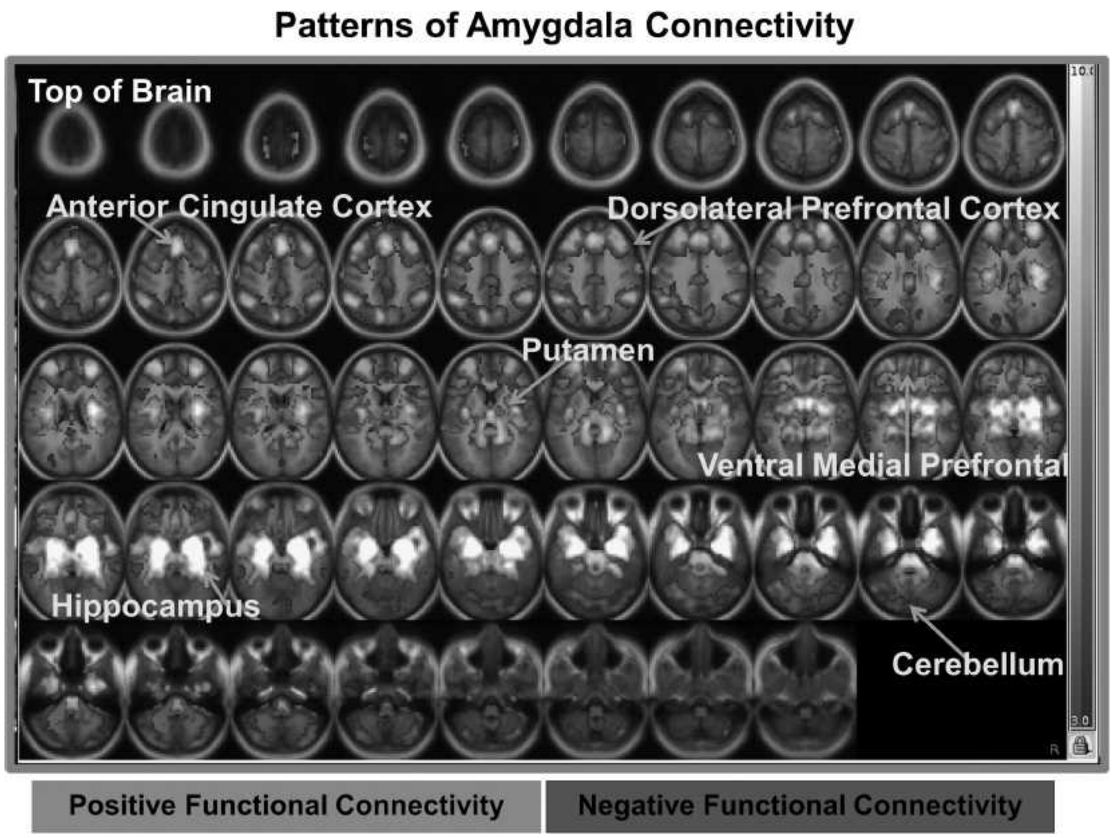
Patterns of Amygdala Connectivity
Top of Brain
Anterior Cingulate Cortex
Dorsolateral Prefrontal Cortex
Putamen
Ventral Medial Prefrontal
Hippocampus
Cerebellum
10.0
3.0
R
Positive Functional Connectivity
Negative Functional Connectivity

FIGURE 1.28–4. Patterns of amygdala connectivity. Unpublished data showing regions (in *light gray*) demonstrating positive functional connectivity with the amygdala and regions (in *dark gray*) showing negative functional connectivity or "anticorrelations" with the amygdala. The regions showing positive functional connectivity with the amygdala include the hippocampus, the ventral medial prefrontal cortex, and the anterior temporal poles. The regions showing negative functional connectivity with the amygdala include the dorsolateral prefrontal cortex.

Importantly, several studies have shown that early life stress and/or maternal deprivation are associated with altered connectivity between the amygdala and ventral medial prefrontal cortex and pregenual anterior cingulate, as well as other regions, connections that are thought to be critical for effective emotion regulation. Further, patterns of altered amygdala connectivity with components of both the salience network and the frontal–parietal and cingulo-opercular networks have also been found in individuals with major depression. The results have been somewhat variable across studies, but one of the more consistent findings has been reduced connectivity between the amygdala and various regions of the prefrontal cortex, including dorsolateral prefrontal, rostral prefrontal, and ventral prefrontal regions, as well as part of the anterior cingulate. As noted above, one hypothesis is that such reduced amygdala to prefrontal/cingulate connectivity reflects impaired top-down regulation of emotion processing and that such reduced connectivity contributes to emotion regulation difficulties in this form of psychopathology. Interestingly, these patterns of altered connectivity have been found across the life span, including among individuals with depression who are children, adolescents, young and middle-aged adults, and older adults, suggesting that altered amygdala to prefrontal/cingulate connectivity may be a life span characteristic of depressive disorders. Further, such connectivity alterations are present even in those who are at risk of the development of depression by virtue of having a mother with depression, suggesting that such connectivity disruptions precede the onset of manifest illness and might even be part of the pathophysiologic pathway. For example, among preschool children with early signs and symptoms of depression, reduced connectivity between the amygdala and medial prefrontal cortex is associated with impaired emotion regulation and increased negative emotionality (see Fig. 1.28–5).

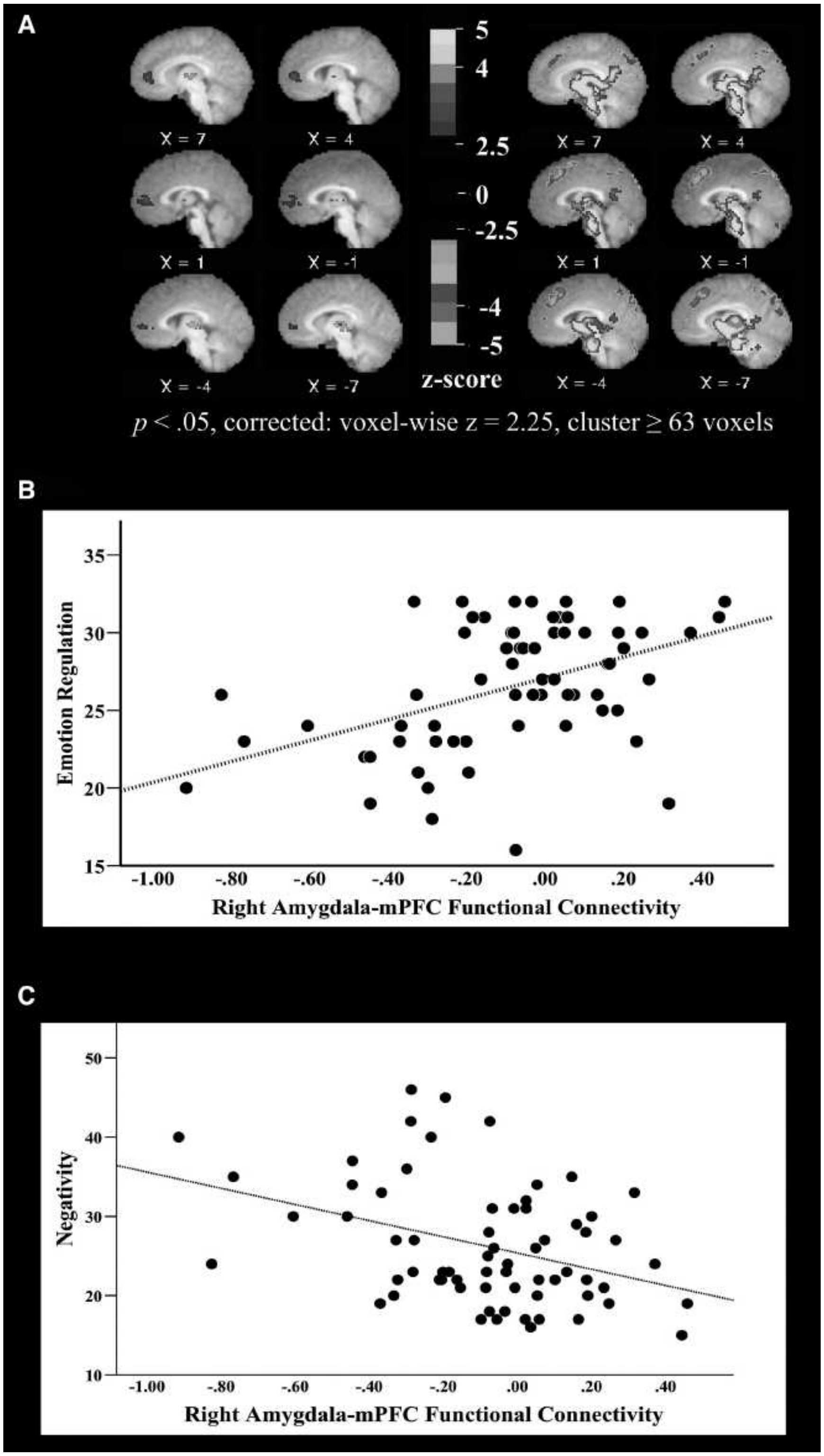

FIGURE 1.28–5. Amygdala to medial prefrontal connectivity in relation to emotion regulation in children. The **A panel** of the figure is an image depicting regions from whole-brain analyses with preschools showing early signs of depression where altered connectivity with the amygdala was related to disrupted emotion regulation. **Panels B** and **C** demonstrate the relationships between reduced amygdala–medial

prefrontal cortex connectivity relate to parent-reported disruptions in emotion regulation and increased negative emotionality. (See eBook for color images.) (Reprinted from Gaffrey MS, Barch DM, Luby JL, Petersen SE. Amygdala-functional connectivity is associated with emotion regulation and amygdala reactivity in 4-6-year-olds. *JAACAP.* 2021;60(1):176–185. ©2021 American Academy of Child and Adolescent Psychiatry. doi: 10.1016/j.jaac.2020.01.024.)

Default Mode Network Connectivity and Rumination. The second area of consistent findings of altered connectivity associated with major depression is in the domain of DMN connectivity. As described above, one of the major functions ascribed to the DMN is to modulate attention to internal emotional states and to govern self-referential processing. Phenomenologically, one of the major characteristics of depression is a preoccupation with negative mood, negative self-evaluation, and frequently a ruminative style of coping with negative events and thoughts. It is not surprising, then, that research points to an *increase* in connectivity among default mode regions in major depression, especially those forming the more anterior part of the DMN, including the dorsomedial prefrontal cortex. Interestingly, work has also shown an increase in connectivity between the dorsomedial prefrontal cortex and the rostral anterior insula, part of the salience network, among individuals with major depression. This increased connectivity between parts of the DMN and the salience network has been interpreted as potentially reflecting an overly strong integration or dominance of DMN activity overactivity related to attention to salient events, potentially contributing to a negative evaluation bias in regard to salient information in the surrounding environment.

A number of studies have also found evidence for increased connectivity of the subgenual prefrontal cortex with the DMN. The subgenual prefrontal cortex is a brain region that has received consistent attention in the literature on mood disorders, as studies have demonstrated that it shows glial cell reduction, altered metabolism, and reduced volume and/or thinning among individuals with MDD, again across the life span. It has been hypothesized that the subgenual prefrontal cortex is important for mechanisms of behavioral withdrawal and effective negative appraisal, with activation of this region associated with conserving resources (i.e., withdrawal), promoting safety-oriented behaviors, and experiencing guilt. Importantly, research has shown that the degree of hyperconnectivity between the subgenual prefrontal cortex and the DMN predicts the level of rumination among individuals with depression. This work has led to the hypothesis that this overly strong integration serves to link the self-referential processing supported by the DMN with the negative affectivity and withdrawal supported by the subgenual prefrontal cortex, leading to a cycle of rumination and impaired emotion regulation among individuals with MDD. As with disrupted amygdala connectivity, such default mode alterations are present even in those who are at risk of the development of depression by virtue of having a family history of depression, again suggesting that such connectivity disruptions precede the onset of manifest illness and might even be part of the pathophysiologic pathway.

Striatal Connectivity and Anhedonia. In recent years, a growing number of studies have been examining functional connectivity correlates of anhedonia both within depression and transdiagnostically. Anhedonia is defined as either a reduction in the ability to experience pleasure or a reduction in the ability to anticipate pleasure. Much of this work has focused on connectivity from regions in the striatum, such as the nucleus accumbens or caudate, with either whole networks (e.g., default mode, cingulo-opercular, salience, etc.) or specific regions in these networks (e.g., the pregenual cingulate of the DMN, the DLPFC in the frontal–parietal to network). This focus on the striatal connectivity has been driven by the evidence suggesting that functional activity and connectivity of the striatum may be important for understanding individual differences in reward responsivity, a construct central to thinking about anhedonia. In individuals with major depression, the findings have been somewhat mixed, but a number of studies have found decreased connectivity of the striatum with other networks, including the default mode, frontal–parietal, and cingulo-opercular networks, to be associated with greater anhedonia.

Summary of Functional Connectivity Literature in Major Depression. As with the literature on functional connectivity in schizophrenia, the literature on connectivity in depression is large and somewhat mixed. However, several findings again stand out as somewhat replicable. The first is evidence of altered connectivity between the amygdala and prefrontal/cingulate regions that are part of the frontal–parietal and cingulo-opercular networks. These connections are thought to be crucial for top-down emotion regulation and control over regions such as the amygdala, which are important for responding to stress and salient environmental events. These connectivity alterations are present in individuals experiencing various forms of early stress or trauma, are present across the life span in individuals with major depression, and are present even among those at risk of depression. As such, that may represent an important clue as to the etiologic pathway linking stress and trauma to subsequent development of emotion dysregulation and mood pathology. The second area of relatively consistent results is evidence for increased connectivity of the DMN, including links between the subgenual prefrontal cortex and other regions of this network. This increased connectivity has been specifically linked to rumination among individuals with major depression and may reflect an overly tight integration between brain regions supporting self-referential processing and those influenced by negative effect processing and withdrawal. A third area is evidence for a link between disrupted striatal connectivity with default–model, frontal–parietal, and cingulo-opercular networks and anhedonia in depression and transdiagnostically, though more work is needed to establish the robustness and replicability of such findings.

CONCLUSION

The goal of this chapter was to provide a brief overview of the current state of the field in terms of methods for assessing the human functional connectome, the existing knowledge about the core intrinsic functional connectivity networks present in the human brain, and the ways in which connectivity abnormalities in these networks are relevant for understanding the pathophysiology of major mental illnesses. Although in many ways, this field is still in its infancy, the remarkable consistency and robustness of the identified functional brain networks across imaging modalities, functional connectivity methods, and populations attest to the importance and relevance of understanding their impact on human behavior. Work from the Human Connectome Projects across the lifespan has confirmed that individual variation in intrinsic network connectivity relates to a wide range of both positive and negative traits and characteristics in humans. Further, this work has shown that core aspects of the brain's functional organization arise early in life, though with developmental evolution across the lifespan. Further, the work selectively reviewed in this chapter illustrates that there are robust, replicable, and differentiable alterations in functional connectivity associated with different forms of psychopathology. This includes robust evidence for altered thalamic connectivity associated with psychosis, impairments in global and local efficiency of frontal–parietal and cingulo-opercular networks in relation to cognitive function in psychosis, and alterations in the hub structure of regions such as the anterior insula in schizophrenia. In major depression, there is consistent evidence for the altered amygdala to prefrontal–cingulate connectivity relevant for understanding emotion regulation, as well as evidence for increased default mode connectivity in relation to rumination. The next key steps, several of which are ongoing, are to examine the role that these connectivity alterations play in the development of psychosis and depression, or specific symptom dimensions, as well as the degree to which connectivity can be modified by treatment and whether it serves as a mediator of symptom improvement. Further, it will be crucial to link alterations at this neural circuit level of analysis with alterations at other levels of analysis, such as genes, gene expression, neurotransmitter signaling, and/or cellular architecture.

FURTHER READINGS

Andreasen NC, Paradiso S, O'Leary DS. "Cognitive dysmetria" as an integrative theory of schizophrenia: a dysfunction in cortical-subcortical-cerebellar circuitry? *Schizophr Bull*. 1998;24(2):203–218.

*Anticevic A, Haut K, Murray JD, et al. Association of thalamic dysconnectivity and conversion to psychosis in youth and young adults at elevated clinical risk. *JAMA Psychiatry*. 2015;72(9):882–891.

Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med*. 1995;34(4):537–541.

Chen JE, Glover GH, Greicius MD, Chang C. Dissociated patterns of anticorrelations with dorsal and ventral default-mode networks at rest. *Hum Brain Mapp*. 2017;38(5):2454–2465.

Cheng W, Rolls ET, Qiu J, et al. Functional connectivity of the human amygdala in health and in depression. *Soc Cogn Affect Neurosci*. 2018;13(6):557–568.

Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE. Intrinsic and task-evoked network architectures of the human brain. *Neuron*. 2014;83(1):238–251.

Dauvermann MR, Whalley HC, Romaniuk L, et al. The application of nonlinear dynamic causal modelling for fMRI in subjects at high genetic risk of schizophrenia. *Neuroimage*. 2013;73:16–29.

Finn ES, Bandettini PA. Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *Neuroimage*. 2021;235:117963.

Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci*. 2007;8(9):700–711.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ. Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*. 1997;6(3):218–229.

Gabbay V, Ely BA, Li Q, et al. Striatum-based circuitry of adolescent depression and anhedonia. *J Am Acad Child Adolesc Psychiatry*. 2013;52(6):628–641.e13.

Geller WN, Liu K, Warren SL, et al. Specificity of anhedonic alterations in resting-state network connectivity and structure: a transdiagnostic approach. *Psychiatry Res Neuroimaging*. 2021;317:111349.

Giraldo-Chica M, Woodward ND. Review of thalamocortical resting-state fMRI studies in schizophrenia. *Schizophr Res*. 2017;180:58–63.

Gong L, He C, Zhang H, Zhang H, Zhang Z, Xie C. Disrupted reward and cognitive control networks contribute to anhedonia in depression. *J Psychiatr Res*. 2018;103:61–68.

Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb Cortex*. 2016;26(1):288–303.

Gordon EM, Laumann TO, Gilmore AW, et al. Precision functional mapping of individual human brains. *Neuron*. 2017;95(4):791–807. e7.

*Hamilton JP, Farmer M, Fogelman P, Gotlib IH. Depressive rumination, the default-mode network, and the dark matter of clinical neuroscience. *Biol Psychiatry*. 2015;78(4):224–230.

He Z, Lu F, Sheng W, et al. Abnormal functional connectivity as neural biological substrate of trait and state characteristics in major depressive disorder. *Prog Neuropsychopharmacol Biol Psychiatry*. 2020;102:109949.

Hyatt CJ, Calhoun VD, Pearlson GD, Assaf M. Specific default mode subnetworks support mentalizing as revealed through opposing network recruitment by social and semantic FMRI tasks. *Hum Brain Mapp*. 2015;36(8):3047–3063.

Li BJ, Friston K, Mody M, Wang HN, Lu HB, Hu DW. A brain network model for depression: from symptom understanding to disease intervention. *CNS Neurosci Ther*. 2018;24(11):1004–1019.

Liu R, Wang Y, Chen X, Zhang Z, Xiao L, Zhou Y. Anhedonia correlates with functional connectivity of the nucleus accumbens subregions in patients with major depressive disorder. *Neuroimage Clin*. 2021;30:102599.

Ma S, Calhoun VD, Phlypo R, Adali T. Dynamic changes of spatial functional network connectivity in healthy individuals and schizophrenia patients using independent vector analysis. *Neuroimage*. 2014;90:196–206.

Marek S, Tervo-Clemmens B, Nielsen AN, et al. “Identifying reproducible individual differences in childhood functional brain networks: an ABCD study.” *Dev Cogn Neurosci*. 2019;40:100706.

*Mulders PC, van Eijndhoven PF, Schene AH, Beckmann CF, Tendolkar I. Resting-state functional connectivity in major depressive disorder: a review. *Neurosci Biobehav Rev*. 2015;56:330–344.

Noble S, Scheinost D, Constable RT. A decade of test-retest reliability of functional connectivity: a systematic review and meta-analysis. *Neuroimage*. 2019;203:116157.

Noble S, Scheinost D, Constable RT. A guide to the measurement and interpretation of fMRI test-retest reliability. *Curr Opin Behav Sci*. 2021;40:27–32.

Palaniyappan L, Deshpande G, Lanka P, et al. Effective connectivity within a triple network brain system discriminates schizophrenia spectrum disorders from psychotic bipolar disorder at the single-subject level. *Schizophr Res*. 2019;214:24–33.

Perino MT, Myers MJ, Wheelock MD, et al. Altered whole-brain resting-state functional connectivity patterns associated with pediatric anxiety and involuntary attention capture. *Biol Psychiatry Glob Open Sci*. 2021;1(3):229–238.

Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage*. 2014;84:320–341.

Rzepa E, McCabe C. Decreased anticipated pleasure correlates with increased salience network resting state functional connectivity in adolescents with depressive symptomatology. *J Psychiatr Res*. 2016;82:40–47.

Rzepa E, McCabe C. Anhedonia and depression severity dissociated by dmPFC resting-state functional connectivity in adolescents. *J Psychopharmacol*. 2018;32(10):1067–1074.

Scheinost D, Tokoglu F, Hampson M, Hoffman R, Constable RT. Data-driven analysis of functional connectivity reveals a potential auditory verbal hallucination network. *Schizophr Bull*. 2019;45(2):415–424.

Sendi MSE, Zendehrouh E, Ellis CA, et al. Aberrant dynamic functional connectivity of default mode network in schizophrenia and links to symptom severity. *Front Neural Circuits*, 2021;15:649417.

Sharma A, Wolf DH, Ciric R, et al. Common dimensional reward deficits across mood and psychotic disorders: a connectome-wide association study. *Am J Psychiatry*. 2017;174(7):657–666.

Sheffield JM, Kandala S, Tamminga CA, et al. Transdiagnostic associations between functional brain network integrity and cognition. *JAMA Psychiatry*. 2017;74(6):605–613.

Smith SM, Miller KL, Salimi-Khorshidi G, et al. Network modelling methods for fMRI. *Neuroimage*. 2011;54(2):875–891.

*Smith SM, Nichols TE, Vidaurre D, et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat Neurosci*. 2015;18(11):1565–1567.

Sylvester CM, Yu Q, Srivastava AB, et al. Individual-specific functional connectivity of the amygdala: a substrate for precision psychiatry. *Proc Natl Acad Sci U S A*. 2020;117(7):3808–3818.

Tang Y, Kong L, Wu F, et al. Decreased functional connectivity between the amygdala and the left ventral prefrontal cortex in treatment-naive patients with major depressive disorder: a resting-state functional magnetic resonance imaging study. *Psychol Med*. 2013;43(9):1921–1927.

van den Heuvel MP, Sporns P, Collin G, et al. Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA Psychiatry*. 2013;70(8):783–792.

VanTieghem MR, Tottenham N. Neurobiological programming of early life stress: functional development of amygdala-prefrontal circuitry and vulnerability for stress-related psychopathology. *Curr Top Behav Neurosci*. 2018;38:117–136.

Yang XH, Tian K, Wang DF, et al. “Anhedonia correlates with abnormal functional connectivity of the superior temporal gyrus and the caudate nucleus in patients with first-episode drug-naive major depressive disorder.” *J Affect Disord*. 2017;218:284–290.

Yeo BTT, Krienen FM, Sepulcre J, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol*. 2011;106(3):1125–1165.

1.29 Neuropsychiatry of Self

David R. Vago, Ph.D., Debra A. Gusnard, M.D., and David A. Silbersweig, M.D.

This chapter is intended to provide an overview of considerations for what may be regarded as a translational science of self and its relevance to psychiatry. Conceptual issues are discussed from both historical and theoretical perspectives to inform our current understanding of the self. Evolutionary, neurobiologic, and clinical perspectives inform the phenomenology and function supporting the act of being a self in the world. A single paradigm for research on “self” does not exist, as it is not necessarily a unitary phenomenon. Diverse disciplines inform the study of the individual (qua self) and its varied representations and relevance uniquely. However, conceptual convergence in this scientific enterprise addressing self “issues” continues to emerge along with signs of where a more advanced program for a science of self may be heading. Implications for a critical interdisciplinary analysis of self on the field of psychiatry and into the nature of consciousness are provided.

INTRODUCTION

Stated simply, self-representation is central to human behavior in health and disease. That is, an individual’s abilities to experience and recognize their physical and mental being as distinctly their own and to act on various kinds of self-knowledge are critical for their being able to organize and regulate their own behavior and engage in social relationships. Until recently, the importance of self-representation and the value of its study for understanding consciousness and clinical health and well-being have been generally underappreciated in scientific circles. But this has begun to change as clinical and psychological disciplines concerned with various disorders bearing on aspects of self-representation have started to collaborate and communicate more directly. Philosophers, religious scholars, computational cognitive scientists, and neuroscientists are improving methods and technologies that can be applied to the explanatory gaps found between objective, observable properties, first-person experience, and biophysical systems embedded in particular contexts. This chapter aims to clarify the various models of the self that have been developed from diverse academic disciplines, with the goal of informing an integrated, comprehensive approach to clinical treatment and neuropsychiatric research. The multifaceted nature of the self, with its profound impact on individual well-being and collective human flourishing, holds an essential place in our understanding of consciousness and evolution of the human race.

Historically, the relative dearth of scientific studies targeting self-representation has been largely a consequence of the fact that unlike most other domains in science where commonly agreed upon and well-elaborated conceptual frameworks guide investigation and constrain the interpretation of data, the scientific investigation of the self has not been grounded by a universal framework and, as a consequence, has been a relatively fragmented enterprise. Yet, there is increasing acceptance that a science of self-representation provides insights critical for understanding a fundamental human nature of consciousness. In this view, forms of self-representation involve operations that support the distinguishing and recognition of an individual's spatiotemporal perspective *as* their own as well as an individual's self-specifying sense of ownership and agency in relation to their own mental and bodily states. Many models of self-representation and conscious experience have been proposed and are now currently contributing to the paradigmatic development of a coherent scientific approach to the expressly multidimensional self.

One general distinction is made between substantialist and nonsubstantialist theories that affirm or deny the existence of a material self, respectively. Substantialist theories see the self as the subject of experience with specific physical (e.g., biologic) properties that support conscious states. Nonsubstantialist theories describe a self as consciousness itself, devoid of any particular material properties. Nonsubstantialist theories that propose a “nonself” stem mostly from Indian Buddhist traditions. These traditions propose any self-identification with material substance (e.g., body, brain) is an illusion and the cause of our mental suffering. One common feature of substantialist theories of self-representation and their constitutive operations is the naturalistic emphasis on neural processes to support the conscious experience of self. Although neurobiologic evidence has been rapidly accumulating, there are also clear reasons not to overemphasize a view limited exclusively to the brain or “brain in isolation” but rather emphasize a theoretical grounding from an interdisciplinary and complex systems perspective instead.

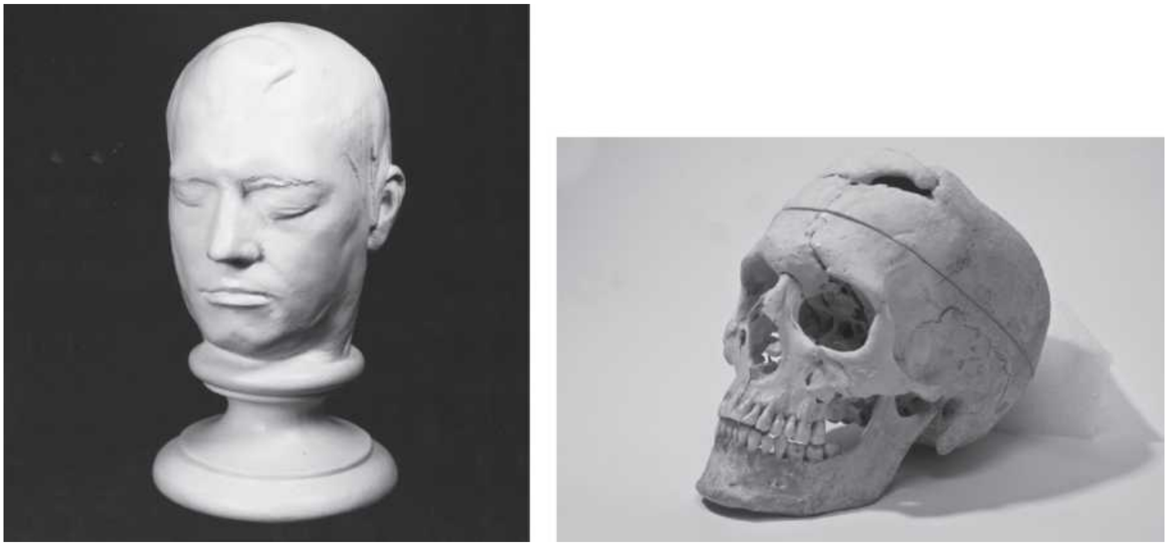

FIGURE 1.29–1. The life mask and skull of Phineas Gage. Note damage to the frontal region. “A famous case illustrating the result of frontal lobe damage involves Phineas Gage, a 25-year-old railroad worker. While he was working with explosives, an accident drove an iron rod through Gage’s head. He survived, but both frontal lobes were severely damaged. After the accident, his behavior changed dramatically. The case was written up by J. M. Harlow, M.D., in 1868, as follows: [Gage] is fitful, irreverent, indulging at times in the grossest profanity (which was not previously his custom), manifesting but little deference for his fellows, impatient of restraint or advice when it conflicts his desires… His mind was radically changed, so decidedly that his friends and acquaintances said he was ‘no longer Gage.’” (Warren Anatomical Museum Collection, Center for the History of Medicine in the Francis A. Countway Library of Medicine, Harvard University.)

Specific historical understandings of the self have evolved in tandem with advancements in scientific methodologies. These developments have enabled an increasingly accepted line of inquiry that links self-representation, functional brain anatomy, and theories of neural networks. For instance, pioneering work in functional anatomy carried out by Franz Joseph Gall and J. C. Spurzheim in the late 18th and early 19th centuries provided early insights into the distribution of mental abilities across different brain areas. This foundational research paved the way for subsequent phrenologic theories advanced by Pierre Flourens and other prominent neurophysiologists. Early brain lesion studies by Paul Broca and Carl Wernicke described a variety of aphasias that led to connectionist models of localized function. The case of Phineas Gage, a railroad worker suffering from orbitofrontal cortex (OFC) damage (see Fig. 1.29–1), provided profound insight into our understanding of frontal lobe function in emotional expression and regulation. Equally impactful was the case of H. M., who underwent the removal of a significant portion of his hippocampus and adjacent tissue to address intractable seizures. This case became a landmark in shedding light on the specific brain structures responsible for memory functions. Earlier, Karl Lashley published *Brain Mechanisms and Intelligence* (1929) and argued a theory of localization focused on a principle of equipotentiality. Lashley reasoned that complex functions "are largely carried out in independence of structural differentiation." Split-brain studies performed by Roger Sperry and colleagues in the 1960s on individuals who had undergone separation of their two cerebral hemispheres by corpus callosotomy for intractable epilepsy revealed significant disconnection of the perceptions and decisions of one hemisphere from the other. These findings, among other functional neurologic case studies, contributed to deflating the notion of a necessary unity and coherence in self-representation and the recognition that the normal experience of such unity masks underlying modular neural organizational processes that may be subject to disruption and scientific manipulation.

Recent transdisciplinary collaborations have stimulated new thinking and inherently different, though still functionalist, perspectives on considerations of self-representation. For example, cognitive and computational neuroscientists, neuropsychiatrists, and cognitive behavioral neurologists sometimes work with philosophers of mind and religious scholars to refine and clarify concepts enhancing the design and interpretation of experimental studies. Many contemporary models posit that the self exists beyond some individual entity and elaborate upon the phenomenology and conscious experience of being a self, the relation of self and others, the self-organization of living things, and the naturalistic substrates in the brain and body that support an existence of the self situated in the world. The perspectival perspective emphasizes the viewpoint-dependent nature of our self-concept and identity. This perspective holds that the way we understand and experience our self is inherently shaped by various contexts, such as cultural, social, and even moment-to-moment situational factors. Other theories emphasize an embodied perspective on our self-representation, which is constructed through how we move through space, including our physical interactions, and even biologic states like hunger or fatigue. One model by Christoff and colleagues (2011) proposes that the reafferent–efferent processes of sensorimotor integration and homeostatic regulation implement a self-specific, agentive perspective at the bodily level of perception and feeling without the need for conceptual processing. Here, a distinction is made between self-specific and self-related representations of self. Processes that are self-specific are intrinsically related to the agent’s own action without an explicit self-concept (e.g., the self is specified through interaction with the world) whereas self-related processes involve evaluation of some feature of oneself in relation to one’s perceptual image or mental concept one’s self-identity. The first-person experience of what it is like to be a self in the world is also described by Francisco Varela and other phenomenologists as “direct knowledge” supported by a science of mind embodiment that incorporates the experience of being there. Philosophers such as Martin Heidegger, Edmund Husserl, and Maurice Merleau-Ponty established phenomenology as a philosophical tradition and analytical account of the self through direct experience that does not exist in separation from the stream of consciousness and is nonconceptual and prereflective in nature. A number of representational systems and theories from this perspective have been proposed. Thomas Metzinger (2010) attempts to provide a synthesis of these perspectives through his phenomenal self-model (PSM). The Projective Consciousness Model proposed by David Rudrauf, Karl Friston, and colleagues (2017) builds upon a perspectival phenomenological structure supporting consciousness with a free energy minimization model of active inference describing the modulation of cognitive and affective dynamics to control the embodied agentic self. A global workspace model was proposed by Bernard Baars (2005) that attempts to account for both phenomenal and access forms of consciousness. Metzinger’s PSM theory of subjectivity, Karl Friston’s free energy principle theory, and Anil Seth’s utilization of hierarchical Bayesian inference have all contributed to philosophical and computational understandings and ontologic claims about the nature of the self. Scientific inquiry into the nature of self-representation is clearly broadening across levels of description (from psychological and behavioral levels of description to computational, neurophysiologic, and systemic levels of description) and in terms of relevant capacities that change across human development and may exist in other animal species as well.

HISTORICAL CONSIDERATIONS

The Hindu Vedic and Buddhist texts dating back between 500 to 1000 BCE all describe insight into the nature of self as the ultimate form of salvation. The Hindu scripture, the Bhagavad Gita, written around 500 BCE, says, “True knowledge of the self does not lead to salvation: it is salvation.” The founder of the Sōtō school of Zen, Dōgen Kigen, claimed, “To study the Way is to study the self. To study the self is to forget the self. To forget the self is to be enlightened by all things. To be enlightened by all things is to remove the barriers between one’s self and others.” Vedic descriptions of the self and the soteriologic aspects of self-reflection describe an immutable self (Sanskrit, ātman) akin to the soul, while the Buddhist descriptions, similar to Dōgen, further elaborate upon the nondual nature of the self, its impermanency, and how mental suffering manifests through attachment to the illusory concept of self, self-identity, and the distorted interpretation of the world through the lens of mental habits and social conditioning.

Although extensive work had been done through the soteriologic lens of Hindu philosophy and Buddhist spiritual teachings, socially accepted intellectual interest in the self has clearly varied across time due to societal and cultural forces. For example, in the West, before the decline of the Middle Ages and the attendant decline of religion as the dominant organizational culture, emphasis was placed on the concept of community rather than on self in all aspects of social life. The significance of the community tended to overshadow the significance of individuals and their personal interests, which were viewed by the dominant culture as arising from a human nature that was inherently biased and “selfish.” Ideally, self-interest was to be suppressed in the interests of the common good and spiritual attainments. Major shifts in social thought began to occur with the dawning of the Renaissance in the 14th century with its greater focus on the personal and temporal, which was then followed by the Enlightenment period in the 18th century with its emphasis on rationality, systematized thinking, and science. During these periods, the individual, personal identity, and personal expression became increasingly valued. Descartes (1637) proposed that the self is not identical to one's body or, indeed, to any physical thing. Instead, he famously concluded that the essential self —that claims, "I exist"—is a nonphysical, conscious thing—creating a Cartesian duality of immaterial and material. "Cogito, ergo sum; I think therefore I am." Technologic developments also played a role. Interestingly, it has been suggested that one of the more significant technologic developments in this regard was the rise of the glass industry in Western Europe and the establishment of mirrors as a common commodity. The widespread availability of mirrors provided individuals for the first time in human history with a stable and accurate representation of their own faces, which allowed them to regard themselves in a different way and encouraged greater self-inspection, including more frequent introspection. Since then, political and economic changes, as well as the rise of capitalist societies, have promoted a veritable culture of individualism, particularly in North America and Western Europe. Such forces have all contributed to a modern intellectual climate that has made the concept of the individual and the nature of the self and subjectivity available for consideration and conceptual elaboration as never before through a scientific lens.

Conceptualizations of self have been far from homogeneous as discussions in the modern era have unfolded, however. One of the fundamental questions targeted in the intellectual debate has, in fact, been whether there is such a thing as the self, adding support to nonsubstantialist, nonself theories. David Hume, in his *Treatise on Human Nature* in 1739, notoriously stated that, when he looked inside himself, he could find many perceptions but nothing linking them together. Introspection, Hume concluded, reveals only a continuously changing flux of sensory, emotional, cognitive, and other bodily processes. Since Hume's denial of the self, other influential philosophers, including Friedrich Nietzsche in the German tradition and Ludwig Wittgenstein in the analytical tradition, have also argued that the self is most reasonably understood to be a fiction that may have some utility for those using the term but that it has no actual substance or basis in reality, that "the self [is simply] a center of narrative gravity," as Daniel Dennett, the American philosopher has famously put it. The French novelist and philosopher Jean-Paul Sartre summarizes this incommensurability by stating we are "condemned" to believe in the self.

For some scientists, the very term “self” has been associated with this problematic notion of an entity within the brain controlling brain processes and behavior. For example, the term “homunculus” is used to describe maps in the brain that correspond to sensory input and motor output for different parts of the body. These are often portrayed as distorted human figures, where the size of each body part is proportional to the amount of brain area. Yet, the presence of a homunculus does not explain the subjective nature of experience and inadvertently supports Cartesian dualism—the idea that the mind and the body are separate entities. Many scientists have focused on the processes supporting an individual organism’s subjectivity and sense of personal-state ownership via consideration of biologic and evolutionary pressures. Neuroscientists Rodolfo Llinás and Walter Freeman, for example, have both used the term “self” to refer to that locus of properties that permit a mobile organism to sustain its viability *as* an individual organism. Autopoiesis, or “self-production,” is a concept introduced in the 1970s by the biologists Humberto Maturana and Francisco Varela to describe a sytem capable of reproducing and maintaining itself and distinguishing it from nonself. Varela suggested that the human organism should be considered an autopoietic entity where mental processes are intrinsically linked to bodily ones. For example, emotions are not just abstract experiences but have physiologic components that influence the entire organism. Similarly, the way we perceive and engage with the world is informed by our embodied and enactive experiences, which in turn are shaped by our cognitive and emotional states. This creates a circular or recursive relationship between the body and mind, typical of autopoietic systems. Niklas Luhmann’s autopoietic sociologic theory (1995) further distinguishes between society and functionally differentiated social systems that are interconnected through complex communication, suggesting group dynamics may contribute to self-identity at individual and collective levels. Though the details of their formulations differ, many theories consider the self from a comparative perspective and regard the primary *need* for a “self” to be related to action or mobile organisms’ capacities for goal-directed behavior in an environment. That is, in contrast to organisms incapable of active movement, mobile organisms need to be able to predict the effects of their actions on their environment. Otherwise, the outcomes of their actions would be purposeless and potentially dangerous. They need to be able to “own” their behavior (though this need not be conscious) and distinguish their behavior’s effects from effects attributable to other forces in their environment.

Such basic biologic processes have also been distinguished from more complex cognitive processes that help elicit a self-narrative over time. Evolutionary theory and the field of paleoneurology have described the evolution of biologic processes to support the demands of survival for Homo sapiens. For example, specific basic and core architecture has been described to evolve through layers of complexity (e.g., proto- and core-self) with unique regulatory neurophysiology. With the advent of complex cognition in human evolution, pressures for reasoning, problem solving, and learning influenced the development of the frontal lobe and capacity for reflection. Theoretical neuroscientists like Endel Tulving have attempted to understand consciousness through the perspective of conscious self-awareness and memory. For example, Tulving (1983) describes his theoretical position of self-knowing as autonoetic consciousness. Tulving defines this awareness of episodic events as a veridical part of one’s own past existence. Karl Friston and Patricia Churchland extend and elaborate on the idea of the self-concept as something fundamentally rooted in memory, or the “mnemonic concept.” In their framework, the self is not just a static concept or identity but a dynamic, self-organizing biologic system. This system functions as a complex array of organizational tools that serve to bring coherence to the various operations of the brain, including planning, decision making, and perception. In essence, they propose that the self acts as a kind of internal coordinator, working to align different cognitive processes into a unified, coherent experience. This approach sees the brain as an adaptive system constantly seeking to reduce uncertainty and inconsistency. The self, in this model, helps to navigate this uncertainty by “coherencing” or synchronizing various neural and cognitive processes. Bridging the gap between biologic systems and first-person phenomenology, Francisco Varela (1971) advocated for a neurophenomenological perspective that accounts for an emergent, multilayered self, arising from the dynamic interactions between brain, body, and environment and not merely through neural computations within the brain. This is in line with his broader framework of enaction, which postulates that organisms enact or bring forth their worlds through their actions and perceptions. Within these opportunities for investigation at the first-, second-, and third-person perspectives of observation, cognitive science, neuroscience, and the cross-disciplinary field of neuropsychiatry are integrating the neurophysiologic substrates supporting the range of self-representation with the evolutionary, therapeutic, and transformative value of contemplative traditions.

Developmental approaches to healthy, age-appropriate, psychological functioning contribute greatly to our understanding of the genetic, sensory–motor, and behavioral evolution of the self as a function of individuals' embeddedness in social structures. Current conceptions of self and its role in human culture, society, developmental, and clinical health are very relevant to an understanding of the progressive evolution of the human mind and partial fulfillment of its fate as a uniquely self-aware and self-conscious species (Homo sapiens).

MULTIDIMENSIONALITY OF SELF-REPRESENTATION

In scientific literature, over 25 notions or aspects of the self have been specified, including the cognitive self, conceptual self, core self, minimal self, ecologic self, embodied self, emergent self, the material self, social self, interpersonal self, and others. William James made the distinction between self-as-subject (the I-self) and self-as-object (the Me-self) as two separable but intimately interrelated aspects of the self. He also identified subcomponents in which the self-as-subject included self-agency (the sense of ownership of one's thoughts and actions), self-awareness (an appreciation of one's internal states, needs, thoughts, and emotions), self-continuity (the sense that one remains the same person over time), and self-coherence (a sense that one is a single, coherent, bounded entity). Components of his self-as-object included what he referred to as the "material me," the "social me," and the "spiritual me," which were conceptualized as aggregates of various kinds of knowledge about the self that one acquired by adopting an observer's perspective, the details of which also had the potential for coming into conflict with each other.

While details of the Jamesian framework have been challenged, his general distinctions have remained a viable heuristic for much subsequent research in this area. Until recently, more empirical attention had been devoted to the latter, the self as an object of one's knowledge and evaluation, because of the relative ease with which it and its derivatives (e.g., individuals' "self-concepts" and "self-esteem") are able to be queried and articulated by investigators and study participants. The more implicit and thereby elusive cognitive processes that define the individual as a subject have only recently gained increasing emphasis in theories and studies of self-development. This has happened largely because of advances in interdisciplinary collaboration, experimental paradigms, methods, and technology, facilitating insight into neurophysiologic and psychosocial mechanisms supporting the complex ways individuals think and act.

Shaun Gallagher (2013), for example, proposed a pattern theory of self to organize these dimensions further through situated, extended, narrative, psychological, intersubjective, minimal experiential, and minimal embodied aspects of self-processing. In those areas of scientific discourse, clinical health, and philosophy where progress has been made, it has become increasingly clear that a single discrete referent for the concept “self” does not exist, nor does a unitary, fixed, or independent self within the world of phenomenology. There is also no property of the concept “self” that is present in all instances of “self” processing and absent in all instances of “nonself” or “other” processing. Properties or characteristics by which “self” is sometimes defined (e.g., being a particular kind of mental phenomenon or having a particular kind of function) do not apply to all instances in which the concept is used, and none is unique to it.

Work that has been undertaken from comparative evolutionary and human developmental perspectives is particularly useful in illustrating this with regard to notions of the changing self. Importantly, both of these perspectives are grounded in considerations of the biology and adaptive functioning of organisms. Across the animal kingdom and in human ontogeny, self-representational processes that are operationalized in the nervous system clearly vary in their degrees of complexity and in the self-representational phenomena that they support. Yet it is also the case that, from these evolutionary and developmental perspectives, “self” may be regarded as the set of mechanisms or means by which individuals (and, more specifically, their nervous systems) organize their perceptions, plans, and decisions, allowing them to act as a coherent, embodied whole rather than as a group of independently operating systems with competing interests. It is, in fact, when such coherence significantly breaks down that self-pathology is thought to exist, and affected individuals (if they are capable of sufficient insight) may actually report that aspects of their experience or behavior seem “foreign” to them or unmanageable by them.

Comparative Evolutionary and Paleoneurologic Perspectives

Scientists who study various animal species are among those that have been interested in the question of how organisms represent themselves to themselves. Because such scientists consider the conservation or loss of certain bodily structures and capacities across species (in this case, conservation or loss of organisms' capacities to recognize aspects of themselves as being or belonging to themselves), they are often able to offer physiologic (proximate) and evolutionary (ultimate) explanations for "why" such capacities exist.

In contrast to organisms of greater complexity, simple vertebrates are observed to be very limited in their self-representational capacities. In these simple animals, such capacities are confined to mechanisms that coordinate visceral and other internal signals with perceptions of the outside world that allow them to distinguish the boundaries of themselves or the effects of their own actions from other phenomena in the environment and generate responses that enable their survival (e.g., an appendage withdrawal when a noxious stimulus is perceived to have impinged upon it). At this simple level, the coordination of responses is automatic and implicit. There is no conscious awareness of a "self" who is the subject of experience or the agent of the action.

The centralization of such coordinated, even if limited, processing is the neurobiologic foundation upon which higher levels of self-representation rest. One widely recognized view on the interdependent relationships between this implicitly coordinated reflexive processing that subserves basic survival functions of the individual organism and the computationally more sophisticated reflective processing that is associated with self-awareness has been provided by Antonio Damasio and Jaak Panksepp. While both scholars agree on the importance of emotion in constituting a foundational, primitive aspect of the self, Panksepp's work focuses on the primal, universal aspects of emotion that are conserved across species as the basis for more complex, secondary emotions found in humans more uniquely. Damasio, on the other hand, tends to focus on the complex interplay between emotion and cognition, especially as it pertains to human consciousness and decision making. In Damasio's model, the "proto-self" is the most basic level of self-representation which consists of a collection of distributed neural patterns that preconsciously map the body's internal state and serves as the foundation for more complex forms of self, such as the "core self" and "autobiographical self." Panksepp does not use the term, “proto-self,” but describes primal emotions that are hardwired into subcortical regions of the brain that interact with sensory-motor input and output, influencing behavior and physiology. The multilayered models of the self proposed by Panksepp and Damasio offer a nuanced understanding of the self that integrates biologic, psychological, and social dimensions that map complex aspects of self that have developed in humans layered upon more basic nonconscious processes that arise from interactions of the body with the environment.

The paleoneurology literature also suggests points of human evolution in which anatomical changes to the brain coincide with the emergence of complex social phenomena like language development, tool use, and cognitive skill development. Brain changes and associated functional advances are implied through morphologic changes found in skull casings from early hominid ancestors. Aside from current human cranial capacity (~1,350 to 1,400 cm^3) being three times the size of our earliest hominid ancestors (~450 cm^3 for australopithecines), the anterior frontal cortex has been found to be twice greater in relative size than other parts of the frontal lobe in our closest ancestors, chimpanzees and bonobos. As described later, the anterior or frontal polar cortex (FPC, Brodmann area 10) is uniquely involved in higher-order, complex cognitive processing (e.g., prospective memory, meta-cognition) and is uniquely positioned to facilitate more complex self-reflective processes. Aspects of the frontal pole’s organization vary slightly across hominid species, including the relative width of its cortical layers and the space available for connections. BA10 is larger in humans than in apes, both in absolute terms and relative to the rest of the brain. Horizontal spacing distance between neurons has been observed to be 30% larger in the human frontal pole than in other species, suggesting more opportunities for complex dendritic branching and interconnectivity across frontal structures. Other perspectives from comparative biology have questioned the social brain hypotheses, suggesting ecologic factors, rather than social complexity, best predict brain size across the evolution of primate species. DeCasien and colleagues (2022) provide modern insights from comparative biology and emphasize that a proportionally large human prefrontal cortex may not represent an adaptive specialization; rather, its role in integrating with wider distributed neural networks may be more relevant for understanding the unique evolution of the self.

Human Developmental Perspective

Developmental approaches to healthy, age-appropriate, psychological functioning contribute greatly to our understanding of the genetic, sensory–motor, and behavioral evolution of the self as a function of individuals' embeddedness in social structures. Researchers in child and adolescent development describe human beings not as unitary but as integrated multidimensional beings whose capacities underlying self-identity emerge and mature over time and in relation to others. This social-developmental perspective contends that our sense of self is not formed in isolation but at least partially through our social roles and how others see us. Concepts like Charles Cooley's "looking-glass self" fall under this category, emphasizing the role of social validation and feedback in the formation of self-identity. The socially constructed aspects of the self are dependent upon one's social roles in societal systems and other aspects of social identity.

Some developmental theorists have invoked William James' framework for considering the complexity of self-representation that fluctuates between increasingly differentiated and integrated perspectives emerging in childhood through adolescence. Self-as-subject and Self-as-object are two general domains of self-representation proposed by James and are considered crucial for healthy development. Specifically, the structure and content of *what* individuals say they think about themselves or how they act in the world at any developmental stage depend on the acquisition of specific cognitive and self-specifying abilities. These learned persepctives influence how individuals are able to implicitly demonstrate and explicitly report on what they reflectively know about themselves across stages of development. For example, one piece of evidence concerning the development of I-self, or self-as-subject, is that infants show relatively long fixation durations between 0 and 2 months of age, whereas they show relatively short fixation durations from 2 to 6 months of age. This may reflect a genetically programmed sensitive period in which the infant automatically (rather than effortfully) learns how to fixate on and differentiate selected objects first (e.g., caregiver's face); followed by a period in which they learn to shift their attention to other objects nearby in time and space and thereby begin linking sets of experiences together. The me-self, or self-as-object, begins to take over more considerably by age 2 through the development of representational constructs. Young children between the ages of 2 and 4 are able to construct separate attributes of themselves that may be physical (e.g., “I have blonde hair”), social (e.g., “I have three brothers”), or psychological (e.g., “I am happy”). However, they are unable to coordinate two such constructs into a coherent self-portrait, in part because of working memory limitations at that age that prevent them from holding several features in mind simultaneously. They are also unable to acknowledge the possibility that they (or others) can possess opposing attributes (e.g., good and bad) or different emotions of the same valence, such as mad and sad, or opposite valence, such as happy and sad, at the same time. It is not until middle childhood, between the ages of 7 and 9, that children acquire the ability to integrate positive and negative concepts about the self and become less likely to engage in the kind of all-or-none thinking about self that they display earlier.

Similarly, while even infants are capable of perceiving that other individuals (particularly their significant socializing agents) have reactions toward them, it is not until early to middle childhood that children acquire an ability for mentalization—a cognitive appreciation for the perspective of others and recognition that others have a particular viewpoint and are actively evaluating them and their behavior. As this perspective-taking and mentalizing ability develop, it begins to function as a guide for children so that they are able to identify more with what others expect of them and are increasingly capable of regulating their own behavior. Consistent with Erik Erikson’s concept of *psychosocial identity* introduced in his seminal work, “Childhood and Society” (1950), Roeser and Peck (2009) describe self-identity developmentally in a relational context between an individual’s intra-individual self-identity structure and the extra-individual self-identity structures to which they are currently exposed. In their Basic Levels of Self (BLoS) model, this relational identity refers to specific self-other part-processes within the larger set of multilevel systems (body, mind, social and physical worlds) within which every human being negotiates their particular lifepath. It is not until adolescence, however, that perspective-taking goes beyond recognition of others’ expectations for the self and extends to sophisticated abilities for evaluating the self independently and abstractly. It is during this developmental phase that dramatic increases in introspection are observed, which in some individuals appear to be associated with enhanced vulnerabilities to depression or social anxiety. The Basic Levels of Self-in-Context (BASIC) Model further expands upon the developmental changes in self from the vantage point of individuals’ selves/identities—where selves/identities are defined as multilevel phenomena involving individuals' embodied ways of being in their physical and social worlds given the affordances and constraints of those worlds, their habitual states of awareness, and their complex system of psychological representations of self, social others, and the world.

Recent developmental cognitive neuroscience studies have similarly described adolescence as a period of profound social cognitive change. Given the emphasis during this period on developing self-identity, self-consciousness, and social relations as an independent person in society, the changes in the developing brain provide clues that identify neural correlates for the self. In terms of brain development, numerous neuroimaging studies have revealed differences in the density of white and grey matter between the brains of prepubescent children and adolescents (12+ years old). In terms of cortical grey matter density, sensory and motor brain regions are found to mature earliest, with frontal and parietal cortex peaking around 12 years old. Adolescents consistently have a higher volume of white matter and a lower volume of grey matter in the frontal cortex and parietal cortex compared with prepubescent children. The increase in white matter continues through the 20s and has been interpreted as reflecting continued axonal myelination and synaptic proliferation. A sharp acceleration in grey matter loss between childhood and adolescence has been consistently found in the dorsal lateral prefrontal cortex (dlFPC) and the parietal cortex, suggesting synaptic pruning in those areas. Other studies of adolescents versus adults during mentalizing and self-knowledge retrieval tasks find that the dorsal medial prefrontal cortex (dmPFC), and specifically FPC, is more active in children than in adults, suggesting this region is heavily involved in developing self-other relations but requires less engagement as these functions become more routine later in development.

Despite the general utility of the Jamesian framework and proposed developmental approaches, there is no clear consensus on a standard developmental taxonomy of self-representational phenomena at this time. This is particularly apparent in research on adult populations, where debates about self and its conceptualizations are, at times, still fairly contentious. Self-identity across the lifespan is thought to be highly influenced by experience in the context of societal and cultural pressures. Significant coordination among the comparative evolutionary, human developmental, contemplative, and traditional (adult human) investigative approaches will likely provide a broadly applicable theoretical framework for research on “self” into the future.

CLINICAL PERSPECTIVE

In her 1961 work, “Seduction of the Minotaur,” French-Cuban-American essayist Anaïs Nin describes, “We don’t see the world as it is, we see the world as we are.” An essential text from the Buddhist canon, The Dhammapada, also describes, “Our life is shaped by our mind, for we become what we think.” Both perspectives allude to perception and thought content that has the potential to influence biologic dispositions and significantly contribute to personality characteristics, enduring behavioral traits, and habits. For example, an emerging area of investigation on mindsets has demonstrated that particular ways of perceiving and thinking contribute to core assumptions and expectations we maintain about ourselves and how we interact with the world. These mindsets contribute to the placebo and nocebo effects with a strong influence on health outcomes and associated neurophysiology. For instance, the influence of the placebo effect—triggered by a specific mindset and mental approach to ailments, exercise, and the prospect of recovery—has been demonstrated to account for 43% to 71% of meaningful clinical improvements in a range of conditions. These include pain management, anxiety and depression relief, respiratory and cardiovascular issues, treatment of neurodegenerative diseases, bolstering of immune function, postsurgical recovery, and even in reducing premature mortality. Similarly, negatively-oriented mindsets and associated habitual thought processes contribute to self-identity schemas that can have a negative impact on mental health.

Much of the earliest work encouraging scientific investigation of self-representation arose in the clinical context of treatment for adolescent and adult patients with psychiatric disorders. Sigmund Freud was most prominently recognized for his contribution of the tripartite model of the human psyche. Freud’s personality theory (1923) saw the psyche structured into three parts, the id, ego, and superego, all developing at different stages in our lives. These are implicit systems without neurophysiologic correlates and are directly related to the development of self-identity based on childhood experiences. In essence, the self was described to be composed of (1) the id, a primitive and instinctual part of the mind that contains sexual and aggressive drives and hidden memories, (2) the superego, a moral conscience, and (3) the ego, a realistic part that negotiates between the desires of the id and the superego. Freud’s theory helped create a psychoanalytic technique for dismantling self-relations to improve mental illness in adulthood. Heinz Hartmann (1946) built upon Freud’s ego psychology and proposed innate capacities of ego that facilitated an individual’s ability to adapt to their environment. These included cognitive, perceptual, and motor coordination. Under normal conditions, which Hartmann called “an average expectable environment,” these capacities developed into ego functions with autonomy from the libidinal and aggressive tendencies. Heinz Kohut, an Austrian-born American psychoanalyst, helped further craft psychodynamic theory for the treatment of psychopathology, emphasizing an understanding the self that focuses on its developmental aspects and the role of interpersonal relationships. Although his work evolved over time, by 1962 he had outlined several key components that contribute to the formation of healthy self-object relations. In Heinz Kohut’s self psychology, the term “self-object” refers to a person or symbolic representation that serves specific psychological functions for the self, particularly in the realms of validation, care, and support. In essence, self-objects are those entities in our external environment—usually significant people like parents, caregivers, teachers, mentors, and later in life, peers and romantic partners—whose functions and responses are essential for developing and maintaining our sense of self. Kohut argued that self-objects are crucial especially during the formative years for providing the experiences that lead to a cohesive, healthy self. These experiences generally fall under three categories: (1) Mirroring: Self-objects in the form of parents or caregivers reflect back the child’s own sense of worth and value. This mirroring validation gives the child a sense of vitality and effectiveness. (2) Idealizing: In this function, the self-object acts as a figure of strength and calm onto which the child can transfer feelings of fear or instability. The child idealizes this powerful, nurturing figure and seeks to be like them, absorbing their qualities into their own sense of self. (3) Twinship (or Alter Ego): Twinship self-objects are those with whom individuals feel a deep sense of similarity or likeness. These relationships often serve to confirm one’s own sense of being and of belonging to a community. Importantly, the need for self-objects doesn’t entirely vanish as we grow older. Adults also seek self-object experiences, often through intimate relationships, friendships, or mentorship, to maintain a cohesive sense of self and to repair disruptions in their self-cohesion. Kohut's theory had revolutionary implications for psychotherapy, suggesting that therapeutic success often hinges on the therapist becoming a self-object for the patient—providing the mirroring, idealizing, and twinship experiences that might have been missing or flawed in the patient's earlier life. Object relations theorists such as Margaret Mahler (1897–1985), Melanie Klein (1896–1971), and Otto Kernberg (1928) were also responsible for developing ego psychology and recognized the importance of early object relations and stages of self-development. Object-relations theorists recognized the centrality of the early mother–infant relationship for personality development, the development of an internal representation of self and others, and for the early understanding of interpersonal relationships. Carl Jung, a Swiss psychiatrist and psychoanalyst who founded analytical psychology, described the self as not just the conscious ego, but as a larger, overarching structure that includes both conscious and unconscious elements. The self, in Jung's framework, serves as the central organizing principle of the total psyche, striving for wholeness and integration. In seminal works like "Man and his Symbols" (1964) and "Memories, Dreams, Reflections" (1962), Jung described the self as inclusive of the conscious mind (ego), the unconscious that contains weakness, desire, and instinct (shadow), the inner contrary gender characteristics (animus), the social mask or facade one assumes in public (persona), and innate, unconscious universal patterns or images that shape human experience (self-archetypes). Last, Jung described a collective unconscious dimension of the self, which reflects a reservoir of archetypes and collective experiences that have accumulated over the course of human history.

The rise of behaviorism and cognitive psychology in the mid-20th century had a significant impact on cognitive and behavioral theories that explained psychopathology. Aaron Beck, one of the founders of Cognitive Behavioral Therapy (CBT), conceptualized the role of "schemas" in shaping one's self-concept and, ultimately, one's emotional and behavioral responses. In Beck's model (1979), "self-schemas" refer specifically to the durable reified cognitive structures that contain our beliefs about ourselves. These enduring self-schemas influence how we interpret experiences, what we pay attention to, and how we react emotionally and behaviorally. A healthy self-schema helps a person adapt effectively to life challenges, whereas maladaptive self-schemas can lead to emotional distress and dysfunctional behavior. Beck proposed that psychopathological self-schemas are enduring and dysfunctional cognitive structures that contribute to the onset and maintenance of psychological disorders. For example, in depression, these schemas might consist of persistent, ruminative beliefs like “I am worthless,” “I am unlovable,” or “I am not good enough.” These schemas become activated by specific situations that are perceived as threatening or confirmatory of the maladaptive beliefs. Once activated, they bias cognitive processing, leading to distortions in perception, memory, and thought, which in turn maintain or exacerbate emotional distress, maladaptive behavior, and lasting depressive attitudes and dispositions. There is now clear evidence that psychopathology is particularly vulnerable to negatively oriented mindsets that perpetuate negative self-schemas. Treatment of psychopathology through CBT involves identifying these dysfunctional self-schemas and cognitive distortions, challenging them, and replacing them with more adaptive beliefs and thought patterns about the self and the world.

Although much research has emerged in the last century to delineate the clinical relevance of the self and how it is represented and embodied in the context of mental illness, the influence of cognitive schemas, mindsets, and mindful awareness has only recently been leveraged in the context of therapy to facilitate a de-reified and more adaptive form of self-representation.

Brain Networks Supporting Self-Representation and Self-Transformation

A significant amount of PET and fMRI data have emerged from functional imaging studies specifically targeting various properties of the self, implicating coordinated activity in a network of specific frontoparietal association cortices (see Gusnard, 2005).

Vago and Silbersweig (2012) proposed a systems-based neurobiologic model for self-transformation that focused on frontoparietal cortices supporting the reduction in emotion dysregulation and the development of a sustainable healthy mind through changes in self-representation, including self-awareness, self-regulation, and self-transcendence (S-ART). The S-ART model describes transformative changes in the self through systematic training of the mind using specific meditation techniques and mindful awareness. Such techniques facilitate insight and familiarity with the conditions which cause (and remove) distortions or biases in the individual's construction of their external or internal experience through interoceptive attention, cognitive and emotional regulation, and prosocial development. Meta-awareness is described as a fundamental mechanism by which the individual gains awareness and control over pathologic thought processes common to emotion dysregulation and most mental illnesses. Meta-awareness is a combination of reduced reactivity to and disidentification from one's internal experience, improving the maladaptive emotional and cognitive processes. Specific resting-state networks (RSNs) were identified as comprising different aspects of the self in action or "selfing" (see Fig. 1.29–2), including networks supporting enactive experiential, phenomenological, narrative, and integrated self-related processes.

Enactive experiential brain regions are described as supporting early, prereflective, nonconceptual processing of the world. The self-specifying sensory–motor convergence proposed to contribute to enactive self-processing is localized to a distributed set of interconnected spinothalamocortical regions (including the periaqueductal gray [PAG], midbrain colliculi [MC], thalamus, hypothalamus, and posterior insular cortex [PIC]), all of which are known to be closely involved in monitoring (orienting and tracking), deciphering, and controlling interoceptive feedback at a nonconscious level. Frontal eye fields (FEF) and pre- and supplementary motor areas (PMA/SMA) were also described as critical to enactive self-processing. Experiential phenomenological self-processing involving the sense of being the immediate subject of experience without reflection or evaluation is described to involve primary sensory cortices (S1), the dorsal lateral PFC (dlPFC), dorsal anterior cingulate cortex (dACC), precuneus, and temporoparietal junction (TPJ). Together, these regions are believed to facilitate "executive" monitoring, vigilance, and alertness—monitoring performance or arousal levels and maintaining response readiness. The narrative self refers to the empirical or autobiographical construction of self-identity proposed by Daniel Dennett (1992), and also described in the pattern theory, to highlight the continuity over time as it extends in a coherent way into the future and reflectively into the past. The brain regions that have come to be most identified with the narrative self comprise the default mode network (DMN), an RSN that consistently incorporates the ventral posteromedial cortex (vPMC), including posterior cingulate cortex (PCC) and retrosplenial cortex (RSC), ventral medial prefrontal cortex (vmPFC), posterior inferior parietal lobe (pIPL), hippocampus, and lateral temporal lobe. Sometimes referred to as the hippocampal–cortical memory system, the DMN has occasionally been reported to also include the dorsomedial/rostromedial PFC (including BA 8, 9, and 10), rostral anterior cingulate cortex (rACC, or anterior medial PFC), insular cortices, and temporal pole. The DMN and other RSNs were first described in early functional connectivity studies of passive resting states using fMRI to probe spontaneous low-frequency (<0.1 Hz) blood oxygen level–dependent (BOLD) fluctuations that demonstrate consistent spatially and temporally coherent connectivity among large-scale functional brain networks. It has been hypothesized that DMN regions—especially the vmPFC—support the introspective self-reflective and evaluative mental processes that are characteristic of resting-state mind-wandering conditions. This conception of the core DMN as a “self-making” network is reinforced by findings that these regions are among the brain’s most highly integrated and play a major role in processing autobiographical events and representations of individuals, including oneself. Jonathan Smallwood and Jonathan Schooler (2015) propose mind wandering involves engagement of episodic and emotional processes supporting the narrative self in addition to *perceptual decoupling*, in which attention is disengaged from perception during periods of self-generated or internally directed cognition.

The final network proposed by Vago and Silbersweig to support the integration of self-related informational processing is the frontoparietal control network (FPCN) (rostral frontopolar PFC [FPC], right ventrolateral PFC [vlPFC], dmPFC, dorsal anterior cingulate cortex [ACC], dlPFC, anterior insular cortex [AIC], lateral cerebellum, and anterior inferior parietal lobe [aIPL]). Some central executive and FPCN regions often overlap with other RSNs and are depicted in Figure 1.29–2 as cross-modal areas. The FPCN has been described as a hub and supervisory attentional gateway coordinating information processing between RSNs. Described to specifically facilitate relational binding and controlling stimulus-independent and stimulus-oriented cognition, the FPCN is uniquely situated to integrate information coming from other RSNs and facilitate global reorganization or plasticity among networks. Both global reorganization and integration across networks appear to support self-transformations and improved clinical health outcomes. For example, FPCN-DMN connectivity has correlated with improved clinical outcomes in support of mindfulness meditation training and psychedelic-assisted psychotherapy. Studies of mindfulness-trained individuals, including a seminal paper by Norman Farb and colleagues (2007), found unique engagement of experiential self-networks and suppression of DMN regions supporting a narrative self-focus. A number of meta-analyses have identified mindfulness-mediated targeted activity, related morphologic changes, and functional connectivity changes to the FPCN, associated regions, and their interaction with the DMN.

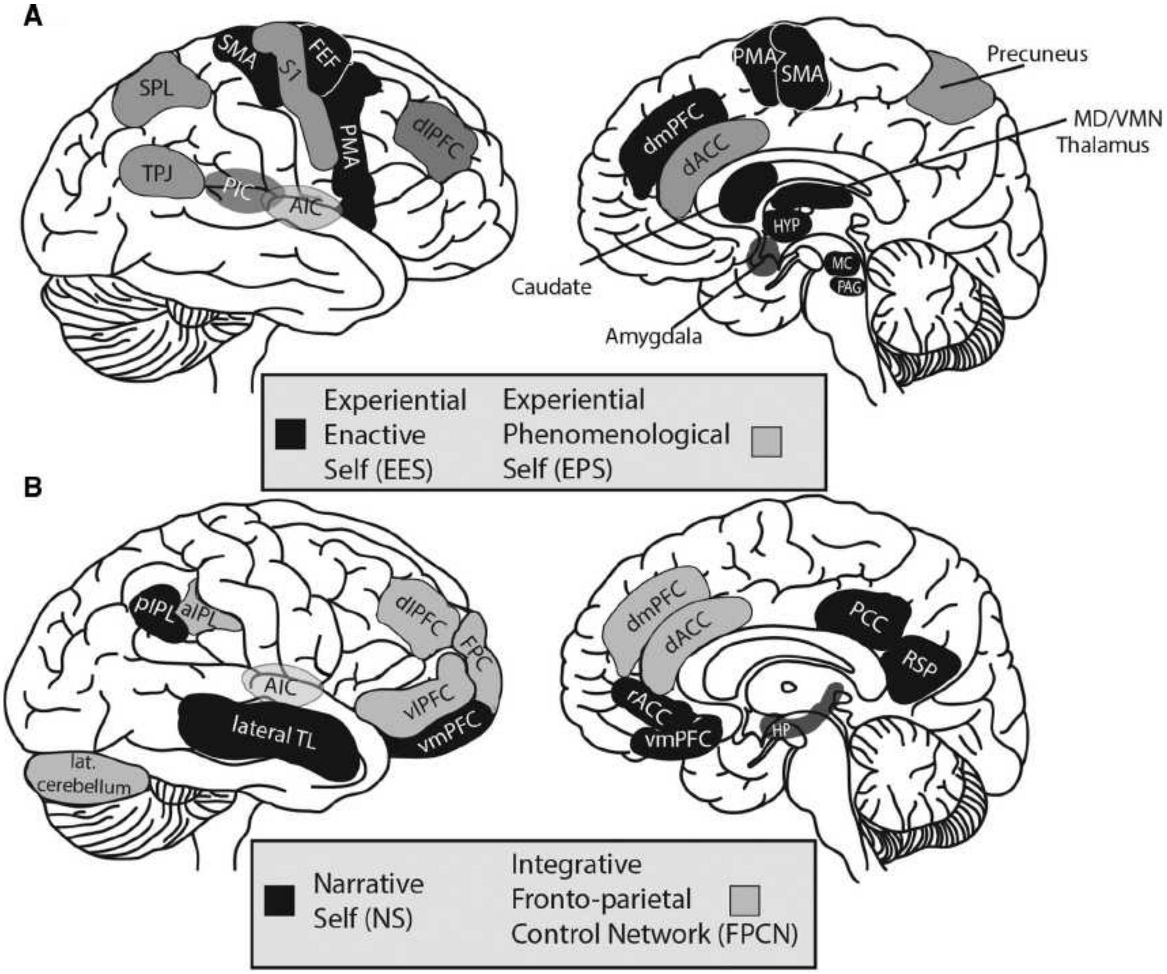
A
SMA
S1
FEF
SPL
PMA
dlPFC
TPJ
PIC
AIC
PMA
SMA
Precuneus
dmPFC
dACC
MD/VMN
Thalamus
HYP
MC
PAG
Caudate
Amygdala
■ Experiential Enactive Self (EES)
Experiential Phenomenological Self (EPS) ▨
B
pIPL
aIPL
dlPFC
FPC
AIC
vlPFC
lateral TL
vmPFC
lat. cerebellum
dmPFC
dACC
PCC
RSP
rACC
vmPFC
HP
■ Narrative Self (NS)
Integrative Fronto-parietal Control Network (FPCN) ▨

FIGURE 1.29–2. Functional networks shown to support different aspects of the self in action or “selfing.” Functional networks originally described by Vago and Silbersweig, 2012 include an experiential enactive self (EES) network, experiential phenomenologic (EPS), narrative self (NS), and integrative frontoparietal control network (FPCN), each of which involves parcellation nodes previously identified in resting-state networks. The self-specifying sensory–motor convergence proposed to contribute to EES is localized to a distributed set of interconnected spinothalamocortical regions (including the periaqueductal gray [PAG], midbrain colliculi [MC], thalamus, hypothalamus, and posterior insular cortex PIC]), all of which are known to be closely involved in monitoring (orienting and tracking), deciphering, and controlling interoceptive feedback at a nonconscious level. Frontal eye fields (FEF), pre- and supplementary motor areas (PMA/SMA) are also described as critical to EES processes. The EPS involves the sense of being the immediate subject of experience without reflection or evaluation, and is described to involve primary sensory cortices (S1), the dorsal lateral PFC (dlPFC), dorsal anterior cingulate cortex (dACC), precuneus, and temporoparietal junction (TPJ). Together, these regions facilitate “executive” monitoring, vigilance, and alertness—monitoring performance or arousal levels and maintaining response readiness. The NS refers to the empirical or autobiographical

construction of self-identity mostly supported by the default mode network (DMN), a resting state network that consistently incorporates the ventral posteromedial cortex (vPMC), including posterior cingulate cortex (PCC) and retrosplenial cortex (RSC), ventral medial prefrontal cortex (vmPFC), posterior inferior parietal lobe (pIPL), hippocampus, and lateral temporal lobe. Sometimes referred to as the hippocampal–cortical memory system, the DMN has occasionally been reported to also include the dorsomedial/rostromedial PFC (including BA 8, 9, and 10) rostral anterior cingulate cortex (rACC, or anterior medial PFC), insular cortices, and temporal pole. The FPCN supports the integration of self-related information processing and includes the rostral frontopolar PFC (FPC), right vlPFC, dmPFC, dorsal ACC, dlPFC, AIC, lateral cerebellum, and anterior inferior parietal lobe (aIPL). Some central executive, FPCN, and DMN regions often overlap in functional parcellations with other RSNs identified by Yeo et al. 2011 and are depicted in figure as cross-modal areas. (Adapted from Vago DR, Silbersweig DA. Self-awareness, self-regulation, and self-transcendence (S-ART): a framework for understanding the neurobiological mechanisms of mindfulness. *Front Hum Neruosci.* 2012;6:296. doi: 10.3389/fnhum.2012.00296. License: https://creativecommons.org/licenses/by/4.0/.)

Schizophrenia and Other Psychiatric Disorders

Neurologic syndromes, severe psychiatric illness, and known structural brain lesions have also contributed to illustrating many of the phenomena that will need to be explained by any scientific formulation of neural substrates of self-representation. In psychiatry, beyond depression, schizophrenia may be viewed in part as a disorder of the self. This is presumably because symptoms observed in patients with schizophrenia, such as auditory hallucinations ("voices" that likely represent patients' own inner speech), delusions (that commonly involve patients' own identities and powers), thought withdrawal and insertion (delusions regarding ownership and control of patients' own thought processes), and negative symptoms (that include deficits in emotional responsiveness, spontaneous speech, and volition) typify dysfunction of some of the traditional subject and object conceptions of the self already noted.

In the very earliest days of psychiatry, Emil Kraepelin (1896, 1913) claimed that disunity in schizophrenic patients' consciousness ("an orchestra without a conductor") was a core feature of the illness and that this disunity was linked to "a peculiar destruction of the psychic personality's inner integrity, whereby emotion and volition, in particular, are impaired." Such a framing of a core dysfunction in schizophrenia is striking for its resonance with current thinking from dynamical systems and theoretical neuroscience perspectives that consciousness, including aspects of self-representation, fundamentally depends on the brain's *capacity to "properly" integrate and coordinate* information within and across widely separated brain regions.

Several other psychiatric disorders exhibit features that current thinking (according to DSM-5 diagnostic criteria) would suggest involve critical dysfunction of some aspect or aspects of self-representation. Examples include anorexia nervosa (with its disturbance in the way in which one's body weight or shape is experienced), various dissociative disorders (with their losses of continuity in subjective experience, such as fragmentation of identity, depersonalization, and derealization and/or inability to access information or control mental functions that normally are amenable to such access or control), social anxiety disorder (with its marked fear or anxiety of social situations in which the individual is exposed to possible scrutiny by others), and various personality disorders, notably borderline personality disorder, with its pattern of instability of self-images, affects, and interpersonal relationships. This nosologic "list" is likely to become more refined and expanded as research progresses and the current understanding of the neural bases of self-representational capacities increases.

Neurologic Syndromes

Seizure disorders have been shown to present with experiences of depersonalization and derealization, or even autoscopy, suggesting self-representation may be disrupted by focal or global electrical disruption. Epilepsy with "ecstatic auras" have been described as experiences of intense feelings, including bliss, joy, and heightened self-awareness. This unique seizure disorder has been associated with abnormalities in the anterior temporal pole (TP) and AIC. This region has also been implicated in representing the "sentient self" or interoceptive awareness. Some patients who are known to have sustained brain damage, usually as a result of a cerebrovascular accident, display dissociations between aspects of their behavior and their reported experiences or stated beliefs about themselves. For example, there are patients with alien limb and alien hand syndromes who, by definition, exhibit movements of all or a portion of one of their upper extremities that are dissociated from their stated intentions and are often in conflict with those of their opposite extremity. The patients react with surprise and concern at such movements. These involuntary movements are not spontaneous but are in response to stimuli in patients' peripersonal space. The qualities of these movements are different depending on the location of their brain lesions. Patients with damage to the medial surface of the frontal lobe, for example, typically display intermanual conflict or grasping, groping movements of the hand represented by the lesioned cortex, and an inability to release grasped objects. Patients with parieto-occipital lesions, by contrast, display involuntary movements that involve avoidance of contacts so that the affected hand and limb often levitate away from support surfaces. Damage to the precuneus and surrounding lateral and medial parietal structures appears to be critical to the control of goal-directed visually guided reaching in humans as well as motor, sensory, or visual field–related disorders like optic ataxia and ability to perceive the visual field as a whole as seen in simultagnosia.

In addition to neurologic disorders that exhibit dissociations between self-perceived will and purposive action, there are other disorders that involve an apparent lack of self-awareness with regard to some acquired impairment of the body that is clearly apparent to observers. In some cases, this lack of awareness is associated with unshakeable false beliefs about the alteration in their bodies or their impairments. The syndrome of unilateral spatial neglect, for example, is seen in a heterogeneous group of patients who have suffered damage, usually to the right parietal lobe (right frontal and left parietal variants are uncommon). Though specific symptoms and their severity vary, all patients exhibit some failure to detect or respond to stimuli on the side opposite to the brain damage. Along with this, patients may have an isolated delusion where they deny ownership of the limb or entire side of the body represented by the lesioned cortex. At the same time, they continue to retain awareness of subjective sensations such as whether they are tired or hungry and continue to have normal autobiographical memory.

Patients who have sustained brain injuries in various locations (again, typically in the right hemisphere) may also remain unaware of and deny their handicaps (anosognosia). This is seen acutely in 20% to 30 % of cases of hemiplegia/hemiparesis after stroke but can occur with virtually any neurologic impairment, including bilateral cortical blindness, prosopagnosia (face blindness), amnesia, aphasia, or apraxia, as well as in cases of multiple impairments where lack of awareness may be for all or limited to just one of the impairments. The amputation of a limb represents a serious disruption of body integrity and is associated with postamputation “phantom limb” pain. Almost all amputees report ongoing awareness of a phantom limb, with 60% to 80% complaining about phantom limb pain, a painful sensation perceived in the amputated limb and thus located outside the physical borders of the body. Such denials of one’s own impairments do not appear to be directly related to sensory loss but rather to damage to higher-level cognitive processes, including attentional mechanisms, which likely disrupt some capacity for self-monitoring. Nonlocalizability (except to the right hemisphere) also suggests a disturbance of a brain “dynamical” mechanism (e.g., an integrative mechanism, such as one that permits the coordination of sensory information with other processes that support spatial or bodily representations) as its basis. Such attentional and “dynamical” mechanisms need not be mutually exclusive.

There are also patients who exhibit dissociations between their behavior and reported sense of themselves of another sort. These patients have often been characterized as having disorders of self-regulation. Adult patients with acquired lesions involving the prefrontal cortex, particularly the ventromedial region, often exhibit changes in their ability to control their behavior. While in testing situations, they can verbalize norms for acceptable social behavior, in actual practice, they behave in ways that conflict with their reportable knowledge. It is as if, despite knowing “the rules,” they are unable to implement them. Such patients often appear impulsive, socially insensitive, and as if their decision making, particularly in social contexts, is impaired. They are commonly described by those familiar with them as having undergone a “personality” change. This was seen in the case of Phineas Gage, mentioned earlier, who sustained a traumatic injury to his OFC (see Fig. 1.29–1). When the prefrontal damage is of childhood onset, affected individuals exhibit defective social and moral reasoning as well, suggesting that capacities for acquiring knowledge of complex social conventions and making moral judgments have also been impaired. Such “moral self-effect” changes that shift characteristics like social reasoning, honesty, empathy, or virtuous behavior have been suggested to constitute the greatest weight and importance to self-identity in comparison to other aspects of change to mental life (e.g., memories, preferences, desires).

These various clinical syndromes demonstrate some of the ways in which impairments in capacities for self-representation manifest themselves and result in the disunity of functioning that is perceptible to the subject in some cases and only to outside observers in others. Patients' verbal reports and clinical observations of patients' behavior are clearly limited in what they can contribute to the development of a taxonomy of self-representation functions and underlying mechanisms, however. Although a pattern of self-specific neural substrates appears to be emerging (see Brain Networks Supporting Self-Representation and Self-Transformation), more experimental work is needed with neurologic cases and those not afflicted to better characterize the instantiation of such impaired aspects of self-representation in nervous systems and elucidate an even wider variety of pathology involving the disordering of aspects of human subjectivity and behavior.

Split-Brain Studies

Some of the most significant data providing conceptual traction on the issue of self-representation in the human nervous system arose from the aforementioned ground-breaking studies of patients who had undergone complete transection of the corpus callosum for the treatment of medically intractable epilepsy that were first undertaken in the 1960s. Functional specializations of the two cerebral hemispheres (e.g., the dominant role of the left hemisphere in supporting most language functions) were the most widely reported findings of the initial studies that were conducted by Roger Sperry and for which he won a Nobel Prize in 1981. Sperry used several ingenious tasks to investigate the lateralization of brain function. The experiments all involved presenting tasks separately to the two cerebral hemispheres in each experimental subject. For example, after blindfolding one of the subject's eyes, subjects would be asked to fixate with the seeing eye on a point in the middle of a screen to the left or right of which a stimulus would subsequently be presented for less than one-tenth of a second. This very short duration ensured that there would not be time for subjects to move their eyes and permit the visual information to "spread" across both left and right visual fields and, therefore, across both sides of the brain. Analogous procedures were used for presenting unseen tactile stimuli to one or the other of the subjects' hands or auditory and olfactory stimuli to only one side of their brains. The subjects would then be asked to name the various stimuli. Because language is largely processed in the left hemisphere, it was only when stimuli were presented so that processing occurred in the left hemisphere that subjects were able to name the stimuli (i.e., because of fiber crossing, this would occur when stimuli were presented to subjects' right visual fields or right hands). It was also observed that when subjects were presented with a stimulus first in one of the two visual hemifields and then the other, they responded in the second case as if they had never seen the stimulus before. But when the stimulus was represented to the original hemifield, subjects were able to recognize it as the one they had seen before. Other examples of the separability of processing in the two hemispheres were seen in related experiments where different objects would be placed in each of the participants' hands at the same time and then placed for retrieval in a pile of test items. Each hand was observed to search out its own object, rejecting the other item for which the other hand was searching.

Besides revealing specialization of the two hemispheres, these studies also showed that a manipulation (in this case, a disconnection of the two hemispheres) was able to result in a real-time separability of aspects of an individual person's consciousness. Long-term observations of such patients showed that each disconnected hemisphere possesses not only a separate sensorimotor interface with the environment, with its own perceptual, mnestic, and linguistic repertoires, but also its own characteristic likes and dislikes and styles of decision making. Researchers involved in conducting experiments with such patients also reported that the experience of "interacting" with each of the separated hemispheres in individual patients felt akin to interacting with a distinct personality.

Recognizing that the partially disconnected right hemisphere demonstrates awareness in the absence of an ability to express such awareness verbally has provided additional evidence that language is not necessary for human consciousness. Also, while the normal experience of consciousness is often thought to be inherently unified, it now appears that normal consciousness may be fundamentally "dual" (if not "multiple"). That is, it is grounded in parallel processes in the two hemispheres that tend to interact and accommodate the resources of each other in most settings, but whose partially separable workings may sometimes be in evidence, as in the common human experience of subjective conflict.

And, despite the fact that the two cerebral hemispheres in a single split-brain individual clearly exhibit signs of distinctiveness in processing styles that occur in parallel, other mechanisms exist for "driving" coherence in the nervous system in natural settings. For example, there are also pathways at levels below that of the corpus callosum that indirectly connect the two hemispheres via the cerebellum, brainstem, and the thalamus. These may facilitate interhemispheric cooperation for compatible stimuli in the two visual fields and the typically observed unity in motor responses. Findings such as these and others from real-world settings that appear to engage other compensatory processes have led increasing numbers of theoretical neuroscientists and others to begin viewing the brain and its operations as a kind of “complex adaptive system,” subject to self-organizing principles and dynamics similar to those observed in other domains.

Distinguishing Self from Nonself

It is critical for the integrity and survival of all mobile organisms that they be able to distinguish “self” from “nonself” and act on the basis of such a distinction. The evolution of nervous systems from the simplest vertebrates to humans has resulted in marked increases in complexity and diversity of capacities for achieving various forms of such “self-recognition.” Brains, and the nervous systems of which they are a part, are thus being increasingly viewed as inherently dynamical. Their ongoing activity, the patterns of communication between brain areas residing in such activity, and adaptive “problems” they “solve” likely subserve various forms of self-representation. For example, the fundamental problem of distinguishing self from nonself exists for all mobile organisms. It has been studied in lower animals as well as humans. One well-studied animal model is that of the electric fish. Studies of this animal have been useful for identifying strategies that a relatively simple sensorimotor system may use for solving the inverse problem of determining which of all the patterns of neural activity in its sensory system represent those arising from environmental events (nonself). All such strategies are made possible by the animal’s movement and exploration of its environment, which result in modification of the kind of sensory information that it acquires and help constrain the solutions to this problem.

Some of the modifications that help obtain such a solution arise from nervous systems’ determination of different temporal relationships between, and integration of, sensory afferent and motor efferent signals. As a result of interaction with the environment, an organism must constantly integrate efferent and afferent signals to distinguish between reafference—afferent signals arising as a result of the organism’s own efferent processes (self)—and exafference—afferent signals arising as a result of environmental events (nonself). This functional yet implicit self–nonself distinction specifies the self as a bodily agent. Internal reports of an organism's own movements are referred to as corollary discharge signals. Corollary discharge information informs sensory areas about the upcoming movements and allows them to prepare for the sensory consequences of the movement. These prediction signals arising from the motor system and their comparison with sensory signals are thought to play a major role in neural processes that permit the separation of self-generated movements from movements caused by another individual (passive movement) or from other activity that is going on in the environment. That is, well-established theoretical frameworks (called feed-forward models) have been developed for computational considerations of processes describing such action execution and motor control by nervous systems. In these frameworks, at least one kind of neural signal that is thought likely to be a component of this preparatory activity for movement execution is one *predicting* the sensory consequences of the movement.

Sarah-Jayne Blakemore, Daniel Wolpert, and Chris Frith conducted a now-classic experiment (2000) in which normal subjects underwent self-produced tactile stimulation and robotic stimulation using the same stimulus (a sinusoidal movement having a particular amplitude and frequency of a piece of soft foam across the right palm). In a related experiment, subjects underwent the same kind of stimulation, but it was achieved via a robotic interface that the subjects manipulated with their own left hands and into which different time delays were introduced (between when the subjects moved their left hand and produced the stimulation on their right palm). Under these controlled conditions, subjects rated the self-produced tactile stimulation as significantly less tickling and pleasant than that produced by the robot. They also reported a progressive increase in the tickling and pleasant sensation as the temporal delay increased from no delay to a 200-ms delay. Such data are consistent with the feed-forward model's hypothesis of perception of self-produced tactile stimulation being able to be attenuated due to precise sensory predictions. A follow-up neuroimaging study of the self versus robotic stimulation comparison supported this by showing that there was relatively attenuated activity in the parietal operculum (secondary somatosensory cortex) and anterior cingulate cortex (ACC) in the self-condition.

In similar experiments as well as those using other designs, individuals with delusions of control have been shown to differ in their responses from individuals without such delusions. Other designs include the “rubber-hand illusion,” an established method for manipulating the sense of body(part) ownership, where subjects see feedback of their own hand movement or that of an experimenter’s hand making similar movements. In such experiments, individuals with delusions of control have found it more difficult than control subjects to distinguish sensations arising from self-generated actions from those arising from actions that were not self-generated. They also required greater temporal discrepancies between such actions to distinguish between actions arising from themselves and those not arising from themselves. These and other results have suggested that some component of the feed-forward modeling process for making such discriminations is impaired in these individuals.

Self versus Other: Perspective-Taking and Mental State Attribution

In human beings, there is another set of functions by which an individual’s experience of self may be uniquely characterized. These involve the manipulation of perspective-taking and mental state attribution. First- versus third-person perspectives are quite different from each other. First-person perspective refers to the centralization of the subjective multimodal experiential space around one’s own body, while third-person perspective refers to a variety of processes that all include ascribing mental states to others and shared experience with others.

Individuals may adopt first- versus third-person perspectives with regard to movement through or location in space. Spatially, an individual is situated in what is called an egocentric reference frame, where locations of objects are represented relative to that individual’s body. However, when individuals try to understand what someone else sees, for example, they need to be aware that others’ percepts are actually different from their own, *and* they need to be able to perform the mental computation that permits reorienting the egocentric reference frame and centers it on that other person’s body. Electrophysiologic studies in primates have yielded much data on the role of the posterior parietal cortex, including the precuneus, in multisensory integration and spatial coordinate transformations for this purpose. These processes are also required for converting sensory input centered on an organism’s body into motor output directed at objects located in an allocentric reference frame that references object–object relations with internal coordinates that are independent of an organism’s body. In human brain imaging studies, the posterior parietal cortex, usually on the right side, has routinely been activated when people have been asked to execute tasks that involve adopting a third-person perspective.

The capacity to be aware that others have mental states that are different from one’s own is critical for individuals’ ability to participate successfully in human social interaction. Much of human cooperative behavior depends on the ability and motivation to exchange information, the recognition that others lack important knowledge, have beliefs that are false, or have intentions that complement or conflict with one’s own. Human social interaction is also supported by individuals’ abilities to understand the emotional and affective states of selves and others. Abilities to reflect on one’s own feelings as well as to be empathic, feel sympathy, and have compassion for others contribute to moral decision making and help promote social cohesiveness. Cognitive empathy or mentalizing is often characterized uniquely from affective empathy and is more often associated with shared experience through perspective-taking. The cognitive form has been measured by items describing the understanding of others’ feelings (e.g., “I’m able to recognize, before many other children, that other people’s feelings have changed”), and affective empathy by items about sharing others’ feelings (e.g., “When somebody tells me a nice story, I feel as if the story is happening to me”). Powerful position and status among peers are increasingly important for adolescents, and both forms of empathic ability have been shown to also contribute to increasing and maintaining social status.

Research using neuroimaging techniques has shed light on the neural mechanisms underlying the process of understanding others’ mental states, a key aspect of perspective-taking. These studies often distinguish between tasks and stimuli related to the self and those that are not. Notably, evaluating one’s own personality traits, physical features like facial appearance, or tasks that require interpreting emotions or mental states, consistently activate certain brain regions. These include the medial prefrontal cortex (mPFC), the precuneus, the posterior cingulate cortex (PCC), the temporoparietal junction (TPJ), and the temporal pole (TP) (see Fig. 1.29–2). For example, self-face recognition and mental state attribution in comparison to other familiar faces has been found to have particular neural substrates supporting this process, including the middle and superior frontal gyri in the right hemisphere. The mPFC has shown activation when subjects have been asked to perform various acts of self-evaluation. For example, mPFC and experimental tasks have included asking subjects to make trait judgments about themselves ("does this word [e.g., friendly, unhappy, industrious] apply to you") and make preference judgments ("which do you prefer—item 1 or item 2").

These data support a model developed by Gordon Gallup Jr. (1982) that posits that self-awareness and mental state attribution are part of a shared neurocognitive suite of processing and that neural architecture implicated in processing knowledge about the self is called upon when inferring knowledge in others. Gallup devised mirror recognition tasks in nonhuman primates as a distinguishing factor for self-awareness capacity. Other studies have focused on distinguishing self from nonself through identifying a social network of functioning that includes mentalizing and empathic behavior. Brain areas that are involved in mentalizing include TPJ, TP, AIC, precuneus, dmPFC, while the ACC, inferior frontal gyrus (IFG), posterior superior temporal sulcus (pSTS), and inferior parietal lobe (IPL) have been implicated more with prosociality, empathic concern and experience sharing. Imagined self-action relative to imagined experimenter action strongly activates IPL (including TPJ), somatosensory cortex, and precuneus. The dmPFC has also shown activation when subjects have been asked to assess what others think or feel, as in so-called theory-of-mind (ToM) tasks. ToM refers to the cognitive ability to attribute mental states, such as beliefs, desires, and intentions, to oneself and others and to understand that others have mental states that are different from one's own. The right IPL is more often associated with ToM, whereas the left IPL is more often involved in representing one's own mental states. The posterior IFG has been shown to be involved in emotional judgment and has been suggested to play a role in emotion recognition, predicting emotion, affective empathy, and mirroring action.

Cognitive studies and functional imaging experiments targeting empathy have also been done. Experimental work on this issue has shown that the same network of brain regions is often engaged when subjects experience something (e.g., pain) *and* when they observe someone else appearing to be in the same experiential state (e.g., when shown pictures of others in pain). Such data again support the view that certain brain networks *share* in representing self and others' states and, in this case, account for humans being able to achieve empathic understanding. Brain networks that have been identified in such contexts include but are not limited to portions of PFC, medial parietal cortex, portions of the ACC, and insula. Yan Fan and colleagues (2011) found that the dACC and anterior midcingulate cortex (aMCC) were recruited more frequently in the cognitive–evaluative form of empathy, while the right AIC was found to be involved in the affective–perceptual form of empathy only, and the left AIC was found to be active in both forms of empathy. Tania Singer and Claus Lamm (2009) propose the AIC serves a dual role, a primary mapping of internal states with respect to the self and the predictive representations or simulation of how the same emotional stimuli feel to others.

A growing number of behavioral and neurophysiologic studies demonstrate shared representations between perception, action, and predicted action. There is evidence that perception of an action activates motor representations to the degree that the perceived and the represented action are similar. As such, these representations may be shared between individuals through a theoretical mirror neuron system. Indeed, a mirror neuron system has been identified in the ventral premotor cortex and posterior parietal cortex of the macaque monkey that fires both when it carries out a goal-directed action and when it observes the same action performed by another individual. This common coding reflecting shared representations between self and others helps clarify the meaning of a given object, action, or social situation that may be common to several individuals and therefore activate corresponding distributed patterns of neural activation in similar brain networks of perception–action and inference across individual brains. This shared-representation explanation is also consistent with, though not completely explained by, the tenets of “simulation theory,” which is a framework for describing how individuals understand others’ behavior and subjective states. Identical circuitry for “sharing” or “simulating” others’ experiences is unlikely to be the sole mechanism underlying abilities to appreciate and recognize others’ subjective states as such, however. Being empathic also requires maintaining a separateness of oneself from another so that one can act. In settings such as psychotherapy, for example, both must occur for progress to be made. Thus, mechanisms underlying personal agency likely need to be integrated with those that underlie this kind of equivalence between self and other to accomplish this in practice.

Current evidence suggests that the relatively primitive capacity to recognize features of one's physical self, such as features of one's face or body, is limited to a small set of primate species, including humans. The behavioral mirror self-recognition test has been evidence of such self-recognition capacity. More cognitively abstract forms of self-awareness, such as introspection and mental state attribution, are likely unique to humans. Interestingly, von Economo neurons (VENs) and fork cell neurons (FNs), two morphologically "atypical" types of neurons found predominantly in the AIC and ACC in humans, and some nonhuman primates (e.g., great apes, gibbon, rhesus) have been identified as unique to primates who have self-awareness and capacity for self-reflection. While their exact function has yet to be clearly elucidated, the VENs and FNs have been uniquely associated with the evolutionary development of self-reflective and socially relevant forms of subjective awareness. They uniquely provide connections between cognition and internal homeostatic bodily states implicated in this unique human capacity. The precuneus has also been implicated in self-specified conscious awareness, agency, visuospatial imagery, sensory integration, and episodic memory retrieval. It is heavily interconnected with visuospatial areas in the inferior and superior parietal areas, extensively with dlPFC, premotor areas, and subcortical regions like the thalamus, striatum, claustrum, and brainstem. Interestingly, the precuneus and surrounding posteromedial areas are among the brain structures with the highest resting metabolic rates displaying transient decreases in tonic activity during engagement in non–self-referential goal-directed actions. As one progressively lowers conscious awareness toward states of sleep, drug-induced anesthesia, and vegetative states, the metabolic rate linearly decreases its metabolic output.

Although a neural network supporting social cognitive processes for self–nonself distinctions is gaining more traction in the science of self and social cognition (see description of S-ART model earlier), it remains unwarranted at this juncture to describe any of these brain regions as uniquely self-specific, given the wide range of other cognitive processes these regions support in other contexts. It does shed light on one aspect of self-awareness in which the self is seen as an object, not as the agent or knowing subject. In a large meta-analysis conducted by Dorothée Legrand and Perrine Ruby (2009), self-specific brain regions and networks of brain regions have been identified for self-representation;

however, they emphasized a framework for describing self-processing as a product of more general evaluative mechanisms that, taken separately, are not self-processing specific, but in their aggregate are self-defining. More specifically, they proposed that these mechanisms together operationalize a subjective perspective and permit relating represented objects to a representing subject. Their proposal aligns in several respects with other investigators' emphases on aspects of agency, underscoring the role of functional processes that effectively integrate an individual's motor activity with its predicted sensory consequences in the course of acting in the world.

Another attempt at a broad theoretical framework has used a different approach with potential applicability beyond the human species, approaching the problem of self-recognition specifically. Matthew Apps and Manos Tsakiris (2014) focused on the relatively low-level problem of how the brain processes incoming visual, auditory, somatosensory, or interoceptive input as "self" ("me") and distinguishes it from "nonself" ("not-me"), allowing recognition of an individual's physical self (face, body and movement, and voice) as their own. They describe how the recently developed "free-energy principle" (itself an attempt at a unifying theory of the brain viewed as a self-organizing biologic system developed by Karl Friston and colleagues) can be used to explain many of the current findings in self-recognition research and also relate to higher cognitive processes of self-awareness. Though the details of their theoretical account (particularly the Bayesian principles and mathematical descriptions upon which it is based) lie outside the purview of the present chapter, the basic notion is that there are mechanisms by which the brain optimizes predictions about the sensory consequences of events occurring in the environment (including the results of subjects' own actions) in order to minimize the amount of "free-energy" (i.e., "surprise" or deviation from expectation) and engender states of high predictability. Anil Seth and Karl Friston (2016) expand on this concept to explain how prediction is continually updated through sensory input and how Bayesian principles may further explain our active inferences, hypotheses, or beliefs as neuronally encoded probability distributions. They apply these principles to predictive coding in the context of interoception, in particular, and suggest perceptual predictions arise from agranular motor cortices and are updated based on prediction errors and sensory input. Apps and Tsakiris also argued that several of the brain regions identified by others for so-called self-processing could be interpreted as participating in the mechanisms of minimizing prediction errors and integrating multisensory information for recognizing the “self” versus “nonself” that are critical to their account. Like Legrand and Ruby, they propose that there are general mechanisms underlying the defining of a particular sense of self that are not self-processing specific but are grounded in the operations of brain regions whose dynamic and coordinated activity are what are, in fact, responsible for such self-definition. Clearly, additional work will be required to disambiguate self-other distinctions to more dynamic notions of self-*processing* (i.e., *how* such representations are constructed).

The Self as Agent

With modern scientific discussions of self moving away from a relatively “static” concept to more dynamical, process-based accounts, there are new considerations of self as an agent promoting investigations that target *experiences* of agency (e.g., experiences of mental causation or of being the “source” of intentions and motor commands) or the actual *structure* of agency (the organizational relationships among perception, an intention or goal, planning or selecting a movement consciously or unconsciously, and a motor command that may or may not require subsequent monitoring for reaching its goal). Experiences of agency represent individual agents and their actions as being in a certain way. For example, normal individuals do not simply find themselves walking toward a stairway and then, on the basis of this, acquire a belief that they must be intending to go down the stairs. Rather, the experience is of walking toward the stairway in order to go down the stairs. Individuals usually experience a sense of purposiveness with their actions, which may be related to intentions that are conscious or not. When there is no awareness of a prior intention, it is still the case that normal individuals’ actions involve experiences of having acted in the service of a goal and thus having implemented an intention that they experience as their own. In normal individuals, the sense of intention and of being the “owner” and source of their actions is typically very strong. Normal individuals tend not to doubt that their actions are self-generated. There is now a variety of data on how normal individuals experience their actions as self-generated and the form that this self-generation takes.

Research has investigated the experience of agency by probing awareness. In one study, subjects were typically asked to draw a straight line on a computer screen while they were unable to see their arm or hand. They would be given false feedback about the trajectories of their movements so that they would have to make significant deviations from straight movements in order to reach their goal. However, when asked about their experiences, subjects' verbal reports indicated that they had been unaware that they were making deviant movements. Other experiments involved discrepancies in the timing between expected and actual movements with similar results. Investigators also determined thresholds for the degrees of discrepancy that could exist before individuals reported awareness of a discrepancy when they were asked. When they were not asked, subjects appeared not to notice even when the degrees of discrepancy were higher than these thresholds. Thus, it appears that there are aspects of their own movements about which even normal individuals are unaware. In addition, individuals may be more aware of movements they *intend* to make than those they actually make in particular settings.

Disorders of agency are also described as disorders of volition—that is, the psychiatric or neurologic conditions where individuals' own actions are either denied having been self-generated or are reported as feeling as if they have not been self-generated. Drug- or trauma-induced dissociative experiences similarly result in an experience of disconnection from one's own action, thoughts, feelings, memories, or sense of identity. Agency is also reduced or experienced as illusory during an out-of-body experience in which there is a feeling of disembodiment (location of the self outside one's body), the impression of seeing the world from a distant and elevated visuospatial perspective (extracorporeal egocentric perspective), and the impression of seeing one's own body (or autoscopy) from this elevated perspective. These dissociative and out-of-body experiences demonstrate how multisensory afferent signals interact with body representations to generate a sense of body ownership, and efferent signals may modulate this sense of ownership through the sense of agency. There is some evidence that the sense of agency originates in neural processes responsible for the motor aspects of action, including visual and proprioceptive/kinaesthetic information that provides information about movement in space. Studies of self-attribution of action and ownership most consistently activate a right frontoparietal network, including the superior parietal cortex, the inferior parietal cortex, and the insular cortex. Yet, dysfunction of the TPJ leads to the related failure to integrate multisensory information from one’s own body, also resulting in a disrupted sense of agency. Thus, there appears to be a fundamental relationship between brain activity in frontoparietal regions and conscious perception of one’s agentic self that can be modulated by the dissociative experience. These data suggest interpretation of dissociative experiences by the individual may contribute to expanding consciousness beyond normal limits of perception and boundaries of bodily ownership or becoming constricted and fear-inducing through traumatic forms of dissociative experience. Future research may better distinguish between drug-induced dissociation and that which is associated with post-traumatic stress. More clarity may thus be provided on dissociation as a potential defensive mechanism for the organism with psychotherapeutic implications.

Sense of Agency and Voluntary (“Willful”) Action

There are common human intuitions as well as folk-psychological beliefs that intentions and thoughts *cause* human action. As already noted, however, scientific accounts of intention and its relationship to action reject dualistic accounts, where the “mind” is thought to cause changes in the body, resulting in action. In the reductionist scientific view, both conscious experience (“thought”) and action are *results* of brain activity. Despite the inherent challenges in identifying organized relationships in neural processes that instantiate this process, some investigators have specifically attempted to identify neurophysiologic mechanisms for the human experience of agency.

A classic study by Benjamin Libet and colleagues (1983) has been commonly cited as evidence that consciousness (specifically, individuals’ experience of the intention to move) is not the initial cause of behavior and that behavior occurs instead as the result of a chain of events initiated by nonconscious brain events. Libet monitored subjects’ electroencephalogram (EEG) and muscle activity while subjects performed simple finger movements and asked them when they were aware of the “urge” to move. Though subjects did consistently anticipate the starting time of the movement by over 50 ms, there were signs of preparatory brain activity for this movement that preceded these “awareness judgments” by several hundred milliseconds. Many authors have had concerns about some of the technical details of this experiment, however, and have consequently also disagreed that the results of this study support the idea that “free will” does not exist.

Other investigators have undertaken experiments based on the Libet paradigm to pursue this issue further. These experiments are notable in two respects. First, the basic result that the generation of action (from the beginning of preparatory brain activity to the onset of muscular activity) extends over time and has a phase that is unavailable to awareness has been replicated. However, the details of the relationships among brain activity, conscious intention, and action have also been determined to be more complex than initially surmised. What is now recognized is that action generation is not a “simple” process. That is, it is not simply an underlying neural process of preparation to move and its execution that may (or may not) be experienced. Rather it now appears that the brain processes of preparation, execution of movement, and conscious experience of these phenomena each have several components.

The actual structure of agency itself is thought to depend upon particular organizational relationships among several variables, including intention or goal, action selection, a motor command, and its outcome. Models have been proposed for this structure, with a particularly notable one being a “comparator model” formulated by Daniel Wolpert and colleagues in the 1990s. This model and its derivatives were originally articulated as models of motor control. An important question, however, is whether or not the sense of agency associated with or emerging from this structure depends solely upon comparing the degree of matching between *predicted* and actually experienced consequences of movement (so that sense of agency is strong when there is a close match between predicted and actual sensory consequences of an action and is reduced or absent when the match is poor). Such an account would necessarily imply that the sense of agency occurs *after* an action has been performed and information about its sensory consequences becomes available. While the “comparator model” and its retrospective account of agency have received significant behavioral and neuroimaging support (recall the “tickling” and “rubber-hand illusion” experiments, as examples), other experimental data challenge the notion that sense of agency is *fully* explained by such an account.

The sense of agency may also be generated *prospectively* via the process by which intentions (here defined as prospective signals involved in action preparation) are transformed into specific actions to reach desired effects. In a variant of the Libet paradigm, for example, they stimulated the motor cortex using TMS and observed that, relative to the nonstimulated condition, the perceived time of movement was only slightly delayed (~75 ms) while signs of initiation of the movement (muscular activity) were more significantly delayed (~200 ms). Such data support the idea that awareness of initiating a movement is not derived from any sensory signals arising from the moving limb (because such signals are not available until after the limb has started moving). Rather awareness appears to be linked, at least in part, to some signal that precedes the movement. Additional behavioral and neuroimaging experiments by this group have shown that sense of agency is also able to be modulated by subliminal priming maneuvers that affect the fluency of action selection processes (i.e., the sense that something may or may not be "wrong" when selecting or following a course of action). Thus, at least under certain circumstances, individuals' sense of agency may be more an effect of such prospective signals than an effect of the consequences of the movements themselves.

As previously noted, prediction signals are crucial in the nervous systems of organisms that are capable of movement. For human beings, who are capable of both voluntary and involuntary movement, being able to compare signals that predict the likely effects of commands arising from the motor system with sensory information acquired during movement is particularly important for adjusting movements easily and checking whether movements are being completed as planned. That is, such a system includes both feed-forward (prediction) and feedback (signal comparison) components. Haggard and colleagues (2005) suggest, however, that even earlier prospective signals involving action *selection* (and action programming) may participate along with these prediction signals and signal comparisons in acting as cues that have a role in the experience of agency. Additional research will be required to further clarify the details of these processes and the circumstances under which each may have a particularly important role.

Self-Transcendence: Nondualistic Notions of Self and Spiritual Self-Development

The ultimate experience in being human is often described by the empowerment that is experienced in connection with some force greater than our individual self. The self is thus transcended and no longer conceptually or dualistically represented in space and time. In *Varieties of Religious Experience* (1902), William James wrote, “we can experience union with something larger than ourselves and in that union find our greatest peace.” Some models of self that are rooted in a developmental perspective have focused on this uniquely human characteristic underlying self-transcendence at the top of a hierarchical structure and part of personal growth.

Self-transcendence has been used to describe contextually similar phenomenology, arising spontaneously, in connection with religious or other culturally embedded rituals, as part of meditation, yoga, or other mind–body practice, and in pharmacologically induced states such as those influenced by psychedelic pharmacy. In addition to James’ description of self-transcendence as a “religious” or “mystical” experience, self-transcendence has been operationalized in a myriad of ways, including hypoegoic states, transpersonal experiences, unitary experiences, spiritual experiences, sacred moments, epiphanic experiences, ecstatic experiences, anomalous experiences, nondual awareness, near-death experiences, persistent nonsymbolic experiences, among others. David Yaden and colleagues (2017) define self-transcendent experiences (STEs) as experiences of unity with other people or one’s surroundings, involving the dissolution of boundaries between the sense of self and “other.” Such transient mental states of decreased self-salience involve a positive relationship between self and other that transcends self-focused needs, increases prosocial dispositions (e.g., compassion, empathy, and forgiveness), and increases the frequency of nondual (subject-object) phenomenological experience. Schoenberg and Vago (2019) further elucidate states and stages of self-development common to training across styles of meditation and other mind–body practices, including yoga breathwork, bodywork, and movement-based practices. Briefly, the first stage is “neurovisceral integration” in which autonomic, attentional, and affective systems are regulated through improved inhibitory tone. A continuum of contemplative exercises that focus attention on cardiorespiratory regulation and musculoskeletal strengthening facilitate the relaxation response to arise more readily, improve embodied awareness of interoceptive experience, and gradually improve the capacity for strengthening inhibitory processes across systems. This stage of improved parasympathetic control through integration of body awareness with body movement is essential for preparing the mind for sustained concentration. The second stage is described as “concentration practices” characterized by training the mind with stable concentration and/or flexible open, receptive attention. Inhibitory control processes engage negative feedback circuits that interrupt ongoing behavior. Inhibitory circuits and coordination with frontoparietal networks facilitate rapid toggling between resources supporting conceptual, sensory, and nonconscious processes driving contextually specified demands of the situation. The next two stages of “insight” and “nondual” practices serve to integrate more abstract contemplative practices for ordinary and extraordinary insight into the nature of self, mental habits, and self-transcendence. The ultimate stage termed “unified compassion practices” describes a concrete understanding of ultimate reality where one’s formal practice transforms into everyday living, integrating the preceding tiers and insights. This is described as the highest state and stage of development where altruistic motivations for compassionate action, service to others, and unconditional love are more readily experienced. This state is similarly described in contemplative pedagogy as psychological fulfillment characterized by an “unconditional compassion or love for the world, joined with perfect insight or wisdom.”

The hypothesis here, common to many contemplative spiritual traditions, is that normal waking consciousness is a construction that we “live through” that is obfuscating with respect to the world as it truly is. That is, there is a gap between “the world as we construct and perceive it” and the “world as it truly is.” The spiritual path, from a contemplative or pharmacologically induced vantage point, is a means of reducing this gap by “cleansing the doors of perception” such that we can “see the world as it is and know it again for the first time.” While both contemplative, mind–body practices and psychedelics may disrupt self-consciousness and underlying neural processes in a phenomenologically similar fashion, there are unique differences reported, and each experience of self-transcendence is likely to involve unique mechanisms.

Psychedelics

Psychedelics are a pharmacologic technology with a mixed history, venerated as sacred medicines and vilified as narcotics with no medicinal value; yet, the profound effects on consciousness and self-representation are very well documented.

Psychedelic compounds fall into two general categories: tryptamines and phenethylamines. Lysergic acid diethylamide (LSD), psilocybin, and its metabolite psilocin are classic tryptamines. *N*,*N*-dimethyltryptamine (DMT), 5-methoxy-DMT (5-MeO-DMT), ibogaine, and bufotenine are also examples of tryptamine compounds. Phenethylamine hallucinogens are amphetamine analogs such as 2,5-Dimethoxy-4-iodoamphetamine (DOI) and 2,5-Dimethoxy-4-methylamphetamine (DOM, also known as "serenity, tranquility, and peace" or STP), 2,5-dimethoxy-4-bromophenethylamine (2C-B), and mescaline. The two classes of psychedelics are structurally distinct but produce hallucinations and other psychedelic effects via serotonin-2A receptor stimulation. They are distinct structurally and mechanistically from other drugs that are often also considered hallucinogens, such as cannabinoids, N-Methyl-D-Aspartate (NMDA) glutamate receptor antagonists (ketamine, phencyclidine [PCP], dextrorphan, nitrous oxide), and kappa-opiate receptor agonists (salvinorin-A). Although structurally related, the phenethylamine hallucinogens are structurally and mechanistically distinct from the so-called "entactogenic" drugs like 3,4-Methylenedioxymethamphetamine (MDMA), which produce stimulant effects, feelings of relatedness, and elevated mood but do not produce hallucinations. Although the neurophysiology is still being characterized, it is clear that most of the classic psychedelic tryptamine compounds are agonists or partial agonists at brain serotonin 5-hydroxytryptamine 2A (5-HT_{2A}) receptors widely distributed across cortical networks, with particular importance on those expressed on apical dendrites of layer 5 pyramidal cells. There is considerable evidence demonstrating correlations between 5-HT_{2A} receptor occupancy across the FPCN-DMN and subjective intensity of psychoactive effects. Much of the self-transformative effects of psychedelics are proposed to be due to the increased excitability and neuroplastic action of 5-HT_{2A} activation, including dendritic arborization, complexity, spine growth, synapse formation, increases in brain-derived neurotrophic factor, and an increased rate of learning/extinction. Other pharmacodynamics have been identified, including activation of 5-HT_1 and 5-HT_{2C} receptors, increases in glutamate availability, agonistic and antagonistic activity at the D1 and D2 dopamine receptors, anticholinergic, kappa-opioid agonists, and activation of tryptophan metabolic pathways in the gastrointestinal tract.

In 1938, while working at Sandoz laboratories, Albert Hofman synthesized the first tryptamine psychedelic, LSD, from the ergot fungus and later reported on its profound psychological effects in 1943. Shortly thereafter, the Swiss pharmaceutical company Sandoz chemicals and the National Institute of Mental Health began researching and supporting the promising therapeutic effects of LSD and the similar serotonergic psychedelic, psilocybin, as Delysid and Indocybin, respectively. These pharmaceuticals were investigated as psychiatric treatments for a variety of diseases, including end-of-life anxiety, traumatic stress, depression, and alcoholism. In 1952, Charles Savage published the first study using LSD for the treatment of depression. The word “psychedelic” was first coined by psychiatrist Humphrey Osmond in a New York Academy of Sciences meeting in 1957, drawing from the Greek words for “mind” or “soul” (psychē) and “to manifest” (delos), suggesting these drugs make manifest latent aspects of the mind. During the period of time between the 1950s and mid-1960s, approximately 1,000 clinical research studies were published on the benefits of psychedelics in psychiatric treatment, and a number of other therapeutic mechanisms were proposed. For example, one popular psychological theory described that psychedelics may weaken psychological defenses, allowing personal conflicts to surface and be worked out with a therapist. The dynamics underlying psychological processes experienced in the context of psychedelic psychotherapy were well documented by two prominent psychiatrists, Hanscar Leuner (1962) and Stanislav Grof (1972). They suggested ego-dissolution as key to the psychodynamic therapeutic process through integrating increased sensory and basic emotional activity with regression of ego functions as they emerged through autobiographic memories and autoproduction of symbolic imagery. Grof professed that the use of psychedelics in proper treatment contexts can be a revolutionary tool for psychiatry, similar to what the microscope is for biology and medicine or the telescope is for astronomy. Savage (1962) also proposed that psychedelics allowed the individual to let go of the ego, with or without volition, and experience intense existential or peak moments that facilitate long-term positive effects on behavior and attitude. One must also consider how this sense of ego or consciousness dissolution relates to similar phenomenology found in brain-lesioned patients or as might occur with meditation. Evidence for the temporary suspension of effortful cognition or spontaneous thought brought on through psychedelics or meditation may prove to be similar to the phenomenology experienced through unique agnosias or lesion-induced deficits such as athyhormia, a disorder of motivation. Initial studies have begun to explore the use of psychedelics as potential therapeutic aids for brain damage and dysfunction; yet, distinctions between these states have yet to be clarified.

Due to a combination of factors, including social stigma, poor translation of benefits, and recreational use and misuse by high-profile researchers, psychedelics became an illicit substance by 1971. After decades of prohibition, growing failures of conventional mental health care, pioneering work of the Multidisciplinary Association for Psychedelic Science (MAPS), and research from a few labs at prominent universities, a new wave of psychedelic research emerged (See Nutt & Carharhart-Harris, 2021). Models for how psychedelics function and contribute to mental health are currently evolving and describing self-related insights uniquely induced by specific chemical variants of hallucinogenic pharmacology, with targeted dose-related effects and integration with psychotherapy and other psychiatric treatment modalities. Some of the most prominent theories supporting clinical outcomes focus on increased 5-HT_{2A} receptor excitability and plasticity, increased integration, and decreased modularization of brain networks. Most recently, a randomized controlled trial investigating psilocybin versus escitalopram for treatment-resistant depression found the antidepressant response to psilocybin was rapid, sustained, and correlated with decreased DMN and FPCN modularity.

Many leaders in psychiatry now value the therapeutic potential of psychedelics in proper psychotherapeutic contexts and with proper support and training for providers. New protocols that mix psychotherapy with supervised psychedelic treatment are at various stages of randomized clinical trials. Yet, there is some cautious criticism of the growth in the industry. There are now more than 50 publicly traded companies related to the development of the administration of psychedelic-like drugs in the United States alone, with a market of $2 billion in 2020 to projected growth of over $10 billion by 2027. With the expansion of research, large growth in the psychedelic industry, and deregulation at federal levels, global brands promise a psychiatric revolution in treatment and mental health outcomes over the next decade. As of April 2021, there are more than 80 psychedelic clinical trials registered through ClinicalTrials.gov investigating either psilocybin, LSD, or MDMA with clinical psychiatric patients. These studies include treatment for anorexia, headache, obsessive-compulsive disorder, caregiver burnout, PTSD, depression, and addiction. Dosage titration has been key to this process, and data are emerging in support of such protocols across the neuropsychiatric spectrum. The “psycholytic” method uses lower doses and frequent sessions to enhance the standard psychotherapeutic process. Low-dose psychedelics, also referred to as “microdosing,” are beginning to show potential for well-being enhancement in nonclinical samples. The “psychedelic” method uses higher doses in fewer sessions to induce a mystical experience and moments of intense catharsis, enabling participants to work through and integrate difficult feelings and situations, thereby reducing anxiety and depression. As therapeutic mechanisms emerge with more clarity, and newer protocols also emerge with specific cocktails of tryptamine or phenethylamine psychedelics, the self and how it may be transformed through a neuropsychiatric lens will clearly be impacted.

FUTURE DIRECTIONS

Through growing interdisciplinary approaches, the science supporting the nature of the self and its clinical relevance to psychiatry is developing rapidly. Future translational science will leverage the results of studies emerging from neuroscience, computational and cognitive science, along with the insights from other diverse disciplines. Conceptual issues are discussed here from both historical and theoretical perspectives to inform our current understanding of the self using evolutionary, neurobiologic, and clinical perspectives. Representative scientific work provides some of the conceptual convergence in this scientific enterprise with its diverse moves toward more dynamical and multidimensional considerations of a socially and otherwise “systemically” embedded self, along with signs of where a more advanced program for a science of self might be heading.

Novel psychotherapeutic approaches introducing mindfulness, psychedelic drugs, and plant-based entheogens in appropriate therapeutic contexts provide the opportunity to advance both the science of mental health and the therapeutic relevance of the mind in and of itself. Mindfulness-oriented therapies include a focus on context, processes, and functions by which a person relates to their internal experience rather than focusing on the content of those thoughts. Mindfulness training is being used in the context of a new wave of behavioral and cognitive psychotherapy (e.g., mindfulness-based cognitive therapy) to improve clinical outcomes (e.g., depression, anxiety, pain). These new mindfulness-oriented therapies (e.g., mindfulness-based cognitive therapy, acceptance and commitment therapy, and mindfulness-oriented recovery enhancement) are referred to as third-wave therapies, supporting an evolution of traditional cognitive behavioral treatments. Psychedelic medicine is demonstrating effectiveness in augmenting the therapeutic process, offering direct insight and STEs in which to resolve maladaptive mental habits and integrate the dissolution of self-other boundaries that may improve prosocial relations with others. Yet, major challenges remain, such as scaling treatment approaches given the intensive time invested in therapist-guided treatment and the cost-prohibitive issues associated with the use of medicine that remains illegal under most international conventions. Augmentation approaches that utilize both psychedelics and mindfulness in psychiatric contexts are likely to leverage a science of self for therapeutic purposes. One recent review describes the potential synergistic effects psychedelics and mindfulness-based interventions may have through common contextual, methodologic, experiential, neurobiologic, and clinical outcomes, as well as an emphasis on insight into the nature of the self.

Brain–computer interfaces (BCIs) are promising technologies for extending the self, supplementing and augmenting cognition, remote communication and control, and social integration with games and entertainment. A powerful communication tool between users and systems, BCI promises to transcend boundaries of self-other relations, improve physical limitations found in neurologic disorders or deficits, and will likely provide a research platform from which to better understand consciousness and self-representation. Brain stimulation and physiologic sensor technology are transforming health care and the limitations of our own biology. Altering brain network dynamics and peripheral physiology through repetitive transcranial magnetic stimulation, transcranial direct current stimulation, focused ultrasound, and vagus nerve stimulation will continue to improve therapeutic opportunities and change our perception and limitations of self-functioning.

Neuroimaging and big data science methods for assessing anatomical and functional connectivity now exist in data-rich and open science collaborative networks, but they are continuing to undergo development, particularly in terms of the speed and accuracy with which the imaging data may be acquired and processed and how patterns are identified to predict outcomes through the use of machine learning analytics. Researchers and computational neuroscientists are also beginning to use this kind of data to assess patterns of connectivity disturbance and network breakdown in various developmental, degenerative, and mental disorders during passive resting states. These studies are providing preliminary markers for what constitutes robust versus disordered patterns of brain network activity, which also have implications for our understanding of neural (information) processing and self-representation. These brain network and other brain dynamical approaches are highly sophisticated and currently relatively labor-intensive techniques, however, so their ramifications for furthering the understanding of self-processing remain in their infancy.

So, while understanding “the self” in all the various and nuanced ways in which that term is used provides challenges, continued pursuit of the scientific agenda on self-processing within the limits and with some of the caveats outlined in this chapter holds promise for continued progress to enrich the understanding of mechanisms underlying many psychiatric disorders. The next generation of neuropsychiatric studies will rely heavily on clarifying some of the pertinent issues that have been identified as limitations and on new emerging technologies and methods that capture dynamic complex neural dynamics and associated physiologic and embodied phenomena underlying the nature of suffering for the human mind and some of its remediation. There is clearly increasing emphasis on mechanistic and therapeutic considerations of self-representation and associated processing for psychiatry, including insight into the experiential, autobiographical, introspective, and integrative aspects of self-experience from the third-person perspective of naturalistic science and through the phenomenology of first-person experience. Indeed, adaptive self-transformative processes across contexts of human experience may be critical to the therapeutic lens by which psychiatric treatment advances. As methods and current technologies improve, progress will be made in identifying potential biologically based diagnostic and therapeutic targets and strategies for those suffering from mental illness and clarifying our understanding related to the nature of mind and consciousness.

Further Readings

Apps MAJ, Tsakiris M. The free-energy self: a predictive coding account of self-recognition. *Neurosci Biobehav Rev*. 2014;41:85–97.

Baars BJ. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. In Laureys S, ed. *Progress in Brain Research*. Vol. 150. Elsevier; 2005:45–53 https://doi.org/https://doi.org/10.1016/S0079-6123(05)50004-9

Beck AT, Rush AJ, Shaw BF, Emery G. *Cognitive Therapy of depression.* Guilford Press; 1979.

Blakemore S-J, Wolpert DM, Frith CD. Central cancellation of self-produced tickle sensation. *Nat Neurosci*. 1998;1(7):635–640.

Churchland PS. Self-representation in nervous systems. *Science*. 2002*;*296:308–310.

Decety J, Chaminade T. When the self represents the other: a new cognitive neuroscience view on psychological identification. *Conscious Cogn*. 2003;12:577–596.

Dennett D. The self as a center of narrative gravity. In: Kessel FS, Cole PM, Johnson DL, eds. *Self and Consciousness: Multiple Perspectives*. Lawrence Earlbaum Associates; 1992.

Erikson EH. *Identity: Youth and Crisis*. Norton & Co; 1968.

Farb NAS, Segal ZV, Mayberg H, et al. Attending to the present: mindfulness meditation reveals distinct neural modes of self-reference. *Soc Cogn Affect Neurosci*. 2007;2(4):313–322.

Friston K. Am I self-conscious? (or does self-organization entail self-consciousness?). *Front Psychol*. 2018;9:579.

Gallagher S. A pattern theory of self. *Front Hum Neurosci*. 2013;7:443.

Gallup GG Jr. Self-awareness and the emergence of mind in primates. *Am J Primatol*. 1982;2:237–248.

Grof S. Varieties of transpersonal experiences: observations from LSD psychotherapy. *The Journal of Transpersonal Psychology*. 1972;4:45.

Gusnard DA. Being a self: considerations from functional imaging. *Conscious Cogn*. 2005;14:679–697.

Kohut H. The psychoanalytic curriculum. *J Am Psychoanal Assoc.* 1962;10:153-163.

Lashley KS. *Brain Mechanisms and Intelligence: A Quantitative Study of Injuries to the Brain*. 1929.

Legrand D, Ruby P. What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychol Rev*. 2009;116(1):252–282.

Libet B, Gleason CA, Wright EW, Pearl DK. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*. 1983;106(Pt 3):623–642.

Luhmann N. *Social Systems*. Stanford University Press; 1995.

Maturana HR, Varela FJ. *Autopoiesis and Cognition: The Realization of the Living*. Springer Netherlands; 2012.

Metzinger T. The self-model theory of subjectivity: a brief summary with examples. *Humana Mente—Q J Philos*. 2010;14:25–53.

Nutt D, Carhart-Harris R. The current status of psychedelics in psychiatry. *JAMA Psychiatry*. 2021;78:121–122.

Roeser R, Peck SC. An education in awareness: self, motivation, and self-regulated learning in contemplative perspective. *Educ Psychol*. 2009;44:119–136.

Ruby P, Decety J. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat Neurosci*. 2001;4:546–550.

Rudrauf D, Bennequin D, Granic I, Landini G, Friston K, Williford K. A mathematical model of embodied consciousness. *J Theor Biol*. 2017;428:106–131.

Seth AK, Friston KJ. Active interoceptive inference and the emotional brain. *Philos Trans R Soc Lond B Biol Sci*. 2016;371(1708):20160007.

Singer T, Lamm C. The social neuroscience of empathy. *Ann N Y Acad Sci*. 2009;1156:81–96.

Smallwood J, Schooler JW. The science of mind wandering: empirically navigating the stream of consciousness. *Annu Rev Psychol*. 2015;66:487–518.

Strawson G. The phenomenology and ontology of the self. In: Zahavi D, ed. *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-Experience*. John Benjamins; 2000.

Tulving E. *Elements of episodic memory*. Oxford University Press; 1983.

Vago DR, Silbersweig DA. Self-awareness, self-regulation, and self-transcendence (S-ART): a framework for understanding the neurobiological mechanisms of mindfulness. *Front Hum Neurosci*. 2012;6:296.

Varela FG. Self-consciousness: adaptation or epiphenomenon? *Stud Gen (Berl)*. 1971;24:426–439.

Varela FJ, Thompson E, Rosch E. *The Embodied Mind, Revised Edition: Cognitive Science and Human Experience*. MIT Press; 2017.

Yaden DB, Haidt J, Hood RW Jr, Vago DR, Newberg AB. The varieties of self-transcendent experience. *Rev Gen Psychol*. 2017;21:143–160.

Zahavi D. Phenomenology of self. *The Self in Neuroscience and Psychiatry*. 2003;56:C75.

1.30 Learning Theory and Psychosis, Anxiety, and Addiction

Philip R. Corlett, Ph.D.

Associative learning theorists concern themselves with how humans and other animals learn and use information about the relationships between events. Ivan Pavlov is perhaps the most famous learning theorist; his empirical studies of anticipatory learning about food rewards in dogs gave us a framework to examine how relationships between environmental stimuli (lights, tones) and salient events (food, electric shocks) are acquired. In a curious mistranslation from his original Russian, we call this process conditioning and the cues conditioned stimuli (they ought to have been conditional, as the salient events were conditional upon those cues).

Conditioning has obvious relevance to psychiatry (which is broadly considered as the study and treatment of mental illness, emotional disturbance, and abnormal behavior), given that conditioning guides behavior and emotions—abnormal conditioning may portend emotional disturbance, abnormal behavior, and thus mental illness. Anxiety and addiction have been studied most extensively from a learning theory perspective. More recently, psychosis has been likewise investigated in learning theory terms. Given that learning theory is inherently translational studying mental illnesses from this perspective may portend a translational or reductionist understanding of mental illness. We will review the impact of learning theory on our understanding of psychosis, addiction, and anxiety. Furthermore, we shall conclude by highlighting how computational neuroscience and, thereby, computational psychiatry are grounded in learning theory.

Pavlov himself appreciated the role of conditioning in mental illness. Toward the end of his career, he sought to explain hysteria and neuroses in conditioning terms. He attributed emotional perturbations to dysfunctions in the low-level first signaling system that dealt with unconscious, preverbal associations. The verbal, second signaling system Pavlov considered to be unique to humans, and he asserted that dysfunction in the second signaling system led to psychosis. This distinction—between simple associations and complex propositions continues to be a focus of contemporary learning theory, and it may have relevance to mental illness. In the first part of this chapter, we will review the history of associative learning and learning theory addressing the development and history of learning theory.

A BRIEF HISTORY OF LEARNING THEORIES

Learning theory is sometimes referred to as associative learning theory, its currency being associations between mental concepts, although this is the subject of ongoing debate. Proponents of learning theory are sometimes called associationists.

Associationism began with Plato. Aristotle developed the concept and outlined the first associative learning theory by expounding the laws of association: **The law of contiguity**; things that occur close to each other in space or time are associated in mind. **The law of frequency**; the more often two things occur together, the more strongly they will be associated. If you think of Oliver Hardy, it is rather difficult not to think of Stan Laurel. **The law of similarity;** thoughts about one thing will trigger thoughts of similar things. **The law of contrast**, on the other hand, recollecting something will also trigger its complete opposite. When thinking of the tallest person you know, you may suddenly recall the shortest one as well.

Although John Locke is considered the first to discuss the role of improper association of ideas in mental illness, David Hartley believed that the principles of association were the fundamental operations of cognition and a means through which to relate psychology and physiology. In a neural doctrine inspired by Newton, Hartley argued that neural vibrations passing from the representation of one idea to another formed the physiologic basis for their association (preempting the contemporary Hebbian notion of neurons firing together and wiring together). Hume added cause and effect (contiguity in time) as a law of association. British empiricist philosophers rejected Descartes's notion of innate faculties. At birth, they thought of the mind as a tabula rasa—a blank slate—that accumulated ideas via sensory experience and built more complex ideas from simpler ones through the laws of association.

Pavlov was perhaps the first experimental associationist. He explored the physiologic and psychological mechanisms of association. He developed paradigms for presenting associations to experimental animals and tracking their learning, as well as theories for explaining how the cortex and subcortex underpinned that learning. Pavlov's work provided a framework for broad experimental exploration, some of which indicated that mere contiguity was not sufficient for learning. For example, Leon Kamin blocking effect. When a novel cue (e.g., a tone) is paired with a stimulus (e.g., a light) that already predicts an outcome (electric shock)—the pretrained cue—the light–"blocks" new learning about the novel cue—the tone; clearly a problem for contiguity-based theories of association.

Blocking requires that associative learning is sensitive to surprise—the degree of predictability of one associate to another modulates the formation of new associations. Hebb and Hartley's physiologic explanations of learning are driven by contiguity and hence cannot explain blocking. It seems error correction– earning from mistakes–is crucial to the formation and updating of associations. For example, an algorithm was developed to solve engineering problems using a simple connectionist neural network learned via error correction; the networks represented their inputs and outputs across distributed nodes. Associations or connections were formed and strengthened between nodes. The weights of connections between particular nodes were updated using the mismatch between the desired output from a given input and the output that actually occurred. By reducing this error signal, the network's responses to particular inputs improved—the network learned associations.

A similar algorithm was proposed to explain conditioning in rodents by Robert Rescorla and Allan Wagner. Here, environmental stimuli engender a set of expectations about potential subsequent states of the world by exciting representations of those states (e.g., the availability of food or the presence of a predator). Any mismatch, or prediction error, between those expectancies and actual experience, is used as a teaching signal to update future expectancies by changing the associative strength. They explain Kamin blocking in terms of a lack of prediction error. The outcome is already completely predicted by the pretrained cue. When the novel cue is added, there is no prediction error and hence no new learning.

Attention is also critical for associative learning. Associationists theorize that attention changes associability—the readiness with which an event or its representation (cue, outcome) can enter into associative relationships. Some theorists argue that cues that are predictably associated with important outcomes are more associable. On the other hand, stimuli with an uncertain predictive history also garner attention. Hence, cues whose outcomes were surprising in the past are allocated more attention and learned more readily. These theories are incompatible, but they are both supported by behavioral evidence. Attention must be important to association formation in different ways under different circumstances. Theories have reconciled these observations by positing a role for both confirmation and violation of expectancy in associative learning; the occurrence of an event that matches an active expectancy amplifies its representation in short-term memory, increasing the likelihood that it will be consolidated within long-term memory. By contrast, when an unexpected event violates the active expectancy, an orienting system is activated, which resets short-term memory (dropping active expectancies) and engages an orienting response, permitting the acquisition of new explanatory associations. Using both of these mechanisms, organisms learn associations between stimuli, events, thoughts, and percepts to build an internal model of their environment from which to make predictions.

Note the increasing cognitive complexity of the models beyond the formation of simple links between representations. This has been the cause for some to reject any role of associations in predictive learning and the formation of causal beliefs. Many such critics favor a propositional alternative in which directional links between concepts are formed and used to make predictive and causal inferences. Indeed, some comparative psychologists imbue the humble laboratory rat with rational inferential abilities. For some, the causal models used in inferential learning are innately specified. The British empiricist philosophers and later the behaviorists were opposed to any innate endowment; indeed, the demise of behaviorism was confirmed by Chomsky’s scathing critique of Skinner’s behaviorist analysis of language (and behaviorism in general) that denied any innate linguistic representations. However, across species, there are clearly certain associations that are easier to learn, some of which appear to be evolutionarily prepared and endowed upon young organisms from birth. One can conceive of mechanisms through which preparedness could be learned very early on through environmental interaction. Much more probable is the sort of neo-constructivist approach whereby a set of genetically conferred allowable connections is endowed, and then the weights of those connections are influenced by sensory experience.

We suggest that the principles of contiguity, similarity, and prediction error drive association formation and, like the British empiricists, more complex cognitive structures emerge by combining these simple representational units. There is some evidence for such an approach in human causal learning; subjects learn new causal models (propositional knowledge structures that they use to learn about causes) by minimizing prediction errors within and across possible causal models, that is, associative links can be used to form more complex causal representations and the principles of associative learning (namely minimizing prediction error) may aid decisions about which structures to impose on an inference task. We note, too, that simple connectionist architectures in which Widrow–Hoff/Rescorla–Wagner-like error-correcting mechanisms are embedded often demonstrate rather complex emergent properties akin to reasoning and inference. In some of the earliest computational models of cognition, McCulloch and Pitts elaborated upon Gottfried Leibniz, the 17th-century philosopher who attempted to create an alphabet of human thought, each letter of which represented a concept and could be combined and manipulated according to a set of logical rules. Their building block was the proposition—the simplest possible statement, either true or false. From there, they employed the fundamental operations of logic, like the conjunction (“and”), disjunction (“or”), and negation (“not”), to link propositions into increasingly complicated networks. McCulloch–Pitts networks have also been combined with Widrow–Hoff error correction in a foreshadowing of propositional causal model theory.

Causal model theories also highlight an important new direction in cognitive science, the rise of Bayesianism. Thomas Bayes was a British mathematician whose doctrine of probabilities (published posthumously in 1873) has had a striking impact on science in recent years. The theorem embodies a formal approach to reasoning about data using hypotheses and captures the probabilistic nature of many of the tasks faced by organisms; we believe a Bayesian approach to the mind and brain might bridge David Marr’s levels of analysis and achieve what Fodor thought to be impossible, a consilient understanding of cognition and its implementation in the brain. To wit, the problem faced by an organism is to predict its environment and respond appropriately. The organism’s brain does this by minimizing its uncertainty about subsequent inputs by forming and maintaining a set of prior expectancies (predictive associations) about incoming stimulation.

Such prediction error minimization is even manifest in the signaling properties of single neurons that shuttle ion channels and receptors into and out of their membranes to modulate the cell's receptivity to incident stimuli. It is also embedded in the hierarchical neuroanatomy of the brain, in which stimuli enter through sense organs, and their neural representations are communicated through the brain. Top-down expectations are communicated from areas with more abstract representations downward through the hierarchy. Any mismatch between prior expectation and current input (prediction error) needs to be dealt with. It is either canceled by the top-down expectancy or propagated and used to update subsequent expectations with new learning. The neuroanatomy and neurochemistry of backward and forward connections across cortical layers are exquisitely suited to this approach. Such a Bayesian analysis of the neurobiology of vision predicts various qualities of receptive fields in the visual cortex, such as "end stopping." Even the retinal protein signaling cascades involved in signaling light incident on the eye have been subjected to Bayesian analysis with favorable results.

The task of visual perception, conceived of as unconscious inference about current inputs based on a world model, embodies similar Bayesian principles. Indeed, the various visual illusions to which we are susceptible profoundly illustrate the overriding role of prior visual experience and expectancies on current perception. Pavlov believed that Helmholtz's unconscious inferences were aligned with his conditioned responses, an assertion realized at the algorithmic level by the use of similar Bayesian formalisms to explain Kamin blocking, perceptual learning, and visual perception. Indeed, color-after effects in the visual system appear to be influenced to some degree by selective learning. They are subject to Kamin blocking, underlining the crucial role of learning across cognitive, perceptual, and neural systems.

Taken together, these theories conceive of perception, cognition, learning, action, and belief as processes of inference about latent causes, latent causes that unite environmental stimuli and neural responses, cues and outcomes, choices and rewards. These latent causes form the basis of the associations on which associative learning theories are and, by extension, minds are predicated. Disruptions to these processes of inference can lead to perturbed world models, percepts, beliefs, and actions. In the remainder of this chapter, we shall use learning theory to explain the symptoms of psychosis, addiction, and anxiety.

Psychosis

Psychotic symptoms involve a profound departure from consensual reality, manifest as perceptual experiences (hallucinations) and beliefs (delusions) that others in one's culture do not share. We will address one of the cardinal delusions, which occurs both after neurologic injury and as a symptom of psychiatric illness. The Capgras delusion, originally termed the illusion of doubles is the belief that familiar individuals (usually family members) have been replaced by imposters. While the individual confronting the sufferer looks very much like the familiar person, they are described as subtly different in appearance, and crucially, the sufferer reports lacking the feeling of emotional recognition that characterizes our experience of those familiar to us.

First, we will explain delusions using learning theory as a theoretical framework that bridges mind and brain, perception and belief. The unit through which this bridging takes place is the association. The following processes are central to our general explanation of delusions; first, the notion of expectation and that, through experience, we learn to expect a certain set of circumstances. Second, the consequences when these expectancies are violated are either to ignore or discard the dissonant experience or to embrace it and incorporate it into our world model, engendering an updated set of expectancies. That is, when confronted with someone whose features very much resemble a loved one, we expect to feel a sense of familiarity. The absence of such a feeling induces a prediction error that demands explanation.

Absent but expected events are crucial in the phenomenology of psychosis. They contribute greatly to the feeling of delusional mood or atmosphere that attends the formative stages of delusions. In learning theory, too, absent but expected events garner new learning. Indeed, events that involve updating a causal belief about an absent but expected event engender a prediction error brain signal in the same circuit engaged when we learn beliefs about events that were actually experienced.

This error signal guides the Capgras sufferer toward a particular conclusion; that a replica has replaced their loved one. Our learning model challenges the strict distinction between perception and belief and, therefore, the necessity for two factors to explain delusions.

The prediction error model provides little insight into the exact form that the new learning will take; the particular beliefs it will reinforce. The simplest associative learning models argue that the belief formed will track the evidence experienced; that the weight given to a particular belief will be proportional to the prediction error that it minimizes. Nearly all delusions are socially relevant: they are ideas about oneself and one's relationships with others. Their content is crucially related to the individual's specific fears, needs, or security.

Also important are the particulars of past and present experiences, including personal and cultural context. Associations made in early childhood are likely critical to the specific content of delusions, though in many cases, those experiences will be difficult to identify. A recent review of associative learning mechanisms of synesthesia (associations across sensory modalities, e.g., people for whom "A" is red): though some of those people recall a red A in a much-loved set of letter magnets, for many others, it is reasonable to think that early childhood experience included this pairing. With regard to the cultural context, it was common during the cold war for persecutory delusions to involve KGB agents; now, they more commonly involve the CIA. Delusions also keep up with technology and popular culture: delusions of reference have evolved from worries about monitoring via radio and then satellite to worries about monitoring via the internet; five years ago, our patients were in The Matrix, and now they are in Transcendence associated with the delusional mood is perhaps best explained in terms of the Truman Show starring Jim Carey —in which the protagonist discovers that his life is a reality TV show and everything, every social interaction, and seemingly insignificant event, is arranged for him by a director and cast. Such experiences (perceiving others as actors wearing masks) predate the movie. However, it has become the framework through which many patients explain their delusions.

Historically, associative learning research has also failed to consider mental representations of causality, that is, going beyond how correlations are represented to try to understand how people make causal attributions. For example, how do people learn about and represent foods causing allergies or drugs causing side effects? Also, these causal relationships can be complex. For example, in the case of foods causing allergic reactions, do two foods that cause allergies interact to cause a stronger allergy? As described above, individuals bring a rich set of expectancies regarding causal interactions to bear on the causal relationships with which they are confronted and, in the presence of competing candidate causes, they make all kinds of inferences about those candidates and how they might interact. There may be minimization of prediction error within a causal model (when learning about particular cues) as well as across causal models (when forming beliefs about causal mechanisms) as beliefs form. When research subjects see implausible candidate mechanisms but strong covariation between candidate cause and effect, executive prefrontal brain regions are activated, including the left dorsolateral prefrontal cortex (DLPFC) and the anterior cingulate cortex (ACC). These regions have been shown to respond to prediction errors. A note about laterality: Devinsky has postulated, after Ramachandran and Blakeslee and according to our own data, that the right DLPFC is responsible for signaling when events have violated expectations and that the left DLPFC is responsible for generating explanations for those expectancy violations and producing the delusional narrative. In Fugelsang and Dunbar's data, the combination of an implausible theory and strong evidence in favor of that theory drives the "explanation generator" (left DLPFC) more robustly, as do delusional mood experiences during delusion formation.

One remarkable feature of delusional beliefs is their elasticity: they expand and morph to include new contradictory data. The person with Capgras, claiming their spouse is an imposter, might respond to other family members who greet the spouse warmly by saying, "*of course, they hug her [the impostor spouse]–they're in on it!"* This can be particularly difficult to understand, as the patients can often learn about other new things (they don't have an all-encompassing deficit in learning). How can associative learning, and in particular prediction error, account for delusion formation and maintenance account for elasticity?

Given their explanatory utility, delusions are rehearsed extensively. When delusions are questioned, *bringing them to mind* may actually serve to reinforce rather than disrupt the memory.

In summary, then, delusions form in response to aberrant prediction errors that imbue stimuli, thoughts, and percepts with inappropriate salience. Delusions formed to explain these experiences relieve the uncertainty associated with the prodromal "delusional mood." These beliefs are maintained in memory more strongly through cycles of recall, reactivation, and reconsolidation-based strengthening. Confronted with contradictory evidence, people are more likely to discount the evidence than relinquish the belief (whether or not it is a delusion).

In light of these challenges, we think it prudent to briefly review the burgeoning literature on belief updating in people with delusions and psychosis more broadly. This has, until recently, focused on potential biases and heuristic accounts of how people with delusional beliefs behave. For example, hasty decision-making, when confronted with evolving evidence, has been termed a jumping to conclusions (JTC) bias. Rigid anchoring of one's belief in the face of contradictory evidence is a BADE. The advent of formal computational modeling—wherein we write a generative model of participant beliefs and convolve the model with their behavior, estimating the model parameters and comparing them between groups (with and without delusions, for example). This work allows us to assess features of belief updating that may not be immediately visible from a heuristic stance. For example, JTC and promiscuous belief updating may be conceived of as a strong prior belief in environmental volatility—people with delusions (particularly paranoia) expect that reward learning tasks have an unstable generative process, and as such, they switch choices even after being reinforced. They also fail to update that volatility belief in light of the evidence they are experiencing—manifesting as a slower learning rate from volatility. Relatedly, in the perceptual domain, people with hallucinations and delusions are less sensitive to increases in volatility and thus stick with their prior perceptual beliefs—compared with healthy voice-hearers who have the string priors but can update them with alacrity when the task changes. Furthermore, even in the context of the same task, people with schizophrenia engage in "all or none" belief updating—they both jump to conclusions on some trials and fail to jump at all on others in the same session. The circumstances that lead to such updating and the mechanisms that govern are worthy of much more attention from theorists and experimentalists alike. The picture is much more complex than strong or weak priors, slow or fast updating, and it likely varies both with the task at hand, as well as the state and trait features of the participants. There is no brain region that stores our prior beliefs and no monolithic update mechanisms.

Hallucinations

Hallucinations (percepts without external stimulus, often auditory and verbal in patients with schizophrenia—auditory verbal hallucinations, AVH) may be explicable in terms of aberrant associative learning. Horace Barlow, after Helmholz, pointed out the similarities between Pavlovian conditioning and the unconscious inference processes that underpin perception.

Statistical regularities abound in the world around us, and many of them are actually or potentially important for our survival. Barlow suggested that perception also adapted to statistical regularities of the environment and that, indeed, all learning can be regarded as the internalization of environmental regularities, for it is driven by the statistically regular occurrence of reinforcement following particular sensory stimuli or self-initiated actions. Helmholtz argued that perception results from the interaction of apperception–the immediate impact of sensory messages–with remembered ideas resulting from past experience. He wrote of perceptions that *"… by their peculiar nature they may be classed as conclusions, inductive conclusions unconsciously formed.*" Thus he held the view that experience of the environment was internalized or remembered and provided the basis, together with current sensory messages, for the statistical conclusions—mostly valid—that constitute our current perception of the world. These conclusions constitute symbolic working models based on the associative structure of objects and events in the environment and are therefore expressions of environmental regularities. This is similar to Tolman's notion of cognitive maps. Can those models be misappropriated to give rise to perception without stimulus?

We highlight the "*conditioned hallucination*" research, published most extensively by Douglas Ellson in the 1940s. Here, subjects believed they were undergoing a test of auditory acuity: every time they saw a lightbulb illuminate, they would be presented with an auditory stimulus. After a series of pairings between tone and light, the light was presented. Subjects reported continuing to perceive the tone. Patients who hallucinate are more susceptible to this effect, suggesting a common mechanism between pathologic and conditioned hallucinations. At first, it seems conditioning theories might have some difficulty explaining the phenomenology of AVH—for example, why are they voices rather than other sounds? Theorists like Chris Frith have proposed self-monitoring theories whereby AVH arise as a result of misattribution of inner speech to an external source. Can we reconcile prediction error and conditioning with this notion of source monitoring?

In perhaps the paradigmatic example of computational neuroscience, Wolfram Schultz has demonstrated in a series of studies that the prediction error signals that drive conditioning are generated in midbrain dopamine neurons. The signals are communicated to the striatum and prefrontal cortex via glutamate corelease, where they have a role in ascribing salience to coincident stimuli, thoughts, and percepts and determining agency for important events. Peter Redgrave argued that the dopamine signals in VTA do respond to expectancy violations. However, their role in the behavior is one of agency attribution—did I cause that surprising event to happen with my actions and is it something I want to happen again? This is achieved by discerning whether the VTA prediction error was coincident with forward models of action, reverberating representations of intended actions in the striatum. When predictions and prediction errors are disrupted, agency for actions can be misattributed externally, leading to delusions of control (feeling that one's actions are under the control of an external agent). Since speech is a motor act, it too may be generated and attributed via prediction error mechanisms, and through similar mechanisms, inner speech may be externalized.

In terms of the Bayesian brain theory, slower neuromodulators like acetylcholine and dopamine gate the impact of a particular prediction error on new learning and future percepts. It has much in common with associability in earlier learning theories or attention in more common parlance. Hallucinations may arise top-down due to higher precision-weighting of predictions such that they shape expectations in a sort of listening attitude. Exactly how that expectation more commonly gives rise to voices than other sounds remains to be established. There are some intriguing findings from functional neuroimaging studies of patients who hear voices. First, in a predictive learning task, aberrant resting activity in the auditory cortex and thus weakened responses to unexpected speech were observed in patients with auditory hallucinations. Similarly, the connectivity between Wernicke's area and Broca areas is disrupted during inner speech perception in patients who hear voices. Specifically, there is reduced input from temporal to frontal language areas in schizophrenia patients with AVH, suggesting that Broca activity may be less constrained by perceptual information received from the temporal cortex that could, in turn, prime increased top-down efforts from Broca area. Indeed, it has been suggested that reduced perceptual input triggers top-down influences in perception. With regard to speech processing, an fMRI study showed that only Broca area (BA44) was activated by unintelligible speech presented at low signal-to-noise ratios, whereas an extended frontotemporal network (Wernicke areas) was active during intelligible speech at high signal-to-noise ratios. The agency mechanism hypothesized by Redgrave and Gurney has yet to be examined in the context of speech studies in patients with schizophrenia. However, there is evidence for auditory cortex prediction error dysfunction and evidence (in healthy controls) that striatal prediction errors can train perceptual expectations in the auditory cortex. Given the growing role that linguistic representations, like propositions, have in associative learning, the links between voice-hearing and conditioning bear further investigation.

The predictive processing account of psychosis posits that hallucinations—percepts without an external stimulus—arise when prior beliefs are overweighted relative to sensory evidence so that expectations override inputs and prediction errors are discarded. According to some models, this weighting is based on the relative precision of the incoming data relative to prior beliefs. If the incoming data are less precise (i.e., they have a higher inverse variance), then inference will conform to prior beliefs, and hallucinations will result. Empirical data support this hypothesis: Participants who hallucinate tend to evince overweighting of prior beliefs in experimental tasks. For example, they can be more readily conditioned to hallucinate an absent tone on the basis of previous conditioning that tones are predicted by salient visual checkerboard stimuli. Ongoing work aims to delineate the underlying mechanisms of prior overweighting in order to better deliver pharmacologic and psychological treatments for hallucinations.

Negative Symptoms

Under a conditioning account, the negative symptoms of schizophrenia (like anhedonia, amotivation, and alogia) might occur when predictions are unreliable and sensory data are noisy. This would lead to a state in which decisions are difficult, and ultimately, actions are fruitless. One can easily imagine how such a situation would lead, for example, to the apathy and a-motivation that are debilitating, strongly associated with poor functional outcomes, and particularly poorly treated with current pharmacotherapies. Furthermore, some current pharmacotherapies interact with the specification of reward predictions and prediction errors to exacerbate negative symptoms iatrogenically, which may explain some of the unpleasant side effects of these drugs like weight gain and sexual dysfunction.

Learning theory may further shed light on the relationships between psychotic symptoms. Philosophers and clinicians have challenged the notion that delusions are even real beliefs, equating them with “empty speech acts” without manifest conviction. That is, patients with delusions espouse their beliefs verbally but do not act upon them in a manner consistent with such fantastical or terrifying claims; for example, a patient may claim that their food is being poisoned, but they will nevertheless eat it. This is termed double bookkeeping. Of course, some patients do act on their delusions with disastrous consequences, both suicidal and homicidal. The cognitive neuropsychiatric account of delusions is committed to conceiving delusions as a broken form of belief, but belief nevertheless. One answer is that even those without delusions fail to act consistently on their beliefs. Perhaps a more satisfactory explanation is that motivation plays a crucial mediating role, dictating whether we act upon our beliefs or not. Given that patients who have delusions in the context of schizophrenia often have co-occurring negative symptoms, including motivational deficits, double bookkeeping may represent a confluence of delusions and negative symptoms.

ADDICTION

Laboratory animals will self-administer drugs and learn the cues and actions that predict drug access. Those cues drive subsequent drug-seeking, even after periods of abstinence. Thus, associative models of addiction in animals have face and construct validity for human drug abuse. However, clinical research determining how human subjects learn about drug-predicting stimuli is lacking. First, there are ethical considerations when administering drugs to human subjects. Furthermore, it is difficult to know exactly what might serve as a drug-predicting cue to a human addict, for example, the places in which drugs were procured or consumed, the sight and smell of drugs and drug paraphernalia, or even the friends with whom drugs have been taken. In one study, smokers recorded the place, situation, and feeling experienced during smoking.

Different contexts were associated with differing degrees of craving and enjoyment.

But how does this learning take place? Thus far, with a few notable exceptions, clinical research into cue processing in addiction has focused on exposing drug users to generic cues associated with drug use, such as stimuli from the International Affective Pictures System or dramatized video representations of cocaine consumption. However, conditioning histories are unique to each subject, and hence, these standard pictures will vary in their relevance to subjects (for example, a cocaine user may not snort their cocaine from a mirror). Nevertheless, stimuli related to the abused substance evoke neural activation in the striatum of smokers, alcoholics, and cocaine users, exactly as one would predict from the preclinical literature and from an aberrant learning and memory model of addiction (note, however, that one should be careful of making reverse inferences with functional imaging data; merely observing striatal activation in a subject does not mean that the subject was necessarily experiencing reward, motivation or engaging in learning).

Positron emission tomography (PET) scans with radiolabelled tracers sensitive to dopamine binding reveal dopamine release in the dorsal striatum in cocaine users in response to cocaine-associated cues. Alcoholics too release dopamine in the striatum in response to alcohol cues, particularly in situations when alcohol was expected but not delivered. Such expectancy violations or prediction errors may provide a strong impetus to drug-seeking and may drive new learning of Pavlovian cue-drug associations and the instrumental actions involved in drug procurement.

Learning to predict important environmental events and therefore behave adaptively toward natural rewards, aversive experiences, and other salient shifts in environmental contingencies are highly conserved across phylogeny. Drugs and drug-predicting stimuli appear to usurp the systems responsible for such adaptive anticipation, driving behavior toward anticipation and consumption of drug rewards and biasing memory consolidation toward associations between cues, drugs, and the actions that procure them, perhaps by spuriously elevating reward prediction error signals in the midbrain, striatum, and frontal cortex. These aberrant errors may drive addiction-related plasticity toward a particular behavioral repertoire, but once strongly consolidated, this learning becomes insensitive to updating and is maintained despite large discrepancies between expected and experienced reinforcement.

During the earliest phases of drug use, individuals appear to show sensitized responses to all salient events but with repeated drug use, and hence conditioning, behavioral repertoires become more and more focused on the procurement and consumption of drug rewards. This conditioned attentional orienting response (CAOR) is elicited by initial sensory processing of the cues and engenders a shift in the focus of attention toward the spatial location of drug-predicting cues. It is speculated further that the CAOR gates other drug-associated conditioned responses and Pavlovian-to-instrumental-transfer (PIT)—the energizing effect of drug-predicting cues on drug-seeking behavior, which support and constitute drug-seeking and drug-taking.

In a series of studies with smokers, Hogarth, Dickinson, and Duka have established the utility of eye-tracking as an outcome measure in studies of predictive learning in addiction. They trained a set of novel arbitrary visual cues as discriminative stimuli. That is, subjects were trained that instrumental key-press responses in the presence of these stimuli were always rewarded with smoke delivery. In a subsequent test phase, they found that subjects demonstrated preferential attention to those stimuli relative to cues that did not predict smoke delivery. There are competing theories of the role of attentional allocation in conditioning and suggestions from Hogarth and colleagues that attentional biases may differ depending on the stage of drug use and abuse.

Folk Psychology conceives of drug use as a bad habit, something in which addicts indulge despite an expressed desire not to. Preclinical behavioral neuroscience supports this conception of addiction. As described above, ascertaining whether a behavior is goal-directed or habitual involves a simple yet elegant procedure—if behavior is governed by the incentive value of a reward, then devaluing that reward (by, for example, prefeeding to satiation or associating reward consumption with gastric illness) will decrease the vigor with which an organism engages in reward-seeking behaviors. However, if behavior is habitual, it is no longer governed by a representation of reward value. Instead, it involves reflexive responses to environmental cues previously associated with drug use; hence devaluing the drug reward has no effect on behavioral choice or vigor, and the organism continues to engage in reward-seeking. Although much of this work has been conducted in rodents, recent reports have confirmed similar behavioral and neural processes are involved in human habit learning. That is, with overtraining, devaluing a particular food will not prevent human subjects from engaging in the instrumental action that procures that food, and such behavior is associated with activation in the dorsal caudate and putamen the regions aligned with habitual responding in laboratory animals.

To date, no studies have examined the neural mechanisms of associative learning where drugs are the reinforcers. However, studies using monetary outcomes associated with previously neutral cues have revealed neural differences between addicted individuals and nonsubstance-abusing controls. These studies indicate a more global dysfunction in learning and memory suggestive of rapid habit learning and insensitivity of actions to their consequences. For example, in a financial investment game, smokers' brains compute a fictive error signal (indicating what they would have won had they chosen an alternative option). However, they do not use that signal to guide their choices in the same way nonsmokers do. Likewise, alcoholics learning action choices generate striatal prediction errors, but those errors appear to not be communicated to the frontal cortex effectively such that they are not used to update action choices. Interestingly, in this study, the degree of frontostriatal dysconnectivity predicted current alcohol craving in the problem-drinking group. Of course, these effects with monetary rewards may indicate a relative lack of salience of monetary reinforcers when compared with drugs, but, nevertheless, they implicate associative learning dysfunction in the pathology of addiction. Furthermore, these studies suggest an insensitivity to the consequences of actions in addicts that resonates with the theoretical conception of habitual responding.

Some learning-based treatments for addiction have emerged. Psychotherapies based upon cue reactivity involve repeatedly exposing a client to the sights, smells, or paraphernalia associated with a substance in the absence of actual consumption, thus attenuating the power of those cues to elicit an expectation of drug effects. In other words, they no longer readily predict a “high” is coming, so the cue-drug associations are extinguished. Such therapies are often effective in the therapeutic setting, but they can fail to transfer to the real world outside of the clinic. That is, they are context-dependent.

ANXIETY

Watson and Rayner (1920) originally argued that phobias are associative fears that develop when a neutral stimulus is paired with a traumatic event. They conditioned a phobia in their subject, Little Albert (a 9-month-old boy), by pairing a loud and frightening gong with the presence of a rat. It is said that the boy became afraid of white furry animals subsequently, and that fear even generalized to Santa Claus—whose beard was white and fluffy.

One criticism of conditioning theories, though, is the observation that many people with phobias do not appear to recall a specific conditioning episode. It is possible that simply observing others experiencing trauma or behaving fearfully could be sufficient for some phobias to develop. Such vicarious conditioning of a fear of snakes fear has been observed in primates. Laboratory-reared monkeys without initial fear of snakes watched unrelated monkeys reacting fearfully to snakes. The observers rapidly acquired an intense fear of snakes. This vicarious conditioning also occurred via watching videotapes of model monkeys behaving fearfully. This has interesting implications for human susceptibility to vicarious conditioning from watching movies and TV.

However, many individuals who actually have traumatic experiences do not develop phobias. There may be a genetic vulnerability rendering some individuals more susceptible than others. There are also relevant differences in life experiences that may occur before, during, or following a fear-conditioning experience, and they can affect how much fear is experienced, acquired, or maintained over time.

Latent inhibition is a conditioning phenomenon in which prior exposure to a conditioned stimulus before it is ever paired with an unconditioned stimulus retards subsequent conditioning. Children who have had more previous nontraumatic encounters with a dentist are less likely to develop dental anxiety if subsequently traumatized at the dentist’s office. Likewise, monkeys that had previously seen a conspecific unafraid of snakes did not acquire that fear when they subsequently observed it.

Instrumental learning (learning to make responses rather than passively experience contingencies) is also relevant to fear conditioning and anxiety. As we have discussed so far, Pavlovian conditioning elicits fear in response to situational cues. That fear response then mediates avoidance responses. The avoidance response is maintained, according to two-factor theory, by escape from the conditioned stimulus and by an immediate reduction in its fear-eliciting properties. These ideas aided the development of behavior therapy, particularly exposure and response prevention approaches.

In addressing pathophysiologic and potentially therapeutic processes, we should also consider what happens after fear learning. Extinction refers to the well-demonstrated effect that if a conditioned predictor that engenders a response is presented alone (without the salient outcome), then response strength will decline over successive trials until the stimulus no longer elicits a response. Exposure to the cue without the traumatic outcome should allow the anxious response to extinguish, which is the basis for exposure-based treatments for anxiety.

Despite this success, conditioning as an explanation of how phobias develop has come under fire: basic conditioning theory predicts that fear should decrease over successive nonreinforced presentations of the cue. However, the opposite is often true. When someone with an established spider phobia subsequently comes into contact with spiders without traumatic consequences, fear should extinguish. Instead, spider fear increases, and it incubates. Incubation may proceed through memory reconsolidation. Reconsolidation occurs when memories are recalled into a labile state, integrated with new information, and consolidated once more. This process depends on prediction error and evokes competing processes: on the one hand, there is extinction initiated by a negative prediction error signal and engendering competing learning that overrides the original memory. On the other hand, the positive prediction error response to predictive cues leads to the reminder-based strengthening of the belief. The precise balance between these processes is driven by the magnitude, sign, and timing of the accompanying prediction error signal. Incubation may occur because of this reminder-based reconsolidation. That is, fear memories strengthen because they are reactivated, ruminated upon, reconsolidated, and laid down more strongly.

Although exposure and other interventions based on conditioning have already proved successful, it is clear that therapeutic interventions are relatively simplistic and do not incorporate, for example, the possibility for incubation. Of course, therapists already try to discern cues (e.g., in exposure therapy), but realizing that such cues might also trigger incubation and understanding the role of reconsolidation in fear memory maintenance might add to and improve our therapeutic arsenal. For example, although reconsolidation might contribute to fear of memory maintenance, it can be blocked pharmacologically and psychologically and might therefore be employed therapeutically.

SYNTHESIS AND FUTURE DIRECTIONS

Associative learning lends itself to the burgeoning field of computational psychiatry in which patients' behavioral and brain responses can be compared to and combined with algorithmic predictions of optimal performance in order to characterize which processes are related to (or are perhaps responsible for) their psychopathology. If we can characterize the underlying psychological and neural mechanism of those processes, and we can ameliorate those deficits through behavioral and pharmacologic means, it stands to reason that computational modeling (of behavioral and brain responses) in individual psychiatric patients could perhaps be used to divert patients toward particular treatment approaches that address their specific deficits. Thus far, goal-directed and habitual behavior has largely been modeled in terms of model-based and model-free reinforcement learning, respectively. These computational approaches capture the initial flexibility and outcome sensitivity (model-based/goal-directed) and its transition to inflexible stimulus-driven responding (model-free/habitual). However, many do not appreciate that, in order for these models to work, they necessitate an assumption that agents have complete knowledge of the task and environmental contingencies. This is not realistic. Rather than assuming that our subjects will learn associations (e.g., between stimuli-actions and outcomes), they instead learn beliefs or causal hypotheses—that the same causal factor leads to the stimuli and outcomes and that actions influence those factors. Hence transfer occurs because of the agent's assumptions about common causes between the factors generating the relationships between Pavlovian cues and outcomes and the factors influencing whether (and which) actions lead to outcomes. Similar latent cause models have been applied to Pavlovian conditioning and attentional allocation. Furthermore, these latent cause models can discern goal-directed and habitual behavior. We can apply these models to the behavioral and neural data generated in our studies to discern the representational structures and processes that are guiding learning in clinical disorders.

FURTHER READINGS

Beckers T, De Houwer J, Pineno O, Miller RR. Outcome additivity and outcome maximality influence cue competition in human causal learning. *J Exp Psychol Learn Mem Cogn*. 2005;31(2):238–249.

Blaisdell AP, Sawa K, Leising KJ, Waldmann MR. Causal reasoning in rats. *Science*. 2006;311(5763):1020–1022.

Bortolotti L, Broome M. Affective dimensions of the phenomenon of double bookkeeping in delusions. *Emot Rev*. 2012;4(2):187–191.

Cartoni E Puglisi-Allegra S, Baldassarre G. The three principles of action: a Pavlovian-instrumental transfer hypothesis. *Front Behav Neurosci*. 2013;7:153.

Coltheart M, Langdon R, McKay R. Schizophrenia and monothematic delusions. *Schizophr Bull*. 2007;33(3):642–647.

Corlett PR, Cambridge V, Gardner JM, et al. Ketamine effects on memory reconsolidation favor a learning model of delusions. *PLoS One*. 2013;8(6):e65088.

Corlett PR, Fletcher PC. Computational psychiatry: a rosetta stone linking the brain to mental illness. *Lancet Psychiatry*. 2014;1(5):399–402.

Corlett PR, Taylor JR, Wang XJ, Fletcher PC, Krystal JH. Toward a neurobiology of delusions. *Prog Neurobiol*. 2010;92(3):345–369.

Courville AC, Daw ND, Touretzky DS. Bayesian theories of conditioning in a changing world. *Trends Cogn Sci*. 2006;10(7):294–300.

Curcic-Blake B, Liemburg E, Vercammen A, et al. When Broca goes uninformed: reduced information flow to Broca's area in schizophrenia patients with auditory hallucinations. *Schizophr Bull*. 2013;39(5):1087–1095.

Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*. 2005;8(12):1704–1711.

den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE. A dual role for prediction error in associative learning. *Cereb Cortex*. 2009;19(5):1175–1185.

Devinsky O. Delusional misidentifications and duplications: right brain lesions, left brain delusions. *Neurology*. 2009;72(1):80–87.

Dickinson A. The 28th Bartlett Memorial Lecture. Causal learning: an associative analysis. *Q J Exp Psychol B*. 2001;54(1):3–25.

Eisenhardt D, Menzel R. Extinction learning, reconsolidation and the internal reinforcement hypothesis. *Neurobiol Learn Mem*. 2007;87(2):167–173.

Fiorillo CD. Towards a general theory of neural computation based on prediction by single neurons. *PLoS One*. 2008;3(10):e3298.

Fiser J, Berkes P, Orban G, Lengyel M. Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn Sci*. 2010;14(3):119–130.

FitzGerald TH, Dolan RJ, Friston KJ. Model averaging, optimal inference, and habit formation. *Front Hum Neurosci*. 2014;8:457.

Friston K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*. 2005;360(1456):815–836.

Friston K. The free-energy principle: a rough guide to the brain? *Trends Cogn Sci*. 2009;13(7):293–301.

Fugelsang JA, Stein CB, Green AE, Dunbar KN. Theory and data interactions of the scientific mind: evidence from the molecular and the cognitive laboratory. *Can J Exp Psychol*. 2004;58(2):86–95.

Gershman SJ, Niv Y. Exploring a latent cause theory of classical conditioning. *Learn Behav*. 2012;40(3):255–268.

Gold J, Gold I. The “Truman Show” delusion: psychosis in the global village. *Cogn Neuropsychiatry*. 2012;17(6):455–472.

Haselgrove M, Hogarth L. *Clinical Applications of Learning Theory*. Psychology Press; 2011.

Hoffman RE. Revisiting Arieti’s “listening attitude” and hallucinated voices. *Schizophr Bull*. 2010;36(3):440–442.

Horga G, Schatz KC, Abi-Dargham A, Peterson BS. Deficits in predictive coding underlie hallucinations in schizophrenia. *J Neurosci*. 2014;34(24):8072–8082.

Houillon A, Bessiere P, Droulez J. The probabilistic cell: implementation of a probabilistic inference by the biochemical mechanisms of phototransduction. *Acta Biotheor*. 2010;58(2-3):103–120.

Mesulam M. Representation, inference, and transcendent encoding in neurocognitive networks of the human brain. *Ann Neurol*. 2008;64(4):367–378.

Mineka S, Zinbarg R. A contemporary learning theory perspective on the etiology of anxiety disorders: it’s not what you thought it was. *Am Psychol.* 2006;61(1):10–26.

Mitchell CJ, De Houwer J, Lovibond PF. The propositional nature of human associative learning. *Behav Brain Sci*. 2009;32(2):183–198; discussion 198–246.

Monfils MH, Cowansage KK, Klann E, LeDoux JE. Extinction-reconsolidation boundaries: key to persistent attenuation of fear memories. *Science*. 2009;324(5929):951–955.

Quirk GJ, Pare D, Richardson R, et al. Erasing fear memories with extinction training. *J Neurosci*. 2010;30(45):14993–14997.

Redgrave P, Gurney K. The short-latency dopamine signal: a role in discovering novel actions? *Nat Rev Neurosci*. 2006;7(12):967–975.

Schiller D, Monfils MH, Raio CM, Johnson DC, Ledoux JE, Phelps EA. Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*. 2010;463(7277):49–53.

Schultz W, Dickinson A. Neuronal coding of prediction errors. *Annu Rev Neurosci*. 2000;23:473–500.

Schwitzgebel E. Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pac Philos Q*. 2010;91(4):531–553.

Shanks DR. The associative nature of human associative learning. *Behav Brain Sci*. 2009;32:226–227.

Simms V, McCormack T, Beckers T. Additivity pretraining and cue competition effects: developmental evidence for a reasoning-based account of causal learning. *J Exp Psychol Anim Behav Process*. 2012;38(2):180–190.

Tenenbaum JB, Griffiths TL, Niyogi S. Intuitive theories as grammars for causal inference. In: Gopnik A, Schulz L, eds. *Causal Learning: Psychology, Philosophy, and Computation*. University Press; 2007.

Tronson NC, Taylor JR. Molecular mechanisms of memory reconsolidation. *Nat Rev Neurosci*. 2007;8(4):262–275.

Turner DC, Aitken MR, Shanks DR, et al. The role of the lateral frontal cortex in causal associative learning: exploring preventative and super-learning. *Cereb Cortex*. 2004;14(8):872–880.

Vandorpe S, De Houwer J. Beckers T. Outcome maximality and additivity training also influence cue competition in causal learning when learning involves many cues and events. *Q J Exp Psychol (Hove)*. 2007;60(3):356–368.

Wilkinson S. Accounting for the phenomenology and varieties of auditory verbal hallucination within a predictive processing framework. *Conscious Cogn*. 2014;30:142–155.

Yon D, Press C. Back to the future: synesthesia could be due to associative learning. *Front Psychol*. 2014;5:702.

Zekveld AA, Heslenfeld DJ, Festen JM, Schoonhoven R. Top-down and bottom-up processes in speech comprehension. *Neuroimage*. 2006;32(4):1826–1836.

▲ 1.31 Neuroscience of Substance Use Disorders

KHALED MOUSSAWI, M.D., PH.D.

"Drug addiction" will likely remain in the current vernacular, but it fails as a medical term because it is laden with semantics and negative stigma. In particular, it is impossible to ascribe a precise point at which drug misuse can be classified as drug addiction. Likewise, "drug dependence" fails as a medical term for a disorder because dependence can be a normal physiologic response to a substance. Substance use disorder (SUD), as defined in DSM-5, solves these issues by combining the DSM-IV categories of substance misuse and dependence into a single disorder measured on a continuum from mild to severe. In DSM-5, each specific substance (hereafter referred to as "drug") class (e.g., alcohol, stimulants) is addressed separately, and diagnostic criteria are provided for three distinct drug-related phenomena: SUD, substance intoxication, and substance withdrawal. Although disorders with the majority of drugs are diagnosed based on the same overall criteria (Table 1.31–1), it is increasingly clear that chronic exposure to different drugs produces different patterns of neuroadaptations.

Clinically, severe SUD is defined as meeting six or more of the DSM-5 criteria listed in Table 1.31–1. From a neurobiologic perspective, severe

SUD is a relapsing CNS disorder characterized by craving and compulsive drug seeking and taking, along with negative emotional states (e.g., dysphoria, anxiety, irritability) when access to the drug is prevented. Neural adaptations persist beyond detoxification and are responsible for craving and relapse when an individual has drug access. Research over the last four decades has significantly advanced the understanding of the neurobiology underlying SUD, which has led to improved treatments. Through the use of increasingly sophisticated experimental methods in laboratory animals and humans, detailed information from molecular to behavioral levels exists on the neural pathways that underlie each phase of the addiction process. Although animal models cannot fully reconstitute psychiatric disorders in humans, studies in laboratory animals permit investigation of specific elements of SUD. As neuroscience and psychiatry have evolved, these elements have become more concrete and tractable.

Table 1.31–1.
Generic Diagnostic Criteria for Substance Use Disorder, DSM-5

1. Drug is taken in larger amounts or over a longer period than intended.
2. There is a persistent desire or unsuccessful effort to reduce drug use.
3. A great deal of time is spent in activities necessary to obtain, use, or recover from the drug.
4. Craving, or strong desire or urge to use the drug.
5. Recurrent drug use resulting in a failure to fulfill major obligations at work, home, or school.
6. Continued drug use despite persistent of recurrent social or interpersonal problems caused or exacerbated by drug use.
7. Important social, occupational, or recreational activities are given up or reduced because of drug use.
8. Recurrent drug use in situations where it is physically hazardous.
9. Continued drug use despite knowledge of having a persistent or recurring physical or psychological problem that is likely to have been caused by or exacerbated by the drug.
10. Tolerance: either a need for increased amounts of drug to achieve intoxication or desired effects, or a markedly diminished effect with continued use of the same amount of drug.
11. Withdrawal: either experiencing a drug withdrawal syndrome or taking drug to relieve or avoid the withdrawal syndrome.

As described in the DSM-5, SUDs comprise a collection of cognitive, behavioral, and physiologic symptoms suggesting that an individual continues to misuse drugs in a pathologic pattern despite adverse consequences. SUD diagnostic criteria (see Table 1.31–1) can be roughly grouped into four categories:

1. Impaired control (criteria 1 to 4)
2. Social impairment (criteria 5 to 7)
3. Risky use (criteria 8 and 9)
4. Pharmacologic (criteria 10 and 11)

PHASES OF THE ADDICTION PROCESS

The distinction between drug use, misuse, and SUD is critical to understand. Drug use includes legal use of substances such as caffeine, nicotine, and alcohol, as well as medically prescribed use of drugs such as benzodiazepines, marijuana, and opioid painkillers. Drug misuse includes illicit drug use or when a psychoactive drug is taken for nonmedical use. It is estimated that 15% to 16% of the U.S. adult population engage in drug misuse at some point in their lives. Of those, approximately 3% go on to develop SUD. Implicit in this is the idea that drug use (drug exposure) is required for the expression of SUD, wherein increasing levels of drug consumption lead to compulsive drug use despite adverse consequences.

The transition from casual drug use to SUD has been neurobiologically conceptualized as a shift from impulsive to compulsive use and hedonic dysregulation driven by a transition from positive to negative reinforcement mechanisms. Reinforcement is a motivational process by which a response-contingent delivery or removal of a stimulus increases the probability of a subsequent response. There are two forms: positive reinforcement, which is associated with euphoria (e.g., initial drug "high"), and negative reinforcement, which is associated with cessation of an aversive state (e.g., drug withdrawal). Casual drug users who go on to develop SUD display characteristics of impulse control disorders, in which there is an increasing sense of tension or arousal before the impulsive act, followed by a sense of relief or pleasure when the act is committed. Impulse control disorders are thought to involve positive reinforcement mechanisms. In contrast, individuals who meet criteria for SUD display characteristics of compulsive disorders, in which there are feelings of anxiety and stress before engaging in a compulsive and repetitive behavior, followed by relief from stress when the compulsive behavior is enacted. Compulsive disorders are thought to involve negative reinforcement mechanisms and habitual, automatic behaviors. When viewed within this psychiatric-motivational framework, the development of SUD can be described as a three-stage cycle: binge/intoxication, withdrawal/negative affect, and preoccupation/anticipation (Fig. 1.31–1). Impulsivity is most evident in the early stages, whereas compulsivity dominates in later stages. The transition from substance use to SUD involves pathologic neuroplasticity at each stage within neural circuits that regulate affective state, motor patterns, and cognition.

Binge/Intoxication

Initial drug use may begin in vulnerable individuals. However, it is difficult to study this early stage of substance use because what constitutes vulnerability and reasons for initiating drug use vary between individuals and across cultures. Furthermore, only a small subset of people (20%) who use drugs casually will develop SUD, indicating that the increased vulnerability to SUD is not universal. Indeed, the DSM-5 provides diagnostic criteria for drug intoxication that are separate from those of SUD. Whether or not a person tries a drug and what the hedonic valence of initial intoxication will be is influenced by a confluence of genetic factors, innate traits (e.g., decisional and motor impulsivity, sensation seeking, negative urgency, risk-taking), social and environmental influences (e.g., socioeconomic status, education, stress, exposure to adversity like trauma and abuse, peer pressure, drug availability), and psychiatric comorbidities (e.g., depression, anxiety). Despite these complexities, recent studies have begun to identify some of the critical behavioral and biologic markers that may predict a propensity to SUD. Animal models have been developed that reflect the innate or environmental factors thought to contribute to acquisition of drug taking. These models continue to be tested in behavioral paradigms that reflect the binge/intoxication stage of the SUD cycle (Table 1.31–2).

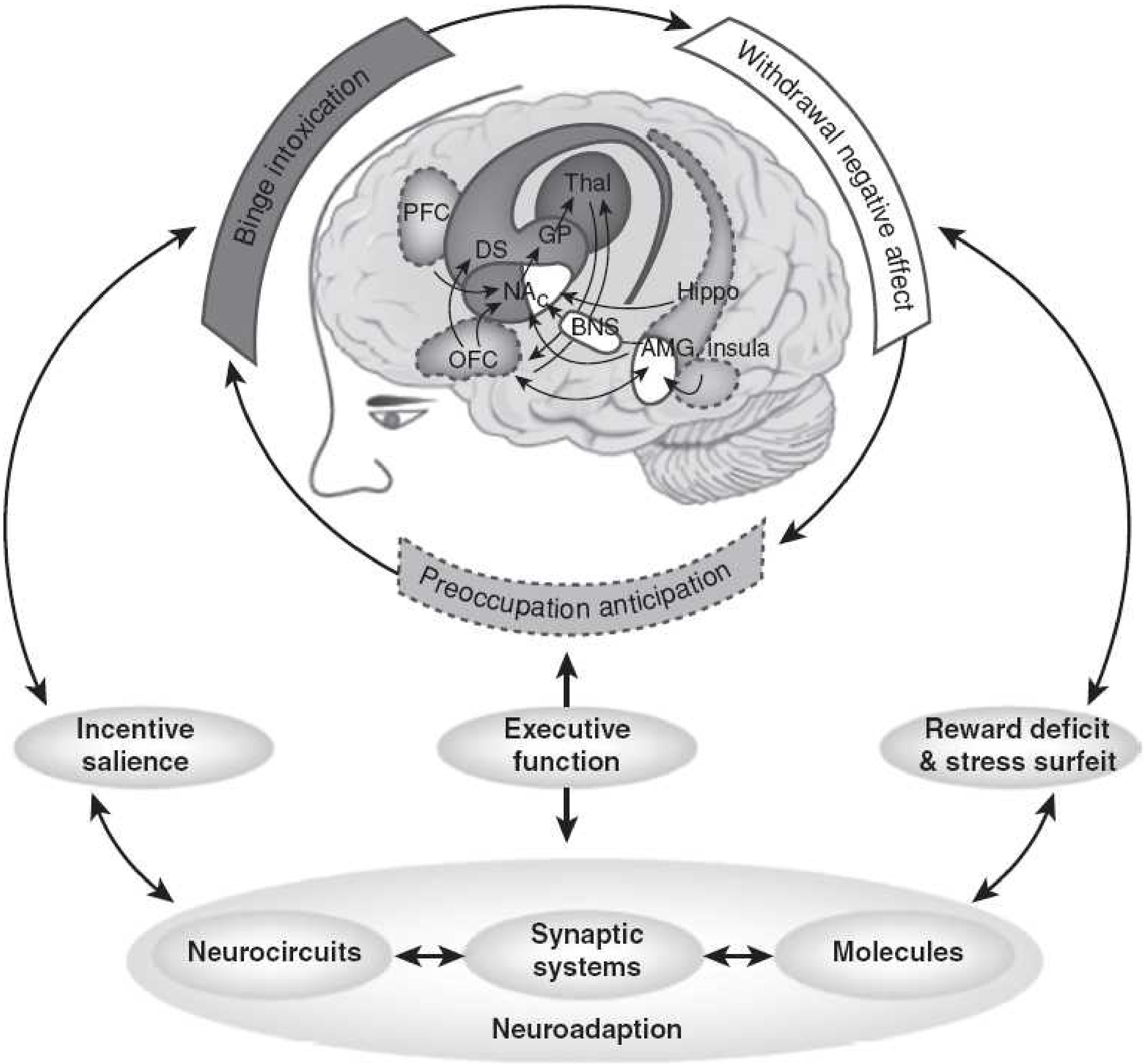
Binge intoxication
Withdrawal negative affect
Preoccupation anticipation
PFC
Thal
GP
DS
NAc
Hippo
BNS
AMG, insula
OFC
Incentive salience
Executive function
Reward deficit & stress surfeit
Neurocircuits
Synaptic systems
Molecules
Neuroadaption

FIGURE 1.31–1. Diagram showing the neurocircuitry of the addiction cycle divided into three stages: binge/intoxication (*dark gray*); withdrawal/negative affect (*white*); and preoccupation/anticipation (*light gray*). The neurocircuits involved are also color coded, with the basal ganglia, including the nucleus accumbens (NAc), dorsal striatum (DS), globus pallidum (GP), and thalamus (Thal) as key elements of the binge/intoxication stage; the extended amygdala, including the central nucleus of the amygdala (AMG), bed nucleus of the stria terminalis (BNST), and a transition area in the shell of the nucleus accumbens (NAc) as key elements of the withdrawal/negative affect stage; and the frontal cortex and allocortex, including the prefrontal cortex (PFC), orbitofrontal cortex (OFC), hippocampus (Hippo), and insula (Insula) as key elements of the preoccupation/anticipation stage. Molecular, synaptic, and neurocircuitry neuroadaptations combine to render the four key elements of the transition to addiction: increased incentive salience, decreased reward, increased stress, and decreased executive function. (From Wise RA, Koob GF. The development and maintenance of drug addiction.

Neuropsychopharmacology. 2014;39:254–262, Springer Nature. doi:10.1038/npp.2013.261.)

Withdrawal/Negative Affect

As substance use progresses toward SUD, periods of chronic drug use are typically followed by periods of abstinence and increasingly severe somatic and/or psychological withdrawal signs. People with SUD have described withdrawal states as a “hunger” or a primary need that is only satisfied by taking more drug (i.e., negative reinforcement). As drug use escalates and becomes compulsive in nature, neuroadaptations develop and result in executive dysfunction (e.g., decreased inhibitory control) and increased motivation for drugs that outweigh the motivation for natural rewards. Motivation is defined as the biologic, emotional, social, and cognitive forces that activate behavior, and it can be modulated by drug-induced changes in hedonic and affective states that occur with the transition from substance use to SUD. This has been conceptualized in the opponent process theory of motivation (Fig. 1.31–2). According to Solomon and Corbit (1974), “hedonic, affective, or emotional states are automatically opposed by CNS mechanisms which reduce the intensity of hedonic feelings, both pleasant and aversive.” Positive hedonic responses to drug use (i.e., euphoria, reward) occur shortly after taking the drug, correlate closely with the intensity, quality, and duration of the drug, and show tolerance. Upon drug withdrawal, negative affective states are unmasked. These are slow to develop, slow to decay, and grow in magnitude with repeated exposure to drug. In the context of SUD, opponent processes begin early in drug taking and reflect compensatory changes in the neural circuitry that controls reward and stress responses.

Table 1.31–2.
Using Animal Models to Study the Stages of Substance Use Disorders

Stage	Source of Reinforcement	Animal Models
Binge/intoxication	Positive reinforcement	Conditioned place preference, drug self-administration, choice procedure, increased brain stimulation reward
Withdrawal/negative affect	Negative reinforcement	Increased anxiety-like responses, decreased brain stimulation reward, conditioned place aversion, increased motivation for drug self-administration during dependence
Preoccupation/anticipation	Conditioned positive reinforcement	Drug-, cue-, stress-induced reinstatement of drug seeking
	Conditioned negative reinforcement	Protracted abstinence

Adapted from Koob G, Kenneth Lloyd G, Mason B. Development of pharmacotherapies for drug addiction: a Rosetta Stone approach. *Nat Rev Drug Discov.* 2009;8:500–515, Springer Nature. doi: 10.1038/nrd2828.

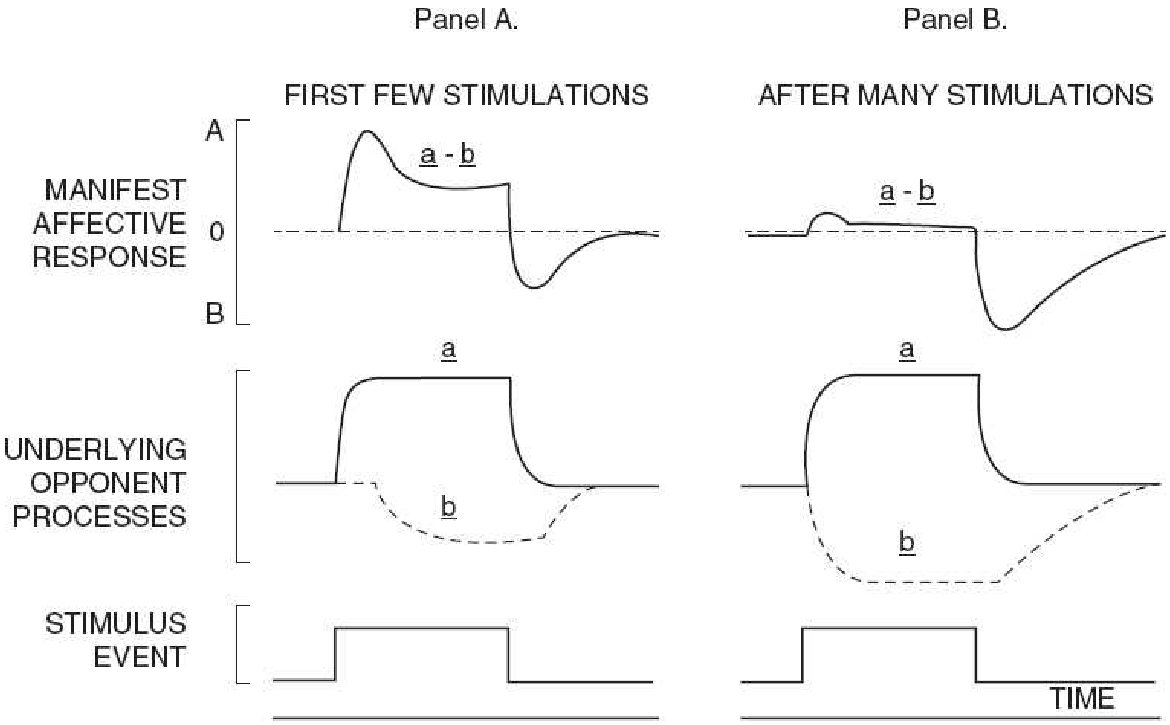

FIGURE 1.31–2. Opponent process theory. **Panel A:** The operation of the summing device for the first few stimulations. **Panel B:** The operation of the summing device after repeated stimulations. The summation of the underlying opponent processes (schematics in **middle row**), “a” and “b,” yields the manifest affective response (schematics on **top row**). In the context of drugs of abuse, manifest affective response “A” represents acute drug reward and “B” represents drug-opposite effects (aversion) that emerge on stimulus offset (schematics on **lower row**). (Copyright © 1974, American Psychological Association. Reproduced with permission from Solomon RL, Corbit JD. An opponent-process theory of motivation: I. Temporal dynamics of affect. *Psychol Rev.* 1974;81(2):119–145. doi:10.1037/h0036128)

Preoccupation/Anticipation

Periods of drug withdrawal are followed by instances of craving and relapse, whereby various triggers such as stress, exposure to the drug itself, or previously drug-paired environmental cues reinstate drug seeking and drug taking. The subjective experience of craving frequently precedes relapse. Craving can be defined as an acquired desire characterized by wanting/anticipating drug use. Its importance is exemplified by its inclusion in the DSM-5 criteria for the diagnosis of SUD (see Table 1.31–1). In studies of people with psychostimulant use disorder, increasing levels of craving have been associated with increasing levels of anxiety, irritability, restlessness, increases in heart rate, nausea, and other symptoms that overlap with stress-related arousal and are transiently relieved with drug use.

The objective measurement of craving in the laboratory has historically presented a major challenge for researchers; however, animal models provide a variety of objective and quantifiable indices of drug-taking and drug-seeking behavior from which neurobiologic substrates of craving may be deduced. The best empirical model to study mechanisms of drug-seeking and drug-taking behavior is self-administration (intravenous [IV], vapor, or oral routes), an established operant procedure for measuring the reinforcing effects of drugs. Drug self-administration in animals meets face, construct, and predictive validity criteria. Animals contingently administer drugs akin to human drug users. IV and inhalation routes allow for rapid drug delivery to the brain and therefore are highly reinforcing. Drug self-administration studies in animals are also widely used for preclinical evaluation of putative pharmacotherapies for the treatment of SUD. Preclinical studies suggest that the underlying mechanisms and neural circuits mediating the reinforcing effects of drugs in animals overlap substantially with those mediating the drug-related effects in humans. These studies have also increased the understanding of the normal function of the neural substrates of reward-related behavior.

Clinical interventions have been successful at getting SUD patients off drugs while minimizing withdrawal symptoms (detoxification). However, clinical interventions to maintain abstinence in those patients have been of limited efficacy. Therefore, the study of the neurobiologic substrates of relapse has received particularly intense focus because this point in the cycle of SUD underlies the chronicity of the disorder, and preventing relapse is critical for successful treatment approaches. Drug self-administration paradigms have been adapted to model relapse by focusing on the reinstatement of operant behavior (i.e., lever pressing or nose poking) previously associated with drug delivery. “Reinstatement”—a common abbreviation for “reinstatement of drug-seeking behaviors”—is an animal model of drug craving and relapse. In this procedure, laboratory animals are trained to self-administer a drug of abuse and then put through either extinction training, in which operant responding is no longer paired with drug delivery, or abstinence. When predetermined criteria of operant responding under extinction conditions is achieved, the animals are tested for reinstatement of drug-seeking behavior by the presentation of a drug prime, discrete cues, contextual cues, or a stressor. Although reinstatement models in animals can be further refined to more closely approximate relapse in humans, they possess good face and construct validity for modeling the activation of drug craving and relapse produced by various stimuli in drug-dependent humans.

THEORIES OF ADDICTION

A number of theories have been proposed to explain the development and maintenance of SUD. Given the heterogeneity of the disorder (e.g., contributions of genes, environment, and the drugs themselves), it is likely that aspects of each theory are applicable under given conditions. The major theories are summarized below and framed in the context of the SUD cycle described above and in Figure 1.31–1. It is important to note that *these theories are not mutually exclusive* but rather tend to emphasize particular aspects of the SUD cycle. There is a general consensus that positive reinforcement kicks off the cycle and leads to the initial repetition of drug taking that eventually becomes habitual and compulsive.

Psychomotor Stimulant Theory

Addictive drugs share common features: they cause psychomotor activation, are reinforcing, and elicit approach behaviors. Seminal work done in the 1960s to 1980s demonstrated that the mesocorticolimbic dopamine system is a major neural substrate of the natural approach system. In this theory, people initiate and maintain drug-taking behavior because drug-induced dopamine release has reinforcing effects and activates motor systems that result in repetitive drug “approaches.”

Incentive Sensitization Theory

This theory builds upon the psychomotor stimulant theory of addiction in that drugs are hypothesized to hijack neural systems (e.g., mesocorticolimbic dopamine system) that normally motivate animals to approach salient stimuli that are important for preservation of the species (e.g., scent of prey). Contexts, cues, and other stimuli—including drugs of abuse—that activate these neural systems become tagged with incentive salience (motivational importance, “wanting”). This theory posits that repeated exposure to drugs of abuse results in persistent neural adaptations in these systems such that the animal becomes hypersensitive to drugs and to any cues that have become associated with the drug. Thus, as SUD develops, a person’s motivational focus narrows to drug seeking at the expense of natural reward seeking. A key aspect of this theory is that sensitization of neural systems that control incentive salience (drug “wanting”) is independent of changes in neural systems that control reward (drug “liking”). Finally, proponents of this theory argue that the systems that mediate incentive salience are distinct from those controlling drug withdrawal–induced negative hedonic states.

Aberrant Learning Theory

The increase in the incentive salience of drugs and drug cues at the expense of natural rewards is explained in the context of aberrant learning. As with incentive sensitization theory, this theory holds that repeated exposure to drugs heightens responsiveness to drug-associated cues by modulating neural systems that control normal responses to nondrug-associated cues. This aberrant learning process is thought to represent changes in associative brain structures such that drug seeking and taking becomes automatic or habitual. Evidence supports the idea that this is mediated by a gradual shift from ventral to dorsal striatal mechanisms. An important aspect of this theory is that the sensitized responsiveness to drugs and drug cues is immune to punishment, resulting in a core symptom of SUDs: continued drug use despite adverse consequences.

Frontostriatal-Dysfunction Theory

This theory incorporates mechanisms to explain the clinical observation that people with SUD tend to have poor impulse control by proposing that repeated drug exposure leads to alterations in cortical and limbic circuits such that an individual displays pathologic responses to drugs and drug cues while simultaneously displaying impaired inhibitory, “top-down” executive control over drug-taking behavior. This leads to loss of impulse control, poor decision making, heightened responsiveness to drug-associated cues, and compulsive drug use.

Hedonic Allostasis Theory

This theory is rooted in the opponent process theory of motivation and postulates that initial drug misuse is motivated by the positive reinforcement obtained from the drug's rewarding effects, whereas chronic drug misuse is motivated by the negative reinforcement resulting from the recruitment of stress-related systems and emergence of hypernegative hedonic and emotional states referred to as "hyperkatifeia." This chronic drug-induced allostatic hedonic state reflects a persistent change in the animal's normal reward set point. Hedonic allostasis contributes to compulsive drug taking through negative reinforcement mechanisms that involve circuits connecting cortical, striatal, and thalamic regions.

BRAIN PATHWAYS IN ADDICTION

There is considerable evidence from animal models and human studies that all drugs of abuse converge on a common circuitry in the brain's limbic system. Synthesizing (and simplifying) decades of work, there are three primary circuits upon which chronic exposure to all drugs of abuse acts. First, chronic exposure to addictive drugs causes a homeostatic impairment of the dopamine system, in which natural rewards (e.g., food, water, sex, social companionship) become less effective at increasing dopamine release in a typical fashion. These homeostatic changes are thought to contribute to negative affective states observed between periods of drug taking or during periods of drug withdrawal and abstinence. It is also observed that chronic, intermittent drug exposure can sensitize the dopamine system, such that exposure to the drug or drug-associated stimuli elicits greater increases in dopaminergic transmission. This sensitization of dopamine responses can persist long after drug taking ceases and may relate to drug craving and relapse. A second neural circuit common to all drugs of abuse involves the extrahypothalamic actions of corticotropin-releasing factor (CRF). Acute drug withdrawal activates CRF-containing neurons in the amygdala, which project to numerous forebrain and brainstem regions discussed below. The stress-like effects of CRF are thought to contribute to the negative affective and somatic symptoms that occur upon drug withdrawal. Another adaptation common to chronic exposure to most drug classes is reduced activity in various brain networks (e.g., salience, reward, executive, memory, habit). Of particular relevance is cortical "hypofrontality," in which there is reduced basal activity of cortical neurons and a decreased ability of natural rewards to activate certain cortical regions such as the prefrontal cortex, ACC, and orbitofrontal cortex. These cortical regions control executive function, working memory, attention, and behavioral inhibition through glutamatergic projections to the nucleus accumbens and ventral tegmental area (VTA). Chronic drug-induced neural adaptations in these frontal cortical regions and their glutamatergic outputs are thought to contribute to both impulsive and compulsive aspects of SUD.

Neural Pathways Underlying Acute Drug Reward

The DSM-5 separates substance intoxication, withdrawal, and SUD into distinct diagnostic categories. Intoxication comprises four criteria (Table 1.31–3), two of which include feelings of euphoria that occur during or shortly after substance ingestion. Although effects other than euphoria can certainly be reinforcing, this section will focus on brain pathways known to play a role in the acute rewarding effects of drugs of abuse. It is this feature of intoxication that is thought to reinforce drug taking and lead to SUD in vulnerable individuals.

Table 1.31–3.
Generic Diagnostic Criteria for Substance Intoxication, DSM-5

A. Recent ingestion of substance.
B. Clinically significant problematic behavioral or psychological changes that develop shortly during or shortly after substance ingestion.[a]
C. A minimum number of physiological/psychological symptoms that are substance-specific.[a]
D. The signs or symptoms are not attributable to another medical condition and are not better explained by another mental disorder, including intoxication with another substance.

[a] DSM-5 includes euphoria for cannabis, inhalants, opioids, and stimulants.

The discovery in the 1950s that rats would learn to press a lever to self-deliver electrical stimulation into specific brain regions (an operant behavior known as intracranial self-stimulation [ICSS]) allowed for the discovery of neural circuits involved in reward function. The efficacy of ICSS to produce brain stimulation reward is quantified by measures such as the minimum frequency of electrical stimulation that supports operant responding. Although brain stimulation reward involves multiple brain regions, the most sensitive site—defined by the lowest stimulation thresholds—is within the medial forebrain bundle at the level of the lateral hypothalamus. The medial forebrain bundle connects the VTA to the basal forebrain and carries ascending norepinephrine and dopamine monoaminergic fibers (Fig. 1.31–3). Acute administration of most drugs of abuse decreases brain stimulation reward thresholds (i.e., increase reward), whereas chronic administration increases reward thresholds during withdrawal (i.e., decrease reward). In summary, these early studies provided strong support for monoaminergic neurotransmission as a critical substrate of the reinforcing effects of drugs of abuse.

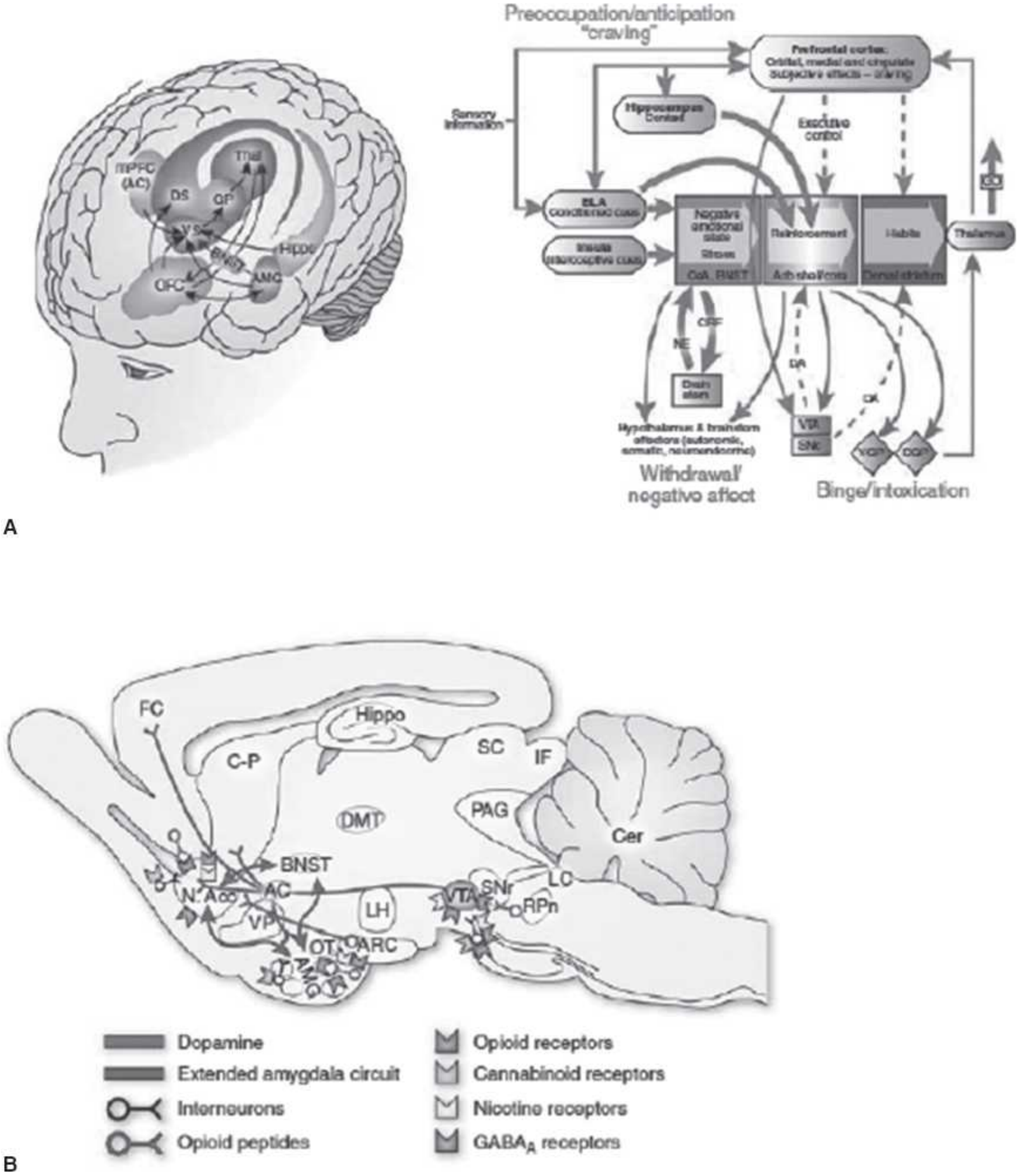
mPFC (AC)
DS
GP
Thal
VS
BNST
Hippo
AMG
OFC
Preoccupation/anticipation "craving"
Prefrontal cortex: Orbital, medial and cingulate subjective effects – craving
Hippocampus Context
Sensory information
Executive control
BLA conditioned cues
Insula interoceptive cues
Negative emotional state
Stress
CeA, BNST
Reinforcement
Acb shell/core
Habits
Dorsal striatum
Thalamus
GO
CRF
NE
Brain stem
DA
DA
Hypothalamus & brainstem effectors (autonomic, somatic, neuroendocrine)
VTA
SNc
VGP
DGP
Withdrawal/ negative affect
Binge/intoxication
A
FC
Hippo
C-P
SC
IF
DMT
PAG
Cer
BNST
N. Acc
AC
VP
LH
VTA
SNr
LC
RPn
OT
ARC
AMG
Dopamine
Extended amygdala circuit
Interneurons
Opioid peptides
Opioid receptors
Cannabinoid receptors
Nicotine receptors
GABA$_A$ receptors
B

FIGURE 1.31–3. Diagram of neural circuits implicated in substance abuse and substance use disorders. **A:** Schematic illustrating the combination of neuroadaptations that are thought to occur within brain circuitry required for the three stages of the addiction cycle. The frontal cortex system is compromised, producing deficits in executive function and contributing to the incentive salience of drugs compared to natural reinforcers. Dopamine systems are compromised, and brain stress systems such as CRF are activated to reset further the salience of drugs and drug-related stimuli in the context of an aversive dysphoric state. (A left, reprinted from Koob GF, Everitt BJ, Robbins TW. Reward, motivation, and addiction. In: Squire LG, Berg D,

Bloom FE, Du Lac S, Ghosh A, Spitzer N, eds. *Fundamental Neuroscience*. 3rd ed. Academic Press;2008:987–1016.) **B:** Sagittal section through a representative rodent brain illustrating the pathways and receptor systems implicated in the acute reinforcing actions of drugs of abuse. The blue arrows represent interactions within the extended amygdala system hypothesized to have a key function in drug reinforcement. AC, anterior commissure; AMG, amygdala; ARC, arcuate nucleus; BNST, bed nucleus of the stria terminalis; Cer, cerebellum; CP, caudate–putamen; DMT, dorsomedial thalamus; FC, frontal cortex; Hippo, hippocampus; IF, inferior colliculus; LC, locus coeruleus; LH, lateral hypothalamus; N Acc., nucleus accumbens; OT, olfactory tract; PAG, periaqueductal gray; RPn, reticular pontine nucleus; SC, superior colliculus; SNr, substantia nigra pars reticulata; VP, ventral pallidum; VTA, ventral tegmental area. (See eBook for color images.) (B, reprinted from Koob GF. The neurocircuitry of addiction: implications for treatment. *Clin Neurosci Res.* 2005;5(2–4):89–101. Copyright © 2005 Association for Research in Nervous and Mental Disease. doi: 10.1016/j.cnr.2005.08.005.)

Emerging data is accumulating that mesolimbic dopamine pathway activity is required for the primary reinforcing effects of psychostimulants but has a more enabling function for other drug classes. These differences are due to the varied pharmacologic mechanisms of action of each drug class. Psychostimulants such as cocaine and amphetamine block monoamine reuptake transporters at nerve terminals to directly increase levels of dopamine and other monoamines, whereas other drugs indirectly activate mesocorticolimbic pathways (Fig. 1.31–4). Opioids, such as oxycodone and heroin, act on μ-opioid receptors in the VTA to decrease the activity of inhibitory GABA interneurons, which results in a greater release of dopamine in forebrain regions. Other drugs, including nicotine and cannabinoids, activate their respective receptors, which cause disinhibition or excitation of dopamine neurons. Drugs with more complex pharmacologic profiles, including alcohol, also lead to increased dopamine release, albeit through actions at a variety of receptor subtypes, including serotonergic, glutamatergic, and GABAergic receptors. While somewhat controversial, there are also theories that the rewarding effects of some of these substances (e.g., opioids, phencyclidine, alcohol, benzodiazepines) may be regulated by actions directly upon the medium spiny neurons of the nucleus accumbens, independent of dopamine.

The mesocorticolimbic dopamine pathway is consistently identified as a common neural substrate of all drugs of abuse. Indeed, studies in rodents have shown that all drugs with addictive potential increase dopamine release in forebrain structures such as the nucleus accumbens (see

Fig. 1.31–3). The majority of dopamine cells reside in the ventral mesencephalon, particularly the VTA, and project to forebrain regions, including the nucleus accumbens, prefrontal cortex, and amygdala. As an example of the importance of this pathway in acute drug reward, it has been shown that lesioning the VTA or nucleus accumbens prevents the self-administration of drugs in rodents. Also, drugs like cocaine, heroin, or alcohol are self-administered directly into the VTA, while inhibition of dopamine neurons in the VTA suppresses drug self-administration. In general, repeated exposure to nondrug rewards (e.g., food) activates the mesocorticolimbic dopamine pathway in a consistent manner that does not result in abnormal (too much or too little) dopamine release. In contrast, drugs of abuse “hijack” this pathway such that acute drug ingestion produces dramatic and supranormative increases in dopamine release, with subsequently profound and long-lasting adaptive changes following repeated exposure to the drug and intervening withdrawal periods. The last several decades of research have identified multiple roles for dopamine: it stamps in salience to stimuli in the environment and promotes behavioral activation and performance of goal-directed behavior. The time course of dopamine signaling is also a key factor. Rapid, phasic dopamine neuronal firing encodes the reward experience and value and is necessary to learn the association of predictive stimuli with rewards and punishers. On the other hand, tonic firing of dopamine neurons determines the overall motivational state in responding to learned cues.

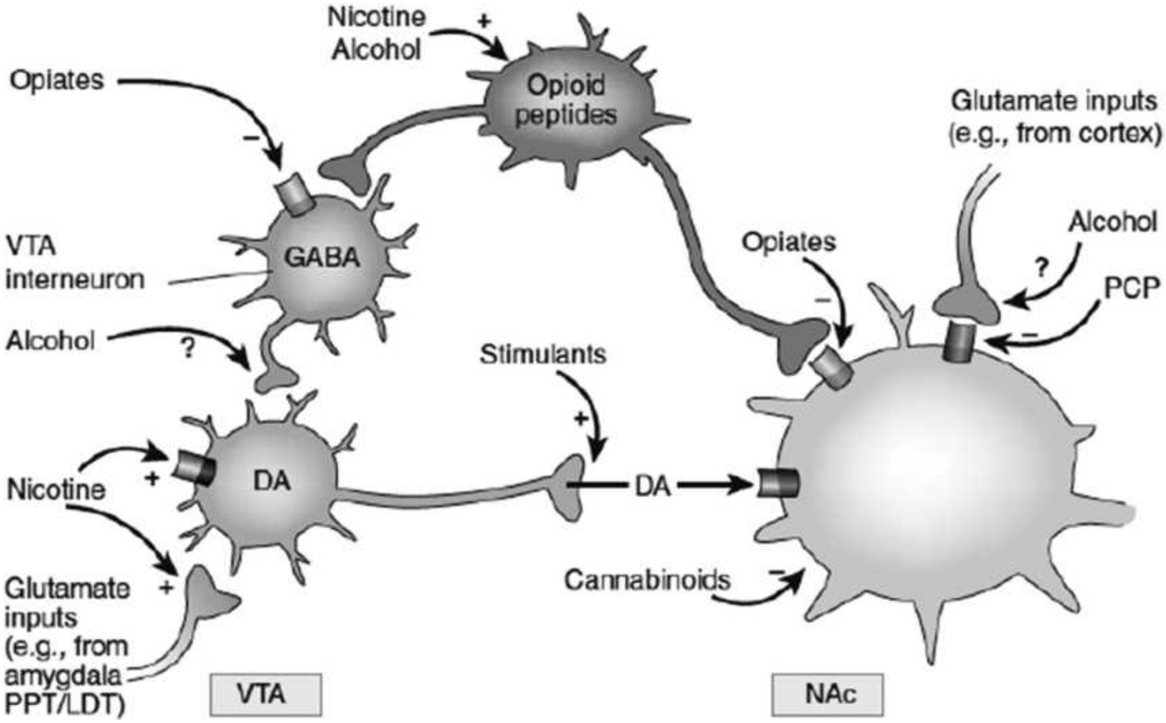
Nicotine
Alcohol
+
Opioid peptides
Opiates
–
Glutamate inputs
(e.g., from cortex)
VTA interneuron
GABA
Opiates
–
Alcohol
?
PCP
–
Alcohol
?
Stimulants
+
Nicotine
+
DA
DA
Glutamate inputs
(e.g., from amygdala PPT/LDT)
+
Cannabinoids
VTA
NAc

FIGURE 1.31–4. Simplified schematic of converging acute actions of drugs of abuse on the VTA–NAc pathway. Stimulants directly increase dopaminergic transmission in the nucleus accumbens (NAc). Opiates indirectly increase dopaminergic transmission in the NAc by inhibiting GABAergic interneurons in the ventral tegmental area (VTA), which disinhibits VTA dopamine (DA) neurons. Opiates also act directly on μ-opioid receptors on NAc neurons. μ-Opioid receptors, like D2 dopamine receptors, signal via the inhibitory G-protein Gαi; hence, the two mechanisms converge within some NAc neurons. The actions of other drugs of abuse remain more conjectural. Nicotine seems to activate VTA dopamine neurons directly via stimulation of nicotinic cholinergic receptors on those neurons and indirectly via stimulation of its receptors on glutamatergic nerve terminals that innervate the dopamine cells. Alcohol promotes GABAA receptor function, which may inhibit GABAergic terminals in the VTA and hence disinhibit VTA dopamine neurons. It may similarly inhibit glutamatergic terminals that innervate NAc neurons. Many additional mechanisms (not shown) are proposed for alcohol. Cannabinoid mechanisms are complex and involve activation of CB1 receptors (which, like D2 and μ-opioid receptors, are Gαı linked) on glutamatergic and GABAergic nerve terminals in the NAc, and on NAc neurons themselves. Phencyclidine (PCP) may act by inhibiting postsynaptic NMDA glutamate receptors in the NAc. Finally, there is some evidence that nicotine and alcohol may activate endogenous opioid pathways and that these and other drugs of abuse (such as opiates) may activate endogenous cannabinoid pathways (not shown). PPT/LDT, peduncular pontine tegmentum/lateral dorsal tegmentum. (Reprinted from Nestler E. Is there a common molecular pathway for

addiction? *Nat Neurosci*. 2005;8:1445–1449, Springer Nature. doi: 10.1038/nn1578.)

The nucleus accumbens is a small subcomponent of the ventral striatum that is a substrate for both dopamine-dependent and dopamine-independent drug reinforcement. The nucleus accumbens receives information related to hedonic states from glutamatergic afferents originating in the amygdala, frontal cortex, and hippocampus and converts it to motivational action via connections with the extrapyramidal motor system. Glutamate neurotransmission is tightly regulated: too much or too little can have deleterious effects (e.g., cytotoxicity), whereas stimulus-dependent changes in glutamatergic transmission are necessary for learning and behavioral engagement. Indeed, a long-standing concept is that the nucleus accumbens is a "motivation to movement interface." The nucleus accumbens itself can be subdivided into core and shell regions whose functional connectivity and neuronal phenotypes reflect this functional integration of motivation to movement. The shell has an important role in acute drug reward, motivated behavior, behavioral sensitization, and changes in affective state. The core is a central portion of the ventral striatum and is a functional continuation of the neighboring, motor-related dorsal striatum.

Neural Pathways Underlying Tolerance and Withdrawal

When drugs are used repeatedly over time, tolerance may develop. Tolerance is defined as a reduced sensitivity to the drug effects (e.g., euphoria) such that a higher dose of the drug is required to achieve the same level of response achieved initially. Importantly, the development of tolerance is not equivalent to addiction. In the context of addiction, tolerance engages mechanisms that serve to oppose reward and, paradoxically, reinforce drug-seeking behavior. Opponent process theory has been used as a framework to understand tolerance and the transition from occasional drug use to addiction: immediate rewarding effects of a drug are followed by delayed aversive effects that are due to drug-dependent recruitment of opponent processes (see Fig. 1.31–2). These aversive effects emerge in force during drug withdrawal and include acute and protracted aversive states, including anxiety, anhedonia, irritability, and depressed mood that resemble MDD. The mechanisms underlying tolerance to drugs and opponent processes occur at the molecular, cellular, and circuit levels and can be classified as within- or between-system adaptations.

A primary example of a “within-system” neural adaptation to chronic drug administration is hypofunction of the mesocorticolimbic dopamine system during withdrawal, which leads to decreased motivation for nondrug-related stimuli and increased sensitivity to, and motivation for, the misused drug (see Fig. 1.31–3A). Consistent with hypofunction of mesocorticolimbic dopamine systems, psychostimulant withdrawal in humans is associated with fatigue, depressed mood, and sluggishness, and in animals is associated with decreased motivation to work for natural rewards and decreased locomotor activity. Similarly, withdrawal from chronic opioids essentially unmasks all the neural adaptations the brain produced in its attempts to equilibrate in the presence of drug. Consequently, neural circuits regulating everything from gastrointestinal function to affective states are instantly unbalanced, and an opioid withdrawal syndrome emerges. One cause of affective withdrawal signs, such as anxiety, dysphoria, depression, and irritability, is thought to be the reduction in dopamine neuron firing and dopamine release in efferent targets such as the nucleus accumbens.

Despite the long tenure as the “reward centers” of the brain, increasing evidence supports the idea that the nucleus accumbens and ventral pallidum are bivalent structures that process positive and negative environmental stimuli into either approach or avoidance behavior. Electrophysiologic data collected over the last several years have shown that the nucleus accumbens may encode aversive states via increases in firing rates of intrinsic medium spiny neurons and that individual nucleus accumbens neurons can track the hedonic valence of a stimulus by varying their firing-rate response to it. In the ventral pallidum, appetitive versus aversive stimuli are encoded by different neuronal populations, namely GABA versus glutamate neurons, respectively. This has important ramifications for understanding addiction since drugs of abuse provide hyperbolic positive (drug “high”) and negative (drug withdrawal, “crash”) emotional signals to the nucleus accumbens and pallidum, and since drug addiction is characterized by insensitivity to aversive outcomes.

Between-system neuroadaptations also contribute to tolerance and withdrawal syndromes. These involve neurochemical systems different from those directly targeted by drugs of abuse, such as norepinephrine, CRF, orexin, dynorphin, substance P, vasopressin, neuropeptide Y (NPY), endocannabinoids, orphanin, and many others. The extended amygdala, which is composed of the central nucleus of the amygdala, bed nucleus of the stria terminalis (BNST), and a medial portion of the nucleus accumbens shell, represents a common anatomical substrate that integrates stress and reward systems to produce drug withdrawal—induced hyperkatifeia and promote negative reinforcement processes associated with the development of SUD (see Fig. 1.31–3A). Engagement of the extended amygdala is one mechanism by which the brain attempts to restore normal function despite the chronic presence of drug. The stress peptide CRF mediates stress responses of both the HPA axis and the extended amygdala, and CRF is released during withdrawal from all major drugs of abuse. In summary, drug withdrawal is associated with within-system decreases in mesocorticolimbic dopamine activity and between-system recruitment of stress-like neurotransmitter systems.

Chronic exposure to drugs of abuse increases the activity of CRF and dynorphin systems within mesocorticolimbic and extended amygdala circuits. In general, CRF and dynorphin produce states that have characteristics of both withdrawal and stress. Dynorphin binds to and activates κ-opioid receptors (KORs), and CRF binds to and activates CRF type 1 and 2 receptors (CRF_1-R, CRF_2-R) in the brain. KOR activation in mesolimbic structures, including the VTA and nucleus accumbens, decreases dopamine transmission, whereas the effects of CRF are more complex. On balance, evidence supports a serial CRF to dynorphin directional relationship between the two stress peptide systems such that stress triggers CRF release, which then functions to increase dynorphin release and subsequent activation of the KOR system within neural stress circuits. Withdrawal from drugs of abuse can be conceptualized as a stressor that causes a dysphoric state encoded by CRF and dynorphin release. This is supported by studies showing that CRF infusions into the VTA can trigger reinstatement of drug seeking—perhaps through activation of dopamine neurons—and CRF antagonists reduce stress-induced but not drug-induced reinstatement of drug seeking. Similarly, KOR antagonism blocks the development and expression of cocaine withdrawal–induced anhedonia and attenuates the expression of nicotine and morphine withdrawal signs in rats. KOR-induced dysphoria is thought to enhance the rewarding valence of drugs of abuse—although empirical evidence centers primarily on psychostimulants. This role of KORs in addictive behavior has been expanded to include a role in stress-induced reinstatement of drug seeking. In rodents, KOR antagonism and genetic deletion of the dynorphin gene block stress-induced, but not cocaine prime-induced, reinstatement of drug seeking.

Neural Pathways Underlying Substance Use Disorder

Impaired Control (Impulsive → Compulsive Use). Repeated exposure to drugs results in potentiation of the pathway between the prefrontal cortex and D1-receptor–expressing medium spiny neurons in the ventral striatum. This is thought to underly the invigoration of drug-seeking behaviors after repeated drug exposure. Another aspect of the development and progression of SUD is the transition from voluntary (goal-oriented, model-based) to habitual (model-free) and compulsive drug use with repeated drug exposure. Recent data show this is encoded through drug-dependent transitions from ventral to dorsal striatal control over drug-seeking and drug-taking behaviors (Fig. 1.31–5). This has been demonstrated in stimulants, nicotine, and alcohol SUDs. The basic organization of cortical inputs to the striatum comprises a functional gradient from limbic to associative to motor domains. A similar gradient is found in the midbrain to nucleus accumbens connections. There is topographical organization such that a medial (i.e., shell) to lateral (i.e., core/dorsal striatum) series of projection loops allows emotion-based information from limbic-associated structures to transfer to motor-related areas of the basal ganglia. It has been conceptualized that the nucleus accumbens (ventral striatum) is an especially critical substrate for the rewarding drug effects that contribute to the development of SUDs, whereas the dorsal striatum is an especially critical substrate for the motor habits that maintain SUDs. This phenomenon of a transition from voluntary to habitual drug use as SUD develops was shown to be associated with increased activity in the orbitofrontal to medial dorsal striatum circuit, which is thought to reflect the overestimation of drug value relative to associated adverse outcomes. Such findings are broadly consistent with functional neuroimaging studies in humans showing increased dopamine release and cue reactivity in the dorsal striatum in patients with alcohol or stimulant SUD, but increased dopamine release and cue reactivity only in the ventral striatum in subjects who use drugs recreationally.

In summary, the connections among prefrontal cortex, the striatum, and midbrain dopamine neurons are organized in an ascending spiral that creates a feed-forward organization from reward-related regions of the striatum to cognitive and motor areas. Importantly, the neural networks underlying the transition from voluntary to habitual drug use are drug dependent. For example, it has been shown in heroin self-administration studies that information in ventral medial cortical to ventral and dorsal striatal circuits may be processed in parallel rather than in a spiraling, feed-forward series.

Craving. Drug craving is an important construct in addiction, and it is now included in the DSM-5 as a clinical symptom of SUD. Craving can be defined as an acquired desire characterized by wanting to use drug in that moment. Drug craving is evident after chronic drug use, along with the associated increased salience of drug and drug cues. Clinical studies show that acute increases in anxiety or distress that occur during periods of drug abstinence or withdrawal do not in themselves predict drug relapse as proposed in models of negative reinforcement. Rather, concomitant increases in aversive states and drug craving predict subsequent drug taking and relapse.

Craving is difficult to study in animal models of SUD as it is a highly personal and subjective, emotion-based state. Recent advances in neuroimaging techniques and ecologic momentary assessment approaches to studying craving and drug use episodes in real time in individuals with SUD have allowed identification of neural circuits involved in craving. While activity in different brain networks is reduced in SUDs in a nondrug context (e.g., salience, reward, executive, memory, habit networks), exposure to drug cues results in a significant increase in neuronal activity of these circuits, which is referred to as cue reactivity and has been shown to be correlated with craving. Similarly, stress-induced craving activates the different brain circuits.

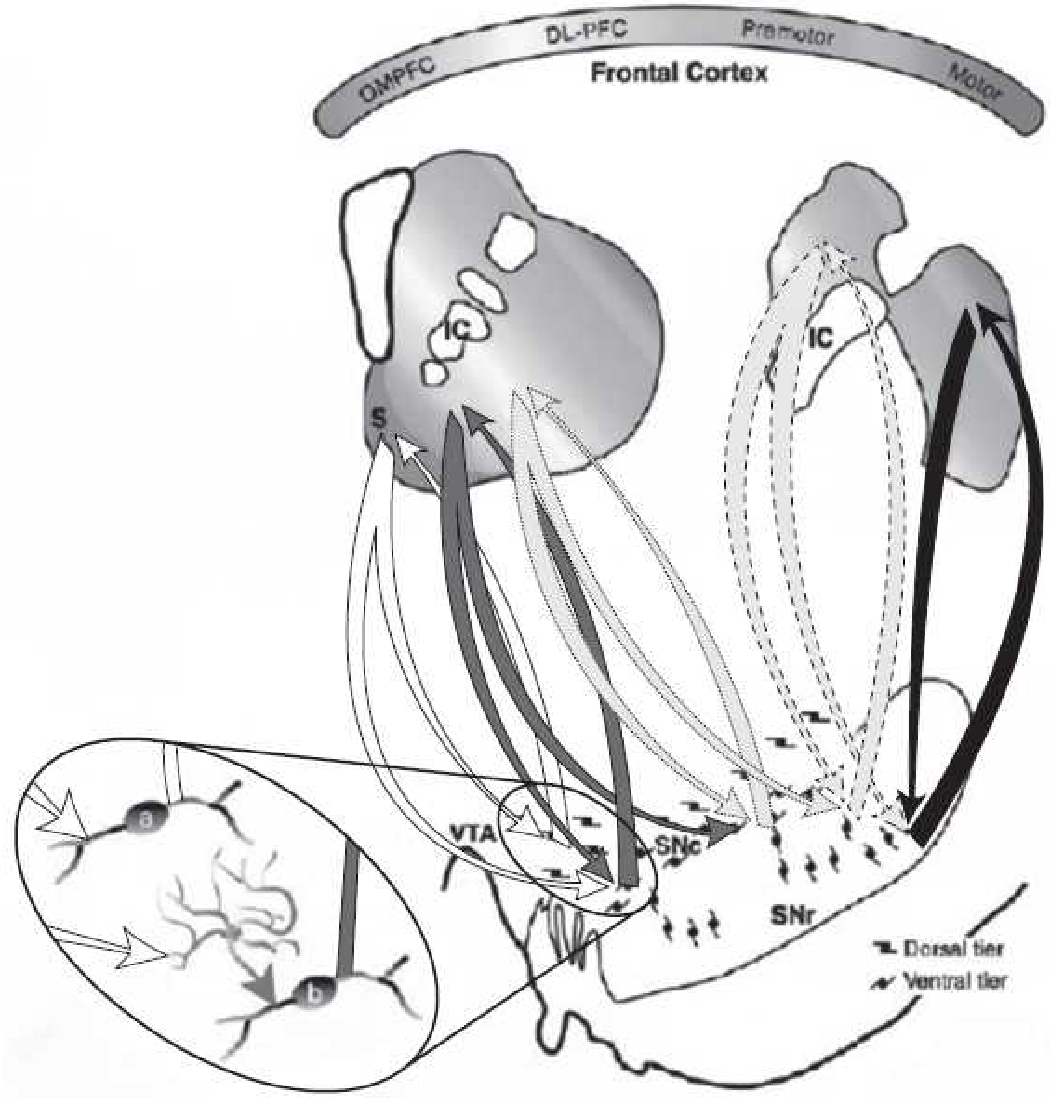

FIGURE 1.31–5. Diagram of the organization of striatonigralstriatal (SNS) projections that comprise spiraling limbic–motor loops. The colored gradient in rostral and caudal schematics of the striatum illustrates the organization of functional corticostriatal inputs (*white* = limbic, *light gray* = associative, *black* = motor). The shell receives forebrain input primarily from the amygdala, hippocampus, and cortical areas 25 and Ia. The core receives input from the entire orbital and medial prefrontal cortex (OMPFC). The dorsolateral prefrontal cortex (DL-PFC) projects to the central striatum and premotor and motor cortex project to the dorsolateral striatum. Midbrain projections from the shell target both the ventral tegmental area (VTA) and ventromedial substantia nigra pars compacta (SNc) (*white arrows*). Midbrain projections from the VTA to the nucleus accumbens shell form a “closed,” reciprocal SNS loop (*white arrow*). Projections from the medial SN feed forward to the nucleus accumbens core forming the first part of a spiral (*dark gray arrow*). The spiral continues through the SNS projections (*light gray arrows* and *dotted arrows*) with pathways originating in the core and projecting more dorsally (*black*

arrows). In this way, ventral striatal regions influence more dorsal striatal regions via spiraling SNS projections. Magnified oval region shows a hypothetical model of the synaptic interactions of SNS projections in reciprocal versus feed-forward loops. The reciprocal component (*white arrows*) of each limb of the SNS projection terminates directly (a) on a dopamine cell, resulting in inhibition. The nonreciprocal, or feed-forward, component (*dark gray arrow*) terminates indirectly (b) on a dopamine cell via a GABAergic interneuron (*light gray cell*), resulting in disinhibition and facilitation of dopaminergic cell burst firing. IC, internal capsule; S, nucleus accumbens shell; SNr, substantia nigra, pars reticulata. (Reprinted from Haber SN, Fudge JH, McFarland NR. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J Neurosci.* 2000;20(6):2369–2382. doi:10.1523/JNEUROSCI.20-06-02369.2000. Copyright © 2000 Society for Neuroscience.)

Relapse. A major problem in the treatment of SUD is relapse to drug use even after periods of abstinence. Unlike craving, which is difficult to model in laboratory settings, the phenomenon of relapse is well modeled in laboratory animals using the reinstatement procedure (described above). Evidence from animal studies suggests that neural circuit changes associated with drug- and cue-induced reinstatement after extinction of operant responding for drug are linked to glutamatergic projections from the prefrontal cortex to the nucleus accumbens core, dopamine projections from the VTA to the medial prefrontal cortex, and GABA projections from the nucleus accumbens to the ventral pallidum. In contrast, stress-induced reinstatement of drug-seeking depends more on activation of CRF and norepinephrine systems within the extended amygdala.

Drug-Primed Relapse. Small doses of a misused drug (often called “priming”) can initiate subjective states of drug desire that prompt renewed drug consumption in humans. Early studies indicated that dopamine neurons in the VTA and glutamatergic neurons in the prefrontal cortex (prelimbic, anterior cingulate, media) that project to the nucleus accumbens shell and core, respectively, play critical roles in cocaine- and heroin-induced reinstatement of drug seeking. Indeed, a key feature of the circuitry underlying cocaine and heroin seeking is enhanced glutamate release into the nucleus accumbens core by afferents from the prelimbic cortex. Inhibiting α-amino-3-hydroxy-5-methyl-4-isoxazoleproprionic acid (AMPA)-type glutamate receptors or activating presynaptic GluR2 autoreceptors in the nucleus accumbens core blocks both cocaine- and cue-induced cocaine seeking. Dopamine transmission within mesocorticolimbic circuitry is also required for drug-primed relapse as blockade of dopamine D1 receptors in the medial prefrontal cortex decreases reinstatement. Additional brain areas have been implicated in drug-primed reinstatement through reversible, pharmacologic inactivation studies. These include the ventral subiculum, ventral pallidum, and central and basolateral amygdala.

Converging evidence from all domains of neuroscience research on addiction highlights the point that the actions and consequences of different drug classes can be distinct from one another. As a case in point, the neural circuitry that underlies heroin-primed reinstatement of heroin seeking has both distinct and overlapping components. Compared to cocaine-primed reinstatement, a more diffuse circuit appears to be engaged during heroin-primed reinstatement because inactivation of multiple cortical and limbic structures abolishes reinstatement of heroin seeking. Some of these heroin-specific brain regions include the substantia nigra, BNST, and mediodorsal thalamus. Furthermore, dopamine has distinct roles. It has been shown that blockade of D1 or D2 dopamine receptors in the nucleus accumbens prevented heroin- and cocaine-primed reinstatement of drug seeking.

Conditioned Cue–Induced Relapse. In the animal model of conditioned cue–induced reinstatement of drug seeking, several lines of research have implicated corticostriatolimbic pathways in the development and maintenance of drug–cue associations that drive drug-seeking behavior after periods of withdrawal. Of particular interest has been the amygdala, which is well established as a critical structure in the learning of affectively relevant stimuli for both appetitive and negative reinforcers. Studies have found that permanent excitotoxic lesions or temporary pharmacologic inhibition of the basolateral amygdala does not affect cocaine taking during daily cocaine self-administration but abolishes the reinstatement of cocaine seeking produced by drug-paired cues long after the cessation of drug self-administration. Additional studies have demonstrated that amygdala modulation of conditioned-cued reinstatement is dopamine dependent. Blockade of dopamine D1 receptors in the amygdala blocks cue-induced reinstatement, whereas increasing dopamine signaling in the amygdala during cue presentation potentiates cocaine seeking.

Discrete subregions of the prefrontal cortex (anterior cingulate and prelimbic cortex) and striatum are also involved in conditioned cue-induced reinstatement. Pharmacologic inactivation of the dorsal medial prefrontal cortex, the lateral orbitofrontal cortex, or the nucleus accumbens core attenuates cue-induced cocaine seeking. Similarly, lesions or inactivation of the ventral pallidum inhibits relapse to drug seeking. The amygdala, prefrontal cortex, ventral pallidum, and nucleus accumbens core circuitry in the animal model of cue-induced relapse shows striking homology with data from humans. Brain imaging studies in cocaine-dependent human subjects have shown that cocaine-paired cues increase metabolic activation of the amygdala, the anterior cingulate region of the cortex, the nucleus accumbens, and the orbitofrontal cortex.

While limited in scope, imaging studies on the neural circuitry of cue-induced relapse in humans have generally found overlapping patterns of brain activation of cortical and limbic structures across various drugs of abuse, including methamphetamine, alcohol, opioids, and nicotine. In rodent models, the amygdala is necessary for both heroin- and methamphetamine-paired cue-induced relapse in animal models. However, a more diffuse circuitry is engaged during heroin-paired cue reinstatement, suggesting mechanistic differences that can be best parsed out in animal models.

In humans, craving and relapse can occur after prolonged periods of abstinence. For unknown reasons, drug craving waxes and wanes over time. In an attempt to account for temporal oscillations of craving and the persistent capacity for relapse, it was hypothesized that cue-induced drug craving progressively increases over time during abstinence and can remain high over several months (>6 months). This effect was first described in the context of alcoholism and was called the “alcohol deprivation effect.” More recently, this effect has been more broadly conceptualized as “incubation of drug craving” and can be modeled in rodents, as shown by time-dependent increases in cue-induced cocaine or heroin seeking that occur after withdrawal from the drug. Many of the same brain regions implicated in cue-induced reinstatement of drug seeking (discussed above) also contribute to incubation of drug craving. In the first few days of withdrawal, the number of silent synapses increases; these synapses contain NMDA but not AMPA glutamate receptors and hence are not active under normal synaptic transmission conditions. However, these silent synapses are then “filled” with the high-conductance calcium-permeable AMPA glutamate receptors resulting in potentiation of glutamatergic synapses in the ventral striatum medium spiny neurons. This finding of increased expression of calcium-permeable AMPA receptors is one of the most robust drug-induced adaptations associated with incubation of craving. In addition, increased neuronal activity has been demonstrated in the ventral medial prefrontal cortex, nucleus accumbens core, and central nucleus of the amygdala when drug craving —as assessed in reinstatement tests after 1 or 30 days of drug withdrawal (extinction)—is measured.

Stress-Induced Relapse. Stress plays a role in all facets of drug misuse and SUD. People with SUD have an increased behavioral sensitivity to stress, even after prolonged abstinence, which supports the clinical relevance of negative emotionality and hyperkatifeia for relapse. By definition, stress is the nonspecific response of an organism to any demand (stressor) placed upon it that requires adaptation. Stressors typically engage the HPA axis to produce physiologic and emotional responses. The biologic effects of stress are relatively hardwired and can be measured objectively, whereas emotional effects can depend on an individual's perception of their ability to cope with the stressor and are, therefore, more subjective. A popular conceptualization of stress is that it is a constellation of negative, depressive-like emotions (anhedonia, anxiety, irritability, feeling overwhelmed) that should be avoided, although stimuli that would seem to qualify as stressors can be experienced as positive or even thrilling (e.g., new job, skydiving). Controlled laboratory studies in patients with SUD have shown that drug craving can be elicited by stressors and that this stress-induced response predicts relapse. The fact that the same stressor may trigger pleasure in one individual but aversion in another has made it very difficult to develop blanket "antistress" pharmacotherapeutics as a panacea for addiction treatment.

Part of the circuitry required for drug-primed or conditioned cue–induced reinstatement of drug seeking is also necessary for stress-induced reinstatement, including the prelimbic cortex and nucleus accumbens, as well as structures within the extended amygdala. Inactivation of BNST and central, but not basolateral, amygdala attenuates stress-induced reinstatement. For intermittent footshock-induced reinstatement, studies indicated a role of CRF and glutamate in the VTA and a role of CRF projections from the central amygdala to the

BNST. Central infusions of CRF produce reinstatement, whereas CRF receptor and ionotropic glutamate receptor antagonists block stress-induced reinstatement of cocaine and heroin seeking, as do noradrenergic α_2-receptor agonists.

Numerous other neurotransmitter systems known to play a role in stress responses likely contribute to stress-induced relapse. For example, enhanced vulnerability to stress during withdrawal and protracted withdrawal may be due in part to a hypodopaminergic state in the mesocorticolimbic system. Drug dependence and early drug withdrawal have been associated with decreased dopamine levels and dopamine receptors in limbic areas such as the nucleus accumbens. Reduced tonic dopamine levels during withdrawal may provide a better contrast, or “signal-to-noise ratio,” to a phasic dopamine signal in response to a stressor. As discussed earlier, the neuropeptide dynorphin is increased with chronic drug administration and acts at KORs on dopaminergic nerve terminals to decrease dopamine release. This raises the possibility that chronically elevated dynorphin expression and release could contribute to drug withdrawal–induced hypodopaminergic state and increased stress sensitivity. Consistent with this, blockade of KORs in the VTA with the selective antagonist nor-BNI decreases swim stress–induced reinstatement of cocaine seeking in rats. A summary schematic of the shared and distinct neurocircuitry for the reinstatement of drug-seeking behavior produced by conditioned cues, drug priming, and stress is illustrated in Figure 1.31–6.

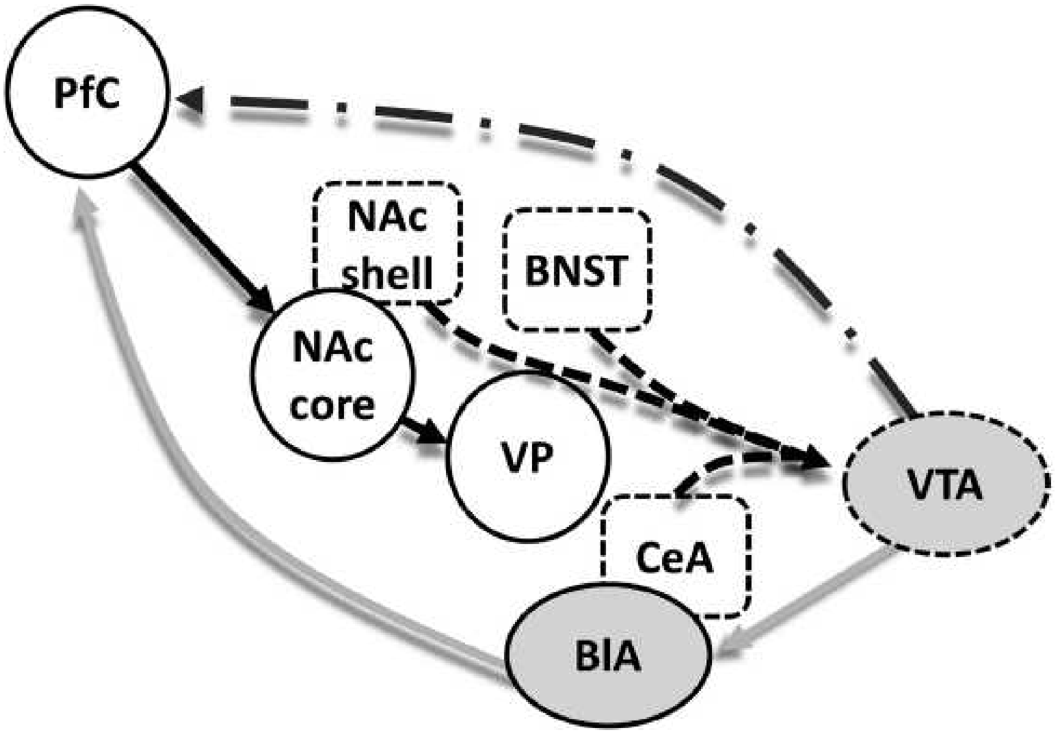

FIGURE 1.31–6. Neural circuitry mediating drug seeking. Serial projections from the prefrontal cortex to the nucleus accumbens core to the ventral pallidum (*white circles*) is a final common pathway for stress-induced drug seeking initiated within extended amygdala circuits (*white, dashed line* boxes), by drug-associated cues (*gray ovals*), or by the drug itself (*dash-dot line* from ventral tegmental area to prefrontal cortex indicating increased dopamine release). PfC, prefrontal cortex; NAc, nucleus accumbens; VP, ventral pallidum; BNST, bed nucleus of the stria terminalis; CeA, central amygdala; BlA, basolateral amygdala; VTA, ventral tegmental area.

CELLULAR AND MOLECULAR SUBSTRATES OF ADDICTION

Defining SUD as a relapsing CNS disorder accompanied by adaptations in the brain that persist beyond detoxification and are responsible for continued drug use, craving, and relapse require identifying and understanding such neural adaptations. This has proven to be difficult, as chronic exposure to drugs of abuse elicits a plethora of cellular and molecular adaptations that vary depending on the type and route of administration of the drug, the brain region studied, developmental history, and biologic sex, to name a few. A current struggle is understanding how different drug classes (e.g., psychostimulants and opioids) can produce similar behaviors but seemingly different molecular and cellular adaptations. Following are a few representative examples of drug-induced neuroadaptations and current theories on their functional significance. This is a rapidly changing field of study with extraordinarily high stakes: development of effective SUD treatments.

Gene Expression

The majority of studies examining drug-induced molecular changes have focused on the mesocorticolimbic system (VTA, nucleus accumbens, prefrontal cortex) and the consequences of altered dopamine signaling. As discussed above, a common initial action of all drugs of abuse is an increase in extracellular dopamine levels in the nucleus accumbens. Dopamine binds to and activates D1- or D2-like G-protein–coupled dopamine receptors, which leads to either stimulation or inhibition, respectively, of adenylate cyclase and cyclic adenosine monophosphate (cAMP) production. As a molecular second messenger, cAMP has multiple intracellular effects, including activation of cAMP-dependent protein kinase A (PKA) and subsequent phosphorylation of an activating site on cAMP response element-binding protein (CREB), a transcription factor (Fig. 1.31–7). As such, a reliable acute effect of most drugs of abuse is modest D1-receptor–mediated CREB phosphorylation in the nucleus accumbens. More importantly, for the development of SUD, chronic administration of drugs of abuse produces compensatory adaptations in cAMP-PKA signaling pathways. Repeated exposure to cocaine, amphetamine, or opioids significantly induces CREB activity in the nucleus accumbens, whereas alcohol and nicotine reduce CREB phosphorylation in this region. Drug-induced modulation of CREB activity is not limited to the nucleus accumbens. For example, acute morphine binds to inhibitory µ-opioid receptors and inhibits CREB phosphorylation in the locus coeruleus, a brain region important for opioid physical dependence and withdrawal. This inhibition undergoes tolerance with chronic morphine exposure due to compensatory upregulation of adenylate cyclase–mediated cAMP production and activation of PKA. Upon morphine withdrawal, this upregulated cAMP/PKA pathway results in a spike in CREB activation, which is thought to contribute to withdrawal signs. Finally, there is evidence that chronic administration of several drugs of abuse activates CREB in the VTA, an effect that directly impacts dopamine synthesis and excitability of VTA dopamine neurons. CREB induces expression of tyrosine hydroxylase (TH), the rate-limiting enzyme in dopamine biosynthesis, in the VTA. Despite increased TH gene expression in the VTA, decreased levels of TH protein and/or activity have been reported in VTA nerve terminals in the nucleus accumbens under certain experimental conditions. Decreases in TH protein levels could mediate chronic drug-induced hypodopaminergic states and emergence of negative affective states.

Although CREB phosphorylation is a transient effect on the order of minutes, repeated brief activations of CREB trigger numerous changes in the neuronal transcriptome that, in turn, cause significant alterations in the physiology of individual cells as well as the circuits in which they are embedded. Some examples of CREB-regulated genes that likely play important roles in SUD include other transcription factors such as c-Fos, intracellular messengers (e.g., adenylate cyclase VI. VIII), neurotransmitter biosynthetic enzymes (e.g., TH), peptide neuromodulators (e.g., CRF and somatostatin), growth factors (e.g., brain-derived neurotrophic factor [BDNF]), opioid peptides (e.g., dynorphin and enkephalin), and neurotransmitter receptor subunits (e.g., GluR1) (see Fig. 1.31–7). The behavioral effects of chronic drug-induced activation of CREB in the nucleus accumbens tend to be similar across drugs. Increased CREB activity in the nucleus accumbens decreases behavioral responses to cocaine, opioids, and alcohol, whereas decreased CREB activity increases such responses. This finding, which has been observed in many experimental systems with a variety of methods, has led to the hypothesis that nucleus accumbens CREB activation is a molecular opponent process engaged to counteract the continued presence of drug. In support of this hypothesis, CREB-induced dynorphin activates KORs on VTA dopamine neurons and their nerve terminals, resulting in decreased dopaminergic tone and opposition of drug reward in the presence of drug or production of negative affective states in the absence of drug (i.e., withdrawal).

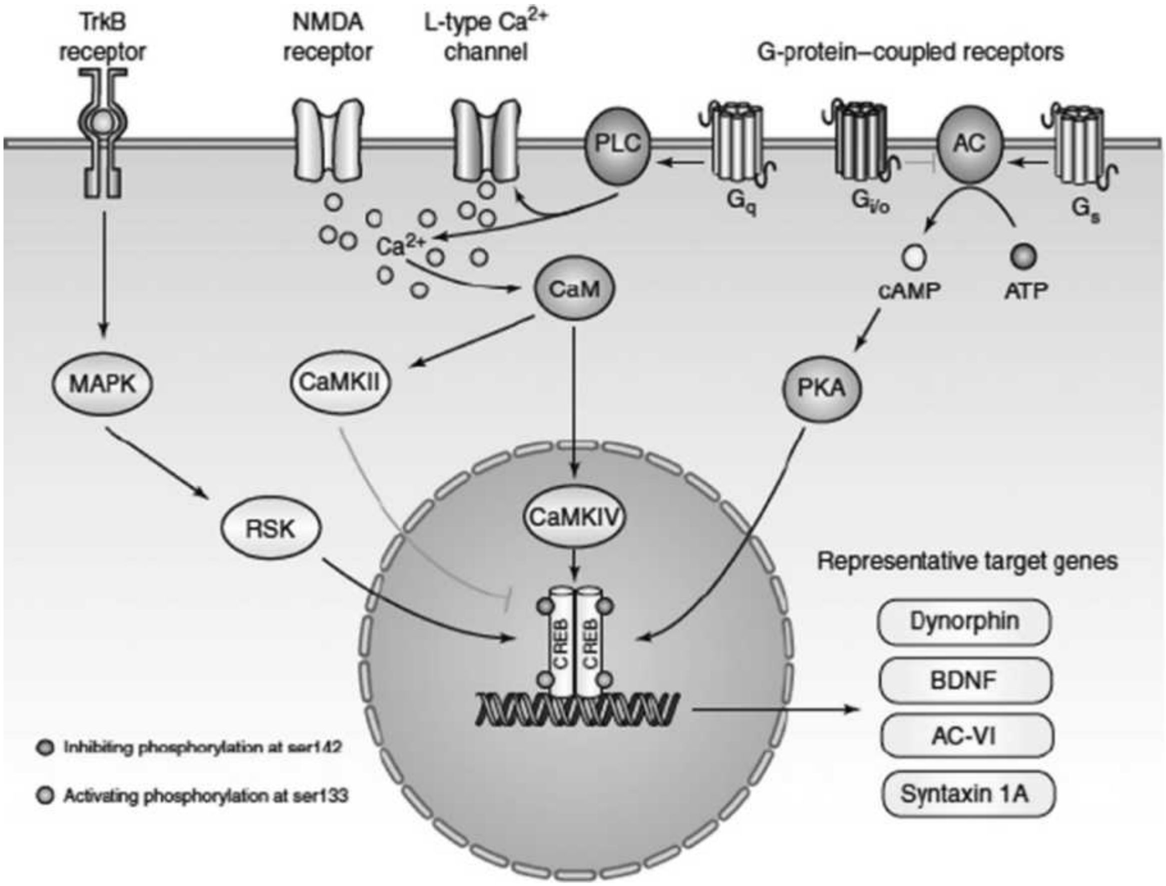
TrkB receptor
NMDA receptor
L-type Ca2+ channel
G-protein–coupled receptors
PLC
Gq
Gi/o
AC
Gs
Ca2+
CaM
cAMP
ATP
MAPK
CaMKII
PKA
RSK
CaMKIV
CREB
CREB
Representative target genes
Dynorphin
BDNF
AC-VI
Syntaxin 1A
Inhibiting phosphorylation at ser142
Activating phosphorylation at ser133

FIGURE 1.31–7. Simplified depiction of how extracellular signals influence gene transcription via CREB. Phosphorylation of CREB at ser133 increases its transcriptional activity and target gene expression. CREB activity can be attenuated by phosphorylation of CREB at ser142. Within the nucleus accumbens (NAc), neurons expressing Gαs-coupled receptors (e.g., dopamine D1) activate adenylate cyclase (AC), whereas neurons expressing Gαi/o-coupled receptors (e.g., dopamine D2) inhibit AC. Exposure to drug-induced increases in dopamine in the NAc can also trigger activation of MAPK- and calcium- (Ca^{2+}) sensitive signal transduction pathways. For clarity, cross-talk between second messengers is not included. *Black lines* signify activation; *gray lines* signify inhibition. (Reprinted with permission from Muschamp JW, Carlezon WA, Jr. Roles of nucleus accumbens CREB and dynorphin in dysregulation of motivation. *Cold Spring Harb Perspect Med*. 2013;3(2):a012005.© Cold Spring Harbor Laboratory Press. doi:10.1101/cshperspect.a012005.)

In addition to CREB activation, drugs of abuse activate other transcription factors, including members of the Fos protein family (c-fos, FosB, Fra-1, and Fra-2) and NAC-1 in the nucleus accumbens. Direct involvement of these gene products in acute drug reward is doubtful, as this depends upon more immediate signaling events associated with dopamine and opioid receptor stimulation. ΔFosB, a highly stable form of FosB, has been shown to accumulate over time with repeated administration of virtually all classes of drugs of abuse. Animals in which ΔFosB is activated have been shown to have exaggerated sensitivity to the rewarding effects of morphine and cocaine. This has led to the hypothesis that ΔFosB may be a sustained molecular “switch” that helps to initiate and maintain a state of addiction. As with CREB, ΔFosB itself does not contribute to cellular functioning but facilitates transcription of numerous genes that encode proteins involved in neural plasticity. Taken together, the current understanding of how drugs of abuse modulate gene expression comes primarily from studies that focus on one drug class (e.g., psychostimulants or opioids). A postmortem study in people with cocaine and heroin use disorder showed that out of approximately 39,000 gene transcripts that were investigated in the nucleus accumbens, only 25 genes had altered expression in both cocaine and heroin users. For about half of these genes, the drugs had opposite effects on expression.

A final example of how drugs of abuse regulate gene expression is the finding that drugs of abuse upregulate BDNF in the VTA, nucleus accumbens shell, and amygdala, where it is thought to act as a “proaddictive” modulator. BDNF promotes presynaptic axonal branching and organization as well as postsynaptic dendritic arborization, number, size, and motility of dendritic spines and enhances the synthesis of synaptic proteins. BDNF is increased by acute drug administration, but BDNF protein levels continue to increase in the nucleus accumbens during withdrawal from cocaine self-administration, with maximum levels observed after approximately 30 days of withdrawal. BDNF can remain elevated for months. The temporal pattern of BDNF accumulation maps onto the phenomenon of incubation of drug craving (as discussed above in Conditioned Cue–Induced Relapse), and intra-VTA or nucleus accumbens shell BDNF injections cause delayed and long-lasting increases in cocaine seeking after withdrawal. However, it has been shown that BDNF projections from the prefrontal cortex to the nucleus accumbens reduce vulnerability to cocaine seeking. This provides another important example of how the behavioral effects of proteins and other modulators depend on the neural circuits within which they act.

Neuronal Morphology and Synaptic Plasticity

In addition to dopamine- and opioid receptor–mediated changes in gene expression within mesocorticolimbic pathways after repeated drug exposure, the transition from drug use to addiction involves persistent restructuring of cortical circuitry. Drug-induced structural plasticity includes changes in the size of cell bodies and changes in dendritic arborizations or spine morphology. The most consistent and recognizable cellular adaptations observed are morphologic changes in dendritic spine density. These changes are most likely due to drug-induced regulation of genes that encode cytoskeleton regulatory proteins, including decreased levels of Homer 1 and postsynaptic density 95 (PSD95), both of which are scaffolding proteins associated with the postsynaptic actin cytoskeleton. In addition, morphine and cocaine have been shown to decrease RhoA, Rac1, and cell division cycle 42 (Cdc42), which are small GTPases that regulate the actin cytoskeleton.

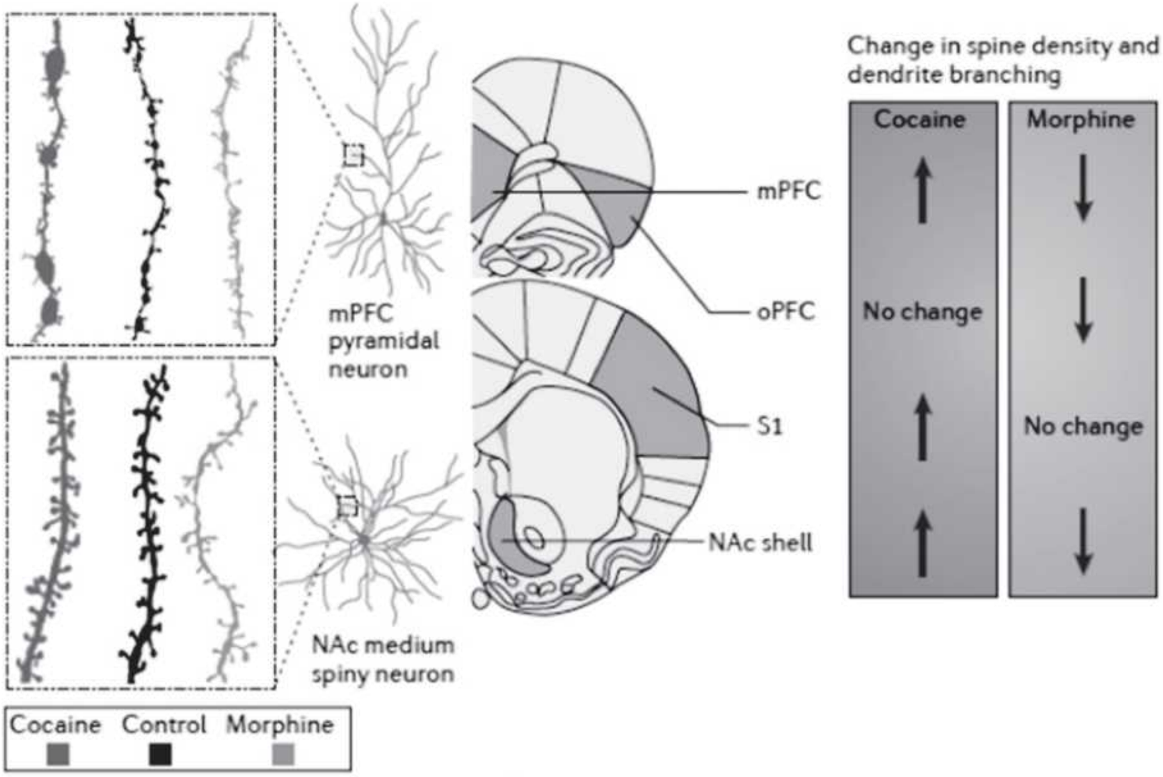
mPFC pyramidal neuron
NAc medium spiny neuron
Cocaine
Control
Morphine
mPFC
oPFC
S1
NAc shell
Change in spine density and dendrite branching
Cocaine
Morphine
No change
No change
A
B

FIGURE 1.31–8. Morphine and cocaine have opposite effects on structural neuroplasticity in the nucleus accumbens (NAc) and medial prefrontal cortex (mPFC). **A:** Groups of rats were trained to self-administer morphine or cocaine intravenously for several weeks. The control groups were given daily intravenous infusions of vehicle for the same period of time. After 1 month of withdrawal from the drugs, the rats' brains were processed using the Golgi staining procedure. Rats that were exposed to cocaine showed increased dendritic branching and increased spine density in both NAc medium spiny neurons and mPFC pyramidal neurons. By contrast, rats that were exposed to morphine had both reduced dendritic branching and reduced spine density in these brain regions. **B:** A summary of changes in spine density and dendritic branching that occur after exposure to cocaine or morphine relative to controls. A dissociation between the effects of cocaine and morphine was also observed in the orbital prefrontal cortex (oPFC) and in the primary somatosensory cortex (S1). (From Badiani A, Belin D, Epstein D, Calu D, Shaham Y. Opiate versus psychostimulant addiction: the differences do matter. *Nat Rev Neurosci*. 2011;12:685–700, Springer Nature. doi: 10.1038/nrn3104.)

A current paradox in addiction research involves the generalization that opioids decrease, whereas psychostimulants increase, the number and complexity of dendritic spines on neurons in the nucleus accumbens, medial prefrontal cortex, and on hippocampus pyramidal neurons

(Fig. 1.31–8). Likewise, opioids decrease, whereas psychostimulants increase, the soma size of VTA dopaminergic neurons. These changes appear to normalize within 2 weeks after the last drug exposure. One hypothesis toward reconciliation of these different cellular effects of opioids and psychostimulants involves the finding that both repeated morphine and cocaine administration increase actin cycling. This can theoretically lead to increases or decreases in spine density, but more importantly, it is associated with greater plasticity in responses to drug-induced changes in glutamate release.

Despite the apparent opposite effects of different drug classes on neuronal morphology and synaptic strength, an implicit assumption has been that neuroadaptations that occur independently of drug class are particularly important for core elements of SUD that exist across drug classes. Relapse is a core element of SUD, and it has been shown that, regardless of drug class, cue-induced reinstatement is accompanied by a rapid, transient synaptic potentiation of glutamatergic synapses in the core of the nucleus accumbens, especially on D1-receptor–expressing neurons. In addition, enlargement of dendritic spine heads and increases in AMPA-mediated glutamatergic transmission are observed. In addition, the extracellular matrix constitutes an environment replete with neurotransmission modulators—many of which are likely involved in addiction. As one example, matrix metalloproteinases (MMPs) are a family of extracellular proteases that strongly modulate synaptic plasticity. Drugs of abuse and reinstatement of drug-seeking rapidly induce MMP activity and subsequent enlargement of dendritic spine heads and increased synaptic strength in the nucleus accumbens core. Likewise, inhibiting MMP activity prevents reinstatement to drug seeking. Finally, a relatively consistent finding is that drug seeking is associated with increased glutamate release in the nucleus accumbens. The mechanisms underlying enhanced glutamate release include downregulation of cystine–glutamate exchange in glia, decreased signaling through presynaptic metabotropic glutamate receptors (mGluRs), and reduced elimination of glutamate through glutamate transporters.

Epigenetics

Epigenetics can be broadly defined as a series of biochemical processes through which changes in gene expression are achieved throughout the life of an animal without a change in DNA sequence. Epigenetics is a means through which environment and genes can interact to produce enduring structural and functional changes underlying experience-dependent neuronal development and plasticity, as well as addiction-related behaviors. DNA is organized into chromatin, and the fundamental unit of chromatin is the nucleosome, which consists of DNA wrapped around a core group of histones. Epigenetic mechanisms control internucleosome spacing and the degree to which they are condensed, which is a determinant of gene activity. Heterochromatin refers to the condensed, inactivated state, which does not allow gene transcription, and euchromatin refers to the activated, open state, which allows individual genes to be transcribed. There are three major types of epigenetic regulation: histone modifications (e.g., acetylation and methylation), DNA methylation, and regulatory noncoding RNAs.

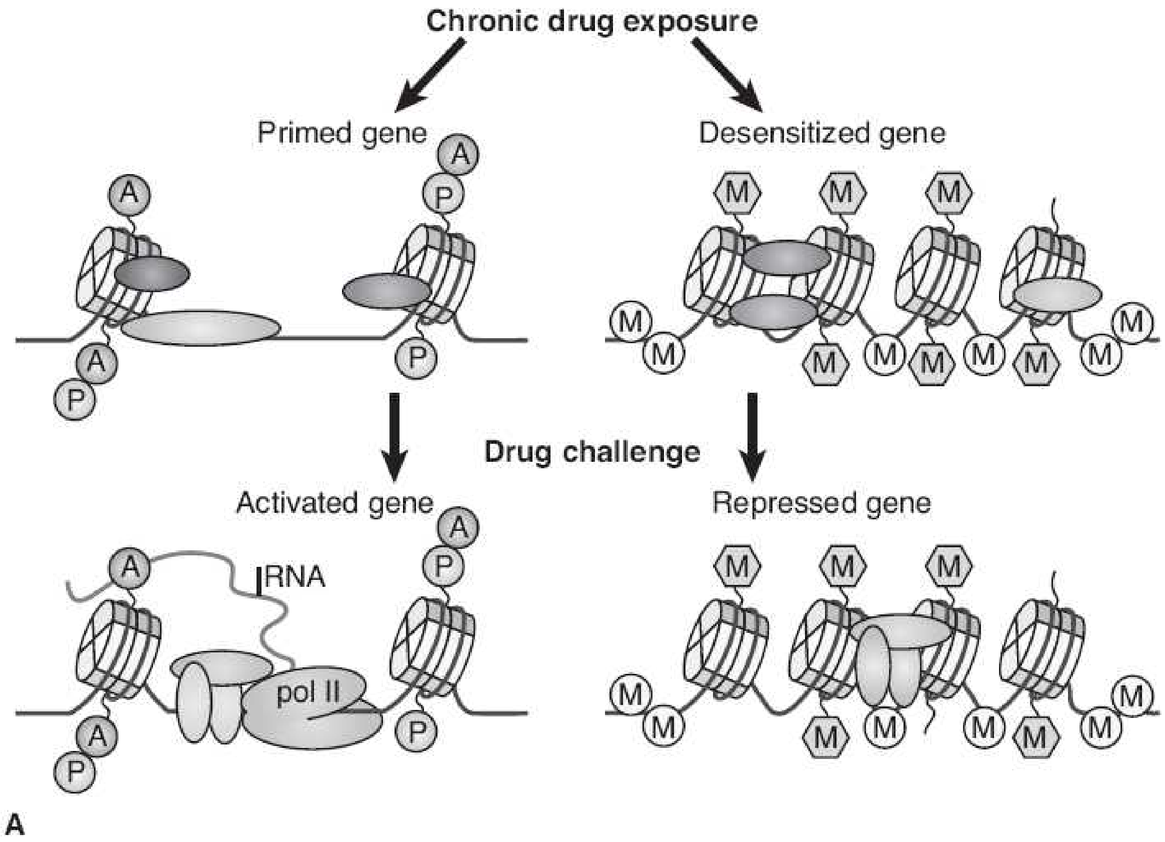
Chronic drug exposure
Primed gene
Desensitized gene
A
P
M
Drug challenge
Activated gene
RNA
pol II
Repressed gene
A

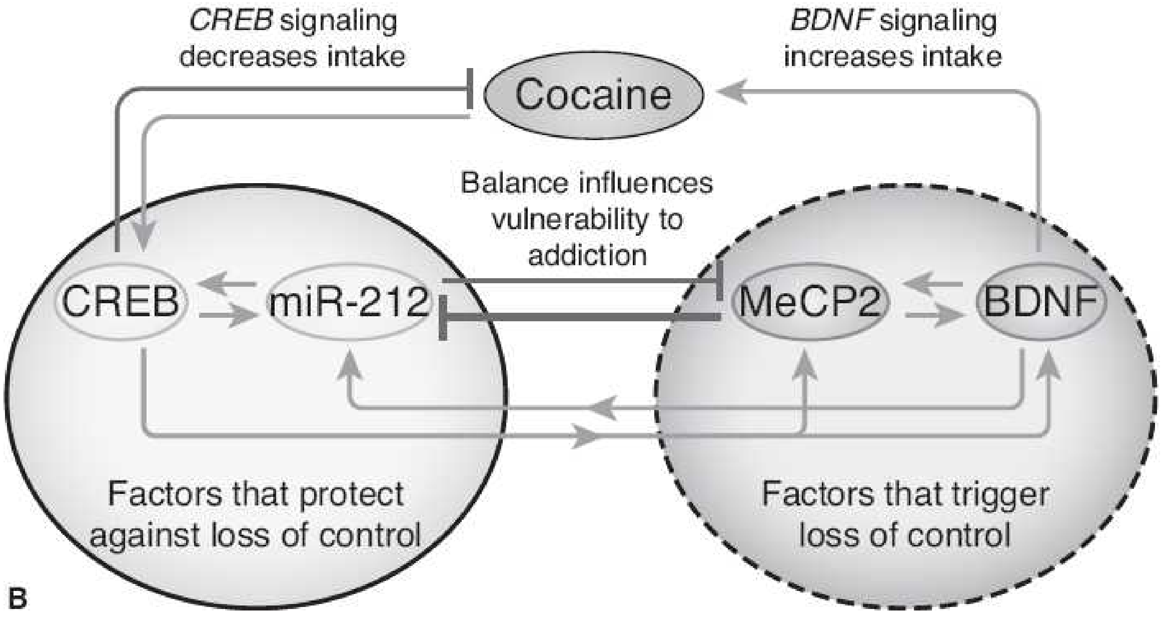
CREB signaling decreases intake
BDNF signaling increases intake
Cocaine
Balance influences vulnerability to addiction
CREB
miR-212
MeCP2
BDNF
Factors that protect against loss of control
Factors that trigger loss of control
B

FIGURE 1.31–9. A: Epigenetic mechanisms implicated in acute and chronic drug effects on gene priming and desensitization. Many epigenetic changes are latent, meaning that they are not reflected by stable changes in steady-state mRNA levels. Instead, such changes alter chromatin structure, so that a later drug challenge induces a given gene to a greater (primed) or lesser

(desensitized) extent based on the epigenetic modifications that are induced by previous chronic drug exposure. A, acetylation; M, methylation; P, phosphorylation; pol II, RNA polymerase II. (A, from Robison AJ, Nestler EJ. Transcriptional and epigenetic mechanisms of addiction. *Nat Rev Neurosci*. 2011;12:632–637, Springer Nature. doi:10.1038/nrn3111.) **B:** The complex interactions between micro RNA 212 (miR-212), cAMP response element-binding protein (CREB), methyl CpG-binding protein 2 (MeCP2), and brain-derived neurotrophic factor (BDNF) and their putative effects on cocaine intake. The *black outlined circle* implies that CREB signaling protects against the development of escalating cocaine intake, whereas the *dotted circle* implies that MeCP2-BDNF signaling promotes escalation of intake. *Light gray arrows* indicate a stimulatory relationship, whereas *dark gray lines* indicate an inhibitory relationship. Cocaine is shown to activate both CREB and MeCP2-BDNF signaling, and the balance between these two pathways, coordinated by miR-212, likely regulates escalation of cocaine intake in extended access rats, and perhaps vulnerability to cocaine addiction. (B, from Jonkman S, Kenny PJ. Molecular, cellular, and structural mechanisms of cocaine addiction: A key role for microRNAs. *Neuropsychopharmacology.* 2013;38:198–211, Springer Nature. doi:10.1038/npp.2012.120.)

Epigenetics likely plays a major role in addiction. Repeated exposure to drugs of abuse or other environmental insults can engage epigenetic processes in early life (or adulthood). These epigenetic processes act at specific genes or sets of genes resulting in stable changes in gene expression or changes in the inducibility of other genes that can lead to addiction in vulnerable individuals (Fig. 1.31–9A). It is also possible that drugs or other environmental insults produce epigenetic changes in gametes, which can be passed on to offspring and increase their vulnerability to SUDs. There is increasing evidence supporting this idea of transgenerational epigenetic inheritance of vulnerability to addiction, but it remains controversial. As such, two salient examples of epigenetic changes are described below.

Histone Acetylation. Histone acetyltransferases (HATs) and deacetylases (HDACs) catalyze acetylation and deacetylation of histones, respectively. Acetylation promotes decondensation of chromatin and—in general—increases gene activity. The bulk of addiction research on histone acetylation has been with psychostimulants, which increase total cellular levels of acetylated histones in the nucleus accumbens. The c-Fos gene promoter is acetylated briefly after drug treatment, with no observable chronic changes, which is consistent with the rapid on–off expression pattern of c-Fos mRNA. In contrast, BDNF promoters become acetylated only after chronic cocaine, which is consistent with the accumulation of BDNF protein only after chronic cocaine and during withdrawal. However, most genes that show altered histone acetylation in response to cocaine do not exhibit changes in mRNA expression expected with hyper- or hypoacetylation. Furthermore, the behavioral effects of histone acetylation are difficult to parse out because each gene promoter has multiple sites at which epigenetic regulation can occur.

Micro RNAs. Micro RNAs (miRNAs) are small, noncoding RNA transcripts (~22 nucleotides) that bind to target mRNAs and either inhibit their translation or induce their degradation. To date, hundreds of miRNAs have been identified, and each miRNA can potentially target hundreds to thousands of mRNA transcripts. In a series of studies focused on the miRNA miR-212 (Fig. 1.31–9B), it was shown that extended access to cocaine self-administration in rats increases levels of miR-212 in the striatum. Experimenter-controlled overexpression of miR-212 in the striatum reduces the motivation to self-administer cocaine, specifically in rats given extended access to the drug, but had no effect on self-administration of cocaine in rats with restricted drug access. CREB increases the expression of miR-212, which can, itself, dramatically increase CREB activity in the striatum. Although BDNF is a CREB target gene (see above), elevations in miR-212 actually repress BDNF expression in the striatum. Taken together, the epigenetic regulator, miR-212, is increased with chronic cocaine self-administration, activates CREB, and suppresses BDNF. These findings suggest that miR-212 may protect against the development of cocaine addiction.

SEX DIFFERENCES IN SUBSTANCE USE DISORDERS

In humans, there are numerous sex differences in almost every facet of SUD (see Table 1.31–4). For most drugs, men consistently have higher rates of use, misuse, and dependence compared to women, with an overall odds ratio of 2:1 (men:women). However, these differences have been declining. For example, the alcohol misuse rate in women is rapidly approaching the rate in men. This suggests that many of the population sex differences in SUDs could be cultural rather than biologic in nature. Having said that, clinical and preclinical studies reveal significant sex differences related to the interaction of gonadal hormones with the stress and mesolimbic dopaminergic systems. Accumulating evidence suggests that women progress from substance use to SUD faster than men, a phenomenon termed “telescoping.” In animal models of SUD, sex differences have been found in drug sensitivity and self-administration behaviors. Repeated drug exposure, especially psychostimulants, enhances incentive sensitization to a larger extent in females compared to males. In general, females are more vulnerable than males to the reinforcing effects of psychostimulants, opioids, and nicotine during each phase of the addiction cycle: acquisition, maintenance, escalation, and relapse.

Table 1.31–4.
Summary of Human Studies on Sex and Ovarian Hormone Influences on Drug Misuse

Dependent Measure/Phase	Drug	Finding[a]
Subjective effects	Cocaine	Women ↑↑ nervousness
		Women ↓↓ euphoria and dysphoria
		↑↑ high during follicular
	Amphetamine	↑↑↑ euphoria and high during follicular
		↑ pleasant stimulation with estrogen
Levels/patterns of use	Cocaine	No sex difference in total levels
		Women ↓ abstinence
	Opioids	No sex difference in total levels
		No sex difference in patterns of use
Relapse	Cocaine	Women ↑ cue reactivity
		Women ↑↑↑ craving during abstinence

↑ Small, ↑↑ Moderate, ↑↑↑ Large effect.

[a] Adapted from Lynch WJ, Roth ME, Carroll ME. Biological basis of sex differences in drug abuse: preclinical and clinical studies. *Psychopharmacology*. 2002;164:121–137, Springer Nature. doi: 10.1007/s00213-002-1183-2.

The biologic mechanisms that underlie sex differences in substance use and SUD are still being elucidated. While the basic neural systems involved in positive and negative reinforcement processes that contribute to SUD development are similar in males and females, sex differences have been described in the organization and connectivity of these systems. When thinking about how sex can influence SUD, it is important to keep in mind that sex differences can occur at behavioral and/or neural levels. Given that rigorous examination of the role of sex in SUD is relatively recent, there are still many gaps in the current knowledge. It has become clear that factors such as drug type, experimental paradigm, animal species, and estrous cycle contribute to behavioral and neurobiologic responses to drugs of abuse. As such, much of the data currently available are widely variable. Following are two relatively robust examples of putative mechanisms underlying sex differences in substance use and addiction.

Activational Effects of Gonadal Hormones

Evidence from both preclinical and clinical studies demonstrates that fluctuations in gonadal sex hormones, particularly estrogen and progesterone, over the course of the female reproductive cycle (Fig. 1.31–10) play important roles in sex differences in the response to drugs. For example, several of the reinforcing effects of amphetamine, such as euphoria and increased energy, are potentiated in women during the follicular phase compared to the luteal phase and are positively correlated with estrogen levels. Consistent with this, administration of exogenous estrogen during the follicular phase enhances the subjective effects of psychostimulants in women. In contrast to the potentiating effects of estrogen, progesterone levels are negatively correlated with the subjective effects of psychostimulants. Furthermore, administration of exogenous progesterone during the follicular phase decreases many of the positive subjective effects of cocaine. The use of animal models allows a more precise examination of the effects of gonadal hormones on behavior because gonadectomy can eliminate the source of hormones, which can then be reintroduced with experimental control of dosing and timing. Using these methods, it has been shown that ovariectomy reduces cocaine self-administration, and exogenous estrogen treatment is sufficient to facilitate acquisition of cocaine, morphine, and heroin self-administration in ovariectomized rats. Interestingly, there is no effect of castration on acquisition of cocaine self-administration in male rats.

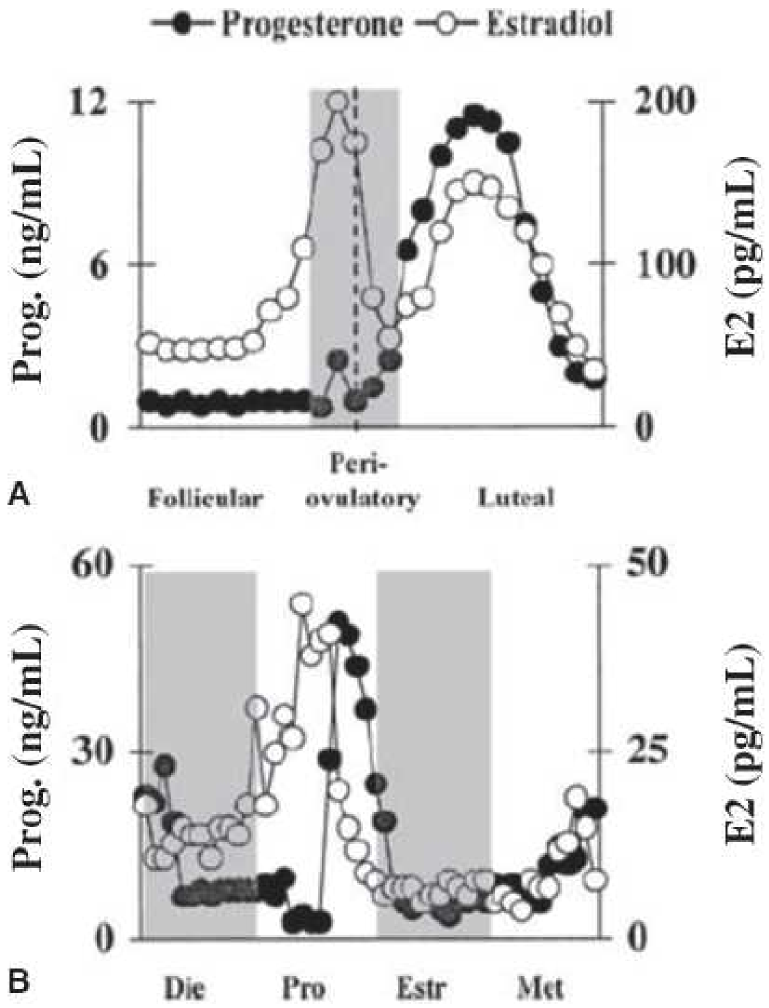

FIGURE 1.31–10. Gonadal hormone fluctuations throughout the female reproductive cycle. **A:** Changes in levels of estrogen (E2) and progesterone (Prog) as a function of days from the preovulatory luteinizing hormone (LH) surge for the human menstrual cycle. The *shaded bar* separates the periovulatory phase from the follicular and luteal phases. **B:** Changes in levels of estrogen and progesterone throughout the phases of the rat estrous cycle. The *shaded bars* separate the successive estrous cycle phases diestrus (DIE), proestrus (PRO), estrus (ESTR), and metestrus (MET) to identify the start and end of each phase. (From Lynch WJ, Roth ME, Carroll ME. Biological basis of sex differences in drug abuse: preclinical and clinical studies. *Psychopharmacology*. 2002;164:121–137, Springer Nature. doi:10.1007/s00213-002-1183-2.)

Sex Differences in Dopamine Signaling. In a second example of a possible mechanism underlying sex differences in addiction, the number and function of midbrain dopamine neurons have been reported to be sexually dimorphic. For example, women have been reported to have a greater capacity for dopamine synthesis and reuptake through the dopamine transporter (DAT) relative to men. Gonadal hormones can regulate the density of dopamine terminals in the nucleus accumbens and dorsal striatum and the rate of firing of dopamine neurons. The density of dopamine terminals impacts basal dopamine tone, which has been shown to fluctuate in females over the course of the reproductive cycle but is relatively stable in males. For example, female rats show higher firing rates and more burst firing in dopamine neurons during the estrous cycle. There is also evidence that—regardless of the estrous cycle—females show greater basal and cocaine-stimulated dopamine release and reuptake compared to males.

Table 1.31–5.
Evidence-Based Clinical Interventions for Substance Use Disorder and Their Hypothesized Functional Neurobiologic Targets

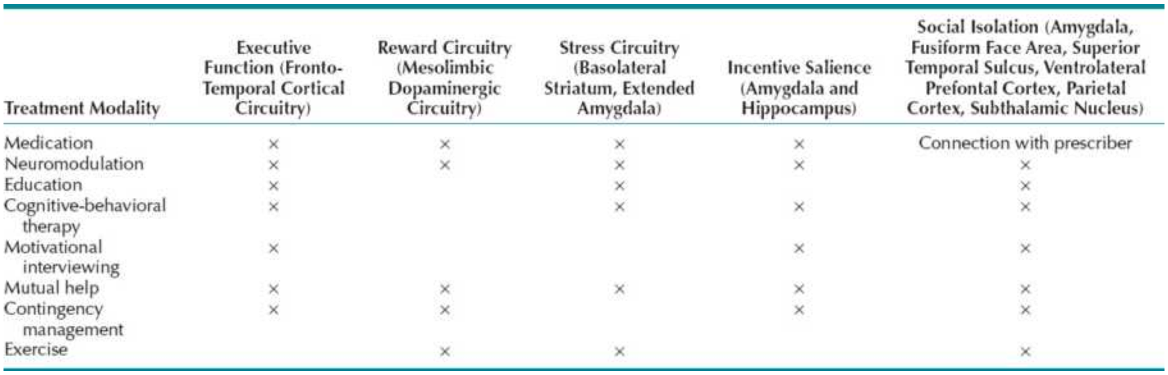

Treatment Modality	Executive Function (Fronto-Temporal Cortical Circuitry)	Reward Circuitry (Mesolimbic Dopaminergic Circuitry)	Stress Circuitry (Basolateral Striatum, Extended Amygdala)	Incentive Salience (Amygdala and Hippocampus)	Social Isolation (Amygdala, Fusiform Face Area, Superior Temporal Sulcus, Ventrolateral Prefontal Cortex, Parietal Cortex, Subthalamic Nucleus)
Medication	×	×	×	×	Connection with prescriber
Neuromodulation	×	×	×	×	×
Education	×		×		×
Cognitive-behavioral therapy	×		×	×	×
Motivational interviewing	×			×	×
Mutual help	×	×	×	×	×
Contingency management	×	×		×	×
Exercise		×	×		×

CURRENT TREATMENTS

Treatment of SUD occurs in three phases: detoxification, initial recovery, and relapse prevention. After detoxification, the goal of the initial recovery phase is to learn to tolerate craving induced by external or internal cues and to learn new patterns of behavior that help the individual avoid relapse. The goal of relapse prevention treatments is for the individual to develop long-term strategies that allow them to replace addictive behaviors with new, healthy behaviors. Treatments can be pharmacologic, behavioral, social, and/or cognitive. There is a general consensus among clinical and preclinical investigators that the most effective treatments will be based on scientific evidence that comes from the understanding of basic neural mechanisms underlying SUD.

Evidence-based clinical interventions may be understood as effective interventions according to the neurobiologic substrates targeted during the engagement process and early recovery treatment of those with SUDs (Table 1.31–5). For example, medication interventions are designed to ameliorate allostatic states associated with chronic substance exposure and the development of physiologic dependence and withdrawal syndromes. As such, stabilization of these systems includes the improvement of withdrawal symptoms, substance craving, and hyperarousal associated with early abstinence. This translates clinically to improved individual capacity for executive functioning and impulse regulation, learning and attention, lowered stress and cue reactivity, and social normalization.

Psychosocial interventions may be similarly conceptualized based on the components of their structure (see Table 1.31–5). Psychosocial interventions include both community-based interventions such as education about SUDs and their treatments, access to exercise programs as a component of health and wellness, and mutual help peer support programs aimed at supporting abstinence (e.g., the 12-step model of recovery founded as Alcoholics Anonymous), as well as professional interventions delivered by trained clinicians such as motivational interviewing, cognitive behavioral relapse prevention, and positive reinforcer contingency management. For instance, education is a prevention and treatment modality that instructs individuals about risk factors for SUDs and also provides a basic understanding of SUD as a brain-based illness that may be arrested with abstaining from substance use. As such, it will primarily assist individuals in making healthy choices, promoting healthy social networks, and avoiding the use of substances as a means of coping with stressors. It will not have much impact on altering reward-seeking behavior and related incentive salience, nor will it help to stabilize withdrawal syndromes. An engagement intervention, known as motivational interviewing, refers to a trained process of dialogue with an individual to help that person better define their goals, values, and current incentivized behavioral patterns, with a goal of assisting that person in focusing on incentivized behaviors congruent with their goals, and diminishing incentivized behaviors (such as substance misuse) resulting in consequences that are inconsistent with personal goals and values. While motivational interviewing may successfully engage an individual in planning for change, it is unlikely on its own to significantly alter subjective reward-driven behavior and/or stress reduction, and thus motivational interviewing best serves as a platform to support engagement in other therapies, such as cognitive behavioral relapse prevention or contingency management, which are more specifically aimed to alter these aspects during the abstinence recovery phase of treatment. Interestingly, mutual help peer support may provide the broadest neural substrate “coverage” in recovery by virtue of its being a socially facilitated system of training the individual to understand addiction cues (“people, places, and things”), avoid occasions of substance use, develop healthier alternatives for stress reduction and natural reward, and integrate meaningful social connection into planned recovery and meaningful life pursuits.

Neuromodulation, defined as an intervention that changes neuronal activity in brain circuits, is emerging as a promising therapeutic approach for SUD. Neuromodulatory tools include invasive approaches like deep brain stimulation (DBS) and noninvasive approaches like transcranial magnetic stimulation (TMS) or transcranial direct current stimulation (tDCS). TMS is currently FDA approved for the treatment of nicotine use disorder.

Below are two clinical vignettes of opioid and stimulant SUD and the clinically recommended treatments currently available.

A 23-year-old single woman with no children, working part-time in a retail department store, and on the weekend as a dancer, presents for first-time treatment of heroin use disorder with IV drug self-administration. She has a significant history of childhood and adult sexual trauma. She began using substances at age 14 with cigarette smoking, drinking with friends episodically, and using marijuana regularly. At age 18, her boyfriend introduced her to using illicit prescription opioids, first orally and then intranasally. At age 19, he left the relationship, and she experienced opioid withdrawal for the first time. This led her to increasingly risky drug-seeking behaviors, including dancing at a strip club, and at age 20, she was introduced to IV heroin use. She has maintained a daily intake of 1 to 2 g IV for 3 years and has not had any medical or mental health care. She presents for treatment in the context of severe weight loss and having been assaulted two nights ago. She is in early opioid withdrawal and has been smoking marijuana for relief. She is a one-pack-per-day cigarette smoker.

Treatment approach: This young woman has multiple life-threatening problems related to her SUDs: lack of stable housing, injection drug use, malnutrition, and sexual assault. The first step in treatment is securing environmental safety through residentially contained housing to support engagement in health care and prevent access to substances and dangerous peers. The severity of early opioid withdrawal symptoms will not be adequately diminished by cannabis use, and consideration of agonist stabilization as a medication-assisted treatment (i.e., buprenorphine or methadone maintenance) would be recommended to engage her in care and to reduce opioid craving that would drive her to leave the residential setting. Treatment with naltrexone is not appropriate in the first 7 to 14 days after the last opioid use to reduce the risk of withdrawal symptoms. Effective care must also be comprehensive and would include assessment of depression and PTSD as well as medical assessment of nutritional status, infection (endocarditis, HCV, HIV, sexually transmitted diseases), and smoking cessation. Women’s peer support would provide important social structure for breaking dangerous behavioral patterns and optimizing reproductive health. Vocational counseling is necessary to support independent recovery in a safe manner. In the setting of injection opioid use disorder, best practices would include a component of contingency management to incentivize recovery-oriented behaviors and training and access to intranasal naloxone rescue in the event of future relapse and opioid overdose.

A 55-year-old divorced man and father of three adult children is referred by his primary care provider for stabilization of cocaine use disorder. He has been followed medically for treatment of recurrent sinus and upper respiratory infections and was ultimately found to have nasal perforation associated with intranasal cocaine use. He owns a plumbing business that has been successful and flourished over the past 20 years. He presents with significant grief, depressed mood, and passive thoughts of suicide in the context of the recent loss of his homosexual partner of 5 years, with whom he has been using cocaine 3 days per week on average. His family is unaware of his loss and drug use, and he lacks any personal support. Recently, he has been using online sites for rapid "hookups" with other men who use cocaine, and he has begun to experience financial strain in supporting some of these evenings. His last cocaine use was 1 week ago, and he presents with severe cravings. He does not use other drugs, has never injected any drugs, and drinks alcohol infrequently.

Treatment approach: In this case, in addition to counseling immediate disengagement with potentially dangerous online partner-seeking sites and assessing his environmental safety, the most imminent risk of harm in his bereaved, early abstinent state is that related to suicidal ideation, which could progress at any time to suicidal planning and suicidal behavior. Careful assessment of the need to hospitalize or otherwise admit him to a residential treatment setting is appropriate, as SUDs are associated with significantly increased suicidality. While there are currently no specific medication-assisted treatments for cocaine or other stimulant use disorders, aggressive treatment of dysphoria/depression, a hallmark of early stimulant abstinence, is warranted; consideration of bupropion would be optimal due to its weak dopaminergic activity, which may contribute to amelioration of cocaine craving. Culturally sensitive therapy addressing "coming out" and recent complicated bereavement is essential for recovery, and mutual help peer support would provide social support that facilitates coping and healing. Although family education is typically an integral component in care, some aspects concerning this case would suggest caution and consideration that the individual be further along in his recovery process prior to integrating family directly into the treatment planning of his health care.

FURTHER READINGS

Becker JB, Koob GF. Sex differences in animal models: focus on addiction. *Pharmacol Rev*. 2016;68:242–263.

Berridge KC, Robinson TE. Liking, wanting, and the incentive-sensitization theory of addiction. *Am Psychol*. 2016;71:670–679.

Bossert JM, Marchant NJ, Calu DJ, Shaham Y. The reinstatement model of drug relapse: recent neurobiological findings, emerging research topics, and translational research. *Psychopharmacology (Berl)*. 2013;229:453–476.

Bruchas MR, Land BB, Chavkin C. The dynorphin/kappa opioid system as a modulator of stress-induced and pro-addictive behaviors. *Brain Res.* 2010;1314:44–55.

Carlezon WA Jr, Thomas MJ. Biological substrates of reward and aversion: a nucleus accumbens activity hypothesis. *Neuropharmacology*. 2009;56(Suppl 1):122–132.

Chartoff EH, Carlezon WA Jr. Drug withdrawal conceptualized as a stressor. *Behav Pharmacol*. 2014;25:473–492.

Cheron J, de Kerchove d'Exaerde A. Drug addiction: from bench to bedside. *Transl Psychiatry*. 2021;11:424.

Dong Y, Taylor JR, Wolf ME, Shaham Y. Circuit and synaptic plasticity mechanisms of drug relapse. *J Neurosci*. 2017;37:10867–10876.

Everitt BJ, Robbins TW. Drug addiction: updating actions to habits to compulsions ten years on. *Annu Rev Psychol*. 2016;67:23–50.

Haber SN, Knutson B. The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*. 2010;35:4–26.

Heyer MP, Kenny PJ. Corticostriatal microRNAs in addiction. *Brain Res*. 2015;1628(PtA):2–16.

*Jonkman S, Kenny PJ. Molecular, cellular, and structural mechanisms of cocaine addiction: a key role for microRNAs. *Neuropsychopharmacology*. 2013;38:198–211.

Kalivas PW. The glutamate homeostasis hypothesis of addiction. *Nat Rev Neurosci*. 2009;10:561–572.

Koob GF. Neurobiology of opioid addiction: opponent process, hyperkatifeia, and negative reinforcement. *Biol Psychiatry*. 2020;87:44–53.

Koob GF, Volkow ND. Neurobiology of addiction: a neurocircuitry analysis. *Lancet Psychiatry*. 2016;3:760–773.

Lüscher C, Malenka RC. Drug-evoked synaptic plasticity in addiction: from molecular changes to circuit remodeling. *Neuron*. 2011;69:650–663.

Lüscher C, Robbins TW, Everitt BJ. The transition to compulsion in addiction. *Nat Rev Neurosci*. 2020;21(5):247–263.

Moussawi K, Kalivas PW. Group II metabotropic glutamate receptors (mGlu2/3) in drug addiction. *Eur J Pharmacol*. 2010;639:115–122.

Muschamp JW, Carlezon WA Jr. Roles of nucleus accumbens CREB and dynorphin in dysregulation of motivation. *Cold Spring Harb Perspect Med*. 2013;3:a012005.

Nestler EJ, Lüscher C. The molecular basis of drug addiction: linking epigenetic to synaptic and circuit mechanisms. *Neuron*. 2019;102:48–59.

Pickens CL, Airavaara M, Theberge F, Fanous S, Hope BT, Shaham Y. Neurobiology of the incubation of drug craving. *Trends Neurosci*. 2011;34:411–420.

*Potenza MN, Sofuoglu M, Carroll KM, Rounsaville BJ. Neuroscience of behavioral and pharmacological treatments for addictions. *Neuron*. 2011;69:695–712.

Prasad AA, McNally GP. The ventral pallidum and relapse in alcohol seeking. *Br J Pharmacol*. 2020;177:3855–3864.

Quigley JA. Logsdon MK, Turner CA, Gonzalez IL, Leonardo NB, Becker JB. Sex differences in vulnerability to addiction. *Neuropharmacology*. 2021;187:108491.

Robison AJ, Nestler EJ. Transcriptional and epigenetic mechanisms of addiction. *Nat Rev Neurosci*. 2011;12:623–637.

Sinha R. The clinical neurobiology of drug craving. *Curr Opin Neurobiol*. 2013;23:649–654.

Smith ACW, Scofield MD, Kalivas PW. The tetrapartite synapse: extracellular matrix remodeling contributes to corticoaccumbens plasticity underlying drug addiction. *Brain Res*. 2015;1628(Pt A):29–39.

Wise RA, Koob GF. The development and maintenance of drug addiction. *Neuropsychopharmacology*. 2014;39:254–262.

Wise RA, Robble MA. Dopamine and addiction. *Annu Rev Psychol*. 2020;71:79–106.

Wolf ME. Synaptic mechanisms underlying persistent cocaine craving. *Nat Rev Neurosci*. 2016;17:351–365.

Zilverstand A, Huang AS, Alia-Klein N, Goldstein RZ. Neuroimaging impaired response inhibition and salience attribution in human drug addiction: a systematic review. *Neuron*. 2018;98:886–903.

2

Neuropsychiatry and Behavioral Neurology

2.1 The Neuropsychiatric Approach to the Patient

FRED OVSIEW, M.D.

Psychiatry has eliminated the term “organic” from the official nomenclature, but the word persists in vernacular use because of the care of patients with identifiable, acquired brain diseases—such as those with epilepsy, movement disorders, traumatic brain injury, and dementia due to neurodegenerative disease—requires the physician to have a knowledge base and a familiarity with assessment and treatment methods not usually required for patients with primary psychiatric disorders. Patients with organic mental syndromes are common in clinical practice and often difficult to manage for the general psychiatrist, even with consultation from other specialists who may themselves not be experts in the mental and emotional phenomena accompanying brain disease.

Neuropsychiatry is the psychiatric subspecialty that deals with the psychological and behavioral manifestations of brain disease. Neuropsychiatry is closely allied with cognitive and behavioral neurology, the neurologic subspecialty that interests itself in psychological phenomena in patients with brain disease, especially cognitive phenomena. In addition to expert management of patients with organic mental disorders, from its clinical vantage point neuropsychiatry can offer a distinctive perspective on idiopathic psychiatric disorders, although later in this section the limitations on the usefulness of this perspective will be noted.

A few preliminary words about the history of neuropsychiatry and its terminology will help to sketch the neuropsychiatric perspective. The seemingly obvious view that neuropsychiatry is the offspring of psychiatry and neurology is historically mistaken. Psychiatry differentiated itself as a medical specialty in the early part of the 19th century; neurology did so somewhat later. Evidence is ample to show that early asylum physicians, the precursors of psychiatrists, considered their patients to be suffering from brain diseases, and moreover that a large proportion of their patients evinced organic disease, even with the limited tools of that time. General paresis of the insane (neurosyphilis, as it was later discovered to be), epilepsy, intellectual disability, and the complications of alcohol abuse were all common in the 19th-century asylum. Based on this evidence, one might say that general psychiatry, as it was understood for the larger part of the 20th century, derived from earlier neuropsychiatry.

An early neurology, on the other hand, took little part in the care of patients with major psychiatric disorders, at least those requiring hospitalization; but what would later become outpatient psychiatry—the care of patients with milder mood and anxiety disorders not requiring asylum management, for example—fell into the province of the early neurologists. The theories they deployed to understand their patients have fortunately been consigned to the dustbin of history. The mainstream of Anglo-American neurology was ill-equipped to give rise to scientific neuropsychiatry, and it was not until (as a convenient and meaningful landmark) Norman Geschwind in 1965 awakened interest in the continental tradition of a behavioral neurology *avant la lettre* that the contributions of John Hughlings-Jackson, Ludwig Lichtheim, Hugo Liepmann, Karl Wernicke, and others could provide impetus to the development of a clinical specialty devoted to the scientific understanding of the cerebral basis of mental and behavioral disorder.

To say that neuropsychiatry is devoted to the care of patients with brain disease is not to depreciate the role of psychological and social factors in the understanding of the genesis of symptoms or in the formulation of interventions to assist patients. On the contrary, patients with brain disease are often inordinately reactive to or dependent on influences from the outside world, notably the social world. To the extent that patients suffer from brain-based impairments in processing information from their environment, their need for assistance in dealing with instrumental and interpersonal tasks increases. Much of the brain, after all, is devoted to processing social information and devising ways of meeting internal needs in a social context. The neuropsychiatrist requires a detailed assessment of the patient's functional deficits and the contexts in which they arise.

To say that the neuropsychiatrist regards deficits, or for that matter intact behavior, as the manifestations of brain-based processes is not to imply that idiopathic psychiatric disorders occur in people with normal brains, nor that general psychiatrists are unaware of the cerebral origin of these disorders. On the contrary, the evidence for abnormal brain structure and function in major psychiatric disorders is unmistakable, and general psychiatrists often assert the neurobiologic nature of these illnesses. However, evidence for such assertions is often not demonstrable in the individual case, with all available laboratory investigations characteristically falling within the broad range of normal. Moreover, the neurobiologic abnormalities in question are believed to be, at least in large part and at least in most illnesses, genetic in nature and developmental in pathogenesis. The recognition and understanding of the mental consequences of acquired diseases of the brain, which form the bulk of the neuropsychiatrist's concern, are likely to require different tools from those required by the general psychiatrist treating idiopathic disorders. Although a bright line between the two situations is not possible and although many neuropsychiatrists maintain a lively interest in such disorders as schizophrenia and autism, the distinction supports the continued use of the term "organic" to refer to these acquired disorders with pathology identifiable at the bedside and by the clinical laboratory.

NEUROPSYCHIATRIC NEUROANATOMY

The "neuropsychiatric brain" is more complex than the "general-psychiatric brain." The latter is a soup of neurotransmitters, perhaps in "chemical imbalance," with considerable pharmacologic but little anatomic specificity. Although the benefits of psychopharmacological intervention are indisputable, the locus of these effects is rarely of concern to clinicians. A neuropsychiatric approach relies on greater differentiation among brain circuits and systems.

Lateralization

The two hemispheres differentially subserve many cerebral functions, although, in many instances, both hemispheres participate in naturally occurring behavior, albeit contributing differently to the complex outcome. Brain asymmetries arose early in vertebrate evolution, and the two hemispheres display regional lateral asymmetries in size and differentially innervate viscera and peripheral endocrine tissues. For example, pars opercularis of the third frontal gyrus (Broca area) and planum temporale (infolded cortex in the posterior portion of the Sylvian fissure) are typically larger on the left, with greater dendritic branching of the neurons therein. (For simplicity's sake, "left" and "right" here refer to the situation in the average dextral patient.) These cortical regions are parts of the substrate of language processing. The insular cortex of the right hemisphere regulates cardiac sympathetic drive, of the left hemisphere parasympathetic drive. In consequence, left hemisphere stroke involving insula produces more cardiac destabilization and morbidity than right, and lateralization of seizure discharges may have implications for autonomic function and unexplained sudden death in patients with epilepsy. Lateral differences in limbic (hypothalamic and amygdalar) regulation of sexual function also have clinical implications; for example, polycystic ovary syndrome in women may be more commonly associated with left-sided limbic epilepsy. The hemispheric side of the lesion also affects the immunologic consequences of brain injury.

Whether a single tag can accurately contrast the processing "styles" of the hemispheres across multiple functions—local versus global or linear versus context-dependent, for example—is doubtful. Although left lateralization of language and right lateralization of visuospatial function is widely recognized, lateral specialization in the prefrontal regions is less obvious but of clinical significance. Frontal lobe degeneration involving the right more than the left frontal lobe is particularly associated with disinhibition. Traumatic injury to the right hemisphere is more associated with depression and anxiety, to the left with anger and hostility. Women and sinistrals tend to show less lateralization of language (and perhaps of other functions) so that left hemisphere lesions are less likely to produce severe impairment.

Of considerable importance for neuropsychiatric practice is the question of lateralization of emotional processing. An array of evidence supports the notion of differential emotional valences in the two hemispheres. On this account, the left hemisphere is specialized for positive emotions, the right for negative emotions. Thus, it has been claimed that left hemisphere destructive lesions are associated with pathologic crying, right hemisphere ones with pathologic laughing; contrariwise, left hemisphere discharging lesions produce gelastic (laughing) epilepsy, right hemisphere ones dacrystic (crying) epilepsy. However, these associations have not been widely confirmed.

Much evidence favors assigning a prepotent role in emotional processing in general to the right hemisphere. Patients with right hemisphere damage appear to be more impaired at perceiving emotion, regardless of the valence or input medium. Lesions of the right hemisphere are associated with impairments in processing emotion in speech, a defect known as aprosodia. Patients may lack the capacity to modulate prosody so as to encode emotional information into speech or the capacity to recognize emotional intonations produced by others. Subtler clinically may be deficits in recognizing emotion in faces or visual scenes.

Fronto-Subcortical Circuits

The projection of the prefrontal cortex to subcortical structures in multiple closed loops is a crucial feature of behavioral neuroanatomy. The key concept is that, in each loop, a distinct region of the prefrontal cortex projects to a distinct portion of the striatum, then to an output nucleus of the basal ganglia, then in turn to a specific region of the thalamus, which in turn projects to the given area of cortex. Thus, a set of parallel (relatively) closed loops of fronto-subcortical connections process information in separate domains. In the motor system, the premotor cortex and supplementary motor area project primarily to the putamen, the output of which projects via ventrolateral globus pallidus and caudolateral substantia nigra pars reticulata (SNr) to ventrolateral/ventroanterior and centromedianum nuclei of the thalamus and then back to the originating cortical structures. Of particular interest to neuropsychiatrists are the loops involving dorsolateral prefrontal, medial and lateral orbitofrontal, and anterior cingulate cortex:

- Dorsolateral prefrontal cortex projects to dorsolateral caudate; projections from caudate go to dorsolateral globus pallidus and SNr. The output from basal ganglia flows primarily to ventrolateral and ventral anterior nuclei of the thalamus (but also to the dorsomedial nucleus of the thalamus), whence it projects to areas 9 and 46 of the dorsolateral prefrontal cortex.

- Lateral orbitofrontal cortex projects to ventromedial caudate, thence to the caudomedial aspect of SNr. The thalamic level of this loop is represented in ventral anterior and dorsomedial nuclei, whence projections arise back to the lateral aspect of area 12 in the orbitofrontal cortex.
- The medial orbitofrontal cortex loop features projections from the gyrus rectus and medial orbital gyrus to ventromedial caudate; output from the basal ganglia arises in SNr and flows to the dorsomedial nucleus of the thalamus as well as ventrolateral and ventral anterior nuclei, then back to the medial orbitofrontal cortex.
- The anterior cingulate cortex, in the dorsomedial aspect of the hemisphere, projects to the ventral striatum, including the nucleus accumbens and olfactory tubercle (termini of the mesolimbic dopamine system); with output from SNr flowing through the ventral anterior thalamus on its way back to anterior cingulate cortex.

Disruption of each of these loops produces a distinctive clinical syndrome. As is implied by the concept of a circuit, deficits similar to those produced by cortical damage can also occur with damage to the subcortical connections of the cortical region. Most naturally occurring lesions do not respect the anatomic boundaries, so clinical presentations are commonly mixed. Nonetheless, for analytic purposes, anatomic specificity is of interest and importance.

Interference with the loop involving the dorsolateral prefrontal cortex prominently produces executive cognitive impairment, with decrements in working memory, problem-solving, and related capacities. Damage to this loop commonly arises from traumatic brain injury, stroke, and basal ganglion degenerative diseases, such as Parkinson disease. Involvement of the white matter of the frontal lobes by small vessel disease commonly leads to interruption of cortico-subcortical connections in this circuit, resulting in the picture of subcortical dementia.

Damage to the orbitofrontal cortex and its connections produces impulsivity, disinhibition, dampening of the experience of emotion, irritability and lability of affect, poor judgment and decision making (especially in regard to social behavior), and insightlessness about these impairments. These impairments are generally seen with bilateral damage, although a unilateral right-sided injury may also produce them. As a neighborhood sign, damage may involve the olfactory nerve (which runs along the orbital surface of the brain) with consequent anosmia; at times, the only elemental neurologic sign. Cognitive function, as tested by the usual bedside or neuropsychological probes, may be unaffected even in the presence of devastating personality change. Trauma is a common etiology.

Damage to dorsomedial prefrontal structures may arise from a tumor or stroke. Abulia and apathy, disorders of initiation of action and the experience of motivation, are the result. Abnormalities of initiation of movement, with akinetic mutism as the most extreme state, may occur. The cingulate cortex is a structure of particular interest. Evidence from animal and imaging studies demonstrates its importance in orienting attention under conflicting stimulus demands, modulating focused problem-solving and monitoring performance to optimize reward. A cell type seen only in the cingulate cortex and fronto-insular cortex, the spindle cell (or von Economo neuron), appears in ontogeny only at age 4 months, concomitant with the infant's increasing capacity to focus attention. These cells appear to be specifically damaged in frontotemporal dementia. Interference with the output of the cingulate gyrus, namely by interrupting the cingulum—the procedure of cingulotomy—appears to be beneficial in a disorder of excessive attention, namely obsessive-compulsive disorder (OCD).

Limbic System

Le grand lobe limbique was delineated in the mid-19th century (by Broca of aphasia fame) as a ring of cortical and subcortical structures on the medial aspect of the hemispheres. Papez drew attention to the circuit formed by projections from the hippocampus via fornix to mamillary bodies of the hypothalamus, then to the anterior nucleus of the thalamus, then via anterior limb of the internal capsule to the cingulate gyrus, then back to the hippocampus via presubiculum, entorhinal cortex, and the perforant pathway. In addition to this "Papez circuit," the amygdala and its reciprocally connected orbitofrontal cortex are taken to form part of a limbic system, a term first used by MacLean.

The core limbic structures are characterized by rich reciprocal monosynaptic connections with the hypothalamus. These are (1) hippocampus, (2) amygdala, (3) piriform cortex, anterior to amygdala on the medial surface of the temporal lobe, (4) septal nuclei, in the medial wall of the hemispheres immediately rostral to lamina terminalis, and (5) substantia innominata in the basal forebrain. Paralimbic cortices reside in temporopolar, insular, and orbitofrontal regions, which have primary affiliations with the amygdala, and in parahippocampal, retrosplenial/posterior cingulate, and subcallosal regions, with primary affiliations with the hippocampus.

In the limbic system, broad and direct input from sensory cortices into the amygdala and hippocampus is extensively processed on its way to effector neurons in the hypothalamus that regulate autonomic and endocrine activity. In addition to this mediation of the regulation of the internal milieu, the limbic system gates the activity of the motor systems in the basal ganglia, regulating action in the external milieu. This occurs by prefrontal cortical integration of information in the limbic fronto-subcortical circuit, which reaches the cortex via projections from the ventral pallidum to the dorsomedial nucleus of the thalamus.

One reason for the central importance of the limbic system in neuropsychiatry is that the threshold for production of epileptic discharges is lowest in the amygdala and hippocampus. Thus most epilepsy in adults is limbic epilepsy. One consequence is the "voluminous mental state" first identified by Hughlings-Jackson. This refers to the range of experiential phenomena encountered as auras in limbic epilepsy: déjà vu, depersonalization/derealization, micropsia and macropsia, and so on. Such symptoms are seen not only in epilepsy but also in mood disorders and as putative pointers to limbic involvement in paroxysmal disorders not of clear epileptic nature, including those associated with borderline personality disorder and with childhood abuse. Their presence, therefore, does not unequivocally mark an epileptic phenomenon. Another reason for the centrality of the limbic system is that the hippocampus, in particular, has a crucial role in explicit memory, further discussed below.

Cerebellum

Anatomic evidence shows that cerebellar inputs access areas of the prefrontal cortex, with a relay in the thalamus. These areas of the cortex project reciprocally to the cerebellum, creating, as with the prefrontal-basal ganglia circuits previously discussed, a set of parallel (relatively) closed loops. These crossed connections from the cerebellar hemispheres and the further crossing of descending cerebrofugal long tracts mean that motor deficits manifest ipsilateral to lateralized cerebellar injuries. Additional reciprocal connections link the cerebellum with the hypothalamus monosynaptically and with other areas of the limbic system via a relay in basis pontis. The phylogenetically older vermis and fastigial nucleus can be differentiated from the neocerebellum of the cerebellar hemispheres and considered a “limbic cerebellum.”

Growing evidence of cerebellar contributions to cognition and affect comes from clinical data and neuropsychological studies. Interpretation of some of the data is uncertain, however, because many cerebellar patients have disorders that may not be limited to the cerebellum; for example, cerebellar degenerations may include cortical degeneration and tumors (and their treatment with radiation and chemotherapy) may have remote effects. Moreover, the phenomenon of crossed cerebellar diaschisis—the reduction in blood flow to connected neocortical areas after cerebellar damage—means that interpretation of deficits as due to abnormal cerebellar processing, as opposed to shut-down of cerebral cortical processing, is treacherous. Nonetheless, patients with stroke lesions clinically and by neuroimaging limited to the cerebellum may have deficits in executive cognitive function, memory, language, and visuospatial function. The data suggest that lateralized cerebellar damage is associated with the predicted lateralized cognitive phenomena (right cerebellar damage with language impairment, left with visuospatial impairment). Defects in affect regulation, with irritability and lability, are proposed to be associated with damage to the limbic cerebellum, notably the vermis.

Cerebral Cortex

The cerebral cortex develops through complex but increasingly well-understood processes of cell proliferation and migration, axonal projection, and dendritic proliferation and pruning. Abnormalities in these processes lead to cortical dysplasia, with clinical consequences including intellectual disability and epilepsy.

Some 10% of intractable epilepsy may be due to such disorders, and increasingly migration abnormalities are recognizable by imaging prior to neuropathologic examination. Failure of normal pruning of synapses by elimination of dendrites is now known to be crucial in the pathogenesis of the fragile X syndrome and has been speculatively linked to schizophrenia. Rarely, cortical dysplasia may be present without epilepsy or intellectual disability; the neurobehavioral consequences of this abnormality are just coming under investigation.

The organization of sensory cortices follows a regular plan. Each primary sensory cortical area projects to unimodal association cortices specialized for the extraction of features in that particular modality; the unimodal association cortices are densely and reciprocally interconnected. For example, the visual association cortex has specialized regions for color, motion, and shape. Unimodal association cortices project in turn to heteromodal cortices which receive inputs from more than a single sensory modality. Heteromodal cortices are located in prefrontal, posterior parietal, lateral temporal, and parahippocampal regions. Unimodal cortices do not project to unimodal cortices in other modalities, only to the higher-level heteromodal cortices. Further, widespread hippocampal projections to the cortex arrive only at association cortices, not primary sensory cortices. These structural features amount to the isolation of sensory processing from top-down influences over the first several synaptic stages and presumably increase its fidelity to external phenomena.

Lesions of cortical association areas produce an array of behavioral and cognitive disorders of intriguing specificity. The specificity can be demonstrated by the occurrence of double dissociations: a lesion in area A produces a deficit in function X but not Y; a lesion in area B produces a deficit in function Y but not X. This pattern of findings provides crucial confirmation that the deficits arise not from task difficulty (if Y were simply more difficult than X, then Y would always be disturbed when X was disturbed) but from separable processing components. For example, some patients show a greater impairment for naming living things than for naming artifacts after a brain injury. However, occasionally patients show the opposite pattern, greater impairment in naming living things: a double dissociation. The explanation of the discrepancy thus cannot depend on insufficient processing resources but must reveal a property of the organization of the semantic system.

Cognitive disorders of the visual system can serve as a paradigm of the syndromes seen with damage to the association cortex. Lesions of the primary visual cortex (V1, or Brodmann area [BA] 17) produce cortical blindness in a quadrant, hemifield, or the entire visual field. Despite the genuine blindness, accuracy above chance in localizing visual stimuli can be achieved without awareness of vision, the phenomenon of blindsight, which testifies to subcortical visual processing inaccessible to consciousness. V1 projects to adjacent cortical regions (BA18 and BA19), which contain neurons that respond to specific features of visual stimuli such as color, movement, or shape. Lesions in these cortices produce deficits in the identification of these features. Thus arise syndromes such as central achromatopsia, demonstrated by the inability to sort (as well as to name) colors. The information transfer divides into dorsal and ventral streams, with the former specialized for localization of visual stimuli (“where”) and the latter for identification of the stimuli (“what”). Dorsal lesions involving the superior parietal lobule can produce impaired reaching under visual guidance (optic ataxia), a part of the Balint syndrome; the deficit testifies to the integration of visual information with motor output in the association cortex. Ventral lesions involving the inferotemporal cortex produce defects in recognition (agnosia). Agnosic patients are not only unable to name elements within the domain of agnosia but also unable to demonstrate their use or show recognition of the items in other nonverbal ways.

Central auditory disorders include cortical (or central) deafness; pure word deafness, the inability to recognize words presented in the auditory modality despite preserved visual-verbal function; and auditory agnosia, the inability to recognize words or complex sounds (e.g., the meaning of the ringing of a telephone). Central deafness requires bilateral lesions involving the primary auditory cortex in the superior temporal gyrus or auditory radiations in white matter. Patients with pure word deafness generally have bilateral lesions of the association cortex more anteriorly in the superior temporal gyrus, although unilateral left lesions, presumably disconnecting left from right cortices by subcortical damage, also are reported. The primary auditory cortex is partially spared in these cases. In agnosia for nonverbal environmental sounds, right-hemisphere damage is sufficient to produce the deficit. Amusia, the incapacity to recognize musical sounds, is associated with cortical damage, but the issue of laterality is complex, dependent in part on the preinjury level of musical skills.

White Matter and Cerebral Connectivity

Although the volume of the neocortex has increased over the course of phylogenetic history, the volume of white matter has increased disproportionately. In human beings, the white matter tracts occupy some 42% of the volume of the hemispheres. The great majority of these fibers serve cortico-cortical connectivity rather than projections between cortical regions and subcortical sites; for example, thalamic input is estimated to represent only 5% of the total input into the primary sensory cortex, remainder being from other cortical areas.

The fibers in white matter are of several types. First are the longer intrahemispheric fiber tracts:

- arcuate fasciculus, which connects superior and middle frontal gyri to the temporal lobe and (via a superior portion of the fasciculus called superior longitudinal fasciculus) the parietal and occipital lobes;
- uncinate fasciculus, which connects the orbitofrontal cortex to the temporal cortex and (via an inferior portion of the fasciculus called inferior occipitofrontal fasciculus) the occipital lobe;
- cingulum, which lies medially beneath the cingulate cortex in the cingulate gyrus and connects frontal and parietal lobes with the parahippocampal gyrus and

adjacent structures.

Second are the long projection systems linking the cortex, subcortical nuclei, and lower portions of the neuraxis. The medial forebrain bundle is the primary connection between limbic structures and the brainstem and carries projections from the monoaminergic cells in the midbrain and pons. Others are the thalamic peduncle, with reciprocal fibers between the thalamus and parietal lobe, and the corticopontine and corticospinal tracts descending through corona radiata and internal capsule. Fibers from the prefrontal cortex descend into the anterior limb of the internal capsule so that lesions there may have predominant behavioral and a paucity of elementary sensorimotor effects.

Lacunes and degeneration of the white matter due to hypertensive small vessel disease (along with strokes, known as Binswanger disease) interrupt these cortico-cortical fibers and cortico-subcortical projections. The result of progressive loss of communication among cortical regions and between cortex and subcortical gray matter is the clinical state of subcortical dementia, which is prominently characterized by slowing of mental processing and failure of executive control processes. The latter may be explained in part by the preferential occurrence of lacunes in frontal locations but also by the impairment of connectivity.

Third, U-fibers are the short, juxtacortical fibers connecting adjacent cortical regions. These fibers are characteristically spared in Binswanger disease, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), and certain other disease processes.

Fourth are the many specific projection systems linking delimited regions, such as the mammillothalamic tract, which connects mamillary bodies with the anterior nucleus of the thalamus, and the fornix, which connects mammillary bodies with the hippocampus. Interesting neurobehavioral syndromes have been described related to rare cases of focal interruption of such pathways. For example, interruption of the mammillothalamic tract or of the fornix is implicated in amnesia.

Fifth, several pathways connect the two hemispheres, notably the corpus callosum but also anterior and posterior commissures and massa intermedia of the thalamus. Syndromes due to interruption of the smaller commissures have not yet been described, although the absence of massa intermedia is reported to be associated with schizophrenia in women, and anterior commissure and massa intermedia are larger in women than in men. Corpus callosum is congenitally absent in numerous neurodevelopmental syndromes, and its absence has been associated with schizophrenia. Congenital absence is not, however, associated with the interesting disconnection symptoms seen in lesional interruption of the callosum, such as by anterior or posterior cerebral artery stroke or by surgical callosotomy for control of epilepsy.

Two callosal disconnection syndromes are worthy of specific mention. After anterior cerebral artery occlusion with anterior callosal infarction, the right hemisphere is deprived of verbal information; left-hand apraxia is seen, and the patient cannot name unseen objects placed in the left hand. Reciprocally, the right hand shows constructional apraxia. This is termed the anterior disconnection syndrome. After occlusion of the left posterior cerebral artery with infarction of the left occipital lobe and splenium (posterior portion) of the corpus callosum, the language cortices of the left hemisphere lose access to visual information: The left visual cortex is damaged, as are the projections from the right visual cortex, which cross in splenium. Thus, reading becomes impossible, although other language functions are unaffected—the syndrome of alexia without agraphia.

Modulators of Brain States

This account of cognitive processing in the cortex will seem to many psychiatrists to leave out of consideration the matters with which they are most concerned, namely pervasive states of altered mood, drive, and behavior. That such states are behaviorally pervasive does not entail that they are anatomically global. The limbic structures discussed above provide in part the anatomic substrate for emotional states. Further, several systems with diffuse cortical projections have the capacity to modulate processing in widespread brain regions. These originate in:

- Intralaminar thalamic nuclei, which project to the cortex (especially prefrontal and cingulate cortex) and to the striatum;
- Histaminergic cells in posterior hypothalamus;
- Serotonergic cells in pontine raphe nuclei;
- Noradrenergic cells in locus ceruleus;
- Dopaminergic cells in the midbrain ventral tegmental area, giving rise to the mesocortical and mesolimbic systems;
- Cholinergic cells in basal forebrain nuclei, such as nucleus basalis of Meynert.

The last of these is of relevance to cholinesterase inhibitor treatment of dementia and the preceding three are of relevance to the treatment of mood, anxiety, and psychotic disorders. The hypothalamic histaminergic projections are involved in arousal. “Nonspecific” thalamic projections may have an important role in executive dysfunction seen after thalamic lesions.

Brain Networks

The description above of white matter tracts is an account of *structural* connectivity. Correlations of activity as measured by functional imaging yield patterns of *functional* connectivity. A finite number of circuits so-defined have been identified, some clearly of relevance to behavior:

- Default mode network (DMN): involving the medial prefrontal cortex, posterior cingulate cortex, precuneus, and angular gyrus
- Salience network: centered on anterior cingulate and anterior insular cortices, with nodes in the amygdala, ventral striatum, and thalamus
- Central executive network: linking dorsolateral prefrontal and parietal regions.

These networks are also distinguished by atrophy patterns in neurodegenerative disease. In Alzheimer disease, the DMN shows atrophy corresponding to deficits in episodic memory. The salience network is affected by behavioral-variant frontotemporal dementia (bvFTD), with deficits in emotional and compartmental processing.

This way of looking at brain function—as the output of multiple regions organized as networks—begins to solve a problem posed by studies of lesion localization: not all cases have lesions in the presumed “right” location. The technique of lesion network mapping looks at the connectivity (as known from studies of the connectome of healthy control subjects) of each of the lesions in patients with a given disorder and then examines the overlap of those connectivity maps rather than the overlap of the lesions themselves. The technique has been used to show the underlying brain networks for disorders as various as hallucinations, movement disorders, depression, mania, free will, and criminal behavior.

An additional aspect of the network approach relates to individual differences in brain function. Studies of groups of affected subjects compared with healthy controls may reveal regions underlying a given function or deficit. Only large data sets of healthy subjects studied over time can reveal that different individual subjects perform the task by enlisting different networks, putatively reflecting different strategies in performing the task, different strategies that are trait-like and stable over time for each person. Individual differences in the organization of the language cortex are evident clinically in the unusual, but not negligible, occurrence of crossed aphasia (aphasia due to right hemisphere injury in a dextral), crossed nonaphasia (lack of aphasia in a dextral with a left hemisphere injury that “should” cause aphasia), and aphasic deficits anomalous in respect to the predicted effects of lesions in both dextrals and sinistrals. While this approach is in its early stages, the day may come when our understanding of the varying instantiation of function by brain networks is able to reflect the complexity of human behavior as studied in neuropsychiatry.

MODULARITY AND NEUROPSYCHIATRY

Focal behavioral syndromes compel attention to local processing in the brain and almost irresistibly suggest a particular model of brain organization. One imagines a box-and-arrow diagram in which each box—representing an elementary cognitive function—maps onto a specialized region of the cortex. Each area of the cortex has its job to do and a lesion of any area produces a distinctive, delimited, and predictable deficit.

This model raises the issue of the modular organization of the brain. The general topic of modularity in cognitive processing deserves further consideration because it is crucial to the theoretic perspective of neuropsychiatry, and in particular because it bears on the value of neuropsychiatric data for the understanding of idiopathic psychiatric disorders. Modularity in cognitive neuroscience refers to a brain organization characterized by multiple computational devices, each of which operates on characteristically encapsulated input with prewired (perhaps innate) rules, thus being rapid, efficient, and reliable. For example, elementary visual processing can be considered modular inasmuch as it utilizes restricted input with hardwired feature extraction (e.g., motion, color, shape). In another domain, consider that it is easier to teach an animal to associate a taste than a visual stimulus with the aversive effects of an ingested toxin. This finding implies domain-specific, innate learning constraints. The classic cognitive example of domain-specific prewiring is language, for example, Chomsky's observation that children generate language errors that they have never heard: "He bringed me here," the small child might say, although they have never heard an adult say "bringed." The implication is that a language-processing module possesses innate grammatical rules that have generated a grammatical form without an experiential foundation.

Evolutionary psychologists have forcefully argued the case for modular processing, as opposed to domain-general problem-solving devices. The core of the evolutionary argument is that cerebral organization is the result of natural selection operating on the adaptational fitness of mankind's Pleistocene hunter–gatherer ancestors. Domain-specific processing has advantages of speed and efficiency that necessarily lead to an advantage in fitness. The availability of pre-experiential information about the content of domain-specific processing carries a large advantage over the "combinatorial explosion" of informational possibilities requiring evaluation by a domain-general processor. For example, detection of cheating in social exchanges is an essential element of adaptation in a population group featuring cooperative behavior. Is it a function of a domain-general logical problem-solving device, or is there a cheater-detection module? Cross-cultural evidence shows that people are better at detecting violations of social exchange rules than at solving problems of equivalent logical complexity when posed in other terms, and focal lesions can differentially affect cheater detection. The implication is that prewired mechanisms, presumably located in a particular brain area, are "tuned" to recognize and reason about this adaptationally crucial behavior, just as innate mechanisms subserve language learning and toxin recognition. One of the strengths of the evolutionary approach is to direct attention to those processing domains, the modularity of which is plausible on adaptational grounds.

However, many of the "modules" that have attracted clinical interest are not plausibly directly the objects of natural selection. Reading and writing are clear examples. These have arisen too recently in evolutionary time to have been the product of natural selection and thus must depend on the workings of processors that are, at least to this extent, domain-general.

Moreover, much of the literature on modularity is written from a cognitive psychological or philosophical perspective, with less attention to the "wetware" (i.e., actual brain substance) implementing the processing devices. A foundation in cognitive neuroscience and evolutionary biology can enrich clinical theories, but it creates the potential for misunderstanding by clinicians interested in the functioning of patients with brain lesions. The modules of the evolutionary biologists and philosophically inclined cognitivists do not map directly to brain areas. One of the striking results of functional neuroimaging experiments is that, although the function under study is delimited, multiple areas of brain activation are found. Even functions that seem psychologically fundamental are not implemented in a simple way, and local processing components may be recruited into multiple networks subserving a variety of tasks. This seems to be the case with respect to the limited number of frontal sites involved in a wide range of executive tasks. The specialization of regions is dependent on input from other regions; specialization is not entirely dependent on intrinsic properties but partially on top-down influences.

The issue is not whether different cerebral regions carry out different modes of information processing. This is unquestionably so, and neither unreconstructed holists who believe in the equipotentiality of the cortex nor strict localizationists who believe only in fully autonomous processing devices figure on the current neuroscientific scene. The question is how regions are linked in carrying out tasks. Functions are implemented by networks, most or all of the nodes of which participate in multiple functional networks. This pattern of cerebral organization has been termed selectively distributed processing or sparsely distributed networks. Although cortical regions have specialized capacities for information processing, functions cannot be localized to regions (as Hughlings-Jackson explicitly warned a century and a half ago). It would be erroneous, for example, to believe that an area crucial for face recognition contained all of the neurons, and only neurons, that respond to faces.

Another critique of the modularity hypothesis crucial for neuropsychiatry derives from developmental psychology. Trenchant arguments contradict the assumption that a mapping of deficits to specific brain structures could be static over developmental time. On the contrary, how the brain performs cognitive tasks changes with development. Development entails changing patterns of interaction among brain components, and localization may alter as neurons and regions become “tuned” in responsiveness, based on their initial characteristic processing biases and their patterns of inputs and connectivity. This reorganization of cortical function could mean that the same behavior has different substrates at different developmental epochs. For example, in adult subjects with Williams syndrome, poor function at number processing and good language skills are characteristic; however, in infancy, the opposite pattern is seen. Whatever the fundamental processing disorder of genetic origin may be, it cannot be seen as having knocked out a module. A large expanse of nonlinear brain development lies between the gene and the clinical phenomena, an expanse that can be understood only with a better theory than neophrenology. The very development of modularity can be anomalous. Indeed, in the Williams syndrome cases, functional magnetic resonance imaging (fMRI) data disclose an anomalous, diffuse pattern of activation for music perception, an area of preserved or enhanced ability in these patients. Focal syndromes in adults provide an appropriate place to start formulating hypotheses, but a deficit seen in an idiopathic disorder cannot be assumed to have its basis in dysfunction in the same simple locus as a phenomenologically similar deficit seen after a focal brain lesion occurring in an adult.

Nothing in this line of argument diminishes the interest in focal neurobehavioral syndromes, which are clinical facts that have a substantial heuristic value for cognitive neurosciences. Neuropsychiatry, along with other brain specialties, has the task of importing into clinical theory the understanding of the mind and brain elaborated in cognitive neuroscience. The search for psychopathological understanding based on the identification of deficits in cognitive modules that are relatively well understood in normal subjects has been termed cognitive neuropsychiatry. This pursuit inevitably results in the deconstruction of the psychiatric diagnoses of the Diagnostic and Statistical Manual of Mental Disorders (DSM) or the International Statistical Classification of Diseases and Related Health Problems (ICD) into symptoms or syndromes because the standard diagnostic categories are generally based on folk-psychological notions (such as the division between “thought” and “affective” disorders).

Much of this section is devoted to the anatomical mode of thought practiced by neuropsychiatrists. However, some clinicians hope that neuropsychiatry will provide a localizing taxonomy of behavioral syndromes so that particular psychiatric disorders will carry the same localizing power as, say, the Babinski sign for the corticospinal tract: the nuclear syndrome of schizophrenia to the left temporal lobe, for example. From the contemporary cognitive neuroscience perspective just reviewed, this seems likely to be false hope. The Babinski sign is a limiting, not a paradigmatic, case of brain-behavior relationships. For the fullest understanding of complex mental syndromes, notably those traditionally in the realm of psychiatry, a more adequate theory of brain function is needed than can be offered by the localizationist tradition.

CLINICAL EVALUATION

The neuropsychiatric perspective places great reliance on information that can be gathered at the bedside. No practical inquiry and examination can include all possible items; rather, the clinician selects from a toolbox of probes of the history and the patient’s functioning in the examination room in order to confirm or refute hypotheses generated by the emerging clinical picture. Screening items should have high sensitivity but not necessarily high specificity. Beyond screening, elements of the history and examination that might potentially elucidate the nature of the disease process under consideration form the entire corpus of medical assessment.

The Neuropsychiatric History

The initial steps in screening for the presence of organic disease in patients with mental symptoms are easily taken. The physician should obtain a general medical history, including a history of diseases possibly relevant to the neuropsychiatric symptoms under consideration and a review of systems in potentially relevant areas. With a cognitively impaired or psychotic patient, such history-taking may be unreliable. An account from a family member or other informant and a review of medical records are almost always essential.

With virtually every patient, the clinician should inquire as to a history of (1) heart, lung, liver, kidney, skin, joint, or eye disease; (2) hypertension; (3) diabetes; (4) traumatic brain injury; (5) seizures, including febrile convulsions in childhood; (6) sleep disturbance; (7) unexplained medical symptoms; (8) substance use; (9) current medication; and (10) family history of neuropsychiatric disorder. The inquiry about these disorders in some settings can be quite general. For example, the question "Have you ever had heart problems?" along with a few questions in the review of systems may suffice to screen for heart disease in a young, apparently healthy patient. In other circumstances, more detailed information must be gathered.

The review of systems as well should vary according to the setting. Positive responses should lead to further inquiry. The clinician should be practiced at inquiring about:

- Constitutional symptoms: fever, malaise, weight loss, pain complaints;
- Neurologic symptoms: headache, blurred or double vision, impairment of balance, impairments of visual or auditory acuity, swallowing disturbance, focal or transient weakness or sensory loss, clumsiness, gait disturbance, altered urinary or defecatory function, altered sexual function;
- Paroxysmal limbic phenomena: micropsia, macropsia, metamorphopsia, déjà vu, jamais vu, forced thoughts or emotions, depersonalization/derealization, autoscopy, paranormal experiences such as clairvoyance or telepathy;
- Thyroid symptoms: heat or cold sensitivity, constipation or diarrhea, rapid heart rate, alopecia or change in the texture of hair;
- Rheumatic disease symptoms: joint pain or swelling, mouth ulcers, dry mouth or eyes, rash, past spontaneous abortions.

Birth History and Early Development

Because brain development starts before birth, so too does the neuropsychiatric history. The clinician should note:

- Maternal substance use, bleeding, and infections during the pregnancy;
- The course of labor;
- Fetal distress at birth, including Apgar scores if available;
- Perinatal infection or jaundice;
- Motor and cognitive milestones, such as the age at which the child crawled, walked, spoke words, spoke sentences;
- The infant's temperament;

- The child’s school performance (including special education and anomalous profiles of intellectual strengths and weaknesses) is usually the best guide (absent psychometric data) to premorbid intellectual function.

The role of perinatal injury in cerebral palsy and intellectual disability has commonly been overestimated; in many instances, the developmental disorder was present in gestation prior to the perinatal misadventure, which may, in fact, have arisen from the pre-existing abnormality. However, perinatal injury, in particular hypoxic injury, is probably associated with later schizophrenia.

Head Injury and Its Sequelae

Head injury is a potential factor in late-onset mood and psychotic disorders, as well as cognitive impairment, epilepsy, and post-traumatic stress disorder (PTSD). The clinician should inquire about a history of head injury in virtually every patient. The nature of the injury should be clarified by eliciting the circumstances, including risk-taking behaviors that may have predisposed to injury and others who were injured in the same incident, often an emotionally powerful aspect of the event. Loss of consciousness is not a prerequisite to important sequelae; even a period of being stunned, “seeing stars,” can presage later neuropsychiatric symptoms. The period of loss of consciousness, or coma, should be established, ideally with the assistance of contemporaneous medical records. The period of retrograde amnesia —from the last memory before the injury to the injury itself—and of anterograde amnesia—from injury to the recovery of the capacity for consecutive memory—should be noted.

Paroxysmal Disorders

Paroxysmal disorders of neuropsychiatric interest include epilepsy, migraine, panic attacks, and episodic dyscontrol of aggression. Taking a history of an episode has common features irrespective of the nature of the disorder. The clinician should track the chronology of the attack. This starts with the possible presence of a prodrome, a warning of an impending episode in the hours or days prior to one. The episode itself may be presaged by an aura lasting seconds to minutes. In the case of an epileptic seizure, this represents the core of the seizure itself and may carry important localizing information about the hemispheric side and site of the focus. The pace of buildup, from onset to peak of the ictus, is of differential diagnostic importance. For example, epileptic seizures begin abruptly; panic attacks may have a more gradual development to peak intensity. The mental and behavioral features of the ictus itself should be elicited in detail, if possible, from collateral informants as well as from the patient. The duration of the spell and the mode of its termination should be elicited. Inquiring whether the patient has just one sort of spell or more than one is an essential prelude to establishing the frequency of episodes, both at present and at maximum and minimum in the past.

By interviewing the patient and collateral informants, information necessary to make a differential diagnosis can usually be elicited. The differential diagnosis between epilepsy and psychogenic nonepileptic seizures (PNES) can be difficult; but at times, if asked properly, the patient will make the diagnosis for the clinician by reporting “two kinds of seizures,” one of which is clearly epileptic and the other whose semiology is inconsistent with epileptic seizures.

Cognitive Symptoms

Recognizing cognitive symptoms in patients without established dementia is a crucial element of neuropsychiatric history-taking. Such symptoms may be outweighed by more dramatic behavior or mood change, but identification of cognitive impairment can reorient the diagnostic evaluation of late-life depression, for example. No doubt, the commonest complaint along cognitive lines is memory problems. In the setting of depression, the more intense the complaint of memory impairment, the less likely it is to have an organic basis and the more likely to testify to depressive ideation and attentional failure. The clinician should establish whether forgotten material (say, an acquaintance’s name or a task meant to be performed) comes to the patient later as a matter of absentmindedness rather than memory failure; recall when cued by a conversational partner likewise suggests memory impairment at the retrieval stage. Certain other complaints are highly characteristic of organic disease. These include a loss of the capacity for divided attention or for the automatic performance of familiar tasks. A patient might report, for example, no longer being able to read and listen to the radio at the same time. Getting lost or beginning to use aids for recall, such as a notebook, is suggestive of organic cognitive failure. A discrepancy between the examination and the patient’s everyday functioning (e.g., a patient who follows his favorite sports teams effectively but in the consulting room cannot count to 10) should be noted. Patients with memory or other cognitive deficits should be assessed for the safety of their living arrangements.

Appetitive Symptoms and Personality Change

Alterations of sleep, appetite, and energy are common in idiopathic psychiatric disorders as well as transiently in the healthy population and cannot be interpreted as implying brain disease. Certain patterns of altered sleeping and eating behavior and personality, however, are pointers to organic disease. Excessive daytime sleepiness or sleep attacks raise the question of sleep apnea or narcolepsy, or in a different temporal pattern, Kleine–Levin syndrome. Sleep apnea should be routinely screened for, with questions about daytime sleepiness, snoring, and morning headache, because it is a potential contributor to cognitive impairment and treatment resistance in depression. Abnormal behavior during sleep raises the question of parasomnia. Dream-enacting behavior during rapid eye movement (REM) sleep, accompanied by loss of the characteristic atonia of REM sleep, is termed REM-sleep behavior disorder. Typically the behavior is aggressive—flailing the arms, kicking, punching—and corresponds to aggressive dream content. This may be due to a pontine lesion or an autoimmune encephalopathy, but the absence of such disorders strongly points to an ingravescent synucleinopathy such as Lewy body disease. Much more rarely, nocturnal dream-enacting behavior, the so-called oneiric stupor, portrays with simple motor behaviors patient's daytime activities (such as, in one personal patient, arm movements like an orchestra conductor's, the patient, in fact, being an orchestra conductor). This usually discloses a prion disease, namely fatal familial insomnia. Loss of dreaming occurs with parietal or bifrontal damage; loss of visual imagery in dreams occurs with ventral occipitotemporal damage, part of the Charcot–Wilbrand syndrome (loss of visual imagery with brain damage).

In medial hypothalamic disease, eating behavior is marked by a lack of satiety and resultant obesity. In the Klüver–Bucy syndrome of bilateral anterior temporal damage (involving the amygdala), patients mouth nonfood items. With frontal damage, patients may stuff food into the mouth, a form of utilization behavior, sometimes with alarming or even fatal consequences. A change in food preferences with a new "sweet tooth" suggests frontotemporal dementia. A "gourmand" syndrome of excessive concern with fine eating has been associated with right anterior injury.

Changes in sexual behavior are common consequences of brain disease. Hyposexuality is common in epilepsy, possibly as a consequence of limbic discharges. A change in habitual sexual interests, quantitative or qualitative, developing in midlife suggests organic disease. It is possible, although understudied, that relevant organic disease, such as the sequelae of traumatic brain injury, is common in sexual offenders. Other changes in personality, such as the development of shallowness of affect, irritability, loss of sense of humor, reduced empathy for intimates, or a coarsening of sensibilities, may indicate ingravescent organic disease, for example, frontotemporal dementia.

Handedness

About 90% of people designate themselves as dextral, almost all the rest as sinistral, and a very few as ambidextrous. The true state of affairs is somewhat more complicated in that handedness may be considered dimensionally (i.e., as a matter of degree rather than categories). Patients may call themselves right-handed but use the left hand preferentially for certain tasks. Inquiring about a few specific tasks—writing, throwing, drawing, using scissors or toothbrushes—yields helpful information. A family history of sinistrality may also be relevant.

THE NEUROPSYCHIATRIC PHYSICAL EXAMINATION

In principle, any aspect of the general physical or neurologic examination may be relevant to neuropsychiatric diagnosis, if only in revealing an incidental clinical problem in a neuropsychiatric patient. Here, the focus is on elements of the physical examination with specific relevance to the detection and identification of organic disease in patients with mental presentations. The mental examination, including the cognitive examination, is discussed below in association with syndromes of behavioral disorder and insofar as it can elicit or elucidate these syndromes in the consultation room.

The General Physical Examination

General Appearance. Dysmorphic features include so-called minor physical anomalies, some of which are captured in the widely used Waldrop scale. These are associated with developmental disorders, including schizophrenia. These features center on the head, hands, and feet. No single minor anomaly is diagnostic of pathologic development, only the concurrence of multiple anomalies.

Many specific developmental disabilities syndromes can be diagnosed by the constellation of dysmorphic features presented. Cleft lip or palate may be associated with brain malformations and frontal cognitive impairment. Asymmetry of the extremities, often best seen in the thumbnails or of the cranial vault, points to a developmental abnormality. Occasionally, a patient even reports wearing shoes of different sizes on the two feet. The larger extremity and the smaller side of the head are ipsilateral to the abnormal cerebral hemisphere. Short stature is an important feature of many developmental syndromes, both common, such as fetal alcohol syndrome and Down syndrome, and uncommon, such as mitochondrial cytopathies. Abnormal habitus, such as the marfanoid habitus of homocystinuria, may be a clue to the diagnosis.

Vital Signs. Elevated temperature, blood pressure, or heart or respiratory rate should never be ignored, even in a patient whose agitation or anxiety might seem to explain the abnormality. Doing so risks missing infection, neuroleptic malignant syndrome, serotonin syndrome, connective tissue disease, or other important causes of morbidity. Abnormal respiratory patterns occur in hyperkinetic movement disorders, including tardive dyskinesia. Yawning is a feature of opiate withdrawal and serotonergic toxicity. Weight loss is an important clue to systemic diseases, such as neoplasia; it should not be dismissed even in a patient with depression, which may—but may not—account for the weight loss. Weight gain equally may point to the brain or systemic disease, especially an endocrinopathy, or may reflect the toxicity of psychotropic drugs.

Skin. Alopecia or rash may point to systemic connective tissue disease. Alopecia is also a feature of drug toxicity and hypothyroidism (where thinning of the lateral part of the eyebrow is characteristic). The malar rash of systemic lupus erythematosus is typically slightly raised and tender and extends to both cheeks in a “butterfly” pattern while sparing the nasolabial folds. Discoid rashes in lupus are characterized by hyperkeratosis, atrophy, and loss of pigment; the strong tendency to scarring means that the presence of a discoid rash does not necessarily indicate active disease. A pink periungual rash is also characteristic of lupus. A vasculitic rash is classically palpable purpura and may be seen in lupus or other connective tissue diseases. Livedo reticularis, a net-like violaceous pattern on the trunk and lower extremities, is not specific but raises the question of Sneddon syndrome when stroke or dementia is a clinical accompaniment. The neurocutaneous syndromes have typical skin manifestations: adenoma sebaceum (facial angiofibromas), ash-leaf macules, depigmented nevi, and shagreen patches (thickened, yellowish skin over the lumbosacral area) in tuberous sclerosis; a port-wine stain (typically involving both upper and lower eyelids) in Sturge–Weber syndrome; neurofibromas, café-au-lait spots, and axillary freckling in neurofibromatosis.

Head. Head circumference should be measured in patients with a question of developmental disorder. Most reference works give normal ranges for head circumference in developing children but not for adults. Fortunately, adequate data to establish normal ranges do exist. Although height and weight need to be taken into account along with sex, roughly the normal range for adult males is 54 to 60 cm (21.25 to 23.5 in); for females, 52 to 58 cm (20.5 to 22.75 in). An old skull fracture or intracranial surgery usually leaves palpable evidence.

Eyes. Exophthalmos usually indicates Graves disease; especially if unilateral, it may reveal a space-occupying lesion. Dry eyes, along with dry mouth, raise the question of Sjögren syndrome, although drug toxicity or the aging process are common confounds. Inflammation in the anterior portion of the eye, uveitis, can be appreciated at the bedside by the presence of pain, redness, and a constricted pupil; this is commonly associated with connective tissue disease. The Kayser–Fleischer ring is a brownish-green discoloration at the limbus of the cornea; it sensitively and specifically indicates Wilson disease. The examination of pupils, optic discs, visual fields, and eye movements is discussed below.

Mouth. Oral ulcers can be seen in lupus, Behçet disease, and other connective tissue diseases. Dry mouth is a part of the sicca syndrome, along with dry eyes discussed above. Vitamin B_{12} deficiency produces atrophic glossitis, a smooth, painful, red tongue.

Heart and Vessels. A carotid bruit indicates turbulent flow in the vessel but is a poor predictor of the degree or potential risk of the vascular lesion. A thickened, tender temporal artery points to giant cell arteritis; here, the physical examination is an excellent guide to clinical significance. Cardiac valvular disease, marked by cardiac murmurs, is important in assessing the cause of stroke and congestive heart failure may be relevant in delirium. In a schizophrenic patient, a murmur may imply the velo-cardio-facial syndrome (a deletion at 22q11, also called DiGeorge syndrome). Patients with developmental disabilities may have multiple anomalies, including structural heart disease.

Extremities

Joint inflammation as a pointer to systemic rheumatic disease is distinguished from noninflammatory degenerative joint disease (osteoarthritis) by the presence of swelling, warmth, and erythema and is characteristically seen in wrists, ankles, and metacarpophalangeal joints, as opposed to the involvement of the base of the thumb, distal interphalangeal joints, and spine in degenerative joint disease. Raynaud phenomenon and sclerodactyly are signs of connective tissue disease. Joint laxity, such as is seen in the Ehlers–Danlos syndrome, is surprisingly associated with anxiety disorders.

The Neurologic Examination

Olfaction. Hyposmia is common in neurologic diseases but even more common in local disease of the nasal mucosa, which must be excluded before a defect is taken to be of neuropsychiatric significance. Assessment of olfaction is often ignored ("cranial nerves II through XII normal"), but it is easily performed and gives clues to the integrity of regions otherwise hard to assess, notably the orbitofrontal cortex. The olfactory nerve lies underneath the orbitofrontal cortex; projections go to the olfactory tubercle, entorhinal, and piriform cortex in the temporal lobes, amygdala, and orbitofrontal cortex. Testing of olfaction is best performed using a floral odorant, such as scented lip balms, which are inexpensive and simple to carry. Although a distinction can be made between the threshold for odor detection and that for identification of the stimulus, with differing anatomies, at the bedside without special equipment, the best one can achieve is recognition of a decrement in sensitivity (i.e., whether the patient smells anything, even without being able to identify it).

Eyes. Pupillary dilation may indicate anticholinergic toxicity; pupillary constriction is a characteristic feature of opiate toxicity. Argyll Robertson pupils are bilateral, small, irregular, and reactive to accommodation but not to light; the finding is characteristic of paretic neurosyphilis but also present in other conditions. Papilledema indicates elevated intracranial pressure; the earliest and most sensitive feature is the loss of venous pulsations at the optic disc. A homonymous upper quadrantic field defect is present when temporal lobe disease affects the loop of Meyer, the portion of the optic radiation that dips into the temporal lobe. A field defect in a delirious patient may point to an etiology in focal vascular disease (as discussed below). The normal spontaneous blink rate is 16 ± 8 per minute. Hypodopaminergia is associated with a reduction in blink rate. Impairment of voluntary eye-opening is seen in association with extrapyramidal signs, making the common appellation "apraxia" of eye-opening a misnomer. Impairment of voluntary eye closure is seen after frontal or basal ganglia damage.

Both saccadic and pursuit eye movements should be examined. The former is assessed by asking the patient to look to the left and the right, up and down, and at the examiner's finger on the left, right, up, and down. Pursuit eye movements are examined by asking the patient to follow a moving stimulus in both the horizontal and vertical planes. These maneuvers test supranuclear control of eye movements; the oculocephalic maneuver (doll's head eyes), that is, moving the patient's head tests the brainstem pathways and may be added to the assessment if saccades or pursuit is abnormal. Limitation of voluntary upgaze is common in the normal older adults. A limitation of voluntary downgaze, however, in a patient with extrapyramidal signs or frontal cognitive impairment may suggest progressive supranuclear palsy. Slowed saccades are characteristic of Huntington disease. Impairment of initiation of voluntary saccades, requiring a head thrust or head-turning, amounts to apraxia of gaze and is seen in developmental disorders as well as Huntington disease and parietal damage. Contrariwise, impairment of inhibition of saccades represents a visual grasp, with forced gaze at environmental stimuli. This can be usefully tested by placing stimuli (a finger and a fist) in the left and right visual fields of the patient and asking the patient to look at the fist when the finger moves and vice versa. The patient's inability to perform horizontal pursuit or saccadic movements without turning the head (head-eye synkinesia) may represent the same impairment of inhibition.

Facial Movement. Both spontaneous movements of emotional expression and movement to command should be tested. In pyramidal disorders, spontaneous movements may be relatively spared when the face is hemiparetic for voluntary movements. Contrariwise, in nonpyramidal motor disorders, voluntary movement may be possible despite a hemiparesis of spontaneous movement. The latter situation is seen *inter alia* in temporal lobe disease, including temporal lobe epilepsy, for which it has lateralizing value. Vertical furrows between the eyebrows are known as Veraguth folds and are associated with depression.

Speech. A number of speech abnormalities are listed in Table 2.1–1. A systematic examination of speech may include asking the patient to produce a sustained vowel ("ahhh"), the performance is assessed for voice quality, steadiness, and loudness; then strings of consonants ("puh-puh-puh") and alternating consonants ("puh-tuh-kuh-puh-tuh-kuh"), the performance is assessed for rate, rhythm, and clarity. Difficulty with the alternating sequence in the presence of normal production of individual consonants amounts to apraxia of speech.

Table 2.1–1.
Speech Syndromes

Syndrome	Output	Characteristic Lesion Location or Associations
Aphemia	Initial mutism, recovery without agrammatism	Broca area (BA44), foot of left third frontal gyrus
Apraxia of speech	Inconsistent and slowed articulation, flattened volume, abnormal prosody	Left insula
Ataxic dysarthria	Slowed, equalization of or erratic stress (scanning), imprecise articulation	Cerebellum, especially superior-anterior vermis, left hemisphere to right
Pyramidal dysarthria	Slowed, strained, slurred	Anterior hemispheres, usually bilateral; may be accompanied by pseudobulbar palsy (dysphagia, drooling, pathologic laughing and crying)
Extrapyramidal dysarthria	Hypophonia, monotony of intonation, tailing off with longer phrases	Basal ganglia
Bulbar dysarthria	Nasality, breathiness, slurred articulation	Brainstem
Expressive aprosodia	Loss of emotional “melody of speech”	Right hemisphere
Foreign accent syndrome	Phonetic and prosodic alterations like those of dysarthria after cortical damage but giving listener feeling of foreign accent	Motor or premotor cortex or subjacent white matter of left hemisphere
Developmental stuttering	Repetition, prolongation, arrest of sounds; if overcome in childhood, may re-emerge after stroke, onset of Parkinson disease	Various hemisphere sites
Acquired stuttering	No dystonic facial movements as are seen in developmental stuttering	Various hemisphere sites
Cessation of stuttering	Not an abnormality but the reversal of an abnormality	Various hemisphere sites
Echolalia	Automatic repetition of interlocutor’s speech or words	Various anatomies, but seen in frontotemporal

	heard in environment, sometimes with reversal of pronouns, correction of grammar, completion of well-known phrases	dementia, transcortical aphasias, other settings
Palilalia	Automatic repetition of own final word or phrase, with increasing rapidity and decreasing volume	Usually extrapyramidal system
“Blurting,” “echoing approval”	Automatic utterance of stereotyped or simple responses (e.g., “yes, yes”)	Frontal system

The mute patient poses a special problem in neuropsychiatric assessment. Mutism may occur at the onset of aphemia or transcortical aphasia due to vascular lesions, and it commonly develops late in the course of patients with frontotemporal dementia or primary progressive aphasia. The examiner should assess nonspeech movements of the relevant musculature, for example, tongue movements, swallowing, and coughing. Other means of communication should be attempted, such as gestures, writing, or pointing on a letter board or word board.

Abnormalities of Movement. The neuropsychiatric examiner should pay attention to weakness, abnormality of muscle tone, abnormal gait, and involuntary movements. Weakness due to muscle, peripheral nerve, or lower motor neuron disease is associated with atrophy, fasciculations, characteristic distributions, loss of reflexes, and tenderness in the case of muscle disease. Of greater relevance to cerebral mechanisms, pyramidal weakness, greatest in the distal musculature, is accompanied by increased muscle tone in a spastic pattern (flexors in the upper extremity, extensors in the lower extremity, with the sudden loss of increased tone during passive movement, the “clasp-knife” phenomenon), loss of control of fine movements, brisk tendon jerks, and the presence of abnormal reflexes such as the Babinski sign (discussed below). Less well recognized is the nonpyramidal motor syndrome, such as is seen in caudate or premotor cortical lesions: clumsiness, decreased spontaneous use of affected limbs, apparent weakness but production of full strength with coaxing. Mild degrees of impairment can be elicited with the pronator test by seeking pronation of the outstretched supinated arms; the forearm rolling test, by asking the patient to roll the forearms around each other first in one direction and then in the other, looking for one side that moves less thus appearing to be an axis that the other “orbits” around; rapid tapping of each hand and each foot; or fine finger movements, with the hands resting facing up on the thighs, the patient touching each finger to the thumb in turn and repeatedly.

Muscle tone can be increased not only in the pyramidal fashion just described but also as a manifestation of extrapyramidal or diffuse brain disease. In the latter case, paratonic rigidity, or Gegenhalten, is manifested by an erratic, “pseudoactive” increase in resistance to passive movement. The fluctuating quality of the resistance reflects the presence of both oppositional and facilitatory aspects of the patient's response to passive movement. The facilitatory aspect can be evoked by repeatedly flexing and extending the patient's arm at the elbow, then abruptly ceasing and letting go when the arm is extended; the abnormal response, facilitory paratonia, is for the patient to continue the sequence by flexion. In the case of extrapyramidal disease, the tone is increased in both extensors and flexors and throughout the range of movement, so-called lead-pipe rigidity. The "cogwheel" or ratchety feel to the rigidity is imparted by a coexisting tremor and is not intrinsic to the hypertonus; when paratonic rigidity co-occurs with a metabolic tremor, a delirious patient may mistakenly be thought to have Parkinson disease. Extrapyramidal rigidity can be amplified by activation techniques, such as movement of the limb contralateral to the one being examined (the Froment maneuver). However, this can occur bilaterally in healthy subjects, so it is most informative if unilateral.

Gait should always be tested, if only by focused attention to the patient's entering or leaving the room. Attention should be paid to the patient's station, postural reflexes, stride length and base, and turning. Postural reflexes can be assessed by asking the patient to stand in a comfortable fashion, then pushing gently on the chest or back, with care taken to avoid a fall. Gait should be stressed by asking the patient to walk in a tandem fashion and on the outer aspects of the feet. This may reveal not only mild ataxia (representing cerebellar vermis dysfunction) but also asymmetric posturing of the upper extremity (in nonpyramidal motor dysfunction). A short-stepped, "magnetic" gait in the context of unimpaired elementary movements of the legs (e.g., showing how to pedal a bicycle) is considered apraxia of gait.

Akinesia is manifested by delay in initiation, slowness of execution, and difficulty with complex or simultaneous movements. Mild akinesia may be observed in the patient's lack of spontaneous movements of the body while sitting or of the face or elicited by asking the patient to make repeated large-amplitude taps of the forefinger on the thumb (looking for decay of the amplitude). Akinesia is characteristically accompanied by rigidity. These plus rest tremor and postural instability represent the core features of the parkinsonian syndrome, seen not only in idiopathic Parkinson disease (IPD) but in several other degenerative, "Parkinson-plus" disorders such as progressive supranuclear palsy and multiple system atrophy as well as in vascular white-matter disease. Parkinsonian signs are not uncommon in frontotemporal dementia. Rest tremor is less common in these other disorders than in IPD.

Dystonia is sustained muscle contraction with consequent twisting movements or abnormal postures. Typically dystonia in the upper extremity is manifested as hyperpronation, in the lower extremity as an inversion of the foot with plantar flexion. Dystonia may occur only with certain actions, such as writer's cramp; focally, such as blepharospasm or oculogyric crisis; or in a generalized pattern, such as torsion dystonia associated with mutations in the DYT1 gene. The symptoms and signs often do not comport with a naïve idea of how things should be in organic disease; only expert knowledge will suffice for recognition. For example, a patient with early torsion dystonia may be able to run but not walk because the latter action elicits leg dystonia. Or a patient with intense neck muscle contraction may be able to bring the head to the midline by a light touch on the chin, a *geste antagoniste* diagnostic for dystonia.

A tremor is a regular oscillating movement around a joint. In rest tremor, the movement occurs in a relaxed, supported extremity, and is reduced by action. Often an upper extremity rest tremor is amplified by ambulation. The frequency is usually 4 to 8 Hz. This is the distinctive tremor of Parkinson disease. In postural tremor, sustained posture elicits tremor. It may be made more visible by placing a sheet of paper over the outstretched hand. Hereditary essential tremor presents as postural tremor, predominantly in upper extremities but also at times involving the head, jaw, and voice. A coarse, irregular, and rapid postural tremor is often seen in metabolic encephalopathy. In intention tremor, the active limb oscillates more prominently when approaching its target, such as touching with the index finger the examiner's finger. Maximizing the range of the movement increases the sensitivity of the test. Intention tremor is one form of kinetic tremor, which is a tremor elicited by movement; another sort of kinetic tremor is that elicited by a specific action, such as writing tremor or orthostatic tremor upon standing upright. The examiner can characterize tremor by observing the patient with arms supported and fully at rest, then with arms outstretched and pronated, then with arms abducted to 90 degrees at the shoulders and bent at the elbows while the hands are held palms down with the fingers pointing at each other in front of the chest. The patient should also be observed during ambulation. Anxiety exaggerates tremor; this normal phenomenon, for example, when the patient is conscious of being observed, should not be mistaken for a functional etiology. A good test for psychogenic tremor relies on the fact that although organic tremor may vary in amplitude, it varies little in frequency. A patient can be asked to tap a hand at a frequency different from the tremor frequency; if another tremulous body part entrains to the tapped frequency, a psychogenic tremor is likely.

Choreic movements are random and arrhythmic movements of small amplitude that dance over the patient's body. They may be more evident when the patient is engaged in an activity such as ambulation. When the movements are of large amplitude and forceful, the disorder is called ballism. Ballistic movements are usually unilateral.

Myoclonus is a sudden, jerky, shock-like movement. It is more discontinuous than chorea or tremor. The negative of myoclonus is asterixis, a sudden lapse of muscle contraction in the context of attempted maintenance of posture. Both phenomena, but more sensitively asterixis, are common in toxic-metabolic encephalopathy (not just hepatic encephalopathy). Asterixis should be carefully sought by observation of the patient's attempt to maintain extension of the hands with the arms outstretched ("like you are stopping traffic") because it is pathognomonic for organic disease and is never seen in acute idiopathic psychosis or other nonorganic disorders. Myoclonus is additionally an important feature of nonconvulsive generalized status epilepticus, serotonin syndrome, and Creutzfeldt–Jakob disease (CJD). Unilateral asterixis may be rarely seen in parietal, frontal, or (most often) thalamic structural disease.

Tics are sudden, jerky movements as well, but they may be more complex than myoclonic jerks and are subjectively characterized by an impulse to perform the act and a sense of relief for having done so (or mounting tension if restrained from doing so). Compulsions are not easy to differentiate from complex tics; the tiqueur may, like the patient with compulsions, report deliberately performing the act. Repetitive behavior superficially like compulsions may occur in organic disease but represent environment-driven behavior rather than having the same subjective structure as compulsive behavior. For example, a patient with frontal disease may repeatedly touch an alluring object without an elicitable subjective impulse and without anxiety if separated from the object. Organic obsessions and compulsions occur as well and have been associated with globus pallidus lesions, among others.

Akathisia is defined by both its subjective and its objective features. The patient exhibits motor restlessness, for example, by shifting weight from foot to foot while standing, and expresses an urge to move. At times psychotic or cognitively impaired patients cannot convey the subjective experience clearly, and the examiner must be alert for the objective signs in order to differentiate akathisia from agitation due to anxiety or psychosis. The complaints and the signs in akathisia are referable to the lower, not the upper, extremities; the anxious patient may wring the hands, and the akathisic patient shuffles the feet. Myoclonic jerks of the legs may be evident in the recumbent patient. The phenomenon occurs in IPD and with drug-induced dopamine blockade, but also rarely with extensive frontal or temporal structural lesions.

Ataxia is a disorder of coordinating the rate, range, and force of movement and is characteristic of diseases of the cerebellum and its connections. In the limbs, dysmetria represents a disordered determination of the distance to be moved so that the patient overshoots or undershoots the target; if the reaching limb oscillates in the process, the clinician observes an intention tremor. Asking patients to touch the examiner’s finger, then their own nose tests this system. Accurately touching one’s own nose with eyes closed requires both cerebellar and proprioceptive function. Eye movements also may be hypermetric or hypometric. The patient’s difficulty in performing rapid alternating movements, such as supination/pronation of the hand or tapping of the foot, is called dysdiadochokinesia. The failure of coordination of movement is also demonstrated by loss of check, which should not be elicited by arranging for the patient to hit themselves when the examiner’s hand is removed. In a normal situation, if the outstretched arms are tapped, only a slight waver is produced; the ataxic patient fails to dampen the movement. Gait may be affected by midline cerebellar (vermis) disease in the absence of limb ataxia, which is related to cerebellar hemisphere disease. Gait is unsteady, with irregular stride length and a widened base. (In the normal subject, the feet nearly touch at their nearest point; even a few inches of separation represent a widening of the base.) Gait and limb ataxia may be complemented by cerebellar dysarthria (described in Table 2.1–1) and by eye movement disorders, including nystagmus (usually gaze-paretic), slowed saccades, saccadic pursuit, and gaze apraxia.

The catatonic syndrome has been variously defined. The core of the syndrome is a mute, motionless state; variably added are abnormal movements, including grimacing, stereotypy, echopraxia, and catalepsy. The latter, also known as *flexibilitas cerea* (waxy flexibility), refers to posturing of a limb in the position in which it is placed by the examiner or in some other unnatural position. It is not seen in all or even most cases of catatonia, and it can be seen apart from the catatonic syndrome in patients with contralateral parietal lesions (as described below as the avoidance sign of parietal disease). Catatonic excitement refers to the sudden eruption into over-activity of a catatonic patient; this probably usually represents psychotic mania. The catatonic syndrome occurs in the course of schizophrenia or mood disorder as a manifestation of autoimmune encephalopathy due to antibodies to the NMDA receptor, in nonconvulsive generalized status epilepticus, in acute cerebral metabolic or structural derangements, or without other psychopathology as idiopathic catatonia. In the latter case, it is best thought of as a nonspecific reaction pattern, such as delirium, requiring a comprehensive clinical and laboratory evaluation to seek the cause of the behavioral disturbance. An important instance is catatonia as part of the neuroleptic malignant syndrome, the diagnosis of which requires exclusion of other metabolic encephalopathies, notably systemic infection. Catatonia is thus a medical emergency, requiring prompt attention to diagnostic evaluation as well as supportive care (fluids, nutrition, and measures to avoid complications of immobility, including venous thrombosis).

Motor sequencing tests assay the function of premotor cortical areas and striatum and are related to deficits in executive cognitive function seen with dysfunction of the dorsolateral prefrontal loop. The ring/fist test involves asking the patient to alternate between making a ring with his thumb and first finger and making a fist with the same hand (“ring, fist, ring, fist”). The abnormal response is perseveration of one or the other posture or disorganization of the sequence. At times, patients will be unable to perform the correct series even when repeating the verbal cues aloud. A more complex alternation is between striking the table gently with the fist, then the edge of the hand, then the palm (“fist, edge, palm, fist, edge, palm”). A different approach is to ask the patient to extend the arms, make a fist with one hand while keeping the other hand flat, and then switching hands. The abnormal response has the patient ending up with two fists or two palms outstretched.

A review of the material in this section on the motor system will reveal how much can be accomplished in the neurologic examination by asking the patient to stretch out their arms. With a few additional maneuvers (tapping the outstretched pronated hands, supinating them and asking the patient to close the eyes, then asking the patient to touch the nose with the eyes still closed, then asking the patient to perform the alternating fists test) all of the following can be assessed in a matter of a minute or so: postural and intention tremor, loss of check, asterixis and myoclonus, a pronator sign, dysmetria, and motor sequencing. Doing this, plus testing muscle tone, plus observing the patient's natural and stressed gait, and checking tendon jerks and abnormal reflexes, takes just a few minutes and does not elucidate disorders of muscle, nerve, and spinal cord but represents a rather extensive assessment of the central organization of the motor system.

Abnormalities of Sensation. Disorders of sensation are sometimes difficult to assess reliably in patients with cognitive and behavioral disorders. Nonetheless, several points should be familiar to the neuropsychiatrist. Distal loss of sensation, often accompanied by loss of ankle jerks, is characteristic of peripheral neuropathy. Often, all modalities of sensation are disturbed. If proprioception is sufficiently severely reduced, a Romberg sign will be present. The Romberg sign means that closing the eyes produces substantial impairment of balance; it is elicited by asking the patient to stand, allowing the patient to seek a comfortably balanced position, then asking the patient to close the eyes (ensuring against a fall).

Loss of sensation from sensory cortex injury is classically limited to complex discriminations, such as graphesthesia (recognizing numbers written on the palm), stereognosis (identifying unseen objects in hand), and two-point discrimination (telling whether the examiner is touching with one or two points, as these come closer together in space). However, patients with parietal stroke may have a pseudothalamic sensory syndrome (with impairments in elementary sensory modalities and subsequent dysesthesia) or other anomalous patterns of sensory loss. At times these patients will present with pseudomotor deficits: ataxia, fluctuating muscle tone and strength (dependent in part on visual cueing), "levitation," and awkward positioning of the arm contralateral (or at times ipsilateral) to the lesion. In the acute phase, the combination of deficits can amount to motor helplessness. These deficits result from the loss of sensory input to regions in which motor programs arise. The lessons here are that "cortical" sensory deficits should be sought if there is a question of cortical involvement and that more dramatic or unusual sensory abnormalities may also occur with cortical lesions.

Soft Signs. An extensive literature about the "soft signs" of neurologic dysfunction is difficult to comprehend because of the varied definitions and batteries used in the various studies. Most of the signs sought in these batteries are discussed in this section under their more specific headings, such as graphesthesia under abnormalities of sensation and the alternating fist (Oseretsky) test under motor sequencing. From the corpus of test batteries, a few simple maneuvers can be extracted that can contribute to the neurologic examination of the patient with a neuropsychiatric presentation.

While the patient is touching each finger to the thumb, as described above in the section on weakness, the examiner can watch the opposite hand for mirror movements. Obligatory bimanual synkinesia is seen specifically in disorders of the pyramidal pathways, such as the Klippel–Feil syndrome, and in agenesis of the corpus callosum, but also in putative neurodevelopmental disorders such as schizophrenia. Asking the patient, with eyes closed, to report whether the examiner is touching one or the other hand (with the patient's hands on the patient's lap), or one or the other sides of the face, or a combination, makes up the face–hand test. The examiner touches the left hand and right face simultaneously. If the patient reports only the touch on the face (i.e., extinguishing the peripheral stimulus), then the examiner can prompt (once), "Anywhere else?" Then the examiner touches the right hand and left cheek, left hand and left cheek, right hand and right cheek, both hands, and both cheeks. Extinction of the peripheral stimulus is the pathologic response and has been associated with schizophrenia and dementia.

Abnormal Reflexes. The Babinski sign is the shibboleth of the neurologic examination. It should be elicited by stroking the lateral aspect of the foot from back to front, with the leg extended at the knee, using a pointed object such as an orange stick or a key. The response of extension of the great toe with or without fanning of the other toes indicates corticospinal tract disease. Two confounding factors in the assessment of the Babinski sign are the striatal toe and the plantar grasp. The striatal toe is an extension of the hallux without fanning of the other toes or a flexion synergy in the other muscles of flexion of the leg. It may occur spontaneously or upon elicitation in patients with Parkinson disease in the absence of evidence of pyramidal dysfunction. The plantar grasp, the equivalent of the familiar palmar grasp, may mask an extensor response to lateral foot stimulation when stimulation in the midfoot brings about flexion of the toes.

Other important reflexes for the neuropsychiatrist are:

- Myerson sign, a failure to habituate to regular, one per second, taps on the glabella (with the tapping hand outside of the patient's visual field), present in parkinsonism and diffuse brain disease;
- Hoffmann sign, flexion of the thumb with snapping of the distal phalanx of the patient's middle finger, an upper extremity sign of pyramidal dysfunction though sometimes present bilaterally in normal subjects;
- Grasp, flexion of the fingers with stroking of the patient's palm toward the fingers during distraction, despite instructions to relax, associated with disease of the contralateral supplementary motor area;
- Avoidance, extension of the wrist and fingers to the same stimulus as the grasp, a less well-known sign that points to contralateral parietal cortex abnormality.

Several other "primitive reflexes" are less specific in that they are commonly present in the normal subject and are thus less useful for diagnostic purposes. These include the suck, snout, and palmomental reflexes.

FOCAL NEUROBEHAVIORAL SYNDROMES

The idea that the brain is regionally specialized had a difficult gestation in the 19th century, and the key to its acceptance lies with the recognition of the effects of focal brain lesions. At the end of the 18th and early in the 19th centuries, phrenology drew adherents to the claim that personality traits could be inferred by inspection of the cranium. This claim was faulty, but phrenology had an underlying theory that was an important step forward for the brain sciences; in particular, the beliefs that the brain was the organ of the mind and that the mind could be fractionated into functions gave impetus to the development of neuroscience in a modern form. In the middle of that century, the gradual realization that aphasia occurred with damage to specific areas of the left hemisphere was another crucial step. The subsequent identification of numerous syndromes of localized damage—syndromes such as apraxia, agnosia, visuospatial impairment in its various forms, and so on—is a fascinating story of astute and painstaking clinicopathologic, and later clinicoradiologic, correlation.

Patients' introspective access to their deficits may be limited. Much of cognitive processing is unconscious, not in the sense of being excluded from awareness by motivated defense, but in the sense that it is not even in principle open to introspection. Jonathan Miller, in his television show *The Body in Question*, displayed this point by asking passersby, in a man-in-the-street interview, "Sir, where is your spleen?" No one can say from introspection where the spleen is. The same is true of much of cognitive processing. Explanations provided by patients may be confabulations that fill in such introspective gaps in a situation where the brain is functioning abnormally in a way not foreseen in its design. In neuropsychiatry, subjective experience and behavior are separate explicanda. The patient's appraisal of his or her own situation is always relevant to collaboration with treatment and its outcome and should be explored in every clinical encounter.

Dementia

The term dementia, like the term organic, has been officially replaced in the case of dementia by Major Neurocognitive Disorder. Psychologists and psychiatrists are more sensitive to this change than other clinicians, such as neurologists, so for the purposes of interdisciplinary communication, as with the term organic, this chapter uses the older term. Although commonly considered a syndrome of "global" cognitive impairment, implying global or diffuse brain dysfunction, in fact, each dementing disorder produces a distinct pattern of brain pathology and a corresponding pattern of cognitive dysfunction. For this reason, and against tradition, dementia will be discussed under the rubric of focal neurobehavioral syndromes. (Here, the science, not a committee, elicits the change.)

In Alzheimer disease, the earliest neuropathologic abnormality is a characteristically medial temporal accumulation of plaques and tangles, initially in the entorhinal cortex and subiculum (the input and output zones of the hippocampus). The disease progressively involves association cortices in temporoparietal and prefrontal regions. This burden of pathology determines the characteristic early memory impairment with ensuing anomia, failure of grasp, and coarsening of personality. On occasion, Alzheimer disease pathology is predominantly posterior, with a concomitant predominance of visuospatial impairment in the clinical course (the syndrome of posterior cortical atrophy), or predominantly in frontal networks, with executive dysfunction and characteristic behavioral alterations (the syndrome of frontal or executive variant Alzheimer disease). By contrast, in frontotemporal dementia, the earliest disease manifestations are pathologically in the frontal or temporal cortex, clinically presenting as primary progressive aphasia, semantic dementia, or a frontal apathy or disinhibition syndrome (bvFTD). In none of these situations is a view of dementia as a "global" impairment of brain function justified by the clinical or pathologic facts; rather, selective disruption of anatomic networks corresponds to symptomatic features.

Extensive subcortical white and gray matter damage due to small vessel disease is a common cause of dementia, and in this situation, the clinical picture is dominated by slowed mental processing, forgetfulness with relative preservation of recognition memory (as opposed to free recall), and executive cognitive dysfunction. "Strategically" located single infarctions can also produce dementia. These strokes can involve the left angular gyrus, genu of the internal capsule, and (perhaps most commonly) medial thalamus. The thalamic and internal capsule strokes may produce cognitive impairment by interfering with frontal networks.

Delirium

Classically a syndrome of "global" brain dysfunction due to toxic-metabolic or infectious encephalopathy, delirium may also point to focal brain disease. Delirium may be caused by infarction in the right posterior superior temporal gyrus due to occlusion of the inferior division of the right middle cerebral artery, or by infarction in the inferior temporo-occipital cortex, on the left or bilaterally due to posterior cerebral artery occlusion. In both instances, focal neurologic signs may be limited to a visual field cut or absent entirely. Finding bilateral asterixis or multifocal myoclonus strongly indicates a toxic-metabolic brain derangement, and the history and physical examination should provide pointers for the essential laboratory confirmation of the abnormality. Features of the mental state are of little use in determining the cause of the syndrome, except that agitation is far more common in certain disorders, such as substance withdrawal, hypoxia, and the syndromes of left posterior cerebral artery stroke or right-middle cerebral artery stroke with involvement of the temporal lobe. Agitation with prominent insomnia and sympathetic hyperactivity—together called *agrypnia excitata*—may raise the question of Morvan syndrome, associated with antibodies to CASPR-2 or, less commonly, LGI1.

Aphasia

Acquired impairment of lexical or syntactic performance is termed aphasia. Lexicon and syntax do not exhaust the domain of language and attention is devoted below to prosody and discourse pragmatics. At the bedside, the clinician should be able to distinguish language impairment from other sources of abnormal discourse (such as psychosis), delineate the features of abnormality in the patient's linguistic function, and tentatively identify the locus of brain injury.

A simple distinction between "expressive" and "receptive" defects has some power to distinguish between anterior and posterior lesion sites, but it is not in current use in aphasiology because most aphasiogenic lesions produce some impairment in both production and comprehension of language, and these impairments are of multiple sorts.

A widely accepted approach to examination and classification of aphasia identifies six domains for elucidation: spontaneous speech, naming, comprehension, repetition, reading, and writing. Attention to spontaneous speech reveals dysfluency and word-finding difficulties. The dysfluent speaker produces shorter phrases and utterances without a natural flow. Substantives (nouns and verbs) may be preserved at the expense of function words (such as prepositions) and grammatical morphemes (such as tense endings), leading to agrammatical utterances that are nonetheless relatively information-rich. Lesions disrupting fluency are characteristically anterior in the left hemisphere or involve the putamen. Naming performance requires the adequate functioning of a network, including posterior temporal, temporoparietal, and inferior frontal sites. This is ordinarily tested by confrontation ("What do you call this?"), which can be conveniently done using body parts or common items at the bedside. Naming from description ("What do you call the vehicle that travels underwater?") is an alternative mode of testing, particularly useful for visually impaired or agnosic patients. Comprehension is tested best using probes with minimal demand on output, so "yes/no" questions ("Does a stone sink in water?") are better than motor commands, which may be impaired by concurrent apraxia. Impairment of comprehension results from posterior temporal lesions. Disordered repetition is disclosed by asking the patient to produce progressively longer utterances reiterating the examiner: "Airplane, he and she are here, the spy fled to Greece…." Repetition may be surprisingly spared (the so-called transcortical aphasias) or disproportionately impaired (conduction aphasia). The latter depends on lesions of the insula or external capsule. Reading comprehension (not reading aloud, a different skill) tests comprehension with a different input modality from aural comprehension and some patients will show significant dissociations. Similarly, writing tests output in a different modality from speech. Writing is a particularly sensitive probe for the anomia seen in early Alzheimer disease and the disorganization seen in delirium.

A classification of the aphasias using the data from an examination, as just outlined, is shown in Table 2.1–2. Clinicians recognize, however, that many patients will not fit well into the categories created by this scheme.

Ideomotor apraxia commonly occurs together with aphasia. This is a disorder of performance of skilled movements to command in the absence of explanatory elementary sensory and motor disturbances. Oral apraxia is revealed by the patient's incapacity to show, for example, how to blow out a match or lick an envelope. Limb apraxia is shown by the patient's incapacity to show, for example, how to wave goodbye or use a hammer or screwdriver. Patients with these deficits may nonetheless be able to follow whole body commands: "Show me how a boxer stands," for example. Patients rarely complain of apraxic deficits, in part because they are artifacts of the examination, in the sense that they may not be present in the utilization of real-world items.

Attention

Several related phenomena cluster under the clinical description of attentional disorders. At the most fundamental level, alertness represents a continuum ranging from a coma to normal wakefulness. The clinician faced with a patient who is less than fully alert should quantify the disorder by assessing the patient's response to a graded series of probes: does the patient orient to the examiner's presence in the room, what does the patient do when his or her name is called or when touched or when shaken or when a (harmless) painful stimulus is applied, and so on. These responses should be recorded in detail rather than summarized by an ambiguous term such as "lethargic."

Alert patients may show deficits in sustained attention to external stimuli (vigilance) or internal stimuli (concentration). Attentional deficits of these sorts are characteristic of delirium. Vigilance can be assessed with a bedside adaptation of a continuous performance task, for example, by asking the patient to lift a hand each time the examiner says the letter "A" or (to raise the difficulty) each time the examiner says the letter "A" after the letter "D." The examiner then produces a series of random letters at a deliberate and steady rate over an extended period of time. Any error of omission or commission represents a failure. By asking the patient to count from 20 to 1 or give the days of the week or the months of the year in reverse, the examiner can appraise concentration. Digit span—asking the patient to repeat a list of numbers spoken at a slow, steady rate without separation into chunks—is a classic test of attention; the lower limit of normal digit span is five.

A "higher" level of attentional function is the capacity to manipulate information kept in consciousness over a short period of time: working memory. An excellent example is alphanumeric sequencing. The patient is asked to alternate between numbers and letters; the examiner provides "1-A-2-B-3-C" as a model, then allows 30 seconds for the patient to start at 1 and give as many alternations as possible. If only the number of correct alternations is counted (ignoring errors), the lower limit of normal is 15. A comparable, simple task is alphabetizing the letters of the word "world" (or any similar word). Working memory is known to require intact processing in the dorsolateral prefrontal cortex.

Hemineglect is a focal disorder of attention, almost always of left hemispace in a patient with acute right hemisphere disease. Most characteristically, the lesion is parietal, but distinguishable patterns of hemineglect occur with frontal and cingulate lesions. Gross neglect can be recognized in the patient's ignoring, or even denying the ownership of, the left limbs or failing to attend to objects and people in the left hemispace. Subtler degrees of neglect can be elicited by presenting an array in which the patient must search into both hemifields to point to items, for example, all the yellow dots in a stimulus card with dots of varied colors to both left and right.

In the phenomenon of hypermetamorphosis, part of the Klüver–Bucy syndrome of bilateral anterior temporal damage, animals or patients exhibit an increased level of attention to individual items in the environment.

Amnesia

The term *memory* is used in several ways by clinicians and psychologists. The amnestic syndrome features impairment of learning of new material (anterograde amnesia) and a variable period of impaired recall prior to the onset of the syndrome (retrograde amnesia). It is due to damage to the hippocampus or to the anterior thalamus (including the mammillothalamic tract). Memory proper is distinguished from retention in the consciousness of material over the course of a few seconds, which may be called "working," "iconic," or "short-term" memory. This function is spared in the amnestic syndrome because it depends on frontoparietal mechanisms distinct from the hippocampal and thalamic pathways damaged in amnesia. Deficits due to hippocampal and thalamic lesions are dependent on the lateralization of the damage, left-sided damage producing predominantly verbal and right-sided damage figural memory impairments.

A distinction between free recall and recognition memory is of clinical significance and generally can be made adequately if imperfectly at the bedside. Hippocampal and thalamic patients show accelerated forgetting so that cues (such as providing the semantic category) are relatively ineffective in aiding recall. This is the characteristic pattern of Alzheimer disease. Recognition memory is always better than free recall on an absolute scale; exaggeration of the disparity (i.e., sparing of recognition memory) is characteristic of memory impairment due to dysfunction of frontal mechanisms of effortful search.

In addition, frontal patients show impairments of memory for the temporal context or source of information. This deficit is probably relevant to the occurrence of confabulation. Spontaneous confabulation occurs in only a minority of amnestic patients and depends on ventromedial frontal damage.

Memory is so commonly impaired in brain disorders that it should be tested in all patients undergoing neuropsychiatric evaluation. Recall of a test phrase (e.g., a name and address) or a word list over a several-minute distraction is a valid and simple screening test. However, a more detailed analysis of memory is necessary for patients with disorders likely to affect memory mechanisms, including (among many others) head injury, epilepsy, and dementing disorders. Testing should include both verbal and nonverbal material. For example, testing recall over several minutes' delay of three words and three shapes, or three words and three-pointed directions, or three object names and the locations in the room in which those objects were hidden is easily performed. Further, the examiner should be prepared with cues, including appropriate (incorrect) foils, to assess sparing of the capacity to make use of cues in frontal memory impairment. For example, if one of the words provided is "piano," the examiner could cue, "One was a musical instrument" and further provide "guitar, piano, violin" as multiple-choice options. Only rough inferences can be drawn from this bedside assessment as compared with neuropsychological evaluation.

Table 2.1–2.
Aphasia Syndromes

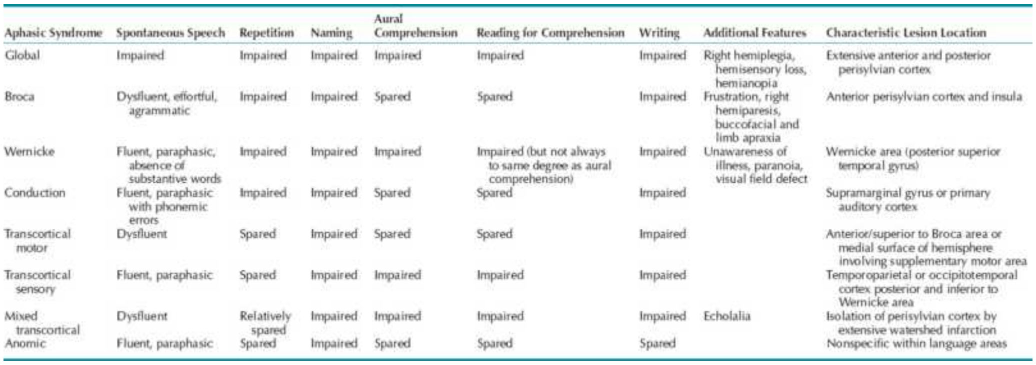

Aphasic Syndrome	Spontaneous Speech	Repetition	Naming	Aural Comprehension	Reading for Comprehension	Writing	Additional Features	Characteristic Lesion Location
Global	Impaired	Impaired	Impaired	Impaired	Impaired	Impaired	Right hemiplegia, hemisensory loss, hemianopia	Extensive anterior and posterior perisylvian cortex
Broca	Dysfluent, effortful, agrammatic	Impaired	Impaired	Spared	Spared	Impaired	Frustration, right hemiparesis, buccofacial and limb apraxia	Anterior perisylvian cortex and insula
Wernicke	Fluent, paraphasic, absence of substantive words	Impaired	Impaired	Impaired	Impaired (but not always to same degree as aural comprehension)	Impaired	Unawareness of illness, paranoia, visual field defect	Wernicke area (posterior superior temporal gyrus)
Conduction	Fluent, paraphasic with phonemic errors	Impaired	Impaired	Spared	Spared	Impaired		Supramarginal gyrus or primary auditory cortex
Transcortical motor	Dysfluent	Spared	Impaired	Spared	Spared	Impaired		Anterior/superior to Broca area or medial surface of hemisphere involving supplementary motor area
Transcortical sensory	Fluent, paraphasic	Spared	Impaired	Impaired	Impaired	Impaired		Temporoparietal or occipitotemporal cortex posterior and inferior to Wernicke area
Mixed transcortical	Dysfluent	Relatively spared	Impaired	Impaired	Impaired	Impaired	Echolalia	Isolation of perisylvian cortex by extensive watershed infarction
Anomic	Fluent, paraphasic	Spared	Impaired	Spared	Spared	Spared		Nonspecific within language areas

Apart from patients with persisting amnestic syndromes, the neuropsychiatrist may be presented with patients who suffered an amnestic state transiently or rarely may see one during a transient amnestic state. The syndromes of transient global amnesia and transient epileptic amnesia, and their differentiation, have been fully described and require thorough history taking, neuropsychological evaluation, and electroencephalogram (EEG) recordings. The neuropsychiatrist should also know that amnesia for criminal offenses is common; certainly, it is not confined to those who committed a crime while in a delirious, ictal, or postictal state, as is sometimes claimed for exculpatory reasons.

Visuospatial Dysfunction

The requirement to test visual as well as verbal memory has just been mentioned. Drawing and copying tasks can further the assessment of visuospatial function. Copying intersecting pentagons (as in the Mini-Mental State Examination) or the shapes used in the memory task begins the assessment. With more complex figures, failures with a slavish element-by-element strategy are characteristic of patients with right hemisphere damage, as is the neglect of the left side of the stimulus. Recognition of drawn objects presented in atypical ways, such as the Poppelreuter overlapping figures, tests visuospatial analysis without praxis.

The variety of disorders of higher visual function has already been mentioned in describing the complex structure of visual association cortices. Prosopagnosia is a defect in the recognition of faces. Such a defect may be obvious from the history or may be a more subtle abnormality; it can be spotted at the bedside, albeit insensitively, with the use of a few pictures of famous people. Defects of topographic skill, although rarely presenting in an isolated form, also occur with right hemisphere dysfunction. The patient can be asked to describe a route between familiar places or a geographical question thought to be within premorbid capacities ("If you're going from New York to Los Angeles, is the Atlantic Ocean in front of you, behind you, to your left, to your right?").

The incapacity to grasp in attention multiple visual objects at once is known as simultanagnosia. The patient may fail in describing a complex visual scene by reporting only a single, perhaps peripheral, element. Together with psychic paralysis of gaze (inability to direct gaze voluntarily, or ocular apraxia) and optic ataxia (misreaching under visual guidance), it makes up Balint syndrome, the archetypal disorder of the dorsal visual pathway. The patient with impairment of reaching under visual guidance should be examined reaching without visual guidance (e.g., pointing to parts of their own body with eyes closed) to confirm the defect.

Executive Cognitive Dysfunction

This term refers to the initiation of cognition and action, their maintenance in the face of distraction, organized but flexible pursuit of goals, and self-monitoring with error correction. Executive processes are crucial for adaptive function, and performance in this realm is better correlated with real-world outcomes of brain-injured patients than are many other domains traditionally analyzed in neuropsychology or many psychosocial variables. Exploration of executive function is of central importance in the neuropsychiatric examination. Analysis of behavioral disturbance and neuropsychological deficits in patients with cerebral injury suggests that multiple dissociable processes compose executive cognitive function, and certainly, these processes are instantiated by anatomically distributed systems. However, many different tasks recruit a similar or identical set of regions in the mid-dorsolateral prefrontal cortex, the mid-ventrolateral prefrontal cortex, and the anterior cingulate. Nonetheless, the clinical examiner must know that no single probe can screen for all dysfunctions.

Many aspects of executive function are illuminated by attention to the patient's performance of elements of the history taking and examination. Disinhibition may be noted in abnormalities of comportment during social interaction. Motor impersistence, the failure to sustain actions that can be initiated properly, may be noted in the patient's peeking when asked to keep the eyes closed, repeatedly looking back at the examiner's face when lateral gaze (especially to the left) is attempted, or failing to keep the arms extended or the tongue protruded when instructed to do so. Perseveration is the continuation of elements of past actions into present activity. Perseverative responses may be noted when testing naming or attention. Echopraxia, for example, the patient’s crossing the arms when the examiner (spontaneously) does so, even when some other behavior has been requested, can be observed during the interview and examination. Similarly, echolalia refers to the automatic repetition of an interlocutor’s speech. Blurting, also called echoing approval or yes/no reversal, is the automatic response to a question, followed immediately by the patient’s correction of the automatic response (“yes, yes”) to the intended or correct response (“no”). Utilization behavior is an automatic tendency to make use of objects in the environment, for example, picking up a pen and starting to write, despite this behavior’s being inappropriate to the setting.

More focused efforts to assess executive function are almost always indicated in the neuropsychiatric examination. Perseveration may be specifically sought in the motor sequencing tasks described above or with a sample of spontaneous writing. A tapping task with conflicting instructions may illuminate the inflexibility of goal-directed behavior that gives rise to perseverative responding. The patient is instructed to tap once if the examiner taps twice and twice if the examiner taps once. The examiner then taps on the table in a random series of one tap or two taps. This can be directly followed by a go/no-go tapping task, in which the patient is instructed to tap once if the examiner taps once, not at all if the examiner taps twice. Intrusions from the previous task’s instructions represent perseverative responding; echopraxic responses (tapping just like the examiner) represent failures of inhibition. Inhibition of reflexive gaze can be tested during the examination of eye movements, as described above. Looking at the moving stimulus rather than in the opposite direction as instructed amounts to a visual grasp response and represents a failure of inhibition. Spontaneous word-list generation (“Tell me all the animals you can think of,” or “Tell me all the words that start with ‘s’”) depends on the capacity for an effortful search of semantic stores. A greater decrement in fluency to semantic cues (“animals”) than to phonemic cues (“words with ‘s’”) is seen in Alzheimer disease because of the degradation of semantic stores due to temporoparietal damage. Working memory can be assessed with the alphanumeric sequencing task described above.

Anatomic inferences from dissociations in performance on these tasks are limited. Go/no-go tasks depend on the integrity of the orbitofrontal cortex and other tasks on the dorsolateral prefrontal cortex and its circuit. Impersistence is associated with right hemisphere dysfunction. A further dissociation between executive cognitive impairments and personality change can be seen in frontal injury. Especially with orbitofrontal lesions, executive function can be spared even in the face of grave alterations in emotions and comportment. Nonetheless, it bears repeating that neuropsychiatric examiners always should consider executive cognitive function in their formulation of cases.

Dysnarration

The patient’s narrative also gives clues to brain dysfunction. Clinicians are familiar with the information-poor accounts provided by patients with memory impairment, perhaps accompanied by glances at the companion seeking assistance. Patients with dorsolateral prefrontal lesions may be laconic, aspontaneous, and apathetic in ways that are evident in their accounts of symptoms. Patients with ventromedial prefrontal lesions may be impulsive, disorganized, and insightless in ways that are observable in the interview. Pragmatic disorders due to frontal and right hemisphere damage may impair narrative coherence through verbosity, vagueness, and disregard for the listener's informational needs. Thus, for example, pronoun use may be syntactically correct, but the referents of pronouns are obscure to the listener. Although language itself is normal, the way language is embedded in social interaction is not.

Disordered Mood and Emotion

Several syndromes of disordered emotion in organic disease can be delineated. Disturbances of recognition and expression of emotion with right hemisphere lesions have already been mentioned. Patients and their families are rarely aware of these deficits and do not complain about them; rather, the examiner must recognize the deficit and figure it into a formulation of the patient's social and functional decline. Impairment of prosodic expression should not be mistaken for depressed affect. Testing of affective prosody can be undertaken at the bedside without special equipment. The examiner can ask the patient to say emotionally loaded sentences in an emotional manner, expressing surprise, fear, pleasure, and anger. People vary considerably in their native acting talents, and the range of normal performance is wide. The examiner also can utter neutral sentences in various emotional tones: "I am going to the store," stated with surprise, and so on, with the examiner's face turned away from the patient to avoid providing a second input channel. Patients should be able to recognize the affect. Separately, if testing materials are available, the examiner can assess the patient's capacity to identify emotions in visual scenes and facial expressions. Lesions that involve both limbic and heteromodal cortices in the right hemisphere especially impair performance in recognizing emotional facial expressions.

Pathologic laughing and crying are stereotyped displays of affect incongruent with inner experience and elicited by inappropriate, nonemotional, or inadequate stimuli. The examiner may, in the extreme, be able to elicit full displays of affect by waving a hand in front of the patient's face. The patient is often embarrassed by the pathologic expression of affect. Features of pseudobulbar palsy are often present in these patients. In these cases, damage to a volitional control system has disinhibited a network subserving emotional display. This network includes the amygdala, hypothalamus, periaqueductal grey, and dorsal tegmentum. A broader form of affective dysregulation, which may be called emotionalism or emotional lability, is more commonly seen, usually in the direction of tearfulness. Patients report that they are more emotional than previously and that the tears are sudden, unexpected, and uncontrollable. However, they are generally congruent with the patient's subjective state. Such patients are often cognitively impaired. The rare phenomenon of *fou rire prodromique* (mad prodromal laughter) presages acute vascular lesions of the brainstem or thalamus.

Apathy is an emotional disturbance marked by a reduction of affect and motivation. Goal-directed behavior is reduced, and emotional responses are lacking. The distinction from depression is crucial: patients do not report negative emotional states or ideational content. Although they may meet the criteria for depression because of the loss of interest in activities, they are mentally empty rather than full of distress. Recognizing apathy rather than mistaking it for depression may imply treatment with different pharmacologic agents, for example, the use of dopamine agonists. Euphoria refers to a persistent and unrealistic sense of well-being without the increased mental or motor rate of mania. Although often mentioned in connection with multiple sclerosis (MS), it is unusual and almost always associated with extensive disease and substantial cognitive impairment.

The Klüver–Bucy syndrome, as described in the captive monkey, includes a reduction in aggression (tameness), excessive and indiscriminate sexual behavior, hypermetamorphosis (forced attention to environmental stimuli), and hyperorality (mouthing nonfood items). This mixture of emotional, perceptual, and motivational changes is dependent on bilateral damage to the amygdala. In human patients, pathologies including trauma, herpes simplex encephalitis, and frontotemporal dementia can produce the syndrome, usually in partial form.

Depression is common in patients with brain diseases including stroke, MS, traumatic brain injury, epilepsy, and Parkinson disease. Certainly, this is in part a reaction to altered circumstances and distressing disability. Nonetheless, the syndromal nature of the depressed state and its imperfect correlation with measures of disability have prompted extensive efforts to seek anatomic correlations. Converging evidence leads to a model of alterations in a distributed network involving neocortical and limbic elements. In particular, a dorsal compartment involving the dorsolateral prefrontal cortex, inferior parietal cortex, and the dorsal and posterior portions of the cingulate gyrus show underactivity in the depressed state; these regions are thought to mediate the cognitive alterations and impairments of depression. Inversely, a ventral compartment containing the anterior insula, subgenual cingulate, hippocampus, and hypothalamus are overactive; these elements are thought to mediate somatic (“vegetative”) features of the depressed state. Interactions between the two compartments are mediated through the thalamus, basal ganglia, and especially rostral cingulate.

Mania is substantially less common than depression after brain injury. Mania is associated with right-sided lesions involving paralimbic cortices in orbitofrontal or basotemporal regions or subcortical sites in the caudate or thalamus. Some evidence suggests that subcortical lesions are more likely to produce a bipolar picture and cortical lesions unipolar mania. As with depression, the abnormal mood state does not necessarily appear in immediate temporal association with the injury, so determining whether the mood disorder is organic or idiopathic is not always straightforward. The absence of a personal history of mood disorder is an obvious criterion, but the presence of a family history of a mood disorder may mark a vulnerability factor not operative in the absence of the brain lesion. Age of onset is relevant, especially for mania: the onset of idiopathic mania after age 40 is rare.

A common issue in neuropsychiatric assessment is the patient with late-onset depression, in whom evaluation reveals executive cognitive dysfunction and subcortical white-matter disease. This state of vascular depression is marked by the presence of vascular risk factors, notably hypertension, a tendency to psychomotor retardation and anhedonia and not psychosis or guilty ideation, and poor outcome with usual treatments. Some but not all of these patients have apathy rather than depression.

Abnormalities in Agency

Ordinarily, the person performing an action has the sense of being the one performing it. The prototype abnormality of this normal subjective sense is the “alien hand” phenomenon. Patients with parietal lesions may report a sense of foreignness of the hand and the limb may exhibit levitation or avoidance reactions. More dramatically, with medial frontal or callosal lesions, the hand may engage in unwilled behavior (representing unilateral utilization behavior), or intermanual conflict may occur.

Abnormal Social Behavior

The multitude of behaviors exhibited in social interaction has multiple underpinnings. Several behavioral complexes, the neurobiology of which has come under scrutiny, can be observed in their abnormal form in patients and, at times, understood from an anatomic and physiologic point of view.

Failures of empathic understanding are common in patients with frontal injury. These impairments, in part, result both from cognitive inflexibility in assessing complex social situations, especially in patients with dorsolateral prefrontal lesions, and from emotional impoverishment, especially in patients with orbitofrontal lesions. The intensity of social interaction manifested by patients with temporal lobe epilepsy may be due to deficits in social cognition or to a limbic lesion that reinforces social cohesiveness.

The capacity of human beings to understand the mental states of others—and thus to recognize not just another’s goals or intentions but also the other’s deceptions or pretense—has been termed mentalization or “theory of mind.” Deficits in theory of mind are thought to be fundamental in autism spectrum disorders. Imaging and lesion data suggest that this capacity depends critically on the prefrontal cortex (particularly the right medial prefrontal cortex adjacent to the anterior cingulate gyrus), right temporoparietal cortices, and amygdala. Patients with isolated lesions of the amygdala are rare, but deficits in theory of mind are seen in patients with frontal disease and may contribute to their social failure. Patients with right hemisphere lesions have a range of deficits in social interaction that may be characterized as a disorder of pragmatics. Although they may grasp the propositional content of language correctly, they mistake aspects of communication that require an appraisal of the interlocutor’s intent, for example, whether an utterance was intended as a joke.

Abnormal Beliefs and Experiences

Hallucinations are a common feature of diseases of the brain. Visual hallucinations in the absence of auditory hallucinations are suggestive of organic disease. Visual hallucinations may occur in a hemifield blind from cerebral disease, so-called release hallucinations. Visual hallucinations in the setting of visual impairment due to ocular disease, usually in the older adults, are known as the Charles Bonnet syndrome. The hallucinations are characteristically vivid images of living figures, and the patient is aware of their unreality. Other psychopathology is absent, and treatment aimed at the hallucinations is usually ineffective. Elaborate formed visual hallucinations may occur with lesions of the thalamus or upper brainstem, so-called peduncular hallucinosis. The symptoms are worse in the evening (crepuscular), and again the patient is aware of the unreality of the visual experiences. Prominent, early visual hallucinations in the context of progressive dementia suggest dementia with Lewy bodies (DLB).

Auditory hallucinations rarely occur with pontine lesions. More common are musical hallucinations in the setting of hearing impairment, akin to the Bonnet syndrome. Unilateral hallucinations are characteristically experienced as ipsilateral to the deaf ear. Olfactory hallucinations occur as a limbic aura in focal epilepsy, but they also occur in idiopathic psychiatric illness.

Palinopsia and palinacousis refer to persisting or recurrent perceptual experiences after the object is gone, in the visual and auditory domains, respectively. Lesions in the association cortex—parieto-occipital and temporal, respectively—are responsible, although (for the visual sphere more than the auditory) drug toxicity is often the explanation.

The content of delusions may yield clues to causative organic disease and its nature. Most notably, misidentification delusions have been associated with dysfunction of face processing and clearly linked—in some but not all cases—to right hemisphere dysfunction. Misidentification of place is regularly associated with visuospatial and executive cognitive dysfunction. Misidentification delusions have been of special interest in cognitive neuropsychiatry. Perceptual recognition without a sense of familiarity (as in the Capgras syndrome and perhaps the nihilistic delusions of the Cotard syndrome) may reflect a disruption of visual-limbic connections. In a sense, it is the reverse of déjà vu, which amounts to familiarity without perceptual recognition. However, many patients with misidentification delusions have no evidence of organic disease. Although such patients may have dysfunction of underlying mechanisms similar to patients with ascertainable organic disease, the similarity of clinical phenomena cannot be taken to prove an identity of mechanism.

Particular delusional themes may mark delirious thinking, such as a focus on danger to or harm to others, as opposed to the more self-centered constructions in idiopathic psychotic disorders. However, most delusions in patients with brain disease are of a more banal nature, often with persecutory elements that bespeak cognitive failure (the reported theft of one's purse, e.g., deriving from a failure of memory as to its location). Complexity or elaborateness of delusional ideation is associated with the preservation of intellect, and delusions tend to become less complex with the progression of dementia.

Paraclinical Investigations

Specialized paraclinical investigations form a major part of the neuropsychiatrist's toolbox. Sometimes patients are referred for neuropsychiatric consultation when a routine investigation—such as a screening MRI or EEG—gives an unexpected abnormal result; the neuropsychiatrist is called on to assess the meaning of the finding in the psychiatric context. Atypical age of onset, atypical features or course (such as particularly rapid progression of symptoms), absence of relevant psychiatric family history or presence of a relevant general medical history, presence of cognitive impairment, and presence of abnormal physical signs should prompt increased scrutiny for organic disease in a patient presenting psychiatrically. Catatonia should also lead to exploration for organic disease: data from several centers suggest that one catatonic patient in five has an explanatory organic disease.

Neuroimaging

Structural neuroimaging with computed tomography (CT) and later with MRI revolutionized practice in the clinical neurosciences. No longer was the organ of interest invisible within the carapace of the skull. CT relies on the differential absorption of x-rays by brain tissues and on the power of computerized methods to integrate data from multiple perspectives. The strengths of CT are its speed and its sensitivity to blood and bone. Thus for neuropsychiatric purposes, situations in which a patient cannot tolerate a prolonged imaging procedure may mandate CT. This problem often arises with an agitated demented or psychotic patient. Bony abnormalities, parenchymal deposition of calcium, and intracranial hemorrhage are particularly well assessed by CT. Such questions arise in the acute aftermath of trauma in particular.

The advent of MRI was an advance over CT in several respects. The anatomic resolution is substantially better, and the discrimination of white matter abnormalities exceptionally so. The capacity to display data from a single acquisition in multiple views—sagittal, axial, and coronal—allows improved anatomic understanding. Gradient-echo and susceptibility-weighted images sensitively reveal the sequelae of hemorrhage and may be useful in assessing the damage from trauma. Infusion of gadolinium for contrast enhancement is not necessary for delineation of nonvascular anatomic structures, such as is the goal in the case of atrophy or old stroke or trauma but can identify areas of breakdown of the blood–brain barrier, such as in the meninges in meningitis or in parenchymal lesions of active multiple sclerosis, tumor, or acute stroke. Special imaging sequences should be used for the identification of cortical dysplasia or mesial temporal sclerosis. Volumetric MRI allows diagnosis by quantitative assessment of delineated brain structures, such as the hippocampus, in the case of temporal lobe epilepsy and potentially Alzheimer disease. MR angiography allows the delineation of medium and large vessels without the administration of contrast material, as is required for conventional angiography. Stenosis of these vessels, such as the vessels of the neck, or the presence of vascular malformations or aneurysms is reliably ascertained. However, the resolution is not sufficient to allow the assessment of small vessels; thus, some forms of vasculitis cannot be excluded with MR angiography and require contrast angiography.

Additional MRI sequences include diffusion-weighted imaging, which captures acute vascular injury; diffusion-tensor imaging, which discloses patterns of connectivity in white matter; and magnetization-transfer imaging, which promises even greater sensitivity to brain lesions than FLAIR imaging. Diffusion-weighted imaging is essential in the evaluation of stroke, and high intensities in the cortex ("ribboning") are characteristic of CJD. Magnetic resonance spectroscopy (MRS) is a method for analyzing the regional chemical composition of the brain. The benefits of its ability to identify neuronal loss and glial proliferation are still under investigation, although in certain circumstances—such as distinguishing radiation necrosis from a recurrent brain tumor—it is of proven utility. Dopamine-transporter imaging (DAT-scan) sensitively and specifically shows loss of nigrostriatal neurons and aids in the differential diagnosis of DLB and drug-induced parkinsonism. However, loss of nigral cells occurs in frontotemporal lobar degeneration as well as in DLB, and DAT-scanning can be abnormal in frontotemporal dementia.

Functional Neuroimaging

Four methods of functional neuroimaging are available: single-photon emission computed tomography (SPECT), positron emission tomography (PET), functional MRI, and brain mapping by quantitative electroencephalography. All are exciting research avenues, but the established clinical role for functional imaging is limited. All the techniques have a place in the presurgical evaluation of epileptic patients. SPECT and PET in the patient with frontotemporal dementia or Alzheimer disease typically disclose the focal nature of the dysfunction. The demonstration of occipital hypoperfusion strongly supports a diagnosis of DLB. The evidence for other clinical uses of functional imaging is at present limited or anecdotal.

Electroencephalography

The expectation of the originators of EEG was that it would allow the tracking of mental processes. This hope has not been realized. EEG does have the advantage over other clinically available brain imaging tools in that it reflects function at high temporal resolution, a resolution corresponding to the time course of mental processing. Thus, at least from a research perspective, measurement of brain potentials in relation to stimuli—the technique of evoked potentials—has the capacity to identify anomalous modes of cerebral processing. Recordings from electrode placements in subdural or cortical sites provide irreplaceable information about the origin and spread of epileptic discharges, but this invasive technique is justified only under special circumstances. From today's practical point of view, scalp EEG can distinguish epileptic from nonepileptic spells and clarify the type of epilepsy. It can also distinguish acute encephalopathy from nonorganic psychosis.

Only 30% to 50% of patients with epilepsy show an epileptic abnormality on a single interictal waking EEG. With sleep deprivation, sleep during the recording, and repeated recordings, sensitivity improves to 70% to 80%. Anterior temporal electrodes add to the sensitivity and localizing power of the EEG, but nasopharyngeal electrodes, which are quite uncomfortable for the patient, do not provide additional sensitivity and are not recommended. A reasonable protocol would start with a routine EEG including anterior temporal leads; if this is negative, but suspicion remains high, a second EEG with sleep deprivation can be undertaken. A third and fourth EEG may be useful, but the rate of discovery of abnormalities declines after that. Even then, some epileptic patients will not have been shown to have interictal abnormalities. At times ambulatory EEG is of use to ascertain the epileptic nature of undiagnosed events, but the restricted montage of the ambulatory equipment limits its utility. Hospitalization for video EEG recording may be essential for clarifying the nature of puzzling spells.

Delirium is characterized by slowing of the EEG, a finding never seen in acute idiopathic psychosis. This differential point can be decisive in a confusing clinical setting. However, EEG is not indicated as routine in the screening of psychotic patients, except that the EEG is abnormal in almost all (>95%) patients with anti-NMDAR encephalopathy. Among the dementing disorders, frontotemporal dementia is distinctive in having a normal EEG even as the clinical state becomes moderately severe. In CJD, the EEG is always slow and may ultimately (not necessarily immediately) show the diagnostic feature of pseudoperiodic complexes.

Evoked potentials can identify abnormalities in neural transmission along myelinated pathways, such as the visual pathway or the sensory pathways of the spinal cord and brainstem. This can help in the diagnosis of disorders such as multiple sclerosis or B_{12} deficiency.

Polysomnography

Identifying sleep-related disorders by multimodal overnight recordings is helpful in identifying obstructive sleep apnea, which may produce cognitive or affective symptoms; REM sleep behavior disorder (RBD), which when idiopathic is a strong indicator of an alpha-synucleinopathy such as DLB; or narcolepsy, which produces a panoply of neuropsychiatric symptoms. Other sleep or circadian rhythm disorders may be of psychiatric relevance, for example, in explaining fatigue.

Laboratory Investigations

In general, empirical evidence for the utility of laboratory studies supports only a limited role for “routine” or screening investigations; for the most part, laboratory tests should be performed as guided by the history and examination. A full discussion of laboratory strategies for all neuropsychiatric situations is beyond the scope of this section. In regard to dementia, a complete blood count (CBC), chemistry panel, B_{12} assay, perhaps methylmalonic acid, and thyrotropin (TSH) assay are indicated as screening tests, in addition to a test for syphilis (the fluorescent treponemal antibody test [FTA] or other antitreponemal antibody tests) in those areas of the United States in which the prevalence of syphilis justifies the testing. (The region of high incidence is a broad belt across the South in addition to some urban areas in the North; 30 U.S. counties contribute more than half the national total of cases.) The reason the FTA is the test of choice is that reagin tests (the venereal disease research laboratory test [VDRL] or rapid plasma reagin [RPR]) may revert to normal after intercurrent antibiotic treatment or with the passage of time and thus are insufficiently sensitive to serve as screening tests for neurosyphilis.

Appropriate screening tests for mental presentations other than dementia, for example, first-episode psychosis (FEP), are less well established. Recent studies confirm one’s clinical impression that organic cases are rare, probably fewer than 5%, of patients with FEP. The first step should be a neuropsychiatric history and examination. No laboratory approach substitutes for a psychiatrist’s expert sense that the patient with FEP is atypical of idiopathic mental illness. The factors mentioned above (atypical age of onset, atypical features, the rapid evolution of symptoms, absence of relevant psychiatric family history or presence of a relevant general medical history, presence of cognitive impairment, and presence of abnormal physical signs) should prompt consideration of organic disease. However, these clinical “red flags” have not been shown to have adequate sensitivity and specificity to rule out autoimmune encephalopathy in particular. A reasonable laboratory screen might include CBC, chemistry panel, TSH, urinalysis, and urine toxicology. (A fasting lipid panel and, in women of reproductive potential, a pregnancy test are appropriate as well.) If it is considered justified to screen for rheumatic disease, an antinuclear antibody test (ANA) is adequate for this purpose, being abnormal in almost all cases of lupus, although not sufficient to confirm that diagnosis. (False positives from psychotropic drug-induced positive ANA assays will be an important confound.)

Of particular recent interest is the question of evaluating patients, especially patients with FEP, for autoimmune encephalopathy. Such an evaluation is mandated by the presence of catatonic features. EEG is a reasonable step given its very high sensitivity for anti-NMDAR antibody-mediated encephalopathy. CSF examination is necessary to make the diagnosis, so a serum antibody panel can be considered inadequate. This may create a difficult situation if the patient declines lumbar puncture. The literature suggests that most patients will agree to a spinal tap, and no doubt the clinician’s manner and explanation will play a role in how the situation evolves. Nonetheless, in some situations, the data may be inadequate for a definite diagnosis, and how to proceed will be a matter of clinical judgment.

The clinician must remember that the utility of a test depends not just on its sensitivity and specificity but also on the prevalence of the disease tested for in the population tested (pretest probability). In a population with a very low frequency of disease, a positive test, even a highly specific one, is more likely to be a false positive than a true positive. Though the adverse consequences of missing an autoimmune encephalopathy are unquestioned, given the demonstrated infrequency of autoimmune encephalopathy in psychiatric populations, the clinician must realize that routine testing by means of an antibody panel in patients without organic features (such as cognitive impairment or motor abnormalities) can result in a cascade of positive but uninformative tests, frightening to the patient, misleading to the patient’s family, and disruptive of good psychiatric care. In the patient without such organic features, the clinician can remember the advice given in what is now an old-fashioned textbook: “…[T]he second examination is the most helpful diagnostic test in a difficult neurological case.”

Excessive laboratory testing is to be deplored; on the other hand, limiting laboratory testing to generally familiar diseases is inexpert. Many rare metabolic diseases can present in adolescence or young adulthood with purely psychiatric manifestations, with psychiatric manifestations on the background of developmental intellectual impairment, or with prominent psychiatric symptoms mixed with motor or systemic abnormalities. The family history, which may be of severe intellectual deficit or early mortality in a sibling, may be a pointer to the need for a more thorough investigation. Some of these disorders are listed in Table 2.1–3.

Genetic testing is relevant in certain neuropsychiatric situations, some of which are mentioned in Table 2.1–3. “Premutation”-range repeat expansions in the fragile X gene may at times provide an explanation for an anxiety disorder. Identifying a mutation in the prion protein gene or finding a mutation (such as in presenilin-1) or repeat expansion (such as in Huntington disease or a disorder due to C9orf72) can clinch an otherwise uncertain diagnosis of progressive cognitive impairment and put paid to otherwise plausible diagnostic avenues. Because of the implications of a positive test result, often genetic counseling is appropriate before testing.

A 28-year-old man was referred for cognitive impairment. He said, “I don’t have a big thought process right now.” At age 16, he was in a car crash, but his mother reported that he had not lost consciousness. A month later, he began to show myoclonic jerks at the breakfast table, and 3 months after the crash, he had a generalized seizure. After appropriate evaluation, including normal brain imaging, he was diagnosed with juvenile myoclonic epilepsy, which responded well to treatment. His mother had noticed a reduction in mental acuity about 5 years before evaluation. Two years before evaluation, he was unable to carry on working as a cashier, and by a year later was unable to operate a microwave. He displayed repetitive behaviors, such as switching a light on and off, which had been considered to be due to obsessive-compulsive disorder, but the behaviors were not driven by anxiety. By the time of evaluation, he required cues for personal activities of daily living.

His perinatal history was unremarkable, but his speech was delayed. He had been an average student and an average athlete. A college counselor raised the question of an autism spectrum disorder because of social awkwardness.

On examination, he had short 5th digits of both hands but no other dysmorphic features. Except for echolalia and a pronounced visual grasp, the neurologic examination was unrevealing. He was nearly unable to complete a coherent utterance of more than a few words. Detailed cognitive examination, insofar as it was feasible, disclosed a severe disorder of working memory, with other deficits seemingly consequent to that.

An extensive evaluation, undertaken before referral and after, was unrevealing. A panel of genetic tests disclosed a mutation in the serpin gene, and in retrospect, neuroserpinosis could have been suspected because of the history of juvenile myoclonic epilepsy. His mother proved to show mosaicism of the serpin gene.

Table 2.1–3.
Metabolic Disorders with Potential Psychiatric Presentations

Disease	Psychiatric Features	Other Features (All May Have Cognitive Impairment)	Diagnostic Test
Metachromatic leukodystrophy	Psychosis	Peripheral neuropathy, cerebellar and pyramidal signs, anterior white matter abnormality on MRI	Arylsulfatase A, genetic testing
GM2 gangliosidosis	Psychosis, mood disorder	Pyramidal, cerebellar, lower motor neuron signs	Hexosaminidase A, genetic testing
Niemann–Pick type C	Mood disorder, psychosis	Hepatomegaly, splenomegaly, cerebellar signs, dystonia, vertical oculomotor apraxia	Fibroblast culture, genetic testing
Adrenoleukodystrophy	Mood disorder, psychosis	Pyramidal signs, peripheral neuropathy, visual loss, adrenal insufficiency, posterior white matter abnormality on MRI	Very long-chain fatty acids, genetic testing
Neuronal ceroid lipofuscinosis (Kufs disease in adolescents)	Mood disorder, psychosis	Extrapyramidal, cerebellar signs	Enzyme assay, genetic testing
Cerebrotendinous xanthomatosis	Psychosis	Cerebellar, pyramidal signs, peripheral	Cholestanol, genetic testing

		neuropathy, cataracts, xanthomas	
Urea cycle disorders, organic acidemias	Acute confusion/psychosis	Protein intolerance, headache, nausea, liver disease, pyramidal signs, seizures	Ammonia, urinary organic acids, plasma amino acids
Porphyria (acute intermittent or variegate)	Acute confusion/psychosis	Peripheral neuropathy, abdominal pain	Urine porphobilinogen, genetic testing
Wilson disease	Personality change, rarely psychosis	Extrapyramidal signs, Kayser–Fleischer ring, liver disease, basal ganglia hyperintensity on MRI	Ceruloplasmin, 24-hour urinary copper, liver biopsy, genetic testing

Examination of the Cerebrospinal Fluid

Examination of CSF obtained through lumbar puncture is sometimes a crucial element of the diagnostic process, in particular to diagnose infection or inflammation, more rarely in neuropsychiatric practice to seek evidence of neoplasia (such as meningeal carcinomatosis). Specific assays are available for the diagnosis of neuropsychiatrically relevant infectious agents. In rheumatic diseases involving the brain, the white cell count may not be elevated, but elevated protein and evidence of intrathecal elaboration of antibodies may give evidence of inflammatory activity. The latter is sought by the ratio of immunoglobulin-G (IgG) to albumin or, better, by the IgG index which requires measurement of serum IgG, and by the presence of oligoclonal bands specific to the CSF. As mentioned, for the autoimmune encephalopathies, the relevant antibodies may not be present in serum and CSF examination is required.

Measurement of the neuron-derived 14–3–3 protein in CSF was used to assist in the diagnosis of CJD, but the ratio of phosphorylated to total tau protein in spinal fluid is more sensitive and highly specific. RT-QUIC directly identifies prions in the CSF and gives even greater sensitivity and specificity for the diagnosis of CJD (though it is less sensitive for other presentations of prion disease). Increased CSF total and phosphorylated tau and reduced amyloid protein, in combination, show high sensitivity and high specificity for Alzheimer disease.

Removal of CSF by lumbar puncture or external drainage is part of the evaluation of patients suspected of shunt-reversible normal-pressure hydrocephalus. Patients undergoing this procedure should have CSF assayed for amyloid and tau.

Neuropsychological Assessment

Neuropsychological evaluation has an important role to play in neuropsychiatric care, both for diagnosis and for management. Sound use of the clinical neuropsychologist as a consultant requires a first step formulation of a cogent consultative question. The more specific the consultant's question, the more able the neuropsychologist is to integrate the psychometric data with the rest of the clinical picture. Much of the early literature on neuropsychological assessment focused on identifying and localizing organic brain disease. With the advent of neuroimaging, neuropsychological testing is seldom the most powerful means of addressing this issue, although it certainly continues to play such a role, for example in lateralizing cognitive deficits as a preoperative tool in epileptic patients. Nor is the role of the neuropsychologist to make a disease diagnosis, although at times the psychometric picture is strongly suggestive of a particular diagnosis.

If substantial confounds make bedside evaluation difficult, neuropsychological data may be of considerable assistance. For example, identifying supervening cognitive impairment in an intellectually disabled or poorly educated patient or subtle impairment in a highly intelligent patient may be impossible for the clinician to do with confidence, while neuropsychological assessment may provide these diagnoses. Another example of utilizing neuropsychological assessment as a probe of brain function is disclosing a pattern of cognitive strengths and weaknesses amounting to right hemisphere learning disabilities in a patient with a clinical picture suggestive of autism spectrum disorder or a cluster A personality disorder. Obtaining neuropsychological data often allows for more precise targeting of behavioral interventions, more specific education of families, and a more confident assessment of decline or of benefit from pharmacologic treatments.

One common use of neuropsychological assessment requires a word of caution: identifying cognitive impairment in an older patient presenting with mood disorder or psychosis. No neuropsychological findings should deter the clinician from aggressive treatment of the psychiatric symptoms, and nonspecific state-dependent attentional and motivational factors may confound the neuropsychological results. Rather than devoting resources of time and energy to pinpointing a moving target in the acute phase, deferring the assessment until symptoms are reduced is often the wiser course. Similar caution applies to the situation of an acutely psychotic young person.

Another caution about neuropsychological assessment falls under the rubric of ecologic validity. This term refers to the extrapolation of results obtained in the neuropsychological laboratory by artificial paper-and-pencil methods to real-world performance. The concern arises in particular with orbitofrontal lesions, which may produce a paucity of cognitive findings but devastating personality change. Deriving clinical measures suitable to characterize such patients from the developing realm of affective neuroscience is a current challenge to neuropsychology.

Brain Biopsy

Biopsy of the brain has a limited role in neuropsychiatric evaluation. The morbidity and mortality of the procedure, as performed by an experienced neurosurgeon, are low, but the sensitivity of the procedure is lower than one might expect. For example, the sensitivity of biopsy for primary angiitis of the central nervous system (CNS) may be only 75%. In some circumstances, a biopsy of a peripheral tissue can substitute for a brain biopsy in a patient with primarily cerebral symptoms at lower risk. For example, lung or muscle biopsy may make a diagnosis of sarcoid, skin biopsy a diagnosis of vasculitis if a rash is present or of CADASIL, and temporal artery biopsy of giant cell arteritis. In neuropsychiatric situations, the major indication for biopsy of the brain is consideration of inflammatory disease, when the nature and aggressiveness of treatment depend on a tissue-proven diagnosis.

Although it cannot be considered a clinical diagnostic test, the neuropsychiatrist should not neglect the autopsy as a learning tool.

COMMON NEUROPSYCHIATRIC CONDITIONS

This section provides a survey of some issues commonly brought to the attention of neuropsychiatrists. The emphasis is on the priorities for clinical and laboratory assessment for a variety of presentations. The organization is by disease and syndrome as a complement to the anatomic and symptom-oriented discussion provided so far. This perspective is distinctive for neuropsychiatry within psychiatry; the disease processes underlying symptoms in the idiopathic disorders are unknown. For neuropsychiatric patients, one can hope and work to uncover the disease-causing the symptoms and, on fortunate occasions, to provide disease-specific treatment.

Dementia

Many diseases produce the clinical state of dementia. A shotgun laboratory approach to “ruling out treatable disease” is unwise, not only because finding reversibility is unusual but because, to a considerable degree, differential diagnosis among neurodegenerative diseases is possible and of clinical importance. Moreover, clinical clues to reversible disease are available in the history and examination: use of psychotoxic medicines, rapid course, mildness of cognitive impairment (even short of fully meeting criteria for dementia), subcortical features of the cognitive disorder, and presence of motor signs. Table 2.1–4 provides specific guidance to be gained from clinical clues. Progressive dementia supervening on developmental intellectual disability raises a particular differential diagnosis, including many of the diseases listed in Table 2.1–3.

The clinician needs to gather data relevant to management issues other than reversibility, such as safety of living arrangements, driving safety, and preparation of a will and advance directives. Dementing patients often develop psychiatric symptoms, which respond to pharmacologic and behavioral treatment. All these considerations should prompt the clinician to cast a wide net in data gathering regarding the demented patient.

A 67-year-old woman presented with at least 1 year of progressive memory impairment, confusion, then irritability and suspiciousness. The mental state was typical of Alzheimer disease, and the physical examination disclosed only brisk tendon jerks. An EEG, done earlier because of a spell of uncertain nature, had shown left temporal spikes. The neuropsychological assessment had shown a pattern typical for Alzheimer disease, with memory impairment characterized by rapid forgetting, semantic more than phonemic verbal fluency deficits, and anomia. MRI, however, demonstrated extensive white-matter disease with bilateral confluent hyperintensities, which extended into the gyri and involved U-fibers. CSF examination was entirely normal. Skin biopsy for CADASIL and screening genetic assay for CADASIL were negative. Repeat EEG showed bilateral temporal spikes and carbamazepine (Tegretol) was begun. The clinical diagnosis was leukoencephalopathy due to cerebral amyloid angiopathy, possibly with Alzheimer disease. The patient's mother had died at age 86, having suffered from "the same thing" as the patient. Four of the mother's five siblings became demented in the eighth or ninth decade of life, none earlier, in most cases with a clinical diagnosis of Alzheimer disease. The patient herself was an only child. The patient's two daughters, both young adults, were very concerned that they might inherit the same disease as their mother, and they insisted the patient undergo the brain biopsy that a geriatrician had recommended. This disclosed pronounced amyloid angiopathy. Immunostaining for A-β confirmed showed neuropil plaques; immunostaining for tau did not reveal neuritic plaques. Nonetheless, Alzheimer disease could not be excluded. No inflammation was seen. Unfortunately, several days after the biopsy, she developed status epilepticus.

Among the dementias, bvFTD is of particular neuropsychiatric interest. In contrast to Alzheimer's disease (AD), the incidence of which increases with age, in FTD, the peak incidence is in the fifth and sixth decades of life. In contrast to AD, which characteristically has a cognitive presentation with memory failure most prominent, bvFTD presents with altered personality and behavior; typically, executive cognitive dysfunction can be detected. Other neuropsychiatric features may be present early in the course, including psychotic symptoms, especially in bvFTD due to repeat expansions in the C9orf72 gene.

Epilepsy

Major concerns in patients with epilepsy include differential diagnosis, psychosis, personality change, depression, violence and other episodic behaviors, and PNES. The last will be dealt with below, along with other conversion disorders.

Patients with paroxysmal disorders can be misdiagnosed as having epilepsy when they do not have it or as having a different disorder when epilepsy is the correct diagnosis. Paroxysmal symptoms from panics, cardiovascular disease with syncope or near-syncope, endocrine disorders (pheochromocytoma, carcinoid, systemic mastocytosis), or conversion disorder can be mistakenly labeled epileptic; contrariwise, epilepsy can be missed when a diagnosis of panic disorder, in particular, is accepted.

A 59-year-old man was evaluated for 7 years of memory problems and spells refractory to treatment on a diagnosis of panic disorder. These spells characteristically lasted 5 to 10 seconds and recurred as often as hourly; he was sometimes amnestic for the spells afterward. During an attack, he had gooseflesh, and his speech became garbled; once at church, he was thought to be speaking in tongues. Extensive treatment trials with benzodiazepines and serotonergic drugs had given no consistent benefit. Apart from hyperlipidemia, he had no significant medical history. EEG had been negative on three occasions, MRI on two occasions, and Holter monitoring and SPECT on one occasion each. The neurologic and mental state examinations were normal. An episode was witnessed during the examination: He showed 10 seconds of facial flushing and stereotyped hand movements. The attacks were subsequently abolished by a trial of an antiseizure medicine.

Table 2.1–4.
Clues to the Differential Diagnosis of Dementia in the Neurologic Examination

Abnormal eye findings	Celiac disease Gaucher disease type 3 Mitochondrial cytopathy Multiple sclerosis Niemann–Pick disease type C Progressive supranuclear palsy	Syphilis Vascular dementia Wernicke–Korsakoff syndrome Whipple disease Wilson disease
Ataxia	Celiac disease Cerebellar degenerations Chronic traumatic encephalopathy GM2 gangliosidosis Hypothyroidism Multiple sclerosis Niemann–Pick disease type C Prion disease	Progressive multifocal Leukoencephalopathy Toxic-metabolic encephalopathy Wernicke–Korsakoff syndrome Wilson disease
Dysarthria	Cerebellar degenerations Chronic traumatic encephalopathy Dialysis dementia Motor neuron disease Multiple sclerosis Niemann–Pick disease type C Neuroacanthocytosis	Pantothenate kinase-associated Neurodegeneration Progressive multifocal Leukoencephalopathy Progressive supranuclear palsy Wilson disease
Extrapyramidal signs	Alzheimer disease Cerebellar degenerations Chronic traumatic encephalopathy Dementia with Lewy bodies Fahr syndrome GM1 gangliosidosis type 3 Huntington disease Multiple system atrophy Neuroacanthocytosis Niemann–Pick disease type C	Parkinson disease Progressive supranuclear palsy Postencephalitic parkinsonism Subacute sclerosing panencephalitis Toxic-metabolic encephalopathy Vascular dementia Wilson disease

	Normal-pressure hydrocephalus Pantothenate kinase-associated Neurodegeneration	
Gait disorder	Adrenomyeloneuropathy Cerebellar degenerations Chronic traumatic encephalopathy HIV encephalopathy Multiple sclerosis Normal-pressure hydrocephalus	Parkinson disease Progressive supranuclear palsy Syphilis Vascular dementia Wernicke–Korsakoff syndrome
Myoclonus	Alzheimer disease Celiac disease Dialysis dementia Kufs disease Lafora body disease	Mitochondrial cytopathy Prion disease Subacute sclerosing panencephalitis
Peripheral neuropathy	Adrenomyeloneuropathy B_{12} deficiency HIV encephalopathy Metachromatic leukodystrophy	Porphyria Toxic-metabolic encephalopathy
Pyramidal signs	Adrenomyeloneuropathy B_{12} deficiency Cerebellar degenerations GM2 gangliosidosis HIV encephalopathy Kufs disease Metachromatic leukodystrophy Motor neuron disease Multiple sclerosis	Pantothenate kinase-associated Neurodegeneration Polyglucosan body disease Progressive multifocal Leukoencephalopathy Syphilis Vascular dementia

HIV, human immunodeficiency virus.

Modified with permission from Sandson TA, Price BH. Diagnostic testing and dementia. *Neurol Clin.* 1996;14(1) 45–59. Copyright © 1996 W. B. Saunders Company. doi:10.1016/S0733-8619(05)70242-5.

A 52-year-old woman was referred for the evaluation of spells. In her 30s, she had been hospitalized for depression and was subsequently treated intermittently as an outpatient. The family history included several members with depression or bipolar disorder. Two years before evaluation, she presented with a headache and proved to have an unruptured aneurysm, which was clipped through a craniotomy. Several months later, she had a generalized convulsion. She went on to have spells at a rate of up to six a day. They were stereotyped and abrupt in onset and termination; she could not identify provocative factors or social contexts. During a spell, she would feel cold and have gooseflesh for about 3 minutes. Then she would become rigid and unable to speak or interact, although she was able to hear others' speech. This would last several minutes. Then she would begin to cry. The whole sequence would last some 6 to 10 minutes. She was on phenytoin (Dilantin) with a therapeutic serum level. Previous trials of divalproex (Depakote) and topiramate (Topamax) were not tolerated. EEGs had shown only right frontal slowing with no epileptic features on several tracings. The neurologic and cognitive examinations were normal, and she was not depressed at the time of evaluation. The clinical picture was inconclusive. In favor of epilepsy were the abrupt onset and termination, stereotyped nature of the spells, and background of craniotomy; against epilepsy were the weeping, length of the ictus (if all the phenomena were taken to be ictal), failure of response to treatment, relatively inactive EEG, and background of depression. Video EEG done after medication withdrawal recorded three focal seizures with impaired awareness with right anterior-inferior temporal onset with her typical semiology.

In exploring the psychiatric concomitants of epilepsy, the clinician needs to be aware of the nature of epilepsy. Most adult epilepsy is focal (localization-related epilepsy), with the ictal onset in the temporal lobe. However, other forms of epilepsy, including frontal epilepsy and primary generalized epilepsy, are common. These distinctions, and the laterality of the focus, can often be inferred from the semiology of the seizures as reported or as observed clinically. A history of febrile convulsions in childhood and the age of onset of epilepsy are relevant to the likelihood of mesial temporal sclerosis as the underlying pathology. Body asymmetry and dissociated facial paresis should be sought as indicators of laterality. The MRI and EEG provide crucial information on pathology and seizure type. Many of the findings relating psychiatric disorders to epilepsy are concerned with focal epilepsy of temporal onset; linking psychiatric symptoms to epileptic syndromes other than temporal lobe or limbic epilepsy may go beyond the evidence. Further, to what extent psychopathology is associated with seizures themselves and to what extent with the brain disease that gives rise to them remains controversial. Without question, cognitive impairments are related to the lateralization of the temporal focus.

Psychotic states in epileptic patients are usually divided into those occurring during the epileptic ictus, often called epileptic twilight states; those occurring for a delimited period in the aftermath of a seizure or, more commonly, a flurry of seizures, called postictal psychosis; and those that are chronic, called interictal psychosis. Usually, this chronology can be ascertained by inquiry, but at times EEG monitoring is necessary to identify the occurrence of seizures in relation to psychopathological phenomena, especially because patients can be amnestic for focal seizures with impaired awareness. A further issue to be elucidated from the history is the relationship between seizure treatment and control and the level of psychopathology, especially psychosis. An inverse relationship is sometimes noted. Better seizure control is associated with the occurrence of psychosis, a phenomenon known as forced normalization. On the other hand, frequent seizures certainly can cause an increase in confusion and related failure in functional capacity.

A 35-year-old woman with lupus and intractable epilepsy was admitted several times with persecutory and nihilistic delusions ("I'm dead") and depressive symptoms. Investigations to identify active cerebral lupus were unrevealing, even when she had evidence of peripheral activity of the disease. In fact, MRI disclosed the findings of hippocampal sclerosis, suggesting that the epilepsy was idiopathic and not due to cerebral lupus. Without specific treatment, the psychotic symptoms diminished over the course of several days; this also was thought to make a diagnosis of cerebral lupus unlikely and a postictal psychosis likely. Between episodes, she showed no psychotic phenomena.

The interictal personality syndrome of temporal lobe epilepsy (the Gastaut–Geschwind syndrome) is characterized by hypergraphia, religiosity or deepened metaphysical interest, aggressiveness, and hyposexuality. Also described are altered social behavior depending on the intensity of interaction, an inability to end interactions, and circumstantiality of discourse (phenomena confusingly denominated "viscosity"). The syndrome remains controversial; what is of importance for neuropsychiatric assessment is that inquiry be directed to phenomena such as hypergraphia that are not included in the usual review of symptoms of idiopathic psychiatric disorder.

Episodic aggression is often suspected of being ictal but rarely is. Aggression occurring during seizures is almost always disorganized, not deliberately directed. A high threshold is justified in attributing a violent act to epilepsy in the absence of typical epileptic features. Amnesia for serious violence is common and not a strong pointer to an epileptic origin.

A special issue in the neuropsychiatric assessment of the epileptic patient is the presurgical evaluation. Surgical treatment, especially of temporal lobe epilepsy due to mesial temporal sclerosis, is underutilized; ideally, more and more patients with medically refractory epilepsy will be evaluated for their suitability for surgery. Along with an intensive electroencephalographic evaluation, volumetric MRI, functional neuroimaging, and neuropsychological assessment, the patient's psychiatric state should be systematically evaluated. The patient's ability to consent and issues such as the patient's capacity to cope with the stress of monitoring and surgery, as well as the expectations held for surgery, should be addressed. Neither depression nor psychosis is an absolute contraindication to surgery, although a chronic psychosis probably will not be alleviated by surgery. Indeed few if any psychiatric findings will contraindicate surgery, but psychiatric evaluation may well reveal deficits that need to be taken into account in developing a treatment plan.

Traumatic Brain Injury

Traumatic brain injury is epidemic in our society, with advances in emergency medical care leading to growth in the prevalence of survivors of severe injury. Issues commonly facing the neuropsychiatrist include aggression, depression and anxiety, and the delineation of deficits (sometimes for legal purposes) in patients with a mild traumatic injury. The features of the head injury should be ascertained, ideally with confirmation from medical records.

The altered behavior and personality common after traumatic brain injury are more burdensome for families than are the physical disabilities. Disinhibition and aggression are particularly uncomfortable and often hard to treat. A complicating factor is that preinjury impulsivity and substance abuse are common, as they predispose to head injury.

A 24-year-old woman was seen 19 months after a car crash in which she was an unrestrained passenger. She had been comatose for 3 months and underwent surgical evacuation of a left-sided intracranial hematoma. She had a few weeks of rehabilitation after regaining consciousness and returned home after spending most of a year in a nursing home. The family was at their wits' end over episodes of aggression, which appeared to be directed angry behavior elicited by frustration. She did not have depressive symptoms. On examination, she showed severe bilateral spasticity, including spastic dysarthria, drooling, and a brisk jaw jerk. She was able to recall dates and other details of her illness accurately, but she disclaimed behavioral or emotional alterations. Her language comprehension was adequate, but the output was telegraphic. Affect was labile. Behavior during the consultation was initially appropriate, with an obvious effort to cooperate with the evaluation, although she had greeted the examiner with, “I love you.” At the end of the examination, however, she urgently requested the examiner's business card and rammed him with her wheelchair while cutting off his access to the door. Chronic phase CT showed atrophy and left temporal encephalomalacia.

A 59-year-old man was referred by a court for an assessment of his ability to take part in proceedings related to his divorce. Three months prior to the evaluation, he was struck several times by an unknown assailant during an altercation regarding who was to get the use of a taxi. He suffered contusions of the left periorbital area but no other overt injury. He was able to recount the events in some detail, but he explained that this was because over time, by comparing notes with others, he had “put it all back together”; of his own recollection, he could remember the first punch that struck him but not the second or subsequent events of the altercation. Although he could not be certain of the duration of the gap in his recollection, it was clearly a matter of some seconds, possibly a minute or two, and at no time was he unconscious. In the aftermath, he was “confused” and had a headache. He found that he could not come up with names, dates, or numbers, although this information would generally come back to him later or with considerable unaccustomed effort. He also noticed that he “could not visualize” geographical scenes, so in planning to go to a familiar place, he was unable to picture it in his mind. Although he did not get lost, he found that when driving, he would turn the wrong way or miss a turn. He noted that his memory, previously highly trustworthy, could not be counted on: “I had to write everything down.” He found that he had to “take time to think” and “strain my brain to focus.” He distanced himself from business decisions and relied on trusted subordinates to counsel him. He acknowledged sensitivity to light, noting that he had begun to wear sunglasses even when the weather was cloudy and to turn off the room lights when he was watching television. To a lesser extent, he was bothered by noise. He noted that he was more readily irritated than was characteristic of him. He did not have depressive symptoms, intrusive recall of the altercation, or nightmares. The symptoms had gotten gradually less severe. His history included several head injuries in adolescence, with two of which he had a loss of consciousness for a few hours without recognized sequelae. The noncognitive mental state and neurologic examinations were normal. He scored 21 on the Mental Alternations Test, a normal performance on a task of mental speed and working memory. He scored 16 of 18 on the Frontal Assessment Battery, a collection of tasks assessing executive cognitive function. The two lost points were on the go/no-go task, on which he made perseverative errors. He was mildly disorganized in performing the ring/fist test of motor sequencing. MRI and EEG were normal. The picture was thought to be consistent with organic sequelae of traumatic brain injury. The case underlines the importance of prior traumatic brain injury in determining the effects of seemingly mild trauma and that loss of consciousness is not a prerequisite for significant sequelae.

Movement Disorders

Cognitive impairment due to the involvement of subcortical structures is a common neuropsychiatric feature of movement disorders. This applies to cerebellar as well as basal ganglia diseases for the anatomic reasons described above. The anatomy of the close relationship between emotion and movement was also described above. Clinically, mood disorders are common in IPD and other movement disorders. Anxiety disorders, although less emphasized in the literature than depression, are also common. The evaluator should take into account that mood and anxiety can fluctuate according to the timing of doses of dopaminergic drugs. A mood disorder can occasionally present in advance of overt movement abnormalities, so IPD must be considered in the differential of late-onset mood disorders.

A 43-year-old woman with no personal or family history of psychiatric illness developed a psychotic depression. She had a severe extrapyramidal reaction to risperidone (Risperdal). Two years later, when euthymic and unmedicated, she developed progressive shuffling gait, upper extremity tremor, and micrographia. She then suffered another episode of depression. Three years later, she had severe anxiety, no cognitive impairment, and the motor features of IPD.

Psychotic reactions to dopaminergic drugs are an important feature of movement disorders. Sometimes this is the result of overuse of prescribed dopaminergic agents in an effort to increase time in the “on” state.

A 63-year-old man with long-standing IPD developed delusions while being treated with high-dose levodopa and carbidopa (Sinemet) on a five-times-a-day schedule, pramipexole (Mirapex), tolcapone (Tasmar), and amantadine (Symmetrel). Under inpatient observation for several days on the prescribed doses, he remained psychotic. He responded well to quetiapine (Seroquel).

Developmental Disabilities

Adult patients with developmental disabilities are underserved by the medical and social service communities and are frequently referred for neuropsychiatric attention. Some of these patients will have had an inadequate diagnostic evaluation for the cause of the disability. Beyond clinical assessment, with particular attention to dysmorphology, because features of the mental state and neurologic examination are generally nonspecific, the most useful diagnostic tests are MRI and genetic testing. Specific genetic probes can confirm tentative clinical recognition of syndromes of mental retardation. Patients with developmental disabilities are vulnerable—indeed especially vulnerable—to the mood, anxiety, and psychotic disorders that can afflict anyone and can be treated effectively for these; diagnostic overshadowing (attributing all psychological and behavioral disturbance to developmental disability *tout court*) is to be avoided. These syndromes may present atypically in the developmentally disabled population, and the clinician must be alert to indirect indicators of mood disturbance or psychotic experience. For example, while the patient may not report a depressed mood verbally, the caregivers may report the loss of interest in favorite activities and the other features of a depressive syndrome.

Of particular neuropsychiatric interest is the question of behavioral phenotypes and specific psychological correlates of developmental syndromes. Syndromes recognized to have behavioral phenotypes (and their correlates) include:

- Lesch–Nyhan syndrome (self-injury);
- Prader–Willi syndrome (excessive eating);
- Williams syndrome (anomalous cognitive profile, elevated sociability);
- Velo-cardio-facial (DiGeorge or 22q11 deletion) syndrome (schizophrenia).

Infectious and Inflammatory Diseases

Infectious and inflammatory diseases of the brain always need to be considered in acute or subacutely evolving mental disorders. Among the infectious diseases, HSV encephalitis has a particular claim on attention because delay in diagnosis, even by hours, can lead to substantially increased morbidity and mortality. Definitive diagnosis is possible without biopsy by assaying for HSV in the CSF with the PCR, but treatment may be indicated if suspicion is high in advance of a firm diagnosis. Chronic meningitis, for example, from infection with fungi, is a rare consideration in subacutely evolving dementia; the definitive diagnostic tests are CSF assays or serologic tests (e.g., for toxoplasmosis). In the acquired immunodeficiency syndrome (AIDS) era, infection with opportunistic agents and with the human immunodeficiency virus itself needs to be kept in mind. In more recent times, the question of long-lasting brain dysfunction due to SARS-CoV-2 (CoViD-19) has assumed great importance. Present evidence is insufficient to judge the pathophysiology in many of these patients.

Noninfectious inflammatory diseases include rheumatic diseases, of which the prototype is systemic lupus. A rheumatic disease review of systems is always of importance in exploring the differential diagnosis of a puzzling case, especially in a young woman. Although psychosis is often considered the psychiatric hallmark of lupus, in fact, psychotic states (other than delirium) are unusual and a variety of other psychiatric pictures need to be included in the clinician’s consideration. Few clinical features of lupus are risk factors for cerebral disease, not even disease activity, which may be misleading in either a positive or a negative direction. One feature that is a risk factor for neuropsychiatric symptoms, including cognitive impairment, is the presence of antiphospholipid (aPL) antibodies; the primary aPL syndrome similarly carries mental risk.

A 40-year-old woman presented with the typical features of psychotic depression. There was a family history of depression, and she had suffered two episodes of depression earlier in her adult life, both of which were brief, nonpsychotic, and responsive to treatment. For the previous year, however, her depression had been poorly responsive to pharmacologic treatment and electroconvulsive therapy (ECT). Examination disclosed no definite cognitive abnormality and brisker reflexes on the left. MRI showed striking areas of white matter abnormality in the right hemisphere. Extensive laboratory investigation, short of angiography and biopsy, revealed only high-titer IgA anti–β2-glycoprotein-1 antibodies. Neuropsychological assessment performed after partial remission of the depression showed deficits in attention and mental processing speed. The working diagnostic formulation was that an otherwise mild idiopathic depressive disorder had been rendered treatment-resistant and severe by a wave of cerebral injury due to the aPL syndrome.

Other rheumatic diseases, such as Sjögren syndrome and the vasculitides, are also of neuropsychiatric importance. Autoimmune encephalopathy in the absence of systemic autoimmune disease often but not always presents as limbic encephalitis comprising memory impairment, anxiety, and seizures. Some but not all of these patients have neoplasms, the frequency depending on the particular autoantibody present. An expanding list of antigenic targets for the autoimmune process allows for specific diagnosis. No one believes that the list is complete; antibodies of yet-unknown specificity may in the future be found to explain neuropsychiatric illnesses.

A 56-year-old man without prior psychiatric history presented because of “a few months of memory problems.” Beginning 5 months prior to the neuropsychiatric evaluation, he seemed confused and forgetful to his family. For example, he repeatedly called his daughter at college and could not recall what she was studying. He declined over the ensuing months, with anxiety, depression, reduced energy, and difficulty in maintaining sleep but without anhedonia or diurnal variation of mood. A psychiatrist treated him with quetiapine and lithium on a diagnosis of an affective mixed state but without benefit. The neurologic examination was normal. On cognitive examination, he showed mild memory impairment but no disturbance of naming or executive function. MRI was unremarkable. The neuropsychological assessment showed impaired memory with both rapid forgetting and reduced encoding. A panel of autoimmune serologies showed antibodies to the voltage-gated potassium channel, and he responded to immunomodulatory treatment.

When and how to screen for autoimmune encephalopathy in psychiatric patients, for example, those with FEP has been controversial, as discussed above.

Conversion Disorder (Functional Neurologic Symptom Disorder)

Neuropsychiatrists often see patients whose symptoms appear to arise from brain disease but do not. These patients' condition has been described under various names: hysteria, functional disorder, psychogenic disorder, conversion disorder, or medically unexplained symptoms. None of the designations is entirely satisfactory. Whatever the designation, such patients are not uncommon. Complicating matters is the not uncommon coexistence of organic disease and conversion symptoms. For example, a sizable minority of patients with PNES have epilepsy as well. Brain disease may have produced, in some patients with conversion disorder, organic personality change with a reduction in the maturity of defenses and the too-easy resort to somatization.

Various techniques have been advocated for the identification of nonorganic diseases by means of physical examination. These have several shortcomings. First, they easily lend themselves to a counter-therapeutic alliance in which the examiner is trying to trick the patient—not a good start for the treatment, whatever the findings. Second, they fail to distinguish deliberate falsification on the patient's part (i.e., malingering) from conversion disorder. Third, most such findings are commonly present in patients with an organic disease who are trying to help the examiner make the diagnosis. That is, they may show a patient to be histrionic or suggestible but fail to rule out organic disease. Thus, for example, reporting a difference in vibratory sensation between the two sides of the sternum is by no means confined to patients with conversion disorder but may arise from the demand characteristics of the examination setting. Exceptions to this caution occur in cases where the nonphysiologic finding is precisely the phenomenon of the complaint.

Even then, however, the phenomena of brain disease are sometimes sufficiently odd that the examiner should maintain an attitude of humility about achieving diagnostic certainty by recognizing the nonphysiologic at a glance. Of the described "signs of hysteria," perhaps the best is the Hoover sign. The examiner places a hand underneath the heel of the affected leg of a supine patient who complains of leg weakness. Asked to press down with the heel, the patient fails to generate power with the leg. Asked to raise the opposing leg, however, the patient produces an automatic synergistic downward movement of the affected leg.

Recent systematic findings of progressively greater methodologic sophistication confirm the belief that experiences of neglect and abuse in childhood, and other childhood adversities are common in the background of patients with conversion disorder. This may indirectly account for another progressively more solidly substantiated finding, namely that the prognosis of conversion disorder is sometimes poor. Although a given symptom may wax and wane or disappear, patients commonly have a chronic course of disability, interpersonal difficulties, emotional symptoms, and fruitless seeking of medical help. Although hysterical symptoms have often been taken to represent a psychological conflict symbolically, the fundamental difficulty is that patients who make prominent use of somatization have a disorder of the symbolic function itself. The goal of the psychiatric examiner should be not to expose the patient but to establish an alliance that allows exploration of areas of the patient’s life outside the presenting symptoms and the construction of a plan to reduce distress and manage emotions (including focused treatment of commonly coexisting depressive disorder) and to develop alternative ways of seeking attention and assistance for distress.

THE NEUROPSYCHIATRIC PERSPECTIVE

This section has surveyed a neuropsychiatric approach to the patient. The neuropsychiatrist thinks anatomically about mental state disorders, even as cognitive neuroscientists attempt to construct a sufficiently sophisticated model of large-scale brain function to do justice to the complex mental states of neuropsychiatric interest. The neuropsychiatrist relies on rich data gathered at the bedside and on laboratory methods of investigating brain structure and function and diagnosing disease. The effort is to identify not just behavioral syndromes as found in DSM or ICD, but the pathologic processes underlying them in two senses. First, neuropsychiatry seeks medical diagnoses of systemic or brain diseases that account for the patient’s illness. Second, neuropsychiatry seeks to understand clinical phenomena in terms of the disruption of elementary mental processes, the nature of which is beginning to be elucidated by the cognitive neurosciences. The result is a highly differentiated diagnostic enterprise. With continual refreshment from a multidisciplinary base—ranging from cognitive neuroscience to general medicine—the neuropsychiatric approach to the patient is certain to remain exciting.

Further Readings

Argyropoulos GPD, van Dun K, Adamaszek M, et al. The cerebellar cognitive affective/Schmahmann syndrome: a task force paper. *Cerebellum*. 2020;19(1):102–125.

Brier MR, Bucelli RC, Day GS. Reader response: unintended consequences of Mayo paraneoplastic evaluations. *Neurology*. 2019;93(13):603.

Cotovio G, Talmasov D, Barahona-Corrêa JB, et al. Mapping mania symptoms based on focal brain damage. *J Clin Invest*. 2020;130(10):5209–5222.

D’Esposito M. Are individual differences in human brain organization measured with functional MRI meaningful? *Proc Natl Acad Sci U S A*. 2019;116(45):22432–22434.

Ducharme S, Dols A, Laforce R, et al. Recommendations to distinguish behavioural variant frontotemporal dementia from psychiatric disorders. *Brain*. 2020;143(6):1632–1650.

Endres D, Matysik M, Feige B, et al. Diagnosing organic causes of schizophrenia spectrum disorders: findings from a one-year cohort of the Freiburg Diagnostic Protocol in Psychosis (FDPP). *Diagnostics (Basel)*. 2020;10(9):691.

Frith C. In praise of cognitive neuropsychiatry. *Cogn Neuropsychiatry*. 2008;13:1–7.

Gainotti G. The role of the right hemisphere in emotional and behavioral disorders of patients with frontotemporal lobar degeneration: an updated review. *Front Aging Neurosci*. 2019;11:55.

Geschwind N. Disconnexion syndromes in animals and man. I. *Brain*. 1965;88:237–294.

Geschwind N. Disconnexion syndromes in animals and man. II. *Brain*. 1965;88:585–644.

Graus F, Dalmau J. Autoimmune encephalitis or autoimmune psychosis? *Eur Neuropsychopharmacol*. 2021;50:112–114.

Guasp M, Giné-Servén E, Maudes E, et al. Clinical, neuroimmunologic, and CSF investigations in first episode psychosis. *Neurology*. 2021;97(1):e61–e75.

Hassan A, Josephs KA. Alien hand syndrome. *Curr Neurol Neurosci Rep*. 2016;16(8):73.

Heimer L, Van Hoesen GW, Trimble M, Zahm DS. *Anatomy of Neuropsychiatry: The New Anatomy of the Basal Forebrain and its Implications for Neuropsychiatric Illness*. Academic Press/Elsevier; 2008.

Kelleher E, McNamara P, Dunne J, et al. Prevalence of N-methyl-D-aspartate receptor antibody (NMDAR-Ab) encephalitis in patients with first episode psychosis and treatment resistant schizophrenia on clozapine, a population based study. *Schizophr Res*. 2020;222:455–461.

Lamont PJ. Cognitive decline in a young adult with pre-existent developmental delay–what the adult neurologist needs to know. *Pract Neurol*. 2004;4:70–87.

Le Heron C, Apps MAJ, Husain M. The anatomy of apathy: a neurocognitive framework for amotivated behaviour. *Neuropsychologia*. 2018;118(Pt B):54–67.

Mendez MF. *The Mental Status Examination Handbook*. Elsevier; 2022.

Mesulam MM. Behavioral neuroanatomy: large-scale networks, association cortex, frontal syndromes, the limbic system, and hemispheric specializations. In: Mesulam M-M, ed. *Principles of Behavioral and Cognitive Neurology*. 2nd ed. Oxford University Press; 2000:1–120.

Mesulam MM. From sensation to cognition. *Brain*. 1998;121:1013–1052.

Ovsiew F. Hysteria in neurological practice: the somatoform and dissociative disorders. In: Jeste DV, Friedman JH, eds. *Psychiatry for Neurologists*. Humana Press; 2006:67–79.

Ovsiew F, Jobe T. Neuropsychiatry in the history of mental health services. In: Ovsiew F, ed. *Neuropsychiatry and Mental Health Services*. American Psychiatric Press, 1999:1–21.

Ovsiew F, Murray ED, Price BH. Neuropsychiatric approach to the psychiatric inpatient. In: Ovsiew F, Munich RL, eds. *Principles of Inpatient Psychiatry*. Lippincott, Williams & Wilkins; 2009:97–124.

Padmanabhan JL, Cooke D, Joutsa J, et al. A human depression circuit derived from focal brain lesions. *Biol Psychiatry*. 2019;86(10):749–758.

Pievani M, de Haan W, Wu T, Seeley WW, Frisoni GB. Functional network disruption in the degenerative dementias. *Lancet Neurology*. 2011;10(9):829–843.

Schmahmann JD, Pandya DN. *Fiber Pathways of the Brain*. Oxford University Press; 2006.

Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD. Neurodegenerative diseases target large-scale human brain networks. *Neuron*. 2009;62:42–52.

Trakadis YJ, Fulginiti V, Walterfang M. Inborn errors of metabolism associated with psychosis: literature review and case-control study using exome data from 5090 adult individuals. *J Inherit Metab Dis*. 2018;41(4):613–621.

Wortzel HS, Arciniegas DB. A forensic neuropsychiatric approach to traumatic brain injury, aggression, and suicide. *J Am Acad Psychiatry Law*. 2013;41:274–286.

2.2 Neuropsychiatric Aspects of Cerebrovascular Disorders

Stanley Lyndon, M.D., and Juan Carlos Urizar, M.D.

Definition

Cerebrovascular disorders consist of a large number of disorders that interrupt the blood supply to the brain. These disorders are divided into two groups based on etiology, ischemic (i.e., absence of blood flow) and hemorrhagic (i.e., bleeding in or around the coverings of the brain). They may also be classified according to the size of the affected blood vessel. Large artery disease produces large lesions of the brain, while small artery (arteriolar) disease results in small lacunar lesions. Vascular lesions may occur in any region of the brain.

Clinicians have recognized the association between neuropsychiatric disorders and cerebrovascular disease for over a century. However, this subject has only been the focus of systematic scientific scrutiny over the last 40 years. This chapter will focus on the neuropsychiatric correlates of stroke.

A stroke is defined as the sudden loss of blood supply to the brain leading to permanent tissue damage. Nearly 90% of strokes are due to ischemia, while the remaining 10% are caused by hemorrhagic events. While less common, hemorrhagic strokes tend to be more severe and are associated with greater mortality than ischemic strokes.

As the most common serious neurologic disorder in the world, stroke represents a global public health burden. The World Health Organization estimates that 15 million people are diagnosed with a stroke each year worldwide. The age-specific incidence of stroke varies dramatically over the lifespan. According to the 2021 Update of Heart Disease and Stroke Statistics published by the American Heart Association, the overall incidence of perinatal stroke in children in the United States is 29 per 100,000 live births. Stroke incidence increases considerably with age, and the incidence of first stroke is greatest in those 85 years or older.

However, approximately 10% of all strokes occur in individuals 18 to 50 years of age. In adults ages 55 to 64, the annual incidence of first-ever stroke is 2.2 to 5.3 per 1,000 persons, while the incidence is 8.9 to 13.8 per 1,000 persons in those 75 to 84 years old. Patients with stroke >85 years of age make up 17% of all patients with stroke. In the United States alone, 795,000 strokes and approximately 130,000 stroke-related deaths occur annually. Stroke is the fourth-leading cause of death in the United States and the second-leading cause of death globally. Over the period 2010 to 2050, the number of incident strokes is expected to be more than double, with the majority of the increase among older adults (≥75 years of age) and minority groups. Very older adult patients have a higher risk–adjusted mortality, have greater disability, have longer hospitalizations, receive less evidence-based care, and are less likely to be discharged to their original place of residence.

History

Early reports of depression after brain damage (usually caused by cerebrovascular disease) were made by neurologists and psychiatrists in case descriptions. Adolf Meyer warned that new discoveries of cerebral localization, such as language function in the early 1900s, led to an overly hasty identification of centers and functions of the brain. He identified that several clinical syndromes such as delirium, dementia, and aphasia were the direct result of brain injury. However, in keeping with his view of biopsychosocial causes for most mental “reactions,” he saw manic-depressive illness and paranoiac conditions as arising from a combination of brain injury (specifically citing left frontal lobe and cortical convexities), family history of psychiatric disorder, and premorbid personal psychiatric disorders. Eugen Bleuler described “melancholic moods lasting for months and sometimes longer appear frequently” in patients after strokes. Emil Kraepelin recognized an association between manic-depressive insanity and cerebrovascular disease. He stated, “the diagnosis of states of depression may offer difficulties, especially when arteriosclerosis is involved.” Kraepelin concluded that cerebrovascular disorder may be an accompanying phenomenon of manic-depressive disease or may itself produce depressive disorder.

Pseudobulbar affect (PBA) is another emotional disorder that has been associated with brain injuries, including cerebral infarction, and represents one of the differential diagnoses for depression in stroke patients. In 1956, Redvers Ironside described the clinical manifestations of this condition, noting that patients' emotional displays were characteristically unrelated to their reported inner emotional state. Laughing and/or crying may occur spontaneously or after some seemingly minor provocation. The emotion expressed is exaggerated relative to both the stimulus and the experienced emotional state and sometimes completely incongruent with the experienced emotional state. The expression of emotion here is involuntary and uncontrollable. This phenomenon has been given various names, including pathologic laughter and crying (PLAC), emotional incontinence, emotional lability, pathologic emotionalism, and involuntary emotional expression disorder. For the purpose of uniformity, this chapter will use the more widely used term in clinical settings, PBA.

Another emotional abnormality, also thought to be characteristic of brain injury, is the indifference reaction described by Derek Denny-Brown in 1952. Associated with right-hemispheric lesions, this reaction was described as indifference toward failures, lack of interest in family and friends, enjoyment of foolish jokes, and minimization of physical difficulties. In the late 19th century, Leonore Welt first described euphoria and loquaciousness associated with orbital frontal injury. Hermann Oppenheim used the term "witzelsucht" to refer to the inappropriate humor in these patients, while Karl Kleist stated that the orbital frontal cortex was the center of emotional life and the dorsal lateral frontal cortex was the source of psychomotor and intellectual activity.

Kurt Goldstein described another neuropsychiatric disorder historically associated with brain injury such as stroke, and he called it the catastrophic reaction. He characterized this reaction as an emotional outburst involving various degrees of anger, frustration, depression, tearfulness, refusal, shouting, swearing, and sometimes aggressive behavior. Goldstein ascribed this reaction to the inability of the organism to cope when faced with a serious defect in its physical or cognitive functions.

Joseph Jules François Félix Babinski introduced the term anosognosia to indicate a patient's lack of awareness of their hemiplegia. Babinski described this condition as "a mental disorder which I had the opportunity to observe in cerebral hemiplegia, consisting in the fact that the patients ignore or seem to ignore the paralysis which affects them." This term has been extended to unawareness of other deficits after strokes, such as cortical blindness, hemianopia, and amnesia.

Comparative Nosology. The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) defines poststroke psychotic, mood, and anxiety disorders as disorders due to another medical condition (stroke). Relevant specifiers are applied to the diagnosis to provide additional details about the patient's clinical presentation. For example, as per DSM-5, a poststroke depressive syndrome meeting full criteria for a major depressive episode would be called depressive disorder due to stroke, with major depressive-like episode. Poststroke subsyndromal depression, also known as minor depression, provides another instructive case example. Common following stroke, subsyndromal depressive symptoms are characterized by more than two but less than five of the required symptoms for a major depressive episode and persist for at least 2 weeks. According to DSM-5 nomenclature, such a clinical scenario is classified as a depressive disorder due to stroke with depressive features.

The diagnosis of poststroke anxiety based on DSM-5 criteria is termed "anxiety disorder due to stroke" regardless of the phenomenologic features of the anxiety disorder. For example, one patient may have anxiety symptoms suggestive of panic disorder and another patient may have symptoms more consistent with generalized anxiety disorder (GAD). Both of these patients would receive diagnoses of "anxiety disorder due to stroke" according to DSM-5.

The only DSM-5 disorders that are specific to cerebrovascular disease are the major or minor vascular neurocognitive disorders. DSM-5 identifies two forms of cognitive impairment due to vascular disease: major and minor vascular neurocognitive disorders. The term major neurocognitive disorder due to vascular disease replaces the term vascular dementia, which was used in the fourth edition of the DSM (DSM-IV). Minor vascular neurocognitive disorder is now used in lieu of mild cognitive impairment. The other aforementioned neuropsychiatric disorders which are specific to brain injuries, such as PBA or catastrophic reactions, do not have defined diagnostic criteria in DSM-5.

Investigators of depression that co-occurs with physical illness have also debated the most appropriate method for the diagnosis of these disorders when some symptoms (e.g., sleep or appetite disturbances) could result directly from the physical illness. The following four approaches have been used to assess depression in this complicated scenario:

1. *The inclusive approach:* depressive diagnostic symptoms are counted regardless of whether they may be related to physical illness.
2. *The etiologic approach:* a depressive symptom is counted only if the diagnostician feels that it is not caused by the physical illness. Of note, this is the method to be used when making DSM-5 diagnoses.
3. *The substitutive approach:* other psychological symptoms of depression replace the neurovegetative symptoms.
4. *The exclusive approach:* symptoms are removed from the diagnostic criteria if they are not found to be more frequent in depressed than nondepressed patients.

In 1997, Sergio Paradiso and colleagues examined the utility of these methods in the diagnosis of depression during the first 2 years following stroke. Among 205 patients with acute stroke, 142 patients were followed up for examination at 3, 6, 12, or 24 months following stroke. Of the 142 patients with follow-up, 60 (42%) reported the presence of a depressed mood (depressed group) while they were in the hospital, and the remaining 82 patients were nondepressed. Depressive symptoms were divided into vegetative (e.g., autonomic anxiety or subjective anergia) and psychological (e.g., worrying or hopelessness) subtypes. The only vegetative symptoms that were not significantly more frequent in the depressed compared to the nondepressed patients at the different follow-up intervals were as follows: at the initial evaluation, weight loss and early morning awakening; at 6 months, weight loss and early morning awakening; at 1 year, weight loss, early morning awakening, anxious foreboding, and loss of libido; and at 2 years, weight loss and loss of libido. Among the psychological symptoms, the depressed patients had a higher frequency of most of the psychological symptoms throughout the 2-year follow-up. The only psychological symptoms that were not significantly more frequent in the depressed than in the nondepressed group were as follows: at 3 months, suicidal plans, simple ideas of reference, and pathologic guilt; at 6 months, pathologic guilt; at 1 year, pathologic guilt, suicidal plans, guilty ideas of reference, and irritability; and at 2 years, pathologic guilt and self-depreciation.

These findings are supported by a more recent study by Cumming and colleagues from 2010. This study included 149 stroke patients at 18 months poststroke and 745 age- and sex-matched controls from the general population. A major depressive episode was diagnosed in 25 of 149 (17%) stroke patients and 61 of 745 (8%) of controls. When symptom profiles were compared across these two depressed groups; the controls reported a greater “inability to feel” in comparison to the stroke group. Otherwise, there were no statistically significant differences between symptom profiles in the groups. Moreover, a factor analysis of 10 psychological and somatic symptoms of depression showed no difference between the stroke and control groups, regardless of depression diagnosis.

In sum, these two studies suggest that the phenomenology of poststroke depression (PSD) is similar to depression in the general population. In other words, the use of the same diagnostic criteria for depression in these two populations appears to be valid. Thus, DSM-5 defines depressive disorder due to stroke according to the same symptoms as major depression occurring in patients without stroke.

EPIDEMIOLOGY

Vascular Major Neurocognitive Disorder

As discussed in the nosology section, DSM-5 now categorizes vascular dementia as an etiologic subtype of either major or mild neurocognitive disorder. Dementia and vascular dementia continue to be the popular terminologies used in published research, so we will use the terms “dementia” and “major neurocognitive disorder” (and “vascular dementia” and “vascular major neurocognitive disorder”) interchangeably in the following sections reviewing published work. In 2000, Antonio Lobo and colleagues performed a review of the prevalence of dementia in Europe based on population-based studies. Based on a total of 2,346 cases of mild to severe dementia, the age-standardized prevalence was 6.4% for all-cause dementia, 4.4% for dementia due to Alzheimer disease (dAD), and 1.6% for vascular dementia. The authors found a consistent increase in the lifetime prevalence of vascular dementia with advancing age. The pooled prevalences for those aged 65 to 69 years were 0.5% in men and 0.1% in women, while the rates were 3.6% in men and 5.8% in women 90 years and older. In 2007, the Aging, Demographics, and Memory Study estimated the prevalence of dementia in the United States. Composed of approximately 850 elders aged 71 and older, the prevalence of dementia overall was 13.9%, equating to 3.4 million people. The prevalence of vascular dementia was 2.43%. In Asia, vascular dementia was more common than dAD until the last two decades. In a review of the epidemiology of dementias in developing countries, Raj Kalaria and coauthors concluded that the prevalence of vascular dementia appears to be low except for some Latin American and Asian countries. Overall, they estimate the prevalence of vascular dementia in developing countries to be 0.6% to 2.1%. Finally, a 2021 meta-analysis by Una Clancy et al. that included 81 studies found associations between worse white matter hyperintensity (WMH) severity and apathy. Worse WMH severity was also associated with delirium and fatigue.

Hypertension has been shown to be the most common modifiable risk factor for stroke worldwide, and it may potentially have a mediating effect on the strong association between vascular-related cognitive impairment and Alzheimer disease. A longitudinal study of 173 patients with vascular dementia examined the frequency of risk factors for stroke and cerebral atherosclerosis. Although hypertension was the most common risk factor for cerebral atherosclerosis and stroke, hypotension was the most common risk factor for vascular dementia in this sample. Heart disease of the atherosclerotic type, with or without cardiac arrhythmia, was also present in the majority of cases of vascular dementia. Cardiac disease may provide a source of cerebral emboli leading to vascular dementia. Cigarette smoking of one or more packs per day was a risk factor in 21%. Hyperlipidemia type 4, that is, hypertriglyceridemia, was present in 29% of cases. Diabetes mellitus of sufficient clinical severity to require medical treatment was found in 20% of the cases, while symptomatic peripheral vascular disease with ischemic symptoms referable to the lower extremities was present in 6%. Vascular dementia was also associated with limited education suggesting that prevention efforts that are related to education may be effective in slowing or preventing the disease. Alternatively, the association with limited education may suggest some social or neurobiologic benefits of learning that may inhibit the development of this disease.

Depression

Several reviews have examined the prevalence of depression following stroke. In the most recent, Luis Ayerbe and colleagues evaluated data from 43 studies, including 20,293 patients. They found the pooled prevalence of depression observed at any time point was 29%. The prevalence rates did not differ significantly over time or in studies in different settings. The cumulative incidence of PSD was 39% to 52%

within 5 years of stroke. In another large review of 51 cohorts and 25,207 patients, Maree L. Hackett's group estimated that 33% of stroke patients experience depression. Both reviews emphasize that the estimation of the prevalence of PSD is confounded by heterogeneity in important clinical variables, such as the type of depression (e.g., syndromic vs. subsyndromic depression) or population type (e.g., hospital-based vs. outpatient sample). To address these limitations, Bob Robinson and coinvestigators conducted a systematic review of these factors in 2006. Among the 2,108 patients in community-based settings, the prevalence rate of major depression was 14.1%, while the rate of minor depression was 9.1%. A total of 2,769 patients were treated in acute or rehabilitation hospitals. Of these patients, 21.6% met the criteria for major depression, and an additional 20.0% had minor depression. Among studies of outpatients, the prevalence rates of major and minor depression were 24.0% and 23.9%, respectively. Suicidal thoughts may develop shortly after the onset of acute stroke, especially in patients with a lower educational level, a previous mood disorder, and those who developed depressive symptoms in the acute phase of stroke. In a 2015 review, Maurizio Pompili and colleagues concluded that PSD, pre-existing mood disorder, prior history of stroke, and cognitive impairment were found to be the most important risk factors for poststroke suicide.

Mania

Compared to PSD, mania after stroke is a rare neuropsychiatric occurrence, and the literature consists primarily of case reports or small case series. Catarina Santosa and colleagues performed a systematic review of poststroke mania in 2011 to better understand its demographic and clinical features. Their literature search covered approximately 50 years and yielded a total of 74 cases of mania in adult stroke patients. The authors observed the typical patient was male without a personal or family history of psychiatric illness, with at least one vascular risk factor, without subcortical atrophy, and with a right cerebral infarct. Overall, this study demonstrates the rarity of mania after stroke.

Anxiety

In their 2018 review and meta-analysis, Lena Rafsten and colleagues estimated the pooled prevalence of anxiety disorders at 29.3% during the first year poststroke. The risk was highest immediately after stroke, reducing over the year, with frequency estimation of 36.7% 0 to 2 weeks poststroke, 24.1% 2 weeks to 3 months poststroke, and 23.8% 3 to 12 months poststroke.

A meta-analysis from 2013 of 44 reports and 5,760 stroke patients found the pooled estimate of poststroke anxiety disorders, as determined by clinical interview, was 18%. When anxiety symptoms were assessed by a rating scale, the prevalence of anxiety was 25%. Three of the included studies examined the relative frequency of the different anxiety disorders and found that phobic disorders and GAD were the most common. In addition, the prevalence of anxiety tended to increase with time. The prevalence was 20% in the month following stroke, 23% between 1 and 5 months, and 24% had anxiety symptoms at 6 months or more. These results suggest that anxiety symptoms are common after stroke and tend to persist. The review also found similar prevalence estimates regardless of study setting (population-based vs. hospital vs. rehabilitation vs. community-based).

A significant number of patients with a poststroke anxiety disorder also have depression. A review of 1,445 patients within 3 months following stroke found that 49% met the criteria for GAD without comorbid depression, while 14% had comorbid GAD and major depression. Among community samples, the rate of GAD alone was 2%, and GAD with depression was 8%. Hospital and outpatient samples found that GAD alone occurred in 5.5% of cases, while GAD with depression occurred in 15.2%. Thus, anxiety disorders following stroke are frequently comorbid with a depressive disorder, although a significant number of patients have anxiety alone.

While numerous case reports exist, a paucity of data is available from systematic studies of the frequency of different types of anxiety disorders (e.g., phobia or panic disorder) after stroke. However, a growing literature now pertains to the occurrence of PTSD after stroke. For instance, Goldfinger and coinvestigators evaluated 535 patients who had a stroke or transient ischemic attack (TIA) and found that 18% likely had PTSD. In general, those with likely PTSD had worse physical and mental health. Predictors of poststroke PTSD included low income, a history of recurrent stroke or TIA, more disability, and greater medical comorbidity. Older age, the presence of a romantic partner, and having emotional support were protective against developing PTSD. While this literature is still in its infancy, such studies are a valuable contribution to studies of PTSD after a variety of acute medical events.

Psychosis

Although rare, case reports and empirical studies have documented that psychosis, primarily identified by the presence of delusions or hallucinations, may occur after a stroke. Two hospital registry studies initially evaluated the epidemiologic characteristics of poststroke psychosis. The reported frequency of psychotic disorders was 0.4% and 3.1% of stroke patients with an incidence of 1.1 per 1,000 person-years and a median time to onset of 6.1 months following stroke. Psychosis following stroke was also evaluated in a recent systematic review by Helle Stangeland et al. They included 76 studies. The average age for poststroke psychosis was 66.6 years, and delayed onset was common. The neurologic presentation was typical for stroke, but a significant minority had otherwise “silent strokes.” The most common psychosis was delusional disorder, followed by schizophrenia-like psychosis and mood disorder with psychotic features. Estimated delusion prevalence was 4.67% and hallucinations 5.05%. Twelve-year incidence was 6.7%. Lesions were typically in the right hemisphere, particularly frontal, temporal; and parietal regions, and the right caudate nucleus. In general, poststroke psychosis was associated with poor functional outcomes and high mortality. Psychosis considerably adds to the illness burden of stroke.

Apathy

Although DSM-5 does not include diagnostic criteria for this condition, apathy is typified by the absence or lack of feeling, emotion, interest, concern, or motivation and has been reported frequently among patients with brain injury, including stroke. Over the last 5 years, apathy has become one of the most intensely studied stroke-related disorders. Table 2.2–1, a summary of several recent studies, shows the reported prevalence of apathy in poststroke patients.

Table 2.2–1.
Recent Studies of the Prevalence of Poststroke Apathy

Study	N	Apathy Frequency	Associated Variables based on Multivariate Models
Brodaty et al. (2013)	Index: 135 Year 5: 70	Index: 32% Year 1: 33% Year 3: 37% Year 5: 37%	Dementia New stroke ADLs impairment
Caeiro et al. (2012)	94	38%	Cerebral hemorrhage Low education
Caeiro et al. (2013)	Meta-analysis: 2,221	36%	Depression Cognitive impairment
Mayo et al. (2009)	408	50% low score 3% high score 7% improve	High score related to poor recovery
Tang et al. (2013)	75	33%	Depression
Tang et al. (2014)	391	9%	Poor mental quality of life
van Dalen et al. (2013)	Meta-analysis: 2,706	35%	Female gender Older age Cognitive impairment Greater disability Depression

ADLs, activities of daily living.

Brodaty H, Liu Z, Withall A, Sachdev PS. The longitudinal course of post-stroke apathy over five years. *J Neuropsychiatry Clin Neurosci*. 2013;25(4):283.

Caeiro L, Ferro JM, Costa J. Apathy secondary to stroke: a systematic review and meta-analysis. *Cerebrovasc Dis*. 2013;35(1):23–39.

Caeiro L, Ferro JM, Figueira ML. Apathy in acute stroke patients. *Eur J Neurol*. 2012;19(2):291–297.

Mayo NE, Fellows LK, Scott SC, Cameron J, Wood-Dauphinee S. A longitudinal view of apathy and its impact after stroke. *Stroke*. 2009;40(10):3299–3307.

Tang WK, Chen YK, Liang HJ, et al. Location of infarcts and apathy in ischemic stroke. *Cerebrovasc Dis*. 2013;35(6):566–571.

Tang WK, Lau CG, Mok V, Ungvari GS, Wong KS. Apathy and health-related quality of life in stroke. *Arch Phys Med Rehabil*. 2014;95(5):857–861.

van Dalen JW, Moll van Charante EP, Nederkoorn PJ van Gool WA, Richard E. Poststroke apathy. *Stroke*. 2013;44:851–860.

A 2013 meta-analysis of 2,221 stroke patients found that apathy was associated with older age, depression, and cognitive impairment. The frequency of apathy did not differ when the following factors were considered: gender, stroke type (i.e., ischemic vs. hemorrhagic), and lesion location (i.e., left- vs. right-hemispheric lesions). Interestingly, the presence of apathy did not appear to affect the clinical outcome when measures of activities of daily living (ADLs) were examined. In a study of 80 acute stroke patients, 18 (22.5%) had apathy compared to 62 stroke patients without apathy. Nine (11.3%) had comorbid apathy and depression. The apathetic patients, with or without depression, were significantly older than the nonapathetic patients. Also, the apathetic patients showed significantly more severe deficits in ADLs. The nine cases who met the criteria for both apathy and depression were significantly more impaired in their ADLs than patients with apathy or depression alone. Seiji Hama and colleagues further examined the differential impact of depression and apathy on functional recovery. Based on a sample size of 237 patients with hemorrhagic or ischemic stroke, the authors found that apathy was a stronger predictor of poor functional recovery than depression. Douven et al.'s meta-analysis of studies of imaging markers also concluded that lesion type rather than lesion laterality or location might be an important factor in determining who is at risk of developing PSA in the acute as well as postacute phase.

A recent meta-analysis of 2,706 patients from 24 cohorts found a mean prevalence of apathy of 34.6% at a median of 120 days following stroke. Depression occurred in 40.1% of patients with apathy which suggests that apathy often occurs distinctly from depression. The estimated relative risk of depression among apathetic patients was 1.8. A very recent longitudinal, 1-year study found that 41.1% of patients met diagnostic criteria for apathy during the first year after stroke. Further, the mean months to onset was 3.8 ± 3.3 months, and the mean duration of meeting diagnostic criteria for apathy was 5.6 ± 2.3 months, as seen in Figure 2.2–1. In a linear mixed model, the authors controlled for age, initial severity of impairment, and major depression and found that patients in the apathy group had significantly less recovery in cognition and ADLs during the first year after stroke compared with nonapathetic patients.

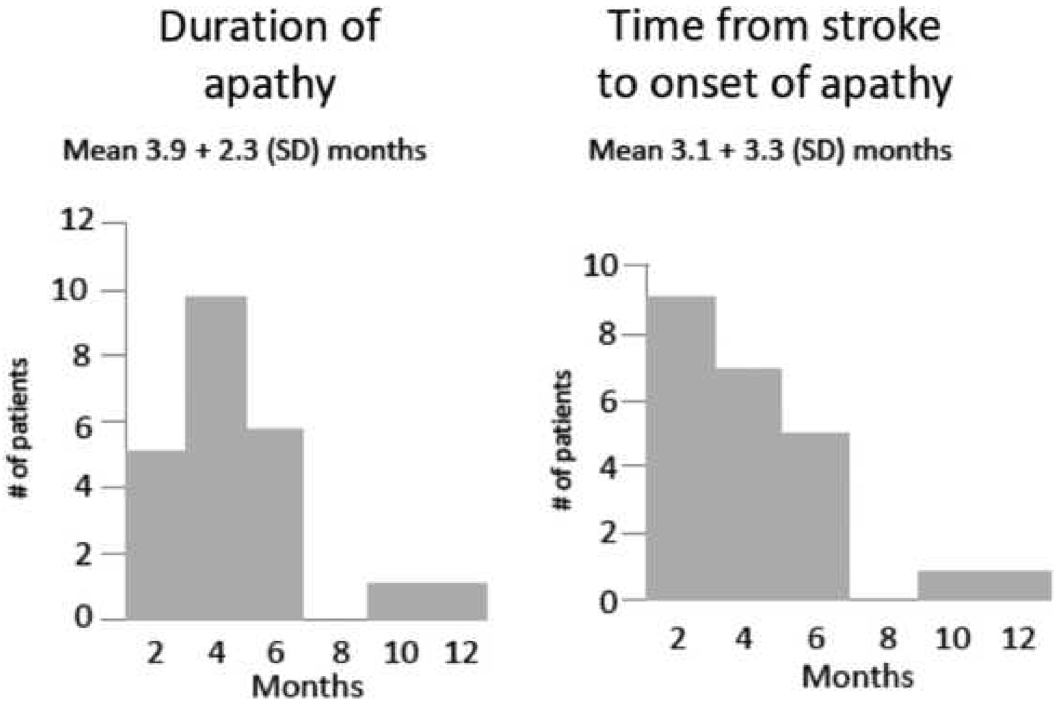

FIGURE 2.2–1. Incident apathy after stroke. **Left:** The number of patients who met diagnostic criteria as proposed by Robert et al. for apathy for the duration indicated by the right edge of the column (e.g., 5 patients had 3 months' duration, whereas in the next column 10 patients had 4.5 months' duration). **Right:** The approximate time from initial evaluation to onset of apathy. The column on the **left** shows nine patients had onset at the initial evaluation, whereas the next column shows seven patients had onset at 1.5 months (i.e., no apathy at initial evaluation but present at 3 months). Thus, except in a few cases, apathy occurred within the first 5 months after stroke and lasted almost 6 months. (Reprinted with permission from Mikami KI, Jorge RE, Moser DJ, Jang M, Robinson RG. Incident apathy during the first year after stroke and its effect on physical and cognitive recovery. *Am J Geriatr Psychiatry.* 2013;21(9):848. doi:10.1016/j.jagp.2013.03.012.)

Catastrophic Reaction

First described by Goldstein, catastrophic reactions are characterized by anxiety, tears, aggressive behavior, swearing, displacement, refusal, renouncement, and/or, sometimes, compensatory boasting. In an early study, the Catastrophic Reaction Scale (CRS), which was developed to assess the existence and severity of catastrophic reactions, was introduced. Based on a survey of 62 consecutive patients, 12 (19%) with acute stroke lesions were found to have catastrophic reactions. Of these 12, 9 (75%) met the criteria for major depression, and an additional 2 had minor depression. Further, the 12 patients with catastrophic reactions showed elevated scores on a measure of anxiety, the Hamilton Anxiety Scale. In a larger prospective study, 202 consecutive patients received the CRS within 4 days of stroke. Anger was seen in 71 (35%) patients, and 26 of these were severely angry. In a third study, the 10-item Spielberger Trait Anger Scale was used, and 47 of 145 (32%) of outpatients had evidence of the inability to control anger or aggression 3 to 12 months poststroke.

Pseudobulbar Affect

In a cohort study of 508 consecutive patients with ischemic stroke, 9.4% of patients had evidence of pathologic emotions during their admission for the care of acute stroke. Three months later, the frequency increased to 11.7%. Yet, in a separate study of 516 patients 3 months after stroke, the estimate of PBA was 7.6%. In a third study, 127 stroke patients who initially received care at a university-affiliated regional hospital were evaluated for PBA 3 months poststroke. When caregivers were involved in the assessment, the prevalence of PBA was 17.9%. In contrast, the estimate fell to 6.3% when only patients were evaluated. Such studies illustrate how relevant clinical factors (e.g., timing of study; involvement of informant) can significantly influence prevalence rates.

Anosognosia

Estimates of the prevalence of anosognosia vary considerably in the literature. Several factors contribute to this heterogeneity. For instance, the inclusion of patients with aphasia or dementia tends to conflate estimates due to significant language and cognitive limitations, respectively. Further, anosognosia is uncommon beyond 3 months poststroke. Therefore, the timing of the study assessment is another important factor in this literature. In the only meta-analysis to date, Lorenzo Pia and colleagues identified 52 studies from 1938 to 2001 for study inclusion. They found the frequency of anosognosia for hemiplegia to range from 20% to 44%, depending on when the study was conducted relative to stroke occurrence. Further, in patients with complete contralateral motor impairment, the prevalence of anosognosia was 32.3%.

ETIOLOGY

Depression

Several hypotheses about the etiology of PSD have been postulated. Numerous studies have found abnormalities in biomarkers linked to mood dysregulation, such as the serotonin transporter gene and certain inflammatory cytokines, particularly interleukins 6 and 18. Elevated levels of glucocorticoids, which affect glutamate transmission, and stroke location, which results in the interruption of biogenic amine transmission, are additional findings. Of these factors, lesion location has been a topic of considerable interest.

A systematic review and meta-analysis of studies on imaging markers for PSD by Douven et al. suggested that lesion location rather than lesion laterality or type may be an important risk factor for PSD in the postacute stroke phase. In studies performed 3 to 12 months after stroke, the prevalence of PSD is similar in those with left- and right-hemispheric lesions. However, in the first few months after stroke, lesion location does appear to increase the risk of PSD. This relationship has been the subject of a meta-analysis conducted by Robinson and colleagues who evaluated the frequency of depression within 2 months following stroke. Of the 128 patients with left anterior and left posterior strokes, the presence of a left anterior stroke increased the odds for PSD by 2.3 in comparison to the posterior lesion. In a comparison of right anterior and left anterior strokes, having a left anterior lesion increased the odds of PSD by 2.18. In a separate meta-analysis of 8 studies conducted within 6 months of stroke, the proximity of the anterior border of the stroke lesion to the left or right frontal pole and severity of PSD was evaluated. The authors found a significant inverse relationship between severity of depression and distance of the stroke from the left frontal pole, but not the right frontal pole. A third meta-analysis paints a different picture. Performed by Na Wei and colleagues, the meta-analysis consisted of 43 studies and approximately 5,500 patients. The overall analysis failed to show a relationship between lesion location and PSD. However, the impact of time since stroke was evaluated in subsequent subgroup analyses, and a modest, albeit statistically significant relationship (odds ratio of 0.79), was found between right-hemispheric stroke and PSD risk within the subacute (i.e., 1 to 6 months poststroke) group. The fourth and most recent meta-analysis performed by Douven et al., however, included 74 studies with PSD and showed a clear risk of PSD with frontal and basal ganglia lesions.

In sum, while patients appear to be at elevated risk for PSD within the first few months after stroke, meta-analyses have reported somewhat conflicting results regarding the association between PSD and lesion location. Such conflicting results exemplify a fundamental concept seen across neuropsychiatry: specific clinical symptoms are unlikely to be the product of an isolated lesion but instead due to dysfunction of key neurocircuitry. Further, a major problem with these meta-analyses, as indicated earlier, is that combining findings across studies frequently overlooks potential modifying variables. For example, specific types (i.e., major vs. minor depression), specific symptoms of depression (i.e., neurovegetative vs. psychological), and type of brain injury (e.g., ischemic vs. hemorrhagic) may be other crucial variables that are overlooked by meta-analytic techniques.

Several studies have evaluated the clinical correlates of PSD. The most consistent finding has been increased severity of impairment in ADLs in those with PSD. In 2005, Hackett and colleagues performed a systematic review of observational studies. In 9 of 11 studies evaluating this factor, disability after stroke showed a positive and significant association with PSD. In a separate review by Robinson’s group, 15 out of 18 studies (83%) found a statistically significant relationship between the severity or existence of PSD and the severity of impairment in ADLs. However, the temporal relationship between these factors (i.e., whether the depression preceded or followed the severe functional impairment) was not examined in most of these studies.

In addition, numerous studies have demonstrated that a personal history or a family history of psychiatric disorders is an important risk factor for PSD. This combination of biologic, clinical, and psychological correlations has led many investigators to hypothesize a biopsychosocial etiology for PSD. This may be a helpful clinical construct for the explanation of PSD as a whole, but it remains to be tested systematically.

Mania

Like other neuropsychiatric disorders, the etiology of mania remains unknown. The relatively rare occurrence of mania following stroke suggests that a combination of variables may be needed to produce this syndrome. For example, a study of 17 patients with stroke and mania (i.e., DSM-IV diagnosis of mood disorder due to stroke, with manic features) found that 12 had unilateral right-hemispheric lesions. The frequency of right-hemispheric lesions was significantly higher compared to 28 patients with major depression who tended to have left frontal or basal ganglia lesions, or patients with no mood disorder following stroke. Lesions associated with mania were either cortical (basotemporal or orbitofrontal cortices) or subcortical (frontal white matter, basal ganglia, or thalamus). A PET study using (^{18}F)fluorodeoxyglucose (FDG) showed focal hypometabolic deficiency in the right basotemporal cortex in three manic patients with right subcortical lesions, but this hypometabolism was not seen in seven, age-comparable, normal controls. Another systematic review of single or small series of cases of mania by Santos et al. found, among 49 patients, associations with male sex, vascular risk factors, and right-hemispheric infarcts.

Thus, mania appears to be provoked by injury to specific right-hemispheric structures that have connections to the limbic system. The right basotemporal cortex may be particularly important because direct lesions as well as distant hypometabolic effects of this cortical region, which is suggestive of diaschisis, were associated with secondary mania. The basotemporal cortex has strong efferent connections to the orbitofrontal cortex, suggesting that the lateral orbitofrontal circuitry in the right hemisphere may play a role in the etiology of mania. A combination of biogenic amine system dysfunction and release of tonic inhibitory input to the orbitofrontal–thalamic circuit may lead to the production of mania.

Another study by Starkstein et al. compared patients with secondary mania to patients with no mood disturbances. Participants were matched for size, location, and etiology of brain lesions. Patients with secondary mania had a significantly greater degree of subcortical atrophy as measured by the ratio of bifrontal and third ventricular volume to brain volume. Moreover, of the patients who developed secondary mania, those with a positive family history of psychiatric disorders had significantly less atrophy than those without such a family history. This implies that genetic predisposition to affective disorders and brain atrophy may be independent risk factors for poststroke mania. On the other hand, the systematic review of 49 cases by Santos et al. reported no associations with subcortical atrophy or family history of psychiatric disorder, and many of the studies included in this review did not examine those two factors. Therefore it may have been inadequately powered to explore those associations.

Psychosis

Information about the mechanism of poststroke psychosis is derived from anecdotal reports or small case series. One study of five patients with psychosis following stroke found that all patients had right-hemispheric lesions, primarily involving frontoparietal regions. When compared with five patients matched for age, education, lesion size, and location but no psychosis, patients with secondary psychosis had significantly greater subcortical atrophy. This atrophy was manifested by larger areas of both the frontal horn of the lateral ventricle and the body of the lateral ventricle on the side contralateral to the brain lesion. However, the majority of patients with right hemisphere stroke do not develop delusions. Neuroimaging analyses reveal that lesions of the right lateral prefrontal cortex and lesions that affect connectivity to this area correlate with the development of delusional beliefs. Several investigators have also reported a high frequency of seizures among patients with secondary psychosis. These seizures usually started after the brain lesion but before the onset of psychosis. The aforementioned study of patients with poststroke psychosis compared with lesion-matched controls found a seizure disorder among three of five patients with poststroke psychosis, as compared to zero of five stroke patients without psychosis. Overall, a right-hemispheric lesion, especially the right lateral prefrontal cortex or lesions that affect connectivity to this area, seizures, and subcortical brain atrophy appear to be important factors in the pathogenesis of poststroke psychosis.

Apathy

The mechanism of poststroke apathy has been the focus of many studies. For example, in a sample of 55 chronic stroke patients, patients who showed little behavioral response to positive reinforcement in the laboratory reported higher levels of apathy. These patients were also found to have lesions of the ventral putamen and globus pallidus, as well as the dorsal thalamus and prefrontal cortex. Another study of 106 acute stroke patients found that apathy at the time of acute stroke predicted apathy at 1-year follow-up. A meta-analysis reported that apathy among 2,221 patients was associated with PSD and greater cognitive impairment. Imaging studies have found that apathy is significantly associated with lesions of the pons, frontal cortex, basal ganglia, dorsal thalamus, posterior limb of the internal capsule, and temporal cortex (Table 2.2–2).

Overall, lesions along the anterior cingulate subcortical circuit (including cingulate gyrus, frontal cortex, ventral striatum, ventral pallidum, and magnocellular dorsomedial thalamus) have been frequently associated with the occurrence of apathy and may constitute an anatomical substrate for poststroke apathy.

Table 2.2–2.
Recent Imaging Findings in Poststroke Apathy

Author	N	Caseness	Main Findings
Hama et al. (2007)	243	126 with apathy or depression	Apathy associated with bilateral BG damage; PSD associated with left frontal lobe damage
Murakami et al. (2013)	149	149 total cases	Apathy associated with lesions of brainstem and bilateral striatum
Onoda et al. (2011)	102	37 with apathy, 65 controls	CBF reduced in basal ganglia in apathetic cases; hypoperfusion in bilateral-basal ganglia associated with presence of apathy
Tang et al. (2013)	185	20 with apathy, 165 controls	Pontine infarcts associated with apathy; volume of pontine infarct significantly predicted occurrence of poststroke apathy

BG, basal ganglia; CBF, cerebral blood flow; PSD, poststroke depression.

Hama S, Yamashita H, Shigenobu M, et al. Post-stroke affective or apathetic depression and lesion location: left frontal lobe and bilateral basal ganglia. *Eur Arch Psychiatry Clin Neurosci*. 2007;257(3):149–152.

Murakami T, Hama S, Yamashita H, et al. Neuroanatomic pathways associated with poststroke affective and apathetic depression. *Am J Geriatr Psychiatry*. 2013;21(9):840–847.

Onoda K, Kuroda Y, Yamamoto Y, et al. Post-stroke apathy and hypoperfusion in basal ganglia: SPECT study. *Cerebrovascular Dis*. 2011;31(1):6–11.

Tang WK, Chen YK, Liang HJ, et al. location of infarcts and apathy in ischemic stroke. *Cerebrovascular Dis*. 2013;35(6):566–571.

Catastrophic Reaction

In a study of 62 patients with acute stroke, those demonstrating catastrophic reactions had a significantly higher frequency of lesions involving the basal ganglia compared to acute stroke controls. When 10 depressed patients with a catastrophic reaction were compared to 10 depressed patients without a catastrophic reaction, the catastrophic reaction group had significantly more anterior lesions, which were located primarily in subcortical regions (i.e., eight of nine depressed patients with catastrophic reaction had subcortical lesions; three of nine depressed patients without catastrophic reaction had subcortical lesions).

In another study, catastrophic reactions occurred in 12 of 62 patients (19%) admitted to the hospital with acute stroke. Patients with catastrophic reactions were found to have a significantly higher frequency of familial and personal history of psychiatric disorders, mostly depression, than patients without catastrophic reactions. Catastrophic reactions, however, were not significantly more frequent among aphasic compared with nonaphasic patients. This finding did not support the contention that catastrophic reactions represent an understandable psychological response of “frustrated” aphasic patients. Furthermore, 9 of 12 (75%) patients with catastrophic reactions also had PSD, 2 (17%) had minor depression, and only 1 was not depressed. Thus, the catastrophic reaction was closely linked with PSD.

Based on these findings, catastrophic reactions may result from neurophysiologic dysfunction rather than the realization of intellectual impairment. Catastrophic reactions occurred predominantly in patients with major depression associated with anterior subcortical lesions. Subcortical damage has also been hypothesized to underlie the “release” of an emotional display by removing inhibitory input to limbic areas of the cortex.

Pseudobulbar Affect

Pathologic emotions have traditionally been explained as secondary to the bilateral interruption of descending neocortical upper motor neuron innervation of bulbar motor nuclei. Some patients with pathologic emotions have such bilateral lesions, but this is not a uniform finding. For instance, one study found that patients with frontal or temporal lesions in either hemisphere had a significantly increased frequency of pathologic emotions. Examination of lesion size and location in 12 patients with pathologic crying found that patients with the most frequent crying episodes had relatively large bilateral pontine lesions. The intermediate group had large bilateral lesions. The least affected patients had relatively large unilateral subcortical lesions. Such evidence has led to the hypothesis that pathologic emotions may arise from the partial destruction of raphe serotonergic neurons or their projections. There is also increasing evidence of cerebellar involvement, and investigators at the

University of Iowa have suggested that the critical lesions eliciting PBA are located along fronto-ponto-cerebellar pathways.

DIAGNOSIS AND CLINICAL FEATURES

Mild and Major Vascular Neurocognitive Disorders

The DSM-5 classification of vascular cognitive impairment (VCI) includes both patients with dementia (i.e., major vascular neurocognitive disorder) and its precursor (i.e., mild neurocognitive disorder). The following three conditions must be met to reach diagnostic criteria for major vascular neurocognitive disorder:

1. Criteria are met for major neurocognitive disorder (i.e., the presence of acquired impairment in one or more cognitive domains and significant impairment in ADLs not better explained by another medical condition);
2. The clinical features are consistent with a vascular etiology as suggested by either (a) cognitive deficits that are temporally related to one or more cerebrovascular events or (b) presence of prominent executive dysfunction (e.g., inattentiveness or slow processing speed);
3. The patient’s history, physical examination, and/or neuroimaging findings are consistent with cerebrovascular disease of sufficient severity to account for patient’s cognitive impairment.

Patients with mild vascular neurocognitive disorder have impairment in one or more cognitive domains, but daily functioning remains relatively intact. However, affected individuals may take longer than usual or engage in compensatory techniques to overcome their cognitive disturbances. Cognitive impairment due to vascular disease is expected to occur in the presence of multifocal motor and/or sensory deficits on the physical examination. A single vascular lesion could result in a clinical picture consistent with major vascular neurocognitive disorder.

Depression

The assessment for depression in patients with stroke or other physical illness has been a topic of considerable debate. According to DSM-5, two types of poststroke depressive disorders are recognized: depressive disorder due to stroke with major depression-like episode and depressive disorder due to stroke with depressive symptoms (i.e., subsyndromal depression). The diagnostic criteria, however, require the clinician to determine whether they believe that the mood disorder is the direct physiologic consequence of the stroke. If this judgment is made, then the patient is diagnosed with “depression due to stroke.”

The problem with this diagnostic schema is that even the most experienced clinician cannot be certain if the depression is due to the direct physiologic consequences of the stroke. Even if the patient has a history of depression prior to the stroke, the physiologic response to brain injury may or may not provoke a new depressive episode. Furthermore, the clinical manifestations of depression appear to be the same regardless of whether the depression followed a stroke. That is, empiric evidence has shown that all the symptoms of major depression with poststroke onset have been shown to be significantly more frequent in depressed compared with nondepressed stroke patients. Thus, the symptoms of depressed mood, anhedonia, weight loss, insomnia, psychomotor agitation or retardation, loss of energy, worthlessness, poor concentration, and suicidal thoughts characterize patients with PSD just as well as depressed patients without structural brain injury. Thus, a better diagnostic system may be to make stroke a specifier akin to *postpartum onset.* Namely, the diagnosis of “major” or “minor” depression would be qualified with *poststroke onset.* Empirical data appears to support such nomenclature more so than the current diagnostic classification.

To promote the early detection of PSD in patients hospitalized for acute stroke, Janneke de Man-van Ginkel and colleagues developed the Post-Stroke Depression Prediction Scale. The authors validated this scale in a study of 410 stroke patients. Within the first week after stroke, participants were assessed for various PSD predictive factors and then reassessed 6 to 8 weeks after stroke. Based on a multivariate model, a medical history of depression or other psychiatric disorder was the most important predictor (odds ratio 7.2) of PSD. Other predictive factors included a medical history of hypertension, angina pectoris, and impaired ability to dress oneself based on the dressing item of the Barthel Index of ADLs. Further, the lowest risk category (<10 sum score) at 1-week poststroke was associated with a 2% risk of depression at 6 to 8 weeks follow-up, while 82% of the highest-risk category (>21 sum score) developed depression at 6 to 8 weeks.

Mania

The symptoms of mania were examined in a series of 25 consecutive patients who met DSM-IV criteria for a mood disorder due to brain injury with manic features. These patients, who developed mania after a stroke, traumatic brain injury, or tumors, were compared to 25 patients with primary mania (i.e., no known neuropathology). Both groups of patients showed similar frequencies of elation, pressured speech, flight of ideas, grandiose thoughts, insomnia, hallucinations, and paranoid delusions. Thus, the symptoms of mania that occurred after stroke (now diagnosed by DSM-5 as bipolar disorder due to stroke with manic-like episode if full criteria for mania are met, or bipolar disorder due to stroke with manic features if full criteria for mania are not met) appear to be the same as those found in idiopathic cases of mania (i.e., bipolar disorder). As with depression, a better diagnostic classification might be mania with *poststroke onset.*

Anxiety

A 2013 study including 374 patients found that 23% had anxiety defined by a score of eight or more on the Anxiety Section of the Hospital Anxiety Depression Scale. Patients with poststroke anxiety had a poor stroke-related quality of life in general and in several areas, including energy, mood, personality, thinking, and work. Another 2013 study looked at long-term outcomes of poststroke anxiety in 220 patients followed longitudinally. At 5 years poststroke, 64 (29%) of patients were deemed anxious, and 73 (33%) were depressed, with anxiety and depression comorbidity in 43 (20%). Higher anxiety scores at 6 months following stroke were associated with anxiety at 5 years poststroke but not ADLs scores at 6 months. Finally, an imaging study of 693 patients with acute stroke reported that the anxious patients, as defined by a score of eight or more on the Hospital Anxiety Depression Scale, had a significantly increased frequency of right-frontal infarcts even after controlling for other relevant variables. In sum, anxiety is common after a stroke and its prevalence remains elevated even several years out from the stroke. Since anxiety (along with depression) is associated with poor recovery, regular assessments for anxiety (and PSD) are warranted.

The following case illustrates a typical clinical course of poststroke anxiety and its frequent comorbidity with depression:

A 71-year-old farmer suffered a thrombosis of the basilar artery. He developed blurring of his vision, gait impairment, and paresthesias of his face. Within 2 months of the stroke, the patient developed panic attacks and GAD. His GAD was characterized by almost constant anxiety about minor issues, insomnia, restlessness, and poor concentration. The panic attacks were typified by the rapid onset of anxiety with tachycardia, shortness of breath, sweating, and fearfulness that he would pass out or die from another stroke or a heart attack. The panic attacks occurred first when he was away from home and later while he was at home. The panic attacks were controlled by alprazolam, but in spite of taking this medication four times per day, he continued to have significant anxiety marked by worrying, restlessness, muscle tension, poor concentration, and insomnia. About 2 months later, the patient also developed depression with symptoms of low mood, loss of interest, poor concentration, self-blame, hopelessness, and psychomotor slowing. He responded to electroconvulsive therapy but relapsed quickly. He was then treated with nortriptyline, which led to remission of both his depressive and anxiety disorders.

Apathy

The clinical manifestations of apathy were utilized by Marin and Starkstein in the 1990s to develop an apathy rating scale and to propose diagnostic criteria. The essential symptom of apathy is the loss of motivation or drive which is the required symptom for the diagnosis of apathy. Lack of motivation may be manifested emotionally, cognitively, or behaviorally as reflected in the recently proposed diagnostic criteria for apathy by Roberts and colleagues, as reflected in Table 2.2–3.

The following vignette illustrates a case of poststroke apathy and some of its differences from other conditions:

A 55-year-old office manager was recovering from a right-prefrontal ischemic stroke, believed to be a consequence of poorly controlled hypertension and obesity. A week after his admission, a psychiatric consultation was requested to evaluate him for depression. His wife and physical therapist reported that he was not engaging in rehabilitation as expected and that he seemed “disconnected.” He avoided therapy by saying, “I don’t have energy now,” even when he maintained a cooperative and polite attitude. He denied feeling sad but reported “feeling disinterested and flat.” He was otherwise hopeful about his recovery. Visits from his family did not seem to cheer him up as they did previously, and his daughter commented that he seemed “as if he had lost his sparkle.” The patient denied feeling like crying and his family concurred that he didn’t display crying spells or abnormal laughing. While he did not have a history of psychiatric or neurologic conditions, the staff was concerned about poststroke depression. His routine blood work-up was unremarkable and imaging studies did not yield any new findings. His most recent neurologic examination revealed improved mild-left hemiparesis, stronger muscle strength, and steadier ambulation. The confusional state that was documented at the time of his admission seemed to have resolved and he was completely oriented to himself, the date, his surroundings, and the circumstances of his hospitalization. He recognized all his family members and clinicians and was not suspicious. He was attentive during conversations but seemed easily tired. There were no fluctuations in his mental state throughout the day, and at night, his sleep was reported as sound. His appetite was reduced but was gradually improving. It was thought that recovering from delirium could have been the most likely cause of the patient’s symptoms, and therefore observation and monitoring as an outpatient were recommended. Unfortunately, in his first ambulatory medical follow-up 3 weeks later, his family reported that his lack of energy and motivation had worsened and that he would get angry and refused to engage in his recovery meaningfully. The wife was concerned about cancelation of rehabilitation services by the medical insurance. Once again, the screening for depression was negative. A diagnosis of poststroke apathy was made, and he was initiated a low dose of methylphenidate. Within a couple of weeks of this intervention, the patient started to participate in regular sessions of physical therapy and perform the prescribed exercises at home in between them.

Table 2.2–3.
Questions from the Apathy Scale Used to Determine the Existence of Lack of Emotion, Slowed Activity, and Impaired Cognitive Activity Seen in Apathy

Lack of Emotion	Slowed Motor Activity	Impaired Cognitive Activity
Are you concerned about your condition?	Does someone have to tell you what to do each day?	Are you interested in learning new things?
Are you indifferent to things?	Do you put much effort into things?	Do you have anything that interests you?

The proposed criteria for the diagnosis of apathy is lack of motivation (e.g., “Do you have motivation?”) plus two of the three symptoms above (i.e., lack of emotion, slowed motor activity, and/or impaired cognitive activity) present most of the time for at least 4 weeks based.

Data from Mikami K, Jorge RE, Moser DJ, et al. Prevention of poststroke apathy using escitalopram or problem-solving therapy. *Am J Geriatr Psychiatry*. 2013;21(9):855–862.

Catastrophic Reactions

Currently, no diagnostic criteria are available for catastrophic reactions. As described earlier in the chapter, patients typically present with sudden outbursts of refusal, swearing, tears, and sometimes aggressive behavior. This intense emotional reaction is usually provoked by stressful situations such as cognitive testing and typically subsides within a few minutes.

Pseudobulbar Affect

There are no established diagnostic criteria for PBA. This condition is characterized by the sudden onset of crying, or more rarely, laughing, which is out of proportion to the conversation or situation in which the emotional reaction occurred. The emotion may last from a few seconds to a few minutes with no residual feelings of sadness or happiness.

At least five studies have examined the clinical features of pathologic emotions. The hallmark features of PBA are as follows: an inability to control crying or laughter; an increased frequency of such emotional display; and recognition by the patient that the emotional display is inconsistent or excessive to their underlying emotional feelings. Several rating scales are available for PBA, but only the PLACS scale was specifically developed for poststroke pathologic emotions. As seen in Figure 2.2–2, this scale assesses the existence and severity of pathologic emotion among patients with stroke.

PATHOLOGY AND LABORATORY EXAMINATION

Mild and Major Vascular Neurocognitive Disorders

The clinical identification of vascular neurocognitive disorders requires a medical history, neurologic examination, psychiatric interview, and neuropsychological assessment. Diagnostic testing also plays an important role and is informed by blood chemistries (chiefly B_{12}, folate, and thyroid function), head CT and/or brain MRI, and, in certain cases, cerebral angiography. These laboratory and imaging data are used to evaluate for potentially treatable forms of dementias caused by tumor, vascular malformation, cerebral hematoma, normal pressure hydrocephalus, infections, endocrine abnormalities, or vitamin deficiencies. Structural imaging studies (i.e., head CT or brain MRI) should document the existence of one or, more likely, several cerebrovascular lesions and, when present, the distribution of the cerebral volume loss associated with those lesions. Cerebral angiography may identify vascular lesions (e.g., aneurysms or cerebral artery stenosis) that may be reversible through interventional radiology or neurosurgical procedures. Additional tests that may be helpful include CSF analysis to rule out infectious causes as well as other causes of dementia like Alzheimer disease and paraneoplastic syndromes, electroencephalogram (EEG) to help with prognostication after acute ischemic stroke, differentiate delirium from dementia, and also differentiate vascular dementia from Alzheimer disease, and EEG with evoked responses to evaluate for functional limitations after stroke (visual-evoked potential evaluates the visual pathway, brainstem evoked potential evaluates stroke that occurs in brain structures supplied by the vertebrobasilar system, and somatosensory and motor evoked potentials represent a useful early prognostic marker of functional recovery in ischemic stroke).

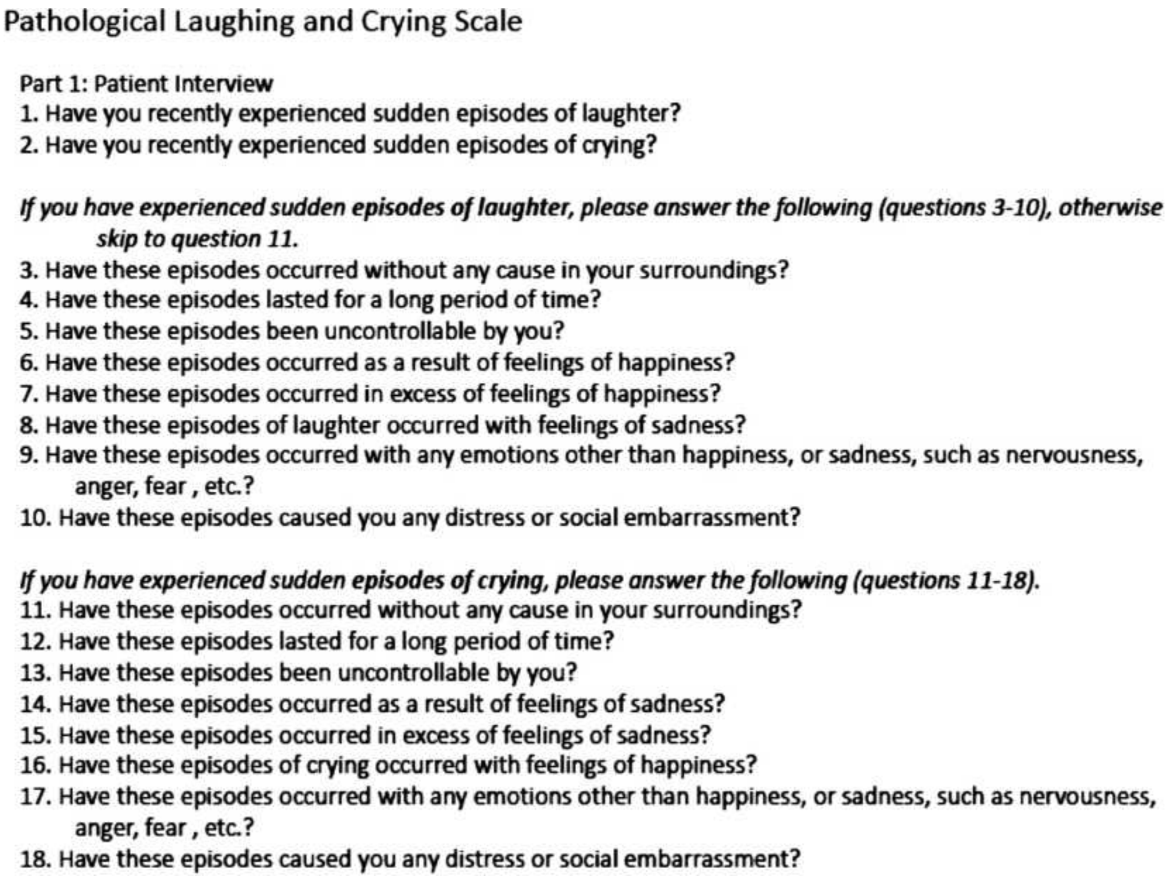

Pathological Laughing and Crying Scale

Part 1: Patient Interview

1. Have you recently experienced sudden episodes of laughter?
2. Have you recently experienced sudden episodes of crying?

If you have experienced sudden episodes of laughter, please answer the following (questions 3-10), otherwise skip to question 11.

3. Have these episodes occurred without any cause in your surroundings?
4. Have these episodes lasted for a long period of time?
5. Have these episodes been uncontrollable by you?
6. Have these episodes occurred as a result of feelings of happiness?
7. Have these episodes occurred in excess of feelings of happiness?
8. Have these episodes of laughter occurred with feelings of sadness?
9. Have these episodes occurred with any emotions other than happiness, or sadness, such as nervousness, anger, fear , etc.?
10. Have these episodes caused you any distress or social embarrassment?

If you have experienced sudden episodes of crying, please answer the following (questions 11-18).

11. Have these episodes occurred without any cause in your surroundings?
12. Have these episodes lasted for a long period of time?
13. Have these episodes been uncontrollable by you?
14. Have these episodes occurred as a result of feelings of sadness?
15. Have these episodes occurred in excess of feelings of sadness?
16. Have these episodes of crying occurred with feelings of happiness?
17. Have these episodes occurred with any emotions other than happiness, or sadness, such as nervousness, anger, fear , etc.?
18. Have these episodes caused you any distress or social embarrassment?

FIGURE 2.2–2. Pathological Laughter and Crying Scale. (Reprinted with permission from Robinson RG. *The Clinical Neuropsychiatry of Stroke: Cognitive, Behavioral and Emotional Disorders following Vascular Brain Injury*. 2nd ed. Cambridge University Press; 2006:430.)

Poststroke Depression

Currently, no laboratory or neuroimaging studies are routinely used to inform the clinical diagnosis of PSD (except for B_{12}, folate, hemoglobin, and thyroid function, which are routinely done to rule out other causes of depression symptoms). This topic, nevertheless, is an area of considerable interest for the scientific community. Studies to date have largely focused on endocrine markers linked to major depression. For instance, the dexamethasone suppression test (DST) has been investigated as a possible biologic marker. A total of nine studies involving 327 patients demonstrated a statistical association between PSD and a failure to suppress serum cortisol in response to the administration of dexamethasone. The mean pooled data specificity of the test was 87%, while the sensitivity was 47%. These values are insufficient to be diagnostically useful. The inadequacy of the DST as a diagnostic tool was also seen in a study of 65 patients whose acute strokes had occurred within the preceding year. Of these cases, 67% with major depression failed to suppress serum cortisol compared to 25% of patients with minor depression and 32% of nondepressed patients. The sensitivity of the DST for major depression was 67%, with a positive-predictive value of 50%. The specificity was only 70%, with a negative-predictive value of 80%. False-positive tests, found in 33% of patients, were related to large lesion volumes.

In a different line of work, a study of growth-hormone response to desipramine found that growth-hormone responses were significantly blunted in patients with PSD. These results suggest that diminished α_2-adrenergic receptor function may be an important marker for PSD. The sensitivity of the test was 100%, and the specificity was 75%. Future studies of endocrine responses as markers of PSD may prove to have diagnostic validity. Alternatively, such markers may help define more homogeneous subtypes of PSD for research purposes.

Another prominent line of inquiry revolves around identifying various biomarkers that predict the risk of PSD. The majority of this research has focused on the numerous inflammatory biomarkers that are deranged in patients with PSD compared to patients with stroke who do not develop depression, including decreased serum BDNF concentrations, increased early markers of inflammation like C-reactive protein and ferritin, serum proinflammatory cytokines (TNF-α, IFN-γ), as well as proinflammatory/anti-inflammatory ratios, lowered complement expression, decreased serum vitamin D levels, hypercortisolemia and blunted cortisol-awakening response.

Neuroimaging markers have also been identified. Lesions affecting the frontal-subcortical circuitry of mood regulation (prefrontal cortex, basal nuclei, and thalamus) predominantly in the left hemisphere appear to predict PSD 1 year after stroke. Additional pontine and lobar cerebral microbleeds in acute stroke patients, as well as severe microvascular lesions of the brain, increase the likelihood of PSD.

Other Disorders

The utility of laboratory examinations in the diagnosis or prognosis of poststroke mania, anxiety disorders, psychosis, apathy, catastrophic reactions, and anosognosia has not been established except as discussed under etiology.

COURSE AND PROGNOSIS

Mild and Major Vascular Neurocognitive Disorders

The course of major vascular neurocognitive disorder is characterized by a close temporal relationship between stroke occurrence and deterioration of cognitive function. The probability of recurrent stroke is approximately 7% per year. The course and prognosis of poststroke cognitive impairment, however, can be influenced by preventative measures.

The Framingham Heart Study was the first to prospectively examine the effect of new-onset stroke on cognition, as reported in 2014. In this study, 132 new-onset, dementia-free stroke patients were prospectively followed and received serial cognitive testing. At 6 months after stroke, significantly greater impairment in the following cognitive domains was seen: immediate recall of logical and visual memory, verbal learning, naming, executive function, visuospatial, and motor skills. These deficits were independent of prestroke cognitive function and vascular risk factors.

A 2014 update on subcortical vascular neurocognitive disorder noted that subcortical vascular dementia is driven by severe stenosis of the small vasculature of white matter and small vessel disease. In turn, these processes lead to white matter ischemia, slow progression of symptoms, and cholinergic neuronal damage that makes it difficult to differentiate cognitive impairment due to subcortical vascular disease from dAD.

In their prospective study on survivors of incident stroke, Levine and colleagues found an association between acute decline in cognitive function and accelerated and persistent cognitive decline over 6 years.

Depression

The longitudinal course of PSD has been examined in a number of studies. At the time of the initial acute or rehabilitation hospital evaluation (e.g., 1-week poststroke), 21.6% of patients had a symptom cluster consistent with major depression, and 20% had a symptom cluster of minor depression (i.e., subsyndromal depression). Although both major and minor depressive disorders appear to continue for months, the observed course has varied from one study to the next. Philip Morris and colleagues calculated a mean duration of major depression of 39.0 ± 31.8 weeks and a mean duration of minor depression of 12.2 ± 18.2 weeks. Another report by Robinson and colleagues of 142 patients over 2 years following acute stroke found a mean duration of 31.2 weeks for major depression but 11.9 months for minor depression. A mean of 26% of patients continued to have major depression 1 year after initial diagnosis. A study of 1,233 stroke patients followed over 15 years found that 50% of patients with PSD at 3 months had recovered by 1 year. A 2013 meta-analysis of 4,022 patients from 50 studies found that among patients depressed a few months following a stroke, recovery at 1 year ranged from 15% to 57%. In sum, these findings indicate that a significant number of patients who develop either major or minor depression following stroke may experience persistent depression for 2 or more years after stroke.

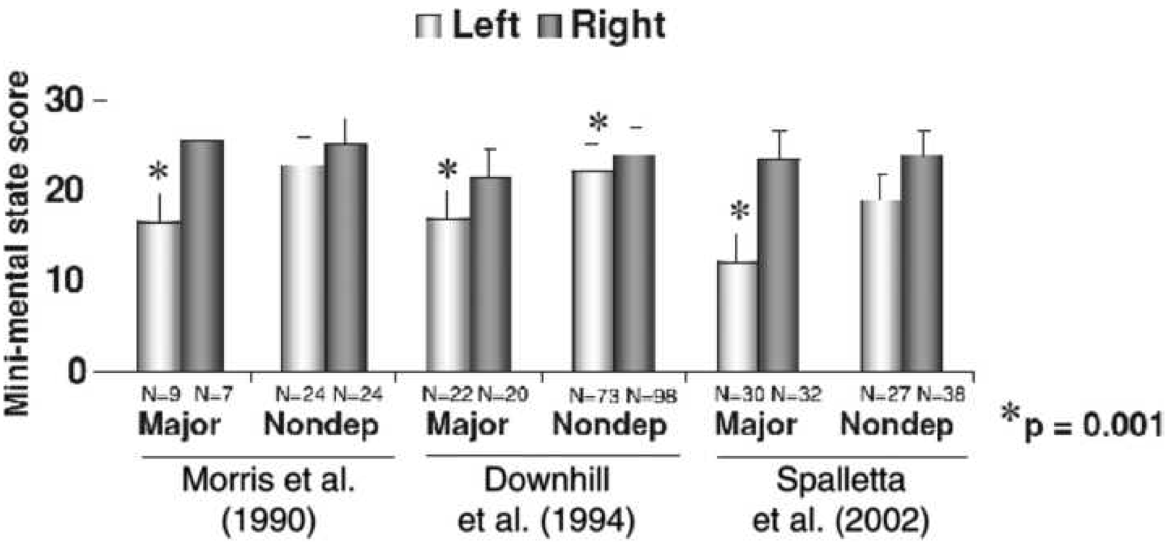

FIGURE 2.2–3. Cognitive impairment related to depression and lesion site. Mini-mental state examination scores following acute stroke in three studies among patients with major or no mood disturbance grouped according to the hemisphere of ischemia. In all three studies, there was a significant difference between patients with major depression following left-hemispheric stroke and nondepressed patients with similar lesions (*P* = .001). Major depression following right-hemispheric lesions did not lead to the same phenomenon. (Reprinted with permission from Robinson RG. *The Clinical Neuropsychiatry of Stroke: Cognitive, Behavioral and Emotional Disorders following Vascular Brain Injury*. 2nd ed. Cambridge University Press; 2006:155.)

Nearly all the studies that have examined the relationship between depression and physical/functional recovery from stroke (as measured by ADLs) have found that the more severe the depression, the more impaired patients were in ADLs. Furthermore, a subset of these studies has demonstrated that depression severity after acute stroke is an independent predictor of the severity of ADLs impairment 1 or more years after stroke. Thus, the prognosis for recovery in ADLs is significantly worse if a patient has depression following their acute stroke. Interestingly, Cevdet Bilge and colleagues found that patients with PSD who responded to treatment with citalopram showed significantly better improvement in ADLs than patients who were never depressed.

Similarly, the prognosis for cognitive functioning is also dependent, in part, upon the existence of PSD. Three separate studies have shown that major depression following acute stroke is associated with more severe cognitive impairment if the stroke occurred in the left hemisphere (Fig. 2.2–3). This effect of right- versus left-hemispheric stroke is not seen among nondepressed patients with similar lesions.

Longitudinal studies have also shown that patients with major depression following left-hemispheric stroke have greater severity of cognitive impairment in the 12 months following a stroke in comparison to patients with comparable stroke but no major depression. Between 12 and 24 months following a stroke, however, cognitive function among these patients tends to improve. By 24 months of follow-up, no difference exists in the severity of cognitive impairment among patients regardless of lesion location (i.e., right- vs. left-hemispheric strokes) or depression status (i.e., depressed or nondepressed patients) following the acute stroke.

Importantly, PSD has a dramatic association with increased mortality. The first study to report this phenomenon using standardized interviews and diagnosis was performed in 1993 on a sample of 103 acute stroke patients. At 10 years poststroke, patients with major or minor depression had a 3.4-fold increase in the odds of mortality in comparison to those with no depression. Another study of 976 patients with stroke found that patients with depression, assessed at 3 weeks poststroke using the Wakefield Self-Assessment Depression Inventory, had 50% higher mortality at 1 year compared to nondepressed patients. A 2013 meta-analysis of 50 studies found that patients with depression within 3 months of stroke had poor clinical outcomes defined by lower quality of life, increased disability, and increased mortality at follow-up compared with nondepressed stroke patients.

Perhaps the most provocative finding, however, is the relationship between mortality following PSD and treatment with antidepressant therapy. A 9-year follow-up of patients who had been treated for PSD found that active treatment with nortriptyline or fluoxetine ($n = 53$) versus placebo ($n = 28$) over 12 weeks resulted in an increased probability of survival (i.e., 59.2% for treated vs. 36.4% for placebo patients) at 6 years follow-up (Fig. 2.2–4). A logistic regression that examined the effects of age, diabetes, relapsing depression, and antidepressant use found that antidepressant use and the existence of diabetes mellitus independently and significantly predicted survival.

A series of recent findings demonstrated that the use of fluoxetine (20 mg/day) or nortriptyline (100 mg/day) was associated with improved recovery from a stroke at 3 or 12 months, respectively, after 3 months of treatment. Interestingly, this effect was independent of PSD.

The course of poststroke mood disorders is exemplified by the following case history:

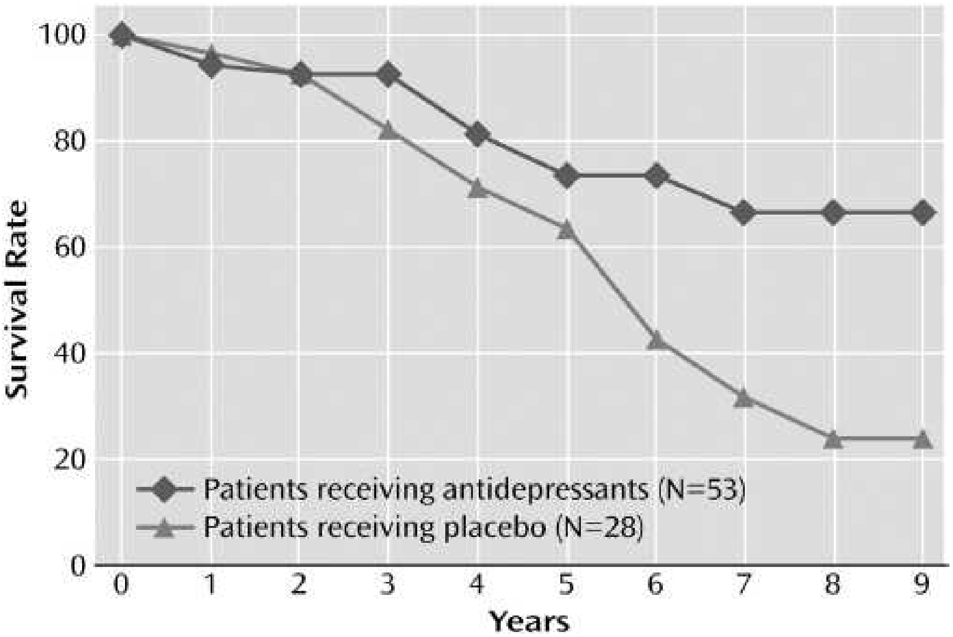

FIGURE 2.2–4. Poststroke mortality after 3 months of treatment. Survival rates over 9-year follow-up for stroke patients who received a 12-week course of antidepressants or placebo during the first 6 months following stroke. Probability of survival was significantly greater in the patients receiving antidepressants ($\chi 2$ = 8.2, df = 1, *P* =.004, Kaplan–Meier survival analysis, log-rank test). (Reprinted with permission from Jorge RE, Robinson RG, Arndt S, Starkstein S. Mortality and poststroke depression: a placebo-controlled trial of antidepressants. *Am J Psychiat.* 2003;160(10):1827. (Copyright © 2003). American Psychiatric Association. All Rights Reserved. doi: 10.1176/appi.ajp.160.10.1823.)

Ms. A. was a 35-year-old woman who had been the regional director of marketing for a national company. Deadlines, frequent travel, and sales quotas were all part of her high-pressure work. She developed hypertension during her first pregnancy, but in spite of this, she kept up her hectic work schedule. While on a business trip during her subsequent pregnancy, however, she suffered a stroke that caused mild weakness on her right side as well as aphasia characterized by difficulty producing speech but not comprehension (i.e., nonfluent, expressive aphasia). These motor and language impairments were relatively mild and cleared up within several months of the stroke.

When she was approximately 6 months poststroke, she was convinced that there was still something wrong with her as a result of the stroke. Several physicians told her that there was nothing physically wrong with her and all she needed to do was to get back to work. She had never experienced prolonged depressive symptoms prior to the stroke and her affect did not appear depressed at the time of the assessment. She was talkative, and her thoughts and speech were not slowed, as often occurs in depression. She was not tearful or suicidal. She did, however, feel depressed and had a loss of interest, concentration, and motivation. She had returned to work for a couple of hours a day but was unable to concentrate well enough to accomplish even simple tasks. She had lost interest and pleasure in virtually all of her work and social activities. She no longer had the ambition to climb the corporate hierarchy. She also had sleep disturbance with early morning awakening, loss of appetite and weight, decreased sexual interest, and decreased energy. A diagnosis of PSD was made, and antidepressant therapy with nortriptyline was started. Her response to antidepressant treatment was dramatic. Between 4 and 6 weeks after beginning nortriptyline, her mood greatly improved and she returned to work. She found that she was able to concentrate and experience interest and pleasure in her work. Over a period of 2 to 3 months, she reverted from somebody who was virtually immobilized by depression vocationally and socially to an effective, energetic woman. She also had a return of some of her previous ambition, though she still lacked her prior drive to reach the top of the corporate hierarchy.

After 9 months of taking nortriptyline, she wanted to discontinue this medication. Her rationale was that she felt she had fully recovered and did not want to continue taking medication that produced a dry mouth and constipation. The medication was tapered over a period of about 6 weeks and then stopped. She remained well for approximately 1 year but then had a recurrence of the same symptoms that were initially observed. Namely, she was uninterested in work, had no feeling of pleasure in any of her usual activities, and was unable to attend to the demands of work or home. She also had difficulty sleeping, lost her appetite, and felt depressed. These symptoms subsided after restarting her antidepressant medication, which she continued to take for another year. After that year, she again insisted on stopping her antidepressant medication. Over the next 2 years of follow-up, she remained free of depressive symptoms, but she was counseled that her previous history placed her at increased risk for another recurrence of depression. She was advised to remain vigilant for any signs or symptoms of her depression and to report them immediately so resumption of treatment could be discussed in a timely fashion.

Mania

The course of mania following stroke has not been systematically examined. Anecdotal cases have been reported indicating that recurrent episodes of mania or depression may occur in these patients (Fig. 2.2–5).

Anxiety

The prevalence of GAD and major depression based on standardized interviews and diagnostic criteria was documented in two separate studies (Fig. 2.2–6). Note that the rates of GAD with and without comorbid depression are stable at about 20% over 3 years poststroke. A 5-year longitudinal study of 220 patients found that at 5 years poststroke, the prevalence of anxiety was 29%, and those with anxiety at 6 months were significantly more likely than nonanxious patients to have anxiety at 5 years. Another 10-year study reported that anxiety ranged from 17% to 24%, with a cumulative incidence of 57% over 5 years. Further, comorbid anxiety and depression each year ranged from 57% to 73%.

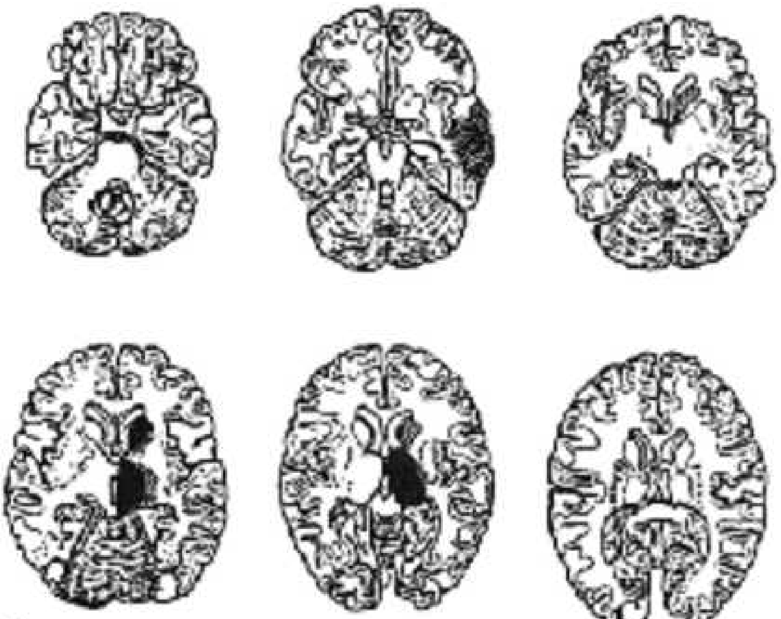

Bipolar Group

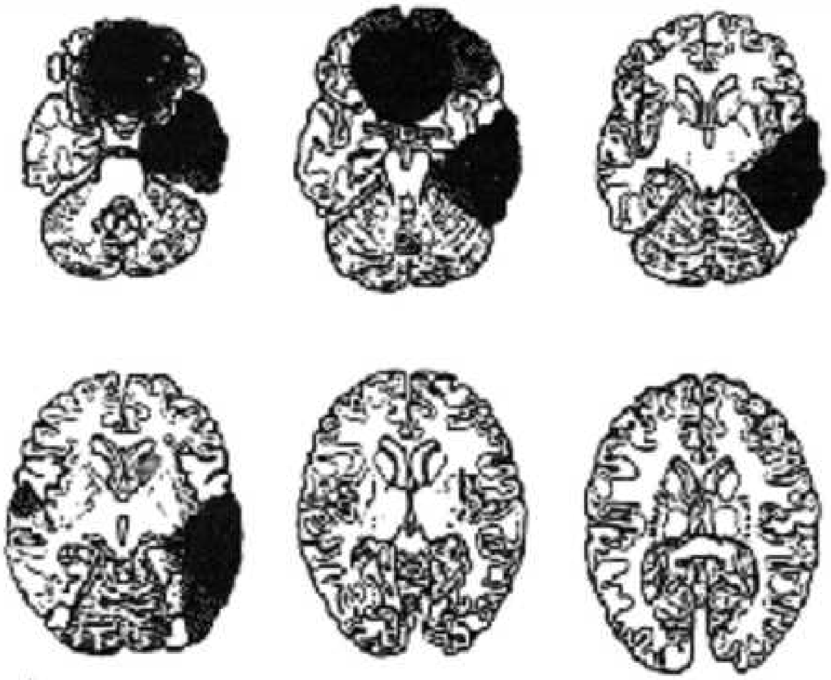

Mania only group

FIGURE 2.2–5. Brain imaging of stroke lesions for patients with bipolar disorder or mania only. Bipolar Group: schematic templates of computed tomographic scan slices showing largest cross-sectional area of lesion for manic-depressive patients. Lesions mainly involved the right head of the caudate and right thalamus. Mania only group: schematic templates of computed tomography scan slices showing largest cross-sectional area of lesion for mania only patients. Lesions mainly involved the orbitofrontal cortex and the right basitemporal cortex. (Reprinted with permission from Robinson

RG. *The Clinical Neuropsychiatry of Stroke: Cognitive, Behavioral and Emotional Disorders following Vascular Brain Injury*. 2nd ed. Cambridge University Press; 2006:302.)

Another study of 142 patients over 2 years found that early-onset (i.e., within 3 months after stroke) anxiety disorder without depression had a mean duration of 1.5 months while delayed onset (i.e., more than 6 months after stroke) GAD without depression had a mean duration of 3 months. In addition, the existence of anxiety also influenced the duration of depression. Patients with GAD and major depression had a mean duration of depression that was significantly longer than the duration of depression without a comorbid anxiety disorder.

The longitudinal study of poststroke anxiety by Luis Ayerbe and colleagues in 2014 reported that anxiety at 3 months poststroke was associated with lower quality of life at 10-year follow-up. In 2014, Maya Lambiase and coauthors found among 6,019 patients with stroke that higher scores of baseline anxiety were associated with an increased risk of incident stroke within the 16-year follow-up period (Fig. 2.2–7).

Apathy

A prospective cohort study found an association between baseline VCI and increasing apathy levels from baseline to the chronic stroke phase. Interestingly, no association was found between baseline VCI and the course of depressive symptoms.

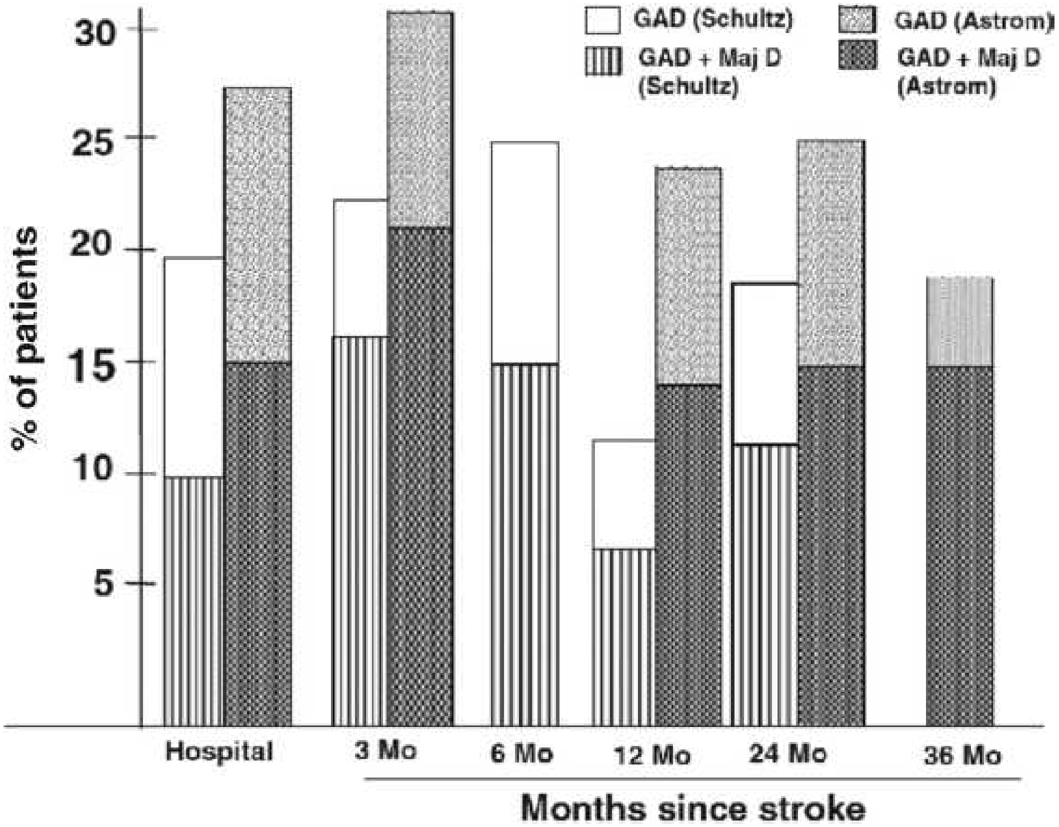

FIGURE 2.2–6. Longitudinal association of major depression with anxiety disorder after stroke. The frequency of generalized anxiety disorder (GAD) with and without major depression over the 3 years following acute stroke. Results obtained from Schultz et al. (1997) using the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) criteria were compared to the results of Astrom (1996) using the third revised edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R) criteria. Results indicate a slightly lower frequency of GAD using the DSM-IV criteria and emphasize the prominence of major depression in this population of patients with poststroke anxiety disorder. (Reprinted with permission from Robinson RG. *The Clinical Neuropsychiatry of Stroke: Cognitive, Behavioral and Emotional Disorders following Vascular Brain Injury*. 2nd ed. Cambridge University Press; 2006:336.)

Other Disorders

The course and prognosis of patients with psychosis, apathy, catastrophic reaction, PBA, and anosognosia have not been systematically studied.

TREATMENT

Mild and Major Vascular Neurocognitive Disorders

Some of the risk factors for stroke can be effectively treated, thus giving hope that cognitive impairment due to vascular disease may someday be a preventable condition. In the past decade, treatment with antiplatelet aggregate drugs has reduced the number of repeated ischemic vascular episodes in patients with TIAs. Acetylsalicylic acid (ASA) and other antiplatelet drugs have been shown to be effective in the secondary prevention of stroke. The United Kingdom-TIA Aspirin Trial, with 2,435 patients using two different dosages of ASA (300 mg daily and 1,200 mg daily), found that there were 21.7% and 25.1% reductions, based on dosage, in the risk of nonfatal strokes, myocardial infarction, or death compared with placebo treatment.

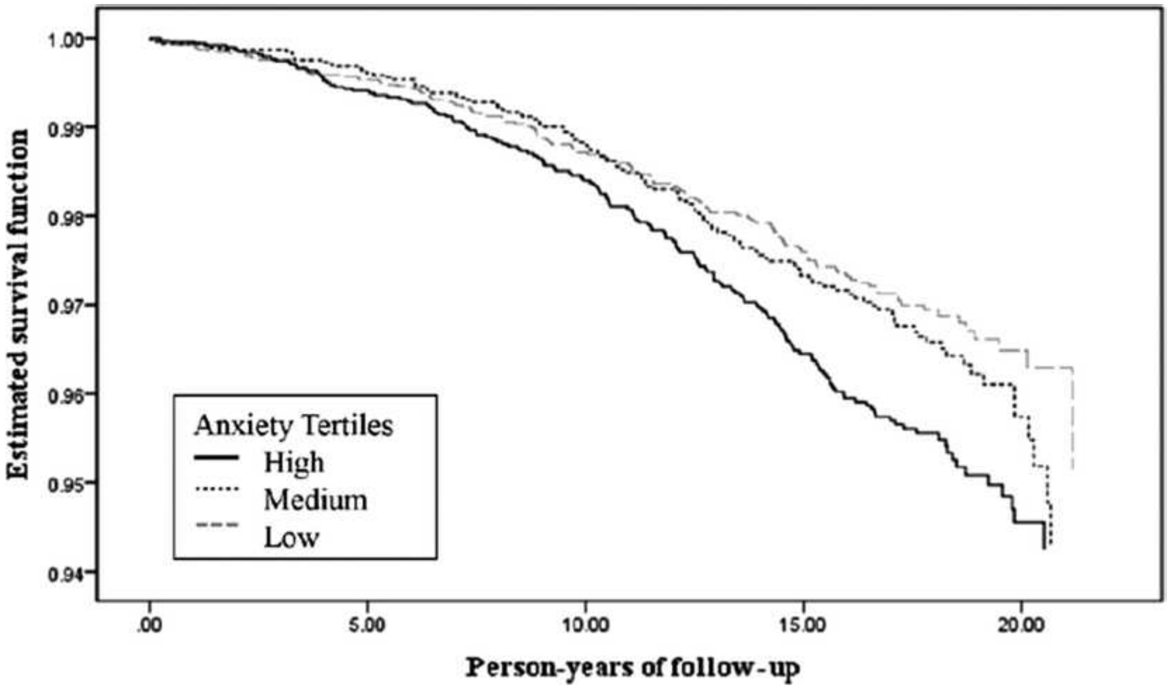

FIGURE 2.2–7. Estimated survival function based on the relationship between anxiety symptoms and incident stroke after adjustment for age and gender. Higher anxiety symptoms levels were associated with higher risk of stroke. (Reprinted with permission from Lambiase MJ, Kubzansky LD, Thurston RC. Prospective study of anxiety and incident stroke. *Stroke*. 2014;45(2):438. doi:10.1161/ STROKEAHA.113.003741)

Given that hypertension is the biggest risk factor for stroke, its impact on the incidence of vascular dementia has been an area of great interest. A 2011 meta-analysis by Sally Sharp and coscientists of 768 cases of vascular dementia and nearly 10,000 control cases found a significant association between hypertension and risk of vascular dementia with an odds ratio of 1.6 and 4.8, respectively, based on longitudinal studies and cross-sectional studies. A separate meta-analysis from 2011 based on 14 studies demonstrated that individuals prescribed antihypertensive medications had a reduced risk of vascular dementia (pooled relative risk of 0.67). A 2018 Cochrane systematic review that evaluated 38,742 participants found a pooled risk ratio of 0.88 for dementia among people taking a blood pressure-lowering agent compared with people taking a placebo. To date, it is unclear if this benefit is attributable to a certain class(es) of antihypertensives and/or from the treatment of hypertension alone.

For patients who are at particular risk for stroke-related cognitive impairment but are yet to show signs of dementia (i.e., those with a history of TIAs, stroke, or silent cerebral infarctions), carotid endarterectomy, when carotid stenosis ranges from 70% to 99%, may be indicated to minimize the risk of future strokes. After a cardioembolic stroke, anticoagulation is an effective treatment to reduce the risk of stroke recurrence. Stroke of cardioembolic origin is responsible for about 15% of all ischemic strokes, and this percentage is even higher among younger patients.

In patients with vascular dementia, ongoing treatment of vascular risk factors is generally recommended unless the risks clearly outweigh the benefits. For example, generally, antihypertensive treatment is recommended to minimize the risk of further stroke and thus further stroke-related cognitive decline. However, if antihypertensive treatment is leading to orthostatic hypotension, the risk of falls may outweigh the benefit of such treatment. Therapeutic measures aimed at reducing vascular risk factors include antihypertensives (e.g., β-blockers or calcium channel blockers), lipid-lowering agents such as statins, smoking cessation, and prevention or careful management of diabetes mellitus. Finally, treatment may also include cholinergic agonists (e.g., donepezil and galantamine) for cognition, antidepressant agents, antiplatelet aggregation agents, and statins for other comorbidities.

Depression

As reported in 1984, the first randomized, blinded, placebo-controlled study in PSD found that 14 patients treated with nortriptyline (50 to 100 mg/day) had a significantly greater reduction in Hamilton Depression Scores over 6 weeks than 20 patients given a placebo. The first double-blind, placebo-controlled trial to examine the efficacy of selective serotonin inhibitors (SSRIs) was reported by Grethe Anderson and colleagues in 1993. Among 33 poststroke patients given citalopram (20 mg/day under age 65; 10 mg/day over age 65), a significantly greater reduction in Hamilton Depression Scores over 6 weeks was observed in comparison to 33 similar patients given a placebo. Consisting of 16 trials and 1,655 subjects, a Cochrane Review from 2008 by Hackett and colleagues found an odds ratio of 0.47, favoring pharmacologic treatment over control for the treatment of PSD. Most recently, a 2020 meta-analysis that included 5,370 patients showed an RR of 0.78 for patients taking SSRIs. Thus, the efficacy of antidepressant treatment for PSD, particularly severe depression, has been established. However, a 2019 Cochrane review showed that SSRIs do not appear to aid functional recovery of neurologic impairment after stroke.

Cognitive-behavioral therapy (CBT) appears to have some positive effects on depression, but there is not enough evidence for it to be used as a monotherapy in PSD.

Perhaps the most promising recent advance in the treatment of PSD has been the study of preventative treatments (Fig. 2.2–8). To date, at least four randomized studies have been designed to evaluate the prevention of PSD using SSRI antidepressants. The results of these studies are shown in Figure 2.2–8. Although three of the four studies failed to show a significant effect on prevention, the similarities in the rate of prevention suggest that lack of power may have been the cause of the lack of an effect. The fourth study, conducted by Robinson and colleagues, involved 58 nondepressed acute stroke patients who were treated with escitalopram (5 mg/day over age 65; 10 mg/day under age 65) over 1 year. The incidence of PSD was 8.5% compared with 11.9% among 57 patients receiving problem-solving therapy (PST) and 22.4% among 58 patients receiving placebo (escitalopram vs. placebo, $P \leq .001$; PST vs. placebo, $P \leq .001$).

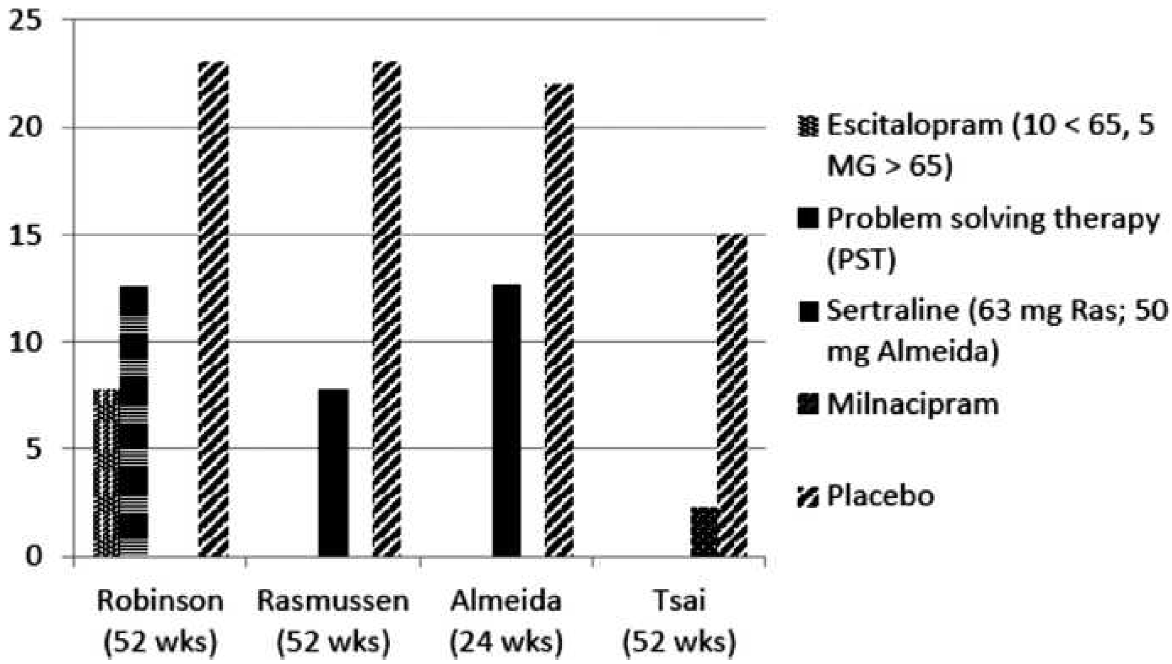

FIGURE 2.2–8. A summary of four randomized trials for the prevention of poststroke depression. For each study, the percent of patients who developed depression during the duration of the prevention trial is shown. The Robinson and Rasmussen studies showed remarkably similar findings. The difference between antidepressant and placebo were statistically significant for the Robinson and the Tsai studies. (The data were extracted from Robinson RG, Jorge RE, Moser DJ, et al. Escitalopram and problem-solving therapy for prevention of poststroke depression: a randomized controlled trial. *JAMA*. 2008;299(20):2391–2400; Rasmussen A, Lunde M, Poulsen DL, Sorensen K, Qvitzau S, Bech P. A double-blind, placebo-controlled study of sertraline in the prevention of depression in stroke patients. *Psychosomatics*. 2003;44(3):216–221; Almeida OP, Waterreus A, Hankey GJ. Preventing depression after stroke: results from a randomized placebo-controlled trial. *J Clin Psychiatry*. 2006;67(7):1104–1109; and Tsai CS, Wu CL, Chou SY, et al. Prevention of poststroke depression with milnacipran in patients with acute ischemic stroke: a double-blind randomized placebo-controlled trial. *Int Clin Psychopharmacol*. 2011;26 (5):263–267.)

Another promising recent advance includes the use of noninvasive brain stimulation like repetitive transcranial magnetic stimulation (rTMS) and transcranial direct current stimulation (tDCS) for the treatment of PSD. Data from rTMS trials are especially promising. However, larger randomized controlled studies need to be conducted once protocols are standardized before they can be considered viable treatment options.

Mania

Data on individual patients with single or recurrent episodes of mania suggest that they respond to lithium, although some fail to respond to either lithium or carbamazepine. No randomized controlled studies of the treatment of poststroke mania have been conducted, but one longitudinal study of 23 patients showed that most patients' manic episodes remit within a few months.

Anxiety Disorders

Data from three randomized, double-blind treatment studies were merged to evaluate nortriptyline ($n = 13$) versus placebo ($n = 14$) in the treatment of patients with comorbid GAD and depression following stroke. The severity of anxiety was measured using the Hamilton Rating Scale for Anxiety, and the severity of depression was measured using the Hamilton Rating Scale for Depression. Although there were no significant differences between the nortriptyline and placebo groups in demographic characteristics, stroke type, and neurologic findings, patients receiving nortriptyline treatment showed significantly more rapid improvement on the Hamilton anxiety scores compared with the Hamilton depression scores, suggesting that anxiety may respond more quickly to treatment than depression. Furthermore, the anxiety symptoms showed greater improvement in response to nortriptyline treatment compared with placebo. Katsunaka Mikami and colleagues performed a secondary analysis of patients in the previously described prevention trial of PSD conducted by Robinson's group. Of 52 patients without poststroke GAD treated for 1 year with escitalopram, only three cases developed GAD compared with 3 of 53 treated with PST and 9 of 49 patients given placebo ($P = .0005$). A 2017 Cochrane Review of treatment of poststroke anxiety without placebo control concluded that there is overall insufficient evidence to guide the treatment of anxiety after stroke.

Psychosis

There are no controlled treatment trials among patients with delusions or hallucinations following stroke. Anecdotal reports have suggested two basic approaches to treatment: utilization of antiseizure medication or antipsychotic medication. The use of antiseizure medications has its rationale in the frequent coexistence of seizures with psychotic disorders following stroke. Antipsychotic medications would need to be used with caution given their propensity to lower the seizure threshold and patients with strokes having a higher risk of seizures.

Apathy

Apathy following stroke has been treated with nortriptyline, bromocriptine, methylphenidate, amantadine, selegiline, and tacrine with some success. Nefiracetam is a novel GABA-A agonist, and in a double-blind trial for the treatment of comorbid apathy and depression by Robinson and colleagues, nefiracetam (900 mg/day) was significantly better than placebo in reducing scores on the apathy rating scale. Of note, nefiracetam has not been approved by the U.S. Food and Drug Administration (FDA) for any indications. Since apathy is a relatively common consequence of stroke and can have devastating effects on poststroke recovery, treatment trials are urgently needed.

Mikami and colleagues performed a secondary analysis of the previously mentioned depression trial to evaluate the prevention of apathy after stroke. Among 51 nonapathetic patients treated for a year with escitalopram (10 mg/day for those less than age 65; 5 mg/day for those greater than age 65), 2 developed apathy compared with 3 of 56 treated with PST and 7 of 47 given placebo ($P = .0031$). Thus, treatment trials with a primary outcome of apathy prevention are needed to determine if the administration of antidepressant medication such as escitalopram may prove helpful for the prevention of poststroke apathy.

Pseudobulbar Affect

A Cochrane Review of treatment trials of pathological affect in 2019 included seven trials totaling 239 people with stroke. Antidepressant treatment, including both tricyclic, SSRI and selective serotonin-norepinephrine reuptake inhibitors (SNRIs), significantly reduces the frequency and severity of PBA. For example, a double-blind, placebo-controlled cross-over study of 16 patients reported a greater than 50% reduction in the number of crying episodes while receiving citalopram. Eight patients responded to treatment within 24 hours of taking citalopram (20 mg/day); three patients responded within 3 days, and four patients took more than a week to respond. Further, an FDA-approved combination drug consisting of dextromethorphan and quinidine (Nuedexta) is now available with a specific indication for the treatment of

PBA. Thus, treatments that have shown to be effective in treating PBA are available for clinical use.

The clinical manifestations of PBA can be appreciated in this case history:

> A 64-year-old right-handed, retired, married woman with no prior history of a stroke suffered a thrombotic right-hemispheric stroke with hemiparesis but no sensory deficit. Beginning within a few days of the stroke, the patient had uncontrollable crying episodes that occurred 5 to 10 times per day, and each lasted about 1 to 2 minutes. In addition to the crying episodes, the patient had major depression with a Hamilton Depression Scale score of 19. She stated that she felt sad, but the crying was greatly in excess of her sadness at the time of her episodes and she had no sense of being able to control the crying. The pathological emotional expression was more troublesome to her than the depressed mood. Although she and her husband previously had an active social life, she stopped seeing friends and leaving the house for fear of being embarrassed socially by these crying episodes. Her PLACS score was 24, which was in the severe range. She improved greatly after a course of nortriptyline.

Other Disorders

Effective treatments have not been established for catastrophic reactions or anosognosia.

ACKNOWLEDGMENTS

We would like to acknowledge the significant contributions made by the authors of this chapter in the previous edition of this book, Sarah T. McCutcheon, M.D. and Robert G. Robinson, M.D.

FURTHER READINGS

Ayerbe L, Ayis S, Wolfe CD, Rudd AG. Natural history, predictors and outcomes of depression after stroke: systematic review and meta-analysis. *Br J Psychiatry*. 2013;202(1):14–21.

Bucur M, Papagno C. A systematic review of noninvasive brain stimulation for post-stroke depression. *J Affect Disord*. 2018;238:69–78.

Clancy U, Gilmartin D, Jochems ACC, Knox L, Doubal FN, Wardlaw JM. Neuropsychiatric symptoms associated with cerebral small vessel disease: a systematic review and meta-analysis. *Lancet Psychiatry*. 2021;8(3):225–236.

Denny-Brown D, Meyer JS, Horenstein S. The significance of perceptual rivalry resulting from parietal lesions. *Brain*. 1952;75:433–471.

Douven E, Köhler S, Rodriguez MMF, Staals J, Verhey FRJ, Aalten P. Imaging markers of post-stroke depression and apathy: a systematic review and meta-analysis. *Neuropsychol Rev*. 2017;27(3):202–219.

Frey J, Najib U, Lilly C, Adcock A. Novel TMS for stroke and depression (NoTSAD): accelerated repetitive transcranial magnetic stimulation as a safe and effective treatment for post-stroke depression. *Front Neurol*. 2020;11:788.

Goldstein K. *The Organism: A Holistic Approach to Biology Derived from Pathological Data in Man*. American Books; 1939.

Hackett ML, Kohler S, O'Brien JT, Mead GE. Neuropsychiatric outcomes of stroke. *The Lancet Neurology*. 2014;13(5):525–534.

Knapp P, Campbell Burton CA, Holmes J, et al. Interventions for treating anxiety after stroke. *Cochrane Database Syst Rev*. 2017;5(5):CD008860.

Kohler S, Hackett ML, O'Brien JT, Mead GE. Neuropsychiatric outcomes after stroke–authors' reply. *Lancet Neurol*. 2014;13(12):1168–1169.

Kraepelin E. *Manic Depressive Insanity and Paranoia*. E & S Livingstone; 1921.

Legg LA, Tilney R, Hsieh C-F, et al. Selective serotonin reuptake inhibitors (SSRIs) for stroke recovery. *Cochrane Database Syst Rev*. 2019;2019(11):CD009286.

Levada OA, Troyan AS. Poststroke depression biomarkers: a narrative review. *Front Neurol*. 2018;9:577.

Levine DA, Galecki AT, Langa KM, et al. Trajectory of cognitive decline after incident stroke. *JAMA*. 2015;314(1):41–51.

Li W, Ling S, Yang Y, Hu Z, Davies H, Fang M. Systematic hypothesis for post-stroke depression caused inflammation and neurotransmission and resultant on possible treatments. *Neuro Endocrinol Lett*. 2014;35(2):104–109.

Lincoln NB, Brinkmann N, Cunningham S, et al. Anxiety and depression after stroke: a 5 year follow-up. *Disabil Rehabil*. 2013;35(2):140–145.

Meyer A. The anatomical facts and clinical varieties of traumatic insanity. *Am J Insanity*. 1904;60:373.

Pompili M, Venturini P, Lamis DA, et al. Suicide in stroke survivors: epidemiology and prevention. *Drugs Aging*. 2015;32(1):21–29.

Rafsten L, Danielsson A, Sunnerhagen KS. Anxiety after stroke: a systematic review and meta-analysis. *J Rehabil Med*. 2018;50(9):769–778.

Robinson RG, Jorge RE. Post-stroke depression: a review. *Am J Psychiatry*. 2016;173(3):221–231.

Robinson RG, Jorge RE, Long J. Prevention of post-stroke mortality using problem solving therapy or escitalopram. *Am J Geriat Psychiatry*. 2017;25:512–519.

Rochat L, Van der Linden M, Renaud O, et al. Poor reward sensitivity and apathy after stroke: implication of basal ganglia. *Neurology*. 2013;81(19):1674–1680.

Stangeland H, Orgeta V, Bell V. Poststroke psychosis: a systematic review. *J Neurol Neurosurg Psychiatry*. 2018;89(8):879–885.

Starkstein SE, Hayhow BD. Treatment of post-stroke depression. *Curr Treat Options Neurol*. 2019;21(7):31.

Tang WK, Lau CG, Mok V, Ungvari GS, Wong K-S. Impact of anxiety on health-related quality of life after stroke: a cross-sectional study. *Arch Phys Med Rehabil*. 2013;94(12):2535–2541.

Virani SS, Alonso A, Aparicio HJ, et al. Heart Disease and Stroke Statistics-2021 Update: a report from the American Heart Association. *Circulation*. 2021;143(8):e254–e743.

Wang S-B, Wang Y-Y, Zhang Q-E, et al. Cognitive behavioral therapy for post-stroke depression: a meta-analysis. *J Affect Disord*. 2018;235:589–596.

Wiesmann M, Kiliaan AJ, Claassen JA. Vascular aspects of cognitive impairment and dementia. *J Cereb Blood Flow Metab*. 2013;33(11):1696–1706.

Zonneveld TP, Richard E, Vergouwen MD, et al. Blood pressure-lowering treatment for preventing recurrent stroke, major vascular events, and dementia in patients with a history of stroke or transient ischaemic attack. *Cochrane Database Syst Rev*. 2018;7(7):CD007858.

2.3 The Neuropsychiatry of Brain Tumors

ALASDAIR G. ROONEY, MB.CHB., M.D., PH.D.

Neuropsychiatric disorders are common in brain tumor patients and can present particular diagnostic and treatment challenges. Identifying an affective psychosis in someone on high-dose corticosteroids may be straightforward, but deciding when natural grief crosses the line into pathologic depression may not; neurocognitive impairment can be easily missed unless directly examined and personality change may only come to light via collateral history. Meanwhile, the treatment of common psychiatric disorders mostly lacks a brain tumor-specific evidence base, and management of more specialized disorders such as neurocognitive impairment and personality or behavioral change can be difficult. There is a tremendous clinical need for psychiatric input to neuro-oncology teams. In this chapter, we will discuss some clinical aspects and themes on the neuropsychiatry of brain tumors.

PSYCHIATRIC CLASSIFICATION

The diagnostic validity of psychiatric classification remains unclear when applied to brain tumor patients. In these patients—as with many other neurologic disorders—the balance and causes of neuropsychiatric symptoms can differ in certain ways from psychiatric patients who are otherwise physically well. When the DSM system was developed in the 1970s, there was (and to a large extent still is) an absence of careful studies of the phenomenology of psychiatric syndromes as they present in brain tumor patients. The absence of disease-specific studies permits a range of opinions about the nature of psychiatric symptoms in these patients. Two arguments will be made briefly here to illustrate this point, hopefully without over-emphasizing its importance.

In the DSM and the ICD (WHO) classification systems, the core symptoms of listed psychiatric syndromes relate to emotions or inner psychic realities (fear, sadness, anger, pleasure, paranoia, delusions, hallucinations, and so on). We infer that psychopathology must arise in some way from the brain. In that context, the presence of a metabolically active tumor invading, altering, distorting, and destroying brain tissue presents a unique theoretical challenge. At the least, it must be unclear whether the pathophysiologic causes of psychiatric symptoms are similar in those with a brain tumor compared to those without. If they are not, then the founding assumption of an operationalized classification system—that covarying symptoms reflect a common underlying pathophysiology—is challenged. First, therefore, one should keep at least an open mind about the validity and reliability of psychiatric syndromes in brain tumor patients. These patients have—confoundingly—a tumor growing in the seat of the psyche.

Second, much of the context of symptomatology is now stripped away. The jettisoning in DSM-5 of the “bereavement exclusion” criterion for major depressive disorder (MDD) is one example of this strategy. Minimizing context has several advantages, including better reproducibility and, in some ways, less theoretical speculation. The problem, however—as any clinician familiar with brain tumor patients will agree—is that it is quite difficult to think of a neurologic patient group for whom context is more relevant. It is hard to think of a good reason to ignore the context of a catastrophic existential threat arising out of the blue, where toxic treatments have side effects identical to key psychiatric symptoms, which cause multiple and progressive losses in multiple domains of personhood and life and is associated with cognitive impairment, epilepsy, personality change, and in many cases, death. This is a challenging test for the philosophy of minimizing or abandoning context. Few other psychiatric patient groups face quite the same sudden, severe, persisting, and progressive mix of stressors.

Do these theoretical objections matter? The only way to find out is to conduct prospective studies of the reliability and validity of psychiatric diagnoses in large cohorts of brain tumor patients of all ages and at all stages of disease. As a community, unfortunately, we are nowhere near this ideal. Brain tumor-specific studies of psychiatric symptom validity and reliability remain rare. Brain tumor-specific randomized controlled trials (RCTs) of psychiatric interventions are similarly scarce.

In general, the day-to-day psychiatric management of brain tumor patients is therefore pragmatic more than it is dogmatic. It is an open field with great unmet need and an even greater challenge to psychiatrists: to systematically study the abnormal emotional, behavioral, and neurocognitive consequences of brain tumors.

BRAIN TUMORS

Epidemiology

Primary brain and central nervous system (CNS) tumors are reportable diseases in many countries, including the United States. The word “primary” in this context refers to a tumor that arises de novo in the brain and is not the result of a “secondary” spread from a site outside the CNS. The Central Brain Tumor Registry of the United States (CBTRUS) is a nonprofit organization that collaborates with state registries to provide an annual statistical analysis. It is an authoritative resource for epidemiologic statistics on new primary brain tumor diagnoses, although limited in scope to the United States.

The most recent publicly available CBTRUS data show that the average annual incidence of all adult primary brain and other CNS tumors is 23.8/100,000. In the United States alone, this translates to nearly 84,000 new adult cases of brain and other CNS tumors expected to be diagnosed in 2020, with over 16,000 deaths per year. The median age at diagnosis for all primary brain and CNS tumors is 59 years. In young adults (aged 18 to 40), the incidence is most similar to that of testicular or thyroid cancer, and in older adults (aged 40+) to that of bladder cancer or melanoma. In children and adolescents, meanwhile, primary brain and

CNS tumors are the single most common cancer and the most common cause of cancer death. Primary brain tumors overall are slightly more common in females, and incidence rates for certain tumors vary according to ethnicity (see CBTRUS annual reports).

Brain tumors are a heterogeneous group of diseases with a wide range of etiologies, ages of onset, anatomical locations, clinical courses, responses to treatment, and prognosis. The simplest division is into nonmalignant and malignant tumors. Most nonmalignant brain tumors arise in the meninges, pituitary gland, or cranial nerve sheaths. Psychiatrists in routine practice are more likely to encounter nonmalignant tumors in the course of their work because they are more common. Many such tumors picked up on routine brain scans will be incidental findings such as a small calcified meningioma. Most malignant brain tumors, meanwhile, are gliomas—primary cancer arising from neuroepithelial tissue. Malignant tumors may be seen more often by psychiatrists linked to general hospitals or neuro-oncology services because, unlike the typical nonmalignant tumor, which is cured by surgical excision and does not require long-term follow-up; glioma is generally invasive and incurable. Patients with glioma often become markedly physically and cognitively impaired and require ongoing clinical follow-up.

The gliomas are categorized by the World Health Organization (WHO) according to the predominant tissue histologic profile (astrocytic, oligodendrocytic, etc.) and by malignancy. Malignancy is judged according to histologic features known to confer a poorer prognosis, such as the extent of proliferating cancer cells in a biopsy or the presence of neovascularization and tumor necrosis indicative of very rapid growth. The WHO grades glioma on a level from I to IV, where IV is the most malignant type, also known as glioblastoma multiforme (GBM). GBM, which has the worst prognosis, is sadly also the most common malignant primary brain tumor. By convention, WHO grades I and II are called "low-grade" while WHO grades III and IV are "high-grade." Most low-grade tumors eventually progress and transform into a higher grade; they are therefore not truly "benign." In adults, most gliomas are supratentorial and situated in the frontal, temporal, or parietal lobes. In children, by contrast, glioma is usually infratentorial (Fig. 2.3–1).

The incidence of most histologic classes of primary brain tumors rises continuously with advancing age. The main exceptions are pilocytic astrocytoma, germ cell tumors, and embryonal tumors such as medulloblastoma, which are more common in younger children and rare in adults. The most common tumor histologic cell of origin across different ages is shown in Table 2.3–1.

In contrast to primary brain tumors, the epidemiology of metastatic brain tumors is less well understood. There is no requirement on national cancer registries to document their incidence. There is a general assumption that metastatic or "secondary" tumors outnumber primary tumors. Metastatic brain tumors occur in a sizeable minority of cancer patients, are usually multiple, and are especially more likely to occur when the primary site is lung, breast, skin, kidney, or colon. Because their presentation and management are therefore scattered across multiple clinical pathways, and because the prognosis of metastatic brain tumors is particularly poor, studies of their psychiatric effects are difficult. Here, we will focus only on primary brain tumors and mainly on glioma.

Management of Brain Tumors

In many countries, treatment decisions are made by a multidisciplinary neuro-oncology team. In the United Kingdom, the minimum core personnel requirement for this team is stated in NHS clinical guidance and includes a team coordinator, neurosurgeon, neuroradiologist, neuropathologist, neurologist, oncologist, clinical nurse specialist, palliative care specialist, and neuropsychologist. Psychiatry or neuropsychiatry input is optional. One consequence of this arrangement is that, at least in the United Kingdom, psychiatric involvement is sporadic and often restricted to tertiary neuro-oncology centers with well-developed links to liaison or neuropsychiatry services. U.K. neuro-oncology teams have said clearly, however, that more input from psychiatry would be desirable.

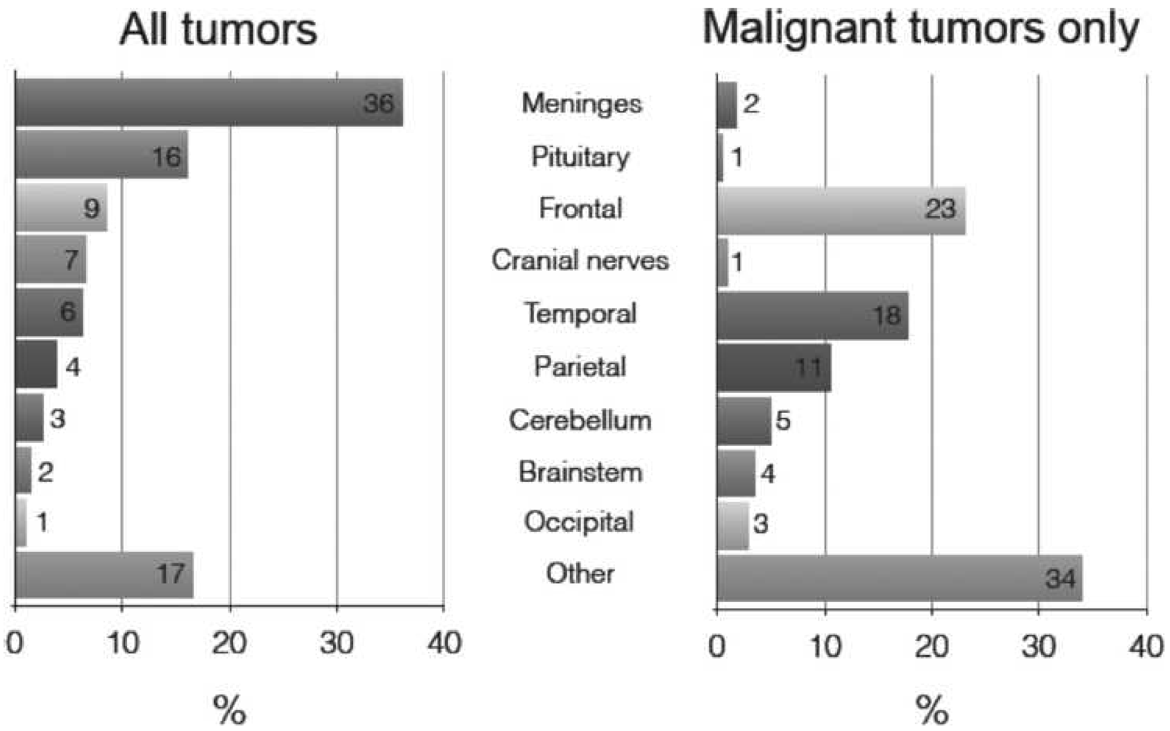

FIGURE 2.3–1. Most brain tumors occur in the meninges or pituitary gland, but malignant tumors are more common in the parenchyma of the brain. (Source data: CBTRUS statistical report 2015.)

Table 2.3–1.
The Main Histologic Cell of Origin of Primary Brain Tumors Differs at Different Ages

Age	Most Common Histology	Second Most Common Histology	Third Most Common Histology
0–4	Malignant glioma NOS	Pilocytic astrocytoma	Medulloblastoma
5–9	Malignant glioma NOS	Pilocytic astrocytoma	Medulloblastoma
10–14	Malignant glioma NOS	Pilocytic astrocytoma	Pituitary tumor
15–19	Pituitary tumor	Malignant glioma NOS	Pilocytic astrocytoma
15–39	Pituitary tumor	Meningioma	Nerve sheath tumor
20–34	Pituitary tumor	Meningioma	Nerve sheath tumor
35–44	Meningioma	Pituitary tumor	Nerve sheath tumor
45–54	Meningioma	Pituitary tumor	Glioblastoma
55–64	Meningioma	Glioblastoma	Pituitary tumor
65–74	Meningioma	Glioblastoma	Pituitary tumor
75–84	Meningioma	Glioblastoma	Pituitary tumor
85+	Meningioma	Glioblastoma	Pituitary tumor
Total	*Meningioma*	*Glioblastoma*	*Pituitary tumor*

Data from Ostrom QT, Patil N, Cioffi G, et al. CBTRUS Statistical Report: Primary brain and other Central Nervous System tumors diagnosed in the United States in 2013–2017. *Neuro Oncol*. 2020; 22(12 Suppl 2):iv1–iv96. doi: 10.1093/neuonc/noaa200.

A postal questionnaire survey of U.K. neuro-oncology multidisciplinary teams (MDTs) was conducted to determine: the reported proportion of new patients referred for psychiatric assessment in the previous year, whether neuropsychiatry services were available locally, and whether the team perceived an unmet need for neuropsychiatry. Of 39 neuro-oncology MDTs, 29 (74%) responded, treating a total estimated 3,500 new patients/year. A median of 1% (mean 4%, range 0% to 25%) of these new patients were reported as having been referred to psychiatry in the previous year. Eight MDTs had made no psychiatric referrals. Most often, significant emotional difficulties were either first identified by the clinical nurse specialist or brought to clinical attention by the patient's relatives. A minority of MDTs had access to neuropsychiatry (12/29; 41%), and 24/29 (82%) perceived an unmet need for neuropsychiatry. In summary, therefore, nurse specialists were central to referral procedures in most U.K. centers. Most neuro-oncology MDTs had no regular access to neuropsychiatry assessment. There was a clear national and clinical demand for more neuropsychiatry input to neuro-oncology services. *Data collected by Cliodhna Carroll (University of Cambridge, UK) and the author, with supervision from Dr. Howard Ring (Consultant neuropsychiatrist, University of Cambridge, UK) and Dr. Robin Grant (Consultant neurologist, NHS Lothian, UK).*

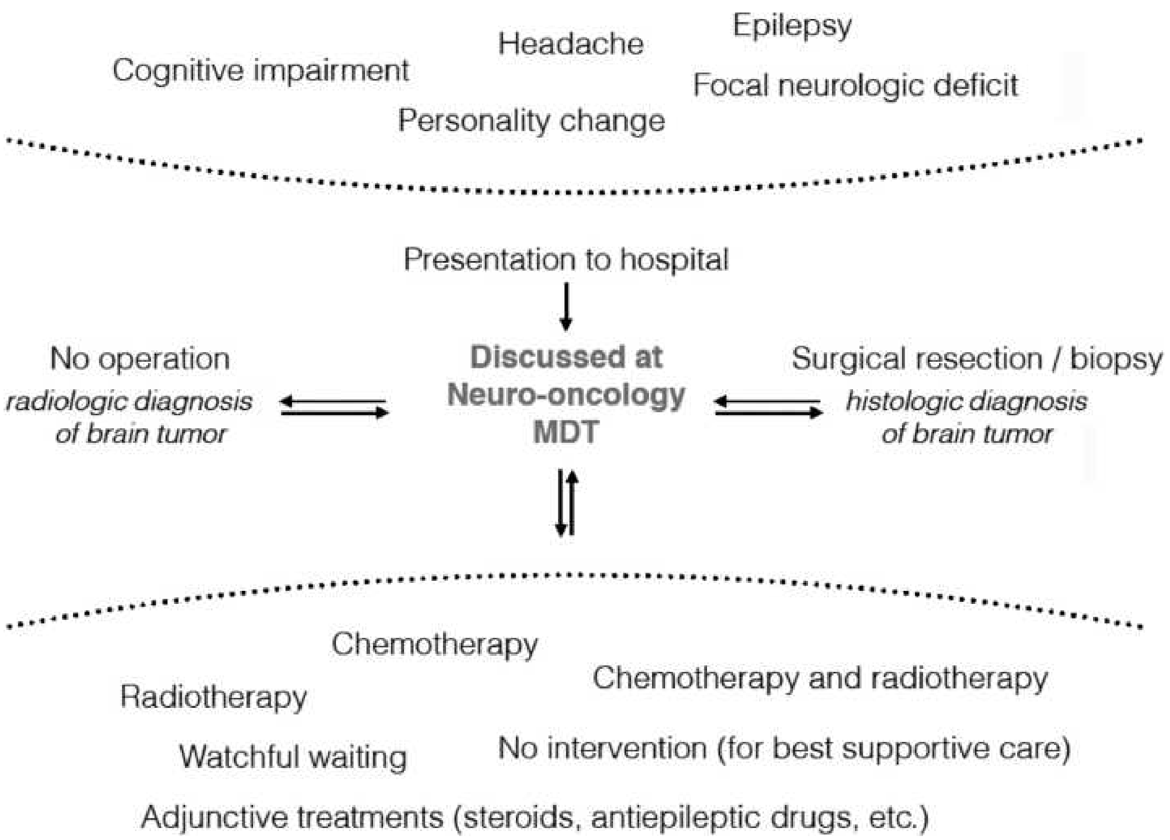
Headache
Epilepsy
Cognitive impairment
Focal neurologic deficit
Personality change
Presentation to hospital
No operation
radiologic diagnosis of brain tumor
Discussed at Neuro-oncology MDT
Surgical resection / biopsy
histologic diagnosis of brain tumor
Chemotherapy
Chemotherapy and radiotherapy
Radiotherapy
Watchful waiting
No intervention (for best supportive care)
Adjunctive treatments (steroids, antiepileptic drugs, etc.)

FIGURE 2.3–2. The neuro-oncology multidisciplinary team (MDT) is key to planning the care and treatment of brain tumor patients.

Allowing for some differences in international healthcare models, the main treatment decisions are shown diagrammatically in Figure 2.3–2. Newly diagnosed patients can be treated with varying combinations of established treatments: surgery, radiotherapy, chemotherapy, watchful waiting, or supportive (palliative) care. The search for curative treatments continues and the range of "experimental" additional therapies is wide. Currently, the most common active treatment for GBM—the most malignant form of glioma—is a combination of maximal surgical resection, radiotherapy, and chemotherapy. High-dose corticosteroids may be started with the intention of reducing cerebral edema and alleviating symptoms caused by mass effect and raised intracerebral pressure. Antiseizure medications are often also prescribed, though if the patient has not presented with seizures, their value in prophylaxis is unclear and the subject of ongoing research. Childhood brain tumors are often managed with treatments focused on the posterior fossa or brainstem, again comprising surgery, radiotherapy, and/or chemotherapy.

Patients with a simple meningioma may be cured by surgery and discharged from follow-up. Patients with glioma, meanwhile, will receive follow-up during which they are regularly reviewed by the treating neuro-oncology team. Many patients report that the transition from active treatment to follow-up is an unexpectedly difficult period. The energy of the primary treatment phase is over, and they must readjust to a life that may now be very different. Physical disability, epilepsy, cognitive impairment, emotional difficulties, and behavioral changes are so common as to be normal. Many patients—having had brain surgery and possibly epilepsy—will be forbidden to drive by national authorities and unable to work, placing sudden and continuing strain on previously stable family routines. Although some patients—especially younger patients with low-grade tumors and well-controlled epilepsy—can regain much of their independence, the reality for many is partial dependence on carers and an uncertain future.

Prognosis

The prognosis of a brain tumor depends on the type. Nonmalignant tumors such as meningioma and some pituitary tumors can be cured surgically with relatively little in the way of long-term deficit. Malignant tumors such as glioma present an impossible challenge to surgical cure because, by the time of clinical presentation, tumor cells are microscopically dispersed throughout much of the brain (not just in the macroscopic area seen on an MRI or CT scan).

Low-grade gliomas (WHO grades I–II) typically follow an indolent course before eventually transforming into a higher grade after a highly variable period ranging from a few years to many years. Higher grade tumors (WHO grades III–IV) are more rapidly progressive and fatal. The median survival of patients with GBM in the field-leading RCT was around 14 months: around one-quarter of these patients survived to 2 years postdiagnosis and only 10% were alive after 5 years. These numbers apply at the population level: it is impossible to give a precise duration of survival for an individual patient at the point of diagnosis, although younger patients and those with better preserved physical and cognitive functioning tend to survive longer.

It is clear that tumor molecular characteristics play an additional prognostic role, with some mutations conferring a better prognosis than others. A seminal example of this feature was the discovery that tumors in which the O^6-methylguanine-DNA methyltransferase (MGMT) promoter region is methylated (meaning their DNA-repair mechanisms are switched off) are more sensitive to the DNA-disrupting, tumor cell-killing effects of radiation and chemotherapy. MGMT testing is now routine for new glioma diagnoses. Unsurprisingly, molecular characterization in neuro-oncology is extremely complex and well beyond the scope of this chapter; interested readers are, however, directed to further reading (see Further Readings).

Thus, the tempo of the clinical course of a malignant primary brain tumor is characterized by sudden presentation, a flurry of multidisciplinary management amid a period of intensive assessment and primary treatment lasting for several months, and then a variable period of follow-up in which patients must attempt to rediscover a normality of life. The profile of disability in the patient is initially moderate and often relieved for a while by primary treatment, before a progressive but unpredictable series of impairments that sadly but typically end in the death of the patient after months or a modest number of years. It is in this context that we will consider the neuropsychiatry of brain tumors.

BRAIN TUMORS IN PSYCHIATRIC PATIENTS

Does your psychiatric patient have a brain tumor? The idea that psychiatric symptoms are in some way indicative of an underlying brain tumor is certainly reinforced by popular culture. In the movie "Phenomenon," the main character experiences unusual psychic changes, which are ultimately attributed to a developing astrocytoma. The movies "The Green Mile" and "Saw" feature characters whose motivations and actions are violently twisted by brain tumors. A character in the TV series "Grey's Anatomy" experiences visual hallucinations secondary to brain metastasis. Real-life examples, meanwhile, include George Gershwin, the musical composer and pianist, whose olfactory hallucinations and erratic behavior were initially attributed to hysteria before he collapsed and died from a brain tumor aged 37. Speculative associations between a brain tumor and some form of psychopathology are drawn by hundreds of single case studies in the academic literature.

Indeed, large population-based epidemiologic datasets do suggest that among all occult developing cancers, the risk for psychiatric hospitalization in the period before a cancer diagnosis is highest for brain tumors. The effect is not marked compared to other cancers, but it does seem that some patients will experience symptoms and end up in a psychiatric hospital because, in fact, they have a brain tumor. Many such patients (and their doctors) may initially attribute psychiatric symptoms to various environmental or situational stresses. The problem for the general psychiatrist is that these “secret tumor” patients are a vanishing minority: almost certainly under 1% of their caseload over an entire professional career. How should psychiatrists attempt to identify the small number of patients whose symptoms are due to a brain tumor? Can investigations be targeted, or is indiscriminate screening required?

The importance of targeted investigation is illustrated by autopsy data from chronic–psychiatric patients dying in mental hospitals from other causes. These data show unsuspected and undiagnosed brain tumors in roughly 3% of patients examined. More recent studies converge on a similar conclusion: the frequency of clinically unsuspected space-occupying lesions on brain CT or MRI scanning of inpatient psychiatric populations is in the order of 3% to 4%. Similarly, the frequency of incidental abnormalities observed on brain scans in the general nonpsychiatric population is approximately 3%. The likely conclusion is that many abnormalities encountered in routine scanning of psychiatric patients will be incidental (functionally silent). Because at least some, if not most, of these tumors will never progress to cause any kind of clinical problem, they do not require treatment, and indiscriminate scanning may increase the risk of unnecessary intervention. So how should investigations be targeted?

We identified adult psychiatric inpatients referred consecutively over a 4-year period for a head CT on suspicion of organic pathology. Patients were excluded if they were diagnosed with a brain tumor before the psychiatric admission. The primary aim was to determine the prevalence of brain tumors. A secondary aim was to describe presenting psychiatric symptoms and neurologic signs. Consultant radiologist CT reports were obtained from computerized records and reviewed. Psychiatric notes were traced for patients noted to have suspected or possible primary brain tumor on CT report. To simplify the analysis, presenting psychiatric symptoms were grouped heuristically into the following broad categories: *low mood* (anxiety, depression, or psychomotor retardation); *cognitive dysfunction* (confusion or cognitive impairment); *personality/behavior change*; and *psychosis* (delusions, mania, or paranoia).

A total of 328 psychiatric inpatients (mean age 65; range 37 to 86; 43.3% male) were identified as referred for a head CT. Space occupying lesions were identified from the CT report in 19 cases, of which 12/328 (3.7%) were newly discovered tumors. The most common radiologic diagnosis was meningioma (5/12). Eight tumors involved the cerebral hemispheres (five right-sided and three left-sided). The rest were located in the midline or involved the ventricular system.

Most patients with newly discovered tumors had symptoms from more than one psychiatric category. The most common were *low mood* and *cognitive dysfunction*, each seen in 9/12 patients. *Personality/behavior change* was reported in 6/12 and *psychosis* in 5/12. In five patients, no neurologic signs were documented in the psychiatric notes. In the remainder, brisk reflexes (five patients) and speech deficits (four patients) were the most common signs. Papilledema was seen in one patient and was associated with a ventricular tumor. In summary, therefore, few (in this study, roughly 1 in 27) head CTs for suspected brain pathology in adult psychiatric inpatients revealed a new case of brain tumor. *Data were collected by Anita Vas-Falcao (year 4 medical student, University of Edinburgh) with supervision from the author and Dr. Robin Grant, Consultant Neurologist.*

“Red flag” Symptoms and Signs

Brain tumors can present with nonspecific psychiatric and behavioral changes. It is wise to maintain a low threshold for investigating the possibility of a brain tumor in any patient with new-onset or newly different psychiatric symptoms. A high index of suspicion is particularly warranted in certain situations: if the patient is a child; if the patient has a negative past or/and family history of psychiatric illness; if the psychiatric symptoms have atypical features; if they are associated with additional unexplained symptoms; or if there is evidence on examination of neurologic or neurocognitive dysfunction. The psychiatrist should take a careful history from the patient and, ideally, an informant, asking for physical symptoms associated with brain tumors and do a neurologic examination. As a trainee, the author clinically predicted a right-sided parietal lobe tumor in a middle-aged lady who was brought to the hospital wandering and confused, wearing a jumper back-to-front, with dysgraphesthesia of the left palm.

Look particularly for a history of recent mental state change in association with *unexplained*:

1. Headache. The classical headache of raised intracranial pressure (ICP) is worse on wakening secondary to cerebral vasodilatation during sleep. It may transiently worsen with coughing, straining or sneezing. Not all brain tumor-associated headaches fit this pattern (nor do all tumors cause headaches). As in the general population, too, the vast majority of headaches in psychiatric patients will not be caused by a brain tumor. Other common explanations may include opiate medication overuse, migraine, somatization, tension headache, and selective serotonin reuptake inhibitor (SSRI) discontinuation syndrome. Nevertheless, headache with new-onset psychiatric symptoms comprise the neurologic combination of “headache +” (“headache plus…” unusual symptoms), which is a useful rule of thumb to trigger suspicion of an organic cause.
2. Nausea or vomiting. Presumed secondary to raised ICP compressing vagal centers in the medulla, nausea or vomiting is a classic “red flag” which is, however absent in many brain tumor patients.
3. Cognitive impairment. Some degree of cognitive impairment is detectable in most brain tumor patients. Impairment may range in severity from subtle to profound. The more subtle deficits may only be picked up on detailed neuropsychological testing or a careful history.

Slower processing speed is very common. Patients may appear sluggish and vague. Since poor concentration and executive function —features of many psychiatric conditions—may mimic and overlap with slower processing speed, cognitively impaired psychiatric patients need a full neurologic history and examination.

4. Epilepsy. Most brain tumor patients will have seizures at some point in the disease trajectory, but epilepsy is more commonly a presenting symptom for patients with low-grade (WHO grades I–II) glioma. Common alternative explanations for seizures in psychiatric patients may include alcohol or benzodiazepine withdrawal, a lower seizure threshold secondary to antipsychotic use, hyponatremia, or functional nonepileptic attacks.
5. Weakness or clumsiness. Motor weakness or clumsiness due to a brain tumor is often subacute (onset of duration over a few weeks) and gradually progressive. Again a careful history is valuable. Weakness may be described as seemingly mild difficulties in dressing, handling the TV remote control, or taking lids off jars. A history of unexplained falls should prompt examination for foot drop or ataxia. Apraxia (such as difficulty dressing correctly or manipulating gadgets) may indicate a parietal lobe lesion, although it is not specific for this location.
6. Visual or sensory disturbance. The patient may not be fully aware of any deficit. Hemisensory or visual neglect may manifest in the history as the patient bumping into doorways on the same side or walking out into traffic, or in the examination by the patient ignoring people on one side of the bed (sometimes a very subtle sign), asymmetry of personal grooming, altered visual fields, or visual or sensory inattention.
7. Unusual (“noncanonical”) psychiatric symptoms. Case reports provide anecdotal evidence only, but many odd symptoms that “don’t quite fit” with the classical psychiatric syndromes have been reported in association with an occult brain tumor. These symptoms have included visual or musical hallucinations, palinacousis, palinopsia or similar “simple” perceptual phenomena, with a scattering of rare and exotic symptoms (e.g., polyopic heautoscopy). The unexplained development of personality abnormalities in adulthood is another indication to order neuroimaging.
8. Objective neurologic deficit. A comprehensive neurologic examination can be difficult to perform in patients who are acutely psychiatrically unwell, but is obviously a key part of the full assessment. Readers are directed to further reading on a full neurologic examination. More

obvious abnormalities may be picked up in a short 1- to 2-minute screening examination by: testing for visual and sensory inattention; examining eye movements (the cranial nerves supplying the eye have long intracranial courses and may be particularly affected by raised ICP); looking for pronator drift (a sensitive test for pyramidal weakness in the upper limbs); assessing muscle tone and reflexes; assessing gait including a heel-toe test; and screening for dysmetria with the finger-nose test. These examination components are a basic skim: a full examination (including fundoscopy) should be done if the patient allows. In isolation, neurologic "soft signs" such as sensory abnormalities or the subtle persistence of developmental reflexes may lack specificity. If there is doubt a neurologist or neuropsychiatrist can give a second opinion.

9. Papilledema. A difficult neurologic sign to elicit in acutely psychiatrically ill patients, papilledema indicates raised ICP. The presence of papilledema should always trigger further assessment.
10. Endocrine abnormality. Pituitary tumors can cause a variety of endocrine disturbances. Systemic physical or blood biochemical signs of acromegaly, Cushing disease, abnormal lactation, or altered menstruation would raise suspicion. The textbook neurologic deficit of pituitary tumor is a bitemporal hemianopia secondary to midline compression of the optic chasm. These patients may, also, walk out into traffic.

This list is not exhaustive and neurologic examination may be normal in the presence of a brain tumor. A normal examination may especially occur with slow-growing tumors involving a relatively neurologically silent brain region such as the corpus callosum, nondominant temporal and parietal lobes, and the frontal pole.

Investigations

Brain CT or MRI scanning is indicated in psychiatric patients with focal neurologic findings and/or symptoms such as those listed above. When brain tumor is suspected, lumbar puncture should be avoided until imaging confirms that there is no obstruction to the free flow of cerebrospinal fluid (CSF). There is a risk of causing fatal coning by lumbar puncture if CSF cannot circulate freely out of the ventricles in the context of raised intracerebral pressure. Lumbar puncture may however be helpful when cytology studies are required in the assessment of established tumors such as leukemias, lymphomas, and meningeal metastasis.

CT scanning will pick up about 90% of cerebral tumors and is enhanced further by the use of intravenous (IV) contrast. CT scan may be especially useful in the evaluation of tumors which are calcified, eroding bony intracranial structures, or causing shifts in middle cerebral structures. CT scans can detect abnormalities involving the ventricular system such as obstructive hydrocephalus, and gauge the extent of cerebral edema. Certain types of tumors remain difficult to identify with CT scans. These include: lesions smaller than 0.5 cm in diameter; tumors occurring next to bony structures (such as acoustic neuromas, pituitary tumors, some meningiomas, and tumors deep in the posterior fossa); low-grade astrocytomas; and meningeal carcinomatosis. CT scanning also involves a degree of radiation exposure.

MRI outperforms CT scanning in those latter respects. Owing to greater image resolution MRI is also better at identifying the architecture of brain tumors (whether they are solid or cystic, or both), and in more precisely defining the relationship of a tumor to nearby vascular structures. However MRI scanning can be claustrophobia-inducing and takes longer: some psychiatric patients may better tolerate CT, and a CT scan costs less.

Newer quantitative diagnostic techniques can provide further information. These techniques include single photon emission computed tomography (SPECT), positron emission tomography (PET), magnetoencephalography (MEG), and diffusion tensor imaging (DTI) tractography. Although—at least in the United Kingdom—these techniques are not yet in routine clinical use, they may be useful in certain situations. SPECT and PET can enhance the ability to differentiate tumor recurrence from radiation necrosis in patients who have received prior radiotherapy and in whom structural imaging suggests progression. They may also help discriminate between CNS lymphoma and opportunistic infections such as toxoplasmic encephalitis in patients with acquired immunodeficiency syndrome. MEG and DTI tractography may be helpful in more precisely characterizing cerebral “disconnection” syndromes which may occur. MEG, along with functional MRI, is used in the preoperative localization of specialized cortical function such as motor, speech, and vision. These modalities aid neurosurgeons in planning stereotactic surgery that removes as much tumor as possible while minimizing the risk of critical cortical damage. Increasingly such location-critical operations are performed as "awake craniotomies" under sedation but with consciousness maintained. With the patient awake, surgeons can stimulate an area of brain before its removal and check whether it is functionally vital.

PSYCHIATRIC SYMPTOMS IN BRAIN TUMOR PATIENTS

Switching focus to the neuro-oncology team, we will consider the neuropsychiatric symptoms that may arise in someone who has a primary brain tumor.

Predisposing Factors

Tumor Location. The notion that brain tumors in particular lobes of the brain cause specific psychiatric symptoms is attractive but simplistic. Although there can be a relationship of sorts, anatomical tumor location is only one factor to consider in the assessment. A classically "frontal lobe" dysexecutive syndrome can arise in the context of a brainstem tumor. Symptoms may arise from structures seemingly far removed from the tumor due to the functional vascular phenomenon of diaschisis, or from structural white matter disconnection syndromes secondary to tumor-associated damage.

Older textbooks might hold that tumors in the left hemisphere cause depression whereas right-hemisphere tumors cause euphoria. Right-sided lesions of the temporal skull base have indeed been associated with a disinhibition syndrome in one small but interesting prospective cohort study conducted after CT scanning became widespread. Overall, however, the association between laterality and psychopathology is mostly inconsistent. Most large studies of this question predate the more recent psychiatric classification systems, many the advent of widespread CT scanning, and a few even the introduction of modern research standards of methodology and reporting. Most recent studies are limited to case studies or small series. It remains difficult to interpret the literature on how tumor laterality affects psychopathology.

It is likely that unpicking the causal relationships between tumor location and psychopathology will require more sophisticated functional neuroimaging techniques than CT or MRI. To date, few studies have utilized the likes of DTI tractography or SPECT to address this issue, however (see Jutten et al. for an interesting introduction to this topic).

Tumor Grade. Does tumor grade influence the development of neuropsychiatric symptoms? Like tumor location, various authors have proposed a causal relationship between WHO grade and psychopathology. In some ways, this is a plausible hypothesis. Advocates could postulate high/low-grade differences between: edema, ICP, and mass effect; metabolic profiles in the cellular microenvironment; or impairment of neuronal functioning by contact from invading cancer cells. Again, however, the evidence suggests a more complex picture. There is a lack of robust supporting evidence from prospective studies that directly compare high- and low-grade glioma patients. A few studies do make this comparison but lack power because low-grade glioma is rarer than high-grade glioma, and it is usually difficult to accrue large sample sizes. As it stands, therefore, evidence supporting a role for tumor grade is inconsistent. There is about as much supporting evidence for greater generalized distress in *low-grade* tumors, an association that could be rationalized a posteriori by the younger median age of low-grade brain tumor patients and the greater prevalence of epilepsy.

This slightly deflating appraisal of the roles of location and malignancy is perhaps a caution against making sweeping statements about how anatomo-clinical tumor characteristics affect psychiatric outcomes. Such statements risk: (1) ignoring the role of sociodemographic variables; (2) conflating cognitive with psychiatric psychopathology; (3) glossing over methodologic limitations in the literature; and (4) sounding so plausible that the first three things are overlooked. One of the most consistent findings is the absence of evidence correlating tumor- and treatment-related variables with neurotic psychopathology, but even this statement is based mostly on underpowered studies. Properly powered studies are required to identify brain tumor patients at high risk of future psychiatric complications.

The evidence that tumor and treatment characteristics can influence neurocognitive functioning is stronger but it remains difficult to compare clinical studies against each other. There is little doubt that brain surgery, radiotherapy, and chemotherapy can impair cognitive function, however. A recent retrospective study proposed a relationship between a certain molecular subtype of faster growing tumors and greater cognitive impairment. If confirmed by prospective research, insights like these may offer useful perspectives on mechanisms of neurocognitive impairment in other clinical populations and identify groups to target with early interventions.

Premorbid and Psychosocial Factors. In contrast to most tumor-related variables, the patient's premorbid psychiatric and psychosocial histories may well influence their risk of postoperative psychiatric complications including anxiety, depression, and distress. In some ways, this makes intuitive sense. Even after adjusting to the diagnosis there will be many sources of ongoing stress: neurocognitive and physical disability, side effects of surgery, radiotherapy, or chemotherapy, the threat of clinical progression, and many changes to lifestyle. One can expect that the patient's premorbid vulnerabilities, cognitive abilities, behavioral coping mechanisms, and psychosocial support systems would play important roles in determining the impact and degree of neuropsychiatric symptoms.

Contributory factors that determine psychiatric and behavioral symptoms co-occurring with brain tumors are therefore multiple and complex. To varying degrees and depending on the symptom of interest they may include tumor type, the rate and extent of tumor growth, anatomical location, the presence or absence of increased ICP, the type of treatment used and the severity of treatment complications, premorbid patient characteristics, psychiatric history, the adequacy of coping skills, and the quality of psychosocial and family support systems. In adults, on balance and with a risk of over-generalizing one might be particularly alert for neuropsychiatric and/or neurocognitive dysfunction in patients with any of the following: frontotemporolimbic, bifrontal, or deep midline tumors; large (>5 cm) or multifocal tumors; aggressively growing malignant tumors; raised ICP or obstructive hydrocephalus; significant tumor-related physical functional impairment; a history of any psychiatric disorder; prior brain injury; a family history of psychiatric illness; lower premorbid intellectual capabilities; poorer adaptive coping skills; or lower psychosocial support.

General Treatment

General Approach. When a psychiatric disturbance is directly caused by a brain tumor, its surgical removal may associate with remission of the psychiatric symptoms. Such improvements are however reported by only a few handfuls of single case studies and comprise a tiny minority of brain tumor-associated neuropsychiatry. Most psychiatric morbidity in brain tumor patients occurs after surgery during primary chemo- or radiotherapy and the subsequent follow-up. Depending on the problem, psychiatric input is most likely to be requested when a patient's behavior or neurocognitive symptoms start, worsen, or persist during primary treatment or follow-up. Dealing with the issue may reduce symptomatic distress, improve functional ability, and enhance overall well-being and quality of life in the time remaining to the patient. Accordingly, referrals should never be dismissed or evaded as "understandable/normal/futile because they have a brain tumor." There is nearly always something that can be done. Treatment typically involves nonpharmacologic support, pharmacotherapy when necessary, psychoeducation and support of the family, and clear communication about treatment between the psychiatrist and the neuro-oncology team.

The lifetime prevalence of psychiatric disorders is as high among people who will eventually develop brain tumors as it is in the general population. Patients with a history of previous depressive or anxiety disorders are at high risk of relapse due to the fear and ongoing stress of being diagnosed with a brain tumor, among other causes. Neurocognitive impairment is the exception to this general rule: this very common problem can often be attributed directly to the tumor or from side effects or complications related to its treatment. It is also one of the most challenging problems to treat.

In planning a therapeutic approach, the psychiatrist may find it useful to attempt to characterize the patient's psychiatric and behavioral symptoms as (1) a relapse or a continuation of pre-existing psychiatric disorder, or (2) arising de novo in response to either the psychological and/or biologic stress of having a brain tumor. This broad framework can be clinically helpful in communicating with patients and their families, and in planning optimal drug and nondrug interventions.

Pre-Existing Psychiatric Medication. The management of a relapse or a continuation of pre-existing psychiatric disorders is grounded in common sense and best practice. The balance of risks and benefits of treatment should be fully discussed with the patient. Drug treatment follows the same general principles as in the general psychiatric population. The prescriber should adjust doses (and occasionally drugs) as necessary to accommodate for the increased risk of CNS side effects. Common side effects include delirium and the increased potential for an epileptic seizure. Prescribers should anticipate cytochrome P450 (CYP450) system pharmacokinetic interactions between psychiatric drugs and the array of medication these patients may be on.

Many older psychopharmacological agents are potentially deliriogenic. Examples include the tricyclic antidepressants (TCAs), typical antipsychotics, anticholinergic antiparkinsonian drugs, the benzodiazepines as a group, and lithium carbonate. As a rule of thumb, these older agents are first-line only when they have previously been effective and well tolerated by the patient. The drug may be reintroduced at a low dose and titrated to efficacy. In many patients newer drugs may be better as first-line treatment. These include the SSRIs, other relatively nonanticholinergic antidepressants, atypical antipsychotics, nonbenzodiazepine anxiolytics, and anticonvulsant mood stabilizers. Caution is especially needed for older adults and those with multiple medical conditions who may be on many other medications.

It is common to hear patients and doctors voice concern about the risk of causing epilepsy. Many psychotropic drugs have the potential to lower seizure threshold, particularly in overdose. The risk of inducing seizures in brain tumor patients with different psychotropic medications at routine prescription doses is unknown. There is little brain tumor-specific evidence to inform practice. There is a lot of clinical anecdote suggesting that drugs are usually safe, but such general and uncontrolled views are difficult to extrapolate to individual cases.

Newer atypical antipsychotics may have less epileptogenic potential at comparable doses than older typicals. If high-potency older antipsychotics are required, haloperidol may be less likely to lower seizure threshold than others in its class. For antidepressants, licensing data from RCTs in patients without a brain tumor suggest that most antidepressants do not affect seizure threshold in therapeutic doses, apart from bupropion and maprotiline which have a relatively higher risk of inducing seizures. These particular drugs may therefore be best avoided where possible in brain tumor patients. A reasonable first choice when considering treatment of depression would be an SSRI with few effects on the CYP450 system, started at half-dose and monitored if there are particular concerns.

Pharmacokinetic interactions may occur through the CYP450 system; for example, some antidepressants are CYP450 system inhibitors. If antiseizure medication is metabolized by the same enzyme, administration of a directly competing SSRI could predispose to an interaction. In this particular example, the main theoretical risk would be of antiseizure medication toxicity (rather than increased metabolism and inefficacy leading to seizures). In the reverse direction, coprescription of an enzyme-inducing antiseizure medication might necessitate higher therapeutic levels of an antidepressant. This example is extremely general: prescribers should consider interactions on a case-by-case basis, consulting a pharmacist for specialist advice when in any doubt. If interactions cannot be avoided clinicians may wish to monitor antiseizure medication levels. Others may choose to closely monitor the patient.

In brain tumor patients with a history of schizophrenia or other psychotic disorders, atypical antipsychotics should be used where possible. In some circumstances it may still be necessary to use high-potency agents with a long history of successful use in medical disorders (such as haloperidol), or other individual drugs known to be well-suited to a particular patient. With appropriate dosage adjustments these may well be tolerated.

Patients with longstanding type I bipolar disorder have the same risk of developing a brain tumor as everyone else. As is also the case with schizophrenia, the combination of these two devastating illnesses is rare. A complicating factor in this population is that lithium carbonate can, at toxic or high blood concentrations, cause seizures and delirium and for this reason may be better avoided if possible. There may be cases where lithium has been used for many years and where the relative risks and benefits can only be weighed on an individual basis. Mostly though, in acute type I bipolar mania a mood-stabilizing agent with documented efficacy in acute mania such as valproic acid would be a reasonable first choice. An anticonvulsant mood stabilizer may have obvious additional advantages in patients with epilepsy. An atypical antipsychotic may be added if required. In type I bipolar depression there is a well-known risk of precipitating secondary mania or rapid cycling when traditional antidepressants are used. The anticonvulsant lamotrigine may be effective here and appears to have the advantage of not precipitating secondary mania or causing rapid cycling, although it also has the potential for causing rare but serious and potentially life-threatening dermatologic side effects.

Specific Neuropsychiatric Syndromes

Neuropsychiatric problems often arise after the diagnosis of a primary brain tumor. Here the most frequent ones are discussed for adults with a primary brain tumor (except delirium, which is also a common problem and is discussed in Section 10.2).

Depressive Disorder Due to Another Medical Condition (DSM-5 293.83)

CLASSIFICATION. When assessing a brain tumor patient for depression you will have to make a series of judgments that broadly fall into two groups. One relates to whether somatic symptoms (weight, sleep, and psychomotor changes, and fatigue) should count toward a diagnosis of depression given the considerable risk of confounding from biologic effects of the tumor and its treatment. The other group of judgments relates to whether psychological symptoms (depressed mood, anhedonia, guilt, poor concentration, and suicidal ideation) are abnormally severe in the context of multiple personal losses and a profound existential threat. Because these judgments are difficult, they can cause variability in clinical practice. This variability might threaten the intended reliability of the DSM system: indeed, that is probably what happens.

When assessing patients with a brain tumor for depression, doctors have to decide whether to attribute symptoms to a likely mood disorder, or else to the direct consequences of the tumor and its treatment. Different doctors may interpret the same diagnostic information in different ways, causing unreliability which may go on to influence their downstream treatment decisions. To test this, a written vignette was constructed for a hypothetical brain tumor patient meeting criteria for major depressive episode (DSM-IV-TR 296.2). The vignette was sent to NHS Consultants in five neuro-oncology-related specialties in Scotland. The aims of the study were to determine how many doctors diagnosed depression; whether their diagnosis influence their preferred treatment; and among doctors who diagnosed depression, whether concern over a history of epilepsy influenced their readiness to recommend an antidepressant.

The vignette contained eight of the nine symptoms of DSM-IV-TR MDD. However, each symptom also had a realistic alternative explanation directly related to the tumor or its treatment. All NHS consultant neurologists, clinical oncologists with neuro-oncology interest, neurosurgeons, and liaison psychiatrists working in Scotland, as well as the general practitioners (GPs) of all local glioma patients were invited to participate. Respondents were asked to state the diagnosis ("major depression" or "an understandable reaction"); specify treatment; and state to what extent the history of epilepsy had influenced their treatment decision ("not at all," "a little," "moderately," or "markedly"). The vignette read as follows:

You are seeing Mrs. Smith, a previously fit and well 54-year-old language teacher, in your clinic. Four months ago she presented with focal seizures with impaired awareness, followed by a prolonged generalized tonic–clonic seizure which required admission to ITU. CT head revealed a left temporal lobe tumor which was debulked and found to be a high-grade glioma. She completed a course of radical radiotherapy last month and continues on chemotherapy. There is no other past medical history and no past psychiatric history.

About a month ago Mrs. Smith found a website on the internet saying chemotherapy gives her a 20% chance of being alive in 2 years. She has thus been thinking a lot about her death and how it would affect her family, and has been tearful and feeling sad since then. Her hair fell out toward the end of the radiotherapy, which she found distressing. She used to enjoy sports, and meeting friends, but says that both the self-consciousness and a marked increase in tiredness have left her totally disinterested in her old activities for some weeks. She feels bad about this, worries that she is letting her friends down, and struggles with thoughts that she has become a “weak” person instead of being her usual strong self. She has put on 5 kg in weight since she stopped going out.

Mrs. Smith is currently unable to return to her teaching job due to persisting word-finding difficulties. She has begun to lie awake at night worrying about her finances. Since the radiotherapy finished, she has also been much more forgetful and finds it hard to read the newspaper. She smokes 10 cigarettes per day, does not drink and has a supportive relationship with her husband and two adult daughters. One has just become pregnant and Mrs. Smith finds the thought of not seeing her first grandchild particularly painful. She would consider going to Switzerland for euthanasia if she deteriorated and was no longer able to walk, but otherwise hopes to live to see the child.

Steroids were stopped soon after the operation. She was initially started on phenytoin but switched to a low dose of lamotrigine after the operation. It was increased slightly to control focal seizures a few weeks later. She remains on the same dose and has had no seizures since. She is on no other medicines and has no allergies.

Today’s brain MRI shows postoperative scarring with a small amount of residual tumor. At interview, Mrs. Smith sits with little spontaneous movement, and cries when talking about her family. She is not psychotic. Her speech is quiet and halting at times. Her concentration occasionally wanders but she is orientated to time and place. Mrs. Smith asks you if this is a normal reaction to her situation and treatment. She has no strong feelings about how you manage her.

Of 228 questionnaires distributed, 105 were returned complete (46% response rate). There was a clear split in opinion. Sixty-four doctors (61%) diagnosed major depression and forty-one (39%) diagnosed an “understandable reaction”. There were no significant interspeciality differences. Choice of diagnosis strongly influenced treatment. Antidepressants were more likely to be suggested when the doctor had preferred the diagnosis of depression (38/64 vs. 2/41; $P < .00001$). For an “understandable reaction,” meanwhile, counseling (23/41 vs. 15/64; $P < .001$) or watchful waiting (13/41 vs. 0/64; $P < .00001$) were more likely to be suggested as the first-line management. Eleven distinct types of counseling were identified with no consensus on the best sort. Just over half of the doctors diagnosing depression stated that the history of epilepsy had “moderately” or “markedly” influenced their choice of treatment (34/64). There was a trend for concern about epilepsy to be associated with avoidance of antidepressant prescription. GPs and psychiatrists appeared to be more concerned about epilepsy than neurologists. In summary, therefore, a sample of over 100 consultants presented with exactly the same clinical information showed marked disagreement in whether to diagnose MDD in a brain tumor patient. Preferred diagnosis strongly influenced preferred management. Psychiatrists were significantly more concerned about a history of epilepsy than neurologists. *Data collected by the author with supervision from Dr. Robin Grant and Dr. Alan Carson (Consultant neuropsychiatrist, NHS Lothian).*

Attempting to mitigate these problems by setting diagnostic “rules” for individual symptoms often raises further questions. One could, for example, decide to exclude all reports of insomnia if the patient is also on high-dose corticosteroids. But insomnia is a common symptom of depression and the relationship between steroids and sleep is not straightforward. In a local series of over 150 glioma patients, most of those taking dexamethasone shortly after primary surgery reported either no sleep changes at all or *hyper*somnia at clinical interview (author’s non–

peer-reviewed data). It may be impossible to confidently attribute causality to individual depressive symptoms.

What can be done? The “general comments” section above considered whether the patient’s depressive disorder arose (1) as a recurrence of prior depressive disorder, or (2) de novo in response to either the psychological or biologic stress of having a brain tumor. This provides a clinical framework by which to communicate the meaning of a diagnosis to patients. It may also suggest particular therapeutic strategies. However as can be seen above, the DSM entities of MDE or MDD do not make it easy to secure the diagnosis in the first place.

DSM-5 instructs users to discount depressive symptoms if the symptoms are clearly due to another medical condition. Yet, as shown in the vignette above, almost every depressive symptom can often quite plausibly be attributed to the brain tumor or its treatment. It may therefore be simpler instead to assess potentially depressed brain tumor patients for the simpler syndrome of “**Depressive Disorder Due To Another Medical Condition**” **(DSM-5 293.83**). This syndrome is more agnostic about the cause of symptoms. The criteria are discussed below: they are different from MDD in that they do not rest on theoretical judgments about the causes of symptoms. The presence or absence of a prior history of MDD/E can be labeled as secondary if desired.

A blanket approach like this could raise objections, including: (1) depressive disorder arising in patients with a prior history of depression may be a simple relapse rather than caused de novo by the tumor; (2) the lower diagnostic threshold could lead to overdiagnosis; and (3) individual symptoms still have to be weighed when assigning specifiers. But the strategy: (1) is testable by comparing the clinical course of brain tumor patients with, and without, a prior history of depression; (2) minimizes inter-clinician variability in practice, helping to standardize diagnosis; and (3) side-steps the questions of whether to count somatic and psychological symptoms toward the main diagnostic label. It is also easily communicable.

Apart from the upcoming section on “Diagnosis and clinical features” which for consistency also focuses on the suggested pragmatic diagnosis of **Depressive Disorder Due To Another Medical Condition**, the rest of this section is based on the historic literature. The authors of this literature, including the present author, used either traditional DSM diagnoses of MDD or questionnaires which were validated against DSM-IV MDD (or its equivalent at the time).

EPIDEMIOLOGY. MDD is more common in glioma patients than in the healthy general population. Retrospective clinical data from the U.S. Nationwide Readmissions Database (NRD) suggest that among people receiving craniotomy for glioma, the frequency of MDD is 11.4%; comparable to lung or breast cancer patients and greater than patients with head and neck, colorectal, and prostate cancers (see Rumalla et al.). The biggest individual study to use prospective “clinical interview” diagnostic methodology in these patients found that the 6-month prevalence following diagnosis of glioma was 20%; while reviews of studies using clinical interview methodology return a median frequency of 15%. If depressive symptoms are measured by rating scale, frequency estimates vary depending on the scale that is used. For example, the Beck Depression Inventory (BDI) returns a median frequency of 39%, whereas the Hospital Anxiety and Depression Scale (HADS) suggests a median of 16%.

Original data and reviews suggest that MDD is most consistently associated with impaired physical and cognitive functioning, longer-term corticosteroid use, a prior history of depression, and reduced quality of life. Unlike in the general population, female sex is not a clear risk factor. A concurrent clinical diagnosis of MDD has been associated with worse perioperative outcomes and greater depressive symptoms may predict increased mortality from the tumor, but further studies are needed on this issue. Longitudinal studies of depression in glioma are rare: based on very limited samples, a clinical diagnosis of depression can persist in about half of patients, for at least 3 months. The extent to which biologic symptoms of tumor or its treatment contribute to artificially inflate these estimates is not known.

ETIOLOGY. The etiology of MDD in brain tumor patients is likely to be multifactorial. As is the case in other forms of cancer, a prior history of depression is likely to confer increased risk of depression after diagnosis of a brain tumor such as glioma (“brain cancer”). In some cases therefore the cause is likely to be a relapse of a pre-existing tendency to depressive disorder. Genetic influences could be inferred when there is a family history of depression. Post-tumor diagnosis, a new case of depression could have a number of speculated possible causes, including tumor- or edema-associated disruption of critical tracts; a functional diaschisis; critical metabolic alterations of the neuronal microenvironment; the putative depressogenic effect of long-term corticosteroids; or the long-term effects of radiotherapy or chemotherapy on hippocampal neurogenesis or white matter tracts. A good case can often be made for emphasizing a psychological reaction to the devastating losses that accompany the onset, treatment, and natural course of many brain tumors. Various neuroendocrine abnormalities may be etiologically implicated in pituitary tumors including hypercortisolemia and hypothyroidism. It is, of course, usually very difficult to know which of these theories may be correct.

A subset of brain tumor patients meets criteria for a depressive disorder by having marked anhedonia but no depressed mood. This clinical picture is recognized in both pituitary tumor and glioma, and implies causal mechanisms distinct from the classical depressive picture of pervasively sad mood. The biologic basis of any such difference remains to be studied.

DIAGNOSIS AND CLINICAL FEATURES. **Depressive Disorder Due To Another Medical Condition** (Table 2.3–2) requires clinical interview evidence of a prominent, persistent, and socially disruptive change of mood. The change is depressive, apathetic, or both, and occurs in the context of a radiologically or histologically diagnosed brain tumor. Adjustment disorder with depressed mood should be excluded along with delirium. It can be further specified with respect to the full criteria for MDD/E: w*ith major depressive-like episode* if they are met, and *with depressive features* if not.

The traditionally required minimum duration of symptoms is 2 weeks. Some psychiatrists may prefer to allow at least 1 month to pass from the point of brain tumor diagnosis before considering whether a mood disorder is present, unless extremely severe symptoms manifest before then. Waves of grief are normal postdiagnosis. Social disruption can be difficult to assess: many brain tumor patients are unable to drive. They are often more socially isolated than previously and are unlikely, at least initially, to be able to work. Questions about social interaction should focus on what the patient *is* able do. The psychiatrist can assess their daily interaction with carers, family and friends, and whether this has changed along with mood. If anhedonia is the main clinical feature a similar problem arises: questions about “enjoying things they normally enjoy” will often be met with the reply that they “can’t do those things now” because they have a brain tumor. A useful question may be whether they still feel the same pleasure from spending time with grandchildren or other close family. Symptoms should be new and not attributable to the patient’s long-standing personality. Excluding adjustment disorder is a matter of assessing severity and persistence of the cardinal depressive symptoms. Excluding hypoactive delirium may be difficult if the patient is already cognitively impaired but in the history, the absence of fluctuating agitation or disorientation in time should be noted. Proxy report (e.g., from the patient’s main carer) is important and is likely to add to diagnostic reliability.

Table 2.3–2.
DSM-5 Criteria for Mood Disorder Due to Another Medical Condition (293.83)

A prominent and persistent disturbance in mood predominates in the clinical picture and is characterized by either (or both) of the following:

(1) Depressed mood or markedly diminished interest or pleasure in all, or almost all, activities or
(2) elevated, expansive, or irritable mood

There is evidence from the history, physical examination, or laboratory findings that the disturbance is the direct physiologic consequence of having a brain tumor

The disturbance is not better accounted for by another mental disorder (e.g., adjustment disorder With depressed mood in response to the stress of having a brain tumor)

The disturbance does not occur exclusively during the course of a delirium.

The following specifiers can be used to characterize the disorder:

- *With Depressive Features* if the predominant mood is depressed but the full criteria are not met for a Major Depressive Episode; *With Major Depressive-Like Episode* if the full criteria are met for a Major Depressive Episode; *With Manic Features* if the predominant mood is elevated, euphoric, or irritable; *With Mixed Features* if the symptoms of both mania and depression are present but neither predominates

This is a suggested diagnosis with some advantages in these patients (see text); psychiatrists may alternatively diagnose Major Depressive Episode/Disorder according to standard criteria.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:206.

LABORATORY INVESTIGATIONS. Investigations are as those required to exclude delirium and a physical cause for MDD. An endocrine screen is indicated in patients with a remote history of radiotherapy to the brain, if the radiation field included the pituitary gland.

DIFFERENTIAL DIAGNOSIS. The main competing psychiatric diagnoses are delirium, neurocognitive disorder due to a brain tumor, and (probably most commonly) adjustment disorder with depressed mood. The medical differential is as broad as that for MDD in patients without a brain tumor. In brain tumor patients specifically, potential medical causes include: postepileptic personality change; neuroendocrine abnormalities (including systemic diseases such as secondary hypothyroidism); and steroid-induced depressive symptoms.

COURSE AND PROGNOSIS. Longitudinal studies of depression in brain tumor patients are rare. This is partly because few have asked the question, but also partly because depression is more likely in functionally impaired patients with a poorer prognosis who are difficult to retain in long-term studies. Available evidence suggests that among patients who can be followed up, clinically significant depressive disorders can persist for at least three months. This duration is roughly in keeping with the median duration of depression in the non–brain tumor population. It lends some support to the argument that depressive symptoms cause significant clinical burden in these patients.

MDD and moderate–severe depressive symptoms have been shown to correlate with shorter survival times in meta-analyses of studies of cancer. They also correlate with functional impairment, which is a strong negative prognostic factor for brain tumor patients. An open question is therefore whether depression is independently predictive of mortality in brain tumor patients after controlling for functional impairment. Again only a few studies have asked and results are mixed. One difficulty in demonstrating such an impact is that median survival time for the most often-studied tumor (high-grade glioma) is measured only in months. A large sample would likely be required to show even a modest impact of depression on mortality. An individual patient data meta-analysis may be one way to clarify this issue, or more obliquely examining survival times in cohorts of people who receive antidepressants versus those who do not. Studies examining this issue must take care to untangle the prognostic effect of depression from that of poorer physical function (which correlates with depression and poorer survival and is therefore a prominent confound).

Clinical depression is associated with an increased risk of epileptic seizures in the general population. It is possible that left untreated, it may predispose these vulnerable patients further to epilepsy. Suicidal ideation is also a complication of uncontrolled MDD. Brain tumor patients have been reported to have one of the highest-risk ratios for completed suicide in the year following diagnosis of tumor. Reported suicidal ideation was however rare in other cohorts which examined the clinically busy first 6 months after brain tumor diagnosis. Overall the frequency of suicidal ideation is also unclear and deserves further study.

TREATMENT. No phase III RCT has examined the effectiveness of antidepressants for MDD in brain tumor patients. Although the same can be said for nearly every psychiatric condition in this patient group, the absence is felt more acutely here. Depression is a common condition which many psychiatrists will feel relatively comfortable recognizing and prescribing antidepressants for, yet in brain tumor patients, one moves a little closer to clinical equipoise between their effectiveness and risk. The equipoise is on the one hand whether these drugs are more effective than placebo in the grossly abnormal cerebral environment of a primary tumor (glioma especially). On the other hand, the likelihoods that they cause AED or chemotherapeutic toxicities by CYP450 interaction, worsen fatigue as per the main recognized side-effect profiles of all antidepressants, contribute via anticholinergic mechanisms to cognitive impairment, worsen seizure control in this highly susceptible population, or even by contrast improve seizure control by treating depression, are in each case also unknown. These factors (professional confidence, and for lack of a better term academic ignorance) might combine to create a “blind optimism” in favor of automatic prescribing for MDD. Conversely, doctors might fear unquantified consequences such as epilepsy, and hesitate to prescribe when treatment may actually be effective. The solution will be to conduct RCTs in depressed brain tumor patients who meet treatment thresholds at enrolment. In the meantime, treatment is based on commonly accepted principles and empirical best practice. Patients should be fully informed about the possible risks and benefits, treated cautiously, and reviewed regularly.

In some patients, antidepressants do appear clinically to be helpful in treating new depressive symptoms. The author routinely offers them to depressed brain tumor patients (subject to caveats and safeguards above), and they do often seem to help. Given that depression plays a major role in decreasing the overall quality of life that these patients experience, its early recognition and rapid treatment is potentially of great benefit if life is shortened by the tumor. An SSRI with a relatively clean CYP450 side-effect profile such as sertraline may be a reasonable first-line treatment. The SSRIs are increasingly clinically familiar, generally well-tolerated, and may be less likely to cause delirium than older antidepressants. A particular clinical drawback with these agents can be sexual side effects and brain tumor patients are already at high risk of sexual difficulties. Experience with newer drugs like SNRIs (e.g., venlafaxine) or Noradrenergic and specific serotonergic antidepressants (NaSSAs) (e.g., mirtazapine) in this patient group—in the research literature at least—is limited to sparse case reports. As in the case of prescribing any antidepressant, care should be taken to minimize the risk of a gastrointestinal bleed, a risk which may be higher in the older adults and in patients who have been prescribed corticosteroids to reduce brain edema. An ECG should be checked before prescribing to confirm that the QTc interval has not been prolonged by chemotherapy.

TCAs have a side-effect profile which includes sedation, anticholinergic effects, and orthostatic hypotension. For this reason, they are often considered second line unless previously effective for the patient. If TCAs are considered for refractory depression, some clinicians report that nortriptyline may be well tolerated. In some circumstances, such as severe refractory depression or where there is a long history of previous effective treatment, monoamine oxidase inhibitors (MAOIs) may be appropriate. Caution is required however as procarbazine—another drug of the MAOI class—may form part of certain chemotherapy regimens used to treat glioma. Before using these older agents, the psychiatrist should therefore discuss with the treating clinical team to confirm the choice of cancer treatment and hospital pharmacy to explore potential interactions.

Rarely and in severely depressed brain tumor patients refractory to drug treatment, ECT may be considered. ECT was once thought to be absolutely contraindicated in the presence of brain tumor. A recent review, however, examined published and indexed case reports of 75 patients who underwent ECT in the presence of a brain tumor over the last 80 years. The most common type of tumor was a meningioma or arachnoid cyst, indicating that the case report literature is strongest for patients with benign, small, or clinically insignificant tumors. In this context, unilateral brief pulse ECT has been reported to be safe, effective, and mostly well tolerated. Precautions have included close attention to the patient's capacity to consent to treatment, careful dose titration, and the need to monitor the patient carefully for post-ECT delirium. Dexamethasone was used pre-emptively in a substantial number of patients to reduce the risk that increased blood flow and edema around the tumor would lead to neurologic signs. Phenytoin was administered to one patient to prevent a prolonged seizure. Since the presence of a brain tumor may not be known when patients start ECT, the authors of the review suggest ordering neuroimaging in any patient displaying new neurologic symptoms manifesting after ECT, such as a significant lateralization of the seizure, secondary or prolonged seizures, Todd paresis, aphasia, or eye-movement disorders.

Psychostimulants can be effective in treating cancer-related fatigue and have been proposed as a treatment for depression in patients with brain tumor. Systematic trial evidence is currently limited. Modafinil is effective for cancer-related fatigue in general, but trials in brain tumor patients specifically are negative or absent. At present, there is no good evidence that modafinil is effective for depression in these patients. Methylphenidate may reduce subthreshold depressive symptoms, but studies to date measured these only as secondary outcomes and conclusions are inconsistent. Whether these drugs are more effective than placebo in treating clinical MDD is not known. Anecdotally, some clinicians find them helpful in some patients, but studies suggest a considerable placebo effect. If a trial of treatment is considered, methylphenidate has the advantages of a lower effect on seizure threshold and no major sedating or deliriant properties. Adverse effects are however estimated to be three times more likely than with placebo. In a recent meta-analysis, vertigo, anxiety, anorexia, and nausea were significantly more frequent in the treatment arm.

For all patients with MDD, psychological treatments should be considered and offered wherever possible alongside antidepressants. Once more it is unclear whether these treatments work (and if so which ones, and in what circumstances) in patients with progressive cognitive and communication deficits of varying severity. Decisions about their suitability for individual patients should be made empirically. Patients with a lower severity of subthreshold depressive symptoms and distress, meanwhile, may benefit from supportive or group therapy or referral to external services such as a brain tumor charity (nonprofit or patient advocacy support group, such as in the United States, the American Brain Tumor Association or the National Brain Tumor Society).

Anxiety Disorders

EPIDEMIOLOGY. As with depression, screening for anxiety disorders may capture elements of "normal" distress as well as the more severe and specific symptoms that present to adult psychiatry. In the brain tumor literature, anxiety has generally been measured using general self-report questionnaires. Frequency estimates usually range between 30% and 50%: how much would be regarded as clinically pathologic in structured interviews is not well characterized. These numbers probably reflect, in part, reaction to the natural fear, agitation, uncertainty and stress of having a brain tumor. Anxious reaction may be more likely in low-grade glioma, in younger adult patients, in women, and in those with lower premorbid IQ.

ETIOLOGY. Rarely, anxiety can be caused directly by the brain tumor. One example would be a coarsening of pre-existing anxious personality traits as the tumor disrupts higher inhibitory circuits. Another direct tumor cause would be stereotyped panic-like attacks occurring in the context of focal seizures with impaired awareness. Isolated case reports record the association of new-onset obsessive or compulsive symptoms with a brain tumor or its treatment in the brainstem, right parietal lobe, and left frontotemporal lobe (a tumor site that has also been associated with greater obsessional thinking in a small cohort study). A large part of anxiety and distress however usually arises as a psychological reaction to the stress of having a brain tumor. Anxiety presenting during outpatient follow-up may be related to occupational or social stress, especially in well-functioning patients with a low-grade glioma. For example, a hairdresser who has returned to work may have subtle cognitive deficits that cause difficulty in such a high-pressure environment. Or a patient may be under pressure from bullying work colleagues and find it harder to contain the anxiety this situation naturally creates. Both are real-life examples from the author's caseload.

DIAGNOSIS AND CLINICAL FEATURES. Generalized anxiety disorder (GAD, 300.02) requires symptoms of anxiety to be present for at least 6 months. Additionally and as with MDD, several further symptoms (e.g., being easily fatigued, or difficulty concentrating) can also be caused by the physiologic effects of the tumor or its treatment. This combination of features means GAD is unlikely often to be the preferred diagnosis. A more commonly diagnosable anxiety disorder in brain tumor patients, and perhaps the most common psychiatric disorder of all, will be **Adjustment Disorder (309.xx)**. The diagnosis (Table 2.3–3) requires emotional or behavioral symptoms occurring in response to the stress of having a brain tumor. The specific stressor may be the diagnosis itself or one or more physical or social complications that can arise during treatment and follow-up. Symptoms should arise within 3 months of the stressor. Criteria of “disproportionate” distress and occupational impact may be difficult to assess in the circumstances. Impairment in social functioning can be partly explored by asking about the impact of symptoms on contacts with friends. Additional specifiers code for prevailing mood and behavioral state. Differentiating an adjustment disorder (plus or minus specifiers for depressed mood and anxiety) from a depressive disorder is a clinical judgment that may turn on the severity or persistence of cardinal depressive symptoms such as low mood and anhedonia.

Table 2.3–3.
DSM-5 Criteria for Adjustment Disorder (309.xx)

Emotional or behavioral symptoms develop in response to an identifiable stressor or stressors within 3 months of the onset of the stressor(s) plus:

Marked distress that is out of proportion to the severity or intensity of the stressor, even when external context and cultural factors that might influence symptom severity and presentation are taken into account, AND/OR significant impairment in social, occupational, or other areas of functioning;

The stress-related disturbance does not meet criteria for another mental disorder and is not merely an exacerbation of a pre-existing mental disorder;

The symptoms do not represent normal bereavement;

After the termination of the stressor (or its consequences), the symptoms persist for no longer than an additional 6 months

The following six specifiers are used to identify subtypes of adjustment disorder:

- With depressed mood
- With anxious mood
- With mixed anxiety and depressed mood
- With disturbance of conduct
- With mixed disturbance of emotions and conduct
- Unspecified

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:319, 320.

Diagnostic features of panic disorder (300.xx) are discussed in later chapters. In the general population, classical panic attacks have a characteristically sudden onset, stereotyped (repeating) pattern of features, and relatively short duration. In brain tumor patients, similar things can be said about some types of epileptic seizure. Although it is not a very common occurrence it may be difficult, purely on the basis of history, to distinguish between these two causes of paroxysmal panic symptoms. One symptom which may point toward focal epilepsy is the presence of an unusual prodromal aura (such as deja vu, rising gastric sensations, a strange smell, or visual disturbance). Another is the presence of orofacial movements during the attack. For this reason, a corroborative eyewitness account is recommended when assessing panic in a brain tumor patient. With the patient's consent carers may be asked to video-record an attack on their mobile phone for the psychiatrist's review.

DIFFERENTIAL DIAGNOSIS. Psychosis and depressive disorder should be excluded. Alongside the medical differential for anxiety disorders, it may be helpful to consider in particular focal epileptic seizures, pituitary-related endocrinologic abnormalities (hypopituitarism being a late consequence of certain radiotherapy regimes), and PE/DVT presenting as difficulty breathing or hyperventilation.

LABORATORY INVESTIGATION. Most anxiety will be—for lack of a better term—psychological in origin. It is still valuable to exclude potential reversible physical causes. Brain tumor-specific investigations may include EEG, endocrine blood screen, and a chest x-ray as per the medical differential.

COURSE AND PROGNOSIS. In cancer generally, distress characteristically peaks at the times of diagnosis, end of primary treatment, relapse, and the terminal phase. Distress occurring at these points can at first be observed supportively unless it is very severe. Patients can be reassured that it is normal and may pass. For example, much of the generalized distress and grief at diagnosis resolves (as one might expect) within the first few months. Brain tumor patients often face the added and recurring stressor of routine interval scanning: repeat neuroimaging at intervals usually ranging from every 3 months to every year and tracking progress of the tumor. These times can be highly distressing for those affected. They are temporary and supportive care is again indicated.

Symptoms may persist preferentially in the minority of patients with other risk factors for psychiatric disorders, a pre-morbid anxious predisposition, or a direct tumor-related cause of symptoms. In these patients, or in those with severe levels of distress or anxiety at the points noted above, more systematic treatment may be indicated.

TREATMENT. Currently, no phase III RCTs have studied interventions for anxiety in brain tumor patients. General management principles from adult psychiatry should be applied. In a demonstration of their power, anxiety management and relaxation techniques—including listening to music, good communication, aromatherapy, slow breathing, distraction, visualization, and prayer—have been used intraoperatively to facilitate awake craniotomy without sedation and only minimal analgesia. A biopsychosocial formulation may be a useful starting point for treatment decisions. Full and detailed explanations of all diagnostic procedures and proposed treatments may help. There is observational evidence that information-giving lowers anxiety, for example, in the perioperative phase. The patient and family should have the opportunity to ask questions. Patients can also be referred to a local brain tumor charity, patient advocacy support group, or non-profit organization.

In patients whose anxiety persists or is moderately severe, talking therapies may be indicated. Not all patients may be suitable candidates, however. Severe cognitive, functional, or practical concerns may hamper their delivery. Where possible, adjustments may need to be made to allow the patient to participate. Questions to consider include: (1) Does the patient have cognitive dysfunction that is too severe for them to engage with treatment, remember, and internalize therapeutic strategies? (2) Do they have any physical impairment that may limit their participation such as dysphasia or uncontrolled epilepsy? (3) If the patient is unable to drive, how will they get to regular sessions? and (4) If there *is* a cognitive, physical, or practical difficulty, is it possible to adapt therapy to compensate? Questions of this sort should ideally be considered by the psychiatrist or psychologist together in liaison with the neuro-oncology team.

Anxiolytic drug treatment in brain tumor patients may include SSRI antidepressants, low-dose regular trazodone, or buspirone, in conjunction with supportive or structured psychotherapy. Alternative medications such as pregabalin (Lyrica) and propranolol (if tolerated and safe to prescribe) may be helpful. Where essential for short-term control of severe anxiety, low-dose, long-acting benzodiazepines may have a role. Antipsychotics should probably be avoided as treatments for anxiety unless specific psychotic symptoms are associated with the patient’s anxiety and agitation, or they have been first-line for that patient before.

The general principle for patients with a pre-existing, longstanding anxiety disorder who subsequently develop a brain tumor is to continue previously effective treatment where possible. In line with the general comments above attention should be paid to the possibility of interactions with other prescribed medicines. One or more of the selective serotonin reuptake inhibitors, buspirone, propranolol, or clonazepam may be effective and well tolerated in treating acute symptomatic exacerbations. Patients with fear or panic disorder symptoms occurring as a part of a temporal lobe tumor-induced focal seizure may respond to treatment with ASMs. If such symptoms occur during the interictal period, anxiolytic agents discussed previously may be helpful. Close collaboration with the neuro-oncology team is essential.

Steroid-Induced Psychotic Disorder (292.9)

EPIDEMIOLOGY. Corticosteroids are prescribed to brain tumor patients to reduce brain edema, lower ICP and thereby improve physical and cognitive functioning. The typically high doses (in the United Kingdom, the standard initial dose is 8-mg dexamethasone orally, twice per day) and often urgent necessity mean that newly diagnosed brain tumor patients are at risk of developing a steroid-induced psychotic disorder (hereafter "steroid-induced psychosis"). This sounds worrying but most brain tumor patients do not become psychotic from any cause, including steroids. The frequency of steroid-induced psychosis is estimated at under 5% from review of case and clinical study literature. Much of this literature is focused on rheumatology patients receiving prednisolone. It suggests that women may be more at risk, but the most consistently identified risk factor is higher steroid dose.

ETIOLOGY. The relationship between dose and emergence of psychotic symptoms is incompletely understood: while probably not linear, psychotic symptoms are clearly more likely in patients on very high doses. The molecular pathologic mechanism is unknown and may be due to a primary effect of the corticosteroid on brain function or a secondary effect (e.g., severe sleep disruption), coupled with particular—and currently unknown—biologic predisposition(s).

DIAGNOSIS AND CLINICAL FEATURES. The diagnosis of steroid-induced psychosis (Table 2.3–4) rests upon psychotic symptoms (delusions and/or hallucinations, together with other symptoms consistent with psychosis) occurring in association with the use of corticosteroids. The risk is highest soon after starting treatment and most patients destined to develop psychosis are thought to do so in the first fortnight from steroid onset. On stopping the offending drug, most cases of psychosis could be expected to settle within a week. Associated affective symptoms may take longer to resolve. High dose (e.g., >40-mg prednisolone equivalent/day) makes the syndrome more likely but is not essential: the reaction appears to be idiosyncratic. The psychiatrist should exclude a prior history of primary psychosis and, if possible, delirium. The latter may not be straightforward as the clinical picture may fluctuate from both psychosis and delirium, and white cell count is likely to be raised secondary to the corticosteroid. To help exclude delirium clues to infection may be sought in the physical examination. These might include fever, tachycardia, and a raised respiratory rate, as well as various focal signs of infection. Investigations can help include metabolic and infective causes (electrolyte or liver function abnormalities, urinalysis, chest x-ray if possible). The drug prescription sheet should be reviewed for any potentially deliriogenic medicines. The tempo of onset in relation to steroid prescription should be considered.

Anecdotally, steroid-induced psychosis often has a prominent affective rather than schizophreniform flavor. There is however, no sole characteristic phenotype other than the presence of delusions or hallucinations. Anecdotally oncologists treating brain tumor patients comment that many of their patients are often subtly elated for the first few weeks of corticosteroid treatment, without ever becoming frankly manic or psychotic.

Table 2.3–4.
DSM-5 Criteria for Steroid-Induced Psychotic Disorder (292.9)

Evidence of symptom(s) of psychosis:

- Abnormal psychomotor behavior
- Negative symptoms
- Impaired cognition
- Depression
- Mania
- Disorganized speech

PLUS at least one of the following:

- Delusions
- Hallucinations

The symptom(s) developed during or soon after substance use or withdrawal, AND the involved substance is capable of producing the symptom(s);

The disturbance in not better explained by a psychotic disorder that is not substance-induced;

The disturbance does not occur exclusively during the course of a delirium;

The disturbance causes clinically significant distress or impairment in important areas of functioning

Data from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

DIFFERENTIAL DIAGNOSIS. The differential diagnosis of steroid-induced psychosis in brain tumor patients includes delirium, psychosis secondary to focal epilepsy, MDD with psychotic features, and a stress-induced relapse of a pre-existing primary/affective psychosis. There is no clear evidence that either paranoid schizophrenia or type I bipolar disorder are associated with brain tumor whether in cause or effect. In the fast-moving environment after presentation to hospital however, on occasion the clerking medical history may be incomplete, especially for historical psychiatric episodes.

LABORATORY INVESTIGATION. Treat the patient as per any severe psychotic episode, with the addition of trying to reduce or stop the offending steroid.

Physical investigations to consider as the clinical situation allows are those relevant to delirium and the major primary psychotic syndromes. If obtainable, an EEG may be clinically useful. Epileptogenic foci in deep frontal lobe structures can occasionally present with a bizarre behavioral syndrome that looks quite unlike either a classical generalized or focal seizure with impaired awareness.

COURSE AND PROGNOSIS. Psychotic symptoms should abate with the combination of antipsychotic and a lowering or reduction of steroid dose. If there is no evidence of improvement, perhaps particularly in patients with schizophreniform features, consider the possibility of an emergent or previously undiagnosed psychotic illness such as schizophrenia or bipolar I disorder. In this situation, a detailed family history may be of additional value. Coemergence of the catastrophic events of brain tumor and a primary psychosis is rare but does occasionally happen in younger patients. The literature on this topic consists almost entirely of single case reports and there is currently no robust evidence that it is anything other than a coincidence of timing.

The literature is similarly sparse on prognostic guidance. What evidence there is suggests that prior steroid-induced psychosis may not accurately predict the likelihood of future episodes. Unlike, say, puerperal psychosis where a previous episode greatly increases the future risk, florid psychic reactions to corticosteroids appear to be largely idiosyncratic both between individuals, and within the same individual over time.

TREATMENT. If possible, the corticosteroid should be reduced or stopped. This may well have consequences such as the re-emergence of cognitive or neurologic symptoms and the decision rests on a balance of risks specific to each individual case. Check with the treating medical or surgical team before changing the steroid dose. Sometimes the risk of changing it may be deemed too high, for example, in the case of a brainstem tumor threatening to compress respiratory centers.

Standard nonpharmacologic methods to reduce patient agitation, anxiety, and disorientation can be applied. These methods are identical to those used widely for delirium and dementia and include nursing by familiar faces, adequate ambient lighting, regular reassurance and reorientation, and de-escalation and gentle distraction when necessary. Antipsychotic medication is also indicated. Brain tumor patients are likely to be antipsychotic naive; they are also at high risk of epilepsy. Neither efficacy nor the risks of causing a dystonic reaction or epileptic seizure by prescribing any antipsychotic to brain tumor patients are conclusively known. Case studies in varying clinical populations have reported successful treatment with haloperidol, quetiapine, risperidone, olanzapine, or aripiprazole. If the patient is manic, a mood stabilizing agent may also be required (e.g., sodium valproate, or lithium). If the patient is depressed, TCAs are best reserved for second line: there is some evidence from older studies that these drugs may worsen symptoms. The successful use of ECT for steroid-induced psychotic depression has been reported.

The acutely disturbed brain tumor patient who presents an immediate danger to themselves or others should be managed according to local rapid tranquillization protocols. Such extreme situations are rare. If an intramuscular injection is necessary to rapidly sedate the patient, close nursing observation is essential afterwards. Always consider risk and communicate and document decisions on risk mitigation, as for any psychiatric patient. The patient should be reviewed regularly by the psychiatrist, since the neuro-oncology team may wish to restart steroids. It is probably sensible to seek to minimize the future doses of both corticosteroid and any ongoing antipsychotic as a starting point. Clinicians should also check that the patient's carers know how to contact services in the event of any future emergency.

Personality, Cognitive, and Behavioral Change. Personality, cognitive, and behavioral changes are among the most distressing symptoms for families and carers. Evidence specific to brain tumor patients on how best to manage these problems has however been lacking. Defining "personality change" is not entirely straightforward and studies make variable use of questionnaires about changes in attitudes, objective behavior, and/or cognitive function. It can be especially challenging to untangle the influence upon personality of cognitive function, which may be markedly affected by a brain tumor. In some cases it may be simpler to focus on the identification, formulation, and management of observable behavior.

EPIDEMIOLOGY. Personality changes are very frequent and reported across tumor subtypes, ages, and stages of disease. Studies conducted in various primary brain tumor populations report a point prevalence ranging from between 16% in patients with low-grade oligodendroglioma during follow-up to over 70% in patients with primary CNS lymphoma. In general terms, only a minority of patients will present with clear behavioral symptoms from the outset and they probably become more common as the disease progresses. Behavioral symptoms are highly prevalent in the palliative stage of glioma. There is no evidence in adult patients that men and women are affected differently.

In practice, cognitive symptoms are meanwhile very common following radiotherapy. A heuristic working model is that radiotherapy can particularly affect elements of executive function and information processing. The academic literature on postradiation neurocognitive impairment, however and perhaps surprisingly for an anecdotally prevalent issue, is contradictory and difficult to concisely summarize. Discrepancy between study outcomes may arise from the many confounding variables including: often retrospective study designs; tumor type, location and extent of response to radiotherapy; level of patient fatigue, depression, or effort; and differences in cognitive outcome measures. Higher radiation doses, broader radiation fields, and radiation delivered to children or the older adults may confer higher risk of cognitive impairment. Interested readers are directed to the recent review by Koutsarnakis et al. as an introduction to the difficulties of accurately studying this topic.

ETIOLOGY. Patients with brain tumors report significantly more post-operative subjective emotional changes compared to non–brain tumor postsurgical controls. Irritability and mental inflexibility may be particularly common in the early period postsurgery. In the longer term, it is not uncommon to see a “coarsening” of pre-existing personality traits, or childlike dispositions.

Unlike in neurology—where destruction of a critical circuit or nucleus may cause predictable “hard” neurologic signs—in neuropsychiatry the relationship is often less clear. One does see references linking tumor lobe location (e.g., “the frontal lobe”) with specific psychopathology, but these are usually general statements repeated from venerable older textbooks or historical small case series. More recently, it has been recognized that the brain is organized primarily in circuits and small-world networks, not siloed lobes.

It is possible that behavioral correlates of frontotemporal–subcortical circuits may include apathy (ventromedial prefrontal cortex), lability or disinhibition (orbitofrontal cortex), or aggression and paranoia (medial temporal lobes). However, unpicking the role of tumor upon the functionality of a specific brain circuit is difficult and consequently, the question of whether a tumor is “causing” personality or behavior change is often unresolved. There is a lack of high-quality evidence to underpin simplistic or blanket assertions. Recent studies have used sophisticated neuroimaging to link tumor location with behavioral phenotype, but consist mainly of case reports. It is plausible that critically located tumors may have specific neuropsychiatric consequences (see case study below), but in practice the etiology of personality change is messier, difficult to prove, and potentially encompassing biologic, psychological, social, and iatrogenic causes. Depressive and anxiety symptoms, too, are associated with poorer attention and processing speed in primary brain tumor patients receiving radiotherapy.

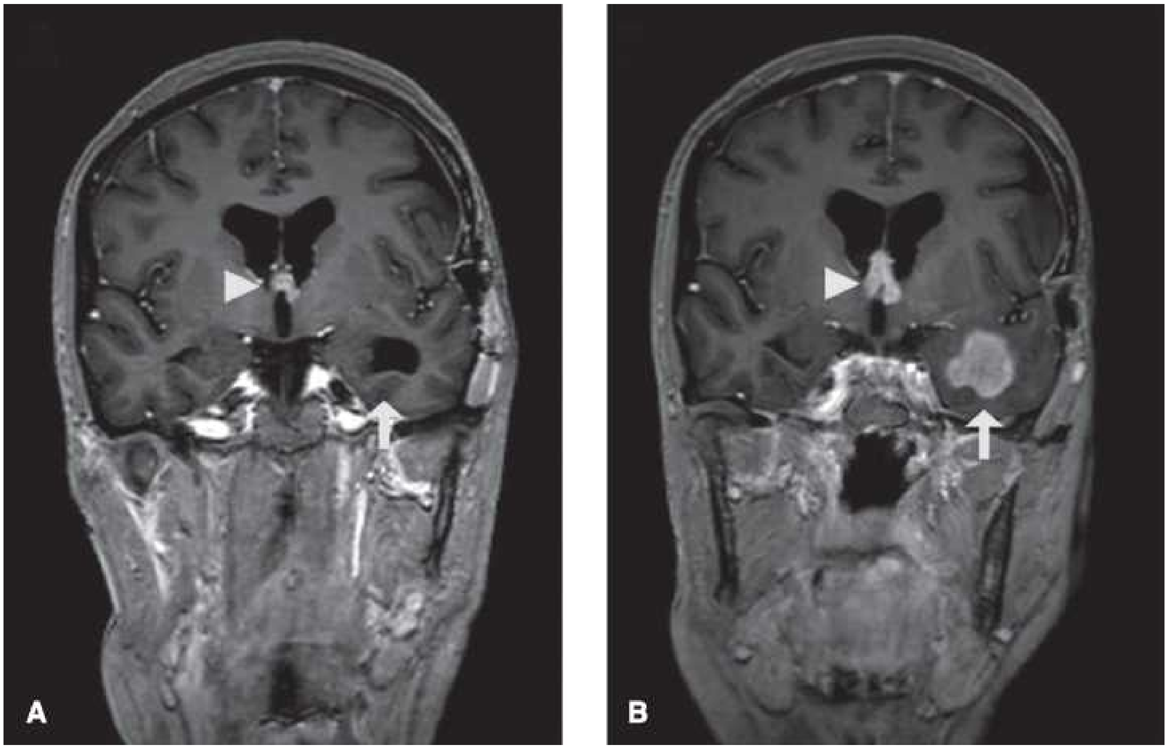

FIGURE 2.3–3. T1-weighted post-Gadolinium coronal MRI scans of the same patient after primary chemoradiotherapy (panel **A**), then 4 months later (panel **B**). Note progressive forniceal tumor (*arrowheads*) which correlated with the onset of dense anterograde amnesia. Recurrent tumor in the left temporal lobe is also present in panel **B** (*arrow*). (Courtesy of Dr. David Summers, Consultant Neuroradiologist, Department of Clinical Neurosciences, Royal Infirmary of Edinburgh)

Table 2.3–5.

Major Neurocognitive Disorder Due to Another Medical Condition (294.11)

Evidence of significant cognitive decline from a previous level of performance in one or more cognitive domains—such as complex attention, executive function, learning, memory, language, perceptual-motor, or social cognition

This evidence should consist of:

- Concern of the individual, a knowledgeable informant (such as a friend or family member), or the clinician that there's been a significant decline in cognitive function; and
- A substantial impairment in cognitive performance, preferably documented by standardized neuropsychological testing. If neuropsychological testing is not available, another type of qualified assessment is accepted

The cognitive deficits interfere with independence in everyday activities (e.g., requiring assistance with complex instrumental activities of daily living, such as paying bills or managing medications)

The cognitive deficits do not occur exclusively in context of a delirium, and are not better explained by another mental disorder

This diagnosis should be used to code for brain tumor-associated personality and behavioral change which is felt to be driven primarily by cognitive impairment.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:729, 730.

DIAGNOSIS AND CLINICAL FEATURES. For some patients the appropriate DSM-5 diagnosis will be **Personality Change Due To Another Medical Condition** (310.1). Where the dominant feature underlying the behavioral syndrome is judged to be neurocognitive dysfunction, a more appropriate diagnosis may be **Major (or Mild) Neurocognitive Disorder Due To Another Medical Condition** (294.11 or 331.83, Table 2.3–5). To illustrate, a case study of one of the author's patients is briefly presented in the box below and Figure 2.3–3.

A 54-year-old female professional manager, who had been previously well, presented abruptly with a cluster of generalized tonic–clonic seizures. Brain MRI revealed a nonenhancing diffuse cortical lesion in the medial aspect of the left temporal lobe. The left hippocampus was abnormal and swollen. A glial series tumor was suspected, but acute changes related to seizures could not be excluded. A short period of follow-up with repeat scanning was recommended. Repeat MRI subsequently showed an increase in size and enhancement of the mass. Following surgical biopsy, neuropathologic testing was unable to fully characterize the tumor. Further surgery was considered too risky due to lesion location. A grade III (anaplastic) glioma was presumed from key molecular characteristics although lymphoma could not be excluded.

Initially, the only psychiatric symptoms displayed by the patient were anxiety attributed to the uncertainty regarding diagnosis, and perhaps some mild forgetfulness which was regarded as very subtle by her husband and clinically negligible by her treating team. She remained well with no physical complaints or clinical evidence of disease. She received treatment of the presumed anaplastic glioma with radiotherapy and chemotherapy.

A few months after initially presenting she was referred to neuropsychiatry. She had symptoms of normal grief—denial, waves of sadness, anger, and bargaining—but also some new depressive symptoms which were unusual for most people in her situation: anhedonia, early morning wakening, deriving no pleasure from food, and worsening anxiety. She commenced mirtazapine for a presumed depressive disorder. This led to improvements in sleep, appetite, and partially mood, though grief at the loss of future time with her family remained prominent. She was referred to neuropsychology but declined cognitive assessment. Six months after her initial presentation, she had returned to work—from home and part-time—successfully doing remote administrative tasks. She felt back to around 80% of her old cognitive functioning. She continued to participate in psychiatric follow-up with no evidence of significant cognitive impairment in general conversation. MRI scans from this time revealed reduced tumor volume in the left temporal region and minimal residual enhancing disease elsewhere (Fig 2.3–3A).

Her personality markedly changed over the next 4 months. Gradually, anxiety increased further with the onset in particular of repetitive questioning. She could not be satisfied with answers, asking her husband questions again only minutes after they had been answered. She became very agitated at times and was admitted to hospital in high distress on two separate occasions. On closer questioning her husband reported the gradual but progressive onset of memory difficulties. Further psychiatric review was arranged. On assessment, there was a new dense anterograde and partial retrograde amnesia. She was aware of having a brain tumor but was now unable to remember previous psychiatric appointments or discussions. Her ability to register and learn new material was intact, but she was completely unable to recall or even recognize it after a few minutes. Strikingly when shown a clock face she had drawn minutes before, she looked bewildered and said, “When did I draw that?”

Repeat MRI and neuroradiologic review confirmed tumor recurrence in the left temporal lobe, notably with extensive, progressive tumor now also in the midline, at the anterior columns of the fornices and foramina of Munro (Fig 2.3–3B). The clinical finding of new-onset anterograde amnesia—a form of nonalcoholic Korsakoff syndrome—was most likely attributable to the forniceal location of progressive tumor.

This lady’s case is an unusual example of tumor in an anatomically critical location causing a discrete and identifiable neuropsychiatric syndrome. Her personality changes of intractable anxiety, repetitive questioning, and severe agitation were secondary to a near-complete anterograde amnestic deficit. It is more commonly the case that tumor location and psychiatric phenotype are not so clearly linked. *The author is extremely grateful to the patient and her husband who both consented to this anonymized summary being presented for teaching purposes.*

Assessment of behavioral change should be person-centered, encompassing behavioral analysis, cognitive functioning, and assessment of the patient’s emotional state. Assessment by both neuropsychiatry and neuropsychology is recommended. Reversible causes should be excluded or addressed including depression, anxiety, delirium, epilepsy, hypoxia, hypoglycemia, pain, confusion, or frustration at physical or communication deficits. Brain tumor medication may contribute to unwanted behavioral symptoms, including a disinhibition syndrome secondary to corticosteroids or ASMs. The ASM levetiracetam has been reported to have association with increased irritability in some patients. However, these medications may also be essential for control of brain edema or epilepsy, and full discussion with the treating team is vital.

One should aim to identify and characterize the main behavioral syndrome. A collateral history from a carer who knows the patient well is essential. The problems most likely to be reported are anger, apathy, emotional lability, or disinhibited behavior. Establish the patient's premorbid personality and behavioral characteristics and confirm that the emergence of the troublesome behavior aligns with the timeline for presentation or treatment of the brain tumor. The new behavior may be context-specific but should persist over time. An assessment of exacerbating and alleviating factors may suggest strategies for treatment.

DIFFERENTIAL DIAGNOSIS. The psychiatric differential includes delirium, chronic pain, substance dependence, MDD, and GAD. The psychiatrist should consider reversible causes and the possible role of any medication.

LABORATORY INVESTIGATION. Physical investigations should be targeted at reversible medical causes. They may include as appropriate: standard hematologic and biochemical screens for causes of delirium; pulse oximeter reading to exclude hypoxia (DVT/PE is very common in these patients); blood-glucose measurement to exclude steroid-associated diabetes and consequent hypoglycemic episodes; chest x-ray to exclude chest infection or DVT/PE; and a repeat head CT or MRI to exclude tumor progression or intratumoral bleed.

COURSE AND PROGNOSIS. There is limited observational evidence of the longitudinal evolution of personality changes in adult brain tumor patients. Anecdotally once established these symptoms tend to persist and are difficult to treat. The patient's family can learn effective behavioral management strategies to mitigate the causes and consequences of the patient's symptoms. Psychiatry and psychology services may helpfully contribute by teaching and reinforcing these strategies, prescribing medication if necessary to manage any unacceptable risks to the patient or carers, and signposting carers to trusted support services.

TREATMENT. Few RCTs have studied interventions for behavioral syndromes in brain tumor patients. A single-center randomized control trial showed some benefit from a home-based psychosocial intervention combining neuropsychological assessment, psychological therapy, and a person-centered approach to treatment (see Ownsworth et al.). In general however, there is a lack of specific therapy resources tailored to problematic behaviors in brain tumor patients. Clinical management is mostly based on discussion with the patient and family and a tailored behavioral approach with carer education and support. Many situations can be managed acceptably by treating reversible causes, educating the patient and carer, and using behavioral strategies. A range of nonpharmacologic management strategies, which are not specific to brain tumor but likely to be generalizable across patient groups, are provided in the narrative review by Leppla et al. (see Further Readings).

In terms of medication, there is currently little high-quality brain tumor-specific evidence to guide practice. As always, one would weigh the risks and benefits of treatment and consider empirical therapeutic trials. Depression and anxiety symptoms should be sought and treated where possible. Where the main problem is cognitive impairment, a trial of a psychostimulant such as methylphenidate or modafinil should be weighed against the risk of causing seizures in these vulnerable patients. In settings of severe uncontrolled disinhibition, options may include treatment with a mood stabilizer such as lithium, carbamazepine, or valproic acid. Explosive anger characteristic of acquired brain injury might respond to β-blocker such as propranolol, if the patient is physically able to tolerate it. Low-dose risperidone has been reported as of some benefit in a case study set in palliative care. Other single-case studies have reported improvement in psychomotor retardation, abulia, and apathy after experimental treatment with dopamine agonists such as bromocriptine. Behavioral agitation associated with the terminal stages of disease can be complex to manage, and palliative care / hospice services should be involved.

A small number of RCTs have examined cognitive rehabilitation strategies for tumor- and treatment-associated cognitive dysfunction (e.g., Gehring et al.). These trials were not aimed specifically at the problem of personality change, but cognitive changes may often affect behavior. Cognitive rehabilitation techniques might therefore be of use in treating personality or behavior change in these patients. They can be delivered successfully both in the early postoperative and the later outpatient phases of tumor treatment. In general, studies show that while these treatments can lead to significant improvements, they are usually limited to narrow psychometric parameters (e.g., attention and visual memory). It is unclear whether cognitive retraining is effective for more complex social behaviors underpinned by executive function and impulse control, and “real-world” outcomes like return to work are largely unstudied.

Some studies have also tested the efficacy of medications for the treatment or prevention of cognitive changes in brain tumor patients. Methylphenidate, modafinil, and donepezil have shown some efficacy in small phase II RCTs. These studies recruited general samples of brain tumor patients rather than targeting those with frank behavioral disturbances. They showed modest improvements in relatively narrow neuropsychological parameters, again mostly attention and some aspects of memory. The clinical utility of these treatments has not been studied in large phase III studies. Psychiatrists may consider therapeutic trials as appropriate in individual patients.

Childhood Brain Tumor Survivors. Brain tumors are the leading cause of mortality from childhood cancer. Childhood brain tumors however show major differences from adult tumors in molecular characterization, histology, intracerebral location, and prognosis. Epidemiologic datasets report 5-year survival rates of pediatric brain tumor in the region of 80%. Increasing life expectancies have potential to increase the personal and societal impact of neuropsychiatric morbidity on quality of life and cognitive function. Survivors of childhood brain tumors can display complex symptoms and disabilities, at the crux of which are the long-term cognitive effects of surgery, radiotherapy, and chemotherapy on the young developing brain.

EPIDEMIOLOGY. Keng et al. reported a recent systematic review of the neuropsychiatric consequences of pediatric brain tumor (see References). They discussed acute and longer-term cognitive deficits, and a range of emotional and behavioral disorders including depression, anxiety, somatization, distress, and ADHD.

Acute Cognitive Deficits. Cerebellar mutism (posterior fossa syndrome) is a common early complication of childhood brain tumor resection, arising in the first few days postoperatively, and characterized by postoperative mutism and reduced speech. Reported risk factors include medulloblastoma histopathology, brainstem involvement, midline location, larger tumors, periventricular ischemia, higher mean body temperature in the first few days after surgery, and lower socioeconomic status. Cerebellar mutism usually recovers within months but has the potential for ongoing deficits in IQ, memory, and learning.

Longer-Term Cognitive Deficits. Late (>1 year) cognitive outcomes have received the most attention. Although some recovery can be expected, childhood deficits have long-term impacts on psychosocial development. IQ is hard hit with scores approximately 1 to 2 *standard deviations* lower than population norms. Other studies report deficits across cognitive domains with impaired attention, visuospatial function, memory, learning, processing speed, and/or executive function/planning. Some clinical variables have been associated with poorer cognitive outcomes including younger age of treatment, medulloblastoma histopathology, supratentorial location, receiving chemotherapy, and a history of uncontrolled epilepsy. However at an individual level it is difficult to predict which children will face the greater impairments later in life.

Emotional and Behavioral Disorders. A recent meta-analysis indicates that survivors of pediatric brain tumors have relatively higher levels of depressive symptoms, anxious symptoms, and general distress than healthy comparison groups. Other reviews report that behavioral and social problems may be common at school, and that survivors tend to report lower amounts of physical activity than peers. Estimates of ADHD prevalence in childhood brain tumor survivors range from 9% to 23.8% in association particularly with a history of adjuvant cranial radiation, intrathecal chemotherapy, and corticosteroid use.

Among adult survivors, frequency estimates for affective symptoms range from 9% to 18% with potentially a high prevalence of suicidal ideation. Survivors with depression, anxiety, or somatization report lower quality of life and greater psychoactive medication use. Evidence is currently scarce to draw firm conclusions about PTSD, eating disorders, mania, psychosis, or catatonia.

ETIOLOGY. In acute pediatric neuro-oncology, posterior fossa syndrome has been associated with abnormal functional neuroimaging in mute postoperative patients, with observation of global supratentorial cortical hypoperfusion akin to a cerebello–cerebral diaschisis. These abnormalities affect multiple supratentorial sites during the acute episode and appear to resolve as the mutism resolves.

Several cellular mechanisms may explain persisting neurocognitive impairment in childhood survivors. Radiotherapy to the brain adversely affects neural stem cell proliferation and neurogenesis in the rodent hippocampal subgranular zone. The precise functional contribution of postnatal neurogenesis to human cognition is not yet clear, but it is likely that the impact of radiotherapy is functionally relevant in humans too, given key hippocampal roles in modulating learning, memory, and emotion. A second vulnerable cell population may be oligodendrocyte precursor cells (OPCs). These cells generate oligodendrocytes which ensheath neurons in myelin. They support the biologic integration and normal functioning of white matter tracts which may be critical to cognition. Animal and human studies suggest that OPCs are vulnerable to the effects of chemotherapy.

Genetic studies provide some preliminary evidence that children with a history of cranial radiotherapy and COMT gene heterozygosity (Val/Met) may have better working memory than patients with COMT gene homozygosity (Met/Met or Val/Val). Structural candidates include supratentorial white matter tract changes with loss of volume and atrophy of cortical and subcortical structures correlating to poorer functioning. Meanwhile functional magnetoencephalography (MEG) data suggest that loss of high-frequency resting-state and task-oriented gamma oscillations may be associated with greater cognitive impairment and slower processing speeds in children undergoing radiotherapy. The clinical utility of these and other recent discoveries remains to be shown.

DIAGNOSIS AND CLINICAL FEATURES. Posterior fossa syndrome consists of cognitive, emotional, and behavioral disturbances characteristically presenting in the days following cerebellar surgery. Mutism can be present to varying degrees (milder cases may be dysarthric) but is often complete. Patients may have shown similar, if less marked, symptomatology preoperatively. Whether such preoperative abnormalities might indicate a higher risk of developing the post-surgical syndrome is currently unknown.

The PFS typically moves from complete mutism, through a period of cerebellar dysarthria, to resolution. The few case studies or small series charting its recovery suggest that this period can often be measured in weeks, rather than days or months. Even after 1 year however, many children who have recovered from PFS may remain cognitively impaired in multiple domains. Whether the PFS is a risk factor for even greater cognitive impairment than normal for adult survivors is currently unclear.

In brain tumor survivors, psychiatrists should carefully consider the contribution of subtle cognitive impairments which can persist for many years. These impairments are often dysexecutive, with working memory, attention, and processing speed commonly lower than normal levels for the patient's age and developmental stage. Deficits may become more apparent with age.

DIFFERENTIAL DIAGNOSIS. Children presenting acutely with an apparent PFS should receive an MRI brain scan to rule out post-operative bleeding or cerebral edema. In the longer term the psychiatric differential will depend on the particular syndrome presented by the individual patient. Considering a general syndrome of subtle cognitive dysfunction in young adult, treatable conditions to be excluded may include MDD, anxiety, and alcohol or drug use or dependence. Medical causes to exclude will depend on the specific presentation, but the general health of these patients is often poor. In particular, hypertension, diabetes, and endocrine abnormalities secondary to radiation-induced pituitary dysfunction should be excluded. Some forms of chemotherapy can cause deafness which may contribute to psychosocial or emotional difficulties. It may help to quantify cognitive deficits through formal neuropsychological assessment.

COURSE AND PROGNOSIS. The literature on the neuropsychiatric sequelae of childhood brain tumor is richer than it is for adults, both in the number of longitudinal studies and in the broader emphasis on psychosocial outcomes. It is clear that many childhood brain tumor survivors have considerable impairments of cognitive and emotional functioning. Attention, working memory, and processing speed may all remain affected. These impairments may manifest as social, educational, and vocational disability. Reading, spelling, and arithmetic may all be affected. Impaired emotional functioning with increased anxiety is seen in some children with some evidence of defective cognitive processing of emotions. Behavioral problems may manifest at school including aggression and possibly a higher level of somatization. Peer relationships suffer accordingly and adaptive and developmentally important behaviors such as communication and socialization are often adversely affected. Patients may report cognitive and behavioral difficulties that do not fully meet diagnostic thresholds, suggesting a large hidden morbidity.

In late adolescence and adulthood, the general picture is similar. Medulloblastoma survivors are highly likely to need some form of accommodation for acquired impairments in learning skills. The physical health of this group can be very poor. Hearing impairment, secondary cancers, diabetes, hypertension, and endocrinopathies may be common. Many patients are likely to remain dependent on parental income into their fourth decade. One-third of adult survivors of posterior fossa tumors display clinically significant levels of apathy. Themes of impaired identity formation, social isolation, difficulty securing financial independence, and limited career options are common in young adult survivors generally. Completion of higher education is relatively rare. Unemployment is often the norm, although by no means universal and if there is a glimmer of silver lining it is that despite an often-formidable array of psychosocial disadvantages, most adult patients will be functionally independent in daily life.

TREATMENT. Treatment of Posterior Fossa Syndrome is mostly supportive. In cases that do not recover as expected, a trial of the hypnotic Zolpidem may be of value. Families should be informed of the often gradual process of recovery, and where needed should be supported in making arrangements for their child returning to school. The condition can have a considerable impact on cognitive skills at a critical period for educational and social development, so detailed neuropsychological testing is valuable.

Treating chronic, biologically mediated, multimodal impairments that present in the longer term is difficult. A person-centered, problem-focused, and realistic approach is recommended. In adult survivors, neuropsychological assessment will help inform treatment plans ideally developed in a collaboration of psychology, psychiatry, the patient, and carers. Other professions including occupational therapy, physiotherapy, and dietetics may be involved. If the neuro-oncology team is still following the patient up they should be consulted; they will know the patient better than nearly any other professional service. Recommendations and adaptations should be feasible to implement.

Specifically, cognitive rehabilitation—structured exercises to boost cognitive skills—has some evidence for efficacy in childhood survivors. Improved executive and memory skills are seen particularly in the domains of attention, processing speed, cognitive flexibility, and declarative memory. As with adults, the “real-life” impact of these interventions is largely unknown: the literature consists mainly of small pilot studies that did not examined impact on broader educational and social attainment.

Empirical medication trials can be considered for neurocognitive impairment in survivors. There is some evidence that the psychostimulant methylphenidate may improve attention and behavior, in the short term, in childhood survivors of brain tumor. Boys, those of older age, and a higher baseline IQ may respond better. It has not been consistently shown to improve educational performance. The acetylcholinesterase inhibitor donepezil has limited evidence of efficacy for executive and memory impairments. It is important to stress that evidence underpinning drug treatment is, again, limited in this situation. A therapeutic trial may form part of a holistic treatment package, but given the empirical nature it should be time-limited, reviewed regularly, and discontinued if unsuccessful. Primary psychiatric syndromes arising in adult survivors should be treated as outlined above.

As well as treating established dysfunction, researchers are exploring strategies that may help minimize neurologic damage at the time of primary treatment. In particular, hippocampal-sparing radiotherapy schedules have recently been associated in both pediatric and adult brain tumor patients with better preservation of cognitive functioning. A full understanding of the clinical scope and limitations of hippocampal-sparing radiotherapy will require further study, but they hold promise for the future.

FUTURE DIRECTIONS

Clinical

There is increasing awareness of the potential for benefit of multidisciplinary neurorehabilitation for adult and pediatric brain tumor patients. Among the wide range of patients with differing types of primary brain tumor, those with persisting disability but a medium- to long-term life expectancy may derive considerable benefit. The evidence in support of multidisciplinary interventions is largely observational and therefore at a high risk of bias. Available studies tentatively suggest that multidisciplinary rehabilitation for brain tumors can be effective, particularly for improving some functional outcomes, vocational skills, and quality of life. It is difficult to design and successfully conduct trials that provide integrated care to such a complex population. A recent U.K. study has however shown the feasibility of delivering multimodal lifestyle coaching interventions to highly fatigued brain tumor patients. This study is, notably, the first to selectively target highly impaired patients in an attempt to modify aspects of quality of life—an important nettle to grasp when one seeks evidence of how to treat highly impaired patients in clinic.

A strong case can be made that psychiatrists have a critical role to play in growing this evidence base. Psychiatrists, especially consultation-liaison and neuropsychiatrists, have sophisticated transferable professional skills that are relevant to both neuro-oncology and rehabilitation. These skills include experience working in a multidisciplinary environment; the mindset, technical abilities, and time required to conduct comprehensive and holistic person-centered assessments; and knowledge of how to manage chronic and complex illnesses affecting vulnerable individuals. They also have direct expertise in the assessment and management of the many cognitive and psychiatric problems which very commonly affect these patients. They may have experience of developing and delivering behavioral treatment plans.

Despite the strong case for involving psychiatry in neuro-oncology, the current clinical reality in most centers is that neuro-oncology, neurorehabilitation, and neuropsychiatric services remain separate. There are exceptions but mostly these systemic divisions are a natural consequence of separate service evolution. They create practical barriers to clinical integration. Resistance to the idea that brain cancer patients can be meaningfully rehabilitated is common, if almost certainly wrong. One consequence is that psychiatric diagnoses may be missed in surviving patients. Skilled and caring clinicians nonetheless may think something like “I’d be depressed if I had a brain tumor”; patients themselves may be reluctant to “bother the doctor” with emotional symptoms; and the pressure of time in a busy neuro-oncology clinic may well mitigate against a thorough exploration of psychiatric symptoms. One solution to these problems would be for psychiatrists themselves to recognize and make the case for their skills in managing brain tumor patients. There is a clear need for psychiatry, both in established acute neuro-oncology, and in emerging rehabilitation neuro-oncology teams.

Research

The presence of a brain tumor distorting and destroying critical cerebral pathways poses questions for traditional diagnoses of some kinds of mental disorder. Part of the challenge therefore lies in testing and where necessary refining diagnostic criteria. Large, longitudinal studies will inform better prognostication. The relative rarity of brain tumors should promote multicenter collaborations. Research is also needed into mechanisms that contribute to the risk, onset, maintenance, and resolution of brain tumor-associated neuropsychiatric disorders. The unique clinical context means these mechanisms should be studied in brain tumor patients specifically. A third high-value area will be in finding effective treatments for neuropsychiatric disorders, by RCTs conducted in brain tumor patients specifically. These are by no means easy to conduct in such an impaired population, but will be essential to be sure that treatments work in the context of the many difficulties of having a brain tumor.

ACKNOWLEDGEMENTS

With thanks to Dr. Robin Grant, Prof. Alan Carson, and the Edinburgh Centre for Neuro-Oncology for many years of support.

Further Readings

Alper K, Schwartz KA, Kolts RL, Khan A. Seizure incidence in psychopharmacological clinical trials: an analysis of Food and Drug Administration (FDA) summary basis of approval reports. *Biol Psychiatry*. 2007;62:345–354.

Beevers Z, Hussain S, Boele FW, Rooney AG. Pharmacological treatment of depression in people with a primary brain tumour. *Cochrane Database Syst Rev*. 2020;7(7):CD006932.

Benros ME, Laursen TM, Dalton SO, Mortensen PB. Psychiatric disorder as the first manifestation of cancer: a 10-year population-based study. *Int J Cancer*. 2009;124:2917–2922.

Buday J, Albrecht J, Mareš T, et al. Brain tumors and electroconvulsive therapy: a literature overview of the last 80 years. *Front Neurol*. 2020;11:723.

Carroll C, Watson P, Spoudeas HA, et al. Prevalence, associations, and predictors of apathy in adult survivors of infantile (<5 years of age) posterior fossa brain tumors. *Neuro Oncol*. 2013;15:497–505.

Catsman-Berrevoets C, Patay Z. Cerebellar mutism syndrome. *Handb Clin Neurol*. 2018;155:273–288.

Day J, Yust-Katz S, Cachia D, et al. Interventions for the management of fatigue in adults with a primary brain tumour. *Cochrane Database Syst Rev*. 2016;4:CD011376.

Gehring K, Sitskoorn MM, Gundy CM, et al. Cognitive rehabilitation in patients with gliomas: a randomized, controlled trial. *J Clin Oncol*. 2009;27:3712–3722.

Gibson AW, Graber JJ. Distinguishing and treating depression, anxiety, adjustment, and post-traumatic stress disorders in brain tumor patients. *Ann Palliat Med*. 2021;10:875–892.

Huang J, Zeng C, Xiao J, et al. Association between depression and brain tumor: a systematic review and meta-analysis. *Oncotarget*. 2017;8(55):94932–94943.

Jütten K, Mainz V, Gauggel S, et al. Diffusion tensor imaging reveals microstructural heterogeneity of normal-appearing white matter and related cognitive dysfunction in glioma patients. *Front Oncol*. 2019;26:536.

Keng A, Stewart DE, Sheehan KA. Neuropsychiatric symptoms after brain tumor resection in children and adolescents: a scoping review. *J Acad Consult Liaison Psychiatry*. 2022;63(2):110–118.

Koutsarnakis C, Neromyliotis E, Komaitis S, et al. Effects of brain radiotherapy on cognitive performance in adult low-grade glioma patients: a systematic review. *Radiother Oncol*. 2021;160:202–211.

Knudsen-Baas KM, Johannesen TB, Myklebust TÅ, et al. Antiepileptic and psychiatric medication in a nationwide cohort of patients with glioma WHO grade II-IV. *J Neurooncol*. 2018;140:739–748.

Leppla I, Fishman D, Kalra I, Oldham HA. Clinical approach to personality change due to another medical condition. *J Acad Consult Liaison Psychiatry*. 2021;62(1):14–21.

Monje M, Mizumatsu S, Fike JR, Palmer TD. Irradiation induces neural precursor-cell dysfunction. *Nat Med*. 2002;8:955–962.

Noll KR, Sullaway CM, Wefel JS. Depressive symptoms and executive function in relation to survival in patients with glioblastoma. *J Neurooncol*. 2019;142:183–191.

Ostrom QT, Patil N, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2013–2017. *Neuro Oncol*. 2020;22(12 Suppl 2):iv1–iv96.

Ownsworth T, Chambers S, Damborg E, Casey L, Walker DG, Shum DH. Evaluation of the making sense of brain tumor program: a randomized controlled trial of a home-based psychosocial intervention. *Psychooncology*. 2015;24(5):540–547.

Pace A, Dirven L, Koekkoek JAF, et al. European Association for Neuro-Oncology (EANO) guidelines for palliative care in adults with glioma. *Lancet Oncol*. 2017;18(6):e330–e340.

Padovani L, André N, Constine LS, Muracciole X. Neurocognitive function after radiotherapy for paediatric brain tumours. *Nat Rev Neurol*. 2012;8:578–588.

Rooney AG, Brown PD, Reijneveld JC, Grant R. Depression in glioma: a primer for clinicians and researchers. *J Neurol Neurosurg Psychiatry*. 2014;85:230–235.

Rooney AG, McNamara S, Mackinnon M, et al. Frequency, clinical associations, and longitudinal course of major depressive disorder in adults with cerebral glioma. *J Clin Oncol*. 2011;29:4307–4312.

Rumalla K, Lin M, Orloff E, et al. Effect of comorbid depression on surgical outcomes after craniotomy for malignant brain tumors: a nationwide readmission database analysis. *World Neurosurg*. 2020;142:e458–e473.

Sharkey CM, Espeleta HC, Traino KA, et al. Psychological adjustment outcomes among pediatric brain tumor survivors: a meta-analysis. *Pediatr Blood Cancer*. 2020;67(10):e28644.

Shi C, Lamba N, Zheng LJ, et al. Depression and survival of glioma patients: a systematic review and meta-analysis. *Clin Neurol Neurosurg*. 2018;172:8–19.

Smithson EF, Phillips R, Harvey DW, Morrall MC. The use of stimulant medication to improve neurocognitive and learning outcomes in children diagnosed with brain tumours: a systematic review. *Eur J Cancer*. 2013;49:3029–3040.

Starkstein SE, Pearlson GD, Boston JD, Robinson RG. Mania after brain injury. A controlled study of causative factors. *Arch Neurol*. 1987;44:1069.

Tibbs MD, Huynh-Le M-P, Reyes A, et al. Longitudinal analysis of depression and anxiety symptoms as independent predictors of neurocognitive function in primary brain tumor patients. *Int J Radiat Oncol Biol Phys*. 2020;108:1229–1239.

2.4 Neuropsychiatric Aspects of Epilepsy and Psychogenic Nonepileptic Seizures

Mario F. Mendez, M.D., Ph.D., Leila Parand, M.D., and Gaston Baslet, M.D.

NEUROPSYCHIATRIC ASPECTS OF EPILEPSY

Clinicians have reported associations between epilepsy and psychiatric disorders since antiquity. In modern times, however, many clinicians are unfamiliar with these associations and the unique challenges involved in their management. The development of new antiepileptic and psychiatric therapies and novel neuroimaging techniques makes understanding the association between epileptic seizures and psychopathology increasingly important. Psychiatrists and neurologists are increasingly investigating the neuropsychiatric aspects of epilepsy, as exemplified by the amount of new research and publications in this field.

Epidemiologic data support an increased risk for psychiatric comorbidity among epilepsy patients as compared to nonepileptic patients. The most established association is between epilepsy and depression or dysthymia, but a range of psychopathology occurs in one-fourth or more epilepsy patients. For example, epilepsy is associated with anxiety disorders, somatic symptom disorder, personality changes, hyposexuality, and, perhaps most dramatically, several forms of psychosis. These behaviors and others have different potential relationships to the ictus or seizure itself. Psychiatric manifestations may result directly from ictal discharges (e.g., psychic auras), as peri-ictal phenomena (e.g., postictal confusion), interictally or between seizures (e.g., interictal psychosis), or with a variable and less-established relationship to the seizure discharges (e.g., most mood disorders).

Psychiatric comorbidity has a serious impact on the quality of life and well-being of patients with epilepsy. Psychiatrists and neurologists need to maximize positive (e.g., mood-stabilizing) and minimize negative psychotropic effects of antiepileptic drugs (AEDs), consider the seizure threshold lowering effects of some psychotropic medications, and monitor the potential interaction of antiepileptic and psychotropic drugs. Before committing patients to antiepileptic treatment, psychiatrists and neurologists must also be able to distinguish epileptic seizures from other spells, particularly psychogenic nonepileptic seizures.

Definition

Seizures are sudden, involuntary behavioral events associated with excessive or hypersynchronous electrical discharges in the brain. The seizure itself is known as the *ictus.* The *interictal period* refers to the period between the postictal abnormalities and the next ictus, and the *peri-ictal period* refers to the period just before or after the ictus and is applied when there is insufficient information to know when the ictus ends or begins. Seizures can be symptomatic of a transient or physiologic disturbance. They may also have a genetic basis or can be caused by an acquired insult such as trauma, stroke, inflammation, or neoplasm. Symptomatic seizures are not equivalent to epilepsy. *Epilepsy* is a disease of the brain most commonly characterized as having at least two unprovoked or reflex seizures occurring greater than 24 hours apart. *Status epilepticus* is prolonged (>5 minutes) or two or more seizures in a 30-minute period without an intervening recovery.

In epilepsy, abnormal electrical discharges are due to hyperexcitable neurons with sustained postsynaptic depolarization. Proposed mechanisms for this sustained depolarization include changes in ionic conductance, decreased gamma-aminobutyric acid (GABA) inhibition of cortical excitability, and increased glutamate-mediated cortical excitation. In animals, alumina-induced membrane changes alter the ratio of intracellular to extracellular ionic concentrations and result in abnormal neuronal firing. AEDs, such as phenytoin (Dilantin), carbamazepine, and valproate, reduce this repetitive firing through effects on sodium channels. Ethosuximide (Zarontin) works through the blockage of calcium currents. Penicillin-induced cortical injury causes seizures through decreased GABA inhibition. Barbiturates and benzodiazepines may reduce seizures by enhancing GABA receptor current, and valproate may also block GABA catabolism. Kainic acid, a glutamate agonist, induces seizures through increased synaptic action at its *N*-methyl-D-aspartate (NMDA) receptors. Work is underway on potential AEDs that may work through inhibition of this excitatory receptor mechanism.

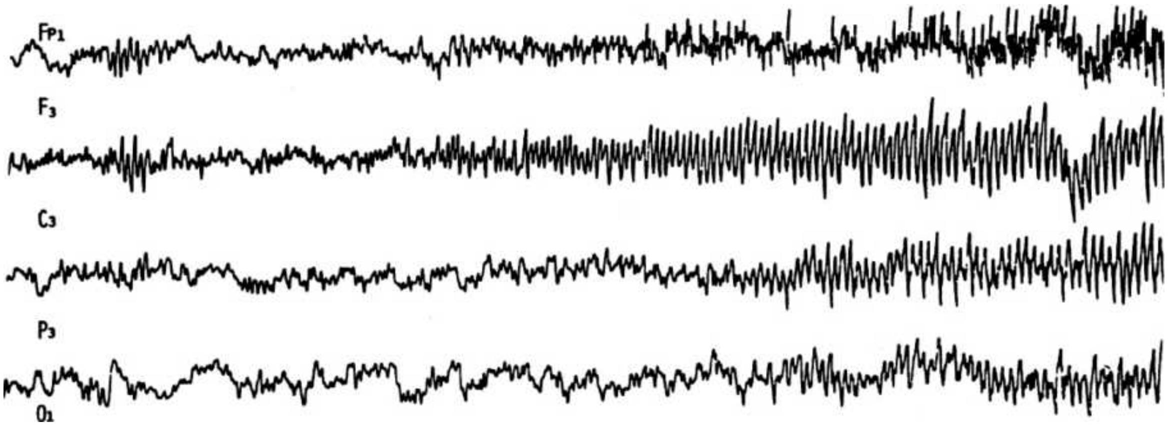

FIGURE 2.4–1. Electroencephalogram demonstrating the focal onset of seizure discharges from the left frontotemporal region that are consistent with the onset of focal seizures with impaired awareness.

The electroencephalogram (EEG) is a surface recording of brain wave activity used in the evaluation of seizures. Basic waves include normal waking alpha waves (8 to 13 Hz), which are most prominent over the occipital region, high-frequency beta waves (greater than 13 Hz), theta waves (4.0 to 7.5 Hz), and delta slowing (3.5 Hz or less). Seizures are manifest as multiple spikes or spike-and-wave discharges on the EEG (Fig. 2.4–1). A spike is a sharp transient with a duration of 20 to 70 milliseconds. Interictally, single spikes and other markers of abnormal electrical activity may be seen, often emanating from a temporal lobe.

History

In his book on epilepsy, *On the Sacred Disease,* Hippocrates (460 to 377 BC) attacked the prevailing belief that those afflicted with epilepsy were possessed by gods or goddesses. He proposed that epilepsy was a brain disease caused by the blockage by phlegm of air-carrying vessels to the brain. Despite this initial view, throughout most of human history, epilepsy constituted demonic possession or the accumulation of bad humors, and attempts at exorcism involved trephination, cautery of the back of the skull, diuretics, emetics, bloodletting, purging, sweating, and even intercourse to release sperm. In the 18th century, the first so-called scientific treatise on epilepsy since ancient times attributed seizures to masturbation. By happenstance, bromides, which were introduced to diminish libido and masturbation, proved to be the first successful antiepileptic medication. With the development of effective AEDs and the introduction of EEG, physicians have come full circle to Hippocrates' belief that epilepsy is rooted in organic brain disease.

The purported association of epilepsy with behavioral disorders also dates to antiquity. The brain was the seat of "the falling sickness" and of "madness," and both were related to phlegm. With demonic possession as a form of punishment, unusual or abnormal behaviors became associated with seizure patients, even during their seizure-free periods. At the turn of the 20th century, the psychiatric writings of Emil Kraepelin emphasized that epilepsy patients possessed personality changes and a predisposition to psychosis. With a greater understanding of the physical basis of epilepsy, many clinicians sought to protect epilepsy patients from the demonic stigma of their disease; in their view, psychiatric problems resulted from the psychosocial difficulties associated with having seizures rather than any unique relationship of epilepsy with psychiatric illness. The current age was initiated by the definition of temporal lobe epilepsy and the concept of a physiologic disturbance in the limbic or emotional brain.

Nosology

The classification and terminology of seizures were revised by the International League Against Epilepsy in 2017 to reflect updated advances in the understanding of epilepsy (Table 2.4–1). Seizures are classified based on the mode of etiology and mode of onset. Etiology can be grouped into genetic, structural–metabolic, and unknown. The mode of onset can be *generalized* or *focal.* Generalized onset seizures are those with an initial widespread bihemispheric involvement, and focal onset seizures are those that emanate from a focus limited to part of one hemisphere. In adults, most generalized seizures are *tonic–clonic* seizures (grand mal seizures or convulsions) and are characterized by an abrupt loss of consciousness with tonic rigidity followed by a synchronous, clonic release. Focal seizures (previously partial or localized seizures) can occur with or without impairment of consciousness or awareness. Simple partial seizure is an old term for focal onset without impairment of consciousness, whereas complex partial or dyscognitive seizure is an old term used for seizures with focal onset with impaired awareness. Focal seizures produce isolated motor, sensory, autonomic, psychic, or mixed symptoms in a clear sensorium. "*Auras*" refer to focal seizures without impaired awareness that involve subjective sensory or psychic phenomena without observed motor or autonomic changes. Focal seizures with impaired awareness usually manifest as motionless staring combined with simple automatisms or automatic motor activity and last approximately 1 minute. Focal seizures with impaired awareness can evolve to become a tonic, clonic, or tonic–clonic seizure. Finally, there is a form of generalized onset seizures, *absence* (petit mal) seizures, which occur less commonly in adults and manifest as brief lapses of consciousness. Absence seizures differ from focal seizures with impaired awareness in that they are generalized in the mode of onset. They also differ in that they are short (10 seconds in length) and repetitive; lack auras, postictal confusion, or complex automatisms; and have characteristic generalized two to four counts per second spike-and-wave discharges on EEG.

Table 2.4–1.
2017 ILAE Classification of Seizure Types

Focal Onset Seizures	
Motor	**Nonmotor**
Awareness[a] Impaired awareness[b]	Awareness[a] Impaired awareness[b]
Automatisms	Autonomic
Atonic*	Behavioral arrest
Clonic	Cognitive
Epileptic spasms*	Emotional
Hyperkinetic	Sensory
Myoclonic	
Tonic	
Focal to bilateral tonic–clonic[c]	Focal to bilateral tonic–clonic[c]
Generalized Onset Seizures	
Motor	**Nonmotor**
Tonic–clonic	Typical
Clonic	Atypical
Tonic	Myoclonic
Myoclonic	Eyelid myoclonia
Myoclonic–tonic–clonic	
Myoclonic–atonic	
Atonic	
Epileptic spasms	
Unknown Onset Of Seizures	
Motor	**Nonmotor**
Tonic–clonic	Behavior arrest
Epileptic spasms	
Unclassified Seizures^	

[a] Previously, simple partial seizure.

[b] Previously, corresponds to “complex partial seizure” or “dyscognitive” seizure.

[c] Previously, partial seizure with secondary generalization.

*Degree of awareness is usually not specified.

^Due to inadequate information or inability to place in other categories.

Adapted with permission of John Wiley & Sons from Fisher RS, Cross JH, French JA, Higurashi N, et al. Operational classification of seizure types by the International League Against Epilepsy: Position paper of the ILAE Commission for Classification and Terminology. *Epilepsia*. 2017;58(4):522–530. doi: 10.1111/epi.13670.

Most adults with epilepsy, regardless of seizure type, have an epileptic focus in the temporal lobe, and many generalized tonic–clonic seizures are secondarily generalized from a temporal lobe focus without a preceding focal seizure with impaired awareness.

Epidemiology

Seizure disorders are common and usually have an early onset. Epilepsy affects over 45 million people worldwide and has a prevalence of at least 0.62% and an annual incidence of approximately 0.06%. The overall incidence is high in the first year, drops to a minimum in the third and fourth decades of life, then peaks again after age 75 years. More than 75% of patients have their first seizure before 18 years of age, and 12% to 20% have a familial incidence of seizures. Among adults, the most common seizures are focal with impaired awareness and generalized tonic–clonic seizures. Focal onset is most frequent in the temporal lobes. The most common cause of new-onset seizures in adults is cerebrovascular disease.

Etiology

Diagnosing the specific cause can have important implications for management. Previous descriptions of idiopathic epilepsy are increasingly found to have evidence of specific genetic defects. Other frequent secondary causes include trauma in the third and fourth decades of life, neoplasms in the fifth and sixth decades of life, and cerebrovascular disease in the elderly. Although some focal seizures with impaired awareness originate from the frontal or temporal neocortex and other areas, at least two-thirds of these focal seizures and generalized tonic–clonic seizures originate from the mediobasal temporal limbic structures (hippocampus, amygdala, and parahippocampal gyrus).

Psychopathology. Epidemiologic studies from communities, psychiatric hospitals, and epilepsy clinics report a 20% to 60% prevalence of psychiatric problems among epilepsy patients. Psychiatric disorders often occur at rates two- to threefold or higher than in the general population without epilepsy. Epilepsy patients are prone to psychosis, mood disorders, especially depression, anxiety disorders, somatic symptom disorders, personality disorders, hyposexuality, and dissociative symptoms. These problems are approximately equally divided between those that occur actually or peri-ictally and those that occur interictally or are variably related to the ictus. The percentage of epilepsy patients in psychiatric hospitals is also higher than the general prevalence of epilepsy and ranges from 4.7% of all inpatients in a British psychiatric hospital to 9.7% in a U.S. Veterans Affairs psychiatric facility. Among patients attending epilepsy clinics, approximately 30% have a prior psychiatric hospitalization, and 10% to 20% are on at least one psychotropic drug. Furthermore, epidemiologic studies indicate increased interictal psychopathology among head-injured patients with epilepsy compared to head-injured patients without epilepsy. There exists a bidirectional relationship between psychiatric disorders and epilepsy, such that patients with epilepsy are more likely to develop depression than the general population, and patients with depression, anxiety, and psychosis are more likely to develop epilepsy. Despite criticisms of selection bias, these studies constitute a broad spectrum of sources that indicate greater overall psychopathology among epilepsy patients.

Do epilepsy patients have greater psychopathology than other similarly impaired patients? If this were so, then it would suggest that the psychopathology is of biologic origin rather than a less specific reaction to chronic disease. Although disputed by some investigators, several studies report more psychopathology among epilepsy patients than among patients with chronic diseases that do not directly affect the brain. Furthermore, the pattern of behavioral changes in seizure patients appears specific to epilepsy. For example, on the Minnesota Multiphasic Personality Inventory (MMPI) 2, patients with epilepsy score higher on the schizophrenia scale and paranoia scale than patients with other neurologic disabilities. In addition to a potential genetic overlap between epilepsy and psychiatric conditions, the neural networks for epilepsy and for psychoses appear to overlap considerably, potentially explaining some of the comorbidity.

Many studies found a special relationship to psychopathology in patients whose seizures emanate from mediobasal temporal lesions.

Psychiatric disturbances, primarily psychosis and personality disorders, are two to three times more common in patients with focal seizures with impaired awareness, most of temporal lobe origin, compared to those with generalized tonic–clonic seizures. Other studies have failed to find a difference in psychopathology based on focal or generalized seizure onset. Major depression, dissociative symptoms, obsession, and compulsion were also seen more frequently in temporal lobe epilepsy compared to generalized epilepsy. Psychic auras from the temporal lobe, particularly if associated with negative feelings (e.g., *jamais vu* and fear), predispose to psychosis or personality disorders.

Psychosis. Psychosis is a specific psychiatric disorder associated with epilepsy. The lifelong prevalence of all psychotic disorders among epilepsy patients ranges as high as 12%. In a follow-up of 100 children with focal seizures with impaired awareness for as long as 30 years, of the 87 patients who survived to adulthood and who did not have intellectual impairment, 9 (10%) experienced a psychotic illness. In temporal lobectomy studies, in which there is the surgical removal of an epileptic focus in the temporal lobe, psychosis occurred in 7% to 8% of patients, even long after the seizures were arrested. The data from a meta-analysis also indicate that the prevalence of psychosis in temporal lobe epilepsy was 7%. That percentage is higher than in the other types of epilepsy (5.6%) and represents approximately a five- to eightfold greater risk of psychosis for epilepsy patients than for the general population.

In some patients, autoantibodies may play a role in psychotic symptoms in epilepsy. Data from studies found that specific autoantibodies, for example, anti-NMDA, anti–voltage-gated potassium channel (VGKC), and anti–GABA-B receptor ($GABA_B$), can be identified in up to 10% of patients with epilepsy and psychosis. Moreover, those patients who have autoimmune encephalitides usually develop convulsive seizures during the course of the illness.

Regarding lateralization and neuropathology, results from many studies suggest an association of a left-sided focus and hippocampal sclerosis with psychosis. Although conclusions derived from surface EEG recording are open to criticism, depth recordings of presurgical patients show that twice as many patients with left temporal lesions have psychosis. Data from studies using voxel-based morphometry analysis of magnetic resonance imaging (MRI) also show that patients with temporal lobe epilepsy and psychosis have more severe gray matter loss in the entire temporal lobe as well as in the left hippocampus, insulae, anterior cingulum, and left inferior parietal lobule. The results from white matter data further demonstrate white matter reduction in widespread brain regions. Finally, positron emission tomography (PET) scans and single-photon emission computed tomography (SPECT) scans may show predominant left temporal hypometabolism among epilepsy patients with psychosis.

Depression. The prevalence of depression in different studies varies from 7.5% to 34% of patients with epilepsy. This discrepancy is attributed to various diagnostic measurements and different definitions of depressive disorders across studies. Major depressive episodes can be found in up to 20% of the patients. Those with focal seizures with impaired awareness originating from the mesial temporal lobe and those with poor seizure control are more likely to have mood disorders. Psychological studies also suggest a greater incidence of self-criticism and depression among epilepsy patients with a left hemisphere focus. Patients with focal seizures of temporolimbic origin also have a higher incidence of depression than patients with other types of seizure disorders. Specific instruments are available for assessing depression among epilepsy patients, such as the Neurological Disorders Depression Inventory for Epilepsy (NDDI-E).

Table 2.4–2.
Psychic Auras

Type	Symptoms	Probable Source
Dysphasic[a]	Nonfluent	Left perisylvian language areas
	Impaired comprehension	
Dysmnesic	*Déjà vu, déjà vécu, déjà pensé, déjà entendu, jamais vu,* etc., prescience, illusion of memory	Mesobasal temporal,[b] especially on right
Cognitive	Dreamy state, altered time sense, derealization, depersonalization	Mesobasal temporal and temporal neocortex
	Forced thinking, forced actions, and altered or obscure thoughts	Frontal association cortex
Affective	Fear, anxiety, apprehension, depression, pleasure, displeasure	Mesobasal temporal and temporal neocortex
Illusions[c]	Macropsia, micropsia, teleopsia, movement, metamorphopsia, increased color intensity, increased stereopsis intensity	Lateral superior temporal neocortex, especially on right for visual illusions
Hallucinations[c]	Structured, hallucinatory remembrances, autoscopy	Mesobasal temporal and temporal neocortex

[a] Does not include speech arrest or simple vocalizations.

[b] Includes hippocampus, amygdala, and the parahippocampal gyrus.

[c] Includes interpretive (size, motion, shape, and stereopsis) or experiential (elements of past experience or involvement).

Other Behaviors. The prevalence of other specific behavioral disorders among patients with epilepsy is less well established. However, there is evidence to suggest that personality disorders, suicidal behavior, anxiety disorder, and hyposexuality are more prevalent among epilepsy patients than among those without seizure disorders.

Pathophysiology. The relationship between seizures, psychiatric syndromes, and the mediobasal temporal lobes implies that many behavioral changes are more than psychological reactions to the psychosocial stressors of epilepsy. Stimulation and ablation studies in humans and animals link temporal limbic structures to emotion and behavior.

For example, temporal limbic stimulation in a person evokes psychic auras (Table 2.4–2) and automatisms, and amygdala stimulation and ablation in animals result in aggression or placidity. Moreover, psychotic behavior in cats occurs when their limbic structures undergo kindling (i.e., the repeated application of epileptogenic agents to induce lasting behavioral changes).

There are several potential causes of psychiatric disturbances in epilepsy (Table 2.4–3). First, the pathology itself could be the source of seizures and behavioral changes. Left hemisphere and temporal lobe lesions may be associated with interictal psychosis, and psychosis in epilepsy may be particularly frequent if there is specific underlying pathology, such as ventricular enlargement. Psychotic disorders may be more common with temporal dysplasia or neurodevelopmental abnormalities and depression with mesial temporal sclerosis. Second, ictal or subictal epileptiform activity may promote behavioral changes by facilitating distributed neuronal connections, increasing limbic–sensory associations, or changing the overall balance between excitation and inhibition. This may occur not only with temporal lobe seizures but also with those that originate in the frontal lobes. Third, the absence or hyperactivity of function, such as the interictal hypometabolism observed on PET scans (Fig. 2.4–2), may lead to depression or other interictal behavioral changes. For example, a reduction in cerebral blood flow in the left medial temporal region from SPECT may be associated with interictal psychosis. In contrast, auditory hallucination during a focal onset seizure without impairment of consciousness may demonstrate a cerebral hyperperfusion in the primary auditory cortex. Fourth, seizures may result in neuroendocrine or neurotransmitter changes, such as increased dopaminergic or inhibitory transmitters, decreased prolactin, increased testosterone, or increased endogenous opioids, all of which can affect behavior. Furthermore, neurobiologic factors may be potentiated by psychodynamic factors, such as feelings of helplessness, learned helplessness, dependency, low self-esteem, and the disruption of reality testing. In summary, the psychiatric manifestations of epilepsy are heterogeneous disorders with potentially different causes.

Table 2.4–3.
Proposed Relationships of Psychiatric Disturbances to Epilepsy

- Common neuropathology, genetics, or developmental disturbance
- Ictal or subictal discharges potentiate abnormal behavior
    - Kindling or facilitation of a distributed neuronal matrix
    - Changes in spike frequency or inhibitory–excitatory balance
    - Altered receptor sensitivity, e.g., dopamine receptors
    - Secondary epileptogenesis
- Absence of function at the seizure focus
    - Inhibition and hypometabolism surrounding the focus
    - Release or abnormal activity of remaining neurons
    - Dysfunction or downregulation of associated areas
- Neurochemical
    - Dopamine and other neurotransmitters
    - Endorphins
- Gonadotropins and other endocrine hormones
- Psychodynamic and psychosocial effects of living with epilepsy
    - Dependence, learned helplessness, low self-esteem, weak defense mechanisms
    - Disruption of reality testing
- Neurobiologic and psychodynamic factors potentiate each other
- Sleep disturbance
- Antiepileptic drug related

Diagnosis

The Diagnostic and Statistical Manual of Mental Disorders, 5th edition, text revision (DSM-5-TR) does not have an exclusive diagnostic entity for epilepsy-related psychiatric disorders. Each psychiatric syndrome must be judged by a physician whether it is a direct consequence of epilepsy or not. If the syndrome is clearly associated with epilepsy, it will fall under the "mental disorder due to another medical condition" diagnosis; if not, the diagnosis should be listed as a primary psychiatric disorder. Nevertheless, since the DSM removed the multiaxial system, all the psychiatric diagnoses have to be recorded alongside a diagnosis of epilepsy. The DSM-5 also encourages clinicians to perform a dimensional assessment, measuring the severity of specific psychiatric symptoms and recording the disability of patients.

The other way to categorize psychiatric behaviors associated with epilepsy is their relationship to the ictus or seizure discharges. These behaviors may occur as part of the ictus, peri-ictally, or during the interictal period (Table 2.4–4). A range of other behaviors appear to have some relation to the ictus but do not clearly fall into one of the former three categories.

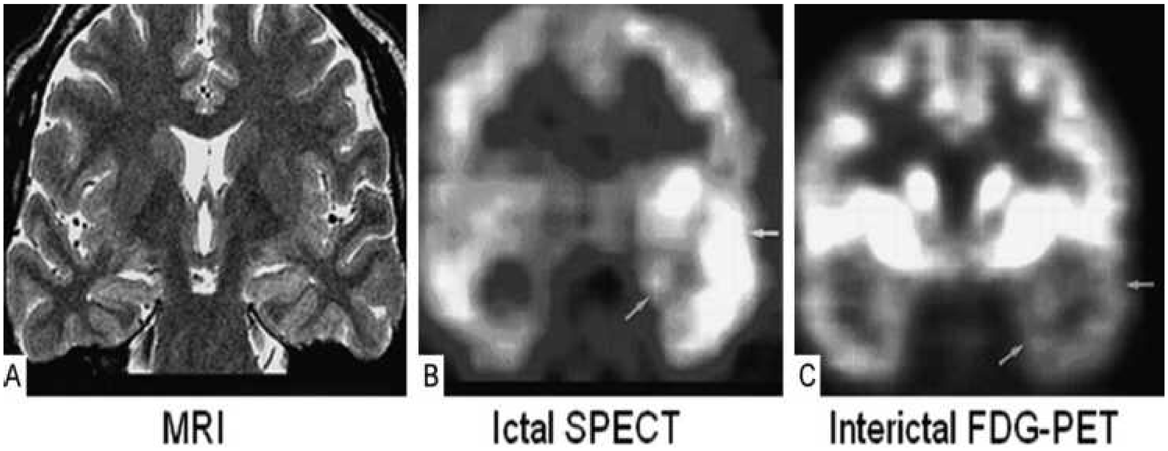

FIGURE 2.4–2. Neuroimaging from a 30-year-old man with focal seizures with impaired awareness of temporal lobe origin and ictal activity on invasive EEG. **A:** Coronal T2-weighted MR does not disclose abnormalities. **B:** Ictal single photon emission tomography image shows left temporal hyperperfusion. **C:** Fluorodeoxyglucose positron emission tomography confirms left temporal hypometabolism. The patient had a left temporal lobectomy and hippocampal sclerosis. (Reprinted with permission from Won HJ, Chang KH, Cheon JE, et al. Comparison of MR imaging with PET and ictal SPECT in 118 patients with intractable epilepsy. *AJNR Am J Neuroradiol.* 1999;20(4):593–599. Copyright © American Society of Neuroradiology.)

Ictal Features. Seizure discharges can produce semipurposeful automatisms and psychic auras, such as mood changes, derealization and depersonalization, and forced thinking. Ictal fear, which ranges from a vague apprehension to abject fright, has occurred without any other seizure manifestation, and ictal depression has extended days or longer after the seizure has passed. Some patients have pleasurable auras. Fyodor Dostoyevsky had “ecstatic auras” in which he felt in perfect harmony with the entire universe and “would give 10 years of this life, perhaps all of it, for a few seconds of such bliss.” The experience of epileptic derealization or depersonalization could impair reality testing. Another psychic aura is “forced thinking,” characterized by recurrent intrusive thoughts, ideas, or crowding of thoughts. Forced thinking must be distinguished from obsessional thoughts and compulsive urges. Some patients may experience vivid auditory and visual hallucinations during an epileptic discharge. Epilepsy patients with forced thinking experience their thoughts as stereotypical, out of context, brief, and irrational, but not necessarily as ego-dystonic.

Table 2.4–4.
Behavioral Disorders in Epilepsy

- **Ictal**
  - Ictal psychic symptoms (aura)
  - Nonconvulsive status: focal onset seizures without impaired awareness, focal onset seizures with impaired awareness, and periodic lateralizing epileptiform discharges
  - Catatonia
- **Peri-ictal** (includes prodromal, postictal, and mixed ictal)
  - Prodromal symptoms: irritability, depression, headache, etc.
  - Postictal confusion
  - Peri-ictal psychoses
  - Concomitant with increased seizure frequency
  - Concomitant with decreased seizure frequency
  - Postictal psychoses
- **Interictal**
  - Interictal psychosis and personality disturbances
  - Interictal psychosis (schizophreniform psychosis)
- **Other behavioral disorders associated with epilepsy**
  - Personality disorders
  - Gastaut–Geschwind syndrome
  - Behavioral disturbances variably related to ictus
  - Mood disorders (depression and mania)
  - Anxiety disorders including panic and posttraumatic stress disorder
  - Aggression and violence
  - Hyposexuality
  - Suicide

A 36-year-old right-handed man presented with frontal headaches and 5 years of focal seizures with impaired awareness. His seizures began with 15 seconds of a sense of “impending doom,” speech arrest, and orobuccal movements followed by 30 seconds of altered consciousness. At seizure onset, the patient felt forced to think the phrase “tell me yes.” The phrase was repeated several times without his being able to control it. Concomitantly, his mouth would open, and he would attempt to say the phrase but could utter only unintelligible sounds. The patient interpreted this phrase as a call for help. On examination, he had a mild memory deficit, normal language testing, a right facial droop, and brisk right-sided reflexes with a right Babinski sign. Neuroimaging revealed a left frontal mass lesion. EEGs showed amplitude attenuation and polymorphic delta in the left frontal area, and intraoperative electrocorticography disclosed polyspike and spike–wave discharges associated with impaired language from just below the lesion. The patient underwent subtotal resection of a 4.3- × 3- × 3-cm oligodendroglioma. Postoperatively, his forced thinking and seizures resolved, but he had nonfluent aphasia.

Cognitive disorders follow status epilepticus with focal or absence seizures. Recurrent or prolonged focal seizures may not result in alteration of consciousness or invariable abnormalities on EEG, and, if manifested by psychic auras, focal seizures may be difficult to distinguish from primary psychiatric disturbances. Status epilepticus from focal seizures with impaired awareness and absence of seizures results in prolonged alterations of responsiveness. With the addition of various ictal auras, focal status epilepticus (epilepsia partialis continua) can appear psychotic. Nonconvulsive status epilepticus can also present with immobility, waxy flexibility, and behavioral negativism resembling catatonia. Occasionally, EEGs and a therapeutic trial of antiepileptic medications may be the only way to distinguish behavioral disturbances due to nonconvulsive status epilepticus. Finally, recurrent interictal EEG complexes, known as *periodic lateralizing epileptiform discharges,* may also be associated with prolonged confusional behavior and focal cognitive changes.

A 68-year-old man had a left temporal–parietal hemorrhagic stroke. Initial fluent aphasia and right hemiparesis completely resolved, but he developed poststroke epilepsy. His seizures began with speech arrest and were followed by secondary generalization to tonic–clonic seizures. The postictal periods lasted days due to continued left-hemisphere periodic lateralizing epileptiform discharges. During these prolonged postictal periods, he was confused and placid and had a return of his aphasia. One year later, after achieving seizure control, the patient developed mania for the first time in his life. His mania was in a clear sensorium without a change in his neurologic examination or epileptiform activity on EEG. He did not sleep, had a flight of ideas, and had grandiose ideation, including beliefs that he was a three-star general, had killed Adolf Hitler, and was now a millionaire. He exposed himself to everyone, including his daughter, and inserted pencils up his penis because he believed that he needed catheterization. His psychosis lasted for 3 months until he had two generalized tonic–clonic seizures. Postictally, for 10 days, he remained placid, confused, and aphasic, with right beating nystagmus and periodic lateralizing epileptiform discharges maximal in the left temporal region (Fig. 2.4–3). With new antiepileptic medication, he returned to normal with a total resolution of his mania.

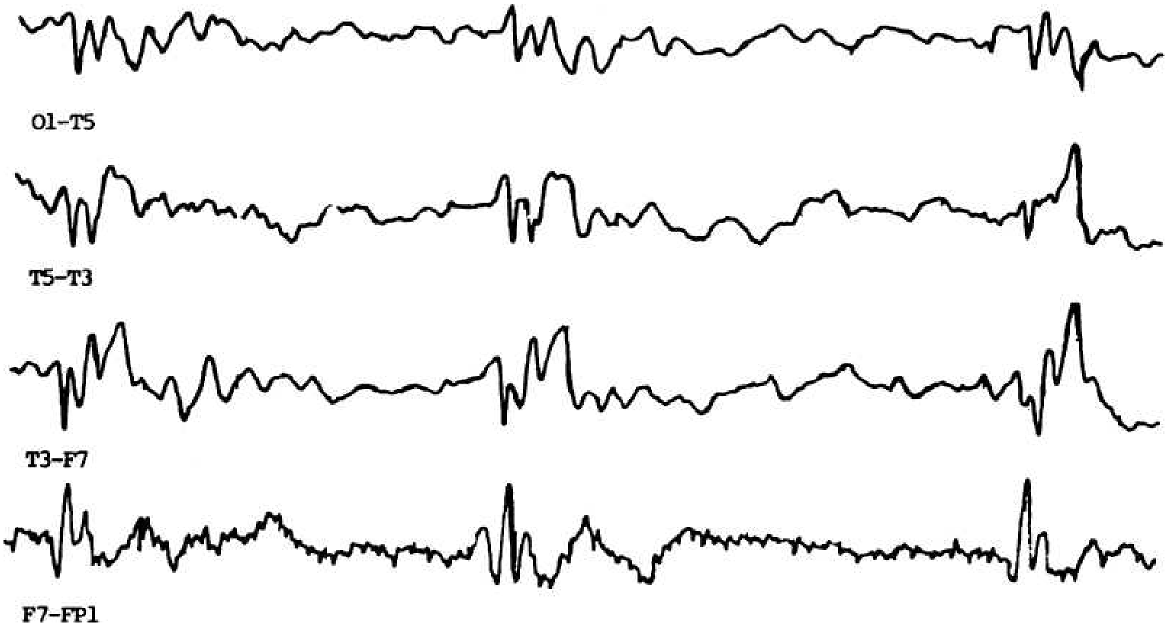

FIGURE 2.4–3. Periodic lateralizing epileptiform discharges. Left-sided electrodes per the International 10/20 System.

Peri-Ictal Features. Psychiatric disturbances can occur before seizures (prodromal), after seizures (postictal), or during intermittent seizure activity. Some patients experience prodromal symptoms that begin at least 30 minutes before seizure onset, last 10 minutes to 3 days, and are continuous with irritability, depression, headache, confusion, and other symptoms. The postictal period is characterized by a confusional state lasting minutes to hours or, occasionally, days. Prolonged, postictal confusion may particularly follow right temporal focal seizures. Some “twilight states” result from a protracted period of intermixed ictal and postictal changes.

Peri-ictal psychotic symptoms often worsen with increasing seizure activity. Rarely psychotic symptoms alternate with seizure activity. In this *alternating psychosis,* when patients are having seizures, they are free of psychotic symptoms, but when they are seizure free, and their EEG has *forced* or *paradoxical normalization,* they manifest psychotic symptoms. The mechanism for this phenomenon is unclear, although dopamine may play a major factor. This alternating pattern is much less common than the increased emergence of psychotic behavior with increasing seizure activity.

An important peri-ictal psychiatric disorder consists of brief psychotic episodes that follow clusters of generalized tonic–clonic seizures (i.e., postictal psychosis). These psychotic episodes occur in patients who have focal seizures with impaired awareness, frequent secondary generalization to tonic–clonic seizures, bilateral interictal discharges, and frequent discharges involving the left amygdala. The postictal psychosis of epilepsy emerges after a lucid interval of 2 to 72 hours (with a mean of 1 day), during which the immediate postictal confusion resolves, and the patient appears to return to normal. The postictal psychotic episodes last 16 to 432 hours (with a mean of 3.5 days) and often include grandiose or religious delusions, elevated moods or sudden mood swings, agitation, paranoia, and impulsive behaviors, usually without perceptual delusions or voices. The postictal psychoses remit spontaneously or with the use of low-dose psychotropic medication.

A 33-year-old man with a 15-year history of generalized tonic–clonic seizures and a 4-year history of peri-ictal psychotic episodes had several hospitalizations for recurrent postictal psychosis. The initial flurry of generalized tonic–clonic seizures was followed by a 24- to 48-hour latency period and, subsequently, 2 to 7 days of delusions, hallucinations, and disordered thought processes. He believed that people could transmit messages and could read his thoughts, and voices commanded him to love his neighbor. The patient claimed to read the future and to communicate with a dead grandfather who voiced dissatisfaction with things on Earth. During these episodes, the patient had loose associations, euphoria, agitation, and occasional spike and waves on EEGs. Between psychotic episodes, he was psychiatrically and neurologically normal, and his EEGs showed left temporal interictal spikes. After the postictal psychosis, the patient returned to baseline without residual changes in behavior.

Interictal Features

INTERICTAL PSYCHOSIS (SCHIZOPHRENIFORM/SCHIZOPHRENIA LIKE PSYCHOSIS). Most psychotic episodes that occur in epilepsy patients are interictal psychosis which usually has no known direct relationship to seizure events or ictal discharge. The duration of psychotic episodes varies greatly from days to years but usually lasts longer than 1 month in the vast majority of cases. However, many of these patients develop worsening psychotic symptoms that are concomitant with an increase in seizure frequency or with AED withdrawal, and a few others have worsening psychotic symptoms on control of the seizures (alternating psychosis). The terms *alternating psychosis* and *forced* or *paradoxical normalization* refer to this demonstrable antagonism between the psychosis and the seizures or EEG discharges. Although a spontaneous resolution of interictal psychosis is not common, it may occur in patients with a shorter duration of epilepsy history and later onset of epilepsy. Patients with chronic interictal psychosis often have an early age of onset of seizures and a decade or more of poorly controlled focal seizures with impaired awareness with secondary generalized tonic–clonic seizures. Interictal psychosis may evolve from prior recurrent postictal psychotic episodes. Seizure control with AEDs or removal of the seizure focus does not prevent the development of interictal psychosis, which occasionally emerges for the first time after successful seizure treatment. This disorder sometimes resembles schizophreniform psychosis and schizoaffective psychosis with intermixed affective symptoms. In addition, there are prominent paranoid delusions, relatively preserved affect, normal premorbid personality, and no family history of schizophrenia. Other reported differences with idiopathic schizophrenia are outlined in Table 2.4–5.

Table 2.4–5.
Proposed Predisposing Factors for the Interictal Psychosis of Epilepsy

- **Epilepsy characteristics**
  - Focal onset seizures with impaired awareness and with secondary generalized tonic–clonic seizures
  - More auras and automatisms than nonpsychotic epilepsy patients
  - Epilepsy present for 11–15 years before psychosis
  - Long interval of poorly controlled seizures
  - Recently diminished seizure frequency, especially generalized tonic–clonic seizures
  - Left temporal focus
  - Mediobasal temporal lesions, especially tumors
- **Psychosis characteristics**
  - Atypical paranoid psychosis–paranoia with sudden onset
  - Psychosis alternating with seizures
  - Preserved affective warmth
  - Failure of personality deterioration
  - Less social withdrawal than schizophrenia
  - Less systematized delusions than schizophrenia
  - More hallucinations and affective symptoms than schizophrenia
  - More religiosity than schizophrenia
  - More positive, as opposed to negative, symptoms
  - Few schneiderian first-rank symptoms

A 23-year-old man developed paranoid delusions after his daily seizures were controlled for the first time. His focal seizures with impaired awareness dated from 8 years of age and consisted of a rising epigastric sensation and facial flushing followed by a motionless stare and automatisms, often culminating in secondary generalized tonic–clonic seizures. Before initiating antiepileptic therapy, there was no history of paranoid or psychotic behavior. Afterward, he believed that people were sending energy to him through small concealed batteries. He believed that he was able to work this energy off with his fluorescent watch dial and a one-armed plastic crucifix in his boot. The patient also felt that people were observing him, trying to manipulate him, and threatening him through telephone lines and telephone poles. His examination was remarkable for the degree of emotion when relating his bizarre ideas. He had a lesion in the left anterior temporal area, probably consistent with an old calcified cyst, and left temporal spikes on EEG. His paranoid delusions subsequently abated with antipsychotic therapy.

PERSONALITY DISORDERS. Among epilepsy patients, there is a high prevalence of personality disorders, including borderline, atypical or mixed, histrionic, and dependent disorders. Although many seizure patients show dependent and avoidant personality traits, the most common personality disorder in epilepsy is a borderline personality. Not surprisingly, epilepsy patients frequently lack a stable character structure and can be immature and impulsive. This personality constellation partially explains the increased incidence of irritability, suicide attempts, and intermittent explosive disorder. Those with epilepsy are stigmatized, feared, and subject to difficulties in obtaining a job, driving an automobile, and maintaining a marriage. Their experience, stigmatization, and psychosocial difficulties, along with any associated mental retardation, contribute to the dependency, low self-esteem, and overall borderline personality traits present in many such patients. In addition, the experience of epileptic auras may contribute to the development of personality disorders.

Another way to understand personality changes in patients with epilepsy is from the evidence of impaired social cognition in this population. Recent studies consistently showed that patients with temporal lobe epilepsy usually have an impairment in the ability to understand others' feelings, thoughts, and perspectives. Such ability is known as the "theory of mind (ToM)." It is also common for patients with temporal lobe epilepsy to have deficits in other aspects of social cognition, including affective and cognitive empathy, processing of implied meaning, and semantic social concept. These components may eventually lead to decreased personal warmth and inappropriate social behaviors.

GASTAUT–GESCHWIND SYNDROME. Although there is no general epileptic personality, a group of traits termed the *Gastaut–Geschwind syndrome* occurs in a subset of patients with focal seizures with impaired awareness. Some epilepsy patients with a temporal limbic focus develop a sense of the heightened significance of things. These patients are serious, humorless, and overinclusive and have an intense interest in philosophical, moral, or religious issues. Occasionally, epilepsy patients experience multiple religious conversions or experiences. In interpersonal encounters, they demonstrate *viscosity,* the tendency to talk repetitively and circumstantially about a restricted range of topics. They can spend a long time getting to the point, give detailed background information with multiple quotations, or write copiously about their thoughts and feelings (hypergraphia). Viscosity may particularly occur in patients with left-sided or bilateral temporal foci.

A 39-year-old man developed seizures after a contusion of the left temporal region. His seizures began with stereotypical voices and generalized to tonic–clonic seizures. He was extremely circumstantial and tangential, stressed every detail, and had difficulty getting to the point. Ironic and minor philosophical insights were fascinating to him. He wrote 30-page rambling letters to his physician, and his writings were full of metaphors and quotes. An example of his writing was as follows: “I became overwhelmed by the sentiment of a letter composed in my head before reaching paper. The sentiment of this letter continued to expand in all dimensions until it seemed no longer connected to any specific ideal but more to an all-pervasive color, yellow, and a smell, like burning leaves. I felt deliriously happy, but I felt in danger as well. Afterward, I got an acute attack of aphasia and could do nothing but shrug. My prior prophet voices, which went away with the Dilantin, were saying something profound that I needed to get down on paper. It seems as though I am a prophet, and I will never have another problem for the rest of my life.”

Conclusive proof that epilepsy patients with temporal lobe foci are disproportionately prone to the Gastaut–Geschwind syndrome has remained elusive. Most of the early studies used the MMPI, a test that proved insensitive to most of the specific traits attributed to epilepsy. Studies with the Bear–Fedio Inventory, an MMPI-like instrument developed to assess these “epileptic traits,” found that epilepsy patients with temporal lobe foci were sober and humorless, dependent and circumstantial, and had strong philosophical interests. In addition, those with a left-sided focus had a more reflective ideational style and maximized their problems, whereas those with a right-sided focus had emotional tendencies and minimized their problems. Further investigations with the Bear-Fedio Inventory described these seizure patients as having a viscosity in interactions, prominent religious interests, a pronounced sense of personal destiny, and deepened affect. However, other applications of this inventory found the same characteristics in nonepileptic patients with psychiatric disorders or with comparable physical disabilities. Although these personality characteristics do occur in some epilepsy patients, they may not be specific to patients with seizure disorders.

Behavioral Disturbances Variably Related to Ictus

MOOD DISORDERS. Depressive disorder is the most prevalent neuropsychiatric disorder in epilepsy; it is also the main diagnosis among epilepsy patients in psychiatric hospitals. Depression is twice as common in epilepsy patients as in comparably disabled populations, suggesting that much of the depression in epilepsy patients is more than just a psychological reaction to a disability.

Depression can manifest in several forms; MDD, persistent depressive disorder (formerly dysthymic disorder), adjustment disorder, and depressive episodes of bipolar disorder. Most patients with mood disorders and epilepsy have dysthymia; some investigators refer to this condition as chronic interictal depression (*interictal dysphoric disorder of epilepsy*) and emphasize its association with paroxysmal irritability and agitation along with a good therapeutic response to antidepressant medications. These patients may have accompanying paranoia and hallucinations, emphasizing the continuum with psychotic disorders. Patients with interictal dysphoria tend to have frequent focal seizures with impaired awareness, possibly with greater left-sided temporal foci, although this lateralization is not well established. The experience of certain psychic auras, especially those with cognitive content, may predispose to interictal depression. Several investigators also report increased seizure control or a decrease in seizures before the onset of interictal depressive symptoms. Patients with this “alternating depression” experience relief with a seizure or Electroconvulsive therapy (ECT).

There are other associations between depression with epilepsy. The rare occurrence of ictal depression may not only outlast the actual ictus but also may lead to suicide. Depression also occurs peri-ictally. Episodic mood disturbances, often with agitation, suicidal behavior, and psychotic symptoms, may occur with increased seizure activity. Finally, postictal depression is common, and a prolonged depressive state occasionally follows focal seizures with impaired awareness, even when ictal experiences do not include depression.

Mood disorder due to epilepsy with manic features or with mixed features is much rarer than mood disorder due to epilepsy with depressive features or with major depressive-like features. Rarely, manic symptoms may emerge with an increase in seizure frequency or after seizure control.

Although a right temporal focus is a possible source of mania in epilepsy, this laterality is not well established.

ANXIETY DISORDERS. Anxiety disorders, including generalized anxiety disorder (GAD), panic disorder, obsessive-compulsive disorder (OCD), posttraumatic stress disorder (PTSD), phobia, and social anxiety disorder, are the second most common psychiatric comorbidities in patients with epilepsy after depressive disorders. GAD appears to be the most common anxiety disorder among patients with epilepsy. The prevalence of anxiety disorders in epilepsy patients may be as high as 30%. Anxiety disorders that occur during the interictal phase must be distinguished from focal seizures without impaired awareness manifesting as anxiety or panic attacks. Anxiety may be present with depression or other psychopathology, as part of Cluster C personality disorders or independently as a GAD. Some patients with epilepsy clearly have PTSD from the psychological trauma of their recurrent seizures. Finally, among women with epilepsy, the risk of both anxiety and depression is higher during pregnancy and after childbirth compared to the general population.

AGGRESSION. Lay people have attributed aggressive and violent acts to epilepsy and have even used this epilepsy defense in criminal proceedings. This belief peaked in the 19th century when the criminologist Cesare Lombroso promoted the association of epilepsy with aggressive, sociopathic tendencies. Investigators have bolstered this association with studies showing aggressive verbalizations with stimulation of the amygdala and interictal defensive rage in cats with epileptic hippocampal lesions. A minority of violent epilepsy patients have left-sided amygdala atrophy, probably from prior encephalitis. Among patients in a maximum-security psychiatric hospital, the violent patients had focal temporal slowing or sharp waves on EEG and dilated temporal horns or small temporal lobes on CT scans. Moreover, on neuropsychological tests, patients with left temporal lobe seizure foci have higher scores on hostile feelings than other patients with epilepsy. These results suggest that aggression is associated with abnormal temporal structure and electrical discharges.

A 37-year-old left-handed man with epilepsy presented with aggressive episodes. The seizures consisted of an olfactory aura followed by “spacing out” or alteration of consciousness for approximately 1 minute. In addition to these focal seizures with impaired awareness, the patient had occasional secondary generalized tonic–clonic seizures with urinary incontinence and tongue biting. During the postictal period, as he began to recover consciousness, he experienced an overwhelming sense of threat or of having been harmed. These feelings became focused on any individual who was in his immediate environment. That person was believed to have beaten or otherwise hurt him and was going to harm him further. The patient felt compelled to attack these individuals, often inflicting significant physical injury. Although his postictal confusion would clear in approximately 1 hour, his sense of being harmed or threatened slowly diminished over approximately 24 hours after a seizure. After the resolution of these feelings, he felt great remorse over the harm that he had done. Nevertheless, on several occasions, he was charged with aggravated assault. Sleep-deprived EEGs confirmed the presence of left anterior temporal epileptiform activity. The patient’s aggressive postictal episodes abated with control of his focal seizures with carbamazepine.

Although aggression can occur in relation to an ictus, as exemplified by this patient’s subacute postictal aggression, most aggression among epilepsy patients is not related to epileptiform activity. Aggression in epilepsy is usually associated with psychosis or with intermittent explosive disorder and correlates with subnormal intelligence, lower socioeconomic status (SES), childhood behavior problems, prior head injuries, and possible orbital frontal damage. Moreover, although the prevalence of epilepsy among prison inmates has been two to four times that among the general population, studies from the United Kingdom and the United States have not found more violent crimes among prisoners with epilepsy than among prisoners without epilepsy. Antiepileptic medication may also trigger aggression and mood lability in some patients, especially levetiracetam.

Can violence itself be a seizure? After the 1976 case of a New York City policeman who had never had seizures and successfully claimed the epilepsy defense, criteria for ictal violence were proposed that included video-EEG (VEEG) telemetry (Table 2.4–6). Since then, epilepsy has rarely, if ever, been proved to directly result in well-organized, premeditated, purposeful violence. Such acts require a series of coordinated steps that rarely occur as manifestations of seizures. Simple violent automatisms, such as spitting, verbal profanities, or flailing the arms, can occur at the onset of focal seizures with impaired awareness, and secondary violent automatisms can occur as a response to an unpleasant or emotional aura or peri-ictal sensation (Table 2.4–7). More commonly, nondirected violent movements, aimless destructive behavior, or angry verbal outbursts occur during postictal delirium when patients misinterpret attempts to protect or restrain them (resistive violence) or as a manifestation of postictal psychosis and subacute postictal aggression.

Table 2.4–6.
Criteria for the Assessment of Ictal Violence in Epilepsy

The diagnosis of epilepsy is established by at least one specialist in epilepsy.

The presence of epileptic automatisms is documented by history and by closed circuit television EEG telemetry.

The presence of violence during epileptic automatisms is verified in a videotape-recorded seizure in which ictal epileptiform patterns are also recorded on the EEG.

The aggressive act is characteristic of the patient’s habitual seizures, as elicited by history.

A clinical judgment is made by the epilepsy specialist attesting to the possibility that the aggressive act was part of a seizure.

EEG, electroencephalogram.

SEXUALITY. Patients with epilepsy tend to be hyposexual. Men and women experience disturbances of sexual arousal and a lower sexual drive. Some patients have a disinterest in “all the usual libidinous aspects of life,” including loss of erotic fantasies or dreams, and may experience impotence or frigidity. In general, sexual dysfunction is most likely to occur among patients with uncontrolled and frequent seizures, longer duration of epilepsy, focal seizures with impaired awareness, and those receiving enzyme-inducing and multiple AEDs.

Although comorbid depression and anxiety may affect sexual behavior, men with epilepsy have a two- to threefold increased risk of erectile dysfunction, suggesting a neurophysiologic component, and studies of sex hormones suggest the possibility of subclinical hypogonadotropic hypogonadism. Moreover, there can be undesirable effects on the sexual performance of antiepileptic medication, including erectile dysfunction, decreased sexual desire, retrograde ejaculation, and altered semen quality. AEDs such as oxcarbazepine, lamotrigine, and levetiracetam may be associated with significant sexual dysfunction. Hormone replacement therapy may normalize sexual drive and sexual performance in some patients. Substantial improvement to the point of public hypersexuality can occur after seizures are brought under control. Before temporal lobectomy, most epilepsy patients were hyposexual, but nearly one-third of them had an increase in libido after the operation, providing that their seizures were controlled.

Table 2.4–7.
Mechanisms of Aggression among Epilepsy Patients

Period	Cause
Interictal	Impulse control disorder
	Intellectual disability or cognitive impairments
	Personality disorders
	Schizophreniform/schizophrenia-like psychosis
	Medication related
Prodrome	Mounting tension, irritability
Ictal	Direct manifestation of the seizure
	Violent automatism
	Reaction to a negative aura
	Subtle seizure equivalents
Postictal	Resistive
	Postictal psychosis
	Subacute postictal aggression
	Poriomania and somnambulism

Other sexuality changes are rare. Individual cases of homosexuality, transvestism, fetishism, and gender dysphoria are not frequent enough to exclude a coincident association. True ictal sexual manifestations are also unusual; however, libidinous feelings, erotic sensations, sexual remembrances, and even orgasm rarely occur, primarily in women and probably from seizure discharges in the amygdala. In addition, ictal masturbation has occurred with absence status. A woman with nymphomania proved to have incidental sexuality from sensory focal seizures caused by a tumor in the sensory cortex representing her genital region.

SUICIDE AND SELF-HARM. Patients with epilepsy are at risk of higher rates of self-harm, suicidal ideation, and complete suicide. The risk of completed suicide in epilepsy patients is 4 to 5 times greater than that among the nonepileptic population, and those with focal seizures of temporal lobe origin have a particularly high risk, as much as 25 times greater. Death by suicide occurs in 3% to 7% of epilepsy patients. A comparison of suicide attempts among patients with epilepsy and comparably handicapped nonepileptic controls has reported that 30% of those with epilepsy had attempted suicide as compared to 7% of the controls. This increased risk of suicide continues even long after temporal lobectomy and successful control of seizures. Although many studies showed that the severity of depressive symptoms and psychosocial status influence suicidal ideation in patients with epilepsy, most suicidal behavior among epilepsy patients is not directly due to reactions to the psychosocial stressors of having a seizure disorder. Rather, these patients are likely to attempt suicide in conjunction with borderline personality behaviors and are likely to complete suicide during postictal psychosis. In addition, some evidence suggests that antiepileptic medications may also increase suicidality. This conclusion is still controversial. Contributors to successful suicides include paranoid hallucinations, agitated compunction to kill themselves, and occasional ictal command hallucinations to commit suicide. In sum, the cause of self-injurious behavior in patients with epilepsy is often multifactorial, and clinicians should consider any potential cause for an early intervention to prevent self-harm and suicide.

A 26-year-old woman had her initial seizure during her first pregnancy at 18 years of age. Her seizures included echoing sounds "like walking in a cave," a motionless stare with stereotypical automatisms, postictal confusion, and occasional secondary generalized tonic–clonic seizures. Because of variable AED response, she underwent EEG and closed-circuit television video-EEG (CCTV-EEG) telemetry that documented focal seizures from a right temporal focus and nonepileptic seizures. The patient, who had six children by six different individuals, had prominent feelings of inadequacy and isolation and was considering cutting her wrists "just to see if anyone cared." Her multiple suicide attempts and threats resulted in five psychiatric hospitalizations. During one period of time, she complained of decreased menses, weight gain, stretch marks, increased appetite and sleep, and exhaustion. She insisted that she was pregnant despite six negative pregnancy tests and multiple evaluations.

OTHER BEHAVIORAL CHANGES. Other psychiatric disorders may be associated with epilepsy or epileptiform EEG activity. A specific association of epilepsy with dissociative identity disorder, depersonalization disorders, possession states, fugue states, and psychogenic amnesia is intriguing but unresolved. In one intracarotid amobarbital (Amytal) study of dissociative identity disorder in two seizure patients, the different personalities were precipitated without seizure activity. Persistent alterations in the experience of self and feelings of being taken over by others may occur in patients with auras of derealization and depersonalization. In epilepsy, prolonged periods of compulsive wandering with amnesia have resulted from an admixture of ictal and postictal changes and have been termed *poriomania.* Finally, patients with epilepsy are subject to other behavioral difficulties stemming from their epilepsy, such as adjustment disorders, subtle cognitive effects of seizures, and the potential behavioral effects of antiepileptic medications.

Pathology and Laboratory Examination

Neurodiagnostic Tests. Seizure and epilepsy are clinical diagnoses. Additional tests can help identify provocative factors and diagnose seizure syndromes. In addition to the routine laboratory data and toxicology screens used to exclude reactive seizures, several neurodiagnostic tests are useful in the assessment of epilepsy. Scalp EEG is the most common confirmatory test for seizures; however, sensitivity ranges from 29% to 55%. This is because the source area of epileptic discharge has to be large enough in order to be detected by scalp EEG. Only 10% of cortical spikes having a source area of less than 10 cm^2 produced scalp potentials; the source area, which is less than 6 cm^2 of gyral cortex, is rarely captured by scalp EEG. Single routine EEGs are frequently normal and must be repeated, particularly with provocative maneuvers, such as photic stimulation, hyperventilation, and sleep deprivation. Occasionally, VEEG telemetry for an extended period of time is necessary to characterize spells of unknown etiology, for localization of seizure focus, and to the classification of seizures in a patient with known epilepsy. Inpatient epilepsy monitoring units are very helpful in distinguishing epilepsy from Psychogenic nonepileptic seizures (PNES). Neuroimaging procedures, such as CT scans and MRI, can more precisely visualize a seizure focus or even mesial temporal sclerosis (Fig. 2.4–4). If the patient is a candidate for epilepsy surgery due to medication-resistant epilepsy, other tests that occasionally aid in localizing the seizure focus include invasive EEG, SPECT scans, and PET scans. PET scans may show interictal hypometabolism around the temporal seizure focus and are also useful in the presurgical assessment of medically intractable seizure patients. Neuropsychological examinations, particularly during a Wada test, further help in localizing and lateralizing memory and language before surgery.

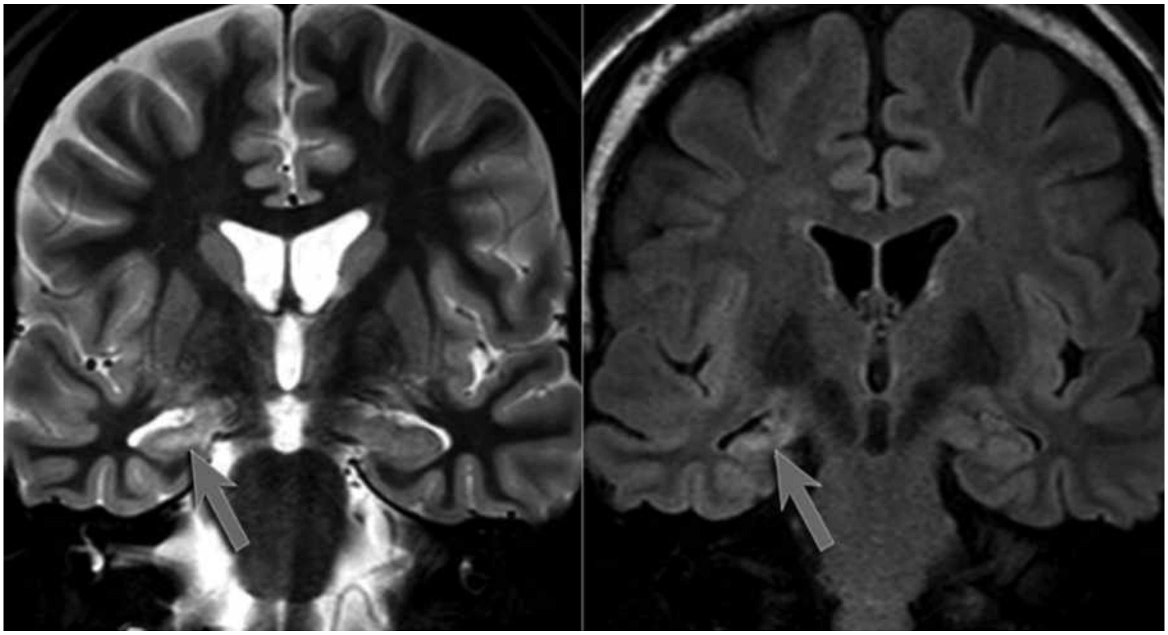

FIGURE 2.4–4. Coronal T2-weighted (**left**) and FLAIR (**right**) MR images showing prominent right-sided mesial temporal sclerosis. The high signal indicates gliosis. There is associated atrophy with enlargement of the right temporal horn of the lateral ventricle. (From Robin Smithius, MD. http://www.radiologyassistant.nl/)

Neuropathology. The most common pathologic findings in epilepsy are mediobasal temporal lobe lesions. Approximately two-thirds of epileptic adults have a temporal lobe focus, and two-thirds of these have mesial temporal sclerosis with pyramidal cell loss in the hippocampus. Theories about the cause of mesial temporal sclerosis include perinatal insults, dysgenesis, and kindling from reactive seizures. Another 20% to 25% of those with temporal lobe lesions have tumors, such as hamartomas and gangliogliomas. The rest have scars from trauma and other causes or lack a distinct histologic lesion.

Course and Prognosis

Most epileptic patients have a good prognosis. The majority of seizures can be controlled sufficiently with antiepileptic medications so that the patient can live a productive life. Some seizures, such as absence seizures, tend to disappear by adulthood. For epilepsy patients who are medically intractable, epilepsy surgery offers a good alternative (e.g., temporal lobectomy), provided that the focus can be localized. In addition, most epileptic patients do not have psychiatric disorders, and others have psychiatric difficulties only if they endure many years of poorly controlled seizures. For those with behavioral problems, AEDs or epilepsy surgery may relieve some symptoms, such as aggression and impulsivity, but may not affect the emergence of others, such as psychosis and suicidal behavior. Existing cognitive-behavioral therapy (CBT) and mindfulness-based interventions appear to have some benefit for psychiatric symptoms in epilepsy. Psychological interventions targeting the enhancement of health-related quality of life (HRQOL) and adherence and a decrease in comorbidity symptoms (anxiety, depression) need to be incorporated as part of a comprehensive epilepsy care program. Clinicians can also address and minimize the "stress profile" of patients with epilepsy, which may not only protect against psychiatric comorbidities but also reduce seizure frequency.

Treatment

Antiepileptic Medications. In the treatment of psychiatrically disturbed epilepsy patients, the first consideration is the behavioral effects of antiepileptic medications. Carbamazepine and valproate have significant antimanic and modest antidepressant properties. They have efficacy in the long-term prophylaxis of manic episodes and may also ameliorate some dyscontrolled, aggressive behavior in brain-injured patients. Lamotrigine has well-established efficacy in preventing the recurrence of depressive episodes in bipolar disorder, and lacosamide may additionally exert a beneficial effect on mood. Clonazepam, in addition to its anxiolytic properties, can serve as a supplement to other antimanic therapies. Gabapentin and pregabalin also decrease anxiety and improve general well-being in some epilepsy patients. Carbamazepine and ethosuximide may have value for borderline personality disorder in reducing mood swings and behavioral disinhibition.

Encephalopathic changes occur at toxic levels of all AEDs. Even at therapeutic levels, barbiturates may need discontinuation because of drug-induced depression, suicidal ideation, sedation, psychomotor slowing, and paradoxical hyperactivity in the very young and the very old. The exact incidence of aggressive behavior as an adverse effect of AEDs is difficult to estimate, with some reports of up to 16%. Treatment with levetiracetam, perampanel, and topiramate, which have diverse mechanisms of action, can result in aggressive behavior. Psychiatric and behavioral side effects occur in up to 8% of patients with drug-resistant epilepsy with treatment-emergent psychiatric adverse events regardless of the AED, although most frequent with levetiracetam and zonisamide. Levetiracetam is a well-known cause of irritability and aggression and may need to be avoided in emotionally unstable patients. Gabapentin may induce aggressive behavior or hypomania, and vigabatrin (Sabril) may precipitate depression. In addition, clinicians need to be aware of the potential emergence of psychopathology on withdrawal of antiepileptic medications. Anxiety and depression are the most common emergent symptoms of withdrawal, but psychosis and other behaviors may also occur (Table 2.4–8).

Table 2.4–8.
Behavioral Effects of Antiepileptic Drugs

Antiepileptics	Treatment-Emergent Behavioral Adverse Effects	Potentially Positive Behavioral Effects
Brivaracetam	Anxiety, depression, sedation, less aggression, and irritability than levetiracetam, rare psychosis	Sedation
Carbamazepine–Oxcarbazepine	Irritability, impaired attention, rare suicidality	Mood stabilization
Clobazam	Lethargy, sedation, impaired cognition, occasional suicidality	Sedation
Clonazepam	Lethargy, depression, impaired cognition, irritability, disinhibition	Sedation
Ethosuximide	Behavioral abnormalities, rare psychosis	
Felbamate	Depression, anxiety, irritability	
Gabapentin	Behavioral problems in children	
Lacosamide	Depression, rare suicidality	
Lamotrigine	Insomnia, agitation, occasional emotional lability	Mood stabilization
Levetiracetam	Aggression, irritability, common emotional lability	
Phenobarbital	Depression, irritability, aggression, impaired cognition and attention, sedation, hyperactivity	Sedation
Phenytoin	Encephalopathy, depression, impaired attention	
Pregabalin	Less likely to cause behavioral problems	
Primidone	Emotional disturbance, irritability, impaired cognition, sedation	Sedation
Tiagabine	Depression (nonconvulsive status epilepticus), anxiety, irritability,	Sedation

	sedation	
Topiramate	Depression, psychomotor slowing, psychosis, impaired cognition (word-finding and memory)	
Valproate	Encephalopathy, depression	Mood stabilization
Vigabatrin	Depression, aggression, psychosis	
Zonisamide	Agitation, depression, psychosis, somnolence, impaired cognition, rare psychosis	

Psychotropic Medications

ANTIPSYCHOTICS. A small number (up to 15%) of the interictal psychosis resolves without any treatment, but a majority of patients require antipsychotics. There are no distinct differences in efficacies between second- antipsychotics and first-generation antipsychotics. Therefore, a second consideration is the seizure threshold lowering effect of psychotropic medications (Table 2.4–9). This is usually not a problem but can occasionally reach clinical significance in poorly controlled epilepsy. Haloperidol, risperidone, paliperidone, and aripiprazole may be the least seizure threshold lowering agents in contrast to clozapine, olanzapine, quetiapine, chlorpromazine, thioridazine, and perphenazine. Quetiapine has clinically significant interaction with enzyme-inducing drugs, even for dosages below 700 mg/day. These psychotropic drugs are most convulsive with the rapid introduction of the drug and in high doses. Clozapine, for example, had induced seizures in 1% to 4.4% of patients, particularly when the dose was rapidly increased. In general, mild-to-moderate psychotic episode patients usually require 300 to 1,000 mg/day of chlorpromazine equivalent (CE) dose (Table 2.4–10). Although patients with a severe psychotic episode, for example, unshakable delusion with severe aggression and suicidal ideation, may require dosages higher than 1,000 mg/day of CE, it is still the best to initiate with a low dose and titrate up slowly while monitoring antiepileptic levels and EEGs.

ANTIDEPRESSANTS. Prescribing antidepressants can alleviate both anxiety and mood symptoms at the same time. Given that the overall efficacies of antidepressants in treating depressive symptoms are the same, the principle of choosing antidepressants will be based on adverse effects and tolerability of the patients. The most concerning issue for antidepressants in epilepsy is the seizure threshold lowering property. Although therapeutic doses do not usually promote seizures, overdose and intoxication of tricyclic (TCA) and tetracyclic antidepressants (TeCA), selective serotonin reuptake inhibitors (SSRIs), and serotonin–norepinephrine reuptake inhibitors (SNRIs) are potential causes of seizures. Other antidepressants known to lower the seizure threshold usually avoided among patients with epilepsy are amoxapine, bupropion, clomipramine, and maprotiline. Paradoxically, some studies demonstrated that antidepressants might help in reducing the frequency of seizures in some patients.

Table 2.4–9.

Seizure Threshold Lowering Effect of Psychotropic Medications

Potential	Antipsychotic	Antidepressant	Other Psychotropic
High	Chlorpromazine (Thorazine)	Bupropion (Wellbutrin)	
	Clozapine (Clozaril)	Imipramine (Norfranil)	
	Thioridazine (Mellaril)	Maprotiline (Ludiomil)	
	Perphenazine (Trilafon)	Amitriptyline (Elavil)	
	Olanzapine (Zyprexa)	Amoxapine (Asendin)	
	Quetiapine (Seroquel)	Nortriptyline (Aventyl)	
Moderate	Most piperazines	Protriptyline (Vivactil)	Lithium (Eskalith)
	Paliperidone (Invega)		
	Thiothixene (Navane)	Clomipramine (Anafranil)	
	Ziprasidone (Geodon)		
Low	Fluphenazine (Prolixin)	Doxepin (Sinequan)	Ethchlorvynol (Placidol)
	Haloperidol (Haldol)	Desipramine (Norpramin)	Glutethimide (Doriden)
	Loxapine (Loxitane)	Trazodone (Desyrel)	Hydroxyzine (Vistaril)
	Molindone (Moban)	Trimipramine (Surmontil)	Meprobamate (Equanil)
	Pimozide (Orap)	Selective serotonin reuptake inhibitors	Methaqualone (Quaalude)
	Risperidone	Serotonin–norepinephrine	

(Risperdal)	reuptake inhibitors
Paliperidone (Invega)	
Aripiprazole (Abilify)	
Pimavanserin (Nuplazid)	
Lumateperone (Caplyta)	
Asenapine (Saphris, Secuado)	
Cariprazine (Vraylar)	
Lurasidone (Latuda)	

According to their better tolerability profile, most SSRIs and SNRIs can be given as first-line antidepressants in patients with epilepsy (Table 2.4–11). SSRIs may be more suitable in patients with a depressive disorder associated with comorbid anxiety disorder, while SNRIs may be more beneficial in patients with depressive disorder with psychomotor retardation, fatigue, anhedonia, and cognitive slowness. These medications should not be discontinued abruptly because of withdrawal symptoms.

ANTIANXIETY. Although anxiety disorders are common comorbidities in epilepsy, there is a lack of evidence from good quality randomized controlled trials (RCTs) with epilepsy patients. Therefore, regimens of psychotropic prescription largely come from evidence in general psychiatric populations. Benzodiazepines may be particularly helpful in the rapid relief of symptoms of anxiety, but a recent study demonstrated that chronic use of benzodiazepines in epilepsy patients with GAD did not decrease anxiety severity. If patients have comorbid panic disorder, GAD, OCD, or PTSD, SSRIs are probably the best choice for treatment. Pregabalin recently showed an antianxiety effect in a small open-label trial in patients with epilepsy and GAD. Therefore, it can be an alternative choice if the patient does not respond to an SSRI.

Drug Interactions. A third treatment consideration is a potential for interaction of antiepileptic and psychotropic medications. Most commonly, AEDs can increase the metabolism of a psychotropic drug with a consequent decrease in its therapeutic efficiency. Conversely, withdrawal of an AED can precipitate rebound elevations in psychotropic drug levels. Moreover, the initiation of a psychotropic drug may result in competitive inhibition of AED metabolism with elevations of AED levels to toxicity. These interactions should be considered carefully, especially in older antiepileptic medications. However, the newer antiepileptic medications have fewer potential interactions with psychotropic medications. Lacosamide (Vimpat), levetiracetam (Keppra), zonisamide (Zonegran), gabapentin (Neurontin), clobazam (Onfi), perampanel (Fycompa), ethosuximide (Zarontin), vigabatrin (Sabril), and tiagabine (Gabitril) are relatively free of enzyme-inducing or -inhibiting properties. Lamotrigine (Lamictal) and topiramate (Topamax) have interactions as described in Table 2.4–12.

Surgery. Up to a third of epilepsy patients may have seizures that are not fully controlled with medications. For these patients, resection of epileptogenic brain regions may offer a far better chance of seizure control than medication alone. The most commonly performed operation is a temporal lobectomy, involving resection or removal of 4 to 6 cm of the anterior temporal lobe. More than 80% of temporal lobectomy patients experience some reduction in their seizure frequency, and more than 60% of patients are entirely seizure free. Less common epilepsy surgeries include resection of cortical dysplasia, tumors, vascular malformations, areas of old stroke, removal of the epileptogenic hemisphere, and ligation of the corpus callosum. Neocortical resections result in about 50% of seizure freedom.

Neuropsychiatric manifestations of peri-surgery are beginning to be characterized, and data is most robust in temporal lobectomies. Similar to medically managed patients, surgical patients suffer from depression and anxiety. Most depression occurs during the first 3 months following epilepsy surgery, with the majority manifesting depressive symptoms during the 3- to 12-month postoperative period. Depressive symptoms may correlate with dysfunction in temporolimbic structures contralateral to the epileptogenic zone. Studies have shown that independent of postoperative seizure outcome, the severity of depression and anxiety independently predict the quality of life. These psychiatric symptoms gradually improve after surgery; only a minority of patients will suffer from refractory depression and anxiety.

Table 2.4–10.
Chlorpromazine Equivalent (CE) Dosages of Antipsychotics

First Generation Antipsychotics	**Chlorpromazine Equivalents (mg)**
Chlorpromazine	100
Fluphenazine	2
Haloperidol	2
Loxapine	10
Perphenazine	8
Pimozide	2
Prochlorperazine	15
Trifluoperazine	2–5
Thioridazine	100
Thiothixene	4
Second Generation Antipsychotics	
Aripiprazole	7.5
Asenapine	4
Brexpiprazole	N/A
Cariprazine	N/A
Clozapine	100
Iloperidone	3–4
Lurasidone	16
Olanzapine	5
Paliperidone	2
Quetiapine	75
Risperidone	1
Ziprasidone	60

CE: 1 mg of each antipsychotic can be converted into a certain dose of chlorpromazine, e.g., 1 mg of haloperidol is about 50 mg of chlorpromazine.

Adapted from American Association of Psychiatric Pharmacists. Psychiatric pharmacy essentials: Antipsychotic dose equivalents.

https://aapp.org/guideline/essentials/antipsychotic-dose-equivalents.

References for this table:

Leucht S, Samara M, Heres S, Patel MX, Woods SW, Davis JM. Dose equivalents for second-generation antipsychotics: the minimum effective dose method. *Schizophr Bull*. 2014;40(2):314–326.

Leucht S, Samara M, Heres S, Davis JM. Dose equivalents for antipsychotic drugs: the DDD method. *Schizophr Bull*. 2016;42 (Suppl 1):S90–S94.

Rothe PH, Heres S, Leucht S. Dose equivalents for second generation long-acting injectable antipsychotics: The minimum effective dose method. *Schizophr Res*. 2018;193:23–28.

Leucht S, Samara M, Heres S, et al. Dose equivalents for second-generation antipsychotic drugs: the classical mean dose method. *Schizophr Bull*. 2015;41(6):1397–1402.

Patel MX, Arista IA, Taylor M, Barnes TRE. How to compare doses of different antipsychotics: a systematic review of methods. *Schizophr Res*. 2013;149(1–3):141–148.

Taylor DM, Paton C, Kapur S. *The Maudsley Prescribing Guidelines in Psychiatry*. 11th ed. Wiley-Blackwell; 2011.

Schatzberg AF, Cole JO, DeBattista C. *Manual of Clinical Psychopharmacology*. 7th ed. American Psychiatric Publishing; 2010.

Woods SW. Chlorpromazine equivalent doses for the newer atypical antipsychotics. *J Clin Psychiatry*. 2003;64(6):663–667.

Removal of the amygdala and most of the hippocampus may have postoperative behavioral effects. Some patients develop anomia or a verbal memory deficit after resection of the dominant hemisphere, and patients occasionally develop a transient postoperative affective disorder. Others experience a reduction in postictal psychosis, depression, and hyposexuality, but epilepsy patients may continue to develop interictal psychosis, personality changes, and suicidal behavior even long after the temporal lobectomy. Patients with preoperative psychotic symptoms are at higher risk for poor surgical outcomes and postoperative psychosis. Corpus callosotomy, which aims to prevent the interhemispheric spread of seizures, results in a unique, transient disconnection syndrome of mutism, apathy, agnosia, apraxia of the nondominant limbs, difficulty naming, and difficulty writing with the nondominant hand.

Table 2.4–11.
Suggested Antidepressants for Patients with Epilepsy

Suggested Level	Drug Name and Classification	Dose Range
First-line	SSRI	
	Fluoxetine	20–60
	Sertraline	50–150
	Escitalopram	10–20
	SNRI	
	Venlafaxine	75–300
	Duloxetine	30–90
Second-line	TCA	
	Amitriptyline	150–300
	Nortriptyline	50–150
	Imipramine	75–150
	Clomipramine	75–150
	TeCA	
	Maprotiline	75–150
	Mirtazapine	7.5–30

SNRI, serotonin and norepinephrine reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressants; TeCA, tetracyclic antidepressants.

Seizure Management. In treating the neuropsychiatric disorders of epilepsy, a final consideration is altering the seizure management itself. In addition to the occasional behavior alleviated by strict seizure control, neuromodulation techniques may offer promise. ECT, vagus nerve stimulation (VNS), and repetitive transcranial magnetic stimulation (rTMS) may relieve some cases of peri-ictal psychosis, depression, or other behaviors. Studies also suggest that VNS ameliorates depressive symptoms in drug-resistant epilepsy patients and that the VNS effect on depression is not correlated to seizure response. On the horizon, promising new treatments for epilepsy include the use of stem cells to treat both the biologic and psychiatric components. All patients with epilepsy should have optimal care by addressing seizure effects on their quality of life, such as emotional health, physical health, social health, and independent lifestyle. Ultimately, the optimal management of the neuropsychiatric comorbidities of epilepsy involves an interdisciplinary team that may include neurologists, psychiatrists, psychologists, social workers, and, when indicated, neurosurgeons.

PSYCHOGENIC NONEPILEPTIC SEIZURES

PNES are sudden involuntary episodes of any combination of altered movement, sensation, or awareness that bear a resemblance to epileptic seizures but are not explained by epileptiform electrical activity and are instead presumed to be associated with a psychological origin. PNES are classified as a subtype of functional neurologic symptom disorder (FNSD, also known as conversion disorder) in DSM-5, and they are also called dissociative seizures or functional seizures. Some clinicians refer to PNES episodes as “events” or “spells” instead of “seizures” to create less confusion for patients who associate the word “seizure” with epilepsy.

PNES are frequently misdiagnosed as epilepsy. The mean latency from initial manifestation to proper diagnosis is 5 to 7 years. PNES can represent up to 30% of admissions to epilepsy monitoring units. PNES are associated with significant medical resource utilization, disability, and low quality of life. They affect women during their reproductive years at a 3:1 ratio compared to men. While more frequently identified in young adult patients, PNES can occur across the lifespan.

Table 2.4–12.
Antiepileptic–Psychotropic Drug Effects on Blood Levels

Antiepileptic	Type of Seizure	Effects of Psychotropic Drug on Antiepileptic Drug[a]	Effects of Antiepileptic Drug on Psychotropic Drug[a]
Carbamazepine (Tegretol)	Focal, generalized	Potentially decreased	Decreased
Oxcarbazepine (Trileptal)	Focal, generalized	Mild increased, rarely toxic levels	Decreased
Eslicarbazepine (Aptiom)	Focal	Potentially increased	May increase aripiprazole
Phenytoin (Dilantin)	Focal, generalized	Potentially decreased or increased, rarely toxic levels	Decreased
Phenobarbital (Barbita) and (Myidone)	Focal, generalized	Potentially decreased	Significantly decreased
Valproic acid (Depakene)	Focal, generalized, absence	Potentially increased, rarely toxic levels	Potentially decreased
Ethosuximide (Zarontin)	Absence	None known	None known
Clonazepam (Klonopin)	Generalized	Potentially decreased	Potentially decreased
Gabapentin (Neurontin)	Focal, generalized	No significant interactions known	No significant interactions known
Lamotrigine (Lamictal)	Focal, generalized	No significant interactions known	Decrease quetiapine, increase olanzapine,

			Increase aripiprazole
Topiramate (Topamax)	Focal, generalized	No significant interactions known	Increase haloperidol, decrease risperidone
Vigabatrin (Sabril)	Focal, generalized	No significant interactions known	No significant interactions known
Tiagabine (Gabitril)	Focal, generalized	No significant interactions known	No significant interactions known
Felbamate (Felbatol)	Focal, generalized	No significant interactions known	No significant interactions known
Brivaracetam (Briviact)	Focal	No significant interactions known	No significant interactions known
Levetiracetam (Keppra)	Focal, generalized	No significant interactions known	No significant interactions known
Pregabalin (Lyrica)	Focal	No significant interactions known	No significant interactions known
Tiagabine (Gabitril)	Focal	No significant interactions known	No significant interactions known
Cannabidiol (Epidiolex)	Lennox-Gastaut, Dravet, Tuberous Sclerosis Complex	No significant Interactions known	May increase clozapine levels
Cenobamate (Xcopri)	Focal	No significant interactions known	May reduce pimavanserin,

			clozapine, and aripiprazole levels
Vigabatrin (Sabril)	Focal, generalized	No significant interactions known	No significant interactions known
Zonisamide (Zonegran)	Focal, generalized	No significant interactions known	No significant interactions known
Lacosamide (Vimpat)	Focal, generalized	No significant interactions known	No significant interactions known
Perampanel (Fycompa)	Focal, generalized	No significant interactions known	No significant interactions known
Primidone (Mysoline)	Focal, generalized	No significant interactions known	May decrease levels of aripiprazole, clozapine, quetiapine, and risperidone
Rufinamide (Banzel)	Lennox-Gastaut syndrome	No significant interactions known	May decrease clozapine levels
Stiripentol (Diacomit)	Dravet syndrome(with clobazam)	No significant interactions known	May increase clozapine levels

[a] Antipsychotic and antidepressant drugs; lithium and benzodiazepines have few drug interactions with antiepileptic drugs.

Focal, focal onset seizure; generalized, generalized onset seizure; absence, absence seizure.

Data from Patsalos PN. Drug interactions with the newer antiepileptic drugs (AEDs)—Part 2: pharmacokinetic and pharmacodynamic interactions between AEDs and drugs used to treat non-epilepsy disorders. *Clin Pharmacokinet.* 2013;52:1045–1061. doi: 10.1007/s40262-013-0088-z.

Diagnosis

Diagnosis is established by capturing semiology characteristics of PNES without accompanying electrographic changes during VEEG monitoring. Semiologic signs suggestive of PNES include preserved awareness during bilateral movements, eye fluttering or forced eye closure, responsiveness to bystander intervention during the episode, prolonged duration (>2 minutes), waxing and waning course during the episode, and rapid full recovery immediately following the episode. Frontal lobe seizures may present a particular diagnostic challenge as clinical manifestations may be reminiscent of PNES and show minimal or no changes in EEG at the time of the episode. Semiologic features typical of frontal lobe seizures that distinguish them from PNES include short duration (<30 seconds), occurrence from EEG-verified sleep, open eyes, tonic posturing, and ictal grasping. Table 2.4–13 lists semiologic signs that can help distinguish epileptic seizures from PNES. Induction techniques using conventional activation procedures (hyperventilation and photic stimulation) can accelerate the capture of a typical episode during VEEG and offer a diagnostic admission of a shorter duration. Interictal abnormalities and even epileptiform discharges can be seen in patients with sole PNES and should not lead to overinterpretation. EEG results should be contextualized with history and semiology.

Table 2.4–13.
Differentiating Semiologic Signs of Epileptic Seizures versus PNES

	Epileptic Seizures	PNES
Duration	Shorter (less than 1 minute usually)	Longer (more than 2 minutes usually)
Eyes	More commonly open	More commonly closed
Bodily injury	Relatively more common	Not common
Urinary incontinence	Can be seen	Can be seen
Ictal weeping	Unlikely	Relatively common
Occurrence from sleep (EEG-confirmed)	Can occur	No
Pelvic thrusting	Can be seen in frontal lobe seizures	Can be seen
Asynchronous body movement	Can be seen in frontal lobe seizures	Can be seen
Influence by others	Unlikely	More likely
Postictal confusion	Common	Not common
Postictal breathing	Stertorous	Shallow

Table 2.4–14.
Physiologic Paroxysms that Need to be Differentiated from Epileptic Seizures and PNES

Physiologic Paroxysm	**How to Identify and Diagnose?**
Syncope	Very short (seconds) loss of consciousness, prodromal signs (lightheadedness, pallor, diaphoresis, nausea/vomiting, tunnel vision), can be accompanied by myoclonic jerks, no postepisode confusion, can be triggered by changes in position or specific situations (micturition, blood sight). *Recommended tests to support diagnosis: electrocardiogram (ECG), holter ECG, tilt table test.*
Parasomnias	Complex behaviors during sleep (eating, walking, etc.), behaviors are not stereotyped, usually last minutes, little or no recollection, onset typically in childhood. *Recommended tests to support diagnosis: polysomnography or video polysomnography.*
Narcolepsy	Excessive daytime sleepiness, sleep attacks, sleep paralysis, hypnogogic hallucinations, cataplexy (sudden transient episodes of loss of tone with preserved consciousness, triggered by strong emotions), onset in childhood or young adulthood. *Recommended tests to support diagnosis: polysomnography, multiple sleep latency test.*
Transient ischemic attack	Usually presents with negative symptoms (weakness, aphasia, sensory loss), preserved consciousness, can last minutes to hours, cardiovascular risk factors present. *Recommended tests to support diagnosis: brain imaging, carotid ultrasound, ECG, echocardiogram, other tests depending on suspected etiology.*
Migraines	Auras include visual, sensory, motor, and speech symptoms, headache is often unilateral, throbbing, worsens with exercise, associated with nausea/vomiting, phono/photophobia, fully reversible weakness in hemiplegic migraines. *Recommended tests to support diagnosis: clinical diagnosis, imaging as clinically indicated.*

Transient global amnesia	Repetitive questions demonstrating poor recent memory, can last hours, can carry out complex tasks. *Recommended tests to support diagnosis: imaging as clinically indicated.*
Paroxysmal movement disorder	Movements can include dyskinesias, dystonia, myoclonus, ballismus, tics, tremors; there is preserved consciousness, movements improve during sleep, some may be triggered by exercise. *Recommended tests to support diagnosis: imaging as clinically indicated.*
Vestibular disorder	Severe vertigo, tinnitus, preserved awareness. *Recommended tests to support diagnosis: vestibular examination.*
Breath holding spells	Occurs in young children (1–3 years old), brief loss of consciousness, no postevent confusion, pallor and cyanosis may be present. *Recommended tests to support diagnosis: clinical diagnosis.*

Other physiologic paroxysms can resemble epileptic seizures and should be considered in the differential diagnosis of PNES and epileptic seizures. Some may call these paroxysms “physiologic nonepileptic seizures,” substituting the “psychogenic” prefix with the word “physiologic.” Table 2.4–14 lists common physiologic paroxysms and how to clinically distinguish and diagnose them.

Psychiatric comorbidities are common in PNES. Mood disorders, for instance, are frequent in PNES but not significantly more frequent than in epilepsy. Prior exposure to traumatic experiences, PTSD, and personality disorders are usually more frequent in PNES than in epilepsy. Comorbid epilepsy (electrographically and clinically confirmed) is variably reported in 10% to 20% of adult patients with PNES, and the comorbidity rate is higher in pediatric patients. Conversely, PNES is reported in approximately 10% to 20% of adult epilepsy patients. Mild traumatic brain injury, multiple allergies, and somatic symptom disorder, including chronic pain syndromes, are commonly reported in PNES. While important to formulate a diagnosis and treatment plan, clinicians should not assume a

PNES diagnosis solely based on the presence of comorbid psychiatric diagnoses such as PTSD or personality disorders.

Pathophysiology

The underlying pathophysiology of FNSD, of which PNES is a subtype, involves alterations in the activity and functional connectivity in brain regions involved in body awareness, emotion processing, self-monitoring, cognitive control, agency and motor intention, initiation, and inhibition. These findings are restricted to a few studies with phenotypically heterogeneous subjects but provide insight into a nascent neurobiologic explanation for PNES. Conversion and dissociative hypotheses have predominantly explained the phenomenon psychologically. A more modern and integrative cognitive hypothesis describes PNES as the result of automatic, reflex-like execution of a learned mental representation of seizures, typically in the context of dysfunction of inhibitory processes due to chronic stress, arousal, and other factors that compromise high-level cognitive processes.

Treatment

Once the diagnosis of PNES is confirmed, it has to be clearly communicated to patients, family members, and involved clinicians. Communication of the diagnosis can lead to improvement in symptom frequency or severity and medical utilization for a small subset of patients, at least temporarily. Elements to be covered during the diagnosis explanation include (1) reassurance that symptoms are genuine and not considered fabricated; (2) explanation of how the diagnosis was arrived at and names for the condition; (3) discussion of hypothetical mechanisms for the disease (including a neurobiologic model) and potential role of identified risk factors if any; and (4) review of effective treatments. Multidisciplinary collaboration is essential for a successful outcome. The diagnosing neurologist can help manage comorbid neurologic conditions (such as headaches or epileptic seizures), manage or taper unnecessary medications (such as antiseizure medications), and provide recommendations related to work and driving restrictions. Mental health clinicians are best suited to deliver treatment.

Evidence-based treatment for PNES favors the use of psychotherapy. Multicenter RCTs have shown the efficacy of CBT in a number of functioning and symptom severity measures. Other psychotherapeutic interventions studied in PNES in either small RCTs or uncontrolled trials with promising results include paradoxical intention (using imagery exposure to induce PNES), operant conditioning (positive reinforcement and punishment through withdrawal of privileges), long-term interpersonal psychodynamic psychotherapy, group psychotherapy, mindfulness-based psychotherapy, CBT-based inpatient programs, and prolonged exposure for PNES with comorbid PTSD.

There is no convincing evidence that pharmacologic treatment can reduce PNES frequency and/or severity. Psychopharmacologic interventions should be limited to the treatment of comorbid psychiatric conditions. Noninvasive stimulation therapies lack sufficient evidence for the treatment of PNES at this point, with case reports highlighting the need for further exploration.

Outcomes

Most currently available data on long-term outcomes do not consider participation in evidence-based treatment. Participation in psychotherapy is, however, associated with improved PNES frequency 1 to 2 years after diagnosis. Disability rates do not decrease years after the diagnosis. Poor outcomes are correlated with comorbid epilepsy and psychiatric diagnoses, dramatic seizure semiology, and younger age of onset, while better outcomes are correlated with higher education, shorter duration of illness, employment, acceptance of the diagnosis, and social support.

FURTHER READINGS

Agrawal N, Mula M. Treatment of psychosis in patients with epilepsy: an update. *Ther Adv Psychopharmacol*. 2019;9:1–10.

Barry JJ, Ettinger AB, Friel P, Gilliam FG, Harden CL. Advisory group of the epilepsy foundation as part of its mood disorder: consensus statement: the evaluation and treatment of people with epilepsy and affective disorders. *Epilepsy Behav*. 2008;13(Suppl 1):S1–S29.

*Baslet G, Bajestan SN, Aybek S, et al. Evidence-based practice for the clinical assessment of psychogenic nonepileptic seizures: a report from the American Neuropsychiatric Association Committee on Research. *J Neuropsychiatry Clin Neurosci*. 2021;33:27–42.

Baslet G, O’Neal MA, Dworetzky BA. Functional neurological symptom disorders. In: Silbersweig DA, Safar LT, Daffner KR, eds. *Neuropsychiatry and Behavioral Neurology: Principles and Practice*. McGraw-Hill; 2021:361–378.

Beghi E, Giussani G, Nichols E, Abd-Allah F, et al. Global, regional, and national burden of epilepsy 1990-2016: a systematic analysis for the Global Burden of Disease Study

2016. *Lancet Neurol*. 2019;18(4):357–375.
*Blumer D, Montouris G, Davies K. The interictal dysphoric disorder: recognition, pathogenesis, and treatment of the major psychiatric disorder of epilepsy. *Epilepsy Behav*. 2004;5:826–840.
Chen B, Choi H, Hirsch LJ, et al. Psychiatric and behavioral side effects of antiepileptic drugs in adults with epilepsy. *Epilepsy Behav*. 2017;76:24–31.
*Clancy MJ, Clarke MC, Connor DJ, Cannon M, Cotter DR. The prevalence of psychosis in epilepsy; a systematic review and meta-analysis. *BMC Psychiatry*. 2014;14:75.
Conway CR, Udaiyar A, Schachter SC. Neurostimulation for depression in epilepsy. *Epilepsy Behav*. 2018;88S:25–32.
De Toffol B, Trimble M, Hesdorffer DC, et al. Pharmacotherapy in patients with epilepsy and psychosis. *Epilepsy Behav*. 2018;88:54–60.
Durrant J, Rickards H, Cavanna AE. Prognosis and outcome predictors in psychogenic nonepileptic seizures. *Epilepsy Res Treat*. 2011:274736.
Ettinger AB. Psychotropic effects of antiepileptic drugs. *Neurology*. 2006;67:1916–1925.
Fisher RS, Cross JH, French JA, et al. Operational classification of seizure types by the International League Against Epilepsy: Position paper of the ILAE Commission for Classification and Terminology. *Epilepsia*. 2017;58(4):522–530.
Goldstein LH, Robinson EJ, Mellers JDC, et al; CODES study group. Cognitive behavioural therapy for adults with dissociative seizures (CODES): a pragmatic, multicentre, randomised controlled trial. *Lancet Psychiatry*. 2020;7:491–505.
Hall-Patch L, Brown R, House A, et al; NEST collaborators. Acceptability and effectiveness of a strategy for the communication of the diagnosis of psychogenic nonepileptic seizures. *Epilepsia*. 2010;51:70–78.
Hamid H, Blackmon K, Cong X, et al. Mood, anxiety, and incomplete seizure control affect quality of life after epilepsy surgery. *Neurology*. 2014;82(10):887–894.
*Hesdorffer DC, Ishihara L, Mynepalli L, Webb DJ, Weil J, Hauser WA. Epilepsy, suicidality, and psychiatric disorders: a bidirectional association. *Ann Neurol*. 2012;72(2):184–191.
Irwin LG, Fortune DG. Risk factors for psychosis secondary to temporal lobe epilepsy: a systematic review. *J Neuropsychiatry Clin Neurosci. Winter* 2014;26(1):5–23.
Josephson CB, Jetté N. Psychiatric comorbidities in epilepsy. *Int Rev Psychiatry*. 2017;29(5):409–424.
Kawakami Y, Itoh Y. Forced normalization: antagonism between epilepsy and psychosis. *Pediatr Neurol*. 2017;70: 16–19.
LaFrance WC Jr, Baird GL, Barry JJ, et al; NES Treatment Trial (NEST-T) Consortium. Multicenter pilot treatment trial for psychogenic nonepileptic seizures: a randomized clinical trial. *JAMA Psychiatry*. 2014;71(9):997–1005.
*LaFrance WC Jr, Baker GA, Duncan R, Goldstein LH, Reuber M. Minimum requirements for the diagnosis of psychogenic nonepileptic seizures: a staged approach: a report from the International League Against Epilepsy Nonepileptic Seizures Task Force. *Epilepsia*. 2013;54(11):2005–2018.
Lopez MR, Schachter SC, Kanner AM. Psychiatric comorbidities go unrecognized in patients with epilepsy: “you see what you know.” *Epilepsy Behav*. 2019;98:302–305.
Michaelis R, Tang V, Goldstein LH, et al. Psychological treatments for adults and children with epilepsy: Evidence-based recommendations by the International League Against

Epilepsy Psychology Task Force. *Epilepsia*. 2018;59(7):1282–1302.
Michaelis R, Tang V, Wagner JL, et al. Psychological treatments for people with epilepsy. *Cochrane Database Syst Rev*. 2017;10:CD012081.
Mula M. Pharmacological treatment of anxiety disorders in adults with epilepsy. *Expert Opin Pharmacother*. 2018;(17)19:1867–1874.
Rathore C, Henning OJ, Luef G, Radhakrishnan, K. Sexual dysfunction in people with epilepsy. *Epilepsy Behav*. 2019;100 (Pt A):106495.
Reuber M, Brown RJ. Understanding psychogenic nonepileptic seizures-phenomenology, semiology and the integrative cognitive model. *Seizure*. 2017;44:199–205.
Scott A, Sharpe L, Hunt C, Gandy M. Anxiety and depressive disorders in people with epilepsy: a meta-analysis. *Epilepsia*. 2017;58(6):973–982.
Seethalakshmi R, Krishnamoorthy ES. Depression in epilepsy: phenomenology, diagnosis and management. *Epileptic Disord*. 2007;9(1):1–10.
Tolchin B, Dworetzky BA, Martino S, Blumenfeld H, Hirsch LJ, Baslet G. Adherence with psychotherapy and treatment outcomes for psychogenic nonepileptic seizures. *Neurology*. 2019;92:e675–e679.

2.5 Neuropsychiatric Consequences of Traumatic Brain Injury

Melissa B. Jones, M.D., Amy R. Corcoran, M.D., Joanne A. Byars, M.D., and Ricardo E. Jorge, M.D.

Traumatic brain injury (TBI) is defined as an acute disruption of brain function and/or structure induced by an external force to the head. In the United States, an estimated 1.7 million TBIs occur annually and contribute to 30.5% of injury-related deaths. For survivors of TBI, the cognitive and neuropsychiatric sequelae of TBI contribute to postinjury disability, poorer functional recovery, and increased health care utilization.

This chapter summarizes the emotional, cognitive, and behavioral sequelae associated with TBI most pertinent to psychiatric practice. The chapter begins with brief descriptions of the epidemiology and pathophysiology of TBI, followed by an overview of neuroimaging, biomarker, electrophysiology, and neuroendocrine-based assessments of its neurobehavioral outcomes. We then introduce TBI severity grading as a precursor to understanding the range and patterns of cognitive sequelae following TBI of different severities. Finally, the chapter concludes with a section on the clinical features and management of individual neuropsychiatric disorders post TBI.

EPIDEMIOLOGY OF TBI

In 2014, approximately 2.87 million TBI-related emergency department visits, hospitalizations, and deaths occurred in the United States. Direct and indirect costs of TBI exceed $76 billion per year (expressed in 2010 U.S. dollars).

Most TBIs occur among infants, adolescents/young adults, and elderly individuals. Males are 40% more likely to suffer brain injury than females. African Americans and individuals of low socioeconomic status also experience higher rates of TBI than other groups. In the United States, falls represent the most common cause of TBI, both overall and in the specific populations of children and older adults. Among young adults, motor vehicle accidents and assaults produce the most TBIs.

TBI may be categorized as closed or penetrating, depending on the integrity of the meningeal coverings. The most common cause of penetrating injuries is gunshot wounds. However, closed head injuries constitute the vast majority of TBIs. TBI may also be categorized by severity (see “Categorization of TBI Severity” below). Approximately 80% of patients with TBI have mild head injury, 10% have moderate head injury, and the remaining 10% severe.

Patients with TBI have a greater risk of recurrent TBI than noninjured controls; about 10% will sustain a second TBI within the year following the initial event. The single greatest risk factor, however, is alcohol/drug abuse. One epidemiologic study reported that close to one-third of brain injury patients had an identifiable alcohol problem before the trauma, and more than 50% were intoxicated at the time of injury.

In general, military service members experience higher rates of TBI compared to civilian populations. TBI was coined the “signature wound” of Operation Iraqi Freedom (OIF), Operation Enduring Freedom (OEF), and Operation New Dawn (OND) conflicts involving U.S. troops in Iraq and Afghanistan. Approximately 10% of veterans of these conflicts seeking care at Veterans Affairs (VA) medical facilities experienced a TBI while deployed. While military-related TBI can occur from training activities, sports, and motor vehicle accidents, blast exposure represents the most common cause of deployment-related TBI in veterans of recent eras. For instance, improvised explosive devices (IEDs) or mines produced the majority of injuries (97%) observed among a Marine unit in Iraq. In addition, an estimated 59% of soldiers with blast injuries sustain a TBI.

MECHANISMS OF TBI

The definition of TBI requires damage inflicted by an external force to the head. Such external forces may include rapid acceleration/deceleration, being struck in the head by an object, penetration of a foreign object into the brain, or blast-related forces. Trauma-induced neuroimaging and neuropathologic abnormalities may also be categorized as primary or secondary injuries. Primary injuries refer to the damage occurring immediately at the time of injury. Secondary brain damage refers to physiologic processes initiated by the injury but spanning a variable period following the traumatic episode (such as edema, seizures, and hydrocephalus).

Closed injuries produce regional patterns of injury due to variations in brain tissue densities, elasticities, and proximity to bony and fibrous restrictions of the skull and dura, respectively. On the other hand, penetrating injuries destroy, displace, and/or deposit bone fragments or other infectious materials into bone or brain tissue. Regional vulnerabilities to penetrating injuries depend on such features as the object's location of entry, trajectory, and size. While this section provides an overview of closed injuries, it is worth noting that patients may experience a combination of closed and penetrating-type injuries. For example, a blast exposure from an IED may induce combinations of closed and penetrating TBIs (see the section on blasts below).

When linear and/or rotational forces are applied to the head, inertia causes the brain's movement to lag relative to that of the skull. The bony protuberances of the skull and rigid, tough fibers of the falx and tentorium cerebri may deform and damage brain parenchyma. Furthermore, the brain's elaborate and complex neurovascular and axonal structures are particularly susceptible to rotational and shearing–tensile forces.

This deformation, shearing, and/or stretching of blood vessels may lead to intracranial and intracerebral hemorrhages that can expand over an interval of hours to days after the injury. Contusions may occur at the crest of cerebral gyri, with a predilection for the frontal and temporal lobes (housed within the anterior and middle fossae, respectively).

Vascular shearing injury may also produce intraparenchymal hemorrhages visualized with computed tomography (CT) or magnetic resonance imaging (MRI); their presence indicates that enough force was applied to the brain to induce underlying traumatic axonal injury (see Fig. 2.5–1). Other vascular injuries include intracranial–extracerebral hemorrhages in the epidural (i.e., between the skull and the dura), subdural (i.e., between the dura and the arachnoid), or subarachnoid spaces. Finally, ischemia may occur from other traumatic vascular lesions (e.g., carotid arterial dissection), intracranial hypertension, or vasospasm following subarachnoid hemorrhage.

Traumatic axonal injury, also known as “diffuse axonal injury,” refers to multifocal regions of white matter damage caused by traumatic rotational and tensile-shearing forces. Injured regions commonly include upper brainstem and diencephalic structures, gray–white matter junctions, the parasagittal white matter of the cerebrum, and the corpus callosum. The latter white matter tract is particularly sensitive to TBI due to its location near the falx. Although a traumatic axonal injury is usually associated with immediate and persistent coma, lesser degrees of axonal disruption may also be seen in patients who suffered mild brain injuries (Fig. 2.5–1).

The injuries produced at the time of impact may continue to expand via several mechanisms. For instance, increasing edema may augment tissue deformation, raise intracranial pressure, reduce cerebral perfusion pressure, and ultimately cause ischemia and/or herniation. Another example of progressive injury may include hydrocephalus resulting from an intraventricular expansion of hemorrhage.

The macroscopic primary and secondary injuries described above reflect a complex series of microscopic cellular cascades initiated by the trauma. Mechanical deformation of neurons induces ion homeostasis disruptions, excessive release of glutamate and other neurotransmitters, intracellular calcium release, and the activation of cytotoxic cascades. Other complex delayed post-TBI processes include microglial activation, the release of inflammatory cytokines, and alterations in cerebral blood flow.

Evidence from experimental models suggests that TBI generates widespread neuropathologic changes in diverse cortical areas, subcortical nuclei, and white matter tracts. Neuronal damage and cell loss may progress over weeks to months after the initial insult in selectively vulnerable regions of the prefrontal cortex, hippocampus, thalamus, striatum, amygdala, and forebrain nuclei. Damage to ascending biogenic amine and cholinergic pathways may also influence the pathophysiologic processes contributing to neuropsychiatric presentations and the long-term outcome of patients with TBI. Furthermore, network-wide disruptions may result from Wallerian degeneration, diaschisis, or repair and regeneration mechanisms, such as synaptogenesis and axonal sprouting, altering network organization.

Blast Injuries

TBI from blast exposure occurs via several mechanisms. Primary blast injury results from the initial pressure wave of the explosion causing mechanical tissue damage and the release of heat, acoustic, and electromagnetic energies. Further injury may also be caused by the reflection of the pressure wave when it hits the skull. Primary blast injuries are unique to blasts. Of note, air–fluid interfaces of the tympanic membrane, heart, lungs, and gastrointestinal tract are potentially susceptible to blasts. Lung injury accounts for a high percentage of blast fatalities due to the disruption of the alveoli–capillary interface.

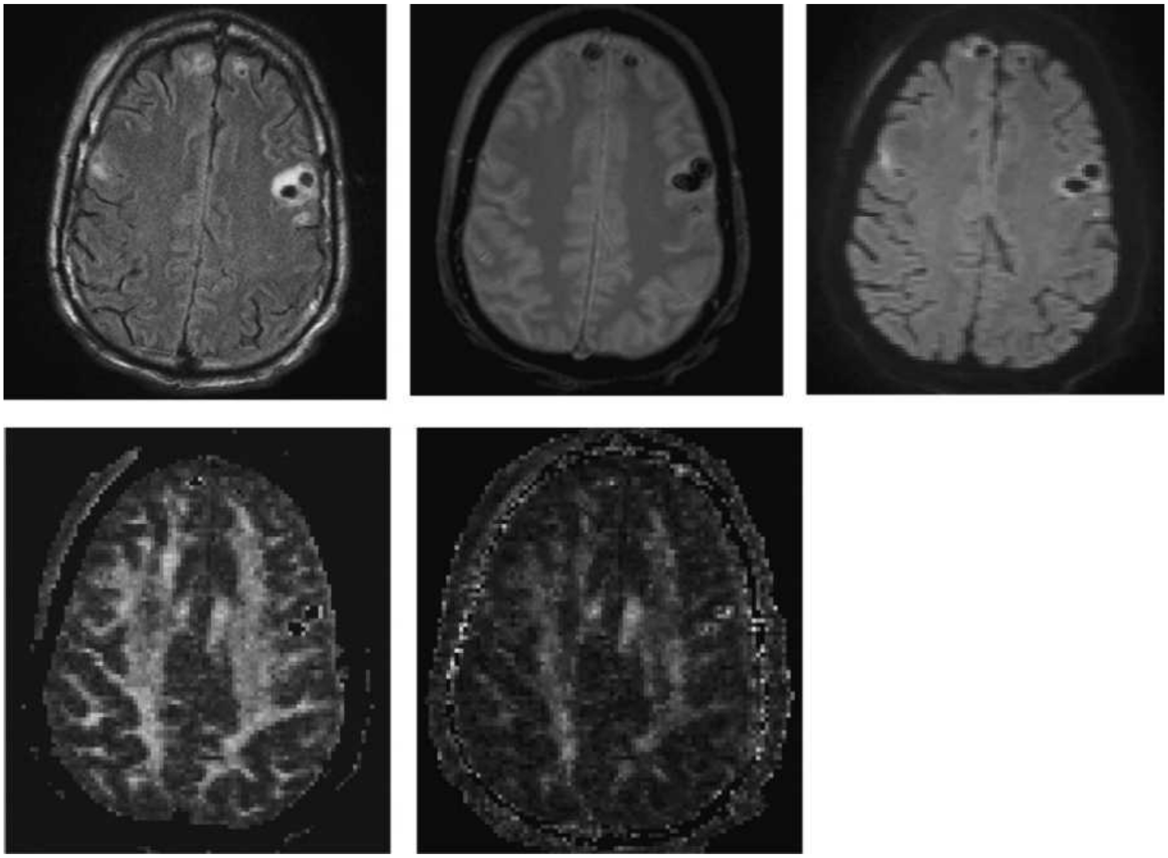

FIGURE 2.5–1. Diffuse axonal injury. Magnetic resonance images consistent with the presence of diffuse axonal injury at the gray–white matter junction and frontal intracerebral hemorrhages.

Secondary mechanisms of injury may result from the release of high-energy fragments that may penetrate the skull, eye, or other body parts. Tertiary injury refers to injuries occurring from acceleration/deceleration or blunt-force injuries (e.g., from being thrown from the blast and hitting a solid object). Tertiary injury may result in a closed TBI via a similar mechanism to that of a motor vehicle accident. Quaternary injury is caused by toxic or radiological substances emitted from the blast. Gas inhalation or toxin-induced burns may result from this type of injury.

The growing number of TBIs incurred by soldiers deployed to the wars in Iraq and Afghanistan has heightened interest in understanding mechanisms of blast injury. Research groups are actively investigating thresholds required to induce blast-related TBI and the potential for unique neuropsychiatric outcomes associated with blasts. For instance, the Translational Research Center for TBI and Stress Disorders (TRACTS), a multi-VA cohort study of previously deployed, post-9/11 veterans, is actively collecting structured histories of blast and other TBI exposures and longitudinal psychiatric, cognitive, and neuroimaging data to understand long-term outcomes of blast injuries and optimize the assessment of blast-related TBI.

NEUROIMAGING AND LABORATORY INVESTIGATIONS

This section provides a brief summary of neuroimaging techniques, serum biomarkers, electrophysiologic studies, and neuroendocrine evaluations of brain alterations associated with TBI. These modalities have contributed to current understandings of the pathophysiology of injury and neural circuits implicated in its acute and chronic neuropsychiatric sequelae.

Neuroimaging

CT and MRI are routinely incorporated into the clinical evaluations of patients with TBI. Common data elements guiding technical considerations and definitions of TBI-related neuroimaging findings were published in 2010 and have aided in the standardization of data reporting for clinical care and outcomes research.

CT is the most efficient means of detecting injuries requiring neurosurgical intervention. Thus, CT remains the initial study of choice in the emergency room (ER). Structural abnormalities visualized on CT include skull fractures, epidural or subdural hematomas, subarachnoid hemorrhage, intraparenchymal contusions, and mass effect producing herniation. The Marshall and Rotterdam CT classification systems utilize these CT abnormalities to predict outcomes (e.g., mortality) following more severe TBIs.

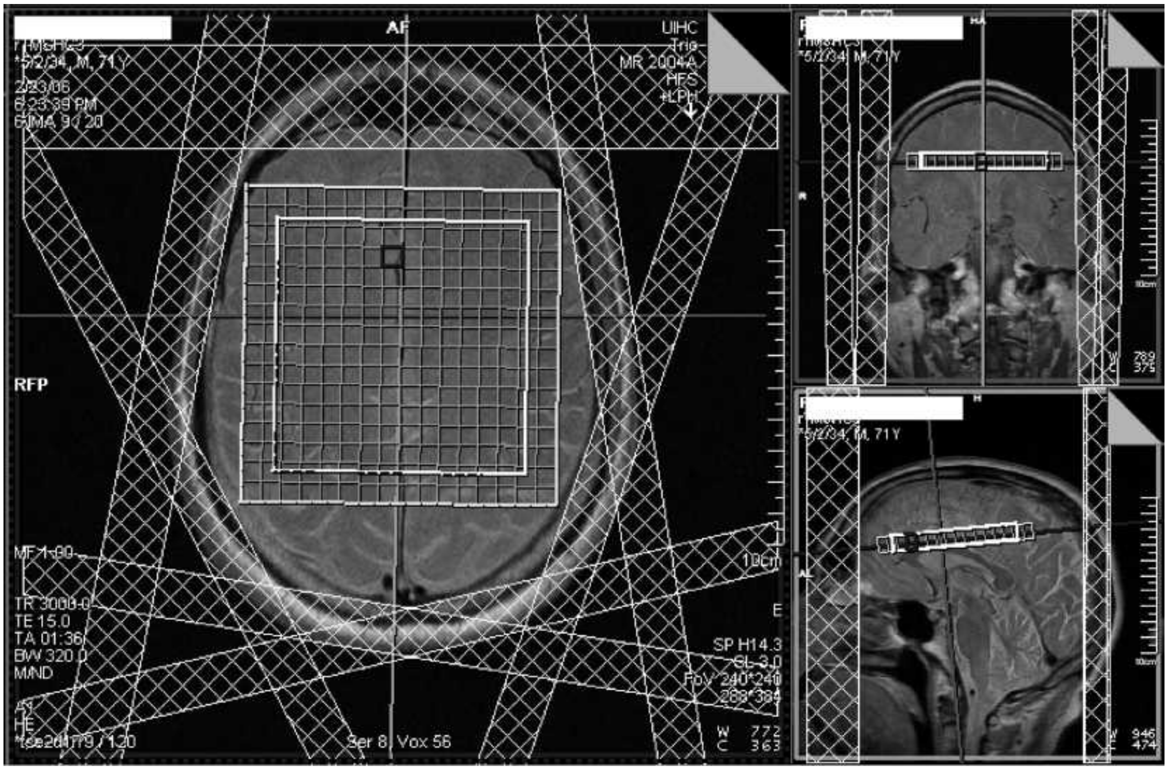
[redacted]
AF
UIHC
Trio
MR 2004A
HFS
+LPH
*5/2/34, M, 71Y
2/23/06
6:23:39 PM
RFP
MF 1.00
TR 3000.0
TE 15.0
TA 01:36
BW 320.0
M/ND
10cm
E
SP H14.3
SL 3.0
FoV 240*240
288*384
W 772
C 363
Ser 8, Vox 56
[redacted]
*5/2/34, M, 71Y
W 789
C 375
[redacted]
*5/2/34, M, 71Y
W 946
C 474

FIGURE 2.5–2. Proton magnetic resonance spectroscopy (dorsal–rostral cingulate gyrus PRESS CSI).

MRI is more sensitive than CT for visualizing the extent of acute and chronic post-traumatic lesions (Fig. 2.5–1), such as small hemorrhages associated with traumatic axonal injury. MRI is also the study of choice for visualizing volumetric changes (e.g., hippocampal atrophy) over time. The use of diffusion-weighted imaging (DWI), apparent diffusion coefficient (ADC), gradient echo (GRE), and susceptibility-weighted imaging (SWI) sequences also improves the detection of axonal injury. Finally, CT angiography (CTA), MR angiography (MRA), and catheter angiography may be required to visualize traumatic vascular injuries, such as arterial dissections and traumatic aneurysms.

CT and routine MRI lack the sensitivity to identify the whole extent of the pathologic changes associated with TBI. Structural volumetric MRI studies, however, provide quantitative data on volumetric changes observed in gray matter, white matter, and cerebrospinal fluid. Volumetric MRI may be particularly useful for monitoring the progression of atrophy over time.

Diffusion tensor imaging (DTI) enhances the sensitivity of MRI to detect white matter changes after TBI. DTI provides an objective measure of directional water diffusion in space (e.g., fractional anisotropy [FA]), which in white matter regions reflects the integrity of myelin sheaths. DTI reveals changes in the white matter that correlate with the severity of TBI and with the functional outcome at the time of hospital discharge. Diffusion tensor tractography, which allows visualization and analysis of specific white matter pathways, may extend the potential of DTI in TBI. However, DTI and tractography techniques are primarily reserved for the research setting due to varying methodologies among investigators and lacking normative data.

Functional imaging studies, such as functional MRI (fMRI), positron emission tomography (PET), single-photon emission computed tomography (SPECT), and magnetic resonance spectroscopy (MRS), are also used in research settings to detect in vivo brain changes not captured by structural imaging. Both task-based and resting-state fMRI has been used to explore the integrity of neural networks subserving cognitive, emotional, and behavioral functions post TBI. For instance, in task-based fMRI studies of working memory and attention, individuals with TBI show different patterns of regional brain activation than healthy controls, even when performance on these cognitive tasks does not substantially differ from normal. This suggests that injury-related changes in the allocation of working memory and attentional resources might underlie some of the cognitive complaints after mild head injury.

In addition, Fluorodeoxyglucose (FDG) - PET and SPECT examine regional differences in metabolism and blood flow, respectively. Like fMRI, FDG-PET and SPECT may identify circuits involved in processing cognitive and emotional information, including regions activated during the implementation and monitoring of behavioral responses.

MRS techniques enable the simultaneous generation of maps of metabolite distributions over wide brain regions (Fig. 2.5–2). Multislice short echo time (TE) H_1-MRS allows for the in vivo evaluation of the metabolic profiles of large brain regions, including the neocortex, deep white matter, and basal ganglia. Specific metabolites studied with H_1-MRS include *N*-acetyl aspartate (NAA), a marker of neuron density and metabolism; choline-containing compounds (Cho), which reflect phospholipid metabolism; creatine-containing compounds (Cre), which reflect the cellular energetic status; myoinositol (mI), a putative glial cell marker; the sum of glutamate and glutamine (Glx); lactic acid (Lac); and other peaks. In the research setting, MRS has demonstrated associations between metabolic abnormalities and cognitive, emotional, and behavioral changes in patients with otherwise normal-appearing white matter after TBI.

Biomarkers

The identification of cerebrospinal fluid and serum biomarkers of TBI severity and prognostication has inspired much research. However, the interpretation of biomarker levels requires consideration of their specificity for brain tissue damage, temporal kinetics, and optimal cut-off values.

S100B, a calcium-binding protein involved in cell cycle progression and differentiation, is the most studied of these biomarkers. Serum S100B measurements have been incorporated into some national guidelines for triaging CT imaging in the ER following a mild TBI. Compared to other biomarkers, its relatively shorter half-life may enable serial sampling to detect the progression of secondary injury in the acute setting. However, this marker is nonspecific for cerebral injury, having also been studied as a biomarker for idiopathic psychiatric disorders, for instance. Glial fibrillary acidic protein (GFAP), a protein expressed in astrocytes, is more specific to brain tissue but is cleared less rapidly than S100B.

Recently, investigators of the Collaborative European Neurotrauma Effectiveness Research (CENTER-TBI) study compared the incremental value of measuring serum biomarkers to clinical CT decision-making rules for predicting day-of-injury CT abnormalities. Compared to other serum biomarkers (S100B, neuron-specific enolase, ubiquitin C-terminal hydrolase-L1, neurofilament light chain, total tau), GFAP had the highest discriminative value for predicting CT abnormalities, with a 99% likelihood of improved discrimination of CT-positive patients compared to clinical decision-making rules. GFAP also enhanced the prediction of CT abnormalities in a subset of patients with mild TBI. These research efforts may contribute to our understanding of TBI prognostication and severity categorization.

Electrophysiologic Studies

Conventional electroencephalography (EEG) recordings are currently used in trauma intensive care units for monitoring procedures and for brain death diagnosis. EEG is also invaluable in the diagnosis of status epilepticus and is the primary diagnostic procedure to localize a post-traumatic epileptic focus. The use of nasopharyngeal, anterior temporal, or sphenoidal leads increases the chance of finding an abnormality on routine EEG. Video-EEG monitoring and 24-hour ambulatory recordings aid in the differential diagnosis of patients presenting with unclear paroxysmal behavioral disturbances. Quantitative EEG (Q- EEG) is currently used as an adjunctive diagnostic technique in the evaluation of slow-wave abnormalities associated with brain injuries and in the diagnosis of post-traumatic temporal lobe epilepsy.

Polysomnography permits the diagnosis of sleep disturbances that may occur in patients with TBI. These include sleep-disordered breathing, nocturnal epilepsy, rapid-eye-movement sleep behavior disorder, sleep-related movement disorders, and narcolepsy. The latter may also require a multiple sleep latency test.

Visual-, auditory-, and somatosensory-evoked potentials help localize the level of injury to the central nervous system. Auditory-evoked potentials are particularly effective in detecting brain stem pathology. More complex electrophysiologic responses may provide insight into the integrity of neural systems subserving cognitive and emotional functions post TBI. For example, one study using P50 physiology found evidence of deficits in sensory gating mechanisms in patients with TBI.

Neuroendocrine Testing

Hypopituitarism may contribute to the neuropsychiatric sequelae in the acute and chronic phases of TBI. The pooled prevalence estimate of hypopituitarism following TBI is 29%. However, rates vary widely depending upon TBI severity, time post injury, hormones of study, and testing methodologies.

Hypopituitarism may result from primary or secondary injury mechanisms. In terms of the former, traumatic forces may directly injure the pituitary gland, stalk, hypothalamus, and/or their complex vasculatures. In terms of the latter, elevated intracranial pressure, hypotension, and hypoxia may lead to pituitary ischemia and infarction.

During the acute phase following a mild-complicated, moderate, or severe TBI, screening for hypocortisolism (via early morning cortisol) may prevent hypotension, hyponatremia, and increased vasopressor needs. A posterior pituitary injury may also be suspected in patients presenting with potentially life-threatening hypernatremia and hypotonic polyuria from diabetes insipidus. Conversely, the syndrome of inappropriate antidiuretic hormone (SIADH) may cause hyponatremia in the acute setting.

In the chronic phase, the risk of developing hormonal deficiencies appears to be highest within the first 12 months postinjury; however, some deficiencies may spontaneously remit while others begin at later time points. The emergence of the syndrome of growth hormone deficiency (GHD) in adults—characterized by decreased strength, energy, and sense of well-being—highlights the potential importance of such deficiencies. Estimated rates of GHD following TBI range from 10% to 30%, making it likely the most common endocrinopathy. Rates of luteinizing hormone (LH), follicle-stimulating hormone (FSH), adrenocorticotropic hormone (ACTH), thyroid-stimulating hormone (TSH), and prolactin abnormalities range from 2% to 17%. About 3% to 5% of patients with TBI may show multiple neuroendocrine abnormalities.

Currently available recommendations for screening in the subacute and chronic phases apply to mild-complicated, moderate, and severe TBIs. Recommendations for appropriate testing after mild, uncomplicated TBI require further research. According to some guidelines, screening for anterior pituitary deficiencies with dynamic ACTH testing (i.e., ACTH stimulation test), TSH levels, and FSH/LH levels is indicated at 3, 6, and 12 months. While GHD may be the most common anterior pituitary hormone abnormality post TBI, dynamic growth hormone testing (e.g., insulin tolerance test or, more recently, macimorelin test) is generally not recommended until 12 months postinjury, as its deficiency may be transient or appear later in the recovery course. After 12 months, if no hormonal deficiencies develop, then repeat testing is unnecessary in the absence of clinical symptoms.

CATEGORIZATION OF TRAUMATIC BRAIN INJURY SEVERITY

Head injuries span a wide range of severity, from patients who die at the time of trauma to those who do not require medical evaluation or assistance. Severity categorization, while arbitrary in several respects, provides a useful heuristic to predict the likelihood of persistent deficits and group patients into relatively homogenous strata for research purposes.

Most commonly, indices of TBI severity are based on the estimated durations of loss of consciousness (LOC) and post-traumatic amnesia

(PTA); presence or absence of neuroimaging findings; and Glasgow Coma Scale (GCS) scores. (see Table 2.5–1).

The GCS gives a quantitative estimate of the level of consciousness and neurologic status based on patterns of eye-opening, as well as best verbal and motor responses. GCS scores between 13 and 15 define mild brain injury; scores between 9 and 12 define moderate injury; and scores between 3 and 8 define severe injury. The best score in the 24 hours postinjury is a more useful indicator than the initial score immediately after injury (e.g., during a 45-second period of LOC that then resolves).

LOC refers to a period immediately after the injury where the patient is unresponsive and not awake. PTA refers to loss of memory for events preceding the injury or the injury itself (retrograde amnesia) and/or an inability to produce new memories immediately after the injury (anterograde amnesia). The term alteration of consciousness (AOC) typically refers to the disruption in brain function associated with mild TBI (discussed in the mild TBI section below); in the severity classification system presented in Table 2.5–1, it may also refer to cognitive deficits following resolution of PTA after a mild-complicated, moderate, or severe TBI.

Table 2.5–1.
Grading of TBI Severity

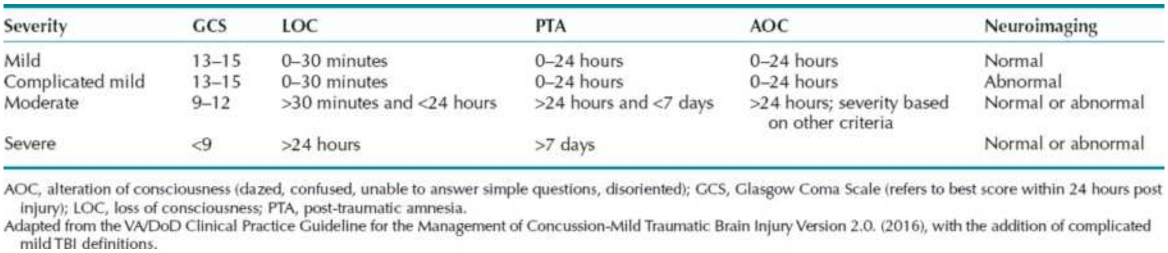

Severity	GCS	LOC	PTA	AOC	Neuroimaging
Mild	13–15	0–30 minutes	0–24 hours	0–24 hours	Normal
Complicated mild	13–15	0–30 minutes	0–24 hours	0–24 hours	Abnormal
Moderate	9–12	>30 minutes and <24 hours	>24 hours and <7 days	>24 hours; severity based on other criteria	Normal or abnormal
Severe	<9	>24 hours	>7 days		Normal or abnormal

AOC, alteration of consciousness (dazed, confused, unable to answer simple questions, disoriented); GCS, Glasgow Coma Scale (refers to best score within 24 hours post injury); LOC, loss of consciousness; PTA, post-traumatic amnesia.

Adapted from the VA/DoD Clinical Practice Guideline for the Management of Concussion-Mild Traumatic Brain Injury Version 2.0. (2016), with the addition of complicated mild TBI definitions.

Finally, neuroimaging studies (CT or MRI) may demonstrate the presence of structural brain lesions in some patients otherwise meeting the criteria for a mild TBI—these TBIs are then classified as mild-complicated. The trajectory of recovery following a mild-complicated TBI is akin to a moderate TBI, leading some researchers and clinicians to classify this subtype as a moderate TBI.

Caveats

There are several limitations to the above classification scheme worthy of mention. Durations of LOC, PTA, and AOC are useful for predicting the presence of abnormal imaging findings and persistent functional impairments post TBI. However, their arbitrary durations may prevent clinicians from viewing TBI severity as occurring on a spectrum. Instead, patients with mild TBI may also be conceptualized as having relatively briefer and less severe manifestations of the acute post-traumatic encephalopathy (discussed further below) that, by definition, fully recovers without sequelae. Other markers of injury, such as serum biomarkers, may improve the predictive value of this classification scheme, particularly for mild TBI.

Furthermore, establishing the durations of LOC, PTA, and AOC from retrospective self-report is challenging, especially when the TBI occurred decades earlier. Memory gaps occurring during a period of PTA may be misinterpreted as an LOC. In addition, patients and/or observers may not disentangle the effects of TBI from other potential causes of altered mental status, such as alcohol or substance intoxication, iatrogenic sedation, pain medications, or other neurologic complications (e.g., seizures, hypoxic–ischemic brain injury), and even severe emotional trauma.

Clinical Assessment of TBI Severity

In clinical practice, obtaining a step-by-step inquiry into what happened before, during, and after the injury, as well as obtaining observer accounts of the injury and any available records, may assist with determining the presence, severity, and duration of mental status alterations. This method of assessment may also assist with determining whether a patient was truly not awake and unresponsive (i.e., had LOC) after the injury. Other useful information may include descriptions of the medical care the patient received in the acute and subacute time frame after the injury; whether neuroimaging demonstrated abnormalities or was even performed; any concurrent alcohol or substance use at the time of the injury; and the course of cognitive and functional recovery postinjury.

From a practical standpoint, standardized questionnaires of TBI history, such as the Boston Assessment of TBI-Lifetime or Ohio State University TBI Identification Method, are rather lengthy for clinical purposes. Referral to neuropsychology, behavioral neurology and neuropsychiatry, or other

TBI subspecialists may further assist with diagnostic clarification and symptom evaluation.

COGNITIVE MANIFESTATIONS AND SEQUELAE OF TBI

Regional vulnerabilities to closed head injury translate into relative patterns of cognitive impairments and recovery of cognitive domains following TBI. The following sections provide an overview of these cognitive sequelae. Of note, these cognitive impairments begin immediately after the injury and may persist in accordance with injury severity (and other individual factors – see “Major and Minor Neurocognitive Disorder due to TBI”). Also, emotional and behavioral changes following TBI (particularly moderate and severe TBI) often co-occur with cognitive impairment and may reflect damage to shared neural circuitry. However, neuropsychiatric sequelae have been separated in this chapter for the purpose of clarity.

Acute Traumatic Encephalopathy

The term “acute traumatic encephalopathy” refers to the hierarchical recovery of cognitive domains and mental states emerging in the acute period following mild-complicated/moderate or severe TBIs. This nosology is typically employed to characterize the cognitive sequelae occurring immediately after a TBI requiring acute hospitalization. While patients with mild, uncomplicated TBI may also experience similar patterns of cognitive impairments (e.g., PTA), their durations are brief, and, by definition, recovery is relatively full (see Fig. 2.5–3). Persistent cognitive symptoms following mild TBI are discussed in more detail in a separate section.

Post-Traumatic Disorders of Consciousness

As stated above, when a patient experiences LOC, they are not responsive, their eyes are closed, and they are not awake. In the case of mild TBI, this may be transient. For severe TBIs, persistent and widespread damage, including to the reticular activating system and/or thalamocortical circuitry, may lead to a protracted disorder of consciousness, such as coma, persistent vegetative state, or minimally conscious state.

Coma is characterized by a lack of voluntary movement, behaviors limited to reflex activity; closed eyes; and absent sleep–wake cycles. A persistent vegetative state implies re-emergence of the sleep–wake cycle and spontaneous periods of eye-opening. Minimally conscious states are characterized by inconsistent but clearly present behavioral responses to environmental stimuli (i.e., visual tracking).

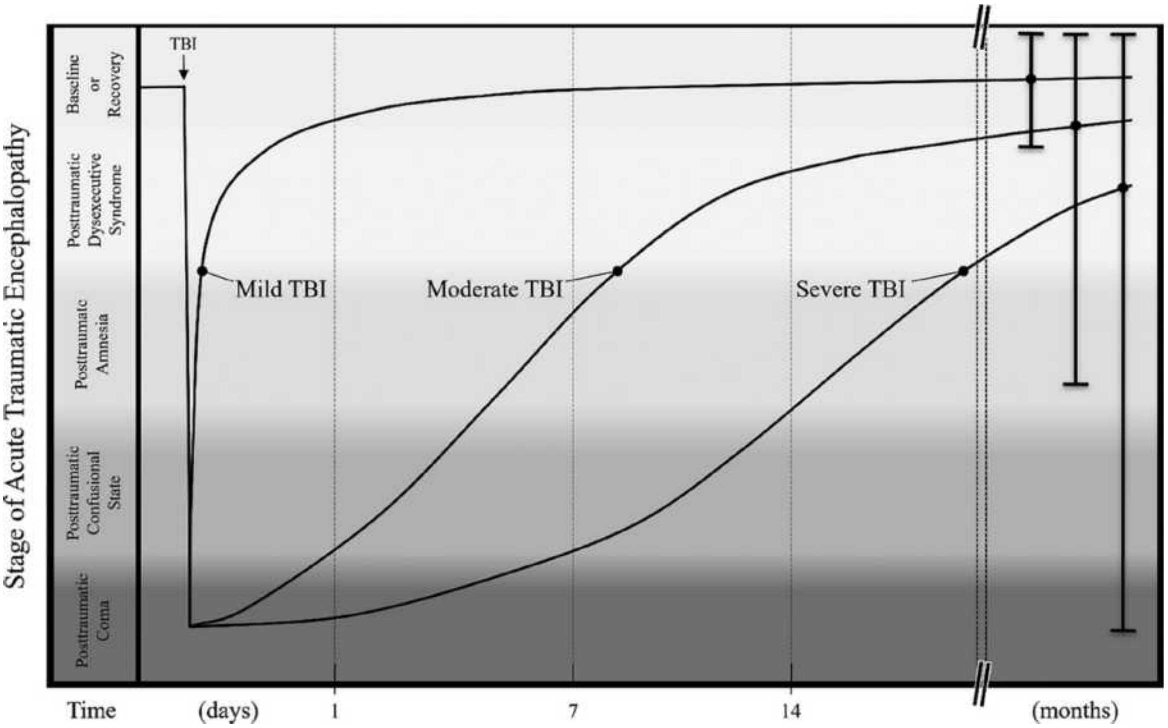

FIGURE 2.5–3. The stages and typical outcomes of the acute traumatic encephalopathy. Reprinted with permission from the Textbook of Traumatic Brain Injury, Third Edition, (Copyright © 2019). American Psychiatric Association. All Rights Reserved.

Post-Traumatic Confusional State

After emerging from a disorder of consciousness, patients can experience a post-traumatic confusional state akin to delirium. As such, patients demonstrate impaired simple and sustained attention and may experience perceptual disturbances (i.e., illusions or hallucinations), delusional thoughts, psychomotor agitation or retardation, and affective lability. Symptoms usually have an acute onset and a fluctuating course. Importantly, affective fluctuations and psychosis occurring during this confusional state should be distinguished from discrete episodes of idiopathic mood or psychotic disorders.

Clinical Vignette: Post-traumatic Acute Confusional State

A 19-year-old man was admitted to a trauma center after a motorcycle accident. He presented with a right epidural hematoma and bilateral contusions in the anterior temporal lobes. He also had a scalp laceration and right maxillary and zygomatic fractures. The postresuscitation GCS score was 8. Following the surgical evacuation of the hematoma, the patient was transferred to the intensive care unit. Two days later, the patient became restless and agitated. He removed his intravenous (IV) lines and monitoring devices and tried to get out of bed. He was disoriented, incoherent, and aggressive. His behavior suggested that he was experiencing frightening visual hallucinations. He appeared to grab at nonexistent objects in the environment and picked at the bedclothes. He responded well to a short course of a low-dose atypical antipsychotic agent.

Other potential conditions may also lead to the development of acute encephalopathy/delirium in patients with TBI and require consideration in the differential diagnosis. Other causes may include acute hydrocephalus, nonconvulsive seizures or status epilepticus, electrolyte imbalance, infections, medications (e.g., barbiturates, opiates, or steroids), and drug or alcohol withdrawal.

Post-Traumatic Amnesia

PTA, as in the above-mentioned classification scheme, refers to the period of recovery that most prominently features an impaired ability to form new memories after the injury and a loss of memory immediately before or during the event, despite the relative recovery of arousal and simple attention. Inability to remember the traumatic event, events immediately preceding the trauma, and the inability to form new memories after the trauma reflects injury to hippocampal systems subserving mnemonic functioning.

The duration and severity of PTA may be assessed using the Galveston Orientation and Amnesia Test (GOAT), which evaluates orientation to person, place, and time, as well as awareness of the accident and its consequences. Alternatively, retrospective structured questionnaires have shown an excellent correlation with this prospective determination.

Of note, measures of PTA duration tend to blur this phase of recovery with the duration of the post-traumatic confusional state, as attentional impairments occurring in the context of an encephalopathy/delirium also impede memory formation. Despite this, the duration of PTA has proved to be a good predictor of disability, vocational outcome, and severity of comportment changes following TBI.

Post-Traumatic Dysexecutive Syndrome

Following recovery from PTA, executive dysfunction tends to predominate (and frequently persists). This is due to the predilection of close head injuries for frontal–subcortical circuitry and their concomitant white matter connections.

Executive functions disrupted by TBI include goal formation, planning, selection of adequate response patterns, monitoring of ongoing behavior, and control of other cognitive domains. When confronted with demanding or novel environments, the adaptive functioning of patients with TBI is also often impaired.

Bedside tests of executive function—such as the Frontal Assessment Battery, Oral Trail Making Test, the Executive Interview 25, elements of the clock-drawing test, and verbal or figural fluency tests—may be employed to screen for executive deficits and/or their progression over time.

If a more detailed evaluation is appropriate, there are several neuropsychological tasks specifically designed to quantify executive deficits. The Wisconsin Card Sorting Test, the Goldberg Executive Control Battery, the Tower of London Test, and the Trail Making Test can measure impairments in goal formation, set-shifting, and monitoring of behavioral responses.

TBI-related impairments in executive attentional control may also manifest as restricted verbal or visuospatial attentional spans and altered vigilance patterns (i.e., sustained attention deficits). The most consistent findings, however, are associated with performance in the most demanding tasks (e.g., in divided attention paradigms such as the Paced Auditory Serial Addition Task).

During this acute (or chronically persisting) dysexecutive phase of recovery, patients and family members still frequently report “memory” problems. However, the nature of the memory impairment shifts from an inability to form new memories, as in PTA, to difficulty with retrieving memories, reflecting impaired executive function. Additional impairments may include faulty sequencing of events and inefficient encoding and storage strategies.

TBI also frequently affects executive control of language and social cognition (the latter is discussed further below). Patients with TBI may demonstrate higher-order language alterations and present with a defective narrative discourse, a lack of semantic coherence, and impaired pragmatics of communication. All of these result in impoverished and disorganized language and in reduced communication proficiency.

Other Cognitive Sequelae of Moderate–Severe TBI

Mild-complicated, moderate, and severe TBIs are more likely to produce persistent cognitive deficits than mild, uncomplicated TBIs. In addition to the deficits in arousal, attention, memory, and executive domains discussed above, processing speed is frequently impaired during acute and chronic recovery phases. Slowed processing speed may be conceptualized as resulting from damage to white matter tracts required for efficient information processing.

Patients post TBI, particularly those with right hemisphere involvement, frequently exhibit unawareness (anosognosia) or denial of deficits. Unawareness of deficits impedes realistic goal setting; interferes with the rehabilitation progress; and may put patients at risk of additional TBIs if they fail to recognize the need to take precautions to compensate for cognitive and other deficits.

While impairments in primary visuospatial, praxis, and language functions are not classically associated with closed TBI, focal damage from penetrating injuries or other complications may result in additional cognitive impairments beyond those typical for closed head injuries. For instance, a bullet passing through the right parietal lobe may produce visuospatial dysfunction and hemineglect. A large subdural hematoma compressing left parietal regions may produce apraxia or aphasia syndromes. Furthermore, subcortical forms of cortical syndromes may occur, such as subcortical aphasia manifesting from injury to left thalamic circuits.

Importantly, peripheral injuries, motivational syndromes, and primary motor or sensory impairments must be considered when interpreting the nature and extent of cognitive impairments. Neuropsychiatric and neuropsychological consultation can assist with the evaluation of these deficits and the development of rehabilitation strategies to aid in recovery.

Major or Mild Neurocognitive Disorder Due to TBI

Major or mild neurocognitive disorder due to TBI may be characterized by the incomplete recovery and persistence of cognitive impairments occurring during the “acute traumatic encephalopathy.” Thus, these protracted cognitive impairments may manifest as chronic disorders of consciousness; persistent alterations of arousal and attention (as seen in the acute post-traumatic confusional state); impaired new memory formation (as seen in acute PTA); persistent executive and processing speed abnormalities; and/or other chronic focal deficits resulting from additional injuries. For closed head injuries, visuospatial, praxis, and primary linguistic functions may be relatively preserved. Although beyond the scope of this chapter, post-traumatic parkinsonism or dystonia are also possible.

Clinical Vignette: Major Neurocognitive Disorder Due to TBI

A 40-year-old man sustained severe injuries in a motor vehicle accident and experienced a protracted coma of 1-month duration. MRI performed 2 weeks postinjury revealed the presence of widespread diffuse axonal injury. At 15-month follow-up, the patient presented with a mild left hemiparesis, right hemidystonia, and a left lower motor neuron facial palsy. Neuropsychological testing showed profoundly impaired frontal lobe functions and processing speed. Although the assessment was limited by apathy and slowed processing speed, his visuospatial, episodic memory, and linguistic skills were mildly impaired. Treatment with methylphenidate led to modest improvements in speed of information processing speed and apathy.

Of note, the degrees of neurobehavioral and functional recovery (and hence, the likelihood of developing a persistent cognitive disorder) are influenced by several factors unique to the individual experiencing the TBI. Beyond the severity of intracranial pathology, other negative prognosticators include older age; lower premorbid cognitive reserve; poorer socioeconomic statuses; limited social support; participation in legal or work compensation proceedings pertaining to the injury; and pre-

or postinjury neurologic/psychiatric conditions. Several genetic factors influencing preinjury cognitive and sensorimotor functions, injury-related inflammatory and vascular responses, and postinjury repair mechanisms may also play a role in cognitive outcome. For instance, several studies have demonstrated associations between the APOE-epsilon4 allele and a poor clinical outcome, implying genetic susceptibility to the effect of brain injury.

Mild Traumatic Brain Injury

The categorization of “mild” TBI implies that this form of head injury is at the least end of the severity spectrum of TBI. However, mild TBI severity can be quite broad, ranging from momentary confusion or disorientation to hours of PTA (see Table 2.5–1). The term “concussion” falls within the mild TBI categorization and can be used to imply the least severe form of TBI.

Akin to moderate and severe TBI, LOC and PTA may occur in the context of a mild TBI. AOC also can occur without associated LOC or PTA. While AOC definitions vary, patients present as dazed, confused, disoriented to time of day or current location, perseverative, incoherent, and unable to answer simple questions (e.g. that one plus two equals three). Note that AOC may easily be confused with surprise, fear, or substance intoxication.

While essentially all researchers and clinicians agree that moderate and severe TBI can produce significant and permanent cognitive changes, controversy exists as to whether mild TBI can as well. Patients may report concentration and memory deficits during the first weeks following mild TBI, and some studies have found objective cognitive deficits persisting up to 3 to 6 months postinjury. However, most of these patients spontaneously recover within the first 3 months.

The presence, severity, and functional contributions of persistent cognitive impairments occurring in the context of repetitive, subconcussive head injuries are also controversial. Protracted cognitive deficits are more frequent in patients with mild-complicated TBI. Of note, although CT is more likely to be performed in the ER than MRI, the latter has greater sensitivity for detecting evidence of shear injury. The development of serum biomarkers may also influence severity grading and prognostication of mild TBI. The long-term outcomes for patients with multiple mild TBIs and subconcussive blows are active areas of research.

Psychiatric disorders preceding mild TBI are also associated with the presence and severity of prolonged concussive symptoms. It is unclear whether this association represents a predisposition to psychiatric disorders more generally or premorbid disruptions to neurocircuitry that lower reserve in response to the injury.

Persistent symptoms following mild TBI, or postconcussive symptoms, may be categorized according to cognitive (e.g., slowed processing speed, attentional difficulties, executive dysfunction, and memory problems), affective (e.g., depression, anxiety, irritability), somatic (e.g., headache, photo- and phonophobia, dizziness, imbalance), or sleep–arousal (e.g., fatigue, insomnia) domains. However, these symptoms are nonspecific for brain injuries. For instance, similar types of symptoms are reported in healthy and orthopedically injured subjects. Furthermore, a patient with untreated obstructive sleep apnea may also experience daytime fatigue, insomnia, and irritability.

For patients with prolonged (i.e., lasting for at least 3 months beyond the initial injury) postconcussive symptoms, reviewing the history of the initial injury and available medical records may provide insights into the severity of TBI. Additional inquiries may include whether neurosurgical interventions were required; neuroimaging was performed; medical complications or alcohol/substance use contributed to mental status alterations; and inpatient hospitalization or rehabilitation was required. Of note, cognitive impairments post TBI begin immediately after the injury and tend to improve with time; for patients with symptoms beginning sometime after the TBI occurred, the causality of the TBI is less clear.

TREATMENT OF COGNITIVE IMPAIRMENTS

Treatment strategies for post-TBI cognitive impairments require consideration of injury severity (e.g., mild vs. severe), timing postinjury (e.g., acute hospitalization vs. subacute inpatient rehabilitation), goals of treatment (e.g., enhancing rehabilitation vs. compensating for chronic impairments), and the medical/psychiatric comorbidities of the patients. The literature guiding management decisions also varies in terms of subject characteristics, settings, time postinjury, and outcomes of interest. With these limitations in mind, the following section briefly summarizes interventions aimed at enhancing cognitive recovery and/or facilitating neural network reorganization in the acute and chronic periods following a

TBI of sufficient severity to require hospitalization. A brief overview of psychiatric management of persistent postconcussive symptoms is also provided.

Optimal treatment approaches for cognitive impairments following a mild-complicated, moderate, or severe TBI often involve combinations of environmental, cognitive training, and/or pharmacologic interventions. For instance, environmental interventions used to enhance rehabilitation during the post-traumatic confusional state may include providing consistent staffing patterns; developing predictable schedules for activities; minimizing environmental disruptions; proactively addressing internal sources of disruptions (e.g., hunger, fatigue); allowing rest periods; utilizing external reminders or cues; and educating family members.

In terms of cognitive rehabilitation, strategies with a good evidence base for efficacy in TBI include direct attention training, memory strategy training and compensatory strategies (such as notebook use), and pragmatic conversation training. Metacognitive training, including self-monitoring and regulating strategies, may remediate executive function, memory, and attention deficits. In addition, a comprehensive rehabilitation approach involving training in emotional, social, and functional skills in addition to cognitive skills may enhance outcomes.

Pharmacologic interventions for disorders of consciousness and other post-TBI cognitive disturbances typically target catecholaminergic and cholinergic neurotransmitter pathways. Methylphenidate, a dopamine and norepinephrine reuptake inhibitor, may target hypoarousal, speed of information processing, motivational deficits, and fatigue. This medication may also have beneficial effects for depressive symptoms. Of note, studies have not provided support for the commonly held concern that methylphenidate decreases the seizure threshold in patients with acquired brain injuries

Studies of amantadine, a drug with complex pharmacologic effects on dopaminergic, cholinergic, and glutamatergic NMDA receptors, have shown mixed results for enhancing attentional and other cognitive impairments in the subacute and chronic periods post TBI. In the inpatient rehabilitation setting, amantadine is commonly utilized to enhance recovery from a disorder of consciousness, based on the results of a multicenter, randomized controlled trial (RCT) of 184 patients in vegetative or minimally conscious states. Of note, patients in this study were 4 to 16 weeks postinjury, and the benefits of amantadine (200 to 400 mg/day for 4 weeks) over placebo dissipated after amantadine’s discontinuation.

Cholinesterase inhibitors (e.g., donepezil) may target motivation, arousal, and memory deficits. In terms of the latter, a recent multi-VA trial comparing rivastigmine 9.5 mg/24-hour patches to placebo for chronic memory deficits in 94 veterans (with TBIs of mixed severities) was negative. However, the inclusion of subjects with high rates of concurrent opioid and anxiolytic/hypnotic use, mixed TBI severities, and relatively intact functional and cognitive statuses producing a ceiling effect may account for these negative findings.

Persistent Postconcussive Symptoms

For persistent symptoms following a mild TBI, a general approach to psychiatric management includes providing education on the expected course of improvement post mild TBI (complete resolution of symptoms in approximately 90% of patients within 12 months of a single, uncomplicated mild TBI) and the treatment of comorbid psychiatric conditions. As the evidence supporting pharmacologic management of persistent cognitive symptoms following mild TBI is limited to small RCTs and uncontrolled studies, these symptoms should rarely be treated with cognitive enhancers alone. On the other hand, medications with anticholinergic or other cognitively harmful side effects, adverse metabolic effects, and sedating properties should be tapered off, discontinued, and/or replaced. Other causes of cognitive impairments, such as obstructive sleep apnea, vitamin deficiencies, or another neurodegenerative process, should be considered in the differential diagnosis. Finally, referrals to specialists may assist in further assessment, treatment, and/or rehabilitation of cognitive, somatic, and sleep complaints, as appropriate.

TBI AND LATER-LIFE NEURODEGENERATIVE DEMENTIA

Overall, most of the evidence—including retrospective, prospective, and meta-analytic studies—indicates that TBI increases the risk of developing neurodegenerative disorders later in life. The mechanisms by which TBI, including mild TBI, increase and/or accelerate the onset of various dementias are unclear. In predisposed individuals, the injury may accelerate or augment the aggregation of pathologic proteins. Alternatively, the injury may lower cognitive reserve. In some cases, reverse causation could also be true-for example, the TBI could reflect an already present neurodegenerative process impairing impulse control and balance. Furthermore, one large epidemiologic study found that TBI increases the risk of subsequent ischemic stroke; as cerebrovascular disease is a known risk factor for vascular neurodegenerative dementias, this could potentially represent another pathway for TBI to predispose to several types of dementia.

The preponderance of evidence gleaned from epidemiologic studies indicates that TBI confers a greater risk of developing Alzheimer disease (AD). Some of the pathology of AD (primarily amyloid deposits) may be overrepresented in the brains of individuals with TBI. However, most of these neuropathologic analyses were carried out in cases with severe TBI who died within a few days of trauma. Recently, amyloid-beta PET performed on 241 subjects enrolled in the Alzheimer’s Disease Neuroimaging Initiative showed that TBI history was associated with greater amyloid-beta deposition, more pronounced cortical thinning, and a 3- to 4-year earlier age of cognitive impairment onset.

Lifetime TBI history has also been linked to non-Alzheimer, neurodegenerative dementias. Researchers utilizing data from 609 subjects in the National Alzheimer Coordinating Center (NACC) found an age of onset accelerating effect of TBI on several different dementia diagnoses, including mild cognitive impairment, AD, Parkinson disease and related disorders, and frontotemporal dementia. Overall, subjects with self-reported TBI (including single or repetitive TBI, with or without losses of consciousness) were 3.25 years younger, on average, at their first cognitive decline assessment than subjects without TBI, independently of sex, race, attained education, APOE genotype, and clinical diagnosis. Another large, retrospective database analysis demonstrated a synergistic effect of TBI and post-traumatic stress disorder (PTSD) on Parkinson disease risk in U.S. veterans. Advances in neuroimaging, cerebrospinal fluid biomarkers, and serum biomarkers of neurodegeneration may provide new insights into the association of TBI with Alzheimer, Parkinson disease, and other neurodegenerative dementias in living subjects.

Chronic Traumatic Encephalopathy

In addition to facing higher risks of AD and other neurodegenerative disorders, individuals with repeated mild TBI may develop a unique cognitive disorder known as chronic traumatic encephalopathy (CTE), a dementia syndrome not explainable by an alternative pathophysiologic process. Media coverage of cases of CTE in professional athletes has heightened interest in the relationship between TBI and dementia. This line of research has important implications for military service members, given the higher rates of repetitive TBI in this population compared to the general public.

Modern descriptions of CTE may represent a different condition from the classically described CTE. Classic CTE, also known as dementia pugilistica and initially described in professional boxers, primarily manifests with parkinsonism and nonprogressive cognitive deficits. Modern CTE presents with prominent psychiatric symptoms, and the cognitive deficit tends to progress, though parkinsonism occurs in both conditions.

Clinical descriptions of modern CTE do not clearly distinguish it from other types of dementia. Thus far, no distinguishing pattern of cognitive deficits has been described in CTE; this differs from other dementia subtypes such as AD, where early memory involvement is prominent, or behavioral variant frontotemporal dementia, which tends to present with executive function deficits. The presence of early prominent psychiatric symptoms also does not distinguish CTE from other dementia types—psychiatric changes are common in all types of dementia and may present as an initial symptom. Even motor deficits are not unique to CTE; for instance, motor deficits occur in about one in five patients with AD, the prototypical cortical dementia. A history of TBI is also not specific to CTE, as TBI appears to increase the risk for conventional dementia

At this time, CTE is only definitively diagnosed at autopsy. The pathognomonic CTE finding is the presence of irregular, phosphorylated tau aggregates in neurons and astrocytes around small vessels at the depths of the cortical sulci. Part of the difficulty in determining whether CTE represents a clinically distinct dementia syndrome is due to the postmortem ascertainment of cases, which means that the clinical phenomenology is often determined by a retrospective interview with family members rather than with detailed examination and testing during life. Currently, prospective studies of individuals with repeated TBI, such as professional athletes, are underway, which may help clarify the phenomenology of CTE and validate clinical criteria.

NEUROPSYCHIATRIC DISTURBANCES FOLLOWING TRAUMATIC BRAIN INJURY

The following sections summarize the frequencies, presentations, and management of selected neuropsychiatric disorders post TBI. These disorders are intricately linked to functional, cognitive, and psychosocial outcomes.

The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) classifies many of the neurobehavioral consequences of TBI as disorders due to another medical condition, in this case, TBI. Contemporary research supports the notion that the relationship between TBI and neuropsychiatric disorders is indeed bidirectional: not only can TBI directly cause these disorders, but they appear to increase the risk of TBI as well.

Individuals with premorbid psychiatric disorders have a 70% higher risk of experiencing TBI, possibly because conditions such as substance use disorders, attention-deficit/hyperactivity disorder (ADHD), mania, and psychosis may place them in situations in which TBIs are more likely to occur. Thus, the development of cognitive and psychiatric disorders post TBI must be considered in the context of the individual’s pre-existing psychiatric/medical history, cognitive status, genetic/epigenetic predispositions, and social functioning.

GENERAL MANAGEMENT PRINCIPLES

Before proceeding with summaries of individual disorders, several management principles applicable to all post-TBI neuropsychiatric disorders are worthy of mention. Akin to the treatment of cognitive impairments, treatment strategies for post-TBI neuropsychiatric disorders are best approached from a multidisciplinary perspective. Pharmacologic interventions are ideally employed as adjunctive to nonpharmacologic ones. There is also a paucity of high-quality RCTs providing a solid scientific basis for psychiatric treatments post TBI. Furthermore, studies in this population vary in terms of injury severities, times post injury, and other sample characteristics. Currently, anecdotal cases, small uncontrolled trials, and clinical experience comprise the only evidence bases supporting daily treatment decisions.

Nonpharmacologic therapies—that is, psychotherapy, environmental and behavioral interventions—should be prioritized wherever possible. Given the limited evidence base for many interventions, clinicians must make sure to avoid iatrogenic harms related to medications.

Finally, care must be taken when increasing medications—“start low and go slow.” However, adequate dosage and duration should be trialed prior to tapering off or adding adjunctive medications. A partial response to a specific agent may require augmentation therapy, ideally with a drug whose mechanism of action and side-effect profile does not mimic the former’s. Finally, clinicians should frequently reassess brain-injured patients for side effects and interactions with other medications.

Agitation and Aggression

Agitation is defined as excessive, usually nonproductive, motor activity associated with inner tension. Agitation is frequently used to describe various behaviors without reference to the patient’s subjective experience or intention, such as yelling or pulling out lines. On the other hand, aggression refers to overtly threatening or injurious behaviors directed toward another person or object. Aggressive behaviors tend to be reactive and impulsive, as opposed to proactive, when occurring in the context of TBI-related cognitive impairments.

Agitation and aggression pose frequent and significant problems in acute hospital and rehabilitation settings. In the acute phase of recovery, agitation, and aggression may occur during the post-traumatic confusional state. In the chronic setting, aggression and agitation may impede care and reflect protracted impairments in impulse control or affective lability. The identification of iatrogenic or other medical complications that may contribute to aggression, agitation, or delirium should precede treatment initiation.

Typical environmental strategies for agitation and aggression may include frequent reorientation, placing the patient in a private room, and reducing ambient noise. Applied behavioral analysis and management techniques may identify and intervene upon triggers and possible reinforcements. Other nonpharmacologic interventions include caregiver support and education, supportive psychotherapy, and cognitive behavioral therapy (CBT) tailored to the patient’s cognitive status.

Standardized measures of agitation and aggression (i.e., the Overt Aggression Scale [OAS]) may track response or nonresponse to possible interventions.

Pharmacologic therapy should be considered adjunctive to nonpharmacologic strategies. Evidence from animal studies suggests that both typical (e.g., haloperidol) and atypical neuroleptics (e.g., risperidone) might have a negative impact on cognitive recovery. One animal study demonstrated that haloperidol, but not olanzapine, impaired cognitive performance in rats post TBI. In general, a prudent strategy for treating acute agitation in the context of TBI is to avoid anything that could worsen the patient’s cognition—and thus ability to control their behavior—such as anticholinergic drugs, other deliriogenic medications like benzodiazepines, and, possibly, typical antipsychotics. Finally, care must be taken to ensure that aggression and agitation are not induced by iatrogenic side effects (e.g., akathisia from antipsychotics).

Pharmacologic treatments for chronic agitation or aggression following mild-complicated, moderate and severe TBI include amantadine, beta-blockers (e.g., propranolol and pindolol), and selective serotonin reuptake inhibitors (SSRIs) (e.g., sertraline). Anticonvulsants may be considered despite limited data supporting their efficacy and safety in this population. Recently, RCTs of amantadine (200 mg per day) and carbamazepine (up to 400 mg per day) for post-TBI neuropsychiatric symptoms failed to find significant differences between active and placebo groups on most measures of irritability, aggression, and anger. On the other hand, the placebo response rates observed in these studies highlight the benefits of nonpharmacologic aspects of treatment on TBI-associated agitation and aggression, such as developing therapeutic alliances and tracking symptoms.

Of note, while prophylactic administration of antiseizure medications (ASMs) for the first 7 days following severe TBI effectively reduces the incidence of seizures during this timeframe, longer administration of ASMs or administration in mild or moderate TBI has no effect on the subsequent development of post-traumatic epilepsy. As ASMs can have significant side effects and may potentially inhibit the neuroplasticity needed for optimal recovery, clinicians should avoid prophylactic use of ASMs beyond 7 days postinjury in patients with severe TBI and avoid prophylactic use completely in patients with mild or moderate TBI.

Comportment Changes

Patients with TBI may experience deficits in comportment, defined as the set of mental processes facilitating contextually appropriate social adaptation and judgment. Clinically, caregivers may report the patient's "personality" has changed. For example: "Before the injury, he/she was polite and introverted. Now he/she curses, blurts out hurtful comments, and hits people when he/she is told no in public."

The famous history of Phineas Gage exemplifies the potentially devastating impact of TBI on comportment and social cognition. Gage was a 25-year-old construction foreman for a railroad company when he suffered a severe, penetrating TBI from a blast at a railroad construction site. A 43-inch long, 13.5-pound iron rod passed through the left frontal lobe before exiting the top of his skull. Gage was able to talk and ambulate away from the accident with assistance from coworkers. However, his physician described Gage as inconsiderate, capricious, obstinate, and profane after the accident. Despite having a relatively intact intellect, "He was no longer Gage." Ultimately, Gage was dismissed from this job, but he was able to work in various others before his death from status epilepticus.

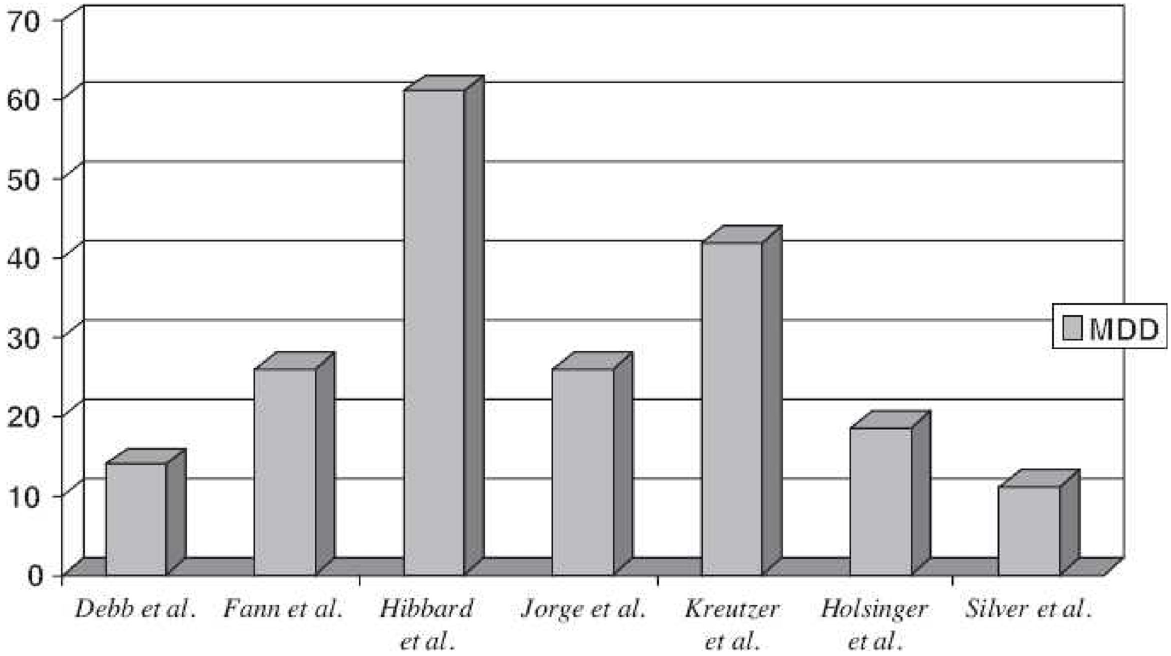

FIGURE 2.5–4. Frequency of mood disorder due to traumatic brain injury with major depressive features.

Broadly speaking, TBI-induced changes in comportment reflect injury to the prefrontal neural circuitry required for modulating reward/punishment-based decision making, emotional expression and generation, and insight. In the case of Phineas Gage, modern imaging techniques have demonstrated rod-inflicted damage to the orbitofrontal and medial prefrontal cortical regions that contribute to these functions.

Treatment strategies for impaired social cognition prioritize behavioral interventions. Providing psychoeducation to the patient and family is an integral component of management. Therapeutic interventions may include recalling premorbid emotional events and using stimuli that trigger strong emotions to assist with reconnecting to these emotions. Role-playing and image interpretation can assist with social skills training. In addition, partner training has been shown to improve communication appropriateness and effectiveness. Finally, behavioral analysis may quantify the frequency of unwanted behaviors and identify modifiable triggers.

Pharmacologic interventions may target disinhibited, agitated, or aggressive behaviors. SSRIs may reduce sexual drive and are utilized for disinhibited behaviors in other neuropsychiatric conditions. Anticonvulsants, like those used for mania, are other options for managing disinhibition and aggression. Furthermore, addressing cognitive impairments may concomitantly improve behavioral control. More research is needed to guide pharmacologic strategies targeting comportment changes following TBI.

Clinical Vignette: Comportment Changes due to TBI

A 42-year-old construction worker fell from the second floor of a new building. He spent 3 days in a coma and remained amnestic and disoriented for approximately 3 weeks. A CT scan showed bilateral orbitofrontal and anterior temporal hemorrhagic contusions. Six months after the injury, he exhibited a significant personality change. His wife complained of his frequent and often inappropriate sexual demands and stressed his lack of intimacy. He also became easily upset, shouting and making threats when he felt provoked. He was less sensitive to other people’s feelings. A trial of sertraline and the maintenance of a numerical record of outbursts resulted in significantly reduced irritability, outbursts, and hypersexuality.

Depressive Disorders

Depressive disorders are common neuropsychiatric complications of TBI. The frequency of post-TBI depression ranges from 6% to 77% (Fig. 2.5–4). This variability in the reported frequency of depressive disorders, particularly major depression, may reflect the lack of uniformity in the psychiatric diagnosis, sampling times postinjury, and TBI severities.

The risk of developing a depressive disorder appears to be highest in the first year. Approximately 29.4%, 19.6%, and 1% of patients are estimated to meet DSM-IV criteria for major depressive disorder (MDD), depressive disorder not otherwise specified, and dysthymic disorder, respectively, at 12 months postinjury. In comparison, the annual rate of MDD in the general population has been quoted at 6.7%.

Approximately 25% of those without depression in the first year will go on to develop it in the second year, and the risk of depression remains elevated for decades. Furthermore, post-TBI depression tends to be persistent, with one study finding that two-thirds of patients experience symptoms for 1 year or longer. Overall, patients with TBI are at increased risk of developing recurrent depressive disorders throughout their lifetimes compared to patients without TBI.

The DSM-5 classifies depressive disorders associated with TBI as "Depressive Disorder due to Traumatic Brain Injury" with the following subtypes: (1) "with depressive features" (prominent depressed mood but full criteria for a major depressive episode are not met), (2) "with major depressive-like episode" (if the full criteria for a major depressive episode are met), or (3) "with mixed features" (manic or hypomanic symptoms present but not predominant). The standard DSM-5 criteria have a high sensitivity and specificity for identifying depressed patients when compared with alternative diagnostic criteria. One study found that the three symptoms of hopelessness, feeling worthless, and anhedonia best differentiated depressed from nondepressed patients.

Other conditions requiring consideration in the differential diagnosis of post-TBI depression include adjustment disorder with depressed mood; apathetic syndromes; emotional lability or pathologic laughing and crying (PLC); delirium occurring in the context of a post-traumatic confusional state or another cause; and expressive dysprosodia (i.e., impaired ability to utilize intonation and inflection to convey meaning in language associated with right hemispheric injury). Patients with adjustment disorders develop short-lived and relatively mild emotional disturbances within 3 months of a stressful life event, and symptoms do not persist longer than 6 months after the stressful event or its consequences cease.

It is well known that age and severity of brain injury are major determinants of recovery from TBI. However, mood disorders have consistently shown independent deleterious effects on recovery of patients with TBI. Patients with post-TBI depression are more likely to have poorer executive function, social integration, and vocational outcomes. Importantly, post-TBI depression appears to increase the risk of subsequent suicide. Thus, the identification, treatment, and/or prevention of depression represent essential elements of post-TBI care.

Clinical Vignette: Post-TBI Depression

A 42-year-old engineer had a motor vehicle accident when returning from a convention. He had multiple injuries, including a diaphragmatic rupture and a left frontotemporoparietal subdural hematoma. When admitted to the hospital, the patient was hypotensive and hypoxic. With the urgent intervention of two surgical teams, he underwent repair of his diaphragm and evacuation of the subdural hematoma. The patient remained in a coma during the following 72 hours. PTA lasted for almost 3 weeks. At this point, the neurologic examination disclosed a right hemiparesis, a left lateral rectus palsy, mild hypophonia, and dysarthria. Forty days after the accident, he transferred to a specialized rehabilitation unit. Once his PTA had cleared, neuropsychological tests neared normal limits. The patient conveyed a profoundly depressed mood and feelings of hopelessness. He stated that he would never be able to recover, that his career was ruined, and that it would have been better if he had died in the accident. He had no appetite and refused to participate in physical rehabilitation. He also had significant sleep problems. Treatment of depression was initiated with sertraline at a dosage of 50 mg per day. After 3 weeks, the patient’s mood significantly improved, and he became involved in the rehabilitation program. At 6-month follow-up, he was no longer depressed and had returned to work.

The relationships between depression and psychosocial stressors are often bidirectional. Data from a longitudinal cohort study suggests that pathological processes associated with the brain injury contribute, at least in part, to the development of mood disorders. In this study, patients with TBIs of mixed severities were recruited from three academic centers (n = 158) and compared to an orthopedic injury control group (n = 27) with similar demographic characteristics, psychiatric histories, and functional impairments. At 3, 6, and 12 months of follow-up, patients with TBI exhibited a significantly greater frequency of mood disorders than the comparison group of patients with orthopedic trauma (Fig. 2.5–5). In the TBI group, 54% of patients developed a mood disorder at some time during the first year after injury, compared to 22% of patients with multiple traumatic injuries but without central nervous system involvement. In addition, the frequency of MDD was significantly greater in the TBI versus the control group. Thus, mood disorders were significantly more frequent in patients who suffered TBIs than in patients with similar background characteristics who underwent similar levels of stress (e.g., motor vehicle accidents) but who did not sustain a brain injury.

The pathologic mechanisms initiated by TBI and contributing to depression risk remain unclear. In one series, post-TBI major depression was associated with reduced gray matter volume in the lateral aspects of the left prefrontal cortex (Fig. 2.5–6). Major depression could result from the deactivation of more lateral and dorsal frontal cortex and increased activation in ventral limbic and paralimbic structures.

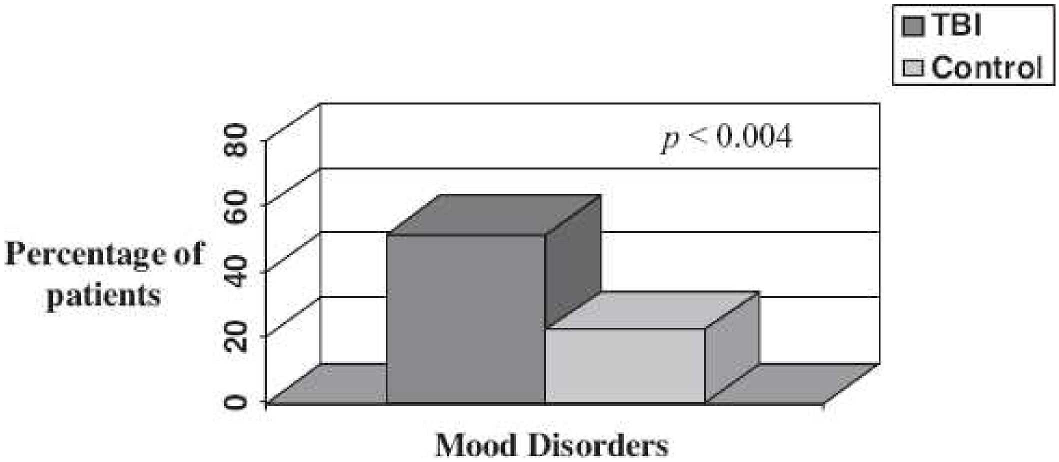

FIGURE 2.5–5. Frequency of mood disorders among patients with traumatic brain injury (TBI) compared to controls with orthopedic injuries. The frequency of mood disorders was significantly greater in TBI patients than in a comparison group of patients with orthopedic trauma.

Faulty prefrontal modulation of medial limbic structures could explain the co-occurrence of anxiety and aggression in this cohort. In a subgroup analysis, MDD was significantly associated with the presence of anxiety disorders (Fig. 2.5–7). Furthermore, major depression was also associated with the occurrence of aggressive behavior, as categorized by the OAS. Of the 30 patients with major depression, 17 (57%) demonstrated significant aggressive behavior compared with 10 of 44 patients who showed the same level of aggression without mood disorder during the first year after TBI (*p* = .004) (Fig. 2.5–7).

In addition to the injury itself, several non-injury factors do appear to influence post TBI depression risk. These factors include personal histories of mood and anxiety disorders, impaired social functioning, current or past alcohol use, unemployment, and poverty. In this same longitudinal cohort, patients with post BI major depression we more likely to have personal histories of mood disorders than non-depressed patients. Groups did not differ significantly in respect to types or severities of brain injuries, family histories of psychiatric disorders, or degrees of physical impairments. Patients with major depression had significantly poorer premorbid social functioning than the nondepressed group, and poor social functioning was the strongest and most consistent clinical correlate of major depression during follow-up (Fig. 2.5–8).

As limited data regarding the treatment of depression in the TBI population exist, clinical care generally follows principles derived from the treatment of idiopathic MDD. Psychotherapeutic interventions with some evidence for TBI-related depression include CBT, mindfulness-based CBT, behavioral activation, problem-solving therapy, acceptance and commitment therapy, and supportive psychotherapy. However, most evidence is derived from small and uncontrolled studies. There is not enough data supporting the efficacy of one psychotherapy over another. One RCT found that both CBT and supportive psychotherapy equally improved depression following TBI. Of note, psychotherapies may require tailoring to suit the cognitive and physical needs of the patient. For example, patients with post-TBI cognitive impairments may benefit from receiving written summaries of session contents, handouts or visual aids, simplified instructions, and caregiver support.

Selection among competing antidepressants is usually guided by their side-effect profiles. In general, it is best to choose drugs with minimal anticholinergic properties, sedative effects, and lowering of the seizure threshold. Currently, SSRIs are considered first-line therapies. The most studied SSRI in the TBI population is sertraline. However, the two largest

RCTs of sertraline to date have not supported its efficacy for the treatment of post-TBI depression. In the most recent of these two RCTs, sertraline was associated with more side effects and dropouts than placebo. Despite being the "largest" trials for post-TBI depression, the number of subjects in both studies (n = 52 and n = 62) were still relatively small.

Additional evidence supporting the use of sertraline and other SSRIs (e.g., escitalopram, citalopram, and fluoxetine) for post-TBI depression are derived from small RCTs or uncontrolled studies. Methylphenidate may benefit depression, cognition, and fatigue, according to small RCTs. Tricyclic antidepressants (TCAs) are not preferred due to their side-effect profiles, including lowering of the seizure threshold and anticholinergic activities. Other antidepressants used in the treatment of idiopathic MDD (e.g., selective serotonin and norepinephrine inhibitors) may be reasonable alternatives despite lacking systematic investigations in the TBI population. Bupropion, while not absolutely contraindicated in TBI, can lower the seizure threshold; using the lowest effective dose, the extended- or sustained-release formulation, and careful monitoring may help reduce the risk.

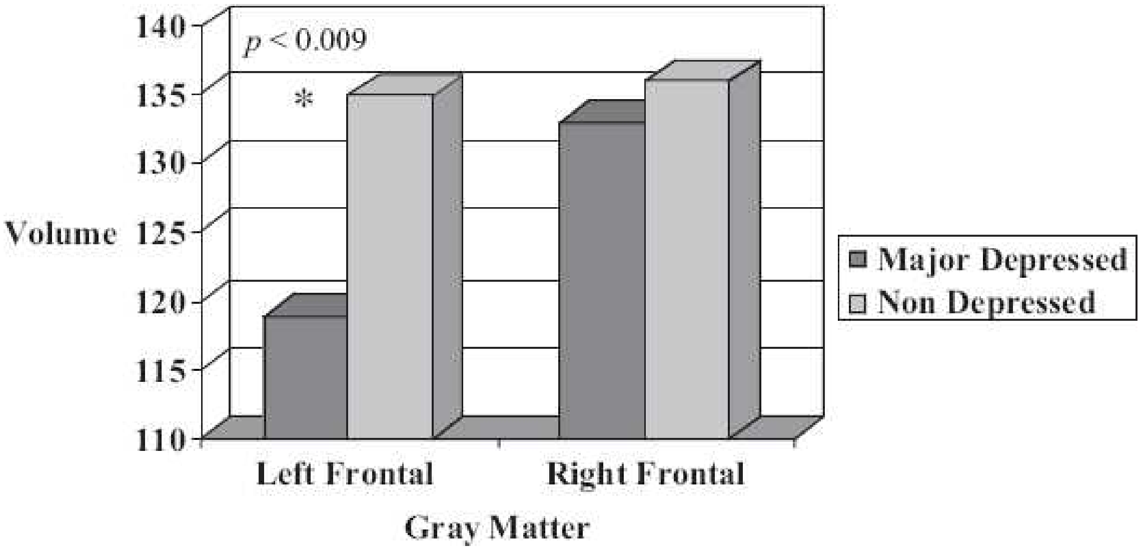

FIGURE 2.5–6. Frontal gray matter volumes in traumatic brain injury (TBI) patients with major depression. Magnetic resonance imaging volumetric studies showing that, after controlling for age and severity of trauma, left frontal gray matter volumes are significantly decreased in the group of depressed TBI patients compared with the nondepressed group.

ECT is not contraindicated in patients with TBI and may be considered if other methods of treatment prove to be unsuccessful, though there are no systematic studies of its effects. ECT should be administered with the lowest possible effective energy, using brief-pulse, nondominant unilateral currents, and with an interval of 2 to 5 days between treatments. Neurostimulation therapies such as repetitive transcranial magnetic stimulation (rTMS) and transcranial direct current stimulation (tDCS) are also options, although evidence for the latter is inconsistent. A recent meta-analysis of three RCTs and two quasi-experimental pre- and postintervention studies found a positive pooled effect of rTMS on post-TBI depression severity. However, studies to date have utilized heterogeneous stimulation parameters and research methodologies. Furthermore, rTMS should be used with caution, as it may increase the risk of seizure.

Prevention is regarded as more effective than treatment. The prevention of post-TBI depression has received little attention compared to other medical conditions. One RCT compared sertraline 100 mg/day versus placebo for the prevention of depression in 94 patients in the subacute recovery period of TBI. In this study, the hazard of developing depression during 24 weeks of treatment was four times greater in the placebo than sertraline group, and the number needed to treat with sertraline to prevent depression at 24 weeks was 5.9, 95% confidence interval (CI) [3.1, 7.1]. However, another RCT comparing sertraline 50 mg/day versus placebo for 3 months in the subacute recovery period did not find a significant difference in the number of patients who developed depressive episodes between groups. Additional studies of preventive interventions are needed.

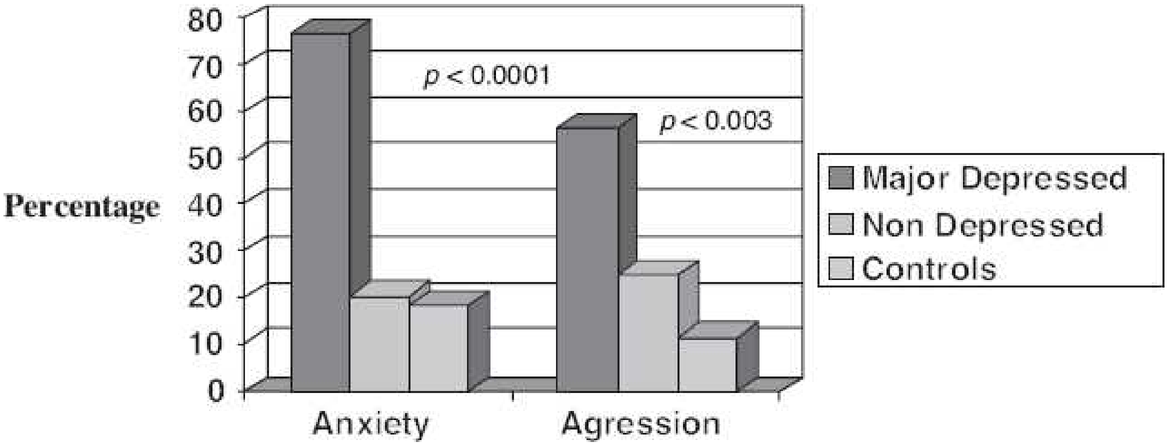

FIGURE 2.5–7. Coexistent anxiety and aggression among depressed traumatic brain injury (TBI) patients. Frequency of anxiety and aggressive disorders among depressed TBI patients, nondepressed TBI patients, and patients with orthopedic injuries. Major depressive disorder following TBI was significantly associated with the presence of anxiety disorders and aggressive behavior.

Pathologic Laughing and Crying

PLC manifests as sudden and uncontrollable affective outbursts which may be congruent or incongruent with the patient's mood and can occur spontaneously or may be triggered by minor stimuli. Patients recognize the emotional displays as unrelated or excessive to the underlying mood. Affective displays in PLC are not necessarily reflective of a pervasive alteration of mood or specific vegetative symptoms associated with a major depressive episode.

One study examined the prevalence and clinical correlates of this disorder, assessed using the Pathological Laughter and Crying Scale, in a group of 92 consecutive patients with TBI. Eleven percent of the patients developed PLC during the first year after TBI. PLC was associated with the presence of anxiety disorders and frontal lobe lesions involving the lateral and ventral aspects of the prefrontal region (Fig. 2.5–9). In addition, patients with PLC had significantly more frequent aggressive outbursts and poorer social functioning.

Treatment options for PLC in TBI remain understudied; however, based on their efficacy for this condition in other neurologic disorders and their generally benign side-effect profile, SSRIs represent a reasonable first-line strategy. However, treatment response varies greatly among brain-injured patients, with some showing a rapid response at relatively low doses while others require more time and higher doses. Combination therapy with dextromethorphan and quinidine is another option, but quinidine may inhibit the metabolism of other medications. In addition, the proprietary combined pill is quite expensive and may impose a financial burden on some patients.

Apathy

Apathy consists of impairment of motivation, initiation, interest, activity, emotional responsiveness, and/or spontaneity. Apathy, regardless of the underlying etiology, frequently occurs comorbidly with depression, but the two conditions constitute distinct entities with different implications for prognosis and treatment. While patients with apathy often do not appear distressed by their symptoms, apathy can markedly impair patients' ability to participate in rehabilitation therapies, function in the community, and maintain interpersonal relationships.

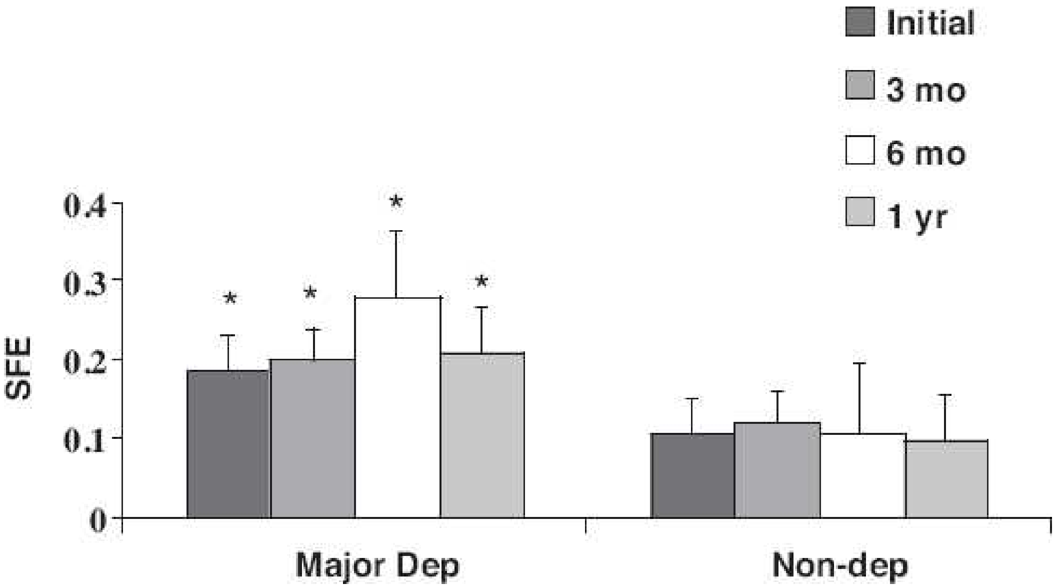

FIGURE 2.5–8. Comparison of Social Functioning Examination (SFE) scores of major depressed and non-depressed patients during the first year following traumatic brain injury. Higher scores are associated with greater impairment. Poor social functioning was the strongest and most consistent clinical correlate of major depression during follow-up.

Apathy is a frequent complication of TBI: studies have quoted prevalence ranges nearing 75%. Studies investigating potential associations between severity of TBI and apathy have produced conflicting results some studies have found no relationship, but others have found that more severely injured patients are more likely to show apathy. One imaging study in patients with TBI found that patients with apathy showed damage to the left hemisphere, specifically in the frontal lobe, anterior cingulate, insula, corpus callosum, and corona radiata, consistent with these regions' involvement in emotion and goal-directed behavior.

Currently, treatments for apathy after TBI are lacking an evidence base. Providing psychoeducation on the nature of apathy and why it does not represent laziness or withdrawal of love may improve caregivers' understanding of the condition and expected treatment outcomes. As patients with apathy may participate in and enjoy activities initiated by others, even if they do not initiate them on their own behalf, family members and friends can provide a structured environment and utilize verbal or visual reminders to encourage the person with apathy to engage in activities. Based on case reports and case series, pharmacologic agents with prodopaminergic activity (e.g., methylphenidate, amantadine, bromocriptine, bupropion) may be trialed as well. Some studies in the general population suggest that SSRIs can worsen apathy, although others have found the opposite. In general, rigorous studies examining treatments for apathy post TBI are greatly needed.

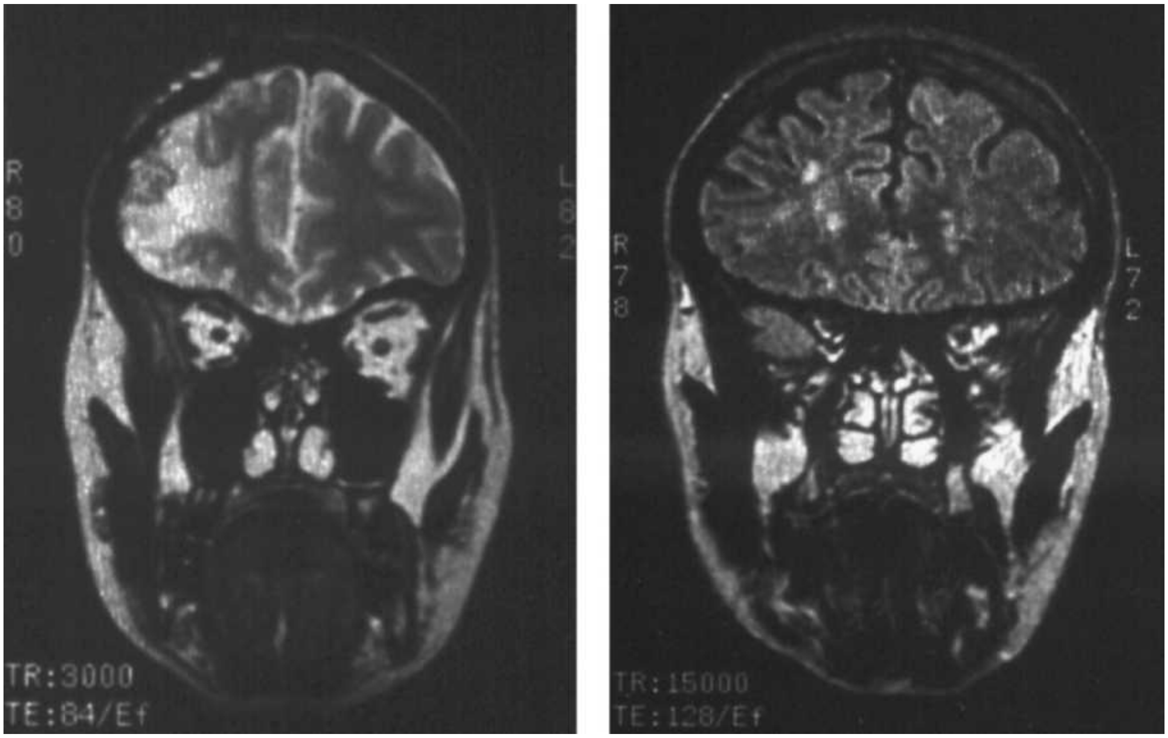

FIGURE 2.5–9. Emotional deregulation following traumatic brain injury. Frontal lobe lesions involving the lateral and ventral aspects of the prefrontal region in a patient with recurrent episodes of pathologic crying consistent with a diagnosis of pathologic laughing and crying.

Bipolar and Related Disorders

Bipolar mood disorders are relatively less frequent compared to depressive disorders post TBI. Estimated rates of bipolar-type disorders in the first-year postinjury range from 2% to 9%. The diagnosis of mania following TBI is based on the DSM-5 criteria for bipolar and related disorders due to TBI, with three subtypes: (1) “with manic features” (full criteria not met for a manic or hypomanic episode), (2) “with manic- or hypomanic-like episode” (full criteria met), and (3) “with mixed features” (depressive symptoms present but not predominant). As in the case of depressive disorders, the mania must appear to stem directly from the physiologic consequences of the brain injury. The diagnosis only requires that the predominant mood be elevated, euphoric, or irritable and accompanied by abnormally elevated energy and/or activity. This does not require the same specificity of the syndrome as a manic episode or a major depressive episode.

A 1987 study reported on 20 patients who developed manic syndromes after closed head trauma. There was a significant association between mania and the presence of post-traumatic seizures, predominantly focal onset with impaired awareness (temporal lobe epilepsy). There was no association, however, with a family history of bipolar disorder among first-degree relatives.

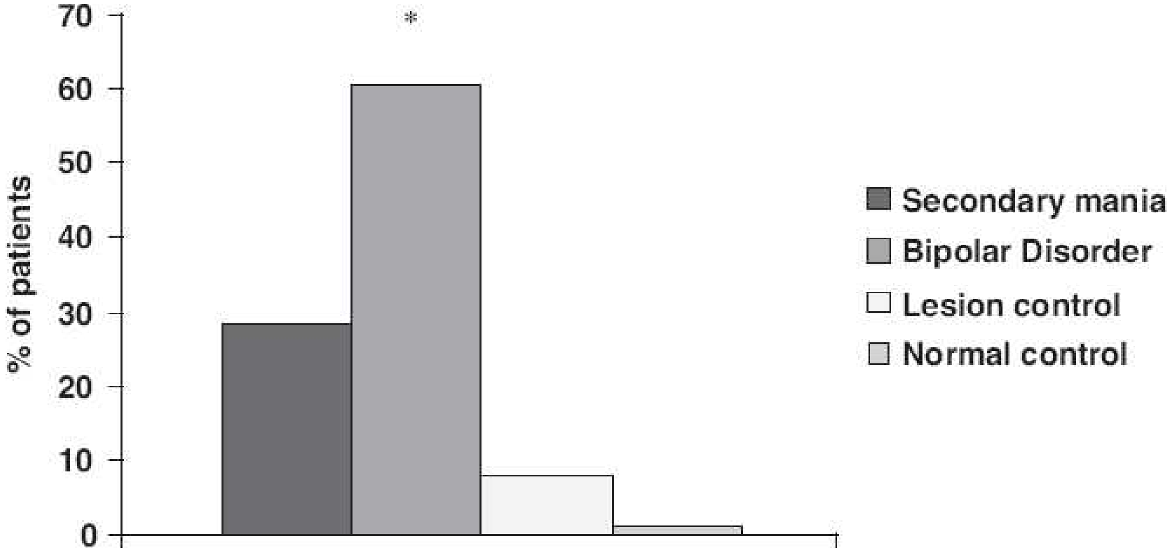

FIGURE 2.5–10. Family history of mood disorder among patients with manic syndromes due to traumatic brain injury (TBI). The frequency of family history of mood disorders is significantly higher among patients with bipolar disorder compared to patients with secondary mania, TBI patients without mood disorders, and healthy controls. However, family history of mood disorders was greater among patients with secondary mania compared to TBI patients without mood disorders and healthy controls.

One study found that patients with TBI showed a higher rate of secondary mania during the year after traumatic injury than patients with orthopedic injuries. Although manic episodes lasted only approximately 2 months, the presence of an elevated, expansive, or irritable mood had a mean duration of 5.7 months. Manic syndromes were significantly associated with hyperactivity, disinhibition, aggressive behavior, and hypersexuality. Secondary mania was not related to the type or severity of brain injury, degree of physical or intellectual impairment, level of social functioning, or the presence of family or personal history of psychiatric disorder (Fig. 2.5–10). The development of abnormal electrical activation patterns in limbic networks, functional changes in aminergic inhibitory systems, and the presence of aberrant regeneration patterns may all play a role in the genesis of post-TBI mania.

The differential diagnosis of mania following TBI should include substance/medication-induced bipolar and related disorder, delirium, and comportment change due to TBI. Impairments in social cognition, conversational pragmatics, and executive dysfunction may appear “manic like.” Affective and behavioral manifestations may also feature prominently in delirium (including during the acute post-traumatic confusional state). Substance-induced mood disorder occurs following intoxication or withdrawal from drugs and is another important differential diagnosis among patients with TBI, who frequently have a prior history of substance abuse.

There have been no systematic studies for the treatment of secondary mania. Thus, it is reasonable to treat along the lines of idiopathic bipolar disorder. Some researchers have theorized that lithium and valproate have neuroprotective effects, which would certainly constitute an important therapeutic advantage among brain-injured populations; however, there is currently no evidence to support if this actually occurs, as data from a controlled trial of valproate in TBI failed to identify a beneficial effect on cognitive and functional outcomes. Patients with persistent cognitive impairments following TBI may also be more susceptible to neurologic side effects of lithium and anticonvulsants. The role of the anticonvulsant lamotrigine for post-TBI bipolar disorder has not been studied. Although data is limited to case reports, atypical antipsychotics (e.g., quetiapine, olanzapine) may also be considered.

Clinical Vignette: Bipolar Disorder Due to TBI

A 31-year-old married man with two children was struck in the head when a welding tank exploded, sending large and small pieces of metal in all directions. He experienced intracerebral bleeding in the right basal ganglia and white matter surrounding the anterior horn of the right lateral ventricle. Months after the injury, he became manic, staying up all night, selling things he owned for ridiculously low prices, and buying a new car he could not afford. He believed that he would become a famous songwriter and exhibited hypersexual behavior. He was admitted to a hospital, where he was treated successfully with low-dose lithium and olanzapine. He remained euthymic for approximately 6 months, and he discontinued his medications. Although he continued to experience significant behavioral changes, manic symptoms did not recur during a 2-year follow-up period.

Trauma- and Stressor-Related Disorders

PTSD, the prototypical trauma-related disorder, is characterized by recurrent intrusive recollections, distressing dreams, and flashbacks of the traumatic event. Patients with PTSD show features of increased arousal such as difficulty falling or staying asleep, exaggerated startle response, and irritability. Patients lose interest in social activities and avoid thoughts, feelings, or circumstances associated with the event.

Clinical Vignette: PTSD Following TBI

A 33-year-old nurse was assaulted while walking to her job in a small hospital. She received several blows to her head and was admitted unconscious to the ER. A CT scan showed a depressed skull fracture and right temporal and parietal contusions. She regained consciousness during the first 12 hours of admission and was confused and disoriented for the following 3 days. After this, she had an uneventful course and a good recovery. One month after trauma, she received a complete neuropsychiatric evaluation. She complained of intrusive recollections of the moments preceding the assault, which were associated with great psychological distress, sweating, and palpitations. She was also disturbed by nightmares in which she suddenly found herself in an intensive care unit connected to a ventilator, incapable of talking or asking for help. She avoided the hospital's surroundings and was unable to walk alone. She also complained of decreased concentration and memory impairments. The results of her neuropsychological tests, however, fell within the normal range. A course of modified prolonged exposure (PE) therapy greatly improved the distress associated with intrusive memories and the frequency of nightmares.

Individuals who sustain a TBI are more likely to develop PTSD than those who suffer other types of traumatic injuries. Veterans who experienced a TBI show a threefold greater rate of PTSD than those without a TBI; civilians with TBI have almost twice the likelihood of developing PTSD as those with other types of traumatic injuries. Estimates of how many individuals with TBI have comorbid PTSD vary widely, from 12% to 30% in civilians and from 12% to 89% in veterans. In addition to variance introduced by different methods of assessment, part of the uncertainty lies in the overlap between PTSD and TBI symptoms, such as impaired concentration/attention and executive function and poor sleep.

Some researchers have speculated that TBI is incompatible with the development of PTSD, because amnesia for the traumatic event would be protective. More recent studies have not born this out. It is possible that even patients with moderate or severe TBI can encode trauma memories at an implicit level, and these memories can influence ongoing emotions and behaviors. Lack of declarative memory of the traumatic event could also interfere with attempts to place the trauma in the context of an overall life story and mobilize coping resources to address it.

TBI may predispose to PTSD by causing damage to neural circuits responsible for modulating fear responses. Cognitive deficits from PTSD could impair the ability to effectively utilize coping skills to deal with PTSD. Conversely, ongoing psychological stressors from PTSD could impair cognitive recovery from the TBI. Alternately, the types of incidents that result in TBI could produce more feelings of fear and helplessness than those responsible for other types of injuries.

The combination of TBI and PTSD appears to worsen outcomes compared to either one alone. Patients with PTSD at 6 months following severe TBI had significantly more impaired community reintegration than TBI patients without PTSD. Individuals with TBI complicated by PTSD self-assess their cognitive and health status as worse than those with TBI alone, and some data suggests that the combination of PTSD and TBI impairs objective cognitive performance more than TBI alone. Patients with both disorders report more symptoms of PTSD than those with PTSD alone and experience greater PTSD symptom severity.

Currently, guidelines for the treatment of PTSD in the general population favor psychotherapeutic interventions, such as PE or cognitive processing therapies (CPTs), over pharmacologic ones. Similarly, CBT for insomnia (CBT-I) is favored over pharmacologic interventions for PTSD-related insomnia. While more research is needed to assess the efficacy of these interventions for patients with TBI, particularly moderate and severe TBI, preliminary studies suggest that evidence-based psychotherapies for PTSD may still be useful in this population. However, therapies may require modifications for patients with persistent cognitive and physical impairments. For instance, exposure therapies may require reconstructing memories to offset amnesia, simplification of homework, and more frequent repetition.

Given their safety and tolerability in patients with TBI, SSRIs represent a reasonable first-line treatment strategy in PTSD associated with TBI. An open-label study of prazosin—effective for PTSD-related nightmares in the general population—in patients with TBI found improvements in cognitive performance, sleep, and headaches; however, the largest RCT to date of prazosin versus placebo for PTSD-related nightmares was negative. Given the possibility of adverse cognitive side effects, anticholinergic medications and benzodiazepines should be avoided. Finally, consideration of other conditions contributing to poor sleep quality and PTSD symptoms, such as obstructive sleep apnea, should be incorporated into psychiatric management.

Anxiety Disorders

The available data suggests that post-TBI anxiety disorders are common; may take the form of any anxiety-spectrum disorder; and tend to co-occur with depression. Of the individual anxiety disorders, generalized anxiety disorder may be the most common following TBI, but data has been inconsistent. A meta-analysis of 22 studies, comprised of 1,145 subjects with TBIs of mixed severities and postinjury durations, estimated the prevalence of generalized anxiety disorder to range from 2% to 28%, with a pooled estimate of 11%.

Rates of panic disorder following TBI range from 6% to 14%. One study found that TBI conferred double the risk of developing panic disorder compared to other traumatic injuries. TBI also doubled the risk of developing agoraphobia, with 7% of patients with TBI experiencing this condition.

The estimated prevalence of specific phobias post TBI, while based on limited data, appears to be similar to that of the general population (1–10% versus 8–12%, respectively). One study found that among patients who experienced TBI due to a motor vehicle accident, the specific phobias that subsequently developed involved transportation. Another group found that a substantial proportion of individuals with TBI experience a phobic-like fear of falling, which at times may interfere with activities. One study found that 5% of patients with TBI developed social anxiety disorder, twice the rate of those who experienced other types of injuries. Although obsessive-compulsive disorder is no longer classified with other anxiety disorders by the DSM-5, its estimated prevalence (1% to 5%) does not appear significantly elevated above that of the general population.

Previous experimental and clinical studies on the neurobiology of anxiety disorders have emphasized the etiologic role of disrupted prefrontal modulation of limbic and paralimbic circuits involving medial and orbital prefrontal cortices, the amygdala, and the hippocampus. A recent meta-analysis of structural and functional neuroimaging correlates of syndromal post-TBI anxiety found a broad range of regional and circuit alterations reported among studies. Although the most frequently implicated regional alterations involved components of the default mode network, the methodologic heterogeneity and a limited number of included studies prevent definitive conclusions. Additional investigations on the neurobiological correlates of anxiety disorders post TBI are needed.

The evidence base for the treatment of anxiety disorders in TBI remains preliminary. As with trauma-related disorders in TBI, SSRIs are likely the best first-line pharmacotherapy, and medications that could worsen cognition in this vulnerable population, such as benzodiazepines and drugs with anticholinergic effects, should be avoided. Few studies specifically examining psychotherapy (e.g., CBT) for TBI-related anxiety suggest this may be effective as well.

Substance-Related and Addictive Disorders

Substance-related and addictive disorders are major risk factors for the occurrence of TBI and major determinants of the clinical and psychosocial outcomes post TBI. A prospective study of 197 patients with TBI showed that 42% had blood alcohol levels consistent with legal intoxication at the time of admission to the emergency department and that blood alcohol levels suggested a history of problem drinking among these patients. The authors also observed that the frequency of alcohol abuse and associated disorders diminished immediately after TBI but increased again at 1-year follow-up, although not to preinjury levels. Although alcohol use significantly decreased during the first year after injury, about 25% of the patients reported alcohol-related problems. As expected, preinjury problematic alcohol use predicted alcohol-related problems after TBI.

In one study, 60% of patients with alcohol use disorders prior to TBI had resumed problematic alcohol use within 1 year of their injury. Patients who relapsed had significantly less education, showed significantly more severe injury as demonstrated by lower GCS scores and a greater frequency of moderate and severe injuries, had significantly higher frequencies of focal lesions on their initial CT or MRI scans and showed impaired cognitive performance on memory and executive tasks.

Conceivably, the combined effects of traumatic injury and the toxicity of chronic substance abuse (particularly alcohol) might produce greater disruption of the neural circuits involved in mood regulation, motivation, and reward processing. In turn, these changes could result in more severe psychopathology, uncontrolled addictive behavior, and poor psychosocial outcome. One study found that inmates with head injuries had higher levels of alcohol and marijuana use, a significantly higher frequency of depression, anxiety, and suicidal thinking, as well as more difficulties in concentrating and controlling violent behavior.

Substance use disorders are associated with worse post-TBI outcomes. Individuals with TBI who abuse substances show worse cognitive performance, higher rates of incident depression, decreased employment, and poorer rehabilitation participation in the years following injury than patients with TBI alone. One large study found that, among individuals with TBI, those who had consumed alcohol around the time of their injury were 20% more likely to die of their injuries and were less likely to be able to return home at hospital discharge.

The optimal treatment strategy for comorbid substance use disorders and TBI remains to be determined. One study of systematic motivational counseling (SMC) to treat substance abuse comorbidity among patients with TBI found that the SMC group showed significant improvements in motivational structure and a significant reduction in negative affect and the use of addictive substances. However, two RCTs of a brief motivational interviewing and educational intervention targeting alcohol use disorders among patients with TBI did not find effects on alcohol consumption, but one of the studies did indicate that the intervention increased awareness of the negative consequences of alcohol use. Physical disabilities, such as those which may accompany TBI, can pose barriers to substance abuse treatment, as many treatment programs do not accommodate them.

The evidence supporting pharmacologic intervention for alcohol or substance use is limited in the TBI population. A 12-week pilot RCT demonstrated that topiramate significantly reduced the number of drinks per week compared to placebo in veterans with mild TBI and alcohol use disorder. However, topiramate may induce adverse cognitive effects in patients with moderate and severe TBI. Another recent trial comparing valproate and naltrexone in veterans with and without TBI did not find one to be superior to the other; however, patients with moderate and severe TBI were more likely to relapse than those without TBI. Expanding substance abuse treatment access for these patients remains critical.

Psychotic Disorders

According to DSM-5 criteria, psychotic disorder due to TBI consists of prominent hallucinations or delusions occurring as a direct physiologic consequence of the TBI. Symptoms occurring exclusively in the context of delirium are excluded.

The reported incidence of psychosis following TBI varies widely. Retrospective reviews of World War II records found a rate of 7%, while one prospective study in an inpatient rehabilitation setting found that 64% of individuals with moderate or severe TBI experienced hallucinations and/or delusions at some point in their postinjury course. Psychotic symptoms were associated with poorer sleep and worse cognitive impairment and tended to occur earlier in the post-TBI recovery period.

Much of the literature on psychosis following TBI takes the form of case studies. Nineteen published reports describe patients with delusions only, including Capgras syndrome, delusional jealousy, Cotard syndrome, Fregoli syndrome, and somatic delusions. Seventy-one published reports describe patients with psychotic symptoms thought to resemble those of schizophrenia—that is, hallucinations and delusions; almost all of these patients experienced auditory hallucinations, one in five also experienced visual hallucinations, 85% experienced delusions (most commonly persecutory), and 43% experienced negative symptoms. On average, these 90 patients developed psychotic symptoms 3.3 years after the TBI. Thus, attributing psychosis directly to the TBI is challenging.

Care must be taken to consider other causes of psychosis, including iatrogenic effects, substance intoxication/withdrawal symptoms, epilepsy-related psychoses, and causes of delirium. Psychotic symptoms occurring during an acute post-traumatic confusional state should not be labeled as “psychotic disorder due to TBI” or another idiopathic psychotic disorder.

In line with the current hypothesis on the pathophysiology of psychotic disorders, post-traumatic psychosis has been associated with the presence of cognitive disturbance and imaging and EEG abnormalities in the frontal and temporal lobes. In addition, post-traumatic psychosis has been associated with seizure disorder in approximately one-third of cases.

A related clinical problem is whether TBI constitutes a risk factor for later schizophrenia. A recent meta-analysis of nine studies with a total of 180,859 persons found an elevated risk of subsequent development of schizophrenia in individuals who had experienced a TBI (RR 1.65, 95% CI 1.17–2.32). Individuals with a family history of schizophrenia appeared to face a particularly heightened risk after TBI; conversely, a family history of schizophrenia also represented a risk factor for experiencing a TBI, suggesting the possibility of a complex interaction between the two disorders.

To date, no clinical trials have examined the pharmacotherapy of psychosis due to TBI. By extrapolation from the treatment of idiopathic psychosis, atypical antipsychotic medications will likely represent the mainstay of treatment. Psychosocial treatments useful in patients with schizophrenia, such as psychoeducation of patients and family, vocational rehabilitation, psychotherapy, and assertive community treatment, may also help patients with psychosis due to TBI, though no studies have yet explored this.

FUTURE DIRECTIONS

In summary, cognitive and neuropsychiatric disturbances frequently complicate functional outcomes and quality of life following TBI. While the above sections briefly describe these disturbances and several treatment strategies, the evidence base supporting daily clinical decisions requires much expansion. Future research is needed to improve the identification of neuropsychiatric disorders, predict functional outcomes, optimize available treatments, and understand the relationships between TBI and neurodegeneration.

Ongoing research efforts are attempting to address some of these knowledge gaps. For instance, point-of-care serum biomarker devices are being investigated for the enhancement of TBI prognostication and case definitions in the acute setting. Other exciting areas of research involve the applications of noninvasive neurostimulation techniques to treat disorders of consciousness, cognitive impairments, and neuropsychiatric disorders. Studies of TBI-related sleep disturbances are increasingly demonstrating the detrimental impact of poor sleep on functional outcomes and longitudinal neurodegeneration risk. Current efforts to characterize the prevalence, risk factors, and clinical features of CTE will hopefully guide the clinical management of suspected cases.

The complex interplay between post-TBI neuronal degeneration and repair processes may produce long-lasting changes in neuronal activity and circuitry. Mechanisms of secondary injury continue to be investigated for potential targets of interventions that may reduce injury progression. Moreover, research efforts to characterize the functional changes observed in neuronal networks may provide insight into the neurologic substrates of post-TBI cognitive and behavioral deficits.

Multicenter consortia are critical for future study of the longitudinal course of mild TBI, relationships between TBI and neurodegeneration, and the development of biomarkers and other predictors of recovery postinjury. The Transforming Research and Clinical Knowledge in TBI (TRACK-TBI) study is an NINDS-funded, multicenter TBI data repository of neuroimaging, blood biospecimens, and detailed clinical outcomes post TBI. Similarly, the previously mentioned TRACTS cohort study collects a repository of neuroimaging, blood specimens, and cognitive and psychiatric outcomes for the purpose of investigating the longitudinal course of blast-related and other TBIs in post-9/11 veterans. Large, multicenter data sets obtained from these and other consortia promote the advancement of precision medicine by improving TBI classification and outcomes, particularly for mild TBI.

Finally, insights gleaned from research are made clinically applicable by the providers directly caring for patients with TBI. Additional efforts to incorporate principles of TBI pathophysiology and management into medical student and postgraduate education are needed to ensure that clinicians are attuned to the transdisciplinary needs of these patients.

ACKNOWLEDGMENTS

The authors report no proprietary or commercial interest in any product mentioned or concept discussed in this article. This work was supported by resources and facilities at the Michael E. DeBakey VA Medical Center. The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

Further Readings

Arciniegas DB, Frey KL, McAllister TW. Acute traumatic encephalopathy. In: Silver JM, McAllister TW, Arciniegas DB, eds. *Textbook of Traumatic Brain Injury*. 3rd ed. American Psychiatric Association; 2019:217–246.

Bahraini NH, Breshears RE, Hernández TD, Schneider AL, Forster JE, Brenner LA. Traumatic brain injury and posttraumatic stress disorder. *Psychiatr Clin North Am*. 2014;37:55–75.

Beedham W, Belli A, Ingaralingam S, Haque S, Upthegrove R. The management of depression following traumatic brain injury: a systematic review with meta-analysis. *Brain Inj*. 2020;34(10):1287–1304.

Brawman-Mintzer O, Tang XC, Bizien M, et al. Rivastigmine transdermal patch treatment for moderate to severe cognitive impairment in veterans with traumatic brain injury (RiVET study): a randomized clinical trial. *J Neurotrauma*. 2021;38(14):1943–1952.

Cicerone KD, Goldin Y, Ganci K, et al. Evidence-based cognitive rehabilitation: systematic review of the literature from 2009 through 2014. *Arch Phys Med Rehabil*. 2019;100(8):1515–1533.

Czeiter E, Amrein K, Gravesteijn BY, et al; CENTER-TBI Participants and Investigators. Blood biomarkers on admission in acute traumatic brain injury: relations to severity, CT findings and care path in the CENTER-TBI study. *EBioMedicine*. 2020;56:102785.

Fann JR, Bombardier CH, Temkin N, et al. Sertraline for major depression during the year following traumatic brain injury: a randomized controlled trial. *J Head Trauma Rehabil*. 2017;32(5):332–342.

Fann JR, Quinn DK, Hart T. Treatment of psychiatric problems after traumatic brain injury. *Biol Psychiatry*. 2022;91(5):508–521.

Fujii DE, Ahmed I. Psychotic disorder caused by traumatic brain injury. *Psychiatr Clin North Am*. 2014;37(1):113–124.

Gardner A, Iverson GL, McCrory P. Chronic traumatic encephalopathy in sport: a systematic review. *Br J Sports Med*. 2014;48(2):84–90.

Haacke EM, Duhaime AC, Gean AD, et al. Common data elements in radiologic imaging of traumatic brain injury. *J Magn Reson Imaging*. 2010;32(3):516–543.

Hammond FM, Malec JF, Zafonte RD, et al. Potential impact of amantadine on aggression in chronic traumatic brain injury. *J Head Trauma Rehabil*. 2017;32(5):308–318.

Hammond FM, Zafonte RD, Tang Q, Jang JH. Carbamazepine for irritability and aggression after traumatic brain injury: a randomized, placebo-controlled study. *J Neurotrauma*. 2021;38(16):2238–2246.

Hicks AJ, Clay FJ, Hopwood M, et al. Efficacy and harms of pharmacological interventions for anxiety after traumatic brain injury: systematic review. *J Neurotrauma*. 2021;38(5):519–528.

Iacono D, Raiciulescu S, Olsen C, Perl DB. Traumatic brain injury exposure lowers age of cognitive decline in AD and non-AD conditions. *Front Neurol*. 2021;12:573401.

Jahed S, Daneshvari NO, Liang AL, et al. Neuroimaging correlates of syndromal anxiety following traumatic brain injury: a systematic review of the literature. *J Acad Consult Liaison Psychiatry*. 2022;63(2):119–132.

Jorge RE, Li R, Liu X, et al. treating alcohol use disorder in U.S. veterans: the role of traumatic brain injury. *J Neuropsychiatry Clin Neurosci*. 2019;31(4):319–327.

Jorge RE, Robinson RG, Moser D, Tateno A, Crespo-Facorro B, Arndt S. Major depression following traumatic brain injury. *Arch Gen Psychiatry*. 2004;61:42–50.

Kakehi S, Tompkins DM. A review of pharmacologic neurostimulant use during rehabilitation and recovery after brain injury. *Ann Pharmacother*. 2021;55(10):1254–1266.

Loggini A, Tangonan R, El Ammar F, et al. The role of amantadine in cognitive recovery early after traumatic brain injury: a systematic review. *Clin Neurol Neurosurg*. 2020;194:105815.

Mayer AR, Quinn DK, Master CL. The spectrum of mild traumatic brain injury: a review. *Neurology*. 2017;89(6):623–632.

Mohamed AZ, Nestor PJ, Cumming P, Nasrallah FA; Alzheimer's Disease Neuroimaging Initiative. Traumatic brain injury fast-forwards Alzheimer's pathology: evidence from amyloid positron emission tomography imaging. *J Neurol*. 2022;269(2):873–884.

Neumann DM, Zupan B, Eberle R. Social cognition. In: Silver JM, McAllister TW, Arciniegas DB, eds. *Textbook of Traumatic Brain Injury*. 3rd ed. American Psychiatric Association; 2019:281–298.

Neurobehavioral Guidelines Working Grou; Warden DL, Gordon B, McAllister TW, et al. Guidelines for the pharmacologic treatment of neurobehavioral sequelae of traumatic brain injury. *J Neurotrauma*. 2006;23(10):1468–1501.

Peterson AB, Xu L, Daughtery J, Breiding MJ. Surveillance report of traumatic brain injury-related emergency department visits. *Hospitalizations, and Deaths*. Accessed July 06, 2021. https://www.cdc.gov/traumaticbraininjury/pdf/TBI-Surveillance-Report-FINAL_ 508.pdf

Polich G, Iaccarino MA, Zafonte R. Psychopharmacology of traumatic brain injury. *Handb Clin Neurol*. 2019;165:253–267.

Ponsford J, Alway Y, Gould KR. Epidemiology and natural history of psychiatric disorders after TBI. *J Neuropsychiatry Clin Neurosci*. 2018;30(4):262–270.

Quinn M, Agha A. Post-traumatic hypopituitarism-who should be screened, when, and how? *Front Endocrinol (Lausanne)*. 2018;9:8.

Rahmani E, Lemelle TM, Samarbafzadeh E, Kablinger AS. Pharmacological treatment of agitation and/or aggression in patients with traumatic brain injury: a systematic review of reviews. *J Head Trauma Rehabil*. 2021;36(4):E262–E283.

White DL, Kunik ME, Yu H, et al. Post-traumatic stress disorder is associated with further increased Parkinson's disease risk in veterans with traumatic brain injury. *Ann Neurol*. 2020;88(1):33–41.

2.6 Neuropsychiatric Aspects of Movement Disorders

Laura Marsh, M.D., and Russell L. Margolis, M.D.

Most often, and consistent with the focus of this chapter, the term *movement disorders* refers to a group of central nervous system (CNS) diseases that primarily involve neurodegeneration of the basal ganglia, cerebellum, or both. From the clinical perspective, however, movement disorders mark the crossroads of the fields of psychiatry and neurology and underscore the importance of interdisciplinary care and perspectives. This is because psychiatric and cognitive abnormalities figure prominently in many patients, resulting in excess morbidity and caregiver burden. Diagnosis and treatment of neuropsychiatric conditions in patients with primary movement disorders are also challenged by the interplay of motor, cognitive, and psychiatric features over the course of the disease and their degree of overlap at the time of an examination. Examples are the masked face of parkinsonism and flat affect in depressive disorders. Furthermore, management of psychiatric disorders in patients with movement disorders is potentially complex; treatments of motor symptoms can exacerbate or cause cognitive or psychiatric symptoms, and, conversely, psychiatric treatments may result in adverse motor and cognitive side effects. Appreciation of the distinctive motor, cognitive, mood, and behavioral aspects of the various movement disorders and their underlying pathophysiology helps to target therapies that are effective and limited in their risks.

Comparative Nosology

Classification of Movement Disorders. Several approaches are used to classify movement disorders (Table 2.6–1). From the phenomenologic standpoint, movement disorders are generally grouped into three main categories. Parkinsonian disorders are characterized by reduced amplitude of movements (hypokinesia), slowness of movements (bradykinesia), and difficulty initiating movements (akinesia). Dyskinetic disorders are characterized by increased motor activity and excessive (hyperkinetic) involuntary movements such as tremor, chorea, ballism, tics, dystonia, or myoclonus. Ataxias involve gait disturbances, dysmetria, and dysdiadokinesis. Nosologically, the phenomenologic scheme is problematic because clinical features overlap across diseases. In addition, syndromes frequently involve several types of movement abnormalities, such as akinesia, dystonia, and tremor in Wilson disease (WD), while others, such as focal dystonia, consist of a single movement abnormality. Classification schemes based on neuropathologic findings generally distinguish diseases in which the primary pathology involves the basal ganglia from diseases that primarily affect cerebellum, cerebral cortex, or motor neurons. Again, common clinical features and neuronal system involvement across different neuropathologic groups limit this approach. An etiologic classification has the advantage of linking a syndrome directly to its cause and directing attention at specific modes of pathogenesis, treatment, and prevention. However, the cause of many diseases is unknown (e.g., idiopathic Parkinson disease [PD]), and an etiologic nosology can lead to confusion since entities with clinical similarities are separated into different categories.

Table 2.6–1.

Classification Schemes for Movement Disorders, with Examples

Major Clinical Feature	Location of Primary Neuropathology	Etiologic
Hypokinetic movements Idiopathic Parkinson disease Parkinsonian syndromes Progressive supranuclear palsy Corticobasal degeneration Frontotemporal dementia Dementia with Lewy bodies Multiple system atrophy—parkinsonian type *Hyperkinetic movements* (See also Table 2.6–4) Huntington disease Wilson disease Essential tremor Dystonias Ataxia Spinocerebellar ataxia Multiple system atrophy—cerebellar type	*Basal ganglia* Idiopathic Parkinson disease Parkinsonian syndromes Progressive supranuclear palsy Huntington disease Wilson disease Dystonias *Cerebellum* (see also Table 2.6–5) Friedreich ataxia Some spinocerebellar ataxias *Cerebral cortex* Frontotemporal dementia *Motor neuron* Amyotrophic lateral sclerosis *Multiple affected regions* Multiple system atrophy—both types Corticobasal degeneration	*Genetic* Huntington disease Parkinson disease secondary to α-synuclein or Parkin mutations Friedreich ataxia Some spinocerebellar ataxias Frontotemporal dementia with parkinsonism linked to chromosome 17 Familial amyotrophic lateral sclerosis Wilsons disease Familial essential tremor Some dystonias *Medication induced* Neuroleptic-induced parkinsonism or dystonia Toxin induced MPTP, manganese, and carbon monoxide–induced parkinsonism *Infectious* Postencephalitic parkinsonism Traumatic Post-traumatic parkinsonism, tremor, dystonia, or frontotemporal dementia *Vascular* Vascular parkinsonism or ataxia *Idiopathic* Most Parkinson disease Progressive supranuclear palsy

	Dementia with Lewy bodies Some spinocerebellar ataxias	Corticobasal degeneration Some cases of frontotemporal dementia Dementia with Lewy bodies Multiple system atrophy Essential tremor Sporadic spinocerebellar ataxia

Classification of Psychiatric Disturbances Associated with Movement Disorders. In accordance with the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5), psychiatric disturbances associated with movement disorders can generally be classified as “due to another medical condition,” such as depression due to Huntington disease (HD), psychotic disorder secondary to dementia with Lewy bodies (DLBs), or anxiety disorder secondary to PD. Additional qualifiers designate specific features of the presentation, such as *with hallucinations or with major depressive–like episode*. The prominence of psychiatric and cognitive manifestations in movement disorders, however, points out the arbitrariness of this scheme; it could be equally valid for many of the movement disorders to be classified as neurocognitive disorders with associated movement and psychiatric phenomena, for example, probable major neurocognitive disorder with Lewy bodies, with behavioral disturbance. For that reason, many research studies establish diagnoses using the DSM phenomenologic criteria for idiopathic psychiatric conditions, that is, from the revised fourth edition of DSM (DSM-IV-TR), and now the DSM-5, while ignoring the criterion stating that the disturbance is “not attributable to the physiological effects of a substance or to another medical condition.” Another shortcoming of the DSM nosologic system is that its terminology does not capture the nuances of many psychiatric disturbances. Examples are the disruptive perseverative behaviors seen with HD or pathologic laughter or crying syndromes that may occur with progressive supranuclear palsy (PSP).

Bedside Assessment of Movement Disorders

Emotion, cognition, and movement are controlled by a series of interrelated and interdependent circuits involving the basal ganglia and frontal–subcortical circuits. It is therefore not surprising that patients with motor abnormalities often have emotional or cognitive disorders and patients with disorders of emotion or cognition frequently have motor abnormalities.

While often relegated to the specialty of neuropsychiatry, the careful appraisal and appreciation of movement abnormalities and movement disorders in psychiatric patients should be considered an important aspect of general psychiatric practice. First, side effects of many psychiatric medications include abnormal movements, some of which may be long-lasting or even permanent. Second, some patients demonstrate abnormal movements before exposure to psychiatric drugs, underscoring the need for baseline motor screening in all patients. Abnormal movements in such patients may indicate the presence of gross congenital disorders (such as cerebral malformations), developmental disorders without gross brain malformation (such as the “soft signs” seen in patients with schizophrenia or autism), neurodegenerative disorders (such as HD or PD), other cerebral insults (including strokes and tumors), or even normal aging. Third, powerful emotions can lead to abnormal movements in otherwise healthy individuals as well as those with CNS disturbances (e.g., tremulousness with anxiety). Fourth, primary psychiatric disorders may present with movement abnormalities, as illustrated by motor slowing seen with depression or the motor hyperactivity with mania. When the motor examination is accompanied by an adequate history and suspicion, overlapping motor and psychiatric signs can be recognized and distinguished according to the relative influences of a primary motor disorder, a psychiatric disturbance, a drug effect, or some interaction among them. For example, an older adult patient with melancholic depression may walk slowly with a hunched posture, but also display unilateral upper extremity rigidity, bradykinesia, and a rest tremor, suggesting a diagnosis of PD and the need for additional treatment beyond antidepressant agents.

In patients with known movement disorders, the initial history and physical examination should confirm the presence of the diagnosed movement disorder. The causes, course, and treatment of the different movement disorders are diverse, but most share similar features. Given this, there is frequent misdiagnosis of movement disorders; only 75% to 80% of patients with parkinsonism are correctly diagnosed with idiopathic PD during life by movement disorder specialists and the rate of misdiagnosis is probably higher among nonspecialists. Examples of disorders that mimic PD, especially early in their course, are PSP and DLB, as well as medication-induced parkinsonism.

A second goal of the motor examination is to evaluate the extent of deficits relative to the reported impairment. Complaints of motor dysfunction can be exaggerated in patients with depression, whereas dementia or lack of insight, as is often the case with HD, can contribute to denial of illness or impulsive behaviors that increase the risk of falls. Finally, as mentioned above, serial motor examinations provide a barometer of motor side effects caused by agents used to treat psychiatric conditions.

Table 2.6–2.
Screening Examination for Motor System Dysfunction

Gait-observe walking, including on tiptoes and heels, and turning
Station-observe balance, pull test
Abnormal movements—with patient sitting at rest and with arms outstretched, observe limbs, trunk, face, and head for abnormal and excessive movements or hypokinesia
Coordination-rapid finger tapping, finger-to-nose test
Tone-passive movement of limbs and neck
Strength

The main elements of a motor function screen (Table 2.6–2) include assessment of gait, involuntary movements, coordination, muscle tone, and strength, as discussed further in the chapter on neuropsychological assessment. Gait provides clues to the presence of upper versus lower motor neuron disease (MND), sensory loss, and cerebellar versus extrapyramidal pathology and should first be observed when a patient walks to the examining room. Ability to rise independently from a chair and without pushing off the armrests is a common but nonspecific screen for gait difficulties and increased fall risk in the geriatric population. Balance should be examined separately. Hyper- and hypokinetic movement abnormalities can be observed during the interview while the patient rests comfortably in a chair. The five main types of abnormal hyperkinetic involuntary movements are tremor, chorea, myoclonus, tics, and dystonia. Having the patient hold their arms outstretched will reveal forms of nonrest tremor, other adventitious movements, and motor weakness. Finger-to-nose testing permits examination of large proximal arm movements, kinetic tremor, and dysmetria. Hypokinesia, characteristic of parkinsonian syndromes, is manifest by bradykinesia and akinesia. There is a relative absence of spontaneous movements or gestures, slowed movements, decreased arm swing, and diminished facial expressiveness (masked face) and spontaneous blinking. Gait disturbances are related to difficulties initiating and sustaining movements and turning. Coordination can be assessed with finger-thumb tapping. Motor tone is tested by passive movement of an extremity or the neck while the patient is at rest. When assessing tone, the examiner should note whether there is increased resistance to the passive movements, and whether that resistance is active, spastic, or “lead pipe” in character. Cogwheeling, a feature of a tremor, is often mistaken for a measure of tone.

PARKINSON DISEASE

Introduction and Definition

PD is a progressive neurodegenerative disorder that is associated with loss of dopaminergic neurons in the substantia nigra pars compacta (SNpc). The hallmarks of the disease are its triad of motor features—resting tremor, rigidity, and akinesia/bradykinesia (inability to initiate movement and slowness of movement, respectively). Gait and postural disturbances also characterize the disease. Parkinsonism or parkinsonian syndromes are less specific and nonetiologic terms that refer to the clinical appearance of PD. Examples are parkinsonism secondary to antidopaminergic medications, without pathology of the substantia nigra, or atypical parkinsonian syndromes such as multiple system atrophy (MSA).

History

Descriptions of parkinsonian syndromes are found in the earliest medical writings. Idiopathic PD is named after James Parkinson, a general practitioner in London who wrote about *paralysis agitans* as a distinct disease separate from other causes of tremor in his classic 1817 article, “Essay on the Shaking Palsy.” Parkinson’s description of six cases highlighted the cardinal motor signs that remain characteristic of the disease today. Parkinson’s case descriptions included neuropsychiatric as well as other physical changes, such as constipation. However, his conclusion of the “senses and intellect being uninjured” led to relatively less emphasis on these aspects of PD. Subsequent descriptions recognized memory loss and other neuropsychiatric features, but it is only relatively recently that these have received greater attention. It is worth noting that profound akinesia and speech difficulties may have limited evaluation of the mental state in patients with PD until effective treatment with levodopa became available in the 1960s, thus leaving psychiatric disorders often unrecognized. Although anticholinergic belladonna alkaloids were used to treat PD in the 19th century, the discovery that dopaminergic precursors reversed reserpine-induced parkinsonism led to the treatment of PD with levodopa, still the gold standard of antiparkinsonian medications.

Epidemiology

Parkinsonian syndromes frequently present in older adults and their prevalence increases with advancing age, affecting up to 15% of individuals up to age 74 and over half of those over age 85. The most common parkinsonian syndrome is PD. It is the second most common neurodegenerative disease after Alzheimer disease (AD), and it affects over 1% of the population over age 55 and nearly 3% of the population over age 70. Worldwide, an estimated 4 to 6 million people have PD. The mean age of onset for PD is 60 years of age and 80% of individuals develop the disorder between the ages of 40 to 70 years. About 5% of patients have symptom onset before age 40 and, in such cases, the disease is classified as young onset PD. Juvenile onset occurs in less than 1% of cases. The wide range in the age of onset of PD contributes to misdiagnosis, especially in younger patients.

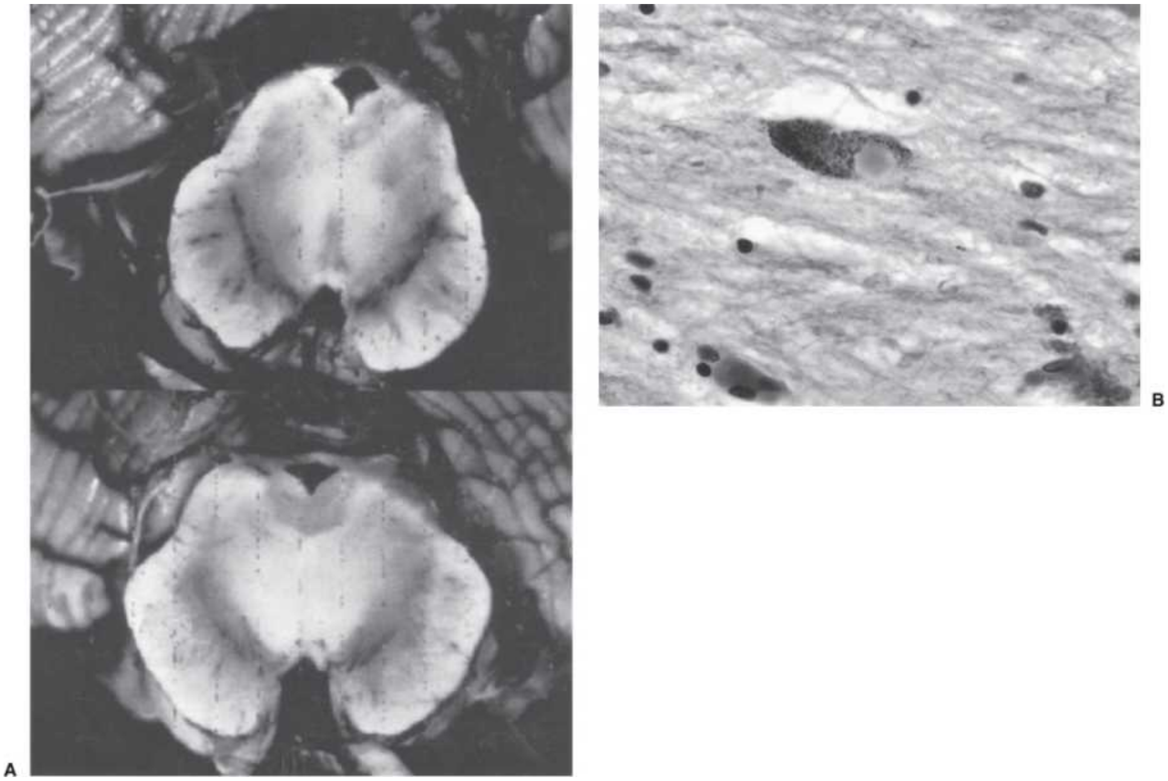

FIGURE 2.6–1. Parkinson disease neuropathology. **A:** Gross pathology in cross-section of the midbrain. Note pallor and depigmentation of the substantia nigra in PD (**bottom**) compare to control (**top**) brain. **B:** Lewy body in a pigmented neuron of the substantia nigra. Lewy bodies are cytoplasmic inclusions with a dense eosinophilic core surrounded by a clear halo. (Photographs courtesy of The Morris K. Udall Parkinson's Disease Research Center and the Alzheimer's Disease Research Center at Johns Hopkins and Drs. Olga Pletnikova and Juan Troncoso.)

Sex and ethnic differences in the frequency of PD can be an important source of variation because of the association of PD with genetic and environmental factors. Taken together, prevalence studies indicate some age-range–associated differences in PD prevalence by geographic location and sex. For example, prevalence among individuals in their eighth decade is higher in European, North American, and Australian populations relative to Asians. Prevalence of PD, with and without dementia, is higher in men as compared to women. With increasing age, incidence of PD increases in both sexes. Across all age groups, incidence is consistently higher among males, with differences in age-specific pooled male:female incidence ratios achieving statistical significance among those between 60 to 79 years of age. Incidence of PD has been reported as lower among African Americans than Caucasians; discrepant rates may be a function of socioeconomic factors, such as access to health care and perception of the disease.

Etiology

The neuropathologic hallmarks of PD include progressive loss of dopaminergic neurons in the ventral tier of the SNpc and intraneuronal aggregation of the α-synuclein protein in the form of Lewy bodies and Lewy neurites (Fig. 2.6–1A). When approximately 60% to 80% of the SNpc neurons degenerate and dopamine production is inadequate, the motor symptoms of PD emerge. What leads to selective and progressive cell death is unknown. Transformation of α-synuclein, a presynaptic protein, to fibrillary and oligomeric forms and their aggregation appears central to neurotoxicity. As such, PD belongs to the family of neurodegenerative disorders referred to as synucleinopathies, which also includes DLB and MSA.

A prevailing theory of the pathophysiology of PD, referred to as the Braak hypothesis, suggests that initial symptoms of PD involve nondopaminergic neuronal degeneration in the enteric nervous system, olfactory bulb, and other brain regions (e.g., raphe nuclei, locus coeruleus, and dorsal motor nucleus of the vagus) followed by rostral progression to the SNpc over many years. This phase, beginning up to 20 years before onset of motor symptoms, is termed “prodromal.” Clinical studies support the Braak hypothesis in that nonmotor symptoms, such as depressive and anxiety disturbances, olfactory dysfunction, rapid eye movement sleep behavior disorder (RBD), constipation, and autonomic disturbances, often precede onset of motor symptoms by several years. These conditions represent potential risk markers for PD and provide a basis for alternative therapeutic interventions that modulate noradrenergic, serotonergic, glutamatergic, and other pathways

Research over the last two decades demonstrates that the contribution of genetic factors to the pathophysiology of PD is much greater than previously expected. Most cases of PD (>85%) are regarded as sporadic (not inherited), but, now, over 20 monogenic forms of PD are described and genome-wide association studies (GWAS) have identified over 100 loci as risk factors for PD. Taken together, it appears that most cases of PD result from effects of multiple predisposing genetic factors, possibly in combination with select environmental insults. Proposed mechanisms of neuronal degeneration include α-synuclein misfolding and aggregation, mitochondrial dysfunction, oxidative stress, excitotoxicity, perturbations of neuronal cytoskeleton and axonal transport, neuroinflammation, and programmed cell death.

Though several risk factors have been proposed, the only unequivocal risk factor for PD is increasing age. Dopaminergic cell loss also occurs with normal aging, but the loss is less severe and involves different subpopulations of SNpc neurons than PD. Toxins associated with parkinsonism include viral infection of the CNS and industrial exposures, including manganese, carbon monoxide, organophosphates, and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP). Higher prevalence in industrialized regions and associations with well water consumption, farming, and rural living may reflect exposure to neurotoxins contained in pesticides and other agricultural agents that contain homologs of MPTP. By contrast, other environmental factors and behaviors, such as caffeine use and cigarette smoking, are associated with a reduced risk to develop PD.

Familial forms of PD are present in about 15% of individuals with the disease. Studies of large family and twin samples with heritable parkinsonian conditions have led to identification of over 20 causative genes, to date. One of the most prevalent genes associated with familial and sporadic PD is leucine-rich repeat kinase 2 (LRRK2), identified in 2004 and inherited in an autosomal dominant pattern. The locus for LRRK2, called PARK8, is located on chromosome 12q12. Its most common pathogenic mutation, G2019S, is found in up to 5% of Caucasian patients with familial PD. In Ashkenazi Jews and African Arabs, G2019S mutations are present in up to 30% of familial PD cases and 13% to 40% of sporadic PD cases. The G2385R LRRK2 variant is the most common genetic risk factor identified among Asian people, accounting for 23% of familial cases as well as 4% to 9% of early and typical late-onset cases. Most LRRK2 carriers with PD resemble noncarriers with PD, though LRRK2 carriers who develop PD may have a more benign course than noncarriers with PD. The LRRK2 protein has a number of functions; its proposed mechanism in PD is thought to involve dysregulation of kinase and GTPase activities.

The α-synuclein gene (SNCA) mutation, located on chromosome 4q21-q23, was the first mutation identified in association with autosomal dominant PD. Although SNCA mutations are a rare cause of PD, the study of α-synuclein led to the discovery that it is a major component of

Lewy bodies and demonstrated the role of protein aggregation in the pathophysiology of PD. Since patients with sporadic PD, who have normal α-synuclein genes, have Lewy bodies that stain positive for α-synuclein and ubiquitin, neuronal cell death in PD is attributed to abnormal accumulations of these proteins. Gene mutations on chromosome 6 that code for Parkin (PARK2), a protein involved in the ubiquitination pathway, are the most common inherited defect and are associated with a juvenile autosomal recessive form of early-onset PD. Alterations in gene loci on chromosome 4 that affect ubiquitin processing (UCH-L1) and on chromosome 1 that affect β-glucocerebrosidase and lysosomal function (GBA) are also linked to PD. While rare, other identified genes are associated with autosomal dominant and autosomal recessive parkinsonian syndromes, the latter including the oncogene DJ-1 (PARK7), the MAPT (tau) gene, progranulin (PGRN), FBXO7 (PARK-15), and PINK1.

Current models of basal ganglia/cortical physiology suggest the presence of five adjacent but anatomically distinct frontocorticostriatothalamic loops with motor, oculomotor, limbic, anterior cingulate, and prefrontal targets and that dysfunction of these neural loops is linked to the motor and nonmotor features in PD and other basal ganglia disorders. A common feature of these cortical–subcortical loops is their projection from higher-level cortical areas to the basal ganglia via striatal areas with outflow through the internal globus pallidum (Gpi) or its histologic analog, the substantia nigra pars reticulata (SNpr) (Fig. 2.6–2). Gpi/SNpr outputs then project to specific areas of the thalamus that complete the loops through projections to the original cortical area. Within each loop, the striatonigral component includes a "direct" pathway that projects directly from striatum to Gpi/SNpr and an "indirect" pathway that flows from the striatum through the subthalamic nucleus (STN) to reach Gpi/SNpr.

Motor Manifestations. Human and animal models of PD demonstrate that loss of SNpc dopaminergic neurons affects frontocorticostriatothalamic circuits sufficiently to result in PD motor manifestations. The best studied of these neuronal circuits is the "motor circuit" as illustrated in Figure 2.6–2. In PD, dopaminergic degeneration appears to have differential effects on the nigrostriatal pathway. Projections from SNpc to striatal targets are inhibitory in the dopamine type 1 (D1) receptor–mediated direct path and excitatory in the D2 receptor–mediated indirect path. Normal movement results from coordinated regulation of thalamocortical excitation by the direct and indirect pathways, with the presence of nigrostriatal dopaminergic projections tending toward increased movement by facilitation of the direct pathway and inhibition of the indirect pathway. Loss of dopamine results in hypokinetic symptoms secondary to overactivity of the STN and Gpi via the indirect path, which increases inhibitory input to the thalamus and then reduces the excitatory thalamocortical activity that ordinarily facilitates motor movements. Hyperkinetic movements, such as levodopa-induced dyskinesias, occur with overactivation of the direct pathway.

Cognitive Manifestations. Nearly all patients with PD experience some degree of cognitive change, with deficits ranging from slightly abnormal with minimal impact to very severe and disabling and different rates of progression. While there is considerable heterogeneity across patients, a typical cognitive profile early in the course of PD involves preserved intellectual abilities with mild-to-moderate impairments in memory retrieval, executive functions, and visuospatial functions. Word-finding difficulties may be present, but language functions are generally intact. Mild cognitive impairment in PD (PD-MCI), affecting about 80% of PD patients, is defined on the basis of single- or multidomain (e.g., mild memory and executive dysfunction) deficits on neuropsychological testing that are not associated with significant interference with functional independence. The symptoms are frustrating and may preclude gainful employment, especially with disease progression but awareness is usually retained so that compensatory strategies can be used to mitigate the impact of cognitive and motor deficits on daily functioning. PD-MCI is demonstrated in about 20% of patients upon diagnosis of PD and in 40% to 50% after 5 years. The neuropathologic underpinnings of cognitive dysfunction in PD are complex and not fully understood. Executive dysfunction in PD, primarily associated with frontal lobe dysfunction and disrupted frontostriatal networks, may also involve dysfunction of nonmotor neural circuits, including mesocortical and mesolimbic dopaminergic and nondopaminergic projections that subserve memory, motivational, and response inhibition functions. Bradyphrenia may be associated with noradrenergic cell loss.

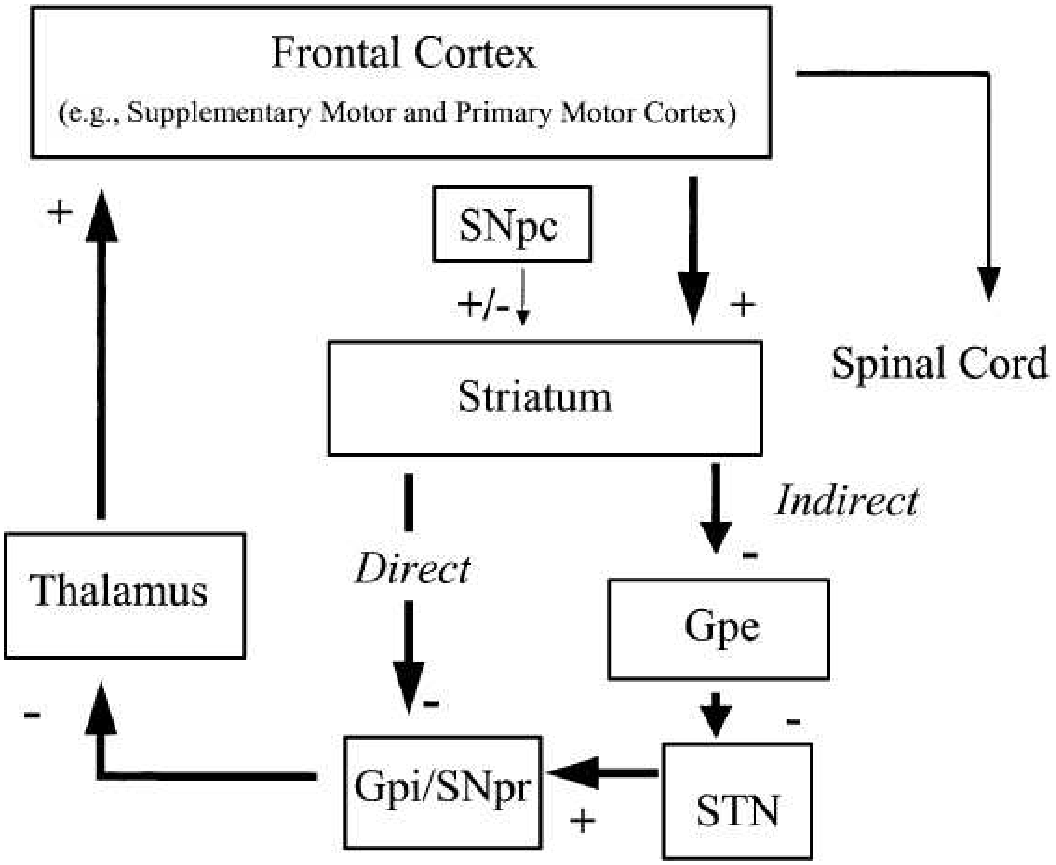

FIGURE 2.6–2. Corticostriatalthalamic circuits. In the motor circuit, the supplementary motor area and primary motor cortices in frontal regions project mainly to putamen and then “directly” to Gpi or “indirectly” via the external globus pallidus (Gpe) and then the STN to reach Gpi. Excitatory (glutamatergic) cortical projections generally decrease Gpi activity through the direct pathway and increase Gpi activity through the indirect pathway. Inhibitory (GABAergic) Gpi projections to thalamus and excitatory thalamocortical projections result in net excitation (increased motor activity) or net inhibition (decreased movement) through the direct and indirect pathways, respectively. Dopamine has differential inhibitory and excitatory effects on the nigrostriatal pathway.

Over time, risk for developing dementia in PD (PD-dementia) is sixfold higher in comparison to the general population after controlling for age, gender, and educational level. Point prevalence is estimated to be 24% to 30%; prevalence is lower within the first 5 years after diagnosis (<5%) and increases with disease duration, affecting over 80% of patients followed for 20 years. In patients with PD-MCI, an estimated 20% to 60% progress to dementia within 5 years, but there is considerable heterogeneity and severity of PD-dementia. The risk of PD-dementia is associated with increasing age, a family history of dementia, depression, and more severe motor dysfunction. Neuropathologic and imaging studies in patients with PD-dementia are associated with neuronal loss in cholinergic cells of the basal forebrain, implicating cholinergic deficits in PD-related cognitive impairment. Cortical Lewy bodies may also be associated with dementia. The histopathologic changes of AD—neurofibrillary tangles and senile plaques—are seen in some patients with PD and dementia.

Psychiatric Manifestations. Other than PD itself, there are no known specific risk factors for psychiatric disturbances in PD. Though psychological reactions to the experience of living with PD can influence occurrence of depression, most evidence suggests that depressive disorders, like motor and cognitive features, are related to the underlying disease processes of PD. Most compelling is evidence that depressive and anxiety disorders frequently develop *before* the onset of motor symptoms, suggesting that these mood phenomena are either neurologic signs of the disease (i.e., prodromal PD) or a risk or causative factor for development of PD. Cross-sectional studies do not show consistent relationships between depression onset and severity to age onset or duration of PD, motor severity, disease subtype (tremor dominant vs. akinetic-rigid), or disease stage. Several studies link depression with excessive serotonergic loss from the midbrain raphe. Depression in PD is also associated with decreased blood flow in frontal regions, a finding also seen in idiopathic depression. In apathetic syndromes, noradrenergic dysfunction is implicated by the relationship between bradyphrenia and neuronal loss in the locus coeruleus. Noradrenergic dysfunction is also evident in anxiety disorders, including an imbalance in noradrenergic/dopaminergic tone. Psychosis in PD is often associated with dopaminergic therapy, which may lead to hypersensitivity of mesocortical and mesolimbic dopaminergic receptors. Other factors related to the incidence of psychosis are cognitive impairment, disease severity, mood disorders, impaired visual acuity, other psychoactive medications, and delirium. An association of psychosis with cognitive impairment and mood disorders implicates more widespread brain pathology involving other neurotransmitter systems, including cholinergic deficits or a serotonergic/dopaminergic imbalance.

Diagnosis and Clinical Features

Motor Manifestations. Motor manifestations of PD initially affect one side of the body and gradually progress to the contralateral side, though remaining asymmetric. A slow (3- to 6-Hz) classic resting tremor is often the presenting feature of PD, but other types of tremor may be present and 20% to 30% of cases have no tremor at all. Bradykinesia and akinesia are the most disabling aspects of PD and contribute to problems with dexterity and gait. Patients typically walk in a flexed forward posture without swinging their arms. The gait is slow and shuffling with small steps and a tendency to fall forward and hasten the rate of walking (festination). Difficulty initiating movements leads to start hesitation and "freezing." For example, a patient may suddenly be unable to move their legs while walking or their hands when reaching into a purse. Deteriorated handwriting, weakness, and clumsiness in one limb are also common initial symptoms of bradykinesia. Rigidity is appreciated as increased muscle tone on examination. Patients may note muscle cramps, weakness, or stiffness as the first indication of rigidity.

In the absence of a biologic marker for PD, its diagnosis is based on the clinical examination and associated history and course. Most experts require presence of two of the three cardinal motor signs to diagnose PD along with a robust response to an initial trial of levodopa. Early in its course, PD can be difficult to diagnose since signs of the disease may be subtle and not even perceived by the patient. Parkinsonian signs can also be a feature of normal aging or other parkinsonian syndromes. The initial symptoms in up to 20% of patients are nonmotor, including fatigue, musculoskeletal complaints, and depression. Other nonmotor somatic symptoms of PD include sialorrhea, dysarthria, visual and genitourinary dysfunction, sleep disturbances, sweating, seborrhea, edema, constipation, paresthesias, and hyposmia. As mentioned, the prodromal phase may be characterized by depression, anxiety, and musculoskeletal discomfort. Heightened vigilance for emerging clinical signs of PD during this phase might result in earlier disease detection, especially in older adults who present with new-onset depression or anxiety.

Cognitive Manifestations. Cognitive impairment in PD ranges from mild executive dysfunction to dementia and affects nearly all patients with PD. Commonly, cognitive decline is not obvious clinically and scores on tests that screen for dementia, such as the Mini Mental State Examination (MMSE), remain in the "normal" or nondemented range for some patients who nonetheless have dementia. The concept of “subcortical dementia” has been used to describe the cognitive impairment that is frequently associated with the various movement disorders characterized primarily by subcortical neuropathology. Executive dysfunction is especially common, resulting from deficits that affect the ability to process new information and anticipate, plan, initiate, maintain, and change behaviors. Memory, visuospatial skills, information processing speed, attention, explicit recall, spatial planning, perseveration, verbal fluency, and poverty of thought are also affected. Compared to patients with AD, recognition memory, aphasia, agnosia, apraxia, and higher language functions are relatively spared. Some patients with PD, however, demonstrate wider involvement of cortical functions, including aphasia, apraxia, and memory deficits, and may have a cognitive presentation indistinguishable from AD. Task forces, commissioned by the Movement Disorder Society, delineated consensus diagnostic criteria for PD-MCI and PD-dementia and procedures that can be used to identify and distinguish cognitive decline in PD at its earliest stages. These diagnostic criteria and procedures, while subject to refinement with further research, including validation, benefit clinical care and research by facilitating communication about the clinical cognitive syndromes in PD.

Psychiatric Manifestations. Psychiatric disturbances affect up to 90% of patients at some point during the course of PD and more than one psychiatric disturbance is often present. Early in the disease process, adult-onset anxiety and depressive disorders precede the obvious onset of motor symptoms in up to 30% of patients.

DEPRESSION. Normal psychological reactions to the illness can result in low mood, grief, demoralization, frustration, and embarrassment. Depressive disorders in PD involve more pervasive mood changes and generally resemble idiopathic forms of depression.

Reported prevalence ranges from 20% to 90%, with an average reported frequency of 40% to 50%. Up to half of those with depression, in most studies, have major depression. The majority of PD patients with depression have nonmajor depressive syndromes that are variably referred to as minor depression, persistent depressive disorder (dysthymia), or subsyndromal depression. Nonmajor forms of depression often progress to major depression if untreated. Anxiety symptoms and comorbid anxiety syndromes such as panic disorder are especially prominent, but self-blame, guilt, delusions, a negative self-attitude, self-destructive thoughts, a sense of failure, and anhedonia may be less evident. Suicide can occur, but at a similar rate as for the general population. Other psychological, vegetative, and autonomic features of depression are also common, but early morning awakening, anergia, and retardation occur at comparable rates in nondepressed patients with PD. Depression is not consistently associated with a family history of mood disorders.

APATHY. Apathy occurs in at least 25% of patients and often coexists with depression. It manifests as indifference and a lack of motivation, initiative, perseverance, interest in new things, or concern for ones health. In most cases, apathy is not distressing to the patient, but the absence of spontaneous effort and interest can be frustrating for caregivers, who may perceive the patient as depressed. Inactivity, deconditioning, and fatigue contribute to further loss of function and increased disability and suffering. Other neurologic signs and symptoms such as akinesia, hypomimia, hypophonia, cognitive dysfunction, and bradyphrenia can confound recognition of apathy. Selective serotonin reuptake inhibitors (SSRIs), the most frequently prescribed antidepressants for patients with PD, can also cause apathy, though this has not been demonstrated in PD specifically.

ANXIETY. Difficulties initiating movements can frequently lead to anxiety-provoking circumstances in the course of living with PD, but pathologic anxiety is still more common in PD compared to other parkinsonian syndromes. Particularly common are generalized anxiety disorder, social phobia, and panic disorder, which have a prevalence of 25% in some series. In general, anxiety syndromes in PD resemble those in idiopathic conditions and frequently co-occur with depression. There are also anxiety disturbances unique to PD, particularly anxiety associated with fluctuations in levodopa levels and motor function, although there is not always a strict temporal relationship between anxiety and motor “off” states. Some patients describe episodic panic attack–like states that develop at the same time every day and can be associated with motor deficits. Sleep difficulties are very common and seem to aggravate daytime anxiety.

PSYCHOSIS. Psychosis affects about 25% to 40% of patients, with increased prevalence as the disease progresses. The temporal pattern and clinical profile of PD psychotic symptoms in PD differs from other psychotic disorders such as substance-induced psychosis or schizophrenia. Lack of standardized diagnostic criteria for PD-associated psychosis contributed to variability in reported prevalence. As studies were often restricted to visual hallucinations, the most common psychotic symptom in PD, new criteria for PD-psychosis were proposed for in 2007 by a National Institutes of Health (NIH)-sponsored work group.

In PD, psychotic symptoms include hallucinations and delusions. Atypical symptoms, referred to as minor psychotic phenomena, affect over two-thirds of patients and include sense of presence, passage hallucinations, and illusions. It is a misnomer to refer to minor symptoms as “benign,” as they, like other psychotic symptoms in PD, are associated with more depressive symptoms and worse quality of life. While visual hallucinations are most common, and are often related to dopaminergic medications, other sensory modalities can be involved. Typically, in earlier stages of PD, visual hallucinations include well-delineated animals or people that occur in a clear sensorium. They are not necessarily distressing, and insight may be retained. Later in the disease course, some patients experience persistent hallucinations and/or delusions in a clear sensorium but with diminished insight. Aggression, delusional jealousy and other paranoid accusations, delusional fears about eating or medications, inappropriate sexual behaviors, and confusion are frequent manifestations that can be difficult to manage and may require hospitalization. Some patients develop chronic paranoid syndromes that resemble schizophrenia. Hallucinations and delusions also occur with major depression, mania, or anxiety disorders, and may be overlooked when patients are especially agitated. Fragmented sleep, vivid dreams, and nightmares may herald the onset of chronic psychosis. Hallucinations and delusions associated with advanced PD are a major reason for nursing home placement. Another category includes hallucinations or delusions in the context of delirium related to antiparkinsonian medications and/or other causes.

IMPULSE CONTROL DISORDERS. Impulse control disorders (ICDs), most commonly pathologic gambling, hypersexuality, uncontrolled spending, and binge eating, are observed in up to 14% of patients with PD. About one-third of affected patients have more than one ICD. Sporadic cases of hypersexuality were reported in the medical literature in the 1970s, after the introduction of levodopa. However, ICDs are most strongly associated with use of dopamine agonists for PD (e.g., pramipexole, ropinirole, and rotigotine) or for other indications, such as restless legs syndrome, underscoring the importance of patient (and family) education plus close monitoring for ICDs whenever dopamine agonists are prescribed. While not all individuals taking dopamine agonists develop ICDs, monitoring is especially prudent when there are other risk factors associated with development of ICDs, including younger age of PD onset, unmarried status, family history of problem gambling, cigarette smoking, caffeine use, and a novelty-seeking temperament.

In most cases, ICDs resolve after the offending medications are discontinued, but there can be devastating social consequences, including serious financial loss, bankruptcy, divorce, other legal problems, and loss of employment. Similar to other addictive disorders, ICDs may elude detection or, when referred for psychiatric care, the link to a potential side effect of PD medications is not appreciated. In part, this is because ICD behaviors can begin as early as a few months after starting dopamine agonists, but may also develop only after years of treatment, such as when the dose is increased, or when levodopa and dopamine agonists are used together. In the absence of a collateral informant, patients are unlikely to divulge their uncontrollable urges to engage in sexual activity, gamble, or spend excessively, and the behaviors may be unusual or idiosyncratic. For example, hypersexuality, seen in 2% to 6% of patients with PD, usually presents as an increase from baseline sexual activity, but paraphilias including transvestitism, pedophilia, and flagrant promiscuity are reported. Spending may involve purchases of large quantities of unnecessary items, for example, lamps, or excessive and reckless generosity. Such behaviors may be misinterpreted as a way of “coping” with PD.

Other significantly disruptive behaviors, often referred to as compulsive behaviors, are associated with antiparkinsonian treatment, namely levodopa. Punding refers to intense time-consuming preoccupation and fascination with seemingly meaningless rituals or stereotyped activities, for example, arranging and rearranging objects. Hobbyism involves intense time-consuming preoccupation with specific hobbies or activities, for example, endlessly working on projects or hobbies, or computer addiction. Walkabout refers to excessive aimless walking or driving. The dopamine dysregulation syndrome is a challenging condition involving an addiction-like overuse of dopaminergic medications that exceeds what is needed to control motor symptoms and persists despite harmful side effects. The dopamine agonist withdrawal syndrome (DAWS), also challenging to manage, includes addiction-like withdrawal symptoms (anxiety, panicky feelings, diaphoresis, irritability, dysphoria, drug craving, pain, and fatigue) when dopamine agonists are eliminated. Mania or euphoric syndromes are generally related to dopaminergic medications and are often associated with mood swings and on–off phenomena, but can also reflect a premorbid (pre-PD) mood disorder. Emotionalism refers to fleeting states of heightened and excessive sentimentality, often with pathologic tearfulness (also referred to as pseudobulbar affect) that can be precipitated by a variety of emotions; it can be a feature of a depressive disorder.

PREMORBID PERSONALITY. There is controversy over whether certain premorbid personality characteristics prevail in patients with PD. Observers have noted premorbid personality features that include industriousness, cautiousness, perfectionism, and punctuality in individuals who ultimately develop PD, with persistence of these traits over the course of the illness. One argument is that these traits represent a consequence of premorbid reductions in dopaminergic tone.

Pathology and Laboratory Examination. Unlike other common neurodegenerative disorders, there are no accepted neuropathologic criteria for PD. Most neuropathologists agree that dopaminergic neuronal loss in the SNpc is a requisite finding. However, the pathologic process precedes the clinical presentation since classic motor signs of PD are not evident until there is 70% to 80% loss of nigral neurons. In addition to dopaminergic cell loss, Lewy bodies, which are intraneuronal aggregates of the protein α-synuclein, are present in most forms of PD (Fig. 2.6–1B) and are seen in the substantia nigra, locus coeruleus, dorsal vagal nucleus, and nucleus basalis of Meynert. Lewy bodies are not specific to PD, and are seen in a variety of conditions, including AD. PD is also accompanied by degeneration of noradrenergic neurons in the locus coeruleus, serotonergic neurons in the dorsal raphe, and cholinergic neurons in the nucleus basalis and their attendant projection systems.

Alzheimer type neuropathologic changes (i.e., senile plaques and neurofibrillary tangles) are present in some patients with PD and dementia. Differential degrees of pathology among affected neuronal systems are thought to underlie the heterogeneous motor, cognitive, and psychiatric features of PD.

It is important to rule out other causes of parkinsonism. Laboratory tests should be conducted to rule out WD and hypothyroidism. Neuropsychological testing can help clarify the extent of cognitive deficits, evaluate cognitive status in potential candidates for neurosurgical treatments, or establish alternative diagnoses. Dementia is an exclusionary criterion for surgical therapies since it is associated with a poor motor outcome and limits compliance. A cognitive profile atypical for PD also predicts a poor outcome. Structural neuroimaging studies may help exclude other diagnoses, such as vascular disease, normal pressure hydrocephalus, and other Parkinsonian conditions, but are generally not needed to diagnose classic PD. Brain magnetic resonance imaging (MRI) is expected to be normal in PD. Functional imaging with dopamine transporter single photon emission computerized tomography (DaT SPECT) or PET can be used to assess presynaptic dopaminergic integrity as a marker of disease progression or diagnose PD in unclear cases, for example, presence of atypical signs and a poor response to levodopa, a prominent essential tremor (ET), or parkinsonism from dopamine antagonist medications. Decreased dopaminergic uptake in the putamen is characteristic of PD, but subclinical dopaminergic changes are also evident in some at-risk relatives of kindreds with familial PD.

Differential Diagnosis. Several clinical features help distinguish PD from other parkinsonian disorders. PD progresses more slowly, postural instability is a later finding, and has an excellent and sustained response to levodopa. Alternative diagnoses, such as PSP or MSA, are suggested by rapid progression with development of postural instability and early falls (within 3 years), a poor or transient response to levodopa, orthostatic hypotension and other signs of autonomic dysfunction, early dysphagia, and supranuclear gaze palsy. Early hallucinations, unrelated to treatment, and early dementia suggest a diagnosis of DLB or AD with associated PD or with parkinsonism. Myoclonic jerks, pyramidal signs (e.g., Babinski sign), and cerebellar signs (e.g., wide-based gait), are not features of PD, but are seen in other parkinsonian disorders. Other non-PD–associated findings include alien limb syndrome or ideomotor apraxia.

It is always important to consider alternative diagnoses to depression. Apathetic syndromes, dementia, or delirium can all result in changes in mood that are usually more apparent to families than the patient. The differential diagnosis of depression in PD also includes bipolar disorder, adjustment disorders, anxiety syndromes, psychosis, pathologic tearfulness (emotional incontinence), substance abuse or dependence, mood lability associated with parkinsonian motor fluctuations, and increased medical debility. Independent of a depressive disorder, sleep disturbances and daytime fatigue in PD can impair function and lead patients to feel discouraged. PD-related depression is also frequently associated with anxiety disorders (e.g., panic disorder) and dementia. The somatic and autonomic symptoms of anxiety can overlap with features of other medical conditions. Psychosis in the early stages of PD suggests alternative diagnoses, such as DLB or AD with extrapyramidal features. In some cases, pre-existing psychopathology emerges with dopaminergic treatment.

Course and Prognosis. PD is clinically heterogeneous. In each case, there is a different clinical presentation, age onset, natural history, and response to levodopa. Many cases of PD can be subdivided into either a tremor predominant form, or one characterized by prominent gait disorder and postural instability, the former having a more benign prognosis. Heterogeneity in clinical onset, progression, and levodopa response also influence the natural history of PD. This includes the onset of motor fluctuations and other treatment-related complications, longevity, and functional impairment. The rate and course of cognitive deterioration is also highly variable, ranging from mild to severe. Successful treatment of depression improves motor status but medical treatment of motor symptoms does not usually improve depressive syndromes. Even in the absence of depression, antidepressant medications and ECT have been shown to improve motor function. As in the general population, anxiety disorders often coexist with depressive episodes, but anxiety disorders tend to be chronic. The need for nursing home placement, often triggered by psychosis, is associated with a higher mortality rate.

Treatment

Motor Symptoms

PHARMACOLOGIC TREATMENTS. Current treatment options for PD provide only symptomatic relief of the motor deficits; there are no known neuroprotective or regenerative strategies, though this is an active area of research. Consequently, the gold standard of treatment for PD remains levodopa, which increases the life expectancy among patients with PD and significantly improves function and quality of life. During early stages, anticholinergic medications such as trihexyphenidyl or benztropine, amantadine, and selegiline provide mild-to-moderate reductions in motor symptoms. Anticholinergic medications are used primarily to reduce tremor; their adverse effects, for example, cognitive impairment, urinary retention, and lack of benefit on bradykinesia limit their utility in the treatment of PD. Dopamine agonists, such as pramipexole, ropinirole, and pergolide, and/or levodopa may then be added to address progressive disability. Dopamine agonists are especially effective early treatments that can delay initiation of levodopa therapy. Postural instability does not usually respond to antiparkinsonian medications.

There is debate over when to begin levodopa therapy since it is associated with dose-limiting motor side effects (in addition to psychosis). Initially, there is tremendous improvement in motor function. As PD progresses (along with the underlying neurodegeneration), higher levodopa doses are required to achieve similar benefit and the duration of the motor response wanes. A roller coaster–like experience of fluctuating motor effects then replaces the sustained motor benefits: patients experience states of improved motor function after taking dopaminergic medications (the “on” state) followed by reduced mobility or even immobility when the medication effects subside (the “off” state). Patients also develop abnormal involuntary hyperkinetic choreiform or dystonic movements during the “on” state, referred to as drug-induced dyskinesias. Agents that inhibit dopamine metabolism via catechol-o-methyltransferase inhibition, such as tolcapone and entacapone, can reduce motor fluctuations.

Psychiatric Disturbances Related to Fluctuating Effects of Dopaminergic Therapy. Emotional influences can trigger or intensify the “on–off” motor fluctuations and drug-induced dyskinesias, themselves, can be distressing or embarrassing. In the extreme, “on–off” fluctuations with dyskinesias range from states of complete immobility or “freezing” to severe writhing. Fluctuating motor states can also be accompanied by nonmotor phenomena such as akathisia, anxiety, dysphoria, autonomic symptoms, irritability, apathy, hallucinosis, psychosis, screaming, and cognitive slowing during “off” periods and hypersexuality, hypomania, and thought racing during “on” periods.

SURGICAL TREATMENTS. Before availability of levodopa in the 1960s, neurosurgical procedures were standard treatments for tremor and rigidity in PD. More refined stereotactic neurosurgical therapies currently used for PD involve high-frequency deep brain stimulation (DBS) of the STN and globus pallidus pars interna (Gpi) or ablative procedures that target the GPi (pallidotomy) or thalamus (thalamotomy). Both improve motor function because they reduce the hyperactivity of the STN and its efferent targets, thus partially restoring thalamic outflow. Thalamotomy and thalamic DBS are indicated for disabling tremor. STN DBS and pallidal interventions also improve tremor, but are mainly indicated for severe dyskinesias, bradykinesia/akinesia, and rigidity. The procedures are aimed at alleviation of contralateral symptoms, but bilateral benefits are seen for dyskinesias and bradykinesia. In 2018, U.S. Food and Drug Administration (FDA) approved unilateral high-intensity focused ultrasound (FUS) of the STN or thalamus for treatment of tremor-dominant PD. The indications were subsequently expanded to include bradykinesia, rigidity, and dyskinesias. Embryonic cell transplantation of dopamine neurons, first performed in the 1980s, remains experimental, as are trials on surgical delivery of gene therapies.

Neuropsychiatric Effects of Neurosurgical Therapy for PD. Initial studies showed limited, if any, adverse cognitive effects after ablative or DBS procedures when surgical candidates underwent careful screening for dementia. Some pallidotomy studies found changes in executive function, memory, verbal fluency, and attentional set shifting, but lesion size may have been influential. Stimulation parameters, laterality and site-specific effects, electrode placement, and stimulation sites may also affect cognitive outcomes.

While many patients report improved quality of life after surgery, acute stimulation with DBS can be associated with profound depressive or manic symptoms, including in patients with no prior psychiatric history. In some patients, pre-existing disorders, such as depression, anxiety, or drug dependence, and interpersonal conflicts worsen and can overshadow motoric benefits. DBS treatment-emergent sexually deviant behaviors, gambling, mood lability, apathy, and aggression are also described. Reductions in anxiety and depression, unrelated to improved motor status, are reported after unilateral pallidotomy, whereas depression and apathy are reported after bilateral pallidotomy. Pallidotomy studies show no consistent effects on depression or anxiety, but a history of a prior or active affective disturbance at the time of surgery may be associated with development of a mood disorder.

COGNITIVE MANIFESTATIONS. Several large randomized controlled trials on treatments of cognitive impairment or dementia in PD or DLB have been conducted. Presence of cholinergic deficits in PD and their relationship to dementia suggests a role for cholinesterase inhibitors. Intervention trials for PD-related dementia or DLBs show that cognition and psychiatric symptoms, including psychosis, improved with donepezil and rivastigmine, though the data are more limited for galantamine. Cholinesterase inhibitors are not always well tolerated; the most common side effects are nausea, vomiting, anorexia, dizziness, drowsiness, and insomnia and may be more common with rivastigmine. Memantine, while well tolerated, has not demonstrated consistent benefit, though perhaps more so in DLB. A recent systematic review indicates that both cholinesterase inhibitors and memantine result in slight improvement in global impression, but cognitive performance improved only with cholinesterase inhibitors.

A few studies investigated treatment of PD-MCI, though inconsistent definitions of PD-MCI confound interpretation of the results. Cholinesterase inhibitors appear to have some potential benefits, but targeted treatment of noradrenergic deficits may also be beneficial. Cognitive rehabilitation programs are increasingly recognized as beneficial to cognitive performance and functional abilities in patients with PD, including as adjunctive therapy to medications. In addition, patients and families benefit from education about cognitive deficits in PD, especially their consequences on executive functioning and strategies for minimizing distress related to selective cognitive impairments, and, later in the course, dementia.

PSYCHIATRIC MANIFESTATIONS. A first step in the treatment of psychiatric disorders is to evaluate the contribution of the antiparkinsonian medications, fluctuating medication effects, other medical conditions, cognitive deficits, and caregiver and other psychosocial variables to the presenting problem. The common occurrence of more than one psychiatric disturbance also influences treatment. Sleep is commonly disrupted in PD, independent of psychiatric problems, and addressing it alone can reduce daytime mood lability, psychosis, and motor impairments. Initial treatment involves consideration of the antiparkinsonian regimen, sleep hygiene practices, and use of a low-dose sedating antidepressant or hypnotic such as trazodone (e.g., 25 to 100 mg). Benzodiazepines can be used judiciously but are not considered first-line therapy for anxiety, agitation, or sleep disturbances because of their effects on cognition, falls, and interdose withdrawal symptoms. Patient and caregiver education, especially about the psychiatric condition and its overlap with parkinsonian features, and supportive therapy are cornerstones of effective treatment. Parkinson support groups can be invaluable sources of education and community support. Speech, physical, and occupational therapies can help maximize function, which facilitates socialization and quality of life. Even early in the disease, and then periodically as the disease progresses, rehabilitative therapies provide patients and caregivers opportunities to develop new strategies and skills that enhance communication, gait, balance, and daily self-care and function as new impairments emerge. Regular exercise helps motor symptoms and overall well-being.

DEPRESSION. For many years, despite the prevalence and seriousness of PD-related depression, published controlled trials on antidepressant treatment in PD were of poor quality. More recent research on various modalities to treat depression in PD provides some guidance, though there remains a need for larger, well-designed trials including behavioral and pharmacologic treatments. The most recent systematic review and meta-analysis indicates that pharmacologic treatments, especially using serotonin reuptake inhibitors (citalopram, sertraline, fluoxetine, and paroxetine), and CBT, result in significant improvements in depressive symptoms in both major and nonmajor depressive syndromes. Longer-term benefits of these treatments have not been assessed. However, a randomized controlled trial of CBT for PD depression demonstrated efficacy of individual CBT as monotherapy or in combination with antidepressant medication, including for reducing residual symptoms in patients already treated with antidepressants. In-person and remote telephone- or video-based modalities for delivering CBT are all shown to be efficacious. Tricyclic antidepressants (nortriptyline, desipramine, and doxepin) showed a large pooled effect that was not statistically significant. Trials of nonantidepressant medications (pramipexole, pergolide, atomoxetine, and memantine) were not efficacious. rTMS studies showed improvement in depressive symptoms, but were not sufficiently powered to yield firm conclusions.

As with primary depression, antidepressant selection is usually based on consideration of side-effect profiles. Serotonin reuptake inhibitors can aggravate motor symptoms, but this is not especially common. Use of selegiline with antidepressants carries the risk of inducing a serotonin syndrome or hypertensive reactions, but most patients tolerate this combination. In clinical practice, ECT is safe and effective for depression in PD, especially in the context of psychotic symptoms.

ANXIETY. In spite of the high prevalence of anxiety disturbances in PD and their impact, there is limited treatment data for this population. The most extensive data shows improvement of anxiety symptoms with CBT, though impact on specific anxiety disorders was not assessed. In the one randomized controlled medication trial, bromazepam, a long-acting benzodiazepine, improved psychological and somatic anxiety symptoms. While benzodiazepines can be effective in reducing anxiety, their unfavorable effects on alertness, cognition, and gait, and fall risk should discourage their use, with careful evaluation of known risks and putative benefits before prescribing even for short-term use. SSRI antidepressants have been shown to improve anxiety when measured as a secondary outcome in depression treatment trials in patients with PD. Buspirone, a nonbenzodiazepine anxiolytic with partial 5-HT agonism, has been used in PD but its specific effects on anxiety are not obvious. Gabapentin is used for treatment of anxiety disorders, but its use in PD has not been examined. The utility of antipsychotic medications as adjunctive therapy is also not known.

APATHY. Studies on treatment of apathy in PD are limited, in part because of practical and conceptual barriers, such as lack of established diagnostic criteria or a gold-standard assessment scale, and confounding conditions. Treatment of comorbid depression is an obvious initial approach, but there may be residual or SSRI-induced apathy. Antidepressants that inhibit dopamine or norepinephrine reuptake are suggested by neurochemical theories of apathy. If depression is not an issue, dopaminergic agents, including levodopa, dopamine receptor agonists, selegiline, amphetamines, and amantadine, have been most consistently effective; the absence of cognitive dysfunction predicts a better response. Among PD patients without dementia or comorbid depression, apathy improved in double-blind placebo-controlled trials of medications that either enhanced acetylcholine (rivastigmine) or stimulated D2 or D3 receptors (piribedil). In Alzheimer and PD, cholinesterase inhibitors may reduce apathy. Subthalamic DBS can induce or aggravate apathy in some patients

PSYCHOSIS. Treatment of psychosis in PD is challenged by the delicate balancing act between the use of medications that can relieve psychosis without aggravating motor or cognitive dysfunction and medications that improve motor symptoms without exacerbating the mental status. Caregiver safety is also an issue because of immobility, directed aggression, and combativeness during care. Reducing or simplifying the antiparkinsonian regimen usually involves elimination of anticholinergics, amantadine, and selegiline; in addition, when there is psychosis, levodopa is generally preferred to dopamine agonists or controlled-release levodopa. When higher doses of levodopa are required or disruptive symptoms persist, treatment usually requires addition of an atypical antipsychotic.

There is a solid evidence base for treatment of psychosis in PD. The gold standard is clozapine; it is effective even at low doses (<25 mg), does not block D2 receptors, and has been evaluated in several controlled trials. Because of the requisite blood monitoring with clozapine, quetiapine is usually the initial drug of choice, but some patients eventually experience worsened motor function and patients with dementia may be especially vulnerable. Controlled trials of quetiapine, however, have not been shown to be efficacious. Orthostatic hypotension, headache, and nausea also contribute to its intolerance. Other “atypical” neuroleptic agents (olanzapine and risperidone) reduce or eliminate psychosis at low doses, but their effects on striatal D2 receptors are frequently associated with significant motor deterioration. Case reports and open-label studies demonstrate that aripiprazole, a partial agonist at serotonergic and dopaminergic receptors, has variable effects on psychotic symptoms but its propensity to worsen motor function indicates it is not suitable for use in PD. Controlled trials of ziprasidone are needed to evaluate its efficacy and safety, though case reports and case series report psychiatric benefits without significant adverse effects. Pimavanserin, a selective serotonin 5-HT2A inverse agonist without adverse motor effects, is FDA approved for treating PD psychosis, including both hallucinations and delusions. Pimavanserin may also improve cognition and sleep, even before the antipsychotic effect, which may not be evident for 3 to 4 weeks. ECT is effective for major depression with psychosis and nonpsychotic depression; it also benefits parkinsonism and can be used for dopamine-induced psychosis. Cholinergic agents can also have antipsychotic effects.

Other Parkinsonian Conditions

Progressive Supranuclear Palsy

EPIDEMIOLOGY AND ETIOLOGY. PSP, an important cause of parkinsonism, was first described in 1964 and accounts for 4% of all cases with parkinsonism. Prevalence of the classic PSP–Richardson syndrome (PSP-RS) is five to six cases per 100,000 with onset of symptoms usually after age 60 in all races and both genders, but men tend to be affected more often than women. Disease duration is approximately 6 years. Recognition that there is spectrum of clinical phenotypes corresponding to PSP pathology. There are rare familial cases of PSP, but its development is usually sporadic. PSP is classified as a tauopathy, a term referring to neurodegenerative disorders characterized by intracellular aggregations of abnormally phosphorylated tau protein. Tau normally stabilizes microtubules and its aggregation leads to formation of abnormal cytoskeletal fibrils. The cause of the tauopathy in PSP, and its role in neurodegeneration, remains uncertain.

CLINICAL FEATURES. The defining features of PSP-RS are of a gradually progressive neurodegenerative condition that is manifest in affected individuals over age 40 years as ocular motor dysfunction (vertical supranuclear gaze palsy) or postural instability and falls within the initial year of onset. More recent criteria by the Movement Disorder Society also include akinesia (progressive gait freezing) and cognitive dysfunction as core features. In contrast to PD, the parkinsonism progresses with symmetric bradykinesia and axial rigidity along with paralysis of downward vertical gaze. The problems with ocular motility may not be evident for several years, but cause problems with walking downstairs. Dysphagia develops over the course of the illness. Unlike PD, postural instability leading to falls and dysarthria are early and often presenting features. Also, there is a poor response to levodopa, tremor is uncommon and stride length is erratic. Posture tends to be upright or sometimes opisthotonic because of axial rigidity. Instead of a lack of facial expression (hypomimia), patients appear “worried” or “astonished,” resulting from dystonic contractions of facial muscles between the eyebrows, reduced blinking, lid retraction, and gaze palsy.

The clinicopathologic complexity of PSP is notable, leading to recognition of its phenotypic variability, in which initial phenomena are atypical for PSP. With progression, the typical motor features, that is, postural instability and vertical gaze palsy, emerge and the ultimate diagnosis of “definite” PSP is based on evidence of PSP neuropathology. However, the varied initial presentations confound early and accurate diagnosis, underscoring the need for specific biomarkers that facilitate diagnosis. Examples of phenotypic variability include PSP presenting with apraxia of speech (PSP-AOS), as Parkinson disease like (PSP-P) with a jerky postural tremor and levodopa response, as the corticobasal syndrome (PSB-CBS), with cerebellar ataxia (PSP-C), with exclusive ocular motor dysfunction (PSP-OM), with primary lateral sclerosis (PSP-PLS); with progressive akinesia and gait freezing (PSP-PAGF), and with the behavioral variant of frontotemporal dementia (PSP-F).

Cognitively, there is also phenotypic variability, with features of a subcortical dementia with bradyphrenia, memory deficits, and predominant frontal lobe dysfunction with executive, attentional deficits, memory, visuospatial, and language and social cognitive deficits. Approximately 10% present with cognitive symptoms and up to 70% develop a dementia over the disease course. Speech and language dysfunction presenting as a nonfluent/agrammatic variant of primary progressive aphasia (naPPA) or progressive apraxia of speech increases diagnostic certainty for PSP. Aphasia, apraxia, and agnosia, the classic features of “cortical dementias,” are not evident. Apathy, irritability, childishness, impulsivity, perseveration, and disinhibition are compatible with damage to orbitofrontal and medial frontal circuits and are the most common psychiatric or behavioral features. There are also sleep disturbances, depression, emotional lability including abrupt laughter or crying episodes that are usually inappropriate or mood incongruent (also referred to as pseudobulbar affect or emotional incontinence), and occasionally schizophreniform psychotic symptoms.

PATHOLOGY AND LABORATORY EXAMINATION. Neuropathologically, there is broader involvement of the basal ganglia in PSP than in PD. PSP affects the substantia nigra as well as the striatum, thalamus, subthalamus, other brainstem nuclei, and periaqueductal gray matter. The main pathologic findings are abnormal accumulations of tau in what are described as "tufted" astrocytes, globose high-density tau-positive neurofibrillary tangles composed solely of the 4R tau isoform, coiled bodies in oligodendrocytes in white matter, neuronal loss, and gliosis. There is pronounced dopamine depletion in the striatum, decreased density of striatal D2 receptors, widespread cholinergic neuronal damage, and effects on serotonergic and GABAergic systems. Structural imaging studies initially reveal midbrain atrophy followed by third ventricular enlargement and pontine and frontotemporal atrophy. Diffusion tensor imaging studies demonstrate white matter tract degeneration connecting the cerebellum, midbrain, thalamus, and premotor cortex. PET studies show decreased striatal dopamine binding and fluorodopa uptake and bilateral premotor hypometabolism extending into prefrontal cortex, supplementary motor area, and midbrain. A recent GWAS of autopsy-confirmed PSP implicates three genes, STC6, EIF2AK3, and MOBP, as associated with PSP. However, PSP is rarely associated with a family history and is considered sporadic at this point.

DIFFERENTIAL DIAGNOSIS. After PD, which is at least 10 times more prevalent, PSP is the second most common parkinsonian syndrome. Frontotemporal dementia (FTD) and corticobasal ganglionic degeneration manifest similar clinical and pathologic features, including parkinsonism, frontal executive impairments, apathy, and tau-positive inclusions. While profound apathy is the predominant psychiatric feature, it can be misdiagnosed as depression, which occurs with apathy in a minority of cases.

COURSE AND PROGNOSIS. PSP is a rapidly progressive and devastating degenerative disease. Life expectancy is generally 5 to 7 years after onset, with death often related to dysphagia.

TREATMENT. There are no specific treatments for the motor or cognitive aspects of PSP and care is largely supportive and symptomatic. Occupational, physical, and speech therapy interventions address safety and function as the disease progresses. Clinical trials targeting tau represent significant progress and antibody and gene therapies are under development. Both tideglusib, a glycogen synthase kinase 3 inhibitor, and davunetide, which affects microtubule stability, were safe and well tolerated, but did not demonstrate clinical efficacy on the PSP rating scale. Levodopa is ineffective in PSP, probably because of the reduced number of postsynaptic striatal D2 receptors, although there can be an initial response in some patients. Despite extensive cholinergic damage, cholinergic enhancing agents provide little or no benefit for cognition and can be poorly tolerated. Psychiatric symptoms in PSP can be difficult to treat, but include standard targeted pharmacologic and psychosocial approaches aimed at reducing apathy, irritability, agitation, mood lability, etc., in neuropsychiatric populations.

Corticobasal Degeneration

Originally termed corticodentatonigral degeneration with neuronal achromasia, corticobasal degeneration (CBD) was first described in 1968 as a syndrome of involuntary movements and slowed voluntary movements. CBD is classified as a sporadic tauopathy characterized by deposition of hyperphosphorylated 4-repeat tau in neurons and glia, referred to as astrocytic plaques. CBD also shares clinical and pathologic features with other tauopathies, including PSP and certain forms of FTD. Although numerous clinical criteria have been applied to CBD, these predict pathologic CBD in 25% to 56% of cases and diagnostic criteria are evolving. Review of pathologically confirmed cases of CBD suggest four CBD phenotypes: the initially reported phenotype, now referred to as CBS, a frontal behavioral–spatial syndrome (FBS), naPPA, and progressive supranuclear palsy syndrome (PSPS). To facilitate development of potential neuroprotective therapies for tauopathies, two sets of clinical diagnostic criteria comprising these phenotypes for CBD have been proposed, one for probable CBD and a second broader set of criteria for possible CBD with less specificity for CBD pathology, but still likely representing tau pathology.

CBD affects adults with gradual onset usually after age 40 and steady progression until death within 2 to 12 years. Mean age of disease onset is reported as 64 years and the median survival is around 6 to 7 years. The pathologic findings include ballooned achromatic neurons, frontoparietal atrophy, pronounced cortical neuronal loss, cortical astrocytic plaques, and tau inclusions in the substantia nigra, basal ganglia, and along dentatorubrothalamic tracts. Neuroimaging studies show asymmetric frontoparietal atrophy and subcortical hypometabolism contralateral to the more affected side along with bifrontal–parietal hypometabolism.

The clinical manifestations reflect both basal ganglia and cortical dysfunction. Motor signs almost invariably include highly asymmetric or unilateral progressive motor signs of levodopa-resistant parkinsonism (rigidity, brady/akinesia, and postural or action tremor). There is also postural instability, focal reflex myoclonus, and limb dystonia, which is present in about 40% of patients. Oculomotor motility may be impaired, similar to PSP. The most prominent neuropsychiatric features and clinical hallmarks of CBS are lateralized ideomotor and ideational apraxia and the “alien hand syndrome.” Cognitive deficits, originally regarded as late-stage phenomena, are evident in most patients early in the disease course, along with behavioral features. These may be overlooked, however, when there is prominent motor dysfunction. As with other subcortical motor disorders, there are multiple cognitive deficits, including predominantly executive, memory, and attentional dysfunction that may be severe enough to amount to dementia. Parietal lobe dysfunction, manifest as sensory or visual neglect and visuospatial deficits, is apparent on cognitive testing, as are deficits in social cognition. Aphasia may also be present.

Depression is much more common in CBD relative to PSP, which is characterized by apathy. In CBD, apathy, disinhibition, and delusions are also common and may co-occur, but appear distinct from depression, suggesting different underlying mechanisms. There are no definitive treatments for the motor or cognitive symptoms in CBD, but the parkinsonism may have a limited response to levodopa. Antidepressants are used for depression and anxiety but are not always effective. Caregiver education and assessment for depression are important aspects of patient care. Physical therapy interventions address daily functioning, improve rigidity, and prevent contractures.

PRIMARY DEMENTIA SYNDROMES WITH PARKINSONIAN SIGNS

Frontotemporal Dementia

Social cognition is one of the six cognitive domains assessed in the diagnosis of major neurocognitive disorder in the DSM-5. FTD is a clinical syndrome characterized by prominent behavioral or personality changes involving social cognition, language dysfunction, and neuronal loss, spongiosis, and gliosis in the frontal and temporal cortices. Onset of FTD is usually after age 50, but ranges from 35 to 75 years. Life expectancy after diagnosis is 3 to 15 years.

FTD is neuropathologically, clinically, and genetically heterogenous, with approximately 30% to 50% of cases having a positive family history. FTD also overlaps clinically and neuropathologically with CBD, PSP, and motor neuron disease/amyotrophic lateral sclerosis (MND/ALS), which are often considered part of the FTD complex. The characterization and genetics of the FTDs have evolved rapidly over the last decade. Neuropathologic consensus criteria include three major subgroups based on protein pathology: tau, transactive response DNA-binding protein 43 (TDP-43), and fused in sarcoma (FUS) and there are three major FTD disease genes (microtubule-associated protein tau [MAPT], progranulin [GRN], and C9*orf*72) and many other rarer disease genes. Mutations in MAPT, located on chromosome 17, are typically associated with bvFTD and parkinsonism, and are generally referred to as FTD and parkinsonism linked to chromosome 17 (FTDP-17).

Symptom onset in FTD is insidious and many patients with FTD are initially misdiagnosed with a primary psychiatric illness. The behavioral variant (bvFTD) involves prominent and altered social behavior and personality changes such as profound apathy, impulsivity, disinhibition (e.g., shoplifting or grabbing food off another person's plate), loss of personal awareness (e.g., neglect of personal grooming), stereotyped and perseverative behaviors, delusions, depression, affective lability, overeating and other signs of hyperorality, social inappropriateness, and lack of tact or empathy. The most striking cognitive change is anosognosia and indifference toward their cognitive and social deficits. Otherwise, cognitive testing reveals profound impairment on tests of executive function, but relative preservation of other cognitive domains, including memory. Two other variants, progressive nonfluent aphasia (PNFA) and semantic dementia (SD) present with gradual language deterioration and can be difficult to distinguish from one another early in their course. However, the insidious onset of language deficits in PNFA and SD distinguishes them from an acute aphasia caused by a cerebrovascular accident. In PNFA, there are deficits in expressive language and articulation, with effortful nonfluent spontaneous speech, agrammatism, phonemic paraphasias, and marked word-finding deficits. Language comprehension, memory, and other cognitive domains are relatively preserved. A third variant, SD, involves profound language dysfunction and a fluent aphasia notable for its spontaneous, circumlocutory, empty speech resulting from loss of semantic memory in both and the ability to link words and objects to their meanings. Both verbal and nonverbal domains are affected and there is pronounced anomia. With disease progression, behavioral and personality changes akin to bvFTD may develop.

Motor deficits may include symptoms of MND or parkinsonism of the akinetic-rigid type that is common in bvFTD or PNFA and rarely seen in SD. Structural imaging studies may show frontal and anterior temporal atrophy; functional imaging typically reveals hypoperfusion in the same regions, though there may be pronounced asymmetry. More widespread involvement of parietal regions is suggestive of AD. The electroencephalogram may be normal in early stages. The differential diagnosis in younger persons includes schizophrenia, mood disorders, addictions, and personality disorders. When onset is later, the diagnosis is often confused with AD, late-life depression, and, when there are prominent motor features, CBD and PSP. These conditions resemble one another in their advanced stages, but are distinguished by the course of the disease prior to the advanced stage. There are no treatments for the underlying disease process, and no definitive evidence supporting the use of cholinergic treatments. The behavioral symptoms may improve with serotonin reuptake inhibitors. Aggressive behaviors can be managed with low doses of atypical antipsychotics or trazodone and antiseizure medications can reduce agitation and aggression. Family counseling is a critical part of treatment so that the safety of the patient and others is maintained in the setting of their poor judgment and to provide caregiver support and assist with psychosocial issues.

Dementia with Lewy Bodies

DLB is a dementia syndrome with Lewy bodies in the cerebral cortex, brainstem, thalamus, and striatum. It is now regarded as the second most common form of dementia after AD and second most common synucleinopathy after PD. Symptom onset is usually after age 50 and the disease is more common in men. Average survival is less than 10 years. There are both genetic and sporadic forms of DLB, and there is overlap in the clinical and pathologic findings of DLB, AD, and PD and other α-synucleinopathies. About half of the patients with DLB also have Alzheimer pathology. Up to one-third of cases show microvascular ischemic pathology. Neurochemically, there is evidence of a marked cholinergic deficit.

While the defining feature of DLB is progressive cognitive impairment, its association with basal ganglia signs and hallucinations leads to its inclusion in the differential diagnosis of parkinsonian movement disorders with psychiatric presentations. For the diagnosis of probable DLB, the most recently updated consensus criteria require (1) progressively disabling cognitive deficits in which the domains of attention, executive function, and visual processing are disproportionately affected relative to memory and naming and (2) at least two of the four core clinical features: delirium-like cognitive fluctuations, recurrent complex visual hallucinations, parkinsonism, and REM sleep behavior disorder. Early cognitive deficits often involve attentional, visuospatial, or frontal–subcortical skills, and memory impairment may only become apparent with progression. With progression, there are the usual cortical features of dementia (aphasia, apraxia, and agnosia). The psychiatric symptoms present early in the course and often include dramatic well-formed visual hallucinations, depression, or paranoid delusions. Motor features include myoclonus and parkinsonian signs of bradykinesia, rigidity, and gait difficulties with early falls in about two-thirds of patients. Resting tremor is uncommon. There may be an early but limited response to levodopa. As a general rule, the occurrence of psychosis unrelated to medications and early dementia with spontaneous parkinsonism, especially within the first year, should lead to consideration of DLB.

The diagnosis of DLB remains clinically based and neuroimaging studies are not especially helpful in distinguishing DLB from AD. Treatment of DLB requires a targeted approach. Motor symptoms can be treated with antiparkinsonian therapies, though this may aggravate psychiatric symptoms, especially psychosis. Atypical neuroleptics can reduce psychosis, but are often not tolerated and cases of neuroleptic malignant syndrome are reported. The most promising treatments include cholinergic therapies, which can benefit cognitive and psychiatric symptoms. Antidepressants and stimulants can also be helpful.

Multiple System Atrophy

The term MSA was introduced in 1969 and operationally defined in 1999 as a clinically heterogeneous neurodegenerative disorder with variably present parkinsonism, early autonomic dysfunction, cerebellar, and pyramidal symptoms. These are manifested as rapidly progressive parkinsonism, gait ataxia with cerebellar dysarthria, limb ataxia, cerebellar oculomotor dysfunction, Babinski sign with hyperreflexia, postural instability, and dysphagia. At least one sign of autonomic dysfunction should be evident. MSA is now used in place of three older diagnoses, olivopontocerebellar atrophy (OPCA), Shy–Drager syndrome, and striatonigral degeneration, which could not be reliably distinguished from each other. The two MSA subtypes are autonomic dysfunction with predominant cerebellar findings (MSA-C) and autonomic dysfunction with predominant parkinsonian findings (MSA-P). The disease is rare and affects both sexes equally. It typically begins in late middle age, with a median survival of about 9 to 10 years after symptom onset. The cause is unknown, but almost all cases are sporadic. The neuropathology consists of glial cytoplasmic inclusions (containing tau, synuclein, and ubiquitin, among other proteins) in the presence of neurodegeneration in a variable combination of CNS structures, potentially including the putamen, caudate, globus pallidus, substantia nigra, inferior olives, brain stem nuclei, cerebellum, and spinal cord. MRI may reveal typical signs (hyperintense portions of the putamen, “hot cross bun” sign in the basis pontis), but diagnosis is essentially made clinically or neuropathologically.

A prodromal premotor phase is evident in up to 75% of cases and, consistent with multisystem pathology, includes sexual and urinary dysfunction, orthostatic hypotension, respiratory disturbances, and RBD. Initial parkinsonism in MSA is often difficult to distinguish from PD.

Psychiatric and cognitive manifestations of MSA have received little systematic attention. Dementia and visual hallucinations are considered nonsupportive diagnostic criteria and suggestive of DLB. However, more recent data from patients surviving longer than 8 years show evidence for significant cognitive impairment in up to 50% of cases, with a profile of cognitive deficits similar to other parkinsonian disorders and dementia in up to one-third of cases. Executive dysfunctions are most commonly seen, more so in MSA-P, whereas learning and visuospatial impairments are more prominent in MSA-C. Pathologic laughter and crying, depression, anxiety, panic attacks, and suicidal ideation are also reported. Treatment is symptomatic. A transient response to levodopa is seen in up to 40% of patients in the early disease stage. Orthostatic hypotension and urinary symptoms may respond to both pharmacologic agents and environmental manipulations, and such treatment may substantially improve patient and caregiver quality of life. There are no clear guidelines for treating psychiatric disorders in MSA; it is likely that strategies useful for PD, HD, or cerebellar disease will also prove of value in MSA. A small case series demonstrated the usefulness and safety of ECT in patients with MSA.

Encephalitis Lethargica

Encephalitis lethargica, described by Constantin von Economo, presented as a worldwide epidemic between 1917 and 1929. While putative sporadic cases continue to be reported, encephalitis lethargica is important historically as an early example of the relationship between brain disease and psychopathology. The acute phase was characterized by profound lethargy, hence its name, along with other mental status changes such as psychosis or catatonia and various movement abnormalities, including parkinsonism, choreoathetosis, and myoclonus. The mortality rate was approximately 30% during the acute phase. Survivors returned to normal for several months to years, then many developed a chronic postencephalitic syndrome characterized most commonly by parkinsonism (usually without tremor), and less commonly by hyperkinetic movements resembling tardive dyskinesia and episodic oculogyric and respiratory crises. Psychiatric manifestations included disinhibition, mania, attentional deficits, apathy, narcolepsy, verbal tic-like behaviors, executive dysfunction, and bradyphrenia. Hyperactivity and oppositional behaviors developed in children. Postmortem studies of acute victims showed inflammatory changes in the substantia nigra, globus pallidus, and subthalamic, red, and brainstem nuclei. Postencephalitic cases showed degenerative changes and depigmentation of the same subcortical areas, with neurofibrillary tangles but without Lewy bodies. The etiology remains unknown.

Huntington Disease

Introduction and Definition. HD (formerly Huntington chorea) is an autosomal dominant neurodegenerative disorder characterized by midlife onset, a relentlessly progressive course, and a combination of motor, psychiatric, and cognitive symptoms. The disease is caused by a CAG repeat expansion mutation in the huntingtin (HTT) gene on chromosome 4. For 130 years, HD has been recognized as one of the most terrible yet enlightening of all medical disorders, and now, as in the past, the approach to HD provides a model for the development of clinical, scientific, and ethical strategies in other genetic diseases of the brain. HD remains an important example of subcortical dementia and the prominence of the psychiatric consequences of neurodegenerative disorders.

History. George Huntington first described the disease that now bears his name in 1872. His original paper, which noted the salient features that still define HD, remains a model of clarity and clinical observation. After serving as the textbook case of dominant inheritance for 100 years, HD again made news in 1983 as the first disorder mapped to a specific chromosomal locus using anonymous DNA markers, and then again in 1993 when the genetic etiology was discovered to be a member of a new class of mutations referred to as repeat expansion or dynamic mutations. Current research focuses on understanding HD pathogenesis and the use of this knowledge to develop rationale therapeutics for the underlying disease process.

Epidemiology. The prevalence of HD in the United States is about five cases per 100,000 individuals, with similar rates found in most European countries. The prevalence in Japan is about 10-fold less, and prevalence also appears to be low in China, Finland, and Africa. As with many rare autosomal disorders, pockets of high prevalence exist. The most famous concentration is in the villages surrounding Lake Maracaibo in Venezuela;

analysis of this population led to the discovery of the HD mutation and has contributed greatly to the clinical understanding of HD.

Etiology. HD is one of a group of 13 neurodegenerative diseases caused by a mutation in which repeating units of DNA consisting of the nucleotides cytosine, adenine, and guanine (a so-called CAG repeat; cytosine, thymine, guanine—CTG—on the opposite strand) increase in number. This is referred to as a CAG, or CAG/CTG, trinucleotide repeat expansion mutation. In most of these disorders, including HD, the CAG repeat is in-frame to encode for the amino acid glutamine. Genes containing this type of expansion, therefore, express an RNA transcript with an expanded CUG repeat, which in turn is translated into a protein with an excessively long stretch of glutamine residues. The CAG repeat expansion occurs in different and unrelated genes in each of the CAG repeat diseases. The only similarity among the genes is the repeat itself. This and other lines of evidence have led most investigators to conclude that proteins with stretches of more than about 35 to 40 consecutive glutamine residues become neurotoxins, though there is increasing recognition that RNA transcripts with long CUG or CTG repeats may also have toxic properties. The differences among the CAG repeat expansion diseases, particularly concerning which brain regions are most prominently affected and the precise threshold for toxicity, are thought to derive from other aspects of protein expression and structure.

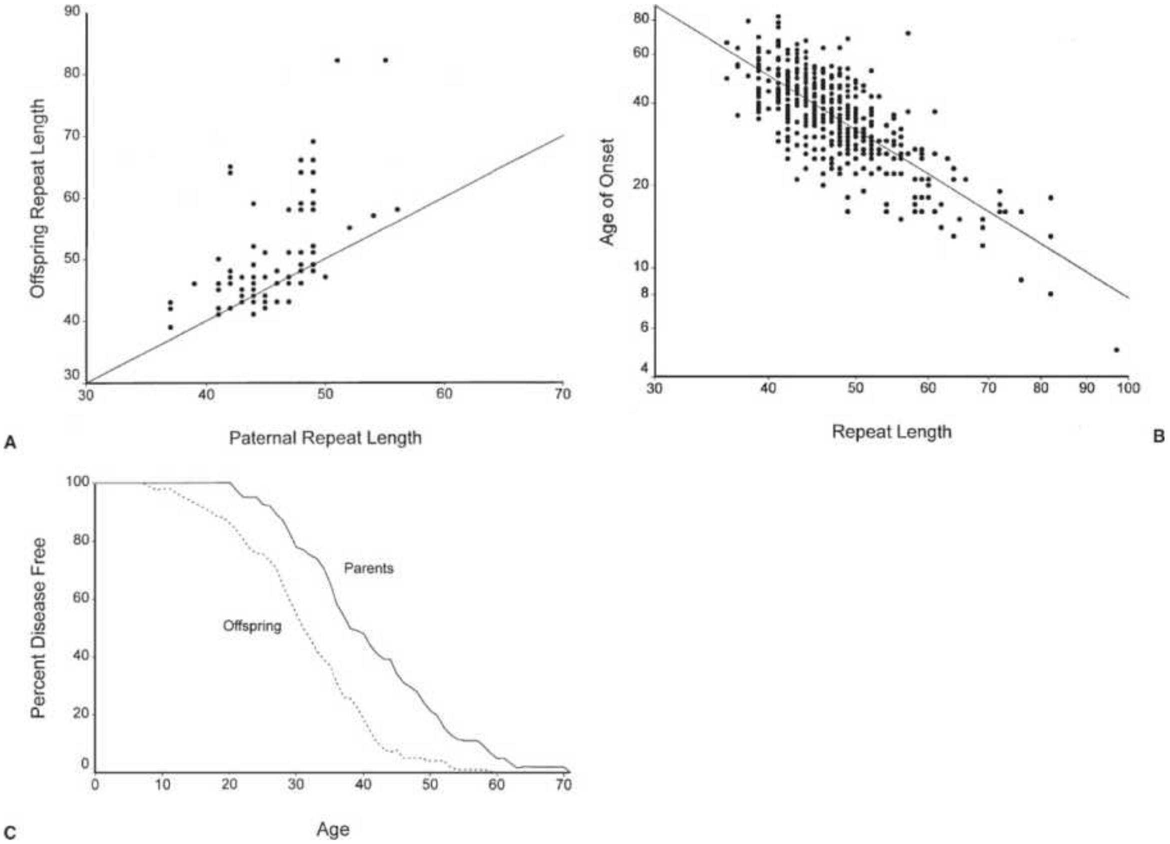

FIGURE 2.6–3. Repeat length as an important variable in Huntington disease. **A:** Repeat length increases in paternal transmission of the expanded allele. Points above the diagonal line represent cases in which the repeat length increases during paternal transmission. Expansion occurs in most cases, and in some instances the expansion is substantial. N = 84 pairs, mean +/− SD increase of repeat length = 4.2 +/− 0.8 triplets. **B:** Correlation of repeat length with a younger age of onset. As repeat length increases, age at disease onset decreases. N = 480, r^2 = 0.57. **C:** Anticipation in HD. The age at which affected parents and their affected offspring first develop disease symptoms is depicted as a survival curve. The younger generation develops the disease at a substantially younger age, a result of increasing repeat length during paternal transmission and the correlation of repeat length with younger age of onset. N = 61 parents, 82 offspring. (All data from the Baltimore Huntington Disease Center, courtesy of Drs Adam Rosenblatt and Christopher A. Ross.)

The CAG repeat expansion in HD is located in exon 1 of the *HTT* gene (formerly referred to as *IT15*) on the short arm of chromosome 4. Like the genes associated with other polyglutamine disorders, HTT is widely expressed, and is not limited to the CNS. The normal function of HTT is only partially understood, though it has been associated with embryogenesis, intracellular organelle transport, and regulation of growth factors. Like many of the other repeat expansion diseases, the phenomenon of anticipation (younger age of onset in subsequent generations of a family) occurs in HD. The explanation derives from a strong correlation between repeat length and younger age of onset and a tendency for repeat length to increase during paternal transmission of the gene (Fig. 2.6–3A, B). There is also increasing recognition that understanding disease pathogenesis requires understanding of the molecular complexity of CAG/CTG repeat regions of the human genome, including the HD locus, in which an RNA transcript is encoded by the DNA strand antisense to HTT. The antisense transcript may regulate HTT expression, and could also independently contribute to disease pathogenesis.

The investigation into the pathway that leads from the CAG repeat expansion to clinical phenotype in HD and related diseases had led to dozens of candidate pathogenic pathways. Early pathogenic steps appear to include abnormal phosphorylation and cleavage of the HTT protein into small fragments by cellular enzymes, and subsequent misfolding of fragments with long glutamine tracts. Small fragments of HTT may also be generated by abnormal splicing during the transcriptional process. These fragments eventually aggregate into structures that can be visualized with immunohistochemical stains in postmortem brain from patients with HD (Fig. 2.6–4A). Many mechanisms have been invoked to explain subsequent steps in pathogenesis. For instance, proteins with long glutamine stretches sequester other proteins, including transcription factors. These proteins become unavailable to perform their normal cellular duties, with potentially lethal consequences. Alternatively, proteins with long polyglutamine expansions, or aggregates of these proteins, may interfere with the process of autophagy, through which cells degrade abnormal protein and recycle organelles. Failure of this system might lead to an accumulation of a variety of toxic proteins. Other proposed pathogenic mechanisms include interference with mitochondrial energy metabolism, polyglutamine-induced vulnerability to excitotoxicity and oxidative stress, loss of brain-derived neurotrophic factor (BDNF) release, and abnormalities in the nucleocytoplasmic transport system. Recent work has also detected cell-to-cell spread of mutant HTT protein, with some similarities to the action of prion proteins. Further, a powerful convergence of genetic evidence and cell biology has demonstrated that neuronal repeat length in HD patients may increase over time, as normal DNA repair mechanisms go awry at the site of the expanded repeat in *HTT*. It is hypothesized that neurotoxicity arises once repeat length reaches an as yet undetermined threshold. This concept may explain why rare individuals with expansion mutations in both copies of their *HTT* gene do not have an unusually severe course of illness.

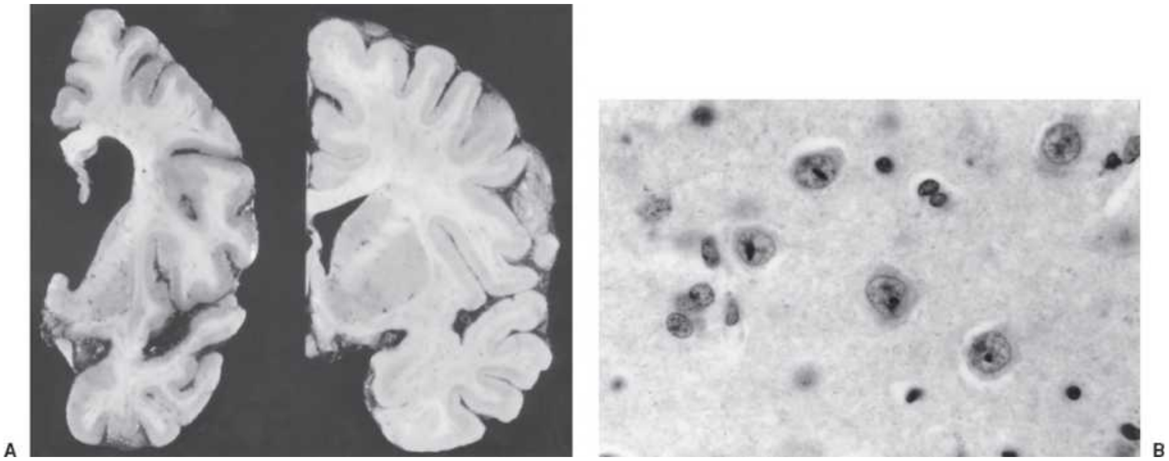

FIGURE 2.6–4. Huntington disease neuropathology. **A:** Gross pathology. HD case on the right, control brain on the left. Note the enlarged lateral ventricle, nearly absent caudate and putamen, and atrophy of the cerebral cortex. **B:** Intranuclear inclusions in HD. Immunohistochemical staining of cerebral cortex from an HD patient with an antibody that labels the huntingtin protein demonstrates prominent intranuclear inclusions in neurons. (Photographs courtesy of Dr Christopher A. Ross.)

Diagnosis and Clinical Features. Typically, the diagnosis of HD is made by a genetic test in the setting of motor signs and symptoms in an individual with a family history of HD. The advent of genetic testing now makes diagnosis possible in individuals with a family history of HD who do not yet have symptoms themselves, or in patients with unusual presentations or HD-like symptoms in the absence of a known family history of a movement disorder.

MOTOR MANIFESTATIONS. The primary *involuntary* movement abnormality is chorea or choreoathetosis, continuous and irregular writhing and jerking movements. The limbs and trunk are most prominently affected, but respiratory, laryngeal, pharyngeal, oral, and nasal musculature may also be involved. Chorea may be among the early symptoms, but often plateaus and may even diminish during the course of the disease. Voluntary movement abnormalities typically affect visual tracking (slow hypometric saccades and catchy pursuit), fine motor movements (slow, poorly coordinated, arrhythmic), speech and swallowing, tone (rigidity), and gait. The voluntary motor problems are most associated with functional disability and are most likely to lead to serious illness or death. Unlike chorea, voluntary movement abnormalities progressively worsen.

COGNITIVE MANIFESTATIONS. Cognitive decline generally parallels development of movement abnormalities. For example, when examined an average of 24 years before the predicted age of onset of clinical HD (as predicted by their repeat length and age), carriers of the HD gene in a systematic study had no detectable cognitive deficits. Carefully controlled studies indicate that detectable cognitive abnormalities likely emerge at about 10 years prior to clinical onset of disease, though these will progress slowly and may be difficult to detect in an individual patient. Early cognitive changes, which may be subtle, often affect cognitive flexibility and speed, such that switching from one task to another, or tracking multiple tasks simultaneously, becomes difficult or slow. Work or school performance may suffer to the point that colleagues notice. A four-domain model of cognitive deficits suggests that in early- to middle-stage HD, visuospatial deficits are the most pronounced deficits, followed by executive dysfunction, and then less prominent memory and verbal deficits. The dementia becomes more global as the disease progresses, but, as in other subcortical dementias, aphasia and agnosia are less evident than in AD. MMSE scores typically decline by 1 to 2 points/year.

PSYCHIATRIC MANIFESTATIONS. As many as 80% of patients with HD develop some form of noncognitive psychiatric disorder within 10 to 15 years of disease onset, and up to 20% may present with psychiatric symptoms. Since it is the psychiatric aspects of HD that are most amenable to treatment at present, psychopathology should be sought diligently. The consequences of these disorders may be devastating to the patient and family, perhaps more so than either the cognitive or motor dysfunction. The suicide rate is much higher than in the general population. Lifetime history of suicide ideation has been estimated at 20%, with a 6.6% lifetime rate of completed suicide. Data from the ENROLL-HD study revealed an incident rate of suicide and suicide attempts in HD gene carriers of 72 per 100,000 person-years and 306 to 375 per 100,000 person-years, respectively, approximately 10 times greater than non–HD gene carriers. The lifetime prevalence of psychotic phenomena is 11% to

13%, and perhaps higher in some populations. Thirty to 40% of patients with HD will develop major depression, and another 10% to 20% will have a different depressive syndrome. Up to 10% of patients with HD develop mania, though this diagnosis may be confused with disinhibition and other personality changes. Less frequently, patients with HD develop disorders that are clinically indistinguishable from idiopathic obsessive-compulsive disorder, schizophrenia, or delusional disorder. The overall prevalence of sleep disorder in HD is 15%, but increases with disease duration, so that by midcourse 27% of patients are affected. As in patients with other types of brain damage, patients with HD are vulnerable to delirium from pharmacologic or medical illness.

Personality change is very common, occurring in up to 50% of patients with HD by midcourse. Most patients with HD with personality change also have cognitive impairment or dementia, but cognitive deficits and personality change do not always occur together. Other patients may develop a naive, superficial, child-like personality that does not fit into current DSM terminology. Apathy and irritability are particularly common and troublesome states, may occur prior to overt clinical disease, and may exist as either separate syndromes or as symptoms of other disorders.

The many psychosocial issues that accompany HD increase the chance for adjustment disorders, relationship disorders, and other behavioral disorders. Caring for a cognitively impaired, irritable, or apathetic family member may result in emotional turmoil and financial hardship. Frequent additional stressors include social stigma (real or feared), parental guilt over passing the disease gene on to children, a family legacy of secrecy, and the burden faced by young gene-positive adults who can predict their own future in light of the progressive decline of a parent or other relative. These issues and their consequences must be recognized and treated, but should not be used to “explain away” other psychiatric disorders for which other treatments may be more appropriate.

Pathology and Laboratory Examination

Pathologic Examination. Gross pathology of HD (Fig. 2.6–4B) is limited to the brain, though microscopic or metabolic involvement has been detected in other tissue, including muscle, liver, adipocytes, and lymphocytes. In advanced cases, brain weight may be reduced by 25% to 30%. The most prominent atrophy involves the caudate and putamen, which is readily apparent on MRI scans or on gross pathology, along with accompanying hydrocephalus ex vacuo. The cerebral cortex becomes quite atrophied, with less pronounced atrophy in other regions of the basal ganglia (especially the globus pallidus and STN) and to a lesser and more variable extent in other brain regions.

On the microscopic level, selective neuronal vulnerability is striking. Striatal cell loss primarily affects medium spiny projection neurons, and in a gradient such that the earliest and most severe loss is in the dorsomedial region. Cortical cell loss consists mostly of large neurons in layers III, IV, and V. Surviving neurons in the striatum and cortex may have an increased number of dendrites and dendritic spines, but some cells will have evidence of dystrophic processes. Gliosis is present in regions with significant cell loss. Intranuclear inclusions are evident in striatum and other brain regions (Fig. 2.6–4B). The inclusions consist of amyloid-like fibrils that contain, among other proteins, mutant HTT, ubiquitin, and synuclein, but not tau. Inclusions may also be detected in neuronal cytoplasm and processes.

Laboratory Evaluation. It has been possible to directly test for the HD repeat expansion since the discovery of the gene in 1993, using a simple PCR-based assay on DNA obtained from a blood sample. The test is available from a number of commercial and academic centers. Interpretation of results is shown in Table 2.6–3. While the test is generally reliable, ambiguity may arise if a repeat length is on the borderline of one of the categories, since the results of even the best laboratories have an error of +/− 1 triplet. Also, occasionally, only a single allele of normal length is detected. This usually means that both alleles contain a repeat of the same length, but the possibility exists that an expanded allele went undetected, which requires additional assays.

Table 2.6–3.
Interpretation of HD Repeat Length (Longest of the Two Alleles)

Repeat Length (CAG Triplets)	Interpretation	Comments
<29	Not at risk	Repeat length stable
29–35	Not at risk	Reports of a few HD cases
36–39	May develop HD	Increasing risk with longer repeats
40 and greater	Will develop HD	Earlier age of onset with longer expansions

Ancillary Tests. MRI, CT, SPECT, and PET were previously of diagnostic importance, but are now mainly used to exclude strokes, tumors, and subdural hematomas, etc., that may coexist with HD. MRI scans should be considered if the patient’s course rapidly deteriorates or in the presence of signs and symptoms not typical for HD, as falls are common in HD, and the risk of a complicating but treatable brain injury is relatively high.

Premanifest Testing. The availability of a diagnostic test for HD has made testing of individuals who are at risk but do not yet have clinical disease testing possible. So far, only a minority of individuals at risk of HD request testing, often for psychological and practical reasons related to childbearing, long-term planning about caring for a relative with HD, and finances. Perhaps because of the self-selected nature of those seeking premanifest testing and the exclusion from testing of those who are suffering from major depression, most patients feel relieved to learn of their status, whether testing positive or negative. Careful protocols for pre- and post-test counseling have been developed to assure that patients understand the potential psychological risks and risks of discrimination (employment and insurance) that may arise from the test results. Expert counseling is mandatory before premanifest testing is obtained. For known gene carriers, it is possible to obtain prenatal diagnosis early in pregnancy. A few centers offer preimplantation testing of embryos generated by in vitro fertilization. Premanifest testing is likely to increase in the future, as the availability of clinical trials for premanifest individuals will increase incentive for learning about gene status.

Differential Diagnosis. The availability of a definitive diagnostic tests leaves two issues in HD differential diagnosis: when to consider a diagnosis of HD and what diagnoses to consider if a patient with HD-like symptoms tests negative for HD. HD may occasionally present with rigidity, dystonia, athetosis, or tic-like movements rather than chorea, sometimes in every member of a pedigree. At other times, patients present with typical clinical manifestations, but they have no known family history of HD. This may arise when disease onset occurs earlier in a child than their parent, a parent carrying the mutation has died from other causes prior to the onset of clinical HD, there is nonpaternity or adoption, or, in very rare cases, a new mutation arises that is not present in previous generations. Genetic testing should be considered in either situation. Table 2.6–4 lists some of the disorders that present with HD-like symptoms. Notably, any disease or insult that affects the basal ganglia might lead to chorea or dystonia. Very few diseases fully mimic the clinical features of HD coupled with autosomal dominant inheritance; examples include Huntington disease-like 2 (HDL2), which is not clinically distinguishable from HD, and spinocerebellar ataxia type 17.

Course and Prognosis. HD is relentlessly progressive and fatal (Fig. 2.6–3C). Onset is typically insidious, with early complaints of clumsiness, jerks, and tremor. While the chorea tends to plateau and may even decline later in the course, the voluntary motor abnormalities progress, paralleled by cognitive decline. Patients with late-stage HD require complete care for all activities of daily living. They are mostly nonverbal, rigid, and akinetic, but may make whole-body movements resembling myoclonic jerks when disturbed. Dementia is global and severe, but may be less severe than the patient’s appearance would suggest. Death usually occurs 15 to 20 years after disease onset, often from the consequences of swallowing difficulties (suffocation or aspiration pneumonia) or general debilitation. Juvenile-onset HD often presents with bradykinesia, rigidity, dystonia, and sometimes seizures or myoclonus. Chorea may be minimal or absent. The course seems to be more rapid and severe than typical HD. Large-scale longitudinal studies in Europe and the United States, including ENROLL, PREDICT, and PHAROS, have attempted to define the course of HD in sufficient detail that the effectiveness of interventions can be precisely measured against the anticipated course of disease in any given individual.

Table 2.6–4.
Differential Diagnosis of Choreiform Disorder

Genetic Causes of Chorea	Comment
Dentatorubralpallidoluysian atrophy (DRPLA)	Autosomal dominant, rare in United States, more common in Japan, variable presentation
Huntington disease-like 2	Rare autosomal dominant disorder clinically indistinguishable from HD
Autosomal dominant spinocerebellar ataxias	Most commonly in SCAs 2,3, but also in others
Hereditary Creutzfeldt–Jakob disease	Autosomal dominant; 85% of cases are sporadic
Wilson disease	Autosomal recessive, distinguished by eye and liver abnormalities, abnormal copper metabolism
Ceroid neuronal lipofuscinoses	Usually childhood onset autosomal recessive, rarely adult-onset autosomal dominant
Neuroacanthocytosis	Recessive and dominant forms, distinguished by abnormal red blood cells (acanthocytes)
Benign hereditary chorea	Usually childhood onset
Pantothenate kinase-associated neurodegeneration	Childhood onset, recessive
Fahr disease	Basal ganglia calcification, usually autosomal dominant, adult onset
Mitochondrial diseases	Maternal inheritance, heterogenous phenotypes
Paroxysmal dyskinesias	Induced by external factors, multiple genetic subtypes, some sporadic forms
Lesch–Nyhan syndrome	Boys only, self-mutilation
Other metabolic disorders	Phenylketonuria, guaric acidemia type I, methylglutaconic aciduria type III, various aminoacidopathies

C9ORF72 disease	Reports of prominent chorea in some mutation carriers
Neuroferritinopathy	Middle-age onset, associated with brain iron accumulation
Sporadic forms of chorea	Comment
Tardive dyskinesia	Common
Sydenham chorea	Poststreptococcal complication
Systemic lupus erythematosis	May occur in 1–7% of cases, chorea also seen in other autoimmune disorders
HIV	Direct encephalitis or secondary lesions
Thyroid disease	Typically hyperthyroidism
Neurosyphilis	May also complicate HIV
Paraneoplastic syndromes	Mechanism appears to be autoantibodies
Drug induced	Anticonvulsants, lithium, stimulants, levodopa

Treatment. There is currently no treatment of the underlying neurodegeneration of HD. Large-scale clinical trials of creatine and co-enzyme Q10 were both terminated early based on strong evidence that the drugs had no beneficial effect. Multiple other drugs have shown some promise only to fail at phase III trials. As of 2021, drugs under active investigation include the sigma-1 receptor agonist pridopidine, and the RNA splicing regulator branapram. Surgical treatments, either via DBS or fetal striatal transplantation, remain in early stages. Perhaps the most promising therapeutic approach is to suppress expression of the mutant HTT gene, using one of several methods that prevent gene transcription or translation. Many hurdles to this strategy must be overcome, including sufficient delivery to the necessary brain regions, off-target effects (i.e., incidental suppression of other genes), the amount of mutant *HTT* suppression that must be achieved to obtain a clinical effect, dosing, and time course of treatment. Indeed, the first phase III trial of a suppression strategy, in which an oligonucleotide (a short piece of DNA) that targeted *HTT* RNA for degradation was infused into CSF via lumbar puncture, was not successful. Nonetheless, multiple other suppression strategies are under investigation, and there remains considerable optimism in the field for the ultimate success of this strategy. In the meantime, more sophisticated study design, and improved measures of treatment outcome including highly sensitive cognitive, motor, and neuroimaging techniques, will likely increase the sensitivity and throughput of clinical trials.

Despite the absence of treatments to modify the fundamental course of the disease, the quality of life for the patient and their family can be substantially improved by medical intervention. Some patients are concerned about the social implications of chorea and others may find chorea functionally impairing. Chorea may be effectively treated by tetrabenazine or its derivatives deutetrabenazine and valbenazine which block dopamine release via action at the vesicular monoamine transporter 2. Chorea may also respond to low doses (0.5 to 2.0 mg/day) of typical high-potency antipsychotics such as haloperidol or fluphenazine, other neuroleptics, benzodiazepines, or other dopamine-depleting agents such as reserpine or amantadine. Care must be taken, since dopamine-depleting agents increase the risk of depression, and other agents may worsen voluntary movement abnormalities or apathy, or cause akathisia or excess sedation. The simple mechanical strategy of weights strapped to the wrists can often prove of considerable value without causing side effects. Environmental interventions, such as removal of throw rugs, adding grab bars in bathrooms, and avoidance of stairs, are always essential to reduce falls.

Other common symptoms can also be at least partially ameliorated. Swallowing abnormalities can be assessed by pharyngeal esophogram and treated with thickened liquid and softened solid diet. Botulinum injections may be useful for dystonia. Myoclonus has been successfully treated with valproate and clonazepam and valproate may also be effective for seizures.

There are no established pharmacologic treatments for cognition. In particular, there is little evidence for an effect of anticholinesterase inhibitors. Simple education interventions may prove helpful, as in other forms of dementia. Early in the disease, patients can be taught to write down notes and to limit the number of tasks undertaken at once. Patients, and especially family members and care providers, need to be told about the progressive cognitive decline, so that they can learn to phrase questions more simply, provide cues for recall, and keep cognitive demands on the patient at a level commensurate with cognitive capacity. The topic of guardianship or power of attorney for legal and medical matters should be raised before the patient is too incapacitated to participate in the discussion.

Depression may be treated with any of the standard antidepressants, with the caveat that as in any patient with brain damage, starting doses should be low and doses should be increased slowly, along with careful monitoring for delirium. SSRIs are often useful and, given the higher suicide risk in patients with HD, are notably less toxic in overdose than tricyclic antidepressants or MAOIs. SSRIs may occasionally cause akathisia, insomnia, or dyskinesias. Other antidepressants have been used successfully. ECT has been used quite successfully to treat life-threatening or intractable depression, and other than an increased risk of delirium, is generally well tolerated. Mild mood elevations may not warrant treatment, but mania, like depression, can be treated with standard pharmacologic agents; limited data and anecdotal accounts suggest that divalproex sodium or carbamazepine is better tolerated and more effective than lithium. Benzodiazepines or antipsychotics (the atypical agents may have fewer negative effects on movement than the typical agents) may be helpful adjuncts.

Aggression and irritability stem from multiple causes, including psychiatric syndromes, cognitive and personality changes, and environmental precipitants such as hunger, thirst, pain, frustration over an inability to communicate, boredom, or unexpected change in routine. Treatment of any underlying psychiatric disorder and careful attention to possible precipitants is the first approach. When necessary, anecdotal evidence supports the effectiveness, in some patients, of SSRIs, antipsychotics, anticonvulsants, β-blockers, and benzodiazepines. Apathy probably responds best to family education combined with a structured and stimulating environment. SSRIs, stimulants, and dopaminergic agents such as amantadine and bromocriptine may be of value, though the latter agents may increase chorea.

Wilson Disease

Introduction and Definition. WD, also named hepatolenticular degeneration, is an autosomal recessive disorder in which mutations in the copper-transporting gene ATP7B result in abnormal copper accumulation in the liver, brain, and other tissues, with consequent hepatic cirrhosis and neuronal degeneration. Psychiatric and movement abnormalities are common, and patients may present to a psychiatrist first. WD can be effectively treated in acute and chronic stages, yet fatal without treatment. Earliest recognition of the characteristic signs, symptoms, and laboratory abnormalities is therefore essential.

History. The disease was named after Kinnear Wilson, who first recognized the syndrome, including a high rate of psychiatric abnormalities, in 1912. In 1948, Cummings recognized the role of copper in disease pathogenesis; Schenberg and Gitlin, in 1952, discovered the deficiency of ceruloplasmin in patients with WD (establishing a test that is still in use), and, in 1960, Bearn definitively established the recessive nature of the disorder. The molecular basis of the disease emerged first with linkage to the esterase D locus on chromosome 13 in 1985, and then, in 1993, with the discovery of causative mutations in ATP7B.

Epidemiology. WD is found worldwide, with a frequency estimated at about 1/30,000. Approximately 1 in 90 individuals are heterozygous carriers.

Etiology. Over 600 disease-causing mutations have been identified in ATP7B. A few mutations are common (H1069Q accounts for more than 40% of mutations in Northern European populations), while most are rare. The result is that most individuals with WD are compound heterozygotes, carrying a different mutation on each allele. The Wilson protein, an ATPase involved in copper transport, functions to maintain copper homeostasis by facilitating excretion of copper from hepatocytes into bile. In patients with WD, this function fails, resulting in accumulation of copper in the liver, brain, and other tissues, eventually leading to hepatic failure and neurologic and psychiatric signs and symptoms. In the brain, the basal ganglia appear to be particularly affected.

Clinical Features. About 40% of patients first present with liver abnormalities, 40% with neurologic abnormalities, and up to 20% with primarily psychiatric abnormalities. Liver dysfunction, the most common childhood presentation of WD, may be manifest as mild elevations of serum transaminases, chronic active hepatitis, cirrhosis, or acute and fulminant liver failure. Copper toxicity may also lead to cardiac arrhythmias, renal disease, rhabdomyolysis, joint disease, anemia, and thyroid and parathyroid dysfunction. Neurologic signs and symptoms, which strongly correlate with brain and CSF copper levels, generally reflect the multiple brain regions involved, and include dysdiadokinesis, dysarthria, tremor, dystonia, rigidity, choreoathetosis, bradykinesia, masked facies, and micrographia. Seizures occur in 6% of patients.

Cognition in WD has not been well studied. A relatively mild degree of cognitive impairment akin to other basal ganglia diseases is thought to be present in about 25% of patients, worsening with time and in conjunction with accumulating neurologic symptoms.

Two-thirds of patients may have psychiatric symptoms by first presentation, and 10% to 20% of patients seek psychiatric treatment before WD is diagnosed (a stage at which, tragically, the diagnosis is often missed). Eventually, psychiatric manifestations affect nearly all patients with WD and the psychiatric profile generally resembles HD. The most common symptoms are personality changes (often described as irritability, aggressiveness, and emotional lability) and depression (ranging from subsyndromal mood changes to major depression, at times with suicidality). Anxiety disorders, psychotic disorders (and individual psychotic symptoms), and catatonia appear at higher rates than in the general population. About 10% of patients present with psychosis and are frequently misdiagnosed with primary schizophrenia or bipolar disorder. Psychiatric symptom severity correlates more strongly with neurologic deficits than to liver disease, indicating that the psychiatric phenomena arise from brain damage versus psychological reactions to the illness.

Diagnosis. Since the hepatic, neurologic, and psychiatric manifestations of copper toxicity are only partially reversible, prompt diagnosis of WD is essential. The key step is considering the diagnosis. Elevated liver transaminases are nonspecific, but maybe the first diagnostic clue. Kayser–Fleischer rings, copper deposits in the limbus of the cornea, may be detected by inspection or with a slit lamp examination, but are often absent early in the disease. Serum ceruloplasmin is below normal in 95% of patients with WD, but is also low in newborns and 20% of heterozygous WD carriers, and may be misleadingly normal in the setting of infection or inflammation. The absence of serum ceruloplasmin oxidase activity is a useful test in the latter cases. Elevated 24-hour urinary copper concentration is generally considered diagnostic of WD. Liver biopsy, the gold standard for diagnostic confirmation, may also be required and show increased hepatic copper concentration and evidence of hepatitis or cirrhosis. Genetic testing with sequencing of the ATP7B gene is confirmatory when the disease-causing mutations are identified, but negative genetic testing, in the absence of supportive laboratory data, does not exclude WD because of heterogeneity of ATPB7 mutations.

In patients with neurologic or psychiatric disorders, and sometimes in patients without clinical evidence of brain disease, head MRI will reveal generalized atrophy and increased T2-weighted signal in the basal ganglia secondary to copper deposition. The midbrain “panda sign,” common in WD but nonspecific, is the result of hyperintense signal around the red nucleus and putamen. Atrophy may be present in cerebral cortex (especially the frontal lobe), brainstem, and cerebellum. PET, SPECT, and MR spectroscopy are used to evaluate functional abnormalities; on SPECT, functional assessment of dopaminergic innervation shows deficits in both pre- and postsynaptic markers, which is relatively specific for WD versus other causes of striatonigral degeneration. The most prominent pathologic changes involve cavitary degeneration of the basal ganglia, with neuronal loss, gliosis, and high levels of copper.

Differential Diagnosis. Elevated hepatic copper may also result from cholestatic liver disease (normal or elevated ceruloplasmin) or idiopathic childhood cirrhosis (elevated ceruloplasmin). Aceruloplasminemia may result in neurologic symptoms but results in elevated liver iron rather than copper. Urine copper excretion may be elevated secondary to metal chelation therapy or in the presence of acute liver failure or cirrhosis with cholestasis or nephrotic syndrome.

Course, Prognosis, and Treatment. Onset of WD is usually in the second or third decade with mean age of onset at 17 years. Atypical late-onset cases are reported in individuals over the age of 60. Without treatment, death may occur within 5 years, and sooner with CNS disease. As WD is one of the few neurodegenerative diseases in which the fundamental pathogenic process, copper accumulation, is amenable to treatment, early consideration of WD, including in older patients, is essential; treatment outcomes are best when treatment begins within 1 month after symptom onset. Unfortunately, average symptom duration before correct diagnosis and treatment is about 1 year. Treatment involves either removing copper (chelation therapy) using D-penicillamine, trientine, and tetrathiomolybdate or blocking copper absorption by zinc salts. Except in areas with high levels of copper in the drinking water, dietary restrictions are inadequate to prevent copper overload. D-penicillamine, which is poorly tolerated by some patients, is prone to induce hypersensitivity reactions, and trientine can paradoxically increase neurologic and psychiatric symptoms acutely, as hepatic copper is mobilized and deposited in the brain. For patients with end-stage liver disease, orthotopic liver transplantation has proven successful. All treatment modalities appear to improve neurologic and psychiatric symptoms, but response may be partial if therapy began at an advanced stage.

There are no controlled trials regarding symptomatic treatment of psychiatric manifestations of WD. Anecdotal evidence suggests that patients with WD may be particularly vulnerable to motor side effects of antipsychotic medications. It seems prudent to treat psychiatric syndromes with standard psychiatric treatment modalities, both pharmacologic and psychotherapeutic, but to choose lower doses of medications, increase doses slowly, and avoid high-potency antipsychotics.

CEREBELLAR DISORDERS

Introduction

The cerebellum has long been considered a modulator of motor function. As the cerebellum also modulates cognition and affect, damage to the cerebellum from many different causes can result in cognitive deficits and psychiatric syndromes.

History

A role for the cerebellum in regulating motor function was established early in the 19th century, beginning with the studies of Luigi Rolando. By the early 20th century, the effect of cerebellar degenerative disease on coordination of the limbs, gait, posture, and speech was appreciated. By 1939, the work of Holmes established much of the terminology and methods of examination of the cerebellum still in use. Simultaneously, diseases that affect the cerebellum were delineated, including hereditary disorders described by Friedreich in 1863 and Marie in 1893. Anecdotal reports of cognitive and psychiatric disorders arising from cerebellar damage were noted in early cerebellar research; systematic investigation of this link is more recent, stimulated by a series of neuroanatomical studies that demonstrated connections between the cerebellum and multiple other brain regions, and frequent observations of odd behavior and mutism in patients following resection of posterior fossa tumors. Substantial evidence now supports an important role of the cerebellum in normal cognition and emotion and that cerebellar damage leads to neuropsychiatric phenomena, including those seen in common psychiatric disorders as major depression, schizophrenia, and autism.

Epidemiology

The overall prevalence of cerebellar disorders is difficult to determine, since many different diseases may affect the cerebellum. As many as 50% of patients who have alcohol dependence syndromes develop clinical or histologic evidence of cerebellar pathology; this may be the single most common cause of cerebellar damage worldwide. The cerebellum is affected in 1.5% to 8% of all strokes, so this is probably the second most common cause of cerebellar damage. In children, the annual incidence of tumors affecting the cerebellum is between 0.25 and 1/100,000/year. In adults, primary cerebellar tumors are rarer and metastases more common. Prevalence of recessive (primarily Friedreich ataxia), dominant, and sporadic neurodegenerative diseases that primarily affect the cerebellum is, in each case, about 2/100,000. Other causes of cerebellar lesions are rare.

Etiology

Table 2.6–5 lists many of the diseases that cause permanent injury to the cerebellum. With the exception of some strokes, tumors, and degenerative conditions, most of these disorders are unlikely to affect the cerebellum in isolation. Therefore, their cognitive and psychiatric manifestations likely derive from damage to multiple brain regions. However, established neuroanatomical links between cerebellum and other regions of the brain plus analysis of focal cerebellar lesions make it clear that damage to the cerebellum itself is sufficient to cause cognitive and emotional disturbance.

The pathway from the cerebral cortex to the cerebellum originates with projections from neurons in cortical layer Vb in motor, somatosensory, association, and paralimbic regions, including posterior parietal association cortex, portions of the superior temporal sulcus and gyrus, parastriate and parahippocampal regions, and dorsolateral, medial, and, to a lesser extent, ventral lateral prefrontal regions. After synapse in the pons, where topographical distinctions among the different sources of input are maintained, the pathway continues as projections into the cerebellar cortex. Feedback to the cortex begins with axons projecting from cerebellar Purkinje cells to the deep cerebellar nuclei, where the pathway again remains topographically defined. These nuclei then project to both “motor” and “nonspecific” nuclei of the thalamus. Axons in neurons from both types of thalamic nuclei project to posterior parietal, superior temporal, and prefrontal association regions, and the intralaminar nuclei also project to limbic regions. This cerebrocerebellar pathway appears to consist of a series of discrete and relatively closed loops, similar to the loops connecting the basal ganglia and the cerebral cortex. Other pathways of potential relevance include reciprocal connections between the cerebellum and brainstem catecholaminergic (locus coeruleus) and serotoninergic nuclei (dorsal raphe), the hypothalamus, and the brainstem reticular formation. fMRI studies now suggest that anterior cerebellar regions subserve sensorimotor processing, while more posterior regions subserve emotional and cognitive processing, with a specific functional association of different regions to relevant areas of cerebral cortex.

Table 2.6–5.
Diseases That Cause Cerebellar Damage: A partial list

Hereditary Degenerative Autosomal Dominant
Spinocerebellar ataxia types 1–42+
Dentatorubropallidoluysian atrophy
Huntington disease (uncommonly)
Hereditary prion diseases
Episodic ataxias types 1–7

Hereditary Autosomal Recessive (Complex Pathogenesis)
Friedreich ataxia (most common autosomal recessive cerebellar ataxia)
Hereditary Autosomal Recessive (Metabolic)
Abetalipoproteinemia
Niemann–Pick type C
Hexosaminidase A deficiency (Tay–Sachs)
Glutamic aciduria
Hydroxyglutaric aciduria
Cholestanolosis (cerebrotendinous xanthomatosis)
Metachromatic leukodystrophy
Galactocereberosidase deficiency (Krabbe disease)
Alpha-tocopherol deficiency (Vitamin E deficiency)
Hartnup disease
Refsum disease (phytanic acid oxidase deficiency)
Biotinidase deficiency

Hereditary Autosomal Recessive, DNA Repair
Ataxia-telangiectasia
Xeroderma pigmentosum
Cockayne syndrome

Hereditary X-Linked
Pelizaeus–Merzbacher disease
Adrenoleukodystrophy
Rett syndrome
Fragile X–associated tremor/ataxia syndrome.

Mitochondrial
Kearns–Sayre syndrome
Myoclonic epilepsy associated with ragged-red fibers (MERRF)
Neuropathy, ataxia, and retinitis pigmentosa (NARP)

Leiber hereditary optic neuropathy (LHON)

Nutritional/Dietary

Vitamin E
Thiamine (often in association with alcoholism)
Celiac disease

Systemic Diseases That May Cause Cerebellar Damage

Toxic
Ethanol
Phenytoin
Lithium
Organic solvents
Oncologic
Primary tumors
Metastases
Paraneoplastic (small cell carcinoma of lung, breast, female genital tract, lymphomas)
Infectious/postinfectious
Abscesses
Whipple disease
Subacute sclerosing panencephalitis
Progressive rubella panencephalitis
Infectious prion disease
Other
Multiple sclerosis
Sporadic late-onset cerebellar degeneration
Multisystem atrophy
Trauma
Radiation
Vascular
Embolic, hemorrhagic, or occlusive
Congenital (some autosomal recessive or X-linked hereditary)
Marinesco–Sjögren syndrome
Joubert syndrome
Disequilibrium syndrome
Early-onset ataxia with hypogonadism, retinopathy, or deafness

In addition to its connections to multiple other brain regions, the other aspect of cerebellar anatomy important to affect and cognition are the “cerebellar cortical modules.” These modules exist as radially oriented, structurally identical units that comprise the cerebellar cortex. Each module consists of a Purkinje cell (numbering about 15 to 30 million in the human cerebellum), granule cells that receive input from outside of the cerebellum and innervate the Purkinje cell, and inhibitory interneurons (basket, stellate, and Golgi cells). The Purkinje cell thus serves to integrate excitatory and inhibitory input.

It is generally believed that these units perform the same type of transform for all information coming into the cerebellum, with the specificity of information processing provided by the unique connections of each region of the cerebellum to other brain structures. Schmahmann has therefore proposed that just as the “cerebellum regulates the rate, force, rhythm, and accuracy of movements, so may it regulate the speed, capacity, consistency, and appropriateness of mental or cognitive processes.” He has termed this the “dysmetria of thought” hypothesis and suggests that damage to the cerebellum will lead to an “unpredictability and illogic to social and societal interaction” just as it leads to poorly coordinated and misjudged movements, using the term “cerebellar cognitive affective syndrome” to broadly define the clinical phenotype. He has suggested the cerebellar cognitive affective syndrome includes five axes: attention control, emotional control, autism spectrum disorders, psychotic disorders, and social skills. While intriguing, it remains uncertain to what extent this hypothesis can explain the multitude of cognitive and psychiatric phenomena that accompany cerebellar disease.

Diagnosis and Clinical Features

The diagnosis of cerebellar disorders involves narrowing the complicated differential: Is onset age at birth (congenital), childhood (metabolic, oncologic, Friedreich ataxia), middle age (spinocerebellar ataxias), or older adults (oncologic, vascular)? What other organ systems or regions of the nervous system are affected? Is the process acute (vascular, toxic, infectious), subacute (infectious, oncologic), chronic (hereditary, toxic, nutritional), or episodic? Is there progression or have symptoms remained static? Is there a family history of a similar disorder? Have there been exposures to drugs, alcohol, or toxins? The history and examination, followed by an MRI scan (and potentially other imaging modalities such as

PET, SPECT, or angiography) and then specific genetic or metabolic tests will often, but by no means always, provide a final diagnosis.

Motor Features. Problems with balance and speech are often the presenting complaints in disorders with primarily cerebellar involvement. The "cerebellar gait" is wide based and staggering. Patients may complain of walking as if they were drunk, and stairs are particularly difficult. On examination, patients are unable to perform a tandem walk and may sway or display titubation, a low-frequency tremor of the body. The dysarthria is characterized by slurred speech of irregular rhythm and with incorrect emphasis of syllables. Speech may be slowed and monotonic (scanning speech) in compensation.

Cerebellar damage is also manifest in other ways. Eye movement abnormalities vary, and may include nystagmus, square wave jerks (inability to maintain fixation when gazing at a specified point), dysmetria (over- or undershoot when making rapid eye movements from one target to another), and uneven smooth pursuit. Limb incoordination may be manifest as trouble with fine motor tasks. Tremor may be either postural or kinetic; the latter is often observable as an intention tremor during finger-to-nose testing. The presence of other signs or symptoms signals the possibility that other regions of the CNS are also involved, more the rule than the exception with diseases affecting the cerebellum.

Cognitive Features. The cerebellum is involved in a variety of cognitive tasks, including procedural learning, executive function, language processing, visual–spatial orientation, sensory processing, timing, and attention. Within 10 to 15 years of disease onset, about 20% of patients with cerebellar degeneration will meet criteria for cognitive decline or dementia, and the rate presumably increases significantly later in the disease course. The pattern of deficits is consistent with a subcortical dementia, with deficits in verbal and nonverbal intelligence and especially executive function in the absence of aphasia, apraxia, or agnosia. Cognitive deficits in these patients, particularly when subtle, may not be detected on simple clinical mental state examinations that focus on language and memory. Executive functions impaired in cerebellar degeneration include attention, sequencing, and timing. Executive, visuospatial, verbal and visual memory, and language deficits also frequently follow focal lesions of the cerebellum. It has been suggested that cognitive deficits are more likely to follow lesions of the cerebellar hemisphere, while vermal lesions result in psychiatric disturbances.

Psychiatric Features. The frequency and nature of psychiatric syndromes in patients with cerebellar disease was initially noted in case reports and a few small studies. The findings of these studies have now been replicated and expanded in more systematic investigations. Following strokes restricted to the posterior circulation (cerebellum and brainstem), about 35% of patient develop either a major or minor depression acutely, with the frequency declining to about 20% 6 months after the stroke and 10% 12 to 24 months later. Personality changes, particularly blunted affect and disinhibition, may be prominent in patients with large posterior circulation infarcts, and potentially other lesions, while anterior lobe or very small cerebellar infarcts elsewhere are less likely to produce personality change.

In degenerative disorders predominately limited to the cerebellum, about 75% of patients develop one or more psychiatric syndromes within 10 to 15 years of disease onset. As many as one-third experience at least one episode of major depression and another third experience other forms of depression (including brief recurrent depression, minor depressive disorder, and dysthymia). Suicidality may be high, with The Clinical Research Consortium for Spinocerebellar ataxias reporting a rate of suicidal ideation as high as 65% in spinocerebellar ataxia type 3. Self-awareness of depression is frequently absent. About 25% develop persistent personality changes, similar in type to those found in HD, and, as in HD, not limited to patients with cognitive decline. The frequency of mania, anxiety syndromes, and psychotic syndromes may all be somewhat elevated. The neuropsychiatric findings in disorders in which cerebellar degeneration is accompanied by degeneration in other brain regions may be particularly severe. For instance, neuropsychiatric findings in the late-onset autosomal dominant disorder spinocerebellar ataxia type 2 (loss of cerebellar Purkinje and granule cells, loss of inferior olive, pontocerebellar nuclei, substantia nigra, and motor neurons, and white matter tract degeneration) include executive dysfunction, impaired memory, apathy, depression, and potentially psychosis.

Pathology and Laboratory Examination. The pathology of cerebellar disorders depends on the underlying disease. Even with isolated cerebellar diseases, pathologic examination is likely to reveal noncerebellar findings. The choice of laboratory evaluation, including genetic testing for hereditary disorders, metabolic screens for metabolic disorders, and immunologic screens for immune-related disorders (especially of paraneoplastic origin) will be based on clinical and MRI findings. Neuropsychological testing may help detect subtle cognitive difficulties of considerable clinical relevance. In addition to the diagnostic workup, imaging studies are important if the nature or severity of signs and symptoms change unexpectedly, to exclude the presence of new subdural hematomas, infarcts, or tumors.

Differential Diagnosis. The differential diagnosis of the underlying disease for a patient presenting with cerebellar signs and symptoms is complex. Various schemes are proposed for either nosologic purposes or for the differential diagnosis in an individual patient. One practical and logical approach is to divide disorders into nonhereditary and hereditary, and to further classify each of these broad groups into subcategories. This approach is depicted in Table 2.6–5. Similar to other movement disorders, the differential diagnosis of affective symptoms is perhaps the most important issue facing the psychiatrist. Delirium must always be considered, regardless of whether criteria for DSM-5 major depressive disorder are met. Isolated affective symptoms in the absence of delirium and major depression may reflect a subsyndromal depressive disorder, but could also be manifestations of a personality change or cognitive deficits either resulting in affective change (particularly apathy or irritability) or fueling demoralization.

Course and Prognosis. The course of cerebellar diseases varies widely by etiology, ranging from steadily progressive to intermittent to completely static. Personality and cognitive change are likely to follow the underlying disease course, while major depression may be more intermittent, suggesting differences in pathophysiology among these disorders.

Treatment. Determining the cause of the underlying disorder is essential, since cerebellar degeneration secondary to nutritional, dietary, and some metabolic disorders may be treatable and at least partially reversible. Cognitive treatments are supportive and educational, as described for HD. Little is known about the treatment of psychiatric disorders. Anecdotal accounts indicate that patients respond to standard treatments, including various classes of antidepressants and ECT. Treatable conditions, especially mood disorders, are likely under-recognized and undertreated. Caution is necessary with agents that may increase ataxia. Large-scale international registries of patients with specific disorders (e.g., European Friedreich ataxia database registry) may help elucidate disease course, establish biomarkers, and facilitate clinical trials of likely agents.

Essential Tremor

ET has been well characterized since 1887, but controversy remains over its definition, whether it represents a symptom, syndrome, or disease entity, and the basis for associated neuropsychiatric abnormalities. ET affects 1% of the population and increases at least fivefold with age, making it the most common movement disorder. Age onset is bimodal. Early onset has a male predominance and is typically in children and before age 25 years; it may even be evident at birth. Older onset begins around age 45 or older and a male predominance is not definitive. In 50% to 70% of patients, ET is a familial disorder, and genetic factors are important, but no causative gene mutations have been consistently reproduced. Growing evidence demonstrates the heterogeneity of ET with respect to its clinical presentation, age onset, progression, pharmacologic response, genetics, pathology, and relationships with other diseases.

The main clinical sign in ET is gradual evolution of a bilateral postural and kinetic (action) tremor affecting the upper limbs. Tremor of the head, voice, or lower limbs may also be present. ET is either sporadic or genetic in origin and is not associated with other known causes of tremor. Activities of daily living can be significantly affected.

The etiology and pathophysiology of ET are unclear. Traditionally, ET was regarded as a benign condition presenting as a monosymptomatic disorder without other neurologic involvement or nonmotor effects. Whether the heterogeneity of ET represents a complex neurodegenerative disorder or a functional reversible disorder of neuronal oscillation is debated. Recent postmortem studies suggest that ET may be a cerebellar disease affecting Purkinje cells. This is supported by increasing evidence that ET is associated with development of other motor abnormalities, including comorbid PD, dystonia, and myoclonus, cognitive and psychiatric features, and olfactory and oculomotor deficits. The kinetic component of ET plus subtle signs of ataxia suggest cerebellar involvement and neuroimaging studies suggest nigrostriatal damage. Other studies identified higher prevalence of MCI or selective cognitive deficits indicative of frontocerebellar dysfunction in association with ET, including executive dysfunction, and impairment in attention, verbal fluency, verbal learning, and verbal memory. Higher rates of depressive and anxiety symptoms, sleep disturbances, and fatigue are reported in ET, but the epidemiology and features of discrete psychopathologic syndromes have not been examined. Self-consciousness and embarrassment about symptoms are common. The sensation of internal tremor can be experienced as anxiety. No studies have addressed specific treatment of neuropsychiatric features in ET.

Treatment of ET is usually initiated when symptoms are socially problematic or interfere with daily functioning. The first-line medication treatments used are β-blockers and antiseizure medications, but response is often suboptimal. Chemodenervation with botulinum toxin injections, customized to the muscles associated with the tremor, can improve function and tremor. DBS and FUS, typically targeting the thalamic ventral intermediate nucleus (ViM), can improve action tremor and associated daily functioning for several years, but data on the duration of FUS effects is limited given the technique is relatively new.

Dystonias

Dystonia is characterized by involuntary, sustained muscle contractions that result in twisting, torsional, repetitive movements of one or more body parts and frequently appear as abnormal and sometimes painful or awkward positions and postures. It is the third most common movement disorder after PD and ET. Like ET, it can present as a monosymptomatic condition or as a feature of broader syndrome. Any part of the body can be affected, including the limbs, trunk, neck, eyelids, face, or vocal cords. In 2013, an international consensus committee developed a new classification system for the dystonias based on two axes: Axis 1, the clinical characteristics (age of onset, body distribution of symptoms, temporal pattern of symptoms, and associated features) and Axis 2, the etiology based on shared biologic mechanisms (e.g., neuropathology, inherited, acquired, idiopathic). Age onset is defined as childhood, adolescence, or adulthood (>28 years) and is an important predictor of whether symptoms will spread to other body parts. Distribution of dystonia is categorized as generalized, focal, segmental, multifocal, or hemidystonic and is one indicator of disease severity. The etiologic classification scheme is evolving with ongoing discovery of different genetic forms of dystonia. Isolated dystonias are defined by the existence of dystonia as the major symptom and include both hereditary and sporadic forms. In combined syndromes, dystonia is accompanied by other prominent neurologic features, such as parkinsonism or myoclonus, or other systemic features, including psychiatric symptoms. At least 10 genes or chromosomal locations are associated with isolated and combined dystonias, including early-onset generalized dystonia, dopa-responsive dystonia, paroxysmal dystonia, X-linked dystonia-parkinsonism (Lubag), myoclonic dystonia, and rapid-onset dystonia-parkinsonism. Acquired dystonias can be attributed to cerebral insults, including perinatal injury, trauma, toxins, or infarction or as a dopaminergic medication-related effect in PD or schizophrenia. Dystonia also occurs in neurodegenerative conditions, such as WD and PD. Neurophysiologic dysfunction in dopaminergic neurotransmission and basal ganglia output activity is thought to underlie dystonic disorders in which there is no evidence of nerve cell loss.

Dystonic movements begin with a particular motor action, but eventually generalize to multiple actions and are then present at rest. The contractions may be extreme, usually resulting in muscle hypertrophy and often pain or an oppositional tremor. Certain postures or sensory tricks can reduce dystonic contractions. Anticholinergic medications used to treat dystonia can affect memory, but cognitive impairment is not a general feature of dystonias.

Increased attention is being paid to the neuropsychiatric features of dystonia. Until 1997, with evidence that dystonia was an inherited neurologic disorder, it is noteworthy that there was sustained controversy as to whether dystonia was a primary psychiatric condition (a “motor neurosis”) versus a neurologic condition related to basal ganglia dysfunction. Pain, sleep disturbances, psychiatric symptoms, fatigue, and cognitive impairment impact quality of life and aggravate dystonia-associated disability and motor symptoms. Compared to healthy controls, patients with focal dystonia have over twice the rate of psychiatric disorders, with increased lifetime and point prevalence for major depression and anxiety disorders, particularly social phobia (71% in a series of patients with cervical dystonia) and panic disorder. Akin to PD, the psychiatric disturbance begins many years before onset of dystonia and there is not a clear correlation between psychiatric and dystonia symptom severity, indicating that the mood disturbance emerges from underlying brain pathology and is not solely an emotional reaction to motor symptoms. Similarly, obsessive-compulsive disorder may be over-represented in families with myoclonic dystonia. Shared pathophysiologic mechanisms in dystonia and psychiatric disturbances are further supported by recent transcriptome analyses showing gene co-expression in specific brain areas significantly enriched for dystonia genes and for depression, obsessive-compulsive disorder, and schizophrenia heritability. A reactive element, nonetheless, also exacerbates psychiatric distress, and can be a result of visible and painful motor symptoms. Alcohol may be misused to self-treat motor symptoms. The differential diagnosis of the various dystonias also includes functional movement disorders, including functional dystonias, which, not infrequently, may co-occur with other dystonias.

The prognosis of dystonia and rate of progression vary with the underlying condition and age of onset. Often, it stabilizes within 5 years of onset, but stressful circumstances provoke symptoms. Most therapy is symptomatic. Strategies include medications to reduce spasms and pain (e.g., anticholinergics, benzodiazepines, and muscle relaxants). Dopamine and dopamine-blocking agents have either negative or beneficial effects on dystonias. Botulinum toxin injections into selected muscles are used extensively to treat the various forms of dystonia. Exercise programs are an important component of therapy. When medical approaches are inadequate, surgery includes thalamotomy, pallidotomy, and DBS procedures. There are no studies on treatment of psychiatric disorders associated with dystonia, but care must focus on the identifying and treating psychiatric disorders or psychological distress. For example, botulism injections may reduce torticollis and depressive symptoms, but self-perceptions remain impaired. SSRIs can precipitate, aggravate, or improve dystonia. ECT may benefit motor and mood symptoms. Initial studies indicate that DBS procedures for dystonia do not have adverse (or positive) effects on rates of cognitive or psychiatric disturbances. Persistent anxiety after DBS-related improvement in dystonia suggest that anxiety is modulated by different networks.

FURTHER READINGS

Akbar U, Ashizawa T. Ataxia. *Neurol Clin*. 2015;33(1):225–248.

Austgen G, Marsh L. Cognitive dysfunction and neuropsychiatric aspects of Parkinson's disease. *Prog Brain Res*. 2022;269(1):59–90. doi: 10.1016/bs.pbr.2022.01.009

Barton B, Grabli D, Bernard B, et al. Clinical validation of Movement Disorder Society-recommended diagnostic criteria for Parkinson's disease with dementia. *Mov Disord*. 2012;27(2):248–253. doi: 10.1002/mds.24059

Coughlin DG, Litvan I. Progressive supranuclear palsy: advances in diagnosis and management. *Parkinsonism Relat Disord*. 2020;73:105–116.

Cummings J, Friedman JH, Garibaldi G, et al. Apathy in neurodegenerative diseases: recommendations on the design of clinical trials. *J Geriatr Psychiatry Neurol*. 2015;28(3):159–173.

DeLong MR, Wichman T. Basal ganglia circuits as targets for neuromodulation in Parkinson's disease. *JAMA Neurol*. 2015;72(11):1354–1360.

Divya KP, Kishore A. Treatable cerebellar ataxias. *Clin Park Relat Disord*. 2020;3:100053

Goodarzi Z, Mrklas KJ, Roberts DJ, Jette N, Pringsheim T, Holroyd-Leduc J. Detecting depression in Parkinson disease: a systematic review and meta-analysis. *Neurology*. 2016;87(4):426–437.

Hedera P. Wilson's disease: a master of disguise. *Parkinsonism Relat Disord*. 2019;59:140–145.

Heinzel S, Berg D, Gasser T, Chen H, Yao C, Postuma RB; MDS Task Force on the Definition of Parkinson's Disease. Update of the MDS research criteria for prodromal Parkinson's disease. *Mov Disord*. 2019;34:1464–1470.

Hoffman LA, Vilensky JA. Encephalitis lethargica: 100 years after the epidemic. *Brain*. 2017;140(8):2246–2251.

Höglinger GU, Respondek G, Stamelou M, et al; Movement Disorder Society-endorsed PSP Study Group. Clinical diagnosis of progressive supranuclear palsy: the movement disorder society criteria. *Mov Disord* 2017;32(6):853–864.

Jankovic J, Tan EK. Parkinson's disease: etiopathogenesis and treatment *J Neurol Neurosurg Psychiatry*. 2020;91:795–808.

Litvan I, Goldman JG, Tröster AI, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Mov Disord*. 2012;27(3):349–356. doi: 10.1002/mds.24893

Lo RY, Figueroa KP, Pulst SM, et al. Depression and clinical progression in spinocerebellar ataxias. *Parkinsonism Relat Disorder*. 2016;22(1):87–92.

McKeith IG, Boeve BF, Dickson DW, et al. Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB Consortium. *Neurology*. 2017;89(1):88–100.

Medina Escobar A, Pringsheim T, Goodarzi Z, Martino D. The prevalence of depression in adult onset idiopathic dystonia: systematic review and metaanalysis. *Neurosci Biobehav Rev*. 2021;125:221–230.

Medina Escobar A, Martino D, Goodarzi Z. The prevalence of anxiety in adult-onset isolated dystonia: a systematic review and meta-analysis. *Eur J Neurol*. 2021;28(12):4238–4250.

Mencacci NE, Reynolds R, Ruiz SG, et al; UK Brain Expression Consortium, International Parkinson's Disease Genomics Consortium. Dystonia genes functionally converge in specific neurons and share neurobiology with psychiatric disorders. *Brain*. 2020;143(9):2771–2787.

Pahwa R, Isaacson SH, Small GW, Torres-Yaghi Y, Pagan F, Sabbagh M. Screening, diagnosis, and management of Parkinson's disease psychosis: recommendations from an expert panel. *Neurol Ther*. 2022;11(4):1571–1582. doi: 10.1007/s40120-022-00388-y

Pontone GM, Mills K. Optimal treatment of depression and anxiety in Parkinson's disease. *Am J Geriatr Psychiatry*. 2021;29(6):530–540.

Rocha NP, Mwangi B, Gutierrez Candano CA, Sampaio C, Furr Stimming E, Teixeira AL. The clinical picture of psychosis in manifest Huntington's disease: a comprehensive analysis of the enroll-HD database. *Front Neurol*. 2018;9:930.

Rosness TA, Engedal K, Chemali Z. Frontotemporal dementia: an updated clinician's guide. *J Geriatr Psychiatry Neurol*. 2016;29(5):271–280.

Sanders RD, Gillig PM. Gait and its assessment in psychiatry. *Psychiatry (Edgemont)*. 2010;7(7):38–43.

Schmahmann JD. Emotional disorders and the cerebellum: neurobiological substrates, neuropsychiatry, and therapeutic implications. *Handb Clin Neurol*. 2021;183:109–154.

Sinoff G, Badarny S. Mild cognitive impairment, dementia, and affective disorders in essential tremor: a prospective study. *Tremor Other Hyperkinet Mov (NY)*. 2014;4:227.

Stankovic I, Krismer F, Jesic A, et al; Movement Disorders Society MSA (MODIMSA) Study Group. Cognitive impairment in multiple system atrophy: a position statement by the Neuropsychology Task Force of the MDS Multiple System Atrophy (MODIMSA) study group. *Mov Disord*. 2014;29(7):857–867.

Stimming EF, Bega D. Chorea. *Continuum (Minneap Minn)*. 2022;28:1379–1408.

Tabrizi SJ, Estevez-Fraga C, van Roon-Mom WMC, et al. Potential disease-modifying therapies for Huntington's disease: lessons learned and future opportunities. *Lancet Neurol*. 2022;21(7):645–658.

van Duijn E, Fernandes AR, Abreu D, Ware JJ, Neacy E, Sampaio C. Incidence of completed suicide and suicide attempts in a global prospective study of Huntington's disease. *BJPsych Open*. 2021;7(5):e158.

2.7 Neuropsychiatric Aspects of Multiple Sclerosis and Other Demyelinating Disorders

Joshua J. Rodgers, M.D.

The higher brain functions referred to as "neuropsychiatric," including complex processes such as emotion, cognition, and behavior, are resultant of distributed cortical–cortical networks modulated by cortical–subcortical and cerebellar–subcortical circuits, all of which depend on myelin-insulated axonal connections (i.e., white matter) for efficient functioning. As such, these neuropsychiatric functions are, more often than not, disrupted in the setting of white matter disorders, including demyelinating disorders. The cerebral white matter consisting of neuronal axons and glial cells—one type of which, oligodendrocytes, creates the myelin sheath around axons—comprises half of the volume of the human brain and can be vulnerable to a wide array of afflictions of which the reader should be aware (Table 2.7–1). Many of these disorders from the various categories can cause damage to myelin and thus be said to be demyelinating.

The demyelinating disorders are heterogeneous in etiology and phenomenology (depending largely on lesion location and what networks are disrupted) but broadly share a common basic pathology, namely, the partial or complete loss of the myelin sheath surrounding axons in the white matter of the brain and spinal cord. As white matter tracts are also essential for the proper functioning of motor and sensory functions (e.g., the corticospinal tract, optic nerve, spinothalamic tract, etc.), demyelinating disorders can also disrupt these neurologic functions as well and result in abnormalities in standard neurologic examination (e.g., weakness, vision loss, somatosensory loss, etc.). However, while these motor and sensory abnormalities are easily assessed by the standard neurologic examination and are often the focus of assessments, the neuropsychiatric sequelae of demyelination account for the majority of the disability in these disorders, and thus psychiatrists and neurologists should be skilled in assessing and treating these aspects as well.

Table 2.7–1.
Classification of Select White Matter Disorders

1. Demyelinating diseases	Multiple sclerosis, neuromyelitis optica spectrum disorders (Devic syndrome), acute disseminated encephalomyelitis, acute hemorrhagic leukoencephalitis, Schilder disease, Marburg disease, Balò concentric sclerosis
2. Genetic disorders	Leukodystrophies, aminoacidurias, phakomatoses, mucopolysaccharidoses, muscular dystrophy, callosal agenesis
3. Infectious diseases	Acquired immunodeficiency syndrome dementia complex or HIV-associated neurodegeneration, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, progressive rubella panencephalitis, varicella zoster encephalitis, cytomegalovirus encephalitis, Lyme disease, COVID-19 encephalitis
4. Inflammatory diseases	Systemic lupus erythematosus, Behçet disease, Sjögren syndrome, Wegener granulomatosis, temporal arteritis, polyarteritis nodosa, scleroderma, isolated angiitis of the central nervous system, sarcoidosis
5. Toxic leukoencephalopathies	Irradiation toxicity, chemotherapy and other therapeutic drug toxicity, drugs of abuse, environmental toxins
6. Metabolic disorders	Cobalmin (B_{12}) deficiency, folate deficiency, central pontine myelinolysis, hypoxia, hypertensive encephalopathy, eclampsia, high-altitude cerebral edema
7. Vascular diseases	Migraine, stroke and Binswanger disease, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy, leukoaraiosis, cerebral amyloid angiopathy, prematurity, and birth trauma–associated disorders

	(periventricular leukomalacia, periventricular hemorrhage—cerebral palsy)
8. Traumatic disorders	Traumatic brain injury, shaken baby syndrome, corpus callosotomy
9. Neoplasms	Gliomatosis cerebri, gliomas, primary cerebral lymphoma
10. Hydrocephalus	Congenital hydrocephalus, acute hydrocephalus, normal pressure hydrocephalus

Data from Filley CM. *The Behavioral Neurology of White Matter*. 2nd ed. Oxford University Press. 2012.

By far, the most common and well-studied demyelinating disorder is multiple sclerosis (MS). In fact, MS is the most common nontraumatic neurologic disease of young and middle-aged adults and, in the majority of cases, will manifest with neuropsychiatric symptoms either on presentation or over the disease course.

MS has been described in the literature for more than a century and a half, and over this time, the emotional, cognitive, and behavioral disturbances of MS have been recognized along with the motor and sensory dysfunction, and substantial efforts have been made to characterize the full spectrum of MS-related neuropsychiatric sequelae since. This interest has been sustained by an increasing awareness that different forms of psychopathology are very common in these disorders and that the prevalence exceeds what would be expected to occur in other similar chronic and disabling conditions. Moreover, it has also been hoped that these disorders, especially MS, may represent a biologic model for primary mood disorders, which may facilitate the elucidation of the etiology and neuroanatomy of primary mood and other psychiatric disorders. Indeed, these findings have led to research into the role of white matter abnormalities in the pathophysiology underlying primary psychiatric disorders, including schizophrenia, depression, bipolar disorder, attention deficit hyperactivity disorder, and autism.

MULTIPLE SCLEROSIS

MS is a chronic, as of yet, incurable disease of the central nervous system (CNS), which can have an onset at nearly any age and is a highly variable course. MS has a complex and not fully known etiopathogenesis characterized pathologically by inflammation and demyelination as well as neuronal degeneration and cortical involvement thought to be due to autoimmune attack of oligodendrocytes and the myelin sheath they create. The lesions appear to occur at random locations, mostly in frontal periventricular white matter, and thus result in high variability in the clinical presentation.

Epidemiology

Nearly 3 million people worldwide are afflicted by MS, and its prevalence is increasing. The highest prevalence rates appear to occur in Caucasians of northern European descent, and in the United States, a clear north–south gradient of prevalence has been demonstrated. The disease is less common in certain Asian populations, such as Uzbekistanis, native Siberians, North and South Native Americans, Chinese, and Japanese, as well as African Americans and New Zealand Maoris. The data suggest more than just a geographical vulnerability but also strong genetic influences. The disease is far more common in women, as are other autoimmune disorders. MS is most commonly diagnosed in the age group of 20 to 50 years of age, but the prevalence of pediatric-onset appears to be increasing dramatically.

Clinical Features and Diagnosis

The French neurologist Jean-Martin Charcot first established diagnostic criteria for MS in the middle of the 19th century. His original diagnostic triad is now known not to be specific, and in fact, over the years, numerous other diagnostic criteria were developed, all of which had a lack of specificity for the illness, thus including many false positives. The chief components of the diagnostic criteria have remained (1) dissemination of distinct attacks of focal neurologic deficits in time (relapsing/remitting features separated by a period of at least 1 month) and space (distinct CNS lesions and deficits) and lasting a minimum of 24 hours, and (2) exclusion of other causes. The 2010 McDonald criteria, consisting largely of clinical and radiographic (magnetic resonance imaging [MRI]) proof of dissemination of disease in time and space, were revised in 2017 to include the latest research and improve specificity. These revisions include expanding the possible evidence of disease dissemination in time (such as the presence of intrathecal oligoclonal bands) and space (such as the inclusion of both symptomatic and asymptomatic lesions and the inclusion of cortical lesions as equivalent to juxtacortical lesions) (see Table 2.7–2).

The clinical diagnostic criteria specified allow one to distinguish MS from disseminated encephalomyelitis, which is most commonly a single episode of illness, with all lesions occurring within a few weeks, although multiple separate lesions do sometimes occur. Once believed to be a subtype of MS, the evolving concept of neuromyelitis optica spectrum disorders (NMOSD) is an important clinical entity to distinguish from MS along with other disorders in the broad differential diagnoses (see below).

The diagnosis of MS remains largely clinical. The initial symptoms of MS are nonspecific and heterogeneous and may include a broad range of somatosensory, ocular, motor symptoms, and neuropsychiatric syndromes. The motor symptoms could include transient hemiparesis and various forms of seizure, among others. Symptoms of cerebellar dysfunction may occur, including scanning speech, gait ataxia, and various tremors. Although these mentioned symptoms and signs are common in the initial presentation, almost any neurologic symptom or sign can be the first presentation of MS.

Table 2.7–2.
The 2017 McDonald Criteria for Diagnosis of Multiple Sclerosis in Patients with an Attack at Onset

	Number of Lesions with Objective Clinical Evidence	Additional Data Needed for a Diagnosis of Multiple Sclerosis
≥2 clinical attacks	≥2	None*
≥2 clinical attacks	1 (as well as clear-cut historical evidence of a previous attack involving a lesion in a distinct anatomical location†)	None*
≥2 clinical attacks	1	Dissemination in space demonstrated by an additional clinical attack implicating a different CNS site or by MRI‡
1 clinical attack	≥2	Dissemination in time demonstrated by an additional clinical attack or by MRI§ OR demonstration of CSF-specific oligoclonal bands¶
1 clinical attack	1	Dissemination in space demonstrated by an additional clinical attack implicating a different CNS site or by MRI‡ AND Dissemination in time demonstrated by an additional clinical attack or by MRI§ OR demonstration of CSF-specific oligoclonal bands¶

If the 2017 McDonald Criteria are fulfilled and there is no better explanation for the clinical presentation, the diagnosis is multiple sclerosis. If multiple sclerosis is suspected by virtue of a clinically isolated syndrome but the 2017 McDonald Criteria are not completely met, the diagnosis is possible multiple sclerosis. If another diagnosis arises during the evaluation that better explains the clinical presentation, the diagnosis is not multiple sclerosis. An attack is defined in panel 1.

*No additional tests are required to demonstrate dissemination in space and time. However, unless MRI is not possible, brain MRI should be obtained in all patients in whom the diagnosis of multiple sclerosis is being considered. In addition, spinal cord MRI or CSF examination should be considered in patients with insufficient clinical and MRI evidence supporting multiple sclerosis, with a presentation other than a typical clinically isolated syndrome, or with atypical features. If imaging or other tests (e.g., CSF) are undertaken and are negative, caution needs to be taken before making a diagnosis of multiple sclerosis, and alternative diagnoses should be considered.

†Clinical diagnosis based on objective clinical findings for two attacks is most secure. Reasonable historical evidence for one past attack, in the absence of documented objective neurological findings, can include historical events with symptoms and evolution characteristic for a previous inflammatory demyelinating attack; at least one attack, however, must be supported by objective findings. In the absence of residual objective evidence, caution is needed.

‡The MRI criteria for dissemination in space are described in panel 5.

§The MRI criteria for dissemination in time are described in panel 5.

¶The presence of CSF-specific oligoclonal bands does not demonstrate dissemination in time per se but can substitute for the requirement for demonstration of this measure.

Reprinted from Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 2017;17(2)162–173. © 2017, with permission from Elsevier. doi:10.1016/S1474-4422(17)30470-2.

MRI can greatly aid the diagnosis through the demonstration of characteristic T2 hyperintensities (see MRI below) but does not supersede the clinical diagnosis. MRI is not always specific for the diagnosis of MS and may lead to a false-positive diagnosis if it is relied on exclusively.

Laboratory support for the diagnosis of MS consists of demonstration in the cerebrospinal fluid (CSF) of at least two oligoclonal bands or increased CNS synthesis of immunoglobulin G (IgG). In addition to this evidence, evoked potentials, particularly visual or somatosensory evoked potentials, are slowed or absent and may contribute to or support the clinical diagnosis.

Course and Etiology

MS can have a widely variable clinical course. The most common course or type, defining 85% to 90% of MS cases, is relapsing–remitting, where the patient has acute relapses of neurologic/neuropsychiatric symptoms lasting at least 24 hours, but usually several days to weeks, with eventual remission and almost complete recovery. A secondary progressive type may evolve from the relapsing–remitting type, leading to incomplete recovery after acute exacerbations of illness. In about 10% of cases, signs and symptoms may persist, worsen, and expand from the outset of the illness, and the disorder then takes on a chronic progressive nature known as the primary progressive type. This form may be rapidly progressive and is rarely seen in patients younger than 40 years of age.

The relationship between these different forms is not clearly understood, and it is uncertain to what extent they represent different types of illness with regard to etiology. The etiology of MS is becoming more fully understood as an autoimmune disorder leading to damage to the myelin with an inflammatory reaction, the resolution of which leads to the formation of plaque in various parts of the brain. Recent research has revealed the presence of peripheral lymphatic ducts in the CNS, which was previously thought to be "immune privileged," and will likely lead to a new understanding of the pathophysiology of MS and other immune system–related neurologic disorders.

Neuropsychiatric Aspects

Fatigue. Various neuropsychiatric syndromes are overrepresented in MS (Table 2.7–3). Fatigue is one of the most common presenting symptoms of patients with all demyelinating disorders and has been extensively described in MS. Fatigue substantially contributes to the disability and poor quality of life that these patients experience. Fatigue may be profound and usually includes both physical fatigue, which makes the patient feel less able to perform physical activity, as well as mental fatigue, which makes the patient unable to participate in their usual activities, including work. Fatigue is one of the most prominent and enduring symptoms of MS, and it is estimated to occur in almost all patients with MS and to be the predominant and often the presenting symptom at the time of diagnosis in up to 50% of cases. Fatigue in MS can have a reverse diurnal rhythm in that it is worse as the day progresses, so many patients with MS alter their lifestyle to do much of their activities early in the day before the fatigue of the later day and evening becomes disabling.

Table 2.7–3.
Neuropsychiatric Syndromes Associated with MS, Rates, Mechanisms, and Treatment

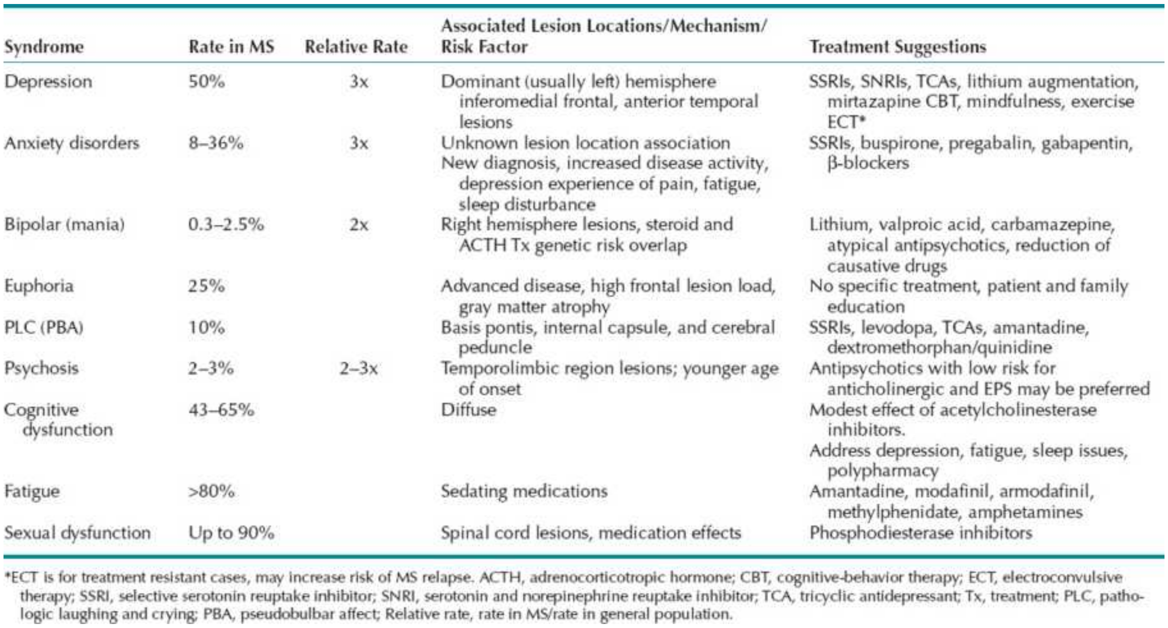

Syndrome	Rate in MS	Relative Rate	Associated Lesion Locations/Mechanism/ Risk Factor	Treatment Suggestions
Depression	50%	3x	Dominant (usually left) hemisphere inferomedial frontal, anterior temporal lesions	SSRIs, SNRIs, TCAs, lithium augmentation, mirtazapine CBT, mindfulness, exercise ECT*
Anxiety disorders	8–36%	3x	Unknown lesion location association New diagnosis, increased disease activity, depression experience of pain, fatigue, sleep disturbance	SSRIs, buspirone, pregabalin, gabapentin, β-blockers
Bipolar (mania)	0.3–2.5%	2x	Right hemisphere lesions, steroid and ACTH Tx genetic risk overlap	Lithium, valproic acid, carbamazepine, atypical antipsychotics, reduction of causative drugs
Euphoria	25%		Advanced disease, high frontal lesion load, gray matter atrophy	No specific treatment, patient and family education
PLC (PBA)	10%		Basis pontis, internal capsule, and cerebral peduncle	SSRIs, levodopa, TCAs, amantadine, dextromethorphan/quinidine
Psychosis	2–3%	2–3x	Temporolimbic region lesions; younger age of onset	Antipsychotics with low risk for anticholinergic and EPS may be preferred
Cognitive dysfunction	43–65%		Diffuse	Modest effect of acetylcholinesterase inhibitors. Address depression, fatigue, sleep issues, polypharmacy
Fatigue	>80%		Sedating medications	Amantadine, modafinil, armodafinil, methylphenidate, amphetamines
Sexual dysfunction	Up to 90%		Spinal cord lesions, medication effects	Phosphodiesterase inhibitors

*ECT is for treatment resistant cases, may increase risk of MS relapse. ACTH, adrenocorticotropic hormone; CBT, cognitive-behavior therapy; ECT, electroconvulsive therapy; SSRI, selective serotonin reuptake inhibitor; SNRI, serotonin and norepinephrine reuptake inhibitor; TCA, tricyclic antidepressant; Tx, treatment; PLC, pathologic laughing and crying; PBA, pseudobulbar affect; Relative rate, rate in MS/rate in general population.

The etiology of fatigue in MS is complex and multidetermined and remains an area of active research. There are no clear etiologic factors identified, but neurologic and endocrine factors may contribute. Other factors may also worsen fatigue, such as comorbid medical disorders, associated depression, sleep difficulties, and the side effects of medications used to treat the symptoms of MS and comorbid conditions (Table 2.7–4). In particular, fatigue may be related to comorbid depression, where fatigue is also a common symptom and treatment of depression may, to a varying extent, alleviate the symptom of fatigue.

Depression and Anxiety. Of the broad range of neuropsychiatric symptoms associated with MS (see Table 2.7–3), depression is the most prominent and common but remains underrecognized and underdiagnosed or misdiagnosed as a separate entity rather than a result of the underlying pathology (demyelination) disrupting mood regulation networks and circuits. Over the course of their illness, approximately half of patients with MS will have at least one episode of signs and symptoms meeting the diagnostic criteria for major depressive disorder. Point prevalence rates may be as high as 20%. Depression can be the presenting feature of MS and is often the dominant complaint of MS patients. Neurovegetative symptoms, frustration intolerance, discouragement, and irritability may more likely characterize the depression of MS than depression in the general population.

Table 2.7–4.
Potential of MS Therapeutics Neuropsychiatric Benefits and Adverse Effects

Therapeutic Agent	Potential Benefits	Potential Adverse Effects
Corticosteroids adrenocorticotropic hormone	Reduced flair symptoms	Mania, insomnia, psychosis, depression, hyperglycemia, weight gain, osteoperosis
Immunomodulatory drugs ▸ Interferon β-1a ▸ Interferon β-1b ▸ Glatiramer ▸ Monoclonal antibodies (natalizumab, alemtuzumab, daclizumab, ocrelizumab, rituzimab*)	Improved cognition, information processing, memory, visuospatial ability, and executive function. Reduced relapse rate and lower white matter disease burden	Fatigue, depression, infusion reactions
Chemotherapeutics (mitoxantrone)	Reduced disease activity	Fatigue, infection, nausea
Small-molecule oral agents (Fingolimod, dimethyl fumarate, teriflunomide)	Reduced disease activity	Seizures (rare), infections, progressive multifocal leukoencephalopathy, malignancies
Dalfampridine	Gait improvement	Anxiety, focal seizure
Baclofen, dantrolene, tizanidine	Reduced spasticity	(Hypo)mania

*Off-label.

Data from Filley CM. *The Behavioral Neurology of White Matter.* 2nd ed. Oxford University Press. 2012; Reich DS, Lucchinetti CF, Calabresi PA. Multiple sclerosis. *N Engl J Med.* 2018;378(2):169–180; and Tobin WO. Management of multiple sclerosis symptoms and comorbidities. *Continuum.* 2019;25 (3, multiple sclerosis and other CNS inflammatory diseases): 753–772.

As described by G.M. Galeazzi and colleagues in a review, the cause of depression in MS patients may be due to (1) depression as a prodromal feature of MS, (2) depression and MS as coincidental but independent conditions, (3) depression and MS as the same condition with common etiology, (4) depression as a reaction to the chronicity and disability of MS, (5) depression as a side effect of MS treatment, particularly corticosteroids and interferon, or (6) any combination of these.

Table 2.7–5.
Reported MS White Matter Lesion Location and Associated Focal Neurobehavioral Syndrome

Bifrontal	Executive dysfunction
Left frontal	Broca aphasia, transcortical motor aphasia
Left temporal	Amnesia
Left temporoparietal	Alexia with agraphia
Left occipital, splenium of corpus callosum	Pure alexia
Left arcuate fasciculus	Conduction aphasia
Left periventricular	Global aphasia
Bilateral occipitotemporal	Visual agnosia
Right hemispheric	Left hemineglect
Corpus callosum an bihemispheric	Left tactile anomia, agraphia, and apraxia
Diffuse	Neurocognitive disorder (dementia), mixed transcortical aphasia

Adapted with permission from Filley, CM. *The Behavioral Neurology of White Matter*. 2nd ed. Oxford University Press; 2012.

Table 2.7–5 presents the associated lesion location for each major neurobehavioral syndrome associated with MS. Based on multiple neuroimaging and other epidemiologic studies, it is generally agreed that dominant hemisphere (usually left), inferior medial frontal, and anterior temporal MS plaque burden is linked with depression in MS. Indeed, lesions in these areas accounted for 42% of depression variance in a regression analysis. Likewise, other demyelinating, vascular, traumatic, infectious, inflammatory, toxic, metabolic, and neoplastic white matter disorders involving these networks frequently cause depression. Smaller hippocampal volumes and associated cortisol hypersecretion, and a possible role for cytokine regulation, have also been related to depression in MS and suggest a more complex neuroendocrine-limbic etiology.

The occurrence of major depressive syndrome in MS patients may substantially add to the disability and the burden of illness associated with MS and dramatically increases the risk for suicide to seven times that of normal populations. Suicide, tragically, is a frequent cause of death for MS patients that can be prevented with recognition and treatment.

The research to date on the antidepressant treatment of depressive syndromes associated with MS is limited, and the efficacy of antidepressants has not been clearly demonstrated. In a Cochrane review, only two studies on pharmacologic treatment of depression in MS met quality inclusion criteria and showed only a trend toward efficacy for desipramine and paroxetine but with the expected adverse effects common with these agents limiting their usefulness. More tolerable agents, including fluoxetine, sertraline, and duloxetine, were shown to be beneficial in open-label studies.

Randomized controlled trials have shown a significant benefit of cognitive-behavior therapy (CBT) for depression in MS. Mindfulness training has also shown benefits. Expert recommendations for treatment of depression in MS are similar to those for the general population and include antidepressants, CBT, mindfulness training, and lifestyle and social interventions, as well as electroconvulsive therapy (ECT) in treatment-refractory cases, though it is cautioned that ECT may exacerbate active disease. Psychoeducation for patients and their families on the prevalence and cause of depression in MS may be especially helpful.

Anxiety in MS remains poorly studied. Anxiety is frequently comorbid with depression in the general population and in MS; clinically significant anxiety is present in about half of MS patients with depression and is correlated to thoughts of self-harm, somatic complaints, and social dysfunction. Pathologic anxiety disorders, including generalized anxiety disorder, panic disorder, and obsessive-compulsive disorder, are roughly three times more common in MS patients than in the general population. Social phobia is found in about 8% of MS patients and is thus less common than in the general population (about 13%). Guidelines for treating anxiety disorders in MS are lacking, and thus the clinician must rely on clinical judgment and guidelines for treating anxiety in the general population. SSRIs such as escitalopram, fluoxetine, and sertraline are generally preferred.

Mania and Bipolar Disorder. Mania-like episodes or the full syndrome of bipolar mood disorder can be the presenting and predominant feature of MS. Bipolar disorder has been estimated to be 2 to 10 times more frequent in MS patients than in the general population. Feinstein associated mania with bilateral temporal lesions. In general, neurologic disorders resulting in right hemisphere lesions are associated with mania, and lesions within the right anterior temporal, orbitofrontal, and cingulate regions may be more highly correlated with mania. In bipolar patients with a rapid-cycling course, the possibility of MS as a course-modifying factor for bipolar disorder should be considered. Mania secondary to corticosteroid treatment in MS patients is also a relevant concern affecting approximately 30% of MS patients.

Due to the dearth of research on the treatment of mania in MS, treatment recommendations follow those for the general population. Of course, corticosteroid or adrenocorticotropic hormone treatments may require special consideration for discontinuation in this context.

Euphoria, Disinhibition, and Other Frontal Personality Changes. Euphoria, a persistent state of impaired insight and judgment with apparent unconcern and even elation regarding disability, is a striking syndrome associated predominantly with MS, with a prevalence of about 10% in the disease. The presence of euphoria is generally indicative of more advanced disease with more disseminated plaque burden and atrophy and is associated with more impaired neurologic and cognitive functioning and social disability. The persistent cheerfulness and unconcern of euphoria may overlap somewhat with mania but is distinct from the full mania syndrome and does not represent a “climate shift” in the mood nor fluctuate, and may even co-occur with depression.

Cortical or subcortical gray or white matter lesions from MS or other causes can result in the other classic frontal lobe syndromes, including dysexecutive, disinhibited, or apathetic if the lesion burden is in the associated dorsolateral–prefrontal, orbitofrontal, or medial frontal/cingulate systems, respectively.

A 41-year-old, right-handed, married male dentist with a family history of bipolar disorder and a past psychiatric history of an isolated episode of depression 3 years prior presented for voluntary inpatient hospitalization at the recommendation of his outpatient psychiatrist for treatment of a first-ever manic episode beginning 4 months prior to presentation and in the context of the onset of use of THC pills. He denied the use of nitrous oxide or other inhalants. On evaluation, he exhibited an elevated and expansive mood with notable euphoria, decreased need for sleep, grandiosity, and pressured speech with flight of ideas, but with notable retained insight regarding the abnormality of his current state. He was empirically started on lithium and Seroquel, and his manic symptoms gradually improved. MRI of the brain (Fig. 2.7–1) was ordered as part of the workup and revealed numerous T2 intense intracranial white matter lesions in the posterior fossa and supratentorial brain, including corpus callosum, and a newer-appearing lesion in the white matter deep to the right cingulate gyrus, all consistent with the diagnosis of demyelinating disease, likely MS. A follow-up MRI of brain and spinal cord with contrast revealed additional T2 hyperintense lesions consistent with demyelination in the cervical cord at the C2 level. No enhancement suggesting active inflammation was seen with contrast administration. A more detailed history was obtained, and the patient recalled when he was in college, he had an acute episode of weakness and paresthesia ascending from his feet and ultimately involving most of his body, which was diagnosed at the time as peripheral neuropathy. This episode resolved spontaneously after a couple of weeks. With this history and the imaging findings, the patient was referred to a neurologist who specialized in MS and was diagnosed with MS based on the history of the prior clinical attack and the objective evidence of corresponding two T2 lesions in the dossal cord at C2–C3 as well as other T2 disseminated in time and space. The right frontal lesion was hypothesized to be potentially causative of his manic symptoms, and the older appearing left hemisphere lesions could have contributed to his prior depressive syndrome, but a definitive causal relationship cannot be asserted. He was started on Copaxone as this was considered the DMA least likely to affect mood, and it was recommended that he avoid cannabinoids and continue treatment with lithium. He was referred to an MS clinic at a university hospital for further evaluation.

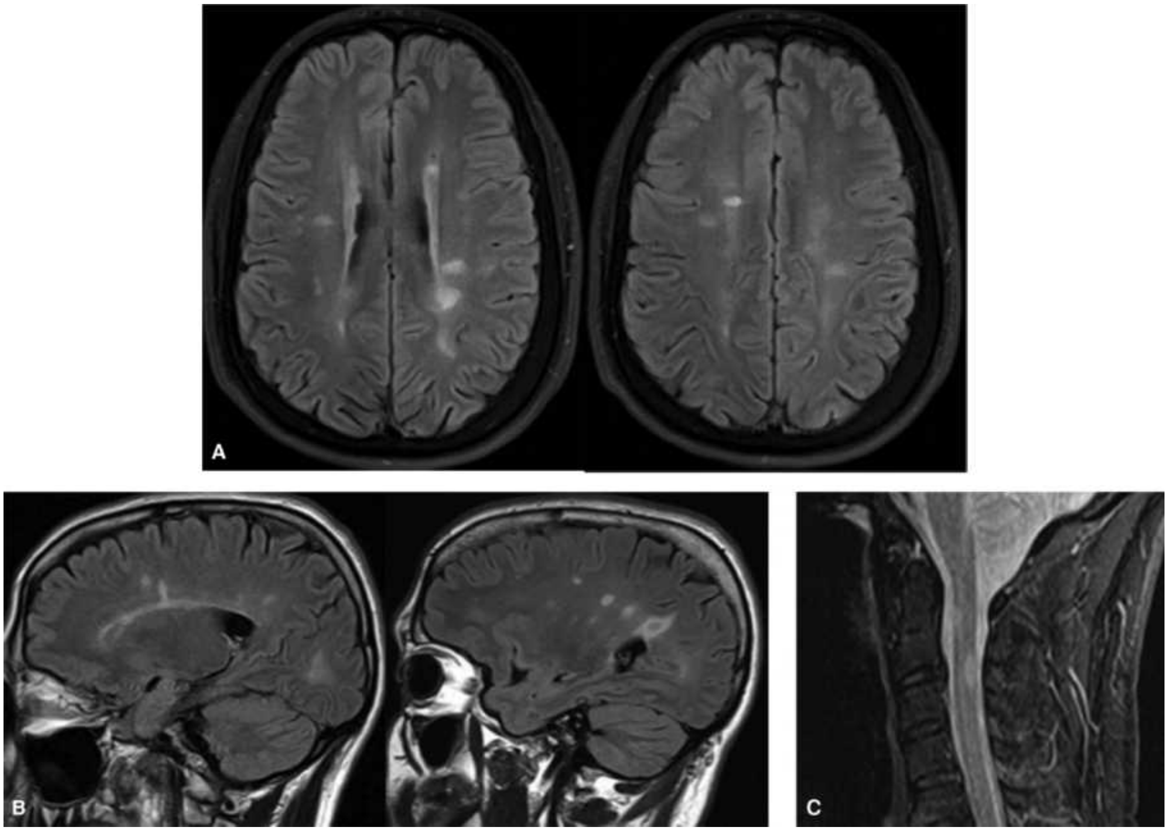

FIGURE 2.7–1. MRI for case presentation. Axial (**A**) and sagittal (**B**) T2 FLAIR MRI of the brain revealing numerous T2 intense intracranial white matter lesions in the posterior fossa and supratentorial brain consistent with the diagnosis of demyelinating disease, likely multiple sclerosis. Sagittal images (**B**) show the relatively MS-specific “Dawson fingers”—periventricular T2 hyperintensities (indicative of demyelinating plaques) perpendicular to the body of the lateral ventricles (along the axis of medullary veins). **C:** Sagittal T2 MRI spinal cord revealing additional T2 hyperintense lesions consistent with demyelination in the cervical cord at the C2 level.

Affect Disorders. As distinct from the more persistent abnormalities of mood discussed above, MS and other white matter disorders can cause disordered affect (defined as the experienced and expressed transient emotional state; see DSM-5 Glossary of Clinical Terms). This includes emotional lability, pathologic laughing, and crying (PLC; also known as pseudobulbar affect, emotional incontinence, and other descriptors). PLC in MS is associated with a heavier lesion burden, usually in the basis pontis, internal capsule, and cerebral peduncle.

Psychosis. The syndrome of psychosis in MS is more frequent than in the general population, occurring in up to 5% of cases, and maybe the initial presenting feature of the disease. Psychosis is a more common presenting feature of other white matter disorders, including metachromatic leukodystrophy. In MS patients, psychosis has been associated with earlier onset of disease and plaque burden in bilateral temporolimbic regions, especially the left temporal horn and left trigone. Persecutory delusions are more common than auditory or visual hallucinations, with incidences of about 70%, 20%, and 20%, respectively. In general, treatment with an antipsychotic with a low risk for anticholinergic and extrapyramidal side effects is preferred.

Sleep Disorders. Various disturbances of sleep/wake are associated with MS, including insomnia, sleep-disordered breathing, restless legs syndrome (RLS), rapid eye movement (REM) sleep behavior disorder, and narcolepsy, and can have a significant impact on fatigue and general health-related quality of life, as well as on cognitive impairment, depression, anxiety, and other neuropsychiatric symptoms.

Cognitive Impairment. It has been estimated that up to 50% to 70% of patients with MS show evidence of cognitive impairment on an appropriate battery of neuropsychological tests. However, unlike the typical pattern of amnesia and aphasia seen in Alzheimer disease, for example, impairments in processing speed and attention are more likely in MS and may not be assessed adequately by cognitive screening instruments. Cognitive symptoms contribute substantially to the dysfunction and to poor quality of life that occurs in patients with MS. Cognitive dysfunction in MS may be broad based and involve multiple neurocognitive domains, and in nearly a quarter of patients, the severity of deficit may surpass the threshold for diagnosis of a neurocognitive disorder (dementia).

In general, white matter dementias negatively affect the speed or efficiency of a given neurocognitive domain rather than cause complete loss of that domain. For example, all aspects of memory function, especially memory retrieval, may be slowed, but procedural memory is generally normal, and the patient would not usually have dense amnesia as seen in cortical dementias. Higher-order functions, such as complex attention, information processing, and executive function, may likely be impaired, and these may contribute substantially to impairments at work, driving, and in relationships and limit the ability to benefit from physical rehabilitation. However, overall intellectual abilities are spared, as are language abilities (aphasias are uncommon), although some dysfunction in these domains may be determined on broad-based neuropsychological tests.

Most of the studies of cognitive impairment in MS are cross-sectional in nature. There are limited data on the longitudinal course of cognitive difficulty. In general, however, these studies show a gradual decline in all aspects of cognitive function in the patient with MS.

Full neuropsychological testing is the gold standard for cognitive assessment but is time-consuming and may not be available in some settings. The Minimal Assessment of Cognitive Function in MS (MACFIMS) is a relatively brief (90-minute) validated battery assessing domains usually affected in MS and is commonly utilized in research and in MS clinics for clinical monitoring.

Cerebellar Involvement. Cerebellar white matter is commonly involved in MS, 85% of cases in one study, and is associated with various kinds of dysfunction, including problems with motor coordination as well as problems with regulation of emotion, cognition, and behavior, depending on the cerebellar region involved with the cerebellar motor syndrome being associated with anterior lobe lesions and the cerebellar cognitive affective syndrome (CCAS) being associated with posterior lobe lesions. In MS, the disease burden in the middle cerebellar peduncle (MCP) is most highly correlated with cognitive dysfunction.

Differential Diagnosis

MS presents a substantial challenge as far as differential diagnosis is concerned. In those with the common relapsing-remitting course of illness, together with typical clinical, laboratory, and MRI findings, the diagnosis is more easily made. However, in atypical cases, differential diagnosis can be difficult. This applies particularly at the onset of the disease, during the first episode when the conventional clinical criteria are hard to apply. With the first attack of MS, it is difficult to differentiate this illness from other demyelinating disorders.

There are a variety of other conditions to consider in the differential diagnosis of MS. These include the following:

1. Disseminated encephalomyelitis: Factors that may point to this diagnosis include systemic symptoms, signs of cortical dysfunction, and, perhaps, bilateral optic neuritis. MRI may be quite characteristic of this disorder and helps distinguish it from MS.
2. Cerebral vasculitis: This has numerous causes. They include primary cerebral vasculitis, which is extremely rare, associated with systemic lupus erythematosus, AIDS, lymphoproliferative disorders, and Wegener granulomatosis. Various ancillary laboratory findings may help distinguish these various forms of the disease. Rarer forms of vasculitis may include Behçet disease and also CNS sarcoidosis.
3. Leukodystrophies: These disorders are characterized by the accumulation of unchanged adrenoleukodystrophy, which may have an adult-onset form. These conditions are rare and can usually be identified by the presence of family history.
4. Chronic fatigue syndrome: It has been previously emphasized that fatigue is one of the most predominant symptoms of MS. Chronic fatigue syndrome, also known as fibromyalgia, requires differentiation, especially if no focal neurologic symptoms are found. Failure to elicit specific neurologic symptoms, as well as the presence of other associated features of fibromyalgia, including sleep disturbance, myalgias, and arthralgias, among other relatively nonspecific symptoms, is supportive of this diagnosis.

Treatment

Treatment of the clinical neuropsychiatric syndromes discussed above is generally initiated empirically in a symptom-targeted approach while workup for a potential identifiable cause is conducted. Care should be taken to educate the patient and family on the potential causal relationship between the MS flair and neuropsychiatric symptoms. Clinicians should avoid the erroneous description of “inorganic” or assume that these symptoms are a manifestation of a psychological reaction to the illness.

CONCLUSION AND FUTURE DIRECTIONS

While the demyelinating disorders present etiopathologic and diagnostic challenges, it is clear that they can produce phenomenology across the spectrum of neuropsychiatric disorders. The most common demyelinating disorder, MS, requires differentiation from other disorders because of its distinct clinical presentation and course of illness, as well as specific treatments that may reduce relapse rates and degree of dysfunction. MS usually presents with neuropsychiatric symptoms, and depressive syndromes and symptoms are especially common. While the broad range of pharmacologic and nonpharmacologic interventions used in primary psychiatric disorders are often used, with good effect, in MS patients, large-scale, well-designed studies are needed to provide a good evidence base for the treatment of psychiatric symptoms and disorders associated with this condition, which may require specific intervention or psychotropic agents. The association of MS with specific major histocompatibility complex (MHC) class II alleles and the relatively recent discovery of peripheral lymphatic duct in the CNS, previously thought to be immune privileged, may open new avenues for research on these disorders.

ACKNOWLEDGMENT

This chapter represents an update and revision of previous chapters in this series, most recently by Dr Russell T. Joffe from the last edition. I am indebted to Dr Joffe for his work and retain much of his prose in this current edition.

Further Readings

Arnett PA, Barwick FH, Beeney JE. Depression in multiple sclerosis: review and theoretical proposal. *J Int Neuropsychol Soc*. 2008;14(5):691–724.

Benedict RHB, Weinstock-Guttman B, Fishman I, Sharma J, Tjoa CW, Bakshi R. Prediction of neuropsychological impairment in multiple sclerosis: comparison of conventional magnetic resonance imaging measures of atrophy and lesion burden. *Arch Neurol*. 2004;61(2):226–230.

Brown RF, Valpiani EM, Tennant CC, et al. Longitudinal assessment of anxiety, depression, and fatigue in people with multiple sclerosis. *Psychol Psychother*. 2008;82(Pt 1):41–56.

Brownlee WJ, Hardy TA, Fazekas F, Miller DH. Diagnosis of multiple sclerosis: progress and challenges. *Lancet*. 2017;389(10076):1336–1346.

Filley CM. *The Behavioral Neurology of White Matter*. 2nd ed. Oxford University Press; 2012.

Filley CM. White matter: beyond focal disconnection. *Neurol Clin*. 2011;29(1):81–97.

Costei C, Barbarosie M, Bernard G, Brais B, La Piana R. Adult hereditary white matter diseases with psychiatric presentation: clinical pointers and MRI algorithm to guide the diagnostic process. *J Neuropsychiatry Clin Neurosci*. 2021;33(3):180–193.

Dietz AG, Goldman SA, Nedergaard M. Glial cells in schizophrenia: a unified hypothesis. *Lancet Psychiatry* 2020;7(3):272–281.

Feinstein A. *The Clinical Neuropsychiatry of Multiple Sclerosis*, 2nd ed. Cambridge University Press; 2007.

Feinstein A, Magalhaes S, Richard JF, Audet B, Moore C. The link between multiple sclerosis and depression. *Nat Rev Neurol*. 2014;10(9):507–517.

Figved N, Benedict R, Klevan G, et al. Relationship of cognitive impairment to psychiatric symptoms in multiple sclerosis. *Mult Scler*. 2008;14(8):1084–1090.

Galeazzi GM, Ferrari S, Giaroli G, et al. Psychiatric disorders and depression in multiple sclerosis outpatients: impact of disability and interferon beta therapy. *Neurol Sci*. 2005;26(4):255–262.

Ghaffar O, Feinstein A. The neuropsychiatry of multiple sclerosis: a review of recent developments. *Curr Opin Psychiatry*. 2007;20(3):278–285.

Greim B, Engel C, Apel A, Zettl UK. Fatigue in neuroimmunological diseases. *J Neurol*. 2007;254(Suppl 2):II102–II106.

Hartung HP, Graf J, Aktas O, Mares J, Barnett MH. Diagnosis of multiple sclerosis: revisions of the McDonald criteria 2017 – continuity and change. *Curr Opin Neurol*. 2019;32(3):327–337.

Hoffmann S, Tittgemeyer M, von Cramon DY. Cognitive impairment in multiple sclerosis. *Curr Opin Neurol*. 2007;20(3):275–280.

Janssens ACJW, van Doorn PA, de Boer JB, et al. Anxiety and depression influence the relation between disability status and quality of life in multiple sclerosis. *Mult Scler*. 2003;9(4):397–403.

Koch M, Uyttenboogaart M, van Harten A, Heerings M, De Keyser J. Fatigue, depression and progression in multiple sclerosis. *Mult Scler*. 2008;14(6):815–822.

Koch MW, Glazenborg A, Uyttenboogaart M, Mostert J, De Keyser J. Pharmacologic treatment of depression in multiple sclerosis. *Cochrane Database Syst Rev*. 2011;16(2):CD007295.

Korostil M, Feinstein A. Anxiety disorders and their clinical correlates in multiple sclerosis patients. *Mult Scler*. 2007;13(1):67–72.

Louveau A, Smirnov I, Keyes TJ, et al. Structural and functional features of central nervous system lymphatic vessels. *Nature*. 2015;523(7560):337–341.

Minden SL, Feinstein A, Kalb RC, et al; Guideline Development Subcommittee of the American Academy of Neurology. Evidence-based guideline: assessment and management of psychiatric disorders in individuals with MS: report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology*. 2014;82(2):174–181.

Reich DS, Lucchinetti CF, Calabresi PA. Multiple sclerosis. *N Engl J Med*. 2018;378(2):169–180.

Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*. 2018;17(2):162–173.

Tobin WO. Management of multiple sclerosis symptoms and comorbidities. *Continuum (Minneap Minn)*. 2019;25(3):753–772.

Walton C, King R, Rechtman L, et al. Rising prevalence of multiple sclerosis worldwide: insights from the Atlas of MS, third edition. *Mult Scler*. 2020;26(14):1816–1821.

Zephir H, De Seze J, Stojkovic T, et al. Multiple sclerosis and depression: influence of interferon β therapy. *Mult Scler*. 2003;9(3):284–288.

Zorzon M, de Masi R, Nasuelli D, et al. Depression and anxiety in multiple sclerosis. A clinical and MRI study in 95 subjects. *J Neurol*. 2001;248(5):416–421.

2.8 Neuropsychiatric Aspects of HIV Infection and AIDS

GLENN JORDAN TREISMAN, M.D., PH.D., ANDREW F. ANGELINO, M.D., JEFFREY HSU, M.D., ANNA MOUKOURI KOUOH, C.R.N.P., AND NICHOLAS P. SCHWEIZER, ED.D., L.C.A.D.C., L.C.P.C.

Since the original description of acquired immune deficiency syndrome (AIDS) in 1981, the human immunodeficiency virus (HIV) epidemic has remained a public health concern due to continued infections, comorbid conditions, and specialty treatment. In the mid-1990s, the discovery of combined antiviral therapy (ART) radically changed the course of the illness from an inevitably fatal illness to a chronic treatable condition. This provoked optimism and spurred research in a way that has transformed the field of medicine. The virus has now become a chronic disease for most patients, and effective multidrug therapy (initially referred to as HAART—**h**ighly **a**ctive **a**nti**r**etroviral **t**herapy but now usually just ART for antiviral therapy), inhibits viral replication so that it is undetectable in the blood of most effectively treated patients. Over the years, alterations in treatment options, prophylactic treatment of opportunistic infections, prevention of maternal–fetal transmission and the advent of pre-exposure prophylaxis (PrEP) have all contributed to improving the quality of life and prognosis of persons living with HIV (PLWH), and made it so that this patient population has a life expectancy near that of the general population when diagnosed early and treated effectively.

BOX 2.8–1.

Acquired Immune Deficient Syndrome (AIDS)-Defining Conditions

Bacterial infections, multiple or recurrent (only among children aged <6 years old)
Candidiasis of bronchi, trachea, or lungs
Candidiasis of esophagus
Cervical cancer, invasive (only among children aged ≥6 years old, adolescents, and adults)
Coccidioidomycosis, disseminated or extrapulmonary
Cryptococcosis, extrapulmonary
Cryptosporidiosis, chronic intestinal (>1 month's duration)
Cytomegalovirus disease (other than liver, spleen, or nodes), onset at age >1 month
Cytomegalovirus retinitis (with loss of vision)
Encephalopathy, HIV related
Herpes simplex: chronic ulcers (>1 month's duration) or bronchitis, pneumonitis, or esophagitis (onset at age >1 month)
Histoplasmosis, disseminated or extrapulmonary
Isosporiasis, chronic intestinal (>1 month's duration)
Kaposi sarcoma
Lymphoid interstitial pneumonia or pulmonary lymphoid hyperplasia complex
Lymphoma, Burkitt (or equivalent term)
Lymphoma, immunoblastic (or equivalent term)
Lymphoma, primary, of brain
Mycobacterium avium complex or Mycobacterium kansasii, disseminated or extrapulmonary
Mycobacterium tuberculosis of any site, pulmonary, disseminated, or extrapulmonary
Mycobacterium, other species or unidentified species, disseminated or extrapulmonary
Pneumocystis jirovecii pneumonia
Pneumonia, recurrent
Progressive multifocal leukoencephalopathy
Salmonella septicemia, recurrent
Toxoplasmosis of brain, onset at age >1 month
Wasting syndrome attributed to HIV

Psychiatric disorders play several roles in the transmission of HIV and the continued epidemic, in elements of treatment, and in comorbidity. Psychiatric conditions are associated with increased risk behaviors for infection, decreased access to care, decreased adherence to HIV therapies, and increased medical comorbidity. Because of this, some authors have gone as far as saying that HIV is a psychiatric epidemic.

Prevention of HIV transmission became a goal in the mid-1980s when the routes of transmission were defined and testing became available. The massive effort at education resulted in a decrease in the spread of HIV in people able to modify their risky behaviors, but vulnerable people who were either unable or unwilling to modify their behavior or unable to understand the risks continued to spread the infection. While there are still individuals who contract HIV as a result of ignorance and bad luck, the vast majority of new infections arise in individuals who are aware of HIV but continue to engage in high-risk behaviors. As HIV treatment has become accessible and with more effective and better-tolerated antiviral cocktails, some people have become less cautious and there has been a resurgence of risky behaviors in some populations. A major breakthrough in prevention came with the development of PrEP, an antiviral combination that prevents HIV infection if taken before exposure. Again, a very effective intervention but one that vulnerable people may not be able to access or effectively utilize.

PrEP is available as either of two oral formulations or a long-acting injectable formulation, is well tolerated, and has the remarkable potential to decrease the HIV epidemic. Unfortunately, patients at risk for HIV often have limited interactions with health care and PrEP has not yet had the impact that it could. Many patients present to mental health settings, particularly those treating substance abuse, with HIV risk behaviors and the potential to benefit from PrEP, and mental health providers could provide access to PrEP for at-risk patients. Newer forms of PrEP are coming and may be even more effective in helping prevent infections.

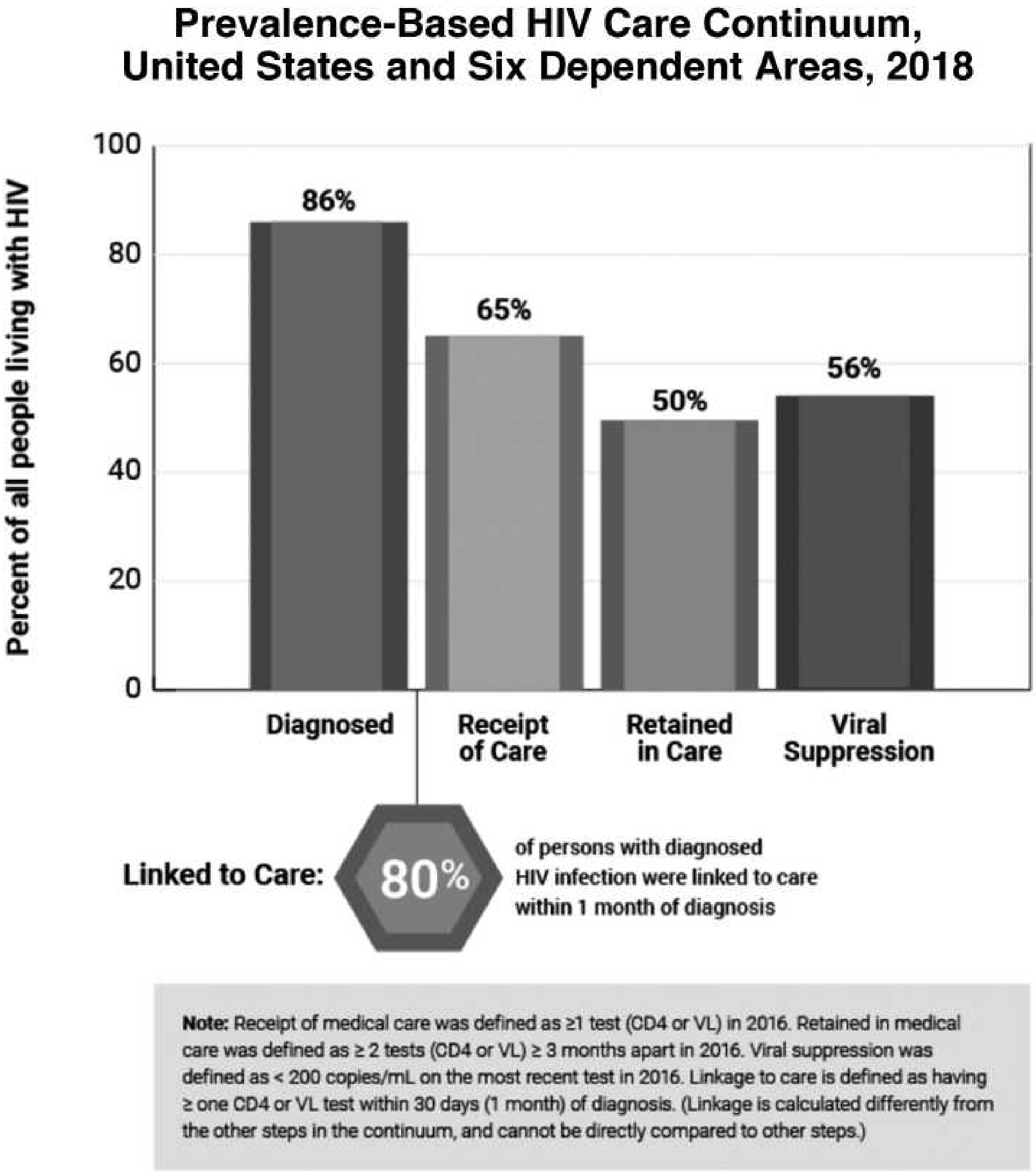

Note: Receipt of medical care was defined as ≥1 test (CD4 or VL) in 2016. Retained in medical care was defined as ≥ 2 tests (CD4 or VL) ≥ 3 months apart in 2016. Viral suppression was defined as < 200 copies/mL on the most recent test in 2016. Linkage to care is defined as having ≥ one CD4 or VL test within 30 days (1 month) of diagnosis. (Linkage is calculated differently from the other steps in the continuum, and cannot be directly compared to other steps.)

FIGURE 2.8–1. Stages of engagement in care. This figure shows the percentage of patients who are receiving various levels of care for their HIV infection. The original publication in 2007 showed that less than 20% of patients had achieved undetectable viral loads. (Reprinted from HIV Care Continuum. https://www.hiv.gov/federal-response/policies-issues/hiv-aids-care-continuum/. Data from CDC, Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2019. HIV Surveillance Supplemental Report 2021; 26(No. 2). Published May 2021.)

Since many of the high-risk behaviors that can result in HIV infection are those that are associated with the brain’s reward systems, it is understandable that humans would be drawn to these behaviors and that they would be common. Sex and substance use are indeed common behaviors. Individuals with various types of untreated mental illness are vulnerable to engage in such behaviors unsafely at a higher frequency, and thus increase their risk of acquiring HIV. Several major efforts have been successful in curbing the epidemic. The concept of “treatment as prevention,” greatly encouraged more aggressive identification and treatment of infected individuals and encouraged clinicians to come up with ways to get patients into treatment. This has resulted in “test and treat” programs to try to accelerate the goal of achieving an undetectable viral load, a condition where patients are minimally infectious to others. Once infected, there is a stepwise hierarchy for accessing care (see Fig. 2.8–1). Based on the description by Eldred and Malitz, it is possible to stratify the degree of engagement in care in various populations. Vulnerable populations (such as those with psychiatric conditions) tend to lag behind in accessing care. Figure 2.8–1 shows the current level of care continuum from the Centers for Disease Control and Prevention (CDC) in 2019.

Once an individual is engaged in care, they must adhere to the antiviral regimen prescribed for them. This has been greatly aided by efforts to simplify treatment, and many regimens have achieved the goal of “one pill once a day.” Untreated mental illness has been shown to lead to nonadherence, higher viral loads, more morbidity, and a higher mortality rate. The continued presence of virus in the blood of those with inadequately controlled HIV infection makes it possible to transmit HIV and perpetuates the epidemic. Recognition and treatment of mental illness is crucial to ending the spread of HIV and implies a critical role for psychiatry in integrated care for these patients.

The introductory part of the chapter is a medical overview of HIV disease. The second part considers psychiatric conditions associated with HIV, and the third part considers related neurologic issues that have an impact on psychiatric function and are part of the differential diagnosis of any patient presenting with HIV and psychiatric complaints.

OVERVIEW OF HIV INFECTION AND AIDS

Epidemiology

HIV was originally recognized through a series of case descriptions involving young homosexual men with *Pneumocystis carinii* (now *Pneumocystis jirovecii*) pneumonia and other rare conditions, including an aggressive form of Kaposi sarcoma, (a usually indolent condition seen prior to that time most often in older Caucasian men of middle eastern extraction), in the early 1980s in Los Angeles (see Fig. 2.8–2 for the first report of the epidemic). These patients suffered from unusual infections associated with immune systems that had been compromised, termed “opportunistic” infections. Subsequently, it became clear that these patients were vulnerable to infections that were usually restricted to immunocompromised individuals.

The World Health Organization (WHO) has case definitions for HIV surveillance and clinical staging based on clinical conditions; patients are sorted into one of four stages with higher stages indicating worse clinical conditions, Stage 1 consists of asymptomatic or with persistent generalized lymphadenopathy of at least two sites for longer than 6 months, and up to stage 4 indicating AIDS-defining illnesses like HIV wasting syndrome, pneumocystis pneumonia (PCP), recurrent severe or radiologic bacterial pneumonia, extrapulmonary tuberculosis, HIV encephalopathy, central nervous system (CNS) toxoplasmosis, chronic (more than 1 month) orolabial herpes simplex infection, esophageal candidiasis, and Kaposi sarcoma. The CDC describes three stages of HIV infection: acute, chronic, and AIDS (see Box 2.8–1). In Stage 1, the acute HIV infection produces large amounts of virus in the blood and flu-like symptoms. During this stage, there are high titers of viral products and patients are most infectious. In Stage 2 called chronic HIV infection, patients are asymptomatic and while the virus is still active it produces lower viral titers and patients are less infectious. In Stage 3, the number of CD4 cells is low and the patient becomes vulnerable to opportunistic infections. This is now defined as having AIDS, similar to the WHO definition.

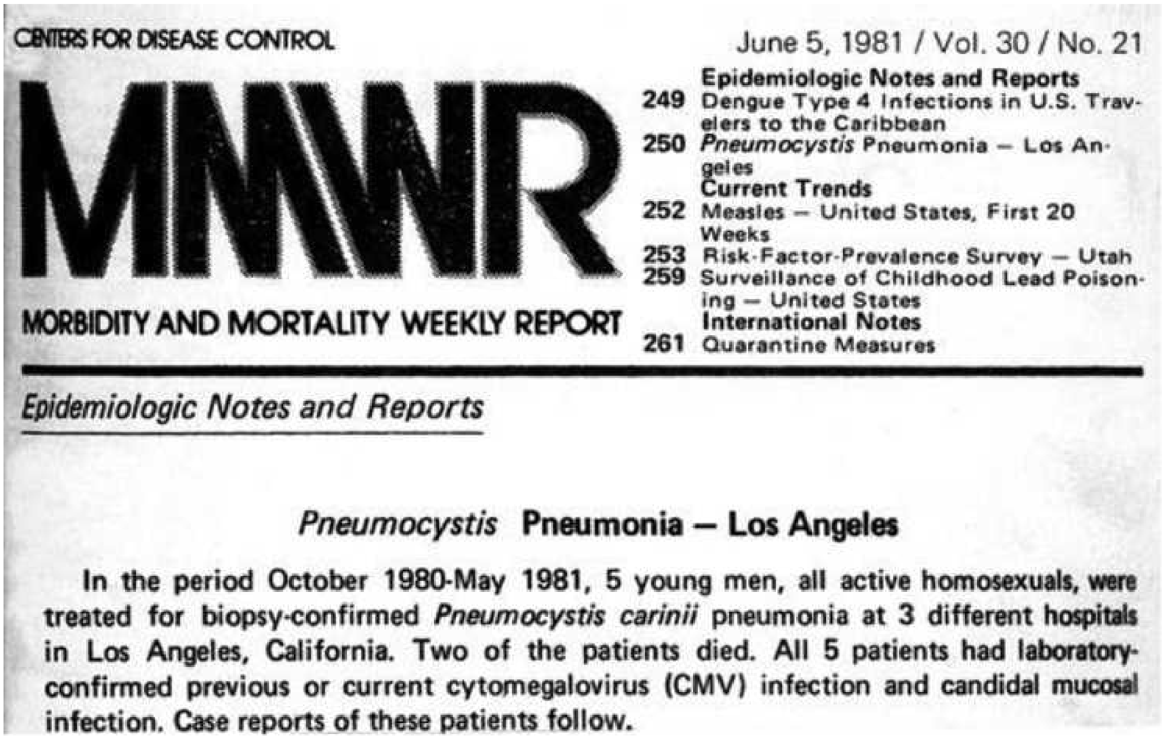

CENTERS FOR DISEASE CONTROL

MMWR

MORBIDITY AND MORTALITY WEEKLY REPORT

June 5, 1981 / Vol. 30 / No. 21

Epidemiologic Notes and Reports
249 Dengue Type 4 Infections in U.S. Travelers to the Caribbean
250 *Pneumocystis* Pneumonia – Los Angeles
Current Trends
252 Measles – United States, First 20 Weeks
253 Risk-Factor-Prevalence Survey – Utah
259 Surveillance of Childhood Lead Poisoning – United States
International Notes
261 Quarantine Measures

Epidemiologic Notes and Reports

***Pneumocystis* Pneumonia – Los Angeles**

In the period October 1980-May 1981, 5 young men, all active homosexuals, were treated for biopsy-confirmed *Pneumocystis carinii* pneumonia at 3 different hospitals in Los Angeles, California. Two of the patients died. All 5 patients had laboratory-confirmed previous or current cytomegalovirus (CMV) infection and candidal mucosal infection. Case reports of these patients follow.

FIGURE 2.8–2. Image from the front page of the CDC weekly morbidity and mortality report showing the first report of five cases of pneumocystis pneumonia reported as part of what became the HIV epidemic. All five also had CMV and candidiasis. (From Centers for Disease Control and Prevention. (2022). HIV and AIDS timeline. https://npin.cdc.gov/pages/hiv-and-aids-timeline#1980)

The 2019 the joint united nations programme on HIV/AIDS (UNAIDS) report suggests that the overall prevalence of HIV globally appears to have stabilized (at 0.7%) but continues to increase in some countries. In 2019, 36.2 million adults and 1.8 million children (<15 years old) were living with HIV/AIDS, 1.5 million adults and 150,000 children were newly infected with HIV, and 600,000 adults and 95,000 children died of AIDS that year. The 2019 incidence of new HIV infections represented a decline of 50% from 2001 when there were 3.4 million new infections. Nearly 60% of the world's population living with HIV are located in sub-Saharan Africa. Children bear a substantial proportion of the burden of HIV with an estimated 70% of the children and infants living with HIV also being born in sub-Saharan Africa. Although HIV/AIDS has been one of the top 10 causes of death since 2010, there has been a decline in the AIDS-related mortality by 39% in the last decade. Currently, the epidemic is escalating in other parts of the world like Southeast Asia and Eastern Europe.

In the United States, there were an estimated 1.2 million people living with HIV in 2019, approximately 14% of whom were not aware of their infection. There were an estimated 36,400 new HIV diagnoses and although this is an overall 7% decrease compared with 2014, there is evidence that cases are rising again over the last 2 to 3 years. Despite the decrease in rates, certain populations are disproportionately affected by HIV, including men who have sex with men (MSM), transgender people, African Americans, and Hispanic/Latin Americans. The highest rates of new infections are accounted for in the MSM population and people who inject drugs (PWID). There has been a great decrease in incidence of mother-to-baby transmission.

Early in the epidemic, transfusion-associated infection represented a significant risk, particularly among those who required pooled or high-frequency transfusions. The transfusion-related risk has now been reduced to less than one in 1.5 million. Currently, the populations at highest risk are MSM, PWID, female partners of IV drug users, and those who trade sex for money or drugs. Mother-to-fetus transmission remains a serious problem in developing countries but has been dramatically curtailed in the United States to less than 2% by the use of antiviral treatment during pregnancy and prophylaxis at parturition.

Efforts at promoting safer sex practices, clean needle use, bleaching of drug paraphernalia, and prenatal HIV testing had an impact on behaviors on a great number of individuals, as evidenced by the flattening of the epidemic curve from an exponential shape to a linear shape. Unfortunately, psychiatric disorders continue to play a significant role in thwarting prevention efforts by making vulnerable patients unable to access or benefit from prevention measures. MSM remain the group with the highest risk of new HIV infection, and psychiatric comorbidity, including addictions, remains one of the major drivers of new infections. Internationally, psychiatric disorders have been identified in both the developed and developing countries as an important comorbidity associated to the risk of infection, lower likelihood of successful treatment, and higher morbidity and mortality.

The retrovirus subfamily to which HIV belongs is called lentivirus (lenti: slow). Early in the HIV epidemic, it was thought that HIV was a “slow virus” that integrated into the cell genome and then was latent until activated much later. In the early 1990s, studies demonstrated very high rates of turnover of CD4-expressing T-lymphocyte cells during the course of HIV infection, with the production of vast quantities of virus. Without treatment, approximately 10 billion viral particles are made on an average day. Untreated HIV infection can be described as a war of attrition in which the virus gradually depletes the immune system. During this fight, the body mobilizes enormous immune system resources, producing up to 2 billion CD4 cells a day. HIV has a relatively high error rate when it is transcribed, and therefore produces large numbers of mutants, thus giving rise to strains of virus that can be resistant to drugs, target different cell receptors, and evade the body’s strategies for defense. More remarkably, the virus has several specific mechanisms of disrupting the immune defenses of the host.

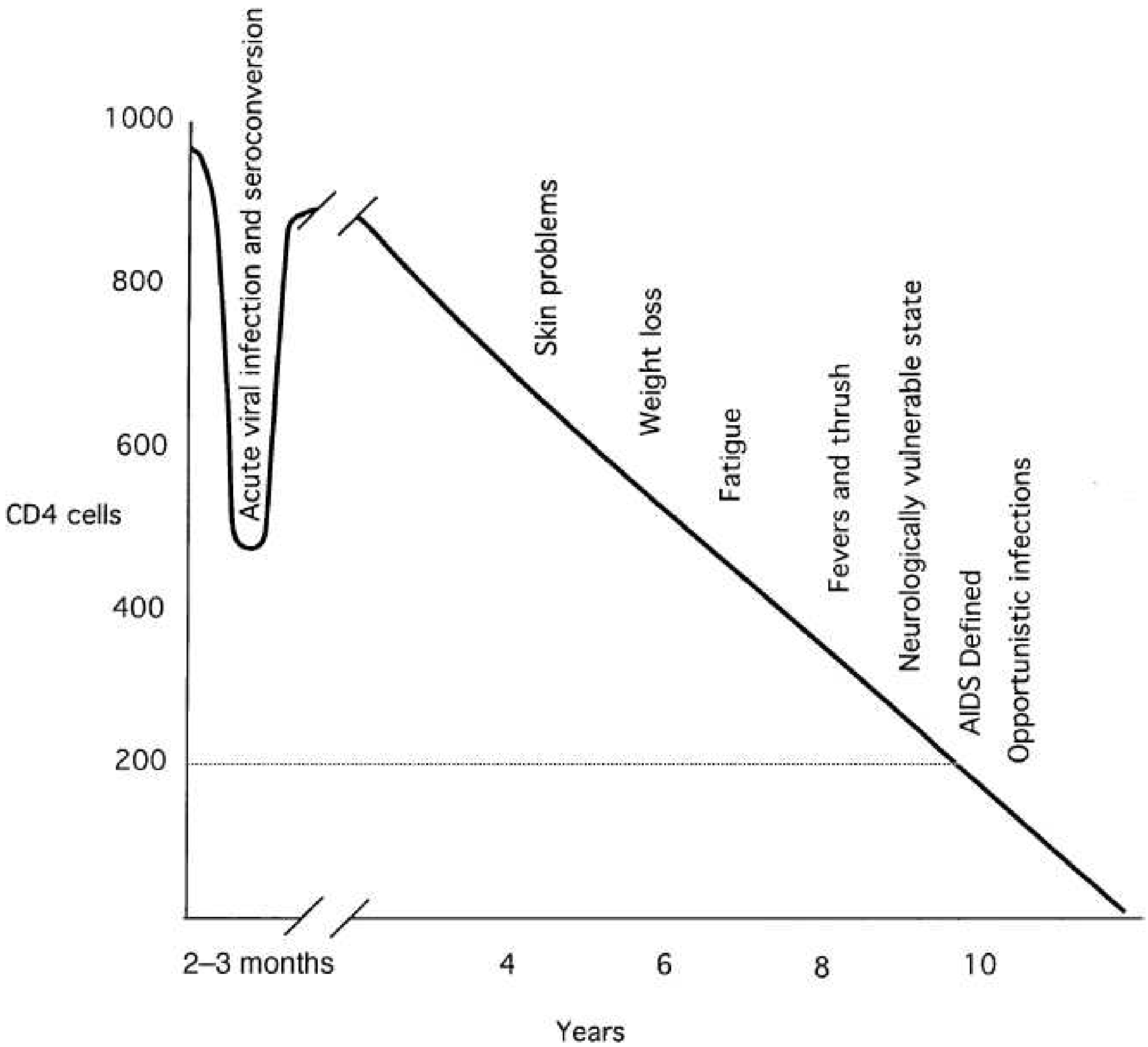

FIGURE 2.8–3. Schematic representation of the course of HIV if left untreated. On the y-axis is the decline in CD4 cell counts and the years of infection are on the x-axis. As noted, after an initial precipitous drop in CD4 cells, the cell count stabilizes and then declines by about 100 per year with death occurring in approximately 10 years. Common clinical issues at each CD4 level are notes on the curve. (Courtesy of John G. Bartlett, M.D.)

HIV infects cells expressing the CD4 surface protein (referred to as the CD4 antigen or receptor), mostly T lymphocytes and macrophages. The main disease mediator is the infection and destruction of T4 helper cells or CD4 cells critical in cell-mediated immunity. The average person has approximately 1,000 CD4 lymphocytes per cubic millimeter (or microliter) of blood. Before the development of effective treatment for HIV, the average patient's CD4 cell count dropped approximately 100 CD4 cells per cubic millimeter (or microliter) per year, so over a 10-year average course, the infected person would lose effective immune function and develop infections that are usually neutralized by cell-mediated immunity (Fig. 2.8–3). The immune system becomes quite compromised as the CD4 cell count drops below 200. At these levels, "cellular" immunity diminishes and is finally lost. At this point, the patient becomes vulnerable to opportunistic infections. There are also several tumors that occur in immunocompromised patients that are also found in patients with AIDS. The term "AIDS" (acquired immunodeficiency syndrome) is used to describe the syndrome in which patients develop these opportunistic medical conditions.

There were several major breakthroughs that changed the overall understanding of the infection: the ability to measure CD4 cells, the ability to measure and quantitate the "viral load" (the quantity of viral particles in the blood), and the ability to detect HIV mutant virus and predict sensitivity to antiviral medications. Currently, certain tests are routinely run at the time of initiation of therapy to make certain that the chosen regimen will be effective. Regular laboratory studies to assess viral load are used to make certain that the virus remains well controlled and that the presence of viral particles in the blood remains undetectable with the current clinical assay.

The HIV virus itself is thought to have originated in Africa. It is genetically related to a family of retroviruses that all infect immune cells that have been extensively researched and mapped. The generic term HIV usually refers to HIV-1, while a second genetically distinct type of human-specific immune virus called HIV-2 also targets humans but accounts for less than 5% of the infected people and is mostly limited to Africa. HIV is a zoonotic infection that mostly likely spread to humans from chimpanzees and gorillas. Current molecular biologic evidence suggests that simian immunodeficiency virus (SIV) "jumped" to humans from both gorillas and chimpanzees and that the jump happened on several occasions leading to different subtypes of infection. SIV infections in chimpanzees and other large apes produce illness, but many monkeys are endemically infected and the virus is tolerated as a chronic infection by those species which are probably the source of infection in large apes. Based on genetic mapping of the changes in the HIV virus over time, it appears that HIV jumped to humans in the early 20th century. Confirmed cases from the 1960s support the view that HIV was slowly moving toward epidemic spread for many years before it came to our attention. The numerous conspiracy theories that persist are often believed to be proven and many people cannot be dissuaded from their viewpoint.

HIV is a lentivirus, a single-stranded, positive-sense, enveloped RNA retrovirus that requires the action of an enzyme (reverse transcriptase, an RNA-dependent DNA polymerase not found in normal cells) to transcribe RNA into DNA in a process counter to normal information flow from nuclear DNA to RNA. The HIV viral particle is an icosahedral particle that is made up of an outer envelope and a viral core. The envelope is mostly derived from the host cell membrane. It is studded with viral proteins including gp120 and the transmembrane gp41 proteins, which perform specific roles in the viral life cycle. The viral envelope is lined with the matrix protein (MA) referred to as p17. The cone-shaped core is made of capsid protein p24 and contains nucleic capsid protein p7 associated with two copies of single-stranded viral RNA and the enzymes reverse transcriptase and integrase.

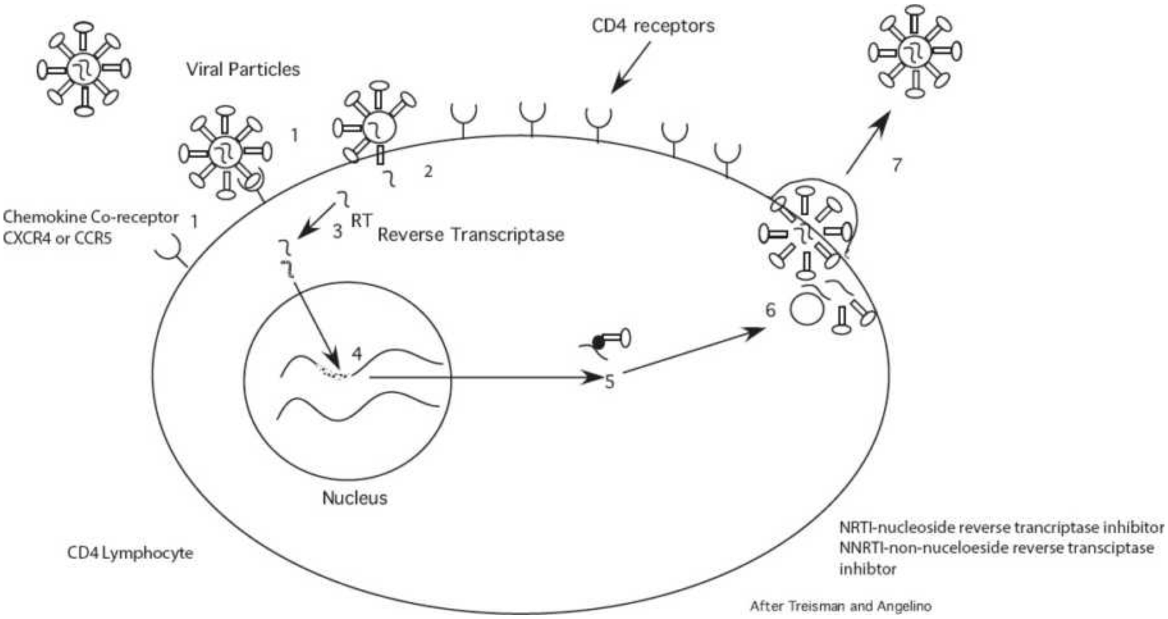

After Treisman and Angelino

FIGURE 2.8–4. The HIV life cycle and sites of action of ART. **1.** During binding and internalization, the viral gp120 protein binds to CD4 receptor–expressing cells. This is the site of action of the attachment inhibitor fostemsivir. Once bound, the virus binds to a second receptor, either CCR5 (most commonly) or CXCR4, and can then be internalized. The CCR5 antagonist maraviroc works at this stage. During internalization, the outer coating of the viral particle fuses with the cell membrane. This is the site of action of the drug enfuvurtide. Once inside the cell, the virus is uncoated and the contents released. **2.** During reverse transcription the enzyme reverse transcriptase which was carried within the viral particle transcribes the RNA into a double-stranded piece of DNA. This is the site of action of all NRTI's and NNRTI's that comprise the largest group of HIV drugs (see Table 2.8–1 for the list). **3.** During translocation, the transcribed DNA is moved into the nucleus with the help of the enzyme integrase. Once in the nucleus, integrase can insert the viral DNA into the human. This is the target of the integrase strand transfer inhibitors. **4.** After translocation, the viral DNA that has been inserted into the host DNA can direct the cell machinery to produce large quantities of viral messenger RNA. The viral RNA then uses the cells own machinery to produce the proteins required to make viral particles and other viral proteins. The viral DNA can also sit silently within the host DNA as a provirus and get activated when the cell is stimulated to perform immune function and then take over the cell. **5.** After activation, viral RNA and viral proteins are made within the cell in large quantity. Using the machinery of the cell, the proteins are assembled around the viral RNA and the immature virions bud outward from the cells surface onto the outer membrane. **6.** The viral particles "mature" due to the cleavage action of the viral protease and released from the cell surface into the host to infect other immune cells. This is the site of action of the protease inhibitors. (Courtesy of Glenn Treisman, M.D., Ph.D, and Andrew F. Angelino, M.D.)

HIV selectively infects immune cells, particularly T lymphocytes and macrophages. Infection requires direct contact between a viral particle with immune cells displaying the CD4 receptor. Infection takes place when viral particles are bound to CD4 receptors on immune cells. There are two primary cell types that get infected: T cells that express CD4 receptors (T4 cells or CD4 cells) and macrophage-derived cells. The usual route of initial infection is viral contact with dendritic cells in mucus membranes that then present the viral particles to T cells. The viral particles bind to the CD4 receptors on the T-cell surface and binding is aided by a co-receptor, either the chemokine receptor (CCR5) found on macrophages or the chemokine receptor (CXCR4) found on some T cells. New cells can be infected by virus being released into the blood and binding to the new cell, but probably occurs more often by T-cell to T-cell contact, by macrophage to T-cell contact, or by dendritic cells presenting virus to an uninfected cell.

Viral Life Cycle (Replication) and Treatment of HIV Disease and AIDS

The steps by which the virus infects a cell are summarized in Figure 2.8–4 and can be seen as a series of six different steps: the first step involves entry of the virus into the cell, which includes binding and fusion. The second step is reverse transcription followed by integration and replication where transcription and translation occur. Finally, in the last two steps, there is assembly, budding, and maturation. Figure 2.8–4 outlines the HIV life cycle and these steps. Each of the steps in the process of HIV infection provides a target for antiviral agents. There are now many drug classes available in various combinations which allow treatment to be tailored to a variety of patient-specific issues. See Table 2.8–1 for a list of the various drugs in each class discussed below. The sites of action for each class are shown in the figure.

Binding of the viral particle to the CD4 receptor on the cell surface and binding to the co-receptor are both targets for inhibition. Maraviroc is a drug that inhibits binding to the CCR5 co-receptor and is an effective antiviral drug in patients where the HIV virus has evolved to use the CCR5 receptor. An assay can be used to look for virus that can use the CXCR4 co-receptor which makes maraviroc less effective. The fusion of the viral particle with the cell membrane can be inhibited by enfuvirtide (also called T-20) and although it is rarely used now, it was lifesaving when it was developed. Fostemsavir is an attachment inhibitor that binds directly to the

HIV-1 gp120 protein and prevents viral attachment and entry of virus into host T cells. It was approved for use in the United states in 2020.

Once the viral RNA enters the cell, it is “reverse transcribed” by an enzyme carried within the viral particle and not normally found in eukaryotic cells. This enzyme was the target for the first effective HIV drugs. The reverse transcriptase step is the most utilized target for HIV therapy. The nucleotide and nucleoside reverse transcriptase inhibitors (collectively known as “nucs” or NRTIs) were the first drugs that were effective antivirals for HIV. This enzyme is also targeted by non-nucleoside reverse transcriptase inhibitors (“non-nucs” or NNRTIs).

The DNA copy made in this step is then integrated into the nucleus and the host genome by an enzyme called integrase, which is targeted by the drug class called integrase strand transfer inhibitors (INSTIs).

The viral DNA is transcribed to viral messenger RNA and is translated into viral proteins, assembled into virions, and finally cleaved from the cell membrane by a protease. This is the target point for the drug class protease inhibitors (PIs) (see Fig. 2.8–4).

Antiviral treatment recommendations change frequently and can be found on numerous websites. Guidelines for treatment are issued by several organizations and the Johns Hopkins POC-IT online guide summarizes most of the current information as does the CDC and the national institutes of health (NIH) guidelines.

The initial clinical use of these agents began in the late 1980s with AZT (originally called azidothymidine, now called zidovudine). When the drugs were given as monotherapy, eventually the virus became resistant to the drugs through mutation. In the mid-1990s, combination therapies (usually two “nucs” and a third agent, either a PI or a “non-nuc”), were shown to be effective in stopping the progress of HIV infection and essentially suppressed the virus. At first called “triple therapy,” then later HAART, it is currently called ART. While treatment suppresses viral replication so effectively that the virus does produce enough copies to develop viral resistance mutations and remains suppressed, extended treatment does not result in a cure. The clinical goal is suppression of viral replication to a level where virus particles are undetectable using the current standard assay. The newest treatments allow most patients to take one pill once a day or twice a day. The newer drugs provide treatment for even those who have been infected for a long time and are resistant to older therapies.

The Food and Drug Administration (FDA)-approved drugs for HIV at the time of this writing are shown in Table 2.8–1.

Table 2.8–1.

Antiretroviral Medications and Drug Classes (Combination Drugs Included)

Classes and Generic Names	Brand Names
CCR5 Antagonist	
Maraviroc	Selzentry
Fusion Inhibitor	
Enfuvurtide (T-20)	Fuzeon
Attachment Inhibitor	
Fostemsivir	Rukobia
Post-Attachment Inhibitor	
Ibalizumab-uiyk (injectable monoclonal antibody)	Trogarzo
Neucleoside/Neucleotide Reverse Transcriptase Inhibitors (NRTIs)	
Abacavir	Ziagen
Zidovudine	Retrovir
Stavudine (seldom used now)	Zerit
Didanosine (seldom used now)	Videx
Emtricitabine	Emtrivia
Lamivudine	Epivir
Tenofovir alafenamide	Vemlidy
Tenofovir disoproxil fumarate	Viread
Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)	
Rilpivirine	Edurant
Doravirine	Pifeltro
Efavirenz	Sustiva
Nevirapine	Viramune
Etravirine	Intelence
Integrase Strand Transfer Inhibitors (INSTIs)	
Raltegravir	Isentress
Cabotegravir	Vocabria
Elvitegravir	Vitekta

Dolutegravir	Tivicay
Protease Inhibitors	
Atazanavir	Reyataz
Tipranavir	Aptivus
Atazanavir-cobicistat (cobicistat to inhibit breakdown of atazanavir)	Evotaz
Darunavir-cobicistat (cobicistat to inhibit breakdown of darunavir)	Prezcobix
Fosamprenavir	Lexiva
Indinavir	Crixivan
Darunavir	Prezista
Nelfinavir	Viracept
Lopinavir/ritonavir (ritonavir to inhibit breakdown of lopinavir)	Kaletra
Saquinavir	Invirase
Ritonavir (antiviral now used to inhibit breakdown of other drugs)	Norvir
Combination Drugs	
Abacavir-lamivudine	Epzicom
Abacavir-lamivudine-zidovudine	Trizivir
Bictagravir-emtricitabine-tenofovir alafenadmide	Biktarvy
Darunavir-cobicistat-emtricitabine-tenofovir alafenamide	Symtuza
Dolutegravir-abacavir-lamivudine	Triumeq
Dolutegravir-lamivudine	Dovato
Dolutegravir-rilpivirine	Juluca
Doravurine-lamivudine-tenofovir disoproxil fumarate	Delstrigo
Efavirenz-emtricitabine-tenofovir disoproxil fumarate	Atripla
Elvitegravir-cobicistat-emtricitabine-tenofovir alafenamide	Genvoya
Elvitegravir-cobicistat-emtricitabine-tenofovir disoproxil fumarate	Stribild

Rilpivirine-emtricitabine-tenofovir alafenamide	Odefsey
Rilpivirine-emtricitabine-tenofovir disoproxil fumarate	Complera
Tenofovir alafenamide-emtricitabine	Descovy
Tenofovir disoproxil fumarate-emtricitabine	Truvada
Zidovadine-lamivudine	Combivir
Injectable Long-Acting Combination Drug	
Cabotegravir in combination with rilpivirine	Cabenuva

Several key studies have shown that the single most important factor in successful treatment is the ability to accurately and completely take the prescribed medications. Adherence correlates highly with treatment success in terms of viral load reduction. Resources are often scarce in HIV clinics and patients with psychiatric barriers to adherence must be identified and treated to effectively prevent treatment failure, transmission of HIV, and comorbidity. The ability to provide adequate psychiatric care to PLWH is critical for effective treatment of HIV.

Psychiatric disorders compromise the ability to take medications, adhere to treatment, practice safer sexual behaviors, and stop using IV drugs. While they are not the only factors driving the HIV epidemic, they have been largely overlooked in the past as a critical factor in the continued spread of HIV.

Psychiatry is clearly a necessary component of treatment in HIV settings. Psychiatrists can help manage the patients most likely to fail treatment and transmit the infection. An important question is what the psychiatrist's role should be in caring for patients at risk for HIV, or those who may have dropped out of treatment, or those with barriers to care? Should we test patients with unknown status? Should we prescribe PrEP to our patients at risk for infection? Do we go as far as to prescribe HIV medications to patients who have no access otherwise? The amazing simplification of PrEP down to two currently approved combination fixed-dose drugs with more safety than most psychiatric drugs argues that we should be considering it. Testing for HIV and helping patients access treatment seems appropriate in substance abuse treatment programs, when treating patients with chronic mental illness, as well as other patients at high risk for infection, but probably should be part of a standard psychiatric evaluation. If we are going to stop the epidemic spread of HIV, there must be psychiatric collaboration and inclusion in the medical care of these patients.

PSYCHIATRIC CONDITIONS IN THE HIV CARE

Delirium

Delirium is a state of global derangement of cerebral function. The term encephalopathy is often used for the same condition. It is often confused with psychosis in clinical settings and distinguishing delirium from psychotic disorders is crucial as delirium has high lethality. It occurs more frequently in medically ill, brain-injured, or metabolically unstable patients. Prevalence of delirium in HIV-infected populations has been reported between 43% and 65%. The clinical presentation and significance of delirium in patients with HIV are the same as in non–HIV-infected individuals, and is characterized by inattention, disorganized thinking or confusion, and fluctuations in level of consciousness. Emotional changes are common and often unpredictable, and hallucinations and delusions are frequently seen. The syndrome has an acute or subacute onset, and remits fairly rapidly once the underlying etiology is treated. If untreated, a marked increased risk of mortality is seen, with estimates of about 20% mortality rate in hospitalized delirious inpatients.

Aside from general risk factors such as older age, multiple medical problems, multiple medications, impaired visual acuity, and previous episodes of delirium, patients with HIV-associated dementia (HAD) are at increased risk to develop delirium. The differential diagnosis of delirium includes HAD, especially with AIDS mania, minor cognitive–motor disorder, major depression, bipolar disorder, panic disorder, and schizophrenia. Delirium can usually be differentiated from the above conditions based on its rapid onset, fluctuating level of consciousness, and link to a medical etiology. Delirium is most commonly seen in patients with advanced HIV disease, complex medical comorbidity with polypharmacy, or patients with limited resilience due to brain pathology.

The cause of delirium should be aggressively sought by intensive medical examination. Vital signs and oxygen saturation, careful history and physical examination, laboratory tests, electrocardiogram (EKG), radiologic examinations, and critical review of all medications are essential to the workup. Possible etiologies include toxic (poisonings, drugs, and new, recently changed or interacting medications, especially medications with potent anticholinergic activity), metabolic (electrolyte disturbances), infectious (CNS infections or sepsis), endocrine (especially thyroid and adrenal axes), neoplastic (especially CNS), cardiovascular (myocardial infarction, arrhythmia), neurologic (seizure, stroke), pulmonary (hypoxia, hypercapnia), and traumatic (head injury, burns) causes. In toxic/metabolic causes especially, the electroencephalogram may show diffuse slowing of the background alpha rhythm, which resolves as confusion clears.

Treatment consists of three parts: the first is the identification and removal of the underlying cause. The second is the reorientation of the patient by maintaining a normal diurnal variation of light cycles, getting patients up and moving, providing orienting stimuli, such as calendars, clocks, and a view of the outside world, and active engagement and reorientation by staff members, family, and friends. The third, if necessary, is the management of behavior or psychosis. In this third step, low doses of high-potency antipsychotic agents work well. Newer, atypical antipsychotics are currently being used with some success, but those with more anticholinergic activity may worsen the condition. Benzodiazepines should be used with caution, as they may contribute to delirium in some patients, but are of particular use in delirium caused by alcohol or benzodiazepine withdrawal. Physical restraint may be necessary if the patient becomes violent, but all attempts should be made to use this at a minimum, not only because of the legal issues it poses, but because restriction of movement or access to orienting stimuli worsens delirium and restraints have been known to cause death in some cases when patients are not monitored properly.

Over the last several years, there has been increasing recognition of postdelirium-related psychiatric conditions, particularly post-traumatic stress disorder (PTSD). Other psychiatric conditions such as depression, anxiety, and cognitive dysfunction also appear at increased incidence. This is not specific to HIV, but seems to be a consequence of delirium in general patient populations and appears to be related to severity of the underlying comorbidity.

Delirium in HIV-infected patients has a broadened differential diagnosis and greater number of possible causes. In immunosuppressed or advanced HIV illness, patients may suffer from unsuspected opportunistic causes of delirium. These are described at the end of this chapter.

HIV-ASSOCIATED DEMENTIA

Introduction and Neuropathology

Early in the AIDS epidemic, some patients presented with rapidly progressing neurocognitive disturbances leading to an intensive search for etiology. Several CNS opportunistic conditions such as cytomegalovirus (CMV) encephalitis, progressive multifocal leukoencephalopathy (PML), cerebral toxoplasmosis, cryptococcal meningitis, and CNS lymphoma were suspected. However, a subset of patients remained for which no identifiable pathogen causing the neurocognitive disturbances could be found and it was deduced that HIV itself was the causative factor behind the decline. These patients had a subcortical dementia characterized by increasing loss of attention and concentration, marked motor slowing, and variable behavioral components that usually progressed to death in less than a year. Autopsy studies of demented patients with AIDS revealed characteristic white matter changes and demyelination, microglial nodules, multinucleated giant cells, and perivascular infiltrates, but a marked absence of HIV within neurons. This has led to the current theories of neuronal loss due to inflammation through action of macrophages and microglial cells and/or through activation of cytokines and chemokines that trigger abnormal neuronal pruning. It appears that basal ganglia and nigrostriatal structures are affected early in the dementia process, with diffuse neuronal losses following. Typical late findings show an approximate 40% reduction in frontal and temporal neurons with atrophy appreciable on imaging studies.

With the development of ART, there has been a dramatic reduction in HIV-induced dementia of the type seen earlier in the epidemic. It is now uncommon in developed countries and less common even in underdeveloped settings. There is, however, still a high prevalence of cognitive impairment in HIV-infected patients, often subtler and with an unpredictable course, and with some marked clinical differences.

Epidemiology, Risk Factors, Clinical Course, and Treatment

With the development of ART, there was a hope that HIV dementia would become a thing of the past. Before ART, HIV dementia presented as a rapidly progressive subcortical dementia that usually ended in death. In the era of effective antivirals, HIV-associated cognitive impairment is harder to assess, and there is considerable debate about the prevalence, causes, and methods of assessment. There is also no effective treatment if the patient is well controlled on ART. Although there is even debate about the terminology, to assist with diagnosis and classification of patients for research and clinical purposes, most authors divide patients with HIV-associated neurocognitive disorder (HAND) by severity into three groups, applied when the patient's presentation cannot be explained by other conditions. Asymptomatic neurocognitive impairment (ANI), mild neurocognitive disorder (MND), and HAD are the three groups in order of increasing severity. One of the problems with assessment of these conditions is that accurate control groups are not well studied and much of the current prevalent data come from cohort studies. The currently used nosology was published in 2007 and was directed toward research but has been widely adopted clinically and has been the basis of significant discussion. The changing nosology has led to changes in the measure of prevalence of HIV-associated cognitive disorders. Although up to 50% of HIV-infected individuals can be shown to have cognitive impairment on careful testing, most patients are clinically asymptomatic. More importantly, unlike the steady downhill course that characterized HIV dementia prior to ART, cognitive impairment may get better with retesting and patients are often unaware of the changes. There has also been a significant increase in cortical dysfunction rather than the subcortical deficits that characterized HIV-induced cognitive impairment in the past.

Although the prevalence of HAD has declined, studies show overall rates of HAND are still remarkably high, especially milder forms. For example, reported data from 15,380 patients with HIV followed in the CASCADE cohort (Concerted Action on Seroconversion to AIDS and Death in Europe) showed a decrease in the incidence of HAD from 6.49 per 1,000 person-years before ART was available declining to 0.66 per 1,000 person-years by 2003 to 2006. A clinically diverse cross-sectional study of 1,555 patients at six academic centers showed the overall rate of cognitive impairment at 52% with 33% ANI, 12% MND, and only 2% HAD.

Various reasons for this continued prevalence of milder cognitive impairments have been hypothesized, with some being: possible subclinical brain injury that developed before treatment with ART was initiated, other comorbidities especially those that are vascular in nature, substance use, untreated psychiatric comorbidities, and effects of aging that may advance faster in PLWH. Higher medical comorbidity may result in increased HAND. Severity of disease, measured by current CD4 cell count and historical CD4 nadir as well as opportunistic infections, contribute to risk. Serum viral load and detectable virus in the CNS, all appear to contribute to the risk as well.

The persistent development of HIV cognitive impairment in treated patients could be caused by many factors, including persistent damage from the presence of latent HIV virus, damage from the immune system after the virus is controlled (immune reconstitution), toxicity from antiviral drugs, or persistent viral replication in the CNS because of poor entry of antiviral drugs. Initially, identification of variable drug penetration into the CNS appeared to be an exciting and remediable cause of continued viral injury. The expected finding that better penetrating antivirals would be associated with less HAND has not materialized, perhaps because of the confounds of poor correlation between CSF concentrations and those in the brain, direct mitochondrial toxicity of the drugs, degree of immune reactivation in patients with previous immunosuppression from HIV, and subtle ongoing viral replication in the CNS. There has also been the suggestion that the chronic immune activation associated with continued HIV infection, even with complete viral suppression, causes more rapid "aging" and therefore cognitive impairments that are usually associated with older age groups. In fact, the aging of the entire population of patients with HIV is a confound.

HAND in the current era involves both cortical and subcortical dysfunction. Prior to ART, patients had predominantly subcortical loss of function with greater impairments in motor skills, cognitive speed, and verbal fluency (as well as apathy and personality coarsening), whereas current patients tend to show more impairment in memory (learning) and executive function including deficits is prospective memory (the ability to "remember to remember") that has been shown to produce problems with medication adherence. Patients now have characteristics in common with cortical dementias such as Alzheimer disease. The course of HIV dementia prior to ART was almost always clinical progressive worsening of dysfunction, while in the current longitudinal studies, some patients decline while others actually improve.

The best treatment for patients with any HAND is aggressive ART in patients who are untreated or inadequately treated. Typical responses after initiation of ART may produce dramatic improvement, although patients with long-term deficits may have slower recovery. Patients with prominent subcortical features may respond to stimulants, particularly if there is a lot of apathy. There is at least some support for switching patients to a better CNS-penetrating regimen of antiviral medications. Safety assessments should be done when indicated and patients requiring intensive monitoring should be referred to appropriate facilities. To date, despite numerous trials, no clear effective treatment for HAND has emerged. There are patients who individually respond to interventions, sometimes dramatically, but clinical trials have not yet shown reliable results in populations, probably because of the heterogeneity of the underlying causes.

MAJOR DEPRESSION IN PATIENTS WITH HIV DISEASE

Epidemiologic Considerations

Depressive symptoms are the most common complication of chronic medical illness. Depression has a negative impact on adherence with medical treatments, quality of life, and treatment outcome. Despite education and efforts to improve treatment, depression remains underrecognized, underdiagnosed, and undertreated in medical clinics. Depression is a significant problem in HIV/AIDS. There are several long-standing issues for patients with comorbid medical conditions and depression. Major depression as a disease state is different from the depressive symptoms in patients who are demoralized (psychologically distressed due to their circumstances). Most research uses checklists and mood questionnaires to measure depression, making it difficult to clearly assess prevalence of and associations with major depression. To add to the complexity, severe psychological stress may trigger episodes of major depression in vulnerable individuals, and the body of research suggests a role for both the stress axis and the immune axis in major depression.

Despite these complexities, early studies showed that patients with depression had an increased likelihood of getting infected with HIV. Several risk factors for HIV infection associated with depression may explain this, including increased high-risk sexual behavior, increased substance use, and increased high-risk injecting practices. In addition, epidemiologic studies showed that advanced HIV disease correlated with the development of depression. Interpreting these studies is complicated by ascertainment methods using different depressive symptom rating scales, clinical records, and actual prospective clinical evaluation. Despite this, the data show that HIV is associated with a three- to fourfold increase in the prevalence of depression compared to the general population, and that depression has a negative influence on medical care and medical outcome. Depression is associated with poorer adherence to antiviral drugs, poorer engagement with health care, and higher risk of behaviors associated with transmission. Meta-analyses of data on the increased risk of depression in patients with HIV and the negative impact of depression on HIV outcome have been published over the course of the epidemic, with studies showing increased mortality associated with depression and several studies demonstrating the effectiveness of treatment of depression in terms of outcome. Although increased rates of depression have been found in MSM and patients with substance use disorders, a recent meta-analysis of 10 studies comparing HIV-positive and at-risk HIV-negative patients demonstrated a twofold increase in the prevalence of major depression in patients infected with HIV.

Major depression is a risk factor for HIV infection by virtue of its impact on behavior, intensification of substance abuse, exacerbation of self-destructive behaviors, and promotion of poor partner choice in relationships. In this way, depression can be seen as a vector of HIV transmission thus fueling the epidemic. Depression has been clearly shown to hinder effective treatment of infected individuals. Patients with major depression are at increased risk of disease progression and mortality. HIV increases the risk of developing major depression through a variety of mechanisms, including direct injury to subcortical areas of the brain, chronic neuroinflammation, chronic stress, worsening social isolation, and intense demoralization. While direct evidence for a relationship between worsening HIV disease and the development of major depression is limited, there are several studies that support this link, particularly the study based in the Multicenter AIDS Cohort Study (MACS) showing that there is a two-and-a-half-fold increase in rates of depression as the CD4 count falls below 200 just before patients develop AIDS.

Taken together, these lines of evidence suggest that HIV is a causal factor in depression, and that depression is a causal factor in HIV transmission and its morbidity, making the patients with these disorders a treatable vector for the HIV epidemic, and suggesting an important role of mental health care in HIV treatment and prevention.

Suicidal ideas, suicidal behavior, and successfully completed suicide have been shown to be increased in HIV-infected patients. The increased risk appears similar to other chronic diseases. Several of the risk factors for suicide (substance abuse, major depression, personality vulnerabilities, nongender conforming and homosexual orientation, and other factors) have been associated with both HIV infection and suicide risk. It appears that HIV-infected patients have an elevated risk for suicide beyond the contribution of these risk factors when comparing those at risk with those who are infected. The development and use of ART has been associated with a decrease in suicide risk but not to the level of background populations.

Differential Diagnosis of Major Depression

The differential diagnosis of depression includes nonpathologic states of grief and mourning (sometimes made quite severe by the vulnerabilities of the person) and a variety of disorders related to both psychological and physiologic disturbances. Patients with complaints of depressive syndromes can have dysthymia, dementia, delirium, demoralization, intoxication, withdrawal, CNS injury or infection, acute medical illness, and a variety of other conditions. AIDS dementia and other HIV-related CNS conditions can produce a flat, apathetic state that is often misdiagnosed as depression. Cocaine withdrawal produces a depressive syndrome, and delirium can mimic many psychiatric conditions including depression. CNS syphilis, a condition that had become quite rare has been reappearing in medical centers with HIV specialty services and remains “the great imitator” as it was called when it was originally described and can present with depression.

The diagnosis of major depression in the HIV clinic is complicated by the high frequency of depressive symptoms that are associated with chronic illness, significant losses and isolation, and complex medical treatments. The diagnosis may be further complicated by the presence of comorbid neurologic illness, comorbid substance use, and the use of many medications that can alter mental function. Despite this, studies of patients with major depression have shown that the response to treatment is similar to that expected in other populations.

Mood is the feature of affective life that describes the prevailing emotional tone a patient experiences. Patients express depression as feeling blue, low, sad, or sometimes flat, devoid of emotion, or empty.

Some say that they are miserable, in agony, anxious, worried, angry, upset, and irritable. Although some are apathetic, most find the state extremely uncomfortable. Moods may vary with the time of day, usually worse in the morning and better in the afternoon, but sometimes in other patterns. Pleasure (or a sense of reward) is associated with many behaviors, including those driven by appetite like sleeping, eating, and sex, and those associated with function, such as work, hobbies, dress, social activity, and artistic expression. Hedonic responsiveness refers to the enjoyment or positive reinforcement related to activities and experiences of life. Anhedonia describes the state in which hedonic responsiveness is blunted or absent, and is a principle feature of depression. Vital sense is the subjective sense of well-being. Depressed patients say they are sick, have a heavy pressure in their chest, have low energy, and even feel themselves dying. Self-attitude is the feeling directed at self. Depressed patients feel guilty, think that they are bad or evil, that they are undeserving of things they have, or that they have failed those they love.

Neurovegetative symptoms are commonly associated with this cardinal syndrome. These include difficulties with sleep, appetite, concentration, and memory. Sleep disturbances can include insomnia or hypersomnia, but most characteristically, patients describe waking early in the morning with difficulty falling back asleep. Appetite may be decreased or increased significantly, and patients often complain that food has lost its flavor. Appetite changes result in corresponding weight changes if untreated for long periods of time. Patients report slowed thought processes, with impairments in concentration and short-term memory, and, occasionally, generalized confusion.

Patients with HIV suffering from major depression frequently present to internists and family medicine doctors with multiple somatic symptoms. These include, but are not limited to, headache, gastrointestinal (GI) disturbances, inexplicable musculoskeletal or visceral pain, cardiac symptoms, dizziness, tinnitus, weakness, and anesthesia. Given the burdens of HIV, the medical problems associated with it, and the side effects of medications, depression may be very low on the list of considered causes of the patient’s complaint. Even patients complaining of depressive symptoms may have depression overlooked or discounted due to a plethora of other diagnoses. Nonspecific somatic symptoms may be the result of depression rather than HIV infection in patients who do not have concurrent medical illness.

As an example, fatigue has been found to be associated with depression and not HIV disease progression. Worsening of fatigue and insomnia at 6-month follow-up was highly correlated with worsening depression but not CD4 count, change in CD4 count, or disease progression by CDC category. These findings support the notion that somatic symptoms generally suggestive of depression should trigger a full psychiatric evaluation.

In the later stage of HIV infection, a variety of illnesses are common, often moving depression down on the differential diagnosis list. Somatic symptoms should always be evaluated carefully, and considered in context, that is, either with other indicators of progression of HIV disease or with other indicators of depression. Certain HIV-related medical conditions and medications can cause depressive symptoms. These include CNS disorders such as toxoplasmosis, cryptococcal meningitis, lymphoma, syphilis, and others. Some investigators have found significant rates of depressive symptoms among male patients with HIV with low testosterone levels and male patients with hypogonadism and major depression often have treatment-refractory depressive episodes until the testosterone is repleted. Of the antiretroviral agents, efavirenz is most often associated with depressive symptoms, which may remit spontaneously in a number of patients but persist in a few. More recently, rilpivirine and efavirenz (both NNRTI) and dolutagravir (INSTI) have been implicated in depression as well. Efavirenz (NNRTI) has been reported to cause frank psychosis in some patients, necessitating its withdrawal in these cases. Patients receiving interferon are at risk of depression from this agent, as are patients receiving chronic steroids (both anabolic and corticosteroids) for comorbid medical problems. Because of high coinfection rates, hepatitis C treatment with interferon-containing regimens often complicates HIV-related depression. The development of hepatitis C–directed antiviral agents, has allowed interferon use for hepatitis C to become a largely historical problem. A number of other agents, including metoclopramide, clonidine, propranolol, sulfonamides, muscle relaxants, and many others have all been reported to produce major depression or similar syndromes.

While these depressive syndromes often respond to withdrawal of the offending drug, when they do not, they should be treated as major depression with appropriate antidepressant medication. HIV-infected patients with major depression respond to antidepressant medication similarly to other patients with major depression. Routine screening for psychiatric disease in HIV treatment settings can effectively preempt urgent referrals for rapidly progressing disorders. Several screening tools for depression in medical settings have been studied and in one study specifically, looking at HIV-infected patients, the combination of two brief self-administered questionnaires, the Beck Depression Inventory (BDI) and the General Health Questionnaire (GHQ), prospectively predicted a psychiatric disorder (other than substance abuse) with a sensitivity of 81%, specificity of 61%, and a positive predictive value of 71%.

Treatment of Major Depression

Pharmacotherapy is the mainstay of treatment for major depression. Several studies have demonstrated efficacy of various antidepressant agents in patients with HIV, but no single antidepressant has been found superior in treating HIV-infected patients as a group. In a recent study examining the overall efficacy of antidepressants in patients with major depression and HIV, technical variations among studies were found to account for the largest portion of the variance of results, but overall, the use of antidepressants was supported.

Aside from how well the pharmacology of the antidepressant matches a patient’s disease, the engine that drives effectiveness is patient adherence. Patients who reliably take adequate doses of antidepressants have the best chance of improvement. A general rule is to start at low doses of any medication, titrating slowly up to a “full” dose or therapeutic serum level (when meaningful), so as to minimize early side effects that may act as obstacles to adherence.

Although some papers have suggested that selective serotonin reuptake inhibitors (SSRIs) antidepressants are the “first-line treatment,” the data do not show any advantage of one antidepressant over another, and most recommend using a medication that best fits the other elements of the case. Tables 2.8–2 and 2.8–3 list antidepressants often used in HIV disease, their dosages, and common side effects. In many patients, side-effect profiles of certain agents may be used to an advantage, and these “good match” agents should be tried first. Second choices are medications that do not have side effects that might worsen the patient’s symptoms (e.g., a medication that is not associated with insomnia for a patient having difficulty falling asleep). Later choices include medications that have side-effect profiles that would likely exacerbate a patient's symptoms (e.g., a medication associated with increased GI motility for a patient with chronic diarrhea).

The first week of treatment with a medication usually determines whether a patient will be able to tolerate the medication. Following this brief period, the dosage of any antidepressant should be increased slowly to either a typical full dose or a therapeutic serum level. Once at this level, patients should be encouraged to wait as long as possible for the therapeutic effect, which may take longer than 6 weeks to achieve in some patients. Side effects should be assessed at every visit, and attempts should be made to treat any side effects the patient finds bothersome. For example, insomnia caused by SSRIs may respond well to low doses of trazodone (25 to 150 mg at bedtime). Constipation from a tricyclic antidepressant (TCA) is often relieved by increased water and fiber intake and polyethylene glycol. Sexual side effects from SSRIs are common; erectile difficulty may be treated with nitric oxide synthase inhibitors (NOS inhibitors include sildenafil, tadalafil, and vardenafil) in select patients, while decreased libido and delayed orgasm may be alleviated with a drug holiday or by the addition of bupropion.

Patients who show only partial response to antidepressant medication should be offered an augmentation strategy or an alternative treatment. There are several other agents that have been used for antidepressant augmentation. The best studied is lithium, but its side-effect profile may limit its use in the HIV setting. Thyroid preparations, especially triiodothyronine, have also been shown to be of benefit, and may be of particular advantage in patients complaining of fatigue. Atypical neuroleptics have also been shown to be effective augmenting agents. Many other strategies have also been used with less support from the literature, including a second antidepressant, other mood stabilizers, trazodone, other antipsychotic medications, methylphenidate, benzodiazepines, sleep deprivation, and bright light therapy.

Table 2.8–2.
Side Effects of Antidepressant Medications

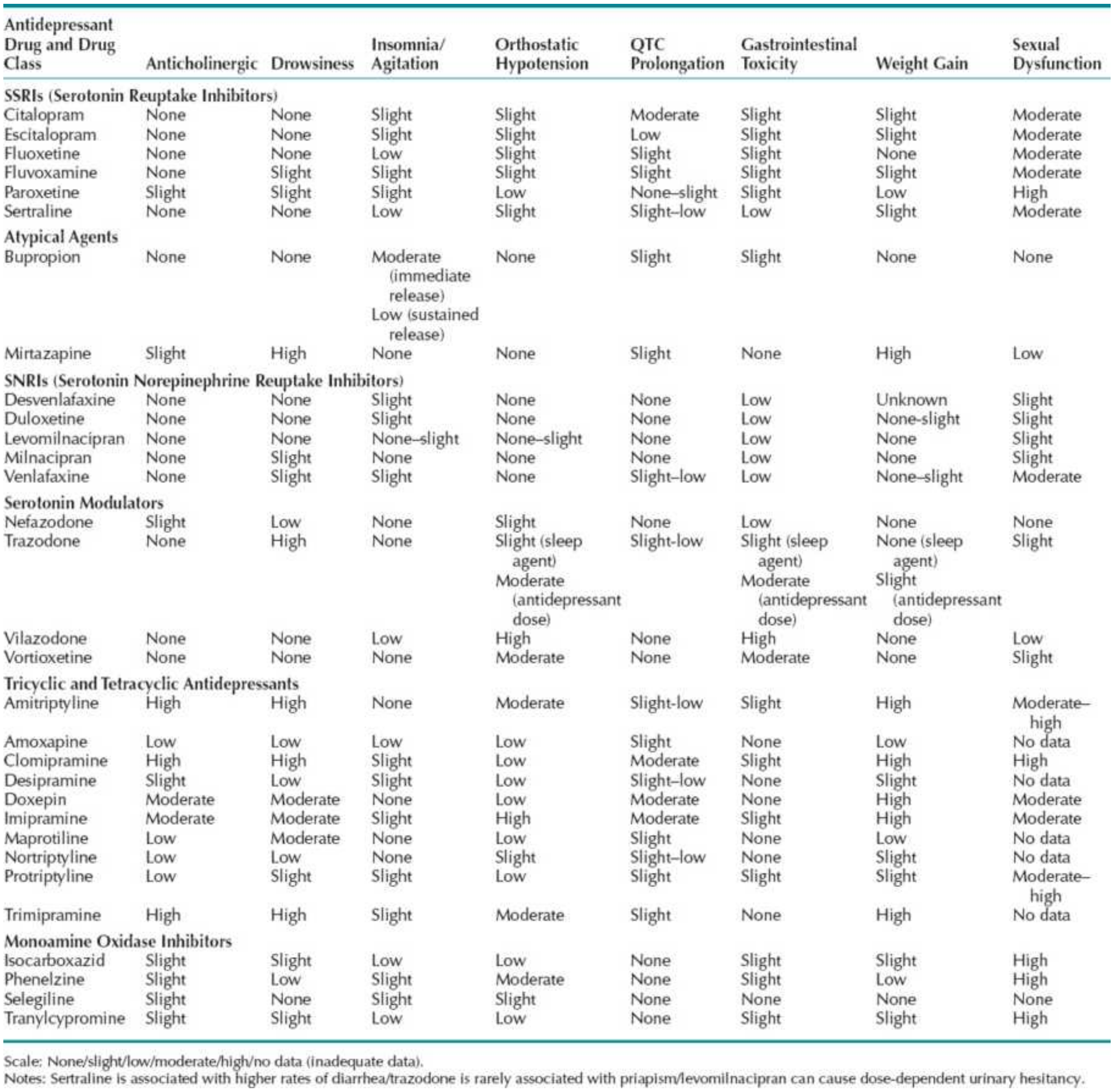

Antidepressant Drug and Drug Class	Anticholinergic	Drowsiness	Insomnia/ Agitation	Orthostatic Hypotension	QTC Prolongation	Gastrointestinal Toxicity	Weight Gain	Sexual Dysfunction
SSRIs (Serotonin Reuptake Inhibitors)								
Citalopram	None	None	Slight	Slight	Moderate	Slight	Slight	Moderate
Escitalopram	None	None	Slight	Slight	Low	Slight	Slight	Moderate
Fluoxetine	None	None	Low	Slight	Slight	Slight	None	Moderate
Fluvoxamine	None	Slight	Slight	Slight	Slight	Slight	Slight	Moderate
Paroxetine	Slight	Slight	Slight	Low	None–slight	Slight	Low	High
Sertraline	None	None	Low	Slight	Slight–low	Low	Slight	Moderate
Atypical Agents								
Bupropion	None	None	Moderate (immediate release) Low (sustained release)	None	Slight	Slight	None	None
Mirtazapine	Slight	High	None	None	Slight	None	High	Low
SNRIs (Serotonin Norepinephrine Reuptake Inhibitors)								
Desvenlafaxine	None	None	Slight	None	None	Low	Unknown	Slight
Duloxetine	None	None	Slight	None	None	Low	None-slight	Slight
Levomilnacipran	None	None	None–slight	None–slight	None	Low	None	Slight
Milnacipran	None	Slight	None	None	None	Low	None	Slight
Venlafaxine	None	Slight	Slight	None	Slight–low	Low	None–slight	Moderate
Serotonin Modulators								
Nefazodone	Slight	Low	None	Slight	None	Low	None	None
Trazodone	None	High	None	Slight (sleep agent) Moderate (antidepressant dose)	Slight-low	Slight (sleep agent) Moderate (antidepressant dose)	None (sleep agent) Slight (antidepressant dose)	Slight
Vilazodone	None	None	Low	High	None	High	None	Low
Vortioxetine	None	None	None	Moderate	None	Moderate	None	Slight
Tricyclic and Tetracyclic Antidepressants								
Amitriptyline	High	High	None	Moderate	Slight-low	Slight	High	Moderate–high
Amoxapine	Low	Low	Low	Low	Slight	None	Low	No data
Clomipramine	High	High	Slight	Low	Moderate	Slight	High	High
Desipramine	Slight	Low	Slight	Low	Slight–low	None	Slight	No data
Doxepin	Moderate	Moderate	None	Low	Moderate	None	High	Moderate
Imipramine	Moderate	Moderate	Slight	High	Moderate	Slight	High	Moderate
Maprotiline	Low	Moderate	None	Low	Slight	None	Low	No data
Nortriptyline	Low	Low	None	Slight	Slight–low	None	Slight	No data
Protriptyline	Low	Slight	Slight	Low	Slight	Slight	Slight	Moderate–high
Trimipramine	High	High	Slight	Moderate	Slight	None	High	No data
Monoamine Oxidase Inhibitors								
Isocarboxazid	Slight	Slight	Low	Low	None	Slight	Slight	High
Phenelzine	Slight	Low	Slight	Moderate	None	Slight	Low	High
Selegiline	Slight	None	Slight	Slight	None	None	None	None
Tranylcypromine	Slight	Slight	Low	Low	None	Slight	Slight	High

Scale: None/slight/low/moderate/high/no data (inadequate data).
Notes: Sertraline is associated with higher rates of diarrhea/trazodone is rarely associated with priapism/levomilnacipran can cause dose-dependent urinary hesitancy.

If no benefit is gained from the primary antidepressant, even after augmentation, or if it must be abandoned due to intolerable side effects, a new primary agent should be chosen, titrated slowly, and augmented as necessary. Although medications in the same class may produce similar side effects and therefore may not be tolerated, there is evidence to suggest that a response may be seen from one drug where none was seen from another in the same class.

Finally, clinicians often wonder about the interaction of antidepressant medications and ART medications. Some of these interactions may increase or decrease serum concentrations of either the antidepressant or the antiretroviral drug, which in turn, may make the side effects worse or decrease the effectiveness of either drug. First, since depression is associated with reductions in adherence to ART, untreated depression may be equally or more detrimental to disease progression than any medication interaction. Second, our experience in working with comorbid HIV and depression has not yet shown most predicted drug–drug interactions to be clinically significant, and it is unclear if dose adjustments for either antidepressants or ART are needed for successful outcomes to occur. Drugs with narrow therapeutic windows benefit from blood level assessment any time therapy is changed. If patients are better on their current treatment and there is a needed change in ART or antidepressant, it may be useful to assess blood levels prior to changing therapy so that subsequently if therapeutic response is altered, levels can be checked and adjusted to see if this will provide therapeutic benefit.

Table 2.8–3.
Antidepressants and Doses

Drug	**Recommended Daily Dose Range (Adults)**
Selective Serotonin Reuptake Inhibitors (SSRIs)	
Citalopram (Celexa)	10–40 mg
Escitalopram (Lexapro)	10–20 mg
Fluoxetine (Prozac)[a] has CR formulation	10–80 mg
Paroxetine (Paxil)[a] has CR formulation	10–50 mg
Sertraline (Zoloft)	25–200 mg
Serotonin Norepinephrine Inhibitors (SNRIs)	
Effexor (Venlafaxine)[a] has XR formulation	37.5–300 mg
Cymbalta (Duloxetine)	20–120 mg
Pristiq (Desvenlafaxine)	25–100 mg
Savella (Milnacipran)	12.5–200 mg
Atypical Antidepressants	
Wellbutrin (Buproprion)[a] SR and XL formulations available	75–450 mg
Remeron (Mirtazapine)	7.5–45 mg
Tricyclic Antidepressants (TCAs)	**Most Require Drug Levels**
Amitriptyline (Elavil)	10–300 mg
Nortriptyline (Pamelor)	10–150 mg
Protriptyline (Vivactil)	5–60 mg
Imipramine (Tofranil)	25–300 mg
Desipramine (Norpramin)	10–300 mg
Doxepin (Sinequan and Silenor) used at low doses for sleep	3–300 mg
Monoamine Oxidase Inhibitors	
Isocarboxazid (Marplan)	10–40 mg
Phenelzine (Nardil)	15–90 mg

Selegiline (Emsam)	6–12 mg/24-hour patch
Tranylcypromine (Parnate)	10–60 mg
Serotonin Modulators	
Nefazodone (Serzone)	100–600 mg
Vilazodone (Viibryd)	10–40 mg
Trazodone (Oleptro)	25–600 mg
Vortioxetine (Trintellix)	5–20 mg

[a] Dosing varies based on formulation. CR (controlled release), SR (sustained release), XL (Extended release).

Some medications have extreme daily dosing recommendations that may be used based on provider comfort and reason for medication usage.

Drug level monitoring done to determine dosing.

Data from Lexicomp Online. Copyright 1978-2021 Lexicomp, Inc. All Rights Reserved and Epocrates. Copyright 2013-2021.

Psychotherapeutic Treatment. Psychotherapy is an important and integral part of the treatment of major depression. Treatment with medication plus psychotherapy has been shown to be more effective for patients than either modality alone. The type of psychotherapy most appropriate to provide remains a major open question. Among the individual psychotherapies, interpersonal psychotherapy and cognitive behavioral psychotherapy are quite popular for the treatment of depression and have the best evidence to support their efficacy. The literature on the use of psychotherapy for the treatment of depression in patients with HIV is extensive but clinical trials data are sparse. There are multiple meta-analyses favoring the combination of psychotherapy with pharmacotherapy and less discontinuation of treatment in randomized controlled trials.

Improvements have been demonstrated as well for patients with HIV treated with group cognitive behavioral therapy either as a single treatment modality or combined with medication. Supportive psychotherapy helps patients with major depression who interpret their suffering to be a reaction to the diagnosis or morbidity of HIV infection. These patients often believe they can pull themselves out of depression and get frustrated when they continue to expend effort with little result. Education about the disease nature of their depression, encouragement to keep going, and therapeutic optimism that the treatments will work can help to decrease frustration.

In addition, psychotherapy applied judiciously and in combination with effective antidepressant medication provides patients with a framework for the provider–patient relationship that is crucial to success. The medical providers who keep the concept of psychotherapy in mind will structure their interactions with patients to slowly empower and enable the patients to take control of their lives, thus relying on the providers less and less.

BIPOLAR (MANIC-DEPRESSIVE) ILLNESSES IN PATIENTS WITH HIV DISEASE

Clinical Course of Bipolar Illness

Bipolar disorder is a condition in which patients experience episodic alterations in mood that cause disorder. Manic episodes are associated with increased rates of substance abuse and impulsive behavior, and there has been speculation that bipolar disorder may be a risk factor for HIV infection. To date, there has been no unequivocal evidence to show that bipolar illness directly increases the risk of HIV infection, but the technical difficulties in demonstrating this link are considerable.

In the classic presentation, patients alternate between extended episodes of depression similar to major depression and briefer episodes of increased mood, energy, and confidence and well-being, often with grandiose ideas about themselves and their circumstances. The synonymous appellation “manic-depressive insanity” is a reminder that many patients suffer from auditory hallucinations and frank delusions when they are ill. Most often, patients cycle from one type of mood state to the other, at times interspersed with periods of normal mood, but occasionally show features of both depressive and elevated mood states simultaneously (mixed states) or in very rapid succession (rapid cycling). Milder forms of mania are seen in a condition termed bipolar illness type II. The spectrum of bipolar illness is broad, ranging from a severely crippling and chronic mental illness to a mildly disordering alternating experience of prevailing mood. This has made it difficult to accurately measure prevalence and incidence and to explore the relationship of bipolar illness with HIV disease. In addition, it is often difficult to distinguish severe bipolar illness from schizophrenia even in studies that utilize rigorous research approaches. Thus, investigators looking at the relationships between HIV and mental illness often will use the term “chronically mentally ill” for patients with severe disability from either schizophrenia or bipolar disease.

The elevated mood states form a continuum from increased energy, euphoria, irritability, and decreased sleep called hypomania to a more extreme condition complicated by hallucinations, delusions, disordered thinking, disorganization, and, sometimes, violently agitated behavior called mania. Hypomania is characterized by euphoria, an improved self-attitude, and an elevated vital sense. Patients feel elated, energized, and as if they are functioning better than usual. Thoughts speed up and horizons expand such that the patient feels many brilliant things are coming to them in rapid succession. Often, there is a noticeable increase in the amount and speed of speech. It may be necessary (although difficult) to interrupt these patients. Since energy is so high, patients feel a decreased need for sleep and occasionally do not sleep at all. When these symptoms impair judgment and function is lost, the patient may develop frank mania. Manic patients not only have pressured speech, they often demonstrate a thought disorder in which ideas come so quickly that it is impossible to see the connections between them, the so-called flight of ideas. The expansive self-attitude may take on proportions outside the realm of reality, known as grandiose delusions. Paranoid delusional thoughts may be seen, and hallucinatory experiences occur in some patients.

AIDS Mania

In the early days of the HIV epidemic, some patients developed a type of mania that was specifically associated with late-stage HIV infection and cognitive impairment. This syndrome was termed AIDS mania and often occurred in patients with a lack of previous manic episodes or family history of bipolar disorder. This form of mania is now rarely seen as ART has had a significant impact on both the frequency of AIDS dementia and mania. Mania can occur anytime in the course of HIV infection for individuals with pre-existing bipolar disorder, but AIDS mania is usually confined to late HIV infection, thus appearing to be a consequence of HIV brain involvement. In general, manic syndromes in patients with HIV occur with higher frequencies after the onset of AIDS. Furthermore, patients with AIDS develop mania at rates substantially greater than the general population: in one series, mania occurred in 8% of all patients with AIDS seen at the HIV clinic over 17 months (more than 10 times the 6-month general population prevalence). The study grouped mania patients into those whose first manic episode came late in their HIV course with CD4 count <200 and those whose episode came early with CD4 count >200. The late-onset patients were less likely to have a personal or a family history of mania or any mood disorder, which presumably means they were less likely to have bipolar disorder or a genetic predisposition to mania. They were also more likely to have dementia or other cognitive impairment indicating brain damage.

AIDS mania has several features that distinguish the condition from the usual appearance of bipolar-associated mania. The history will usually reveal progressive cognitive decline prior to the onset of mania and patients will usually progress to the subcortical dementia associated with AIDS. Although most cases of cognitive impairment in current practice have mixed features of cortical and subcortical dementia, the dementia seen in very advanced HIV disease is still predominantly subcortical and progressive unless aggressive intervention with ART occurs. In AIDS mania, irritable mood is more characteristic than euphoria. Sometimes, prominent psychomotor slowing accompanying the cognitive slowing of AIDS dementia will replace the expected hyperactivity of mania, which complicates the differential diagnosis. Clinical experience has suggested that AIDS mania is usually quite severe in its presentation and malignant in its course. In one series, late-onset patients had a greater total number of manic symptoms than early-onset patients. They were also more commonly irritable and less commonly hypertalkative. AIDS mania seems to be more characteristically chronic than episodic, has infrequent spontaneous remissions, and usually relapses with cessation of psychiatric treatment. Because of their cognitive deficits, patients have little functional reserve to begin with. Also, they are less able to pursue treatment independently or consistently. One common clinically described manifestation is the delusional belief that one has discovered a cure for HIV or has been cured of HIV. While this may serve to cheer otherwise demoralized and depressed patients, it may also result in the resumption of high-risk behavior and lead to the spread of HIV and exposure to other infectious entities. When euphoria is a prominent symptom in otherwise debilitated late-stage patients, caregivers may wistfully question the humaneness of robbing patients of the illusion of happiness. It is the clearly impairing, often devastating effects of the other symptoms of mania that tips the balance of the risk/benefit equation toward treatment.

Treatment of Bipolar (Manic-Depressive) Illnesses

The treatment of mania in early-stage HIV infection is not substantially different than the standard treatment of bipolar disorder. It relies on the use of mood-stabilizing medications, particularly lithium salts and the anticonvulsants like valproic acid, lamotrigine and carbamazepine, and antipsychotic agents, now more commonly atypical antipsychotics. These medications decrease manic symptoms and may prevent recurrence.

AIDS mania in advanced HIV disease with low CD4 counts, more medical illnesses, more CNS involvement, and greater overall physiologic vulnerability may require different treatment strategies. While treatment with traditional mood stabilizers may be preferred, it can be very difficult in patients with advanced disease. There has been considerable experience with traditional mood-stabilizing agents in selected AIDS mania patients, but with relatively sparse documentation. Lithium use has been problematic for several reasons including high rates of associated delirium and cognitive difficulty, GI symptoms including nausea and diarrhea, and polyuria resulting in dehydration. Lithium is also associated with the development of diabetes insipidus in rare patients and may be difficult to use in combination with tenofovir, an NRTI with known renal side effects. The major problem with lithium in patients with AIDS has been rapid fluctuations in blood level, occurring even in the hospital on previously stable doses, causing lithium intoxication. Valproic acid has been used with success, titrating to the usual therapeutic serum levels of 50 to 100 ng/dL. Enteric-coated Depakote is better tolerated in most patients. This is sometimes limited by side effects, especially hepatotoxicity in the setting of chronic viral hepatitis. Monitoring of liver function tests is essential, but hepatic toxicity is not often a problem. In cases of severe hepatic *Mycobacterium avium* complex (MAC) infiltration, for example, with portal hypertension, valproic acid should likely be avoided, but this and related considerations have not been formally studied. Carbamazepine may also be effective but more poorly tolerated because of sedation, and because of the presumed potential for synergistic bone marrow suppression in combination with antiviral medications and HIV itself. Depakote and carbamazepine can also affect hematopoietic function, so white blood cell and platelet counts must be monitored. Lamotrigine is often used to successfully treat bipolar disorder and may have special promise in treating patients with prominent depressive episodes. There are no studies showing an increased incidence of rash or Stevens–Johnson syndrome in patients with HIV receiving lamotrigine, but prescribers should monitor for these adverse events closely, as with any patient. Other agents such as gabapentin, oxcarbazepine, and topiramate have been tried, but the reports of success remain anecdotal.

AIDS mania patients respond to treatment with relatively low doses of antipsychotic agents and these became the mainstay of treatment. Second-generation antipsychotics were reported to be better tolerated. In general, late-stage patients are far more sensitive to the therapeutic effects, but even more so to the toxic side effects of antipsychotic agents. AIDS mania patients are very sensitive to the side effects of antipsychotic medications, even those in the second generation, and extrapyramidal symptoms treated with even modest doses of anticholinergic agents can result in delirium. Because AIDS mania is seldom seen now in developed countries, the treatment has had little study in the last two decades. AIDS mania is seen in developing countries with high prevalence of HIV and can confound the use of ART and make patients difficult to treat.

SCHIZOPHRENIA IN PATIENTS WITH HIV DISEASE

The literature on patients with severe and chronic mental illnesses, accounted for primarily by schizophrenia and bipolar I disorder, reports prevalence rates of between 4% and 19% in both inpatient and outpatient samples. There is no evidence that HIV infection causes schizophrenia, but there are data to show that schizophrenia contributes to behaviors that may lead to HIV infection. Although injection drug use accounts for the majority of infections in many studies, there has also been a wealth of information published regarding the sexual risk factors for patients with schizophrenia. In particular, data reveal high rates of unprotected sex, multiple sex partners, trading sex for money or other goods, and sex while intoxicated. Further, there is evidence that patients with more positive symptoms and impulse control problems are at increased risk of high-risk sexual behavior despite demonstration of adequate knowledge of HIV risk factors. Practitioners that see patients with schizophrenia should be sensitive to the risk of acquiring HIV and should screen patients carefully for high-risk behaviors in addition to inquiring about patients' knowledge of HIV transmission routes.

The principles of treatment for HIV-infected patients with schizophrenia follow the same basic principles as any other patient with schizophrenia, namely control of symptoms with medications and psychosocial support and rehabilitation. In these cases, however, close ties with HIV providers are strongly suggested, so that HIV treatment can be coordinated and monitored. In a recent survey, HIV care providers were just as likely to recommend antiretroviral therapy for patients with schizophrenia who met criteria as those without schizophrenia, but to recommend avoiding efavirenz-based regimens due to higher risk of neuropsychiatric side effects. Further, these practitioners were more likely to seek collaboration with a mental health care provider to coordinate treatment for patients with schizophrenia, suggesting that this population is recognized as outside the scope of practice for solo management by an HIV primary care provider.

ISSUES OF SUBSTANCE USE DISORDERS AND ADDICTION IN HIV DISEASE

The term "addiction" has been criticized and changed in the literature but remains the most widely used term for the ongoing and disordered use of psychoactive substances. The DSM-5 recognizes that substance use disorders exist on a continuum from mild to severe consisting of escalating substance use, physiologic and psychological dependence, and increasing impairment of function. For example, it is often difficult to define precisely when the transition from heavy drinker to alcoholic occurs. Some patients use heavily but never actually become disordered by their use. Others are disordered by surprisingly modest use of a substance. The key feature of addiction is the habitual, compulsive use of substances where drug procurement and use become the main focus and goal of the person's life and continues despite negative physical, psychological, and social consequences. Physiologic dependence defined by tolerance and/or withdrawal may be present but should not be confused with addiction or a substance use disorder as defined by DSM-5.

A multifactorial matrix of influences initiates, drives, and sustains substance use and addiction. Many of these are due to the intrinsic rewarding effect of the substances themselves, but others are characteristic of the host or the environment. Behavioral approaches to understanding addiction have been particularly fruitful, as seen in the work of Joseph Brady and his colleagues. Their approach developed animal models of self-administration and measured the reinforcing properties of different drugs, many of which are profoundly predictive of human behavior. However, environmental influences and the understanding of the individual and cultural and social forces that impact substance use are essential to the translation of those models into human settings.

Psychiatric and psychological comorbidity can also increase vulnerability to substance use disorders. For example, patients with major depressive disorder may find the ordinary experiences of life to be less rewarding and are more sensitized to the positive reinforcement of drugs. Patients with cluster B personality traits may engage in more risk-taking behavior, have a greater likelihood to experiment with novel sensations, and experience an increased sensitivity to the rewarding effects of drugs and alcohol, leading to a greater likelihood of developing a substance use disorder. Other personality types are consequence and risk avoidant, and are relatively protected from addiction. However, these risk avoidant types may become addicted because of an underlying affective or anxiety disorder and turn to the rewarding properties of drugs and alcohol to “self-medicate” dysphoria and anhedonia. Life experiences that expose people to drugs such as social acceptance of drug use within peer groups or family members, particularly during adolescence, can also increase the risk of addiction.

Individual biology is involved in several ways. In the case of alcohol, genetic factors may affect the degree to which alcohol is rewarding, so that some patients report that their first drink was so gratifying that they began a lifetime of heavy drinking immediately. Others will say they never really liked drinking all that much, and therefore are surprised as they gradually become more dependent on alcohol to control the emotional discomforts of their lives. Genetic predisposition may play a lesser role in stimulant and opioid use. These substances are intrinsically more rewarding and patients exposed to these substances have a higher use-to-addiction ratio. Finally, co-occurring medical problems common in HIV such as chronic pain, opportunistic infections, and surgical procedures may result in exposure to opioids and/or sedatives, which can lead to addiction in a vulnerable individual. Substance use disorder is a primary vector for the spread of HIV through engagement in a number of high-risk behaviors during states of intoxication or engaging in actions necessary to use or obtain drugs. Such behaviors include the following:

- Sharing of contaminated needles
- Prostitution
- Risky sexual behavior (e.g., not using condoms)
- Sexual contact within a high-risk population of injection drug users

At the end of 2018, injection drug use accounted for approximately 15% of all HIV diagnoses in the United States. Heroin (80%) was reported to be the most commonly injected drug, followed by heroin and cocaine combined (61%), cocaine (48%), methamphetamine (32%), and prescription opioids (15%). Although new diagnoses of HIV in injection drug users have decreased by 48% from 2008 to 2014, the rate of new infections has remained steady over recent years at approximately 10%. There has been a rise in reported outbreaks of injection-related HIV infection in certain rural communities due to increased use of heroin, illicit fentanyl, and oxymorphone driven by the opioid epidemic, and decreased prescribing of opioid pain medications by providers. Strategies to decrease HIV transmission in injection drug users include needle exchange programs, HIV PrEP, medication-assisted treatment (MAT) for opioid use disorder, and linkage to HIV care. Patients with opioid use disorder should also be prescribed naloxone to reduce risk of death from opioid overdose.

In alcohol and noninjecting drug users, substance use also plays a major, although subtler, role in HIV transmission. Addiction and high-risk sexual behavior have been linked across a wide range of settings. For example, female crack cocaine abusers are more likely to engage in prostitution to obtain money for drugs. MSM who use crack cocaine or methamphetamine are more likely to engage in anal sex without condoms and have multiple sexual partners. Alcohol use, which is very prevalent in the HIV population, can lead to risky sexual behaviors during intoxication by way of cognitive impairment and disinhibition.

Ongoing substance use can not only facilitate HIV transmission, but can also negatively affect outcomes in HIV treatment. Once infected by HIV, patients are often demoralized, become hopeless, and are more likely to continue using drugs and engage in high-risk behaviors. Patients with substance use disorders may not seek health care or may be excluded from health care because of stigmatization. In addition, intoxication and the behaviors necessary to obtain drugs interfere with adherence to medication regimens and medical appointments.

Substance Use Disorders and Their Interaction with HIV Disease and Treatment

Ongoing substance use has grave medical implications for HIV-infected individuals. Screening for alcohol and drug use disorders is essential during the initial clinical evaluation and a number of quick and easy screening tools are available to clinicians, such as the CAGE, AUDIT-C, and DAST-10. Substance use may be difficult to detect because physical symptoms of HIV infection overlap with those of substance intoxication and/or withdrawal, including malaise, fatigue, weight loss, fevers, and night sweats. In addition, the accumulation of medical sequelae from chronic substance abuse can accelerate the process of immunocompromise and amplify the progressive burdens of HIV infection itself. Injection drug users, for example, are at higher risk of developing bacterial infections such as pneumonia, sepsis, soft tissue infections, and endocarditis. Tuberculosis, sexually transmitted disease (STDs), viral hepatitis infection, and coinfection with human CD4 cell lymphotropic virus occur more commonly in injection drug users who are infected with HIV. Certain malignancies, lymphomas in particular, occur more frequently in HIV-infected drug users. Alcohol users may experience faster progression of HIV disease and poorer response to antiretroviral therapy secondary to the immunosuppressive effects of alcohol. In addition to the direct physical effects caused by drugs, active substance use is highly associated with both nonadherence and reduced access to antiretroviral medication.

Neurologic symptoms can overlap between HIV infection and substance abuse. For instance, both AIDS dementia and drug intoxication can present with apathy, disorientation, aggression, and an altered level of consciousness. Drug withdrawal can present with seizures and neurovegetative symptoms, as can opportunistic infections of the CNS. HIV-infected injection drug users tend to be at higher risk of developing fungal or bacterial infections of the brain and spinal cord. HIV-infected patients who drink alcohol, particularly those older than 50, may be more vulnerable to cognitive decline and structural brain changes on neuroimaging than patients who are nondrinkers.

The duration of survival in HIV-infected PWIDs has improved substantially since 1997 with the introduction of effective antiretroviral therapy, reaching rates similar to those in HIV-seronegative PWIDs. However, because the HIV-infected patient is likely to be on a variety of antiretroviral agents and prophylactic agents for opportunistic infections, the clinician must be especially mindful of interactions between these medications and the abused substances. For example, the so-called “D drugs,” didanosine (ddI) and stavudine (d4T), which are seldom used today but which many older patients have been exposed to, can cause peripheral neuropathies as a side effect which may be worsened by the neurotoxic effects of alcohol and malnutrition related to chronic substance abuse. Opioid users on methadone maintenance treatment are at particular risk of medication interactions. Rifampin, for example, increases the elimination of methadone from the body and may result in the rapid onset of withdrawal symptoms. Decreased plasma levels of methadone also occur with concurrent administration of abacavir, amprenavir, ritonavir, nelfinavir, loprinavir/ritonavir, elvitegravir/ritonavir, rilpivirine, efavirenz, fosamprenavir, and nevirapine; necessitating adjustments in methadone dosage should withdrawal symptoms occur. This has important implications for treatment adherence in that the patient on methadone maintenance may be less likely to take a medication because of the fear of going into opioid withdrawal. Patients may be more likely to relapse if opioid withdrawal symptoms do occur when they are started on antiretrovirals without proper communication with their methadone provider. Sublingual buprenorphine may be alternatively used for treatment of opioid use disorder. Studies examining drug interactions between buprenorphine and antiretrovirals have shown no clinically significant interactions with the exception of atazanavir which may cause increased buprenorphine concentrations and sedation.

Substance Use Disorders, HIV, and Co-Occurring Psychiatric Disorders

Co-occurring psychiatric disorders are common in the drug-using HIV population. In the past, the term “dual diagnosis” referred to a patient diagnosed with both a drug use disorder and another psychiatric disorder and “triple diagnoses referred to a dual-diagnosis patient who also has HIV infection. Such patients are overrepresented in treatment settings because of their symptom severity and chronicity. For instance, in inner city Baltimore, as many as 44% of new entrants to the HIV medical clinic at Johns Hopkins Hospital had an active substance use disorder. Twenty-four percent of these patients had both a current substance use disorder and a co-occurring nonsubstance-related Axis I diagnosis.

Affective disorders, especially major depression, are the most common co-occurring psychiatric disorder in this population with studies estimating a prevalence of 15% to 30%. Diagnosing affective disorders (and other psychiatric disorders) in drug users can be difficult. It may be necessary to observe the patient over a period of abstinence before the presence of an independent mood disorder can be established. Family psychiatric history, temporal relationship of symptoms to periods of drug use and abstinence as well as a longitudinal history of the patient’s symptoms obtained from family members, friends, and other psychiatric providers can help clarify the presence of a separate affective disorder. Given the prevalence of overlapping substance use and affective disorders in clinical settings, as well as the poor prognosis associated with both disorders if left untreated, a treatment approach should necessarily emphasize simultaneous and equal treatment of both entities.

The importance of identifying affective disorders lies not only in their own well-known sequelae, including suicide, but also in their complex interactions with addiction and HIV disease. Depression is associated with higher severity of addiction, resistance to treatment, nonadherence to antiretrovirals, higher viral loads, and lower CD4 counts. The anhedonia of depression can be a sustaining factor in addiction. Addicts with major depressive disorder find it difficult to respond to and enjoy life’s other rewards that pale in comparison to the intense, though, ephemeral rewarding effect of intoxicating drugs. Depressed patients are also more difficult to engage, invest in, and sustain treatment, given their anergy and negativism. It is essential, therefore, for the clinician to recognize and treat depression early to maximize a successful outcome and improve patient adherence and retention in treatment.

Difficulties in the realm of personality (Axis II) are also commonly seen in this population. Although personality disorder diagnoses are currently described in a categorical fashion in the DSM-5, it is probably more useful to view personality as being dimensional in nature. Using this model, personality traits exist along a continuum, which predicts habitual maladaptive approaches to life’s difficulties.

Many HIV-positive substance abusers would be classified as “unstable extraverts,” (see below) with personality traits of risk taking, reward seeking, emotionally driven, and present-time oriented. These traits are commonly present in the cluster B personality disorders in the DSM-5 (antisocial, borderline, narcissistic, and histrionic) and can be found in as many as 49% of all substance abusers. Not only do these traits result in a vulnerability to addiction and other risky behaviors that predispose one to become infected with HIV, but they also pose significant barriers to both HIV and addiction treatment. These patients tend to act on strong, impulsive feelings rather than on carefully considered treatment instructions. Their behavior tends to be driven by the transient, immediate rewards of drugs rather than by their lasting future consequences. Such patients tend to get bored easily and treatment is often unexciting. They tend to “want what they want when they want it” rather than when it may be good for them. It is critical to identify these personality vulnerabilities because they can have a profound effect on treatment engagement and prognosis.

Treatment of Substance Use Disorders in Patients Infected with HIV

Although oversimplified, the steps for the treatment of substance use can be outlined in this simple way. These steps often occur simultaneously as treatment begins, but will be described as a sequence.

1. Role induction and motivation to change
2. Detoxification
3. Treatment of co-occurring conditions
4. Rehabilitation
5. Relapse prevention

Role Induction and Motivation to Change. The initial and often most daunting task of treating the addict are engagement and induction of the patient role. The general rule is that addicts and treatment providers begin with differing agendas—addicts tend to come to treatment settings seeking comfort and crisis relief, whereas physicians and other health care providers look at long-term goals of improvement in the patient’s health and overall function. One of the critical initial tasks of the health care provider is to engender the gradual evolution of the patient’s attitudes to coincide with those of the treatment team.

Prochaska and DiClemente have described a “transtheoretical stages of change” model to explain the addiction and recovery process. The patient is viewed as progressing through several different stages of change: (1) precontemplation—the patient has no intention to change their addictive behavior; (2) contemplation—the patient is considering change because of the negative consequences of drug use but is ambivalent about it; (3) preparation—the patient shows intention to change and takes initial steps to seek treatment; (4) action—the patient decides to modify behavior, environment, and circumstances in order to relinquish the addictive lifestyle; (5) maintenance—the patient works to prevent relapse and consolidate his changed behavior and lifestyle. The clinician’s job is to assist the patient in moving from one stage to the next in order to facilitate the recovery process. A technique known as “motivational interviewing” developed by Miller and Rollnick can be used to heighten the patient’s readiness to change by using empathy and gentle confrontation to amplify the discrepancy between the substance abuser’s current lifestyle and long-term life-enhancing goals.

Detoxification. In order for intoxicated patients to understand and process the cognitive steps needed for recovery, detoxification is the first step. Many HIV-positive substance abusers benefit from a brief hospital stay to stabilize their psychiatric and medical comorbidities. Slowly tapering either the drug of dependence or using a cross-dependent drug that has a similar pharmacologic mechanism of action best accomplishes detoxification.

Detoxification is often unpleasant, and there has been no evidence to support the idea that noxious withdrawal during detoxification improves outcome. In fact, some behavioral studies suggest that patients suffering through severe withdrawal during detoxification may actually develop conditioned withdrawal, such that exposure to environments similar to the one experienced during “cold turkey withdrawal” can bring about subacute withdrawal symptoms months later leading to relapse. It should be noted that benzodiazepine, barbiturate, and alcohol withdrawal can be life-threatening and clinicians should have a low threshold for admitting these patients for inpatient detoxification. Active tapers using a slow downward titration of medication from the class to which the patient is addicted is recommended for opiates and sedative-hypnotics. Some authors have used antidepressants to help patients detoxify from psychomotor stimulant classes of drugs (amphetamines and cocaine) but data to support this practice is controversial. This approach may work best for patients with clear evidence of major depression.

Sedative-hypnotics or alcohol are best detoxified through the use of a long-acting benzodiazepine such as diazepam or chlordiazepoxide. Lorazepam or oxazepam should be used in patients with liver disease to prevent the accumulation of active metabolites. Detoxification of opiates is accomplished by starting patients on a taper of sublingual buprenorphine or oral methadone. Clonidine, methocarbamol, dicyclomine, loperamide, and ibuprofen can be used adjunctively to provide symptomatic relief.

Treatment of Co-Occurring Psychiatric Conditions. Many patients with HIV and addictions have co-occurring psychiatric conditions, which need to be treated in order to maximize treatment adherence and abstinence. Conditions such as major depression, bipolar disorder, and schizophrenia are best managed with pharmacologic treatment. Because these patients tend to have multiple medical complications, it is important to remember to start medications at low dosages and to titrate slowly to minimize the risk of developing adverse side effects and delirium.

Disorders of personality, in particular, unstable extraversion, are managed with cognitive behavioral forms of psychotherapy. The ways that unstable extraverts may sabotage treatment include staff splitting, doctor shopping, general nonadherence, and manipulative behavior. Therapy addressing these personality issues should include firm limit setting and consistency on the part of all health care providers involved. To this end, a documented treatment plan with clear goals agreed upon by all the treatment staff is essential. The treatment plan should be reviewed with the patient at the initiation of treatment and regularly during treatment so that he understands clearly what is expected of him and what he can expect from his treatment providers if he is adherent to these goals. Frequent and consistent communication among all treatment providers during the course of treatment can minimize splitting and address issues of nonadherence as they arise.

Maintenance Treatment and Relapse Prevention. Long-term treatment is necessary for patients to begin the process of lifestyle change and recovery since addiction is a chronic disorder. Because this patient population is complicated and especially vulnerable to relapse, the most useful model of treatment is integrated care. To this end, an HIV clinic with comprehensive care is especially useful to engage and maintain patients in treatment. Ideally, a clinic treating HIV-positive addicts should include medical providers, psychiatrists, social workers, housing counselors, peer support workers, and substance abuse counselors. Integrated settings can help the patient access needed services, adhere to the overall treatment plan, and improve provider communication. Specific strategies integrating substance abuse treatment with HIV care include the on-site provision of sublingual buprenorphine for opioid maintenance treatment at HIV clinics (see Buprenorphine-HIV Evaluation and Support [BHIVES] Collaborative model below) and providing directly observed antiretroviral therapy and HIV care at opioid treatment centers.

It is important to remember that addiction treatment is active rather than passive and involves transforming previously held beliefs, attitudes, and personal identity into a new way of life. To this end, group therapy should be included as part of all substance abuse treatment. Various group modalities are available, including 12-step meetings (Alcoholics or Narcotics Anonymous), network therapy, rational recovery, therapeutic community, or SMART recovery. Group principles are similar in that more experienced members of the group provide both confrontation and support for the newly initiated member. Group support also provides the newly recovering addict with a hopeful view of the benefits to be achieved with recovery, exemplified by the lifestyle and achievements of group members with longer periods of abstinence. A commitment to a community of recovery assists the patient in severing ties from the drug-using community and provides the patient with new bonds that help maintain a sense of purposefulness and hopefulness. Specific HIV-positive recovery groups are widely available and may be helpful for addicts who are uncomfortable with their HIV status in a regular group setting.

Patient-individualized therapy should focus on identifying triggers to substance use, on minimizing or decreasing exposure to substances, and on defining a clear plan of action if relapse occurs. It is important to realize that relapse is often the rule and not the exception, and plans should be in place for early intervention. Objective measures such as urine and serum toxicologies and Breathalyzer tests should be used to monitor drug use and can help guide treatment. Contingency management using a variety of positive and negative reinforcers tied to urine toxicology results has also been shown to be effective in maintaining sobriety.

Individual and family therapies can enhance the effectiveness of treatment but should not take the place of group therapy. In individual treatment, it is important that the treatment provider remain flexible in the treatment approach. While regular psychotherapy may work for some patients, others may need a more “hands-on” approach and benefit from being referred to vocational rehabilitation, occupational therapy, and social skills training. Treatment failures often result when the therapist has a “one-size-fits-all” mentality and adheres too rigidly to one model of therapy.

Pharmacologic treatments are available for alcohol, opioid, and nicotine use disorders, can be used as adjuncts to the overall treatment plan and are particularly helpful for patients with opioid use disorder. Pharmacologic treatment can be divided into the following categories:

1. Aversive conditioning
    1. Disulfiram (alcohol)
2. Blockade of positive or negative reinforcement
    1. Naltrexone (opioids, alcohol)
    2. Acamprosate (alcohol)
3. Drive suppression
    1. Bupropion (tobacco)
    2. Buprenorphine (opioids)
    3. Methadone (opioids)
    4. Naltrexone (opioids)
    5. Varenicline (tobacco)
    6. Nicotine replacement therapies (tobacco)

Relapse to alcohol use can occur during a single impulsive moment, and some patients find pharmacologic therapy helpful as a type of “insurance policy” against cravings. Disulfiram, an inhibitor of acetaldehyde dehydrogenase, is taken once daily at dosages from 250 to 500 mg and causes an unpleasant reaction when alcohol is ingested due to the buildup of acetaldehyde in the body. Symptoms include nausea, flushing, headaches, and hypotension. Liver enzymes should be monitored because of the risk of hepatotoxicity. Naltrexone orally once a day has been shown to reduce alcohol cravings in a number of studies, probably by reducing the pleasurable effects of alcohol in alcoholics. Naltrexone is also available in an extended-release injectable formulation

(Vivitrol) given once a month, which may help patients who have problems with medication adherence. As with disulfiram, liver function tests are monitored in patients on naltrexone because of the risk of hepatotoxicity. Acamprosate reduces alcohol relapse by inhibiting excitatory glutamatergic transmissions that can cause subacute withdrawal symptoms in newly abstinent individuals. This medication is excreted renally so it can be particularly useful in patients with liver disease or impairment; however, it is dosed three times a day which makes medication adherence more problematic.

For patients with opioid use disorder, pharmacotherapy or MAT with opioid agonists has been shown to be highly effective in preventing relapse to opioid use as well as reduction in HIV injecting risk. Treatment with methadone, a mu-opioid receptor agonist, is considered the "gold standard" for opioid use disorder use and has been extensively studied for MAT. The typical therapeutic dose ranges from 60 to 120 mg daily and can only be dispensed in a federally regulated opioid treatment program (OTP). Patients on methadone maintenance treatment show better levels of antiretroviral adherence and reduced HIV risk behaviors and transmission. Several agents used to treat HIV and related infections induce methadone metabolism so providers should be aware of drug–drug interactions and monitor patients closely for signs of withdrawal once therapy has been initiated. Buprenorphine, a partial mu-opioid receptor agonist has been shown to be as effective as methadone in opioid maintenance treatment and has the advantage of being more accessible because it can be prescribed in an outpatient clinic setting by a qualified provider with a Drug Enforcement Agency waiver. Buprenorphine is available in several transmucosal formulations combined with naloxone to reduce the risk of diversion and misuse. Buprenorphine/naloxone (bup/nx) is well tolerated by HIV patients and has minimal drug–drug interactions with antiretroviral medications. The patient should show signs of mild to moderate opioid withdrawal before initiating bup/nx to avoid precipitated withdrawal. A long-acting injectable formulation of buprenorphine (Sublocade) is also available and is a subcutaneous abdominal injection given monthly after a patient has been on a stable dosage of transmucosal bup/nx for 7 days. Naltrexone is an opioid antagonist that has a high affinity for blocking mu-opioid receptors. Patients with opioid use disorder maintained on naltrexone experience little or no euphoria when opiates are used. In particular, the extended-release injectable form of naltrexone (Vivitrol) is effective in reducing opioid cravings and has shown comparable effectiveness to bup/nx in decreasing opioid use and preventing relapse. The medication should only be started in patients when they are opioid free for 7 to 10 days to avoid precipitated withdrawal.

Integrating pharmacologic treatment into HIV care can result in improved treatment outcomes and patient satisfaction with care. The BHIVES Collaborative was a multisite demonstration and evaluation project funded by the U.S. Department of Health Resources and Services Administration from 2005 to 2007 that examined the integration of sublingual bup/nx treatment into HIV primary care clinics. Patients participating in integrated treatment were 52% less likely than nontreated patients to use illicit opioids during each treatment quarter and reported overall decreased illicit stimulant use. Patients not on antiretroviral therapy at baseline were found more likely to be started on treatment after initiating bup/nx and to achieve nondetectable viral loads over time. Improvements were also noted in their health-related quality of life. Patients receiving integrated treatment subjectively reported being highly satisfied with the effects of bup/nx and with their treatment outcomes. One study site found clear advantages to integrated treatment compared to referral for treatment to an OTP offsite: patients prescribed bup/nx in the HIV clinic were more likely to start on opioid agonist therapy, less frequently found to have positive opioid and cocaine test results, and attended significantly more HIV primary care visits.

ISSUES OF PERSONALITY IN PATIENTS INFECTED WITH HIV

A disturbing trend in the HIV epidemic has been the persistence of high-risk behaviors among individuals who are infected with HIV. Such individuals, who report high rates of sex and/or drug risk behaviors include HIV-infected drug users, patients presenting at HIV primary care clinics for medical treatment, and HIV-infected men who have sex with other men. Apparently, knowledge of HIV and its transmission is insufficient to deter these individuals from engaging in HIV risk behaviors, suggesting that certain personality characteristics may enhance their vulnerability to practice such behaviors.

Traditional approaches in risk reduction counseling emphasize the avoidance of negative consequences in the future, such as using a condom during sexual intercourse to prevent sexually transmitted infections (STIs). Such educational approaches have proved ineffective for individuals with certain personality characteristics. Most theoretical models of HIV risk behavior have not considered the role of personality factors and few studies have examined mechanisms accounting for dispositional influences on sexual risk taking. Effective prevention and treatment programs for HIV-infected individuals must consider specific personality factors that render them vulnerable to practicing risky behaviors that further endanger their health as well as the health of others. This section will:

1. Discuss extraversion/introversion and stability/instability, measurable traits of temperament that drive personality and behavior
2. Outline the role of personality characteristics and personality disorder in HIV risk behavior
3. Highlight specific interventions to reduce HIV risk behaviors that are formulated for individuals whose personality characteristics place them at increased risk

Dimensional Traits and Personality

Personality is defined by the emotional and behavioral characteristics or traits that constitute stable and predictable ways that an individual relates to, perceives, and thinks about the environment and the self. Personality emerges as the underlying disposition or temperament of an organism and is shaped and affected by development and environment. Traits that may appear to be positive or negative are in fact adaptive in one setting and maladaptive in another. Individuals vary in the degree to which they possess a given trait and in the way it influences their behavior. When traits found in certain individuals exceed the levels found in most of society and are sufficiently rigid and maladaptive to cause subjective distress or functional impairment, a personality disorder is usually diagnosed.

Most personality models depict individuals along affective temperamental dimensions of extraversion–introversion and stability–instability or variations on this theme. The dimension of extraversion–introversion refers to the individual’s basic tendency to respond to stimuli with either excitation or inhibition. Individuals who are extraverted are present oriented, feeling directed, and reward seeking. Their chief focus is their immediate and emotional experience. Feelings dominate thoughts and the primary motivation is immediate gratification or relief from discomfort. Extraverts are sociable, crave excitement, take risks, and act impulsively. They tend to be carefree, inconsistent, and optimistic. By contrast, introverted individuals are future and past oriented, cognition directed, and consequence avoidant. Logic and function predominate over feelings. Introverts are motivated by appraisal of past experience and avoidance of future adverse consequences. They are unlikely to engage in a pleasurable activity if it might pose a threat in the future. Introverted individuals are quiet, dislike excitement, and distrust the impulse of the moment. They tend to be orderly, reliable, and rather pessimistic.

The second personality dimension, stability–instability, defines the degree of emotionality or lability. The emotions of stable individuals are aroused slowly and minimally and return quickly to baseline. By contrast, unstable individuals have intense, mercurial emotions that are easily aroused and return slowly to baseline. If these two personality dimensions are juxtaposed, four personality types emerge as described from the time of the ancient Greeks: phlegmatic (stable introverts), sanguine (stable extraverts), melancholic (unstable introverts), and choleric (unstable extraverts).

Traditional Personality Models, Their Instruments and Findings Related to HIV Risk Behaviors

After the higher-order traits of extraversion and neuroticism, personality models have elaborated other trait domains and developed personality inventories to empirically assess these traits. The major models and their inventories are as follows:

1. The three-factor model (Eysenck & Eysenck) Extraversion, Neuroticism, and Psychoticism, measured by the Eysenck Personality Questionnaire (EPQ).
2. The five-factor model (Costa & McCrae) Neuroticism, Extraversion, Openness, Agreeableness, Conscientiousness, measured by the NEO Personality Inventory (NEO).
3. The alternative five factor model (Zuckerman) Neuroticism–Anxiety, Sociability, Impulsive Sensation Seeking, Aggression–Hostility, Activity,

measured by the Zuckerman Kuhlman Personality Questionnaire (ZKPQ).
4. The seven-dimensional model (Cloninger) Novelty Seeking, Harm Avoidance, Reward Dependence, Persistence, Self-Directedness, Cooperativeness, and Self-Transcendence, measured by the Tridimensional Personality Questionnaire (TPQ).

Personality traits appear to influence a variety of sexual risk behaviors, yet there is relatively little research on sexual risk taking from the major personality models. On the EPQ, extraversion is associated with sexual promiscuity, desire for sexual novelty, and multiple sex partners. Neuroticism is related to unprotected anal sex. Psychoticism is associated with number of sexual partners and unprotected sex in several studies. On the NEO, neuroticism is associated with unprotected sex and to a lesser extent, sex with multiple partners. Low conscientiousness is also associated with unprotected sex. Low openness to experience is associated with the denial of risk of HIV infection. The TPQ has minimal research on sexual risk taking, but one study shows that novelty seeking is associated with unprotected sex. The impulsive sensation seeking scale of the ZTPQ has received the most research attention and predicts number of sex partners, unprotected sex, and high-risk sex encounters, such as sex with a stranger, across a variety of populations.

Research on personality traits links extraversion and neuroticism to drug and alcohol addiction. On the EPQ, psychoticism and to a lesser extent, neuroticism has been linked prospectively to alcohol dependence in a 6-year study. On the ZTPQ, impulsive sensation seeking is consistently associated with addiction severity as well as amount and variety of illegal drug use. While there is no specific “alcoholic” or “drug-using” personality, there appears to be a modest link between substance abuse and either impulsivity/high novelty seeking or high on neuroticism/negative emotionality. Individuals with both these traits may be at the greatest risk of addiction.

Sensation Seeking. A significant empirical contribution to the understanding of the role of personality factors and HIV risk is the conceptual model connecting sexual sensation seeking, alcohol expectancies, and drinking before sex as key predictors of risk.

In a series of studies, Kalichman et al. adapted Zuckerman sensation seeking scale that measures preference for exciting, optimal, and novel levels of stimulation or arousal. Sensation seeking functions as the “third variable” connecting alcohol use to sexual risk behaviors. Using path analysis in a series of experiments, Kalichman et al. tested a model that predicts the association among sensation seeking, alcohol use expectancies, alcohol use, and sexual risk behavior in both men at risk of HIV and HIV-positive men. Sensation seeking is associated with alcohol outcome expectancies (or the beliefs that the individual has about the effects of alcohol on experience or behaviors). Having positive expectancies about the effect of alcohol on sexual pleasure or sexual behavior increases the likelihood that alcohol will be used in sexual situations. In turn, having sex when under the influence of alcohol is associated with an increased likelihood of having unprotected sex. The importance of this model is identifying a marker, sensation seeking, for multiple risk practices as well as identifying alcohol expectancies that can be a point of intervention for prevention and treatment. Recently, this model has been corroborated in heterosexual women and men. With the use of a cross-lagged panel design to explore the relationship between perceived susceptibility to AIDS and HIV risk behavior preincarceration and post release, it was found that HIV/AIDS patients that rate high in sensation seeking, per Zuckerman scale, have a high association with having unprotected sex. Interestingly there was not found to be a high association between patients who rated high in terms of impulsivity and unprotected sex, unless there was the addition of methamphetamines. A systematic review in 2014 looked at personality as a predictor of unprotected sexual behavior and found similar results to the study where the cross-lagged panel design was used. The results of these studies showed that personality constructs that were more compatible with sexual risk taking and a focus on the immediate intense sexual pleasure reward, such as sexual compulsivity and sex-related subcomponents of sensation seeking, had relatively direct associations with unprotected sex.

Implications for HIV Risk Behavior: Clinical Observations

There has been relatively little empirical investigation of the influence of personality characteristics on HIV risk behavior; however, the authors’ clinical observation suggests that of the four temperaments, unstable extraverts are the most prone to engage in HIV risk behavior. The authors estimate that in Psychiatry Service of the Johns Hopkins AIDS Service, about 60% of the authors’ patients present with this blend of extraversion and emotional instability (unpublished observation). These individuals are preoccupied by and act upon their feelings, which are evanescent and changeable. Thus, their actions tend to be unpredictable and inconsistent. They tend to be sensitive to emotional stimuli; easily becoming emotionally labile and subsequently have a slow return to baseline. Most striking is the inconsistency found between thought and action.

Regardless of intellectual ability or knowledge of HIV, unstable extraverts can engage in extreme behaviors that put them at increased risk of acquiring HIV. Past experience and future consequences have little salience in decision making for the individual who is ruled by feeling; the present is paramount. They live their lives through the lens of their emotions and believe their feelings to be factual and thus their feelings are their reality. Their overarching goal is to achieve immediate pleasure or removal of pain, regardless of circumstances. Furthermore, as part of their emotional instability, they experience intense fluctuations in their mood. It is difficult for them to tolerate anything considered painful like boredom, sadness, or unresolved drive; they want to escape or avoid these feelings as quickly and easily as possible. Thus, they are motivated to pursue pleasurable experiences however risky and eliminate low moods.

Unstable extraverts are less likely to plan ahead and carry condoms and more likely to have unprotected vaginal or anal sex. They are more fixated on the reward of sex and remarkably inattentive to the STI they may contract without use of protection. These patients may consistently test positive for syphilis due to the choice of partners and lack of condom use, but refuse to change their behavior. Unstable extraverts are also less likely to accept the diminution of pleasure associated with the use of condoms or once aroused, to interrupt the “heat of the moment” to use condoms. Similarly, unstable extraverts are more vulnerable to alcohol and drug abuse. They are drawn to alcohol and drugs as a quick route to pleasure and reward. They are more likely to experiment with different kinds of drugs and to use greater quantities. Unstable extraverts are also more likely to become injection drug users because the experience is more intense. They are also less likely to defer this intensity in the interest of safety.

The second most common personality type that the authors have observed, which may represent about 25% of Johns Hopkins AIDS Service patients referred to psychiatry is the stable extravert. Stable extraverts are also present oriented and pleasure seeking; however, their emotions are not as intense, as easily provoked, or mercurial. Hence, they are not as strongly driven to achieve pleasure. Their emotional imperturbability (described by many as sanguine) may generate a kind of indifference to HIV risk more than a drive to seek pleasure at any cost. Stable extraverts may be at risk because they are too optimistic or sanguine to believe that they will become HIV infected.

Introverted personalities appear to be less common among the authors' psychiatric patients. Their focus on the future, avoidance of negative consequences, and preference for cognition over feeling render them more likely to engage in protective and preventive behaviors. HIV risk of introverts is determined by the dimension of emotional instability–stability. About 14% of the authors' patients present with a blend of introversion and instability. Unstable introverts are anxious, moody, and pessimistic. Typically, these patients seek drugs and/or sex not for pleasure*,* but for relief or distraction from pain. They are concerned about the future and adverse outcomes, but believe that they have little control over their fates. Stable introverts comprise of the remaining 1% of patients. These patients with their controlled, even-tempered personalities are less likely to engage in risky or hedonistic behaviors.

Personality Disorders in HIV

Personality disorders represent extremes of normal personality characteristics and are disabling conditions. Prevalence rates of personality disorders among HIV-infected (19% to 36%) and HIV at-risk (15% to 20%) individuals are high and significantly exceed rates found in the general population (10%). Antisocial personality disorder (ASPD) is the most common and is a risk factor for HIV infection. Individuals with personality disorders, particularly ASPD, have high rates of substance use disorder and are more likely to inject drugs and share needles compared to those without a personality disorder diagnosis. Approximately half of drug abusers meet criteria for a diagnosis of ASPD. Individuals with ASPD are also more likely to have higher numbers of lifetime sexual partners, engage in unprotected anal sex, and contract STIs compared to individuals without ASPD.

Clinically, it has been useful to characterize patients along the dimensions of extraversion/introversion and emotional stability/instability rather than in the discrete categories such as narcissistic, histrionic, and antisocial. This approach is useful for several reasons. First, it is easier and quicker for all staff to determine where a patient falls along two dimensions than to evaluate, for example, each of the criteria to make a personality disorder diagnosis. Second, a diagnosis of antisocial or borderline personality disorder can be stigmatizing, particularly in a general medical clinic where health care providers may have less experience managing such patients. Also, it is not clinically good practice to diagnose a personality disorder in a patient that has other mental health disorders that have not been fully treated. In some case studies written for the Journal of the National Medical Association, it was noted that the clinical team performing the case studies found it extremely valuable to understand a PLWH's personality type in order to most effectively treat them medically. The authors have found a dimensional model of personality as described to be more clinically useful and a better predictor of risk behavior than models that categorically define personality disorders.

Implications for Medication Adherence

Average medication adherence across a variety of diseases and patient populations has been consistently estimated at 50%. Adherence is especially challenging in HIV disease, which is associated with all of the components of low adherence like long duration of treatment, preventative rather than curative treatment, asymptomatic periods, and frequent and complex medication dosing. Average rates of nonadherence to antiretroviral therapy range from 50% to 70% with adherence rates of <80% associated with detectable viremia in a majority of patients.

Personality factors have received little investigation in relation to adherence and antiretroviral therapy. Personality traits such as neuroticism were significantly associated with poorer quality of life whereas conscientiousness and extraversion were associated with better quality of life. In contrast, personality traits were not directly related to ART adherence. However, the authors' clinical experience suggests that nonadherence is more common among their extraverted or unstable patients. The same personality characteristics that place them at risk of HIV also reduce their ability to adhere to demanding drug regimens. Specifically, their present-time orientation, combined with reward seeking, makes it more difficult for these patients to tolerate uncomfortable side effects from PI drugs whose treatment effects may not be immediately apparent. It is also difficult for feeling-driven individuals to maintain consistent, well-ordered routines. Hence, following frequent, rigid dosing schedules can also be problematic. The authors' unstable, extraverted patients are usually intent upon following the schedule, but their chaotic and mercurial emotions are more likely to interfere and disrupt daily routines. For example, a patient may report that he felt very upset and nihilistic after a fight with a family member and miss several doses of his ART medicines. Missed doses of ART can increase the chance of HIV resistance developing. Identifying factors that influence adherence in HIV disease is important in improving overall health outcomes.

Treatment Implications

Psychiatric and medical treatment of patients with extraverted and/or emotionally unstable personalities is challenging. Such patients are often baffling or frustrating for physicians and other medical care providers because they engage in high-risk sex and drug behaviors in spite of knowing the risks, or fail to adhere to treatment regimens for HIV infection in spite of knowing the consequences. A patient may stop his antidepressant medicine because of a headache while being perfectly willing to inject heroin into multiple sites on his body. After 6 months of missed medical appointments, a patient may impulsively leave the clinic if the primary care provider is 15 minutes late for the appointment. Such personality traits reflect relatively stable, lifelong modes of responding; thus, direct efforts to change these traits are unlikely to be successful. It is possible, however, to modify the behavior that is an expression of the trait. By recognizing individual differences in risk-related personality characteristics, interventions can be better targeted and their impact maximized.

The authors have found that a cognitive behavioral approach is most effective in treating patients who present with extraverted and/or emotionally unstable personalities. Following are the five principles that guide the authors' standard care:

1. Focus on thoughts, not feelings. Individuals with unstable, extraverted personalities benefit from learning how they are predisposed to act in certain ways. Often, they do not recognize the extent to which their actions are driven by the impulse or feeling of the moment. These patients can fail to understand why they intend to stay clean but later

find themselves “shooting dope.” Treatment helps to identify the role that strong feelings play, so that these patients can begin the process of understanding their own chaotic, often irrational behavior.

Simultaneously, treatment encourages the patient’s cognitive, logical side. This process begins by identifying what the patient thinks in a given situation, as opposed to what they think about the situation: “I deserve some cocaine because I have had a difficult day.” The influence of the patient’s assumption upon feelings, behavior, and ultimately the consequences of behavior are examined. Through the treatment dialogue, patients understand that the notion of a “cocaine reward” creates a feeling of urgency and entitlement to getting high. As a consequence of using cocaine, the development of other more constructive reward systems or coping methods is preempted, a relapse into cocaine dependence can occur, and/or family or employers may be alienated. Through treatment, patients learn to identify maladaptive assumptions that drive feeling and behaviors so that they can either lessen the force of these assumptions or substitute more constructive assumptions to guide their life experience.

2. Use a behavioral contract. A behavioral contract is developed with all patients. The contract outlines goals for treatment, often only a day or a week at a time. While patients and mental health professionals may develop the contract, the focus of treatment is not on what patients want or are willing to do to get off of drugs, but rather on established methods, such as drug treatment and Narcotics or Alcoholics Anonymous. The importance of the behavioral contract lies in the creation of a stable plan that supersedes the emotional meanderings of these patients. Unstable extraverts present an ever-changing array of concerns and priorities. The task of the treatment is to order the priorities with patients and help them follow through on these, regardless of changing emotions. In short, behavioral contacts provide consistent, cognitive focus to patients’ bewildering emotional experience of life.
3. Emphasize constructive rewards. In developing the behavioral contract and in treatment, the purpose is cast in terms of the rewards that will follow from their behavioral change. Positive outcomes, not adverse consequences, are salient to extraverts. Most of the patients have already experienced negative consequences from their behavior, HIV, drug addiction, homelessness, and such. Exhortations to use condoms to avoid STDs are unpersuasive. More success has been achieved

with extraverts by eroticizing the use of condoms or by the addition of novel sexual techniques (erotic massage, use of sex toys) into sexual repertoires. Similarly, the rewards of abstaining from drugs or alcohol are emphasized, such as having money to buy clothing, having a stable home, or maintaining positive relationships with children.

   In building adherence to antiretroviral therapies, the focus is on the rewards of an increased CD4 count and reduced viral load, rather than avoiding illness. Using the viral load as a strategy to build adherence can increase acceptance in all patients but is especially effective in reward-driven extraverts.

4. Use relapse prevention techniques. The relapse prevention model, originally developed for treatment of substance abuse behavior, is an effective method for changing any habitual way of behaving. This intervention trains individuals to recognize and interrupt the sequence of behaviors that link to the final high-risk behavior. Behavioral therapy methods are also used to teach individuals how to recognize and avoid situations that trigger high-risk behaviors.
5. Coordinate with medical care providers. Medical care providers are often frustrated or discouraged when treating unstable, extraverted patients. It is useful to provide education about a patient’s personality and how it influences behavior. Particularly effective is the development of a coordinated treatment plan, where medical care provider and mental health professional work in tandem to develop behavioral contracts to reduce HIV risk behaviors and build medication adherence.

Conclusion

Personality characteristics and personality disorders reflect relatively stable, lifelong propensities that are difficult to change. This does not mean, however, that HIV risk reduction efforts are necessarily futile. Rather, by understanding personality characteristics and their role in HIV risk behavior and medication adherence, the mental health professional can develop more effective, specific treatment strategies. Similarly, the HIV-infected patient who can identify aspects of their personality that might interfere with intentions to practice safer behavior, and who knows strategies for dealing with these situations, is less likely to practice high-risk behaviors. Finally, the mental health professional can provide valuable assistance to medical care providers to improve health outcomes for these patients.

PSYCHOLOGICAL PROBLEMS IN PATIENTS INFECTED WITH HIV

In the context of the psychological realm of living with HIV, it is difficult to distinguish which came first, the variety of psychiatric disorders the patient has, the HIV risk behaviors, the losses and psychological traumas, or the impact of having HIV itself. In fact, it is a common sentiment from patients to echo one of the patients at the authors' clinic who said "HIV is not even one of my biggest problems." Each of these factors seems to be both a consequence and an antecedent to all the other factors. The DSM-5 provides psychiatrists with the opportunity to choose between "adjustment disorder" and "PTSD" for patients living with violence, despair, and loss of unthinkable proportion. Unfortunately, the psychological difficulties of living with HIV are infinitely more complex than this.

Patients with HIV report more trauma and more severe trauma than the general population and at-risk controls. They have high rates of PTSD and anxiety, economic and social disenfranchisement, and high rates of interpersonal instability. Cognizance of the psychological issues in the care of patients with HIV is essential. The treatment of these problems is well within the scope of most clinicians in our field. The psychotherapy for patients infected with HIV can be broadly divided into three categories:

1. Psychotherapy for the problems of life encountered by patients infected with HIV
2. Psychiatric treatment for specific psychiatric conditions associated with HIV infections
3. Specific psychoeducational and psychotherapeutic interventions associated with specific types of interactions with HIV-infected patients

For numbers 1 and 2 above, the psychotherapeutic issues are largely similar to those seen in patients who are not infected with HIV. While specific papers have been written regarding coping skills, social isolation, grief, PTSD, intimate relationships, family problems, and many other issues, these recommendations are largely similar to those found in patients coping with other chronic medical conditions, or chronic impoverishment elsewhere in this text. The psychotherapy issues of conditions commonly seen in HIV clinics have been reiterated where appropriate, but are also similar to those issues for patients with the same conditions but without HIV infection. This chapter includes a section on

PTSD because of the high prevalence and special issues for patients with HIV.

Post-Traumatic Stress Disorder

Traumatic events such as those associated with actual or threatened death, serious injury, or sexual violence often provoke terror, anxiety, and stress and some individuals develop chronic psychiatrically disabling symptoms now called PTSD. Characteristics of the condition include (1) vivid recollection and re-experiencing the trauma (often in the form of nightmares or vivid flashbacks); (2) persistent avoidance of thoughts, places, people, or feelings associated with the trauma; (3) persistent negative emotional changes often with distortions of the event and its significance; and (4) persistent hyperarousal often including insomnia, irritability, and hypervigilance. PTSD has a prevalence of less than 1% in the general population and a lifetime prevalence of 1% to 9% with a female-to-male lifetime prevalence ratio of 2:1. In civilian populations, rape is the event most likely to produce PTSD, particularly if it occurs before or during adolescence.

PTSD increases the likelihood of engaging in destructive behaviors such as alcohol and other drug abuse, sexual promiscuity, or prostitution. PTSD is of particular concern in HIV treatment and research because it may engender or exacerbate HIV risk behaviors and worsen health outcomes. Cross-sectional research has shown that symptoms of PTSD have been associated with HIV risk behaviors and markers of HIV progression. Symptoms of PTSD in adolescence have been associated with prostitution, injection drug use, and choice of a high-risk sex partner in young adults. HIV-infected adults who have a history of childhood sexual or physical abuse have reported engaging in more HIV risk behaviors such as drug abuse and sexual compulsivity than persons with no history of trauma. A high prevalence (42%) of HIV-infected women attending county medical clinics had PTSD symptoms sufficiently severe to meet a diagnosis of PTSD. In patients infected with HIV, traumatic stressors and PTSD symptoms have also been associated with lower CD4-to-CD-8 T cell ratio at follow-up visits.

Veterans with a diagnosis of PTSD are at greatly increased risk of HIV infection, particularly if they are also diagnosed with a substance abuse disorder. Women prisoners with a lifetime history of PTSD are more likely to have engaged in prostitution and receptive anal intercourse prior to incarceration compared to women prisoners without PTSD. These studies of PTSD symptoms and HIV risk behaviors have been cross-sectional; thus, a causal relationship cannot be inferred. It may be that HIV risk behaviors, such as prostitution or drug abuse, increase exposure to trauma, and thus the likelihood of developing PTSD. Alternatively, PTSD that stems from early trauma may predispose an individual to engage in sex or drug behaviors that can increase the risk of HIV infection.

The presence of PTSD in an at-risk or HIV-infected positive individual is of particular concern because of high rates of comorbidity (up to 80%) with other psychiatric disorders. Specifically, PTSD is most often comorbid with depression and cocaine/opioid abuse—both risk factors for HIV. Prior depressive episodes may be either a risk factor for the development of PTSD following a traumatic event or a co-occurring response with PTSD to a traumatic event. Substance abuse may be either an attempt to "self-medicate" suffering after a traumatic experience or a lifestyle that increases exposure to traumatic events, such as robbery or assault.

PTSD and substance use disorders that occur together can also adversely affect treatment. Comorbid conditions have been associated with poorer treatment adherence and motivation, quicker relapse, more inpatient hospitalizations and medical problems, and lower global functioning than either disorder alone. Thus, treatment of PTSD that does not address coexisting depression or substance abuse may be insufficient or even worsen psychiatric status. PTSD treatment that typically involves behavioral exposure and flooding has been reported to exacerbate emotional arousal and precipitate relapse, although this had not yet been shown in experimentally controlled investigations. Likewise, the AA philosophy of surrender and sharing one's story may be counterproductive to substance-abusing HIV-infected individuals with PTSD.

HIV at-risk or HIV-infected individuals should be routinely screened for PTSD. Instruments such as the Trauma History Questionnaire and the PTSD checklist have increased detection rates of the disorder. Similarly, any individuals presenting with symptoms of PTSD should routinely be screened for depression and substance abuse. For HIV-infected individuals with PTSD and another concurrent psychiatric disorder, treatment that simultaneously addresses both disorders is likely to be more effective and practical.

HIV-Specific Psychotherapeutic Issues

There are a number of specific circumstances regarding HIV-infected patients that should be discussed here. They include the following:

1. Pretest, test, and posttest counseling issues
2. Risk behavior reduction in patients at risk or infected with HIV
3. Issues of partner notification in patients infected with HIV
4. Impaired patients with issues of capacity and competence
5. ART adherence issues

Pretest/Test Counseling and Education

Patients at risk of HIV infection are often reticent to get testing. Surveys suggest they fear the results of the test or are too overwhelmed by the issues of their current life and behavior to present for testing. Prior to 1993, patients who received a positive HIV test result saw their diagnosis as essentially fatal. In addition, many patients thought it would be burdensome to know that they were placing others at risk of infection. In more recent years with the advent of ART, HIV has become a chronic treatable illness. In this setting, it seems more reasonable for patients to get tested and engage in treatment. Nonetheless, survey data show that a large number of patients that are at significant risk are not getting tested. Psychoeducational psychotherapy directed at encouraging patients to get tested has been offered to a variety of at-risk populations of patients. Intervention studies using psychoeducational psychotherapy suggest that patients receiving therapy are more likely to get tested, more likely to receive a diagnosis, and more likely to begin treatment, particularly in settings where rapid-testng and same-day medication-start are offered.

Pretest counseling has been described in a number of papers. When HIV testing became available in the 1980s, a positive test was associated with an almost certain fatal outcome and HIV testing was regulated and required extensive pretest counseling and usually required a separate informed consent process. Now, most clinicians see HIV testing as a routine test with implications no worse than many other tests in medicine. Different forms of HIV testing are used and therefore patients should be informed about the particular elements of the test they are getting. The most rapid tests and the home testing kits test for antibodies to HIV, while more sophisticated tests also look for viral elements in blood. Depending on which test is used, results remain negative for a time after infection and therefore after a recent exposure, a patient may have a negative test but be in fact infected. Patients may also present immediately after an exposure and may benefit from postexposure prophylaxis treatment with ART. It is therefore important to ask patients about recent potential exposures if they come for testing.

Pretest counseling should also include information on safe sex, safe needle exchange, and other risk-reduction interventions as well as preventative PrEP. This is because a significant percentage of patients who obtain testing do not return for their results for extended periods of time and sometimes not at all. The development of rapid testing using cheek cell swabs has provided a much faster turn-around time, thus enabling a quicker result for the patient, but confirmation with a blood test is usually recommended. So, a combination of pre- and posttest counseling and then another pretest counseling session may be required in short order.

For patients with HIV infection, a number of monitoring tests are necessary, including CD4 counts and HIV RNA loads, and sometimes resistance testing. In similar fashion to the counseling required before HIV antibody testing, patients should be counseled before each of these types of tests, explaining what the test is for, what the possible results will be, and what the test results could mean as to the overall treatment and prognosis. Time should be allowed for questions, and many clinics have developed take-home pamphlets on these tests. The most important initiatives are to integrate mental health professionals in clinics for these types of visits in order to have wraparound type of services.

Posttest Counseling

A number of articles have described posttest counseling psychotherapy issues and interventions. These include both psychoeducational interventions regarding the meaning of test results, recommendations for treatment, and importantly, risk reduction interventions to stem the spread of HIV infection. These posttest interventions should occur in both HIV-negative and HIV-positive patients.

At the time of test results being given to patients, it is not uncommon for patients to have a variety of intense psychological reactions including suicidal feelings, anger, homicidal thoughts directed at potentially infecting partners, overwhelming grief, and complete psychological breakdown. Patients with poor coping skills, poor impulse control, history of suicidal feelings and behaviors, substance abuse disorders, and lack of social supports are at increased risk of impulsive behaviors and self-destructive behaviors. Because of these circumstances, availability of psychological interventions at the time of HIV testing and result provision is critical.

For established HIV-positive patients, bad news regarding progression of HIV disease or detection of resistance can provoke similar responses and should be considered in a similar way. Transition from the asymptomatic phase to the development of an opportunistic infection or to a formal diagnosis of AIDS because of a decline in CD4 cells may provoke denial, anger, depressive feelings, anxiety, hopelessness, or a myriad of other emotions. Again, the presence of psychological evaluation and intervention is extremely critical in this setting. Patients may be overwhelmed by the news that they need to start antiretroviral drugs, and therefore need emergent attention at this time. In particular, discussions around detection of viral resistance may evoke a number of defensive responses from the patient, as there may be a perception of failed adherence and guilt.

Psychotherapy to Prevent HIV Transmission in Selected Populations of Patients

Men who have sexual contacts with other men may either be exclusively homosexual, bisexual, or heterosexual men. Men with sexual contacts with other men were the largest subgroup in terms of new AIDS diagnoses in the United States in the year 2000. In states where HIV is reported, it continues to be the largest subgroup of newly reported HIV infections. In intervention studies looking at men who have sex with other men, many interventions have shown a decrease in either risk behaviors or infection. Studied psychotherapeutic interventions focus on education, behavioral change, management of emotions, and assertiveness including some tailored to specific high-risk populations or environments. Although there are fewer data, similar studies have been done targeting heterosexually transmitted HIV, substance-related risk behaviors, women, and IV drug users. It is unclear from the data what the best intervention is and how to stratify the interventions. The use of PrEP in this at-risk patient population has been a big risk reduction intervention now especially when taken correctly.

Studies of rates of psychiatric disorders in at-risk population show impressively high rates of affective disorders, substance use disorders, personality disorders, and psychological distress. As of now, no systematic study on treatment and targeted intervention methods based on psychiatric diagnosis has been reported. It is clear from the data on risk and epidemiology that this is the direction that needs to be taken to try to improve prevention.

Partner Notification

The landmark legal decision in the Tarasoff case in California has resulted in the increased scope of responsibility for health care providers. A variety of legislation differing from state to state has afforded practitioners an increased number of options with regard to confidential situations. In some states, Tarasoff statutes (those statutes providing a duty to warn vulnerable individuals of imminent danger from a patient overriding issues of confidentiality) have completely changed the way in which mental health professionals handle confidential issues.

Numerous articles have been written about the issues of ethics, confidentiality, duty to warn, and medical/legal aspects of this element of practice. Although no clear consensus has been reached, recommendations are that patients who are sexually active and infected with HIV be counseled about potential risk to their sexual partners. In addition, known partners should be notified of exposure risk and potential infection as well. Partner notification has been an extremely hotly debated topic. However, many states have developed legislation requiring or allowing either physicians or health department officials to notify partners of HIV-infected patients of their risk. The current standard, despite the controversy, appears to be an obligation on the part of health care professionals (in some areas, health departments) to ensure notification to anyone who could be construed as clearly at risk and who may be unaware of their risk.

A particularly difficult situation is that of sex workers, known to be HIV infected and known to be working actively as prostitutes. There are public health issues that pose a risk both for these patients and depending on the politics of the circumstances for their potential partners, clients, customers, victims, or victimizers. The responses to this problem have ranged from a sense that sex care workers and their clients can make their own decisions and should be responsible for their own behavior all the way to the sentiment that HIV-infected sex workers should be arrested and jailed for attempted murder. It has additionally been noted that some sex workers are impaired by a variety of psychiatric conditions including cognitive impairment, major mental illness, personality disorder, and substance use disorders. These may further contribute to the sense that some sex workers may be less than fully responsible for their behavior. Recommendations have been made for voluntary and involuntary interventions regarding these patients. Specific psychiatric interventions regarding competency, ability to consent, capacity, and most importantly treatment for the conditions that impair such people are critical to the mental health needs of patients with HIV.

Capacity to Consent/Competence

Patients with psychiatric disorders in HIV clinics often have a variety of difficulties with medical care provision. In some of these settings, the patient’s capacity or competence to make medical decisions regarding their health care can be in doubt. Provision of mental health care evaluations to determine this is an often-unmet need in HIV clinics. The issues of competence and capacity in these patients are often no different than those described elsewhere in this textbook; however, the consequences of inadequate assessment of patients’ competency and capacity can have grave consequences in this clinical setting. HIV dementia and delirium can be overlooked in this population, as can intoxication with substances, creating problems in obtaining meaningful consent.

The question of capacity and consent involves several issues, but most important is the specific question that the patient is being asked. In many cases, medical care providers want a judgment about competence as a general rule, a judgment impossible to make. To have capacity to give a particular informed consent, a patient must understand that there is a decision before them regarding some aspect of care, and must understand the consequences not only of each option but also of refusal to make a choice. The patient should be able to repeat the benefits and risks of each possible option and must be able to clearly communicate the decision and have the ability to maintain the choice made over time. Finally, in the process of making the decision, the patient must be able to manipulate the information in a rational way. Thus, patients who are delirious or intoxicated may not maintain choices over time as concentration and memory are often impaired. Demented patients may have difficulty maintaining choices due to impaired memory as well. Psychotic patients may not base decisions in reality, occasionally arriving at choices that derive from delusional constructs. While physicians from many disciplines may be able to determine capacity in many situations, it is in these patients with psychotic illness that psychiatrists are often especially helpful.

In many cases, the danger to the person, patterns of prior behavior, severity of illness, poor judgment, and psychiatric vulnerabilities complicate these decisions and play an important role in tempering the way in which patients are managed. The ethics of a particular case may get very complex when a patient understands the issues cognitively, but their judgment is colored by their affective state, temperament, drug cravings, social situation, or simply difficulty with tolerating discomfort. These cases often divide medical teams and require consultation and a group conference to resolve. It is critical to get all providers to discuss the most difficult cases, clarify the issues, and come to a decision based on the patients' best interests, not the most expedient management.

Adherence Counseling

The single most important factor regarding outcome of HIV treatment is the patient's ability to adhere to prescribed regiment. While this has been debated in the literature, a study by Margaret Fischel looking at HIV-infected prisoners revealed that 100% of patients who received directly observed therapy in a prison setting developed undetectable viral loads. This strongly supports adherence as the major feature of treatment. There are compelling studies suggesting that major depression, substance abuse, personality disorder, and psychosocial disruption all affect adherence. Interventions to address these conditions are presented above. More subtle factors affecting adherence include psychosocial support networks, individual coping skills, life structure, access to resources, and behavioral control. Interventions such as cognitive behavioral psychotherapy, structured psychoeducational psychotherapy, supportive psychotherapy, and group interventions have all been used to improve patient adherence to office visits and medication regimens.

In the current literature on HIV, medication adherence focuses on technical interventions such as pill box and timer reminders, less complex pharmacologic interventions, decreased pill burden, and increased access to care. A growing literature examines psychosocial interventions, relationships with health care providers, case management, and psychiatric disorders as barriers to adherence. It is in this arena that mental health care can have an enormous impact on outcome. Psychotherapy has been shown to improve clinic visit adherence, the best indirect predictor of medication adherence.

NEUROLOGIC COMPLICATIONS OF HIV AND AIDS

Opportunistic Infections

Toxoplasmosis. *Toxoplasma gondii is* a protozoan acquired most commonly from cat feces or uncooked meat. Approximately a third of people are latently infected with toxoplasmosis, but in the United States, it is approximately 11%. In immunocompetent hosts, the organism is encysted and infection is usually clinically silent unless the person becomes infected during pregnancy, where there can be catastrophic results. Recent studies have shown substantial influences on personality and behavior in aggregate populations suggesting infection may be less silent than previously thought. Active infection in patients with HIV generally occurs in patients who are immunocompromised with less than 200 CD4 cells/mL. In patients with AIDS, toxoplasmosis is the most common reason for intracranial masses, affecting between 2% and 4% of the AIDS population, with the most common findings being pulmonary and ocular involvement. Other manifestations are possible, including chorioretinitis and pneumonitis. Lymphadenopathy may be present in cutaneous cases. Symptoms of CNS infection are fever, change in level of alertness, headache, focal neurologic signs (approximately 80% of cases), and focal or generalized seizures (approximately 30% of cases). CT and MRI scans usually show multiple, ring-enhancing lesions in the basal ganglia or at the gray–white matter junction. CSF studies are normal in 20% to 30% of cases, but more often show a mild monocytosis. Serum *T. gondii* IgG is generally helpful in the diagnosis, but has a false-negative rate of 5% to 10%. Brain biopsy provides the definitive diagnosis, but since this invasive procedure carries some risks, empiric treatment is often offered if the clinical and radiographic pictures suggest infection.

The preferred treatment consists of sulfadiazine and pyrimethamine because with this combination, patients do **not** require additional trimethoprim-sulfamethoxazole (TMP-SMX) for prophylaxis against *P. jirovecii* infections. In the care of a sulfadiazine allergy, clindamycin in addition to pyrimethamine **plus** leucovorin is an alternate treatment option. There are other alternate regimens such as use of dapsone plus pyrimethamine and leucovorin in case of intolerance. Clinical and radiologic improvement is seen in 75% to 80% of patients. Since these medications are effective only against the tachyzoite form of the protozoan, maintenance therapy using a combination of sulfadiazine and pyrimethamine should be continued, but at lower doses than those used for initial treatment as the goal is to prevent recrudescence. The use of trimethoprim-sulfamethoxazole as prophylaxis has reduced the incidence of *T. gondii* infection. Patients with hypersensitivity to sulfa drugs may use pyrimethamine plus dapsone for maintenance.

Cytomegalovirus. CMV infection is an uncommon complication of AIDS and prior to ART, presented in about 2% of HIV patients before the advent of ART. The development of clinically evident CMV encephalitis is fairly rare and most usually occurs only in patients with CD4 counts less than 50 cells/μL with close to 80% reduction in cases since the development of ART. Symptoms of CMV neurologic disease are dependent on the anatomic location involved. There have been several cases of immune reconstitution inflammatory syndromes (IRIS) reported after start of ART. These patients often have presentations consistent with encephalitis and CNS involvement of infection. The most common symptoms in CMV encephalitis include delirium, confusion, apathy and focal neurologic deficits. Other manifestations involve myelitis, polyradiculopathy, and peripheral neuropathy.

Investigation of CMV encephalitis begins with examination for signs of CMV infection of the retinae, electrolyte studies to look for adrenal insufficiency, and viral blood cultures. CT scan may show ventriculomegaly or decreased attenuation diffusely throughout the parenchyma. MRI may show increased signal intensity around the ventricles. CSF studies may be normal or show high protein, low glucose, and pleocytosis. CSF CMV cultures are usually negative but PCR may reveal the presence of the virus. Brain biopsy may be needed for a definitive diagnosis.

Treatment consists of getting patients on optimal ART but otherwise depends on neurologic involvement and severity of disease. Ganciclovir and foscarnet have been the primary treatment for CNS CMV, but are not necessarily beneficial. Cidofovir is an alternative therapy especially if patients cannot tolerate any of the other above options.

Cryptococcal Meningitis. While meningitis caused by *Cryptococcus neoformans* is rare in immunocompetent persons, it was one of the leading opportunistic infections in patients with untreated HIV. The incidence is now around five cases per 1,000 PLWH and it is usually seen in untreated patients with advanced HIV. Patients generally present with fever and delirium. Meningeal signs (headache, stiff neck, photophobia, and nausea) are not universally seen. Seizures and focal neurologic deficits occur in about 10% of patients. CT scans are normal, but gadolinium-enhanced MRI may show meningeal inflammation. Intracranial pressure is elevated in 50% of patients. CSF studies are normal in about 20%, but otherwise show mild-to-moderate monocytosis, elevated protein, decreased glucose, and positive fungal cultures. The fungus can be seen on India ink stain of CSF about 60% to 80% of the time. There is also a test for *C. neoformans* antigen, which is usually positive in both serum and CSF.

Treatment of cryptococcal meningitis requires amphotericin B and flucytosine. Patients who survive must receive prophylaxis against recurrence, since this is very common. Some authors suggest that patients who receive ART for 6 months with a rise in CD4 count to >100 cells/mL may terminate secondary prophylaxis. Prophylaxis can be prescribed as oral fluconazole or intermittent IV amphotericin B. Primary prophylaxis for *C. neoformans* is not recommended.

Progressive Multifocal Leukoencephalopathy. PML is a demyelinating disease of white matter in immunocompromised patients. First described in cancer patients, the causative agent is a polyomavirus, named JC virus after a patient (not to be confused with Creutzfeldt–Jakob disease, caused by a prion). Its transmission route is unclear but may be respiratory and the virus is commonly present in an asymptomatic form in immunocompetent people. The prevalence of PML prior to ART in AIDS patients was between 1% and 10%. Typically, PML affects patients with AIDS with fewer than 100 CD4 cells/μL. Although rare, it has been reported in patients after initiation of ART with increasing T-4 cell counts.

The pathology of PML consists of demyelination and death of astrocytes and oligodendroglia, with a multifocal presentation. The clinical syndrome consists of multiple focal neurologic deficits, such as mono- or hemiparetic limb weakness, dysarthria, gait disturbances or sensory deficits, and progressive dementia, with eventual coma and death.

Occasionally, there may be seizures or visual losses. There is usually no fever or headache.

MRI is more useful than CT in diagnosis, displaying multiple areas of attenuated signal on T2 images primarily in the white matter of that brain, although gray matter, brain stem, cerebellar, and spinal cord lesions are possible. CSF studies are generally unhelpful, except for PCR evaluation for the presence of JC virus, which is sensitive and specific. Brain biopsy provides the definitive diagnosis, but is rarely used. There is no specific ART for JC virus. Treatment of PML includes support for the patient and ART. There are no data suggesting that higher CNS-penetrating antiretrovirals are of any particular benefit, and this logical conclusion is often followed by providers. The prognosis is poor for patients with PML but the course is more variable in the small group of patients who develop PML after starting ART.

CNS Neoplasms

There has been much data showing that cancer is diagnosed in over 40% of HIV patients with over 28% of HIV-related deaths being caused by some sort of malignancy. Lymphoma, specifically non-Hodgkin lymphoma (NHL), is the most common neoplasm seen in patients with AIDS, affecting up to 10% of patients. AIDS is the most common condition associated with NHL and systemic NHL accounts for most of the HIV-related lymphoma cases. The patient is generally afebrile and may develop a single lesion with focal neurologic signs or small, multifocal lesions most commonly presenting with a mental status change. Seizures are present in about 15% of patients. CNS lymphoma is at times misdiagnosed as toxoplasmosis, HIV dementia, or other encephalopathy. CT scan of the brain may be normal or show multiple hypodense or patchy, nodular enhancing lesions. MRI generally shows enhanced lesions that may be difficult to differentiate from CNS toxoplasmosis, but thallium SPECT scanning may help differentiate the two disorders and is 90% sensitive and specific for lymphoma. CSF studies may be normal or show a moderate monocytosis; cytology studies reveal lymphoma cells in less than 5% of patients. Brain biopsy is required for the confirmation of the diagnosis of CNS lymphoma. As this procedure carries some morbidity, clinicians should weigh the clinical presentation carefully, suspecting lymphoma in afebrile patients with a negative toxoplasma IgG screening test, patients with a single lesion, and patients who fail to respond to empiric therapy for toxoplasmosis as demonstrated by clinical examination and repeat MRI at 2 weeks. The differential diagnosis of CNS neoplasm also includes metastatic Kaposi sarcoma and primary glial tumors. Lymphoma may respond in part to radiation therapy and steroids, thus alleviating high intracranial pressure and its associated symptoms. Chemotherapy is generally adjunctive for lymphoma. While CNS lymphoma had a grim prognosis with an average survival of 3 to 5 months prior to the advent of ART, the prognosis is now dependent on the ART response, with considerable improvement possible in patients who respond to ART.

Direct CNS Manifestations of HIV

1. Guillain–Barré syndrome. A small percentage of patients, usually young men, will present with Guillain–Barré, a syndrome associated with early HIV infection. Guillain–Barré syndrome is an inflammatory demyelinating polyneuropathy causing symmetrical paralysis and few, if any, sensory symptoms, usually beginning in the lower extremities and progressing upward. The condition becomes especially serious if abdominal musculature is involved, as it may impair respiration. The disorder is thought to be autoimmune in etiology and is generally self-limited. IV immunoglobulin and plasmapheresis have been used to shorten the course, but neither treatment has been studied well in HIV-infected individuals.
2. Vacuolar myelopathy. Vacuolar myelopathy is highly prevalent among patients with AIDS, being found in approximately 50% of patients at autopsy. Clinical manifestation of this disease is much less common, affecting 20% to 30% of end-stage patients with AIDS. The presence of vacuolar myelopathy has been associated with history of *P. carinii* and *Mycobacterium avium-intracellulare* infections, suggesting that the development of vacuolar myelopathy is related to more severe immunosuppression. The mechanism of the disease is unclear, but appears similar to the myelopathy of combined systems disease associated with vitamin B_{12} deficiency. Multinucleated giant cells are seen on histologic examination, and theories about mechanism focus on immunologic activation damage, direct toxicity of HIV products, and metabolic dysfunction of transmethylation processes.

The clinical manifestations of vacuolar myelopathy appear when the disease progresses to affect the lateral and posterior columns, and thus includes spastic paraparesis, loss of proprioception and vibration sense, bowel and bladder urgency or incontinence, and impotence. There are now reports of improvement with the initiation of ART.

Peripheral Nervous System Disorders in HIV

Peripheral Neuropathy. Patients with HIV infection may develop peripheral neuropathy, most often in the feet but occasionally in the hands. The neuropathy may range from paresthesia to burning pain, and patients will have a vibratory-sense gradient with decreased sensation in the distal extremity compared to more proximal points. The incidence of peripheral neuropathy increases as HIV disease progresses, but cases are found in patients with well-preserved immune function. Treatment of peripheral neuropathy may include TCAs, pregabalin, gabapentin, or other antiepileptic drugs used to treat neuropathic pain. Opiate analgesics should be used sparingly, and longer-acting agents are preferable, since long-term use of opiates presents the risk of eventual tolerance and dependence. Benzodiazepines are of no use.

Special Issues in HIV

Fatigue. Fatigue is a common symptom in HIV-infected patients, which is often overlooked, improperly assessed, or inadequately investigated. Several authors have commented on the high prevalence of fatigue as a symptom of HIV infection, especially in later stages. Fatigue may be mild and annoying, or it may be severe enough to impair function. Several scales have been published to assess fatigue symptoms and severity. Fatigue is a nonspecific symptom and may have a single or multifactorial etiology. Medical causes include pneumonia, bronchitis, hypothyroidism, hepatitis, heart failure, renal failure, many cancers, and myopathy among others. In a sample of ambulatory patients with AIDS, fatigue is significantly correlated with anemia and pain. In addition to disease causes, patients may present with fatigue as a side effect of medications such as antihypertensives, anticonvulsants, benzodiazepines, antidepressants, narcotic analgesics, antipsychotics, antiemetics, antihistamines, and most importantly for patients with HIV, ARTs. In fact, fatigue has been found to be one of the most common side effects of PIs and may be a reason for nonadherence.

Fatigue may also be the result of a psychiatric disorder. Alcohol and substance use disorders may lead to fatigue either related to the use or withdrawal of the substance or as a symptom of demoralization in addicts. Most importantly, fatigue is caused by major depression. Patients with HIV and major depression are much more likely to complain of fatigue than patients without depression. Many depression screening tools, such as the BDI and Hamilton Depression Rating Scale have not been very useful in distinguishing fatigue from major depression usually because fatigue symptoms are present on the screening tools.

In general, the evaluation of a patient complaining of fatigue should include a careful history of its temporal characteristics, severity, and associated symptoms. It should also include careful review of current and recent medications, physical examination, and a mental status examination. The latter should carefully examine for anhedonia, diminished sense of self-worth, guilty feelings, sleep disturbance—especially early morning awakening with inability to return to sleep, appetite changes—a recent, more than 5% change in body weight, thoughts of death or suicide, and impairments in concentration or memory. Certain laboratory studies should also be obtained, including complete blood count, electrolytes, liver function tests, oxygen saturation, and thyroid function tests.

If fatigue is thought to be related to a medical or medication cause, all attempts should be made to treat the illness or modify the medication so as to alleviate the fatigue. In this context, testosterone is a successful treatment of fatigue in HIV-infected men even when depressive symptoms are present. Of course, a clinical major depression should be treated with standard therapies as discussed above. More activating antidepressants such as fluoxetine, escitalopram, venlafaxine XR, or bupropion SR or XL may be better tolerated by fatigued depressed patients. Some authors have reported that dextroamphetamine may be useful in treating fatigue and depression in HIV. Care must be exercised in using stimulants as long-term use may lead to dependence or worsening depression on some cases.

HIV/HCV Coinfection

Hepatitis C virus (HCV) is a blood-borne pathogen that is currently most commonly transmitted by injection drug use, but may be transmitted sexually, although far less commonly than HIV. Some studies have reported that up to 30% of HIV-infected patients are also infected with HCV in the United States and Western Europe. The natural history of HCV infection in HIV-negative individuals is that 15% of patients clear the infection after the acute phase, while 85% progress to a chronic infection. Hepatic fibrosis develops, often requiring about 10 years to reach significant levels, with cirrhosis following about 20 years from time of infection. Chronic HCV infection is the most common etiology of hepatocellular carcinoma (HCC), which usually develops about 30 years after infection in HIV-negative patients. Unfortunately, HIV infection is likely to make individuals more susceptible to contract HCV if exposed, likely due to immunosuppression and also to cause more rapid progression of liver disease, cutting the above approximate timetable in half (i.e., fibrosis in 5 years, cirrhosis in 10, HCC in 15). The leading causes of mortality in HIV-infected patients are hepatic diseases, most often related to HCV infection.

Chronic hepatitis C is associated with increased rates of depression. In addition, treatment with interferon—ribavirin combinations (previously used) increases risk of major depression and can produce mania. In particular, it appears that interferon-alpha has been associated with depressive syndromes, suicide, and, on rare occasions, mania. Patients with pre-existing depression or bipolar disorder are more likely to develop affective symptoms while receiving the drug, but may not be more likely to stop treatment than patients developing these symptoms de novo. Further, depressive symptoms associated with interferon-alpha have been successfully treated with both SSRIs and TCAs. Because of the development of direct-acting antivirals that target and cure hepatitis C in greater than 90% of patients without the use of interferon, the current preferred antiviral treatment option for most patients includes interferon-free regimens like ledipasvir-sofosbuvir (Harvoni), a combination of an NS5A inhibitor and NS5B inhibitor typically taken for an average of 12 to 24 weeks, depending on severity of disease.

Coinfected HIV/HCV patients should be screened for the presence of psychiatric disturbance like any other patient with HIV, but special monitoring should be performed especially if the patient is being treated with interferon-containing drugs for the purpose of early recognition and treatment of affective symptoms. Alcohol hastens the progression of HCV disease and should be strongly discouraged in any amount. Drug use, in general, exacerbates neuropsychiatric symptoms and patients should be stabilized prior, during, and after antiviral treatment.

CONCLUSIONS

HIV and AIDS are conditions intimately linked to psychiatry. In a sense, psychiatric disorders can be seen as vectors of HIV transmission and additionally complicate the treatment of HIV. Also, HIV produces a number of psychiatric conditions and exacerbates many others. The authors have attempted to show the significant comorbidity and links between various types of psychiatric conditions: the way depression exacerbates addiction, the way personality disorder exacerbates addiction, and the way in which addiction exacerbates both personality vulnerabilities and depression. HIV disease is driven by behaviors that are intimately connected with all these conditions.

The authors see the integration of psychiatry into HIV care as a model that psychiatry may apply when interacting with general medicine and the health care system in general. There is abundant data showing the need for presence of psychiatry in every phase of HIV care, and yet, sadly, there is mostly inadequate funding that limits the availability of psychiatric care in HIV clinics.

The authors' experience in caring for patients with HIV is that by developing a comprehensive diagnostic formulation on which to base treatment, especially in an interdisciplinary setting, they have significant success with even challenging clinical scenarios. The authors' formulation includes disease syndromes such as major depression and schizophrenia, personality vulnerabilities such as unstable extraversion, substance use disorders/addiction, and problems with life experience such as trauma and trust issues. Each problem has the potential to sabotage treatment for all the remaining conditions. The treatment plan must be comprehensive in scope in order to address the whole person.

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the contribution by Dr Joyce E King whose suggestions and edits were essential to this chapter.

We would also like to thank the Reginald F Lewis Foundation for their generous support of our mission and work.

FURTHER READINGS

Afacan Y, Chaudhry F, Santangelo V, Shkolnik A. They cannot hear, talk, and see in sensible ways: the impact of personality disorders on HIV and other sexually transmitted infections. *J Natl Med Assoc*. 2015;107(2):60–73.

Arseniou S, Arvaniti A, Samakouri M. HIV infection and depression: HIV infection and depression. *Psychiatry Clin Neurosci*. 2014;68(2):96–109.

Batkis MF, Treisman GJ, Angelino AF. Integrated opioid use disorder and HIV treatment: rationale, clinical guidelines for addiction treatment, and review of interactions of antiretroviral agents and opioid agonist therapies. *AIDS Patient Care STDS*. 2010;24(1):15–22.

Bositis CM, Louis JS. HIV and substance use disorder: role of the HIV physician. *Infect Dis Clin North Am*.2019;33(3):835–855.

Brooner RK, Greenfield L, Schmidt CW, Bigelow GE. Antisocial personality disorder and HIV infection among intravenous drug abusers. *Am J Psychiatry*. 1993;150(1):53–58.

Centers for Disease Control and Prevention. HIV infection risk, prevention, and testing behaviors among men who have sex with men—National HIV Behavioral Surveillance, 23 U.S. Cities, 2017. HIV Surveillance Special Report 22. Published February 2019. https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html

Ciccarelli N. Considerations on nosology for HIV-associated neurocognitive disorders: it is time to update? *Infection*. 2020;48(1):37–42.

Clifford DB. HIV-associated neurocognitive disorders: epidemiology, clinical manifestations, and diagnosis. In: Gandhi R, ed. *UpToDate*. Updated March 31, 2023.

Fazeli PL, Cheatwood JD, Hopkins C, et al. Association between personality and cognitive performance in middle-aged and older adults with human immunodeficiency virus (HIV). *Appl Neuropsychol Adult*. 2023;30(3):279–288.

Gardner EM, McLees MP, Steiner JF, Del Rio C, Burman WJ. The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. *Clin Infect Dis*. 2011;52(6):793–800.

Ghandi T. Toxoplasmosis in patents with HIV. In: Sax P, ed. *UpToDate*. Updated March 31, 2023.

Global HIV & AIDS statistics—Fact sheet. 2022. UNAIDS (The Joint United Nations Programme on HIV/AIDS). Accessed July 3, 2023. https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf

Grover S, Sahoo S, Chakrabarti S, Avasthi A. Post-traumatic stress disorder (PTSD) related symptoms following an experience of delirium. *J Psychosom Res*. 2019;123:109725.

Heaton RK, Clifford DB, Franklin DR Jr, et al; CHARTER Group. HIV-associated neurocognitive disorders persist in the era of potent antiretroviral therapy: CHARTER Study. *Neurology*. 2010;75(23):2087–2096.

Hernández-Huerta D, Parro-Torres C, Madoz-Gúrpide A, Pérez-Elías MJ, Moreno-Guillén S, Ochoa-Mangado E. Personality and adherence to antiretroviral therapy in HIV-infected adult patients. *J Psychosom Res*. 2021;144:110413.

Hutton HE, Treisman GJ, Hunt WR, et al. HIV risk behaviors and their relationship to posttraumatic stress disorder among women prisoners. *Psychiatr Serv*. 2001;52(4):508–513.

Ickovics JR, Hamburger ME, Vlahov D, et al; HIV Epidemiology Research Study Group. Mortality, CD4 cell count decline, and depressive symptoms among HIV-seropositive women: longitudinal analysis from the HIV Epidemiology Research Study. *JAMA*. 2001;285(11):1466–1474.

Keiser O, Spoerri A, Brinkhof MWG, et al; Swiss HIV Cohort Study, Swiss National Cohort. Suicide in HIV-infected individuals and the general population in Switzerland, 1988–2008. *Am J Psychiatry*. 2010;167(2):143–150

Mayer KH, Krakower D. Administration of pre-exposure prophylaxis against HIV infection. In: Gulick R, ed. *UpToDate*. Updated February 24, 2023.

Peters PJ, Pontones P, Hoover KW, et al; Indiana HIV Outbreak Investigation Team. HIV infection linked to injection use of oxymorphone in Indiana, 2014–2015. *N Engl J Med*. 2016;375(3):229–239.

Pieper AA, Treisman GJ. Depression, mania, and schizophrenia in HIV-infected patients. In: Silver J, ed. *UpToDate*. Updated March 13, 2023.

Pieper AA, Treisman GJ, Overview of the neuropsychiatric aspects of HIV infection and AIDS. In: Silver J. *UpToDate*. Updated October 19, 2021.

Pieper AA, Treisman GJ. Substance abuse and addiction in HIV-infected patients. In: Silver J, ed. *UpToDate*. Updated March 15, 2023.

Rzeszutek M, Gruszczyńska E. Personality types and subjective well-being among people living with HIV: a latent profile analysis. *Qual Life Res*. 2020;29(1):57–67.

Shuper PA, Joharchi N, Rehm J. Personality as a predictor of unprotected sexual behavior among people living with HIV/AIDS: a systematic review. *AIDS Behav*. 2014;18(2):398–410.

Tang C, Goldsamt L, Meng J, et al. Global estimate of the prevalence of post-traumatic stress disorder among adults living with HIV: a systematic review and meta-analysis. *BMJ Open*. 2020;10(4):e032435. https://doi.org/10.1136/bmjopen-2019-032435

Wagner EE, Miller AL, Greene LI, Winiarski MG. Dialectical behavior therapy for substance abusers adapted for persons living with HIV/AIDS with substance use diagnoses and borderline personality disorder. *Cogn Behav Pract*. 2004;11(2):202–212.

Watkins CC, Pieper AA, Treisman GJ. Safety considerations in drug treatment of depression in HIV-positive patients. *Drug Saf*. 2011;34(8):623–639.

Weiss NH, Tull MT, Dixon-Gordon K, Gratz KL. Assessing the negative and positive emotion-dependent nature of risky behaviors among substance dependent patients. *Assessment*. 2018;25(6):702–715.

Wejnert C, Hess KL, Hall HI, et al. Vital signs: trends in HIV diagnoses, risk behaviors, and prevention among persons who inject drugs—United States. *MMWR Morb Mortal Wkly Rep*. 2016;65(47):1336–1342.

2.9 Neuropsychiatric Aspects of Other Infectious Diseases (Non-HIV)

JESSICA HARDER, M.D., RISHAB GUPTA, M.B.B.S., M.D., RUTH MCCANN, M.D., AND IRENE GONSALVEZ, M.D.

Infectious disease represents a critical and yet often overlooked contributor to nervous system malfunction and neuropsychiatric symptoms. All major classes of infectious agents, including viruses, fungi, bacteria, protozoa, and helminths can have neuropsychiatric manifestations, whether via direct neurotropism, associated immune or vascular processes, or some combination (Table 2.9–1). This chapter will address some of the infectious conditions with the most prominent central nervous system (CNS) effects (with the exception of HIV/AIDS and prion disease which are addressed separately).

VIRUSES

COVID-19

Epidemiology. COVID-19, or coronavirus disease 2019, is a viral illness caused by severe acute respiratory syndrome coronavirus 2, or SARS-CoV-2. SARS-CoV-2 is a positive-sense, single-stranded RNA virus, belonging to the family *Coronaviridae* (see Table 2.9–1). The first case of this disease was reported from Wuhan, China in December 2019, and it was declared a global pandemic by March 2020. At the time of this writing, it had infected 380 million people worldwide, killing 5.6 million of them. The United States has been one of the worst affected countries by this virus, with more than 75 million people being infected and more than 885,000 deaths.

Pathophysiology. Despite vigorous research efforts, the pathology of COVID-19 is still not fully understood. SARS-CoV-2 is known to enter various cells of the body, mainly in the upper and lower respiratory tract epithelium and endothelium via angiotensin-converting enzyme 2 (ACE-2) receptors, which have moderate to high expression in a wide variety of tissues throughout the body. Despite being a vital entry point for SARS-CoV-2, ACE-2 expression does not account for the significant interindividual variability in the clinical presentation and outcomes seen in the patients. SARS-COV-2 quickly replicates in the infected cells, causing cellular damage directly, and indirectly via generalized inflammation, and through alterations in blood supply via endothelial dysfunction. In some patients, it also leads to a cytokine storm, an aggressive, systemic inflammatory response, sometimes leading to multiorgan dysfunction. SARS-CoV-2 is known to affect nearly every major organ in the body, including the lungs, heart, brain, kidneys, GI system, and skin. ACE-2 receptors are expressed at low levels in neurons, but they are highly expressed in the pericytes, which are derived from neural crest stem cells, physically connect endothelial cells and astrocytes, and help to maintain the blood–brain barrier. SARS-CoV-2 RNA has been found in human brain tissues and the viral proteins have been detected in the endothelial cells within the olfactory bulb among the patients who died due to COVID-19. Converging evidence for the direct neuroinvasion is slowly emerging from the human brain organoid (self-organizing, 3D cell cultures that histologically resemble the human brain) models, mouse models, and human autopsy studies. Direct neuroinvasion may lead to local hypoxia and vessel dysfunction, which, in turn, may lead to ischemic infarcts and enhance the susceptibility of local regions to viral invasion. A recent study using human brain organoid demonstrated that pericytes may act as “trojan horses,” carrying the virus to the neurons and glia. Various other mechanisms through which the virus may enter the brain are through a leaky blood–brain barrier, direct penetration through olfactory neurons, and axonal retrograde transport of the viral particles through nerve fibers. Cytokine increase and microglia activation lead to increased kynurenine, quinolinic acid, and glutamate, and reduction of monoamines and neurotrophins. These are some biochemical underpinnings of how COVID-19 may lead to occurrence of new-onset neuropsychiatric symptoms or exacerbate preexisting neuropsychiatric disorders.

Clinical Presentation. Clinical severity of COVID-19 is quite variable, ranging from asymptomatic infection to acute respiratory distress syndrome (ARDS). In more than 90% of the patients, COVID-19 presents with a mild upper respiratory or a flu-like illness, with fever, cough, malaise, nasal congestion, sore throat, myalgias, arthralgias, fatigue, anosmia, ageusia, and headaches, with shortness of breath seen in moderate to severe cases (around 15–20% of cases). Some patients present with GI symptoms, like nausea, vomiting, abdominal pain, and diarrhea. A large number (up to 40%) of patients may develop rashes. It has protean clinical manifestations due to its ability to affect practically any organ system. Among patients hospitalized due to COVID-19, relatively common extrapulmonary complications include delirium (55%), hepatic dysfunction (characterized by transaminitis, hyperbilirubinemia, seen in 14% to 53%), venous thromboembolism (31% to 40%), heart failure (23%), myocarditis (8% to 28%), arrhythmias (16%), acute renal failure (1% to 46%). Some of the infrequently reported extrapulmonary presentations associated with COVID-19 are myocardial infarction (0.33%), ischemic stroke (0.86% to 1.3%), encephalitis, seizures, neuropathy, myopathy, Guillain–Barré syndrome, coagulopathy, and precipitation or worsening of diabetes mellitus or thyroid dysfunction. In around 5% of patients, initial respiratory symptoms complicate into a severe disease, requiring oxygenation, ICU admission, mechanical ventilation, and sometimes, intubation. Between 0.1% and 10% of people infected with COVID-19 die, with the variability explained by patients' age, sex, comorbidities, ethnicity, and quality of and access to healthcare.

There seems to be a strong relationship between COVID-19 and psychiatric disorders. Available literature demonstrates a bidirectional relationship between the two, such that COVID-19 increases the risk of various psychiatric illnesses and the latter increases the risk of COVID-19 infection and associated mortality. In a study utilizing a nationwide database of electronic health records of 61 million adult patients across all 50 states in the United States, patients recently diagnosed with a mental disorder had an elevated risk for COVID-19 infection. This effect was strongest for those diagnosed with depression and schizophrenia. Female sex and African American ancestry were important covariates in this relationship. COVID-19 infection among those with a recent psychiatric illness diagnosis was associated with higher death rate and hospitalization rate compared to those with no psychiatric disorder. Similarly, patients with dementia are more vulnerable to SARS-CoV-2 infection and are more likely to develop severe disease compared to other individuals.

Table 2.9–1.
Infectious Diseases with Neuropsychiatric Manifestations

Disease	Infectious Agent	Mode of Infection	Incidence	Typical Neuropsychiatric Symptoms	Diagnosis
VIRUSES					
COVID-19	RNA Virus, *Coronaviridae* family (SARS-CoV-2)	Inhalation; droplet and airborne spread	Variable; 380 million people worldwide affected so far	Delirium, depressive symptoms, anxiety symptoms, PTSD symptoms, insomnia, fatigue, cognitive symptoms; rarely, psychosis, catatonia, mania	RT-qPCT assay from nasal swab
Chikungunya	RNA Virus, *Togaviridae* family (Chikungunya virus)	*Aedes* mosquito bite	Wide variability across countries; >200,000 cases in 2021 (mostly from Brazil and India)	Encephalitis, encephalopathy in adults Meningoencephalitis, developmental delay in affected neonates	Clinical syndrome, ± RT-PCR (week 1), ELISA for IgM antibodies later
Western equine encephalitis (WEE)	RNA virus, *Togaviridae* family (WEE virus)	*Culex* mosquito bite	1–2 cases per year in the United States, rarely occurs in the rest of countries in the Americas	Encephalitis (worse in infants)	Detection of serum IgM antibodies
Eastern equine encephalitis (EEE)	RNA virus, *Togaviridae* family (EEE virus)	*Aedes, Coquillettidia, Culex* mosquito bite	About 11 cases per year in the United States	Encephalitis, seizures	Detection of CSF or serum IgM antibodies
Dengue fever	RNA virus, *Flaviviridae* family (Dengue virus)	*Aedes* mosquito bite	100–400 million cases estimated per year, worldwide	Postinfection persistent fatigue, intracerebral hemorrhage, ischemic stroke, encephalitis, Guillain–Barré syndrome, cranial neuropathy. Rarely, psychosis and mania have been associated	Detection of serum antidengue-virus IgM antibodies, serum NS-1 antigen assay
Powassan virus disease	RNA virus, *Flaviviridae* family (Powassan virus)	Hard ticks, *Ixodes scapularis* and *Ixodes cookie* bite	Rising over time, 100 cases reported between 2003–2018 in the United States	Encephalitis	ELISA to exclude negative samples and followed by more specific, Plaque Reduction Neutralization Test, to confirm the exposure
West Nile virus (WNV) disease	RNA virus, *Flaviviridae* family (WNV)	*Culex* mosquito bite	1,000–2,000 cases per year in the United States	Meningitis, encephalitis	Detection of WNV-specific IgM antibodies in CSF or serum
Rabies	RNA virus, *Rhabdoviridae* family (Rabies virus)	Dog bite in endemic regions; in the United States, exposure to aerosols from racoons and bats	Around 60,000 deaths per year, worldwide	Encephalitis, acute anxiety and psychosis-like symptoms, hydrophobia, Guillain–Barré-like symptoms, cranial neuropathy, seizures, abnormal movements	Detecting viral RNA by hemi-nested RT-PCR from saliva, CSF, urine, or hair follicles
Herpes simplex virus encephalitis (HSVE)	DNA virus, *Herpesviridae* family (primarily herpes simplex virus-1)	Exposure is endemic; viral DNA becomes integrated into host cell and remains latent; encephalitis may occur with latent or reactivated infection	Estimated 2.2 cases per million population per year	Protean manifestations acutely, including mood lability, irritability, agitation, mania, psychosis, amnesia, apathy, and catatonia. Long-term sequelae include depression, anxiety, memory impairment, and other cognitive deficits	CSF PCR
Mononucleosis	DNA virus, *Herpesviridae* family (Epstein–Barr virus [EBV], or human herpesvirus 4 [HHV-4])	Human saliva, transfusion/transplant, possibly also breastfeeding and sexual transmission. Can reactivate	Highest incidence in late adolescence/early 20s. ~90–95% of adults worldwide have EBV antibodies	Fatigue, depressed mood (~28%), uncommonly meningitis/encephalitis	Heterophile antibody test (first-line) or EBV-specific antibody test

Cytomegalovirus (CMV) infection	DNA virus, *Herpesviridae* family (CMV or human herpesvirus 5 [HHV-5])	Close contact with children, transfusion/transplant, sexual transmission, perinatal transmission. Can reactivate	About 1.6 cases per 100 U.S. adults (ages 12–49) per year. Between 30–40,000 U.S. infants per year affected by congenital CMV.	Congenital CMV: Developmental delay, seizures, inattention, hyperactivity. CMV in immunocompromised adults: Encephalitis, meningitis	PCR (serum, CSF). In congenital CMV: Urine/saliva PCR, prenatal ultrasound
HHV-6 infection/Roseola infantum	DNA virus, *Herpesviridae* family (Human herpesvirus 6 [HHV-6])	Saliva, transplant (rarely), prenatal transmission. Can reactivate	Common in early childhood; over 70% of adults in developed countries have antibodies	In immunocompromised patients: Encephalitis, delirium, inattention	In adults: PCR (plasma, CSF). In children: clinical diagnosis
Measles (Rubeola)	RNA virus, *Paramyxoviridae* family (measles morbillivirus)	Airborne and droplet transmission, person-to-person contact, vertical transmission	49 cases in the United States in 2021, more in developing countries	In ADEM: confusion, seizures, psychosis. In SSPE: behavior change, seizures, dementia	Measles antibodies, CSF (lymphocytic pleocytosis/elevated protein, measles IgM)
Rubella (German measles)	RNA virus, *Matonaviridae* family (rubella virus)	Airborne and droplet transmission, vertical transmission	Less than 10 cases in the United States per year, more in developing countries	Autism, aggression, impulsivity, self-injury (later childhood manifestations of congenital Rubella)	Rubella IgM
Influenza viruses	RNA viruses, *Orthomyxoviridae* family	Droplet transmission, contact with contaminated surfaces	About 35 million cases in the United States in 2019–2020 season	Encephalitis, aseptic meningitis, ADEM (in children)	RT-PCR or rapid antigen tests on respiratory specimens
Zika	RNA virus, *Flaviviridae* family (Zika virus)	Mosquito bites, vertical transmission, sexual contact	57 cases in the United States territories in 2020	Adults: Meningoencephalitis Congenital: Microcephaly, developmental delay, seizures	RT-PCR (serum, urine)
BACTERIA					
Listeriosis	Aerobic Gram-positive rod (facultatively anaerobic) (*Listeria monocytogenes*)	Ingestion of contaminated soil or food (unwashed produce, soft cheese), vertical transmission	~0.29 per 100,000 people per year in the United States, higher among pregnant and older adult individuals	Meningoencephalitis. Sequelae of neonatal infection: Cognitive impairment, behavioral problems, depression, anxiety	Blood cultures and stool studies (in pregnant women), CSF culture (if concern for meningitis/encephalitis)
Syphilis	Spirochete bacteria (*Treponema pallidum*)	Sexual contact, vertical transmission	~7 million people with new syphilis infection in 2020 worldwide; variable surveillance for neurosyphilis worldwide leads to unclear incidence, however, estimated annual prevalence in United States between 0.8–2.5%	Early neurosyphilis: syphilitic meningitis/meningovascular meningitis, personality change, emotional lability, insomnia, and poor memory. Late neurosyphilis: tabes dorsalis, general paresis, syphilitic gummas, cognitive deficits, speech disorder, headache/dizziness, facial/tongue tremor, epileptic seizures, dementia	Serologic and CSF tests: VDRL, RPR
Lyme disease	Spirochete bacteria (*Borrelia burgdorferi sensu lato, B. afzeli/B. garinii* [Europe, Asia]; *B. burgdorferi sensu stricto* [United States])	Transmitted by *Ixodes* species ticks bite	~476,000 cases are diagnosed and treated per year in the United States and >200,000 cases per year in Western Europe	Early neurologic symptoms: facial nerve palsy, headache, lymphocytic meningitis, motor/sensory radiculoneuritis, encephalitis Neuropsychiatric symptoms: Depression, irritability, insomnia, cognitive problems, psychosis, mania, or dementia	Two-tiered serology/CSF protocol: 1. ELISA 2. Western Blot

Leptospirosis (Weil Disease)	Spirochete bacteria (*Leptospira*)	Direct or indirect contact with the urine, blood, and tissues of infected animals; Can occur via mucosal surfaces or damaged skin exposed to contaminated soil or water	0.5–1 million infections annually worldwide	Aseptic meningitis (most common); Also, encephalitis, myelitis, encephalomyelitis Intracranial bleeds, cerebellitis, movement disorders, flaccid paraplegias, mononeuritis, neuralgias, autonomic lability, polymyositis	Confirmed exposure+ Clinical findings+ Lab testing: MAT-gold standard ELISA PCR Dark-field microscopy Blood culture
Sydenham chorea	Anaerobic gram-positive bacteria (*Streptococcus pyogenes* [Group A beta-hemolytic streptococcus])	Upper respiratory tract (streptococcal pharyngitis) leading to autoantibody production	Up to 40% of patients with rheumatic fever (RF); RF incidence is 8 to 51 per 100,000 people worldwide	Choreoathetoid movements (St. Vitus dance), hypotonia, behavioral problems, and OCD symptoms	Clinical diagnosis
PANDAS	Anaerobic gram-positive bacteria (*Streptococcus pyogenes* [Group A beta-hemolytic streptococcus])	Autoimmune response to streptococcal infection	Prevalence unknown; estimated by one group to be 1:200 children in the United States	Obsessive-compulsive symptoms, tics (remitting- relapsing pattern). Cognitive decline and behavioral changes (irritability, aggression), behavioral regression.	Clinical diagnosis; five clinical criteria required for diagnosis (history of streptococcal infection required)
MULTIPLE ETIOLOGIC ORGANISMS					
PANS	Multiple infectious agents, including *Mycoplasma,* influenza, varicella, and *Borrelia*	Autoimmune response associated with multiple infectious agents	Prevalence unknown; estimated by one group to be 1:200 children in the United States	Obsessive-compulsive symptoms, restrictive eating, tics, cognitive decline, behavioral changes/regression.	Clinical diagnosis; history of infection does not have to be proven to diagnose PANS
PROTOZOA					
Malaria	Protozoa (*Plasmodium,* including *P. vivax, P. ovale, P. malariae, P. knowlesi, P. falciparum*)	Bite of the *Anopheles* mosquito	Estimated 240 million cases per year worldwide; increased risk for cerebral malaria in the immunocompromised	Cerebral malaria, marked by encephalopathy, seizures, alterations in consciousness, retinal hemorrhages; survivors may have persistent cognitive deficits, behavioral problems, seizures, deafness, blindness; postrecovery autoimmune encephalitis may occur	Blood smear microscopy
Human African Trypanosomiasis	Protozoa (*Trypansoma brucei*)	Bite of the tsetse fly; vertical mother-to-child transmission; sexual transmission	Declining; <1,000 reported cases in 2019	Prominent sleep disturbance with daytime somnolence, seizures, cognitive decline, personality change, sensory disturbance, psychosis	Trypanosome identification in blood or CSF on microscopy
Toxoplasmosis	Protozoa (*Toxoplasma gondii*)	Accidental ingestion of food and water contaminated with the protozoa oocysts, transplacental	Around one-third of the world's population is infected	Symptomatic in immunocompromised, encephalitis, seizures, abnormal movements, psychosis, cognitive deficits; congenital infection causes hydrocephalus, intellectual disability	Serum or CSF PCR, CT head or MRI brain for cerebral toxoplasmosis
Primary amoebic meningoencephalitis (PAM)	Amoeba (*Naegleria fowleri*)	Freshwater recreation, inhaled water	Rare (~2.6 cases per million exposures)	Altered mental status, behavioral changes	Trophozoites observed in CSF

Granulomatous amoebic encephalitis	Amoeba (*Acanthamoeba* in immunocompromised, *B. mandrillaris* in immunocompetent)	Soil or freshwater exposure	Rare	Visual disturbances, behavioral changes, focal neurology deficits	Trophozoites and cysts on brain biopsy
HELMINTHS					
Neurocysticercosis	Pork Tapeworm (*Taenia solium*) larvae	Fecal-oral route via close contact with a *T. solium* carrier excreting *T. solium* eggs in their feces	Serum antibodies in >10% of the developing world	Seizures, encephalitis, hydrocephalus, cognitive deficits	CT head/MRI brain; serology to confirm diagnostically challenging cases
FUNGI					
Mucormycosis	Fungus (*Mucorales*)	Opportunistic infection among immunocompromised, through airborne fungal spores	Wide variability across countries, from 0.02 per 100,000 in the Philippines to 14 per 100,000 in India	Stroke, intracranial hemorrhage, brain abscess	Demonstration of fungal hyphae in scrapings or biopsy of affected tissue

COVID-19 has increasingly been shown to have various neuropsychiatric complications, including new-onset depression, anxiety, psychosis, PTSD symptoms, chronic fatigue, and dysexecutive syndrome. A recent, large sample (N = 62,354) study among COVID-19-positive patients assessed the incidence and risk for psychiatric disorders during the first 14 to 90 days after COVID-19 diagnosis. The recovering patients had a significantly higher rate of psychiatric disorders, with a 5.8% probability of being newly diagnosed with a psychiatric illness within 90 days post-COVID-19 diagnosis. The most frequently acquired psychiatric diagnoses were anxiety disorders and trauma-related disorders, with a probability of occurrence of 4.7% within 90 days. Adjustment disorder, generalized anxiety disorder (GAD), PTSD, and panic disorder were the most frequent. Occurrence of a new mood disorder diagnosis (mostly depressive disorder) within 14 to 90 days after COVID-19 was 2%. The risk of new-onset psychiatric illness was higher among those with more severe COVID-19 disease, but it remained significantly elevated even after controlling for the disease severity.

A cross-sectional study of 402 adult COVID-19 survivors of varying time out from illness found the prevalence of PTSD symptoms, anxiety, and obsessive-compulsive symptoms to be 28%, 42%, and 20%, respectively, using self-rated questionnaires. A major limitation of most studies assessing neuropsychiatric disorders among COVID-19 survivors has been their reliance on standardized questionnaires (mostly self-report) to assess psychiatric disorders instead of using clinical interviews. There are a few reports of new-onset psychosis and mania emerging in the context of COVID-19 infection, but so far, the evidence for increased risk of psychosis or mania due to COVID-19 is very limited. Some studies have found increased risk of suicidality and suicidal behavior among patients with COVID-19, but the evidence is still inconclusive.

It is yet to be determined whether some of the aforementioned psychiatric syndromes are due to the neurobiologic effect of the virus or occur as a psychological reaction to the life-threatening complications of COVID-19. Most studies assessing cognitive deficits after COVID-19 infection have employed small sizes and have not undertaken comprehensive cognitive assessment. Some studies have not controlled for the recent severe medical complications due to COVID-19, including ARDS, which may be independently associated with cognitive deficits. Nonetheless, many of them have highlighted deficits in sustained attention.

As COVID-19 is being researched more thoroughly, both clinicians and researchers are taking note of persistent neuropsychiatric sequelae in the form of a loosely described syndrome termed "long-haul COVID," "long COVID" or more formally, post-acute sequelae of COVID (PASC). Some of the symptoms attributed to PASC include fatigue, dyspnea, brain fog, subjective cognitive symptoms like poor concentration and memory, fever, GI symptoms, headache, sleep impairment, anxiety, and depression. These symptoms may last for months, and vary in their severity, from mild to debilitating. A smartphone app-based prospective, observational multinational study of over 4,000 patients found that 13% had symptoms lasting ≥28 days, 5% for ≥8 weeks, and 2% for ≥12 weeks. Long COVID was characterized by fatigue, headache, shortness of breath, and anosmia. Its risk factors were increasing age, high BMI, and female sex. Having >5 symptoms during the first week of illness was also associated with long COVID.

A 1-year prospective cohort study was conducted at a single center in Wuhan, China among COVID-19 survivors (laboratory-confirmed diagnosis) who had required hospitalization and age, sex, and comorbidity-matched controls. Most survivors recovered functional status; 88% of previously employed participants returned to work. However, sequelae including fatigue, sleep difficulty, hair loss, anosmia, palpitations, arthralgia, low appetite, dysgeusia, dizziness, diarrhea, vomiting, chest pain, sore throat, rash, myalgia, headache, and lung diffusion impairment persisted for 12 months in some patients, especially in those who were critically ill. Anxiety or depression symptoms were more frequently seen at 12-month (26%) than 6-month (23%) follow-up. This increase in psychiatric symptoms may be multifactorial in etiology, with contributions from persistent dysimmune response as well as psychosocial complications of severe COVID-19 disease like social isolation, loss of employment, and poor physical health.

Diagnosis. The most commonly performed diagnostic test for COVID-19 infection is the detection and quantification of viral RNA using reverse transcriptase-polymerase chain reaction (RT-qPCR) assay. It is highly sensitive and specific, and detects early infection as well. Antibody testing using ELISA and antigen detection may be employed to get rapid results. CT chest may indicate the presence of lung involvement. More recently, CRISPR-based nucleic acid detection has been used by some laboratories. To diagnose the neuropsychiatric complications of COVID-19, clinical assessment is the gold-standard. Comprehensive assessment, including routine blood counts and chemistries, thyroid hormone assay, EEG, and CT head or MRI brain should be done to rule out the presence of any other underlying causes. A lumbar puncture for CSF studies may be warranted depending on the clinical presentation.

Management. Disease prevention methods using vaccines, social distancing measures, and the use of surgical masks and N95 respirators have been found to be very effective in reducing the spread of infection at both the individual level and in the community setting. Of the variety available, including mRNA vaccines, DNA vaccines, viral vector vaccines, and inactivated virus vaccines, mRNA vaccines have been shown to be the most effective.

Most patients either remain asymptomatic or develop mild disease which subsides spontaneously or with symptomatic treatment, including antipyretics, anti-inflammatory drugs, antitussives, and bronchodilators. Those with moderate to severe disease may require inhalational and/or systemic steroids, oxygen therapy, ICU care, and mechanical ventilation. The US FDA has approved remdesivir for the treatment of COVID-19 requiring hospitalization for those 12 years and above and weighing at least 40 kg. In addition, the IL-6 receptor antagonist tocilizumab and the JAK (Janus kinase) inhibitor baricitinib have been approved for certain hospitalized adults with COVID-19. An antiviral combination of nirmatrelvir and ritonavir, called Paxlovid, is US FDA-approved to treat mild-moderate COVID-19 in adults at high risk for progression to severe COVID-19, including hospitalization or death. High-dose steroids used to treat

COVID-19 can be lifesaving but may increase the risk of neuropsychiatric disorders including delirium, psychosis, mania, and depression as well as increasing susceptibility to infections.

There are no specific treatments available for the neuropsychiatric complications of COVID-19. Depressive and anxiety syndromes have been treated by a combination of antidepressants and psychotherapy. For psychotic or manic symptoms, one may use antipsychotics or a mood-stabilizing drug. So far, no treatment trials have been performed to address the cognitive symptoms associated with COVID-19. It is conceivable that they will respond to cognitive rehabilitation and regular physical activity, based on the efficacy of these approaches in other conditions (e.g., the benefit of cognitive rehabilitation in HIV/AIDS). It will be extremely important to conduct systematic studies on the neuropsychiatric complications, particularly suicidality, associated with COVID-19 to understand its true psychological burden on the survivors. In a similar vein, it is critical to investigate PASC, considering that it may become the next "pandemic" once the COVID-19 pandemic subsides.

There is burgeoning literature on the higher risk of severe COVID-19 disease and resultant mortality among patients with Alzheimer disease (AD), and other dementias. On the other hand, it is unknown whether having COVID-19 increases the risk of neurodegeneration and occurrence of dementia. Based on the neurobiologic changes seen in COVID-19 (for example, elevated IL-6 levels) that are known to be associated with hippocampal atrophy, microglial activation, and vascular damage, and a few cases of new-onset parkinsonism occurring in the context of COVID-19 infection, several experts have speculated that COVID-19 infection may trigger neurodegenerative diseases such as Alzheimer and Parkinson in some vulnerable patients.

Mucormycosis: An Opportunistic Fungal Infection in the Setting of COVID-19. Mucormycosis is an opportunistic and potentially fatal fungal infection caused by fungi of the order Mucorales (see Table 2.9–1). These ubiquitous pathogens are saprobes, found in soil and decaying matter, and spread through spores that become airborne. Blood vessel invasion causes thrombosis due to endothelial injury, leading to infarction and necrosis of different tissues. Infection mainly affects immunocompromised patients, especially those with diabetes mellitus, prolonged corticosteroid use, organ transplant recipients, neutropenia, and hematologic cancers.

Severe manifestations like rhino-orbital mucormycosis have a very poor prognosis, with a mortality rate of 50% despite appropriate treatment.

Affected individuals often develop fever, nasal congestion, purulent nasal discharge, and headache initially. If not treated early, contiguous spread to adjacent structures like the orbit may occur. Intracranial spread can occur through the cribriform plate, supraorbital fissure, and perineural routes. Cavernous sinus or sagittal sinus thrombosis, carotid occlusion, cerebral infarction, intracranial aneurysm, intracranial hemorrhage, and cerebral abscesses are potential complications.

Prevalence varies across countries, from 0.02 per 100,000 in the Philippines to 14 per 100,000 in India. Various countries have reported mucormycosis associated with COVID-19, including the United States, but in India it has reached epidemic proportions, affecting 1.8% of patients with COVID-19 admitted to tertiary care centers. SARS-CoV-2 impairs cell-mediated immunity via decreases in CD4$^+$ and CD8+ cell counts, thereby increasing the risk of fungal infections. In one study of mucormycosis after COVID-19, the mortality rate was over 23%. Comorbid medical conditions including diabetes mellitus, diabetic ketoacidosis, and ARDS, and the use of corticosteroids and broad-spectrum antibiotics are additional risk factors. Of those affected, almost 90% had developed moderate-to-severe pneumonia, and over 40% had required invasive ventilation in one study; all were given steroids and broad-spectrum antibiotics. On average, 12 days elapse between COVID-19 diagnosis and the onset of symptoms (most common symptom being nasal congestion with or without discharge) suggestive of mucormycosis. Men seem to have a higher risk than women. The average age of affected patients is 55 years.

The infection is diagnosed by demonstrating the presence of fungal hyphae (nonseptate, broad-based hyphal structures with 90-degree branching), by direct examination in scrapings and biopsy of the affected tissue. Imaging modalities like CT and/or MRI of the orbit, brain, and/or paranasal sinuses can be used in suspected cases of rhino-orbital disease. It is treated using systemic antifungals such as amphotericin B, posaconazole, and isavuconazole and surgical debridement of the necrotic tissue.

Alphaviridae

The *Alphaviridae* genus of the *Togaviridae* family includes several viruses known to be pathologic and neurotropic in humans, including Chikungunya virus, Eastern equine encephalitis (EEE) virus, Western equine encephalitis (WEE) virus, and Venezuelan equine encephalitis virus. Several of these are addressed below.

Chikungunya. Chikungunya is an enveloped, positive-sense, single-stranded RNA arbovirus, spread by the *Aedes* mosquito (see Table 2.9–1). Since being first described in 1952 in Tanzania, this virus has spread across the world and into other mosquito vectors. It is predominant in Africa, Asia, including the Indian subcontinent, and in recent years, Europe and the Americas (especially Caribbean countries, Brazil, and Colombia) have also been affected.

The virus appears to enter the CNS via the choroid plexus, but the extent to which it affects neurons and astrocytes is unknown.

Most (72% to 95%) patients develop symptoms, mainly fever and arthralgias, along with headache, fatigue, myalgias, joint swelling, rash, and GI symptoms. Some patients develop chronic disease leading to arthralgias lasting for several weeks to months, mimicking rheumatoid arthritis. Neurologic complications occur in the acute phase in approximately 0.1% of the cases, mostly encephalitis and encephalopathy. Apart from that, it may also lead to myelopathy, acute flaccid paralysis, meningoencephalitis, Guillain–Barré syndrome, uveitis, retinitis, and optic neuritis.

Vertical transmission from a mother to child does occur at a rate of about 50% for both vaginal and cesarean deliveries; antepartum fetal infections do not occur as the virus does not cross the placenta. Symptomatic newborns develop severe typical symptoms of the disease, including fever and arthralgia, rashes, and can have shock and hemorrhage, as well as occasionally meningoencephalitis and long-term neurodevelopmental delays. One study of apparently asymptomatic infected children of mothers who had Chikungunya infection during pregnancy demonstrated significantly worse neurocognitive functioning compared to uninfected children, and survivors of symptomatic neonatal infections have high rates of neurodevelopmental delay (51% in one study). Diagnosis is typically made based on clinical suspicion in endemic regions. RT-PCR is highly specific but sensitive mainly in the first week of infection. After that, paired ELISA may be used to detect a four-fold rise in anti-CHIKV IgM antibodies to make a diagnosis. Treatment is supportive. Survivors of Chikungunya encephalitis have 17% to 43% rates of persistent neurologic sequelae (cognitive deficits, epilepsy, language impairment).

Western Equine Encephalitis. WEE is caused by WEE virus, a positive-sense, single-stranded RNA virus, which is transmitted by *Culex* mosquitoes (see Table 2.9–1). Up till the mid-20th century, it led to huge outbreaks of encephalitis in the United States. Fortunately, and mysteriously, its incidence has reduced drastically over the years, ascribed to reduction in the rural population. Its annual incidence in the last two decades has been one to two cases. Humans and horses are its dead-end hosts. Its incubation period is 1 to 2 weeks. It mostly causes a subclinical infection. In its severe form, WEE manifests with fever, vomiting, headache, neck stiffness, and altered sensorium; some patients may develop tremors. Some develop coma, that can last for up to 4 days, and coma duration predicts the severity of neurologic sequelae. Acute illness lasts up to 2 weeks. Infants are hit the hardest as the infection leaves them with cognitive, behavioral, motor deficits, and seizures. Overall, disease mortality is below 5%. CSF shows the typical profile of viral infection with lymphocytic pleocytosis after the first two days, normal or elevated protein, and normal glucose. EEG findings are usually nonspecific. Diagnosis is confirmed using serology. Management of acute cases is supportive since there are no effective antiviral agents known. A vaccine is available for persons at occupational risk of acquiring the infection (i.e., laboratory workers studying the virus).

Eastern Equine Encephalitis. About 11 cases of EEE are reported annually in the United States, with cases mostly in Atlantic and Gulf Coast regions. It is caused by EEE virus and is transmitted by *Aedes, Coquillettidia, and Culex* mosquitoes (see Table 2.9–1). Infection has a 30% mortality rate, making it the deadliest arboviral infection in the United States. EEE virus causes more severe disease in older adults and children, and asymptomatic cases are uncommon. It mainly affects basal ganglia, thalami, substantia nigra, cerebral cortex, and the hippocampus, and may also involve adjacent white matter. Incubation period is 3 to 10 days, with a prodromal phase characterized by fever, headache, and abdominal distress. Some patients experience rapid deterioration with high-grade fever, delirium, vomiting, focal neurologic signs, and seizures.

A vast majority (~70%) of the patients develop coma within two days of hospitalization; as above, approximately one-third of patients die, and another one-third become severely disabled. On exam, patients manifest fever, leukocytosis with neutrophilia, and hyponatremia. CSF also shows neutrophilia. MRI shows basal ganglia and thalamic involvement in most patients. Some patients have the “parenthesis sign,” linear hyperintensities on T2/FLAIR in the external and internal capsules with sparing of the lentiform nuclei. Detecting IgM antibodies in the serum or CSF is used to make a confirmatory diagnosis. CSF PCR is another method employed. Treatment is supportive, as no specific treatments are available against this infection.

Dengue

Dengue fever is caused by the dengue virus, a single-stranded, positive-sense RNA virus belonging to the *Flaviviridae* family, transmitted by *Aedes* mosquitoes (see Table 2.9–1). It has four serotypes. It is endemic in >100 countries in the tropical/subtropical climate range. In the United States, Florida, Texas, and Hawaii have witnessed recent outbreaks, and it regularly occurs in the U.S. territories of Puerto Rico, Virgin Islands, and Samoa. Its global incidence has increased by 30 times in the last five decades. An estimated 100 to 400 million cases of dengue infection occur annually worldwide. The incubation period is 4 to 8 days. Most infections are asymptomatic or subclinical, and symptomatic disease often manifests with self-limiting flu-like symptoms. Dengue is also associated with the cytokine storm phenomenon. Only a small proportion of the patients develop a severe disease, usually after being infected for the second time. Severe disease manifests with plasma leakage, and hypovolemic shock may set in, leading to multiorgan dysfunction, disseminated intravascular coagulopathy, and severe bleeding. It may also be complicated by hepatitis, neurologic involvement, and myocarditis. Untreated severe disease can lead to mortality in up to 20% of the cases. Various neurologic presentations associated with dengue include intracranial bleeding, ischemic stroke, encephalitis, Guillain–Barré syndrome, and isolated cranial nerve damage. Case reports and case series have described neuropsychiatric complications like mania and acute psychosis following dengue infection. More than 50% of patients (two-third women and one-third men) may experience persistent symptoms (arthralgia, fatigue, etc.) for up to 2 years after the resolution of acute infection.

Infection with one serotype provides lifelong immunity against that strain but renders partial and temporary crossimmunity to other serotypes. Subsequent infections by other serotypes increase the risk of severe disease.

Presence of serum antidengue-virus IgM antibodies indicates active infection. Alternatively, serum NS-1 antigen assay may be used as a rapid diagnostic test. RT-PCR can detect the virus within the first week of illness.

There are no specific treatments or prophylactic agents available. Symptomatic patients are managed conservatively. At present, the key strategy remains mosquito control. A vaccine exists for children with previous infection living in endemic areas.

Powassan Virus Disease

Powassan virus disease is caused by the Powassan virus, a positive-sense, single-stranded flavivirus, mainly transmitted by hard ticks, *Ixodes scapularis* and *Ixodes cookie* (see Table 2.9–1). It is mostly reported from North America and Eastern Russia. It has seen a dramatic rise over the last two decades in the United States (with over 100 cases reported between 2003 and 2018), attributed to proliferation of *I. scapularis* ticks and its reproductive host, the white-tailed deer, in the northeastern and midwestern United States. It has been increasingly affecting middle-aged and older adults. The incubation period ranges from 1 to 5 weeks. Severe cases develop encephalitis, which may be complicated by seizures, paralysis, and coma. Significant neurologic disability occurs in around half of the severe cases, with mortality risk of around 10%. It's unclear what proportion of the infections comprises severe disease, since nonsevere cases may go undetected. CSF may show pleocytosis and elevated protein. MRI findings are nonspecific, showing T2/FLAIR hyperintensities in the brainstem, basal ganglia, and thalami. Diagnosis involves ELISA to exclude negative samples and then using a more specific assay, Plaque Reduction Neutralization Test, to confirm the exposure. Treatment of infected cases is mainly supportive.

West Nile Encephalitis

West Nile Encephalitis is caused by West Nile Virus (WNV) which is an enveloped positive-stranded RNA virus of the family *Flaviviridae* (see Table 2.9–1). It was first isolated from a febrile patient in the West Nile district of Uganda in 1937. The virus has a wide geographical distribution, including southern Europe, Africa, central and south Asia, and Oceania. WNV has become endemic in all 48 contiguous states in the United States, and Canada since its discovery in North America in New York City in 1999. The annual number of reported WNV disease cases can fluctuate widely due to periodic epidemics, with most cases reported in the summer-fall season (July through September).

Although the virus infects a wide variety of animals, including horses, cattle, and humans, the major vertebrate hosts are wild birds, and migratory birds are important in the spread of the disease owing to long-term high-titer viremia. The principal vectors are mosquitoes of the *Culex* genus.

Most (~80%) infections are asymptomatic. Clinical manifestations include West Nile fever in 25% of those infected with symptoms consisting of headache, generalized weakness, morbilliform or maculopapular rash (often at time of defervescence), fever (often low grade), and myalgia; and neuroinvasive disease in less than 1% of those infected. Neuroinvasive disease can manifest with three different clinical presentations: (1) meningitis characterized by clinical signs of meningeal inflammation, including nuchal rigidity, Kernig or Brudzinski sign, or photophobia or phonophobia; (2) encephalitis characterized by depressed or altered level of consciousness, lethargy or personality change lasting more than 24 hours; and (3) acute flaccid paralysis characterized by abrupt onset limb weakness with marked progression over 48 hours, typically asymmetric, areflexic or hyporeflexic, and without sensory abnormalities. All neurologic presentations may be accompanied by other symptoms and signs: nausea and vomiting, myalgia, chills, rash (less common than in West Nile fever), arthralgia, ataxia, visual disturbance, tremors, myoclonus, and bulbar dysfunction (dysarthria, dysphagia).

Diagnosis is most readily made by detection of WNV–specific IgM antibody in cerebrospinal fluid (CSF) or serum. CSF generally shows normal glucose, elevated protein, and pleocytosis (>5 leukocytes/μL). Treatment is supportive for both WNV fever and neuroinvasive disease. Illness duration is days to weeks with complete recovery in case of WNV fever (prolonged fatigue may continue for weeks to months) and weeks to months in neuroinvasive disease. Long-term functional and cognitive difficulties are common in patients with encephalitis and paralysis.

Rabies

Rabies virus, a neurotropic RNA virus, belongs to the *Lyssavirus* genus in the family *Rhabdoviridae* (see Table 2.9–1). Stray dogs are the major cause of rabies globally, while raccoons and bats are common reservoirs in the United States. Around 60,000 deaths occur annually due to rabies worldwide, mostly due to lack of postexposure prophylaxis. The risk of infection from a bite is much higher than from a scratch. Aerosol spread is also possible through exposure to infected bats. The incubation period lasts anywhere from 1 to 2 months up to 8 years. The most common virus variant–that found in dogs–multiplies inside myocytes after binding to the alpha subunit of the nicotinic receptor, hence deep bites into muscle are high risk. CNS entry occurs through retrograde axonal transport. Soon, it spreads through motor and sensory axons leading to extraneural organ involvement, including the heart and autonomic plexuses, skin, and salivary glands. When symptoms develop, the virus has already disseminated throughout the CNS.

There are two main clinical presentations—classical and nonclassical. The classical furious and paralytic (resembling Guillain–Barré syndrome) forms occur mostly due to the dog rabies variant. Nonclassical type occurs in patients exposed to bats. In furious rabies, the initial presentation resembles intense anxiety or acute psychosis, worsened by thirst, fear, bright light, or loud noise. Fever is common, starting after the prodromal phase. Soon, the patients develop fluctuating consciousness (alternating periods of severe agitation and normalcy or depression), phobic and inspiratory spasms, and autonomic dysfunction. Aggressive behavior is common. Coma ensues as the disease progresses. Paralytic rabies does not involve aggression or impaired consciousness. Most people die within a week due to furious rabies and within 2 weeks due to paralytic rabies. In the classical type, encephalopathy, hydrophobia, and aerophobia are more common.

Nonclassical type leads to choreiform movements of the bitten extremity during the prodromal phase. The patients may have focal brainstem signs and myoclonus, convulsive and nonconvulsive seizures, and hallucinations. Cranial, motor and sensory deficits, tremor, myoclonus, and exposure-site symptoms (burning, itching, piloerection) are more common in the nonclassical type.

CSF is typically normal. Most cases, both furious, and paralytic, have MRI abnormalities in the brainstem and hippocampus. Late-stage blood–

brain barrier breakdown shows up as midline gadolinium contrast-enhanced lesions. The diagnosis is confirmed by detecting viral RNA by heminested RT-PCR from saliva, CSF, urine, or hair follicles (viremia does not occur).

There is no effective treatment once the clinical signs develop. The best way to manage rabies is by using pre-exposure vaccination for people at high risk of exposure or postexposure prophylaxis which includes aggressive wound cleaning followed by administering an approved rabies vaccine. Transdermal wound also needs local infiltration with rabies immunoglobulin.

Human Herpesviruses

There are over 100 known herpesviruses, of which eight (the human herpesviruses or HHV) are known to cause infection in humans. These include herpes simplex virus (HSV) 1 and 2, varicella-zoster virus (VZV or HHV-3), Epstein–Barr virus (EBV or HHV-4), cytomegalovirus (CMV or HHV-5), HHV-6 and -7 (both of which can cause roseola), and HHV-8, which is the cause of Kaposi sarcoma. The human herpesviruses are made up of a linear, double-stranded DNA genome enclosed in an icosahedral capsid beneath a lipid envelope. The viral DNA becomes integrated into the host cell genome. Following primary infection, human herpesviruses establish latency and can later be reactivated, especially in the setting of immunologic compromise. Almost all human herpesviruses can result in CNS infection in adults and children, including HSV-1 (most commonly), and CMV (in the setting of AIDS).

Herpes Virus Encephalitis

EPIDEMIOLOGY. HSV type 1 (HSV-1) and HSV type 2 (HSV-2) can both cause severe neuropsychiatric disease, with HSV-2 representing a common cause of viral meningitis. HSV-1 however can spontaneously develop into herpes simplex virus encephalitis (HSVE) in the immunocompetent patient, making it an important viral cause of neuropsychiatric symptoms (see Table 2.9–1). Viral encephalitis overall affects about 20,000 people per year in the United States, and HSV-1 is the most commonly identified cause of sporadic acute viral encephalitis, representing 8% of encephalitis and meningitis cases in a large-scale study of over 26,000 patients, but up to 25% to 50% of cases in other studies. HSVE affects all age groups, with mortality rates up to 70% in the untreated and 25% to 50% for those receiving treatment.

Neuropsychiatric symptoms are common in HSVE, varying from depression and anxiety to irritability, agitation, mood lability, mania, and psychosis, to amnesia, apathy, and more. Overall behavioral changes appear to be the presenting symptom of HSVE in about a quarter of cases. Long-term, there are significant neuropsychiatric sequelae of HSVE, including depression (18%), anxiety (45%), panic attacks (27%), irritability (64%), and apathy (1%) when measured about 6 months after acute illness in one study. Other work has identified long-term sequelae that include memory impairment in 69%, behavioral problems in 45%, OCD-like symptoms in 7%, and aggression and hyperphagia each in 3%. Kluver–Bucy syndrome can be a consequence but is less common when acute HSVE is treated with acyclovir. Persistent functional impairment is usually secondary to emotional or behavioral sequelae rather than memory problems.

PATHOPHYSIOLOGY. HSVE appears to plausibly arise from both primary (30%) and latent or reactivated (70%) infection. Entry into the parenchyma may occur via the olfactory nerve or trigeminal nerve; entry via the olfactory nerve may explain the virus's predilection for the adjacent temporal lobes, or alternatively, this may reflect a viral affinity for this region. In addition to the temporal lobes (including hippocampus) and insula, HSVE may also affect the cingulate gyrus and orbitofrontal cortex. Damage to frontal and amygdalar regions may be responsible for persistent behavioral problems in survivors, whereas hippocampal and amygdalar damage likely underwrite persistent memory problems. The virus may impair major histocompatibility complex antigen expression on infected neurons and inhibit autophagy in order to evade host immune response. This leads to tissue edema evolving into hemorrhagic necrosis. Individuals with Toll-like receptor mutations may have increased vulnerability to developing HSVE.

CLINICAL PRESENTATION. A prodrome of fever and headache may precede the abrupt onset of changes in personality and behavior, alterations in consciousness, seizures, and motor abnormalities. Neuropsychiatric symptoms may fit broadly into four categories: psychosis, catatonia, depression with psychotic features, and mania, but with considerable overlap between categories rather than strictly fitting into conventional diagnostic specifications. With regards to psychotic symptoms, delusional thinking can occur, as well as auditory, visual, olfactory, and gustatory hallucinations. Irritability, aggression, and violence, restlessness or hyperactivity, anxiety, disinhibition, perseveration, and confusion can also occur. Memory disorders are prominent, with anterograde memory markedly more affected than retrograde. Disturbances in executive functioning, visuospatial abilities, and language may also develop. Some research on post-HSVE neuropsychological function suggests universal impairment in memory encoding, naming, and semantic memory, with the majority (~75%) affected by executive function, working memory, and autobiographical memory dysfunction, and about half showing visuospatial dysfunction.

DIAGNOSIS. A clinical history of fever together with focal neurologic signs, seizures, confusion, or decreased consciousness, should raise clinical suspicion for HSVE. Lumbar puncture will show elevated opening pressure, xanthrochromic CSF, elevated protein, normal glucose, and a lymphocyte-predominant elevated leukocyte count. PCR more often shows HSV-1, and is positive 24 hours after symptoms begin and until 7 days after antivirals are started. It is also important to rule out other viruses, bacteria, fungi, and autoimmune causes of encephalitis. EEG findings are variable, ranging from nonspecific slowing, asymmetric spikes, slow waves, to intermittent periodic lateralizing epileptiform discharges (PLEDs). Early on, brain imaging may not show abnormalities, but later temporal lobe signal abnormalities, edema, and sometimes hemorrhage can be observed on MRI. However, treatment should not wait for confirmation of the diagnosis given that delaying can adversely impact morbidity and mortality.

After resolution of the acute illness, it may be helpful to obtain neuropsychological testing to characterize deficits and guide rehabilitation efforts, as well as to monitor for seizures as postencephalitic epilepsy can develop.

MANAGEMENT. Suspected or confirmed HSVE should be treated with intravenous acyclovir for up to 21 days. There is very limited data to guide treatment of neuropsychiatric symptoms. Case reports have suggested that atypical antipsychotics and mood stabilizers may help with specific symptom domains, although these patients may be at increased risk for neuroleptic malignant syndrome so antipsychotics should be dosed cautiously. Ropinirole may be helpful for postencephalitic apathy and depression. Bupropion should be avoided given its effect on the seizure threshold and high rates of epilepsy in post-HSVE patients. Antiseizure medications can sometimes address both seizures and behavioral problems. There may be a role for empiric trials of other classes as well, such as antidepressants, benzodiazepines, cholinesterase inhibitors, and stimulants. ECT can help with catatonia.

If cognition is not too severely impaired after HSVE then psychotherapy may provide some benefit. For those with more severe residual memory deficits, undesirable behaviors may respond to behavioral tokens and reinforcers. Management of sources of anxiety and agitation (such as disorientation to time) may also benefit the severely impaired patient. For encoding new material, these individuals may respond to repeated learning trials, cuing, and visually presented material.

Overall, early acyclovir treatment and rehabilitation in multiple domains may help to prevent and treat neuropsychiatric sequelae of HSVE.

Epstein–Barr Virus

Epstein–Barr virus (EBV, or HHV-4) is typically transmitted through saliva, though EBV can also be acquired through blood transfusion, organ transplantation, and possibly also sexual contact. EBV is tropic for B- and T lymphocytes, NK cells, myocytes, and epithelial cells (see Table 2.9–1). Approximately 90% to 95% of adults worldwide have antibodies to EBV, with rates being higher in developing countries than in the United States.

Acute EBV infection is typically subclinical but can also manifest as infectious mononucleosis, which presents with 1 to 3 weeks of fever, tonsillitis/pharyngitis, cervical lymphadenopathy, and fatigue (which can last months). Additional complications can include mild hepatitis (~90%), splenomegaly (up to 50%), and depression (28%). Following acute infection, EBV becomes a chronic, latent infection. A so-called transforming virus, EBV is associated with a higher prevalence of malignancies, including B- and T-cell lymphomas, Hodgkin lymphoma, and nasopharyngeal and gastric carcinomas. This is particularly true in immunocompromised individuals (e.g., patients who have HIV or are immunosuppressed post-transplant).

In 0.5% to 7.5% of cases, acute EBV can have neurologic manifestations, most commonly aseptic or viral meningitis. Other neurologic complications can include encephalitis, myelitis, cranial and peripheral neuropathies, Guillain–Barré syndrome, and primary CNS lymphoma in patients with AIDS. There is some evidence to suggest a relationship between EBV exposure and the development of autoimmune diseases, including multiple sclerosis, possibly through interactions between viral proteins and human genes related to inflammation. Based on limited data, prior EBV infection has also been posited as a risk factor for psychosis, depression, fibromyalgia, and chronic fatigue syndrome.

Serum positivity for heterophile antibodies or EBV-specific antibodies confirms the diagnosis of infectious mononucleosis. Treatment is typically supportive. In acute CNS involvement, CSF can show lymphocytic pleocytosis and EBV positivity on PCR, and patients are typically treated with corticosteroids.

Cytomegalovirus

Like other human herpesviruses, CMV can establish latency and reactivate later in life, particularly in the setting of immunologic compromise (see Table 2.9–1). Worldwide, the prevalence of CMV IgG antibodies is between 40% (in industrialized countries) and 100% (in some developing countries). Approximately 0.6% of infants worldwide are born with congenital CMV. Symptomatic congenital CMV typically results from primary CMV infection, usually in the first trimester. CMV infection can occur through contact with young children, sexual contact, contact with blood, and vertical transmission.

In immunocompetent hosts, primary CMV infection can be asymptomatic or can resemble infectious mononucleosis. Among individuals with immunologic compromise (due to post-transplant immunosuppression, AIDS, etc.), CMV can result in tissue-invasive disease, including enteritis, colitis, and retinitis. Usually in the setting of immunologic compromise, CMV can have CNS manifestations, including encephalitis, meningitis, and Guillain–Barré syndrome. Congenital CMV can present with petechiae, hepatosplenomegaly, microcephaly, chorioretinitis, and sensorineural hearing loss. Symptoms of congenital CMV can also present later in development, and can include developmental delay, abnormal tone, seizures, and inattention/hyperactivity.

A growing body of evidence suggests links between CMV and neuropsychiatric symptoms. A leading hypothesis is that CMV adversely affects CNS function through chronic neuroinflammation, potentially mediated by interferon-gamma. CMV IgG positivity has been associated with higher rates of psychiatric illness (including somatoform symptoms and bipolar disorder with later-life onset) and suicide attempts. Associations have also been found between CMV IgG levels and the presence of Alzheimer neuropathology in postmortem brain tissue.

Primary CMV is diagnosed via serum PCR. Acute CNS CMV is diagnosed through CSF PCR. In congenital CMV, prenatal ultrasound may show periventricular calcifications, ventriculomegaly, and evidence of abnormal brain migration, for example, pachygyria. Treatment in acute infection involves use of an antiviral (e.g., ganciclovir, valganciclovir, or foscarnet).

HHV-6

Like most herpesviruses, HHV-6 can establish latency before reactivating in later life (see Table 2.9–1). Acute infection with HHV-6 can cause roseola infantum in children but typically does not cause symptoms in immunocompetent adults. Immunocompromised patients—particularly, hematopoietic cell transplantation (HCT) recipients—are vulnerable to HHV-6 reactivation, which may precipitate hepatitis, pneumonitis, and encephalitis. HHV-6 has known neurotropism, and HCT patients with HHV-6 reactivation are acutely at higher risk of delirium and impaired attention. HHV-6 has also been proposed as a causative factor in medial temporal lobe epilepsy (given increased prevalence of HHV-6 in brain tissue from affected patients). Connections between HHV-6 and ME-CFS and psychosis have been proposed but have not been borne out.

Measles (Rubeola)

A single-stranded, negative-sense RNA virus, measles (rubeola) is an extremely contagious disease that spreads through airborne transmission or interpersonal contact (see Table 2.9–1). After infecting the host through the respiratory mucosa or conjunctivae, measles can travel through the lymphatic system and bloodstream to various organs, including the lungs and brain. Although the introduction of the measles vaccine in the 1960s dramatically decreased the number of cases worldwide, outbreaks continue to occur in areas with lower rates of immunization (including, recently, Nigeria and Pakistan). Measles was declared eliminated in the United States in 2000, but regions of the United States with lower vaccination rates continue to be vulnerable; the year 2019 saw the highest number of U.S. measles cases since 1992.

Primary measles infection typically involves a prodrome of fever, anorexia, and malaise, followed by cough. Other symptoms can include rhinitis, conjunctivitis, Koplik spots, and a maculopapular exanthem. Complications such as pneumonia, keratitis, and (rarely) encephalitis can also occur. In resource-limited settings, acute measles can be fatal in up to 10% of cases. Infection during pregnancy can result in fetal or maternal death, spontaneous abortion, premature birth, and low–birth-weight.

In approximately 1 in 1,000 cases, measles can cause acute disseminated encephalomyelitis (ADEM), a demyelinating syndrome thought to be an autoimmune response to an infectious trigger. ADEM typically presents within 2 weeks of the measles rash and confers a 10% to 20% mortality rate. Symptoms can include fever, headache, meningismus, confusion, coma, and seizures. Rarely, psychosis can be a presenting symptom.

Up to 7 to 10 years after measles infection, patients can develop subacute sclerosing panencephalitis (SSPE), a progressive, fatal complication thought to result from persistent CNS infection. SSPE can present with behavioral changes (lasting weeks to years), seizures, and lethargy, followed by myoclonus, motor and sensory symptoms, and dementia.

Vaccination with the MMR (measles, mumps, and rubella) vaccine is a valuable preventative measure. While there has been some public concern that the MMR vaccine can lead to autism, evidence does not support this association. Diagnostic testing involves serum assay for IgM antibodies, or isolation of virus in nasal or oropharyngeal culture. In the case of neurologic symptoms, CSF studies can show a lymphocytic pleocytosis, elevated protein, and positive IgM antibodies. Vitamin A supplementation is recommended in children with acute measles, but treatment is otherwise supportive.

Rubella

Also known as “German measles,” rubella is caused by a togavirus that spreads through both airborne and droplet transmission (see Table 2.9–1). Following a 2-week incubation period, patients can develop maculopapular rash, mild fever, and lymphadenopathy. Potential complications include arthritis and, rarely, postinfectious encephalitis. However, many people with acute rubella are asymptomatic or have a subclinical presentation.

While not a particularly threatening illness in children and adults, rubella poses great risks to fetuses infected through vertical transmission, especially during the first trimester. Acute rubella in pregnancy can result in spontaneous abortion, intrauterine growth restriction, and preterm delivery. Symptoms of the congenital rubella syndrome include cataracts, retinopathy, hearing loss, hydrocephalus, cardiac defects, petechiae, endocrine abnormalities, meningoencephalitis, and seizures. Later manifestations can include intellectual disability (in about one-third of patients), autism, aggression, impulsivity, self-injury, and motor delay, even in infants who are asymptomatic at birth. Rarely, patients with the congenital rubella syndrome can develop a progressive panencephalitis resembling SSPE. Use of the live attenuated rubella virus vaccine (part of the MMR) is key in preventing congenital rubella.

Other Prenatal Infections (Influenza, Zika, and Listeria)

Maternal infection during pregnancy with various pathogens (including several discussed elsewhere in this chapter) is known or suspected to cause neuropsychiatric symptoms in some affected children. Influenza virus, Zika virus, and *Listeria* bacteria fall into this category and will be discussed briefly here.

Connections between prenatal maternal infection and schizophrenia were first hypothesized in the 1980s by researchers drawing on data from flu epidemics. Although most influenza viruses are not neurotropic, it is thought that influenza (an orthomyxovirus) results in a maternal inflammatory cascade that can adversely impact the developing fetal brain. Birth in the winter or spring months is a risk factor for schizophrenia, and influenza has been proposed as a potential mediator of this association. Limited data suggest that maternal influenza confers an increased risk of bipolar disorder with (but not without) psychotic features. While initial data also suggested a link between maternal influenza and schizophrenia, a more recent study did not replicate those findings, and the relationship between influenza and psychosis remains unclear.

Although adults with acute influenza typically present with symptoms such as fever, malaise, and dry cough, influenza can also cause CNS manifestations, such as encephalitis, aseptic meningitis, and Guillain–Barré syndrome. Children with influenza can develop febrile seizures or ADEM (see Table 2.9–1).

The 2015 to 2016 outbreak of the Zika virus raised the global profile of this flavivirus (a relative of dengue and West Nile). Zika is typically transmitted by mosquitoes but can also be spread vertically and through sexual contact. About 20% to 25% of infected adults experience clinical symptoms, which are typically mild and include fever, pruritic rash, and arthralgias. Zika is neurotropic and can also have CNS manifestations, such as meningoencephalitis and Guillain–Barré syndrome (see Table 2.9–1).

The congenital Zika syndrome can affect up to 40% of children whose mothers contracted Zika during pregnancy. The risk is greatest with 1st- and 2nd-trimester infection. Congenital Zika syndrome is thought to be the result of Zika's known toxicity to neural progenitor cells. Affected infants can present with microcephaly, intracranial calcifications, hypertonia, hearing loss, ocular abnormalities, developmental delay, and seizures. The effects of in-utero Zika exposure are still being characterized, but early data suggests that exposed children who were asymptomatic at birth may nonetheless exhibit deficits in motor and cognitive functioning starting in early childhood.

A Gram-positive rod, *Listeria monocytogenes* is transmitted through ingestion of contaminated food (e.g., unwashed produce, soft cheeses). Adults exposed to *Listeria* can develop febrile gastroenteritis, and bacteremia may lead to complications (e.g., meningoencephalitis, cerebritis). Listeriosis has a 10-fold higher incidence in pregnancy. Maternal infection can result in fetal or neonatal infection (via bacteremia) with adverse outcomes including fetal death, preterm labor, neonatal sepsis, and meningitis. Prognosis is most favorable for infection occurring in the 3rd trimester. Long-term sequelae of early-life meningitis (due to *Listeria* or other causes) can include cognitive and motor impairment, hydrocephalus, hearing loss, and higher incidence of behavioral problems, depression, and anxiety.

BACTERIA

Spirochetes

The family Spirochaetaceae includes *Treponema*, *Borrelia,* and *Leptospira*. Spirochetes have a wavelike body and flagella enclosed between the outer and inner membranes.

Syphilis

EPIDEMIOLOGY. *Treponema pallidum*, the causative agent of syphilis, is a bacterial spirochete that was first identified in 1905. Infection occurs through sexual contact with infected individuals, though may also be transmitted vertically from the infected mother to the fetus in utero or in neonates during passage through the infected birth canal during delivery (see Table 2.9–1). On rare occasions, syphilis may be acquired by transfusion of infected human blood or via epithelial lesions.

During the second half of the 20th century syphilis rates declined due to use of penicillin. However, since reaching a historic low in 2000 to 2001, rates have been steadily climbing almost every year, with unprotected sex largely driving the increase (rates increased 11.2% during 2018 to 2019). In 2019, over 129,000 cases of all stages of syphilis were reported in the United States. Currently, men who have sex with men (MSM) account for approximately 80% of cases in the United States. Sex workers, transgender women, and incarcerated individuals have also been identified as at elevated risk for contracting syphilis. Individuals with HIV infection who are untreated, have low $CD4^+$ counts, or detectable HIV RNA levels are also at high risk due to both increased susceptibility and overlapping risk factors such as unprotected anal intercourse (in 2002 risk was reported to be 77 times greater in HIV-infected individuals than in the general population by the CDC).

Although syphilis is a reportable condition, rates of neurosyphilis in the United States are not known partially due to difficulties with data requirements to meet surveillance definitions. Similarly, although infectious syphilis is reportable in many countries, the rates of neurosyphilis are unknown as this stage of infection is often not reportable at the national level or insufficient data may be available to classify this stage of infection. Globally, there were an estimated 7 million new syphilis infections in 2020, with particularly high incidence in Africa. Rates of progression to neurosyphilis have been described to be up to 10% of patients with untreated syphilis, being more likely in HIV-infected individuals than those without HIV due to level of immunosuppression (patients with a CD4 count of ≤350 cells/μL have a threefold increase in neurologic involvement). The risk of neurosyphilis is two to three times greater in white than in black populations and is two times more common in men than in women.

PATHOPHYSIOLOGY. *Treponema pallidum* gains access to the body through mucosal membranes as well as epithelial breaks in the skin. This is the early stage of syphilis, *primary syphilis*, when rough induration occurs at the entry point of the pathogen approximately 10 days to 3 months after infection. This results in a painless ulcer, known as "chancre," with regional lymphadenopathy. In 60% to 70% of all patients, the primary ulcer remains the only manifestation of syphilis.

Mucopolysaccharidase helps *T. pallidum* attach to host cells, leading to obliterative endarteritis and necrotic changes in the terminal arterioles. Such an inflammation of the intima, the inner lining of the artery, can actually occlude the lumen of the vessel, leading to focal ischemia and a wide variety of focal neurologic deficits. On the other hand, some of the late neurologic complications of syphilis (i.e., general paresis and tabes dorsalis) are attributed to parenchymal damage from syphilis infection. This phase of hematogenic and lymphogenic spread (systemic infection) which happens within a few months of infection, is known as *secondary syphilis*. Approximately one-quarter of those with untreated infection develop the systemic illness which presents with constitutional symptoms (fever, fatigue, headache, joint pain, and/or muscle pain), adenopathy, *Condylomata lata* (hypertrophic papules most commonly found on the mucous membranes), and a characteristic rash, classically maculopapular that affects the palms and the soles. All organs can be affected, including the CNS. Approximately 25% to 40% of patients have "neuroinvasion" at some point, typically during the primary or secondary stage of infection, with up to 10% of syphilis cases developing symptomatic neurosyphilis. In most cases, the infection from the CSF is spontaneously cleared without requiring treatment for neurosyphilis and potentially without having any symptoms (*asymptomatic neurosyphilis),* while others may present with aseptic meningitis including lymphocytic pleocytosis in the CSF.

In the late phase, after a latent period (many years after infection), disease is known as *tertiary syphilis.* Presents with tuberous skin changes, ulcerating granulomatous lesions in different organs (gummas), and cardiovascular changes (mesaortitis, aneurysms), along with neuropsychiatric symptoms including mood disturbance (both depressed mood and euphoria), anxiety, hallucinations, delusional thought, personality change, apathy, dementia, disinhibition, and irritability.

CLINICAL PRESENTATION OF NEUROSYPHILIS. Early neurosyphilis can include *syphilitic meningitis*, characterized by headache, meningism, nausea/vomiting, and cranial nerve lesions (affecting one or more of the third, seventh, and eighth nerves), along with potential optic nerve involvement, polyradicular symptoms, and vascular brainstem syndromes; *meningovascular neurosyphilis* resulting from the ischemic changes caused by proliferative endarteritis, causing permanent CNS damage, and presenting most commonly as a stroke syndrome. The latter can also occur many years after infection and is therefore often included as a manifestation of late neurosyphilis as well. During the early stage of neurosyphilis, there may be mild neuropsychiatric symptoms, including personality change, emotional lability, insomnia, and decreased memory.

Late neurosyphilis, which generally starts 10 to 20 years after infection can include *tabes dorsalis:* chronically progressive dorsal radiculoganglionitis with loss of reflexes in the lower extremities, pallanesthesia, pupil disorders, hyperextensibility of the knee and hip joints, micturition disorders, and optic damage; *paralytic neurosyphilis/general paresis:* peaking in incidence 20 to 30 years after infection, a chronic progressive frontotemporal meningoencephalitis with ongoing loss of global cortical function; *syphilitic gummas:* circumscribed granulomas that develop from the meninges of the cerebral convexity–described as a brain mass with irregular enhancement, dura-contact, and surrounding edema (if intracranial pressure is increased, brain edema, or aggravated neurologic dysfunction may need surgical intervention); and/or neuropsychiatric symptoms: typical symptoms are increasing cognitive deficits, weakness of discrimination and judgment, psychotic episodes, speech disorders, headache and dizziness, abnormal pupillary reaction, tremor of the tongue, facial tremor, epileptic seizures, reflex anomalies, and finally severe dementia, urinary and fecal incontinence, and marasmus. The disease can mimic any psychiatric disorder, making it impossible to distinguish between primary psychiatric illness and psychiatric presentations of neurosyphilis based on clinical phenomenology alone. As per data on neurocognitive testing, patients with neurosyphilis (more specifically those with paralytic neurosyphilis/general paresis) indicated worse scores than the normal control group in memory, language, attention, executive function, and spatial ability. This result suggested that there is an overall decline in cognitive function. Moreover, when compared to a group of AD patients (matched for age, education, and MMSE scores), the group with neurosyphilis presented a similar pattern of cognitive impairment to those with AD. More advanced cognitive symptoms associated with neurosyphilis are associated with increased risk for neuropsychiatric symptoms such as mood disturbance and psychosis.

Atypical neurosyphilis: The more recent literature includes cases of temporal lobe encephalitis that resembled herpes simplex virus 1 (HSV-1) infection or limbic encephalitis, in which the main clinical manifestations are cognitive deficits and epileptic seizures.

DIAGNOSIS. There is no gold standard for the diagnosis of neurosyphilis, which relies on a combination of clinical findings, serologic tests for syphilis, CSF analysis, and, on occasion, neuroimaging. Serologic tests for syphilis include the nontreponemal venereal disease research laboratory (VDRL) test or rapid plasma reagin (RPR) test and, to confirm positive results, the fluorescent treponemal antibody absorption (FTA-ABS) test. More recently, real-time PCR techniques have been developed, which allow for the detection of *T. pallidum* with estimates of sensitivity of 80% and specificity of 98%. CSF studies are essential for diagnosis of neurosyphilis and can also serve to detect asymptomatic involvement so that treatment can be started and efficacy followed. In most cases, a normal CSF excludes neurosyphilis. The different CSF tests are as follows: CSF VDRL—highly specific and is generally accepted as the diagnostic test of neurosyphilis. A positive test and appropriate clinical history establishes the diagnosis of neurosyphilis. CSF RPR—lower sensitivity than VDRL. A positive result and appropriate clinical history also establish the diagnosis of neurosyphilis. CSF treponemal tests (FTA-ABS)—highly sensitive; however, it is nonspecific and thus more useful to rule out neurosyphilis when the pretest probability is moderate to low. CSF pleocytosis: the diagnosis of neurosyphilis usually needs a CSF white blood cell (WBC) count of 20 cells/μL or more. Sensitive but nonspecific (increased cell count occurs in both infectious and noninfectious causes). Patients with comorbid HIV may have pleocytosis regardless of the presence of neurosyphilis, particularly if they are not on antiretroviral therapy (ART). CSF protein: elevated levels support the diagnosis, but are neither sensitive nor specific.

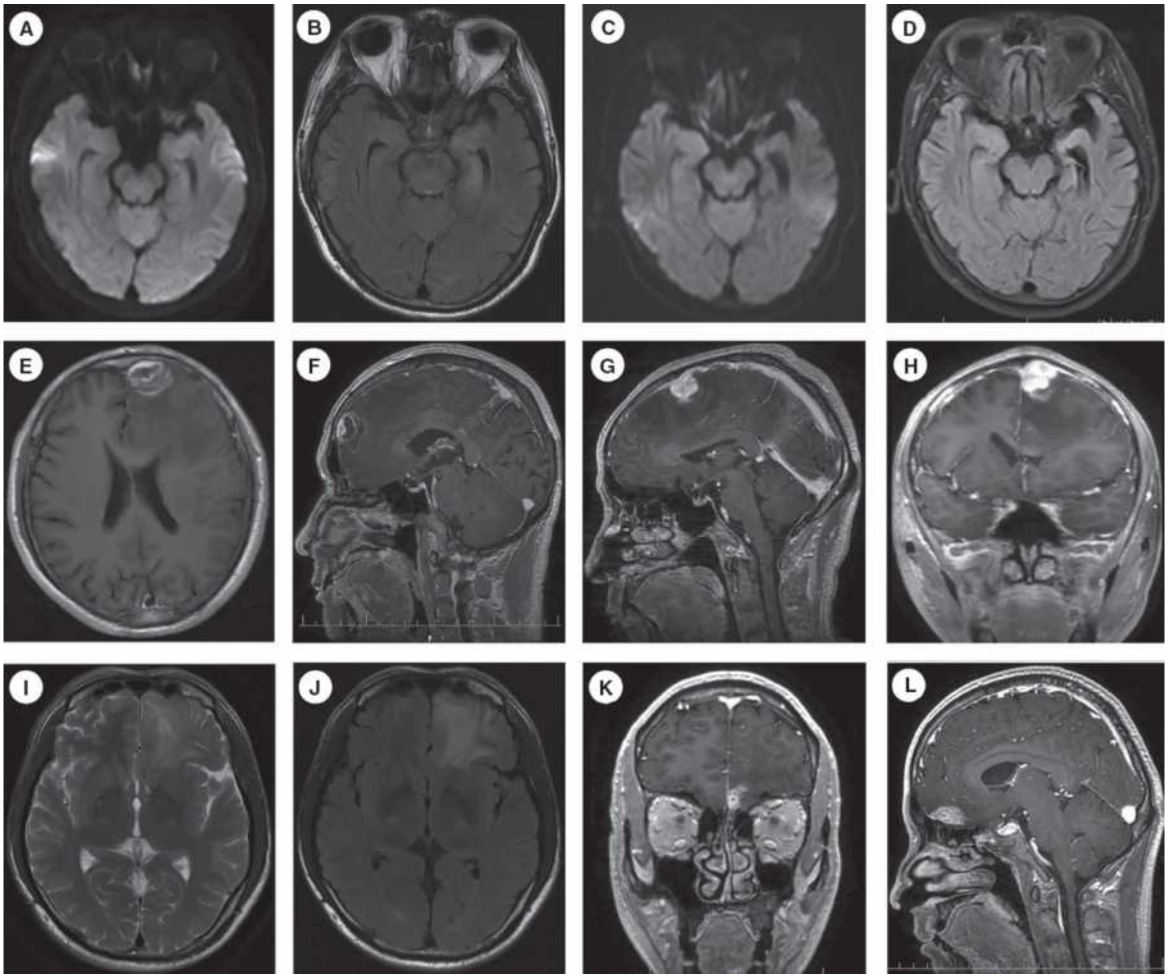

FIGURE 2.9–1. Neuroimaging (MRI) of neurosyphilis patients illustrating a variety of manifestations. Images **A–D** demonstrate abnormal signal in left temporal lobe and hippocampus (**A, B**) progressing to left temporal lobe atrophy and hippocampal sclerosis (**C, D**) 2 years later despite treatment for syphilis. Images **E–F** and **G–H** show two separate patients with left frontal lobe nodular enhanced lesion with surrounding edema. Images **I–L** exhibit an example of a cortical nodular lesion at the base of the frontal lobe. (Reprinted with permission from Shang XJ, He CF, Tang B. et al. Neuroimaging features, follow-up analyses, and comparisons between asymptomatic and symptomatic neurosyphilis. *Dermatol Ther (Heidelb).* 2020;10:273–283. doi: 10.1007/s13555-020-00361-3.)

Neuroimaging studies can be helpful in the diagnosis and management, but findings are generally nonspecific and can be diverse. The most common findings are frontal and temporoparietal atrophy. Patients with general paresis reveal frontocortical atrophy and nonspecific white matter lesions in a frontal distribution. Figure 2.9–1 shows examples of some neuroimaging findings in neurosyphilis, including temporal atrophy and frontal lesions. SPECT imaging in general paresis reveals a marked reduction in cerebral perfusion, particularly in the bilateral frontal and temporal cortices. Changes consistent with infarction are usually encountered in patients with meningovascular disease.

Novel biomarkers are being studied such as beta-2 microglobulin. Its presence in CSF may help in the diagnosis of CNS involvement and may aid in monitoring the response to therapy.

MANAGEMENT. CDC guidelines recommend parenteral penicillin G for all stages of syphilis and for neurosyphilis. Standard courses of antibiotics for 10 to 14 days consist of IV aqueous penicillin G (divided doses every 4 hours), or alternatively intramuscular injections of benzathine penicillin (once weekly) or procaine penicillin (four times a day).

The goal in clinical neurosyphilis is to reverse the manifestations or arrest the disease progression, although, in some patients, antibiotic therapy may not be able to achieve these goals. Certain conditions, such as comorbid HIV infection, may place patients at greater risk for persistence of treponemal infection after antibiotic treatment. A large review found that in HIV-positive patients treatment with antibiotics is associated with serologic failure (defined as a less than fourfold decline in RPR/VDRL titer) in 0% to 33% of cases. ART reduces this risk by roughly 60%.

In terms of psychiatric symptoms, there are no definitive guidelines for proceeding with the utilization of psychotropic medication in combination with appropriate antibiotic therapy, although in published case reports it has been suggested to have been effective in reducing symptoms. Haloperidol, as well as the atypical agents risperidone and quetiapine at or below recommended doses for schizophrenia, appear to have clinical benefits for psychosis in this population. Furthermore, the addition of divalproex sodium appeared to exert some benefit in mood stabilization and agitation. As in AD, no clear standard exists for the management of noncognitive neuropsychiatric symptoms of syphilis.

After commencing treatment, tests of cure via VDRL/RPR are recommended in all patients. Unlike other infectious diseases, the definition of cure cannot include documentation of complete eradication of the infecting organism because *T. pallidum* cannot be cultured in vitro. A fourfold drop in titer is considered an appropriate reduction at 12 months after treatment of primary, secondary, or early latent syphilis (early syphilis) and at 24 months for late latent syphilis and syphilis of unknown duration (late syphilis) if the initial titer was high. Serum monitoring should be performed monthly for three months, and at 6 and 12 months for early syphilis. For patients with late syphilis, this should take place every 6 months until negative. Due to the strong association with HIV, patients that are HIV-positive should have repeat testing yearly. Regarding neurosyphilis, CSF studies should be tested every 6 months until negative.

The likelihood of marked improvement for patients with general paresis is less than that for patients with syphilitic meningitis or meningovascular syphilis, reflecting the pathologic process, which in the former is irreversible neuron damage and in the latter CNS inflammation.

Lyme Disease

Epidemiology. Lyme disease is a multisystem infection caused by *Borrelia burgdorferi* (see Table 2.9–1). Lyme disease was first described in the United States in the late 1970s during an investigation of an unusual epidemic of what appeared to be juvenile rheumatoid arthritis in Lyme, Connecticut. However, retrospective serologic data suggest that Lyme disease was endemic in the United States as early as 1962, and the clinical manifestations of this infection have been known in Europe since the early 1900s. The causative agent was isolated from *Ixodes* ticks in 1982, allowing the start of a better understanding of this disease. Lyme disease is the most common tick-borne disease in both the United States and Europe; an estimated 476,000 cases are diagnosed and treated per year in the United States and >200,000 cases per year in western Europe. The risk for acquiring Lyme disease in the United States varies with the distribution, density, and prevalence of infection in vector ticks. Most cases of Lyme disease occur in the northeastern and north-central states. The incidence of Lyme disease in Europe is highest in the Scandinavian and Baltic states in northern Europe and in Austria, the Czech Republic, Germany, and Slovenia in central Europe. Lyme disease also occurs in Russia, Japan, and China.

Lyme disease is a reportable disease in the United States; for surveillance purposes, confirmed infection is defined as either (1) physician-diagnosed erythema migrans greater than 5 cm in diameter in a person exposed to ticks in a Lyme-endemic region or laboratory evidence of infection or (2) one or more objective late manifestations of (arthritis, neurologic signs, or cardiac conduction defects) with laboratory evidence of infection with *B. burgdorferi*.

Pathophysiology. *B. burgdorferi* is transmitted to man by *Ixodes* ticks that become infected while feeding on infected animals. Laboratory studies indicate that transmission of *B. burgdorferi* by the ticks requires a minimum of 36 to 48 hours of attachment. The natural hosts for *B. burgdorferi* are wild and domesticated animals, including mice and other rodents, deer, sheep, cattle, horses, and dogs. Infection in these animals may not be apparent, although clinical infection has been observed in cattle, horses, and dogs. In areas endemic for Lyme disease, 2% to 50% of ticks may carry *B. burgdorferi*. Most Lyme disease cases result from bites by infected nymphs, as their small size (about the size of a poppy seed) allows them to easily go unnoticed. Infections are most common during the months of May through August (northern hemisphere), when both the nymph ticks' activity and human outdoor activity are at their peak.

Of note, congenital infection may occur with serious, potentially fatal, consequences for the fetus.

Clinical Presentation. Clinical manifestations may differ in the United States and Europe, mostly due to significant differences in the bacterial strains causing infection in the two continents. As a result, *B. burgdorferi* is divided into three different genospecies distributed in two antigenic distinct groups.

- *B. afzeli* and *B. garinii* are responsible for most Lyme disease in Europe and Asia and are associated with chronic skin and neurologic symptoms, respectively.
- *B. burgdorferi sensu stricto* is the sole cause of Lyme disease in the United States where Lyme arthritis is a common complication of infection.

Clinically, Lyme disease is divided into early and late disease. Lyme disease usually begins with the characteristic skin lesion: a spreading annular rash (sometimes appearing as a bull's eye) known as erythema chronicum migrans (ECM), at the site of the tick bite between 3 and 22 days after infection (ECM may be absent in about 10% of patients with early Lyme disease). In the United States, secondary skin lesions similar to those of ECM are common as *B. burgdorferi* rapidly disseminates by the blood stream throughout the body. The rash and secondary lesions fade within 3 to 4 weeks. The most common early neurologic manifestations are cranial neuropathy, particularly facial nerve palsy (cranial nerve VII palsy, which may be bilateral), headache, lymphocytic meningitis, and motor or sensory radiculoneuritis, in addition to encephalitis characterized by mood lability and disturbances of memory or sleep. *B. burgdorferi* has been found in the CNS as soon as 2 weeks after initial skin infection. Between 15% and 40% of patients may have neurologic signs as their presenting feature.

Cardiac involvement occurs in 4% to 8% of untreated patients, the most common feature being a fluctuating atrioventricular block. Constitutional symptoms, arthralgias, myalgias, and severe fatigue are common, and regional lymphadenopathy also may occur. Asymmetric oligoarticular arthritis, especially involving the knees, occurs in about 60% of untreated patients at a mean of 6 months from the onset of the disease.

Patients may also have asymptomatic infection or only nonspecific symptoms.

In Europe, *B. afzelii* and *B. garinii* infection cause a firm, painless, bluish-red nodular lesion usually localized on the earlobe or nipple, known as Borrelia lymphocytoma.

B. burgdorferi may remain latent, causing illness months to years later (late disease). Partly because of this latency in disease expression, patients may be unable to recall the initial tick bite or rash.

Late neurologic Lyme disease may present as a subacute mild encephalopathy affecting memory and concentration; chronic mild axonal polyneuropathy manifested as distal paresthesias and/or, less commonly, as radicular pain.

Lyme disease has been reported to cause a vast array of neuropsychiatric disorders, ranging from depression/irritability and cognitive problems (short-term memory loss, impairment in processing speech, and/or verbal fluency problems) to, less commonly, psychosis, mania, or dementia. Although the full spectrum of psychiatric disorders has been associated with *B. burgdorferi* infection, by far the most frequent are disturbances of mood, characterized by irritability, dysthymia, and sleep loss, which are usually seen during the disseminated phase of the infection. Studies have looked at both the presence of Lyme in identified psychiatric patients and the emergence of neuropsychiatric symptoms in identified Lyme patients after becoming infected. In identified psychiatric patients, one study of 517 hospitalized psychiatric patients in a highly endemic area failed to find sufficient immune diagnostic reactivity to *B. burgdorferi* meeting CDC surveillance criteria in a single patient, while another study of 1,810 demonstrated a higher prevalence of antibodies to *B. burgdorferi* in hospitalized psychiatric patients when compared to matched pairs of healthy subjects (33% vs. 19%). Given these somewhat discordant findings, additional research would be beneficial.

In a recent study, United States and Danish researchers report that patients with diagnosis of Lyme disease had a 28% higher rate of mental disorders and were twice as likely to have attempted suicide postinfection, compared to individuals without the diagnosis. In addition, patients with diagnosed Lyme also had a 42% higher rate of affective disorders, such as depression and bipolar disorder, and a 75% higher rate of death by suicide than those without the diagnosis.

POST-TREATMENT LYME SYNDROME VERSUS CHRONIC LYME DISEASE. Two distinct terms refer to the problem of persistent symptoms after antibiotic treatment: *Post-treatment Lyme syndrome* refers to those patients who experience persistent nonspecific symptoms including fatigue, sleep disorders, headache, memory and concentration difficulties, myalgias, and arthralgias after antibiotic treatment of well-defined prior Lyme disease. *Chronic Lyme disease* can include a vastly heterogeneous patient population, including patients with late Lyme disease, those with post-Lyme disease syndrome, and those who have no objective evidence of prior Lyme disease (patients exhibiting non-specific illness who do not test positive on FDA-approved serologic tests). While the exact cause of persistent symptoms after treatment is unclear and most likely multidetermined, it is possible that some symptoms represent ongoing postinfectious secondary immune effects.

Diagnosis. Currently in the United States, a two-tiered serology protocol is the only validated diagnostic approach for LD diagnosis recommended by the CDC. This is typically an enzyme-linked immunoassay (ELISA) to detect IgM or IgG antibodies to *B. burgdorferi* in serum and if the sample is positive or equivocal on the screening assay, then a Western Blot is used to detect serum IgM (during first 30 days of infection) or IgG antibodies (after 30 days of infection) to *B. burgdorferi*. Use of IgM testing is recommended during the first 30 days of infection, after which only IgG tests should be used.

False positive results may occur due to cross-reactivity with other microorganisms; a problem particularly among patients with syphilis infection, but also found in multiple other infections (leptospirosis, periodontitis, relapsing fever, tularemia, Southern tick-associated rash illness (STARI), and Epstein–Barr virus infection) and with some autoimmune disorders (rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis). Thus, if the clinical syndrome does not fit, clinician judgment should be used, and other more specific tests should be performed (i.e., indirect hemagglutination antibody (IHA) test using *B. burgdorferi* strains).

False-negative results may occur if the patient is tested before the opportunity for an appropriate antibody response (too soon after infection). Therefore, if clinical suspicion is strong and symptom onset is recent, repeat testing is indicated.

Other laboratory studies include: polymerase chain reaction (PCR) for borrelial DNA, which is highly specific for *B. burgdorferi* DNA but rarely positive beyond the stage of the rash (the organism is tissue-tropic and does not reside for prolonged periods in the bloodstream) and antigen detection assays. CSF analysis (all prior laboratory studies can be performed in CSF) has a critical role ruling out other possible causes of CNS disease and establishing diagnosis of Lyme meningitis or encephalitis. Lymphocytic pleocytosis, mildly increased protein, and an elevated *Borrelia*-specific IgG index are observed. ELISA of CSF should be drawn at the same time as the serum ELISA so that a Lyme index can be calculated. In rare cases, Lyme-specific antibodies at diagnostically high levels may be found in the CSF but not in the serum. It has been reported that in as high as 20% of cases of late-stage neurologic Lyme disease, the CSF may appear normal. MRI may show punctate white matter lesions on T2-weighted images, EEG is usually normal, but diffuse slowing or epileptiform discharges have been described, and SPECT/PET can demonstrate a non-specific pattern of diffuse heterogeneous hypoperfusion in the late-stage CNS Lyme disease.

Management. Antibiotic therapy including penicillins, macrolides, cephalosporins, and tetracyclines have all been used successfully. However, it appears that tetracyclines (e.g., doxycycline) may produce fewer late complications than penicillin therapy. For early Lyme disease without CNS involvement, a 2- to 3-week course of oral antibiotics is recommended, while for Lyme disease with CNS involvement, the recommendation is IV antibiotic treatment for 3 to 4 weeks. Patients with Lyme encephalopathy have gradual improvement in their symptoms, usually starting a few months after completion of therapy, and continue to improve slowly for up to 1 to 2 years. While early treatment with antibiotics at the time of the skin rash (erythema migrans) often results in rapid resolution of the illness, delayed treatment may result in a more entrenched set of symptoms and a longer course of illness. Postexposure antimicrobial prophylaxis with a single 200-mg dose of doxycycline has been shown to reduce the risk for Lyme disease after an *I. scapularis* tick bite and is recommended for consideration in the United States.

Leptospirosis (Weil Disease)

Leptospira is the agent of Weil disease, which can have CNS manifestations (see Table 2.9–1). Leptospirosis is a zoonotic disease caused by pathogenic spirochetes of the genus Leptospira, and carriers include wild and domesticated animals such as rodents, cattle, dogs, and pigs. It is transmitted by direct or indirect contact with the urine, blood, and tissues of infected animals (most commonly found in rodents, cattle, dogs, and pigs; though any mammal can be a carrier). In addition, infection can occur via mucosal surfaces or damaged skin exposed to contaminated environments such as moist soil or contaminated water. Spirochetes colonize the proximal renal tubules of infected reservoir hosts and are excreted in the urine. Between half to one million infections occur annually worldwide with approximately 60,000 deaths a year. Leptospirosis is more prevalent in tropical countries primarily due to occupational exposure than in developed countries where it is associated with water-related recreational activities. The disease is expected to become more prevalent in developing countries where slum settlements, climate change, and rapid urbanization lead to increased transmission.

The disease spectrum ranges from mild anicteric leptospirosis, characterized by fever, headache, myalgia, sore throat, and cough to a more serious Weil syndrome manifested as jaundice, renal dysfunction, and bleeding diathesis including pulmonary hemorrhage. Neurologic manifestations of leptospirosis include encephalitis (altered sensorium, psychosis, seizures, hemiplegia), intracranial bleeds, cerebellitis, movement disorders, myelitis, flaccid paraplegias (including Guillain–Barré syndrome-like presentation), mononeuritis, neuralgias, autonomic lability, polymyositis, and ADEM. However, the most common manifestation of neuroleptospirosis is aseptic meningitis and could account for 5% to 13% of all aseptic meningitis cases (presence of clinical and laboratory evidence of meningeal inflammation in the absence of positive routine bacterial cultures in CSF).

Diagnosis relies on medical history of possible occupational or recreational exposure and appropriate clinical findings, confirmed by laboratory testing. The microscopic agglutination test (MAT), which can indicate the likely infecting serogroup or serovar, is generally accepted as the “gold standard.” ELISA (detecting IgM in early infection) and PCR are also used. The detection of leptospires with dark-field microscopy offers the earliest confirmation of infection (in blood within the first week of illness and, less reliably, in urine during the second week). Blood culture may be useful in severe, fulminant disease.

Antibiotics are preferably started within 24 to 48 hours of illness onset. In severe illness, intravenous benzylpenicillin is the drug of choice. For milder infections a 7- to 10-day course of oral amoxicillin is appropriate. Patients allergic to penicillins can be treated with erythromycin. The value of antibiotic treatment has limited support from clinical trials, however, supportive management to maintain tissue and organ function, such as the temporary maintenance of renal function by dialysis, may be life-saving.

Group A β-Hemolytic Streptococcus

Group A beta-hemolytic streptococcus (GAS) can cause several postinfectious, immune-mediated diseases including acute rheumatic fever (with Sydenham chorea as only neurologic manifestation), poststreptococcal reactive arthritis, pediatric autoimmune neuropsychiatric disorders associated with streptococcus infection (PANDAS), and poststreptococcal glomerulonephritis. This chapter will focus on Sydenham chorea and PANDAS, both of which share an array of neuropsychiatric symptoms (see Table 2.9–1). However, distinguishing one from the other, especially at onset of symptoms, can prove challenging.

Epidemiology. Since GAS infections are found worldwide in humans, they are considered a global health problem. However, the distribution among different populations and the incidence of infection vary by time and geographic region, which presumably reflects a population's susceptibility to particular strains, but also the natural variation in the predominant types. Although GAS infections may be observed in persons of any age, the prevalence of infection is higher in children, presumably because of the combination of multiple exposures (in schools or nurseries, for example) and host immunity. The age of onset in both neuropsychiatric manifestations of GAS is typically prepubertal. Sydenham chorea usually occurs between the ages of 5 and 15 years and PANDAS symptoms are most often found in patients between 3 years of age and onset of puberty (with mean age of 6). Regarding gender differences, a female predominance (2:1) is found in Sydenham chorea studies whereas in PANDAS, boys outnumber girls by a ratio of 2.6:1. Sydenham chorea is the only recognized neurologic manifestation of rheumatic fever, occurring in about 20% of patients, most often weeks to months after group A streptococcal infection. It is important to note that diagnosis for these pathologies is often delayed by an average of 8 years from disease onset, highlighting the need to increase clinicians' awareness of this condition. Indeed, the majority of patients receive the diagnosis after consulting with multiple specialists.

Pathophysiology. Streptococcus pyogenes (group A beta-hemolytic streptococcus [GABHS]) is an anaerobic gram-positive coccus whose only reservoir is humans. The pathophysiology of both PANDAS and Sydenham chorea is thought to be autoimmune. Antibodies produced against GABHS in the context of acute infection are proposed to cross-react with basal ganglia epitopes, disrupting cortical-basal ganglia–thalamo–cortical (CBGTC) circuits, leading to motor, behavioral, and cognitive symptoms. Genetic vulnerability seems to play an important role in disease expression. The rate of tic disorders and OCD in first-degree relatives of pediatric probands with PANDAS is higher than those reported in the general population. The majority of children exposed to GABHS infections remain unaffected which suggests genetic or other vulnerability or susceptibility in some individuals. Inquiring about a family history of tics and OCD in first-degree relatives would therefore be deemed important in both conditions.

Clinical Presentation. PANDAS encompasses a complex of obsessive-compulsive disorders (OCD) and/or tics, both characterized by a pattern of remitting and relapsing symptoms, with the initial episode occurring abruptly or even overnight. Caregivers may describe their child as ‘‘completely new and unrecognizable.’’ Within a short time, children with PANDAS may also develop cognitive decline and behavioral regression, including attention difficulty, severe anxiety, deterioration in handwriting or other fine-motor skills, decline in school performance, night terror, and occasionally even psychosis. Urinary urgency or frequency may be observed as an initial manifestation, frequently misdiagnosed as urinary tract infection in the absence of other typical neurologic symptoms.

Sydenham Chorea is characterized by the abrupt onset of choreoathetoid movements. Although chorea is the most important neurologic symptom, hypotonia may be prominent—and both combined can lead to severe problems with feeding, dressing, and walking—and tics can occur as well. In severe cases, speech may become dysarthric with a slurred and explosive character. Psychiatric symptoms can also be prominent (present in up to 70% of patients with Sydenham chorea) and include emotional lability, anxiety, depression, obsessive–compulsive symptoms, as well as ADHD features, and even psychosis. OCD has been extensively described in relation with Sydenham chorea, with different time courses of onset (before, concomitantly, and after the characteristic choreatic movements had begun). Prevalence of patients with OCD has been reported up to 40% in patients with Sydenham chorea, with the higher percentages observed in those with persistent chorea (more than 2 years since onset).

Diagnosis. To date, there is no single clinical manifestation, laboratory test, or imaging technique that is pathognomonic for either PANDAS or Sydenham chorea. Clinically, five criteria have been proposed for diagnosis of PANDAS: (1) presence of obsessive-compulsive or tic disorder; (2) prepubertal symptom onset; (3) abrupt onset (24 to 48 hours) and episodic course (relapse–remission cycle); (4) neurologic abnormalities, such as choreiform movements during symptom exacerbations; and (5) a temporal relationship between GABHS infection and symptom onset (typically 7 to 14 days). In addition, careful observation of the electrophysiologic nature of the abnormal movements may prove informative. The choreiform movements associated with Sydenham can be described as irregular, semipurposeful, avolitional, adventitious in nature, and almost always result in functional impairment. In contrast, the choreiform movements in PANDAS cause no functional impairment.

In terms of laboratory testing, serum autoantibody profile with elevated Antistreptolysin O and Anti-DNase B antibody titers have been described in both PANDAS and SC (elevated in up to 80% of PANDAS confirmed cases). Also, the presence of some other markers may suggest or support the diagnosis. Antibasal ganglial antibody (ABGA) are found in PANDAS in 100% of acute disease and 69% of persistent forms as well as at an undefined frequency in SC. Specific monoclonal antibodies (D8/17) have been observed in 85% of PANDAS cases and 89% of SC cases versus 17% of healthy controls. In addition, structural neuroimaging has shown volumetric increases in basal ganglia structures (caudate, putamen, globus pallidus) in both PANDAS and Sydenham chorea cases.

Management. As there are multiple neurologic, behavioral, and emotional symptoms involvement in both conditions (PANDAS and Sydenham chorea), a multimodal treatment approach is recommended, including cognitive behavior therapy, selective serotonin reuptake inhibitors, steroids, intravenous immunoglobulins, and plasmapheresis.

Pharmacologically, antibiotics penicillin, cephalosporin, and amoxicillin are all considered appropriate treatments for GAS infections. There has been some evidence that antibiotic treatment for GAS infections both prevents neurologic manifestations (e.g., tics) and improves OCD symptoms in PANDAS, but the data available are inconclusive and cannot determine the full efficacy of antibiotics in prophylaxing against tics and OCD. Children with Sydenham chorea require long-term benzathine penicillin prophylaxis to reduce the risk of rheumatic heart disease. Immunomodulatory treatments (IV immunoglobulin IVIG) and plasmapheresis have been shown to reduce neuropsychiatric symptom severity in children with PANDAS, and more specifically OCD symptoms and tics. Tonsillectomy may also represent an effective treatment option in children severely affected by PANDAS.

Pediatric Acute Onset Neuropsychiatric Symptoms

More recently a new broader diagnostic category was established, known as pediatric acute-onset neuropsychiatric symptoms (PANS) to include similar clinical presentations to PANDAS but that are more etiologically diverse and can involve infectious agents that are not exclusively due to Group A Streptococcal infections (see Table 2.9–1). Unlike PANDAS that requires a temporal relationship with GAS infection specifically, PANS do not require a specific infectious agent as a precursor to symptom onset, as other infections (e.g., mycoplasma pneumonia, viral infections, *B. burgdorferi*) have also been linked to new-onset OCD.

Criteria include (1) abrupt and dramatic onset of OCD or severely restricted food intake; (2) concurrent presentation of 2 out of 7 other neuropsychiatric symptoms: anxiety, emotional lability and/or depression, irritability, aggression, and/or severe oppositional behaviors, behavioral (developmental) regression, deterioration in school performance, sensory or motor abnormalities including heightened sensitivity to sensory stimuli, and somatic signs or symptoms such as sleep disturbance or enuresis; and (3) symptoms not better described by an otherwise known neurologic or medical condition such as Tourette disorder, Sydenham chorea, or systemic lupus erythematosus. No laboratory or imagining studies have been shown to help establish diagnosis.

As in patients with PANDAS, long-term antibiotic prophylaxis with benzathine benzylpenicillin has been shown to result in a substantial benefit to patients' outcome, along with potentially preventing relapses of neurological manifestations over time.

PROTOZOA

Malaria

Plasmodium is the genus of parasitic protozoans that cause malaria, and disease-causing species include *Plasmodium vivax* (the most common), *P. ovale*, *P. malariae*, *P. knowlesi*, and *P. falciparum* (see Table 2.9–1). Infection is acquired via the bite of the Anopheles mosquito. Malaria as a whole affects an estimated 240 million people annually, killing over 500,000. One study found neurologic involvement in 48% of children with malaria; those living in endemic areas of sub-Saharan Africa are at highest risk. Older adults, pregnant women, HIV-infected individuals, and patients with a history of splenectomy are also at increased risk for cerebral involvement. Parasitized red blood cells sequestered in cerebral capillaries, cytokine production/immune cell accumulation, and BBB endothelial lesions leading to edema and vascular injury all contribute to the mechanism of neurologic disease. Presentations of malaria involve cyclic spikes of high fever and chills, with frequency varying by species, as well as diaphoresis, fatigue, headache, myalgia, jaundice, and spleno- or hepato-megaly. *P. falciparum* carries the special risk of being associated with prominent neuropsychiatric symptoms, or cerebral malaria.

Cerebral malaria presents as encephalopathy that occurs as impaired consciousness, delirium, or seizures. Retinal hemorrhages occur in 30% to 40% of cases. Untreated, cerebral malaria is generally fatal; with treatment mortality drops to 15% to 20%. Survivors of pediatric cerebral malaria may have neuropsychiatric sequelae including cognitive and language deficits and behavioral problems, as well as seizures, deafness, blindness, and other neurologic problems. Autoimmune encephalitis may rarely develop after resolution of cerebral malaria. Diagnosis is made by blood smear microscopy. Low CSF glucose also suggests cerebral malaria over viral encephalitis. Intravenous artesunate is the treatment of choice. Management of cerebral edema and seizures are other neuropsychiatrically relevant considerations, but no agents have been demonstrated to be unequivocally beneficial in clinical trials.

Human African Trypanosomiasis

Trypanosomes are protozoan parasites found in Africa; *Trypanosoma brucei rhodesiense* causes a more acute and *T. brucei gambiense,* a more chronic form of trypanosomiasis (see Table 2.9–1). *T. b. gambiense* is the more common cause of illness in humans. Exposure to the tsetse fly, which transmits the infection, is greater in rural and poor communities. Initially, trypanosomes penetrate the skin and circulate via the blood and lymphatic system; subsequently they infect the CNS. Presentation may involve fever, weight loss, lymphadenopathy, and neurologic manifestations; CNS involvement may not occur for 2 years. Sleep disturbance with daytime somnolence may be striking, earning trypanosomiasis the designation “sleeping sickness.” Seizures, cognitive decline, psychosis, personality change, and sensory disturbance can also occur. Identification of trypanosomes in blood or CSF by microscopy is diagnostic. Treatment is with antitrypanosomal agents such as fexinidazole, pentamidine, nifurtimox, or eflornithine.

Toxoplasmosis

Toxoplasmosis is caused by *Toxoplasma gondii*, an obligate intracellular, neurotropic protozoan parasite (see Table 2.9–1). Cats are its definitive host, producing oocysts, which are found in their feces. Oocysts are most often transmitted to humans through accidental ingestion of contaminated food or water, for example with insufficient hand hygiene after handling cat litter or consumption of raw or undercooked contaminated meat. Infection can also occur transplacentally. Around one-third of the world's population is infected by *T. gondii*, making it one of the most successful human parasites. Infection leads to cyst formation, predominantly inside neurons and glial cells, and may be unifocal or multifocal. Symptomatic infection is most common in immunosuppressed hosts. Clinical presentation depends on the location and number of lesions. Usual symptoms include headache, fever, focal neurologic deficits, seizures, delirium, ataxia, lethargy, and visual alterations. Some patients may present with psychotic symptoms, cognitive dysfunction, and involuntary movements.

Immunocompetent individuals are usually asymptomatic. However, latent infection may predispose people to neuropsychiatric symptoms. This concept is gradually evolving based on converging evidence from studies among humans and rodents. Rodents infected with *T. gondii* have elevated dopamine levels, lose their natural aversion to the odor of cat urine, and may develop an attraction to cat's pheromones. Some human studies have found that *T. gondii* seropositivity is associated with higher likelihood of risk-taking, both financial (entrepreneurship) and physical (increased rate of motor vehicle accidents), as well as reduced memory performance. However, studies are inconsistent on whether *Toxoplasma* infection is associated with risk-taking. A recent meta-analysis did find a significant association between IgM antibodies against *T. gondii* and acute psychosis compared to controls. One strain of *T. gondii* (Type I) has been associated with an increased risk of psychosis in the offspring of infected mothers and demonstrates a greater effect on neuroepithelial host cell expression of genes involved in brain development and synaptic plasticity compared to other strains. Differences in strain may explain why, despite high seroprevalence of *Toxoplasma*, only a small proportion of the population has psychotic illness.

Congenital toxoplasmosis occurs in 1/1,000 to 1/10,000 live births. Maternal-fetal transmission risk is <20 % during the first trimester but increases to around 80% in the third trimester. Infection acquired earlier in gestation typically leads to more severe fetal outcomes, including spontaneous abortion, hydrocephaly, and intellectual disability, while infection occurring later in gestation leads to mild disease, with most cases being asymptomatic or presenting with recurrent chorioretinitis. Toxoplasmosis is diagnosed using a combination of clinical, serologic, and radiologic evidence. A presumptive diagnosis is often confirmed with a positive response to antiprotozoal drugs, with a favorable clinical response expected within 2 weeks. Serum or CSF PCR is also a useful method to diagnose *T. gondii* infection. CT head or MRI brain is critical for the diagnosis of cerebral toxoplasmosis. Single or multifocal lesions may be seen, with a typical pattern of a hypodense lesion with ring-enhancing and perilesional edema. Those with latent infection maintain IgG antibodies in their serum. Active infection is treated with pyrimethamine and sulfadiazine, pyrimethamine and clindamycin or atovaquone, or trimethoprim/sulfamethoxazole for at least 6 weeks. If immunosuppression is present, secondary prophylaxis with trimethoprim/sulfamethoxazole or pyrimethamine/sulfadiazine is continued indefinitely.

Amoebiasis

Amoebas are free-living environmental protozoa that do not require a host for survival, nor a vector for transmission. Four genera cause human disease, and all four can cause CNS infections: *Naegleria, Acanthamoeba, Balamuthia,* and *Sappinia.* CNS syndromes associated with amoebic infection are primary amoebic meningoencephalitis (PAM) and granulomatous amoebic encephalitis (GAE) (see Table 2.9–1).

Naegleria fowleri is found in freshwater and soil worldwide, and recreational water activities are a risk factor for infection, particularly if water is inhaled. Prevalence is low, with approximately 2.6 cases per million exposures. Presentation is typically PAM, an acute hemorrhagic meningoencephalitis with a mortality rate that approaches 100%. Symptoms include fever, headache, nausea/vomiting, photophobia, seizures, and altered mental status/behavioral changes. Uncal herniation can occur, and death typically occurs 10 days after exposure. Serum shows a neutrophilic leukocytosis, CSF shows frank hemorrhage and a markedly elevated white count, with elevated pressure; imaging is nonspecific. The exclusion of bacterial and viral encephalitis should raise suspicion for PAM, and the diagnosis is confirmed by observation of trophozoites in CSF. Optimal treatment regimen is unclear, but should include amoebicidal drugs that can cross the blood–brain barrier and steroids to control edema.

Acanthamoeba, Balamuthia mandrillaris, and *Sappinia pedata* can cause GAE, a subacute/chronic CNS infection. *Acanthamoeba* largely affects the immunocompromised, and is often transmitted via inhalation or via skin contact leading to hematogenous spread. Headache, low-grade fever, visual disturbances, behavioral abnormalities, and focal neurologic deficits can progressively worsen over weeks to months, eventually leading to seizures, elevated ICP, coma, and death. Brain biopsy is required for diagnosis, and will show trophozoites and cysts. Treatment relies on multiple drugs with anti-amoebic activity. *B. mandrillaris* can cause disease in the immunocompetent, and soil or water exposure may be related. An initial skin lesion is followed by development of neurologic manifestations and progresses to death within 3 months. *S. pedata* has been the identified cause of a single case of GAE.

HELMINTHS

Neurocysticercosis

Neurocysticercosis is caused by CNS penetration of *Taenia solium* larvae and is endemic to most of the developing world, including Central and South America, sub-Saharan Africa, the Indian subcontinent, most of Southeast Asia, and parts of China (see Table 2.9–1). It has a major public health impact since 30% of the epilepsy cases in these regions can be attributed to this disease. Epidemiologic studies have shown that specific serum antibodies may be seen in >10% and intraparenchymal brain calcifications on CT may be present in 10% to 20% of the general population in the endemic zones, suggesting high prevalence rates, though most cases are asymptomatic.

Humans acquire neurocysticercosis through fecal-oral route via close contact with a *T. solium* carrier excreting *T. solium* eggs in their feces. The larval cysts can form in most tissues, including the brain, and develop to their usual size in 2 to 3 months. CNS lesions can vary from a few millimeters to 1 to 2 centimeters in size and are found at the gray/white matter junction in the cerebral hemispheres or in the deep gray matter. Clinical symptoms appear years after the infection. The clinical symptoms, management, and prognosis depend on the CNS parasite burden, parasite size, location of the lesions, and host inflammatory response.

Patients with exclusively intraparenchymal brain lesions typically present with seizures only, which tend to subside over the years. Neurocysticercosis may sometimes present with "cysticercotic encephalitis" where hundreds of small cysts are seen with severe diffuse inflammation, seizures, and raised intracranial pressure. This presentation occurs more commonly in children and young women, reflecting an intense host immune response. Antiparasitic treatment alone in these cases may worsen the symptoms and lead to mortality.

Lesions in the ventricular and basal subarachnoid space cause hydrocephalus and/or intracranial hypertension, leading to significant morbidity and mortality. Mass effect may lead to focal neurologic deficits, endocrinologic, and ophthalmologic abnormalities. Fourth ventricle cysts may cause Bruns syndrome, presenting with episodic severe headache, vomiting, vertigo, and sudden loss of consciousness related to head movement due to ball valve mechanism. Cognitive deficits are common, ranging from mildly impaired cognitive performance to severe dementia. Psychotic episodes can occur, likely as a complication of epilepsy.

The diagnosis is made using neuroimaging, with CT head used for screening and MRI brain, where available, used for better lesion characterization, while serology is employed to confirm diagnostically challenging cases.

In most patients, seizures respond well to monotherapy with first-line antiseizure medications. The optimal duration of antiseizure medication use is not very clear. More than 50% of patients develop recurrent seizures after discontinuing antiseizure medications. Multiple lesions, a history of recurrent seizures, and the presence of brain calcification(s) all predict a higher likelihood of seizure relapse. The underlying parasitic infection, including CNS lesions, is treated with anthelmintics like praziquantel and albendazole. Corticosteroids are used to reduce inflammation in case of encephalitis, angiitis, and arachnoiditis, and to control the host immune response when anthelmintic treatment leads to exposed parasitic antigens causing local inflammation.

FURTHER READINGS

Bandara A, Kalaivarny G, Perera N, Indrakumar J. Aseptic meningitis as the initial presentation of Leptospira borgpetersenii serovar Tarassovi: two case reports and a literature review. *BMC Infect Dis*. 2021; 21(1):488.

Baringer JR. Herpes simplex infections of the nervous system. *Neurol Clin*. 2008;26(3):657–674.

Campbell O, Krause PJ. The emergence of human Powassan virus infection in North America. *Ticks Tick Borne Dis*. 2020;11(6):101540.

Chams N, Chams S, Badran R, et al. COVID-19: a multidisciplinary review. *Front Public Health*. 2020;8:383.

Cheslack-Postava K, Brown AS. Prenatal infection and schizophrenia: a decade of further progress. *Schizophr Res*. 2021;247:7–15.

Christensen SS, Eslick GD. Cerebral malaria as a risk factor for the development of epilepsy and other long-term neurological conditions: a meta-analysis. *Trans R Soc Trop Med Hyg*. 2015;109(4):233–238.

Desmettre T. Toxoplasmosis and behavioural changes. *J Fr Ophtalmol*. 2020;43(3):e89–93.

Forrester NL, Kenney JL, Deardorff E, Wang E, Weaver SC. Western equine encephalitis submergence: lack of evidence for a decline in virus virulence. *Virology*. 2008;380(2):170–172.

Frye MA, Coombes BJ, McElroy SL. Association of cytomegalovirus and toxoplasma gondii antibody titers with bipolar disorder. *JAMA Psychiatry.* 2019;76(12):1285–1293.

Garcia HH, Tanowitz HB, Brutto OHD. *Neuroparasitology and Tropical Neurology*. 1st ed. Elsevier; 2013.

Greenwood D, Slack R, Peutherer J, Barer M. *Medical Microbiology A Guide to Microbial Infections*. Churchill Livingstone; 2007.

Guzman MG, Harris E. Dengue. *Lancet*. 2015;385(9966):453–465.

Ha T, Tadi P, Dubensky L. Neurosyphilis. [Updated 2021 Sep 29]. In: *StatPearls [Internet].* StatPearls Publishing; 2021.

Hampton MM. Congenital toxoplasmosis: a review. *Neonatal Netw*. 2015;34(5):274–278.

Jain A, Marshall J, Buikema A. Autism occurrence by MMR vaccine status among US children with older siblings with and without autism. *JAMA.* 2015;313(15):1534–1540.

Kennedy PGE. Update on human African trypanosomiasis (sleeping sickness). *J Neurol*. 2019;266(9):2334–2337.

Luzolo AL, Ngoyi DM. Cerebral malaria. *Brain Res Bull*. 2019;145:53–58.

Manolis TA, Apostolopoulos EJ, Manolis AA, et al. COVID-19 infection: a neuropsychiatric perspective. *J Neuropsychiatry Clin Neurosci*. 2021;33(4):266–279.

Marques AR, Strle F, Wormser GP. Comparison of Lyme disease in the United States and Europe. *Emerging Infect Dis*. 2021;27(8):2017–2024.

Mehta R, Gerardin P, de Brito CAA, et al. The neurological complications of chikungunya virus: a systematic review. *Rev Med Virol*. 2018;28(3):e1978.

Morens DM, Folkers GK, Fauci AS. Eastern equine encephalitis virus–another emergent arbovirus in the United States. *N Engl J Med*. 2019;381(21):1989–1992.

Muthu V, Rudramurthy SM, Chakrabarti A, Agarwal R. Epidemiology and pathophysiology of COVID-19-associated mucormycosis: India versus the rest of the world. *Mycopathologia.* 2021;186:1–16.

Nath A, Berger JR, eds. *Clinical Neurovirology*. 2nd ed. CRC Press, Taylor & Francis Group; 2020.

Petersen LR, Brault AC, Nasci RS. West Nile virus: review of the literature. *JAMA*. 2013;310(3):308–315.

Portegies P, Corssmit N. Epstein-Barr virus and the nervous system. *Curr Opin Neurol*. 2000;13(3):301–304.
Ropper AH. Neurosyphilis. *N Engl J Med*. 2019;381(14):1358–1363. Erratum in: N Engl J Med. 2019 Oct 31;381(18):1789.
Seidel J. Primary amebic meningoencephalitis. *Pediatr Clin North Am*. 1985;32(4):881–892.
Singer HS, Loiselle C. PANDAS: a commentary. *J Psychosom Res*. 2003;55(1):31–39.
Sudre CH, Murray B, Varsavsky T, et al. Attributes and predictors of long COVID. *Nat Med*. 2021;27(4):626–631.
Teixeira AL, Vasconcelos LP, Nunes M, Singer H. Sydenham's chorea: from pathophysiology to therapeutics. *Expert Rev Neurother*. 2021;21(8):913–922.
Wakelin D. Helminths: pathogenesis and defenses. *Medical Microbiology*, 4th ed. Chapter 87.

2.10 Neuropsychiatric Aspects of Prion Disease

Brian Giunta, M.D., Ph.D., Alireza Minagar, M.D., M.B.A., Paul Shapshak, Ph.D., Nadejda Alekseeva, M.D., and Francisco Fernandez, M.D.

Transmissible spongiform encephalopathies (TSEs), also known as prion diseases, are an unusual and uncommon group of neurodegenerative disorders that are caused by conformational changes, misprocessing, and malfunction of the cellular prion protein (PrP^{C}). In 1982, Stanley B. Prusiner proposed the term prions to distinguish the spongiform group of nervous system disorders (TSEs) from other conventional infectious disorders caused by infective agents such as viruses. Prusiner pointed out that prions cause TSEs and are *pro*teinaceous *in*fectious particles that resist inactivation by most procedures that denature proteins and nucleic acids.

Prions (a member of a group of pathologies known as "proteinopathies") self-aggregate and are capable of generation of abnormal molecular aggregates with toxic intracellular signaling. The scrapie isoform of the PrP^{C} (PrP^{Sc}) is the infectious agent behind diseases such as Creutzfeldt–Jakob disease (CJD) in humans. After transmission to a naive host, prions initiate misfolding of the PrP^{C} in an autocatalytic manner, resulting in an exponential increase in PrP^{Sc} in the brain and spinal cord, culminating in neuronal death. The primary amino acid sequence of PrP^{Sc} is governed by host PrP^{C}. PrP^{C} is encoded by the prion gene, *PRNP*, which is located on the short arm of chromosome 20 (20pter-p12). In humans, neurodegenerative disorders caused by prions can present as hereditary/genetic (about 14% of cases), sporadic (85%), or acquired infectious disorders of the central nervous system (CNS; 1%).

The human prion diseases are sporadic Creutzfeldt–Jakob disease (sCJD), the variant Creutzfeldt–Jakob disease (vCJD), iatrogenic Creutzfeldt–Jakob disease (iCJD), familial CJD (fCJD), fatal familial insomnia (FFI), Gerstmann–Sträussler–Scheinker (GSS) disease, and Kuru. sCJD is rare and 100% fatal; it typically presents in persons with no known risk factors or inherited genetic mutations. vCJD is believed to be acquired through the consumption of beef products made from cattle infected with bovine spongiform encephalopathy (BSE), an animal prion disease commonly referred to as “mad cow” disease. iCJD is known to have been caused by corneal and meningeal grafts contaminated with prions, human growth hormone (hGH), prion-contaminated surgical instruments, and blood transfusions. Familial forms of CJD are transmitted as autosomal dominant traits with *PRNP* mutations, although other genetic loci have been implicated in the pathogenesis of human prion diseases. FFI is characterized by progressive insomnia and has been diagnosed in only about 25 to 40 families worldwide. Almost always inherited, GSS disease is an extremely rare and fatal prion disease observed in only a few families worldwide. Kuru is found among the Fore people of New Guinea. It was transmitted among members of the Fore tribe through ritualistic cannibalism when the Fore people prepared and consumed infected brain tissue of deceased family members. Today Kuru has nearly disappeared as a result of governmental discouragement of the practice of cannibalism.

The prion diseases of animals include scrapie, BSE or “mad cow disease,” chronic wasting disease (CWD), and transmissible mink encephalopathy (TME). In both humans and animals, these diseases are variations on a theme. That is, PrP^{C} conformational changes result in several phenotypes but have common end points: rapidly progressive dementia and death.

In humans, prion diseases are an uncommon cause of dementia and present with rapidly progressive cognitive and behavioral changes associated with visual and cerebellar impairments, extrapyramidal and pyramidal signs and symptoms, and akinetic mutism. In addition, as a group, these disorders share a spongiform degeneration of the brain and variable amyloid plaque formation. During the last two decades of the 20th century, worldwide concerns increased regarding BSE outbreaks in various regions of the world, particularly the United Kingdom, and its concomitant transmission of vCJD to humans. Thus, vCJD is a major concern for physicians and scientists, although its incidence has declined significantly since the 20th century. Clinically, sCJD differs from vCJD: in the latter, patients present with early psychiatric symptoms including anxiety, depression, apathy, and social withdrawal, with dementia being a late complication. In contrast, clinical features of sCJD include rapidly progressive dementia, ataxia, and myoclonic involuntary movements (Table 2.10–1). The emergence of multifocal neurologic deficits over days to weeks differentiates typical cases of sCJD from more common forms of dementia. Early diagnosis is not typical as cases may present initially with isolated syndromes, including cerebellar ataxia, acute stroke-like deficits, and cortical visual impairment (for review, see Will and Ironside, 2017). It should be noted that the causative agents of vCJD in humans and BSE in cattle share a common origin and differ from the causative agent of sCJD. This discovery stimulated interest in TSEs in general and in the possibility of their transmission from animals to humans.

NONFAMILIAL PRION DISEASES

Sporadic (Nonfamilial) CJD

sCJD has been found around the globe. The distribution of cases within countries or regions appears random, with no good evidence of clustering to suggest an environmental, infectious source. This is consistent with case-control studies of risk factors for sCJD development, which have been negative for the most part. Positive findings have been suggested to result from methodologic biases as opposed to authentic biologic risk factors. However, this is not the case for individuals with a history of prior surgery where an increased frequency of sCJD is observed. However, this is not consistent across all studies. For example, there is a lack of evidence of an increased risk from working in medical or paramedical occupations. Moreover, there is strong evidence that sCJD is not transmitted through blood transfusion. Indeed, the epidemiologic evidence points to sCJD developing because of a random conversion of PrP^{C} to PrP^{Sc} rather than from an external infectious source.

FAMILIAL GENETIC PRION DISEASES

Genetic prion diseases are categorized by their clinical symptoms and neuropathology. They include fCJD, FFI, and GSS disease. *PRNP* mutations are autosomal dominant, highly penetrant, and are composed of missense mutations, insertions, and deletions. Disease onset typically starts in the fifth or sixth decade of life.

Familial CJD

fCJD is defined as a form of CJD that results from mutation(s) in the gene that encodes PrP^{C} and exhibits autosomal dominant inheritance. Approximately half of the individuals carrying the mutated form of the PrP^{C} gene develop CJD. Although clinically, the presentation of fCJD varies with the mutation site within the *PrP* gene, it most closely resembles sCJD in its manifestations.

Fatal Familial Insomnia

FFI is a hereditary prion disease that is identified by disrupted sleep, motor abnormalities, and hyperactivation of the autonomic nervous system. As with fCJD, it is inherited by autosomal dominance with an atypical presentation and exclusive bilateral degeneration of the thalamus without the neuropathologic changes of spongiform degeneration or amyloid deposits. FFI is a midlife disease that affects both genders equally. The age of onset of FFI ranges from 36 to 62 years, and the disease is uniformly fatal and causes death in 8 to 72 months with an average of 18 months.

Table 2.10–1.
Characteristics of CJD Subtypes

Acquired	Kuru	Progressive cerebellar ataxia	▶ Chromosome location: 20pter-p12 ▶ OMIM#: 245300	▶ Atrophy of the cerebellar vermis and flocculonodular lobe, ▶ Microscopically, loss of granule and Purkinje cells, fusiform swelling of the proximal portion of Purkinje cells' axons, and severe radial gliosis of Bergmann astrocytes
	Iatrogenic CJD	Ataxia	Dependent on the mutations of the prion protein gene of the original contaminant	Dependent on the mutations of the prion protein gene of the original contaminant
	Variant	▶ Psychiatric and behavioral symptoms, dysesthesias ▶ Median age of onset: 26 years (range 12–74) ▶ Median duration: 13 months (range 6–39) ▶ Cerebellar signs in 100% of patients	▶ Chromosome location: 20pter-p12 ▶ OMIM #: 176640, 123400 ▶ val129 provides resistance to vCJD prion	▶ Abundant PAS-reactive, PrP amyloid plaques in the cerebrum and cerebellum ▶ Diffuse PrP plaques ▶ Cerebral plaques may be surrounded by a halo of spongiform vacuoles and form florid plaques. ▶ Other PrP plaques and amorphous pericellular and perivascular PrP deposits in the cerebellar molecular layer. ▶ Caudate nucleus and putamen with spongiform alterations ▶ Thalamus with severe neuronal loss and intense gliosis; most prominent thalamic nuclei ▶ No periodic EEG complexes ▶ CSF 14–3–3 protein in 33% of patients ▶ Readily detected in lymphoid tissue ▶ Frequently absent hyperintense signal in the caudate and putamen on diffusion-weighted and FLAIR sequences of MRI ▶ Pulvinar sign on brain MRI are frequently present ▶ Marked increased ratio of protease-resistant glycoform of prion protein on immunoblot analysis
Inherited	Familial CJD	Psychiatric and behavioral abnormalities, dementia, ataxia, and myoclonus	▶ Mutation(s) in the gene that encodes PrP. ▶ Autosomal dominant pattern of inheritance. ▶ Chromosome location: 20pter-p12 ▶ OMIM #: 176640, 123400 ▶ Mutation: met129	▶ Presentation of fCJD varies with site of mutation within the PrP gene ▶ Resembles sCJD in its manifestations.
	Fatal familial insomnia	Sleep disturbance, autonomic instability	▶ Mutation(s) in the gene that encodes PrP. ▶ Autosomal dominant pattern of inheritance. ▶ Chromosome location: 20pter-p12 ▶ OMIM#: 600072 ▶ Mutation: asp178-to-asn (D178N)	▶ Exclusive bilateral degeneration of the thalamus without the typical neuropathologic changes of spongiform degeneration or amyloid deposits
	Gerstmann–Sträussler–Scheinker (GSS) disease	Cerebellar abnormalities, hyporeflexia, dysesthesias	▶ Mutation(s) in the gene that encodes PrP. ▶ Conglomerate of at least six dominantly inherited syndromes, each of which is linked to a different mutation and GSS (Q217R) ▶ Chromosome location: 20pter-p12 ▶ OMIM#: 137440	▶ multicentric amyloid (PrP) plaques in the molecular layer of the cerebellar cortex, cerebral cortex, and basal ganglia. ▶ Plaques with a central larger mass surrounded by smaller satellite amyloid deposits. ▶ Each mutation coupled with different subtypes of PrP plaques ▶ Some mutations cause considerable neurofibrillary tangle neuronal degeneration and generation of neuritic plaques ▶ Amyloid plaques consist of highly truncated PrP

Sporadic	Sporadic CJD	▶ Progressive dementia, ataxia, myoclonus, 40% of patients with cerebellar signs ▶ Median age of onset: 65 years (range 15–94) Median duration 4 months (range 1–74)	▶ Chromosome location: 20pter-p12 ▶ OMIM #: 176640, 123400 ▶ Mutation: met129 and met/val129	▶ Macroscopically, spongiform encephalopathy in cerebral and/or cerebellar cortex and/or subcortical gray matter, and/or encephalopathy with rare PrP plaque immunoreactivity ▶ Microscopically, spongiform degeneration of neurons, neuronal loss, severe astrocytosis, formation of amyloid plaques ▶ >90% of patients with periodic EEG complexes ▶ CSF 14–3–3 protein in 99% of patients ▶ Not readily detected in lymphoid tissue ▶ Frequent presence of hyperintense signal in the caudate and putamen on diffusion-weighted and FLAIR sequences of MRI

In terms of neuropathology, FFI is ambiguous but has been observed to manifest as a focal neuronal loss in the thalamus, inferior olivary nucleus, cerebellum, and varying degrees of spongiform changes in the cerebral cortex. FFI is described to mainly affect the thalamus with a propensity for anterior ventral and mediodorsal thalamic nuclei. Other parts of the brain, including the inferior olives of the medulla oblongata and the cerebral cortex, have also been shown to be involved. The PrP^{Sc} deposition pattern favors the brainstem and thalamus earlier in the disease, but the thalamus is understood to be more vulnerable to subsequent degenerative changes. The reason for this involvement pattern is poorly understood but can explain the variety of symptoms seen in the disease.

Gerstmann–Sträussler–Scheinker Disease

GSS disease is another human prion disease that typically occurs in the fourth or fifth decade and mainly manifests with cerebellar ataxia and motor deficits. It has a familial autosomal dominant pattern of inheritance. GSS disease, like CJD, is not one single entity. There are at least six dominantly inherited syndromes with dissimilar clinical features, underlying neuropathology, and linkage to different mutations of the *PRNP* gene. GSS disease is an extremely rare neurodegenerative disorder with an incidence of 2 to 5 cases per 100 million.

Acquired Prion Diseases

Kuru is a prion disease that was discovered among geographically and ethnically isolated tribes in the Fore highlands of New Guinea. In the Fore language, Kuru means “trembling with fear.” Carlton Gajdusek and colleagues discovered that endocannibalistic funeral practices, by which deceased tribal members were ritualistically cooked and consumed by other family members, had a causal relationship to disease contraction.

During this ceremony, female relatives and children typically consumed the deceased's brain. It was demonstrated that women and children had an increased incidence of Kuru compared to others.

The development of CJD in small groups of individuals exposed to human tissues during medical procedures lends substantial support to the hypothesis of iatrogenic transmission of CJD. Indeed, iCJD has occurred in more than 200 patients treated with human pituitary hormones and in an almost equal number of patients receiving human dura mater grafts. A small number of cases of iCJD have been traced to corneal grafts, depth electrodes, as well as contaminated neurosurgical instruments. Cases linked to hGH and human dura mater grafts continue to be observed, indicating extended incubation periods of up to 42 and 30 years, respectively. The presentation of iCJD also varies in that it typically presents with motor findings (ataxia and gait abnormalities) associated with only mild dementing signs and symptoms toward the later phase of the disease. It is assumed that tissues from sCJD cases entered the production process and that pooling of pituitary glands or dura mater grafts led to the dissemination of infection. The incidence of hGH CJD varies by country, likely in relation to the level of contamination entering the production process. On the other hand, the rates for dura mater CJD likely reflect the frequency of use of this material in neurosurgery. For the most part, recombinant pituitary hormones and synthetic or autologous grafts are now used rather than human-derived treatments. As such, new iatrogenic exposures are unlikely to develop.

Nonhuman Mammalian Spongiform Encephalopathies

Scrapie is a fatal disease of sheep and goats that is characterized by chronic itching, loss of coordination, and progressive degeneration of the CNS. Studies during the 1940s and 1950s by Björn Sigurdsson, an Icelandic veterinarian, defined the new concept of slow infections, utilizing the Koch postulates via experimental verification.

TMEs are defined as infectious neurodegenerative diseases that affect human and animal hosts. It has been identified in the United States, Canada, Russia, Finland, and Germany and is believed to be a result of dietary exposure to a prion-infected animal. Nevertheless, the origin of the disease remains uncertain, and no cases have been reported for over 30 years.

CWD is a TSE that affects deer, elk, and moose. CWD is another slowly developing prion disease with a prolonged incubation period. Experimental inoculation of mule deer resulted in the discovery of a 2- to 4-year incubation time. Symptoms include physical and behavioral changes. Detection can be difficult, and these symptoms can be subtle early in the disease course or even fall within the normal range of behavior or seasonal body mass fluctuations. As such, diagnosis based on clinical signs is unreliable. Pathologic or biochemical analyses of the brain or lymphatic tissues are required. These would demonstrate neuronal vacuolation and spongiform changes, astrocytosis, and PrP^{Sc} accumulation in the brain or PrP^{Sc} deposits in lymphoid tissues. In most prion diseases, PrP^{Sc} and infectivity are primarily within the brain. However, in CWD, PrP^{Sc} can be identified in an array of extraneural tissues, body fluids, and excreta. This, in turn, promotes horizontal transmission.

Prions have caused multispecies epidemics and have spread as a zoonosis. BSE was first described in 1986, and within 10 years of its discovery, over 180,000 cases were diagnosed in cattle, zoo bovids, felids, and primates in the United Kingdom. It is currently classified as a potentially lethal zoonotic disease acquired *via* contaminated food. Indeed, the utilization of ruminant-derived meat and bone meal (MBM) as a food source for animals is the most well-accepted hypothesis for its spread in bovine populations. MBM contains CNS tissues and other tissues that contain the protease-resistant, disease-associated isoform of the PrP^{Sc}. BSE spread from the United Kingdom to at least 28 other countries, the majority of which are in Europe, with sporadic cases confirmed in Asia, the Middle East, and North America. BSE has an incubation period of approximately 2.5 to 8 years, with clinical manifestations usually observable in adult cattle at a peak age onset of 4 to 5 years. BSE is characterized by altered behavior, abnormal posture, incoordination and difficulty in rising, reductions in milk production, and weight loss. Histopathologic analysis reveals neuronal and neuropil vacuolation, glial reaction, and an absence of inflammatory lesions.

It should be mentioned that prion-like proteins also occur in fungi such as yeast. This demonstrates the importance of prion-like proteins in normal cell function as such proteins are present in fungi. For example, two yeast proteins, including ure2p and sup35p, are partially protease resistant and promote amyloid-like filament aggregation. Moreover, yeast “prion” proteins serve as paradigms for animal and human prion diseases. They also undergo conformational changes in their function, propagation, and strain variation. Yeast prions are infectious when propagated in the bacterial cytoplasm. Such findings are supportive of a wider significance than previously hypothesized for prions in biology.

HISTORY OF PRION DISEASES

The history of prion diseases commenced in the year 1920 when Hans Gerhard Creutzfeldt reported a 22-year-old woman with a mysterious and progressive focal syndrome of the CNS that was clinically identified by psychomotor abnormalities and cortical symptoms. Autopsy of this patient revealed prominent gliosis with noninflammatory focal lesions of the cerebral cortex. In 1921, Alfons Maria Jakob observed three additional cases of chronic neurologic disease, which he considered a new entity. In these patients, Jakob reported a spastic pseudosclerosis and encephalomyelitis with the spread of focal degeneration. A few years later, Walter Spielmeyer reviewed these cases and, based on their similarities, coined the name CJD. Subsequently, GSS disease was described in 1936 and FFI in 1986. In 1957, in the Kuru tribal region of New Guinea, Carlton Gajdusek and Walter Zigas characterized the clinical and neuropathologic features of Kuru and reported that the disease may be associated with cannibalism. In 1959, Hadlow distinguished certain similarities between Kuru and the spongiform encephalopathy of scrapie in sheep. This astute observation led to a series of experiments to transmit Kuru to nonhuman primates, accomplished by Gajdusek and Gibbs, and further discovery of salient similarities between Kuru and the spongiform encephalopathy of sheep scrapie. sCJD and sheep scrapie were found similarly related as well. Further research into the nature and pathogenesis of prion diseases was kindled after the discovery that BSE in the United Kingdom was causally related to vCJD. It is to Gajdusek’s credit that he was able to reduce the incidence of Kuru among the Fore people by convincing them to halt ritual cannibalism.

The precise nature of the transmissible infective agent in TSEs has been investigated for several decades. Initially, it was hypothesized that slow viruses cause TSEs. However, a lack of normative immune responses against TSE agents and the unusual resistance to certain treatment procedures that normally denature and inactivate nucleic acids led scientists to propose that “an agent devoid of nucleic acids” or a protein may be the causative agent of TSEs. In 1982, Bolton isolated a protease-resistant sialoglycoprotein, designated as PrP, from brain homogenates. Contemporaneously, Stanley Prusiner coined the term prion (from proteinaceous infectious particle) and characterized prions in molecular and biologic detail. It should be noted that since then, immune responses against prions have been accomplished.

EPIDEMIOLOGY OF PRION DISEASES

Ritual cannibalism, which was practiced by Fore tribes in New Guinea, transmitted Kuru to females as well as children (of both genders) who consumed the brain tissue of deceased individuals. Brain tissue possesses elevated quantities of transmissible infective agents, and its ingestion transmitted disease to these two groups more than to other groups who did not consume brain tissue. Although Kuru is transmitted horizontally, after the cessation of cannibalism in New Guinea, Kuru has no longer been observed.

The epidemiologic assessment of CJD only exists in certain countries where surveillance centers document every new case. As noted previously, CJD can be sporadic, acquired, genetic, or iatrogenic. Patients with CJD die rapidly, and annual statistics for incidence, prevalence, and mortality rates are problematic and not easily obtained. However, generally, the worldwide annual incidence of CJD is estimated at 1 million individuals per year. Human prion diseases affect most populations around the globe with a frequency of 1 to 1.5 cases per million, with some variation internationally. The United States averages approximately 300 diagnosed cases of prion diseases annually.

PRION MOLECULAR BIOLOGY AND PATHOGENESIS

Prion Introduction

Analysis of the biologic, biochemical, and molecular activity and properties of prion proteins PrP^{C} (cellular) and PrP^{Sc} (scrapie) enables one to better understand the pathogenesis of prion diseases. Research since the discovery and identification of prions supports the concept that prion diseases result from the conversion of the normal host-encoded cell-surface glycoprotein, PrP^{C}, into abnormally and conformationally altered isoforms PrP^{Sc}. These studies also analyze mechanisms of how the structure and function of prions of form PrP^{C} are altered to prions of form PrP^{Sc} and how these processes lead to the development of inexorable terminal, progressive neurodegenerative disease. A fundamental finding that sets prions apart from any other known life forms and also is contradictory to the original Watson–Crick “fundamental dogma” of molecular biology is that PrP^{Sc} proteins are infectious and can propagate in the absence of nucleic acids.

Molecular Pathogenesis

During prion disease, characteristic destructive CNS damage occurs in the form of spongiform vacuolation, widespread neuronal loss, microglial activation, the proliferation of astrocytes, and PrP^{Sc} production. Great strides are being made in scientists’ understanding of prion diseases, tropism, and molecular pathogenesis. There is extensive data pertaining to the mechanisms by which PrP^{Sc} prions gain access to the CNS after infection. Prions are generally transmissible by inoculation into, for example, subhuman primates, hamsters, and rodents. Intracerebral inoculation of brain homogenate is the most rapid method for inducing spongiform encephalopathy. In addition, TSE can be induced through peripheral routes, including intravenous (IV) and intraperitoneal injections, feeding, and via various aspects of introduction into the eye. Iatrogenic needle sticks, scalpel cuts, blood transfusions, and tissue grafts also result in human disease transmission of CJD and, as mentioned above, cannibalism-transmitted Kuru.

Cerebral accumulation of misfolded PrP^{Sc} is central to the pathophysiology of prion diseases. Several mechanisms are thought to underlie the toxicity of prion proteins. Dysregulated neuronal proteostasis takes place in prion disease at various levels. PrP^{Sc} disrupts the ubiquitin/proteasome system early in the disease progression, thus enhancing the buildup of PrP^{Sc} through impaired protein degradation. In turn, the accumulation of PrP^{Sc} disrupts the autophagy/lysosome pathway, which normally degrades aggregated proteins. Moreover, exhaustion of unfolded protein response (protection of the endoplasmic reticulum’s function in protein synthesis and sorting pathways) occurs early in the disease process. This initiates a feed-forward cycle wherein disrupted PrP^{C} trafficking, impaired PrP^{C} function, and translational cessation deteriorates the neurons, triggering synaptic loss and, eventually, inducing cell death pathways.

Studies have revealed the close interaction between PrP^{Sc} deposits and the neuronal plasma membrane. There are two major mechanisms through which toxicity may occur from this interaction. First, PrP^{Sc} aggregates cause a significant membrane disruption, conceivably by disturbing the functionality of neuronal receptors and thereby changing plasma membrane permeability. Glycosylphosphatidylinositol (GPI)-anchored PrP^{C} can proficiently transduce neurotoxicity, and prion disease is accelerated in mice in which PrP^{C} shedding is disrupted. In the second mechanism, membrane-bound PrP^{C} is thought to function as a receptor of prion toxicity. A direct interaction between PrP^{Sc} and PrP^{C} induces neurotoxicity analogous to the mechanism first described in Alzheimer disease (AD), wherein oligomeric amyloid β-peptides bind membrane PrP^{C} complexed to the metabotropic glutamate receptor mGluR5. This, in turn, sets off an intracellular signaling cascade where Fyn kinase is activated and culminates in synaptotoxicity. PrP^{C} has also been implicated in neurotoxicity. Antibody binding to the C-terminal globular domain results in neurotoxic signal generation through the N-terminus, inducing calpain activation and overproduction of reactive oxygen species (ROS). Likewise, PrP^{Sc} has also been demonstrated to cause similar toxic signaling transduction with calpain activation and ROS generation. It should be noted that the sequence of events leading to receptor-mediated neurotoxicity has yet to be fully elucidated, and GPI-anchored PrP^{C} would require a coreceptor to make intraneuronal signal transduction possible.

After contacting the intestinal epithelium, prions are able to spread via M cells. Once they have penetrated the subepithelial area, neurotropic prions, are believed to further transmit through retrograde axonal transport into the brain stem through the autonomic peripheral nervous system. Lymphotropic prions, including sheep scrapie, deer CWD, and possibly vCJD, spread within hours to Peyer patches and draining lymph nodes. In the lymphatic tissue, PrP^{Sc} builds up in the germinal centers of lymphoid follicles, where it persists. Follicular dendritic cells (FDCs) capture antigens on the plasma membrane for presentation to B cells. FDCs have also been shown to be productive replicators of prions, resulting in sustained lymphoid prion infections for months to years. On the surface of FDCs, CD21/35 receptors are likely to bind a PrP^{Sc} complement complex.

Subsequently, the hematogenous spread is promoted by eosinophils, neutrophils, monocytes, and mast cells. Ultimately, these cells cross the blood–brain barrier via multiple Trojan horse mechanisms. Nonetheless, it should be noted additionally, that cell-free blood and lymph spread occur as well. The timeframe and detailed mechanisms are under continued investigation.

Once prions have penetrated the brain and spinal cord, their numbers are further increased by neurons and astrocytes. Astrocytes are highly prone to in vitro prion infection and transfer prions to neurons. On the other hand, microglia are not a major factor in replication but rather are key for the clearance of prions. Oligodendrocytes do not possess any known substantial contribution to prion replication or CNS spread.

PrP^{c} concentration is highest in CNS (brain and spinal cord) neurons at both early stages of embryogenesis and in the adult, specifically in association with synaptic membranes as well as nicotinic acid/acetyl choline and NMDA receptors. Tissues infected by PrP^{Sc} prions extend beyond the CNS, including the spleen and muscle tissue of patients with sCJD. In addition, chronic inflammation can broaden the tropism of prion infectivity to other tissues, including the liver, pancreas, kidney, muscle, and mammary glands, although these tissues were originally considered to be prion free.

Cell Function Properties

It has been proposed that PrP^{C} is a multifunctional protein. PrP^{C} functions include its role in neural development, cell adhesion, axon guidance, synapse formation, neuroprotection, regulation of circadian rhythm, maintenance of myelin, maintenance of ion homeostasis, and signaling. Some of the most well-characterized functions are not directly due to PrP^{C} in its membrane-bound, GPI-anchored form but instead arise from soluble PrP^{C} fragments. For example, this has been reported for soluble PrP in regulating myelin homeostasis or in promoting neurite outgrowth. In the case of myelin maintenance, binding of N-terminal soluble PrP agonizes a Schwann cell G-protein–coupled receptor, Adgrg6 (Gpr126), whereas the molecular mechanisms subserving neurite outgrowth–promotion are unclear. It has been suggested that membrane-bound PrP^{C} functions as a receptor via homophilic interactions. In both cases, regulated proteolysis could serve as a form of functional regulation for transmitting the information. Indeed, such is the case for proteolytic cleavage fragments from the amyloid precursor protein (APP).

PrP^{C} may exert antiapoptotic protective activity in mammalian cells, mice, and yeast, which is due to internal or environmental stress factors that initiate apoptosis. For example, PrP^{C} is able to protect human fetal neurons in culture against Bax pathway apoptosis.

Oxidative stress affects several interconnected cellular pathways and can potentially cause mitochondrial dysfunction, damage to the ubiquitin protease system, aggregation of proteins, altered iron metabolism, and excitotoxicity. In contrast to the potential neurotoxicity discussed previously, several observations support a potential cytoprotective role for PrP^{C} against oxidative stress. PrP^{C} may protect cells from oxidative stress that is associated with superoxide dismutase (SOD) activity. It has been hypothesized that chronic oxidative stress of neurons is a major promoter of neurodegenerative disorders; this is exemplified in neuro-AIDS and drug abuse, in which there is an increase in neuronal nitric oxide synthase expression. It has been demonstrated that neurons (cerebellar granular and neocortical) cultured from *Prn-p*$^{0/0}$ mice are more susceptible than neurons from wild-type mice to treatments with oxidative agents such as hydrogen peroxide xanthine oxidase and copper irons. PrP^{C} appears to have both antiapoptotic and SOD activities that are obviously crucial neuron protective functions. However, it should be noted that work by other investigators finds no SOD activity associated with PrP^{C}. The still puzzling observation is that PrP^{C} has SOD activity but also contains copper ions within its normal structure. A discussion of the copper association with *PrP*^{C} follows.

Table 2.10–2.
Normal and Abnormal Forms of Prion Proteins

Features	Normal	Pathologic/Disease
Name	*PRNP, PrP*^{C}	*PrP*^{Sc}, CJD, GSS, FFI, etc.
Molecular weight	Monomer, 22–36 Kd	Stacked aggregates, >400 Kd
Cellular location	External, GPI linked at cell surface, synaptic clefts	Internal, Golgi, vesicles
Sensitivity to protease K	Yes	No
Biochemical structure	40% α-helices	40% β-pleated sheets
	3% β-sheets	30% α-helices
Cellular location	GPI linked at cell surface, synaptic clefts	Internal Golgi, vesicles, extracellular
Detergent miscibility	Miscible	Fibrils and rods form
Human tissue distribution	Brain, lymphocytes, heart, lung	Brain, lymphoid tissue
Concentration in normal brain	Variable, 1–5 μg/g tissues	Absent
Sensitivity to typical autoclave	—	No
Sensitivity to 4N NaOH hydrolysis	—	Yes
Apoptosis	Protective	Not protective
Superoxide dismutase	Yes	No
Binds Cu^{+2} cations	Yes	No
Coupled to tyrosine kinase Fyn	Yes	No

CJD, Creutzfeldt–Jakob disease; FFI, familial fatal insomnia; GPI, glycosyl phosphatidyl inositol; GSS, Gerstmann–Sträussler–Scheinker disease; *PrP*^{C}, prion protein cell; *PrP*^{Sc}, prion protein scrapie.

Data from Irani DN, Johnson RT. Diagnosis and prevention of bovine spongiform encephalopathy and variant Creutzfeldt-Jakob disease. *Annu Rev Med*. 2003;54:305; Johnson RT. *Viral Infections of the Nervous System*. 2nd ed. Lippincott-Raven; 1998; Roucou X, Gains M, LeBlanc AC. Neuroprotective functions of prion protein. *J Neurosci Res*. 2004;75(2):153; Van Rheede T, Smolenaars MMW, Madsen O, de Jong WW. Molecular evolution of the mammalian prion protein. *Mol Biol Evol*. 2003;20:111.

PrP^{C} may be a copper-binding protein. The histidine-containing octapeptide repeats (OR) bind up to four Cu^{+2} ions in a pH-dependent and negatively cooperative style. This binding interaction between *PrP*^{C} and Cu^{+2} ions changes the biologic and biochemical functions of the PrP^{C} (Table 2.10–2). Endocytosis of PrP^{C} via clathrin-coated pits is stimulated by micromolar concentrations of Cu^{+2} ions.

Structure

The physiologic form of the prion protein is GPI anchored and possesses two N-linked glycosylation sites. Mature PrP^{C} is composed of approximately 210 amino acids in a disordered N-terminal domain and a globular C-terminal domain of three α-helices and an antiparallel β-pleated sheet. PrP^{C} is largely present as a diglycosylated protein at the outer leaflet of the plasma membrane in lipid-enriched microdomains in its mature form. After internalization, PrP^{C} is recycled to the plasma membrane or the Golgi apparatus. Alternatively, it can be transported to late endosomes and then to intraluminal vesicles within multivesicular bodies (MVBs) for release as exosomes or for lysosomal degradation. PrP^{C} is a protease-sensitive sialoglycoprotein that possesses two N-linked complex oligosaccharide chains, like other GPI-anchored proteins; most of *PrP*^{C} is found in cholesterol-rich lipid membrane rafts.

Importantly, PrP^{C} is subject to α-cleavage and shedding of the protein. α-Cleavage results in an unstructured N-terminal protein fragment and a C-terminal globular fragment embedded in the cell membrane. These events occur during vesicular trafficking of PrP^{C} within the secretory pathway. The precise nature of the responsible protease remains uncertain. Cleavage occurring at the distal C-terminus of PrP^{C} and releasing almost full-length PrP is known as PrP shedding. This occurs only on the plasma membrane, and ADAM10 (a disintegrin and metalloproteinase-10) is the only relevant PrP sheddase, with diglycosylated PrP^{C} being the major substrate.

There are 23 enzymes involved in the biosynthesis of human GPI-linked proteins that are on 11 chromosomes, including two on chromosome 2. There are at least six GPI-linked proteins on neurons. A search for GPI-anchored human proteins at NCBI indicates that there are several hundred. Thus, compared to the entire proteome, there are few GPI-linked proteins. The function of this linkage is under investigation. The PrP^{C} molecule contains 253 amino acids and possesses two consensus sequences for N-linked glycosylation with molecular variants that have different degrees of glycosylation: unglycosylated, monoglycosylated, and diglycosylated. Some PrP^{C} is transferred to clathrin-coated pits where it undergoes endocytosis and recycling. PrP^{C} from humans, mice, Syrian hamsters, and cattle share common properties, including a flexible amino-terminal tail (residues 23 to 128), three α-helices, and a two-stranded antiparallel β-sheet that flanks the α-helix. A large flexible loop connects the second β-sheet and the third α-helix. The N-terminal portion of the *PrP*^{C} contains two conserved regions: a region of five repeats of octameric amino acid sequence, identified as the OR region (OR, octa-repeat region), and a second region with a high hydrophobicity (hydrophobic core), which is preceded by a hydrophilic domain known as charge cluster.

The OR region of PrP^{C} is significant for several reasons. This region contains the histidine residues that are capable of binding copper with copper-binding locations within the OR region. Copper complexing is associated with endocytosis of PrP^{C} and, through this, plays a role in the metabolism of this surface glycoprotein. The other reason for the significance of the OR region is that the expansion of the OR domain of up to 13 total repeats is also associated with genetic CJD and GSS disease (Table 2.10–3).

Table 2.10–3.
Prion Protein Structure

Location, Amino Acid Number[a]	Name	Comments
1–22	Signal peptide	Cleaved as part of protein maturation and migration
23–28	NLS-I	Nuclear localization signal
50	Pro	Proline hydroxylation site
56–128	Contains octapeptide repeats (variable)	Repeats of GWGQPHGG, with homology to TWIQDNGG (Bcl-2 BH2), and GWIQDNGG (Bax BH2)
151–165	Hydrophobic region	—
146/147	Lysine/histidine	Cleavage site
216 and 252	Cysteines	Disulfide bridge
218 and 234	ASN	Asparagine glycosylation site
164–168	B1	β strands
180–190	H1	α helical region
197–203	B2	β strands
208–233	H2	α helical region
237–269	H3	α helical region
275	GPI	Glycolipid-phospho-inositol anchor
276–298	Signal peptide	Cleaved as part of protein maturation and migration

[a] It should be noted that there is variation in numbering among different publications. One cause is variation in the number of octapeptide repeats.

Data from van Rheede T, Smolenaars MMW, Madsen O, de Jong WW. Molecular evolution of the mammalian prion protein. *Mol Biol Evol.* 2003;20(1):111–121; Roucou X, Gains M, LeBlanc AC. Neuroprotective functions of prion protein. *J Neurosci Res.* 2004;75(2):153–161.

PrP^{C} and PrP^{Sc} have the same primary amino acid sequence, and their different secondary structures impart dissimilar physiologic and physicochemical properties. A study of their secondary structures demonstrated that PrP^{C} contains a high α-helical content (roughly 42%) and small β-sheet content (3%). The tertiary structure of PrP^{C} shows that the normal molecule possesses three α-helical domains and two β-strands. It is believed that PrP^{Sc} resistance to degradation by protease is related to the alteration of its structure from a molecule with the α-helical structure into a different conformation molecule with a large β-sheet content (Ross et al., 2005). Accumulation of this protease-resistant isoform within neurons disrupts their normal function and results in vacuolization and widespread cell death. PrP^{Sc}, through a self-promoting or -catalyzing mechanism, converts normal PrP^{C} into PrP^{Sc} (Tables 2.10–2 and 2.10–3).

Strains

Different strains of prions are associated with different prion diseases and their separate neuropathologies and the clinical signs and symptoms as described in this section. In animal model studies, biochemical traits were preserved through several passages in prion disease rodent models. Work by I. H. Pattison and G. C. Millson in 1961 provided the first evidence for the presence and production of various prion strains. Goats infected by the same preparation of infectious scrapie agent developed two different clinical syndromes: scratching and drowsy. Prion strains are infectious isolates that show distinct prion-disease phenotypes, including incubation periods, neuropathology, and specific neuronal targets in several regions of the brain, in the same hosts, and with serial transmission. Several mechanisms have been proposed, including that prion strain-specific phenotypic traits may be encoded by an ancillary genome due to specific PrP^{Sc} genetics and conformations in donor inocula, selected postinfection, or by host genetics.

Host Protein Interactions

PrP^{C} interacts with other cellular proteins, and some of these interactions may relate to the pathogenesis of prion diseases. Many of these potential interaction molecules are primarily or entirely cytoplasmic proteins, and their interactions with PrP^{C} are under study. As mentioned, PrP^{C} is a GPI-anchored protein, and its entire polypeptide chain is external to the cytoplasmic membrane. Direct PrP^C interactions are probable with other receptors, secreted or transmembrane proteins. However, PrP^C contains a conserved hydrophobic sequence that can span the lipid bilayer in both directions. This results in two transmembrane variants, ^{Ntm}PrP (N-transmembrane) and ^{Ctm}PrP (C-transmembrane). In general, without predisposing mutations in *PRNP,* these variants are present in very small quantities.

Genetics

There are three prion-related genes, *PRNP, PRND,* and *SPRN,* and the detailed study of these genes commenced after the cloning and sequencing of the human, hamster, and mouse prion proteins, first accomplished in 1986.

PRNP. The *PRNP* gene in humans is located on chromosome 20, and its product is PrP^C (see the Online Mendelian Inheritance in Man [OMIM 1 www.omim.org]). Polymorphisms of *PRNP* have been extensively studied, and the polymorphisms in codon 129 may play a critical role in host susceptibility to prion diseases. For example, the presence of the V129 allele is not associated with vCJD, which raises the possibility that the presence of this allele is protective against vCJD superinfection. Some mutations in *PRNP,* including D178N, are associated with fCJD or FFI (depending on the residue *cis* at codon 129), and E200K is associated with fCJD. Moreover, P102L has been linked to GSS disease.

Prion Gene Complex, Downstream. *PRND* (prion gene complex, downstream) encodes the Doppel protein and is another gene related to prion diseases. This is a paradigm known for the human genome that is due to gene tandem duplications. The following are terms used for the Doppel PrP: DPL, prion-like protein Doppel precursor; MGC41841, prion gene complex, downstream; PrPLP, PrP2 (doublet); and dJ1068H6.4, prion-like protein Doppel. The observation that $Prnp^{0/0}$ knockout mice exhibited late-onset ataxia led to the discovery of the *PRNP* Doppel. These mutants had underlying neuropathology of cerebellar Purkinje cell loss. *PRNP* Doppel is located 16 kb downstream of the mouse *PRNP.* The significance of Doppel protein and prion pathogenesis is under further investigation.

Shadoo. The third prion-related gene encodes a short protein, known as Shadoo (*SPRN*), with similarities to the alanine-valine–rich central hydrophobic domain of PrP^C. The following are terms used for Shadoo (shadow precursor): FLJ41197, shadow of PrP homolog; and SHO, shadow of PrP. *SPRN* gene is not part of the PrP genomic complex and resides on the human chromosome 10. The significance of this protein and prion pathogenesis is under investigation.

Neuropathology

Neuropathologic studies of patients with Kuru have revealed that except for atrophy of the cerebellar vermis and flocculonodular lobe, the brain macroscopically may appear normal. Microscopically, prominent neuropathologic features manifest in the cerebellum with loss of granule and Purkinje cells, fusiform swelling of the proximal portion of Purkinje cells' axons, and severe radial gliosis of Bergmann astrocytes.

Macroscopically, the brains of patients with CJD may offer no specific diagnostic clues, or alternatively, one may observe various levels of cerebral cortical, striatal, and cerebellar atrophy. The WHO proposed neuropathologic criteria for CJD, including the presence of spongiform encephalopathy in cerebral and/or cerebellar cortex and/or subcortical gray matter and/or encephalopathy with PrP immunoreactivity (plaque and/or diffuse synaptic and/or patchy/perivacuolar types). Microscopically, the prominent neuropathologic features of CJD consist of spongiform degeneration of neurons and neuronal processes, significant neuronal loss, presence of severe astrocytosis, and formation of amyloid plaques. Spongiform degeneration that occurs in the context of CJD consists of the development of abundant rounded vacuoles within the neuritic processes and synapses and can be observed in the cerebral cortex (invariably in all cases regardless of the clinical manifestations), the subiculum of the hippocampus, caudate nucleus, thalamus, putamen, and the molecular layer of the cerebellum. Interestingly, the degree of spongiform degeneration and vacuolation in the same cortical section may vary from one area to another. The reactive astrocytosis that occurs during CJD involves the widespread presence of enlarged astrocytes in the cerebral cortex. In 10% of CJD cases, amyloid plaques that are positively immunoreactive with PrP antibodies and not immunoreactive with β-amyloid antibodies are present.

The neuropathology of vCJD demonstrates morphologic and immunocytochemical features that set it apart from other prion diseases. The most prominent neuropathologic feature of vCJD is the abundant presence of PrP^{Res} in the form of PAS-reactive, *PrP* amyloid plaques in the cerebrum and cerebellum. Many of the cerebral plaques are surrounded by a halo of spongiform vacuoles and form florid plaques. Other PrP plaques and amorphous pericellular and perivascular PrP deposits are prominent in the cerebellar molecular layer. The caudate nucleus and putamen are the focus of spongiform alterations, while the thalamus demonstrates severe neuronal loss and intense gliosis. These abnormalities are more pronounced in the posterior thalamic nuclei.

GSS disease is a TSE that is neuropathologically characterized by the presence of multicentric amyloid (PrP) plaques that are present in the molecular layer of the cerebellar cortex, cerebral cortex, and basal ganglia. The GSS disease–amyloid plaque characteristically contains a central larger mass surrounded by smaller satellite amyloid deposits. Like CJD, GSS disease is a conglomerate of at least six dominantly inherited syndromes, each of which is linked to a different mutation of the *PRNP* gene. These include GSS (P102L), GSS (P105L), GSS (A117V), GSS (Y145Stop), GSS (F198S), and GSS (Q217R), and these dissimilar syndromes possess four general properties on neuropathologic and molecular genetic grounds: (1) the clinical manifestations resulting from each mutation differs with the others; (2) each mutation is coupled with different subtypes of PrP plaques; (3) certain mutations are linked with considerable neurofibrillary tangle neuronal degeneration and generation of neuritic plaques; and (4) amyloid plaques present in the neuropathology of GSS syndromes consist of highly truncated PrP peptides.

The main neuropathologic feature of FFI is thalamic atrophy. The affected thalamic nuclei include anterior ventral, dorsal medial, pulvinar, and centromedian. Although atrophy of the cerebellar cortex is minimal, the inferior olives demonstrate atrophic changes. Some neuroscientists have argued that neuronal loss observed in the neuropathology of FFI may be caused by apoptosis.

DIAGNOSIS AND CLINICAL FEATURES

Prion diseases, as a unique group of spongiform encephalopathies that affect human and animal hosts, have become the subject of intense media and popular interest. The significance of this group of invariably fatal disorders is further demonstrated by two Nobel prizes won by D. Carleton Gajdusek and Stanley B. Prusiner in 1976 and 1997, respectively. Prion diseases are a rare but fatal group of neurologic disorders with unique pathophysiology. As a rule, in any patient with rapidly progressive dementia or other neurologic deficits and in the absence of another reasonable diagnosis, the presence of prion diseases should be investigated.

Neurologic Manifestations

A healthy 38-year-old right-handed female complained of weakness of the lower extremities, imbalance, and memory loss. Within a few weeks, this neurologic syndrome was followed by increasing ataxia, startle myoclonus, and moderate dysarthria. Three months later, she was admitted to the neurology service with profound lethargy, dementia, lip-smacking, spasticity of the lower extremities, seizures, and hallucinations. Her electroencephalogram (EEG) demonstrated the presence of periodic sharp wave complexes (PSWC) (Fig. 2.10–1), while her brain magnetic resonance imaging (MRI) was normal. She died within 6 months from the initiation of her neurologic disease, and an autopsy confirmed the diagnosis of CJD.

A 63-year-old previously healthy woman was admitted with visual agnosia, visual hallucinations, and cognitive decline followed by ataxia and depression. CT scan of the brain revealed mild brain atrophy, and her EEG was interpreted as slow and disorganized. Within 2 months, she developed tonic–clonic seizures, and her dementia rapidly deteriorated so that she developed akinetic mutism. MRI of the brain, axial T_2-weighted images revealed the presence of hyperintense signals in the caudate nuclei and putamen (Fig. 2.10–2), and upon examination of her cerebrospinal fluid (CSF), 14–3–3 proteins were detected. Four months after the onset of her neurologic symptoms, she died. Postmortem examination of her brain confirmed sCJD.

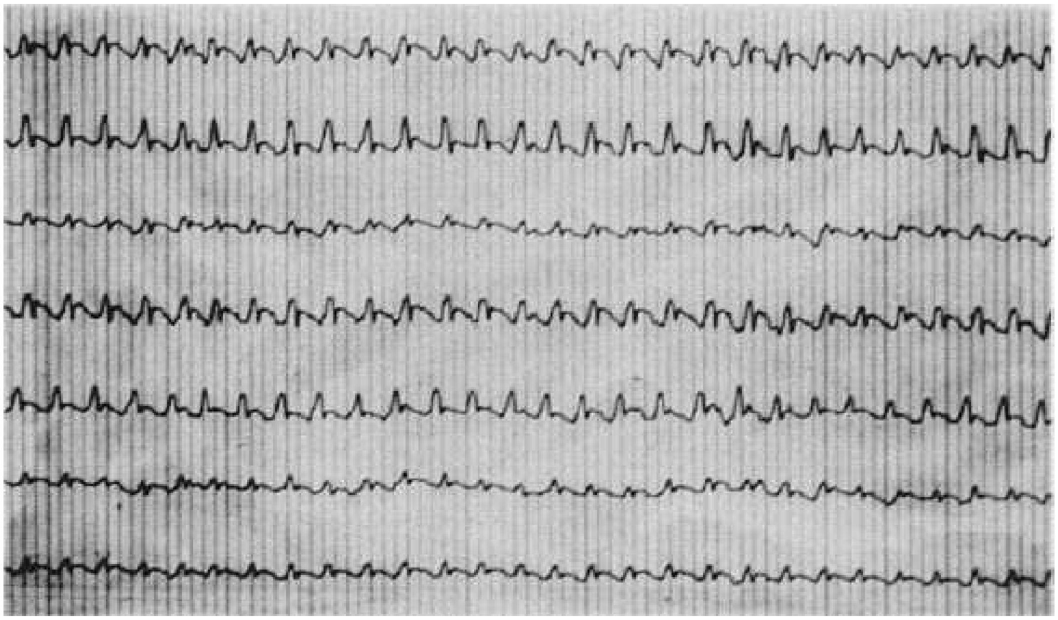

FIGURE 2.10–1. Typical triphasic periodic electroencephalogram of sporadic Creutzfeldt–Jakob disease. This shows the typical 1- to 2-Hz periodic electroencephalogram with triphasic waves that are seen diffusely in 70% of patients with sporadic Creutzfeldt–Jakob disease in the latter stages of illness. (Courtesy of Paul Brown.)

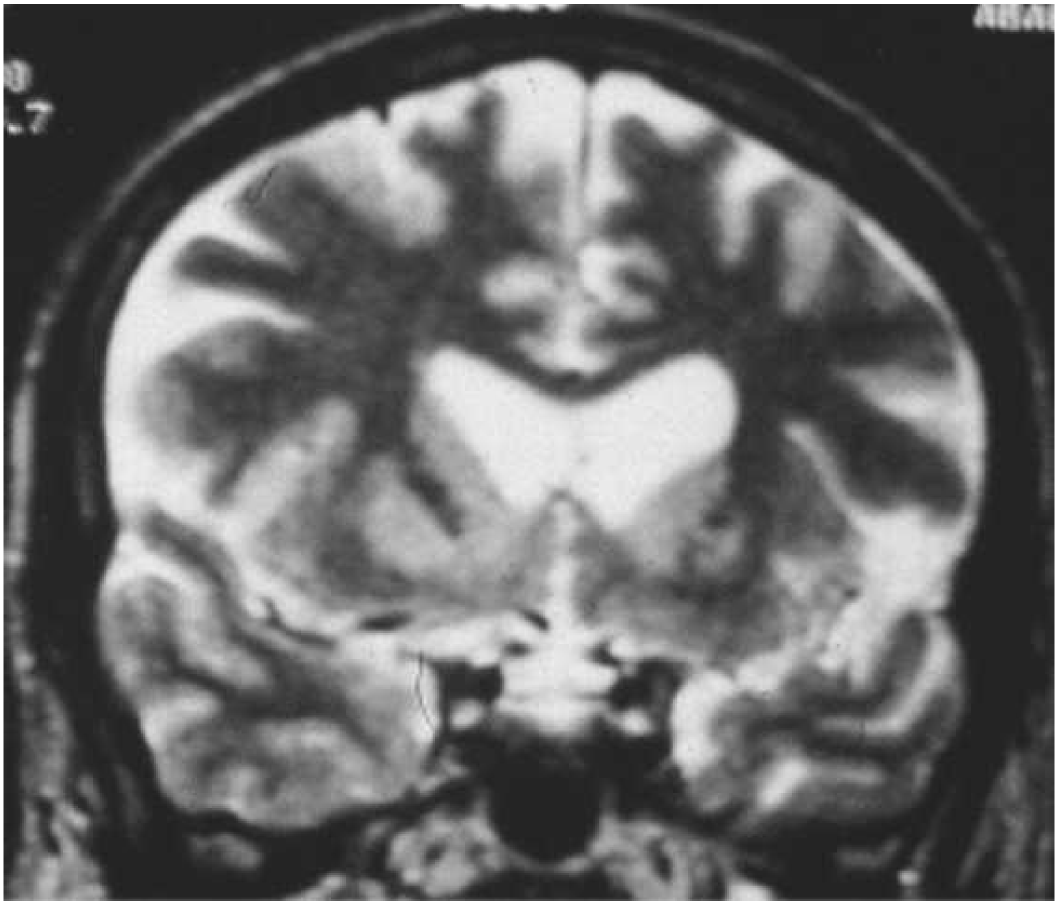

FIGURE 2.10–2. Magnetic resonance imaging coronal T_2-weighted image demonstrating hyperintense signals in the basal ganglia of a male patient with sporadic Creutzfeldt–Jakob disease with rapidly progressive dementia and myoclonic jerks. (Courtesy of Dr. Eduardo Gonzalez-Toledo.)

Sporadic CJD. sCJD is still the most common and most prominent human prion disease, which usually manifests insidiously and affects both genders equally. sCJD is a disease of late middle age that causes an invariably fatal disease. Usually, sCJD develops insidiously with nonspecific behavioral abnormalities, including anxiety, asthenia, depression, loss of appetite, alteration of the sleep pattern, weight loss, fatigue, dizziness, and social regression. Neurologically, many of these patients develop worsening forgetfulness and progressive decline of higher cortical functions such as reasoning, abstract thinking, calculation, and judgment. Almost one-third of these patients initially present with purely neurologic deficits, mainly cerebellar gait ataxia. A minority of patients present with a combination of cognitive impairment and focal neurologic deficits. With further progression of the disease process, patients rapidly become demented, aphasic, or apraxic, and demonstrate choreiform–athetoid movements, myoclonic jerks, and pyramidal and extrapyramidal signs. Initially, most patients do not manifest myoclonus; however, with disease progression, most patients develop myoclonus. Startle myoclonus is a prominent feature of CJD, which is precipitated by certain stimuli such as loud noises or touch. Other symptoms that appear with disease progression include paratonic rigidity, primitive reflexes, cortical blindness, and oculomotor disturbances. Other visual symptoms include diplopia, blurred vision, and visual agnosia. Terminal patients develop akinetic mutism, collapse into a coma, and die of infections or thromboembolic complications. Certain clinical features such as epileptic seizures (only in 10% of patients), sensory deficits, and lower motor neuron features are uncommon among these patients. In addition, the presence of certain signs or symptoms, such as early-onset seizures during the clinical course, acute onset motor deficits, cranial nerves palsy (except for visual loss or diplopia), and significant ataxia without concomitant cognitive decline, raise serious questions about the accuracy of the diagnosis of CJD.

The two rare clinical variants of CJD include Heidenhain and Brownell–Oppenheimer. In the Heidenhain variant of CJD, the neuropathologic changes mainly affect the occipital cortex and result in cortical blindness, optic hallucinations, and agnosia. Brownell–Oppenheimer variant manifests with a pure cerebellar syndrome secondary to widespread neuropathologic changes of the cerebellum (WHO diagnostic criteria for CJD: https://www.who.int/zoonoses/diseases/Creutzfeldt.pdf).

Variant CJD. In 1995, the vCJD form of human CJD emerged from the United Kingdom. vCJD, as a distinct variation, differs substantially from the more typical sCJD because of the following characteristics: Disease clinical picture, development in younger individuals as compared to CJD, initial dominance of psychiatric manifestations, underlying neuropathology, the absence of any mutations in the *PRNP* gene and almost exclusive occurrence in those with methionine homozygosity at codon 129, and its linkage to BSE. The exact mechanism of transmission to humans remains unknown; however, it is hypothesized that consumption of meat or meat products that were contaminated with BSE infectious agent in the late 1980s caused BSE migration of BSE prion from other species to humans, generating the vCJD epidemic. The potential causal link between vCJD and BSE has attracted public interest. In addition, to date, it remains unknown why the clinical and neuropathologic abnormalities of vCJD tend to affect younger adults.

In contrast to sCJD, vCJD can be transmitted through blood transfusion and presents with a relatively slow clinical course. The average age of onset of vCJD is 29 years as compared to the sCJD average age at onset of 65 years. Psychiatric symptoms such as depression, anxiety, agitation, delusions, and hallucinations are typically the initial manifestations, and within 6 months, they are followed by more characteristic neurologic features of CJD such as gait ataxia, dementia, dystonia, chorea, myoclonus, upward gaze paresis, and sensory abnormalities. In patients with vCJD, a nerve conduction study is usually normal; however, somatosensory-evoked responses may demonstrate minor abnormalities, indicating central involvement of pain pathways or thalamic origin for the pain. Like what is observed in sCJD, the end-stage patients with vCJD develop profound dementia along with akinetic mutism. Patients die after a median duration of 13 months (WHO diagnostic criteria for vCJD: https://www.who.int/zoonoses/diseases/Creutzfeldt.pdf).

Iatrogenic CJD. As noted previously, the accidental transmission of CJD to humans has been reported and well documented with corneal transplantation, implantation of contaminated EEG electrodes, implantation of dura matter grafts, and following use of contaminated HGH preparations derived from human pituitaries. In addition, one case of CJD following repair of an eardrum perforation with a pericardium graft has been reported. The disease can also be transmitted to humans by using inadequately sterilized neurosurgical instruments. Interestingly, in cases of central transmission (use of inadequately sterilized neurosurgical instruments or implanted EEG electrodes), patients mainly manifest with cognitive decline, while in peripheral cases such as the transmission of CJD via injection of contaminated HGH or gonadotropins, progressive ataxia followed by dementia predominate the clinical picture. Many of these patients develop myoclonic jerks.

Kuru. The word *Kuru* originates from the Fore word for shiver and refers to a progressive cerebellar syndrome that occurs among the highlanders of Papua New Guinea. Kuru affects individuals between the ages 5 and 60 years, with an equal ratio among male and female children but prominent excess in female adults. Kuru is transmitted by ritual cannibalism. Women and children, most of the victims of Kuru, participate in the cannibalistic feasts and consume the brain and intestines of the dead relative due to a social order within the tribe (Fig. 2.10–3). It is unclear when the first case of Kuru occurred in this population; however, it is believed that the first case began as sCJD, and then consumption of this infected deceased individual by other tribe members eventually led to the recycling of the prions within this isolated population and occurrence of Kuru at an epidemic proportion. By the late 1950s, cannibalism ceased, and the frequency of Kuru dropped significantly. If Kuru has continued with an even low incidence, then it is not purely due to cannibalism but may have a genetic component as well. In fact, that could be an alternative situation. That is a mutant develops in this highly homogeneous and isolated set of tribes in New Guinea. Their degree of isolation from each other is supported by the very large number of unrelated languages spoken there until they were discovered. Once the mutant prion gene was transmitted, then cannibalism that may have existed prior to that or occurred as a result then propagated the disease and its further spread. In that case, one would expect a continued low level of Kuru to linger after cannibalism ceased. However, secret cannibalism may still be practiced and could contribute to a very low incidence of Kuru. Interestingly, prior to the cessation of cannibalism, each potential patient was exposed to prions many times, which in turn made it difficult to precisely determine the incubation period of Kuru. The Kuru disease process commences with a vague prodrome characterized by malaise, arthralgia, and headache. Neurologically, Kuru usually manifests with gait instability followed by balance problems and frequent falls until the individual can no longer walk independently or sit without support. Gait ataxia is accompanied by dysmetria, dysarthria, incoordination of the upper extremities, various movement disorders such as clonus, chorea, and athetosis, emotional incontinence with inappropriate laughter, and convergent strabismus. Interestingly, weakness and rigidity are absent. Kuru patients' cognition is initially intact, and usually, in the advanced stages of the disease, dementia manifests. As the disease advances, dementia usually occurs, and patients become apathetic and withdrawn. In the end, patients become completely incapacitated and dependent on one another for daily life activities. Usually, within 1 year of disease onset, patients die of aspiration pneumonia, sepsis from decubitus ulcers, or inanition.

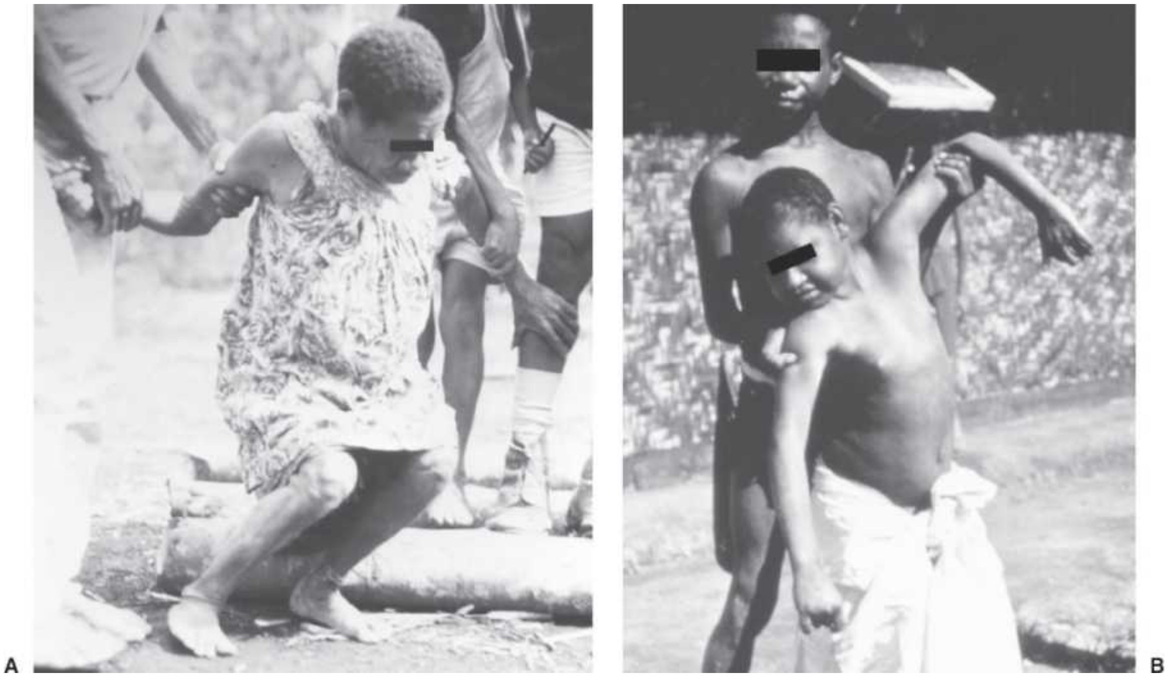

FIGURE 2.10–3. Two kuru patients. These figures show a young woman (**A**) and a child (**B**) in the late ambulatory phases of kuru. (Courtesy of Kimbra Kenney, M.D.)

Gerstmann–Sträussler–Scheinker Disease. GSS disease was initially described by Josef Gerstmann and colleagues in 1928 and 1936 in the members of a large Austrian family with multiple affected generations and autosomal dominant inheritance. However, GSS disease was categorized as a TSE in 1981 when its transmissibility was discovered. Affected patients present with clinical manifestations of cerebellar involvement such as ataxia and dysarthria, and later dementia manifests. In some of the reported families, parkinsonism predominates, while others demonstrate cortical blindness, deafness, and gaze palsies. A codon 102 mutation of *PRNP,* which results in a substitution of a proline for leucine in the *PrP* molecule, has been linked to the ataxic form of GSS disease. Recently, Jones et al. reported a case of GSS disease associated with strongly seropositive VGKC-complex antibodies, which emphasizes the point that these antibodies should be searched for in patients with GSS disease.

Fatal Familial Insomnia. FFI is a rare disease that was initially reported by Medori and colleagues in 1992. FFI is the third most common genetic prion disease with a worldwide occurrence and with analogous clinical manifestations also in different genetic settings. This dominantly inherited prion disease manifests with sleep–wake cycle disturbances and insomnia followed by hallucinations, and eventually, patients plunge into a coma. Initially, most patients complain of disturbances of vigilance, such as being unable to go to sleep, attention, visuomotor function, or having uninvigorating sleep and demonstrate personality changes such as apathy. Insomnia is an early manifestation of FFI, which in certain patients may progress to almost complete inability to sleep. The insomnia of FFI is associated with major and lasting disorganization of sleep cycles. Loss of ability to sleep is more obvious and more progressive in the early disease course in subjects who are methionine homozygous at codon 129 than in those who are methionine and valine heterozygous. Some patients also develop visual fatigue with diplopia and sympathetic activation. Other autonomic disturbances, including pyrexia, hyperhidrosis, tachycardia, hypertension, and cardiac arrhythmia, develop. Major motor abnormalities consist of ataxia, spontaneous and evoked myoclonus, and pyramidal signs. In contrast to sCJD, where dementia rapidly develops, in patients with FFI, dementia may or may not develop. Other abnormalities include loss of circadian rhythm for secretion of melatonin, prolactin, and growth hormone, as well as a decrease in the secretion of adrenocorticotropin and an increase in the secretion of cortisol.

FFI rarely affects the thalamus without the involvement of other regions of the forebrain, and the underlying neuropathology extends throughout much of the thalamus and cerebral cortex. Neuropsychological assessment of the original FFI families demonstrated increasing disturbances of attention and vigilance and impairment of working memory and temporal ordering of events. Patients’ intelligence was preserved at least until vigilance levels permitted meaningful neuropsychology testing; however, frontal lobe functions were impaired.

A 67-year-old right-handed man developed depression, visual agnosia, and visual hallucinations. In 3 months, his neurologic syndrome deteriorated, and he developed progressive dementia, myoclonus, and ataxia. CT scan of the brain showed only brain atrophy, and EEG did not reveal any significant abnormalities. Within the next 2 months, his dementia progressed, and he became bedbound and eventually became unresponsive. Examination of CSF demonstrated the presence of protein 14–3–3, and within 7 months from the onset of his first neurologic symptoms, he passed away. Neuropathologic examination of the brain during autopsy confirmed sCJD.

A previously healthy 29-year-old right-handed female developed brief episodes of forgetfulness and lapse of attention. She was initially diagnosed with major depression and was treated with antidepressants. Her neurologic condition deteriorated with progressive further loss of communication and social skills. Within the next 4 months, she became severely ataxic and globally demented. The rapidly progressive dementia led to akinetic mute condition followed by coma and death. EEG revealed rhythmic triphasic sharp wave activity more obvious on the right frontal and temporal areas on a slow background. MRI of the brain on axial T_2-weighted images revealed bilateral hyperintense signal in caudate nuclei and putamen (Fig. 2.10–4). During postmortem examination, typical neuropathologic findings of sCJD were detected.

Psychiatric Manifestations

Psychiatric disorders commonly occur in patients with sCJD and generally manifest during the early phase of this progressive neurodegenerative disorder. In a retrospective review of 126 patients with sCJD who were assessed at the Mayo Clinic, 80% of patients suffered from psychiatric symptoms within the first 100 days of their illness, with 26% having them at presentation. The most frequently reported psychiatric symptoms were sleep disturbances, psychotic symptoms, and depression. The investigators concluded that psychiatric symptoms are common in sCJD and occur early prior to formal diagnosis.

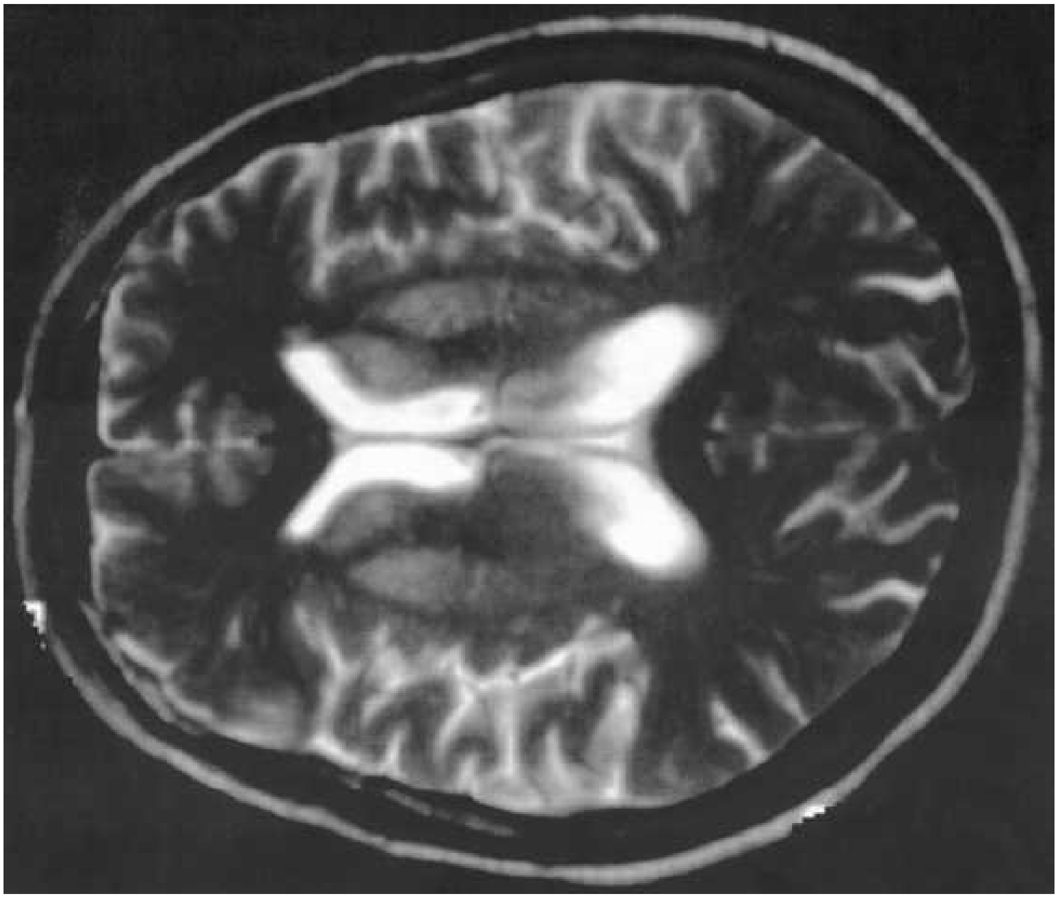

FIGURE 2.10–4. T_2-weighted magnetic resonance image of a patient with sporadic Creutzfeldt–Jakob disease. There is a hyperintense signal in the caudate nuclei and putamen. (Courtesy of Donald Collie.)

vCJD, which is less common than sCJD, affects younger individuals and has distinct clinical and neuropathologic features, with more neuropsychiatric manifestations compared to sCJD (Table 2.10–1).

In one series, all vCJD patients had psychiatric symptoms during the early phase of their disease process, and the majority were diagnosed with depression or depression secondary to organic disease. Two patients had symptoms suggestive of psychosis, while the majority manifested transient delusions and auditory or visual hallucinations. Thus, psychiatric symptoms, as consistent early components of vCJD, occur in a majority of these patients. However, they do not manifest any characteristic features that can differentiate them from other common psychiatric disorders. Only the occurrence of associated persistent sensory symptoms may raise the possibility of vCJD. Nonetheless, the evaluation of any patient with subtle neurobehavioral signs and symptoms should include the systematic approach described below.

NEUROPSYCHIATRIC WORKUP: THE STRUCTURED PSYCHOLOGICAL AND NEUROPSYCHOLOGICAL HISTORY

A comprehensive cognitive history is essential for initiating the assessment and directing which tests should be used in the evaluation battery. In combination with the clinical pattern of evolving dementia, the history can be key in the differential diagnosis of the etiologies of behavioral and cognitive dysfunction in patients with CJD with psychiatric complaints. Sometimes, the interview identifies the basis of the problem when neuropsychological testing is not possible or available. Table 2.10–4 lists the essentials of the interview for formulating this history. Answers to these questions can help associate patient complaints with cognitive disorders and dementia.

The Mental Status Examination

This introductory cognitive history taking should include an appropriate mental status examination. A standard examination, such as the Mini-Mental State Examination (MMSE), is frequently done because it is well known, and many practitioners have developed a sense of what its score means in terms of overall cognitive functioning. This test was developed to help identify the cortical type of dementia associated with AD. It has a concentration of items associated with language and orientation and has a visuospatial task. As such, it may miss the types of memory, speed of information processing, and attention/concentration problems often associated with early manifestations of CJD. Another screening task, the High Sensitivity Cognitive Screen (HSCS), may be useful in screening for the simple presence or absence of cognitive dysfunction. The creators of this measure report high correlations between this test and the overall result of neuropsychological testing. They note that the HSCS is also correlated with EEG results in medical psychiatric inpatients and with functional status. Results of this measure may then establish the eligibility of the patient for a more in-depth neuropsychological assessment.

Table 2.10–4.
Cognitive History

Name

Age and birthday
Handedness
First language at home
Educational background
Best subjects, grades
Worst subjects
Occupational background
How long
Medical history
Childhood diseases or injuries
Head injuries with loss of consciousness
Strokes
High fevers
Toxin exposure
Major illness, injuries, or surgeries
Medicines: prescription, nonprescription
Duration of changes in behavior or functioning
Current problem: change in thinking functions: how long, or over what period of time
Any change in ability to concentrate
Any periods of confusion or mental “fuzziness”
When talking with people, or on the phone, watching TV or a movie, reading
Any problem with following the train of thought
Any difficulties with handwriting
Any word-finding problems; difficulties with slurring or stammering
Any slowing of thinking or understanding, trouble with mental arithmetic such as making change or balancing checkbook
Wear glasses
Any blurring vision, double vision, or flashing lights in eyes
Any change in understanding what is seen; do things look right in their relation to each other
Overlook things when right in front of you

Hearing any unusual sounds; see unusual things; have any strange feelings
Any changes in any other senses
Decreased hearing, ringing, or buzzing sounds
Change in smell or taste
Any numbness, “pins or needles,” loss of feeling, tingling, or burning feelings
Any severe pain
Memory
Any areas of memory that are better or worse
Memory for recent information
Information from way back in life
Any difference in memory for *situations* vs. rote facts and figures
Kinds of things most easily forgotten: names, addresses, directions, reading
How long can things be remembered, more notes written than used to
Any lapses noted
Any getting lost or forgetting where one is
Any new difficulties with thinking through problems or solving them, decisions making, staying organized—on job, at home
How is sleep: any trouble getting to sleep; night vs. daytime; any awakenings from which one cannot immediately return to sleep
Any inability to move any parts of the body
Muscle weakness, twitching, spasms, trouble walking, coordination problems, tremors or shakiness, problems with dropping things, feeling like moving more slowly, difficulty using tools or household utensils, getting dressed, telling right from left
Headaches or dizziness, instances thought to be seizures (staring off into space for a long time, uncontrollable movements, periods where one seemed “lose” time, incontinence)
Changes in mood, feelings, ideas
Mood swings, loss of patience or change in temper, increase in irritability, change in amount of worry, sense of panic

Neuropsychological Assessment

The most prevalent component of cognitive impairment related to CJD includes general cognitive decline. Early problems with abstraction, attention, concentration, learning and memory, and psychomotor speed progress to more serious difficulties with these functions, as well as impaired executive functions, nonverbal problem solving, and visuospatial integration and construction.

As with the language-heavy MMSE, neuropsychological tests for the assessment of other forms of dementia include tasks that gauge such cortical functions as complex language-associated functions (such as aphasia and apraxia), higher-level cognitive functions of verbal and nonverbal abstract reasoning, and problem solving and perceptual functioning. These tasks are still necessary when one suspects or has information that the patient is experiencing dysfunction related to focal disturbances in the CNS. These can be caused by varying conditions such as infection, tumor, stroke, vasculitis, multiple sclerosis (MS), HIV, and other etiologies. Other tasks often used in comprehensive batteries for the evaluation of memory, attention, concentration, and psychomotor speed are more useful for detecting the often-subtle impairments of the early stages of cognitive impairment as may occur in vCJD. Neuropsychologists have often tried to summarize or digest multiple scores into indices that correspond to diagnostic levels (e.g., the Halstead–Reitan Impairment Index). Regarding one of the most generally used instruments, the Rey 15-Item Test, several investigators have shown that genuine conditions may yield false-positive results, and clinical judgment must ultimately determine the validity of this task's performance. Another method, the forced two-choice selection task, in which correct selections below the 50% level (actually, from the research, a much higher cutoff score; 69) may imply an attempt to deceptively manipulate the task, appears to be less problematic. These tasks have not been validated in the CJD population, and unless the patient is obviously severely demented or delirious, there can be high confidence that results reflect the patient's actual effort.

Psychological Functioning

Psychiatric conditions have received much attention as they often arise in vCJD neurobehavioral disorders. The initial manifestations of emotional lability, anxiety, depression, aggression, and psychosis, along with the emotional distress caused by a change in one's behavior and all the influences it has on a person, and possible pre-existing psychiatric disorders and/or neurologic syndromes caused by other medical circumstances may necessitate a psychiatric consultation. Several screenings and more comprehensive measures have been used but are not validated in CJD.

Although not formally researched in CJD, screening instruments such as the Beck Depression Inventory and the Symptom Checklist-90 (or its brief form, the Brief Symptom Inventory; Derogatis, 1993) have appeared frequently in the literature with medically ill patients and are simple to administer and score. These instruments and the more comprehensive Minnesota Multiphasic Personality Inventory 2 (MMPI-2) have been used to assess psychological difficulties in medically ill patients. Common problems and points of debate are the overlap of somatic or vegetative symptoms (e.g., fatigue) of depression and the systemic effects of any disease and/or its treatments. A newer, multifactor screening instrument, the Hospital Anxiety Depression Scale (HADS), attempts to avoid this problem by focusing on the psychological distress elements. For psychological evaluations beyond the screening level, a structured interview, as well as administration of the MMPI-2 as a secondary measure, would provide the information required for a full psychological diagnosis.

Clinical Assessment

Psychiatrists will best diagnose CJD by recognizing its characteristic symptoms. The approach described above will assist only in excluding a few other illnesses that mimic or overlap with CJD. Clinically, one would likely make the diagnosis of a mild neurocognitive disorder or fully developed dementia.

Unfortunately, there are several issues with the category for mild neurocognitive disorder in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM). For example, the basis for the diagnosis is mostly focused on the presence of neuropsychological impairment instead of clinical findings. Also, there is no listing of motor signs or symptoms and thus is not very specific for prion, progressive supranuclear palsy, Parkinson’s disease (PD), HIV, or other etiologies.

Although early diagnosis may be made by clinical psychiatrists, it is the progressive nature of the neurologic dysfunction that leads to neurologic consultation and a definitive assessment and diagnosis made. In one survey in the United Kingdom, only 30% of psychiatrists were aware of the diagnostic criteria and the surveillance project for CJD. Although in all the cases of vCJD, psychiatric symptomatology is the rule, it is nearly impossible to make the diagnosis in this early phase of the illness as there is no specific psychiatric phenotype to warn clinicians. Thus, the development and progression of neurologic aspects of the disease and either molecular or neuropathologic examination are necessary to properly identify all CJD cases.

Neurodiagnostic Workup

Laboratory diagnosis of CJD rests on abnormalities found in one or more of these tests: examination of CSF, EEG, or brain MRI. Table 2.10–5 provides a summary of the initial diagnostic workup for CJD.

Routine examination of the CSF is usually within normal limits, although occasionally, slightly elevated protein levels or minimal pleocytosis (less than 10 cells/mL) may be detected. However, the CSF protein level is generally less than 100 mg/dL. A class of 14–3–3 proteinase inhibitor proteins is currently being measured within CSF as surrogate markers for neuronal injury that occurs during CJD. Even though elevated CSF proteins 14–3–3 lend support to a diagnosis of CJD, the elevated levels may occur in other pathologic conditions such as recent stroke, subarachnoid hemorrhage, hypoxic brain damage, glioblastoma, postictal, inflammation, corticobasal degeneration, or paraneoplastic syndromes. However, where CJD is suspected, the detection of protein 14–4–3 improves the accuracy and confidence of diagnosis. Another method for PrP^{Sc} assay in the CSF has been developed, which is known as “protein misfolding cyclic amplification” (PMCA) and is being utilized as a diagnostic test. In certain aspects, PMCA is like a polymerase chain reaction and works as an amplification procedure to multiply misfolded prions. It is presently used for the diagnosis of prion diseases. A modified version of this technique has been developed as real-time quaking-induced conversion (RT-QuIC), which is capable of separating patients with sCJD from AD with 80% sensitivity and no false-positive cases. CSF from patients with sporadic as well as inherited CJD, GSS disease, and FFI have been examined utilizing RT-QuIC, and the results demonstrate that RT-QuIC is as sensitive as a measurement of protein 14–3–3 but more specific than the latter assay.

EEG of patients with CJD characteristically demonstrates the presence of bi- and triphasic PSWC in the context of slow background. These classic EEG abnormalities are present in up to 70% of patients with sCJD but may take up to 3 months to appear. Other diseases that may imitate EEG abnormalities observed in CJD include AD, HIV-associated neurocognitive disorders, Lewy body disease, MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes), anoxic/ischemic encephalopathy, hypoglycemia, hypo- and hypernatremia, hepatic encephalopathy, hyperammonemia, and lithium toxicity.

CT scan of the brain remains normal in most patients; however, in some, it may reveal cerebral atrophy with enlargement of ventricles and cisterns or cerebellar atrophy. Brain MRI usually demonstrates distinguishing hyperintense signals in the caudate and putamen with less involvement of the cortex or thalamus. These abnormal hyperintense signals can be observed on T_2-weighted proton density and FLAIR sequences.

Compared to sCJD, studies of CSF and EEG are less useful in the diagnosis of vCJD. CSF 14–3–3 proteins are less sensitive in vCJD and are positive in only half of the patients, and EEG traces do not demonstrate the typical PSWC that is seen in sCJD.

MRI of the brain in patients with sCJD demonstrates the presence of hyperintense signals in the caudate and lenticular nuclei or at least two cortical areas (temporal, parietal, or occipital in diffusion-weighted and/or FLAIR images) (Fig. 2.10–5). These abnormal areas on brain MRI correlated with the areas with the most spongiform vacuolation and neuronal loss than gliosis. MRI of the brain in patients with vCJD shows bilateral hyperintense signals in the pulvinar of the thalamus (the pulvinar sign) on T_2-weighted and proton-density images (Fig. 2.10–5). Pulvinar hyperintense signals correlate with gliosis. In patients with FFI routing, neuroimaging does not yield any characteristic features. Brain imaging with fluorodeoxyglucose positron emission tomography ([^{18}F]-FDG PET) demonstrates remarkable thalamic hypometabolism, which is sometimes associated with diffuse hypometabolism of the basal ganglia, cerebral cortex, particularly the frontotemporal cingulate regions, and the cerebellum. Bilateral thalamic with less pronounced cingulate cortex hypometabolism is a typical aspect of FFI. Polysomnographic studies of these patients have shown the absence of sleep spindles and δ-sleep, along with prominent autonomic and motor hyperactivity (also known as agrypnia excitata). In FFI patients, stage 1 nonrapid eye movement (NREM) sleep is substantially preserved. In patients with FFI, an increase in 5-hydroxytryptamine synthesizing neurons within the median raphe system has been reported, and examination of CSF of patients with FFI has revealed a four- to fivefold increase in the 5-hydroxyindoleacetic acid (5-HIAA) catabolites compared to normal controls.

Table 2.10–5.
Initial Diagnostic Tests for Creutzfeldt–Jakob Disease

Blood tests: Complete blood count, comprehensive metabolic panel, serum ammonia, RPR, sedimentation rate, C reactive protein, serum vitamin B_{12}, HIV test, Lyme test, PT/INR

Urine tests: Urine analysis, urine drug screen

Cerebrospinal fluid analysis: Cell count and cytology examination for malignant cells, presence of oligoclonal bands, IgG index, protein and glucose level, VDRL

Neuroimaging: MRI of brain with and without contrast including

FLAIR and DWI images

EEG

CT scan of body with and without contrast searching for malignancy

Brain biopsy

CT, computed tomography; DWI, diffusion-weighted imaging; EEG, electroencephalogram; FLAIR, fluid-attenuated inversion recovery; HIV, human immunodeficiency virus; IgG, immunoglobulin G; MRI, magnetic resonance imaging; PT/INR, prothrombin time/international normalized ratio; RPR, rapid plasma reagin; VDRL, Venereal Disease Research Laboratory test.

Differential Diagnosis

Kuru does not pose a diagnostic challenge for clinicians since there are not too many other disorders that can cause relentlessly progressive cerebellar disease in Fore tribe members of New Guinea. Similarly, FFI is not a major diagnostic dilemma once it occurs in a member of an FFI pedigree. However, in certain cases, positive family history is absent, or insomnia develops only with disease worsening. In these cases, other prion-disease imitators such as familial AD should be ruled out. However, this may not be the case for other prion diseases, and in certain cases, comprehensive diagnostic procedures should be performed.

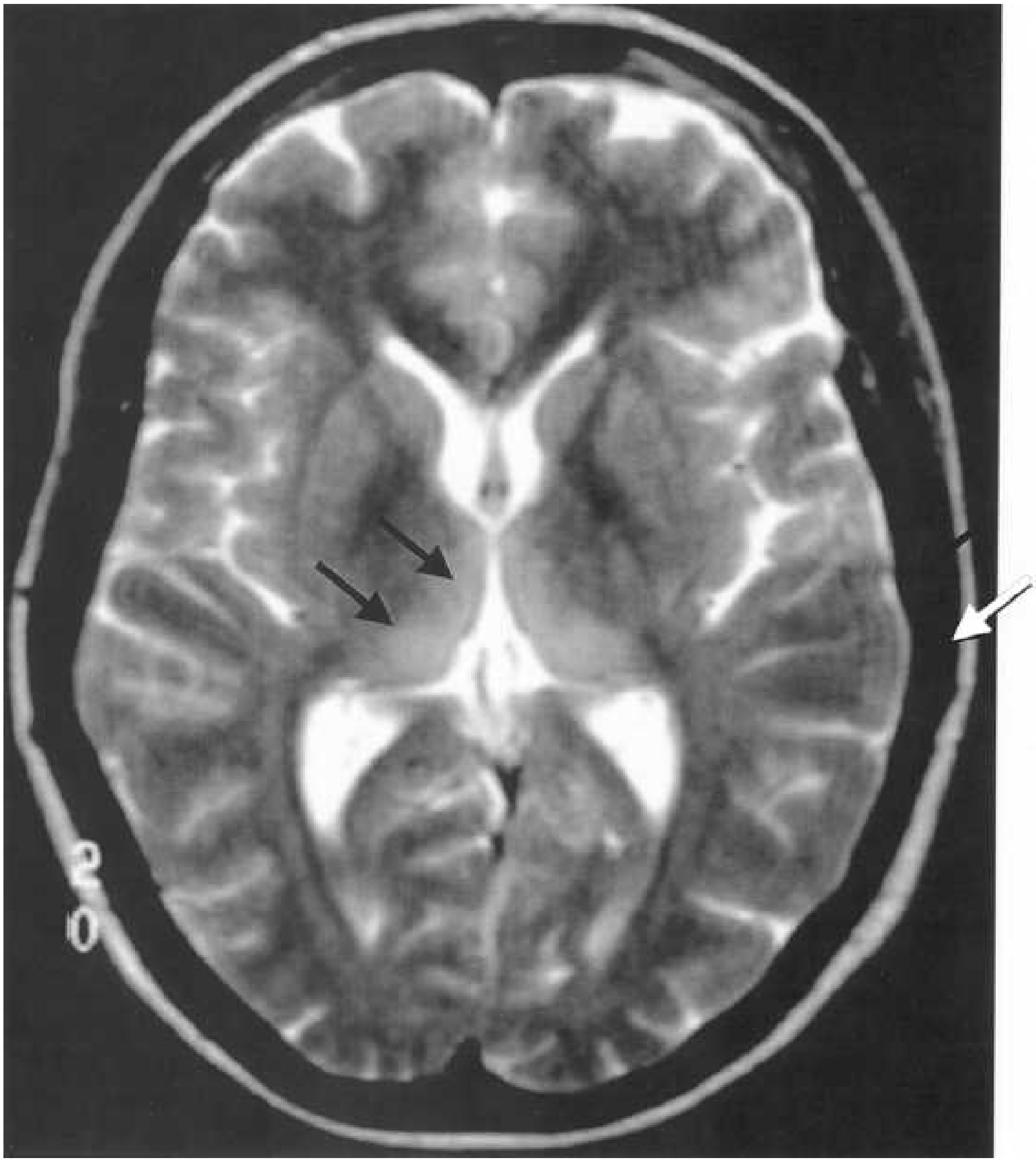

FIGURE 2.10–5. T_2-weighted magnetic resonance image from a patient with variant Creutzfeldt–Jakob disease. There is a hyperintensity (*black arrows*) of the pulvinar nuclei extending into the more anterior thalamus, leading to the “hockey-stick” sign of variant Creutzfeldt–Jakob disease. (Courtesy of Donald Collie.)

Table 2.10–6.
Differential Diagnosis of Creutzfeldt–Jakob Disease

Toxic/metabolic disorders
- Endocrine disorders (thyroid, parathyroid, adrenal gland)
- Electrolyte imbalance (sodium, calcium, phosphate, magnesium)
- Deficiency of vitamin B_{12}, B_1, E, niacin, folate
- Uremic encephalopathy
- Wilson disease
- Porphyria
- Neuroacanthocytosis
- Metal intoxication (lithium, manganese, mercury)

Infections
- HIV-associated dementia
- Viral encephalitis
- Lyme disease
- Progressive multifocal leukoencephalopathy
- Fungal CNS infections
- Whipple disease
- Subacute sclerosing panencephalitis

Neurodegenerative disorders
- Alzheimer disease
- Huntington disease
- Pick disease
- Lewy body disorders
- Amyotrophic lateral sclerosis
- Progressive subcortical gliosis

Immune-mediated disorders
- Multiple sclerosis
- Hashimoto encephalopathy
- Systemic lupus erythematosus cerebritis
- Anti-GAD syndrome
- CNS vasculitis (primary or secondary)
- Sprue
- Acute disseminated encephalomyelitis
- Anti-VGKC syndrome

Neoplasms
- Metastatic encephalopathy
- Primary CNS lymphoma

	Glioblastoma cerebri
	Intravascular lymphoma
	Paraneoplastic syndromes
	Metastatic CNS tumors
Vascular neurologic disorders	
	Multi-infarct dementia
	Cerebral amyloid angiopathy
	Thalamic or corpus callosum infarcts

HIV, human immunodeficiency virus; CNS, central nervous system.

Differential diagnosis of sCJD includes subacute encephalopathies secondary to intoxication with certain drug toxins such as lithium, mercury, or bismuth, and other diseases including CNS vasculitis, Hashimoto thyroiditis, paraneoplastic syndromes, infectious and granulomatous disorders such as neurosyphilis, sarcoidosis, HIV-1–associated neurocognitive disorders, CNS fungal infections, other neurodegenerative disorders such as AD, Lewy body dementia, PD with dementia, corticobasal degeneration, amyotrophic lateral sclerosis, and frontotemporal dementia. An extensive list of differential diagnoses of CJD is presented in Table 2.10–6.

COURSE AND PROGNOSIS

Despite dissimilar ages of onset, clinical manifestations, and durations of survival, CJD is a relentlessly progressive disease with an almost daily compromise of neurologic and psychiatric function that culminates into death. The mean duration of CJD is approximately 7 months, and a disease duration of more than 2 years is rare. Similar to other prion diseases, vCJD is invariably fatal. Patients with FFI may experience worsening of their dysautonomia syndrome and eventually succumb to their fatal illness. To date, no information exists on the effect(s) of pregnancy on prion diseases.

TREATMENT

To date, there is no known treatment for prion diseases, although investigational pharmacotherapies are emerging. Palliative therapy is the cornerstone of treatment for these patients. Despite the differences in their age of onset, clinical manifestations, and disease course and duration, prion diseases remain invariably fatal. Symptomatic therapy includes the use of opioids and antiepileptic agents for pain management, selective serotonin reuptake inhibitors (SSRIs) for management of depression and agitation, atypical antipsychotics for the treatment of agitation and psychosis, and clonazepam to manage myoclonus and agitation. Referring patients and families to the CJD Foundation for caregiving planning and end-of-life care should be a priority from the onset of prion disease symptoms.

The basis of experimental therapeutics is the assumption that disease processes arise from the transformation of PrP^C to PrP^{Sc} and its subsequent accumulation. Direct inhibition of this conversion is an important therapeutic strategy. Targeting PrP^C may make it possible to knockdown or totally remove substrates for prion propagation. In theory, this strategy could be applied to all forms of prion diseases. On the other hand, targeting PrP^{Sc} itself could have no impact on disease progression or may even enhance or prolong disease. Thus, current investigational therapeutic strategies that may assist scientists in developing effective treatments of human prion disease target the conversion of PrP^C to PrP^{Sc} and include disruption of PrP^C plasma membrane trafficking, stabilization of PrP^C structure with chemical chaperones, manipulation of PrP^C-related signaling pathways, and disruption of the PrP^C–PrP^{Sc} physical interaction.

At present, families of identified patients with CJD may inquire or request genetic testing, especially in the context of familial prion disease. Since there is no available treatment for CJD, it is debatable whether such testing is warranted. Although there are no specific treatment guidelines for genetic counseling in such situations, one can adapt recommendations that exist (see https://cjdfoundation.org/genetic-counseling). If you wish to consider genetic counseling, one can find an expert genetic counselor and consult privately in person or via telehealth through the National Society of Genetic Counselors (see https://findageneticcounselor.nsgc.org/?reload = timezone).

As with any other type of genetic testing and counseling, the risk–benefit ratio of testing should be carefully reviewed with the family and the medical team. If one proceeds with such testing, it should be done under the strictest confidentiality and privacy rules unless otherwise stipulated, as with the U.S. Food and Drug Administration (FDA) ruling on blood donations and genetic analysis (https://www.fda.gov/regulatory-information/search-fda-guidance-documents/recommendations-reduce-possible-risk-transmission-creutzfeldt-jakob-disease-and-variant-creutzfeldt). Consistent with these developments, the following recommendations are provided for the clinician:

1. Testing conditions should be voluntary and follow local statutes where available. Pre- and posttest counseling should not be combined in the same session.
2. Either genetic counselors or psychiatrists should be competent in counseling patients about seeking genetic testing and should be well informed about the need for pre- and posttest counseling, provide information about the limits of confidentiality, should be aware of CJD surveillance procedures in their local jurisdictions, and should be aware of resources available to patients and their families.
3. A complete behavioral health assessment should be considered in all cases in order to fully address the individual's reaction to the results. This should include an assessment that the patient is indeed ready to be tested. Patients with high risks of untoward psychological reactions or destructive behaviors should be a special concern to anyone screening for CJD, and a referral to psychiatry should be initiated.
4. Genetic counseling involving pre- and posttest counseling of patients should include an assessment of capacity and discussion of the risks and benefits of testing, previous experience with genetic testing, the implications of a positive or negative result, the limits of confidentiality, strategies for reducing anxiety and depression, and help with the availability of and referral to appropriate resources for further counseling and assistance.
5. Genetic testing should not be performed solely for the purpose of medical team and staff awareness.
6. All CJD testing must be done with appropriate informed consent. It is not sufficient simply to have a consent form signed; it must also be documented that the person is informed and understands the consequences of both a positive and negative result.
7. The confidentiality of information regarding genetic testing for CJD should be protected.
8. Sharing of CJD status should be compliant with applicable state and federal statutes.
9. Patients should be made aware of program policies regarding documentation of CJD status in the medical record before initiating

genetic testing.

10. Patients should be aware that when third parties pay for genetic testing, both positive and negative test results may be available to the Medical Information Bureau and can subsequently affect eligibility for future insurance.

Living in the era of molecular medicine forces all health professionals to confront difficult legal and ethical issues. Problem solving around issues of genetic testing requires an educated balancing of diverse interests and a thoughtful approach that rationally weighs the benefits and disadvantages of standard as well as new solutions. This field is rapidly changing, and as such, there are no easy answers in terms of whether asymptomatic family members should be tested for an untreatable disease.

PREVENTIVE MEASURES

Since prion diseases remain incurable, universal precautionary measures need to be applied during the care and management of affected patients. Even though the infectious proteins causing CJD can be found in many organs, it appears less frequently in certain body fluids such as tears, saliva, sweat, urine, or feces. However, this infective prion can be detected in CSF and rarely in blood. In terms of infectivity, human tissues are divided into three groups: high-infectivity tissues: CNS tissues that attain a high titer of infectivity in the later stages of all TSEs, and certain tissues that are anatomically associated with the CNS; lower-infectivity tissues: peripheral tissues that have been tested positive for infectivity and or PrP^{TSE} in at least one form of TSE; and tissues with no detectable infectivity: tissues that have been examined for infectivity and/or PrP^{TSE} with negative results.

Health workers should wear gloves while handling these patients' biologic specimens and should avoid any penetrating injuries from contaminated sources that can potentially transmit the infected material. Any accidental exposure of intact skin to these infective proteins must be irrigated with fresh undiluted bleach or 1 N sodium hydroxide, followed by washing and cleansing with copious amounts of water and soap. For medical and surgical procedures, disposable materials and instruments should preferably be used, and once they have been used, they must be destroyed by incineration. In case certain instruments or devices cannot be discarded or destroyed, they should be sterilized based on proposed WHO decontamination guidelines. Autopsies and cremation of patients affected with prion diseases must be carried out based on the guidelines published by the Centers for Disease Control (CDC) (https://www.cdc.gov/prions/cjd/funeral-directors.html).

Universal Precautions Regarding CJD Risk

It should also be noted that frequently the patient is unaware of having CJD or other infectious diseases when coming to the doctor's office or emergency room. Under these circumstances, universal precautions should be used. The psychiatry staff inpatient unit, trainees, and attending physicians are responsible for laboratory procedures that are part of the medical workup of the patients admitted to the unit, including the psychiatry emergency room. Moreover, this includes nursing personnel who must provide parenteral injections to patients without knowing their CJD, HIV, or hepatitis C status. Under these conditions, universal precautions must be followed according to the CDC (https://www.cdc.gov/prions/cjd/infection-control.html) as well as the *American Psychiatric Association Compendium of Practice Guidelines* (www.psych.org/AIDS/), the U.S. FDA (http://www.fda.gov/cber/gdlns/cjdvcjd.htm), and the WHO Infection Control Guidelines for TSE (www.who.inf.csr/resources).

Biohazards: Prion Transmissibility and Infectiousness

Overall, the exercise of universal precautions related to biohazards is essential for handling and processing all specimens of human tissues and fluids. The examination of this topic in PubMed indicates that this is discussed in the literature due to ubiquitous biohazard risks. The dangers, risks, remedies, and countermeasures in the laboratory and clinical settings have been well developed through the use and institution of (biosafety level) BSL-3 and BSL-4 laboratories. This has become most apparent during the past Ebola virus outbreaks.

There are several corrective laboratory measures regarding human prion diseases. Thirty-five years ago, it was already recognized that prions derived from spongiform encephalopathies including Kuru, CJD, TME, and scrapie posed biohazard risks for humans. Based upon that, laboratory design proceeded.

Tissues and fluids infected or even potentially infected (and that includes all human specimens) are treated and handled with care and universal precautions. These precautions are well known for tissues and fluids at risk of HIV and HCV infection. Universal precautions emphasize the risk of infectivity due to unanticipated biologic agents. However, risks of infection by prions are an additional risk that is associated with any human specimens—derived from symptomatic or asymptomatic individuals. Similarly, animal specimens are treated with similar care. The goal is a reduction in the potential exposure risk. In the United States, several accrediting agencies deal with such risks and including the NIH, CDC, and Occupational Safety and Health Administration (OSHA) (see https://www.cdc.gov/prions/cjd/infection-control.html).

The central concept in the infectiousness of PrPs is that they are in misfolded conformations that are self-replicative and can infect, seed, and propagate their strain characteristics in dissimilar host species as well as homologous host species. The seeding infecting PrP converts much of the host's PrP^C into infectious PrP proteinase K–resistant particles. This process is also termed PMCA.

Current investigations also implicate a process of "seeding" misfolded α-synuclein protein involvement in some aspects of PD. This possibility is related to the prion disease CJD. In fact, it has further been proposed that there is a component of prison-related pathogenesis (gain of toxic function) in AD as well as in PD. It has been suggested that these diseases should be included with TSEs or prion diseases with the well-known prion diseases, including CJD and its variant, Kuru, as well as with FFI and GSS disease. In addition, amyotrophic leukospongiosis pathogenesis may involve 18 to 20 kD prion-like globular structures.

As a postscript, it may be noted that the scientific attitude to potential infection by microbial life forms or nonviable particles, for that matter, has wide-reaching implications. For example, there was, apparently, an absence of any sensible biocontainment protocols during the many years of the Apollo Space Program missions. There was also an absence of a functional particle containment protocol even though the lunar surface was composed of dust. There were findings of extensive human and mechanical contamination by lunar dust that have been well documented. Thus, it is not always obvious to the nonbiologically oriented scientist what types of exposure there may be, biologic or mechanical. This is of relevance because of the unanticipated situations, including prion diseases in which higher levels of biocontainment or mechanical containment are necessary, though originally not perceived.

FURTHER READINGS

Aguzzi A. Prion diseases of humans and farm animals: epidemiology, genetics, and pathogenesis. *J Neurochem*. 2006;97:1726–1793.

Atarashi R, Satoh K, Sano K, et al. Ultrasensitive human prion detection in cerebrospinal fluid by real-time quaking-induced conversion. *Nat Med*. 2011;17(2):175–178.

Ayuso Blanco T, Urriza Mena J, Caballero Martínez C, Iriarte Franco J, Munoz R, García-Bragado F. [Fatal familiar insomnia: clinical, neurophysiological and histopathological study of two cases]. *Neurologia*. 2006;21(8):414–420.

Collins SJ, Lawson VA, Masters CL. Transmissible spongiform encephalopathies. *Lancet*. 2004;363(9402):51–61.

Cordery RJ, Alner K, Cipolotti L, et al. The neuropsychology of variant CJD: a comparative study with inherited and sporadic forms of prion disease. *J Neurol Neurosurg Psychiatry*. 2005;76:330–336.

Daus ML, Wagenführ K, Thomzig A, et al. Infrared microspectroscopy detects protein misfolding cyclic amplification (PMCA)-induced conformational alterations in hamster scrapie progeny seeds. *J Biol Chem*. 2013;288(49):35068–35080.

Derogatis LR. Brief Symptom Inventory. Pearson Assessments; 1993.

Chen C, Dong X. Therapeutic implications of prion diseases. *Biosafety and Health*. 2021;3(2):92–100.

Grizzle WE, Fredenburgh J. Avoiding biohazards in medical, veterinary and research laboratories. *Biotech Histochem*. 2001;76(4):183–206.

Glatzel M, Stoeck K, Seeger H, Lührs T, Aguzzi A. Human prion diseases: molecular and clinical aspects. *JAMA Neurol*. 2005;62(4):545.

Hall DA, Leehey MA, Filley CM, et al. PRNP H187R mutation associated with neuropsychiatric disorders in childhood and dementia. *Neurology*. 2005;64:1304–1306.

Hannaoui S, Schatzl HM, Gilch S. Chronic wasting disease: emerging prions and their potential risk. *PLoS Pathog*. 2017;13(11):e1006619.

Hilker R, Brotchie JM, Chapman J. Pros and cons of a prion-like pathogenesis in Parkinson's disease. *BMC Neurol*. 2011;11:74.

Jones M, Odunsi S, du Plessis D, et al. Gerstmann-Straüssler-Scheinker disease: novel PRNP mutation and VGKC-complex antibodies. *Neurology*. 2014;82(23):2107–2111.

Khan Z, Bollu PC. Fatal Familial Insomnia. [Updated 2020 May 27]. In: *StatPearls [Internet]*. StatPearls Publishing. January 2021. https://www.ncbi.nlm.nih.gov/books/NBK482208/

Korth C, Peters PJ. Emerging pharmacotherapies for Creutzfeldt-Jakob disease. *Arch Neurol*. 2006;63(4):497–501.

Lawson VA, Collins SJ, Masters CL, Hill AF. Prion protein glycosylation. *J Neurochem*. 2005;93(4):793–801.

Ligios C, Sigurdson CJ, Santucciu C, et al. PrPSc in mammary glands of sheep affected by scrapie and mastitis. *Nat Med*. 2005;11(11):1137–1138.

McGuire LI, Peden AH, Orrú CD, et al. Real time quaking-induced conversion analysis of cerebrospinal fluid in sporadic Creutzfeldt-Jakob disease. *Ann Neurol*. 2012;72(2):278–285.

Nazor Friberg K, Hung G, Wancewicz E, et al. Intracerebral infusion of antisense oligonucleotides into prion-infected mice. *Mol Ther Nucleic Acids*. 2012;1:e9.

Panegyres PK, Armari E. Therapies for human prion diseases. *Am J Neurodegener Dis*. 2013;2(3):176–186.

Poleshchuk NN, Kapitulets SP, Kapitulets NN, Kvacheva EB, Eremin VF, Votiakov VP. [The infectiousness of 18- to 20-kd proteins isolated from the brain of people who have died from amyotrophic leukospongiosis]. *Bull Eksp Biol Med*. 1993;116(10):409–412.

Ross ED, Minton A, Wickner RB. Prion domains: sequences, structures and interactions. *Nat Cell Biol*. 2005;7(11):1039–1044.

Sano K, Satoh K, Atarashi R, et al. Early detection of abnormal prion protein in genetic human prion diseases now possible using real-time QUIC assay. *PLoS One*. 2013;8(1):e54915.

Sigurdson CJ, Bartz JC, Glatzel M. Cellular and molecular mechanisms of prion disease. *Annu Rev Pathol*. 2019;14:497–516.

Toupet K, Compan V, Crozet C, et al. Effective gene therapy in a mouse model of prion diseases. *PLoS ONE*. 2008;3(7):e2773.

U.S. Food and Drug Administration. Guidance for Industry: Recommendations to Reduce the Possible Risk of Transmission of Creutzfeldt-Jakob Disease and Variant Creutzfeldt-Jakob Disease by Blood and Blood Components. 2022. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/recommendations-reduce-possible-risk-transmission-creutzfeldt-jakob-disease-and-variant-creutzfeldt

Vandevelde M, Zurbriggen A, Fatzer R. Spongiform encephalopathies with special reference to bovine spongiform encephalopathy. *Schweiz Med Wochenschr*. 1992;122(23):887–892.

Wadsworth JDF, Joiner S, Linehan JM, et al. Phenotypic heterogeneity in inherited prion disease (P102L) is associated with differential propagation of protease-resistant wild-type and mutant prion protein. *Brain*. 2006;129(Pt 6):1557–1569.

Walter ED, Chattopadhyay M, Millhauser GL. The affinity of copper binding to the prion protein octa-repeat domain: evidence for negative cooperativity. *Biochemistry*. 2006;45(43):13083–13092.

Will RG, Ironside JW. Sporadic and Infectious Human Prion Diseases. *Cold Spring Harb Perspect Med*. 2017;7(1):a024364.

World Health Organization. Human transmissible spongiform encephalopathies. *Wkly Epidemiol Rec*. 1988;73:361

▲ 2.11 Neuropsychiatric Aspects of Headache

KATHLEEN RIES MERIKANGAS, PH.D., TARANNUM LATEEF, M.D., M.P.H., AND JAMES R. MERIKANGAS, M.D.

During the past decade, there has been growing attention to the enormous public health impact of migraine. Headache disorders in 2019 ranked 14th among global disability-adjusted life years (DALYs) and were the third greatest source of years lost to disability (YLDs), particularly among young females. However, many of those troubled by headaches do not receive effective care. For example, in the United States of America and the United Kingdom, only half of those identified with migraine had seen a doctor for headache-related reasons in the previous 12 months, and only two-thirds had been correctly diagnosed. Most were solely reliant on over-the-counter self-treatment.

There have been multiple developments in the diagnosis and treatment of headaches. First, there has been a revision of the standardized diagnostic criteria for some of the primary headache subtypes in adults and youth. Second, there is growing awareness of the differential manifestations of migraine across the life span, with increasing effort to target both preventive efforts and treatment to different strata of the population. Third, there have been remarkable developments in pharmacologic treatments of headache, with the introduction of a new class of drugs that has rapidly come into widespread use. Fourth, there is a growing effort to track symptoms of headache in real time using electronic diaries in order to identify targets for prevention and intervention, Finally, there has been a burgeoning effort to consider the pervasive comorbidity of headache with both mental and physical disorders in the evaluation and treatment of migraine. Spurred by establishment of large international collaborative studies, there has also been progress in identifying the influence of genetic factors on migraine and other headaches. The goals of this chapter are to (1) provide an overview of the diagnostic nomenclature for headache syndromes; (2) describe the magnitude, risk factors, and patterns of comorbidity of migraine and other headache subtypes; (3) summarize the clinical and laboratory examination necessary for making a diagnosis of headache and the neuropsychiatric manifestations of migraine and other headache syndromes; and (4) review current acute and prophylactic treatment strategies for the major headache syndromes.

DEFINITIONS

The International Headache Society (IHS) recently released the third edition of the *International Classification of Headache Disorders* (ICHD-3) (https://ichd-3.org/) that has revised criteria for the following categories: (1) the primary headache syndromes (i.e., migraine without aura, migraine with aura, tension-type headache, and cluster headache), (2) secondary headache syndromes include eight types of headaches secondary to other acute and chronic conditions, and (3) cranial neuralgias and other causes of facial pain. The eight types of secondary causes of headache are shown in Table 2.11–1. These primarily involve brain injuries or conditions that lead to headache among other pathophysiologic manifestations. Below, we consider the three primary headache disorders that include migraine, tension-type and cluster headaches, and a brief discussion of post-traumatic headache.

Migraine

Migraine is a disorder characterized by recurrent attacks or episodes of headache accompanied by other neurologic and gastrointestinal (GI) systems. Migraine presentation is multifaceted with symptoms emanating from multiple systems, including vascular, neurologic, GI, endocrine, and visual. There is general agreement that a comprehensive neuropsychiatric evaluation is required for all patients presenting with headache complaints.

Table 2.11–1.
Secondary Causes of Headache

- **Disorders of the Cranium or Face**
  - Cluster headache
  - Cranial neuralgias
  - Neck, eyes, ears, nose, sinus, teeth, mouth
- **Head Trauma**
  - Headache within 7 days
- **Intracranial Disorders**
  - Hemorrhage (subarachnoid; subdural)
- **Infections (Noncephalic)**
  - Fever, sinus, viral
- **Metabolic Disorders**
  - Arterial hypertension
  - Cardiac cephalalgia
  - Dialysis
  - Fasting
  - Hypothyroidism
  - Hypoxia/hypercapnia (high altitude, diving, sleep apnea)
- **Substance Use or Withdrawal**
  - Caffeine, alcohol, other
- **Temporomandibular Joint Dysfunction**
  - Bruxism, joint injury, or inflammation
- **Vascular Disorders**
  - Arteriovenous malformation, stroke, vasculitis

Table 2.11–2.
ICHD-3 Criteria for Migraine without Aura

A. At least 5 attacks fulfilling criteria B-D
B. Duration between 4 and 72 hours (untreated or unsuccessfully treated)
C. At least 2 of the following:
   1. Unilateral
   2. Pulsating pain
   3. Moderate to severe intensity
   4. Aggravation by or causing avoidance of routine physical activity (walking or climbing stairs)
D. During headache at least one of the following:
   1. Nausea and/or vomiting
   2. Photophobia and phonophobia
E. Not better accounted for by another disorder

The ICHD-3 criteria for migraine with and without aura are presented in Tables 2.11–2 and 2.11–3. The core features of most definitions of migraine include a recurrent headache that is often unilateral, GI symptoms such as nausea or vomiting, and hyperesthesia manifested by photophobia or phonophobia. The headache generally has a pulsatile or throbbing quality, and the pain is exacerbated by routine physical activity involving movement of the head. The ICHD-3 criteria operationalize these features of a headache to draw common thresholds and distinctions between migraine and other types of headaches. Migraine was formerly divided into two major subtypes, common and classic, with the latter being distinguished by the presence of neurologic symptoms that precede the onset of the headache. The ICHD-3 no longer includes the common–classic distinction; instead, migraine is subtyped according to the presence or absence of aura symptoms (reversible neurologic dysfunction). Approximately 20% of migraine sufferers experience aura.

Despite recent progress in the standardization of the classification of migraine by the ICHD-3, the diagnostic criteria have not been subjected to intensive investigation with respect to reliability or validity. There are still several features unique to the headache syndromes that constitute impediments to developing a valid set of diagnostic criteria for headache syndromes that need to be addressed. These include the co-occurrence of multiple headache syndromes within individual persons; the tendency for headache characteristics to change across the life span; the effects of professional and self-treatment of headache in obscuring the manifestations of the underlying headache syndrome(s); and the lack of generalizability of treated samples from which the diagnostic criteria were derived. Specific areas of the classification system have also been identified as requiring additional clarification, notably the specification of procedures for ensuring standardized methodology for the ascertainment of the diagnostic criteria; methods for assessing and coding multiple headache syndromes within individuals; and the development of standardized methods for discriminating between primary headache syndromes and those for which the secondary etiology is known.

Table 2.11–3.
ICHD-3 Criteria for Migraine with Aura

- **A. At least 5 attacks fulfilling criteria B-D**
- **B. Aura consisting of at least one of the following but no motor weakness:**
  1. Fully reversible visual symptoms including positive features (e.g., flickering lights, spots, or lines) and/or negative features (i.e., loss of vision)
  2. Fully reversible sensory symptoms including positive features (i.e., pins and needles) and/or negative features (i.e., numbness)
  3. Fully reversible dysphasic speech disturbance
- **C. At least two of the following:**
  1. Homonymous visual symptoms and/or unilateral sensory symptoms
  2. At least one aura symptom develops gradually over 5 minutes and/or different aura symptoms occur in succession over 5 or more minutes
  3. Each symptom lasts 5 or more minutes and less than or equal to 60 minutes
- **D. Headache fulfilling criteria B-D for migraine without aura begins during the aura or follows within 60 minutes**
- **E. Not attributed to another disorder**

Tension-Type Headache

The definition of tension-type headache according to ICHD-3 criteria is presented in Table 2.11–4. Briefly, tension-type headache is characterized by episodes of stable bilateral pain lasting several days at a time. It is distinguished from migraine headaches by its generally longer duration, the lack of pulsating quality of the pain, the lack of worsening with physical activity, and the absence of GI concomitants. However, migraine and tension-type headaches may often coexist, either simultaneously or alternating over time. It is no longer believed that tension-type headache results from muscle tension. Indeed, neck pain may result from head movement to reduce headache pain.

Cluster Headache

Cluster headache is a distinct syndrome characterized by frequent intense attacks (often several per day) over a 1- to 2-month period, separated by headache-free intervals for as long as 1 or 2 years. Although it is commonly grouped with migraine, current evidence, including epidemiologic data, treatment response, and clinical features, suggests that cluster headaches may comprise a distinct syndrome.

Table 2.11–4.
ICHD-3 Criteria for Tension-Type Headache

Frequent Episodic Tension-Type Headache

- **A. At least 10 episodes occurring on greater than or equal to one but less than 15 days per month for at least 3 months (greater than or equal to 12 and less than 180 days per year) and fulfilling criteria for B-D**
- **B. Headache lasting from 30 minutes to 7 days**
- **C. Headache has at least two of the following characteristics:**
  1. Bilateral location
  2. Pressing/tightening (nonpulsating) quality
  3. Mild or moderate intensity
  4. Not aggravated by routine physical activity such as walking or climbing stairs
- **D. Both of the following:**
  1. No nausea or vomiting (anorexia may occur)
  2. No more than one of photophobia or phonophobia
- **E. Not attributed to another disorder**

Chronic Tension-Type Headache

- **A. Headache occurring on greater than or equal to 15 days per month on an average for more than 3 months (greater than or equal to 180 days per year) and fulfilling criteria B-D**
- **B. Headache lasts hours and may be continuous**
- **C. Headache has at least two of the following characteristics:**
  1. Bilateral location
  2. Pressing/tightening (nonpulsating) quality
  3. Mild or moderate intensity
  4. Not aggravated by routine physical activity such as walking or climbing stairs
- **D. Both of the following:**
  1. No more than one of photophobia, phonophobia, or mild nausea
  2. Neither moderate or severe nausea nor vomiting
- **E. Not attributed to another disorder**

Cluster refers to a “clustering in time,” with the headache bouts occurring every day to several times a day over a period of days to weeks, followed by a lengthy headache-free interval. Cluster headache is generally retroorbital in location and is accompanied by autonomic changes such as lacrimation, rhinorrhea, erythema of the eye, and agitation. Males tend to suffer more from cluster headaches than females. Patients with cluster headaches do not retire to dark rooms and lie down to avoid the stimulation but may, in fact, do quite the opposite, appearing almost manic in their agitation. The pain can be so intense that the sufferer may appear to be psychotic because of the screaming and thrashing that may be associated with the pain. Prior smoking and alcohol use has been associated with cluster headache, with alcohol often triggering the onset of the headache.

Chronic paroxysmal hemicrania is a type of cluster headache, specifically responsive to treatment with indomethacin and characterized by many daily focal attacks of pain lasting for short periods, generally about 15 or 20 minutes per attack.

Headache Attributable to Head or Neck Trauma

The key symptoms of ICHD-3 post-traumatic headache include a headache following head trauma that develops within 7 days after head trauma accompanied by a loss of consciousness, post-traumatic amnesia, and abnormal laboratory tests. Post-traumatic headache is variable in symptom presentation, severity, and duration.

Although a headache following a traumatic head injury has often been attributed to emotional factors, empirical evidence suggests that emotional factors are more likely to be a sequela rather than a cause of post-traumatic headache. Nevertheless, the pathogenesis of post-traumatic headaches is unknown. The major hypotheses include cerebral edema, cortical spreading depression, innate vulnerability to cerebral vasospasm, and transient elevation of intracranial pressure. There is no direct relationship between the prevalence or chronicity of post-traumatic headache and several indicators of the severity of the head injury, including duration of unconsciousness, post-traumatic amnesia, electroencephalographic abnormalities, presence of skull fracture, or the presence of blood in the cerebrospinal fluid. There appears to be an inverse relationship between the severity of the head injury and the development of posthead injury headache; post-traumatic headache is more common after injuries that do *not* result in a skull fracture.

The onset of typical migraine attacks following acute head trauma occurs so frequently that it has been hypothesized that head trauma serves as a trigger for migraine in persons with underlying susceptibility to migraine or with a personal or family history of migraine. Moreover, relatives of post-traumatic migraine subjects have an increased prevalence of neurologic symptoms, suggesting a propensity to neurologic manifestations of migraine.

EPIDEMIOLOGY AND COURSE

Prevalence

Headache is very common in the general population, with about 70% of adults endorsing a history of headaches in a given year. The average lifetime prevalence of migraine of about 12% in international studies has remained stable for the past several decades and is very close to the 10% to 12% rates in pre–ICHD-2 studies. By contrast, chronic migraine is quite rare, with an aggregate estimate of 0.5%. Recent reviews of the prevalence of tension-type headaches estimate the aggregate lifetime prevalence of tension-type headaches of about 30%. These rates are substantially greater than those of migraine, indicating that tension-type headache is the most frequent primary headache disorder in the general population.

Estimates of migraine incidence in the United States converge in demonstrating an earlier peak incidence of migraine, particularly migraine with aura in males in the midteens than in females in whom the peak incidence is in the early 20s. The disability of migraine is most significant in young adults from ages 20 to 30. Despite the high prevalence rates, only 20% continue to manifest chronic migraine over time. In general, both the frequency and duration of migraine decrease at midlife in both men and women and the symptomatic manifestations may change substantially over time.

The severity of migraine ranges from mild to nearly total disability. Over 80% of those with migraine report some degree of disability. Recent community studies have underscored the enormous personal and social burden of migraine in terms of both direct and indirect costs. Although there is an increasing proportion of those with headaches seeking professional treatment, only approximately one-half of those individuals who suffer from debilitating migraine seek professional help.

Migraine in childhood is more likely to be associated with gastrointestinal complaints, particularly episodic bouts of stomach pain, vomiting, or diarrhea, and the duration is shorter than that commonly observed in adults. In women, migraine is strongly associated with reproductive system function, with increased incidence during puberty and the first trimester of pregnancy, and is associated with exogenous hormone use. After menopause, the frequency of migraine attacks generally decreases dramatically unless estrogen replacement therapy is administered.

Aside from sex and age, a family history of migraine is one of the most potent and consistent risk factors for migraine. The results of twin studies implicate genetic factors underlying approximately one-third of the familial clustering of migraine, but the mode of inheritance is clearly complex. Despite an increasing number of candidate gene association studies of migraine, to date, no replicated linkage or associations between specific genes and migraine has emerged, except for hemiplegic migraine. To date, the application of genome-wide association studies in cases and controls has not identified significant associations between migraine and genetic markers.

Migraine is strongly associated with a variety of medical disorders, especially asthma, eczema, allergies, epilepsy, cardiovascular disease, cerebrovascular disease, and particularly ischemic stroke. Anxiety and mood disorders are strongly associated with migraine. Prospective data from community studies of youth reveal that anxiety in childhood is associated with the subsequent development of headaches in young adulthood.

There have been numerous factors that have been consistently implicated as precipitants of acute headache attacks, including hormonal changes, stress, or its cessation, fasting fatigue, oversleeping, particular foods and beverages, drug intake, chemical additives, bright light, weather changes, and exercise, but these agents/situations vary dramatically within and between individuals in prospective research. Many of these factors may be nonspecifically related to headaches in general rather than migraine. There has been a growing effort to examine the role of sleep and circadian rhythms in both migraine pathogenesis and incident attacks.

The prevalence of tension-type headaches is greater in women, but the sex difference is far less pronounced than that of migraine. Although tension-type headache is most common in young adults, there is a less steep decrement in prevalence with age.

By contrast, post-traumatic headache is quite rare in the general population (i.e., about 1% lifetime prevalence). However, it is likely that this is an underestimate because of the lack of systematic data on post-traumatic headaches in population-based samples. Based on retrospective reports of those who suffer from a severe head injury, the prevalence of severe and chronic headaches ranges from 28% to 62%. Children and young adults appear to be particularly susceptible to the development of headaches after head trauma.

The results of prospective studies of the incidence of headache following severe head injury, usually defined as postconcussion headache, reveal that approximately 50% of each series of admissions continue to suffer from headache at the time of discharge from the index admission, with a gradual dissipation to 20% 1 year later. Persistence of headache has been related to female sex, age over 45, the presence of dizziness, lack of skull fracture, intracranial hematoma, disorders of smell, hearing, or vision, depression, and impaired concentration.

The population prevalence of cluster headaches is very low (less than 1% of the general population) and occurs nearly exclusively in males. The age at onset of cluster headache is somewhat later than that of migraine and tension-type headache; the first attack of cluster usually begins in the late 20s or 30s and may recur intermittently throughout life. Risk factors include smoking and heavy alcohol use. There is some family study and twin research demonstrating the role of genetic factors underlying the etiology of cluster headache.

Comorbidity

Although comorbidity of migraine and numerous other conditions has long been reported in clinical samples, etiologic inferences are precluded by the role of comorbid disorders in treatment seeking for primary disorders. There is now strong evidence from population-based surveys regarding associations between migraine and musculoskeletal conditions, cardiovascular disease, particularly ischemic stroke, asthma, allergies and other immune and inflammatory disorders, and epilepsy. However, the only comorbid condition that appears to be specifically associated with migraine is cardiovascular disease. By contrast, the association between migraine and obesity has been shown to be attributable to headaches in general rather than migraine specifically.

Comorbidity between migraine with mental disorders has also been well established in population samples of adults and children. About 40% of young adults with major depression and/or anxiety have migraine, and about 60% of those with migraine have a lifetime history of either anxiety or mood disorder. The presence of a pre-existing physical or medical disorder may also elevate the risk of migraine. Prospective research of both adults and children demonstrates that migraine is associated with an increased risk for the development of depression rather than the converse. However, anxiety disorders, particularly phobias, are associated with an increased risk of incident migraine in prospective research. This link may be a manifestation of underlying autonomic reactivity that may represent a common underlying diathesis for migraine. The elevated rates of infantile colic in children in treatment for migraine would be consistent with this explanation. As such, comorbid disorders may reflect underlying etiology rather than environmental triggers of migraine. Several studies have now confirmed that there is a syndromic association between migraine, depression, and anxiety, with anxiety preceding the onset of migraine followed by the subsequent development of mood disorders. Studies that investigated the specificity of these associations have concluded that migraine was specifically associated with phobias and hypomania/bipolar spectrum conditions, whereas the association with depression was found for headaches in general. Prospective studies of children with headaches have indicated an increased risk for both headaches and somatic and psychiatric symptoms in adulthood. Research into the possible mechanisms underlying these associations, however, remains limited. Studies examining the order of onset and cross-transmission of migraine and psychiatric disorders in controlled family studies have been unable to clearly distinguish between causal and common etiologic models of association.

Comorbidity has a major impact on migraine sufferers’ quality of life and disability and is associated with a poorer course and outcome of migraine. As such, comorbidity may comprise an index of the severity of migraine. For example, simultaneous investigation of somatic and psychiatric comorbidity in headaches and migraine has demonstrated greater health care utilization and negative health perception among those with comorbidity. In fact, about 65% of migraine-associated disabilities can be attributed to comorbid physical and mental disorders.

Impact and Health Seeking of Migraine

Community studies of migraine have documented the enormous personal and social burden of migraine in terms of both direct and indirect costs. Over 80% of those with migraine report some degree of disability. The finding that young adults with migraine suffer from migraine for an average of 1 month of every year across 30 years of prospective follow-up highlights the cumulative impact of migraine during the peak period of attainment of educational, occupational years, and social milestones of adult life. About one-third of people in the general population with active migraine tend to have three or more attacks per month, and more than half reported severe impairment requiring bed rest. Children with migraine have been consistently shown to have more school absences, decreased academic performance, social stigma, and impaired ability to establish and maintain peer relationships. In fact, the quality of life in children with migraine is impaired to a degree similar to that in children with arthritis or cancer.

Severe headaches and migraine not only have a substantial impact on the affected person, but they also have a major economic impact due to medical expenses and employer costs. The bulk of the expenses attributable to migraine derives from its high population prevalence and indirect costs. There have also been expanded efforts to characterize the magnitude of migraine in areas of the world that had been previously underrepresented.

Despite the high magnitude of disability associated with migraine, only approximately one-half of those individuals who suffer from debilitating migraine seek professional help. The gap in treatment is remarkably similar across the world despite variations in health systems across the world. Of those who do seek treatment, many do not continue in treatment. Only a minority of those with migraine in the general population ever seek treatment with clinicians with expertise in headaches. As expected, those who seek professional treatment are characterized by greater severity, longer duration, more disability, and more comorbidity. In light of the overwhelming evidence regarding the tremendous burden of migraine, leaders in the headache field have called for increased awareness of the availability of preventive efforts. Operational criteria for prevention have been developed based on headache frequency and attack-related disability, yet few of those with migraine have received preventive interventions. For example, only 25% of those with migraine in the general population of the United States seek professional treatment and receive appropriate medications, whereas a much smaller proportion (e.g., 11% to 12%) of those with tension-type headaches seek treatment specifically for headaches.

ETIOLOGY

Although the etiology of the major types of headaches is still unknown, recent advances in brain imaging have advanced our understanding of the pathogenesis of migraine. Most theories of migraine focus on the activation of the trigeminovascular system and its central projections. Functional imaging studies have shown that migraine is associated with brainstem activation, particularly the dorsolateral pons. Diffusion tensor imaging studies demonstrate abnormalities in primary and modulatory components of the central nociceptive pathways, such as the periaqueductal gray, an important central modulatory component of the pain pathway, the somatosensory cortex, and in the occipital lobe in subjects with migraine with aura. What remains unclear, however, is whether these findings represent causative abnormalities or whether they have occurred as the consequence of frequent repetitive attacks in migraine sufferers.

Psychophysiologic studies have also contributed to a better understanding of mechanisms underlying migraine pathogenesis. In particular, interictal abnormalities of evoked and event-related potentials have been reported for different **cortical** areas in migraineurs, such as the sensory cortices. Several studies suggest that the excitability of neurons in the visual cortex plays a fundamental role in the brain's susceptibility to **migraine** attacks. Other reports include increased amplitudes of visual-evoked potentials, reduced habituation of **cortical**-evoked responses, and an increased contingent negative variation in migraine patients.

Current theories of the etiology of cluster headaches posit that hypothalamic and central pain control regions trigger a cascade of events in the brain stem comprising afferent pain and efferent parasympathetic pathways. Positron emission tomography (PET) has shown vasodilatation of the major basal arteries during the acute pain attack in cluster headaches, representing the first convincing demonstration of the activation of neuronal vasodilator mechanisms in humans.

Far less is known about the etiology of tension-type headaches. It is clear, however, that tension-type headache is a misnomer since there is no evidence that muscle tension is the underlying cause of this headache subtype.

DIFFERENTIAL DIAGNOSIS AND CLINICAL EVALUATION

A very skillful workup is essential because headache is such a nonspecific complaint with an enormous number of etiologies ranging from trivial to acutely life-threatening. A thorough examination should include a description of the type and location of pain, timing, precipitants, prodromal events, and associated symptoms. Although diaries of headache symptoms have been widely used in the treatment of migraine for several decades, the availability of smartphone applications that can be used to track symptoms, triggers, and treatments is expected to revolutionize our ability to develop individualized treatments in the future. Factors that should be evaluated in the diagnosis of migraine include the following: (1) onset; (2) frequency; (3) location; (4) duration; (5) quality; (6) severity; (7) precipitants (factors that may elevate the risk of a migraine attack such as sleep deprivation or stress); (8) precursors (changes that may occur early in the development of an episode of headache such as visual symptoms or yawning); (9) triggers (factors that precipitate an acute attack such as dietary agents, exposure to particular smells or light); (10) phenomena that worsen or relieve the pain; (11) warning signs (specific precursors particularly visual changes); (12) specific symptoms including visual changes, GI symptoms, or neurologic symptoms; (13) sensitivity to light, noise, sounds, or touch; mood changes; and (14) cognitive changes. In addition, it is important to obtain a detailed family history, description of course, and a history of previous evaluation and treatment. Differential diagnosis of headache is based on a neurologic examination to rule out pathognomonic signs that might indicate other brain disorders. Some of the signs and symptoms that should signal immediate evaluation of headache include first headache, particularly in midlife, worst headache of a person’s life, gradual worsening over days and weeks, vomiting prior to the headache onset, abnormal neurologic evaluation, particularly hemiplegia, ongoing systemic illness, onset after age 50, accompanied by fever, or occurs during sleep.

A migraine is more than a headache. There are a variety of manifestations of migraine that may mimic several neurologic or psychiatric disorders, including epilepsy, psychosis, and functional neurologic disorder. Visual and auditory hallucinations may occur, especially in children. Migraine may have autonomic manifestations suggesting cardiac disease, irritable bowel syndrome, or even acute abdominal emergency. Migraine may be associated with irritability, mood swings, and in some cases, impulsive temper outbursts. In a sense, migraine may be considered as lying on a continuum between the rapid neurophysiologic changes of epilepsy and the less rapid state changes of bipolar disorder. Basilar artery migraine, defined by a particular vascular distribution, may produce stupor and coma or paralysis, blindness, ataxia, dysarthria, or perceptual abnormalities. The extent to which such manifestations may be attributed to comorbid disorders has not been established.

In addition to a history and physical examination, laboratory studies are critical. Even if the results are negative and do not uncover a metabolic, endocrine, or autoimmune etiology, this information may serve as a baseline for subsequent drug therapy. Headache onset over age 50 is particularly ominous and requires careful investigation. Temporal arteritis, or giant cell arteritis, must be ruled out, as it has the potential to cause stroke or blindness if not urgently treated. A sedimentation rate and serum alkaline phosphatase are useful screening tests followed by corticosteroids and antiviral medication. Application of the ICHD-3 requires that all the potential causes of headaches shown in Table 2.11–1 be considered. The diagnosis of headache requires the exclusion of other conditions, including structural lesion, vascular malformation, viral or bacterial meningitis, or encephalitis, intracranial abscess or hemorrhage, cerebral contusion, metabolic disorders (urea cycle disorders, aminoacidopathies, mitochondrial disorders), pseudotumor cerebri, vasculitis, brain tumors, sinusitis, or ocular disorders, any of which may be concurrent rather than causal. One of the most important findings of the past decade is the converging evidence that there is an increased risk of ischemic (but not hemorrhagic) stroke among young women with migraine. This finding supports the importance of the reduction of other stroke risk factors, including oral contraceptive use, hypertension, and smoking among young women with migraine.

Although imaging procedures may not be considered necessary in the evaluation of primary headache syndromes, an image of the brain is mandatory for the evaluation of patients with severe or persistent headaches, the “first” or “worst” headache, or when a subdural hematoma is suspected. An abnormal neurologic examination or seizure mandates brain imaging. A CT scan is indicated to rule out acute hemorrhage, while magnetic resonance imaging (MRI) is indicated when hydrocephalus, brain tumor, sinusitis, vasculitis, or posterior fossa lesions are suspected. X-rays of the jaw and cervical spine are useful to rule out malocclusions and degenerative changes of arthritis.

TREATMENT OF HEADACHE SYNDROMES

Migraine

The mainstay of migraine treatment is pharmacologic intervention. Treatment of migraine is divided into medications that prevent future attacks (prophylactic treatment) and interventions during the acute attack that provide symptom relief and/or progression.

Pharmacologic Prevention Treatments. Preventative treatment of migraine should be instituted when either the frequency or severity of migraine attacks leads to impairment in functioning. As shown in Table 2.11–5, the major classes of drugs that have been investigated in the prophylaxis of migraine include antidepressants, antiseizure medications, β-adrenergic blocking agents, triptans, and Omabotulatxin A. The recent introduction of a new class of drugs, calcitonin gene-related peptide antagonists (CGRP), known as the gepants, has now expanded our ability to both treat and prevent migraine.

Clinical trials of migraine treatment are complicated by the high placebo response rate among subjects with migraine, the heterogeneity of diagnostic subtypes of headache, the intermittent nature of the condition, and the frequent use of additional analgesics to treat headache pain. Due to the highly variable placebo effect in prophylactic migraine RCTs, the European Medicines Agency and the International Headache Society have recommended the inclusion of a placebo in comparative RCTs of migraine prophylaxis.

PHARMACOLOGIC TREATMENTS. Antidepressants have been a mainstay of preventative treatment of migraine for more than 30 years, yet there have been a limited number of controlled trials of either tricyclics or monoamine oxidase inhibitors that were reported to be the most efficacious treatments of migraine, even in the absence of comorbid mood disorders. The two drugs that have shown some evidence of superior performance over a placebo are amitriptyline and venlafaxine. The secondary amines (e.g., nortriptyline and desipramine) that have been shown to be efficacious in the treatment of depression but have fewer side effects than do the parent tertiary amines (e.g., amitriptyline, imipramine) have also been shown to reduce headache in patients with depression. However, the relative efficacy of the various tricyclic antidepressants in migraine prevention has not been examined. The more widely used selective serotonin uptake inhibitors do not have demonstrated efficacy in migraine. In fact, many patients complain of headaches as a secondary effect of the latter class of drugs. Headache is a common symptom encountered in clinical psychiatric practice, particularly among patients with anxiety and depression. This association should be systematically evaluated because headaches can be a side effect of many psychotropic drugs or could be a comorbid condition. If patients with anxiety and/or depression suffer from migraine, it is critical to treat the entire syndrome rather than limiting the treatment goal to headache cessation. In general, comorbid depression and anxiety are more important in the selection of migraine prophylaxis than in the treatment of an acute attack of migraine. The use of prophylactic medications with side effects of lassitude, fatigue, or depression should be avoided, if possible. If not, careful clinical evaluation of the above-cited manifestations of depression, including anergia, hypersomnia, and irritability, should be monitored. Given the overlap of symptoms of the actual migraine episode, including acute changes in energy, appetite, mood, and level of anxiety, as well as those that occur between attacks, and those on the anxiety/depression spectrum, it is not surprising that similar pharmacologic agents have been successfully employed in the treatment of migraine and anxiety/depression.

Table 2.11–5.
Pharmacologic Treatments for the Prevention of Migraine

Class	Established Efficacy	Probable Efficacy
Antidepressants	—	Amitriptyline Venlafaxine
Antiseizure	Divalproex sodium Topiramate Valproate sodium	
β-Blockers	Metropolol Propranolol Timolol	Atenolol Nadolol
CGRP monoclonal antibody	Eptinezumab Fremanezumab Galcanezumab	
CGRP receptor monoclonal antibody	Erenumab	
Neurotoxins	OnabotulinumtoxinA	
Triptans	Frovatriptan	

Three antiseizure drugs that have shown superior efficacy to controls are divalproex sodium, topiramate, and valproate sodium. However, the side effects of these agents often lead to discontinuation, so they, therefore, require careful monitoring.

The three β-blockers that have been shown to be efficacious in migraine are metoprolol, propranolol, and timolol, and there is probable efficacy for atenolol and nadolol. This class of medications has few side effects and is particularly helpful among patients with high levels of autonomic reactivity.

One of the most exciting developments in the treatment of migraine over the past few years has been the development of agents that work directly on the CGRP level or its receptor that is involved in the reactivity of the vasculature. There are now three FDA-approved CGRP monoclonal antibody agents (i.e., eptinezumab, fremanezumab, and galcanezumab) and another CGRP receptor monoclonal antibody named erenumab. Their use has grown rapidly in specialty headache clinical settings because they are reported to have moderate efficacy yet limited side effects, yet their comparative efficacy both within this class and between other prophylactic agents has not yet been demonstrated.

OnabotulinumtoxinA has also been effective for a variety of headache conditions and has primarily been used in controlled trials of chronic migraine. Finally, the only 5-HT 1B agonist that has been shown to be superior to a placebo in the prevention of migraine is frovatriptan. As described below, this class of drugs has been most widely used to treat acute attacks of migraine.

Treatment of Acute Migraine Attacks. Table 2.11–6 shows the current treatments for acute attacks of migraine. The choice of treatment should be based on the severity of the attacks, history of treatment, concomitant conditions, and susceptibility to side effects. The three main classes of treatments are the triptans (5-HT 1B/1D agonists), ergot alkaloids including intranasal dihydroergotamine and ergotamine, the nonsteroidal anti-inflammatories (NSAIDs) including diclofenac potassium, naproxen and aspirin, and acetaminophen. Doses shown in the table are those that yielded significantly better efficacy than placebo in controlled trials. There are multiple triptans that have been introduced, and all appear to have about equivalent efficacy in reducing the acute symptoms of migraine attacks compared to placebos. The NSAIDs and acetaminophen are in widespread over-the-counter use for pain in general. However, they all can have serious side effects, particularly when used in higher doses and over longer periods than indicated.

Table 2.11–6.
Acute Treatment of Migraine

CLASS	Drug (Dose/mg)
Triptans (5-HT 1B/1D Agonists)	Sumatriptan (25; 50) Zolmitriptan (2.5; 5) Naratriptan (2.5) Rizatriptan (5; 10) Eletriptan (20; 40; 80)
Ergot Alkaloids	Dihydroergotamine (intranasal) Ergotamine
NSAIDs	Diclofenac potassium (50–150) Naproxen (500–550) Aspirin (325)
Acetaminophen	Acetaminophen (1,000)

NSAIDs, nonsteroidal anti-inflammatory drugs.

Drug Treatment for Children and Adolescents. There are few well-designed large clinical trials that have examined the efficacy of prophylactic migraine drugs for children. The only effective agents that have been effective in migraine prophylaxis for children in multiple controlled trials are topiramate and flunarizine. The remainder of the studies failed to show the superiority of migraine treatments over placebo or were open-label studies or retrospective reviews. Despite the lack of robust evidence, cyproheptadine is commonly used as a preventative medication in this age group, particularly for children under 12 years of age.

Nonpharmacologic Treatments. There has been substantial research on the use of behavioral treatments, including biofeedback, relaxation training, and cognitive behavioral therapy (CBT), for migraine prevention. Although most of the studies show greater reductions in migraine for the intervention compared to the no-treatment group, the lack of both clinician and patient blinding in these studies diminishes the extent to which such studies may provide conclusive evidence for efficacy. Likewise, herbal treatments such as feverfew (*Tanacetum parthenium*) and dietary supplements such as magnesium and vitamin D have only been assessed in a few studies. Finally, whereas massage and relaxation therapy have been shown to be equivalent to propranolol and other medications in the management of migraine, there is no evidence from controlled trials for cervical manipulation in the treatment of migraine.

The American Academy of Neurology and the American Headache Society developed the latest practice recommendations for pediatric migraine prevention, in 2019, based on a review of 15 class I to III studies. They found that topiramate and cinnarizine (an antihistamine), compared to placebo, caused a decrease in the frequency of migraine or headache days. Similarly, propranolol, compared to placebo, caused a 50% reduction in headache attacks. While amitriptyline alone was not shown to be efficacious in migraine prevention, compared to placebo, children and adolescents with chronic migraine who receive amitriptyline and CBT are more likely than those who receive amitriptyline and headache education from having a reduction in headache frequency. There is insufficient evidence to support the use of onabotulinumtoxinA in adolescents with chronic migraine.

Disease prevention must remain the cornerstone of migraine management in youth. Multiple behavioral factors are known to influence headache frequency. Recurrent headache in adolescents is associated with increased BMI, caffeine and alcohol use, lack of physical activity, poor sleep habits, and tobacco exposure. Headache disability is higher among adolescents with depression. In addition, weight loss can decrease the headache burden in children who are overweight. Thus, recognition and avoidance of factors that contribute to headache risk can reduce migraine frequency.

Tension-Type Headache

Pharmacologic Treatments. At present, ibuprofen (800 mg), which is associated with the lowest risk of GI bleeding or perforation, is the first choice for acute treatment of tension-type headache followed by naproxen sodium (825 mg) which has a higher risk of GI bleed. These recommendations are based on several drug trials that fulfilled the criteria set by the International Headache Society. Taken together, these studies suggest that although aspirin (500 or 1,000 mg) and various NSAIDs are more effective than placebo in aborting tension headaches, the effectiveness of aspirin is comparable to that of paracetamol (500 to 1,000 mg) while that of NSAIDs is superior to simple analgesics. The combination of analgesics and caffeine, sedatives, or tranquilizers might be more effective in some patients than simple analgesics or NSAIDs. The adjunction of caffeine (130 or 200 mg) significantly increases the efficacy of simple analgesics and ibuprofen in controlled trials.

Individuals with chronic tension-type headaches can use preventative treatments such as amitriptyline or topiramate. The initial dose of tricyclic antidepressants should be low (10 to 25 mg amitriptyline or clomipramine before bed) and increased gradually. The average required dose of amitriptyline for patients with chronic tension-type headaches, however, is 50 to 75 mg per day.

Most authorities recommend the discontinuation of treatment after 6 months, regardless of efficacy. A decrease in the daily dose by 20% to 25% every 2 to 3 days might avoid rebound headaches. The mode of action of antidepressants in chronic tension-type headaches is unclear. Tricyclic antidepressants' effect on increasing serotonin and endorphin levels, or inhibiting NMDA receptors, all important chemical components of the pain pathway, might all play a role in alleviating pain in tension-type headaches. Selective serotonin reuptake inhibitors (SSRIs) have not yet been convincingly proven effective for the prevention of tension-type headaches, however.

Nonpharmacologic Treatments. There are numerous behavioral treatments that have been tested in the treatment of tension-type headaches, including relaxation and biofeedback alone or in combination. Cognitive behavioral treatments, such as stress management, are also effective in treating tension-type headaches, especially when combined with biofeedback or relaxation therapies in subjects with high levels of stress.

The improvements produced by behavioral treatment might appear slowly compared with those produced pharmacologically; however, improvement is maintained for longer periods—up to several years—without monthly sessions or contact with the therapist.

The combination of stress management techniques and tricyclic antidepressants has been shown to be superior to either treatment alone in treating chronic tension-type headaches.

A systematic review of randomized clinical trials with physiotherapy and spinal manipulation in patients with tension-type headaches suggests that there is insufficient evidence to support the effectiveness of such techniques. Treatments such as massage, transcutaneous electrical nerve stimulation, and the application of heat or cold, have not been shown to be effective in the long-term treatment of tension-type headaches either. In addition, as already mentioned in the section on nonpharmacologic treatments for migraine, cervical manipulations are associated with the possibility of vertebral artery dissection.

Cluster Headache

First-line abortive therapy is subcutaneous sumatriptan, with most people obtaining relief quickly. Home oxygen can be very useful at high flow, 100% concentration via a nonrebreathable mask at 12 to 15 L/minute for up to 20 minutes. Transitional treatments, such as oral steroids, can help the patient through the acute phase while starting preventative strategies. Prophylactic medicine is almost always indicated for treating cluster headaches because of the extreme severity of pain induced by an acute attack, which often occurs at night. Medications that have been shown to be effective in preventing attacks of cluster headaches are lithium, corticosteroids, methysergide, calcium channel blockers, β-blockers, and valproic acid. Side effects can be severe, and combinations of these agents are often necessary to achieve success. Some may benefit from adjuvant topiramate. Histamine desensitization (that involves the repeated intravenous administration of low dose of histamine in an inpatient setting) and surgical intervention are options upon exhaustion of traditional agents.

Post-Traumatic Headache

There have been no randomized, double-blind, placebo-controlled studies of post-traumatic headaches in either adults or children. Therefore, treatment should be tailored to the symptoms of the individual patient. Most people with post-traumatic headaches have characteristics of either tension-type headaches or migraine, and so treatment is directed toward those entities. If a peripheral mechanism, such as muscular or ligamentous changes, is identified, physical therapy may be indicated. Clinicians should be aware of legal action that may diminish the motivation of the patient to improve.

FURTHER READINGS

Ailani J, Burch RC, Robbins MS; Board of Directors of the American Headache Society. The American Headache Society Consensus Statement: update on integrating new migraine treatments into clinical practice. *Headache*. 2021;61(7):1021–1039.

Burch R. Antidepressants for preventive treatment of migraine. *Curr Treat Options Neurol*. 2019;21(4):18.

Burch R. Epidemiology and treatment of menstrual migraine and migraine during pregnancy and lactation: a narrative review. *Headache*. 2020;60(1):200–216.

Burch R. Migraine and tension-type headache: diagnosis and treatment. *Med Clin North Am*. 2019;103(2):215–233.

Burch RC, Buse DC, Lipton RB. Migraine: epidemiology, burden, and comorbidity. *Neurol Clin*. 2019;37(4):631–649.

Charles A. The pathophysiology of migraine: implications for clinical management. *Lancet Neurol*. 2018;17(2):174–182.

Dao JM, Qubty W. Headache diagnosis in children and adolescents. *Curr Pain Headache Rep*. 2018;22(3):17.

de Boer I, van den Maagdenberg AMJM, Terwindt GM. Advance in genetics of migraine. *Curr Opin Neurol*. 2019;32(3):413–421.

Diener HC, Tassorelli C, Dodick DW, et al; International Headache Society Clinical Trials Standing Committee. Guidelines of the International Headache Society for controlled trials of acute treatment of migraine attacks in adults: fourth edition. *Cephalalgia*. 2019;39(6):687–710.

Edvinsson L, Haanes KA, Warfvinge K, Krause DN. CGRP as the target of new migraine therapies–successful translation from bench to clinic. *Nat Rev Neurol*. 2018;14(6):338–350.

GBD 2017 US Neurological Disorders Collaborators; Feigin VL, Vos T, Alahdab F, et al. Burden of neurological disorders across the US from 1990–2017: a global burden of disease study. *JAMA Neurol*. 2021;78(2):165–176.

Hamilton KT, Singh RBH, Ailani J, Butterfield RJ, Robbins MS. Migraine treatment in pregnancy: an American Headache Society survey. *Headache*. 2020;60(10):2594–2596.

Hansen JM, Charles A. Differences in treatment response between migraine with aura and migraine without aura: lessons from clinical practice and RCTs. *J Headache Pain*. 2019;20(1):96.

Herd CP, Tomlinson CL, Rick C, et al. Botulinum toxins for the prevention of migraine in adults. *Cochrane Database Syst Rev*. 2018;6(6):CD011616.

Jackson JL, Kuriyama A, Kuwatsuka Y, et al. Beta-blockers for the prevention of headache in adults, a systematic review and meta-analysis. *PLoS One*. 2019;14(3):e0212785.

Krause DN, Warfvinge K, Haanes KA, Edvinsson L. Hormonal influences in migraine – interactions of oestrogen, oxytocin and CGRP. *Nat Rev Neurol*. 2021;17(10):621–633.

Lateef T, He JP, Nelson K, et al. Physical-mental comorbidity of pediatric migraine in the Philadelphia Neurodevelopmental Cohort. *J Pediatr*. 2019;205:210–217.

Loder E, Weizenbaum E, Frishberg B, Silberstein S, American Headache Society Choosing Wisely Task Force. Choosing wisely in headache medicine: the American Headache Society's list of five things physicians and patients should question. *Headache*. 2013;53(10):1651–1659.

Merikangas KR. Contributions of epidemiology to our understanding of migraine. *Headache*. 2013;53(2):230–246.

Minen MT, De Dhaem OB, Kroon Van Diest A, et al. Migraine and its psychiatric comorbidities. *J Neurol Neurosurg Psychiatry*. 2016;87(7):741–749.

Nierenburg HDC, Ailani J, Malloy M, Siavoshi S, Hu NN, Yusuf N. Systematic review of preventive and acute treatment of menstrual migraine. *Headache*. 2015;55(8):1052–1071.

Nye BL, Thadani VM. Migraine and epilepsy: review of the literature. *Headache*. 2015;55(3):359–380.

Ong JJY, De Felice M. Migraine treatment: current acute medications and their potential mechanisms of action. *Neurotherapeutics*. 2018;15(2):274–290.

Patterson-Gentile C, Szperka CL. The changing landscape of pediatric migraine therapy: a review. *JAMA Neurol*. 2018;75(7):881–887.

Suri H, Ailani J. Cluster headache: a review and update in treatment. *Curr Neurol Neurosci Rep*. 2021;21(7):31.

Szperka CL, VanderPluym J, Orr SL, et al. Recommendations on the use of anti-CGRP monoclonal antibodies in children and adolescents. *Headache*. 2018;58(10):1658–1669.

Todd C, Lagman-Bartolome AM, Lay C. Women and migraine: the role of hormones. *Curr Neurol Neurosci Rep*. 2018;18(7):42.

Vgontzas A, Pavlović JM. Sleep disorders and migraine: review of literature and potential pathophysiology mechanisms. *Headache*. 2018;58(7):1030–1039.

Yang CP, Liang CS, Chang CM, et al. Comparison of new pharmacologic agents with triptans for treatment of migraine: a systematic review and meta-analysis. *JAMA Netw Open*. 2021;4(10):e2128544.

2.12 Neuropsychiatric Aspects of Neuromuscular Diseases

TIMOTHY P. WUERZ, D.O., M.SC., RAJEET SHRESTHA, M.D., AND ALAN J. LERNER, M.D.

The peripheral nervous system (PNS) is a complex system of interconnected motor, sensory, and autonomic nerves and pathways that allow for rapid information passage to and from the central nervous system (CNS). In daily life, it subserves vital functions, including sensation leading to perception, movement around joints throughout the body, and coordination of automatic functions.

A wide variety of disorders occur in the PNS, and two major classes of neuropsychiatric aspects can be considered. Some disorders, such as amyotrophic lateral sclerosis (ALS) or systemic vasculitis, may affect both PNS and CNS simultaneously. Other disorders, such as myasthenia gravis

(MG), affect the PNS solely, and their neuropsychiatric aspects arise directly from the disabling injury, its treatment, or its psychosocial effects. Medications and toxic exposures may also affect both systems. This chapter will cover the structure of the PNS and major disorders with neuropsychiatric aspects important for clinical practice.

STRUCTURE OF THE PERIPHERAL NERVOUS SYSTEM

The traditional separation of the nervous system into PNS and CNS is not always anatomically straightforward and at times results in ambiguous distinctions. Some portions of the PNS have their cell bodies located in the CNS (e.g., anterior horn cells or motor neurons that innervate skeletal muscle fibers), whereas other neurons originating in the PNS (e.g., sensory neurons located in the dorsal root ganglia) project axons into the CNS where they synapse in the dorsal horn of the spinal cord (e.g., ascending nerve tracts in the spinal cord). The autonomic nervous system can be divided into sympathetic and parasympathetic units and has central and peripheral components. Nevertheless, this imperfect separation is well established and has resulted in distinct neurologic subspecialties, including those related to peripheral neuromuscular diseases and clinical electrodiagnostic medicine. Advances in diagnostic testing such as neuroimaging are providing new insights into nervous system structure, which may change how we understand and categorize CNS and PNS structures.

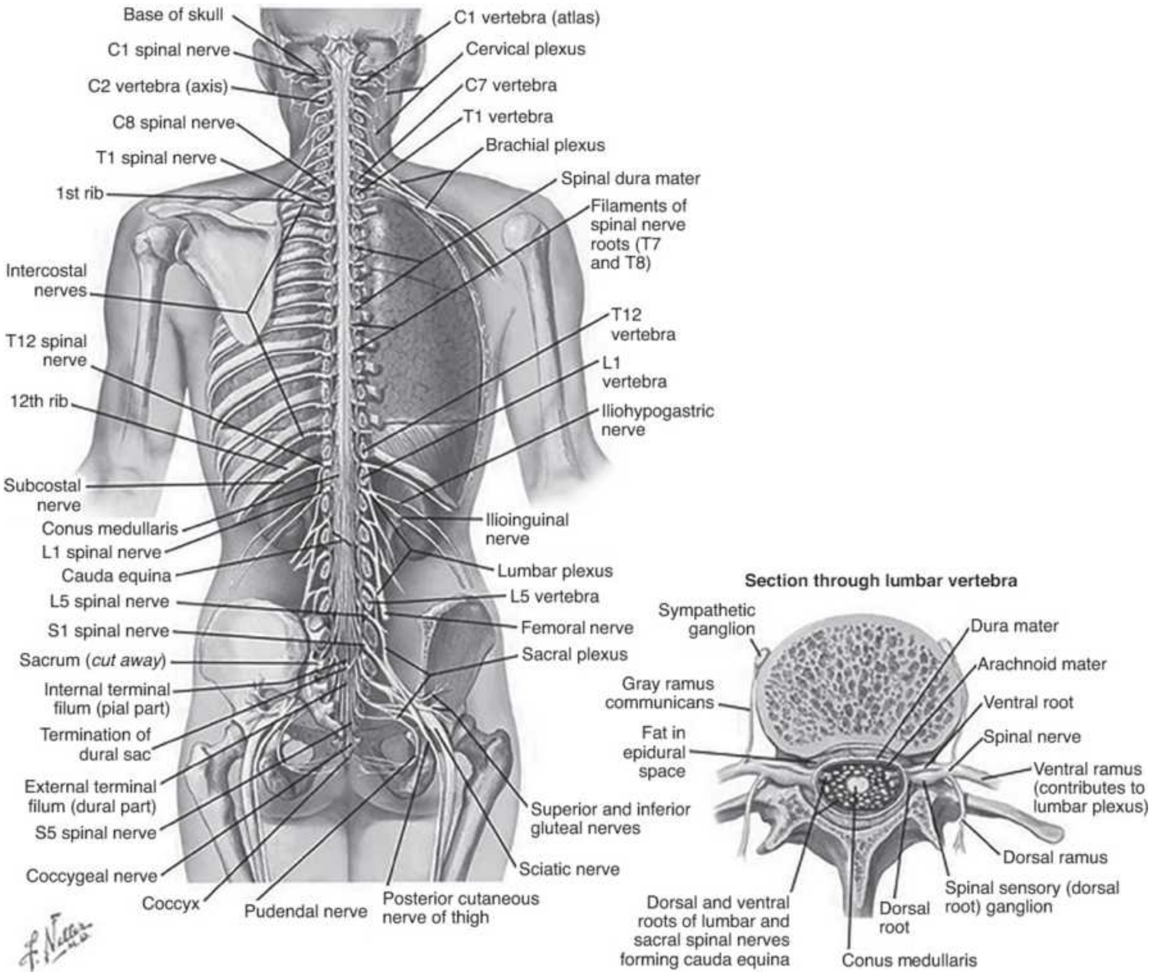
Base of skull
C1 spinal nerve
C2 vertebra (axis)
C8 spinal nerve
T1 spinal nerve
1st rib
Intercostal nerves
T12 spinal nerve
12th rib
Subcostal nerve
Conus medullaris
L1 spinal nerve
Cauda equina
L5 spinal nerve
S1 spinal nerve
Sacrum (cut away)
Internal terminal filum (pial part)
Termination of dural sac
External terminal filum (dural part)
S5 spinal nerve
Coccygeal nerve
Coccyx
Pudendal nerve
Posterior cutaneous nerve of thigh
C1 vertebra (atlas)
Cervical plexus
C7 vertebra
T1 vertebra
Brachial plexus
Spinal dura mater
Filaments of spinal nerve roots (T7 and T8)
T12 vertebra
L1 vertebra
Iliohypogastric nerve
Ilioinguinal nerve
Lumbar plexus
L5 vertebra
Femoral nerve
Sacral plexus
Superior and inferior gluteal nerves
Sciatic nerve
Section through lumbar vertebra
Sympathetic ganglion
Gray ramus communicans
Fat in epidural space
Dura mater
Arachnoid mater
Ventral root
Spinal nerve
Ventral ramus (contributes to lumbar plexus)
Dorsal ramus
Spinal sensory (dorsal root) ganglion
Dorsal root
Conus medullaris
Dorsal and ventral roots of lumbar and sacral spinal nerves forming cauda equina

FIGURE 2.12–1. Peripheral nervous system (*Netter's Neurology*. 2nd ed.) (Asset provided by Anatomical Chart Co.).

The PNS is defined to include the motor neurons and their axons projecting into the periphery through peripheral nerves, the neuromuscular junction, and the innervated musculature. A motor unit encompasses all muscle fibers innervated by a single motor axon. This may vary from a few fibers in an eye muscle to several hundred in the quadriceps, allowing for graded contraction responses. The peripheral sensory system consists of various modality-specific receptors, including touch, pain, temperature, vibration and joint position sense, afferent sensory axons, their cell bodies in the dorsal root ganglion, and the portion of the axon projected centrally that comprises the dorsal nerve roots.

The ventral (motor) and dorsal (sensory) segmental nerve roots merge to form the segmental spinal nerves, which in turn combine to form the brachial and lumbosacral plexuses before dividing into the individual peripheral nerves (sensory, motor, or mixed sensorimotor) and their autonomic nervous system components. The cranial nerves (except olfactory and optic nerves) are similar to the segmental nerves, comprising peripheral and central segments but traversing or originating in the brainstem instead of the spinal cord (Fig. 2.12–1).

NEURONOPATHIES AND DISORDERS OF THE DORSAL ROOT GANGLION

Poliomyelitis and Other Disorders of Anterior Horn Cells

Poliomyelitis has been eliminated throughout much of the world, although pockets of infection remain in rural areas in Pakistan, Afghanistan, and scattered cases in the Middle East and Africa. Polio is the quintessential lower motor neuron disease (MND) affecting motor neurons in the spinal cord, although bulbar symptoms may occur. Syndromes similar to polio may occur with arthropod-borne diseases such as West Nile virus (WNV) infection and with idiopathic cases of flaccid paralysis. WNV may be limited to a segment of the spinal cord with concomitant weakness, but encephalitis with mental status changes and coma may occur. The prognosis for recovery from WNV infection is guarded, and supportive care and counseling throughout are important components of care for patients with this condition.

Postpolio Syndrome

Postpolio muscle dysfunction disorder has an unclear etiology. A consensus workshop in 1996 proposed the following diagnostic criteria: (1) history of paralytic Polio: confirmed or not confirmed with partial or fairly complete functional recovery; (2) after a period of functional stability of at least 15 years, development of new muscle dysfunction emerges; (3) muscle weakness; (4) muscle atrophy; (5) muscle pain; and (6) fatigue. Neurologic examination in postpolio muscle dysfunction should be compatible with prior poliomyelitis with lower motor neuron lesion, decreased or absent tendon reflexes, no sensory loss, and compatible findings on electromyography (EMG) and/or MRI.

Postpolio syndrome can lead to significant distress, anxiety, and depression as the pain and fatigue cause marked changes in the level of functioning. In a sample of polio survivors, more than a quarter had depression, while those with postpolio syndrome and lower family functioning were at higher risk of depression. It is important to carefully assess postpolio patients for psychiatric complications and to consider interventions, including counseling and antidepressants. Sedatives and alcohol should be avoided due to the increased risk of falls and hypoventilation. Physical rehabilitation may provide benefit, and seral medications have been used empirically without clear sustained benefit.

Amyotrophic Lateral Sclerosis

Long considered a very rare disease, ALS, the most common MND, has an annual incidence of roughly 2 per 100,000 and a prevalence of 6 per 100,000 persons. The typical clinical presentation is the combination of both upper motor neuron and lower motor neuron signs. Rarer forms include prominent bulbar dysfunction and pure lower- or upper-motor neuron presentation at least initially. Lower motor signs include weakness, muscle atrophy, and the presence of fasciculations, which are painless spontaneous discharges of motor units and involuntary rapid twitches in limb or trunk muscles. EMG and nerve conduction studies (NCS) play critical roles in the evaluation and may reveal fibrillations, a sign of active denervation. Upper motor neuron signs include spasticity, loss of dexterity, hyperreflexia, and pseudobulbar palsy. Death from respiratory failure or aspiration typically occurs within 3 years of diagnosis, although occasionally, patients may have a prolonged course, particularly if a decision is made by the patient to accept respiratory support. Because of the uniformly fatal course of ALS, a wide-ranging differential diagnosis should be considered, encompassing virtually the entire neuraxis.

The pathology of ALS involves degeneration of the motor neurons of the spinal cord (anterior horn cells) and brainstem, as well as the upper motor neurons of the cerebrum. Gliosis and neuronal loss occur in superficial layers of the dorsomedial neocortex of the frontal lobes and may occur in the hippocampal and parahippocampal regions in the temporal lobes. About 5% to 20% of cases are familial, and several different genes have been found in association with ALS. These include *SOD1, C9ORF72, VCP, TARDBP, FUS, CHMP2B*, and *ATXN2*.

There has been increasing recognition of the neuropsychiatric aspects of ALS, most notably its association with frontotemporal degeneration (FTD). FTD and ALS occur on a spectrum, and detailed examination may show a cognitive decline in patients presenting with classic ALS symptoms. FTD is clinically heterogeneous and may present as a behavioral and dysexecutive syndrome, primary progressive aphasia, or parkinsonism, with features of progressive supranuclear palsy or corticobasal syndrome

(Table 2.12–1). ALS may develop in patients with previously diagnosed FTD or vice versa. Patients with the *C9ORF72* expanded hexanucleotide repeat mutation may present with behavioral syndromes, especially psychosis and agitation. These seemingly disparate disorders are linked by genetic and pathologic commonalities, with ubiquitin-positive neuronal cytoplasmic inclusions in lower motor neurons, hippocampus, and neocortex in both conditions. Gliosis and neuronal loss in superficial layers of the dorsomedial neocortex of the frontal lobes also occur across the FTD-ALS spectrum.

One-third to one-half of patients with ALS perform poorly on neuropsychological testing compared with matched control subjects. The pattern of cognitive change generally includes mild memory loss, associated word finding and semantic fluency difficulties, and executive function impairments. Facial emotional recognition is also impaired in ALS patients. These patterns of cognitive impairment differ substantially from that seen in Alzheimer disease, wherein memory loss and apraxia predominate. The presence of cognitive impairment has been associated with decreased survival time. Acetylcholinesterase inhibitors are of doubtful value in individuals with cognitive impairment in this population.

Pseudobulbar affect (PBA) is a syndrome of disinhibited affective display in which patients demonstrate inappropriate and exaggerated laughing or weeping responses to environmental stimuli and may affect up to 50% of ALS patients. The behaviors are not usually associated with disturbances of the underlying mood, and the behavioral phenotype tends to be stereotypical for each patient. Successful treatment of PBA in ALS patients has been reported with tricyclic compounds or selective serotonin reuptake inhibitors. A combination of dextromethorphan and quinidine (Nuedexta) has been found to be effective in many cases and is U.S. Food and Drug Administration (FDA) approved for the treatment of PBA.

There are several medications approved by the U.S. FDA for ALS treatment. Riluzole blocks glutamate release and prolongs survival by about 3 months in clinical studies. Tiglutek is a thickened liquid formulation of Riluzole designed to avoid problems that may be caused by crushing tablets. Riluzole's side effects include bladder pain, dark urine and difficulty, burning, or painful urination, tachycardia, malaise, cough, itching skin, and muscle aches and pains. Riluzole overdose may cause agitation.

Edaravone is administered intravenously at a dose of 60 mg according to the following schedule: an initial treatment cycle with daily dosing for 14 days, followed by a 14-day drug-free period. Subsequent treatment cycles with daily dosing for 10 days out of 14-day periods, followed by 14-day drug-free periods. Clinical studies favored Edaravone on the ALS Functional Rating Scale-Revised (ALSFRS-R). The ALSFRS-R scale consists of 12 questions evaluating fine motor, gross motor, bulbar, and respiratory function of patients with ALS (speech, salivation, swallowing, handwriting, cutting food, dressing/hygiene, turning in bed, walking, climbing stairs, dyspnea, orthopnea, and respiratory insufficiency). Each item is scored from 0 to 4, with higher scores representing greater functional ability.

Table 2.12–1.
Toxic Agents and Impact on CNS and PNS

Toxic Agents	**Means of Exposure**	**Impact to CNS/PNS**	**Treatment**
Lead	Various trades including plumbing, antiquated water supply, illegal moonshine production.	Cognitive dysfunction, acute encephalopathy with ataxia, tremulousness, choreiform movements, and hallucinations. Can progress to seizures, coma, or death.	Utilizing chelating agents and preventing or minimizing exposure.
Arsenic	Massive ingestion, chronic low-grade exposure from smelting of lead and copper ore, contaminated well water, tainted illicit drugs, and burning of preserved wood.	Death, organic psychosis, CNS dysfunction. Peripheral neuropathy symptoms include burning and numbness of the feet followed by hands.	Utilizing chelating agents and preventing or minimizing exposure.
Mercury	Antiquated thermometers, latex paints as preservative of latex paint, manufacture of chlorine, in dental amalgam, and	Encephalopathy [“mad as a hatter”], ataxia, dystonic movements, numbness in the hands and feet, general muscle weakness,	Removal from exposure, use of chelating agents, and preventive education.

	in the natural gas industry.	narrowing of the field of vision, and damage to hearing and speech.	
Carbon disulfide	Industrial and chemical polar solvent, factory workers, major metabolite in the breakdown of the drug disulfiram (Antabuse).	Memory impairment, confusion, hallucinations, and emotional lability.	Pyridoxine may have a role in treatment.
Organophosphates	Insecticides, flame retardants, various other industrial manufacturing components, and in chemical warfare. Accidental exposure from agricultural spraying is the most common cause of individual intoxication.	CNS involvement includes emotional lability, decreased alertness, cognitive impairment, convulsions, coma in extreme cases, weakness of the neck flexors and proximal muscles, cranial muscles including the extraocular muscles may be involved, dystonic posturing may be seen, neuropathy, and organophosphate-	Administration of atropine. Cathartics and gastric lavage are indicated in the case of acute ingestion.

		induced delayed polyneuropathy.	
Hexacarbons	Components of glues and lacquers and solvents, production and refining of petroleum products. High-level exposure can also occur with intentional inhalation such as “glue sniffing.”	CNS depression or stupor. Repeated exposure via “glue sniffing” leads to cognitive deficits, motor and sensory neuropathy, cranial nerve and autonomic dysfunction.	Cessation of hexacarbon exposure.

Spinal Muscular Atrophies

Spinal muscular atrophies (SMAs) are disorders with lower motor neuron dysfunction due to the loss of anterior horn cells. The syndromes may present in infancy (Werdnig–Hoffman syndrome; SMA I), childhood (chronic infantile; SMA II), adolescence (Kugelberg–Welander disease, SMA III), or adulthood (SMA IV). Developmental arrest and limited survival are features of SMA type I and II. SMA is an autosomal recessive disorder, with an incidence of about 1 per 10,000 live births. The differentiation of these complex disorders is increasingly being done by genetic analysis. The autosomal recessive proximal SMAs have been linked to mutations or deletions in the survival motor neuron (SMN) gene on chromosome 5q11.12–13.3. Variants of SMA also linked to chromosome 5q include SMA with arthrogryposis and a very severe form with survival measured in weeks after birth. There are also SMA syndromes, often with other neurologic features, that are not linked to 5q mutations. These include Kennedy disease (X-linked bulbospinal muscular atrophy with androgen resistance), Fazio–Londe (progressive bulbar atrophy of childhood), late-onset Tay–Sachs disease, focal atrophy syndromes, and distal SMAs (“spinal” Charcot–Marie–Tooth disease, and autosomal dominant adult-onset SMA).

Long considered completely untreatable, a remarkable development is a new treatment for SMA with nusinersen, a survival motor neuron-2 (SMN2)-directed antisense oligonucleotide indicated for the treatment of SMA in pediatric and adult patients. SMA gene therapy is a one-time intrathecal infusion that replaces the defective or missing *SMN1* gene with a working copy. This new gene increases SMN protein levels, which improves motor neuron function and increases the likelihood of survival. Other gene-directed therapies are in development.

Dorsal Root Ganglion Disorders

Disorders of the dorsal root ganglia present as pure sensory syndromes. Paraneoplastic sensory neuropathies due to inflammation of the dorsal root ganglion may be associated with the production of other autoimmune antibodies producing paraneoplastic encephalomyelitis. Symptoms of encephalomyelitis include cognitive and behavioral disturbances (also seen in limbic encephalitis), ataxia, seizures, cerebellar ataxia, opsoclonus, myoclonus, and gut dysmotility.

Table 2.12–2.
General Categories of Sensory or Sensorimotor Neuropathies

Acute	Subacute/Chronic	Subacute/Chronic
Guillain–Barré syndrome (acute inflammatory demyelinating polyneuropathy [AIDP]) Toxic HIV-associated AIDP Vasculitis Lyme disease Acute pandysautonomia Porphyria Critical illness neuromyopathy	Vitamin B_{12} deficiency Environmental toxicity (arsenic, lead, N-hexane, organophosphates, taxol, pyridoxine, cisplatin, vincristine, dapsone, chloroquine, many other medication and toxins) Paraneoplastic Diabetes Hypothyroidism Monoclonal gammopathies (especially associated with multiple myeloma or other neoplasms) Sarcoidosis Connective tissue disorders (SLE, Sjogren, scleroderma) Hereditary (Charcot–Marie–Tooth, Fabry disease, Tangier disease) POEMS syndrome Amyloidosis	Cryoglobulinemia Acromegaly Graft vs. host disease Thiamine deficiency Gastric bypass

SLE, systemic lupus erythematosus. POEMS, polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, skin abnormalities.

PERIPHERAL NEUROPATHIES

There are many ways to classify peripheral neuropathies. A clinically useful two-pronged approach is to consider (1) the onset and course as acute, subacute, or chronic; (2) and whether the involvement is pure motor, mixed motor and sensory, or pure sensory, and with or without autonomic involvement. This classification applied to specific disorders is presented in Table 2.12–2.

Sensory or Sensorimotor Neuropathies

Neuropathies involving sensory nerves or their cell bodies typically present with symptoms and signs of sensory loss involving the distal limbs. Sensory symptoms are very hard to describe in ordinary language: symptoms described as “burning,” “freezing,” or “cutting” may represent identical pathophysiology and, indeed, may be described alternately or even simultaneously in the same patient. Despite decreased sensation, patients paradoxically show hyperesthesia and often complain of uncomfortable sensations, characterized as burning and tingling, spontaneously or in response to a stimulus such as light touch. Traditional modality-specific sensory testing is quite crude and is aimed primarily at detecting changes in sensory thresholds. Clinical testing does not address the stimulus–response relationship, which may help explain this apparent paradox of decreased sensation and hyperesthesia. Most neuropathies are sensorimotor in nature, as most peripheral nerves subserve both functions.

Diagnosis of Neuropathies

The diagnostic evaluation of neuropathy is a highly specialized process, with a bewildering number of conditions possible with overlapping clinical phenotypes (Table 2.12–2). The clinical history may suggest positive family history, the presence of early onset, and slow progression, consistent with an inherited etiology. History may also suggest toxic exposures and systemic conditions (neoplasms, autoimmune disorders, and medications) associated with neuropathies. A thorough examination can help identify symptom patterns to aid the diagnostic process. The American Academy of Neurology recommends initial screening for diabetes, checking serum electrophoresis for underlying monoclonal gammopathy and vitamin B_{12} level, and using electrodiagnostic testing such as NCS and EMG for further classification. Nerve and muscle biopsy may be indicated in certain situations, and skin biopsy may help confirm small fiber neuropathy. Nonetheless, some neuropathies may remain undiagnosed. Neuromuscular ultrasound is being increasingly recognized as a valuable adjunct to NCS/EMG. It can show local pathologies, impacting nerve structure and function that would not otherwise be suspected, especially in focal syndromes such as entrapment neuropathies, traumatic injuries, and plexopathies.

Treatment of Neuropathies

Treatment of neuropathy is primarily directed at treating the underlying cause. Neuropathic pain is one common problem associated with neuropathies. Treatment of neuropathic pain is typically resistant to traditional analgesics (i.e., NSAIDs or opiates) and requires a multimodal approach, often involving adjunctive pain medications. Antidepressants with mixed pharmacologic effects are useful in chronic pain syndromes, including the pain associated with some peripheral neuropathies. Amitriptyline (Elavil) is inexpensive and effective for this indication and is favored by many as a first-choice agent. Doses in the range of 25 to 75 mg per day are usually sufficient to generate some symptomatic improvement. Imipramine may also be used and has fewer anticholinergic adverse effects. Duloxetine (Cymbalta) is FDA approved in the United States for the treatment of painful diabetic neuropathy. Daily doses for duloxetine range from 30 to 120 mg per day. Venlafaxine has also been recommended as a first-line treatment for neuropathic pain at doses similar to those used for the treatment of depression. Pregabalin (Lyrica) is indicated for several neuromuscular conditions, specifically, management of neuropathic pain associated with diabetic peripheral neuropathy, management of postherpetic neuralgia (PHN), management of fibromyalgia, and management of neuropathic pain associated with spinal cord injury. Neuropsychiatric adverse effects of pregabalin include mood changes, anxiety, panic attacks, sedation or insomnia, impulsivity, agitation or aggression, and suicidal ideation. Dosage ranges from 50 mg three times daily up to 100 mg three times a day are used in patients with creatinine clearance of at least 60 mL/min, with reduced dosage based on renal function. In addition to being frequently used for neuropathic pain, gabapentin (Neurontin) is also indicated for usage in PHN at doses ranging from 100 mg/day up to 600 mg three times daily and must be adjusted for renal insufficiency. The neuropsychiatric side-effect profile is similar to that seen with pregabalin.

Clinicians must be aware that the approved indications for the use of duloxetine, pregabalin, and gabapentin with regard to neuropathic pain are limited. Their usage, or the use of many antidepressant medications for modulation of neurotransmitter systems involved in pain transmission and perception, in other pain syndromes, other neuropathies or for vague indications, is encountered commonly and may be of limited efficacy. Correction of metabolic abnormalities, removal of toxic exposures, and nonpharmacologic approaches such as physical therapy, psychological interventions, and exercise need to be integrated into a holistic care model.

Physical and occupational therapy can address motor weakness and provide techniques and adaptive equipment to help overcome sensory loss.

Diabetic Neuropathy

Diabetes mellitus is the most common cause of sensory or sensorimotor neuropathy in the United States. In primary care settings, more than half of diabetic patients have symptoms and signs of peripheral neuropathy, and diabetes may be newly diagnosed during the evaluation of a new-onset painful neuropathy. Improved glycemic control slows the progression of diabetic neuropathy, but symptomatic improvement of unpleasant sensory symptoms remains a major clinical challenge.

Guillain–Barré Syndrome

Guillain–Barré syndrome (GBS) is an acute inflammatory demyelinating polyneuropathy (AIDP). It is classically described as ascending paralysis, frequently with sensory loss and autonomic instability. The pattern of presentation is somewhat variable and may be misdiagnosed as a nonspecific illness at first. Onset may follow a nonspecific illness, such as an upper respiratory infection. Cases of GBS, especially those associated with *Campylobacter jejuni* infection, may have an axonal injury in addition to demyelinating features. The presence of axonal injury and loss may predict a poorer prognosis. During the acute phase, the disorder can be life-threatening, with about a quarter requiring admission to an intensive care unit (ICU) for mechanical ventilation. The majority of individuals affected with GBS recover and similar degrees of improvement rate are seen with both γ-globulin infusions and with plasmapheresis. Both plasmapheresis and intravenous γ-globulin are used to hasten the return of nerve function caused by GBS, whereas steroids have not proven efficacious in this regard.

It is important for the psychiatrist to be aware of the associated neuropsychiatric manifestations both directly and indirectly from an acute GBS syndrome with a resulting ICU admission in terms of anxiety, acute stress disorder, depressive episodes, and brief reactive psychosis. Psychotic episodes are strongly associated with severe quadriparesis, artificial ventilation use, and multiple cranial nerve dysfunctions. Patients express that experiencing and coping with the loss of communication is the most difficult aspect of GBS. Therapeutically, patients often feel reassured by the ICU environment and contact with relatives and visitors is helpful. Supportive counseling is a valuable adjunct, in addition to psychopharmacologic measures, in helping patients cope with this illness, which can be devastating in its effects despite a good long-term prognosis.

Chronic Inflammatory Demyelinating Polyneuropathy

Chronic inflammatory demyelinating polyneuropathy (CIDP) is an immune-based neuropathy of unknown cause and may resemble GBS in its early stages. Both cellular and humoral components of the immune response contribute to pathogenesis, although the precipitating antigen(s) is unknown. There is a breakdown of the blood–nerve barrier and recruitment of macrophages. Secretion of toxic factors then damages myelin sheaths and may produce axonal injury. In HIV-positive patients, there may be as many as 50 WBC/cubic mm in the CSF, mimicking a chronic CNS inflammatory process.

CIDP may occur at any age and may occur in connection with other autoimmune disorders and monoclonal gammopathy of unknown significance. Successful treatments of CIDP with corticosteroids, immunoglobulin infusions, plasmapheresis, and immunosuppressants have been reported. The uncertainties regarding prognosis present a special challenge to the patients, and dealing with these iatrogenic and situational problems can benefit from psychiatric consultation.

Nutritional and Alcoholic Neuropathies

Alcohol can affect every level of the PNS and CNS. Persons with alcohol-related neuropathies frequently have a history of alcohol dependence and other comorbid vitamin and nutritional deficiencies. Alcohol-related major neurocognitive disorder is complicated by the various syndromes described in individuals who abuse alcohol, as well as other possible comorbidities contributing to cognitive dysfunction in these individuals, including vitamin B_{12} deficiency, head injuries and subdural hematomas cerebrovascular disease, and so forth. It is also unclear to what degree alcohol-related major neurocognitive disorder overlaps with Wernicke–Korsakoff syndrome. Alcohol-related major neurocognitive disorder is diagnosed much more often than Wernicke–Korsakoff syndrome, but many people diagnosed with alcohol-related syndromes have a thiamine deficiency often missed by treating health professionals. Research has identified other metabolic disturbances that result in functional thiamine deficiency.

Knowledge about whether alcohol abuse may be a risk factor for other dementias such as Alzheimer disease is also sparse.

Vitamin B_{12} deficiency affects multiple levels of the nervous system, which may present as peripheral neuropathy, myelopathy, or cognitive syndrome. Vitamin B_{12} is essentially found only in animal products, and deficiency can be due to decreased intake or failure of absorption since intrinsic factor is a gastric protein important for facilitating dietary absorption. Sensitivity in determining vitamin B_{12} deficiency may be enhanced by the measurement of intermediate metabolites such as methylmalonic acid and serum homocysteine. Rapid repletion of vitamin B_{12} is usually accomplished by intramuscular administration of 1 g daily for a week, then weekly for the first month, followed by maintenance therapy on a monthly basis. Chronic oral repletion is possible, even in patients with low intrinsic factor activity, through a low-affinity uptake mechanism.

Cranial Nerve Disorders

The 12 cranial nerves are largely analogous to the segmental peripheral nerves and have their cell bodies within the CNS. A variety of conditions can selectively affect cranial nerves, and occasionally these are associated with neuropsychiatric complications. Examples include cranial neuropathies secondary to concomitant meningeal inflammation, which may occur with granulomatous diseases such as tuberculosis or sarcoidosis. Cranial neuropathies may be associated with acute or subacute alterations in baseline mental status, hydrocephalus, or major neurocognitive disorder.

A clinical challenge may occur with cranial neuropathies in association with neurodegenerative disorders. Anosmia (olfactory nerve: CN I) may occur years before the classic motor symptoms of Parkinson disease or memory loss in Alzheimer disease. Visual hallucinations or complex visuospatial disturbances, seen in major neurocognitive disorder secondary to Alzheimer disease or dementia with Lewy bodies, often calls for ophthalmologic investigation of retinal and optic nerve (CN II) function as part of the neuropsychiatric treatment plan. Visual hallucinations can occur in the visually impaired and are termed Charles Bonnet syndrome (CBS). The visual impairment in CBS can be secondary to any cause of visual impairment and is considered to be a cortical release phenomenon. Patients with pre-existing neurodegenerative disorders may be predisposed to CBS. The treatment of CBS is based on the improvement of vision impairment.

The importance of sensory loss due to hearing loss and visual loss (e.g., cataracts, macular degeneration, and sensorineural hearing loss) is being increasingly recognized as an important component of multidisciplinary evaluations since memory is critically dependent on sensory input for later recall.

Peripheral Nerve Injury

Traumatic peripheral nerve injuries range from stretch injury without rupture of the nerve sheath (e.g., foot drop after hip surgery) to complete transection of the affected nerve(s). While neural repair with reinnervation of affected nerves and their targets often occurs, the estimated daily repair rate is about 1 mm/day; thus, the time required for full functional return may be weeks to months. Patient expectations and secondary mood adjustments may respond to psychiatric interventions.

Concomitant traumatic brain injury (TBI), spinal cord, or severe musculoskeletal injury greatly complicates peripheral nerve injury, particularly in the acute post-traumatic period. Concurrent therapies are frequently needed as the presence of multiple injuries is often associated with greater mood effects; the range of symptoms with TBI include altered mental state, psychosis, disinhibition, mood effects, and focal neurologic injuries and should be treated as appropriate.

Metabolic, Endocrine Neuropathies, and Ischemic Neuropathies

This diverse group of disorders is of interest to psychiatrists because of concurrent psychiatric comorbidity. Metabolic and endocrine disorders associated with neuropathy include uremia, hepatic failure, hyper- and hypothyroidism, and neuropathy associated with acromegaly. Pituitary tumors, in general, are often associated with psychiatric symptoms, including depression, anxiety, apathy, behavioral and personality changes, and cognitive impairment. Changes in attention, memory, and executive function have been reported to be common in acromegalic individuals.

Environmental and Toxic Neuropathies

Toxic neuropathies can be caused by exposure to substances, including chemicals or heavy metals (Table 2.12–1). These exposures may be accidental (in the workplace or through drinking water) or intentional (homicide or suicide). Most toxic neuropathies produce symmetric deficits, but the symptoms and signs may vary based on dose and the intensity or duration of exposure to the toxic substance. A clue to toxic exposures is that the symptoms usually involve multiple organ systems and rarely present as an isolated peripheral neuropathy. The first step in identifying a neuropathy as toxin related is simply having a high degree of suspicion.

Frostbite

Frostbite is an ischemic neuropathy, and high-risk groups include outdoor workers, the older adults, the homeless, drug or alcohol abusers, psychiatric diseases, high-altitude or cold-weather athletes, and military personnel. It should be easily diagnosed, although several grades of severity are recognized. Psychiatric illness is very pervasive in victims of frostbite, and psychiatric screening for patients with frostbite has been highly encouraged. Acute treatments include rewarming and treatment of ischemic injury, often resulting in the need for amputation. Long-term consequences include sensitivity to cold and chronic limb pain.

Lead. Lead toxicity is a well-known cause of neurologic dysfunction. Potential for exposure is usually via inorganic lead and may occur in automobile factory workers, plumbers, exposure to outdated paint products, and consuming illegal “moonshine” whiskey. Exposure via lead-contaminated drinking water led to widespread population exposure in Flint, Michigan. Lead intoxication in children is most commonly associated with CNS dysfunction. This may lead to chronic cognitive dysfunction (developmental delay or loss of milestones) or acute encephalopathy with ataxia, tremulousness, choreiform movements, and hallucinations, which may progress to seizures, coma, or death. Acute or chronic exposure in adults can also lead to encephalopathy. Chronic lead exposure presentation usually includes gastrointestinal (GI) disturbance, anemia, weight loss, fatigue, renal dysfunction, and bluish discoloration of the gums (“lead lines”), in addition to a neuropathic picture. Porphyria should be included in the differential. Treatment consists of utilizing chelating agents and preventing or minimizing exposure.

Arsenic. Arsenic is famous as a poison used for suicide or homicide, usually resulting from massive ingestion, although it can also result from chronic low-grade exposure from the smelting of lead and copper ore, contaminated well water (a common occurrence in Bangladesh), tainted illicit drugs, and burning of preserved wood. High-level acute exposure causes multiple organ dysfunction, organic psychosis, CNS dysfunction, and possible death. Peripheral neuropathy manifests within weeks and may continue for a period of weeks after removal of the exposure. Peripheral neuropathy symptoms include burning and numbness of the feet followed by hands. Treatment of acute ingestion requires ICU care. Chronic exposure treatment includes chelating agents and education of those at risk.

Mercury. Mercury toxicity can be of several types: (1) elemental mercury found in outdated thermometers, barometers, and other gauges; (2) organic mercury found in latex paints as a preservative or as an environmental pollutant, exemplified by the so-called "Minamata disease" following massive organic mercury pollution in Japan; (3) and inorganic mercury salts used in the manufacture of chlorine, in dental amalgam, and in the natural gas industry.

Intoxication with mercury results in predominantly CNS dysfunction (encephalopathy ["mad as a hatter"], ataxia, dystonic movements, numbness in the hands and feet, general muscle weakness, narrowing of the field of vision, and damage to hearing and speech). Associated systemic toxicity includes GI symptoms and nephrotic syndrome. PNS dysfunction is seen in all forms of mercury poisoning. Treatment of mercury intoxication is primarily by removal from exposure, use of chelating agents, and preventive education.

Carbon Disulfide. Carbon disulfide is a clear liquid that vaporizes at room temperature and is readily absorbed by inhalation but is also absorbed via the GI tract and skin. It is also a major metabolite in the breakdown of the drug disulfiram (Antabuse). Acute or subacute exposure to high levels of carbon disulfide causes CNS dysfunction, including memory impairment, confusion, hallucinations, and emotional lability primarily. Pyridoxine may have a role in treatment.

Organophosphates. Organophosphorus (OP) compounds are used as insecticides, flame retardants, various other industrial manufacturing components, and in chemical warfare. Accidental exposure from agricultural spraying is the most common cause of individual intoxication. OP compounds are a frequent agent for suicide and are implicated in almost 20% of suicides in India.

The sarin gas attack in the Tokyo subways has been well studied, and a small number of survivors had a short-lived neuropathy. Post-traumatic stress disorder was common in those survivors. OPs have also been implicated in the Gulf War syndrome, with various combinations of PNS and CNS symptoms.

OPs prevent the degradation of acetylcholine (ACh) by disrupting acetylcholinesterase resulting in excessive stimulation of both muscarinic and nicotinic receptors. There are acute (type I) and intermediate (type II) cholinergic syndromes depending on which receptor is preferentially affected. Type I results from excessive stimulation of muscarinic cholinergic receptors. An acute syndrome frequently begins hours after exposure and always within 1 day. Symptoms include GI effects, micturition, sweating, and tachycardia or bradycardia. CNS involvement includes emotional lability, decreased alertness, cognitive impairment, convulsions, and coma in extreme cases. Type II occurs within 12 to 96 hours after exposure and is due to nicotinic ACh receptor activation in skeletal muscle. The initial feature is usually respiratory insufficiency, followed by weakness of the neck flexors and proximal muscles. The cranial muscles, including the extraocular muscles, may be involved. Dystonic posturing may be seen. This neuropathy has been labeled the organophosphate-induced delayed polyneuropathy (OPIDP) since there is an interval of 7 to 21 days after exposure before peripheral neuropathy symptoms are seen. Although less common than the cholinergic syndromes, OPIDP causes significant morbidity, including foot drop, claw hand deformity, or atrophy in extreme cases of severe neuropathy. Administration of atropine only treats the muscarinic type I syndrome and does not prevent the intermediate type II nicotinic cholinergic syndrome. Reactivation of acetylcholinesterase may be accelerated by the administration of pralidoxime. Cathartics and gastric lavage are indicated in the case of acute ingestion. Chronic low-level exposure to OP can result in mood dysregulation, increased risk of suicidality, and increased impulsivity. Psychiatric and psychological support is an important adjunct to more direct treatments and should be initiated early in the treatment regimen.

Hexacarbons. Hexacarbons are components of glues and lacquers, and solvents. Hexacarbon exposure occurs in the production and refining of petroleum products, and high-level exposure also occurs with intentional inhalation such as “glue sniffing.” Acute exposure causes CNS depression and stupor. Repeated exposure via “glue sniffing” leads to cognitive deficits, motor neuropathy, and cranial nerve and autonomic dysfunction. Severe neuropathies can result in sensory loss, atrophy, and distal weakness. The only treatment is the cessation of hexacarbon exposure. Psychiatric concerns from “glue sniffing” include the emergence of violent behavior and/or self-mutilation, functional deterioration, impairment of intelligence, and possible development of psychosis. Glue sniffing is often considered a “gateway drug” to other forms of substance abuse, and specialized addiction treatments should be considered in these cases.

Vasculitic Neuropathies and Neuropathies of Connective Tissue Disorders. Systemic disorders, especially primary vasculitides and connective tissue disorders (many of which are autoimmune in etiology), often affect peripheral nerves. Concomitant neuropsychiatric complications may be primary such as vasculitis affecting CNS structures, or secondary to treatment (e.g., steroid-related mood disturbance) or adjustment to severe chronic illness. Prominent syndromes affecting the CNS are systemic lupus erythematosus (SLE), Sjögren syndrome, giant cell arteritis, and disorders of circulating autoantibodies, especially the antineutrophil cytoplasmic antibodies (ANCA) classically seen in granulomatosis with polyangiitis (Wegener granulomatosis).

Infectious and Granulomatous Neuropathies. The infectious and granulomatous neuropathies result from a variety of inflammatory conditions. Peripheral nerve involvement in these disorders may manifest clinically as mononeuropathy, mononeuropathy multiplex, polyneuropathy, or polyradiculopathy. Nerve pathology may result from an inflammatory reaction induced by the infection and from the host’s immunologic reaction.

HIV-ASSOCIATED NEUROPATHIES. More than one-third of AIDS patients are estimated to have peripheral neuropathy. HIV neuropathy may affect proximal and distal components of the motor, sensory, and autonomic PNS. Other AIDS-related neurologic conditions, such as myelopathy, myopathy, and dementia, may coexist.

The diagnosis of HIV-associated neuropathy may be particularly complicated and challenging, as a wide range of potential etiologies underlies peripheral neuropathy in HIV-infected patients. Etiologies of HIV-associated neuropathy include cytotoxic effects due to direct HIV infection, cytokine dysregulation, opportunistic infection, neoplastic processes, autoimmune disorders, metabolic abnormalities, and toxic effects of medications.

AIDS patients present with a variety of PNS syndromes. Distal symmetrical polyneuropathy (DSP) is a painful neuropathy occurring in 30% of AIDS patients. Patients complain of numbness, burning, and paresthesias of the feet, with pain usually beginning in the toes and soles of the feet. Inflammatory demyelinating polyneuropathy produces rapidly progressive muscle weakness of two or more limbs, often starting in distal muscles and may resemble GBS or CIDP. Mononeuropathy multiplex is a neuropathy that affects single or multiple nerves causing asymmetric motor and sensory complaints presenting with weakness or dysesthesias in the distribution of cutaneous nerves, mixed nerves, and nerve roots. Progressive polyradiculopathy is a serious condition presenting as a rapidly progressive cauda equina syndrome, characterized as leg weakness, difficulty walking, bowel and bladder dysfunction, low back pain, and leg paresthesias.

Prompt recognition and treatment of neuropathy may significantly benefit a patient's quality of life. Possible etiologies range from relatively benign facial neuropathy to severe opportunistic infections, including lymphomatous meningitis, CMV encephalitis, and *Cryptococcal* meningitis. In addition, cranial neuropathies occasionally are caused by primary CNS lymphoma, non-Hodgkin lymphoma, or brain stem glioma. Treatable conditions in HIV populations include B_{12} deficiency, ethanol abuse, neurotoxic medications, uremia, or diabetes.

Evaluation of AIDS or HIV-related neuropathies consists of ancillary tests, including NCS, EMG, autonomic testing, and blood, CSF, and neuroimaging studies. Treatment of HIV-related peripheral neuropathy depends on etiology but has four main components to be considered: antiviral treatment, immunomodulation, treatment of underlying causes, and symptom reduction.

HIV-associated peripheral neuropathy is often refractory to standard treatments for pain. For example, a placebo-controlled study of amitriptyline (tricyclic antidepressant) and mexiletine (antiarrhythmic) in patients with HIV-associated peripheral neuropathy found no significant improvement in pain intensity relative to placebo with either treatment. Similarly, a randomized, placebo-controlled study of the efficacy of acupuncture, amitriptyline, and a combination of the two treatments for pain relief in patients with HIV-associated peripheral neuropathy found that none of the treatments worked better than placebo. However, one study showed that lamotrigine worked significantly better than a placebo to decrease neuropathic pain. Severe rash, an important adverse effect of lamotrigine treatment, occurred more frequently than in studies of lamotrigine treatment for epilepsy. Studies suggest initial trials of gabapentin may be helpful, titrated up to 1,200 or 2,400 mg per day, divided into two or three times daily doses. Topical capsaicin can be used in cases that have contraindications, cannot tolerate, or prefer to avoid systemic medication.

LEPROUS NEUROPATHY. Leprosy is caused by *Mycobacterium leprae,* an acid-fast bacillus, and is characterized as a chronic infectious disease that affects the skin and peripheral nerves. The prevalence of leprosy has declined greatly in developed countries and is seen only rarely in immigrants from endemic areas. Sensory loss due to intracutaneous nerve damage is the hallmark of leprous neuropathy. The goal of treatment for leprous neuropathy is the eradication of *M. leprae.* The prognosis for neuropathy depends on the type of leprosy, the time of diagnosis, chemotherapeutic response, and reactive sites.

LYME NEUROPATHY. Lyme disease, the most common tick-borne disease in the United States, is caused by *Borrelia burgdorferi,* a gram-negative spirochete, and transmitted by *Ixodes dammini* and related species. After the incubation period of 3 to 30 days, three stages emerge. Stage I is characterized by a rash at the tick attachment site, causing constitutional symptoms, although many patients do not recall a tick bite. Neurologic complications develop during stages II (early) and III (late). Lymphocytic meningitis and lymphocytic meningoradiculitis (LMR) are common in early neuroborreliosis. Patients with LMR usually have variable motor and sensory involvement of multiple nerves, and symptoms may wax and wane over several months. Focal and multifocal peripheral and cranial nerve involvement is observed in half of patients. Facial nerve paralysis is the most common sign.

In Europe, neuroborreliosis has a somewhat different presentation with meningoradiculoneuritis, characterized by severe radicular pain with or without paresis. Late neuroborreliosis is marked by peripheral neuropathy and encephalomyelitis. The PNS abnormalities of this late stage last for weeks to months. Predominant sensory neuropathy may be seen in up to two-thirds of untreated patients. Stage III is characterized by asymmetric oligoarthritis, cardiac involvement, and a variety of CNS syndromes. Evaluation of Lyme disease consists of a laboratory two-test protocol (ELISA) and Western blot performed on serum or CSF. PNS complications of Lyme disease are treated by Borrelia eradication. Neurologic manifestations carry a good prognosis after antimicrobial therapy.

HERPES ZOSTER NEUROPATHY. Approximately 95% of the U.S. population has been infected with VZV, a Herpes virus that is the causative agent for chickenpox. Latent, lifelong infection persists in sensory nerve ganglia and can reactivate to cause zoster neuropathy, or “shingles.” VZV neuropathy results from the reactivation of latent virus from a primary infection. VZV infection is more commonly recurrent and chronic in AIDS and cancer patients receiving chemotherapy and radiotherapy. The most common PNS manifestations of VZV infection include PHN, radiculopathy, zoster sine herpete (zoster without rash), and cranial nerve palsies. PHN can be diagnosed when pain persists more than 4 months after VZV eruption and is more common in older adults. Vaccines are available for the prevention of chickenpox in children and for risk reduction of zoster and PHN in adults. Antiviral treatment is also available for acute attacks, prophylaxis against recurrent attacks, and decreasing the incidence of HSV. Antidepressants, such as amitriptyline, desipramine, and nortriptyline, may relieve pain associated with PHN, but the side-effect risks limit their use. Anticonvulsants are of limited value in the management of neuritic pain. Opioids are the most effective agents for severe neuritic pain. In addition, topical lidocaine 5% gel or patch, nerve blocks, skin infiltration, transcutaneous electrical nerve stimulation, IV lidocaine, and capsaicin dermal patch have been shown to provide adjunctive relief in severe PHN.

COVID-19–RELATED NEUROMUSCULAR DISORDERS

Adults. The COVID-19 pandemic has resulted in numerous unusual neuromuscular disorders both in children and adults. It is also associated with substantial worsening in patients with pre-existing neuromuscular disorders, especially those who have pre-existing breathing or swallowing dysfunction. A wide variety of neuromuscular disorders have now been associated with COVID-19, although the precise etiology and pathophysiology are often unclear.

There have several case reports of GBS with COVID-19, and GBS can be seen in conjunction with several viral diseases. Molecular mimicry between viral proteins and peripheral nerve epitopes on gangliosides has been postulated to underlie this association. There have been few clear-cut cases of motor neuron invasion or inflammation, as can be seen with WNV or poliomyelitis.

Myopathy may be a common problem, with reports from China of frequent increased serum creatine kinase, and myositis is common in conjunction with other coronaviruses. Critical illness neuromyopathy in ICU patients with prolonged sedation with nondepolarizing neuromuscular blockade is possible, and weakness from disuse atrophy in patients with prolonged illness is also likely. Fatigue is common after COVID-19 infection in symptomatic patients and may be due to a combination of factors of both PNS and CNS origin, as well as nonneurologic factors such as reduced pulmonary function.

COVID-19 therapies may also increase the risk of neuromuscular dysfunction. Patients may be on prior immunosuppressive therapies, and protease inhibitors may increase the risk of neuropathy in HIV patients. Chloroquine and hydroxychloroquine have been discredited as COVID-19 treatments, but both may be associated with a slowly progressive syndrome of proximal weakness affecting legs more than arms.

Children and Adolescents. A multisystem inflammatory syndrome in children (MIS-C) has been detailed, and diagnostic criteria have been published. The CDC definition includes clinically severe illness requiring hospitalization; and multisystem (two or more) organ involvement (cardiac, renal, respiratory, hematologic, GI, dermatologic, or neurologic). MIS-C has some features similar yet distinct from other childhood multisystem inflammatory disorders such as Kawasaki syndrome and toxic shock syndrome.

Neurologic manifestations may affect as much as 50% of MIS-C patients. These may be of CNS origin (headache, encephalopathy) or PNS such as severe proximal weakness with reduced reflexes. In a series of 82 children from New York City early in the pandemic, 43% had neurologic signs and symptoms. Neuromuscular complications included fatigue in 25% and weakness in 14%, and three patients with cranial nerve VI palsy. Most had evidence of recent or ongoing COVID-19. The full spectrum of this unusual disorder remains to be fully appreciated as cases accumulate in the ongoing COVID-19 pandemic.

SARCOID NEUROPATHY. Sarcoidosis is a multisystem granulomatous disorder that affects young and middle-aged adults and is manifested clinically with bilateral hilar adenopathy, pulmonary infiltration, reticuloendothelial system, ocular, and cutaneous involvement. It is estimated that neurosarcoidosis affects 5% to 16% of sarcoidosis patients, and of these, the PNS is affected in approximately 85% of patients. Sarcoid neuropathy can affect large nerve fibers or small nerve fibers. Large fiber neuropathy may present as cranial neuropathy, mononeuropathy multiplex, radiculopathy, or symmetric polyneuropathy. It may mimic GBS or CIDP or cause a small fiber neuropathy with neuropathic pain or autonomic symptoms. Cranial neuropathy is most common and often involves the facial nerve, which may be bilateral. The optic nerve may be affected, and hearing loss is not uncommon. Intra- and extradural granulomatous masses may cause radiculopathy. Endocrine dysfunction may occur due to the involvement of the hypothalamus or pituitary gland. Focal-onset or generalized tonic–clonic seizures may also occur. Neuropsychiatric symptoms include memory loss, changes in mood or behavior, delirium, and psychosis. The diagnosis of sarcoid neuropathy is straightforward when accompanied by recognized systemic involvement. In spite of their potential side effects, corticosteroids remain the mainstay of therapy.

NEUROMUSCULAR JUNCTION DISORDERS

Myasthenia Gravis

MG is the most common disease of the neuromuscular junction. MG is an autoimmune disease, with several identified autoantibodies directed mainly against postsynaptic ACh receptors on the muscle membrane. Two major clinical subtypes are ocular myasthenia and generalized myasthenia. Most frequently, cranial nerves are affected initially, especially extraocular muscles presenting as fluctuating double vision, facial weakness, or difficulty swallowing. Ocular MG may progress to generalized MG or remain an isolated syndrome.

Weakness in MG is classically effort dependent, so weakness increases with exertion and frequently worsens as the day progresses. MG severity also fluctuates over time with frequent remissions and exacerbations, and there are no objective findings on standard neurologic examination such as muscle atrophy or altered reflexes. Perhaps for these reasons, patients with MG frequently are thought to have psychiatric or psychological conditions or even seemingly unlikely diagnoses such as “psychogenic respiratory arrest” before the appropriate tests establish the diagnosis (e.g., elevated ACh receptor antibody levels, abnormal response to repetitive motor nerve stimulation, or abnormal single-fiber electromyogram).

Just as the fatigable weakness of MG should not be casually attributed to psychiatric disorders, vague symptoms such as “tiredness” or “fatigue,” which can be seen in many psychiatric conditions, should not necessarily be considered a manifestation of MG. Patients with somatization disorder have been misdiagnosed as MG, sometimes with disastrous outcomes.

Dysphagia is a prominent and sometimes isolated symptom in early MG. Dysphagia is a nonspecific symptom and is frequently associated with many anxiety disorders, and several figures of speech recognize this connection between anxiety and difficulty swallowing (“I choked,” “I had a lump in my throat,” “I can’t swallow that”). Accompanying facial weakness characterized by intermittent ptosis, disconjugate gaze, and inability to express facial expression is seemingly inconsequential compared to the life-threatening components of MG. However, for many patients, these cosmetic and communicative features are far more disconcerting, often reaching pathologic levels and requiring psychiatric support. Myasthenic patients demonstrate high rates of anxiety symptoms when followed longitudinally through the course of their illnesses, especially anxiety symptoms focused on respiratory status. Research does not presently view these psychiatric symptoms as generated directly by the neurobiology of MG.

Acute episodes of worsening (myasthenia crises) can be life-threatening if they involve respiratory paralysis or dysphagia with aspiration. MG treatment involves the use of anticholinesterase medications to prolong the availability of ACh in the neuromuscular junction, increasing the likelihood that the transmitter will find an available ACh receptor and the use of long-term immunosuppressant medications, most commonly corticosteroids, with their attendant adverse behavioral side effects. Many medications can exacerbate weakness in MG, and clinicians should always check new medications to determine if they may worsen neuromuscular transmission in patients with MG. Lists of these medications to be avoided or used cautiously in MG patients are widely available, or clinicians should coordinate care with a treating neurologist.

Eculizumab is an FDA-approved monoclonal antibody for generalized myasthenia in antibody-positive patients and is also indicated for paroxysmal nocturnal hemoglobinuria (PNH) to reduce hemolysis atypical hemolytic uremic syndrome (AHUS), and neuromyelitis optica spectrum disorder in adult patients who are anti–aquaporin-4 antibody positive.

Neuropsychiatric side effects may include headache, nausea, vomiting, or fatigue.

Lambert–Eaton Myasthenic Syndrome

Lambert–Eaton myasthenic syndrome (LEMS) is sometimes confused with either MG, chronic fatigue syndrome, or fibromyalgia. LEMS is an autoimmune paraneoplastic syndrome that results in progressive weakness frequently with autonomic symptoms and may progress to respiratory insufficiency. The majority of LEMS patients have a circulating voltage-gated calcium channel antibody. LEMS may be a presenting symptom of a neoplasm, with small-cell lung cancer and breast cancers most common. Although LEMS is not associated with psychiatric symptoms, weakness may be misattributed to psychiatric disorders. Amifampridine phosphate is a newly FDA-approved treatment for LEMS. Side effects may include seizures, allergic reactions, paresthesias, and insomnia and should not be used in addition to other aminopyridine compounds. Many medications lower seizure thresholds, and they should be used with caution in this context to prevent epileptic seizures.

MUSCLE DISEASES

Myotonic Dystrophy

Myotonic dystrophy, the most frequent *adult* muscular dystrophy, is often easily recognized in adults by the characteristic facial features, grip myotonia (e.g., difficulty releasing the hand grip), percussion myotonia (percussion of a muscle evokes myotonia), and distal weakness. Myotonic dystrophy occurs in 15 of 100,000 live births and is autosomal dominant with incomplete penetrance; heterozygotes may present at any age. The genetic basis for myotonic dystrophy type I is an expanded trinucleotide repeat in a protein kinase gene on chromosome 19. The rarer myotonic dystrophy type 2 (DM2) presents with proximal weakness and maps to chromosome 3. *Anticipation* may occur with the development of more severe symptoms in successive generations due to expansion of the length of trinucleotide repeat between generations. Systemic manifestations of myotonic dystrophy include cataracts, diabetes, temporal baldness, and testicular atrophy. Cardiac conduction defects, cardiomyopathy, and potentially fatal arrhythmias may occur and should be treated aggressively.

Respiratory muscle weakness may develop later in the illness and is often the cause of death.

The interference in ribonucleic acid (RNA) processing resulting from expanded trinucleotide repeats accounts for the peripheral manifestations and the behavioral syndromes that appear predictably among patients with the disease. Cognitive impairment of variable severity occurs in patients with myotonic dystrophy. The pattern of cognitive impairment most commonly includes personality change, alterations of visuospatial and constructional abilities, and prefrontal or executive cognitive domains. The severity of cognitive impairment does not correlate with age or severity of muscle disease but correlates most closely with the length of the trinucleotide repeat expansion in the kinase gene. The cognitive impairment syndrome may remain stable over a prolonged period. Along with cognitive dysfunction, a range of personality traits from avoidant to obsessive-compulsive and passive-aggressive tendencies may be found.

Patients with myotonic dystrophy frequently complain of excessive daily sleepiness (hypersomnia), sleeping 12 or more hours per day. They may have hypnagogic hallucinations upon awakening and may resemble narcolepsy. As many as 39% of myotonic dystrophy patients have such symptoms, and there is a positive correlation between complaints of hypersomnia and functional disability due to muscular impairment. Central and obstructive sleep apnea with chronic hypercapnia may further complicate excessive daytime somnolence in myotonic dystrophy patients. The neurophysiologic basis of this complaint is somewhat similar to that found in narcolepsy, with reports of decreased hypocretin-1 in cerebrospinal fluid, suggestive of dysfunction of the hypothalamic hypocretin system. Therapy of hypersomnolence has generally relied on stimulants such as modafinil (Provigil) or methylphenidate. Many drugs have been used for myotonia, but large-scale randomized clinical trial data are lacking.

Dystrophinopathies

Different mutations in the dystrophin gene on the X chromosome produce Duchenne muscular dystrophy (DMD) and the milder Becker muscular dystrophy phenotype. Dystrophin is one component of a glycoprotein complex located in the sarcolemmal membrane of muscle fibers. In DMD, dystrophin is virtually absent, while in Becker dystrophy, there is a partial deficiency of dystrophin. Since dystrophin is X linked, males are predominantly affected, while female carriers are typically asymptomatic.

DMD affects 1 of 3,500 live male births. Up to one-third of cases of DMD are sporadic.

Symptoms in DMD start with proximal muscle weakness and calf muscle hypertrophy developing at 3 to 4 years of age. Weakness is progressive and becomes more diffuse, with loss of ambulation by 12 years of age and death by the age of 20 due to respiratory failure. In Becker dystrophy, the clinical syndrome is less severe, with symptomatic weakness beginning between ages 5 and 15. Loss of ambulation typically occurs after 15 years of age, and death often occurs in the third or fourth decade. A mild form of Becker dystrophy may present in middle age with myalgias and exertional muscle cramping. Either disorder may be associated with symptomatic or asymptomatic cardiomyopathy.

Although usually considered an isolated muscle disease, a variety of cognitive and behavioral syndromes among patients with dystrophinopathies may be directly related to the neurobiology of the diseases. The dystrophin gene product is expressed in the brain as well as in muscle, although its CNS function is unknown.

A range of depressive and even psychotic syndromes has been reported. Approximately 30% of Duchenne patients have cognitive impairments by comparison with age-matched controls, but also by comparison with matched subjects with other spinal muscular disorders. Verbal function is more affected than visuospatial functions. Prior studies reported an average intelligence quotient (IQ) of 85 in Duchenne patients. The cognitive deficit is usually not progressive.

A glucocorticoid with anti-inflammatory and immunosuppressive properties, Deflazacort, is approved for the treatment of DMD in patients 5 years of age and older. Glucocorticoids can change the natural history of DMD, prolonging independent ambulation, preserving cardiac and pulmonary functioning in the teen years, and preventing severe scoliosis. However, side effects may include mood swings and difficult behavior. While anecdotal accounts suggest that these negative reactions may improve over time, they continue to remain one of the primary reasons for the discontinuation of therapy.

The impact of treatment with steroids on learning and memory is currently unclear. Acute steroid treatment for any condition has been shown to result in reversible impairment of long-term memory functions and retrieval of negative words, suggesting that emotional material is especially sensitive to the memory-modulating effects of steroids. Some research indicates that glucocorticoid-associated memory effects may be dose dependent.

Several medications have been approved that target increase in normal intracellular dystrophin levels in DMD patients with specific mutations amenable to exon skipping (exons are pieces of DNA that provide information for making proteins in a person's genome). The following medications are available: Eteplirsen for those with confirmed dystrophin mutations amenable to exon 51 skipping, Golodirsen for those with mutations amenable to exon 53 skipping, and Casimersen for those patients with a mutation amenable to exon 45 skipping. A clinical benefit of these medications, including improved motor function, has not been established. Nonetheless, consideration of the potential risks associated with these medications, the life-threatening and debilitating nature of DMD, and the lack of available therapy has opened new hope for what has been considered an incurable and untreatable disorder.

Inflammatory Myopathies

There are three major syndromes of inflammatory myopathies: polymyositis (PM), dermatomyositis (DM), and inclusion body myositis (IBM). All may be associated with autoimmune disorders such as SLE, and DM, in particular, may be associated with occult neoplasms, suggesting a paraneoplastic syndrome. Neuropsychiatric complications may occur secondary to the associated autoimmune disorder. The weakness in PM/DM is usually symmetric and proximal, whereas weakness in IBM may be asymmetric and distal, with prominent early involvement of quadriceps, finger and wrist flexors, and ankle dorsiflexors. The weakness is usually slowly progressive, serum creatine phosphokinase (CPK) level is elevated, and many patients are not diagnosed for years after onset. PM and DM are treated with steroids, immunosuppressant agents, and γ-globulin, with their attendant neuropsychiatric risks, whereas IBM responds less well to immunosuppression.

Inclusion Body Myositis. IBM is a disease of adults, most commonly over the age of 50 years. It affects men more than women. The nature of the inclusion bodies and the finding of β-amyloid in muscle fibers has led to theories that IBM may be a degenerative disorder of muscle (with homologies to Alzheimer disease) or possibly of viral origin. Rarely, IBM may be familial. Familial IBM is associated with Paget disease of bone, and frontotemporal dementia is associated with valosin-containing protein

(VCP) gene mutations. Neuropsychiatric features of this syndrome may include behavioral variant FTD, primary progressive aphasia, and MND. Although no approved treatments are available, a study of Rapamycin showed promise on several secondary endpoints in a small clinical study.

Mitochondrial Myopathies

Muscle tissue, unlike peripheral nerves, is very energetically active. A number of myopathies, collectively called mitochondrial myopathies, can present with symptoms from childhood onward. Mitochondria are the “powerhouse of the cell” and contain their own DNA, which is uniquely transmitted by the mother (since sperm do not contribute mitochondria), so family histories show no male-to-male transmission. However, many mitochondrial proteins are of nuclear origin and show classic Mendelian inheritance, so not all mitochondrial myopathies are due to intrinsic mitochondrial gene mutations. Mitochondrial disorders affect bioenergetic processes with dysfunction in multiple body systems, resulting in effects such as developmental short stature, learning deficits, visual loss and blindness, eye movement abnormalities, deafness, cardiomyopathy and conduction defects, diabetes, liver and kidney problems, GI motility defects, and stroke-like symptoms.

Psychiatric features are found within the clinical spectrum of mitochondrial syndromes. Case reports have linked depression and obsessive-compulsive disorder to specific mitochondrial mutations. It is increasingly recognized that mitochondrial dysfunction may be associated with neuropsychiatric abnormalities such as dementia, major depression, and bipolar disorder. Furthermore, several lines of evidence suggest that there is mitochondrial dysfunction in schizophrenia, including documented alterations in brain energy metabolism, electron transport chain activity, and expression of genes involved in mitochondrial function. Treatments for mitochondrial disorders include coenzyme Q10 and the amino acid L-arginine. Patients with mitochondrial encephalopathy with lactic acidosis and stroke-like episodes (MELAS) who also have muscle weakness have defective respiratory chain enzymes, resulting in a high $NADH/NAD^+$ ratio that inhibits the nitric oxide (NO) synthetase reaction to cause a decreased production of NO in vascular endothelium or arterial smooth muscle. Low plasma concentration of L-arginine, a potent donor of NO, has been observed in patients with MELAS, suggesting the possibility of L-arginine’s therapeutic potential.

Endocrine Myopathies

Endocrine myopathies include weakness due to hormonal imbalance in the pituitary, adrenal, thyroid, or parathyroid glands. Examples of this class of muscle disorders include steroid myopathy with weakness caused by either endogenous or exogenous glucocorticoid; psychiatric comorbidities of mood or thought disturbance are common. Adrenal insufficiency is associated with generalized weakness and fatigue. Treatment is aimed at normalizing hormonal levels or removal of toxic compounds as tolerated while working with the patient and medical team to optimize quality of life.

MISCELLANEOUS NEUROMUSCULAR DISORDERS

Stiff-Person Syndrome

Stiff-person syndrome is a rare disorder of the CNS, characterized as having progressive and fluctuating muscle rigidity with painful spasms. The spasms may occur at rest and may be triggered by sensory stimuli. The etiology of stiff-person syndrome is autoimmune and is often associated with type 1 diabetes and with dysfunction in the GABA inhibitory system. A majority of patients have autoantibodies to glutamic acid decarboxylase (GAD-65). Psychiatric symptoms in this disorder are prominent anxiety, probably mediated by changes in the GABAergic system.

Treatment is symptomatic, and high-dose benzodiazepine like diazepam, up to 100 mg/day or clonazepam, up to 6 mg/day, is usually the first-line treatment. Other GABA-mediating drugs such as baclofen, vigabatrin, tiagabine, and valproic acid (the latter up to 3,000 mg/day in divided doses) may also be helpful. Levetiracetam or pregabalin may be used if symptoms persist. Immunotherapies have included rituximab and tacrolimus. For patients with refractory disease, intrathecal baclofen, intravenous immunoglobulin (IVIG), or plasmapheresis may be administered.

Neuroleptic Malignant Syndrome

Neuroleptic malignant syndrome (NMS) is usually conceived of as a syndrome of fever, and catatonic rigidity, with increased serum CPK levels in the setting of neuroleptic administration or its rapid withdrawal. The etiology of this disorder is obscure, and the lack of consistent case definitions has hampered the fuller study of this syndrome. The initial description of NMS referred to it as pallor and hyperthermia and isolated primary symptoms of fever, akinesia or stupor with hypertonicity, and pulmonary congestion. Since then, a number of diagnostic criteria have been proposed. In essence, the syndromic definitions stress the presence of rigidity, fever, and autonomic changes in the presence of neuroleptic exposure. Alteration in sensorium is also a common feature of several of these criteria. Conditions with clinical similarity to NMS include serotonin syndrome (SS), malignant hyperthermia, drug toxicity (e.g., lithium), heat stroke and related syndromes, and malignant catatonia. Treatment of NMS involves immediately discontinuing the offending agent, aggressive supportive care to manage and prevent complications, and pharmacologic therapy. Bromocriptine mesylate, a dopamine agonist, and dantrolene sodium, a muscle relaxant, are most commonly used. If NMS is due to the rapid dopaminergic medication withdrawal, prompt and rapid medication reinstitution may improve symptoms.

Serotonin Syndrome

SS occurs due to excessive serotonergic activity in the CNS and PNS, usually in the context of initiation or dose increase of a serotonergic agent. Although this may occur with a single agent, it is usually the result of a combination of medications, supplements, illicit drugs with serotonergic activity (e.g., LSD), or proserotonergic foods. The symptoms fall under three broad categories—altered mental status, neuromuscular hyperactivity, and autonomic instability. Altered mental status may range from hyperactivity, anxiety, restlessness, mood elevation and excitement to lethargy, confusion, disorientation, hallucinations, and coma. Neuromuscular hyperactivity can manifest as myoclonus, hypertonicity, ocular clonus, hyperreflexia, and tremors. Autonomic instability may include diaphoresis, tachycardia, hypertension, and in moderate or severe cases, hyperthermia. In severe cases, symptoms can progress to generalized tonic–clonic seizures, multiorgan failure, and coma. Although clinical criteria such as Sternbach criteria and Hunter Serotonin Toxicity Criteria exist, there is not much concordance among these criteria. A heightened awareness of signs and symptoms, obtaining a detailed medication history, and conducting a careful physical examination are crucial for diagnostic discrimination of SS and other similar syndromes such as NMS, malignant hyperthermia, and anticholinergic toxicity. Treatment of SS involves cessation of the offending agent, supportive therapy, and use of the serotonin antagonist cyproheptadine. Given the widespread use of serotonergic antidepressants and the large number of substances that have been implicated in SS, it is of paramount importance that clinicians thoroughly evaluate the indication for adding a serotonergic agent to a medication regimen that might already contain serotonergic agent(s) and monitor closely for signs and symptoms of SS in those who are at increased risk.

Fibromyalgia

Fibromyalgia is a syndrome characterized by chronic widespread pain localized in various muscle groups, accompanied by fatigue, sleep disturbances, mood changes, and cognitive deficits. It is thought of as a centralized pain state in that it involves the central amplification of peripheral nociceptive stimuli. It is more common in females in their 30s to 50s. The provisional diagnostic criteria for fibromyalgia published by the American College of Rheumatology in 2010 are shown in Table 2.12–3.

Table 2.12–3.
2010 Fibromyalgia Diagnostic Criteria

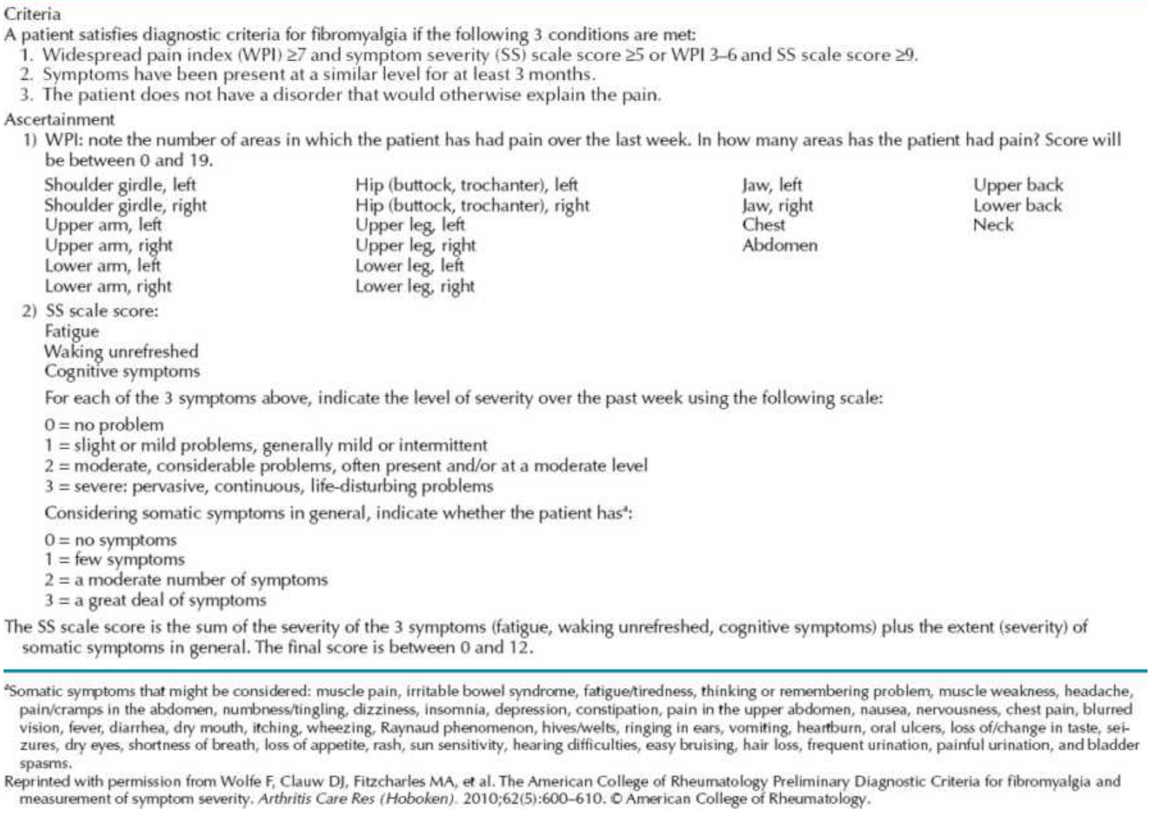

Criteria

A patient satisfies diagnostic criteria for fibromyalgia if the following 3 conditions are met:

1. Widespread pain index (WPI) ≥ 7 and symptom severity (SS) scale score ≥ 5 or WPI 3–6 and SS scale score ≥ 9.
2. Symptoms have been present at a similar level for at least 3 months.
3. The patient does not have a disorder that would otherwise explain the pain.

Ascertainment

1) WPI: note the number of areas in which the patient has had pain over the last week. In how many areas has the patient had pain? Score will be between 0 and 19.

Shoulder girdle, left
Shoulder girdle, right
Upper arm, left
Upper arm, right
Lower arm, left
Lower arm, right
Hip (buttock, trochanter), left
Hip (buttock, trochanter), right
Upper leg, left
Upper leg, right
Lower leg, left
Lower leg, right
Jaw, left
Jaw, right
Chest
Abdomen
Upper back
Lower back
Neck

2) SS scale score:

Fatigue
Waking unrefreshed
Cognitive symptoms

For each of the 3 symptoms above, indicate the level of severity over the past week using the following scale:

0 = no problem
1 = slight or mild problems, generally mild or intermittent
2 = moderate, considerable problems, often present and/or at a moderate level
3 = severe: pervasive, continuous, life-disturbing problems

Considering somatic symptoms in general, indicate whether the patient has[a]:

0 = no symptoms
1 = few symptoms
2 = a moderate number of symptoms
3 = a great deal of symptoms

The SS scale score is the sum of the severity of the 3 symptoms (fatigue, waking unrefreshed, cognitive symptoms) plus the extent (severity) of somatic symptoms in general. The final score is between 0 and 12.

[a] Somatic symptoms that might be considered: muscle pain, irritable bowel syndrome, fatigue/tiredness, thinking or remembering problem, muscle weakness, headache, pain/cramps in the abdomen, numbness/tingling, dizziness, insomnia, depression, constipation, pain in the upper abdomen, nausea, nervousness, chest pain, blurred vision, fever, diarrhea, dry mouth, itching, wheezing, Raynaud phenomenon, hives/welts, ringing in ears, vomiting, heartburn, oral ulcers, loss of/change in taste, seizures, dry eyes, shortness of breath, loss of appetite, rash, sun sensitivity, hearing difficulties, easy bruising, hair loss, frequent urination, painful urination, and bladder spasms.

Reprinted with permission from Wolfe F, Clauw DJ, Fitzcharles MA, et al. The American College of Rheumatology Preliminary Diagnostic Criteria for fibromyalgia and measurement of symptom severity. *Arthritis Care Res (Hoboken)*. 2010;62(5):600–610. © American College of Rheumatology.

There is significant overlap and comorbidity between symptoms of fibromyalgia and psychiatric disorders such as depression and anxiety. The shared putative mechanisms underlying depression, anxiety, and pain, namely dysregulation of serotonergic, noradrenergic, and dopaminergic neurotransmitter systems, may explain the co-occurrence of these disorders. This has important treatment implications—antidepressants that modulate the serotonin and norepinephrine neurotransmitter systems have been found to be effective in treating symptoms of fibromyalgia; furthermore, the secondary psychiatric syndromes can provide therapeutic targets for psychopharmacology. There is also significant comorbidity between patients with fibromyalgia and rheumatologic disorders, such as rheumatoid arthritis, systemic lupus, and others.

There is a significant disability associated with fibromyalgia, with 35% of patients receiving Social Security disability income in one study. Although treatment with multiple modalities may lead to improvements in symptoms and quality of life, their impact on functional recovery is limited. Duloxetine, milnacipran, and pregabalin have been approved by the U.S. FDA for the treatment of fibromyalgia. Duloxetine and milnacipran are both SNRIs that mediate the antinociceptive effect through modulation of serotonin and norepinephrine in the synaptic cleft, whereas pregabalin modulates neuronal calcium influx by binding to α_2-δ subunits of voltage-gated calcium channels. Duloxetine is also approved for other psychiatric and pain conditions, while milnacipran is only approved for the treatment of symptoms in fibromyalgia. Other pharmacologic agents that have been shown to improve symptoms include tricyclic compounds such as amitriptyline and nortriptyline, cyclobenzaprine, and selective serotonin reuptake inhibitors. The choice of medication should be based on targeting predominant symptoms and side-effect profiles. Nonpharmacologic treatment plans include graded exercise regimens, rehabilitation programs, cognitive behavioral therapy, and patient education.

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a multisystem disorder of unknown etiology. A more recent term used for this condition is systemic exertional intolerance disease (SEID). This complicated entity has an unclear relationship to other disorders such as fibromyalgia or myofascial pain syndrome. It is estimated to affect about 2 million people in the United States.

ME/CFS is characterized by extreme fatigue lasting more than 6 months, without a clear explanation as due to another underlying medical condition.

The fatigue worsens with mental or physical exertion and is not improved with rest. Besides unrefreshing sleep, there is prominent difficulty with memory, attention and concentration, and dizziness that worsens with moving from lying down or sitting to standing. Other physical examination findings may include sore throat, headaches, enlarged regional lymph nodes, and muscle or joint pain.

Despite extensive investigations, the etiology of ME/CFS is unknown. It may occur at any age, but most often in young to middle-aged adults. There is a gender disparity with ME/CFS diagnosed much more often in women than men, but this may be confounded by ascertainment bias.

Possibilities etiologies include infections such as Epstein–Barr virus, Human Herpes virus 6, undiagnosed HIV infection, COVID-19, Lyme disease, autoimmunity or immune system imbalance, hormonal imbalance, neoplasms or paraneoplastic processes, functional and psychological stress, or some combination thereof. Physical or emotional trauma has been implicated in some, but not all cases, in temporal proximity to syndrome onset. Genetic susceptibility is also suspected in some cases. Diagnostic testing can rule out some disorders, but there is no single confirmatory test. Treatment for ME/CFS focuses on improving symptoms. There are multiple published diagnostic criteria, and many criteria have been used to assess “recovery,” leading to more confusion for both healthcare providers and patients.

Treatment with cognitive behavioral therapy and graded exercise therapy have been the most widely used treatment modalities and provided some patient responses, although many patients do not achieve sustained benefit. The medical literature abounds with small studies of limited duration, leading to inconclusive results in diagnosis, treatment, and recovery.

ME/CFS is a debilitating disorder, and as the name suggests, it is a chronic disorder frequently lasting many years. There are frequent lifestyle restrictions due to fatigue and inability to work, resulting in the need for disability assistance. Social isolation and depression are common. Severe cases can lead to being bedbound with severe neurologic deficits. It is important to understand that despite the lack of a defined cause, suggesting that symptoms are of psychological origin (“it’s all in your head”) is fraught with complications. This simplistic formulation could be simply wrong, and patients often feel rejected by the medical system where they are seeking a diagnosis, explanation of symptoms, and treatment. Patients seeking “alternative” medical solutions such as holistic healing, homeopathic remedies, acupuncture, invasive treatment, multiple supplements, unproven therapies, medical marijuana, etc., are frequent and may help some obtain temporary relief but add a layer of complexity to an already complicated medical and psychosocial picture.

FUTURE DIRECTIONS

Advances in the understanding of the molecular and structural basis of many neuromuscular diseases and their relationship to neuropsychiatric symptoms have ushered in a new era in psychiatry. The natural history and prognosis of these disorders are generally well understood. Even in less well-understood conditions such as fibromyalgia, the focus should be on using available clinical evidence and building excellent physician–patient relationships, a key factor in providing improved neuropsychiatric outcomes. Further advances and insights into the integrated neurobiology of these disorders raise hope that psychiatrists will find themselves intimately involved in all phases of care targeting these challenging disorders.

Further Readings

Andrei G, Snoeck R. Advances and perspectives in the management of varicella-zoster virus infections. *Molecules*. 2021;26(4):1132.

*Arnold LM, Hudson JI, Keck PE, Auchenbach MB, Javaras KN, Hess EV. Comorbidity of fibromyalgia and psychiatric disorders. *J Clin Psychiatry*. 2006;67:1219.

*Bates D, Schultheis BC, Hanes MC, et al. A comprehensive algorithm for management of neuropathic pain. *Pain Med*. 2019;20(Suppl 1):S2–S12.

Bennion Callister J, Pickering-Brown SM. Pathogenesis/genetics of frontotemporal dementia and how it relates to ALS. *Exp Neurol*. 2014;262:84.

Crofford LJ. Psychological aspects of chronic musculoskeletal pain. *Best Pract Res Clin Rheumatol*. 2015;29:147.

Fortunato F, Rossi R, Falzarano MS, Ferlini A. Innovative therapeutic approaches for Duchenne muscular dystrophy. *J Clin Med*. 2021;10(4):820.

Galvez-Sánchez CM, Reyes Del Paso GA. Diagnostic criteria for fibromyalgia: critical review and future perspectives. *J Clin Med*. 2020;9(4):1219.

Guidon AC, Amato AA. COVID-19 and neuromuscular disorders. *Neurology*. 2020;94(22):959.

Hendriksen JGM, Vles JSH. Neuropsychiatric disorders in males with Duchenne muscular dystrophy: frequency rate of attention-deficit/hyperactive disorder (ADHD), autism spectrum disorder, and obsessive-compulsive disorder. *J Child Neurol*. 2008;23:477.

*Hulse EJ, Haslam JD, Emmett SR, Woolley T. Organophosphorus nerve agent poisoning: managing the poisoned patient. *Br J Anaesth*. 2019;123(4):457.

Ickmans K, Meeus M, De Kooning M, Lambrecht L, Pattyn N, Nijs J. Associations between cognitive performance and pain in chronic fatigue syndrome: comorbidity with fibromyalgia does matter. *Pain Physician*. 2015;18:E841.

Jiang L, Tang K, Levin M, et al. COVID-19 and multisystem inflammatory syndrome in children and adolescents. *Lancet Infect Dis*. 2020;20(11):e276–e288.

Kim DY, Lee JS, Son CG. Systematic review of primary outcome measurements for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) in randomized controlled trials. *J Clin Med*. 2020;9(11):3463.

Li Hi Shing S, Chipika RH, Finegan E, Murray D, Hardiman O, Bede P. Post-polio syndrome: more than just a lower motor neuron disease. *Front Neurol*. 2019;10:773.

*Macfarlane GJ, Kronisch C, Dean LE, et al. EULAR revised recommendations for the management of fibromyalgia. *Ann Rheum Dis*. 2017;76:318–328.

Meola G. Clinical aspects, molecular pathomechanisms and management of myotonic dystrophies. *Acta Myol*. 2013;32:154.

National Clinical Guideline Centre (UK). *Motor Neurone Disease: Assessment and Management*. National Institute for Health and Care Excellence (UK); 2016.

Orisakwe OE. The role of lead and cadmium in psychiatry. *N Am J Med Sci*. 2014;6:370.

*Rajabally YA, Seri S, Cavanna AE. Neuropsychiatric manifestations in inflammatory neuropathies: a systematic review. *Muscle Nerve*. 2016;54(1):1.

Simon LV, Keenaghan M. Serotonin Syndrome. January 23, 2021. In: *StatPearls [Internet]*. StatPearls Publishing; 2021.

Tormoehlen LM, Tekulve KJ, Nañagas KA. Hydrocarbon toxicity: a review. *Clin Toxicol (Phila)*. 2014;52(5):479.

*Wolfe F, Clauw DJ, Fitzcharles MA, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. *Arthritis Care Res (Hoboken)*. 2010;62:600.

Wolfe F, Walitt BT, Katz RS, Häuser W. Social security work disability and its predictors in patients with fibromyalgia. *Arthritis Care Res (Hoboken)*. 2014;66:1354.

2.13 Psychiatric Aspects of Child Neurology

Jeffrey E. Max, M.D., M.B.B.Ch., Sarah Mattson, Ph.D., Yvonne E. Vaucher, M.D., M.P.H.J., Sharon Nichols, Ph.D., and Mark P. Nespeca, M.D.

The topic “Psychiatric Aspects of Childhood Neurology” is large, and for justice to be done, a textbook is required. However, the approach taken in this chapter was to succinctly provide an updated review of five common and important pediatric neurologic conditions rather than review a much higher number of related topics in a cursory fashion. The topics (primary authors) chosen for this chapter are (1) Fetal Alcohol Spectrum Disorders (Sarah N. Mattson, Ph.D.); (2) Premature Birth and Low Birth Weight (Yvonne E. Vaucher, M.D.); (3) Human Immunodeficiency Virus (Sharon Nichols, Ph.D.); (4) Epilepsy (Mark P. Nespeca, M.D.); (5) Traumatic Brain Injury (Jeffrey E. Max, MBBCh). The focus of the chapter is on psychopathological sequelae, although mention will be made of associated problems in other domains of function.

FETAL ALCOHOL SPECTRUM DISORDERS

Children born to women who consume alcohol during pregnancy are at risk for physical, behavioral, and cognitive changes that can affect everyday function. These functional effects are collectively termed fetal alcohol spectrum disorders (FASD). The United States prevalence of FASD is between 1.1% and 5.0%, though estimates vary based on sampling techniques and study population. Criteria for diagnosis were first set forth by Jones and Smith in 1973 and have been updated by the Institute of Medicine and Hoyme. While multiple methods exist for clinical diagnosis, those set forth by Hoyme et al. will be described here. FASD includes five subdiagnoses: fetal alcohol syndrome (FAS), partial fetal alcohol syndrome (PFAS) with or without documented prenatal alcohol exposure, alcohol-related neurodevelopmental disorder (ARND), alcohol-related birth defects (ARBD), and neurobehavioral disorder associated with prenatal alcohol exposure (ND-PAE). ARBD is primarily a physical diagnosis and will not be described here.

Table 2.13–1.
Summary of Diagnostic Features That Comprise Fetal Alcohol Spectrum Disorders (FASD)

Domain	Feature	Requirement	Additional Detail
Prenatal alcohol exposure		Indication of documented prenatal alcohol exposure as evidenced by maternal report or a reliable collateral report	Can include drinking levels reported by the mother 3 months before pregnancy recognition
Face	Characteristic pattern of minor facial anomalies	Requires two or more facial features:	
	Palpebral fissures	≤10th centile	
	Thin vermilion border	Rank 4 or 5 on a racially normed lip/philtrum guide	
	Smooth philtrum	Rank 4 or 5 on a racially normed lip/philtrum guide	
Growth	Pre- or postnatal growth deficiency	Requires one or more indicators:	
	Height or weight	≤10th centile	
Brain	Deficient brain growth, abnormal morphogenesis, or neurophysiology	Requires one or more indicators: Head circumference ≤10th centile Structural brain abnormalities	

		Recurrent nonfebrile seizures	
Neurobehavioral impairment	Cognitive or behavioral impairment	Requires either cognitive or behavioral impairment	
	Cognitive impairment	Global impairment OR	GCA or IQ estimate ≥1.5 SD below mean
		One or more neurobehavioral domain ≥1.5 SD below mean	Executive functioning, specific learning impairment, memory impairment, or visual–spatial impairment
	Behavioral impairment (without cognitive impairment)	One or more behavioral domain ≥1.5 SD below mean	Self-regulation: mood or behavioral regulation impairment, attention deficit, or impulse control

GCA, general cognitive ability; IQ, intelligence quotient.

Diagnostic Criteria

The diagnoses under the umbrella of FASD share five critical domains. These are documentation of prenatal alcohol exposure, a characteristic pattern of facial anomalies, growth deficiency, deficient brain growth or abnormal morphogenesis, and neurobehavioral impairment. These domains are defined in Table 2.13–1. The diagnoses each require a different configuration of these domains, as indicated in Table 2.13–2. At the severe end of the spectrum, heavy prenatal alcohol exposure can cause characteristic facial features, resulting in a diagnosis of FAS. A diagnosis of FAS requires four diagnostic domains. Because of the specificity of this combination, documentation of confirmed prenatal alcohol exposure is not required. A diagnosis of partial FAS (PFAS) can be made with or without documentation of prenatal alcohol exposure. When alcohol exposure is confirmed, facial anomalies and neurobehavioral impairment are required. When alcohol exposure is not confirmed, impairment in either growth or brain growth is additionally required. A diagnosis of ARND is made in the absence of physical indicators (face and growth) and requires confirmed prenatal alcohol exposure and neurobehavioral impairment. Thus, ARND is intended to capture the broader population of children affected by prenatal alcohol exposure.

Table 2.13–2.
Summary of Criteria for Disorders Included in FASD

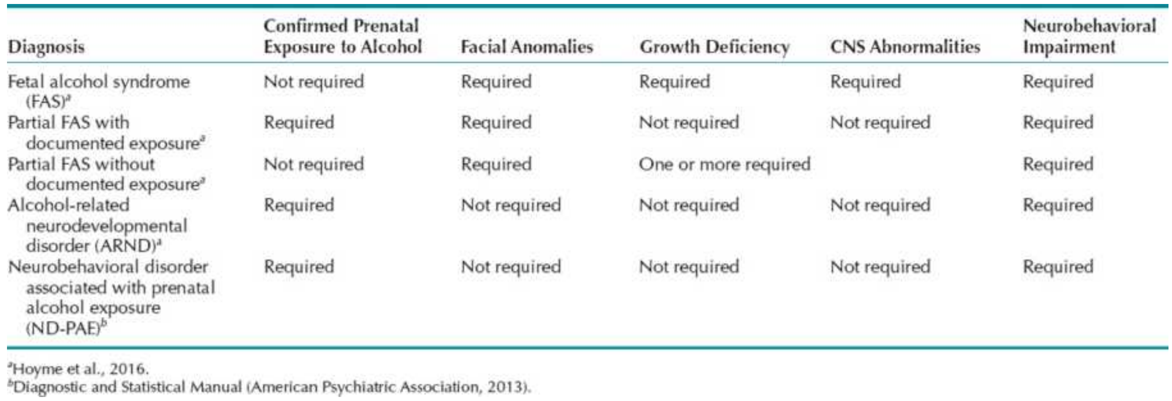

Diagnosis	Confirmed Prenatal Exposure to Alcohol	Facial Anomalies	Growth Deficiency	CNS Abnormalities	Neurobehavioral Impairment
Fetal alcohol syndrome (FAS)[a]	Not required	Required	Required	Required	Required
Partial FAS with documented exposure[a]	Required	Required	Not required	Not required	Required
Partial FAS without documented exposure[a]	Not required	Required	One or more required		Required
Alcohol-related neurodevelopmental disorder (ARND)[a]	Required	Not required	Not required	Not required	Required
Neurobehavioral disorder associated with prenatal alcohol exposure (ND-PAE)[b]	Required	Not required	Not required	Not required	Required

[a] Hoyme et al., 2016.
[b] Diagnostic and Statistical Manual (American Psychiatric Association, 2013).

ND-PAE has been proposed as an addition to those specified by Hoyme et al. The most recent revision of the Diagnostic and Statistical Manual (DSM-5) includes ND-PAE in the appendix under Conditions for Further Study. This proposed diagnosis reflects the importance of the neurobehavioral difficulties caused by prenatal alcohol exposure and addresses the critical need to better identify children affected by prenatal alcohol exposure. The proposed criteria for ND-PAE are listed in Table 2.13–3. ND-PAE fills a diagnostic gap by providing specific criteria for the neurobehavioral and cognitive effects seen in children affected by prenatal alcohol exposure. The specification of these criteria enables a clinical diagnosis for these children, which can improve identification, aid in treatment planning, as well as provide an etiologic context for behavioral concerns. Importantly the diagnosis of ND-PAE is independent of the diagnosis of FAS (i.e., it can occur in exposed children with or without the FAS).

Table 2.13–3.
Criteria for Neurobehavioral Disorder Associated with Prenatal Alcohol Exposure (ND-PAE)

Criterion	Requirement
Prenatal Alcohol Exposure	More than minimal prenatal alcohol exposure required
Neurocognitive impairment	**One or More Required** Impairment in global intellectual functioning Impairment in executive functioning Impairment in learning Impairment in memory Impairment in visual–spatial reasoning
Self-Regulation Impairment	**One or More Required** Impairment in mood or behavioral regulation Impaired attention Impairment in impulse control
Adaptive Functioning	**Two or More Required**[a] Impairment in communication Impairment in social interactions and communication Impairment in daily living skills Impairment in motor skills
Developmental Timing	Onset of symptoms in childhood required

[a] Two or more adaptive function difficulties required for diagnosis, but see also Kable and Coles, 2018.

Behavioral Findings

Disordered behavior is characteristic of FASD and may represent the most clinically relevant deficit. Parent reports of behavioral disturbance are pervasive and severe. Many studies indicate the presence of behavioral problems in children with FASD, including issues with impulsivity, rule-breaking, anxiety, and social interactions, which persist even when compared to IQ-matched controls. Adaptive behavior deficits have also been noted across all domains of adaptive function: communication, socialization, and daily living skills. In particular, deficits in socialization and communication appear to worsen with age. As a result, behavioral problems in childhood may translate into difficulties in adulthood, including legal problems, substance abuse, psychiatric disorders, and elevated suicide risk.

Cognitive Findings

The cognitive effects of prenatal alcohol exposure are well documented. Deficits occur in multiple domains, including general intellectual function, verbal learning, executive function, language, visual–spatial, attention, and academic function. Children with FASD have a broad range of IQ scores, ranging from severely intellectually deficient to average, although the majority of children do not have IQ scores below 70. The average IQ score for children with FAS is around 72, while that for children with heavy prenatal alcohol exposure without FAS is somewhat higher at around 80, although there is considerable variability in both groups.

In addition to overall intellectual performance, children with FASD also display a broad array of neuropsychological and cognitive deficits. Verbal learning and memory are consistently impaired in this population, with deficits occurring predominantly in encoding with relatively intact retention. Repeated exposure and implicit strategy aid in improving performance on learning and memory tasks. Attention is a hallmark deficit of FASD. Children in this population typically display deficits in vigilance, reaction time, and information processing across multiple tasks, although visual attention may be more impacted than auditory. Deficits in attention are supported by both laboratory measures (e.g., continuous performance testing) and parent reports. Other aspects of executive function are also impacted in children with FASD. This domain has been particularly well-studied, and deficits occur in planning, working memory, response inhibition, set-shifting, concept formation, and verbal and nonverbal fluency. Deficits in executive function appear especially useful in the identification of alcohol-affected children.

Academic achievement is one domain that requires further study. Increased school failure has been documented, and limited research indicates deficits in reading, spelling, and mathematic achievement. Two studies support deficits across these three domains and provide additional details on underlying mechanisms. Results indicated that deficits in reading and spelling may relate to working memory deficits, and deficits in mathematics may relate to spatial processing deficits.

Comorbidity

Children with FASD are often diagnosed with psychiatric disorders and have elevated rates of concurrent psychiatric disorders, especially in younger children and boys. Across multiple studies, the most common psychiatric diagnosis made in children with FASD is attention-deficit/hyperactivity disorder (ADHD); the prevalence of ADHD in this population ranges from 44% to 96%. Other than ADHD, the most common conditions are oppositional defiant disorder (ODD), conduct disorder, depressive and anxiety disorders.

Because ADHD is so common in children with FASD, several studies have compared children with FASD to children with idiopathic ADHD (i.e., without FASD or prenatal alcohol exposure). These studies have examined adaptive behavior, psychopathology, and neuropsychological functioning (e.g., executive function, verbal learning and memory, and attention). Although some similarities between FASD and ADHD have been noted, e.g., performance on the Wisconsin Card Sorting Test and parent reports of attention, more studies show differences between these clinical groups. For example, group differences have been noted in overall IQ, verbal fluency, trail making, verbal encoding, attention, and adaptive behavior; generally, children with FASD perform worse than children with idiopathic ADHD. On measures of attention, children with FASD show deficits in shifting between attentional sets or types of stimuli and encoding information, while those with ADHD have more difficulty focusing and sustaining attention, although differences between groups may also relate to the timing of stimuli and regulation of activation. On parent reports of adaptive function, children with FASD, but not idiopathic ADHD, show arrested development of socialization and communication skills, failing to improve with age. In one study, even though IQ and adaptive function profiles were similar, children with FASD and ADHD had worse cognitive functioning than those with idiopathic ADHD. Thus, overall, it appears that the cognitive-behavioral profiles of FASD show more impairment than what is seen in ADHD.

Another important question is whether the presence of ADHD in many children with FASD affects behavioral or neuropsychological outcomes. That is, do children with FASD and ADHD have greater impairment than either condition alone. Three studies compared children with FASD with and without ADHD, children with idiopathic ADHD, and children with neither ADHD nor prenatal alcohol exposure. Both studies used a 2 (exposure: yes/no) × 2 (ADHD: yes/no) factorial design. In the first study, parent reports of psychopathology and problem behaviors were compared. Results suggested that the presence of ADHD in children with FASD elevated rates of conduct disorder and externalizing behaviors. A second study of adaptive behavior, using the same design, indicated that children with both ADHD and FASD had the greatest impairment in the communication domain of adaptive behavior, while the effects in socialization and daily living skills suggested independent effects of both ADHD and FASD. In contrast, the third study of this type examined a broad neuropsychological test battery and did not support an additive or synergistic effect of FASD and ADHD on neuropsychological functioning. Instead, neuropsychological performance was similar in the alcohol-exposed groups, regardless of ADHD diagnosis. An additional study documented similar results with children with FASD or FASD and ADHD demonstrating greater cognitive deficits than children with ADHD alone.

Summary

Research conducted over nearly 50 years has demonstrated that alcohol is a potent physical and behavioral teratogen. While some children will not be affected by the physical manifestations of this exposure, all children with histories of prenatal alcohol exposure are at risk for cognitive and behavioral impairment at great personal and societal cost. Debate continues over whether any safe amounts or periods of alcohol exposure exist, while empirical data clearly show that the effects of this exposure can be devastating. Additional research is needed on long-term outcomes for individuals with FASD. Limited research suggests continued negative outcomes, while some anecdotal reports suggest that more positive outcomes are possible.

PREMATURE BIRTH AND LOW BIRTH WEIGHT

Ten percent (383,061) of U.S. children in 2019 were born prematurely between 23 and 37 weeks gestation, of whom 52,330 (1.38%) were very low–birth-weight (VLBW), weighing less than 1,500 g at birth, and 24,758 (0.66%) were extremely premature (EP), born before 28 weeks gestation (Table 2.13–4).

More than 90% of these premature newborns, including the majority of EP children born at 24 weeks gestation, now survive as the result of remarkable medical advances in neonatal care over the past 30 years. However, those born prematurely are at increased risk for adverse neurologic, cognitive, neurosensory, behavioral, and psychiatric outcomes (Table 2.13–5). The risk is multifactorial, influenced by numerous antenatal, intrapartum, and postnatal factors. Major independent determinants of short-term adverse outcomes include gestational age, birth weight, gender, intrauterine growth restriction, severity and type of neonatal illness, and neonatal brain injury. Barring significant brain injury, socioeconomic factors and maternal education become the most important predictors of long-term neurodevelopmental outcomes.

Despite significant advances in medical care and dramatically improved survival over the past 60 years, the incidence of adverse neurologic, cognitive, behavioral, and psychiatric outcomes, with the possible exception of decreased clinically significant cerebral palsy in the most immature, has remained essentially unchanged, raising the question of whether being born preterm independently influences normal brain development. Preterm infants are born into a dramatically different vestibular, motor, and neurosensory extra-uterine environment. The extent to which this alters subsequent brain development is unknown. In addition, all premature organ systems are physiologically immature and vulnerable to injury after birth. The resulting medical problems, such as chronic lung disease, patent ductus arteriosus, sepsis, necrotizing enterocolitis, intracranial hemorrhage, white matter injury (WMI), and retinopathy of prematurity, are each associated with adverse neurodevelopmental outcomes. At the time of their child's preterm birth, parents are primarily concerned about survival and the risk of severe motor or cognitive impairment. The latter is apparent within the first 24 months after birth. However, the more frequent behavioral, emotional, psychiatric, and educational problems are usually not evident until later in childhood.

Table 2.13–4.
Common Nomenclature for Preterm Infants

Weight	
Low–birth-weight (LBW)	<2,500 g
Very low–birth-weight (VLBW)	<1,500 g
Extremely low–birth-weight (ELBW)	<1,000 g
Gestational age	
Late preterm	$34^{0/7}$ to $36^{6/7}$ weeks
Moderate preterm	$32^{0/7}$ to $33^{6/7}$ weeks
Very preterm (VP)	$28^{0/7}$ to $31^{6/7}$ weeks
Extremely preterm (EP)	$<28^{0/7}$ weeks

Table 2.13–5.
Adverse Neurodevelopmental and Psychiatric Outcomes of VLBW Children

Cerebral palsy
Developmental coordination disorder
Minor neurologic dysfunction
Visual impairment
Hearing impairment
Cognitive impairment
Speech and language delay
Perceptual motor dysfunction
Low academic achievement
Learning disability
Poor social interaction
Attention-deficit/hyperactivity disorder (ADHD)
Anxiety
Depression
Autism spectrum disorder (ASD)

Brain Development

Preterm infants are born from the end of the 2nd through the 3rd trimester when the brain is undergoing very rapid growth and organization. Between 25 weeks and term, the cerebral cortical volume increases fourfold. At 35 weeks gestation, the brain is one-third smaller than it will be at term. Neurons proliferate primarily in the 1st trimester and migrate radially outward from their origin in the subependymal germinal matrix beneath the floor of the lateral ventricles to the cerebral cortex during the 2nd trimester. A smaller wave of neuronal proliferation and migration occurs from 20 to 27 weeks gestation, overlapping the time when the most immature preterm infants are critically ill and receive medical care in neonatal intensive care units. The 3rd trimester is characterized by active neuronal organization and exponentially increasing synaptic connectivity, which is dependent upon complex interrelationships between structural relationships, functional activity, sensory input, growth factors, hormonal and neurotransmitter activity. This is also a critical period, between 24 and 32 weeks gestation when transient subplate neurons are in place just beneath the cerebral cortex, waiting to connect with and guide late ascending neurons from the thalamus and commissural axons to the cortex. By 36 weeks gestation, these thalamocortical axons, callosal, and corticocortical association afferents have reached their ultimate destination in the cortex.

Neuronal survival and function are dependent upon the glia, the most common cell type in the brain. Glia determine and maintain the structural and functional integrity of the neurons, mediate the inflammatory response, and facilitate neuronal communication. Specifically, astrocytes transfer nutrients from capillaries to neurons, modulate the strength of synaptic transmission, and participate in synaptic repair; oligodendrocytes are responsible for axonal myelination, and microglia mediate the inflammatory response. Glia creates the topographic boundaries and maintains the structure of the white matter projection, commissural, and association axon tracts. Immature, preoligodendrocytes predominate from 23 to 32 weeks gestation and are extremely vulnerable to inflammatory, excitotoxic, and free radical injury due to hypoxemia, ischemia, hemorrhage, cytokines, and the toxic by-products of neuronal injury, all of which may result in acute glial cell injury or death as well as later apoptosis, delayed or abnormal maturation, and impaired function. Lack of glial support and hypomyelination results in neuronal injury, loss, and dysfunction. A reciprocal relationship between neuronal and glial injury has been proposed by Volpe, whereby injury to the glia leads secondarily to injury to the neuron and vice versa. He proposes that this “encephalopathy of prematurity,” a combination of primary destructive disease and secondary trophic and maturational disturbances, is responsible for the adverse neurodevelopmental outcome.

The cerebellum also undergoes very active development between 24 to 40 weeks gestation, increasing fivefold in volume and 30-fold in surface area, with peak cerebellar growth, differentiation, and neuronal migration occurring from 30 to 40 weeks gestation. Hemorrhagic injury, arising in the cerebellar germinal matrix, is more often unilateral and asymmetric, whereas ischemic infarction is more likely to be bilateral. The sequence of events is similar to that of cerebral injury-acute destruction, resulting in focal or diffuse volume loss with impaired neuronal development and hypomyelination. Trophic interaction (crossed diaschisis) between the contralateral cerebellar and cerebral hemispheres is such that injury to one is associated with decreased volume in the other, which may explain why cerebellar injury is associated with a wide range of cognitive, language, and motor deficits.

Preterm Brain Injury

Intracranial or intraventricular hemorrhage (IVH) currently occurs in approximately 25% of all VLBW infants. The widespread use of antenatal steroids to accelerate lung maturity has had the additional benefit of reducing the incidence of intracranial hemorrhage by half. Hemorrhage occurs within the first few days after birth, originating from the fragile capillaries in the subependymal germinal matrix just under the floor of the lateral ventricles at the caudo-thalamic groove. IVH is most often unilateral and, when bilateral, usually asymmetric in size. The mildest form (Grade 1 IVH) is contained within the subependymal germinal matrix (Fig. 2.13–1). The hemorrhage may rupture through the floor into the lateral ventricle (Grade 2 IVH) and may become large enough to distend the lateral ventricles with blood (Grade 3 IVH). The most severe form of IVH occurs when a large subependymal hemorrhage occludes the terminal vein draining the adjacent periventricular white matter resulting in intraparenchymal venous infarction (Grade 4 IVH, intraparenchymal hemorrhage [IPH] or periventricular hemorrhagic infarction [PVHI]). The associated white matter destruction may evolve into many small cysts, a large porencephalic cyst, or into an irregular outpouching of the ventricle into the adjacent area of white matter volume loss (Fig. 2.13–2). Blood within the ventricular system, especially with Grade 3 IVH, may result in posthemorrhagic hydrocephalus.

The majority (~85%) of hemorrhages are in Grades 1 or 2. Approximately 10% are Grade 3, and 5% are Grade 4 IVH/IPH. As expected, severe IVH/IPH (Grades 3 and 4) occurs most frequently in the smallest and most immature infants, affecting about 20% of those born at 23 to 24 weeks gestation. Fortunately, Grades 1 to 2 IVH are not associated with an increased risk of motor impairment. For infants with Grades 3 and 4 IVH/IPH, however, the risk of cerebral palsy is much higher–35% and 80%, respectively. Neurologic sequelae of IVH are primarily due to associated white matter damage, although free iron may also contribute to cellular injury. Almost all hemorrhages, except the most massive, are clinically silent, being recognized only by cranial imaging. Except for those in the posterior fossa, intracranial hemorrhages are easily visualized by cranial ultrasound. It was not until the advent of MRI that the incidence of cerebellar hemorrhage, occurring in up to 15% of VLBW infants, was appreciated.

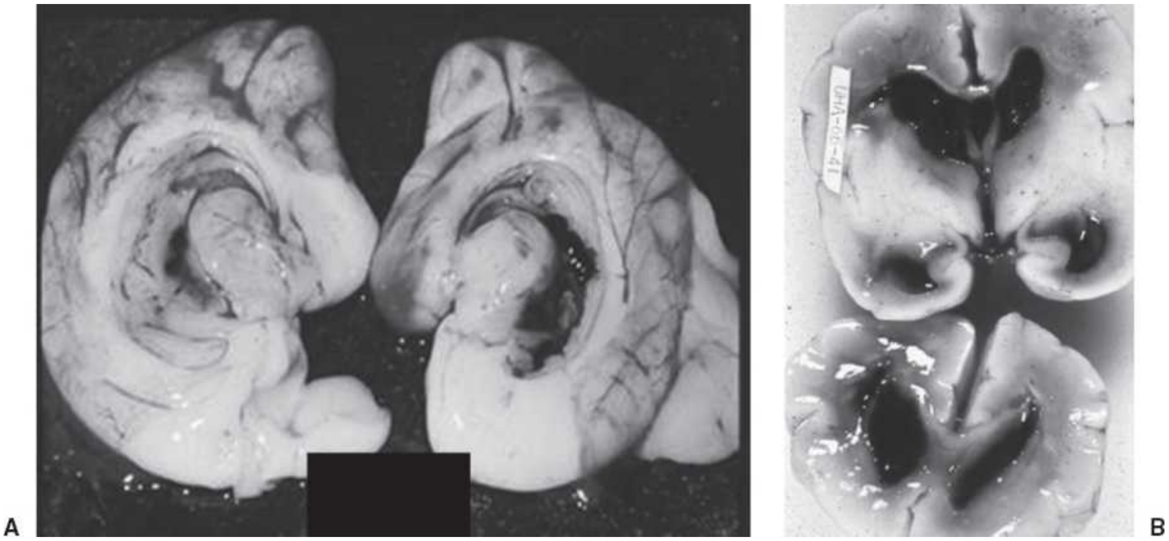

FIGURE 2.13–1. Types of intraventricular hemorrhage (IVH). **A:** Grade 1 IVH **(left)**, Grade 2 IVH **(right)**. **B:** Grade 3 IVH **(right and left)**; Grade 4 IVH-Periventricular Intraparenchymal infarction **(left).**

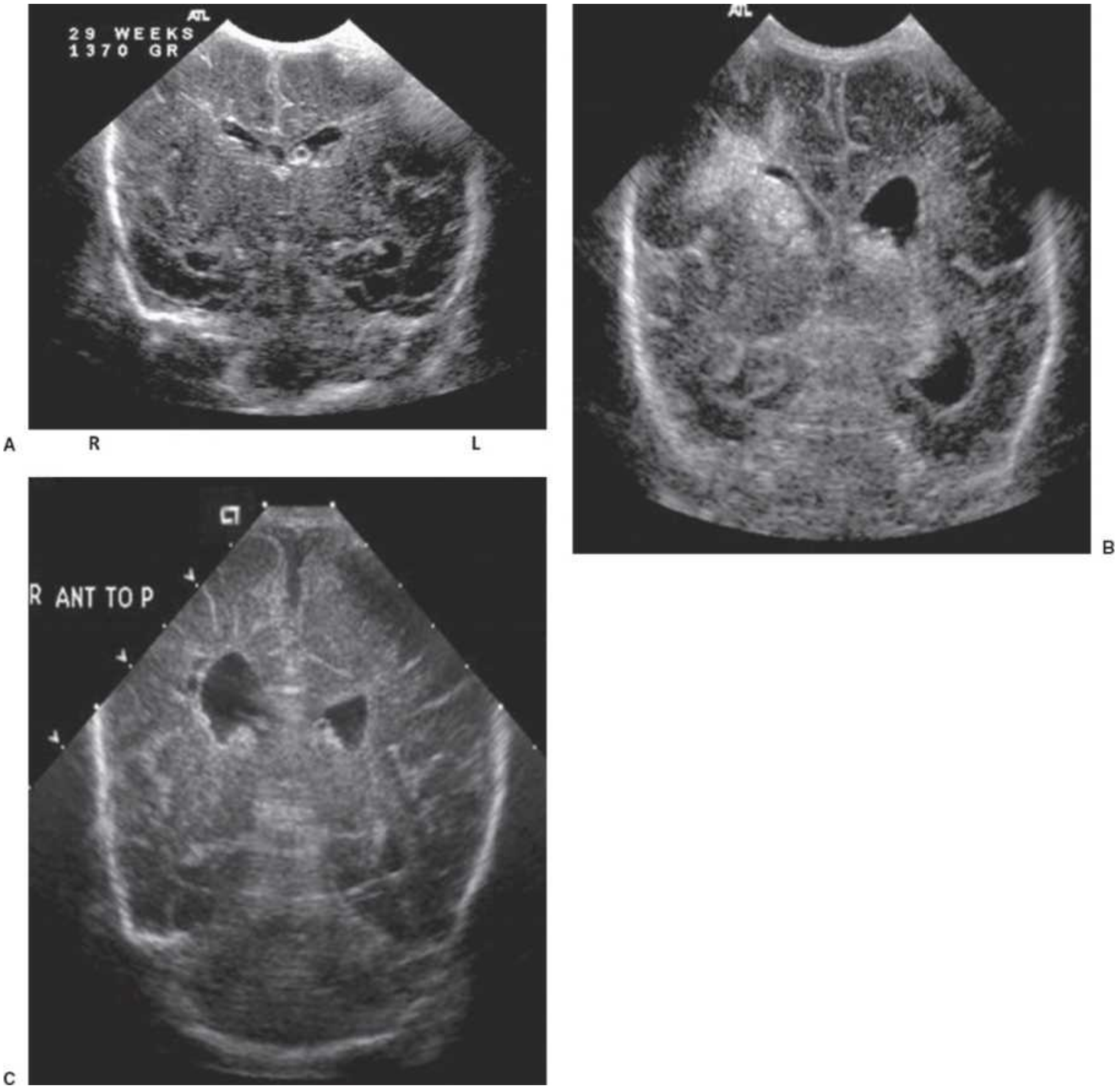

FIGURE 2.13–2. Evolving periventricular hemorraghic infarction (IPH/PVHI) Coronal views. **A:** DOL (day of life) 1: Small left Grade 1 subependymal hemorrhage (SEH); normal ventricular size. **B:** DOL 6: Large right Grade 1 subependymal hemorrhage (SEH), small left Grade 1 subependymal hemorrhage IVH, bilateral Grades 2–3 IVH with ventricular dilatation, right intraparenchymal hemorrhagic infarction (IPH/PVHI). **C:** DOL 55: Right parenchymal cysts with right ventricular outpouching into area of white matter loss.

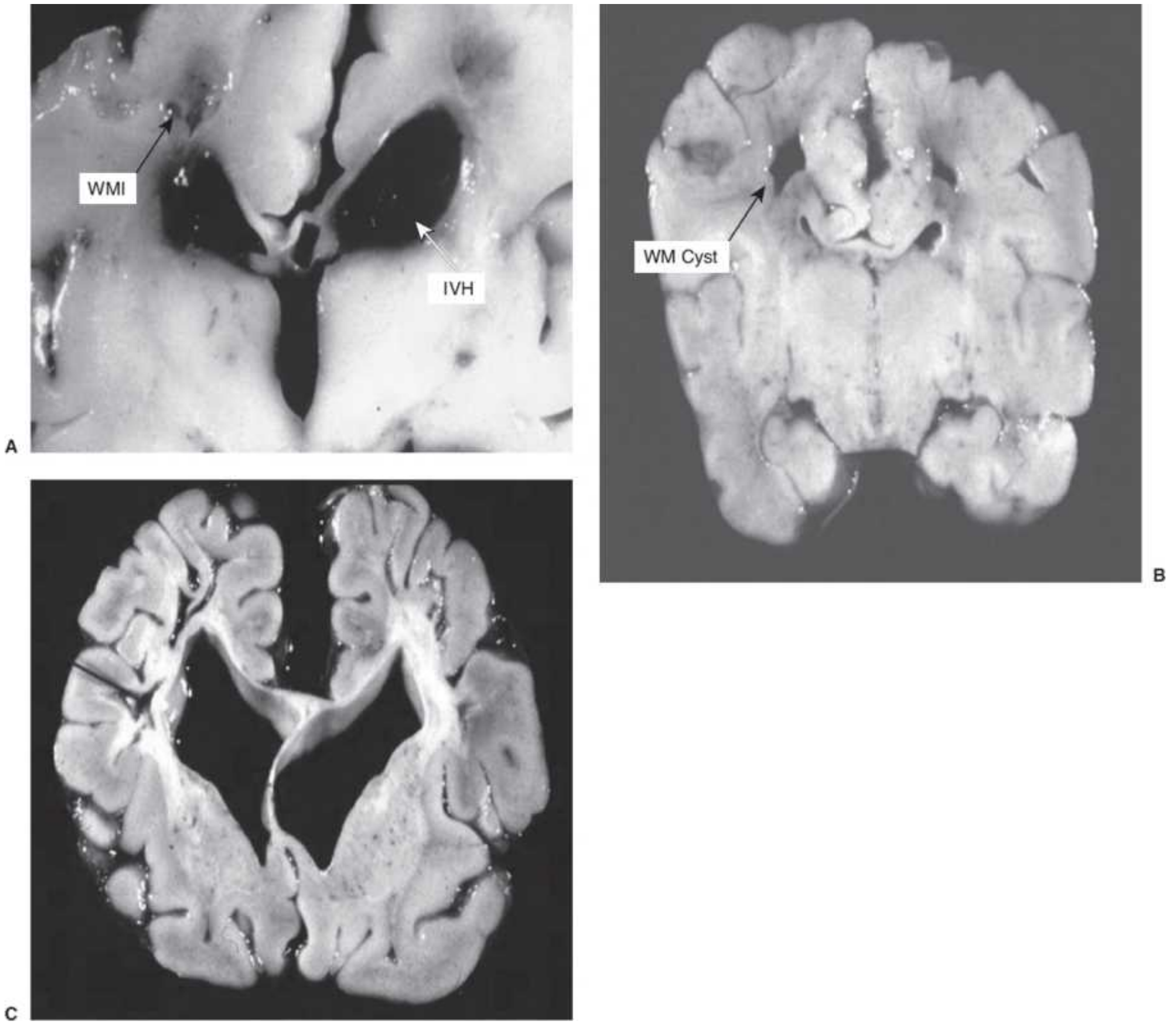

FIGURE 2.13–3. White matter injury (WMI)/periventricular leukomalacia (PVL) Coronal sections. **A:** Acute, focal, bilateral, white matter infarction (WMI); plus bilateral Grade 3 IVH. **B:** Subacute, focal, bilateral white matter loss with cystic degeneration. **C:** Chronic, diffuse, bilateral white matter loss with marked ventricular dilatation and brain atrophy.

Periventricular leukomalacia (PVL) or WMI occurs in the white matter without associated hemorrhage and is symmetric, bilateral, and clinically silent. Watershed areas with a decreased arterial blood supply in the region of the internal capsule near the angles of the lateral ventricles are especially vulnerable to ischemic injury. Acute ischemia results in focal necrosis and the formation of transient cysts; moderate ischemia results in more gradual injury due to maldevelopment and diffuse apoptosis with varying degrees of progressive brain atrophy due to glial and axonal loss associated with low pressure, ventricular dilatation into the areas of volume loss (Fig. 2.13–3). Macroscopic cyst degeneration, visible on cranial ultrasound, occurs in 5% of VLBW infants. However, microscopic cystic degeneration is much more common, affecting about 50% of VLBW infants. In one study, over 70% of infants born at ≤30 weeks gestation had abnormalities in the white matter detected by MRI at term equivalent gestation. Whereas focal necrosis with loss of all cellular elements is more likely to result in spastic cerebral palsy due to the destruction of the motor tracts crossing through the area of acute injury, diffuse white matter loss is associated with more generalized cognitive and behavioral deficits. Signs of white matter and associated axonal loss include thinning of the corpus callosum, delayed myelination, immature gyral development, enlarged subarachnoid space, and volume loss in the cortical and deep grey matter, peristriatal regions, and the temporal lobes.

A unifying, common path for brain injury may be antenatal, intrapartum, or postnatal hypoxia, ischemia, or inflammation, which in turn results in brain, especially white matter, injury in the physiologically vulnerable, preterm group. Subsequent adverse neurodevelopmental outcomes, including cerebral palsy, epilepsy, and cognitive impairment, are all thought to relate primarily to the adverse effects of WMI on both glial and neuronal structure, maturation, and function.

Role of Brain Imaging

Detection of neonatal brain injury after preterm birth relies on neonatal cranial ultrasound and MRI imaging. Although both modalities comparably identify severe cerebral brain injury, neonatal MRI also detects cerebellar abnormalities as well as more subtle signs of brain injury. Neonatal imaging abnormalities, including moderate to severe intracranial hemorrhage, PVL, and ventriculomegaly, are highly predictive of cognitive, motor, and behavioral outcomes at 2 years of age in very preterm (VP) children. By 10 years of age, however, moderate-severe neonatal brain injury predicts only poorer cognitive outcome. There are no specific neonatal anatomical markers for later behavioral, socio-emotional, and psychiatric problems. Despite advances in neonatal neuroimaging, it is not possible to accurately predict the long-term neurodevelopmental outcome for an individual child.

In later childhood, structural and functional MRI studies demonstrate substantial differences in regional brain volumes and neural connectivity, including the smaller hippocampal, thalamic, caudate, corpus callosum, and cerebellar volumes, as well as thinner cortices in infants, children, and adolescents born preterm. Disorganization of tracts in the internal capsule and posterior corpus callosum reflects widespread alterations in neural connectivity. Functional MRI DTI studies demonstrate that preterm children utilize novel neural pathways to complete tasks suggesting the development of alternative networks to compensate for abnormal brain development or injury.

Neurodevelopmental Outcome

The risk of adverse neurodevelopmental outcome increases progressively with increasing prematurity, with the lowest risk for those born nearest term at 36 weeks gestation and the highest risk for those born at the margins of viability at 22 to 23 weeks gestation. Thus far, overall individual outcomes during infancy and childhood appear to be longitudinally stable.

The best outcome studies are prospective and comprise large, population or geographically based cohorts followed longitudinally from infancy through adulthood. Comparison of outcomes with normal birth-weight (NBW), term-born, control children is critical. Most such studies have been performed in high-resource countries where children can be tracked easily within a national healthcare system over long periods of time (e.g., Australia, Canada, Scandinavia, and Great Britain). However, reports from single centers, cross-sectional and smaller samples are remarkably congruent with the results from larger, longitudinally followed cohorts.

The majority of neurodevelopmental outcome studies have focused on the highest risk preterm infants-either VLBW (<1,500 g), ELBW (<1,000 g), VP (<32 weeks gestation), or EP (<28 weeks gestation) in very early childhood at 2 years of age. Approximately one-third of EP children born at 25 weeks gestation demonstrate moderate-severe neurodevelopmental impairment at this time. Multinational outcome studies for EP children in midchildhood, now becoming available, report similar rates of impairment. Recently more attention has been paid to the neurodevelopmental outcomes of larger, more mature, moderate, and late preterm children born at 32 to 36 weeks gestation whose risk of adverse cognitive and behavioral outcomes, while substantially lower than that of more immature preterm infants, is significantly higher than that of term infants. Due to their much greater numbers representing over three-quarters of all preterm newborns, moderate to late preterm children contribute disproportionately to the burden of adverse neurodevelopmental outcomes associated with preterm birth.

A myriad of demographic, antenatal, intrapartum, and postnatal variables have been associated with adverse neurodevelopmental and psychiatric outcomes. Many of these factors, such as SES, parental stress, maternal depression, and infection, are also associated with an increased risk of preterm birth. Epigenetic markers of prenatal and postnatal environmental exposures also likely influence brain development and neurodevelopmental outcome. Whether there will be substantive differences in long-term outcomes based on birth before and after the currently widespread use of antenatal steroids, intrapartum neurologic protection, and postnatal surfactant beginning in the early 1990s remains to be seen. Whereas studies of childhood and adolescent outcomes reflect current obstetrical and neonatal practice, most adult outcome studies are derived from the earlier time period. Of note, although antenatal steroids decrease the risk of severe illness in preterm newborns, a recent population-based study in Scandinavia found that whereas antenatal steroids decreased the risk of intellectual disability, the risk of mental and behavioral disorders in preterm children was increased.

Despite the physical, emotional, behavioral, and cognitive adversities facing many children and adults born prematurely, the majority self-report a satisfactory quality of life. Often self-perception is more positive than parent or teacher reports, particularly with regard to ADHD and depression. Whether this is a lack of self-awareness or reflects resilient coping mechanisms is not clear. The majority of preterm adults are able to compensate for their difficulties and go on to lead full, productive lives.

Cognitive Outcomes

Both the risk and magnitude of cognitive deficits are inversely correlated with gestational age. VLBW or VP children have a decrement of 0.8 SD on measures of IQ and an increased risk of moderate to severe cognitive deficits. Moderate and late preterm children born at ≥32 weeks gestation score an average of 8 points lower, VP children born from 28 to 31 weeks gestation score 11 points lower, and EP children born at <28 weeks gestation score 14 points lower—a decrease of approximately 1.5 IQ points/week if born before 33 weeks gestation. EP/ELBW are two to three times more likely to have impaired attention than term controls with deficits including delays in selective, sustained, encoding, divided, and shifting attention. Processing speed, the time required to interpret and respond to incoming stimuli, is also slower. Compared to term controls, VP/VLBW children are 2 to 3.5 times more likely to have deficits in all aspects of memory, including short-term working memory, long-term episodic (events) memory, and semantic (facts) memory. Impaired visual acuity and visual/perceptual processing problems are common. It is, therefore, not surprising that executive function is also often impaired. These deficits do not appear to improve over time. Some, such as difficulties related to attention and executive function, tend to worsen as intellectual challenges become increasingly complex throughout childhood and adolescence.

Academic performance, especially for EP/ELBW children, is moderately to severely affected and due in part to specific learning disabilities in mathematics, reading, and spelling, often being worse than expected for a given IQ. In one study, compared with term controls, children born at <26 weeks gestation were more likely to have impairment at 11 years of age in reading (52% vs. 11%) and mathematics (70% vs. 14%). In another study, 65% of ELBW children at 8 to 9 years of age had at least one learning disability compared to 13% of those born at term. Although many VP/VLBW children may require some special educational resources at some point in their education, most VP/VLBW children are in mainstream classes and graduate from high school; many attend college and pursue postgraduate degrees.

Family Impact of Preterm Birth

The emotional effect of preterm birth, especially extremely preterm (EP) birth, on sibling and parental behavior and emotional well-being may be profound and pervasive, including post-traumatic stress disorder, anxiety, depression, family dysfunction, and divorce. Preterm birth is traumatic for parents, disrupting the cultural idealization of the perfect child, the competent mother, and the protective father. Physical separation and medicalized parenting interfere with normal parent–infant bonding. The most immature preterm infants are critically ill for many weeks, with multiple ups and downs involving difficult decisions by both the family and the medical team. EP children are often survivors of multiple gestations, adding the burden of grieving a fetal or neonatal loss to the family dynamic. Parents of preterm infants are more likely to report moderate to severe anxiety, depressive symptoms, and higher levels of parenting stress many years after their VP child’s birth. Maternal depression not only increases the risk of preterm delivery but also is associated with internalizing behavior, anxiety, depression, and low social competence in their offspring. Maternal and paternal emotional distress related to preterm birth persisting for many years may negatively affect family function, parental perceptions of their child, long-term parent/child relationships, and child development.

Psychopathologic Outcome

Behavioral and socio-emotional problems are three to eight times more common in the preterm child compared to their term counterparts. Preterm children are more likely to have persistent internalizing and less likely to have externalizing problems. The preterm behavioral phenotype has been described as being inattentive, anxious, and socially withdrawn. At 4 years of age, poor social competence in VP children was associated with birth at <28 weeks gestation, socioeconomic disadvantage, the severity of brain injury, and early exposure to high maternal anxiety and negative parenting. Poor peer relationships and bullying are reported more often by preterm children.

Studies of preterm young adults compared with their term-born peers have noted decreased risk-taking behavior, delinquency, and incarceration and an increased risk of depression, anxiety, and social difficulties, including introversion, autistic features, and neuroticism. The reasons for less high-risk behavior are unclear but may reflect a lack of interest or opportunity, decreased social interactions, or increased protection by the family. Unlike term-born adults, longitudinal studies of ELBW adults found no decrease over time in internalizing problems from midchildhood to young adulthood. Behavioral and psychiatric diagnoses in preterm children are generally stable through adolescence.

Being born preterm substantially increases the risk of psychopathology, especially anxiety, depression, ADHD, and Autism. As with other outcomes, the risk increases with decreasing gestational age, rising sharply for those born before 31 weeks gestation. The risks of autism spectrum disorder (ASD) and emotional and behavioral problems are each 10 times higher if born at 23 to 27 weeks compared to ≥37 weeks gestation. VP children are two to three times more likely than moderate to late preterm children to have nonaffective psychosis, depressive disorder, and bipolar disorder. From a Scandinavian National Registry database, the risk of hospitalization for any psychiatric diagnosis doubles below 33 weeks gestation. Psychiatric disorders are often comorbid with other preterm-related cognitive, motor, and neurosensory impairments making diagnosis and treatment more difficult.

In midchildhood, compared to term-born children, EP children are approximately 3 times as likely to have a psychiatric diagnosis, 4 times as likely to have ADHD, 2.6 times as likely to have emotional disorders, and 8 times as likely to have ASD. In a U.S.-based longitudinal study, extremely low gestational age newborns (ELGAN) at 10 years of age were much more likely to screen positive for internalizing disorders (i.e., anxiety, dysthymic disorders, and social phobias) and other psychiatric disorders (i.e., ASD, ADHD) compared to NBW controls. Overall, 29% screened positive for one to four psychiatric disorders and were four times more likely to be taking prescribed psychotropic medications. In this same cohort, ASD determined by validated diagnostic criteria was highest (15%) in the ELGANS born at 23 to 24 weeks gestation, decreasing to 3% in those born at 27 weeks gestation.

In a meta-analysis including five studies of LBW, preterm adolescents, one in four had a psychiatric diagnosis (OR 3.66). In another study of midadolescent VLBW children, whereas 11% had cerebral palsy and 18% had moderate intellectual deficiency, 46% had emotional or behavioral problems and 25% had psychiatric symptoms fulfilling at least 75% of diagnostic criteria.

For moderately preterm young adults, the prevalence of a psychiatric disorder at 19 years of age was higher than their full-term counterparts (30% vs. 19%). The primary disorders were mood, ADHD, and antisocial personality, with males and females being equally affected. The risk of having a psychiatric problem increased from 11 to 19 years. More than 50% of those with a disorder at 19 years had not had a psychiatric diagnosis when seen at 11 years of age.

Anxiety/Depression. In a recent longitudinal U.S. study of an EP cohort at 10 years of age, parent and teacher reports of anxiety and depression were associated with the increased degree of prematurity, public insurance, Hispanic ethnicity, obesity, and physical health. Compared to term controls, a multicultural meta-analysis of VLBW adult cohorts born between 1977 to 1988 reported more internalizing and avoidant personality problems and less externalizing and antisocial personality problems. In LBW (<2,500 g) preterm adolescents, the attributable risk for depression in female adolescents based on psychiatric assessment was 18% to 24% of LBW females versus 3% of NBW females. LBW adolescent females were also found to be less resilient than their NBW peers. Compared to NBW females, LBW females have more risk factors associated with depression: (i.e., 20% of LBW vs. 5% of NBW for one risk factor and 85% of LBW vs. 20% of NBW for two risk factors). In this study and others, the risk of depression was not increased in preterm male adolescents. Absolute rates of adolescent depression were slightly lower in an Australian population-based cohort where 15% of preterm LBW adolescent girls and 1% of preterm LBW adolescent males had stable depression in comparison to 2% of term females and 0.25% of term male adolescents. The OR for LBW adolescents developing a stable depressive disorder was 11.6 after adjustment for background confounders. Other studies indicate that moderate to late preterm young adults born at 32 to 36 weeks gestation are 2.7 times more likely, and VP young adults born before 32 weeks gestation are 7.4 times more likely to develop bipolar disorder.

Attention-Deficit/Hyperactivity Disorder. The prevalence of ADHD is 2 to 4 times greater in preterm children and occurs most often in the most premature children. The majority of preterm children are more likely to have the inattentive ADHD subtype rather than combined or hyperactive/impulsive subtypes. In one large population-based study, EP children had a high rate of generalized attention impairment at 11 years of age–77% were impaired in at least one area of attention, and 50% were impaired in 2 areas. In the same cohort seen at 17 years of age, 18% had clinical ADHD compared to 7% of the term adolescents. Unlike in term children, where there is a male predominance, preterm girls and boys are equally affected. Specific neonatal factors are weak predictors of attention impairment and ADHD although poor intrauterine growth may contribute to the risk of hyperactivity. In a large twin study, the lower–birth-weight twin scored higher on a hyperactivity scale. In preterm children with ADHD, neuroanatomical and functional abnormalities are present on MRI in specific structures and in the fronto–striatal–cerebellar circuitry with evidence of hyperactivation in some areas and decreased activation in others.

Autism/Autism Spectrum Disorder. The prevalence of autism in preterm children is two to three times the prevalence in the overall population. There is a gradual increase in the risk of ASD associated with decreased length of gestation, with the most immature and the smallest at the highest risk. Clinically, preterm and LBW children have predominately social and communication problems, with less stereotypic, repetitive movements and more motor and cognitive delays.

Although a high proportion (21% to 41%) of VP children fail the Modified Checklist for Autism in Toddlers (MCHAT) screener at 18 to 24 months of age, these results are strongly influenced by concomitant cognitive, motor, and neurosensory impairments. The percentage of positive screens is much lower following confirmatory interviews. In later childhood and adolescence, the prevalence of ASD fulfilling diagnostic criteria is 5% overall in LBW children and up to 8% in EP children. Although the prevalence of autism is four times higher in males, the risk is higher for preterm females born at less than 33 weeks gestation. Preterm females are also more likely to have comorbid conditions.

Many variables that have been associated with ASD occur more frequently with preterm delivery. See Table 2.13–6 for a list of these risk factors. Most of these variables are either primarily or secondarily associated with hypoxic/ischemic, inflammatory, or toxic brain injury. Both genetic and environmental components of autism have been demonstrated in twin studies which have high rates of concordance in identical but not dizygotic twins. When identical twins are discordant for ASD, the smaller twin, presumably with the less optimal intrauterine environment, is more likely to be affected. Autism is associated with neonatal cranial imaging abnormalities, including ventricular dilatation, intracranial hemorrhage, parenchymal white matter volume loss, isolated cerebellar injury, and hypoplasia of the cerebellar vermis.

Table 2.13–6.
Risk Factors Associated with Autistic Spectrum Disorder

Risk Factors		
Maternal/Antenatal	**Fetal/Intrapartum**	**Neonatal/Postnatal**
Gestational diabetes	Preterm delivery	Low Apgar score
Advanced maternal age	Fetal distress	Need for intensive care
Maternal hypertension	Breech presentation	Hypoxemia
Maternal antidepressants	Umbilical cord complications	Respiratory distress
Antenatal bleeding	Intrauterine growth restriction	Meconium aspiration
Antenatal infection	Twin gestation	Jaundice
	Birth order	Anemia
	Induced labor	Hemolytic disease of newborn
	Summer delivery	Lack of human milk feeding
		Congenital malformation

Psychotic Disorders. Given the low prevalence of nonaffective psychosis, little information is available. The few studies from mental health or large birth registries associate preterm birth, hypoxia, and related obstetrical complications (e.g., perinatal asphyxia), fetal undergrowth, and pre-eclampsia with schizophrenia. Moderate to late preterm young adults born at 32 to 36 weeks gestation are 1.6 times more likely, and VP young adults born before 32 weeks gestation are 2.5 times more likely to develop nonaffective psychosis. Brain abnormalities associated with schizophrenia include cerebellar hypoplasia, severe IVH, and ventricular dilatation.

In conclusion, psychopathology, including behavioral problems, depression, anxiety, autism, ADHD, and nonaffective psychosis, are all more prevalent in children, adolescents, and adults born prematurely. The risk increases steadily with increasing prematurity such that those born at less than 28 weeks gestation are at the highest risk. Abnormalities in brain structure and function incurred as a result of preterm birth have been implicated in adverse neurodevelopmental and psychiatric outcomes.

HUMAN IMMUNODEFICIENCY VIRUS IN CHILDREN AND ADOLESCENTS

Epidemiology

The human immunodeficiency virus type-1 (HIV) is a retrovirus acquired through exposure to infected bodily fluids, most commonly in adults via unprotected sexual contact or injection drug use; infection via blood products, a significant route of exposure early in the epidemic, has been reduced to negligible levels in the United States by donor screening and treatment of donated blood. For infants born to a woman with HIV, exposure can occur in utero, during birth, or via breastfeeding. There is regional variation in the most common route of infection and the population affected; in the United States, new infections are most likely among men who have sex with men (MSM). HIV primarily infects and destroys CD4$^+$ T-lymphocytes and, untreated results in progressive immunosuppression and vulnerability to opportunistic infections and neoplasms. Centers for Disease Control (CDC) surveillance classification of disease severity assigns categories according to age-specific CD4$^+$ T-lymphocyte count and whether an opportunistic infection has been diagnosed; Stage 3, or acquired immunodeficiency syndrome (AIDS), refers to CD4$^+$ count below 200 or diagnosis of a Stage 3-defining opportunistic infection. The development of effective antiretroviral therapy (ART), particularly combinations of ARTs from different drug classes, has resulted in a dramatic decrease in AIDS incidence and increased focus on more subtle effects of HIV, such as those in the brain; long-term toxicities associated with ART; interactions of HIV with other comorbidities such as substance use, depression, concomitant infections, and aging; and functional cure of HIV through the elimination of viral reservoirs. A discussion of virology and pathogenesis-related to HIV, as well as neuropsychiatric findings in adults, are presented in the chapter in this volume, “Neuropsychiatric Aspects of HIV Infection and AIDS.”

Over the more than 30 years of the HIV epidemic, children and adolescents have been significantly affected. Globally, an estimated 37.7 million people, including 1.7 million children under 15 years of age, were living with HIV in 2020, with 1.5 million new infections and 680,000 deaths from AIDS-related illnesses that year. Since the administration of ART during pregnancy and to the infant after birth was discovered to prevent most cases of mother-to-child HIV transmission, the incidence of new infant infections has dropped dramatically in high-resource environments such as the United States, with the majority of perinatal infections occurring in limited resource settings, particularly sub-Saharan Africa. However, in all settings, numbers of older perinatally infected children and youth have grown as effective ART has extended their lives; many of them are now adolescents or young adults and experiencing long-term neurodevelopmental, behavioral, and mental health issues associated with HIV as well as mental health and substance use problems common to their age group and/or demographic. Among individuals who acquire HIV infection postnatally, adolescents (age 15 to 24) have a relatively high rate of infection.

If taken appropriately, ART can suppress HIV to undetectable levels in the blood for most people across the lifespan, preventing immunosuppression with its associated health risks, some of which remain following immune reconstitution, and dramatically reducing the risk of transmitting HIV to others. Currently, recommendations for initiating ART include all individuals upon diagnosis, including infants at birth. In order to reap the benefits of ART and achieve viral suppression, however, an individual must first be identified as HIV infected, linked to and engaged in medical care, prescribed ART, and maintain high ART adherence (the so-called HIV treatment cascade). Youth aged 13 to 24 with HIV have lower rates of success at all stages of the treatment cascade than other age groups, and as a result, only an estimated 33% in the United States were virally suppressed in 2018, according to the CDC. Globally, access to ART continues to be lower for youth than for adults, and HIV is among the leading causes of death for adolescents.

Mental Health and Psychiatric Disorders

The interconnections of HIV with mental health and psychiatric disorders, including neurocognitive impairment, psychopathology, and substance use, are well established in the adult population but are less well understood among children and youth. The likelihood of sexual and injection drug use risk behaviors associated with HIV acquisition, as well as other infections, is already greater among youth by virtue of the age-typical tendency toward risk-taking and is increased in the presence of mental illness and substance use. Conversely, HIV infection increases the risk for mental health and substance use problems both directly in the case of neurocognitive decline and indirectly through the emotional and psychosocial impacts of having HIV. Furthermore, neurocognitive and psychiatric disorders complicate health and medication management for individuals with HIV, increasing morbidity and likelihood of HIV transmission, and the presence of multiple comorbidities challenges healthcare systems, particularly in lower resource settings. This suggests the important role of psychiatrists and other mental health professionals in preventing HIV infection as well as increasing the quality of life in those already infected. Addressing the gap in mental health care for individuals with HIV is a necessary component of reducing and ultimately eliminating the HIV epidemic.

Adolescents with Behaviorally Acquired HIV

Neurocognitive Impact of HIV. The neuropathogenesis, as well as neurocognitive and psychiatric sequelae of HIV in adults, are reviewed in detail in the chapter “Neuropsychiatric Aspects of HIV Infection and AIDS” and will be summarized only briefly here, along with differences between youth and adults. The CNS is vulnerable to HIV, and neuropsychiatric sequelae of HIV have been described since the beginning of the epidemic. The virus can enter the CNS early in infection via infected monocyte-derived macrophages and lymphocytes and establish a locally productive infection that may be relatively independent of other loci, representing a viral reservoir. The effects of HIV on neural functioning are largely indirect, through exposure to viral proteins, production of proinflammatory cytokines by infected or activated monocytes and other cells, and alteration of the blood–brain barrier rather than through direct infection of neurons. HIV can lead to a dynamic state of chronic inflammation and immune activation that impacts the CNS and other systems and may be influenced by other lifestyle factors such as substance abuse and may interact with neurodevelopment in perinatally infected children and possibly in adolescent infection. HIV-related immune activation and inflammation may become established early in infection and persist despite viral suppression using ARTs.

Before the advent of highly active ART, HIV-associated dementia occurred in 20% to 30% of adults with HIV by the time of death, typically in the context of significant immunosuppression, where it can still occur. In addition, opportunistic infections in the CNS, such as progressive multifocal leukoencephalopathy, can occur with immunosuppression.

Amongst severely immunocompromised individuals, delirium, coma, seizures, and intracranial hypertension can be seen. In the era of ART, the incidence of HIV dementia and other severe sequelae has declined dramatically; however, the prevalence of milder impairments continues to be high and is particularly likely in individuals who have experienced profound immunosuppression at some point despite later immune reconstitution (i.e., “legacy” effect). Preventing long-lasting effects of immunosuppression, as well as the establishment of inflammation, have led to the recommendation to initiate ART upon diagnosis with HIV, including in infants.

A “subcortical” presentation, with slowing, impaired retrieval from memory but more intact recognition, fine and gross motor impairment, apathy, and decrease in emotional expressivity, has been described among individuals with untreated HIV. With ART, impairments are more subtle and have been suggested to have a more “cortical” presentation, with executive function difficulties assuming more prominence. Several mechanisms have been proposed to explain the presence of impairments in virally suppressed or early-infected individuals, including inflammation and higher viral activity in the CNS than in the periphery. Although it is often assumed that behaviorally HIV-infected adolescents and young adults would follow the course typically seen among older adults, the few studies of this population suggest some differences in presentation. In particular, striking impairments in verbal and nonverbal learning, and lesser difficulties in other cognitive domains, have been observed among young adults with behaviorally acquired HIV. The presence of some associations with HIV disease severity and markers of inflammation and immune activation raises the question of subtle, HIV-associated neurocognitive function in this population, but associations with demographics and psychiatric and educational covariates suggest that other, possibly pre-existing causes for the impairments are contributory. Distinguishing subtle HIV-associated impairments from those related to comorbidities such as depression or substance use requires careful assessment. Regardless of their origin, however, neurocognitive impairments have been found to be associated with poorer daily functioning and medication adherence; monitoring is indicated in all individuals with HIV. This is particularly true since viral suppression through good medication adherence is an important tool for preventing HIV transmission. Because both reduced medication adherence and increased risk behaviors are more likely among adolescents, adherence interventions take on increased importance for this population. Unfortunately, efforts to develop CNS-specific therapies have been disappointing, and results of using ART with greater penetrance across the blood–brain barrier into the CNS have been mixed. Although cognitive impairments have shown modest improvement with ART and systemic viral suppression in some studies (however, findings for adolescents are less compelling), choosing regimens with less potential neurotoxicity is of particular relevance for this population facing many decades of therapy. Nevertheless, at this time, viral suppression with ART is considered the most effective prevention and treatment for HIV-associated neurocognitive disorders.

Mental Illness and Substance Use. Though prevalence estimates vary across settings and populations, in general, the literature demonstrates a higher prevalence of mental health disorders among adolescents and adults with HIV than among the general population, with rates up to 75% to 83% in some studies. Several classes of disorders are of particular concern due to their prevalence or potential to impact health outcomes or HIV transmission.

DEPRESSION. Depression is a common referral issue for mental health treatment for individuals with HIV, with reported prevalence rates two to three times higher than among the general population and a significant unmet need for treatment, particularly among minority youth. Assessment of depression is important in the clinical care of youth with HIV for several reasons, even in the absence of depression as the presenting complaint. The American Academy of Pediatrics notes the higher rates of depression and suicidal ideation among sexual minority (lesbian, gay, bisexual, transgender, and questioning) youth and the mental health treatment disparities resulting from discrimination and stigma related to sexual orientation and gender identity, and recommends screening for depression as part of a complete psychosocial history. Depressive symptoms can be a side effect of certain antiretroviral medications (e.g., efavirenz) and indicate the need for a change of medication regimen. Untreated depression can contribute to sexual and substance use risk behaviors and to poorer adherence and engagement in care, thus increasing morbidity and HIV transmission likelihood and suicide. Finally, cognitive symptoms associated with depression (e.g., psychomotor slowing, apathy) can be similar to those of HIV-associated neurocognitive disorders, and thus, depression must be considered as a differential diagnosis during the neuropsychological assessment.

Pharmacologic treatment of depression in youth with HIV can follow principles and guidelines for other populations since there is no evidence of differential efficacy or side effect profiles in the presence of HIV; however, attention to potential interactions with ART is important (below). As in other populations, psychopharmacologic treatment can be combined with other supportive and evidence-based therapies. Evidence-based treatments such as cognitive behavior therapy have been adapted to incorporate concerns relevant to youth with HIV and address associated issues such as medication adherence.

ANXIETY DISORDERS. Although the lifetime prevalence of anxiety disorders may be higher in adults with HIV, anxiety disorders have not received as much focus in studies of HIV. In a survey study of youth with HIV, 16% to 17% reported clinically significant symptoms of anxiety. The exception is post-traumatic stress disorder, which appears to have a higher prevalence (30% to 64%) among adults with HIV than in the general population in North America and Africa. Youth with HIV may have been premorbidly subject to chronic stress associated with discrimination as members of racial and sexual minority groups, poverty, histories of physical or sexual abuse, immigrant or refugee status, and in some cases, homelessness, among other stressors. Following their diagnosis, they may experience trauma related to receiving the diagnosis and anxiety related to HIV stigma, disclosure to sexual partners and others, and fear of HIV sequelae and mortality. The association of stress and trauma with inflammation and decreased immune functioning amplifies concern about trauma, stress, and PTSD in youth with HIV. Individual and group therapy for anxiety and PTSD should be sensitive to the acute and chronic stresses common to this population. Recommendations regarding pharmacotherapy are based largely on studies with non-HIV–infected populations, but clinicians should be aware of the possibility of additional confusion or sedation in individuals with HIV-associated cognitive impairment.

PSYCHOSIS. Individuals with chronic mental illness have an increased risk of becoming infected with HIV. Those with substance use or unstable housing may exchange sex for money or drugs or have decreased access to condoms or clean needles. HIV screening, risk reduction counseling, and, where needed, substance abuse treatment should be considered for this population. Although less commonly relevant for adolescent populations, new-onset psychotic symptoms or mania can occur in late-stage disease as part of HIV disease or opportunistic infections in the CNS; these symptoms may respond to low-dose neuroleptics.

SUBSTANCE USE. Injection drug use is a vector for HIV infection, particularly in South and Southeast Asia but to varying degrees worldwide. In addition, substance use is relevant for adolescents with HIV as a common comorbidity of other psychiatric disorders as well as a disorder in its own right, as a factor in sexual risk behaviors and poor medication management, and because of its potential interactions with immune and neurocognitive functioning. Among adults, some substances, such as alcohol and methamphetamine, have been found to exacerbate HIV-associated neurocognitive impairment. Less research on this topic has been done with adolescents; however, studies showing the increased vulnerability of the developing adolescent brain to the effects of substance use raise concern, and one study showed that HIV and alcohol risk interacted in association with cognitive functioning in youth with HIV. Furthermore, drugs of abuse may interact with some ARTs. Screening for substance use is a critical part of psychiatric and primary care for youth with HIV.

Although estimates vary, substance use overall is generally found to be more prevalent among adults with HIV than among uninfected adults. Among HIV-infected youth engaged in care, the most common substances of abuse have been reported to be cannabis, tobacco, and alcohol; however, the use of other drugs such as methamphetamine, opioids, club drugs, and cocaine varies regionally and represents a significant concern. Cannabis is of particular interest due to its wide availability, including legal use in some states; its increasing acceptability, particularly among youth; immune interactions and unknown interactions with cognitive impairment; use by some individuals to manage HIV symptoms and ART side effects; and, specifically for youth, the literature suggests a long-term impact of cannabis use during adolescence on continuing neurodevelopment.

Facilitating engagement in the care of youth with substance use is vitally important for clinical and public health reasons. Examples of strategies to encourage engagement in care include integrating treatment for substance use disorders with mental health treatment and HIV services where possible and reducing the stigma of acknowledging substance use to adolescent providers. Where abstinence is not achievable, harm reduction (e.g., reducing behaviors with the potential to transmit HIV) is an applicable strategy. Evidence-based youth-friendly therapeutic approaches such as motivational interviewing can be implemented in a variety of settings.

Children and Youth with Perinatally Acquired HIV

Neurocognitive Impact of HIV. Our understanding of the neurodevelopmental effects of HIV among perinatally infected children is complicated by the changing state of their CNS and the numerous other developmental risks present in this population. These can be environmental (e.g., economic and educational disadvantage, chronic stress related to poverty), genetic (e.g., familial psychiatric and learning disorders), prenatal or birth-related (e.g., low–birth-weight, prenatal exposure to substances of abuse and ARTs, other congenital infections such as cytomegalovirus), or related to their own and/or caregivers' chronic illness (e.g., death or changes in caregiver, missing school). Prior to the advent of combination ART, progressive HIV encephalopathy characterized by delay or loss of developmental milestones or cognitive functioning, pyramidal tract motor deficits, and microcephaly (or atrophy in older children) was seen in 24% to 35% of children with HIV. A plateau course with deviation from the expected developmental trajectory rather than a loss of milestones or decline in cognitive functioning was also seen. Currently, as with adults, encephalopathy is less common (2% to 15%), and impairments are most likely in children or youth who have had immunosuppression sufficient for an AIDS diagnosis in the past, regardless of current immune functioning, particularly if the most severe disease impact was early in life (e.g., prior to age 3) or encephalopathy was present. Although most studies have utilized measures of global cognitive functioning such as I.Q., impairments in specific cognitive domains such as language, executive functions, or processing speed have been noted. According to some studies, children and adolescents who have had consistently good viral control perform at a level comparable to those who were perinatally exposed to HIV but uninfected, with both groups performing in the low average range overall, possibly due to the other developmental risks outlined above. However, the role of prenatal exposure to HIV and ARTs cannot be ruled out by these studies.

Mental Illness and Substance Use. Children and youth with perinatally acquired HIV face numerous risks for emotional and behavioral issues. These include high rates of familial mental illness and substance use, potential or actual death of a parent, concerns about HIV-related stigma and disclosure to others, and living with the actual or potential effects of HIV itself, such as illness, cognitive impairment, delayed puberty, pain, and death. Perinatal HIV differentially affects minority and economically disadvantaged populations who may, as a result, already be subject to high levels of chronic stress. Reviews by Mellins and Malee, 2013 and Vreeman et al. noted higher rates of psychiatric disorders and behavioral problems in children and youth with perinatally acquired HIV than seen in the general population. Studies using diagnostic tools such as structured interviews have found that up to 70% have met the criteria for a nonsubstance-use psychiatric diagnosis at some time point, most commonly anxiety, ADHD, disruptive behaviors, and depression. However, some studies have found similar rates among youth who were perinatally HIV-exposed but uninfected, bringing into question the relationship between the high rate of diagnoses to HIV infection. Similarly, the results of studies using caregiver- or youth-report symptom checklists vary, some demonstrating higher rates of reported mental health problems than in the general population or comparison groups, with others showing no differences or even lower ratings of mental health problems among infected youth, possibly reflecting the close monitoring of and availability of resources for youth with PHIV in some settings.

Only in recent years have significant numbers of youth with perinatally acquired HIV survived into adolescence and reached the age of substance use initiation. They have been found to initiate substance use at similar rates as other adolescents, with higher severity of emotional symptoms or meeting symptom criteria for ADHD, major depressive disorder, and oppositional or conduct disorder associated with a greater likelihood of substance use.

Regardless of the origin of mental health problems in this population, the research demonstrates a clear need for intervention. In addition to the importance of intervention for decreasing suffering and improving quality of life, as youth with perinatally acquired HIV grow into adolescence, those with mental health problems are more likely to engage in substance use and sexual risk behaviors and to have difficulty adhering to their medication regimens. Thus, the intervention has the potential to improve long-term health outcomes and reduce HIV transmission. Mental health services are commonly provided by pediatric and adolescent HIV programs. However, little data exists regarding the differential effects of either pharmacologic or behavioral treatment in perinatally infected youth compared to uninfected youth.

Factors that have been found to increase the risk for mental health problems in the context of perinatally acquired HIV include caregiver psychiatric disorder and limit-setting problems, lower child cognitive functioning, and exposure to stressful life events. Conversely, resilience to mental health problems has been associated with higher self-esteem, positive caregiver–child relationships and parental monitoring and involvement, and family, peer, and community support. Thus, points of intervention can be multisystemic and include, for example, family-based therapy to facilitate communication, including disclosure of the child's or mother's HIV infection, peer support groups, and community programs to reduce HIV stigma. At the individual level, interventions to boost resilience through increasing coping skills and self-esteem can complement the use of pharmacologic and evidence-based therapeutic approaches.

Issues Related to Psychopharmacologic and Antiretroviral Treatment

Key issues for psychiatrists monitoring youth with HIV and prescribing psychopharmacologic medications are potential psychiatric side effects of ART and interactions between ART and psychotropic medications. Psychiatric side effects of some antiretroviral agents are well known; the use of efavirenz, in particular, can be associated with mood disturbance, insomnia and vivid dreams, impaired concentration, and in some cases, hallucinations, among other symptoms. Familiarity with the side effect profiles of patients' ART regimens is critical for the treating psychiatrist. An important issue for psychiatrists is the common pharmacokinetic and pharmacodynamic interactions between psychopharmacologic and antiretroviral agents. Some antiretroviral medications, such as protease inhibitors and non-nucleoside reverse transcriptase inhibitors, involve the CYP450 system and thus have the potential to affect drug levels of antidepressants and other medications. An additional consideration in prescribing psychopharmacologic treatments in the context of HIV is that individuals using specific ARTs or who have neurocognitive impairment may be more sensitive to the side effects of psychiatric medications. In general, the treating psychiatrist is advised to take a careful drug history, anticipate drug interactions and toxicities, start with lower dosages, and take a cautious approach when increasing them. Drug interactions are discussed in detail in the chapter “Neuropsychiatric Aspects of HIV Infection and AIDS.”

EPILEPSY

Seizures and Epilepsy Syndromes in Children

It is estimated that 1% to 1.5% of children will have a nonfebrile seizure at some time in their life. The most common seizure condition in pediatric medicine is febrile seizures which can occur in 4% of infants, toddlers, and preschoolers overall, with some ethnic groups, such as Japanese (10%) and Guamanians (20%), having a much higher incidence. Seizures can begin at any time of a child’s life, from prenatally to perinatal to onset in late adolescence. Epilepsy has been defined in the past as having two or more unprovoked seizures. A slightly different operational definition of epilepsy has recently been stated by the International League against Epilepsy, along with an updated Classification of Seizure Types.

Based on this new definition, epilepsy is a disease of the brain defined by any of the following conditions:

1. At least two unprovoked (or reflex) seizures >24 hours apart.
2. One unprovoked (or reflex) seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after two unprovoked seizures occurring over the next 10 years.
3. Diagnosis of an epilepsy syndrome.

Child neurologists and epileptologists consider in their evaluation of a child with a single or recurrent seizures the historical and examination findings, seizure type, age at onset, and EEG findings. In doing so, roughly 50% of children can be characterized as having a particular epilepsy syndrome. Making a syndrome diagnosis can provide prognostication and often guides discussion about therapeutic options.

The latter is based upon accumulated experience, clinical trials of single agents versus placebo, or rarely as in the case of childhood absence epilepsy, well-conducted comparative clinical trials. Table 2.13–7 provides information about seizure type classification, and Table 2.13–8 provides a list of pediatric epilepsy syndromes.

In past decades epileptologists and geneticists hoped for a single gene to be associated with electroclinical epilepsy syndrome, but as data have accumulated over the past 25 or more years, this idea has not been supported by available data from genetic association studies. Many individual electroclinical syndromes have been found to be associated with a given single gene mutation–often de novo, while a particular electroclinical syndrome may be associated with two or three or more different genetic mutations. Knowledge of the specific gene mutation for a given epilepsy syndrome has largely failed to inform the optimal choice of therapy. Exceptions to this include SLC2A1 mutation (glucose transporter deficiency) and SCN1A mutation associated with Dravet syndrome. The former condition responds to ketogenic diet therapy, and the latter is often worsened by lamotrigine therapy which is usually avoided in such patients.

Table 2.13–7.
Current Classification of Seizure Types (International League Against Epilepsy 2017)

Focal Onset	Generalized Onset	Unknown Onset
Aware	**Motor**	**Motor**
Impaired awareness	Tonic clonic	Tonic clonic
Motor onset	Clonic	Epileptic spasms
Automatisms	Tonic	**Nonmotor**
Atonic	Myoclonic	Behavioral arrest
Clonic	Myoclonic-tonic-clonic	**Unclassified**
Epileptic spasms	Myoclonic-atonic	
Hyperkinetic	Atonic	
Myoclonic	Epileptic spasms	
Tonic	**Nonmotor (absence)**	
Nonmotor onset	Typical	
Autonomic	Atypical	
Behavioral arrest	Myoclonic	
Cognitive	Eyelid myoclonia	
Emotional		
Sensory		
Focal to bilateral tonic–clonic		

Table 2.13–8.

Organization of the Epilepsies–Electroclinical Syndromes by Age of Onset

Neonatal Period

Benign familial neonatal epilepsy
Early myoclonic encephalopathy
Ohtahara Syndrome

Infancy

Epilepsy of infancy with migrating focal features
West syndrome
Myoclonic epilepsy of infancy
Benign infantile epilepsy
Benign familial infantile epilepsy
Dravet syndrome
Myoclonic encephalopathy in nonprogressive disorders

Childhood-onset

Febrile seizures plus (FS+)
Panayiotopolous syndrome
Epilepsy with myoclonic atonic (previously termed myoclonic astatic) seizures
Benign epilepsy with centrotemporal spikes (BECTS)
Autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE)
Late-onset childhood occipital epilepsy (Gastaut type)
Epilepsy with myoclonic absences
Lennox–Gastaut syndrome
Epileptic encephalopathy with continuous spike-and-wave in sleep (CSWS)
Landau–Kleffner syndrome
Childhood absence epilepsy

Adolescent/Adult-onset

Juvenile absence epilepsy
Juvenile myoclonic epilepsy
Epilepsy with generalized tonic–clonic seizures alone
Progressive myoclonic epilepsies
Autosomal dominant epilepsy with auditory features
Other familial temporal lobe epilepsies

Less specific age relationship

Familial focal epilepsy with variable foci (onset childhood to adult)
Reflex epilepsies

Distinctive constellations
Mesial temporal lobe epilepsy with hippocampal sclerosis
Rasmussen syndrome
Gelastic seizures with hypothalamic hamartoma
Hemiconvulsion-hemiplegia-epilepsy

Epidemiology of Psychiatric Issues in Childhood Epilepsy

The best prospective assessment of psychiatric issues in pediatric epilepsy comes from a multiyear study of new-onset seizures/epilepsy in Connecticut. In this study, 501 children were enrolled and followed for several years. One hundred fifty-nine children were judged to have a “complicated” profile with either a brain lesion evident on imaging or an IQ less than 80 or both. When brain MRI showed no etiologic structural lesion and IQ was 80 or higher in 342 children, this constituted the “uncomplicated” group.

	Complicated *N* = 159	**Uncomplicated *N* = 342**	***P*-Value**
No psychiatric disorder	98 (62%)	251 (73%)	0.008
Psychiatric disorder present	61 (38%)	91 (27%)	

Internalizing disorders (depression, bipolar disorder, obsessive-compulsive disorder, schizophrenia) were reported in 17% of the complicated group and 15% of the uncomplicated group. Externalizing disorders (ADHD, ODD, and conduct disorder) were present in 31% of the complicated group and 18% of the uncomplicated group ($P < .002$).

When specific features were evaluated, it was found that the age of seizure onset influenced the occurrence of any psychiatric disorder. Of those with onset <2 years old, 11% had a psychiatric disorder with onset at 2 to 5 years old, 24% had a psychiatric diagnosis, and if onset was >10 years old, 36% had a psychiatric disorder. Sex was not a factor influencing the incidence of psychiatric disorders. Of those children who had childhood absence epilepsy, 23% had a psychiatric diagnosis. Thirty percent of children who had benign epilepsy with centro-temporal spikes (aka Benign Rolandic Epilepsy) had a psychiatric diagnosis. Thirty-seven percent of children with juvenile absence epilepsy and juvenile myoclonic epilepsy had a psychiatric diagnosis.

More recently, questions have been raised regarding parental/proxy reporting of mood and anxiety symptoms in children with epilepsy. The point is made that when parental factors of stress and parental mood disorders are factored into the assessment of a child's behavioral problems, the differences in reported psychiatric diagnoses among different childhood epilepsy syndromes may not be apparent. Therefore, caution is advised when relying solely on parental reporting of behavior for both research and clinical assessments.

Depression and Anxiety in Pediatric Epilepsy

In the population-based study referenced above, often patients in both the complicated and uncomplicated groups had more than one psychiatric diagnosis–most often depression + anxiety or ADHD + anxiety or depression. A surprising finding in this study was no increase in the occurrence of psychiatric disorders in children who had failed >2 antiseizure medication trials versus those well-controlled on the first or second medication used.

Other studies have found by KSADS (Kiddie Schedule for Affective Disorders and Schizophrenia) assessment of children with focal unaware seizures and absence seizures, 12% and 13%, respectively, had features of depression. When assessed by the Children's Depression Inventory, 25% of children in a community-based cohort of 115 had depression, and this did not differ from those in a tertiary referral center.

Surveys of the general childhood population have shown 9.5% to have an anxiety disorder, and in children with epilepsy, figures have ranged from 5% to 49%. A wide array of disorders has been identified, including generalized anxiety disorder, specific phobias, and social anxiety, with one study finding specific phobias to be the most common. No specific data regarding optimal therapy of anxiety disorders in pediatric epilepsy have been published as of this writing, though cognitive behavioral therapy is usually recommended as first-line treatment and consideration of combined CBT with an SSRI if needed.

Data regarding the incidence/prevalence of obsessive-compulsive disorder or psychosis in children who have epilepsy have not been published to date.

Interictal and Postictal Psychosis

Only anecdotal reports are available about psychosis in childhood epilepsy, and we must turn to the adult literature for more informative and relevant data. Up to 9% of adults who have temporal lobe epilepsy will have an associated interictal psychosis, and this has been observed to be more common with left versus right temporal lobe seizure origin. Such patients most often display positive symptoms with delusions and may demonstrate hallucinations in the setting of otherwise clear consciousness. They often have the absence of a formal thought disorder, a better premorbid state, and less personality deterioration.

An uncommon situation in which psychosis coincides with improved seizure control, often with improved EEG findings via medication, vagal nerve stimulation, or resective surgery, has been termed “forced normalization” and seems to be a rare and controversial phenomenon. Great variability (<1% to 28%) has been reported in the occurrence of de novo psychosis after epilepsy surgery. When this occurs, it usually begins within 5 months and is more likely to occur in individuals who have a family history of psychosis or are older than 30 years of age at the time of surgery.

Proposed criteria for defining postictal psychosis include the following:

1. within one week of a seizure or seizure cluster
2. duration >1 day and <3 months
3. disorientation, delusions, delirium, or hallucinations in clear or clouded consciousness
4. no evidence of anticonvulsant toxicity, nonconvulsive status epilepticus on EEG, prior interictal psychosis, recent head trauma, or drug/alcohol intoxication (reference)

Postictal psychosis has been sparsely reported in children and adolescents. In the adult epilepsy literature, the risk of postictal psychosis is increased with age >30 years old, male gender, focal seizures, rapid secondary generalization of focal seizures, prior encephalitis, slowing on EEG, and prior personal or family history of psychosis. Thought insertion, commenting/command hallucinations, and negative symptoms are rare. While there are literature reports of levetiracetam, vigabatrin, and zonisamide therapy started vigorously for repeated new onset seizures as a cause for psychosis, child neurologists and consulting child psychiatrists must consider that in some patients, new onset repetitive seizures may be a presenting feature of NMDAR antibody encephalitis, and there may be an alternate explanation for psychotic behavior in a child with new-onset seizures.

It is advisable to avoid as much as possible the use of clozapine, chlorpromazine, and loxapine for pharmacologic therapy of psychosis in patients with epilepsy due to their risk of lowering seizure threshold and association with seizure occurrence in individuals not known to have epilepsy. It is important to be aware of potential drug-drug interactions when neuroleptic agents are used; for example, ziprasidone levels are reduced by enzyme-inducing antiseizure medications such as carbamazepine, phenytoin, and phenobarbital.

Aggression

Little data about aggression is reported in the pediatric epilepsy literature. Up to 5% of adults with epilepsy have problems with aggressive behavior. Risk factors include intellectual disability, onset of seizures earlier than age 10 years old, traumatic brain injury (TBI), psychosis, and lower socioeconomic status. Medications, including gabapentin, phenobarbital, lamotrigine, and levetiracetam, have been associated in case series with the occurrence of aggressive behavior that improves after cessation of that particular medication. Ictal and postictal aggression are a rare phenomenon, and in video-EEG studies, they have been associated primarily with staff or family member physical restraint attempts.

Attention-Deficit Hyperactivity Disorder and Sorting Out Staring Spells

Even prior to the initiation of antiseizure medication, symptoms of attention deficits are commonly present in children who develop epilepsy. Most studies have found approximately a 30% to 35% prevalence of attention deficit in children with epilepsy versus a 5% prevalence in the general childhood populations. A prospective cohort study of baseline neuropsychological testing in children with new onset seizure(s), including half not yet started on antiseizure medication compared with unaffected siblings, found a significantly worse performance on Conners Continuous Performance Task III in patients versus nonaffected siblings. Medication may exacerbate or precipitate attentional problems, such as has been demonstrated in a large comparative blinded trial of valproate, ethosuximide, and lamotrigine in the therapy of childhood absence epilepsy. Thirty-six percent of 393 children with new onset childhood absence epilepsy had attention deficit as gauged by Conners Performance Test II. After 16 to 20 weeks of the assigned antiseizure medication, 49% of the valproate group showed attention deficits compared to 32% of the ethosuximide and 24% of the lamotrigine-treated groups. This was statistically significant for valproate versus others, but no significant difference was present between the lamotrigine and the ethosuximide group.

In the past, there have been statements advising the avoidance of stimulant medication in children who have seizures. Numerous open-label studies and two randomized placebo-controlled studies, however, have failed to show a worsening of seizures in children with epilepsy taking either methylphenidate or atomoxetine. Although the issue has not been well studied in pediatric epilepsy patients with unstable, poorly controlled seizures, stimulant medications have been well tolerated when introduced to children taking antiseizure medications.

Not uncommonly, children who have ADHD present to healthcare providers with a history of staring spells. When observant parents or teachers can provide a detailed history, it is usually found that the staring off does not interrupt ongoing speech, play, eating, walking, or running as is usually the case in absence or focal unaware types of seizures. When staring spells are brief, on the order of seconds, and not associated with versive head or eye movements or somnolence/confusion after and there is concern that a child might have absence seizures, a routine awake EEG with two periods of hyperventilation will detect generalized 3 Hz spike-wave activity of absence seizures in 95% of affected children. If there is lingering doubt, long-term ambulatory EEG or inpatient VEEG may be required to settle the issue but will be most helpful only if the staring spells occur on a daily basis.

Psychogenic Nonepileptic Seizures

Estimates from epilepsy centers caring for children using video EEG (VEEG) for documentation of psychogenic nonepileptic seizures (PNES) indicate an estimated prevalence of PNES to vary between 1% to 9%. This is lower than 15% to 20% incidence of PNES in patients referred to adult epilepsy centers for VEEG monitoring for refractory seizures. Nonepileptic events in children vary according to age and can include syncope, parasomnias, various movement disorders, inattention/daydreaming, breath-holding spells, hypnic jerks, and sleep myoclonus. In general, the older the child, the greater likelihood that an event is a psychogenic nonepileptic event as opposed to one of the above. Older children are more likely to have motor features of their PNES, and these may have a variety of manifestations–focal, upper or lower extremities alone, whole body, etc. Clinical videography of such events to determine if the habitually observed characteristics of true epileptic seizures are a critical part of the evaluation. Younger children have been found to more often simply display a behavioral pattern of staring and unresponsiveness without motor features.

The gold standard in diagnosing PNES is VEEG monitoring with capture of a typical episode in VEEG with semiologic features not seen in epilepsy with normal EEG at the time of the event. Such patients will usually have preservation of normal EEG background rhythms and be nonresponsive to commands during PNES spells. Behavioral characteristics of PNES often include longer duration of events (>10 minutes), fluctuating course during events, side-to-side head or body movements, pelvic thrusting, asynchronous movements (excluding abrupt asymmetric tonic postures as are seen in frontal lobe partial seizures), crying during the event, eyes remaining closed during the event and memory recall within the event. Tongue biting, incontinence, flailing or thrashing movements, opisthotonus, and gradual onset of events may occur in either epileptic seizures or psychogenic nonepileptic events. They usually arise from an awake state or from apparent sleep that is not corroborated by EEG. Some patients presenting to emergency rooms may receive multiple doses of antiseizure medications or even be intubated. For convulsive events, a major rise in serum CPK 8 to 16 hours after a convulsive event supports that an event was epileptic. The absence of such a CPK rise argues against such an event being a tonic–clonic epileptic seizure but would not exclude such a possibility. Prolactin measurements comparing a baseline nonictal sample to a sample drawn 10 to 20 minutes after the onset of the seizure may be of help when the postictal prolactin measure is twice that of the baseline measure. Results may be unreliable in the setting of dopamine antagonist or agonist use, tricyclic antidepressant use, breast stimulation, and syncope.

Unfortunately, surveys of psychiatrists reveal that a substantial portion of psychiatrists believe that VEEG monitoring is not an accurate tool for determining if an event is a psychogenic nonepileptic seizure. A well-explicated characterization of the level of certainty in making the diagnosis of PNES is detailed in an ILAE task force report and can be applied in a variety of different scenarios, including healthcare facilities that do not have VEEG capability.

As most teens and adults who have PNES appear to have somatic symptom disorders as opposed to major depressive, anxiety, or bipolar disorder, psychopharmacologic therapy often has a limited role in habilitating such patients. Psychiatric consultation is sought to determine in a structured psychiatric interview if other diagnoses are present such as depression or anxiety, which may respond to psychopharmacologic intervention. Such consultation, however, is often of great benefit in identifying precipitating factors. In published studies, factors include school-related difficulties (9% to 46%); stressful family environment (8% to 67%); sexual/physical abuse (8% to 38%); fear of rejection/attention-getting (19% to 26%), and no precipitant found in 12% to 29% in various studies. DSM IV and ICD-10 psychiatric diagnoses have been reported in 6 studies. Results indicate 19% to 62% were found to have depression or anxiety, 7% to 41% had separation anxiety/school phobia, 8% to 12% have panic disorder, 6% to 12% have an exclusive general anxiety disorder, 2% to 8% have PTSD, and 16% to 59% had no psychopathologic diagnosis or was not reported in others.

A key element in the treatment of PNES is making a firm diagnosis of the habitual type of spell. This should be presented in a positive fashion to the parent or parents by the child neurologist/epileptologist. At the same time, it must be communicated to the family and child is not “faking” their seizures and that these involuntary symptoms have a biopsychosocial basis. Treatment in some situations may be as easy as simply identifying the nonepileptic nature of the events. Some series report a high rate of resolution within a few weeks of diagnosis of PNES. There should be clear communication that the treatment program must focus on the underlying psychological causes. In adults, multimodality therapies, including cognitive behavioral therapy (once weekly for 12 weeks), have been found to be more effective, with or without sertraline, compared to sertraline alone or treatment as usual. It may be important for some patients to continue to see the child neurologist to wean off antiseizure medications. Of course, ongoing care by neurology is required when patients have both PNES and epileptic seizures, which is not uncommon. Primary care physicians, school personnel, and social groups should be engaged in providing support to the treatment effort. This often requires communication about the appropriate response to the PNES spells in environments outside the home. Psychiatrists can play a critical role in the habilitation of such children not only by providing psychopharmacological treatment but also by assisting the neurology team in identifying appropriate local resources for cognitive behavioral therapy. Interventions in addition to cognitive behavioral therapy have included other modalities, including hypnosis, occupational therapy, art therapy, music therapy, and family therapy. Despite such efforts, some patients may still be refractory to treatment. Nonresponder rates are often lower in children and teens than in adults. It is reasonable in such situations to ask the neurologist/epileptologist to reassess the patient in light of the lack of improvement, especially if new symptoms arise.

Presurgical Psychiatric Evaluation

In surveys of Epilepsy Centers, 20% performed presurgical psychiatric evaluations on all patients prior to resective epilepsy surgery. Forty-five percent of centers referred for psychiatric evaluation only if prior psychiatric history is present or if presurgical neuropsychological evaluation advised psychiatric evaluation. One-fourth of such centers had their own dedicated psychiatrist, while 45% utilized different psychiatrists for this purpose. The rationale for performing such evaluations is that postoperative psychiatric complications are not rare, and identification of personal or family history of prior psychoses, depression, and anxiety can aid in the evaluation and therapy of any postoperative psychiatric disorder. Decades ago, some patients were denied epilepsy surgery if there was a major psychiatric disorder, but this is no longer judged to be a specific contraindication to surgery in an otherwise good surgical candidate. One study found, however, that the presence of major depression was a significant factor in reducing the likelihood of seizure freedom after temporal lobectomy. A single-center study utilizing both CBCL (Child Behavior Checklist) and clinical psychiatric diagnostic evaluation reported results on 40 children (aged 6 to 17 years who also had IQ above 70) who were undergoing evaluation for surgery for refractory partial epilepsy. Eighty-two percent had at least one axis I psychiatric diagnosis based on the clinical interview. ADHD was found in 40%. Anxiety was found in 32%, and depression was found in 20%. Two or more diagnoses were found in 28% of this group. The use of CBCL found broadband categories to be in the clinically significant range of 35% for internalizing problems and 25% for externalizing problems. Analysis of temporal lobe versus extratemporal foci in this group revealed that a similar proportion of both groups had attention and anxiety diagnoses, whereas CBCL data revealed that the temporal lobe foci patients had higher scores in depression-specific categories.

Adjustment Issues

Surveys have found that children with epilepsy have lower self-concepts and lower self-esteem than children with other chronic illnesses, such as diabetes or asthma. Lower social competence has been associated with lower IQ range, prior disruptive behaviors, and not so much with seizure variables. Half of the teens who have epilepsy state they do not feel stigmatized. However, they still often “keep it a secret.” Surveys of childhood and teen peers about epilepsy indicate major misperceptions supporting such stigmatization. Forty-six percent were not sure if epilepsy was contagious. Forty percent were not sure if people with epilepsy are dangerous, and only 31% would date a peer who is known to have epilepsy. Unfortunately, one survey revealed that 24% of parents believed that having a child with epilepsy in their own child’s classroom would be disruptive. Teachers tend to underestimate the learning abilities and athletic abilities of students with epilepsy and commonly report a lack of knowledge about how to handle a seizure in the classroom. As a consequence, many teachers and school personnel often advise unnecessary restrictions for students who have epilepsy. Many local chapters of the Epilepsy Foundation of America, many pediatric epilepsy centers, and other organizations provide educational programs for school nurses and school staff and offer teen and child support groups, parent support groups, and summer camp programs to help children with epilepsy overcome such stigma and promote a better quality of life.

TRAUMATIC BRAIN INJURY

TBI in children and adolescents is a major public health problem. The annual incidence of TBI in this age group is 400/100,000, and the male:female incidence ratio is at least 1.8:1. Incidence rates are related to median family income independent of age, race, or ethnicity. The proportion of TBI caused by motor vehicle accidents increases with age. The mechanism of injury in infants, toddlers, and young children is related to child abuse or falls in almost 50% of cases. Injury severity may be classified as mild (90%), moderate (7% to 8%), and severe (5% to 8%).

Pathophysiology

Diffuse injury and brain swelling that result in intracranial hypertension is more common in children than adults whereas focal injuries including intracerebral, subdural, and epidural hematomas are less common in children versus adults. Children have a higher frequency than adults of lesions in frontal white matter, orbitofrontal, and dorsolateral frontal regions. The reader is referred to a review by Serpa et al. on the pathophysiology of pediatric TBI.

Relationship of Psychiatric Disorders and Pediatric TBI

This review of psychiatric disorders in relation to pediatric TBI focuses primarily on studies that have used gold-standard psychiatric interview research assessments and less so on studies relying on parent reports and self-reports of behavior.

Preinjury Psychiatric Disorders. Prospective studies that used standardized psychiatric interviews found that 33% to 52% of consecutively treated children had a preinjury lifetime psychiatric disorder. Birth cohort prospective data have also supported a disproportionate increase in preinjury behavior problems in children who go on to have an injury compared with uninjured children. This finding may be related to higher impulsivity and risk-taking in children with externalizing disorders.

Postinjury Psychiatric Disorders. New-onset psychiatric disorders, termed “novel psychiatric disorders,” are characteristically heterogeneous. The term novel psychiatric disorders originated in the late 1990s, specifically in contrast to the term “new psychiatric disorder” used by Michael Rutter and his team in the 1970s and 1980s. New psychiatric disorders had been defined as psychiatric disorders which developed only in children with no preinjury psychiatric disorder. Novel psychiatric disorders could also occur in children with a preinjury psychiatric disorder if the child developed a different psychiatric disorder after the injury. At least one-third of children who incur a TBI have a preinjury psychiatric disorder, and therefore the broadening the definition of the population of children who might develop psychiatric complications after TBI resulted in more representative research samples. The only disorder that is unique to brain injury is personality change due to TBI, which will be discussed below. Novel psychiatric disorders have been documented to be related to injury variables. For example, they occurred in 54% to 83% of children after severe TBI, in 7% to 25% after mild/moderate TBI, and in 3% to 14% of children with orthopedic injuries without TBI. The finding that a novel psychiatric disorder is related to lower fractional anisotropy in bilateral frontal and temporal lobes, bilateral centrum semiovale, and bilateral uncinated fasciculi suggests that diffuse white matter damage is central to the emergence of psychiatric disorders after TBI. Recently it was reported that gray matter lesions of the superior frontal gyrus, and orbital gyrus, in addition to frontal white matter lesions, were independently significantly associated with novel psychiatric disorders. Noninjury variables have also been related to novel psychiatric disorders, and these included preinjury psychiatric disorders, socioeconomic status/preinjury intellectual function, preinjury adaptive function, preinjury family function, and family psychiatric history. The relationship between novel psychiatric disorders or worsening of behavioral symptomatology and family function has been shown to be a bidirectional process. Across the range of injury severity, family outcome is associated with lower preinjury family function, the presence of novel psychiatric disorders, fewer sources of support, and more stressors. The only long-term prospective longitudinal psychiatric interview study of pediatric TBI found that a novel psychiatric disorder that was present at a 24-year follow-up was independently significantly related to the severity of the injury and preinjury lifetime psychiatric disorder. These findings are consistent with findings from a Swedish birth cohort study encompassing 13 years, over one million individuals, 9% of whom had a registry entry of at least one TBI. Mean follow-up from age 8 to 26 years found that TBI contributed to absolute risks of over 10% for specialist diagnoses of psychiatric disorders and low educational attainment, about 5% for disability pension, 2% for premature mortality, and a high relative risk for psychiatric inpatient hospitalization

Mild TBI in children and adolescents has become an area of public concern, especially in the area of sports injury. Recently, the first prospective controlled longitudinal psychiatric interview study of mild pediatric TBI versus orthopedic injury showed that novel psychiatric disorders occurred at a significantly higher frequency in the first 3 months postinjury, even when analyses were adjusted for potential confounders. Multipredictor analyses found that aside from mild TBI, other variables independently accounting for novel psychiatric disorders were preinjury lifetime psychiatric disorder, high versus low family psychiatric history, and worse socio-economic status. After the first 3 months postinjury, mild TBI no longer exhibited a higher frequency of novel psychiatric disorders. These new findings add to those of a 2016 systematic review which found that psychiatric and psychological problems were more prevalent when the mild TBI was associated with hospitalization, when the evaluation occurs earlier in the recovery phase, when there were multiple previous mild TBIs, in children with pre-existing psychiatric disorders when outcomes depended on retrospective recall, and when the comparison group was uninjured children rather than children with injuries not involving the head, for example, orthopedic injuries. It has been demonstrated that children with mild TBI are more likely than orthopedically injured controls to display transient or persistent increases in postconcussive symptoms in the first year after injury, especially with more serious injuries. The families of children with mild TBI experienced more distress and burden than their orthopedic injury counterparts, and this was closely related to postconcussive symptoms.

Specific Novel Psychiatric Disorders

PERSONALITY CHANGE DUE TO TRAUMATIC BRAIN INJURY. Personality change due to TBI (PC) is the most common and important novel psychiatric disorder, especially after severe TBI. Despite its name and relatively new classification under personality disorders in DSM5, it is arguably not a personality disorder but rather a disorder of clinically significant emotional and/or behavioral problems, including affective instability, aggression, disinhibition, apathy, and paranoia. Persistent PC has been found in approximately 40% of consecutively hospitalized children with severe TBI in an average of 2 years postinjury. PC (diagnosed by waiving a 1-year duration criterion) occurred in 5% of mild/moderate TBI but was always transient, suggesting that the duration of symptoms criterion has validity. The most common subtypes are those with significant affective instability, aggression, and disinhibition presentations. The severity of the injury and superior frontal gyrus lesions are related to PC in the first year postinjury, consistent with models of affective regulation implicating failure of dorsal prefrontal cortex modulation of affect generated from a ventral brain network. Second-year postinjury correlates of PC include the severity of the injury, frontal white matter lesions, and preinjury adaptive function, suggesting that recovery is limited by network damage and related to a preinjury reserve of functioning. PC is significantly comorbid with postinjury or secondary ADHD. The neurocognitive profile of PC that emerged from research comparing children with TBI meeting diagnostic criteria of PC versus no PC was characterized by lower performance on attention, processing speed, verbal memory, IQ, and executive function measures.

SECONDARY ATTENTION-DEFICIT/HYPERACTIVITY DISORDER. Secondary ADHD (SADHD) is the term given to ADHD that develops after TBI. The phenomenology includes all subtypes of ADHD, although inattentive symptomatology is the most pronounced. SADHD is related to the increasing severity of the injury and the concurrent adaptive and intellectual function, memory deficits, and family dysfunction in samples ranging from severe to mild TBI. In samples limited in range from moderate to severe, the findings are mixed except for adaptive function differences being still evident. The categorical diagnosis of SADHD, as well as postinjury ADHD symptoms, have been associated with greater psychosocial adversity and poorer preinjury family function. Imaging correlates of SADHD included orbital frontal gyrus, right putamen, and thalamic lesions, suggesting that damage to different locations along the same cortico-striatal-pallidal-thalamic loop can result in a similar phenotypic expression. A study of neurocognitive correlates of SADHD has shown that children with SADHD exhibited deficient attention, working memory, and psychomotor speed compared to children with preinjury ADHD. In addition, children with SADHD and the injured children who had neither SADHD nor preinjury ADHD were impaired compared to the children with preinjury ADHD with regard to planning. There were no group differences related to response inhibition which is considered a signature deficit in pre-injury ADHD.

OPPOSITIONAL DEFIANT DISORDER/CONDUCT DISORDER. Novel ODD and conduct disorder have been associated with preinjury psychosocial adversity, lower socio-economic status, worse preinjury family function, lower early postinjury processing speed, lower preinjury adaptive function, and concurrent emotional lability symptoms. ODD symptoms have also been found to be associated with preinjury ODD symptoms, social class, and preinjury family function. Severity of injury predicted ODD symptoms after the first year postinjury and significantly accounted for change in ODD symptoms from before injury.

POST-TRAUMATIC STRESS DISORDER. Post-traumatic stress disorder (PTSD) can occur notwithstanding neurogenic amnesia. Studies have documented that new-onset PTSD occurs in 4% to 13% of pediatric TBI but that the presence of subsyndromal PTSD symptoms is quite common (68%) in the weeks after injury before declining to 12% by 24 months postinjury. PTSD symptoms were related to preinjury anxiety or mood disorder/symptoms, greater injury severity, postinjury anxiety and depressive symptoms, female gender, preinjury psychosocial adversity, and nonanxiety psychiatric disorders. Re-experiencing symptoms were associated with relatively intact right cingulum and hippocampus possibly because these structures are necessary in the elaboration of these particular symptoms. Hyperarousal symptoms were associated with relatively intact left orbitofrontal lesions and were associated with left temporal lesions.

OBSESSIVE-COMPULSIVE DISORDER AND ANXIETY DISORDERS. Novel obsessive-compulsive disorder has been associated with frontal and temporal lobe lesions and has occurred in the absence of striatal injury. Novel obsessive symptoms were associated with psychosocial adversity, female gender, and mesial frontal and temporal lesions. Novel anxiety disorders have included specific phobia, generalized anxiety disorder, and separation anxiety disorder. Preinjury anxiety symptoms and younger age at injury were related to postinjury anxiety symptoms. In addition, postinjury levels of stress and severity of brain injury were associated with mood/anxiety disorders. Superior frontal gyrus lesions have been associated with postinjury anxiety symptoms suggesting that damage to this region can result in affective dysregulation that includes anxiety as well as the affective instability of PC. One study found an association between fewer anxiety symptoms and orbitofrontal lesions.

MANIA/HYPOMANIA. Manic and hypomanic syndromes have occurred after pediatric TBI documented in case reports and in a prospective study in 8% of consecutively injured children. Careful consideration of the overlapping and more common syndromes of PC and SADHD should be taken before diagnosing a manic syndrome. Manic/hypomanic syndromes have been associated with frontal and temporal lesions, greater severity of injury, and a family history of mood disorder. The episodes are usually long-lasting, and there has been a similar frequency of irritability versus elation as the primary form of affective symptomatology.

DEPRESSIVE DISORDER. Novel depressive disorder has occurred in 11% to 33% of previous studies. Children with social disadvantages were at higher risk of developing depressive symptoms. Novel depressive disorder can be classified as nonanxious depression (no comorbid new-onset anxiety disorder) or anxious depressive disorder (comorbid novel anxiety disorder). Nonanxious depression was related to left inferior frontal gyrus and left temporal tip lesions, whereas anxious depression was related to right frontal lobe white matter and left parietal lesions. Anxious depression was related to PC and a family history of anxiety disorder.

AUTISM AND PSYCHOSIS. Nonaffective psychosis has been documented in only two cases in studies of consecutive admissions of children with TBI that used standardized psychiatric interviews. Autism has not been described after pediatric TBI in contrast to other forms of early brain injury. It is important clinically to distinguish autism from qualitatively distinct aberrations of pragmatic communication and social communication that have been complications of severe TBI in young children.

Case Formulation

A useful model suggested by Asarnow et al. that has guided the clinical approach to the psychiatric formulation in the assessment of children and adolescents with a history of TBI involves six pathways. (1) The psychiatric problem predates the injury and may increase the risk of its occurrence postinjury. (2) A preinjury psychiatric problem is worsened by the TBI. (3) The psychiatric problem is a direct (biologic) effect of the TBI. (4) The psychiatric problem is an immediate secondary effect of the TBI (e.g., a traumatic emotional response to the incident, such as PTSD). (5) The psychiatric problem is a long-term secondary effect of the TBI (e.g., disruptive behavior problems and lowered motivation from frustration resulting from new-onset cognitive problems). (6) The psychiatric problems are caused by factors other than the TBI.

Intervention

Prevention. Prevention is the most critical intervention with regard to pediatric TBI. This includes education with respect to the use of bicycle helmets, improving motor vehicle safety, limiting alcohol-related motor vehicle accidents, and reducing the risk of child abuse and neglect.

Nonpharmacologic Treatment

FAMILY THERAPY. Web-based family and teen problem-solving therapies have been developed to target behavioral problems after TBI. These included Online Family Problem-Solving Therapy (F-PST), either self-guided or therapist-guided, resulting in similar improvement based on parent ratings of adolescent behavior and executive function, whereas the corresponding face-to-face treatment was not associated with improvement. Earlier iterations of this type of therapy included counselor-assisted problem solving (CAPS), which is a 6-month web-based, family-centered treatment that focuses on problem-solving, communication, and self-regulation. CAPS and a similar treatment called teen online problem solving (TOPS) have been effective in multiple ways, including improving parent-rated behavioral executive functioning and externalizing problems in the children, as well as enhancing problem-solving and decreasing depressive symptoms in parents. Pilot research has demonstrated that cognitive-behavioral therapy for insomnia (CBT-I) in children with sleep disruption and persistent postconcussion symptoms leads to improved sleep quality, shorter sleep-onset latency, better sleep efficiency, fewer dysfunctional beliefs about sleep, and longer sleep time compared with children receiving treatment as usual. Other effective approaches have included general cognitive behavioral therapy and the restructuring of the child’s environment to optimize behavioral supports.

SCHOOL. The Individuals with Disabilities Education Act entitles children, especially those with severe TBI, to services. Furthermore, special education interventions target poor academic function related to a variety of permutations of academic skills deficits and emotional and behavioral problems, which may compound corresponding preinjury problems. For a more detailed review of educational and neuropsychological correlates and interventions for pediatric TBI, the reader is referred to Crowe et al.

Pharmacologic Therapies. A comprehensive review of pharmacologic treatment in pediatric brain injury by Feinberg et al. has concluded that there was a limited number of high-quality, clinically meaningful studies in the acute stage of injury and, therefore, evidence-based analysis that would inform clinicians' decisions on medication management was not feasible. Anecdotal guidelines only are therefore noted here. Mood stabilizing antiseizure medications and selective serotonin reuptake inhibitors can be helpful together with behavioral treatment for affective instability and aggression typical in PC. The disinhibited type of PC is difficult to treat. Neuroleptic medication may be helpful for the agitated or paranoid patient when the behavior is complicating rehabilitation and recovery, although usage must be balanced against potential risks in the alteration of neuronal recovery. There have been several positive reports of stimulant use for children with ADHD-like symptoms after TBI, although overall findings are mixed. In the absence of research data, depressive disorders after TBI may be treated with selective serotonin reuptake inhibitors as in corresponding syndromes in the absence of TBI.

Conclusion

Pediatric TBI is a major cause of morbidity and mortality and is thus an important public health problem. Particularly with severe TBI, the sequelae may be life-long. Mental health professionals have a key role to play in this field because psychiatric disorders are common both before and after injury. As in circumstances unrelated to TBI, psychiatric disorders and family function have a close relationship. Novel psychiatric disorders are associated with injury and psychosocial variables that can be recorded soon after injury, and this opens potential options for secondary prevention and early recognition of predictable outcomes. Neurocognitive, school, and adaptive function problems often occur after severe TBI, and effective treatment for the patient must address each of these domains of function. Research in psychosocial treatments is more advanced than psychopharmacological approaches at this point.

OVERALL SUMMARY

There are several striking similarities and differences with regard to psychopathology in children and adolescents with varying brain injuries discussed in this chapter. Psychiatric disorders occur commonly in each of the brain injury groups we have reviewed. The rates of psychiatric disorders are at least 27% (in uncomplicated epilepsy) and generally higher in the range of 50% to 60% in children with the other kinds of brain injury reviewed. The psychiatric disorders that manifest are typically heterogeneous, that is, there is no signature disorder that occurs in all brain-injured children. Comorbidity is also the rule rather than the exception. The disorder specific to brain injury is Personality change due to a general medical condition. Aside from this disorder, it is quite clear that the most common disorder in brain-injured children is ADHD, typically around 25% to 30%, except for FASD, where ADHD occurs at at least double that rate. The very high rate of ADHD associated with prenatal alcohol exposure may be due to a confluence of factors that include a particularly virulent teratogen, genetic factors, and poverty. There is accumulating evidence that ADHD associated with brain injury has different neuropsychological characteristics compared with developmental ADHD. Other psychiatric disorders such as depressive, anxiety, and ODDs that are present occur at similarly high rates across brain injury categories. Of the brain disorders reviewed in this chapter, Autism is most closely associated with prematurity and low–birth-weight. While it is true that children with autism have a relatively high rate of epilepsy, only a small proportion of children with epilepsy have autism. Standardized psychiatric interviews are more sensitive in detecting psychiatric disorders compared with behavior checklists in brain-injured pediatric populations, possibly because parents who complete the checklists have accommodated their child’s limitations and often do not keep in mind a normative expectation.

IQ tends to be low average in children who are significantly affected by the brain injuries discussed in this chapter. Intellectual disability or IQ ≤70 is unusual. Deficits occur in multiple neuropsychological domains and again there is no signature neuropsychological deficit profile across brain injury groups.

Treatment for the psychiatric disorders associated with the brain injury groups generally rely on treatments designed for children and adolescents in general or even those used in adults with brain injury. This will remain the standard of care until more intervention research is conducted. The public health impact of brain injury-related psychopathology in children and adolescents is high because of its high frequency as well as the many years of illness burden when brain injury is acquired at a young age. There is an exceptional need for mental health researchers and clinicians to understand, develop treatments, and to intervene to improve the lives of young people with brain injury.

ACKNOWLEDGMENT

Preparation was supported by NIAAA grant U01 AA014834 (S. Mattson).

Further Readings

Aarnoudse-Moens CS, Weisglas-Kuperus N, van Goudoever JB, Oosterlaan J. Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Pediatrics*. 2009;124:717–728.

Anderson PJ. Neuropsychological outcomes of children born very preterm. *Semin Fetal Neonatal Med*. 2014;19:90–96.

*Berg AT, Altalib HH, Devinsky O. Psychiatric and behavioral comorbidities in epilepsy: a critical reappraisal. *Epilepsia*. 2017;58:1123–1130.

Brown LK, Whiteley L, Harper GW, Nichols S, Nieves A; ATN 086 Protocol Team for The Adolescent Medicine Trials Network for HIV/AIDS Interventions. Psychological symptoms among 2032 youth living with HIV: a multisite study. *AIDS Patient Care STDS*. 2015;29(4):212–219.

Crocker N, Riley EP, Mattson SN. Visual-spatial abilities relate to mathematics achievement in children with heavy prenatal alcohol exposure. *Neuropsychology*. 2015;29:108–116.

Crowe LM, Catroppa C, Anderson V. Sequelae in children: developmental consequences. *Handb Clin Neurol*. 2015;128:661–677.

Dunn DW, Besag F, Caplan R, Aldenkamp A, Gobbi G, Sillanpaa M. Psychiatric and behavioural disorders in children with epilepsy (ILAE Task Force Report): anxiety, depression and childhood epilepsy. *Epileptic Disord*. 2016;18(Suppl 1):1–86.

Emery CA, Barlow KM, Brooks BL, et al. A systematic review of psychiatric, psychological, and behavioural outcomes following mild traumatic brain injury in children and adolescents. *Can J Psychiatry*. 2016;61:259–269.

Fastenau PS, Johnson CS, Perkins SM, et al. Neuropsychological status at seizure onset in children: risk factors for early cognitive deficits. *Neurology*. 2009;73:526–534.

Hadaya L, Nosarti C. The neurobiological correlates of cognitive outcomes in adolescence and adulthood following very preterm birth. *Semin Fetal Neonatal Med*. 2020;25:101117.

Jansen L, van Steenis A, van den Berg-Huysmans AA, et al. Associations between neonatal magnetic resonance imaging and short- and long-term neurodevelopmental outcomes in a longitudinal cohort of very preterm children. *J Pediatr*. 2021;234:46–53.e2.

Johnson S, Marlow N. Growing up after extremely preterm birth: lifespan mental health outcomes. *Semin Fetal Neonatal Med*. 2014;19:97–104.

Joseph RM, O'Shea TM, Allred EN, et al. Prevalence and associated features of autism spectrum disorder in extremely low gestational age newborns at age 10 years. *Autism Res*. 2017;10(2):224–232.

Khoury JE, Milligan K. Comparing executive functioning in children and adolescents with fetal alcohol spectrum disorders and ADHD: a meta-analysis. *J Atten Disord*. 2019;23:1801–1815.

Kochanek PM, Bell MJ, Bayir H. Quo vadis 2010? – carpe diem: challenges and opportunities in pediatric traumatic brain injury. *Dev Neurosci*. 2010;32:335–342.

Kurowski BG, Taylor HG, McNally KA, et al. Online family problem-solving therapy (F-PST) for executive and behavioral dysfunction after traumatic brain injury in adolescents: a randomized, multicenter, comparative effectiveness clinical trial. *J Head Trauma Rehabil*. 2020;35:165–174.

*Mathewson KJ, Chow CH, Dobson KG, Pope EI, Schmidt LA, Van Lieshout RJ. Mental health of extremely low birth weight survivors: a systematic review and meta-analysis. *Psychol Bull*. 2017;143:347–383.

*Mattson SN, Roesch SC, Glass L, et al. Further development of a neurobehavioral profile of fetal alcohol spectrum disorders. *Alcohol Clin Exp Res*. 2013;37:517–528.

Max JE, Troyer EA, Arif H, et al. Traumatic brain injury in children and adolescents: psychiatric disorders 24 years later. *J Neuropsychiatry Clin Neurosci*. 2022;34(1):60–67.

Max JE, Wilde EA, Bigler ED, et al. Personality change due to traumatic brain injury in children and adolescents: neurocognitive correlates. *J Neuropsychiatry Clin Neurosci*. 2015;27(4):272–279.

McDougall S, Finlay-Jones A, Arney F, Gordon A. A qualitative examination of the cognitive and behavioural challenges experienced by children with fetal alcohol spectrum disorder. *Res Dev Disabil*. 2020;104:103683.

McLachlan K, Flannigan K, Temple V, Unsworth K, Cook JL. Difficulties in daily living experienced by adolescents, transition-aged youth, and adults with fetal alcohol spectrum disorder. *Alcohol Clin Exp Res*. 2020;44:1609–1624.

Moore PS, Mokrova I, Frazier JA, et al. Anxiety and depression correlates at age 10 in children born extremely preterm. *J Pediatr Psychol*. 2021;46:422–432.

Nichols SL, Bethel J, Garvie PA, et al. Neurocognitive functioning in antiretroviral therapy-naive youth with behaviorally acquired human immunodeficiency virus. *J Adolesc Health*. 2013;53:763–771.

Pangilinan PH, Giacoletti-Argento A, Shellhaas R, Hurvitz EA, Hornyak JE. Neuropharmacology in pediatric brain injury: a review. *PM R*. 2010;2:1127–1140.

Popova S, Temple V, Dozet D, O'Hanlon G, Toews C, Rehm J. Health, social and legal outcomes of individuals with diagnosed or at risk for fetal alcohol spectrum disorder: Canadian example. *Drug Alcohol Depend*. 2021;219:108487.

Pyhala R, Wolford E, Kautiainen H, et al. Self-reported mental health problems among adults born preterm: a meta-analysis. *Pediatrics*. 2017;139(4):e20162690.

Raldiris TL, Bowers TG, Towsey C. Comparisons of intelligence and behavior in children with fetal alcohol spectrum disorder and ADHD. *J Atten Disord*. 2018;22:959–970.

Saigal S. Functional outcomes of very premature infants into adulthood. *Semin Fetal Neonatal Med*. 2014;19:125–130.

Salpekar JA, Mishra G. Key issues in addressing the comorbidity of attention deficit hyperactivity disorder and pediatric epilepsy. *Epilepsy Behav*. 2014;37:310–315.

Sariaslan A, Sharp DJ, D'Onofrio BM, Larsson H, Fazel S. Long-term outcomes associated with traumatic brain injury in childhood and adolescence: a nationwide Swedish cohort study of a wide range of medical and social outcomes. *PLoS Med*. 2016;13:e1002103.

Serpa RO, Ferguson L, Larson C, et al. Pathophysiology of pediatric traumatic brain injury. *Front Neurol*. 2021;12:696510.

Skorka K, McBryde C, Copley J, Meredith PJ, Reid N. Experiences of children with fetal alcohol spectrum disorder and their families: a critical review. *Alcohol Clin Exp Res*. 2020;44:1175–1188.

Taylor HG. Neurodevelopmental origins of social competence in very preterm children. *Semin Fetal Neonatal Med*. 2020;25:101108.

Torchin H, Morgan AS, Ancel PY. International comparisons of neurodevelopmental outcomes in infants born very preterm. *Semin Fetal Neonatal Med*. 2020;25:101109.

*Vreeman RC, McCoy BM, Lee S. Mental health challenges among adolescents living with HIV. *J Int AIDS Soc*. 2017;20:21497.

Ware AL, Glass L, Crocker N, et al. Effects of prenatal alcohol exposure and attention-deficit/hyperactivity disorder on adaptive functioning. *Alcohol Clin Exp Res*. 2014;38:1439–1447.

Yates R, Treyvaud K, Doyle LW, et al. Rates and stability of mental health disorders in children born very preterm at 7 and 13 years. *Pediatrics*. 2020;145(5):e20192699.

2.14 Neuropsychiatry of Neurometabolic and Neuroendocrine Disorders

MARK WALTERFANG, M.B.B.S. (HONS), PH.D., F.R.A.N.Z.C.P., RAMON MOCELLIN, M.B.B.S., M.SC., F.R.A.N.Z.C.P., OLIVIER BONNOT, M.D., PH.D., AND DENNIS VELAKOULIS, M.B.B.S., M.P.M., DIP.CRIM., D.MED.SCI., F.R.A.N.Z.C.P.

Metabolic and endocrine disturbances can have wide-ranging effects on the central nervous system (CNS). Neurons, because of their high metabolic level of activity and demand, are often acutely sensitive to derangements in systemic metabolic processes. As a result, disorders affecting systemic metabolism often have a very high rate of associated CNS disturbance, ranging from the severe—with gross neurodevelopmental disruption, encephalopathy, seizures, delirium, and/or coma—to the mild, with subtle cognitive and behavioral disturbance. A number of these disorders cause recognizable and significant psychiatric illnesses, including psychotic disorders, affective disturbance, anxiety disorders, attention-deficit disturbance, and other behavioral disturbances such as apathy, dysexecutive syndromes, and catatonia. For the purposes of this section, those disorders of metabolism and endocrine function associated with significant CNS disturbance will be referred to as *neurometabolic* and *neuroendocrine* disorders, respectively, and those associated with major neuropsychiatric syndromes will be highlighted.

The relevance of recognition of neuropsychiatric comorbidity in neurometabolic and neuroendocrine syndromes has both diagnostic and treatment importance. A number of these disorders can produce an accurate *phenocopy* of psychiatric illness, where illness presentation carries most or all the characteristic features of a psychiatric diagnosis. For example, a schizophrenia-like psychosis may develop in metachromatic leukodystrophy (MLD) that is otherwise indistinguishable from “typical” schizophrenia, at least until other features more typical of the primary neurometabolic illness supervene. Awareness of the types of disorders that may present as psychiatric illness phenocopies and their associated physical, cognitive, or neurologic concomitants allows for the appropriate recognition and diagnostic confirmation of an underlying metabolic or endocrine illness. In some circumstances, this may mean medical treatment of an endocrine illness, enzyme replacement, or substrate reduction therapy for a metabolic illness. Failure to recognize the underlying illness while focusing on psychiatric treatment alone can delay the institution of appropriate management and result in potentially irreversible CNS changes. However, it is not uncommon for these disorders to present with associated, and often atypical, features—such as cognitive or motor impairment or other systemic organ involvement (such as hepatosplenomegaly)—and the recognition of “atypicality” in the presentation of psychiatric illness, and the associated disorders that may result in secondary atypical presentations, is an important skill for all psychiatrists working in healthcare settings.

Most clinicians will undertake a routine set of investigations for underlying “organic” causes of psychiatric illness at the initial presentation, including screening for thyroid illness, infective disease, electrolyte disturbance, and renal and kidney function, in addition to brain imaging and/or electroencephalography (EEG). Although these tests are relatively high yield and detect a number of illnesses producing secondary psychiatric syndromes, less common but no less significant conditions are often missed. Literally, hundreds of metabolic disorders have been described, and only a very limited number of these are screened for in the neonatal period (usually less than 10 in most centers). This may provide the clinician with a false assurance that any significant and/or serious metabolic disorders will already have been detected or excluded. This situation is further complicated by the fact that the age of onset and presentation, in addition to disease expression, vary greatly across and within these disorders, and disorders that present later or in a less severe form are often missed by screening processes during the neonatal and childhood periods, or are not suspected in the setting of non-specific presentations (e.g., recurrent emergency department presentations with abdominal pain in a patient with acute intermittent porphyria [AIP]).

Those metabolic disorders that present initially, or predominantly, with neuropsychiatric syndromes generally have their onset in the period of the life cycle associated with the onset of the majority of psychiatric illnesses, that is, adolescence or early adulthood—the “psychiatric window” where the majority of psychiatric illnesses manifest. The disorders that present during this window include inborn errors of metabolism that affect cellular function in the CNS. A wide range of phenotype–genotype correlations occur in such disorders, as differing mutations in the gene for the protein serving a key function in metabolism will result in different structural or conformational changes to the protein product and differential metabolic effects. For example, some mutations in the gene encoding for the NPC1 protein in Niemann–Pick disease type C are associated with an early onset of illness, severe intellectual disability, and death in childhood, whereas others are associated with a less significant deficit in the function of NPC1, and thus less disruption to intracellular cholesterol metabolism (even falling into the “normal” range), and a presentation in adolescence or early adulthood as opposed to early childhood. The adult-onset forms of these diseases often form a minority of cases of these disorders, although they will generally respond as favorably to primary treatments for the disorder; these late, “nonclassical” presentations are often, unfortunately, missed. Additionally, some carriers with autosomal recessive disorders who were previously considered asymptomatic (such as female carriers with adrenomyeloneuropathy) have been shown to have attenuated syndromes often involving subtle psychiatric disturbance. For most of these disorders, gross metabolic disturbance results in severe impairment of CNS function (such as delirium, seizures, or encephalopathic presentations), with moderate impairments often resulting in dementia (or intellectual disability in children) and movement disorders. More subtle impairments, however, are less likely to disrupt these core vegetative and functional systems of the brain but rather disrupt higher-order functions that are much more dependent on highly synchronous cortico-cortical and subcortical connectivity, with the result being a disturbance to higher cognitive functions and a predilection for major mood and psychotic disorders. As these systems tend to mature later in the brain’s neurodevelopmental trajectory, those disorders that alter or interrupt later neurodevelopment are more likely to cause neuropsychiatric syndromes. For progressive neurometabolic disorders, it is not uncommon for these disorders to present initially with a neuropsychiatric syndrome, which may be diagnosed and treated as a primary psychiatric illness. As the pathologic effect of the metabolic derangement impinges further on the CNS and begins to result in degenerative change, frank neurologic illness, and dementia often supervene.

In addition to those metabolic disorders that impact neurodevelopment, some metabolic disorders are associated with episodic but reversible metabolic disturbances (such as the acute porphyrias or urea cycle disorders [UCDs]). As opposed to impacting late-maturing developmental networks, these disorders impact metabolically on presumably mature or normally developed brain systems. Those disorders that show a predilection for neuropsychiatric disturbance are more likely to affect brain regions (such as the frontal or temporal cortical regions) or transmitter systems (particularly monoaminergic or glutamatergic systems) that are strongly associated with psychiatric illness. Similarly, the majority of major endocrine disorders that are associated with psychiatric disturbance do so in the setting of an otherwise intact CNS. These endocrine systems are crucially involved in energy metabolism, cell turnover, and downstream metabolic effects. Elevated rates of psychiatric disturbance in these disorders occur via a direct effect of the altered hormonal system on particularly vulnerable cellular populations or neurotransmitter systems, such as the effects of elevated cortisol levels on hippocampal neurons or thyroid hormone on serotonin turnover.

NEUROMETABOLIC SYNDROMES

Lysosomal Disorders

The lysosome is a subcellular organelle that is manufactured by the Golgi apparatus and contains a number of hydrolytic enzymes such as proteases, lipases, nucleases, and polysaccharidases. They function as the “garbage disposal” system of the cell via phagocytosis (digestion of extracellular material), endocytosis (digestion of cell surface proteins), and autophagy (digestion of old or damaged intracellular organelles or structures). Macromolecules (DNA, RNA, proteins, glycoproteins, lipids, and phospholipids) transported to lysosomes are degraded by enzymatic “factories,” which then pass out their monomeric components for reutilization. Impairment in the function of a lysosomal enzyme occurs if it is structurally altered, alteration occurs to a cofactor protein, or enzyme transport is affected. Each enzyme is specific for breaking a particular chemical bond, as opposed to a particular substrate macromolecule. Over

70 lysosomal enzymes are known, and more than 40 disease syndromes involving defective enzyme function have been characterized. Many of these disorders present both in childhood and adulthood, and for those that present in adolescence or early adulthood—the period of onset of most major psychiatric disorders—the rate of major psychiatric illness is not unexpectedly elevated, whereas childhood presentations commonly result in major intellectual disability and older adult presentations in dementia, in addition to frank neurologic disturbance. Of those lysosomal storage disorders (LSDs) that present in adult life, those strongly associated with neuropsychiatric presentations include those involving defective breakdown of sphingolipid components (MLD, Fabry disease, GM2-gangliosidosis/Tay–Sachs disease (TSD), Niemann–Pick type C disease), glycoproteins (α-mannosidosis [AM]), and cholesterol and lipids (neuronal ceroid lipofuscinosis [NCLs]).

Metachromatic Leukodystrophy

MLD is an autosomal recessive, incompletely penetrant genetic deficiency of the lysosomal enzyme arylsulfatase A (ASA) that occurs in 1.5 per 100,000 individuals and was first described in 1933. ASA hydrolyzes various sulfatides, including sulfate-containing lipids of the CNS. Lysosomal sulfatide accumulates in the brain, peripheral nerves, kidney, and gallbladder, but particularly in myelinated structures, seen as metachromatic granules on histologic examination and widespread loss of myelin. MLD is protean in its presentations such that in younger patients, seizures and motor symptoms predominate with psychiatric manifestations and dementia frequently occurring prior to the onset of motor symptoms in adult patients, particularly those with the I179S mutation; the severity is inversely correlated with the degree of residual ASA activity. The adult form appears to cleave into two distinct phenotypes, one with a predominantly motor cerebello-pyramidal presentation and the other with a predominantly psychiatric presentation. Up to half of the patients with illness onset between 10 and 30 years of age present with psychotic symptoms, including auditory hallucinations, systematized delusions, formal thought disorder, catatonic posturing, and inappropriate affect. As the illness progresses, other neurologic symptoms supervene, such as seizures, chorea, or dystonia. Dementia is an inexorably progressive component of the illness, initially presenting as attentional disturbance, reduced speed of processing and executive impairment, and may present as a phenocopy of frontotemporal dementia; if disinhibition is a prominent feature, it may present as a phenocopy of hypomania. Diagnosis is made by demonstrating reduced enzyme activity in leukocytes or skin fibroblasts. Magnetic resonance imaging (MRI) generally demonstrates typical periventricular white matter changes sparing subcortical U-fibers and often shows pathology with a frontotemporal preponderance (Fig. 2.14–1).

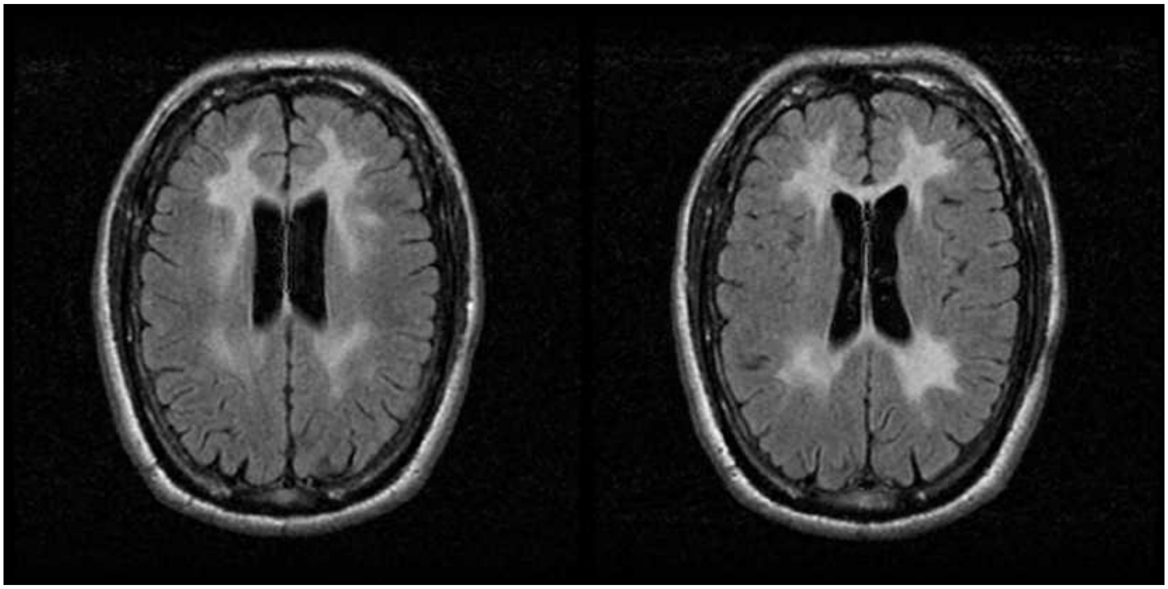

FIGURE 2.14–1. Two-axial FLAIR images from a 32-year-old male with genetically confirmed metachromatic leukodystrophy (MLD). White matter involvement is symmetrical, with an anterior preponderance superiorly, and characteristic involvement of the patient’s corpus callosum. (Supplied by author.)

Treatment is generally symptomatic, although bone marrow transplantation has shown benefit in some patients, and intrathecal enzyme replacement therapy is in phase II trials. Adolescent/adult MLD provides an intriguing model for the understanding of the neurobiology of psychosis, as it interrupts myelinative processes that occur during this critical period of neurodevelopment, in particular frontotemporal myelination. As frontotemporal connectivity is known to be impaired at both functional and structural levels in schizophrenia, MLD appears to have an almost uniquely psychotogenic propensity in this age group and suggests that a neuropathologic process that interrupts the normal development of connectivity between these cortical regions in this crucial late neurodevelopmental window can result in psychosis. It has also been shown that some treatment-refractory schizophrenia patients have ASA pseudo-deficiency or intermediate ASA levels (assessable with leucocyte enzyme assay), suggesting that this may be a disease modifier or potentially an independent risk factor for the development of psychosis through its disruption of myelination.

Fabry Disease

Fabry disease is an X-linked recessive disorder of the lysosomal enzyme α-galactosidase A (GLA), resulting in the accumulation of the glycolipid globotriaosylceramide (Gb3) in blood vessels and other tissues (Fig. 2.14–2), which occurs in 1 in 40,000 to 60,000 males. Hemizygous males are particularly affected, with females varying from asymptomatic to severely affected due to the degree of X-inactivation. The major clinical manifestations of Fabry disease reflect the particular impact of the disorder on the vascular endothelium. Early clinical features with onset in childhood or adolescence include angiokeratomas and acroparesthesias. Angiokeratomas are dark, punctate vascular lesions most prevalent between the umbilicus and the knees but also seen on the oral mucosa and conjunctiva. Acroparesthesias are acute episodes of severe pain in the fingers and toes lasting days to weeks precipitated by exercise, fatigue, or fever, which are very disabling. Chronic pain syndromes associated with acroparesthesias may continue into adulthood and account for much of the associated psychiatric comorbidity. Corneal or lenticular opacities and hypohidrosis are also early manifestations, and patients may also develop tinnitus and hearing loss. Disease progression is marked by vascular disease, resulting in kidney impairment, cardiovascular disease, and stroke in adulthood. Brain MRI often demonstrates subcortical changes in periventricular white matter and the basal ganglia, thalamus, and brainstem. Cognitive impairment may be detected via formal neuropsychological testing and often presents as impaired processing speed, attention and executive functioning, likely reflecting the impact of subcortical vascular ischemia. Depressive disorders, often meeting the criteria for severe clinical depression, occur in up to half of all sufferers and are most strongly associated with the degree of peripheral pain and anhidrosis, suggesting that it reflects the burden of disease rather than being a poststroke phenomenon or associated with subcortical vascular change. Anxiety is also commonly reported; other psychiatric syndromes appear relatively rare. The severity of depression and anxiety also appears to predict the degree of social and occupational dysfunction. Symptomatic treatment of depression is often effective; however, enzyme replacement therapy with agalsidase-alpha or -beta may prevent some of the physical manifestations of Fabry disease that appear to be causally related to depression in this disorder; other treatments are symptomatic, including anticonvulsants, anticoagulants, and antihypertensives.

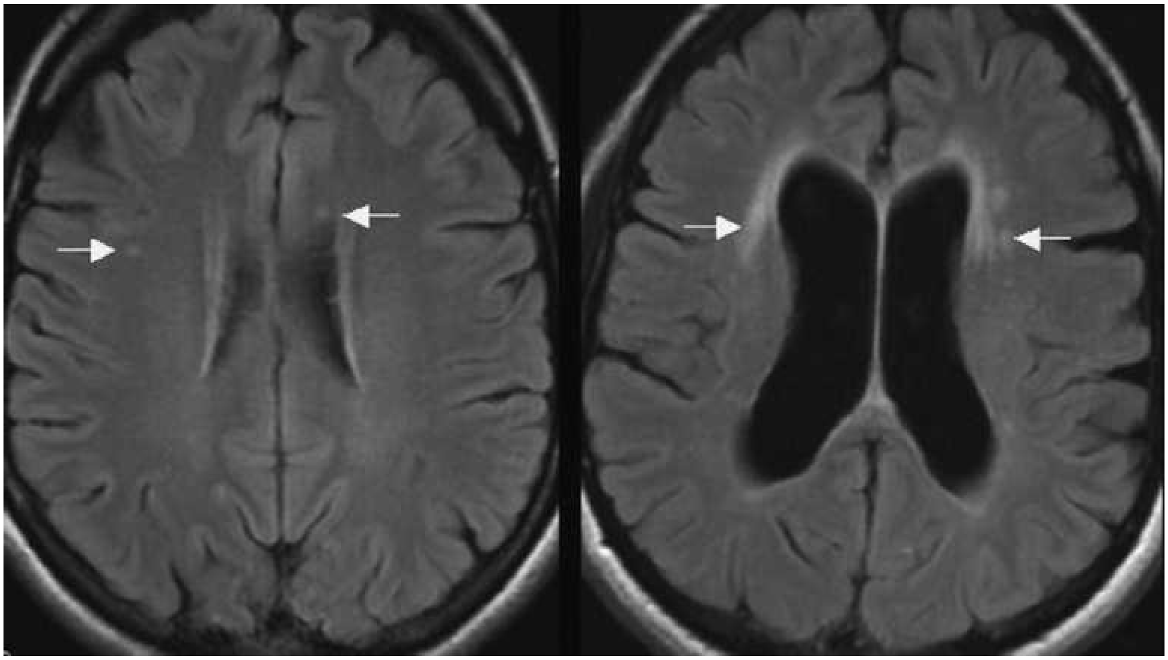

FIGURE 2.14–2. Fluid-attenuated inversion recovery magnetic resonance image demonstrating mild (*left*) white matter lesions (*arrows*) in a 39-year-old woman with Fabry disease, and more significant (*right*) white matter lesions in a 41-year-old woman with the same disease. (From Fellgiebel A, Müller MJ, Mazanek M, Baron K, Beck M, Stoeter P. White matter lesion severity in male and female patients with Fabry disease. *Neurology*. 2005;65(4):600, Figure 1, with permission. doi: 10.1212/01.wnl.0000173030.70057.eb. © American Academy of Neurology.)

Gaucher Disease

Gaucher disease (GD) is the most common hereditary LSD due to a deficiency in β-glucocerebrosidase (GBA1) occurring in 1 in 50,000 individuals, which results in the accumulation of glucocerebroside in macrophage lysosomes. GD carrier status has more recently been shown to be the most common genetic risk factor for Parkinson disease (PD) because of the molecular relationship between glucocerebrosides and alpha-synuclein. These are deposited in a variety of tissues, including marrow, liver, spleen, and brain. Type I GD is “non-neuronopathic” as the brain, and spinal cord are thought to be spared; this presents with hepatosplenomegaly, anemia, leukopenia, and thrombocytopenia in addition to skeletal bone lesions. Types 2 and 3 (“neuronopathic”) affect the CNS, with type II causing progressive severe disease in infancy and type III being more indolent and often associated with survival into adulthood, with the development of seizures, ataxia, and gaze palsy being common neurologic presentations. It is known, however, that even GD1 patients have some neuronal involvement, with astrogliosis and alpha-synuclein deposition occurring in neuronal populations, but with only type I patients being spared neuronal loss. GD1 patients may demonstrate subtle cognitive deficits, largely in attention and memory; in GD3 patients, cognitive function can vary widely—from severe impairment to normal functioning—according to the severity of the disease. Depression and anxiety are not uncommon in GD1 patients, at approximately three times the rate of the normal population. Depression in GD patients is known to respond to antidepressant medication. Enzyme replacement therapy in GD is associated with significant improvement in mood and psychological functioning, with similar improvements reported in quality of life, even though intravenous enzyme therapy does not significantly cross the blood–brain barrier. Notably, heterozygosity for GBA1 disease-causing mutations is now known to be the most significant genetic predisposition to PD, with 5% of PD patients carrying a GBA1 mutation, compared to 1% of the healthy population.

GM2 Gangliosidosis (Tay–Sachs Disease)

TSD is an autosomal recessive lipid storage disorder caused by the accumulation of GM2-gangliosides within neurons due to a deficiency in β-hexosaminidase A (HEX-A) and occurs in 1 in 300,000 individuals, although it is as common as 1 in 3,000 in the Ashkenazi Jewish population—and thus is part of carrier screening for some individuals prior to childrearing. HEX-A deficiency in lysosomes impairs the catabolism of gangliosides from the neuronal cell membrane, resulting in the accumulation of lysosomal gangliosides (Fig. 2.14–3). This leads to secondary axoneuronal changes, particularly axon hillock outgrowth to form “meganeurites” with ectopic dendritogenesis and focal axonal enlargements known as *axonal spheroids*, both of which may alter neuron-to-neuron microconnectivity. Additionally, ganglioside accumulation results in direct neurotoxicity, altered neuronal electrical properties, inappropriate apoptosis, or an inflammatory response. In infantile or childhood forms, severe neurologic impairment usually results in death within 3 to 10 years. A later, or adult-onset, form of TSD has been described and may be associated with having at least one allele (most commonly the G269S mutation) associated with residual enzyme activity. In this late-onset form, psychiatric symptoms may co-present with or predate the development of neurologic disturbances in early adulthood. Patients present with a speech disorder, gait disturbance, saccadic eye movement abnormalities and tremor, with normal or near-normal cognitive function, although subtle deficits in executive function, processing speed, and memory may be present in up to half of patients. Brain MRI often shows significant cerebellar atrophy, although cortical atrophy may also be seen. Neuropsychiatric presentations occur in up to half of late-onset TSD patients, predominantly psychosis, which may occur in 30% to 50% of adult patients, marked by disorganization, auditory and visual hallucinations, and catatonia. Affective disorders, including mania and depression, have been occasionally reported. Patients only partially respond to neuroleptics or lithium and are often very sensitive to motor side effects of these medications; neurologic worsening has been described with typical antipsychotics and tricyclic antidepressants (TCAs). Importantly patients with psychotic illness appear to respond to electroconvulsive therapy (ECT).

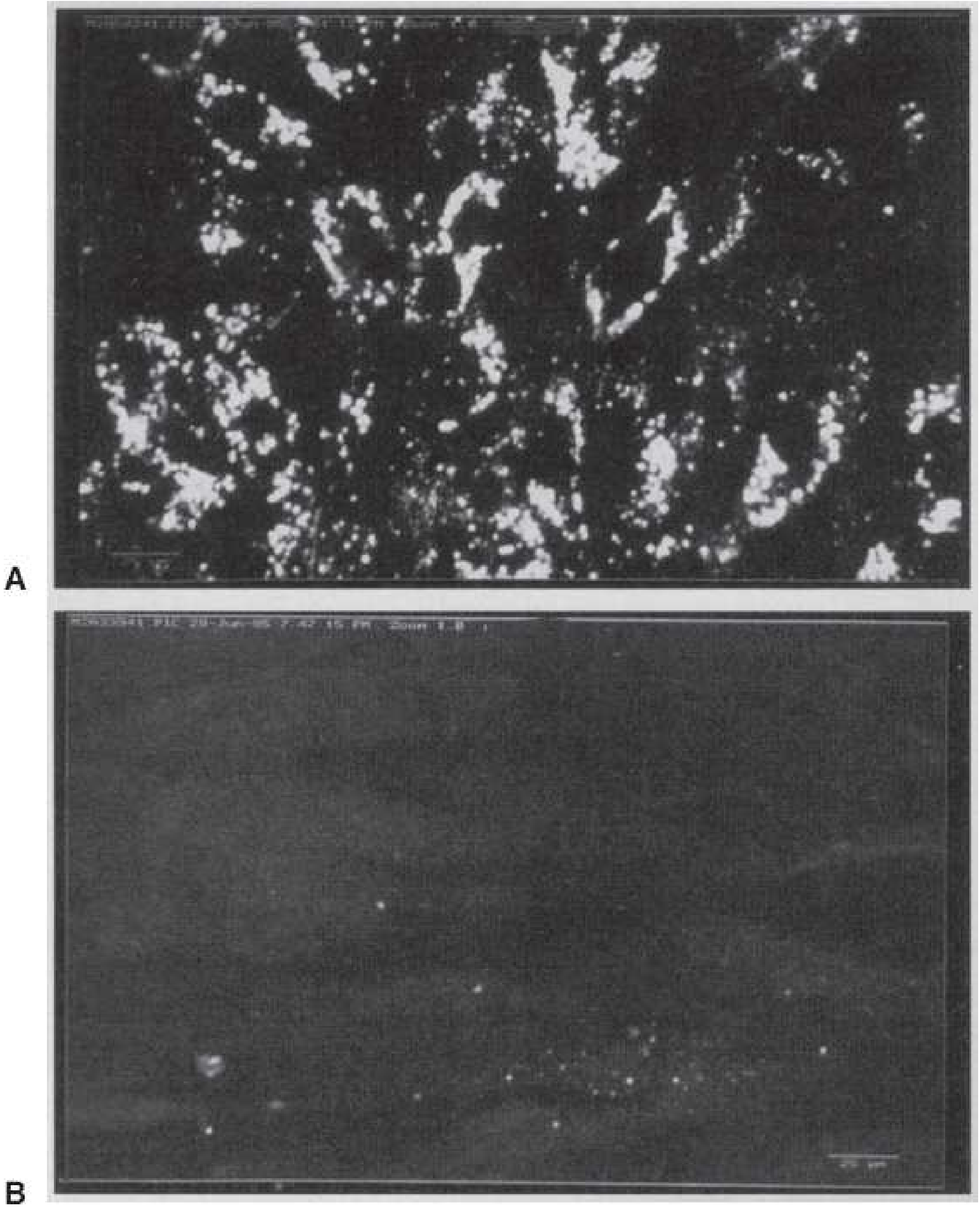

FIGURE 2.14–3. Indirect immunofluorescence staining of cultured fibroblasts with an anti-GM2 antibody in a patient with Tay–Sachs disease (*top*), showing significant accumulation of GM2 gangliosides compared to a healthy control (*bottom*). (From Sakuraba H, Itoh K, Shimmoto M, Utsumi K, et al. GM2 gangliosidosis AB variant: clinical and biochemical studies of a Japanese patient. *Neurology*. 1999;52(2):372, Figure 1, with permission. doi: 10.1212/WNL.52.2.372. © American Academy of Neurology.)

Neuronal Ceroid Lipofuscinosis

The NCLs, sometimes referred to as Batten disease in childhood, are a group of neuronal storage disorders that are collectively the most common neurodegenerative disorders in childhood, with a frequency of 1 in 20,000 to 40,000. Accumulation of abnormal lipopigments with characteristic ultrastructure in lysosomes, particularly in neurons, characterizes NCLs. Mutations in at least 10 genes that encode for a range of proteins, all involved in lysosomal catabolism and recycling of proteins and lipids, have been shown to be responsible for this group of disorders. These include lysosomal enzymes and proteins (CNL1, 2, 5, 10, and 13), membrane-associated cytoplasmic proteins (CLN4 and 14), a range of transmembrane proteins (CLN3, 6–8 and 12) and a soluble lysosomal protein (CLN5). Lipopigment accumulation diffusely affects cortical and subcortical neurons, but particularly the hippocampus and entorhinal cortex, and is associated with significant glial activation that precedes rather than follows neuronal loss. Alteration in axonal and synaptic function due to ultrastructural change occurs across the NCLs.

Adult neuronal ceroid lipofuscinosis (ANCL), also known as Kufs disease, is generally inherited in a recessive fashion (with the rare dominant Parry disease [CLN4] being dominantly inherited) and makes up between 2% and 10% of all NCLs, and unlike childhood forms of the disease, adults rarely have retinal pathology. Dominant pedigrees have been shown to have pathogenic variants in the *DNAJC5* gene, and recessive pedigrees have variants in *CLN6* or *CTSF* and less commonly in *GRN* or *CLN5*. Symptoms most commonly appear at the beginning of the fourth decade but may be present early in the second decade. Two clinical forms of ANCL have been described: a progressive myoclonic epilepsy form associated with debilitating action myoclonus, photosensitivity and ataxia (type A), typically associated with pathogenic variants in *CLN6* or *DNAJC5*, and a neuropsychiatric/cognitive form (type B) characterized by dementia, psychiatric disturbance, and dyskinesia, with epilepsy usually absent, usually due to pathogenic variants in *CTSF*.

Psychosis occurs in up to 25% of patients, particularly in adolescent and early adult patients, with characteristic hallucinations, delusions and thought disorder, and occasionally catatonia. A mood disturbance is particularly common, and OCD has also been described. Patients present cognitively with slowing, attentional disturbance, and impaired new learning; older patients may present with a frontotemporal dementia-like picture or a Kluver–Bucy type syndrome.

Diagnosis traditionally rested on the identification of characteristic inclusions in brain or skin punch biopsy, but pathologic interpretation in peripheral tissues can be challenging and is being supplanted by molecular genetic diagnosis. MRI often shows cerebral and cerebellar atrophy, and callosal thinning, while SPECT shows regional cortical hypoperfusion. Treatment is symptomatic, although they may be sensitive to extrapyramidal side effects such as dystonia and neuroleptic malignant syndrome, as NCL patients may have muscle membrane pathology comorbid with changes in the CNS.

α-Mannosidosis Type II

AM is a recessively inherited and rare LSD that results from a deficiency of α-mannosidase, which causes lysosomal accumulation or undigested oligosaccharides, with a frequency of 1 in 500,000 individuals. AM is characterized by mild to moderate intellectual disability, hearing loss, skeletal changes, and recurrent infections secondary to immunodeficiency, with an indolent form (type II) occurring in the minority of patients who survive to adulthood. The deficiency of AM results in the intralysosomal accumulation of mannose-rich oligosaccharides and the formation of storage vacuoles in neuronal and glial cells, which impairs myelin formation. Neuropathologically, AM appears to initially affect myelinated structures before progressing to involve the neuronal body. In type II AM, the predominant clinical features are cerebellar ataxia, hearing loss, neuropsychological impairment, and retinopathy. Diagnosis rests on the demonstration of a pattern of urinary oligosaccharides and reduced enzyme activity in leukocytes or fibroblasts. MRI scanning shows periventricular T_2 hyperintensities and cortical and cerebellar atrophy (Fig. 2.14–4) and may show iron accumulation in the basal ganglia. Up to 25% of type II AM patients over the age of 15 develop clear psychiatric illness, predominantly a psychotic disorder characterized by delusions, hallucinations, and confusion. Generally, psychosis presents with neurologic manifestations, although it may rarely present as a prelude to frank neurologic disturbance. Like many LSDs, bone marrow or peripheral blood cell transplantation is the only current viable treatment option, although enzyme replacement therapy using velmanase-α, a recombinant form of the enzyme, may assist patients with mild to moderate disease.

Peroxisomal Disorders

The peroxisome is a subcellular organelle that plays a role in the breakdown of fatty acids, the degradation of hydrogen peroxide (released via oxidation of fatty acids), membrane phospholipid and cholesterol synthesis, and the metabolism of amino acids. Unlike lysosomes, the peroxisomes bud off from the endoplasmic reticulum. Disorders of biogenesis (formation) of the peroxisome are generally fatal during infancy. Defects in single enzymes of the peroxisome may also cause disease compatible with survival well into adult life, and at least one of these, X-linked adrenoleukodystrophy (X-ALD), the most common peroxisomal disorder, is known to be associated with psychiatric illness.

X-linked Adrenoleukodystrophy. X-ALD is an X-linked recessive disorder occurring in 1 in 20,000 births and is caused by mutations to *ABCD1*, the gene for a peroxisomal membrane protein that β oxidizes very–long-chain fatty acids (VLCFAs). This leads to the accumulation of saturated VLCFAs in brain white matter and adrenocortical cells predominantly, which impairs membrane stability, but oxidative stress and bioenergetic failure, particularly in mitochondria, are thought to play major roles in the pathophysiology of the disease. It predominantly affects males, although some female carriers can be affected. In adults, it presents in either a predominantly cerebral form (5% of cases), marked by an inflammatory demyelinating process, or an adrenomyeloneuropathic form (AMN, 45% of cases), in which neuronal dysfunction is predominantly a distal axonopathy affecting the dorsal columns and corticospinal tract. The majority of other presentations are of the childhood cerebral form, which is rapidly progressive over 2 to 3 years. Both adult forms also present with adrenocortical insufficiency, often indistinguishable from primary Addison disease. The adult cerebral form shows a predilection for neuropsychiatric presentations, although the neurobiology of this is unclear. Demyelinative changes are most prominent in the parietal and occipital cortex as well as the thalamus, callosum, and brainstem (Fig. 2.14–5), although more anterior cortical regions may be involved. MRI frequently shows symmetrical white matter hyperintensity that begins in the splenium of the corpus callosum and spreads into neighboring posterior regions on T_2-weighted imaging. At presentation, a significant proportion of adult-onset patients present with psychiatric disturbance, most commonly behavioral changes. Mania and affective psychosis appear to be the most common neuropsychiatric presentations, more so than schizophreniform illnesses, although the latter does occur. Early attentional and executive changes are also common. When white matter changes occur in a more anterior distribution, patients may even show a frontotemporal dementia-like presentation. Psychiatric presentations may precede motor or cognitive changes by some years, and as in other organic psychiatric illnesses, patients often show treatment resistance. The AMN form, the most common form of ALD among adults, was thought to affect only the peripheral nervous system; however, subtle cerebral manifestations of the disorder are often present, and the rate of depressive illness appears to be elevated at least twofold. Some patients may present with mood changes subsequent to adrenal insufficiency, which reverse with appropriate corticosteroid replacement therapy. AMN patients may also show subtle neuropsychological changes that affect the speed of processing and executive functioning. There is no primary treatment for X-ALD, although bone marrow transplantation has provided some stabilization in younger patients with early disease, and the oral administration of 4:1 glyceryl trioleate and glyceryl trierucate (“Lorenzo’s Oil”) normalizes plasma VLCFA levels but does not improve neurologic function in already symptomatic or adult patients. A current trial is examining the role of pioglitazone, an agonist of the PPARγ pathway of mitochondrial biogenesis which has been shown to halt axonal degeneration in a mouse model of the disease in X-ALD patients.

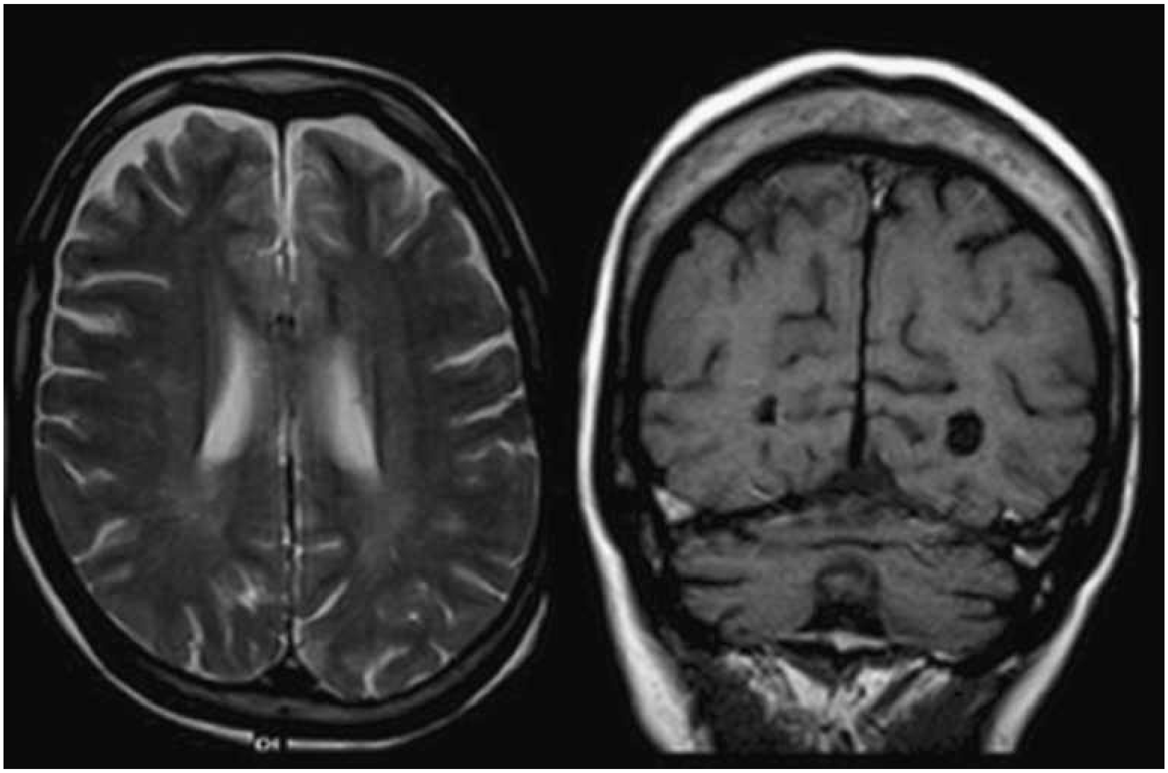

FIGURE 2.14–4. Magnetic resonance imaging scan on a patient with psychosis and α-mannosidosis. *Left*, T_2-weighted imaging shows parietal and occipital white matter hyperintensities and frontal atrophy. *Right*, coronal T_2-weighted imaging at the level of the cerebellum shows atrophy of the cerebellar hemispheres. (From Gutschalk A, Harting I, Cantz M, Springer C, Rohrschneider K, Meinck HM. Adult α-mannosidosis: Clinical progression in the absence of demyelination. *Neurology*. 2004;63(9):1744, Figure 1, with permission. doi:10.1212/01.WNL.0000143057.25471.4F. © American Academy of Neurology.)

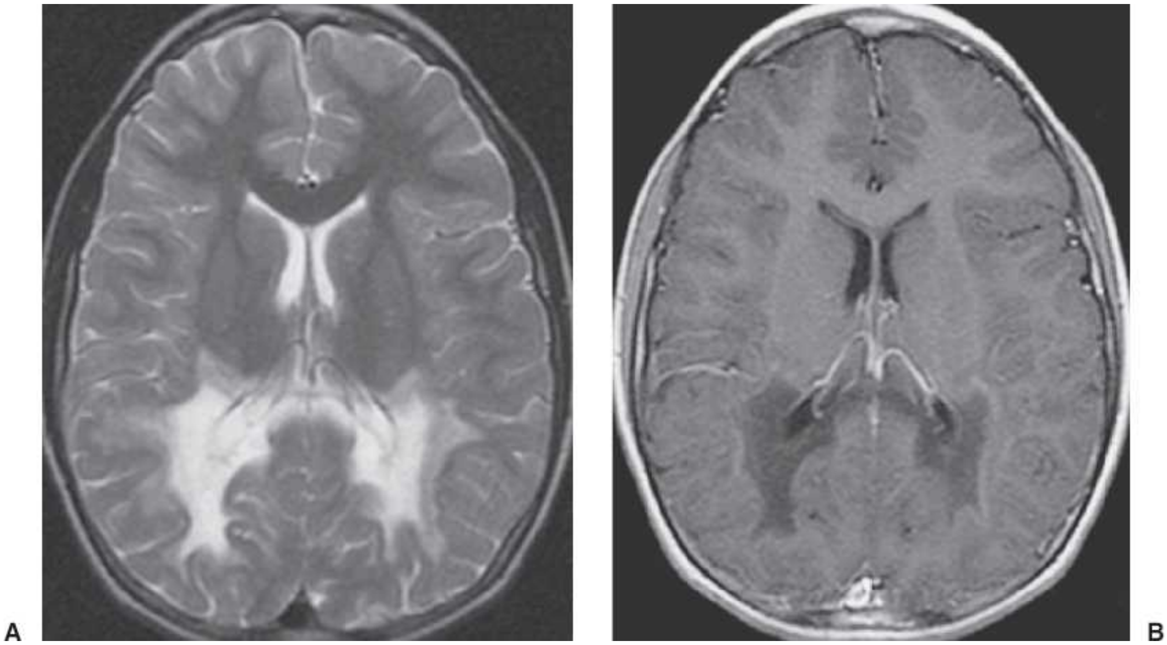

FIGURE 2.14–5. T_2-weighted (**A**) and T_1-weighted (**B**) axial magnetic resonance images (MRI) from an adolescent with adrenoleukodystrophy who presented with confusion and psychosis. Symmetrical and confluent hyperintensity is seen in the posterior white matter of both hemispheres, a typical MRI finding. (From Hesselink JR: Differential diagnostic approach to MR imaging of white matter diseases. *Top Magn Reson Imaging.* 2006;17(4):243, Figure 19, with permission. doi: 10.1097/01.rmr.0000248666.91834.af.)

Other Enzyme Deficiency Disorders

Acute Intermittent Porphyria. AIP is one of the porphyria disorders group, where defects in heme biosynthesis result in excessive secretion of urinary porphyrins and their precursors; a number of heme synthesis pathway intermediates are neurotoxic. Most of the porphyrias are inherited in a dominant fashion. The incompletely penetrant, autosomal-dominant AIP results from defects in the enzyme porphobilinogen (PBG) deaminase, which speeds the conversion of PBG to hydroxymethylbilane, and occurs in 1 in 20,000 individuals. Deficient activity of this enzyme results in the accumulation of porphyrin precursors PBG and aminolevulinic acid (ALA), which are thought to be neurotoxic. This enzymatic deficit becomes apparent in situations that require an increase in heme synthesis, including fasting, menstruation, intercurrent medical illness, and drugs that induce the cytochrome P450 system, such as alcohol estrogens, barbiturates, and sulfonamides, and presents most commonly in women of child-rearing age. Coproporphyria, a dominantly inherited deficiency of coproporphyrinogen oxidase, can present identically to AIP in addition to cutaneous findings (bullous lesions, skin fragility, and scarring).

The periodic “madness” of King George III has in recent decades been considered to be secondary to AIP, in addition to being implicated as illnesses affecting the work of Van Gogh and Soren Kierkegaard. AIP has been shown to be significantly overrepresented in a sample of 4,000 psychiatric inpatients (1 in 500 compared to a community rate of 1 in 100,000), emphasizing its predilection for psychiatric illness. The “classical triad” consists of abdominal pain, psychiatric disturbance, and peripheral neuropathies (mostly motor and often mimicking Guillain–Barré syndrome) during episodes, although psychiatric symptoms alone may be the single presenting feature. Of clinically symptomatic cases, a psychiatric disturbance occurs in up to half of all cases, half of which are psychotic episodes, although depression, anxiety, and delirium may also be the main presenting symptoms. The intermittent “attacks” of neuropsychiatric disturbance may result in a misdiagnosis of schizophrenia; additionally, the high likelihood of abdominal pain during attacks can result in a diagnosis of a somatoform disorder. More chronic porphyrias have been associated with elevated rates of anxiety and depression. How PBG and ALA accumulation causes neuropsychiatric disturbance is unclear. Explanatory hypotheses have included oxidative stress, vascular change, and demyelination, although it may be that ALA’s structural similarity to γ-aminobutyric acid (GABA) results in an impaired release of GABA from synapses of GABAergic inhibitory neurons and reductions in heme-dependent enzymes causing increased serotonin turnover and reduced nitric oxide activity.

Diagnosis of AIP rests on the demonstration of elevated urinary ALA and PBG and red cell PBG deaminase activity; for coproporphyria, urinary coproporphyrin levels should be assessed. Gross elevations of urinary ALA and PBG will often turn urine amber or purple in direct sunlight; it is imperative to collect specimens carefully (such as a 24-hour collection of urine in light-protected containers and correctly preserved). Management involves correct identification and avoidance of precipitants if possible. During an attack, treatment includes the reversal of contributing illnesses, intravenous hydration, carbohydrate loading to inhibit heme synthesis, and hematin or heme arginate to provide negative feedback to the heme synthetic pathway. Psychopharmacological management of AIP involves judicious use of medication that will not worsen the biochemical deficit, which for psychosis includes chlorpromazine (Thorazine) and droperidol (Inapsine), fluoxetine (Prozac) for depression, lithium for mania, and lorazepam (Ativan), triazolam (Halcion), and temazepam (Restoril) for anxiolysis and sedation.

Phenylketonuria. Phenylketonuria (PKU), an autosomal recessive disorder with an incidence of 1 in 10,000 to 20,000, is caused by mutations in the gene for phenylalanine hydroxylase (*PAH*) that converts the amino acid phenylalanine (PhE) to tyrosine. Mutations in both copies of the gene for *PAH* result in inactive or deficient enzyme levels, and accumulated PhE is converted to phenylketones, which are detectable in the urine. Resultant low levels of tyrosine, the precursor for the monoamines dopamine and serotonin, causes monoaminergic depletion in the CNS and severely disrupt normal neurodevelopment. Additionally, reduced large neutral amino acid (LNAA) availability causes reduced protein and cholesterol synthesis, impaired synaptogenesis and myelination, and altered glutamatergic neurotransmission. Hypomyelination occurs in untreated and early-treated patients; increased T_2 signal on MRI scanning is seen in most patients in frontal but particularly parieto-occipital zones with magnetic resonance spectroscopy (MRS) demonstrating increased cerebral PhE that tightly correlates with serum PhE levels. Functional MRI demonstrates altered connectivity within and between prefrontal cortical regions; diffusion tensor imaging (DTI) has also demonstrated impairments in white matter integrity, and these appear to underpin the executive deficits seen in PKU patients.

PKU sufferers require an individually tailored diet of foods low in PhE and supplemented with tyrosine. Routine screening of newborns and the universal initiation of dietary restriction within the first 3 months of life have essentially eliminated the severe, irreversible cognitive deficits, and behavioral disturbance associated with untreated PKU. A reduced PhE diet maintained for the first decade of life is associated with essentially normal cognitive outcomes, although some controversy exists as to whether relaxing dietary restrictions in the preadolescent phase (“early treated” patients) is associated with greater psychiatric disturbance than in PKU sufferers in whom dietary control is not relaxed (“consistently treated” patients). Early treated patients are described as showing elevated rates of depression, anxiety disorders (particularly agoraphobia), attention-deficit disorder, and more nonspecific psychosocial adjustment issues in adolescence when compared to matched healthy individuals. Compared to individuals with chronic illnesses such as diabetes, they show elevated rates of anxiety disorders, suggesting that this is unlikely to be an effect of adjustment to chronic illness. Elevated rates of psychiatric and cognitive illness appear to directly negatively affect psychosocial outcomes, with a higher proportion of PKU patients remaining unmarried, childless and/or living with their parents than their unaffected peers. Cognitive impairment also relates directly to the degree of past treatment, with untreated patients showing intellectual disability and developmental delay; early-treated patients may show subtle reductions in IQ and executive and attentional impairment. Prefrontal dopaminergic neurons are particularly vulnerable to decreased tyrosine availability, and mild elevations in PhE to tyrosine ratios (a marker of dopamine availability) in these patients in whom a dietary restriction is not maintained are associated with executive impairment and may underpin the adolescent and early adulthood psychiatric disturbance seen in these patients. Acute tyrosine depletion in healthy adults is associated with depression, anxiety, and executive disturbance and may provide a model for psychiatric and cognitive disturbance as a result of a prefrontal hypomonoaminergic state as a result of elevated PhE. Additionally, impairment of oligodendrocyte function may impair myelination in PKU (Fig. 2.14–6). Frontotemporal myelination, in addition to the dopaminergic innervation of the prefrontal cortex, peaks during adolescence, and the disruption of these processes may result in neuropsychiatric illness in this patient group. Animal models show that hypomyelination has been partially restored with dietary control, and some human studies show an improvement of psychiatric, cognitive, and imaging parameters with improved dietary control—even in individuals who have been “off diet” for years. However, perhaps less than 20% of adults maintain dietary control throughout adulthood, with the impact of diet on social interactions being the most commonly reported impediment for patients; discontinuation in adulthood is associated with poorer social, educational, and health outcomes, but re-engagement of “lost” patients may allow for some reversal.

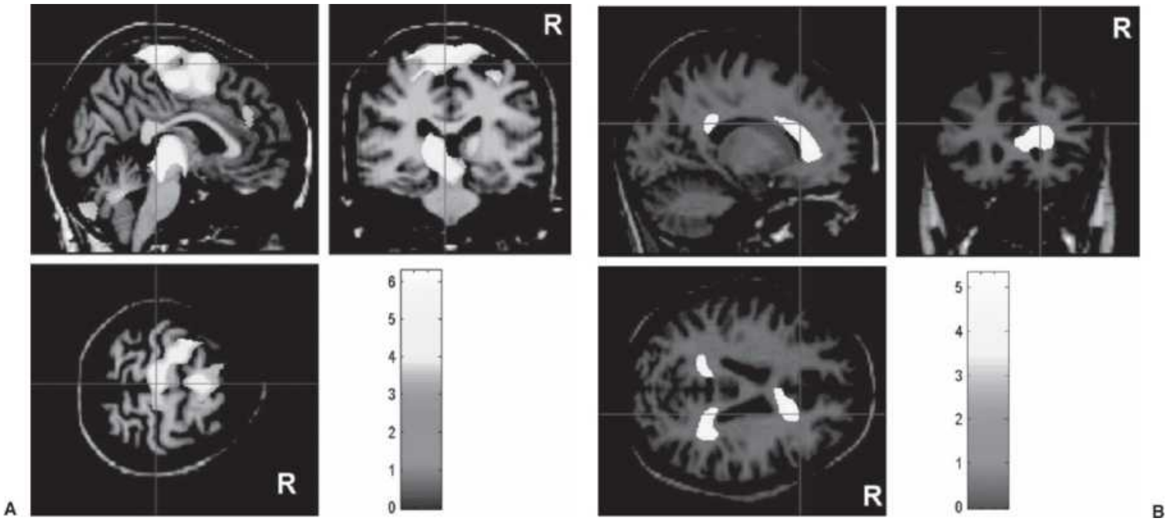

FIGURE 2.14–6. Results of voxel-based analysis of magnetic resonance image changes in phenylketonuria. **A:** Statistical parametric map of gray matter volume reduction of phenylketonuria (PKU) patients compared to control in motor and prefrontal cortex and thalamus. **B:** White matter volume reductions in the same comparison in anterior and posterior forceps of the corpus callosum. R, right side. (From Pérez-Dueñas B, Pujol J, Soriano-Mas C, Ortiz H, et al. Global and regional volume changes in the brains of patients with phenylketonuria. *Neurology*. 2006;66(7):1074, Figure 1, with permission. doi: 10.1212/01.wnl.0000204415.39853.4a. © American Academy of Neurology.)

Urea Cycle Disorders. UCDs result from aberrant metabolism of waste nitrogen, created by the breakdown of proteins and related nitrogenous molecules; the cycle is dependent on six co-ordinated enzymes and two transporters within the mitochondria and the cytosol and are found in hepatocytes only. They are recessive disorders, other than OTC deficiency (see below), which is an X-linked disorder; their prevalence varies from 1 in 10,000 to 35,000. While a third of patients present early in life, the remainder present outside the postnatal period, and adult presentation—particularly with neuropsychiatric illness—is not uncommon. The cardinal biochemical manifestation is hyperammonemia, and in the neonatal period—when the deficit is usually severe—this may present with somnolence and failure to feed, progressing to seizures and cerebral edema. When the biochemical deficit is milder, it may present only in situations of metabolic stress (such as systemic infection or pregnancy) and hyperammonemia when milder is associated with vomiting, headache, lethargy, sleep disturbance, and a range of neuropsychiatric symptoms, including agitation, psychosis, and confusion.

In very mild patients, the diagnosis is often missed, but a clue can be the avoidance of a high-protein diet, with some patients becoming vegetarian due to the association between protein intake and mild encephalopathic states, or the use of valproate which can trigger hyperammonemia in mild UCD patients. Acutely, sodium phenylbutyrate or sodium benzoate is used to scavenge ammonia; and chronically, UCD patients utilize urea cycle precursors such as arginine and citrulline to reduce ammonia levels.

Acute hyperammonemia results in neuronal damage akin to that seen in hypoxia or stroke. The insular cortex is vulnerable early, with other cortical regions affected with prolonged hyperammonemia; after severe decompensation, occasionally gliosis and atrophy occur. On neuroimaging in less severely affected adults, T_2-weighted imaging may show hyperintensities and reductions in white matter integrity detected using DTI. Even mildly or minimally affected adults may develop executive and/or attentional deficits in the setting of mild and/or intermittent hyperammonemia.

Psychiatric symptoms are common in UCDs, and often present comorbid with neurologic symptoms such as ataxia, lethargy, and seizures. Agitation and hallucinations are common during hyperammonemic states; in patients with mild disease, when not grossly hyperammonemic, patients may be diagnosed with anxiety, depression, anorexia (often driven by protein avoidance) or somatization.

Ornithine transcarbamylase (OTC) deficiency is the most common UCD and is an X-linked disorder that usually presents in young male children up to 5 years of age with encephalopathy, ataxia, and focal neurologic signs; in heterozygous females, it often only presents during periods of metabolic compromise and high protein diet, and females are often asymptomatic for years, making diagnosis difficult. It is not uncommonly diagnosed as postpartum psychosis in females.

A 33-year-old married mother of two children (7 and 2) and a graduate nurse was diagnosed in childhood with OTC deficiency following maternal diagnosis. She had two significant episodes of hyperammonemia during pregnancy; in the first, she suffered from confusion, headache, and vomiting in the setting of mild hyperammonemia (130–160 µmol/L); in the second, with more significant hyperammonemia (>220 µmol/L), she required an intensive care admission, a PEG for amino acid infusions due to persistent vomiting, and then developed an acute psychotic episode which resolved without antipsychotic treatment. She was, however, referred in the setting of persistent anxiety, which began in her mid-teens and which had been treated with serotonergic antidepressants since the age of 18 at slowly escalating doses. Because of the association between food and the symptoms she experienced during metabolic crises, she at times became anxious about food intake and would struggle to maintain food intake and manage her citrulline and arginine supplementation. Her anxiety appeared to worsen at times when her ammonia levels were slightly increased (50 to 100 µmol/L). She ultimately improved with the addition of pregabalin to her antidepressant treatment, which allowed her to resume more regular supplementation and lower her ammonia level.

Carbamoyl phosphate synthetase I (CPSI) deficiency is a recessive disorder affecting the first enzyme in the cycle, located in the mitochondria that commonly presents in infancy with severe encephalopathy progressing to coma; however, milder forms present later in childhood or adulthood and may be asymptomatic until triggered by metabolic stress.

Citrullinemia (CTL1) is also recessive and caused by a deficiency in arginosuccinate synthetase (ASS). It similarly presents severely in neonates with typical findings of a UCD but can present as an asymptomatic illness punctuated by encephalopathy during times of metabolic stress in mildly affected adults. It has been known to present with schizophrenia-like psychosis but also as psychosis during pregnancy.

Maple Syrup Urine Disease. Maple syrup urine disease (MSUD) is an autosomal recessive disorder caused by a defect in the branched-chain α-ketoacid dehydrogenase enzyme complex, with resultant abnormalities in branched-chain amino acid (BCAA) catabolism. In adult-presenting disease, altered leucine metabolism results in impaired glial function, reduced synaptogenesis and dendritic arborization, impaired myelination and the induction of neuronal apoptosis. Additionally, elevated leucine reduces transport across the blood–brain barrier of large neutral amino acids, including most monoamine precursors, which results in disruptions to dopaminergic, noradrenergic, histaminergic, and serotonergic transmission. Altered BCAA metabolism also results in depletion of glutamate, GABA, and aspartate, which alters excitatory/inhibitory transmission throughout the CNS. Although quite rare, occurring in 1 in 185,000 individuals, the incidence is significant (1 in 200) in Amish and Mennonite populations. Clinical manifestations include body fluid odor that resembles maple syrup and overwhelming illness in the first week of life, beginning with vomiting and lethargy and progressing to seizures, coma, and death if untreated. In milder forms of the disease, the illness may manifest symptoms only during stress (such as infection or following surgery). MSUD is diagnosed by elevated plasma BCAAs, particularly leucine, and profound ketosis and acidosis. Long-term management is via restriction of dietary BCAAs, although small amounts are required for normal metabolic function. Decompensation can be prevented in some patients via liver transplantation. Like PKU, the advent of dietary restriction has modified the illness course significantly, and MSUD patients who survive into adulthood have demonstrated subtle cognitive deficits and an elevated rate of some neuropsychiatric disorders. Like PKU, adult neuropsychological impairment appears more related to age at the institution and adherence to treatment rather than persistent BCAA levels. Whereas in PKU, the relative tyrosine deficiency results in altered myelination and monoaminergic transmission, in MSUD, the BCAA-restricted diet results in chronic cerebral valine deficiency, which itself may impair neuronal and oligodendrocyte function. Similar to PKU, however, hyperleucinemia inhibits the transport of tyrosine and tryptophan, which results in reduced monoamine production, and the ketoacid derivative of leucine causes glutamate depletion. Bioenergetic and osmotic dysregulation also occurs, particularly in encephalopathic states, which results in cerebral edema. Despite strict metabolic control, many children suffer from attention-deficit disorders, whereas adolescents and adults commonly present with depression and anxiety. These are described as responding to psychostimulants and antidepressants, respectively.

Cognitive impairment is usually seen in executive and attentional domains.

A 32-year-old partnered mother of an 11-year-old child was referred for assessment of mood, anxiety, and cognition. She worked in a factory steaming second-hand clothes and making manila folders 3 days per week. She was diagnosed with MSUD at the age of 17 days old in the setting of persistent vomiting and failure to thrive, spent the first 3 months of her life in hospital care, was managed on a low BCAA diet throughout childhood, and suffered epilepsy throughout her development. She had struggled with dietary compliance in adulthood and frequently presented to the hospital emergency department with nausea, abdominal pain and vomiting, and at other times with obtundation and seizures. In between encephalopathic episodes, she suffered from significant worsening anxiety and panic, triggered by social conflict, medical procedures and enclosed spaces, and road travel. Neuropsychological assessment revealed mild intellectual disability (full-scale IQ of 62) with particular difficulties in memory, executive function, and attention. She was prescribed escitalopram syrup but struggled to remain compliant.

Cerebrotendinous Xanthomatosis. Cerebrotendinous xanthomatosis (CTX) is an autosomal recessive disorder of cholestanol metabolism caused by mutations in the sterol-27 hydroxylase gene (*CYP27A1*) on chromosome 2. It is exceptionally rare, with a frequency of 1 in 1 million, but is much more frequent in the Moroccan Jewish population. CTX was first described in 1937 and is also known by the eponymous name of van Bogaert disease. Since then, over 300 patients with CTX have been described, and about 50 mutations have been identified in the *CYP27A1* gene, just under half the mutations being missense mutations. Deficiency of the mitochondrial enzyme sterol 27-hydroxylase leads to reduced hepatic production of bile acid and reduced chenodeoxycholic acid (CDCA) production. The absence of CDCA-driven negative feedback on bile acid synthesis results in the accumulation of cholestanol precursors, increased cholestanol and 7-hydroxycholesterol production. Accumulation of cholestanol and cholesterol in the eyes, tendons (often the Achilles, but also elbow/hand extensor tendons, patellar and neck tendons), and brain

—particularly the cerebellar hemispheres (Fig. 2.14–7) lead to the classic clinical triad of juvenile cataracts, adult tendinous xanthomas, especially of the Achilles tendon, and progressive neurologic impairment. The neurologic manifestations are due to the deposition of large granulomatous lipid deposits in basal/cerebellar nuclei and cerebral white matter or to demyelination and include peripheral neuropathy, intellectual disability, seizures, cerebellar ataxia, pyramidal signs, and dementia. Early-onset illness tends to present with behavioral and personality changes associated with intellectual disability, whereas later-onset illness is characterized by neuropsychiatric illness associated with frontal, frontotemporal, or orbitofrontal syndromes. Psychiatric presentations include depression, psychosis, and personality or behavioral change. Case series of CTX patients with psychiatric disturbance have described 3 of 35 cases as suffering a neuroleptic-responsive psychotic disorder; 7 of 10 CTX patients demonstrated agitation and psychosis; and a series of 13 patients at a single site found 6 presented with an organic psychiatric presentation, usually associated with dementia. A review of 425 published CTX cases found that 54 presented with neuropsychiatric illness, and in a series of 78 patients, 10 had a confirmed retrospective diagnosis of an autism spectrum disorder (ASD)—suggesting that ASD symptoms may be present in childhood and herald later neuropsychiatric illness. Accelerated cardiovascular disease can result in strokes, further impacting cognitive and psychiatric function.

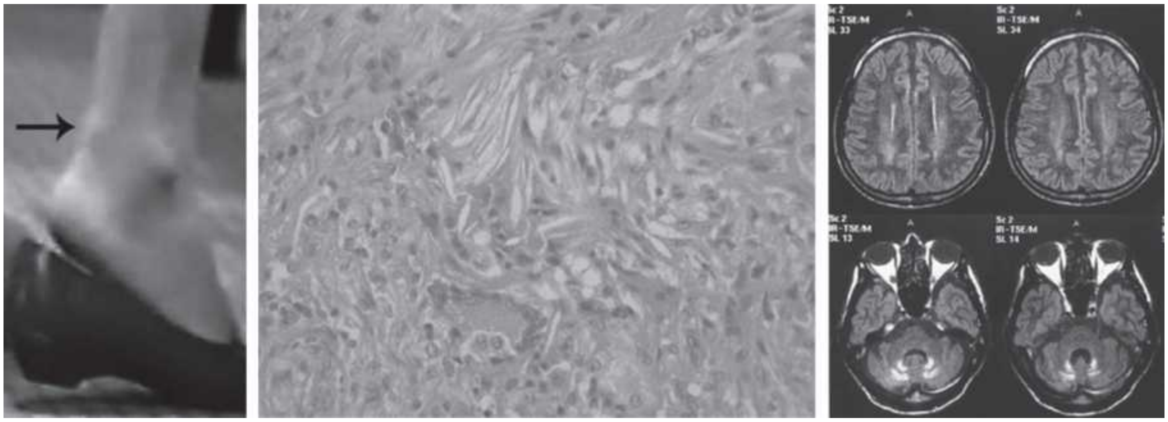

FIGURE 2.14–7. Cerebrotendinous xanthomatosis. *Left*, image of left ankle of a woman described in text demonstrating Achilles tendinous xanthoma. *Middle*, hematoxylin and eosin stain of xanthoma of a 42-year-old woman with early-onset dementia and behavioral changes, showing cholesterol crystals. (From Wang Z, Yuan Y, Zhang W, Zhang Y, Feng L. Cerebrotendinous xanthomatosis with a compound heterozygote mutation and severe polyneuropathy. *Neuropathology*. 2007;27(1):62, with permission. doi: 10.1111/j.1440-1789.2006.00739.x. © Japanese Society of Neuropathology.) *Right*, transverse fluid-attenuated inversion recovery magnetic resonance imaging on two adult cerebrotendinous xanthomatosis patients showing mild hyperintensity of subcortical white matter and significant hyperintensity of the dentate nucleus of the cerebellum. (From De Stefano N, Dotti MT, Mortilla M, Federico A. Magnetic resonance imaging and spectroscopic changes in brains of patients with cerebrotendinous xanthomatosis. *Brain*. 2001;124:121, with permission of Oxford University Press. doi: 10.1093/brain/124.1.121.)

A 47-year-old woman with a 2-year history of treatment-resistant depression, unusual behavior, and memory problems was referred for neuropsychiatric assessment. On admission, she fluctuated between periods of withdrawal and noncommunicativeness and periods of joviality associated with child-like comments and fatuous affect. She was unable to provide a detailed history and responded in a stereotyped way with single sentences: “I am confused,” “This is what I am like at the moment.” Cognitive assessment was not possible. Her past history included cataracts and a hip fracture. Physical examination revealed enlarged, thickened Achilles tendons bilaterally (Fig. 2.14–7). Neurologic examination revealed primitive reflexes and clonus in the right ankle, together with a right plantar response. A diagnosis of cerebrotendinous xanthomatosis was made on the basis of the clinical picture and elevated cholestanol levels. Treatment with chenodeoxycholic acid was instituted.

The diagnosis of CTX in adults should be suspected when the clinical picture includes cataracts, xanthomas, chronic diarrhea, and progressive neurologic or cognitive impairment. In the presence of the typical clinical picture, the identification of elevated plasma cholestanol levels with low or normal cholesterol levels plus reduced urinary excretion of bile alcohols is usually sufficient to confirm the diagnosis. MRI using fluid-attenuated inversion recovery (FLAIR) sequences can identify bilateral hyperintensities of the cerebellar dentate nuclei that mirror the known neuropathologic site of disease involvement (Fig. 2.14–7), usually associated with folia atrophy and lesions in the brain stem and periventricular regions. Global gray and white matter volume is frequently reduced. Treatment consists of lifelong replacement of CDCA (750 mg per day, often combined with a statin) and is aimed at limiting the long-term damage caused by cholestanol and cholesterol deposition, which improves neurologic and neuroradiologic markers of disease, particularly when initiated at an early stage. Treatment with CDCA alone has also been shown to result in an improvement in psychiatric symptoms and cognition in CTX patients.

Tetrahydrobiopterin Deficiency. Tetrahydrobiopterin deficiency (THD) is an extremely rare recessive disease (1/ 500,000 to 1,000,000 newborns) caused by the mutation of several genes, in particular GCH1, PCBD1, SPR, PTS, and QDRP, which encode enzymes involved in the production and recycling of tetrahydrobiopterin (BH4). BH4 is an essential cofactor for numerous enzymes involved in the synthesis and maintenance of monoaminergic neurotransmitters (via tyrosine-3-hydroxylase and tryptophan-5-hydroxylase), the regulation of nitric oxide synthesis, and the modulation of glutamatergic transmission. It is also a cofactor for the enzyme PAH, impairments that cause PKU. Mutations in MTHFR and DHFR genes can also cause a mild deficiency. Impaired enzymatic function leads to altered levels of PhE and central monoamines. PhE hydroxylase, tyrosine hydroxylase, and tryptophan hydroxylase all require BH4 as a cofactor. Inherited defects that reduce the concentration of BH4, therefore, lead to PKU and to reduced CNS levels of dopamine and serotonin, as tyrosine hydroxylase and tryptophan hydroxylase are enzymes required for the synthesis of these neurotransmitters.

Classical symptoms are progressive and include intellectual disability (IQ between 60 and 90 in most patients), movement disorder, dysphagia, seizures, developmental delay, and behavioral disturbance in younger patients. Various psychiatric symptoms (including depression, anxiety, psychosis, obsessive-compulsive features, impulsivity, and attention-deficit disorder), as well as sleep disturbance, are reported infrequently in all of the THDs. However, psychiatric and behavioral problems are likely underdiagnosed, except in sepiapterin reductase deficiency, where they were documented in around 45% of patients. BH4 deficit has been associated with depression and psychosis, and reduced BH4 levels have been demonstrated in both schizophrenia and schizoaffective disorder, and supplementation with the BH4 analog sapropterin is being investigated as a potential therapeutic agent in psychotic disorders.

Hyperprolinemia. Hyperprolinemia (HP) is a recessive disease caused by mutation of either of the two genes, *ALDH4A1* and *PRODH*. These genes code for pyrroline-5-carboxylate dehydrogenase and proline oxidase, respectively, which are implicated in the degradation of proline. Proline is converted to glutamate, and HP thus alters glutamatergic transmission, alters synaptic activity, and thus impacts cognition. HP, which has no definitive treatment, is divided into two subtypes based on the severity of biochemical deficit; HP2 (with proline blood levels elevated

10–15X) is considered more severe in patients presenting in childhood with seizures and mild intellectual disability. In HP1 (proline levels elevated 3–10X), patients have few symptoms or may be asymptomatic—rendering the true frequency of HP difficult to determine. Neuroimaging is often normal. There has long been a speculative association between HP1 and psychosis; the 22q11 deletion, known for its strong association with schizophrenia, results in a hemizygous deletion of the PRODH gene, and up to 50% of patients with the 22q11 deletion have elevated plasma proline levels. In a recent study, schizophrenia patients demonstrated a mild to moderate increase in serum levels of proline when compared to sex and gender-matched controls, and in a large study of patients with major psychiatric illness, HP due to missense mutations or deletions in *PRODH* was found to be a risk factor for schizoaffective disorder (OR = 4.6). A recent systematic review of 35 case reports and case-control studies found a common psychiatric phenotype of developmental delay, intellectual disability, autism spectrum and psychotic disorders.

Propionic Acidemia. Propionic acidemia (PA), also called ketotic glycinemia and propionic aciduria is a recessive disorder affecting 1:100,000 births, caused by the deficiency of activity of the enzyme propionyl CoA carboxylase, a mitochondrial biotin-dependent enzyme that is essential for the catabolism of the amino acids threonine, methionine, isoleucine, valine, as well as cholesterol and some fatty acids. This results in a buildup of organic acids in blood and tissues, in particular propionic acid. Patients with early-onset PA present with acute metabolic decompensation in childhood. Late-onset PA, still largely a childhood disorder, is more heterogeneous and may fluctuate (particularly in times of increased metabolic demand), with symptoms including vomiting and feeding difficulties with impairment in psychomotor development and movement disorders such as choreoathetosis and dystonia.

Acute acidosis with encephalopathy and hyperammonaemia may occur, during which brain MRI often demonstrates cortical edema and signal changes in the basal ganglia. It may rarely present in the third or fourth decade of life. Diagnosis requires urine organic acid analysis and a plasma or blood spot acylcarnitine profile, and treatment is largely symptomatic. PA has been associated with adolescent or young-adult psychosis, isolated visual hallucinations, associated with metabolic decompensation, and with ASD in preadolescent patients. In pediatric patients, it has been associated with ASD-type presentations in a significant majority of patients. Treatment with antipsychotics may require careful consideration as patients may have long QT syndrome.

AMACR Deficiency. Alpha-methylacyl-CoA racemase (AMACR) deficiency is an exceptionally rare and recently discovered recessive disorder caused by mutations in the *AMACR* gene and has no specific treatment. The AMACR enzyme is found in mitochondria and peroxisomes. In peroxisomes, lack of AMACR leads to elevation of pristanic acid, a fatty acid usually supplied from meat and dairy foods in the diet. The role of AMACR in mitochondria is felt to relate to the breakdown of molecules derived from pristanic acid. Accumulation of pristanic acid influences functions of brain cells, leading to neuronal and glial cell death and leads to alterations of brain structure with encephalopathy, progressive demyelinating neuropathy and xanthomas with sterol storage in the CNS. Diagnosis is made by demonstrating elevated pristanic acid in plasma. The disorder can present across the lifespan but usually presents in early adulthood with cognitive decline, seizures, and migraine; acute decompensations may present as stroke-like episodes. More chronic neurologic signs, such as ataxia, spasticity, and sensorimotor neuropathy, may also present, and it has been associated with late-onset ataxia presenting in the 5th decade. MRI tends to demonstrate increased T_2 signal in cerebello-thalamic pathway structures.

Disturbances in fatty acid and oxidative phosphorylation pathways have been described in brain regions of patients with schizophrenia, including the prefrontal cortex. The AMACR gene is also located on chromosome 5p13, a locus that has been shown to be a putative schizophrenia susceptibility locus in a large sample of patients from Puerto Rico with 3 missense variants segregating with schizophrenia. A larger study of exome sequences of multiplex and multigenerational families has also implicated the AMACR gene in schizophrenia.

Cerebral Creatine Deficiency Syndrome. The cerebral creatine deficiency syndromes (CCDS) encompass two autosomal-recessive creatine biosynthesis disorders: GAMT deficiency and L-arginine:glycine amidinotransferase (AGAT or GATM) deficiency, and the X-linked creatine transporter (SLC6A8) deficiency. Intellectual disability and seizures are common in all three CCDS. Creatine plays a key role in cellular energy availability through the regeneration of ATP and is partly synthesized endogenously and supplemented by dietary intake. Creatine is synthesized from L-arginine, glycine, and methionine through a two-step process involving AGAT and GAMT. When synthesized, creatine is then transported to tissues with a high-energy demand via the creatine transporter (SLC6A8). A diagnosis is suggested by a reduction or absence of the creatine peak visualized by proton MRS and increased urinary creatine excretion.

Almost all patients with GAMT deficiency have a behavioral disorder that may include autistic behavior and self-mutilation associated with pyramidal and extrapyramidal signs, with onset usually between the ages of 3 months and 3 years. Patients present with mild intellectual disability and speech delay to severe intellectual disability, seizures, and behavioral disorders that may become more marked during the course of the disease. AGAT deficiency usually presents as intellectual disability, autistic behavior and severe language disorder without epilepsy. Autistic behavior is also a common feature in SLC6A8, although presentation tends to occur later in childhood or even in adulthood. SLC6A8 deficiency is the cause of approximately 2% of X-linked intellectual disabilities. Clinical characteristics include psychomotor and expressive speech delay, autism, epilepsy, and muscle hypotonia and hypotrophy. Semantic pragmatic language disorders and oral dyspraxia have been described. CCDS patients may also present with a choreoathetoid movement disorder associated with dystonia and/or ataxia, and patients may demonstrate abnormal signal intensities in the basal ganglia with or without myelination changes on MRI although imaging may also be normal. Patients are often treated with supplementation with creatine monohydrate (and, in GAMT, L-ornithine), although SLC6A8 deficiency has no treatment.

Glycine Encephalopathy. Glycine Encephalopathy, also known as nonketotic hyperglycemia (NKH), is caused by mutation of the glycine ecarboxylase (GLDC, 80% of NHK) or aminomethyltransferase (AMT, 10% to 15%) genes; its prevalence is estimated at 1:75 000 newborns. These genes produce proteins implicated in the glycine cleavage enzyme complex that is responsible for breaking down glycine. A dysfunction in this complex leads to a toxic level of glycine in various tissues. As glycine, an amino acid is also a neurotransmitter, the brain is a key site of toxicity. The classical form of glycine encephalopathy appears shortly after birth presenting with lethargy, feeding difficulties, hypotonia, abnormal movements and breathing and is not infrequently fatal. Outside the neonatal period, individuals may present with profound intellectual disability and treatment-resistant seizures. The chief clinical phenotype is intellectual deficiency and childhood ADHD. Due to the intrinsic role glycine plays in excitatory neurotransmission by modulating the activity of the *N*-methyl-D-aspartate (NMDA) subtype of glutamate receptors, it has been suggested that NKH theoretically could present with psychosis in older patients, although there are no published reports associating psychosis and NKH. Biologic diagnosis is made by glycine measurement in the CNS and confirmed by mutation analysis of the GLDC and AMT genes. Treatment is with sodium benzoate, which reduces serum glycine levels, and dextromethorphan (and ketamine) can be utilized to reduce excessive NMDA-mediated glutamatergic activity.

SSADH Deficiency. Succinic semialdehyde dehydrogenase (SSADH) deficiency is a rare recessive metabolic disorder with a nonspecific and variable phenotype characterized by reduced activity of the enzyme SSADH, involved in the degradation of GABA, leading to an accumulation of γ-hydroxybutyric acid (GHB) and coded for by the ALDH5A1 gene. It is suggested to occur in 1: 1,000000 individuals. GABA is the major inhibitory neurotransmitter in the brain (although it is excitatory in the embryo) and is derived from glutamate, the major excitatory neurotransmitter. Impairments to SSADH activity result in an alteration of the balance between excitatory and inhibitory neurotransmission. Diagnosis rests on the demonstration of elevated urinary GABA and leukocyte SSADH assay. MRI commonly shows T_2 hyperintensity in the pallidum, dentate nucleus and white matter, and in 2/3 of patients, EEG shows background slowing and spike discharges (Fig. 2.14–8).

Patients with SSADH deficiency present intellectual disability with deficits in expressive language, hypotonia, choreoathetosis, nonprogressive ataxia, and hyporeflexia. Neuropsychiatric symptoms are prominent and include sleep disorders, inattention, hyperactivity, autism-like behavior and anxiety or obsessive-compulsive symptoms. Symptoms occur before age 4 in most cases, although diagnosis has been made as late as the third decade. Developmental delay is present in all cases, and

10% of patients will present movement disorders such as chorea, myoclonus and dystonia. Delayed language is very common and frequently associated with intellectual disability. A recent study on a sample of transcortical sensory aphasia (TSA) patients (n = 187) found two patients with SSADH deficiency, suggesting enrichment of this disorder in some clinical populations. In a review of 37 cases of SSADH deficiency with psychiatric symptoms (70% of 53 patients), hyperactivity was present in 40% of patients, inattention in 50%, OCD in 35% and aggressive behavior in 15%. Four patients in this sample also had hallucinations. The majority of neuropsychiatric presentations were seen in older children and adults, and these symptoms tended to respond to stimulants, serotonergic medication and antipsychotics. The enzyme deficiency itself has no specific treatment, although trials of vigabatrin, an irreversible inhibitor of GABA transaminase (which thus inhibits the formation of succinic semialdehyde), led to improvement in some patients but worsening in others.

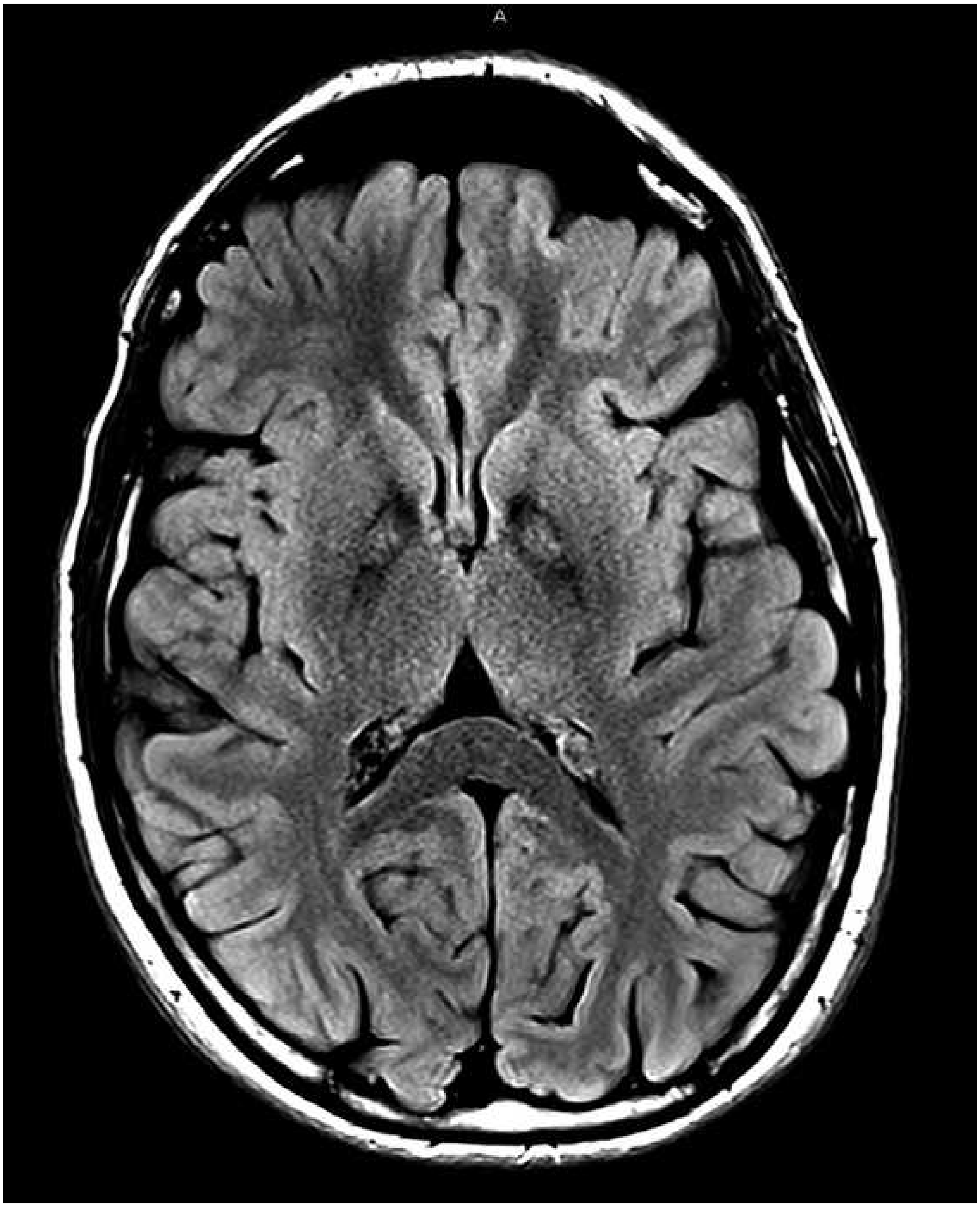

FIGURE 2.14–8. **SSADH**. Global pallidal lesions in SSADH, seen on axial T_2-weighted FLAIR image. A 26-year-old woman with SSADH deficiency, obsessive-compulsive disorder and seizures.

Serine Deficiency Syndrome. Serine deficiency syndromes (SDS) are a group of threee diseases implicating disturbance in the L-serine synthesis pathway. Three serine-deficiency syndromes have been described: 3-phosphoglycerate dehydrogenase (3-PGDH) deficiency, 3-phosphoserine phosphatase (3-PSP) deficiency, and phosphoserine aminotransferase (PSAT) deficiency. L-serine plays a key role in the development of the CNS. Its conversion to glycine provides a source of one-carbon units for methylation processes, nucleotide and phospholipid synthesis, and cell proliferation; it also facilitates the role of glycine and D-serine in excitatory neurotransmission through activation of NMDA receptors.

Neurodevelopmental disorders are associated with SDS, and clinical presentations differ according to the enzyme affected. PSAT deficiency patients present with acquired microcephaly, psychomotor retardation, treatment-resistant seizures, and hypertonia. PSH deficiency patients have congenital microcephaly and intrauterine growth retardation, feeding difficulties, and psychomotor retardation, and the disorder has been associated with Williams Syndrome. The infantile form of 3-PGDH deficiency presents akin to PSH deficiency, whereas the juvenile form is more insidious with absence seizures, moderate developmental delay and abnormal behavior. This last clinical form may be seen by child psychiatrists in adolescence with developmental delay and aggressive behavior with no clinical neurologic signs.

Diagnosis is made by demonstration of low plasma and CSF serine; treatment with L-serine (+/– glycine) improves seizures and behavioral disturbance but not developmental delay.

Cerebral Folate Deficiency. Cerebral folate deficiency is defined as any neurologic syndrome where CNS levels of 5-methyltetrahydrofolate are low in the presence of normal extra-CNS levels of folate. The two main causes are autoantibodies to the folate receptor that transports folate (also known as vitamin B_9) into the CSF or a mutation in the gene encoding this receptor; it has also been described in Aicardi and Rett syndromes and mitochondriopathies. Folate plays a key role in myelination, in the generation of biogenic amines and pterins, and is involved in the production and repair of DNA and gene expression. CFD is generally treated with folinic acid, which ameliorates symptoms if initiated early.

Cerebral folate transport deficiency (CFTD) is a rare recessive disorder caused by a mutation in the *FOLR1* gene encoding for the folate receptor FR1 that allows folate to enter the cell and is found on the basolateral endothelial surface of the choroid plexus. Impaired function of (or antibodies blocking) this receptor prevents the receptor-mediated endocytosis of folate and thus its entry into the CSF. Symptoms of CFTD do not begin until late infancy as until this period the folate receptor FR2 is responsible for folate transport before birth and in early infancy. Affected children have onset of the disease classically around age 2 with psychomotor delay and regression, speech difficulties, seizures and ataxia. Leukodystrophy is usually apparent on MRI scanning secondary to marked impairment of myelination. Folate deficiency may also interact with a polymorphism in the MTHFR gene to increase the risk of ASD in CFD patients.

Autoimmune cerebral folate deficiency (ACFD) is an acquired disorder with a wider spectrum of age of onset of disease and has been reported in children, adolescents, and rarely in adults. The vast majority are infants who develop symptoms after 4 months of age with irritability, sleep disturbance and then dyskinesia, ataxia and seizures. A number of patients present with autistic behavioral features, and in adolescents, it has been associated with catatonic schizophrenia. In these patients, FR1 and FR2 receptor encoding are normal, but autoantibodies develop in the receptors. It has been suggested that this may develop via induction by soluble folate-binding proteins in cow's milk in infants or by sensitization by other similar epitopes in older patients. In one sample of ASD patients, FR1 and FR2 autoantibodies were found in almost 75% of patients, and folinic acid treatment improved one-third of patients. Similarly, FR1/2 autoantibodies were detected in 80% of a small sample of treatment-resistant schizophrenia patients, a number of whom showed symptomatic improvement to treatment with folinic acid.

Other Neurometabolic Disorders

Niemann–Pick Disease Type C. Niemann–Pick type C disease (NPC) is an autosomal recessive neurovisceral disorder of lipid storage with a frequency of 1 in 100,000 live births, with 95% of sufferers having aberrations on the *NPC1* gene (18q11–12) and 5% on the *NPC2* gene (14q24.3). It is biochemically and phenotypically distinct from Niemann–Pick diseases type A and B, which result from a deficiency of lysosomal sphingomyelinase. *NPC1* and *NPC2* are involved in the cyclical movement of sterols within cells, and their impairment results in late endosomal accumulation of cholesterol and gangliosides; excess GM2 and GM3 gangliosides cause ultrastructural change to neurons, and disrupted cholesterol metabolism results in altered tau and beta-amyloid processing (with Alzheimer-like neurofibrillary tangles, NFTs), and disrupted calcium and metal metabolism. Axonal structures are particularly vulnerable and are affected early with subsequent involvement of cerebellar Purkinje cells, basal ganglia, and thalamic neurons, followed by hippocampal and neocortical regions later (Fig. 2.14–9). The most fulminant neuronal loss occurs in the cerebellum, causing early cerebellar signs. Diagnosis is usually confirmed by demonstrating a low esterification rate of exogenous cholesterol in fibroblasts or by testing for lysosomal accumulation of free cholesterol by filipin staining (Fig. 2.14–10); more recently, oxidized cholesterol levels in plasma and/or direct NPC1 sequencing have been utilized. NPC may present in infancy, adolescence, or adulthood with a clinically variable picture, although its core features include dementia, dysarthria, ataxia, vertical supranuclear ophthalmoplegia, and hepatosplenomegaly. Seizures, dysphagia, and pyramidal signs may appear with disease progression. Psychosis occurs in 25% to 40% of adolescent and adult-onset cases and may precede motor and cognitive disturbance by many years. Some cases have been described with onset in middle age associated with cognitive impairment alone. MRI often shows diffuse cerebral and/or cerebellar atrophy, atrophy of the striatum, hippocampus and thalamus, and thinning of the callosum. Positron emission tomography (PET) imaging often shows fronto-thalamic hypometabolism and has also demonstrated significant neuroinflammation and NFT accumulation in frontotemporal regions. Executive function is impaired early in adults, before memory impairment and speed of processing becomes impaired. The co-occurrence of vertical gaze palsy and psychosis should prompt the clinician to consider NPC, so that appropriate treatment can be instituted. Psychosis may require higher doses of antipsychotics and may require treatment with a combination of neuroleptics and mood stabilizers. The development of substrate reduction therapy using the imino sugar miglustat (Zavesca), which penetrates the blood–brain barrier and reduces ganglioside accumulation, has shown promise in reversing the motor and cognitive deficits in NPC and may also reduce or prevent psychiatric disturbance in this disorder, and current trials are underway involving intrathecal and intravenous cyclodextrin, a toroidal sugar that appears to restore the biochemical cholesterol trafficking deficit in animal and cellular models, and acetyl-L-leucine, a modified amino acid that appears to stabilize Purkinje cell function in NPC and other ataxias.

A 26-year-old man presented with movement disturbance and dysarthric speech following a 10-year history of a treatment-resistant psychotic disorder. He suffered persistent auditory hallucinations and referential and persecutory delusions, which only abated when treated with olanzapine (Zyprexa) 60 mg per day and valproic acid (Depakene) 2,000 g per day. On examination, he showed dysarthric speech, gait ataxia, and disturbed eye movements with jerky saccades and grossly impaired downgaze. He was cognitively rigid with significant memory and executive impairment. Tests for Tay–Sachs disease and other enzyme disorders were negative. Filipin staining of cultured fibroblasts showed an elevated number of cells, demonstrating perinuclear filipin staining of cholesterol (60% to 70%, normally less than 5%), and cholesterol esterification rate was mildly abnormal (2.9 pmol/hr/mg; less than 2 abnormal, 2 to 3 equivocal, greater than 3 normal). Mutation analysis of the NPC1 gene revealed a compound heterozygote status of G992R/R1186H, genetically confirming the diagnosis of Niemann–Pick disease type C.

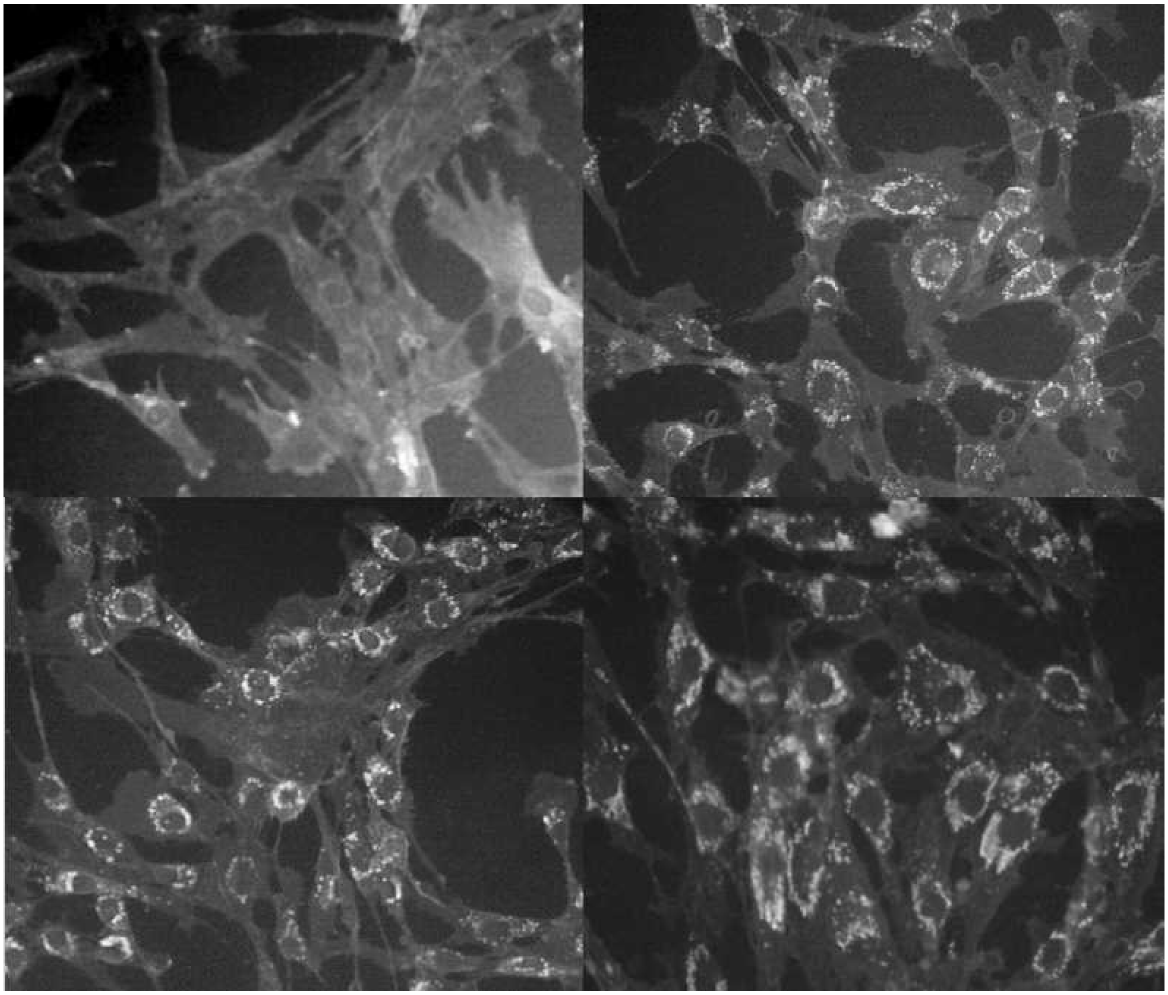

FIGURE 2.14–9. Filipin staining on fibroblasts for three patients with Niemann–Pick disease type C. *Top left* stain on healthy control reveals lack of filipin staining of accumulated cholesterol; on three other cases, each of whom presented with psychosis, punctate perinuclear accumulation of cholesterol is identified.

Pelizaeus-Merzbacher Disease. Pelizaeus-Merzbacher disease (PMD) is an X-linked recessive disorder with a frequency of 1:200,000 to 1:500,000 due to abnormalities in the gene encoding for the proteolipid protein (PLP), the major structural protein of CNS myelin, resulting in patchy myelin loss as a result of oligodendrocyte apoptosis and/or axonal damage. PMD is also variable in its onset and clinical manifestations, with childhood-onset PMD resulting in intellectual disability, nystagmus, and spastic paraparesis, and adult-onset cases most commonly associated with mild, adult-onset lower limb spastic paraparesis. Dementia and psychiatric dysfunction, including psychosis, are common in adult-onset cases, often associated with subtle or partial interruption to myelination

(Fig. 2.14–11), although adult-onset PMD is rare. Psychosis, when reported, occurs in the fourth or fifth decade. PLP has been implicated in schizophrenia: downregulation of the gene encoding PLP has been reported in schizophrenia and in animals treated with NMDA antagonists in experimental models of schizophrenia, implicating abnormal formation of myelin in psychosis. Treatment of PMD is symptomatic and supportive.

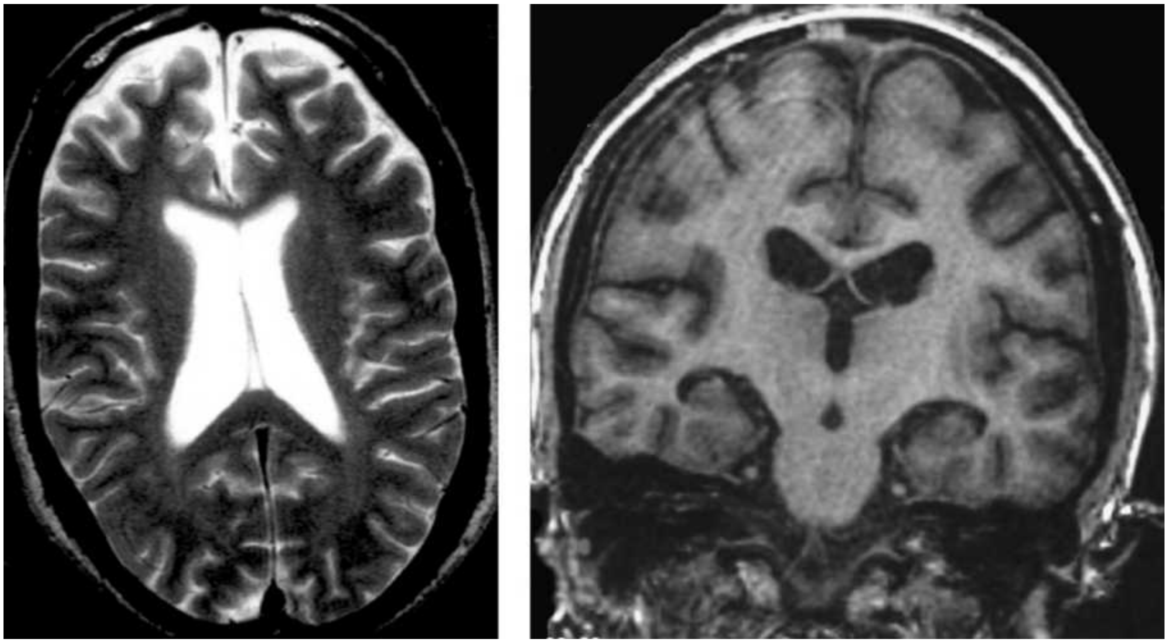

FIGURE 2.14–10. Magnetic resonance image scans on two individuals with Niemann–Pick disease type C who presented with major psychiatric disorders. *Left*, T_2-weighted axial image of a young adult male who presented with psychosis at age 16 prior to the onset of neurologic disturbance at age 25, showing frontal atrophy, ventricular enlargement, and a cavum septum pellucidum. *Right*, a coronal T_1-weighted image of a young male who presented with a rapidly cycling bipolar disorder in his early 20s. Note the disproportionate hippocampal atrophy in comparison to subtle global cerebral atrophy.

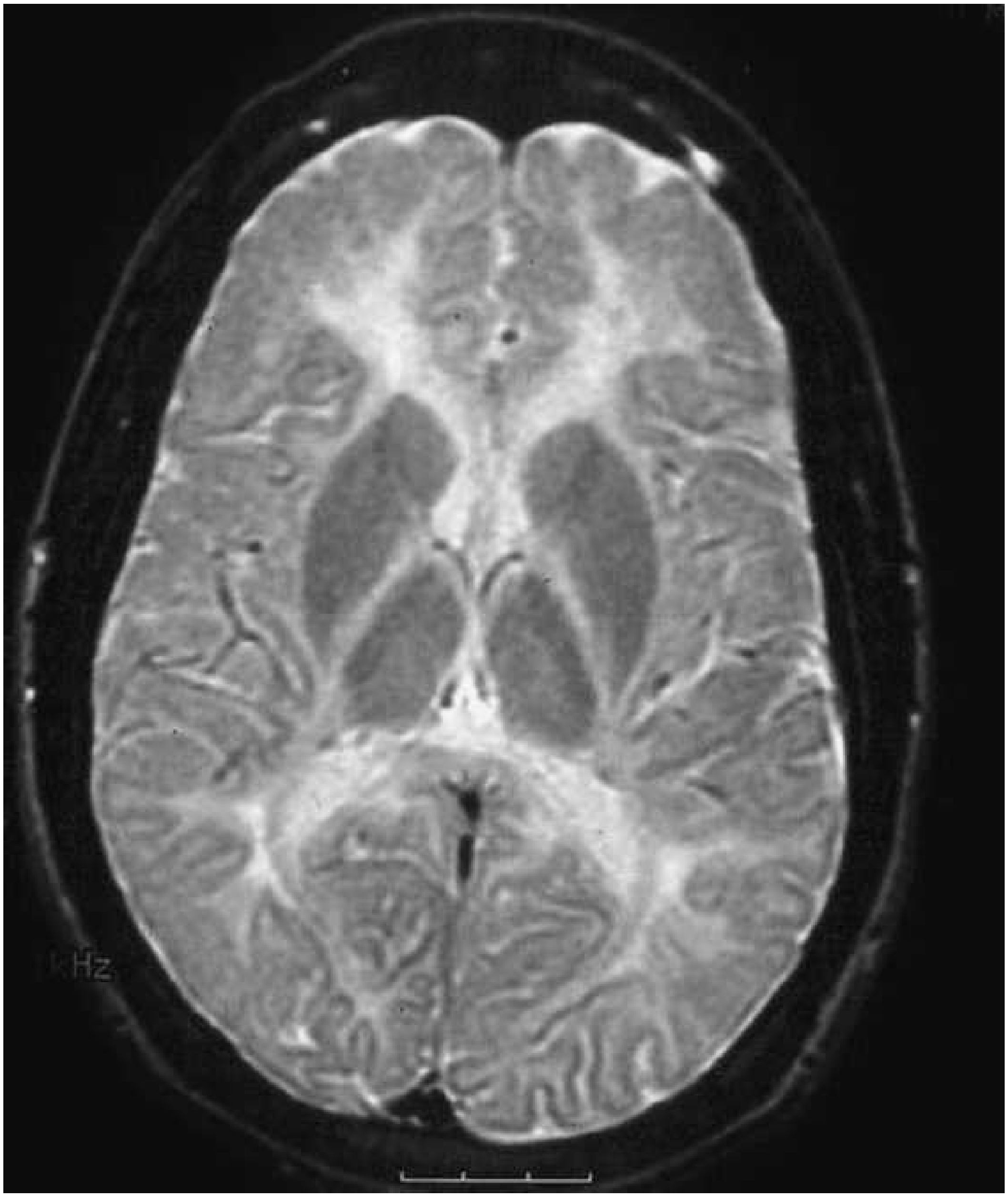

FIGURE 2.14–11. T_2-weighted magnetic resonance image scan in a 20-year-old individual with Perlizaeus–Merzbacher disease, demonstrating abnormally high signal in the internal capsule and posterior corpus callosum. The “mottled” frontal white matter represents “islands” of normally formed myelin. (From Koeppen AH, Robitaille Y. Pelizaeus-Merzbacher disease. *J Neuropathol Exp Neurol.* 2002;61(9):747, with permission of Oxford University Press. doi: 10.1093/jnen/61.9.747.)

Chorea-Acanthocytosis. Chorea-acanthocytosis (ChAc) is a form of neuroacanthocytosis, a group of disorders that present with neurologic and psychiatric manifestations secondary to changes in the basal ganglia and acanthocytes: spiculated red blood cells. ChAc is an autosomal recessive disorder associated with mutations or deletions in the *VPS13A* gene on chromosome 9q, which codes for the membrane protein chorein, a protein expressed in all tissues but particularly in the brain, skeletal muscle, and erythroid cell precursors. The loss of function of chorein, which interacts with the membrane cytoskeletal proteins adducing and actin, causes destabilization of the membrane skeleton in red blood cells and neurons; additionally, adducing and actin are also crucially involved in synaptic function and plasticity, and synaptic vesicle trafficking. Cell loss and astrocytic gliosis then occur in the basal ganglia in ChAc patients, most particularly in the caudate, but also the ventrolateral substantia nigra and globus pallidus (Fig. 2.14–11).

The onset of neurologic disturbance in ChAc is usually between the third and fifth decades, commonly with limb and orobuccal chorea that may be indistinguishable from Huntington chorea. It can be differentiated from the other main neuroacanthocytosis disorder, McLeod syndrome, by the absence of Kell antigens in red blood cells and the presence of cardiomyopathy in the latter. Patients with ChAc frequently present with mutilation of the tongue, lips, and cheeks, which is generally not a feature of Huntington chorea and can help to clinically distinguish the two disorders. Seizures, dystonia, and denervation atrophy occur in up to half of patients. Significant psychopathology is common in ChAc patients, occurring in up to two-thirds, and may precede the onset of frank neurologic disturbance by up to a decade. In the original series of 19 individuals described by Hardie et al. (of a mixed ChAc-dominant group of neuroacanthocytosis patients), the most prominent psychiatric feature was a behavioral and cognitive change (apathy, disinhibition, and poor judgment and planning) consistent with a dysexecutive syndrome in more than half the patients, with obsessive-compulsive disorder-like symptoms occurring in two patients. When psychiatric symptoms precede frank neurologic disturbance, an overrepresentation of OCD-type disorders occurs, in addition to behavioral syndromes secondary to frontal lobe disturbance. These both appear to be the result of the predilection of neuropathology in ChAc for the head of the caudate nucleus, which is a key relay in a basal ganglia–thalamocortical loop known as the lateral orbitofrontal loop. This circuit integrates information from the anterior cingulate, orbitofrontal, and dorsolateral prefrontal cortex to determine behavioral and motor programs that occur to resolve conflict or facilitate decision-making. Disruptions to this circuitry lead to apathy, disinhibition, and poor judgment and planning. The motor compulsions seen in many ChAc sufferers may be secondary to behavioral dysregulation of motor acts through a loss of motor inhibition. Less commonly, patients with ChAc have been known to present with a schizophrenia-like psychosis. When compulsive disorders occur in ChAc, they have been known to respond to both selective serotonin reuptake inhibitors (SSRIs) and TCAs. The disorder itself usually responds to antiseizure medication, and dopamine blockers/depleters for chorea, in addition to deep brain stimulation and botulinum toxin treatment.

A 38-year-old woman with a history of focal seizures with impaired awareness with secondary generalization since the age of 21 presented with a history of movement disorder. She first presented at age 16 with contamination fears and compulsive picking at her skin, which responded to antidepressant treatment. At the age of 27, she developed involuntary movements of the limbs, head, and neck, which were severe at the time of assessment. MRI demonstrated gross bilateral caudate atrophy, particularly affecting the head of the caudate, and a blood film showed 5% acanthocytes (Fig. 2.14–12). A Western blot on peripheral blood demonstrated the absence of normal chorein, confirming the diagnosis of neuroacanthocytosis. Her cognitive function deteriorated, and she remained on antidepressant medication while haloperidol (Haldol) was added, which significantly improved her worsening chorea. Her cognitive and behavioral decline continued over subsequent years, and she died of an unrelated medical illness 2 years later.

Cysteine and Homocysteine Disorders

Homocystinuria. Disorders that lead to increased blood and urinary levels of homocysteine are grouped under the term homocystinurias. The classic form of homocystinuria is an autosomal recessive disorder caused by a defect in the chromosome 21 gene coding for cystathionine beta-synthase, an enzyme that converts homocysteine and serine into cystathionine. This leads to the accumulation of homocysteine and methionine. Excessive homocysteine disrupts the structure of fibrillin-1, an extracellular matrix protein, and leads to damage of collagen and elastic fibers; this leads to significant arterial and venous diseases such as stroke, myocardial infarction or venous thrombosis. In neurons, it alters the release of monoamines and may potentiate neurotoxicity through NMDA receptor agonism; it can also induce DNA strand breakage and cellular apoptosis. The primary clinical effects of the disorder reflect this pathology. The presenting features are usually developmental delay or intellectual disability associated with optic lens dislocation within the first 10 years. Optic lens dislocation is seen in almost 100% of patients older than 10, and other ophthalmologic complications may ensue, such as cataracts, glaucoma, and optic nerve atrophy. Intellectual disability is seen in about half of the patients, and intellectual decline tends to be slowly progressive. Patients may appear Marfanoid with a long narrow head, arachnodactyly, kyphoscoliosis, and pectus excavatum. Patients will often have pale, pink skin together with fine, fragile, light-colored hair and malar rash. Abnormalities in the clotting cascade and increased platelet adhesiveness lead to occlusion of and thromboembolism from arterial and venous vessels, with very high mortality. Other neurologic complications include seizures and dystonias. Although early case reports identified patients with psychosis and homocystinuria, studies investigating large case series of patients have reported that personality disorder, behavioral disturbance (such as aggression), depression, and OCD are the most common psychiatric findings and that aggression and behavioral disturbances were more common in patients with intellectual disability. The diagnosis of classic homocystinuria is made on the basis of elevated plasma methionine and elevated plasma and urine homocysteine. Brain imaging may show the effects of cerebrovascular accidents, atrophy, or venous occlusions. Treatment with methionine restriction and cysteine supplementation will generally prevent any long-term sequelae if the disorder is diagnosed at birth. Oral pyridoxine (vitamin B_6), which remethylates homocysteine to methionine, leads to a reduction in levels of methionine and homocysteine in about half of patients, while the addition of folic acid and B_{12} may be of further benefit.

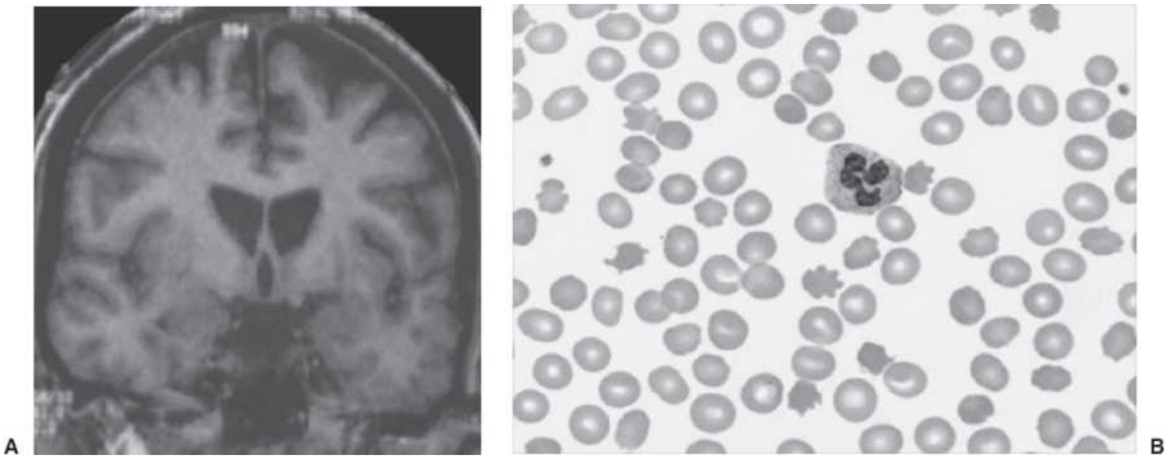

FIGURE 2.14–12. A 38-year-old woman with neuroacanthocytosis who first presented with adolescent obsessive-compulsive symptoms prior to adult-onset seizures and chorea. **A:** T_1-weighted coronal magnetic resonance image showing an almost total loss of the caudate head. **B:** Blood smear showing characteristic red cell acanthocytes. (See Color Plate.)

The metabolic pathway that converts homocysteine and serine to cystathionine is the most common pathway for the metabolism of homocysteine (Fig. 2.14–13). Other autosomal recessive homocystinurias are caused by defects in an alternative remethylation pathway in which homocysteine is catalyzed to methionine by either methionine synthetase or methylenetetrahydrofolate reductase (MTHFR). An inherited defect in methylcobalamin (methyl-B_{12}) that acts as a cofactor for methionine synthetase will also result in homocystinuria, together with methylmalonic aciduria. Deficiencies in these enzymes will lead to high levels of homocysteine but, in contrast to cystathionine synthase deficiencies, low levels of methionine. The clinical picture of these enzyme defects is one of intellectual disability, seizures, and hypotonia, while megaloblastic anemia is an additional feature of the disorder of methionine synthetase deficiency. The less severe form of MTHFR deficiency has been strongly associated with psychosis in adolescents and adults; this may be underpinned by alterations to dopamine synthesis and myelination. The C677T variant of the MTHFR gene, which causes a 35% reduction in enzyme function, has been shown to be an independent risk factor for the development of psychosis. The clinical course and MRI findings can resemble that seen in childhood leukodystrophies in these two disorders. Neuropsychiatric symptoms figure most prominently in the presenting symptoms of patients with methyl-B_{12} deficiency, with delays of up to 13 years between the onset of neuropsychiatric symptoms and the ultimate diagnosis.

Mitochondrial Diseases

Mitochondrial diseases are a heterogeneous group of diseases caused by defects in oxidative phosphorylation. They are characteristically multisystem disorders that overlap clinically and should be considered as a differential diagnosis across a range of neurologic and neuropsychiatric presentations. The genetic and phenotypic complexity of these disorders can be best understood in the context of a description of the mitochondrial genome, which is inherited maternally. Each human cell has up to several thousand copies of the mitochondrial genome, which is organized into a circular double-stranded structure. Human mitochondrial DNA was first sequenced in 1981 and consisted of 37 genes, with no introns and a small noncoding portion. Thirteen proteins, translated from mitochondrial DNA, are involved in the respiratory chain/oxidative phosphorylation system. The mitochondrial respiratory chain consists of five enzyme complexes (complexes I, II, II, IV, V) made up of polypeptides encoded by nuclear and mitochondrial genes, except for complex II, which is entirely encoded in the cell nucleus. As a result, the mitochondrion is under the genetic control of both nuclear and mitochondrial DNA, and mitochondrial disorders can result from mitochondrial or nuclear DNA mutations. These enzyme complexes participate in a chain of metabolic processes that lead to ATP production, the overall process being referred to as oxidative phosphorylation. ATP is used in the vast majority of cellular metabolic processes as an energy source, and the respiratory chain responds to the energy needs of cells, which in some cases may be quite stable, while in others (e.g., muscle), they may vary dramatically over time. Other functions of the mitochondria include cellular homeostasis, fatty acid oxidation, the urea cycle, intracellular signaling, calcium storage and signaling, apoptosis, and the metabolism of amino acids, lipids, cholesterol, steroids, and nucleotides.

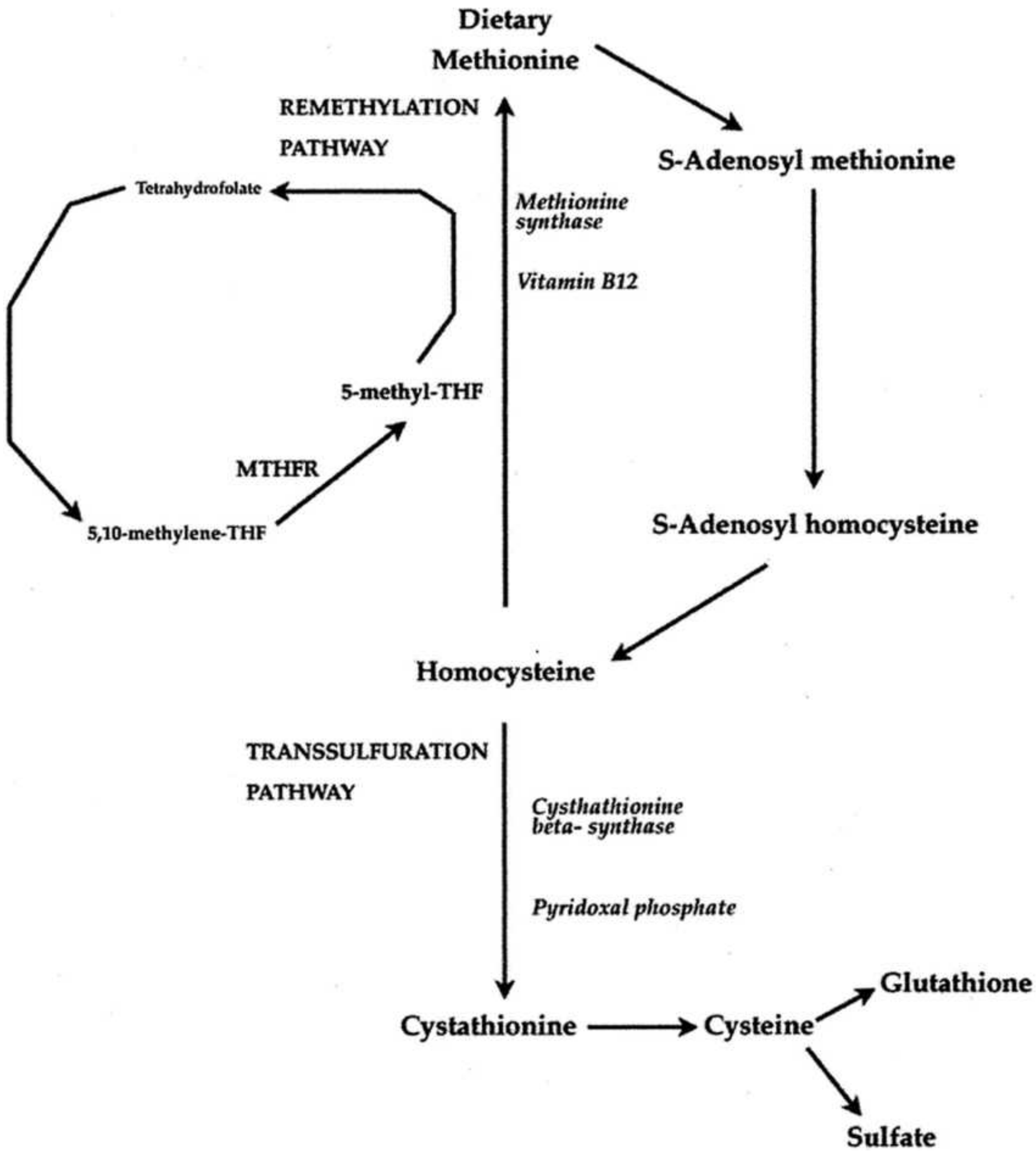
Dietary Methionine
S-Adenosyl methionine
S-Adenosyl homocysteine
Homocysteine
REMETHYLATION PATHWAY
Methionine synthase
Vitamin B12
Tetrahydrofolate
5-methyl-THF
MTHFR
5,10-methylene-THF
TRANSSULFURATION PATHWAY
Cysthathionine beta- synthase
Pyridoxal phosphate
Cystathionine
Cysteine
Glutathione
Sulfate

FIGURE 2.14–13. Metabolism of homocysteine in vivo. THF, tetrahydrofolate. (From Kelly PJ, Furie KL, Kistler JP, et al. Stroke in young patients with hyperhomocysteinemia due to cystathionine beta-synthase deficiency. *Neurology*. 2003;60(2):275, with permission. doi: 10.1212/01.WNL.0000042479.55406.B3. © American Academy of Neurology.)

The genetics of mitochondrial disorders are complex, but the following principles can be generally applied:

1. Mitochondrial disorders may be sporadic, maternally inherited, or inherited in an autosomal recessive or dominant pattern.
2. Due to the polyploid nature of the mitochondrial genome, the one cell may include normal and mutated mitochondrial DNA (heteroplasmy), and thus siblings may show a very broad range of clinical variability due to differences in the inheritance of such heteroplasmic mitochondria. The mitochondrial function will generally begin to fail once the proportion of mutated mitochondrial DNA has crossed a threshold of 70% to 80%.
3. Mitochondrial respiratory chain disorders will mostly affect tissues with high metabolic needs such as muscle, central and peripheral nervous system, heart, endocrine and eye.
4. The clinical expression of mitochondrial disorders may vary widely from individual to individual with the same mutation depending on the proportion of mitochondria affected in different tissues, the interaction of that individual with the environment, and the differential metabolic energy needs of different tissues within the one individual.

Mitochondrial diseases are the most common inborn errors of metabolism, accounting for 1 in 5,000 live births with nearly 300 causative genes identified. One-third of these genes have been identified through next-generation sequencing in the last decade. About 15% of these genes are mitochondrial genes, with the remainder being genes encoded in the nucleus. For the reasons outlined above, there is no clear genotype–phenotype relationship for mitochondrial genome disorders. Clinical features common to all mitochondrial disorders include dysfunction of endocrine (short stature, diabetes, thyroid and adrenal disorders), neurologic (deafness, myopathy, peripheral neuropathy, retinopathy, optic atrophy, ophthalmoplegia, seizures, ataxia, dementia), and cardiac (cardiomyopathy, cardiac block) systems. The patterns of clinical presentation vary significantly with regard to the age of onset, the temporal order of symptoms and conditions, and the progress of the disorders. The combination of maternal history, multisystem involvement, and a progressive course should arouse clinical suspicion of a mitochondrial disorder. In patients with atypical psychiatric presentations, physical signs such as muscle weakness, hearing loss, seizures, short stature, diabetes, Wolff–Parkinson–White syndrome, or migraines should alert clinicians to the possibility of a mitochondrial disorder. The disorders reviewed below are not an exhaustive review of mitochondrial diseases and are restricted to those disorders with clinical neuropsychiatric phenotypes.

Mitochondrial Encephalomyopathy, Lactic Acidosis, and Stroke-Like Episodes (MELAS). MELAS, the most common of the mitochondrial disorders, presents before early adulthood after a period of normal development. The majority (about 80%) of MELAS cases are due to an A to G substitution at nucleotide 3243 of tRNA leucine (A3243G $\mathrm{tRNA}^{\mathrm{Leu(UUR)}}$). A further 10% of mutations are in other regions of this same gene, while the remaining 10% of mutations occur in six other mitochondrial genes. The onset is most common in childhood but can first present in early adulthood. The characteristic features of MELAS are stroke-like episodes whose lesions do not conform to vascular territories and may involve gray or subcortical white matter. They typically occur in tempero–parieto–occipital regions, basal ganglia, brainstem, and cerebellum and lead to hemiparesis, hemianopia, and cortical blindness. Other clinical features include short stature, migraines, vomiting, and seizures. Lactic acid levels are elevated and have been correlated with the level of neurologic symptoms. The course of MELAS is highly variable, ranging from single stroke-like episodes to a progressive course characterized by one or more multiple strokes, deafness, diabetes, retinopathy, seizures, and cardiac abnormalities. Cases of schizophrenia or schizophrenia-like psychosis associated with MELAS have been reported. Such cases typically show onset in the third decade, several years to a decade before the clinical diagnosis of MELAS is made. Similar cases of depression, obsessive-compulsive disorder, bipolar disorder, and borderline personality disorder have been reported, but like the schizophrenia cases, the course of the illness is not typical. A retrospective review of the patient’s history will often reveal previously unappreciated features of MELAS, such as short stature, diabetes, or unexplained somatic symptoms. The development of neurologic signs or symptoms, cognitive decline, and evidence of strokes on imaging usually leads to the definitive diagnosis.

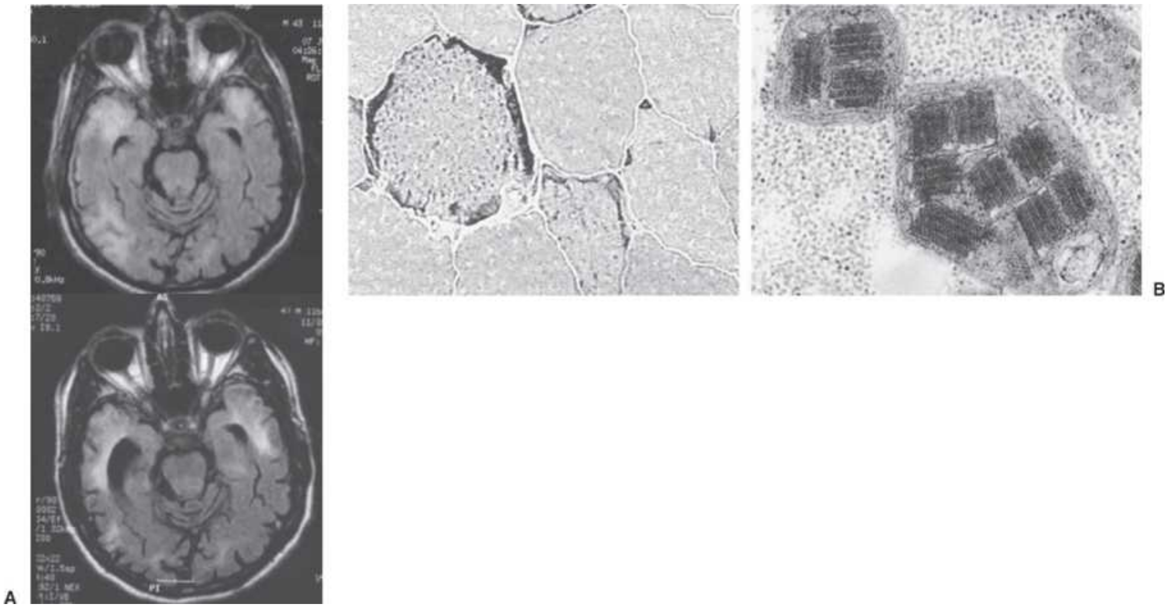

FIGURE 2.14–14. Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS). *Left*, axial fluid-attenuated inversion recovery magnetic resonance image scans of a patient at age 43 and age 47, showing early gliosis of anterior temporal and parietal zones (*top*) with progression over 2 years to more advanced gliosis and cortical volume loss (*bottom*). *Right*, skeletal muscle from a 29-year-old man with early-onset stroke. **A:** Gomori trichrome stain demonstrates the typical appearance of ragged red fibers, consisting of abnormal subsarcolemmal proliferation of mitochondria. **B:** Electron micrograph of skeletal muscle demonstrates "parking lot" inclusions, consisting of dystrophic mitochondria.

A 42-year-old man with deafness and diabetes was admitted with a confusional episode and found to have extensive posterior cortical and subcortical changes on MRI. The diagnosis of MELAS was confirmed by genetic testing. Five years later, he was rereferred due to altered behavior, aggression, and having become very "fixed" in his ideas. He performed poorly on bedside executive function tests. Repeat MRI revealed extensive inferior frontal and parietotemporal atrophy (Fig. 2.14–14).

The diagnosis of MELAS is based on the clinical syndrome, elevated serum lactic acid levels, muscle biopsy showing ragged red fibers and molecular genetic testing. Ragged red fibers are muscle fibers exhibiting mitochondrial proliferation in response to mitochondrial failure. MRI scanning typically shows cerebral stroke-like lesions. There is currently no available treatment for MELAS. A study of 13 patients treated with oral or intravenous L-arginine during acute stroke-like episodes identified a reduction in their frequency and severity but did not alter the disease trajectory over the 9-year follow-up period. Antioxidants, respiratory chain substrates, and cofactors have been studied in trials with varying results. Symptomatic treatments for MELAS include cochlear ear implants for deafness, anticonvulsants, antipsychotics and diabetic treatments. Antipsychotics with a higher risk of glucose intolerance (e.g., olanzapine) should be avoided or used with caution.

Myoclonic Epilepsy With Ragged Red Fibers (MERRF). MERRF typically begins in middle adulthood with photosensitive myoclonic seizures and is associated with limb-girdle weakness, dementia, cerebellar ataxia, cardiac abnormalities, and neuropathy. Muscle biopsy reveals ragged red fibers, and the electroencephalogram (EEG) is usually abnormal. In 90% of cases, the causative mutation is an A8344G transition in the tRNA *lysine*. There is no specific treatment, but appropriate management of the epilepsy is critical to clinical outcome, with levetiracetam and clonazepam recommended as first-line treatments.

Mitochondrial Diseases with Progressive External Ophthalmoplegia. Progressive external ophthalmoplegia, a clinical syndrome of progressive ptosis with paresis of eye movement, is associated with a number of mitochondrial diseases. Mutations in the nuclear PLOG1 gene are the commonest causes of the syndrome with a phenotype of adult presentation, dementia, depression, parkinsonism, ataxia, and occipital seizures. Kearns-Sayre syndrome (KSS) is a sporadic mitochondrial deletion disorder characterized by a triad of onset before 20, progressive external ophthalmoplegia plus one of the following: pigmentary retinopathy, heart block, cerebellar ataxia, or elevated cerebrospinal fluid (CSF) protein. It is associated with short stature, cognitive impairment, and death due to cardiac or neurologic morbidity before the age of 40. Muscle biopsy reveals ragged red fibers, and serum lactate is often elevated. MRI will often show diffuse central white matter abnormalities and basal ganglia calcification.

Mitochondrial Diseases With Cerebral White Matter Disease. Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE) is an autosomal recessive disease caused by mutations in the thymidine phosphorylase gene (TYMP), leading to the accumulation of thymidine. The clinical presentation is in adolescence with gastrointestinal dysmotility symptoms, progressive external ophthalmoplegia, peripheral neuropathy, and leukoencephalopathy progressing to death in the fourth decade. Thymidine phosphorylase replacement with platelet or erythrocyte transfusion has been of limited value. Increasing thymidine clearance through hematopoietic stem cell transplantation and orthotopic liver transplantation has been associated with significant medical complications, particularly due to the need for life-long immunosuppression. Leber hereditary optic neuropathy (LHON) is a maternally inherited disorder presenting as bilateral visual neuropathy in young adults. Although usually confined to the optic nerve, it has been described together with white matter pathology. Maternally inherited Leigh syndrome (MILS), also known as subacute necrotizing encephalomyelopathy, is an infantile encephalomyopathy characterized by seizures, heart block, dystonia, and optic atrophy.

Many overlapping mitochondrial syndromes have been described that include features typical of MELAS plus one of the other syndromes (e.g., LHON/MELAS, MELAS/MERRF, Leigh/MELAS).

Wolfram Disease Type I and Type II. Wolfram disease Type I is an autosomal recessive disorder caused by a mutation on the short arm of chromosome 4 (4p16.1) in the *WFDS1* (wolframin) gene. The wolframin protein acts to regulate cellular calcium. Wolfram disease is also known by the acronym DIDMOAD (*d*iabetes *i*nsipidus, *d*iabetes *m*ellitus, *o*ptic *a*trophy, *d*eafness) and often displays characteristic atrophy of the optic tract and loss of signal of the neurohypophysis on MRI (Fig. 2.14–15). Neurologic symptoms include dysarthria, areflexia, epilepsy, and anosmia. In a 1991 study, 68 patients with Wolfram disease were reviewed. Sixty percent (41 of 68) had had psychiatric symptoms, and 25% (17 of 68) were classified as showing severe psychiatric illness. Eleven (16%) patients had a history of psychotic symptoms. Heterozygous family members exhibited a high rate of psychiatric illness, approximately eight times greater than for noncarriers of the wolframin gene. Subsequent studies and systematic reviews of the clinical phenotype have not replicated these high rates of psychiatric morbidity in patients with Wolfram syndrome.

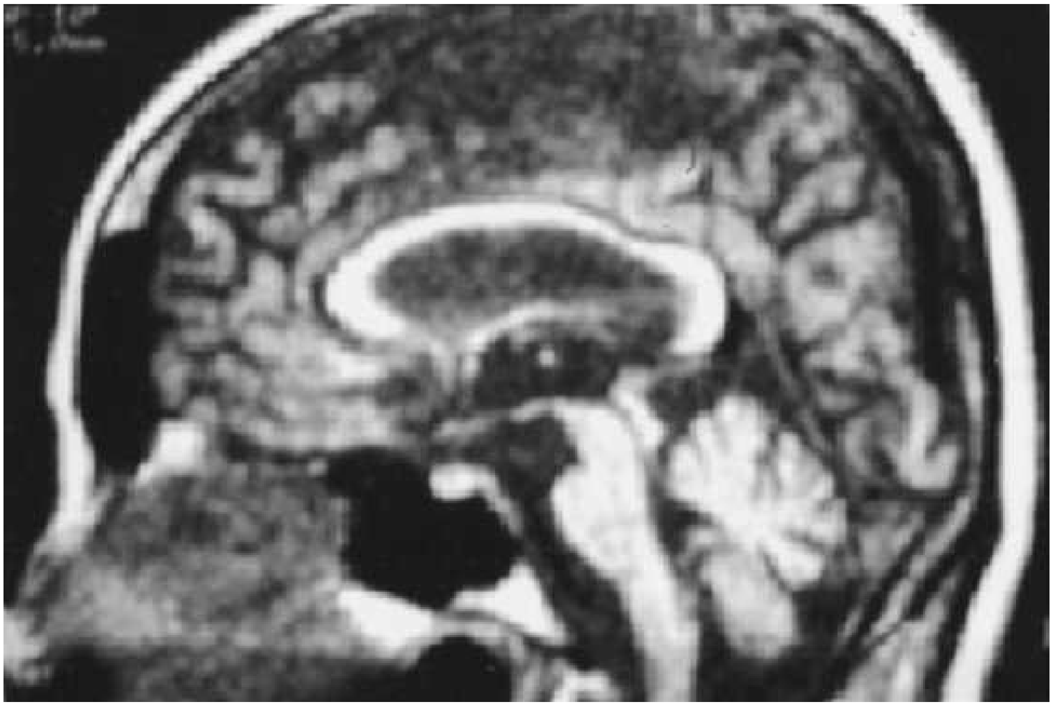

FIGURE 2.14–15. Wolfram syndrome. Sagittal T_1-weighted magnetic resonance image scan on a 32-year-old female with deafness and depression. Evident are frontal atrophy, thinning of the optic chiasm and tracts, atrophy of the brainstem and vermis and absence of physiologic high signal of the neurohypophysis. (From Pakdemerli E, Karabulut N, Bir LS, Sermez Y. Cranial magnetic resonance imaging of Wolfram (DIDMOAD) syndrome. *Australas Radiol*. 2005;49(2):189, with permission. © The Royal Australian and New Zealand College of Radiologists. doi: 10.1111/j.1440-1673.2005.01420.x.)

Disorders of Metal Metabolism

Wilson Disease. Wilson disease is an autosomal recessive disorder caused by mutations in a copper transporting ATPase encoded by the *ATP7B* gene on chromosome 13q14.3. About 1% of the population carries an *ATP7B* mutation, of which almost 1,000 have been identified. The frequency of the disorder is estimated to be about 1 in 40,000. The gene defect leads to the accumulation of copper in the liver through impaired copper excretion and impaired binding of copper to ceruloplasmin. The subsequent catabolism of ceruloplasmin leads to low ceruloplasmin plasma levels and increased free copper. Free copper then accumulates in the brain and leads to the neurologic and neuropsychiatric manifestations of this disorder. Copper deposition occurs in astrocytes but not neurons or the extracellular matrix and is particularly evident in the basal ganglia.

The most common presentation of Wilson disease is hepatic disease anywhere between the first and fourth decades. About 50% of patients are symptomatic by age 15. Up to two-thirds of patients will present with neurologic disease in the second or third decade without clinical evidence of liver disease. The *dystonic* form is the commonest presenting neurologic syndrome with dysarthria, dysphagia, drooling, and a rigid open mouth. In the *pseudosclerotic* form, patients present with incoordination, clumsiness, unsteadiness of gait, or dysarthria. As the illness progresses, patients may exhibit features of both forms with rigidity, tremor, choreic, athetoid, or dystonic movements. Such movements may be exacerbated by stress in the early stages of illness and be interpreted as functional. In particular, the tremor associated with Wilson disease, the so-called wing beating, may be interpreted as a functional movement disorder. This tremor is characteristically absent at rest and develops after a short period of the arm extension. The arms beat in a wide violent arc, and the tremor may be altered by the position of the arms.

Wilson's original description in 1912 emphasized the importance of psychiatric changes in the disease. Since then, many authors have described the high prevalence of psychiatric symptoms, with estimates ranging from 30% to 100% of symptomatic patients experiencing psychiatric symptoms at some point during their illnesses. Up to two-thirds of patients with Wilson disease will present with psychiatric symptoms, and one-third will have received psychiatric treatment prior to the diagnosis being made. The most common reasons for psychiatric referral are behavioral and personality changes, with disinhibition, bizarre, or impulsive behavior being present in about a quarter of patients and depression in about one-fifth. Personality and behavioral changes, but not depression, correlate with the degree of neurologic impairment. Although early reports suggested that psychotic presentation was a feature of Wilson disease, more recent investigations have shown that psychotic symptoms are relatively rare. The importance of recognizing the early psychiatric presentations of Wilson disease lies in the proven benefits of early treatment intervention. Diagnostic delays may be exacerbated if extrapyramidal movements are attributed to psychotropic medications or interpreted as functional. Patients with neurologic involvement show a subcortical pattern of cognitive impairment with frontal executive deficits that correlate with the extent of cerebral involvement.

The presence of Kaiser Fleischer (KF) rings (copper deposits in the outer rim of the cornea that are brown or gray-green) on slit-lamp examination is the single most important clinical diagnostic sign and is observed in about half of patients with hepatic presentation and almost all patients with a neurologic or psychiatric presentation. Low serum levels of ceruloplasmin have high sensitivity (over 80%) and low specificity, and normal ceruloplasmin may occur in 10% of patients in the presence of high urinary copper excretion. Low serum copper or free copper and high urine copper can be seen but are less reliable. Liver biopsy with staining for total hepatic copper has a sensitivity of over 80%, but liver biopsy is subject to false-negative results due to sampling errors. Genetic testing of the ATP7B gene is the gold standard for diagnosis but requires direct sequencing of the entire gene and may detect variants of unclear significance. MRI shows reduced T_1 signal and increased T_2 signal in basal ganglia, thalamus, and brainstem (Fig. 2.14–16). The European Association for the Study of the Liver developed the Leipzig criteria, incorporating clinical and laboratory findings, in recognition of the variability of clinical and laboratory findings in Wilson disease. The criteria are valued in clinical and research settings. The aims of treatment in Wilson disease are to remove accumulated copper, prevent re-accumulation through maintenance treatment and reduce its absorption. Treatment with d-penicillamine combined with pyridoxine has been the mainstay of initial treatment but may lead to a worsening of neurologic symptoms in up to 50% of patients. This has led to the recommended use of less neurotoxic copper chelators such as trientine. Zinc acetate therapy which induces metallothionine in the gastrointestinal epithelium and inhibits copper absorption, has been used as a preventative treatment in patients who are presymptomatic and for maintenance treatment. Restriction of high copper foods (shellfish, liver, chocolate, nuts, beans, mushrooms) is encouraged. A liver transplant can restore normal copper excretion but is limited due to the need for lifelong immunosuppression and the complexity of the procedure. Neurologic and psychiatric symptoms will improve with copper chelation over 1 to 2 years, with regression of MRI lesions on serial imaging. In patients with neurologic forms of Wilson disease, the best outcomes are seen in patients with the pseudosclerotic form (82% symptom-free), with worse outcomes in patients with chorea (75% of patients), parkinsonism (63%), and dystonia (53%).

Neurodegeneration With Brain Iron (NBIA). Neurodegeneration with brain iron (NBIA) refers to a group of disorders that were previously referred to as Hallevorden–Spatz syndrome. Common to these disorders is the accumulation and deposition of iron in the brain in association with clinical, radiologic, and pathologic evidence of neurodegeneration. The clinical phenotype of these disorders include neurologic (parkinsonism, dystonia, choreoathetosis), cognitive impairment and neuropsychiatric symptoms. Five of the ten recognized NBIA disorders which have adolescent or adult-onset phenotypes are discussed below (PKAN, PLAN, MPAN, aceruloplasminemia, neuroferritinopathy).

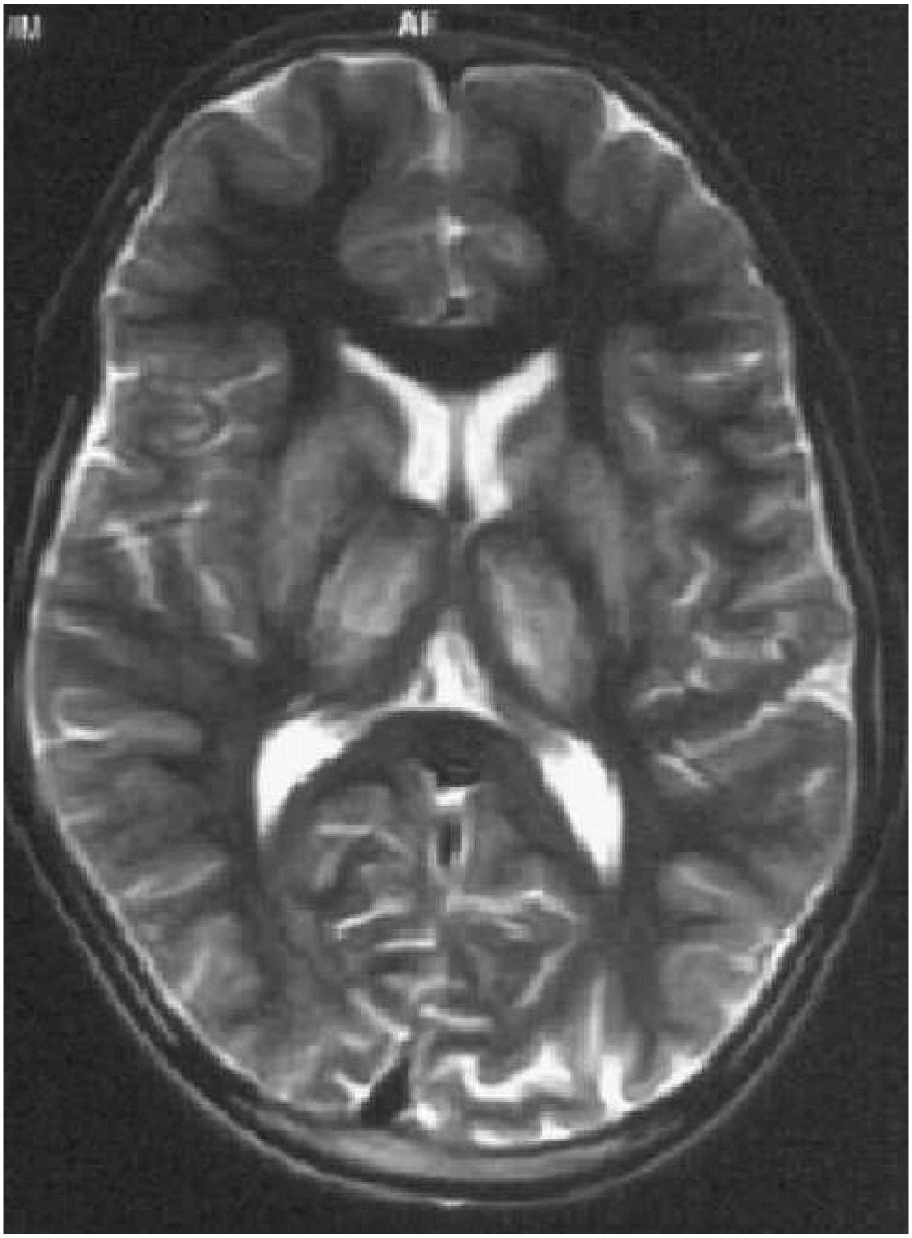

FIGURE 2.14–16. T_2-weighted axial magnetic resonance image of an adolescent patient with Wilson disease, showing characteristic bilateral hyperintensity of the thalamus, caudate, and putamen. (From Das M, Mirsa UK, Kalita J. A study of clinical, MRI and multimodality evoked potentials in neurologic Wilson disease. *Eur J Neurol.* 2007;14(5):498, with permission. © EAN. doi: 10.1111/j.1468-1331.2006.01676.x.)

PANTOTHENATE KINASE-ASSOCIATED NEURODEGENERATION (PKAN). The identification of mutations in the gene *PANK2* on chromosome 20p13 led to the descriptive term “pantothenate kinase-associated neurodegeneration” for what is recognized as the most prevalent form of

NBIA. Pantothenate kinase 2 regulates the mitochondrial synthesis of coenzyme A (CoA), which is involved in energy and fatty acid metabolism. *PANK2* catalyzes the phosphorylation of pantothenate (vitamin B_5) to phosphopantothenate, which condenses with cysteine in the next step of CoA biosynthesis. Mutations in *PANK2* lead to an accumulation of cysteine, which binds iron and leads to free radical production, triggering cell membrane damage and death. The retina and basal ganglia appear particularly sensitive to these effects of PKAN mutations, and the primary clinical features reflect this retinal and basal ganglia involvement.

The clinical phenotype of *PKAN* can be divided into three types: The *classic* syndrome first described by Hallevorden and Spatz consisted of early childhood (3 to 4 years) onset of dystonia, dysarthria, rigidity, pyramidal signs, pigmentary retinopathy, and cognitive decline, and with a progressive fulminant course such that patients became nonambulatory by their mid to late teens. All patients with this classic phenotype have been found to have a *PANK2* mutation. The *atypical clinical phenotype* (about 25% of PKAN) is characterized by presentation in the early to mid-teenage years with palilalia, tachylalia, dysarthria, and psychiatric symptoms, including depression, emotional lability, personality changes, and cognitive decline. Extrapyramidal rigidity, dystonia, and pyramidal spasticity develop subsequently and result in progressive loss of mobility over 15 to 40 years. About one-third of this *atypical clinical phenotype* exhibit *PANK2* mutations. Interestingly, the patients with *PANK2* mutations and late-onset appear to be more likely to exhibit speech and psychiatric symptoms at onset. The third clinical phenotype, previously termed *HARP* (hypobetalipoproteinemia, acanthocytosis, retinopathy, and pallidal degeneration), has now been identified as being caused by a *PANK2* mutation. Together with a strongly suggestive clinical picture, the finding of the characteristic MRI “eye of the tiger” sign (a low signal intensity region caused by iron deposition and a high signal area that corresponds to the axonal spheroid formation, as seen in Fig. 2.14–17) is almost pathognomonic of a *PANK2* mutation and should lead to genetic testing. Treatment aimed at reducing iron levels using deferiprone, an iron chelator that crosses the blood–brain barrier, has been relatively ineffective in the classic phenotype but may slow progression in the atypical phenotype. A phosphopanthonate replacement trial using fosmetpantotenate did not improve function in PKAN patients (Table 2.14–1).

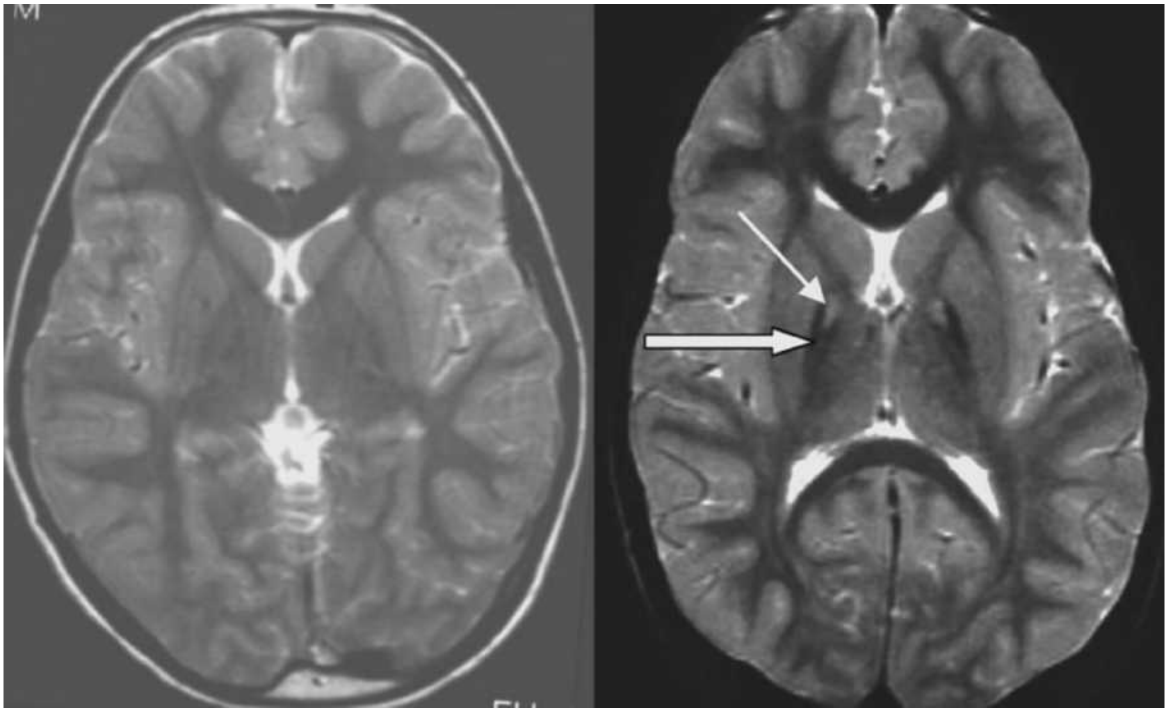

FIGURE 2.14–17. Pantothenate kinase-associated neurodegeneration. *Left*, T_2-weighted axial magnetic resonance image of a normal control shows isodense globus pallidus. *Right*, *PANK2* mutation-positive patient with neurodegeneration with brain iron accumulation (NBIA) shows hypointensity (*thick arrow*) with a central region of hyperintensity (*thin arrow*) in the medial globus pallidus, known as "the eye of the tiger" sign. (From Hayflick SJ: Unraveling the Hallervorden-Spatz syndrome: Pantothenate kinase-associated neurodegeneration is the name. *Curr Opin Pediatr*. 2003;15[6]:572, with permission.)

PLA2G6-ASSOCIATED NEURODEGENERATION (PLAN). About 200 autosomal recessive mutations have been identified in the phospholipase A2 gene (PLA2G6, chromosome 22q13.11) and can lead to infantile, childhood-onset and adult-onset disorders. Pathologically this disorder is characterized by the accumulation of spheroids in peripheral axons and nerve terminals. The adult-onset disorder is typically one of dystonia-parkinsonism with associated abnormalities in eye movements, pyramidal signs, cognitive impairment, and psychiatric symptoms.

MITOCHONDRIAL MEMBRANE PROTEIN-ASSOCIATED NEURODEGENERATION (MPAN). MPAN is associated with approximately 50 mutations (autosomal recessive or dominant) in the (chromosome 19 open reading frame 12) gene. The function of the C19ORF12c protein, a mitochondrial membrane protein, remains unknown, as does the mechanism by which mutations lead to iron accumulation. The clinical phenotype is one of progressive dementia, pyramidal and extrapyramidal symptoms, optic atrophy, peripheral motor neuropathy, and neuropsychiatric symptoms.

ACERULOPLASMINEMIA. Aceruloplasminemia is an autosomal recessive disorder associated with reduced or absent levels of ceruloplasmin and tissue iron deposition. It can be caused by about 40 mutations in the ceruloplasmin gene on chromosome 3q25 and occurs in 1 per 2 million births. There are two types of ceruloplasmin. Serum ceruloplasmin is predominantly synthesized in the liver but does not cross the blood–brain barrier, while a second form (GPI-anchored form) is produced by astrocytes in the CNS. Ceruloplasmin promotes the loading of iron onto transferrin, allowing Fe^{3+} efflux out of cells and preventing oxidative damage caused by Fe^{2+}. GPI-anchored ceruloplasmin plays a key role in regulating iron levels in the CNS and in preventing free radical injury. The disease leads to the deposition of iron in the CNS, retina, pancreatic cells, liver, spleen, and ovaries. The major sites of CNS iron deposition in aceruloplasminemia are similar to the sites of greatest iron concentration in healthy individuals: the basal ganglia, cerebellar dentate nuclei, red nucleus, thalamus, and hippocampus.

Table 2.14–1. Metabolic Disorders Presenting as Neuropsychiatric Syndromes, With Pattern of Inheritance, Gene Location, and Product, Prevalence, Onset, and Neuropsychiatric Features

Disorder	Inheritance	Location	Gene	Gene Product	Stored or Altered Substance	Prevalence per Million	Diagnosis	Onset	Neuropsychiatric Features
Metachromatic leukodystrophy	AR	22q13.31	ARSA	Arylsulfatase A	Cerebroside sulfate	5–25	Enzyme levels (leukocytes), genetic testing	Childhood–adulthood	Psychosis in adolescent onset/chorea/seizures/dementia
Fabry disease	X-linked recessive	Xq22	α-Gal	α galactosidase A	Globotriaosylceramide	25 (males)	Enzyme levels (leukocytes), genetic testing	Childhood–adolescence	Depression
Tay–Sachs disease	AR	15q23–q24	HEXA	β–hexosaminidase A	GM2-gangliosides	150–400 (Jewish) 3 (general population)	Enzyme levels (leukocytes), genetic testing	Childhood–adulthood	Psychosis in adult-onset, speech, gait disorder, tremor
Neuronal ceroid lipofuscinosis (Kufs)	AD/AR	Not yet identified	CLN4	Unknown—more than one may be involved (CLN1&2—lysosomal thioesterase and protease; CLN3&8 are membrane proteins of unknown function)	Unknown	1–2	Biopsy—muscle, vessel	Adult (Kufs)	Psychosis/depression/myoclonic epilepsy/dyskinesias
α-mannosidosis	AR	19q13.1	MANB	α-Mannosidase	Oligosaccharides	2	Enzyme levels (leukocytes), genetic testing	Childhood–adulthood	Psychosis in adult onset/ataxia/deafness/retinopathy
X-linked adrenoleukodystrophy	X-linked recessive	Xq28	ABCD1	ALDP peroxisomal membrane protein	Saturated very-long-chain fatty acids (VLCFAs)	50 (males)	Serum VLCFA levels, genetic testing	Childhood–adulthood	Affective psychosis in adult-onset/demyelination
Acute intermittent porphyria	AD	11q23.3	HMBS	Porphobilinogen-deaminase	Porphobilinogen/amino levulinic acid	40–100	Urine porphobilinogens	Adulthood	Psychosis/delirium/depression/peripheral neuropathy
Phenylketonuria	AR	12q23.2	PAH	Phenylalanine hydroxylase	Phenylketones	50–100	Phenylalanine levels	Childhood	Intellectual disability
Maple syrup urine disease	AR	19q13.1–13.2 6p22–p21 1p31	BCKDHA BCKDHB DBT	Branched-chain ketoacid dehydrogenase	Branched-chain amino acids (BCAAs)	5	Urine keto acids, serum BCAAs, genetic testing	Childhood–adulthood	Depression and anxiety in adults
NPC	AR	18q11–12 14q24.3	NPC1 NPC2	NPC1 NPC2	Cholesterol and gangliosides	10	Plasma oxysterols or lysosphingomyelin, genetic testing	Childhood–adulthood	Intellectual disability in childhood-onset/psychosis in adolescent-onset/dementia in adult-onset
Pelizaeus-Merzbacher Disease	X-linked recessive	Xq22	PLP1	Proteolipid protein	n/a	2–5	Neuroimaging, genetic testing	Childhood–adulthood	Intellectual disability in childhood-onset/psychosis and dementia in adult-onset
Cerebrotendinous xanthomatosis	AR	2q33	CYP27A1	Sterol 27 hydroxylase	Cholestanol and cholesterol	300 reported cases	Plasma cholestanol, genetic testing	Adulthood	Psychosis/agitation/depression
Cystinosis	AR	17p13	CTNS	Cystonin	Cystine	5	Cysteine levels (leukocytes), genetic testing	Childhood	Cystinosis encephalopathy
Homocystinuria	AR	21q22.3	CBS	Cystathionine synthase	Homocysteine and methionine	3–5	Plasma homocysteine, liver biopsy, genetic testing	Childhood	Intellectual disability/seizures/dystonias/personality change/behavioral disturbance/depression/OCD
Tetrahydrobiopterin deficiency	AR	Multiple	GCH1 PCBD1 PTS QDRP	Multiple	BH4	2	Urinary pterin levels, BH4 and PhE loading tests, genetic testing	Childhood–Adulthood	Behavioral disturbance, intellectual disability; adult mood and psychotic disorders

Hyperprolinemia	AR	1p36	ALDH4A1	Pyrroline-5-carboxylate dehydrogenase	Proline	<1	Plasma proline level, genetic testing	Childhood–adulthood	Intellectual disability, psychosis
Propionic academia	AR	13q32, 3q21-22	PCCA PCCB	Propionyl-CoA carboxylase	Propionic acid	10	Plasma propionyl-carnitine and methionine, genetic testing	Childhood–adulthood	Encephalopathy, intellectual disability, autism, psychosis
AMACR deficiency	AR	5p13	AMACR	α-Methylacyl-CoA racemase	Pristanic acid	<1	Plasma pristanic acid, genetic testing	Early adulthood	Cognitive decline, psychosis
Cerebral creating deficiency syndrome (CCDS)	AR and XL	19p13.3, 15q21.1, Xq28	GAMT AGAT SLC6A8	Guanidinoacetate *N*-methyl transferase, glycine amidinotransferase, sodium and chloride-dependent creatine transporter	Creatine	<1	Plasma/urine creatine, guanidinoacetate, creatinine; genetic testing	Childhood	Autism, self-mutilation, intellectual disability
Non-ketotic hyperglycinemia (NKH)	AR	9p22, 3p21.31	GLDC AMT	Glycine dehydrogenase, aminomethyltransferase	Glycine	20	Glycine levels in the blood, plasma, CSF; genetic testing	Childhood	Intellectual disability, attention deficit
SSADH deficiency	AR	6p22	ALDH5A1	Succinic semialdehyde dehydrogenase	GABA, GHB	<1	Urine levels of 4-hydroxy-butyric acid; enzyme levels (leukocytes); genetic testing	Childhood-adulthood	Intellectual disability, autism, OCD, psychosis
Serine deficiency	AR	1p12, 7p11.2, 9q21.2	PHGDH PSPH PSAT1	Phosphoglycerate dehydrogenase, phosphoserine phosphatase, phosphoserine aminotransferase	Serine	<1	Plasma/CSF serine levels; genetic testing	Childhood	Aggression, intellectual disability
Cerebral folate deficiency	AR, auto-immune	11q13.3	FOLR1	FR1	Folate	<1	CSF 5-methyl-tetrahydrofolate levels	Childhood, adulthood	Intellectual disability, autism, psychosis
MELAS	Maternal	mtDNA	MT-TL1	Mitochondrial tRNA leucine 1	n/a	60	Serum lactate/pyruvate, neuroimaging, genetic testing (leukocytes, urinary sediment, tissue)	Adolescence/early adulthood	Stroke-like lesions/psychosis/depression
Wolfram disease	AR/Maternal	4p/mitochondrial	WFS1	Wolframin	n/a	170 reported cases	Genetic testing	Childhood–adolescence	Schizophrenia
Wilson disease	AR	13q14.2–q21	ATP7B	Copper transporting ATPase	Copper	25	Serum copper and ceruloplasmin, urine 24-hour copper excretion, genetic testing, liver biopsy	Adolescence/early adulthood	Extrapyramidal movements/psychosis/parkinsonism/dysphagia/dysrathria
Aceruloplasminemia	AR	3q25	CP	Ceruloplasmin	Iron	.5	Serum ceruloplasmin, genetic testing, neuroimaging	Adolescence/early adulthood	Ataxia, dystonia, dyskinesia, subcortical dementia
Pantothenate kinase–associated neurodegeration	AR	20p13	PANK2	Pantothenate kinase	Cysteine	1	Neuroimaging, genetic testing	Childhood–adolescence	Childhood dystonia, rigidity/adolescent palilalia, dysarthria, depression, lability, personality change, cognitive decline

AR, autosomal recessive; AD, autosomal dominant; tRNA, transfer ribonucleic acid; ATP, adenosine triphosphate; MELAS, Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes; OCD, obsessive-compulsive disorder.

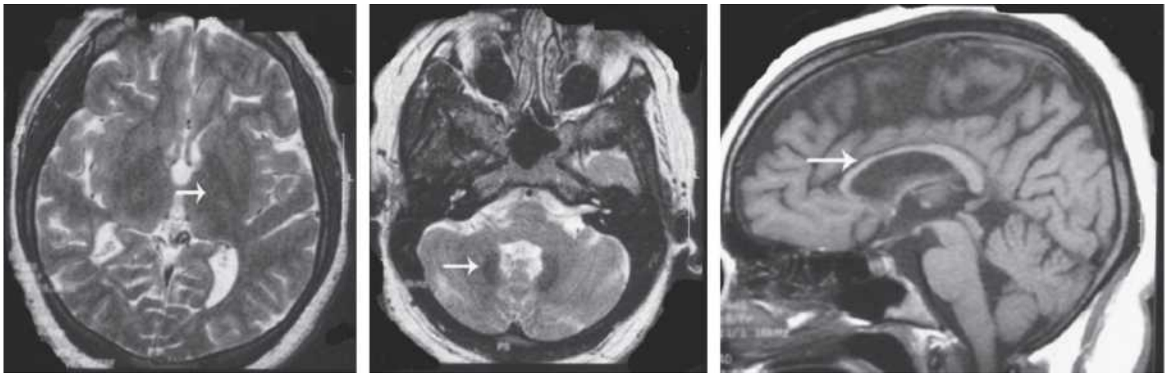

FIGURE 2.14–18. Scans of a 22-year-old woman patient with aceruloplasminemia and a schizophrenia-like psychosis. Reduced intensity of the basal ganglia (*left*) and dentate nucleus of the cerebellum (*middle*) on axial T_2-weighted imaging representing iron deposition (*arrows*) and T_1-weighted sagittal scan demonstrating anterior callosal thinning (*right, arrow*).

Aceruloplasminemia presents with a triad of diabetes mellitus, retinal degeneration, and neurologic symptoms. Neurologic signs may be preceded for many years by diabetes mellitus and anemia due to inefficient iron delivery. Ataxia and extrapyramidal movements such as blepharospasm, dystonia, dyskinesia, grimacing, and parkinsonism usually develop in the fifth decade. A subcortical picture of cognitive decline then follows with personality change, amotivation, psychomotor slowing, and executive deficits. Psychosis has been reported, although psychiatric symptoms have not been commonly described in aceruloplasminemia due to the rarity of the disorder. MRI findings typically show marked T_2 hypointensity in the regions of maximal iron deposition, posterior white matter tract hyperintensity, and superficial cerebral and cerebellar cortical hypointensity.

A 21-year-old woman on treatment for aceruloplasminemia was referred for neuropsychiatric assessment. There was a family history of aceruloplasminemia, and the diagnosis was made 12 months earlier after she was found to have abnormal liver function tests. She now presented with an 18-month history of schizophrenia-like psychosis and declining function in the absence of neurologic signs. Neuropsychological testing showed significant dominant hemisphere deficits. MRI showed bilateral iron deposition in the cerebellar dentate nuclei and thalami, frontal atrophy, and periventricular white matter hyperintensities (Fig. 2.14–18).

The diagnosis of aceruloplasminemia can be made biochemically with findings of absent ceruloplasmin, low serum copper, normal serum total iron-binding capacity, and moderately elevated ferritin. Early treatment with the combination of the iron-chelating agent deferiprone and phlebotomy has been shown to delay the onset of neurologic symptoms in patients with early or pre-symptomatic disease.

NEUROFERRITONOPATHY. Neuroferritinopathy is an extremely rare autosomal dominant disorder with about 120 reported cases. It is caused by a mutation in the ferritin light chain (FTL) gene on 19q13.33, with 7 mutations identified to date. The FTL gene codes for the FTL, a subunit of ferritin, which plays a role in the storage and release of iron from cells. In neuroferritinopathy, ferritin is unable to store iron, and excess iron is deposited in brain cells, especially the globus pallidus, putamen and dentate nuclei of the cerebellum. The clinical phenotype begins in the third or fourth decade with extrapyramidal symptoms and signs including chorea, ataxia, dystonia, tremor, dysphagia, and dysarthria. This clinical syndrome may often be misdiagnosed as Huntington disease, but the presence of orofacial action-specific dystonia is characteristic of neuroferritinopathy. A review of the neuropsychiatric features of neuroferritinopathy described 22 cases. This review identified that subcortical type cognitive changes began about 4 years after the onset of motor symptoms (mean onset age 47.2). Individual cases of obsessive-compulsive disorder, psychosis, emotional lability and aggression were noted.

The clinical diagnosis of neuroferritinopathy should be suspected in cases of adult-onset movement disorder, MRI findings of excess iron deposition in basal ganglia or cystic degeneration on T_2* MRI images. The presence of cystic degeneration may help distinguish neuroferritinopathy from aceruloplasminemia. Serum ferritin may be low. The genetic diagnosis is based on the identification of a mutation in the FTL gene. Treatment using iron chelation (deferiprone or desferrioxamine) with phlebotomy has been unsuccessful. Symptomatic treatment of the movement disorder using conventional treatments (tetrabenazine, benzhexol, l-dopa, benzodiazepines) is variably successful. Botulinum toxin may be of benefit in dystonia.

Ion Channel Disorders

Disorders of ion channels, or channelopathies, have become an increasingly recognized group of disorders affecting cellular ion channels involving Na^+, Ca^{2+}, K^+, and Cl^- in electrically excitable tissue, such as heart, muscle, and brain. A number of genetic ion channel diseases have now been well described, such as the long QT syndrome (LQT), periodic paralysis, and a range of monogenic seizure syndromes. The cardinal feature of ion channel disease is the disturbance of rhythmic function, best illustrated by epilepsy, with an abnormally synchronous discharge causing a seizure; other rhythmic CNS disturbances include ataxia, paralysis, and sensorineural deafness. Additionally, the possibility of developing an *acquired* abnormality of ion channel function has been recognized, particularly in autoimmune disorders.

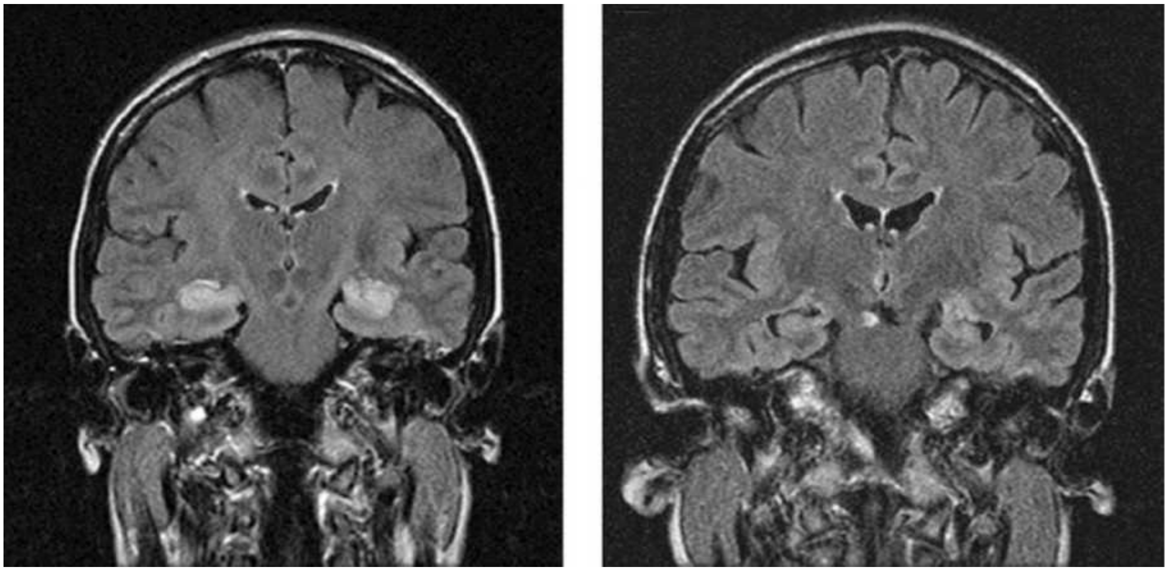

FIGURE 2.14–19. Voltage-gated potassium channel antibodies causing limbic encephalitis. *Left*, coronal fluid-attenuated inversion recovery magnetic resonance image shows bilateral mesial temporal hyperintensity. *Right*, identical coronal slice 3 months later, after treatment, shows resolution of hyperintensities but resultant hippocampal atrophy.

Voltage-Gated Potassium Channel Encephalopathy. Voltage-gated potassium channel (VGKC) antibodies are linked to a group of rare disorders characterized by abnormal neuromuscular excitability and CNS manifestations. These disorders form part of a larger group of CNS autoimmune disorders involving antibodies directed against targets on the neuronal surface. Isaac syndrome (acquired neuromyotonia) is associated with thymoma and VGKC antibodies but has no CNS manifestations. Morvan syndrome is a very rare disorder characterized by neuromyotonia, severe insomnia, excessive sweating, hypersalivation, and subacute encephalopathy commonly accompanied by psychotic features including delusions and hallucinations. High titers of VGKC antibodies are also detected. A group of patients with nonparaneoplastic limbic encephalitis have high titers of VGKC antibodies. This group presents with focal seizures with impaired awareness and a progressive amnestic syndrome with sleep disturbance as well as a fluctuating conscious state, personality change, and variable psychotic symptoms. MRI frequently demonstrates bilateral high signals in the hippocampal region. These symptoms are steroid responsive, including reversal of cognitive deficits and MRI changes once the diagnosis has been made (Fig. 2.14–19).

Timothy Syndrome. Timothy syndrome is a relatively recently recognized multisystem Ca^{2+} channelopathy in which ASDs occur in 80% of affected individuals, alongside cardiac arrhythmias and syndactyly. The affected $Ca_v1.2$ gene is widely expressed, particularly in the heart, brain, smooth muscle, and pituitary and adrenal glands. In the CNS, the highest expression is in the granular layer of the dentate gyrus of the hippocampus and cerebellum. It is not clear how changes to neuronal tissue excitability contribute to the development of autism, but it is likely to represent the endpoint of the interaction between tissue excitability and normal neurodevelopment. A greater understanding of the pathophysiology of this disorder may open up new avenues of research into the possible contribution of ion channel disturbance to polygenic neuropsychiatric disorders.

NEUROENDOCRINE DISORDERS

Hypothalamic Disorders

The relationship between the hypothalamus and the pituitary gland is complex. In brief, the release of six anterior pituitary hormones prolactin (PRL), growth hormone (GH), follicle-stimulating hormone (FSH), luteinizing hormone (LH), adrenocorticotrophin (ACTH), and thyrotropin or thyroid-stimulating hormone (TSH) is under the tonic influence of hypothalamic neuropeptides which travel from hypothalamic neurons via the portal system of the anterior pituitary to influence pituitary hormone-producing cells. AVP (vasopressin or ADH, antidiuretic hormone) and prolactin are produced in the hypothalamus and stored in the posterior pituitary. As a consequence of this relationship, endocrine abnormalities of the hypothalamus are largely manifest as abnormalities of pituitary function, as described below. However, the hypothalamus also has a number of nonendocrine functions that, when disturbed, may manifest as neuropsychiatric disorders.

The hypothalamus contains a number of unique neurons that create two neuropeptides known as *orexins* (formerly *hypocretins*). They are synthesized only in the hypothalamic neurons and share some homology to the gut hormone secretin. Orexinergic neurons project from the hypothalamus to a number of monoaminergic centers, including the locus ceruleus, raphe nuclei, and ventral tegmentum. A number of these monoaminergic systems are involved in the regulation of sleep, in which the hypothalamus is now recognized as a key center. The contribution of the hypothalamus is via sleep-promoting GABAergic neurons of the VLPO (ventrolateral preoptic area) and the wakefulness-promoting orexinergic neurons in the lateral hypothalamus. These pathways are closely linked to the circadian pacemaker in the suprachiasmatic nuclei and the regulation of other hypothalamic functions such as temperature, food intake, metabolism, and hormone secretion.

The orexinergic system is also important in the regulation of food intake and energy expenditure. Orexin production can increase food craving but is also inhibited by leptin, a hormone produced by adipocytes and stimulated by ghrelin. Ghrelin is secreted by the stomach just prior to a meal and is known to stimulate caloric intake. This provides a biochemical basis for the well-demonstrated phenomenon of sleep-deprivation-related catabolism despite adequate caloric intake described in animal models.

Narcolepsy

Narcolepsy is a sleep disorder characterized by excessive daytime somnolence, cataplexy (a sudden loss of muscle tone, often triggered by strong emotional reactions) and manifestations of disordered rapid eye movement (REM) sleep such as hypnogogic hallucinations, automatic behavior, and sleep paralysis. Although narcolepsy in dogs is linked to mutations in hypocretin-related genes, no such mutations have been described in humans, although in approximately 90% of patients, orexin levels in the CSF are low or nonexistent. Strong links to the human leukocyte antigen (HLA) system suggest an autoimmune basis. HLA-DR15 and HLA-DQ6 have been described in up to 85% of patients, but only 20% of controls and the DQB1*602 allele has been detected in 98% of patients with cataplexy. Gene–environment interaction is likely given the <25% concordance rate between monozygotic twins. Depression has been described in up to 25% of patients in some series, although others suggest a rate no higher than in the healthy population.

There appear to be three main groups of patients who experience psychosis in narcolepsy. First, sleep-related hallucinations have been described in up to 80% of patients with typical narcolepsy, most often nonverbal in form and associated with insight. It has been suggested that these symptoms are “narcoleptic” in origin and respond to treatment with stimulants and that REM-intrusion into wakefulness may be misdiagnosed as schizophrenia. Second, in an atypical form of narcolepsy, vivid hallucinatory experiences and daytime dream experiences associated with reduced insight may result in delusional preoccupations or misinterpretations resembling the delusional symptoms described in schizophrenia. Third, there is a small group of patients who have a dual diagnosis of schizophrenia and narcolepsy. The vast majority of psychoses associated with narcolepsy do appear to relate to stimulant use/misuse, as historically, dopaminergic agents such as dexamphetamine have been used to promote wakefulness. Newer agents such as modafinil may be less psychotogenic. Neuroleptics, however, worsen narcoleptic somnolence, as dopamine outflow is central to cortical wakefulness.

A 23-year-old man was referred for assessment of psychotic symptoms occurring in the setting of narcolepsy. At age 13, narcolepsy was diagnosed on the basis of daytime somnolence, cataplexy, sleep paralysis, and hallucinations. He showed an HLA-DR15 haplotype. He had begun dexamphetamine 2 years prior to assessment and started to experience auditory hallucinations, thought broadcasting, and a complex persecutory delusional system (whereby he was being studied through a device implanted in his head) within 6 months. His family described a precipitous psychosocial decline over 12 months, as the patient became unemployed, socially withdrawn, and increasingly disorganized. On mental state, he presented as fatuous and inappropriate, with a clear thought disorder. After commencing fluphenazine 4 mg and ceasing dexamphetamine, the hallucinations and delusions diminished, although he remained fatuous and amotivated. A diagnosis of narcolepsy-related psychosis secondary to dexamphetamine treatment was made, with a differential diagnosis of schizophrenia.

Hypothalamic Lesions

Hypothalamic obesity has been associated with lesions of the ventromedial nucleus and may initially involve aggressive behavior and hyperphagia until a new set weight is reached, at which time reduced appetite and activity may manifest. Rage reactions are also well described in animals with lesions in this region of the hypothalamus, although less so in humans with such lesions. Lateral lesions have been reported to result in an apathetic state. Thirst may also be impaired if ADH production is reduced by a hypothalamic lesion. Short-term memory dysfunction has been reported, particularly in lesions of the ventromedial and premamillary areas of the hypothalamus. Extensive hypothalamic lesions may produce features consistent with a dementing illness.

Hypothalamic disease may also result in abnormalities of thermoregulation. Temperature-sensitive neurons are located in the anterior hypothalamus, whereas the posterior hypothalamus mediates heat loss mechanisms. Acute lesions such as hemorrhage, infarction or those from surgical procedures may result in acute and paroxysmal hypothermia. Conversely, posterior hypothalamic lesions may result in paroxysmal hyperthermia with associated fevers and rigors.

Childhood tumors invading the anterior and basal hypothalamus, such as gliomas, midline cerebellar astrocytomas, and suprasellar ependymomas, may result in diencephalic syndrome which is manifested as motor hyperactivity, euphoria or inappropriate effect, increased alertness and emaciation despite normal caloric intake. If death does not ensue, the clinical picture may change to one of obesity and intermittent rage reactions. In adulthood, slower-growing tumors usually result in a dementia syndrome, endocrine dysfunction, and food intake dysregulation. More rapid or destructive processes present with disturbances of consciousness, temperature, and autonomic dysregulation. Craniopharyngiomas, germinomas, and gliomas are the most frequently reported tumors. Craniopharyngiomas, remnants of the embryonic Rathke pouch, can manifest in both childhood (most common) and adulthood, usually involving the posterior hypothalamus. As most are suprasellar, patients usually present with visual abnormalities and headaches. Hypogonadism, hyperprolactinemia, diabetes insipidus, and weight gain are common, as is cognitive deterioration and personality change without evidence of psychosis (Fig. 2.14–20).

Pituitary Disorders

The pituitary gland, or *hypophysis*, is an endocrine structure that sits in the midline sella turcica at the base of the brain in the middle cranial fossa and is covered by a dural fold (the *diaphragma sellae*). It secretes hormones regulating homeostasis and, through the release of trophic hormones, stimulates other distal endocrine structures. Its anterior lobe, the *adenohypophysis,* is under the direct functional control of the hypothalamus via the hypophysial-portal vascular connection in the pituitary stalk, through which stimulatory and inhibitory signals are sent to control the five distinct endocrine cell types that release pituitary hormones. The posterior lobe, the *neurohypophysis*, is predominantly a collection of axons from the supraoptic and paraventricular nuclei of the hypothalamus that secretes peptide hormones into the hypophyseal circulation. The posterior lobe is connected by the infundibulum in the pituitary stalk, and the release of oxytocin and vasopressin is controlled through the tuberoinfundibular pathway. Given the complex nature of the hypothalamic–pituitary axis, it is not surprising that clinical manifestations of pituitary disease are protean. Pituitary dysfunction may result from pituitary tumors or destructive disease processes.

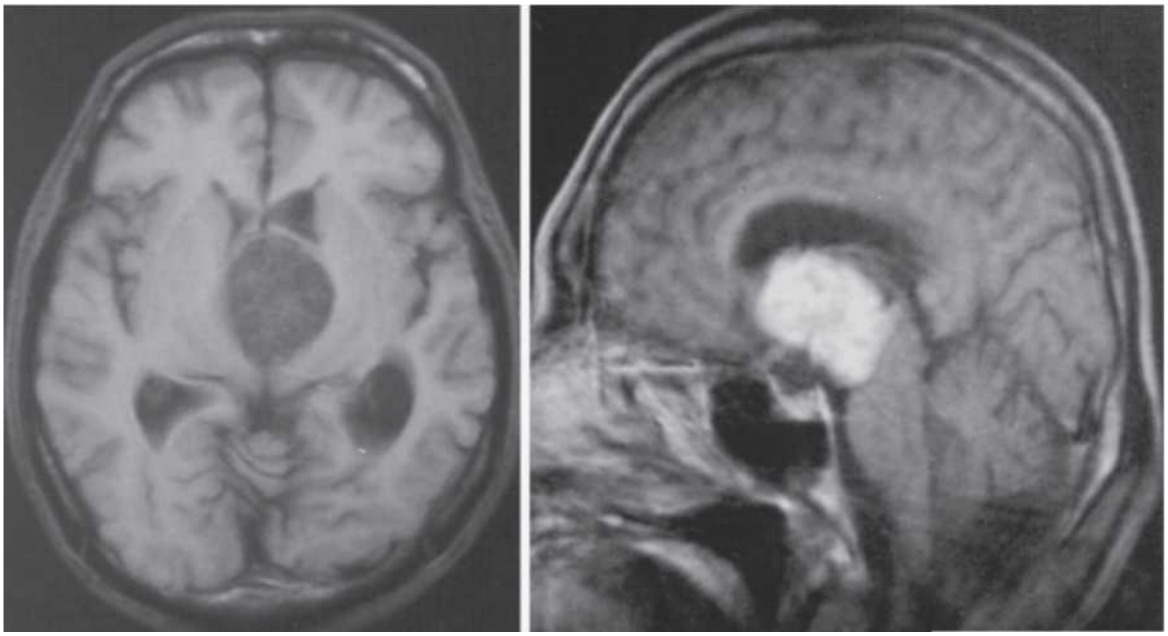

FIGURE 2.14–20. Craniopharyngioma causing the diencephalic syndrome. T_1-weighted magnetic resonance imaging scans of a patient who presented with emaciation and features of the diencephalic syndrome. *Left*, axial image demonstrating a slightly hypointense tumor within the third ventricle accompanied by moderate dilation of the lateral ventricles. *Right*, sagittal image obtained after the administration of gadolinium revealing that the tumor is homogeneously enhanced and is entirely confined within the third ventricle. (From Miyoshi Y, Yunoki M, Yano A, Nishimoto K, Konovalov AN. Diencephalic syndrome of emaciation in an adult associated with a third ventricle intrinsic craniopharyngioma: a case report. *Neurosurgery*. 2003;52(1):224, with permission. © Congress of Neurological Surgeons.)

Pituitary Tumors

Pituitary tumors are found in up to 20% of adults at autopsy and are frequently incidental findings in neuroimaging. They may result in an increase or decrease of hormone levels and produce symptoms by invading surrounding structures such as the hypothalamus.

Prolactinomas. The most common secretory tumors are prolactinomas. Dysregulated secretion of prolactin results in amenorrhea, infertility and galactorrhea in women and impotence and occasionally galactorrhea or gynecomastia in men. Data regarding the neuropsychiatric manifestations of hyperprolactinemia are lacking. There is some evidence of increased aggression in lactating animals associated with elevated prolactin levels and in hyperprolactinemic human subjects. Depression and anxiety symptoms also occur with greater frequency in this group of patients, which may respond to treatment with the dopamine agonist bromocriptine. Psychotic symptoms have been described in neuroleptic naive patients with hyperprolactinemia at a case report level. Treatment with antipsychotic agents may then result in further elevation of prolactin. Psychotic symptoms have also been described in patients receiving bromocriptine for the treatment of a prolactinoma, with both delusions and hallucinations recorded. The incidental finding of a pituitary adenoma on neuroimaging can complicate the assessment of patients with a psychotic illness stabilized on an antipsychotic agent with hyperprolactinemia. Impulse control disorders such as pathologic gambling have also been reported in patients receiving cabergoline (a potent dopamine type II receptor agonist) therapy for pituitary adenomas.

Growth-Hormone Secreting Tumors. These are the second most common functional pituitary adenomas (Fig 2.14–21). In adults, they result in *acromegaly* with soft tissue and bone enlargement, particularly involving the hands, feet, jaw and tongue, with a characteristic overall coarsening of facial features. There may be associated hypertension, congestive cardiac failure and obstructive sleep apnea and hypersomnolence. A number of psychiatric symptoms have been associated with acromegaly, largely at the case report level. Acromegalic patients appear to have a lifetime increased risk of affective disorders when compared with control subjects with and without a chronic somatic disease. There is no evidence of a preponderance of psychotic symptoms, although these are well described when acromegalic patients are treated with bromocriptine resulting in delusional symptoms, schizophrenia-like presentations, and visual hallucinations.

ACTH Secreting Tumors. ACTH secreting tumors are the next most common pituitary tumor type, resulting in excessive cortisol production or Cushing disease. The neuropsychiatric manifestations of increased cortisol release are discussed below in the description of adrenal disease.

Other Pituitary Tumors. Clinical symptoms resulting from tumors producing LH or FSH are very rare. Increased TSH production can occasionally result in hyperthyroidism, the psychiatric aspects of which are discussed below. As many as 30% of pituitary tumors are non-secretory and are often larger at diagnosis because of the lack of endocrine manifestations. If large enough, pituitary tumors may impinge on surrounding structures resulting in the classic visual field defect of bitemporal hemianopia, oculomotor palsies headache and occasionally hypothalamic syndromes (see above). Visual hallucinations in the context of visual field defects related to pituitary tumors impinging the optic chiasm have been reported. More often, these are of the simple, non-formed type, are associated with insight and can be exacerbated by treatment with bromocriptine.

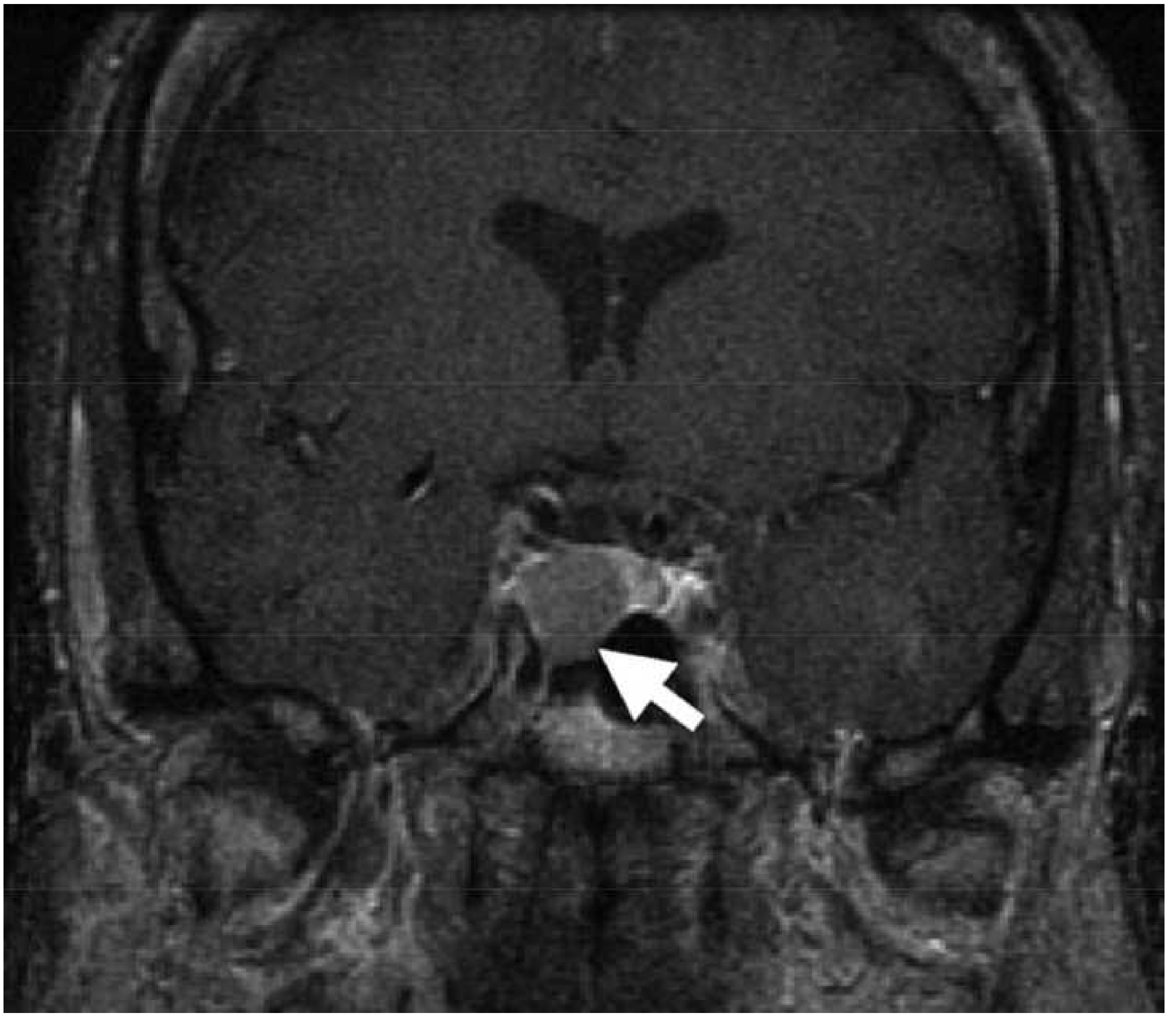

FIGURE 2.14–21. Pituitary adenoma causing acromegaly. Gadolinium-enhanced magnetic resonance imaging of the brain showing a well-defined mass in the right lateral aspect of the sella with a lack of enhancement (see white arrow) compared with the normal pituitary gland. Endocrinologic evaluation revealed elevated human growth hormone and insulin-like growth factor levels consistent with acromegaly. (From Koo CW, Bhargava P, Rajagopalan V, Ghesani M, Sims-Childs H, Kagetsu NJ. Incidental detection of clinically occult pituitary adenoma on whole-body FDG PET imaging. *Clin Nucl Med.* 2006;31(1):42, with permission. doi: 10.1097/01.rlu.0000191779.75532.05.)

Hypopituitarism

Deficiency or dysfunction of one or more of the pituitary hormones is referred to as hypopituitarism. In adulthood, this is usually caused by an acquired destructive process that may be traumatic (related to head injury), inflammatory, immune-mediated, vascular or as the result of compression from an adjacent tumor. Clinical manifestations are dependent on the hormones involved. GH is often first affected, followed by the gonadotrophins with associated amenorrhea and infertility in women and reduced libido and body hair loss in men. Low TSH can result in hypothyroidism, ACTH deficiency in fatigue, reduced appetite, weight loss, and impaired stress response. Vasopressin deficiency may also ensue, resulting in polyuria and thirst. A number of neuropsychiatric manifestations have been described in hypopituitarism in the absence of delirium. Memory impairment, sleep disturbance and personality change are commonly but variably reported. Visual and auditory hallucinations have been described at a case report level, and several observers have reported an absence of affect on the mental state examination. A systematic study of psychiatric co-morbidity in hypopituitarism found more psychiatric symptoms than those expected in chronic diseases such as diabetes mellitus, most particularly depression and anxiety. Psychosis is uncommon, but depression appears to be more common in those who have hypopituitarism associated with a traumatic brain injury.

More specific symptoms are attributable to particular hormone deficiencies. GH deficiency may result in reduced energy, depressed mood, anxiety, emotional lability, and impulsivity or impaired self-control. Testosterone deficiency has been associated with depression, irritability and insomnia. Fatigue, social withdrawal and negativism have been reported when ACTH is low. Finally, as in hypothyroid states, fatigue, depression, insomnia or hypersomnia and psychotic symptoms have been described when TSH levels are low. There can be considerable variation in the type of psychotic symptoms described. In general, appropriate correction of the hormone deficiency and addressing the underlying cause results in the resolution of these features.

Thyroid and Parathyroid Disorders

Normal thyroid function involves the production of the two thyroid hormones, L-thyroxine (T_4) and 3,5,3′-triiodo-L-thyronine (T_3), which are required for the regulation of a number of metabolic processes. Increased thyroid hormone production results in hypermetabolism or increased caloric utilization and the other clinical features of hyperthyroidism. Hypometabolism and the features of hypothyroidism (sometimes referred to as myxedema) result from reduced hormone production. Psychiatric symptoms that accompany either of these states have been subject to more systematic study than those associated with pituitary disorders, as primary thyroid disorders have a much higher incidence in the population.

Hyperthyroidism. Hyperthyroidism is the result of the excessive production of thyroid hormones. This may result from a toxic multinodular goiter, a single functioning adenoma or from the presence of a thyroid stimulator, such as a thyroid-stimulating antibody in Grave disease. Exogenous thyroid hormone can produce a similar picture as can disorders of thyroid hormone storage consequent to autoimmune thyroiditis.

Depressive symptoms are not only the most common psychiatric features seen in hyperthyroidism, occurring in up to 30% of patients, but frequently occur prior to the onset of the other physical features. This includes lowering of mood as well as neurovegetative disturbance such as insomnia, reduced libido, weight loss, and fatigue. However, as distinct from depression, appetite is invariably increased in hyperthyroidism. The severity of the depressive symptoms has not been found to be related to the severity of hyperthyroidism as measured by subsequent thyrotoxic features and the extent of the biochemical abnormalities.

Older adult patients with thyrotoxicosis may present with predominantly apathy, depression, and weight loss rather than increased psychomotor activity. Typically this presentation is of slower onset, the neurologic and ophthalmologic features are less prominent, but cardiovascular events such as exacerbation of angina, cardiac failure, and atrial fibrillation are more prevalent. Anxiety symptoms presenting as generalized anxiety are also common, with a prevalence of between 10% and 20%, but panic and agoraphobia are relatively uncommon. Like depressive symptoms, anxiety may occur prior to the other features of hyperthyroidism but more often correlates with the severity of the thyrotoxic features. Anxiety and depressive features are often comorbid. Manic symptoms are less common in hyperthyroid states with a prevalence between 2% and 5% but may be difficult to distinguish from psychomotor agitation and anxiety. Psychotic symptoms, including paranoid delusions and auditory hallucinations, while historically reported as common, have a true prevalence of between 2% and 5%. The depressive syndrome seen in hyperthyroidism rarely requires treatment other than which is used to restore the euthyroid state. However, some features, such as anxiety, fatigue and loss of function, may persist for as long as 12 months after a euthyroid state has been achieved.

Cognitive dysfunction is reported between 5% and10% of patients with thyrotoxicosis, but to a lesser degree than seen in hypothyroidism. Patients may present with slow processing speeds, impairments in immediate memory, defective higher-level problem solving or frank delirium. Mild disorders of attention and concentration are common, but their severity does not always correlate with the severity of thyrotoxicosis. These deficits invariably respond to the reversal of the thyrotoxic state.

Hypothyroidism. The most common cause of hypothyroidism in adults is primary autoimmune hypothyroidism related to antithyroid antibodies. Other causes include treatment for hyperthyroidism, drug-related effects, and iodine deficiency. Supra-thyroid causes (hypothalamic or pituitary dysfunction) account for less than 5% of cases. The symptoms of hypothyroidism may include fatigue, lethargy, weight gain, constipation, cold intolerance, stiffness and cramping of muscles, hair loss, cognitive slowing, and depression. Signs include hypothermia, bradycardia, dry skin, sparse hair, periorbital swelling, thickening of the tongue, coarsening and deepening of the voice, and a characteristic prolonged relaxation phase of deep tendon reflexes. This clinical picture is often referred to as *myxedema*. Hypothyroid patients may present with a variety of psychiatric symptoms ranging from mild cognitive slowing and depression to frank encephalopathy, which often predates other physical features. Functional neuroimaging studies have demonstrated global hypometabolism and more specific areas of hypoperfusion in the posterior cingulate, insula, fusiform gyrus, and right parieto-occipital and primary motor cortex.

Cognitive deficits are the most common neuropsychiatric features of hypothyroidism, occurring in up to 50% of cases. Psychomotor speed, memory, and visual-perceptual skills are often impaired. Difficulties with constructional skills and reduced performance in trail-making and maze tasks also suggest prominent executive deficits. The severity of these disorders is correlated with the degree of biochemical abnormality, and although largely corrected by a return to a euthyroid state, some cognitive deficits may remain, particularly in the older adults or those with reduced cognitive reserve. Severe disturbance of consciousness, including coma and delirium, may be encountered but may be associated with other metabolic changes related to the hypothyroid state, such as hyponatremia. Depression is only slightly less common than cognitive disturbance, reported in approximately 40% of patients, but appears less closely related to the severity of biochemical hypothyroidism. Low mood, fatigue, anhedonia, reduced concentration, and hypersomnolence are the most commonly described features of the depressive syndrome in hypothyroidism. These features predictably respond to treatment of the hypothyroid state. The origin of depression in hypothyroidism appears to relate to the role of thyroxine in serotonergic transmission, such that reduced thyroid input reduces serotonergic tone and lowers the threshold toward the development of depressive symptoms. Conversely, thyroid replacement restores central serotonin activity in concert with improvement in depressive symptoms. This may also underpin the adjunctive antidepressant effect of thyroxine.

In contrast, manic and hypomanic symptoms have been infrequently reported in association with hypothyroidism. However, hypothyroidism may be a risk factor for the development of bipolar disorder, particularly the rapid cycling form and treatment of otherwise refractory mood disorders with thyroid hormones has occasionally been shown to be effective. Generalized anxiety symptoms are described in up to 30% of patients and are strongly correlated with depressive symptoms but not with biochemical severity. There is some evidence that hypothyroidism is a risk factor for suicidal ideation in patients with depression. Psychotic symptoms, including paranoid ideas, misidentification, visual and auditory hallucinations, and thought disorder, were originally thought to be common (and described as “myxedematous madness”) but likely occur in less than 5% of all patients with hypothyroidism and tend to emerge after the onset of physical symptoms. Although these symptoms also respond to appropriate thyroid hormone treatment within days to weeks, rapid titration of hormone doses may exacerbate psychosis. Careful addition of a low dose of antipsychotic to thyroxine has been reported to be well-tolerated and results in an earlier remission of psychosis.

“Subclinical” Hypothyroidism. Thyroid hormone abnormities may occur without overt functional hypothyroidism. Designated *subclinical hypothyroidism*, these scenarios can be further classified into elevated TSH without changes in thyroid hormones (Grade II hypothyroidism), abnormal TSH response to stimulation with TRH (Grade III) and the presence of antithyroid antibodies with no thyroid hormone system abnormalities (Grade IV). Grade II hypothyroidism has been associated with depressive disorders. Patients with major depressive disorders have an increased incidence of Grade II hypothyroidism, and some studies show these patients respond poorly to conventional treatment. Grade II hypothyroidism may be a risk factor for major depressive disorders. Cognitive disturbance, particularly memory and psychotic symptoms, have also been reported in Grade II hypothyroidism. Although there is some evidence of improvement in cognition, mood and psychosis, treatment of subclinical hypothyroidism is controversial.

Hashimoto Encephalopathy. Hashimoto encephalopathy (HE), also described as steroid-responsive encephalopathy associated with autoimmune thyroiditis (SREAT), is best defined as an uncommon autoimmune encephalopathy of unknown etiology associated with high titers of serum antithyroid (usually antithyroid peroxidase ± anti-thyroglobulin) antibodies. The role of antithyroid antibodies is controversial. There is no direct evidence that these auto-antibodies exert direct effects on CNS tissue, and they may be epiphenomena of another autoimmune process. Additionally, the base rate of elevations of these autoantibodies in the population is as high as 10%. Elevated serum antithyroid antibodies are associated with other thyroid disorders (Grave disease, de Quervain thyroiditis, primary hypothyroidism, and colloidal goiter) and other auto-immune disorders (including diabetes mellitus, Addison disease, and pernicious anemia).

HE is more common in women with a mean age of onset between 45 and 50 years. There may be a history of other autoimmune diseases. The clinical picture is one of encephalopathy with progressive cognitive decline, although the course may also be relapsing and remitting. Common features include seizures (>95% of cases), stroke-like episodes (> 65% of cases), and memory dysfunction. Neuropsychiatric features include agitation and restlessness, apathy, and social isolation. Visual hallucinations are frequently reported (>90%), as are other disorders of perception and delusions. HE may also present as a depressive illness, although typical presentations of psychiatric illness, such as depression, in the setting of mildly elevated antithyroid peroxidase antibodies are unlikely to be due to HE. However, when antithyroid peroxidase levels are very high (>1,000 IU/l), the likelihood of them being associated with neuropsychiatric HE is much greater.

Routine investigations are often normal, although patients will often show slowing on EEG, with the degree of slowing being proportional to the clinical severity of the syndrome, and some patients may show triphasic waves (Fig. 2.14–22). The majority of patients respond to corticosteroid treatment with complete resolution of the neuropsychiatric symptoms.

A 66-year-old woman presented with a 4-month history of cognitive decline, including short-term memory and language deficits. She had a generalized tremor and developed a delusional belief that her husband wished to harm her. She became agitated and disoriented, particularly at night, and also described visual hallucinations of small animals in her room. No neurologic features, including myoclonus, were described. At this time, all routine blood investigations (including complete blood count, serum urea, electrolytes, creatinine, liver enzymes, thyroid hormone assays, folate and B_{12} level and CT brain) were normal. One month later, she suffered a generalized tonic–clonic seizure. Routine blood investigations and examination of the CSF were normal. Angiography of the anterior and posterior cerebral circulation was normal. MRI scan was normal, and EEG was unremarkable. On mental state examination, she was disoriented with impaired attention and markedly poor short-term memory (particularly nonverbal memory). Her speech was characterized by impaired verbal fluency and word-finding problems. There was a concrete thinking style with reduced abstract reasoning as well as poor judgment. There were no depressive or psychotic features at this time. Her then mental state rapidly deteriorated. She became virtually mute, with poor concentration and attention and required assistance with all aspects of personal care. She had bilateral grasp and pout reflexes as well as diffuse hyperreflexia. She also had bilateral and multifocal myoclonus. In the face of previously normal investigations, the diagnosis of Hashimoto encephalopathy was considered. Thyroid hormone (TSH, T4, T3) levels were within normal limits. Antithyroid peroxidase antibodies (anti-TPO) were raised in titer (640 IU/ml, normal <50 IU/mL), consistent with this diagnosis. A course of prednisolone 60 mg was commenced with marked improvement in her mental state within 2 weeks.

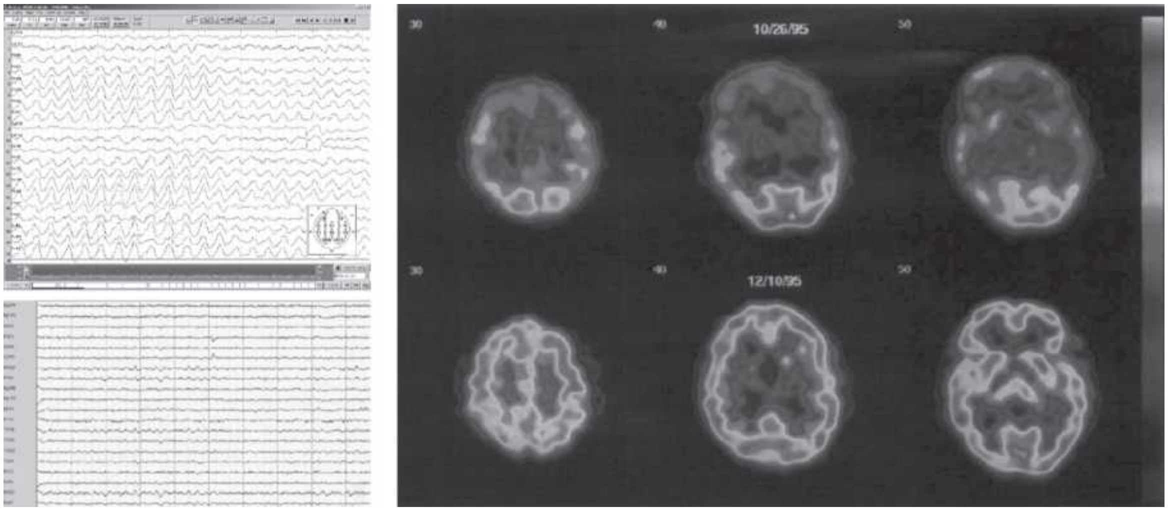

FIGURE 2.14–22. Reversibility in Hashimoto encephalopathy. *Left*, electroencephalogram (EEG) before and after treatment in a 38-year-old woman with psychosis and an antithyroid peroxidase antibody titer of 779. *Top*, pretreatment with corticosteroids showing general slowing with high-voltage (2 to 3 Hz) δ biphasic and triphasic waves. *Bottom*, EEG after corticosteroid treatment showing α frequency, with occasional θ waves (5 to 6 Hz), mainly in posterior regions. (From Sporis D, Habek M, Mubrin Z, Poljakovic Z, Hajnsek S, Bence-Zigman Z. Psychosis and EEG abnormalities as manifestations of Hashimoto encephalopathy. *Cog Behav Neurol.* 2007;20(2):138, with permission. doi: 10.1097/WNN.0b013e31804c702e.) *Right*, single-photon emission computed tomography (SPECT) showing gross global hypoperfusion in all nonoccipital regions in a 59-year-old woman with rapidly progressive cognitive impairment and myoclonus, with a Mini-Mental State Examination (MMSE) score of 20 (*top*) and after significant clinical improvement 6 weeks later, when MMSE score was 27 (*bottom*). (See Color Plate.) (From Forchetti CM, Katsamakis G, Garron DC. Autoimmune thyroiditis and a rapidly progressive dementia: Global hypoperfusion on SPECT scanning suggests a possible mechanism. *Neurology*. 1997;49:623, with permission. doi: 10.1212/WNL.49.2.623. © American Academy of Neurology.)

Parathyroid Disorders

The parathyroid glands are small accessory glands that are anatomically associated with but functionally distinct from the thyroid gland. Their sole function is to maintain serum calcium levels within a narrow range to permit optimum functioning of the nervous and muscular systems through the release of parathyroid hormone (PTH). The release of PTH increases serum calcium via stimulating osteoclasts in bone to release calcium and by increasing its absorption in the gut and kidneys.

Hyperparathyroidism. Hyperparathyroidism is usually diagnosed after an incidental finding of hypercalcemia on routine blood tests, and 50% of patients with hyperparathyroidism are asymptomatic. Increased PTH release is usually caused by a single functioning adenoma, although multiple adenomas may occur as part of Multiple Endocrine Neoplasia Syndrome (MENS). Symptoms are principally those secondary to hypercalcemia. These include fatigue, general malaise, proximal muscle weakness, renal colic, abdominal pain, and cognitive decline. There is often little to find on examination, although real calculi, nephrocalcinosis, or bone changes (osteitis fibrosa cystica, now rare) may be seen on radiologic investigations ordered for other reasons. Although a number of psychiatric symptoms have been described in hyperparathyroidism (as part of the symptomatic tetrad of “bones, stones, moans and psychic groans”), the prevalence overall is likely to be less than 10%. These symptoms correlate with both the duration and severity of the associated hypercalcemia. Affective disorders predominately of the depressive type are the most frequently recorded, with anxiety often comorbid. Psychotic disorders are reported, although infrequent, with persecutory and paranoid delusions predominating. Cognitive changes, usually related to short-term memory loss, disorders of attention and acute confusional states often occur, and severe derangement of calcium metabolism may present with somnolence and coma. Neuropsychiatric symptoms are often the initial presentation in the older adults or those with limited cognitive reserve. The more severe the hypercalcemia, the more severe the psychiatric disturbance, but symptoms generally respond to appropriate treatment such as parathyroidectomy. When PTH levels are mildly raised in the setting of normocalcemia (most commonly due to vitamin D insufficiency, increasingly prevalent in developed countries), a psychiatric disturbance is uncommon.

Hypoparathyroidism. The most common cause of impaired PTH production in adults is inadvertent surgical removal during thyroid surgery or excessive removal for hyperparathyroidism. Clinical features related to hypocalcemia, particularly symptoms of neuromuscular excitability, including paresthesias, muscle cramps, carpopedal spasm, facial grimacing progressing to laryngeal spasm and convulsions. An examination may reveal features of tetany, reduced or absent deep tendon reflexes, papilledema, and QT interval prolongation on ECG.

Delirium is now understood to be the most common neuropsychiatric manifestation. One large study available demonstrated cognitive impairment in 39% of patients, affective or neurotic symptoms in 12%, psychotic symptoms in 11% and nonspecific affective disturbance in 21% of patients. Again, the severity of symptoms directly relates to the degree of hypocalcemia and appropriate normalization results in the resolution of these symptoms, although persistent psychosis has been noted when associated hypomagnesemia was not addressed.

Adrenal Disorders

The adrenal glands produce both adrenal steroids from the cortex and catecholamines from the medulla. The adrenal cortex produces glucocorticoids, principally cortisol, under the stimulatory effects of ACTH from the anterior pituitary controlled by a negative feedback loop. Aldosterone, a mineralocorticoid, originates from the adrenal cortex but is controlled by the renin-angiotensin system influenced by body volume and potassium balance. Adrenal androgens, predominately DHEA (dehydroepiandrosterone), are also regulated by the ACTH system and undergo peripheral conversion to sex-determining androgens. Hyperfunction or hypofunction of these adrenal systems results in distinct clinical syndromes with complex physical and neuropsychiatric manifestations.

Hyperadrenalism. Hyperadrenalism (Cushing syndrome) occurs when the adrenal gland produces excess corticosteroids, usually as a result of ACTH overproduction from the anterior pituitary secondary to a pituitary adenoma (Cushing disease). Other causes include ACTH production from a non-endocrine tumor, an adrenal neoplasm or, most commonly, exogenous steroids, usually prescribed for treatment of a steroid-sensitive medical condition. Patients with Cushing syndrome present with the classic features of hypertension, muscle weakness and fatigue, osteoporosis, cutaneous striae and easy bruising. A characteristic pattern of obesity is seen involving the upper face (resulting in a “moon face”), back (“buffalo hump”), and mesentery resulting in truncal obesity. Women may experience hirsutism, acne and amenorrhea, men, decreased libido and impotence. Neuropsychiatric features are well described, with depression the most frequently noted psychiatric symptom in up to 70% of patients. Anxiety is commonly comorbid with depression in up to 50% of sufferers. Depressive symptoms may present prior to physical symptoms. They respond to treatment of the underlying cause, and the response correlates with lowering plasma cortisol levels. Elevated mood is infrequently reported in Cushing syndrome, reported in less than 10% in most case series. Delirium is also relatively uncommon and is usually a marker of a supervening infection or other metabolic disorder such as metabolic alkalosis. Although dramatic case reports of Cushing syndrome presenting with psychosis and obfuscating the underlying condition are reported, this form of presentation is rare. When present, psychotic symptoms are almost always mood-congruent delusional beliefs and derogatory auditory hallucinations associated with a depressed mood. Cognitive impairment occurs in over 50% of patients, presenting as deficits in verbal memory, attention, and visuomotor and visuospatial function. The degree of memory impairment, and its reversibility, appear to correlate with hippocampal volume, suggesting that cognitive impairment is partially driven by the effect of excess glucocorticoids on hippocampal neurons. A more general reduction in gray matter volume, cortical thickness, and white matter integrity has been demonstrated and can persist after correction of the hypercortisolism.

With exogenous administration of steroids, manic symptoms are most frequently reported, followed by delirium, depression, and psychotic symptoms. Mixed affective states also appear over-represented. The severity of these symptoms may be dose-related, reproducible in the individual and minimized with the use of divided doses of steroids.

Manic symptoms may respond to both cessation of the exogenous steroid administration and pharmacotherapy with mood-stabilizing agents such as lithium and valproate or treatment with antipsychotics.

Hypoadrenalism. Primary hypofunction of the adrenocortical system may result from a primary process at the level of the gland, such as destruction by an autoimmune process, referred to as Addison disease (most common), other inflammatory or destructive processes, or an inborn failure of enzyme function. It may also arise secondary to dysfunction of the hypothalamic–pituitary axis or withdrawal of exogenous steroids. The onset of symptoms is usually insidious with progressive fatigue, weakness, anorexia, nausea, abdominal pain, weight loss, cutaneous and mucosal pigmentation, hypotension and hypoglycemia. Other findings include hyponatremia, hyperkalemia, and metabolic acidosis. Reports of neuropsychiatric symptoms in Addison disease are uncommon, although the true prevalence is uncertain due to probable under-reporting. Depression, reduced motivation and energy and behavioral changes predominate. Memory dysfunction is the most common form of cognitive disturbance. Paranoid symptoms and delusions are less common. Catatonia and self-mutilation are rare but dramatic presentations. Auditory and visual hallucinations, changes in conscious state, irritability, insomnia, and nightmares often herald an addisonian crisis with frank delirium, coma, and seizures. Hyponatremia and metabolic acidosis may contribute to delirium and cognitive deficits. These symptoms appear to largely resolve, including those of adrenal crisis when treatment with adequate doses of corticosteroids is commenced. Recent data indicates that higher corticosteroid doses reduce the occurrence and severity of neuropsychiatric symptoms, particularly depressive symptoms. The use of other psychotropic agents is rarely indicated.

Hyperaldosteronism. Excess of the major adrenal mineralocorticoid aldosterone can arise from the adrenal gland (primary aldosteronism) or from an extra-adrenal site (secondary aldosteronism). Primary aldosteronism is usually due to an aldosterone-producing adenoma (Conn syndrome) or bilateral cortical hyperplasia. Secondary aldosteronism relates to disturbance of the renin-angiotensin system and hypovolemia. The characteristic features of Conn syndrome are hypokalemia, hypertension, muscle weakness, fatigue, polyuria, and polydipsia. Metabolic alkalosis and hypomagnesemia may also occur, and edema is characteristically absent. Depression has been identified as one of the major features, but more recent studies have demonstrated features of generalized anxiety with individual cases of associated depression, features of obsessive-compulsive disorder and panic without specific gender differences. Untreated primary hyperaldosteronism and resulting severe hypertension can result in vascular dementia if left untreated. Although treatments, such as surgical excision of a functioning adenoma, are effective in reversing the physical symptoms of hyperaldosteronism, the effect on anxiety and other symptoms is unknown.

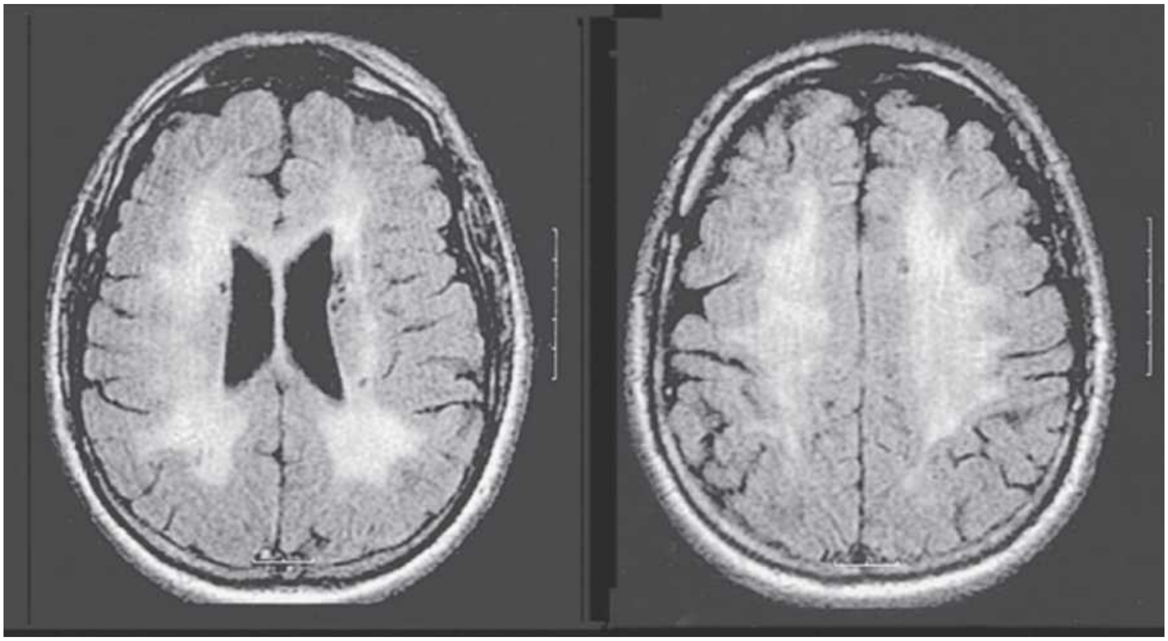

FIGURE 2.14–23. Subcortical dementia due to hyperaldosteronism in a 48-year-old man with untreated hyperaldosteronism. Axial fluid-attenuated inversion recovery images show bilateral extensive periventricular and subcortical white matter hyperintensity with multiple lacunar infarcts in the basal ganglia, consistent with long-standing untreated hypertension.

A 48-year-old man with no past psychiatric history had raised family concerns following 3 years of unusual behavior, social withdrawal, and poor self-care. He lost contact with friends, began to gamble heavily, had a number of car accidents, and was disinhibited and socially inappropriate. He had recently started a fire in his residence with one of many discarded cigarettes. On examination, he was hypertensive (240/120 mm Hg) and had signs of left ventricular hypertrophy but no neurologic abnormalities. On cognitive assessment, he had a marked dysexecutive syndrome. Serologically he had a mild metabolic alkalosis and hypokalemia (3.2 mmol/L). Serum cortisol, ACTH, and vasculitic screen were normal. ECG showed first-degree heart block and left ventricular hypertrophy, which was confirmed on echocardiography. There was a failure of aldosterone suppression with saline challenge and bilateral adrenal hyperplasia on abdominal CT. Brain MRI showed extensive severe periventricular and subcortical white matter disease (Fig. 2.14–23). He was diagnosed with Binswanger dementia secondary to hypertension associated with hyperaldosteronism.

Pheochromocytoma. Pheochromocytomas are tumors that secrete catecholamines and most commonly originate in the chromaffin cells of the adrenal medulla but may rarely arise from similar cells in sympathetic ganglia. Familial forms are associated with MENS type IIa and b and von Recklinghausen neurofibromatosis. Secretion of catecholamines may be continuous or sporadic, sometimes resulting in characteristic paroxysmal symptoms of headache, profuse sweating, palpitations, Raynaud phenomenon, tremor, nausea, vomiting, abdominal and chest pain. The triad of palpitations, headache, and profuse sweating is the most sensitive and specific for pheochromocytoma. Clinical findings include hypertension, pallor, postural hypotension, and signs of chronic hypertension such as retinopathy. Paroxysmal symptoms occur in 40% of pheochromocytomas and may present as a phenocopy of anxiety symptoms, particularly a panic episode, and are often diagnosed as such. These may occur spontaneously or may be precipitated by exercise, postural change, raised intra-abdominal pressure, or emotional excitement or shock. The severity of the episodes may also vary, and anxiety may persist for some time after the attack. Psychosis and cognitive deficits have not been described, although there are several case reports of a secretory pheochromocytoma causing a relapse of a previously treated or quiescent psychotic disorder. Diagnosis is made by detecting increased urinary secretion of catecholamines or catecholamine metabolites. Location of the underlying tumor and surgical excision with alpha-adrenergic blockade is effective in resolving most anxiety symptoms if paroxysmal episodes are terminated.

Neuroendocrine Tumors

While the term “neuroendocrine” is commonly used to refer to the interaction of the endocrine and nervous systems, histologically, it refers to a particular type of cell.

Neuroendocrine cells are cells that release a hormone or regulatory peptide into the circulation in response to a neural stimulus. The most extensive neuroendocrine system is in the gastrointestinal tract and associated organs, and when these tissues release excess hormones, a range of neuropsychiatric syndromes can occur.

Carcinoid Syndrome. A carcinoid is a neoplasm of neuroendocrine cells that synthesize and secrete serotonin in the respiratory system and gastrointestinal tract. The main symptoms are flushing (often severe facial flushing with bronchial tumors), diarrhea, wheezing, and hypotension. Although peripherally secreted serotonin does not cross the blood–brain barrier, a number of psychiatric symptoms have been described. These include depression (in up to half of patients), anxiety, sleep disorders and, rarely, psychosis. These symptoms may be more common with metastatic disease. A series of 23 patients found that over half experienced personality change consisting of increased irritability and impulsive aggressive thoughts or behavior, some meeting criteria for impulse control disorder. These changes often preceded the other physical symptoms. Depression was far less common, and psychotic symptoms were not detected.

Insulinoma. Insulinomas are functioning b islet cell tumors of the pancreas, which result in unregulated insulin secretion, at times with abrupt fluctuations. Clinical features include fasting hypoglycemia (relieved by glucose ingestion) and weight gain. Hypoglycemia is characterized by hunger, restlessness, palpitations, flushing and ataxia but may also include malaise, anxiety, depersonalization and derealization or seizures. Neuroglycopenic symptoms such as confusion, personality change, amnesia, and bizarre behavior are also reported. A more subacute syndrome consisting of clumsiness, disinhibited or aggressive behavior (and associated amnesia for these episodes) may mimic alcohol intoxication or a seizure disorder. Global and irreversible cognitive deficits may result if hypoglycemia and neuroglycopenia are long-standing. Surgical therapy is the most definitive treatment, but hyperglycaemic treatment with the somatostatin analog octreotide can be helpful.

Glucagonoma. This is a rare pancreatic islet cell that secretes glucagon. The presenting features are impaired glucose tolerance and diabetes, and a severe migratory and necrolytic erythema. Anxiety and agitation, as well as depression, may accompany these features.

FUTURE DIRECTIONS

An awareness of the relationship between neurometabolic and neuroendocrine disease and psychiatric illness by psychiatrists and other physicians provides for more diagnostic precision, in addition to allowing for improved treatment of psychiatric comorbidity in these disorders. Psychiatric symptoms can have a profound effect on long-term quality of life, and the recognition of a comorbid psychiatric illness in these disorders can improve health outcomes and quality of life for both patient and caregiver alike. As psychiatrists are best placed to manage major psychiatric disturbances, the involvement of a psychiatrist in the care of these patients is crucial, whether a psychiatric illness is the first or only presentation of illness or if it develops later in the course of the illness. In Table 2.14–1, we have included a summary of the key neurometabolic disorders, their presentation and diagnosis.

Recognizing, understanding, and exploring the links between neurometabolic and neuroendocrine disorders and major psychiatric syndromes can also provide insights into the neurobiologic basis of psychiatric illness. For example, the recognition of the elevated rates of schizophrenia-like psychosis in MLD has highlighted the possible role of myelinated structures as a possible anatomical substrate for the functional disconnectivity that has been well described in schizophrenia. Following a group of publications by Hyde and Weinberger on the links between schizophrenia and leukodystrophies, a large body of research in the subsequent 20 years has produced real insights into the role of white matter structures in schizophrenia at the genetic, developmental, and structural levels. Additionally, the emerging data on the exceedingly rare movement disorder ChAc suggesting significantly elevated rates of OCD and its unique predilection for neuropathology in the caudate and putamen has further highlighted the role of disturbed prefrontal-subcortical circuitry in OCD “proper.” Further delineation of the neurobiologic link between the cellular and metabolic deficit in these syndromes and the psychiatric disturbance they are commonly associated with may yet yield further insights into the neurobiologic basis of those disorders that present most commonly in psychiatric practice but that have only yet afforded limited insights into their underlying pathophysiology.

FURTHER READINGS

Anglin RE, Rosebush PI, Mazurek MF. The neuropsychiatric profile of Addison’s disease: Revisiting a forgotten phenomenon. *J Neuropsychiatry Clin Neurosci*. 2006;18:450.

Bigot A, Brunault P, Lavigne C, et al. Psychiatric adult-onset of urea cycle disorders: A case series. *Molecular Genetics and Metabolism*. 2017;12:103–109.

Bonnot O, Kluenemann HH, Sedel F, Tordjman S, Cohen D, Walterfang M. Diagnostic and treatment implications of psychosis secondary to treatable metabolic disorders in adults: a systematic review. *Orphanet J Rare Dis*. 2014;9:65.

Borer MS, Bhanot VK. Hyperparathyroidism: Neuropsychiatric manifestations. *Psychosomatics*. 1985;26:597.

Fraidakis MJ. Psychiatric features in cerebrotendinous xanthomatosis. *Trans Psychiatry*. 2013;3:e302.

Gargus J. Ion channel functional candidate genes in multigenic neuropsychiatric disorders. *Biol Psychiatry*. 2006;60:177.

Hanin C, Arnulf I, Maranci J, et al. Narcolepsy and Psychosis: A systematic review. *Acta Psychiatr Scand*. 2021;144(1):28–41.

Hayflick SJ, Westaway SK, Levinson B, Zhou B, Johnson MA. Genetic, clinical, and radiographic delineation of Hallervorden-Spatz syndrome. *N Engl J Med*. 2003;348:33.

Heinrich TW, Grahm G. Hypothyroidism presenting as psychosis: Myxedema madness revisited. *Prim Care Companion J Clin Psychiatry*. 2003;5:260.

Lynch S, Merson S, Beshyah SA, Skinner E, Sharp P. Psychiatric morbidity in adults with hypopituitarism. *J R Soc Med*. 1994;87:445.

Medvei VC, Cattell WR. Mental symptoms presenting in phaeochromocytoma: A case report and review. *J R Soc Med*. 1988;81:550.

Miyajima H. Aceruloplasminemia, an iron metabolic disorder. *Neuropathology*. 2003;23:345.

Mocellin R, Walterfang M, Velakoulis D. Hashimoto’s encephalopathy: Epidemiology, pathogenesis and management. *CNS Drugs*. 2007;21:799.

Muelly 1 ER, Moore GJ, Bunce SC, et al. Biochemical correlates of neuropsychiatric illness in maple syrup urine disease. *J Clin Invest*. 2013;123(4):1809–1820.

Nia S. Psychiatric signs and symptoms in treatable inborn errors of metabolism. *J Neurol*. 2014;261:S559–S568.

Pavlovic D, Pekic S, Stojanovic M, Popovic V. Traumatic brain injury: neuropathological, *Neurocog Neurobehav.l Sequelae/Pituit*. 2019;23(3)270–282.

Piasecka M, Papakokkinou E, Valassi E, et al. Psychiatric and neurocognitive consequences of endogenous hypercortisolism. *Intern Med*. 2020;288(2):168–182.

Russo S, Boon JC, Kema IP, Willemse PH, den Boer JA. Patients with carcinoid syndrome exhibit symptoms of aggressive impulse dysregulation. *Psychosom Med*. 2004;66:422.

Sardar S, Habib M, Sukik A, et al. Myxedema Psychosis: Neuropsychiatric Manifestations and Rhabdomyloysis Unmasking Hypothyroidism. *Case Rep Psychiatry*. 2020;2020:7801953.

Sedel F, Baumann N, Turpin J, Lyon-Caen O, Saudubray J. Psychiatric manifestations revealing inborn errors of metabolism in adolescents and adults. *J Inherit Metab Dis*. 2007;30:631.

Seniow J, Bak T, Gajda J, Poniatowska R, Czlonkowska A. Cognitive functioning in neurologically symptomatic and asymptomatic forms of Wilson’s disease. *Mov Disord*. 2002;17:1077.

Sheth S, Brittenham GM. Genetic disorders affecting proteins of iron metabolism: Clinical implications. *Annu Rev Med*. 2000;51:443.

Sonino N, Fallo F, Fava GA. Psychological aspects of primary aldosteronism. *Psychother Psychosom*. 2006;75:327.

Swift RG, Perkins DO, Chase CL, Sadler DB, Swift M. Psychiatric disorders in 36 families with Wolfram syndrome. *Am J Psychiatry*. 1991;148:775.

Velasco PJ, Manshadi M, Breen K, Lippmann S. Psychiatric aspects of parathyroid disease. *Psychosomatics*. 1999;40:486.

Vincent A, Lang B, Kleopa KA. Autoimmune channelopathies and related neurological disorders. *Neuron*. 2006;52:123.

Walterfang M, Fietz M, Fahey M, Sullivan D, Leane P. The neuropsychiatry of Niemann-Pick type C disease in adulthood. *J Neuropsychiatry Clin Neurosci*. 2006;18:158.

Walterfang M, Velakoulis D. Storage disorders and psychosis. In: Sachdev P, Keshavan M, eds: *Secondary Schizophrenia*. Cambridge University Press; 2008.

Walterfang M, Bonnot O, Mocellin R, Velakoulis D. The neuropsychiatry of inborn errors of metabolism. *J Inher Met Dis*. 2013;36:687–702.

Walterfang M, Yucel M, Walker R, Evans A, Bader B. Adolescent obsessive compulsive disorder heralding chorea-acanthocytosis. *Mov Disord*. 2008;23(3):422–425.

2.15 Catatonia

JOSHUA J. RODGERS, M.D., AND THANH THUY TRUONG, M.D.

Catatonia is a peculiar, fascinating, and important neuropsychiatric syndrome defined clinically by abnormalities in the initiation and/or termination of motoric and behavioral functions, but cognitive, emotional, and autonomic dysregulation can also be involved. Catalepsy (maintenance of fixed postures) and waxy flexibility are easily recognizable classic signs, but withdrawal, mutism, and negativism are more common features and may go overlooked. Catatonic features may be insidious or acute, transient or chronic, and persist for days to years. The seemingly contradictory features of stuporous (retarded) or hyperactive types, as well as the typically robust but paradoxical response to benzodiazepines, may be understood from a perspective of dysregulation of cortical–subcortical (direct and indirect pathways) circuits and cortical–cortical network functions. Neuroleptic malignant syndrome (NMS) can be understood as drug-induced catatonia, and there is also a syndromic and mechanistic overlap with akinetic Parkinsonism. These later syndromes reflect "bottom-up" subcortical–cortical disordered dopamine modulation of the involved circuitry, while catatonia reflects "top-down" cortical-subcortical disordered GABA modulation of the circuitry. Catatonia is nonspecific and may be associated with any psychiatric, neurologic, or general medical disorder that may cause dysfunction in these neural pathways. Catatonia is highly prevalent but often unrecognized and untreated, which can have fatal consequences. Fortunately, the gold standard intervention of electroconvulsive therapy (ECT) is highly effective, but unfortunately, unjustified legal regulations in many U.S. States make ECT very difficult to administer in the most vulnerable catatonic patients. Understanding the syndrome of catatonia as a deficit of "executive volition" and its underlying mechanisms may help elucidate the underlying mechanisms of such essentially human constructs of conscious awareness and free will.

BRIEF HISTORY AND NOSOLOGY

Catatonia has been interpreted through many lenses over the past century and a half, depending on the prevailing model of psychiatry at the time; in a way, the syndrome is a projective test for nosologic conflicts within the fields of psychiatry and neurology. Kahlbaum described mental illnesses as "state-course entities" and, in 1874, provided the first systematic description of catatonia as "tension insanity." He described motoric, behavioral, and affective impairments that had a relapsing–remitting course with a good prognosis. He also documented the syndrome’s association with mood disorders and organic (medical and neurologic) disorders. In 1886, Schüle differentiated “true” catatonia from catatonia secondary to organic causes. In 1899, Kraepelin included catatonia in “dementia praecox” and emphasized negativism/positivism signs over motor features. Unlike Kahlbaum, Kraepelin documented a chronic course with a poor prognosis. In the 1900s, Wernicke and other behavioral neurologists understood catatonia through a neurologic dysfunction model and focused on motor signs.

However, around this time, Bleuler, based on a psychoanalytic model, accepted only extreme subconscious ambivalence (conflicts) as the basis for catatonic features and largely rejected organic etiologies or pathophysiologic explanations. Bleuler considered catatonic features to be accessory symptoms of schizophrenia (“splitting of the mind”), his coined term and reconceptualization of dementia praecox. This view persisted for decades, even into the DSM-IV, which continued to describe catatonia as a subtype of schizophrenia, though also recognizing it as a specifier of mood disorders or as secondary to a medical disorder.

The use of sodium amobarbital for catatonia in the 1930s heralded the era of psychopharmacology, and together with a general failure to examine or recognize the motor signs of catatonia, led Mahendra and many in the field to believe that the syndrome of catatonia had disappeared.

Fortunately, in the 1970s and 1980s, several clinician-researchers, including Morrison, Abrams, and Taylor, maintained focus on catatonia and noted its common association with mood disorders rather than only schizophrenia. In 1976, Gelenberg revived the recognition of neurologic, general medical, and psychiatric etiologies to the syndrome, and in 1991, Taylor and Fink summarized the evidence supporting that the classification of catatonia should reflect its association with many illnesses and later continued to argue that it warrants recognition as an independent syndrome. Bräunig and Krüger reconciled the various views of catatonia by regarding it “… as part of a motor-behavior continuum … an independent dimension reflecting a disturbance of executive function that cuts across diagnostic categories.” We suggest this dimension be called “executive volition” as it is the executive control of desired initiation and termination of motor, cognitive, and emotional functions which is disturbed in catatonia. The latest edition of the DSM (DSM-5-TR, released in 2022) defines catatonia as a distinct syndrome that can be a specifier of psychiatric disorders, secondary to a medical condition, or as “unspecified catatonia.”

EPIDEMIOLOGY AND ETIOLOGY

Catatonia is common and underdiagnosed. Reports on the incidence of catatonia vary widely depending on the population studied and the methodologies used. Incidence studies have found catatonia in 7% to 31% of psychiatric inpatients and 1.6% to 5.5% of psychiatric consultations in general medical settings. A research team using consistent standard screening instruments diagnosed catatonia nine times more than the admitting team in a psychiatric hospital.

Table 2.15–1 provides a partial list of the possible etiologies of catatonia. The underlying diagnosis in catatonic patients may be a psychotic disorder about 15% of the time (range 4% to 67%), a mood disorder about 30% of the time, with a bipolar mixed manic state constituting 14% to 71% of cases. Other associated psychiatric disorders include functional neurologic disorder, obsessive-compulsive disorder, personality disorder, and dissociative disorders. Neurologic and medical disorders account for 4% to 46% of cases of catatonia. Medications that increase the risk for drug-induced catatonia include typical and atypical neuroleptics and vesicular monoamine transporter 2 inhibitors (i.e., tetrabenazine). Withdrawal from substances such as benzodiazepines/sedative–hypnotics and dopaminergic agents (i.e., levodopa) can induce catatonic symptoms to emerge. Other agents associated with catatonia from use or withdrawal include alcohol, antidepressants, lithium, antiseizure medications (carbamazepine, primidone), aspirin, disulfiram, metoclopramide, morphine, steroids, hallucinogens, amphetamines, methylphenidate, and cocaine. Numerous medical conditions have been associated with catatonia (Table 2.15–1). Given the high frequency of catatonic features and their prognostic and treatment implications (see below), we suggest that clinicians look for and freely call out and note catatonic features when observed as they would for atypical or psychotic features of a mood disorder, for example.

DIAGNOSIS AND FEATURES

None of the commonly used diagnostic schemas identify a cardinal feature of catatonia a notable absence that begs remedy. As noted above, catatonia is recognized as a cluster of objective signs (note the absence of symptoms in the diagnostic schemas, unique in psychiatric disorders) signifying dysfunction of what we suggest should be termed executive volition in motoric/behavioral, cognitive, and affective domains. Executive volition is a broadly distributed function of cortical–cortical, largely frontal lobe, and cortical–subcortical circuits. Dysfunction of these circuits is thought to result in the features of catatonia (discussed further below). Table 2.15–2 describes the subtypes of catatonia: retarded, excited, malignant, and periodic catatonia. Each may involve a disruption in specific circuits.

Diagnosis

DSM-5 criteria for the diagnosis of catatonia as a specifier of another mental disorder requires the presence of 3 or more of 12 signs: stupor, catalepsy, waxy flexibility, mutism, negativism, posturing, mannerisms, stereotype, agitation, grimacing, echolalia, and echopraxia. The same criteria are used for diagnosing catatonia due to another medical condition as long as there is sufficient evidence that the features are the direct physiologic consequence of a medical condition and not more likely due to a psychiatric disorder or delirium.

Multiple instruments and scales have been developed to assist with screening for catatonia and rating its severity. The most widely used clinically and in research is the Bush–Francis catatonia screening instrument (BFCSI) and rating scale (BFCRS) (Table 2.15–3). The BFCSI (14 items) and BFCRS (23 items) are accompanied by a standardized examination procedure to increase interrater validity and reliability. Screening can also be conducted using the DSM-5 criteria, which does not assess severity. Other scales include the Rogers Catatonia Rating Scale, designed to distinguish catatonia from depression and Parkinson disease; the Northoff Catatonia Rating Scale, which includes motor, behavioral, and affective domains; and the Braunig Catatonia Scale.

Table 2.15–1.
Potential Causes of Catatonia

Psychiatric Conditions	Psychotic disorder about 15% of the time (range 4–67% depending on the population), mood disorder ~30% of the time, with a bipolar mixed manic state constituting 14–71%. Other associated psychiatric disorders include functional neurologic disorder, obsessive-compulsive disorder, personality disorder, and dissociative disorders
Medications/substances	Neuroleptics and vesicular monoamine transporter 2 inhibitors (i.e., tetrabenazine), benzodiazepines/sedative–hypnotics and dopaminergic agents, antibiotics (macrolides and fluoroquinolones) phencyclidine, ecstasy, synthetic cannabinoids
Poisoning	Coal gas, organic fluorides, tetraethyl lead poisoning
Neuromedical	Aneurysms, arteriovenous malformations, cerebral infarcts (temporal, bilateral, parietal, hemorrhagic), thrombosis, hemorrhage (subdural hematoma, subarachnoid hemorrhage, third ventricular hemorrhage), multiple sclerosis
Other central nervous system causes	Encephalitis, traumatic brain injury, alcohol degeneration Wernicke encephalopathy, epilepsy (nonconvulsive status epilepticus), tuberous sclerosis, frontal lobe atrophy, cerebellar degeneration, anoxia, hydrocephalus, basal ganglia lesions, narcolepsy, central nervous system lupus
Metabolic/endocrine	Diabetic ketoacidosis, acute intermittent porphyria, Addison disease, Cushing syndrome, pellagra, hepatic dysfunction, hyper and hypoparathyroidism, idiopathic hyperadrenergic state, thyrotoxicosis, pheochromocytoma
Neoplasm	Frontal lobe tumors, paraneoplastic encephalitis,

	angioma, glioma, Langerhans cell histiocytosis, periventricular diffuse pinealoma
Infectious	Acquired immunodeficiency syndrome (AIDS), bacterial meningoencephalitis, viral encephalitis (i.e., herpes), sepsis, malaria, mono, tertiary syphilis, tuberculosis, typhoid, subacute sclerosing panencephalitis

Table 2.15–2.
Subtypes of Catatonia

Subtype	Symptoms
Stuporous	Mutism, withdrawal, negativism, inhibited movement, posturing, rigidity, and staring
Excited	Excessive and purposeless motor activity, restlessness, stereotypy, disorganized speech, disorientation, impulsivity, agitation, and combativeness
Malignant	Fever, autonomic instability (labile or elevated blood pressure, tachycardia, tachypnea, and diaphoresis), delirium, and rigidity. Mortality 75–100% before 1960 (before ECT); now ~10%
Periodic catatonia	Follows cyclical relapsing and remitting course with psychomotor symptoms, mixed stupor, and excitement. Autosomal dominant inheritance with variable penetrance; disease locus has been linked to chromosome 15q15. Chromosome 15 polymorphisms have also been associated with schizophrenia and autism.

Table 2.15–3.
Description of Signs of Catatonia

Category	Sign Name	Description
Motor	Posturing°*+	Spontaneous maintenance of posture(s) for long periods
Motor	Catalepsy°*+	Passive induction (from an external person or object; >1 minute) of a posture with persistence/inability to return to initial posture
Motor	Stereotypy°*+	Inherently normal but repetitive (>3) and unchanged, non–goal-directed motor activity
Motor	Festination+	Paroxysms of involuntary, jerky, uncoordinated, inappropriate, hasty movements
Motor	Dyskinesias+	Abnormal involuntary fast movements
Motor	Athetosis+	Choreatic-like movements with a screw-shaped character
Motor	Mannerisms°*+	Inherently abnormal, odd, bizarre execution of purposeful movements with a disturbance in the harmony of movements

Motor	Rigidity°+	Maintenance of a rigid position despite efforts to be moved
Motor	Muscular hypotonus+	Slack and loose active movement; decreased muscle tone in passive movements.
Motor	Sudden muscular tone alterations+	Rapid switches between muscular normotonus, hypotonus, and hypertonus, with or without apparent external trigger
Motor	Akinesia+	Complete absence or paucity of movements (for at least 30 minutes)
Motor	Waxy flexibility°* (Flexibilitas cerea+)	Initial resistance before allowing repositioning (like bending a candle)
Motor	Grasp reflex°	Reflexive bending of fingers in response to palmer tactile stimulation
Motor/Behavior	Mitgehen°+/Mitmachen+	Arm raising in response to light pressure despite contrary instruction (tone induced by passive manipulation going with the examiner)
Motor/Behavior	Gegenhalten°+	Resistance to passive movement

		proportional to stimulus strength (tone induced by passive manipulation going against the examiner)
Behavior	Negativism°* Negativismus+	Resistance to instructions or attempts to move/examine patient. Contrary behavior does the opposite of the instruction
Behavior	Automatic obedience°+	Exaggerated cooperation with the examiner’s request
Behavior	Mutism°*+	Persistent verbal unresponsiveness or minimal responsiveness (for at least 30 minutes+)
Behavior	Grimacing°*+	Maintenance of odd facial expressions
Behavior	Echopraxia°*+/echolalia°*+	Mimicking an examiner’s movements/speech
Behavior	Verbigeration°+	Repetition of phrases or sentences (like a scratched record)
Behavior	Perseveration°+	Repeatedly returns to the same topic or movement; Non–goal-directed repetition of thoughts and/or actions

Behavior	Abnormal speech[+]	Qualitative abnormalities in volume and intonation of speech
Behavior	Withdrawal° Autism/Withdrawal[+]	Refusal to eat, drink, or make eye contact Isolates from or avoids social contacts
Behavior	Immobility/Stupor°[*+]	Persistent extreme hypoactivity, immobility, minimal response to stimuli (for at least 30 minutes[+])
Behavior/Affective	Loss of initiative[+]	Subjective/ objective (expressed or observed) loss of initiative concerning things they usually do/daily Routine/relation to the environment and/or other persons.
Behavior/Affective	Staring°[+]	Fixed gaze, decreased/no scanning of the environment, or blinking (at least 20 seconds[+])
Behavior/Affective	Excitement° (agitation[*+])	Uncontrollable extreme hyperactivity, constant motor unrest (nonpurposeful, not goal-directed, not akathisia; may be related to intense

		emotional experiences+)
Behavior/Affective	Impulsivity°+	Sudden inappropriate behaviors (running, taking off clothes)
Behavior/Affective	Compulsive behavior or speech+	Patient feels driven to perform repetitive patterns (i.e., >5 times) of behavior or speech that they cannot control or relate to themselves
Behavior/Affective	Ambitendency°/Ambivalence+	The appearance of being “stuck” in indecisive, hesitant movement; exhibits conflicting emotions and/or thoughts so that they appear “stuck,” indecisive, and hesitant+
Behavior/Affective	Combativeness°/Aggression+	Verbally or physically aggressive in an undirected manner—unable to explain, but apparently induced by internal emotional states or external events/triggers
Affective/Behavior	Affective related behavior+	Abnormal behavioral reactions apparently related to particular emotional states and/or discharges
Affective	Compulsive emotions+	Uncontrollable and ego-dystonic abnormal

		affective reactions
Affective	Affective/emotional lability+	Unstable and quickly alternating affective reactions
Affective	Flat affect+	Considerable reduction in quantity and quality of emotional reactivity (i.e., flat facial expression, aprosodic speech)
Affective	Affective latency+	Delayed emotional initiation or reactivity to external stimulus
Affective	Anxiety+	Signs of intense, uncontrollable anxiety (i.e., facial expression, verbal, and/or vegetative (i.e., perspiration)
Autonomic	Dysautonomia° Vegetative abnormalities+	Abnormal temperature, blood pressure, heart rate, respiratory rate, diaphoresis

Signs marked * are included in the DSM-5 diagnostic schema, which requires three or more signs to diagnose Catatonia. See DSM-5 for details. American Psychiatric Association. DSM-5 Task Force. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Signs marked ° are included in the Bush-Frances Catatonia Rating Scale.

Signs marked + are included in the Northoff Catatonia Rating Scale.

Differential Diagnosis

Catatonic features are nonspecific. When a patient presents with catatonic features, the clinician must have a broad differential in mind. There is substantial symptom overlap between several disorders, including serotonin syndrome, NMS, akinetic mutism, and disorders of consciousness (DOC). Distinguishing between these presentations is essential to guide the next steps in management (Table 2.15–4). For example, a patient presenting with catatonic features may have been recently prescribed neuroleptics in their history. Likewise, NMS is differentiated from serotonin syndrome by the latter exhibiting hyperreflexia rather than hyporeflexia, myoclonus, and gastrointestinal distress.

Table 2.15–4.
Differential Diagnoses for Patients Presenting with Possible Catatonia

Differential Diagnosis	Features
Neuroleptic malignant syndrome (NMS)	Drug-induced malignant catatonia caused by neuroleptic (antipsychotic) medications; presents after high doses of neuroleptics are given
Malignant hyperthermia	Rare genetic disorder with exposure to inhaled anesthetic (e.g., halothane) or a depolarizing muscle relaxant (e.g., succinylcholine)
Serotonin syndrome	Distinguished by the prominence of hyperreflexia, myoclonus, nausea, vomiting, and diarrhea
Akinetic mutism	Unlike catatonia, there is no posturing, echolalia, or echopraxia, and it is not responsive to benzos. Medial frontal lesions are usually detectable on imaging. Dysfunction is of motivation and not executive volition.
Selective mutism	Patients can be observed to be mute with some and not others; does not exhibit other catatonic features. Presentation is seen in personality disorders and major stressful periods.
Nonconvulsive status epilepticus	EEG would show epileptic activity, whereas there is no specific EEG abnormality in catatonia.
Locked-in syndrome	Presents with immobility and mutism but no other catatonic features; neuroimaging shows brainstem lesions. The patient may make attempts to communicate.
Stiff-person syndrome	Rigidity, dysautonomia, and response to benzodiazepines. However, this is differentiated by a fixed deformity of the spine and the presence of anti-GAD antibodies. There is no mutism, posturing, and other catatonic features.
Parkinson disease	Characteristic stooped posture, not odd. No benzodiazepine response. Presence of resting tremor, postural instability, response to levodopa.

Stroke	Focal neurologic signs, neuroimaging.
Delirium	No posturing, rigidity, or negativism; generally worsens with benzos.
Dementia	No posturing, rigidity, or negativism generally worsens with benzos.

GAD, glutamic acid decarboxylase.

NEUROPATHOPHYSIOLOGY

It is generally accepted that the features of catatonia arise from disruption of the function of distributed cortical–cortical (frontoparietal and frontolimbic) networks and the cortical–subcortical and cerebellar–subcortical circuits that regulate cortical functioning (see Fig. 2.15–1). This disruption can be caused by psychological (functional) factors, neurologic lesions (e.g., ablations in motor systems or motor–perceptual integration systems, or excitatory lesions in the thalamus), neurochemical imbalances (e.g., reduced dopamine or GABA-A activity or increased GABA-B activity due to drugs or disease), metabolic derangements, epilepsy, and other causes.

Disruptions in the motor circuitry may be involved in rigidity, anterior cingulate–medial orbitofrontal circuitry in akinetic mutism, and dorso- and ventromedial prefrontal cortex in perseverative-compulsive behaviors in catatonia. Such dysfunctions may be due to reduced $GABA_A$ inhibition of $GABA_B$ interneurons in the substantia nigra and ventral tegmental area. This reduced GABAergic activity would induce a state of hypodopaminergia in the dorsal and ventral striatum and paralimbic cortex. Another network involves upregulation of *N*-methyl-D-aspartate (NMDA, a glutamate receptor) activity in the dorsal striatum secondary to reduced $GABA_A$ inhibition of frontal corticostriatal tracts (Fig. 2.15–2). Activation of $GABA_A$ through the use of agonists such as lorazepam would cause disinhibition of dopamine activity in the striatum and paralimbic cortex, restoring normal regulation of the functions of these systems. Likewise, glutamate antagonists such as amantadine have been effective in some lorazepam-refractory cases. Serotonin disequilibrium with upregulated serotonin 1A (5-HT1A) and downregulated serotonin 2A (5-HT2A) receptors have also been implicated in catatonia. In NMS and akinetic Parkinsonism, dysfunction in the system is due to “bottom-up” dysregulation of striatal D2 receptor activity by blockade in NMS, usually by neuroleptics with high affinity for this receptor, or by loss of substantia nigra dopaminergic inputs to the striatum in Parkinson disease.

A "gridlock analogy" may be helpful in understanding the pathophysiology of catatonia. Cortical–subcortical (frontal cortex to striatum to globus pallidus to thalamus and back to cortex) circuits forming the direct and indirect pathways serve to regulate cortical function, essentially modulating the flow of information in the system, disinhibiting desired functions (direct pathway) and inhibiting undesired function (indirect pathway), much the same way a traffic light system regulates the flow of traffic through a city street grid. Continuing with this analogy, if all the traffic lights are green, then all traffic is going at the same time resulting in gridlock. Proper timing and balance of green lights (disinhibition) and red lights (inhibition) are necessary for normal traffic flow. Restoring inhibition by giving GABA agonists in catatonia can thus free up the gridlock from a relative deficit of inhibitory modulation.

Neuroimaging Findings in Catatonia

Computed tomography (CT) studies show cortical enlargement (in catatonic schizophrenia) and cerebellar atrophy in catatonic patients relative to controls. Cerebral blood flow studies have shown imbalanced perfusion with hyperperfusion of the left basal ganglia and hypoperfusion in the left medial temporal region and right parietal cortex, which may normalize following ECT. Functional magnetic resonance imaging (fMRI) has shown reduced right motor cortex activation related to motor signs and reduced orbitofrontal cortical activity related to behavioral and affective symptoms. Single-photon emission computerized tomography (SPECT) studies have shown lower GABA-A binding and imbalances in the left sensorimotor cortex and right lateral orbitofrontal, and right posterior parietal cortex in patients with catatonia.

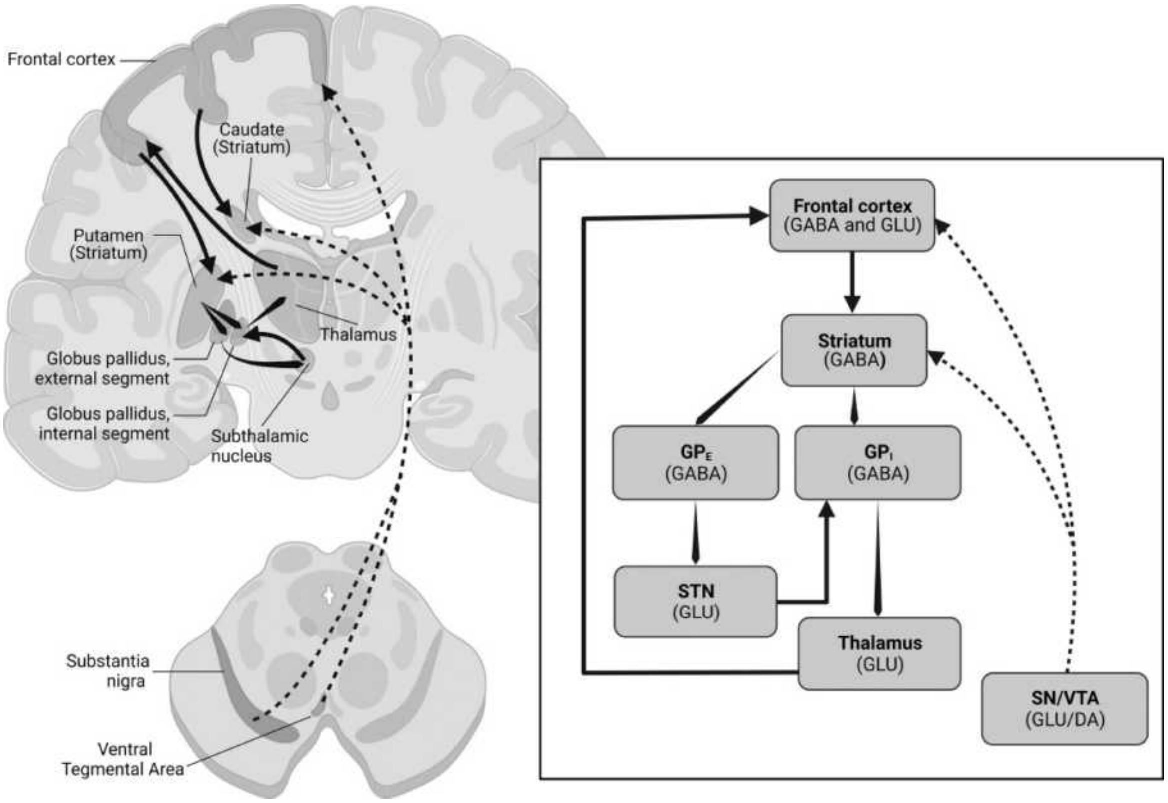
Frontal cortex
Caudate (Striatum)
Putamen (Striatum)
Thalamus
Globus pallidus, external segment
Globus pallidus, internal segment
Subthalamic nucleus
Substantia nigra
Ventral Tegmental Area
Frontal cortex (GABA and GLU)
Striatum (GABA)
GP$_E$ (GABA)
GP$_I$ (GABA)
STN (GLU)
Thalamus (GLU)
SN/VTA (GLU/DA)

FIGURE 2.15–1. Typical cortical–subcortical circuits, including direct and indirect pathways. DA, dopamine; GLU, glutamate; GP_E, globus pallidus externa; GP_I, globus pallidus interna; STN, subthalamic nucleus; SN, substantia nigra; VTA, ventral tegmental area. *Solid arrow*: glutaminergic (excitatory); *tapered arrow*: GABAergic (inhibitory); *dashed arrow*: dopaminergic (modulatory) (Note: arrow paths denote circuit start and end points and not anatomically accurate pathways). (Courtesy of Robert Boland MD. Created with BioRender.com.)

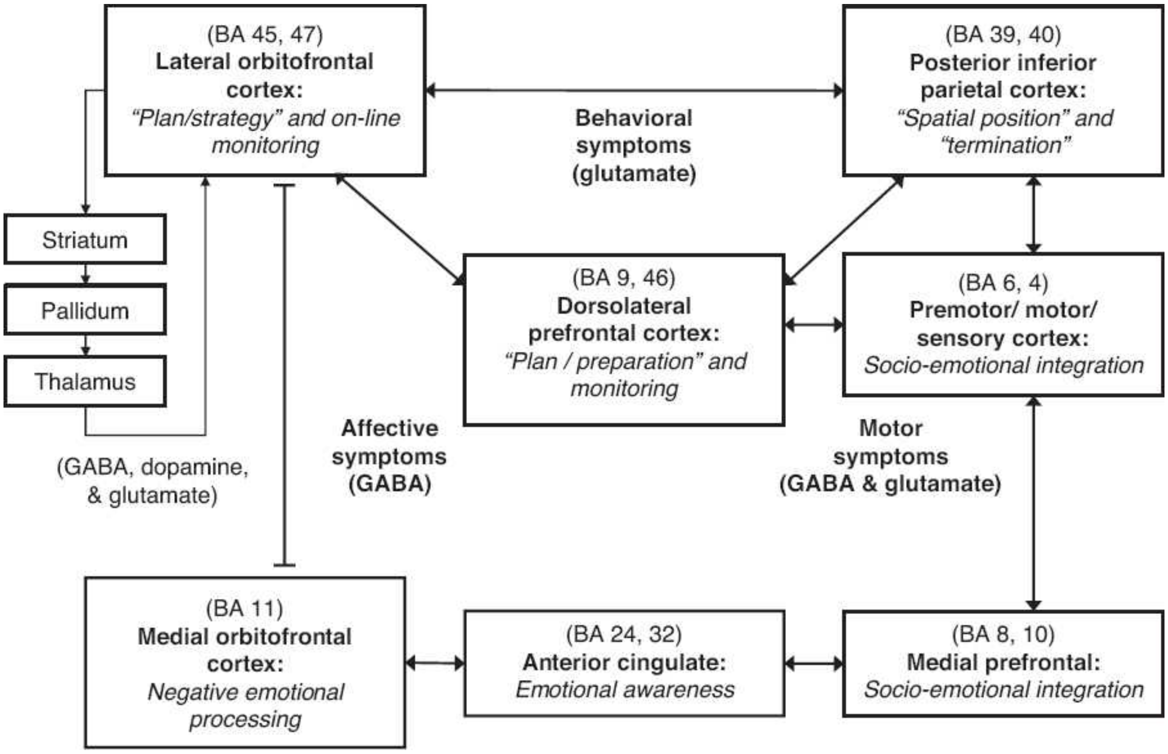
(BA 45, 47)
Lateral orbitofrontal cortex:
"Plan/strategy" and on-line monitoring
(BA 39, 40)
Posterior inferior parietal cortex:
"Spatial position" and "termination"
Behavioral symptoms (glutamate)
Striatum
Pallidum
Thalamus
(BA 9, 46)
Dorsolateral prefrontal cortex:
"Plan / preparation" and monitoring
(BA 6, 4)
Premotor/ motor/ sensory cortex:
Socio-emotional integration
(GABA, dopamine, & glutamate)
Affective symptoms (GABA)
Motor symptoms (GABA & glutamate)
(BA 11)
Medial orbitofrontal cortex:
Negative emotional processing
(BA 24, 32)
Anterior cingulate:
Emotional awareness
(BA 8, 10)
Medial prefrontal:
Socio-emotional integration

FIGURE 2.15–2. Proposed catatonia model. A simplified cortical–subcortical circuit is shown for the regulation of one cortical region for simplicity. Each cortical region has subcortical circuit regulation. BA, Broca area. (Adapted from Northoff G. What catatonia can tell us about "top-down modulation": A neuropsychiatric hypothesis. *Behav Brain Sci*. 2002;25(5):555–604, with permission of Cambridge University Press. doi:10.1017/S0140525X02000109.)

MANAGEMENT AND TREATMENT

Catatonia can be a life-threatening illness with significant morbidity and mortality. Thus, the clinician would be wise to "treat first, ask questions later" with an emphasis on supportive interventions. The presence of dysautonomia denotes "malignant catatonia," with a 75% to 100% mortality rate. Now with ECT and benzodiazepines, the mortality rate has decreased to 10%. If an etiology is found, then treatment for the underlying condition follows the initial interventions for catatonic symptoms. If dopamine antagonists (i.e., antipsychotics, metoclopramide) or dopamine-depleting agents (i.e., tetrabenazine) are suspected, they should be discontinued. If symptoms are suspected to be from the withdrawal of dopamine agonists or GABAergic agents, they should be resumed.

Supportive interventions should be provided for consequences of catatonia: hyperthermia, dehydration, malnourishment, bed sores, flexion contractures, rhabdomyolysis (acidosis, renal failure), deep venous thrombi, pulmonary emboli, urinary retention, infection, and aspiration pneumonia. Management may involve attentive nursing care, intravenous (IV) fluids, anticoagulation, urinary catheterization, nasogastric tube/percutaneous endoscopic gastrostomy (PEG) tube feeds, oral care, suction, cooling blanket, and autonomic control.

Lorazepam is the first-line treatment regardless of etiology and carries an efficacy of 70% to 80% for a reduction in symptoms. IV lorazepam is preferred due to its quick onset of action and ease of administration. Intramuscular (IM) and sublingual administration may be used if IV access is not available since patients may not be able to swallow the oral formulation. Clinicians should be aware of IM administration potentially causing pain and increased fear in patients with catatonia. An initial lorazepam dose of 2 mg (lorazepam challenge) is used for both diagnostic and treatment purposes. Lower doses of lorazepam may be used in pediatric and frail older adult patients, but 2 mg is recommended for most adults to avoid an equivocal response. A robust “Lazarus” response may be seen within 30 minutes of administration, though some patients may require up to 120 minutes to see effects. A positive response (increased interaction, reduction of catatonic features) is confirmatory of catatonia, whereas a negative response (sedation) to the initial dose does not rule out catatonia. If the response is equivocal, that is, there is no change, the patient is neither sedated nor more interactive, and clinical suspicion remains high, another dose of lorazepam should be given. In healthy individuals, 2 mg of lorazepam should be quite sedating; no response may be a significant observation. Multiple doses may be required for a response. Once a positive response to lorazepam has been established, lorazepam should be scheduled (i.e., 2 mg q6h) at least for the first 24 to 48 hours. Some patients need a lorazepam drip (1 to 2 mg q4h), titrated to a total daily dose that reduces symptoms without excessive sedation. Some patients may require up to a total daily dose of 30 mg or more. It is important for nursing staff to administer medications even when the patient appears to be sleeping because the “sleep” may represent stupor and indicate a recurrence of symptoms from lorazepam wearing off. Once there is a significant improvement in symptoms, a slow taper of benzodiazepines may be considered (10% to 25% decrease in a dose per day). Lorazepam treatment should be considered a temporary bridge to definitive ECT in many cases.

For all patients with catatonia, ECT should be considered and arranged if there is high clinical suspicion and/or if there is a positive response to lorazepam. This foresight prevents delays in treatment if the patient does not adequately respond to lorazepam or progresses into malignant catatonia. ECT is the gold standard treatment with a 90% efficacy in catatonia. There are no absolute contraindications for ECT, and it is well tolerated in even medically ill patients. However, patients with recent myocardial infarction, stroke, and other cardiovascular conditions are at increased risk of complications from ECT. Appropriate consultations (i.e., cardiologist) prior to ECT can mitigate risks. Altogether, ECT is the most effective treatment for catatonia, and risks should be weighed against potential benefits. Most patients improve within the first two to three treatments. More treatments are often recommended to prevent a recurrence, with some needing monthly maintenance ECT to prevent relapse into catatonia. The requirements for informed consent for ECT vary in each state. In some states, patients must voluntarily consent to ECT, and the use of surrogate decision-makers is not permitted, which limits its use in nonresponsive catatonic patients. In such situations, if the patient becomes alert with lorazepam, discussion of ECT during these periods may allow for it as an option.

Alternative treatments for catatonia that have shown efficacy include hypnotics such as zolpidem 5 to 10 mg, antiepileptics that increase GABA (topiramate, valproate, carbamazepine, zonisamide), dopamine agonists (bromocriptine), and NMDAR antagonists (amantadine, memantine). Amantadine (100 to 600 mg daily) and memantine (10 to 20 mg daily) are used as adjunctive treatments to lorazepam rather than monotherapy. Some case reports have described the success of targeting the dorsolateral prefrontal cortex (DLPFC) using transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (TDCS), though this effect appears to be inconsistent. Other treatments that have had mixed success include muscle relaxants, calcium channel blockers, anticholinergics, lithium, thyroid meds, stimulants, and corticosteroids. Antipsychotics were frequently used to treat catatonia in the past, but in many cases, this worsened symptoms and increased mortality. Therefore, for the treatment of underlying psychosis, the use of antipsychotics should be accompanied by lorazepam, and second-generation antipsychotics, especially agents with partial D2 agonist activity (i.e., aripiprazole), are preferred.

Workup of Underlying Causes of Catatonia

The initial evaluation and treatment should occur at the same time or prior to the workup of underlying medical or psychiatric causes. Catatonia is not associated with the specific laboratory, imaging, or EEG findings. Initial labs, useful for looking for underlying causes or monitoring for complications of catatonia, can include complete blood count (CBC) with differential, complete metabolic panel (CMP), serum iron, serum creatine kinase (CK) (Table 2.15–5 for a list of workup considerations). Serum CK is especially elevated in malignant catatonia and NMS and can increase the risk of renal failure. Serum CK should be tracked to ensure a downtrend. Neuroimaging, such as MRI or computed tomography (CT) brain can be obtained to detect lesions. Electroencephalography (EEG) showing diffuse background slowing may indicate delirium. If there are signs of encephalitis, a lumbar puncture is indicated. The cerebrospinal fluid (CSF) may be sent for an autoimmune encephalitis panel. Other workup is guided by clinical suspicions of underlying conditions. If testing results are unremarkable, the patient is likely to have an underlying psychiatric cause of symptoms.

Catatonia may masquerade as DOC as there is considerable overlap in symptoms (mutism, staring, etc.). Thus, the diagnosis of catatonia may be missed in settings where the focus is on DOC (minimally responsive, comatose). Whyte et al. published a placebo-controlled, double-blind, single-dose, crossover study with 84 patients who had DOC based on observations that some vegetative and minimally conscious patients paradoxically experienced temporary recovery of consciousness when given zolpidem. Their objective was to determine the frequency and predictability of this response. While 4.8% of the participants responded to zolpidem, the responders could not be distinguished from the nonresponders. Catatonia was not evaluated in these patients. Based on the response to zolpidem, some patients with presumed DOC may actually have catatonia. In a pilot study, DOC patients in a neurorehabilitation hospital were assessed with the BFCRS and treated with zolpidem or other GABA-A agonist (Rodgers and Arciniegas, 2015). Patients who responded were noted to exhibit posturing/catalepsy, grimacing, rigidity, and dysautonomia. Nonresponders did not exhibit these signs but did exhibit other catatonic features. Subcortical networks may have been too severely damaged (e.g., thalamic stroke) to allow for a therapeutic target in nonresponders.

Table 2.15–5.
Workup Considerations for Catatonia

Test	Comments
CBC with differential	Look for leukocytosis
CMP	Look for electrolyte abnormalities, changes in nitrogen balance
Serum CK	Elevated in malignant catatonia and NMS
Serum iron	May be low in NMS and catatonia
MRI brain	Look for acute and chronic lesions, usually correlated with a focal deficit
EEG	May be normal in psychiatric catatonia but abnormal in other causes (i.e., diffuse slowing in encephalitis, nonconvulsive status epilepticus)
CSF	Indicated if there are signs of encephalitis; CSF may be tested for autoimmune encephalitis

CBC, complete blood count; CK, creatine kinase; CMP, complete metabolic panel; CSF, cerebrospinal fluid; NMS, neuroleptic malignant syndrome.

SUMMARY AND FUTURE DIRECTIONS

Catatonia is a common clinical syndrome diagnosed on the basis of observed and elicited motoric and behavioral signs. There is no established cardinal feature, but dysfunction of executive volition in motoric, behavioral, and affective domains is a central theme. Catatonic features can result from any disruption in the function of cortical–subcortical (basal ganglia–thalamocortical; direct and indirect pathway) circuits regulating cortical–cortical networks and lead to neurotransmitter abnormalities: reduced dopaminergic and GABAergic activity and upregulation of NMDA activity. Etiologies are widespread and include acute and chronic medical conditions, psychiatric conditions, medications, substance use, and genetic predisposition. Benzodiazepines and ECT are the mainstays of treatment. There is a considerable overlap of features between catatonia and DOC, which is a diagnostic dilemma. However, the presence of catatonic features in DOC may predict response to GABA agonists such as zolpidem and benzodiazepines.

Psychotherapeutic Interventions for Catatonia

Our understanding of catatonia has largely focused on clinician observations rather than the subjective experience of the patient. Limited case literature examining the psychological state of catatonia describes themes of death, fear, and overwhelming defeat. Northoff's Catatonia Rating Scale includes affective symptoms, including emotional lability, anxiety, ambivalence, and other symptoms. Fink and Shorter posited that persisting fear may play a role in sustaining catatonia, where defensive mechanisms of withdrawal and denial are activated (Fink and Shorter, 2017). Like any psychiatric condition, there is not a clear boundary between the biologic and psychological causes of catatonia. In patients with chronic relapsing and remitting catatonia, long-term maintenance treatment with ECT and benzodiazepines is often needed. However, psychotherapeutic interventions to help the person manage stress, tolerate anxiety, manage ambivalence, and regain agency may be part of the treatment (Heckers and Walther, 2021). As far as the authors are aware, there is no literature on psychotherapeutic interventions that could improve or prevent relapse into a catatonic state. Various psychotherapies such as cognitive behavioral therapy, dialectical behavioral therapy, and psychodynamic therapy, all have evidence for enhancing coping strategies and increasing insight, which could theoretically improve the quality of life and potentially reduce the risk of a relapse into catatonia. Further research may elucidate methods for assessment of the subjective experience, which may narrow down the etiology (psychiatric vs. medical).

FURTHER READINGS

American Psychiatric Association. DSM-5 Task Force. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Bush G, Fink M, Petrides G, Dowling F, Francis A. Catatonia I. Rating scale and standardized examination. *Acta Psychiatr Scand*. 1996;93:129–136.

Caroff SN, Mann SC, Francis A, Fricchione GL. *Catatonia: from Psychopathology to Neurobiology*. American Psychiatric Publishing; 2004.

Caroff SN, Mann SC, Francis A, Fricchione GL. Epidemiology. In: Caroff SN, Mann SC, Francis A, Fricchione GL, eds. *Catatonia: from Psychopathology to Neurobiology*. American Psychiatric Publishing; 2004.

Carroll BT, Anfinson TJ, Kennedy JC, Yendrek R, Boutros M, Bilon A. Catatonic disorder due to general medical conditions. *J Neuropsychiatry Clin Neurosci*. 1994;6:122–133.

Carroll BT, Thomas C, Jayanti K, et al. Treating persistent catatonia when benzodiazepines fail: neural circuit changes help explain syndrome's signs, suggest potential therapies. *Curr. Psychiatry*. 2005;4(3):56–64.

Coffey MJ. Catatonia: Treatment and Prognosis. Last updated: 8/15/22. https://www.uptodate.com/contents/catatonia-treatment-and-prognosis

Daniels J. Catatonia: clinical aspects and neurobiological correlates. *J Neuropsychiatry Clin Neurosci*. 2009;21:371–380.

Dhossche DM, Stoppelbein L, Rout UK. Etiopathogenesis of catatonia: generalizations and working hypothesis. *J ECT*. 2010;26(4):253–258.

Fink M, Shorter E. Does persisting fear sustain catatonia? *Acta Psychiatr Scand*. 2017;136(5):441–444. https://doi.org/10.1111/acps.12796

Fink M, Taylor MA. *Catatonia: A Clinician's Guide to Diagnosis and Treatment*. Cambridge University Press; 2003.

Fink M, Taylor MA. Catatonia in psychiatric classification: a home of its own. *Am J Psychiatry*. 2003;160:1233–1241.

Heckers S, Walther S. Caring for the patient with catatonia. *JAMA Psychiatry*. 2021;78(5):560–561. https://doi.org/10.1001/jamapsychiatry.2021.0096

Northoff G. What catatonia can tell us about "top-down modulation": a neuropsychiatric hypothesis. *Behav Brain Sci*. 2002;25:555–577.

Northoff G, Koch A, Wenke J, et al. Catatonia as a psychomotor syndrome: a rating scale and extrapyramidal motor symptoms. *Mov Disord*. 1999;14(3):404–416.

Northoff G, Kötter R, Baumgart F, et al. Orbitofrontal cortical dysfunction in akinetic catatonia: a functional magnetic resonance imaging study during negative emotional stimulation. *Schizophr Bull*. 2004;30(2):405–427.

Philbrick KL, Rummans TA. Malignant catatonia. *J Neuropsychiatry Clin Neurosci*. 1994;6:1–13.

Purves D, Augustine GJ, Fitcpatrick D, et al. *Neuroscience*. 2nd ed. Sinauer Associates; 2001. Projections from the Basal Ganglia to Other Brain Regions. http://www.ncbi.nlm.nih.gov/books/NBK10860/

Rodgers JJ, Arciniegas DB. Is response to zolpidem in disorders of consciousness predicted by the presence of catatonic features? *J Neuropsychiatry Clin Neurosci*. 2015;27(2).

Saddawi-Konefka D, Berg SM, Nejad SH, Bittner EA. Catatonia in the ICU: an important and underdiagnosed cause of altered mental status. a case series and review of the literature. *Crit Care Med*. 2014;42(3):e234–e241.

Smith JH, Smith VD, Philbrick KL, Kumar N. Catatonic disorder due to a general medical or psychiatric condition. *J Neuropsychiatry Clin Neurosci*. 2012;24:198–207.

Stern TA, Fava M, Wilens TE, Rosenbaum JF. *Massachusetts General Hospital Comprehensive Clinical Psychiatry*. 2nd ed. Elsevier Inc; 2016.

van der Heijden FMMA, Tuinier S, Arts NJM, Hoogendoorn MLC, Kahn RS, Verhoeven WMA. Catatonia: disappeared or under-diagnosed? *Psychopathology*. 2005;38:3–8.

Whyte J, Rajan R, Rosenbaum A, et al. Zolpidem and restoration of consciousness. *Am J Phys Med Rehabil*. 2014;93(2):101–113.

3

Contributions of the Psychological Sciences

3.1 Contributions of the Psychological Sciences: Sensation, Perception, and Cognition

Aldrich Chan, Psy.D., Louis J. Cozolino, Ph.D., and Daniel J. Siegel, M.D.

The brain is a social organ of adaptation, built and maintained through the interaction of social, psychological, and biologic forces. Because of the vast complexity of this synergistic process, the study of sensation, perception, and cognition is necessarily broad and far-reaching. The rapid increase in knowledge of neuroscience and the development of new hypothetical models through which one can understand the brain's functioning make this an exciting time in the cognitive sciences.

Mental processes such as thinking, feeling, and remembering are not synonymous with brain activity, but they may be dependent upon them. At a minimum, the subjective, inner felt experience of these fundamental mental processes, the feeling of being alive and awake, cannot be reduced to neural firing, even if subjective experience arises from that brain activity. Further, the mind may be more than just "an output" of what the brain does, with the possibility that it is an emergent, self-organizing process that is both fully embodied and embedded in one's relationships.

In this way, the system of mind may be characterized by patterns of energy with a symbolic value called “information.” Energy and information flow happens both within humans and between humans. Seen this way, the mind is an emergent aspect of a complex system arising from energy and information flow, and as a self-organizing emergent process, it may regulate that flow. This regulation is reflected in mental processes, beginning with sensation and perception, to cognition, then into the initiation of behavior. Nevertheless, science is clear: Mental processes are profoundly shaped by neural activity, even if they are not limited to what goes on in the head. For this reason, understanding neural processing is an important part of comprehending information processing, one of many important aspects of the mind.

The terms s*ensation, perception,* and *cognition* are used to describe the three broadening tiers of human information processing. Think of s*ensation* as the immediate result of the stimulation of sensory neurons and *perception* as involving the organization and conscious awareness of these sensations by distinct regions of the brain. *Cognition* refers to the set of interwoven processes, including memory, language, and problem-solving, which both comprise and organize our conscious processing. Although distinguishing among sensation, perception, and cognition has a long academic history, the ability to separate them anatomically and functionally in the brain grows increasingly difficult as more is learned about their interdependence in the dynamic functioning nervous system.

COGNITIVE SCIENCE

The fields relevant to this overview are part of the interdisciplinary studies of cognitive science, which include cognitive and developmental psychology, psycholinguistics, computational science, and the emerging field of interpersonal neurobiology. Biologic, psychodynamic, and social psychiatry find a common home within cognitive science whereby the usual divisions of nature versus nurture and of biology versus psychology disappear on examination of brain development and the origins of mental processes.

In recent years, discoveries in the neurosciences have revealed a wide range of findings relevant to psychiatry. One such discovery showed that the brain’s structure and function are a result of the transaction among genetic, physiologic, and experiential influences. In particular, brain development requires specific forms of experience to foster the growth of neural circuits involved in a wide array of mental processes, including attention, memory, emotion, attachment, and self-reflection. Whereas genes function as a template of information and a mediator of transcription of the proteins that determine the neural structure, experience directly shapes the selection and timing of how gene activity is translated into the structure of the brain; it is now known that experiences influence the epigenetic regulators that modulate which, when, and how genes are expressed in response to experience and thus shape the unfolding of genetically programmed development of the central nervous system (CNS).

The human brain, especially the cerebral cortex, is undeveloped at birth. This functional immaturity requires the use of the caregiver's brain to help the child's brain grow and organize to mature functioning—findings from developmental neuroscience point to the centrality of interpersonal relationships in the development of the brain. The cooperative communication of infant–caregiver attachments is thought to shape neural infrastructure, not only for emotional development but also for abstract reasoning and cognitive abilities. The patterns of interaction between child and caregiver have a direct impact on the development of the child's brain and the functioning of the mind—the emergence of energy and information flow within and between. Consequently, cognitive processes and the internal flow of energy and information are an expression of the genetic, physiologic, social, and experiential factors that shape the development and maintenance of mental function.

Hot and Cold Cognition

Traditionally, cognition was studied by experimental psychologists in university laboratories, whereas emotion was explored clinically by psychoanalysts in consulting rooms. Striving to avoid the subjective and imprecise nature of emotions, cognitive psychologists devised flowcharts and algorithms similar to those used to describe computer programs. Input was calibrated, output was measured, and theories were generated concerning what might be happening within the "black box" of the brain. The increasing complexity of these models did not improve their explanatory power, and questions concerning motivation and emotion invariably arose.

As knowledge of brain functioning increased, it has become more and more evident that neural networks involved in perception and cognition are inextricably interwoven with other networks responsible for processing somatic states, survival value, emotion, meaning, and motivation. The myth of *cold cognition,* or cognitive processes devoid of affective and somatic influence, is gradually fading. Flowcharts depicting linear input–output processes are being replaced with more sophisticated models reflecting the reality of the complex neural systems being discovered. When this change is added to the notion that mental life is also shaped by the context, especially the social context, then it can be seen that the mind emerges from factors within the brain, the body, and the larger interactive environment. This is the “mind in context” that needs to be considered to understand all mental processes.

Some cognition is “hot,” such as recalling traumatic memories or spotting someone that is sexually attractive, and some cognition may be “cool,” such as adding rows of figures or stacking dishes. These same cool tasks, however, if performed before an Internal Revenue Service auditor or in front of a strict teacher, may become warm or even hot. The fundamental principle, therefore, is that sensation, perception, and cognition all occur within the context of feedforward and feedback neural influences interwoven with and guided by complex contextual and emotional determinants.

Emotion

What is an emotion? The answer to this question is as complex as the mind itself. Although the lack of clear definitions and good animal models have hindered empirical research, it is clear that emotions play a central role in many cognitive processes. In fact, a broad view suggests that an emotion is a profoundly integrative process. Emotions connect the body to the brain, bring continuity to states of mind across time, and link one person to another within the “emotional connections” that create the fabric of human social lives.

One view considers emotion as a primary value system of the brain, engaging attentional resources, appraising the significance of an event, and allowing activations to be selectively elaborated and reinforced as they are encoded and stored, such that emotionally charged experiences may be more readily recalled later on than uneventful ones. According to this view, the most fundamental aspect of emotion is the arousal–

appraisal system in which the brain responds to a given stimulus with the signal of “this is important—take note and pay attention now!” Emotion thus gives value to a representation by arousing attentional mechanisms and focusing a spotlight of attention on the stimulus. The second stage would then appraise the meaning of such emotional arousal by assessing its hedonic tone: “Is this good or bad? Should this be approached or avoided?” Emotion thus directs the flow of energy—the activations within specific circuits of the brain—as the arousal–appraisal system focuses cognitive processes on elements of the internal and external environments. The third level of emotional processing is the elaboration of this appraisal into a more specific form called a *categorical emotion,* which includes joy, interest, surprise, fear, anger, sadness, or shame. These categorical emotions have distinct psychophysiologic manifestations and are found across cultures.

The research of Jaak Panksepp revealed seven distinct primary affective systems that converge at a nucleus nested in the midbrain called the periaqueductal gray (PAG). These primary systems govern emotions that induce seeking, lust, care, play, fear, panic, and rage. Higher-order cognitive systems order themselves according to the systems that may be dominant at any time. Feedback and feedforward loops in response to these affective systems may subsequently modify the expression of select systems. This is in contrast to LeDoux and Brown’s perspective, who argue that emotions are, in fact, a consequence of higher-order cognition. This theory relegates subcortical networks to nonconscious inputs that may influence emotional experiences but are not central to their conscious expression. Their proposal is still considered controversial given the current consensus in overall emotional research.

Another examination of emotion investigates the way in which changes in the body’s state are represented in the brain in the form of what Antonio Damasio called a *somatic marker.* According to this perspective, the bodily responses to a situation or a choice that needs to be made lets the brain know how the individual feels about an experience. Such a somatic marker can then be used as a gut reaction to an experience, providing additional input into cognitive processing. A part of the brain called the *ventromedial prefrontal cortex* is implicated as the site of somatic marker processing, or what is called *intuition.* Allan Schore noted the importance of early experiences with caregivers in the maturation of this region and its role during early development in coordinating self-regulatory functions with basic emotional reactions and social functioning. A closely related region, the anterior insular cortex, mediates the interoception that allows internal somatic states to be represented in the higher cortical areas and the mind to experience awareness of the interior of the body. Being open to monitoring the body's signals permits the modulation of the somatic elements of emotions and allows for the development of emotional regulation. Disorders in self-organization and social functioning may be better understood by examining the central role of emotion and, perhaps, the orbitofrontal and insular cortex and related regions in the development and maintenance of dysfunctional mental states.

Studies also suggest that the orbital and medial prefrontal regions are responsible for subjective experience and self-awareness, enabling the mind to reflect on the self in the past, present, and the potential future, and even to consider the internal states of others. Inborn and experiential factors may play important roles in allowing this region to develop the capacity to integrate a wide range of important functions of the mind, including the appraisal of meaning, emotional regulation, social cognition, and autobiographical consciousness.

NEURAL NETWORK GROWTH AND INTEGRATION

The growth and selective connectivity of neurons is the basic mechanism of all learning and adaptation. Learning can be reflected in neural changes in a number of ways: (1) the growth of new neurons, (2) the expansion of existing neurons, (3) the changes in the connectivity among existing neurons, and (4) alterations in epigenetic regulation of specific regions of the genome. All of these changes are expressions of plasticity or the ability of the nervous system to change.

There is now sufficient evidence demonstrating neuronal growth and change in reaction to new experiences and learning. Existing neurons grow through the expansion and branching of the dendrites that they project to other neurons. Neurons interconnect to form neural networks, which, in turn, integrate with one another to perform increasingly complex tasks. For example, networks that participate in language, emotion, and memory interact and integrate, allowing for an emotionally meaningful story to be recalled and told with the proper affect, correct details, and appropriate words.

Association areas within the brain serve the role of bridging, coordinating, and directing the multiple neural circuits to which they are connected in what is now called "*the connectome.*" Although the mechanisms of this integration are not yet known, they are likely to include some combination of (1) biochemical processing within neurons, (2) synaptic connections among neurons, (3) relationships among local neuronal circuits, and (4) interactions among functional brain systems. Changes in the synchrony of activation of multiple neural networks may also play a role in the coordination of their activity in a process that has been named by Kringelbach and colleagues as "connectome harmonics."

If everything humans experience internally is represented within neural networks, then psychopathology of all kinds, from the mildest neurotic symptoms to the most severe psychosis, must be represented within and among neural networks. The social environment, too, plays a crucial role in well-being and can itself support healthy or unhealthy states. Healthy functioning requires the potential for proper development and functioning of neural networks that are responsible for organizing conscious awareness, behavior, emotion, and sensation. Severe trauma early in life, in situations of abuse or neglect, can alter the healthy development of the brain's integrative potential. Innate neurologic conditions unrelated to an experiential cause may also result in impediments to the connectome functioning as an integrative whole. With integration, flexibility and harmony arise, and health may ensue. Impaired integration may result in chaos and/or rigidity when the differentiation and/or linkage of areas are impeded.

Symptoms of the various syndromes seen through this lens may be envisioned as the chaos and rigidity of impaired integration within the brain. Psychopathology, then, can be conceptualized as an expression of suboptimal integration and coordination among neural networks—whatever the etiology. Given that experience shapes the brain's development, as do nonexperiential factors such as genetics, infection, inflammation, and random events, the finding that individuals with various forms of psychiatric disorders have suboptimal integration in the brain is supportive of this proposal. Patterns of dysregulation of brain activation in specific disorders also support the theory of a brain-based explanation for the symptoms of psychopathology.

In general, psychological integration suggests that the cognitive functions of the executive brain have a high degree of access to information across networks of sensation, behavior, and emotion.

Integration requires both differentiation of elements of a system and their linkage. And integration appears to be essential for optimal regulation. Anything that impedes either differentiation or linkage can disrupt integration at the moment as a function, or over the long run, as a structural impediment leading to impaired self-regulation. Functional dissociation among these processes can occur when biochemical changes caused by high levels of stress inhibit or disrupt the brain's integrative abilities. With brain growth influenced by a range of processes, structural impairments to integration may ensue. Physical trauma, disease processes, or genetic predispositions that disrupt the development and functioning of integrative neural networks can all result in neural dysregulation and psychiatric symptomatology.

In applying this integration and brain-based model to treatment, psychotherapy, psychopharmacology, and psychosurgery can be viewed as ways of creating or restoring integration and the resultant coordination among various neural networks. For example, research has demonstrated that successful psychotherapy correlates with changes in activation in areas of the brain hypothesized to be involved in psychiatric disorders, such as obsessive-compulsive disorder (OCD) and depression. The return to normal levels of activation results in reestablishing positive reciprocal control among relevant neural structures and networks and a reduction of symptomology.

MIND AND BRAIN

What is this activity of the brain, and how does it give rise to such mental processes as perception and cognition? How do the human experiences of perception, thought, emotion, attention, self-reflection, and memory emerge from neural processes?

A generally accepted view of the conscious mind is that it emanates from cortical networks; yet, there are several clinical cases of individuals who have suffered the destruction of a variety of cortical regions who retain consciousness. Moreover, a minor lesion to subcortical areas such as the PAG can result in the complete loss of intentionality, a significant attribute often associated with conscious mental processing. Cognition, emotion, and even personality have further been demonstrated to be influenced by alterations in gut microbiota. These sorts of findings result in the necessity to revise theories that no longer suit the new discoveries. Brain-based perspectives may thus be limited, further necessitating the involvement of evidence supporting the importance of social relationships. It is for these reasons that the prevailing definition of the mind as purely neuronal firing is insufficient. A more flexible definition that incorporates these sorts of findings is the mind is an embodied and relational process that regulates the flow of energy and information. In other words, envisioning human experience as an interaction of mind, brain, and relationships allows mental health practitioners to free themselves from the overly simplistic view that the mind is "just the activity of the brain." Conceptualizing sensation, perception, and cognition as arising from a synergy of mind, brain, and relationships, provides a perspective that is broad enough to integrate findings across various fields of study.

ENERGY AND INFORMATION

The brain is composed of tens of billions of neurons and perhaps ten times that number of supportive cells called glia. When regions such as the cortex (which may have around 20 billion neurons) and the cerebellum (which may have 100 billion) are combined with other areas, an estimate of over 120 billion neurons is derived with the distribution of glia varying in different areas. An average neuron is connected to approximately 10,000 other neurons at synaptic junctions. With hundreds of trillions of connections within and among thousands of web-like neural networks, there are countless combinations of possible activation profiles. The term *neural net profile* is used to describe a certain pattern of activation of the complex layers of neural circuits, which is the fundamental way in which mental processes are created. These activations can lead to further neural processes in a cascade of dynamic interactions that produce a range of internal events and external behaviors.

The essential components of the mind come directly from how these neural events shape the flow of energy and information. How one subjectively feels the texture of this flow from the shifts of energy and information within the matter of the brain and other regions of the body is something no individual or theory has yet to be able to fully explain. Emergence is a property of complex systems, a way in which the components interacting give rise to something more than the components themselves. It may simply be that an irreducible aspect of embodied energy flow, the emergent aspect of that flow, is a subjective experience. The consciousness of that subjective experience and information processing may be emergent properties of energy flow. And one of the emergent aspects of a complex system is self-organization. We can propose that one of the key facets of the mind is a self-organizing emergent process of embodied and relational energy flow; this may be how the mind functions as a regulatory process that influences and is influenced by both embodied and relational energy and information flow.

Table 3.1–1.
Basic Ideas of the Mind

The mind is a processor and regulator of energy and information
Energy is contained within and among the activations of neural circuits as well as in the communication between people
Information is contained within and among the patterns of activation, termed a *neural net profile* or *mental representation,* and information is embedded in energy patterns shared as communication between people
These representations are energy patterns that serve as symbols that cause further effects in the mind, leading to the processing of information both within the nervous system and in relationships

The mind is a processor of patterns in the flow of energy and information within the brain, the body as a whole, and our relationships with our surroundings—people and with nature (Table 3.1–1). Within the brain, activations of individual neurons, as well as groups, circuits, or networks of neurons all, involve the flow of energy through this complex system. On one level, the system includes the brain and its body. On another level, this system may involve communications with one another. Within the brain, this energy reflects the flow of ions across membranes, the consumption of oxygen and nutrients by neural cells, and the active transport of molecules into and out of nervous tissue.

Information is created within the brain by a process of representation. The essential feature of information processing in the brain is that the patterns of activation of neural circuits (the neural net profile) contain information. Recent research from Seth and Friston characterizes the brain as Bayesian. By this, it is meant that there are predictive neural mechanisms influencing how sensory data is being represented. Feedback from alternate circuits even occurs at the earliest stages of visual perception. These mental representations, in turn, produce further neural events. The location and pattern of neural activations determine the nature of what the neural net profile represents. For example, activity in the optic nerve in response to light leads to a cascade of neural responses within the visual cortex, generating a visual sensation. Future activation of those layers in the visual cortex in that general pattern is the recollection of the visual image. Pattern and localization determine the kind of representation and the information that it specifically contains. For example, when an individual sees the Eiffel Tower, the visual system responds with the activation of a neural net profile. When the Eiffel Tower is recalled at a later time, the visual cortex activates a similar neural net pattern, and the Eiffel Tower is visualized. The activation of a particular pattern of neural firing thus contains representations of information about something, in this case, the Eiffel Tower. Examples of representations include perceptual, emotional, sensory, and linguistic forms, as well as more abstract concepts and categories. Extant literature introduces the notion of the brain as an active agent in processing information rather than a passive recipient. Indeed, Raichle notes that 60% to 90% of data processed in the brain occurs outside of external events.

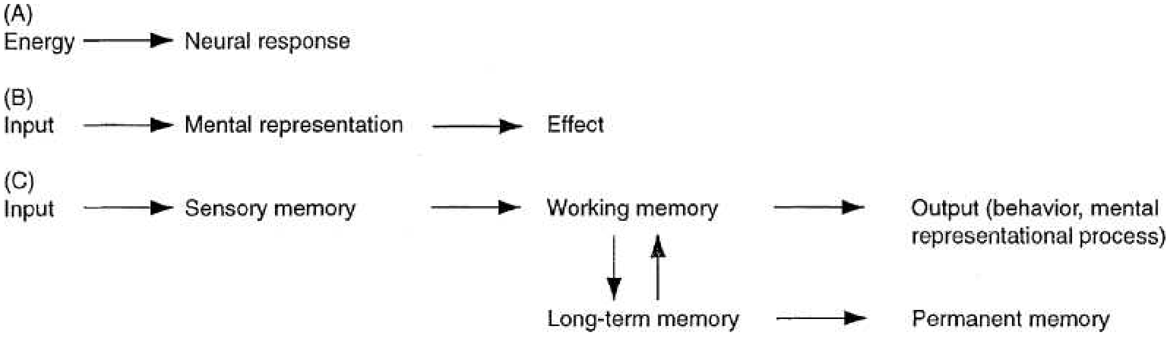

FIGURE 3.1–1. Information-processing models.

INFORMATION PROCESSING

Several elements of the brain’s function as an information processor can be described (Fig. 3.1–1): At the most basic level (Fig. 3.1–1A), energy leads to neural responses. This energy can be in the form of light stimulating the rods and cones of the retina or sound waves vibrating the tympanic membrane. It may also take an internal form in which the flow of energy within neural activations produces subsequent neural responses.

A second level of conceptualizing information processing (Fig. 3.1–1B) lies in the idea that an input (internal or external) leads to a representational response (a neural net profile of activation), which, in turn, produces a downstream effect or output. This output can be internal, such as the generation of other representations, or external, such as in the form of observable behavior. Within cognitive psychology, these information-processing events can be analyzed in terms of contrasting, comparing, generalizing, chunking, clustering, differentiating, and extracting processes, all of which lead to increasingly complex mental representations.

A third level of understanding information processing in the mind (Fig. 3.1–1C) is the conceptualization of sensation, perception, attention, and memory. According to this view, external energy is sensed by the peripheral nervous system and is registered as a sensation within the brain. The selective processing of aspects of these sensations, called filtering, leads to the production of perception. These perceptions are subject to further filtering during which only a select few are placed within working memory; this is sometimes called the "chalkboard of the mind." It is within working memory that representations can be consciously manipulated, contrasted, clustered, and reassembled. Thus, consciousness may be intimately related to this aspect of mental functioning.

Sensation refers to the initial stages of the basic information-processing model (Fig. 3.1–1). In traditional experimental paradigms, sensory memory is conceptualized as lasting for approximately 0.25 seconds. Items in sensory memory are then filtered into working or short-term memory, where they last for approximately 0.5 minutes. When humans consciously attempt to learn new information, working memory is able to handle approximately seven items unless further processing creates links to other items within longer-term memory. *Rehearsal* allows these representations to remain in working memory for longer periods of time. Cognitive processes that can group bits of information into large chunks (chunking) can increase the capacity of working memory by making each unit more information rich. Representations are then processed and placed within long-term memory, where they can be retrieved for future use.

ATTENTION

Attention is the process that controls the focus and flow of information processing. Three components of attention (selectivity, capacity, and sustained concentration) have traditionally been used to describe cognitive deficits seen in psychiatric disorders, such as schizophrenia and attention-deficit/hyperactivity disorder (ADHD). All aspects of attention in clinical or nonclinical populations are influenced by the emotional or motivational value of the stimulus.

Early conceptualizations of attention were based on Donald Broadbent's idea of a filter that selects a limited number of incoming stimuli to be further processed. The limited capacity of attention was thus seen as being attributable to the inability to process the overwhelming number of incoming stimuli. An attention "bottleneck" was described as occurring early in the sensory process (automatic) or late in the perceptual processing stage (identification and classification).

Selective Attention

Attention focuses a metaphorical spotlight on external stimuli or internal mental representations. In Broadbent's conceptualization, selectivity has three dimensions: (1) filtering—focusing on specific attributes (e.g., large squares vs. small squares); (2) categorizing—recognizing information based on stimulus class (e.g., attending to letters in whatever script they are written); and (3) pigeonholing—reducing perceptual information needed to place a stimulus into a specified category (e.g., using only long hair to classify individuals as female). Each of these aspects of attention acts on incoming stimuli to make a determination of fit for the sought-after characteristic.

Attention Capacity

The concept of processing capacity involves the idea that a given task makes a demand on a limited pool of resources: A task with a high-processing load draws more resources from the finite pool than does a task with a low-processing load and thus will inhibit the accessibility of resources for other simultaneous functions drawing from the same pool. Focal attention requires cognitive effort and thus has a high-processing load demand. Cognitive models describing several resource pools suggest an executive process that distributes resources to various cognitive functions.

Optimal performance is attained when there are moderate levels of arousal because this allows for the establishment of task goals, feedback, and appropriate resource allocation. Low levels of arousal impair those processes and lead to inadequate resource allocation, whereas high levels of arousal may be detrimental to the performance because of poor discrimination of stimuli and diminished efficiency of allocation, resulting in poor attention functioning.

Sustained Attention

The ability to sustain attention is called *vigilance* and can be tested with task demands for alertness and concentration over a period of a few minutes to an hour. The tests usually involve detection requirements for target stimuli that infrequently occur at random intervals. An example of such a test is the Continuous Performance Test. Defining aspects of the tests are derived from signal detection theory and include the factors of *sensitivity* and *response criterion.* Sensitivity is the distinguishing of target stimuli from nontarget stimuli; the response criterion is the amount of perceptual evidence required to support the decision regarding a target item versus a nontarget item.

SENSATION AND PERCEPTION

Forms of representations, including sensory and perceptual representations, derive from input from the external world via the peripheral sensory nervous system. The initial stage of encoding a visual representation is called an iconic image and is held within sensory memory for a brief period. Features of the initial stimulus, such as its size, direction, and color, are held within this sensory representation. This is termed bottom-up processing and is as direct as the nervous system allows. As the initial sensory activations are processed, they become influenced by higher-order, prior-experience–shaped influences called top-down processes and are organized as perceptual representations. In other words, learning from the past embedded in synaptic neural patterns of memory shapes perceptual streams of representations in the moment. This can be seen as an important distinction between sensation and perception.

Attentional processes, at the level of sensory memory, act on the initial image with higher cognitive functions, such as classifications, memory, and chunking. In their essence, these top-down processes compare, contrast, and transform the initial representation from sensation to create new perceptual images within working memory. These processes are often automatic, rapid, and outside of awareness.

Perception is thus created by the top-down transformations of sensory images but does not necessarily involve the experience of consciousness. This has important clinical implications in that patients may be influenced by events and stimuli that they cannot consciously recall. Take, for example, the experience of a young girl with anorexia or body dysmorphic disorder. When she looks at herself in the mirror, she becomes consciously aware of the perception of her reflection in approximately half a second (500 ms). In contrast, the sensation of her visual reflection has already reached her eyes almost instantaneously and has resulted in nonconscious neural activation in only 50 ms. As neural networks are being instantiated within her visual cortex, activation in other brain regions, such as the amygdala, is also occurring, streaming top-down emotional and appraisal information back to the visual cortex.

In the 450 ms between initial sensory activation and conscious perceptual awareness, the visual experience of her reflection is constructed through a combination of bottom-up sensations and top-down emotion, learning history, and other visual memories to which it is compared and contrasted. What she focuses on, how she feels about it, and what it means to her—in essence, how she is seeing what she is seeing—is a blend of bottom-up and top-down constructed in automatic, nonconscious processes. At 500 ms, she experiences her brain's construction of her body, this perceptual image, as an external reality. In other words, a great deal happens on the way from sensation to perception.

MEMORY SYSTEMS

The neural networks of the brain are capable of responding to experience through the activation of neural net profiles—particular patterns of distributed neural activation. Donald Hebb described a basic principle of memory, paraphrased by Carla Shatz, that has been repeatedly supported by research: "Neurons that fire together, wire together." Neurons that are activated in a particular pattern at one time tend to fire together in a similar pattern in the future—this long-term potentiation (LTP) is the essence of memory.

Table 3.1–2.
Memory Systems

Implicit

A behavioral, emotional, and perceptual form of memory devoid of the subjective internal sense of something originating in the past. Can include schema or mental models that are summations of representations from numerous experiences

Also known as *early, procedural,* or *nondeclarative memory.* Cannot be directly expressed in words

Present from birth. Does not involve the hippocampus or require focal, conscious attention. Probably involves various circuits, including those of the basal ganglia, limbic system (amygdala, anterior cingulate, and orbitofrontal cortex), and perceptual cortices

Explicit

A form of memory requiring conscious awareness for encoding and involving the subjective sense of recollection and, if autobiographical, of self and past

Also known as *late, episodic* or *semantic,* or *declarative memory.* Can be directly expressed in words

The autobiographical component of explicit memory does not fully develop until past the first 2 years of life, as the hippocampus and orbitofrontal cortex, on which it depends, are maturing.

The brain has various circuits responsible for different systems of memory (Table 3.1–2): The form of memory most commonly thought of as *memory* is termed *explicit* or *declarative memory.* This form involves the conscious sensation of something coming from the past, a feeling of remembering, focal attention (attention with awareness), recall of autobiographical or factual knowledge, and hippocampal activation. In these circuits, items focally attended to are placed in working memory, processed further, and then placed in long-term memory. After a period of weeks or months, items are thought to undergo a process called *cortical* *consolidation* that places them in permanent memory, where their retrieval no longer requires the hippocampus.

Explicit autobiographical memory becomes reliably available after 3 years of life once the hippocampus and cerebral cortex have sufficiently matured. Before this, a form of memory that does not require conscious awareness and focal attention for encoding, called *implicit* or *nondeclarative memory,* is already in place and remains active throughout the life span. This form of memory involves a wide range of neural regions, including behavioral, emotional, and perceptual systems. When these circuits are activated in the retrieval of pure implicit memory, they do not include the conscious sensation of something being recalled. For example, when riding a bicycle, a person may not recall having learned to ride and may not even feel that anything is being recalled. Similarly, a person with a fear of dogs may be unable to explicitly recall any event that may explain such an emotional response.

The existence of intact implicit recollection in the absence of explicit memory is found in various conditions, including being under surgical anesthesia or having neurologic conditions such as Korsakoff syndrome and bilateral hippocampal strokes. Dissociation of memory systems may also occur in response to trauma. For example, patients with posttraumatic stress disorder (PTSD) may not explicitly recall a traumatic event, yet they may implicitly avoid stimuli similar to the initial trauma. They may also evidence implicit symptoms such as a startle response, anxiety, and intrusive perceptual images for that event in the absence of an explicit autobiographical memory. These latter symptoms are thought to reflect the intrusion into conscious awareness of implicit memory in the absence of the subjective sensation that something is being recalled.

CONSCIOUSNESS

The vast majority of mental processes are outside of awareness as energy and information flow without consciousness. A simple way of defining conscious awareness is the subjective sense of knowing. When someone is aware, there is a knowing. And the object that one is aware of is the known, such as visual input, a memory, or a thought. In general, many authors' views converge on the idea that there exist two fundamental forms of consciousness, enabling a knowing of two forms of what can be known in consciousness: A *here-and-now* and a *past–*

present–future process of awareness. These two forms of consciousness are likely mediated via the integration of different neural circuits in the brain, and the experience of the known is distinct, one in the moment, the other focusing on events across time.

Two hypotheses focus on the way in which representational processes are linked or are bound together during the flow of informational transformations within the mind: Crick and Koch suggest that a 40-cycle-per-second sweeping process extends from the thalamus to the neocortex. This sweep may serve to bind representational processes together in the internal experience of consciousness. Processes that are active at the time of the sweep then become linked within consciousness. In more recent research, they have suggested the claustrum as a convergence zone that may be involved with unifying consciousness into a singular experience. A second view implicates the lateral prefrontal cortex and its role in working memory—working memory serves as the chalkboard of the mind, and representational processes that become integrated to a certain degree and linked to the activity in this region fall under the attentional spotlight of conscious awareness called *focal attention*.

Theories of consciousness continue to be unresolved and are by far one of the most controversial subjects of debate. Luckily, discord within the community has spawned creative theories which may provide the proper input towards some form of resolution. Northoff proposes that consciousness is a product of relational connections, specifically between the spatiotemporal alignment of the brain with the world. His theory was influenced by research into mental illnesses, such as schizophrenia, whereby disturbances in consciousness are evident. Gerald Edelman asserts that consciousness is epiphenomenal, arising through natural selection via complex neural processes such as reentrant signaling in interaction with the environment. Dynamic recursive activity among distributed neuronal ensembles induces the layering of multiple sensory-motor maps, which, when integrated with memory and prospective systems, may eventually result in experiential continuity.

Beyond materialist and physicalist views of consciousness, there are two theories that have emerged. Tononi, a previous student of Edelman's and whom Koch supports, views consciousness as integrated information. His theory translates phenomenologic tenets into quantifiable units. Thus far, his theory has been used to quantify potential consciousness in coma patients. Hoffman asserts that perception favors fitness over truth; in other words, we do not see reality as it is, rather we see it as it best fits our survival. This conclusion was reached through computer-simulated games that mimic the process of evolution. Digital units which operated in accordance with the truth of what composed their reality perished, in contrast to those which operated according to natural selection. His conclusion, supported by mathematical formulae, is that conscious agents, not matter, are the fundamental elements of the world. Thus, to him, even brain processing is akin to holding a magnifying glass to view pixels of a computer screen rather than understanding the coding that gives rise to its presentation.

Disturbances of Consciousness

Consciousness provides a sense of continuity—yet many psychiatric patients experience a profound sense of discontinuity and confusion that may be related to a dysfunction in the sense-making, continuity-creating process of consciousness. Some psychiatric symptoms, including derealization and depersonalization, intrusive memories and flashback phenomena (as in PTSD), and hallucinations (as in psychotic states), may be conceptualized in terms of dysfunctions in conscious functioning.

Misidentification syndromes are other manifestations of disturbances of subjective, conscious experience. In prosopagnosia, for example, patients are unable to consciously access memories regarding persons familiar to them. Patients suffering from Capgras syndrome, on the other hand, are able to recognize a familiar person’s face but cannot experience the feelings of recognition and emotional arousal associated with that person. This lack of emotional recognition then results in their conviction that their loved ones have been replaced by imposters. “Being certain,” as in the process of recognition, is one aspect of consciousness that is a cognitive process. Theoretically, then, the pathologic uncertainty of patients with OCD can be viewed as a disturbance in that aspect of conscious functioning.

Another example of disturbance of conscious functioning is found in cortically blind patients, who state that they cannot see visual stimuli but behaviorally respond as if they were fully sighted. Cortically blind patients describe being unaware of visual perception yet are able to make eye and hand movements that reflect the processing of information about stimulus location, shape, orientation, and direction of motion. In information-processing terms, behavioral tests reveal that these blind-sighted patients do sense and perceive visual stimuli but do not have conscious awareness of this perceptual process. Cortical blindness, then, is an example of the dissociation of the usually associated processes of perception and consciousness or awareness of phenomena.

Mental Models and Schemata

Studies of perception and memory support the view that the mind has organizational structures that influence the interpretation of sensory data, shape the encoding of information into long-term memory, bias the retrieval of items stored in memory, and help to determine the behavioral responses. These organizing cognitive functions are called *mental models* or *schemata.*

Mental models are highly organized, implicit, top-down processes derived from past experiences that guide the interpretation of present stimuli and influence the direction of behavior. The adaptational value of a given mental model depends on an accurate reading of the survival demands of the situation. The downside of these models is seen when their unconscious and automatic activation occurs in situations in which they are inappropriate, such as in PTSD, or when they no longer sufficiently meet the demands of a given situation.

Aaron Beck's theory of depression is based on the idea that mental models or schemata can guide depressive thinking as they trigger and enhance depressed moods. John Bowlby and Mary Ainsworth used the concept of internal working models to describe the development of early forms of schemata for relationships: Difficulties in intimate relationships and related behavioral dysregulation can be seen as derivatives of models of inadequate early attachment and the presence of multiple, conflictual models. The *inner objects* of psychodynamic theory are also examples of such mental models. Some psychiatric signs and symptoms can be seen as derivatives of conflicted schemata and situations. Classic descriptions of interpersonal patterns in some patients with personality disorders, such as idealization and devaluation, can be seen as maladaptive schema functions.

Thought and Language

Although there is no universally accepted definition of *thought,* thinking most generally involves the mental representation of some aspect of the world or of the self and the manipulation of those representations. Thinking depends on explicit and implicit memory of prior experiences and is influenced by a person's emotional state, mental models, and other unconscious determinants. The basic components of thinking include categorization, judgment, decision making, and problem-solving. Cognitive processes, such as thoughts, are often directly known only through their experience in consciousness and translation into language. As a result, there is always a question of whether thought and language can be fully separated.

Cognitive science has traditionally viewed language as a dominant influence on subjective experience. Language shapes the ways in which the world is perceived, the manner in which desires are communicated and satiated, and the way in which society responds. For example, the language used to describe and define psychiatric disorders can shape how mental health professionals perceive what a patient may be experiencing as top-down filters shape and constrain the sensory input from the patient's presentation and mold a clinician's assessment and treatment planning.

Rational thought contributes to the ability to judge the probability of uncertain events and to choose among various options. These processes contribute to problem-solving in which data are assessed, classified, transformed, and compared on the basis of logical rules to produce a choice that solves a problem; failures in these steps can result in limitations and distortions in normal thought processes.

Modes of Processing and Laterality

The mind is capable of distinct modes of processing mental representations: A serial mode uses sequential processing in a linear fashion but is thought to be slow and energy-consuming because only a few items can be processed at a time; focal, conscious attention is believed to occur in this serial fashion. A parallel mode involves the simultaneous manipulation of large numbers of representations in a nonlinear fashion and occurs in a rapid, low–energy-consuming process; pattern recognition is one example of parallel processing that can deal with a wide array of stimuli at the same time.

Another distinction in contrasting modes of processing has been identified in the type of mental processes primarily attributed to the right and left cerebral hemispheres. Studies in which patients whose corpus callosum have been surgically severed, who have had unilateral neurologic lesions, or who have undergone brain-imaging protocols have revealed a remarkable consistency in trends of left- versus right-hemisphere functioning. Some general principles from these studies suggest that whereas many processes involve both hemispheres, there are also distinct patterns primarily originating from one side of the brain or the other.

The forthcoming generalizations relate to right-handed individuals and to most left-handed people: In the right hemisphere, there are fast-acting, parallel, holistic processes, including perception and communication of nonverbal signals. The right side specializes in representations with a wide spectrum of values, such as images and sensations, and in the nonverbal meaning of words, sometimes referred to as analogic representations. The right hemisphere is thought to work as a pattern recognition center, capable of assessing the Gestalt context of a scene and providing a synthetic interpretation. In the left hemisphere, there are primarily more slowly acting, linear, time-dependent, serial processes. Left-hemispheric processes manipulate the verbal meaning of words in a logical, analytical mode of processing, focusing on the digital or clearly demarcated meaning of terms categorically organized, such as up or down, right or wrong, left or right. A generalization from a number of studies is that the right hemisphere tends to note the patterns in the world and creates contextual meaning, whereas the left hemisphere notes the details of what it perceives and attempts to organize and make sense of the world, creating a cause–effect explanation of experience.

The processing of emotion appears to be biased toward the right hemisphere. Facial recognition of the affective expression of others also appears to be a specialty of the right hemisphere. Notably, the right hemisphere appears to have a more integrated representation of the body’s internal status, or interoception, information that may be essential for individuals to know how they feel. The hemispheres also seem to demonstrate biases with regard to emotional tone, with the left having a positive affect bias and approach state, whereas the right has a negative bias and a withdrawal state. Integration and balance of hemispheric influences appear to be necessary for adequate affect regulation and mental health.

Jerome Bruner described the distinction between an earlier mode of thought, which he referred to as *narrative cognition,* versus a later mode, called scientific, logical, or paradigmatic cognition. Narrative thinking is a context-dependent form of processing that incorporates the internal experiences of the teller with the perceived expectations of the listener in the production of a story. Stories also involve the subjective experiences of the characters involved in the unfolding sequence of events. Children develop narrative thinking by 2 years of age, and the coconstruction of stories between parent and child is a primary mode of communication across cultures. Drawing on context, emotions, and linear explanations, such narratives help make sense of experiences in ways that involve both the right and left hemispheres of each person.

Logicoscientific paradigmatic processing is said to occur in a context-independent manner that focuses on abstract concepts and their logical, cause-and-effect relationships and thus may rely primarily on left hemisphere dominant processing. The unfolding of each of these distinct processes coincides with the specialized processing of the right and left hemispheres and the sequential process of developmental maturation. See McGilchrist (2009) for an expansive and fascinating exploration of the origins and impact of cerebral lateral specialization.

Metacognition and Self-Reflective Capacity

Metacognition concerns conscious processes that act on cognitive processes—thinking about thinking. Awareness of cognition appears to develop by approximately 6 years of age and takes various forms, including the appearance–reality distinction (things may not be as they appear). This is well documented since Piaget's formulations on the conservation of liquid task. Two components of this awareness are representational diversity (the same object may appear to be different to different people) and representational change (thoughts today are different from those of yesterday and may be different again tomorrow). This form of knowledge about the person-specific meaning of cognitive representation requires some sense of the person's awareness of the separateness of minds, a theoretical domain in developmental cognitive psychology called the *theory of mind.*

The regulation of cognition, also called *metacognitive monitoring,* includes such processes as planning activities, monitoring activities, and checking outcomes. Metacognitive monitoring may involve the assessment of thinking sequences for fallacious logic, factual errors, and contradictions in the content of speech. Peter Fonagy and colleagues explored the development of the reflective function or experiencing an internal observer of mental life. Parents teach children how to be self-reflective by including their own internal state in interactions with them and by encouraging children to share their own. Children who have been taught to tell stories that include mental states usually demonstrate a greater frequency of secure attachment. Being able to understand and consider the mental states of self and others has also been shown to decrease dependency on defensive strategies.

Research suggests that what is created in parent–child narratives about experience is not just a story; embedded within the storytelling is the selection of information to be included, how it is to be processed and understood, and whether it is egocentric or has multiple subjective centers. Children who can appreciate more than one perspective in a story may develop greater empathic capacity.

Social Cognition

Bridging the fields of social psychology and cognitive psychology, the study of social cognition focuses on the mental processes involved in social interactions. Its domains include the study of empathy, interpersonal communication (verbal and nonverbal), person perception, relationship scripts, and group processes. Other related areas include studies of attribution bias, memory for social interactions, stereotyping, mental control of social cognitive processes, and cognitive origins of a sense of self. Social cognition can be seen as a domain of social psychology that uses information processing theory to assess the components of attention, perception, encoding, memory, retrieval, and schemata. A dominant theme in social cognition research has been that top-down, theory-driven processing influences interpretations of and behaviors in social situations. Developmental psychologists have focused on the origins of social cognition and its deviation to reveal, for example, that children with autism spectrum disorder (ASD) have significant deficits in empathic capacity and in the ability to interpret social cues. Social cognitive deficits are also present in other psychiatric disorders, such as bipolar disorder, PTSD, and social anxiety.

There are two theories proposed to explain how we are capable of understanding another person's point of view: "Theory-theory" and "Simulation Theory." Theory-theory posits that we develop internal social models based on an individual and use them to predict their behavior.

Simulation theory suggests the idea that we put ourselves in the other person’s position in the process of empathy. Simulation theory has gained favor as a result of recent research on mirror neurons, which are hypothesized to be involved in an intricate process creating resonance, sympathy, and “mirrored states” in the perceiver. Mirror neurons activated during the perception of another’s intentional acts stimulate neural networks that organize the same motor actions. This work has led to several hypotheses involving a process by which these cortical maps of behaviors are communicated to the limbic areas and down into the body itself. The resulting emotional and bodily states are then registered in the middle areas of the prefrontal cortex involved in self-experience. The “resonance” of the perceiver’s limbic and somatic states with that of the person being observed then serves as an unconscious template through which the prefrontal region assesses how the self is “feeling.” Thus, based on simulation theory, what one experiences as self-states is highly influenced by the experiences of those around them.

These networks are hypothesized to be the substrate of sensing another person’s mind—those intentions and emotions experienced by another are now felt within the observer. Maps of others’ minds are thought to be created in the medial prefrontal cortex and related areas allowing for the empathic understanding of another person’s experience in what can simply be called mindsight or mentalization. One hypothesis regarding the social cognition deficits in autistic individuals suggests that there may be difficulties in the activation of mirror neuron-related areas, either from impairment in input or functioning to these social circuits of the brain. Further research is needed to validate this preliminary view, but it serves to highlight how impairments in neural mechanisms underlying empathy and the ability to map out another person’s mind may be at work in a range of disorders.

Discourse and Narrative

Discourse is communication from one person to another; it is thought to involve a sense of intention or plan. Normal discourse follows a set of rules that ensures the coherence and effectiveness of communication as what is intended to be communicated by the sender is understood by the listener or receiver. Some researchers support the idea that discourse is a cognitive function that follows the basic principles of information processing, including a schema for effective communication that takes into account the listener’s perspective. Incoherent discourse can be noted by analyzing violations of the primary maxims of discourse in a process called *discourse analysis,* which examines the ways in which discourse deviates from an assumed discourse plan.

The exact method to quantify abnormalities in discourse is controversial, but clinical impressions of incoherence are important for assessing deficits in thought and social communication. The deficits may result from learned behavior, inherent cognitive abnormalities in thought or language, or deviations in social cognitive functioning, and they are clinically evident in psychosis, specifically in schizophrenia, and in individuals with unresolved trauma.

Narrative is a broad domain of research ranging from the literary study of fiction to investigations of the origin of autobiographical accounts in developmental psychology. From a cognitive point of view, narrative is important in understanding the relationships among language, memory, consciousness, mental models, self-schemata, and social cognition. *Narrative* can be generally defined as the way in which a person creates a verbal account of a sequence of events in the world and an internal experience of the characters of the story.

Studies of early monologues find that young children interpret and assign meaning to events in their world from an early age and that narrative helps to record and make sense of the past, interpret the present, and anticipate the future. The brain has been called an “anticipation machine,” and mental models, prospective memory, and narrative are all ways in which networks of top-down processing prepare for possible futures. The enactment of narrative themes guides the way in which individuals live out their lives.

Cognitive Development

Developmental theories and research can be divided into several views: Stage theories (Jean Piaget, neo-Piagetian approaches, and the sociocultural school of Alexander Luria and Lev Vygotsky) describe discontinuous periods of development, with periods of stability and consolidation alternating with periods of instability and transition. Information-processing models postulate a nonstage theory, in which the emergence of cognitive capacity is a continuous process rather than a set of invariant sequences. Both stage and nonstage views embrace the idea that hierarchical integration and ongoing differentiation are fundamental aspects of cognitive development.

A distinguishing feature of each of these theories is the relative weight given to the role of innate, biologic factors on the one hand and culturally determined social learning experiences on the other. In other words, do cognitive capacities emerge from a genetically determined plan, as in the Piagetian view, or do they develop in response to experience, as in the sociocultural view? More recent conceptualizations have drawn on the functioning of complex systems, supporting the idea of a transaction between innate factors and environmental experiences in the emergence of ever-more-complex capacities.

Psychiatric disturbances in cognition may reflect arrested patterns of typical cognition (mental retardation), atypical developmental pathways (ASD), and specific cognitive impairments (schizophrenia) that may have been present early on or that only became evident after life requirements such as school, became demanding. Investigations into the developmental features of these disorders are a major focus of the field of developmental psychopathology.

Self-Organizational Processes

An understanding of the development and subjective experience of cognitive processes has been greatly informed by insights from the fields of evolutionary neurobiology and the nonlinear dynamics of complex systems, otherwise known as *chaos theory.* With billions of neurons and trillions of synaptic connections, the brain is capable of organizing an incomprehensible number of activation patterns. In selectionist theory, the billions of neurons become organized into networks with similar functions that can be selected and reinforced by interaction with the environment in an experience-dependent fashion. In addition, the brain has value systems that selectively reinforce the activity of neuronal groups that enhance survival. In this way, the brain's neuronal groups compete and differentiate within the brain in an ever-evolving adaptational system.

Chaos theory suggests that complex systems adhere to a specific set of principles. Three of these principles, nonlinearity, self-organizational processes, and movement toward complexity, are especially relevant to psychiatry. Nonlinear refers to the finding that small changes in input (or initial conditions) can lead to large and difficult to predict changes in output. Complex systems function on the rules of probability, which predict that certain combinations of activity within the system are more likely than others and that these combinations will tend to move the system toward self-organization. This probability also predicts that the system moves toward increasingly complex states of functioning. A complex system is a collection of elements that have the three characteristics of being open to influences from outside of the system itself, being capable of chaotic or essentially random behavior, and being nonlinear.

Complexity theory may offer a foundation for proposing a useful working definition of mental health applicable to individuals, families, and larger social systems. In complex systems, self-organizational processes that move the system's states toward maximal complexity are mathematically shown to be the most flexible, adaptive, coherent, energized, and stable. The movement toward complexity lies between the extremes of sameness, with rigidity and order on the one side, and change, with randomness and chaos on the other. Complexity is achieved when the components of the system achieve a balance between the two fundamental processes of differentiation (specialization in function) and linkage (connected together as a functional whole). Please note that in some neuroscience studies, the term "integration" is used to indicate linkage, and the term "segregation" is used to indicate differentiation. Here we are using the same notions but with the overall blending of the linkage of differentiated components to be designated with the term integration. When a system links differentiated parts, it becomes integrated and achieves a state of complexity and harmoniousness, advancing adaptive functioning bordered on either side by chaos and rigidity.

Examination of psychiatric syndromes reveals examples of deviations from this integrated complex state in which an individual exhibits states and traits of chaos and/or rigidity. For a given individual, such a balance can be achieved when the genetically and experientially influenced growth of neural circuits combines the differentiation of specialized regions with linkage via neural fibers connecting widely distributed areas into a functional, integrated whole. According to this view from the interdisciplinary field of "interpersonal neurobiology," disorder can be seen as occurring when a system's integration is impaired, and its flow toward complexity is impeded as revealed in its movement toward either rigidity or chaos.

Trauma may impair integration by blocking neural linkages within an individual, as seen in the negative effects on the integrative regions of the corpus callosum, prefrontal cortex, and hippocampus in abused and neglected individuals. A dysfunctional family system would be thought to occur when the individuals are excessively differentiated (without emotional connections) or linked (enmeshing that inhibits individuality from being expressed). Such stressed systems are limited in their movement toward complexity and, hence, in their stability, flexibility, and adaptability. Such impediments to integration make a system prone to chaos and rigidity.

Psychiatric disturbances may be conceptualized as disturbances in self-organizational processes. Both inherited and experiential internal determinants, as well as ongoing external, environmental, and social influences that place constraints on the system, can directly affect the development and effective use of integrative self-regulatory mechanisms. Integration is the underlying mechanism of optimal self-regulation. And interpersonal integration likely supports the growth of integrative regions of the brain. Clinical interventions may thus function at the level of external constraints (psychotherapy with integrative communication within a relationship) or internal constraints (pharmacologic treatments) that alter the ways in which the individual’s mind is able to achieve integration and healthy forms of self-organization. Optimal self-organization is created with integration, the linkage of differentiated parts of a system. By seeing the mind as something beyond “brain activity” and “subjective experience” alone, a definition of one aspect of mind as “the embodied and relational, self-organizing emergent process that regulates the flow of energy and information” makes it possible to understand the interconnected nature of the broad range of disorders and interventions as related to the fundamental process of integration. Viewing psychiatric disturbances in this way also allows for a synthesis of the views of psychodynamic, biologic, and social psychiatry.

States of Mind

The state of activation of the various parts of the system can cluster into repeated patterns called *states.* In the brain, a *state of mind* or *mental state* describes the way in which various neuronal groups may become activated at a given time. Repeated patterns of neuronal group activation—a neural net profile—can become reinforced if they occur frequently or if the value system of the brain ingrains their profile. These ingrained patterns of activation are called *attractor states*; those states that are least likely to occur are called *repellor states.* These mental states are determined by the constraints on the system, and modification of constraints allows the nature of attractor and repellor states to be altered. Constraints on the system are both external and internal: Features of the external environment, such as the way in which other people behave and relate to an individual, can directly affect which mental state is more likely to be activated within the person. Internal constraints include the synaptic strengths of association, as determined by constitutional features and genetics, and those learned from experience, as encoded within memory processes.

Repeatedly reinforced patterns of neuronal group firing link the cognitive processes of attention, perceptual bias, memory, mental models, behavioral response patterns, emotional tone, and regulation. These states of mind result in the patterns of cognitive, emotional, and behavioral symptoms seen in both mental health and various psychiatric disorders; in a depressed state of mind, for example, one may pay conscious attention to negative aspects of experience, interpret incoming stimuli in a pessimistic manner, have greater access to depressing past experiences, activate mental models of the self as bad or guilty, have behavioral patterns of withdrawal, and have a depressed mood with difficulty regulating intense affect.

Healthy mental functioning may depend on a flow of state of mind through time, which can then allow for flexible adaptation to an ever-changing environment. This relates back to chaos theory, which suggests that nonlinear complex systems must move continually toward maximizing complexity by balancing predictability and novelty. This is how mental health can be seen to emerge from integration, both within the individual and between the individual and the environment, especially the social world.

SENSATION, PERCEPTION, AND COGNITION IN PSYCHIATRIC DISORDERS

Since the time of Emil Kraepelin and Eugen Bleuler, psychiatrists have known that certain disorders include profound disturbances in cognitive functioning. Since the 1950s, researchers have attempted to determine the exact nature of such deficits. With advances in computer technology and an increasing technical ability to analyze stimulus presentation and response times on the order of tens of milliseconds, cognitive psychologists have been able to devise research paradigms capable of testing increasingly subtle aspects of cognitive processing.

Processing research has focused on all three domains of sensation, perception, and cognition. Sensory-processing studies, using simple stimuli, focus on poststimulus events for as long as 1 second; perceptual studies, using slightly more complex stimuli, examine processing after a period of as long as approximately 5 seconds. Cognitive processing experiments can examine the early aspects of processing (e.g., phenomena occurring within the first 30 seconds), as well as long-term processes that occur over hours, days, or even years. Recent attempts have been made to correlate complex cognitive findings with clinical presentations.

A general problem with correlating cognitive processes with clinical populations lies in the symptomatic heterogeneity of patients falling within the same diagnosis. A related problem is a distinction between general and specific deficits: Do psychiatrically ill patients perform less well on a given task because they are ill or because of a deficit specific to the disorder? For example, psychomotor slowing, as measured by reaction time and response rate, is seen in schizophrenia, depression, and other psychiatric and neurobehavioral disorders. The side effects of medications can also influence processing and response speed. Researchers rely on the creative design of experimental tasks to help distinguish between general and specific deficits. A comparison of target-patient populations with matched healthy persons and other psychiatric patients can help to determine disorder-specific cognitive dysfunction. New research surrounding deficiencies of a protein called SV2A has revealed a linkage of this state to deficiencies in cognitive tasks in patients with schizophrenia. Medications address positive symptoms with clinical efficacy, but cognitive impairments and negative symptoms are still ineffectively addressed to date.

Another general issue is that of state markers versus trait markers; a patient with schizophrenia, for example, may have a cognitive deficit only when actively psychotic (state) or all the time (trait). These results have been found in certain cognitive tests of attention that correlate with improvement on medications. Some atypical results are also found in the nonill first-degree relatives of patients with schizophrenia. Is the marker of genetic vulnerability a coincidental finding or part of the core deficit in schizophrenia? An exploration of the implications of these cognitive differences for the daily life of the patient is an important application of the research findings to clinical psychiatry.

Schizophrenia

Three general hypotheses inform our understanding of the biologic foundations of schizophrenia: NMDA-Glutamate hyperactivity, hyperactivity in the mesolimbic system, and hypoactivity in the mesocortical system. Positive symptoms are attributed to mesolimbic activity and negative symptoms to mesocortical activity (along with nucleus accumbens/reward circuitry). Schizophrenia is often associated with positive and negative symptoms, yet cognitive and affective symptoms are just as significant, especially in relation to problems adjusting to real-world demands. In general, cognitive symptoms have been correlated with inefficiencies in the dorsolateral prefrontal cortex, whereas affective symptoms relate to the ventromedial prefrontal cortex, with aggressive symptoms corresponding to the orbitofrontal cortex and amygdala.

Even before higher-order cognitive processes, individuals afflicted with schizophrenia suffer from issues with sensorimotor gating. Studies have repeatedly found that schizophrenic patients have differences in their capacity to tune out irrelevant data. This is usually an automatic process that occurs in healthy individuals. Attention may become compromised when these implicit processes are malfunctioning. They also show a markedly diminished capacity to habituate, for example, demonstrating a persistent acoustic startle reflex with repetitive tones. In general, deficits in habituation and sensorimotor gating support the idea that schizophrenic patients have a diminished capacity to regulate the flow of rapidly presented information; they experience being inundated by stimuli that are regularly filtered out in the brain of a nonclinical person. Deficits in processing may also occur with internally generated stimuli, producing an overwhelming and confusing experience of the internal and external world.

Three of the most prominent cognitive problems include executive dysfunction, social cognition, and attention. Executive dysfunction is typically registered in assessments evaluating cognitive flexibility, conceptualization, complex problem solving, and verbal fluency. Problems in these domains highly correlate to reduced adaptive potential in daily functioning. For example: to successfully maintain and/or ascend in the domain of occupation, processes involved with strategizing and decision-making are critical. Fostering relationships are just as important, and in order to do so, individuals must have an accurate perception of others and the meaning of their interactions. In this regard, individuals who suffer from schizophrenia have been shown to exhibit hyperactive amygdala activation when viewing neutral faces and hypoactivity when viewing fearful and emotionally charged facial expressions. This may result in social misperceptions such as perceiving a threat in a neutral environment or neutrality in a potentially threatening environment.

Selective attention is a necessary component to learning and responding appropriately to any given context. Patients diagnosed with schizophrenia have an impairment in their ability to avoid distracting stimuli (to filter) and to pigeonhole. Recent conceptualizations have also examined a generalized impairment in the information-processing capacity in schizophrenia. The capacity model examines the way in which a pool of attention is allocated across mental activities. Two components of this model are the *quantity of resources available* (capacity) and *executive allocation policy.* Schizophrenic patients may also suffer from an impaired response selection process and thus demonstrate atypical responses on tasks.

Sustained attention, or *vigilance,* is required to process stimuli of long duration. The most common research paradigm used to examine sustained attention is the *Continuous Performance Test.* This test consists in rapidly presenting a set of letters with varied spacing and timing of target and nontarget stimuli. The analysis of response features is important in the interpretation of results and has led to the finding that patients with schizophrenia have a deficit in their ability to distinguish target stimuli from nontarget stimuli, especially when the stimuli are presented as brief signals at a rapid pace. When compared to nonclinical controls, positron emission tomography (PET) in patients performing a Continuous Performance Test revealed lower metabolic activity in the prefrontal cortex bilaterally but normal or elevated activation in the occipital region. This finding is consistent in supporting the idea of impaired executive functioning in schizophrenia.

Attention-Deficit/Hyperactivity Disorder

Peterson and Posner identified three neural systems critical for attention: These are alerting, orienting, and dual executive networks. Attention is initiated with alertness, which elicits the reticular activating system in the brain stem maintained by adrenergic projections from locus coeruleus in the pons of the brain stem to the forebrain. What follows is the necessity to orient towards a stimulus, which induces activity in two networks, the dorsal and ventral. The dorsal network includes the parietal and frontal eye fields and is involved with synchronizing visuospatial orienting activity. The ventral network houses the temporoparietal junction and ventral frontal cortex, which selects which competing input should be given priority. Given the range of potential surprises, these networks allow us to focus on anomalous activity that may be a threat or of potential benefit to our survival. The dual executive network includes the DLPFC-parietal network involved with task initiation, switching, and adjustments and the ACC-anterior insula network, which assists with sustaining attention. Individuals with ADHD tend to demonstrate deficits in the alerting and executive attention networks.

Data from numerous studies show that patients with ADHD have dysfunctions on a variety of tasks, including those that involve monitoring, perception, memory, and motor control. Intact performance has been found on a number of memory tasks requiring verbal processes (digit span, word tests, and story recall) and nonverbal processes (recall, visual arrays, and block series).

In general, patients with ADHD evidence behavior that is consistent with patients suffering from frontal lobe damage, specifically with regard to deficits in the control of motor responses, in the execution of fine-motor movements, and in the inhibition of ongoing response patterns. Although memory tasks and basic aspects of information processing remain intact, these patients have impaired performance across several modalities (auditory, visual, motor, and perceptual–motor), suggesting some global deficit. These patients also appear to be unusually susceptible to boredom when the required tasks are long and repetitive.

One view examines two proposed systems, an underactive behavioral inhibition system, and an impaired behavioral reward system, to explain the behavioral challenges of ADHD. A related view is that the rule-governed behavioral system is not intact, as evidenced by the fact that patients with ADHD appear to do especially poorly with delayed or nonexistent rewards, including those tasks that require sustained attention, accuracy, or task-directed activity governed by another person's direction or rules. Under these conditions, the patient's poor regulation and inability to meet functional demands are revealed. A related issue is a diminished motivational drive and, possibly, a diminished arousal regulation system.

Another perspective supported by research data is that ADHD patients have metacognitive differences. According to this view, the metacognitive processes that help to plan, monitor, and regulate performance are impaired, and the patient's ability to assess the task and determine strategies is deficient. This is an example of a top-down aspect of attention because there is impairment of the higher cognitive processes that regulate information flow. Bottom-up deficiencies, on the other hand, would involve impairment in basic aspects of attentional focus due to abnormalities in arousal, capacity, and selectivity.

The finding that numerous factors can influence the appearance of deficits led Virginia Douglas to hypothesize that ADHD is a self-regulatory disorder with pervasive effects. According to Douglas, the impairments affect each of four domains—attention, inhibition, reinforcement, and arousal—and result in deficits in several aspects of self-regulation: (1) the organization of information processing, including the making of plans, metacognition, executive functions, the adoption of appropriate cognitive sets for a given task, the regulation of arousal levels and alertness, and self-monitoring and self-correction; (2) the mobilization, deployment, and maintenance of adequate attention; and (3) the inhibition of inappropriate responses to extraneous stimuli and reinforcers. These deficits in self-regulation imply that increased processing demands result in the diffusion of attention and the impairment of the in-depth, coherent acquisition of knowledge and understanding.

The diversity of research findings and theoretical explanations is paralleled by the clinical finding that children who are severely impaired in the classroom may have no attention problems in the confined, one-on-one setting of the psychiatrist's or psychoeducational examiner's office. Such children may also be able to attend for indefinite periods to video games and yet be unable to follow complex conceptual information. This is an important reminder that cognitive dysfunction in psychopathologic conditions may be task, context, or relationship specific, depending on the nature of the cognitive impairment. The patient's clinical history and evaluation must therefore consider potentially hidden domains of atypical cognition.

Autistic Spectrum Disorder

Although early descriptions of autism delineated deficits in social functioning, later studies found cognitive dysfunctions involving abstraction, sequencing, language, and comprehension. Researchers are now focusing on the nature of the core deficit in the disorder, and how the cognitive domains relate to the social-affective deficits is of particular interest. However, there may be many "autisms," and finding a single process may prove unhelpful. In general, three broad areas appear to be atypical for those on the autism spectrum: sensory integration, emotion regulation, and social cognition.

A majority of patients with ASD are assessed as having mental or cognitive deficiencies, although cognitive impairments are especially difficult to assess if language functioning is severely limited. Mental retardation involves global, cognitive, and language impairments that may make it difficult to distinguish autistic disorder features. Studies of high-functioning patients with autistic disorder have permitted a deeper exploration of a variety of cognitive deficits. An array of dysfunctional areas, including numerous language problems, excessive or impaired responsiveness to stimuli of various modalities, different encoding of auditory stimuli, and impairment in the ability to extract important features from incoming information, have all been found; in contrast, some patients with autistic disorder have relatively intact visuospatial and Gestalt functions, musical abilities, and rote memory. Performances on standardized tests reveal relatively good results in object assembly and block design but poor results in comprehension. Language deficits vary and include syntactic, phonologic, prosodic, and pragmatic domains, also suggesting right-hemisphere involvement. The wide array of dysfunctions in autistic disorder may be due to the existence of a variety of subtypes with dysfunction in different loci in the brain. One general finding is that of a larger brain in individuals with autism, with some suggestions that differentiation has become impeded, thus impairing the capacity for integration and the achievement of complex states of cognition and behavioral social interactions.

Other studies have focused on the social cognition of patients with ASD; they have examined the nature of the patient's emotional behavior and level of understanding in order to assess the earliest manifestations of an atypical emotional connection between autistic children and their parents. Recent work in neurobiology suggests a role of the orbitofrontal cortex and the cerebellum in mediating some of these deficits. Two findings support these initial impressions: (1) Autistic children are much less likely than nonautistic children to imitate adult vocalizations and gestures, and (2) they show much less sophisticated representational play with objects than do nonautistic children. These findings have led to the suggestion that a core deficit in ASD exists in the representation of representations (meta-representations), leading to deficits in symbolic play and to the inability to understand the mental states of others. The inability to transfer cognitive representations into language symbols may be a related meta-representational deficit.

A series of studies explored the relation of these possible cognitive impairments to socioemotional behavior. In contrast to clinical lore, children with ASD were found to look at their parents; they had eye contact with their parents when social interactions were parentally elicited, and they revealed a typical distribution of behavioral patterns of attachment, even if the behaviors were not typical themselves. These studies did find a marked lack of social referencing (looking to parents for emotional cues in ambiguous situations) and protodeclarative gestures (pointing to objects and showing objects to familiar adults); three hypotheses have been proposed to explain these findings: (1) Autistic children may not have the capacity to have a representation of another person as having ideas, perspectives, or emotions that can be shared. This proposal is consistent with a “theory-of-mind” hypothesis, in which a core deficit in autism is believed to be the inability to have a sense of another’s mind; (2) autistic patients may have an impaired ability to perceive or to comprehend the emotional (usually facial) signals of others; and (3) the core deficit may involve a lack of interest in others or an aversion to responding to others. Studies have found that although children with autistic disorder do express emotions, they have less positive affect during periods of joint attention with another person. Furthermore, they reveal impairments in their responsivity to the display of strong emotion by another, whether it is of distress or of pleasure.

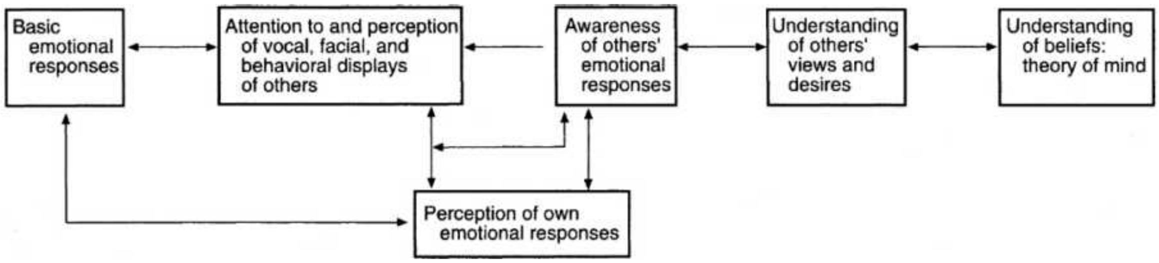
Basic emotional responses
Attention to and perception of vocal, facial, and behavioral displays of others
Awareness of others' emotional responses
Understanding of others' views and desires
Understanding of beliefs: theory of mind
Perception of own emotional responses

FIGURE 3.1–2. Sigman's model for the development of socioemotional understanding. (Courtesy of M. Sigman.)

Tests of high-functioning autistic patients reveal poor performance on emotion-recognition tasks, little comprehension of and empathy with depictions of social situations, and difficulty in talking about socially derived emotions, such as pride and embarrassment. Autistic patients with relatively high intelligence use adaptive cognitive strategies to interpret social stimuli to compensate for impaired emotion-processing abilities. The development of social cognition and socioemotional understanding requires complex interactions among the cognitive, perceptual, and emotional processes.

A series of interactive elements essential to the development of social understanding has been proposed (Fig. 3.1–2) to describe the basic precursors of emotional responsiveness: The ability to attend to, encode and interpret verbal and nonverbal social stimuli, the awareness of one's own and others' emotional responses, and the ability to contrast oneself with others. Out of that matrix develops the ability to understand others' views, desires, and beliefs. Accordingly, a deficit in any one of those basic elements may explain the characteristic deficits observed in the social cognition of persons with ASD. As discussed in the section on Social Cognition, one hypothesis in need of further validation is whether an atypical functioning of mirror neurons and related circuits plays a role in the social deficits of autistic individuals.

Mood Disorders

In contrast to schizophrenia, ADHD, and autistic disorder, mood disorders do not appear to have core cognitive deficits that are diagnosis specific. Instead, cognitive abnormalities appear to be related to the degree of psychopathology and the severity of the mood disturbance. In other words, they are state-dependent. Most studies have examined patients with depression, whereas only a few studies have assessed cognitive functioning in patients with bipolar I disorder during a manic state. In depressed patients, the severity of the depression has ranged from mild depression in students to severe illness in hospitalized patients with major depressive disorders. The studies have primarily examined attention and memory for neutral and emotionally toned stimuli; for the majority of these studies, the concept of self-organizational processes and state regulation has not been the primary focus of attention. These recent conceptualizations of the brain's functioning as a nonlinear complex system capable of self-organization, however, may aid in the future investigation of the primary dysregulatory aspects of mood disorders.

Depressive Disorders. Depressed patients often complain of difficulties with concentrating, learning, and remembering. Studies have documented that such patients perform poorly on tasks that require sustained attention or effortful and elaborate rehearsal, and thus, controlled limited-capacity processes appear to be impaired in major depressive disorder. This limitation of access to capacity-demanding resources appears to be directly related to the severity of the depression and normalizes with remission from a depressive episode.

Depressed patients have also been found to require increased supportive data from the presented stimuli before responding in a test situation. Whether this pause before responding is a psychological response to being depressed, an emotional "need to be certain," or a specific feature of the depression itself is yet to be determined. The reluctance to respond must be considered in interpreting research data and clinical interview findings.

The attention and memory findings have suggested several theoretical frameworks that are clinically useful: A schema theory of depression outlines a positive feedback loop in which negative self-schemata prime persons to have negative thoughts, to recall negative events in their lives, and to interpret present events with a negative bias. Whether as a cause or as a maintaining influence, these depressogenic schemata are thought to create a series of cognitive functions that produce and maintain a depressed mood.

A network theory of memory and emotion, supported by research on depressed and nondepressed persons, posits that mood leads to a spreading activation of items in memory that are congruent with the mood. Thus, emotion directly influences retrieval by the process of state-dependent learning and memory. Furthermore, while depressed, patients are likely to encode items in a form that makes them readily accessible when retrieved in a depressed state. Depressed mood, then, becomes an internal context cue that accesses depression-related memories.

The network and schemata theories of depression provide a framework for understanding how emotions and moods influence cognitive processing. Emotions in health and illness can shape the mental state instantiated at a given time by triggering the activation of previously formed depressive schemata. The activated schemata, in turn, can produce retrieval biasing and behavioral responses fundamental to a mental state, which can then further elicit a negative emotional response; this self-reinforcing loop is established and supports the continuation of both depressed cognition and mood.

Some patients may be especially prone to marked cognitive alterations based on rapid shifts in mood, and these shifts may be a learned or constitutional feature of the individual. From the nonlinear dynamics view, rapid shifts in mental state can be conceptualized as sudden and intense changes in the constraints on the system, which determine the flow of states of mind across time. The depressed state may become deeply ingrained as an attractor state that is difficult to alter. Pharmacologic interventions may aid in directly altering the synaptic constraints that maintain such a state, thereby promoting neural plasticity. Psychotherapeutic interventions, such as cognitive-behavioral, interpersonal, and mindfulness-based therapies, can all be theorized to produce changes in the internal processing of constraints, as well as in the ways in which external constraints from the social world interact to shape the overall shifts and stability of mood.

Bipolar I Disorder. Patients with manic episodes present with a spectrum of state-related cognitive dysfunctions tied to the severity of their episode. Manic disturbances have been described clinically as rapidly paced thinking and speech, quick associations to self-generated or other-generated stimuli, grandiosity, and increased distractibility.

Studies have suggested a high rate of combinatory thinking and the inclusion of loosely associated but related intrusions. The clinical impression of humor (even in the face of underlying dysphoria) in some patients is corroborated by playful, extravagant, or flippant elaborations, as well as intrusions in speech, and seems to indicate primarily state-dependent symptoms that improve on recovery. Forced-choice, span-of-apprehension, and backward masking tasks all reveal impairments similar to those seen in actively psychotic patients with schizophrenia and yet are unlike those seen in schizophrenic patients who have persistent deficits in the remitted state. Bipolar disorders, then, can also be viewed as disorders of self-organization in which the system fluctuates between the extremes of highly activated manic states and highly deactivated depressed states of mind. Structural brain imaging of patients with both disorders has revealed impairments to integration in the form of impediments to linkage.

Two networks are involved in the pathology of bipolar disorder, one involving the patient's homeostatic network, which includes internally directed amygdala-based feeling states, and the second network involving the external regulatory network assisting with cognitive emotion regulation while suppressing maladaptive affect. Compared to healthy individuals, where these two networks would normally regulate the emotional responses collaboratively, in those diagnosed with bipolar disorder, these networks fail at regulating emotional responses with the complexity and flexibility of healthy individuals. Research-based on a better understanding of neural origins of complex emotional symptoms of bipolar disorder is still needed to better understand the overall link between physical symptoms and emotional responsiveness.

Anxiety Disorders

Cognitive studies of patients with anxiety disorders have focused primarily on attention or memory. Various research paradigms have been applied to assess attention bias and memory retrieval for neutral and emotionally activating stimuli. Studies find that anxious patients have an increased tendency to attend to fear- and threat-related words. One research approach includes a dichotic listening task in which an anxious patient is more easily distracted than are controls by fear-related stimuli in the nonattended channel. This finding suggests that anxious patients have automatic, parallel attention processes that are primed to detect certain types of stimuli.

Another approach uses the Stroop paradigm, in which words are presented in different-colored inks, and the person must look at the word and state the color of the ink. Anxious patients have a significant delay in their response times for fear-related words, a finding that suggests that increased attention capacity or cognitive processing is necessary when those “fear” words are perceived, and the color of the ink must subsequently be determined.

One theory that explains these aforementioned findings posits the idea of a “fear network” that encodes fear-related information in a memory structure that is readily accessible and able to influence cognitive, motor, and psychophysiologic responses. The theoretical fear networks, which may be similar to mental models, are thought to contain three related forms of information: (1) fear-eliciting stimulus cues, (2) specific response patterns, and (3) the meaning of the cues for that particular person. According to this theory, patients with anxiety disorders are believed to have fear networks that are especially coherent and stable, easily activated, and require few environmental cues.

Posttraumatic Stress Disorder

Patients with PTSD have been found to have attentional biases toward threat-related stimuli specific to the experienced traumatic event. Cognitive aspects of the disorder are characterized by intrusive processes (memories, images, emotions, and thoughts) and avoidance elements (numbing, amnesia, and withdrawal), which are driven by increased arousal. These patients are thought to have a unique configuration of fear networks containing stimulus cues (environmental stimuli), response components (cognitive, motoric, and psychophysiologic), and meaning elements (the moral implications of the trauma, survivor guilt, and the meaning of intentional trauma vs. accidental trauma).

Some theories argue that states of excessive arousal during trauma impair attention capacity and memory encoding during that event. Emotional processing during and after a traumatic experience may also be hampered by the heightened state of psychophysiologic arousal and altered memory storage. In fact, the part of the brain necessary for explicit memory processing, the hippocampus, has been shown to be smaller in individuals experiencing chronic PTSD. Patients who have used dissociative mechanisms during and after a traumatic experience appear to be at far greater risk for developing PTSD.

One line of research that has expanded upon PTSD has been the examination of neural networks such as The Default Mode Network (DMN), the Salience Network (SN), and the Central Executive Network (CEN). The DMN, a network related to internal cognition such as mental time travel, the experience of self and others, has been found to be one of the first networks to be disrupted upon the experience of a traumatic event. Studies have found that the extent of these disruptions may be predictive of the consequent diagnosis and severity of PTSD. With the diagnosis of PTSD, the flexibility with which these networks operate in a nonclinical population has been found to be disrupted. During resting state, there has been evidence of intrusions from the SN, which typically allows one to toggle between the DMN and the CEN. The CEN is operative when an individual is engaged in a goal-directed task. In a healthy population, the DMN and CEN are mutually inhibitory. In PTSD, the CEN exhibits intrusions from the DMN. Put simply, intrinsic reflection may be characterized by a lack of safety resulting in hypervigilance, whereas internal cognition intrudes upon goal-directed tasks. Some of the underlying cognitive deficits found in PTSD are likely attributable to these dynamic challenges.

PTSD may have some aspects that are adaptive: Attention biases primed to detect fear-related stimuli may permit the early detection of threatening situations that, if not avoided, would produce incapacitating psychophysiologic arousal. Automatic, nonconscious behavioral avoidance response patterns embedded in the proposed fear networks or mental models allow these patients to minimize excessive arousal by avoiding trauma-related situations.

The psychopathologic aspects of PTSD can be cognitively exemplified as follows: A combat veteran with chronic PTSD may have no direct recall (impaired explicit memory) of a helicopter crash in which his best friend, who was seated next to him, was killed. Years later, continued avoidance of airports, amnesia for combat, general apathy, and social withdrawal (avoidance elements), combined with an exaggerated startle response, panic attacks, intrusive images, and nightmares (intrusive components), all suggest intact implicit memory of the combat trauma; the veteran has become emotionally, behaviorally, and cognitively impaired.

Speechless Terror. Scott L. Rauch and colleagues explored the neurobiology of intense fear using patients with PTSD. They took eight patients with PTSD and exposed them to two audiotapes. One tape was emotionally neutral, and the other was a script of a traumatic experience. While patients listened to these tapes, measures of their heart rate and regional cerebral blood flow (rCBF) were measured via PET scans.

Studies found that, during the play of traumatic audiotapes, rCBF was greater in right-sided structures, including the amygdala, posterior medial, orbitofrontal, insular, anterior, and medial temporal lobes, and anterior cingulate cortex. These are the areas thought to be involved with intense emotion. Such findings are consistent with the view of unresolved PTSD as a state of impaired integration prone to the chaotic (intrusive) and rigid (avoidant) symptoms. In developmental trauma, too, those who have experienced abuse or neglect have been found to have structural differences in the brain's integrative regions, including the hippocampus, prefrontal cortex, and corpus callosum, as described below. Psychotherapeutic intervention in posttraumatic patients may yield significant healing by way of promoting the growth of integrative functioning, which future research may reveal as including the growth of integrative structures in the brain of those who resolve their previously unresolved states.

An extremely interesting and potentially important clinical finding was the decrease of rCBF in the Broca area (left inferior frontal and middle temporal cortex) in states of high arousal. These findings suggest an active inhibition of expressive language during trauma and high levels of stress. Based on these results, speechless terror, often reported by victims of trauma, may have neurobiologic correlates consistent with what is known about brain-behavior relationships. This inhibitory effect on the Broca area also impairs the encoding of conscious memory for traumatic events at the time that they occur. It then naturally interferes with the development of narratives that later serve to process the experience and lead to neural network integration and psychological healing. Activating the Broca area and left cortical networks may thus be essential in psychotherapy with patients experiencing PTSD and other anxiety-based disorders.

Approaches to the treatment of patients with PTSD need careful evaluation but generally include the view that the impaired emotional processing of the traumatic event requires the active recollection, in explicit terms, of the details of the experience. The process of effectively treating unresolved trauma usually involves the active cognitive processing of specific memories, including emotional responses, derived belief systems, and psychophysiologic arousal at the time of the traumatic event and during recall.

The provision of new cognitive information in the course of psychotherapy, as well as the relative safety of the therapist's office, can, in theory, widen the "windows of tolerance" for certain emotions and memories to be held flexibly in consciousness and to alter the configuration of the fear networks and allow previously inaccessible information to be explicitly processed and made available to consciousness for incorporation into memory and an ongoing autobiographical narrative. Specific techniques, such as those that facilitate cognitive processing of previously dissociated or, in other ways, isolated cognitions—such as beliefs, images, and sensations—may be useful in alleviating the symptoms and dysfunction after acute or chronic trauma. Such therapy and changes in mental processing may reduce the avoidant and intrusive components of the clinical syndrome of PTSD, and they also may improve social, emotional, and cognitive functioning.

The evaluation and treatment of PTSD patients can be greatly enhanced by an understanding of the relevant cognitive processes and a careful application of the assessment to presenting symptoms. Several related areas that have interested clinicians for decades, and which have become a source of great societal concern and controversy, include the delayed recall of repressed memories of traumatic events and the suggestibility of patients influenced by clinicians, society, or friends to believe that they are the victims of childhood trauma.

Delayed Recall. Many scientists and clinicians believe in the cognitive capacity of patients to be unaware, for years or even decades, of severely traumatic experiences that took place in their childhoods. Other researchers disagree and emphasize the paucity of studies of corroborated cases of childhood trauma that have been followed prospectively into adulthood and have documented impaired access of consciousness to events presumably stored in memory. Two distinct mechanisms that may explain delayed recall of childhood trauma are *repressed* memories and *dissociated* memories.

A repressed memory can be thought of as originating from the active, intentional suppression of memory from consciousness. The mechanisms that underlie this process may then become automatic, and the contents of memory may become inhibited from retrieval into consciousness. This blockage may exist to avoid flooding the person's awareness with information that is associated with excessive anxiety or fear that would impair effective functioning. This is an example of knowledge isolation in which information may be layered in the nervous system, and certain aspects may be kept from conscious awareness.

In contrast, a traumatic event may be so overwhelming that adaptive processing may be impaired. If focal attention is divided or if the hippocampus is shut down temporarily with high cortisol secretion, the nonfocally attended (traumatic) material will only be processed implicitly. Thus, to adapt to a traumatic event, some persons may have the capacity to focus their attention on a nonthreatening aspect of the environment or on their imagination during the trauma; this may be an underlying mechanism in a process called *dissociation.* Traumatic memory that has only been implicitly encoded affects behavior and emotions and possibly contains intrusive images and bodily sensations that are devoid of a sense of past, of self, or of something being recalled. This may partly explain such symptoms of PTSD as amnesia (blocked conscious access to a memory or its origin), avoidance behaviors, hyperarousal, intrusive images, and flashbacks.

A person may use both mechanisms of repression or dissociation for different aspects of a traumatic event, and recollection of the traumatic memory may take different forms. Repressed memories may have been processed to some degree in narrative form, whereas dissociated memories probably lack integrative processing. This latter form may thus be experienced as nonpast and nonself, making the intrusive retrieval of dissociated memories a confusing and frightening experience. This is implicit memory retrieval of traumatic events.

Studies of the development of memory in children suggest that the shared construction of narratives about experienced events is often crucial in making the events accessible to long-term retrieval. The establishment of a personal memory system appears to be a function of such “memory talk,” in which parents discuss with children the contents of their memory. In children who have been forced to keep traumatic events a secret, as may occur with childhood abuse, the important developmental process of narration may be blocked. This may be an additional cognitive mechanism underlying the inaccessibility of some forms of childhood trauma to consciousness in adult patients.

Suggestibility. Numerous studies have demonstrated that human thought and memory are easily influenced by suggestion from others.

Suggestibility is adaptive for a social being who relies on the experiences of others for knowledge of the world. Therefore, listening to the stories of others, reading a textbook, and being coached in athletics all require the receiver to accept data from the sender. Critical analysis of the data received is also an important component of learning; however, the metacognitive function of assessing the accuracy or usefulness of newly acquired information may be suspended under certain conditions, including hypnosis, drug-altered states, and conditions of severe threat.

Studies of human suggestibility indicate that postevent questioning can bias the metamemory processes that help to determine the source and accuracy of a retrieved memory. The verbal and nonverbal cues given by the interviewer may influence a person to believe that actual aspects of an event never took place. A person can also be convinced of the accuracy of an event despite its lack of correspondence with actual experience. Factors that may influence the biasing of interviewees include a belief in the trustworthiness and authority of the interviewer, not being aware that "I don't know" is a permissible response, repetition of a question that has already been answered, and the interviewer's beliefs as communicated through emotional tone and nonverbal gestures.

It is crucial for clinicians to be aware of human suggestibility to avoid iatrogenic distortions. Similarly, it is important for persons who experienced severe trauma early in life to receive informed and empathic evaluations and treatment. There is a delicate balance between supportive neutrality and active advocacy in assessment and intervention; awareness of these fundamental cognitive processes may help to guide the clinician toward achieving this goal.

Complex PTSD. Complex PTSD results from prolonged and inescapable stress and trauma during early childhood. It is complex because of its extensive impact on all areas of development and functioning. Trauma-shaped personality traits and coping strategies tend to increase the individual's vulnerability to future trauma through engagement in abusive relationships, cognitive distortions, poor judgment, and a lack of self-protection. Long-term PTSD has been shown to correlate with the presence of *neurologic soft signs,* suggesting either a vulnerability to the development of PTSD or the long-term impact of the physiologic dysregulation caused by PTSD.

Research on individuals with documented early developmental trauma in the form of severe neglect or abuse (emotional, physical, sexual) has revealed that the regions of the brain that are impaired in their growth are those that link widely separated areas to each other. These integrative regions are affected differentially depending on the timing of the trauma and include the hippocampus, which links various forms of memory into explicit factual and autobiographical memory, the corpus callosum linking left and right sides of the brain, and the prefrontal cortex, which links the cortex, limbic region, brainstem, bodily signals, and the social input from other brains into a coherent whole. In many ways, developmental trauma can be thought of as a severe example of impaired interpersonal integration. With neglect, there is an impairment to linkage; with abuse, there is an impairment in honoring differential needs of child and adult.

This clinical finding of developmental trauma leading to impaired neural integration supports the following possibility: Interpersonal integration (honoring differences and promoting compassionate linkages with other people) stimulates the growth of neural integration (the linkage of differentiated regions of the brain to each other). If this proposal is true, it may mean that integrative psychotherapy techniques may promote a healing process in those with complex PTSD that helps stimulate the activation and growth of neural integration. Future research will need to validate this hypothesis of a potential core mechanism in how therapeutic relationships work to improve the lives of those who have experienced severe forms of trauma early in development.

Why would developmental experiences that overwhelm an individual’s ability to cope have such a profound impact on their growth? When confronted with a threat under usual adult circumstances, the processes related to arousal and the fight-or-flight-or-freeze-or-faint reactivity states become activated; the threat is dealt with and soon passes. For obvious reasons, children are not well equipped to cope with a threat in this way. Because their survival is linked to dependency, fighting and fleeing with an attachment figure may actually decrease their chances for survival. When a child first cries for help, but no help arrives, or when trauma is being inflicted by a caretaker, the child may shift from hyperarousal to dissociation. Traumatized children who are agitated may be misdiagnosed as having an attention-deficit disorder, and the numbing response in infants can be misinterpreted as a lack of sensitivity to pain.

Dissociation allows the traumatized individual to escape the trauma via a number of biologic and psychological processes. Increased levels of endogenous opioids, for example, create a sense of well-being, pain reduction, and a decrease in explicit processing of the overwhelming traumatic situation. Psychological processes such as derealization and depersonalization allow the victim to avoid the reality of his or her situation or to watch it as an observer. These processes provide the experience of leaving the body within awareness, traveling to other worlds, or immersing oneself in other objects in the environment. Hyperarousal and dissociation in childhood create an inner biopsychological environment primed to establish inflexible boundaries between different emotional states and experiences for a lifetime. If it is too painful to experience the world from inside one’s body, self-identity can become organized outside of the physical self.

Early traumatic experiences shape neurophysiology and neuroanatomical networking, thus impacting experience and adaptation throughout development. The tendency to dissociate and disconnect various tracks of processing creates a bias toward unintegrated information processing across conscious awareness, sensation, affect, and behavior. General dissociative defenses result in an aberrant organization of networks of memory, fear, and the social brain and contribute to deficits of affect regulation, attachment, and executive functioning. One way of understanding such fragmentation of mental processes is in the simultaneous activation of two mutually incompatible motivational drives, mediated by two differentiated circuitries. One is the mammalian limbically mediated drive to move toward the caregiver when the infant or child is in a state of distress. The other is the brainstem-mediated drive to move away from the source of terror.

If the attachment figure is the source of terror, then the two circuits driving the body to simultaneously move both toward and away from the same individual create a biologic and psychological dilemma. This has been called “fear without solution” and is a biologic paradox thought to be the mediating mechanism of disorganized attachment common in developmental trauma and often associated with clinical degrees of dissociation. Without a solution, the organized strategy to cope with stress is overwhelmed, and the mind of the child cannot achieve integration as the differentiated circuitries are unable to become linked in a coherent way to deal with the abuse or neglect. Disorganized attachment experiences may be one common developmental origin of clinical dissociation and complex PTSD.

The lack of coherent integration of the interdependent systems within an individual, which may be due to extreme early stress, results in many disorders: Compulsive disorders related to eating or gambling, for example, and somatization disorders in which emotions are converted into physical symptoms can all be understood in this way. Some forms of borderline personality disorder and conversion disorder may also reflect complex adaptation to early trauma, as does dissociative identity disorder.

Cognition in Racial Prejudice

The modern notion of the brain as a human-computer is both alluring and misleading. Unlike computers, human brains have been programmed over millions of years to process the environment in ways that enhance human survival. Adaptations that enhance survival trump accuracy nearly every time. In Ortega y Gasset’s words, “*We do not live to think; we think in order that we may succeed in surviving.*” Because of this evolutionary strategy, cognition is as much distortion as it is representation. One of the consequences of this distortion is racial prejudice.

The ubiquity of implicit bias, prejudice, and racism seems to suggest that humans are hardwired to dislike, distrust, and even hate those who are different. The fact that conflicts and slaughter based on race, tribal affiliations, and religion continue unabated in the “modern” world only seems to strengthen the argument for prejudice as a basic component of human nature. Standing in stark contrast to this view are countless examples of harmony, cooperation, and acts of altruism that seem to contradict racism as a basic human truth.

When we imagine the lives of man’s tribal ancestors, it is easy to see the survival benefit of developing a reflexive system for the classification of unfamiliar others that trigger fear circuitry and a fight–flight response. It would then follow that these primitive reactions would lead one to distortions of sensory perceptions and social judgments that parallel those seen in social phobia. These distorted attitudes and perceptions would then become woven into the culture, guiding everything from interpersonal behavior to public policy to international relations. Therefore, the most relevant questions facing psychiatrists in this area of study are how do human brains process social information, and how can one counteract those perceptions that are incorrect and harmful to others?

In an attempt to identify the neural substrate of prejudice, a number of studies have examined brain activation during exposure to pictures of same/other group faces. It turns out that the old statement “they all look alike,” where “they” are another racial group, may reflect an underlying truth of cognitive processing. Across studies, subjects demonstrate a clear advantage in recognizing faces and distinguishing among faces for same-race faces. Neural processing for the recognition of other-race faces results in longer periods in frontal activation, suggesting that they are processed less automatically and with greater effort.

Supporting this idea is the fact that interracial interactions have been those that show impaired executive functions when confronted with people of other races. It is possible that the extra frontal lobe activation required for facial processing may divert energies and attention away from other frontal functions. Consistent with these findings are those that show that stereotypes are automatically activated and require conscious inhibition to produce low-prejudice responses. The overall result of these processing differences may lead one to think about individuals of other races in less complicated and more stereotyped ways.

Assuming that fear is at the heart of prejudice, a second line of research directly explored amygdala activation in same versus other race encounters. It was found that when Caucasians and African Americans are shown pictures of other-race faces, they exhibit more amygdala activation and an enhanced startle response compared to same-race faces. Although it takes approximately 500 ms for a visual stimulus to reach conscious awareness, information about race is encoded during the earliest stages of visual processing, well before conscious perceptions of who is being seen. The rapidity of this “race response” results in the automatic and instantaneous frontloading of all past biases into the experience of each new person.

This activation of a rapid and unconscious fear response influences the conscious experience in many ways. A finding that directly parallels the experience of social phobics. For example, individuals from other groups are more readily associated with aversive stimuli. This is the likely reason why African Americans are more readily misperceived as carrying weapons, and black faces are more likely to be perceived as angry by white subjects. Interestingly, both African Americans and Caucasians have greater amygdala activation to African-American faces, which may be due to the consistent pairing of African-American faces with criminal behavior in the media. This conditioning may not have as extreme of an impact on African-American viewers because of their greater exposure to other African-American faces. African Americans are also better at recognizing Caucasian faces than vice versa for this same reason. Thus, while one's responses to people of other races are guided by bottom-up amygdala activation, they are also modulated by a variety of top-down cognitive and social factors.

Greater liberalism has been associated with greater gray matter volume in the anterior cingulate, while greater conservatism is associated with increased volume in the right amygdala. Conservatives demonstrate more amygdala activation and persistent cognitive styles compared to liberals, who demonstrate more sensitivity for cues to alternate responses. Along with these differences in cognitive style, greater liberalism was associated with stronger conflict-related anterior cingulate activity reflecting their continued engagement with the complexity of the problems posed to them.

To explore the issue of whether prejudice is learned or innate, faces of well-known and well-liked African Americans were shown to Caucasian subjects who demonstrated an absence of amygdala activation. These findings strongly suggest that exposure and positive experiences can counteract the effects of general cultural prejudice. Even simple exposure to other-race faces increases liking for strangers of that race, and an experience of empathy for one member generalizes to other members of that group. It is also known that through exposure and paying conscious attention to facial details, Caucasian adults can be successfully trained to distinguish among African-American faces and that this learning process results in a decrease of implicit racial bias.

Once stereotypical attitudes and beliefs are inculcated, change requires "intention, attention, and time." Because most children are exposed to some degree of racial prejudice, open discussion and increased interracial exposure work against prejudice, turning into conscious beliefs and negative behaviors. Increased focus and awareness of the handicapped have also been found to decrease discrimination in the same way. It is known from studies of anxiety and fear that conscious labeling activates cortical neural networks, which can regulate and inhibit emotional activity.

People from other cultures, with different degrees of exposure, proximity, and prejudice need to be studied with the same depth that one has studied their own society before one can arrive at any conclusions about the origins of racial prejudice in the human brain. With the science of genetics revealing that the notion of race is more of a cultural than a biologic construct, future study of how human brains react to one another can only help shed light not just on the nature and etiology of the neuroscience of prejudice but on how everyone who lives on this increasingly crowded planet may learn to overcome their differences and discover the true wealth of human diversity.

Cognition in Social Phobia

It is clear that those that struggle with social phobia do not perceive the same interpersonal landscape as others. Social phobia disrupts and distorts the ability to gather and process social information with attentional and memory processes biased toward detecting, orienting toward, exaggerating, and remembering negative responses by others. Once focused on negative input, they fixate on it and find it difficult to disengage, even when it is irrelevant or harmful to their well-being.

Social phobia is most likely the result of an imbalance of the alarm and feedback systems that regulate the balance of approach and avoidance in social situations. At the same time, social phobics show a decreased activation in left-hemisphere regions involved in other aspects of facial evaluation and reality testing. In essence, their brains react to others as if they are in extreme danger no matter what the situation. Fear of direct eye gaze is an evolutionary conserved submissive behavior that is woven into all human cultures but exaggerated in those with social phobia. Social phobics employ visual gaze avoidance and elaborate processing strategies to modulate the fear evoked by being seen.

Children with social phobia experience greater anxiety while looking at faces, and as adults, their memory for faces, especially those with negative expressions, is far better than average. It makes sense, then, that social phobics avoid looking directly at faces, especially eyes, preferring to focus on other aspects of those with whom they interact. Turning away from others has a number of negative consequences. First, it keeps them from viewing potentially positive expressions. Second, it keeps these children from learning to process facial expressions in a more sophisticated manner, especially during development. Finally, turning away from others correlates with using their negative self-image when developing a theory of what is on another’s mind. This may be why social phobics have been found to come to conclusions faster in social situations; recycling old information is faster than acquiring new data.

In social phobia, the amygdala shows decreased coupling with the frontal lobes and higher sustained activation in face processing networks. The combination of increased orientation to and fixation on social evaluation makes the experience of the social phobic, one of constantly being on trial: Every interaction is experienced as a negative evaluation of their worth. Thus, the memories and emotional impact of negative interpersonal responses cut deeper and last longer.

The most consistent findings have tied social phobia to hyperactivation of the amygdala. Subjects with social phobia show increased metabolism in the amygdala, fusiform gyrus, insula, anterior cingulate, and the prefrontal cortex when subject to direct eye gaze. Simultaneously, they demonstrate decreased cortical activation that could inhibit and regulate their subcortically generated fear.

Decoupling of cortical and subcortical processing is likely a major contributor to the disruption of current social information processing and the tendency to become overly reliant on past experience and negative internal self-images in interpreting the present. One group of researchers found that a combination of right prefrontal and left parietal activation was specific to social phobia. This led them to hypothesize that social phobia is related to a unique interaction between the processing of autobiographical memory and social presentation. Perhaps such neural wiring makes it difficult or impossible to overcome self-consciousness and the expectation of negative evaluation about the self.

More recent research has revealed a deactivation of the precuneus and posterior cingulate in these individuals suggestive of inhibition of the default mode processing that is used to process information about oneself and others. Mindfulness- and cognitive-based interventions have proven helpful with social phobia because they simultaneously result in decreased amygdala activation and increased activation in brain regions involved in the deployment of attention.

While individuals who suffer from social phobia share the neurochemical profiles of other anxiety disorders (heightened adrenergic activity combined with decreased GABA activity), they appear to also have evidence of dysfunction in serotonin and dopamine systems. Hypersensitivity to negative social stimuli appears to be linked to variation in serotonin receptor pathways. Lower dopamine (D2) binding has been found in the striatum of social phobics, which in animals correlates with a lack of exploration and lower social status. Remember, dopamine is an important neurochemical in the social reward system related to approaching others, sustaining attachment, and a sense of well-being. As with so many other findings, it is not known if biologic differences are a cause or effect of social phobia.

Social phobia could reflect an innate inability to compete socially or an experience-dependent shutdown of activity in the face of anxiety and shame. When parents dominate children through physical or psychological punishment, some young brains may be shaped to be of persistently lower social rank. On the other hand, social phobia may be an evolutionary continuation of the behavior of lower-ranking individuals in mammal and primate groups who “know their place” and stay in it.

The finding that social rejection is mediated in the same part of the brain as physical pain—the dorsal anterior cingulate cortex—suggests that the fear of social engagement is tantamount to a fear of being injured. Some studies suggest that the experimental use of pain-relieving medications even temporarily reduces social phobia, pointing to this common pathway of somatic and social pain.

FUTURE DIRECTIONS

Clinicians must keep in mind that every client strives to make sense of the world, albeit through information-processing apparatus that are negatively impacted by the acute and chronic aspects of their disorder. Cognitive science offers an array of conceptualizations for understanding the ways in which the mind functions in health and disease. One broad approach to the mind is the approach we’ve been taking here called interpersonal neurobiology—a framework that synthesizes a wide array of research into one perspective. This broad, interdisciplinary field provides numerous research paradigms that are helpful in further elucidating the nature of psychopathology, from the techniques of the neurosciences to the biologic applications of chaos theory. It also provides ways of thinking that can be helpful in the consulting room in understanding the cognitive, emotional, and social impact of the ways in which the brains of maladaptive information processing. The cognitive understanding of emotions and consciousness may also expand psychiatry’s framework for understanding the subjective human experience. Clinical tools, from medications to in-depth psychotherapy, may also find wider and more finely tuned applications as the processes of self-organization and psychological change are better understood.

Psychiatry, in turn, also has much to offer the field of cognitive science. The long history of descriptive psychopathology and the attempt to synthesize views of the mind and the brain can provide nonclinical cognitive scientists with unique data and relevant questions. This may be especially true in the understanding of the social nature of the brain and mind and the impact of both positive and negative relationships on mental health and illness. Psychiatry is invited to join in the search for understanding the cognitive processes of the human mind.

These findings support the view that the mind is more than simply a synonym for brain activity. Our mental lives are fully embodied, and they are fully relational. While the brain in our heads is a core locale for how energy and information both flow and are regulated in our lives, the wholeness of the body proper and the relational systems—families, communities, cultures, and the natural world—in which we live each play a central role in our mental lives. Energy and information flow are not limited by skull or skin. Mind is not just inside the skull, nor limited to only inside the skin-encased body. We are, in our mental health and dysfunction, fully embodied and fully relational beings. When that flow is disrupted in achieving optimal self-organization by a blockage in its differentiation or linkage, the foundations of integration, then the chaos and rigidity of unhealthy conditions arise. And when differentiation and linkage, in the body and in our relational lives, can be cultivated, then we flourish and thrive individually and collectively, here on our shared home, this precious planet we’ve named Earth.

FURTHER READINGS

Baddeley A. Working memory: looking back and looking forward. *Nat Rev Neurosci*. 2003;4:829.

Cacioppo J, Visser P, Pickett C, eds. *Social Neuroscience: People Thinking about Thinking People*. MIT Press; 2006.

*Chan A. *Reassembling Models of Reality: Theory and Clinical Practice*. W. W. Norton; 2021.

*Cozolino LC. *The Neuroscience of Psychotherapy: Building and Rebuilding the Human Brain*. 2nd ed. W. W. Norton; 2010.

*Cozolino LC. *The Neuroscience of Human Relationships: Attachment and the Developing Social Brain*. 2nd ed. W. W. Norton; 2014.

Cozolino LC. *Why Therapy Works: Using Our Mind to Change Our Brains*. W. W. Norton; 2015.

Crick FC, Koch C. What Is the Function of the Claustrum? *Philos Trans R Soc Lond B Biol Sci*. 2005;360(1458):1271–1279.

Crick F, Koch C. Towards a neurobiological theory of consciousness. *Semin Neurosci*. 1990;2:263–275.

Damasio A. *Self Comes to Mind: Constructing the Conscious Brain*. Vintage Books; 2010.

Hoffman DD, Prakash C. Objects of consciousness. *Front Psychol*. 2014;5:577.

Kandel ER. A new intellectual framework for psychiatry. *Am J Psychiatry*. 1998;155:457.

Lane RD, Nadel L. *Cognitive Neuroscience of Emotion*. Oxford University Press; 2000.

Ledoux JE, Brown R. A higher-order theory of emotional consciousness. *Proc Natl Acad Sci U S A*. 2017;114(10):E2016–E2025.

McGilchrist I. *The Master and His Emissary: The Divided Brain and the Making of the Western World*. Yale University Press; 2009.

Northoff G. *The Spontaneous Brain from the Mind-Body to the World-Brain Problem*. The MIT Press; 2018.

Panksepp J. *The Foundations of Human Animal Emotions*. Oxford University Press; 1998.

Petersen SE, a Posner MI. The attention system of the human brain: 20 years after. *Annu Rev Neurosci*. 2012;35(1):73–89.

*Posner MI, ed. *Foundations of Cognitive Science*. MIT Press; 1989.

Raichle ME. Two views of brain function. *Trends Cogn Sci*. 2010;14(4):180–190.

*Schore AN. *Affect Dysregulation and the Damage to the Self*. W. W. Norton; 2002.

Seth AK, Friston KJ. Active interoceptive inference and the emotional brain. *Philos Trans R Soc Lond B Biol Sci*. 2016;371(1708):20160007.

*Siegel DJ. *Mindsight: The New Science of Personal Transformation*. Random House; 2010.

*Siegel, DJ. *IntraConnected: MWe (Me + We) as the integration of self, identity, and belonging*. W.W. Norton; 2023.

*Siegel DJ. *The Developing Mind: How Relationships and the Brain Interact to Shape Who We Are*. 3rd ed. Guilford Press; 2020.

*Siegel DJ. *Pocket Guide to Interpersonal Neurobiology: An Integrative Handbook of the Mind*. W.W. Norton; 2012.

Siegel DJ. *Mind: A Journey to the Heart of Being Human*. WW Norton; 2017.

Siegel DJ. *Aware: The science and practice of presence*. Tarcher/Penguin; 2018.

Tononi, G. Integrated information theory of consciousness: an updated account. *Arch Ital Biol*. 2012;150(2–3):56–90.

Taber K, Rausch S, Lanius R, Hurley RA. Functional magnetic resonance imaging: application to posttraumatic stress disorder. *J Neuropsychiatry Clin Neurosci*. 2003;15:125.

Van der Hart O, Nijenhuis ERS, Steele K. *The Haunted Self: Structural Dissociation and the Treatment of Chronic Traumatization*. W.W. Norton; 2006.

Wilson EO. *The Social Conquest of Earth*. Liveright Publishing; 2013.

3.2 Piaget and Cognitive Development: History, Theory, Extensions, and Modifications

Harrell Woodson, Ph.D.

Jean Piaget (1896 to 1980) was, at various points in his career, considered to be a natural scientist, psychologist, sociologist, and historian of the philosophy of science. His major enterprise was what he termed “genetic epistemology,” the study of the development of logic, reasoning, and higher-level thinking. He had points of contact with European psychiatry and even based his own research method in part upon the psychiatric interview. However, he avoided the emphasis on individual differences or interventional psychology, as practiced by clinical diagnosticians and therapists, and focused instead on the most universal and general characteristics of children’s cognitive development. His influence on education was considerable, not through a specific teaching method, but through his view of the child as an active constructor of his or her own knowledge rather than a passive recipient of information.

Piaget was born in 1896 in Neuchâtel, Switzerland. Growing up, he was influenced by social groups and unique opportunities that supported his interest in natural science, including the Swiss YMCA, the local Friends of Nature, and the director of the natural history museum, who allowed him to work with the specimens of mollusks, the topic of his 1918 doctoral thesis. For a year, he then worked in Zurich, combining laboratory research with clinical psychiatric interviewing supervised by Bleuler. He then moved to Paris to study at the Sorbonne and was recruited by Theodore Simon to work in Alfred Binet’s school-based laboratory, standardizing a reasoning test developed by Cyril Burt. In 1921 he began a long association with the Jean Jacques Rousseau Institute, with parallel appointments at varying times at the Universities of Geneva, Lausanne, and the Sorbonne. In 1955, with support from the Rockefeller Foundation, he founded the Center for Genetic Epistemology, where he continued his research and teaching with European and international colleagues, including prominent American psychologists.

Bennour and Voneche note several of the major intellectual and personal influences on Piaget. His interest in evolution stemmed from exposure to Henri Bergson's work, whereas his interest in psychiatry, psychology, and psychoanalysis stemmed from multiple sources. As mentioned above, Bleuler supervised him in clinical interviewing; Jung and others gave psychoanalytic lectures that he attended, and Simon mentored him in standardized psychometric assessment and influenced his view of cognition as adaptation rooted in biology. Janet was a strong influence on Piaget's thought in terms of his emphasis on psychology as a science of action and representation, with affect as a driver of both action and thought.

Piaget's contacts with American psychology and social science are fascinating, considering the prominence of experimental methodology and of behaviorism in the United States. In 1936, he was awarded an honorary doctorate from Harvard, but Hsueh has documented that the impetus for this came from the business school and sociology faculty rather than from the psychology faculty. Social scientists outside of psychology were impressed with Piaget's method of close observation and detailed "clinical" note-taking and his emphasis on cooperation in problem-solving. In 1969, however, Piaget became the first European to receive the American Psychological Association's Distinguished Scientific Contribution Award. The award noted his creation of an empirical science of epistemology, a field previously restricted to philosophy, as well as both his observations and his controlled data on human cognition, which by then was a major influence on all scientists concerned with cognitive development.

PIAGET'S CLINICAL METHOD

Susan Mayer has shown that Piaget's "Clinical Method," which had brought him to the attention of Harvard social scientists relatively early in his career, was a synthesis of three sources: naturalistic observation, psychometric evaluation, and psychiatric interviewing. Naturalistic observation was practiced by Piaget even as a youth with his biologic studies. He continued this approach when he began to be trained in the psychological testing of children. Under Simon, Piaget learned the psychometric approach, which requires a standardized presentation of specific probes, tasks, and questions. Strictly speaking, psychometric assessment depends on quantitative comparisons of a given child with an adequate normative sample and on intrasubject comparisons across tasks. It requires little interest in the reasons why children make cognitive errors.

Piaget, however, brought to bear his earlier learning of psychiatric interviewing to follow up on erroneous child responses to psychometric items. This served the purpose of identifying forms of thinking that lay between clearly faulty (immature) and fully logical responses. Careful attention to children's language, explanations, and the social functions of their utterances was a further contribution of naturalistic observation to the overall clinical method.

Later in his career, as the clinical method was criticized, Piaget and Inhelder added more experimentally controlled methods to their studies of logical reasoning. However, they did not abandon the observational, psychometric, and individualized probing aspects of the clinical method.

GENETIC EPISTEMOLOGY

Piaget's overarching task was the understanding of how one comes to know things. He took a developmental approach to this question, inquiring how knowledge becomes more adequate over time. He designated himself a genetic epistemologist—one who studies the development of knowledge from infancy to adolescence. By "genetic," he did not mean an approach based on genes but rather an approach based on development over time. It remains important to distinguish Piaget's approach from that of nativism, on the one hand, and empiricism, on the other. Nativism is a theory proposing that at least certain kinds of knowledge are innate or given. In philosophy, Kant took a nativist position regarding the categories of space, time, and causality, arguing that these categories on the data of sensory perception be imposed. In psychology, Gesell theory was essentially based on a nativist position. According to this perspective, certain types of thinking and behavior are expected at certain ages as development unfolds in a predetermined manner.

Empiricism, at the opposite end of the spectrum, proposes that all knowledge is accumulated from the data of sensory perception. In psychology, empiricism is equated with behaviorism and learning theory.

Little attention is paid to predetermined (innate) factors, and the emphasis is on common processes of learning, such as classical and operant conditioning, that lead to accumulated knowledge.

Piaget's approach was, in his words, "constructivist structuralism," according to which mental structures are built through the interaction of the child and the world. Mental structures originate through the actions of the child on objects as the child strives to adapt to the environment. In the educational realm, this implies that guided discovery is preferable to memorization. Only through action and interaction can the child develop a comprehensive and stable structure that enables understanding. What Piaget deemed innate was only an intelligent functioning that makes possible the construction of progressively more adequate structures of knowledge. This improvement of structures is based on abstractions from actions performed over time, and qualitative improvements mark progressive stages of development. The action takes precedence over language in Piaget model. Abstractions from repeated actions lead to the construction of robust and adequate structures. For example, the concept of space is a fundamental mental structure developed in the earliest period of children's lives. In earliest infancy, the child is aware not of one homogeneous space but of several heterogeneous spaces, and each is centered on a certain part of the child's body (e.g., visual space and tactile space). As the child acts on objects that may traverse these various spaces (e.g., a rattle occupying visual, tactile, and auditory spaces), he or she comes to coordinate these individual spaces. Eventually, actions representing displacements in space are organized mentally into the general concept of space. That concept is a structure.

PIAGET THEORY

Piaget provided a succinct summary of his work in a chapter prepared for the 1970 edition of *Carmichael's Manual of Child Psychology.* The interested reader is referred to that source, which was reprinted in the 1983 edition of the *Handbook of Child Psychology.* The summary is based on this and other primary and secondary sources. Piaget theory is broad and comprehensive with regard to cognitive development. It is analogous in this sense to Sigmund Freud or Erik Erikson theories of emotional or personality development. Like those comprehensive theories, it now seems to belong to an earlier era as numerous specific predictions have been modified or falsified, and a great deal of more detailed research has accumulated regarding children's cognitive capacities and processes. This topic will be discussed again after presenting the overarching theory.

Equilibration

For Piaget, the general criterion for intelligent functioning is *equilibration,* briefly defined as "compensation for an external disturbance." Hans Furth described equilibration as "the factor that internally structures the developing intelligence. It provides the self-regulation by which intelligence develops in adapting to external and internal changes." At every level of development, the equilibration mechanism operates to further adaptation, but as development proceeds toward the highest level of cognitive functioning, equilibration becomes progressively more adequate in enabling the organism to adapt to a wider range of internal and external disturbances. Piaget's notion of intelligence as adaptation is, therefore, essentially bound to an equilibration model of intelligent functioning. Equilibration can be thought of as a dynamic exchange as the organism strives to "take in" new and challenging information and "adjust" to this new input.

Equilibration is the final of four necessary components that lead to cognitive development. Piaget recognized the importance of biologic maturation, and more specifically the maturation of the CNS, as essential for cognitive development. He also incorporated learning experiences, including learning based on conditioning, as a necessary component. The third essential component in his theory is social interaction. Not surprisingly, this component appears to be particularly essential for social intelligence. Thus, for Piaget, biologic maturation, learning, and social interaction flow together to facilitate development and crystallize in the process of equilibration as the developing child reacts to cognitive challenges generated by new input.

Assimilation and Accommodation

The biologic foundation of Piaget's developmental theory is clearly evident in his discussion of assimilation and accommodation; these processes are considered functional invariants of all intelligent behavior. At every level of intellectual development, from infancy to adulthood, the processes operate in adaptation.

The assimilation–accommodation account of development stresses the interaction between organism and environment. A certain readiness within the organism is necessary for change or development to occur. In Piaget's view, associationism (empiricism) in psychology commits the fallacy of crediting only one-half of those conditions necessary for learning with all of the explanatory power. A full account of human development must include not only the influence of stimuli on respondents (S → R) but also the influence of the responding organism on incoming stimuli (S ← R). Such an account is provided by Piaget's assimilation–accommodation viewpoint": from a biologic point of view, assimilation is the integration of external elements into evolving or completed structures of an organism. In its usual connotation, the assimilation of food consists of a chemical transformation that incorporates it into the substance of the organism."

Furth referred to assimilation as "an inward-directed tendency of a structure to draw environmental events toward itself." Assimilation is the conservative side of intellectual development, ensuring continuity and coherence by incorporating new aliments (stimuli) into existing mental structures. However, assimilation alone cannot account for growth or change within those structures.

Accommodation occurs during the developmental periods when new data cannot be wholly assimilated into the child's existing mental structures, and yet the data are not so foreign to those structures that they can be ignored. Furth referred to accommodation as "an organism-outward tendency of the inner structure to adapt itself to a particular environmental event." This line of thought reveals the sense in which Piaget defined *intelligence* in terms of equilibration. His equilibration is not a static, balanced system but instead is a dynamic, or mobile, equilibration between assimilation and accommodation as the child responds to the environment.

Structuralism

The term Piaget used for a cognitive structure is *scheme* (*schema*): "A scheme is the structure or organization of actions as they are transferred or generalized by repetition in similar or analogous circumstances." Schemata exist in the infant as perceptual-motor behavior patterns (e.g., the grasping reflex). They also exist in mature intelligence, although, as Furth pointed out, a *schema* is more commonly used to refer to an early mental structure, whereas general schemata resulting from the use of higher intelligence are referred to as *operations.*

The abstraction process that leads to the formation of cognitive structures is called *reflective* or *formal abstraction.* It is an abstraction from *actions,* according to which the similarities inherent in various behavioral acts are dissociated from their particularized contexts (e.g., the earlier example of the rattle in space). These can be contrasted with *simple abstractions,* which are more familiar to psychologists and psychiatrists. Simple abstractions are similarities between or among several things. For instance, when learning colors, the child will note that "green" refers to numerous objects, all of which have the same color; when learning shapes, the child will become able to discriminate circles from squares. Although such abstractions are obviously critical for learning, they do not occur at the same depth of cognitive development as reflective abstractions. The latter contributes not to the content of knowledge but to its progressively more adequate structure.

Theory of Stages

An integral part of Piaget theory of genetic epistemology is a psychology of cognition that seeks to describe how knowledge develops and changes. The genetic framework for that and the process of intellectual adaptation during the major early periods of life is provided in Piaget theory of the stages of cognitive development.

The stages of cognitive development that Piaget and his associates delineated empirically are not defined merely by the dominance of some aspect that remains present but less dominant throughout development.

Piaget was not entirely consistent concerning his stages of cognitive development, but the primary source of confusion in his later writings is whether the so-called preoperational period is to be considered apart from the period of concrete operations in which it culminates. John Flavell's 1963 study and Piaget's 1983 summary defined three major periods and one subperiod of intellectual development (Table 3.2–1). These periods contain subdivisions called *stages.* The major developmental periods are as follows:

1. The sensorimotor period extends from birth until approximately 1.5 years of age. The period is divided into six stages, which are described

in general in the following discussion with reference to the development of the concept of the permanent object.

2. A period of preparation for and acquisition of concrete operations. This period extends from the appearance (at approximately 2 years of age) of the symbolic (semiotic) function to the beginning (at approximately 7 years of age) of higher mental operations applied to concrete objects.
3. The period of formal operations begins at approximately 11 years of age. During this period, full adult intelligence develops as the operations are extended to apply to propositional or hypothetical thinking.

Sensorimotor Period. The *sensorimotor period* of intelligence is so named because the child's construction of mental schemata is in no way aided by representations, symbols, or thoughts. Rather, schemata depend totally on perceptions and bodily movements.

Stage 1 of sensorimotor development is marked by a relatively few organized reflexes that stand out from the spontaneous general activity of the neonate, such as the sucking reflex and the palmar reflex. These primitive reflexes demonstrate three types of assimilation: (1) reproductive (repeating the actions), (2) generalizing (repeating the actions on new objects), and (3) recognitory (performing different varieties of actions on different objects).

Table 3.2–1.
Stages of Intellectual Development Postulated by Piaget

Age (Years)	Period	Cognitive Developmental Characteristics
0–1.5 (to 2)	Sensorimotor	Divided into six stages, characterized by: 1. Inborn motor and sensory reflexes 2. Primary circular reaction 3. Secondary circular reaction 4. Use of familiar means to obtain ends 5. Tertiary circular reaction and discovery through active experimentation 6. Insight and object permanence
2–7	Preoperations subperiod[a]	Deferred imitation, symbolic play, graphic imagery (drawing), mental imagery, and language
7–11	Concrete operations	Conservation of quantity, weight, volume, length, and time based on reversibility by inversion or reciprocity; operations; class inclusion; and seriation
11 through the end of adolescence	Formal operations	Combinatorial system, whereby variables are isolated and all possible combinations are examined; hypothetical-deductive thinking

[a] This subperiod is considered by some authors to be a separate developmental period.

Stage 2 contains the first habits and the primary circular reactions. The first habits develop out of the original schemata as they are applied to objects in the environment or parts of the infant's body without any differentiation between means and ends. In a primitive state of consciousness, the infant is aware only of action sequences and is not even aware of the self. Primary circular reactions occur when, by chance, the infant experiences a new consequence of a motor act and tries to repeat the act.

In stage 3, an initial distinction between means and ends becomes apparent, but in a primitive sense. The infant repeats a particular action pattern that achieves one end for the purpose of achieving other (unrelated) ends. For example, a baby who succeeds in shaking a rattle by pulling a string may repeatedly pull the string in an attempt to bring about other sounds or results.

In stages 4 and 5, infants use a variety of means to obtain particular goals. The distinction between stages 4 and 5 lies in the relative creativity or newness of the means. Stage 4 is marked by the use of familiar means. Stage 5 is marked by a search for new means based on further differentiation of already known schemata and by tertiary circular reactions. The latter differs from secondary circular actions in that the child no longer produces schemata that were effective in one situation to produce magically efficacious results in every situation. Instead, the child explores the environment utilizing various means to achieve an outcome. Discovery is a hallmark of stage 5. For example, a child may use a stick to move an object that is not within reach.

Stage 6 is transitional, leading into the preoperational subperiod. In stage 6, the child becomes capable of inventing new means, not by direct actions on objects but by mental combination. Whereas discovery marked stage 5, insight is a characteristic of stage 6. For example, a child who has seen the father bang on a drawer to loosen it may bang on a toy box to make it easier to open.

During the sensorimotor period, a number of significant concepts are developed, including the child's concepts of space, time, and causality. These categorical concepts develop in a process parallel to the sequence of the six stages outlined in the previous discussion. Most important, during the sensorimotor phase, the child develops the schema of object permanence, the first major victory of conservation and the foundation of all future knowledge.

SCHEMA OF OBJECT PERMANENCE. The knowledge that objects in the external world have an existence independent of the child's actions on them or interactions with them is a major accomplishment of the sensorimotor period. Piaget described his observations of infants' reactions to the disappearance of interesting objects as the foundation for his theory of the development of object permanence. In stages 1 and 2, for example, a child simply continues to look at the place where the object was last seen. In stage 3, if an object such as a spoon drops to the floor, the infant will look for it (e.g., by leaning over and looking at the floor). In stage 4, if an object is repeatedly hidden at point A (visible to the child) and then hidden at point B (also visible to the child), the child searches for it at A, not B. In stages 5 and 6, the infant can follow multiple displacements of the object through points in space, even if the object is hidden within another object.

Preoperational Subperiod and Semiotic Function. The advent of the preoperational subperiod is marked by the appearance of what Piaget called the *semiotic function.* This new ability was defined by Piaget and Inhelder as follows: "It consists in the ability to represent something (a signified object, event, conceptual scheme, etc.) by means of a signifier which is differentiated and which serves only a representative purpose: language, mental image, symbolic gesture and so on."

During the sensorimotor period, a thing could be represented in a limited sense by a part of itself (e.g., the mother's voice might represent the presence of the mother in the room). However, such signifiers are part of what they signify. Symbols and signs are signifiers that are differentiated from what they signify. A drawing of a person, for example, is not part of the person. A word is different from the object to which it refers. These become available to the child only with the appearance of the semiotic function, which makes representational thought possible.

CHARACTERISTIC BEHAVIOR PATTERNS. The semiotic function is heralded by five characteristic behavior patterns in evidence during the second year of life: (1) deferred imitation (imitation that starts after the disappearance of the model); (2) symbolic play, or the game of pretending; (3) drawing, or the use of graphic imagery; (4) the presence of a mental image, which appears as an internalized imitation and not as a function of perception; and (5) the verbal evocation of events not occurring at the time. For

Piaget, the semiotic function, which so enlarges children's worlds—liberating them from the bonds of immediate space and time and enabling them to begin to manipulate symbols and to think rather than just to act on immediately present objects—is rooted in imitation.

Imitation. One can follow the development of imitation through the same six sensorimotor stages delineated for the concept of object permanence. Piaget did this in his volume *Play, Dreams, and Imitation in Childhood.* A radically new form of imitation occurs during the second year of life: deferred imitation. For example, a child may put on their father's hat and walk as their father does, even hours after their father has gone off to work.

For Piaget, intelligence is an equilibration process in which assimilation and accommodation are in balance. However, in imitation, accommodation outweighs assimilation. According to Piaget, imitation is behavior in which "the subject's schemes of action are modified by the external world without his utilizing this external world." In imitation, the child's cognitive structures undergo temporary change without simultaneously incorporating new aliment.

Symbolic Play. A second new behavior pattern that now appears is symbolic play. In imitation, the imbalance between assimilation and accommodation is weighted in favor of accommodation; however, the opposite is true in symbolic play, which is a lessening of the demand of the adaptive process.

Symbols are used in play only after the sensorimotor period, in the type of play characterized by games of pretending. For example, a little girl may pretend that she is asleep, that a box is her pet cat, or that she is in a church. In each instance, symbols are generated "to express everything in the child's life experience that cannot be formulated and assimilated by means of language alone."

According to Piaget, these symbols are created by the same process of imitation that gives rise to deferred imitation at this time (a process in which accommodation outweighs assimilation). Instead of being used accurately (realistically), however, they are used in a process in which liberating assimilation outweighs accommodation.

Drawing. A third behavior pattern associated with the rise of the semiotic function is drawing or the use of graphic imagery. Piaget sees elements of play and imitation in this activity. In developmental terms, he considers drawing “halfway between symbolic play and the mental image,” appearing at approximately 2 or 2.5 years of age. The child enjoys producing drawings for their own sake (assimilation). However, the graphic play also has accommodative elements, especially as the child grows older and attempts to draw not just formless scribble but something.

Mental Image. Closely related to drawing is the presence of a mental image. Piaget tied the genesis of mental imagery to accommodation and imitation. He explicitly denied that mental images could be the product of perception; they are a construction, something the child creates. The mental image is not directly given by perceptual input; it is constructed by the process of accommodation.

Verbal Evocation of Events. The fifth behavior pattern associated with the rise of the semiotic function concerns language, the verbal evocation of events that are not present. Piaget gave the example of a little girl saying, “Anpa, bye-bye.” (Grandpa went away) while pointing to the path he had taken when he left. The parallel with deferred imitation is clear, but the new representational ability is supported by the social system of language.

Concrete Operations. A crucial difference between preoperational and concrete-operational thought is the presence within operative thinking of concepts of conservation. When concrete operations have been organized into a system, the child can conserve, that is, “discover what values do remain invariant … in the course of any given kind of change or transformation.” The progressive and continual structure building that occurs in the concrete-operational period is evident in the increase, with development and age, in the scope of such concepts, such as conservation of quantity, substance, and number. In other words, conservation indicates a stable cognitive structure.

CONSERVATION OF QUANTITY. If a liquid is poured from a short, wide glass into a tall, narrow one, the child in the preoperational stage thinks that the amount of liquid has changed. At the level of concrete operations, however, children are no longer overwhelmed by the perceptual discrepancy between the two configurations. They begin to reason about the transformation, and their correct judgments regarding the conservation of quantity of liquid are accompanied by explanations grounded in logical properties. It is assumed that children are not aware of the logic they use. When problems of conservation begin to be solved, the child passes from the preoperational subperiod into the period of concrete operations.

CONSERVATION OF WEIGHT AND VOLUME. At approximately 7 or 8 years of age, the child can solve the conservation-of-quantity problem and can perform similar judgments of conservation when, for example, a lump of clay is transformed in shape. Between 9 and 10 years of age, the child discovers that the weight of a given object is also conserved, even when its shape is transformed. However, not until approximately 11 or 12 years of age do children have a logical comprehension that the volume displaced by a given object is conserved even after a transformation of the object's shape. Conservation entails the logical certainty that one characteristic of an object remains invariant while the object itself undergoes some type of perceived transformation.

CARDINAL NUMBERS. The concept of cardinal numbers also develops from an initially nonconserving to a conserving stage. For example, consider children in the preoperational subperiod presented with two horizontal rows of colored dots in one-to-one correspondence (i.e., imaginary vertical lines could be constructed between each red dot and its corresponding blue dot). When the experimenter destroys this optical correspondence by spreading out one of the rows of dots, the child in the preoperational period thinks that the larger row contains more dots. Only after the conservation of cardinal numbers has been established as a logical necessity does the child maintain the numerical equivalence of the spread-out row to the other row. Clearly, preoperational concepts of numbers provide inadequate bases for arithmetic skills. It is possible that a lag in the development of number conservation could underlie certain types of arithmetic-related learning disabilities.

Operations. Notions of conservation are the mark of well-established concrete-operational thinking; thus, one must understand the meaning of operation in Piaget's thought. Operations themselves constitute essential thinking. For Piaget, an operation is an action that is (1) interiorized, (2) reversible, and (3) part of an organized system of such actions.

The operations that form this system are interiorized actions. In the sensorimotor period, external behavior patterns give rise through a process of abstraction to the construction of sensorimotor schemata. In a similar fashion, internal thinking patterns later give rise to operations. According to Furth, the possibly generalized aspects of actions, “those which can be found in any coordination of action,” enter into the construction of operations. Saying that the crucial aspect of actions in this regard is their ability to be generalized explains the importance of interiorization in the construction of operations. *Interiorization* refers to “the increasing dissociation of general form from particular content.” In other words, the notions of generalization and interiorization merely point out the process of abstraction that is occurring. For example, a child adds two apples and three apples to obtain five apples. In another instance, a child adds seven blocks and one block to obtain eight blocks. In a third instance, a child combines the category of fathers with that of mothers to obtain the category of parents. The operation abstracted from these three mental actions is an aggregation or combining, without reference to the particular content of numbers, objects, or categories.

Not only must an operation be interiorized action, but it must also be reversible. The action of combining (addition) is not an operation until its relationship to the action of separating (subtraction) is comprehended. To understand reversibility is to understand the third criterion of an operation, its inclusion in a system.

The reversibility essential to operatory thought may be inversion or reciprocity. In reversibility by inversion, an action +A is reversed by −A. For example, in the conservation-of-quantity example, pouring liquid into container 2 (+A) may be mentally reversed, that is, by mentally pouring it back into container 1 (−A). In reversibility by reciprocity, a relation $A < B$ is reversed by a relation $B < A$. Referring again to the conservation-of-quantity example, let A stand for container 1 and B stand for container 2. The rising height of liquid in container 2 ($A < B$) is offset by its narrower width ($B < A$).

Corresponding to these two types of reversibility are the two major categories of concrete operations: those pertaining to classes and those pertaining to relations. In the system of operations performed on classes, reversibility is by inversion; in those performed on relations, it is by reciprocity. For example, subtraction and addition relate to inversion; comparing sticks of different sizes relates to reciprocity.

CLASS INCLUSION. The concrete operation demonstrating an understanding of classes is the class inclusion task. In this task, a child is shown, for example, an array of pets (superordinate class) consisting of dogs and cats (subordinate classes). After counting the number of dogs, cats, and pets, the child is asked whether there are more dogs or more pets. Children in the preoperational subperiod cannot keep in mind the superordinate class while perceiving only the subordinate classes. Thus, they fail the task frequently over a series of such arrays. The concrete operation of class inclusion underlies one's ability to think categorically.

RELATIONS. The concrete operation that demonstrates an understanding of relations is seriation. Children are asked, for example, to arrange a set of rods in order of increasing size. Children in the preoperational subperiod may subgroup the rods but have difficulty completing an entire array along the required dimension. They may understand the concept of smaller versus larger but have difficulty comprehending the gradual nature of change. The concrete operation of seriation underlies one's ability to think dimensionally.

Formal Operations. In the third and final stage in Piaget's conception of intellectual development in the child, the logical structures of concrete operations are superseded by structures referred to as formal operations. The relationship between the real and the possible that characterizes adolescent thinking represents a reversal of that relationship in the thinking of the concrete-operational child. Inhelder and Piaget noted that the real has priority for the younger child, and that possibility is conceived of merely as a prolongation or extension of real operations, "as, for example, when, after having ordered several objects in a series, the subject knows that he could do the same with others." For the adolescent, however, the possible has priority, and the real is seen as a particular instance of it. "Henceforth, they conceive of the given facts as that sector of a set of possible transformations that has actually come about." This immediately presupposes that the adolescent can take a given empirical event (such as "the long, thin rod bends") and categorize it within a system of possible combinations of events (e.g., long rods or short rods, thin rods or thick rods, bending or not bending). Three characteristics follow from this fundamental reorientation in thought: (1) adolescent thought is hypothetical-deductive; (2) it deals in propositions rather than in concrete events; and (3) it can isolate variables and examine all possible combinations of variables.

HYPOTHETICAL-DEDUCTIVE THOUGHT. As a hypothetical-deductive form of thought, formal operational intelligence proceeds from the possible to the real. In this sense, it mirrors scientific reasoning. The implications of a propositional statement are drawn and then tested against reality. Rather than building up a proposition by induction from disparate concrete examples to a loose generalization, formal intelligence operates systematically from a general statement to a specific instance by means of testable hypotheses.

PROPOSITIONAL THOUGHT. That formal operations deal in propositions rather than in concrete events implies increased freedom from immediate content. At one level, this freedom implies the ability to manipulate abstractions that have been tied to concrete examples or events. The adolescent, for example, can perform a transitive inference ($A < B$, $B < C$; therefore, $A < C$) without any empirical demonstration of referents for the terms A and B. At another level, this freedom implies that having performed a concrete operation, the adolescent can abstract the results of that operation and can perform further operations on them. For example, an adolescent can perform the concrete operation of combining two liquids to observe the color of the resultant mix and then take the result of this operation and systematically relate it to the results of all other combinations of available liquids.

ISOLATING VARIABLES AND EXAMINING COMBINATIONS. Instead of dealing with disparate concrete experiments, hypothetical-deductive adolescents can organize their investigations into a coherent pattern from the outset and then perform all relevant combinations of variables to test their hypotheses, thus isolating causal factors. Piaget's theory of formal operational cognition focused on scientific thinking. For example, the weight, speed, shape, and size of an object may all be seen to contribute to the size of a hole the object makes when it hits the ground.

A complete combinatorial system appears only during the period of formal operations. Instead of focusing on empirical givens, as a child in the concrete-operational stage does, the adolescent using formal operational thinking, constructs a hypothetical system comprising the empirical givens. Whereas the younger child could classify events according to various categories, such as length, width, and weight, the adolescent uses that classification as a basis for abstracting all possible combinations of variables. Having done this, the adolescent can then test hypotheses derived from the combinatorial system. The result of this new ability is the capacity to test the causal significance of each individual factor in succession by holding all other factors constant.

Piaget interpreted the rise of formal operational thought in the context of his equilibrium model of cognitive development. Thus, he considered neurologic maturation and experience of the object and interpersonal world as necessary but not enough to explain this qualitative improvement in thinking. The equilibration explanation is as follows: During the stage of concrete operations, several qualitatively heterogeneous factors are constructed by the child, resulting in the achievement of conservation of the factor in question even in the face of perceptual transformations. Such factors include quantity, weight, volume, time, and length. Eventually, the child discovers that, in many concrete instances, the operation of these factors is interrelated. Thus, although the factors have been constructed mentally in relative isolation from one another, their presence in real objects is mixed. Through experience with impersonal and interpersonal objects, the child's concrete operational understanding of these factors is shown to be insufficient, and a more comprehensive, more intelligent understanding is stimulated.

Egocentrism

Each major period of cognitive development is characterized by a qualitative shift toward more-comprehensive and more-adaptive cognitive structures. In this sense, the adolescent is more intelligent than the infant. However, each transition to a higher level of cognitive organization is initially accompanied by a lack of full differentiation between self and object. The most relevant of these episodes for psychiatry occurs with the shift to formal operational reasoning in adolescence. The adolescent has the capacity to engage in hypothetical thinking and to understand others' points of view. However, this is initially accompanied by characteristic patterns of thought in which others are unrealistically presumed to be focusing on the adolescent. This may lead to increased social anxiety, narcissistic reactions, or hypersensitivity. The capacity for formal operations also implies that real, individual people, including parents and teachers, are one instance of what ideal people could potentially be like. This may underlie some features of adolescent rebellion and idealism.

INTELLIGENCE: PIAGETIAN MODEL AND ITS ALTERNATIVES

The Piagetian view of intelligence can be further understood by contrasting it with other contemporary models of intelligence. In the European and American traditions of clinical and educational psychology, intelligence had a pragmatic meaning. Alfred Binet's work in Paris was intended to determine which children were delayed in learning and, thus, in need of special educational assistance. With Théodore Simon, he developed the first *intelligence test* by choosing a variety of tasks and arranging them in order according to the age at which most children master them. This enabled the clinician to assess at what mental age the child was functioning. Further developments in this tradition involved dividing the mental age by the chronologic age to determine the child's intelligence quotient (IQ). This age-corrected notion of intelligence emphasized individual differences through comparisons of children with their peers instead of emphasizing the most general and progressively adaptive competencies of individuals during childhood and adolescence, as Piaget had done. Age-corrected intelligence came to be viewed as a trait, with some children ahead of or behind their age mates in what was believed to be a basic capacity for learning.

Psychologists who espouse a trait model of intelligence differ among themselves on such key issues as to what extent the trait is heritable and whether there is a single general trait or numerous specific traits (types) of intelligence. Nevertheless, all trait theorists differ from Piaget by focusing on individual differences rather than on the most general aspects of cognitive adaptation.

Piaget's view of intelligence also differs from Lev Vygotsky's social model. In the latter, intelligence is viewed chiefly as the outcome of a process through which children internalize the cultural tools and practices represented in their social context. The contrast between Piaget and Vygotsky may at least superficially appear to be one between a theorist who emphasizes the most general cognitive processes and a theorist who emphasizes the most culturally specific cognitive content. As noted earlier, Piaget did include social experience and interaction as a factor contributing to cognitive development, along with biologic maturation, learning, and equilibration. However, he did not emphasize the social and cultural context to the same extent as Vygotsky.

Finally, Piaget's construct of intelligence may be contrasted with the cognitive social learning construct of competencies. Although both constructs involve active, developing processes rather than fixed knowledge, the former pertains to the most general structures of logic derived from coordination of actions, and the latter refers to more-specific, concrete sets of skills, including social and problem-solving skills. Attempts to integrate Piagetian and cognitive social learning models may hold some promise for psychotherapeutic interventions. The latter model points to certain areas that are developmentally crucial for adjustment (goal setting, planning, problem-solving), whereas Piaget model emphasizes the importance of active engagement in the development of structures.

Extensions of Piaget Theory

Piaget theory focuses primarily on logical-deductive reasoning. Little attention is paid to social intelligence, emotional intelligence, or alternative types of intelligence, such as artistic or athletic intelligence. Others have used a Piagetian framework to investigate some, but not all, of these realms. These efforts are referred to as "extensions" of Piaget theory. The first and perhaps best-known attempt to extend Piaget theory in this way was made by Lawrence Kohlberg, who studied the moral reasoning of children.

Although Piaget wrote *The Moral Judgment of the Child,* he did not further develop this area in subsequent work. Kohlberg developed a stage model in which the child's stage of moral reasoning depended on his or her stage of (Piagetian) cognitive development. For instance, to demonstrate moral reasoning at the conventional level, the child must have already entered the concrete-operational stage of cognitive development. Likewise, principled moral reasoning demanded a foundation in formal operational thought.

Kohlberg described three major stages of moral reasoning, each divided into two substages. The three major stages included the morality of the preschool period (based on notions of avoiding punishment and striving for reward), conventional morality (based on notions of authority or mutual benefit), and principled morality (based on general internalized moral principles). As was the case for Piagetian stages, Kohlberg's stages were proposed as progressively more adequate and more highly structured. That is, a child would move only in a forward direction through the stages. However, it was not proposed that all or even most people would necessarily attain the highest level of moral reasoning. Kohlberg investigated the moral reasoning of individuals by presenting them with moral dilemmas and then observing their thinking as they attempted to resolve the dilemmas. The theory of development pertained to the form of their thinking and not to the content of their solution. The best known of the dilemmas involved a husband who needed to obtain a rare medication to save the life of his wife. The druggist who had the medication was charging an exorbitant price for it. The question is whether the husband would be justified in stealing it. In addressing this dilemma, the research participant might have recourse to such considerations as fear of punishment (preconventional), concerns for social order (conventional), or relative values of life and property (principled). Kohlberg was not concerned with the participant's final choice but only with the methods of reasoning used to reach that choice.

Kohlberg's extension of Piaget work was subject to several criticisms. The moral subject was criticized as detached from moral content. In fact, individual differences in the predominant forms of moral reasoning did not seem to correlate highly or consistently with individual differences in actual moral behavior. The research methodology's approach may have confounded the level of verbal skills, and hence general intelligence, with moral reasoning. Finally, because his theory was originally developed from research with an entirely male sample, Kohlberg was criticized by Carol Gilligan for having proposed a theory that was gender-biased. She argued that a woman's moral reasoning proceeds from a primary concern with relationships, whereas a man's moral reasoning proceeds from a primary concern with justice. However, subsequent empirical research has demonstrated that women score as high as, or higher than, men on measures of Kohlberg's stages of moral reasoning.

Following Kohlberg theory, attempts were made to relate Piaget model to social cognition. Initial efforts followed Kohlberg's lead. That is, stages of social cognition were proposed as being dependent on stages of intellectual development. Interpersonal perspective-taking was hypothesized as being dependent on perceptual and cognitive perspective-taking. However, it became clear that children's concepts about other people did not always follow such a layered structural design. James Youniss then developed a theory of children's concepts of other people that appropriated a different key element from Piaget's work. Rather than attempting to layer stages of social cognition on stages of intellectual development, Youniss proposed that social cognition has its own process of development but that these are based on abstractions from interpersonal interactions. Integrating Piagetian psychology with the interpersonal psychiatric theory of Harry Stack Sullivan, Youniss proposed two major categories of children's social cognition: schemas about peers and schemas about authority figures. Each of these develops as a function of repeated interactions with peers or elders. As can be seen in this example, the second type of extension of Piaget's work into the interpersonal domain abandoned the position that stages of social cognition depended on broad structures of intellectual development in favor of the position that the active abstracting processes discovered by Piaget were operative across both types of cognition.

The third type of extension is directed toward the child's acquisition of social knowledge. Again, the Piagetian roots of this extension are to be found in *The Moral Judgment of the Child.* The extension is stimulated by cross-fertilization between adherents of Piaget model and adherents of Vygotsky theory. As Duveen argued, in *The Moral Judgment of the Child,* Piaget depicted two methods by which the child acquires knowledge of social and cultural content. The first is through social transmission, based on learning from an authority figure. The second is through processes of argumentation and debate with peers. Piaget's first process is similar to Vygotsky's notion that the child internalizes from elders a system of symbols that represents a social or collective product. Perret–Clermont pointed out that Piaget emphasized the second process much more than the first over the course of his career and that there was a definite sense in which he viewed the first as potentially constricting. Indeed, it appeared to him to be based on authority but was lacking in the stimulation of new learning that characterized debate with peers and the social construction of knowledge.

To the extent that Piaget neglected or underemphasized the acquisition of knowledge through social transmission, his theory would be deficient in accounting for children's knowledge of social and cultural material that is passed on rather than actively constructed by the individual. Such materials would range from the obvious (socially and culturally acceptable behavior) to the subtle (mathematical inventions, such as the current number system). The implications for education are significant. Discovery methods of education, including experimentation and peer discussion, may be more appropriate when the goal is to assist children in formulating the most general principles of reasoning or to facilitate creative processes, but other methods, including lecture, reading, and modeling, may be more appropriate when the goal is to convey specific social content.

The Post-Piaget Era

Piaget has been criticized for failing to account for individual differences among children in their intellectual development and for failing to account for the effects of cultural settings on the cognitive content of children's intelligence. Educators and clinicians may find Piaget's work of invaluable interest for its broad depiction of developmental stages and processes and yet frustrating because of its failure to provide guidance for interventions that would facilitate learning or correct deficiencies in individual children.

Cross-cultural research has called into question the generality or universality of Piaget's proposed stages. Different tasks appear to be mastered at different ages across cultures. Even within Western culture, there is evidence that modifications in training paradigms or in the way in which cognitive structures are operationalized can lead to significantly different results than those obtained by Piaget regarding children's abilities at different ages. More specifically, the criticism has been made that Piaget relied too heavily on the child's verbalizations to document understanding of a cognitive task. In fact, mastery of concrete and formal operational schema was only acknowledged in Piagetian research if the child could explain verbally why the correct answer was logically necessary.

Moreover, the superiority of formal to concrete operational thinking appears to be culturally relative, as reviewed by Marie-Anne Suizzo. In some cultures, social intelligence, or what has been termed "emotional intelligence," is valued over technical, scholarly forms of cognition, or analogic reasoning is as valued as propositional reasoning.

Moreover, the notion of a developmental stage that may last for as long as 7 years or more, during which schemas of conservation are sequentially developed across areas as divergent as quantity and volume, may simply be too broad to represent a cohesive period. Empirical psychologists have also questioned the notion of cognitive structures by noting the low magnitude of correlations that are purported to measure the same underlying intellectual structure. Of course, similar criticisms have been made of proposed traits and of purportedly related functions in other areas of psychology as diverse as personality theory and the neuropsychology of executive functions.

As noted earlier, the era of “big theories” in psychology and psychiatry appears to be over. Given the explosion of knowledge across such diverse areas as neuroscience, cognitive science, personality and social processes, and cultural and ethnic diversity, it is unlikely that a big theory can possibly hold in such a way as to account for such a complex arena as cognitive development. Gopnik referred to the post-Piaget era, pointing out that cognitive development appears to be specific to certain domains rather than integrated across vast stages. She advocated for developmental pluralism, with a continuing emphasis on testing the validity of each proposed theory. Two theories she discussed are “theory theory” and “script theory,” both of which seem to have relevance for psychotherapy.

“Theory theory” proposes that cognitive development proceeds by way of the same mechanisms that scientists use to test, modify, and confirm or disconfirm theories. Like Piaget model, this model emphasizes the active construction of intelligence on the part of the child or adolescent (or adult) and the role of evidence in leading to modifications. Theories provide predictions, descriptions, interpretations, and explanations of the ensuing evidence. Psychologists using “theory theory” have found that children’s theories about phenomena are domain-specific rather than general or stage-specific (which Piaget proposed) and that the capacities for induction and causal thinking are present earlier in life than Piaget proposed.

Script theory is a more empirical approach to the learning of complex ideas and skills than Piagetian theory. Scripts (or narratives) provide coherence and predictability that are not derived from abstractions of repeated actions but from repeated sequences of actions. Scripts are explanatory in a subjective sense in that they represent the child’s “story” about what happened in a given instance. A common example of a script is the sequence of events that one expects to occur on entering a restaurant: being greeted and seated, ordering courses in sequence, requesting the check, and making payment.

NEW CONCEPTS OF INTELLIGENCE: EMOTIONAL BASIS OF INTELLIGENCE

Human emotions have traditionally been viewed as somewhat separate from cognition and as a minor concern to overall development. However, the research reviewed by Suizzo has shown that children's performance on cognitive tasks improves with higher levels of social support and is affected by the dynamics of interpersonal relationships and by feelings of self-efficacy. New clinical observations and theoretical formulations by Stanley Greenspan and emerging findings from several recent studies suggest that emotions are central to cognition and that they may regulate and orchestrate cognitive capacities. They may also be critical to the development of cognitive capacities. It is suggested that infants' emotional exchanges with their caregivers, rather than their ability to complete cognitive tasks, should become the primary measure of developmental and intellectual competence.

Twelve-month-old Cara sat in her mother's lap at a table, eyes locked onto the psychologist, who tried to get her to follow the bean he placed under the cup and search for it. Cara knocked over the cup. Is this little girl, as her mother fears, cognitively delayed? Does a 1-year-old child who never babbles like other children her age and who violently flings food and toys away from her show signs of a significant intellectual deficit? After a battery of similarly frustrating tests, the evaluator concluded that cognitive delay was the likely diagnosis in Cara's case.

When an evaluator schooled in this new thinking assessed Cara, he focused on her spontaneous interactions with her caregiver. He looked at each of Cara's intentional behaviors as a sign of her emotional interests. For instance, he observed her delight in yanking her mother's nose. At the assessor's suggestion, the mother permitted the tugging on her nose to continue and playfully responded, "Toot, toot." Cara smiled and pulled again. The baby was rewarded with another "Toot, toot" and a big smile from her mother. Cara soon began to copy her mother's gestures and eagerly thrust her nose toward mom. When the mother squeezed Cara's nose, the 1-year-old girl chirped, "Mo, mo," which were her first words.

Cara showed that she could initiate social interactions and could comprehend their consequences. That demonstrated a degree of understanding that put Cara at least at the 12-month level of cognitive development. Further observation revealed an extremely energetic, active, highly physical toddler who liked to have her way and control her surroundings. With the consultant's help, Cara's mother later altered her parenting style. She learned to follow Cara's behavioral lead, then enthusiastically engaged her daughter in creative interactions while simultaneously setting firm limits. Cara's energy quickly became more focused; her babbling became richer. Before long, she was saying real words and actively cooperating with her parents.

For 50 years, developmental testers have expected babies to sit still in their mothers' laps, to pay attention, and to perform prescribed tasks while adults assess their basic intelligence. Traditional wisdom has long insisted that carefully scoring how well a tiny child fits pegs into boards, sorts cards by shape, or hunts beans under cups can reveal an accurate measure of intelligence and developmental competence. However, recent results from research and clinical practice by Stanley Greenspan and others suggest that this entire approach to assessing children's capacities rests on false premises and has inadvertently led to mistaken diagnoses that can stigmatize children throughout their school years.

If a series of such simple, pleasurable interactions with her mother could reveal and foster Cara's language development and organizational ability, then any conception of intellect that marked her as cognitively delayed because of an inability to search for a bean has serious flaws. Those flaws are based on a long-standing mistaken belief that the intellect is superior to and supervises the passions. Clearly, as Cara's linguistic debut demonstrates, analysis of a child's early relationships and sensory and emotional experiences is a vital key to an accurate assessment of intelligence and developmental competence.

Until now, no one has provided an explanation of how emotions give birth to intelligence. In fact, a baby's earliest feelings play a pivotal role in all later intellectual development. Unlikely as the connection between feeling states and intelligence may seem, emotions orchestrate a vast array of cognitive operations throughout an individual's life span. Indeed, they make possible all creative thought.

Results from four distinct lines of inquiry have recently shed new light on the importance of emotions for intelligence. In work with Arnold Sameroff, Greenspan found that children with four or more family emotional risk factors had 24 times the chance of scoring an IQ of less than 80 than children without those risks. Stephen Porges and Greenspan showed that measurements in 8-month-old infants of a part of the brain that regulates emotions correlate with these same children's IQ scores at 4 years of age.

Greenspan's work with a group of children with autistic disorder, who experience some of the most severe thinking and language problems imaginable, also confirmed the inextricable linkage between emotional and cognitive development. Therapeutic programs for these severely challenged children have traditionally concentrated on trying to stimulate their cognition and teach them language. However, a program based on emotional cueing (like the one that revealed Cara's true abilities) proved to be more effective for a number of these children in fostering empathy, warmth, and creative thinking.

One young patient, Ashley, neither spoke nor made any response or eye contact with those around her. The 2-year-old child spent hours staring into space, rubbing persistently at the same patch of rug. Her abnormal repetition was viewed by the clinician observing her as more than just a distressing symptom of her autism. That symptom revealed an underlying interest and motivation that could be harnessed and redirected toward interacting with others. To initiate her cognitive progress, the clinician first had to motivate her to communicate with the simplest of emotional gestures—a smile, smirk, or purposeful hand movements. He suggested that her mother place her hand next to Ashley on the favorite stretch of rug. When Ashley pushed it away, the mother gently put her hand back. Each time the child pushed, the mother's hand would return. A cat-and-mouse game ensued, and after three sessions of these rudimentary interactions, Ashley was looking, smiling, and anticipating.

From that beginning, through a comprehensive therapeutic program, grew a bridge to emotional relationships and eventual verbal exchanges. For example, as therapy progressed, the therapist helped Ashley use her imagination by repeatedly initiating pretend play. He recognized that each time Ashley repeatedly flung herself on her mother, the child was deriving sensory-based simple pleasure from her behavior. He instructed the mother to whinny like a horse each time Ashley lunged at her. Soon Ashley imitated her mother's sounds and then started initiating her own sounds and words. In that way, the therapist helped the mother stretch a pleasant sensation for Ashley into a richer, more complex interaction. Over time, mother and child pretended to be neighing horses, mooing cows, and barking dogs. Their social and emotional interchange grew increasingly complex, passing through the same series of developmental stages identified in children without difficulties. At 7 years of age, Ashley enjoyed warm friendships, argued as well with her lawyer father, and scored in the low-superior IQ range.

The fourth line of inquiry—microscopic clinical observations of children's thinking—further clarifies the relationships between emotion and reason by revealing two necessary elements of thinking. The first process—creating a new idea—stems from the ability to use one's emotional experience to assign meaning and significance to daily events or concepts. The second process—reflection and logical analysis—examines the newly created idea according to whatever principles of logic the person possesses and places it in a wider frame of reference.

To understand those processes in action, Greenspan and Curry put a simple question to two young boys seen in therapy not long ago. When asked by the clinician, “What do you think about people who act bossy to you?” Chris replied, “Well, teachers are bosses, babysitters are bosses, policemen are bosses.” That articulate 7-year-old child lacked the emotional pathways that permit creative and intuitive thought. He could provide a formal classification of different types of bosses but could not relate these categories to his own life. However, 7-year-old Josh had no such difficulties. In response to the same question about bosses, he announced, “Most of the time, I don’t like being bossed, especially when my parents try to tell me when I can watch TV and when I have to go to sleep. I’m big enough to decide for myself. Sometimes when I’m being bad, I guess I need bossing, though. Maybe bosses are okay some of the time, and some of the time they’re not.” Josh found his answer in his, apparently generally irritating, brushes with bosses. Rather than simply listing categories or incidents, he could abstract a principle from the emotional core of those incidents.

How exactly did Josh’s ability to think and abstract develop? A baby’s experience begins with sensations like touch and sound. Each sensation, however, also gives rise to an emotion. A toy may feel interesting or boring; a voice may feel soothing or jarring. Even young infants react to sensations emotionally. They prefer the sound or smell of their mother, for example, to any others and, by 4 months of age, can react to certain persons with fear. Furthermore, contrary to long-held assumptions, basic sensations, such as touch and sound, can be perceived differently by different people, giving rise to emotional differences.

Emotional meaning also adheres to early concepts like *big* and *little, more* and *less, near* and *far,* and *now* and *later. A lot* is a bit more than makes a child happy. *Near* is snuggled next to a child in bed. *Later* is a frustrating stretch of waiting. For a child without an intuitive sense of *few* and *many,* numbers have no meaning. Furthermore, a young child's experience of any sensation always occurs within the context of a relationship that gives it broader meaning. Playing with mother's hair, for example, may evoke smiles and hugs or an angry scolding.

Each sensory experience has this dual aspect and is labeled by its physical properties and its emotional qualities. This double coding helps the child to place the memory or experience in a catalog of experiences and to retrieve or reconstruct it when needed. As the child grows, emotional reactions come to operate as a sixth sense that allows the child to recognize and understand situations.

Emotion orchestrates complex judgments as well. One of modern psychology's main enigmas is how children learn to discriminate among situations ("When can I yell and kick?") and to generalize from one to another ("Should I behave at school like I do at home?"). Consider how a child makes a seemingly simple judgment about when to say "hello." He or she does not learn a set of cognitive rules, such as greeting only those who live on his or her street or only those who wave at him or her. Instead, the child abstracts an emotional pattern from numerous social encounters; there is a feeling of warmth and friendliness in situations that rate "hello." The child's interactions create an emotional signaling system that tells him or her when to say "hello" and that it is okay to punt the football but not to kick Sarah or Charlie in the shins.

Looking at Piaget theory from this perspective reveals the limitations of a cognitive theory that did not adequately deal with the central role of emotions. Piaget's experiments focused on how children comprehend the relationship between physical objects, developing the ability to classify them by such parameters as shape or size, but most children can classify their emotions and emotionally relevant relationships far earlier than they can classify physical objects. For example, they know members of their families from those who are not members, classifying the family as a unit. Some of Piaget's observations are limited because he depended so heavily on children's perceptual and motor performance to signal cognitive advances, even though motor skills often lag behind other skills.

More important, however, is Piaget's relative lack of focus on the role of emotions. He emphasized learning through doing but did not realize that the doing generates formative emotional reactions as well as perceptual, motor, and cognitive ones. Consider how a child learns what an apple is. You can have the child handle an apple and determine it is something red and round, bigger than a peanut, and smaller than a watermelon. Alternatively, the child can observe the aspects mentioned previously, experience more satisfaction in eating an apple when he or she is hungry than when he or she is full and know the pleasure of giving one to his or her favorite teacher.

The child can also imagine how the teacher feels when he or she gives the teacher the apple. The child may catalog how he or she feels when the child throws one at his or her younger brother and gets him on the shoulder, as well as the child's disgust when an apple rots or when the child discovers half of a worm inside. If you ask creative adults to write an essay on apples, they will probably bring a large amount of personal affective experience to their reflections.

Piaget did emphasize how children's thinking comes to incorporate multiple perspectives as they grow older. A classic Piagetian experiment shows how school-age children learn to solve a problem involving weights on a seesaw. By assessing the heaviness of the weights and noticing where they are placed, they can figure out how a seesaw works. But what was not appreciated by Piaget was that the children's postulated perspectives incorporated the additional perceptions afforded by affective experiences. To neglect this element is, therefore, to fail to appreciate the rich array of experiences that contribute to forming abstract concepts.

TOWARD A GENERAL DEVELOPMENTAL MODEL

The diagnosis and treatment of emotional and developmental disorders in infants and young children require that clinicians address all facets of the child's experience. One needs a model with which to look at how constitutional–maturational (regulatory), family, and interactive factors work together as the child progresses through each developmental phase, and each phase must be viewed from affective and cognitive perspectives.

A developmental, structuralist model formulated by Greenspan integrates cognitive and affective development and applies the types of structure Piaget described to a range of experiences. Most importantly, the model also considers individual differences in terms of biology and interaction. New findings suggest that early interaction can alter the structure and wiring of the CNS. The model can be visualized with the infant's constitutional–maturational patterns on one side and the infant's environment, including caregivers, family, community, and culture, on the other side. Both sets of factors operate through the infant–caregiver relationship, which can be pictured in the middle. Those factors and the infant–caregiver relationship, in turn, contribute to the organization of experience at each of six developmental levels (consistent with cognitive and affective milestones), which may be pictured just beneath the infant–caregiver relationship.

This clinical and research model enables the user to look at the mutual influence of highly specific, verifiable constitutional–maturational factors on interactive and family patterns and vice versa, in relationship to specific developmental processes (and to relate these processes to later developmental and psychopathologic disorders).

Developmental Levels

The model contains six developmental levels, which include the ability of the infant and the child to accomplish the following:

1. Attend to multisensory affective experience and, at the same time, organize a calm, regulated state, and experience pleasure.
2. Engage with and show affective preference and pleasure for a caregiver.
3. Initiate and respond to two-way presymbolic gestural communication.
4. Organize chains of two-way communication (opening and closing many circles of communication in a row), maintain communication across space, integrate affective polarities, and synthesize an emerging prerepresentational organization of self and others.
5. Represent (symbolize) affective experience (e.g., pretend play and functional use of language), which calls for higher-level auditory and verbal sequencing ability.
6. Create representational (symbolic) categories and gradually build conceptual bridges between these categories. This ability creates the foundation for such basic personality functions as reality testing, impulse control, self–other representational differentiation, affect labeling and discrimination, stable mood, and a sense of time and space that allows logical planning. This ability rests not only on complex auditory and verbal processing abilities but also on visual-spatial abstracting.

At each level, one looks at the range of emotional themes organized (e.g., can the child play out [symbolize] only dependency themes and not aggressive ones? Is aggression behaved out and dealt with presymbolically?). One also looks at the stability of each level. Does minor stress lead a child to lose the ability to represent, interact, engage, or attend?

In their use in day-to-day clinical work, the six developmental levels can be collapsed into four essential processes that characterize development in infants and young children. These processes concern how an infant and the parents or caregivers negotiate the various phases of their early interactions, and they serve as a basis for diagnosis and treatment.

A 12-month-old infant’s mother worried that “he cries any time I try to leave him, even for a second. If I’m not standing right next to him when he is sitting on the floor, he cries, and I have to pick him up. He’s a tyrant. He’s waking up four times at night and is a fussy eater. He eats for short bursts (breastfeeding) and then stops eating. I’m feeding him all the time.”

The mother was feeling cornered, controlled, manipulated, and bossed around. Her baby was like a “fearful dictator” (therapist’s term). She said, “that’s the perfect way to describe him.” The father was impatient with the mother; he felt that she indulged the baby too much. He was getting “fed up” because she had no time for him.

The baby was interactive and sensitive to every emotional nuance. As he came into the room, he immediately caught the clinician’s eye. They exchanged smiles and motor gestures. He interacted with his parents with smiles, coos, and motor movements. Father intruded somewhat. He would roughhouse until the baby would cry, put the baby down, and then roughhouse again. Mother, in contrast, was ever so gentle, but long silences passed between her vocalizations. During her long silences, the baby would rev up, get more irritable, and start whining. He whined with his mother and cried fearfully with his father. Even before he could finish his motor gestures or vocalizations, his mother moved in and picked him up, gave him a rattle, or spoke for him. In this way, she undermined his initiative. Even while whining, however, he was interactive and contingent.

On physical examination, this baby was sensitive to loud noises and light touch on the arms, legs, abdomen, and back. He had a mild degree of low motor tone and was posturally insecure. He was not yet ready to crawl.

His constitutional and maturational patterns did not compromise his mastering the first developmental challenge of shared attention and engagement. He was an attentive, engaged baby. However, in the second developmental stage, intentional communication and assertiveness, he was a passive reactor. He was not learning to initiate two-way communication, to be assertive, and to take charge of his interactions. His low motor tone was compromising his ability to control his motor movements. His sensory hyperreactivity was compromising his ability to regulate sensation. He was frequently overloaded by just the basic sensations of touch and sound, and he was not receiving support from his mother through the nurturing and rhythmic caretaking that would foster self-initiative.

This family required therapeutic work on several tasks simultaneously. The infant's special constitutional–maturational patterns were discussed. Hands-on practice helped the parents help their baby be attentive and calm. Those tasks included helping the mother be more patient, waiting for the baby to finish what he started, and supporting his initiative (e.g., putting something in front of him while he was on his tummy to motivate him to crawl and reach); getting the mother to put more affect into her voice and to increase the rhythm and speed of her vocalizations, and getting the father to be more gentle. The parents' feelings about the interactions were explored: the father's tough-guy background, the mother's fear of her own assertiveness, her fear of her baby being injured, and their own associated family patterns.

Gradually, the baby began to sleep through the night, and he became more assertive and less clinging and fearful. He also became happier. He was slow to reach his motor milestones, so an occupational therapist worked with him and gave the parents advice on motor development and normalizing his sensory over-reactivity. In 4 months, this infant was functioning in an age-appropriate manner with a tendency toward a cautious but happy and assertive approach to life's developmental challenges.

As developmental clinicians and researchers build on Piaget's findings and formulations, the developmental model serves as a basis for understanding social and emotional development and provides a framework for clinical and educational intervention.

Implications of Piaget's Work for Psychotherapy

Piaget was not an applied psychologist and did not develop the implications of his cognitive model for psychotherapeutic intervention. Nevertheless, his work formed one of the foundations of the *cognitive revolution* in psychology. One aspect of this revolution was an increasing emphasis on the cognitive components of the therapeutic endeavor. In contrast to classical psychodynamic therapy, which focuses primarily on drives and affects, and in contrast to behavior therapy, which focuses on overt actions, cognitive approaches to therapy focus on thoughts, including automatic assumptions, beliefs, plans, and intentions. By including “theory theory” and “script theory,” additional applications to psychotherapy can be seen.

Cognitive development theory has influenced psychotherapeutic approaches in multiple ways. Some therapists have taken developmental notions from Piaget’s work and developed intervention techniques. Others have developed cognitive models of treatment independent of Piaget but with heavy reliance on the role of cognition. Others have included Piaget’s concepts in a broader set of constructs to undergird new developmental approaches to psychotherapy.

First, some psychotherapists applied Piagetian notions directly to child interventions. Susan Harter, for example, discussed techniques for helping young children become aware of divergent or contradictory emotions and integrate these complex emotions within a more abstract or higher class of emotions. One of Harter techniques is to ask the young child to make a drawing that shows different and conflicting feelings in one person. This technique represents an application of the concrete operation of class inclusion to the realm of emotions. Harter’s work applied Piagetian findings to the common therapeutic problem of helping children to recognize, tolerate, and integrate mixed or ambivalent affects within stable object relations. As such, it drew on cognitive theory and psychodynamic theory. Similar techniques are important in working with children who have been exposed to trauma or to sexual abuse. It is an essential component of such work to assist them in labeling, differentiating, and accepting the full range of emotions stemming from these experiences.

Second, other psychotherapists developed treatment models that, although not directly dependent on Piagetian psychology, emphasized core ideas quite similar to those Piaget discovered in his naturalistic observations of cognitive development. These models are even more closely aligned with recent developments in “theory theory.” Aaron Beck, for example, developed an entire school of cognitive therapy that focuses on the role of cognitions in causing or maintaining psychopathology. Cognitive therapy has been shown to be an effective treatment for problems as diverse as depression, anxiety disorders, and substance abuse.

A core idea in cognitive therapy is that the patient has developed certain core beliefs, aspects of the self-schema, and conditional probability beliefs because of developmental experiences, and these contribute to emotional or behavioral problems. For example, depressed persons may have the core belief that “I am unlovable.” Addicted persons may have the belief, “Unless I drink, I cannot feel happy.” In cognitive therapy, the person can be assisted in identifying the negative automatic thoughts and underlying dysfunctional attitudes or beliefs that contribute to emotional distress or addictive behavior. The key therapeutic process after identification of the maladaptive thoughts is to help the patient view these thoughts more objectively, not take them in an unquestioning manner as veridical. Here, cognitive therapy emphasizes evidence, consistent both with Piagetian theory and “theory theory.” The patient is assisted in seeking out evidence to test negative thinking; active involvement, rather than passive listening, is required.

What the cognitive therapist accomplishes through such techniques as Socratic questioning and asking if there are other ways to look at the same event is similar to what the talented teacher does in guiding children to a more adequate, more intelligent understanding of operational tasks. The notion of equilibration is relevant in both instances. By helping the individual see that previous cognitive structures are in some ways inadequate, the therapist or teacher disturbs the old cognitive structure, and the patient or student experiences a disruption that leads to the search for more-adequate structures. The compensation for external disturbance is what Piaget termed *equilibration.* New structures can only be constructed through a process of accommodation, enabling the subject to assimilate a wider array of data, a new perspective, or more complex information.

Because it requires *thinking about thinking,* cognitive therapy seems to require formal operational thinking, although this has not been empirically tested. At the least, it requires the ability to recognize and articulate affects, to recognize and label events that give rise to affects, and to translate into thought the mediating process that occurs rapidly between the event and the affect. Cognitive-behavioral models of psychotherapy include cognitive techniques and more-behavioral, interactive techniques, such as increasing pleasant activities and improving communication and problem-solving skills. It is possible that the less-cognitive, more-behavioral techniques, although requiring a lower level of cognitive development, can also lead to the garnering of evidence and modification of specific expectancies, attributions, and self-schemata.

Because “script theory” or narrative approaches to cognition in psychotherapy are empirically based, generated by repetitive experiences rather than by reflective abstraction, and domain-specific, they may have even more general application to psychotherapy than classic Piagetian theories or “theory theory.” For example, in dialectical behavior therapy, patients provide a “chain analysis” of events, feelings, thoughts, situational stimuli, and interpersonal factors that led up to negative or self-damaging behavior. This narrative provides guidance to the patient and the therapist about where and how to intervene to prevent subsequent similar behavior.

A third approach is to use Piagetian insights in psychotherapy by integrating Piaget’s findings into a broader model. Greenspan, for example, has articulated a developmentally based psychotherapy that takes account of earlier, presymbolic levels of functioning that precede the ability to recognize, label, and articulate affects and their mediating cognitions.

Developmentally Based Psychotherapy

Developmentally based psychotherapy, as developed by Greenspan, integrates cognitive, affective, drive, and relationship-based approaches with a new understanding of the stages of human development. Different therapies look at different aspects of the proverbial elephant, whether from a psychodynamic, object relation, self-psychology, behavioral, or cognitive-behavioral point of view. A comprehensive, cohesive developmental framework integrates elements from these approaches with a broader understanding of the developmental processes essential for emotional or mental health. It formulates a series of principles that an understanding of human development says are prerequisites for emotional growth.

Developmentally based psychotherapy constructs its therapeutic strategies from these principles of human development and growth. The clinician first determines the level of the patient’s ego or personality development and the presence or absence of deficits or constrictions. For example, can the person regulate activity and sensations, relate to others, read nonverbal affective symbols, represent experience, build bridges between representations, integrate emotional polarities and abstract feelings, as well as reflect on internal wishes and feelings?

After determining the developmental level, the clinician looks for constitutional and maturational contributions and difficulties with sensory processing, modulation, or motor planning. The clinician looks for interactive and family contributions. Each of these is explored in the present, the past, and the anticipated future. The patient’s fantasies, sense of self and others, and conflicts are understood in the context of all of these influences. These are how patients make sense of their ego structure, physical makeup, family patterns, and interactions with others. Developmentally oriented therapists do not focus on just one set of variables, such as inner fantasies, family dynamics, biologic proclivities, or prior experience. Similarly, the formulated therapeutic strategy cannot deal only with one or two factors. It must deal with all critical factors that influence the developmental process. As collaborators in the construction of experience, therapists use their understanding of the patient’s development to help the patient to construct interactions that provide growth and overcome difficulties.

Often, it is assumed that critical aspects of development occur through the maturation of the nervous system along with routine, expectable experiences. It is also assumed that, from these routine, expectable maturational sequences and experiences, certain psychological structures having to do with the ability to regulate, engage, interact, represent (symbolize) experience, and reflect and compare experiences are present in most people. With these capacities in place, it is believed that the therapeutic process can focus on conflicts and anxieties and selected maladaptive behaviors or thoughts. It has been observed, however, that only a small percentage of individuals have these core capacities. For most, such capacities must be learned as part of the therapeutic process.

The developmental perspective shows how one learns these capacities during development. It suggests strategies that can be used in the psychotherapeutic process so that adults and children who have not achieved these capacities can learn them. From a developmental point of view, the integral parts of the therapeutic process include learning how to regulate experience; engage more fully and deeply in relationships; read and respond to boundary-defining behaviors and affects; perceive, comprehend, and respond to complex self- and object-defining affects, behaviors, and interactive patterns; to represent experience; to differentiate represented experience; and to form higher-level differentiations, including the capacity to engage in the ever-changing opportunities, tasks, and challenges during the course of life (e.g., adulthood and aging) and, throughout, to observe and reflect on one's and others' experiences. Mastering these core developmental processes makes dealing with conflicts, anxieties, maladaptive behaviors, and thoughts possible.

ACKNOWLEDGMENTS

This chapter represents an update and revision of previous chapters in this series, most recently by Stanley I. Greenspan and John F. Curry from the last edition. We are indebted to Drs. Greenspan and Curry for their work and retain much of their prose in this current edition.

FURTHER READINGS

*American Psychological Association. Distinguished scientific contribution awards. *Am Psychol*. 1970;25:65–79.

*Bennour M, Voneche J. *The historical context of Piaget's ideas*. In: Muller U, Carpendale J, Smith L, eds. The Cambridge Companion to Piaget. Cambridge University Press; 2010.

Gilligan C. *In a Different Voice: Psychological Theory and Women's Development*. Harvard University Press; 1982.

Gopnik A. The post-Piaget era. *Psychol Sci*. 1996;7:221.

Greenspan SI. *Developmentally Based Psychotherapy*. International Universities Press; 1997.

Greenspan SI. *Infancy and Early Childhood: The Practice of Clinical Assessment and Intervention with Emotional and Developmental Challenges*. International Universities Press; 1992.

Greenspan SI. *The Clinical Interview of the Child*. McGraw-Hill; 1981.

Greenspan SI. *The Development of the Ego: Implications for Personality Theory, Psychopathology, and the Psychotherapeutic Process*. International Universities Press; 1989.

Greenspan SI. *The Growth of the Mind and the Endangered Origins of Intelligence*. Addison Wesley Longman; 1997.

Greenspan SI, Shanker S. *The Evolution of Intelligence: How Language, Consciousness, and Social Groups Come About*. Perseus Books; 2003.

Harter S. A cognitive-developmental approach to children's expression of conflicting feelings and a technique to facilitate such expression in play therapy. *J Consult Clin Psychol*. 1977;45:417.

*Hsueh Y. "He sees the development of children's concepts upon a background of sociology": Jean Piaget's honorary degree at Harvard University in 1936. *Hist Psychol*.

2004;7:20–44.

Inhelder B, Piaget J. *The Growth of Logical Thinking from Childhood to Adolescence*. Basic Books; 1958.

Kohlberg L. The development of children's orientations toward a moral order: I. Sequence in the development of moral thought. *Vita Humana*. 1963;6:11.

Mayer SJ. The early evolution of Jean Piaget's clinical method. *Hist Psychol*. 2005;8:362–382.

Piaget J. *The Moral Judgment of the Child. London: Kegan Paul, Trench*, Trubner and Co.; 1932.

Piaget J. *Play, Dreams and Imitation in Childhood*. Norton; 1951.

Piaget J. The stages of the intellectual development of the child. *Bull Menninger Clin*. 1962;26:120.

Piaget J. *The Early Growth of Logic in the Child*. Norton; 1969.

Piaget J. Piaget's theory. In: Mussen PH, ed. *Carmichael's Manual of Child Psychology*. 3rd ed. Vol. 1. Wiley; 1970:703.

Piaget J. Piaget's theory. In: Mussen PH, ed. *Handbook of Child Psychology*. 4th ed. Vol. 1. Wiley; 1983:103;128.

Piaget J, Inhelder B. *The Psychology of the Child*. Basic Books; 1969.

Pinard A, Laurendeau M. Stage in Piaget's cognitive-developmental theory: exegesis of a concept. In: Elkind D, Flavell JH, eds. *Studies in Cognitive Development: Essays in Honor of Piaget*. Oxford University Press; 1969.

Shayer M. Not just Piaget, not just Vygotsky, and certainly not Vygotsky as alternative to *Piaget. Learn Instruct.* 2003;13:465.

Smith L, Dockrell J, Tomlinson P, eds. *Piaget, Vygotsky, and Beyond: Future Issues for Developmental Psychology and Education*. Routledge; 1997.

Sternberg RJ, Berg C, eds. *Intellectual Development*. Cambridge University Press; 1992.

Suizzo MA. The social-emotional and cultural contexts of cognitive development: neo-Piagetian perspectives. *Child Develop*. 2000;71:846–849.

Tudge J, Rogoff B. Peer influences on cognitive development: Piagetian and Vygotskian perspectives. In: Lloyd P, Fernyhough C, eds. *Lev Vygotsky: Critical Assessments: The Zone of Proximal Development*. Vol. 3. Routledge; 1999.

Youniss J. *Parent and Peers in Social Development: A Sullivan–Piaget Perspective*. University of Chicago Press; 1980.

3.3 Learning Theory

Mark E. Bouton, Ph.D.

The principles of learning are always operating and influencing human activity. Because so much human behavior (including overt behavior, thought patterns, and emotion) is acquired through learning, learning principles are often deeply involved in the etiology and maintenance of psychiatric disorders. And because behavior change, the goal of all psychotherapy, is also strongly influenced by learning processes, learning principles can influence the effectiveness of a therapy. In fact, no therapy can be said to be immune to the effects of learning. Even the simple prescription of a medication can bring learning processes into play, because the patient will have opportunities to learn about the drug's effects and side effects, will need to learn to comply with the instructions and directions for taking it, and will need to learn to overcome any resistance to compliance. Learning theory is worth knowing, it has been said, because "much like the law of gravity, the laws of learning are always in effect."

HISTORICAL CONTEXT

The modern study of learning can be traced back to roughly the turn of the 20th century, when two great traditions in biology began to focus on behavior and learning. Working in the reflex tradition, the Russian physiologist Ivan Pavlov first saw the importance of what is now called *classical (or Pavlovian) conditioning*. In the famous description of his simplest experiment, Pavlov rang a bell and then gave a dog some food. After a few pairings of bell and food, the dog began to salivate to the bell and thus anticipated the presentation of food. Today, classical conditioning is seen as an important and ubiquitous behavioral phenomenon, as well as a powerful method for studying basic associative learning processes (see later discussion). In addition, from its earliest days, classical conditioning was also applied to psychiatric phenomena. For example, in the 1920s (before most of Pavlov’s work was available in English), John B. Watson showed that human emotions are influenced by classical conditioning. Watson showed an infant boy a stimulus (which happened to be a laboratory rat) and then made a frightening noise. After a few pairings of the rat and the noise, the child became afraid whenever the rat was presented. This fear generalized to a rabbit, a dog, and a fur coat. Watson saw conditioning as a means by which human emotions could be elicited by an expanding range of cues.

Also working at the turn of the 20th century, but in the tradition of the early comparative psychologists (who were interested in studying the evolution of the mind by studying the behavior of animals), Edward Thorndike ran some of the earliest systematic experiments on *instrumental (or operant) conditioning*. In his studies, cats were placed in “puzzle boxes” and required to learn to move a latch to get out of the box and find reward. Although the method was originally seen as a way to study the cognitive abilities of animals, the instrumental conditioning experiment was soon seen as a tool for studying how the effects of a behavior can strengthen or weaken it. In the late 1930s, B.F. Skinner expanded Thorndike’s method by enclosing rats in a box and allowing them to operate a lever attached to the wall to earn a food-pellet reward. In this case, the animal was free to perform the behavior repeatedly, whenever it wanted to. Nowadays, and in parallel with Pavlovian conditioning, operant learning is seen as (1) an important and ubiquitous behavioral phenomenon, (2) a method for rigorously studying learning and motivation processes (see below), and (3) a phenomenon with applicability to clinical issues.

The application of learning theory to clinical issues became especially central to the *behavior therapy* movement that began in the 1950s and 1960s (e.g., Wolpe, 1958). The specific application of operant conditioning principles became known as *applied behavioral analysis.* The fundamental idea was that psychiatric disorders could be understood and treated using scientifically established principles of learning. The relationship between lab-based research on learning and the practice of clinical psychologists was viewed as analogous to the relationship between lab-based medical research and the practice of physicians. Since the 1970s, the field of behavior therapy (now called *cognitive behavior therapy*) has accepted what seem to be a wider range of explanatory principles, often grounded in social psychology and at least nominally in an information-processing (“cognitive”) perspective rather than an explicitly “behavioral” perspective. At the same time, the basic science of learning theory has also advanced and expanded in ways that have not always been appreciated by clinical scientists. Beginning in the late 1960s and continuing today, new findings led to conceptual changes in how learning processes were analyzed—with an increasing emphasis on cognitive processes and evolution. These developments further deepened the understanding of learning processes and widened the possible implications and applications to clinical problems and issues.

BASIC CONCEPTS AND CONSIDERATIONS

A great deal of modern research on learning still focuses on Pavlovian and operant learning. Pavlovian conditioning occurs when neutral stimuli are associated with a psychologically significant event. The main result is that the stimuli come to evoke a set of responses or emotions that may contribute to many clinical disorders, including (but not limited to) anxiety disorders and drug dependence. The events in Pavlov’s experiment are often described using terms designed to make the experiment applicable to any situation. The food is the *unconditional stimulus*, or *US*, because it unconditionally elicits salivation before the experiment begins. The bell is known as the *conditional stimulus*, or *CS*, because it only elicits the salivary response conditional on the bell–food pairings. The new response to the bell is correspondingly called the *conditional response* (*CR*), while the natural response to the food itself is the *unconditional response* (*UR*). Of course, as will be illustrated below, modern laboratory studies of conditioning use a very wide range of CSs and USs and measure a wide range of conditioned responses.

Operant conditioning occurs when a behavior (instead of a stimulus) is associated with a psychologically significant event. In the laboratory, the most famous experimental arrangement is the one in which a rat presses a lever to earn food pellets. In this case, as opposed to Pavlov’s, the behavior is said to be an *operant* because it operates on the environment. And the food pellet is a *reinforcer*, an event that increases the strength of the behavior it is made a consequence of. A major idea behind this method is that the rat’s behavior is “voluntary,” in the sense that the animal is not compelled to make the response (it can perform it whenever it “wants” to). In this sense, it is similar to the thousands of operant behaviors that humans choose to commit—freely—in any day. Of course, the even larger idea is that even though the rat's behavior appears as though it is voluntary, it is lawfully controlled by its consequences: If the experimenter were to stop delivering the food pellet, the rat would stop pressing the lever, and if the experimenter were to allow the lever press to produce larger pellets, or perhaps pellets at a higher probability or rate, then the rate of the behavior might increase. The point of operant conditioning experiments, then, is largely to understand the relation of behavior to its payoff.

Pavlovian and operant conditioning differ in several ways. One of the most fundamental differences is that the responses observed in Pavlov's experiment are *elicited* and thus controlled by presentation of an antecedent stimulus. In contrast, the "response" observed in Skinner's experiment is not elicited or compelled by an antecedent stimulus in any obvious way—it is instead controlled by its consequences. This distinction between *operants* and *respondents* is important in clinical settings. If a young patient is referred to the clinic for acting out in the classroom, an initial goal of the clinician will be to determine whether the behavior is a respondent or an operant, and then go about changing either its antecedents or its consequences, respectively, to reduce its probability of occurrence.

Despite the academic separation of operant and respondent conditioning, they have an important common function: Both learning processes are designed by evolution to allow organisms to adapt to the environment. The idea is illustrated by considering the *law of effect* (Fig. 3.3–1), which says that whether an operant behavior increases or decreases in strength depends on the effect it has on the environment. When the action leads to a positive outcome, the action is strengthened; conversely, when the action leads to a negative outcome, we have *punishment*, and the action is weakened. In a similar manner, when an action decreases the probability of a positive event, behavior also declines. (Such a procedure is now widely known as *time-out* from reinforcement.) When an action terminates or prevents the occurrence of a negative event, the behavior will strengthen. By thus enabling the organism to maximize its interaction with positive events and minimize its interaction with negative ones, operant conditioning allows the organism to optimize its interaction with the environment. Of course, events that were once positive in the human's earlier evolutionary history are so prevalent in modern society that they do not always seem adaptive today. Thus, reward learning also provides a framework for understanding the development of rather maladaptive behaviors like overeating (in which behavior is reinforced by food) and drug taking (in which behaviors are reinforced by the pharmacologic effects of drugs)—cases where reward principles lead to psychopathology.

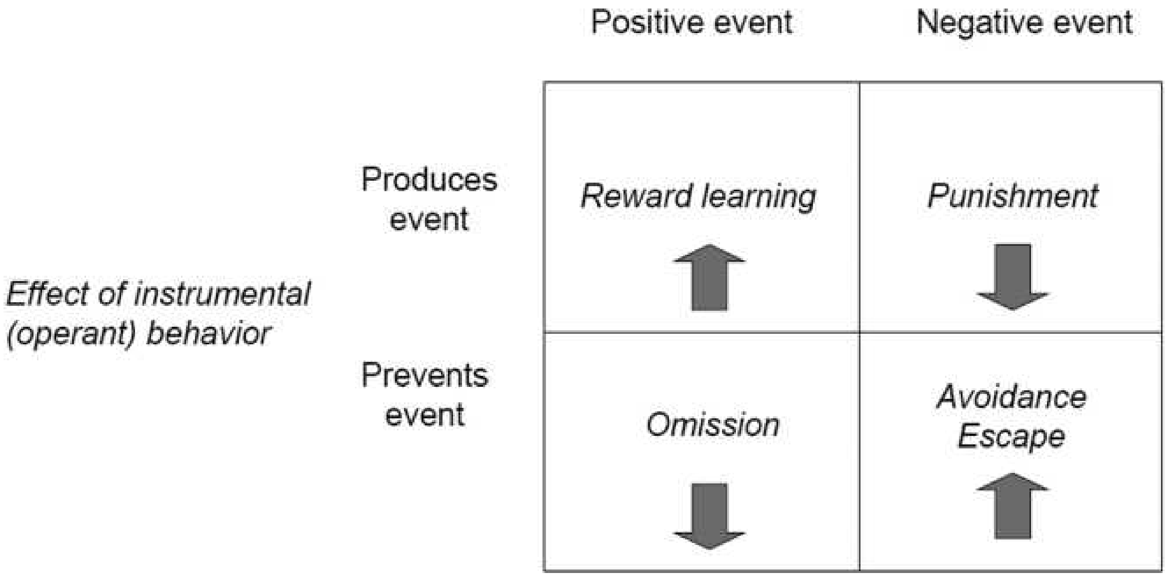

FIGURE 3.3–1. The law of effect in instrumental/operant learning. Actions either produce or prevent good or bad events, and the strength of the action changes accordingly (indicated by *arrow*). “Reinforcement” refers to a strengthening of behavior. *Positive reinforcement* occurs when an action produces a positive event, while *negative reinforcement* occurs when an action prevents or eliminates a negative event.

A parallel to Figure 3.3–1 exists in Pavlovian conditioning, in which one can likewise think of whether the CS is associated with positive or negative events (Fig. 3.3–2). Although such learning can lead to a wide constellation or system of behaviors (as illustrated later), in a very general way, it also leads to behavioral tendencies of approach or withdrawal. Thus, when a CS signals a positive US, the CS will tend to evoke approach behaviors—called *sign tracking*. For example, an organism will approach a signal for food. Analogously, when a CS signals a negative US, it will evoke behaviors that tend to move the organism away from the CS. Conversely, CSs associated with a decrease in the probability of a good thing will elicit withdrawal behaviors, while CSs associated with the decrease in the probability of a bad thing can elicit approach. An example of the latter case might be a stimulus that signals safety, or the decrease in probability of an aversive event, which evokes approach in a frightened organism. In the end, these very basic behavioral effects of both operant (Fig. 3.3–1) and Pavlovian (Fig. 3.3–2) learning serve to maximize the organism's contact with good things and minimize the contact with bad things.

Perhaps because they have such similar functions, Pavlovian learning and operant learning are both influenced by similar variables. For example, in either case, behavior is especially strong if the magnitude of the US or reinforcer is large, or if the US or reinforcer occurs relatively close to the CS or operant response in time. And in either case, the learned behavior decreases if the US or reinforcer that was once paired with the CS or the response is eliminated from the situation. This phenomenon, called *extinction*, provides a means of eliminating unwanted behaviors or behavioral excesses that were learned through either form of conditioning and has led to a number of very effective cognitive behavioral therapies.

Pavlovian and operant conditioning are usually described separately, in part because the separation has helped to simplify their analysis in the laboratory. However, it is important to bear in mind that both processes are occurring all the time, and that both are also always occurring together. For example, at the same time the drug user is learning to associate various operant behaviors with the reinforcing effects of drugs, they are also learning to associate the drug with various stimuli, such as room cues or drug paraphernalia, that are also present. As another example, when a person with a weight problem learns behaviors that lead them to a fatty meal, they are also learning where that food is obtainable, and to connect the flavors of the food with their (too-desirable) caloric consequences. Behavior results because both the tendency to repeat behaviors is reinforced and because the Pavlovian signals for drugs or food can evoke approach. But there is more than this, as we will now see.

		Positive event	Negative event
CS signals	Increase in probability of event	*Approach CS*	*Withdraw from CS*
	Decrease in probability of event	*Withdraw from CS*	*Approach CS*

FIGURE 3.3–2. Sign tracking in Pavlovian learning. CSs either signal an increase or a decrease in the probability of good or bad events, and the CS generally engages approach or withdrawal behaviors accordingly.

PAVLOVIAN CONDITIONING

Effects of Conditioning on Behavior

As noted in an article by Bouton in 2001, where some of the following discussion was first presented, many laypeople have the mistaken impression that Pavlovian learning is a rigid affair in which a fixed stimulus comes to elicit a fixed response. In fact, conditioning is considerably more complex and dynamic than that. For example, signals for food may evoke a large set of responses that function to prepare the organism to digest food: They can elicit secretion of gastric acid, pancreatic enzymes, and insulin in addition to Pavlov's famous salivary response. The CS can also elicit approach behavior (as above), an increase in body temperature, and a state of arousal and excitement. When a signal for food is presented to a satiated animal or human, they may get up and eat more food. Some of these effects may be motivational; for example, an additional effect of presenting a CS for food is that it can invigorate ongoing operant behaviors that have been reinforced with food. CSs thus have powerful behavioral potential. Signals for food evoke a whole *behavior system* that is functionally organized to find, procure, and consume food.

As suggested earlier, Pavlovian conditioning is also involved in other aspects of eating. Through conditioning, humans and other animals may learn to like or dislike different foods. In animals like rats, flavors associated with nutrients (sugars, starches, calories, proteins, or fats) come to be preferred. Flavors associated with sweet tastes are also preferred, while flavors associated with bitter tastes are avoided (these tendencies may be seen as additional examples of the approach and withdrawal tendencies summarized in Figure 3.3–2). At least as important, flavors associated with illness become disliked, as illustrated by the person who gets sick drinking an alcoholic beverage and consequently learns to hate the flavor. The fact that flavor CSs can be associated with such a range of biologic consequences (USs) is important for omnivorous animals that need to learn about new foods. And it has some clinical implications. For example, chemotherapy can make cancer patients sick, and can therefore cause the conditioning of an aversion to a food that was eaten recently (or to the clinic itself). Other evidence suggests that animals may learn to dislike food that is associated with becoming sick with cancer itself. On the flip side, conditioning can enable external cues to trigger food consumption and craving, a potential influence on overeating and obesity.

Pavlovian conditioning, as noted in the previous section, also occurs when organisms ingest drugs. Whenever a drug is taken, in addition to reinforcing the behaviors that lead to its ingestion, the drug constitutes a US, and may be associated with potential CSs that are present at the time (rooms, odors, injection paraphernalia, etc.). CSs that are associated with drug USs can sometimes have an interesting property: They can elicit a conditioned response that seems opposite to the unconditional effect of the drug. For example, although morphine causes a rat to feel less pain, a CS associated with morphine elicits an opposite increase, not a decrease in pain sensitivity. Similarly, although alcohol can cause a drop in body temperature, a conditioned response to a CS associated with alcohol is typically an increase in body temperature. In these cases, the conditioned response is said to be *compensatory* because it counteracts the drug effect. Compensatory responses are another example of how classical (Pavlovian) conditioning helps organisms prepare for a biologically significant US.

Compensatory conditioned responses have implications for drug abuse. First, they can cause drug tolerance, in which repeated administration of a drug reduces its effectiveness. As a drug and a CS are repeatedly paired, the compensatory response to the CS becomes stronger and more effective at counteracting the effect of the drug. The drug therefore has less impact. One implication is that tolerance will be lost if the drug is taken without being signaled by the usual CS. Consistent with this idea, administering a drug in a new environment can cause a loss of drug tolerance and make drug overdose more likely. A second implication stems from the fact that compensatory responses may be unpleasant or aversive. A CS associated with an opiate may elicit several compensatory responses—it may cause the drug user to be more sensitive to pain, undergo a change in body temperature, and perhaps become hyperactive (the opposite of another unconditional opiate effect). The unpleasantness of these responses may motivate the user to take the drug again to get rid of them, an example of escape learning, or negative reinforcement (Fig. 3.3–1), and a classic example of how Pavlovian and operant learning processes might readily interact (see below). The idea is that the urge to take drugs may be strongest in the presence of CSs that have been associated with the drug. The hypothesis is consistent with self-reports of abusers who, after a period of abstinence, are tempted to take the drug again when they are re-exposed to drug-associated cues.

USs do not always cause the conditioning of compensatory responses, as is illustrated by the fact that Pavlov's CS for food caused the dog to salivate, rather than acquire a dry mouth. Moreover, not all drugs cause the conditioning of compensatory responses either; for example, when cocaine or amphetamine are associated with an environment, the environment tends to evoke a drug-like increase in activity rather than its opposite. Whether the CR appears similar or opposite to the unconditional effect of the drug (the UR) appears to depend on how the UR is actually mediated by the nervous system. Compensatory responses appear to develop as the CR when what we label as the "unconditional response" is actually independent of the brain or spinal cord. For example, although an injection of insulin will cause a drop in the level of blood glucose, it does so because of its direct effect on the body's cells, and not via a reaction by the nervous system. The "response" is actually a stimulus to the nervous system, which unconditionally responds by recruiting an increase in blood glucose. The CR that consequently develops to CSs associated with insulin can therefore be a compensatory increase in blood glucose. The CR, however, is the same as the brain's actual reaction—the UR, correctly identified.

USs, however, may have multiple effects on the body, some of which may be mediated by the nervous system, and some of which may not be. For example, an injection of atropine causes both dilation of the pupils and a drying of the mouth. The first effect is mediated by the brain, but the second effect is not; atropine suppresses activity of the salivary glands, which presumably initiates a brain reaction that compensates. Consistent with the aforementioned, a CS associated with injection of atropine will elicit a conditioned dilation of the pupils as well as salivation, the latter being a compensatory response.

Pavlovian learning may potentially be involved in anxiety disorders. We now know that CSs associated with frightening USs can elicit a whole system of conditioned fear responses, broadly designed to help the organism cope. In animals, cues associated with frightening events (such as a brief footshock) elicit changes in respiration, heart rate, and blood pressure, and even a (compensatory) decrease in sensitivity to pain. Brief CSs that occur close to the US in time can also elicit adaptively timed protective reflexes. For example, the rabbit blinks to a brief signal that predicts a mild electric shock near the eye. The same CS, when lengthened in duration and paired with the same US, elicits mainly fear responses, and fear elicited by a CS may potentiate the conditioned eyeblink response elicited by another CS or a startle response to a sudden noise. Once again, CSs do not merely elicit a simple reflex, but evoke a complex and interactive set of responses.

Classical fear conditioning can contribute to phobias (where specific objects may be associated with a traumatic US), as well as other anxiety disorders such as panic disorder and post-traumatic stress disorder (PTSD). In panic disorder, people who have unexpected panic attacks can become anxious about having another one. In this case, the panic attack (the US or UR) may condition anxiety to the external situation in which it occurs (e.g., a crowded bus) and also internal (“interoceptive”) CSs created by early symptoms of the attack (e.g., dizziness or a sudden pounding of the heart). These CSs may then come to evoke anxiety or panic responses. Panic disorder may begin because external cues associated with panic can arouse anxiety, which may then exacerbate the next unconditional panic attack and/or panic response elicited by an interoceptive CS. It is possible that the emotional reactions elicited by CSs may not require conscious awareness for their occurrence or development. Indeed, fear conditioning may be independent of conscious awareness.

As suggested earlier, in addition to eliciting conditioned responses, CSs also motivate ongoing instrumental or operant behavior, a phenomenon known as *Pavlovian-Instrumental transfer*. For example, presenting a CS that elicits anxiety can increase the vigor of operant behaviors that have been learned to avoid or escape the frightening US. Thus, an individual with an anxiety disorder will be more likely to express avoidance in the presence of anxiety or fear cues. Similar effects may occur with CSs that predict other USs (such as drugs or food)—as already mentioned, a drug-associated CS may motivate the drug abuser to take more drugs. The motivating effects of CSs may stem from the fact that CSs may be associated with both the sensory and emotional properties of USs. For example, the survivor of a traumatic train derailment might associate stimuli that occur immediately before derailment (such as the blue flash that occurs when the train separates from its overhead power supply) with both the emotional and the sensory aspects of the crash. Consequently, when the survivor later encounters another flashing blue light (e.g., the lights on a police car), the CS might evoke both emotional responses (mediated by association with the trauma's emotional qualities) and sensory associations (mediated by association with the trauma's sensory qualities). Both might play a role in the nightmares and "re-experiencing" phenomena that are characteristic of PTSD.

The Nature of the Learning Process

Research beginning in the late 1960s began to uncover some important details about the learning process behind Pavlovian conditioning. Several findings proved especially important. It was shown, for example, that conditioning is not an inevitable consequence of pairing a CS with a US. Such pairings will not cause conditioning if there is a second CS present that already predicts the US. This finding (known as *blocking*) suggests that a CS must provide new information about the US if learning is to occur. The importance of the CS's information value is also suggested by the fact that a CS will not be treated as a signal for the US if the US occurs equally often (or is equally probable) in the presence and the absence of the CS. Instead, the organism treats the CS as a signal for the US if the probability of the US is greater in the presence of the CS than in its absence. In addition, the organism will treat the CS as a signal for "no

US” if the probability of the US is less in the presence of the CS than in its absence. In the latter case, the signal is called a *conditioned inhibitor* because it will inhibit performance elicited by other CSs. The conditioned inhibition phenomenon is clinically relevant because inhibitory CSs may hold pathologic CRs like fear or anxiety at bay. A loss of the inhibition would allow the anxiety response to emerge.

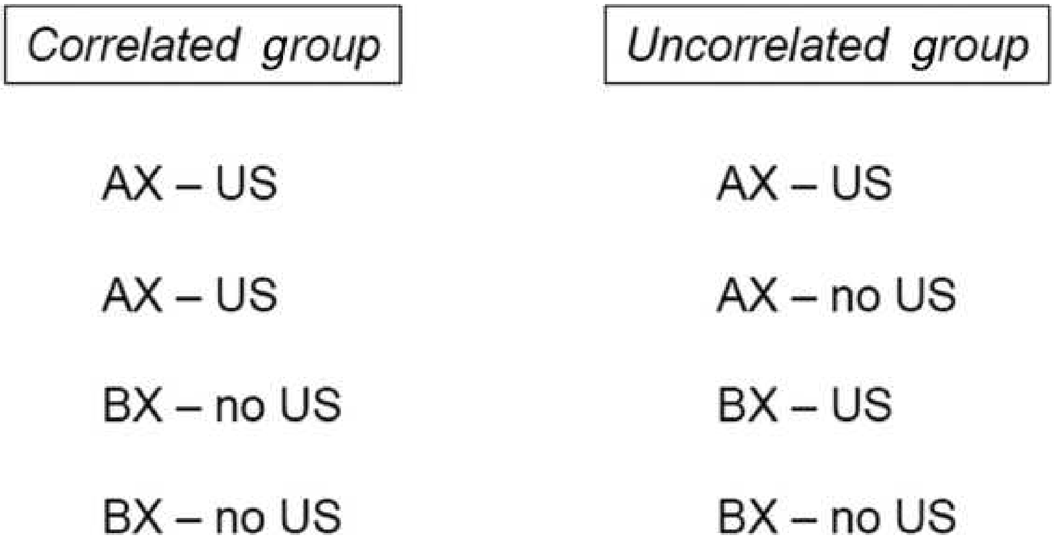

Correlated group	Uncorrelated group
AX – US	AX – US
AX – US	AX – no US
BX – no US	BX – US
BX – no US	BX – no US

FIGURE 3.3–3. Types of trials administered to different groups of subjects described in the text. For the correlated group, A is the most valid predictor of the US, and it consequently controls most conditioned responding—for example, blocking conditioning of X. For the uncorrelated group, despite the overall (but superficial) similarity in types of trials, all CSs are associated with the US on 50% of their presentations, and all CSs therefore come to control conditioned responding. (Because X is paired with the US more times than A or B, it is actually the best-conditioned CS in this arrangement.) Conditioning depends on the relative informativeness, validity, or predictive value of the CS.

Consistent with the role of “information value,” the conditioning process is organized so that conditioning mainly accrues to the most valid predictors of the US. For example, in the experimental design illustrated in Figure 3.3–3, two groups of subjects received subtly different treatments that had markedly different consequences on learning and behavior. For the correlated group, on some of the trials, CSs A and X were presented together (in a “compound”) and always paired with a US. On other trials, B and X occurred together, and the US was never presented. After a number of such AX and BX trials, the learning process treated A as the best predictor of the US—that stimulus blocked learning to X, despite the fact that X was paired with the US on half the occasions on which it was presented. In contrast, in the uncorrelated group, X was also paired with the US half the time, but the results were very different. For this group, X was also combined with cues A and B half the time, but A and B were no more predictive of the US than X (the US was also presented on half the trials with either A or B). Because X was as valid a predictor as A or B, there was substantial learning to X. (In fact, because X was actually paired with the US twice as often as A or B, it was the best predictor of the US.) The pattern of results has been observed over a very wide range of conditioning methods, from fear conditioning in rats, category learning in humans, and sucrose conditioning in honey bees. It seems consistent with the view that, to understand conditioning, “the organism is [best] seen as an information seeker using logical and perceptual relations among events … to form a sophisticated representation of its world.” The fact that conditioning tends to accrue to the most valid predictors of a US suggests that subtle differences in conditioning history can have a profound influence on an individual’s learning and behavior. Thus, if two different drug users were to take a drug in a manner analogous to that of the correlated and uncorrelated groups, one abuser would have tolerance and craving elicited almost exclusively by CS A, whereas the other would have tolerance and craving elicited strongly by A, B, and especially X.

To account for Pavlovian learning, most theorists now suppose that conditioning is not determined by CS–US pairings, but by the discrepancy between (1) the US predicted by all CSs present on a trial, and (2) the US that actually happens on the trial. That discrepancy is often referred to as *prediction error.* One implication is that, depending on the size and direction of the discrepancy or error, the pairing of a CS and traumatic US, for example, can cause an increase in fear conditioning, no change in conditioning, or even a decrease in conditioning. The latter idea has implications for understanding inhibition. A new CS can acquire a negative value if the US is smaller than that which other CSs present predict. Casually speaking, then, a conditioned inhibitor is really a CS that predicts “less US than expected.” And a conditioned excitor is a CS that predicts “more US than expected.”

These principles have been given formal power in the Rescorla–Wagner model, which asserts that the change in associative strength to a CS on any conditioning trial will be a function of the difference between the strength of the US that occurs on that trial and the extent to which the US is predicted by other CSs that are also present:

$$\Delta V_{CS} = \alpha\beta(\lambda - \Sigma V)$$

where Δ_{CS} is the change in associative strength (V) to the CS, λ is the strength of the US, and ΣV is the summed associative strengths of all other CSs that are present on the trial (the extent to which the US is already predicted). The α and β terms are merely fractions that influence the rate of learning by influencing the amount of associative change that will occur on the trial (they are based on the salience of the CS and US, respectively). In this theory, CSs can acquire either positive or negative values which correspond to excitation and inhibition, respectively. Several later theories have expanded this conceptualization and have emphasized a number of cognitive psychological processes, such as surprisingness of the US and the CS, short-term memory, memory priming, and also attention to the CS in accounting for this deceptively simple form of learning.

Classical conditioning is most robust if the CS and US are intense or salient. It is also best if the CS and US are novel. For example, in *latent inhibition*, repeated exposure to the CS alone before conditioning can diminish its ability to elicit responding when it is paired with the US. In the *US pre-exposure effect*, repeated exposure to the US before conditioning can likewise decrease the conditioning that later occurs when a CS and the US are paired. One idea is that the CS and the US must be “surprising” at the time of their pairing for learning to occur. Thus, the effects of pairing a CS with trauma or drug USs may depend in subtle ways on the individual’s prior experience with the CS and the US. Conditioning theories that followed the Rescorla–Wagner model captured this idea by emphasizing that CSs or USs that were already “primed” in short-term memory by either recent presentation or by a retrieval cue were not surprising and commanded less mental processing necessary for learning. Other theories have also emphasized a critical role for attention to the CS, which is thought to increase under circumstances where its consequence is not known and decrease when its consequence is known. Thus, with extended conditioning, the organism may respond to the CS automatically without much attention or mental processing devoted to it.

There are also important variants of classical conditioning. In *sensory preconditioning*, two stimuli (A and B) are first paired, and then one of them (A) is later paired with the US. Stimulus A evokes conditioned responding, of course, but so does stimulus B—indirectly, through its association with A. One implication is that exposure to a potent US like a panic attack may influence our reactions to stimuli that have never been paired with the US directly; the sudden anxiety to stimulus B might seem spontaneous and mysterious. A related finding is *second-order conditioning*. Here, A is paired with a US first and then subsequently paired with stimulus B. Once again, both A and B will evoke responding. Sensory preconditioning and second-order conditioning increase the range of stimuli that can control conditioned response. A third variant worth mentioning occurs, as indicated above, when the onset of a stimulus becomes associated with the rest of itself, as when a sudden increase in heart rate caused by the onset of a panic attack comes to predict the rest of the panic or feeling the onset of a drug may predict the rest of the drug effect. Such intraevent associations may play a role in many of the body's regulatory functions, such that an initial change in some variable (e.g., blood pressure or blood glucose level) may come to signal a further increase in that variable and therefore initiate a conditioned compensatory response.

Emotional responses can also be conditioned through observation. For example, a monkey that merely observes another monkey being frightened by a snake can learn to be afraid of the snake. The observer learns to associate the snake (CS) with its own emotional reaction (US/UR) to the other monkey being afraid. Although monkeys readily learn to fear snakes, they are less likely to associate other salient cues (such as colorful flowers) with fear in the same way. This is an example of *preparedness* in classical conditioning—some stimuli are especially effective signals for some USs because evolution has made them that way. Another example is the fact that tastes are easily associated with illness but not shock, whereas auditory and visual cues are easily associated with shock but not illness. Preparedness may explain why human phobias tend to be for certain objects (snakes or spiders) and not others (knives or electric sockets) that may as often be paired with pain or trauma.

Erasing Pavlovian Learning

If Pavlovian learning plays a role in the etiology of behavioral and emotional disorders, a natural question is how to eliminate it or undo it. As mentioned previously, Pavlov studied *extinction*: Conditioned responding decreases if the CS is presented repeatedly without the US after conditioning. Extinction is the basis of many behavioral or cognitive behavioral therapies designed to reduce pathologic conditioned responding through repeated exposure to the CS (*exposure therapy*), and it is presumably a consequence of any form of therapy in the course of which the patient learns that previous harmful cues are no longer harmful. Another elimination procedure is *counterconditioning*, in which the CS is paired with a very different US/UR. Counterconditioning was the inspiration for *systematic desensitization*, a behavior therapy technique in which frightening CSs are deliberately associated with relaxation during therapy.

Although extinction and counterconditioning reduce unwanted conditioned responses, they do not destroy the original learning, which remains in the brain, ready to return to behavior under the right circumstances. For example, conditioned responses that have been eliminated by extinction or counterconditioning can recover if time passes before the CS is presented again (*spontaneous recovery*). Conditioned responses can also return if the patient returns to the context of conditioning after extinction in another context or if the CS is encountered in a context that differs from the one in which extinction has occurred (all are examples of the *renewal effect*). The renewal effect is important because it illustrates the principle that extinction performance depends on the organism being in the context in which extinction was learned. If the CS is encountered in a different context, the extinguished behavior may relapse or return. Recovery and relapse can also occur if the current context is associated again with the US (“reinstatement”) or if the CS is paired with the US again (“rapid reacquisition”). One theoretical approach assumes that extinction and counterconditioning do not destroy the original learning, but instead entail new learning that gives the CS a second meaning (e.g., “the CS is safe” in addition to “the CS is dangerous”). As with an ambiguous word, which likewise has more than one meaning, responding evoked by an extinguished or counter-conditioned CS depends fundamentally on the current context.

Table 3.3–1.
Effective Contextual Stimuli Studied in Animal and Human Research Laboratories

- Exteroceptive context:
  - Room, place, environment, other external background stimuli
- Interoceptive context:
  - Drug state
  - Hormonal state
  - Mood state
  - Stress state
  - Deprivation state
  - Recent behaviors
  - Recent events
  - Expectation of events
  - Passage of time

Research on context effects in both animal and human learning and memory suggests that a wide variety of stimuli can play the role of context (Table 3.3–1). Drugs, for example, can be very salient in this regard. When rats are given fear extinction while under the influence of a benzodiazepine tranquilizer or alcohol, fear is renewed when the CS is tested in the absence of the context provided by the drug. This is an example of *state-dependent learning*, in which retention of information is best when tested in the same state in which it was originally learned. State-dependent fear extinction has obvious implications for combining therapy with drugs. It also has implications for the administration of drugs more generally. For example, if a person were to take a drug to reduce anxiety, the anxiety reduction would reinforce drug taking. State-dependent extinction might further preserve any anxiety that might otherwise extinguish during natural exposure to the anxiety-eliciting cues. Thus, drug use could paradoxically preserve the original anxiety, creating a self-perpetuating cycle that could provide a possible explanation of the link between anxiety disorders and substance abuse. One point of this discussion is that drugs can play multiple roles in learning: They can be USs or reinforcers on one hand, and CSs or contexts on the other. The possible complex behavioral effects of drugs are worth bearing in mind.

Another general message is that contemporary theory emphasizes the fact that extinction (and other processes like counterconditioning) entails new learning rather than a destruction of the old, an idea that has been incorporated in contemporary cognitive behavioral therapies that emphasize inhibitory learning during exposure. Recent psychopharmacologic research has also built on this idea: If extinction and therapy constitute new learning, then drugs that might facilitate new learning might also facilitate the therapy process. For example, there has been considerable recent interest in D-cycloserine, a partial agonist of the N-methyl-D-aspartate (NMDA) glutamate receptor. The NMDA receptor is involved in long-term potentiation, a synaptic facilitation phenomenon that has been implicated in several examples of learning. Of interest, there is evidence that administration of D-cycloserine can facilitate extinction learning in rats and possibly humans undergoing exposure therapy for anxiety disorders. In the studies supporting this conclusion, the administration of the drug increased the amount of extinction that was apparent after a small (and incomplete) number of extinction trials. Although such findings are promising, it is important to remember that the context dependence of extinction, and thus the possibility of relapse with a change of context, may easily remain. Consistent with this possibility, although D-cycloserine allows fear extinction to be learned in fewer trials, it does not appear to prevent or reduce the strength of the renewal effect. Such results further underscore the importance of behavioral research—and behavioral theory—in understanding the effects of drugs on therapy. Nonetheless, the search for drugs that might enhance the *learning* that occurs in therapy situations will continue to be an important area of research.

Another process that might theoretically modify or erase a memory is illustrated by a phenomenon called *reconsolidation*. Newly learned memories are temporarily labile and easy to disrupt before they are consolidated into a more stable form in the brain. The consolidation of memory requires the synthesis of new proteins and can be blocked by administration of protein synthesis inhibitors (e.g., anisomycin). Animal research suggests that consolidated memories that have recently been reactivated might also return briefly to a similarly vulnerable state; their "reconsolidation" can likewise be blocked by protein synthesis inhibitors. For example, several studies have shown that the reactivation of a conditioned fear by one or two presentations of the CS after a brief fear conditioning experience can allow it to be disrupted by anisomycin. When the CS is tested later, there is little evidence of fear—as if reactivation and then drug administration diminished the strength of the original memory. However, like the effects of extinction, these fear-diminishing effects do not necessarily mean that the original learning has been destroyed or erased. There is some evidence that fear of the CS that has been diminished this way can still return over time (i.e., spontaneously recover) or with reminder treatments. This sort of result suggests that the effect of the drug is to somehow interfere with retrieval or access to the memory, rather than an actual “reconsolidation.”

Generally speaking, the elimination of a behavior after therapy should not be interpreted as erasure of the underlying knowledge. For the time being, it may be safest to assume that after any therapeutic treatment, a part of the original learning may remain in the brain, ready to produce relapse if retrieved. Instead of trying to find treatments that destroy the original memory, another therapeutic strategy might be to accept the possible retention of the original learning and build therapies that allow the organism to prevent or cope with its retrieval. One possibility is to conduct extinction exposure in the contexts where relapse might be most problematic to the patient. Others are to conduct therapy in multiple contexts (to encourage generalization across contexts) and to encourage retrieval strategies (such as retrieval cues like reminder cards) that might help remind the patient of the therapy experience.

OPERANT/INSTRUMENTAL LEARNING

The Relation Between Behavior and Payoff

As noted earlier, operant learning has many parallels with Pavlovian learning. As one example, extinction also occurs in operant learning if the reinforcer is omitted following training. Although extinction is once again a useful technique for eliminating unwanted behaviors, just as we saw with Pavlovian learning, it does not destroy the original learning—spontaneous recovery, renewal, reinstatement, and rapid reacquisition effects still obtain. Another phenomenon known as *resurgence* also occurs: If a new behavior is taught to replace an old behavior while the old one undergoes extinction, the old behavior can return (“resurge”) when the replacement behavior is itself extinguished. All of these effects are consistent with the idea that extinction of an operant behavior, like a Pavlovian response, can be specific to the context in which it is learned.

Although early accounts of instrumental learning beginning with Thorndike emphasized the role of the reinforcer as “stamping in” the action, more modern approaches tend to view the reinforcer as a sort of guide or motivator of behavior. A modern, “synthetic” view of operant conditioning (see later discussion) holds that the organism associates the action with the outcome in much the way that stimulus–outcome learning is thought to be involved in Pavlovian learning.

Operant behavior has traditionally been studied in its own right, in part because so much human behavior seems voluntary, rather than directly elicited by antecedent cues. Human behavior is influenced by a wide variety of reinforcers, including social ones. For example, simple attention from teachers or hospital staff members has been shown to reinforce disruptive or problematic behavior in students or patients. In either case, when the attention is withdrawn and redirected toward other activities, the problematic behaviors can decrease (i.e., undergo extinction). Human behavior is also influenced by verbal reinforcers, like praise, and more generally, by *conditioned reinforcers,* such as money, which have no intrinsic value except for the value derived through association with more basic, “primary,” rewards. Conditioned reinforcers have been used in schools and institutional settings in so-called *token economies* in which positive behaviors are reinforced with tokens that can be used to purchase other valued items. In more natural settings, reinforcers are always delivered in social relationships, where their effects are dynamic and reciprocal. For example, the relationship between a parent and child is full of interacting and reciprocating operant contingencies in which the delivery (and withholding) of reinforcers and punishers shapes the behavior of each. Like Pavlovian learning, operant learning is always operating and always influencing behavior.

Research on operant conditioning in the laboratory has offered many insights into how action relates to its payoff. In the natural world, few actions are reinforced every time they are emitted; instead, most actions are reinforced only intermittently. There are several well-studied ways in which reinforcers can be scheduled intermittently. In a *ratio reinforcement schedule*, the reinforcer is directly related to the amount of work or responding that the organism emits. That is, there is some work requirement that determines when the next reinforcer will be presented. In a “fixed ratio schedule,” every $\underline{x}$th action is reinforced; in a “variable ratio schedule,” there is an average ratio requirement, but the number of responses required for each successive reinforcer varies. Ratio schedules, especially variable ratio schedules, can generate high rates of behavior, as one observes in behavior directed at a casino slot machine. In an *interval reinforcement schedule*, presentation of each reinforcer depends on the organism emitting the response after some period of time has also elapsed. In a “fixed interval schedule,” the first response after x seconds have elapsed is reinforced. In a “variable interval schedule,” there is an interval requirement for each reinforcer, but the length of that interval varies. A person checking for email throughout the day is being reinforced on a variable interval schedule—a new message is not present to reinforce each checking response, but the presence of a new message becomes available after variable timepoints throughout the day. Interestingly, on interval schedules, response rate can vary substantially without influencing the overall rate of reinforcement. (On ratio schedules, there is a more direct relationship between behavior rate and reinforcement rate.) In part because of this, interval schedules tend to generate slower response rates than ratio schedules.

Classic research on operant behavior underscores the fact that the performance of any action always involves *choice*. That is, whenever the individual performs a particular behavior, they choose to engage in that action over many other possible alternatives. When choice has been studied by allowing the organism to perform either of two different operant behaviors (paying off with their own separate schedules of reinforcement), the rate of operant behavior depends not only on the behavior’s own rate of reinforcement, but also on the rate of reinforcement of all other behaviors in the situation. Put most generally, the strength of Behavior 1 (e.g., the rate at which Behavior 1 is performed) is given by

$$B_1 = K^{*} R_1/(R_1 + R_O)$$

where B_1 can be seen as the strength of Behavior 1, R_1 is the rate at which B_1 has been reinforced, and R_O is the rate at which all alternative (or “other”) behaviors in the environment have been reinforced; K is a constant that corresponds to all behavior in the situation—and may have a different value for different individuals. This principle, known as the *quantitative law of effect,* captures several ideas that are relevant to psychiatrists and clinical psychologists. It indicates that an action can be strengthened either by increasing its rate of reinforcement (R_1) or by decreasing the rate of reinforcement for alternative behaviors (R_O). Conversely, an action can be weakened either by reducing its own rate of reinforcement (R_1) or by increasing the rate of reinforcement for alternative behaviors (R_O). The latter point has an especially important implication: In principle, one can slowdown the strengthening of new undesirable behavior by providing an environment that is otherwise rich in reinforcement (high R_O). Thus, an adolescent who experiments with drugs or alcohol would be less likely to engage in this behavior at a high rate (high B_1) if their environment were otherwise rich with reinforcers (e.g., provided by extracurricular activities, outside interests, and so forth). Figure 3.3–4 illustrates this point by showing how the rate of an action (B_1) is predicted to increase as a function of its rate of reinforcement in environments with high or low R_O. The principle can also be used in treatment. For example, *contingency management* treatments have been successful at reducing smoking and drug taking (among other behaviors) by providing vouchers and lottery tickets as reinforcers for alternative behaviors. One effect of such treatments is to increase R_O, although in this case the reinforcers are also usually provided only if the client abstains from the original behavior.

Choice between actions is also influenced by the size of their corresponding reinforcers and how soon the reinforcers will occur. For example, individuals sometimes have to choose between an action that yields a small but immediate reward (e.g., taking a hit of a drug) versus another that yields a larger but delayed reward (going to a class and earning credit toward a GED). Individuals who choose the more immediate reward are often said to be “impulsive,” whereas those who choose the delayed reward are said to exercise “self-control.” Of interest, organisms often choose immediate small rewards over delayed larger ones, even though it may be maladaptive to do so in the long run. Such “impulsive” choices are especially difficult to resist when reward is very imminent. A leading explanation is illustrated by Figure 3.3–5. Choice is thought to be determined by the relative value of the two rewards, with that value being influenced by both the reinforcer’s size and its delay. As the figure illustrates, the bigger the reinforcer the better is the value, and the more immediate the reinforcer the better too: When a reward is delayed, its value decreases or is “discounted” over time. When offered a choice, the organism will always choose the action that leads to the reward whose value is currently higher. Thus, if choice is made available at Time 1, when both rewards are coming soon, the value of the more immediate reward is higher than the value of the delayed one, and the organism may impulsively choose the immediate reward. If choice is available at Time 2, however, when the rewards are more distant in time, their relative values are reversed. The organism might now choose the delayed reward and demonstrate self-control. One implication is that “impulse control” can be achieved if choice (and commitment to that choice) can be made long before the rewards are about to happen. If the client can commit themself early (further left on the abscissa), it will be easier to choose the larger delayed reward. According to theory, other techniques that would help achieve impulse control would be to further increase the value of the delayed reward (e.g., by providing additional reinforcers for choosing it) and/or decrease the value of the immediate reward (e.g., by adding a cost or punishment for choosing it).

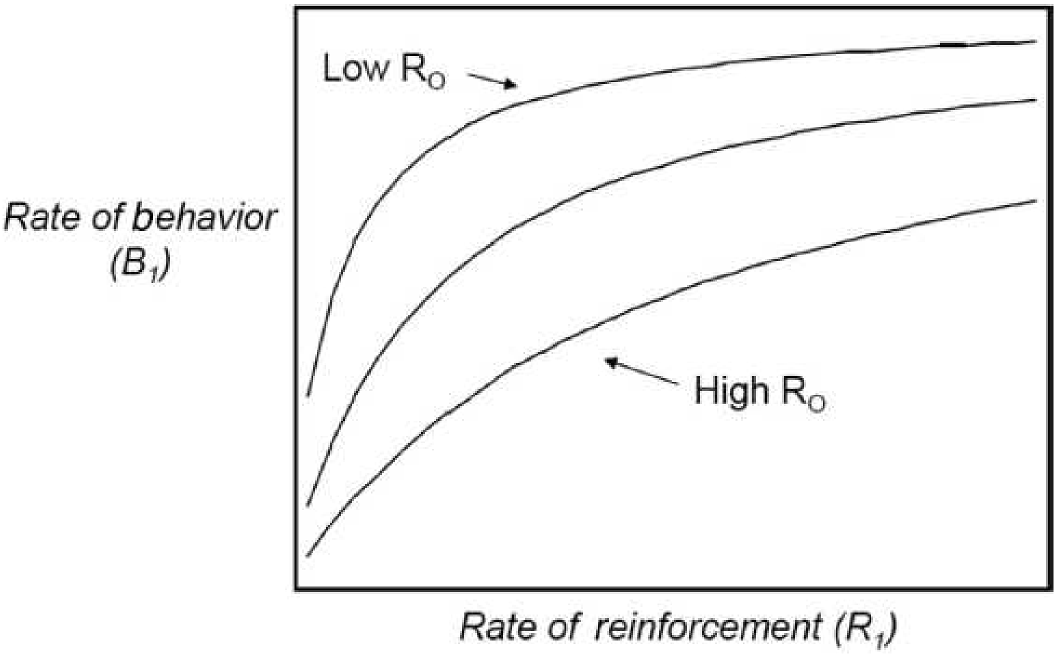

FIGURE 3.3–4. The quantitative law of effect predicts that the extent to which a behavior (B_1) is influenced by its reinforcement rate (R_1) depends crucially on the rate of reinforcement of other behaviors in the background (R_O).

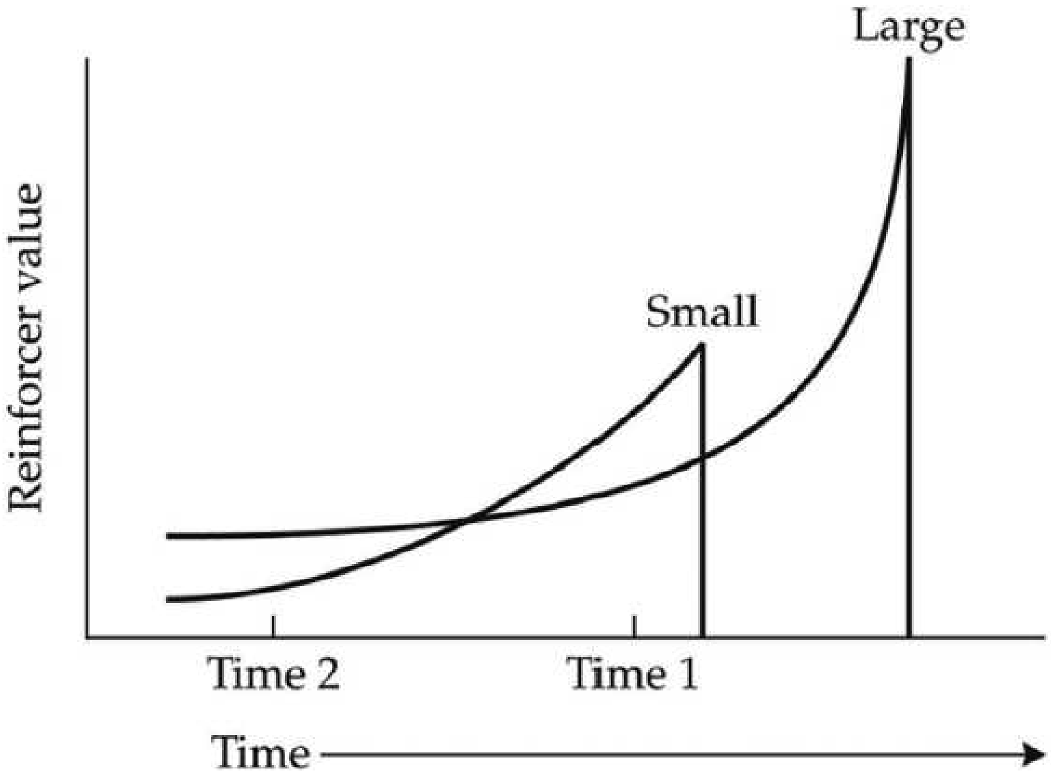

FIGURE 3.3–5. The current value of a reinforcer depends on both its magnitude and its delay. Choice is influenced by current value. Thus, at Time 1, a smaller and more immediate reward has more value than a large delayed reward. However, at Time 2, the larger reward has more value than the smaller reward, despite the same difference in their delay. Behavior is more "impulsive" if choice is made possible at Time 1 than Time 2. (Reproduced from Bouton ME. *Learning and Behavior: A Contemporary Synthesis*. 2nd ed. Copyright © 2007 by Sinauer Associates, Inc, with permission of Oxford University Press.)

Theories of Reinforcement

It is possible to use the foregoing principles of operant conditioning without knowing in advance what kind of event or stimulus will be reinforcing for the individual patient. None of the reinforcement rules described in the previous section say much about what sorts of events in an organism's world will play the role of reinforcer. B. F. Skinner was famously silent about this. He defined a reinforcer empirically, by considering the effect it had on an operant behavior. A reinforcer was defined as any event that could be shown to increase the strength of an operant if it was made a consequence of the operant. This empirical (some would say "atheoretical") view can be valuable because it allows idiosyncratic reinforcers for idiosyncratic individuals. For instance, if a therapist works with a child who is injuring himself, the approach advises the therapist merely to search for the consequences of the behavior and then manipulate them in order to bring the behavior under control. So if, for example, the child's self-injurious behavior decreases when the parent stops scolding him for doing it, then the scold is the reinforcer, which might seem counterintuitive to everyone (including the parent who thinks the scold should function as a punisher). On the other hand, it would also be useful to know what kind of event will reinforce an individual before the therapist has to try everything.

This void is filled by several approaches to reinforcement that allow predictions ahead of time. Perhaps the most useful is the *Premack principle* (named for researcher David Premack), which claims that, for any individual, reinforcers can be identified by giving the individual a preference test in which they are free to engage in any number of activities. The individual might spend the most time engaging in activity A, the second-most time engaged in activity B, and the third-most time engaged in activity C. Behavior A can thus said to be preferred to B and C, and B is preferred to C. The Premack principle asserts that access to a preferred action will reinforce any action that is less preferred. In the example above, if doing activity C allowed access to doing A or B, activity C will be reinforced—it will increase in strength or probability. Similarly, activity B will be reinforced by activity A (but not C). The principle accepts large individual differences. For example, in an early study, some children given a choice spent more time eating candy than playing pinball, while others spent more time playing pinball than eating candy. Candy eating reinforced pinball playing in the former group. In contrast, pinball playing reinforced candy eating in the latter group. There is nothing particularly special about food (eating) or any particular kind of activity as a possible reinforcer. Any behavior that is preferred to a second behavior will theoretically reinforce the second behavior.

The principle has been refined over the years. At the present point in time, it is recognized that even a less-preferred behavior can reinforce a more-preferred behavior if the organism has been deprived of doing the low-preferred behavior below its ordinary level. In the foregoing example, even the low-preference activity C could reinforce A or B if it were suppressed for a while below its baseline level of preference. However, the main implication is that, in the long run, a person's reinforcers can be discovered by simply looking at how they allocate their activities when access to them is free and unconstrained.

Motivational Factors

Instrumental action is often said to be goal oriented. As Edward Tolman illustrated in many experiments conducted in the 1930s and 1940s, organisms may flexibly perform any of several actions in order to get to a goal; instrumental learning thus provides a variable means to a fixed end. Tolman's perspective on the effects of reinforcers has returned to favor recently. He argued that reinforcers are not necessary for learning, but instead are important for motivating instrumental behavior. The classic illustration of this point is the *latent learning* experiment. Rats received several trials in a complex maze in which they were removed from the maze without reward once they got to a particular goal location. When arriving at the goal was suddenly rewarded, the rats suddenly began working through the maze with very few errors. Thus, they had learned about the maze without benefit of the food reinforcer, but the reinforcer was nonetheless important for motivating them to get through the maze efficiently. The reinforcer was not necessary for learning, but it gave the organism a reason to translate its knowledge into action.

Subsequent research has identified many motivating effects of reward. For example, organisms that have had experience receiving a small reward may show *positive contrast,* when they are suddenly reinforced with a larger reward. That is, their instrumental behavior may become more vigorous than that in control subjects who have received the larger reward all along. Conversely, organisms show *negative contrast* when they are switched from a high reward to a lower reward—their behavior becomes weaker than control subjects who have received the same smaller reward all along. Negative contrast involves frustration and emotionality. Both types of contrast are consistent with the idea that the current effectiveness of a reinforcer depends on what the organism has learned to expect; an increase from expectation causes elation, whereas a decrease from expectation causes frustration. There is a sense in which receiving a reward that is smaller than expected might actually seem punishing.

Negative contrast is an example of a *paradoxical reward effect*—a set of behavioral phenomena given the name because they show that rewards can sometimes weaken behavior and that nonreward can sometimes strengthen it. The best known is the *partial reinforcement extinction effect*, in which actions that have been intermittently (or "partially") reinforced persist longer when reinforcers are completely withdrawn than actions that have been continuously reinforced. The finding is considered paradoxical because an action that has been reinforced (say) half as often as another action may nonetheless be more persistent. One explanation is that the action that has been partially reinforced has been reinforced in the presence of some frustration—and is thus persistent in new adversity or sources of frustration. Other evidence suggests that effortfulness is a dimension of behavior that can be reinforced. That is, human and animal participants that have been reinforced for performing effortful responses learn a sort of "industriousness" that transfers to new behaviors. One implication is that new behaviors learned in therapy will be more persistent over time if high effort has been deliberately reinforced.

The effectiveness of a reinforcer is also influenced by the organism's current motivational state. For example, food is more reinforcing for a hungry organism and water is more reinforcing for a thirsty one. Such results are consistent with many theories of reinforcement (e.g., the Premack principle), because the presence of hunger or thirst would undoubtedly increase the organism's preference ranking for food or water. But recent research indicates that effects of motivational states on instrumental actions are not this automatic. Specifically, if a motivational state is going to influence an instrumental action, the individual needs first to learn how the action's reinforcer will influence the motivational state. The process of learning about the effects the reinforcer has on the motivational state is called *incentive learning*.

Incentive learning is best illustrated by an experimental example. In 1992, Balleine trained rats that were not hungry to lever press to earn a novel food pellet. The animals were then food deprived and tested for their lever pressing under conditions where the lever press no longer produced the pellet. The hunger state had no impact on lever press rate; that is, hungry rats did not lever press any more than rats that were not food deprived. On the other hand, if the rat had been given separate experience eating the pellets while it was food deprived, during the test it lever pressed at a high rate. Thus, hunger invigorated the instrumental action only if the animal had previously experienced the reinforcer in that state—which allowed it to learn that the specific substance influenced the state (incentive learning). The interpretation of this result, and others like it, will be developed further later in this section. The main idea is that individuals will perform an instrumental action when they know that it produces an outcome that is desirable in the current motivational state. The clinical implications are underexplored but could be significant. For example, persons who abuse drugs will need to learn that the drug makes them feel better in the withdrawal state before withdrawal will motivate drug seeking. Persons suffering from anxiety might not be motivated to take a beneficial medication while anxious until they have actually had the opportunity to learn how the medication makes them feel when they are in the anxious state, and a person suffering from depression may need to learn what natural reinforcers really make them feel better while they are depressed. According to theory, direct experience with a reinforcer's effect on depressed mood might be necessary before the person will be interested in performing actions that help ameliorate the depressed state.

PAVLOVIAN AND OPERANT LEARNING TOGETHER

Avoidance Learning

From the previous discussion, it is easy to appreciate that there is more going on in operant learning than merely associating an action with a reinforcer. Indeed, theories of the motivating effects of reinforcers have usually emphasized the fact that Pavlovian CSs in the background are also associated with the reinforcer, and that the expectancy of the reinforcer (or conditioned motivational state) they arouse increases the vigor of the operant response. This is *two-factor* or *two-process theory:* Pavlovian learning occurs simultaneously and motivates behavior during operant learning.

The interaction of Pavlovian and instrumental factors is especially important in understanding avoidance learning (Fig. 3.3–1). In avoidance situations, organisms learn to perform actions that prevent the delivery or presentation of an aversive event. The explanation of avoidance learning is subtle because it is difficult to identify an obvious reinforcer. Although preventing the occurrence of the aversive event is obviously important, how can the nonoccurrence of that event reinforce? The answer is that cues in the environment (Pavlovian CSs) come to predict the occurrence of the aversive event, and consequently arouse anxiety or fear. The avoidance response can therefore be reinforced if it escapes from or reduces that fear. Pavlovian and operant factors are thus both important: Pavlovian fear conditioning motivates and allows reinforcement of an instrumental action through its reduction. Escape from fear or anxiety is thought to play a significant role in many human behavior disorders, including the anxiety disorders. Thus, the obsessive-compulsive patient checks or washes their hands repeatedly to reduce anxiety, the agoraphobic stays home to escape fear of places associated with panic attacks, and the bulimic learns to vomit after a meal in order to reduce the learned anxiety evoked by eating the meal.

Although two-factor theory remains an important view of avoidance learning, excellent avoidance can be obtained in the laboratory without reinforcement. Specifically, if an animal is required to perform an action that resembles one of its natural and prepared fear responses—so-called *species-specific defensive reactions* (*SSDRs*)—it will quickly learn to avoid. For example, rats will readily learn to freeze (remain motionless) or flee (run to another environment) to avoid shock, two behaviors that have evolved to escape or avoid predation. Freezing and fleeing are also respondents rather than operants; they are controlled by their antecedents (Pavlovian CSs that predict shock) rather than being reinforced by their consequences (escape from fear). Thus, when the rat can use an SSDR to avoid, the only necessary learning is Pavlovian—the rat learns about environmental cues associated with danger, and these arouse fear and evoke natural defensive behaviors including withdrawal (negative sign tracking; Fig. 3.3–2). To learn to perform an action that is not similar to a natural SSDR requires more feedback or reinforcement through fear reduction. A good example is lever pressing, which is easy for the rat to learn when the reinforcer is a food pellet but difficult to learn when the same action avoids shock. More recent work with avoidance in humans suggests an important role for CS–aversive event and response–no aversive event expectancies. The larger point is that Pavlovian learning is important in avoidance learning; when the animal can avoid with an SSDR, it is the only learning necessary; when the required action is not an SSDR, Pavlovian learning permits the expectation of something bad.

A cognitive perspective on aversive learning is also encouraged by studies of *learned helplessness.* In this phenomenon, organisms exposed to either controllable or uncontrollable aversive events differ in their reactivity to later aversive events. For example, the typical finding is that a subject exposed to inescapable shock in one phase of an experiment is less successful at learning to escape shock with an altogether new behavior in a second phase, whereas subjects exposed to escapable shock are normal. Both types of subjects are exposed to the same shock, but its psychological dimension (its controllability) creates a difference, perhaps because subjects exposed to inescapable shock learn the independence of their actions and the outcome. Although this finding (and the interpretation that the inescapably shocked organism actively learned "helplessness") was once seen as a model of depression, the current view is that the ability to control stressors primarily reduces their stressfulness and modulates their negative impact.

One of the main implications of work on aversive learning is that there are both biologic (i.e., evolutionary as well as physiologic) and cognitive dimensions to instrumental learning. The possibility that so much instrumental learning in the laboratory can be controlled by Pavlovian contingencies, as suggested by the role of SSDRs, is also consistent with research in which animals have learned to respond for positive reinforcers. For example, pigeon subjects have been widely used in operant learning experiments since the 1940s. In the typical experiment, the bird learns to peck at a plastic disk on a wall of the chamber (a response "key") in order to earn food. Although pecking seems to be an operant response, it turns out that the pigeon's peck can be entrained by merely illuminating the key for a few seconds before presenting the reinforcer on a number of trials. Although there is no requirement for the bird to peck the key, the bird will begin to peck at the illuminated key—a Pavlovian predictor of food—anyway. It does so even if the experimenter arranges things so that pecks actually prevent delivery of food (which is otherwise delivered on trials without pecks); here the birds will continue to peck over many trials. (Although the peck has a negative correlation with food, key illumination remains a weakly positive predictor of food.) Thus, this classic "operant" behavior is at least partly a Pavlovian one. Pavlovian contingencies cannot be ignored. When rats are punished with mild footshock for pressing a lever that otherwise produces food, they stop lever pressing at least partly (and perhaps predominantly) because they learn that the lever now predicts shock and withdraw from it. A child might likewise learn to stay away from the parent who delivers punishment, rather than refrain from performing the punished behavior. A great deal of behavior in operant learning settings may actually be controlled by Pavlovian learning and sign tracking rather than true operant learning.

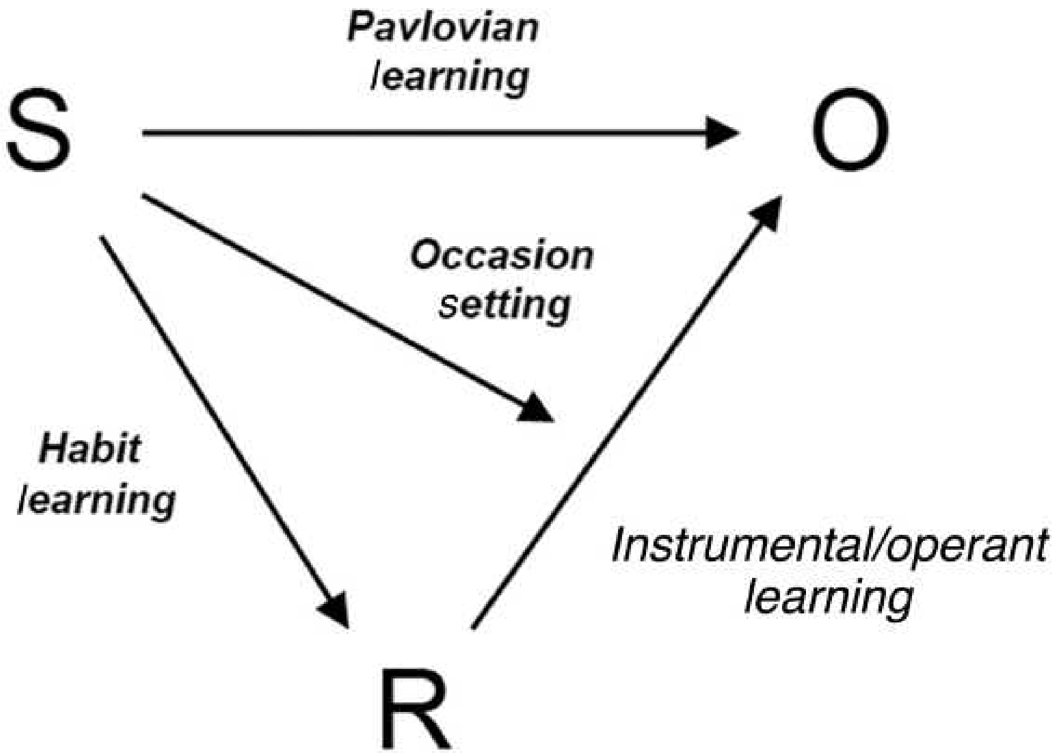

FIGURE 3.3–6. Any instrumental/operant learning situation permits a number of types of learning, which are always occurring all the time. S, stimulus in the background; R, operant behavior or instrumental action; O, biologically significant outcome (e.g., reinforcer, US, etc.).

A Synthetic View of Instrumental Action

The idea, then, is that behavior in any instrumental learning situation is controlled by several hypothetical associations that are illustrated in Figure 3.3–6. As just discussed, much behavior in an instrumental learning arrangement can be controlled by a Pavlovian factor, in which the organism associates background cues (CSs) with the reinforcing outcome (O). As discussed earlier, this type of learning can allow the CS to evoke a variety of behaviors as well as emotional reactions (and motivational states) that can additionally motivate instrumental action.

In modern terms, the instrumental factor is represented by the organism learning a direct (and similar) association between the instrumental action (R) and the reinforcer (O). Evidence for this sort of learning comes from experiments on *reinforcer devaluation* (Fig. 3.3–7). In such experiments, the organism can first be trained to perform two instrumental actions (e.g., pressing a lever and pulling a chain), each paired with a different reinforcer (e.g., food pellet vs. a liquid sucrose solution). In a separate second phase, one of the reinforcers (e.g., pellet) is paired with illness, which creates the conditioning of a powerful taste aversion to the reinforcer. In a final test, the organism is returned to the instrumental situation and is allowed to perform either instrumental action. No reinforcers are presented during the test. The result is that the organism no longer emits the action that produced the reinforcer that is now aversive. In order to perform this way, the organism must have (1) learned which action produced which reinforcer and (2) combined this knowledge with the knowledge that it no longer likes or values that reinforcer. The result cannot be explained by the simpler, more traditional view that reinforcers merely stamp in or strengthen instrumental actions.

We have already seen that organisms also need to learn how reinforcers influence a particular motivational state—the process called "incentive learning." Incentive learning is crucially involved in instrumental learning as a process through which the animal learns the value of the reinforcer. Thus, in the reinforcer devaluation experiment shown in Figure 3.3–7, an important thing that occurs in the second phase is that the organism must actually contact the reinforcer and learn that it does not like it. As described earlier, incentive learning is probably always involved in making outcomes (and the associated actions that produce them) more or less desirable.

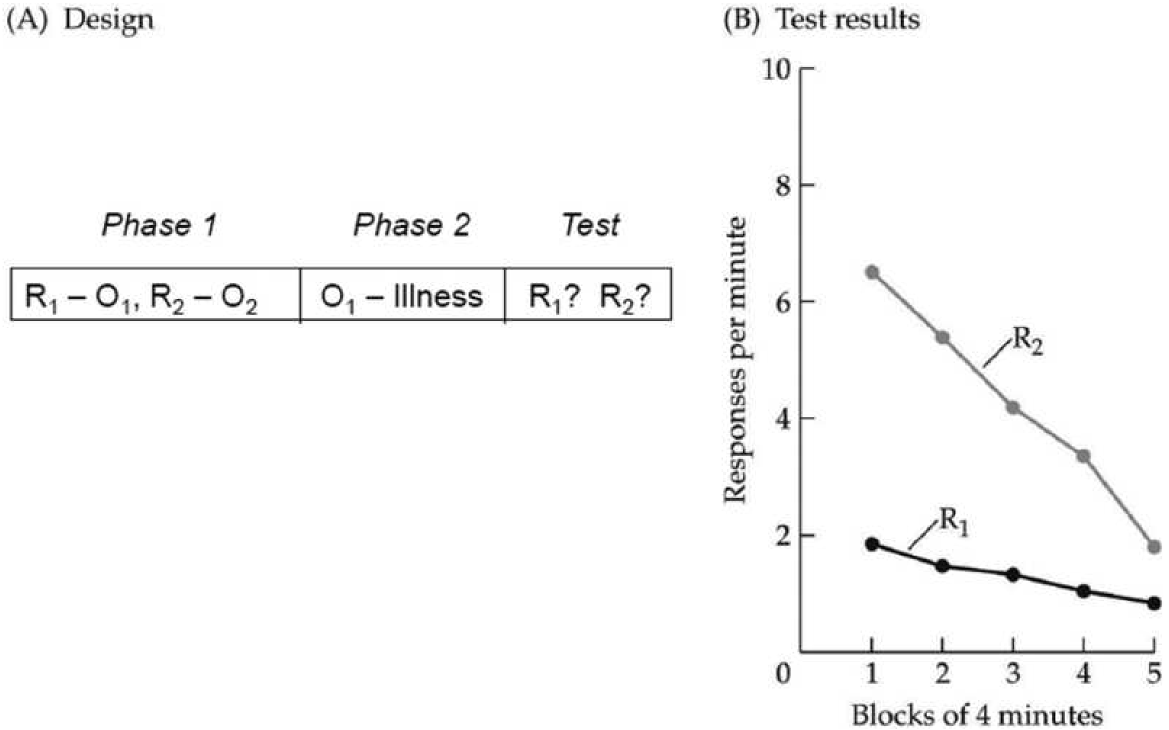

FIGURE 3.3–7. The reinforcer devaluation effect. **A:** Experimental design (see text). **B:** Results of the test session. The result indicates the importance of the response–reinforcer association in operant learning. For the organism to perform the way it does during testing, it must learn which action leads to which reinforcer and then choose to perform the action that leads to the outcome it currently liked or valued. (Reproduced from Bouton ME. *Learning and Behavior: A Contemporary Synthesis*. 2nd ed. Copyright © 2007 by Sinauer Associates, Inc, with permission of Oxford University Press. Data from Colwill RM, Rescorla RA. Associative Structures in Instrumental Learning. In: Bower GH, ed. *The Psychology of Learning and Motivation, Vol. 20.* Academic Press; 1986:55–104.)

The Pavlovian factor is represented by the S-O association in Figure 3.3–6. As noted earlier, whenever organisms earn reinforcers, they can also associate cues in the environment with them. Such S-O (Pavlovian) learning can have the many effects on behavior that were reviewed earlier: S may evoke approach behaviors, and S may elicit many other responses that generally get the organism ready for O. S may also remind the organism of O—it serves as a retrieval cue. In addition, as also mentioned earlier, the presentation of S may excite or motivate the instrumental response through Pavlovian-instrumental transfer: When a CS that is associated with a US (O) is presented while an organism performs an instrumental action that earns O, the rate of responding increases. CSs that predict the O (excitors) increase the rate of the instrumental behavior, whereas CSs that predict the nonoccurrence of O (inhibitors) decrease it. An S that predicts O can excite instrumental behaviors that lead to other Os that are motivationally similar—such as two different types of food. On the other hand, CSs are especially likely to increase responses that are associated with the *same* O. Thus, in addition to motivating instrumental responding in a general way, Pavlovian cues can influence choice of some actions over others. This capacity is presumably exploited by advertisers who create Ss in the form of television commercials, for example, whose presentation can motivate the consumer to buy a particular brand of pizza or a particular brand of beer.

Other experiments have illustrated the other associations to the stimulus that are represented in Figure 3.3–6. In addition to being directly associated with the reinforcer, a stimulus can signal a relation between an action and an outcome. This is called *occasion setting*: Instead of eliciting a response directly, stimuli in operant situations can set the occasion for the operant response. There is good evidence that they do so by signaling a specific response–reinforcer relationship. For example, in one experiment, rats learned to lever press and chain pull in the presence of a background noise and a background light. When the noise was present, lever pressing yielded a pellet reinforcer and chain pulling yielded sucrose. In contrast, when the light was present, the relations were reversed: Lever pressing yielded sucrose and chain pulling yielded pellet. There was evidence that the subjects (rats) learned the corresponding relationships. In a second phase, pellets were associated with illness, so the rat no longer valued the pellet. In a final test, the rats were allowed to lever press or chain pull in extinction in the presence of noise or light present during separate tests. In the presence of the noise, the animals chain pulled rather than lever pressed. When the light was present, the animals lever pressed rather than chain pulled. Thus, the noise informed the rat that lever press yielded pellet, and the light informed the rat that the chain pull did. This is the occasion-setting function illustrated in Figure 3.3–6.

It is worth noting that other stimuli besides lights and noises set the occasion for operant behavior. Modern research on learning in animals has underscored the importance of other stimuli such as contexts (see Table 3.3–1), temporal and spatial cues, and certain perception and memory processes. A particularly interesting example of research on the *stimulus control* of operant behavior is categorization. Pigeons can be shown images of different cars, chairs, flowers, and humans on a computer screen positioned on the wall of a Skinner box. Pecking one of four keys in the presence of these images is reinforced in the presence of any picture containing a car, a chair, a flower, or a human. Of interest, as the number of exemplars in each category increases, the pigeon makes more errors as it learns the discrimination. However, more exemplars creates better learning in the sense that it is more ready to transfer to new test images—after many examples of each category, the pigeon is more accurate at categorizing (and responding accurately to) new stimuli. One implication is that training new behaviors in a variety of settings or ways will enhance generalization to new situations.

The final association in Figure 3.3–6 is simple *habit learning*, or a direct association between the stimulus and the response. Through this association, triggering cues in the background may elicit the instrumental action quite directly, without the intervening cognition of R-O and the valuation of O. Although S-R learning was once thought to dominate learning, the current Zeitgeist envisions it as developing only after extensive and consistent practice. In effect, actions that have been performed repeatedly (and repeatedly associated with the reinforcer) become automatic and routine. Habit learning has been recognized by psychologists dating back (at least) to William James in the 19th century. It is thought to be functional, because automatic, “mindless” responses can free the behaving individual to process other things besides moment-to-moment behavior. Of course, habits can also become maladaptive and are thought to play a role in obsessive-compulsive and addictive behaviors. Interestingly, habit learning can be hastened in laboratory animals by exposure to many drugs of abuse.

Habits can be distinguished from goal-directed actions (those that depend on processing R-O) experimentally. One source of evidence is the fact that the reinforcer devaluation effect—which implicates a kind of cognitive mediation of operant behavior—no longer occurs after extensive instrumental training, as if the animal reflexively engages in the response without remembering the actual outcome it produces. It seems reasonable to expect that many pathologic behaviors that come into the clinic might also be automatic and blind through repetition. Interestingly, the evidence suggests that the eventual dominance of the S-R habit in behavior does not replace or destroy more cognitive mediation by learned S-O, R-O, and/or S-(R-O) relations. Under some conditions, even a habitual response might be brought back under the control of the action–reinforcer association and become goal-directed again. The conversion of actions to habits, the role of habit in addiction, and the relation of habit to cognition is currently an active area of research.

ACKNOWLEDGMENT

Preparation of the manuscript was supported by Grants R01 MH064847 and RO1 DA033123 from the National Institute of Health.

Further Readings

Balleine B. Instrumental performance following a shift in primary motivation depends on incentive learning. *J Exp Psychol Anim Behav Process*. 1992;18:236–250.

Balleine BW, O'Doherty JP. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*. 2010;35:48–69.

Bolles RC. Species-specific defense reactions and avoidance learning. *Psychol Rev*. 1970;71:32–48.

Bouton ME. Classical conditioning and clinical psychology. In: Smelser NJ, Baltes PB, eds. *International Encyclopedia of the Social and Behavioral Sciences.* Vol. 3. Elsevier Science; 2001. 1942–1945.

Bouton ME. Context, attention, and the switch between habit and goal-direction in behavior. *Learn Behav*. 2021;49:349–362.

**Bouton ME. *Learning and Behavior: A Contemporary Synthesis*, 2nd ed. Sinauer Associates, Inc. (now Oxford University Press); 2016.

Bouton ME, Maren S, McNally GP. Behavioral and neurobiological mechanisms of Pavlovian and instrumental extinction learning. *Physiol Rev*. 2021;101:611–681.

**Bouton ME, Mineka S, Barlow DH. A modern learning theory perspective on the etiology of panic disorder. *Psychol Rev*. 2001;108:4–32.

Cain CK. Avoidance problems reconsidered. *Curr Opin Behav Sci*. 2019;26:9–17.

Colwill RM, Rescorla RA. Evidence for the hierarchical structure of instrumental learning. *Anim Learn Behav*. 1990;18:71–82.

Craske MG, Treanor M, Conway CC, Zbozinek T, Vervliet, B. Maximizing exposure therapy: an inhibitory learning approach. *Behav Res Ther*. 2014;58:10–23.

Davis DR, Kurti AN, Skelly JM, Redner R, White TJ, Higgins ST. A review of the literature on contingency management in the treatment of substance use disorders, 2009–2016. *Prev Med*. 2016;92:36–46.

Davis M, Ressler K, Rothbaum BO, Richardson R. Effects of D-cycloserine on extinction: translation from preclinical to clinical work. *Biol Psychiatry*. 2006;60:369–375.

Dworkin BR. *Learning and Physiological Regulation*. University of Chicago Press;1993.

Haselgrove M, Hogarth L. *Clinical Applications of Learning Theory*. Psychology Press; 2012.

Hogarth L, Balleine BW, Corbit LH, Killcross S. Associative learning mechanisms underpinning the transition from recreational drug use to addiction. *Ann NY Acad Sci*. 2013;1282:12–24.

Maier SF, Seligman MEP. Learned helplessness at fifty: insights from neuroscience. *Psychol Rev*. 2016;123:349–367.

McDowell JJ. On the theoretical and empirical status of the matching law and matching theory. *Psychol Bull*. 2013;139:1000–1028.

McSweeney FK, Murphy ES. *The Wiley Blackwell Handbook of Operant and Classical Conditioning*. John Wiley & Sons;2014.

Nader K, Hardt O. A single standard for memory: the case for reconsolidation. *Nat Rev Neurosci*. 2009;10:224–234.

Ohman A, Mineka S. Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychol Rev*. 2001;108:483–522.

**Rescorla RA. Pavlovian conditioning: it's not what you think it is. *Am Psychol*. 1988;43:151–160.

Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF, eds. *Classical Conditioning II*. Appleton-Century-Crofts;1972.

Saunders BT, Robinson TE. Individual variation in resisting temptation: implications for addiction. *Neurosci Biobehav Rev*. 2013;37:1955–1975.

Siegel, S. Pharmacological conditioning and drug effects. In: Goudie AJ, Emmett-Oglesby MW, eds. *Psychoactive Drugs: Tolerance and Sensititzation. Contemporary Neuroscience*. Humana Press; 1989:115–180.

Spreat, S, & Spreat, SR. Learning principles. In: Voith V, Borchelt PL, eds. *Veterinary Clinics of North America: Small Animal Practice*. W. B. Saunders; 1982:593–606.

Wagner AR, Brandon SE. Evolution of a structured connectionist model of Pavlovian conditioning (AESOP). In: Klein SB, Mowrer RR, eds. *Contemporary Learning Theories: Pavlovian Conditioning and the Status of Traditional Learning Theory*. Lawrence Erlbaum Associates, Inc; 1989:149–189.

Wagner AR, Logan FA, Haberlandt K, Price T. Stimulus selection in animal discrimination learning. *J Exp Psychol*. 1968;76:171–180.

Wasserman EA, Bhatt RS. Conceptualization of natural and artificial stimuli by pigeons. In: Honig WK, Fetterman JG, eds. *Cognitive Aspects of Stimulus Control*. Lawrence Erlbaum, Inc; 1992:203–223.

Wolpe J. *Psychotherapy by Reciprocal Inhibition*. Stanford University Press; 1958.

3.4 Biology of Memory

Shruti Dave, Ph.D., and Joel L. Voss, Ph.D.

The topic of memory is fundamental to the discipline of psychiatry. Memory is the glue that binds mental life, the scaffolding for personal history. Personality is in part an accumulation of habits that have been acquired, many very early in life, which create dispositions and influence behavior. In the same sense, many psychiatric illnesses often involve learning—anxieties, phobias, and maladaptive behaviors that result from particular experiences. Psychotherapy itself is a process by which new habits and skills are acquired through the accumulation of new experiences. In this sense, memory is at the theoretical heart of psychiatry's concern with personality, the consequences of life experience, and the possibility of growth and change.

Memory is also of clinical interest because disorders of memory and complaints about memory are common in neurologic and psychiatric illnesses, such as dementia and schizophrenia. Memory impairment is also a side effect of certain treatments, such as electroconvulsive therapy. Accordingly, the effective clinician needs to understand something of the biology of memory, the varieties of memory dysfunction, and how memory can be evaluated.

MULTIPLE TYPES OF MEMORY

What is ordinarily meant by the term memory in everyday language typically falls within the definition of *declarative memory*. Declarative memory supports the conscious recollection of facts and events, including memory for episodes of experience, routes, lists, faces, melodies, objects, odors, and other verbal and nonverbal material across multiple sensory modalities. However, memory is not a single faculty of the mind but consists of various subtypes. Skill learning, habit learning, and simple forms of conditioning are included in a collective set of abilities termed *nondeclarative memory*.

Initial conceptualizations of human memory systems were derived from patients with amnesia. Individuals with amnesia typically show specific impairments of declarative memory alongside broad preservation of nondeclarative memory. Therefore, early information about the brain regions supporting acquisition and access to memories spotlighted declarative memory.

Declarative Memory in Amnesia

The hallmark of *amnesia* is a loss of new learning ability that extends across all sensory modalities and stimulus domains. This *anterograde amnesia* can be explained by understanding the role of brain structures critical for acquiring information about facts and events. Typically, anterograde amnesia occurs together with *retrograde amnesia,* a loss of knowledge that was acquired before the onset of amnesia. Retrograde deficits often have a temporal gradient, following a principle known as the Ribot law; deficits are most severe for information that was most recently learned. A patient with a presentation of amnesia exhibits severe memory deficits in the context of preservation of other cognitive functions, including language comprehension and production, reasoning, attention, immediate memory, personality, and social skills. The selectivity of the memory deficit in these cases implies that intellectual and perceptual functions of the brain are separated from the capacity to maintain in memory the records that ordinarily result from engaging in intellectual and perceptual work.

The Anatomy of Anterograde Amnesia

Studies of a patient with severe amnesia, known as HM, stimulated intensive investigation of the role of the *medial temporal lobe* in memory. HM became amnesic in 1953, at 27 years of age, when he sustained a bilateral resection of the medial temporal lobe to relieve severe epilepsy. The removal included approximately one-half of the hippocampus, the amygdala, and most of the neighboring entorhinal and perirhinal cortices (Fig. 3.4–1). After the surgery, HM's seizure condition was much improved, but he experienced profound forgetfulness. His intellectual functions were generally preserved. For example, HM exhibited normal immediate memory, and he could maintain his attention during conversations. After an interruption, however, HM could not remember what had recently occurred. HM's dense amnesia was permanent and debilitating. In HM's words, he felt as if he were just waking from a dream because he had no recollection of what had just taken place.

Another well-characterized patient, EP, was diagnosed with herpes simplex encephalitis at 72 years of age. Damage to the medial temporal lobe region (Fig. 3.4–1) produced persistent and profound amnesia. During testing sessions, EP was cordial and talked freely about his life experiences, but he relied exclusively on stories from his childhood and early adulthood. He would repeat the same story many times. Strikingly his performance on tests of recognition memory was no better than would result from guessing (Fig. 3.4–2A). Tests involving facts about his life and autobiographical experiences revealed poor memory for the time leading up to his illness but normal memory for his childhood (Fig. 3.4–2B). EP also had good spatial knowledge about the town in which he lived as a child, but he was unable to learn the layout of the neighborhood where he lived after he became amnesic (Fig. 3.4–2C).

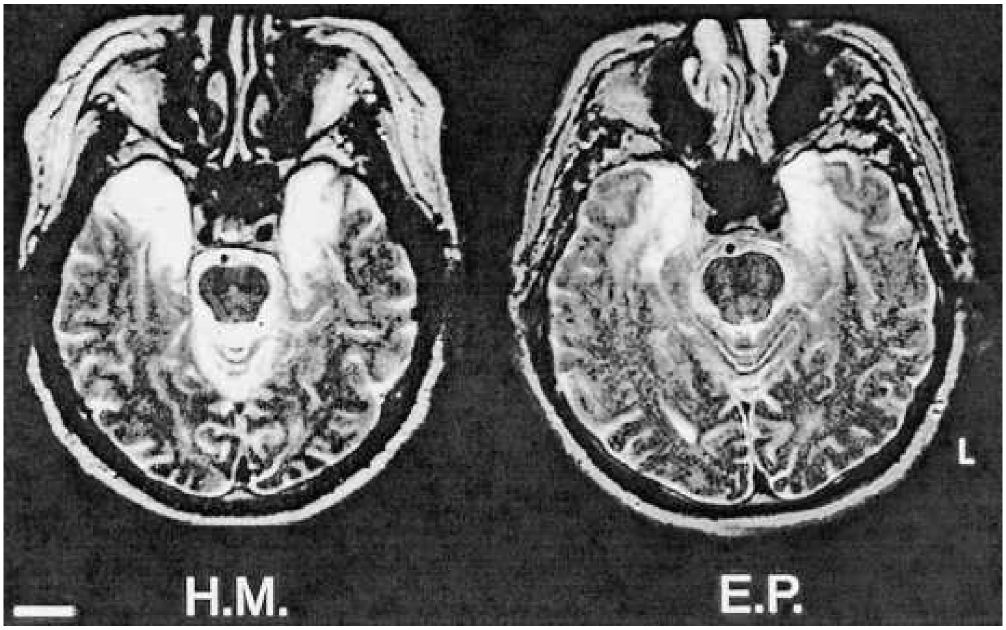

FIGURE 3.4–1. Structural magnetic resonance images of the brains of patients HM and EP through the level of the temporal lobe. Damaged tissue is indicated by bright signal in these T2-weighted axial images. Both patients sustained extensive damage to medial temporal structures as the result of surgery for epilepsy in HM and as the result of viral encephalitis in EP. Scale bar: 2 cm. L, left side of the brain. (H.M. image from Corkin S, Amaral EG, González RG, Johnson KA, Hyman BT. HM's medial temporal lobe lesion: findings from magnetic resonance imaging. *J Neurosci*. 1997;17:3964. doi:10.1523/JNEUROSCI.17-10-03964.1997. Copyright © 1997 Society for Neuroscience.) (E.P. image from Stefanacci L, Buffalo EA, Schmolck H, Squire LR. Profound amnesia after damage to the medial temporal lobe: a neuroanatomical and neuropsychological profile of patient E.P. *J Neurosci.* 2000;20:7024. doi:10.1523/JNEUROSCI.20-18-07024.2000. Copyright © 2000 Society for Neuroscience.)

Given the severity of the memory problems experienced by EP and other patients with amnesia, it is noteworthy that these patients nonetheless perform normally on certain kinds of memory tests, as described below. The impairment selectively concerns memory for factual knowledge and autobiographical events, that is, declarative memory. Amnesia presents as a global deficit in that it involves memory for information presented in any sensory modality, but the deficit is limited in that it covers only memory for facts and events.

In another patient, RB, an episode of global ischemia after cardiac surgery led to moderately severe anterograde amnesia and retrograde amnesia, less severe than in patients HM and EP. After his death 5 years later, a detailed histologic study of his brain revealed that amnesia in RB was the result of a circumscribed bilateral lesion of hippocampal field CA1 (Fig. 3.4–3). Similar histologic evidence of hippocampal damage has been obtained in other patients with amnesia.

Hippocampal pathology in patients with anterograde amnesia was also revealed using high-resolution magnetic resonance imaging (MRI). Such studies indicate that damage limited to the hippocampus results in clinically significant memory impairment. In addition to the hippocampus, other medial temporal lobe regions also make critical contributions to memory. Thus, a moderately severe memory impairment results from CA1 damage (as in patient RB), whereas a more profound and disabling amnesia results from medial temporal lobe damage that includes the hippocampus and adjacent cortex (as in patients HM and EP).

Amnesia due to medial temporal lobe damage is also typical in patients with Alzheimer disease. Hallmark neuropathologic features of Alzheimer disease begin in the medial temporal lobes and then spread throughout the brain with disease progression. Thus, declarative memory impairments are chief cognitive symptoms early in the disease progression, including in the prodromal period of “amnestic mild cognitive impairment,” with substantial cognitive deficits in addition to declarative memory dysfunction increasing with progression.

Amnesia can also result from damage to structures of the *medial diencephalon*. The critical regions damaged in diencephalic amnesia include the mammillary nuclei in the hypothalamus, the dorsomedial nucleus of the thalamus, the anterior nucleus, the internal medullary lamina, and the mammillothalamic tract. However, uncertainty remains regarding which specific lesions are required to produce diencephalic amnesia. *Alcoholic Korsakoff syndrome* is the most prevalent and best-studied example of diencephalic amnesia, and in these cases, the damage is found in brain regions that may be especially sensitive to prolonged bouts of thiamine deficiency and alcohol abuse. Patients with alcoholic Korsakoff syndrome typically exhibit memory impairment due to a combination of diencephalic damage and frontal lobe pathology. Frontal damage alone produces characteristic cognitive deficits along with certain memory problems (e.g., in effortful retrieval and evaluation); in Korsakoff syndrome, the pattern of deficits thus extends beyond what is commonly found in other cases of amnesia (Table 3.4–1).

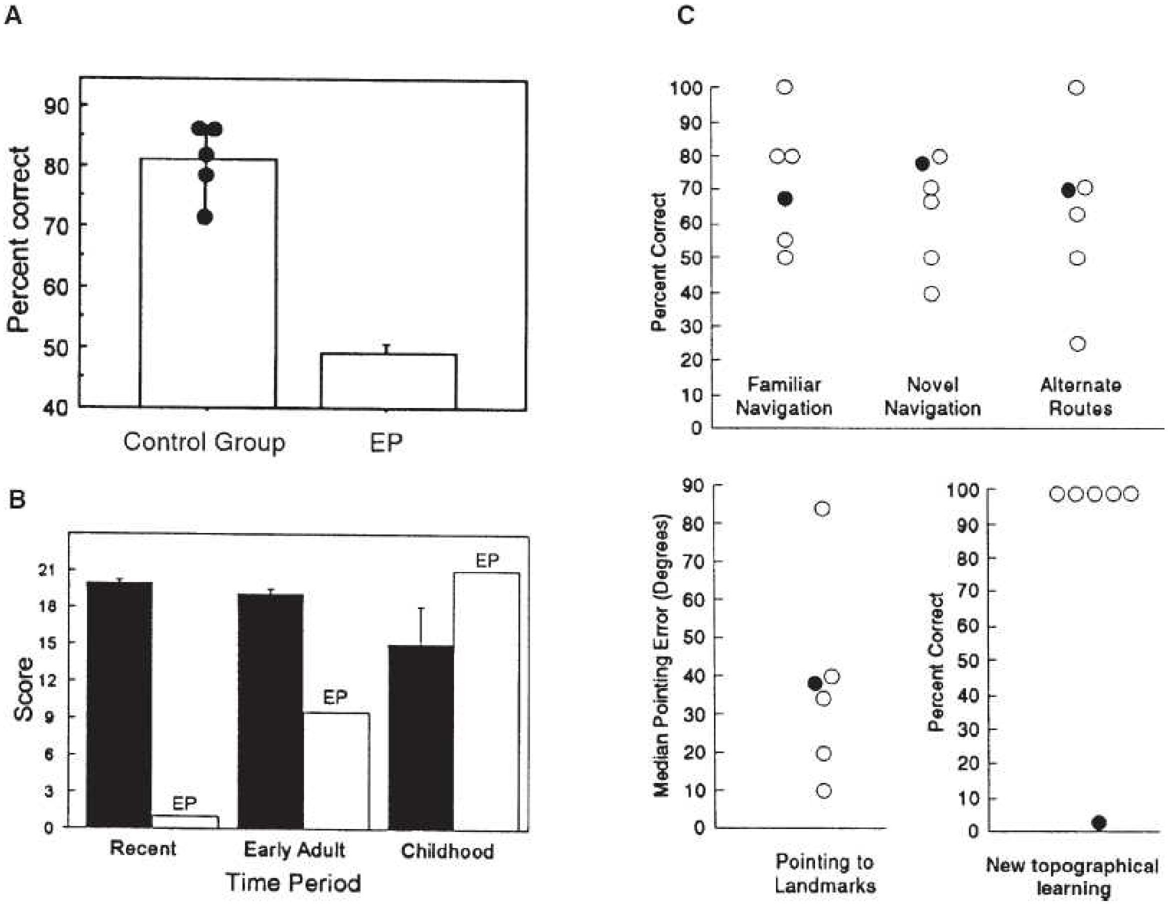

FIGURE 3.4–2. Formal test results from patient EP show severe anterograde and retrograde deficits, with intact remote memory. **A:** Scores were combined from 42 different tests of recognition memory for words given to patient EP and a group of five healthy control subjects. The testing format was either two-alternative forced-choice or yes–no recognition. Brackets for EP indicate the standard error of the mean. Data points for the control group indicate each participant's mean score across all 42 recognition memory tests. EP's average performance (49.3% correct) was not different from chance and was approximately five standard deviations (SDs) below the average performance of control subjects (81.1% correct; SD, 6.3). **B:** Autobiographical remembering was quantified by using a structured interview known as the autobiographical memory interview. Items assessed personal semantic knowledge (maximum score: 21 for each time period). Performance for the recent time period reflects poor memory for information that could have been acquired only subsequent to the onset of his amnesia. For EP, performance for the early adult period reflects retrograde memory deficits. Performance for the childhood period reflects good remote memory. Similar results for semantic and episodic remembering were obtained from these time periods. (Data from Kopelman MD, Wilson BA, Baddeley AD. The autobiographical memory interview: a new assessment of autobiographical and personal semantic memory in amnesic patients. *J Clin Exp Neuropsychol*. 1989;11:724 and Reed JM, Squire LR. Retrograde amnesia for facts and events: findings from four

new cases. *J Neurosci*. 1998;18:3943.) **C:** Assessments of spatial memory demonstrated EP's good memory for spatial knowledge from his childhood along with extremely poor new learning of spatial information. Performance was compared to that of five individuals (*empty circles*) who attended EP's high school at the same time as he did, lived in the region over approximately the same time period, and, like EP (*filled circle*), moved away as young adults. Normal performance was found for navigating from home to different locations in the area (familiar navigation), between different locations in the area (novel navigation), and between these same locations when a main street was blocked (alternative routes). Subjects were also asked to point to particular locations while imagining themselves in a particular location (pointing to landmarks), or they were asked about locations in the neighborhoods in which they currently lived (new topographical learning). EP showed difficulty only in this last test because he moved to his current residence after he became amnesic. (Adapted from Teng E, Squire LR. Memory for places learned long ago is intact after hippocampal damage. *Nature.* 1999;400:675.)

There are also a variety of neocortical areas that work in concert with the medial temporal lobes and diencephalon to accomplish declarative memory. In particular, anatomical studies in nonhuman primates have identified specific areas of medial prefrontal, lateral temporal, and medial and lateral parietal cortex that are reciprocally connected to the medial temporal lobe and noninvasive neuroimaging has confirmed this distributed network of the hippocampus in humans (Fig. 3.4–4). Even small brain lesions within this network can produce a similar amnestic syndrome as that resulting from bilateral hippocampal damage. Thus, declarative memory is not accomplished only by the hippocampus and diencephalon, but rather a large and distributed network makes crucial contributions.

In summary, the ability to remember factual and autobiographical events depends on the integrity of a large-scale brain network centered on the hippocampus. Thus, studies point to a critical role for medial temporal and diencephalic brain areas, in concert with widespread areas of the neocortex, during the formation and storage of declarative memories (Fig. 3.4–5).

Retrograde Amnesia and Memory Consolidation. Whereas memory loss in amnesia typically affects recent memories more than remote memories (Fig. 3.4–6), temporally graded amnesia also has been demonstrated retrospectively in studies of amnesic patients and prospectively in studies of monkeys, rats, mice, and rabbits. These findings have important implications for understanding the nature of the memory storage process. Memories are dynamic, not static. As time passes after learning, some memories are forgotten, whereas others become stronger due to a process of *consolidation* that depends on cortical, medial temporal, and diencephalic structures.

The study of retrograde amnesia has therefore been important for understanding how memory changes over time. The dynamic nature of memory storage can be conceptualized as follows. An event is experienced and encoded by virtue of a collection of cortical regions that are involved in representing a combination of different event features. At the same time, the hippocampus and adjacent cortex receive pertinent high-level information from all sensory modalities. Later, when the original event is recalled, the same set of cortical regions is activated. If a subset of the cortical regions is activated, the hippocampus and related structures can facilitate recall by facilitating the activation of the remaining cortical regions (i.e., pattern completion). When the original event is retrieved and newly associated with other information, hippocampal–cortical networks can be modified. In this way, a gradual consolidation process occurs that changes the nature of memory storage (Fig. 3.4–4). The neocortical components representing some events can become so effectively linked together that ultimately a memory can be retrieved without any contribution from the medial temporal lobe. As a result, amnesic patients can exhibit normal retrieval of remote facts and events, as well as autobiographical memories. Distributed neocortical regions are the permanent repositories of these enduring memories.

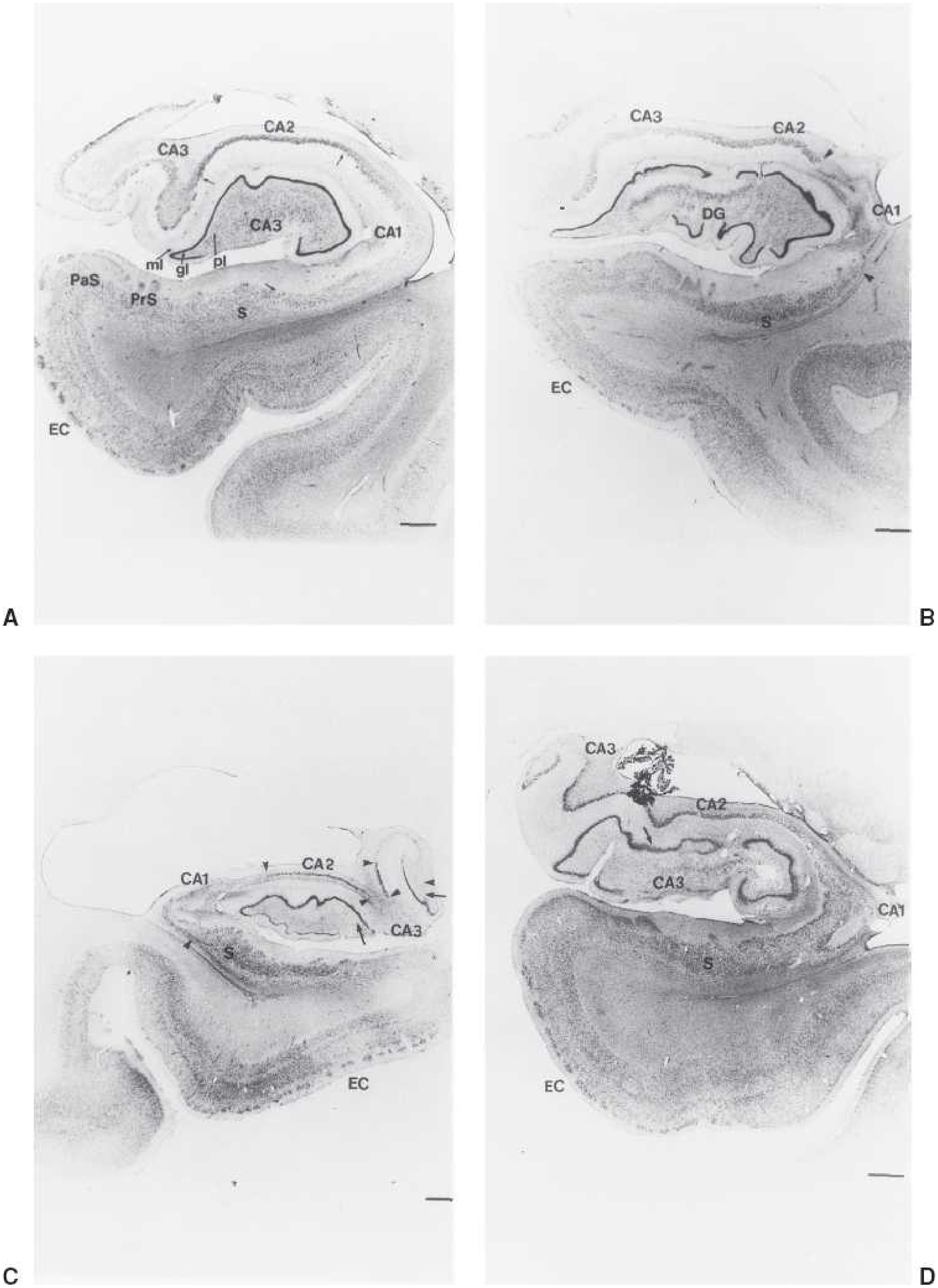
CA2
CA3
CA3
CA1
ml
gl
pl
PaS
PrS
S
EC
A
CA3
CA2
DG
CA1
S
EC
B
CA2
CA1
CA3
S
EC
C
CA3
CA2
CA3
CA1
S
EC
D

FIGURE 3.4–3. A: Coronal sections through the hippocampal region stained with thionin in a healthy individual and (**B–D**) three amnesic patients with bilateral damage to the hippocampal formation, patients GD, LM, and WH.

The hippocampus proper can be divided into three distinct fields: designated CA1, CA2, and CA3. The CA1 field extends to the subiculum (S). Other structures include the dentate gyrus (DG), presubiculum (PrS), parasubiculum (PaS), and entorhinal cortex (EC). In **B**, damage included CA1, marked by *arrowheads*. In **C**, damage included CA1, CA2, CA3, DG, and EC. *Arrowheads* indicate the borders of the CA fields. *Arrows* indicate loss of polymorphic cells in DG. In **D**, damage included CA1, CA2, CA3, DG, S, and EC. *Arrow* indicates abnormal dispersion of granule cells. gl, granular layer; ml, molecular layer; pl, polymorphic layer. (Reprinted with permission from Squire LR, Zola SM. Memory, memory impairment, and the medial temporal lobe. In: *Cold Spring Harbor Symposia on Quantitative Biology*. Vol LXI. Cold Spring Harbor Laboratory Press; 1996:190. © Cold Spring Harbor Laboratory Press.)

Table 3.4–1.
Memory and Cognitive Deficits Associated with Frontal Damage

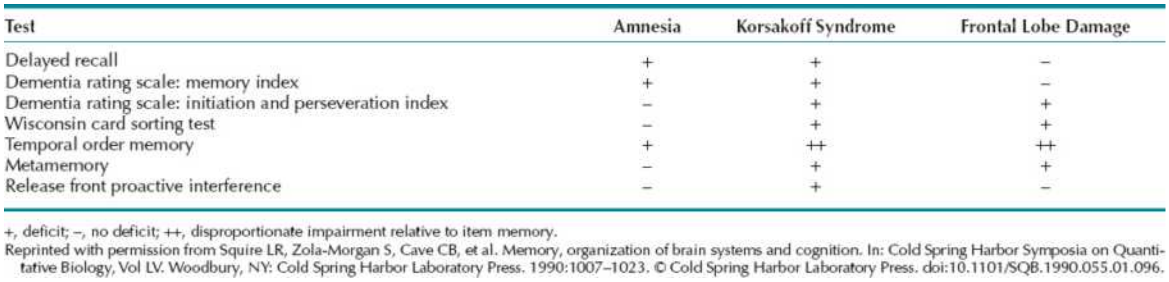

Test	Amnesia	Korsakoff Syndrome	Frontal Lobe Damage
Delayed recall	+	+	–
Dementia rating scale: memory index	+	+	–
Dementia rating scale: initiation and perseveration index	–	+	+
Wisconsin card sorting test	–	+	+
Temporal order memory	+	++	++
Metamemory	–	+	+
Release front proactive interference	–	+	–

+, deficit; –, no deficit; ++, disproportionate impairment relative to item memory.

Reprinted with permission from Squire LR, Zola-Morgan S, Cave CB, et al. Memory, organization of brain systems and cognition. In: Cold Spring Harbor Symposia on Quantitative Biology, Vol LV. Woodbury, NY: Cold Spring Harbor Laboratory Press. 1990:1007–1023. © Cold Spring Harbor Laboratory Press. doi:10.1101/SQB.1990.055.01.096.

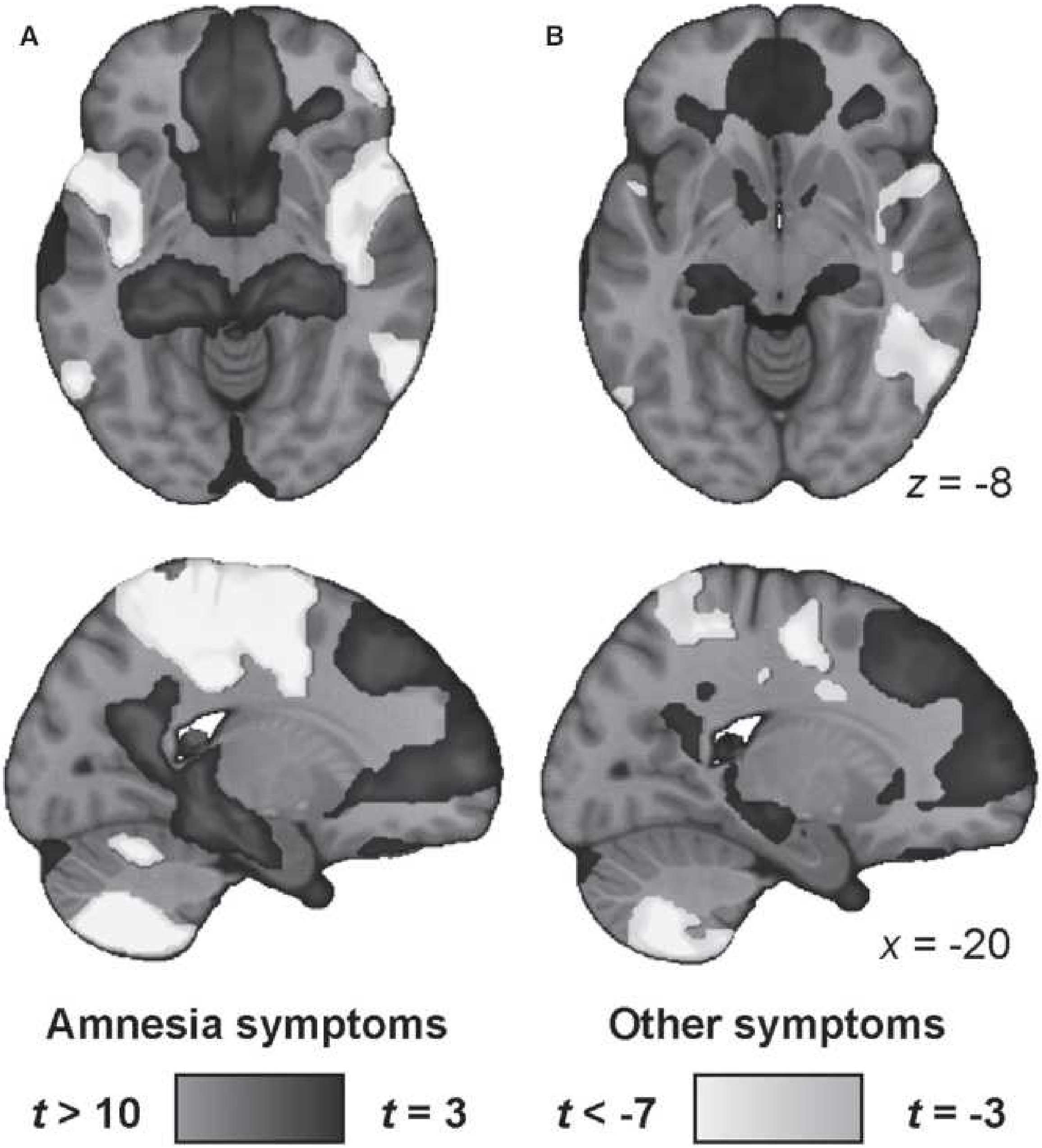

FIGURE 3.4–4. Broader network of the hippocampus defined based on brain locations where damage causes amnesia. Brain lesions causing amnesia were defined in 53 case reports of strokes. Each lesion location was connected to a network of brain regions identified via lesion network mapping (i.e., resting-state connectivity computed using a publicly available connectome dataset of 1,000 healthy individuals). **A–B:** This connectivity profile was specific to lesions causing amnesia relative to lesions resulting in nonspecific symptoms (**A**) or lesions causing symptoms unrelated to amnesia (**B**). Connectivity analyses are reported as voxelwise two-sample t-values corrected for multiple comparisons with a voxel-based family-wise error rate of $P < .05$. Darker grey regions showed greater hippocampal connectivity to lesion locations, causing amnesia relative to locations causing symptoms

other than amnesia, while *white/light grey* regions showed greater hippocampal connectivity to lesion locations causing symptoms other than amnesia relative to locations causing amnesia. (Adapted from Ferguson MA, Lim C, Cooke D, et al. A human memory circuit derived from brain lesions causing amnesia. *Nature Comms*. 2019;10:3497. License: https://creativecommons.org/licenses/by/4.0/)

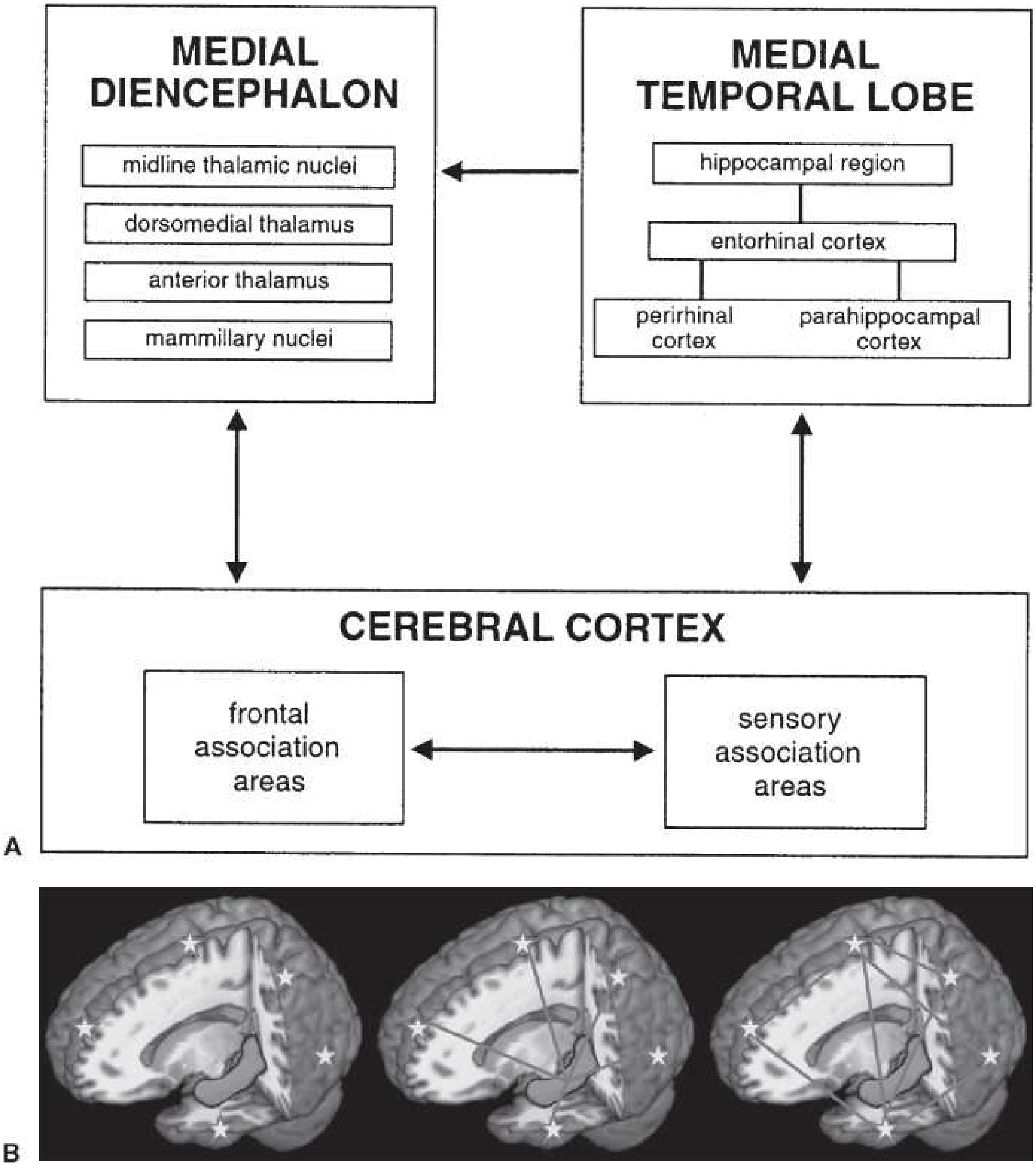
MEDIAL DIENCEPHALON
midline thalamic nuclei
dorsomedial thalamus
anterior thalamus
mammillary nuclei
MEDIAL TEMPORAL LOBE
hippocampal region
entorhinal cortex
perirhinal cortex
parahippocampal cortex
CEREBRAL CORTEX
frontal association areas
sensory association areas
A
B

FIGURE 3.4–5. Brain regions thought to be critical for the formation and storage of declarative memory. **A:** Medial diencephalon and medial temporal regions are critical for declarative memory storage. The entorhinal cortex is the major source of projections for neocortex to hippocampus, and nearly two-thirds of the cortical input to the entorhinal cortex originates in the perirhinal and parahippocampal cortex. The entorhinal cortex also receives direct connections from the cingulate, insula, orbitofrontal, and superior temporal cortices. (Part A adapted from Paller KA. Neurocognitive foundations of human memory. In: Medin DL, ed. *The Psychology of Learning and Motivation*, Vol. 40. Academic Press; 2000:121.) **B:** A schematic view of the cortical components of declarative memory (*stars*) that are initially separate

(**left**) but become linked with the hippocampus and related structures when an event occurs (**middle**) and ultimately can become intrinsically linked together (**right**).

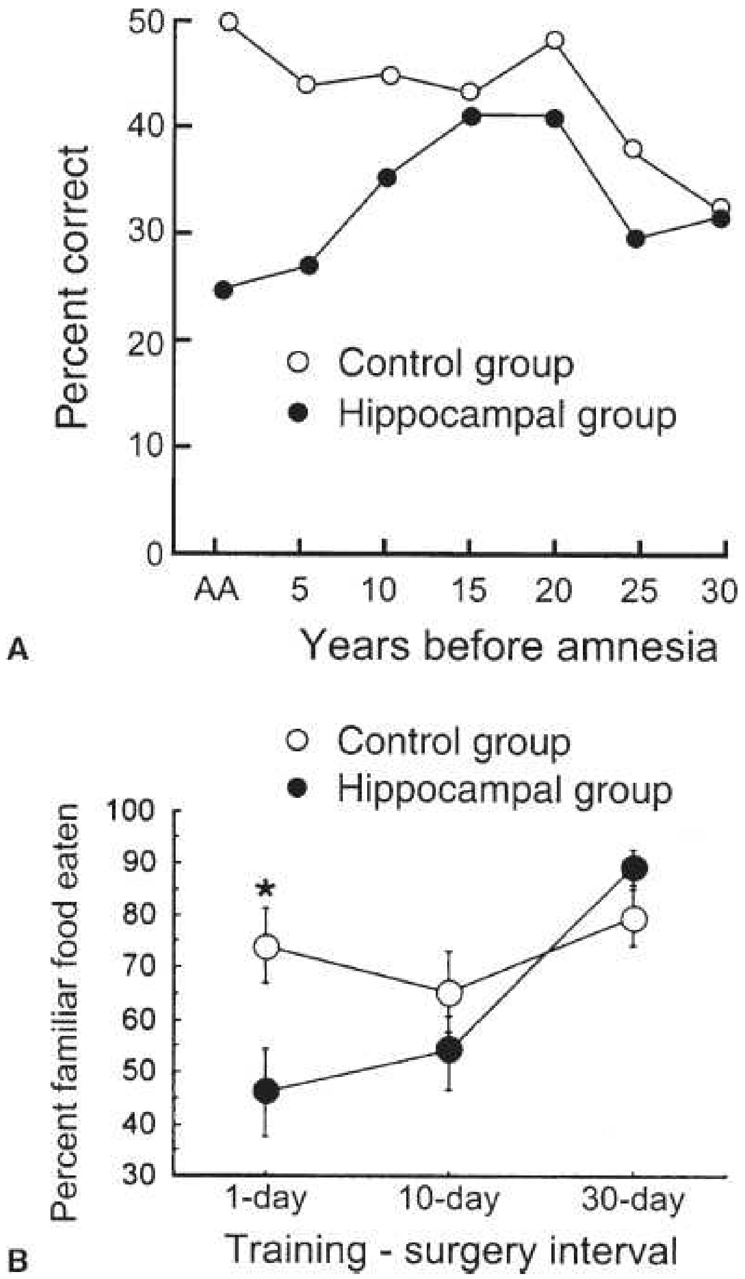

FIGURE 3.4–6. **A:** Temporally limited retrograde amnesia for free recall of 251 news events. Scores were aligned relative to the onset of amnesia in patients (N = 6) and to a corresponding time point in age- and education-matched healthy individuals (N = 12). The time period after the onset of amnesia is labeled AA (anterograde amnesia) to designate that this time point assessed memory for events that occurred after the onset of amnesia. Standard errors ranged from 2% to 10%. Brain damage in the patient group was limited primarily to the hippocampal region. (Part A adapted from Manns JR, Hopkins RO, Squire LR. Semantic memory and the human hippocampus. *Neuron.* 38(1):127–133. doi:10.1016/s0896-6273(03)00146-6. Copyright © 2003 Cell Press, with permission from Elsevier.) **B:** Temporally limited retrograde amnesia in rats with lesions of hippocampus and subiculum. Rats learned to

prefer an odorous food as the result of an encounter with another rat with that odor on its breath. Percent preference for the familiar food was observed for three training-surgery intervals. At 1 day after learning, the control group performed significantly better than the rats with lesions ($P < .05$). At 30 days, the two groups performed similarly, and both groups performed well above chance. *Error bars* show standard errors of the mean. (Part B reprinted from Clark RE, Broadbent NJ, Zola SM, Squire LR. Anterograde amnesia and temporally graded retrograde amnesia for a nonspatial memory task after lesions of hippocampus and subiculum. *J Neurosci.* 2002;22:4663. doi:10.1523/JNEUROSCI.22-11-04663.2002. Copyright © 2002 Society for Neuroscience.)

In contrast to what is observed after damage restricted to the hippocampus, extensive retrograde impairments for facts and events from the distant past can also occur. Damage to the *frontal lobes*, for example, can lead to difficulty in organizing memory retrieval. Accurate retrieval often begins with the activation of lifetime periods and proceeds to the identification of general classes of events and then more specific events, but this process becomes difficult following frontal damage. Damage to other cortical regions can also impair memory storage. Networks in the anterolateral temporal cortex, for example, are critical for retrieving stored information because these areas are important for long-term storage itself. Patients with *focal retrograde amnesia* exhibit substantial retrograde memory impairments together with only moderately impaired new learning ability. Some capacity for new learning remains, presumably because medial temporal lobe structures are able to communicate with other areas of the cortex that remain undamaged.

Although amnesia does not typically occur after limited frontal damage, the frontal lobes are fundamentally important for declarative memory. Patients with frontal lesions have a poor memory for the context in which information was acquired, they have difficulty in unaided recall, and they may even have some mild difficulty on tests of item recognition. More generally, frontal patients have difficulty implementing memory retrieval strategies and evaluating and monitoring their memory performance.

Patient BG suffered an infarction restricted to the right frontal lobe, resulting in substantial false remembering. He had an abnormal tendency to claim that some stimuli were familiar, even though they had not been presented for study. His false responses probably arose because he relied on a general feeling of familiarity for the kind of stimuli that had been presented rather than on specific memories for the stimuli.

THE ROLE OF SLEEP IN MEMORY CONSOLIDATION. The speculation that memories are processed during sleep has a long history. Freud noted that dreams could reveal fragments of recent experiences in the form of day residues. Although many questions about how and why memories may be processed during sleep remain unresolved, recent experiments have provided new empirical support to bolster the idea that memory processing during sleep serves an adaptive function related to memory consolidation. It is now clear that memory performance can be facilitated when sleep occurs after initial learning and that sleep-related facilitation can be observed for many different types of memory.

Memory consolidation appears to be specifically aided by processing during deep sleep in the nights after learning, especially in the stage of sleep known as *slow-wave sleep,* and probably also in other sleep stages. Some results indicate that slow-wave sleep facilitates the storage of declarative memories but not nondeclarative memories. Direct evidence for this proposal has been obtained using stimulation with olfactory stimuli (Fig. 3.4–7), stimulation with direct-current electrical input at the approximate frequency of EEG slow oscillations, and other methods. A fruitful approach for studying memory processing during sleep is to use sensory stimulation to activate specific memories, which then may be subject to modification. Neuronal recordings in animals have revealed a phenomenon of hippocampal replay, whereby activity patterns expressed during the day are later observed during sleep. Research is now focused on delineating the neurophysiologic events that transpire when memories are reactivated and altered during sleep. A novel conjecture, supported by some recent empirical findings, is that adapting these procedures for memory modification during sleep could be used to treat unwanted learned associations, such as those involved in nicotine craving (Fig. 3.4–8). In summary, declarative memories acquired during waking can be processed again during sleep, and this processing can influence the likelihood of subsequent memory retrieval when the individual is awake.

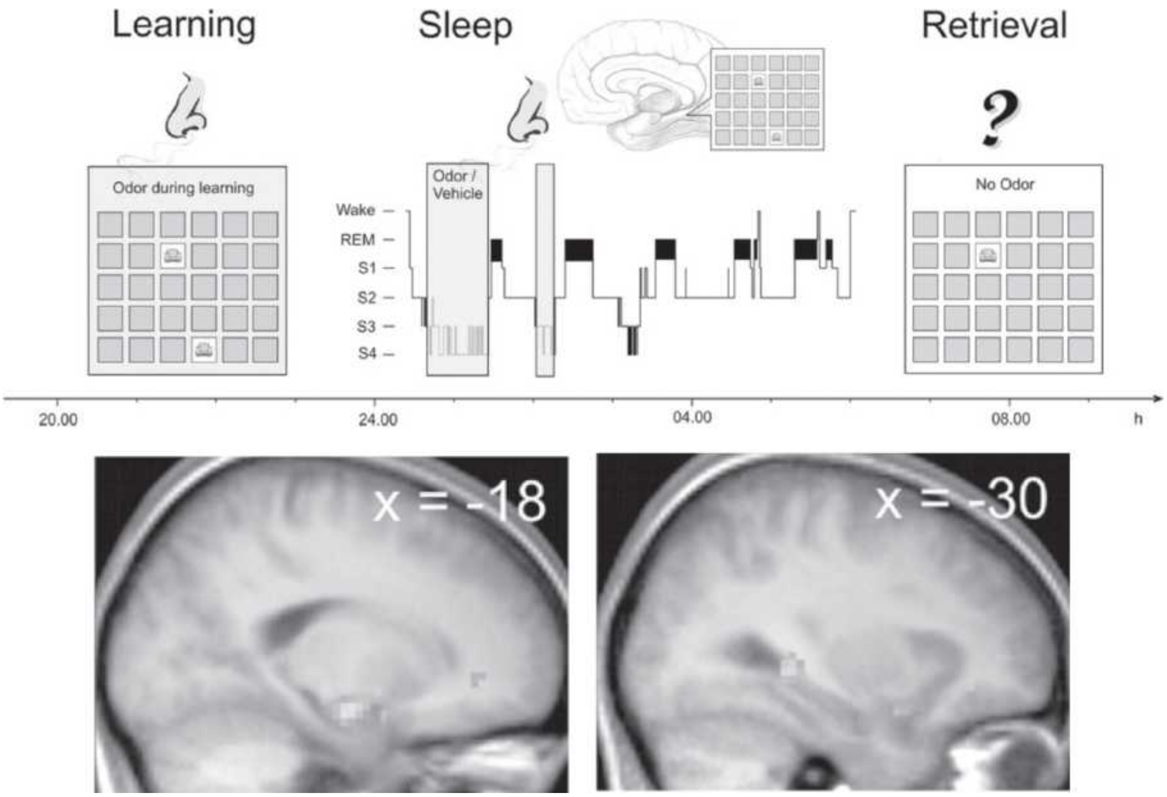

FIGURE 3.4–7. Evidence for memory processing during sleep. Subjects first learned object location associations when a rose odor was present. Following learning, subjects slept wearing a device for delivering odors to the nose, and the rose odor was administered during the first two slow-wave sleep periods of the night (in 30-s periods to prevent habituation). Memory facilitation was observed when object–location associations were tested the following morning in the absence of odor stimulation. Facilitated memory was not found when stimulation occurred during slow-wave sleep but not during learning, when stimulation occurred during learning and then during REM sleep, or when subjects were kept awake. Moreover, odor stimulation during slow-wave sleep was found to produce anterior and posterior hippocampal activation (**lower panels**). (From Rasch B, Büchel C, Gais S, Born J. Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*. 2007;315:1426. Reprinted with permission from AAAS. doi:10.1126/science.1138581.)

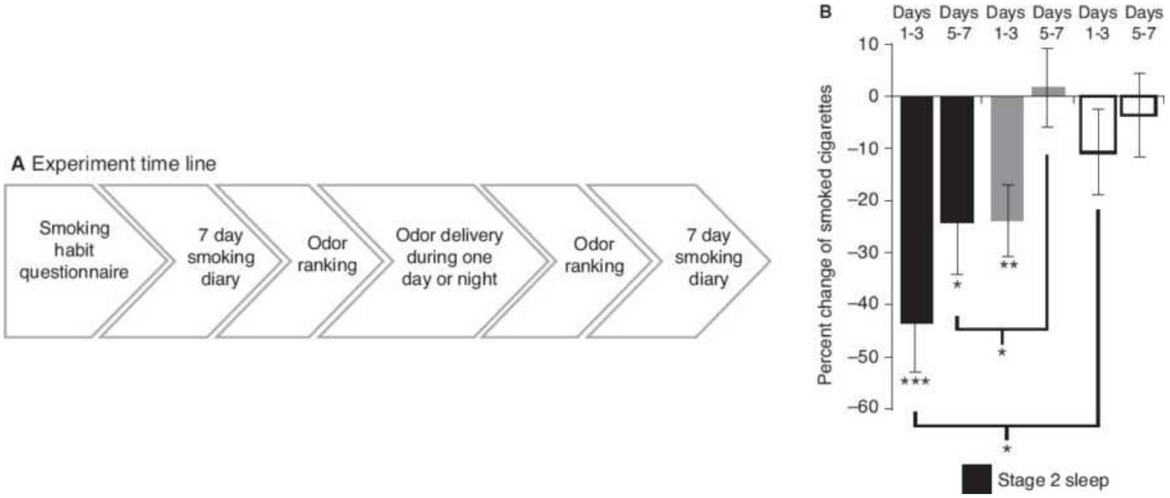

FIGURE 3.4–8. Memory processing during sleep in relation to smoking habits. **A:** Procedure used by Arzi and colleagues (2014) to study aversive conditioning during sleep in individuals who wanted to quit smoking. **B:** Results showed that measures of cigarette smoking during the subsequent week were reduced when conditioning took place during sleep. Conditioning during stage 2 sleep produced the strongest reduction in smoking compared to conditioning during REM sleep and conditioning during wake. (Reprinted from Arzi A, Holtzman Y, Samnon P, Eshel N, Harel E, Sobel N. Olfactory aversive conditions during sleep reduces cigarette-smoking behavior. *J Neurosci.* 2014;34:15382–15393. doi:10.1523/JNEUROSCI.2291-14.2014.)

Nondeclarative Memory

The acquisition of perceptual, perceptual–motor, and cognitive skills—that is, nondeclarative memory—can occur independent from declarative memory, and patients with amnesia acquire these skills at rates equivalent to the rates at which healthy individuals acquire the skills. For example, patients with amnesia can learn to read mirror-reversed text normally, they exhibit the normal facilitation in reading speed with successive readings of normal prose, and they improve as rapidly as healthy individuals at speed reading repeating nonwords. In addition, patients with amnesia can, after seeing strings of letters generated by a finite-state rule system, classify novel strings of letters as rule based or not rule based. Classification performance is normal despite the fact that patients with amnesia are impaired at remembering the events of training or the specific items they have studied.

Priming

Priming refers to the facilitation of the ability to detect or to identify a particular stimulus based on a specific recent experience. Many tests have been used to measure priming in amnesia and show that it is intact. For example, words might be presented in a study phase and then again, after a delay, in a test phase when a priming measure such as reading speed is obtained. Patients are instructed to read words as quickly as possible in such a test, and they are not informed that memory is being assessed.

In one priming test, patients named pictures of previously presented objects reliably faster than they named pictures of new objects, even after a delay of a week. This facilitation occurred at normal levels, despite the fact that the patients were markedly impaired at recognizing which pictures had been presented previously. Particularly striking examples of preserved priming come from studies of patient EP (Fig. 3.4–9), who exhibited intact priming for words but performed at chance levels when asked to recognize which words had been presented for study. This form of memory, termed *perceptual priming,* is thus a distinct class of memory that is independent of the medial temporal lobe regions typically damaged in amnesia.

Another form of priming reflects improved access to meaning rather than percepts. For example, subjects study a list of words, including *tent* and *belt,* and then are asked to free associate to other words. Thus, they are given words such as *canvas* and *strap* and asked to produce the first word that comes to mind. The result is that subjects are more likely to produce *tent* in response to *canvas* and to produce *belt* in response to *strap* than if the words *tent* and *belt* had not been presented recently. This effect, called *conceptual priming,* is also preserved in patients with amnesia, even though they fail to recognize the same words on a conventional memory test (Fig. 3.4–10).

Yet, not all types of priming are preserved in amnesia. Some priming tests have been designed to examine the formation of new associations. When priming tests are based not on pre-existing knowledge but rather on the acquisition of new associative knowledge, priming tends to be impaired. In other words, priming in certain complex situations can require the same type of linkage among multiple cortical regions that is critical for declarative memory.

CORTICAL ORGANIZATION OF MEMORY

Contributions from Animal Models

Findings from human patients with amnesia helped to motivate the creation of animal models examining the locus of memory storage in the brain. In the 1920s, Karl Lashley searched for the site of memory storage by studying the behavior of rats following the removal of different amounts of cerebral cortex. He recorded the number of trials needed to relearn a maze problem that rats learned prior to surgery, and he found that the deficit was proportional to the amount of cortex removed. The deficit did not seem to depend on the particular location of cortical damage. Lashley concluded that the memory resulting from maze learning was not localized in any one part of the brain but instead was distributed equivalently over the entire cortex.

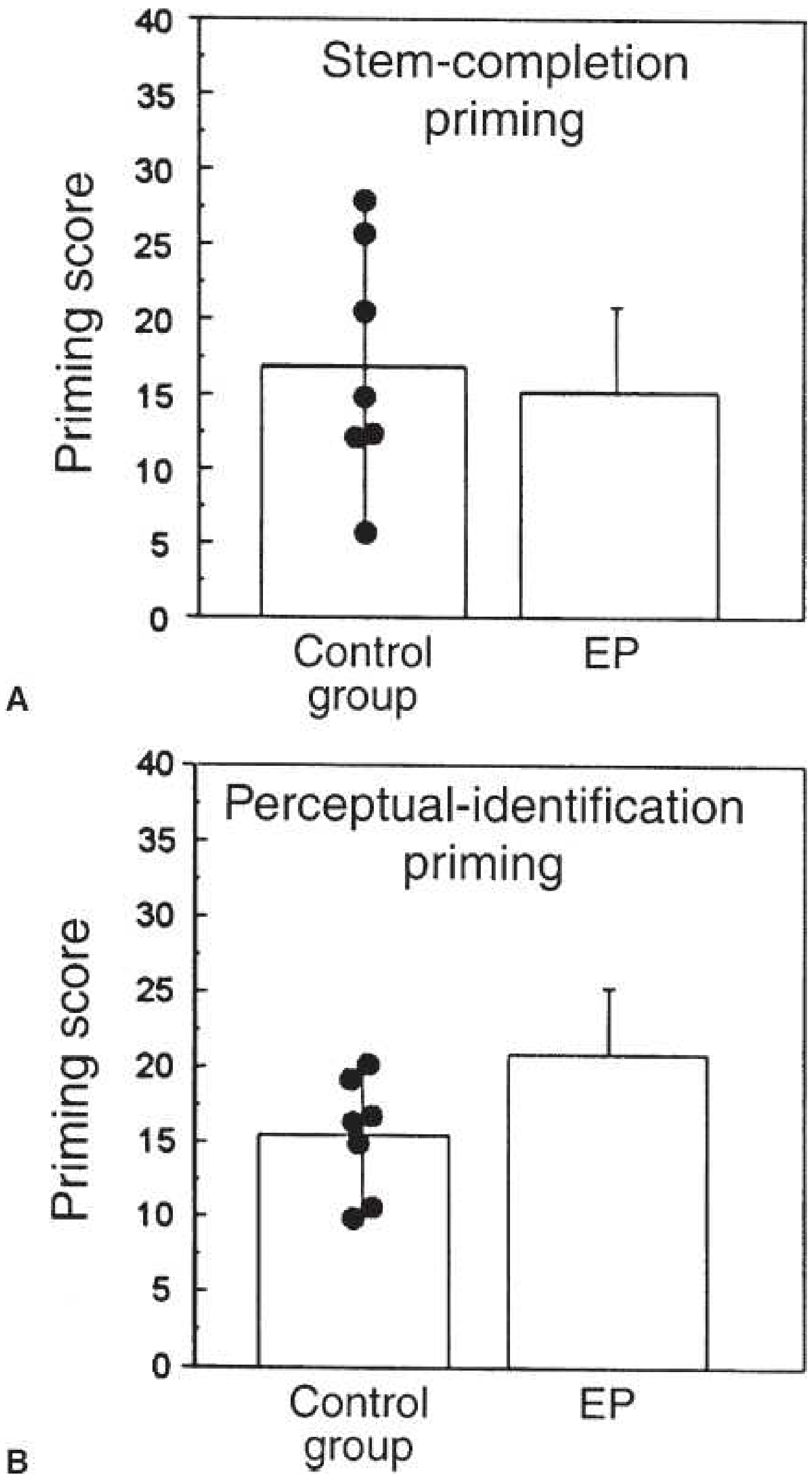

FIGURE 3.4–9. Preserved priming in patient EP relative to seven control subjects. **A:** Stem-completion priming on six separate tests. Priming reflected a tendency for subjects to complete three-letter stems with previously encountered words when they were instructed to produce the first word to come to mind (e.g., MOT___ completed to form MOTEL). Priming scores were calculated as percent correct for studied words minus percent correct for

baseline words (guessing). **B:** Perceptual-identification priming on 12 separate tests. Subjects attempted to read 48 words that were visually degraded. Priming scores were calculated as percent correct identification of previously studied words minus percent correct identification of nonstudied words. Brackets indicate the standard error of the mean. (Copyright © 1997 by American Psychological Association. Reproduced with permission. Hamann SB, Squire LR. Intact perceptual memory in the absence of conscious memory. *Behav Neurosci*. 1997;111:850. No further reproduction or distribution is permitted without written permission from the American Psychological Association.)

Subsequent research has led to reinterpretations of these results. Maze learning in rats depends on different types of information, including visual, tactile, spatial, and olfactory information. Importantly, neurons that process these various types of information are segregated into different areas of the rat cerebral cortex, and memory storage is segregated in a parallel manner. Thus, the correlation between maze learning abilities and lesion size that Lashley observed is a result of progressive encroachment of larger lesions on specialized cortical areas serving the many components of information processing relevant for maze learning.

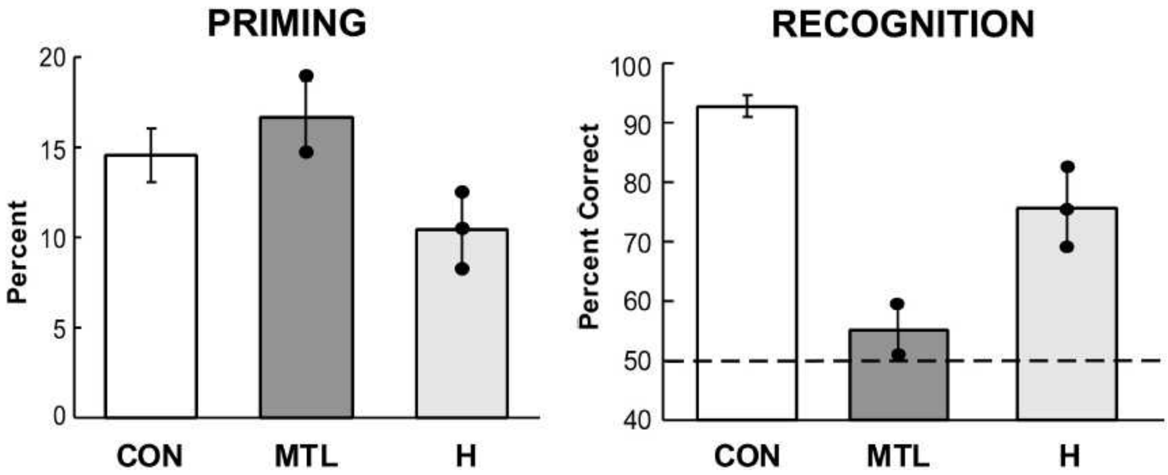

FIGURE 3.4–10. Preserved conceptual priming in amnesia. In the free association test, subjects studied a set of words (e.g., lemon) and 5 minutes later viewed cue words that included associates of studied words (e.g., orange). Subjects were asked to produce the first word that came to mind in response to each cue word. Results are shown separately for the control group (CON, n = 12), amnesic patients with large medial temporal lobe lesions (MTL, n = 2), and amnesic patients with lesions thought to be limited to the hippocampal region (H, n = 3). The **left panel** shows conceptual priming scores computed as the percentage of studied words produced in the free association test minus a baseline measure of the likelihood of producing those words by chance. All the groups performed similarly in the conceptual priming test. The **right panel** shows results from a yes–no recognition memory test using comparable words. Both patient groups were impaired relative to the control group. The *dashed line* indicates chance performance. The data points for the MTL and H groups show the scores of individual patients averaged across four separate tests. *Brackets* show standard errors of the mean for the control group. (From Levy DA, Stark CEL, Squire LR. Intact conceptual priming in the absence of declarative memory. *Psych Sci*. 2004;15:680.)

In monkeys, many parallels to human amnesia have been demonstrated following surgical damage to anatomical components of the medial temporal lobe. Cumulative study of the resulting memory impairment eventually identified the medial temporal structures and connections that are crucial for memory. These include the hippocampus—which includes the dentate gyrus, hippocampal fields CA1, CA2, and CA3, and the subiculum—and also the adjacent cortical regions, including the entorhinal, perirhinal, and parahippocampal cortices.

The cortical areas responsible for processing and storing visual information have been studied most extensively in nonhuman primates, with nearly one-half of the primate neocortex specialized for visual functions. As shown in Figure 3.4–11, the cortical pathways for visual information processing begin in the primary visual cortex (V1) and proceed from there along parallel pathways or streams. One stream projects ventrally to the inferotemporal cortex and is specialized for processing information concerning the identification of visual objects. Another stream projects dorsally to the parietal cortex and is specialized for processing spatiotemporal information. These distinct processing streams converge on the medial temporal lobe and hippocampus, motivating the theoretical view that the hippocampus serves its critical role in memory based on its ability to bind information about what is perceived with its spatiotemporal context.

Specific visual processing areas in the dorsal and ventral streams, together with areas in the prefrontal cortex, register the immediate experience of perceptual processing. The results of perceptual processing are first available as *immediate memory.* Immediate memory refers to the amount of information that can be held in mind (like a telephone number) so that it is available for immediate use. Immediate memory can be extended in time by rehearsing or otherwise manipulating the information, in which case what is stored is now said to be in *working memory.*

Regions of the visual cortex in the forward portions of the dorsal and ventral streams serve as the ultimate repositories of visual memories. The inferotemporal cortex, for example, lies at the end of the ventral stream, and inferotemporal lesions lead to selective impairments in both visual object perception and visual memory. Yet, such lesions do not disrupt elementary visual functions, such as acuity. Electrophysiologic studies in the monkey show that neurons in area TE, which is one part of the inferotemporal cortex, register specific and complex features of visual stimuli, such as shape, and respond selectively to patterns and objects. The inferotemporal cortex can thus be thought of as both a higher-order visual processing system and a storehouse of the visual memories that result from that processing.

In sum, memory is distributed and localized in the cerebral cortex. Memory is distributed in the sense that, as Lashley concluded, there is no cortical center dedicated solely to the storage of memories. Yet, memory is localized in the sense that different aspects or dimensions of events are stored at specific cortical sites—namely, in the same regions that are specialized to analyze and process what is to be stored.

Classical Conditioning

The study of classical conditioning has provided many insights into the biology of memory. Classical conditioning has been especially well studied in rabbits using a tone as the conditioned stimulus and an air puff to the eye (which automatically elicits a blink response) as the unconditioned stimulus. Repeated pairings of the tone and the air puff lead to a conditioned response, in that the tone alone elicits an eye blink. Reversible lesions of the deep nuclei of the cerebellum eliminate the conditioned response without affecting the unconditioned response. These lesions also prevent initial learning from occurring, and when the lesion is reversed, rabbits learn normally. Thus, the cerebellum contains essential circuitry for the learned association. The relevant plasticity appears to be distributed between the cerebellar cortex and the deep nuclei.

An analogous pattern of cerebellar plasticity is thought to underlie motor learning in the vestibulo-ocular reflex and, perhaps, associative learning of motor responses in general. Based on the idea that learned motor responses depend on coordinated control of changes in timing and strength of the response, it has been suggested that synaptic changes in the cerebellar cortex are critical for learned timing, whereas synaptic changes in the deep nuclei are critical for forming an association between a conditioned stimulus and an unconditioned stimulus.

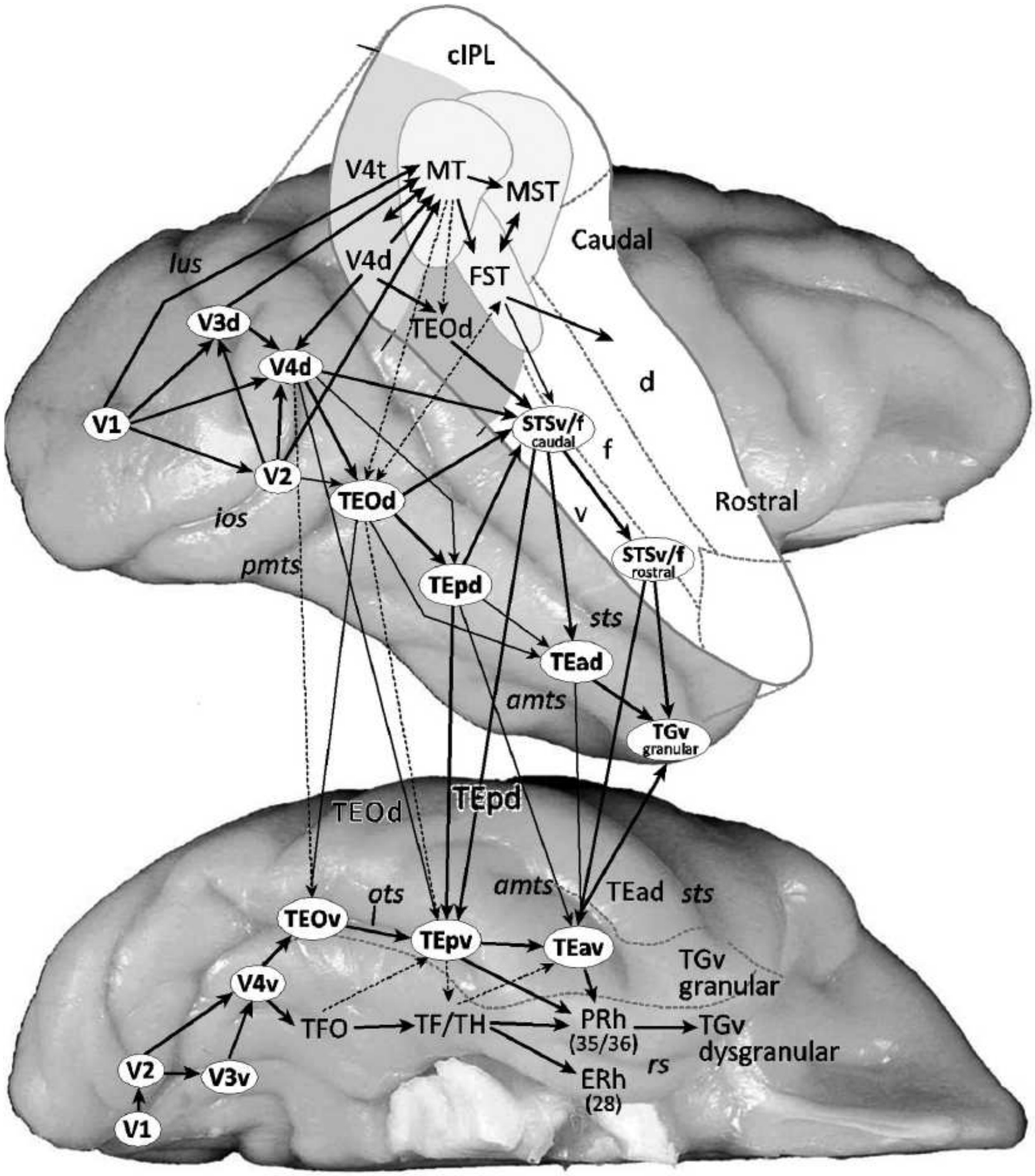
cIPL
V4t
MT
MST
Caudal
lus
V4d
FST
V3d
TEOd
V4d
d
V1
STSv/f
caudal
f
V2
TEOd
Rostral
ios
v
STSv/f
rostral
pmts
TEpd
sts
TEad
amts
TGv
granular
TEOd
TEpd
amts
TEad
sts
ots
TEOv
TEpv
TEav
TGv
granular
V4v
TFO
TF/TH
PRh
(35/36)
TGv
dysgranular
V2
V3v
rs
ERh
(28)
V1

FIGURE 3.4–11. Summary of cortical visual areas and some of their connections. Two major cortical pathways begin from area V1, also known as striate cortex or primary visual cortex, and which receives visual input from the retina by way of the lateral geniculate nucleus of the thalamus. The processing stream for object vision follows a ventral route into the temporal lobe, and the processing stream for spatial vision follows a dorsal route into the parietal lobe, which then projects into the temporal lobe. The convergence of these two processing streams in the medial temporal lobe, including hippocampus, provides an anatomical rationale for the key role of this area in constructing and retrieving memories comprised different sorts of episodic and

semantic information. *Arrows* represent reciprocal anatomical connections, with line width representing the relative strength of interconnection and dotted lines indicating very sparse projections. For anatomical abbreviations, see the original article. (Reprinted from Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn Sci.* 2013;17(1):26-49. Copyright 2013, with permission from Elsevier. doi: 10.1016/j.tics.2012.10.011.)

Fear conditioning and fear-potentiated startle are types of learning that serve as useful models for anxiety disorders and related psychiatric conditions. For example, mice exhibit freezing behavior when returned to the same context in which aversive shock was delivered on an earlier occasion. This type of learning depends on the encoding of contextual features of the learning environment. Acquiring and expressing this type of learning requires neural circuits that include both the amygdala and the hippocampus. The amygdala may be important for associating negative affect with new stimuli. The hippocampus may be important for representing the context. With extinction training, when the context is no longer associated with an aversive stimulus, the conditioned fear response fades. The frontal cortex is thought to play a key role in extinction.

Contributions from Neuroimaging

The understanding of memory derived from studies of amnesia and animal models has been extended through studies using various methods for monitoring brain activity in healthy individuals. For example, studies of brain activity measured using positron emission tomography (PET) and functional MRI show that posterior prefrontal regions are involved in strategic processing during retrieval, as well as in working memory. Anterior frontal regions near the frontal poles have been linked with functions such as evaluating the products of retrieval. Frontal connections with posterior neocortical regions support the organization of retrieval and the manipulation of information in working memory. Consistent with the evidence from patients with frontal lesions, frontal–posterior networks can be viewed as instrumental in the retrieval of declarative memories and in the online processing of new information.

Neuroimaging has also identified contributions to memory made by the parietal cortex. Multiple parietal regions (including inferior and superior parietal lobules, precuneus, posterior cingulate, and retrosplenial cortex) are activated in conjunction with remembering recent experiences. Although many functions have been hypothesized to explain this parietal activity, a consensus position has not yet been reached, and it may be the case that several different functions are relevant.

Neuroimaging studies have also illuminated priming phenomena and how they differ from declarative memory. Perceptual priming appears to reflect changes in the early stages of the cortical pathways that are engaged during perceptual processing. For example, in the case of stem-completion priming, where subjects study a list of words (e.g., MOTEL) and then are tested with a list of stems (e.g., MOT___) and with instructions to complete each stem with the first word to come to mind, neuroimaging and divided visual-field studies have implicated visual processing systems in extrastriate cortex, especially in the right hemisphere. In contrast, conscious recollection of remembered words engages brain areas at later stages of processing. Neural mechanisms supporting priming and declarative memory retrieval have also been distinguished in brain electrical activity recorded from the scalp (Fig. 3.4–12). In sum, priming differs from declarative memory in that it is signaled by brain activity that occurs more efficiently during memory retrieval and that originates in different brain regions. Recent findings suggest that brain activity related to both priming and declarative memory often occurs together, suggesting that multiple co-occurring memory processes are normally involved in memory retrieval. The degree to which processes involved in priming versus declarative memory work together to support healthy memory function is unknown and remains an intriguing focus of contemporary research.

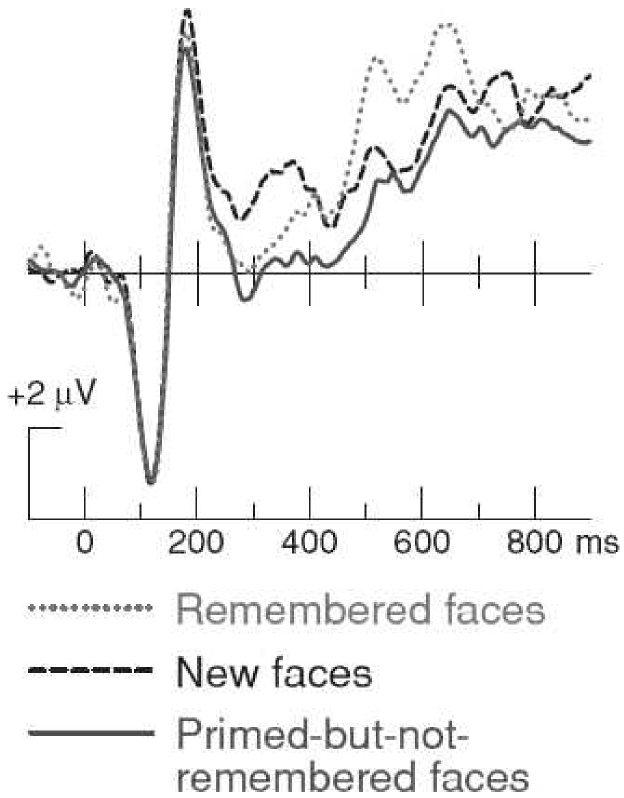

FIGURE 3.4–12. Brain potentials associated with perceptual priming versus declarative memory retrieval. Paller et al. (2003) studied 16 volunteers who took a memory test involving three sorts of faces: new faces, faces they had seen recently and remembered well, and faces they had seen but did not remember because the faces had been presented too briefly to process effectively. In a companion experiment with a priming test, speeded responses were found indicative of priming. Frontal recordings of brain waves elicited by the primed faces included negative potentials from 200 to 400 ms after face presentation that differed from the brain waves elicited by new faces. These differences were particularly robust for trials with the fastest responses (data shown were from trials with responses faster than the median reaction time). Remembered faces uniquely elicited positive brain waves that began about 400 ms after face presentation. Brain potential correlates of face recollection occurred later than those for perceptual priming and were larger over posterior brain regions. (Adapted with permission from Paller KA, Hutson CA, Miller BB, Boehm SG. Neural manifestations of memory with and without awareness. *Neuron.* 38(3):507–516. Copyright © 2003 Cell Press. doi:10.1016/S0896-6273(03)00198-3.)

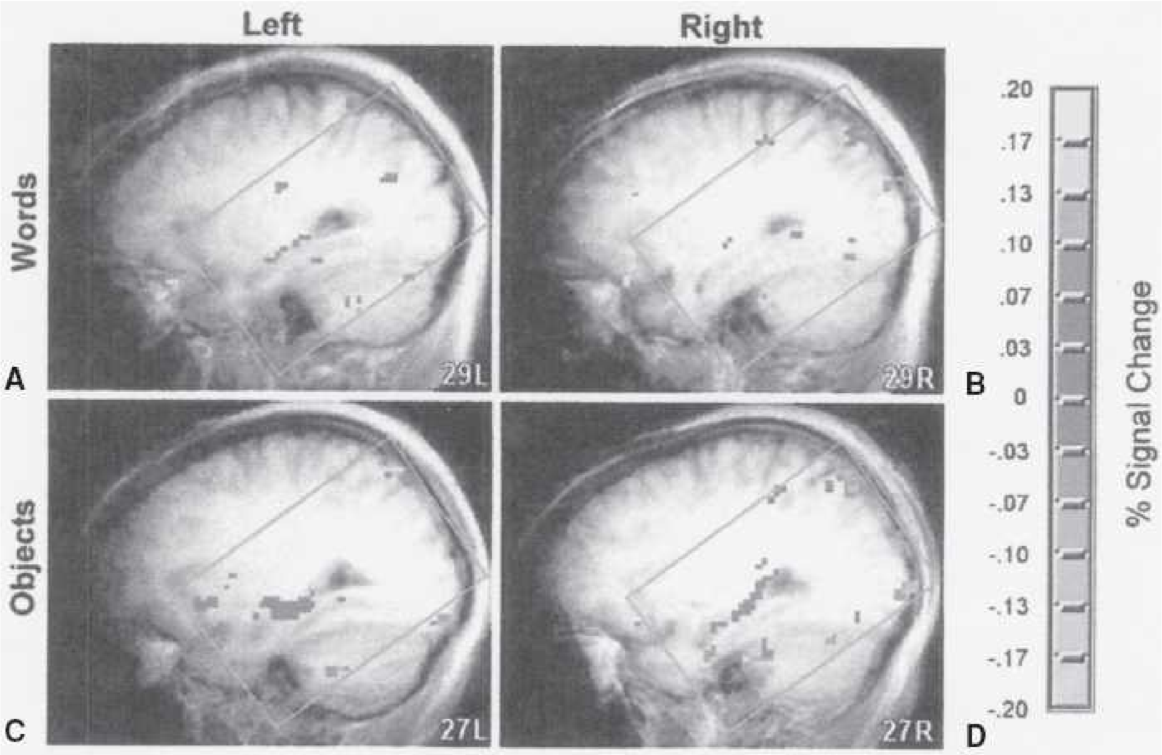

FIGURE 3.4–13. Activity in the left and right hippocampal regions measured with functional magnetic resonance imaging (fMRI) during declarative memory retrieval. Data were collected from 11 participants who saw words at study and test and from 11 different participants who saw pictures of namable objects at study and test. Recognition memory accuracy was 80.2% correct for words and 89.9% correct for objects. Areas of significant fMRI signal change (targets vs. foils) are shown in sagittal sections as color overlays on averaged structural images. The *green box* indicates the area in which reliable data were available for all subjects. **A:** With words, retrieval-related activity was observed in the hippocampus on the left side. **B:** However, retrieval-related activity was not observed in the right hippocampus. **C–D:** With namable objects, retrieval-related activity was observed in the hippocampus on both the left side and the right side. (Adapted from Stark CE, Squire LR. Functional magnetic resonance imaging (fMRI) activity in the hippocampal region during recognition memory. *J Neurosci.* 2000;20:7776. doi: 10.1523/JNEUROSCI.20-20-07776.2000. Copyright © 2000 Society for Neuroscience.)

Hippocampal activity associated with the formation and retrieval of declarative memories has also been investigated with neuroimaging. In keeping with the neuropsychological evidence, the hippocampus appears to be involved in the recollection of recent events (Fig. 3.4–13). Retrieval-related hippocampal activity has been observed in memory tests with many different types of stimuli. The hippocampus is also active during the initial storage of information. Whereas the left inferior prefrontal cortex is engaged as a result of attempts to encode a word, hippocampal activity at encoding is more closely associated with whether encoding leads to a stable memory that can later be retrieved (Fig. 3.4–14). These findings confirm and extend the idea that medial temporal and frontal regions are important for memory storage and that they make different contributions.

Neuroimaging has also been useful in identifying the distributed network of brain regions involved in memory as well as how this network is disrupted by neurologic and psychiatric disorders. For instance, fMRI can be used to define the network of brain areas that presumably interact with the hippocampus by measuring correlations in brain activation while individuals perform memory tasks in addition to when they quietly rest in the MRI scanner. The strength of activity correlation within this hippocampal network predicts declarative memory performance. Many studies have identified reductions or increases in hippocampal network interactions in conditions such as schizophrenia and major depressive disorder (MDD) using fMRI, indicating that this is a promising method for identifying the neurobiologic causes of memory impairment in such disorders.

MEMORY SYSTEMS

One scheme for conceptualizing multiple types of memory appears in Figure 3.4–15. Declarative memory depends on medial temporal and midline diencephalic structures along with large portions of the neocortex. This system provides for the rapid learning of facts (*semantic memory*) and events (*episodic memory*). Nondeclarative memory depends on several different brain systems. Habits depend on the neocortex and the neostriatum. The cerebellum is important for the conditioning of skeletal musculature, the amygdala for emotional learning, and the neocortex for priming. This schematic is an oversimplification in the sense that it does not adequately account for the many examples of memory and other cognitive functions that are shared across regions that are ascribed to distinct “systems.” It is nonetheless a useful starting place for broadly matching observed impairments in individuals with neurologic or psychiatric disorders with potential causes in altered functional neuroanatomy.

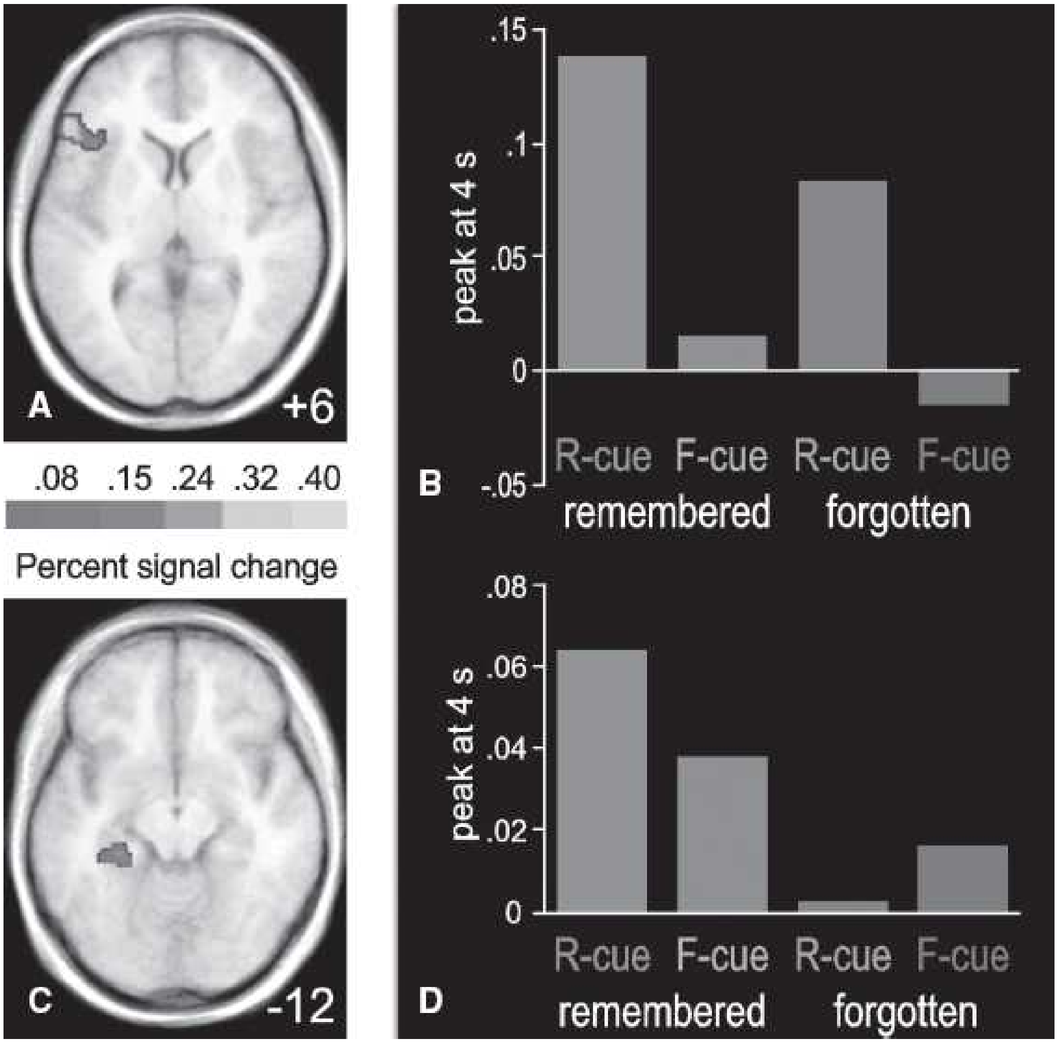
A
+6
.08 .15 .24 .32 .40
Percent signal change
C
-12
B
peak at 4 s
.15
.1
.05
0
-.05
R-cue F-cue R-cue F-cue
remembered forgotten
D
peak at 4 s
.08
.06
.04
.02
0
R-cue F-cue R-cue F-cue
remembered forgotten

FIGURE 3.4–14. Functional activations of prefrontal and medial temporal regions that were predictive of later memory performance. Single words were presented visually, each followed by an instruction to remember (R cue) or to forget (F cue). Trials were sorted based on the remember or forget instruction and on subsequent recognition performance. Activity in the left inferior prefrontal cortex and left hippocampus was predictive of subsequent recognition but for different reasons. **A:** Left inferior prefrontal activation was associated with the encoding attempt, in that responses were largest for trials with a cue to remember whether or not the word was actually recognized later. **B:** The time course of activity in this region was computed based on responses that were time locked to word onset (time 0). Left inferior prefrontal activity increased for words that were later remembered, but there was a stronger association with encoding attempt, as responses were larger for words followed by an R cue that were later forgotten than for words followed by an F cue that were later remembered. **C:** In contrast, left parahippocampal and posterior hippocampal activation was associated with encoding success. **D:** As shown by the time course of activity in this region, responses were largest for words that were subsequently remembered, whether the cue was

to remember or to forget. (Adapted from Reber PJ, Siwiec RM, Gitelman DR, Parrish TB, Mesulam MM, Paller KA. Neural correlates of successful encoding identified using functional magnetic resonance imaging. *J Neurosci.* 2002;22:9541. doi: 10.1523/JNEUROSCI.22-21-09541.2002. Copyright © 2002 Society for Neuroscience.)

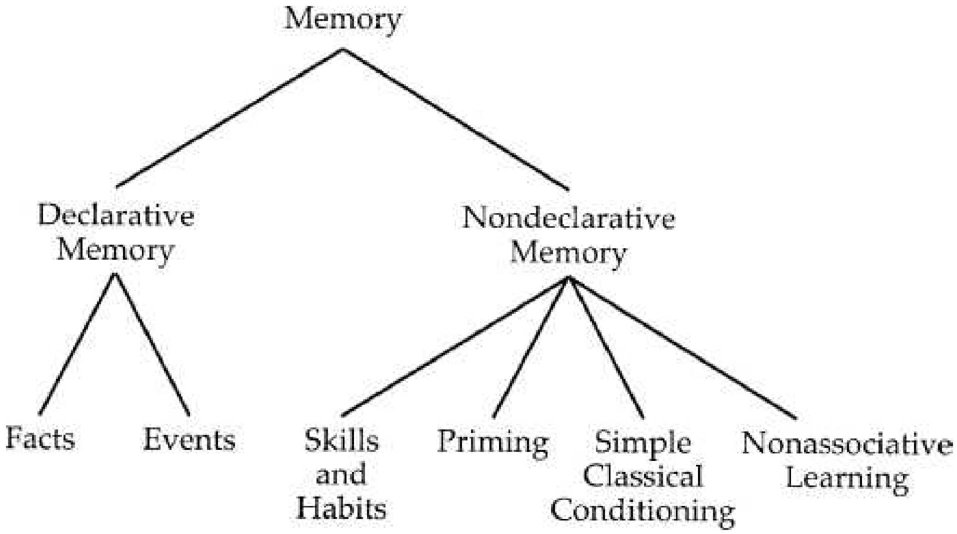

FIGURE 3.4–15. A taxonomy for subtypes of memory. Declarative memory refers to conscious recollection of facts and events. Nondeclarative memory refers to a collection of abilities wherein performance changes as the result of experience but without affording access to conscious memory of the original experience. Nonassociative learning includes habituation and sensitization. (Adapted with permission from Squire LR, Zola-Morgan S. The medial temporal lobe memory system. *Science*. 1991;253(5026):1380–1386. doi:10.1126/science.1896849.)

Declarative and nondeclarative memories differ in important ways. Declarative memory is phylogenetically more recent than nondeclarative memory. In addition, declarative memories are available to conscious recollection. The flexibility of declarative memory permits the retrieved information to be available to multiple response systems. Nondeclarative memory is inaccessible to awareness and is expressed only by engaging specific processing systems. Nondeclarative memories are stored as changes within these processing systems, changes that are encapsulated such that the stored information has limited access to other processing systems.

Semantic memory, which concerns general knowledge of the world, has often been categorized as a separate form of memory. Typically, facts that are committed to memory become independent of the original episodes in which the facts were learned. Patients with amnesia can sometimes acquire information that would ordinarily be learned as facts, but the patients learn it by relying on a different brain system than the system that supports declarative memory.

Consider a test requiring the concurrent learning of eight object pairs. Healthy individuals can rapidly learn which is the correct object in each pair, whereas patients with severe amnesia like EP learn only gradually over many weeks, and at the start of each session, they cannot describe the task, the instructions, or the objects. In patients who do not have severe amnesia, factual information is typically acquired as consciously accessible declarative knowledge. In these cases, the brain structures that remain intact within the medial temporal lobe presumably support learning. In contrast, when factual information is acquired as nondeclarative knowledge, as in the case of EP's learning of pairs of objects, learning likely occurs directly as a habit, perhaps supported by the neostriatum. Humans thus appear to have a robust capacity for habit learning that operates outside of awareness and independently of the medial temporal lobe structures that are damaged in amnesia.

FROM SYNAPSES TO MEMORY

Memory is a special case of the general biologic phenomenon of *neural plasticity.* Neurons can show history-dependent activity by responding differently as a function of prior input, and this plasticity of nerve cells and synapses is the basis of memory. In the last decade of the 19th century, researchers proposed that the persistence of memory could be accounted for by nerve cell growth. This idea has been restated many times since, and the current understanding of the synapse as the critical site of change is founded on extensive experimental studies in animals with relatively simple nervous systems. Experience can lead to structural change at the synapse, including alterations in the strength of existing synapses and alterations in the number of synaptic contacts along specific pathways.

Plasticity

Neurobiologic evidence supports two basic conclusions. First, short-lasting plasticity, which may last for seconds or minutes, depends on specific synaptic events, including an increase in neurotransmitter release. Second, long-lasting memory depends on new protein synthesis, the physical growth of neural processes, and an increase in the number of synaptic connections.

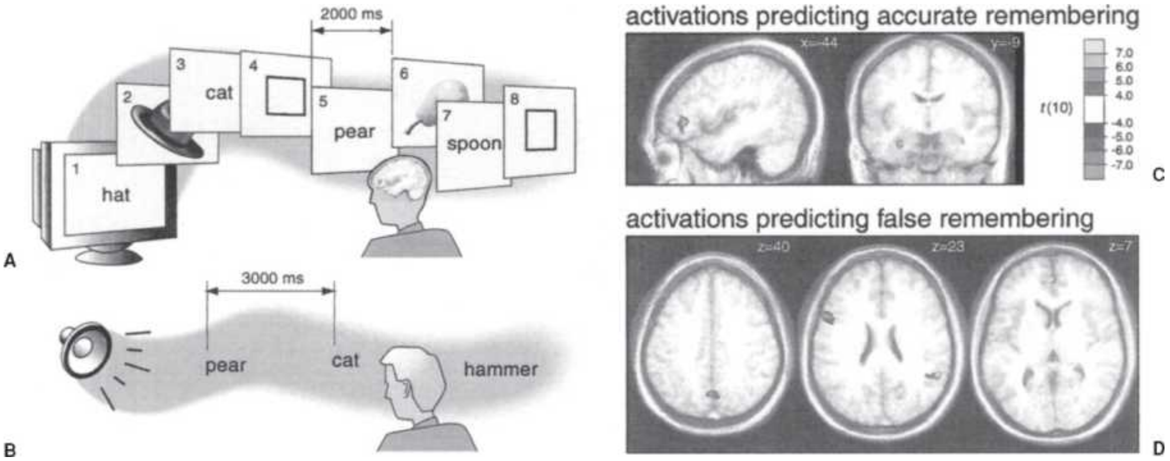

FIGURE 3.4–16. Neural substrates of false memories. **A:** Functional magnetic resonance imaging data were acquired in a learning phase when subjects read names of objects and visualized the referents. One-half of the names were followed 2 seconds later by a picture of the object. **B:** In a surprise memory test given outside the scanner, subjects listened to object names and decided whether they had seen a picture of the corresponding object. On some trials, subjects claimed to have seen a picture of an object that they had only imagined. **C:** Results showed that the left inferior prefrontal cortex and left anterior hippocampus were more active during learning in response to pictures later remembered compared to pictures later forgotten. **D:** Several different brain areas showed a greater response to words in the learning phase that were later falsely remembered as pictures, compared to words not misremembered. Activations that predicted false remembering were found in a brain network important for the generation of visual imagery in response to object names (precuneus, inferior parietal cortex, and anterior cingulate, shown in left, middle, and right images, respectively). (C and D adapted from Gonsalves B, Reber PJ, Gitelman DR, Parrish TB, Mesulam MM, Paller KA. Neural evidence that vivid imagining can lead to false remembering. *Psychol Sci*. 2004;15:655–660.)

A major source of information about memory has come from the extended study of the marine mollusk *Aplysia californica.* Individual neurons and connections between neurons have been identified, and the wiring diagrams of some simple behaviors have been described. *Aplysia* is capable of associative learning (including classic conditioning and operant conditioning) and nonassociative learning (habituation and sensitization). *Sensitization* had been studied using the gill-withdrawal reflex, a defensive reaction whereby tactile stimulation causes the gill and siphon to retract. When tactile stimulation is preceded by sensory stimulation to the head or tail, gill withdrawal is facilitated. The cellular changes underlying this sensitization begin when a sensory neuron activates a modulatory interneuron, which enhances the strength of synapses within the circuitry responsible for the reflex. This modulation depends on a second-messenger system whereby intracellular molecules (including cyclic adenosine monophosphate or cAMP and cAMP-dependent protein kinase) lead to enhanced transmitter release lasting for minutes in the reflex pathway. Both short- and long-lasting plasticity within this circuitry are based on enhanced transmitter release. The long-lasting change uniquely requires the expression of genes and the synthesis of new proteins. Synaptic tagging mechanisms allow gene products that are delivered throughout a neuron to increase synaptic strength selectively at recently active synapses. In addition, the long-term change, but not the short-term change, is accompanied by the growth of neural processes of neurons within the reflex circuit.

In vertebrates, memory cannot be studied quite as directly as in the simple nervous system of *Aplysia.* Nevertheless, it is known that behavioral manipulations can also result in measurable changes in the brain's architecture. For example, rats reared in enriched as opposed to ordinary environments show an increase in the number of synapses ending on individual neurons in the neocortex. These changes are accompanied by small increases in cortical thickness, in the diameter of neuronal cell bodies, and in the number and length of dendritic branches. The behavioral experience thus exerts powerful effects on the wiring of the brain.

Many of these same structural changes have been found in adult rats exposed to an enriched environment, as well as in adult rats given extensive maze training. In studies of maze training, vision was restricted to one eye, and the corpus callosum was transected to prevent information received by one hemisphere from reaching the other hemisphere. The result was that structural changes in neuronal shape and connectivity were observed only in the trained hemisphere. This rules out a number of nonspecific influences, including motor activity, indirect effects of hormones, and overall level of arousal. Long-term memory in vertebrates is thought to be based on morphologic growth and change, including increases in synaptic strength along particular pathways.

Long-Term Potentiation

The phenomenon of long-term potentiation (*LTP*) is a candidate mechanism for mammalian long-term memory. LTP is observed when a postsynaptic neuron is persistently depolarized after a high-frequency burst of presynaptic neural firing. LTP has a number of properties that make it suitable as a physiologic substrate of memory. It is established quickly and then lasts for a long time. It is associative in that it depends on the co-occurrence of presynaptic activity and postsynaptic depolarization. It occurs only at potentiated synapses, not all synapses terminating on the postsynaptic cell. Finally, LTP occurs prominently in the hippocampus, a structure important for memory.

The induction of LTP is known to be mediated postsynaptically and to involve activation of NMDA receptors, which permit the influx of calcium into postsynaptic neurons. Administration of NMDA receptor antagonists (e.g., competitive antagonists, e.g., APV) has been tied to LTP blockage in rat models, including in brain regions known to be critical for memory. Antagonist dosage blocking the induction of LTP impairs the acquisition of novel information.

After initial learning, LTP is expressed and maintained by an increase in the number of postsynaptic AMPA (non-NMDA) receptors and also possibly by increased transmitter release. As such, when NMDA antagonists are delivered *after* the initial acquisition of task information, the dosage has little to no effect on the recall of information learned before antagonist delivery. Postsynaptic NMDA-mediated changes also support the opposing phenomenon of *long-term depression* (LTD), which involves temporary decreases in synaptic efficacy following low-frequency synaptic stimulation. The encoding of long-term memories could involve modification of synaptic connections through LTP and LTD. A promising method for further elucidating molecular mechanisms of memory relies on introducing specific mutations into the genome. By deleting a single gene, mice can be produced with specific receptors or cell-signaling molecules inactivated or altered. For example, in mice with a selective deletion of NMDA receptors in the CA1 field of the hippocampus, many aspects of CA1 physiology remain intact, but the CA1 neurons do not exhibit LTP, and memory impairment is observed in behavioral tasks. Genetic manipulations introduced reversibly in the adult are particularly advantageous in that specific molecular changes can be induced in developmentally normal animals. In particular, by using optogenetics, it is possible to engineer specific populations of neurons such that certain wavelengths of light systematically modify their activity, with specific mutations allowing for light-induced excitation or inhibition. This optogenetic approach for controlling neural activity with high spatial and temporal precision has been used to provide evidence for the involvement of LTP and LTD at specific synapses in long-term memory.

MEMORY IMPAIRMENTS IN PSYCHIATRIC ILLNESS

The biologic description of memory provided by studies of individuals with amnesia, functional neuroimaging, and animal models has deepened the understanding of memory impairments in individuals with a variety of psychiatric illnesses. Abnormal memory is crucial to the etiology of some psychiatric illnesses, such as the pathologic associations between innocuous stimuli and fear that occur in phobias, as well as the intrusive and contextually inappropriate recollections of traumatic events that occur in post-traumatic stress disorder. These disorders have been associated with structural, neurochemical, and functional abnormalities of medial temporal and frontal regions. As further evidence implicating declarative memory, such patients often have moderate impairments of declarative memory for emotional as well as neutral material, suggesting generalized disruption. It remains unclear whether declarative memory disruptions identified in these disorders resulted from premorbid individual differences that conferred risk for the disorder versus whether brain abnormalities accrued over time as part of the disorder. Nonetheless, understanding of biologic processes supporting memory, especially those operating in medial temporal structures, has led to promising developments in the treatment of these disorders. For instance, the extinction of learned fear occurs in a therapeutic setting during cognitive behavioral therapy for phobias and is associated with NMDA-receptor signaling crucial for LTP in the basolateral amygdala. D-cycloserine is a partial agonist of these receptors and has been shown in a number of preclinical and clinical studies to enhance fear extinction and the effectiveness of cognitive behavioral therapy for phobia. Neurochemical modulation of medial temporal structures thus holds promise for the treatment of anxiety disorders and post-traumatic stress disorder.

Memory impairments also occur in psychiatric illnesses that don’t have obvious linkages to memory, such as schizophrenia and major depression. Memory impairments likely occur because these conditions are associated with disruptions of the structure and/or function of medial temporal, midline diencephalic, and frontal neocortical structures that are core components of memory systems. For instance, although schizophrenia is chiefly characterized by delusions, hallucinations, and abnormal emotion, speech, and motivation, patients can experience reduced declarative memory abilities that linger even when core symptoms are pharmacologically treated with success. Patients with schizophrenia perform poorly on memory tests that are highly sensitive to hippocampal structural abnormalities, such as transitive inference and across-domain association learning, and these impairments have been associated with abnormal hippocampal shape and reduced volume and functional activity. These hippocampal structural and functional abnormalities have been shown to worsen with disease progression and are not observed as reliably in other psychotic illnesses, such as psychotic bipolar disorder. Frontal cortex structural and functional abnormality has also been reliably identified in schizophrenia, generally in association to poor working or short-term memory abilities, suggesting impairments in the structuring and organization of information that could further limit declarative memory ability.

Memory impairments are similarly present in major depression. Although memory abnormalities in depression are not core symptoms of the illness and are sometimes understood as mere ramifications of negative affect (e.g., bias toward recall of negative material and rumination on recollection of negative autobiographical episodes), depression patients also exhibit reduced recollection of specific details from previous episodes. Such patients tend to favor overly general or schematic information, even for memories with little emotional content. Again, these memory abnormalities have been associated with structural and functional alterations of medial temporal structures. Thus, schizophrenia and depression are clear examples of conditions in which memory impairments occur in the context of disrupted memory systems and present challenges for patients beyond the symptoms typically associated with the illness. Memory assessment, therefore, could be used to assist clinicians and caregivers in understanding and treating psychiatric patients. Furthermore, lingering cognitive challenges such as memory impairment remain salient barriers to healthy long-term outcomes in patients even when primary symptoms are successfully treated. Greater focus is therefore needed on the development of treatments for memory impairments in a variety of psychiatric disorders.

ASSESSMENT OF MEMORY FUNCTIONS

A variety of quantitative methods are available to assess memory functions in neurologic and psychiatric patients. Quantitative methods are useful for evaluating and following patients longitudinally as well as for carrying out a one-time examination to determine the status of memory function. It is desirable to obtain information about the severity of memory dysfunction, as well as to determine whether memory is selectively affected or whether memory problems are occurring, as they often do, against a background of additional intellectual deficits. The following section presents a rationale for evaluating memory and identifies areas of memory function that should be evaluated in any comprehensive assessment. Although some widely available tests, such as the Wechsler Memory Scale, provide useful measures of memory, most single tests assess memory rather narrowly. Even general-purpose neuropsychological batteries provide for only limited testing of memory functions. A complete assessment of memory usually involves a number of specialized tests that sample intellectual functions, new learning capacity, remote memory, and memory self-report.

The assessment of general intellectual functions is central to any neuropsychological examination. In the case of memory testing, information about intellectual functions provides information about a patient’s general test-taking ability and a way to assess the selectivity of memory impairment. Useful tests include the Wechsler Adult Intelligence Scale; a test of object naming, such as the Boston Naming Test; a rating scale to assess the possibility of global dementia; a test of word fluency; and specialized tests of frontal lobe function.

New Learning Capacity

Memory tests are sensitive to impaired new learning ability when they adhere to either of two important principles. First, tests are sensitive to memory impairment when more information is presented than can be held in immediate memory. For example, one might ask patients to memorize a list of 10 faces, words, sentences, or digits, as 10 items is more than can be held in mind. The paired-associate learning task is an especially sensitive test of this kind. In the paired-associate task, the examiner asks the patient to learn a list of unrelated pairs of words (e.g., queen–garden and office–river) and then to respond to the first word in each pair by recalling the second word.

Second, tests are sensitive to memory impairment when a delay, filled with distraction, is interposed between the learning phase and the test phase. In that case, examiners typically ask patients to learn a small amount of information and then distract them for several minutes with conversation to prevent rehearsal. Recollection is then assessed for the previously presented material. Memory can be tested by the unaided recall of previously studied material (free recall), by presenting a cue for the material to be remembered (cued recall), or by testing recognition memory. In multiple-choice tests of recognition memory, the patient tries to select previously studied items from a group of studied and unstudied items. In yes–no recognition tests, patients see studied and unstudied items one at a time and are asked to say “yes” if the item was presented previously and “no” if it was not. These methods for assessing recently learned material vary in terms of their sensitivity for detecting memory impairment, with free recall being most sensitive, cued recall intermediate, and recognition least sensitive.

The specialization of the function of the two cerebral hemispheres in humans means that left and right unilateral damage is associated with different kinds of memory problems. Accordingly, different kinds of memory tests must be used when unilateral damage is a possibility. In general, damage to medial temporal or diencephalic structures in the left cerebral hemisphere causes difficulty remembering verbal material, such as word lists and stories. Damage to medial temporal or diencephalic structures in the right cerebral hemisphere impairs memory for faces, spatial layouts, and other nonverbal material that is typically encoded without verbal labels. Left medial temporal damage can lead to impaired memory for spoken and written text. Right medial temporal damage can lead to impaired learning of spatial arrays, whether the layouts are examined by vision or by touch. A useful way to test for nonverbal memory is to ask a patient to copy a complex geometric figure and then, after a delay of several minutes, without forewarning, ask the patient to reproduce it.

Remote Memory

Evaluations of retrograde memory loss should attempt to determine the severity of any memory loss and the time period that it covers. Most quantitative tests of remote memory are composed of material in the public domain that can be corroborated. For example, tests have been used that concern news events, photographs of famous persons, or former one-season television programs. An advantage of these methods is that one can sample large numbers of events and can often target particular time periods. A disadvantage is that these tests are not so useful for detecting memory loss for information learned during the weeks or months immediately before the onset of amnesia. Most remote memory tests sample time periods rather coarsely and cannot detect a retrograde memory impairment that covers only a few months. Also, these tests can make it difficult to take into account what the patient might have learned years ago.

In contrast, autobiographical memory tests can potentially provide fine-grained information about a patient's retrograde memory. In the word-probe task, first used by Francis Galton in 1879, patients are asked to recollect specific episodes from their past in response to single-word cues (e.g., bird and ticket) and to date the episodes. The number of episodes recalled tends to be systematically related to the time period from which the episode is taken. Most memories normally come from recent time periods (the past 1 or 2 months), whereas patients with amnesia often exhibit temporally graded retrograde amnesia, drawing few episodic memories from the recent past, but producing as many remote autobiographical memories as normal subjects (Fig. 3.4–6). Endeavors to assess remote memories are hampered by the fact that decades-old autobiographical memories tend to lose some detail, and these memories may also be subject to further modification if memory functions are intact. Patients who have amnesia for a long time may also adopt strategies or habits that promote gist-based memory retrieval that may interfere with retrieving specific details from remote events.

Memory Self-Reports

Patients can often supply descriptions of their memory problems that are extremely useful for understanding the nature of their impairment. Tests of the ability to judge one's own memory abilities are called tests of *metamemory.* Self-rating scales are available that yield quantitative and qualitative information about memory impairment. As a result, it is possible to distinguish memory complaints associated with depression from memory complaints associated with amnesia. Depressed patients tend to rate their memory as poor in a rather undifferentiated way, endorsing all the items equally on a self-rating form. By contrast, patients with amnesia tend to endorse some items more than others; that is, there is a pattern to their memory complaints. Patients with amnesia do not report difficulty in remembering very remote events or in following what is being said to them, but they do report having difficulty remembering an event a few minutes after it happens. Indeed, self-reports can match rather closely the description of memory dysfunction that emerges from objective tests. Specifically, new learning capacity is affected, immediate memory is intact, and very remote memory is intact. Some patients with amnesia, however, tend to have poor awareness of their memory impairment, sometimes overestimating and sometimes underestimating their ability. In patients with Korsakoff syndrome, for example, their poor metamemory stems from frontal lobe dysfunction. In any case, querying patients in some detail about their own sense of impairment and administering self-rating scales are valuable and informative adjuncts to more formal memory testing.

Psychogenic Amnesia

Patients sometimes exhibit memory impairment that differs markedly from the typical patterns of memory loss that follow brain damage. For example, some cases of amnesia present with a sudden onset of retrograde amnesia, a loss of personal identity, and minimal anterograde amnesia. These patients may even be unable to recall their own name. Given the psychological forces that prompt the onset of amnesia in these cases, they are commonly termed *psychogenic amnesia,* or sometimes *hysterical amnesia, functional amnesia,* or *dissociative amnesia.*

Differentiating psychogenic amnesia from a memory disorder that results from frank neurologic injury or disease is often straightforward. Psychogenic amnesias typically do not affect new learning capacity. Patients enter the hospital able to record a continuing procession of daily events. By contrast, new learning problems tend to be at the core of neurologic amnesia. The main positive symptom in psychogenic amnesia is extensive and severe retrograde amnesia. Patients may be unable to recollect pertinent information from childhood or from some part of their past. Formal neuropsychological testing has shown that the pattern of memory deficits varies widely from patient to patient. This variability may reflect a patient's commonsense concepts of memory, even when symptoms are not the result of conscious attempts to simulate amnesia. Some patients may only perform poorly when they are asked to remember past autobiographical events. Other patients may fail as well at remembering past news events. Some patients perform well when memory tests appear to assess general knowledge, such as remembering the names of celebrities or cities. Learning new material is usually intact, perhaps because such tests appear to concern the present moment, not traveling into the past. Occasionally, patients with psychogenic amnesia exhibit broad memory deficits such that they cannot perform previously familiar skills nor identify common objects or common vocabulary words.

By contrast, patients with neurologic amnesia never forget their names, and remote memory for the events of childhood and adolescence is typically normal unless there is damage to the lateral temporal or frontal lobes. Patients with psychogenic amnesia sometimes have evidence of head trauma or brain injury, but nevertheless, the pattern of deficits cannot be taken as a straightforward result of neurologic insult. The clinician's challenge is not to distinguish psychogenic amnesia from neurologic amnesia but to distinguish psychogenic amnesia from malingering. Indeed, the diagnosis of psychogenic amnesia can be difficult to substantiate and may be met with skepticism by hospital staff. Some features that argue in favor of a genuine psychogenic disorder include (1) memory test scores that are not as low as possible and never worse than chance levels; (2) memory access improved by hypnosis or Amytal interview; and (3) a significant premorbid psychiatric history. In some cases, psychogenic amnesia has been observed to clear after a period of days, but in many cases, it has persisted as a potentially permanent feature of the personality.

Patient CB was a 29-year-old man evaluated for memory loss. He reported having severe retrograde amnesia for the past 3months. A medical evaluation revealed swelling of the left side of his head due to recent head trauma. CT head scan was normal. 3months prior to this evaluation, he had been arrested for intoxication and taken to jail. Upon awakening the next day, he was unable to recall his past life. At the time of the evaluation, he indicated that he knew his name, but it was not clear whether he knew his name only because he had relearned it. He also had some difficulty recalling old vocabulary words. He had essentially no memories of events that had occurred before the onset of his amnesia. He was, however, able to relearn about his past life as he was told about it. He also learned new information well. He sometimes did not recognize vocabulary words and had to look them up. He reported that he had recently relearned how to drive a car with a manual transmission.

It was notable that he described himself as having a new personality since the onset of memory problems. Previously he had been very outgoing, a heavy drinker of alcohol, and was feared by some people. He had been “obnoxious but likable.” Since the onset of memory problems, he was less outgoing, more introspective, and better at getting along with other people. He did not like his personality from before the onset of memory problems, and, in fact, he had adopted his former middle name as a new first name. He hoped that he would be able to regain all of his old memories and experiences without the return of his old personality. During a 20-month follow-up period, his severe retrograde amnesia persisted. (Case history from Kritchevsky M, Chang J, Squire LR. Functional amnesia: clinical description and neuropsychological profile of 10 cases. *Learn Mem*. 2004;11:213.)

IMPLICATIONS

Memory Distortion

The current understanding of the biology of memory has significant implications for several fundamental issues in psychiatry. Given the selective and constructive nature of autobiographical remembering and the imperfect nature of memory retrieval more generally, it is surprising that memory is so often accurate. How much can memories be trusted? Subjective feelings of confidence are apparently not perfect indicators of retrieval accuracy. And memory distortion can clearly lead to unfortunate consequences, such as when mistaken eyewitness testimony harms an innocent individual.

The fact is that it is possible to confidently remember events that never happened. For example, it is possible to confuse an event that was only imagined, or dreamed about, with an event that actually occurred. One factor that contributes to memory distortion is that similar brain regions are important both for visual imagery and for the long-term storage of visual memories (Fig. 3.4–16). Another factor that contributes to memory distortion is that memory functions best in remembering the gist of an event, not the particulars from which the gist is derived. In a noted demonstration, people listen to a list of words: *candy, sour, sugar, tooth, heart, taste, dessert, salt, snack, syrup, eat,* and *flavor.* Subsequently, when asked to write down the words they heard, 40% of them wrote down the word *sweet,* even though this word did not appear in the list. Thus, many people in this demonstration failed to discriminate between the words that had been presented and a word that was strongly associated with all the words but had not itself been presented. The word *sweet* can be thought of as a gist word, a word that represents the other words and that captures the meaning of the whole list. Presumably, the words in the study list evoked a thought of the word *sweet,* either at the time of learning or during the memory test, and people then tended to confuse merely thinking of the word with actually hearing it.

The reconstructive nature of recollection means that the interpretation of eyewitness testimony is not straightforward. Whole episodes are not available in the neocortex but rather must be pieced together based on fragmentary components and in the context of potentially misleading influences present at the time of retrieval. Studies in adults and in children have documented that illusory memories can be created. Children are particularly susceptible to these effects, especially when subjected to leading questions and false suggestions.

In view of these features of memory, when a memory of childhood abuse is remembered after many years, it is prudent to ask whether the memory is accurate. Genuine examples of memory recovery have been documented, whereby an individual produces a veridical memory for a past traumatic event after not recalling the event for extended time periods. Numerous examples of apparent memory recovery have also been subsequently discovered to be instances of false memory. Unfortunately, there is no perfect method at present, in the absence of independent corroboration, for determining whether a recollective experience is based on a real event or not.

Infantile Amnesia

The biology of memory has also provided insights relevant to the phenomenon of infantile amnesia, the apparent absence of conscious memory for experiences from approximately the first 3 years of life. Traditional views of infantile amnesia have emphasized repression (psychoanalytic theory) and retrieval failure (developmental psychology). A common assumption has been that adults retain memories of early events but cannot bring them into consciousness. However, it now appears that the capacity for declarative memory does not become fully available until approximately the third year of life, whereas nondeclarative memory emerges early in infancy (e.g., classical conditioning and skill learning). Thus, infantile amnesia results not from the adult's failure to retrieve early memories but from the child's failure to store them adequately in the first place.

Nevertheless, studies in young infants show that a rudimentary capacity for declarative memory is present even at a few months of age. As a child develops, memories can be retained across increasingly long intervals, and what is represented becomes correspondingly richer and more full of detail. Medial temporal and diencephalic regions seem to be sufficiently developed during these early months and years. What limits the capacity for declarative memory appears to be the gradual development and differentiation of the neocortex.

As the neocortex develops, the memories represented there become more complex, language abilities allow for more elaborate verbal descriptions of events, and a developing sense of self supports autobiographical knowledge. As new strategies emerge for organizing incoming information, declarative memories become more persistent, more richly encoded, and better interconnected with other information. It is not the case that fully formed childhood memories are stored but cannot be retrieved. The perspective consistent with the current understanding of the biology of memory is that declarative memories formed very early in life are fragmentary, simple, and tied to the specific context of an infant's understanding of the world. They are unlike typical declarative memories in adults, imbued with meaning and a complex understanding of events.

Technologic Interventions for Memory Impairment

The biology of memory has motivated recent attempts to rescue memory impairments via stimulation-based neurotechnology. For instance, conventional forms of noninvasive brain stimulation such as transcranial magnetic stimulation (TMS) cannot directly affect the core regions that are dysfunctional in memory disorders, such as the hippocampus, but they can readily influence portions of the broader hippocampal network as defined via fMRI. Several studies have shown improvements in declarative memory ability following multisession TMS of the hippocampal network, including positive results from early-stage trials in individuals with memory decline due to aging and early-stage Alzheimer disease. There are also promising findings that memory functions supported by the frontal lobes, such as working memory, can be improved using transcranial electrical stimulation when delivered at appropriate times relative to the targeted cognitive processing. Thus, there are several promising avenues for the rescue of memory impairments via noninvasive stimulation technologies.

Other experiments have evaluated deep brain stimulation (DBS) for memory enhancement. DBS is an FDA-approved treatment for movement disorders such as Parkinson disease, essential tremor, and dystonia. DBS has also recently been approved to treat epilepsy and obsessive-compulsive disorders. These recent advances have bolstered interest in its application for neuropsychiatric and memory disorders.

Converging evidence from stimulation of depth electrodes implanted in patients with epilepsy and mouse models suggests that DBS of the fornix may improve learning or memory consolidation. In sheep, high-frequency DBS of the fornix further effected changes in electrical activity in the hippocampus. However, studies examining DBS of the hippocampus or mesial temporal regions are relatively inconclusive. Further, effects of stimulation applied at the fornix in recent Phase 1, and 2 clinical trials in Alzheimer disease demonstrate that while stimulation is safe and tolerable for patients, DBS has variable long-term outcomes for subsequent memory decline. Exploratory analyses indicate that DBS may be most beneficial for Alzheimer patients who are in the early stages of the disease, but DBS had little to no effect in a large clinical trial of early

Alzheimer patients under the age of 65. The therapeutic efficacy of forniceal DBS thus remains relatively unclear but will continue to be examined across ongoing Phase 2 and 3 clinical trials.

Memories and the Unconscious

The existence of multiple memory systems also has implications for issues central to psychoanalytic theory, including the construct of the *unconscious.* How past experience influences current behavior depends on what view one takes of the nature of memory. By the traditional view, memory is a unitary faculty, and representations in memory vary mainly in strength and accessibility. Material that is unconscious is below some threshold of accessibility but could potentially be made available to consciousness.

The modern, biologic view begins with the distinction between a kind of memory that can be brought to mind—declarative memory—and other kinds of memory that are, by their nature, unconscious. Stored nondeclarative memories are expressed through a performance without affording any conscious memory content. Personalities are shaped by nondeclarative memories in the form of numerous habits and conditioned responses. In this view, one's behavior is indeed affected by events from early life, but the effects of early experience persist in a nondeclarative form without necessarily including an explicit, conscious record of the events. Learned behavior can be expressed through altered dispositions, preferences, conditioned responses, habits, and skills, but exhibiting such behavior need not be accompanied by the awareness that behavior is being influenced by past experience, nor is there a necessity that any particular past experience has been recorded as a complete episode. That is, an influence from early experience does not require the memory of any specific episode. One can be afraid of dogs without remembering being knocked down by a dog as a child. In this case, the fear of dogs is not experienced as a memory. It is experienced as a part of personality. Furthermore, a strong fear of dogs carries with it no implication that the brain retains a specific record of any early experience that subsequently resulted in fear of dogs.

Behavioral change can occur by acquiring new habits that supersede old ones or by becoming sufficiently aware of a habit that one can to some extent isolate, countermand, or limit the stimuli that elicit it. However, one need not become aware of any early formative event in the same sense that one knows the content of a declarative memory. The unconscious does not become conscious. Various forms of nondeclarative memory simply influence behavior without having the additional capacity for these influences to become accessible to conscious awareness.

ACKNOWLEDGMENT

The authors gratefully acknowledge contributions to the present article from Larry Squire and Ken Paller (authors of “Biology of Memory” in prior editions of this book).

Further Readings

Addis DR, Barense M, Duarte A. *The Wiley Handbook on the Cognitive Neuroscience of Memory*. Wiley-Blackwell; 2015.

Baddeley AD, Wilson BA, Watts FN. *Handbook of Memory Disorders*. 2nd ed. John Wiley and Sons; 2002.

Barrett LF, Lewis M, Haviland-Jones JM. *Handbook of Emotions*. 4th ed. Guilford Press; 2016.

Bayley PJ, Frascino JC, Squire LR. Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature*. 2005;436:550.

Bayley PJ, Hopkins RO, Squire LR. Successful recollection of remote autobiographical memories by amnesic patients with medial temporal lobe lesions. *Neuron*. 2003;38:135.

Byrne JH. *Learning and Memory: A Comprehensive Reference*. 2nd ed. Academic Press; 2017.

Kritchevski M, Chang J, Squire LR. Functional amnesia: clinical description and neuropsychological profile of 10 cases. *Learn Mem*. 2004;11:213.

Clark RE, Manns JR, Squire LR. Classical conditioning, awareness, and brain systems. *Trends Cogn Sci*. 2002;6:524.

Corkin S. *Permanent Present Tense: The Unforgettable Life of the Amnesic Patient, H.M*. Basic Books; 2013.

Dalgleish T, Werner-Seidler A. Disruptions in autobiographical memory processing in depression and the emergence of memory therapeutics. *Trends Cogn Sci*. 2014;18:596.

Dew ITZ, Cabeza R. The porous boundaries between explicit and implicit memory: behavioral and neural evidence. *Ann NY Acad Sci*. 2011;1224(1):174–190.

Eichenbaum H. Still searching for the engram. *Learn Behav*. 2016;44:209.

Fuster JM. *Cortex and Mind: Unifying Cognition*. Oxford University Press; 2003.

Hebscher M, Voss JL. Testing network properties of episodic memory using noninvasive brain stimulation. *Curr Opin Behav Sci*. 2020;32:35–42.

Heckers S, Konradi C. Hippocampal pathology in schizophrenia. *Curr Top Behav Neurosci*. 2010;4:529.

Inostroza M, Born J. Sleep for preserving and transforming episodic memory. *Annu Rev Neurosci*. 2013;8:79–102.

Kandel ER. Biology and the future of psychoanalysis: a new intellectual framework for psychiatry revisited. *Am J Psychiatry*. 1999 Apr;156(4):505–24.

Kandel ER. The molecular biology of memory storage: a dialogue between genes and synapses. *Science*. 2001;294:1030.

Kopelman MD. Disorders of memory. *Brain*. 2002;125:2152.

Laxton AW, Lozano AM. Deep brain stimulation for the treatment of Alzheimer disease and dementias. *World Neurosurg*. 2013;80:S28.e1.

Mankin EA, Fried I. Modulation of human memory by deep brain stimulation of the entorhinal-hippocampal circuitry. *Neuron*. 2020;106(2):218–235.

Paller KA. Consolidating dispersed neocortical memories: the missing link in amnesia. *Memory*. 1997;5:73.

Paller KA, Wagner AD. Monitoring the transformation of experience into memory. *Trends Cogn Sci*. 2002;6:93.

Parsons RG, Ressler KJ. Implications of memory modulation for post-traumatic stress and fear disorders. *Nat Neurosci*. 2013;16:146.

Rudoy JD, Weintraub S, Paller KA. Recall of remote episodic memories can appear deficient because of a gist-based retrieval orientation. *Neuropsychologia*. 2009;47:938–941.

Schacter DL. *Memory Distortion: How Minds, Brains, and Societies Reconstruct the Past*. Harvard University Press; 1995.

*Schacter DL. *Searching for Memory: The Brain, the Mind, and the Past*. Basic Books; 1996.

*Squire LR, Kandel E. *Memory: From Mind to Molecules*. 2nd ed. Scientific American Library; 2009.

*Squire LR, Schacter DL, eds. *Neuropsychology of Memory*. 3rd ed. Guilford; 2002.

Squire LR, Zouzounis JA. Self-ratings of memory dysfunction: different findings in depression and amnesia. *J Clin Exp Neuropsychol*. 1988;10:727.

Staniloiu A, Markowitsch HJ. Dissociative amnesia. *Lancet Psychiatry*. 2014;1:226.

Tulving E. Episodic memory: from mind to brain. *Annu Rev Psychol*. 2002;53:1.

*Voss JL, Lucas HD, Paller KA. More than a feeling: pervasive influences of memory processing without awareness of retrieval. *Cogn Neurosci*. 2012;3:193–207.

3.5 Normality and Mental Health

George E. Vaillant, M.D.

Too often, psychiatry has been only preoccupied with mental illness. To paraphrase Mark Twain's quip about the weather, psychiatry is always talking about mental health, but nobody ever does anything about it. Science has conceptualized the building blocks of nuclear fission more readily than the building blocks of mental health. Thus, with the notable exception of the chapter by Daniel Offer and Melvin Sabshin in the early editions of this textbook, a review of other recent major psychiatric textbooks reveals virtually no serious discussion of mental health. One reason for this lack of attention is that the study of positive mental health is a very new field. Only within the last three decades has mental health per se been addressed empirically instead of platonically.

There has been an implicit assumption that mental health could be defined as the antonym of mental illness. In other words, mental health was the absence of psychopathology and synonymous with "normal." Achieving mental health through the alleviation of gross pathologic signs and symptoms of illness is also the definition of the mental health model strongly advocated by third-party payers. Indeed, viewing mental health as simply the absence of mental illness is at the heart of much of the debate concerning mental health policies. The great epidemiological studies of the past half-century have also focused on who was mentally ill and not who was well. Only the Sterling County Studies by Alexander Leighton came close to defining positive mental health operationally. Mental illness, after all, is a condition that is easy to define reliably, and its limits are relatively clear. As a result, it has been argued that achieving above-average mental or physical health is not the province of medicine but of education. Such a definition ignores concepts of positive health like physical fitness and cardiac reserve.

Like physical fitness, however, positive mental health is too important to be ignored. To believe mental health is a Global Assessment of Function score of 70 is to underestimate human potential. Sports medicine does more than mend a skier's broken legs; sports medicine rehabilitates the skier to ski advanced trails once more. Starting early in the last century, internists began studying physiology at high altitudes and devising measures of positive physical health for athletes, pilots, and finally astronauts. In 1929–1930 at the University of California at Berkeley, the Institute of Human Development was founded by Howard Jones, Nancy Bayley, and Jean McFarlane. Originally founded to study healthy child development, the Institute was to provide a seminal influence on Erik Erikson's model of healthy adult development. A decade later, in the late 1930s, Arlie Bock, an internist trained in high-altitude physiology and interested in positive physical health, began at the Harvard University Health Services, The Study of Adult Development. The study was designed as an interdisciplinary study of both mental and physical health. Results from that study, lasting for 70 years, have informed many facets of this chapter.

Although above-average mental health is more difficult to define than above-average physical fitness, it is important for psychiatrists to emulate sports medicine in order to provide precise definitions and measures of positive mental health. Psychologists have already learned to quantify not only "normal" but better than normal intelligence. Thus, the antonym of mental retardation is not regarded as an intelligence quotient (IQ) of 100 but as an IQ over 130. Psychiatry must follow suit.

Certainly, over the last 70 years, psychiatrists have become increasingly involved in mental health consultations with agencies. Rather than merely deciding who is too sick for a job, they are called on to make decisions about who is mentally healthy enough for certain positions—like air controllers and submariners. Analogous to cardiac reserve, the measurement of resilience and the capacity to withstand adversity is on psychiatry's psychometric wish list. In addition, interest in the empirical evaluation of psychiatric therapy outcomes has also brought the issue of positive mental health into focus. Indeed, one of the weaknesses of much existing work on psychotherapy assessment has been a lack of clarity regarding outcome definitions. *For health is not the absence of negatives but the presence of positives.*

But the definition of positive mental health is not easy. Several cautionary steps are necessary. The first step in discussing mental health is to note that "average" is not healthy; it always includes mixing in with the healthy the prevalent amount of psychopathology in the population.

For example, in the general population, “average” weight or eyesight is unhealthy; and if all sources of biopsychosocial pathology were excluded from the population, the average IQ would be significantly above 100. For as with Garrison Keillor’s description of Lake Wobegon youth, mentally healthy children “are all above average.” Being at the center of a normal bell curve of distribution may or may not be healthy. In the case of red blood count, body temperature, or mood, the middle of the bell curve *is* healthy. In the case of eyesight, exercise tolerance, or empathy, only the upper end of the bell curve is healthy; and in the case of serum cholesterol, bilirubin, and narcissism, only the low end of the curve is healthy.

The second cautionary step in discussing mental health is to appreciate the caveat that what is healthy sometimes depends upon geography, culture, and historical moment. Sickle cell trait is unhealthy in New York City, but in the tropics where malaria is endemic, the sickling of red blood cells may be lifesaving. Punctuality is a virtue in Germany and sometimes a failure in Brazil. Indeed, some aboriginal Australian languages do not even have a word for time. General George Patton’s competitive temperament was for him a psychological liability in times of peace but a virtue in two World Wars. In World War II, paranoid personalities made very poor submariners but excellent airplane spotters.

A third cautionary step is to make clear whether one is discussing *trait* or *state.* Who is physically healthier: an Olympic miler disabled by a simple but temporary (state) sprained ankle or a Type 1 diabetic (trait) with temporarily normal blood sugar? In cross-cultural studies, such differences become especially important. Superficially, an Indian mystic in a state of trance may resemble a catatonic schizophrenic but not over time.

In defining mental health, the fourth and most important cautionary step is to appreciate the twofold danger of “contamination by values.” On the one hand, cultural anthropology teaches how fallacious any definition of mental health can be. Competitiveness and scrupulous neatness may be healthy in one culture and regarded as a personality disorder in another. And, if mental health is “good,” what is it good for? The self or the society? For “fitting in” or for creativity? For happiness or survival? And who should be the judge? Since every culture differs in its diet, the World Health Organization (WHO) should never be called upon to design restaurant menus. For ten years, Adolf Hitler was a better leader of his country than

Jimmy Carter; but by every definition that the author knows, Carter had better mental health than Hitler.

Thus, common sense, not postmodernism, must prevail. Biology trumps anthropology. Health is the activity of a living body in accordance with its specific excellences.

"What, for example, is a healthy squirrel?" asked Kass, a research professor in bioethics and neurology, "Not a picture of a squirrel, not really or fully the sleeping squirrel, not even the aggregate of his normal blood pressure, serum calcium, total body zinc, normal digestion, fertility, and the like. Rather, the healthy squirrel is a bushy-tailed fellow, who looks and acts like a squirrel; who leaps through the trees with great daring, who gathers, buries, and covers but later uncovers and recovers his acorns; who perches out on a limb cracking his nuts, sniffing the air for smells of danger, alert, cautious, with his tail beating rhythmically; who chatters and plays and courts and mates and rears his young in large improbable looking homes at the tops of trees; who fights with vigor and forages with cunning; who shows spiritedness, even anger, and more prudence than many human beings … health is a natural standard or norm—not a moral norm, not a 'value' as opposed to a 'fact,' not an obligation but a state of being that reveals itself in activity."

It is true that cultural anthropology teaches that almost no form of behavior is considered abnormal in all cultures, but that does not mean that the tolerated behavior is mentally healthy. It would be an error for the WHO to ignore the universal importance of vitamins or the four basic food groups. Just because, until recently, Portugal did not recognize alcoholism as an illness does not reduce the contribution of alcoholism in Lisbon to morbidity. Just because the American Constitution protected the right to own slaves as inalienable did not make slavery mentally healthy for slaves or for the slave owner. Just because martinis and Coca-Cola outsell carrot juice does not make the former healthier. The best way to enrich the current understanding of what constitutes mental health is to study a variety of healthy populations from different perspectives, in different cultures, and for a long period of time.

This chapter will contrast eight different empirical approaches to mental health. Significantly, the empirical underpinnings of all eight models have emerged only in the past 50 years. First, mental health can be conceptualized as *Above Normal* and a mental state that is objectively desirable—as in Sigmund Freud's alleged definition of mental health was the capacity to work and to love. Second, mental health can be conceptualized as *Positive Psychology*—as epitomized by the presence of multiple human strengths. Third, from the viewpoint of healthy adult development, mental health can be conceptualized as *Maturity.* Fourth, mental health can be conceptualized as *Resilience,* as the capacity for successful adaptation, "grit," and homeostasis. In such a view, mental health, analogous to a competent immune system, allows the individual to function well despite stressful or dangerous environments. Fifth, mental health can be conceptualized as *Emotional Intelligence* and wise and successful object relations. Sixth, mental health can be conceptualized as *Subjective Well-Being* (*SWB*)—a mental state that is subjectively experienced as happy, contented, and desired. Seventh, mental health can be conceptualized as ***Positive or "Spiritual" Emotions***.

MODEL A: MENTAL HEALTH AS ABOVE NORMAL

This first perspective differs from the traditional medical approach to health and illness. Until recently, being without manifest psychopathology equaled mental health. In this medical model, if one were to put all individuals on a continuum, normality would encompass the major portion of young adults, and abnormality would be the small remainder. This definition of health correlates with the traditional role model of the doctor who attempts to free his patient from grossly observable signs of illness. In other words, in this context, health refers to a reasonable, rather than an optimal, state of functioning.

Yet, as already pointed out, mental health is not normal; it is above average. Indeed, some believe that true mental health is the exception, not the rule. Moreover, until recently, some believed mental health was imaginary. Many postwar psychiatrists continued to agree with Sigmund Freud, who had dismissed mental health as "an ideal fiction." In the late 1950s, two of the world's most distinguished psychiatrists could dismiss the term entirely. Sir Aubrey Lewis wrote in 1958, "Mental health is an invincibly obscure concept," and in 1957, Fritz Redlich asserted, "We do not possess any general definition of normality and mental health from either a statistical or a clinical viewpoint."

Certainly, the absence of illness and the presence of health overlap but do not always coincide. In the military, the mental health of a jet pilot must exceed the draft board's 1-A. In primary school teachers and in Supreme

Court justices, it is more than simply freedom from symptoms or Freud's capacity to work and to love that is looked for. However, to avoid quibbling over what traits characterize mental health, it is helpful to adopt the analogy of a decathlon champion. What constitutes a "track star"? Is it muscle strength, or speed, or endurance, or grace, or competitive grit? Does not the definition differ from nation to nation and from century to century? Not really. The salience of a given facet of track or of mental health may vary from culture to culture, but all are important. A high score on the decathlon has conveyed behavioral skill in track for millennia and in every country of the world—even as the hair splitters argue over the right words with which to describe a track athlete.

As with excellence in the decathlon, no single measure defines mental health, but all measures are highly intercorrelated. Table 3.5–1 illustrates how multifaceted and unique each model of positive mental health can be and how very different the semantics can be. But the familial resemblance of each definition to each other is unmistakable.

In 1835 Adolphe Quetelet published the first important book on normality ever written. Rather than focus on pathology, he tried "to approach more closely to what is good and beautiful," and his goal was the statistical analysis of healthy humans. He challenged generations of future investigators with his introductory sentence, "Man is born, grows up, and dies, according to certain laws which have never been properly investigated" (p. 5).

Until World War II, however, Quetelet's challenge to mental health workers went largely unnoticed. When in 1941, John Clausen and his coworkers were commissioned to assess mental health for the draft board, they were embarking on a novel task. As a way of assessing the mental health of recruits, they focused on the absence of psychosomatic symptoms. Although limited, their approach was not abandoned until the 1970s, and questions about psychosomatic symptoms still form an important part of the Hopkins SCL-90 and of scales assessing neuroticism.

Table 3.5–1.
Three Contrasting Definitions of Mental Health

Model A Mental Health/Normality	Model B Positive Psychology	Model C Maturity
Marie Jahoda, as summarized by Daniel Offer and Melvin Sabshin	Christopher Peterson, Martin Seligman	William Menninger
Ability to love, work, and play Possesses empathy adequacy in interpersonal relations	**Love:** Intimacy/reciprocal attachment Kindness/generosity/nurturance Social intelligence/emotional intelligence	**The capacity for love** The capacity for a variety of mutually fulfilling and lasting relationships The need to seek a major source of fulfillment in productive work
Efficient in problem-solving Perceives reality Resistance to stress Environmental mastery	**Temperance:** Forgiveness/mercy Modesty/humility Prudence/caution Self-control/self-regulation	Free of stereotyped and nonproductive patterns of problem-solving Ability to discharge hostility without harming others or oneself The capacity to adapt to change and endure frustration and loss
Invested in life Self-actualization Oriented toward future	**Wisdom and Knowledge:** Curiosity/interest Love of learning Judgment/open-mindedness	A realistic acceptance of the destiny imposed by one’s

	Perspective Creativity/originality/ingenuity	time and place in the world
Autonomy Recognition of needs In touch with their own identity and feelings	**Courage:** Valor Honesty/authenticity Industry/perseverance Zest/enthusiasm	Appropriate expectations and goals for oneself Ability to respond to the uncertainties of reality in a manner consistently free of domination by one’s wishes or peers
	Justice: Citizenship/loyalty/teamwork Equity/fairness Leadership	
	Transcendence: Awe/wonder Gratitude Hope/future mindedness Spirituality/faith Playfulness/humor	**Having the capacity for hope** An altruistic concern for other human beings outside one’s own group and beyond one’s own time and place The capacity to suspend one’s adult identity and engage in childish play at appropriate times

After World War II, influential works on normal adaptive behavior began to be published—Roy Grinker and John Spiegel’s *Men under Stress,* Robert White's *Lives in Progress,* Leo Sroles' *Mental Health in the Metropolis,* and Alexander and Dorothea Leightons' *Cove and Woodlot* were four of the more important. Although all four studies concentrated on the adaptation of nonpatient or normal populations, they still put their emphasis on *not pathological* rather than on *above-normal mental health.*

Then, in 1958, Marie Jahoda's report to the Joint Commission on Mental Illness and Health led to a psychiatric sea change regarding the existence of mental health. As illustrated in Table 3.5–1, Jahoda suggested that mentally healthy individuals should (1) be in touch with their own identity and their own feelings. (2) They should be oriented toward the future, and over time they should remain fruitfully invested in life. (3) Their psyches should be integrated and provide them resistance to stress. (4) They should possess autonomy and recognize what suits their needs. (5) They should perceive reality without distortion and yet possess empathy. (6) They should be masters of their environment—able to work, to love, and to play, and to be efficient in problem-solving. Instead of emphasizing the absence of negative symptoms, Jahoda underscored a number of positive traits. Nevertheless, although the purpose of Jahoda's report was to rid the term, mental health, of "vague, elusive and ambiguous connotation," at the time she published her criteria, there was still no evidence to prove that her plausible definition was more than mere platitudes that reflected her own culture-bound values.

Thus, it was with Roy Grinker's 1962 studies of "homoclites" that investigators began the empirical study of positive mental health. Grinker's homoclites were physical education majors selected for normality but studied only briefly. A second longitudinal study was the elimination process by which out of 130 healthy jet pilots already selected for mental health, the seven original astronauts were selected. The process underscored both the importance and the commonsensical nature of mental health. The final seven astronauts enjoyed not only exemplary work records but also were competent at loving. All came from intact, happy, small-town families; they all were married with children. Although venturesome test pilots, they all had very few accidents during their years of flying or even earlier. They could tolerate both close interdependent association and extreme isolation. They trusted others and were uncomplaining under discomfort. They manifested a great capacity for withstanding frustration; nevertheless, emotions both negative and positive were strongly experienced. Not introspective, the astronauts seldom dwelt on their inner emotions; but they were sensitive to their emotions and could describe them when asked. They were aware of the feelings of others, and they avoided interpersonal difficulties. Unlike most people, their performance improved under stress (e.g., working at a simulated altitude of 65,000 feet in a poorly functioning pressure suit). Their group neuroticism (Maudsley Personality Inventory) score was the lowest of any group reported in the literature. Although each of the astronauts was very different, they all would have starred in a mental health "decathlon."

A still more influential study of mental health was the Menninger Psychotherapy Project. In order to cut through the hair-splitting of projective tests and the subjectivity of pencil and paper tests Menninger psychologist, Lester Luborsky devised a behavioral guide (health-sickness rating scale or HSRS) to assess psychological functioning on a scale of 0 to 100. In 1976 because Luborsky's scale had been designed to evaluate candidates for psychotherapy rather than for general epidemiological studies, two of the architects of DSM-III developed a revision of the HSRS called the global assessment scale (GAS). Rater reliabilities for both instruments and between the two instruments were about 0.85 to 0.95. In cross-cultural comparison, investigators noted that "the usefulness of HSRS as an international thermometer of mental health is strongly supported." A modified version of the GAS was introduced in DSM-III-R as the global assessment of functioning (GAF).

On Luborsky's scale, a score of 95 to 100 reflected "an ideal state of complete functioning integration, of resiliency in the face of stress, of happiness and social effectiveness." On the GAF, a score of 95 to 100 equaled "no symptoms, superior functioning in a wide range of activities; life's problems never seem to get out of hand; the patient is sought out by others because of his warmth and integrity." The words differ, but the melody is the same.

By 1978 *The Report to the President* by the President's Commission on Mental Health forcefully reiterated the importance of clearly defining what is meant by mental health. Nevertheless, it was not until 15 years later, when evidence emerged to support the validity of Axis V, DSM-IV did psychiatry finally possess a metric for the measurement of "above average" mental health. The World Health Organization Disability Assessment Schedule replaced it in the DSM-5, but the new tool is still a measure of mental illness rather than health: all 36 questions contain the words “difficulty,” “problem,” “drain,” or “health condition.”

In order to identify mental health empirically, behavioral data are needed, and a longitudinal biopsychosocial perspective must be maintained. Two very different life histories from Harvard’s Study of Adult Development illustrate what is meant by “above normal” mental health. One man, “Alfred Paine,” received an “average” score for mental health (GAF = 72); he never sought psychiatric care. In contrast, “Richard Luckey” received an above-normal score for mental health (GAF = 95). These two (disguised) case histories help to underscore that objective and subjective mental health are not merely value-ridden “ideal fictions.” Both men had been selected at 18 for mental health and were prospectively observed for a lifetime. Confounding variables like gender, education, social class, ethnicity, age, and birth cohort were held constant. Neither man had ever sought psychiatric help nor been diagnosed with mental illness, but there is no question who was the healthier.

Alfred Paine was a master of cheerful denial. On questionnaires, he described himself as a social drinker, in quite good physical health, and close to his children. It was only by interviewing him personally, talking with his wife, examining his medical record, reading the disappointing questionnaires from his children—and then, finally, by reading his obituary—that Alfred Paine's misery could be fully appreciated

The ancestors of Alfred Paine had been successful New England clipper ship captains. All his grandparents had graduated from high school. One grandfather became a merchant banker, and the other the president of a Stock Exchange. Paine's father had graduated from Harvard and his mother from a fashionable boarding school. When Alfred Paine was two weeks old, his mother died from the complications of childbirth. When he was 2 years old, his father died. Paine was bottle-fed by a variety of surrogates and raised by his grandmother. As a young boy, Paine was a head-banger; in adolescence, he was a "lone wolf."

In college Alfred Paine was often in love, but for Paine being in love meant having someone to care for him. His three marriages were all unhappy—in part because of the alcohol abuse that he maintained that he did not have, and in part because he was frightened of intimacy. At 50, Paine answered "true" to the statements "Sexually most people are animals" and "I would have preferred an asexual marriage."

At 47, Paine recalled the ages from 1 to 13 as the unhappiest in his life. At age 70, he believed that the ages from 20 to 30 were the unhappiest. But there had never been a time that Paine was happy. Nevertheless, on a pencil and paper test of depression, Paine achieved one of the best scores in the study. He had never sought psychotherapy, and none of his doctors ever called him mentally ill. His alcohol abuse never led to an arrest or to missed days of work. At 70, Paine wrote that his own physical health was "excellent," but in fact, he was seriously overweight, afflicted by hypertension and gout, and suffered from obstructive pulmonary disease—the result of life-long smoking. By 75, he had lost all his teeth; both his kidneys and his liver were failing; and he was cursed with mild dementia, the result of a drunken fall.

Although he had made a good living over the years at a job he disliked, by 75, Paine’s trust fund and his pension had been eroded through multiple divorces and self-inflicted tax troubles. His house looked as if furnished from yard sales, and little other than television now absorbed him. On his age 75 questionnaire, Paine refused to answer the part that dealt with life enjoyment. Thus, his subjective lack of joy in life had to be inferred from the fact that over the past 20 years, there was no area of his life other than his religious activities in which he had expressed satisfaction. On average, he saw his three children once every two years, and he grumbled that “They hardly let me see my grandchildren.” Although Paine’s third wife was protective and quite loving towards him, he was quite disrespectful and uncaring towards her. Asked how he and his wife collaborated, he replied, “We don’t. We lead parallel lives.” Since age 50, he had engaged in no pastimes with friends; and at 73, when asked to describe his oldest friend, he growled, “I don’t have any.” Often in his life, Paine had found love, but he could not let it inside.

In contrast, Richard Luckey was a well-loved child who took excellent care of himself. Unlike Alfred Paine, Luckey had come from more modest beginnings. None of his four grandparents had gone beyond grade school. His father graduated from high school and went on to become a successful businessman. After college Luckey became head of two successful businesses (one of which he created), both at the same time. He loved his work.

At 70, when looked at through the eyes of his internist, Richard Luckey's objective physical health had seemed actually worse than Alfred Paine's. Luckey had high blood pressure, atrial fibrillation, a cardiac pacemaker, and pancreatitis. He was even more overweight than Paine. But if Luckey was ill, he certainly did not feel sick. Over the next 10 years, Paine sickened unto death, and Luckey's health got steadily better. Thus, by 76, Lucky had completely recovered from his pancreatitis. He still wore a pacemaker; and, his blood pressure remained high, but in his doctor's words—not just his own—"Mr. Luckey continues to enjoy relatively good health … he continues to be active physically and also mentally." As Luckey, himself, expressed it, "I have done less chainsawing, but I still split wood." At age 76, Luckey spent two months downhill skiing in Austria.

Another crucial difference was that Luckey had never smoked, and he had used alcohol in moderation. "Almost everything we do," Luckey explained, "is family-oriented. His wife amplified, "We rarely go out, but we will have groups for supper such as the church fellowship group or the basketball team for a weekend of skiing or a vestry meeting at the house." Luckey was also commodore of his yacht club, and he had friends with whom he exercised regularly.

Richard Luckey described not only his hobbies, his religion, and his income-producing work as "very satisfying," but, more important, he experienced his relationships with his wife and with his children as "very satisfying." His wife and children's questionnaires revealed a similar satisfaction with him. Luckey's daughter had described her parent's marriage as "better than my friends"; then, for good measure, she had added two pluses. Luckey's wife gave her marriage a "nine" out of "nine," and on her questionnaire, she wrote the study, "My husband is my best friend; I like looking after him. We have grown closer and fonder every year." Asked how his marriage had lasted for 40 years, Luckey replied, "I really love Chrissie, and she loves me. I really respect her, highly esteem her, and she is a real person.” Luckey was close to his brother, with whom he regularly fished in the summer. He stayed in touch with and took great pleasure from his children and his grandchildren. It was so easy for Luckey to take in, to “metabolize,” any love that he was offered.

Asked about his retirement Richard Luckey replied, “I think it is all pretty nice. I don’t have a day when I don’t have something to do that I want to do … creativity is absolutely necessary for someone to be healthy. With painting,” he added dreamily, “you forget everything, and that is why it is so very relaxing.” In church, Luckey sang both solo and in the choir. Despite an indifferent college scholastic record, at age 75, Lucky was making a serious effort to master the technology of a research institute of which he had become a trustee.

In other words, health is based on an active, joyous, energetic engagement with the world. Such a naturalistic model of health is congruent with the increasing attention to “flow,” a concept recently elaborated on and empirically studied by psychologist Mihaly Csikszentmihalyi and his students. “Flow” involves focused attention and psychic absorption, which is characteristic of meditation, but unlike meditation, with “flow,” the “clutch” is engaged, and skilled behavior takes place. With “flow,” the participant feels alive and in the world. In “flow,” experiences with emotions are not just contained and channeled; they are energized and aligned with the consciousness of the task at hand. Action, cognition, and feeling are merged into one. Often, when manifested in intense experiences like advanced tennis, technical rock climbing, or violin playing, the “flow” experience has required hours of prior practice until much of the effort involved has become second nature.

Csikszentmihalyi’s concept of “flow” is distinct from Freud’s libido. “Flow” occurs when a task is challenging and requires both skill and concentration when there are clear goals and immediate involvement when time seems to stop and one’s sense of self vanishes, and when one finds themself both deeply involved—and in control. A species does not survive just by reproducing; it also has to produce and face new challenges.

MODEL B: MENTAL HEALTH AS POSITIVE PSYCHOLOGY

The fact that psychologists have approached mental health somewhat differently from psychiatry has led to the second model. Psychologists, like physiologists, look at continua (traits) rather than categories, while in medicine, you either have an illness, or you do not. In psychology, interventions to improve adequate intelligence or social skills are common, while in medicine, to meddle with adequate thyroid function, a healthy hematocrit, or a normal mood is only to cause trouble. In the healthy rested individual, virtually all psychopharmacological interventions will, over time, make the brain function worse; in contrast, many psychological interventions (e.g., literacy training, stress management) will make the brain function better. Thus, the medical goal of using medication to remove pathology is different from the psychologists' goal of fostering joy, enthusiasm, curiosity, and love for others in an educative model. As a result, physicians and psychologists draw attention to quite different—if complementary—models of mental health.

In the 20th century, medicine became increasingly concerned with concrete pathology, while psychology remained interested in education and in platonic and utopian virtue. The model of positive psychology conceives mental health as "the best possible" and has provided the basis for the positive psychology movement.

The last century witnessed not only psychologist John Dewey's idealism regarding education but also the work of Carl Jung and William James (physicians by training, but psychologists in spirit) concerning the search for and discovery of deeper meanings in life. In the last 70 years, Abraham Maslow's concept of self-actualization and his emphasis on humanistic psychology have drawn attention to the "full use and exploitation of talents, capacities, potentialities." But until the 21st century, such humanistic psychology has not undertaken empirical research and has ignored predictive validity and follow-up. Humanistic psychology has emphasized "happiness" and the "self" at the expense of collective well-being and has spawned myriad self-help movements that have benefited clients more through the placebo effect than from controlled and replicated therapeutic interventions. As early as 1925 psychiatrist, Adolf Meyer was already warning of the need to stop "moralizing" about utopian mental health. Mental health, he suggested, should be studied through "conscientious and impartial study" and "constructive experimentation."

Recently, two popular books by a distinguished psychologist, Martin Seligman, *Learned Optimism,* and *Authentic Happiness,* have served notice that positive psychology is beginning to follow Adolph Meyer's rules of conscientious and impartial study and constructive experimentation. Seligman's concepts of "learned optimism" and positive psychology incorporate the empirical advances in cognitive psychology that have taken place over the past 3 decades. Creating a positive attributional style not only serves as a cognitive-behavioral treatment for depression but can lead to positive mental states.

Positive psychology wishes to learn how to build the qualities that help individuals and communities, not just to endure and survive but also to flourish. Formally introduced in the January 2000 issue of *American Psychologist,* positive psychology hopes to render the psychology of strength and well-being amenable to scientific study and intervention. In that issue Martin Seligman and Mihalyi Csikszentmihalyi wrote:

> "At the individual level, it is about positive individual traits; the capacity for love and vocation, courage, interpersonal skill, aesthetic sensibility, perseverance, forgiveness, originality, future mindedness, spirituality, high talent, and wisdom…Psychology is not just a branch of medicine concerned with illness or health; it is much larger. It is about work, education, insight, love, growth, and play. And in this quest for what is best, positive psychology does not rely on wishful thinking, faith, self-deception, fads, or hand waving; it tries to adapt what is best in the scientific method to the unique problems that human behavior presents to those who wish to understand it in all its complexity." (Copyright © 2000, American Psychological Association. Reproduced with permission. Seligman ME, Csikszentmihalyi M. Positive psychology. An introduction. *Am Psychol*. 2000;55(1):5–14. doi: 10.1037//0003-066x.55.1.5.)

Over the past 40 years, cognitive therapists have demonstrated that altered cognition cannot only change behavior it can also alter brain function. If pessimism is the dominant cognition of the depressed, optimism appears as the dominant cognition of the mentally healthy. If learned helplessness leads to depression, learned optimism and self-efficacy lead to mental health. Even if the depressed often appears to view reality more realistically, such accurate reality testing avails depressed individuals of nothing.

In part, the importance of optimism to positive mental health depends upon an attributional cognitive style that asserts the good things that happen to someone will last forever and are pervasive. The bad things that happen to someone are limited and unlikely to happen again. In addition, optimism includes hope, a facet of mental health as old as the Greek myth of Pandora, yet hope is a topic to which psychiatry up to now has given little formal consideration. In contrast, the explanatory style of many chronically depressed individuals is that bad events, for which one may be responsible, will last forever and will generalize, while he or she bears no responsibility for good events, which not only occur only by chance but are limited and fleeting. How much better to take responsibility for the good events and conceive of them as lasting and pervasive, while the bad events are bad luck, isolated instances, and unlikely to happen again?

Longitudinal studies reveal that an optimistic attributional style improves physical health and wards off depression rather than merely “prolonging human suffering,” as Nietzsche feared. In addition, the so-called illusion of optimism permits one to contemplate and plan for rather than deny the future. As longitudinal studies have repeatedly shown, future-mindedness is a critical ingredient of mental health. In addition, in major depression, the hopeful illusion of the placebo effect accounts for roughly 50% of the variance between the effect size produced by selective serotonin reuptake inhibitors (SSRIs) and no treatment. Thus, experimentally and longitudinally, optimism has been linked to positive mood and good morale, perseverance and effective problem-solving, to longevity, and to succeed in a wide spectrum of activities.

A critical distinction, of course, is the degree to which optimism distorts and to which pessimism accurately perceives reality. The pessimism of psychotic depression, like the psychotic optimism of acute mania, involves reality-defying delusion. But the comforting illusions of the mentally healthy do not interfere with learning that stoves are hot, while the pessimistic perceptions of the dysthymic, while often accurate, can still inhibit future efforts to cook. In the prevention of future mental illness, one of psychiatry’s most powerful tools may be the inculcation in the young of optimism, hope, and future-mindedness. But to be effective, the optimistic must pull into consciousness emotional awareness of future pain and discomfort (see *anticipation* in Model D).

Recently, advocates of positive psychology have divided positive mental health into four components: *Talents, Enablers, Strengths,* and *Outcomes. Talents* are inborn, genetic, and are not much affected by interventions (e.g., high IQ or being an easy baby). *Enablers* reflect benign social conditions, interventions, and environmental good luck (e.g., a strong family, a good school system, living in a democratic meritocracy); enablers can be experimentally modified to enhance strengths. *Strengths* (see Table 3.5–1) are character traits like curiosity and openness that reflect facets of mental health that are amenable to change. The 24 *Strengths* are currently an essential component of modern positive psychology and can be self-assessed via the Internet. *Outcomes* reflect dependent variables (improved GAF, social relationships, SWB) that can be used to provide evidence for the predictive validity that efforts by clinicians to enhance *Strengths* are not just wishful thinking.

Admittedly, which so-called *strengths* are most associated with mental health is open to debate. Wisdom, kindness, and the capacity to love and be loved are strengths with which few would argue. But should courage be included as a strength, and why were intelligence, perfect musical pitch, and punctuality excluded? The answer is that the 24 strengths listed in Table 3.5–1 may be subjected to a variety of tests. First, they must have been recurrent positive values across cultures and across centuries—why punctuality was excluded? Second, they must be valued in their own right and not just as a means to ends. Third, like intelligence and perfect musical pitch, they are not inborn talents.

There are several pitfalls to positive psychology. First, the perspective of mental health as a Utopia is one of the bogeymen of national health policymakers. They are afraid, without directly expressing it, that this perspective will put a backbreaking burden on health insurance. Thus, there is considerable debate within the mental health professions about whether helping people achieve greater SWB is a process that any insurance program should reasonably be expected to cover. Helping people achieve their potential for community participation is perhaps a more laudable goal. But that, too, is highly value-laden and culture-bound. Over time society will have to decide who should pay for positive mental health: the individual, the educational system, third-party payers, religious organizations, or a combination of all four.

A second caution to positive psychology is the danger of the culturally insensitive prescription of parochial virtues. The dangers of value judgments are enormous. “Virtues,” even Aristotelian virtues, need to be distinguished from health. Keeping wounds clean is healthy but not a virtue. Body hygiene in public places is a virtue but not necessarily healthy.

A third controversial facet of positive psychology is its emphasis on optimism. Since the late 19th century, many social scientists, especially those in Europe, have mistrusted optimistic cognition, especially religious optimism, as a maladaptive "American" illusion interfering with an accurate perception of reality. Nietzsche, Freud, Marx, and Darwin all perceived optimism as evidence of ingenuous cultural adolescence, not of mature mental health. Opponents of optimism believed that it was healthier to face the hard facts of life. In addition, for most of the 20th century, investigators like Pavlov and Skinner have viewed the study of cognitions of all kinds, of which optimism is one, as inferior to the study of behavior. Only in the last 20 years has psychiatry begun to see the value of positive psychology.

MODEL C: MENTAL HEALTH AS MATURITY

Unlike other organs of the body that are designed to mature in the first 2 decades and then remain functionally static but gradually less robust until death, the brain is different. The brain can develop new functions until the 8th decade and develop anatomically (more efficiently myelinated) until the 6th. Therefore, just as optimal brain development requires almost a lifetime, so does the assessment of positive mental health. A 10-year-old's lungs and kidneys are more likely to reflect optimal function than are those of a 60-year-old, but that is not true of their central nervous systems. To some extent, then, adult mental health reflects a continuing process of maturational unfolding. Statistically, physically healthy 70-year-olds are mentally healthier than they were at 30; and prospective studies reveal that individuals are less depressed and show greater emotional modulation at age 70 than they did at age 30.

But if prospective studies of adult development reveal that the immature brain functions less well than the mature, does that mean adolescents are mentally healthier than toddlers? And are the middle-aged mentally healthier than adolescents? The answer is both yes and no, but the question illustrates why to understand mental health, one must understand what is meant by maturity.

In some respects, in 1950, Erik Erikson anticipated Jahoda and Grinker when he provided the first model of adult social development. He viewed each of his well-known eight "stages" of human development as a "criterion of mental health." Subsequently, Jane Loevinger provided a model of adult ego development. Lawrence Kohlberg has provided a model of adult moral development, and most recently, James Fowler has provided a model of spiritual development. Implicit in all these models is the assumption that greater maturity reflected better mental health. Arguably, the best definition of mental health that is available is William Menninger's definition of maturity depicted in Table 3.5–1. In his model, maturity is not only the antonym of narcissism, but it is quite congruent with other models of mental health.

To confirm the hypothesis that maturity and positive mental health are almost synonymous, the study of the behavior and feeling states of persons studied over a lifetime becomes necessary. Although such longitudinal studies have come to fruition only recently, all illustrate the association of maturity with increasing mental health. After 50, of course, the association between mental health and maturity is contingent upon a healthy central nervous system. The ravages of brain trauma, major depression, arteriosclerosis, Alzheimer, alcoholism, and schizophrenia must all be avoided. But if so, with good physical health, intellectual functions can remain virtually unimpaired until eighty. Common folklore bemoans that millions of neurons are lost with each decade of adulthood until the value of selective pruning is recognized. Remember, 5-year-olds have twice as many neurons as adults do at 20.

The association of mental health with maturity is mediated not only by progressive brain myelinization into the 6th decade but also by the evolution of emotional and social intelligence through experience. Erikson conceptualized that such development produced a "widening social radius." In such a view, life after age 50 was no longer to be a staircase leading downwards, as in the Pennsylvania Dutch cartoons of lifespan development, but a path leading outwards. In Erikson's model, the adult social radius expanded over time through the mastery of the four tasks of "identity vs. identity diffusion," "intimacy vs. isolation," "generativity vs. stagnation," and "integrity vs. despair." On the basis of empirical data from Harvard's Study of Adult Development, Vaillant has added two more tasks —career consolidation and keeper of the meaning or guardian—to Erikson's four (see Fig. 3.5–1). The mastery of such tasks appears relatively independent of education, gender, social class, and probably culture.

In such a model, the social radius of each earlier adult developmental task fits inside the next. First, adolescents must achieve an *Identity* that allows them to become separate from their parents, for mental health and adult development cannot evolve through a false self. The task of *Identity* requires mastering the last task of childhood: sustained separation from social, residential, economic, and ideological dependence upon the family of origin. Identity is not just a product of egocentricity, of running away from home, or of marrying to get out of a dysfunctional family. There is a world of difference between the instrumental act of running away from home and the developmental task of knowing where one's family values end and one's own values begin. Such separation derives as much from the identification and internalization of important adolescent friends and nonfamily mentors as it does from simple biologic maturation. For example, accents become relatively fixed by age 16 and reflect those of one's adolescent peer group rather than the accents of one's parents.

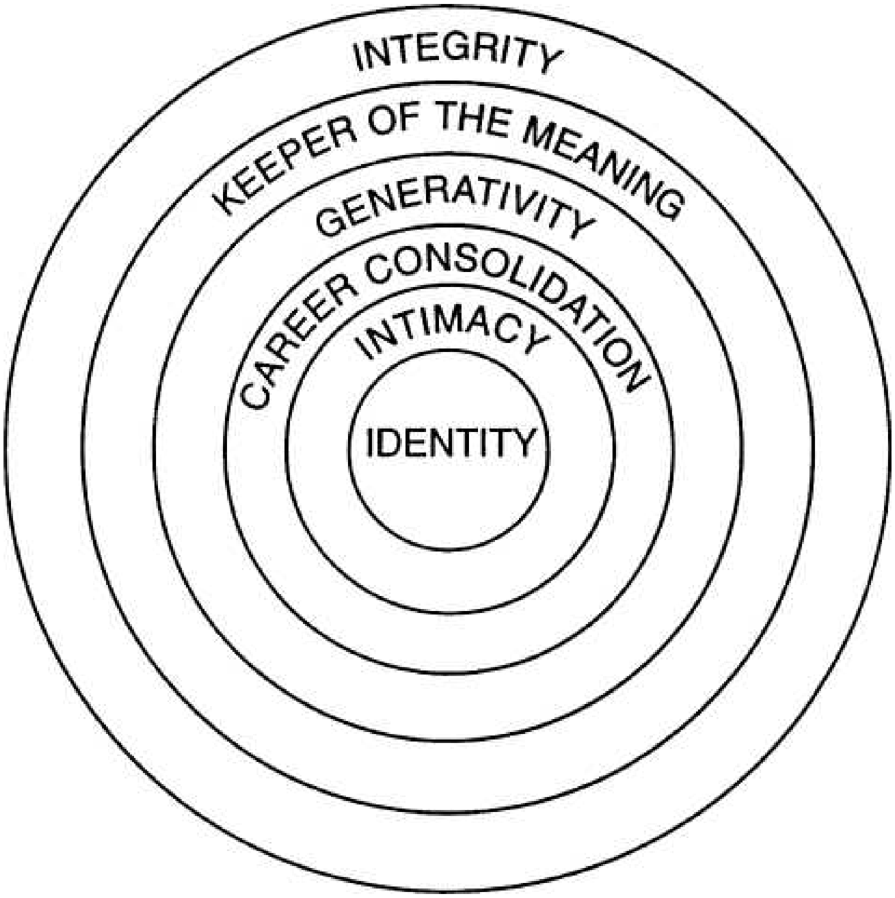

FIGURE 3.5–1. A schematic model of the expanding social radius of maturing individuals during adulthood.

Then, young adults should develop *Intimacy,* which permits them to become reciprocally, and not selfishly, involved with a partner. But living with just one other person in an interdependent, reciprocal, committed, and contented fashion for years and years may seem neither desirable nor possible to a young adult. Once achieved, however, the capacity for intimacy may seem as effortless and desirable as riding a bicycle. Sometimes the relationship is with a person of the same gender; sometimes, it is completely asexual; and sometimes, as in religious orders, the interdependence is with a community. Superficially, mastery of intimacy may take very different guises in different cultures and epochs, but "mating-for-life" and "marriage-type love" are developmental tasks built into the developmental repertoires of many warm-blooded species, including the human species. Nevertheless, recent lifetime studies of marriage reveal "mating" for life is not always the rule—or even desirable. Personalities often change over a lifetime.

Career Consolidation is a task that is usually mastered together with or that follows the mastery of intimacy. Mastery of this task permits adults to find a career as valuable as they once found play. On a desert island, one can have a hobby but not a career, for careers involve being of value to other people. There are four crucial developmental criteria that transform a "job" or hobby into a "career": contentment, compensation, competence, and commitment. Obviously, such a career can be "wife and mother"—or in more recent times, "husband and father." To the outsider, the process of career consolidation often appears "selfish," but without such "selfishness," one becomes "selfless" and has no "self" to give away in the next stage of Generativity. Not only schizophrenics but also individuals with a severe personality disorders often manifest a lifelong inability to achieve either intimacy or sustained, gratifying employment. In short, they fail to grow up.

Mastery of the fourth task, *Generativity,* involves the demonstration of a clear capacity to care for and guide the next generation. Existing research reveals that sometime between the ages 35 and 55, the need for achievement declines, and the need for community and affiliation increases. Depending on the opportunities that society makes available, generativity can mean serving as a consultant, guide, mentor, or coach to young adults in the larger society. Like leadership, generativity means being in a caring relationship in which one gives up much of the control that parents retain over young children. Good mentors learn "to hold loosely" and to share responsibility. Generativity reflects the capacity to give the self—finally completed through mastery of the first three tasks of adult development—away. Its mastery is strongly correlated with successful adaptation to old age. This is because in old age, there are inevitable losses, and these may be overwhelming if one has not continued to grow beyond his or her immediate family.

The penultimate life task is to become a *Guardian* (in anthropological terms, Keeper of the Meaning). Like grandparenthood, this task involves passing on the traditions of the past to the future. Generativity and its virtue, *care,* require taking care of one person rather than another. While keeper of the meaning and its virtues of *wisdom* and *justice* are less selective, justice, unlike care, means not taking sides. Indeed, mastery of this fifth task is epitomized by the role of the wise judge. The focus of a *Guardian* is on the conservation and preservation of the collective products of mankind—the culture in which one lives and its institutions—rather than just the development of its children.

Just as "selfless" coaches play a different role from "selfish" athletes; just so the organizers, judges, and guardians of the Olympic Games play a very different role from the Games' more generative but still parochial coaches. Clearly, caretakers and grandparents are not mentally healthier than caregivers and parents. The only distinction is that grandparents are usually better at the tasks of Keepers of the Meaning than are 30-year-olds.

Finally, in old age, it is common to feel that some life exists after death and that one is part of something greater than one's self. Thus, the last life task in Erikson's words is *Integrity,* the task of achieving some sense of peace and unity with respect both to one's own life and to the whole world. Erikson described integrity as "an experience which conveys some world order and spiritual sense. No matter how dearly paid for, it is the acceptance of one's one and only life cycle as something that had to be and that, by necessity, permitted of no substitutions." One of the study subjects, age 88, described integrity in more mundane terms "I need to go to the Department of Motor Vehicles to discover if the world would be safer if I no longer drove a car." How different the concerns of the elderly adult are from the adolescent seeking identity.

Admittedly, healthy adult development does not follow rigid rules. Some individuals, often due to great stress, tackle developmental tasks out of order or all at once. On the one hand, Alexander the Great, General Lafayette, Napoleon Bonaparte, and Joan of Arc were inspirational and generative leaders in their 20s. On the other hand, William Osler, one of the most generative physicians of all time, did not confront the task of intimacy until he married at age 40—well after he had established his career as a brilliant professor of medicine. Ludwig van Beethoven enjoyed a brilliant, committed career but never enjoyed intimacy.

Finally, it must be kept in mind that mastery of one life task is not necessarily healthier than mastery of another, for adult development is neither a foot race nor a moral imperative. Rather, these six sequential tasks above are offered as a road map to help psychiatrists make sense of where they are and where their patients might be located. One can be a happily married 20-year-old; that is usually healthy. One can be a chronically unemployed single 50-year-old; that is usually unhealthy. Acquiring a social radius that extends beyond oneself by definition allows more flexibility and SWB and thus is always healthier than self-preoccupation.

MODEL D: MENTAL HEALTH AS RESILIENCE

In 1856, Claude Bernard, a French physiologist and a founder of experimental medicine, started the current understanding of positive health when he wrote, “We shall never have a science of medicine as long as we separate the explanation of the pathological from the explanation of normal, vital phenomena.” For example, coughing and pus in response to infection seemed pathologic but, in fact, were often healing. In other words, it is not stress that kills people, but healthy mastery of stress that permits people to survive.

In 1925, Adolf Meyer, a founder of modern American psychiatry, introduced the current understanding of mental health when he asserted that there were no mental diseases; there were only characteristic reaction patterns to stress. Meyer’s point was that although adaptive mental “reaction patterns” like denial, phobias, and even projections can appear to reflect illness, they may, in fact, be “normal, vital phenomena” related to healing. Just as immune mechanisms, clotting mechanisms, and callous formation heal by distorting bodily equilibrium; just so equally involuntary coping mechanisms heal by distorting mental processes.

Of course, the symptoms of organic brain damage usually reflect disease and not adaptation. Manic-depressive psychosis is due to genetic defects; the mental devastation produced by alcoholism is due to poisoning, and the negative symptoms of schizophrenia reflect brain defects and not adaptation. Nevertheless, much so-called mental illness is more like the healthy but red and tender swelling that immobilizes a fracture so that it may heal. Much of what is called mental illness in the current diagnostic nomenclature—the symptoms of many anxiety disorders, some depressions, and most personality disorders—are the outward manifestations of homeostatic struggles to adapt to life. Admittedly, analogous to acne and autoimmune disease, such reaction patterns often reflect rigid, poorly modulated, and pathologic efforts at adaptation.

To set the stage, there are three broad classes of coping mechanisms that humans use to overcome stressful situations. First, there are the ways in which an individual elicits help from appropriate others: namely *consciously seeking social support.* Second, there are *conscious cognitive strategies* that are intentionally used to master stress. Third, there are *adaptive involuntary coping mechanisms* (often called "defense mechanisms") that distort the perception of internal and external reality in order to reduce subjective distress, anxiety, and depression.

By 1970, the term *defense mechanisms,* like many psychoanalytic metaphors, had been largely discarded by empirical social scientists. Consistency of definition and rater reliability was lacking. But over the last 40 years, the idea of involuntary coping has entered the literature of empirical cognitive psychology under such rubrics as "hardiness," "grit," "self-deception," "emotional coping," and "illusion." In recent years experimental strategies for assessing such mechanisms, especially with videotape and the Q-sort, have also improved reliability. Several empirical studies, well-reviewed by Skodol and Perry, have clarified the current understanding of healthy and unhealthy involuntary adaptive mechanisms (a.k.a. defenses). Nevertheless, no one has yet developed a method for assessing defenses, which meets conventional standards for psychometric reliability. As a result, the axis of adaptive involuntary coping mechanisms has been excluded from DSM-5.

This third class of *involuntary adaptive coping mechanisms* reduces conflict and cognitive dissonance during sudden *changes* in internal and external reality. If such changes in reality (e.g., the unanticipated crisis of 9/11) are not "distorted" and "denied," they can result in temporarily disabling anxiety and/or depression. In summary, such homeostatic mental "defenses" shield people from sudden changes in the four lodestars of conflict: affect, reality, relationships, and conscience (see Fig. 3.5–2).

First, such involuntary mental mechanisms can restore psychological homeostasis by ignoring or deflecting sudden increases in the lodestar of impulse, affect, and emotion. (The emotions of terror, grief, longing, and dependency loom just as important as sources of conflict as do lust, fear, and rage.) Psychoanalysts call this lodestar “id,” fundamentalists call it “sin,” cognitive psychologists call it “hot cognition,” and neuroanatomists point to the hypothalamic and limbic regions of the brain.

Second, such involuntary mental mechanisms can provide a mental time out to adjust to sudden changes in *reality* and self-image, which cannot be immediately integrated. Individuals who initially responded to the television images of the sudden destruction of New York’s World Trade Center as if it were a movie provide a vivid example of denial of an external reality that was changing too fast for voluntary adaptation. Sudden good news—the instant transition from a student to a physician or winning the lottery—can evoke involuntary mental mechanisms as often as can an unexpected accident or diagnosis of leukemia.

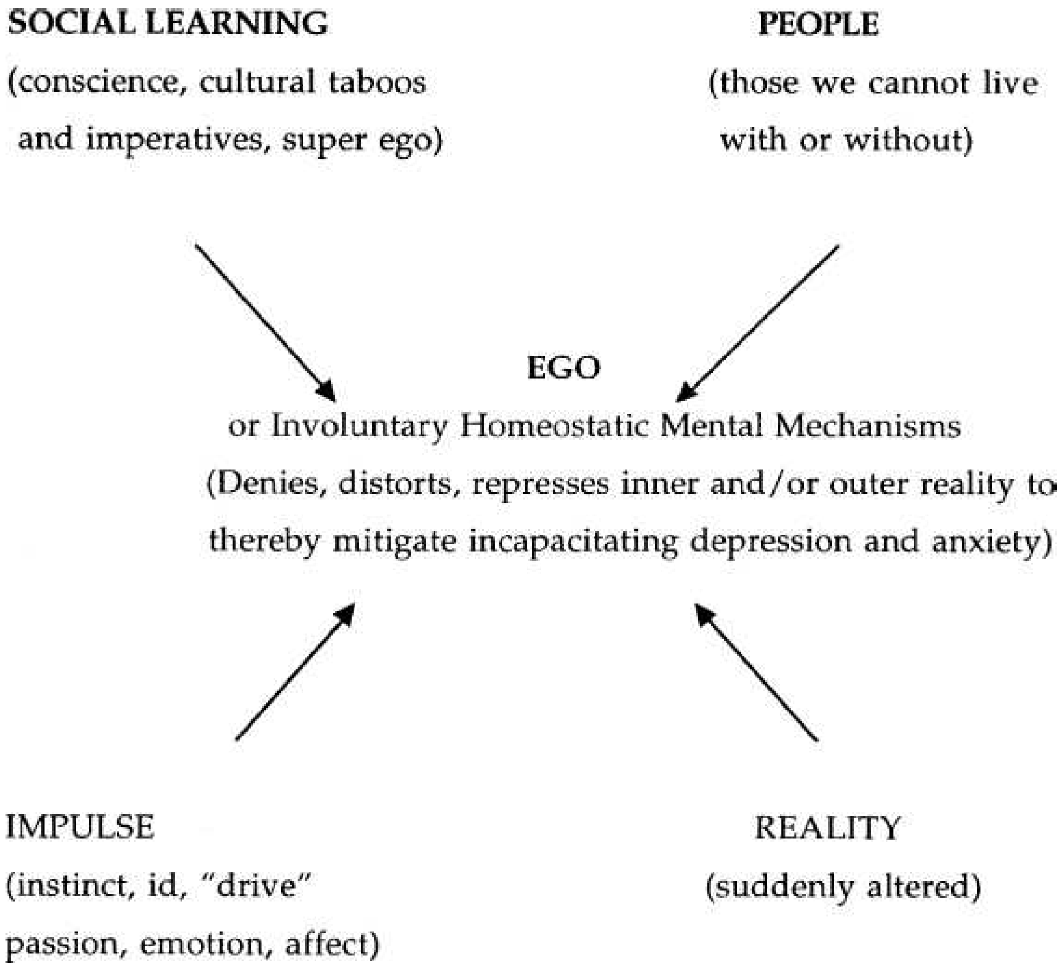
SOCIAL LEARNING
(conscience, cultural taboos and imperatives, super ego)
PEOPLE
(those we cannot live with or without)
EGO
or Involuntary Homeostatic Mental Mechanisms
(Denies, distorts, represses inner and/or outer reality to thereby mitigate incapacitating depression and anxiety)
IMPULSE
(instinct, id, "drive" passion, emotion, affect)
REALITY
(suddenly altered)

FIGURE 3.5–2. The four lodestars of human conflict.

Third, involuntary mental mechanisms can mitigate sudden irresolvable conflict with important *people,* living or dead. People become a lodestar of conflict when one cannot live with them and yet cannot live without them. Death is such an example; another is an unexpected proposal of marriage. And since people exist within one's self as well as without, internal representations of important people may continue to haunt someone and cause conflict for decades after they have ceased to live with them and thus evoke an involuntary mental response.

Finally, the fourth source of conflict or anxious depression is social learning or conscience. Psychoanalysts call it "superego," anthropologists call it "taboos," behaviorists call it "conditioning," and neuroanatomists point to the associative cortex and the amygdala. This lodestar is not just the result of admonitions from one's parents that are absorbed before age 5 but is formed by one's whole identification, with one's culture, and sometimes by irreversible learning resulting from overwhelming trauma.

There are many ways that one can ignore or distort one or more of these four lodestars in order to mitigate intrapsychic conflict. These coping mechanisms or “defenses” can abolish impulse (e.g., by reaction formation), or social learning (e.g., by acting out), or other people (e.g., by schizoid fantasy), or reality (e.g., by psychotic denial). They can abolish one’s conscious recognition of the subject (e.g., by projection), or the object of a conflict (e.g., by turning against the self), or awareness of the idea (e.g., by repression), or of the affect associated with the idea (e.g., isolation of affect).

All classes of adaptive involuntary coping mechanisms are effective in “denying” or defusing conflict and in “repressing” or minimizing stress, but they differ greatly in the psychiatric diagnoses assigned to their users and in their consequences for long-term biopsychosocial adaptation. In the most pathological category, are found distortion and denial of external reality. These mechanisms are common in young children, in one’s dreams, and in psychosis. To breach them requires altering the brain by neuroleptics or waking the dreamer.

More common to everyday life are the relatively maladaptive, “narcissistic,” or “immature” defenses. They include projection, passive aggression, schizoid fantasy, hypochondriasis, dissociation, and “acting out.” Defenses in these categories are common in adolescents, immature adults, and individuals with personality disorders. They often make others more uncomfortable than the user. Such defenses are consistently and negatively correlated with a global assessment of mental health, and they profoundly distort the affective component of interpersonal relationships.

Less maladaptive mechanisms include repression, intellectualization, reaction formation (i.e., turning the other cheek), and displacement (i.e., directing affect at a more neutral object). They are often associated with anxiety disorders and with the psychopathology of everyday life. In contrast to the “immature” defenses, these intermediate defenses are manifested clinically by phobias, compulsions, obsessions, somatizations, and amnesias. Such users often seek psychological help and such “compromise formations” respond more readily to interpretation. Such defenses usually cause more conscious suffering to the user than to other people. They are common to everyone from 5 years old until death. They are neither healthy nor unhealthy.

The “mature” still distort and alter feelings, conscience, relationships, and reality, but they achieve these alterations gracefully and flexibly.

These mechanisms allow the individual consciously to experience the affective component of interpersonal relationships but in a tempered fashion. Thus, the beholder regards the involuntary mature defenses as virtues, just as the beholder may regard the prejudice of projection and the tantrums of acting out as sins.

Doing as one would be done by (altruism), keeping a stiff upper lip or deploying Grit (suppression), keeping future pain in awareness (anticipation), the ability not to take one's self too seriously (humor), and turning lemons into lemonade (sublimation) are the very stuff from which positive mental health is made. Unfortunately, like tightrope walking, without months of practice, mature mechanisms cannot easily be deployed voluntarily and only then by those with innate balance.

Identification of defenses is difficult. Rarely can people identify their own defenses, and people often fail to recognize them in others or, still worse, "project" their own defenses onto them. To identify a defense, an objective observer needs to triangulate past truth with present behavior and subjective report. For example, one woman wins community praise for founding a shelter for battered women (altruism), but she dismisses her behavior as "a need to rent my house." Another woman is imprisoned for breaking her toddler daughter's arm in a tantrum (acting out), but she dismisses her behavior as an accident. In fact, social agency records from 30 years before reveal that both women had been taken at age 2 from the care of physically abusive, alcoholic mothers. Whether one ultimately views another's coping response as healthy or psychopathic depends on the results of their involuntary efforts. Ultimately, like other facets of mental health, the reliable identification of healthy but involuntary coping mechanisms requires longitudinal study.

Healthy Involuntary Mental Mechanisms

Longitudinal studies from both Berkeley's Institute of Human Development and Harvard's Study of Adult Development have illustrated the importance of mature defenses to mental health.

Humor. Everyone recognizes that humor makes life easier. As Freud suggested, "Humor can be regarded as the highest of these defensive processes," for humor "scorns to withdraw the ideational content bearing the distressing affect from conscious attention, as repression does, and thus surmounts the automatism of defense." With humor, one sees all, one feels much, but one does not act. Humor permits the discharge of emotion without individual discomfort and without unpleasant effects upon others. Mature humor allows one to look directly at what is painful, whereas dissociation and slapstick distract one to look somewhere else. Yet, like the other mature defenses, humor requires the same delicacy as building a house of cards—timing is everything.

Altruism. When used to master conflict, altruism involves getting pleasure from giving to others what that person themselves would have liked to receive. For example, using reaction formation, a former alcohol abuser works to ban the sale of alcohol in his town and annoys his social drinking friends. Using altruism, the same former alcoholic serves as an Alcoholics Anonymous sponsor to a new member—achieving a transformative process that may be lifesaving to both the giver and receiver. Obviously, many acts of altruism involve a free will, but others involuntarily soothe unmet needs.

Sublimation. The sign of a successful sublimation is neither careful cost accounting nor shrewd compromise but rather psychic alchemy. By analogy, sublimation permits the oyster to transform an irritating grain of sand into a pearl. In writing his Ninth Symphony, the deaf, angry, and lonely Beethoven transformed his pain into triumph by putting Schiller's "Ode to Joy" to music. But like humor, sublimation does not usually occur on purpose.

Suppression. Suppression is a defense that modulates emotional conflict or internal/external stressors through stoicism. Suppression minimizes and postpones but does not ignore gratification. Empirically, this is the defense most highly associated with other facets of mental health. Used effectively, suppression is analogous to a well-trimmed sail; every restriction is precisely calculated to exploit, not hide, the winds of passion. Evidence that suppression is not simply a conscious "cognitive strategy" is provided by the fact that jails would empty if delinquents could learn to just say "No." Over the last decade, Angela Duckworth has carried out cutting-edge empirical research on suppression, which she relabeled as "Grit."

Anticipation. If suppression reflects the capacity to keep the current impulse in mind and control it, anticipation is the capacity to keep an affective response to an unbearable future event in mind in manageable doses. The defense of anticipation reflects the capacity to perceive future danger affectively as well as cognitively and by this means to master conflict in small steps. Examples would be the fact that moderate amounts of anxiety before surgery promote postsurgical adaptation and that anticipatory mourning facilitates the adaptation of parents of children with leukemia.

As with the use of altruism and suppression, the use of anticipation can often be voluntary and independent of conflict. But it is in cases of "hot cognition" that these defenses can become involuntary and lifesaving.

Just as psychiatry needs to understand how a GAF of 75 might become 90, just so, psychiatry needs to understand how best to facilitate the transmutation of less adaptive defenses into more adaptive defenses. One suggestion has been first to increase social supports and interpersonal safety and second to facilitate the intactness of the central nervous system (e.g., rest, nutrition, and sobriety). The newer forms of integrative psychotherapies using video recording can also catalyze such change by allowing patients to see their involuntary coping style.

MODEL E: MENTAL HEALTH AS SOCIOEMOTIONAL INTELLIGENCE

High socioemotional intelligence reflects above-average mental health in the same way that a high IQ reflects above-average intellectual aptitude. Such emotional intelligence lies at the heart of positive mental health. In the *Nicomachean Ethics,* Aristotle defined socioemotional intelligence as follows: "Anyone can become angry—that is easy. But to be angry with the right person, to the right degree, at the right time, for the right purpose, and in the right way—that is not easy." Nevertheless, as recently as 40 years ago, a textbook on intelligence dismissed the concept of such Aristotelian social intelligence as "useless."

All emotions exist to assist basic survival. While the exact number of primary emotions is arguable, seven emotions are currently distinguished by characteristic facial expressions connoting anger, fear, excitement, interest, surprise, disgust, and sadness. The capacity to identify these different emotions in one's self and in others plays an important role in mental health. The benefits of being able to read feelings from nonverbal cues have been demonstrated in almost a score of countries. These benefits included being better emotionally adjusted, more popular, and more responsive to others. Empathic children do better in school and are more popular than their peers even if they are no more intelligent than the other children. Head Start found that early school success was achieved not by intelligence but by knowing what kind of behavior is expected, how to rein in the impulse to misbehave, being able to wait, and how to get on with other children. At the same time, the child must be able to communicate their own needs and turn to teachers for help.

Ethologically, emotions are critical to mammalian communication. Since such communications are not always consciously recognized, the more skilled one is in identifying his or her own emotions, the more skilled he or she will be in communicating with others and empathically recognizing their emotions. Put differently, the more one is skilled in empathy, the more one will be valued by others, and so the greater will be their social supports, self-esteem, and intimate relationships.

Social and emotional intelligence can be defined by the following criteria:

- Accurate conscious perception and monitoring of one’s own emotions.
- Modification of one’s emotions so that their expression is appropriate. This involves the capacity to self-soothe one’s own anxiety and to shake off hopelessness and gloom.
- Accurate recognition of and response to emotions in others.
- Skill in negotiating close relationships with others.
- Capacity for focusing emotions (motivation) toward the desired goal. This involves delayed gratification and adaptively displacing and channeling impulses.

Some behavioral scientists divide emotions up into positive and negative as if negative emotions were unhealthy. This is probably an error. As with pus, fever, and cough, the negative emotions of sadness, fear, and anger are also important to healthy self-preservation. Positive emotions like joy, love, hope, gratitude, and compassion are associated with subjective contentment. While negative emotions interfere with contentment, their expression can be equally healthy.

Over the last 15 years, three important empirical steps have been taken in the current understanding of the relationship of socioemotional intelligence to positive mental health. The first step is that both functional MRI (fMRI) and ingenious neurophysiological experimentation have led to advances in the current understanding of the integration of the prefrontal cortex with the limbic system, especially with the amygdala and its connections. These research advances have brought psychiatrists closer to understanding emotions as neurophysiological phenomena rather than as platonic abstractions. We have learned that the prefrontal cortex is the region of the brain responsible for working memory and that the frontal lobes, through their connections to the amygdala, hippocampus, and insula, encode emotional learning in a manner quite distinct from both conventional conditioning and declarative memory.

The second step forward has been the slow but steady progress in the conceptualizing and even the measuring of “emotional intelligence.” Over the last decade, measures of emotional intelligence are evolving rapidly.

The third advance is the use of video recording to chart emotional interaction. Videos of sustained family interactions reveal that the most important aspect of healthy infant development, adolescent development, and marital harmony is how partners or parents respond to emotions in others. To ignore, to punish, and to be intimidated or contemptuous of how another feels spells disaster. Children of emotionally attuned parents are better at handling their own emotions and are more effective at soothing themselves when upset. Such children even manifest lower levels of stress hormones and other physiological indicators of emotional arousal.

Where the passive study of positive mental health ends and active primary prevention begins is unclear, but, like the model of positive psychology, the model of socioemotional intelligence is potentially interventionist. Just as someone can have an above-average musical skill or physical coordination, which they can improve with training, someone can learn to enhance emotional modulation. Until late in the Cold War, empirically researched techniques for “getting to yes” in negotiations were not in any academic curricula.

There are now many exercises in handling relationships that help couples, business executives, and diplomats become more skilled at conflict resolution and negotiation. In the past decade, there has also been an increasing effort to teach schoolchildren core emotional and social competencies, sometimes called “emotional literacy.” The relevance of these advances in psychology to psychiatry includes teaching emotion recognition and differentiation in eating disorders, teaching anger modulation, and finding creative solutions to social predicaments for behavior disorders.

MODEL F: MENTAL HEALTH AS SUBJECTIVE WELL-BEING

In this chapter, “happiness” will be construed to refer to *SWB*—a relatively neutral and, more importantly, a long-lasting state of affairs. The term “happiness,” however, has multiple meanings. For example, happiness can also be construed to mean nothing more than selfish pleasure. This interpretation has given Positive Psychology the disparaging reputation of “happyology.” To underscore this aspect, the author prefers to define “happiness” as selfish *drive reduction* and “joy” to refer to the equally satisfying, less autistic, *and longer-lasting interpersonal connection.* The search for happiness can appear selfish, narcissistic, superficial, and banal. Happiness is often based on illusion or on dissociative states. Pleasures can come easily and be soon gone. Illusory happiness is seen in the character structure associated with bipolar and dissociative disorders, with the maladaptive denial of Pollyanna and Voltaire’s Dr. Pangloss and with the exaltation of “me” advocated by much pop psychology. Examples of maladaptive “happiness” can be the excitement of risk-taking, from being “high” on drugs, and from “turning on” to any unmodulated but gratifying primitive need like binge eating, tantrums, promiscuity, and revenge. Such maladaptive happiness can bring temporary bliss but has no sticking power. It is because of such ambiguity of meaning that, throughout this section, the term *SWB* will be substituted for happiness.

National well-being has been used more appropriately than “national happiness” by political scientists to describe the “health” of nations. The demographic variables assessed include women’s rights, lack of corruption, percent live births, and freedom of the press.

The importance of the term *SWB* addresses the question is it better to meet some expert’s definition of mental health, or is it better to feel subjectively fulfilled? The answer is “both.” For positive mental health does not just involve being a joy to others; one must also experience SWB. Long before humankind considered definitions of mental health, they pondered criteria for subjective happiness. For example, objective social support accomplishes little if subjectively the individual cannot feel loved. Thus, the capacity for SWB becomes an important model of mental health. Put differently, happiness and SWB are always congruent if they involve positive emotions. Negative emotions like Greed, Lust, Hate, and

Vengeance are all about *me* and lead only to fleeting satisfactions. The eight great positive emotions: Gratitude, Love, Joy, Compassion, Forgiveness, Trust (Faith), Hope, and Awe are all about the other and can lead to lasting well-being.

SWB is never categorical. Healthy blood pressure is the objective absence of hypotension and hypertension, but SWB is less neutral. SWB is not just the absence of misery but the presence of positive contentment. Nevertheless, if SWB is an inescapable dimension of mental health, it is often regarded with ambivalence or ignored altogether. If through the centuries philosophers have sometimes regarded "happiness" as the highest good, psychologists and psychiatrists have tended to be as oblivious to SWB as they are to mental health. An electronic search of psychological abstracts since 1987 turned up 57,800 articles on anxiety and 70,856 on depression, but only 5,701 mentioned life satisfaction and only 851 mentioned joy.

On the one hand, "no man is happy who does not think himself so," but this will be true only if the "happiness" comes from unselfish positive emotions or "selfish" behaviors that also gratify others:for example, self-efficacy, play, and "flow" (deep but effortless involvement). Spirituality (e.g., prayer, meditation, and belief in a power greater than the self) all rid the individual of narcissistic focus on shame, resentments, and the "poor-me's." Fostering the 24 Strengths in Table 3.5–1, model B is perhaps the most important prescription for SWB from Positive Psychology.

The mental health issues involved in SWB are complicated and clouded by historical relativism, value judgment, and illusion. Europeans have always been skeptical of American concern with happiness. Only in the last decade have investigators like Barbara Fredrickson, Martin Seligman, and David Snowdon pointed out that a primary function of positive emotional states and optimism is that they facilitate self-care. SWB makes available personal resources that can be directed towards innovation and creativity in thought and action. Thus, SWB, like optimism, becomes an antidote to learned helplessness. Again, controlling for income, education, weight, smoking, drinking, and disease, happy people are only half as likely to die at an early age or become disabled as unhappy people.

A distinction can be made between *pleasure* and *gratification.* Pleasure is in the moment, is closely allied with happiness, and involves the satisfaction of impulses and biological needs. Pleasure is highly susceptible to habituation and satiety. In contrast, gratification can be equated with what Aristotle called *eudaimonia* and Csikszentmihalyi terms “flow.” In such a distinction, if pleasure involves satisfaction of the senses and emotions, gratification involves joy, the satisfaction of “being the best you can be,” and of meeting aesthetic and spiritual needs. This can be manifested by a child lost in play, a mountaineer transported by rock climbing, or a father marveling at his daughter’s first solo on a bicycle.

But once again, the role of value judgment clouds the view of whether any given form of happiness is good or bad. Is the happiness of helping others better than the happiness of gratifying creative play? The jury is still out. In addition, all emotions exist to promote survival as well as to communicate with others. Subjective (unhappy) distress can be healthy. As ethologically minded investigators have long pointed out, subjective negative affects (e.g., fear, anger, and sadness) can be healthy reminders to seek environmental safety and not to wallow in SWB. If positive emotions facilitate faith, hope, optimism, and contentment, fear is the first protection against an external threat; sadness protests against loss and summons replacement, and anger signals trespass.

Positive emotion generates exploration and promotes mastery. Unlike negative emotions, which generate zero-sum, win-lose situations, positive emotions generate win-win situations. However, tolerance of negative emotions may characterize some periods of the lifecycle more than others. The relative absence of depression and anxiety over time is healthy in grammar school children from 8 to 10 and in young submariners in their 20s. But increasing capacity to tolerate and bear subjective depression and anxiety distinguishes healthy adolescents from healthy fifth graders and healthy (generative) submarine skippers from healthy enlisted men. There is a time to hope and a time to fear. Human development is complex.

Until recently, the scientific parameters of SWB were as vague as those for objective mental health. A 1967 definition suggested that a happy person is “*young, healthy, well-educated, well paid,* extroverted, optimistic, *worry-free,* religious and married with high self-esteem, a good job, morals, and *modest aspirations.*” In the last 30 years, however, research has shown that such a vacuous generalization only partly corrects and that the italicized adjectives are untrue or true only with qualifications.

The Nun Study by Snowdon and Danner provides perhaps the most convincing link between subjective happiness and physical health. In their 20s, 180 nuns were asked to write a 2 to 3-page autobiography. Of those who expressed the most positive emotion, only 24% had died by 80. In contrast, by age 80, 54% of those who expressed the least positive emotion had died.

Only in the last 3 decades have investigators, especially Edward Diener, made a serious effort to attend to definitional and causal parameters of SWB and thereby address important questions. One such question is: Is SWB more a function of environmental good fortune, or is it more a function of an inborn, genetically based temperament? Put differently, does SWB reflect trait or state? If SWB reflects a safe environment and the absence of stress, it should fluctuate over time, and those individuals who are happy in one domain or time in their lives might not be happy in another.

A second question but one related to the first is what is the cause and what is the effect. Are happy people more likely to achieve enjoyable jobs and good marriages, or do conjugal stability and career contentment lead to SWB? Or are such positive associations the result of still a third factor? For example, the absence of a genetic tendency for alcoholism, for major depression, for trait neuroticism, and even for the presence of a chronic wish to give socially desirable answers (impression management) might conflate both SWB and reports of a good marriage and career contentment.

Research over the last 3 decades has provided a tentative answer to the first question. Modern research has confirmed the aphorism of Duc de La Rochefoucauld, “Happiness and misery depend as much on temperament as on fortune.” SWB is highly heritable and relatively independent of demographic variables. SWB is due more to “top-down” processes—temperamental factors governing SWB—rather than due to “bottom-up factors”—for example, high income leading to SWB.

As with physiological homeostasis, evolution has prepared to make subjective adjustments to environmental conditions. Thus, one can adapt to good and bad events so that he or she does not remain in a state of either elation or despair. But they have a harder time adjusting to their genes. Studies of adopted away twins have demonstrated that half the variance in SWB is due to heritability. The SWB of monozygous twins raised apart is more similar than the SWB of heterozygous twins raised together. Among the heritable factors making a significant contribution to high SWB are low trait neuroticism, high trait extraversion, absence of alcoholism, and absence of major depression. In contrast to tested intelligence, when heritable variables are controlled, SWB is not affected by environmental factors like income, parental social class, age, and education.

For example, investigators have been startled that a significant number of acquired immunodeficiency syndrome (AIDS) victims perceive that their illness has enhanced the quality of their subjective lives. Paraplegic victims of spinal injuries adapt so that within 2 months after injury, their SWB returns to a state where positive emotion exceeds negative emotion. Similarly, after a few weeks of temporary elation, lottery winners also return to baseline. Some of the homeostatic mental mechanisms underlying such adaptation were elaborated on in a prior section.

If SWB were due largely to the meeting of basic needs, then there should be a relatively low correlation between SWB in work and SWB in recreational settings or between SWB in social versus SWB in solitary settings. However, in most studies, the correlations between SWB in different facets of life are very high—higher than the correlation between height and weight. People who are satisfied with their lives at one point in time are more likely to be satisfied with their jobs in the future, and people with high job satisfaction are more likely to report satisfaction with retirement. Admittedly, a few studies exist where correlations are low between satisfaction with different facets of life. Again, more work is needed.

Since women experience more objective clinical depression than men do, the fact that gender is not a determining factor in SWB is interesting. One explanation is that women appear to report both positive and negative affects more vividly than men. In one such study, gender accounted for only 1% of the variance in happiness but 13% of the variance in the intensity of reported emotional experiences.

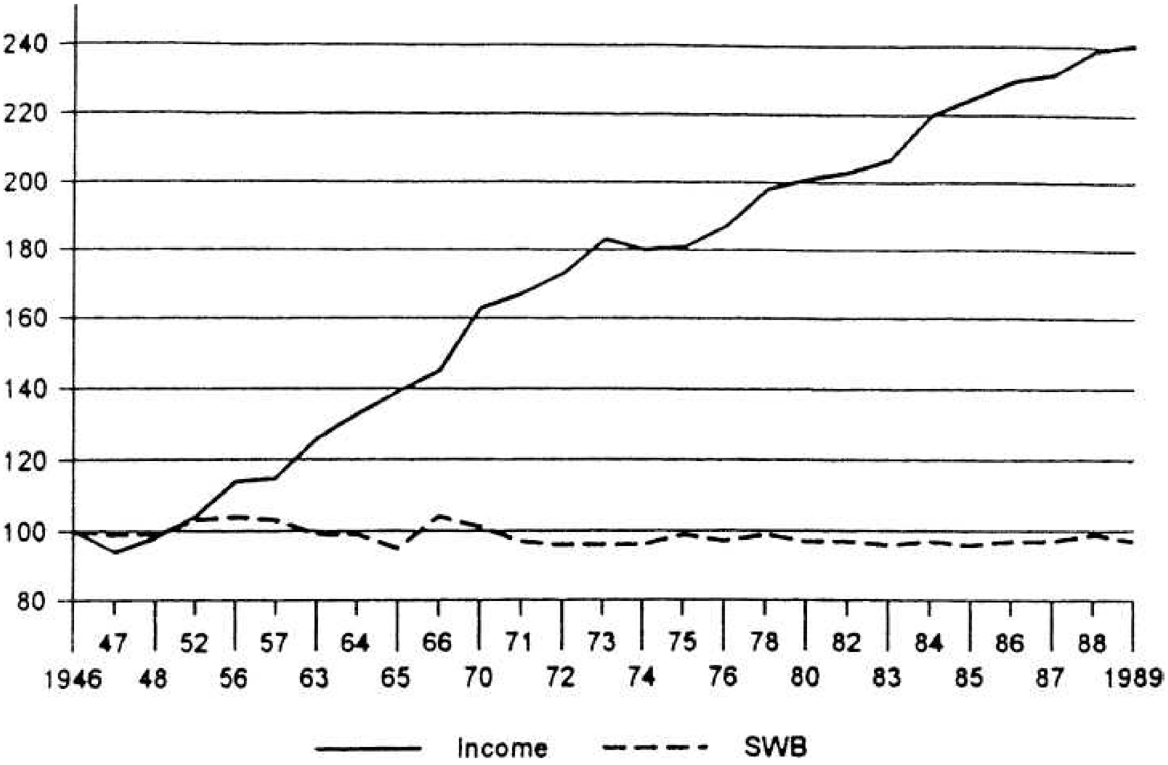

FIGURE 3.5–3. The independence of a 240% increase in US individual income (adjusted for inflation) from subjective well-being over 40 years.

Consistently, relationships are more important to SWB than money. In a representative study of 800 college alumni, respondents who preferred high income, occupational success, and prestige to have very close friends and a close marriage were twice as likely to describe themselves as "fairly" or "very" unhappy. Over the last 2 decades, the doubling of net disposable income (controlling for inflation) in the Western world (Fig. 3.5–3) did not affect SWB. Additional evidence that environmental events are much less important to SWB than might be expected comes from cross-cultural studies. Mean life satisfaction in socioeconomically challenged Brazil and China is higher than in socioeconomically blessed Japan and Germany.

In some instances, the environment *can* be important to SWB. Young widows remain subjectively depressed for years. Even though their poverty has been endured for centuries, respondents in very poor nations, like India and Nigeria, report lower SWB than do other more prosperous nations. The loss of a child never stops aching. Although achieving concrete goals, like money and fame, does not lead to a sustained increase in SWB, social comparison (i.e., watching one's next-door neighbor become richer than you) does exert a negative effect on SWB.

Maintaining self-efficacy, agency, and autonomy make additional environmental contributions to SWB. For example, surprisingly, elders will use their discretionary income to live independently, even though this means living alone rather than with relatives. SWB is usually higher in democracies than in dictatorships. Assuming responsibility for favorable or unfavorable outcomes (internalization) is another major factor leading to SWB. Placing the blame elsewhere (externalization) significantly reduces SWB. In other words, the mental mechanisms of paranoia and projection make people feel worse rather than better.

Over the past 30 years, methodological solutions have accelerated understanding of SWB just as they have accelerated understanding of objective mental health. Longitudinal studies, multivariate analysis, and causal modeling have each done much to free investigators from correlational reasoning. Nevertheless, efforts to measure SWB have been quite varied and lack a gold standard. Some investigators measure SWB as simply global life satisfaction; other investigators assess more specific domains like work, marital satisfaction, and positive affect. However, the question "How do you feel about your life as a whole?" answered on a simple 7-point scale ranging from "delighted" to "terrible" works surprisingly well.

More refined methods of measurement of subjective states of mind have included the the positive and negative affect scale (PANAS), which assesses both positive and negative affect, each with ten affect items. *The Satisfaction with Life Scale* represents the most recent evolution of a general life satisfaction scale. Most recently, the widely validated SF-36 has allowed clinicians to assess the subjective cost–benefits of clinical interventions. Because short-lived environmental variables can distort SWB, the consensus is emerging those naturalistic experience-sampling methods are the most valid way to assess SWB. With such sampling methods, research subjects are contacted by beeper at random times during the day for days or weeks, and at each interval are asked to assess their SWB. This method provides a more stable report of SWB. Finally, in order to tease verbal self-report from actual subjective experience, physiological measures of stress (e.g., measures of galvanic skin response, salivary cortisol, and recording facial expression by concealed cameras) have also proven useful.

In the 21st century, investigators like Fredrickson, Vaillant, and Damasio have found instruments like the PANAS too cognitive and replaced "positive affects" like Joviality, Self-Assurance, Attentiveness, and Serenity with more limbic affects like Joy, Love, Compassion, Trust, and Gratitude. However, these latter, more significant positive affects (described in the next model) do not lend themselves to rating scales.

MODEL G: MENTAL HEALTH AS POSITIVE OR "SPIRITUAL" EMOTIONS

In the 19th century, psychiatrists wrote about concepts like "moral insanity" and "good character," and mental health was deemed related to morality and religious observance. During the 20th century, the rise of cultural anthropology, psychoanalysis, behaviorism, molecular biology, and secularism, in general, all led psychiatrists to doubt there was any relationship between morality and health. Recently, with advances in neuroscience, that viewpoint has begun to change.

Model G as a reflection of positive ("spiritual") emotions, however, was not included in the 8th edition of this textbook because the science was not yet in place to support its inclusion. Recent advances in neuroscience and in the biological understanding of positive emotion have necessitated psychiatry taking "positive psychology" seriously. In the last 10 years, however, positive emotions, previously relegated to popular songs, pastoral counseling, and humanistic psychology, have been rendered scientifically tangible.

Fifty years ago, medical students were taught mainly about hypothalamic emotions like lust, hunger, fear, and rage. Such emotions are common also to lunatics, alligators, and decorticate cats. Prosocial emotions like empathy, compassion, and parental love were thought to be learned behaviors and, therefore, placed in the neocortex and in the curricula of schools of education.

In the 1960s, Paul MacLean, with his focus on the limbic system unique to mammals, and Harry Harlow and John Bowlby, shifting attention to love as an attachment rather than as Eros and lust, created the basis for neuroscience to provide substance to the platitudes of Jahoda, Abraham Maslow, and even of St. Paul with his "theological virtues" of faith hope and love.

Positive emotions arise from the inborn prosocial mammalian capacity for unselfish parental love. Thus, they are grounded in evolutionary heritage. Research over the last 20 years has placed these emotions firmly in the limbic system. Remove a mother hamster's cortex, and she cannot do mazes but remains a competent mother. Damage her limbic system, however, and she can still do mazes but not parent her pups.

Model G defines both mental and spiritual health as the amalgam of the positive emotions that bind to other human beings. Love, hope, joy, forgiveness, compassion, faith, awe, and gratitude comprise the important positive and "moral" emotions included in this model. Of great importance, the eight selected positive emotions all involve human connection. None of the eight emotions listed is just about the self. These positive emotions appear to be a common denominator of all major faiths. Omitted from the list are five other positive emotions—excitement, interest, contentment (happiness), humor, and a sense of mastery, for a person can feel these latter five emotions alone on a desert island.

Negative emotions originating in the hypothalamus, such as fear and anger, are elaborated in the human amygdala (larger in humans than in other mammals). Of tremendous importance to individual survival, negative emotions are all about "me." In contrast, positive emotions, apparently generated in the limbic system and unique to mammals, have the potential to free the self from the self. People feel both the emotions of vengeance and of forgiveness deeply, but the long-term results of these two emotions are very different. Negative emotions are crucial for survival in the time present. The positive emotions are more expansive and help to broaden and build. In time future, they widen one's tolerance for strangers, expand one's moral compass, and enhance one's creativity. Careful experiments by neuroscientists like Jaak Panksepp and Richard Davidson and psychologists like Barbara Fredrickson and Sonja Lyubomirsky document that while negative emotions narrow attention and miss the forest for the trees, positive emotions, especially joy, make thought patterns more flexible, creative, integrative, and efficient.

The effect of positive emotion on the autonomic (visceral) nervous system has much in common with the relaxation response to meditation. In contrast to the metabolic and cardiac arousal that the fight-or-flight response of negative emotion induces in the *sympathetic* autonomic nervous system, positive emotion via the *parasympathetic* nervous system reduces basal metabolism, blood pressure, heart rate, respiratory rate, and muscle tension. fMRI studies of Kundalini yoga demonstrate that meditation increases the activity of the hippocampus and the right lateral amygdala, which in turn leads to parasympathetic stimulation and the sensation of deep peacefulness.

Positive emotions have a biological basis, which means that they have evolved through natural selection. The prosocial emotions probably reflect adaptations, which permitted the survival of relatively defenseless Homo sapiens and their extremely defenseless children in the African savannah 1 to 2 million years ago.

Evidence for Positive Emotions

However, it has taken recent developments in neuroscience and ethology make positive emotions a fit subject for scientific study. For example, infantile autism, a not uncommon genetic disorder of emotional attachment, was not discovered until 1943 by a Johns Hopkins child psychiatrist, Leo Kanner—in his own son. Until then, it was not possible for medicine to articulate a positive emotion as basic, but as cognitively subtle, as an attachment. Today, the congenital lack of empathy and difficulties of attachment in childhood autism can be recognized by any competent pediatrician.

It was in the years 1945 to 1950 that psychoanalyst John Bowlby's ethological studies first convinced physicians that orphans needed affection as much as food. It was in the 1950s that Harry Harlow's objective findings of attachment behavior in rhesus monkeys rendered nonsexual love a tangible biologic reality for psychologists. Two decades later, Paul Ekman, doubting Margaret Mead's assertion that emotions and their facial expression were culturally based, traveled as far afield as the New Guinea highlands to demonstrate that social emotions were biological and not cultural in origin.

Locating positive emotion in the mammalian limbic system has been a slow, arduous process. In 1955, James Olds, an innovative neuropsychologist, observed that 35 out of 41 electrode placements within the limbic system of rats, but only 2 out of 35 placements outside of the limbic system, proved sufficiently rewarding to lead to self-stimulation. In the 1950s, also, neurobiologist Paul MacLean pointed out that the limbic structures govern the mammalian capacity to remember (cognition) but also to play (joy), to cry out at separation (faith/trust), and to take care of own (love). Except for rudimentary memory, reptiles express none of these qualities.

Not until the last 20 years, however, have fMRI studies demonstrated that when individuals subjectively experience existential states of fear, sadness, or pleasure, blood flow increases in limbic areas and decreases in many higher brain areas. Various studies have located pleasurable human experiences (tasting chocolate, winning money, admiring pretty faces, enjoying music, and orgasmic ecstasy) in limbic areas—especially in the orbitofrontal region, anterior cingulate, and insula. These diverse structures are closely integrated and organized to help to seek and to recognize all that falls under the rubric of mammalian love and human spirituality. Only in the last 20 years have neuroscientists like John Allman and Giacomo Rizzolatti identified the limbic spindle (von Economo) neurons and mirror cells that undergird prosocial human mentation.

The *anterior cingulate gyrus* links valence and memory to create attachment. Along with the hippocampus, the anterior cingulate is the brain region most responsible for making the past meaningful. In terms of mediating attachment, the anterior cingulate receives one of the richest dopaminergic innervations of any cortical area. Thus, the cingulate gyrus provides motivational salience not only for lovers but also for drug addicts. The anterior cingulate is crucial in directing whom one should approach and whom one should avoid. Maternal touch, body warmth, and odor via the limbic system and especially via the anterior cingulate regulate a rat pup's behavior, neurochemistry, endocrine release, and circadian rhythm. Brain imaging studies reveal that the anterior cingulate gyrus is aroused neither by facial recognition of friends per se nor by sexual arousal per se. Rather, anterior cingulate fMRI images light up when a lover gazes at a picture of a partner's face or when a new mother hears her own infant's cry.

Perhaps no area of the brain is more ambiguous in its evolutionary heritage or more crucial to mental health than the *prefrontal cortex.* The prefrontal cortex is in charge of estimating rewards and punishments and plays a critical role in adapting and regulating the emotional response to new situations. Thus, the prefrontal lobes are deeply involved in emotional, "moral," and "spiritual" lives.

From an evolutionary standpoint, the human frontal lobes are not different from chimpanzees in terms of the number of neurons. Rather, it is the frontal lobe white matter (the connectivity between neurons through myelinated fibers), which accounts for the larger frontal lobes of humans. This connectivity to the limbic system underscores its “executive” function which includes the ability to delay gratification, comprehend symbolic language, and, most importantly, establish temporal sequencing. By being able to connect memory of the past to “memory of the future,” the frontal lobes establish for Homo sapiens predictable cause and effect.

Surgical or traumatic ablation of the ventromedial prefrontal cortex can turn a conscientious responsible adult into a moral imbecile without any other evidence of intellectual impairment. Destroy the medial prefrontal lobes of the neocortex as was done by the tamping rod that penetrated the skull of the legendary 19th-century railroad worker, Phineas Gage, and one destroys the capacity for obedience to social mores. Phineas Gage had been a responsible foreman until an explosion drove the tamping rod that he was using to place a dynamite charge through the orbit of his left eye and out through the top of his skull. Although he survived, his medial frontal lobes were destroyed, as was his capacity for socially appropriate and empathic behavior. His intellect remained intact, but lacking empathy, he was no longer mentally healthy.

The *insula* is another part of the limbic system that is only beginning to be understood. The insula is a medial cortical gyrus located between the amygdala and the frontal lobe. The brain has no sensation; humans feel emotion only in their bodies. The insula helps to bring these visceral feelings into consciousness: the pain in one’s heart of grief, the warmth in one’s heart of love, and the tightness in one’s gut from fear all make their way into consciousness through the insula.

Both the limbic anterior cingulate and insula appear active in the positive emotions of humor, trust, and empathy. The higher apes are set apart from other mammals by a unique neural component called the spindle cell. Humans have 20 times more spindle cells than either chimps or gorillas. (Adult chimpanzees average about 7,000 spindle cells. Human newborns have four times more, and human adults have almost 200,000 spindle cells.) Monkeys and other mammals, with the possible exception of whales and elephants, are totally lacking in these special cells. These large, cigar-shaped spindle or “von Economo” neurons appear central to the governance of social emotions and moral judgment. Spindle cells may have helped the great apes and humans integrate their mammalian limbic system with their expanding neocortex. Spindle cells are concentrated in the anterior cingulate cortex, the prefrontal cortex, and the insula. More recently, scientists have discovered a special group of “mirror neurons” that reside in the insula and anterior cingulate. These neurons are more highly developed in humans than in primates and appear to mediate empathy—the experience of “feeling” the emotions of another.

While the practical applications of this newest model of mental health are still many years away, these findings provide further evidence that the brain and mind are one. In several studies, the prosocial biological activity of the anterior cingulate cortex and insula was highest in individuals with the highest levels of social awareness (based on objectively scored tests). In other words, there are not only individual biological differences for negative mental health but also for positive mental health.

FUTURE DIRECTIONS

In the 19th century, mental health was deemed related to morality, and psychiatrists wrote of “moral insanity” and “good character.” But in the latter half of the 20th century, with a 19th-century boost from Sigmund Freud, religion and spirituality were expunged from psychiatry.

Religiosity is consistently and positively correlated with well-being, but there may be a chicken-egg relationship between social support and religious observance. For example, among churchgoers, it is difficult to disentangle where spiritual faith ends and community support begins. Nevertheless, it makes sense that positive faith should enhance SWB. Heightened spirituality provides hope that bad events will diminish and lends wonder and joy to ordinary events. While the concept of spirituality still remains too value-laden to serve as a 7th model of mental health, increasingly sophisticated neurological research on the positive emotions by investigators like Jan Panksepp, Marco Iacoboni, Barbara Fredrickson, and Andrew Newberg may find in the 21st century a role for spirituality among the models for mental health.

Table 3.5–2.
Cross-Correlation Between Four Different Models of Mental Health at Midlife (Age 47) and Their Positive Prediction of Late Life (Age 65) Mental Health and Their Independence from Parental Social Class and Childhood Environment

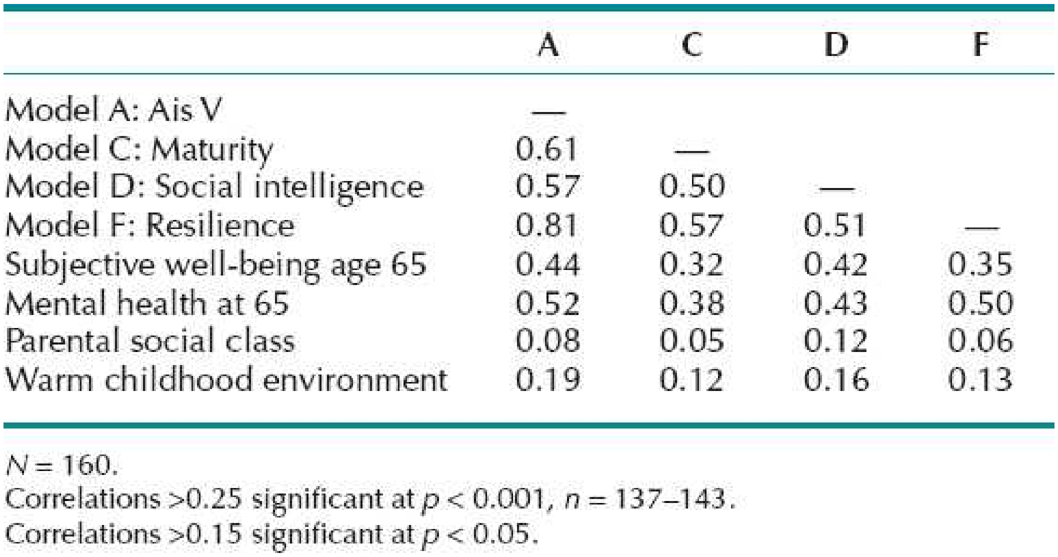

	A	C	D	F
Model A: Ais V	—			
Model C: Maturity	0.61	—		
Model D: Social intelligence	0.57	0.50	—	
Model F: Resilience	0.81	0.57	0.51	—
Subjective well-being age 65	0.44	0.32	0.42	0.35
Mental health at 65	0.52	0.38	0.43	0.50
Parental social class	0.08	0.05	0.12	0.06
Warm childhood environment	0.19	0.12	0.16	0.13

$N = 160$.
Correlations >0.25 significant at $p < 0.001$, $n = 137–143$.
Correlations >0.15 significant at $p < 0.05$.

This chapter has so far suggested six conceptually distinct ways to assess a single construct—mental health. It would be a terrible mistake to believe any one of these six models is superior to the others. From different vantage points, they all measure the same thing. For example, case example, Richard Luckey not only had a GAF of 95, but another independent rater saw him as highly generative, and a third rater saw him using defenses largely from level VII. He radiated SWB. Contrasting case example, Alfred Paine had a GAF of 72, was not generative, used largely level II–V defenses, and was chronically dissatisfied with his life. “Dilip Jeste has recently developed a seventh clear definition and the neurobiological underpinning for “Wisdom” as a construct for mental health.

Table 3.5–2 (from the Study of Adult Development at Harvard) provides an empirical illustration of the interrelationship of these different models in a 75-year prospective study of nondelinquent inner-city men. Not only was each of the four models (measured by independent raters) significantly correlated with the other three, but also each model predicted objective global mental health and SWB assessed 20 years later. Significantly, none of the four models was well predicted by parental social class or a warm childhood environment.

In Table 3.5–2, Axis V was assessed by Luborsky's HSRS. Maturity was determined by an assessment of the presence or absence of Generativity. SWB at 55 was quantified by summing each man's report of satisfaction over the past 20 years (on a 5-point scale) in four life areas (marriage, children, job, and friends) and, then, adding his best score from one of four additional areas (hobbies, sports, community activities, and religion). Social intelligence was assessed by objective evidence of good relations with friends, workmates, spouses, children, and siblings. Resilience was measured by the relative frequency with which men deployed the five **mature defenses** against the relative frequency with which they used **the five immature** (*level II–V*) defenses. Mental Health at 65 reflected success at work, relationships, play, and not using psychiatrists. Parental social class used the method of Hollingshead and Redlich (parental residence, education, and occupation). Warm Childhood Environment reflected good relations with mother, father, siblings, and a cohesive home at age 14.

Currently, four of this chapter's models are capable of being assessed psychometrically—above-average normality by the GAF (Axis V), maturity by the presence or absence of Generativity, SWB by self-report, and resilience by defense level on the optional DSM-IV axis. As already noted, measures to assess psychometrically the other two models—the Strengths of Positive Psychology and socioemotional intelligence are underway.

The concept of mental health also raises the issue of therapeutic interventions to achieve it. What facets of mental health are fixed and which are susceptible to change? With Clozapine or with cognitive-behavioral therapy, one can raise a GAF from 40 to 70, but how would one raise a GAF from 70 to 90? Chemicals can alleviate mental illness but do not improve healthy brain function. One can enhance mental health only through cognitive, behavioral, and psychodynamic education.

One can change some facets of brain function better than others. By analogy, the most intensive educational intervention in individuals who are not severely deprived will raise their IQ by only about seven points, but sustained therapeutic intervention can change individuals utterly illiterate in Italian into fluent Italian conversationalists. Admittedly, a correct accent to go with the words is harder to teach.

In concluding this chapter, it seems important to review some of the safeguards for a study of positive mental health. First, mental health must be broadly defined in terms that are culturally sensitive and inclusive. Second, the criteria for mental health must be empirically and longitudinally validated. Third, validation means paying special attention to cross-cultural studies. In somatic medicine, criteria have been developed so that people of widely varying backgrounds and beliefs can agree upon what constitutes rational therapy for a disease. The same criteria for mental health need to be developed. Fourth, although mental health is one of humanity's important values, it should not be regarded as an ultimate good in itself. One must proceed in efforts toward trying to improve mental health while maintaining due respect for individual autonomy.

There is no logically perfect method of analysis by which one can distinguish private bias and culturally determined value judgment from a culturally universal and scientifically valuable definition of mental health. Predictive validity remains the best guide. A parable offered by the protagonist in Gotthold Lessing's great 18th-century play, *Nathan the Wise,* will help to illustrate why mental health can only be identified by longitudinal study. In Lessing's play, an angry sultan had asked Nathan, a Jew, on pain of death, to identify the one true religion—Christianity, Islam, or Judaism. Nathan gently pointed out the need to maintain a longitudinal perspective.

"A man lived in the East,
Who owned a ring of marvelous worth,
Given to him by a hand beloved.
The stone was opal, and possessed the secret power
To make the owner loved of God and man,
If he but wore it in this faith and confidence…"

The ring was sought as an inheritance by each of Nathan's three sons. Since he loved him all equally, the doting father gave each son an identical ring. After their father's death, the three sons hurried off to a judge and demanded that the judge identify the lucky owner of one true ring. The judge exclaimed,

"But stop! I've just been told that the right ring,
Contains the wondrous gift to make its wearer loved,
Agreeable alike to God and Man.

That must decide, for the false rings will not have this power…
Let each one strive to gain the prize of proving by results
The virtue of his ring and aid its power
With gentleness and heartiest friendliness…
The virtue of the ring will then
Have proved itself among your children's children."

In the future, it is incumbent upon psychiatry to understand the implications of each of these six models of mental health. In conducting research on positive mental health, it is important for psychiatrists to recognize which model they use. Equally important, in the area of national health policy, one needs a clear understanding of not only what constitutes positive mental health but also of who is responsible for intervention and who is responsible for paying for it.

Finally, primary prevention is clearly superior to treating disease once it has occurred. Thus, we need to study individuals with positive mental health the way agronomists study wheat that is resistant to drought and blight. Psychiatry also needs to be able to measure and record mental health. The Global Assessment of Functioning provides the same reliability and has much greater predictive validity than the presence or absence of most psychiatric diagnoses. No psychiatric chart should be without an assessment of functioning. The capacities to work and to love over time are extremely important indices of mental health. They are far more important than the cross-sectional presence or absence of anxiety, depression, or illegal drug use. But the capacity to work and to love must be assessed longitudinally. "How many years since age 21 have you spent employed?" is more useful than, "What is your present job?" Again, "Tell me about your longest intimate relationship" is more useful than "Are you married?"

Finally, the assessment of maturational development provides the best prediction of future clinical courses. Thus, the mental status and diagnostic formulation should both reflect an assessment of social maturation and coping style. If the person is 35, have they mastered Erikson's task of intimacy? If they are 40, have they achieved competence in, commitment to, contentment with, and compensation from their career? If they are over 50, have they mastered Generativity and learned to care less about themselves and more about their children? Again, when the going gets tough, do they eschew less mature mental mechanisms like projection, passive aggression, and dissociation (being in denial), and do they employ mature involuntary coping mechanisms like grit, humor, altruism, and sublimation?

FURTHER READINGS

American Psychiatric Association. Global Assessment at Functioning (GAF) scale. In: *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. American Psychiatric Association; 1994:32.

American Psychiatric Association. defense levels and individual defense mechanisms. In: *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. American Psychiatric Association; 1994:752.

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Csikszentmihalyi M. *Flow: The Psychology of Optimal Experience*. Harper Row; 1990.

Dilip Jeste, *Wiser*. Sounds True, Boulder CO; 2020.

Diener E, Suh EM, Lucas RE, Smith HL. Subjective well being: three decades of progress. *Psychol Bull*. 1999;125:276–302.

Duckworth AL, Peterson C, Mathews MD, Kelly DR. Grit: persistence and passion for long term goals. *J Pers Soc Psychol*. 2007;92:1087–1101.

Erikson E. *Childhood and Society*. Norton; 1950.

Frederickson BL. What good are positive emotions? *Rev Gen Psychol*. 1998;2:300–319.

Goleman, D. *Emotional Intelligence*. Bantam Books; 1995.

Grinker RR Sr, Grinker RR Jr. Mentally healthy young males (homoclites). *Arch Gen Psych*. 1962;6:405–453.

Iacobini M. *Mirroring People*. Farrar, Straus, and Giroux; 2008.

Jahoda M. *Current Concepts of Positive Mental Health*. Basic Books; 1958.

Korchin SJ, Ruff GF. Personality characteristics of the Mercury astronauts. In: *The Massachusetts Institute of Technology, ed. The Threat of Impending Disaster*. MIT Press; 1964.

Ledoux J. *The Emotional Brain*. Simon and Schuster; 1996.

Loevinger J. *Ego Development*. Jossey-Bass; 1976.

Luborsky L. Clinicians' judgment of mental health. *Arch Gen Psychiatry*. 1962;7:407–417.

Mayer JD, Caruso D, Salovey P. Emotional intelligence meets traditional standards for an intelligence. *Intelligence*. 1999;27:267–298.

Menninger WC. *A Psychiatrist for a Troubled World: Selected Papers of William C. Menninger, M.D.* The Viking Press; 1967:788–804.

Offer D, Sabshin M. Chapter 66: Normality. In: Kaplan HI, Freedman AM, Sadock BJ, eds. *Textbook of Psychiatry*. 3rd ed. Williams and Wilkins; 1980:608–613.

Panksepp J. *Affective Neuroscience*. Oxford University Press; 1998.

Peterson C. *A Primer in Positive Psychology*. Oxford University Press; 2006.

Peterson C, Seligman ME. *The VIA Classification of Strengths*. Values in Action Institute; 2003.

Salovey P, Sluyter DJ, eds. *Emotional Development and Emotional Intelligence*. Basic Books; 1997.

Seligman MEP. *Authentic Happiness*. Free Press; 2002.

Seligman MEP, Csikszentmihalyi M. Positive psychology. An introduction. *Am Psychol*. 2000;55:5–14.

Tellegen A, Lykken DT, Bouchard TJ, Wilcox KJ, Segal NL, Rich, S. Personality similarity in twins reared apart and together. *J Pers Soc Psychol*. 1988;54:1031–1039.

Vaillant GE. *Ego Mechanisms of Defense: A Guide for Clinicians and Researchers*. Am Psych Press; 1992.

Vaillant GE. *Aging Well*. Little Brown; 2002.

Vaillant GE. *Spiritual Evolution: A Scientific Defense of Faith*. Doubleday Broadway; 2008.

Vaillant GE. *Triumphs of Experience*. Harvard; 2012.

3.6 Neuroscience of Psychoanalytic Object Relations Theory

Dragan M. Svrakic, M.D., Ph.D., Charles F. Zorumski, M.D., and Otto F. Kernberg, M.D.

Object relations theory is a major modification of classical (Freudian) psychoanalysis. It places the innate instinct to bond, rather than pleasure-seeking, as the driving mechanism of mind development. Early experiences with first “objects” (primary caretakers, usually the mother), so the theory goes, are forgotten as specific episodes but continue to influence adult personality and significant relations in a self-reinforcing fashion. The tenets of object relations theory have been inferred from observations of infants and children. For decades, the sole empirical support for the validity of object relations theory has been the efficacy of psychoanalytic psychotherapy[1] in the treatment of personality disorder, a syndrome paradigmatic for partial (fragmented) object relations. In this selective overview, object relations theory is discussed through the prisms of cognitive- and basic neuroscience, dynamical systems theory, coordination dynamics theory, and complex systems science. Our main postulates are (i) recent advances in the understanding of neural encoding and systems-consolidation of episodic memory provide a framework to outline the neural mechanisms of object relations theory[2]; (ii) dynamical system theory, natural network and computational network science provide a framework to outline object relations as the primordial[3] attractor memory that governs self-organization of BrainMind[4] in a self-reinforcing fashion, and (iii) functional neuroimaging and cognitive neuroscience provide a framework to explain the emergence and the potential neural substrate of the sense of self.

As groundwork for our discussions, we first review basic aspects of brain and mind organization and then the tenets of object relations theory. Each following section is discussed through a dual prism of object relations theory and neuroscience. As the chapter is limited to 30 references, acknowledgments provide a list of authors whose work was used but could not be explicitly cited. Portions of this text have been previously published.

BASICS OF BRAIN AND MIND SELF-ORGANIZATION

The Brain: A Self-Organizing Biological Complex

The brain is understood to be an open, self-organized, dynamical, complex, and adaptive biologic system. Such characteristics beget the brain's ability to generate outcomes new to itself ("emergent properties"), the most global of which is the mind. The concepts of self-organization, complexity, and dynamical systems, attributed to the brain, are the subject of physics, mathematics, biology, cybernetics, and network- and complexity science. Because these fields might be new to some readers, we discuss them briefly as an introduction. We first discuss the terms mind, psyche, mental function, and personality, as these are used inconsistently in the literature.

Mind, Psyche, Mental Function, Personality. The terms mind, mental function, and psyche are equivalent as they derive from Latin mens for mind or from Greek (ψυχη´—psyche) for "soul" or "spirit." The *mind* is best conceptualized as a progressively differentiated but initially global capacity to store previous dynamics in existing brain matter in order to predict the environment, change it, or adjust to it. Functional neuroimaging indicates that *personality* likely reflects a specific configuration of the mind in any given individual, a unique ontogenetic outcome of the general phylogenetic capacity of human beings to develop mind. Caterina Gratton and colleagues report that functional connectivity networks across normal subjects show a stable general structure but also unique and stable individual differences. Individually unique network variants were most pronounced for top-down control networks regulating attention, executive function, salience, and emotion. A network variant shows stability over time within a person (i.e., trait-like) and can accurately identify that specific individual from a large group, likened to a fingerprint. The elementary topography of these networks is present at birth, but network connectivity (topology) continues to interactively self-organize into early adulthood. The outcome is influenced by person-specific factors (genetic polymorphisms, unique environments), thus creating individual differences in mental faculties and behavioral traits called personality.

Self-Organization of Open Biologic Systems. A true self-organizing system achieves and maintains order without a “teacher” but through local interactions of its constituting components governed by intrinsic properties. Self-organization in complex biologic systems improves the system in some respect, for example, unadaptable components are self-organized into an adaptable whole in the given environment. This means that some extrinsic criterion, usually adaptive demands, informs the process. In fact, biologic adaptive systems are thermodynamically open (they operate unbalanced, “far-from-equilibrium,” with a net macroscopic flow of matter or energy) but may be stabilized by input and output. Such systems import order (information, energy) and export entropy or disorder (e.g., the brain dissipates heat by various mechanisms). The brain self-organizes in part based on intrinsic prenatal network development (an immature language network is set in place before birth) and in part by importing order (e.g., the language network self-organizes into its mature form from speech input). Open dynamical biologic systems, in other words, participate in their own organization rather than truly self-organize. With this stipulation in mind, we use the term self-organization in reference to BrainMind.

Complexity and Emergence. Different types of neurons and glia are engaged in different relationships that follow local synaptic rules, for example, interdependencies, competitions, autoregulation, etc. Such microscopic complexity is necessary for higher-order emergence, a macroscopic event that is not a property of and cannot be reduced to any of the microscopic constituting elements. Applied to neuroscience, emergence (e.g., cognitive concepts) is a more or less predictable macroscopic consequence of the underlying collective dynamics of less complex individual neurons. For example, simple semantic concepts called “prototypes” arise in the brain’s convergence zones (e.g., the prefrontal cortex [PFC], among others) from sensory processing within individual neurons at lower levels. The concept of chair encompasses different designs that are unified by the shared essence of an object being intended for sitting. Here, emergence is one level, as the higher level of abstraction emerges from details implemented by the lower sensory level; the concept of chair is thus unexpected given the properties and principles of the lower-level domain, but predictable albeit not in full detail. Emergence can be multilevel, with many positive and negative feedback loops governing the individual activity of billions of neurons. This gives rise to “new world” emergence, with completely new roles while the old roles disappear. The most notable example of such multilevel emergence is the mind (although some consider the mind to be a weak emergence, one level above the neurons–see section on MindBody debate).

Feedback and Learning. Feedback is essential for the self-organization of open biologic systems as it drives learning. Feedback is a circular process (“looping”) in which output is returned (“fed back”) to its input to regulate future output. Negative feedback (higher output reduces further input) levels differences in the collective behavior of neurons leading to stability that counteracts change. Positive feedback (higher output increases further input) amplifies differences and reinforces the direction of change. In complex dynamical systems, where learning has to incorporate proximal consequences, feedback is the means to such learning.

Dynamical, Adaptive, and Scale-free Systems in Biology. A system is said to be *dynamical* if it changes in time through the iterated application of an underlying transition rule (“dynamical rule”). Given the current state of a dynamical system (a collection of points in an appropriate state space), the dynamical rule predicts the immediate future of all state variables, thereby describing the temporal evolution of the system. Applied to neuroscience, a current state in dynamical brain development is defined by phase-specific gene activity (epigenetic programs) and phase-specific environmental factors (change with age) acting upon the previous phase. Each new phase is influenced by the degree of success at the previous phase and indirectly by the degree of success at all previous phases. In other words, the future behavior of a dynamical system, here the brain, depends not only on the initial conditions but also on the history. This is called *path dependence* and is typically mediated by positive feedback loops that top-down channel lower levels to continue to produce input of the same kind.

A biologic dynamical system is *adaptive* if it has the ability to reorganize in response to changing conditions in the environment. New adaptive demands may render the current configuration of an open complex system adaptively suboptimal. The system then learns (imports order) and rapidly self-organizes toward a fitter organization-function (development). Such punctuated transitions are natural and age-selective in BrainMind development, especially frequent in the first decade of life. For example, temperament traits are the main communication avenue of infancy but become suboptimal to meet the growing demand for bilateral interactions in toddlerhood. Around 2 to 3 years of age, the mental sense of "I" (self) in the form of passive noticing of one's own selfhood ("phenomenal self") relatively abruptly differentiates in the mind, now fitter to communicate with itself and the environment.

Scale-free dynamics are characteristic of self-organizing dynamical adaptive systems such as the brain's large-scale networks. A network is said to be scale-free when the distribution of connections between nodes[5] decreases as a power-law function of the number of nodes in the network: a few nodes are highly connected with other nodes in the network ("hubs") while the majority of nodes have only a few connections. Such "power law" distribution is sometimes referred to as the "80–20 law," in which 20% of "rich" nodes have 80% of connections. There is no true average number of connections, making the network difficult to measure or scale ("scale-free"). Self-organization and scale-free connectivity underlie important features of the brain: (i) robustness at a system level against global failure, provided that a sufficient number of vital hubs are not compromised at the same time ("targeted attack"), in which case a cascading failure could occur; (ii) ability to incorporate substantial imperfections, genetic and environmental, and still stay within the range of normalcy, that is, without affecting function in a significant way, again provided that key hubs are not affected. This ability is known as "developmental homeostasis," described by John Alcock.

Power law degree distribution is usually, but not always, suggestive of a brain network functioning *near criticality*, that is, very near the critical point of transition between two different states, one being subcritical (when the network transitions into rigid order, unable to process complexity (e.g., contradictory experience with the same person), the other being supercritical (increased network activity that may result in chaos/disorder, unable to process coherent outcomes). Such operating regime near the critical point of transitions to either rigidity or disorder allows cortical networks to dynamically screen alternatives and process complexity. See section Self-Organizing Criticality of the Brain for more details.

The Concept of Attractors in Dynamical System Theory. All complex adaptive systems, including BrainMind, are also dynamical as they change in time. The evolution of such systems is governed by stable attractors, a conditionally stable state of an evolving system for given internal and/or external conditions. In computational neuroscience, an attractor is a network state that governs the dynamical process of time evolution of the network into its stable, self-sustaining attractor state from large classes of initial network states ("basin of attraction"). Such attractor networks have a high degree of feedback that funnels the collective firing of individual neurons toward desired coherent outcomes. Applied to neuroscience, cortical network activity begins with microscopic complexity where heterogeneously connected heterogeneous types of neurons and glia, interlinked by recurrent excitatory connections, interact in nonlinear and nontrivial fashions. "Nontrivial" means that the final, stable state of an attractor network (stable pattern of firing, self-sustenance) provides feedback, positive or negative, that brakes the symmetry of individual firing and funnels it toward a desirable outcome. For example, differences in luminescence, processed by dedicated neurons in the visual network, are governed by attractor dynamics and feedback toward the collective behavior of all neurons in this network, resulting in a coherent visual percept of an object. Neural networks are thus deterministically self-organized to produce coherent macroscopic complexity (coherent vision, memory, hearing, etc.) out of incoherent microscopic complexity (individually firing neurons). Once the deterministic outcome is achieved, the network relaxes into its stable pattern of baseline firing that sustain the network when neuronal activation by external inputs is absent. The concept of stable attractor networks with persistent self-sustaining baseline activity corresponds to the brain's resting-state functional connectivity (RSFC) networks.

Most Brain Networks Are Pre-assembled In Utero by Intrinsic Activity that Predates Experience. Spontaneous neural activity, either intrinsic to neurons or generated by synaptic interactions, may synchronize local and large-scale embryonic brain networks before afferent signals reach the cortex. The fundamental topography of these earliest networks is genetically predetermined. For example, synchronized bursts of spontaneous activity govern the formation of thalamocortical circuits (these relay sensory information to the cortex, among other functions) and early intracortical networks (sensory-motor, memory, language, etc.). The postnatal surge in sensory input is processed by preassembled thalamocortical circuits and relayed to the cortex, thereby (i) forming a loop to modulate itself (the thalamo–cortico–thalamic loop) and (ii) modulating the a priori intracortical networks with input from the environment. Preassembled intrinsic networks predate experience, as noted. They act as “best guess” or “seed” templates that serve as attractors for selective input, ultimately self-organizing into respective mature networks (language, vision, etc.). For example, in 2-day-old neonates, brain regions that are parts of the auditory language network[6] are fully activated in both hemispheres by speech-input. This immature interhemispheric language network is likely generated by intrinsic synapses governed by genetic programs, before speech input. It provides a biologic basis to develop language, as an attractor network for speech-input that learns to systematize phenomes into words and syntax. The immature inter-hemispheric language network dynamically self-organizes into a left-lateralized intrahemispheric (mature) auditory language network characterized by relative stability (language acquisition). To summarize, intrinsic embryonic networks mainly serve as experience-independent attractors (codes, templates, seeds, molds) for the development of mature brain networks based on subsequent experience-dependent input from their selective internal and external stimuli.

Resting-State Functional Connectivity. RSFC of the brain is spontaneous and maintained in the absence of any particular sensory input or motor output. RSFC forms macroscale spatio–temporal frames of distributed coactive regions linked by function (“functional” or “information coupling”). Takamitsu Watanabe and colleagues show that resting state functional networks (RSNs) involved in attention-demanding cognitive tasks (e.g., the frontoparietal network [FPN]) are (roughly) governed by monostable attractor dynamics with a coarse-grained global basin of attraction, the latter governed by multistable attractor dynamics. In other words, like all complex dynamical systems, RSNs can be characterized by a large number of realizable stable dynamical states. These states are schematically illustrated in Figure 3.6–1 with different levels of coarse-graining; the horizontal axis represents various configurations of the network, while the vertical axis is the corresponding energy. The network evolves toward the lowest energy states (valleys). At the lowest level of coarse-graining (top of the graph), the number of stable attractor states, represented by the valleys, is large, and the whole graph describes multistable attractor states, visualized by a rough dark gray curve. With higher-levels coarse-graining (light gray curve), the number of attractors becomes successively smaller as low-level valleys merge, typically by some mechanism of hierarchical clustering ("deep" learning), and smooth out toward a monostable, single attractor state (the deepest valley in the blue curve). From the original multistability states (dark gray curve), one obtains a funneled representation (light gray curve) in which the number of attractors (valleys) is reduced into a (roughly) monostable state. In contrast to FPN, the Watanabbe study shows, the default mode network (DMN) to which the brain relaxes when not on task, was governed by multistable attractor dynamics. This is suggestive of subcomponent attractor networks that may support memory maintenance with different memory states (see below) as well as that DMN may underlie the conscious sense of self (DMN can be conceived as the largest RSN, composed of all other RSNs).

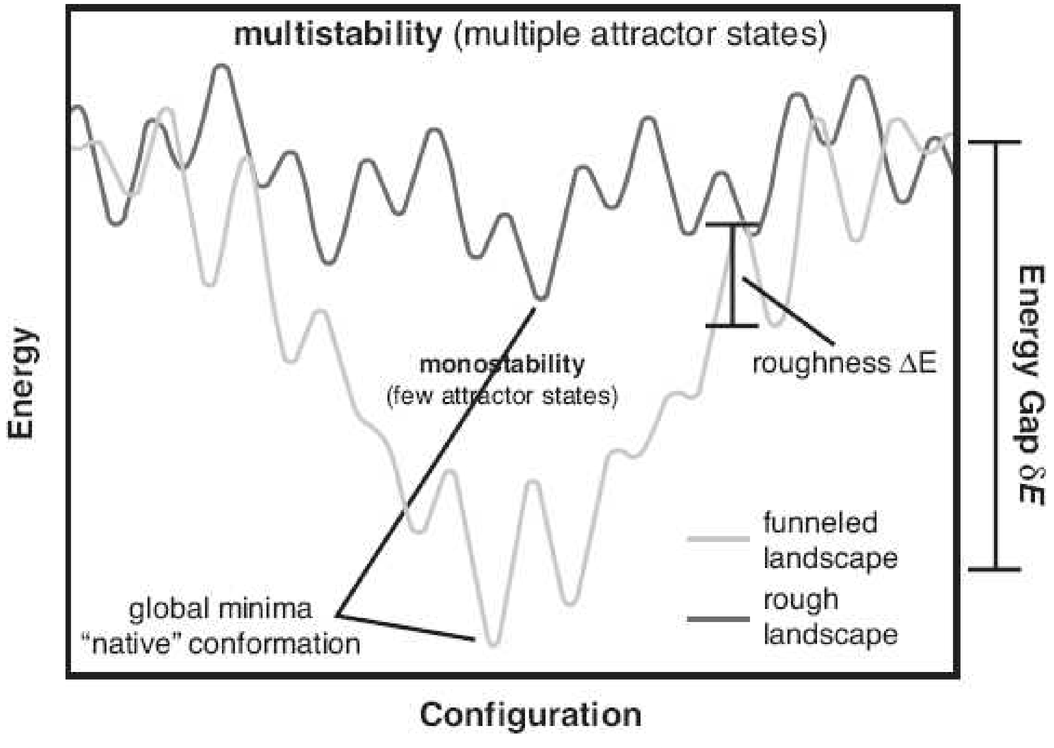
multistability (multiple attractor states)
Energy
monostability
(few attractor states)
roughness ΔE
Energy Gap δE
funneled landscape
rough landscape
global minima
"native" conformation
Configuration

FIGURE 3.6–1. Multistability and monostability of brain networks. delta E, shift in energy mediated by funneling (roughening) of multiple initial network states. (Courtesy of Charles H. Martin, Ph.D.)

The above results may be summarized as follows: when a RSN performs attention-demanding tasks, they are governed by multistable attractor dynamics, allowing the brain to travel from one relatively stable state to another, either in a spontaneous manner or triggered by external input. At rest, however, RSNs show monostable attractor dynamics that serve as algorithms for deep, gist-like learning from a large amount of data. To this end, multiple layers are used to progressively learn higher-level features from the raw input. Simply put, functional brain networks become better at performing their selective tasks (“automation”) as they dynamically reconfigure to incorporate learning by training. Karolina Finc and colleagues report that the autonomy of DMN as well as integration among task-positive networks were modulated by training.

Homeostatic (intrinsic) brain plasticity may play a role in the different appearances of RSNs at work and at rest. An emerging consensus is that resting-state neuronal activity results from the modification of strengths of connections recapitulating common recruitment during task-evoked activity. Such recapitulation tailors RSNs configurations to reflect (i) individual history and experience, and (ii) stable, trait-like individual differences in functional brain networks. The internal “tuning” of these networks by self-organization may involve homeostatic plasticity,[7] aka metaplasticity, that is, the capacity of neurons to autoregulate their own excitability based on network activity. A self-organized system tunes itself as it evolves toward criticality by constantly adjusting the control parameter to the critical value by a decentralized feedback mechanism intrinsic to the network: the control parameter spontaneously decreases when the system is in a supercritical phase and increases when in a subcritical phase. Ma and colleagues recently reported that homeostatic regulation of inhibition plays an important role in generating criticality in excitatory networks.

Moreover, RSFC is defined by covariance at rest, which is independent of mean activity. If anything, this covariance increases when brain regions are not actively engaged in task-specific processing. Abraham Snyder and Timothy Laumann interpret that high spontaneous brain activity at rest may have a role in off-line neural plasticity and homeostasis (e.g., resetting ionic balances, deep learning, memory maintenance) among others.

In the last 15 years, more than 2,500 research papers on RSFC have been published but, individually, report different topology, topography, and taxonomy for these networks. Lucina Q. Udin and colleagues made a commendable attempt to put forward a universal taxonomy for RSNs, but updates will likely be needed to incorporate new data. Thus far, about a dozen macro-scale RSFC attractor networks have been identified, among others are:

- *DMN (or medial FPN) network* is involved in the formation, temporal binding, and dynamic reconfiguration of associative representations based on current goal-states. DMN also detects the associative relevance of internal and external stimuli, provides value coding to perceived events, accommodates predictive coding, semantic associations, and plays a role in continuously monitoring the environment, among other functions; this network is believed to be the neural substrate of the sense of self, or at least a large element of this sense (discussed later in more detail).
- *Dorsal FPN attention network* (driven by internally chosen goals, “internal attention”);
- *Lateral FPN attention network aka central executive network–CEN* (involved in working memory, executive function, external stimulus-driven attention); This network maintains and manipulates information in working memory, one function of which is executive, that is, decision-making and problem-solving in the pursuit of disciplined behavior; working memory, as suggested by Marvin Chun and colleagues, lies closest to the intersection between external and internal attention.
- *Salience network (SN)* or midcingulo-insular network as a combined ventral attention network (detects salient stimuli in the environment, especially when unattended) and cingulo-opercular network (involves the detection of behaviorally relevant environmental stimuli but may include internally generated remembered information). Lucina Udin and colleagues argue that the ventral attention network instantiates the larger, bilateral midcingulo-insular network; the former appears to function mainly during exogenous salience detection, whereas the midcingulo-insular SN plays a broader role, engaging across domains during processing of personally relevant inputs. This network also mediates the dynamic switching between DMN and CEN;
- *Auditory network* (inferior frontal gyrus, superior temporal gyrus, the insula);
- *Visual network* (subcortical geniculate, primary and association visual cortices)
- *Sensory-motor network* (the insula, selective cortical regions);

- *Long- and short-term memory network* (is a multicomponent network involving cortical and subcortical structures, such as posterior–parietal, anterior–prefrontal, the limbic–temporal association areas, and the medial temporal lobe, with input from subcortical survival circuits and brainstem), and, we suggest,
- *The language network* (left-dominant network that comprises parts of the inferior temporal gyrus, premotor cortex, upper temporal lobe, and temporo-parietal interface).

Self-Organized Dynamical Oscillatory Motifs may be the Mechanism of the Brain's Functional Connectivity. In the early 2000s, John Beggs and Dietmar Plenz described spontaneous sharp negative deflections from local field potentials (nLFPs) in cortical neuronal assemblies at rest. As used by the authors, the term "neuronal assemblies" corresponds to attractor networks. Each nLFP self-organizes from local field potentials, the latter representing a summed macroscopic synchrony of microscopic spiking of individual neurons. Such population synchrony results from phase coupling between theta frequencies (4 to 8 Hz, stable attractor rhythm) and gamma frequencies (25 to 35 Hz, inhibitory rhythm, modulates the rhythm of principal neurons). Such phase coupling is known as "theta-gamma nests," where the phase of slow rhythm modulates the power of faster oscillations. These nLFPs are short-lived (milliseconds), separated by several seconds of quiescence, and appear across many spatial scales. They are believed to embed other spontaneous self-organized dynamical motifs that capture macroscopic synchrony of microscopic activity of individual neurons, such as network "up-state," neural avalanches[8] and coherence potentials (both consist of contiguous frames of nLFPs). Neural avalanches are comprised of frames of nLFPs, those with up to two standard deviations above a minimal amplitude threshold of the ongoing LFP amplitude fluctuations. Coherence potentials are composed of frames of nLFPs with ≥3 standard deviations above this amplitude. When arranged as statistical dependencies, contiguous frames of nLFPs self-organize as a cascade event (wave-like neuronal avalanches and coherence potentials) that capture the emergence of population synchrony at many local sites. The above four self-organized dynamical oscillatory motifs occur in a network engaged in spontaneous self-sustaining activity in the absence of input. This spontaneous activity is called "network up-state" in which network supports theta/gamma nests, nLFPs, neural avalanches, and coherence potentials, all involved in self-sustenance. These motifs are only weakly modulated by stimulus presentation, a characteristic of most RSNs.

To summarize, microscopic spiking of individual neurons governed by the attractor state of network near-criticality,[9] reaches synchrony as LFPs (theta/gamma nests) out of which nLFPs, neural avalanches, and coherence potentials self-organize across cortical sites forming diverse spatiotemporal patterns. It is plausible that the distributed brain attractor networks are functionally unified in the context of critical phenomena that maximize information transmission, information storage, dynamic range, metastable states, and computational power.

Multistability and Metastability are Basic Principles of Brain Computation. The brain continuously communicates with facets of internal and external environments for taking specific inputs and giving outputs. The first ingredients in the process are multiple functional attractor networks (e.g., visual, auditory, memory, language, etc.), each with a selective basin of attraction (aka stimulus- or demand-selectivity), each settling in a stable self-sustained pattern of firing once its selective stimulus has been processed (resting state). Energy inputs in the form of changes in control parameters (e.g., a mismatched demand or noise of random firing that indicates the weakness of the attractor network for the given event) may cause switching among functional attractor states, known as multistability[10] of a complex system. Multistability enables the brain to flexibly shift focus between attractor networks as conditions change. During resting states, a network shows (roughly) monostability that reflects deep learning (see Fig. 3.6–1).

Metastability of attractor networks may provide the basis for global brain function. As Scott Kelso points out, the origin of metastability lies in multistability: weak oscillatory coupling and broken asymmetry in discrete attractor networks at near-criticality allow for self-organization of a metastable regime where stable (attractor) and unstable (repellor) synchronization states no longer exist, and activity of individual elements obeys neither the intrinsic dynamics of the elements nor the dynamics dictated by the assembly. Such weakened fixed-point attractors, now “ghost attractors,” enable memory networks to process complexity.

Self-Organizing Criticality of the Brain. In the seminal concept of self-organized criticality, Per Bak posits that stable attractors allow open dynamical systems to function at near-criticality. The brain criticality hypothesis holds that neural networks self-organize into a critical state, at a boundary between different types of dynamics, between stability and instability, order and disorder. As open dynamical systems move away from equilibrium (high entropy, disorder, incoherence) toward stable attractor states (rigidity, order), they eventually reach a point (“critical point, tipping point, bifurcation point, “criticality”) at which the system may either devolve to randomness (disorder) or to the inflexibility of stable attractor state. Right before the critical point (“near-criticality”), an optimally self-organized attractor network may reach the level of complexity to embed a metastable state that hovers between order and disorder, thus able to generate emergent properties new to the network itself. As the largest number of metastable states exists at a point near the bifurcation, the brain can access the largest repertoire of neuronal processing in a flexible way and from distributed processing regions. At this point, network dynamics is weak attraction, close but not equal to neutral. Figuratively, the brain at a near-critical state is metastable but maintains a “sweet spot” of coherent complexity and emergence.

nLFPs, neural avalanches, and coherence potentials are characterized by a power law distribution in size and a branching parameter of 1, in which one nLFP, on average, begets only one other nLFP in the next time step. With branching parameter of 1, the number of repeating avalanches patterns is maximized into effective functional connectivity. This is suggestive of a balanced propagation of activity from one active group of neurons to another in a cascade within and across brain networks operating at near-criticality. Spatiotemporal patterns of avalanches are stable beyond chance and recur over several hours as predictable successive frames of nLFPs to which the brain returns repeatedly. Experimental work has shown that activity in living neural networks can propagate as a critical branching process that revisits many metastable states. Repeated neural avalanches capture the most metastable states of attractor networks. Oren Shriki and colleagues demonstrated neuronal avalanches in RSNs in humans, by using in vivo magnetoencephalography. Of note, criticality has been mainly investigated in neural activity during the resting state dynamics, only occasionally during task-induced changes and in the presence of external stimuli (summarized by Roxana Zeraati and colleagues).

To summarize, self-organized criticality as the general principle of the brain’s function is gaining ground over alternative proposals. Diverse evidence demonstrates a precise regulation between self-sustaining network up-states, nested theta-gamma oscillations, neuronal avalanches, and coherence potentials that follow power laws and a critical branching parameter of 1. These phenomena involve a balanced homeostatic GABA-mediated inhibition and glutamate–dopamine-mediated excitation in superficial layers of the cortex. Self-organized criticality maximizes the brain's information transmission, information storage, dynamic range, metastable states, and computational power.

Basics of Mind Self-Organization

The biologic self-organization of brain networks is believed to be foundational to the global emergent property of the mind. The brain and the mind progress through development in a highly parallel manner, where the maturation of brain networks generally coincides with the emergence of new functions of the mind. Such parallelism is suggestive of BrainMind coevolution. The mechanism of coevolution may involve activity-dependent codetermination, that is, reciprocal modifying influences between structure and function that constrain each to compatible paths of differentiation. Change in either is seen in terms of coevolution rather than as adaptation to separate and distinct environments. The fundamental principles of self-organization of open complex systems apply to both the brain and the mind: modularity, adaptability, robustness/resiliency, functional plasticity, progression from local to global, and dynamical growth. Important differences may reflect differing principles inherent to coevolving structure and function. In contrast to the highly modular organization of the brain, with hierarchical but independent processing units, the mind is modular as well, but in a more modest sense of domain-specificity and dissociability. Each mental faculty (cognition, emotion, motivation), each temperament trait, the sense of self and each of its subsystems (identity, character, morality) are domain-specific and phenomenologically discernible, each is autonoetic (personal happenings) and accessible to introspection.

Autonoesis, Noesis, Anoesis. These concepts are elaborated here to facilitate the discussion of the phenomena of self and consciousness. Neural traces of episodic memory are nonconscious first-order neural states of that memory encoded as "engram complexes." Episodic memories are experienced as autonoetic, that is, that the episode happened to the person remembering ("personal happening," autonoesis), only if, during the retrieval process, the sense of self, ownership, agency, and subjective temporality (the event happened in one's past), are integrated into retrieval. Such integration may be mediated by the works of DMN, the so-called "self-network," which is also coinvolved in memory processes, including recollection. Without autonoesis, episodic memory would merely be retrieved as a context-rich form of semantic memory. Semantic memory involves noesis ("knowing") of factual data not experienced as personal happening. Anoesis ("not knowing") refers to perceptual registration, neural representation, and behavioral responses, temporally and spatially bound to the present environment, that do not involve neocortical cognitive processing and consciousness, such as procedural skills ("not knowing of knowing").

The MindBody Debate. The centuries-old MindBody philosophical debate has evolved into three main approaches. First, the physical-identity approach (MindBody monism, materialism), pioneered by Christian von Wolff, holds that each type of mental state is identical to a particular type of neural state. Second, the functionalist approach (reviewed by Janet Levin in the Stanford Encyclopedia of Philosophy) holds that the identity of a mental state is to be determined by its causal relations to sensory stimulations, other mental states, and behavior. Functionalism leans toward MindBody monism but allows for mental faculties (cognition, emotion, motivation) to be understood as function-to-function emergence, ultimately embedded in the brain. Third, the dual-aspect monism is the view that the mental and the material are different aspects of the same underlying undivided ("neutral") reality in which they do not exist as separate domains. It is left up to the science of metaphysics to define what such neutral, dually-potent reality, shared by BrainMind, might be. The concept of the mind as a multilevel emergent property of interacting brain networks is incompatible with dual-aspect monism. Is the mind an emergent property of the brain (monism, functionalism) or not an emergent property of the brain at all (monism, dual-aspect monism)? In the latter case, the mind's self-organization is independent from the brain's self-organization, but they are inseparable from each other by means of shared ancestor. If functionalism is adopted, does the mind represents a weak, one-level emergence (only unexpected given the principles of the lower-level domain of the brain) or a strong, multilevel emergence (not deducible even in principle from the laws of lower-level domains). Here, we adopt the functionalist model of the mind as a strong, multiple-level emergence not deducible from the laws of lower-level brain networks, but ultimately embedded in these networks. The brain and the mind are best conceptualized as open, self-organizing, complex, and dynamical coevolving systems providing order to each other during their coadaptation to the same environment. If, as noted, intrinsic neonatal networks act as stable seed attractors for brain self-organization, the question is what attractors govern the self-organization of the mind?

Attractor Dynamics May Have Been Predicted by Object Relations Theorists. Melanie Klein's concept of "focal points" for ego integration[11] corresponds in principle to the concept of stable attractors governing evolution of open dynamical systems in biology. In "Notes on some schizoid mechanisms," M Klein introduced the notion of focal points for ego integration and associated it with the first good internal object, that is, the earliest phantasy of the gratifying "good breast" internalized under the hedonic quality of pleasure (sweet breast milk, an unconditioned pleasure). The frustrating "bad breast" (failure to feed) instantiates the unconditioned hedonic quality of displeasure (hunger, fear for survival) and is said to repel normal mind development. In the same paper, M Klein argues that the first good internal object acts as a "focal point in the ego … counteracts the processes of splitting and dispersal, makes for cohesiveness and integration, and is instrumental in building up of the ego" … "a precondition for normal development." The concepts of "comfort zones" (psychology) and "attractor dynamics" (discussed earlier) are helpful to link M Klein's concept of focal points in mind development with psychology and network science. The concept of comfort zones refers to the mental state of ease (low anxiety), underpinned by the brain's physiology of low stress (low cortisol, low epinephrine), in which the person feels they are in control. In Jaak Panksepp's affect theory, comfort zones are understood to be homeostatic set points for survival. Negative hedonic qualities (displeasure), motivation (dread), and behaviors (avoidance, fear, panic, defensive rage) mark deviations from the set point of comfort. These negative affects index danger zones that inform about circumstances that may endanger survival. Positive hedonic qualities (pleasure), motivations (desire), and behaviors (bonding, joy, lust) mark the return toward the set point of comfort and affirm survival. Once the comfort zone is reached, positive affects subside, and the mind idles, de-energized, engaged in familiar routines, not learning, and thus not developing. Through the prism of the dynamical system theory, comfort zones (psychology) and focal points (object relations theory) can be thought of as stable attractors that govern mind development. Between the inertia of comfort zones (de-energized mind) and the threatening anxiety of danger zones, the mind may operate in the zone of "optimal performance" for learning and development.[12] The zone of optimal performance may represent a near criticality regime of the mind, optimally self-organized to reach the level of complexity (e.g., creative thought) while hovering above comfort and anxiety. M Klein recognized the competing tendencies between integration and dispersal in mind development. Thus she might have conceived the mind as an initially dissipative system that evolves around focal stable attractor points but would dissipate without such governance. One can only speculate whether M Klein, based solely on field observations, introduced the concept of focal points for mind development and thus intuitively captured the essence of the forthcoming concepts of "attractors" and the dynamical systems theory.

Object Relations are the Basic Building Units of the Mind. Psychoanalysis postulates that "all psychological experience, from the most fleeting fantasy to the most stable structure, is organized by object relations … the basic units of all experience" (from Auchincloss and Samberg). First we ask the question "what is experience"?

EXPERIENCE, SUBJECTIVE EXPERIENCE, CONSCIOUSNESS. According to the Merriam-Webster dictionary, the term experience comes from Latin experiri (extracting, taking away something by trial or risk). Experience is thus a fundamental function to record nature and the specific form of it. In biology, experience is embedded in events of matching patterns, a fundamental mechanism of acquiring knowledge of all biologic life. Perception, from the Latin perceptio, meaning gathering or receiving, is the means by which an organism gathers experience.[13] Veridical sensory memory[14] is a plausible neural substrate of experience. As recently reviewed by Mario Vaneechoutte, all biologic life comes equipped with innate pattern recognition systems operating at different organismal levels, from enzymes that match their substrates, via gustatory receptors that match their tastant molecules, to attractor memory networks that match their selective basin of attraction. These phylogenetic molds predate real perception and thus enable an organism to predict external environment. Marsel Mesulam reports that unimodal sensory information (vision, hearing, taste, etc.) is processed through a collection of parallel processing networks, each network is modality selective or “unimodal,” that is, specialized for a particular type of information. The selective match between the unimodal input and the recognizing receptor triggers neuronal processing along five successive synaptic processing levels: lower-order (early) synaptic levels produce output that is streamed to later stages (higher-order synaptic levels), making information representation increasingly complex as the stages are passed through. For example, the primary visual cortex feeds information about edges, angles, light contrasts, etc., into higher-order processing cells that create more complex outputs that then converge and diverge onto yet higher-order processing cells (visual association zones), eventually resulting in unimodal visual representation of the entire object. Such within-system convergence creates stable and predictable integrations (emerging macro-property) that by nature cannot be a property of any basic element. Over time, each processing level memorizes conditional dependencies among unobservable causes of sensory input (“hidden states”) and the observed consequences (“sensory states”), and each encodes conditional expectations (Bayesian “degree of beliefs”) about hidden states that best explain sensory input at lower levels. Hidden states and causes are abstract probability estimates that the brain uses to top-down explain or predict actual sensations, akin to predicting the trajectory of a moving object in one’s visual field. For example, top-down connections with lower-level synaptic processing of visual input (“unimodal downstream”) allow higher-level feedback predictions of affordance of lower-level visual input (bars, edges, contrasts, etc.), thereby creating prediction error in the lower levels (mismatches between predicted and actually perceived). Only prediction error and new features (the data not predicted) are feedforwarded to the higher levels (“unimodal upstream”), where they update conditional expectations (novel data are represented as new combinations of familiar features). The updated top-down predictions reduce the occurrences of mismatches in the lower levels, thus enabling the brain to actively construct explanations for its sensory inputs. To summarize, perception is pattern expectation, a product of memory encoded within sensory networks that dictates one's interpretation of sensory data. Up to the point of heteromodal association cortices and posterior insula, sensory processing conveys veridical information about the organism's internal condition, position/movement, and external stimuli, that is, it does include the perceiver's learned pattern expectations, but not the perceiver's emotion and/or motivation. Such veridical reporting by the senses (sensory memory) is what should be called "experience" (*subjective* experience is discussed below). Sensory memory accounts for the brain's ability to retain impressions of sensory information for a few 100's of milliseconds after the original stimulus has ceased. Sensory memory, although vast, cannot hold all aspects of the internal or external environment but can be conceived as a snapshot of veridical information in the unit of time (milliseconds) and allocated to the current event.

The phrase *subjective experience* (a "subject" can be an animal or a human) means that veridical sensory information is top-down edited by personal salience attributed to the incoming sensory perception. Such edited experience reflects the salience properties rather than the surface properties of sensory events. Marsel Mesulam provides a cogent account for such personal editing. The highest level of sensory-fugal synaptic processing occurs within several highly connected "convergence" or "transmodal" zones that lack specificity for any single modality of sensory input, but receive afferents from lower-level unimodal areas and from other transmodal regions ("multimodal convergence"). These transmodal areas include prefrontal cortices ("anterior association area"), mid-temporal cortex including the parahippocampal gyrus and amygdala ("limbic association area") and the junction of occipital, temporal, and parietal lobes including the insula ("posterior association area"), among others. Most convergence zones, but especially the medial and lateral regions of PFC and insula, downstream modulate lower synaptic levels, providing attentional, motivational, historical, and emotional editing of veridical sensory stimuli. Such editing creates a subjective sensory reality rich in salience for each individual, that is, subjective experience.

Consciousness is commonly defined as "subjective experience." The term consciousness comes from Latin "cum" and "sciere,"–"to know about," implies some reflective properties of consciousness. Subjective and personal are synonyms, but neither corresponds to consciousness, the experience about experience, to know about experience. Note that the subjectivity of experience, as described above, means one's unique motivation, attentional priorities, and memory that bias perception in the presence of stimuli. Phenomenal properties of experience, “for there being something that it is like for one … to be in a particular state of having experience” …. e.g., seeing red, require introspection either by passive noticing (“thought-like”, as J LeDoux and R Brown refer to this reflective noticing) or by active scrutiny (deliberate reflection of one’s phenomenal experience), both can occur in the absence of stimuli. Introspection is said to be mediated by attention, working memory, and metacognition. Phenomenal consciousness is noetic (semantic, factual) unless information about the self (one’s self-schema, autobiographical memory, identity, etc.) is included in the retrieval when it becomes “self-aware” (autonoetic or personal happening). Consciousness is thus more accurately conceptualized as one’s passive or deliberate introspection of one’s experience (summarized by Mario Vaneechoutte). Symbolic language, a cognitive function that frames and organizes thought, is pivotal for both passive and deliberate introspection. Higher animals certainly may have subjective (“personal”) experience as defined above. Such experience may be represented in working memory and may guide behavioral priority, but is not accessible to introspective reflection mediated by language. Whether animals are capable of passive noticing of … being in the state of particular sensory experience is actively debated but, in our opinion, doubtful. Even passive noticing involves thought-like reflection as noted, which must involve some form of symbolism or language. Subjective experience that lacks reflection corresponds to awareness, common to all higher animals (see section on Nonconscious vs. Unconscious. Alertness vs. Awareness). In other words, animal awareness is not phenomenally conscious as introspectively accessed qualia. Rather, memorized hedonic qualities (liking vs. disliking), motivations (desire vs. dread), and incentive salience (wanting vs. not wanting) may give rise to affects and reward predictions that bias the animal, say a dog, toward selective foods in the here-and-now awareness.

Going back to the psychoanalytic postulate that object relations represent as the basic units of all experience. Dynamical system theory, cognitive- and systems-neuroscience of schema-forming, and expansion[15] provide a potential mechanism for this theorizing. Early object relations engage peak unconditioned hedonic qualities (displeasure, pleasure), rudimentary motivation (dread, desire), and rudimentary cognition (primitive imagery of the self and the world of objects). Axel Guskjolen and colleagues report that, in infant mice, episodic memories are forgotten as specific events ("infantile amnesia") but are not neurally erased and may be activated in adulthood by optogenetic stimulation of tagged neurons combined with natural re-exposure. These forgotten but neurally preserved cognitive-affective engrams complexes may serve as primordial fixed-point implicit attractor memories to interpret one's survival status and helpfulness of others in a self-reinforcing fashion. Andrzej Nowak and colleagues devise computational models of personality self-organization that support this hypothesis. In such models, fixed-point attractors consisting of cognitive-affective loops (correspondent to object relations) govern personality development. Most brainstem-limbic-medial prefrontal cortex (mPFC) regions involved in early object relations are integrated as nodes into the general networks of cognition–GNC, thus presenting hedonic quality and motivation (liking/desire, disliking/dread) for vertical integration into the infrastructure of cognition. With this in mind, the primordial schema of early object relations may indeed be a *bona fide* basic unit of the mind that governs the self-organization of emotion, cognition, and motivation in a self-reinforcing fashion. This primordial fixed-point attractor is indivisible to subsets that are attractors themselves and has no disjoint basins of attraction. Figuratively early object relations may represent the nonconscious operating system, the BIOS of the mind.

NONCONSCIOUS VERSUS UNCONSCIOUS. ALERTNESS VERSUS AWARENESS. The term "nonconscious" is used to refer to the contents of the mind not reaching active or passive awareness. The term "unconscious" is better suited for medical syndromes where the basal level of neuronal activity for perceptual registration and responses to sensory input is compromised. Brain trauma, for example, may render a person unconscious, that is may result in loss of alertness, interruption in awareness of oneself and one's surroundings, and unresponsiveness to external stimuli. Brain activity can be "nonconscious" (the brain is still online, but mental processes take place outside awareness), while an "unconscious" brain is offline, traumatized. Alertness denotes the basal level of neuronal activity to register perceptions and generate adequate responses to stimuli, for example, to respond to a question. Awareness derives from Old English–gewaer, English ware, and German gewahr, all meaning "aware," having knowledge of something, something one notices. Awareness thus arises from ultrashort sensory memory that is deemed salient by stimulus-driven (perceptual) attention and thus enters short-term memory (15 to 30 seconds). Short-term memory is a cognitive system that holds a limited amount (7 ± 2 items) of salient information, and can be driven by external stimuli or internal goals. Recent work suggests that short-term memory may be a function of phase locking by attention rather than of retention by an active memory trace. Mario Vaneechouette summarizes that awareness has evolved as a crucial evolutionary advantage that filters the unimportant out of all experiencing in sensory memory. Such filtering enables an organism to be aware of only novel or salient sensory information in the here-and-now. The filtering process is influenced by phylogenetically given molds (e.g., attractiveness of sexual partners), age (e.g., growing up changes priorities), previous experiences (e.g., habituation, education), other current experiences (e.g., the current state of internal motivation, like being hungry or not) and the interaction among all of these; the above “decides” which pattern is to be perceived or recognized, which experience is to become aware experience. It follows that, for animals, nonreflective awareness may be the main form of subjective experience that involves memorized affects, motivations, and incentives in prior experiences of the same kind. The concept of awareness in psychology corresponds to the concept of short-term memory in neuroscience.

Mind Development and Differentiation. Nenad Svrakic and colleagues pioneered the concept that mind development proceeds as punctuated homeostasis, a dynamic process whereby periods of relative stability (maintained by homeostasis) are punctuated by more or less abrupt transitions (“optimization of fitness”) toward a higher level of organization and functioning (“development”). Development is both informed and energized by adaptive demands that provide order (new information) to an open system. The term “fitness,” as applied to the mind, reflects the closeness of the mind’s current organization (the one being optimized) to the optimal organization (which guides the optimization process) for the given set of adaptive demands. Stage-like punctuated transitions are frequent early in mind development whenever the existing organization of the mind becomes suboptimal to meet changing adaptive demands. Early adulthood marks the period when significant fitness-driven optimizations of the mind subside. Around this time, adaptation becomes more quantitative (facets of the same) than qualitative (novel). Notwithstanding, neither the brain nor the mind ever completely loose plasticity to learn from and adapt to change.

We submit that mind differentiation involves two inter-related paths, (i) development of mental faculties (cognition, emotion, and motivation) and (ii) development of temperament traits, the sense of self, and its subsystems (character, identity, morality). These two paths are discussed in turn.

Mental faculties (cognition, emotion, motivation) are thought to emerge from interactions of large-scale functional connectivity networks. As such, mental faculties roughly correspond to two-level emergence, as the lower-level emergent functional connectivity implements the details of the higher-level emergent mental faculty by a small number of rules. Jochen Fromm points out that such weak emergence is flexible, predictable in principle but not in every detail, and characterizes open systems with active entities.

Cognition (thought) refers to human capacities such as memory, language, problem-solving, and planning, among others. These faculties are believed to emerge from works of the GNC (defined by J LeDoux and R Brown). The GNC encompasses partially overlapping large-scale functional connectivity networks (and their composite subnetworks) that spread across medial and lateral parietal and frontal regions; these networks dynamically interact to produce various facets of cognition. They include, among others,

1. the lateral FPN or CEN[16] (anchored in dorsolateral prefrontal and posterior parietal cortices) engaged in working memory, top-down goal-driven executive function, and attentional control,
2. the SN (anchored in dorsal anterior cingulate and orbital frontoinsular cortices) is involved in detecting and filtering personally salient stimuli, switching between CEN and DMN, as well as in recruiting relevant functional networks,
3. the DMN, is thought to comprise an integrated system for self-related cognitive activity, including autobiographical, self-monitoring, as well as social cognition, such as the theory of mind of others and, we submit here,
4. the language network[17] (left and right temporal and left frontal cortex, with coactivity in the auditory cortex and hippocampus).

Emotion (meaning) is defined as any conscious mental state with high intensity and hedonic content–pleasure/displeasure. The neural substrate of emotion may involve, among others, the hedonic hot-spots in the brain, notably those in the nucleus accumbens (NAc)-ventral pallidum, the extended amygdala processing network (includes the central medial amygdala, sublenticular substantia innominata, and the bed nucleus of the stria terminalis), the salience- and the explicit memory network.

Intrinsic *motivation* (goals) (the undertaking of an activity for its inherent satisfying qualities) may involve multiple modules:

i. The NAc-ventral pallidum hedonic keyboard mechanism, discovered by Kent Berridge and colleagues, engenders motivations (desire, dread) based on the enhanced hedonic quality of sensory stimuli,

ii. the mesolimbic ventral tegmentum (VTA)–NAc dopaminergic reward pathway (incentive salience, "wanting/not wanting") and

iii. the orbitofrontal cortex (reward-based risky decision-making, valuation of reward, and expected reward). In addition, dorsolateral PFC has been shown to initiate motivated behavior by integrating and transmitting representations of reward to the mesolimbic and mesocortical dopamine systems. In many cases, intrinsic and extrinsic motivations[18] work in concert, as the brain assesses the cost/benefit ratio as to how much effort is needed to obtain the intrinsically wanted goal.

The mental faculties can be thought of as the "tools of the mind" that, each from a different angle and with variable representation in different types of experience, enable the mind to communicate with itself (metacognition), the inner world (introspection) and the outer world (prediction), and to change and/or adapt to both.

The second path of mind development involves the differentiation of temperament traits as well as the sense of self and its subsystems,[19] that is, identity, character, and morality.

Temperament is the most biologic subsystem of the mind. Dragan Svrakic and Mirjana Divac-Jovanovic at first conceived temperament as one of the subsystems of the self. However, this is unlikely because temperament traits, notably harm avoidance (HA), reward dependence (RD), and novelty seeking (NS) emerge around midinfancy, before the emergence of the sense of self. Notwithstanding, temperament does influence one's self-schema, as one's semantic knowledge about one's usual reactions to danger and unfamiliar situations. At the root of temperament are innate dispositions to affect reactivity (biology) which are interactively molded (classical and operant conditioning) into stable mental biases (psychology) toward social gratification, danger, and novelty. Temperament traits refer to behavior tendencies that show relative stability over lifetime. For example, a person's usual, *a priori* approach to unfamiliar situations is either stably cautious (HA), risky (RD), or exploratory (NS). Temperament corresponds to weak, one-level emergence from neural dispositions to affect reactivity.

The sense of self is likely engendered by the works of DMN, itself an emergent functional connectivity (see section "DMN could be the neural substrate of the self") and thus corresponds to two-level emergence, much like the mental faculties of cognition, emotion, and motivation. The *subsystems* of the self (identity, character, morality) can be understood as multiple-level emergence, the first two levels being functional connectivity and mental faculties involved in the sense of morality or character. The subsystems of the self can be understood as emerging coalitions of relevant cognition, emotion, and motivation self-organized into adaptive solutions to selective stimuli. For example, character emerges in adolescence as a self-organized, semistable solution for social demands for cooperation and tolerance, thus facilitates one's social adaptation. Each subsystem carries a specific adaptive function that contributes to the adaptive fitness of the self as a whole. For example, cooperativeness and acceptance of others are character traits that peak in teenagers, who are thus fitter to respond to the pressure for social adaptation.

In parallel with the sense of self emerges the subsystem of *identity*, a repository of autobiographical memories. The main adaptive function of identity may be to provide continuity to the sense of self that, despite numerous updates, has historical coherence and future direction at any phase in one's life. To this end, the subsystem of identity must align past and present self-images and personal goals into a coherent narrative, which usually requires some falsification of autobiographical memories. Such alignment, however, allows prospection, that is, generation and evaluation of mental representations of possible futures, an important adaptive capacity of the self/mind. Identity also encompasses one's consistent view of others, which reinforces the coherent narrative of the self. Complex interconnectedness between the self and autobiographical knowledge is called the self-memory system introduced by Martin Conway: different types of memory provide data that constrain the current working self, but in turn, memory is also reciprocally controlled for accessibility by the current working self. The question is whether identity represents a subsystem of the self or simply one of the associative, “glue” functions of one’s self.

C. Robert Cloninger and Dragan Svrakic report that the subsystem of *character* emerges around 8 to 9 years of age. Character traits, such as cooperativeness, helpfulness, and acceptance of others, are rooted in empathy, prosocial attitudes, and helping behaviors, observable in all group-living animals (bats, bees, wolves, etc.). Karen Wynn and colleagues documented such prosocial behaviors in human infants and toddlers, presumably those who have attained self- and object-constancy and secure attachment (i.e., those not preoccupied with self-related mentation). These emotional forerunners of character are later amended by internalized socio-cultural standards as cognitive facets of character, represent one’s social principles.

In parallel with character and with much overlap, the subsystem of *morality*[20] emerges, enabling moral judgments of our own behavior and that of others. Linguistic expressions that utilize such judgements include “ought, should, right, wrong, good, bad, fair, unfair,” etc. Like character, the subsystem of morality is anchored in the sense of empathy, fairness, justice, and reciprocity of early childhood. These emotional forerunners of morality are later amended by learned moral norms (cognitive component) to become internalized moral principles by mid-to-late–adolescence. Both character and morality maximize the fit between one’s self-centered needs and restrictive social norms,[21] enabling the individual to achieve “functional adaptation,” that is, to fit in the society with a healthy tone of self-interest. As a unique feature, internalized morality enables self-evaluation, the mind’s inner censor of right and wrong. Neurally, the DMN has been implicated in social and moral cognition. The SN may also be important to bring semaniticized knowledge of one’s character and moral traits into conscious working memory to inform decision-making in social and moral tasks. Recently, the CEN and SN were found to be associated with voluntary selection of choices (decisions), including choices that pivot around moral issues.

Adult character traits and moral values reflect voluntary choices that don’t always result in the most gainful outcomes for a person. As discussed, character and morality are piloted by innate prosocial attitudes, empathy, and a sense of fairness, which are then expanded by learning (instruction), and finally internalized as attractor memory that contains one's character and moral schemas. The latter refers to autobiographical memories of one's usual social or moral choices, one's knowledge about social and moral standards in the given culture, etc. Such schemas provide information to working memory where it is used to guide new choices but only to an extent. Namely, new social/moral challenges may harbor new elements that will either revise the pre-exiting character and moral schemas or establish a new attractor for that particular experience.

We understand the subsystems of the self as canonical rather than hierarchical, semistable rather stable. Functional coalitions of emotion, motivation, and cognition that coalesce into the subsystems of the self are not dynamical (constantly reconfiguring) but semistable (functionally adaptable). Adult character and morality are represented by stable principles that retain functional plasticity (a concept close to that of semistability). A semistable equilibrium is one with multiple but not unlimited solutions to stable attractors, here internalized social and moral norms. Such systems are thus capable of transforming by self-organizing around new attractors if sufficiently energized by changing demands. For example, a person's principled morality (a semistable subsystem within the self) may change into opportunism during a war (physical survival takes priority over moral purity), but not at work despite opportunities for a minor promotion that may require some "soft" immorality, for example, use flattery to win over the boss. Such minor promotion is a weaker attractor relative to the stronger attraction of maintaining one's moral integrity.

BASICS OF OBJECT RELATIONS THEORY

Object relations theory is a major modification of classical (Freudian) psychoanalysis. Central to Freudian psychoanalysis is the perpetual conflict between innate "drives," pleasure ("libido") seeking satisfaction ("pleasure principle"), and the restrictive reality of the environment ("reality principle"). Pleasure-seeking (said to be embedded in the "id") inevitably clashes with social restrictions, thus necessitating the development of a balancing structure (the ego) and a controlling structure (the superego). This dynamics is known as the "drive-structure" model, whereby innate drives both instigate the development of psychic structures and are regulated by those same psychic structures. By contrast, central to object relations theory is the innate need to bond ("object-seeking"), a survival instinct conserved across mammalian species. The bonding instinct puts the child–caretaker relationship at the center of mind development. M Klein observed that infants manifest an immediate deep bond with their mother and that these early experiences modulate mind development. M Klein's theorizing departed from the "drive-structure" doctrine of Freud. This departure marked the birth of object relations theory, albeit still rooted in the main postulates of Freudian psychoanalysis, the dual-drive theory (libido and aggression) and tripartite structure (id, ego, and superego). Over the years, a number of more or less radical revisions have molded the theory into its present form. Ronald Fairbairn observed that the innate need for an emotional bond with an object drives mind organization rather than pleasure-seeking as the goal itself. While working in an orphanage, Fairbairn described "schizoid factors in personality,"[22] a major contribution to object relations theory. He theorized that early "splits" between rejection and acceptance allow an orphaned child to repress the rejecting parent, thus opening the way to accept the loving caretaking by others. Such deep splits in the mind (e.g., feeling of artificiality, "plate-glass" feeling about own self or the world, etc.) are seen occasionally in normalcy under severe stress and regularly in schizoid phenomena (hysteria, personality disorder, among others). As pointed out by Stephen Mitchell, Fairbairn's "relational/structure model" posits that "the interpersonal and the intrapsychic realms create … each other." Heinz Hartmann and Edith Jacobson pioneered the concept of self, hitherto unaccounted for by M Klein and Fairbairn, who used the terms "mind," "ego," and "self" interchangeably. Incidentally, Susan Isaacs and Edith Jacobson hinted that episodic memory might be the neural mechanism of object relations, seconded in this overview. Heinz Hartmann, Edith Jacobson, and Otto Kernberg viewed libido and aggression as primary but not innate motivation systems; rather, they cogently argued, the two drives consolidate from the hedonic tone of early interactions, pleasurable affects (pleasure, lust) coalesce into libido, unpleasurable (fear, rage) into aggression. Kernberg further advanced the integration of object relations theory with ego-psychology, revolutionized the concept of personality disorder, and pioneered interest in the neurobiologic mechanisms of object relations. Perhaps most importantly, Fairbairn, Hartmann, and Jacobson reconceptualized object relations insofar that what the infant internalizes is not an image of the object, a position long held by M Klein, but rather the interactive experience with the object, a position now widely accepted in theory and practice. This shift provides the rationale to frame object relations through the prism of affect-enhanced episodic memory, the main point of this overview.

What Is an Object Relation?

The term "object relation" refers to the internalization of early interactive experiences with significant objects. "Significant objects" are those into whom one invests strong emotions, whose interactive impact has been internalized and continues to influence the mind from within, as pointed out by Elizabeth Spillius and colleagues. The first significant object is usually the mother or the primary caretaker. Subsequent significant objects can be anyone or anything into whom/which one invests strong emotions, human beings, systems of knowledge or beliefs (political, scientific, or religious), one's work, an idea, etc. An object relation consists of three mental components: a self-representation, an object representation, and a representation of the affect that characterized the interaction, pleasure, or displeasure.

Our focus at the moment is the most influential object relations, those acquired over the first 2 to 3 years of life, during the most vulnerable phases in BrainMind development. These early object relations can be conceived as primordial cognitive-affective loops (schemas) that serve as implicit templates that regulate subsequent mind development. The intensity of affects (hedonic quality–pleasure, and displeasure) and motivations (desire or dread) evoked by early caretaking are a function of the child's heritable affect-reactivity and/or the quality of caretaking. These earliest interactive episodes occur in the setting of absolute dependency upon a powerful object. In such a setting, interoceptive distress (hunger, pain, etc.) or comfort (satiety, affectionate touch, etc.) narrowly evoke a sense of either a threat to or affirmation of own survival. The experience is processed by the amygdala circuitry involved in "implicit emotional learning and memory," as described by Elizabeth Phelps and Joseph LeDoux. Object relation theory posits that primordial affect-enhanced episodic memories of infancy and toddlerhood are forgotten as specific episodes but are not neurally erased and are foundational to the ensuing BrainMind organization.

Objects in Psychoanalysis: Real, Internal, External. Contemporary interpersonal-, relational-, and many object relations theorists use the term “object” both to refer to a real person and to that real person as it is represented in one’s mind. In Kleinian psychoanalysis, “external object” refers to a mental representation of another person that is modified by projections and reintrojections of one’s own mental issues and thus may not accurately represent that person. The term “internal object” refers to a mental representation of another person that one invests with strong emotions (“significant objects”), whose impact has been internalized as implicit memory and continues to influence the mind from within. A complex interplay, so the theory goes, continues throughout life between internal objects and those in the real world as they are represented in the mind, a solipsistic but plausible view that external objects only exist in the self of the beholder.

Object Relations Through the Lens of Neuroscience and Dynamical Systems Theory

In rodent models of infantile amnesia, enhanced episodic memories are forgotten but are neurally preserved and retrievable in adult rodents, as discussed. In human infants, the rudimentary topology of the long-term memory network is likely in place at birth. Such seed immature network is able to neurally preserve the memory but not to render the memory declarative/explicit.[23] Moreover, the sense of self, which fortifies perception and episodic memory by self-referenced cognition, has yet to fully self-organize in young toddlers. Self-reference provides a form of associative glue for perception, memory, and decision-making and, through this, acts as a central mechanism in information processing, as Jie Sui and Glyn Humphreys cogently argue.

Consistent gratifying and frustrating episodic memories with caretakers, albeit implicit, are thought to semanticize into preconceptual schema-affect loops that, thereafter, top-down regulate the relevance and preferential assimilation of new experience. In the dynamical systems theory parlance, the forgotten memories of object relations may act as implicit attractor memories that set a dynamical rule for a personalized self-organization of brain’s networks. A suggested mechanism of such personalized brain organization is “metaplasticity,” meaning that outcome feedback may enhance synaptic plasticity in similar contexts. Such “plasticity of synaptic plasticity” ensures that one’s personal goals or rules are enabled in the appropriate context. The mechanism of metaplasticity may involve activity-dependent neural changes that modulate their own plasticity threshold; in turn, these threshold changes modify the ability of such primed neurons to express plasticity in subsequent novel but similar experiences. For example, in infants, chronic stress may affect the synaptic organization, dendritic lengths, and dendritic spine density in cortical networks, ultimately modifying cortical metaplasticity. Such changes can prime cortical networks to take an abnormal developmental trajectory in response to later experiences that might otherwise be benign. Inversely, positive interactive experiences may exert a metaplastic effect that ultimately enhances the plasticity of these networks to similar positive social experiences later in life.

THE PRENATAL SOMATOSENSORY ORIGINS OF THE MIND

Emotion and motivation likely arise from the dynamic interaction of large-scale functional networks involved in vertical integration. These networks integrate processing in evolutionarily older regions with more recent brain systems from the brainstem, via the limbic system, to multiple cortical regions. Brainstem and limbic processing engenders the fundamental hedonic aspects of affects, motivation, and incentive salience (pleasure/displeasure, dread/desire, wanting/not wanting) that serve as input for higher-level integrations by rostral cortical systems. Our goal here is not to parcel a few regions for their independent function. Rather, our point is that brainstem and limbic processing valuates the hedonic-motivational-incentive quality of sensory inputs before they get integrated into complex functional networks of emotion, motivation, and cognition. Such caudal valuation may be foundational for innate dispositions to affect reactivity, the most important modulator of BrainMind self-organization in neonates and young infants, including a major impact on early object relations.

The “Emotional” Brainstem[24] and Individual Differences in Affect Reactivity

Many, but not all, brainstem functions are “emotionless,” for example, conduction of proprioceptive and tactile information to the cortex and cerebellum. The “emotional brainstem”:

i. participates in fixed, affect-charged response patterns in the face of threat (fight, flight, freeze)[25] and

ii. contributes to the earliest hedonic valuation of sensory input as appetitive (“liking”) versus aversive (“disliking”) our focus here. Hedonic “hotspots” involve opioid, endocannabinoid, and orexin (but not dopamine) signaling. Per Kent Berridge and Morten Kringelbach, the hotspots are distributed across the brainstem, NAc shell, ventral pallidum, insula, and orbitofrontal cortex, forming a hedonic network that enhances the core liking of a stimulus. The basic sensory-motor circuitry for unconditioned hedonic preference (liking sweet, disliking bitter) operates in the brainstem, mainly guiding ingestion. The brainstem and the limbic system are primary hedonic enhancer in utero and at birth thus foundational for early affect reactivity which, in turn, influences BrainMind self-organization. Caroline Kelsey and colleagues report that, in neonates and 1-month-old infants, innate dispositions to either negative- or positive affect-reactivity contrastingly influence the self-organization of functional networks, especially those involved in attentional regulation and behavior control. Neonates and infants predisposed to negative affect-reactivity show *hyper*-connectivity in DMN (manifested as preoccupation with stimulus-independent self-referential thought and mind-wandering), *hypo*-connectivity in FPN (manifested as slow reorientation of attention and poor executive control of emotion/behavior), and *hyper*-connectivity in the interhemispheric homologous network (HIN) between frontal, temporal, and parietal lobes[26]; the HIN has been implicated in emotion regulation. In contrast, neonates and infants predisposed to positive affect-reactivity show *hypo*-connectivity in DMN (self-referential processes are in balance with attentiveness to the external environment) and *hyper*-connectivity in FPN (manifested as effective regulation of attention and executive control of emotion and behavior). To summarize, functional connectivity that regulates attention, emotion, and behavior in neonates and young infants systematically maps onto individual differences in affect-reactivity and predicts adult personality and social health and pathology. Prenatally established disposition to positive- or negative affect-reactivity may act, respectively, as a protective versus risk factor for mental health versus disorder. Positive affect-reactivity appears to facilitate functional connectivity that is balanced between internally driven mental processes and attentiveness to the external environment, as well as good executive regulation of emotion and behavior. This pattern of network balance may be protective against psychopathology. In contrast, chronic stress, either gestational and/or early postnatal, may shunt functional connectivity toward diverse psychopathologies characterized by a hypofunctional PFC and hyper-functional amygdala, for example, attentional and personality disorders, among many others.

The Neonate: How Intero-, Proprio-, and Exteroception Mediate BrainMind Development

Bodily sensations, exteroceptive (smell, taste, touch), interoceptive (hunger, pain, affectionate touch),[27] and proprioceptive (limb/body position and movement) inaugurate pre- and postnatal BrainMind development. Early bodily sensations engender the neonate's body schema (body from the outside), body homeostasis (body from the inside), and body ownership (body in the mind), each relevant for the sense of physical boundary with others and for the differentiation of the sense of self. These are discussed in turn.

Angelo Maravita and colleagues define *body schema* as a body-part-centered[28] representation of space in one's immediate proximity—"peripersonal space." Multisensory correlations between vision, touch, and proprioception relate the proprioception of limb/body position with the peripersonal space in which they act. Such body schema is updated by movement. Bimodal, visual-and-tactile neurons are believed to be involved in multisensory correlations. These bimodal neurons, characterized by spatial selectivity,[29] are found in the premotor cortex but also other cortical and subcortical regions. Simply put, body schema is a spatial map of body/limb position/configuration as a three-dimensional object moving in space with nearby objects. This spatial map of the body engages the premotor cortex (voluntary motor system), among other regions, which enables motoric navigating of movement around objects in pursuit of behavioral goals. Maria Filippetti and colleagues report that 5-month-old infants may have implicit body schema, a fundamental aspect of social perception, as indicated by differing temporal lobe activity during real-time versus delayed video display of self-initiated face and arm movement. Cortical regions activated in response to body-related contingencies are similar for infants and adults (temporal, parietal, inferior frontal regions, temporoparietal junction), much like other template networks that predate the arrival of their selective input. Incidentally, cortical sensory-motor body maps might be involved in the basic registration of self–other correspondences, which, in turn, facilitate the elementary sense of bidirectionality.

Body ownership (“this body is and will always be mine”) involves multisensory correlations similar to those involved in body schema, that is visual, tactile, and proprioceptive, but these two are different phenomena. To illustrate the difference, an amputee wearing a prosthesis maintains the body schema (movements in peripersonal space are carried out properly) but does not consider the prosthesis as own limb (no sense of body ownership). In contrast, Manos Tsakiris and Patrick Haggard report that a subject watching a rubber hand being touched synchronously with their own hand (the latter positioned to be out-of-sight) causes the rubber hand to be attributed to one’s own body to ‘‘feel like mine.’’ The positive motion of own out-of-sight hand toward the rubber hand (“proprioceptive recalibration”) was understood as a behavioral measure of the sense of rubber hand ownership. However, the multisensory events (visual, tactile) that caused the rubber hand to be attributed to one’s own body are distinct from the personal experience that the “hand is mine.” Such phenomenal incorporation of a rubber hand into one’s internal body model correlates with activity in the right mid-insula and the right frontal operculum. The sense of body ownership gives somatosensory signals a special phenomenal quality of one of being embedded in one’s body and thus is an important contributor to the development of the sense of self and the self-object boundary.

In contrast to the largely emotionless body schema and body ownership, *interoceptive sensation-feelings* (pain, heat, hunger, tickle, affectionate touch, etc.) naturally possess hedonic aspects (pleasant, unpleasant) and motivational goals (dread, desire) that indicate the body’s homeostatic needs.[30] In seminal papers by Bud Craig, interception is understood to represent the ongoing homeostasis of the entire body, not only of the internal organs (visceroception) but also of the skin and brain. The lamina I spinothalamocortical system is a homeostatic afferent pathway that conveys the physiologic status of all tissues. This interoceptive information is integrated by the posterior insula, while the anterior insula generates phenomenal sensation feelings (pain, hunger, coldness) that reflect the interoceptive state of the body coupled with hedonic and motivational aspects. The insular cortex is the first fetal cortical structure to differentiate with a key influence on the self-organization of other cortical regions and networks. In neonates, the anterior insula likely contributes to the phenomenal noticing of the “material me.”

Based on Margaret Mahler's observational analyses, an infant develops a sense of physical separateness from others at about 5 to 7 months of life, a major milestone. Such "self-object boundary" likely does not represent the conscious self, but rather may reflect a nonconscious or an extremely thin phenomenal sense of "material or physical I" that establishes bidirectionality in interactions. This phenomenal sense of self-object boundary attained around midinfancy is likely a converging common outcome of body ownership ("this is my body"), body schema ("agency"—recognition of initiating and controlling own volitional actions), and "material I" (the interoceptive feeling of the body). The evolving recognition memory and selective attention at this period likely aid the attainment.

PHANTASY: THE BRAIN'S PHYLOGENETIC CODE FOR PRIMITIVE MENTAL FORMS

The spelling of phantasy is intended to convey the fundamental difference with fantasy, although they may overlap to some extent (see section "Fantasy" below). M Klein argues that, from the very beginning, recursive cycles of projections and introjections govern mind development. These abstract terms M Klein calls "mechanisms" that mediate exports and imports of mental contents, but provides no elaboration on the mechanisms themselves. The first attempt to reify infantile projection and introjection is the concept of "primary phantasy" (Susan Isaacs) or, as Sigmund Freud called it, "primal phantasy," the term used here. Primal phantasy refers to phylogenetically inherited *a priori* knowledge about the world that predates real experience and is embedded in survival instincts. For example, the sensation-feeling of hunger (an unconditioned displeasure, threat to survival) triggers the survival instinct to forage and the impulse to devour (a survival response), aiming the hungry mammal at the nipple (the rooting instinct finds an appropriate object to feed). These neural processes are said to be mentally represented as phantasy of desire that aims survival instincts toward appropriate means.[31] M Klein and S Freud attributed such primal phantasies to the work of the nonconscious mind (the "id").[32] Nonconscious phantasy is preconceptual and fully sequestered from reality testing. It operates below the level of conscious fantasy, long before the development of symbolic language. The earliest phantasies mainly bind olfactory-gustatory sensations and oral impulses, such as sucking, devouring (incorporation), tearing, biting, etc. With time, phantasy becomes less connected with bodily sensations and more symbolic as the child becomes increasingly able to differentiate between real and phantasized objects. Nonetheless, primal phantasy continues to underlie all mental activity as the mental penumbra inherent to bodily events and survival instincts. Exploration of phantasy as it appears in dreams and transference is an important aspect of psychoanalysis and some psychodynamic approaches.

Fantasy

In contrast to the concept of phantasy, fantasy involves conscious wishful distortions of reality (colloquially “wishful thinking” or “day-dreaming”) that gratify impulses or solve one’s inner problems. Reality testing is intact but temporarily divested in working memory. As either a gratifying or defensive operation, fantasy may and frequently does contain elements of deeper, nonconscious primal phantasies. This may occur when primal phantasies of desire or destruction bleed into conscious wishes of revenge or unconditional love, flavoring them with rage, paranoia, bizarreness, idealization, adoration, etc. Such phantasy/fantasy blend is commonly referred to as “primitive fantasy,” a key compensatory mechanism in personality disorder. Here, reality testing is not lost (as in psychosis) and not temporarily divested (as in daydreaming), but subtly disturbed, especially in interpersonal relationships. Childhood phantasy and fantasy have a structuring influence on self- and object representations.

Science of Phantasy: Plausibility, Analogies, and Survival Circuits

Object relations theory understands phantasy as a phylogenetically inherited code that *predictably* amalgamates sensations, survival instincts, and impulses into archaic motivated behaviors. This is a plausible hypothesis because randomness at early stages of mind development seems implausible from an evolutionary perspective. In the absence of direct evidence, Thomas Ogden reviews examples of “psychological deep structures” that support the concept of phantasy by analogy. Recent work on attractor networks that are created by intrinsic neuronal activity and predate experience corresponds to the concept of psychological deep structures. One robust example is the “innate capacity for language,” universal for all humans, as the pre-existing brain system that governs the acquisition of language from speech input without any knowledge of syntax. This famous thesis of Noam Chomsky has been substantiated by the discovery of an immature “template” language network in neonates two days after birth. Clearly, this immature network predates significant speech input but is trained by speech input to self-organize toward the syntactic and semantic architecture of adult language. Survival circuits are another example of preassembled deep structures in the brain, operational at birth, ready for selective input to activate survival instincts. At this early age, survival instincts are mainly related to fluid- and nutrition balance and thermoregulation. Some functions are automatic reflexes (e.g., shivering when cold), while some involve motivated instincts aimed at a goal (rooting for a nipple when hungry). Survival circuits are said to unify affects, motivation, reinforcement, and arousal into survival functions (survival instincts) that allow organisms to survive and thrive. Such amalgamate of sensation, affects, motivation, reinforcement, arousal, and impulses aimed at a goal inherent to survival circuits is exactly what psychoanalysis refers to as phantasy. It is thus plausible, if not likely, that human phantasy originates in survival circuits and may be in genetic continuity with archaic motivations of desire and destruction shared by all mammals.

Survival Circuits, Survival Functions, and Survival Instincts. J LeDoux introduced the concept of survival circuits and functions that are preserved across mammalian species, notably thermorelgulation, fluid balance, energy supply/nutrition, defense, and reproduction. In his words, focusing on survival functions and circuits allows phenomena related to emotion, motivation, reinforcement, and arousal to be treated as components of a unified process that unfolds when an organism faces a challenge or opportunity. These experiences are not only critical for survival but also for BrainMind development. The original five survival circuits and functions may need to be revised to include the survival instinct to bond, also conserved across mammalian species, and also of major importance in human BrainMind development. The neuroanatomy and manifestations of the instinct to bond with a parent have been outlined by Panksepp. Three out of six survival circuits concern homeostatic maintenance of the body, that is, thermoregulation, energy supply/nutrition, fluid balance), two concern self-preservation (bonding, defense), and one species-preservation (reproduction). The concept of survival functions is more inclusive than that of survival instincts. Survival functions range from very simple (e.g., automatic homeostatic adjustments—adipolysis in starvation, shivering in thermoregulation, etc.), via reflexive (e.g., automatic reflex-pattern responses—fight-flight-freeze), to complex behavior responses, conditioned (e.g., hiding from predators, playing dead) or preplanned (building shelters). Not all survival functions convey the essence of survival instincts, that is, the innate neurosensory dispositions molded into survival behavior by learning. As Konrad Lorenz demonstrated, the instinct to imprint on the mother in precocial birds exemplifies such duality, the innateness and the selective learning of survival instincts: newly hatched chicks are genetically predisposed to modulate the gross sensory input of the imprinting object within a critical period of hours after hatching, which is usually the mother in a stable ecosystem (an instinct to imprint on the mother, vital for survival). With this clarification, the term “survival instincts” is used in this overview.

RECURSIVE CYCLES OF PROJECTION AND INTROJECTION ORGANIZE THE INNER AND THE OUTER WORLD

In psychoanalysis, the term projection (literary, “expelling things in front”) usually refers to a defense mechanism whereby one attributes own thoughts or impulses to another person. The term “introjection” (literary, “taking things in”) is used for the process whereby one “takes in” real or phantasized aspects of the outside world as more or less distorted mental representations of that world.

The hedonic and motivational reality of earliest sensations, neurally processed by subcortical survival circuits, is mentally represented by phantasy, which is, in turn, recursively updated by actual experiences that approximate it. It follows that phantasy is presented for reality testing from the earliest days. For example, the newborn’s phantasy of desire for food (an expectation inherent to survival instincts, a projection) is aimed at the nipple (an appropriate object to feed) and thus brought into contact with realizations (actually experienced) that approximate to the desire (feeding, no feeding). The blend updates the phantasy (an introjection), thus resulting in revised phantasies (revised expectations) which then enter the cycle of projections and reintrojections in future encounters. Similar dynamics characterize the works of cognitive-affective schemas that expect experience based on previous experience (projection) and schema updates via memory (introjection).

The Origin of Inaugural Self- and Object Representations

Kleinian psychoanalysis holds that the first internal objects are inherent in biologic instincts[33] and mentally created by phantasy, the mental penumbra of the instincts. Phantasy, so the theory goes, amalgamates bodily sensations harboring innate hedonic and motivational qualities (pleasure/displeasure, desire/dread) with instinctual impulses (devour, destroy) aimed at controlling the causative sensation felt by the neonate as "presence," "something." This "presence," threatening or nurturing, is what is theorized to be the first object internalized by the mind. The "presence" soon begins to differentiate between the inside and the outside. Phantasies engendered by the brainstem's hedonic valuations of interoceptive sensations such as pleasure or displeasure and the related reflexes to swallow or expulse, may pioneer the first tracings of the inside as the precursors of the self-object boundary. With this in mind, earliest self-representations are embedded in the inaugural internal objects and arise from the neonate's nonconscious recognition of their own sensations. A self-representation thus shares affective valence and develops in tandem with object representation, with which it is initially confused as being its part (first noted by Edith Jacobson in 1964). We refer to these early sensations as "survival-level," as they automatically activate survival circuits and functions in the context of absolute dependency on caretakers.

Object Relations as Episodic Memory

Object relation theorists (Susan Isaacs) and ego-psychologists (Edith Jacobson) pioneered the idea that introjection, a mental reproduction of the outside world, must have its neural substrate in memory. Introjection allows BrainMind to record the history of interactions between the individual and the world of objects or, more precisely, to record the individual's interpretation of these interactions. As such, object relations play a major role in mental life, from early self- and object representations to one's acquisition of culture. Object relations are implicitly learned during survival-level experiences with caretakers and thus evoke intense hedonic qualities, pleasure, and displeasure. The biologic basis of such learning likely involves memory, here affect-enhanced episodic memory. Episodic memory can be conceived as the brain's information processing system by which self-related events like early survival-level episodes are neurally encoded as memory traces ("engrams") and operate inside the mind as a sustained, albeit implicit presence (see section, "Engrams or Neuronal Memory Traces" below). The incidence of experiences narrowly related to survival decreases with increasing mental and physical autonomy. Deborah Talmi and colleagues report that affect-enhanced episodic memories continue after childhood but become mostly concerned with experiences that either promote or frustrate one's current goals.

Engrams or Neural Memory Traces. An "engram" refers to lasting offline physical and/or chemical changes in a subpopulation of neurons elicited by neuronal plasticity that requires temporal precedence (cell A fires just before, not at the same time as cell B). Such Hebbian wiring of neurons (*engram cells*) encodes the newly formed memory associations. Susumu Tonegawa and colleagues define engram cells as populations of cells that constitute critical cellular components of a given engram (i) activated by the learning experience, (ii) physically or chemically modified by the learning experience, and (iii) reactivated by subsequent presentation of the stimuli that were present at learning, resulting in episodic memory retrieval. An "*engram ensemble*" refers to the collection of engram cells distributed across brain regions, with each ensemble supporting a component of multimodal memory. A given memory is supported by an "*engram complex,*" composed of dispersed engram ensembles that are interconnected by an *engram complex pathway.*

PARTIAL OBJECT RELATIONS

Theory

Neonates and infants suffer a great deal of unconditioned fear about their own survival, caused by the trauma of delivery and, post-delivery, by interoceptive sensations (pain, hunger, thirst, etc.). Although normally infrequent, such experiences are inevitable, except in theoretically perfect conditions. Per M Klein, distress is phantasized to represent threats of annihilation by a dangerous persecutor that fails to feed and protect. This early phantasy harbors paranoid fears for own survival and aggression-to-destroy. Aggression is projected not only into the malicious persecutor (of which the prototype is the frustrating breast) but also against the inside of their own body (felt to be a part of the breast). Positive sensations (pleasure) create the phantasy of desire but remain split from the phantasy of dread and destruction (“partial object relations”). Such splitting of contradictory experience, so the theory goes, protects the good internal object from being destroyed by internalized bad objects from within. In 1946, M Klein used the phrase “division between love and hate” to describe the unrealistic but protective split between contradictory experiences toward the same object, internalized and real, and toward their own self (which she refers to as “ego”). As a reminder, self-representations derive from object representations and are thus dichotomized into good and bad (gratified and frustrated), for they share the same hedonic valence with respective gratifying or frustrating object representations. M Klein understood early splitting to be a purely mental mechanism that protects the good internal object, the “focal point” for mind organization, as discussed. In extreme cases, M Klein argued that massive negative self- and object representations might lead to mind disintegration, as “early ego lacks cohesiveness and … a tendency toward integration alternates with a tendency toward disintegration, a falling into bits … especially during the first few months of life.” Mathematical models of personality development, drawing from dynamical systems theory, support this theorizing. Nenad Svrakic and colleagues report that, during the early formation of personality (an evolving dynamical system), the trajectory leading to chaos can be included, a nondesirable outcome, but nevertheless a result of development here failing to self-organize in the context of prevalent adversity and repelling attractor memories.

Neuroscience of Partial Object Relations: Dichotomization of Sensory Stimuli into Dreadful versus Desirable as the First Organizer of the Mind

Studies of contextual fear conditioning in rodents use recombinant DNA technology and optogenetics (described in the next section) to delineate cellular encoding and systems consolidation of affect-enhanced episodic memories. Susumu Tonegawa and colleagues demonstrate that the semantic-like content (“what happened”), contextual semantic-like hedonic valence (factual knowledge that an event was pleasant or unpleasant, not in itself an affect), and spatio-temporal context (“where,” “when”) are processed by the hippocampus. The amygdala process the actual affect evoked by the experience (contextual affect).

Recombinant DNA and Optogenetics Bridge the Gap between Cellular- and Systems-Neuroscience of Episodic Memory. Recombinant DNA refers to engineered DNA, combining sequences from various species and with various roles (e.g., transgenes, promoters, repressors, operators, among others). Manfred Gossen and Herman Bujard introduced "Tetracycline-Controlled Transcriptional Activator," a method of inducible gene expression where transcription is reversibly turned on and off in the respectively, absence or presence of doxycycline, aka "tet-OFF" system. In some current uses, "tet-OFF" is engineered to be transcriptionally active under the control of *c-fos* proto-oncogene promoter, an immediate early gene encoding a transcription factor, c-fos protein, induced by the recent activity of a neuron encoding a memory. The brain region of interest is transfected with an adenovirus-associated viral promoter, under the control of "tet-OFF," to initiate transcription of a transgene that encodes light-activated membrane ion-channels ("channelrhodopsins"). Light-sensitive ion channels open or close in response to light of different wavelengths ("optogenetics"), selectively in active neurons that initiated the cascade, that is, those expressing the c-fos protein. The tagged neurons can then be activated or silenced by light in different training paradigms for associative learning, allowing not only for precise identification of neurons involved in systems consolidation of contextual fear memory but also for selective retrieval of those memories. Retrograde tracing of monosynaptic neural pathways with, for example, cholera B toxin and in vivo imaging of calcium-dependent neuronal activity is combined with the above approaches for, respectively, pathway-level identification and cell-level visualization of neuronal processes in episodic memories.

The hippocampus forms an index of distributed neural events that were active during the actual episodic experience. The contextual affect evoked by the experience is memorized by the amygdala (reported by Takashi Kitamura and colleagues) by neurons genetically programmed to encode either rewarding or frustrating affects (reported by Joshua Kim and colleagues). As they are being encoded via synaptic plasticity, hippocampal engrams simultaneously engage mPFC to form its corresponding engram that indexes distributed component engrams of multimodal information into a coherent episodic memory. The mPFC engram is at first silent but matures[34] within several weeks to become the active episodic memory of the event, now "systems-consolidated."

Optogenetic and/or contextual activation of recent hippocampal episodic memory engrams (day 2 after memory formation) drives firing in the amygdala engrams bearing affect-memory of the episode (here fear, panic) and recreates freezing. The hippocampal originals then become silent (14 days after memory formation), in the sense that they show a decreased density of dendritic spines and can be reactivated only by optogenetic stimulation, not by natural clues; whether the hippocampus eventually loses the memory information is contested. The amygdala also provides input to the mPFC during the formation of mPFC engrams but, unlike the hippocampus, remains active in providing the affect-memory of fear during remote recall of the episodic event. The process of contextual fear memory is described here, but a similar process is likely for reward-enhanced memories, both types qualifying as episodic-like (see the summary). Amygdala activation is enhanced during both pleasant and unpleasant imagery but is correlated with rated emotional *arousal,*[35] thereby enhancing emotional learning and memory. As the hedonic quality of episodic events (pleasure, displeasure) is assigned by the NAc shell and ventral pallidum, an interesting possibility is that amygdala forms an index of distributed regions that coparticipate in an affect (arousal, hedonic quality, motoric reactions).

More recent reports by the Tonegawa group indicate that engram cell ensembles that encode individual components of an episodic memory are distributed brain-wide, across many connected regions, many of which are also functionally connected to hippocampal and amygdala engrams. Together, these inter-connected engram ensembles establish a “unified engram complex” that encodes the whole multimodal episode, including the contextual affects and semantic-like content (“what happened”). This work supports the unified engram complex for a single memory, suggested by Richard Semon in 1904. Remote memory recall induced by chemogenetic activation of multiple engram cell ensembles is most comparable to natural memory. However, some brain regions, particularly the mPFC, may be more active in remote retrieval of episodic memory, particularly the mPFC. Another function of mPFC is to learn associations between contexts, locations, events, and corresponding adaptive responses, particularly emotional responses, as David Euston and colleagues review. This function is consistent with the semantization of episodic memory into a self-referential cognitive-affective schema (discussed in more detail below) and may be the neural substrate of object relations. In other words, the mPFC appears to be involved in systems-consolidation, schematization, and remote retrieval of schematized autobiographical memory. Forgotten but neurally preserved early object relations that are schematized within the mPFC autobiographical memory process the here-and-now episodic, social, and affective information at a preconceptual level and implicitly steer one's actions and emotion based on survival-level experiences of early childhood.

Extrapolated to human infants, contradictory affects evoked by frustrating versus rewarding interactions are likely encoded by separate groups of amygdala neurons, each group dedicated to only one of the valences. As each group provides only one pole of affect-information to mPFC, self- and object representations derived from these episodes are dichotomized by contextual affects into either rewarding or frustrating.[36] Dichotomized experience (gratifying as well as frustrating) are both episodic memory, thus likely coexist within the same mPFC episodic memory network. This may explain the bistable, exclusively gratifying versus exclusively frustrating object relations until around midinfancy, in the context of an immature mPFC. Normally, positive experiences of care, safety, comfort are prevalent and frequent. Attractor memory for positive experience operates as a stable fixed point attractor for new learning (development), thereby expanding its basin of attraction. In contrast, negative experiences of endangerment and discomfort are normally less frequent. Such memories are unstable attractors that repel learning and development. The psychoanalytic concept of splitting between dichotomized self- and object representations (aka partial object relations) may therefore originate in the anatomy and function of the amygdala. Until about midinfancy, partial object relations are maintained by an immature mPFC. Psychoanalysis and neuroscience appear to describe the same phenomenon, a bistable state of the mPFC episodic memory networks, but through the prisms of phenomenology and neuroanatomy. As anatomically segregated groups of amygdala neurons are genetically programmed to irreversibly encode only one affective valence of episodic memories, the dichotomies of partial object relations cannot be integrated in situ, only by higher-order cognitive synthesis, our next point.

INTEGRATED OBJECT RELATIONS: COGNITIVE RECONCILIATION OF AFFECTIVE DICHOTOMIES

Two parallel developments mark the transition from partial (dichotomized) object relations to integrated (whole) object relations, a gradual process evolving from roughly midinfancy to 2- to 3-years of life[37] when self- and object-constancy are established. In the process, the infant/child becomes increasingly adept at:

i. integrating dichotomized experience into a realistic cognitive-affective frame whereby significant objects and own self are experienced as mostly gratifying and occasionally frustrating, and
ii. differentiating self- from object-mental representations; such mental boundary toward others allows the sense of self to be internally maintained and thus increasingly independent from the caretaker's affects and actions.

The phenomenology and mechanisms of whole object relations are discussed in more detail below.

Neuroscience of Whole Object Relations

Episodic-to-Semantic Memory Conversion: Generalization of Knowledge from Specific Experiences. Human thought would be inefficient, if not impossible, if episodic memories retained the level of contextual detail present at encoding. All episodic memory is first encoded by the hippocampus–amygdala memory system and within weeks systems-consolidated by mPFC, as discussed. Recurrent episodes get semanticized into a gist-like neocortical schema based on overlaps, statistical regularities, and abstractions. Such gist-like schemas govern one's expectations in new but similar episodes. The process is known as "schema-learning" or "episodic-to-semantic conversion." Schema-forming may involve *de novo* processing by mPFC, which would imply a new engram encoding the schema existing side-by-side with the episodic memory engram. Such engram has been hypothesized but is yet to be found. Alternatively, schema-forming may arise from "episodic-to-semantic conversion." The phrase, coined by Eric Kandel, refers to the natural tendency of consolidated episodic memories to decay over time, to lose spatio-temporal context-specificity ("when" and "where") but not the semantic-like information ("what happened") and the semantic-like valence information (a frustrating or a gratifying event, not itself an affect). Affects evoked in the event are memorized by the amygdala and are feed-forwarded to mPFC during the current event and in future similar events, as noted. Neurophysiology and epigenetics of schema-forming and self-reinforcing expansion in animals and humans have been described (summarized by Svrakic and Zorumski in 2021). Nikolaos Chrysanthidis and colleagues recently proposed that both Bayesian- and Hebbian synaptic plasticity might underlie the semantization of episodic memory.

Gist-like neocortical schemas can store information within an associative framework. Such schemas are independent of specific learning episodes. Their main function is to extract general knowledge from past experiences that can be applied across similar-but-novel situations. The ventral anterior cingulate, a part of mPFC, is connected with the amygdala, NAc, hypothalamus, hippocampus, and anterior insula, thus believed to be the interface between emotion and cognition. The mPFC is thought to harbor schemas related to the salience of emotion, motivational information, decision-making, and emotional behavior, much like implicit schemas suggested for early object relations.

Intrinsic Propensity of mPFC for Concept-Learning May Pilot the Inaugural Integration of Polarized Experience Around Mid-Infancy. M Klein observed the first signs of integrated object relations around midinfancy. She was criticized for attributing major cognitive capacities to young infants. However, intrinsic processing within mPFC, a convergence zone, is involved in simple rule and categorization memories as well as information on more complex schematic frameworks for episodic memories, as shown by Szu-Han Wang and colleagues. It is thus plausible that the inaugural integration of polarized experience by midinfancy entails a simple linear rule, a nonconscious or vaguely conscious concept that overarches contradictory episodes with the same object. This simple concept may be facilitated by the developing recognition memory, selective attention, and the capacity for object permanence, the latter recently demonstrated in 4-month-old infants by Renee Baillargeon and Julie DeVos, that is, 4 to 8 months earlier than proposed by Jean Piaget.

For normal BrainMind development, at the minimum, “good-enough” parenting[38] (helpful environment) and/or heritably moderate dispositions to negative affects (favorable genetics) are required. In such a setting, positive- and negative episodes with the same caretaker are encoded by the mPFC episodic memory network as dichotomized, occasional negative split off from common positive, each governed by its respective fixed point attractor subnetwork. Likely based on nonconscious Bayesian-type inference, an infant then develops a *degree of belief* in future interactive events. Under good-enough circumstances, the infant's implicit prediction is that interactions with the caretaker will be gratifying despite occasional frustrations. The gradual evolution from partial object relations of midinfancy into phenomenally conscious self- and object-constancy around 2 to 3 years of age corresponds to M Klein's concept of integrated object relations and depressive position. The transition from partial object relations (a bistable state of the mPFC memory network, unable to reconcile contradictions) to integrated object relations (a metastable state of the mPFC memory network, able to reconcile contradictions) allows BrainMind to handle complexity (next section), essential for further progression of mind development and social relations.

Metastability of the mPFC Episodic Memory Network May Be a Mechanism for Conscious Integration of Partial Object Relations in Toddlers. The unconditioned survival-level experiences of neonates and young infants occur in the context of helplessness and total dependency upon caregivers. In toddlers, care evolves toward more bilateral interactions with diversified and more frequent gratification and frustration by the same object (complexity). Bistable partial object relations (love *or* hate) are now adaptively suboptimal, as such exclusive dichotomization fails to meet increasing demands for relational complexity (love *and* hate). The mPFC episodic memory network processes positive and negative episodes that arrive from basins of attraction with so called "fractal basin boundaries." Namely, every point in the basin of positive experiences with the caretaker has pieces of another basin (negative experiences with the same caretaker) arbitrarily nearby; hence, small perturbations in the initial conditions can change the final fixed-point attractor state, positive or negative experiences ("complex memory network"). To handle relational complexity, mPFC must reconcile the dichotomies (integration) which in turn must remain quasi-independent to represent the reality of the actual interaction. A plausible mechanism for cognitive integration of bistable partial object relations may be metastability, a near-critical operating state of the mPFC episodic memory network. Metastability allows mPFC to reconcile the dichotomies by transitioning from bistability (dichotomization) to dynamic monostability (integration by hovering above both poles of dichotomized experiences that remain *quasi*-independent).

The mPFC memory networks may transition from bistability to metastability as a function of a control parameter, here the complexity of contradictory experiences which collide with each other. Right before the critical point, where the network would either devolve into chaotic (experience fragmented to pieces) or become rigid (fixated idealization or hatred), the network reaches a state where neither stable attractor states (positive memories) nor unstable attractor states (negative memories) exist. Instead, S Kelso argues, attraction for their respective sensory-affective inputs remains where these fixed-point attractors used to be. Through the lens of neurophysiology, oscillatory phase-synchrony (theta-gamma nests), governs the attractor dynamics between a fixed-point attractor-memory (usually theta oscillations) and its incoming sensory-affective input (modulated by gamma oscillations). Contradictory inputs (complexity) may break such in-phase attractors by oscillatory mismatches. The network self-organizes to operate in a metastable regime that enables it to process complexity. The remnants of fixed point memory-attractor broken by complexity are called ghost attractors, as these represent weakly-stable partially-synchronized network states. Such ghost attractors may nevertheless reconcile contradicting experiences that now "dwell" around the remnants by means of phase-gathering (quasi-synchrony) but still remain quasi-independent by phase-dispersing. Relative to phase-synchrony and asynchrony, phase gathering and dispersing are weaker associations between a fixed-point attractor and the incoming input. Such self-organized metastable regime may enable the mPFC to dynamically screen network tendencies as "stationary transients,"[39] a more adaptive alternative relative to being locked in dichotomized stationary states. S Kelso viewed metastable states of attractor networks as a pivot toward the emergence of stable collective states, as in integrated object relations and self- and object-constancy.

Successive visits to the remnants of fixed-point attractors are intrinsic to the time course of activation of the mPFC episodic memory network and do not require stochastic noise or additional sources of energy input. In other words, no energy barriers need to be crossed to visit self-organized metastable tendencies, aka second-order, energy-free state transitions. In contrast to physics and chemistry, cortical metastability does not follow the dynamics of lower- and higher-energy states but rather activity-dependent transient oscillatory synchronizations that persist for some period as a deviation from the resting state of an attractor network. Such transient unions readily self-configure and readily dissolve as internal and external demands change. As a result, cortical capacity to deal with complexity is increased: local specialized functions are kept quasi-independent (positive and negative self- and object images are independently realistic) but at the same time coordinated into global integration that provides a more realistic frame for the long-term reality of the relationship.

Jakub Vohryzek and colleagues have demonstrated ghost attractors in RSNs in humans. These authors characterize spontaneous brain activity in the form of trajectories within the state space, calculating the fractional occupancy and the dwell times of each state, as well as the transition probabilities between states. They also demonstrate that within-subject reliability is maximized when including the high-frequency components of the BOLD signal (>0.1 Hz), indicating the existence of individual fingerprints in dynamical patterns evolving at least as fast as the temporal resolution of acquisition. In conclusion, these results reinforce the mechanistic scenario that resting-state networks are the expression of erratic excursions from a baseline synchronous stable state into weakly-stable partially-synchronized states of ghost attractors. The process of continuous transactions between integrative and segregative tendencies by metastability is likely nonconscious, as is the case with most other cognitive processing, but the outcome, that is, the experience of self- and object-constancy, reaches phenomenal consciousness in toddlers. Brain structures and neural processes involved in dichotomization and integration of object relations during normal development are summarized in Figure 3.6–2.

Mental Separation of Self- from Object Representations. In parallel with the integration of contradictory self- and object representations, the sense of self (the “I”), initially embedded in object representations, achieves enough phenomenologic autonomy to emerge as an independent phenomenal experience. Such mental separation between internalized self- and object representations is a developmental milestone first recognizable around 18 to 24 months of age, an approximation based on Margaret Mahler. PFC, and particularly mPFC, major hubs in the default mode “self-network,” develop rapidly during the first 24 months of life, thereby providing neuro-cognitive mechanisms for such differentiation to emerge. The revised higher-order theory of emotional consciousness—HOTEC, devised by J LeDoux and R Brown, has been challenged mainly regarding the origin of phenomenal consciousness (see section, “First-order, higher-order, and revised higher-order theories of consciousness and emotion” below). K Berridge and colleagues define “affect” in a way that may inform this debate (see section “Affect, emotion, feeling” below). Nonetheless, the HOTEC provides a cogent account of the development of the sense of self by cognitive processing of engramized episodic memories, at the core of which is the primordial self-schema embedded in early object relations.

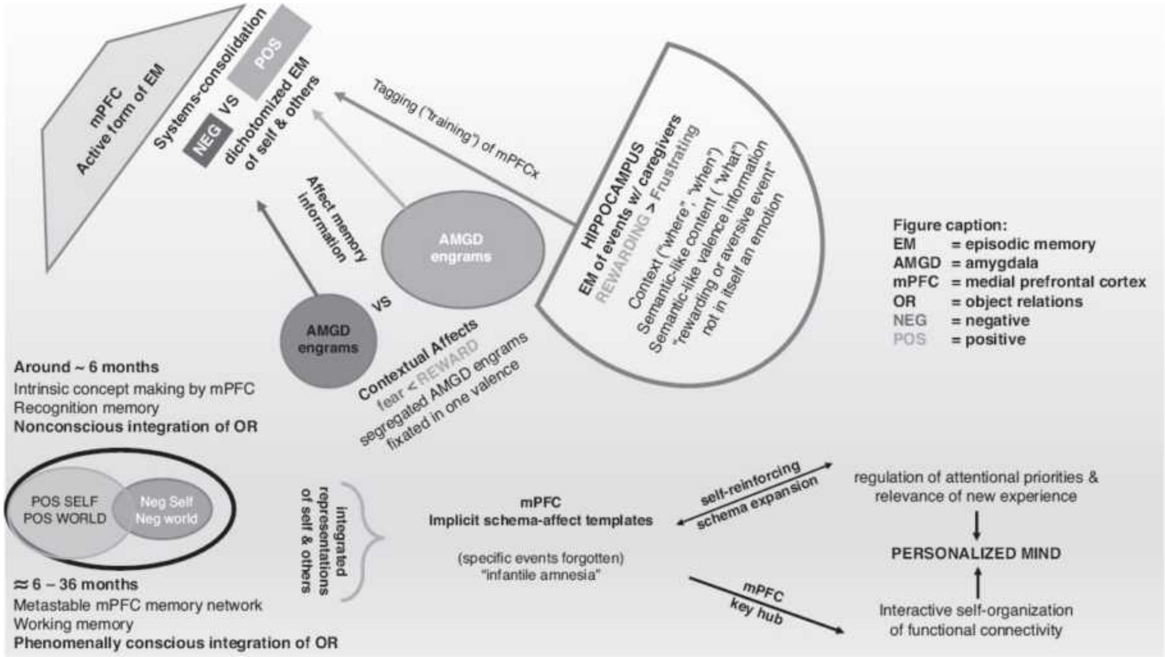
mPFC
Active form of EM
Systems-consolidation
NEG
VS
POS
dichotomized EM
of self & others
Tagging (“training”) of mPFCx
Affect memory
information
AMGD
engrams
VS
AMGD
engrams
Contextual Affects
fear < REWARD
segregated AMGD engrams
fixated in one valence
HIPPOCAMPUS
EM of events w/ caregivers
REWARDING > Frustrating
Context (“where”, “when”)
Semantic-like content (“what”)
Semantic-like valence information
“rewarding or aversive event”
not in itself an emotion
Figure caption:
EM = episodic memory
AMGD = amygdala
mPFC = medial prefrontal cortex
OR = object relations
NEG = negative
POS = positive
Around ~ 6 months
Intrinsic concept making by mPFC
Recognition memory
Nonconscious integration of OR
POS SELF
POS WORLD
Neg Self
Neg world
integrated
representations
of self & others
≈ 6 – 36 months
Metastable mPFC memory network
Working memory
Phenomenally conscious integration of OR
mPFC
Implicit schema-affect templates
(specific events forgotten)
“infantile amnesia”
self-reinforcing
schema expansion
mPFC
key hub
regulation of attentional priorities &
relevance of new experience
PERSONALIZED MIND
Interactive self-organization
of functional connectivity

FIGURE 3.6–2. Brain structures and neural processes involved in dichotomization and integration of object relations (normal development).

FIRST-ORDER, HIGHER-ORDER, AND REVISED HIGHER-ORDER THEORIES OF CONSCIOUSNESS AND EMOTION (FOT, HOT, AND HOTEC). The science of phenomenology studies the structure of experience from the first-person perspective. It does so by focusing on phenomena[40] or the appearances of things as they show up to one, such as colors, taste, memory, imagination, emotion, own thoughts, body awareness, etc. Sensory data about internal and external environments have an essential *first-person perspective*, for example, “what is it like for one” to have experience, to feel pain, to see red, etc., and *phenomenal quality of experience* (e.g., the taste of apples vs. pears, dental pain vs. wound pain). These subjective qualities of things or *qualia* are said to be secondary properties, not contained in objects being represented in experience but interpreted as such by one’s mind and then attributed to the world. J LeDoux and R Brown point out that such phenomenal consciousness involves introspection, either by passive noticing or active scrutiny of qualia. Phenomenal consciousness is not penetrable by others, only reportable by the experiencing person. First-order theories (FOT) of consciousness (Ned Block, Jaak Panksepp, among others) hold that phenomenally conscious experience arises from neural processing of sensory stimuli, for example, from auditory cortex processing sounds, without higher-order monitoring. FOT holds that such “first-order neural states” of stimuli are felt as phenomenally conscious, called “perceptual consciousness” (Ned Block, his term used here) and “core consciousness” (Jaak Panksepp). Perception is traditionally considered anoetic, “not knowing of knowing.” However, the FOT of perceptual consciousness refers to a state of consciousness that implicitly motivates behavior but itself evades conscious experience. In more recent revisions of FOT, Marie Vandekerckhove and Jaak Panksepp take a position that perceptual consciousness infact may not be accessible for introspective examination, may be truly nonconscious, but may become conscious and accessible to introspectiom by additional neocortical processing. Higher-order theories of consciousness (HOT) (discussed by J LeDoux and R Brown) hold that first-order, neural sensory representations alone are insufficient for phenomenal consciousness but need cognitive processing to reflect upon the first-order state in order to become conscious as a meta-representation of neural states. The aforementioned HOTEC holds that the first-order neural states of … tastes, visions, touches … are higher-order represented (HOR) in nonconscious working memory.[41] This first HOR provides a minimal inner awareness of one’s ongoing neural states, hence the name “nonconscious working memory”. Another HOR of this nonconscious HOR is needed in conscious working memory (HOR of HOR, HOROR) for a conscious phenomenal experience to emerge as passive noticing of “what is it like for one … to see red”, figuratively a conscious recognizer of a nonconscious recognizer. Such HOROR allows one to be phenomenally conscious (“what is it like for one”) about being in the first-order neural sensory state. Yet another HOR of HOROR in conscious working memory allows for reflective self-awareness (“I am the one who is in the state of noticing what is it like for me to see red”), a conscious recognizer of a conscious recognizer. Both HOROR and HOR of HOROR are phenomenal states, both are accessible to introspection, either by passive noticing or by active scrutiny.

AFFECT, EMOTION, FEELING. The terms “affect,” “emotion,” and “feeling” are frequently used interchangeably or, yet worse, as confusing blends, “affective feelings” or “emotional feelings.” Such terminologic discord arises from the disparate concepts of affects, emotions, and drives vis-à-vis their origin (innate or learned), biologic role (informants of survival or triggers of survival), and the nature of experience (nonconscious brain states or phenomenally conscious experiences). What appears to be a point of agreement is that affects are associated with biologically relevant sensory stimuli, those that disturb homeostasis or impact the well-being of an organism. Examples include interoceptive sensation-feelings (e.g., hunger, thirst, sensual touch, pain, etc.), exteroceptive sensations (e.g., unconditioned gustatory stimuli—sweet vs. bitter taste), and prepotently conditioned sensory stimuli (the sight of a snake, fear of heights, etc.). K Berridge and colleagues have devised a concept of affects that may be informative to the seemingly endless debate between the FOT, HOT, and HOTEC. They define “affect” as the core hedonic quality of biologically relevant sensations that have been conserved in vertebrate evolution based on their objective consequences for survival and reproduction. These biologically relevant sensations harbor an inherent hedonic quality (pleasure-liking vs. displeasure-disliking), motivational aspects (desire or dread), and incentive salience (wanting or not wanting). Survival circuits might have incorporated the unconditioned hedonic quality of sensations as intrinsic physiologic cues that contribute to incentive salience which, in turn, enables learning by the conditioning of complex survival behaviors. Causative hedonic “hotspots” and “coldspots” in NAc and ventral pallidum involve opiate-mediated enhancement (gain-of-function) of core liking and disliking associated with sensory experience.[42] Such enhanced hedonic qualities of sensations are objective affects (neural affects). In addition, a “keyboard” mechanism embedded in the anatomy and function of the medial shell of NAc attaches motivation to enhanced liking and disliking, that is, desire versus dread as innate “neural motivations.” This NAc hedonic keyboard can generate many mixtures of desire and dread: the most intensive pleasure/desire at the rostral medial shell becomes incrementally more aversive toward the caudal medial shell (dread/fearful escape attempts, distress calls, defensive bite attempts, etc.). It is plausible that all affects, both from the spectrum of pleasures to the spectrum of fears and rage, derive from the NAc-pallidal circuitry and the NAc keyboard mechanism. This circuitry, involved in enhanced core hedonic quality (enhanced liking vs. disliking) and the related enhanced motivations of desire versus dread, is independent of the incentive salience of wanting versus not wanting, the latter a separate mesolimbic dopamine-regulated system. Moreover, such "neural" (objective) hedonic/motivational/incentive qualities of sensory stimuli are HOR in the orbitorontal cortex[43] and anterior insula where they are processed into a phenomenally conscious experience of affects, now conscious emotion and conscious sensation-feelings. These findings seem to support the HOTEC: (i) hedonic quality of sensory experience is embedded in neuronal processing thereby creating a nonconscious bias in reactions without conscious emotion ("objective" or "neural" affects), (ii) the circuitry that processes neural and conscious aspects of affects are neurochemically distinct and (iii) neural affects are brain states that may become conscious emotions only after higher-order representation by the GNC in the orbitofrontal cortex and anterior insula. In support of the distinction between neural affects and conscious emotions, LeDoux and Brown review evidence that conscious emotion (e.g., fear) may not accompany behavioral and physiologic responses to threats regulated by subcortical survival circuits.

Following Bud Craig, the term "*feeling*" is used for sensation-feelings (or somatic feelings), for example, hunger, affectionate touch, temperature, pain, etc., that arise from interoception and are made phenomenally conscious by the anterior insula. The term "*mood* " is best suited for a general prevailing valence of the mind ("feeling good or feeling bad"), of low intensity, specificity, and provokability. As noted, interoception (hunger, temperature, pain, affectionate touch) inheres an affect (pleasure, displeasure), motivation (dread, desire), and incentive salience (wanting/not wanting). Interoception, the physiologic condition of one's own body, is thus a major determinant of feeling good or bad moods. Other determinants are "traditional" emotions, for example, sadness, happiness, fearfulness, unfairness, trust, etc. are conscious experiences that engage anterior insula and anterior cingulate, with input from ventral striatum, amygdala and hippocampus. Hence, these complex emotions combine cognitive and physiologic information, more so than the traditional sensation–feelings (hunger, pain, etc.) where physiologic correlates predominate. With this in mind, one “emotes” love and “feels” pain, although the latter verb can be used for both, due to the sensation–feeling component of traditional emotion. World affairs, weather, economy, and similar events also play a role in one’s mood via their relevance to one’s lifestyle, moral standards, immediate or future plans, etc. All of the above influence one’s subjective mood which in turn influences how people think, behave, make decisions, and interact with others.

The SN, embedding the stimulus-driven attention network, is anchored in the anterior insula and dorsal anterior cingulate, as noted. It is also connected with the amygdala (affect memory), nucleus accumbens, ventral tegmentum, orbitofrontal cortex (hedonic value, reward, incentive salience, emotional decisions) ventromedial, and dorsolateral PFC (executive function, working memory). Such interconnectedness helps integrate emotional, cognitive, and social information necessary for complex moral and social behaviors, among other functions.

COGNITIVE AND NETWORK NEUROSCIENCE OF THE EXPERIENCE OF SELF

A phenomenal experience of self predictably emerges around 2 to 3 years of age, as noted. It is believed to arise from one’s self-schema, a compilation of autobiographical memories and semantic knowledge about one’s own self, composed of:

i. explicit autobiographical events (“what has happened to one in own life”);
ii. semantic memory of one’s identity (“who am I” inside, plus name, date of birth, place of birth, etc.);
iii. semantic knowledge of one’s own personality traits (how one usually feels, acts, and thinks in particular situations); and
iv. one’s material “I” (body schema, body ownership, proprioception, interoceptive sensation-feelings).

These components of self-schema exist as nonconscious engram complexes and nonconscious body-sensing neural systems (posterior insula, cerebellum, sensory cortices) and are thus nonmental, first-order neural states. According to HOTEC, this neural self-schema becomes a conscious sense of self through a series of higher-order representations (self-HOR) and rerepresentations (HOR of self-HOR, or self-HOROR) by the GNC, first in nonconscious and then in conscious working memory.

Each new representation is a higher-order cognitive reflection of the lower-order representation. The first HOR of the neural self-schema (self-HOR) is represented in nonconscious working memory and remains nonconscious. Its rerepresentation in conscious working memory (self-HOROR) allows for a passively noticed phenomenally conscious state of the self to emerge. The phenomenal sense of self is accessible to introspection as the first-person experience of "for there to be something that it is like for one … to have the sense of self"… For a self-aware phenomenal self to emerge ("I am the one who is experiencing what is it like for me … to experience my selfhood"), yet another re-representation of the passively noticed phenomenal self, a HOR of self-HOROR is required. The GNC, via cognitive processes of attention, metacognition, and working memory, may mediate such successive higher-order (re)representations of first-order neural states. Such first-person phenomenality of self, now participating in all experience, instantiates a mental boundary with objects as they are represented in the mind, figuratively a mental "I" coexisting with "you." The latter is at first "physical you" (object possesses no own needs), but in toddlerhood also "mental you" (object possesses own needs, rudimentary theory of mind of others). The capacity for introspection has been demonstrated in 2- to 4-year-old-children and is thought to be a key prerequisite for the capacity to theorize the mind of others (reviewed by Christoper Gonzales and colleagues). Phenomenal experience is only reportable to but not penetrable by a third person, as noted. For example, beginning around 2 years of age, children are increasingly able to keep secrets and tell lies. Both abilities are suggestive of the child's understanding that their mind is impenetrable. First mental clashes with the world also emerge around this age, aka "the terrible twos," coinciding with the emergence of "autonomy and will" in Erik Erikson's terms. The neurophysiology of attention and working memory has been generally outlined, each reflecting complex synchronization patterns between slow (theta, beta) and fast oscillatory frequencies (alpha, gamma). The same cannot be said for metacognition, mainly due to the lack of a good operational definition and methodology to study.

Language as a Cognitive Mechanism of Reflection

The term reflection derives from Latin *reflectere*, "bending backward" and 1670s English "turning of the thought back upon past experiences or ideas." Reflection, a cognitive function, may have been made possible by the development of symbolic language. Jointly considering genetic, archaeologic, paleontologic, and other evidence, language first evolved around 50,000 to 150,000 years ago, during the Middle Stone Age, roughly contemporaneous with the speciation of Homo sapiens. Noam Chomsky cogently argues that natural human language is a computational mechanism central to syntax that processes information into infinite sets of expressions formed by a finite set of rules operating on a finite vocabulary. In this bio-linguistic perspective, a person’s language is a cognitive faculty that creates the world termed mental, while a result of “an organical structure such as the brain.” Daniela Perani and colleagues have discovered the seed language attractor network in 4-day-old newborns, much like other documented intrinsic seed attractor networks. From its initial state that predates experience, this immature language network learns, governed by attractor dynamics, to reach a stable state around puberty. Such internalized symbolic language appears to have evolved primarily as a system of thought, a cognitive mode to create and interpret thought (N Chomsky). In fact, language for thought (“inner speech”) might have evolved long before it was first used for communication (“outer speech”), and its design is unfit and too complex for communication. To summarize, language is primarily a cognitive faculty, only secondarily used to communicate. Animals can communicate information using vocal sounds, but do so only for linear, one-to-one type messages (predator approaching, etc.), while human symbolic language can assemble novel thoughts at any desired level of detail. This type of linear communication in animals may not indicate reflection (no symbolic language) but simply awareness, that is, presentation of salient sensory memory in short-term memory that keeps track of salient stimuli in the here-and-now environment. Such salient short-term memory may become working- and long-term memory by attentional potentiation. As reviewed by Peter Carruthers, primates and perhaps some other mammals like dogs, whales, etc. demonstrate task-dependent working memory in many respects homologous to human working memory, albeit with a more limited capacity of 2 versus 3 to 4 items manipulated at the time, and high susceptibility to interference. Without symbolic language, animals likely operate in the state of non-reflective awareness coupled with conditioned, task-dependent working memory to motivate and prioritize behaviors.

DMN Could Be the Neural Substrate of the Self. If there was a single network that could underlie the sense of self, the DMN is a prime candidate. The topography of DMN is present at birth, and around 2 years of age achieves adult state topology (reviewed by Amanda Hodel). The maturation of DMN roughly coincides with the emergence of the phenomenal experience of own self. The DMN is the largest RSFC network active at wakeful rest, engaging self-related cognition and internally driven attention independent of immediate perceptual input. The main DMN hubs are still a work in progress, but most reports include:

i. mPFC (implicated in social cognition related to self and others, autobiographical memory consolidation and reconsolidation during slow-wave sleep, and sense of agency, among others;
ii. posterior cingulate and precuneus have a pivotal role in DMN due to their involvement in autobiographical memory, self-referential processes, the phenomenal experience of positive hedonic states, the experience of agency;
iii. medial temporal lobe/hippocampus (implicated in episodic and semantic memory);
iv. angular gyrus (implicated in the retrieval of episodic and semantic memories with a contribution to one's experience of remembering);
v. the border between angular and middle temporal gyrus (implicated in bottom-up attention-to-memory processing); and
vi. thalamus and basal forebrain (respectively, a sensory traffic control center and incentive salience/reinforcement learning), both of central importance for the functioning of the DMN.

The brain relaxes to DMN during wakeful rest, disengaged from the external world but reflecting upon one's own autobiographical history, traits, motivations, engaged in social cognition and emotion (e.g., theory of mind of others, empathy, moral reasoning, social evaluations, etc.), and projecting oneself into the future. Together this is called "internal mentation," "a form of travel by mental simulation through our past, present, future, and the minds of others," as noted by Jessica Andrews. The DMN exhibits strong low-frequency (theta, delta) oscillations coherent during the resting state and is deactivated, although not completely, during attention-demanding goal-directed tasks. The DMN with dorsal mPFC as its key hub is considered pivotal for the sense of self and self-related emotional processing. mPFC is among brain regions having the highest baseline metabolic activity, which is attenuated but not absent across attention-demanding tasks, as reported by Debra Gusnard and colleagues. The incomplete deactivation of DMN and mPFC during attention-demanding cognition might be a mechanism of self-referential cognition, that is, of the continuous subpersonal representation of thought and self as a monitoring mechanism.[44] The DMN also appears to support the level of detail in experience during working-memory maintenance of active task states (reported by Mladen Sormaz and colleagues).

Internal Object or Objects: Singular or Plural? Deborah Talmi and colleagues demonstrate that emotionally enhanced episodic memories show increased binding with the encoding context, that is, semantic (what happened), temporal (when), spatial (where) it happened, and contextual affective valence ("how it felt"). The authors attribute such increased binding to the intrinsic neural organization of the encoding process, deficiencies in attention, executive function, and control of emotion, among other psychopathologies. As a plausible mechanism of enhancement, emotionally frustrating experiences activate peripheral stress hormones (epinephrine, cortisol), while enhanced positive experiences may involve a decreased autonomic arousal and processing in the precuneus. The latter appears to mediate the experience of happiness by integrating its emotional and cognitive components, as shown by Wataru Sato and colleagues. Enhanced episodic memories tend to cross-activate each other by semantic (natural) resemblance or by similar affects or by similarly intense affects. Such *en masse* activation of otherwise independent enhanced memories by a single affective or semantic clue might be the mechanism of what M Klein observed as a pervasive activation of a good or bad "internal object" (which also coactivates a pervasively good or bad experience of one's own self). It is plausible that M Klein used the singular to illustrate the cross-activated plurality of self- and object representations with the same affective valence. The above features of enhanced episodic memory may account for the psychoanalytic concepts of libido and aggression. Following Kernberg's concept of stratified motivation systems, the cross-activation of enhanced memories with the same valence may be the mechanism by which unconditioned hedonic qualities of pleasure and displeasure and motivations of dread and desire ("primary motivation system") coalesce as constituent components of libido and aggression ("supraordinate motivation system”), the latter encompassing all constructive versus destructive tendencies of a person.

Aaron Beck Cognitive Theory and Object Relations Theory Are Complementary in Some Instances. Both object relations theory and Beck cognitive theory concern schematized episodic experiences and autoreinforcing expansion patterns during infancy and toddlerhood (object relations) and childhood and adulthood (core beliefs). Beck cognitive theory considers core beliefs, that is, relatively advanced *valuative concepts* about self and others, to be the root source of autoreinforcing biases about self and others (reviewed by Judith Beck, Seth Disner, and colleagues). In contrast, object relations theory addresses *preconceptual* survival-level experiences of infants and toddlers, a developmental phase preceding that of valuative concepts about self and others. Adaptive (maladaptive) core beliefs include themes about self-worth (self-worthlessness), lovability (unlovability), and resourcefulness (helplessness). Maladaptive core beliefs arise from frustrating episodes with significant objects after survival needs are met, that is, after self- and object constancy have been established. Here, self- and object-constancy are eclipsed, but not wiped out in case of subsequent negative experience, and, in fact, may re-emerge after maladaptive core beliefs are restructured in cognitive-behavior therapy. Both adaptive and maladaptive core beliefs can be explicitly formulated if one journals one’s conscious automatic thoughts in challenging situations. In other words, in contrast to implicit object relations, core beliefs are accessible by introspection. Beck cognitive model does not account for the preconceptual sense of endangerment in infants and toddlers exposed to adversity. Self- and object-constancy and adaptive core beliefs rarely develop in this context (“identity diffusion”). Instead, a fantasized compensatory grandiose self (“everyone admires me”) may create the sense of safety and control over others. For normal development, however, Beck’s concept of core beliefs fits into the sequence of progressively more complex cognitive-affective templates for predicting, interpreting, and assimilating experience. Such templates are piloted by object relations (preconceptual and implicit expectation of safety and helpful others), evolve into self-valuative core beliefs (phenomenally conscious themes about self-worth, lovability, and resourcefulness in 3- to 4-year-old children), and finally into cognitive-affective schemas (a compilation of core-beliefs forms broad frameworks of schematized expectations about one’s self and the world). The term “cognitive schema” in Beck theory is somewhat misleading, as all schematized episodic memory is associated with respective affect information. For example, Seth Disner and colleagues note that cognitive theorists talk about *depressive* schemas created by adversity and characterized by negative self-referential beliefs. The limited effectiveness of cognitive-behavior therapy in personality disorders is discussed in Svrakic and Zorumski.

WORKING THROUGH AND BEYOND

Beyond the Integration of Dichotomies: Working through the Anxieties of the Depressive Position

Achievement of integrated object relations has been termed the “depressive position” by M Klein. The depressive position emerges when the dyadic object comes to be experienced by the child as a whole object, at times frustrating but predominantly rewarding. The adjective “depressive” relates to a number of new experiences inherent to integrated object relations that must be “worked through” (resolved) with good-enough parents, a gradual process from about 6 months to about 2 to 3 years of life and thereafter. In the process, paranoid fears about one’s own safety (paranoid–schizoid position, rigidity) are complemented by anxieties about the safety of the object, whose unavailability is now feared to be a loss rather than a persecutory attack (integrated object relations, depressive position). Fear of losing the object laden with guilt about one’s own retaliatory impulses that may destroy the object activate the drive for reparation in the child. Compassionate reassurance by a good-enough mother not only restores but also solidifies the child’s positive internal representations of both. Working through the anxieties of depressive position with good-enough parents transforms guilt-riddled anxieties about the safety of the object into a genuine concern for the welfare of the object, joy in giving and mutual exchanges, empathy, and reciprocity. The destructive envy of the object’s powers evolves into gratitude for the provided care and support, as noted by M Klein in “Envy and Gratitude.” Successful working through sadness, guilt, envy, jealousy, and oedipal issues, is a prerequisite for the capacity for mature love[45] and for reciprocal interpersonal relationships. Unresolved anxieties of depressive position may increase vulnerability to maladaptive core beliefs and schemas that, in turn, underlie depression and anxiety in children and adults, despite integrated object relations. Here, cognitive-behavior therapy is usually effective. The concept of reversible depressive and paranoid–schizoid position is considered by many to be the most important of M Klein’s contributions to understanding the human mind, as mature people may temporarily regress (become paranoid and revengeful) when faced with losses, but most can effectively recover with the help and empathy of close friends and loved ones.

Most Infants/Toddlers Have to Overcome the Headwind of Negative Affect States

Much like in other animals, the human brain appears wired to prioritize negative/aversive over positive/appetitive stimuli. In most cases, the potential impact of negative information outweighs the potential benefits of positive information. In general, emotionally enhanced episodic memories, rewarding *and* aversive, attract attention obligatorily and are prioritized for encoding and retrieval over nonemotional memory. At the tier of hedonic quality, however, aversive states take priority for encoding over rewarding states, as reported by Huang Po-Sheng and colleagues. Negative affect states, such as fear or rage, tend to reach higher intensity in comparison to positive counterparts. Infants and toddlers predisposed to heightened negative affect-reactivity may experience negative states at even higher peaks. To summarize, prenatal disposition to positive affect-reactivity expedites restoration of mental comfort (sense of safety) and neural comfort (lower autonomic arousal) in frustrating events. This dynamics promotes normal self-organization of brain networks and mental function. Heightened negative affect-reactivity (proneness to stress) tends to generate frequent and prolonged departures from comfort into mental and neural discomfort, thus disrupting BrainMind self-organization. All affects have associated ideas and fantasies, but this is especially true of intense survival fears in children predisposed to negative affect reactivity. Here, omnipotent primitive fantasies may safeguard a child from survival fears. Such fantasized safety (invincibility, omnipotence, machoism, grandiosity, etc.) might be the fundamental compensatory mechanism leading to adult personality disorder, as noted. In comparison to positive affects, which quickly subside upon return to comfort zone (a de-energized but safe mind), negative affect-states show inertia, likely mediated by hysteresis, a form of network memory. Hysteresis refers to the tendency of an attractor network, here the episodic memory network, to stay in phase with current sensory-affective input, even if the direction of the control parameter is reversed, for example, even if frustration is comforted by the caretaker. Hysteresis naturally appears when a network contains negative feedback that facilitates inertia rather than change, and most attractor memory networks involve such feedback. Hence, high energy input is required for a distressed infant/toddler to transition more rapidly to positive affect states. Simply put, positive affect states are harder to restore (a threshold effect that requires energy) than for negative states to remain (inertia, no energy required). This dynamics have direct implications for early parenting.

Auto-Regulated Self- and Object Constancy Promote Affect-Constancy

Once the new anxieties of the depressive position are resolved by working through with good-enough parents, a relatively stable, self-sustained good internal object is established in the mind, a successful outcome of the separation-individuation process, described by Margaret Mahler. Simultaneously, an autonomous self, self-sustained and internally-defined, capable of assimilating the objects' frustrating actions as exceptions, is also established. Both attainments set the stage for top-down stabilization of affects across contradictory interactions ("affect-constancy").

Summary

Object relations theory was inferred from observations of infants/children at play or while interacting with significant objects. For a century, it remained metapsychology without a neural basis. Modern neuroscience provides a framework that can help embed object relations theory into neural mechanisms. Although mechanistic studies have been done mainly in rodents, using the contextual fear conditioning paradigm, the processes involve homologous brain systems and fundamental neural functions shared with humans, making these studies potentially informative for human affect-enhanced episodic memory. Contextual fear conditioning may have its limitations to fully representing episodic memory. However, real-world emotional experiences likely engage both Pavlovian classical conditioning and episodic memory mechanisms. Recombinant DNA and optogenetic studies demonstrate that contextual fear conditioning may share neurobiologic systems involved in classical conditioning and episodic memory. The conditioned context (location in the cage) paired with unconditioned stimulus (foot shock) are initially encoded by the hippocampus, while the memory of fear is encoded by the amygdala which also drives avoidant behavior; the episode is systems-consolidated in mPFC. Remote memory expressed by the mPFC engram was “conditioned-context specific, suggesting that it is episodic-like” (Kitamura et al., 2017, p. 5, a position of Howard Eichenbaum as well). This research suggests that contextual fear conditioning may be conceived as a simple form of episodic memory.

The neuroscientific framework not only helps to validate object relations theory but also provides guidance on how to enhance psychodynamic therapy, a derivative of this theory (elaborated in Svrakic and Zorumski, 2021). Psychodynamic therapy is a treatment of choice for personality disorder, a common and severe but preventable and reversible form of BrainMind dysfunction. Perhaps most noteworthy is the question of whether we can take advantage of enhanced plasticity in the developing human brain for preventive purposes. Joan Luby and colleagues combine neuroimaging with income and parenting data to identify infants and toddlers at high risk for psychopathology. Informed early interventions that incorporate good-enough parenting can potentially steer developmental plasticity to dampen if not prevent BrainMind dysfunction that underlies personality disorder and deficiencies in attention, executive function, and control of emotion, among many other psychopathologies.

The plastic hippocampus–amygdala–mPFC circuitry is currently best described, but future studies will likely add other circuits relevant to this subject. Transdisciplinary integrations are heuristic models and thus always involve informed speculation, which can be tested through epistemic iterations as knowledge and technology advance. This overview offers testable directions in research that flow logically from our model:

i. to enhance healthy plasticity in critical neurocircuits to promote well-being,
ii. to activate this plasticity in children for early correction/prevention of multiple forms of psychopathology, and
iii. to enhance psychotherapy of children at risk for personality pathology and adults with personality disorder. The neuroscience of BrainMind development is a work in progress. This proposal is a primer that we hope may set a tone for further study and clarification.

ACKNOWLEDGMENT

We acknowledge authors whose material was used in this chapter, but who were not cited due to reference limits. All authors are listed for publications with up to three authors. For publications with more than three authors, only the first author is listed. Authors whose names are mentioned in the main text are not listed again in acknowledgments. Joëlle Proust, Ilyia R Prigogine, Maike D Hesse, Christian Boy, Oren Shriki, Mark D Morrissey, Angela D Evans, Kang Lee, Heidi M Gordon, Thomas D Lyon, Autumn L Arons, Tahereh S Zarghami, Stanley B Klein, Gregor Leicht, Tiago L Ribeiro, Robert A Renes, Neeltje E M van Haren, Henk Aarts, Stephanie R Miller, Emmanuelle Tognoli, Kathryn S Buss, Tecumseh W Fitch, Kristin A Campos, Stefan Troschütz, Charlotte Murphy, Hao-Ting Wang, Christina Andreou, Harsimrat Singh, Markus Bauer, Wojtek Chowanski, Jessica S Damoiseaux, Frederik Barkhof, Serge A R B Rombouts, Mario Vaneechoutte, Daniel J Amit, Edmund T Rolls, Benjamin A Seitzman, Sheena A Josselyn, Muriel Bruchhage, Richard D Lane, Daniel L Schacter, C.-Y Peter Chiu. Kevin N Ochsner, Jay R. Greenberg, Erin C Dunn, Eren Günseli, Mariam Aly, Peter J Marshall, Andrew N Meltzoff, Helder Bértolo, Eberhard Fuchs, Gabriele Flugge, Leon G Reijmers, Marc W Howard, Lars Nyberg, William J Matthews, Warren H Meck, Christina M Temes, Sébastien Marti, Stanislas Dehaene, Timothy J Teyler, Pascal Discenna, Andrew P Yonelinas, Charan Ranganath, Arne D Ekstrom, Brian J Wiltgen, Alcino J Silva, Yadin Dudai, Susan J Sara, Jonathan L C Lee, Karim Nader, Daniela Schiller, Daoyun Ji, Matthew A Wilson, Fredrik Bergström, Johan Eriksson, Anne L Wheeler, Martin A Conway, Rossella Guerini, John E Mendoza, Steven L Bressler, Pedro Nascimento Alves, Chris Foulon, Vyacheslav Karolis, Elisa Ciaramelli, Cheryl Grady, Brian Levine, Robert A Renes, Christopher R Michael, Janina Jochim, Elkan G Akyürek, Lucy Etherington, Diane Benoit, Marinus H van Ijzendoorn, Carlo Schuengel, Andmarian J Bakermans-Kranenburg, Kenneth N Levy, Joseph E Beeney, Karren Wynn, Paul Bloom, Amy F T Arnsten, Honi Sanders, Samuel J Gershman, Nasrin Esmaeilian, David P Bernstein, John Alcock, G B Huang, Y Q Chen, H A Babri, Annegret Krause-Utz, Raymond Salvador, Sandra Preißler, Giancarlo Dimaggio, Laurel J Gabard-Durnam, Daniela D Pollak, Martijn Meeter, Greg D Reynolds, Alexandra C Romano, Stefano Marini, Catherine Fassbender, Julie B Schweitzer, Ayline Maier, Sara E Keefer,

Gorica D Petrovich, Aryeh Hai Taub, Dylan F Gonzales, G. Gee, Nia Goulden, Farran Briggs, George R Mangun, W Martin Usrey, Gennady G. Knyazev, Vince D. Calhoun, Lucy Etherington, Peter Fonagy, Stephanie D Preston, Frans BM de Waal, Jie Sui, Glyn W Humphreys, Barbel Inhelder, Veikko Tähkä, Serena Chan, Heidi M. Bonnici, Lucy G. Cheke Deborah A.E. Green, Bryce A Mander Vikram Rao, Brandon Lu, Evgueniy V Lubenov, Matthew A Wilson, Benno Roozendaal, James L McGaugh, Peter Fransson, Guillaume Marrelec, May-Britt Moser, Edvard I Moser, Thomas W Soare, Yiwen Zhu. Rosaleena Mohanty, William A Sethares, Veena A Nair, Simon B Eickhoff, Saad Jbabdi, Svenja Caspers, R Matthew Hutchison, Thilo Womelsdorf, Elena A Allen, Abraham Z Snyder, Tyler Blazey, Andrew Kraft, Patrick Wright, Steven M Nelson, Alexander L Cohen, Jonathan D Power, Benjamin A Seitzman, Biyu J He, John M Zempel, James A Henderson, Elie Matar, Peter A Robinson, Sébastien Marti, Stanislas Dehaene, Pietro Morasso, Maura Casadio, Vishwanathan Mohan, Heidi M Gordon, Thomas D Lyon, Kang Lee, R Matthew Hutchison.

FURTHER READINGS

Auchincloss EL, Samberg E. *Psychoanalytic Terms and Concepts*. Yale University Press; 2012.

Berridge KC, Kringelbach ML. Pleasure systems in the brain. *Neuron*. 2015;86(3):646–664.

Chomsky AN. *Language and Mind Third Edition*. Cambridge University Press; 2006:190.

Craig ADB. How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci*. 2009;10:59–70.

Gusnard DA, Akbudak E, Shulman GL, Raichle ME. Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proc Natl Acad Sci U S A*. 2001;98(7):4259–4264.

Isaacs S. The nature and function of phantasy. *Int J Psychoanal*. 1948;29:73–97.

Kandel E, Dudai Y, Mayford MR. The molecular and systems biology of memory. *Cell*. 2014;157:163–186.

Kelso JAS. Multistability and metastability: understanding dynamic coordination in the brain. *Phil Trans R Soc B.* 2012;367:906–918.

Kernberg OF. *Borderline Conditions and Pathological Narcissism*. Jason Aronson; 1975.

Kernberg OF. Neurobiological correlates of object relations theory: the relationship between neurobiological and psychodynamic development. *Int Forum Psychoanal.* 2014;24:38–46.

Kim J, Pignatelli M, Xu S, Tonegawa S. Antagonistic negative and positive neurons of the basolateral amygdala. *Nat Neurosci.* 2016;19:636–1646.

Kitamura T, Ogawa SK, Roy DS, et al. Engrams and circuits crucial for systems consolidation of a memory. *Science*. 2017;356(6333):73–78.

Klein M. A contribution to the psychogenesis of manic-depressive states. *Int J Psychoanal*. 1935;16:145–174.

Klein M. Envy and gratitude. In *Envy and Gratitude and Other Works 1946–1963*. The Free Press; 1957: 176–235.

Klein M. Notes on some schizoid mechanisms. *Int J Psychoanal*. 1946;27:99–110.

Klein M, Riviere JH. *Love, Hate And Reparation*. Hogarth; 1937.

LeDoux JE, Brown R. A higher-order theory of emotional consciousness. *Proc Natl Acad Sci U S A*. 2017;114:E2016–E2025.

Luby JL, Baram TZ, Rogers C, Barch DM. Neurodevelopmental optimization after early life adversity: cross-species studies to elucidate sensitive periods and brain mechanisms to inform early intervention. *Trends Neurosci*. 2020;43(10):744–751.

Luby J, Belden A, Botteron K, et al. The effects of poverty on childhood brain development: the mediating effect of caregiving and stressful life events. *JAMA Pediatrics*. 2013;167:1135–1142.

Panksepp J. The basic emotional circuits of mammalian brains: do animals have affective lives? *Neurosci Biobehav Rev.* 2011;35:1791–1804.

Perani D, Saccuman MS, Friedeerici AD. Neural language networks at birth. *Proc Natl Acad Sci U S A* 2011;108(45):18566.

Redondo RL, Kim J, Arons AL, et al. Bidirectional switch of the valence associated with a hippocampal contextual memory engram. *Nature*. 2014;513:426–430.

Svrakic DM, Divac-Jovanovic M. *The Fragmented Personality*. Oxford University Press; 2019.

Svrakic N, Svrakic D, Cloninger CR. A general quantitative theory of personality development: fundamentals of a self-organizing psychobiological complex. *Dev Psychopathol*. 1996;8:247–272.

Svrakic DM, Zorumski CF. Neuroscience of object relations in health and disorder: a proposal for an integrative model. *Front Psychol*. 2021;12:583743.

Talmi D, Lohnas LJ, Daw ND. A retrieved context model of emotional modulation of memory. *Psychological Review*. 2019;126(4):455–485.

Tonegawa S, Morrissey MD, Kitamura T. The role of engram cells in the systems consolidation of memory. *Nat Rev Neurosci*. 2018;1:485–498.

Tse D, Takeuchi1 T, Kakeyama M, et al. Schema-dependent gene activation and memory encoding in neocortex. *Science.* 2011;333:891–895.

Wang SH, Tse D, Morris RG. Anterior cingulate cortex in schema assimilation and expression. *Learn Mem.* 2012;1:315–318.

Zorumski CF, Rubin E. *Psychiatry and Clinical Neuroscience: A primer*. Oxford University Press; 2011.

---

[1] Psychoanalytic- and psychodynamic psychotherapy are synonyms; both refer to psychotherapy that relies on an emotional relationship between the therapist and the patient to draw conclusions about the nature of the patient’s problems. This less-intensive, more personal approach has evolved from but then largely replaced classical psychoanalysis since mid-20th century.

[2] To our knowledge, Edith Jacobson pioneered the postulate that accumulated episodic "memory traces" could be the neural substrate of the "endopsychic representations of the object world and of the self."

[3] The term "primordial" is used for the episodic events of object relations that are forgotten (infantile amnesia), but remain neurally preserved, thus active as implicit templates that define one's motivational priorities and interpersonal expectations.

[4] The phrase "BrainMind" was coined by Jaak Panksepp to affirm the brain-mind monism aka isomorphism; here, we use the phrase wherever a subject matter applies to or affects both.

[5] The *degree of a node* in a network is the number of connections it has to other nodes while the *degree distribution* is the probability distribution of these degrees over the whole network.

[6] These regions include left and right temporal and left frontal cortex, as well as co-activity in the auditory cortex and hippocampus.

[7] In contrast to Hebbian (extrinsic) plasticity, homeostatic plasticity, aka metaplasticity is intrinsic and facilitates the subsequent plasticity of that synapse, e.g., long-term potentiation or depression.

[8] The term avalanche was chosen to associate with Per Bak's description of relaxation dynamics in the form of avalanches in a sand-pile that is continuously being added grains of sand.

[9] Such near-critical states of brain networks are said to be a stable attractor for external conditions.

[10] Multistability refers to multiple stable states of a network and depends on how quickly the system returns to its stable state following a perturbation.

[11] The term "ego" today would be referred to as the "mind," or more precisely as "personality," a uniquely configured mind for any given individual, as shown by Caterine Gratton and colleagues.

[12] The experience of the difference between positive hedonic states (approaching comfort) and negative hedonic states (departing comfort) may result in learning by conditioning. In infants and toddlers, temperament traits (reward dependence, harm avoidance, novelty seeking) are thought to result from such learning.

[13] The term experience is also used to denote one's knowledge or skills acquired over time by direct observation of or participation in events ("an experienced machinist").

[14] Gyorgi Buzsaki argues that perception is an "action-based process" an exploration initiated by the brain. Things in the world acquire coherence only through brain-initiated actions. Ultimately brain does not represent the world but rather "extracts" aspects that have become relevant to the organism by exploration and Bayesian-type inference. This corresponds to subjective experience, as described above. Up to the level of posterior insula and association sensory cortices, inside-out perception does not include other aspects inherent to perception, for example, emotion and motivation, which are added by convergence zones. With this stipulation, we use the term "veridical".

[15] Marlieke Van Kesteren and colleagues and Dorothy Tse and colleagues have made seminal insights into the neural mechanisms of schema-forming, reinforcement, and expansion in, respectively, humans and rodents.

[16] The CEN is generally the same as the lateral frontoparietal attention network, distinguished from the dorsal attention network involved in voluntary goal-orienting of attention.

[17] Noam Chomsky views symbolic language as a form of cognition that creates and interprets thought, thus enabling self-reflection, among other functions.

[18] Extrinsic motivation describes behaviors driven by external social or material rewards or punishments. Intrinsic and extrinsic motivation likely overlap in their neural substrates.

[19] The term "subsystem" rather than "structure" (Freud's terminology) is used because there are no structures in the mind, which is entirely a function.

[20] The terms ethical and moral are not synonymous. An ethical code doesn't have to be moral. Ethics refers to a set of rules for selected people to follow (e.g., American Bar Association, American Medical Association have created specific ethical codes for their respective fields). The term is also used for the science of morality.

[21] Kernberg (2014) speaks of character as an objective correspondent of identity or, figuratively, "who is me outside" (character) versus "who is me inside" (identity), with some overlap as both to an extent include mental representations of others and the society.

[22] Per Fairbairn some measure of splitting (a defense mechanism by dissociation) is invariably present at the deepest level of everyone's mind/ personality. He argued that "the basic position in the mind is invariably a schizoid position," even in normal individuals under conditions of extreme suffering, hardship, or deprivation. In 1935, M Klein postulated that the earliest position in mind development is "paranoid." The position is characterized by the neonate's aggression against the frustrating breast/mother, re-internalized as the "bad" persecuting internal object which stirs fears of retaliation from within. Influenced by Fairbairn, M Klein in her 1946 paper "Notes on some schizoid mechanisms," revised the paranoid position into "paranoid-schizoid," to account for persecutory fears co-existing with schizoid mechanisms ("splitting") early in mind development ("schism," Latin for division, split).

[23] Part of infantile forgetting is due to high levels of hippocampal neurogenesis for new memories which continually clears memories from the hippocampus by replacing synaptic connections in preexisting hippocampal memory circuits.

[24] Phrase coined by Anand Venkatraman and colleagues.

[25] This "emotional" motor system is distinct from nonemotinal corticospinal motor pathways passing through the brainstem to reach the musculature.

[26] Functional connectivity in neonates and 1-month old infants was studied by functional near-infrared spectroscopy (fNIRS). fNIRS is limited to monitoring activity from (superficial) cortical structures, this study could not fully measure activity from deeper cortical and subcortical regions (e.g., corpus callosum, a passage for most cross-hemispheric connections, among others). This drawback may explain the paradoxical finding of hyper-connectivity between homologous cross-hemispheric regions with infants with negative emotionality and poor emotion regulation.

[27] Traditionally, pain and temperature have been considered to be aspects of touch (exteroception) but are now considered interoception. In contrast to discriminative and crude touch and proprioception, pain and temperature reflect the physiologic state of a

tissue and inherently harbor affective and motivational aspects, much like other interoceptive stimuli. Also, sensual (affectionate) touch (grooming, cuddling, etc.) harbors an emotional and motivational aspects that mediate neurobiologic and emotional effects of parenting.

[28] The "body-schema" and "peripersonal space" are body-part centered. These are network phenomena involving different cortical and subcortical regions of the brain, each processing multisensory information in a reference frame appropriate to the body part concerning which it receives information, and with which responses are to be made.

[29] A bimodal, visual-tactile neuron with a tactile receptive field on one hand (say left) will typically increase its rate of firing as the visual stimulus approaches the left hand.

[30] The same sensory-affective-motivational multiplex characterizes exteroceptive sensations (pleasant vs. unpleasant, desired or dreaded, smell, taste, sounds, etc.).

[31] Phantasy therefore may be framed as "object-seeking" in accord with the Fairbairn's line of thought.

[32] Such sensation-to-phantasy transformation would today be understood to reflect nonconscious processing in subcortical survival circuits, particularly the opiate modulated NAc-ventral pallidum processing (hedonic quality liking/disliking, motivation of desire/dread) as well as the mesolimbic dopaminergic pathway where incentive salience of waning/not wanting is formed.

[33] Psychoanalytic literature inconsistently uses the terms instincts and drives, but the way the term is used in the context of phantasy corresponds to survival instincts or as J LeDoux calls them "survival circuits and functions."

[34] Both the creation of silent cortical engrams and their subsequent maturation require an intact hippocampus and may include, among other mechanisms, immediate early genes, epigenetic modifications, and sharp-wave ripple-mediated replay of activity of hippocampal engrams (reviewed in Tonegawa et al., 2018).

[35] NAc and mPFC activity were each correlated with the rated *pleasure* of imagery.

[36] As split-off contradictory "good" versus "bad" self- and object-representations appear to be genetically hardwired in the anatomy and function of the amygdala (Kitamura et al., 2017), the object is necessarily encoded ("introjected") as dyadic from birth, leaving open to question the Margaret Mahler's concept of full symbiosis (mental fusion with the object) and Fairbairn's concept of an internalized pre-ambivalent all-good object of the earliest oral phase. O Kernberg, in fact, holds that a sharp differentiation between self- and object-representations evolves from birth on, while it takes 2 to 3 years to achieve integration of their dichotomized aspects.

[37] An approximation, based on Margaret Mahler's separation-individuation timeline, likely varies in individual instances.

[38] Donald Winnicott coined the phrase "good-enough" to emphasize the importance of natural instincts of normal devoted parents and healthy families, in an attempt to counterbalance the potential idealization of parenting implicit in Klein's concepts of "all-good breast/mother." The phrase "bad enough" refers to neglectful or abusive caregivers and dysfunctional families, sufficiently stressful to cause BrainMind pathology, as in personality disorder. Occasional manageable frustrations are necessary for the child to develop self-reliance, as observed by Heinz Kohut).

[39] A phrase coined by Scott Kelso.

[40] From Greek, phainomenon, Latin phaenomenon, means “that which shows itself.”

[41] Nonconscious working memory is activity silent, occurs outside of focal attention, but is susceptible to implicit cognitive influences. Its mechanism is actively debated, but may involve intracellular calcium dynamics. This form of memory is understood to be a “stop” in memory retrieval where homeostatic optimizations of memory being retrieved may occur to maintain the mind’s current conditional equilibrium despite challenging reality. The psychonalaytic concept of nonconscious defense mechanism may be neurally explained by nonconscious working memory. Defense mechanisms are in fact implicit memories that develop through early interactions with significant objects for the purpose of regulating anxiety, emotional pain, anger, and self-esteem. These implicit memories intervene in nonconscious working memory to modify disturbing experiences before they enter conscious working memory.

[42] Opioid-stimulated coldspots suppress the “liking” and approach responses and increase dread and avoidant/aggressive responses.

[43] Orbitofrontal cortex (anatomically synonymous with ventromedial PFC—a component of the limbic system) is the rostral projection of the mesolimbic VTA - NAc - mPFC dopamine pathway. Orbitofrontal cortex harbors its own opiate-mediated hedonic hotspots, so it does not only process the reward value and expected reward value but also the subjective pleasantness/unpleasantness of stimuli. Orbitofrontal cortex is said to encode contingencies in a flexible manner, encode value, inferred value, inhibit responses, learning changes in contingency, emotional appraisal, etc. The anatomically distinct mesocortical dopaminergic pathway (VTA-mPFC) is involved in cognition and the regulation of executive functions (e.g., attention, working memory, inhibitory control, planning, etc.).

[44] Alternatively, thought and self may be inherently personal but arrived at automatically, apparently without requiring higher order inference in working memory (reviewed by Richard Brown and colleagues).

[45] Guilt and sadness are inherent parts of love and deepen it profoundly (M Klein and J Riviere, 1937).

4

Contribution of the Social Sciences

4.1 Neurocentrism: Implications for Addiction and the Courtroom

SALLY L. SATEL, M.D., AND STEPHEN J. MORSE, J.D., PH.D.

In the early 21st century, neuroscience has captured the public imagination like never before. Understanding the brain is helpful, perhaps essential, to developing treatments for devastating illnesses like schizophrenia and Parkinson disease. More abstract but no less compelling, the functioning of the brain is intimately tied to one's identity, memories, behavior, and aspirations. But the excitement to explore the brain has spawned a new fixation that one might call *neurocentrism*—the view that all human behavior can be best explained by looking solely or primarily at the brain.

In the view of some critics, neurocentrism poses a threat to psychiatry because it risks oversimplifying a number of complex issues. In its extreme form, it devalues the importance of psychological explanations and environmental factors, such as familial chaos and stressors, for understanding disturbances of mood, thought, and overt behavior. In turn, neurocentrism may pose a distraction from seeking effective and perhaps more enduring solutions, including psychotherapies, and may imply that pharmaceutical approaches should be the strongly preferred first-line interventions for addictions and other behavioral problems.

The prime impetus behind this enthusiasm is noninvasive functional neuroimaging techniques and the now-iconic, vibrant and arresting brain images that grace the science pages of the daily newspaper. Author Tom Wolfe was characteristically prescient when he wrote of fMRI in 1996, 5 years after its introduction, “Anyone who cares to get up early and catch a truly blinding twenty-first-century dawn will want to keep an eye on it.” Why the fixation? First, there is the very subject of the scans: the brain itself. More complex than any structure in the known cosmos, the brain is a masterwork of nature endowed with cognitive powers that far outstrip the capacity of any silicon machine built to emulate it. Containing roughly 85 billion brain cells, or neurons, each of which communicates with thousands of other neurons, this 3-lb universe cradled between the ears has more connections than there are stars in the Milky Way. How this enormous neural edifice gives rise to subjective feelings, a question often called the “hard problem” of consciousness, is one of the greatest mysteries of science and philosophy. Moreover, the brain is more temporally spatially proximal to human behavior than more distal environmental and interpersonal variables. Combine these factors with the fact that multicolored images—in this case, brain scans—can be powerful because vision is the most developed sense and occupies the largest share of cortical space.

There are good evolutionary reasons for this arrangement: The major threats to our primate ancestors were apprehended visually, as were their sources of food. Plausibly, the survival advantage of vision gave rise to our reflexive bias for believing that the world is as they perceive it to be, an error that psychologists and philosophers call naive realism. This misplaced faith in the trustworthiness of one’s perceptions is the wellspring of two of history’s most famously misguided theories: that the world is flat and that the sun revolves around the earth. For thousands of years, people trusted their raw impressions of the heavens. Yet, as Galileo understood all too well, the eyes can deceive. He wrote in his *Dialogues* of 1632 that the Copernican model of the heliocentric universe commits a “rape upon the senses”—it violates everything the eyes reveal.

Trusting the patterns on brain scans to reveal nuanced mental contents is a form of neurorealism, a term coined by University of Montreal researcher Eric Racine. A first cousin of naive realism, neurorealism denotes the fallacious but tempting propensity to regard brain images as inherently more “real” or valid than other types of behavioral data. As

Stanford neuroeconomist Paul Zak has described in his work on the neurobiology of trust, a brain scan “lets me embrace words like ‘morality’ or ‘love’ or ‘compassion’ in a non-squishy way. These are real things.” Or take the psychological impact of combat. A researcher at the Minneapolis VA tells TIME magazine that brain imaging shows that PTSD is “not a made-up disorder.” In both quotes, it is not clear that brain images truly provide novel information; such images are not required to conclude that love or PTSD is genuine. These are “diagnosed” behaviorally in the first instance.

As a tool for exploring the biology of the mind, neuroimaging has given neuroscience not merely a huge scientific boost but a strong cultural presence. As one scientist remarked, brain images are now “replacing Bohr’s planetary atom as the symbol of science.” With its implied promise of decoding the brain, it is easy to see why brain imaging would beguile almost anyone interested in pulling back the curtain on the mental lives of others: politicians hoping to manipulate voter attitudes, marketers tapping the brain to learn what consumers really want to buy, agents of the law seeking an infallible lie detector, addiction researchers trying to gauge the pull of temptations, psychologists and psychiatrists seeking the causes of mental illness, and defense attorneys fighting to prove that their clients lack malign intent or even responsibility (often mistakenly called “free will”). This tendency has produced a torrent of allegedly new fields, such as neuroeconomics, neuromarketing, neuroethics, and neurolaw, to name just a few. At the same time, this fascination is a double-edged sword, as it may lead eager audiences to accept dubious neuroscience claims without adequate scrutiny.

Some misapplications of neuroscience are merely amusing and may be essentially harmless. Take, for instance, the new trend of neuromanagement books, such as one entitled *Your Brain and Business: The Neuroscience of Great Leader.* The latter advises nervous CEOs “to be aware that anxiety centers in the brain connect to thinking centers, including the PFC [prefrontal cortex] and ACC [anterior cingulate cortex].” The fad has, perhaps not surprisingly, also infiltrated the parenting and education markets. Parents and teachers are easy marks for “brain gyms,” “brain-compatible education,” and “brain-based parenting,” not to mention dozens of other unsubstantiated techniques. Although these methods may not be dangerous *per se,* they may incur both direct financial costs and opportunity costs arising from a failure to seek out more efficacious interventions. Once again, unwarranted neurocentrism may blind us to more efficacious interventions.

For the most part, these enterprises merely dress up or repackage good advice with neuroscientific findings that add little or nothing to the overall program. As one cognitive psychologist quipped, “Unable to persuade others about your viewpoint? Take a Neuro-Prefix—influence grows or your money back.” But reading too much into brain scans can be a problem when real-world concerns hang in the balance. Consider the law. When a person commits a crime, who is at fault: the perpetrator or their brain? This is a false choice. If biology teaches anything, it is that “my brain” versus “me” is a false distinction. Still, if biologic roots can be identified—and better yet, captured on a brain scan as juicy blotches of color—it is too easy for nonprofessionals to assume that the behavior under scrutiny must be “biologic” and therefore “hardwired,” that is, involuntary or uncontrollable. Criminal defense lawyers, not surprisingly, are drawing increasingly on brain images supposedly showing a biologic defect that “made” their clients commit murder. At present, such cases are relatively few, but the numbers will increase.

Looking to the future, some neuroscientists envision a dramatic transformation of criminal law. They anticipate a time when scientists will be able to attach a detailed biologic explanation to many types of bad behavior and eventually come to think about the decision making that gave rise to such behavior in the same way one thinks about any physical process, such as diabetes or lung disease. It is plausible that, as highly detailed mechanical accounts become possible in the future. The prediction that jurors will necessarily come to see criminals as less blameworthy if at all accountable, will become far less certain and potentially problematic, as this chapter will show.

But this may not be the correct conclusion to draw from neuroscientific data. After all, if every behavior is eventually traced to detectable correlates of brain activity, does this mean that all troublesome behavior will one day be understood on a “don’t-blame-me-blame-my-brain” of misbehavior? Will no one ever be judged responsible? This chapter is intended to show why thinking through these profoundly important questions hinges on how one understands the relationship between the brain and the mind.

BRAIN AND MIND—WHY THEY ARE DIFFERENT

We must begin with the observation that despite all the advances in neuroscience, we do not remotely understand the basic relation between the brain, the mind, and action. We know that the brain enables the mind and action. A person with a dead brain has no mind and does not act. Despite this limited understanding, it is commonly accepted that “the mind” is basically what the brain (and the rest of the central nervous system) does, but the mind and the brain are not identical.

They are different because they represent different levels of analysis. Virtually all neuroscientists, psychiatrists, and psychologists certainly do not believe that brain and mind occupy materially separate domains. Few are Descartian substance dualists. Every subjective experience, from the ache of nostalgia to the frisson of a Christmas morning, corresponds to physical events in the brain. The mind—the realm of feelings, desires, ideas, memories, intentions, and subjective experience—is produced by the action of neurons and brain circuits. This leads many to claim that the brain and the mind are identical, but this claim is controversial. It is by no means clear that the armchair speculation that the mind is identical to the matter that produces it is true. It simply means that there is no clear scientific or philosophical answer at present, and there is much argument and evidence to the contrary. So, to say that the brain and the mind are different is not necessarily evidence of scientific ignorance, as some have alleged. For example, one cannot use the physical rules from the cellular level to completely predict activity at the psychological or behavioral level, and it seems unlikely that this will ever be true.

As an analogy, consider the words Shakespeare wrote on a page. They are, of course, “material,” but one could not capture the meaning of his plays by conducting a chemical analysis of the ink in which the letters of his words were written. The brain and the mind are different levels of explanation of the same phenomenon; the brain refers to mechanisms enabling perception, emotion, and cognition; the mind to subjective experiences such as awareness and meaning. Because brain and mind are different levels of description, they have different properties. This position is called *property dualism.* It must be distinguished from substance dualism, which implies that brain and mind are materially different and somehow in causal relation with each other.

From the standpoint of property dualism, one cannot rely on the brain alone to predict or understand everything important about human subjectivity or behavior. This is because many psychological phenomena are emergent properties of lower-order constituents, such as neural circuits, neurons, proteins, and genes. “Constitutive” reductionism—reducing complex entities to the sum of their component parts to facilitate study—is not controversial in the scientific community. In contrast, radical or “eliminative” reductionists go a crucial step further. They insist that everything mental will ultimately be *explained* fully at the material level of analysis. This variant of reductionism ultimately eliminates the need for the psychological level of analysis, not to mention all other levels—social, cultural, and so on. (Of course, one may reasonably ask why the reduction should end at the level of the brain, rather than “say” at the level of physics that underlies the brain.)

Adherents of eliminative reductionism speculate that such properties (the painting in full) will ultimately prove reducible to more basic elements (the paint). At one level, the purely physical, that may be true. But at the level that matters esthetically and emotionally, it is false. For the foreseeable future, valuable information is often lost when descending from higher explanatory levels, such as mental states, to lower levels, such as neuronal systems. To be sure, the neural level of explanation is appropriate for certain scientific questions, such as the nature and treatment of Alzheimer disease, wherein research targets amyloid plaques that disrupt communication among neurons, the neurofibrillary tangles that degrade them, and other aspects of brain anatomy and physiology.

Other times, however, a neural explanation can lead one astray. Consider depression. Although depression is clearly influenced by biologic factors, including a genetic predisposition, this does not mean that it is necessarily best treated by medication. In fact, certain psychological treatments, especially cognitive-behavioral and interpersonal psychotherapies, appear to be at least as effective as medications for depression while having no known physical side effects and often showing superior prevention of future episodes. Even treatments from one level of explanation seem helpful in treating depression. It does not follow that the cause or the primary cause can be found at that level. Claiming otherwise is known as the treatment-etiology fallacy.

LEVELS OF ANALYSIS, ADDICTION, AND THE BRAIN DISEASE FALLACY

Distinguishing brain from mind, or criticizing the excesses of brain imaging, is hardly a repudiation of the great insights of human neuroscience: Mental life arises from brain activity. Both levels, the brain and the mind, are essential to a complete understanding of human experience and behavior. The field of addiction psychiatry provides a good example. Within the last decade, in particular, neurocentrism has been ascendant and not always for the better. As a consequence of heavy promotion by the National Institute on Drug Abuse (NIDA), part of the National Institutes of Health, addiction has been labeled a “brain disease” and, moreover, a chronic and relapsing disease. The logic for this designation, as explained by a former director of the NIDA, is that “addiction is tied to changes in brain structure and function.” It is certainly true that repeated use of drugs such as heroin, cocaine, and alcohol alters the neural circuits that mediate the experience of pleasure as well as motivation, memory, inhibition, and planning—modifications that are often visible on brain scans. The critical question, though, is whether this neural disruption renders the addict incapable of self-control. Evidence indicates that they retain substantial control.

The neurocentric model leaves the addicted person in the shadows. Yet, to treat addicts and guide policy, it is important to understand how addicts think. It is the *minds* of addicts that contain the stories of how addiction happens, why they continue to use, and, if they decide to stop, how they cope with life stressors. The answers cannot be divined alone from an examination of addicts’ brains alone, no matter how sophisticated the probe. Viewing addiction as a “chronic and relapsing” condition diverts attention from promising behavioral therapies that challenge the inevitability of relapse by holding patients accountable for their choices. At the same time, because the model implies that addicts cannot stop using drugs until their brain chemistry returns to normal, it overemphasizes the value of brain-level solutions, such as pharmaceutical intervention.

Pragmatically, psychiatrists should be open to using any treatment if there is persuasive evidence that it is effective. But according to some thoughtful critics, NIDA has overemphasized the importance of pharmacologic intervention in the care of addicts. In 1997, the head of NIDA ranked the search for a medication to treat methamphetamine addiction as a “top priority.” A decade later, another director told *Newsweek* magazine that psychiatrists will “treat addiction as a disease [by 2018], and that means with medicine.” The search for a magic bullet is at best dubious—and even NIDA eventually gave up hope of finding a wonder drug—but simplistic forms of the brain-disease narrative continue to inspire unrealistic goals. When British pop star Amy Winehouse succumbed to her high-profile alcoholism in the summer of 2011, a *Psychology Today* blogger asked rhetorically, "Could neuroscience have helped Amy Winehouse?" The author answered in the affirmative, suggesting a dopamine-altering medication of the future because addiction "may be a brain problem that science can eventually solve." One neuroscientist goes even further, asserting that "addiction can be reasonably viewed as a neurological problem that allows for medical solutions, just as pneumonia can be viewed as a lung problem." But the analogy is questionable at best. Changing a behavior such as addiction requires addicts to work hard to change their patterns of thought and behavior. And it must be recalled that addiction is a behavioral condition centrally characterized by persistent seeking and using of substances, which are quintessential human actions. Virtually no other disease includes actions as part of the definition of the disease, although actions can ameliorate or exacerbate many diseases.

In contrast, antibiotic cures for pneumonia work even if the patient is in a coma. The hope of a medical treatment is the logical outgrowth of placing the brain at the center of the addictive process. Overall, success to date has been genuine but modest. When motivated patients take medications—especially patients already armed with relapse-prevention strategies and the support of family and friends—they can sometimes vault into sustained recovery. Methadone, a long-acting synthetic opiate taken once a day to prevent opiate withdrawal, has played a major role in treating addiction to heroin and painkillers since the 1960s. Still, up to half the patients in methadone clinics also fortify themselves with heroin, cocaine, or benzodiazepines sold on the street.

Despite three decades of effort, there is still no medication therapy for cocaine dependence. Cocaine immunotherapy (popularly called a cocaine "vaccine") to prevent cocaine molecules from entering the brain is now in development, but initial data do not look promising for wide-scale use. Other types of medications include blocking agents, such as naltrexone for opiate addiction, which occupy receptors for endorphins and blunt a drug's effect. Aversive agents, such as Antabuse (disulfiram), cause people to feel nauseated and vomit when they ingest alcohol. They can be effective in some cases, although many individuals elect to stop taking them.

NEURAL CORRELATES AND VOLUNTARINESS COEXIST

To be critical of the neurocentric model of addiction is not to dismiss the reality and research the importance of biologic correlates but rather to put them in perspective. The reward circuit is intimately involved in “cue-induced” craving. Such craving is a special species of desire that manifests itself in a sudden, intrusive urge to use drugs brought on by cues associated with use. The mere clink of a whiskey bottle, a whiff of cigarette smoke, or a glimpse of an old drug dealer on the corner can set off an unbidden rush of yearning fueled by dopamine surges. For the addict who is trying to quit, this is typically a tense, not a pleasurable, feeling. Because the rush of desire typically seems to come out of the blue, users may feel blindsided, helpless, and confused.

In a typical demonstration of the biologic correlates of craving, researchers study addicts as they lie in a brain scanner and watch videos of people handling a crack pipe or needle. In response, their prefrontal cortices, amygdalae, and other brain structures typically bloom with activity. (Videos of neutral content, such as landscapes, usually induce no such response.) Even in users who quit several months previously, neuronal alterations may persist, leaving them vulnerable to sudden, strong urges to use drugs.

The familiar late-1980s slogan “This is your brain on drugs” still echoes but now manifests with the brain itself substituting for the fried egg. But that egg is not always sizzling. There is a surprising amount of lucid time in the daily life of addicts. In their classic 1969 study “Taking Care of Business: The Heroin User’s Life on the Street,” criminologists Edward Preble and John J. Casey found that addicts spend only a small fraction of their days getting high. Most of their time is spent either working or hustling. The same is true for many cocaine addicts, as work by ethnographer Philippe Bourgois reveals. Most people tend to think of them, at their worst, frantically gouging their skin with needles, jamming a new rock into a pipe every 15 minutes, or inhaling lines of powder. In the grip of such hunger, an addict cannot be expected blithely to get up and walk away.

These tumultuous states—with neuronal function severely disrupted—is the closest drug use comes to being beyond the user's restraint. But in the days between binges, cocaine users worry about a host of everyday matters, such as whether to find a different job, enroll their children in school, or attend a Narcotics Anonymous meeting. It is during these stretches of relative calm that many addicts could make the decision to get help or quit on their own—and many of them do. But the decision to quit can be long in coming, far too long for those who destroy their health, families, or careers in the meantime.

The paradox at the heart of addiction is this: How can the capacity for choice coexist with self-destructiveness? "I've never come across a single person who was addicted who wanted to be addicted," is a common observation. The implication is that because the state was not desired, it must have been involuntary, that is, beyond the user's control. But how many have ever come across a heavy person who wanted to be overweight? Many undesirable outcomes in life arrive incrementally. "We can imagine an addict choosing to get high each day, though not choosing to be an addict," writes psychologist Gene Heyman. "Yet choosing to get high each day makes one an addict."

Imagine a typical trajectory to see how this dynamic unfolds. In the early phase of addiction, drugs or alcohol become ever more appealing, while once-rewarding activities, such as relationships, work, or family, decline in value. The attraction to the drug starts to fade as consequences accrue—spending too much money, disappointing loved ones, attracting suspicion at work—but the drug retains its allure because it blunts psychic pain, suppresses withdrawal symptoms, and douses intense craving. Addicts find themselves torn between the reasons to use and reasons not to.

Sometimes, a spasm of self-reproach or a flash of self-awareness tips the balance toward quitting. William S. Burroughs, an American novelist and heroin addict, calls this the "naked lunch" experience, "a frozen moment when everyone sees what is on the end of every fork." Christopher Kennedy Lawford, himself in recovery from drugs and alcohol, edited a 2009 collection of essays called *Moments of Clarity* in which the actor Alec Baldwin, singer Judy Collins, and others recounted the events that spurred their recoveries. Some quit on their own; others obtained professional help. A theme in each of their stories is a jolt to self-image: "This is not who I am, not who I want to be." One recovered alcoholic described the process: "You tear yourself apart, examine each individual piece, toss out the useless, rehabilitate the useful, and put your moral self back together again." These are not the sentiments of people in helpless thrall to their diseased brains. Nor are these sentiments the luxury of memoirists. Patients have described similar experiences: "My God, I almost robbed someone!" "What kind of mother am I?" or "I swore I would never switch to the needle."

Data consistently demonstrate that quitting is the rule, not the exception—a fact worth acknowledging, given that the official NIDA formulation is that "addiction is a *chronic and relapsing* [italics added] brain disease." The Epidemiologic Catchment Area Study, done in the early 1980s, surveyed 19,000 people. Among those who had become dependent on drugs by age 24, more than half later reported not a single drug-related symptom. By age 37, roughly 75% reported no drug symptoms. The National Comorbidity Survey, conducted between 1990 and 1992 and again between 2001 and 2003, and the National Epidemiologic Survey on Alcohol and Related Conditions, conducted between 2001 and 2002 with more than 43,000 subjects, found that 77% and 86% of people who said they had once been addicted to drugs or alcohol reported no substance problems during the year before the survey. By comparison, people who were addicted within the year before the survey were more likely to have concurrent ("comorbid") psychiatric disorders. The vast majority of those who quit received no professional services, although they often had the help of family or friends.

In addition, NIDA estimates that relapse rates of treated drug-addicted patients range from 40% to 60%. In other words, they are not representative of the universe of addicts. They are the challenging cases—the chronic and relapsing patients. They are also disproportionately comorbid. Yet these patients often make the biggest impressions on clinicians and shape their views of addiction, if only because clinicians are especially likely to encounter them. Falling prey to what psychologists term an "availability heuristic," researchers and medical professionals may err in generalizing from the sickest subset of people to the overall population of patients.

This caveat applies across the medical spectrum. Just as the clinician wrongly assumes that all addicts must be like the recalcitrant ones who keep stumbling through the clinic doors, psychiatrists and other mental health professionals sometimes view people with schizophrenia as doomed to a life of dysfunction on the basis of their frequent encounters with those whose delusions and hallucinations do not improve with treatment. The error of extrapolating liberally from these subsets of difficult patients is so common that statisticians Patricia and Jacob Cohen dubbed it the “clinician’s illusion.”

At this time, behavioral and psychological approaches have the research edge in treating addiction. In December 1966, Leroy Powell of Austin, Texas, was convicted of public intoxication and fined $20 in the lowest level court. Powell’s lawyer argued that he suffered from “the disease of chronic alcoholism.” Powell’s public display of inebriation, therefore, was “not of his own volition,” and the fine constituted cruel and unusual punishment. A psychiatrist concurred, testifying that Powell was “powerless not to drink.”

Then Powell took the stand. On the morning of his trial, he had a drink at 8 AM that his lawyer gave to him, presumably to stave off morning tremors. Here is an excerpt from the cross-examination:

Q: You took that one [drink] at 8 o’clock [AM] because you wanted to drink?
A: Yes, sir.
Q: And you knew that if you drank it, you could keep on drinking and get drunk?
A: Well, I was supposed to be here on trial, and I didn’t take but that one drink.
Q: You knew you had to be here this afternoon, but this morning you took one drink, and then you knew that you couldn’t afford to drink anymore and come to court; is that right?
A: Yes, sir, that’s right.
Q: Because you knew what you would do if you kept drinking that you would finally pass out or be picked up?
A: Yes, sir.
Q: And you didn’t want that to happen to you today?
A: No, sir.
Q: Not today?
A: No, sir.
Q: So, you only had one drink today?
A: Yes, sir.

The judge let stand Powell’s conviction for public intoxication. Powell appealed directly to the U.S. Supreme Court, the only appeal possible as a result of the procedural posture of the case. It, too, affirmed the constitutionality of punishment for public intoxication. “We are unable to conclude,” said the court, “that chronic alcoholics in general, and Leroy Powell in particular, suffer from such an irresistible compulsion to drink and to get drunk in public that they are utterly unable to control their performance.”

For people like Powell who are not otherwise motivated to quit, consequences can play a powerful role in modifying behavior. Powell took only a single drink on the morning of his trial because of foreseeable and meaningful consequences. Far from being unusual, his ability to curtail his drinking accords with a wealth of studies showing that people addicted to all kinds of drugs—nicotine, alcohol, cocaine, heroin, methamphetamines—can change their behavior in response to rewards or sanctions. Powell had surely experienced many alcohol-induced brain changes, but they did not keep him from making a choice that morning. If Powell came before a judge today, his lawyer might well introduce a scan of his brain “craving” alcohol as evidence of his helplessness. If so, the judge would be wise to reject the scan as proof. After all, a judge, or anyone, can ponder scans of “addicted” brains all day, but they would never consider someone an addict unless that person behaves like one.

Researchers have found that the brains of smokers reporting a strong desire to smoke commonly display enhanced activation of reward circuitry, as would be expected. But they also showed that subjects could reduce the brain’s telltale signs of craving by considering the long-term consequences of smoking, such as cancer or emphysema while observing videos depicting people smoking. When subjects did so, their brains displayed enhanced activity in areas of the prefrontal cortex associated with focusing, shifting attention, and controlling emotions. Simultaneously, activity in regions associated with reward, such as the ventral striatum, decreased. Investigators at NIDA observed the same pattern when they asked cocaine users to inhibit their craving in response to cues. Subjects underwent positron emission tomography (PET) scanning as they watched a video of people preparing drug paraphernalia and smoking crack cocaine. When researchers instructed the addicts to control their responses to the video, they observed inhibition of brain regions normally implicated in drug craving. When not deliberately suppressing their cravings, the addicts reported feeling their typical desire to use drugs, and the PET scans, accordingly, revealed enhanced activation in brain regions that mediate craving.

These powerful findings illuminate the capacity for self-control in addicts. They also underscore the idea that addicts persist not because of an inability to control the desire to use but at least in part because of a failure of motivation. Granted, summoning sustained motivation can be a great challenge: It takes considerable energy and vigilance to resist the craving, especially urges that ambush the addict unexpectedly. Studies on the regulation of craving also help distinguish behaviors that people do not control from behaviors that they cannot control. Imagine, by way of contrast, promising a reward to people with Alzheimer if they can keep their dementia from worsening. That would be both pointless and cruel because the kinds of brain changes intrinsic to dementia are immune to rewards or penalties.

What Powell’s case showed was that even though he sustained brain changes, those changes did not prevent his behavior from being shaped by consequences. Contingency management—the technical term for the behavior therapy technique of adjusting consequences, including incentives—often succeeds with people who face serious losses, such as their livelihood, professional identity, or reputation. When addicted physicians come under the surveillance of their state medical boards and are subject to random urine testing, unannounced workplace visits, and frequent employer evaluations, they fare well: 70% to 90% are employed with their licenses intact 5 years later. Likewise, scores of clinical trials show that addicts who know they will receive a reward, such as cash, gift certificates, or services, are nearly two to three times as likely to submit drug-free urine samples as addicts not offered rewards.

Unfortunately, treatment programs are rarely in a position to offer cash or costly rewards. But the criminal justice system has an ample supply of incentives at its disposal and has been using such leverage for years. One of the most promising demonstrations of contingency management comes from Honolulu in the form of Project HOPE, Hawaii's Opportunity Probation with Enforcement. Project HOPE includes frequent random drug testing of offenders on probation. Those who test positive are subject to an immediate and brief incarceration. Sanctions are fair and transparent: All offenders are treated equally, and participants know what will happen in case of an infraction. The judges expressed a heartfelt faith in offenders' ability to succeed. These basic elements of HOPE's contingency administration—swiftness, sureness, transparency, and fairness combined with the expectation for achievement—are a potent prescription for behavior change in just about anyone. Indeed, after 1 year of enrollment in Project HOPE, participants fared considerably better than probationers in a group who served as a comparison. They were 55% less likely to be arrested for a new crime and 53% less likely to have had their probation revoked. These results are even more impressive in light of the participants' criminal histories and their heavy, chronic exposure to methamphetamine, which can impair aspects of cognitive function.

These findings join a vast body of experimental data attesting to the power of incentives to override the lure of drugs. Yet, because the facts contradict the idea that addiction is analogous to Alzheimer disease, some HOPE personnel objected to incentives, arguing that addicts could not be accountable for their behavior. Indeed, the brain-disease narrative misappropriates language better used to describe such conditions as multiple sclerosis or schizophrenia—afflictions of the brain that are neither brought on by the sufferer nor modifiable by the desire to be well. It offers false hope that an addict’s condition is completely amenable to a medical cure. Finally, it threatens to obscure the vast role of personal agency in perpetuating the cycle of use and relapse.

Addicts embarking on recovery often need to find new, clean, and sober friends, travel new routes home from work to avoid passing near their dealer’s street or deposit their paycheck directly into a spouse’s account to keep from squandering money on drugs. A teacher trying to quit cocaine switched from using a chalkboard—the powdery chalk was too similar to cocaine—and had a whiteboard installed instead. An investment banker who loved injecting speedballs—a cocktail of cocaine and heroin in the same syringe—made himself wear long-sleeved shirts to prevent glimpses of his bare and inviting arms. Former smokers who want to quit need to make many fine adjustments, from not lingering at the table after meals to ridding their homes of the ever present smell of smoke, removing car lighters, and so on.

Thomas Schelling, a 2005 Nobel laureate in economics, refers to these purposeful practices as self-binding. The great self-binder of myth was Odysseus. To keep himself from heeding the overpowering song of the sea sirens—the half-woman, half-bird creatures whose beautiful voices lured sailors to their deaths—Odysseus instructed his men to tie him to the mast of his ship. The famous Romantic English poet Samuel Taylor Coleridge, an opium addict, is said to have hired men to prevent him from entering a pharmacy to purchase opium. Today, one can hire a firm that will provide binding services. It imposes surprise urine tests on the client, collects evidence of attendance at Alcoholics Anonymous (AA) meetings or treatment sessions, and sends a monthly status report (with the good or bad news) to another person, such as a parent, spouse, or boss.

Some addicts devise their own self-binding strategies. Others need the help of therapists, who teach them to identify and anticipate cues that trigger craving. Beyond the classic triad of people, places, and things, addicts may come to realize that internal states, such as stress, bad moods, and boredom, can prompt drug urges. Managing craving matters mightily in recovery, but it usually is not enough. Another important truth is that an addict uses drugs or alcohol because they serve a purpose, or what behavioral psychologists term an operant function. Caroline Knapp, in her powerful 1996 memoir *Drinking: A Love Story,* recounted why she spent two decades of her life as an alcoholic: “You drank to drown out fear, to dilute anxiety and doubt and self-loathing and painful memories.” Knapp doesn’t describe an urge to drink so much as a need to drink. She was not manipulated by an alien desire but by something woven into her being. To say that Knapp’s problem was merely the effect of heavy drinking on her brain is to miss the larger picture.

Heroin and speed helped screenwriter Jerry Stahl, author of *Permanent Midnight,* attain “the soothing hiss of oblivion.” But when the drugs wore off, his vulnerabilities throbbed like a fresh surgical incision. In surveying his life, Stahl wrote, “Everything, bad or good, boil[ed] back to the decade on the needle, and the years before that imbibing everything from cocaine to Romilar, pot to percs, LSD to liquid meth and a pharmacy in between: a lifetime spent altering the single niggling fact that to be alive means being conscious.” Or take Lisa, a 37-year-old woman featured in a 2007 HBO documentary on addiction. When viewers meet her, Lisa is living in a rundown hotel room in Toronto and working as a prostitute. She sits on the bed and talks with the filmmaker behind the camera. Flipping her shiny brown hair and inspecting her well-kept nails, Lisa is animated as she boasts about how much she makes selling sex, how much she spends on cocaine, and the longed-for “oblivion” that drugs help her attain. When Lisa was filmed, she was healthy and engaging; she looked and talked like someone who had recently been abstinent but was back in the early stages of her next downward spiral. She had no interest in stopping things at this point. “Right now, I am in no position to go into recovery. [This way of life] is working for me…. I have money, drugs, business. I’m O.K.” To say that Lisa’s problem is the effect of cocaine on her brain distracts from the true threat to her well-being: Lisa herself. “I always use for a reason. It’s repressing what needs to be repressed,” she says.

These stories highlight one of the shortcomings of the strictly neurocentric view of addiction. This perspective ignores the fact that many people are drawn to drugs because the substances temporarily quell their pain: persistent self-loathing, anxiety, alienation, deep-seated intolerance of stress or boredom, and pervasive loneliness. The brain-disease model is of little use here because it does not accommodate the emotional logic—or illogic—that triggers and sustains addiction.

We conclude this chapter with a discussion of an important case that raises all the issues we have just discussed. For nearly 50 years, the holding in *Powell* that addicts should not be excused simply because they are addicts had scarcely been challenged. Then, in a 2017 Massachusetts case that received wide national attention and newspaper editorials, a full-fledged claim was mounted that, as a constitutional matter, addicts didn’t deserve punishment because they could not control their seeking and using. Defendant Julie Eldred was a 28-year-old fentanyl addict who stole and fenced some jewelry to support her habit. It was her second such offense. She essentially pleaded guilty. The sympathetic trial judge placed her on probation after Eldred promised to stay clean and to submit to random drug tests. About 10 days after being released, she tested positive for fentanyl but denied using it. Eldred was returned to court for her probation violation. Again, the sympathetic judge tried to avoid imprisoning Eldred. Instead, the judge ordered inpatient treatment, but no appropriate inpatient beds were available in the state, so the judge reluctantly sent Eldred to prison where she would not be a danger to herself or others. After about 10 more days, a bed was found, and she was transferred to the inpatient facility.

Ms Eldred then sued the state, claiming that any prison based on her use of the drug constituted cruel and unusual punishment in violation of her 8th and 14th Amendment rights. She claimed that it could not be fair to punish her for behavior symptomatic of her disease that she was helpless to avoid. Rather than decide this claim on the merits, the trial court certified the question to the Massachusetts Supreme Judicial Court, which agreed to hear the case. Virtually all professional medical organizations in the state submitted friend of the court (amicus) advisory briefs on Eldred's behalf. They based their arguments heavily on the neuroscience of addiction. The editorial board of the New York Times urged that relapsing shouldn't be a crime because addiction was a disease. Only two amicus briefs were offered on behalf of the state. The National Association of Drug Court Professionals rejected Eldred's claim because they, the treating professionals, know that incentives matter and fear of negative consequences, such as potential incarceration, are a bedrock of their therapeutic armamentarium. The state's position was also supported by an informal group of addiction experts, including the authors of this chapter, who argued on the basis of the evidence this chapter just addressed that addicts do have sufficient choice to be held accountable for relapsing and for other criminal behaviors associated with their addictions. The neuroscience does not entail the necessity of an excuse. We took no position on what the state's best policy should be, but we were clear that the evidence did not support a one-size-fits-all addiction excuse.

In the event, the Supreme Judicial Court decided unanimously to reject Ms Eldred's claim. It did leave proponents of the excuse an out, however. It held that it could not uphold Ms Eldred's position based on the record before it but seemed willing to reconsider if there were a full airing of the issues in the trial court below. The case has not proceeded, however. Instead, advocates have introduced legislation in Massachusetts to prevent the imprisonment of addicts for relapsing. The legislation has stalled, which we think is a good thing. The legislature is the appropriate branch to decide such issues, but we believe the policy of preventing imprisonment is unwise.

There have been very few such cases since Eldred was decided, but one seems worthy of brief mention. Manning v. Caldwell, a federal case (4th Cir. 2019), involved a constitutional attack on an exceptionally complex Virginia statute applied against homeless addicts. Essentially, the claim was the same as Ms Eldred's, and again neuroscience played a crucial role in the plaintiffs' argument. The bitterly divided court held that the record didn't support the plaintiffs' case, but the case was remanded to the trial court for a full hearing on the merits of the addiction claim. It was clear that the majority thought that a good case could be made and that there were grounds for overturning *Powell.* If there had been a remanded hearing and the case again reached the 4th Circuit and ruled in the plaintiffs' favor, a Supreme Court decision might have ensued. But the case settled, and there was no decision on the merits of the excuse. Nonetheless, the case does indicate that there is pressure to reconsider *Powell* based on the allegedly new scientific understanding of addiction. We expect more such cases to arise in the future. But as Justice Marshall, the author of *Powell*, wrote in that opinion, the petitioner was asking for too much based on too little knowledge. We think that is still true for the reasons given.

NEUROCENTRISM, FREE WILL, AND LEGAL RESPONSIBILITY

It is only natural that advances in knowledge about the brain make people think more mechanistically about themselves. But in one venue, in particular—the courtroom—this bias can be a prescription for confusion. The brain-based defense ("Look at this fMRI scan, your Honor. My client's brain made him do it.") is being used with increasing frequency in capital defenses. The problem with these claims is that, with rare exceptions, neuroscientists cannot yet translate aberrant brain functions into the legal requirements for criminal responsibility. The bedrock principles of criminal responsibility are that human beings are capable of acting intentionally and rationally, that mental states such as intentions play a partial causal role in explaining behavior, and that a reasonable degree of freedom of choice exists when the agent acts. More controversially, many believe that people must have an independent capacity for self-control to be responsible. More controversially, many believe that people must have an independent capacity for self-control to be responsible. For example, no one “blames” the weather or inanimate objects when they cause harm because these are not intentional and potentially rational agents. The severely mentally ill may be acting agents, but their rationality may be so compromised when they do harm that they do not deserve blame and punishment. The very young or the demented may also have their harmdoing mitigated or excused on rationality grounds; the former as a result of developmental immaturity, the latter as a result of brain disease.

What is known about many criminals is that they do not conform to the law and often say that they cannot control themselves. That is very different from being unable to do so. To date, brain science cannot distinguish these alternatives. Moreover, even abnormal-looking brains frequently have owners who are otherwise quite normal. Looking to the future, some neuroscientists envision a dramatic transformation of criminal law. David Eagleman of the Baylor College of Medicine’s Initiative on Neuroscience and Law hopes that “we may someday find that many types of bad behavior have a basic biological explanation [and] eventually think about bad decision making in the same way we think about any physical process, such as diabetes or lung disease.”

But is this the correct conclusion to draw from neuroscience? If every troublesome behavior is eventually traced to correlates of brain activity that can be detected and visualized, will society be able to excuse it on a don’t-blame-me-blame-my-brain theory? Will no one ever be judged responsible? Eagleman’s way of thinking may fall prey to what law professor Stephen Morse calls the “fundamental psycho-legal error,” the powerful temptation to equate cause with excuse. Morse notes that the law excuses criminal behavior only when a causal factor produces an impairment so severe that it deprives the defendant of their rationality. Bad genes, bad parents, or even bad stars are not an excuse.

Nothing less than the legal system’s authority to hold offenders accountable hinges on a clear grasp of the relationship of mental contents and mental capacity to ascriptions of responsibility. How can neuroscientific data assist the law in ascertaining guilt? To answer this question, it is important to understand how the law determines guilt. Some brief background: The criteria for legal (and moral responsibility) are the folk psychological concepts of acts and mental states. American criminal law would hold a person responsible for a crime if they intended to commit a prohibited act. This mental state is called *mens rea,* or guilty mind, which generally requires either intent, knowledge, recklessness (conscious awareness of a risk that one is doing wrong), or negligence (creating an unreasonable risk of doing wrong without awareness that one is doing so). In addition, the criminal law also provides for what are known as affirmative defenses. In these cases, the prohibited act and mental state may be proven, but the defendant claims that his act was right under the circumstances (a justification, such as self-defense) or claims that even if what he did was wrong, he was not responsible (an excuse, such as legal insanity).

Without evidence of *mens rea,* the law cannot hold a person criminally responsible. For example, hitting and killing a pedestrian when one's car goes out of control does not entail *mens rea,* whereas aiming one's car at a pedestrian, stepping on the gas, and running over that person does. Again, even if the defendant intentionally ran over the victim, the defendant may claim that he acted in self-defense (e.g., the victim was trying to shoot me) or acted because he had a delusional belief about the need to kill the victim.

The question for brain imaging is whether it is able to translate information about brain function into the legal requirements for criminal responsibility. Now, neuroscience is an utterly mechanical level of explanation. Neurons and the connectome do not have reasons, aspirations, or a sense of past, present, and future. These are properties of persons. The law, in contrast, is concerned with acts and mental states. The question is whether the mechanical language and concepts of neuroscience can be validly translated into the folk psychological language of the law. The answer is that brain scans cannot shed significant light on most criteria for responsibility. Thus, lawyers, jurors, and judges will still need to rely on traditional methods of assessing the defendant: interviews, observations, witness reports, psychiatric history, and well-established clinical assessments. It is from these methods that a subtler, necessary appreciation of the defendant's mental state can be inferred.

When defense attorneys want to convince the judge and jury that their clients are less worthy of blame or punishment, they try to make the case that the defendant's capacity to reason or to control himself is defective. And because his brain is defective, he made a bad choice, a choice that was unavoidable, and so he should not be held responsible. But what about the argument that no one is truly responsible? What are the implications of brain science for morality and law? Although individuals generally think of themselves as free agents who make choices, a number of prominent scholars claim that they are mistaken. “Our growing knowledge about the brain makes the notions of volition, culpability, and, ultimately, the very premise of the criminal justice system, deeply suspect,” contends Stanford biologist Robert Sapolsky. It should be recognized that critiques such as this are “external” because they reject the premises of the criminal justice system in their entirety. An internal critique, in contrast, accepts the basic validity of the system but attempts to change specific doctrines and practices to achieve greater justice. Often, this debate is framed in terms of whether humans have “free will.” This is an interesting metaphysical question that science cannot answer, but the question exerts its enduring hold, we contend, because so many people believe that having the God-like power to act uncaused by anything other than oneself is necessary for responsibility, and we care deeply about responsibility. The remainder of this section, therefore, explores the interlocking questions of free will and criminal responsibility.

If otherwise healthy individuals can think coherently, plan and change their minds in response to information, should not that render them free enough to be blamed (or praised, for that matter)? Philosophical “compatibilists,” who roughly adopt some version of physicalism (all phenomena start with matter, including immaterial things such as culture) and causal closure (the universe is a causal place; no spooky entities allowed), would say yes. The compatibilist view does not negate material explanations of the world. They concede that behavior and thoughts are the products of a vast array of influences—genes (and the evolutionary history they represent), brain mechanisms, personal upbringing, and the physical and social environments in which people live. They further acknowledge that these forces converge, presumably, to produce one and only one specific act, be it “choosing” soup over salad or murder over mercy. Yet, under this determined view of the universe, people who nonetheless possess the capacities to be guided by reason, to follow rules, deliberate, be deterred, and conform to norms are considered by philosophers to have the freedom to choose—and, thus, to be accountable and perhaps deserving of blame and punishment. Only if the bedrock premises of criminal responsibility are missing will we consider the agent not responsible.

Before continuing, it is important to point out a fundamental error that doctors, lawyers, judges, and others make all the time. It is the misguided belief that if we identify a cause for a particular behavior, especially an abnormal cause such as a mental disorder, then the behavior should be excused or at least mitigated. Because all behavior is caused, the implication of this view is that no one is responsible, but this is inconsistent with the moral and legal world we inhabit. Some harms are caused by neuromuscular spasms; others by intentional actions. Some criminals are severely mentally disordered when they commit their crimes; most are not. Yet all are fully caused even if we don't understand all the causes of the behavior. Yet we morally and legally distinguish these cases on grounds we have good reason to accept: the distinction between no action and action or between rationality and irrationality. Causation, whether by biologic, psychological, sociologic, or astrologic variables, is not per se an excusing condition.

Others, like Sapolsky, disagree. They insist that individuals' deliberations and decisions do not make them free to choose because they are dictated by neuronal circumstances. They say that, as people come to understand the mechanical workings of their brains, they will be compelled by logic to adopt a strictly utilitarian model of justice in which criminals are "punished" solely as a way to change their behavior, not because they deserve blame. Philosophers call this a "hard determinist" view of free will, and some believe that neuroscience can demonstrate that it is the correct understanding of free will.

Philosopher–neuroscientist Joshua Greene and neuroscientist Jonathan Cohen contend that neuroscience has a special role to play in giving the age-old argument about free will more rhetorical bite. "New neuroscience will affect the way we view the law, not by furnishing us with new ideas or arguments about the nature of human action, but by breathing new life into old ones," they write. The familiar argument, of course, is that mental states can be reductively explained mechanistically. The authors embrace the mechanistic conception, believing that neuroscience will accelerate the elucidation of "the mechanical processes that cause behaviour." After predicting that people will finally be convinced by the mechanical account, they make the following claim about the criminal law: "The law will continue to punish misdeeds, as it must for practical reasons, but the idea of distinguishing the truly, deeply guilty from those who are merely victims of neuronal circumstances will, we submit, seem pointless." The most natural reading of this passage in light of the argument about mechanism is that all behavior is mechanically caused in the same way. In short, although we may behave differently, at base, we are all “victims of neuronal circumstances,” or just a “pack of neurons,” as Francis Crick claims. The compatibilist view is most common among the public and even among philosophers. But so what if most people think that they and others have “free will”? Now comes the heart of the matter: The question of whether humans can live in a material world and yet be morally responsible is not empirically testable. It is not a scientific problem. It is a conceptual and ethical impasse that has bedeviled thinkers since antiquity and is still without a resolution.

What is important to establish is that neuroscience has not resolved it either. Those who believe that the absence of ultimate free will means that moral responsibility is an incoherent notion and that society should therefore abolish blame have already staked out a philosophical claim. Simply amassing more data on the workings of the brain may strengthen their conviction that determinism is correct, but it will not make their argument against the coexistence of determinism and moral agency any stronger. The vision of the hard determinist neurocentrist is inconsistent with what common sense, bolstered by scientific confirmation, has taught. But no science, including neuroscience, can prove a metaphysical thesis such as hard determinism. Thus, the neurocentrists are mounting a radical challenge to entrenched concepts and practices about how we should live together, a task in which law and criminal blame and punishment play a central role. In what follows, we suggest the wide range of data and considerations that indicate just how radical the neurocentrist view is. Because concepts and practices are accepted does not mean that they are scientifically or morally justified, but the burden of persuasion must surely be on the radical critics who wish to upend many millennia of human experience and practice based on an unprovable metaphysical thesis that would strip humans of responsibility and its consequences that provide much of the basis for human dignity and worth.

The hard determinist view—a world without “just desert”—raises a set of interesting questions regarding how the criminal justice system should work. Most momentously, are humans even capable of removing blame from the moral and legal domain? Could it somehow be pried loose from entrenched notions of how human beings operate? The prospect of a world without moral responsibility collides head-on with the innate sense that people can choose freely. By the age of 5, most children perceive others' actions in terms of their intentionality and agency. In one experiment, kindergartners observed an investigator sliding open the lid of a box, putting her hand in, and touching the bottom of the box. Asked by the experimenter if she had to touch the bottom or whether she "could have done something else instead," the vast majority of children believed that she could have done something else. But when the experimenter placed a ball on the lid and slid open the lid, and the ball fell to the bottom of the box, only a few of the children said that the ball itself could have "done something else instead."

The intuition that people can do something else persists with age. Adults across cultures, religions, and countries steadfastly reject the idea that the decisions are fixed in such a way that no other actions are possible. Just as consistently, human beings are highly attuned to notions of fairness. Communities buzz with gossip about who did what scandalous thing to whom; people keep score ("I did two favors for you, but you did only one for me") and punish cheaters who fail to reciprocate. Indigenous people in the Amazonian rain forest are as diligent about detecting free riders and are as ready to punish offenders as are college students in the United States, Europe, and Asia. In his 1991 book *Human Universals,* anthropologist Donald E. Brown presented a comprehensive survey of moral concepts shared across the globe. Among them are proscriptions against murder and rape, as well as redress of wrongs.

In an exhaustive review of anthropologic data, psychologists found that within all human cultures, individuals tend to react with quick, automatic feelings of anger, contempt, and indignation when they see people causing suffering. The universality of these responses raises the distinct possibility that intuitions about fairness, harm, and respect for authority are innate to the human mind. The behavioral economics laboratory is a showcase for fairness attitudes as well. Nobel Prize–winning psychologist and behavioral economist Daniel Kahneman and colleagues documented how study participants voluntarily incurred a penalty to punish others for unfair behavior. It is even more striking that they found evidence that people will pay to punish a third person even when participants are not directly involved in the transaction if they perceive that the third person's unfair behavior was intentional.

The stirrings of moral sentiment begin early on. In a series of studies, psychologists Karen Wynn and Paul Bloom of Yale University found that toddlers who watched a puppet show in which one puppet "stole" a ball and another puppet returned the ball to its rightful owner were far more likely to give candy to the helper puppet than to the "bad" puppet and preferentially take it away from the "bad" puppet. More generally, violations of fairness arouse emotions that trigger the urge to retaliate, especially when the violator intentionally injures an innocent person.

Social psychologist Philip Tetlock of the University of Pennsylvania and his collaborators presented subjects with a scenario involving a brutal assault that left the victim permanently brain damaged. No matter what happened to the aggressor—in one hypothetical scenario, he suffered a painful accident; in another, he was cured with medication—subjects' desire for justice was unchanged. Merely rendering him harmless or causing pain accidentally was not sufficient. This is why certain nations, such as Israel, direct their law enforcement officials to pursue octogenarian Nazi war criminals living quietly in South America and elsewhere even though they no longer pose a danger. The deep-seated desire for justice compels that the moral ledger be balanced.

Intrinsic to the idea of retribution—which is not revenge; it is a theory of justice based on giving people what they deserve—is that people must suffer in proportion to the suffering they inflicted. Psychologist Jonathan Haidt of New York University and colleagues showed clips from Hollywood films that portrayed injustice (one involving the rape and murder of a child, and another in which a slave's foot is mutilated by a slave catcher). They next gave subjects a variety of endings and asked which one was the most "satisfying." Among the alternative endings was the "revenge" option: The grieving mother violently kills her daughter's rapist; the hobbled slave chops off part of the foot of the man who mutilated him. In the "catharsis" option, the mother undergoes "primal-scream" therapy; the slave chops wood while visualizing the slave catcher's foot.

In the "forgiveness" ending, the victims join a support group or become more active in church and learn to forgive the transgression that was committed. The viewers derived far less satisfaction from the scenarios in which the victims come to terms with their tragedies and forgive the transgressor. They wanted the perpetrators to pay. It was most satisfying when the punishment matched the crime. At the same time, viewers found gratuitous and less satisfying another ending in which the slave retaliated by murdering his catcher.

This result is consistent with the extensive work of psychologists the late Kevin M. Carlsmith and John M. Darley of Princeton University. In a series of experiments, they found that when it came to sentencing wrongdoers, subjects were exquisitely sensitive to the severity of the offense—say, the theft of $100 to feed a starving child versus using the money to create the world's largest margarita—and largely ignored the likelihood that the person would offend again. Subjects punished people exclusively in proportion to the harm done and not for the harm that might be done in the future. "People want punishment to incapacitate and to deter, but their sense of justice requires sentences proportional to the moral severity of the crime," they concluded. Punishment that violated proportionality offended people's intuitive sense of fairness.

Although hard determinists reject the concept of blameworthiness, they still accept the view that censure can have practical value. As any parent knows, well-calibrated disapproval, along with encouragement, is essential to raising children who are sensitive to the rights of others, behave kindly toward the injured and vulnerable, and reciprocate when they have been helped. No society, whether modern or preliterate, can function and cohere unless its citizens exist within a system of personal accountability that stigmatizes some actions and praises others. Punishment signals to the community who should not be trusted, and its severity reflects the enormity of the transgression. But none of these punishment-related functions requires that anyone be blamed. In utilitarian fashion, punishment could be meted out simply to shape future behavior.

Retribution, however, is a thoroughly different concept. In its pristine, theoretical form, retributive practices are triggered by the simple fact that someone who is mentally competent and uncoerced has committed an offense. The point of punishment is to make perpetrators suffer in proportion to the harm that they have already caused the victim and society. Any incidental benefit to the greater society that comes from reinforcing norms or protecting against future crime is irrelevant.

But when retribution is applied in the real world, it inescapably carries great practical value, too. For one thing, it strengthens people's shared norms of moral obligation to one another. One of those norms is that victims should be valued as human beings. Consider the following vignette: A serial rapist, John, attacks Mary, is found guilty, and is sent to prison. A few months later, John is treated with "Castrex," a fictional new antirape medication guaranteed to permanently eradicate sexually aggressive urges. Castrex works after just a few doses, and several weeks later, John is freed. He is no longer a danger to anyone. His rehabilitation was a success. But John’s light punishment would almost surely have woeful repercussions for Mary, her family, and the community at large.

When society fails to condemn aggressors or simply slaps them on the wrist, victims often feel unavenged and therefore devalued and dishonored. If perpetrators die before they can be judged or are killed in prison before they can be adequately punished, victims and their families usually feel enraged. Criminals who do not “pay their debt” can spur victims and their families to contemplate private retaliation and sometimes even undertake it. Contrary to common perception, such feelings do not always arise out of a sense of white-hot rage or bloodlust vigilantism. The motivation to give wrongdoers what they deserve can instead be motivated by grief or by a solemn sense of duty to set things right.

Communities, too, resist what they perceive to be inadequate punishment. Consider the 2011 trial of Casey Anthony, the 25-year-old Florida woman who never reported her 2-year-old daughter missing. Widely believed to have killed her child or at least to have abetted in her death, Anthony received death threats after she was exonerated for the crime of murder. Likewise, jurists often speak of their moral duty to satisfy the victims and their families, but such satisfaction should not produce disproportionately long sentences.

Restoring the social standing of the victim is another vital function of just punishment. When the law imposes inadequate punishment on wrongdoers or allows them to escape penalties altogether, the demoralizing message to the community is clear: The victim is so inconsequential that their rights, security, and property can be breached with impunity. This is why proportionate punishment must be carried out in full view of the community after open trials, announced in the press, and preceded by a clear message of society’s moral condemnation. Everyone in the community must be reminded that such mistreatment will not be tolerated. In a series of experiments, legal scholar Kenworthey Bilz, now of the Illinois University College of Law, asked subjects to evaluate the moral worth of a female victim whose attacker was punished for rape and a similar woman whose rapists pled guilty to a lesser offense. Both women were judged by participants before their rapists were punished. When participants learned that the rapists were convicted of rape, they rated the victim as more “respected,” “valued,” and “admired” than they were before punishment. When the rapists pled to a lesser nonsexual crime, subjects ranked the victim’s social standing lower than in the prepunishment phase.

What happens when the community is prevented from administering justice? In the mid-1960s, sociologist Melvin Lerner developed the “belief in a just world” hypothesis. According to this influential hypothesis, people harbor a strong need to believe that the world is a place where people get what they deserve and where actions have predictable consequences. A just-world belief is akin to a “contract” with the world regarding the outcomes of one’s conduct: do the right thing and be rewarded—or, at least, know largely what to expect. Lerner understood that the notion of a truly just world is illusory, yet he posited that it helps people to plan their lives and achieve their goals.

In one of his seminal experiments, Lerner asked subjects to observe a 10-minute video of a fellow student as she underwent a learning experiment involving memory. The student was strapped into an apparatus sprouting electrode leads and allegedly received a painful shock whenever she answered a question incorrectly (she was not receiving real shocks but believably feigned distress as if she were). Next, the researchers randomly assigned the observers into groups. One was to vote on whether to remove the victim from the apparatus and reward her with money for correct answers. All but one voted to rescue her. The experimenters told another group of observers that the victim would continue to receive painful shocks; there was no option for compensation. When asked to evaluate the victim at this point, participants in the victim-compensated condition rated her more favorably (e.g., more “attractive,” more “admirable”) than did subjects in the victim-uncompensated condition, in which the victim’s suffering was greater. Lerner’s conclusion is disturbing. “The sight of an innocent person suffering without the possibility of reward or compensation motivated people to devalue the attractiveness of the victim in order to bring about a more appropriate fit between her fate and her character,” he wrote. In other words, when subjects’ intuitions of justice are satisfied, their belief in a just world is supported. But when subjects (or society) are prevented from restoring justice, they typically blame the victim. Somehow, the reasoning goes, she must have asked for it.

Beyond the victims, the legal system itself suffers when criminals are not punished, although not just any punishment will do. The penalty must seem proportionate to the offense. If it is perceived as skewed in either direction—too lenient or too harsh—the law can lose its moral force. Recall Rodney King, the man who led California police on a high-speed chase before being caught and beaten severely. Shortly after the 1992 acquittal of the police officers involved in the assault, half of all Californians surveyed said that they had lost confidence in the court system. In a system perceived as unfair, juries may ignore judges' instructions, and police officers may impose their own judgment on whether to arrest, trump up charges, or abuse suspects.

For their part, witnesses may rebel by refusing to participate in investigations or to testify. Researchers have observed that subjects are more willing to commit minor offenses such as traffic violations, petty theft, and copyright violations when laws more generally do not conform to their sense of right and wrong. Jurors are more susceptible to nullification—that is, to acquit defendants who are legally guilty—when the verdict dictated by law is contrary to their sense of justice, morality, or fairness. Examples of perceived unfairness include suspicions of judges' failure to admit evidence, prosecutorial suppression of evidence, police misconduct, or police lying on the stand.

Finally, intuitions about justice are strong motivators for social change. The victims' rights movement grew out of a powerful sense of unfinished business. Unmet justice alienates juries, judges, and prosecutors, thereby undermining the authority of the law, an institution that depends for its smooth functioning on the good faith of its participants. Pragmatic efforts aimed solely at reducing future crime (and not at all at exacting retribution for past bad deeds) would leave many victims questioning the moral credibility of the law. Those victims do not necessarily seek harsh treatment for their violators, although they surely want judges to administer appropriate punishment. What they want most is for the law to acknowledge that what happened to them was wrong and morally offensive. Indeed, the idea of a victim impact statement was developed so that the sentencing judge could hear directly about the anguish that victims and their loved ones had suffered. Some victims want an apology from their violators—not in place of a penalty, but in addition to it.

In summary, some scientists say that they look forward to a day when neuroscience will explain the mechanical workings of the brain so thoroughly that it will be impossible for society to continue to ignore the "fact" that people do not choose their behavior and, therefore should be absolved of blame. This vision seems to be on a collision course with the psychological and social meanings that criminal justice in general and retribution in particular hold for victims, their loved ones, and societies.

Victims of crime are exquisitely attuned to society's regard for the people who have violated their rights. Unless authorities mete out deserved punishment, victims feel devalued and suffer a loss of moral standing in their community. Society is also harmed when citizens do not see justice being served and thereby lose faith in the moral authority of the law.

Although cloaked in neuroscientific garb, this free will and responsibility question remains one of the great conceptual impasses of all time, far beyond the capacity of brain science alone to resolve. Unless, that is, investigators can show something truly spectacular: that people are *not* agents, not conscious beings whose actions flow from reasons and who are responsive to reason. True, people do not exert as much conscious control over their actions as they think they do. Every student of the mind, including William James and Sigmund Freud, has been aware of this important truism. But it does not mean humankind is powerless to act otherwise than it does. Neuroscience simply cannot prove that responsibility, which is at the heart of our conception of ourselves and others and gives so much meaning and worth to our lives, is an illusion.

CONCLUSION

The brain is said to be the final scientific frontier, and perhaps with some justification. Yet, in many quarters, brain-based explanations appear to be granted a kind of inherent superiority over all other ways of accounting for human behavior—they have exercised the "neurocentric" presumption or the view that human experience and behavior can be best explained from the predominant or even exclusive perspective of the brain. From this popular vantage point, the study of the brain is inherently more "scientific" than the study of human motives, thoughts, feelings, and actions. By making the hidden, allegedly but misleadingly visible, functional brain imaging has been a spectacular boon to neurocentrism. Such imaging has contributed substantially to the knowledge of the correlates of mental illness and personality. The risk is losing sight of the mind in the age of brain science. Although the scans are dazzling and the technology an unqualified marvel, it is important to remember that the brain and the mind, although not differing materially, are two complementary but different frameworks for conceptualizing human behavior.

The neurobiologic domain is one of brains and physical causes, the mechanisms behind one's thoughts and emotions. The psychological domain, the realm of the mind, is one of people—their desires, intentions, ideals, and anxieties. Both are essential to a full understanding of why people act as they do.

ACKNOWLEDGMENT

Scott O. Lilienfeld, who died on September 29, 2020, was a transformative leader in the field of clinical psychological science. An excellent teacher and adored mentor, his work was known for its clarity, astute critical frame, fairness, and intellectual courage.

FURTHER READINGS

Bloom P. Seduced by the flickering lights of the brain. *Seed Magazine*. 2006.

Bourgois P. *Search of Respect: Selling Crack in El Barrio*. Cambridge University Press; 1996, First edition.

Cohen P, Cohen J. The clinician's illusion. *Arch Gen Psychiatry*. 1984;41(12):1178–1182.

Dennett DC. *Elbow Room: The Varieties of Free Will Worth Wanting*. MIT Press; 1984.

DuPont RL, McLellan AT, White WL, Merlo LJ, Gold MS. Setting the standard for recovery: Physician's Health Programs. *J Subst Abuse Treat*. 2009;36(2):159–171.

Eagleman DM, Correro MA, Singh J. Why neuroscience matters for rational drug policy. *Minn J L Sci Tech*. 2010;11(1):7.

Greene J, Cohen J. For the law, neuroscience changes nothing and everything. *Philos Trans R Soc Lond B Biol Sci*. 2004;359:1775–1785.

Hawken A. Behavioral triage: a new model for identifying and treating substance-abusing offenders. *J Drug Policy Anal*. 2010;3(1):1.

*Heyman GM. *Addiction: A Disorder of Choice*. Harvard University Press; 2008.

Higgins ST, Silverman K, Heil SH. *Contingency Management in Substance Abuse Treatment*. Guilford Press; 2008.

Interlandi J. What addicts need. *Newsweek*. 2008;151(9):36–42.

Kagan J. *An Argument for the Mind*. Yale University Press; 2006.

Koob GF, Volkow ND. Neurocircuitry of addiction. *Neuropsychopharmacology*. 2010;35(4):217–238.

*Morse SJ, Roskies AL. *A Primer on Criminal Law and Neuroscience: A Contribution of the Law and Neuroscience Project*. Oxford University Press; 2013.

Richards JR. *Human Nature After Darwin: A Philosophical Introduction*. Routledge; 2000.

*Satel S, Lilienfeld SO. *Brainwashed: The Seductive Appeal of Mindless Neuroscience*. Basic Books; 2013.

*Schelling TC. The intimate contest for self-command. *Public Interest*. 1980;60:94–118.

4.2 Transcultural Psychiatry

KENNETH FUNG, M.D., F.R.C.P.C., M.SC., AND ROBERT KOHN, M.D.

Culture, while often overlooked, is all around us and broadly refers to patterns of human activity among groups of people, giving rise to symbolic systems, shared meaning, and social structures. This includes languages, systems of knowledge, beliefs, norms, values, and other behavioral patterns that are socially acquired and may be transmitted within and across generations. Culture can shape our perception and understanding; thoughts and feelings and their expression; motivation and morality; rituals and other behavioral patterns and practices; identity and sense of belonging; verbal and nonverbal communication; social norms and relationships; social systems and structures including families, kinships, communities, and institutions; and technology and tools.

Culture has eight essential characteristics:

1. It is learned.
2. It is both an individual and group process.
3. It can be passed on from one generation to the next.
4. It involves a set of words, behaviors, events, and symbols whose meanings have been tacitly or explicitly agreed upon by the cultural group.
5. It acts as a template to shape and orient future behaviors and perspectives within and between generations and to take account of novel situations encountered by the group.
6. It exists and evolves in a constant state of change.
7. It includes patterns of both subjective and objective components of human behavior.
8. It is expressed in societal and structural elements, such as institutions and laws.

As cultural influences pervade all human activities, the aim of cultural psychiatry (transcultural psychiatry) includes developing scientific lines of inquiry and clinical approaches to examine and understand the influence of culture and sociocultural contexts on all aspects of mental health and psychopathology, including cultural influences on and interactions among different levels of analysis, including neurobiologic, cognitive-affective, interpersonal, and social systems.

Cultural competence can be broadly defined as the capacity to function effectively taking cultural factors into consideration, including at the micro level of individual clinical care; meso level of programmatic and organizational activities; and macro level of systemic or societal structures and characteristics. Learning objectives in clinical cultural competence training include attitudes (affective), knowledge (cognitive), and skills (psychomotor) domains, as well as the interrogation of the impact of power differentials among cultural groups on the three domains, such as whose knowledge systems are being prioritized and valued. While the concept of cultural competence is useful for providing a framework for training and clinical care for mental health professionals and increasing accountability of the healthcare system to uphold a minimum standard of care for diverse populations, it has been criticized when it is narrowly conceptualized, defined, and taught, such as when it is misconstrued as about developing stereotypic knowledge of specific ethnocultural groups. Other concepts have been developed in view of this potential pitfall.

Cultural safety, arising from work of nurses with indigenous Mao_ ri population in New Zealand, emphasizes the need to recognize historical and ongoing processes of oppression; develop critical consciousness in reflecting on the potential harms in healthcare by the practitioner and health systems due to assumptions, biases, and structural issues; and prioritizing safety from the care recipients and their communities' perspectives.

Structural competence focuses on addressing the impact of social determinants of health on mental health, including promoting the capacity to recognize and address the impact of social structures, such as policies, economic systems, and institutions, on mental health problems, both clinically and in collaboration with other disciplines, communities, and organizations through advocacy and other actions. It is undergirded by developing structural humility to recognize the limitations of physicians to solve structural problems alone, and cultural humility to commit to self-reflection, self-critique, and lifelong learning from others.

Antioppressive practice focuses on disrupting all forms of oppression, such as racism, classism, genderism, heterosexism, and ableism, and addressing inequities. This includes developing a critical consciousness to identify, resist, and disrupt oppressive structures and forces; overcoming barriers to unlearn and relinquish power and privilege; and empowering oppressed individuals and communities to engage in advocacy and activism collaboratively through individual and collective actions, with the aim of promoting liberation of those being marginalized from oppression and systemic change.

The overlapping frameworks above, while varying in perspectives and emphases and each arising from a particular sociocultural and historical context, share foundational values of diversity, equity, and inclusion (DEI), increasingly acknowledged to be essential in medicine. Diversity refers to the recognition, valuing, and representation of differences that may exist among individuals and communities, including race, ethnicity, nationality, sex and gender identity, sexual orientation, (dis)ability, neurodiversity, and other social characteristics. Equity refers to the principle of fairness and that each person should have access to opportunities and resources to maximize their mental wellbeing in accordance with their needs. Equity is different from equality, where every person is treated the same without consideration of different needs caused by historical and ongoing systems of privilege and oppression. Inclusion refers to ensuring that each person is able to fully and meaningfully engage, participate, and have a voice in society, from receiving needed services to being involved in leadership and decision-making processes. These three interrelated values and principles need to be enacted and implemented together.

ETHNOCULTURAL DIVERSITY OF THE U.S. POPULATION

Defining Race, Ethnicity, and Immigration Status

Race has been variously defined and is a social construct that groups people together by superficial physical traits such as skin color and facial features. Some prefer the term racialized to emphasize the social process by which people become defined, which has been used throughout history as justification for exclusion, segregation, discrimination, racism, oppression, slavery, and other acts of violence and atrocities. Manifestations and sequelae of social and structural racism are often falsely misattributed to inherent racial differences, implying biologic differences without scientific evidence, furthering harm and perpetuating oppression. Instead, race-based data may be helpful when utilized to identify needs, inequities, and disparities.

Ethnicity, similarly, is a socially constructed group identity used to define groups of people and communities, rooted in perceived or actual commonality in history, ancestry, geography, language, religion, or other shared social or cultural group characteristics. Ethnicity may be self-assigned or attributed by others. Social and cultural changes, mobility and migration, intermarriages, and intermixing of groups continuously give rise to newly defined, mixed, multiple, or hybrid ethnic identities.

Both race and ethnicity have been variously defined and operationalized in different countries and contexts. This differs significantly even among Western nations with diverse populations.

For the 2020 U.S. Census, race includes White; Black or African American; American Indian or Alaska Native; Asian (which has been broken down in the latest census to include Chinese, Vietnamese, Native Hawaiian, Filipino, Korean, Samoan, Asian Indian, Japanese, and Chamorro); Other Asian; Other Pacific Islander; or some other race. Respondents can identify with one race alone or endorse multiple and can fill in more specific details about their origin under each racial category. Separately, respondents can self-identify their ethnicity as being of Hispanic, Latino, or Spanish origin, defined as ethnic groups originating from Mexico, Puerto Rico, Cuba, and other Spanish-language cultures including those in Central and South America.

In neighboring Canada, another settler country and the only Western country among the top 20 most diverse countries in the world, the 2021 long-form census includes a question that asks for the ethnic or cultural origin of the respondent’s ancestors linked to a list of over 500 examples. It has a separate question on whether the respondent identifies as First Nations, Métis, or Inuk (Inuit), and there is also a self-identification question with a list of selection based on the concept of visible minorities in Canada, including White, South Asian, Chinese, Black, Filipino, Arab, Latin American, Southeast Asian, West Asian, Korean, Japanese, and Other. While the concept of “visible minorities” has been used to promote employment equity in Canada, it has also been criticized, including by the United Nations, because the term’s “lack of precision may pose a barrier to effectively addressing the socio-economic gaps of different ethnic groups.”

As another comparison, in the 2021 Census of England and Wales, the list of ethnic groups included: Asian or Asian British (Indian, Pakistani, Bangladeshi, Chinese, or Any other Asian background); Black, Black

British, Caribbean or African (Caribbean, African, Any other Black, Black British, or Caribbean background); Mixed or multiple ethnic groups (White and Black Caribbean, White and Black African, White and Asian, Any other Mixed or multiple ethnic background); White (English, Welsh, Scottish, Northern Irish or British, Irish, Gypsy or Irish Traveller, Roma, Any other White background); Other Ethnic group (Arab, Any other ethnic group). Scotland and Northern Ireland use different ethnicity classifications for their census.

Evidently, race and ethnicity are arbitrary sociocultural constructs with localized rather than universal applicability or utility. Using them to stereotype an individual is inaccurate and harmful. Nevertheless, they both remain relevant constructs. For some individuals and communities, they may be an important and meaningful aspect of their individual or collective identities. In the interpersonal and social context, these constructs often influence social interactions, such as in the form of racism, and become part of the fabric of social reality with real impacts. At a systems level, disaggregated data by race and ethnicity can surface disparities.

Migrants: Defined broadly as an umbrella term by the United Nations, a migrant refers to any person who is moving or has moved across an international border or within a State away from their habitual place of residence, regardless of the person's legal status, whether the movement is voluntary or involuntary, what the causes of the movements are, and what the length of stay is.

Immigrants: An immigrant can be defined as a person living in a country other than that of their birth. There are four categories of immigration status: naturalized citizens, permanent or conditional residents, nonimmigrants, and undocumented immigrants. People may also migrate to another country without the objective of permanent residence, but rather, for short-term purposes of study or work with the intention of returning to their home countries; these are often referenced as migrants as opposed to immigrants.

The Changing U.S. Population: Race, Ethnicity, and Migration

Racial and Ethnic Diversity. There has been increasing diversity in the U.S. population as captured by the latest census in 2020. Respondents can choose to indicate one race, reported as "alone" in the census report, or multiple races, reported as "in combination." The White alone population accounted for 204.3 million people (61.6% of population), decreased by 8.6% since 2010. The Hispanic or Latino population grew to 62.1 million (18.7%), increased from 16.3% of the population in 2010. The total Black or African American alone and in combination population accounted for 46.9 million people (14.2% of the population), with a growth of 88.7% for the Black or African American in combination population since 2010. The Asian alone or in combination population accounted for 24 million people (7.2% of population). The American Indian and Alaskan Native in combination population grew by 160%, with the alone and in combination population totaling 9.7 million people (2.9% of the population). The Native Hawaiian and Other Pacific Islander alone or in combination population totaled about 1.6 million people (0.5% of the population). The percentage of people reporting two or more races has increased from 2.9% to 10.2% of the population.

Immigration. The United States, like Canada, Australia, and South Africa, is a settler colonial society. Except for indigenous populations, Americans are either immigrants or descendants of immigrants. From the Chinese Exclusion Act of 1882 to the Immigration Act of 1924, early U.S. immigration policies and laws were exclusionary and discriminatory against non-European populations until the Immigration and Nationality Act of 1965, which marked the abolishment of the National Origins Formula. Current U.S. immigration laws are based on the principles of allowing the reunification of families; admitting immigrants with skills that are valuable to the U.S. economy; protecting refugees; and promoting diversity, with a ceiling that family and employment-based immigrants from any single country cannot exceed 7% of the total number of people immigrating to the United States in a year.

During the first half of the 20th century, the two world wars and immigration restrictions resulted in a steady decline in the number of foreign-born Americans, reaching the low point of 4.7% in 1970. During the last three decades of the 20th century and continuing through the beginning of the 21st century, there has been an increasing inflow of immigration. Currently, the United States has the largest immigrant population globally. In 2020, it is estimated that 14.8% of the U.S. population was foreign-born. By 2060, the immigrant population is expected to rise to over 69 million, comprising 17.1% of the projected U.S. population of over 404 million. Viewed from the perspective of how long ago they immigrated, 27% of the total foreign-born population came to the country since 2010, 25% between 2000 and 2009, 21% between 1990 and 1999, 13% between 1980 and 1989, and 13% before 1980.

The immigration rates from source countries to the United States have changed significantly since the 1960s, in part due to the changes in the immigration system, as well as other drivers of migration such as economic, political, demographic, environmental, and sociocultural forces. In 1960, 84% of immigrants were born in Europe, Canada, or other North American countries, while only 6% were from Mexico, 4% from Asia, 3% from the rest of Latin America, and 3% from other areas. From 2011 to 2020, the top five source countries of immigration included: Mexico (14.6%), China (7.0%), India (6.2%), Philippines (4.9%), and Dominican Republic (4.9%). Based on the 2021 American Community Survey, of the 45 million foreign-born U.S. population, about 50% is from Latin America, 30% from Asia, and 11% from Europe, with the top ten countries of origin being Mexico (23.6%), China (6.1%), India (6.0%), Philippines (4.4%), El Salvador (3.1%), Vietnam (3.0%), Cuba (2.8%), Dominican Republic (2.8%), Guatemala (2.4%), and Korea (2.2%).

Starting in the 1960s, opportunities for temporary migration increased steadily in many high-income countries (HICs), including the United States, for temporary workers, students, and technicians. The most significant numbers were for seasonal agricultural workers and factory workers, followed by temporary workers in the service industry, mainly hotels, restaurants, and building maintenance. Between 1997 and 2013, the annual inflow of foreign-born temporary migrants to the United States exceeded the number of permanent immigrants. In 2022, about 3 million temporary workers came to the United States. They are a vulnerable group, as they may experience stress from loss of social status, discrimination, separation from families, and precarious and sometimes dangerous and abusive working conditions.

Refugees and Asylum Seekers. Voluntary migrants are people from rural areas migrating to urban industrial areas in their own and neighboring countries, motivated mainly by job opportunities in the industrializing cities. Involuntary migrants are people who left the regions and countries of their upbringing because of discrimination, persecution, revolution, and war, but followed the same paths of migration to other countries as other migrants. Involuntary migrants are frequently considered refugees.

A refugee as defined by the United Nations High Commissioner for Refugees (UNHCR) is a person who “owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion, is outside the country of his nationality, and is unable to or, owing to such fear, is unwilling to avail himself of the protection of that country.” An internally displaced individual is not considered a refugee. Refugees have the right to international protection. An asylum seeker is a person who has left their country and is seeking protection from persecution and serious human rights violations in another country but is not yet legally recognized as a refugee and is waiting to receive a decision on their asylum claim.

Of the world’s population, 3.6% are international migrants. Europe has the largest proportion of international migrants, 30.9%, followed by Asia, 30.5%, and North America, 20.9%. Africa, Latin America and the Caribbean, and Oceania have 25.4%, 14.8%, and 9.4% of the international migrant population, respectively. There is great variation as to the types of international migrants in a country. In 2017 in the United States approximately 23% were unlawful migrants and 5% were lawful temporary workers. By the end of 2021, according to UNHCR there were 89.3 million displaced people in the world of whom 27.1 million were refugees and 53.2 million were internally displaced by conflict.

The United States admitted 25,465 refugees in 2022, a significant increase from 2021 when the United States admitted only 11,411 refugees, the lowest since the Refugee Act of 1980. The peak year was 1992, with 132,531 refugees admitted. The leading five countries for recent refugees to the United States were the Democratic Republic of the Congo (42.9%), Syria (10.9%), Afghanistan (7.6%), Ukraine (7.0%), and Burma (6.8%). In 2021, the United States granted asylum to only 8,349 individuals, half the number granted relief in 2020. Nevertheless, this number represents an increase in the proportion of successful applications: in 2020, 71% of applications were denied, compared to 63% in 2021. The small number reflects a continued partial court shutdown due to the COVID-19 pandemic–by comparison, the United States granted more than 20,000 applications in 2019 and took in over 120,000 refugees. Before 1983, the United States received waves of Cuban and Indochinese refugees. Between 1985 and 1991, many refugees came from the former

Soviet Union, including Jews and evangelical Christians. From 1992 to 1998, refugees came from Bosnia, Kosovo, and other Balkan states due to the breakup of Yugoslavia. Liberians also were admitted as refugees until the end of their civil war. In 2019, the most common refugee home countries were Venezuela, China, Egypt, Turkey, and Russia.

RACISM AND MENTAL HEALTH

Racial and ethnic discrimination can cause both direct and indirect harm to mental health and can occur at the micro, meso, and macro level.

Interpersonal and Internalized Racism

At the individual micro level, interpersonal acts of racism include both overt actions, such as verbal and physical assaults, and covert acts or microaggressions, such as subtle acts of disrespect, slights, bias, misrecognition, or exclusion, which may or may not be intentional. Both types of racism have been shown to have adverse mental health impacts, and microaggressions are often denied and dismissed, with the victim being blamed as oversensitive, compounding a sense of shame and helplessness.

Based on data from the National Latino and Asian American Study (NLAAS) and the National Survey of American Life (NSAL), discrimination has been found to be associated with psychotic experiences among Latino, Asian, African American, and Afro Caribbean Americans, and Premenstrual Dysphoric Disorder among Asian, Latina, and Black premenopausal women. In the NLAAS, discrimination was associated with lifetime depression, anxiety, and any disorders among Asian Americans; 12-month depressive and anxiety disorders among Asian American women; and lifetime and 12-month depressive and anxiety disorders among Latino immigrants. The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) study found that among Hispanics, discrimination was a risk factor for mood, anxiety, or substance use disorder. The study found a dose-response relationship, where experiencing discrimination across a greater number of domains was associated with an increased likelihood of being diagnosed with a mental disorder. Similarly, the Study of Women Across Nation (SWAN) assessed discrimination experienced by women at six-time points over 10 years and found that Black, Chinese, Hispanic, and non-Hispanic White women who experienced the highest accumulation of discrimination over time, domain, and attributes reported the highest level of depressive symptoms.

A 2015 meta-analysis of 293 studies from 1983 to 2013 found that racism was associated with poorer physical and mental health, including depression, anxiety, and psychological stress. The association between racism and negative mental health was significantly stronger for Asian Americans and U.S. Latinos compared with African Americans, while the association between racism and negative physical health was significantly stronger for Latinos compared with African Americans.

Allostatic load refers to the accumulation of physiologic perturbations from repeated or chronic stress in daily life and has been found to predict increased cardiovascular events and all-cause mortality and may partly account for the increased mortality in Blacks compared to Whites. In a study with African Americans, perceived discrimination was associated with higher allostatic load. In another study, African Americans had a higher allostatic load than Whites. Results of a pathway analysis found that racial group was associated with perceived discrimination, which was associated with increased anger and decreased sleep quality, which were associated with allostatic load; in combination, discrimination, anger, and decreased sleep fully mediated the relationship between race and allostatic load.

In a recent review of studies employing ecologic momentary assessments (EMA), which provide intensive and detailed longitudinal information, race-based discrimination was found to be associated with same-day anxiety for all racialized groups. Associations with discrimination also included somatic symptoms among Asians, same-day depressive symptoms in the days following among African Americans, and next-day depressive symptoms among Latinos. Individuals whose racial identification was more central to their self-concept experienced higher depressive symptoms on the same day compared to those with lower racial centrality. Racial identity exploration exacerbated next-day distress, while racial identity commitment buffered the effect of daily discrimination (see later section on racial identity). Consistent with this, having a strong ethnic identity, as noted in a study of Filipino Americans, helps reduce the stress of discrimination and is associated with fewer depressive symptoms. Having a sense of ethnic pride, involvement in ethnic and communal activities, and having an emotional commitment to one’s ethnic group may be protective factors for mental health.

Racism may become internalized, impacting self-esteem as well as inducing negative perceptions about one's own racial or ethnic group. For example, colorism is the prejudicial treatment against dark-skinned individuals which can occur by in-group (i.e., by other Black people) or out-group (i.e., by Whites) individuals. Using NSAL data, a recent study found that in-group colorism among Black Americans was associated with greater odds of having lifetime psychiatric disorder, as well as substance use disorder, anxiety disorder, and eating disorder.

The myth of the model minority has been used to describe Asian Americans being stereotyped as a successful "intermediary minority" in the social hierarchy with the Whites at the top and the Blacks at the bottom. This has multiple deleterious effects, resulting in the negation of the long history and ongoing anti-Asian racism; ignoring the disparities faced by the diverse Asian populations; promoting a perpetual foreigner identity among Asians; becoming a source of stress and internalized racism for Asians; exacerbating anti-Black racism and perpetuating the denial of systemic racism by promoting a narrative of meritocracy and that hard work alone can lead to success; and promoting divisive intergroup racism.

The concept of intersectionality was first advanced by black feminist scholar Kimberlé Crenshaw. It provides a framework for understanding how multiple social identities, including race, gender, caste, ethnicity, sexuality, religion, disability, class, etc. can intersect and overlap in an individual, which can be empowering or oppressing depending on the social context. For many individuals, in addition to racism, they may encounter sexism, homophobia, classism, and ablism, finding themselves marginalized and oppressed in both the mainstream and their own cultural communities. A recent review of 40 studies on multiple forms of discrimination concluded that the current evidence suggests that multiple discriminated groups exhibit an increased risk of mental health problems, especially depressive symptoms.

Structural and Systemic Racism

At the meso level, or institutional or structural racism, racism is embedded in institutions and structures of the society, including laws, policies, institutional practices, and other norms that scaffold the society. At the macro level, or systemic racism, racism is embedded in the totality of the social system, including its political, legal, economic, health care, school, and criminal justice aspects, as well as all supporting structures, systems, cultural scripts, and peoples of the society. Inequitable structures and systems are mutually reinforcing and interconnected. Systemic racism is sometimes used as an overencompassing term for both meso and macro forms of racism (see Table 4.2–1).

Systemic racism is the result of the historical legacy of slavery and colonialism, as well as ongoing forms of systematized racism that continue to oppress and marginalize indigenous and minoritized ethnoracial groups, while privileging Whites. Examples include political disempowerment through gerrymandering and suppression of voting rights of people of color; segregation of neighborhoods resulting in concentrated economic disadvantages and limiting upward mobility through poorly resourced schools and limited employment opportunities; unfair financial practices, such as racially biased mortgage assessments and underfunded disproportionately segregated neighborhoods deemed hazardous for lending or investments ("redlining"); environmental injustice, such as locating coal-fired power plants and hazardous waste disposals in segregated communities; criminal justice system inequities, resulting in people of color representing about 39% of the general population but making up 60% of incarcerated individuals; police violence, leading to the death of 1 in every 1,000 Black men, with Black victims being more likely than White victims to have been unarmed; a "school-to-prison pipeline" education system, with harsher punishment for minorities in schools and higher involvement of police leading to expulsions and arrests.

Table 4.2–1.
Pathways Between Racism and Mental Health

Economic injustice and social deprivation

Examples include residential, educational, and occupational segregation of marginalized, racialized groups to low-quality neighborhoods, schools, and jobs (both historical de jure discrimination and contemporary de facto discrimination), reduced salary for the same work, and reduced rates of promotion despite similar performance evaluations

Environmental and occupational health inequities

Examples include strategic placement of bus garages and toxic waste sites in or close to neighborhoods where marginalized, racialized groups predominantly reside, selective government failure to prevent lead leaching into drinking water (as in Flint, MI, in 2015–2016), and disproportionate exposure of workers of color to occupational hazards

Psychosocial trauma

Examples include interpersonal racial discrimination, microaggressions (small, often unintentional racial slights and insults, such as a judge asking a black defence attorney “Can you wait outside until your attorney gets here?”), and exposure to racist media coverage, including social media

Targeted marketing of health-harming substances

Examples include legal substances such as cigarettes and sugar-sweetened beverages, and illegal substances such as heroin and illicit opioids

Inadequate health care

Examples include inadequate access to health insurance and healthcare facilities, and substandard medical treatment due to implicit or explicit racial bias or discrimination

State-sanctioned violence and alienation from property and traditional lands

Examples include police violence, forced so-called urban renewal (the use of eminent domain to force the relocation of urban communities of color), and the genocide and forced removal of Native Americans

Political exclusions

Examples include voter-restrictions (e.g., for former felons and through identification requirements)

Maladaptive coping behaviors

Examples include increased tobacco and alcohol consumption on the part of marginalized, racialized groups

Stereotype threats

Examples include stigma of inferiority, leading to physiological arousal, and an impaired patient–provider relationship

Reprinted from Bailey ZD, Krieger N, Agénor M, et al. Structural racism and health inequities in the USA: evidence and interventions. *The Lancet.* 2017;389:1453–1463. © 2017, with permission from Elsevier.

Systemic racism is also embedded in the mental healthcare system from its historical roots to the present time. One of the signers of the Declaration of Independence and considered the “Father of Psychiatry,” Benjamin Rush, taught medical students that black skin was a form of leprosy termed “negritude.” Psychiatrist Samuel Cartwright believed that enslaved Africans wanting to run away from captivity were suffering from drapetomania; their supposed disease of “rascality” or laziness was caused by dyaesthesia aethiopica; and that the treatment for both was whipping. The structural issues continue to the present day, resulting in demonstrable disparities arising from misdiagnosis, decreased access to care, and decreased quality of care. Observed disparities may be misattributed to biologic differences, a form of biologic determinism, or cultural differences, a form of cultural determinism. To address the latter, cultural competence is important, as culture exists in all, not only in “the other.” Rather than pathologizing the cultures of oppressed groups, psychiatrists need to identify and redress embedded cultural norms, assumptions, and biases that inherently perpetuate oppression and systemic racism, which need to be dismantled at the micro, meso, and macro levels. In addition to antiracism strategies, a cultural shift is needed in psychiatry, as “culture eats strategy for breakfast” in institutions.

Table 4.2–2.
Poverty and Ethnoracial Background

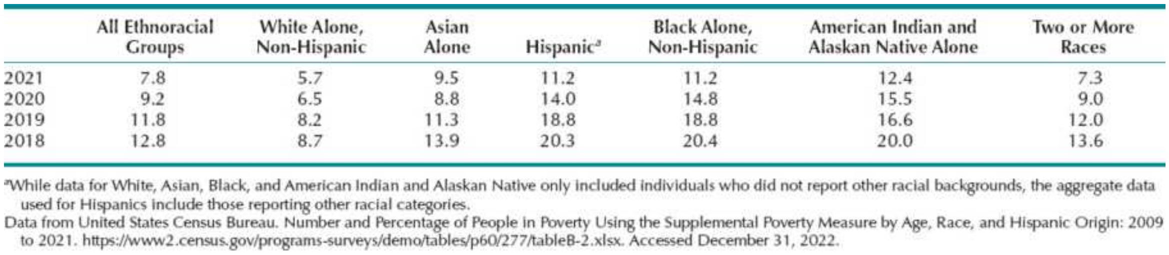

	All Ethnoracial Groups	White Alone, Non-Hispanic	Asian Alone	Hispanic[a]	Black Alone, Non-Hispanic	American Indian and Alaskan Native Alone	Two or More Races
2021	7.8	5.7	9.5	11.2	11.2	12.4	7.3
2020	9.2	6.5	8.8	14.0	14.8	15.5	9.0
2019	11.8	8.2	11.3	18.8	18.8	16.6	12.0
2018	12.8	8.7	13.9	20.3	20.4	20.0	13.6

[a] While data for White, Asian, Black, and American Indian and Alaskan Native only included individuals who did not report other racial backgrounds, the aggregate data used for Hispanics include those reporting other racial categories.

Data from United States Census Bureau. Number and Percentage of People in Poverty Using the Supplemental Poverty Measure by Age, Race, and Hispanic Origin: 2009 to 2021. https://www2.census.gov/programs-surveys/demo/tables/p60/277/tableB-2.xlsx. Accessed December 31, 2022.

Psychiatrists and mental health practitioners can play a key role as allies to fight against systemic racism. Increased awareness through self-reflection and self-education will be the first step, including learning about historical oppression, implicit bias, and unearned privileges that we hold as professionals and through our sociocultural identities. The fundamental DEI principles need to be integrated into the entire mental healthcare system, including psychiatric research, education, and clinical care. This includes ensuring diversity representation and inclusion among clinicians, academics, and leadership; implementing DEI and cultural competence training; and taking steps to improve access to culturally competent mental health programs. Much more focus is needed to address the inequities that impact the social determinants of health and their role in precipitating and perpetuating psychopathology. Advocacy needs to be championed at the clinical level, at the healthcare programs and organizational level, and at the social and health policy level. This work needs to be guided by humility, non-defensiveness, and willingness to collaborate, including engaging and empowering patients and their communities.

Economic Disparities

There is substantial disparity in rates of poverty when race and ethnicity are considered (Table 4.2–2). In 2021, the poverty rate for the total U.S. population was 7.8%, compared with 12.4% for American Indians/American Natives, 11.2% for Blacks, 11.2% for Hispanics (any race), 9.5% for Asian Americans, and 5.7% for non-Hispanic Whites, using the Supplemental Poverty Measure, which takes into account family resources and expenses as well as geographic variation.

Among median household incomes in 2020, Asian Americans had the highest median income, followed by non-Hispanic Whites, Hispanics, and Blacks. A Pew Research Center report analyzed income changes from 1970 to 2016 and estimated an income inequality gap, the gap between the top 10% and bottom 10% of income within a group. In 2016, Asians had the highest income inequality gap at 10.7 times, compared to Blacks (9.8×), non-Hispanic Whites (7.8×), and Hispanics (7.8×). Comparing among the lower-income earners in 2016, defined as the bottom 10% of each group, the lowest income was found among Blacks at 54% of Whites, followed by Hispanics at 66% of Whites, and Asians at 83% of Whites. Among the bottom 10% of each group, percentage gains in income from 1970 to 2016 were lowest among Asians at 11%, followed by Hispanics (37%), Whites (45%), and Blacks (67%). Overall, Hispanics' incomes have increased the least among the groups, likely due to immigration. In another study using longitudinal data from 1989 to 2015 to examine intergenerational changes in income, Blacks and American Indians/American Natives were found to have much lower rates of upward mobility and higher rates of downward mobility than Whites, leading to persistent disparities across generations. The Black–White gap is not explained by parental marital status, education, or wealth, and persists even among boys growing up in the same neighborhood.

Migration and citizenship status are also related to poverty. In 2020, the proportion of foreign-born Americans with incomes below the poverty line was approximately 13.4% compared to U.S.-born Americans at 11.1%. Immigrants who were permanent residents but had not yet become citizens had an especially high poverty rate of about 17.8%. On average, it has been estimated that it may take 10 to 15 years for immigrant families to make the transition from economic dependence to independence.

Social Determinants of Mental Health

Some populations experience an increased risk of mental disorders due to unfavorable social, economic, and environmental factors. These disadvantages and disparities may begin at birth and continue through the life course. Social determinants may be considered risk factors for mental disorders, many of which may be modifiable at the neighborhood, country, or cultural group level. Examples of social determinants include safe housing, transportation, and neighborhoods; racism, discrimination, and violence; education, job opportunities, and income; access to nutritious foods and physical activity opportunities; polluted air and water; and language and literacy skills. Figure 4.2–1 provides an overview of the major social determinants of mental health.

At the individual level, gene-environment interactions may modify the role of social determinants. An inverse relationship between social class and many mental disorders has been found. Education can impact economic prospects. Poverty may initiate and perpetuate mental disorders. Unemployment is considered both a risk factor for and a potential consequence of mental disorders. Food and housing insecurity are associated with economic disadvantage, and may also be associated with climate change, disasters, and war. Food insecurity may affect childhood academic performance, social skills, and other markers of poor mental health. Children in food insecure environments have been found to have increased symptoms of depression, anxiety, and other behavioral problems.

The original ACE study has brought attention to the impact of adverse childhood experiences (ACEs) on lifelong physical and mental health as well as other social outcomes. ACEs include childhood psychological, physical, or sexual abuse, or living with household members who abuse substances, have mental illness, are incarcerated, or exhibit violence against mothers. The initial study conducted involved mainly a White sample (79.8%). A subsequent study done in Philadelphia included a more diverse sample in an inner-city setting and expanded ACEs to include bullying, community violence, low neighbor safety, racism, and living in foster care. The study found that 83.2% reported at least one ACE, with Blacks experiencing significantly higher urban ACE indicators than Whites in witnessed violence, discrimination, adverse neighborhood experiences, and living in foster care. ACEs are associated with mental disorders, substance use, multiple sexual partners, suicide attempts, and worse physical health. Children with higher ACE scores are at a higher risk of being trafficked.

Exposure to violence–whether due to domestic violence, crime, disasters, or war–increases the risk of developing mental disorders, in particular post-traumatic stress disorder (PTSD). Lives lost to natural disasters occur in low- and middle-income countries (LMICs) in part due to poverty, living in unsafe housing, and weak social safety and support systems. Exposure to trauma has been shown in some studies to have an intergenerational impact. Attitudes toward mental health care vary at the individual level and among cultural groups and governments. The WHO 2020 Mental Health Atlas found that 35% of countries do not have stand-alone mental health laws. Lower-income countries spend disproportionality less of their health care budget on mental health.

MENTAL HEALTH OF RACIAL, ETHNIC, AND MIGRANT POPULATIONS

Common Mental Disorders among Minoritized Ethnoracial Populations

Multiple studies in the United States have examined the prevalence of mental disorders across ethnoracial groups. They have generally found lower prevalence of common nonpsychotic mental disorders among minoritized ethnoracial groups compared to White Americans. One of the largest sets of epidemiologic data comes from the Collaborative Psychiatric Epidemiology Surveys (CPES), which joins together three national representative surveys conducted from 2001 to 2003: the National Comorbidity Survey-Replication (NCS-R), The National Latino and Asian American Survey (NLAAS), and The National Survey of American Life (NSAL). The NCS-R is a nationally representative survey of a predominately non-Hispanic White sample (*n* = 9,282). The NLAAS is a survey of Asians (Chinese, Vietnamese, Filipino, and other Asians, *n* = 2,095) and Latinos (Mexicans, Cubans, Puerto Ricans, and other Latinos, *n* = 2,554). The NSAL included African American (*n* = 3,570), Afro-Caribbean (*n* = 1,623), non-Latino White (*n* = 1,006) adults, and Afro-Caribbean and African-American adolescents (*n* = 1,200).

In a 2019 analysis of the consolidated dataset of 21,024 Americans, combining CPES with the National Comorbidity Survey Reinterview (NCS-2), non-Latino Whites were found to have higher rates of mood, anxiety, and substance disorders (22.6%, 26.1%, 18.8%) than Asians (12.0%, 13.1%, 5.1%), Blacks (13.6%, 21.7%, 13.7%), and Latinos (17.3%, 19.8%, 16.2%), consistent with past analyses. These comparative differences remained significant after controlling for nativity, except for Latinos. However, analysis of the 2005 to 2012 U.S. National Health and Nutrition Examination Survey of 12,272 participants aged 40 to 79 years found a higher prevalence of depressive disorders among Latinos (11.3%) and African Americans (10.4%) than non-Latino Whites (7.1%). The more recent All of Us study, a national cross-sectional study collecting voluntary data from 214,206 Americans from 2018 to 2021, found lower prevalence of mood, anxiety, substance use, and personality disorder disorders among Asian, Black and African American, Hispanic/Latino populations compared to non-Hispanic Whites.

It has been proposed that the observed lower prevalence among minoritized ethnoracial groups may be due to the relative protective effects of non–majority-White cultural backgrounds, such as interdependence, familism, collectivism, or religious practices; the relative negative effects of mainstream U.S. society; or the “immigrant paradox” (see next section). Stronger minoritized ethnoracial identity, being foreign-born, and lower English proficiency as a proxy measure of lower acculturation were all negatively associated with mood disorders, while non-Latino White Americans with at least one parent born in the United States had a higher risk of mood disorder than non-Latino Whites who were foreign-born or both of whose parents were foreign-born. In addition, another sociocultural hypothesis is that non-Latino Whites with less education and limited job opportunities may have increased vulnerability compared to other minoritized ethnoracial groups in a system where there is a higher social expectation of success for Whites.

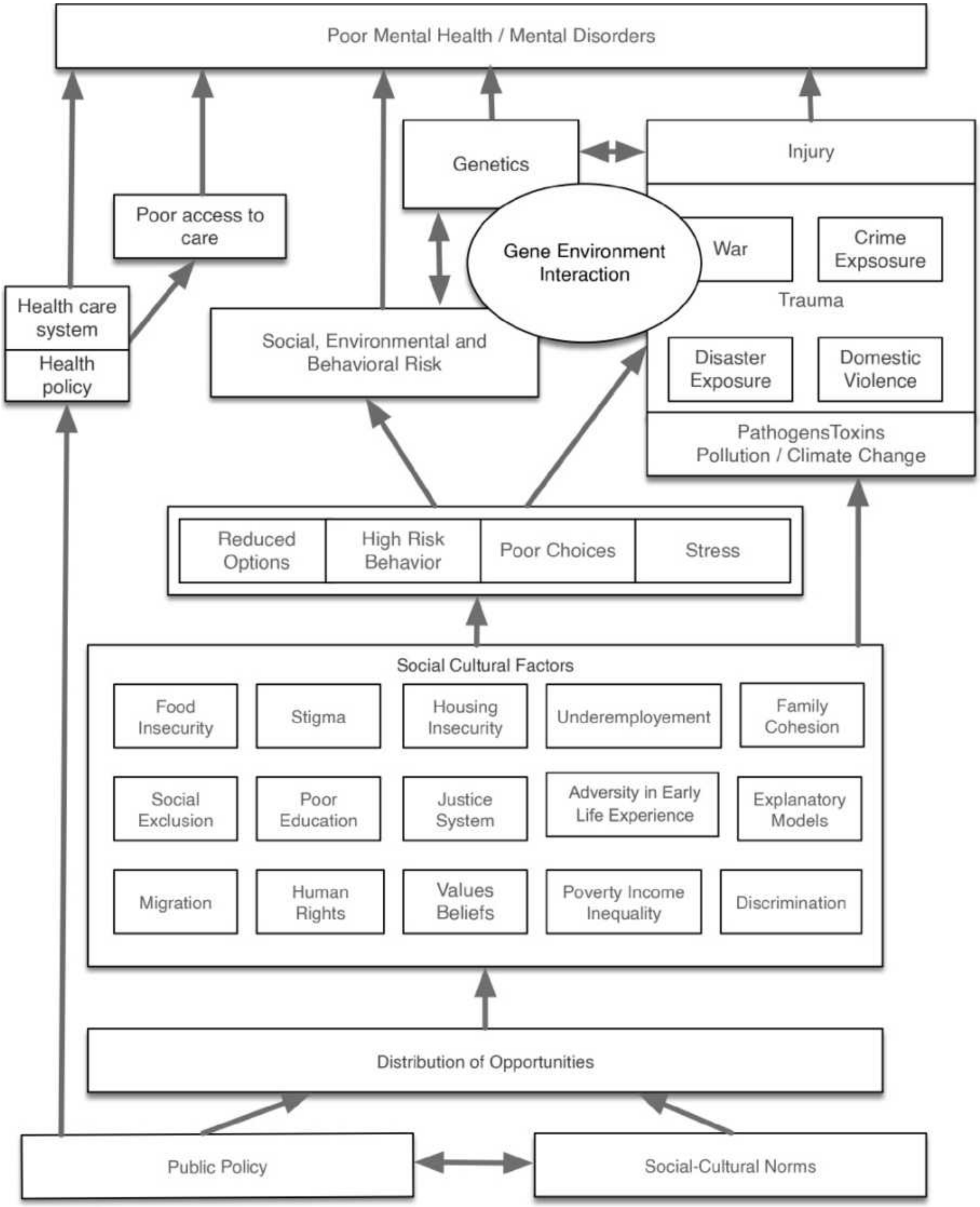
Poor Mental Health / Mental Disorders
Genetics
Injury
Poor access to care
Gene Environment Interaction
War
Crime Expsosure
Trauma
Health care system
Health policy
Social, Environmental and Behavioral Risk
Disaster Exposure
Domestic Violence
PathogensToxins Pollution / Climate Change
Reduced Options
High Risk Behavior
Poor Choices
Stress
Social Cultural Factors
Food Insecurity
Stigma
Housing Insecurity
Underemployement
Family Cohesion
Social Exclusion
Poor Education
Justice System
Adversity in Early Life Experience
Explanatory Models
Migration
Human Rights
Values Beliefs
Poverty Income Inequality
Discrimination
Distribution of Opportunities
Public Policy
Social-Cultural Norms

FIGURE 4.2–1. Social determinants of mental health. (Adapted and reprinted with permission Compton MT, Shim RS. The social determinants of mental health. *Focus*. 2015;13:420.)

Although there is a finding of lower prevalence, analysis of the combined CPES and NCS-2 dataset also found that Asian, Latino, and

Black populations with mood disorders were more likely to be persistent compared to non-Latino Whites. While a longitudinal study of 423 African American and White women from 1996 to 2009 did not replicate this finding, it did find a higher number of women with seven or more depressive episodes among African Americans compared to Whites. These differences, suggesting higher severity or chronicity among racialized populations, may reflect differences in access to care and/or the impact of ongoing stressors due to structural factors.

The NLAAS compared the prevalence of mental disorders among Puerto Ricans, Cuban Americans, Mexican Americans, and other Latinos. Puerto Ricans had the highest overall prevalence of mental disorders among the Latino groups assessed. Increased rates among Puerto Ricans were noted for depressive, anxiety, and substance use disorders. Mexican Americans had the lowest risk among the Latino groups for depressive disorders. Cuban Americans were at the lowest risk for substance use disorders and, among men, for anxiety disorders. Differences among the Latino groups in overall risk for lifetime mental disorders were also noted, with Cuban American men, Mexican American men, and “Other” Latina women having significantly lower risk than their Puerto Rican counterparts. The increased risk for mental disorders among Puerto Ricans compared to other Latino groups has been noted in other studies. More detailed analysis of disaggregated data of the CPES also found African Americans, Puerto Ricans, and Mexicans to have significantly higher chronicity of mental disorders compared to non-Latino Whites. Disaggregated data can yield a finer understanding of health disparities.

Common Mental Disorders among Immigrants

Although immigrants are vulnerable to stressors related to migration, acculturation, discrimination, and other psychosocial predicaments, the healthy immigrant paradox or the healthy immigrant effect hypothesizes that immigrants may have better physical and mental health than native-born individuals because of several factors. These include the somewhat stringent criteria for immigration, as well as selection bias of relatively healthier and more resourced individuals who apply for immigration. Research from the United States, unlike some studies from Europe, is largely consistent with this hypothesis.

The CPES studies generally found immigrants to have lower prevalence of mental disorders compared to U.S.-born individuals. The NCS-R found that immigrants, which included Hispanics (50.7%), non-Hispanic Whites (38.3%), and non-Hispanic Blacks (11%), had a lower lifetime risk of mental disorder than U.S.-born individuals (OR = 0.7), although immigrants who arrived in the United States at age 12 or younger had similar rates of mood and impulse control disorders as the U.S.-born. The NLAAS found that Asian immigrants had a lower risk of lifetime mood, substance, and any mental disorders compared to U.S.-born Asians, with U.S.-born Asian men being at greatest risk of substance use disorder and U.S.-born Asian women being at greatest risk of mood disorders. Among immigrant Asian men, low income and poor English proficiency were associated with a higher risk of mood disorders. The NLAAS found lower rates of most mental disorders among immigrant Latinos compared to U.S.-born Latinos. Disaggregated data found that the lower rates for depressive and anxiety disorders only applied to Mexicans, and lower substance use rates were observed among Mexicans, Cubans, and Other Latinos. No evidence for the immigrant paradox was found among Puerto Ricans. Puerto Ricans are U.S. citizens at birth and would not need to go through the U.S. immigration system. Scholars suggest that they should be classified as internal migrants rather than immigrants, and as noted in other studies, they may have relatively higher mental healthcare needs compared to other Latino immigrants. Data from the NESARC-II (2004 to 2005) confirmed these findings among Latino subgroups, adding that the risk of mental disorders among Mexican-origin respondents increased linearly with rising acculturation indices; these associations were in the same direction, but less consistent for Cuban Americans and other Latinos and no association with acculturation was found among Puerto Ricans. In the NSAL study, first-generation Caribbean Black immigrants had lower prevalence of mental disorders compared to the second- and third-generation U.S.-born Caribbean Blacks, with third-generation respondents having the highest rate. Among women, this difference was limited to substance use disorders.

The NESARC-III (2012 to 2013) found that DSM-5 lifetime mental disorders were significantly less prevalent among all immigrants compared to U.S.-born individuals: 20.3% and 34.9% respectively. The lower prevalence was found for generalized anxiety disorder, major depressive disorder, persistent depressive disorder, panic disorder, social phobia, specific phobia, and post-traumatic stress disorder, but not bipolar disorder. Immigrants who arrived as adolescents (ages 12 to 17) and as adults (age 18+) had significantly lower prevalence of mental disorders compared to the U.S.-born, but the difference was not significant for immigrants who arrived as children. Being in the United States for over ten years was associated with a higher risk of mental disorder compared to more recently arrived immigrants, although the risk for both immigrant groups was still lower than for the U.S.-born. Except for immigrants from Puerto Rico, immigrants from other countries in the study had lower rates of mental disorders compared to native-born, consistent with the NLAAS and NESARC-II findings; immigrants from Vietnam and South Korea also did not differ from their U.S.-born counterparts, which may be explained by smaller sample sizes.

Most recently, the All of Us study similarly found a lower prevalence of mood (OR 0.77), anxiety (OR 0.73), substance use (OR 0.41), stress-related (OR 0.80), schizophrenia (OR 0.57), and personality (OR 0.64) disorders among the non–U.S.-born populations compared to their U.S.-born counterparts.

As proposed by the healthy immigrant effect hypothesis, immigrants who arrive as children may not have an advantage over nonimmigrant children as they are not the main applicants for immigration. Further, child immigrants, as well as second-generation immigrants, are exposed to several adversities. Despite the potential advantages of better cultural adaptation and language skills than their parents, they may experience stress from encountering discrimination, such as being bullied at schools; experiencing social disadvantages such as poverty and growing up in disadvantaged neighborhoods; feeling caught between heritage and mainstream identities and cultures; erosion of protective aspects of their heritage culture or up-taking negative aspects of mainstream cultures, such as substance abuse; experiencing early parentification; or having higher parental pressures to succeed as immigrant families. A recent study using the NCS-R adolescent supplement found that compared to third-generation non-Hispanic White adolescents, there was a higher prevalence of mood and anxiety disorders among second-generation Asians (OR 2.51) and third-generation Hispanics (OR 1.28), but a lower prevalence among first-generation Asians (OR 0.27) and second-generation non-Hispanic Whites (OR 0.5). Behavioral disorders were lower among first-generation Asians (OR 0.26) and all generations of non-

Hispanic Black adolescents compared to non-Hispanic White adolescents. Research has also found that the risk of substance use disorder among Asians and Latinos increases across generations, possibly due in part to the loss of the protective effects of stronger cultural sanctions against substance use reflected in the lower rates of their use in China and Mexico (Table 4.2–3).

Schizophrenia and Psychotic Symptoms. For psychotic disorders, the Alameda County Kaiser Permanente California birth cohort study of 19,044 people followed over 1981 to 1997 found a 3.27-times higher risk of a schizophrenia-spectrum diagnosis for African Americans compared to Whites. Epidemiologic data on 16,423 Americans from the CPES found a higher lifetime prevalence of psychotic symptoms among Blacks (15.3%) and Latinos (13.6%) compared to non-Latino Whites (9.7%) and Asians (9.6%). Low ethnic identity has been found to increase the risk of psychotic symptoms. In a review of empirical studies in the United States from 1989 to 2013, African Americans and Latinos had over three times higher odds of being diagnosed with schizophrenia compared to Whites; rates were also increased for other psychotic disorders.

Table 4.2–3.
Risk for Mental Disorders among Migrants by Nativity and Number of Generations in the United States (Odds Ratios)

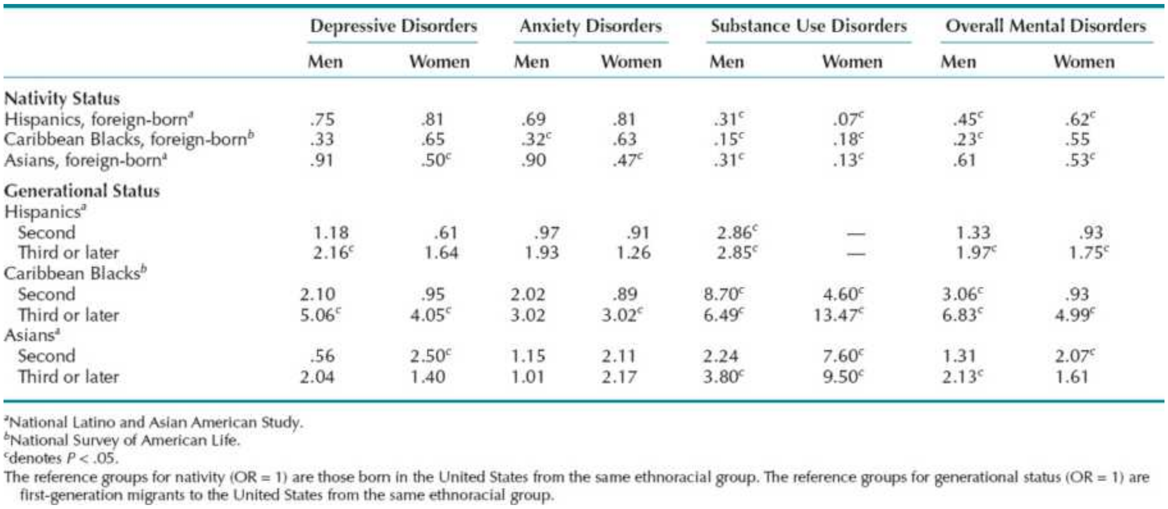

	Depressive Disorders		Anxiety Disorders		Substance Use Disorders		Overall Mental Disorders	
	Men	Women	Men	Women	Men	Women	Men	Women
Nativity Status								
Hispanics, foreign-born[a]	.75	.81	.69	.81	.31[c]	.07[c]	.45[c]	.62[c]
Caribbean Blacks, foreign-born[b]	.33	.65	.32[c]	.63	.15[c]	.18[c]	.23[c]	.55
Asians, foreign-born[a]	.91	.50[c]	.90	.47[c]	.31[c]	.13[c]	.61	.53[c]
Generational Status								
Hispanics[a]								
Second	1.18	.61	.97	.91	2.86[c]	—	1.33	.93
Third or later	2.16[c]	1.64	1.93	1.26	2.85[c]	—	1.97[c]	1.75[c]
Caribbean Blacks[b]								
Second	2.10	.95	2.02	.89	8.70[c]	4.60[c]	3.06[c]	.93
Third or later	5.06[c]	4.05[c]	3.02	3.02[c]	6.49[c]	13.47[c]	6.83[c]	4.99[c]
Asians[a]								
Second	.56	2.50[c]	1.15	2.11	2.24	7.60[c]	1.31	2.07[c]
Third or later	2.04	1.40	1.01	2.17	3.80[c]	9.50[c]	2.13[c]	1.61

[a] National Latino and Asian American Study.
[b] National Survey of American Life.
[c] denotes $P < .05$.
The reference groups for nativity (OR = 1) are those born in the United States from the same ethnoracial group. The reference groups for generational status (OR = 1) are first-generation migrants to the United States from the same ethnoracial group.

A 2011 meta-analysis of 21 population-based studies on migration and psychotic disorders conducted between 1977 and 2008 found that first and second-generation immigrants overall had 2.3- and 2.1-times higher risk of developing psychosis, respectively, compared with nonimmigrants; first and second-generation Black immigrants had 4.0- and 5.4-times higher risk. Respondents were from Australia, Canada, Denmark, Israel, Netherlands, Sweden, and the United Kingdom, but not the United States.

Most recently, a 2019 meta-analysis of 24 studies examining the effect of both ethnicity and migration on psychosis in the general population found that ethnic minorities (defined as ethnic groups representing less than half of the total population) conferred a 1.44-times higher risk for reporting psychotic symptoms and 1.36-times higher risk for reporting psychotic experiences. The latter refers to the presentation of milder psychotic symptoms not reaching the threshold of clinical relevance, operationalized in the meta-analysis as studies that excluded people with psychotic disorders. Among ethnic groups studied, the greatest risk for psychotic symptoms was observed among populations from the Maghreb and the Middle East in Europe (OR 3.3), Hispanics in the United States (OR 1.98), and Black populations (OR 1.85), all of which are populations which are disadvantaged or discriminated against in their local context. The increased risk experienced by ethnic minorities also differed between countries, including the United States (OR 1.59), the United Kingdom (OR 1.38), and the Netherlands (OR 2.74). Studies from HICs (OR 1.55) were higher than low and middle-income countries (OR 1.03), which did not have an elevated risk. Migration was only associated with an increased risk of delusional symptoms.

One of the factors contributing to the observed elevated prevalence of schizophrenia among Black individuals is diagnostic bias, whereby psychotic disorders are overdiagnosed and affective disorders are underdiagnosed, a finding supported by multiple studies over the years. For example, in a study of admitted patients with psychotic symptoms and an expert-consensus diagnosis of a DSM-IV affective disorder, African-American patients were more likely than other patients to be diagnosed with a schizophrenia-spectrum disorder by clinical assessment or structured interviews. A similar pattern of overdiagnosis has been found in various other clinical settings. In a study of 610 in- and out-patients at academic medical centers, African Americans had a higher prevalence of schizophrenia diagnoses even after controlling for the rate of severe affective disorders. A recent study found that among patients with a diagnosis of schizophrenia at a community outpatient clinic, African Americans were 8.2 times more likely than non-Latino Whites to have moderately severe to severe depressive symptoms, suggesting possible misdiagnosis. The race of the clinician may also influence the diagnostic decision-making process. For example, in an inpatient study, clinician's race was not found to impact the prevalence of a schizophrenia diagnosis, but African American clinicians tended to associate positive symptoms with schizophrenia, while non-African American clinicians tended to associate negative symptoms with schizophrenia. In a 2018 meta-analysis of 55 studies from 1983 to 2017, Black Americans were found to be diagnosed with schizophrenia 2.42 times more frequently than White Americans. The use of structured (OR 1.59) over unstructured (OR 2.43) diagnostic assessments appeared to have a small though not statistically significant impact in decreasing this elevated risk. The study year was not a significant predictor of diagnostic disparities, suggesting that there has been no improvement over time despite research attention to this issue for years.

In addition to diagnostic bias, most researchers agree that the observed higher risk of psychosis among migrant and minoritized ethnoracial groups is related to the interrelated effect of social determinants of health and structural racism. In fact, many have argued that overemphasis on diagnostic biases in the literature may detract from a focus on structural causes for higher rates. The elevated risk for a schizophrenia-spectrum diagnosis among African Americans relative to Whites in the Kaiser Permanente California study was reduced to 1.92 after adjustment for socioeconomic status indicators at birth. In U.S. studies, perceived discrimination is associated with elevated psychotic symptoms in the general population, and police discrimination is associated with lifetime increases in psychotic experiences among Black individuals. A 2013 meta-analysis found that the prevalence of psychotic disorders was higher in low-ethnic density areas compared to high-ethnic density areas, implicating the deleterious effect of racism. Recently, a model was proposed to comprehensively examine and address the relationship between structural racism and psychosis in the U.S. context, noting especially the increased exposure to social risk factors in Black and Latino populations. Multiple exposures to neighborhood-level disadvantage, individual-level social disadvantage, collective trauma, and everyday discrimination may lead to cumulative stress, which, through genetic vulnerability, perinatal complications, and altered neurobiology, may result in various forms of psychosis.

MENTAL HEALTH CARE DISPARITIES FOR MIGRANT AND MINORITIZED ETHNORACIAL POPULATIONS

Mental Healthcare Utilization

Lower service utilization rates among U.S. racial and ethnic minorities compared to Whites have been the norm in research studies over several decades, with various access barriers and the lack of culturally competent services being implicated.

In the NLASS study, among Asians, only 8.6% of the sample and 34.1% of the subset with a probable DSM-IV diagnosis of mood, anxiety, or substance disorder in the past 12 months utilized mental health-related services, compared to 17.9% and 41.1% respectively in the total NCS-R sample. In the Latino sample, the utilization rates among Puerto Ricans, Cubans, and Mexicans were 19.9%, 11.5%, and 10.1%, respectively; these rates increased to 43.0%, 41.9%, 34.2% among respondents with probable DSM-IV diagnoses in the past 12 months.

Subanalyses of the Asian sample found that U.S.-born Asians were more likely to utilize services than foreign-born; for example, third-generation individuals with a probable diagnosis had a high utilization rate of 62.6%. Subanalyses of the Latino sample found that nativity, language, age of migration, years of residence in the United States, and generational status were associated with mental health care use only for those without a probable diagnosis; in this subgroup, Puerto Ricans were significantly more likely to use services than Cubans, Mexicans, and other Latinos. Another major factor determining mental healthcare use was insurance. For those with a probable psychiatric disorder, only 19.1% of uninsured Latinos used services, compared to 38.6% with private insurance and 51.6% with public insurance.

In the NSAL, most African Americans with major depression did not seek mental health care. Only 45.0% of African Americans and 24.3% of Caribbean Blacks received treatment of any type. In addition, mental health service utilization did not significantly increase as symptom severity increased. Only 48.5% of African Americans and 21.9% of Caribbean Blacks with severe or very severe symptoms received any psychiatric treatment.

A study using national representative data from the 2004 to 2012 Medical Expenditure Panel Survey (MEPS) found that access to any mental health care increased from 16% to 20% for non-Hispanic Whites, 8% to 9% for Hispanics, 7% to 10% for Blacks, and from 3% to 5% for Asians between 2004 and 2012. The Black/White and Hispanic/White disparities are widening in accessing mental health care and use of psychiatric medications; Asians consistently show the lowest mental healthcare utilization. Overall, the study found no reduction in disparities for any mental health care use, outpatient mental health care, or access to psychiatric medications after controlling for clinical needs.

A study on the impact of the 2008 Mental Health Parity and Addiction Equity Act (MHPAEA) and the 2010 Patient Protection and Affordable Care Act (ACA) examined changes in healthcare utilization patterns from 2006 to 2014, including ethnoracial and gender differences. The 2014 mental health visits for men and women respectively were: 6.7% and 9.2% for Whites; 4.7% and 6.3% for Hispanics; 5.7% and 7.5% for Blacks; and 1.6% and 3.5% for Asians, with no significant difference observed between 2006 and 2014. Subanalyses showed that utilization by Asian men significantly decreased from 2012 to 2014 by 2.4%.

Disparities in healthcare utilization have also been found in individuals with suicidal behaviors. In a study using the National Survey on Drug Use and Health from 2009 to 2016, among suicide attempters, outpatient care usage was lower among non-Hispanic Blacks, Hispanics, and multiracial individuals compared to non-Hispanic Whites, and Hispanics were also less likely to receive inpatient care. Among individuals with suicide ideation but no attempts, outpatient care usage was lower among non-Hispanic Blacks, Hispanics, and Asian, Hawaiian and Pacific Islanders compared to non-Hispanic Whites. The overall pattern suggests that ethnoracial minorities underutilize outpatient services until inpatient treatment is required.

Disparities in Mental Health Treatment

Many studies have consistently documented disparities in the quality of mental health care received by minoritized ethnoracial groups.

A 2016 systematic literature review summarized disparities in the diagnosis and treatment of major depressive disorder among ethnoracial minorities. Consistent with epidemiologic studies, clinical research generally found lower rates of diagnosis of depression, attributable in part to underdiagnosis by clinicians, differences in symptom presentation or help-seeking behaviors, and cultural insensitivity of some psychometric measures. In terms of treatments for depression received, African Americans and Hispanics were less likely to receive adequate care according to evidence-based guidelines compared to Whites. Undertreatment of depression was similarly found in various subpopulations including older adults, adolescents, and postpartum women, and in various healthcare settings, including primary care. African Americans and Hispanics were less likely to receive antidepressants, less likely to fill these prescriptions, and more likely to discontinue medications prematurely. There were similar findings of undertreatment among Asians. African American inpatients were less likely to receive ECT. Utilization of psychotherapy was more mixed; many minoritized ethnoracial groups reported preferring therapy over medications when given the choice and if language and insurance coverage were not barriers. The 2000 Medical Expenditure Panel Survey found that African Americans were more likely than Hispanics and Whites to receive adequate psychotherapy. Hispanics were less likely than Whites to receive counseling during primary care visits, and in another study, Hispanics without adequate insurance coverage to supplement Medicare were less likely to receive psychotherapy. In a depression study conducted in the Veterans Administration system, which would have mitigated against coverage differences, there were no differences in receipt of adequate psychotherapy by race or ethnicity.

A 2017 review of 131 studies summarized key findings on diagnostic, treatment, and outcome disparities affecting ethnoracial groups with severe mental illness (SMI), including schizophrenia, bipolar disorder, and recurrent depression. In medical settings, Black respondents received higher doses of antipsychotics, higher use of depot antipsychotics, and lower use of second-generation antipsychotics (SGA). In terms of outpatient treatment, Blacks in the Veterans Administration system were less likely than Whites to have follow-up visits after an index diagnosis of bipolar disorder or to receive regular follow-up visits after hospitalizations. Both Blacks and Hispanics/Latinos had lower use of case management services. Regarding the quality of care, Blacks, compared to Whites, were more likely to be involuntarily hospitalized, use crisis services, enter emergency treatment by means of law enforcement, receive higher doses of antipsychotic medications, and less likely to receive minimally adequate treatment for SMI, defined as medication management and at least four follow-up visits, or receive treatment for comorbid depression. In treatment engagement, racial and ethnic minorities were found to have a higher risk of premature treatment discontinuation. In terms of outcomes, Blacks showed less improvement in psychotic symptoms, were less likely to return to work, and experienced less improvement in functioning at 1-year post-discharge.

Several recommendations converge from various reviews to better address disparities in providing care for ethnoracial minorities. Upstream social determinants of health need to be addressed to prevent mental illness and enable ethnoracial minorities to receive care. Stigma and health literacy need to be addressed through community engagement strategies using culturally appropriate messaging. Increasing diversity representation in the healthcare system and cultural competency training may increase trust in the system over time. Service development and planning need to be inclusive of the voices of patients, families, and communities, including involving individuals with lived experience in the delivery of care. Treatment interventions, programs, and services need to be culturally competent, and more interventions need to be developed involving families and communities. Spiritual and religious values and practices need to be taken into consideration and integrated into care.

MENTAL HEALTH OF INDIGENOUS PEOPLES OF THE UNITED STATES

For thousands of years, millions of Indigenous Peoples lived as diverse Nations in North America until colonization around 600 years ago by European powers and later the United States government, which sanctioned massacres, forced removals, decimated traditional food systems, and introduced infectious diseases to the populations. From the 1950s to 1970s, policies promoting relocation and forced assimilation were imposed on American Indians/American Native (AI/AN) communities. This included forcibly removing children from their homes into residential schools and foster care services through child protection services to isolate them from their cultures, exposing many to various forms of physical, sexual, or emotional trauma. In 1978, the Indian Child Welfare Act prevented the federal government from continuing to take children away from their homes and the American Indian Religious Freedom Act prevented federal programs from further prohibiting AI/AN religions and cultures. By then, generations had been impacted.

Currently, there are 9.7 million U.S. AI/AN people (i.e., people who report AI/AN background as their only identity or one of several identities) living in the United States, comprising 574 recognized tribes and representing about 2.9% of the total U.S. population. While there are up to 245 indigenous languages recognized in the United States, 65 are extinct and 75 are in jeopardy. Life expectancy at birth of AI/AN in 2021 was 65.2 years compared to 76.1 years for the general U.S. population. AI/AN people are five times more likely to die of alcoholism and liver disease, and twice as likely to have diabetes mellitus compared to Whites. The indigenous population also has higher rates of accidents, violence, suicide, and homicide than the majority U.S. population. Suicide rates are higher among AI/AN groups compared to Whites, with three times the rate of death by suicide among those under age 25. Among all ages and sexes, the death rate due to mental and behavioral disorders for AI/AN is 1.10 to 1.53 times higher than for Whites; between the ages of 15 to 24, AI/AN men have 2.38- to 2.69-times higher risk of death than African Americans. One-third of the Native American population does not receive regular health care.

Studies indicate that AI/AN are at a higher risk for mental disorders. A nationally representative psychiatric epidemiologic study, NESARC, which included a small sample of AI/AN (*n* = 701), found that they have higher rates of any psychiatric disorder, any mood disorder, and any substance use disorder compared to non-Hispanic Whites.

Several studies have focused on the prevalence of mental disorders among specific groups of AI/AN. The American Indian Service Utilization, Psychiatric Epidemiology, Risk and Protective Factors Project (AI-SUPERPFP) worked with Southwest Indian and Northern Plains Indian communities. Compared to other ethnoracial groups, AI/AN was at higher risk of alcohol use disorders and PTSD but at lower risk for major depressive disorder. Among individuals exposed to DSM-IV–defined traumatic events, 15.9% qualified for a lifetime diagnosis of PTSD, comparable to other U.S. studies, suggesting that the higher prevalence rate of PTSD among AI/AN may be due to higher traumatic exposure.

Studies of American Indian Vietnam Veterans from the Northern Plains and the Southwest also found very high rates of PTSD: 31% and 27% for current prevalence and 57% and 45% for lifetime prevalence, respectively. This rate was significantly higher than among Vietnam war veterans of other ethnoracial groups, including Whites, African Americans, and

Japanese Americans. AI/AN veterans were more likely than other groups to have been exposed to war-zone stress in Vietnam, including combat exposure, being wounded, and having noncombat injuries. This higher exposure to war-zone stress accounted for the high rate of PTSD among AI/AN veterans compared to other ethnoracial groups. These veterans also had higher prevalence of alcohol use disorders: over 70% (current) and 80%% (lifetime), compared to much lower prevalence for other ethnoracial groups: 11% to 32% (current) and 33% to 50% (lifetime).

AI/AN children were more likely to receive mental health treatment through the juvenile justice system and inpatient facilities than non-AI/AN children. In 2017, 752 AI/AN youths lived in residential placement, which included detention centers, group homes, and boot camps, according to the National Center for Juvenile Justice. Among adults, AI/AN were more likely to seek help for substance use problems than Whites but less likely to seek help for depressive and anxiety disorders. Less than 30% of those with a mental disorder obtain help from mental health service providers.

Indigenous Historical Trauma (IHT), a synthesis of historical oppression and psychological trauma, is an important concept proposed to help understand the trauma experienced by AI/AN arising from colonization and oppression. Historical trauma is preferred over the term racial trauma as the indigenous identity is closely related to tribal identity, which is an ethnic rather than racial identity. The distinguishing core features of IHT are that it is colonial in origin, collective in impact, cumulative across adverse events, and crossgenerational in the transmission of risk and vulnerability.

A recent systematic review of IHT was conducted on research from the United States and Canada. Although AI/AN consist of diverse and distinct groups of people living throughout North America, they share similar experiences of colonial history and oppression, such as residential schools, with the notable difference that U.S. schools were substantially reformed in the 1930s while Canadian schools were unreformed until the 1970s. Most studies report IHT to be associated with adverse mental health outcomes, but the results are complex and specific, making it challenging to formulate definitive conclusions. Not all AI/AN attended residential schools, nor have family members who did. Among those who attended, including some who experienced abuse, not all reported being adversely impacted. While there can be further refinements in the methodology to conduct research into IHT, the concept should be used as a metaphorical understanding of the complex nature of the trauma experienced by AI/AN rather than as a literal and discrete psychological construct.

Another recent systematic review identified a dearth of empirical studies on mental health interventions that specifically address IHT, revealing mixed results. Findings suggest that sharing IHT memories might be cathartic; treatments integrating traditional healing and tribal practices might be helpful, and incorporating ceremonies into interventions might provide a culturally safe and affirming space for healing.

Clinicians should be aware of the potential role of IHT, the diversity among AI/AN communities and individuals, as well as the social determinants of health impacting AI/AN experiences, including poverty, lack of infrastructure, and ongoing racism. Individual treatment and service planning need to be attuned to their social, cultural, and spiritual needs. Possible interventions include promoting connections with extended families and cultural networks; collaborating with communities and elders including the incorporation of cultural and spiritual teachings as well as traditional knowledge and rituals; and doing advocacy work to address disparities in the social determinants of health and promote indigenous rights.

At a systems level, the Indian Health Services (IHS), established in 1955, currently provides healthcare to 2.6 million AI/AN through a combination of federal IHS programs, Tribally Operated Health Care Services, and Urban Indian healthcare services and resource centers. In many service areas, only about 7% of IHS federal funds are allocated to mental health. Compared to Whites, AI/AN report higher dissatisfaction with the quality of care received and poorer communications with their healthcare providers. Many AI/ANs who moved to the cities encouraged by various federal policies over decades have encountered difficulties accessing IHS. Surveys by AI/AN advocacy groups indicate that better care is received from tribally run programs than those under federal IHS management. The Substance Abuse and Mental Health Services Administration (SAMHSA) administers substance abuse prevention and mental health service grants, including grants to eligible AI/AN entities. SAMHSA has recently launched Circles of Care and Native Connections programs, which provide grant funding for AI/AN-focused services. Overall, key issues that still need to be addressed include underfunding, limited human resources, and lack of culturally and spiritually appropriate services.

A well-known example of a successful indigenous initiative is the Alkali Lake community in Canada, where alcoholism decreased from 95% to 5% over 10 years, and the story is captured in a film. Another example is the work of White Bison, a nonprofit grassroot organization in Colorado Springs, which published The Red Road to Wellbriety, a cultural competent text combining teachings of the Medicine Wheel and 12-step programs.

MENTAL HEALTH OF REFUGEES

Unlike those who emigrate voluntarily, often for primarily economic reasons, refugee populations have been highly traumatized prior to migration and are at a higher risk for mental disorders including depression and PTSD. For example, the NLAAS sample included 660 refugees. Premigration onset of PTSD was 4.86 times higher for refugees (3.5%) compared to voluntary migrants (1.2%), while there was no significant difference in the onset of PTSD postmigration.

Among Southeast Asian refugees, premigration trauma was a significant factor in predicting psychological distress even 5 years after migration to the United States. Some studies have suggested that up to 70% of Vietnamese, Cambodian, and Laotian refugees have PTSD. A community survey among Cambodian refugees found that 45% had PTSD and 51% suffered from depression. Adolescent Cambodian refugees also have high prevalence of depression; 41% had depression even ten years after their traumatic exposure. Among adult refugees arriving in the United States, Vietnamese refugees are at higher risk for depression than Chinese refugees. Even two decades after resettlement, Cambodian migrants have very high rates of PTSD (62%) and major depression (51%). Over 70% experienced violence after migration to the United States. A dose-response relationship has been observed between an increasing degree of exposure to trauma (both premigration and postmigration) and an increased likelihood of a current mental disorder. The most extensive study on Cambodian refugees using a structured diagnostic instrument found that women were no more likely than men to develop PTSD or depression.

The Bhutanese is one of the largest groups of South Asian refugees in the United States; about 90,000 arrived between 2008 and 2012. The prevalence of depression was reported by the CDC to be 21%, with a high suicide rate of 21.5 per 100,000. Being a family provider, self-reporting poor health, and being illiterate in their native language were risk factors for depression. In another recent study of 225 Bhutanese resettled in Massachusetts, 24% and 34.2% had moderate to severe symptoms of depression and anxiety, respectively. High resilience was associated with reduced symptoms of depression and anxiety, with personal competence accounting for most of the effects.

Approximately 91,000 Iraqis refugees arrived in the United States from 2008 to 2013. Most refugees prior to resettlement came from dispersed urban areas in Jordan, Syria, Iraq, Turkey, and Lebanon, rather than through refugee camps. Among Iraqi refugees living in the United States between 8 and 36 months, 27% were depressed and 31% had PTSD. In another study, among a sample of Iraqi refugees assessed 1 year after arrival, 9% reported having been kidnapped; on the SCID-1, 4% met PTSD criteria and 3% met MDD criteria. Those who experienced kidnapping had a higher prevalence of PTSD, but not MDD.

The Syrian conflict began in 2011, resulting in one of the largest refugee crises since World War II, with about 6 million having fled the country. Most have fled to neighboring Lebanon, Jordan, and Turkey, as well as Europe. The United States has taken in more than 23,000 Syrian refugees since 2011. Syrian refugees (n = 157) screened at primary care in Michigan a month after their arrival had a high risk of suspected PTSD (32%), depression (47.7%), and anxiety (40.3%). In a recent systematic review of 15 studies, including 8,176 Syrian refugees resettled in 10 countries, comparable prevalence was found for post-traumatic stress (43.0%; 23.4% to 83.4%), depression (40.9%; 20% to 44.1%), and anxiety (26.6%; 19.30% to 31.8%); the wide ranges were due to diverse research methodologies and country settings. Like Somali and Iraqi refugees, most Syrian refugees are Muslim, and they may be at risk of encountering discrimination due to Islamophobia and challenges with a healthcare system that is not culturally or religiously suitable for their needs.

A 2020 meta-analysis of 26 studies on refugees and asylum seekers around the world reported a prevalence of PTSD at 31.5%, depression at 31.5%, anxiety disorders at 11%, and psychosis at 1.5%. Similarly, in a 2022 meta-analysis of 40 studies on refugees and asylum seekers, the prevalence of PTSD was 31%, MDD 32%, bipolar disorder 5%, and psychosis 1%. Subgroup analyses showed a higher prevalence of MDD in studies conducted in LMICs compared to HICs.

The role of mental health professionals is frequently not well understood by refugees. For example, while Congolese refugees were found to have many mental health problems, including depressive, anxiety, and substance use disorders, as well as domestic violence, many declined mental health services. However, in a controlled trial of psychotherapy with Congolese survivors of sexual violence, group psychotherapy was more effective than individual support in decreasing depression and anxiety symptoms, PTSD symptoms, and functional impairment. This suggests that it may be challenging to build safety and trust in a traditional one-on-one approach with some refugees, while it may be more efficacious to facilitate the collective capacity to heal and develop resilience through group approaches.

Another study of refugees from Bhutan, Burma, Ethiopia, and Somalia found that refugees were willing to report their symptoms when these were described within their political context in focus groups; moreover, their PTSD symptoms varied across cultural settings. The study highlighted the need to ask refugees about their symptoms in context, including political history and struggles and the impact of symptoms on their functioning. Across groups, there were common reports of cognitive symptoms, described as “thinking too much,” and physical symptoms, such as head, stomach, and body pain. Expression of emotional experiences was culturally specific, such as Oromo refugees describing feeling “burning from the inside” and Somalis describing having “air in the brain” or feeling “not dead or alive.” These expressions highlight the combination of universal and culturally specific ways of experiencing and communicating distress.

The premigration life of refugees from war-torn developing countries ill prepares many for the difficulties encountered upon migration, due to limited occupational skills and educational attainment, barriers to cultural integration including language, and significant premigration health and mental health problems. These difficulties often place them in a continuing position of psychosocial vulnerability and at risk for chronic mental illness. Premigration traumatic experiences and postmigration stress are associated with PTSD and depressive or anxiety disorders.

Postmigration factors may mediate the negative impacts of premigration stressors and may have even greater impact on mental health than premigration stressors. These include detention, harsh living conditions, financial stress, length of the asylum process, legal status, sense of belonging, family separation, social isolation, discrimination, and cultural adaptation. Though often studied together with refugees, asylum seekers appear to be at an increased risk of developing mental disorders compared to refugees due to their precarious legal status. In a systematic review of postmigration social and environmental factors facing asylum seekers, discrimination and postmigration stress were linked with increased risk of mental disorders, underscoring the need to provide support for the postmigration settlement process.

For premigration trauma, studies have found empirical support for cognitive behavior therapy, narrative exposure therapy, eye movement desensitization and reprocessing, and other forms of psychotherapy. There have also been a few studies describing the effectiveness of community-based psychosocial programs. Clearly, there is a need to develop an effective and integrated approach to refugee care that takes into consideration their complex needs beyond narrowly focusing on individual psychopathology. A Multitier Mental Health and Psychosocial Support (MHPSS) model has been adapted for humanitarian settings, which provides mental health care for resettled refugees. The model includes four tiers: (1) social adjustment and integration to provide basic services through resettlement programs; (2) family and community support through community-based psychosocial interventions; (3) bereavement and trauma healing through trauma-focused interventions; and (4) specialized mental health treatment through psychiatric services. Being culture-informed and trauma-informed are the two central pillars of the model across tiers, highlighting their critical importance to refugee care. This approach also emphasizes the need to breakdown silos between tiers and between services and the need to engage families and communities in collective empowerment and healing.

CULTURE AND MENTAL HEALTH

Although culture impacts the daily mental functioning of every person, it is often dismissed as irrelevant or only pertaining to racialized or minoritized groups. This presumption has led to the development of theory and knowledge about normal mental functioning that is largely defined by and founded on Western Eurocentric cultural frames of reference and norms. Joseph Henrich has argued that most basic psychological research is based on Western, Educated, Industrialized, Rich, Democratic ("WEIRD") populations, often from White college samples, despite the fact that many of their basic psychological measures such as visual perception, spatial reasoning, categorization, inferential induction, moral reasoning, and self-concepts are outliers when compared to data from most other world populations. Recognition and systematic study of sociocultural differences are key steps toward advancing scientific, clinical, and structural work in mental health.

When exploring group-level cultural differences, a key principle of cultural psychiatry is to avoid stereotyping, which can give a false sense of expertise. Every individual is unique, and their mental wellbeing is affected by their own life experiences as well as particular biologic, psychological, social, cultural, spiritual, and other environmental factors, interacting across time and contexts. Thus, culture should neither be ignored nor used as a simplistic categorical tool to label, stereotype, or stigmatize individual or groups but, rather, as an important lens to promote deeper understanding of the complexity of human suffering and aid the development of more effective approaches to address it.

Cultural Dimensions

Hofstede Cultural Dimensions. One approach to examining sociocultural group differences is the use of cultural dimensions. Geert Hofstede pioneered this in studying IBM employees across different countries, and the model elaborated by him and others consists of six dimensions: individualism-collectivism, defined as the strength of the ties between community members, with individualism focusing more on individual goal attainment and collectivism focusing more on group goals and wellbeing; power distance, the degree to which differences between those with and without power exist and are accepted; uncertainty avoidance, the level of tolerance to ambiguity; masculinity-femininity, the degree of overlap of gendered roles and expectations and the valuing of assertiveness versus modesty; long-term versus short-term orientation (LTO), the relative emphasis on pragmatism, adaptation, perseverance, and thrift (long-term) versus adherence to traditions and principles and focus on short-term gains (short-term); and indulgence versus restraint, the degree of permissiveness to satisfy individuals' desires versus regulation by social norms. This model remains one of the most influential and empirically tested models in crosscultural studies.

At the country level, the U.S. profile on Hofstede's model scores highest on the individualism (91/100) and indulgence (68/100) dimensions and lowest on the LTO dimension (26/100). This stands in contrast to China, where the lowest two dimensions are individualism (20/100) and indulgence (24/100) and the highest dimension is LTO (87/100). The LTO dimension is a concept originally derived from Confucian values to counterbalance the Western bias in the initial model and later became incorporated into it. Caveats of characterizing national cultural profiles include intranational variations, which can account for more variance than across countries, and supranational clusters, which can account for similarity in regions across countries, for example, a Latin American cultural zone. The degree of intranational variations is influenced partly by the looseness-tightness characteristics of the cultural setting, which refers to the relative strength of social pressures and enforcements of cultural norms and behaviors. In general, individualistic cultural areas tend to be looser than collectivistic cultural areas with some notable exceptions. Latin American countries and ex-Soviet countries are considered "loose" though they are collectivistic; Germany is considered "tight" despite the fact that they are relatively individualistic. A country's historical exposure to social and other ecologic challenges may be associated with the tightness of the cultural area as a protective response. For the COVID-19 pandemic, a study has shown that loose cultural settings have 4.99 times the case rate and 8.71 times the death rate compared to tight cultural settings. Another challenge to using cultural models is that culture is fluid and changes over time. A cohort analysis using the World Values Survey (1981 to 2008) suggests that contemporary societies tend to demonstrate increases in individualism and indulgence and decreases in power distance over time with modernization, but the crossnational differences relative to each other were found to remain generally stable across time.

Individualism-Collectivism. At the individual level, among Hofstede's most studied dimension is individualism-collectivism. This has spawned a wealth of crosscultural psychological research, including representation of the self, cognition, emotions, communication, and moral appraisal. Studies often contrasted Western American and European cultural regions, relatively high on individualism, versus Eastern Asian cultural regions, relatively high on collectivism. In a collectivist society, in addition to an independent self, an interdependent self-construal is proposed, where the self is fundamentally defined in relation to others. Western compared to Eastern individuals tend to have a stronger positive bias about themselves. Western individuals tend to reason linearly, analytically, and in a focused manner, while Eastern individuals tend to exhibit more cyclical reasoning and attend to the broader picture, as reflected in differences in relative cognitive strengths in attending to foreground versus background objects. Western individuals tend to demonstrate a greater spontaneous trait inference (i.e., attributing the cause of observed behaviors to an individual person's trait) compared to Eastern individuals.

Research on basic emotions based on facial expressions and cognitive appraisal of emotions has largely found them to be universal, although there are cultural differences demonstrated in in-group cultural advantage in the recognition of emotions. Emotional suppression appears to be maladaptive in Western cultural settings, while valued in Eastern ones. Western individuals tend to both experience and value high-arousal positive motions (e.g., happiness) over low-arousal positive emotions (e.g., calmness), while the reverse has been found for Eastern individuals.

There are limitations to crosscultural psychological findings given that a lot of research only compared United States and European individuals as the Western cultural group versus Japanese and Chinese as the East Asian cultural group, although increasingly more studies are available on Latinx populations. While some studies are limited by the use of self-report instruments, many findings demonstrate convergent results using different psychological experimental paradigms, most recently including neuroscience techniques such as differences in evoked response potential (ERP) and brain imaging. For example, in a meta-analysis, functional imaging suggests that the dorsal medial prefrontal cortex, lateral frontal cortex, and temporoparietal junction are more strongly activated among East Asians during social processing, while anterior cingulate, ventral medial prefrontal cortex, and bilateral insulae are more strongly activated among Westerners, implying that the former is more focused on inferring others' emotional states, while the latter has a relatively greater focus on self-referential processing.

There is emerging research into gene x culture interactions, including a line of research on the DRD4 gene, with the 7- or 2-repeat allele (7R/2R) hypothesized to be associated with increased sensitivity to environmental influences and hence greater capacity to internalize cultural norms. For example, as predicted, East Asians carrying the 7R/2R allele in one study scored higher on an interdependent self than European Americans while noncarriers did not. Similarly, in another study, East-Asian 7R/2R allele carriers experienced greater emotional balance (i.e., weak emotional positivity bias) than noncarriers; among Euro-Americans there was a nonsignificant trend toward the opposite, with positivity bias being stronger among carriers than noncarriers.

In sum, complex gene × environment interactions occur in every individual, highlighting the need not to stereotype. Western individuals may be collectivistic and Eastern individuals may be individualistic. In fact, some Western cultural groups are known to be collectivistic. Adding to the complexity are emerging studies on the health impacts of different types of individualism and collectivism based on differences along another dimension, horizontal, emphasizing equality, versus vertical, emphasizing hierarchy. In a recent study of 1906 university students in the United States (ethnicity not specified), individualism was negatively related to wellbeing, while collectivism was positively correlated to wellbeing, with horizontal collectivism having a more robust positive association with wellbeing than vertical collectivism. This body of research alerts us to potential individualistic biases that may impact clinical care. It also underscores the importance of becoming aware of non-Western cultural concepts that may influence the patient’s behaviors, motivations, or value system, such as collectivistic cultural concepts of filial piety among Confucius cultures or familismo (familism) among Latinx groups.

Cultural Values. Values, which guide and motivate actions and transcend specific activities and situations, differ across cultural settings. To provide basic dimensions to study values differences crossculturally, Schwartz’s theory of basic values has been studied using data from over 80 countries and is an alternative to Hofstede’s dimensions, which also reflect cultural values. Schwartz argues that the existence of values is universal as they help meet the needs of individuals as biologic organisms, coordinate social interactions, and benefits group survival and welfare. He lists 10 basic values: power, achievement, hedonism, stimulation, self-direction, universalism, benevolence, tradition, conformity, and security. His model proposes that values, in general, fall along two orthogonal dimensions: openness to change versus conservatism and self-enhancement versus self-transcendence. Assigning high importance to universalism and openness to change and low importance to conservation values has been found to predict tolerance to immigrants and other minority groups, such as LGBTQ+ individuals. In some studies, immigrants' personal values are shown to change after migration. Across 29 countries, person-country and person-region value congruence predicted six wellbeing measures, including emotional wellbeing, and this was moderated by value types, such that some types of value fit have a positive effect on wellbeing, while others have negative effects.

Racial and Ethnic Identity Development

William Cross Jr., developed the Nigrescence Theory in 1971, which has become an influential model in understanding Black and ethnoracial identity development. His initial model includes five stages for a Black individual: pre-encounter, marked by a neutral stance that race does not matter or by internalized anti-Black sentiments; encounter, the occurrence of an event that challenges a previous identity, ideology, and worldview, such as experiencing a racist incident; immersion-emersion, the person participates in Black culture exclusively, experiencing emotions such as anger and guilt, and then emerges with a more nuanced and balanced view of Blackness; internalization, the new Black identity is accepted, resolving internal conflicts; and internalization-commitment, the person develops a balanced relationship with their own Black identity and with other ethnoracial groups, with a long-term commitment to Blackness, including social activism. The last two stages have later been combined by Cross.

Influenced by Cross, a White Racial Identity Model has been developed by Janet Helms, describing the developmental journey of White individuals toward a nonracist identity. The contact status is marked by a denial of systemic racism and profession of being color-blind while being unaware of internal biases. In the disintegration status, the person becomes conflicted over racial dilemmas, such as believing that one is not a racist, while becoming aware of discomfort about their children marrying a minority group member. The person may experience guilt, depression, or anxiety, and respond by avoiding contact with persons of color, not thinking about race, or seeking reassurance that racism is not the fault of White individuals. Influenced by societal ideology, the reintegration status is marked by a retreat into affirming White superiority and blaming minority groups for their problems. After a painful or insightful event, the individual may enter the pseudo-independence status, learning more about racism cognitively, outreaching to minorities that are most like themselves, and trying to help them conform to White standards. In the immersion-emersion status, there is a deeper experiential and affective understanding of oppression and racism, with a shift in focus from trying to change people of color to changing oneself and other Whites. In the autonomy status, there is awareness of one's Whiteness, acceptance of one's role in perpetuating racism, renewed determination to abandon White entitlement, and the capacity to explore issues of racism without defensiveness.

Similar developmental models have since been proposed for other ethnoracial groups, including Asians and Latinx. Abstracted themes from these models have been used to develop the Racial/Cultural Identity Development Model as an overarching framework for psychotherapy. As in the Cross model, a person of color begins in the conformity phase, characterized by a preference for societally dominant cultural beliefs, lifestyles, values, and other cultural/physical characteristics over one's own ethnoracial cultural background, including experiencing internalized racism. The individual enters the dissonance phase after encountering experiences or information that challenge prior beliefs, such as encountering people from their own ethnic or racial group transcending stereotypes or experiencing racism; the individual becomes increasingly aware of racism and of positive and negative aspects of both dominant and ethnoracial cultures. In the resistance and immersion phase, the person may engage in fervent learning about their ethnoracial culture, which is wholly embraced, while the dominant culture is strongly rejected. They may feel guilt and shame about their previous conforming beliefs and anger toward the dominant group, with a desire to end oppression. As the reactive emotions of this phase can be draining, the need to develop a positive sense of self along with discomfort with some of their own ethnoracial group's more rigid views can propel the individual into the introspection phase: the individual works toward a balance of personal self-identity and ethnocultural group identity, questioning some of their ethnoracial group's more extreme views, becoming more interested in oppression experienced by other marginalized groups, and developing an appreciation of positive aspects of the dominant culture. In the integrative awareness phase, there is an inner sense of security with an appreciation of both the minoritized ethnoracial culture and the dominant culture and a strong commitment to eliminate all forms of oppression. This phasic model can also be adapted to describe White racial identity development.

A nonphasic Latino/a Racial Identity Orientation Model has been proposed to understand the diverse identities and perspectives of the U.S. Latino population, which is influenced by their personal history and experiences with families, peers, institutions, and the larger society. The different types of orientation include undifferentiated, professing being "color-blind," adhering to the dominant culture, and attributing inequities to individual failings; White-identified, identifying with being White and superior to other people of color and rejecting the Latino identity; Latino as Other, identifying as being "not White" and with people of color in a generic sense without a strong connection with Latino cultural values or norms; subgroup-identified, identifying with their own distinct ethnic or nation-of-origin subgroup (e.g., Puerto Ricans, Cubans, Dominicans), and not necessarily with other Latinos or people of color; Latino-identified, identifying with being Latino as an overarching category across subgroups; and Latino-integrated, having an integrated Latino identity along with other social identities, and being able to maintain this identity while challenging racial constructs and other forms of oppression.

Jean Phinney has defined ethnic identity as a dynamic, multidimensional construct that refers to one's identity, or sense of self, as a member of an ethnic group. Two psychological processes in ethnic identity development are: 1) exploration, learning, and participating in ethnic culture and practices, and 2) commitment, positive affirmation and having a sense of belonging to an ethnic group. Based on these processes, and similarly to other developmental models, the ethnic identity status may be classified as unexamined ethnic identity, characterized by a lack of exploration or commitment to one's ethnic culture; ethnic identity search/moratorium, a period marked by active exploration of ethnic culture without commitment; and achieved ethnic identity, when there is a clear commitment to one's ethnic identity after exploration. Her scale has been adapted, drawing on Berry model of acculturation (described below), and recently used in Germany to examine the ethnic and national identities of immigrants and native adults.

Ethnic identity models can help understand identity development of members of oppressed groups in a diverse society. There is a general theme of progression and transformation from ethnic cultural denigration to acceptance; minimization of culture to valuing of diversity; and denial of racism to engagement in activism. However, the narrative may be misused inadvertently to dismiss justified anger, perpetuate structural racism, and translocate faults to the individual of color as not being fully developed. It has been argued that ethnic identity needs to be explored with a critical consciousness and through the lens of intersectionality.

Migration and Acculturation

Acculturation describes the process of psychological, social, and cultural change when two cultures are in prolonged firsthand contact with one another; it has also been used as a broad term to describe the process of adaptation of newcomers to a country. In this context, acculturation has been defined as the process of adopting the culture of the dominant "host" group, while enculturation has been defined as the process of identifying with one's own heritage culture. The colonial stance and early models of acculturation assume that indigenous minoritized groups and newcomers, including immigrants and refugees, will gradually accept and adapt to the normative behaviors and values of the dominant cultural group in a unidirectional acculturative process, with people exhibiting mixed cultures as being "bicultural" until ultimately reaching full cultural assimilation over time or successive generations.

In contrast, Berry's bidimensional model conceptualizes acculturative strategies based on the newcomers' desire to maintain their heritage culture and the desire to interact with the dominant culture. The integration strategy is the process of adopting elements from the dominant culture while maintaining aspects of one's heritage culture; the assimilation strategy is the adoption of the dominant culture and disengagement from one's heritage culture; separation strategy is the retention of heritage culture, while not taking up the dominant culture; and marginalization occurs when there is no uptake of dominant culture while becoming disengaged from one's heritage culture. In this model, it is hypothesized that integration may lead to optimal mental health outcomes, while marginalization may lead to increased vulnerability and the adverse impact on mental health. The other two strategies may be adaptive depending on various factors such as local context. For example, if there is a sufficient critical mass to maintain a heritage cultural community, such as an ethnic enclave, or there is sufficient means to maintain active contact and transaction with the heritage culture, such as through technology or transnational work, the separation strategy may be adaptive for the individual.

In practice, acculturation does not fall neatly into these simple four categories across all domains of life. For example, a newcomer may adopt mainstream dietary and clothing practices in society out of necessity, while retaining heritage cultural values in their family and spiritual life. In the modern world, people are increasingly exposed to multiple sources of cultural influences, some of which may become more salient depending on the situational context. Individuals' acculturative strategies may vary across time, including changes reflective of cognitive, social, and life stages development, or in response to external experiences, such as incidents of discrimination. In addition to internal factors that may influence the choice of acculturative strategies such as personality traits, preferences, and personal history, situational and other external factors are important considerations.

The Bourhis' Interactive Acculturation Model elaborates on Berry's model to consider the attitudes and beliefs of the dominant cultural group toward newcomers with a corresponding model: in integration, newcomers should ideally be able to retain aspects of their heritage culture while also adopting elements of the dominant culture; in assimilation, newcomers should relinquish their heritage culture and be assimilated into the dominant culture; in segregation, newcomers should be segregated into their own communities, thereby retaining their own culture; and in exclusion, newcomers should be excluded from society in general, and not be welcomed to participate in the dominant culture nor supported to retain their heritage culture. Individualism is proposed as another possible attitude held by the newcomer or members of the dominant group, especially those from individualistic cultures, with the belief that each person should be treated as an individual, downplaying the roles of dominant or heritage cultures. Of the five types of attitudes described, segregation and exclusion are the most rejecting toward newcomers. As the dominant group has more structural power, newcomers may not be able to actualize their preferred acculturative strategy, leading to intercultural conflict and increased acculturative stress.

At a systemic level, the attitudes of the dominant group can be reflected in state ideology, influencing national policies. A pluralistic multicultural ideology values the dominant and nondominant cultures and may provide active governmental resources and funding to promote both. A civic ideology, while permissive of diverse cultures, does not actively support or promote nondominant cultures with funding or resources. An assimilation ideology, or “melting pot,” supports the dominant culture and encourages the relinquishment of other competing cultures. An ethnist ideology supports never granting full status to newcomers or people not from the dominant ethnic culture. Over the past 50 years, the United States has seen a shift from an assimilation ideology toward a civic ideology.

There is often a “principle-implementation gap,” in which there is more popular support for abstract principles and ideologies of multiculturalism than for enacted national policies and programs. Two main databases quantify the implementation of national multicultural policies that support the diversity and inclusion of newcomers. The Migrant Integration Policy Index (MIPEX) evaluates 52 countries on three dimensions of integration: whether migrants can enjoy comparable basic rights, have equal opportunities, and can feel secure about having a future in the country. Scored out of 100, the top 5 countries in 2020 were: Sweden (86), Finland (85), Portugal (81), Canada (80), and New Zealand (77); the United States (73) ranked 6th. The Multicultural Policy Index (MPI) consists of three separate indices. The first index reflects policies and programs for immigrant minorities: legislative affirmation of multiculturalism; adoption of multiculturalism in school curricula; ethnic inclusion in public media or media licensing; exemptions from dress codes, Sunday-closing, and other legislations; allowing dual citizenships; funding to support ethnocultural activities; funding of bilingual or mother-tongue instruction; and affirmative action. Of 21 countries, scored using a scale of 0 to 8, the top 5 countries in 2020 were: Australia (8), Canada (7), Finland (7), Sweden (7), and New Zealand (6.5); the United States (3.5) ranks 11th; the lowest is Japan (0). The second index examines policies about indigenous peoples, including land rights, self-government rights, treaties, cultural rights, customary laws, central government representation, legislative affirmation, support for indigenous rights internationally, and affirmative action. Out of 9 countries evaluated on a scale of 0 to 9, Canada (8.5) and the United States (8.5) rank first, and the lowest is Japan (3). The final index examines the country’s policies toward national minorities, defined as subgroups with distinct national identities and goals within a country, such as the Welsh and Scots in Britain, the Catalans and Basques in Spain, the Flemish in Belgium, and the Hispanics in Puerto Rico. This includes policies on territorial autonomy; official language status; representation in the central government or the courts; public funding for the minority language in universities, schools, or media; constitutional affirmation of multinationalism, and representation in international bodies. Of the 11 countries evaluated on a 0 to 6 scale, Canada and the United Kingdom (6) ranked first, the United States (3.5) ranked 9th, and the lowest was Japan (0).

CULTURALLY COMPETENT ASSESSMENT AND TREATMENT

Basic Principles of Culturally Competent Assessment

The principles of cultural competence apply to every clinical assessment, as culture shapes how patients express psychiatric symptoms. In addition, culture influences the meanings attributed to the symptoms and impacts the interaction between the patient and the family, the clinician, and the health care system within which the clinician operates. As culture does not only reside in the patient, the clinician first needs to become aware of and regularly reflect on their own cultural background, assumptions, and biases, and consider how they may impact clinical care.

Throughout the assessment, the process of how the interview unfolds and the patient’s engagement is as important as paying attention to the content of the interview. Accurate assessment is contingent on building a workable therapeutic relationship. Both the clinician and the patient will have unspoken cultural expectations when meeting, including the purpose of the interaction and their respective roles, and this may need to be clarified at the outset.

There are three core aspects to consider culturally in the engagement triad: communication, empathic understanding, and rapport. Communication includes verbal and nonverbal interactions. Verbal interactions may include the use of interpreters, as well as attention to the choice of words, the literal content, and the complexities of the language used. Nonverbal interactions include eye-contact, blinking, facial expressions, body gestures such as nodding, physical contact such as shaking hands, sitting posture, as well as prosody of language such as tone, pitch, rhythm, and so on. Low-context communication is considered typical of Western culture, whereby the message is expected to be literally articulated and objectively understood. In most other cultural settings, high-context communication entails indirect communication with full meaning often embedded or encoded in context. For example, meanings may be conveyed to the patient by the clinician's appearance, including clothing and hairstyle, or the interview room setting such as by the clinician's personal belongings. Empathy refers to openness to perceiving and understanding the other's cultural perspective. Rapport can be built by aiming to increase comfort in the artificial social situation of a clinical encounter that may not be universally understood. These three factors are interrelated, as good rapport facilitates effective communication and disclosure, which facilitates true empathic understanding. Engagement, or the bond between the clinician and the patient, is the cornerstone of building a therapeutic alliance, which consists of the bond as well as alignment of goals and tasks.

Working with an Interpreter. Interpretation refers to rendering one spoken language into another, while translation refers to carrying out this conversion on written materials. Usually, consecutive interpretation is used in clinical settings rather than simultaneous interpretation, which is much harder to achieve high accuracy. The latter would require specialized ability and training of the interpreter, which if available, may be useful in therapeutic settings such as group therapy.

In the United States in 2019, 22% of the population spoke primarily a language other than English at home. Patients should be asked about their language of preference for the psychiatric assessment. As communications of distress can be nuanced and language may make certain internal experiences more accessible, basic proficiency in a language may not be sufficient for assessment, in which case the services of an interpreter should be used to avoid misunderstandings. If there appears to be a need for interpretation but the patient expresses reluctance about involving an interpreter, the clinician should explore whether there are concerns about confidentiality, especially if the patient comes from a relatively small and close-knit community. The interpreter's regional accent may make the patient feel uncomfortable if there has been internal conflict or war in the patient's home country. Children under 18 or family members should not function as interpreters, except for emergency situations. Trained multilingual staff may fulfill the role, although they are not a substitute for trained professional interpreters. Telephone or video interpretation services are becoming increasingly available, although the former has limitations, given the lack of visual nonverbal cues.

Clinicians should hold a prebriefing meeting with the interpreter to ensure mutual understanding of the interpretation process. The interpreter should use first-person language and directly interpret the clinician's questions or statements, such as, "did you hear any voices of people talking to you, even when there is no one around"? rather than the third person voice, such as, "the doctor wants to know if you hear…" Similarly, the clinician should remind the interpreter to directly interpret the patient's response, even if it does not seem coherent, as this may reflect an underlying thought disorder. The interpreter should be provided with essential contextual information, such as the purpose and scope of the interview, including sensitive areas to be explored, or the patient's limitations, such as the presence of a hearing impairment, without divulging unnecessary details. Clinicians should make it clear that the interpreter can ask for clarification if they do not understand what the clinician is asking. An experienced interpreter may provide cultural contextual information–for example, about a faux pas question from the clinician or a culturally nuanced response from the patient–but this should be made apparent to the clinician as the interpreter's opinion. Assessment tools that may be used to evaluate mood and anxiety disorders, substance use disorder, or cognitive status should be reviewed in advance, and where possible validated translations should be used. For bedside cognitive testing, such as the Mini-Mental State Examination, the interpreter needs to understand the purpose of the questions and may need to be reminded of this during the interview. For example, the interpreter may inadvertently assist the patient with the serial-7's task, for example by saying, "subtract 7 again from 93," rather than using neutral prompts like "and next"? The clinician should prepare the room in advance by placing the chairs in a therapeutically strategic manner, preferably though not necessarily, in a triangle. The interview room should not be small, dark, or closed-in, especially if the patient has experienced trauma. Neither the clinician nor the interpreter should come across as interrogating the patient.

Allowances should be made for the additional time needed for an assessment when using an interpreter. The interpreter should be introduced to the patient and their role explained, including reassurances about confidentiality. The clinician should talk directly to the patient and speak in the first person. Statements should be kept brief to ensure interpretation accuracy. Proverbs, sayings, technical language, or other jargon should be avoided. As necessary, the clinician can ask for a break to clarify with the interpreter if problems arise, such as side conversations between the patient and interpreter not being relayed. As a consideration of physical and psychological safety, the interpreter as nonclinically trained staff should not be left alone with the patient if the clinician needs to leave the room.

The clinician should hold a debriefing meeting with the interpreter after the assessment. The interpreter should be asked to provide their perception of the meeting, including nonverbal or cultural expressions noted or other relevant cultural-contextual information. In this capacity, the interpreter may function beyond the role of linguistic conversion, but act also as a cultural broker to help the clinician understand the patient from a cultural perspective, bridging cultural differences. If traumatic material was discussed, the clinician should ensure that the interpreter is not emotionally disturbed. It is not uncommon for interpreters to take notes as a memory aide during the interview. These notes should be destroyed to ensure confidentiality.

Guidelines for working with an interpreter

- Ensure language, dialect, and cultural compatibility between the interpreter and patient
- Brief the interpreter before starting the interview, including asking:
  - The interpreter to speak in the first person (“I”)
  - Not to filter the patient’s statements (e.g., to evaluate thought disorder)
- Introduce the interpreter to the patient and address any concerns (e.g., confidentiality)
- Address the patient directly during the interview
- Do not leave the interpreter alone with the patient
- Use short sentences
- Clarify, summarize, and verify often
- If necessary, ask for a break to discuss with the interpreter about problems such as frequent side conversations
- Debrief after the interview and discuss any concerns
- Obtain any written notes from the interpreter and destroy them

Cultural Formulation

A clinically competent clinician needs to be mindful of and bring a cultural perspective to bear on the patient's lived experience, self-identities, understanding of their own problems or illness, relationships with their families, social network, and the larger society, experiences of discrimination or inequities; psychosocial stressors; strengths, coping strategies, and resources; and spirituality and religion. The DSM-5-TR provides an Outline for Cultural Formulation (OCF) as a framework for organizing cultural information into five major areas. This is detailed below and serves as an essential guide to integrating cultural considerations into assessment and treatment.

The Outline for Cultural Formulation consists of five areas of assessment

- The cultural identity of the individual
- Cultural concepts of distress
- Psychosocial stressors and cultural features of vulnerability and resilience
- Cultural features of the relationship between the individual and the clinician, treatment team, and institution
- Overall cultural assessment

Cultural Identity of the Individual. Cultural identity refers to identification with being a member of sociodemographic or other socioculturally defined groups, such as those based on race, ethnicity, place of current residence, urbanicity, place of origin, language, sex, gender identity, sexual orientation, age, religion, spirituality, socioeconomic class, caste, occupation, migration history, (dis)ability, autism, etc. Self-identification of cultural identities may differ from the identities perceived or ascribed by others, and both may influence social interactions, access to internal and external resources, and developmental and current challenges and conflicts. Every individual has multiple cultural identities, and their intersection may be empowering or disempowering depending on context. As discussed, racial and ethnic cultural identities are not fixed and change over time depending on life-stage and development, acculturative strategies, and response to situational context such as external events and the environment. Assessing language abilities and preferences in different contexts is not only helpful for clinical communication but may be informative about the patient’s cultural identities and social integration.

Cultural Concepts of Distress. Before discussing cultural concepts of distress, it is helpful to reflect on underlying assumptions that may arise from traditional medical training. Western medicine distinguishes between signs and symptoms and between disease and illness, with the former in each set believed to be true pathology and the latter reflective of the patient’s experience. This sharp distinction between objectivity and subjectivity reflects Western medical culture. Increasingly, mental illness is appreciated as a complex interactive dynamic process influenced by biologic, psychological, social, cultural, and spiritual factors. For example, studies have found evidence of physiologic changes resulting from cognitive manipulation, such as in anxiety disorders; differences in brain imaging due to psychological treatment, as in OCD; and differences in brain imaging from culturally patterned ways of thinking, such as individualism versus collectivism. Understanding the patient’s illness experience and cultural concepts of distress is not only an engagement strategy or a way of translating patient’s cultural vocabulary into DSM-5-TR psychopathologies, but an approach for appreciating that the illness experience itself can inherently influence the process of mental illness, including its course and outcomes.

To help the clinician understand the patient’s illness experience from a cultural perspective, the DSM-5-TR uses the term cultural concepts of distress. The three types of cultural concepts include cultural idioms of distress, ways in which the individual expresses or communicates their illness experience to others (e.g., “I have butterflies in my stomach”); cultural explanations, the patient’s explanatory model of illness, or underlying conceptualizations about their illness (e.g., “I have depression because of chemical imbalance in my brain involving serotonin triggered by stress and burnout”); and cultural syndromes, which are clusters of symptoms and attributions recognizable by their cultural group (e.g., “I just had a panic attack”). The patient may present with one or more of these types, and they may or may not cohere together, as they may represent different facets of their illness experience.

Cultural Concepts of Distress (DSM-5-TR):

- *Cultural idiom of distress:* A way of talking about symptoms, problems, or suffering that is common among individuals of a similar cultural background.

- *Cultural explanations:* An explanatory model for certain symptoms, illnesses, or types of distress that is common among individuals of a similar cultural background.
- *Cultural syndrome:* A group of co-occurring symptoms that is common among individuals of a similar cultural background.

A common form of cultural idiom of distress is experiencing and expressing distress through metaphors involving bodily sensations or somatic symptoms such as pain, dizziness, fatigue, gastrointestinal symptoms, or paresthesia. Patients may first present these to primary care or medical specialists and may be most concerned about them. While somatization has been conceptualized as a primitive defense in classical psychoanalysis, studies have since shown that patients from around the world, including Western countries, commonly experience somatic symptoms when distressed. From a diagnostic perspective, patients presenting with somatic symptoms may have common mental disorders such as anxiety or depression but be focusing on their somatic aspects. They may also be expressing an illness anxiety disorder; somatic symptom disorder; other forms of psychosocial distress that may or may not fit under Other Conditions That May Be a Focus of Clinical Attention; or cultural syndromes that may best fit under Other Specified Disorders. In assessment, rather than prematurely arbitrating symptoms as biologic/organic versus psychological/functional/cultural, clinicians need to validate the patient’s experience, while opening up a discussion to fully explore psychological and sociocultural stressors, as well as other symptoms that the patient may be experiencing.

The explanatory model of illness, pioneered by the work of Arthur Kleinman, refers to the patient’s understanding of their illness. A clinician can explore with the patient their perception of their problems or the illness’ name, associated characteristics, severity, underlying causes and mechanisms of illness, exacerbating and ameliorating factors, treatment options, and prognosis. The patient may hold multiple explanatory models; each may be fragmentary; and they may be experienced as intermixed, interacting, or conflicting with one another. The explanatory model held by patient’s families, friends, or cultural communities may also be highly relevant, and they should also be elicited from patient’s perspective.

Based on Eisenbruch's work, a clinical heuristic that may be helpful in exploring cultural causes is to think along two dimensions, Western versus non-Western physiologic (or supernatural) and impersonal versus personal causes. Examples include: imbalance due to too much wind in the body (non-Western physiologic, impersonal); targeted voodoo curses (non-Western physiologic, personal); genetics (Western physiologic, impersonal); and situational stress (Western physiologic, personal). This is not an objective classification system, as the patient's perception may be very different from the clinician's. Exploration of the "personal" nature of a perceived cause may lead to disclosure of feelings of guilt (e.g., "I committed a sin") or anger/fear (e.g., "I am a victim of a malevolent spirit" such as a djinn).

The clinician may explore the patient's worldviews, such as whether they have the capacity to influence the outcomes of events in their lives including their illness. Sue and Sue propose that worldviews on self-appraisals of control may fall along two dimensions, locus of control (internal or external) and locus of responsibility (internal or external). This maps out to four configurations: (1) internal locus of control and responsibility (IC/IR)–believing that one has the capacity to influence outcomes, and that success and failure depend on one's actions; this is commonly reflective of an individualist mindset found in Western culture; (2) external control and internal responsibility (EC/IR)–the experience that one has no control over one's life outcomes, but is still responsible for them; for example, this may reflect a form of internalized racism and self-blame; (3) internal control and external responsibility (IC/ER)–perceiving that one has the capacity to influence outcomes, but that there are external barriers, such as racism, limiting success; (4) external control and responsibility (EC/ER)–perceiving that one has no control over one's own life, and that the fault lies externally; this may exhibit as passivity and apathy, which may reflect a form of learned helplessness or keeping a "low-profile mode" to survive. The clinician may reflect on both their own cultural explanatory model and worldview for a given clinical problem, and how they may be similar or conflicting from the patient's.

The construct of cultural syndrome replaces the former term of culture-bound syndrome, which may exoticize the illness presentations of non-Western cultures while not acknowledging the cultural framing also inherent in Western diagnostic systems. The DSM diagnoses can be understood as evolving Western cultural syndromes formulated to reflect our most current understanding of mental illness from a cultural vantage point grounded in Western science. Clinical observations and research suggest that there are common universal expressions of mental suffering–common natural clusters of symptoms at a high level–as well as culturally patterned ways of experiencing, expressing, understanding, and attending to mental illness that change over time. The DSM-5-TR lists ten examples of cultural concepts of distress, noting that they may be functionally used by the patient as a cultural idiom of distress, cultural explanation, or cultural syndrome. The examples are: ataque de nervios, dhat syndrome, hikikomori, khyâl cap, kufungisisa, maladi dyab, nervios, shenjing shuairuo, susto and taijin kyofusho.

Psychosocial Stressors and Cultural Features of Vulnerability and Resilience. This section includes psychosocial and cultural factors not captured in other sections of the OCF that may predispose, precipitate, or perpetuate the illness (increasing vulnerability) or protect the patient (increasing resilience). These factors should be considered from the patient's, the family's, and communities' past (intergenerational) perspectives, as well as present circumstances. The patient's family, peers, social support network, cultural community, and the larger societal institutions, structures, and systems can all be a source of support or stress. These various levels may interact and conflict with one another. For example, a person from a particular racial, ethnic, cultural, religious, or migration background may hold certain beliefs or values based on their personal choice and/or current state of cultural identity development and acculturative strategies employed, which may be at odds with those from a peer or family member or community. It is not uncommon for patients and family members to experience both an intergenerational gap and a cultural gap or role conflict. For migrants, it is helpful to take a migration history and consider premigration, migration, and postmigration as sources of stress or supports. Under premigration, it is helpful to learn about the circumstances during childhood, such as family composition, attachment figures, early childhood and/or school experiences with peers and teachers, as well as living condition and salient events as an adult. In many cultures, extended family members may play an important role in the patient's life, rather than following a nuclear family structure. Clinicians should use a trauma-informed approach to inquire about political issues, trauma, war, and natural disasters, ensuring not to coerce the patient to share, especially in initial interviews, to prevent retraumatization (Table 4.2–4). It is important to explore reasons and circumstances for the planned or unplanned migration. During the migration journey, especially for refugees, the patient may have experienced additional trauma enroute or at refugee camps (see Table 4.2–4). Under postmigration, the clinician should explore language barriers, acculturative stress and strategies, discrimination, unemployment and underemployment, housing, role reversal within the family, intergenerational conflict, social isolation, and disconnection, as well as other losses. It is useful to examine sources of internal resilience such as coping skills, internal values, and meaning-making, as well as external sources of resilience, such as families, peers, and communities. For all patients, social determinants of health and other structural or systemic discrimination or challenges should be explored. The patient's level of functioning or degree of impairment from their illness can be reviewed from both their own and their cultural community's perspective, including possible stigma.

Cultural Features of the Relationship between the Individual and the Clinician, Treatment Team, and Institution. The cultural identity and worldviews of the clinician, the mental health team, and the institution may have an impact on all aspects of care. Becoming aware of inherent biases and assumptions embedded in cultural scripts or policies at each level, as well as the perception of them by patients or their cultural communities, can help anticipate cultural dynamics that may arise during clinical interactions. There are power imbalances that arise between clinician and patient not only inherent to the dyadic provider–patient relationship but also due to their respective social and cultural identities; this includes any power imbalances between groups the clinician and patient belong to, such as from a historical and present structural and systemic perspective. These factors may influence the development of trust and the establishment of a culturally safe space for assessment and treatment. Cultural dynamics, especially if unidentified and unresolved, may manifest in transference and countertransference-based interactions between the clinician (including the treatment team) and the patient (including the family), resulting in a therapeutic impasse. As a result, patients or families may be blamed or labeled pejoratively as "lacking insight," "uncooperative," "in denial," "noncompliant," "passive-aggressive," "codependent," or "emmeshed."

Table 4.2–4.
Refugee Premigration Experiences and Risk Factors for Psychopathology

Common Premigration Traumatic Experiences of Refugees	Common Risk Factors for Psychological Distress Among Refugees
War and deliberate killings	Past experiences of traumatic events
Genocide	Premigration health problems
Murder of family members/friends/relatives	Lack of accommodation/ overcrowding
Witnessing the murder of family members/friends	Isolation/loneliness
Prolonged mental and physical torture	Lack of supportive social networks
Rape of women and young girls	Lack of supportive friends
Kidnap of children and women	Family separation
Long imprisonment and forced labor	Unemployment status
Cruel amputations	Lack of access to education
Food shortage/starvation	Language problems
Lack of water	Barriers to culturally appropriate health and social services
Looting and daily robbery	Legal uncertainties
Destruction of personal properties	Culture shock and adjustment problems
Destruction of public infrastructures	Substance misuse
	Experiences of racism and discrimination
	Abject poverty

Reprinted with permission from Warfa N, Bhui K. Refugees and mental health. In: Bhugra D, Bhui K, eds. *Textbook of Cultural Psychiatry*. 1st ed. Cambridge University Press; 2007:505.

Overall Cultural Assessment. This section synthesizes the information gathered using the OCF, focusing on how it may impact assessment, diagnosis, formulation, and the treatment and management plan.

Conducting a Cultural Assessment

The Cultural Formulation Interview. The DSM-5-TR includes the Cultural Formulation Interview (CFI), a semistructured set of protocols designed to help the clinician obtain pertinent cultural information to complete the OCF. The CFI follows a person-centered approach to promote a collaborative, shared decision-making process concerning diagnosis and treatment between clinicians and patients (core CFI) and their families (CFI-informant version). In addition, supplementary modules can be used to help clinicians conduct a more comprehensive assessment. The original supplementary models included: (1) explanatory model; (2) level of functioning; (3) social network; (4) psychosocial stressors; (5) spirituality, religion, moral traditions; (6) cultural identity; (7) coping and help-seeking; (8) patient-clinician relationship; (9) school-age children and adolescents; (10) older adults; (11) immigrants and refugees, and (12) caregivers. Two additional modules have since been developed, focused on grief (13) and active-duty military members, veterans, and their families (14).

Under each of its four sections, the core CFI provides detailed information about the section's purpose and instructions on the left-hand column and sample questions on the right-hand column, which can be reworded or elaborated to suit the clinical situation. The CFI is discussed in Chapter 29.2 in this textbook and can be found at: https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/assessment-measures.

In a condensed form, the four sections with the 16 questions are:

Cultural definition of the problem Elicit main concerns and cultural idioms of distress: Cultural Definition of the problem: (1) What brings you here today? (2) How would you describe your problem* to your families, friends, or communities? (3) What troubles you most about it? (* use the patient's term to describe the problem as applicable, e.g., "breakdown").

Cultural perceptions of cause, context, and support Elicit explanatory model of illness, stressors and supports, and cultural identity: Causes: (4) Why do you think this is happening to you? (5) What do others in your family, friends, or communities think? Stressors and Supports: (6) Are there supports that make your problem better, such as from family, friends, or others? (7) Are there stresses that make it worse, such as difficulties with money or family? Cultural Identity: (8) What are the most important aspects of your background or identity? (9) Are there aspects of your identity that affect your problem? (10) Are there aspects of your identity that are causing other problems?

Cultural factors affecting self-coping and past help-seeking Elicit coping response and help-seeking experiences: Coping: (11) What have you done to cope with your problem? Help-Seeking: (12) What past treatments, including non-Western ones, have you tried? Barriers: (13) Has anything prevented you from getting the help you need?

Cultural factors affecting current help-seeking Clarify concerns and expectations regarding treatment: Preferences: (14) What kinds of help do you think would be most useful for you? (15) Are there other kinds of help your family, friends, or community would suggest? Clinician–Patient Relationship: (16) As we have different backgrounds and experiences, do you think there is anything we can do to better provide you with the care you need?

Contextualizing a Culturally Competent Assessment. The CFI is applicable to every patient; its use is especially important when there are concerns that cultural issues and dynamics are impacting the patient's care and impeding therapeutic progress. The CFI may be used in its entirety, or individual components can be integrated into a clinical encounter when there may be time constraints or other limiting factors. A contextualized approach for cultural assessment and formulation has been proposed (see Table 4.2–5). An integrated approach embeds culture-related questions into the format of a standard psychiatric interview (see Table 4.2–6).

One consideration is whether to first start with an exploration of the patient's sociocultural identities or their presenting issues. Each has potential advantages and disadvantages. The first approach may signify to the patient that their sociocultural identities will be valued, respected, and taken into consideration and that they will be treated as more than their presenting issues; it may help the clinician gain a more contextualized understanding of the presenting issues and guide the exploration of its meaning and importance; in some cultural settings, establishing a basic minimal rapport first may help the patient discuss their symptoms and experiences. The second approach may reassure the patient that their presenting problems will be addressed and prioritized; the clinician may ask and frame sociocultural questions in a way that is most pertinent to their problems; in some cultural settings, there may be an expectation that the clinician will focus on their problems and only discuss their background when its relevance is justified. The best approach may depend on the clinician's preference and skills, the patient's or their cultural communities' preference, and the nature of the clinical setting and context.

Table 4.2–5.

A Contextualized Approach to Culturally Competent Assessment and Formulation

Level	Setting/Context	CC Assessment	Cultural Formulation
I	Emergency, acute care, consultation-liaison/medical consultation services	Partial or focused assessments of the presenting issue from a sociocultural perspective: Significance, what does it mean to you? Stressors, is your social support network making things harder for you? Supports, how can you get more support from your social network or community?	The 3 S's from a sociocultural perspective: ▶ Significance ▶ Stressors ▶ Supports
II	Initial standard assessment in: ambulatory (outpatient) care, inpatient care, assessment for psychotherapy	Integrated CC Psychiatric Assessment (Table 4.2–6) and/or DSM-5-TR Cultural Formulation Interview (CFI)	▶ Formulation Grid: Bio-Psycho-Social-Spiritual × Predisposing-Precipitating-Perpetuating-Protecting - with culture considered in every square and/or ▶ DSM-5-TR Outline for Cultural Formulation (OCF)
III	Longitudinal follow-up care of Level II; academic purposes–research and education	Integrated CC Psychiatric Assessment (Table 4.2–6) And DSM-5-TR CFI	▶ Formulation Grid: Bio-Psycho-Social-Spiritual × Predisposing-Precipitating-Perpetuating-Protecting-with

culture considered in every square and

▶ DSM-5-TR OCF

The mental status examination is a central component of the psychiatric examination. However, its interpretation in migrant populations and cultural groups with which the clinician is unfamiliar requires caution, as the assessment may be culturally biased. Patients' cultural backgrounds of origin, educational levels, and types of acculturative adaptation mold their response. The standardized mental status examination components include cooperation, appearance, behavior, speech, mood/affect, thought process, thought content, perception, cognition, insight, and judgment. Cultural differences are wide and varied in dress and grooming. Facial expressions and body movements to express affect may be more reflective of typical cultural manifestations than pathology. If the clinician is unfamiliar with the individual's cultural background and the patient's fluency in the country's dominant language is limited, the clinician must use caution in interpreting disturbances of speech and thought process, perception, and affect. The presence of hallucinations, for example, can be easily misinterpreted, such as hearing encouraging or clarifying comments from deceased family members, which are normative experiences in many cultural groups. The clinician should not assume that the patient understands what the clinician is trying to communicate, and miscommunication is a common problem, with or without the help of interpreters. The cognitive examination may be particularly tricky. Education and literacy have an essential and biasing role. Patients may need adequate time to fully express themselves and the clinician may facilitate this process by repeating and restating questions to reduce miscommunication. Asking about the meaning of proverbs unfamiliar to the patient may be an inappropriate means of determining abstract thinking. Clinicians can carry out an accurate mental status examination if they allow additional time to clarify the cultural equivalence of concepts.

As psychiatric diagnoses are made based on clinical phenomenology, cultural and social factors may lead to over- or underdiagnosis. Use of the DSM-5-TR CFI and OCF helps to mitigate this challenge. The text of

DSM-5-TR has been expanded on social and cultural considerations under specific disorders. In ambiguous cases, it is important to involve patient's peers or families (e.g., with the CFI-Informant version) as well as consider consultation with cultural brokers, cultural consultants, and community or spiritual leaders. In complex cases, the diagnosis may not be clear until the patient is followed over time and social issues are first resolved.

Culturally Competent Management and Intervention

Patient- and Family-Centered Care. Attending to social and cultural issues is an important part of providing patient-centered care, which helps to protect and ensure patient's autonomy in managing their health. In many collectivistic cultures, families are also expected to influence patients' healthcare decisions in addition to providing collateral information. They may request to be present for assessments and may influence whether the patient agrees or adheres to the management plan. The clinician needs to explore the patient's wishes whether to involve their family, the role they would play, and what type of information would be safe to share. In general, it is helpful to have opportunities to meet with the patient alone when culturally permissible and, where appropriate, arrange to meet with the patient and their family together to discuss the management plan. Psychoeducation involving families can support the patient's recovery and promote patient's adherence to the management plan. Family interventions, such as family therapy and multifamily group interventions, are vastly underutilized. Group therapy and community-based psychosocial interventions are also unvalued and need more systemic investment.

Using the LEARN Model. A common framework for providing crosscultural therapeutic intervention is summarized by the mnemonic LEARN (Table 4.2–7). This approach may be used when the clinician recommends any type of therapeutic intervention. It emphasizes acknowledgment of cultural differences while using a collaborative approach to care.

Table 4.2–6.
Integrated Culturally Competent Psychiatric Assessment

Standard Interview Section	Sample Questions to be Integrated into Standard Psychiatric Interviewing[a]	Related Core CFI/Supplementary Modules	Related OCF Domain
ID	I would like to first learn a bit about your background, so I can understand you better and we can figure out together how I can be of help to you. What languages do you speak and prefer to be interviewed in? Who do you live with? What is your current relationship status? Do you have any children? Do you self-identify with any particular racial, ethnic, cultural, or community groups? Where were you born? What is your source of income?	CFI-8–10 CFI-SM Cultural Identity	Cultural identity of the individual

CC/HPI	What is the problem that brought you here today? Do you think it is an illness, and if so, what would you call it? What have you experienced as a part of it? What causes it? What makes it better or worse? How can it be treated? How have you tried to cope or make your illness better? How has it worked?	CFI-1–7, 11 CFI-SM Explanatory Model; Level of functioning; Psychosocial stressors; Coping and help-seeking	Cultural concepts of distress
Past psychiatric history	What do you (and your family) think about psychiatric treatments you received, like treatment programs, hospitalization, medications, or psychotherapy? Besides Western treatments,	CFI-11–13 CFI-SM Coping and Help-Seeking	Cultural concepts of distress

	what else have you tried, like supplements or other ways of healing? How has it worked? Are you still using them?		
Family psychiatric history	Are there any family members that experienced emotional difficulties that may or may not have sought help?	CFI–5	Cultural concepts of distress
Social and developmental history	Where were you born/raised? Was it in the city or the country? What were your mother/father’s jobs? How would you describe your mother/father? How’s your relationship with them? Do you have older or younger siblings? How’s your relationship with them? Were there any	CFI-6–10, 13 CFI-SM Cultural Identity, Social Network, Psychosocial Stressors Depending on age: CFI-SM School-Age Children Adolescents, Older Adults	Psychosocial stressors and cultural features of vulnerability and resilience

others who helped raised you? Where are your parents/sibs now? How closely do you keep in contact with them? I am unfamiliar with the school system you went through-how does it work? What was your experience like at school with peers and teachers? Were you similar or different from others? How would you describe yourself? Your strengths? If it’s okay, would you like to share about any traumatic experiences that you might have had? What jobs have you worked in in your country/here?

	What is your sexual orientation? Tell me a bit about your significant romantic relationships. Tell me about your involvement in the community.		
Migration history	Premigration: What was life like before you came to this country? If you feel it's ok, would you like to share about any trauma you have been through? What do these experiences mean to you? What were the circumstances that led you to come to this country? Migration: What was the journey like coming to this country? Postmigration: How do you find living here? How much do	CFI-6–10, 13 CFI-SM Immigrants and Refugees	Psychosocial stressors and cultural features of vulnerability and resilience

	you identify with the culture here? And your own culture? What are your current social supports like here, like your family, friends, and community, and from back home? What challenges have you experienced in finances, employment, or housing? Have you encountered any discrimination? Have you had difficulties regarding your legal status here?		
Religion and Spirituality	FICA (adapted from Puchalski): Faith and Beliefs: What is your faith or spiritual beliefs? Importance: How important is it to you?	CFI-6–13 CFI-SM Spirituality, Religion, and Moral Traditions	Psychosocial stressors and cultural features of vulnerability and resilience

	Community: Are you part of a faith community? Address in care: How would you like spiritual or religious issues to be part of your care, by me or others?		
Concluding the interview	Summarize Mutual Feedback: As we have different backgrounds and experiences, do you think there is anything we can do to better provide you with the care you need?	CFI-14–16 CFI-SM Patient-Client Relationship	Cultural features of the relationship between the individual and the clinician, treatment team, and institution Overall cultural assessment
Obtaining Collateral History	Obtain collateral history from family, friends, and/or community	CFI-Informant Version CFI-SM Caregivers CFI-School-Age Children Adolescents Addendum for Parents' Interview	All parts of the OCF
Cultural Consultation	Consult with Cultural Informants or Consultants	CFI-Informant Version	

[a] Adapted from Cultural Psychiatry: Focus on Assessment Seminar for Sociocultural Dimensions of Psychiatry Core Seminars, developed by Fung K, Andermann L, Munshi A, Lo T, Department of Psychiatry, University of Toronto.

Table 4.2–7.
LEARN Mnemonic

L—Actively Listen to the patient's problems and their own explanatory model of illness.
E—Explain your understanding of the patient's condition in understandable terms.
A—Acknowledge similarities and differences between your views and the patient's, opening up a dialogue.
R—Recommend appropriate treatment and explain how the treatment plan may help.
N—Negotiate a collaborative treatment plan that the patient (and family) is agreeable to and is inclusive of the patient's own cultural treatment methods, if safe and appropriate.

Pharmacotherapy. There are several sociocultural factors that should be considered in prescribing psychiatric medications.

The field of ethnopsychopharmacology researches the role of ethnicity-related differences in pharmacokinetics and pharmacodynamics. Studies have predominately focused on the former, especially metabolic differences mediated by the cytochrome P-450 system. The activity of CYP 2D6, which metabolizes many antipsychotics and antidepressants, is phenotypically divided into ultrarapid, extensive, intermediate, and poor metabolizers. While Whites tend to have higher rates of poor metabolizers (5% to 10%), up to 70% of East Asians are intermediate metabolizers, accounting for the clinical observation that they may require lower doses of psychotropics. Conversely, up to 20% to 30% of Arabs and Ethiopians may be ultrarapid metabolizers. At a given dose, clozapine tends to have higher plasma concentration in Asians compared to Whites, likely due to differences in CYP 1A2 and CYP 2C19 activity. The activity of enzymes CYP 1A2 and CYP 3A3/4 is significantly affected by inducers and inhibitors, and drug interactions need to be carefully considered. Cultural differences in diet, use of medicinal herbs, and environmental factors become important variables, through their role as potential inducers or inhibitors. For instance, grapefruit juice can inhibit CYP 3A3/4, and should be avoided in patients taking ziprasidone, clozapine, pimozide, buspirone, carbamazepine, tricyclic antidepressants, and sildenafil. In reviewing therapeutic drug monitoring studies, a recent study found that the plasma concentrations of haloperidol, clozapine, quetiapine, aripiprazole, and lurasidone, which are metabolized by CYP 2D6, 1A2, or 3A4, were higher among East Asian populations than Western populations at the same daily dose. Some studies have found that migrants who have adopted Western diets exhibit changes in their metabolism. Other observed ethnic differences, such as variation in the concentration and distribution of alpha 1-acid glycoprotein, may also be significant; in this case, protein-binding may be affected and only the nonbinding fraction of a drug can penetrate the central nervous system.

In terms of pharmacodynamics, a meta-analysis of 30 studies found that genetic polymorphism of the serotonin transporter-linked promoter region (5-HTTLPR) may be associated with antidepressant response and remission among Whites, with the l-allele being associated with better response and remission rates with the SSRIs, but it does not appear to play a major role among Asians. This finding is supported by a 2020 meta-analysis of 49 studies on antidepressant response in major depression and a 2021 meta-analysis of 16 studies on antidepressant response in psychiatric disorders other than major depression. In terms of side effects, African Americans, compared to Whites, tend to have an increased risk of tardive dyskinesia in response to first-generation antipsychotics (FGA). Relative to Whites, African Americans, Asians, Hispanics, and AI/AN groups have a higher prevalence of type-II diabetes, and African Americans may be more prone to new onset of type-II diabetes and diabetic ketoacidosis from some SGA. A recent quasisystematic review found that some genes were associated with altered efficacy and safety of atypical antipsychotics among Blacks but not Whites, while other genes were associated with improved treatment response only among Whites but not Blacks. It was also observed that there is an increased risk of metabolic side effects and neutropenia with atypical antipsychotic use among Blacks compared to Whites. Since September 2015, the FDA has revised recommendations permitting patients with benign ethnic neutropenia (BEN) to continue treatment with clozapine until the absolute neutrophil count is less than 500 cells per microliter. As BEN is more commonly reported in individuals of African, Caribbean, Middle Eastern, or West Indian descent, this ethnic variation has previously contributed to the underutilization of clozapine among African Americans.

Ethno-psychopharmacologic studies have limited implications for clinical practice due to methodologic issues, imprecise definitions of race, impact of nongenetic factors such as diet, and the inapplicability of population data to individuals. For example, advocates have suggested renaming BEN to benign familial neutropenia to move away from race/ethnicity-based medicine. These studies do reveal the problem of ignoring diversity during drug development, resulting in recommended dosages and guidelines that may not be the most appropriate for diverse populations. For example, the generally recommended therapeutic level of lithium for maintenance treatment of bipolar disorder is higher in the United States than in parts of Asia. As studies suggest, ignoring diversity may also lead to erroneous conclusions in pharmacogenomics. Increasingly, the availability of CYP-P450 testing allows clinicians and patients to make more informed decisions about medications. Pragmatically, the “start low and go slow” approach may be considered for every patient.

Psychosocial, cultural, and religious factors are important considerations when prescribing medications. At a societal level, there is growing recognition of clinician disparities in prescription practices, such as increased use of depot antipsychotics and FGAs and decreased use of clozapine and SGAs with African American patients compared to Whites. Patients or families may have cultural beliefs about Western medications that influence treatment. For example, some tenets of traditional Chinese medicine may consider Western medications to be “hot” and dangerous to the homeostatic balance of the body if taken long-term, unless countermeasures are implemented. A nonjudgmental exploration of a patient’s adherence to medications can help uncover these beliefs. For example, the clinician may ask “Most people miss taking medications here and there… on average, how many times a week were you able to take your medications? What were some of the reasons or barriers?” It is also instructive to ask patients to bring in their medications and explain how they are taken. Patients may have idiosyncratic or shared cultural reasons for not taking medications. Religious beliefs should also be considered. For example, although there may be provisions for Muslims and Jews to consume certain food products as medicine that may otherwise be forbidden, many may choose to avoid such medications, such as capsules made with gelatin derived from pork. Fasting, based on personal choice, cultural beliefs, or spiritual practices, such as Islamic practices during Ramadan, may affect patients' choices when taking medications. In most instances, the clinician can collaborate with the patient and/or community/faith leaders to find a workable solution.

Psychotherapy. Social and cultural factors can greatly influence the practice of psychotherapy, including access to culturally appropriate psychotherapy. The culture of psychotherapy itself can be examined using a three-tiered model: transtheoretical, metatheoretical, and intratheoretical. These perspectives can offer the clinician a new perspective on the psychotherapy model that is being employed, as well as a more flexible and culturally competent way of conducting psychotherapy.

From a transtheoretical cultural perspective, psychotherapy has been compared to other forms of healing across cultures. As described by Jerome Frank, common healing methods such as psychotherapy typically help patients combat demoralization by means of an emotionally charged, confiding relationship with a helping person in a healing setting, using a ritual or procedure based on shared beliefs and rationales. This universalizing perspective reveals common therapeutic factors contributing to healing outcomes, of which therapeutic alliance is one of the most important. Identifying common essential psychotherapy elements that may be universal is also a helpful strategy in developing scalable interventions used in global mental health (GMH) efforts to increase access to care.

From an intratheoretical cultural perspective, work has increasingly focused on developing cultural adaptations of psychotherapies for specific populations. Using a mixed-methods approach, this is often done by conducting focus groups with the targeted communities to identify cultural adaptations needed for an evidence-based therapy, followed by examining the adapted therapy's effectiveness in a randomized controlled trial. Clinicians may also carry out case-by-case adaptations based on the social and cultural backgrounds of specific patients to adapt the therapy to the patient's needs. In general, adaptations may involve the use of culturally appropriate metaphors, proverbs, or idioms and simplification of complex abstract concepts, while maintaining fidelity to the original psychotherapy model and not challenging the culture of the model itself.

The meta-theoretical cultural perspective reflects on the underlying cultural philosophy and assumptions of various forms of psychotherapies. This may include two approaches: (i) goals, content, and process, and (ii) time, activity, and locus.

GOALS, CONTENT, AND PROCESS. Goals refer to the targeted endpoints of psychotherapy, content to the specific areas being worked on in therapy, and process to the way in which the therapy is conducted and the therapeutic process is hypothesized to unfold. For example, for a depressed patient, the goal may be to decrease depressive symptoms and restore behavioral functioning, the content may focus discussions on grief over the mother's death, and the process (depending on the type of therapy) may include catharsis, processing interpersonal loss and re-establishing social connections, or challenging cognitive errors. Each of these aspects reflects cultural assumptions; for example, that (a) depressive symptoms form a syndrome that influences behaviors and that both depressive symptoms and behaviors are legitimate targets of change; (b) grief is an appropriate and relevant area to focus on; (c) talking-through processes involving expressing emotions, building social networks, or changing thinking patterns are appropriate and helpful ways to achieve recovery. Different models of psychotherapy may vary in how each component is defined, and this, in part, reflects their cultural origins. For example, elements of classical psychoanalysis are influenced by cultural views related to Judaism, the Victorian era, and the emergence of Western science; cognitive-behavioral therapy is influenced by Greek stoicism; and mindfulness-based therapies, including acceptance and commitment therapy and dialectical behavioral therapy, are influenced by Buddhism and Asian philosophy. Although re-establishing social networks may seem on the surface to be a universal therapeutic process, there is a discernable difference between viewing the therapeutic process as one of increasing the security and social support system of an independent individual versus supporting the relinquishment of self-centered attachments and promoting the recognition that the interdependent self cannot be separately defined from one's present or past social network.

The patient may hold a range of views about the components of goals, content, and process based on personal and cultural values and beliefs, and these may conflict with one another, with the psychotherapy's underlying cultural assumptions, or with the clinician's views. If a relevant cultural issue or conflict is identified, the clinician can modify any of the three components collaboratively with the patient. Based on the patient's cultural background, the clinician can consider exploring with the patient the use of procultural strategies, harnessing culturally reinforcing or congruent strategies from among the diverse elements of the patient's culture, or countercultural strategies, empowering the patient to reflect on and challenge cultural scripts that may be harmful or unworkable in the patient's situational context. This approach embodies the flexible perspective that: cultural influences are embedded in an invisible way in the patient, the clinician, and the tool of psychotherapy itself; there is often a diversity of beliefs and values within any single identified cultural group; and various cultural elements may not always be adaptive in all contexts and can be functionally reshaped.

TIME, LOCUS, AND ACTIVITY. Another approach to examining cultural differences among psychotherapies is to identify how basic elements of psychotherapy, such as the hypothesized problem, solution, and therapeutic process, are defined along the three dimensions of time, locus of therapy, and activity. Along the time axis, therapies may situate the hypothesized problem, solution, or therapeutic process in the past, present, and/or future. In the locus of therapy dimension, the location of the problem, solution, or therapeutic process can be hypothesized as being internal (e.g., intrapsychic) or external (e.g., the sociocultural environment). In the activity dimension, the clinician, patient, or therapeutic process can be relatively active or passive; in this sense passivity is not understood as a form of giving up but as a quiet power of tranquility. This can be understood as akin to the concept of complementary forces of “yin” (passive, dark, feminine) and “yang” (active, light, masculine) in traditional Chinese Medicine and Daoism. For example, cultivating a mindfulness practice is an example of a present, internally focused form of powerful passivity in the clinician, patient, and therapeutic process. A cultural profile can be developed for different types of psychotherapy, which may or may not match that of the patient's. When there is a mismatch, a different therapy model may be chosen (e.g., problem-solving therapy instead of mindfulness practice); the therapy may be modified (e.g., laughing meditation); or the patient may choose to work with the clinician toward overcoming the incongruent experience and exploring a new way of approaching problems.

Countertransference and Transference. As noted in the fourth component of the OCF, cultural issues can manifest in countertransference and transference reactions, related to cultural differences and complicated by historical and present power differentials between the patient and the clinician (see Table 4.2–2). Without awareness, the clinician in a more dominant position may unwittingly oppress the patient in some instances, or in other situations, become ineffective when burdened by guilt from the knowledge of historical oppression perpetuated by the clinician's group of origin. Through the lens of intersectionality, there are instances where the clinician may feel disempowered when meeting with the patient, because of differences due to gender, race/ethnicity, sexual orientation, or religion. Identifying and working through cultural issues may help resolve problematic countertransference and transference issues and turn them into facilitative forces for healing.

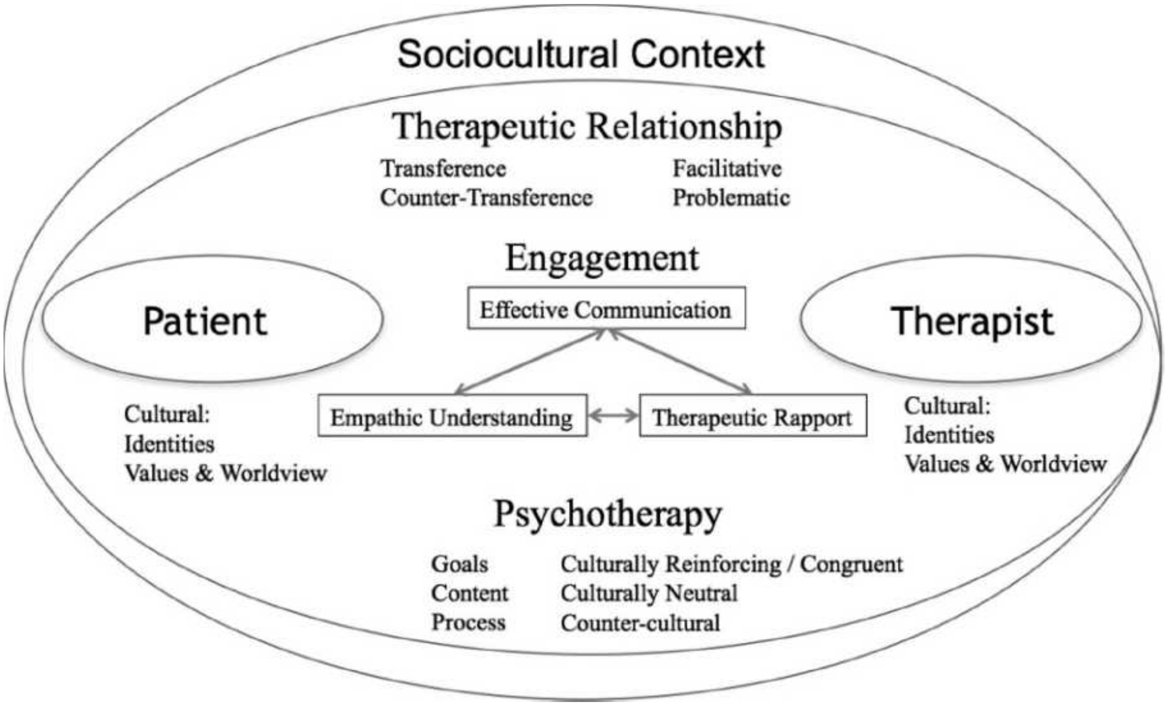
Sociocultural Context
Therapeutic Relationship
Transference
Counter-Transference
Facilitative
Problematic
Engagement
Effective Communication
Patient
Therapist
Empathic Understanding
Therapeutic Rapport
Cultural:
Identities
Values & Worldview
Cultural:
Identities
Values & Worldview
Psychotherapy
Goals
Content
Process
Culturally Reinforcing / Congruent
Culturally Neutral
Counter-cultural

FIGURE 4.2–2. Cultural considerations in a therapeutic relationship. (Reprinted with permission from Springer Nature. Fung K, Lo T. An integrative clinical approach to cultural competent psychotherapy. *J Contemp Psychother.* 2017;47:65–73. doi:10.1007/s10879-016-9341-8.)

Figure 4.2–2 summarizes cultural considerations in a therapeutic relationship. The patient and clinician/therapist each bring their own cultural identities, values, and worldview in a sociocultural context bridged by the engagement triad of communication, empathic understanding, and rapport. The therapeutic intervention’s goals, process, and content may be framed as procultural, countercultural, or cultural-neutral. Transference and countertransference may be facilitative or interfering.

GLOBAL PERSPECTIVE ON MENTAL HEALTH

The Global Burden of Mental Illness

Mental disorders are prevalent and may begin early in life, often resulting in an inability to work or function and, as a result, may engender prolonged periods of disability. Based on the Global Burden of Disease data from 2019, mental disorders, excluding substance use disorders and suicide, accounted for 4.9% of the disease burden as measured by disability-adjusted life years (DALYs), a measure of healthy years lost due to premature mortality and disability, and 14.6% of years lived with disability (YLDs). The estimate is based on data from 204 countries using DSM-IV and ICD-10 diagnostic classification systems. Another estimate that includes mental disorders, specific neurologic disorders such as dementias, substance-related disorders, and suicide re-estimates the total burden at 13.0% of DALYs, comparable to the burden from cardiovascular diseases at 15.4%. The DALYs for mental disorders were higher in many HICs compared to sub-Saharan Africa and Asia. However, interpretation is limited by the relatively lower coverage of data in the latter regions. Among mental disorders, depressive and anxiety disorders were the leading causes of disease burden worldwide, and the rates were higher in countries identified with high rates of childhood sexual abuse, intimate partner violence, and conflict and war. Schizophrenia and other psychotic disorders, while impacting on a smaller proportion of populations, are still concerning, and acute psychoses were accorded the highest disability weighting in estimates.

Based on three influential WHO studies, The International Pilot Study of Schizophrenia, the Determinants of Outcome of Severe Mental Disorder, and the International Study of Schizophrenia, it has been asserted that schizophrenia may have better outcomes in LMICs than in HICs. Possible explanations include supportive kin networks; spiritual beliefs about the causation of mental illness, which may reduce stigma, promote acceptance of those with mental illness, or orient them to religious-focused healing services; greater social acceptance both within families and in society; and opportunities to participate in the microeconomy with more flexible work roles. More recent studies question these findings. In addition to methodologic challenges in measuring recovery, there is wide variation in living conditions, mental illness stigma, and access to care among LMICs, with a likely selection bias in favor of the participating research sites that may not be representative of local populations. In a recent systemic review of studies conducted from 2000 to 2020, there was no difference in complete or social recovery rates from schizophrenia by geographical region. Comparisons between LMICs and HICs found that the only difference was that LMICs had better employment outcomes, lending partial support to previous observations. Outcomes were significantly better among individuals with first-episode psychosis versus multiple-episode psychosis, likely reflecting the heavy investments in early intervention among HICs. The annualized remission rate significantly declined after 2008, possibly due to the financial crash of that period and lending more support to the role of social determinants of health in schizophrenia outcomes.

Suicide, one of the leading causes of death, is estimated to be 1.3% worldwide in 2019, accounting for more deaths than malaria, HIV/AIDS, breast cancer, war, or homicide. In 2019, the global age-standardized suicide rate was 9 per 100,000. Crosscountry differences ranged from 2 to 80 per 100,000. Most of the world's suicides occurred in LMICs (77%), where most of the world's population lives. However, HICs had the highest rate at 10.9 per 100,000 compared to 10.1, 9.9, and 7.3 per 100,000 for low, lower-middle, and upper-middle-income countries respectively. Over the past 20 years, the global suicide rate has decreased by 36%, but the Americas is the only region bucking this trend with a 17% increase. The top five countries with the highest suicide rates are: Lesotho (72.4 per 100,000), Guyana (40.3), Eswatini (29.4), South Korea (28.6), and Kiribati (28.3).

Globally, the rate of death by suicide for males was at 12.6 per 100,000 and for females at 5.4 per 100,000. The gender ratio varied, with the highest ratio at over 3 in HICs, compared to 2.9, 2.6, and 1.8 for low, upper-middle, and lower-middle-income countries respectively. The countries with lower gender ratios appear to be characterized by less egalitarian gender norms and higher gender discrimination against women. A meta-analysis of studies between 2006 and 2017 in Canada, the United States, the United Kingdom, and Australia found that males in rural areas were more likely to die by suicide than their male counterparts in urban areas (RR 1.41), while the difference was not significant for women. The increased risk of rurality for male suicides is consistent with another literature review of 24 studies around the world.

Treatment Gap in Access to Psychiatric Treatment

Despite limitations, the World Mental Health initiative has provided essential insights into mental health in different national populations. The crossnational variation in disorder severity is closely associated with access to treatment. In HICs, 35.5% to 50.3% of the population is diagnosed with some form of mental illness, and, in LMICs, 76.3% to 85.4% has received no treatment for their diagnosed condition in the previous 12 months. The proportion of people diagnosed with a mental illness who receive psychiatric treatment correlates with the percentage of the countries' gross domestic product spent on health care. The seriousness of the disorder correlates with service utilization but varies widely across nations.

The treatment gap is the difference between the true prevalence of a disorder and the treated proportion of individuals affected by the disorder. Data on the treatment gap between high- and low-income countries are beginning to emerge. Data from the Americas illustrate this disparity. The treatment gap for anxiety disorders, mood disorders, and substance use disorders in North America is 56.5%, 58.4%, and 69.1%, while for Latin America, it is 76.7%, 73.7%, and 83.7%, respectively. The treatment gap for schizophrenia in LMICs is alarming. The treatment gap in 50 LMICs was estimated to be 69%, and among the low-income countries, 89%. In an analysis of 35 Latin American and Caribbean countries, the treatment gap for schizophrenia was 56.4%, with the highest rate in Central America (79.4%). To reduce disability and the economic burden associated with mental disorders, it is necessary to bridge the treatment gap and reduce the treatment lag time. The implications of not reducing the treatment gap for the mentally ill are profound, including worsening poverty and lower socioeconomic status for the families of mentally ill individuals, impaired family functioning, decreased educational attainment, poorer quality of life, and increased mortality. To reduce the treatment gap, scaling up services and task shifting need to occur. In addition, many more people need access to effective mental health services that are culturally appropriate, considering community beliefs and attitudes.

To address the global treatment gap, the U.S. National Institute of Mental Health has identified 25 grand challenges and summarized them into six goals:

- Identify root cause, risk, and protective factors of mental disorders
    - Identify modifiable social and biologic risk factors across the life course
    - Understand the impact of poverty, violence, war, migration, and disaster; and identify biomarkers
- Advance prevention and implementation of early intervention strategies
    - Support community environments that promote physical and mental well-being throughout life
    - Reduce the duration of untreated illness by developing culturally sensitive early interventions across settings
    - Develop interventions to reduce the long-term negative impact of low childhood socioeconomic status on cognitive ability and mental health
    - Develop an evidence-based set of primary prevention interventions for a range of mental disorders
    - Develop locally appropriate strategies to eliminate childhood abuse and enhance child protection
- Improve treatments and expand access to care: integrate screening and core packages of mental health services into routine primary health care; reduce the cost and improve the supply of effective medications
    - Develop effective treatments for use by nonspecialists, including lay health workers with minimal training
    - Incorporate functional impairment and disability into the assessment
    - Provide effective and affordable community-based care and rehabilitation
    - Improve children’s access to evidence-based care by trained health providers in LMICs
    - Develop mobile and IT technologies, such as telemedicine, to increase access to evidence-based care
- Raise awareness of the global burden of untreated illness

    - Develop culturally informed methods to eliminate the stigma, discrimination, and social exclusion of patients and families across cultural settings
    - Establish cross-national evidence on the cultural, socioeconomic, and services factors underlying disparities in incidence, diagnosis, treatment, and outcomes
    - Develop valid and reliable definitions, models, and measurement tools for quantitative assessment at the individual and population levels for use across cultural settings
    - Establish shared, standardized global data systems for collecting surveillance data on the prevalence, treatment patterns, and availability of human resources and services
- Build human resource capacity: increase capacity in LMICs
    - Create regional centers for mental health research, education, training, and practice that incorporate the views and needs of local people
    - Develop sustainable models to train and increase the number of culturally and ethnically diverse lay and specialist providers to deliver evidence-based services
    - Strengthen the mental health component in the training of all healthcare personnel
- Transform the health system and policy response: establish and implement minimum healthcare standards for mental disorders around the world
    - Redesign health systems to integrate mental disorders with other chronic disease care
    - Create parity between mental and physical illness through investment in research, training, treatment, and prevention
    - Incorporate a mental health component in international aid and development programs

Global Mental Health

The basic tenet of the GMH movement is that every person in the world has an inherent right to mental health from a CEI and social justice perspective. There has been a focus on addressing treatment gaps by scaling-up evidence-based interventions and increasing access to care. A common strategy employed in low-resource settings is to engage in task shifting or task sharing, where nonmental health specialists such as lay community members or traditional healers may be trained to outreach to communities to engage in stigma-reduction work, promote awareness through psychoeducation, deliver psychosocial intervention, or support access to psychiatric treatment. One successful example is the “Friendship Bench” project, where trained lay community health workers are stationed on benches in accessible places to deliver problem-solving therapy to alleviate kufungisisa, the symptom of “thinking too much,” in addressing the treatment gap for depression and anxiety in Zimbabwe. The intervention group led to greater improvements as measured by the SSQ-14 and PHQ-9 compared to standard psychiatric care enhanced with information, education, and support. Its success has also led to its implementation in New York City.

As the field develops, concerns have been raised that Western conceptions of mental disorders may dominate over other alternative concepts as a form of intellectual colonialism and epistemic injustice; diverse and unique cultural expressions of distress and pathways for healing may be ignored or endangered; interventions may be unduly influenced by pharmaceutical industries; social determinants of health and mental health may be ignored; and initiative agendas and gains may be coopted by Western researchers. These remain legitimate concerns that GMH clinicians and researchers need to be mindful of and reflect on.

In part in response to these criticisms, the most recent Lancet Commission on Global Mental Health report in 2018 broadens the GMH agenda in the context of the U.N. Sustainable Development Goals. The aim is to extend the agenda beyond decreasing the treatment gap to also include closing the quality gap and the prevention gap. It proposes four foundational pillars: (1) mental health is for the global public good and is relevant to all countries; (2) mental health exists in a continuum from distress to disorder to disability rather than a binary approach of diagnosable mental disorders; (3) the mental health of each person is a product of social and environmental influences interacting with genetic, neurodevelopmental, and psychological processes and affecting biologic brain pathways; and (4) mental health is a fundamental human right.

The reframed and broadened agenda entails six key actions: (1) mental health services need to be scaled up as an essential component of universal health coverage and integrated with other global health priorities; (2) barriers to mental health need to be addressed including stigma, climate change, and growing inequalities; (3) mental health needs to be protected by public policies and developmental efforts with intersectoral engagement of stakeholders within and beyond the health sector such as education, workplace, social welfare, criminal justice, child and youth services, humanitarian efforts, and others; (4) innovation need to be embraced, including the use of non-specialist individuals and digital technologies; (5) investments in mental health should be increased in view of the strong evidence of the intersection between economics and health; and (6) investments in research and innovation should be increased in diverse disciplines such as genomics, neuroscience, health services, clinical sciences, and social sciences both to promote discovery of causal mechanisms and research in implementation science.

Implementation science is the scientific study of methods to promote systemic uptake of evidence-based practices to enhance access to care as well as quality and effectiveness of care. It is often used as a framework to guide GMH research or initiatives. An example of a common implementation science framework is the RE-AIM framework, which includes: Reach, examining the number and characteristics of individuals who can or cannot participate in the initiative; Effectiveness, assessing the outcomes of the initiative; Adoption, gauging uptake at organizational and systemic levels; Implementation, examining the process of conducting the interventions in real-world settings, including the balance of local cultural adaptation and fidelity; and Maintenance, evaluating the enduring effects at individual and organizational levels. While being mindful of overarching project goals and ethical considerations, the key to applying the implementation science framework successfully in GMH contexts includes equitable collaboration with local partners and stakeholders; careful cultural adaptation of concepts and interventions; and flexible adjustment of project parameters to suit local needs and changing contexts.

The challenge of the grand mission of GMH, and indeed psychiatry, in alleviating human suffering remains daunting and its methods are not without pitfalls and criticisms. There may always be a tension between an “etic” perspective–an outsider, universalizing approach–versus an “emic” perspective–an insider, particularistic cultural approach. Rather than an either-or position, a dialectical approach is possible. The integration of cultural psychiatry into GMH and the field of psychiatry at large will lead to improved research, training, clinical practice, and ultimately, patient and societal outcomes.

FURTHER READINGS

Alvarez K, Fillbrunn M, Green JG, et al. Race/ethnicity, nativity, and lifetime risk of mental disorders in US adults. *Soc Psychiatry Psychiatr Epidemiol*. 2019;54:553–565.

Bailey ZD, Krieger N, Agénor M, Graves J., Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. *Lancet*. 2017;389:1453.

Barr PB, Bigdeli TB, Meyers JL. Prevalence, comorbidity, and sociodemographic correlates of psychiatric disorders reported in the All of Us Research Program. *JAMA Psychiatry*. 2022;79:622–628.

Beugelsdijk S, Maseland R, van Hoorn A. Are scores on Hofstede's dimensions of national culture stable over time? A Cohort Analysis. *Glob Strategy J*. 2015;5:223–240.

Bourhis RY, Moise LC, Perreault S, Senecal S. Towards an interactive acculturation model: a social psychological approach. *Int J Psychol*. 1997;32:369–386.

Compton MT, Shim RS. The social determinants of mental health. *Am Psychiatr Publ*. 2015;13(5):419–425.

Fung K, Lo T. An integrative clinical approach to cultural competent psychotherapy. *J Contem Psychother*. 2017;47:65–73.

Global Burden of Disease 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry*. 2022;9:137–150.

Gone JP, Hartmann WE, Pomerville A, Wendt DC, Klem SH, Burrage RL. The impact of historical trauma on health outcomes for indigenous populations in the USA and Canada: a systematic review. *Am Psychol*. 2019;74:20–35.

Im H, Rodriguez C, Grumbine JM. A multitier model of refugee mental health and psychosocial support in resettlement: toward trauma-informed and culture-informed systems of care. *Psychol Serv*. 2021;18:345–364.

Jerome RN, Pulley JM, Sathe NA, et al. Exploring biologic predictors response disparities to atypical antipsychotics among Blacks: a quasi-systematic review. *Ethn Dis*. 2020;30:229–240.

Kirmayer LJ, Fung K, Rousseau C, et al. Guidelines for training in cultural psychiatry. *Can J Psychiatry*. 2021;66:195–246.

Kitayama S, Varnum MEW, Salvador CE. Cultural neuroscience. In: Cohen D, Kitayama S, eds. *Handbook Of Cultural Psychology*. Guilford Publications; 2019.

Kohn R, Ali AA, Puac-Polanco V, et al. Mental health in the Americas: an overview of the treatment gap. *Rev Panam Salud Publica*. 2018;42:e165.

Leaune E, Dealberto MJ, Luck D, et al. Ethnic minority position and migrant status as risk factors for psychotic symptoms in the general population: a meta-analysis. *Psychol Med*. 2019;49:545–558.

Lin SK. Racial/ethnic differences in the pharmacokinetics of antipsychotics: focusing on East Asians. *J Pers Med*. 2022;12:1362.

Maura J, Weisman de Mamani A. Mental health disparities, treatment engagement, and attrition among racial/ethnic minorities with severe mental illness: a review. *J Clin Psychol Med Settings*. 2017;24:187–210.

Nisbett RE, Masuda T. Culture and point of view. *Proc Natl Acad Sci USA*. 2003;100:11163–11170.

Olbert CM, Nagendra A, Buck B. Meta-analysis of Black vs. White racial disparity in schizophrenia diagnosis in the United States: do structured assessments attenuate racial disparities. *J Abnorm Psychol*. 2018;127:104–115.

Paradies Y, Ben J, Denson N, et al. Racism as a determinant of health: a systematic review and meta-analysis. *PLoS One*. 2015;10:e0138511.

Patanè M, Ghane S, Karyotaki E, et al. Prevalence of mental disorders in refugees and asylum seekers: a systematic review and meta-analysis. *Global Mental Health*. 2022;9:250–263.

Patel V, Saxena S, Lund C, et al. The Lancet Commission on global mental health and sustainable development. *Lancet*. 2018;392:1553–1598.

Payne HE, Steele M, Bingham JL, Sloan CD. Identifying and reducing disparities in mental health outcomes among American Indians and Alaskan Natives using public health, mental healthcare and legal perspectives. *Adm Policy Ment Health*. 2018;45:5–14.

Ren F, Ma Y, Zhu X, Guo R, Wang J, He L. Pharmacogenetic association of bi- and triallelic polymorphisms of SLC6A4 with antidepressant response in major depressive disorder. *J Affect Disord*. 2020;273:254–264.

Sagiv L, Roccas S, Cieciuch J, Schwartz SH. Personal values in human life. *Nat Hum Behav*. 2017;1:630–639.

Salas-Wright CP, Vaughn MG, Goings TC, Miller DP, Schwartz SJ. Immigrants and mental disorders in the United States: new evidence on the healthy migrant hypothesis. *Psychiatry Res*. 2018;267:438–445.

Shao Z, Richie WD, Bailey RK. Racial and ethnic disparity in major depressive disorder. *J Racial Ethn Health Disparities*. 2016;3:692–705.

Sue DW, Sue D. *Racial/Cultural Identity Development in People of Color: Counseling Implications. Counseling the Culturally Diverse: Theory and Practice*. John Wiley & Sons, Incorporated; 2016.

Vilsaint CL, NeMoyer A, Fillbrunn M, et al. Racial/ethnic differences in 12-month prevalence and persistence of mood, anxiety, and substance use disorders: variation by nativity and socioeconomic status. *Compr Psychiatry*. 2019;89:52–60.

Williams DR, Lawrence JA, Davis BA, Vu C. Understanding how discrimination can affect health. *Health Serv Res*. 2019;54 Suppl 2:1374–1388.

4.3 Evolutionary Foundations for Psychiatric Research and Practice

RANDOLPH M. NESSE, M.D.

FIVE ADVANCES

Major advances in the evolutionary explanation of behavior and emotions from the second half of the 20th century have yet to be fully applied in psychiatry. This chapter summarizes those advances and how understanding them can make clinicians and researchers more effective. Some have direct applications, such as understanding the human mating system, how negative emotions are useful, and why there are no common alleles with large effects on highly heritable diseases such as schizophrenia. However, the big payoffs, in the long run, will come from the integrated framework evolution provides for understanding normal behavior and psychopathology. Evolutionary approaches are not an alternative to understanding mechanisms; they are the crucial complement that explains why mechanisms are the way they are and why they are vulnerable to failure. Evolutionary biology provides a framework that can transform the bio-psycho-social model from a catchphrase into a solid working model. It provides, for psychiatry, what physiology provides for the rest of medicine—an understanding of normal functioning as the foundation for understanding pathology.

Five major advances in evolutionary biology are especially important for psychiatry. Each is described briefly below.

Proximate and Evolutionary Explanations Are Both Essential

Until the middle of the 20th century, the study of animal behavior was mostly descriptive. Meticulous observations by ethologists described behavior in detail, but theory was absent. Everything changed in the mid-1960s, with the recognition that all biologic traits need evolutionary as well as proximate explanations. This insight came first from students of animal behavior, especially Niko Tinbergen. So-called “proximate questions” are about mechanisms; answers to such questions describe how mechanisms work and how they develop. Other “evolutionary questions” are about how traits got to be the way they are; answers to such questions describe how and why the trait changed over evolutionary time. For instance, a proximate explanation of anxiety describes the relevant brain and psychological mechanisms, how they develop, and how they are influenced by events and environmental influences. However, even describing every detail about the proximate mechanism does not explain why the capacity for anxiety exists at all. That requires an evolutionary explanation of how the capacity for anxiety increased fitness for our ancestors.

In a famous article published in 1963, Tinbergen outlined four questions that all must be answered to provide a complete understanding of any trait. Two are evolutionary questions. The first is about phylogeny, the history of a trait. The second is about the adaptive significance of the trait and what forces of selection shaped it. There are also two proximate questions. The first is about the mechanism and how it works at all levels, from molecules to social psychology. The second is about ontogeny, how the mechanism develops in an individual from a zygote to an adult (see Fig. 4.3–1). Almost all research in psychiatry has been to describe mechanisms, their development, and their malfunctions. Only now is attention turning to also address evolutionary questions, such as why emotions exist.

Behaviors Need an Evolutionary Explanation

The second core advance, closely related, was the application of evolutionary thinking to behavior. Initial approaches provided evolutionary explanations for specific behaviors, such as rams butting horns in the mating season and birds removing broken shells from nests to avoid attention from predators. It was soon recognized, however, that it was not behaviors themselves that natural selection shaped but behavior regulation mechanisms. That is, the brain.

A general core principle soon became clear: **brains are shaped by natural selection to give rise to behavior that maximizes Darwinian fitness**. Instead of implying rigid behavior patterns, this principle recognizes the enormous fitness benefits of flexible behaviors that allow organisms to adapt to rapidly changing situations. Learning is not an alternative to an evolutionary explanation. It is a capacity shaped by natural selection. Textbooks of animal behavior are now all grounded in evolutionary biology.

Social Behaviors Need Evolutionary Explanations

The evolutionary analysis of social behavior was central to transforming animal behavior research. The old idea that organisms are shaped to behave in ways that benefit their groups and species was replaced by a recognition that selection generally shapes traits that maximize individual fitness, even if that harms the group. Many have heard about lemmings jumping into the sea for the good of the group when food is too scarce for all to survive. This turns out to be a myth, one promulgated by a Disney film that seems to show lemmings flinging themselves off a cliff when in fact, they were swept off by workers with brooms. An allele that induces individuals to sacrifice themselves for the good of the group will be quickly selected out except in certain special circumstances. Behaviors costly to individuals that benefit groups, such as parents sacrificing for their offspring, wolves sparing those who signal defeat, or crowded populations reducing reproduction, have other explanations.

Tinbergen's Four Questions		Two Objects of Explanation	
		Diachronic The historical/developmental sequence that shaped the trait	**Synchronic (single form)** The trait at one slice in time
Two kinds of explanation	**Proximate** Describes the mechanism and its ontogeny	**Ontogeny** How does the trait develop in individuals?	**Machanism** What is the structure of the trait; how does it work?
	Evolutionary Describes the adaptive significance and phylogeny	**Phylogeny** What is the phylogeny of the trait?	**Adaptive significance** How have variations in the trait influenced fitness?

FIGURE 4.3–1. Tinbergen's Four Questions. (Adapted from Nesse RM. Tinbergen's four questions, organized: a response to Bateson and Laland. *Trends Ecol Evol*. 2013;28(12):681–682. Copyright © 2013, with permission from Elsevier. doi: 10.1016/j.tree.2013.10.008.)

Kin Selection Explains a Lot of Social Behavior

Recognition that costly individual sacrifices cannot routinely be explained by benefits to the group was followed closely by William Hamilton's discovery of the main exception, kin selection. Close relatives share genes that are identical by descent; 50% of a child's genes are identical to each of its parents. Alleles that decrease an individual's number of offspring can be selected for if they sufficiently benefit identical alleles in kin. Hamilton's famous formula, $C < B \times r$, is a crucial foundation for understanding family relationships. An action that costs an amount C can nonetheless give inclusive fitness benefits if C is less than the benefit to a relative, B, times the percentage of genes identical by descent. In the classic example, it is not worthwhile to sacrifice one's life for a sibling, but it is for two siblings or eight cousins. The term "inclusive fitness" incorporates the reproduction of an individual's relatives as well as the individual. Kin selection offers a profound evolutionary explanation for basic human behaviors, such as parents making sacrifices that benefit their children.

This discovery led to thinking in terms of “selfish genes.” The metaphor has been useful to combat naïve group-selection thinking, but it has been somewhat too potent for its own good. If genes are selfish, this seems to imply that organisms must be selfish. However, capacities for generosity and moral behavior exist because they increase inclusive fitness. Selfish genes make individuals who can be generous…selectively.

Evolutionary Medicine

Natural selection shapes traits that work well, so it would seem unable to explain diseases. That’s correct—diseases are not shaped by natural selection, and viewing diseases as adaptations is a serious but common mistake. However, traits that leave individuals vulnerable to disease, such as the appendix, the narrow birth canal, and the capacity for low mood, have been shaped by natural selection. Evolutionary medicine tries to understand, for each disease, why selection has left bodies vulnerable. Six main kinds of explanation help to organize the search. Notice that these are not explanations for why some individuals become ill while others do not. Instead, they are about why all humans share vulnerability to a disorder.

Six evolutionary explanations for vulnerability

1. Mismatch with modern environments
2. Constraints–things selection cannot do
3. Co-evolution with fast-evolving pathogens
4. Tradeoffs
5. Reproduction at a cost to health
6. Defenses that are useful but costly and aversive

Mismatch of the Environment to the Body. Most chronic disease results from the mismatch between our bodies and the modern environments we have created. Atherosclerosis, breast cancer, and autoimmune diseases are vastly more common in technological societies. So are obesity, eating disorders, alcoholism, and drug abuse. Our behavior regulation mechanisms were never shaped to cope with the ready availability of hamburgers, candy bars, whiskey, and heroin. Our distant ancestors also never had to cope with life in bureaucracies, 9-to-5 jobs, alarm clocks, and expectations of lifetime monogamous mating, they were, instead, responding to different, often harsher environments with different survival needs.

Many alleles that predispose to major disorders are not defects; they are merely “quirks” that cause harm only in modern environments. For instance, nearsightedness is highly heritable but nonetheless rare in hunter-gatherer populations not exposed to reading early in life. Other heritable tendencies may have vastly different significance in modern environments. For instance, a short attention span may not have posed serious problems for hunter-gatherers. Some who advocate an evolutionary approach to mental disorders have incorrectly attributed nearly all problems to a mismatch with modern environments; such factors are significant, but the five other factors are also important.

Pathogens Evolve Faster Than Hosts. Infectious disease will always be with us because pathogens evolve much faster than we can. Also, every time selection shapes a host defense, pathogens evolve some way around it. The resulting arms race leaves us with costly defense mechanisms prone to causing disease.

For instance, streptococci have evolved immune markers similar to human tissues, making it risky for our immune system to attack them. Rheumatic fever, Sydenham’s chorea, and some cases of obsessive-compulsive disorder are examples. Selection shapes immunologic defenses to high levels of aggressiveness, whose associated effects can cause mental disorders. For instance, inflammation now appears to cause some depression and most Alzheimer disease.

Constraints. There are many things that natural selection cannot do. It can never completely prevent mutations, although the extraordinary fidelity of DNA replication illustrates the power of selection. Evolution also cannot redesign a trait from scratch; it proceeds only by the process of small “tinkering” changes. So, while birth via the pelvis is painful and dangerous, no alternative route can evolve. Similarly, vertebrates are stuck with the blind spot where nerves and vessels enter in the back of the eyeball, while cephalopods have nerves that traverse their eyeballs without causing a blind spot. Some human cognitive biases and heuristics that limit objectivity result from the continuity of mechanisms shaped to make simpler decisions. Emotion regulation mechanisms likewise often fail because they are products of tinkering with preexisting mechanisms.

Tradeoffs. No trait can be perfect. The vast majority of changes that could make a trait superior in one respect would make it inferior in other respects or would compromise the effectiveness of other traits. Finches with larger beaks can crack tougher seeds, but they have less access to very fine seeds. Individuals with more tendency to anxiety are able to avoid dangers but at the cost of the inability to take advantage of many opportunities.

Reproduction at the Expense of Health. It is disturbing but essential to recognize that natural selection does not shape organisms to maximize health, happiness, or longevity. It instead maximizes inclusive fitness. Usually, health and long life are good for fitness. However, an allele that harms health will nonetheless spread if it increases reproduction.

The best example is the short lives of men compared to women. Male mortality rates in early adulthood are three times those for women in modern societies because tendencies that lead to success in the mating competition are selected for in males even if they harm health. Tendencies to tissue repair and sensible caution are relatively favored in females.

Much research has sought to understand the proximate mechanisms that account for women having more anxiety disorders than men. An evolutionary framework suggests that it is men who have the problem. Women have closer to the optimal amount of anxiety for their own health and well-being, while men tend to be reckless in ways that benefit their genes

Some have speculated that tendencies to mania may have been selected for because they result in increased matings and increased reproductive success. If that were really the case, then tendencies would spread and become universal. This provokes an interesting possible explanation for the high prevalence of cyclothymia. However, the rarity of manic depression suggests that the increased sexuality of individuals with tendencies to mania has not paid off with net increases in the number of grandchildren.

Defenses. Most problems people bring to their physicians are not direct manifestations of disease; they are useful defensive responses, such as pain, fever, cough, nausea, and vomiting. These defenses, shaped by natural selection, are usually expressed only in the presence of some problem or disease, but they are not diseases themselves. This knowledge is crucial for making good clinical decisions about when to block such defenses.

Understanding the smoke detector principle is also crucial for making such decisions well. Natural selection shapes mechanisms that express defenses, such as panic, whenever the benefits are greater than the costs. However, the presence of a threat is often uncertain, so even an optimal system will have some false alarms and some failures to respond when a response would be useful. This is particularly germane for understanding panic attacks. A fight–flight response will be potentially lifesaving if the noise from behind a rock is made by a lion, but it will be wasted if the noise is coming from a monkey. Whether an individual should flee depends on how likely it is that the noise was from a lion. If the cost of flight is, for instance, 100 kcal, and the cost of not fleeing if the lion is actually present is 100,000 kcal, then the optimal system will set off the fight–flight response whenever the noise is loud enough to indicate a probability of a predator being present of greater than one in 1,000. This means that 999 responses out of 1,000 will be false alarms, each useless but completely normal. False alarms in defense regulation systems are expected and normal. This has profound implications for understanding and treating not only panic disorder but also depression, pain, cough, and fever.

CORE PRINCIPLES OF EVOLUTIONARY BIOLOGY

Evolution has two major components, both discovered by Darwin. The first is the unity of all life; all individuals in all species are descendants of ancestors that go back in an unbroken chain to the origins of life on earth. This is phylogeny, now traceable via genomics. It is relevant for tracing human origins and for identifying the history of alleles related to mental disorders, but our focus here will be on Darwin's second discovery: natural selection and how it explains adaptations. It is the only viable explanation for why so many aspects of organisms are well-suited to coping with life's challenges. The rabbit's astute hearing, the woodpecker's barbed tongue, and the eagle's sharp eyes are products of natural selection. So are human empathy, emotions, cognition, capacity for language and love, and jealousy. Or, more exactly, the brain and its mechanisms for regulating thinking, emotions, and behavior.

Natural Selection

Natural selection is best understood in the context of selection in general. If members of a group differ in ways that influence their likelihood of being in the group in the future, the group will change over time. If you put all coins from your pocket into a jar each evening and select only the silver-colored coins in the morning, the jar will gradually take on a copper hue as the concentration of pennies increases. Television producers try out extraordinarily varied offerings. Those that get viewers and advertising dollars persist and are imitated; others are eliminated. Selection explains how groups change when variations among individuals influence their contributions to future groups.

Natural selection is the subtype of generic selection in which the variations are products of genetic variation that influence reproductive success. The principle is simple. If individuals in a population differ genetically in ways that influence how many surviving offspring they leave, the average characteristics of the population will change over the generations. This is not a theory; it is a theorem that is necessarily true. In the classic example, if previously light trees become covered with soot and moth color varies, light-colored months are more likely to be eaten by birds, so the average moth will become darker over succeeding generations. VISTA is a useful mnemonic for the process of natural selection: Variation, Inheritance, Selection, Time → Adaptation.

Dogs offer a familiar example. From their common ancestors in wolves, artificial selection has produced dramatically diverse dog breeds in just a few thousand years. The process of natural selection is the same as breeding, except that differences in offspring numbers result not from decisions made by breeders but from variations in the ability to cope with challenges that influence reproductive success. The products of natural selection look as if they were designed, and the sequence of species makes it seem as if there is something planned, or at least preordained, about the progression. However, there is no plan or endpoint, and traits that appear to be designed were shaped by mindless natural selection.

Standard examples usually describe how natural selection changes a trait. However, natural selection mostly keeps traits the same. If individuals 7 feet tall had more children than others, the mean height would shift to that optimum. If individuals more aggressive or outgoing than others consistently had more children, the population mean would shift. However, at least over the past few tens of thousands of years, individuals at the extremes of height, aggressiveness, and extroversion almost certainly had fewer offspring than those closer to the mean. The persisting wide variation in these traits suggests that fitness is similar across a wide range.

Whether a trait or a gene is advantageous or deleterious depends on the environment. In equatorial climates, dark skin protects against sun damage; in cloudy environments, it can result in rickets from insufficient sunlight-induced vitamin D synthesis. A tendency to prefer fatty foods can be useful where fat is scarce, but in modern environments, it can lead to disease.

Evolution

Evolution involves much more than natural selection. The three other main forces are mutation, migration, and genetic drift. **Mutations** happen despite being minimized to an astounding degree by natural selection; increasing evidence points to their importance in explaining schizophrenia, autism, and bipolar disorder. **Migration** may strongly influence the prevalence of an allele in a population. Most northern Europeans can digest milk because migrants brought the relevant gene from rapidly expanding dairying populations further south. Huntington’s chorea is especially common near Lake Maracaibo in Venezuela because Antonio Justo Doria, a Spanish sailor, brought the mutation there in the late 1800s. **Genetic drift** refers to changes in gene frequency, more correctly allele frequency, that result from random factors. An allele that increases fitness may be lost from an entire population if the individuals carrying it all die in an accident. An allele that is mildly harmful can nonetheless become more common if those who carry it have more children than others just by chance. Alleles whose effects on fitness are close to neutral can drift to fixation or be eliminated, especially in small populations. There are good reasons why population genetics relies on the so-called “null model,” in which frequency changes result only from drift. This is not, however, an alternative to natural selection or a reason to think that most traits are not well-suited to their functions.

What Selection Shapes

The usual textbook examples of selection are fixed physical traits such as bird beaks and moth wing color. Far more important for psychiatry, however, is selection of systems that regulate the body and behavior.

Most regulation systems keep things stable. Selection shapes mechanisms to maintain homeostasis at every level. Regulation of glucose, oxygen tension, acid–base balance, and osmolality are the usual examples.

The association of natural selection with fixed traits and systems that maintain homeostasis fosters the misperception that an evolutionary explanation somehow implies rigidity, especially rigidity of behavior. Nothing could be further from the truth. Natural selection has shaped hundreds of mechanisms that adjust organisms to changing situations. The traits vary in scale from the molecular to large-scale behavior, and across time from instantaneous changes to those that play out over a lifetime. The term *plasticity* often refers to longer-term changes, especially those that shape an individual in relatively irreversible ways as function of early experience, such as the barrel-shaped chest of humans living at high altitudes. In the medium term, high altitude increases hemoglobin concentration, and exposure to sunlight induces skin tanning. In the short term, cold induces shivering, infection causes fever, and threat arouses fear. Some responses, such as blinking, are nearly instantaneous. All are facultative adaptations, mechanisms shaped by selection that adjust phenotypes to changing aspects of environments.

The brain is obviously responsible for many such responses. It monitors internal and external cues as input to mechanisms that adjust physiology, expression, behavior, cognition, and more. These responses vary across individuals, and some of the variation results from genetic variation. To make a long story short, the brain was, like every other organ, shaped by natural selection. We know what it is for. It regulates behavior and other responses in ways that maximize inclusive fitness. Capacities for learning and relationships are especially valuable ways to accomplish that. Humans have, perhaps uniquely among all animals, capacities for abstract representation that make language and causal thinking possible. We also have distinctive capacities for enduring relationships that go beyond merely trading favors. These capacities, in turn, make possible social complexities and cultures that create new selection forces that further shape brains in a feedback process that has shaped our astounding social and cognitive abilities. An evolutionary approach to human nature emphasizes the remarkable plasticity and adaptability of behavior to different circumstances.

An example of potential importance for psychiatry is the possibility that selection has shaped a mechanism that detects and responds to early adversity by shifting development to create phenotypes that can cope with an anticipated short lifespan in a harsh environment. This could help account for the increased risk-taking, stress sensitivity, and inflammation and resulting chronic diseases that are more common in children who were abused early in life. However, it is hard to exclude the possibility that the associations arise from confounding genetic, environmental, or social factors or from individuals simply making good decisions in bad environments.

Genotypes and Phenotypes

Natural selection changes gene frequencies, more accurately, frequencies of specific alleles and other genetic variations. Population genetics describes how selection, migration, and genetic drift increase and decrease allele frequencies in different situations. This sometimes gives the impression that evolutionary explanations are genetic explanations, and that gives the impression that an evolutionary approach to behavior somehow implies a commitment to genetic explanations for variations in personality and behavior. These are misconceptions. It’s instructive to realize that Darwin knew nothing about genes. The principles of natural selection were discovered by observing variations in phenotypes.

Genes carry the information that is acted on by natural selection, but that action is almost entirely via interactions of phenotypes and environments. Just as the adaptive value of a trait depends on the environment, there is no such thing as an allele that is useful in general; the adaptive significance of a genetic variation depends on its interactions with other genes (epistasis) and interactions with environments.

Sexual Selection

The interests of individuals and their genes are generally the same. An allele that makes healthy long-lived phenotypes is more likely to be passed on. There is, however, an exception. Mutations that decrease health, happiness, longevity, or cooperativeness will spread if they increase the number of offspring.

Sexual selection is the best example. It is the subtype of natural selection in which individuals compete for mates, either by displaying extreme traits that lead to being chosen by the other sex, or by traits that lead to success in competing with others of the same sex. Gigantic colorful tails are a costly drag for peacocks, but the responsible genes spread because peahens prefer peacocks with big tails. The process can set up a positive feedback loop in which the advantages of extreme male traits give advantages to female preferences for those traits because sons with extreme traits will have an advantage. The trait becomes more extreme until the costs it imposes are as large as the reproductive benefits.

The relevance of sexual selection to humans is illustrated by the excess mortality rates experienced by men compared to women. In the early reproductive years in modern societies, as mentioned above, men in the early reproductive years are three times more likely to die than women. Much of the excess is from fights and accidents, but almost all causes of death are greater for males because they allocate effort to competitive ability at the expense of tissue repair. Even early childhood mortality rates are 50% higher for boys than girls.

Sexual selection is a subtype of social selection, the more general category of evolutionary change resulting from social choices and competitions. In sexual selection, the resource at stake is mates, with its obvious connection to reproductive success. However, especially for humans, competition for status, friends, and group memberships also influences reproductive success. Individuals preferred as relationship partners or group members get advantages substantial enough to shape extreme human capacities for altruism and empathy unmatched by any other species. These advantages also help to account for extraordinary social sensitivity and the prevalence of social anxiety in humans.

Many people experience extreme concern about what others think about them. One possible explanation is that other people’s opinions can help or harm reproductive success. There are good evolutionary reasons why many psychiatric disorders are associated with internalizing or externalizing personality extremes. There are good evolutionary reasons to expect constant conflicts between (id) impulses to get short-term advantages and (superego) constraints that sacrifice personal short-term gain for longer-term benefits of being a preferred social partner. Rigidity that emphasizes one over the other is disadvantageous and is recognized as sociopathy or neurosis, respectively. Most people benefit from mechanisms shaped by selection that allow them to shift their social strategies as a function of subtle shifts in social circumstances.

Levels of Selection

Recognition that selection works mainly at the level of the gene and the individual has profound implications for psychiatry. While few evolutionary biologists now support the idea that traits are routinely selected to benefit the group or the species, debates about levels of selection persist, and the idea of group selection remains attractive to many nonbiologists. This may be, in part, because it is so often framed as an explanation for “the mystery of altruism.” Genetic variations that induce costly helping of nonrelatives will become less common in each generation, even if the group benefits. But such tendencies exist, thus posing the mystery. As noted already, kin selection solves the mystery when group members are relatives. But how can we explain examples of cooperation between non-kin? Many examples are “mutualisms” that give benefits to both parties with no opportunity for defection, such as two people moving a heavy rock to get food. Often, however, exchanges are not immediate, leaving the possibility that one partner will fail to reciprocate a favor. Such “reciprocity” exchanges have been subject to exhaustive modeling using the prisoner’s dilemma schema. In this model, the largest payoff goes to an actor who defects when the other person cooperates. If both cooperate, they both get a modest benefit. If both defect, neither gets any benefit.

Because the cells in an individual’s body are all genetically identical, their interests are almost always aligned with the welfare of the individual. The machinations of meiosis ensure that only a single copy of each chromosome gets into a gamete, preventing otherwise destructive genetic competition. Metazoans also have manifold mechanisms to suppress lack of cooperation, such as apoptosis which eliminates cells that start dividing out of control. It is possible to view those mechanisms as shaped to benefit the whole organism, and they certainly do, but the genes that mediate such enforcement persist because of benefits to themselves that result from benefit to the individual.

Some exceptions to genetic cooperation are particularly important for psychiatry. In particular, the interests of maternal and paternal genomes differ. Female mammals invest a proportion of their available resources in a pregnancy, reserving some for future offspring. The male parent may or may not be the parent of those future offspring, so the paternal genome gets an advantage if the fetus extracts a bit more nutrition from the mother. It may seem preposterous that maternal and paternal genomes should compete over relatively small differences in investment in an infant, yet intriguing genetic evidence suggests this is exactly what is happening.

IGF-II alleles tend to make offspring bigger. When they are passed on via an egg, they are imprinted, that is, methylated in a way that inhibits expression of the gene, thus making the baby smaller. IGF-IIr alleles counteract the actions of IGF-II, yielding larger offspring. IGF-IIr alleles are imprinted if they come through the paternal line. The balance between these opposing forces yields babies of normal size, but if imprinting is missing for either gene, offspring will be especially small or large.

It has been suggested by Bernard Crespi and colleagues that this phenomenon may be related to schizophrenia and autism. They can be speculatively viewed as flip sides of the same coin, with autism reflecting unbridled influences of the paternal genome, and schizophrenia reflecting excess influences of the maternal genome. Genes that increase the risk of schizophrenia and autism are imprinted in ways remarkably congruent with this theory. A recent study found, as predicted, that babies lighter than average were especially likely to develop schizophrenia, while those born heavier than average were more likely to develop autism. These findings support but do not confirm the theory, and large-scale genetic studies suggest that the issue is much more complex than can be described here, but Crespi’s suggestion is an illustration of how we might test a creative idea inspired by evolutionary thinking.

A more mundane but nonetheless fascinating example is the weaning conflict. There is a period when mammal mothers try to stop nursing, but infants protest vigorously and try to continue. This conflict is predictable from evolutionary principles. The mother and infant share 50% of their genes, but the infant shares 100% of its genes with itself. A mother that fails to nurse her infant loses a huge investment, so providing whatever her baby needs early in life is in the mother’s genes interests. As the baby grows, it is increasingly able to get nutrition from other sources. At some point, this means that the mother’s reproductive success would best be served by another pregnancy. The baby’s interests, however, are best served by continuing to nurse until the point when the benefits via kin selection to a younger sibling (who shares 50% of genes) are as great as the benefits from continuing to nurse. At this point cessation of nursing is in the child’s interest, so the weaning conflict ceases. An extensive literature in psychoanalysis and child development investigates nursing and weaning, but much remains to be done to put it in an evolutionary context.

Misunderstandings About Natural Selection

Some common misunderstandings about evolution deserve mention. Is evolution over for humans? No. Even if all individuals in a species live for the same number of years, natural selection will change the average characteristics of a species if heritable traits influence numbers of offspring. This does not meet that future traits can be predicted with certainty. Death from drunk driving is now a powerful selection force, but it is hard to say whether it will select for disliking alcohol or better driving while drunk.

Can selection have any impact after menopause? Many people think not, but this is because they have not recognized the role of kin selection. Postmenopausal women have a major influence on the proportion of their genes in future generations when they do things that help their grandchildren. Eric Erickson's generative phase of life has an easy evolutionary explanation. It remains uncertain if menopause is an adaptation shaped by the advantages of taking care of existing children instead of undergoing the risks of reproduction at an advanced age that might jeopardize the welfare of existing children.

Do we all have Stone Age bodies and minds, unchanged for the past 10,000 years? No. Substantial changes have been documented; examples include changes in skin color, ability to digest lactase, and the ability of certain human subpopulations to cope with high altitudes. Increasingly complex social structures may also have imposed strong selection forces that have shaped social and emotional traits and language abilities. Selection for behavioral traits may have sped up since the development of agriculture changed human social structures.

The opposite misconception, that selection changes traits in just a few generations, is also prevalent. However, selection is too slow to adjust our bodies to environments that change drastically over just a few generations. Some people think that nearsightedness, a highly heritable trait, has become more common because the invention of eyeglasses has made it less of a disadvantage. However, a few hundred years is not nearly enough time for disadvantageous alleles to drift to a prevalence that would cause disability in about a third of the population. Myopia is a fine example of a "genetic quirk." The responsible alleles probably have disadvantageous consequences only when they interact with some modern environmental factor, possibly early reading, to cause pathology. One wonders how many alleles that predispose individuals to psychiatric disorders are also not defects, but quirks.

Does an evolutionary view of behavior imply some kind of rigidity or genetic determinism? As noted already above, but deserving repetition, natural selection has shaped learning and innumerable other mechanisms that adapt bodies to changing circumstances. Like learning, culture is not an alternative to an evolutionary explanation. Natural selection has shaped a capacity for culture that unleashed new selection forces that have given humans social capacities unmatched by any other species. Gene-culture coevolution escalates human capacities for culture, which in turn makes a more complex culture possible, in a positive feedback cycle whose end is not yet in sight.

Much difficulty incorporating evolutionary biology as a basic science for psychiatry arises because of perceptions that it is associated with atheism. While belief in God is less common among scientists compared to the general public, the fact that humans have been shaped by natural selection does not directly imply anything about the existence of God. It does contradict some myths that are interpreted literally by some religious traditions, however many religions, including the Catholic Church, now officially recognize that humans are products of natural selection.

BEHAVIORAL ECOLOGY

The synthesis of ethology with evolutionary biology produced the field now known as behavioral ecology. It analyzes how behavior patterns influence fitness. Aside from the core principle that behavior tends to maximize inclusive fitness, behavioral ecology focuses on trade-offs between the different kinds of efforts individuals exert to get different kinds of resources. It routinely distinguishes several categories of effort. Somatic effort is to get calories and other resources to grow, sustain, and protect the body. Reproductive effort is to get mates and take care of offspring. Social effort is to recruit allies and to achieve status in the group.

This framework offers a powerful tool clinicians can use to unpack the sources of stress. Just as physicians conduct a Review of Systems to consider all possible sources of a problem, mental health clinicians need to conduct a Review of Social Systems to consider all possible sources of emotional problems. Most stress comes from the inability to get a secure, sufficient supply of some resource, a threat of loss of a resource, or because efforts to get one resource interfere with efforts to get another. Examples abound in the clinic. One person may be working 70 hours a week, and feeling, correctly, that this makes searching for a romantic partner unrealistic. Another may be stressed because a commitment to full-time childcare makes it hard to allocate effort to anything else. Many people are distressed because of inability to find a mate, infertility, insecurity of a marriage, or wanting to leave a relationship. Many others are unable to find a satisfactory job, worried about losing a job, or feeling stuck in a position that provides little respect.

Other problems arise from unbalanced allocations of effort. For some people money is the overwhelming goal, others are dedicated to being good, getting famous, having friends, or taking care of children. The task of balancing efforts to get these various resources is constant, and imbalances create pathology. Narcissism can be viewed as investing all life effort in the pursuit of self and status to the exclusion of all else. Conversely, some patients invest all effort into trying to please others. Some are preoccupied with their appearance, others with their wealth. Stress often arises from health problems that make it impossible to pursue other goals.

Review of S.O.C.I.A.L Systems

Social status, friends, and groups
Occupation and other valued social roles
Children and family
Income and material resources
Abilities, appearance, health, and other personal resources
Love and sex

To identify the specific source of life stress, one can ask, for each area, what the person has, wants, is trying to do, expects, and the person’s feelings about how things are going in that area. No one ever has everything they want securely and to full measure, but it is nonetheless useful to distinguish at least three groups: patients who have generally adequate sources for most of life’s core resources, patients who lack or have lost something crucial or are facing a threat, and patients those who lack sources of most main life resources.

Other animals face the similar challenges. A grazing white-tailed deer looks up from the grass every few minutes to see if enemies might be present. Each time the animal looks up, it takes time away from eating. The allocation of effort between foraging and defense is regulated according to the level of nutritional supplies and the danger from predators. Risk-taking is also adjusted according to circumstances. Birds given a choice between a feeder that provides one seed every visit, and one that provides six seeds every sixth visit, generally prefer the steady payoff. If, however, the temperature is lowered to a point where the steady payoff provides too few calories to survive through the night, they switch to the risky payoff because that at least offers some chance of survival.

Studies of foraging also have important implications for psychiatry. Food for most animals is distributed in patches, such as different fruit trees. Starting in a new patch often provides many calories per minute, but as the patch is depleted, payoffs slow. How quickly should the animal move to the next patch? It depends on how long it will take to find a new one. Staying too long will waste time, resulting in very few calories per minute. Moving too quickly to a new patch will waste time looking for food when it could be spent gathering food. The optimum strategy is to continue in the current patch until the rate of return declines to that averaged over many patches. This is Charnov marginal value theorem. Animals do not have to do calculus to figure out the optimum moving time; their brains have been shaped by selection to optimize such choices to a remarkable degree. This is relevant for attention disorders and for mood. Consider the enthusiasm experienced upon first beginning to pick berries from a new bush, starting a new job, or engaging in a new relationship. Inevitably, payoffs decline with time. Deciding when to make a change is a difficult decision that often is the focus of psychotherapy.

Reproductive Effort

Brains result in behaviors that maximize inclusive fitness, a term that combines the reproductive success of an individual and their relatives. The simplest rule of thumb is that natural selection shapes brains that result in behavior that tended, in ancestral environments, to maximize the number of grandchildren, nieces, and nephews.

This makes it seem as if everything depends on mating often and with many partners. Differences in reproduction are indeed evolution's engine, but the notion that reproductive success is maximized by maximizing mating success is simplistic. Compared to chimpanzees, humans have a remarkable ability to selectively inhibit sexual impulses. This ability must have given a selective advantage in the context of the mating patterns that developed during human evolution.

Sex and mating behaviors have been a natural focus of behavioral ecologic study. There is no room here to expand on this topic, but it is important to recognize that the very existence of sex poses an evolutionary mystery because parthenogenetic populations grow twice as fast as sexual ones. Also important are the forces that shape many small gametes in one sex (e.g., sperm) and fewer larger ones in the other sex (e.g., ova). This is the essence of the difference between males and females. It means that different mating strategies maximize fitness for males as compared to females. Compared to males, female mammals invest much more in each offspring—many calories and the time and effort of pregnancy—so the total number of possible offspring is quite limited. The investment of a male can be as little as a few sperm and a few minutes, so one man can have hundreds of children, an outcome uncommon in technologic cultures, but reported in cultures where agricultural surpluses made it possible for one man to have many mates. Variation in reproductive success is therefore potentially much larger for men; while some have many offspring, others have none.

However, the human mating system is quite different from that of most other primates. In particular, men make major investments in rearing children. In this respect, humans are more like birds than most other mammals in that rearing offspring takes two parents. Chicks left alone in the nest without protection are likely to quickly be eaten by a predator, so a partnership is required for successfully rearing young. Human babies are so helpless, and their rearing takes so many years, that just having sex was rarely enough to successfully reproduce.

Large investments in child-rearing make sense for a man only if he can be relatively certain that the children are his own. So, tendencies to jealousy and mate-guarding are human universals, although the range of accepted behaviors varies enormously between cultures, from death by stoning for what would elsewhere be dismissed as mere flirtation, to extramarital sex being widely accepted.

These distinctive aspects of the human niche are likely related to another human peculiarity—human ovulation is concealed instead of being advertised, as it is in many other primates. This may have increased paternity certainty by drastically lowering the benefits of forced copulation by a man who is not the woman's mate.

These generalizations are too often pushed to an essentialized version that purports to be THE normal human mating pattern, despite massive evidence for huge cultural variations in mating and family structures. Nonetheless, understanding the evolutionary origins of the human niche can be helpful in understanding marital conflicts, jealousy, and the complications of affairs and divorce.

Life History Traits

Traits such as the duration of gestation, rate of growth, age at maturity, interbirth interval, duration of nursing, and rates of aging are all shaped by selection. One particularly important dimension has been called fast-life versus slow-life strategy. In environments where life is short, reproductive success is maximized by taking risks and reproducing early and often, investing less in each offspring. In more secure environments, it is better to invest more in longer-term strategies and invest more in each offspring. The wide range of human variation on this dimension has been proposed to result from a facultative mechanism that responds to different environments by inducing a faster or slower life history strategy.

Attachment

John Bowlby's recognition of the adaptive value of mother-infant attachment marked the beginning of evolutionary approaches to psychiatry. However, he worked before the major advances that revolutionized the study of behavior, so it is not surprising that he and his followers have tended to see secure attachment as the norm, and other kinds as pathologic. Secure attachment is best for mental health, but important recent work has suggested that anxious or avoidant attachment may give selective advantages in some situations.

EMOTIONS

Emotions are special modes of operation shaped by selection to cope with the challenges that arise in situations that have recurred for thousands of years. Like sweating, cough, and pain, they are facultative adaptations that are useful in certain situations. They evolve in concert with mechanisms that monitor for cues associated with the situations in which they are useful. Like all such responses, emotional responses can be pathologically excessive or deficient. Determining if an emotional response is abnormal requires knowing the situation it has been shaped to cope with, and whether or not that situation is present.

When a patient presents with pain, fever, or cough, clinicians search for what might be arousing the response, such as tissue damage, infection, or foreign material in the lungs. When a patient presents with a negative emotion, the same kind of investigation is needed, but the search is more difficult. Instead of being aroused by tissue pathology or infection, emotions are aroused by less tangible situations. Modern emotions theory emphasizes that the associated stimuli are not just cues; emotions arise from an individual's appraisal of what new information means for their ability to pursue personal goals. A thorough investigation of the context is essential to determine if an emotion is normal and useful, normal but useless, excessive, or deficient.

Excesses of depression and anxiety are by far the most common mental disorders. Almost all effort to understand them has focused on what is different about people who experience intense anxiety or low mood, compared with other people. An evolutionary framework suggests different questions. Why do the capacities for anxiety and low mood exist at all? In what situations are they useful? Why are the mechanisms that regulate them so vulnerable to failure?

The most fundamental question is why emotions exist at all. Controversies about emotions distract from a growing consensus that they are special states shaped by natural selection along with regulation mechanisms that express an emotion in the situation where it gives a selective advantage. No one aspect of emotions is primary. Emotional states shift sensation, perception, cognition, behavior, physiology, motivation, and learning. Subjective experience is just one of many aspects of an emotion.

Controversy about emotion has focused on whether they are best viewed as distinct categories (the basic emotions view), or whether they are best understood as positions on dimensions, such as positive versus negative, and aroused versus calm. Neither perspective is consistent with the evolutionary origins of emotions. Emotions are neither fully separate nor fully dimensional; they are overlapping clusters of changes that have evolved from previous emotions (Fig. 4.3–2).

Some attempts to understand the utility of emotions have tried to map specific functions to specific emotions. However, most emotions serve multiple functions including motivation, communication, and changes in physiology and behavior. Emotions were not shaped to serve specific functions; they were shaped to improve coping with the challenges that arise in a **situation** that has recurred over evolutionary time. Being chased by a predator happened often enough that it reliably evokes a coordinated suite of physiologic behavioral and emotional changes, the fight–flight response. Realizing one has broken a promise arouses guilt. Experiencing generous help arouses gratitude. The key to understanding an emotion is to understand the situation it was shaped for.

One profound implication for psychiatry is that negative emotions are useful. Their aversiveness is one aspect that contributes to their utility. Anxiety, low mood, jealousy, anger, and envy exist because our ancestors with a capacity for these emotions got a selective advantage compared with those who lacked them. Negative emotions seem pathological because they are usually associated with untoward circumstances, and because they are so painful and prone to interfere with important daily tasks. However, they exist only because they have been useful… for our genes, if not for ourselves.

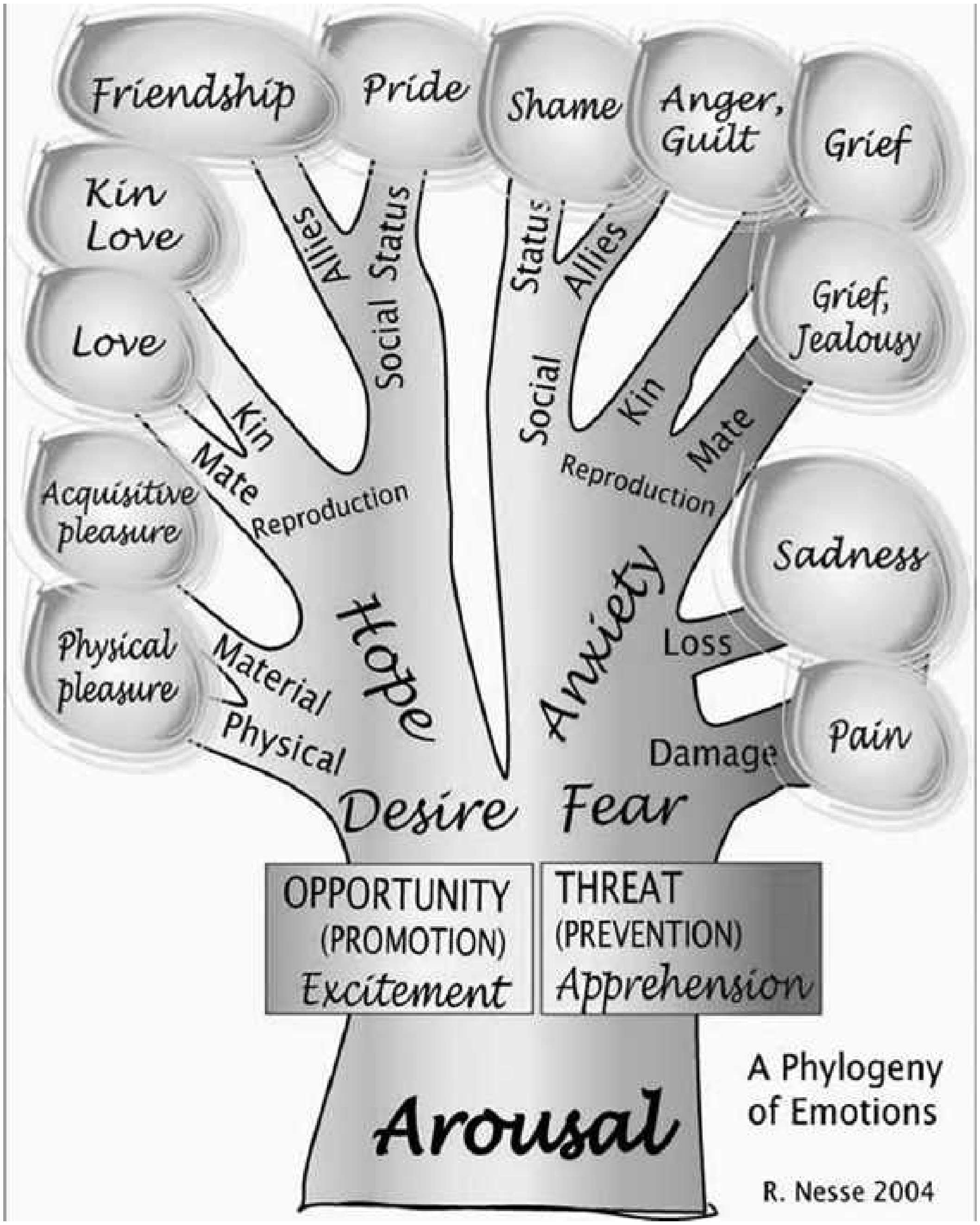
Friendship
Pride
Shame
Anger, Guilt
Grief
Kin Love
Allies
Social Status
Status
Allies
Love
Grief, Jealousy
Kin
Mate
Social
Kin
Mate
Acquisitive pleasure
Reproduction
Reproduction
Sadness
Hope
Anxiety
Physical pleasure
Material
Loss
Physical
Damage
Pain
Desire
Fear
OPPORTUNITY (PROMOTION) Excitement
THREAT (PREVENTION) Apprehension
Arousal
A Phylogeny of Emotions
R. Nesse 2004

FIGURE 4.3–2. A phylogeny of emotions. (Reprinted with permission from The Royal Society. Nesse RM. Natural selection and the elusiveness of happiness. *Phil Trans R Soc Lond.* 2004;359(1449):1333–1347. doi: 10.1098/rstb.2004.1511.)

Fear is useful in the face of threat. It mobilizes energy, cognition, and behavior to escape the danger. Like all facultative adaptations, pathology results from deficiencies as well as excesses. People with apparently excessive anxiety crowd our clinics. People with deficiency of the fear response, hypophobia, cope poorly with dangerous situations, but they rarely come for treatment.

Anxiety is often distinguished from fear because it is aroused by less tangible dangers that are distant or social. It is especially prone to be aroused by situations in which an individual is tempted to do something that might give a short-term advantage at a major social cost. Here too, pathology from excessive anxiety, especially social anxiety, is obvious, while the costs of anxiety deficiencies are far less obvious, even though sometimes more devastating. People with social anxiety tend to seek treatment, those with deficient social sensitivity suffer major social losses.

The value of low mood is much less obvious. The term *low mood* allows a description of states of low motivation, low self-esteem, and pessimism without the terminologic tangles that accompany the word “depression.” In what situation can sadness or low mood be useful? Sadness is aroused by a loss. It helps to recover, replace, or otherwise adjust to a loss. Low mood is aroused when efforts to reach a goal are proving fruitless. That situation has recurred millions of times for individuals in the course of human evolution. While simply persisting with a positive attitude seems intuitively attractive, that is not the best strategy when foraging in winter, trapped in the pursuit of an uninterested partner, or trying for the fifth year in a row to get into medical school. In such situations, it is best to pause, conserve energy, consider other strategies, and, if no route appears viable after all options have been explored, give up and put energies toward a more achievable goal.

Continued pursuit of an unreachable goal can escalate low mood into clinical depression. There are many possible reasons for inability to give up a goal. Few parents are able to give up trying to get a child off of drugs or alcohol. Many individuals have identities so grounded in their careers that they persist despite years of failure. And often, in the face of impending divorce, one partner persists in trying to preserve the relationship, even as demoralization turns into depression.

Other hypotheses about the origins of capacities for depression are also under consideration, many of which refer not to a situation, but to one specific function of depression. One is that depressive rumination pulls effort away from life pursuits so it can be allocated to finding solutions to complex social problems; analyses find little evidence that rumination actually leads to problem solutions. Another is that depression motivates looking inward at what one can do to avoid being expelled from a group.

Another is that depression symptoms give benefits by manipulating others. A particularly influential and plausible proposal is summarized in the phrase “involuntary yielding.” After losing a status competition, persisting in challenging the winner will likely bring more attacks or expulsion from the group. Experiencing oneself as helpless and even lower status than one actually is can prevent such attacks. The relevant situation is pursuing an unreachable goal in the domain of status competition. Subtypes of low mood have been shaped to cope with the different kinds of problems that arise in the pursuit of different kinds of goals. But a general understanding of the evolutionary origins of the capacity for mood variation based on how those variations can help to cope in situations that vary in propitiousness can help to forestall fruitless debates about which function of depression is most important.

Jealousy is aroused by a threat of the loss of a mate or the mate’s fidelity. Responsible for vast interpersonal problems and millions of murders, sexual jealousy is a toxic emotion. However, the situation of a mating relationship being threatened by an outsider has occurred so often over human evolutionary history, with such extreme effects on reproductive success, that it has shaped a specialized pattern of responses. The characteristics of jealousy are often unreasonable in the extreme, but game theory tells us that unpredictability is essential to a successful strategy. Being predictable means being manipulable. Not being fully aware of one’s motives or able to control one’s behavior may well be advantageous. Once again, the benefits may be for one’s genes, not necessarily for oneself.

Positive emotions are equally useful, even if not, unfortunately, so common. In situations characterized by short-term opportunities, it is valuable to experience enthusiasm, risk-taking, and energetic initiative. This suggests the counter-intuitive prediction that short-term opportunities will give rise to more enthusiasm and effort than long-term opportunities. Excesses of positive mood rarely motivate treatment-seeking except when they go to the extremes found in mania.

The resources that influence human reproductive success are overwhelmingly social, so we should expect to see specific emotions shaped to cope with the situations that arise in relationships. The boxes of the prisoner’s dilemma lend themselves well to understanding the origins of some social emotions. The game is named for situations in which the police offer leniency to whichever of two suspects first confesses the details of a crime. The dilemma arises because keeping quiet results in a modest punishment, but each prisoner knows that he will get a higher sentence if the other confesses first in order to get a lesser punishment.

Table 4.3–1.
Emotions Aroused by Situations that Emerge in Reciprocity Relationships

Social Emotions	Other Cooperates	Other Defects
You cooperate	Friendship, trust	Suspicion (before) Anger (after)
You defect	Anxiety (before) Guilt (after)	Rejection, disgust

Repeated cooperation by both parties arouses friendship, loyalty, and trust. If you cooperate when the other defects, anger is the reliable result. It signals that apologies and reparations are required to preserve the relationship. It also may be accompanied by spite, motivations to harm the defector even when the costs are larger for the self. This apparently senseless emotion gives selective advantages by its powerful ability to keep others from defecting when they otherwise would. The game-theoretic best strategy of doing what the other person did on the previous move is called tit-for-tat. It accounts for the runs of cooperation and defection seen in relationships (Table 4.3–1).

RELATIONSHIPS

Views of human relationships as fully explained by kinship and reciprocity are grossly simplistic. A small research industry has grown up to try to explain apparent anomalies of human social behavior, including tendencies to cooperation even without a guarantee of payback, willingness to punish defectors even when that will harm the actor, and eagerness for spiteful revenge that can have no direct payoff.

A key to such anomalies is found in a variation on the core principle. Just as selection shapes bodies to maximize reproductive success, even at a cost to health or longevity, it shapes minds and emotions to maximize reproductive success at the expense of rationality. As Robert Trivers pointed out in his preface to *The Selfish Gene*, the idea that selection shaped us for objective thought is naïve indeed. Often, especially in social situations, commitments beyond justification, and wishes for revenge beyond measure, are more useful.

This has an important implication for psychoanalysis and psychotherapy in general. While self-knowledge is desirable and can be enormously helpful, a lack of conscious access to one’s own motives and emotions may sometimes be adaptive. It has been suggested that the human capacity for psychodynamic defenses that maintain active repression of memories and motives may give selective advantages, not just by making life more bearable but also by influencing others.

The fields of sociobiology and evolutionary psychology study how natural selection shapes social behaviors and resulting group behaviors. There is no room here to explore the full framework. However, when a relationship is at the root of a clinical problem, as is often the case, a formal analysis of how that relationship is or is not working offers a solid starting place. This is very different from trying to understand the personality traits of the participants or the early events that lead to distortions and rigid relationship strategies. Instead, it looks at the history and current state of exchanges and expectations in the relationship.

While this perspective is just beginning to be applied, some clinical phenomena are readily approached from this point of view. For instance, the classic borderline personality pattern of excessive immediate intimacy and adulation, followed by disappointment and distancing, reflects a rigid relationship strategy of intense personal commitment that, when not reciprocated, leads to rejection and anger. The pattern in neurosis is more one of trying always to please others and being angry when they do not live up to expectations.

EXPLAINING VULNERABILITY TO SPECIFIC DISORDERS

As already noted above, every medical disorder needs two kinds of explanation. First, a proximate explanation for what aspect of the body’s mechanism has gone awry and why. Second, an evolutionary explanation for why natural selection has left the body vulnerable to this kind of failure. The six categories from evolutionary medicine help to explain vulnerability to mental disorders, just as they do to other medical disorders.

Disorders of Emotions

Determining whether a particular emotion is normal or abnormal requires understanding the situation in which that emotion is useful, the cues that usually regulate it, and the presence or absence of those cues. Current DSM diagnostic criteria ignore context and attend only to the frequency, severity, and duration of emotional symptoms. This works well at the extremes and increases diagnostic reliability but gives no guidance for cases where it is hard to decide if an emotion is normal or not.

The main evolutionary explanation for negative emotions is that they are useful defenses, with false alarms and excessive expression being common because of the smoke detector principle. Recognizing the utility of negative emotions should not inhibit attempts to relieve them. The rest of medicine routinely provides relief from normal pain, fever, and cough. Psychiatrists should relieve normal anxiety, low mood, and jealousy when that can be done safely. However, good decisions require an understanding of normal function.

> A professional motorcycle racer requested treatment because he was unable to sleep or keep food down the night before a race. The problem started when a friend died in a crash. It became worse after another crash caused another friend to become quadriplegic. He denied symptoms of other disorders and denied any problems with anxiety before his friend’s crashes. He had predictable difficulty recognizing the potentially life-saving nature of his anxiety.

Bad Feelings Are Usually Useless But Normal. It seems intuitively obvious that normal expressions of emotion should almost always be useful, and useless expressions should be products of some abnormality. This has spurred huge searches for neural and psychological abnormalities that cause anxiety and depression. Some instances are indeed products of defective regulation mechanisms. But an evolutionary perspective suggests that useless anxiety and depression usually arise from normal regulation mechanisms.

The smoke detector principle accounts for many instances. Many more arise from novel environments that arouse normal but useless emotions, such as despair from being trapped in a hierarchical bureaucracy or an inescapable marriage. Many bad feelings are harmful to the individual but good for their genes. Examples include sexual desires that cannot be satisfied, jealousy, envy, wishes for revenge, and blind ambitions that cause constant dissatisfaction. Finally, there are simple stochastic variations in an experience that result in normal but useless anxiety and depression. An unlucky sequence of failures sets a pessimistic schema in place that inhibits exploring the full landscape. Threat monitoring mechanisms adjust adaptively to increased sensitivity when repeated exposure to danger indicates they have not been doing their job. But that adaptive feature can easily slip into a positive feedback mode in which every experience of an emotion makes it more likely in the future. Panic disorder is the exemplar in which fear of panic attacks themselves sensitizes the mechanism further. Understanding and describing this vicious cycle for patients can be enormously helpful in itself, and it encourages completing difficult behavior therapy exercises.

Recognition that useless bad feelings often arise from normal mechanisms is a missing foundation for understanding emotional disorders. It helps patients escape from the identity of diseased persons and reframes symptoms as adaptive responses that are out of control. It helps clinicians to identify the root cause of emotional problems.

Schizophrenia

Most of the variation in risk for schizophrenia results from genetic variation. Why didn’t natural selection eliminate the responsible alleles? Many authors have suggested possible advantages to those with the disorder or, more plausibly, to their relatives. However, reproductive success for people with schizophrenia is a fraction of that for other people, and there is no documented advantage for relatives; this hypothesis fails. Others have suggested that it is a disease of modern environments. There are some small differences between cultures, but there is no evidence for dramatically lower rates in nonindustrial societies. New genetic evidence finds that every identified common allele that increases schizophrenia risk has only a tiny effect. Larger effects are mostly due to rare copy-number variations. This pattern is exactly what one would expect if selection were continuously and efficiently purging vulnerability alleles from the genome. Increased rates of schizophrenia associated with older paternal age suggested that de novo mutations might be major factors. They contribute but not that much; most of the association with the father’s age turns out to arise because men with genes that predispose their children to schizophrenia marry and have their first children later than other men.

The big unanswered question is why the cognitive system is so vulnerable to this particular kind of failure. One possibility is that some advantageous cognitive trait has been pushed close to a fitness “cliff edge” beyond which failure is likely, just as strong selection for speed in racehorses has made their lower leg bones long, light, and thin, but vulnerable to breaking. The neural mechanism could involve neuron pruning, cortical folding, or any number of other traits that give major benefits up to some threshold beyond which failure becomes likely.

Substance Abuse

The devastation and early death resulting from substance abuse selects against the alleles that increase vulnerability. However, pure substances and means to administer them have been reliably and readily available for far too few generations for selection to have a big effect. From this perspective, our vulnerability to substance abuse is a product of a mismatch with modern environments. Chemicals mediate the brain’s motivation and learning systems, so it should be no surprise that exogenous chemicals can hijack those systems. The amazing thing is that so many people can use drugs and alcohol without major problems. Active efforts are underway to understand why some people have brain mechanisms that make them especially vulnerable. It would be most interesting to see how people with high and low levels of vulnerability behave in environments where substances of abuse are not readily available. Do they use different strategies when foraging for food? Are their interpersonal relationships different? Answering these questions could lead to new behavioral tests that estimate vulnerability.

Eating Disorders

Anorexia and bulimia have likely become more common in the past century. The predisposing alleles are almost certainly quirks without major effects on fitness until they interact with modern environments. Developmental and personality factors strongly influence who gets an eating disorder. However, the shared human vulnerability seems to result from evolved nutrition regulation mechanisms interacting with novel environmental exposures. Good candidates include the ready availability of every imaginable kind of food, a resulting epidemic of obesity, and media that portray slimmer-than-real bodies. Most serious eating disorders are initiated by a strict diet. The resulting binge eating would be life-saving in a famine, but it causes panic and guilt in dieters and increases resolve to more strictly restrict intake, creating a vicious cycle. Wild fluctuations in calorie intake in ancestral times aroused a mechanism that deposits extra fat stores as insurance against insecure food supplies, so it is unsurprising that dieting often results in weight gain. The knowledge provides some patients with a powerful motive for sensible eating.

Child Abuse

It makes no evolutionary sense for parents to harm their own children. This led evolutionary psychologists to investigate whether the risk of death of a child depends on how it is related to the parents. In some studies, death rates from child abuse are as much as 80 times higher if there is a step-parent in the home. This research was inspired by studies of vervet monkeys that found infanticide to be routine when a new male takes over a harem. Eliminating nursing infants speeds return to estrus in the females, thus increasing the reproductive success of the new male during the period before another male displaces him. The different human mating system means that infanticide in humans gives no such advantage and was not shaped by selection. However, not being with both parents from the start seems to interfere with the bonding that usually protects children from harm.

This research must be viewed very tentatively, as not all data confirms it. In primates more closely related to humans then vervet monkeys, such as chimpanzees, infanticide is very rare. Furthermore, making connections between animal and human data is always complex. And although data on child homicide in, for example, Canada does support this hypothesis, other countries, for example, Sweden, found no difference in child homicide rates between genetic and nongenetically related children.

Extended early contact also seems to be the signal that makes parents and children uninterested in each other as sex partners. For instance, children raised together from their early years in a kibbutz usually experience embarrassment and disgust at the idea of marrying another member of the group. Over evolutionary time, this incest inhibition mechanism has increased reproductive success by reducing the risk of having children with many pairs of identical deleterious recessive genes.

In this example, cultural prohibitions amplify already existing evolved inhibitions. The relevance to psychiatry is all too obvious in cases where such inhibitions are weak because a parent has little early contact with a child. It is interesting to consider if the development of such inhibitions is related to early Oedipal wishes and their subsequent repression.

Sociopathy

The range of prosocial tendencies in humans is breathtaking. Some individuals lie awake at night wondering if some small oversight might have caused others inconvenience, while sociopaths take pleasure in deceiving others and sadistically causing them pain. In this case, selection seems not to have settled on a narrow optimal mean. Frequency-dependent selection has been proposed as a possible explanation.

If people with sociopathic tendencies do well in groups when there are many cooperators to exploit but badly in groups where many others are also sociopathic, this could maintain a small proportion of sociopaths in the population. Their tendency toward sexual license and lack of commitment to families and children could give a reproductive advantage until the proportion of sociopathic individuals becomes high enough that they must deal with each other and until cooperative individuals have developed strong abilities to detect and avoid them. This hypothesis is supported by the high heritability for sociopathy, but it is undermined by evidence for minor neurologic abnormalities in many such individuals and by the role of early abuse and neglect. From a longer-term perspective, the presumed relatively recent evolution of strong prosocial tendencies may turn out to be important. Note that these evolutionary explanations say little about why some people become sociopathic and others do not. They only explain how extreme prosocial traits may give selective advantages and why not all individuals may have such traits.

Personality Disorders

Other personality disorders are also characterized by rigid extremes of social strategies. People with obsessive-compulsive personality traits do their duty to an extreme, expect others to be equally conscientious, and are constantly disappointed. Ordinary neurotics are not so rigid, but they also invest huge proportions of life effort into trying to please others, efforts that often result in disappointment. Those with paranoid tendencies expect that others will deceive or harm them, and their lack of trust means they have difficulty finding relationship partners who can correct their misperception. Those with anxious, avoidant personalities invest vastly excessive proportions of life energy into protecting against possible harm in general. Dependent personality tendencies result in overwhelming investments in particular relationships, and in preserving those relationships. From an evolutionary point of view, personality disorders are not interpreted as adaptations, but as extremes of normal personality dimensions that reflect different social strategies.

FUTURE DIRECTIONS

Evolutionary biology provides psychiatry with a missing solid foundation for understanding behavior, relationships, and emotions. Clinicians who understand the evolutionary origins and functions of emotions have tools that allow them to better understand disorders of emotions. Those who understand the evolutionary origins of capacities for relationships are better able to deal with problematic relationships and the emotions they arouse. Those who understand how natural selection has shaped extreme prosocial tendencies in humans are better able to understand social anxiety and neurosis. Those who recognize personality characteristics as social strategies have a framework for understanding personality disorders. Most globally, clinicians who grasp the distinction between proximate and evolutionary explanations are well prepared to integrate their knowledge of factors increasing risks for individuals with their knowledge about the evolutionary reasons why all members of the species are vulnerable to a particular disorder.

The implications for research are perhaps even more important. Huge efforts to try to identify specific disorders based on specific brain abnormalities will eventually succeed for some disorders, but others will turn out to be, like epilepsy or congestive heart failure, system failures that can have many causes. Severe brain disorders such as autism and schizophrenia may turn out to have a specific, consistent neuropathology, but an evolutionary view suggests that we should be open to the possibility that they result from systems that have been pushed by selection close to some fitness cliff-edge, or that they are otherwise intrinsically vulnerable to failure for other good evolutionary reasons.

Implications for studies of treatment are substantial. Studies of drugs to relieve emotional suffering are often framed as normalizing neurotransmitter abnormalities. However, recognizing that even negative emotions are useful suggests that drugs relieving aversive emotions are like drugs that relieve cough or fever by blocking normal mediating mechanisms. Studies of cognitive-behavioral interventions increasingly go beyond simple learning theory to also consider social factors that influence schemas, and the advantages as well as the disadvantages of distorted thinking. They also increasingly recognize the power of social groups and relationships to get people to carry out behavioral exercises that are difficult, but essential. Interpersonal therapy analyzes social roles and conflicts that almost invariably involve struggles to get important reproductive resources, and trade-offs among conflicting strategies, roles, and relationships.

This chapter provides only the briefest sketch of a basic science whose vast implications for psychiatric research and treatment are just beginning to be explored. Recognizing the opportunities will lead to increased support for research that should result in major advances.

ACKNOWLEDGMENT

Thanks to Adam Hunt for useful suggestions.

Further Readings

*Abed RT, St. John Smith P, eds. *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*. Cambridge University Press; 2022.

*Abrams M. *The New CBT: Clinical Evolutionary Psychology*. Cognella Academic Publishing; 2021.

Alcock J. *Animal Behavior: An Evolutionary Approach.* 10th ed. Sinauer Associates; 2013.

Alcock J. *The Triumph of Sociobiology*. Oxford University Press; 2001.

Baron-Cohen S. *The Maladapted Mind*. Psychology Press; 1997.

Belsky J. Modern evolutionary theory and patterns of attachment. In: Cassidy J, Shaver PR, eds. *Handbook of Attachment: Theory, Research, and Clinical Applications*. Guilford Press; 1999:141–161.

Bjorklund DF, Pellegrini AD. *The Origins of Human Nature: Evolutionary Developmental Psychology*. American Psychological Association; 2002.

*Brüne M. *Textbook of Evolutionary Psychiatry: The Origins of Psychopathology.* 2nd ed. Oxford University Press; 2016.

Brüne M, Belsky J, FaBrega H, et al. The crisis of psychiatry—insights and prospects from evolutionary theory. *World Psychiatry*. 2012;11(1):55–57.

Buss DM. *Evolutionary Psychology: the New Science of the Mind.* 2nd ed. Allyn and Bacon; 2004.

Crespi B, Stead P, Elliot M. Evolution in health and medicine Sackler colloquium: comparative genomics of autism and schizophrenia. *Proc Natl Acad Sci U S A.* 2010; 107(Suppl 1):1736–1741.

Del Giudice M, Ellis BJ. Evolutionary foundations of developmental psychopathology. In: Cicchetti D, ed. *Developmental Psychopathology: Developmental Neuroscience*. Vol. 2. 3rd ed. Wiley; 2016:1–58.

Gilbert P, Bailey KG. *Genes on the Couch: Explorations in Evolutionary Psychotherapy*. Taylor & Francis; 2000.

Gilbert P. Evolution and depression: issues and implications. *Psychol Med.* 2006;36(03):287–297.

Haig D. Transfers and transitions: parent-offspring conflict, genomic imprinting, and the evolution of human life history. *Proc Natl Acad Sci U S A.* 2009;107:1731–1735.

Horwitz AV, Wakefield JC. *The Loss of Sadness: How Psychiatry Transformed Normal Sorrow into Depressive Disorder*. Oxford University Press; 2007.

*Krebs JR, Davies NB. *Behavioural Ecology: An Evolutionary Approach.* 4th ed. Blackwell Science; 1997.

McGuire MT, Troisi A. *Darwinian Psychiatry*. Harvard University Press; 1998.

Nesse RM. Explaining depression: neuroscience is not enough, evolution is essential. *Understanding Depression: A Translational Approach*. 2009:17–36.

*Nesse, RM. Evolutionary psychiatry: Foundations, progress and challenges. *World Psychiatry*. 2023;22(2):177–202.

*Nesse RM. *Good Reasons for Bad Feelings: Insights from the Frontier of Evolutionary Psychiatry*. Dutton; 2019.

*Nesse RM. Ten questions for evolutionary studies of disease vulnerability. *Evol Appl*. 2011;4(2):264–277.

Nesse RM, Bergstrom CT, Ellison PT, et al. Making evolutionary biology a basic science for medicine. *Proc Natl Acad Sci U S A.* 2009;107:1800–1807.

Nesse RM, Berridge KC. Psychoactive drug use in evolutionary perspective. *Science*. 1997;278:63–66.

Nesse RM, Ellsworth PC. Evolution, emotions, and emotional disorders. *Am Psychol.* 2009;64(2):129–139.

Nesse RM, Stein DJ. Towards a genuinely medical model for psychiatric nosology. *BMC Medicine*. 2012;10(1):5.

Nesse RM, Williams GC. *Why We Get Sick – The New Science of Darwinian Medicine*. Times Books; 1994.

Price J, Sloman L, Gardner R, Gilbert P, Rohde P. The social competition hypothesis of depression. *Br J Psychiatry.* 1994;164:309–315.

Stearns SC, Koella JC, eds. *Evolution in Health and Disease.* 2nd ed. Oxford University Press; 2008.

Troisi A. Mental health and well-being: clinical applications of Darwinian psychiatry. *Applied Evol Psychol*. 2011:276–289.

van Dongen J, Boomsma DI. The evolutionary paradox and the missing heritability of schizophrenia. *Am J Med Genet B: Neuropsychiatr Genet*. 2013;162(2):122–136.

Wenegrsat B. *Sociobiological Psychiatry: A New Conceptual Framework*. Lexington; 1990.

5

Quantitative and Experimental Methods in Psychiatry

5.1 Quantitative and Experimental Methods in Psychiatry: Epidemiology

Diana E. Clarke, Ph.D., M.Sc., Lamyaa Yousif, M.D., Ph.D., M.Sc., Nusrat Rahman, Ph.D., Adrienne Grzenda, M.D., Ph.D., M.S., and William E. Narrow, M.D., M.P.H.

BACKGROUND

Epidemiology describes how often disease occurs in populations, the rates at which they change through time, and the factors that explain their occurrence. A field that studies populations as contrasted to individuals, it provides useful information for public health decisions about prevention, treatment, and the social costs of illnesses. This chapter focuses on psychiatric epidemiology, where the illnesses of interest are psychiatric disorders. Because mental disorders more closely follow a chronic pattern, analytical models used in psychiatric epidemiology resemble those used in other chronic illnesses, as opposed to classic acute or infectious disease models.

In 1957, Stuart L. Morris described the uses of epidemiology, which are still relevant today and readily applied to psychiatric epidemiology. These uses of epidemiology included historical study, community diagnosis, health services utilization, completing the clinical picture, identification of syndromes, assessing individual risks, and identifying causes. These seven uses of epidemiology are applied to psychiatric epidemiology below:

Historical study—Findings from epidemiologic studies can be compared through time. A classic example is the finding of a shift in the major causes of mortality that accompanies economic development, from infectious diseases to chronic diseases such as cancer. For psychiatric disorders, such historical comparisons are difficult due to radical changes in diagnostic customs since the 1970s and a relatively recent focus on community diagnosis, as opposed to diagnosis in treated populations such as hospital discharge diagnoses. Research with historical data examines issues such as changes in the incidence or prevalence of disorders over time due to factors such as changing population demographics. For example, as the U.S. “baby boomer” population (those born between 1946 and 1964) continues to age, rates of age-related disorders such as major neurocognitive disorder due to Alzheimer disease are projected to increase for future years compared to current rates.

Community diagnosis—Epidemiologic studies provide a picture of the health status, morbidity, and mortality characteristics of a population and its subgroups. These data permit the identification and characterization of vulnerable populations and their critical health service needs. The burden of an illness on society is estimated in terms of reduced worker productivity, treatment costs, and premature mortality. The World Health Organization Global Burden of Disease project is discussed at the end of this chapter.

Health services utilization—Epidemiologic studies provide valuable information on a population’s treatment needs and its current level of treatment demand. Demand for treatment is defined as the current level of service use in the population. Epidemiologic studies performed since the 1970s have documented the “de facto mental health service system” in the United States as characterized by traditional specialty treatment settings and service use in primary medical care, social services, school, and alternative medicine settings. The primary care setting is an especially important and highly used source of mental health care globally, highlighting the need for workers in these settings to be familiar with accurate psychiatric diagnosis and treatment strategies.

Treatment needs for persons with psychiatric disorders, as opposed to many other medical disorders, have not been well-defined for population-level purposes. For medical disorders such as a bone fracture or sepsis, the assignment of a diagnosis implies an intrinsic treatment need. Psychiatric disorders, on the other hand, are often on a continuum with mood and emotional states characterized by features including duration, frequency, intensity, and appropriateness. The diagnostic threshold distinguishing a typical reaction from a disordered reaction is often unclear and largely dependent on the degree of functional impairment to the individual. Further, psychiatric conditions not reaching full diagnostic criteria may still cause significant distress or disability to an individual. Symptoms may also spontaneously remit without treatment, particularly in mild cases. For these reasons, the need for treatment is an unresolved concept in psychiatric epidemiology. Various combinations of symptoms and symptom severity, levels of distress and disability, and duration of illness are used in current diagnostic criteria. These definitions also carry implications for defining the concepts of treatment and prevention. However, no matter how treatment needs are defined, a large, unmet need for mental health treatment is well documented in the United States and other countries.

Completing the clinical picture—Community-based epidemiologic studies provide data on the natural history of a disorder, including risk and protective factors, prodromal states, the onset of illness, and disease progression. The large number of persons with psychiatric disorders who need treatment but do not receive it has implications for the study of these disorders. Persons who seek treatment are likely to have more severe cases of disorder, a longer duration of the disorder, and more comorbid conditions. Sociodemographic and socioeconomic factors also influence which cases of the disorder are seen for treatment. Hence, samples drawn from clinical settings are different from those drawn from community settings and may thus give biased information about the nature of a disorder. A well-known example of this difference is the Berkson bias, which describes erroneous conclusions about risk factors that may be drawn about a disorder simply due to higher rates of comorbidity in hospitalized patients compared to nonhospitalized patients.

Identification of syndromes—Psychiatric disorders lacked well-defined diagnostic categories until the publication of the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) in 1980. The dearth of objective validating tests for the disorders listed in DSM, across editions, means that included disorders are only approximations of underlying disease states, and the diagnostic criteria represent the best available hypotheses for detecting these underlying diseases. Examination of the diagnostic criteria in clinical and research settings since the release of DSM-III suggests that the categorical, symptom-based approach to diagnosis has significant limitations, both in its ability to describe the clinical complexity of patients and to facilitate the discovery of the underlying causes of mental illnesses. Therefore, efforts were made in the development of DSM-5 to incorporate cross-cutting dimensional symptom assessments as supplements to the categorical diagnostic criteria. The Research Domain Criteria (RDoC) project, sponsored by the National Institute of Mental Health, examines the neuroscience of basic behavioral domains with the goal of facilitating etiologic research and ultimately improving diagnosis precision and reducing time to treatment.

Epidemiologic research can support the scientific investigation of syndromes by assessing their prevalence in the community, associated disabilities, and their relationship to other disorders. Syndromes that are highly prevalent in the community, with little associated disability, are more likely to represent expected variations of human behavior than disorders. A syndrome that is highly associated with another disorder is likely to be a variant or subtype of that disorder. Much attention has been drawn to the relationship between depression, anxiety, and somatization, which are highly comorbid in the community and treatment settings.

Assessing individual risks—The results of epidemiologic studies are important in assessing the likelihood that an individual will develop a disease. A well-known example is that individuals who have high cholesterol levels, hypertension, and smoke exhibit significantly increased cardiovascular disease risk. Such findings are important for treatment and preventive interventions at the individual level, as exemplified by widespread acceptance of efforts to promote a healthy diet, smoking cessation, and treatment of hypertension. Epidemiologic findings on risk also figure into actuarial determinations for insurance coverage purposes and decisions on the marketing of pharmaceuticals to certain population groups. Risk and protective factors for psychiatric disorders are ever-evolving, but the translation to individual-level preventive interventions is still at a nascent stage, despite awareness that early treatment is the key to mitigating lasting adverse consequences.

Identifying causes—One of the ultimate goals of epidemiologic investigations is to elucidate the causes of disorders. Once the cause of a disorder is found, targeted treatments and preventive interventions can be developed to help reduce the burdens of the disorder in the community. For complex conditions, such as psychiatric disorders, causes are often multifactorial and interdependent, involving genetic and environmental factors. The mechanisms by which all of these factors interact to influence disease presentation or its progression are generally not well- elucidated. The relative impact of genetic versus environmental factors varies by disorder. One of the best-known examples of a specific gene-environment interaction in a psychiatric disorder involves a polymorphism of the promoter region of the serotonin transporter gene and stressful life events in the development of major depression as an adult. However, more research is required before these findings are actionable in the treatment and prevention of major depressive disorder.

MEASURES OF DISEASE FREQUENCY

Prevalence

A useful measure to quantify the occurrence of a specific condition in a population is the prevalence estimate, a ratio describing the number of people with a disorder in a specific population in a designated time period. The three elements important to prevalence estimation are case definition, target population, and time.

$$\text{Prevalence rate} = \frac{\text{number of cases identified in a specified period of time}}{\text{number of people in the population at the same time}}$$

Commonly used prevalence rates include:

Current (Point and 1-Month) Prevalence. *Current (Point and 1-Month) Prevalence* describes people with a disorder at a specified point in time and is referred to as the point prevalence. While the point prevalence is the best estimate of the “current” prevalence of a disorder, it is difficult to estimate for mental disorders in the general population since mental disorder diagnosis requires multiple symptoms to cluster together for specific time periods, and these symptoms may wax and wane on a daily basis. Therefore the 1-month prevalence rate has been more frequently used to describe the percentage of people who currently have a mental disorder. This time period describes people in a specified population whose symptoms have met the criteria for a disorder at any point in the past month.

1-Year Prevalence. *1-Year Prevalence* describes people who have had a disorder at any time in the past year. Because service planning is done on an annual basis, planners are most interested in how many people can be expected to have a disorder in a given year. If an assumption is made that mental disorder prevalence rates do not change greatly from year to year, and there have been no significant demographic changes in the population of interest, a 1-year prevalence rate from a recent time can be multiplied by an estimate of the total population for a given year to determine how many people in the population can be expected to have the disorder in that year.

Lifetime Prevalence. *Lifetime Prevalence* describes people who have had a disorder at any point in their lives. An accurate estimate of lifetime prevalence can be a useful tool to describe the risk of developing a disorder in a population. Lifetime prevalence is also valuable for risk factor studies. In studies of risk and protective factors, it is important to determine, as accurately as possible, who has had a disorder and who has not. The relapsing and remitting course of many mental disorders, with potentially years between episodes, makes 1-month and 1-year assessments less valuable for this purpose since true cases are likely to be counted as noncases. Lifetime prevalence is not helpful for service planning, as it will count as cases of single episodes of disorders that occurred in the remote past and have little chance of recurring in the next year. Lifetime prevalence rates are also more subject to recall bias than are prevalence estimates based on shorter time frames; that is, remotely occurring episodes of the disorder are less likely to be accurately recalled than those occurring more recently, particularly if complex diagnostic criteria are used (e.g., multiple symptoms, age of onset, duration, and disability requirements). Lifetime prevalence estimates are generally lower in older age groups, raising the question of whether older generations were more protected from developing mental disorders than younger generations. The role of recall bias cannot be discounted in this deliberation: compared to younger respondents, older respondents have lived more years in which they could have developed a mental disorder, but also, on average, have had more years elapse between their episode of illness and the present time.

Caution is needed when interpreting prevalence rate differences over time. A shift in the distribution of a demographic factor in the population can cause such observed differences. For example, when migration increases the proportion of young people in a population, disorders more frequent in the older adults will have lower prevalence rates in the overall population, even though the prevalence remains stable in the older adult population itself. For this reason, statistical adjustments are made to account for such differences in the two populations being compared.

Incidence

At times, epidemiologists require estimates of how many new events occurred in a period of time, for example, the number of new cases of posttraumatic stress disorder since a natural disaster. Incidence rates are the measure of choice in such instances. The numerator of the incidence rate is the number of new events arising in a specified time period; the denominator is the population at risk of having an event during that time period.

$$\text{Incidence rate} = \frac{\begin{array}{c}\text{number of new cases of disorder}\\ \text{in a specified period of time}\end{array}}{\begin{array}{c}\text{people at risk for the disorder in the}\\ \text{same time period (person-time)}\end{array}}$$

People who already have the disorder at the start of the time period are, by definition, taken out of the denominator for they are not at risk for developing the disorder. A new case is a person who did not have the disorder at the start of the specified time period and developed it afterward. Persons without the disorder at the start of the specified time period may or may not have had the disorder previously. When the new cases are limited to people who have never before had the disorder in question, the measure is referred to as the first incidence. Because most mental disorders are relatively rare in the population, the occurrence of new cases is also relatively rare, and the occurrence of first-incident cases is even more rare. This makes estimation of incidence rates difficult, requiring large populations-at-risk and relatively long periods of observation to accrue new cases. Most incidence rates are based on a 1-year time period. Longer periods of observation are also potentially subject to recall biases.

Measurement of the population at risk over a period of time involves several considerations. In reality, populations are often nonstationary; that is, the population at risk at the beginning of the study changes during the period of observation due to people entering the population after the study period started, dropping out of the study, moving without leaving new contact information, or dying. These individuals will have been observed for different time periods, but the data they contributed is included in the denominator as person-time at risk, an estimate of the actual time at risk that all persons contributed to a study. It is estimated as the product of the number of people at risk multiplied by the time elapsed until they were no longer at risk (became a case) or stopped being followed.

Relationship Between Prevalence and Incidence

Prevalence, incidence, and mean duration are mathematically related. The mean duration of the disorder is an indicator of the chronicity of the disorder. It is estimated as the average of the difference between the age of onset and age at the last time when criteria for the diagnosis was met. Duration is affected not only by the intrinsic characteristics of the disorder but also by external factors such as receipt of treatment. The prevalence of a disorder is proportional to incidence times duration ($P \sim I \times d$). This equation illustrates another reason why caution should be used in the interpretation of prevalence rates, as prevalence rates reflect not only increases in the number of new cases (incidence) but also the length of time that cases have the disorder (duration). When a treatment is found to cure a previously incurable condition, the prevalence of the condition can be expected to decrease. However, a treatment that extends the life of a chronic, incurable condition might actually raise its prevalence rate through the increased duration of the condition. For example, the prevalence of AIDS in the population increased after an effective treatment was found that prolonged the lives of affected patients and reduced mortality. In such cases, it is wise to employ incidence measures; relying solely on prevalence estimates makes it impossible to tell apart an increase in new cases from an improvement in survival.

Comorbidity

In psychiatry, co-occurrence of disorders, meaning individuals diagnosed with more than one mental disorder, is frequently observed in both community and clinical populations. Random co-occurrence of disorders is possible whenever the presence of one disorder does not preclude the presence of the other disorder. For common conditions, such random co-occurrences may be quite common. By definition, random co-occurrence does not imply underlying relationships between the co-occurring disorders.

The most commonly used concept of comorbidity refers to a nonrandom co-occurrence of disorders, in which the prevalence of condition A is higher in people with condition B than the prevalence of condition A in persons without condition B. This type of comorbidity has implications for the underlying processes and boundaries of the two disorders. For example, the comorbid disorders may share common risk factors or may not be separate disorders at all, but rather different presentations of the same disorder. At times, the apparent co-occurrence can be explained by artifacts of the criteria used to make the diagnoses. For example, high rates of co-occurring substance use disorders were observed in persons with DSM-III antisocial personality disorder. However, substance use was one of the symptoms that could be used to meet diagnostic criteria for both antisocial personality disorder and substance use disorder in DSM-III. Sampling bias may also create false comorbidity; for example, comorbidity is higher in treated samples than community samples, but this relationship is not related to shared etiologies.

MEASURES OF EFFECT AND ASSOCIATION

Risk, Protection, and Causation

Epidemiologic studies often aim at estimating the effects of putative risk and protective factors on the occurrence of disorders in the population. A risk factor increases the likelihood that a person possessing that factor will develop a specified outcome (such as a mental disorder) compared to a person who does not possess the factor. A protective factor decreases the likelihood that a person possessing that factor will develop the outcome compared to a person who does not possess the factor. A plausible model accounting for biologic and psychosocial risk and protective factors is a prerequisite for the interpretation of effect measures. Quantifying these effects relies on accurate identification and differentiation of cases and noncases, unambiguous definitions of the risk and protective factors in question, and accurate identification of the exposure of each case and noncase to the factor. The findings of risk or protection should not be viewed as synonymous with causation, although they may provide crucial clues in the search for causes. For example, female sex is an accepted risk factor for the development of major depressive disorder. In this case, one would not cite female sex as a cause of depression. Rather it is a measurable characteristic with a wide variety of underlying biologic and associated psychosocial factors, each of which may play a more direct causal role.

In psychiatric epidemiology, exposures can be genetic or environmental. For example, being a victim of a natural disaster is an environmental exposure and being the offspring of a depressed mother is an exposure that likely has mixed genetic and environmental aspects, even though the genetic exposure is unlikely to be precisely identified. The role of specific genetic exposures in the development of psychiatric disorders is likely to be studied with increasing frequency in the future as candidate genes for disorders are identified, and genomic science advances. Without a plausible model accounting for biologic and environmental factors, observed associations cannot be justified as contributing to causation.

The factors involved in the development of chronic diseases are numerous, complex, and intertwining, and the process by which they operate has been described as a *web of causation*. In this web of causation, causal pathways involving the main effects of interest might be modified by other, noncausal factors. These indirect modifying effects should be considered and taken into account in the initial explanatory model, to the extent possible, since the effects of unknown and unmeasured factors are always a threat to the validity of a risk factor study. The fields of epidemiology and biostatistics have developed study designs, sampling strategies, and analytic techniques to help overcome some of these difficulties in the estimation of effects, although these strategies cannot overcome the disadvantage of a poorly specified model.

Studies of causation are logistically difficult to implement and maintain and tend to be expensive. For many disorders, including psychiatric disorders, the occurrence of a new case in a population is a rare event, requiring the monitoring of large populations in order to gain a sufficient number of new cases. Also, the time span between the exposure to risk and protective factors and the onset of a disorder may be long, requiring periodic monitoring of the population to prevent drop-outs (“loss to follow-up”) and detect new exposures among the previously unexposed.

Table 5.1–1.
2 × 2 Table for Calculation of Relative Risk

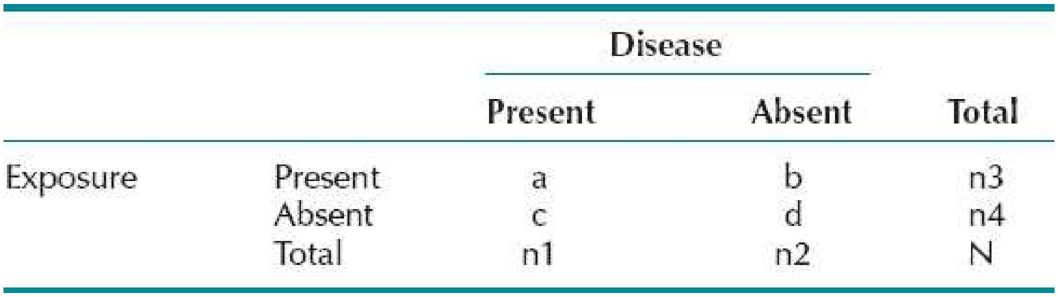

		Disease		
		Present	Absent	Total
Exposure	Present	a	b	n3
	Absent	c	d	n4
	Total	n1	n2	N

Relative Risk

The classic relative effect quantification is *relative risk*, the ratio of two incidence rates. Table 5.1–1 lays out the distribution of exposed and unexposed persons who will go on to develop or not develop the disease of interest. Cell a represents persons with the exposure who go on to develop the illness, while cell b represents exposed individuals who have not developed the disease. Cell c represents unexposed persons who nonetheless have developed the disease, while cell d represents unexposed persons who have not developed the disease. The risk of developing the disease for a person with the exposure is a/a + b. The risk of developing the disease in persons who were not exposed is c/c + d. Therefore, the relative risk of developing the disease for those with compared to those without the exposure is:

$$\frac{(a/a + b)}{(c/c + d)} \text{ or } \frac{\text{incidence}_{\text{exposed}}}{\text{incidence}_{\text{unexposed}}}$$

A relative risk greater than 1 suggests that the exposure is a risk factor, and a relative risk less than 1 indicates that the exposure is a protective factor. A relative risk of 1 suggests that exposure has no effect on the onset of the disorder.

Odds Ratios as Approximations of Relative Risk

For the reasons mentioned above, studies using incident cases are not always feasible. Still, it is possible under some circumstances to design studies where prevalent cases are sampled, and risk and protective factor data are then collected (see Analytic Studies, below). In such cases, the incidence is not directly measured, and risk ratios cannot be calculated. However, with carefully chosen cases (disease present) and controls (disease absent), and accurate measurement of exposures, an approximation to the relative risk can be made by using *odds ratio.*

Table 5.1–1 can also be used to represent the different assignments that can be made in a case-control study. Cases with the exposure present are represented in cell a; cases without the exposure are in cell c. Controls with the exposure are in cell b, and controls without the exposure are in cell d. In this table, the odds of having the disease among those with the exposure is a/b. The odds of having the disease among those without the exposure is c/d. The odds ratio compares the odds of having the disease and the exposure to the odds of having the disease without the exposure.

$$\text{Odds ratio} = \frac{a/b}{c/d}$$

Simple algebraic manipulation of this formula allows a convenient way to calculate the odds ratio, the crossproduct ratio. Care must be used when calculating an odds ratio using this method to make sure that the rows and columns of the table are set up as in Table 5.1–1.

$$\text{Odds ratio (cross-product ratio)} = \frac{ad}{bc}$$

As with the relative risk, odds ratios vary around 1. An odds ratio greater than 1 indicates an association between the exposure and the presence of the disease, and an odds ratio less than 1 indicates an association between the exposure and not having the disease. An odds ratio of 1 suggests that there is no association between exposure and having or not having the disease.

Confounding. *Confounding* occurs when a third factor (A), other than the exposure of interest, is independently associated with the outcome (O) as well as with the exposure of interest (E) and is not in the causal pathway.

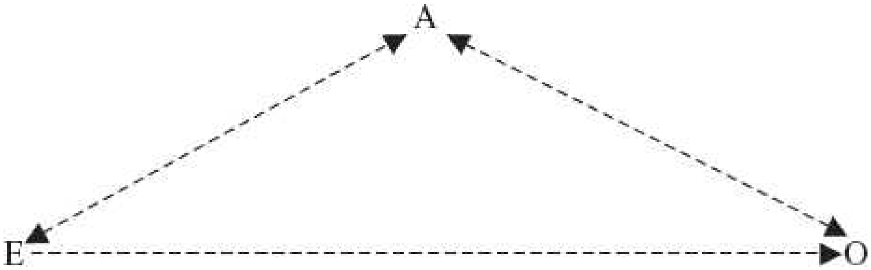

For example, studies have found that smoking is associated with suicidal behaviors. However, alcohol use is associated with smoking in individuals with and without suicidal behaviors and also is independently associated with suicidal behaviors regardless of smoking status. Therefore, alcohol use is considered a confounder of the relationship between smoking and suicidal behaviors. In order to interpret the effect of smoking on the risk of suicidal behaviors, adjustments must be made for alcohol use.

Confounding can distort the relationship between exposure and outcome in any direction, either reducing or augmenting the size of the effect. Therefore, when measuring the effect of an exposure, other factors that might explain the outcome should be measured. When potential confounders have been measured, analytic and statistical modeling techniques such as stratification and regression analysis can help disentangle the “true” contribution of the exposure from that of the confounding factors. When such techniques are used, the results of the analysis are referred to as “adjusted” or “controlled” for the confounding factors (e.g., an adjusted odds ratio for the effect of smoking on suicidal behavior controlling for alcohol use). Unmeasured confounders, by definition, cannot be accounted for by these techniques, and their potential presence should always be considered when interpreting the results of an analysis.

Mediation. Mediation occurs when a third factor (A) in the causal pathway explains the mechanism through which the exposure (E) leads to the outcome (O). The mediator variable (A) must be independently related to, and occur temporally after, the exposure variable (E).

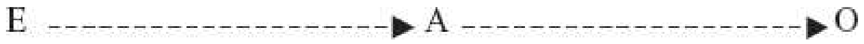

For example, cognitive therapy can reduce depression by reducing anxiety. In this example, anxiety is a mediator of the relationship between cognitive therapy and depression; it provides a mechanism by which cognitive therapy and depression are related. Mediation can be partial, where the third factor only explains some but not all of the relationship between the exposure and outcome, or complete, where the mediator explains all of the relationship between the exposure and the outcome. In mediation analysis, the relationship between exposure and outcome is broken down into direct (i.e., E directly affects O) and indirect (i.e., E affects O only through A) components.

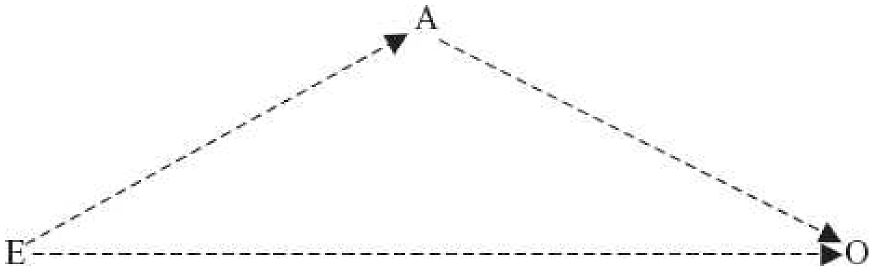

Effect Modification/Moderation. *Moderation* occurs when a third factor (A), at different levels (i.e., A_1 vs. A_2), impacts the relationship between the exposure and outcome differently. Furthermore, for the third factor (A) to be a moderator, it must temporally precede and be independent of the exposure. For example, the relationship between antidepressant use and depression recovery is found to be different for individuals who are adherent compared to those who are nonadherent. In this example, the level of adherence moderates the relationship between antidepressant use and depression recovery. Typically, stratified analyses (i.e., stratification based on different levels of A) or inclusion of an interaction term between the third factor (A) and the exposure is used to identify *effect modification*.

Attributable Fraction. The *attributable fraction* is used to study the effect of a risk factor on the incidence of a disease. Attributable fractions can be calculated for those exposed to a risk factor or for an entire population and are also referred to as etiologic fractions or *population-attributable risks*. The attributable fraction for those exposed to a risk factor shows how much excess disease can be attributed to the risk factor among those exposed. It is calculated by dividing the excess risk for the risk factor by the incidence rate in the exposed group.

$$\text{Attributable fraction for exposed group} = \frac{\text{incidence}_{\text{exposed}} - \text{incidence}_{\text{unexposed}}}{\text{incidence}_{\text{exposed}}}$$

The population-attributable fraction tells us how much excess disease in a population can be attributed to a risk factor. It is calculated as follows:

$$\text{Population attributable fraction} = \frac{\text{incidence}_{\text{total}} - \text{incidence}_{\text{unexposed}}}{\text{incidence}_{\text{total}}}$$

If the risk factor is related to the disease in a causal way, then attributable fractions can be used to determine the amount of disease reduction that would occur if the risk factor was prevented. The prevalence of exposure to the risk factor will affect its population-attributable fraction. A rare exposure, no matter how strongly related to the disease, will result in a small population attributable fraction.

EPIDEMIOLOGIC STUDY DESIGNS

The choice of the design for an epidemiologic study depends on several factors. First and foremost, the study's design must be appropriate to the research question that the study is addressing; not all types of studies can address all types of questions. At the most basic level, study designs can be divided into *experimental* and *observational*. In an experimental study, the researcher can control the exposure, which is usually a treatment or preventive intervention. In nonexperimental, or observational, studies, the researcher is only able to observe and quantify the effect of the exposure, which cannot be manipulated as in an experimental study. Observational studies can be further characterized as descriptive or analytic.

Experimental Studies

In standard clinical trials, the exposure is a treatment (usually medication, but psychotherapies have been increasingly subjected to clinical trial designs), with the goal being to assess the *efficacy* of the treatment, that is, whether the treatment produces the desired outcome for the condition it is supposed to treat. Desired outcomes may include cure or remission of disorders, decrease in symptom levels, prevention of disability, or increased survival time. Efficacy trials tend to use highly selected subjects possessing the disorder of interest, with as few potential confounding factors as possible. The strict inclusion and exclusion criteria for enrollment into an efficacy trial allow for a purer experimental milieu, but they do not reflect clinical reality. Over the past two decades, it has been recognized that treatment outcomes in routine clinical practice do not always mirror the results of efficacy trials and that clinical trials should also reflect the reality of routine practice. In routine practice settings, as opposed to efficacy trials, treatments may be changed or discontinued because of nonresponse, side effects, or patient preference, and patients may have comorbid mental or general medical disorders, take multiple medications, use substances, and not follow treatment instructions to the letter. This recognition has led to the fielding of clinical *effectiveness* trials in which a wider range of subjects is included in an effort to provide a clearer picture of how a treatment will work in clinical practice.

The standard clinical trial is controlled and randomized. Control groups receive either an inactive treatment (placebo) or a treatment already proven to be efficacious for the disorder. Control groups are necessary for comparison purposes and to account for effects that may occur regardless of the effects of experimental treatment. For example, there are often salutary effects of receiving care, even if the patient is receiving an inactive treatment; clinical improvement in the face of an inactive treatment is referred to as a placebo effect. Similarly, patients receiving inactive treatment may report side effects. The use of inactive comparison treatments is now less frequent than in previous years. Research experience has shown that there can be deleterious, long-lasting clinical consequences when psychiatric patients with a treatable disorder are not actively treated, particularly among those with severe symptoms. This raises obvious ethical issues around the use of placebos.

Random assignment of subjects to either the experimental or control condition helps to prevent bias in the outcome of the experiment by maximizing the chance that factors related to the outcome are evenly distributed between the groups. For example, subjects with more severe symptoms may be less likely to respond to treatment; preferential assignment of them to the experimental group may bias the study results in favor of the control treatment, and conversely, preferential assignment to the control group might produce a bias in favor of the experimental treatment. Another method used to increase comparability between groups is to *double-blind* the study so that neither the study staff nor the subjects themselves know to which group the assignments have been made.

The main difference between clinical trials and *prevention trials* is that subjects in prevention trials have not developed the disorder of interest. Subjects for a prevention trial can be selected from the community or from any other facility that is not a treatment center for the disorder of interest. Prevention trials are usually more costly than clinical trials, mainly because sample sizes required for these trials must be adequate to observe a sufficient number of outcomes, usually defined as incident cases of a disorder. Well-thought-out sampling strategies can help to maximize the number of subjects at risk for the study outcome while still maintaining the ability to generalize the results to the population studied. Additionally, because the period of risk for developing a disorder can be lengthy, especially compared to a circumscribed treatment period, prevention trials can take longer to complete, requiring more extensive tracking of study subjects than is needed for clinical trials. Although all guidelines for subject assignment in clinical trials apply to prevention trials, at times, randomization is not feasible in larger studies. For example, studies of the so-called *universal interventions* are applied to all members of the general population and are not feasibly randomized.

Observational Studies

There are two primary nonexperimental, observational studies in epidemiology, analytic and descriptive. Analytic studies, in which the effect of an exposure, uncontrolled by the investigator, is studied relative to a specific outcome, can be subdivided into cohort and case-control studies.

Analytic Studies. A *cohort study* is similar in concept to a clinical trial. In these studies, two cohorts of individuals without the outcome of interest are randomly chosen from a defined population, one cohort with an exposure and one without an exposure. These cohorts are then followed over time, and the rate at which the subjects in each cohort develop the outcome of interest is measured. By comparing the rates of the outcome in the two groups, the effect of the exposure can be estimated. Cohort studies have many advantages inherent in their design: onset of the outcome can be observed in real-time (as opposed to relying on recollection), and potentially confounding variables can be measured before the onset of the outcome. However, because incident cases are being measured, and the base population prevalence of these outcomes is often quite low, the sizes of the cohorts may need to be large, increasing the cost and complexity of the study. Cohort studies are also dependent on accurate measurement of exposure: sometimes such exposures are reliably obtained through administrative records (e.g., employees working in a particular industry or in the military; individuals living in a geographic area at the time of an exposure), at other times measurement of exposure must rely on the subjects' recollections, which may be inaccurate.

In contrast to cohort studies in which selection is based on exposure, with subsequent measurement of the outcome, in *case-control studies*, selection is based first on having the outcome, with subsequent measurement of past exposure. Cases with the outcome of interest are selected randomly from a population, and a control group without the outcome is randomly sampled from the same population that gave rise to the cases. Because the control group is used to estimate the distribution of the unexposed and exposed individuals in the population, it is important that the controls are sampled from the source population independently of their exposure status. Frequently, given the low prevalence of many disorders and the size of the original population, all cases are chosen. Selecting more controls than cases can increase the statistical power of the study. The rates of exposure in the case group and the control group are compared to give an estimate of the effect of the exposure on the development of the disorder. Case-control studies have the advantage of being easier to carry out than cohort studies.

Descriptive Studies. *Descriptive studies*, as their name implies, are aimed at describing the health status of a population. Descriptive studies typically include studies documenting the prevalence and incidence of disorders, the characteristics of the people who have or develop disorders (e.g., age, sex, race, education, income), and the use of health services in a population. They can be cross-sectional (study of a population sampled at a given point in time), repeated cross-sectional (studies of a population sampled and then re-sampled at several points in time), or longitudinal (studies of a population sampled once and then followed over time). Descriptive studies provide the basic data on which all other types of epidemiologic investigations are built, as well as serve as valuable resources for service planners and policymakers. A primarily descriptive focus of an epidemiologic study does not preclude analytic aims. The representative sampling of a large descriptive study can be a valuable asset for the collection of representative data for analytic studies. For example, collection of biologic samples for genetic analysis could be done in a large community sample, provided that the analytic goals are well-specified, the technology for collection and analysis is available, and the study subjects provide informed consent for the use of these samples.

SAMPLING AND SAMPLING STRATEGIES

Sampling refers to the process of selecting a subset (i.e., sample) of the population of interest for a research study (e.g., selecting a subset of college students from among all college students in the United States). The goal is to select a sample that is representative of the population of interest, which is determined by the specific research question. In order to select a representative sample of the population of interest, one needs to have an exhaustive list of the members in the population, called the sampling frame, from which the sample will be drawn. Ideally, the sampling frame should be the list of the entire population of interest—as may be the case for a population of college students—but this may not be the case depending on the size of the population of interest. For instance, the population may be too large to have a list of every member (e.g., the U.S. population). In such an instance, the researcher may rely on a list that approximates a representative sample of the population of interest.

There are two main methods of sampling: *probability* and *nonprobability sampling*. Probability sampling occurs when the sampling frame for the population of interest is known, and each unit/member within the population has a known chance, or probability, of being selected. When a random sample of the population of interest is selected, the chance of missing entire sections of the population is minimized, and the size and effect of sampling error are calculable. The results from a study that uses probability sampling can be generalized to the entire population if the study has an adequate response rate.

Two main types of probability sampling are simple and stratified random sampling. In simple random sampling, each unit in the population, as determined by the sampling frame, has an equal chance of being selected (e.g., randomly selecting 500 college students from the list of all college students in a particular state). However, if there is variability across subsamples of the population with respect to the variable of interest, simple random sampling cannot ensure that all subsamples will be represented within the selected study sample. For instance, if the variable of interest is expected to vary for males and females, and there are five times more female college students than males in the state, simple random sampling might not provide enough males for the study. When there is expected variability across subsamples of the population with respect to the variable of interest, stratified sampling is the preferred method of choice. In stratified random sampling, the sampling frame is divided into a number of nonoverlapping strata based on a factor that may affect the variable of interest (such as sex in the example above), and individuals are randomly selected from within each stratum (i.e., selecting a random sample of 250 males from among all male college students and a random sample of 250 females from among all female college students within the sampling frame). Stratified random sampling can ensure that representative samples of all relevant subsamples of the population are selected (in this example, it would ensure enough males were selected for the study of college students). Stratified random sampling provides greater statistical precision because there is less variability within a stratum compared to the variability when dealing with the entire population, as in simple random sampling.

Nonprobability sampling is used when the sampling frame is not available. With nonprobability sampling, information on entire sections of the population may be missing, which affects the ability to estimate the size and effect of the sampling error. Results from a study that uses nonprobability sampling cannot be generalized to the entire population of interest. In nonprobability sampling, subjects are selected on the basis of their accessibility or by the purposive personal judgment of the researcher. Examples of nonprobability sampling include *convenience*, *consecutive*, *quota*, and *snowball sampling*. Convenience sampling refers to the collection of sample units that are easiest to access until the desired sample size is reached (e.g., a volunteer sample). Like convenience sampling, consecutive sampling is not based on random sampling. However, in consecutive sampling, efforts are made to include all accessible sampling units typically within a specified time period. An example of consecutive sampling would be the recruitment of all patients who present to a mental health clinic from January 1 to June 30 of a specified year. Quota sampling involves the nonrandom selection of sample units based on some fixed quota that is related to one or more specific sample characteristics. Sampling is done until the quota is met, for example, adding new clinic admissions with a diagnosis of bipolar disorder to the study sample until a specified sample size is met. Snowball sampling is the method of choice when trying to sample hidden or hard-to-reach populations, such as undocumented immigrants or homeless individuals. Under such conditions, researchers may identify one or more individuals from their populations of interest and invite them to participate. These individuals are then asked to identify and recruit additional individuals until the desired sample size is attained. For example, undocumented immigrants may notify other undocumented persons to participate.

MEASUREMENT

Measurement instruments are developed in order to identify disorders and their associated characteristics in community and clinical populations. When considering a measure for use, either in a research or clinical setting, the target population on which the instrument was tested should match the population in which the instrument will be used. An instrument cannot be guaranteed to perform well in a population for which it was not developed or tested. This is a particularly important consideration when using an instrument in populations with different sociodemographic characteristics, cultures, or languages. Similar caution must be exercised when using an instrument that has been modified from its original construction. The deletion of items from an instrument, changing of item order, and addition of items may all have an effect on the performance of the instrument, no matter how small or useful the changes may seem.

Reliability

Reliability measures the ability of a measurement to produce the same results on repeated administrations. *Test–retest reliability* assesses the ability of a measure to arrive at the same result for the same subject on repeated administrations. The interval between the test and the retest should be long enough to ensure that the person's responses are based on their current condition rather than on the memory of their responses in the first test administration. If the interval is too long, however, there is an increased risk that the person's condition may have changed between test and retest. Another frequently used assessment of reliability is *inter-rater reliability*. This test describes a measurement instrument's ability to produce the same result when administered by two different raters.

The importance of reliability is obvious: if a measure cannot produce the same results on repeated administrations and across various raters, then it has limited value in assessing the presence or absence of a condition. Test–retest reliability is enhanced when the construct being assessed is stable across time, when it is specifically defined, and when the assessment instrument is standardized. In addition, reliability is more easily demonstrated when the individuals being assessed are clear cases of the disorder, not cases barely meeting the threshold for caseness. The reliability of an instrument, particularly a diagnostic instrument, is therefore generally lower when measured in a community population, where untreated cases around the diagnostic threshold abound, as opposed to a more selected population, such as a clinical population. Individuals chosen from clinical populations are more likely to fall well beyond the diagnostic threshold, are more accustomed to answering questions related to their clinical status, and are less likely to change answers on retest and less likely to fall below a diagnostic threshold.

Table 5.1–2 is used to illustrate the outcomes of inter-rater and test–retest reliability assessments, where the outcome is dichotomous—each individual must be classified as a case or a noncase. Cells-a and d represent individuals for whom there is an agreement between both raters (for an inter-rater reliability assessment) or ratings (for a test–retest assessment). Cells-b and c represent individuals for whom there is a disagreement between raters or ratings. The statistic most commonly used to describe the reliability of an instrument is the *Cohen kappa statistic*, which calculates agreement adjusted for chance. In Table 5.1–2, observed agreement is represented by (a + d)/N, chance agreement is represented by [(n1 × n3) + (n2 × n4)]/N^2.

$$\text{Kappa} = \frac{\text{observed agreement} - \text{chance agreement}}{1 - \text{chance agreement}}$$

The *kappa statistic* can vary between 0 and 1, with 0 representing agreement no better than chance and 1 representing perfect agreement. Interpretation of kappa scores between 0 and 1 has not been standardized, but guidelines suggested by Landis and Koch in 1977 are frequently used. These guidelines divide the kappa range into five equal categories (0 to 0.2, 0.2 to 0.4, 0.4 to 0.6, 0.6 to 0.8, 0.8 to 1.0). The range of 0.4 to 0.6 is termed “fair” or “moderate” agreement, and a kappa coefficient of at least 0.4 is generally considered to represent adequate reliability for a diagnostic instrument.

Validity

Validity is the extent to which an assessment instrument measures what it is purported to measure. There are several types of validity.

Content Validity. *Content Validity* is concerned with how well a measure covers the construct that is being measured. For a diagnostic instrument assessing a psychiatric disorder such as major depression, content validity might at a minimum assess how well the DSM criteria are covered. For example, an instrument that does not inquire about depressed mood or that inquires about spending sprees would have suspect content validity as a measure of major depressive disorder. Content validity also depends on the population that the instrument is meant to cover and the method by which data are gathered. For example, one would not ask men about menstrual symptoms when assessing somatic complaints, nor would one ask a respondent to estimate their current blood pressure as opposed to directly measuring it. Content validity cannot be empirically measured but rather must be assessed by expert consensus after careful examination of the characteristics of the construct that is being measured and comparison with other available instruments.

Criterion Validity. *Criterion Validity* examines the measurement instrument in relationship to a factor that is external to the measurement, known as the criterion. Criteria used to assess the validity of a measurement for a mental disorder are problematic in that there is usually no “gold standard” by which to assess the definitive presence or absence of disorder. Semistructured clinical interviews are often used as a criterion, in the absence of a gold standard, although they are also prone to some degree of error. Two types of criterion validity are commonly assessed, *concurrent validity* and *predictive validity*.

Table 5.1–2.
2 × 2 Table for Reliability Assessments

		Rater 1 or Test		
		Present	Absent	Total
Rater 2 or Retest	Present	a	b	n3
	Absent	c	d	n4
	Total	n1	n2	N

Concurrent validity is assessed when the measure being tested is administered along with another measure known to be related to the construct of interest. An example of concurrent validity was demonstrated in the *Methods for the Epidemiology of Child and Adolescent Mental Disorders* (MECA study), described below. In the MECA study, a highly structured lay interviewer-administered diagnostic instrument, the DISC-2.3, was compared to the results of a clinical interview administered at the same session. Predictive validity is assessed when the measure of interest is compared to a criterion that is measured in the future. In other words, the measure is assessed in terms of how well it predicts an outcome. Predictive validity is particularly important in assessing a test that predicts the later development of a disorder or a person’s response to treatment. Criterion validity is often expressed in terms of the sensitivity, specificity, and predictive value of a test illustrated by Table 5.1–3.

True positives, in which the test correctly identifies a case, are represented in cell a. *True negatives* in which the test correctly identifies a noncase are represented in cell d. A perfect test would categorize all persons in one of these cells. In practice, however, no test is totally accurate; these measurement errors should be minimized. Errors of measurement *include false positives*, in which the test assigns caseness to a noncase (cell b), and *false negatives*, in which the test does not recognize a case (cell c). The degree to which a test correctly assigns cases and noncases is described by the concepts of *sensitivity* and *specificity*.

Table 5.1–3.
2 × 2 Table to Determine Measures of Criterion Validity

		Criterion		
		Positive	Negative	Total
Test	Positive	a	b	n3
	Negative	c	d	n4
	Total	n1	n2	N

Sensitivity is described as the proportion of cases that are correctly assigned as cases by the test.

Sensitivity = a/n1

Specificity can be described as the proportion of noncases that are correctly assigned as noncases by the test.

Specificity = d/n2

Sensitivity and specificity vary between 0 and 1. Maximizing the sensitivity and specificity of an instrument has important ramifications. A high sensitivity assures that a minimal number of persons with a disorder will be missed by the test, an obvious necessity in clinical or screening situations. High specificity minimizes the chance that an individual will receive unnecessary treatment or confirmatory tests.

Predictive value is another measure that can be derived from data in Table 5.1–3. *Positive predictive value* is the proportion of persons who test positive for the disorder who actually have the disorder. *Negative predictive value* is the proportion of persons who test negative for the disorder who do not have the disorder.

Positive predictive value = a/n3

Negative predictive value = d/n4

Construct Validity. *Construct validity* is the most complex type of validity assessment, indicating how well the measure of interest relates to a theoretical set of concepts that, taken together, describe the condition being measured. Implicit in a determination of construct validity is a theoretical model of the condition being measured, with an explanation of how the components of the model are related. The most well-known theoretical model of the qualities of a psychiatric disorder was described in 1972 by Robins and Guze. This model posits the following components of a valid psychiatric diagnosis:

Clinical description—A clear description of the clinical syndrome is needed, which may include not only symptoms but also associated factors such as the age of onset, precipitating factors, and sex ratios.

Laboratory studies—Consistent and reliable findings on tests do much to increase the validity of a diagnosis. Such tests may include psychological tests as well as biochemical, imaging, anatomical, and physiologic tests.

Delimitation from other disorders—Robins and Guze argued that different disorders could have similar clinical descriptions and test results, so exclusion criteria should be specified to differentiate disorders, as well as to exclude borderline and doubtful cases.

Follow-up study—Once diagnosed, the natural history of a disorder (course and outcomes) should be consistent, and certainly, the diagnosis should not change over time.

Family study—An increased prevalence of the same disorder among family members lends credibility to the validity of the diagnosis.

It is clear from this list of validators that establishing construct validity is a process that involves more than a single study, unlike other tests of validity. Ideally, multiple studies should be used to replicate a validator using different measurement tools. The slow progress in establishing construct validity for psychiatric disorders indicates that further research is needed; as noted above, this may include a rethinking of the categorical model of mental disorders to include dimensional symptom assessments, as well as continued work on the neurobiology of basic mental and behavioral processes.

CASE IDENTIFICATION

Background

As in all fields of epidemiology, case identification is a key issue in psychiatric epidemiology. In contrast to illnesses where pathologic processes or lesions can be identified through imaging, cellular pathology, blood tests, and so on, making a psychiatric diagnosis relies on the self-reported history of the individual being assessed, reports of other informants, and ancillary sources of information such as past medical records. It should be noted that this state of affairs is not limited to psychiatric disorders; some neurologic disorders such as migraine headaches and multiple sclerosis, for example, lack definitive diagnostic tests, and diagnosis relies heavily on the accumulation of supportive clinically reported data.

The Diagnostic and Statistical Manual of Mental Disorders

The development of the diagnostic system for mental disorders has obvious implications for the development of psychiatric epidemiology. Although various diagnostic classifications have been in use since ancient times, the development of psychiatric epidemiology has been inextricably linked to the development of the DSM. The first edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-I) was published in 1952 by the American Psychiatric Association. Both DSM-I and its successor, DSM-II, published in 1968, contained glossaries that described the diagnostic categories in the manuals. Clinicians were expected to make diagnoses based on these descriptions and their clinical judgment. Given the latitude for variations in interpretation of glossary entries, along with differences in clinical training and judgment, diagnostic reliability was problematic for DSM-I and DSM-II, and clinical uptake of the diagnostic nomenclature was inconsistent. The problems with the reliability of psychiatric diagnoses were dramatically illustrated in a landmark cross-national study of diagnostic patterns, now known as the “U.S.–U.K. Study.” This study aimed to elucidate the disparities in the rates of schizophrenia and affective disorders in the United States and the United Kingdom by examining diagnostic methods. At the time, schizophrenia was diagnosed more frequently in the United States, and affective disorders were diagnosed more frequently in the United Kingdom. The study trained a group of psychiatrists from the two countries in the administration of the present state examination (PSE), a semistructured diagnostic instrument. When these psychiatrists made diagnoses with the PSE, rather than relying totally on their clinical judgments, their diagnostic patterns were similar; no country-specific differences in assignment of affective disorder or schizophrenia were found.

The transition from DSM-II to DSM-III in 1980 allowed for a more precise case definition in psychiatry and spurred a new generation of studies in psychiatric epidemiology. The DSM-III aimed to provide clinicians with a common language to describe psychiatric disorders so that a diagnosis had the same meaning regardless of the diagnostician. It also aimed to be theoretical and dispensed with many concepts in previous editions that were tied to then-prevalent and untestable psychoanalytic theories. These aims were accomplished by replacing the glossary of DSM-II with specific diagnostic criteria in DSM-III, based on symptoms that were observable or could be elicited directly from the patient. The DSM-III was accepted relatively readily and quickly became the diagnostic standard for mental health clinicians and researchers, as well as others relying on these diagnoses, such as federal, state, and local governments, regulatory agencies, the legal system, and the insurance industry.

A revision of the DSM-III-R was released in 1987 to correct inconsistencies and clarify criteria that were shown to be problematic in DSM-III. DSM-IV was released in 1994, DSM-5 in 2013, and DSM-5-TR in early 2022. DSM-5-TR included text revisions to updated sections on associated features, prevalence, development and course, risk and prognostic factors, culture, diagnostic markers, suicide, and differential diagnosis. DSM-5-TR also included a new diagnosis, prolonged grief disorder, as well as clarifications and corrections to approximately 70 criteria sets and incorporating new findings from research that has accumulated since the release of the prior edition. As discussed below, changes in the DSM diagnostic criteria have had implications for the epidemiologic measurement of mental disorders and for comparisons between studies performed with different diagnostic criteria.

The International Classification of Diseases

The *International Classification of Diseases* (ICD) is a diagnostic nomenclature maintained by the World Health Organization (WHO). It is in use in the United States and around the world as the common language for reporting basic morbidity and mortality statistics. Because psychiatric disorders are reported as morbidity statistics, the ICD and the DSM have been associated to varying degrees for many years. The DSM-I was a variant of the sixth edition of the ICD (ICD-6), and DSM-II was developed to be released in tandem with ICD-8. ICD-9 was published in 1975, only 1 year after work began on DSM-III, and implemented in 1978. The ICD-9 terminology was, therefore, more reflective of DSM-II, and its diagnostic categories did not match those in the new DSM-III. Because of dissatisfaction across all medicine with the lack of specificity in ICD-9, it was modified for use in the United States, resulting in ICD-9-CM (*Clinical Modification*). The development of the DSM-IV and the ICD-10 was more closely coordinated than any previous efforts, and the resulting systems were highly congruent, although not identical. The WHO also, for the first time, developed “Diagnostic Criteria for Research” to accompany the ICD-10 mental disorder diagnoses. As with the disorder list and nomenclature, these diagnostic criteria were usually, but not always, congruent with DSM-IV diagnostic criteria. Several diagnostic instruments have incorporated ICD-10 criteria along with DSM criteria, allowing examination of their differences and refinement of both sets of criteria. ICD-10 was published in 1992, but in contrast to most other countries, it was not quickly adopted as the official code set for the United States. The clinical modification of the ICD-10 (ICD-10-CM) was implemented in the United States on October 1, 2015. The development of DSM-5 was intended to be closely coordinated with the development of the ICD-11. However, because the release date for ICD-11 was pushed back several years, the opportunities for coordination with DSM-5 were limited. ICD-11 was adopted at the World Health Assembly in 2019, nearly 6 years after the publication of DSM-5. Due to diligent efforts by the harmonization team, the organizational structure and classification of mental disorders in the ICD-11 and the DSM-5 are more successfully aligned in comparison to the ICD-10 and DSM-IV classifications. For example, based on the published research, 30% of the disorders were identical in both DSM-5 and ICD-11, and about 10% had only minor differences.

Disability. Over the past 30 years, the consequences of mental disorders have assumed an increasingly important role in the theoretical underpinnings of epidemiologic studies and the interpretation of their data.

The concepts of clinically significant distress or impairment in functioning have been included in the DSM definition of “mental disorder” since the publication of DSM-III. To further emphasize the importance of these concepts in establishing a threshold for diagnosis, the DSM-IV added (for most disorders) a criterion requiring clinically significant distress or impairment in an important domain of functioning before a disorder could be diagnosed. This decision was criticized on conceptual grounds for its blending of two separate constructs, symptomatology and the consequences of symptomatology, as opposed to a conceptualization of a mental disorder as a combination of symptom aggregation, severity, and duration without regard to any resulting disability. The decision was reexamined during the development of DSM-5, and the revised definition of “mental disorder” in the manual states that mental disorders are “usually associated with significant distress or disability in social, occupational, or other important activities.” This change allowed the developers of DSM-5 to remove the “clinical significance criterion” of DSM-IV, which was implemented for some disorders. Although the multiaxial assessment was removed from DSM-5, the manual continued to emphasize the importance of the clinician’s assessment of the patient’s social and occupational functioning and disability for treatment planning and prediction of outcome.

The WHO has taken a leadership role in developing the theory, definitions, and classification of concepts related to health and functioning. The *International Classification of Impairments, Disabilities, and Handicaps: A Manual of Classification Relating to the Consequences of Disease* (ICIDH) was released by the WHO in 1980 and subsequently revised as the *International Classification of Impairments, Activities, and Participation: A Manual of Dimensions of Disablement and Functioning* (ICIDH-2). The next and most current version of this classification was released in 2001 and titled the *International Classification of Functioning, Disability, and Health* (ICF). This document is part of the WHO’s “family of international classifications” that includes the ICD-10 and allows coding of complementary information on health conditions and their consequences to provide a more detailed picture of the health status of an individual or a population. The ICIDH was mainly a classification of “consequences of disease” classification, while the ICF made a determined shift to classify “components of health.” Part I of the ICF covers “Functioning and Disability,” while part II covers “Contextual Factors.” Table 5.1–4 summarizes the further organization of the ICF. The ICF uses an alphanumeric coding system for three of its components: "Body Functions and Structures," "Activities and Participation," and "Environmental Factors." Personal factors (e.g., race, gender, habits, coping styles, etc.) are not associated with codes. The categories of the ICF are nested so that broad categories have a number of subcategories. For example, Chapter 1 of the "Body Functions" section describes mental functions, which are subcategorized as global mental functions and specific mental functions. Global mental functions contain eight subcategories such as temperament and personality, sleep, orientation, and consciousness. There are 14 subcategories of "Specific Mental Functions," including attention, memory, emotional, perceptual, and mental functions of language. Although the ICF is comprehensive in its description of health domains and provides a common language for describing these domains, its main users continue to be in the physical medicine and rehabilitation fields, where interventions focus more on improving functioning and removing barriers, rather than diagnosing and treating illnesses. There is no doubt, however, that the use of a standard classification of disabilities would be helpful for psychiatric research and clinical practice by drawing attention to the conceptual distinctions between disorders and their consequences.

Table 5.1–4.
Overview of the ICF

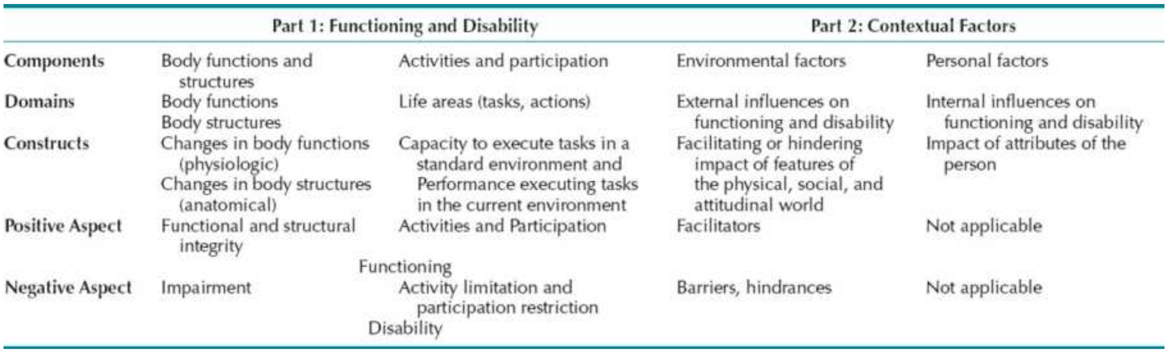

	Part 1: Functioning and Disability		Part 2: Contextual Factors	
Components	Body functions and structures	Activities and participation	Environmental factors	Personal factors
Domains	Body functions Body structures	Life areas (tasks, actions)	External influences on functioning and disability	Internal influences on functioning and disability
Constructs	Changes in body functions (physiologic) Changes in body structures (anatomical)	Capacity to execute tasks in a standard environment and Performance executing tasks in the current environment	Facilitating or hindering impact of features of the physical, social, and attitudinal world	Impact of attributes of the person
Positive Aspect	Functional and structural integrity	Activities and Participation	Facilitators	Not applicable
	Functioning			
Negative Aspect	Impairment	Activity limitation and participation restriction	Barriers, hindrances	Not applicable
	Disability			

ASSESSMENT INSTRUMENTS USED IN EPIDEMIOLOGIC STUDIES OF MENTAL DISORDERS

Diagnostic Instruments for Adult Populations

The Diagnostic Interview Schedule. The *Diagnostic Interview Schedule* (DIS) was developed in 1978 in response to two major impending events: the release of DSM-III, with its criterion-based diagnostic system, and the development of the *National Institute of Mental Health (NIMH) Epidemiologic Catchment Area* (ECA) program, which was being planned to capitalize on DSM-III's innovative new diagnostic system. Lee Robins, Ph.D., the lead developer of the DIS, wrote that there was no existing diagnostic instrument that could be used for the ECA program, not only because the DSM-III criteria sets were totally novel, but also because of limitations with the existing instruments in meeting the needs for a large community survey. Particularly important was the need for the instrument to be administered by interviewers who did not have clinical training, a necessity given the large numbers of respondents estimated for the ECA program. Other requirements for the diagnostic instrument included the ability to cover a broad range of major DSM diagnoses in a single administration, to assess specific DSM-III criteria from respondent reports only, without the use of medical records or outside informants, and also to be acceptable and comprehensible to a general adult population. The *Renard Diagnostic Interview* (RDI), developed at Washington University in St. Louis and based on Feighner diagnostic criteria, was felt to represent a suitable model for the ECA program's diagnostic instrument. The RDI was developed in a clinical setting from observations of psychiatrists' clinical interviews based on Feighner criteria. The instrument was "fully structured," meaning that its questions, response options, and administration patterns were all specified and therefore did not require clinical judgment for administration or data recording. Its reliability and concurrent validity were good in clinical populations, with lay and clinical interviewers. The task of developing a diagnostic instrument for DSM-III was assigned to the RDI's authors. The DIS covered approximately 28 DSM-III diagnoses, as well as diagnoses based on the earlier Feighner criteria and Research Diagnostic Criteria (RDC). The DIS generally assessed symptomatology first on a lifetime basis (i.e., "Have you ever had [symptom…]"?) and then assessed clustering of the positive lifetime symptoms into an episode. If symptoms did cluster together at the same time, an assessment of recency was made, so that period prevalence estimates could be made (e.g., 1-year, 6-month, 1-month prevalence rates). Like the RDI, the DIS attempted to distinguish "true" psychiatric symptoms or syndromes from normal reactions to life events and from conditions stemming directly from general medical conditions or the use of substances. Questions assessing the former situation were referred to as *clinical significance* questions; questions assessing the latter situation were referred to as the *probe flow* questions, named after the “probe flow chart” from which interviewers read these questions. The highly structured nature of the DIS allowed diagnoses to be made by computer via a diagnostic algorithm. This diagnostic algorithm allowed a diagnosis to be made free from any diagnostic custom or clinical interpretation of the interviewer, and along with the operationalized DSM-III criteria and the highly structured interview format, allowed a reliable diagnosis to be made. Testing showed that the DIS achieved the DSM-III’s goal of diagnostic reliability and that concurrent validity with a clinical research interview was reasonably strong. The DIS was updated to account for changes from DSM-IIIR to DSM-IV and to take advantage of new survey techniques and lessons learned from surveys using the DIS and similar diagnostic instruments. For instance, assessment of disability in the context of mental disorders was greatly expanded, diagnostic coverage was expanded to include disorders of childhood and an updated assessment of cognitive impairment, and there is less frequent use of skip-outs (where respondents who deny a cardinal symptom of a disorder are not asked any other symptoms of the disorder). The DIS has been superseded by the Composite International Diagnostic Interview (CIDI).

The Composite International Diagnostic Interview. The *Composite International Diagnostic Interview* (CIDI) was first developed in the early 1980s at the request of the U.S. Alcohol, Drug Abuse, and Mental Health Administration and the WHO, which had formed a Joint Project on Diagnosis and Classification. Originally designed to incorporate questions from both the DIS (DSM-III) and the *Present State Examination* for expanded use in epidemiologic studies worldwide, it was subsequently modified to cover other diagnostic classifications, including DSM-III-R, DSM-IV, and ICD-10. The CIDI was based largely on the DIS and shared many of its questions and characteristics, including highly structured administration and scoring, assessment of psychopathology beginning at a lifetime level, and use of clinical significance and probe questions. Taking advantage of the increasingly rapid uptake of computers throughout the world, a computerized version of the CIDI was developed.

An expanded module for detailed examination of substance use disorders was also developed, called the CIDI-SAM. One of the intended first uses of the CIDI was the *National Comorbidity Survey* (NCS).

The Composite International Diagnostic Interview, University of Michigan Version. The paper-and-pencil administered DSM-III-R/ICD-10 version of the CIDI was modified by the investigators in the NCS, which was conducted out of the University of Michigan. The resulting instrument was called the UM-CIDI. The UM-CIDI was thought to bring major improvements to the CIDI while still maintaining the basic CIDI structure. Nonetheless, the developers of the modified instrument acknowledged that the decision to modify the instrument was not agreed on by all of the original CIDI developers. Among the modifications made in the UM-CIDI were changes in the layout of the instrument form; the addition of an introductory question asking the respondent if they are willing to think carefully in order to provide the interviewer with accurate information; deletion of diagnoses with low reliability or of secondary importance to the goals of the NCS; using the CIDI psychosis module as a screener for follow-up clinical interviews, rather than as a self-contained diagnostic module; moving the diagnostic stem questions to their own section at the beginning of the interview; reordering questions; adding and rewriting questions for clarity; adding notes to help interviewers and reduce the amount of memorization required of them; changing the methods of administering the probe flow questions, and adding a respondent booklet with visual cues. Although many of these modifications in the UM-CIDI did address problems detected in previous usages of the DIS and the CIDI, it was not clear whether they all represented improvements when implemented in a major household survey such as the NCS. The differences between the UM-CIDI and the DIS were, therefore, among the primary reasons cited for the disparities in prevalence rates between the ECA and the NCS. Attempts to reconcile these prevalence rates stimulated a critical examination of the instruments and survey methods in these two landmark surveys, as discussed below.

The Composite International Diagnostic Interview, World Mental Health Surveys Version. The CIDI underwent further modification in anticipation of the WHO WMH Survey Initiative. It had been recognized that although the CIDI provided a common diagnostic assessment, it did not provide guidance on other domains critical to large-scale epidemiologic surveys. Thus, it was difficult to compare risk factors for disorders, consequences of disorders, patterns of disorders, and correlates of service use across surveys. The WMH CIDI was created to address these inconsistencies in assessments. It consists of seven sections: Screening and lifetime review, disorders, functioning, and physical disorders, treatment, risk factors, socio-demographics, and a methodologic section. The interview is designed to be conducted in two parts so that interviews can be ended for a subset of subjects who do not meet the criteria for lifetime psychopathology. A number of changes were made to the CIDI's diagnostic questions to improve the quality of responses, and several new diagnoses were added, such as intermittent explosive disorder, separation anxiety disorder, and attention deficit/hyperactivity disorder. A fuller assessment of disability was instituted, including expanded diagnostic-level clinical significance questions, a modified *WHO Disability Assessment Schedule* (WHODAS), and detailed questions on work performance. New sections on pharmacoepidemiology, nonspecific psychological distress, family burden, and childhood experiences were also included. Research on the validity of this instrument across the diverse range of nations in the WMH Survey Initiative is ongoing. The CIDI 5.0 and WMH CIDI 5.0 were updated with diagnostic criteria after the ICD-10 and the DSM-5.

The Alcohol Use Disorder and Associated Disabilities Interview Schedule. The *Alcohol Use Disorder and Associated Disabilities Interview Schedule* (AUDADIS) was originally developed at the National Institute on Alcoholism and Alcohol Abuse (NIAAA) in 1989, with a focus on substance use disorders and the mental disorders commonly comorbid with them. It was subsequently updated for DSM-IV and DSM-5 diagnostic criteria. The AUDADIS-5, based on DSM-5 diagnostic criteria, was completed before the release of the DSM-5, although the developers of the instrument note that the criteria for the disorders covered in the instrument were substantially finalized by 2011. It provides an assessment of criteria for substance use disorders including alcohol, tobacco, and other drugs; depressive and bipolar disorders; posttraumatic stress disorder; and personality disorders. The AUDADIS has been used to make DSM diagnoses in the NIAAA National Longitudinal Alcohol Epidemiologic Survey (NLAES) and the three versions of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). As with other instruments used in large epidemiologic surveys, it is fully structured, computerized, and can be administered by lay interviewers.

Structured Clinical Interview for DSM-IV (SCID). The *Structured Clinical Interview for DSM-IV* (SCID) is a semistructured interview guide for making diagnoses according to the criteria published in the DSM. The SCID was developed and piloted in the years following the publication of the DSM-III. Initially, it had 2 modules: SCID-P (patient), which was designed for use with patients with mental health issues. The other version, SCID-NP (Nonpatient), was designed for use in surveys. SCID-IV was published in 1994; it was based on DSM-IV criteria and had two versions, SCID-I and SCID-II. SCID-I assessed mental and substance use disorders while SCID-II assessed personality disorders. Following the release of DSM-5, the SCID for DSM-5 (SCID-5) was published in 2013. SCID-5 has many sub-types, including the Research Version (SCID-5-RV), Clinician Version (SCID-5-CV), Clinical Trials Version (SCID-5-CT), Personality Disorders Version (SCID-5-PD) as well as an Alternative Model for Personality Disorders Version (SCID-5-AMPD). SCID-5 is intended for interviewing adult individuals. A child and adolescent version of the SCID is currently being developed. The SCID is administered by either a clinician or trained mental health professional who is familiar with the DSM-5 classification and diagnostic criteria. It can be used to interview mental health patients or individuals with no mental health diagnosis as part of a survey questionnaire.

Diagnostic Interviews for Child and Adolescent Epidemiology

The development of diagnostic interviews for child and adolescent psychiatric epidemiology has paralleled that of instrument development for adult populations. However, the complexities involved in obtaining diagnoses in these populations and funding priorities placed on adult research have resulted in somewhat slower progress.

In 1969 a semistructured instrument patterned after the Renard Diagnostic Interview was developed, the *Diagnostic Interview for Children and Adolescents* (DICA). The development of the DIS for adults in the late 1970s spurred much interest in developing a fully structured lay interview schedule for assessing DSM-III diagnoses in children and adolescents. There is currently no available DSM-5 version of the DICA.

Under the sponsorship of the NIMH, the first version of the *Diagnostic Interview Schedule for Children* (DISC), covering DSM-III diagnoses, was developed in the early 1980s, and work on the DISC-R, covering DSM-III-R diagnoses, was started in 1985. Refinements of the DISC-R led to the DISC-2.1 and -2.3, the latter version being extensively tested for reliability and validity in the *Methods for the Epidemiology of Child and Adolescent Mental Disorders* (MECA) study. Findings from the MECA study led to the DISC-IV, which covers DSM-IV (and also DSM-III-R and ICD-10) disorders. The DISC-IV assesses over 30 disorders, organized into 19 diagnostic sections. The DISC-IV is a fully structured diagnostic interview that is suitable for administration by lay interviewers. A computer-assisted version (C-DISC) has been developed to simplify interviewer administration and reduce errors and costs. The DISC-IV has matching parent and child/youth versions (DISC-P and DISC-C); a partial teacher version (DISC-T) is also available. The DISC-P can be administered to parents of children ages 6 to 17 years; the DISC-C can be administered to children ages 9 to 17 years; the DISC-T is suitable for teachers of elementary school-age children. Symptoms are assessed on a past-year and past-month basis. An optional "whole-life" module was developed for administration at the end of the interview. Questions on age and context of onset, receipt of treatment, and distress and disability are asked at the end of each diagnostic section if a respondent meets a predetermined number of symptoms for that diagnosis. Diagnoses are made by computer algorithms. Diagnostic Interview Schedule for Children, Version 5 (DISC-5) includes updates to both the ADHD and tic disorder modules to reflect changes in DSM-5 criteria. Although the DISC is the most widely used epidemiologic diagnostic instrument in child and adolescent populations, other instruments are in use.

The *Child and Adolescent Psychiatric Assessment* (CAPA), developed at Duke University, was used in the Great Smoky Mountains Study. The CAPA provides DSM-IV and ICD-10 diagnoses with a high level of detail and can be administered by lay interviewers. It requires intensive training to provide the interviewers with a sufficient level of judgment to make decisions on the presence or absence of symptoms (referred to as an "interviewer-based" approach). However, with training and ongoing monitoring of interviewers, the CAPA works well in community surveys with children and adolescents ages 9 to 17 and their parents. The developers of the CAPA have also developed the Young Adult Psychiatric

Assessment (YAPA) for respondents aged 18 and over and the Preschool Age Psychiatric Assessment (PAPA) for administration to parents of 2- to 5-year-old children. The DICA was updated to provide DSM-IV diagnoses; it is a semistructured instrument that allows interviewers to probe responses. A DSM-5 version of the CAPA is available and includes questions related to depression, anxiety, oppositional defiant disorder, conduct disorder, ADHD, and impairment modules.

Kiddie-Schedule for Affective Disorders and Schizophrenia (K-SADS) was adapted from the SADS (Schedule for Affective Disorders and Schizophrenia) questionnaire to be used in children ages 6 to 18 years old. The K-SADS is a semistructured interview that measures past and current symptoms of mood, anxiety, psychotic, and disruptive behavior disorder. There are multiple versions of K-SADS developed by different groups. The lifetime version of K-SADS (K-SADS-PL DSM-5) was developed in 2016 to make it more compatible with DSM-5. Prior to administering the interview portion of K-SADS-PL, participants complete the DSM-5 cross-cutting 25-item symptom rating scales, responses of which are then taken into account in assessing the interview responses.

Instruments to Assess Disability and Impairment

There are very few instruments that have been developed to assess disability in community-based epidemiologic surveys. Most instruments require clinical training to administer and have been developed for patient populations. The development of instruments for child and adolescent populations has seen more activity than the development of instruments for adults, in part because the diagnostic threshold issues for younger populations were recognized as a particular problem.

WHO Disability Assessment Schedule. The *WHO Disability Assessment Schedule* (WHODAS 2.0) was developed to be compatible with the disability model put forth in the ICF, particularly in the activity and participation dimensions. It is fully structured and can either be administered to adult subjects by lay interviewers or be self-administered. The core instrument is self-report, but a version has been developed that can be administered to a proxy respondent. The reporting period is the past 30 days. The full version of the WHODAS 2.0 is 36 items long and assesses six domains: cognition (understanding and communicating), mobility (moving and getting around), self-care (hygiene, dressing, eating, and staying alone), getting along (interacting with other people), life activities (domestic responsibilities, leisure, work, and school), and participation (joining in community activities). There is a short version that is 12 items long. An earlier version of the WHODAS 2.0, the WHODAS II, was substantially modified for use in the WHO World Mental Health Survey Initiative. These modifications appear to have reduced the sensitivity of the instrument to detect milder levels of disability, complicating the interpretation of the results. The full version of the WHODAS 2.0 was tested in the DSM-5 field trials and was found to have good to excellent reliability for the domains and total scores in randomly sampled adults at seven clinical sites.

Disability Assessment in Child and Adolescent Populations. The *Children's Global Assessment Scale* (CGAS) is based on the adult Global Assessment Scale (GAS). Although developed for clinical use, an adaptation of this scale with nonclinical language, for use by lay interviewers, was developed for use in community surveys and was administered in the MECA survey (see below). Like the GAS, the rater assigns a single number that best describes the child's functioning, from 1 (most impaired functioning) to 100 (healthiest). A score of 70 or less indicates at least mild functional impairment. The Brief Impairment Scale (BIS) grew out of work on the Columbia Impairment Scale (CIS), which was used in the MECA survey. It is a fully structured, 23-item scale that assesses the child's functioning without tying it to any specific health conditions that the child may have. The BIS was designed to allow administration by a lay interviewer in community surveys, but it is also suitable for use in clinical settings. It is administered to an adult informant. Three domains are assessed: interpersonal relations, functioning at school or work, and self-fulfillment. Initial testing of the scale has shown that it has acceptable test–retest reliability, internal consistency, and convergent and concurrent reliability. The three domains show low intercorrelations, suggesting that they are related but separate components of the overall global functioning measured by the scale. Several other assessment scales are available but have limitations, including unidimensionality and mixing of symptom and disability constructs, or have practical difficulties for use in community surveys, such as length of administration or a need for clinical judgment in scoring.

MAJOR EPIDEMIOLOGIC SURVEYS

Large surveys do not have the timeframe or opportunity for clinical judgment. Therefore, the considerations for generating prevalence estimates typically include what instruments are used, who administers them, and how generalizable the findings are. Table 5.1–5 provides some examples of such survey instruments.

Table 5.1–5.
Examples of Survey Instruments Assessing Mental Illness

Survey Name	Instrument	Who Administers It	How Long It Takes
Epidemiologic Catchment Area (ECA) Survey of Mental Disorders	Structured interview–Family Life Survey	Trained field workers	1.5 hours
	Structured Interview–Home Interview Survey	Psychologists, social workers	2 hours
	Structured Interview–Schizophrenia and Affective Disorders Schedule-Life time version	Nonpsychiatrist mental health professionals	1.5 hours
NCS	Structured Clinical Interview for DSM-III-R (SCID)/ Clinical reappraisal interviews–based on the SCID	Trained clinical interviewers	NR
	Adult ADHD Self-Report Scale (ASRS)	Trained clinical interviewers	NR
	Health and Work Productivity Questionnaire (HPQ)	Trained clinical interviewers	NR
	K6-K10 Scales	Trained clinical interviewers	NR
	Tier 2 Disaster Survey Pilot Interview Schedule	Trained clinical interviewers	NR
NCS-A	Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS).	Trained lay interviewers	NR
NCS-R; NCS-A	World Mental Health Survey	Trained lay interviewers	2 hours

	Initiative version of the Composite International Diagnostic Interview (WMH-CIDI)/ CIDI Primary Care Screening Tool (PCST) and PCST Scoring rules/CIDI 3.0 Bipolar Screening Scales		
NSDUH	Structured Clinical Interview for DSM-5—Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV)	Clinical interviewers	30–60 minutes
	World Health Organization Disability Assessment Schedule (WHODAS)	Either self-administered or administered by an interviewer (no training required)	15–20 minutes
	Kessler-6 Psychological Distress Scale (K6)	Either self-administered or administered by an interviewer (no training required)	2 minutes
	Nicotine Dependence Syndrome Scale (NDSS) and the Fagerstrom Test of Nicotine Dependence (FTND)	Either self-administered or administered by an interviewer (no training required)	NR
NESARC-III	NIAAA Alcohol Use	Trained	NR

	Disorder and Associated Disabilities Interview Schedule (AUDADIS-5)	interviewers	
YRBSS	YRBS	Self-administered	NR

NR, not reported.

Modified from Bagalman E, Cornell AS. Prevalence of Mental Illness in the United States: Data Sources and Estimates. Congressional Research Service, Library of Congress, 2013. http://www.fas.org/sgp/crs/misc/R43047.pdf Accessed November 1, 2021.

Additional data from Regier DA, Myers JK, Kramer M, et al. The NIMH Epidemiologic Catchment Area program. Historical context, major objectives, and study population characteristics. *Arch Gen Psychiatry.* 1984;41(10):934–941; Substance Abuse and Mental Health Services Administration. Center for Behavioral Health Statistics and Quality, 2021; Youth Risk Behavior Surveillance System (YRBSS): https://www.cdc.gov/healthyyouth/data/yrbs/index.htm: National Comorbidity Survey: https://www.hcp.med.harvard.edu/ncs/index.php.

Surveys of Adult Populations

The Epidemiologic Catchment Area Program. Spurred by the release of the President’s Commission on Mental Health, chaired by First Lady Rosalynn Carter, the *Epidemiologic Catchment Area Program* (ECA) program was developed by the NIMH in conjunction with the simultaneous development of the DSM-III and the DIS. The main aims of the ECA were to estimate the prevalence of DSM-III mental disorders in community and institutional populations and determine the extent of the use of treatment services for these disorders. The ECA was carried out in five catchment areas across the United States: New Haven, Connecticut; Baltimore, Maryland; Durham, North Carolina; St. Louis, Missouri; and Los Angeles, California. In order to analyze specific subgroups of interest, oversampling of the older adults, African Americans, and Hispanics was done at selected sites. There were two samples in each site, a household sample, and an institutional sample, which included jails and nursing homes; respondents were all aged 18 or older. The final sample size was 18,571 household respondents and 2,290 institutional respondents. The ECA was designed as a longitudinal survey with three waves conducted 6 months apart with the same respondents. The first and third waves collected diagnostic and service use data, while the second wave collected service use data to reduce recall bias in service use reports. The ECA program’s sampling strategy was not nationally representative, but the investigators combined the data from the five sites and adjusted them to the U.S. population to create national estimates. Prevalence and service use rates from the ECA program are shown in Table 5.1–6. Over time, the analysis of data generated by the ECA program served to point out weaknesses in the DSM-III and its subsequent revisions, as well as provide valuable data for service planning and analytic studies. The high prevalence of diagnostic comorbidity among DSM-III disorders, particularly in the absence of DSM-III diagnostic hierarchies, was one of the first major issues uncovered by the analyses of ECA data. Surprisingly, a high prevalence of disorders previously thought to be rare, such as social phobia and obsessive-compulsive disorder, was found in the community; low rates of treatment for these and most other psychiatric disorders contributed to this mistaken assumption. The onset of many disorders was found to originate in childhood, contrary to much contemporary thinking that children were relatively protected from psychiatric disorders. Data from the ECA survey were instrumental in educating the general public as well as policymakers and legislators that mental disorders are real, prevalent in U.S. society, disabling, and grossly undertreated. Such basic efforts provided support, through an increasingly detailed series of reports released by NIMH, for parity of mental health insurance benefits with the benefits for other medical or surgical conditions and treatments. A similar purpose was served for the landmark *Surgeon General's Report on Mental Health,* published in 1999. More recently, the DIS clinical significance criteria data provided a crucial piece of information to evaluate the role of clinically significant distress and functional impairment in diagnosis and estimation of prevalence rates, as detailed below. Longitudinal follow-ups of the Baltimore ECA sample continued 16 and 22 years after the sample was originally drawn. Among the findings reported from these follow-ups are reports that major depressive episode is an independent risk factor for the development of hypertension and that cluster B personality disorders and traits are independent risk factors for the development of cardiovascular disease as well as cardiovascular mortality. Finally, the ECA program served as a stimulus for similar surveys in other locations, such as Puerto Rico and the Netherlands, and set the stage for subsequent U.S. and international surveys, such as the NCS and the WMH Survey Initiative.

Table 5.1–6. Effect of Clinical Significance Criteria on 1-Year Prevalence Rates of Mental Disorders in Adults, Age 18 to 54, ECA and NCS Surveys

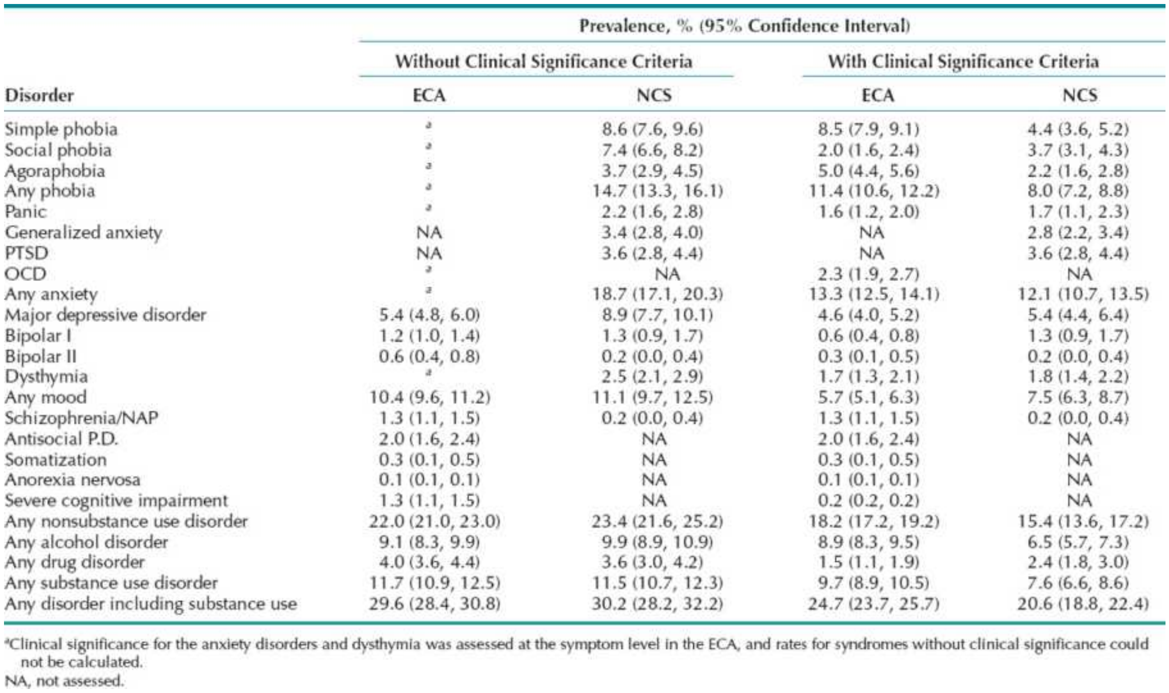

	Prevalence, % (95% Confidence Interval)			
	Without Clinical Significance Criteria		With Clinical Significance Criteria	
Disorder	ECA	NCS	ECA	NCS
Simple phobia	[a]	8.6 (7.6, 9.6)	8.5 (7.9, 9.1)	4.4 (3.6, 5.2)
Social phobia	[a]	7.4 (6.6, 8.2)	2.0 (1.6, 2.4)	3.7 (3.1, 4.3)
Agoraphobia	[a]	3.7 (2.9, 4.5)	5.0 (4.4, 5.6)	2.2 (1.6, 2.8)
Any phobia	[a]	14.7 (13.3, 16.1)	11.4 (10.6, 12.2)	8.0 (7.2, 8.8)
Panic	[a]	2.2 (1.6, 2.8)	1.6 (1.2, 2.0)	1.7 (1.1, 2.3)
Generalized anxiety	NA	3.4 (2.8, 4.0)	NA	2.8 (2.2, 3.4)
PTSD	NA	3.6 (2.8, 4.4)	NA	3.6 (2.8, 4.4)
OCD	[a]	NA	2.3 (1.9, 2.7)	NA
Any anxiety	[a]	18.7 (17.1, 20.3)	13.3 (12.5, 14.1)	12.1 (10.7, 13.5)
Major depressive disorder	5.4 (4.8, 6.0)	8.9 (7.7, 10.1)	4.6 (4.0, 5.2)	5.4 (4.4, 6.4)
Bipolar I	1.2 (1.0, 1.4)	1.3 (0.9, 1.7)	0.6 (0.4, 0.8)	1.3 (0.9, 1.7)
Bipolar II	0.6 (0.4, 0.8)	0.2 (0.0, 0.4)	0.3 (0.1, 0.5)	0.2 (0.0, 0.4)
Dysthymia	[a]	2.5 (2.1, 2.9)	1.7 (1.3, 2.1)	1.8 (1.4, 2.2)
Any mood	10.4 (9.6, 11.2)	11.1 (9.7, 12.5)	5.7 (5.1, 6.3)	7.5 (6.3, 8.7)
Schizophrenia/NAP	1.3 (1.1, 1.5)	0.2 (0.0, 0.4)	1.3 (1.1, 1.5)	0.2 (0.0, 0.4)
Antisocial P.D.	2.0 (1.6, 2.4)	NA	2.0 (1.6, 2.4)	NA
Somatization	0.3 (0.1, 0.5)	NA	0.3 (0.1, 0.5)	NA
Anorexia nervosa	0.1 (0.1, 0.1)	NA	0.1 (0.1, 0.1)	NA
Severe cognitive impairment	1.3 (1.1, 1.5)	NA	0.2 (0.2, 0.2)	NA
Any nonsubstance use disorder	22.0 (21.0, 23.0)	23.4 (21.6, 25.2)	18.2 (17.2, 19.2)	15.4 (13.6, 17.2)
Any alcohol disorder	9.1 (8.3, 9.9)	9.9 (8.9, 10.9)	8.9 (8.3, 9.5)	6.5 (5.7, 7.3)
Any drug disorder	4.0 (3.6, 4.4)	3.6 (3.0, 4.2)	1.5 (1.1, 1.9)	2.4 (1.8, 3.0)
Any substance use disorder	11.7 (10.9, 12.5)	11.5 (10.7, 12.3)	9.7 (8.9, 10.5)	7.6 (6.6, 8.6)
Any disorder including substance use	29.6 (28.4, 30.8)	30.2 (28.2, 32.2)	24.7 (23.7, 25.7)	20.6 (18.8, 22.4)

[a] Clinical significance for the anxiety disorders and dysthymia was assessed at the symptom level in the ECA, and rates for syndromes without clinical significance could not be calculated.

NA, not assessed.

The NCS and Associated Studies. The *National Comorbidity Survey* (NCS) was an NIMH and National Institute on Drug Abuse (NIDA)-sponsored survey designed to study the comorbidity of DSM-III-R substance use disorders and other mental disorders. Unlike the ECA, it was a nationally representative household survey with 8,098 respondents. It was fielded from 1990 to 1992. Because of the relatively low prevalence of substance use disorders in older populations, the age of subjects sampled for the NCS was capped at 54 years to maximize its analytic goals; to minimize recall bias of early-onset disorders, the lower age range of the survey was 15 years. In addition to measuring a wide range of psychiatric disorders with the UM-CIDI and mental health service use, a plethora of other domains were included in the NCS survey, such as human immunodeficiency virus (HIV) risk behaviors, personality constructs, and social life. The initially published prevalence rates of mental disorders from the NCS were generally higher than those previously published from the ECA program, leading to intense scrutiny of both surveys and uncertainty as to how to use the NCS results for health service policy and planning. In particular, it was not evident that the prevalence estimates of the NCS reflected a true, substantial increase in the rates of mental disorders in the United States in the decade since the initiation of the ECA program. Nor was it evident that the ECA substantially underestimated the prevalence rates of mental disorders; indeed, the rates of disorders reported by the ECA program were themselves felt to be nonconservative. Methodologic and diagnostic differences in the two surveys were suspected, at least in part, to underlie the differences in prevalence rates. First, there were some differences between the diagnostic criteria in DSM-III and DSM-III-R, which were reflected in the respective diagnostic instruments and could affect prevalence rates. Second, differences in the age ranges of the respondents in the two surveys could affect prevalence rates; although few previous comparison data were available for the prevalence of mental disorders in adolescents aged 15 to 17 years sampled in the NCS, it had been documented that in epidemiologic surveys, prevalence rates of mental disorders declined with age and were particularly low in the older adults. Third, the effects of several innovations in the UM-CIDI, such as grouping of stem questions and the commitment question, were unknown but potentially could increase rates of positive responses in the NCS. Finally, it was recognized that the clinical significance questions in the NCS were not used to determine prevalence rates and were used inconsistently in the ECA prevalence estimates. In a study that aimed to reconcile the age range differences and apply the unused clinical significance questions, the overall prevalence estimates of the ECA were reduced by about 17%, and estimates of the NCS were reduced by about one-third. As a result, the estimates of the two surveys were brought closer together (see Table 5.1–6).

Several surveys related to the NCS have subsequently been carried out. The NSC-R surveyed another national sample from 2001 to 2003. The sample size was 9,282. This survey allowed comparisons of rates of mental disorders and service use in the decade since the original NCS survey in a repeated cross-sectional design. Many aspects of the NCS survey instrument were kept intact for the NCS-R to promote these comparisons. However, based on knowledge gained in the intervening years, new components were also added to the NCS-R, as described in the section on the WMH-CIDI above. One-year prevalence rates from the NCS-R are found in Table 5.1–7, and rates of service use are found in Table 5.1–8. Although the NCS-R permitted another cross-sectional estimate for the U.S. population, the NCS-2, by contrast, attempted to reinterview all respondents in the original NCS survey. The NCS-2 was conducted from 2001 to 2002. The longitudinal aspect of the survey has allowed investigators to examine risk factors for the new onset of disorders and to study the course of illness in persons who did have disorders in the original NCS survey. Several articles have been published using the NCS/NCS-2 data sets. These include studies of risk factors for suicide-related outcomes, mental disorders as risk factors for substance use disorders, and the role of baseline mental disorder severity in later clinical outcomes.

In the 1990s, increasing attention began to be focused on disparities in health status between the white population and racial and ethnic minority populations in the United States. These disparities appeared across diseases and covered not only differences in prevalence but also differences in the use of health services to treat these conditions. Three nationally representative surveys connected with the NCS-R were conducted to further explore these issues for mental disorders. The *National Survey of American Life* (NSAL) interviewed 3,570 African American subjects, 1,623 African-Caribbean immigrants, and 1,006 non-Hispanic white subjects. The National Latino and Asian American Study (NLAAS) conducted surveys of four Latino groups (Cuban, Mexican, Puerto Rican, and “other Latino”) with a sample size of 2,554, four Asian groups (Chinese, Filipino, Vietnamese, and “other Asian”) with a sample size of 2,095, and a small control group of non-Hispanic whites. The surveys used the WMH-CIDI, with translations into Spanish, Chinese, Vietnamese, and Tagalog for respondents who preferred to take the interview in those languages. Many articles have been published from these data sets on a variety of topics, although 12-month prevalence and service use rates for specific disorders in specific ethnic/racial groups were published in a piecemeal fashion.

Table 5.1–7.

1-Year Prevalence Rates of DSM-IV Mental Disorders in Adults, NCS-R and NESARC and NESARC-III Surveys

Disorder	**Prevalence, % (SE)**		
	NCS-R	**NESARC**[a]	**NESARC-III**[b]
Specific phobia	8.7 (0.4)	7.1 (0.3)	4.6 (0.2)
Social phobia	6.8 (0.3)	2.8 (0.1)	2.9
Agoraphobia w/o panic	0.8 (0.1)		0.9 (0.1)
Panic	2.7 (0.2)		2.7 (0.1)
Panic w/agoraphobia		0.6 (0.1)	
Panic w/o agoraphobia		1.6 (0.1)	
Generalized anxiety	3.1 (0.2)	2.1 (0.1)	4.4 (0.2)
PTSD	3.5 (0.3)		4.7
OCD	1.0 (0.3)		
Any anxiety	18.1 (0.7)	10.4 (0.3)	12.7
Major depressive disorder	6.7 (0.3)	7.1 (0.2)	15.1 (0.3)
Bipolar I or II	2.6 (0.2)		1.4
Mania		1.7 (0.1)	
Hypomania		1.2 (0.1)	
Dysthymia	1.5 (0.1)	1.8 (0.1)	3.2 (0.1)
Any mood	9.5 (0.4)	9.2 (0.2)	12.8
Any impulse control disorder	8.9 (0.5)		
Alcohol abuse	3.1 (0.3)	4.7 (0.2)	
Alcohol dependence	1.3 (0.2)	3.8 (0.1)	
Any alcohol disorder		8.5 (0.2)	13.9
Drug abuse	1.4 (0.1)	1.4 (0.1)	
Drug dependence	0.4 (0.1)	0.6 (0.1)	
Any drug use disorder		2.0 (0.1)	3.9
Any substance use disorder	3.8 (0.3)	9.4 (0.3)	3.9

Any disorder	26.2 (0.8)	29.0

[a] Rates for NESARC mood and anxiety disorders exclude substance-induced disorders.

[b] NESARC-III data based on Simpson et al., 2019; Olfson et al., 2019; Grant 2016 (some of the sources did not provide SE data).

PTSD, post-traumatic stress disorder; OCD, obsessive-compulsive disorder.

The National Survey on Drug Use and Health (NSDUH). The *National Survey on Drug Use and Health* (NSDUH) provides up-to-date information on tobacco, alcohol, drug use, mental health, and other health-related issues in the United States. The survey is designed as a face-to-face household interview targeting civilian, noninstitutionalized populations 12 years or older residing in the United States. The NSDUH survey began in 1971 and is conducted annually in all 50 states and the District of Columbia. In 2019, about 50,731 interviews with adults aged 18 or older were obtained. Diagnoses in the NSDUH survey are currently based on either DSM-IV or DSM-5 criteria depending on the disorder. Table 5.1–9 shows the substance use percentage in 2019 based on the NSDUH. The survey uses various instruments to collect data such as the Structured Clinical Interview for DSM-5—Research Version, the World Health Organization Disability Assessment Schedule (WHODAS 2.0), Kessler-6 Psychological Distress Scale (K6), the Nicotine Dependence Syndrome Scale (NDSS), and the Fagerstrom Test of Nicotine Dependence (FTND) among others.

Over the years, DSM III, III-R, IV, and 5 criteria were used to describe various mental health and substance use disorders. The NSDUH publishes reports annually with past month or past year estimates of substance use and related treatment presented separately for people aged 12 or older, adolescents, and adults. Additionally, estimates of mental health issues and mental health service use are presented for adolescents aged 12 to 17 and adults aged 18 or older.

Table 5.1–8.
1-Year Service Use by Persons in the United States and Overall, With DSM-IV Mental and Addictive Disorders, NCS-R, and World Mental Health Survey

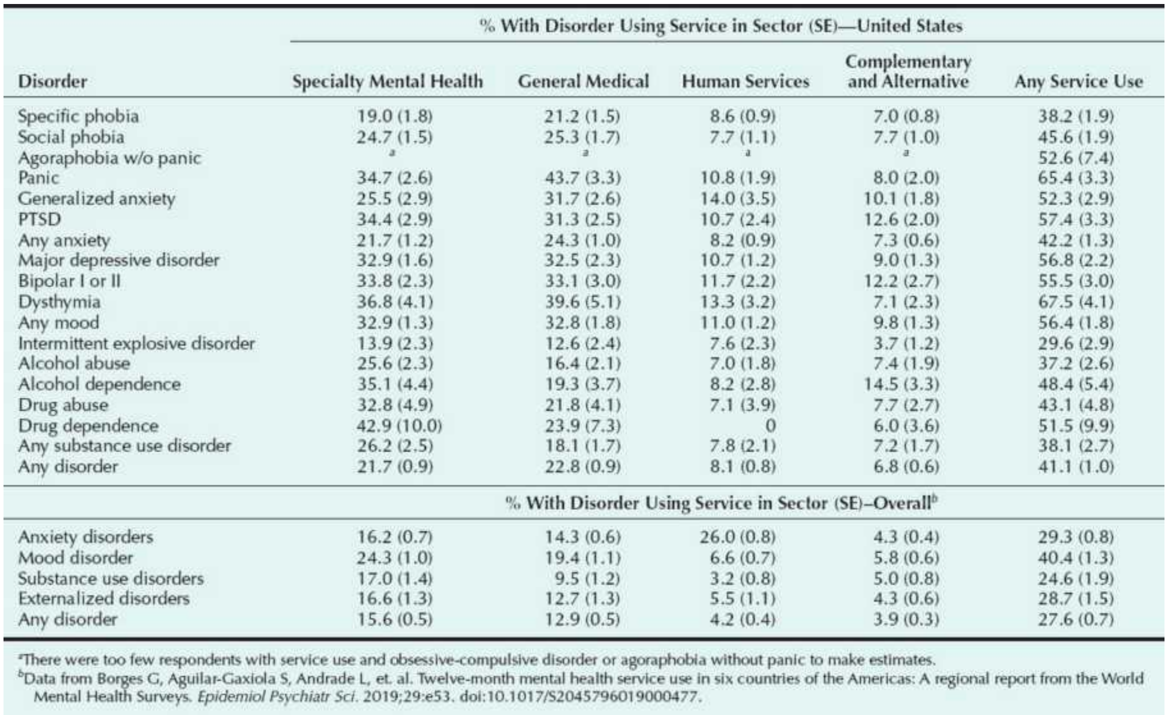

	% With Disorder Using Service in Sector (SE)—United States				
Disorder	Specialty Mental Health	General Medical	Human Services	Complementary and Alternative	Any Service Use
Specific phobia	19.0 (1.8)	21.2 (1.5)	8.6 (0.9)	7.0 (0.8)	38.2 (1.9)
Social phobia	24.7 (1.5)	25.3 (1.7)	7.7 (1.1)	7.7 (1.0)	45.6 (1.9)
Agoraphobia w/o panic	[a]	[a]	[a]	[a]	52.6 (7.4)
Panic	34.7 (2.6)	43.7 (3.3)	10.8 (1.9)	8.0 (2.0)	65.4 (3.3)
Generalized anxiety	25.5 (2.9)	31.7 (2.6)	14.0 (3.5)	10.1 (1.8)	52.3 (2.9)
PTSD	34.4 (2.9)	31.3 (2.5)	10.7 (2.4)	12.6 (2.0)	57.4 (3.3)
Any anxiety	21.7 (1.2)	24.3 (1.0)	8.2 (0.9)	7.3 (0.6)	42.2 (1.3)
Major depressive disorder	32.9 (1.6)	32.5 (2.3)	10.7 (1.2)	9.0 (1.3)	56.8 (2.2)
Bipolar I or II	33.8 (2.3)	33.1 (3.0)	11.7 (2.2)	12.2 (2.7)	55.5 (3.0)
Dysthymia	36.8 (4.1)	39.6 (5.1)	13.3 (3.2)	7.1 (2.3)	67.5 (4.1)
Any mood	32.9 (1.3)	32.8 (1.8)	11.0 (1.2)	9.8 (1.3)	56.4 (1.8)
Intermittent explosive disorder	13.9 (2.3)	12.6 (2.4)	7.6 (2.3)	3.7 (1.2)	29.6 (2.9)
Alcohol abuse	25.6 (2.3)	16.4 (2.1)	7.0 (1.8)	7.4 (1.9)	37.2 (2.6)
Alcohol dependence	35.1 (4.4)	19.3 (3.7)	8.2 (2.8)	14.5 (3.3)	48.4 (5.4)
Drug abuse	32.8 (4.9)	21.8 (4.1)	7.1 (3.9)	7.7 (2.7)	43.1 (4.8)
Drug dependence	42.9 (10.0)	23.9 (7.3)	0	6.0 (3.6)	51.5 (9.9)
Any substance use disorder	26.2 (2.5)	18.1 (1.7)	7.8 (2.1)	7.2 (1.7)	38.1 (2.7)
Any disorder	21.7 (0.9)	22.8 (0.9)	8.1 (0.8)	6.8 (0.6)	41.1 (1.0)
	% With Disorder Using Service in Sector (SE)–Overall[b]				
Anxiety disorders	16.2 (0.7)	14.3 (0.6)	26.0 (0.8)	4.3 (0.4)	29.3 (0.8)
Mood disorder	24.3 (1.0)	19.4 (1.1)	6.6 (0.7)	5.8 (0.6)	40.4 (1.3)
Substance use disorders	17.0 (1.4)	9.5 (1.2)	3.2 (0.8)	5.0 (0.8)	24.6 (1.9)
Externalized disorders	16.6 (1.3)	12.7 (1.3)	5.5 (1.1)	4.3 (0.6)	28.7 (1.5)
Any disorder	15.6 (0.5)	12.9 (0.5)	4.2 (0.4)	3.9 (0.3)	27.6 (0.7)

[a] There were too few respondents with service use and obsessive-compulsive disorder or agoraphobia without panic to make estimates.

[b] Data from Borges G, Aguilar-Gaxiola S, Andrade L, et. al. Twelve-month mental health service use in six countries of the Americas: A regional report from the World Mental Health Surveys. *Epidemiol Psychiatr Sci*. 2019;29:e53. doi:10.1017/S2045796019000477.

Table 5.1–9.
Substance Use Percentage in Past Year 2019 in the NSDUH

Disorders	Percentage		
	12–17 Years	18–25 Years	26 and Older
Illicit drugs	17.2	39.1	18.3
Marijuana	13.3	35.4	15.2
Cocaine	0.4	5.3	1.7
Heroin	NE	0.3	0.3
Hallucinogens	1.8	7.2	1.5
Misuse of Psychotherapeutics	4.3	11.5	5.3
Opioid	2.3	5.3	3.6
Tobacco products	8.3	35.2	26.8
Cigarettes	5.4	25.8	21.0
Smokeless tobacco	2.5	7.7	3.8
Alcohol	21.2	72.0	69.1

NE, not estimated.

Data from the Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: Results from the 2019 National Survey on Drug Use and Health (HHS Publication No. PEP20-07-01-001, NSDUH Series H-55). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2020. https://www.samhsa.gov/data/

World Mental Health Initiative. The paucity and noncomparability of epidemiologic data on mental disorders and their consequences in many countries, particularly in developing nations, was a hindrance to the WHO's Global Burden of Disease study (described below). For this reason, the *World Mental Health* (WMH) *Initiative* was conceived by the WHO to increase precision in future estimates of the societal burden of mental disorders and, in turn, provide a more solid basis for mental health service planning and delivery of effective interventions. The initiative aimed to study the prevalence of mental disorders, their consequences in terms of disabilities and social disadvantages, and patterns of service use for mental disorders. Twenty-nine countries participated in the WMH Initiative, representing all WHO regions (6 countries in the Pan-American region including the NCS-R in the United States, 2 in Africa, 3 in the Eastern Mediterranean, 13 in Europe including Israel, 1 in Southeast Asia, and 5 in the Western Pacific). Some countries chose to perform national surveys, and others chose to sample from smaller geographical units in their countries. All surveys did, however, use multistage household probability sampling to ensure that their results could be generalized to their target populations. In all, over 160,000 respondents were sampled worldwide. The WMH-CIDI was the diagnostic instrument, and both DSM-IV and ICD-10 diagnoses were assessed. To increase comparability in implementation, instrument translations were done according to standard WHO protocols. Interviewers for each country were trained centrally with the same documents and protocols. Informed consent and institutional review board approvals were required. The 12-month prevalence of mental disorders in an initial group of WMH-participating countries is shown in Table 5.1–10. The wide range of rates across countries was suggested to be due to variations in symptom expressions in some countries. The validity of the WMH-CIDI diagnoses was not tested at all sites, so the ability of the instrument to detect specific diagnoses in diverse cultures is uncertain. Idioms of distress and conceptualizations of psychiatric symptoms may vary across cultures, reducing the validity of a translated common measurement instrument. It was also suggested that variations in the translation, implementation, and quality control across sites may account for the prevalence differences. Countries that had the lowest disorder prevalence also had the highest proportion of cases classified as “subthreshold” which did not meet the criteria for DSM-IV/CIDI disorders. Finally, it is not possible to adequately control for the large differences in potentially relevant social factors (e.g., income, education) between countries in the WMH surveys, and these differences, when uncontrolled, could account for some portion of the prevalence differences.

Table 5.1–10.
12-Month Prevalence Estimates of DSM-IV/CIDI Disorders in the W M H Surveys[a,b]

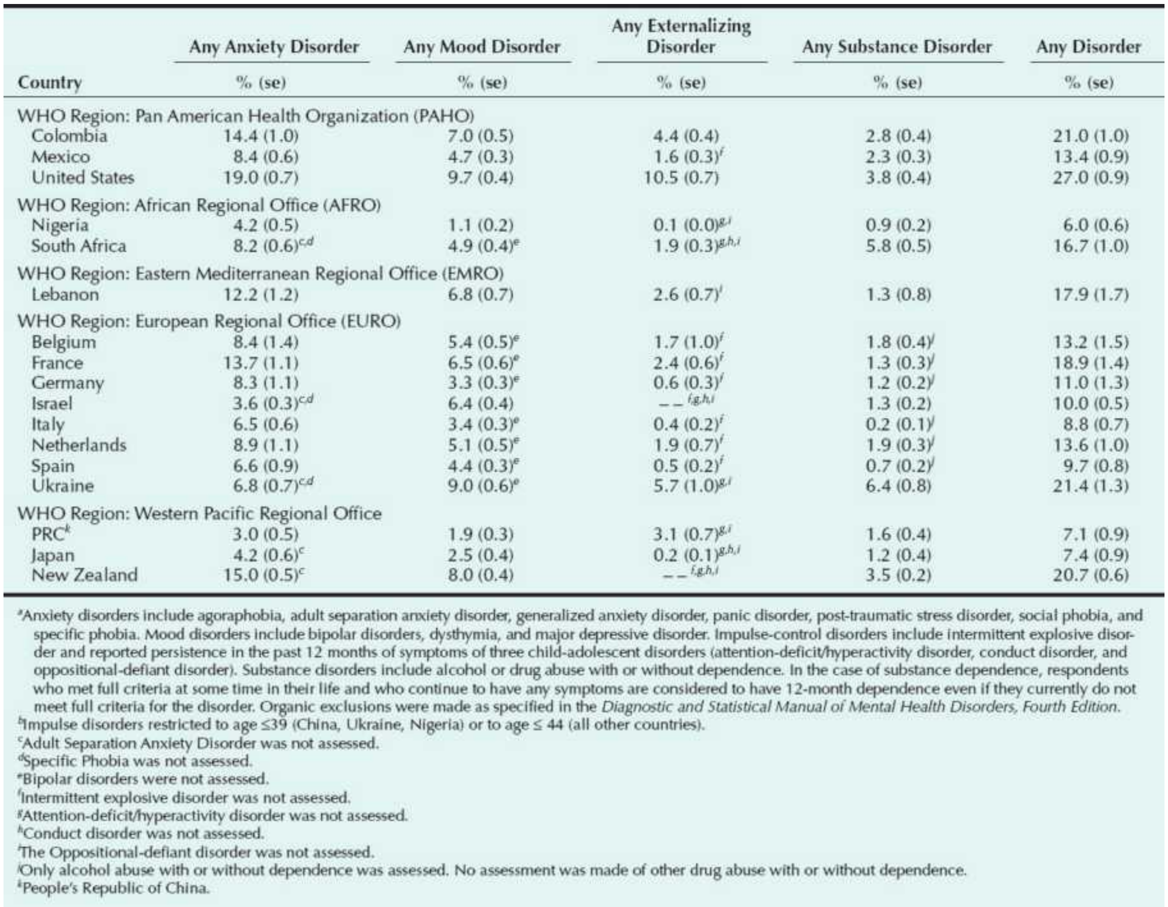

	Any Anxiety Disorder	Any Mood Disorder	Any Externalizing Disorder	Any Substance Disorder	Any Disorder
Country	% (se)	% (se)	% (se)	% (se)	% (se)
WHO Region: Pan American Health Organization (PAHO)					
Colombia	14.4 (1.0)	7.0 (0.5)	4.4 (0.4)	2.8 (0.4)	21.0 (1.0)
Mexico	8.4 (0.6)	4.7 (0.3)	1.6 (0.3)[f]	2.3 (0.3)	13.4 (0.9)
United States	19.0 (0.7)	9.7 (0.4)	10.5 (0.7)	3.8 (0.4)	27.0 (0.9)
WHO Region: African Regional Office (AFRO)					
Nigeria	4.2 (0.5)	1.1 (0.2)	0.1 (0.0)[g,i]	0.9 (0.2)	6.0 (0.6)
South Africa	8.2 (0.6)[c,d]	4.9 (0.4)[e]	1.9 (0.3)[g,h,i]	5.8 (0.5)	16.7 (1.0)
WHO Region: Eastern Mediterranean Regional Office (EMRO)					
Lebanon	12.2 (1.2)	6.8 (0.7)	2.6 (0.7)[i]	1.3 (0.8)	17.9 (1.7)
WHO Region: European Regional Office (EURO)					
Belgium	8.4 (1.4)	5.4 (0.5)[e]	1.7 (1.0)[f]	1.8 (0.4)[j]	13.2 (1.5)
France	13.7 (1.1)	6.5 (0.6)[e]	2.4 (0.6)[f]	1.3 (0.3)[j]	18.9 (1.4)
Germany	8.3 (1.1)	3.3 (0.3)[e]	0.6 (0.3)[f]	1.2 (0.2)[j]	11.0 (1.3)
Israel	3.6 (0.3)[c,d]	6.4 (0.4)	– –[f,g,h,i]	1.3 (0.2)	10.0 (0.5)
Italy	6.5 (0.6)	3.4 (0.3)[e]	0.4 (0.2)[f]	0.2 (0.1)[j]	8.8 (0.7)
Netherlands	8.9 (1.1)	5.1 (0.5)[e]	1.9 (0.7)[f]	1.9 (0.3)[j]	13.6 (1.0)
Spain	6.6 (0.9)	4.4 (0.3)[e]	0.5 (0.2)[f]	0.7 (0.2)[j]	9.7 (0.8)
Ukraine	6.8 (0.7)[c,d]	9.0 (0.6)[e]	5.7 (1.0)[g,i]	6.4 (0.8)	21.4 (1.3)
WHO Region: Western Pacific Regional Office					
PRC[k]	3.0 (0.5)	1.9 (0.3)	3.1 (0.7)[g,i]	1.6 (0.4)	7.1 (0.9)
Japan	4.2 (0.6)[c]	2.5 (0.4)	0.2 (0.1)[g,h,i]	1.2 (0.4)	7.4 (0.9)
New Zealand	15.0 (0.5)[c]	8.0 (0.4)	– –[f,g,h,i]	3.5 (0.2)	20.7 (0.6)

[a] Anxiety disorders include agoraphobia, adult separation anxiety disorder, generalized anxiety disorder, panic disorder, post-traumatic stress disorder, social phobia, and specific phobia. Mood disorders include bipolar disorders, dysthymia, and major depressive disorder. Impulse-control disorders include intermittent explosive disorder and reported persistence in the past 12 months of symptoms of three child-adolescent disorders (attention-deficit/hyperactivity disorder, conduct disorder, and oppositional-defiant disorder). Substance disorders include alcohol or drug abuse with or without dependence. In the case of substance dependence, respondents who met full criteria at some time in their life and who continue to have any symptoms are considered to have 12-month dependence even if they currently do not meet full criteria for the disorder. Organic exclusions were made as specified in the *Diagnostic and Statistical Manual of Mental Health Disorders, Fourth Edition.*

[b] Impulse disorders restricted to age ≤39 (China, Ukraine, Nigeria) or to age ≤ 44 (all other countries).

[c] Adult Separation Anxiety Disorder was not assessed.

[d] Specific Phobia was not assessed.

[e] Bipolar disorders were not assessed.

[f] Intermittent explosive disorder was not assessed.

[g] Attention-deficit/hyperactivity disorder was not assessed.

[h] Conduct disorder was not assessed.

[i] The Oppositional-defiant disorder was not assessed.

[j] Only alcohol abuse with or without dependence was assessed. No assessment was made of other drug abuse with or without dependence.

[k] People's Republic of China.

National Epidemiologic Survey on Alcohol and Related Conditions. The National Institute of Alcohol Abuse and Alcoholism's *National Epidemiologic Survey on Alcohol and Related Conditions* (NESARC) is a series of population-based surveys of alcohol use disorders and other substance use and mental disorders in the United States. Among its goals were to provide prevalence and incidence rates of alcohol-related disorders and associated disabilities; examine comorbidity patterns with other mental disorders, including other substance use disorders; document the natural history of alcohol-related disorders and disabilities; examine service use rates and patterns and barriers to care; and attempt to elucidate and differentiate "alcohol-induced" physical and mental disorders from "independent" disorders. The NESARC has been fielded three times. The first survey, with an unprecedented sample size of 43,093, was fielded from 2001 to 2002. This sample was nationally representative of the adult household population of the United States, also including noninstitutional group quarters such as boarding and rooming houses, nontransient hotels, shelters, worker housing, college quarters, and group homes. The second survey aimed to reinterview all first-wave respondents and was fielded from 2004 to 2005. A total of 34,653 interviews were completed, amounting to a response rate of 86.7%. The data from the second wave of the survey has allowed estimation of incidence rates and risk factors for the development of alcohol, drug, and other mental disorders, as well as examination of the course and outcome of treated and untreated mental and substance use disorders. One-year prevalence rates from the first NESARC survey are listed in Table 5.1–7. One-year incidence rates based on the first two surveys are listed in Table 5.1–11. NESARC-III was fielded from 2012 to 2013. The target population was the U.S. civilian, noninstitutionalized population ages 18 and over, with an oversampling of Black, Hispanic, and Asian populations. The total sample size was 36,309, representing an overall response rate of 60.1%. Genetic data from the NESARC-III that were collected from about 23,000 survey participants are available for researchers.

Table 5.1–11.

12-Month Incidence of DSM-IV Psychiatric Disorders, NESARC Wave 2

DSM-IV Disorder	Unweighted Number of Respondents at Risk	Unweighted Number of 12-Month Incident Cases	Weighted Annual Incidence per 100 Person-Years
Alcohol use disorder[a]	23,713	363	1.66 (0.11)[b]
Alcohol abuse	24,291	233	1.03 (0.10)
Alcohol dependence	30,220	481	1.72 (0.09)
Any drug use disorder[a]	30,830	89	0.31 (0.04)
Any drug abuse	30,887	75	0.28 (0.04)
Any drug dependence	33,639	101	0.32 (0.04)
Any mood disorder[a]	26,867	666	2.24 (0.12)
Major depressive disorder	28,859	489	1.52 (0.09)
Bipolar I	32,990	189	0.53 (0.05)
Bipolar II	34,045	74	0.21 (0.03)
Any anxiety disorder[a]	26,179	452	1.58 (0.09)
Any panic disorder	32,302	224	0.62 (0.05)
Social phobia	32,352	112	0.32 (0.04)
Specific phobia	29,502	147	0.44 (0.05)
Generalized anxiety	32,361	390	1.13 (0.08)

[a] Because the population at risk for each aggregate disorder category excludes all respondents with prior lifetime histories of any disorder in that category, the

denominator of the incidence rate is invariably smaller than the denominator for any component disorder. For the same reason, and because individuals can become incident cases of more than one disorder in each category in the same time period, the numerator for the aggregate rate is not a simple sum of the number of incident cases across the specific disorders.

[b]Numbers in parentheses are standard errors.

Adapted from Grant BF, Goldstein RB, Chou SP, et al. Sociodemographic and psychopathologic predictors of first incidence of DSM-IV substance use, mood, and anxiety disorders: results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *Mol Psychiatry.* 2009;14(11):1051–1066, with permission of Springer Nature.

Surveys of Mental Disorders in Children and Adolescents

Concerted research efforts on child and adolescent mental disorders started later than on adult disorders, and the knowledge base is correspondingly smaller for epidemiologic and treatment studies alike. The reasons for this are multifaceted, but as noted above, mental disorders in these populations were considered to be relatively rare until the ECA and other studies in the 1980s revealed much younger ages of onset for common mental disorders than was previously assumed. Diagnostic criteria for children and adolescents were widely criticized by clinicians and researchers for being difficult to operationalize and not reflecting developmental considerations and clinical reality, deficiencies that the DSM-5 attempted to remedy. In addition, the paucity of treatment services for children and adolescents hid what was later found to be a high prevalence of untreated mental disorders. However, interest in the diagnosis and the epidemiology of child and adolescent mental disorders has increased since the fielding of the ECA program. In fact, researchers have been able to take advantage of, and contribute to, advances in diagnostic criteria and research methods since the DSM-III and the ECA by developing sophisticated diagnostic instruments and study designs. Unfortunately, no global epidemiologic studies have been fielded in child and adolescent populations. However, several national surveys and datasets include information on children’s mental health as described below.

Methods for the Epidemiology of Child and Adolescent Mental Disorders. In the wake of the successes of the ECA program, the need to stimulate the field of child and adolescent psychiatric epidemiology with a “child ECA” was recognized as a crucial next step by the NIMH. Also recognized, however, was that diagnosing children and adolescents in community settings would require different methods than were used for adults in the ECA. It was assumed that depending on the developmental level of the children surveyed, a parent or primary caregiver might be needed as an informant on diagnostic and service use variables, and other informants such as teachers might give valuable additional information. A major issue to be clarified was how to gather this information in a large-scale survey and then how to best combine these multiple-source reports to arrive at reliable and valid diagnoses. The reliability and validity of the available diagnostic instruments also needed to be confirmed. Developing and evaluating assessments of disability and service use also needed to be addressed. NIMH set up a two-phase program, the first phase of which was a methodologic evaluation addressing the issues named above, and the second phase a full multisite survey, the “child ECA.” The MECA survey was a cooperative agreement between the NIMH and four academically based sites in New York City, New Haven, Atlanta, and San Juan, which was funded to address these methodologic issues in child and adolescent psychiatric epidemiology. The MECA surveys included children ages 9 to 17 years. The NIMH DISC, version 2.3, was used as the core diagnostic assessment instrument for DSM-III-R disorders, and an instrument to assess service use, sociodemographic factors, and potential risk and protective factors was developed. The CGAS and the Columbia Impairment Scale were included in the survey to assess disability. Only parent and child reports were obtained, mainly due to a lack of funds to obtain other informant reports. Data collection was completed in 1992, with 1,285 parent–child pairs being interviewed. Reliability and validity tests were set up to assess the DISC’s performance with both clinicians and lay interviewers. The DISC 2.3 was found to have acceptable reliability and criterion validity, and several reports on methodology, diagnosis, disability, and service use were published. Despite the successes of the MECA survey, the anticipated “child ECA” was never fielded. The results of the longitudinal Great Smoky Mountains Study have also provided extremely valuable descriptive and analytic data in a sample that has been followed longitudinally, but this study’s sampling was limited to a geographical area in North Carolina.

National Comorbidity Survey Replication Adolescent Supplement. The National Comorbidity Survey Replication Adolescent Supplement (NCS-A) was a supplemental survey to the NCS-R, conducted from 2001 to 2004. As such, it sampled youth ages 13 to 17 years from the households selected for the NCS-R and from a representative sample of schools located in the counties where the household surveys were conducted. The final sample size from households was 904, with an 86.8% response rate. From the 320 schools sampled, 9,244 youth were selected, with an 82.6% response rate. The youth were administered a modified version of the CIDI that had alterations in phrasing and examples to make it more appropriate to the respondents' language level and life experiences. Some disorders were dropped from the CIDI for this population, such as pathologic gambling, obsessive-compulsive disorder, and psychotic disorders. A parent or adult surrogate was given a self-administered assessment but not directly interviewed. This assessment was relatively brief, covering the youth's symptoms in categories for which adult reports of psychopathology were felt to be particularly important for making a diagnosis (major depressive episode, dysthymia, separation anxiety disorder, attention-deficit/hyperactivity disorder, conduct disorder, and oppositional-defiant disorder). The adult also filled out the Strengths and Difficulties Questionnaire. NCS-A data efforts ceased in 2002. Publications using NCS-A data have been slowly accumulating since the initial publication of findings in 2009. The NCS-A data showed that the 1-year prevalence of any mental, behavioral, or substance use disorder in the adolescent population was 40.2%, with estimates ranging from 8.3% for any substance use disorder to 24.9 for any anxiety disorder.

Youth Risk Behavior Surveillance System (YRBSS). The YRBSS is conducted to identify leading causes of death, disability, and social problems among youth and adults in the United States. The survey is conducted among 9th through 12th-grade students and asks questions about behaviors that contribute to unintentional injury and violence, substance use, and sexual behaviors related to unintended pregnancy and sexually transmitted infections. The survey has collected responses from more than 4.9 million high school students since 1991. Although this survey does not specifically examine mental disorders, analysis of co-occurring health behaviors such as violence, substance use, and unsafe sexual behaviors provide information on risk factors associated with mental disorders. The YRBS questionnaire is 87-item long and is administered every 2 years.

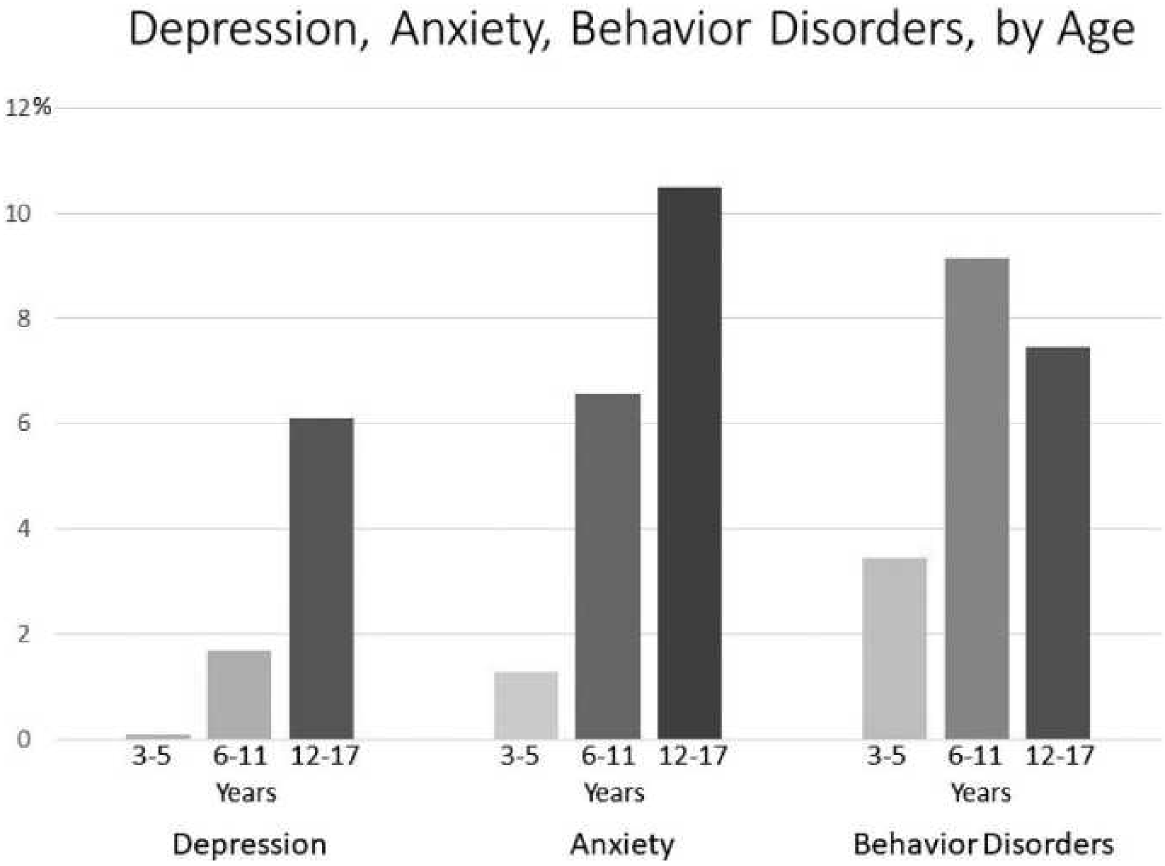

FIGURE 5.1–1. Percentage of children with mental health conditions per age group and disorder categories. (Based on Reprinted from Centers for Disease Control and Prevention. Data and Statistics on Children's Mental Health. https://www.cdc.gov/childrensmentalhealth/data.html.)

The National Survey of Children's Health. The *National Survey of Children's Health* NSCH survey provides data on children's physical and mental health along with access to care, education, and psychosocial context. The survey contains separate questionnaires for the age groups 0 to 5, 6 to 11, and 12 to 17 years and asks about the child's history of mental disorders.

Overall, aggregated data from children's mental health show that anxiety and depression are more common among older subgroups, while behavior problems are more common among the 6 to 11 years age group compared to the other two subgroups, as shown in Figure 5.1–1.

In addition to the surveys mentioned above, there are other important national health surveys that include a mental distress measure such as the Behavioral Risk Factor Surveillance System (BRFSS), which collects information on health risk behaviors, preventative practices, and healthcare access. The *National Health and Nutrition Examination Survey* (NHANES) collected data on depression, anxiety, concerns associated with mental health and substance abuse, and mental health service use and need. The *National Health Interview Survey* (NHIS) collects data on both adult and children's mental health and mental disorders. The *National Survey of the Diagnosis and Treatment of ADHD and Tourette Syndrome* (NS-DATA) collects information about children 2 to 15 years old in 2011 to 2012 who had ever been diagnosed with attention deficit hyperactivity disorder and/or Tourette syndrome to understand diagnostic practices, level of impairment, and treatments for this group of children. More information on these surveys can be accessed through the Centers for Disease Control and Prevention page: https://www.cdc.gov/mentalhealth/data_publications/index.htm

WHO Global Burden of Disease Study

The innovative *Global Burden of Disease* (GBD) studies are perhaps the best examples of the use of population-based epidemiologic surveys to serve the needs of public health policy and planning. There have been six GBD studies, the first published in 1996 and sponsored by the WHO and the World Bank, and subsequent ones (GBD 2010, 2013, 2015, 2017, and 2019) sponsored by the Bill and Melinda Gates Foundation. The GBD studies recognized that the consequences of a disease include not only death but also disability, and the impact of disability should be quantified and factored into any estimates of disease burden. By taking disability into account, these studies served to draw attention to chronic diseases, including mental and substance use disorders, which may ultimately cause premature death but often cause significant societal burden through reduced role functioning. *Disability-adjusted life years* (DALYs) were used to estimate the burden of disease. DALYs represent the sum of years of life lost to premature death (YLL) and years lived with a disability of a specified severity and duration due to the condition (YLD). One DALY represents 1 year of healthy life lost. GBD 1990 showed that five of the top ten causes of disability fell under the neuropsychiatric disorder category–a grouping that included mental and substance use disorders with neurologic disorders. GBD 2010 was a more ambitious project compared to the first study, expanding the number of health conditions, including mental and substance use disorders. Compared to the 1990 study, the 2010 study added six more categories: anxiety disorders, eating disorders, pervasive developmental disorders, childhood behavior disorders, dysthymia, and bipolar disorder subtypes. Substance use disorder was expanded to include burdens for alcohol use disorders and illicit drug use disorders. GBD 2010 also added new computations that accounted for the burden of mental and substance use disorders as risk factors for other disorders. Four risk factors were quantified: major depressive disorder as a risk factor for ischemic heart disease; cannabis use as a risk factor for schizophrenia; major depressive disorder, bipolar disorder, anxiety disorder, anorexia nervosa, schizophrenia, cocaine dependence, amphetamine dependence, and opioid dependence as risk factors for suicide; and injection drug use as a risk factor for hepatitis B and C and HIV.

Table 5.1–12.
Proportion of YLDs, YLLs, and DALYs Explained by the 10 Leading Causes of Total Burden in 2019

	Proportion of Total DALYs (95% UI)	Proportion of Total YLDs (95% UI)	Proportion of Total YLLs (95% UI)
Cardiovascular diseases	15.5% (14.1–16.9)	4.0% (3.4–4.5)	21.4% (20.0—22.6)
Neoplasms	9.9% (9.0-11.0)	0.9% (0.8–1.0)	14.5% (13.6–15.3)
Maternal and neonatal disorders	7.8% (6.8–8.9)	2.3% (1.8–3.0)	10.6% (9.7–11.7)
Respiratory infections and tuberculosis	6.0% (5.5–6.6)	1.5% (1.3–1.8)	8.3% (7.8–8.9)
Other noncommunicable diseases	6.0% (5.3–6.9)	10.1% (8.6–12.1)	3.9% (3.4–4.5)
Musculoskeletal disorders	5.9% (4.6–7.3)	17.1% (14.9–19.8)	0.2% (0.1–0.2)
Mental disorders	4.9% (3.9–6.1)	14.6% (12.2–16.8)	0.0% (0.0–0.0)
Diabetes and kidney diseases	4.4% (4.2–4.7)	5.3% (4.7–5.9)	4.0% (3.8–4.2)
Chronic respiratory diseases	4.1% (3.7–4.4)	3.8% (3.2–4.4)	4.2% (3.8–4.6)
Unintentional injuries	4.1% (3.7–4.4)	4.4% (3.8–5.1)	3.9% (3.4–4.3)

Institute for Health Metrics Evaluation. Used with permission. All rights reserved. http://ghdx.healthdata.org/gbd-results-tool.

The most recent GBD study (GBD 2019) analyzed 286 causes of death, 369 diseases, and injuries in 204 countries and territories. To calculate the global burden of mental and substance use disorders, the study investigators performed extensive literature searches for data on prevalence, incidence, remission and duration, and all-cause excess mortality. Publications were required to use DSM or ICD diagnostic criteria, to employ representative sampling of a community, and to provide sufficient information on study methods and sample characteristics. Prevalence was required to be reported in a current (point or past month) or past-year time frame.

Table 5.1–13. Proportion of YLDs (A), YLLs (B), and DALYs (C) Explained by Each Mental and Substance Use Disorder Group in 2019

	YLDs (95% UI)	YLLs (95% UI)	DALYs (95% UI)
Depressive disorders	31.6% (22.2–43.0)		29.2% (20.5–39.8)
Anxiety disorders	19.3% (13.4–26.5)		17.9% (12.4–24.5)
Drug use disorders	8.2% (5.7–10.9)	48.3% (46.5–50.8)	11.3% (8.9–13.7)
Alcohol use disorders	7.2% (4.9–10.3)	51.5% (44.2–55.0)	10.6% (8.4–13.4)
Schizophrenia	10.2% (7.4–13.0)		9.4% (6.9–12.0)
Other mental disorders	5.8% (3.8–8.8)		5.4% (3.5–8.2)
Bipolar disorder	5.7% (3.5–8.8)		5.3% (3.2–8.1)
Conduct disorder	3.3% (1.8–5.1)		3.0% (1.7–4.8)
Idiopathic developmental intellectual disability	3.0% (1.6–4.9)		2.7% (1.4–4.5)
Autism spectrum disorders	2.9% (1.9–4.2)		2.7% (1.8–3.9)
Eating disorders	2.0% (1.2–2.9)	0.1% (0.1–0.2)	1.8% (1.1–2.7)
Attention-deficit/hyperactivity disorder	0.7% (0.4–1.2)		0.6% (0.4–1.1)

Data are % (95% UI). DALYs, disability-adjusted life years; YLDs, years lived with a disability; YLLs, years of life lost.

Institute for Health Metrics Evaluation. Used with permission. All rights reserved. http://ghdx.healthdata.org/gbd-results-tool.

Tables 5.1–12 and 5.1–13 show some of the main results of GBD 2019. Cardiovascular diseases, neoplasms, and maternal and neonatal disorders were the major causes of disease burden worldwide. However, mental disorders were a major source of burden as well, accounting for roughly 5% of the total disease burden. Looking solely at mental disorders, the leading sources of DALYs were depressive disorders, followed by anxiety disorders, drug and alcohol use disorders, schizophrenia, and other mental disorders. This pattern was also seen for YLDs. Alcohol and drug use disorders accounted for the majority of YLLs; data were sufficient to attribute only three disorders as a cause of death. Countries with the highest DALY rates for mental disorders were Western Europe, sub-Saharan Africa, North Africa, and the Middle East. Locations in Asia were among those with the lowest DALY rates. Not surprisingly, the largest number of acceptable data sources were from high-income regions indicating the need for reviewing the cross-cultural applicability of case definition and data collection methods. Overall, the global distribution and burden of mental disorders have increased steadily since 1990, except for alcohol use disorder (Fig. 5.1–2).

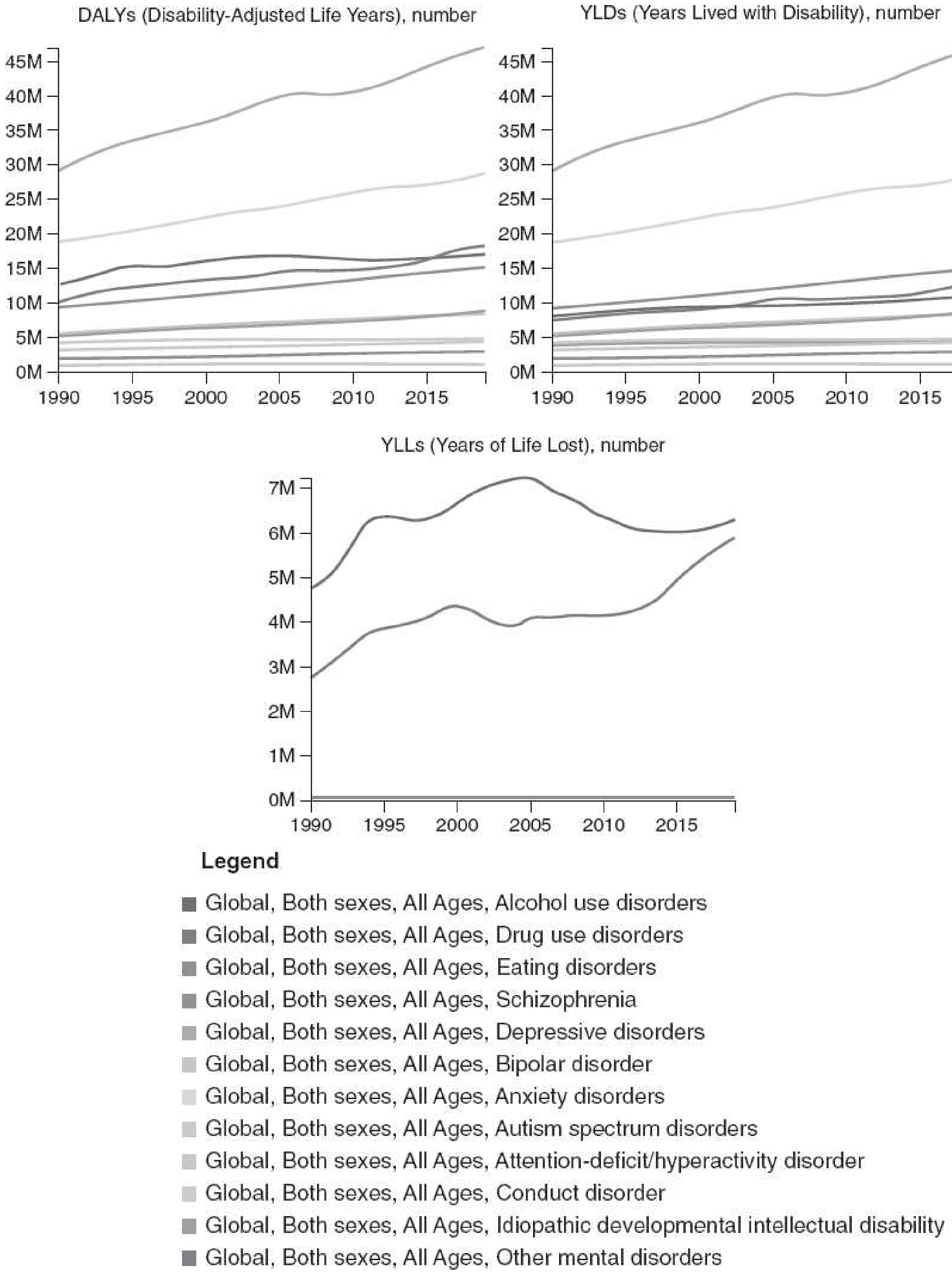

FIGURE 5.1–2. Annual trends of DALYs, YLDs, and YLLs from 1990 through 2016. (See eBook for color image.) (Based on Global Burden of Disease Collaborative Network. *Global Burden of Disease Study 2019 (GBD 2019) Results*. Institute for Health Metrics Evaluation. Used with permission. All rights reserved. http://ghdx.healthdata.org/gbd-results-tool.)

EMERGING ISSUES

Coronavirus Disease-2019 Pandemic (COVID-19)

On March 11, 2020, the World Health Organization declared COVID-19 a global pandemic, leading many countries to close their borders, forcing quarantine, and imposing strict measures (social distancing) to curb the spread of the deadly virus. These unprecedented measures led to disruptions in daily functioning, physical activity, sleep, as well as mental health. In a meta-analysis of published literature, there was a higher reported prevalence of suicide during the pandemic. The study observed that the main risk for suicidal ideations for frontline medical workers included low social support, high physical and mental exhaustion, and poorer self-reported physical health, while for the general population, the risk factors included sleep disturbances, exhaustion, loneliness, and mental health difficulties. In other literature reviews, there was a substantial increase in the prevalence and burden of major depressive disorder and anxiety disorders as well as worsening of eating disorder psychopathology as a result of the COVID-19 pandemic. Other studies reported that burnout, post-traumatic stress disorder, and violence have increased with the pandemic. A study conducted on over 8,000 families between March 2020 and March 2021 in the United Kingdom showed a clear increase in parent-reported symptoms of behavioral and attentional difficulties among children and adolescents during the time of lockdown. In a sample of children and youth, it was found that exposure to greater pandemic-related stressors led to higher levels of internalizing and externalizing psychopathology. This effect was especially prominent at the time of the lockdown through 6 months after. However, having a structured daily routine, low passive screen time use, low pandemic-news consumption, spending more time in nature as well as a healthy sleep routine were associated with better mental health outcomes.

These reports are preliminary since a longer period of time is needed to collect more data and observe the trends concerning the COVID-19 pandemic and its impact on mental health.

Further Readings

Alegria M, Takeuchi D, Canino G, et al. Considering context, place and culture: the National Latino and Asian American Study. *Int J Methods Psychiatr Res*. 2004;13:208–

220.

American Psychiatric Association Publishing. The structured clinical interview for DSM-5®. Accessed November 12, 2021. https://www.appi.org/products/structured-clinical-interview-for-dsm-5-scid-5

Borges G, Aguilar-Gaxiola S, Andrade L, et al. Twelve-month mental health service use in six countries of the Americas: a regional report from the World Mental Health Surveys. *Epidemiol Psychiatr Sci*. 2019;29:e53.

Center for Behavioral Health Statistics and Quality. *2020 National Survey on Drug Use and Health (NSDUH): Methodological Summary and Definitions*. Substance Abuse and Mental Health Services Administration; 2021. Accessed November 1, 2021. https://www.samhsa.gov/data/

Centers for Disease Control and Prevention. Data and Statistics on Children's Mental Health. Accessed November 1, 2021. https://www.cdc.gov/childrensmentalhealth/data.html#ref

Centers for Disease Control and Prevention. Data and Statistics. Accessed November 12, 2021. https://www.cdc.gov/mentalhealth/data_publications/index.htm

Centers for Disease Control and Prevention. Youth Risk Behavior Surveillance System (YRBSS). Accessed November 1, 2021. https://www.cdc.gov/healthyyouth/data/yrbs/index.htm

First MB, Gaebel W, Maj M, et al. An organization- and category-level comparison of diagnostic requirements for mental disorders in ICD-11 and DSM-5. *World Psychiatry*. 2021;20(1):34–51.

Galea S, Merchant RM, Lurie N. The mental health consequences of COVID-19 and physical distancing: the need for prevention and early intervention. *JAMA Intern Med*. 2020;180(6):817–818.

Global Burden of Disease Collaborative Network. *Global Burden of Disease Study 2019 (GBD 2019) Results*. Institute for Health Metrics and Evaluation (IHME); 2020. Accessed November 1, 2021. http://ghdx.healthdata.org/gbd-results-tool

Graft BF. The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). 2019. Accessed November 1, 2021. https://pdfs.semanticscholar.org/defc/ebd4704a293c7c49a7715c1b746132936fc7.pdf

Jackson JS, Torres M, Caldwell CH, et al. The national survey of american life: a study of racial, ethnic and cultural influences on mental disorders and mental health. *Int J Methods Psychiatr Res*. 2004;13:196–207.

*Kessler RC, Avenevoli S, Costello EJ, et al. Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders in the National Comorbidity Survey Replication Adolescent Supplement. *Arch Gen Psychiatry*. 2012;69:372–380.

Khaled SM, Petcu C, Bader L, et al. Conducting a state-of-the-art mental health survey in a traditional setting: challenges and lessons from piloting the World Mental Health Survey in Qatar. *Int J Methods Psychiatr Res*. 2021;30(3):e1885.

Knudsen AKS, Stene-Larsen K, Gustavson K, et al. Prevalence of mental disorders, suicidal ideation and suicides in the general population before and during the COVID-19 pandemic in Norway: a population-based repeated cross-sectional analysis. *Lancet Reg Health Eur*. 2021;4:100071.

Lawrence HR, Burke TA, Sheehan AE, et al. Prevalence and correlates of suicidal ideation and suicide attempts in preadolescent children: a US population-based study.

Transl Psychiatry. 2021;11(1):489.

National Institute on Alcohol Abuse and Alcoholism. NESARC-III genetic data now available to researchers. Accessed November 1, 2021. https://www.niaaa.nih.gov/news-events/announcement/nesarc-iii-genetic-data-now-available-researchers

Olfson M, Blanco C, Wall MM, Liu SM, Grant BF. Treatment of common mental disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions-III. *J Clin Psychiatry*. 2019;80(3):18m12532.

Rosen ML, Rodman AM, Kasparek SW, et al. Promoting youth mental health during the COVID-19 pandemic: a longitudinal study. *PLoS One*. 2021;16(8):e0255294.

Simpson TL, Rise P, Browne KC, Lehavot K, Kaysen D. Clinical presentations, social functioning, and treatment receipt among individuals with comorbid lifetime PTSD and alcohol use disorders versus drug use disorders: findings from NESARC-III. *Addiction*. 2019;114(6):983–993.

Substance Abuse and Mental Health Services Administration. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2019 National Survey on Drug Use and Health (HHS Publication No. PEP20-07-01-001, NSDUH Series H-55)*. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2020. https://www.samhsa.gov/data/

World Health Organization. Disease Burden, 2000–2019. Accessed November 1, 2021. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/global-health-estimates-leading-causes-of-dalys

World Health Organization. ICD–11 Fact Sheet. Accessed November 1, 2021. https://icd.who.int/en/docs/icd11factsheet_en.pdf

World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 –11 March 2020. Accessed November 1, 2021. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020

▲ 5.2 Statistics and Experimental Design

EUGENE LASKA, PH.D., CAROLE E. SIEGEL, PH.D., AND ISAAC GALATZER-LEVY, PH.D.

Jerry Cornfield declared that statistics is the bedfellow of the sciences in his presidential address to the American Statistical Association in 1974. Even a casual reader of articles that appear in the medical and psychiatric literature becomes aware of arcane terms such as chi-square and analysis of variance (ANOVA) and *P* values. The lexicon of statistics is now part of the common parlance of editors, reviewers, and authors, so to appraise and appreciate a scientific article, it is incumbent on the reader to have some familiarity with the lingua franca. Of course, it may be argued, the user of a computer need not understand the bits and bytes and operating system that drive the processor to use it, so why can't the reader simply accept that the referees and editors have done their jobs? Indeed, why can't the reader just peruse the abstract and the discussion section and be done with it? Well, truth be told, he or she can. A word processor can be used successfully even by one with limited computer skills. A scientific report can be skimmed to obtain the essence of its message without needing to appreciate the nuances of the methods utilized. But that strategy will not enable an independent view of the strength of the evidence nor of the credibility of the findings. The reputation of the author and the journal may provide some comfort, and perhaps in reports related only peripherally to the reader's interests, that is all that is required. But at least in one's own field of expertise, a full understanding of the techniques used to obtain the results and to reach the conclusions should be accessible. If the reader wants to truly understand the material, then at least a rudimentary understanding of statistical concepts is necessary.

So, what are the must-know elements of statistics? It is hard to say. It depends on the area of science or therapeutics that is of interest. The discipline of statistics has become so vast that few, if any, professionals can pretend to be an expert in all of its parts. But a psychiatrist should have at least a basic understanding of the rudiments. This chapter will not try to replicate what countless articles, books, and streaming-abbreviated introductory courses do. The basic ideas are described without any assumptions about the mathematical skill level of the reader in many elementary texts and chapters. It is assumed that the reader already has some familiarity with the basic concepts, but nevertheless, this chapter begins with a very brief review of some elementary statistical ideas. We also introduce some topics in machine learning. The remainder of the chapter discusses some of the many topics in statistics selected in the hope that they will increase the capability of a reader to make independent judgments and interpretations about the articles that he or she reads. Necessarily, the presentations are incomplete, the choice of topics is somewhat arbitrary, and there are many other topics that could have been included in this chapter. It is the authors' hope that this presentation will stimulate further study of statistical methods, particularly by individuals interested in psychiatric research. Although the authors have strived to keep technical details to a minimum, occasionally, some complexities are unavoidable.

BASIC IDEAS

In this section, the notions of elementary probability theory are reviewed, and the concepts of a random variable, measures of central tendency, and a measure of dispersion are introduced. To describe the joint behavior of two random variables, the ideas of statistical independence, covariance, and correlation are necessary to enable the introduction of a multivariate distribution.

Probability Theory

The mathematical foundation on which statistical methods are built is probability theory. This theory considers a situation in which the outcomes are random events. The set of all possible simple outcomes is called the *sample space*. For example, throwing a die once gives rise to a sample space containing six elements: 1, 2, 3, 4, 5, 6. A random variable is a mapping of the sample space onto the real line. It is usually denoted by an uppercase Latin letter such as X, and an actual outcome is usually denoted by its corresponding lowercase letter x. Thus, X is the random outcome of the face value resulting from the roll of the die, and $x = 6$ is an observed outcome. Corresponding to each *simple* outcome is a number greater than or equal to zero representing the probability of the outcome. It is defined as the long-run relative frequency of the event. For a fair die, the probability that it lands 6, written $\mathrm{P}(X = 6)$, is 1/6. The sum of the probabilities over all simple outcomes in the sample space is 1. The assignment of a probability to a simple outcome automatically determines the probability of all subsets of the sample space. For example, the probability that the outcome of the roll of the die is an odd number is $\mathrm{P}(X = 1) + \mathrm{P}(X = 3) + \mathrm{P}(X = 5) = 1/2$.

Probability Density Functions

Characterization of the random nature of an experiment is provided by the chance of observing a particular value x. If there are only a finite number of possible outcomes, the probability of each event x, $P(X = x)$ as a function of x, is called a probability density function (PDF). If the range of outcome values includes all values in an interval, the *continuous case*, the PDF, generically denoted by $\mathrm{f}(x)$, has a slightly different meaning. This is because the probability of observing any particular single value of x, possibly with a finite (countable) number of exceptions, is 0. In this case, the area under the density curve in the interval, the integral of $\mathrm{f}(x)$ over the interval, is equal to the probability that an outcome falls in that interval. The *cumulative distribution function*, frequently represented by $\mathrm{F}(x)$, is the probability that the random variable takes a value less than or equal to x. It is obtained in the discrete case by summing the probabilities of all possible outcomes that are less than or equal to x and in the continuous case by integrating the PDF from minus infinity to x. All probabilities must be nonnegative, and their sum over all possible events, or in the continuous case, the integral over the sample space, must be 1. Therefore, as x goes to infinity, $F(x)$ goes to 1. The frequentist interpretation of probability is that in the long run, the fraction of times that the outcome x is observed (or that x falls in a specific interval $[a,b]$ of the sample space) in N replications of the experiment tends to $f(x)$ (or to the integral of the PDF from a to b) as N tends to infinity.

Conditional Probability and Independence

Many compound events can be defined from the simple outcomes of an experiment. In the roll-of-the-die experiment, the event that an odd number occurs or the event that either a 1 or a 5 occurs are two examples. In some cases knowing that one event has occurred can impact the likelihood of another event occurring. For example, if the faces of the die are equally likely (all have a probability of 1 of 6), then the probability of a 1 or 5 occurring given that an odd number has occurred is 2 of 3, while the probability of a 1 or 5 occurring with no other event information is 1 of 3. The random variable X takes the value 1 if the die falls on face 1 or 5, and 0 otherwise, and the random variable Y takes the value 1 if the die falls on an odd-numbered face are said to be dependent. Two random variables X and Y are said to be independent if $P(X = x|Y = y) = P(X = x)$. The notation $P(X = x|Y = y)$ is read “the probability that $X = x$ given (or conditional on) $Y = y$.” It is the probability that the event $X = x$ occurs calculated in the outcome space defined by the event $Y = y$. Formally, $P(X = x|Y = y) = P(X = x \text{ and } Y = y)/P(Y = y)$. Also, two random variables are independent if the probability of the joint event $X = x$ and $Y = y$ is equal to the product of the probabilities of the two events, that is, $P(X = x \text{ and } Y = y) = P(X = x)P(Y = y)$. Independence and conditional probabilities play a major role in statistical thinking. For example, suppose treatment assignment and degree of improvement on a five-level clinical global scale (CGI) is reported in a table with five rows (CGI) and two columns (treatment or placebo). A chi-square statistic is used to analyze such contingency tables for testing the independence of two random variables that define the rows and columns. If the statistic is significant, the data suggest that the variables are dependent, and the distribution of CGI outcomes conditioned on a value of treatment assignment differs depending on the value of the conditioning variable. If treatment and placebo improvement rates differ, then one would expect a chi-square test to be significant because the distribution of CGI conditional on receiving the treatment is not the same as the distribution conditional on receiving a placebo.

Parametric Forms

In parametric analyses, the *form* of the PDF is assumed to be known with the exception of the value of one or more parameters, which, once specified, completely determines the distribution. The most commonly assumed parametric form is the so-called Gaussian or normal distribution. The shape of the normal PDF is the familiar bell-shaped curve, whose peak (mode), mean, and median occur at the same value. In fact, the curve merely translates to the right or left but retains the same form as the mean is respectively increased or decreased. Distributions with this property are called *translation parameter families,* and they have the mathematical form $f(x - \varphi)$, where φ is the mean of the random variable X. Greek letters, such as φ, are often used to represent parameters of the PDF to distinguish them from random variables. The logistic, beta, exponential, F, and t, for continuous random variables and the binomial, multinomial, and Poisson for discrete random variables are PDFs that frequently appear in the scientific literature.

Mean, Median, and Variance

A variety of measures that help characterize the nature of a random variable have been defined. Among the most important are the mean, median, and variance. The *mean* is the average of all possible outcomes, weighted by the probability of the outcome. The mean is thought of as a measure of the central tendency of the random variable. It is also called the *expected value*, and if X is the random variable, then it is usually denoted by $E(X)$. The value m is the median if the probability that the random variable X is less than m equals the probability that X is greater than m. The *variance* is the average of the square of the difference between each outcome x and the mean, weighted by the probability of the outcome. The variance, usually denoted by σ^2, and its square root, σ, the standard deviation, are thought of as measures of the spread of the observations about the mean.

Multivariate Random Variables

Most serious investigations of scientific phenomena cannot be described by the outcome of a single random variable. A multivariate probability distribution describes the chance of various outcomes of an experiment in which k events are observed on each "unit" in a sample. Typically in trials in mental health, a unit is a patient, and observations are made on one measure collected over time, many measures collected at one point in time, or many measures collected at multiple times. Two random variables, X_1 and X_2, may be independent or dependent. As discussed above, they are independent if and only if their joint (simultaneous) probability function, $f(x_1, x_2)$, is always equal to $f_1(x_1)$ times $f_2(x_2)$. The PDFs $f_1(x_1)$ and $f_2(x_2)$ are called the *marginal distributions* of X_1 and X_2.

Covariance and Correlation

One important measure of the relationship between dependent pairs of random variables X_1 and X_2 is the covariance, the average of the product of the difference between each outcome x_1 and its mean and the difference between each outcome x_2 and its mean, weighted by the probability of the joint outcome x_1, x_2. Closely related is the correlation, usually denoted by ρ, which is defined to be the covariance divided by the product of the standard deviation of X_1 and the standard deviation of X_2. The distribution most widely used in multivariate applications is, by far, the multivariate normal. If the mean and variance of each of the random variables and the covariance between all pairs of random variables are known, then the multivariate normal distribution is completely specified.

USING DATA TO ESTIMATE PARAMETERS

The scientific question under consideration must be represented by a statistical model, which at least approximately characterizes the set of outcomes that can be observed after the data collection is complete. If the model, represented in parametric form as the PDF f(x, φ), is fully known, there is little left to do. Neither an experiment nor statistical inference is needed. The probability distribution is the answer to the question: If the experiment is performed, what can be expected with what probability? But the true value of the distribution is rarely known, and the reason for the experiment is to collect data so that something can be said about the nature of the PDF and the value of the parameters that specify it. This section discusses the problem of parameter estimation. It begins with a description of the method of maximum likelihood. Some of the properties that are used to appraise the merits of an estimator are described. Confidence interval (CI) estimators and their interpretation are introduced.

Point Estimators

In many application areas, the general form of the parametric distribution that describes the collected data is well established. After an experiment is performed, the first question an investigator should attempt to answer is what values of the unknown parameters within the assumed PDF form best represent the data. This is the setting of the theory of statistical estimation, which has a rich and long history. Of course, the principal goal is to come as close to the true value of the parameter as possible. Estimation theory is concerned with the problem of how to obtain an estimator and appraise its worth, not in an individual case but over all possible outcomes, weighted by the probability of the outcome. Perhaps the premier approach, introduced by Sir Ronald Fisher, is the maximum likelihood estimation (MLE). The idea is both intuitive and simple. Suppose that the parameter φ is unknown and f(x, φ) is the probability of observing the event x. The likelihood is just f(x^*, φ), where x^* is the value of the observed outcome. Here x^* is fixed, and f is thought of as a function that depends only on φ. For any given φ, the likelihood is a measure of how "likely" the observation x^* is had the outcome been selected randomly from the probability distribution f with parameter φ. If the observed random variables are independent, then the likelihood is the product of the likelihoods of each observation. It is a measure of the probability of the joint occurrence of x_1^*, x_2^*, …, and the maximum likelihood theory posits that φ should be chosen so that the probability of the event that was actually observed is as large as possible. Any other choice of parameters would make the observed data set less likely to have been observed. For example, if a coin is tossed 20 times, landing heads 10 of the times, setting the probability of a head, P, equal to 10 of 20, or .5, is perfectly sensible. In fact, under the assumption that the distribution of the random variable (the number of heads) is binomial, it makes the chance of 10 heads in 20 tosses maximal. Setting it equal to a lower or higher number makes the estimator less consonant with the data. In a great many cases, the properties of the MLE are very desirable.

Properties of Estimators

Since an estimator is itself the result of a random experiment, it is a random variable, and so it has a mean or expected value. The theoretical calculation of its mean involves the use of the density f(x, φ), where the true value of φ is unknown. We refer to the estimator as *unbiased* if, for each value of the parameter φ, the mean of the estimator is φ. Such estimators exhibit a kind of impartiality that is usually desirable. Another property of an estimator concerns its behavior as the sample size of the experiment increases. Imagine a sequence of estimators in which the ith estimator is based on a sample of size i. For example, the ith estimator in the sequence could be the sample mean based on a sample of size i. An estimator is said to be *consistent* if for any true value φ the sequence converges to φ. Another property of an estimator concerns its precision, as measured by its variance. An estimator is said to be *efficient* if it is unbiased and its variance is as small as it can possibly be compared to any other unbiased estimator. In some situations, a lower bound for the variance of an estimator can be determined. In this circumstance, the *relative efficiency* of an estimator can be determined by comparing its variance to the theoretical minimum variance. Consistency suggests that as the sample size increases, the variance decreases, and the estimator approaches φ.

Confidence Interval Estimation

Even if an estimator has all of the most desirable properties, it is not likely that it will be exactly right. To leave a little wiggle room, an interval estimate, in addition to a point estimate, is desirable. These are termed CI estimators. The amount of confidence placed on the estimator, in the sense that the parameter being estimated lies within the interval, is called the *confidence level* or *confidence coefficient*. The usual form of a CI is $P[a \le \varphi \le b] = c$. If $c = .95$, this is read: $[a,b]$ is a 95% CI for the parameter φ or; $[a,b]$ is a CI for φ with confidence coefficient .95. Typically for a fixed sample size, as the confidence coefficient increases, so too does the length of the interval. An interval from minus infinity to plus infinity with confidence coefficient 1 always covers the unknown parameter, but it is completely uninformative.

Interpretation of a Confidence Interval

It is important to understand how to interpret a CI. From a frequentist point of view, the probability of an event is characterized by the frequency with which the event occurs in infinitely many replications of the experiment. The form of the interval before the experiment is conducted is usually given in terms of some not yet observed random variable. For a normal random variable with mean φ and variance one, the probability statement $P[\bar{X} - 1.96/\sqrt{N} \le \varphi \le \bar{X} + 1.96/\sqrt{N}] = 0.95$, where $\bar{X}$ is the arithmetic mean of N observations, has a clear and unambiguous meaning. Suppose that a random sample of $N = 100$ observations is repeatedly drawn from a population governed by a normal distribution with mean φ and variance 1. As the number of repeated draws becomes very large, regardless of the true value of φ, the proportion of times that the interval $\bar{X} - 1.96/\sqrt{N} \le \varphi \le \bar{X} + 1.96/\sqrt{N} = \bar{X} - 0.196 \le \varphi \le \bar{X} + .196$, $\bar{X}$ covers φ approaches .95. Once data are collected, however, the game has changed. The statement $P[0.304 \le \varphi \le 0.696]$ can only take the value 1 or 0, depending on whether or not the true value of φ falls in the interval. So, from a frequentist point of view, it is incorrect to think of the confidence coefficient of a CI estimate as the probability that the observed interval includes the true value of the parameter. Rather, it is a statement about the joint probability that the random variable that defines the lower limit falls below φ and the random variable that defines the upper limit of the interval falls above φ (before data are collected) for any true value of φ.

USING DATA TO TEST HYPOTHESES

A statistical hypothesis is a statement about the possible values of the parameters of a probability distribution. In this section, test statistics are introduced, and the potential errors that may arise in testing a statistical hypothesis are discussed. The roles of power and the ubiquitous *P* value in hypotheses testing are presented.

Hypothesis Testing

Sometimes interest is focused on answering a scientific question about a random variable with a (qualified) yes or no. Is drug 1 more effective than drug 2? In such cases, the goal often is to determine whether the mean of one distribution is larger than the mean of another. Many scientific questions can be framed in statistical terms as a hypothesis test about the parameters of a probability distribution. Suppose that a clinical trial is designed to test the null hypothesis that two treatments have the same mean effect. Formally, the treatments have true mean effects, denoted respectively by parameters φ_1 and φ_2, and our interest lies in testing whether the difference between the parameters of the two treatments $\varphi = \varphi_1 - \varphi_2$ is 0. The null hypothesis then is $Ho{:}\varphi = 0$. Many alternative hypotheses are possible. The two most common in this setting are the two-sided alternative $H_1{:}\varphi \neq 0$ and the one-sided alternative $H_1{:}\varphi > 0$ (or $H_1{:}\varphi < 0$). A test statistic must be chosen to test the null hypothesis against the alternative. Notably, different alternative hypotheses often lead to different rejection regions for the same test statistic. If the data from both treatments are assumed to follow a normal distribution, and both have the same albeit unknown variance, then the test statistic usually used is a student's *t* statistic, which has the form

$$t = (\overline{x}_1 - \overline{x}_2)\Big/\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Here, $\bar{x}_1$ and s_1, and $\bar{x}_2$ and s_2 are estimates of the mean and standard deviation based on samples of size n_1 and n_2 from the two respective populations. The t distribution depends on the expected value of $\bar{x}_1 - \bar{x}_2$, which is $= \varphi_1 - \varphi_2$, the variance of the random variable, and sample sizes, n_1 and n_2.

Type I and Type II Errors

The decision as to whether to accept or reject the null hypothesis is determined by the value of the test statistic, which, of course, is a random variable whose value is unknown before the data are analyzed. If the value of the statistic falls in a prespecified *rejection* or *critical* region *R*, the null hypothesis is rejected. Even if the null hypothesis is true, there is still a chance that the statistic will lie in *R*, in which case a type I error has occurred. That is, a true null hypothesis has been erroneously rejected. Alternatively, the test statistic may not fall in *R*, although H_o is not true, in which case a type II error has occurred. That is, a false hypothesis has not been rejected. Both kinds of errors must be considered in choosing *R*.

Usually, the choice is made so as to ensure that the probability of a type I error, usually denoted by the Greek letter α, is small. In most clinical trials, *R* is chosen so that α is .05, although in some applications choosing a type I error rate of .10 may be reasonable. This value is called the *level* of the test. For the comparison of two means, the calculation of this probability is performed under the assumption that $\varphi = 0$, a *point* or *simple* null hypothesis. If the alternative hypothesis is $H_1{:}\varphi \neq 0$, and the sample sizes are large, then the hypothesis is rejected if the *t* statistic is either less than –.168 or greater than 1.68. These are the *large sample critical values* of a *t* distribution with $n_1 + n_2 - 2$ degrees of freedom with α set equal to .05. This is the so-called two-tailed test. If the alternative hypothesis is $H_1{:}\varphi > 0$, the null hypothesis is rejected if the *t* statistic is greater than 1.96.

If there are only two parameter values under study, that is, two simple hypotheses, which one should be labeled the null and which the alternative? The null hypothesis should be chosen to be the one whose false rejection would be considered to be the more serious error. This is because the testing paradigm that selects α, the level of the test, controls the probability of this error.

Power of a Test

The probability of a type II error, or rather one minus this probability, is called the *power* of the test. Since the alternative hypothesis contains many values of φ, it is called a *composite* hypothesis—power is a function of the possible values of φ in H_1. In our example, power is an increasing function of φ and the sample size and a decreasing function of σ^2. During the design phase of a study, the sample size is chosen so that the power to reject a false null is relatively large, perhaps .8 or greater, for a value of φ that is sufficiently large to be clinically meaningful. The calculation is often performed under an assumed value of σ^2 that is based on data from other similar clinical trials. Suppose that the estimated value of φ turns out to be a number that is smaller than the clinically meaningful value chosen when the power calculation was performed and that the test statistic falls in the critical region. There is no difference in the statistical interpretation of the results of the trial. The hypothesis is still rejected at the designated level of the test.

P Values

Even a casual reader of the psychiatric literature will encounter a study in which *P* values are reported. Unfortunately, in some articles, only *P* values appear without estimates of the parameters themselves or the variance of the estimators. All too often, both the author and the reader misinterpret the *P* values. Does a very small *P* value mean that one of the treatments is very much better than the other?

A good place to start is with an explanation of the meaning of a *P* value. Before the analysis of data commences, the value of the test statistic is, of course, random, and in the example, the comparison of two treatments, it follows a *t* distribution. Suppose that the null hypothesis being tested is true, so the expected value of $\bar{x}_1 - \bar{x}_2 = 0$. In this case, the distribution of the statistic is completely specified. It follows a *t* distribution with $n_1 + n_2 - 2$ degrees of freedom. So, for every possible value of the test statistic, the chance of obtaining a value more extreme can be computed or found in a table. If t^* is the value of the test statistic observed in the trial, then the corresponding probability of observing a more extreme value is defined to be the *P* value. Alternatively, the *P* value may be regarded as the smallest level (had it been chosen in advance) for which the hypothesis would be rejected with the observed data. But a small *P* value and a true null hypothesis are inconsistent, for were the trial repeated over and over again, obtaining the observed value of the test statistic or a value more extreme would be very rare. This compellingly argues against the null hypothesis and suggests that it should be rejected. Scientific tradition or perhaps convention usually defines rare as an occurrence whose frequency is less than or equal to 1 time in 20. Note well that the fact that the null is rejected does not mean that the magnitude of the difference has any clinical or scientific relevance. The sample size could be so large that even small trivial differences achieve statistical but not practical significance. Also, the inference that a large *P* value implies that the means of the two treatments are equal, that is, proving the null cannot be made. A calculation of the power of the test would help to rule out the possibility that the sample size was not adequate to detect a meaningful difference. More helpful still would be a CI for the parameter whose upper limit falls below any values that would be scientifically or clinically meaningful. Finally, it is necessary to emphasize that in the frequentist tradition of statistical inference, the *P* value cannot be interpreted as the probability that the null hypothesis is true. The hypothesis is not a random event, so it is either true or false. In a subsequent section, the Bayesian approach will be introduced. In this framework, a method for forming the probability or perhaps the degree of belief in the truth of a null hypothesis modified by the collected data will be described.

Testing Multiple Hypotheses: The Family Wise Error Rate (FWER)

In the preceding section, testing a single hypothesis was considered. However, it is rare that there is only one outcome variable, one dose of a drug, or one homogeneous population that is of interest. As the costs of even simple clinical trials increase, it is prudent to consider simultaneous investigation of a variety of issues because gathering additional information may represent marginal additional expenditures. The result of such a strategy is many more hypotheses tests to answer the various questions that may be asked of the data. The collection of statistical hypotheses, each derived from some comparison of interest that relates to a common aspect of the research question, is referred to as a *family of hypotheses*. For example, three doses of a drug and a placebo may be simultaneously studied in a clinical trial. The doses are each to be compared to a placebo. The three contrasts, high dose versus placebo, middle dose versus placebo, and low dose versus placebo, are a family of hypotheses. *A priori,* it is possible that all of the null hypotheses in the family are true. Testing each of these at a specified level α ensures that the probability of a type I error of each contrast is α. But, it is prudent to ask, what is the probability that one or more of the tests erroneously rejects the corresponding null hypothesis? The probability of erroneously rejecting one or more of the hypotheses in the family is called the *familywise error rate*. Clearly, in general, this probability is much higher than α. Under the null hypothesis, the probability of not making a type I error when testing a single hypothesis at a .05 level is .95. Suppose that a family is comprised of m individual hypotheses. If the test statistics are independent, the probability that none of the hypotheses are erroneously rejected is $.95^m$. Therefore, the probability of the complementary set, that one or more of the m hypotheses is erroneously rejected is $1 - .95^m$. For $m = 3$, it is .14 and for $m = 5$, it is .23. The FWER approaches 1 as m increases.

In many research settings, it is appropriate to require that the statistical procedures utilized should not result in a large FWER. For example, if multiple doses of a putative drug are studied and the drug is ineffective, it would be unreasonable to allow the chance to be high so that one of the doses is found to be statistically different from placebo if the null hypothesis is true. In this section, statistical tests that control the FWER multiple hypotheses tested are discussed. Several procedures for controlling the size of the FWER are presented and their advantages and disadvantages are discussed.

The most widely known approach to controlling the FWER is the Bonferroni method, which is based on a probabilistic inequality. In fact, the inequality is known as *Boole's inequality* and as often happens, the credit for its utilization in simultaneous statistical inference may have been misplaced. If there are m hypotheses in the family to be tested and the type I error level of each test is set to α/m, then the probability of an FWER $\leq \alpha$. That is, each test is performed at a level that is much smaller, α/m, than α, which would be used if there were only one test. The Bonferroni procedure is conservative in that, while it controls the familywise error rate, it renders the chance of rejecting any one of the individual hypotheses considerably lower. As a consequence, the user of this approach pays a heavy price in terms of loss of power. For example, if the family has five hypotheses, then, in order to achieve a FWER of .05, the individual hypotheses must be tested at the $\alpha = .01$ level. If a t test is used for each of the individual hypotheses, then the value of the t statistic calculated after the experiment is complete must be larger than 2.457 in order to reject the corresponding null. If only a single hypothesis were being tested, then the t statistic would have to be larger than 1.697. This makes the task of rejecting any of the individual hypotheses considerably more difficult. On the other hand, the Bonferroni method can be used to control the FWER in any setting. It is easy to use and does not depend on the distribution of the test statistic. Several improvements to this method have been suggested. Sidak modified the Bonferroni approach providing only a slight improvement in power. If the desired FWER is desired to be α, then instead of using α/K, the Sidak cut point is defined to be $\alpha^* = 1 - (1 - \alpha)^m$, where m is the number of hypotheses tested.

In some studies, for example, one comparing K treatments, the investigator is interested in a contrast of each treatment with every other treatment. If say $K = 3$, and the τ_i is the parameter characterizing treatment i, then the family consists of three pairwise comparisons $(\tau_i - \tau_j)$. The first step in analyzing pairwise comparisons is often an analysis of variance on the K treatments. One of the first tests to control the FWER, the least significant difference (LSD) procedure was developed by Sir Ronald Fisher. He argued that if the analysis of variance rejected the null that all three treatments have equal means (the omnibus null hypothesis), then to test the pairwise contrasts, t tests should be computed between each pair. The estimated variance used in the denominator of the t statistic is based on all the data, not just on the data for the two treatments being compared. If there are 30 subjects per group, then a two-tailed t statistic rejects the null for each contrast if it exceeds 2.042. If the analysis of variance is not significant, then neither the omnibus null hypothesis nor any other null hypothesis about differences between means can be rejected. This procedure is valid in that the FWER is controlled for $K = 3$ but it is not valid for any $K >$

3. Its use for such values does not result in tests whose FWER is protected at the α level.

This approach utilizes a staged approach to testing a family of hypotheses. The omnibus test, in the above case an ANOVA, is performed first. Such tests have been called *gatekeeper* tests because they limit the overall FWER. Only after the omnibus test rejects the global null does further testing take place. These tests can be less demanding in terms of power than Bonferroni-like tests without affecting the FWER. The Newmann-Keuls procedure is another well-known example of such tests.

A variant of the LSD procedure is Tukey test for pairwise comparisons. It utilizes a *t*-like statistic called the *studentized range*. Most statistical software packages that compute ANOVAs include routines to compute this test. Tukey test compares each of the K(K − 1)/2 studentized *t* statistics with a single-tabled critical value. If the sample sizes of treatments are the same and each of the pairwise comparisons is to be tested, then many believe that Tukey method is the method of choice. It exactly achieves the desired FWER and has reasonable power. The theory for this test was developed under the assumption that the sample sizes for each treatment are equal. Various corrections have been proposed to handle the unequal sample size case. For example, the Tukey–Kramer test replaces the common sample size in the formula with the harmonic mean of the respective sample sizes, that is, $(1/n_i + 1/n_j)/2$. Though this modification is widely used, the FWER fails to attain its proscribed level in most cases.

Although a clinical trial may involve K treatments, not all pairwise comparisons may be of interest. For example, K-1 new treatments may be in a trial together with one standard treatment. A new medication for an illness may only be worth further study if it is superior to the standard medication, frequently called the control. Here the comparisons of primary interest are each new treatment against the control. Thus, there are only K-1 comparisons to consider. Dunnett derived a test statistic for this problem that is similar to a *t* statistic. However, in this test, the estimate of variance used in the formula is derived from the data from all of the K treatments. The critical value is based on K and the *degrees of freedom*, which is the total number of subjects on all medications minus K. For a given FWER α the critical value may be obtained from published tables. Dunnett test achieves the exact FWER.

The procedures discussed above for multiple comparisons are all specified in advance, but the order in which testing is performed is irrelevant. *Adaptive* or *sequential procedures* modify the process of testing depending on the outcome of earlier tests. They can provide an increase in power while still controlling the FWER. The Holm procedure for testing a family of m hypotheses begins by testing each hypothesis individually using a statistic such as a student's t. Each test produces a P value without multiplicity adjustment, so the experiment yields $p_1, p_2, \ldots, p_m$. These values are then ordered from the smallest to the largest, $p(1) \leq p(2) \leq \ldots. \leq p(m)$. The notation $p(i)$ refers to the ith smallest P value. Relabel the m hypotheses so that H(i) is the hypothesis associated with $p(i)$. If $p(1) > \alpha/m$, H(1) is not rejected, the process terminates and none of the m null hypotheses are rejected. If $p(1) \leq \alpha/m$, H(1) is rejected and the process proceeds to examine $p(2)$. If $p(2) \leq \alpha/(m - 1)$, H(2) is rejected and the process proceeds to examine $p(3)$. Otherwise testing is stopped and H(2), H(3), … H(K) are not rejected. The procedure continues in this manner until a hypothesis is not rejected or there are no further hypotheses to tests. The Holm procedure is more powerful than the Bonferroni test, which requires that each $p(\mathrm{i}) \leq \alpha/m$ as opposed to $p(\mathrm{i}) < \alpha/(m - i + 1)$. The Holm procedure is called a *step-down procedure* because the cut points are in descending order. Note that the p_i are put in ascending order, which sometimes results in confusion as to what the test actually is.

A dual of this step-down testing procedure is the *step-up procedure*, introduced by Hochberg. Here the quantities $p_1, p_2, \ldots, p_\mathrm{m}$ are reordered in descending order $p[1] \geq p[2] \geq \ldots. \geq p[m]$ where $p[i]$ is the ith largest among the $p_1, p_2, \ldots, p_\mathrm{K}$. The hypotheses are then relabeled so that H[i] is the hypotheses associated with $p[i]$. If $p[1] \leq \alpha$, H[1] is rejected as are all of the remaining hypotheses. If *p*[1] > α then H[1] is not rejected and *p*[2] is examined albeit against a modified threshold. If *p*[2] meets the condition, H[2] and the remaining hypotheses are rejected. Otherwise p[3] is examined. In brief the step-up procedure accepts hypotheses in order and at the first nonacceptance rejects all remaining hypotheses.

A *closed testing procedure* may be used to control the FWER on a "closed" family, and the method has some remarkably desirable properties. The procedure sequentially tests members of the family of hypotheses, yet each test is performed at size α. The family of hypotheses is said to be closed under intersection if the family satisfies the condition that if *H* and *G* are any two hypotheses in the family, the hypotheses represented by *H and G* are also members of the family. The sequential nature of the closed testing procedure requires that prior to rejecting any hypotheses, *H*, one must reject all hypotheses that include *H*. For example, suppose the family contains three hypotheses, *G*, *H*, and *J*. Then closure under intersection requires that it must also contain *G and H*, *G and J*, *H and J*, as well as *G and H and J*. To test *H* at level α, one must first test *G* and *H* and *J*, then test *G* and *H* and test *H* and *J*, and finally test *H*. Each of these tests is carried out at level α, and if any test does not reject, then the process ends. The FWER of the closed testing procedure is α.

This procedure has the very desirable property that the first hypothesis is tested at .05 and the very undesirable property that if any test fails along the way, the process stops. It might very well be that one or more of the hypotheses that are not reached would have been found significant had an alternative procedure for protecting the FWER been used. However, once the choice is made, hopefully before the data are examined, there is no going back. Further examination of *P* values is of course legitimate and desirable, but they no longer represent probabilities of a type I error under the null hypothesis.

Comparing a combination medication to each of its component medications is another example of a testing procedure where each test is performed at the α level and yet the FWER is α. Let φ_i denote the mean difference in efficacy between the combination and medication *i*. Federal regulations require that prior to licensing a combination, one must show that each component medication contributes to the claimed effects. One translation of this requirement is that φ_i is positive for every component medication. This leads to the null hypotheses that for at least one component $\varphi_i \leq 0$. If the outcome variable follows a normal distribution and φ_i is the test statistic for $\varphi_i \leq 0$, then the test statistic for the null hypotheses is min $(\varphi_1, \varphi_2, \ldots, \varphi_K)$. Each φ_i is tested at the same α level. This procedure is referred to as the *min test*. It is a simple example of what is known as the union–intersection principle.

Each of these approaches to controlling the FWER has positive and negative properties. Especially if the test statistics are not independent, the Bonferroni method becomes increasingly conservative as the number of hypotheses increases. The sequential procedures are usually more powerful than the other methods described above. However, when the number of false hypotheses in the family is large, they can be problematic. Tests based on the union–intersection relationship also tend to be quite conservative, and they are dependent on the existence of a satisfactory omnibus statistic.

False Discovery Rate

The shortcomings of the FWER methods are never more apparent than when the number of statements in the family of hypotheses is very large. Imaging data presents a prime example. The number of potential statements is enormous, reaching even into the tens of thousands. Consider, for example, the question of whether a given voxel has been activated by a particular stimulus, or whether in a microarray analysis particular genes are differentially expressed. The statistics related to a pair of adjacent voxels will undoubtedly be statistically dependent, and the compound null hypothesis may be partially or even completely false. Under such circumstances, the use of the methods used to control the FWER for simultaneous testing may be useless. Some have instead conceptualized the problem as one of separating the important effects in the image from those that are small or unimportant. The false discovery rate (FDR) is a relatively new approach to the problem of multiplicity. The FDR controls the expected *proportion* of false positives among the declared rejected hypotheses. Just as in FWER methods, each voxel is tested resulting in a *P* value. On the basis of the distribution of the observed *P* values, a threshold is determined. The hypothesis corresponding to those *P* values falling below the threshold is rejected. However, the interpretation of the *P* values is quite different from the approach described above. If 100 genes with a maximum FDR of .20 are declared by the procedure to be differentially expressed, then the expected value of the maximum number of false positives is 20. There is no comparable interpretation that attaches to *P* values. There are four major factors that affect the characteristics of a study utilizing an FDR criteria: (1) the proportion of null hypotheses that are false, (2) the distribution of the variables being assessed, (3) the measurement variability, and (4) the sample size. The investigator has limited ability to manipulate any of these except for the size of the study. The FDR approach provides much weaker control of the type I errors in the test procedure than do the traditional methods described above. This is because the FDR criteria for false positive outcomes (the expected *proportion* of false positives is less than α) are much weaker than the FWER criteria (the probability of one or more false rejections is less than α). Thus, FDR methods are said to have weak control of the type I error.

THE BAYESIAN APPROACH

The formal setting for incorporating prior beliefs into a statistical analysis in which the data modify such beliefs is the essence of the method presented below.

Bayes Theorem

One of the most important results that emanates from elementary probability theory is Bayes theorem, which states

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{B}|\mathrm{A})P(\mathrm{A})}{P(\mathrm{B})}$$

Bayesian inference played a major role in statistical thinking until the early part of the 20th century. Although the hypothesis-testing ideas of Jersey Neyman, Karl Pearson, and Sir Ronald Fisher played a dominant role in the rest of the century, in the past 25 years the tide has begun to turn. The methodology and intuitive logic of the Bayesian approach, made more feasible by high-speed computers, is now pervasive in the literature of modern science and statistics.

Bayesian Inference

Bayes theorem is the basis on which Bayesian inference is built. Suppose that φ is a parameter and that φ = 0 represents the event that the null hypothesis is true. Then according to Bayes theorem,

$$P(\varphi|\mathrm{data}) = \frac{P(\mathrm{data}|\varphi)P(\varphi)}{P(\mathrm{data})}.$$

The parameter φ may be thought of as a representation of the scientific theory under investigation. The left-hand side of the equation is a quantification of the probability, or perhaps degree of belief, of the truth of the proposition that φ represents the true state of nature, given the observed data. In this setting, the unknown parameters are viewed as random variables and as such they follow a probabilistic law. Remarkably, if the right-hand side can be evaluated, a numerical value for the probability can be determined for this quantity, which arguably is the main issue once data collection is complete. It reflects the current belief, given the data, about the veracity of any value of φ and it is called the *posterior* distribution. The first term on the right is just the likelihood of the observed data, evaluated at the parameter value φ. It is the next term that has made Bayesian inference controversial. What is the value of $P(\varphi)$? This quantity is a function of all of the possible values of the unknown parameter φ, and it is called the *prior probability distribution*, or just the *prior* for short. A prior distribution that has the same functional form as the posterior distributions is called a *conjugate* prior. This is a particularly desirable situation, for it permits efficient sequential updating of the posterior distribution as more data are collected. The posterior distribution at one stage is used as the prior in the next.

Note well that the result of applying the Bayesian method is an entire posterior distribution and not a single estimate of the unknown parameter. The posterior summarizes all of the information, past and present, about the unknown parameter value. In the Bayesian framework, issues such as estimation and hypothesis testing are all based on the posterior distribution. For example, sensible estimators of a parameter are the mean or median of the posterior distribution. For hypothesis testing the posterior distribution can be used to calculate the probabilities that the null and the alternative hypotheses are true. One possible rule is to decide to accept the null as true if the ratio of the posterior probability of the null divided by the posterior probability of the alternative exceeds some constant and reject the null otherwise. The interpretation of a *P* value in this context is more natural than it is in the frequentist paradigm. The value of *P* is in fact the probability that the null hypothesis has been erroneously rejected. Contrast this with the frequentist interpretation that *P* is the chance of obtaining a data set that is more extreme.

The controversy about the use of Bayesian inference hinges on the problem of objectively choosing a prior. In many if not most cases, knowledge gleaned from previous experiments and perhaps scientific theory can guide the mathematical specification of $P(\varphi)$. However, in situations in which little is known or there is controversy about prior experience or relevant theory, the choice of the prior is subjective. There are two schools of thought about choosing the prior. In one approach the prior reflects the individuals' real-world experience and insights. An alternative approach is mechanistic or objective. The prior is chosen to satisfy some property; for example, it is noninformative or it is the conjugate distribution.

What price is paid for introducing subjectivity into the scientific process? Supporters of the Bayesian approach argue that Bayes theorem mathematically demonstrates the inevitability of a subjective component to scientific inquiry. Scientists are the first to admit that insight, intuition, judgment, and beliefs are part of the process that drives their practice. Regulatory bodies are less convinced, but even in such scientifically driven agencies such as the U.S. Food and Drug Administration (FDA), Bayesian approaches frequently appear. A full discussion of how the a priori distribution can be specified and the dependence of the final a posteriori distribution upon this choice is somewhat technical. Needless to say, there are a wide array of formal a priori distributions, including a noninformative prior, that is, one in which values of the parameter are equally likely, that can be used. A commonly used approach is to specify a range of priors and examine the degree to which conclusions are sensitive to the choice. The more robust the finding under different priors, the more believable they are.

RANDOMIZED CLINICAL TRIALS

There are many kinds of research projects in mental health that require the use of statistical methods to appraise the results. Here the discussion is focused on randomized clinical trials because they are important and ubiquitous, and many of the issues apply in other research contexts. The methods discussed are neither exhaustive in the range of topics covered nor are they complete in the detail that is required to appreciate the nuances fully. The goal of many mental health studies is to conclude “causality,” that is, some observed effect is the result of an intervention. To begin, this is a model that leads to conditions under which causality may validly be concluded. This may be accomplished by randomization or by using matched subjects. In the situation where matching is practically difficult, an alternate approach, the propensity score is introduced. The section concludes by introducing randomization tests that are valid regardless of the underlying distribution of the random variable of interest.

Causality

The principal reason for carrying out clinical research is to demonstrate the causal effect of the treatments being studied. Donald B. Rubin described an approach by which causality can be demonstrated in a clinical trial. Suppose that a putative antidepressant is given to a well-diagnosed patient suffering from the illness. If the patient does poorly, assuming the agent does no harm, no evidence is adduced about its effectiveness. Suppose the patient does well and there is a considerable reduction in symptoms. How can it be concluded that the treatment was the causal agent? If the clock could be turned back and the same patient with the same symptoms did not benefit from placebo treatment, it would be reasonable to conclude that administration of the test agent caused clinical improvement in this patient. Each patient has two hypothetical outcomes, O_T and O_P, which would result from being treated with T or P, respectively. If this hypothetical thought experiment were repeated a large number of times, then the frequentist would conclude causality if the average effect of an individual receiving the treatment is larger than the average effect produced by placebo in the same individual. In statistical terms, it would be written $E(O_T - O_P) > 0$, and it would be recognized that the concept is the same whether the hypothetical experiment is repeated over and over with the same patient each time or with different patients. But the generalizability of the conclusion is not the same in the two cases, and from a statistical perspective the interest is in population causality rather than individual causality. Of course, the clock cannot be turned back, so what other possibilities are there? If each of the patients has an identical twin with the same genetic makeup and identical clinical symptom profiles, then one of each pair could be given the test treatment and the other given placebo. If the former improved and the latter did not, a causal relationship has been demonstrated. That is, it has been demonstrated that $E(O_T) - E(O_P) > 0$. Unfortunately, in the real world, each subject has only one outcome, O. If the treatment assignment is T, then the outcome is

O_T, and if it is P, then the outcome is O_P. One of the two is counterfactual. In this case, $E(O|T) - E(O|P) = E(O_T) - E(O_P) = E(O_T - O_P)$. What is necessary to be able to conclude causality is a mechanism that would enable the conclusion that $E(O_T - O_P) = E(O_T|T) - E(O_P|P) > 0$, where the expectation on the left-hand side indicates that each individual received both treatments and the expectation on the right-hand side indicates that each individual received only one of the treatments. The notation beyond the vertical bar indicates which treatment was assigned to the patient, and the expectation, the computation of the average over the population, is restricted to those receiving the indicated treatment. Suppose that the treatment assignment is a random variable determined by some mechanism that is independent of the outcome, O. For example, suppose a patient is assigned to the test treatment if a fair coin (one whose probability of each of the two outcomes is .5) lands heads and to placebo otherwise. Then the treatment assignment is independent of the outcome random variable, O. But then $E(O|T) - E(O|P) = E(O_T) - E(O_P) = E(O_T - O_P)$. So, the use of a random assignment mechanism enables conclusions of causality.

Note that the coin did not have to be a fair coin for the treatment assignment to be random. The same conclusion on causality would be reached were the probability of assignment to the treatment group equal to, say, .6. This is called a *biased coin random allocation*. The only difference is that the result of such a randomization process would be that more individuals would tend to be assigned to the treatment group (about 60%) than to the control group (about 40%). This fact is important for the application of propensity score methods, which are discussed below.

In fact, the idea extends to situations in which treatments are assigned to subjects based on factors z that are observed. For example, z might be severity at baseline or some other stratification factor. Even *response adaptive designs*, where the probability that the next subject is assigned to the test treatment depends on outcomes observed on patients studied earlier in the trial, can demonstrate causal effects so long as the response to treatment and assignment to treatment are independent, conditional on z.

Matching and Randomization

Suppose that a clinical trial is to be conducted to establish a causal relationship between receipt of a treatment believed to be an active drug and clinical improvement, the effect produced. The treatment is to be allocated to one group of patients and a control treatment, a placebo, is to be allocated to another. As was just seen, the control group can be comprised of one member from each set of twins and the test treatment group the other twin and causal conclusions can be reached. Although such a design can only rarely be realized, matching two groups on all variables that are prognostic indicators of the outcome may be feasible. Subjects can be matched on the basis of characteristics such as age, weight, severity, prior episodes, genetic factors, family factors, and on and on. However, in psychiatric research, it will be a rare study indeed in which the investigator and more importantly the target audience of the research believe that the two groups are truly matched. However, there are some circumstances in which this will be the best approach, and such cases need to be appraised individually. In the idealized twin case, treatment assignment need not be random. But here, since there is no guarantee that paired subjects are truly identical, for causal inference the assignment mechanism within the pair must be random.

Randomization helps to avoid systematic bias, for example, evaluating one treatment in older or more severely ill patients and the other in younger or less severely ill patients. It has no effect on selection bias, which chooses individuals to participate in the study who are not representative of the target population. For example, in most drug comparison studies, the various eligibility requirements, both inclusion and exclusion, as well as the option candidates have of declining to participate in the trial bias the sample. However, this does not compromise the validity of the treatment contrasts so long as treatments are randomly assigned to participants.

The very mechanism of assigning treatments to subjects in the clinical trial provides the probabilistic basis for computing, under the null hypothesis, valid probabilities of the various possible experimental outcomes. But this is exactly what is needed to assess the likelihood that the null hypothesis is true. In particular, it enables calculation of a rejection region corresponding to a chosen type I error, which is the basis of hypothesis testing. Further, because randomization guarantees that outcomes for different subjects are independent, a valid estimate of the experimental error can be obtained. This quantity is required for obtaining valid CI estimates of treatment effects and for parametric testing of a null hypothesis. Also, given a sufficiently large sample size, all of the possible known and unknown confounding effects that might possibly affect outcome are, with high probability, eliminated in the sense that they are averaged out. Finally, the objectivity and repeatability of randomized assignment to treatment must be pointed out. There is no point in conducting a clinical trial whose results will not be creditable. The ability to describe exactly how candidate patients were placed into the various treatment groups makes transparent the process and gives credence to the conclusions reached.

The Likely Responder Method of Analysis

The premise of precision medicine is that there is diversity in patient response to a treatment as a function of baseline characteristics. For mental health disorders, whose presentations are highly variable, individual care would substantially improve if treatment were based on patient-specific characteristics. In an RCT, the opportunity to discover the best treatment practice is often eclipsed by obeisance to the classical hypothesis test of mean difference between treatments in the whole sample. Finding responder subtypes is relegated to secondary analyses, designed to provide input to the design of the next trial. This analytic strategy, on its face, makes no sense if the premise of precision medicine pertains. The Likely Responder approach to the analysis of an RCT reverses the order: first the subset of patients likely to respond to the test treatment is identified. Then a test of treatment differences is performed on patients in this group matched on baseline features. This causal analysis of treatment effect is carried out in a potential outcome framework. The Likely Responder approach was used for the first time in a reanalysis of a gabapentin clinical trial for the treatment of alcohol use disorder by Laska, Siegel, Lin, and their colleagues. While there were no treatment differences in the whole RCT sample, there were in the subset of likely responders.

Randomization Tests

Randomization tests make no assumptions as to the distribution of the random variables and so are considered to be distribution free. The results of the test depend only on the data actually collected. Suppose there are two groups of patients, whose group assignment is determined by a randomization procedure, with one group receiving the test treatment and the other receiving a placebo. After a suitable period of time, an outcome is obtained on each subject. The quantity of interest is the difference between the mean outcomes of the two groups. The first step in a randomization test is to compute the statistic, the difference between means, from the sample. The second step is to rerandomize and recalculate the value of the mean difference for the new treatment assignments. Under the null hypothesis, there is no difference between the groups; so reassigning some patients from one group to the other should not change any probabilistic properties of the distribution of treatment differences. Reassigning treatments to patients by rerandomization (or, for small numbers of subjects by exhausting all possible permutations of possible treatment assignments) and recalculating the mean differences induce a probability distribution of mean differences. The percentile in the randomization-induced probability distribution of the actual mean difference observed in the trial is informative about the plausibility of the null hypothesis. If the observed mean difference lies near the middle of the distribution, the null distribution seems likely to be true. If, on the other hand, the observed mean difference falls into one of the extremes of the randomization distribution, say less than .05 (α) of the values are larger, then the null hypothesis seems unlikely to be true. The mean values of the two treatments are likely to be different. The randomization *P* value is the proportion of rerandomizations in which the value of the statistic is more extreme than the value obtained from the original treatment assignment. Randomization tests produce valid significance levels (probability of a type I error) regardless of the usually unknown underlying probability distribution of the outcome measures.

Rank randomization tests are similar to randomization tests. However, first the observations are converted to ranks. That is, the two groups are merged and put in numerical order. The largest is assigned rank one, the second largest rank two, and so on. Next, the two groups are separated into their original configuration, and a randomization test is computed on the ranks. This is a particularly useful approach when the scale used to measure patient response has some nonlinear properties and when outliers in the data might skew estimates of mean effects. The disadvantage is that some information is lost when the actual observed values are converted to ranks. Although rank randomization tests may be somewhat less powerful than randomization tests based on the original data, the loss in power may be a small price to pay to purge the potential contamination caused by exceptional large or small outcomes.

Propensity Scores

In efforts to understand the properties of a new treatment or intervention, it is often impossible or too expensive to conduct a randomized clinical trial. On the other hand, there may be data available on the new treatment as well as on standard practice that has been collected under naturalistic conditions. In earlier sections, considerable attention was paid to the merits of randomization and the benefits that accrue from its use. Randomization of persons to treatment or control groups allows statements of causality to be made, and systematic differences in characteristics between the groups are not likely to occur if the sample size is sufficiently large.

Is it possible to nevertheless obtain valid estimates of effect size from data on patients who were not randomly assigned to treatment? Many standard statistical methods of analysis are not appropriate because they implicitly (at least for the purpose of causal inference) assume that the groups were formed by random assignment.

Suppose that background characteristics, some known, such as severity of illness and age, and others that may not be known, are markers for the assignment of patients to a new treatment or to a control or usual practice group. This might occur as part of usual care procedures or by unconscious hunches designed to individualize and optimize treatment outcome. If the two groups are compared without regard to the treatment assignment mechanism, the results can be biased particularly if there is an imbalance across groups in variables that influence outcome.

If there are only a few confounding variables and they are known, it may be possible to form a stratum in which the treatment and control group look alike. For example, suppose there is a difference in the proportion of males receiving treatments in the two groups. Then a group containing only males and a group containing only females can be formed. The treatments can be compared within groups and if they are similar the results can be pooled.

If there are two known confounding variables, such as age (young, middle aged, and older adults) and gender, then age/gender groups can be formed yielding six strata in which to make comparisons. However, as the number of confounding variables grows, it becomes increasingly difficult to form the multiple strata that would be required. Sample sizes in each stratum would considerably diminish, resulting in a loss of power to detect differences when they truly exist.

Propensity score methods are used to control for a collection of confounding covariates. The propensity score is a probability of treatment group membership based on patient covariates. An estimate of the propensity score for each individual in the study is usually obtained by first fitting a logistic regression and then substituting his or her covariates into the resulting model. Naturally, persons who are actually in the treatment group will tend to have higher propensity scores than those in the control group. If the range of propensity score values in the two groups overlap, then it is possible to match individuals in one group with those in the other group who have like scores. Rather than one-to-one matching, usually the propensity scores for the combined groups are ordered, and five strata, that is, quintiles, are formed. Within a stratum, all individuals have approximately the same estimated likelihood of being assigned to the treatment group, and the multivariate distribution of the covariates used to estimate the propensity score in the two groups do not differ statistically. Across strata, the likelihood of being assigned to the treatment group differs. Within each stratum, some individuals were assigned to treatment and some to the control, so biased coin randomization has in essence been mimicked. If all subjects have a propensity score of about one-half, then treatment assignment appears to have been determined by a fair coin randomization process. Treatment effects are estimated within propensity score strata, and if the results are similar, they can be combined to obtain an estimate of the overall effect size. As Philip Lavori put it, within each observational study there lives a randomized study that can be found with the help of propensity scores.

More on Causality

When possible, causal relationships are inferred by manipulating the causal agent and determining its effect on the target in an experiment. However, there are many instances in which experiments are not possible and only observational or naturalistic data can be obtained. For example, a researcher studying the development of Posttraumatic Stress Disorder (PTSD) may hypothesize that different trauma types (combat exposure, motor vehicle accident, sexual assault) differentially affects the likelihood of the emergence of PTSD.

Causal Graph Models, which utilize directed acyclic graphs (DAGs) to represent causal relationships between variables or events, has been developed to facilitate causal inference in such situations. A DAG is simply a mathematical model described pictorially with arrows (edges) connecting vertices (variables or events) whose chance of occurrence are given by conditional probability distributions. Typically, arrows point in the same direction from *parent* or *ancestors* nodes (the cause) to *child* or *descendent* nodes (the consequence.) *Paths* refer to a sequence of edges connecting one node to another possibly distant node; *causal paths* are those leading from a causal node to an outcome node. *Colliders* are nodes at which two edges converge. Necessarily, such graphs are acyclic in that there are no paths in the graph that start at one node that eventually loops back to the same node. The goal of this technique is not discovery of an *association*, statistical dependence between variables. The goal is determination of causality. The simultaneous presence of the symptoms of streptococcus and streptococcus bacteria is a dependence relationship (they are not independent) which is substantially different from the causal chain: the presence of symptoms is caused by the presence of the bacteria. A dependence relationship cannot differentiate disease causes symptoms and symptoms cause disease.

A *Bayesian network* is a DAG in which the nodes are events together with their probability distributions that depend only on the probabilities of prior events. In this sense, a probabilistic DAG formally encodes knowledge and beliefs about causal and potentially causal relationships among variables through visual representations. Edges represent conditional dependencies while nodes that are not connected are conditionally independent. A probabilistic DAG has two primary purposes: to represent assumptions explicitly in order to reveal testable causal hypotheses and to enable explicit data driven hypotheses testing of these causal effects. This approach sharply contrasts with traditional hypothesis testing approaches that assume a null relationship which is rejected when data supply sufficient evidence of incompatibility. In Bayesian network theory, potential relationships between nodes are specified in a DAG and data is used to sustain or reject the relationships. Many algorithmic computer programs have been written to discover causal graphs by hypothesizing edges between all nodes and removing those that are incompatible with data.

Missing or Incomplete Data

In the conduct of research in mental health settings, having missing data is inevitable. Such data sets are said to be “incomplete,” and the problem of parameter estimation or hypothesis testing is regarded as an “incomplete-data” problem. Missing data problems occur both in clinical trials and in studies that are survey or field based, and the data can be missing for many reasons. In a clinical trial in which data are collected at multiple time points, subjects may not be available for all interviews. Some may drop out of the trial before it is completed because of adverse events or death, or they may move to another home, hospital, or jail or to the streets, or they may just perceive continuing in the trial to be a nuisance. In surveys, respondents may refuse to answer particular questions because of the sensitive nature of the inquiry, or they just may not know the answer.

There is no one approach to handling all missing data situations, but ignoring holes in the data when performing a statistical analysis can have an enormous effect on the results. The possibility of substantial structural biases that might have led to the data losses can lead to analyses that may even be so fatally flawed as to nullify the results. Conceptually, the complete data set consists of two parts, only one of which is observed. The missing part can only be inferred from the observed part. The steps to take depend on the mechanisms that led to the missing data. This is often referred to as the *missingness* of the data. Roderick Little and Donald Rubin have defined three kinds of missing data mechanisms: Missing completely at random (MCAR), missing at random (MAR), and nonignorable missing.

A very simple probabilistic model can be used to conceptualize data that are MCAR. Suppose that a random variable taking the value 0 or one with respective probability P and $1 - P$ was observed for each observation of the variable Y. Assume further that the value of the observation on Y is independent of the binomial random variable. Such a random variable is obtained by tossing a coin whose probability of landing heads is P. If the coin toss corresponding to a particular observation of Y lands heads, the data are lost. The remaining observations are a random subsample of the original sample and so the data remain a random sample.

An example of MAR can be obtained in a similar fashion. Suppose Z is gender, and a coin is tossed for males with probability P_1 of landing heads, and a second coin is tossed for females with probability P_2 of landing heads. For each observed Y, one or another coin is tossed depending on the gender as captured by Z. If the coin lands head, the data are lost. The remaining data in each subsample is still a random sample.

As an example of nonignorably missing, suppose the observations on Y have been sorted into good and bad outcomes, and two coins are tossed with respective probabilities of heads P_1 and P_2 for the two types of outcomes for each observation. If either coin lands heads, then the corresponding observation is lost. The remaining data no longer arise from the same underlying probability distribution that gave rise to the original sample. If $P_1 = P_2$, then the effect of sorting an outcome no longer matters, and the subsample is a random subsample of the original sample. In this case the data would be MAR.

There are several methods used to handle the situation when data are MCAR or MAR. Certainly, the simplest strategy is to delete from the study those records with missing data on the primary variable of interest. If the sample size is sufficiently large, this may be a viable approach. However, if the amount of information lost diminishes the ability of the study to reach confident conclusions, then this method is not recommended. If the data are multivariate and only some of the variables in a record are missing, it would be rather wasteful to dispose of the entire record. Some analysts retain those parts that have data that are present and use them in marginal analyses not involving the variables with missing data. This approach too can be problematic. For example, suppose a clinical trial is designed to compare an active treatment with placebo. Those individuals who respond to the placebo remain in the trial, while the others drop out. For a self-limiting illness, the placebo responders and the treatment responders are not likely to differ toward the end of the trial. Ignoring the early dropouts and limiting the analysis to those who remain in the trial could lead to the erroneous conclusion that the drug loses its efficacy after a while. Filling in or imputing the missing data with a conservative estimate of its value is the current favored approach. However, this strategy too can lead to erroneous conclusions if there are too many records with missing data. Neither strategy is completely satisfactory nor is there a satisfactory rule of thumb as to how much missing data is too much for imputation. Individual variations among studies require critical appraisal in each particular case as to what is an appropriate analytical approach.

Imputation of missing data can be done in several ways. One simple imputation procedure is called *mean/median substitution*, in which the mean or median value of the observed data is substituted for the missing value. This approach tends to reduce the estimated variance, which thereby provides more comfort to the researcher than may be warranted. Several methods rely on a matching strategy. Records with missing data are matched with records of subjects with similar demographics whose outcomes are present. The data from the matched records are used to impute the missing value. In *hot deck imputation* cases the missing data are matched with similar cases with complete data on the basis of defined background variables, for example, age and gender. A critical issue is how to choose the variables that define “similar.” Matching criteria variables are chosen because they are prognostic indicators of the missing data variables. The value substituted for the missing value is randomly selected from data present in the set of matched cases. *Multiple imputation* uses an imputation method such as hot deck to repeatedly generate estimates for the missing data values. For each replication, the full data set, actual and imputed, is analyzed. The results from these multiple analyses are combined into one overall analysis. This approach gives an idea of the variability of the results among the various possible imputations that could have randomly been selected.

For data collected longitudinally, many researchers have used the last observed value carried forward (LOCF) approach to fill in the missing value. If values are available before and after the missing time point, many interpolate a value. These approaches too tend to reduce the estimated variance, which can lead to a greater chance of reporting an effect when there really is none. Hierarchical linear modeling provides a more satisfactory approach for analyzing longitudinal data. In these models, a time curve is estimated for each individual in the study from the available data without imputation, which works well as a substitute for interpolation. If many patients have missing observations in the later stages of a study, there may be a considerable degree of extrapolation. In this case, the method must be used with caution.

More sophisticated methods for imputing missing values are model based. For example, regression models have been used where the variable with missing data is the dependent variable. Parameters in the regression equation are estimated based on data from complete records. The predicted value from the model is used as an estimate of the missing data value. Another approach relies on maximizing a likelihood function using the available data values while taking into account the nature of the missingness. Here great care must be taken in specifying the underlying distribution function of the random variables.

Perhaps the most widely used method of imputation is the expected/maximization (EM) algorithm. It was originally conceptualized as a method for finding maximum likelihood estimators for incomplete data. The algorithm starts by choosing initial estimates of the values of the unknown parameters. The estimates are improved iteratively by first computing the conditional probability distribution of the missing data given the observed data and then the current estimate of the parameters. The expectation of the log likelihood over the previously calculated distribution of missing values is obtained. This is called the *E step* of the algorithm. A new or updated estimate of the unknown parameters may be obtained by maximizing this expectation. This step is called the *maximization step* (M step) after which the E step is repeated. An important feature of the EM algorithm is that it gives the complete probability distribution for the missing data and point estimates for the model parameters. Under some reasonably general conditions, the EM estimates obtained in the iteration procedure converge locally to the MLE.

The methods described above handle missing data under the MAR assumption. Pattern-mixture models have been proposed for nonignorable missing data. Records are stratified into groups with different expected patterns of response and missingness. The group then becomes a predictor variable in a method of analysis (e.g., hierarchical linear models [HLM]) that allows subjects to have incomplete data across time. In this way the degree to which missingness influences outcomes can be assessed. Group membership can also be used in an interaction term in the model. This allows the assessment of the degree to which missingness patterns interact with key effect variables such as treatment group.

In sum, there is no one simple recipe for handling missing data. An understanding of the reason for the missingness is necessary, and a critical eye must be kept on the question of whether the MAR assumption is plausible. Whatever technique is used, some form of sensitivity analysis should be used to see how the imputation carried out in different ways affects the results. The choice of technique to use to handle the missingness must be made on a case-by-case approach.

SPECIFIC STATISTICAL METHODS

In this section, a variety of commonly used analytic techniques are introduced. These include linear regression and logistic regression, contingency tables, the general linear model (GLM), and survival analysis. In mental health research, outcomes are typically observed over time, and a discussion of longitudinal methods to address such situations is given.

Regression Models

Regression models quantify a relationship between a dependent variable and one or more independent, explanatory variables. A statistical model is complete only if both its functional and probabilistic form and its parameters are determined. The procedure for estimating the unknown parameters is known as *model fitting*. Historically, the dependent variable, say Y, is assumed to follow a normal distribution, and the hypothesized model relates its mean or expected value and a function of the independent variables, say x_1, $x_2,\ldots, x_k$. The data consists of an observation on independent and dependent variables on each of N subjects. The most common model assumes that the normal mean is a linear function of the explanatory variables. That is, it is of the form

$$E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k,$$

where the $k + 1$ Greek letters, α, β_1, $\beta_2,\ldots$, β_k are unknown parameters to be determined from the data. In terms of the actual observations rather than the expected value of Y, the form of the model is assumed to be

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon.$$

Here ε represents a random error term that accounts for the deviations among the observations from their mean value. For each subject, the difference between the observation Y and the model fit evaluated at his or her specific value of the independent variables is called the *residual error*. If Y follows a normal distribution with mean $\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$, and variance σ^2, then the error term is also normal with the same variance but with mean zero. In a regression analysis, estimates of the coefficients are obtained by maximizing the likelihood of the observations. Elementary calculus is used to find the value of the estimates, which are the solutions of a set of simultaneous equations, referred to as the *normal equations*. An alternative strategy for finding estimates of the parameters is to minimize the *sum of the squared errors* as a criterion for the goodness of the fit of the model to the data. The errors or the *residuals* are the difference between the observed values and the predicted values, the latter obtained by substituting the values of the *x*s into the regression equation. The *ordinary least squares* (OLS) approach makes no assumption as to the distribution of the error term. The method of OLS was developed at the turn of the 18th century by the mathematician Karl Friedrich Gauss working in Germany, and some historians say by others including Adrien Marie Legendre, Pierre-Simon Laplace, and (possibly) Robert Adrain, France and America, respectively. The least squares solution and the maximum likelihood solution are identical if the dependent variable is normally distributed.

Since specification of the set of explanatory variables is somewhat subjective, regression analyses include tests of whether each coefficient is equal to 0. If the null hypothesis is rejected, then the explanatory variable is included in the model. Stepwise procedures for choosing the variables to include in the model are also commonly used. Goodness of fit may be ascertained in a variety of ways, and model validation is a critical step in building a creditable model. At a bare minimum, the value of an R^2 statistic, which is the fraction of the total variability in the response that is accounted for by the model, should be obtained. But a high value of R^2 is not a guarantee that the model fits the data well, so it is wise to examine the residuals as well. Different types of plots of the residuals provide information on the adequacy of different aspects of the model. If the model fits well, a plot of the residuals will appear to behave randomly. This approximates the random errors that explain the difference between the independent variables and the response variable. If there is a nonrandom structure evident, then the model does not fit the data well.

There is, of course, no reason to limit the function on the right-hand side of the regression equation to linear forms. Many applications use rather complex nonlinear functions to represent the hypothesized relationships. In these cases, numerical methods are the only practical means for obtaining the estimates, and hypothesis tests are usually valid for large samples only. In the past quarter century statisticians have generalized regression models in many directions, and applications incorporating these extensions appear frequently in the psychiatric literature.

Classification: Logistic Regression and Machine Learning Techniques

Many research studies deal with outcome variables that are discrete. The simplest and most common case is dichotomous or binary random variables. A binary variable is often used to define the occurrence of an event, such as a treatment success, a treatment-emergent side effect from a study medication, the disappearance of a symptom, or death. The event of interest might also be a classification problem, that is, predicting membership in one of two groups. The random variable Y takes the value 1 with probability P if the event occurs, and zero with probability $1 - P$ otherwise. This is called a *Bernoulli random variable*, and an experiment in which many Bernoulli random variables are observed is called a *Bernoulli trial*. An investigator may want to know how other explanatory variables (covariates or risk factors) affect the frequency of the occurrence of the event. The idea is to use data, the training set, to find a method that will identify relevant covariates and use them to correctly predict an outcome in the test set. In recent years, there has been an explosion in the amount of available biologic information (genetics and epigenetics, imaging, peripheral neuroendocrinology, clinical self-report, etc.) that can contribute to the classification and prediction accuracy. The limitations of traditional data analytic methods in such circumstances, in which there may be many more potential predictors than subjects has led to the development of “big data” methods. In this section, after introducing classical logistic regression analysis, we introduce machine learning techniques, algorithmic methods that build models based on sample data that do not follow strict rule based or structured statistically modeling approaches.

Logistic Regression

Were the dependent variable continuous and approximately normally distributed, multiple linear regression would likely be used to explore its relationship to the covariates. Here, linear regression is not appropriate. An applicable model for a dichotomous variable is *logistic regression*. In this approach, the model assumes that the log of the odds ratio of the event occurring is linear in the covariates. These independent variables may be continuous variables, categorical variables, or both. The *odds* is the ratio of the probability that the event occurs to the probability that it does not occur. Odds are commonly used in gambling contexts. If the odds that a team will win are three to one, then one would expect the team to win about three times more often than its losses. In this case, the probability of winning is 3/4. Suppose q is the probability of the occurrence of the event. Then odds = $q/(1 - q)$ and q = odds/(1 + odds).

The logistic model is

$$\log q/(1-q) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$

The left-hand side of the equation is called a *logit*. The covariates on the right-hand side of the equation can be continuous, dichotomous, nominal, or ordinal. Tests of the null hypothesis that each parameter is 0 may be made. Interpretations of the coefficients are made in a manner similar to ordinary linear regression. However, logistic regression calculates changes in the log odds of the dependent variable, not changes in the dependent variable itself, as in linear regression. For plotting and interpreting results from logistic regression, it is usually more convenient to express fitted values on the scale of probabilities. The inverse transformation of the model equations yields the logistic function for the probability of the event q.

$$q = \exp[\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k] / \\ (1 + \exp[\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k]).$$

The fit of a logistic regression can be assessed by looking at a classification table that displays whether the model has correctly classified each of the observed outcomes of *Y*. Goodness-of-fit tests are available as indicators of model appropriateness. One method begins by dividing the subjects into ten groups, in an increasing order of estimated probability of occurrence of the event. The first group corresponds to those subjects who have the lowest predicted probability; the next group corresponds to those with the next lowest predicted probability; and so on. A chi-square statistic is used to examine whether the observed counts in each group are close to the expected count under linearity. If they are, the chi-square statistic will be small and the corresponding *P* value will be large. Logistic regression can be extended beyond dichotomous response variables to ordered polychotomous (more than two) categories.

There are many measures of the accuracy of a classifier. The most widely used is the area under the receiver-operating (ROC) curve. The ROC graphically displays the true positive rate (sensitivity or recall) plotted against the false positive rate (1 - specificity, or fallout) at various thresholds. The area under the curve (AUC) from a sample data set is an estimate of the probability that the classifier assigns a higher score to a randomly chosen member of one class than to a randomly chosen member of the other class. The *P* value of a chi-squared test that this probability differs from ½, is often reported.

Supervised Machine Learning

In the classification setting, *machine learning* methods are shorthand for a collection of computer-driven approaches for obtaining a classification rule learned from training data. Among the many machine learning algorithms developed, two prominent and promising approaches, *CART* and its generalization *Random Forests* and *Support Vector Machines (SVMs)* are introduced below. These are popular and have been implemented in a number of software packages.

CART determines a set of logical if-then conditions, instead of a linear equation as in logistic regression, on predictor variables that corresponds to a logical or decision tree leading to terminal nodes.

CART, the abbreviation for the classification and regression tree algorithm, is based on *recursive partitioning,* is often able to uncover complex interactions among predictors or features that may be difficult or impossible to uncover using traditional multivariate techniques. The composition of the training set membership in a specific terminal node is used to estimate the conditional probability of class membership and to predict or classify all subjects in that node. For example, if 80% of the subjects in the node are in one group, then all subjects in the node, are classified, some erroneously, as members of that group. The feature on which to split and the value on which to split (e.g., age greater than 75) at each node is chosen by the software to produce the greatest improvement in predictive accuracy at each step. This is usually driven by a “node impurity” measure such as the Gini Index, which provides an indication of the relative classification homogeneity of subjects that fall in the terminal nodes. The general approach to limiting over-fitting is first to stop generating new splits when further splitting produces only a small improvement in prediction accuracy. The candidate prediction tree is subsequently parsimoniously “pruned back” to further avoid the use of more complex models than are required. An added possibility is the use of linear regression at the terminal nodes.

Random Forests. CART provides the underpinning of the Random Forests (RF) classification method. In the RF approach, a multitude of *decision trees* are constructed during the training phase. The individual trees in the “forest” are grown by taking a random sample of subjects. If there are many more candidate features than subjects, these too are sampled at random as well. The final probability that an individual with a particular vector of covariates is predicted to be a member of a class is by “majority vote,” that is, the average of the probabilities of the terminal nodes into which the predictor vector falls in each tree in the forest. This *bootstrap aggregating approach* is called *bagging* and its purpose is to reduce variance and to prevent overfitting. Each individual tree may be highly sensitive to noise, but model averaging across many trees reduces variability, under the condition that individual trees are not highly correlated. By selecting random subsets of features in growing the forest, referred to as “feature bagging,” correlations among trees is reduced.

Support Vector Machines. SVM algorithms build classifiers by identifying a linear decision surface that separates the classes with the largest distance, also called *gaps* or *margins*, between objects at the borderline. If there are two covariates, the space is two-dimensional and the decision surface is a line. Although an infinite number of lines can separate two classes, only one, called the support vector, uniquely separates classes with the largest gap between borderline objects. Often it is not possible to linearly separate sample objects belonging to two classes. When a linear decision surface cannot be identified, a SVM maps the data into a higher dimensional feature space in which a separating linear decision surface can be identified. This technique is known as the *kernel trick*, which extends SVMs beyond linear to nonlinear classification. This framework is particularly useful in the not uncommon situation in which the dimension of the features is extremely large. SVMs construct a hyperplane, or set of hyperplanes in high dimensional space, which separates the groups and determines the classification of a new subject. Support Vector Regression is an analogous procedure when the variable of interest is a real number instead of a class indicator.

Over-fitting. When there are hundreds or thousands or even millions of *features* under consideration, the potential for over-fitting is profound, particularly when the number of subjects is relatively small. With so many predictors available to use in fitting the model, the algorithms of most classification programs can achieve near-perfect results on the training set, but more often than not fail abysmally on the test set. Many techniques have been developed to avoid or at least detect this over-fitting difficulty while at the same time permitting utilization of the full range of potential predictors. These include evaluating the model’s performance on a data that is independent of the training set, for example, cross-validation, as well as methods that utilize fit criteria that penalize the use of increasing numbers of parameters and more generally, model complexity, for example, regularization and pruning. In a Bayesian framework, regularization is equivalent to imposing specific forms of priors on the parameters. Randomization tests are also used; the class membership in the training sets is randomly scrambled and the classification algorithm is applied. If good sensitivity and specificity are obtained for fictitious class labels, clearly, the algorithm cannot be trusted.

Contingency Tables and Log Linear Models

Many research studies involve discrete random variables. Frequency distributions of two discrete variables tabulated simultaneously are called *contingency tables*. The levels of one variable are listed in the rows of a table, the levels of the other variables are listed in the columns, and the joint frequency is entered in each cell. The sums of the cell frequencies across both the *r* rows and *c* columns are placed in the margins. The lower right-hand corner of the table displays the sum of the row marginal frequencies, which is equal to the sum of the column marginal frequencies, which is equal to the sample size.

Reports of many research studies begin with a description of the sample population, including the distribution of patient characteristics, such as treatment and baseline severity of illness. If there are two treatment groups being compared, then it is important to assess whether the groups were comparable at baseline. Define a treatment group indicator variable for the columns and a severity level variable for the rows. The question then is: Are the two variables independent or is there an association? Effects in a contingency table are relationships between the row and column variables in which levels of the row variable are differentially distributed over levels of the column variables. If the two variables are not independent, the two treatment groups have different distributions of baseline severity level, and the row variable will need to be controlled in any subsequent treatment comparisons in the analyses. Failure to reject the null means that there is not enough evidence to conclude that the variables are dependent; observed differences in cell frequencies are likely due to chance. Nevertheless, even if the test failed to reject the independence hypothesis, it may still be necessary to adjust for imbalances in baseline variables. Important prognostic variables may be sufficiently imbalanced to alter inference and cause statistical errors.

Hypothesis tests of association are often based on a chi-square statistic. The test statistic examines the difference between the observed counts in the table and those counts that would be expected if the variables were independent. The statistic sums the ratio of the squared difference between the actual count and the expected count to the expected count under the assumption of independence. If the differences are small, the data support the hypothesis that the variables are independent. In the 2 × 2 table, the difference of proportions, the relative risk, and the odds ratio are measures of the strength of association.

Contingency tables can involve more than two variables. One of the most common situations is the $2 \times 2 \times K$ table. For example, one variable may be treatment assignment to drug or placebo, the second may be a variable that measures success or failure, and the third may represent a control variable such as research sites, gender, or initial severity. There are three test statistics commonly used to analyze such tables. The Cochran–Mantel–Haenszel (CMH) statistic assumes a common odds ratio and tests the null hypothesis that two of the variables are conditionally independent, while controlling for the possible confounding factors of the rest of the variables. The odds ratio is a measure of the increase in odds of success for those given a drug relative to those given a placebo. If the odds ratio is 1, the drug has no effect. The CMH tests whether the response is conditionally independent of the explanatory variable (treatment assignment) while adjusting for the control variable (site). If the directions in some strata are opposite to the direction of those in other strata, the CMH statistics have low power for detecting an association. The CMH procedure first estimates a common odds ratio across the K strata and tests whether it is equal to 1. The

Breslow–Day statistic tests the null hypothesis of homogeneous odds ratios. That is, it tests whether the odds ratio between the assignment and outcome variables are the same for the *K* levels of the third control variable.

Higher level multidimensional contingency tables and more complex hypotheses are often analyzed using an approach known as *log linear models*. These models are used when two or more variables are measures of response rather than explanatory or control variables. The log linear model is one of the specialized cases of the generalized linear model, discussed below, for Poisson-distributed data.

The General Linear Model

The restriction to a normally distributed dependent random variable in regression theory was removed with the introduction of the GLM by John Nelder and Robert Wedderburn. This model assumes that the dependent random variable has a distribution in what is called a *parametric exponential family*. This family includes not only normally distributed and other continuous random variables, but also many discrete random variables such as the binomial and the Poisson. The latter distributions are used for modeling count data. The classical regression approach that models a normal mean as a linear function of the explanatory variables is generalized in GLM to a regression between a function of the mean and the explanatory variables. The function is called the *link function*. Although there are many possible link functions, associated with each particular parametric distribution is a natural or *canonical link function*, which enables efficient estimation of the coefficients. Here too, estimation is accomplished by maximizing the likelihood of the observations under the assumed parametric exponential family distribution via the link function. For example, in the case of a Bernoulli response variable with mean p, the canonical link function is $\log[p/(1 - p)]$, the log of the odds. Here, GLM assumes that the error term follows a logistic distribution, which is a member of the exponential family of distributions. But this is exactly what is done in logistic regression, which is therefore a special case of GLM. As in ordinary linear regression, tests of the hypotheses that coefficients of the explanatory variables do not differ from 0 are used to determine which ones have an effect on the mean. Similarly, if count data are distributed as Poisson variables and the link function is the log of the counts, then the GLM analysis is equivalent to a log linear model analysis of a contingency table.

Longitudinal Models

In the GLM and regression context, a single dependent random variable for each subject is considered. Especially in studies of mental and other chronic illnesses, subjects are usually observed at multiple time points or occasions. A typical hypothesis in such situations is that the means at all T time points are equal. This corresponds to the statement that there is no change over time. Conceptually, data in this situation may be described as a rectangular array, comprised of N rows and T columns. The rows correspond to the N subjects, and the entries in the columns are the subjects' response at the T time points. In a pre- and postmedication trial with N subjects, there are $T = 2$ columns. In the simplest case, the question is whether the mean effect of the pre- and postmedication responses are the same. A paired t, based on the difference between the pre- and postresponse is the usual statistic for this comparison.

Until fairly recently, the most common approach to analyzing longitudinal data was based on a *repeated measures* model. The statistical assumption for this model requires that the T random responses of each subject follow a multivariate normal distribution in which the covariance structure is completely symmetric. That is, the variances at each time point are equal and the covariance between any two responses, irrespective of their time difference, are equal. Clearly, these are rather demanding assumptions. Also, missing observation presented a significant problem, and in many trials all data on some subjects have been omitted. More recently, methods such as the EM algorithm enabled incomplete repeated measures data sets to be analyzed. However, hierarchical multivariate linear models, a generalization of repeated measures model have become more popular.

In a longitudinal study, despite the best of intensions, the occasions when subjects are observed as well as the number of time points that subjects are observed may vary widely from patient to patient. HLM allows such deviations, which overcomes the missingness problem when the data are MAR. While still retaining the multivariate normal distribution assumption for each subject, these models do not restrict the structure of the covariance matrix. The name hierarchical refers to the view that conceptually the analysis proceeds on different levels: Time is nested within subjects, subjects may be nested within sites, sites may be nested within countries, and so on. The model posits that the random normal observation is a sum of three terms: (1) a sum of the form, $\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$, (2) a sum of random effects, and (3) an independent error term. The first sum is referred to as the *fixed effects* term. Although the random effects comprising the second sum are not observed, they enter the model by contributing to the variances and covariances of the variables. Since subjects are independent, the form of the likelihood is a product, which leads to a set of extended normal equations for estimating the parameters. Parameters may also be estimated using a least squares approach, although within this model a generalized least squares criterion, that weights the residuals in terms of variances and covariances, is used.

In a still more recent development, the multivariate normal assumption has been relaxed through the introduction of a *semiparametric regression* method, which incorporates elements of the GLM. The method of generalized estimating equations (GEE), introduced by King-Yee Liang and Scott Zeger, focuses on the means at each time point through link functions. In this way, each marginal mean response at each observed occasion is modeled. In many psychiatric studies, hypotheses are concerned with the behavior of marginal means. In lieu of modeling the entire joint distribution of the observations over time, the GEE method models only the variance and covariances of the observations over time. The variance at a given occasion is modeled as a function of the marginal mean at that occasion. The method allows missing data provided they are MCAR, that is, they are MAR conditional on the covariates. Although the likelihood function cannot be formed because the parametric form of the distribution is not specified, estimators can be obtained by solving an analogue of the normal equations. The resulting estimators are consistent. That is, as the number of subjects increases, the estimators converge to their true values. Consistency holds even if the structure of the correlation matrix has been misspecified. Also, asymptotically the distribution of the estimators is normal, so that testing whether the coefficients of the independent variables differ from 0 can be accomplished.

Latent Class Growth Analysis and Growth Mixture Modeling

A constantly occurring research question is identification of distinct patterns of longitudinal course and outcome of a disease and the differential characteristics of individuals who follow them. For example, individuals’ response to trauma varies; many resilient individuals rapidly remit, some are slow remitters, some are nonremitting and some “late-onset” victims display low symptom levels initially that increase over time. Latent Class Growth Analysis (LCGA) and Growth Mixture Modeling (GMM) are commonly used statistical methods for identifying trajectories and the associated covariates associated with class membership. GMM extend traditional growth curve modeling, which allows the intercept and slope parameters to vary across individuals. However, growth curve modeling assumes a common pattern of change across the population. In the case of GMM, the latent, that is, the underlying but not directly observable, variables that characterize the time course (the growth curve model, such as the slope and intercept) are utilized to identify an additional second-order categorical latent variable, class membership. Each class has its own growth model with unique covariance structure. The distinction between LCGA and GMM is that the former assumes that the variance and covariance among the growth factors within each class can be set to zero; that is the trajectories of class members are homogeneous. Because of complexities in model identification, analysts often begin with LCGA and constraints on the variance are removed as the search for a best fitting model proceeds. A typical data set is comprised of repeated measures across multiple time points and if data are missing at random, MLE is valid. A trajectory assumed to have a linear form requires, at a minimum, observations at three time points. A trajectory with quadratic term requires at least four time points. Maximum likelihood parameter estimates are obtained iteratively through the use of the EM algorithm. In these models, appropriate model selection methods need to applied to guard against misspecification or over-fitting. Omission of influential variables, inclusion of irrelevant variables, incorrect model form, errors-in variables, incorrect covariance matrices and heteroscedasticity are just some of the issues that need to be considered.

Survival Analysis

Many psychiatric issues are related to events in the time domain, for example, time to onset of effect of a treatment, time in hospital, and time to remission of symptoms of depression. These are generically called *time-to-event* or *survival* variables. The first survival variable that was studied was time to death. *Life table methods* were originally devised for actuarial and insurance uses to describe expected survival time. For a particular group of persons (e.g., males) and selected age intervals (e.g., decades), life tables usually exhibit the number of persons entering the age interval alive, the proportion of these who die in the interval, and the proportion for whom data became unavailable during the interval. The proportion who die in the interval is called the *hazard rate*. Other useful statistics that can be derived from such tables include the *conditional survival rate* (among those who enter the interval, the proportion surviving beyond the interval) and the *survival function* (the cumulative proportion of individuals surviving to the end of the interval). The *survival distribution* is defined to be 1 minus the cumulative distribution function, 1 − F(*x*). The time at which the survival function is equal to .5, the *median survival time*, is the statistic most often used to characterize the distribution. Other percentiles are of course of interest as well. Although life tables provide a great deal of information, they are of little use for predictive purposes or for assessing the impact of other variables on survival.

Sophisticated statistical methods for estimating survival time distributions were developed primarily to model data from clinical trials of cancer treatments, and even more sophisticated methods have emerged in association with the study of acquired immunodeficiency syndrome (AIDS). Survival time methods are widely used in most of the medical and biologic sciences as well as in engineering (where the method is called failure time analysis) and in the social and economic sciences. The methods for estimating survival time frequency distributions from sample data have been developed to take into account the possibility that persons may still be alive at the end of the study period. Such individuals in essence have missing data as to the time of their death, but it would be folly to exclude them from the analysis. Observations on these individuals as well as the times of those who dropped out of the study before the signal event occurred, for example, a subject has moved and contact is lost, are called *censored observations*. If a new treatment is designed to study recurrence of depression and the study lasts 2 years, a censored observation at study completion or at loss to follow-up indicates that the treatment has been successful at least for that period of time. As in the missing data situation, analysis of these latter types of censored observations is facilitated if the data are MCAR, and, are therefore, noninformative as to the distribution of the time to the event being studied. Nonignorably missingness such as systematic censoring, if erroneously assumed to be MCAR, may greatly bias the estimation of the survival time distribution.

The Kaplan–Meier survival function estimation technique is nonparametric because it makes no assumption as to the underlying shape of the distribution. Based on the observed survival times and censoring times, it defines time intervals that contain exactly one observed case (except if there are ties). This is in sharp contrast with life table methods that count the number of deaths in fixed time intervals. The probability of surviving beyond time *t* is the product of the probability of being alive in each interval up to time *t*. If a case is censored at some time point, it is used in the calculation of the estimated probabilities up to the time of censoring and not used in subsequent calculations. Survival distributions generated from two samples, such as two treatment groups, can be compared statistically. The most commonly used tests are nonparametric, including Gehan test, the generalized Wilcoxon test, and the log-rank test.

Parametric survival methods assume an underlying shape of the survival distribution, for example, exponential, Weibull, or Gompertz. Techniques such as MLE or least squares to fit the parameters of the model from the observed data are used to estimate parameters' values.

Many research projects entail an investigation of the influence that covariates have on survival time. Cox proportional hazard models were introduced to enable regression-type analyses to be conducted on time-to-event data. No assumptions are made concerning the underlying form of the survival distribution, so the method is called *semiparametric*. The model assumes that the hazard rate, the probability of dying at time *t* given one is still alive up to that time, is equal to a fixed population or underlying hazard rate multiplied by the exponential of a linear function of the covariates. The log of the exponential multiplier is a linear sum of the covariates. The multiplicative form of the model implies that the hazards are proportional. This is called the *proportionality assumption*, which explains the name of the method.

Proportional hazard models are frequently used to test the null hypothesis of equality of the survival distributions of two treatments. If there is reason to believe that there is a common underlying hazard function for the two groups, a treatment group assignment term is used as a covariate and the null hypothesis that the corresponding parameter is zero is tested. If there is reason to believe that each group has its own underlying hazard, a stratified analysis may be used to test the null hypothesis that the same regression model applies to the two groups.

Cure models have been proposed to account for situations in which the event being studied cannot occur for some part of the population. Cure models were introduced in cancer studies in which time to death was the principal outcome variable, but some fraction of the population was cured. In essence, there are two populations: Those who when given the test treatment will be cured and those who will die from the disease. In a genetics context, these are the susceptible and nonsusceptible groups. In psychiatry, the study of onset of treatment effect parallels this situation. In a study of the effectiveness of an antidepressant, an important parameter is the time for the treatment to take effect. But some in the population may never respond to the treatment and therefore will never have onset. This group is analogous to the cured group in the disease model. The quantities of interest in this model are the proportion cured and among the group not cured, the conditional probability of surviving to time t. For onset, the analogous quantities are the proportion of nonresponders, and among the group of responders, the conditional probability of onset by time t. The estimation of these quantities must be made with care. A subject may survive beyond the end of the study either because he or she is cured or because he or she has a true survival time that exceeds the time that observations cease. The proportion not cured may be estimated nonparametrically by the proportion still surviving at the end of the trial, as estimated in the Kaplan–Meier approach. The Kaplan–Meier curve can also be used to solve for the conditional survival distribution of the population not cured. If two or more groups are being compared, parametric and nonparametric models can be used to test for the equality of the cure fraction and for the equality of the conditional survival distribution. The advantage of parametric cure models is that they can include covariates that affect either the proportion not cured, or the conditional survival distribution, or both. Thus, one can learn about treatment effects as well as the characteristics of a group that predicts cure as well as those that predict shorter or longer survival time.

STUDY DESIGN

Usually, in the pursuit of a research question, there are many ways to design a study to estimate parameters and to test hypotheses. In all fields of inquiry, but in medical research in particular, an experiment must be designed so as to produce the most accurate possible estimates of treatment effect within the limitations of sample size and personal safety. To use a design that is less than optimal is certainly inefficient but also may very well be unethical. The study of the design of experiments is a field of inquiry in statistics that searches for designs that are as efficient as possible. This section examines only one issue, the design of a crossover trial, as an illustration of the kind of considerations that enter the search for optimality.

Crossover Designs

There has been much interest and indeed controversy about the use of crossover designs. In a crossover trial, the comparison of t study treatments is accomplished by observing N subjects on several occasions, usually called periods. A subject may receive a different treatment in each period, or one or more treatments may be repeated, so a subject acts as his or her own control. The best-known crossover design is a two-treatment, two-period, two-sequence trial in which each subject receives both treatments, some in the order AB and the rest in the order BA. Experimental designs in which subjects are crossed over to other treatments have long been recognized as having considerable potential for achieving greater precision than can be obtained from parallel groups (also called completely randomized) designs. This is because between-subject variation is eliminated from the error term. Another advantage is that a crossover design requires fewer subjects to obtain the same number of observations. In a two-period design, each subject contributes two observations. For some types of studies, it is expensive to recruit participants, and once one has been trained in the details of the study procedure, there may be high motivation to continue to obtain observations on the response to other treatments. Although they have not been used frequently in psychopharmacology, their use may well increase as the value of crossover designs is more widely appreciated by psychopharmacologists.

The negative aspect of crossover designs is the possibility that carryover or residual effects, that is, the effects of the previous treatment, whether pharmacologic or psychological, may alter subsequent responses. The standard analysis of a two-period, two-treatment, two-sequence design (AB, BA) is rife with problems. An FDA advisory committee compared this simplest of crossover designs, in terms of precision and cost, to the completely randomized one-period design. Byron Brown, a member of the committee, concluded “the crossover experiment can yield great savings in cost if the assumption of no carryover effect is valid, but the design should not be used if this assumption is in doubt.” In many cases, the sample size needed to test whether there is carryover present exceeds the sample size needed for a similarly powered parallel groups trial. For many years, the crossover design (AB, BA) was in disfavor, particularly by researchers in the pharmaceutical industry, as a result of Brown’s analysis and the presumed concurrence of FDA statisticians. However, enlarging the number of sequences to four so that the two-period design is (AB, BA, AA, BB) avoids the carryover problems of the (AB, BA) design, albeit with diminished efficiency to separate treatments. However, expanding the number of periods results in designs without the inefficiency of the two-period designs. These results have led to considerable renewed interest in finding ways to capitalize on the known advantages of crossover designs with three or more periods without falling prey to the problem of carryover. But there are many sequences possible, even for a three-period, two-treatment design there are eight possible sequences: (AAA, AAB, ABB, BBB, BBA, BAA, ABA, BAB). If there are to be *N* participants in the trial, how should they be allocated to each sequence to achieve the maximum possible efficiency in estimating the parameters that characterize the treatment effects? In a *P* period crossover design, there are *P* responses from each subject, which naturally are correlated. If the $p \times p$ covariance matrix of the responses is known, then the optimal design can be exactly determined. But this is hardly ever the case, and ignorance of the covariance matrix greatly complicates the problem of specifying an optimal design. There are two practical approaches to solving the design dilemma. (1) Based on information from other similar studies, the researcher can assume that the correlation structure is of some specified form, from which an optimal design may be determined. Almost all medical research involving crossover designs has been conducted under some assumption as to the form of the covariance matrix. In this case, the design is called *fixed* because the entire design is specified before the clinical trial begins. (2) A clinical trial with P periods can commence, and as data accumulate the information on the first few of the N subjects to be studied can be used to estimate the covariance matrix. Treating the estimate as if it were the true covariance matrix, the optimal sequences to which the next few subjects will be allocated are determined. Their data are added to the first set of data, and a new estimate of the covariance matrix is obtained. The process continues until all N subjects are studied. For reasonable sample sizes, simulated experiments suggest that the resulting trials are nearly as efficient as the optimal design that would have been used had the covariance matrix been known. Designs in which no specific form of the covariance matrix is assumed are called *response-adaptive* because the subsequent course of the design is specified only after prior responses are obtained.

Notably, the controversy over the use of crossover designs has not subsided. Optimality, results are based not only on the assumed covariance matrix but also on the assumed response model, that is, the set of assumptions as to how carryover and response are related. Different models may yield entirely different results. For example, some authors maintain that the carryover effect that occurs if two consecutive treatments are identical, for example, AA, is not the same as the carryover effect that arises from two consecutive treatments that are different, for example, AB. Recently, statisticians have begun to consider response models that reflect this and other views, and designs that are optimal under some of these models have been found.

EVALUATING TREATMENTS UNDER CONTROLLED CONDITIONS, IN THE REAL WORLD, AND UNDER FINANCIAL CONSTRAINTS

Efficacy studies, designed to determine whether a treatment has an effect, are usually carried out under relatively pristine laboratory-like conditions with rigid adherence to detailed protocols. In contrast, *effectiveness studies* attempt to evaluate how and if a treatment works under real-world conditions. Invariably, treatment choices must involve their costs. Statistical *cost-effectiveness analysis (CEA)* provides information to help support the decision maker.

Efficacy Studies

Randomized double-blind, controlled clinical trials are the gold standard for efficacy studies, enabling, theoretically at least, an unbiased assessment of the relative merits of the study treatments and causal inference. In an efficacy study, extraneous variables that might have an impact on outcome and therefore confound the results and their interpretation are controlled to the greatest extent possible. But when experimental trials are conducted under the best of laboratory-like conditions, it must be kept in mind that the ability to generalize the findings is limited. Such trials provide little evidence as to how the study treatments will work under usual care conditions in the real world. Efficacy studies are usually conducted in a limited number of well-circumscribed patient populations, and many of the patients most likely to receive the treatment if the FDA approves it are excluded. Trial eligibility criteria may exclude, for example, patients with comorbidities, those whose symptoms are too severe or not severe enough, those suspected or known to be “noncompliant,” or those who have a history of violence. The trials may involve care conditions that are quite specialized, delivered by highly trained staff who provide individual attention far in excess of the norm. A positive finding does prove that the experimental treatment works, but the issue of efficacy beyond the well-defined populations on whom it has been tested in the well-defined environments in which the studies were conducted remains open.

Effectiveness Studies

Effectiveness studies examine how well a treatment works under real-world conditions. Many studies designed as randomized clinical trials should actually be considered to be effectiveness studies, and the environmental deviations from controlled conditions should be taken into account. For example, clinical trials of psychotropic agents are rarely conducted under pristine experimental conditions. They may take place for a short time in hospitals, but follow-up is usually in an outpatient setting. They may be based in outpatient clinics or in multiple practitioners' offices so subjects come and go from their homes. They may have multiple exposures to various unreported risk factors that impact the conditions for which they are being treated. Such conditions cannot be completely controlled. Over the course of the study, treatment groups may be differentially affected in ways that can influence outcomes to such a degree that the effect of the treatment they receive is overwhelmed.

Although randomization is preferred, this may not always be possible; many studies are conducted in which patients have been assigned to treatments in the naturalistic course of their care. Although methods for mimicking randomization (e.g., propensity scores) can be used, findings from these studies are often affected by environmental factors unrelated to the study treatment. Sometimes study treatment and locus of care are confounded because of the nature of the intervention. This is almost always true in trials that test nonpharmaceutical treatments, such as new clinical programs, screening models, or complex approaches such as supported housing. In such trials, blinding as to treatment assignment is essentially impossible so that the assumption of unbiased assessment of outcomes may be questionable.

Such situations create studies in which analysis of the data and more importantly their interpretation are substantially more complex than analyses of the traditional randomized, double-blind controlled clinical trial. Nevertheless, the value of these studies can be enormous. Besides having the potential to provide information on efficacy, a well-designed effectiveness study can discover the way to maximize the yield on the use of an efficacious agent under real-world conditions. These studies test acceptability both to consumers and to the health professionals who deliver the treatment. Outcomes of interest include measures of the treatment's acceptability to recipients and to caregivers, the treatment's impact on quality of life, and its costs. Compliance issues need to be examined with respect to events in the clients' world and to the biologic or structural elements of the intervention itself. To understand the factors that affect compliance, data should be collected on basic demographic descriptors and on substance use, work and housing conditions, social relations, and environmental conditions, including those of the caregiver and the treatment setting. Such data items need to be collected at baseline and also over the course of the study.

From a statistical perspective, multiple analyses of multiple outcomes are required. In the comparison of treatment groups based on an intent-to-treat analysis, all subjects, including noncompliers, treatment crossovers, and dropouts remain members of the group to which they were originally assigned. This analysis provides a view of how well the "policy" of assigning the study treatments works in the real world, blemishes and all. An efficacy analysis can be conducted using the same data set by identifying matched persons from each group based on predicted treatment assignment, compliance, or crossover behavior patterns. Treatment outcomes are compared within matched groups. Propensity score methods can be used to accomplish the matching. A third set of analyses might examine the impact of environmental factors on treatment acceptability and compliance.

Cost-Effectiveness

The high cost of health care provides a compelling reason to examine the costs of interventions, not just their effectiveness. CEA is an economic technique for guiding choices among alternative treatments, policies, or program-level interventions so as to maximize effectiveness under a constrained budget. In a cost-effectiveness (CE) study alternative interventions are compared in terms of the costs required to provide a particular level of effect.

For CE purposes, a program *P* is characterized by an ordered pair of positive values (ε, γ), representing respectively its mean effectiveness and its mean cost. Programs are *mutually exclusive* if only one of them may be used at one time in one patient. However, a mixture occurs when two or more mutually exclusive treatments are given to a collection of subjects. For example, if 100 patients are to receive medication treatment A or B and no person can receive both treatments because they are incompatible, then 80% receive drug A and 20% receive drug B, a mixture of programs has occurred. Programs are *independent* if any number of them may be used at one time and neither the cost nor effectiveness of any program is affected by a decision to use or not use any other program(s) in the set. A program is *dominated* if there is another program with greater or equal effectiveness and less cost. A *weakly dominated* program is one that is dominated by a mixture of programs. Within a class of mutually exclusive programs, only those that are neither dominated nor weakly dominated (the admissible class) should be considered for use. The *cost-effectiveness ratio* (CER) of a program, the price per unit of effectiveness, is the ratio γ/ε. If two programs are ranked in order of increasing cost, the *incremental cost-effectiveness ratio* (ICER) of the program with higher cost compared to the program with lower cost, the additional price per additional unit of effectiveness, is the incremental cost divided by the incremental effectiveness.

Subject to conditions on independence and mutual exclusivity among specified programs and *given the values of their mean cost and mean effectiveness*, the *deterministic* resource allocation problem of CEA is the selection of a set of programs for funding whose effectiveness is the maximum that can be achieved without incurring a cost greater than a specified fixed budget, C^*. For independent programs, resources are allocated sequentially in order of increasing CERs until the budget is exhausted. For mutually exclusive programs, *dominated* programs are eliminated and the remaining programs ranked in order of increasing effectiveness, ICERs are computed, weakly dominated programs are eliminated, ICERs are recalculated, and the process continues until all remaining programs are admissible. The remaining programs form the *frontier*, the piecewise line segments connecting adjacent admissible programs. Given a budget constraint, the program that should be funded is the one on the frontier with that budget; if the budget lies between two frontier programs budgets, then a mixture of these two is funded.

When analyzing CE from the societal perspective, the above fixed budget approach is replaced by a fixed price approach. Economists have shown that the greatest feasible health gain is achieved by implementing within each independent set of mutually exclusive programs that program whose ICER is largest among those programs that do not exceed a specified threshold value. This value, denoted by λ, is interpreted as the amount society is willing to pay (WTP) for an additional unit of health gain. Programs with ICERs above the threshold value are not implemented. The fixed budget and fixed price approaches prescribe the same allocation, except that the latter does not allow partial or mixture implementations of programs.

Statistical methods specifically for CEA began to emerge in the mid-1990s centered on methods for obtaining CIs for ICERs. The inherent difficulty of dealing with a random ratio whose denominator may assume the value 0 compounded the difficulties of this approach. Recently, the value of net health benefit (NHB) for statistical CEA has become widely accepted. The *NHB* of program k is defined as $\mathrm{NHB}_k(\lambda) = \varepsilon_k - \lambda_k/\lambda$, $k = 1, 2, \ldots, K$. At each WTP value, λ, the optimal NHB rule for mutually exclusive program funds the program with the largest positive NHB. For resource allocation, both the standard ratio-based allocation rules of CEA and the new NHB-based rule, lead essentially to the same optimal solution.

The NHB rule requires neither rank ordering of programs nor elimination of inadmissible programs, which enables statistical methods that control errors and that emulate the optimal deterministic rules of CEA. For bivariate normally distributed cost and effectiveness variables and a specified WTP, λ, the statistical procedure is based on the method of *constrained multiple comparisons with the best* (CMCB) for determining the program with the largest NHB. This method both controls the pointwise error rate at each λ and provides a confidence set for the programs that are either best or comparable to the best. The statistical frontier, plotted in the λ-NHB plane, displays the program or possibly the set of programs with the largest NHB at each WTP value. Bayesian methods for performing a CEA have also been developed.

ACKNOWLEDGMENT

This work was supported by the National Institute on Alcohol Abuse and Alcoholism [Grant number PO1AA027057-01].

FURTHER READINGS

Agresti A. *An Introduction to Categorical Data Analysis.* 3rd Ed. Wiley; 2018.

Bluman AG. *Elementary Statistics: A Step by Step Approach*. McGraw-Hill Higher Education; 2009.

*Cornfield J. A statistician's apology. *JASA*. 1975;70:7.

Fliess JL. *The Design and Analysis of Clinical Experiments*. Wiley; 2011.

Gehan EH, Lemak NA. *Statistics in Medical Research: Developments in Clinical Trials*. Springer Science & Business Media; 2012.

*Gelman A. *Bayesian Data Analysis*. CRC Press; 2013.

Hsu JC. *Multiple Comparisons, Theory and Methods*. CRC Press; 1996.

Imbens GW, Rubin DB. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press; 2015.

Jones B, Kenward M. *Design and Analysis of Cross-Over Trials*. CRC Press; 2014.

Jung T, Wickrama KAS. An introduction to latent class growth analysis and growth mixture modeling. *Soc Personal Psychol Compass*. 2008;2:302–317.

Kleinbaum DG, Klein M. *Survival analysis*. Springer; 2010.

Laska EM, Klein DF, Lavori PW, Levine J, Robinson DS. Design issues for the clinical evaluation of psychotropic drugs. In: Prien RF, Robinson DS, eds. *Clinical Evaluation of Psychotropic Drugs*. Raven Press; 1994.

Lavori PW, Laska EM, Uhlenhuth EH. Statistical issues for the clinical evaluation of psychiatric drugs. In: Prien RF, Robinson DS, eds. *Clinical Evaluation of Psychotropic Drugs*. Raven Press; 1994.

Liang KY, Zeger S. Longitudinal data analysis using general linear models. *Biometrika*. 1986;73:13.

Little R, Rubin D. *Statistical Analysis with Missing Data* (Vol. 793). John Wiley & Sons; 2019

*Lovric M. *International Encyclopedia of Statistical Science*. Springer; 2011.

McCullagh P. *Generalized Linear Models*. 3rd Ed. Routledge; 2019.

Muthén BO, Muthén LK. Integrating person-centered and variable-centered analysis: growth mixture modeling with latent trajectory classes. *Alcohol Clin Exp Res*. 2000;24:882–891.

Pearl J. Causal inference in statistics: An overview. *Statistics Surveys Volume*. 2009;96–146.

Piantadosi S. *Clinical Trials: A Methodologic Perspective*. John Wiley & Sons; 2017.

Rosenbaum PR. *Design of Observational Studies*. Springer; 2010.

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41.

*Rosenberger WF, Lachin JM. *Randomization in Clinical Trials: Theory and Practice*. John Wiley & Sons; 2015

Rubin D. *Multiple Imputation for Nonresponse in Surveys*. Wiley; 2004.

Zeger S, Liang KY, Albert P. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988;44:1049.

6

Theories of Personality and Psychopathology

6.1 Psychoanalytic Theories and Treatment: Perspective from North America and Great Britain

SYLVIA GONZALEZ CRUZ, M.D., SHWETA SHARMA, PSY.D., CYNTHIA MULDER, L.C.S.W.S., AND DEEBA ASHRAF, M.D.

Psychoanalysis is over a century old. It is a deep, independent, intellectual tradition that has continued to make significant strides in its understanding of human subjectivity. It is a body of scientific knowledge based on fundamental theoretical propositions and an ever-increasing fund of information derived mainly from intensive clinical work with patients. Whether for or against psychoanalytic ideas, the biologic psychiatrist of today cannot afford to ignore unconscious motivation, defenses, early childhood experience, dynamics of repetition compulsion, and the myriad of other psychoanalytic discoveries that complicate, if not outrightly, challenge the mainstream understanding of mental disorders.

Psychoanalysis is also a method of investigation. It is unparalleled in the depth of its inquiry into human experience. Psychoanalytic epistemology allows it to study and explore the most complex human problems and offer a perspective that complements the biologic understanding of human development and psychopathology. As a therapeutic technique, psychodynamic psychotherapy is an evidence-based approach to treating mental disorders based on psychoanalytic ideas that are well-recognized and widely practiced in several parts of the world.

Psychoanalytic theories are ambitious formulations aiming to expand our understanding of human development and suffering. They add explanations of unconscious meanings and influences. Each psychoanalytic theory makes a unique contribution regarding the understanding of unconscious mental functioning and its clinical manifestations. This chapter aims to give a basic understanding of psychoanalytic models of mind that generations of clinicians have arrived at, mainly through intensive study of private thoughts and feelings in the context of an intensive psychotherapeutic relationship. These models represent a vibrant body of ideas that reward detailed study. However, the vastness of the ever evolving and expanding psychoanalytic thinking is near impossible to summarize in a chapter. For pragmatic reasons, this chapter will focus on psychoanalytic thinking from the major North American and British traditions and leave out the developments rooted in French, German, Italian, and Latin American traditions. We hope that the introduction we provide will get interested readers excited to know more. The chapter assumes no prior knowledge of psychoanalytic theories. However, it is intended for readers interested in the clinical practice of psychoanalysis and psychodynamic psychotherapy.

To help orient the readers to extremely rich but often complex and sometimes confusing psychoanalytic ideas, we begin with what is common among various psychoanalytic theories. All psychoanalytic theories share the following common set of assumptions:

- Complex unconscious mental processes motivate and determine human behavior, affect regulation, and adaptation to the social

environment. All conscious human action is influenced by an unconscious desire for instinctual gratification or safety.
- All psychological events, including behaviors, feelings, thoughts, and actions, are influenced by and are the end result of a preceding sequence of causal events.
- All psychiatric problems can be meaningfully analyzed by considering unconsciously held beliefs and feelings that determine an individual’s reaction to their biologic, psychological, and social problems. In other words, symptoms carry multiple meanings that are hidden and unconscious.
- Intrapsychic conflict due to incompatible wishes or adverse childhood experiences is ubiquitous.
- Defense mechanisms are part of normal human development and work to protect the self against threat to safety and unpleasure.
- Interpersonal experiences are internalized and determine the representation of self and others.
- The therapeutic relationship is an essential mechanism of change in treating the suffering mind.

SIGMUND FREUD

Freud gave meaning to mental disorders by linking them to childhood experiences. He developed his theory over several decades. Three main assumptions helped Freud organize his thinking as he revised his theory into successive models of the mind —*psychological determinism, unconscious psychological processes,* and *unconscious psychological conflicts*. The assumption of psychological determinism, derived from Freud’s Helmholtzian convictions, allowed him to develop a causal psychological understanding of apparently arbitrary pathologic behavior. His belief in psychological determinism was reinforced by his clinical observation that apparently meaningless hysterical symptoms could be relieved when linked to repressed past experiences. The assumption of unconscious psychological processes allowed Freud to bring unconscious processes within reach of psychological understanding and explanation. Freud used first hypnosis and then free association to make patients aware of the repressed unconscious past experiences. He found that the unconscious mental processes differed radically from conscious behavior and thought processes in the manner of their regulation. The unconscious material is governed by the pleasure principle and the mechanisms of the primary process as opposed to the secondary process. He discovered that the unconscious psychological conflict between and among psychic forces formed the basic elements underlying neurotic difficulties.

The Trauma Hypothesis

Freud's first major psychoanalytic proposal concerned the nature of hysterical illness. From 1887 to 1897, Freud was immersed in the serious study of patients with *hysteria*. His patients, primarily women, were experiencing physical symptoms like paralysis or blindness with no apparent organic cause. In collaboration with Josef Breuer, he extended the work of French neurologist Jean-Martin Charcot and demonstrated that physical symptoms of his patients were related to the psychological experience of trauma. This approach contrasted with the contemporary understanding that hysterical illness was due to the degeneration of the nervous system. Freud's proposal assumed that his patients had experienced emotional trauma that had become blocked from their awareness because it was unacceptable to their conscious mind. He described this state of dissociated consciousness as the basis of hysterical illness. He postulated that repression of an idea from consciousness and its exclusion from any modification by association with other ideas was an essential condition for the development of symptoms of hysteria. He held that even though out of consciousness, the emotions induced by the traumatic experience continued to press for discharge into consciousness and caused the patient to develop symptoms.

Freud believed that the traumatic experience of his patients generally involved childhood or early adolescent experiences that had aroused feelings that could neither be sufficiently expressed nor resolved. Invariably, when inquiring into past histories of his patients with symptoms of hysteria, Freud found that the repressed traumatic memories had to do with sexual experiences, usually recalled in the form of a sexual seduction occurring before puberty and often relatively early in the child's experience. He concluded that these early traumatic experiences or a series of experiences had been repressed because of their painful or disagreeable nature. However, the excitation involved in sexual stimulation was not extinguished, and traces of it persisted in the unconscious in the form of repressed memories. Freud held that these memories could remain without pathogenic effect until some event, for example, a disturbing love affair, served to revive them. At that juncture, the repressive counterforce's strength was diminished, and the patient experienced what Freud termed the *return of the repressed.* The original sexual excitement was revived and found its way by a new path, allowing it to manifest itself in the form of a neurotic symptom. Specifically, in conversion hysteria, impulses that were not allowed access to consciousness were diverted into paths of somatic innervation, resulting in such hysterical symptoms as paralysis, blindness, disturbances of sensations, and other manifestations. Freud observed that the symptoms of hysteria disappeared when the traumatic event was brought to consciousness, with the patient describing the event in great detail and putting the accompanying affect into words. In this model, the symptom was viewed as a compromise between repressed wish and the ego seeking to keep the repressed idea of consciousness. The process of repression thus involved conflicting forces, the repressed idea struggling to express itself and the "dominant mass of idea constituting the ego." Moreover, the therapy allowed strangulated emotion to discharge through conscious speech by remembering the repressed trauma and integrating it into consciousness.

Based on his clinical work, Freud reconstructed the following sequence of hysteria development:

1. The patient had undergone a traumatic experience. By trauma, Freud meant an experience that stirred up intense emotion and excitation, which was intensely painful or disagreeable to the individual.
2. The traumatic experience represented to the patient some idea or ideas incompatible with the “dominant mass of ideas constituting the ego.”
3. This incompatible idea was intentionally dissociated or repressed from consciousness.
4. The excitation associated with the incompatible idea was converted into somatic pathways, resulting in hysterical manifestations and symptoms.
5. What was left in consciousness was merely a mnemonic symbol only connected with the traumatic event by associative links that were disguised.
6. If the memory of the traumatic experience could be brought into consciousness and the patient could sufficiently release the strangulated affect associated with it, the affect was discharged, and symptoms disappeared.

Freud’s early research into hysteria led him to the striking aspect of psychoanalytic understanding; the seduction experiences were of central importance for understanding the etiology of not just hysterical illness but all neurosis. Freud attributed the basic etiology of the neuroses to sexual factors. Over several years, he continued to collect clinical material that seemed to reinforce this critical hypothesis. Freud extended the basic proposition based on the studies of cases of conversion hysteria that symptoms resulted from a compromise between a repressed impulse and other repressing forces to obsessive-compulsive phenomena and even to paranoid ideation. The logical consequence of this hypothesis was that the treatment focused primarily on enabling the patient to recall repressed sexual experiences. The accompanying excitation could be allowed to find its way into consciousness and be discharged along with the revivified and previously dammed-up affects. At this early stage, Freud believed that psychotherapy achieved its curative effect by ending the emotional force of the idea that had not been sufficiently abreacted in the first instance. Of note, Freud did not regard all hysterical symptoms as psychogenic in origin, so they could not all be effectively treated by a psychotherapeutic procedure.

The trauma hypothesis proposed a physiologic model of mind. Freud regarded repression as intentional and thought it served as the basis for converting a sum of neural excitation. When cut off from more normal paths of psychic association, this sum of excitation found its way even more easily along a deviant path leading to somatic innervation. He argued that the basis for such repression must be a feeling of displeasure derived from the incompatibility between the idea to be repressed and the dominant mass of ideas constituting the ego. Freud's collaboration with Josef Breuer brought about the publication of their *Preliminary Communication* in 1893. The *Preliminary Communication* created considerable interest and was followed in 1895 by *Studies on Hysteria*, in which Breuer and Freud reported on their clinical experience in treating hysteria and proposed a theory of hysterical phenomena. The case of Anna O. provided a critical starting point for Freud's thinking and a crucial juncture in the development of psychoanalysis. Freud's case discussions proved highly significant since they formed an early expression of his early psychoanalytic thinking and technique.

Although Freud quickly abandoned the trauma hypothesis and effort to provide a neurophysiologic model of the mind, many of his prescient ideas have been confirmed by the advanced neurobiologic research of the past few decades. The impact of the trauma hypothesis on the clinical practice of psychiatry can be felt to this day. Both the search for early trauma and the release of pent-up emotions remains part of the popular understanding of psychotherapy. However, at the time, Freud assumed that his patients' reports of memories of actual experiences represented unconscious fantasies. He believed that the counterforce of repression and resistance clashed with the sexual drives and diverted them into fantasies or symptoms and unconscious repression and resistance opposed these fantasies from becoming conscious. This led him to develop the topographical model of the mind.

The Topographic Model

By the turn of the twentieth century, Freud discovered the role of the child's own psychological experience in elaborating the memories of childhood seduction and associated mental contents. He learned that his patients' accounts of forgotten but revived memories of childhood sexual involvement were sometimes fantasies related to unconscious wishes and not consistently accurate. This forced Freud to turn his attention to the inner fantasy life of the child. He began to focus on the innate drives as the motivating force of personality development and distortion of this development in psychopathology. This approach marked the significant shift of attention in psychoanalytic thinking from external events as causes of disorder to drives (biologic tension states) pressing for discharge and gratification against the resistance of consciousness. In this model, Freud understood dreams and *parapraxis*, so-called slips of the tongue, as the indirect or inadvertent expression of the blocked drives or wish fulfillment. In topographic formulation, the psychotherapeutic intervention was to lift the repression, making conscious the unconscious wish.

The Three Systems of the Mind. Beginning with the abandonment of the trauma hypothesis and the concomitant turning of Freud's interests to inner processes of fantasy and dream formation, Freud's thinking was mainly cast in terms of the topographic theory. In the topographic viewpoint, mental events were classified descriptively regarding the quality and degree of awareness. In constructing his psychological model of dreaming, Freud distinguished three layers of mind—*system unconscious*, *system preconscious*, and *system conscious*. These systems were described not in anatomical or spatial terms. Instead, they were specified according to their relationship to consciousness.

The *system unconscious* or dynamic unconscious, the deepest layer of the mind, was thought to be made up of wishes and drive representations that were in some measure unacceptable, threatening, or abhorrent to the individual. Freud held that these wishes, and drives were primarily sexual and sometimes destructive. Specifically, at this stage of theory development, Freud considered drives to consist of sexual and self-preservative (ego) drives—aggression was added later. The system unconscious was assumed to be regulated by the *primary process* demands and the *pleasure principle*. The primary process had as its principal aim the facilitation of wish fulfillments and instinctual discharge.

Consequently, it was intimately associated with—and functioned in terms of—the pleasure principle. System unconscious disregarded logical connections, permitted contradictions to coexist, recognized no negatives, and had no conception of time. The unconscious mental contents of the system are inaccessible to consciousness, except indirectly or inadvertently. Freud identified dreams as the product of primary process thinking. He also held that the quality of motility, characteristic of primary process thinking and unconscious energy, was linked to the capacity for creative thinking. The current understanding of dreams, based on the neuropsychological studies by neuropsychoanalyst Mark Solms, discussed later in this chapter, validates Freud's classical ideas in that it shows that the dreaming parts of the brain are closely related to motivational systems.

The *system preconscious*, the middle layer of the mind, was thought to act to censor the unacceptable, threatening, or abhorrent wishes. The forbidden wishes were only allowed access to consciousness if they were so distorted that their unconscious origins could not be detected. The experimental studies of perceptual defense support Freud's suggestion that the motive for censorship is perhaps the avoidance of displeasure associated with the conflictual idea. These studies have demonstrated that some people have a higher perceptual threshold for distressing stimuli than neutral words. It is assumed that subjects in these studies perceive and evaluate the word preconsciously and block it from entering consciousness.

The *system conscious*, the uppermost layer of the mind, consisted of reality-oriented perceptions, thought sequences, or problem-solving analysis dominated by secondary process thinking. The system conscious was assumed to be organized by logic and reason and was primarily concerned with external reality—safety from danger and civilized behavior.

In the topographic model, the mind conflicts with the repressed forces and the repressing forces, and the symptom is viewed as essentially a compromise between conflicting forces. Repressive ideas reach consciousness under specific conditions, such as when the repressive barrier is overpowered (as in psychoneurotic symptom formation), relaxed (as in dream states), or fooled (as in jokes). Furthermore, traumatic experience occurs when the conscious part of the mind is overwhelmed by pressure to gratify unconscious wishes, followed by intolerable feelings of rejection or punishment. The shift in emphasis on fantasy over reality led Freud to explain all human behavior in terms of the failure of the child's mind to deal adequately with the pressures of biologic drives. Freud discovered during his clinical work that repression of a childhood memory could occur if energy were withdrawn from it and, especially if verbal energy were removed. This repressive force or *countercathexis* manifests itself in psychoanalytic treatment as resistance to remembering. When the words were reconnected to the forgotten memory traits (as during psychoanalytic treatment), it became recathected and could thus reach consciousness once more.

Psychosexual Development. The shift in emphasis from reality to inner motivations and fantasy marked the beginning of psychoanalysis. Psychoanalysis took on a new and highly significant dimension to distinguish psychic reality and fantasy from actual external events. This shift in direction inevitably led to the emergence of a dynamic theory of infantile sexuality in which the child's psychosexual life played a significant and dominant role. According to the theory, sexual instincts seek expression throughout life, which the individual tries to master to conform to society's demands. These biologic tension states are represented mentally in terms of wishes and are directed towards external objects for their satisfaction.

The earliest manifestations of infantile sexuality arose with bodily functions regarded as basically nonsexual, such as feeding and developing bowel and bladder control. However, Freud saw that these functions involved degrees of sensual pleasure, which he interpreted as forms of psychosexual stimulation. Freud used the term *sexual* in an extended sense to mean physically pleasurable. He identified a succession of three developmental phases in developing infantile urges, each of which was thought to build on and subsume accomplishments of the preceding phases—namely, the *oral*, *anal*, and *phallic phases*. The oral phase occupied the first 12 to 18 months of the infant's life; next came the anal phase, lasting until about 3 years of age; finally, the phallic phase, from approximately 3 to 5 years of age. Later, the *latency phase,* from ages 6 to puberty, and the *genital phase*, from puberty onwards, were added to complete the picture.

Freud postulated that phallic erotic activity was essentially a preliminary stage for adult genital activity in boys. In contrast to the male, whose principal sexual organ remained the penis throughout psychosexual development, the female had two leading erotogenic zones, the clitoris, and the vagina. Freud felt that the clitoris was preeminent during the infantile pregenital period but that after puberty, erotic primacy was transposed to the vagina. Recent sexual investigations have cast doubt on a supposed transition from clitoral to vaginal primacy.

Karl Abraham modified and refined Freud's basic schema of the psychosexual stages. He further subdivided the phases of libido development, dividing the oral period into a sucking and biting phase and the anal phase into a destructive–expulsive (anal sadistic) and a mastering–retaining (anal erotic) phase. Finally, he hypothesized that the phallic period consisted of an earlier phase of partial genital love, which was designated as the true phallic phase, and a later, more mature genital phase. For each of the stages of psychosexual development, Freud delineated specific erotogenic zones that gave rise to erotic gratification.

The theory has evolved based on direct empirical and experimental observations of children in child analysis and developmental studies as opposed to relying on the reconstruction of childhood memories in adult analysis. It now focuses less on libidinal phase specificity with a progression of libidinal stages from oral to genital in a prescribed order, and places greater emphasis on the complex integration of multiple developmental influences. These include maturational factors, temperamental dispositions, object relations involvement, affective and cognitive development, language acquisition and related phenomena. Accordingly, there is a greater inclination to view libidinal stages as more loosely organized, intermingled, and not necessarily rigidly sequential.

The topographic theory assumed the existence of *psychological energies* that originated in instinctual drives. This assumption was derived from the observation that recall of traumatic experiences and their accompanying affects resulted in the disappearance of symptoms and anxiety. This suggested, therefore, that a displaceable and transformable quantity of energy was involved in the psychological processes responsible for symptom formation. Freud originally assumed that this quantity was equivalent to the affect, which became dammed up or strangulated when it was not appropriately expressed and, thus, was transformed into anxiety or conversion symptoms. After he had developed his notion of instinctual drives, this quantitative factor was conceived of as drive energy *(cathexis*). The assumption of psychic energies served Freud as a vital heuristic metaphor. The usefulness of the metaphor and its necessity as a basic assumption of analytic theory has subsequently been questioned and found wanting. Refocusing these observations in motivational terms loses nothing in terms of explanatory potential.

Freud’s deepening appreciation of the unconscious dimension of these processes led him to view the ego as, in some part unconscious, one of the reasons for his later formulation of the structural theory in 1923.

Structural Model

Freud gradually realized the limits of the topographic model. The repeated encounters of resistance in his patients, difficulty accessing the defenses responsible for resistance, and feelings of guilt in his patients led him to conclude that the mechanism of repression was itself unconscious. These observations made clear to Freud that the ego has both conscious and unconscious components. Freud rethought his topographic model and conceived of the structural model of the mind, or the *tripartite theory* as it is often called.

Freud first introduced the structural model of the mind in *The Ego and the Id* (1923). The structural model was composed of three distinct entities or organizations within the psychic apparatus—the *id*, the *ego*, and the *superego*. Freud conceived of the id as a completely unconscious, unorganized, primordial reservoir of energy derived from the sexual and aggressive instincts and under the domination of the pleasure principle and primary process. However, id was not synonymous with the unconscious because certain functions of the ego, specifically certain defenses against unconscious instinctual pressures, were also unconscious; and for the most part, the superego also operated on an unconscious level.

In contrast to his concept of the id, Freud conceived of the ego as a coherent structure, having an organized, problem-solving capacity, whose task was to mediate between the competing demands of the id, superego, and the external reality. The ego was partly conscious but largely unconscious. The conscious part of the ego has the capacity for sophisticated psychological processes to deal with the external reality. The unconscious part of the ego regulated pressure and demands of the id and the superego by the mechanisms of defense that operate unconsciously.

The third structural component of the psychic apparatus, the superego, contained the internalized moral values, ideals, prohibitions, and standards of the childhood parental authority figures. In Freud’s view, the superego was the source of conscience and guilt. It was concerned with the ideals and morals based on unconscious learning from parents and hence society. The superego was partly conscious but largely unconscious.

The structural model assumed that most psychological processes could function outside of awareness. Freud’s view of mental processes as largely unconscious has been validated and supported by cognitive and experimental psychology. Freud believed that the ego develops out of modifications of the id, resulting from the demands and limits of the external world. The ego appropriates energies of the id to manage the pressures of the external reality. His observation that repression and resistance, both associated with the ego, could be unconscious, pointed to the need for an expanded concept of the ego. The ego was considered an organization retaining a close relationship to consciousness and external reality and yet performing a variety of unconscious operations in relation to drives and their regulation. With a broadened scope for the ego, consciousness was redefined as a mental quality that, although exclusive to the ego, constitutes only one of its qualities or functional aspects, rather than a separate mental system itself as in the topographic model. The ego controls the apparatuses of motility and perception, contact with reality, and, through defense mechanisms, inhibition, and control of primary instinctual drives.

While the ego developed out of frustrated id wishes, the superego developed out of defense against the id wishes. In the Oedipal period, the child wishes to possess the parent. Each must, however, contend with a powerful rival, the other parent. The frustration of the child’s positive Oedipal wishes by this parent evokes intense hostility, which finds expression not only in overt antagonistic behavior but also in thoughts of killing the parent who stands in the way, along with any brothers or sisters who may also compete for the love of the desired parent. Quite understandably, this hostility on the part of the child is unacceptable to parents and, in fact, eventually becomes unacceptable to the child as well. The moral objection to the id wishes form the basis of the superego. The theory contends the child’s sexual explorations and masturbatory activities may themselves meet with parental disfavor, which may even be underscored by real or implied threats of castration. Consequently, the child turns away from the Oedipal situation and its emotional involvements and enters the latency period of psychosexual development and renounces the sexual impulses of the infantile phase. It was this effort of renunciation that gave the superego its prohibiting character. However, because the superego evolved due to the repression of instinctual desires, it had a closer relation to the id than the ego itself. Its origins were more internal; the ego originated to a greater extent in response to the external world and was its internal representative.

Throughout the latency period and thereafter, the child (and later the adult) continues to build on these early identifications through contact with teachers, heroic figures, and admired persons, who form the sources of the child's moral standards, values, and ultimate aspirations and ideals. The child's conflicts with parents continue but largely internally, between the ego and superego. In other words, the child's superego internalizes the parental standards, restrictions, commands, and punishments. The superego becomes the judge and guide of behavior which operates from within, even in the absence of the parents.

EGO PSYCHOLOGY

Freud's ego was conceived as a structural entity, a coherent organization of mental processes and functions, primarily organized around the Perception–Consciousness System, which also included structures responsible for resistance and unconscious defense. However, the ego was viewed as relatively passive and weak. Its functioning was still a result of mediating the pressures deriving from the id, superego, and external reality. The ego was the helpless rider on the id's horse, more or less obliged to go where the id wished to go. The assumption remained that the ego was dependent on forces of the id and somehow genetically derived and differentiated out of the id. Freud had yet to recognize any actual development of the ego comparable to the phases of libidinal development.

However, Freud's conception of ego as a powerful regulatory force responsible for integrating and controlling behavioral responses, and the concern with the adaptive function of the ego, even though rudimentary and limited to the ego's capacity to avoid danger, provided the foundation for the refinement and advancement of ego psychology. The systematization of ego psychology stemmed from the collective works of Anna Freud, Heinz Hartmann, David Rapaport, and Charles Brenner, to name but a few giants of the ego psychology movement.

Ego Functions

Ego psychologists viewed ego as the preeminent psychic agency, as important as the drives, and studied and theorized the role of ego and its various functions as individuals adapt to their physical and social environment or modify those environments to make them more compatible with personal wishes and needs. The ego was still a subsystem of the personality and not synonymous with the self, the personality, or the character. The ego comprised a class of functions that mediated between instincts and the outside world and we share some of the important functions here.

- Ego *mediates the relation with reality*. The sense of reality originates with the development of the ego. The ability to distinguish internal reality (wishes and fears) and actual reality (events in the real world) depends first on the *primary autonomous functions* of the ego, such as memory and perception. Because of the fundamental importance of *reality testing* for "negotiating" with the outside world, its impairment is associated with a severe mental disorder.
- Ego *makes the judgment* about what is "reasonable" or "appropriate" course of action. It considers id wishes, as well as social expectations and reality considerations to arrive at sound judgment. The capacity for judgment works in tandem with reality testing.
- Ego *controls and regulates instinctual drives*. The capacity to delay the immediate discharge of urgent wishes and impulses helps the ego mediate between the id and the outside world.

Modulation of impulses involves judgment. Both too much control and too little control over impulses is problematic.

- Ego *regulates affect* by keeping out of awareness painful or unacceptable feelings, or by modulating the expression of feelings. Ego matches the feelings with the social and cultural norms to make emotional expression congruent.
- Ego *helps create mutually satisfying relationships*. The evolution in the child's capacity for relationships with others, progressing from initial relations with maternal and other caretaking figures to social relationships within the family and then to relationships within the larger community, is related to this capacity.
- Ego *mobilizes defenses* to bring instinctual demands into conformity with demands of external reality. Defenses operate unconsciously, in response to danger signals like, threatening instinctual motivations and affects. Defenses are dynamically motivated and reversible. They could be functionally adaptive as well as pathologic.
- Ego *helps maintain a steady and reasonable level positive self-regar*d by regulating negative and positive feeling states that may dysregulate *self-esteem*.
- Ego *masters challenges of each successive psychosexual stage* (oral, anal, phallic, genital) and thereby becomes more effective in organizing and synthesizing mental processes.
- *Adaptation to reality* is closely allied to the concept of mastery, both in respect to instincts and external tasks. The function of adaptation to reality is closely related to the defensive functions of the ego.
- Ego's *synthetic function* is concerned with the overall organization and functioning of the ego in the self-system. It unites, organizes, and binds together various drives, motives, tendencies, and functions within the personality, enabling the individual to think, feel, and act in an organized and directed manner. As part of its synthetic function ego must enlist the cooperation of other ego and nonego functions in its operation. Although the *synthetic function* subserves adaptive functioning in the self, it may also bring together various forces that, although

not wholly adaptive, are an optimal solution for the individual in a particular state at a given moment.

- Clinical example: Hysterical conversion combines a forbidden wish and the punishment for it into a physical symptom. On examination, the symptom often turns out to be the only possible compromise under the circumstances.

Ego Development

In addition to the maturation of relatively autonomous ego functions, the ego is built upon the basis of processes by which aspects of the external world are acquired and become part of the ego. These processes where the internal world is built, and the structure is consolidated within the self are called *internalization*. Forms of internalization—*incorporation, introjection,* and *identification*—are variously connected with the development of the ego.

Incorporation was initially conceived of as an instinctual activity derived from and based developmentally on the oral phase and was considered as a genetic precursor of identification. Incorporation as an internalization mechanism involves a primitive oral wish for union with the object. The union is total and global. The object loses all distinction and independent function and is completely assumed into the person's inner world. Incorporation thus comes into play in relatively more infantile or regressive conditions in which the sense of distinction of the object as separate is lost.

Introjection is perhaps the most central process in developing the structural apparatus involving ego and superego. Freud originally described introjection in *Mourning and Melancholia* (1917) as a process of narcissistic identification in which the lost object is introjected and thus retained as a part of the internal structure of the psyche. Freud later applied this mechanism to the genesis of the superego, making introjection the primary internalizing mechanism by which parental *imagos* were internalized at the close of the Oedipal phase. Through introjection, the child tries to retain gratifications derived from object relationships, at least in fantasy. By this mechanism, qualities of the gratifying relationship are internalized and re-established as part of the self's organization. In doing so, the child retains connection with the object, which serves both adaptive and defensive functions. Freud referred to this internalized product as a *precipitate of abandoned object cathexis.* However, because of the degree of residual object-connection, introjects are not fully integrated into the structuring of ego and superego; they retain a quasi-independence and inner presence. They have been aptly described in Kleinian terms as *internal objects*.

Identification has often been confused with introjection, partially because Freud treated the two processes as overlapping and somewhat interchangeable. Identification is an active structuralizing process within the self, by which the self-constructs the inner constituents of regulatory control. What constitutes the model of identification can vary considerably and can include introjects, structural aspects of real objects, or even regulatory components of group structures and group cultures. The identification process is specifically an intrasystemic structuralizing activity, attributed to the ego functions of the self and related to its synthetic function, affecting structural integration in all parts of the psychic apparatus, including ego and superego. By identification, the character of these structures is internally altered, such that their qualities can be regarded as authentic aspects of the self that effectively modifies aspects of autonomous self-functioning.

Heinz Hartmann's Ego Psychology

Hartmann's ego psychology advanced Freud's structural theory by pointing out that some of the functions of the ego develop autonomously of the id. He referred to these ego mechanisms as apparatuses of *primary autonomy*. He postulated that the ego and id originate from a common matrix, the so-called undifferentiated phase, in which the ego's precursors are inborn apparatuses of primary autonomy. These apparatuses are rudimentary, present at birth, and develop outside the area of conflict with the id. Hartmann referred to this as a *conflict-free zone* of ego functioning. He included perception, intuition, comprehension, thinking, language, certain phases of motor development, learning, and intelligence among the functions in this *conflict-free sphere*.

With the introduction of the *primary autonomous functions*, Hartmann provided an independent genetic derivation for at least part of the ego, thus establishing it as an independent realm of psychic organization and functioning that was not entirely dependent on and derived from the instincts. This was a formulation of significant importance because it laid the foundations for the emerging doctrine of ego autonomy and meant that the analysis of ego development would have to consider an entirely new set of variables quite separate from those involved in instinctual development.

Over time it became increasingly clear that each of the autonomous functions of the ego could become involved in conflict secondarily during development and usually did. For example, aggressive, competitive impulses could intrude on the impulse to learn, and evoke inhibitory defensive reactions on the part of the ego, thus interfering with the conflict-free operation of these functions. Hartmann also recognized that ego structures and functions, arising in conflict, could change function to become relatively autonomous from drives. He observed that the conflict-free sphere could be enlarged, and other ego functions could be withdrawn from the domination of drive influences. This was Hartmann’s concept of *secondary autonomy*. A mechanism that may arise initially in the service of defense against instinctual drives may, in time, become an independent structure, such that the drive impulse merely triggers the automatized apparatus. Thus, the apparatus may serve other functions than the original defensive function, such as adaptation or synthesis. Hartmann referred to this removal of specific mechanisms from drive influences and becoming relatively autonomous as a change of function. This view of adult behavior as having different functions, likely independent of their instinctual origins, continues to be relevant today. Hartmann’s treatment of adaptation brought the adaptational point of view into focus and was generally accepted as one of the basic metapsychological assumptions of psychoanalytic theory.

Another crucial aspect of Hartmann’s ego psychology is the importance of reality, in its broadest and most profound meanings, as a significant dimension of psychoanalytic thinking. This is a direct extrapolation of Hartmann’s thinking about adaptation. The adaptive functioning of the organism has directly to do with fitting in with the requirements of external reality and adaptively interacting with the environment, not only the inanimate but also the personal and social environment. Hartmann and his colleagues introduced the concept of *average expectable environment* which recognized the importance of actual parents and thereby the influence of the environment on the development. Along with affirming the importance of the relation between the individual and their environment, ego psychology, following Hartmann’ distinction between self and other, furthered the interest in the self in so far as relations with the outside world occur between self and other. The psychoanalytic interest in self was expanded and advanced by Heinz Kohut’s and other more structurally oriented views of the self.

Anna Freud’s Ego Psychology

Like Hartmann, Anna Freud too emphasized the importance of analyzing the ego. She developed a theory of child development which was carried further by her colleagues like Erik Erikson, Edith Jacobson, and Margaret Mahler. With her interest in child analysis, she emphasized the importance of analyzing the issues that arise in a child. In the absence of conflict, id, ego, and superego work in harmony and are hard to distinguish. However, conflict generates anxiety, which acts as a signal for the defensive function of the ego and brings forth the imperatives of id and superego. Anna Freud and her group worked to list and categorize these defenses. Her efforts opened the way to a more systematic and comprehensive study of ego defenses. In her classic monograph, *The Ego and the Mechanisms of Defense* (1936), she maintained that everyone, whether normal or neurotic, uses a characteristic repertoire of *defense mechanisms*, but to varying degrees. Based on her extensive clinical studies of children, the primacy of drives at various developmental stages evoked anxiety in the ego. This anxiety, in turn, produced a variety of defenses. Concerning adults, her psychoanalytic investigations led her to conclude that although resistance was an obstacle to progress in treatment, to the extent that it impeded the emergence of unconscious material, it also constituted a valuable source of information concerning the ego's defensive operations.

Defenses emerge early in development to help the ego mediate pressures of the id and the requirements and strictures of outside reality. At each phase of libidinal development, associated drive components evoke characteristic ego defenses. For example, introjection, denial, and projection are defense mechanisms associated with oral-incorporative or oral-sadistic impulses, whereas reaction formations, such as shame and disgust, usually develop with anal impulses and pleasures. Defense mechanisms from earlier phases of development persist side by side with those of later periods. When defenses associated with the pregenital phase of development tend to predominate in adult life over more mature mechanisms, such as sublimation and repression, the personality retains an infantile cast.

The defensive forms of ego functioning can be categorized in various ways. Defenses may be classified developmentally by the libidinal phase in which they arise or are associated. Denial, projection, and distortion are assigned to the oral stage of development and the correlative narcissistic stage of object relationships. However, defenses, such as magical thinking and regression, cannot be categorized in this way. Moreover, specific basic developmental processes, such as introjection and projection, may also serve defensive functions under certain specifiable conditions. The defenses have also been classified based on their commonly associated form of psychopathology. The obsessional defenses include isolation, rationalization, intellectualization, and denial. However, defensive operations are not limited to pathologic conditions. Finally, the defenses have been classified as to whether they are simple mechanisms or complex, in which a single defense would involve a combination or composite of simple mechanisms.

Table 6.1–1 gives a brief classification and description of some of the basic defense mechanisms most frequently employed and most thoroughly investigated by psychoanalysts.

Table 6.1–1.
Classification of Defense Mechanisms

Narcissistic-Psychotic Defenses

These defenses are usually found as part of a psychotic process, but may also occur in young children and adult dreams or fantasies. They share the common note of avoiding, negating, or distorting reality.

Projection	Perceiving and reacting to unacceptable inner impulses and their derivatives as though they were outside the self. On a psychotic level, this takes the form of frank delusions about external reality, usually persecutory, includes both perception of one’s own feelings in another, with subsequent acting on the perception (psychotic paranoid delusions). Impulses may derive from id or superego (hallucinated recriminations).
Denial	Psychotic denial of external reality, unlike repression, affects perception of external reality more than perception of internal reality. Seeing, but refusing to acknowledge what one sees, or hearing, and negating what is actually heard, are examples of denial and exemplify the close relationship of denial to sensory experience. Not all denial, however, is necessarily psychotic. Like projection, denial may function in the service of more neurotic or even adaptive objectives. Denial avoids becoming aware of some painful aspect of reality. At the psychotic level, the denied reality may be replaced by a fantasy or delusion.
Distortion	Grossly reshaping the experience of external reality to suit inner needs, including unrealistic megalomanic beliefs, hallucinations, wish-fulfilling delusions, and employing sustained

	feelings of delusional grandiosity, superiority, or entitlement.

Immature Defenses

These mechanisms are fairly common in preadolescent years and in adult character disorders. They are often mobilized by anxieties related to intimacy or its loss. Although they are regarded as socially awkward and undesirable, they often moderate with improvement in interpersonal relationships or with increased personal maturity.

Acting out	The direct expression of an unconscious wish or impulse in action to avoid being conscious of the accompanying affect. The unconscious fantasy, involving objects, is lived out and impulsively enacted in behavior, thus gratifying the impulse more than the prohibition against it. On a chronic level, acting out involves giving in to impulses to avoid the tension that would result from postponement of their expression.
Blocking	An inhibition, usually temporary in nature, of affects especially, but possibly also thinking and impulses. It is close to repression in its effects, but has a component of tension arising from the inhibition of the impulse, affect, or thought.
Hypochondriasis	Transformation of reproach toward others arising from bereavement, loneliness, or unacceptable aggressive impulses, into self-reproach in the form of somatic complaints of pain, illness, and so forth. Real illness may also be overemphasized or exaggerated for its evasive and regressive possibilities. Thus, responsibility may be avoided, guilt may be circumvented, and instinctual impulses may be warded off.
Introjection	In addition to the developmental functions of the process of introjection, it also can serve specific

	defensive functions. The introjection of a loved object involves the internalization of characteristics of the object with the goal of ensuring closeness to and constant presence of the object. Anxiety consequent to separation or tension arising out of ambivalence toward the object is thus diminished. If the object is lost, introjection nullifies or negates the loss by taking on characteristics of the object, thus in a sense internally preserving the object. Even if the object is not lost, the internalization usually involves a shift of cathexis reflecting a significant alteration in the object relationship. Introjection of a feared object serves to avoid anxiety through internalizing the aggressive characteristic of the object, and thereby putting the aggression under one's own control. The aggression is no longer felt as coming from outside, but is taken within and utilized defensively, thus turning the subject's weak, passive position into an active, strong one. The classic example is "identification with the aggressor." Introjection can also take place out of a sense of guilt in which the self-punishing introject is attributable to the hostile–destructive component of an ambivalent tie to an object. Thus, the self-punitive qualities of the object are taken over and established within one's self as a symptom or character trait, which effectively represents both the destruction and the preservation of the object. This is also called identification with the victim.
Passive–aggressive behavior	Aggression toward an object expressed indirectly and ineffectively through passivity, masochism, and turning against the self.

Projection	On a nonpsychotic level, projection involves attributing one’s own unacknowledged feelings to others; it includes severe prejudice, rejection of intimacy through suspiciousness, hypervigilance to external danger, and injustice collecting. Projection operates correlatively to introjection, such that the material of the projection derives from the internalized but usually unconscious configuration of the subject’s introjects. At higher levels of function, projection may take the form of misattributing or misinterpreting motives, attitudes, feelings, or intentions of others.
Regression	A return to a previous stage of development or functioning to avoid the anxieties or hostilities involved in later stages. A return to earlier points of fixation embodying modes of behavior previously given up. This is often the result of a disruption of equilibrium at a later phase of development. This reflects a basic tendency to achieve instinctual gratification or to escape instinctual tension by returning to earlier modes and levels of gratification when later and more differentiated modes fail or involve intolerable conflict.
Schizoid fantasy	The tendency to use fantasy and to indulge in autistic retreat for the purpose of conflict resolution and gratification.
Somatization	The defensive conversion of psychic derivatives into bodily symptoms; tendency to react with somatic rather than psychic manifestations. Infantile somatic responses are replaced by thought and affect during development (desomatization); regression to earlier somatic forms or response (resomatization) may result

	from unresolved conflicts and may play an important role in psychophysiologic and psychosomatic reactions.

Neurotic Defenses

These are common in apparently normal and healthy individuals as well as in neurotic disorders. They function usually in the alleviation of distressing affects and may be expressed in neurotic forms of behavior. Depending on circumstances, they can also have an adaptive or socially acceptable aspect.

Controlling	The excessive attempt to manage or regulate events or objects in the environment in the interest of minimizing anxiety and solving internal conflicts.
Displacement	Involves a purposeful, unconscious shifting of impulses and/or affective investment from one object to another in the interest of solving a conflict. Although the object is changed, the instinctual nature of the impulse and its aim remain unchanged.
Dissociation	A temporary but drastic modification of character or sense of personal identity to avoid emotional distress; it includes fugue states and hysterical conversion reactions.
Externalization	A general term, correlative to internalization, referring to the tendency to perceive in the external world and in external objects components of one’s own personality, including instinctual impulses, conflicts, moods, attitudes, and styles of thinking. It is a more general term than projection, which is defined by its derivation from and correlation with specific introjects.
Inhibition	The unconsciously determined limitation or

	renunciation of specific ego functions, singly or in combination, to avoid anxiety arising out of conflict with instinctual impulses, superego, or environmental forces or figures.
Intellectualization	The control of affects and impulses by way of thinking about them instead of experiencing them. It is a systematic excess of thinking, deprived of its affect, to defend against anxiety caused by unacceptable impulses.
Isolation	The intrapsychic splitting or separation of affect from content resulting in repression of either idea or affect or the displacement of affect to a different or substitute content.
Rationalization	A justification of attitudes, beliefs, or behavior that might otherwise be unacceptable by an incorrect application of justifying reasons or the invention of a convincing fallacy.
Reaction formation	The management of unacceptable impulses by permitting expression of the impulse in antithetical form. This is equivalently an expression of the impulse in the negative. Where instinctual conflict is persistent, reaction formation can become a character trait on a permanent basis, usually as an aspect of obsessional character.
Repression	Consists of the expelling and withholding from conscious awareness of an idea or feeling. It may operate either by excluding from awareness what was once experienced on a conscious level (secondary repression), or it may curb ideas and feelings before they have reached consciousness (primary repression). The “forgetting” associated with repression is unique in that it is often accompanied by highly

	symbolic behavior, which suggests that the repressed is not really forgotten. The important discrimination between repression and the more general concept of defense has been discussed.
Sexualization	The endowing of an object or function with sexual significance that it did not previously have, or possesses to a lesser degree, to ward off anxieties connected with prohibited impulses.

Mature Defenses

These mechanisms are healthy and adaptive throughout the life cycle. They are socially adaptive and useful in the integration of personal needs and motives, social demands, and interpersonal relations. They can underlie seemingly admirable and virtuous patterns of behavior.

Altruism	The vicarious but constructive and instinctually gratifying service to others, even to the detriment of the self. This must be distinguished from altruistic surrender, which involves a masochistic surrender of direct gratification or of instinctual needs in favor of fulfilling the needs of others to the detriment of the self, with vicarious satisfaction only being gained through introjection.
Anticipation	The realistic anticipation of or planning for future inner discomfort: Implies overly concerned planning, worrying, and anticipation of dire and dreadful possible outcomes
Asceticism	The elimination of directly pleasurable affects attributable to an experience. The moral element is implicit in setting values on specific pleasures. Asceticism is directed against all “base” pleasures perceived consciously, and gratification is derived from the renunciation.

Humor	The overt expression of feelings without personal discomfort or immobilization and without unpleasant effect on others. Humor allows one to bear, and yet focus on, what is too terrible to be borne, in contrast to wit, which always involves distraction or displacement away from the affective issue……….
Sublimation	The gratification of an impulse whose goal is retained, but whose aim or object is changed from a socially objectionable one to a socially valued one. Libidinal sublimation involves a desexualization of drive impulses and the placing of a value judgment that substitutes what is valued by the superego or society. Sublimation of aggressive impulses takes place through pleasurable games and sports. Unlike neurotic defenses, sublimation allows instincts to be channeled rather than to be dammed up or diverted. Thus, in sublimation, feelings are acknowledged, modified, and directed toward a relatively significant person or goal so that modest instinctual satisfaction results.
Suppression	The conscious or semiconscious decision to postpone attention to a conscious impulse or conflict.

Data from Vaillant GE. *Adaptation to Life*. Little Brown; 1977; Semrad E. The operation of ego defenses in object loss. In: Moriarity DM, ed. *The Loss of Loved Ones*. Charles C Thomas; 1967; Bibring GL, Dwyer TF, Huntington DS, Valenstein AA. A study of the psychological principles in pregnancy and of the earliest mother–child relationship: Methodological considerations. *Psychoanal Stud Child*. 1961;16:25.

Anna Freud’s view of drives was rooted in the structural model. She gave primacy to instinctual satisfaction over the role of the object. However, she took the theory forward because she was very interested in the role of real parents in shaping the development of the ego and the superego. Her theory of development was significantly more comprehensive and emphasized the interaction between drive and ego–superego development and the environmental determinants. She used the notion of developmental lines to explore the complexity of structural development. The concept of developmental lines offered a way to understand a child's behavior in its developmental context. The child's developmental profile was created, and developmental lines were used to understand problematic behavior in terms of developmental arrest or regression. Clinically, her formulation added the provision of *developmental help* to psychoanalytic work. Addressing both conflict and deficit to restore the path of normal development allowed for a more comprehensive view of development. Anna Freud's work foreshadowed the work of generations of developmental theorists that attended to pathogenic effects of childhood adversity. The notion of developmental lines and focus in clinical work on providing developmental help is reflected in the theory of mentalization and mentalization-based treatment.

Based on her work with children, Anna Freud proposed several categories of childhood disorders. These included a range of developmental disturbances such as regressive dependence, poorly controlled aggression, narcissism, object relatedness, and neurotic compromise formation like phobias and obsessions. She developed a model of anxiety that distinguished between fear of the external world (actual reactions of parents and external events) and fear of the internal world. She and her colleague, Dorothy Burlingham, identified four separate internal sources of anxiety—*annihilation anxiety*, *anxiety related to loss of object*, *anxiety related to loss of objects' love*, and *anxiety related to severe bodily injury or castration*. She linked these anxieties to development, specifically capacities of the ego. A less mature ego is unable to mentalize, regulate, or defend against the psychic pain. The infant research does not support the assumption of an infant's helplessness in the face of primitive anxiety.

MARGARET MAHLER'S DEVELOPMENTAL MODEL

The separation–individuation process advanced by Margaret Mahler and her associates was a significant empirically-based contribution to understanding the developmental process after Freud. Mahler's theory emphasizes the process of separation from the maternal orbit and the establishment of personal autonomy. Her theory, organized in terms of *phases of separation and individuation*, came under severe criticism by later developmentalists. In contrast, developmentalists, mainly following the inspiration of self-psychology and intersubjectivism, emphasized maternal dependence in development and the continuing dependent reliance on self-objects throughout the life cycle. However, these divergent emphases may not be exclusive insofar as mature autonomy does not rule out significant dependence in object relations, nor does mature dependence exclude the possibility of meaningful autonomy.

Autistic Phase

The first phase of Mahler's theory of development described the *autistic phase*: "During the first few weeks of extrauterine life, a stage of absolute primary narcissism, marked by the infant's lack of awareness of a mothering agent, prevails. This is the stage termed *normal autism*. It is followed by a dim awareness stage that needs satisfaction that cannot be provided by oneself but comes from somewhere outside the self." During the first month of life, the infant is not aware of the mother as a separate entity and is not aware that it cannot meet its own needs.

Symbiotic Phase

In the next stage, the *symbiotic phase*, Mahler articulated the origin of the initial differentiation of self and object, in which infants can be said to experience something outside of themselves, to which they can relate, as satisfying their inner needs. This dawning awareness of the external object is a significant state in the psychological development of children and involves not only cognitive and perceptual developments but also goes together with the organization of rudimentary infantile drives and affects concerning emerging object experiences. Mahler indicates that this awareness signals the beginning of *normal symbiosis* “in which the infant behaves and functions as though he and his mother were an omnipotent system—a dual unity within one common boundary.” Boundaries become temporarily differentiated only in the state of *affect hunger* but disappear again when the need is gratified. Consequently, the object is recognized as separate from the self only at moments of need so that, once the need is satisfied, the object ceases to exist—from the infant’s (subjective) point of view—until a need again arises. At this stage, physical need satisfaction and psychological need satisfaction are experienced as the same. It is not until later in development that these two processes separate.

Separation–Individuation

The *separation–individuation process* is divided into successive phases or periods—the *hatching* period, the *practicing* period, *rapprochement*, and the development of *object constancy*—describing the gradual separation from maternal dependency and the increasing autonomy of the child.

Hatching. During the *hatching* period, the child gradually differentiates out of the symbiotic matrix with effort. The first behavioral signs of such differentiation seem to arise at about 4 or 5 months and continue through about 9 months. During this time, the infant is more aware of the surrounding environment and is curious about these surroundings. The infant explores the primary caregiver as a separate entity and, becoming more mobile, starts to explore the world around them. What started as a very internal preoccupation now becomes external, focusing on the primary caregiver.

Practicing. As hatching and separation from the primary caregiver gradually increase, there is a move to the second or *practicing* subphase of separation–individuation. The practicing period can be usefully divided into an early period and a proper period. The early practicing phase begins with the infant's earliest ability to move physically away from the primary caregiver by locomotion: crawling, creeping, climbing, and assuming an upright sitting position. However, moving away from the safe protective orbit of the primary caregiver has its risks and uncertainties. In the early practicing phase, there is frequently a pattern of visually "checking back to mother" or even crawling or paddling back to her to touch or hold on as a form of "emotional refueling."

The attainment of free upright locomotion characterizes the practicing period proper. It is marked by three interrelated developments that contribute to the continuing process of separation and individuation. These are (1) rapid bodily differentiation from the mother (primary caregiver), (2) establishment of a specific bond with her, and (3) growth and functioning of autonomous ego apparatuses in close connection and dependence on the mothering figure.

Rapprochement. As testing of the freedom of individuation proceeds, by about the middle of the second year, the child enters the third subphase of *rapprochement*. The crisis in the rapprochement phase is mainly that of *separation anxiety*. The child's wishes and desires to be separate, autonomous, and omnipotent are tempered by increasing awareness of the need for and dependence on the mother. Ambivalence is also characteristic of the middle phase of the rapprochement subphase. Thus, the mother's availability and the reassurance of her continuing love and support become all the more critical. If this crisis is not resolved well, there will be an increase in extreme clinging or shunning behavior.

Object Constancy. As the conflicts and rapprochement crisis are gradually resolved, the child enters the final phase of separation and individuation; namely, the phase of consolidation of individuality and the beginnings of emotional *object constancy*. At this stage, there are significant developments in the structuralization and integration of the ego and definite signs of internalization of parental demands, reflecting the development of superego precursors.

Attainment of object constancy marks a transition from the stage of need-satisfying relationships to a more mature psychological involvement with objects. Object constancy implies a capacity to differentiate between objects and to maintain a meaningful relationship with one specific object, whether needs are being satisfied or not. Such object constancy also implies stability of object cathexis and specifically the capacity to maintain positive emotional attachments to a particular object in the face of frustration of needs and wishes regarding that object. This achievement also implies the capacity to tolerate ambivalent feelings toward the object and value that object for its qualities over and beyond the functions it may serve in satisfying needs and gratifying drives.

OBJECT RELATIONS THEORY

Object relations theory emerged more or less in parallel to the evolution of psychoanalytic ego psychology, but only gradually over the ensuing years have these parallel and somewhat independent courses of theoretical development converged into complementary rather than oppositional perspectives. The development of classical psychoanalytic theory through the elaboration of systemic ego psychology has led inexorably to understand better the ego's adaptive functions, particularly the close involvement between the ego and reality in its functioning and development. The most critical dimension of the problem of the relation to reality is relationships with others, namely the whole question of object relations. The integration of these complementary currents of analytic thinking provides a more comprehensive basis for thinking about the mind as it functions not only intrapsychically but interpersonally in its relation to others as essential sources of the social environment of the human person.

Object relations theory in its bare essentials contains several fundamental points that differentiate it from classical theory. First, the ego is conceived of as whole or total at birth, becoming split or losing inner unity due to early bad experiences in object relationships, particularly the mothering object. This point differs quite radically from the classical theory, according to which the ego begins as undifferentiated and unintegrated and only achieves unity through the course of development. Second, libido is regarded as a primary life drive of the psyche, the energetic source of the ego's search for relatedness with good objects, which is essential for ego growth. Third, aggression is regarded as a natural defensive reaction to frustration of the libidinal drive, rather than specifically as an independent instinct. Fourth, the structural ego pattern that emerges when the pristine ego unity is lost involves a pattern of ego splitting and the formation of internal ego–object relations.

Melanie Klein

The origin of the object relations view can best be traced from the contribution of Melanie Klein. Klein's theorizing based itself on Freud's later instinct theory, primarily on the *death instinct* as the central theoretical assumption of her metapsychology. Working primarily with very young children, she described instinctual dynamics in the first years of life. Driven by the death instinct, the child was compelled to rid him- or herself of intolerable, destructive impulses (predominantly oral) and to project them externally. The earliest recipient of these projected impulses was the mother's breast, providing need-satisfying nourishment and satiation (*good breast*) but also often depriving and failing to satisfy (*bad breast*). At this stage, the breast images are part-objects that the infant had yet to combine into a single whole object, the mother. Early frustration of oral needs, even in the first year of life, reinforced these trends so that the bad breast became a persecutory object that was hated, feared, and envied.

Experience of the bad breast and its associated persecutory anxiety formed the earliest developmental stage in Klein's theory: *The Paranoid–Schizoid position*. The bad breast withheld gratification and thus stimulated the child's primitive oral envy, provoking sadistic wishes to penetrate and destroy the mother's breasts and body. In boys, these primitive, destructive impulses gave rise to the fear of retaliation (based in part on projection) in the form of castration anxiety; in girls, the primitive envy was expressed in envy of the mother's breast during the oral developmental phase, and later was transformed into penis envy during the genital phase. Klein held that by the time of weaning, the child could recognize the mother as a whole object possessing good and bad qualities. However, the combination of good and bad qualities in a single object—previously separated into part-objects—created a dilemma: Destructive attacks on the bad object would also destroy the good and needed object. This prevented the child's unleashing aggressive impulses against the object. This idea laid the basis where aggression turns against the self rather than the object: *The Depressive Position*. The guilt associated with destructive wishes against the object was the precursor of conscience.

Klein's emphasis in the child's developmental experience fell on the processes of introjection and projection, derived from basic instinctual drives and their interactions with the critical and primary objects of the child's early experience. Projection of destructive superego elements permits accepting good introjects (internalization of good objects), thus alleviating the underlying paranoid anxiety. The projected superego elements were later reintrojected to become the agency of guilt and early forms of obsessional behavior. The Kleinian emphasis on good and bad introjects concentrated on vital relationships to objects at the earliest level of child development, and the delineation of the internal structuring of the child's inner fantasy world in terms of the vicissitudes of these introjects, or internal objects, provided the basis and the rudimentary content for an object relations view of development.

Thus, internal objects that were either good or bad and with whom the individual was involved in intrapsychic interactions and struggles, which were in many ways as real as those carried on with the real objects outside the person, comprised Klein's "inner world." Klein saw external object relations as derived from and influenced by projective content derived from the internal object relations.

Klein's observations and formulations had a tremendous impact, particularly in bringing into prominence the role of aggression in pathologic development, in making theorists of development much more aware of the early developmental precursors of later structural entities, and particularly in providing the basic rudiments and foundations for an emergent theory of object relations.

Klein has been generously criticized for her almost blind interpretation of all forms of aggressive or destructive intent as manifestations of the death instinct. Critics point to her failure to distinguish among the various kinds of intrapsychic content (lumping object representations, self-representations, internal objects, fantasies, and psychic structures of various kinds together indiscriminately and treating them in a unitary fashion). They also note her tendency to substitute theoretical inferences for observations. Finally, they criticize her marked tendency to predate the emergence of intrapsychic organizations that are generally thought by other theorists to be achieved only in later developmental stages, for example, locating the origin of the superego in the first year of life rather than in resolution of the Oedipal situation in latency.

Wilfred Bion

Bion extended and applied Klein's ideas, especially developing the ramifications of the notion of projective identification—a process, initially described by Klein, by which a subject displaces a part of the self into an object and then identifies with that object or elicits a response in the object corresponding to qualities of the projection. Bion applied this notion to a wide range of psychotic and cognitive operations. He developed the metaphor of the *container* and the *contained* to express how projective identification occurs, especially in the contexts of mother–child and analyst–patient interaction. The child–patient projects toxic or destructive contents onto the mother–analyst, who in turn absorbs, modifies, or *contains* it so that it becomes available in a more benign form for subsequent reinternalization by the child–patient, resulting in a healthier modification of the child–patient’s pathogenic introjects.

The British Independent School

Beginning in about 1931, a shift occurred, or a new trend in psychoanalytic thinking emphasized the primacy of the external environment and the influence of objects on development. It was advanced primarily by British theorists, among whom the work of Ronald Fairbairn, Michael Balint, and Donald Winnicott stand out.

Ronald Fairbairn

Fairbairn explored the problem of ego analysis and emphasized bringing personal object relations into the center of the theory. Whereas the ego in Freudian theory had been regarded as a superficial modification of the id, explicitly developed for impulse control and adaptation to the demands of reality, Fairbairn conceived the ego as the core phenomenon of the psyche. Rather than an organization of functions, he conceived it more specifically as embodying an authentic self, that is, as the dynamic center or core of the personality. Instead of basing his theory on the instinctual drives with their implied seeking of gratification, Fairbairn shifted the emphasis to the ego and saw everything in human psychology as specifically an effect of ego functioning.

This reorientation was accompanied by a parallel reformulation of the instinctual perspective. In general, the libido or instincts, rather than mechanisms for energic discharge, were regarded as essentially object seeking. Erotogenic zones were not the primary determinants of libidinal aims but, instead, served as channels mediating the primary relationships with objects, particularly with objects that had been internalized early in life under the pressure of deprivation and frustration. Ego development was characterized by a process whereby an original state of infantile dependence, based on a symbiotic union with the maternal object, was abandoned in favor of a state of adult or mature dependence based on the differentiation between self and object. Thus, Fairbairn conceptualized the developmental process as *vicissitudes of relations with objects* rather than the *vicissitudes of instinctual dynamics*.

The basis for much of Fairbairn's theorizing was his experience with schizoid patients. He contrasted the basic dilemma of the schizoid with that of the neurotic patients he felt were treated primarily by classical psychoanalysis. He developed the view that the schizoid was not primarily concerned with the control of threatening impulses toward significant objects but that the issue for this kind of patient was having an ego capable of forming object relations at all. The relationship to objects presented a difficulty, not because of dangerous instinctual impulses arising in connection with them, but because the ego itself was weak, undeveloped, infantile, and fragile. In the struggle to defend and protect this inner weakness, the schizoid's impulses became antisocial.

Michael Balint

Both Balint and Winnicott were concerned with levels of early developmental failure that were essentially pre-Oedipal. These manifest in personality disorders that are more primitive and more difficult to treat than the usual neurotic disorders and seem to involve critical aspects of the relationships with objects early in development. Correspondingly, they do not fit well with classical psychoanalytic structural theory, with its primary focus on issues of intrapsychic conflict.

Balint envisioned several layers of psychological functioning in analysis. The first is the familiar genital level, centered on triadic relationships and specifically on intrapsychic conflicts. These conflicts and the corresponding quality of relationships were usual and familiar in most analytic processes and could be treated using adult language in verbal interpretations. There was, however, a second, deeper level in which the conventional meaning of words no longer had the same impact, and interpretations were no longer perceived as meaningful by the patient. This was the level of *preverbal experience*. He referred to this level of impairment in object relations as the *basic fault.* Balint recognized that at this level of preverbal experience, any attempt to address or describe the child's experience in adult language is bound to fail. Problems arose in analysis when efforts were made to interpret events from this preverbal level in adult or secondary process terms.

Balint distinguished between forms of regression that he described as benign and malignant. *Benign regression* was more or less equivalent to the usual form of analytic regression to more primitive relationships with primary objects. Such regression was gradual, tempered, and modulated according to the patient's capacity to tolerate and productively integrate the resulting anxiety. During this regression, the analyst maintains the analytic structure and relation, and his empathic responsiveness made it possible for the patient to withstand this unstructured experience and keep the anxiety within manageable limits. At the level of the basic fault, the lost infantile objects can be mourned, and the quality of the relationship with them was open to reworking so that the patient's basic assumptions governing his or her interaction with the internal and external object world could be reformed.

During phases of benign regression to this preverbal and pregenital level of object relations in the analysis, the analyst could usually provide an adequate degree of empathic acceptance and recognition, rather than verbalized interpretations, of this level of the patient's unstructured and regressive experience, without anxiety or any need to escape or subvert this level of experience through interpretation. Balint felt that the dynamics at this level are more primitive than can be adequately expressed in terms of conflict because they derive from the basic form of dual relationship involved in early mother–child interaction, which is the basic fault.

In contrast, *malignant regression* tends to be precipitous and extreme; the ego is prematurely overwhelmed by traumatic and unmanageable anxiety. This anxiety prevented any effective reworking of fundamental disturbances in object relationships, recreating and reinforcing the basic fault rather than creating the conditions for its therapeutic revision. At an even deeper level, beyond the reach of analytic resources, there lies the area of creativity; that is, an idiosyncratic, uncommunicable, and objectless area beyond any conventional or linguistic expression.

Regression to the level of the basic fault was an entirely different and distinct phenomenon than the more usual Oedipal regressions experienced in the analysis of adult neurotics. In the benign Oedipal regression, the aim was to gratify infantile instinctual wishes. Regressing to the level of the basic fault, however, sought a basic recognition by the therapist and protective support and consent to express the inner core of creativity that lies at the heart of the patient's being and accounts for the capacity to become ill or well. Balint used the notion of primary love at this deepest level to describe the withdrawal of libido from a frustrating object in the effort to re-establish a certain inner harmony in which it becomes possible to recover the conditions of early care and tranquility. He referred to a *harmonious interpenetrating mix-up* to describe this early, almost undifferentiated interaction of the infant and environment. Balint's analogy was that of breathing air; the organism cannot exist without air, so air and the organism are seemingly inseparable, but to cut off the supply of air reveals both the organism's need for it and the distinction between air and the organism. In terms of *primary love*, as it came to bear in analysis, the patient would seek a basic form of recognition from the analyst, as he had from significant objects in the patient's early life experience.

Donald Winnicott

Winnicott particularly emphasized the importance of early interactions between mother and child as determining factors in laying down essential components of ego development. He was concerned with the earliest phases of the mother–child relationship and the importance of what he described as *good-enough mothering* for the child's psychic development. Development involved movement from an early stage of total or absolute dependence toward a more adult relative dependence and independence phase. As he saw it, the inherited native potential for growth was strongly influenced by the quality of maternal care. This potential for development is affected even from the moment of conception. Even before birth, the child becomes invested by a strong narcissistic cathexis that allows the mother to identify with the child and become empathically attuned to the child's inner needs as if the child were—and indeed is—an extension of her own self. He called this early prenatal involvement of mother and child-in-the-womb a *primary maternal preoccupation.* This set the stage for the development of a holding relationship, in which the mother becomes sensitively attuned to the infant's needs and sensitivities and is both physically and emotionally responsive to them, thus providing a physical, physiologic, and emotional ambiance, protection, and security for the dependent infant.

As the infant moves from this early stage of *absolute dependence* toward a more relative one, awareness of personal needs and the mother's existence as a caretaking object grows. The optimal relationship at this stage involves a continuation of protective holding, along with an optimal titration of gratification and frustration. As a result of this optimally attuned relationship between the patterning of infantile drives and initiatives and their harmonious fitting in with maternal sensitivities and responsiveness, there is a developing sense of reliable expectation that the infant's needs will be satisfied without the threat of excessive withdrawal of the mothering object and the threatening, overwhelming, and short-circuiting of the infant's initiatives because of excessive maternal impingements.

In normal development, this allowed for the emergence of a certain infantile omnipotence. From this position, the child gradually retreats due to *tolerable degrees of frustration* by and separateness from the maternal object. Although the mother continues her holding at this phase, she must yet allow enough separation between herself and the developing infant to permit expression of the baby's needs and initiatives that form the rudiments of an emerging sense of self. If she is too distant, too unresponsive, or not sufficiently present, anxiety arises and is accompanied by the fading of the infant's internal representations of her.

The transition from a phase of *absolute dependence* to one of *relative dependence* represents a crucial development in the capacity for object relations. It is accompanied by a critical transition from total subjectivity to objectivity in the perception of and relation to objects. The transition from subjectivity to objectivity is accomplished by developing transitional phenomena, expressed in the first instance in the emergence of transitional objects. These objects are the child's first object possessions that are perceived as separate from the emerging self—the first "not me" possessions. From the study of infant behavior, Winnicott argued that the *transitional object* was a substitute for the maternal breast, the first and most significant and substitute object in the environment to which the infant related. He regarded the transitional object as existing in an intermediate realm or space contributed to both by the external reality of the object (the mother's breast) and by the child's own subjectivity. This intermediate realm was both subjective and objective at once without being exclusively either.

Winnicott referred to this realm as the *realm of illusion,* an intermediate area of experience that embraced both inner and external reality and may be retained in areas of adult functioning, involving such imaginative capacities as creativity, religious experience, and art. In its primitive form, however, the transitional object commonly experienced in childhood development may take the form of a particular object, a blanket, a pillow, or a favorite toy or teddy bear to which the child becomes intensely attached and from which it cannot be separated without stirring up severe anxiety and distress. Attachment to this object was an immediate displacement from the mother's figure and represented an essential developmental step insofar as it allowed the child to tolerate increasing degrees of separation from the mother, using the transitional object as a substitute.

The mother participated in this intermediate transitional realm of illusion by responsiveness to the infant's need to continually create her as a good mother. In her sensitivity and responsiveness, she functions as a good-enough mother. However, her failure to provide such adequate mothering, either by excessive withdrawal or by excessive intrusion and control, may result in the child's emergence of a false self-based on compliance with the demands of the external environment. This condition reflected a developmental failure and often severe character pathologies.

When such patients were seen as adults, they were neither neurotic nor psychotic but seemed to relate to the world through a compliant shell that was not quite real to them or the analyst. They were often mistrustful without being specifically paranoid. They appeared withdrawn and disengaged and seemed able to relate only using the protective shell, which seemed obsessive and compliant but that separated and isolated them from meaningful contacts with their fellows, even as it provided their only basis for a relationship. These disturbed personality types, or *false-self adaptations*, reflected a fundamental impairment in very early object relations, particularly in the mutuality and responsiveness of very early mother–child interaction.

Infants who developed in a false self-mode have not experienced a maternal relationship's security and mutual satisfaction. Such mothers are emphatically out of contact with the child and react mainly based on their inner fantasies, narcissistic needs, or neurotic conflicts. The child's survival depends on the capacity to adapt to this pattern of the mother's response, which is often grossly out of phase with the child's needs. This establishes a pattern of gradual training in compliance with whatever the mother was capable of offering, rather than seeking out and finding what is needed and wanted. Consequently, the child's needs, instinctual impulses, wishes, and initiatives, instead of becoming a meaningful guide to satisfying growth experiences and enlarging capacities to interact meaningfully with objects, become from the very beginning threatening to the harmony of the relationship with the mother. She remains unresponsive to affective feedback from the child.

Winnicott's attempts to formulate treatment principles for such impaired patients were built on the model of good-enough mothering. This called for a capacity for holding, empathic responsiveness, and a capacity for creatively playful exchange that allowed the patient's capacities for growth to emerge and flourish. Such playful engagement and interaction between analyst and patient had a facilitating effect on expanding the patient's authentic sense of self, hidden behind an external facade of false–self-compliance.

HEINZ KOHUT'S SELF-PSYCHOLOGY

Beginning in the 1960s, a focus on self-concept was emerging, emphasizing the deepening psychoanalytic understanding of the human psyche. Self-psychology, as a separate movement within psychoanalysis, takes its origin from contributions by Heinz Kohut and his followers. Heinz Kohut discussed that the application of classical drive theory was not providing the clinical results with patients with narcissistic issues, and he began looking at treatment differently. Kohut linked the origin of the self to narcissism, viewing the self as the result of a separate line of narcissistic development that progresses through a series of archaic narcissistic structures toward establishing mature and cohesive self-organization. Kohut credited Heinz Hartmann with stimulating the development of a concept of the self in the context of structural theory. Hartmann's distinction between the ego and the self, casts the self in representational terms, as a complex representation, organized, and synthesized as an ego function.

Kohut posited that the negative connotation of narcissism was limiting conceptually and in practice. He encouraged further understanding and consideration of developmental progression and adaptive uses of narcissism. Adaptive uses of narcissism include self-esteem and self-confidence. He also highlighted that the patient with narcissistic personality structure also had assets, not only deficiencies. Kohut discussed that patients with narcissistic personality organization had a cohesive self, yet this self was less subject to disintegration and considered patients with a narcissistic personality more analyzable than those with a borderline or psychotic organization. Kohut's work led to a focus on the self, more or less exclusively, in terms of its subjective and experiential dimension and has subsequently become a significant stimulus to renewed interest in the self.

In Kohut's view of narcissistic development, the original *primary narcissism* differentiates during development, and in response to lapses in parental empathy differentiates into two archaic configurations, the *grandiose self,* and the *idealized parental imago.* The *grandiose self* involves an exaggerated and exhibitionistic image that becomes the repository for infantile perfection; the *idealized parental imago,* in contrast, transfers the previous perfection that was lost to an admired omnipotent object(s). There is a wish to remain in close union with the idealized parental imago. Normal development of the grandiose self leads to more mature forms of narcissism such as ambition, self-confidence, and pleasure in accomplishment. The idealized parental imago can be healthily integrated into the ego ideal with the mature values, ideals, and standards it represents.

Pathologic persistence of the grandiose self intensifies grandiosity, exhibitionism, shame, envy, depression, hypochondriacal concerns, and undermines self-esteem. Loss of the idealized object or the idealized object's love can result in narcissistic imbalance, leaving the individual vulnerable to depression, depletion, poor self-esteem, failure of ideals and values, and even self-fragmentation.

Kohut based his self-psychology on the need, both during development and during life, for empathic interaction with *self-objects.* The original self-object is the primary caregiver who provides an empathic response to self-object needs in the infant in the form of love, admiration, acceptance, joyful participation, warmth, and responsiveness, communicating a sense of valued and cherished existence to the child. However, Kohut contends that human beings seek objects to fulfill these basic self-object needs throughout life. Failure to fulfill such needs can result in pathologic psychic structures and behavior patterns during development and pathologic character structures during adult life.

Two other concepts discussed by Kohut are now often considered when evaluating patients. *Empathy* is something he valued, but not in a vague sentimental way, but using empathy to interpret what is happening in a person's life. Another concept he discussed is his distinction between self and identity. He wrote that the self refers to the core personality made up of various constituents in the interplay with the child's earliest self-objects. *Identity* is the point of convergence between the self developed between late adolescence and the sociocultural position of the individual. One might have a diffuse identity, but a strong sense of self, which Kohut posited could be the makeup of a psychoanalyst who is resistant to fragmentation but whose diffuse identity could allow them to empathize with many different types of people.

The current prevailing view of the self is Kohut's self psychology. Emphasis is placed on the initial and continuing dependence of the self concerning self-objects. Developmentally, the importance of empathic parental attunement is considered vital to normal psychic growth in the interest of facilitating infantile needs for nurturance, love, and narcissistic approval.

Theoretical gains of the emerging role of the self-concept:

1. The self as a theoretical construct provides a focus for formulating and understanding the complex integrations of functional processes that involve combinations of functions of the respective operational components of the self. This would have specific application to such complex activities as affects, in which all of the psychic systems seem to be in one way or other represented; complex ego–superego integrations reflected in such formations as value systems; and other complex interactions of psychic systems that involve fantasy production, drive-motor integration, or cognitive-affective processes.
2. The self-concept provides a more specific and less ambiguous frame of reference for articulating self-object interrelationships and interactions, including the complex areas of object relations and internalizations.
3. The emergence of a self-concept provides a locus in the theory for articulating the experience of the personal self, either as grasped introspectively and reflectively or experienced as the originating source of personal activity. This sense of the self-as-subject would provide a place within the theory for an account of subjectivity and subjective meaning.

Although a person's experience of themselves and sense of self is now a commonplace concept in psychoanalysis and therapy, Kohut did not gain wide acceptance of his theories during his lifetime. This has been attributed to his seeking to discard drive theory.

OTTO KERNBERG'S INTEGRATION OF OBJECT RELATIONS AND STRUCTURAL SCHOOLS

In contrast to Kohut's view, Otto Kernberg's view of narcissism is embedded in drive theory. Kernberg implied that both libido and aggression are invested in the self, thus constituting elements of narcissism. Aggression may be condensed within the structure of a defensive, pathologic grandiose self, constituting the basis of the most severe constellations of pathologic narcissism. Kohut and Kernberg disagree about the origin of this grandiose self and whether it reflects the fixation of an archaic "normal'' primitive self (Kohut's view) or a pathologic structure, clearly different from normal infantile narcissism (Kernberg's view). Kernberg's work includes his belief in applying libidinal and aggressive drives and applied it to object relations, especially under very activated affect states in early life. He theorizes that repressed drive elements are invested in object relations from a young age. Kernberg believes affect states representing these primitive drives are essential links between self and object representations and feels these affects are building blocks for the drives. He sees unconscious intrapsychic conflict as self and object constellations under the influence of a drive derivative on one side of a conflict and contradictory and opposing units of self and object representation and defensive structure on the other side.

A major contribution to the field was Kernberg's focus on *levels of personality organization* in the context of ego psychology. These levels have been called the *mature*, *neurotic*, *borderline*, and *psychotic levels of organization*. Severe psychopathology involves splitting mechanisms in the ego–id matrix to stabilize affects. Individuals at a higher level of organization, *mature personality organization*, having a relatively well-integrated ego and superego.

The superego is harsh and perfectionistic. The ego utilizes inhibitory defenses such as repression and reaction formation, thereby masking instinctual needs. The patient's ego is restricted, but their adaptation to the social world is not seriously impaired.

The person with a *neurotic level of organization* has fairly deep, stable object relationships and is capable of experiencing guilt, mourning, and a wide variety of affective responses, with only a partial inhibition of sexual drives. At neurotic level "the excessively punitive nature of the patient's superego is even stronger than that of the higher-level disorders, but the superego is less integrated." Often, neurotic individuals manage the contradictory demands of the cruel and prohibitive superego and the primitive and often idealized sense of who they want to become. There is a wish to be great, powerful, physically attractive, and also morally perfect. Thus, there can be a blurring between superego and ego. Mood swings within this level of organization are the result of the superego's difficulty in regulating the ego. Kernberg observed that although repression is still present, there are fewer inhibitory defenses at this higher level, more impulsivity and expression of unacceptable sexual and/or aggressive needs, as well as more reaction formation, intellectualization, rationalization, and undoing. Neurotically organized individuals have a capacity for stable, lasting relationships.

In the next lower level, later termed *borderline personality organization*, the patient's superego integration is minimal and the capacity for projection of primitive, sadistic superego elements is high. The individual uses projection and projective identification as defenses. The boundary between ego and superego is extremely blurred. The ego is unable to integrate experience and instead uses splitting and dissociation as defenses versus repression. The mechanism of splitting is expressed as contradictory ego states alternating with each other, and this dissociation is reinforced by the patient's use of denial and projective identification. These patients have trouble holding in mind contradictory ("good" and "bad") self and object images. Kernberg points to early aggressive drives making it difficult to integrate these concepts. Thus, their inability to integrate aggressive and libidinal drives leads to maintaining object relationships of either a need-gratifying or a threatening nature. They are unable to have empathy for others, trouble experiencing people in their entirety, and often have relationships that address an aspect of the other person.

Finally, at the lowest level, termed *psychotic level of organization*, the person lacks an integrated sense of self. The inner world is composed of the most negative and positive aspects of people who are important to them but are not integrated. And neither their experience of their inner world or their sense of self is integrated, that leads to identity diffusion. Without this coherence of self and experience the ego strength suffers and has trouble tolerating frustration, tolerating anxiety, trouble with impulse control, and difficulty using sublimation as a defense.

ATTACHMENT THEORY

Unlike most psychoanalytic theories that are principally based on the data derived from the clinical experience, the attachment theory is different in that it adds the perspective of systematic observation and empirical research to the development of psychoanalytic ideas. It is grounded in robust infant research that spans over a period of half a century. For an excellent recent review of attachment research that has contributed to the development of attachment theory, please refer to the third edited volume of *Handbook of Attachment* by June Cassidy and Phillip Shaver.

Attachment theory derives from the work of John Bowlby. In his pioneering work with maladjusted and delinquent children, who had experienced early and prolonged separation, Bowlby conceptualized that attachment constituted a central motivational force. Bowlby emphasized the survival value of attachment. He postulated that attachment to the mother was essential for the infant’s survival and had important consequences for later development and personality functioning. Influenced by ethologists like Konrad Lorenz, he proposed that attachment behaviors are part of a biobehavioral system. A biobehavioral system involves inherent motivation and cannot be reduced to another drive. Attachment behaviors enhance safety through proximity to the caregiver and cannot be reduced to the caregiver's feeding function. Bowlby's formulation differs from the object relations perspective in its view of relationship with the object as the primary motivator rather than the drive. Specifically, Bowlby emphasized the significance of caregiver availability and the consequent state of being or feeling close to the caregiver. He asserted that availability referred to the expectation based on repeated experience with the actual caregiver, which was represented as a model. He called it an *internal working model*. The emphasis on the "actual" presence or absence of the caregiver was again a departure from the psychoanalytic perspective of privileging the internal world of fantasy over the external world. Later researchers redefined the goal of attachment system from physical closeness to feeling of *security*, which extended the applicability of the concept of attachment across the lifespan. With the notion of felt security and the focus on internal representations in the infant and the caregiver the attachment theory increasingly recognizes the overlap between the external reality and the psychic reality.

Mary Ainsworth and her colleagues, in their pioneering work, developed a procedure to study Bowlby's conceptualization. In an arrangement for observing an infant's internal working model in action, known as the *Strange Situation*, she had infants briefly separate from their caregiver in an unfamiliar situation and she documented patterns of their attachment behaviors. Her findings led her to identify three different patterns of attachment behavior—*secure*, *anxious/avoidant*, and *anxious/resistant*. Secure infants readily explore in the presence of their caregiver, are anxious in the presence of a stranger, are distressed by the caregiver's absence, rapidly seek contact when caregiver returns, are easily reassured by contact with the caregiver, and return to exploring. Anxious/avoidant infants do not prefer primary caregiver over the stranger, they do not show distress when caregiver leaves, and they do not seek contact with the caregiver upon their return. Anxious/resistant infants are inhibited in their exploration even in the presence of primary caregiver, they are extremely distressed by caregiver's absence, and even though they seek contact when caregiver returns, they are difficult to soothe and struggle to return to exploration and play. Patterns of attachment behavior of *disoriented/disorganized* infants was added by Mary Main and Judith Solomon. This group of infants show undirected and confusing behavior and are understood to have experienced their caregivers as frightened or frightening. The experience of caregivers as both a source of comfort and fear leaves this group of infants in an impossible bind which is then reflected in their conflicting and disorganized attachment behavior.

Subsequent attachment theorists have studied these patterns of early infant attachment and related them to patterns of adult attachment. Using the *Adult Attachment Interview* (AAI), developed by Main and others, they documented the stability of internal working models, like introjections or internal objects, based on early attachment relations. It turned out that these patterns of infantile attachment could then be related to attachment attitudes in adults based on the results of the AAI. Thus, secure infants were found to develop secure and relatively autonomous patterns of attachment and interaction with others in adult relationships and were more likely to become more resilient, self-reliant, empathic, and confident in forming meaningful relationships. Avoidant or dismissing infants were found to be associated with an adult pattern of dismissing object relationships as important or significant, a combination of indifference and independence, and a form of self-reliance akin to counter-dependence and denial of any need for the presence or caring of others. Their investment is in maintaining aloofness and detachment in the service of defending against loss and separation. These infants were seemingly more vulnerable to adult disorders than others, for example, borderline, histrionic, and dependent personality disorders, and may be more challenging to treat in psychotherapy. Resistant or preoccupied infants later seemed to develop a somewhat preoccupied adult stance in which subjects are aware of attachment needs but come to expect that they will be inevitably disappointed, resulting in excessive dependency accompanied by disappointment and fears of loss and abandonment. Childhood experiences of past attachments were often connected with confused, angry, or fearful emotions, especially fears of abandonment. They seem invested in maintaining contact with need-satisfying objects but continually defending against the anxiety related to separation and loss. The disorganized–disoriented infantile pattern seemed to be related to an unresolved and disorganized pattern in adults, suggesting a more severe disturbance in self-object relations. These findings provide an extension and specification of an object relations approach and provide specific empirical and observational methods for a more detailed study of the development of object relations from childhood to adulthood. Attachment studies document the development of internal representations of object relationships and their stability and instability throughout development.

Peter Fonagy’s work on *reflective functioning,* the capacity to perceive and think about the intentionality of self and others, documents intergenerational transmission of attachment security and insecurity. Specifically, high levels of reflective functioning in mothers contribute significantly to their infants’ secure attachment. This capacity reflects the formation of coherent representations of the psychic world of others, especially caregivers, and of the subjects’ own internal states and is measured using the AAI.

Both attachment theorists and object relations theorists emphasize the significance of the mother’s empathic responsivity to the infant for self-development and relatedness, the importance of the mother–child involvement for personality development, and the role of the mother (caregiving context) as the catalyst for age-appropriate development.

Some object relations theorists have advanced objections to the attachment theory approach. Criticisms have focused on the attachment categories: they seem stagnant, lack developmental complexity, give little account of basic instinctual motivations and the role of unconscious fantasy, and provide little or no account of the progression in the evolution of object representations or patterns of internalization as development proceeds. Moreover, attachment theory pays little attention to the relational dynamics between mother and child, especially as the child progresses developmentally. No mention is made of the uniqueness of each mother–child relation and its modification by surrounding influences from other relationships.

Theoretical issues that would seem to call for clarification and integration might include the relation between mental representations and the internal working models of attachment theory. They seem analogous, and both theories remark that both representation and working models of self and others derive from experiences with early caretaking and become guides for the subsequent development of more mature object relations. However, they differ in the quality of epigenetic differentiation, representations following a specified developmental sequence and thus becoming more complex, symbolic, and linguistically mediated in contrast with internal working models that preserve a relatively constant character. This and other theoretical divergences would seem to provide a ripe field for theoretical revision and integration of these two relatively independent perspectives on early object relationships and their developmental course.

RELATIONAL THEORY AND INTERSUBJECTIVE APPROACH

The emergence of relational and intersubjective approaches to the nature of the analytic relationship has in many ways incorporated and extended aspects of Kohutian self psychology. Following Kohut's initiative, the relational perspective began by distinguishing between drive theories and relational theories, implying that viewing the analytic relation as a function of the ongoing interaction between analyst and analysand demanded the abandonment of drive theory. This dichotomy proved false in that one could entertain the relational view without dispensing with internal dynamics and psychic processes operating within the individual participants, both before and within the development of the analytic relation. Even a relational viewpoint, for example, cannot do without a system of intrapsychic motivation since the process of relating to another person and forming a relationship is itself motivated.

These approaches to understanding the analytic interaction employ a form of social constructionist epistemology in which transference is regarded as resulting from the interaction of analyst and patient. This constructivist view contrasts with the more objectivist view of traditional ego psychology and object relations. Thus, the analyst's attention is focused on the here-and-now interaction with the patient rather than only on the inner dynamics of the patient's mental life and experience. This introduces a shift from a one-person to a two-person psychology. Whether conscious or unconscious, the ongoing interactions between the participants become central, and transference and countertransference are regarded as mutually cocreated by both analyst and analysand.

Intersubjective theory further expands the self psychological emphasis on self-selfobject transferences and rejects consideration of the analyst's stance as neutral or observational, to now question the analyst's subjectivity, authority, and capacity to know any objective reality about the patient. Stephen Mitchell, Jessica Benjamin, Robert Stolorow, and Owen Renik, among others, proposed the concept of personality as developing within a relational matrix. It calls for a central focus on the intersubjective field within the relationship between analyst and patient. This aspect of the analytic situation is explored and interpreted to bring about personal growth in the patient. The analyst's technical neutrality and objectivity are rejected in this approach as illusory and little more than expressions of the analyst's authoritarian position. Within a self-selfobject or intersubjective relation approach, neutrality is precluded as potentially traumatizing and destructive to potential consolidation of the self. An inevitable consequence of these approaches is that they do away with the traditional notion of the unconscious and undercut any sense of transference as reflecting unconscious aspects of the patient's inner psychic life. Instead, transference is created anew in the present analytic interaction.

This approach radically departs from classical analytic views in its intersubjective guise. In the *intersubjective field*, cocreated by analyst and patient, both diagnosis and treatment are functions of the mutual interaction between them, implying that neither diagnosis nor prescribed therapy techniques have any objective validity but arise spontaneously from the workings of the intersubjective process. Any contribution from biologic determinants and constitutional factors, or for that matter from the prior history of object relationships and developmental vicissitudes, is discounted. Transference and countertransference arise in the immediacy of the intersubjective interaction. They develop in virtue of mutual here-and-now influences the participants exercise on each other, rather than from displacements or past projections. Questions have arisen about the meaning of the unconscious in this perspective and whether interactional determinants have replaced it. The emphasis on here-and-now interaction has promoted technical divergences such as the analyst's self-revelation and disclosure, even to the point of sharing countertransference elements with the patient. Preferences for spontaneity and self-expression prevail, even "throwing away the book" in the interest of escaping from any implication that the analyst has any knowledge, competence, skill, or authority that would disturb the presumed equality of the intersubjective matrix.

This approach has thus raised some significant epistemologic issues, perhaps the most telling having to do with the question of what a relation might be without a constitutive subject and object. Ordinarily, relations arise between a subject and an object, connecting them in some fashion. However, the relation itself arises from their mutual connection and interaction and has no meaning or existence without subject and object. Many intersubjective theorists address the relation as a third entity existing independently between subject and object. The properties of the analytic interaction arise within the relation separately and independently of the constituting subject and object, that is, patient and analyst.

The relational and intersubjective viewpoints bring into focus and articulate aspects of the analytic situation that have been less frequently discussed. These relational approaches' contribution deepens and expands one’s understanding of the analytic interaction.

NEUROPSYCHOANALYSIS

Another exciting and challenging branch of contemporary psychoanalytic thinking has arisen from the neuroscientific exploration of brain functions. The advent of neuroimaging and microelectronic brain recording devices has crossed a new threshold in understanding the brain’s functioning, from which significant discoveries emerge almost daily. Insofar as psychoanalysis is itself a scientific study of the functioning of the mind, these new developments thrust to the foreground issues related to how brain processes are involved in and related to mental functioning.

Neuropsychoanalysis explores the interface between neurologic knowledge and psychoanalytic models of the human mind. The work of neuroscience primarily involves mapping various processes in the function of the brain, from the molecular level to large-scale networks. With the evolution of neuropsychoanalysis, there is now a bridge between the worlds of neuroscience and psychoanalysis. Advances in technology now allow us to visualize discrete processes that we can now extrapolate onto psychoanalytic theory and practice, such as understanding the deep unconscious layers of the mind, the central role of emotions and interpersonal relations in mental life, and the importance of fantasy and mental representations. Neuroscience is rapidly expanding our understanding of the neural circuits involved with conscious and unconscious processes, motivation, emotion, self-regulation, memory, and interpersonal relations.

This field has been supported by cognitive, behavioral, and affective neuroscientists, developmental neurobiologists, neurologists, cognitive psychologists, neuropsychologists, and neuropsychoanalytic researchers. Some of the leading investigators in the field include Beatrice Beebe, Antonio Damasio, Vittorio Gallese, Eric Kandel, Joseph Ledoux, Helen Mayberg, Jaak Panksepp, Manos Tsarkis, Vilyanur Ramachandran, and Mark

Solms. Their work, in large part, attempts to bridge the gap between the fields of neuroscience and psychoanalysis. From Freud’s original psychoanalysis to today, there has been much debate about the legitimacy of psychoanalytic theory. Advances in neuroscience have “legitimized” various aspects of psychoanalytic thinking and theory. More importantly, they are deepening our understanding of a given analysand/patient/client, arguably the most crucial aspect of analytic work.

There is debate as to whether neurobehavioral discoveries affect analytic technique. For some individual analysts working in their consulting room, it does not matter, nor does it affect their work or approach to the patient in any significant way. Yet for others, neurobehavioral research, for example, has identified multiple and more or less independent forms of memory function, each subserved by different neural systems. Also, the primary distinction between explicit and implicit forms of memory has begun to have an impact on analytic technique, insofar as it brings into focus the problem that free associative and interpretive methods are appropriate for the recovery and processing of explicit forms of memory, for example, autobiographical memory, but not the implicit forms. Analysts are already working to determine what ways to engage with and what therapeutic use can be made of implicit memory systems.

However, although it may not matter which specific neural mechanisms are at work in producing mental actions and functions, it is imperative to identify such mechanisms. The significance, therefore, is less in terms of implications for clinical technique and practice at this point and considerably more in terms of the scientific status and integration of psychoanalytic theory. The critical focus in this area thus falls on the understanding of the mind–brain relation, or more broadly conceived, the mind–body relation.

Neuroscientific research makes it increasingly and abundantly clear that no action or function of the human mind does not relate directly to and correspond with a parallel process in the neural net. It is not surprising that the brain regulates and directs bodily functions: When one moves their hand, the corresponding activation of cells in the motor cortex seems unsurprising. However, when one imagines the outline of a triangle or when one adds a set of numbers, it seems surprising that parallel activity can be identified in the brain. The more complex the system the seemingly harder it is to understand for some. The neural net is vast. We have yet to discern discrete processes for complex thoughts however we are clearer on patterns of connections between the different regions of the brain. Initially, the brain was conceived of discrete areas with very discrete and singular functions, just as Freud had originally conceived of the systems of the mind as three discrete independent entities. Years of vigorous study combined with clinical observations have shown that there are very few discrete processes at play outside of the basic human functions. As we come to understand the vastness of these connections and corresponding behaviors, it can help us to determine or modulate our approach with a patient based on their unique story and experiences.

The stumbling block for many analysts is that there seems to be an inescapable gap between neuroscience on one side and psychology and psychoanalysis on the other. Neuroscience focuses on third-person data, observations, and measurements of neural processes, electrochemical transmission patterns, and the activations of neurophysiologic organization traced by multiple imaging and neural recording techniques. In contrast, analysis studies the inner psychological processes in the mind; the data are subjective, first-person in character, not available simply to external observation but only through subjective processes of introspection, recall, and association. The nature of the data and methods of acquiring the data has been perceived as different and independent. Contemporary neuroscientific studies incorporate clinical observation which is serving to bridge the gap between the channels of knowledge about the human phenomenon. Either one without the other provides no more than a limited and incomplete understanding of what is essentially a complex, integrated, and unified mind–brain entity that constitutes the reality of the human person or self.

Nevertheless, the fact remains that neuroscience currently offers a kind of evidence that was not available to previous generations, including Freud. It is as though the brain going into action in performing mental acts can almost literally be seen. The connection is no longer an inference but more akin to an observation. Current understanding of the functioning of the upper brainstem, as an example, helps to illuminate the innate nature of drives. One can argue that there is direct causality when effects follow acts. It opens the possibility that the brain is directly causing these effects, that mental actions are equivalently actions of the brain. The conclusion would have far-reaching effects on how subjective phenomena are understood. Further, the role of psychoanalysis as a scientific study of the mind would be more clearly established and confirmed. However, these questions and the possible alternatives they open are still matters of active debate and reflection.

PSYCHOANALYTIC THEORY IN PRACTICE

Psychoanalytic ideas have been inspiring psychological theories and practice since their inception. These ideas remain radical and are richer beyond comparison in describing mental functioning. The curiosity about the meaning of mental experience encouraged by psychoanalytic theories provides clinicians with an understanding of their patients that's complex, deep, and particularly useful when making sense of intense and disturbing clinical interactions.

Psychoanalytic practice provides an unparalleled window on human behavior and experience. However, it has not been easy to illustrate the translation of psychoanalytic theory into effective treatment. Psychoanalysis is not yet able to explain with precision the nature of therapeutic action. Also, it has lagged in gathering evidence to support its effectiveness. More broadly, the practice of psychoanalysis has been slow to develop, especially compared to the significant advancements in psychoanalytic theory. Of note, Freud developed his practice on a trial-and-error basis. For example, he discovered free association as he abandoned hypnosis, which he then put forward as a psychoanalytic technique. Psychoanalytic techniques discovered and proposed by psychoanalytic ancestors continue to guide the clinical work. These time-honored techniques and principles have changed little over the past century. The other issue is that the techniques don't have a one-to-one correlation with the theory. All too often one theory can be used to justify varied psychoanalytic practices and the same technique can be linked back to more than one theory. Additionally, psychoanalysts have lagged in gathering evidence to support their claims of effectiveness of psychoanalytic treatment. Although not impossible, the challenge that psychoanalysts face in this regard is twofold. One, it is both difficult and counterproductive to manualize psychoanalytic treatment, which is long term, highly subjective, and unique to the patient–analyst dyad. Two, the research methodology in current use is not up to par in its ability to capture the subjective experience that is at the core of psychoanalytic treatment.

Of Freud's two radical propositions regarding mental suffering and its treatment, the notion that *symptoms have meaning* and can be understood in terms of *unconscious beliefs and affects* has been well supported by research but there is far less evidence to support the claim that it is indeed curative to communicate the understanding of mental functioning to the patient. What effects cure and how remains a serious query for practitioners of psychoanalysis. Even though there remains a gap between psychoanalytic theory and practice, familiarity with psychoanalytic theories such as the one's reviewed in this chapter is crucial for the clinicians interested in appreciating the complexity of the clinical encounter. Moreover, the advancements in genetics and neuroscience have been strengthening psychoanalytic convictions regarding dynamic mental functioning. The recent advances in neurosciences are also leading the way forward in identifying objective indicators to support the practitioners of psychoanalysis in proving the effectiveness of their practice. The psychoanalytic technique, on the other hand, has evolved from clinical experience. All psychoanalytic practice is based on cumulative clinical wisdom supported by theory-driven framework of psychological mechanisms. Below is a brief discussion of generally agreed upon principles of psychoanalytically derived clinical practice and their compatibility with progress being made in related fields of study of mind.

Focus on the Internal World

All psychiatric disturbance involves the interaction of genetics and environmental factors. The knowledge emerging from genetic research over the last three decades, however, has led to a cultural shift in psychiatry. Instead of looking for ways to integrate the psychoanalytic insight with the neurobiologic findings the trend is towards the dominance of the biologic model of mind and psychopathology. The psychoanalytically derived clinical work challenges the equating of brain and mind and values the internal world of wishes, fears, fantasies, and dreams. The move away from reductionist either–or thinking is essential for today's psychiatrist to provide adequate patient care. The studies attempting to control the effect of genetics on development of disturbance demonstrate the considerable influence of early experience and underscore the potential danger of genetic reductionism. Contemporary genetic research in fact, shows that expression of genes is highly regulated by environmental influences. This has substantial clinical significance for psychodynamic practice. The psychodynamic focus on intrapsychic variables has much to contribute to the way gene–environment correlations and interactions may or may not be expressed in the individual patient's psychic disturbance.

Dynamic Unconscious

The hallmark of psychoanalytically oriented clinical work is the attention to dynamically unconscious mental life to account for symptoms and behaviors. Cognitive neuroscience has revealed that most of the work of the brain is nonconscious, which includes nonverbal behaviors, patterns of relatedness, procedural memory, and implicit aspects of cognitive tasks like decision making among others. The psychodynamic clinician offers uniquely significant insight into complex and often contradictory and conflicted human behavior by its focus on unconscious motivational and affective processes.

Unconscious Motivation

A key tenet of psychoanalytically oriented clinical work is the assumption of the ubiquity of *psychic conflict*. All human behavior and symptoms are compromises or adaptations to the internal and external demands. Human behavior and symptoms have multiple antecedents and varied explanations. They are the end result of conflicting and corresponding demands of several different unconscious processes involving drives, defenses, object relations, external reality, and they serve many different functions.

Early Childhood Experience

All psychodynamic formulations assume a correspondence between early childhood experiences and later psychopathology. Advances in infant research have proven beyond doubt the importance of early experiences for healthy development. Moreover, attachment researchers have documented the adverse impact of early disturbances in attachment relationships.

Mental Representation of Relationships

Psychoanalytic object relations theory emphasizes the mental representations of relationships as mediators of healthy and unhealthy behavior. The analytic or therapeutic relation can be conceived as composed of at least three coexistent components, mutually interacting, and influencing, and intermingled at all points in the analytic process—the transference and countertransference, the therapeutic alliance, and the actual relationship. Although constantly interacting to influence the patterns of interaction between analyst and patient and determining the course of the analytic process, they can be usefully distinguished in that they point to differentiable issues and aspects of the therapeutic process and call for different therapeutic responses and interventions. While transference and countertransference play an undoubtedly central—often dominant—role in the analytic process and interaction, the influences from the alliance and real relationship remain continually active and should not be ignored.

Transference. *Transference* is at the heart of psychoanalytic process and technique. Transference phenomena refers to the persistence of early patterns of mental organization in adult life. Through free association, unconscious patterns of the patient's mental organization, which may refer to events or fantasies in the patient's private experience, are brought to life and activated in the relation with the analyst. In the simplest model of transference, the analyst is gradually invested with emotions usually associated with significant figures in the past. The patient displaces or projects feelings originally directed toward these earlier objects onto the analyst, who then is perceived alternately as a friend or enemy, one who is nice or frustrates needs and punishes, one who is loved or hated, and so on, as the original objects were loved or hated. Moreover, this tendency persists, so that the patient unconsciously reenacts the past relationships with the analyst and in so doing introduces the analyst to significant figures from the past. The transference object is the lens through which the patient views the analyst, seeing him or her in the image of the transference representation.

The concept of transference has undergone considerable elaboration over time, resulting in multiple variants, broadening its connotations to include every emotional connection to the analyst, and extending the transference model to encompass the widening range of psychopathology addressed by psychoanalysis. Analysts differ regarding the degree to which emphasis is placed on the role of transference in analysis. Some insist that analysis is incomplete unless a full-blown transference neurosis is developed and ultimately resolved over the course of the analysis. Others feel that transference phenomena are to be expected, but that the development of a transference neurosis as such is neither necessary nor common. Others, especially relational and intersubjective theorists, dismiss the classic notion of transference and focus on the here-and-now interaction between analysts and analysand. In classical terms, the unresolved childhood attitudes and feelings emerge and function as displacements or fantasied projections toward the analyst, and he or she becomes for the patient a phantom composite figure representing various important persons in the patient's early environment or objects represented in his or her inner world. The past relationships are reactivated with some of their original affective vigor, thus exposing in some degree the roots of the patient's disturbance. In the relational approach, transference is thought to be cocreated by both analyst and patient, and the resulting interaction is appreciated as the emerging pattern of relatedness with the analyst within the analysis.

In psychoanalytically oriented clinical work transference represents therapeutic material to be explored and understood. Understanding transference requires exploration of mechanisms involved in their formation and their dynamic interactions. The basic mechanisms by which transferences are affected include displacement, projection, and projective identification (see Table 6.1–2).

Countertransference. Just as the patient has transference, the analyst or psychiatrist has *countertransference*. This means that the analyst unconsciously experiences the patient as someone from the past. This is usually in response to the patient's transference. The concept of countertransference has evolved dramatically since its inception. Originally countertransference was seen as the analyst's transference to the patient or the analyst's response to the patient's transference. That is, the emergence of unresolved conflicts from the analyst's unconscious affecting their view of and reaction to the patient. This early view of countertransference saw it as a hindrance to treatment and a problem for the analyst that required treatment. However, in current thinking it is seen as encompassing the total affective response of the analyst to the patient, whether conscious or unconscious. This broader definition of countertransference as the analyst's natural emotional reaction to the patient conceptualizes countertransference as the therapeutic tool that is used to attune to the patient's internal world. The emphasis is on the possible positive contributions to more effective analytic work arising from attention to and utilization of countertransference responses in an analysis. Countertransference has thus become regarded as inevitable and a possible form of communication within the analytic process, not necessarily destructive to the analytic process.

Table 6.1–2.
Transference Mechanisms

Displacement

The basic mechanism of classic transference paradigms in which an object representation or representations derived from any level or combination of levels of the subject's developmental experience are displaced to the representation of the new object, namely the analyst, in the therapeutic relationship. Displacement is usually the basic mechanism for libidinally based transferences, both positive and erotic, as well as for aggressive and especially negative transferences. By and large, displacement transferences tend to play a more dominant role in neurotic disorders, in which phallic–Oedipal (and to a lesser degree pre-Oedipal) dynamics tend to play a dominant, though not exclusive, role.

Projection

Projection is the process by which qualities or characteristics of the self, specifically the self-as-object, usually involving introjections or self-representations, are attributed to the object representation of an external object, and the subsequent interaction with the object is determined by the projected characteristics. Thus the analyst-object may be seen as sadistic, that is as possessing the sadistic character of the analysand-subject, an aspect of the subject's self that is denied or disowned by the subject. Projection tends to play a more prominent, though again not exclusive, role in formation of transferences in more primitive character disorders, but can be found in variously modified forms throughout the spectrum of neuroses. Since projections derive primarily from the configuration of introjects constituting the patient's self-as-object, the effect of projective or externalizing transferences is that the image of the therapist comes to represent part of the patient's own self-organization rather than simply an object representation as such.

Projections derived from destructive introjects can provide the basis for both negative and paranoid transference reactions. Those based on the victim-introject result in the patient relating to the therapist as a victim and him- or herself assuming a correspondingly hostile or sadistic position as a destructive aggressor or victimizer in relation to the therapist's role as victim. Then again, projection based on the aggressor-introject results in the patient relating to the therapist as an aggressor and him- or herself assuming the role of a weak, vulnerable, or masochistic position in which he or she becomes a passive and vulnerable victim of the therapist's destructive aggression. Similar patterns can take place around narcissistic issues involving introjective configurations of narcissistic superiority and inferiority. From the narcissistically superior perspective, the transference takes the form of seeing the analyst as superior, idealized, valuable, and exceptionally gifted or intelligent, while the patient retains within him- or herself the image of narcissistic inferiority, worthlessness, inadequacy, and diminished self-esteem. The opposite configuration obtains from the narcissistically inferior perspective —the analyst is then seen in the light of the projected image as inferior, inadequate, worthless, and so on, while the patient retains a view of him- or herself as superior, valued and valuable, narcissistically enhanced, even grandiose and exceptional.

However, projective dynamics in self-object transferences seem to involve more than narcissistic projections, since these forms of transference tend to draw the analyst into meeting the pathologic needs of the self. If anything is projected, it would be an infantile wished-for imago, one lacking earlier in the patient's experience, as for example an empathic and idealized parental figure. On the other hand, transitional transferences, despite their considerable overlap with self-object phenomena, tend to involve a more explicit projective element as the self-related contribution to the transitional experience.

Projective Identification

The concept of projective identification was first proposed by Melanie Klein, arguing that the projection of impulses or feelings into another person brought about an identification with that person based on attribution of undesirable dimensions of one's own qualities to that other. This attribution also served as the basis for a sense of empathy and connection with the other. On these terms, projective identification was a fantasy taking place solely in the mind of the one projecting.

Projective identification is often appealed to as a mechanism of transference, or more exactly transference–countertransference interactions. Confusion arises from the failure to clearly distinguish between projection and projective identification. The notion of projective identification added to the basic concept of projection the further notes of diffusion of ego boundaries, a loss or diminishing of self-object differentiation, and inclusion of the self as part of the object.

Later elaborations of the notion of projective identification transformed it from a one-body into a two-body phenomenon, describing interaction between two subjects, one of whom projects something onto or into the other, whereupon the other introjects or internalizes what has been projected. Instead of the projection and introjection taking place in the same subject, the projection now takes place in one and the internalization in the other. This latter usage has led to extensive extrapolation of the concept of projective identification to apply to object relations of all sorts, including transference. The emphasis in Kleinian transferences is less on the influence of the past on the present, but rather the influence of the internal world on the external in the here-and-now interaction with the analyst.

It is worth noting that there can be important clinical differences between displacement and projective transferences, especially when the latter involve projective identification. Generally, in displacement transferences, the emphasis is on the patient’s experience of the analyst object as like a previous object. The process is by and large internal with little in the way of external influences to draw the object into conformity with the image. Projective transferences have a different quality, in that part of the patient’s self is experienced as in the other. This calls for differences in therapeutic handling and interpretation. Moreover, any interpretation without substantial evidence pointing to the presence of the projected qualities within the patient’s self would be less than optimal. When projective identification is in question, the same kind of evidence is required, but in addition, the projection carries with it the effort to influence the analyst to unconsciously internalize and thus enact the role that the patient seeks to impose on him or her. This can and often does elicit some form of countertransference reaction in the analyst, resulting in a transference–countertransference interaction, which can complicate and even derail the analytic relation and process.

Understanding of one’s countertransference pull(s) is quite helpful when working with patients in analysis, psychotherapy, or pharmacotherapy. By being aware we are less likely to act on our feelings (both positive and negative) in a way that is detrimental to the patient and treatment. It can also serve as useful in terms of diagnostic understanding of patients, as well as in gaining insight into the relational difficulties the patient may encounter in their lives.

Real Relationship. Every clinical relationship is a mixture of transference phenomena and the *real relationship*. Reality pervades the analytic relationship by way of the clinician’s irreducible subjectivity. Both the analyst and analysand bring their unique personalities and characteristics to the analytic encounter. These include style of behavior and speech, mannerisms, habits, way of going about the task of managing the therapeutic situation, attitudes toward the patient as a human being, prejudices, moral and political views, personal beliefs and values, gender, racial, and sexual identity among others. These aspects of the analyst’s personality and humanity are realities that play a role in the therapeutic relationship and are entirely distinct from transference, countertransference, and the alliance. The patient is curious about, observant, and sensitive to the smallest details of the analyst’s real person and life. The same considerations concerning the patient operate from the analyst’s side. Important aspects of the reality of the patient’s person and life situation play a decisive role in the analyst’s evaluation of analyzability and continue to present themselves throughout the analysis as subjects of analytic concern, inquiry, and exploration since they contribute to and affect the patient's ongoing life and experience and his or her participation in the work of the analysis. If, for example, a patient is having marital problems, this must have significant reverberations for what happens in the analysis.

Additionally, the realities of time, place, and circumstances external to the analytic setting constantly influence the course of the analytic relationship. These include realities of the location of the analyst's office, the physical surroundings, the furniture and decorations in the room, the geographic location itself, and even how the analyst dresses; they all have their effects on the analytic situation and influence how the patient experiences the person of the analyst. The surrounding circumstances that create the context for the analytic effort—the patient's financial situation, marital status, job demands, payment arrangements, insurance, pressures pushing the patient into treatment, accidental factors like illness, family, or business obligations—are all factors extrinsic to the analysis but exercise significant influence on the analytic relationship.

Therapeutic Alliance. The *therapeutic alliance* is based on the one-to-one collaborative relationship that the patient establishes with the analyst. It encompasses aspects of the analytic frame that enable the patient and analyst to engage meaningfully and productively in the analytic process to achieve therapeutic benefit. The *working alliance* contributes to or sets the stage for the effective working together of the analyst and the analysand. The terms of the alliance are negotiated between the analyst and the patient and includes at least the following elements: empathy, trust, autonomy, continuity, commitment, responsibility, management of time and money, authority, freedom, honesty, and neutrality. All these elements are as pertinent to the role of the patient in analysis as to the analyst.

The working alliance is the permanent aspect of an analysis. It allows a split in the respective selves of both patient and analyst. The observing ego of the patient allows him or her to ally with the analyst in a working relationship. This opens the way for the patient to gradually identify positively with the analyst in the analyzing function and thus modify the pathologic defenses put up by the defensive ego against the anxiety aroused by internal danger situations. Similarly, the split in the analyst allows him or her to engage affectively and emphatically with the patient while at the same time preserving an observant, reflective, and neutral analytic stance from which he or she can understand and effectively interpret the meaning of the patient’s experience and behavior.

The maintenance of the therapeutic split and the relationship to the analyst involved in the therapeutic alliance requires a certain level of ego strength. It involves capacity for self-object differentiation, tolerance, mastery of ambivalence, and separation of fantasy from reality in the relationship. When these ego capacities are weak the alliance requires work and effort to establish, and the alliance serves as one of the objectives of the analytic work. The alliance can never be assumed or taken for granted and misalliances must be carefully looked for and attuned to, to avoid crisis or impasse in analytic work. In more severely disturbed personalities, there is a tendency to disrupt the alliance rather than create misalliances, which can destroy an analytic process and requires extreme efforts to salvage the therapy.

Maintenance of the therapeutic alliance requires that the patient differentiates between the more mature and more infantile aspects of his or her experience with the analyst. The therapeutic alliance serves a double function. On the one hand, it acts as a significant buffer to excessive regression of the ego in the analytic process; on the other hand, it serves as a fundamental aspect of the analytic situation, against which the wishes, feelings, and fantasies evoked by the transference and transference neurosis can be evaluated, measured, and interpreted. In many pathologic conditions—some character neuroses, borderline personalities, and more severe neurotic disorders—it may be challenging to maintain a clinical distinction between therapeutic alliance and the transference neurosis.

The therapeutic alliance derives from the mobilization of specific ego resources relating to the capacity for object relations and reality testing. The analyst must direct attention toward eliciting the patient's capacity to establish such a relationship that will withstand the inevitable distortions and regressive aspects of the transference and transference neurosis. Inevitably, the fundamental features of the therapeutic alliance are carefully evaluated, understood, and ultimately integrated with the analysis of the transference. When therapeutic misalliances arise, as they invariably will, they can become matters of inquiry and interpretation—more often than not, they reflect unconscious and hidden transference dynamics that otherwise have not been previously recognized or explored in the analysis. Analysts should remain alert to such misalliances in that they often provide subtle indications that something is awry in the analytic work that will both require readjustment and offer the potentiality for deeper transference exploration.

Resistance. Even the most conscientious efforts on the part of the patient to engage in the work of analysis and say everything that comes to mind are never entirely successful. No matter how willingly the patient may attempt to free associate, the signs of *resistance* are apparent throughout every analysis. The manifestations of resistance are protean.

The development of resistance in the analysis is automatic and independent of the patient's will just like the development of the transference itself. The sources of resistance are just as unconscious as the sources of transference. However, the emotional forces that give rise to resistance usually defend against or express those that produce the transference. The analysis is the recurring field of conflict between the tendencies toward transference and resistance, manifested by the involuntary inhibition of the patient's efforts to associate freely.

Resistance occurs in all phases of the analysis, but its quality and significance are different, depending on the analytic task at hand. In any case, the patient's resistance enables the analyst to evaluate and become familiar with the defensive organization of the patient's ego. In this way, the pattern of resistance offers valuable information to the analyst about the patient's basic conflicts and offers a channel by which the patient can be approached therapeutically. Of note, the use of the term resistance has been waning in contemporary theory and practice with the emerging understanding of the analytic process. There is broadening of understanding of reasons why patients resist therapeutic change. Manifestations of resistance include:

- Pausing abruptly
- Correcting oneself
- Slips of the tongue
- Stammering
- Remaining silent
- Fidgeting, for example, with a part of clothing
- Asking irrelevant questions
- Intellectualizing
- Arriving late for appointments
- Finding excuses not to keep appointments
- Criticizing the rationale underlying the treatment method
- Being unable to think of anything to say
- Censoring thoughts that occur, deciding they are banal, uninteresting, or irrelevant and not worth mentioning

TECHNICAL ASPECTS

Analytic technique is adapted to the idiosyncrasies of the patient's developmental capacities, needs, conflicts, compromises, and defensive constellation. It is part of the dynamic analytic process intended to induce and achieve significant internal psychic growth. A few of the more salient techniques are as follows.

Free Association

The cornerstone of the psychoanalytic technique is *free association*. As many have noted, the process of associating is not altogether free but is to a certain degree guided by the expectations of the analytic setting and the analyst's responses toward more meaningful or affectively loaded or conflictual material. The patient is encouraged to free associate as far as possible throughout the treatment. Besides providing analytic content, the primary function of free association is to help induce the necessary regression and relatively passive dependence connected with establishing and working through the transference. Thus, free association is conjoined with the other techniques that induce such regression, namely lying on the couch, not seeing the analyst, and conducting the analysis in an atmosphere of quiet reflection and restful tranquility.

The patient's free association reflects the complex influences arising from transference, the real relationship, and the therapeutic alliance. The process of free association is complex, challenging to conceptualize, and increasingly seen in the context of the relationship between analyst and patient. From a contemporary perspective much more is required of a patient than simply free associating. It is not enough for the patient to lie back and self-surrender to passive dependency within the analytic relationship without also being able to mobilize basic ego resources in the service of mastery, insight, mobilization of executive and synthetic capacities, and ultimately assumption of a less passive and more active and autonomous function within the analytic relationship. This reflects the analytic split discussed previously. There is a gradation in the mobilization of these capacities in the patient, which varies during the analytic process, optimally leading to more mature and adaptive levels as the analysis progresses, especially in the progression to the termination.

Interpretation

Interpretation was long regarded as the chief tool of the analyst in transference analysis. Current views tend to give it a more modest role. It is now conceived as part of the relational process of unconscious communication and meaning creation taking place in the clinical field. As such, it is an important technical tool among various other influences, including aspects of the analytic relationship and the quality of the analytic interaction. In the early stages of psychoanalytic therapeutic development, interpretation was used primarily to inform the patient of his or her unconscious wishes. Later, it was designed to help the patient understand the resistance to spontaneous self-awareness. In current psychoanalytic practice, the analyst's function as an interpreter is not limited to simply paraphrasing the patient's verbal reports but rather indicating what is not reported or implicit in what is reported. Consequently, analytic interpretation does not produce immediate symptomatic relief. On the contrary, even after accurate and clarifying interpretations, there may be a heightening of anxiety and emergence of further resistance.

If a correct (mutative) interpretation is given with proper timing, the patient may react immediately or after a period of emotional struggle, during which new associations are offered. These new associations often confirm the validity of previous interpretations and add significant additional data, thus disclosing motivations and experiences of the patient of which the analyst could not previously have been aware. It is not the analyst's insight into the patient's psychodynamics that produces progress in the analysis as the patient can gain their own insight, whether independently or in collaboration with the analyst. The analyst can facilitate this process by reducing unconscious resistance to self-awareness through appropriate, carefully timed interpretations. The most effective interpretation is timed to meet the patient's emerging, if hesitant and half-formed, awareness. Thus, the analyst must gauge the patient's capacity at any given moment to hear, assimilate, and integrate the content of a given interpretation.

Another critical aspect of an interpretation is that it cannot be seen in isolation from the total context of the analytic situation and the analytic process. As given by the analyst and received by the patient, an interpretation occurs within the therapeutic relationship. It may focus on elements of either transference or therapeutic alliance, or both. Giving and receiving interpretations is cloaked with meanings that influence the patient's capacity to accept and integrate interpretations and the analyst's sense of offering and providing such interpretations. Experience has shown that, at best, the therapeutic benefits produced by the analyst's explanations or unilaterally provided insights are only temporary. The most effective interpretations of lasting therapeutic value are those arrived at by the delicate dialectic arising from the mutually facilitated and growing awareness of both patient and analyst.

Interpretations are best offered in a spirit of inquiry, as hypotheses suggested for the patient's consideration and remaining open to his or her further consideration, acceptance, questioning, modification, or rejection. Interpretations, in a sense, are synonymously invitations for the patient to exercise their curiosity, reasoning, questioning, and judgment and quasiexperiments in autonomous self-reflection and analysis. Any interpretation the analyst might offer is no more than a piece of sophisticated guesswork about what the patient is thinking or meaning. Contrary to what some analysts may have thought, the analyst cannot read the patient's mind. Thus, interpretations can have little meaning or effect unless scrutinized, evaluated, and found to have application or relevance that the patient can acknowledge and accept. This cannot be achieved without the patient's active participation in the process of inquiry, discovery, and understanding of his or her unconscious thinking and motivation. In the contemporary view the quality of the relationship between analyst and patient is viewed as promoting beneficial change. The advances in neuroscience, on the other hand, emphasize the role of nonconscious, in therapeutic action. More broadly, there is a trend towards interest in and focus on how people think rather than uncovering the unconscious mental contents.

CONCLUDING REFLECTIONS

Psychoanalysis began as a revolutionary intellectual tradition and psychoanalytic ideas remain radical and relevant even as they continue to evolve and integrate the new knowledge emerging from the developments in genetic, neuroscience, and other related fields. Psychoanalytic theory is very much alive today and taking on the challenge of considering the social and cultural context, becoming more sophisticated in appreciating the gene–environment interaction, and exploring the relevance of psychoanalytic ideas and practice to the community at large. Psychoanalytic ideas continue to inspire important theories of psychological functioning and psychopathology, as well as psychological treatment although the evidence base for effectiveness of psychoanalysis itself is lagging.

Psychodynamic approach to treatment is unique in its systematic attention to the clinician's conscious and unconscious contribution to the process of treatment and evaluation. The exposure to psychoanalytic ideas prepares the psychiatrist to recognize the characterologic resistances to treatment, understand the pharmacologic noncompliance in patients along the lines of transference, countertransference, and resistance or explore the meaning of medications, thus enriching and broadening the scope of psychiatrist's practice.

FURTHER READINGS

Cassidy J, Shaver PR. eds. *Handbook of Attachment: Theory, Research, and Clinical Applications*. Guilford Publications; 2016.

Diamond D, Blatt SJ, eds. Attachment research and psychoanalysis. 1. Theoretical considerations. 2. Clinical implications. *Psychoanal Inquiry*. 1999;19:4.

Diamond D, Blatt SJ, Lichtenberg J. Attachment research and psychoanalysis. 3. Further reflections on theory and clinical experience. *Psychoanalytic Inquiry*. 2003;23:1.

Fenichel O. *The Psychoanalytic Theory of Neurosis*. Norton; 1945.

Fonagy P, Target M. *Psychoanalytic theories: Perspectives from developmental psychopathology*. Whurr publishers; 2003.

Freud A. *The Writings of Anna Freud*. 7 vols. International Universities Press; 1965–1974.

Freud S. *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. 24 vols. Hogarth Press; 1953–1974.

Freud GP. *A Life for Our Time*. Norton; 1988.

Goodman G. *The Internal World and Attachment*. Analytic Press; 2002.

Greenberg JR, Mitchell SA. *Object Relations in Psychoanalytic Theory*. Harvard University Press; 1983.

Hartmann H. *Essays on Ego Psychology*. International Universities Press; 1954.

Hartmann H. *Ego Psychology and the Problem of Adaptation*. International Universities Press; 1958.

Hinshelwood RD. *A Dictionary of Kleinian Thought*. Free Association Books; 1991.

Kernberg OF. A psychoanalytic classification of character pathology. *J Am Psychoanal Assoc.* 1970;18:800–822.
Kernberg OF. Contrasting viewpoints regarding the nature and psychoanalytic treatment of narcissistic personalities: a preliminary communication. *J Am Psychoanal Assoc.* 1974;22:255–267.
Kohut HS. *The Analysis of the Self*. International Universities Press; 1971.
Kohut HS. *The Restoration of the Self*. International Universities Press; 1977.
Kohut HS. *How Does Analysis Cure?* University of Chicago Press; 1984.
Laplanche J, Pontalis J-B. *The Language of Psycho-analysis*. Norton; 1973.
Lazar A, Sandell R, Grant J. Do psychoanalytic treatments have positive effects on health and health care utilization? Further findings of the Stockholm Outcome of Psychotherapy and Psychoanalysis Project (STOPPP). *Psychother Res*. 2006;16:51–66.
Loewald HW. *Papers on Psychoanalysis*. Yale University Press; 1980.
Mahler MS, Pine F, Bergman A. *The Psychological Birth of the Human Infant*. Basic Books; 1975.
Main M, Solomon J. Discovery of an insecure-disorganized/disoriented attachment pattern. In: Brazelton TB, Yogman MW. eds. *Affective development in infancy*. Ablex Publishing; 1986:95–124.
Meissner WW. *The Therapeutic Alliance*. Yale University Press; 1996.
Meissner WW. *The Ethical Dimension of Psychoanalysis—A Dialogue*. State University of New York Press; 2003.
Mitchell SA. *Relational Concepts in Psychoanalysis*. Harvard University Press; 1988.
Shedler J. Where is the evidence for “Evidence-Based” therapy? *Psychiatr Clin North Am*. 2018;41:319–329.
Solms ML. The neurobiological underpinnings of psychoanalytic theory and therapy. *Front Behav Neurosci.* 2018;12:294.
Target M. On psychoanalysis. *New Sci*. 2010;208:30–31.

6.2 Erik H. Erikson

DORIAN S. NEWTON, PH.D., AND SUDHIR KAKAR, PH.D.

“I am what survives of me.”

Erik H. Erikson was an artist, teacher, anthropologic field worker, and psychoanalyst. His legacy includes an original and authoritative theory of individual psychosocial development with cross-cultural and universal applications. Previous conceptions of personality development had been centered on infancy and childhood; Erikson's extended across the life span and was anchored in familial, societal, and historical contexts. His was a model of human experience which emphasized adaptation and growth over fixation and pathology.

A prolific writer, Erikson is best known for three major works. In *Childhood and Society*, Erikson advanced a theory of the life cycle which traced the trajectory of individual psychosocial development through the maturing ego's relations with an expanding social world. In *Young Man Luther*, Erikson reconstructed and analyzed the youthful monk's identity crisis and its creative resolution in early adulthood. *Gandhi's Truth* demonstrated the union of authentic identity and calling in a man at midlife. Through his psychological biographies of these historical figures, Erikson elucidated the interrelations between individual psychodynamics and development on the one hand and social structure and history on the other, deftly avoiding a reductionism in either a psychodynamic or sociologic direction.

Sixty-five years after the publication of *Childhood and Society*, Erikson's ideas on identity development, intimacy, and generativity continue to inspire research and engender debate.

CHILDHOOD AND ADOLESCENCE

Erik Homburger Erikson was born on June 15, 1902, in Karlsruhe, Germany, the old capital of a Lutheran principality. His father was Protestant, and his mother was Jewish. His parents, both Danish, had separated before his birth; his mother was visiting friends in Germany when the baby was born. She stayed in Karlsruhe and a few years later married her child's pediatrician, a well-to-do Jewish doctor, Theodor Homburger, in whose house young Erik grew up. During an alienated adolescence, the tall, blond boy found himself regarded as a gentile in his father's Jewish milieu and as a Jew at school. He remembered his mother as sad, bookish, and artistic; his adoptive father as professionally respected; and both as loving. The boy’s adoption by his stepfather was the formative occasion of what became a lifelong pattern of getting himself adopted by kind men. His last name remained Homburger until he was 37 when he changed it to Erikson.

Erikson attended the *Humanistische Gymnasium* in Karlsruhe, where he studied Greek, Latin, philosophy, literature, ancient history, art, and science. His primary interest was art; impatient with formal study and possessed by a restlessness he never lost, Erikson chose not to go to a university, preferring to travel about the countryside reading, drawing, and making wood carvings. Back home after a year, he tried formal art study with some success, first in Karlsruhe at the Badische Landeskunstschule and then in Munich at the Kunst-Akademie, but in neither case with decisive commitment.

This sort of wandering about was not uncommon among German youth of the period, so Erikson was permitted, as his biographer Robert Coles sagely wrote, “to go through his own years of discontent and confusion without being especially singled out and thereby forced to defend behavior often best granted the limits of its own momentum.” As Erikson confided to Coles, “ … if ever an identity crisis was central and long drawn out in somebody’s life it was so in mine.” Erikson was having what he would later term a *psychosocial moratorium*; in so doing, he was also mitigating the asperities of an *identity crisis*. At about age 21, Erikson went to live in Florence, where he continued his art studies informally. There he enjoyed the friendship of his old *Gymnasium* chum Peter Blos, a writer who later became a famous American child psychoanalyst.

BECOMING A CHILD PSYCHOANALYST

When Erikson was 25 years of age, Peter Blos invited him to become a faculty member in a progressive grammar school in Vienna, where Blos taught language and science. Erikson’s difficult transition from adolescence to early adulthood was over. The year was 1927; Sigmund Freud was 71, and his youngest child, Anna, an educator and psychoanalyst, had started a psychoanalytically enlightened school for children with an American friend named

Dorothy Burlingham. Erikson joined Blos and later recalled Blos's determination to turn him into a disciplined worker: "To make a teacher of me…the highly disciplined Peter first had to teach me to keep regular work hours, a task which was initiated every morning, no matter what time of year, by a cold shower, then the preferred shock treatment for identity confusion." At the school, "Herr Erik" taught his students art and history; together, they studied different cultures and illustrated what they learned with drawings, essays, toys, tools, and exhibits.

Before long, Erikson found himself not only a teacher of children but also an *analysand*, what is now called a *candidate*, at the Vienna Psychoanalytic Institute, in treatment with Anna Freud. Any sort of psychoanalytic approach to treating or educating children was then a radical idea, even an approach as cautious as Freud's. Both educationally and clinically, Freud's group was sufficiently deviant such that a person with Erikson's diverse identity was able to fit in. There was in 1927 a *configurational affinity* between Erikson's personal history and the history of psychoanalysis as a profession. Still, there was the worry: What was a fledgling artist without a university education to do among those high-powered theorists and intellectuals at the Vienna Psychoanalytic Institute?

Erikson recalled this epiphanic exchange with his analyst, Anna Freud: "When I declared once more that I could not see a place for my artistic inclinations in such high intellectual endeavors, she said quietly: "You might help to make them see."" Dreams had been Sigmund Freud's royal road to the unconscious; observing children's play (and later, anthropologic field studies in the role of participant-observer) would be Erikson's path to understanding the ego and its development. Looking back, Erikson thought that it was Anna Freud's "simple mandate" that enabled him to succeed in combining the artistic and the theoretical in *Childhood and Society*.

Erikson remained in Vienna for 6 years until 1933. In those years between the ages of 25 and 31, he turned his artist's eye from the observation of nature to the analysis of children; he learned about psychoanalysis by studying children's play and his own free associations as a patient, and he learned fundamental clinical skills in the supervised treatment of others. As he explained to Coles, Erikson related art to psychoanalysis in this way, “I began to perceive how important visual configurations were, how they actually preceded words and formulations: certainly dreams are visual data, and so is children’s play, not to speak of the “free associations” which often are a series of images, pure and simple–only later put into words.” Erikson trained with a Montessori group in Vienna, graduating from the Montessori teacher’s association, the Lehrerinnenverein. He studied psychoanalysis with August Aichhorn, Edward Bibring, Helene Deutsche, and Ernst Kris. He formed mentor relationships with Anna Freud and Heinz Hartmann and a more distant, admiring-inspirational relationship with Freud’s sick, aging, but still-productive father.

During his own late adulthood, Sigmund Freud had moved beyond clinical problems more narrowly conceived to problems of the ego, society, and history in works such as *Beyond the Pleasure Principle* (written at age 64), *Group Psychology and the Analysis of the Ego* (at 65), and *The Ego and the Id* (at 66); during Erikson’s years in Vienna, Freud was working on *The Future of an Illusion* (at 71) and *Civilization and Its Discontents* (at 73). Erikson seemed to identify with both Freuds, internalizing them as mentors, and they provided him with a strong, inner basis for a lifetime of psychoanalytic treatment and psychosocial research.

In 1929, Erikson met Joan Serson, a woman of mixed Canadian and American backgrounds. She had a Master’s degree in sociology, a special interest in the history and social origins of modern dance, and in psychoanalysis. They married, and Serson joined the faculty of Burlingham’s school, where she taught English, literature, and American history. Serson’s unusual combination of interests and skills—education, psychoanalysis, and sociology—coupled with her ability as a writer gave Erikson a skilled coworker for a lifetime of intellectual work. Several articles were coauthored, Joan Erikson helped with most of the others, and the endowed chair at Harvard that was named after Erikson carried both their first names.

With the birth of their sons, Kai and Jon (daughter, Sue, was to be born five years later, and their last son, Neil, was born with Down syndrome and spent most of his life institutionalized), Erikson had by 31 become a husband, a father, and a child psychoanalyst. He had come a long way from the aimless youth of his early 20s and was productively engaged in the stage of ego development that he later characterized by the polarity *intimacy versus isolation*. He continued to work on that ego stage, but changes were now required in the life that Erikson had begun to build at 25.

EMIGRATING TO AMERICA

By 1933 the Fascist menace in Europe was growing. When Erikson graduated that year from the Vienna Psychoanalytic Institute, he and his family prepared to leave. They considered repatriating to Denmark but encountered problems because Erikson had lost his Danish citizenship when he had been adopted and become a naturalized German. A serendipitous meeting between Erikson and Freud's disciple Hanns Sachs in which Sachs enthusiastically invited Erikson to come to Boston settled the matter. In 1933 the Eriksons emigrated to the United States. At 31, Erikson became Boston's first child psychoanalyst and a member of the faculty of the Harvard Medical School. He took a job as a consultant at Judge Baker Guidance Center, where he helped to diagnose and treat poor and delinquent children with emotional disorders. In Cambridge, Erikson met Margaret Mead, Gregory Bateson, Ruth Benedict, and Kurt Lewin, each of whom influenced his intellectual development in important ways.

Erikson joined a group of personality researchers working under the leadership of Henry Murray, who was the Director of the Harvard Psychological Clinic. Murray's book *Explorations in Personality: A Clinical and Experimental Study of Fifty Men of College Age* was published in 1938. It is one of the few integrative masterpieces in American psychology—experimental and biographical, psychological and sociologic, Freudian and Jungian, developmental and diagnostic—and it provided a luminous example of studying persons both in their origins and in their purposes. Murray's example proved seminal when Erikson later turned his hand to the biographical study of the world-historical figures Sigmund Freud, Martin Luther, and Mohandas Gandhi.

Listed on the title page of Murray's book, among other research associates, was "Erik Homburger." The next year, at age 37, Homburger renamed himself Erikson, retaining Homburger as his middle name. With characteristically bold creativity, Erikson solved the problem of his paternity by adopting himself.

PSYCHOANALYTIC ANTHROPOLOGY

In 1936 Erikson hit the road again, this time to the Institute of Human Relations of the Department of Psychiatry at Yale University. The interdisciplinary work of the institute further shaped Erikson's interest in cross-cultural research, and in 1938 he joined a colleague on a research expedition to the Sioux Indians in South Dakota. On the Pine Ridge reservation, he observed children, interviewed adults, and noted the impact of economic, geographical, and historical factors on child-rearing practices. Based upon this research, Erikson published his paper, "Observations on Sioux Education," in 1939.

In the same year, the peregrinating Eriksons moved to California, where he joined the faculty at the University of California at Berkeley. Erikson affiliated himself with Mount Zion Hospital and the San Francisco Psychoanalytic Institute and ultimately became a U.S. citizen. At Berkeley, Erikson continued his cross-cultural research, this time with the Yurok Indians of northern California. He also lent his research skills to the war effort in articles on submarine habitation, the interrogation of prisoners of war, and difficulties encountered by veterans in returning to civilian life.

Erikson remained at U.C. Berkeley for a decade, the longest period in one place since his preadulthood years in Karlsruhe. During that time, he became his own man, visible first in the outward signs of changing his name and later in refusing to sign the university's loyalty oath of anti-Communist purity. To Erikson, the contract was " … an empty gesture toward meeting the danger of infiltration into academic life of indoctrinators, conspirators, and spies." In 1950, following the dismissal of some of his peers for refusing to declare themselves noncommunists, Erikson resigned from his position as professor of psychology.

At Harvard, Erikson had studied “normal” students and their use of toys and dramatic scenes to enact and illustrate internal conflicts. At the Institute of Child Welfare at the University of California, Erikson continued his research and found important sex-related differences in the use of toys and play space by “normal” adolescents. As Coles noted,

> “Psychological themes—what the child says he is up to, what he can be seen doing and experiencing—are related to “spatial configurations,” namely, objects and forms that exist in a world outside the mind. The mind uses those objects or forms to express wishes, fears, and conflicts.” (Excerpted from Coles R. *Erik H. Erikson: The Growth of His Work*. Little, Brown and Company; 1970.)

In his research which had its roots in his work at Burlinghams’s experimental school, Erikson had once again combined his artistic and psychoanalytic sensibilities in the study of human behavior.

But at age 40, as Erikson was entering middle age, he was becoming more biographical and more interested in the adult years of the life cycle. Like others who manage to continue developing, he found a way to pursue his development through the strictures of exigent reality. World War II was underway, and Erikson was deeply concerned about it. After his initial narrowly framed attempts at scholarly patriotism, he began writing his first psychobiographical essays on Adolf Hitler and the psychosocial dynamics of his appeal to young Germans—“Hitler’s Imagery and German Youth” was published in 1942. It united Erikson’s interests in political science, history, and anthropology. As Coles stated, “Very little that Erikson would do for the next two decades was not in some respects foreshadowed by this paper.”

MIDLIFE TRANSITION

Childhood and Society was the creative product of Erikson’s transition to midlife. He had intended it to be a contribution to the psychiatric education of clinicians from various disciplines, but the book outgrew its author’s intentions and found its way into every corner of the academy and beyond.

Erikson began work on the book when he was 42 and largely completed work on it by 46. Published in 1950, it is at once a product of early research, a prospectus of what was to come, and an initial integration of both. In it, he presented clinical cases in which individual psychodynamics, society, and history are interwoven with a skill not seen before or since; analyses of children’s play and development in various cultures; a theoretical sketch of the entire human life cycle; pieces on the problem of identity; and biographical essays on Adolf Hitler and the Russian writer Maxim Gorky.

As in Karl Abraham’s and Freud’s psychosexual stages, most of Erikson’s stages in ego development occur in childhood and adolescence. Unlike the Abraham-Freud conception, however, Erikson’s stages are psychosocial, describing crucial steps in the maturing ego’s relations with the social world rather than a biologic unfolding of neurophysiologic capacities for excitation. Also, whereas Freud’s developmental theory falls back on itself after adolescence, Erikson’s continued with characterizations of developmental tasks during youth, middle age, and old age.

Erikson’s emphasis, like Freud’s, was both cross-cultural and universal. Erikson’s eight stages were ineluctable parts of the human life cycle, yet each person traversed them in distinct ways determined by culture, concrete circumstance, and personality. With *Childhood and Society*, Erikson became and remained an ego psychologist, shifting the traditional psychoanalytic focus on the drives to adaptation and growth. In so doing, he was furthering the work of his analyst-mentor, Anna Freud, who in 1936 had written ego psychology’s basic theoretical treatise, *The Ego and The Mechanisms of Defense.*

EIGHT STAGES OF THE LIFE CYCLE

Erikson’s conceptualization of psychosocial development across the life cycle is the centerpiece of his life’s work, and he elaborated the conception throughout his subsequent writings. It takes as its model the epigenetic principle of organismic growth in utero. In Erikson's view, psychosocial growth occurs in phases, with individual aspects of development proceeding according to a predetermined timetable. Elements of each phase are present from birth and differentiate over time. Every phase has its own period of quiescence and critical ascendancy, and each is dependent on the proper development of the other phases in the proper sequence. Work on any particular phase, Erikson theorized, is never complete, and old developmental conflicts can be activated by critical life events.

The eight stages of the life cycle represent points along a continuum of development in which physical, cognitive, instinctual, and sexual changes combine to trigger an internal crisis whose resolution results in either psychosocial regression or growth and the development of specific *virtues*. In *Insight and Responsibility,* Erikson defined virtue as "inherent strength," as in the active quality of a medicine or a liquor. He wrote in *Identity: Youth and Crisis* that "crisis" refers not to a "threat of catastrophe, but to a turning point, a crucial period of increased vulnerability and heightened potential, and therefore, the ontogenetic source of generational strength and maladjustment." Elsewhere, Erikson averred, "…we do not consider all development a series of crises: we claim only that psychosocial development proceeds by critical steps—"critical" being a characteristic of turning points, of moments of decision between progress and regression, integration and retardation." A child's experience of each crisis or critical turning point is affected by the attitudes of its parents and the values and customs of its culture.

Trust versus Mistrust (Birth to about 18 Months) In *Identity: Youth and Crisis*, Erikson noted that the infant "lives through and loves with" its mouth. Indeed, the mouth forms the basis of its first *mode* or pattern of behavior, that of incorporation. The infant is taking the world in through the mouth, eyes, ears, and sense of touch. The baby is learning a cultural modality that Erikson termed *to get*, that is, to receive what is offered and elicit what is desired. As the infant's teeth develop and it discovers the pleasure of biting, it enters the second oral stage, the active-incorporative mode. The infant is no longer passively receptive to stimuli; it reaches out for sensation and grasps its surroundings. The social modality shifts to that of *taking and holding on* to things.

The infant's development of basic trust in the world stems from its earliest experiences with its mother or primary caretaker. In *Childhood and Society,* Erikson asserts that trust depends not on "absolute quantities of food or demonstrations of love, but rather on the quality of maternal relationship." A baby whose mother is able to anticipate and respond to its needs in a consistent and timely manner despite its oral aggression will learn to tolerate the inevitable moments of frustration and deprivation. The defense mechanisms of introjection and projection will provide the infant with the means to internalize pleasure and externalize pain such that "consistency, continuity, and sameness of experience provide a rudimentary sense of ego identity." Trust will predominate over mistrust, and hope will crystallize. For Erikson, the element of society corresponding to this stage of ego identity is religion, as both are founded upon "trust born of care."

In keeping with his emphasis on the epigenetic character of psychosocial change, Erikson conceived of many forms of psychopathology as examples of what he termed *aggravated development crisis*, development that, having gone awry at one point, affects subsequent psychosocial change. A person who, as a result of severe disturbances in the earliest dyadic relationships, fails to develop a basic sense of trust or the virtue of hope may be predisposed as an adult to the profound withdrawal and regression characteristic of schizophrenia. Erikson hypothesized that the depressed patient's experience of being empty and worthless is an outgrowth of a developmental derailment that causes oral pessimism to predominate. Addictions may also be traced to the mode of oral incorporation.

Autonomy versus Shame and Doubt (about 18 Months to About 3 Years) In the development of speech and sphincter and muscular control, the toddler practices the social modalities of *holding on and letting go* and experiences the first stirrings of the virtue that Erikson termed *will*. Much depends on the amount and the type of control exercised by adults over the child. Control that is exerted too rigidly or too early defeats the toddler’s attempts to develop its own internal controls, and regression or false progression results. Parental control that fails to protect the toddler from the consequences of his or her own lack of self-control or judgment can be equally disastrous to the child’s development of a healthy sense of autonomy. In *Identity: Youth and Crisis*, Erikson asserted: “This stage, therefore, becomes decisive for the ratio between cooperation and willfulness, and between self-expression and compulsive self-restraint or meek compliance.”

Where that ratio is favorable, the child will develop an appropriate sense of autonomy and the capacity to “have and to hold”; where it is unfavorable, doubt and shame will undermine free will. According to Erikson, the principle of law and order has at its roots this early preoccupation with the protection and regulation of will. In *Childhood and Society*, he concluded, “[T]he sense of autonomy fostered in the child and modified as life progresses, serves (and is served by) the preservation in economic and political life of a sense of justice.”

An individual who becomes fixated on the transition between the development of hope and autonomous will, with its residue of mistrust and doubt, may develop paranoiac fears of persecution. When psychosocial development is derailed in the second stage, delinquency and other forms of pathology may emerge. The perfectionism, inflexibility, and stinginess of the person with an obsessive-compulsive personality disorder may stem from conflicting tendencies to hold on and to let go. The ruminative and ritualistic behavior of the person who suffers from an obsessive-compulsive disorder may be an outcome of the triumph of doubt over autonomy and the subsequent development and retention of a primitively harsh conscience.

Initiative versus Guilt (about 3 years to about 5 Years) The child’s increasing mastery of locomotor and language skills expands its participation in the outside world and stimulates omnipotent fantasies of wider exploration and conquest. Here the youngster’s mode of participation is active and intrusive; its social modality is that of *being on the make*. The intrusiveness is manifested in the child’s fervent curiosity and genital preoccupations, competitiveness, and physical aggression. The Oedipus complex is in ascendance as the child competes with the same-sex parent for the fantasized possession of the other parent. In *Identity: Youth and Crisis*, Erikson wrote that “Jealousy and rivalry… now come to climax in a final contest for a favored position with one of the parents: the inevitable and necessary failure leads to guilt and anxiety.”

Guilt over the drive for conquest and anxiety over the anticipated punishment are both assuaged in the child through repression of the forbidden wishes and the development of a superego to regulate its initiative. This conscience, the faculty of self-observation, self-regulation, and self-punishment, is an internalized version of parental and societal authority. Initially, the conscience is harsh and uncompromising; however, it constitutes the foundation for the subsequent development of morality. Having renounced oedipal ambitions, the child begins to look outside the family for arenas in which it can compete with less conflict and guilt. This is the stage that highlights the child’s expanding initiative and forms the basis for the subsequent development of realistic ambition and the virtue of *purpose*. As Erikson noted in *Childhood and Society*, “[T]he “oedipal” stage…sets the direction toward the possible and the tangible which permits the dreams of early childhood to be attached to the goals of an active adult life.” Toward this end, social institutions provide the youngster with an economic ethos in the form of adult heroes who begin to take the place of their storybook counterparts.

When there has been an inadequate resolution of the conflict between initiative and guilt, the individual may ultimately develop a conversion disorder, inhibition, or phobia. Those who overcompensate for the conflict by driving themselves too hard may experience so much stress as to produce psychosomatic symptoms.

Industry versus Inferiority (about 5 Years to about 13 Years) With the onset of latency, the child discovers the pleasures of production. He or she develops industry by learning new skills and takes pride in the things made. Erikson wrote in *Childhood and Society* that the child’s “ego boundaries include his tools and skills: the work principle…teaches him the pleasure of work completion by steady attention and persevering diligence.” Across cultures, this is a time when the child receives systematic instruction and learns the fundamentals of technology as they pertain to the use of basic utensils and tools. As children work, they identify with their teachers and imagine themselves in various occupational roles.

If the child is unprepared for this stage of psychosocial development, either through an insufficient resolution of previous stages or by current interference, the child may develop a sense of inferiority and inadequacy. Teachers and other role models help to transmit social values and become crucially important in the child's ability to overcome a sense of inferiority and to achieve the virtue known as *competence*. In *Identity: Youth and Crisis*, Erikson noted: "[T]his is socially a most decisive stage. Since industry involves doing things beside and with others, a first sense of division of labor and of differential opportunity, that is, a sense of the technologic ethos of a culture, develops at this time."

The pathologic outcome of a poorly navigated stage of industry versus inferiority is less well defined than in previous stages, but it may concern the emergence of a conformist immersion into the world of production in which creativity is stifled, identity is subsumed under the worker's role, and feelings of inferiority are warded off through a defensive preoccupation with status and compensation.

Identity versus Role Confusion (about 13 Years to about 21 Years) With the onset of puberty and its myriad social and physiologic changes, the adolescent becomes preoccupied with the question of identity. Erikson noted in *Childhood and Society* that youth are now "primarily concerned with what they appear to be in the eyes of others as compared to what they feel they are, and with the question of how to connect the roles and skills cultivated earlier with the occupational prototypes of the day." Childhood roles and fantasies are no longer appropriate, yet the adolescent is far from equipped to become an adult. In *Childhood and Society,* Erikson writes that the integration that occurs in the formation of ego identity encompasses far more than the summation of childhood identifications. "It is the accrued experience of the ego's ability to integrate these identifications with the vicissitudes of the libido, with the aptitudes developed out of endowment, and with the opportunities offered in social roles.” Coles explains that ego identity is an “accrued confidence” which develops over time and crystallizes in early adulthood—or does not. It is the confidence that “… somehow in the midst of change one *is*; that is, one has an “inner sameness and continuity” which others can recognize and which is so certain that it can unselfconsciously be taken for granted.”

The formation of cliques and the intolerance of individual differences are ways in which the young person attempts to ward off a sense of identity confusion. In *Childhood and Society*, Erikson noted:

> For adolescents not only help one another temporarily through much discomfort by forming cliques and by stereotyping themselves, their ideals, and their enemies; they also perversely test each other’s capacity to pledge fidelity. The readiness for such testing also explains the appeal that simple and cruel totalitarian doctrines have on the minds of the youth of such countries and classes as have lost or are losing their group identities (feudal, agrarian, tribal, national) and face worldwide industrialization, emancipation, and wider communication. (Excerpt from Erikson EH. *Childhood and Society*. W.W Norton & Company; 1963.)

Falling in love, a process by which the adolescent may clarify a sense of identity by projecting a diffused self-image onto the partner and seeing it gradually assume a more distinctive shape, and an overidentification with idealized figures are means by which the adolescent seeks self-definition. With the attainment of a more sharply focused identity, the youth develops the virtue of *fidelity*—a faithfulness not only to the nascent self-definition but to an ideology that provides a version of self-in-world. As Erikson, Joan Erikson, and Helen Kivnick wrote in *Vital Involvement in Old Age*: “[F]idelity is the ability to sustain loyalties freely pledged in spite of the inevitable contradictions of value systems. It is the cornerstone of identity and receives inspiration from confirming ideologies and affirming companionships.”

Role confusion ensues when the youth is unable to formulate a sense of identity and belonging. Coles explains:

> In a state of “acute identity diffusion” the young individual may feel isolated, empty, anxious, and unable to make decisions that he (let alone his parents or

> teachers) feels pending. He wants "out," he wants to be away, he wants "time" to think and decide and only later act. (Excerpted from Coles R. *Erik H. Erikson: The Growth of His Work*. Little, Brown and Company; 1970.)

Erikson held that delinquency, gender-related identity disorders, and borderline psychotic episodes can result from such confusion.

Intimacy versus Isolation (about 21 Years to about 40 Years) Freud's famous response to the question of what a normal person should be able to do well, "Lieben und arbeiten" (to love and to work), is one that Erikson often cited in his discussion of this psychosocial stage, and it emphasizes the importance he placed on the virtue of *love* within a balanced identity. Erikson asserted in *Identity: Youth and Crisis* that Freud's use of the term love referred to "the generosity of intimacy as well as genital love; when he said love and work, he meant a general work productivity which would not preoccupy the individual to the extent that he might lose his right or capacity to be a sexual and a loving being."

Intimacy in the young adult is closely tied to fidelity; it is the ability to make and honor commitments to concrete affiliations and partnerships even when that requires sacrifice and compromise. The person who cannot tolerate the fear of ego loss arising out of experiences of self-abandonment (e.g., sexual orgasm, moments of intensity in friendships, aggression, inspiration, and intuition) is apt to become deeply isolated and self-absorbed. *Distantiation,* an awkward term coined by Erikson to mean "the readiness to repudiate, isolate, and, if necessary, destroy those forces and people whose essence seems dangerous to one's own," is the pathologic outcome of conflicts surrounding intimacy and, in the absence of an ethical sense where intimate, competitive, and combative relationships are differentiated, forms the basis for various forms of prejudice, persecution, and psychopathology.

Erikson's separation of the psychosocial task of achieving identity from that of achieving intimacy and his assertion that substantial progress on the former task must precede development on the latter have engendered much criticism and debate. Critics have argued that Erikson's emphasis on separation and occupationally based identity formation fails to take into account the importance for women of continued attachment and the formation of an identity based on relationships.

Generativity versus Stagnation (about 40 Years to about 60 Years) Erikson asserted in *Identity: Youth and Crisis* that "Generativity…is primarily the concern for establishing and guiding the next generation." The term *generativity* applies not so much to rearing and teaching one's offspring as to a protective concern for all the generations and for social institutions. It encompasses productivity and creativity as well. Having previously achieved the capacity to form intimate relationships, the person now broadens the investment of ego and libidinal energy to include groups, organizations, and society. *Care* is the virtue that coalesces at this stage. In *Childhood and Society,* Erikson emphasized the importance to the mature person of feeling needed. "[M]aturity needs guidance as well as encouragement from what has been produced and must be taken care of." Through generative behavior, the individual is able to pass on knowledge and skills while obtaining a measure of satisfaction in having achieved a role with senior authority and responsibility in the tribe.

When people are unable to develop true generativity, they may settle for pseudo-engagement in occupation. Often such people restrict their focus to the technical aspects of their roles at which they may now have become highly skilled, eschewing larger responsibility for the organization or profession. This failure of generativity can lead to profound personal stagnation, masked by a variety of escapisms, such as alcohol and drug abuse and sexual and other infidelities. A midlife crisis or premature invalidism (physical and psychological) may occur. In this case, pathology appears not only in middle-aged persons but also in the organizations that depend upon them for leadership. Thus, the failure to develop at midlife can lead to sick, withered, or destructive organizations that spread the effects of failed generativity throughout society; examples of such failures have become so common as to constitute a defining feature of modernity.

Integrity versus Despair (about 60 Years to Death) In *Identity: Youth and Crisis*, Erikson defined integrity as “the acceptance of one’s one and only life cycle and of the people who have become significant to it as something that had to be and that, by necessity, permitted of no substitutions.” From the vantage point of this stage of psychosocial development, the individual relinquishes the wish that important people in his life had been different and is able to love in a more meaningful way—one that reflects an acceptance of responsibility for one’s own life. The individual in possession of the virtue of *wisdom* and a sense of integrity has room to tolerate the proximity of death and to achieve what Erikson termed in *Identity: Youth and Crisis* a “detached yet active concern with life.” Despair may be present, but it does not predominate.

Erikson underlined the social context for this final stage of growth. In *Childhood and Society*, he wrote, “The style of integrity developed by his culture or civilization thus becomes the “patrimony” of his soul…In such final consolidation, death loses its sting.”

When the attempt to attain integrity has failed, the individual may become deeply disgusted with the external world and contemptuous of persons as well as institutions. Erikson wrote in *Childhood and Society* that such disgust masks a fear of death and a sense of despair that “time is now short, too short for the attempt to start another life and to try out alternate roads to integrity.” In thinking about the relationship between adult integrity and infantile trust, Erikson observed, “[H]ealthy children will not fear life if their elders have integrity enough not to fear death.”

BECOMING A BIOGRAPHER OF YOUTH

In 1949, as *Childhood and Society* was going to press, Erikson’s stand on the University of California loyalty oath had rendered his position at the university untenable. News of his availability spread east, and other institutions vied to capitalize on Berkeley’s mistake. Erikson was well known at the Menninger Clinic in Topeka, Kansas, where he had lectured, so when Robert Knight left Menninger to become the director of the Austen Riggs Center in Stockbridge, Massachusetts, he took Erikson with him. The Austen Riggs Center was devoted to psychoanalytic research and to the treatment of severely disturbed adolescents and young adults—Erikson had been happily adopted once again.

Erikson stayed at Austen Riggs from 1950 to 1960, from age 48 to 58, when he returned to the faculty of Harvard University. During those years, Erikson completed the transition to biographer, forming more fully the approach foreshadowed in his incomplete essays on Hitler, Gorky, and George Bernard Shaw. The transition was not easy; Coles describes that time as constituting a second "identity crisis." Erikson had been a clinician and a theorist of youth; he was now old enough to enact the program promised in *Childhood and Society* and to become a biographer and theorist of the whole life cycle. However, before he wrote *Young Man Luther*, he once again returned to Freud and the origins of his own professional identity. The way to the future was through the past.

In his early 50s, Erikson wrote three biographical essays on Freud: "The Dream Specimen of Psychoanalysis," "Freud's 'The Origins of Psychoanalysis,'" and "The First Psychoanalyst." These essays all concern themselves with the crisis Freud suffered during his own midlife transition as he struggled to leave neuropathology and to define a new professional identity as a psychoanalyst. Strengthened by that re-examination of Freud's successful transition (and perhaps having reassured himself of Freud's blessing), Erikson began writing one of the genuine masterworks of psychoanalysis, *Young Man Luther.*

As always, Erikson's clinical work and observations enriched his theorizing about the life cycle. He acknowledged his debt to Austen Riggs and the Western Psychiatric Institute at the University of Pittsburgh in his preface to *Young Man Luther*. His work there had allowed him to "study the afflictions of young patients as variations on one theme, namely, a life crisis, aggravated in patients, yet in some form normal for all youth. I could identify those acute life tasks that would bring young people to a state of tension in which some would become patients."

However, Erikson would neither make Martin Luther a patient nor reify the psychopathology of the young people who were his patients. Instead, he asserted that comparisons between Martin Luther and his patients were “not restricted to psychiatric diagnosis…but…oriented toward those moments when young patients, like young beings anywhere, prove resourceful and insightful beyond all professional and personal expectation. We will concentrate on the powers of recovery inherent in the young ego.”

Working with the young patients at Austen Riggs helped Erikson hone his understanding of identity formation. As he later put it in “The Problem of Ego Identity,” “[I]dentity…is gradually established by successive ego syntheses and resyntheses throughout childhood. It is a configuration gradually integrating *constitutional givens, idiosyncratic libidinal needs, favored capacities, significant identifications, effective defenses, successful sublimations, and consistent roles*.”

These “ego syntheses and resyntheses” were about to receive great attention as Erikson attempted his biography of Luther. But to do that work, Erikson, the biographer, had to move further away from his own professional origins as a child psychoanalyst. He had criticized psychoanalysis for its contention that something has been explained by finding an analogy to its earliest manifestations. Yet his own conception of the life cycle, with most of its stages occurring within the preadult era, left him still heavily rooted in childhood. In his study of Luther, Erikson was trying to understand a life, not just a personality; from that perspective, childhood and adolescence had to be introductory rather than the story itself.

Seeing that a child-centered view was not adequate for the tasks of biography, without either formally changing the imbalance in his theory or neglecting childhood determinants, Erikson deftly devoted the greatest attention to an explication of development in the adult years. So successful was he in interrelating Luther’s adult problems with those of his early development that *Young Man Luther* provided the seminal inspiration for the next generation of life-cycle investigators. Some of those investigators, such as Daniel Levinson, formally redressed in their own theories the originologic bias vestigial in Erikson's theories.

In Luther's early or mid-20s, according to some of his contemporaries, the young monk had fallen to the floor of the choir of his monastery in a fit and shouted, "ich bin's nit!" or "Non Sum!" "It isn't me," or "I am not." Erikson's task as a psychological biographer was to explain how Luther got from the identity crisis of his 20s to nailing his 95 theses to the door of the church in Wittenberg at the age of 32. Yet *Young Man Luther* is also, as it is subtitled, "a study in psychoanalysis and history." Since it is not a psychoanalysis of history, some nonreductionist connecting concepts between the individual and the collectivity were needed. One of Erikson's connecting concepts was *ideology*, which he defined as "an unconscious tendency underlying religious and scientific as well as political thought: the tendency at a given time to make facts amenable to ideas, and ideas to facts, in order to create a world image convincing enough to support the collective and the individual sense of identity."

Luther had personal and psychiatric problems, but in his defiance of Catholic orthodoxy, he created a new religious ideology that not only supported his own identity but the identities of emerging generations of Europeans as well. "Ich bin's nit" expressed the young monk's identity crisis as he found himself "fatally over-committed" to what he was not, a young man authentically embarking on a future as an orthodox Catholic priest. "In some periods of history," Erikson asserted, "and in some phases of his life cycle, man needs…a new ideologic orientation as surely and as sorely as he must have air and food." As such a man, Luther commanded Erikson's "sympathy and empathy" as he "faced the problems of human existence in the most forward terms of his era."

The reception of the book cemented Erikson's international reputation, initially won by *Childhood and Society*, as a leader in original, humane thought about psychological development. In no other biography have the dynamics of individual conflict and development been so seamlessly interwoven with those of society and history. With *Young Man Luther*, Erikson in his late 50s ascended from the first rank of developmental psychoanalysts to that of a cultural seer.

BECOMING A BIOGRAPHER OF MIDDLE AGE

As Erikson moved through the transition from middle to old age, he abandoned clinical work with patients to turn his attention fully to the study of the life cycle and the survival of the species. His emphasis, emergent in *Young Man Luther* and implicit in all his earlier work, was on the ego virtues that permit a person to live in a constructively critical relationship with the social institutions of his time. Earlier, Erikson had described the stages in the development of the ego across the life cycle. In 1964 he wrote more about the ego virtues in *Insight and Responsibility*. He was getting ready amidst the American moral and political crises of the mid-1960s to embark on his last major work, a study of the great moral leader Mohandas Gandhi. In the end, he dedicated it, *in memoriam*, to Martin Luther King, Jr.

Young Man Luther had been a story of personal choice and historical change wrought by a young man establishing and revising an initial identity and life structure as an adult. In his mid-60s, Erikson had gained sufficient perspective on the life cycle to analyze a case of decisive crisis and change in a man who was going through a midlife transition and solving it by creating a vital life structure for middle age. Gandhi did that, as he himself put it, by realizing “his vocation in life,” leading *satyagraha* truth-force campaigns in his nation, in his family, and in his own soul. Gandhi was entering the stage of the life cycle in which Erikson’s ego polarity of *generativity versus stagnation* becomes active; from his developmental work on this polarity, Gandhi was learning the ego virtue of *care*.

Erikson saw that Gandhi, like Luther, was both a leader and a “religious actualist.” As a leader, he succeeded in articulating inner concerns in a way that struck a collective cord. Self-rule and home rule were inextricably combined in Gandhi’s program. He would have endorsed in his own terms, Erikson believed, Luther’s assertions, "Christ comes today; God's way is what makes us move; we must always be reborn, renewed, regenerated; to do enough means nothing else than always to begin again."

Gandhi had begun and successfully completed his political novitiate as an Oxford-trained barrister in South Africa during his 20s and 30s. By his late 30s, he had developed passive resistance as a political strategy to defeat the Black Act, a law that required all Indians in the Transvaal to register with the government and to carry identification papers on their persons. At 40, Gandhi staked out his claim for leadership in his Indian Home Rule Manifesto, and during the next few years of his midlife transition, he developed the device of the Satyagraha campaign, culminating in the Great March of striking mineworkers in South Africa when he was 44.

When Gandhi returned to India in 1915 at age 45, he did so, Erikson wrote, "like a man who knew the nature and the extent of India's calamity and that of his own fundamental mission." As a mature, middle-aged man, he was, in Erikson's view, a person who has determined what "he does and does not *care for*" as well as "what he *will* and *can* take *care of*. He takes as his baseline what he irreducibly is and reaches out for what only he can, and therefore, *must* do."

What Gandhi had to do was lead a labor strike at the mills in Ahmedabad and the next year, at age 50, a national strike for independence from Britain. Erikson contended that Gandhi's emergence at that time as the "father of his country" underlined "the fact that the middle span of life is under the dominance of the universal human need and strength which I have come to subsume under the term *generativity*." Erikson refers here not merely to the generativity that a parent at 20 or 30 may feel for a child. Erikson's concept of generativity is often mistakenly taken to apply to the child-rearing work of young adults. Instead, he is describing the protective concern for the generations and their retarding-facilitating social institutions, whose leadership is the essential responsibility of the middle-aged individual.

In the early 1970s, Erikson returned to the San Francisco Bay Area to live and begin again. The Eriksons relocated to Cambridge, Massachusetts, some 10 years later. In 1981 the Eriksons and psychologist Helen Kivnick undertook an intensive study of 29 octogenarian parents of children whose lives had been meticulously scrutinized in the Guidance Study begun in 1928 at the Institute of Human Development at the University of California at Berkeley. The results of their research were published in *Vital Involvement in Old Age.*

In this work, Erikson, Erikson, and Kivnick revisit the eight stages of psychosocial development and their attendant virtues or strengths. They emphasize the opportunities that exist at the end of life to integrate “maturing forms of hope, will, purpose, competence, fidelity, love, and care into a comprehensive sense of wisdom.” A successful integration results from a final reworking of the earlier polarities of basic trust versus mistrust; autonomy versus shame and doubt; initiative versus guilt; industry versus inferiority; identity versus confusion; intimacy versus isolation; generativity versus self-absorption; and integrity versus despair. The authors use the term *vital involvement* to identify an involvement with the environment characterized by “actuality” and “mutuality.” The elder who is able to maintain a vital involvement with his world—no matter how small its scope—is better able to tolerate the depredations of aging: the loss of sensory acuity, physical mobility, and stamina; changes in intimate relationships as children marry and move away, and friends and partners sicken and die; the loss of social status, financial security, and sources of pride in professional competence; and the realization that relatively little time remains.

Developed in infancy, hope is the basis upon which the senescent individual attempts to integrate faith in the universe and the relative predictability of its laws, with a realistic mistrust about that which is unpredictable and unreliable. The elder may experience a need to affirm a basic faith—in religion or nature—and to understand more deeply his own place in a generational progression.

Late in life, the individual must struggle to redefine a sphere of autonomy even as he becomes increasingly dependent upon others. Living independently, residing in familiar surroundings, and following established routines can be crucial symbols of autonomy and the means by which signs of impairment (e.g., a walker) may be accepted without undue shame. An individual’s lifelong experience of autonomy influences his or her capacity to adjust to restrictions upon freedom. The ability to allow others to help supports the generative impulses of younger adults and reinforces for both the sense of a generational cycle.

The polarities of initiative versus guilt, first evident in the youngster’s exploration and play and subsequently in adult work and recreation, must be balanced anew in old age. With the passage of time, the person faces a diminution of opportunities for the exercise of initiative and must regret those instances where he or she failed to act decisively or with sufficient concern for others. Successful adaptation to the limitations of old age must draw upon the strength of purposefulness developed in childhood and tempered by maturity.

With the natural decline in physical strength and sensory acuity, the senescent individual increasingly depends upon a lifelong sense of effectiveness. External rewards are usually not as numerous as they were in young adulthood or middle age. The feeling of competence and mastery must be sustained internally. An old person may cope with growing physical limitations by changing the criteria for a sense of accomplishment, pursuing new hobbies or modifying old ones, developing new avenues of involvement with former professions, or living vicariously through the accomplishments of children and grandchildren.

Old age provides the individual with the chance to review a lifetime of beliefs, to come to terms with choices made and opportunities lost, and to make and act upon any final commitments, which now most clearly reflect what Erikson termed “the ‘I’ in the totality of life.” In striking a final balance between identity and identity confusion, the elder concerns himself with both an external image and a personal image. The external image is the way the individual may expect to be remembered; the personal image reconciles the sense of who he or she has been with an evolving sense of whom he or she may yet become. Identification with the successes of younger generations and satisfaction in one’s own part in the transmission of values helps the senescent to stave off feelings of insignificance or obsolescence in the face of technologic and sociologic changes.

The sense of love that can emerge at the end of life is one that is built upon a foundation of love, expressed and unexpressed, throughout the life cycle. A new balance may be forged between intimacy and isolation in the face of decline. Intimate relationships may be recast in more positive terms; reminiscences provide the elder with the opportunity to draw upon and reintegrate earlier experiences of tenderness and sexuality.

In order to reconcile feelings of integrity and despair, the individual must make peace with previous choices, acknowledging what was gained and lost and accepting that it is too late to change. Relationships and values may be reviewed with more room to consider alternate points of view. In balancing the dispositions toward generativity and stagnation, the older adult may be able to build upon years of experiences as a parent, worker, and creatively productive person from middle age, which themselves were shaped by childhood experiences of caring or its absence. Grandparenthood offers another chance at generativity and a way to undertake it more robustly and less ambivalently. Erikson coined the term “grand-generativity” to distinguish the generativity of old age in which the person, freed from the direct responsibility for family, institution, and community, is better able to incorporate care for the present with concern for the future. Such concern may be expressed through advice or financial assistance to family members or, on a larger scale, through a commitment to religious beliefs, political involvement, or community action. Grand-generativity contributes to a feeling of immortality; in caring for the younger generation and allowing oneself to be taken care of, the senescent secures a personal place in a generational history.

The “joint reflections of old age” contained in the book are an effort to elucidate the psychosocial process of vital involvement and to extend Erikson’s observations to the outer limits of the life cycle. The book also brought Erikson back full circle to his professional beginnings—having made almost 40 years earlier, play observations on the Guidance Study children.

ERIKSON AND CULTURE

Erikson has been a pioneer in alerting psychotherapists, especially of the psychodynamic persuasion, to the role of culture in constructing the psyche. This understanding is a consequence of his recommendation to therapists (and psychobiographers) that the meaning of an “event” in a patient’s life needs to be understood in a fourfold configuration of meaning: the meaning of the event at the stage of the patient’s life cycle, its meaning in the patient’s life history, its meaning in the patient’s culture, and its meaning in the history of the culture. He alerts us that the terroir of psychotherapy for more than a century has been and continues to be Western, with its centers located in North America, United Kingdom, and North and Central Europe. It contains many Western cultural ideas and ideals that permeate psychotherapeutic theories and practices. Shared by therapist and patient alike, pervading the therapeutic space in which the two are functioning, fundamental ideas about human relationships, family, marriage, male and female, on what constitutes psychological maturity and so on which are essentially cultural in origin often remain unexamined and regarded as universally valid. Examples include how to behave toward superiors and subordinates in organizations, the kinds of food conducive to health and vitality, and the family’s web of duties and obligations. All these are as much influenced by the cultural part of the mind as are ideas concerning the proper relationship between the sexes or one’s relationship to the Divine.

However, Erikson is not a complete relativist but introduces the notion of psychic universals lying on a continuum rather than occupying a fixed point. To illustrate, take the example of the universal psychic phenomenon of separation-individuation. Between a minimum of separation-individuation needed to function as an adult in a particular society and a maximum that cuts off all ties with family and community, there is a range of positions. Psychoanalysis should recognize the positions as part of being human rather than closely identifying with any particular position on the continuum as the only one that is “healthy and mature.”

Take another example that is fundamental for our psychic makeup and how we view the world: gender.

Unconsciously absorbed since childhood, our culture determines what it means to be, look, think, or behave like a man or a woman. This concept becomes more apparent if one thinks of Greek or Roman sculpture with their hard, muscled men's bodies and chests without any fat (familiar to Europeans and Americans from childhood visits to the museum or reproduced in countless visual mediums) and compare it with the sculpted representations of Hindu gods or the Buddha, where the bodies are softer, suppler, and, in their hint of breasts, nearer to the female form. Between a minimum of sexual differentiation needed to function heterosexually with a modicum of pleasure and a maximum that cuts off any sense of empathy and emotional contact with the other sex experienced as a different species altogether, there is a whole range of positions, each occupied by a culture that assumes it is universal and insists on calling it the only one that is mature and healthy.

In a postcolonial, globalizing world, steering between the extreme universalist and relativistic positions, Erikson's contributions to the role of culture in the formation of the psyche can be invaluable in reconsidering many psychotherapeutic assumptions and models.

APPLICATION OF ERIKSON'S CONCEPTS TO CLINICAL WORK

Erikson's view of individual experience as inexorably embedded in developmental, familial, societal, and historical contexts crucially shaped his ideas concerning mental illness and psychiatric treatment. As noted previously, Erikson was reluctant to pathologize behavior or to rush to judgment about the meaning of any given symptom. He asserted, "Perhaps there are certain stages in the life cycle when even seemingly malignant disturbances are more profitably treated as *aggravated life crises* rather than as diseases subject to routine psychiatric diagnosis."

To understand the meaning of such “disturbances,” Erikson drew upon his training in classical drive theory (with its concepts of the unconscious; id, ego, and superego conflicts; repression, regression, and repetition compulsion) but emphasized the adaptive and synthesizing capacities of the ego in brokering relations between internal drives and external reality. When Erikson listened to his patients, he thought about their symptoms not simply as compromise formations between libidinal drive and superego prohibition but also as expressions of arrested or derailed psychosocial development. Toward this end, Erikson examined many aspects of his patient’s current life (e.g., skills, talents, and aspirations; religious and political commitments; social and intimate relationships; roles as worker, partner/spouse, and parent) and located them within a larger social, cultural, and historical context. His focus was on the individual’s entire life cycle, the development of specific virtues, and those factors which facilitated or retarded psychosocial growth. As Shapiro and Fromm point out:

> He paid attention to the whole life context of any immediate situation. He asked a number of questions. What is the immediate stimulus for the patient’s reaction? What is the acute life conflict, the current developmental stage, the issues that are manifest? In what developmental context did the patient’s reaction first occur? Is it now manifest in the relationship to the therapist, in a repetitive conflict, in a characteristic way that the individual solved earlier developmental struggles? In what social context is the individual embedded, what roles are available?…What defenses does the individual use? What are the individual’s deepest psychological investments?

Erikson conceived of these psychological investments in broad terms. He used concepts of attachment, separation, and mutuality (defined as “a relationship in which partners depend on each other for the development of their respective strengths”) to illustrate ways in which the individual is essentially a social being. He recognized that psychological crises often occurred at times of developmental separation or individuation and that fears regarding dependency and abandonment were frequently the catalyst for such crises. Though Erikson acknowledged the influence of early childhood conflict or trauma on an individual’s progress in negotiating developmental tasks in adolescence and adulthood, he did not think that the issues being played out were simply reiterations of infantile conflict.

Erikson's belief that psychosocial development continues throughout the life cycle led him to caution psychotherapists against a theoretical reductionism in which symptomatic behavior was attributed to a traumatic past and in which insufficient attention was paid to contemporary conflicts in the ego's relations with the world. He was particularly critical of the "originology" in psychoanalytic thought. "I mean by it a habit of thinking which reduces every human situation to an analogy with an earlier one, and most of all to that earliest, simplest and most infantile precursor which is assumed to be its "origin."" Erikson was wary of other theoretical blindspots within his profession. As Coles noted, "…psychiatric and psychoanalytic concepts emerge from and become very much suited to a particular society—which exerts its (moral and puritanical) influence on everything, even the most complicated, rarefied, and "objective" of abstractions."

In his article, *The Nature of Clinical Evidence*, Erikson outlined the relationship between patient and therapist. He described the "clinical core of medical work" as the

> "encounter of two people, one in need of help, the other in the possession of professional methods. Their *contract* is a therapeutic one: in exchange for a fee, and for information revealed in confidence, the therapist promises to act for the benefit of the individual patient, within the ethos of the profession." (Excerpted from Erikson EH. The nature of clinical evidence. *Daedalus*. 1958;87(4):65–87. © 1958 American Academy of Arts & Sciences, with permission. http://www.jstor.org/stable/20026463.)

Though he acknowledged certain parallels between medical and psychiatric consultations (e.g., the patient presents with a list of symptoms, the physician/psychiatrist asks questions designed to elucidate the problems, arrives at a tentative diagnosis, and proposes a treatment plan), Erikson departed from a traditional medical model in important ways. Though according to the therapist the respect and authority appropriate to his special training and expertise, Erikson envisioned the patient as an active participant in a collaborative endeavor. Within the limitations imposed by the treatment setting (e.g., inpatient vs. outpatient) and degree of psychological impairment, the patient was expected to take significant responsibility for his treatment. Erikson noted a natural transformation that occurs in the patient.

> ““Under observation,” he becomes self-observant. As a patient he is inclined, and as a client often encouraged, to historicize his own position by thinking back to the onset of the disturbance, and to ponder what world order (magic, scientific, ethical) was violated and must be restored before his self-regulation can be reassumed.” (Excerpt from Erikson EH. *Insight and Responsibility: Lectures on the Ethical Implications of Psychoanalytic Insight*. W.W. Norton & Company; 1964.)

How does the patient participate in his own “cure”? He is invited to free associate. Erikson noted, “(W)e consider a patient’s “associations” our best leads to the meaning of an as yet obscure item brought up in a clinical encounter, whether it is a strong affect, a stubborn memory, an intensive or recurring dream, or a transitory symptom.” The associations consist of whatever thoughts, feelings, sensations, or images occur to the patient during and after the mention of that item. Erikson continued,

> “Except in cases of stark disorganization of thought, we can assume that what we call the synthesizing function of the ego will tend to associate what “belongs together,” be the associated items ever so remote in history, separate in space, and contradictory in logical terms…It is…this basic synthesizing trend in clinical material itself that permits the clinician to observe with “free-floating attention,” to refrain from undue interference, and to expect sooner or later a confluence of the patient’s search for curative clarification and his own endeavor to recognize and to name what is most relevant, that is, to give an *interpretation*.” (Excerpt from Erikson EH. *Insight and Responsibility: Lectures on the Ethical Implications of Psychoanalytic Insight*. W.W. Norton & Company; 1964.)

Free-floating attention “turns inward to the observer’s ruminations even as it attends the patient’s “free associations’ and which, far from focusing on any item too intentionally, rather waits to be impressed by recurring themes.” A message represented by a patient’s memory, dream image, or symptom can be overdetermined —“a condensed code transmitting a number of other messages, from other life situations, seemingly removed from the therapy. This we call “transference.””

As for the therapist’s role in working with the patient’s free associations, Erikson averred,

> “I have to assume that the patient is (to varying degrees) unconscious of the meaning which I discern in his communications, and that I am helping him by making fully conscious what may be totally repressed, barely conscious, or simply cut off from communication…I take for granted an effective wish on his part (with my help) to see, feel and speak more clearly.” (Excerpt from Erikson EH. *Insight and Responsibility: Lectures on the Ethical Implications of Psychoanalytic Insight*. W.W. Norton & Company; 1964.)

Erikson felt that a “therapeutic” interpretation was one which was clear and concise, yet presented the patient with a “unitary” theme, “…a theme common at the same time to a dominant trend in the patient’s relation to the therapist, to a significant portion of his symptomatology, to an important conflict of his childhood, and to corresponding facets of his work and love life.” How would the therapist determine if an interpretation was “right”? Erikson believed, “(T)he proof lies in the way in which the communication between therapist and patient “keeps moving,” leading to new and surprising insights and to the patient’s greater assumption of responsibility for himself.”

Erikson acknowledged that there was a great deal of subjectivity involved in both the patient’s complaints and the therapist’s interpretations. He cautioned, “…even while facing most intimate and emotional matters, (the therapist) must maintain intellectual inner contract with his conceptual models, however crude they may be.” He must evince “…a *specific self-awareness* in the very act of perceiving his patient’s actions and reactions.” Erikson wrote, “…there is a core of *disciplined subjectivity* in clinical work–and this both on the side of the therapist and of the patient—which it is neither desirable nor possible to replace altogether with seemingly more objective methods.” The therapist was advised to undergo a personal psychoanalysis in order to reduce or eliminate potential blindspots which might compromise this disciplined subjectivity.

Erikson noted “the power of the transference, that is, the patient’s transfer to me of significant problems in his past dealings with the central people in his life” but also conceived of what we would now call a “real” relationship between himself and his patient. In *The Nature of Clinical Evidence*, Erikson provided a lengthy clinical vignette illustrating dream analysis and his use of his emotional reactions in his interpretations to his patient. Because he did not conceive of clinical work as primarily oriented toward the development and interpretation of the transference, Erikson accorded himself and other clinicians the freedom to cultivate “real” aspects of the relationship with the patient. He felt it important and necessary at times that the therapist educate and guide his patient. Erikson averred, “…only by playing my role as a new person in his present stage of life can I clarify the inappropriateness of his transferences from the past.”

A psychotherapy informed by Eriksonian principles is essentially optimistic in nature. For it takes as a given the notion that psychosocial development is universal and continues throughout the life cycle. Natural forces propel the individual forward. Psychiatric symptoms are often best understood as manifestations of aggravated developmental crises—development that has been blocked or derailed. In such cases, the therapeutic task is to release or redirect those built-in forces for growth and adaptation.

As Erikson noted in *Identity: Youth and Crisis*, the object of psychotherapy is not to head off future conflict but to assist the patient in emerging from each crisis “with an increased sense of inner unity, with an increase of good judgment, and an increase in the capacity “to do well” according to his own standards and to the standards of those who are significant to him.”

FURTHER READINGS

Arnett JJ. Identity development from adolescence to emerging adulthood: what we know and (especially) don’t know. In: McLean KC, Syed M, eds. *The Oxford Handbook of Identity Development*. Oxford University Press; 2015:53–64.

*Coles R, Erikson EH. *The Growth of His Work*. Little, Brown and Company; 1970.

*Erikson EH. *Childhood and Society*. Norton; 1950.

Erikson EH. Freud's "The Origins of Psychoanalysis." *Int J Psychoanal*. 1955;36:1.
*Erikson EH. *Gandhi's Truth*. Norton; 1969.
Erikson EH. Hitler's imagery and German youth. *Psychiatry*. 1942;5:475–493.
Erikson EH. *Identity: Youth and Crisis*. Norton; 1968.
Erikson EH. *Insight and Responsibility*. Norton; 1964.
Erikson EH. Observations on Sioux education. *J Psychol*. 1939;7:101–156.
Erikson EH. The dream specimen of psychoanalysis. *J Am Psychoanal Assoc*. 1954;2:5.
Erikson EH. The first psychoanalyst. *Yale Rev*. 1956;46:40.
Erikson EH. The problem of ego identity. *Psychol Issues*. 1959;1:379.
*Erikson EH. *Young Man Luther*. Norton; 1962.
*Erikson EH, Erikson J, Kivnick H. *Vital Involvement in Old Age*. Norton; 1986.
Evans R. *Dialogue with Erik Erikson*. Harper and Row; 1967.
Freud A. *The Ego and Mechanisms of Defense*. International Universities Press; 1966.
Friedman LJ. Erik Erikson on identity, generativity, and pseudospeciation: a biographer's perspective. *Psychoanal Hist*. 2001;3:179.
Friedman LJ. *Identity's Architect: A Biography of Erik Erikson*. Scribner; 1999.
Levinson D, Darrow C, Klein E, Levinson M, McKee B. *The Seasons of a Man's Life*. Knopf; 1978.
McAdams DP, Zapata-Gietl C. Three strands of identity development across the human life course: reading Erik Erikson in full. In: McLean KC, Syed M, eds. *The Oxford Handbook of Identity Development*. Oxford University Press; 2015:81–96.
Roth MS. Psychoanalysis and history. *Psychoanal Psychol.* 2016;33S:S19.
Schein S, ed. *Erik Erikson: A Way of Looking at Things*. Norton; 1987.
Shapiro ER, Fromm MG. Eriksonian clinical theory and psychiatric treatment. In: Sadock BJ, Sadock VA, eds. *Comprehensive Textbook of Psychiatry*. Vol 7. Lippincott Williams & Wilkins; 2000:2200–2207.
Slater C. Generativity versus stagnation: an elaboration of Erikson's adult stage of human development. *J Adult Dev*. 2003;10:53.
Stevens R, Erikson EH. *Explorer of Identity and the Life Cycle*. Palgrave Macmillan; 2008.
Syed M, McLean KC. The future of identity development research: reflections, tensions, and challenges. In: McLean KC, Syed M, eds. *The Oxford Handbook of Identity Development*. Oxford University Press; 2015:562–574.
Tummala-Narra P. Cultural competence as a core emphasis of psychoanalytic psychotherapy. *Psychoanal Psychol*. 2015;32:275.
Wallerstein R, Goldberger L, eds. *Ideas and Identities: The Life and Work of Erik Erikson*. International Universities Press; 1998.

6.3 Other Psychodynamic Schools

KRISTINE L. MCCASLIN, PSY.D., AND ADAM M. BRENNER, M.D.

The professionals discussed in this chapter—Adolf Meyer, Alfred Adler, Carl Gustav Jung, Franz Alexander, Wilhelm Reich, Karen Horney, Erich Fromm, Harry Stack Sullivan, Eric Berne, and Nancy Chodorow—expanded psychiatric thought and practice throughout the 20th century and into the 21st. With knowledge of the neurosciences initially primitive, the ethos of psychiatry was to use clinical observation to search for an overarching theory that would encompass not only psychopathology but all of human behavior. None of these individuals doubted that the mind has a biologic basis, but they all were aware of Sigmund Freud's woeful failure in his *Project for a Scientific Psychology* to develop an understanding of how experience, thoughts, and feelings are transduced to and from material states. Because of this, they were forced to stand back from the physical organism and view humans strictly as psychological entities. What, then, is the relevance of these theories to 21st-century psychiatry? Are these theories of merely historical interest, or do they hold some wisdom for current clinical practice? Until 30 years ago, psychiatrists spent great time and energy comparing and contrasting theorists and arguing the pros and cons of various formulations. Within psychoanalysis and other "schools" of psychotherapy, such theoretical discussions still occur. Now, however, the primary contribution of these theorists is in their broad brush strokes; for example, should one embrace the frustration model of classical psychoanalysis, the actively empathic model of self-psychology, or the interactive model of interpersonal psychiatry?

There is also a historical shift that has occurred. In the past, as a clinician contemplated some new insight into the workings of patients' inner lives, it was often seen as necessary, not only to build a new theory but also to create a new "school" of therapy. This may have been an outgrowth of Freud's personality, which tended to focus on his current line of observations and thought, thus forcing others (e.g., Adler and Jung) to move outside the inner circle to pursue their different lines of thinking. These days, psychoanalysis is a far more open intellectual enterprise, embracing a variety of different models and inputs.

These theories remain relevant in several important ways. First, regardless of what is learned about neurobiology, psychiatry is and will remain concerned with distress in behavior, thoughts, emotions, and relationships. Inevitably, these will be communicated through words and nonverbal signals. Psychiatrists will always need some organizing templates at the behavioral and verbal communicational levels to help order and interpret their observations. A sad, pensive, tearful expression does not automatically translate into a serotonin or norepinephrine deficiency or some other structural abnormality. It must first be translated into the psychological constructs of loss, depression, grief, and so forth. Second, these theorists spent enormous time listening to patients, and all were careful observers. As such, their observations remain helpful in defining key dimensions for psychiatrists to consider in understanding the inner experiences patients attempt to communicate. Finally, all of these theorists were well versed in a rigorous approach to theory construction. Every human being has a “theory of personality,” acknowledged or otherwise. This theory, often intuitive and unrigorous, based on a mixture of reading, personal values, and unsystematic observation, influences much of the way people interact with one another. In the clinical situation, when a psychiatrist mixes supportive management and medication, they are using a personal theory of personality. The advantage of knowing the theorists in this chapter is that models of intellectual rigor can be brought to personal intuitions.

The following case is used as a basis for formulation according to each theorist to illustrate the different understandings and therapeutic approaches of each theorist.

Mr. A was a 26-year-old White man who had a history of bipolar I disorder. He was brought in for treatment after not completing the last required course for his advanced degree and being arrested for disturbing the peace. He had consistently lied to his family about where he stood with his coursework and about having skipped an examination that would have qualified him to use his professional degree. He had also not told them that he had been using marijuana almost daily for a number of years and occasionally used hallucinogens. His arrest for disorderly conduct was for swimming naked in an apartment complex in the middle of the night while under the influence of hallucinogens.

Mr. A's use of marijuana began early in college but became daily during graduate school. He was diagnosed as having bipolar I disorder early in his senior year at college after a clear episode of mania. His mood disorder was well controlled on lithium (Eskalith). During graduate school, he was episodically adherent to taking medications, preferring to try to maintain a state of hypomania. He saw a psychiatrist every 3 to 6 months for medication management. During his 4 years in graduate school, he had two clear episodes of depression and began taking sertraline (Zoloft), 100 mg per day, with questionable benefit. Mr. A believed that he could be a great writer. He spent most of his time reading and trying to write. He dreamed of going to New York or Paris and becoming part of a new wave of novelists. This aspiration and his marijuana abuse predated his development of bipolar I disorder. He attended class episodically, nonetheless performing adequately. His last class had no final examination but required a paper. He planned to write this paper in the form of a play involving a dialogue between two thinkers from different times and cultures. His professor was very excited about this idea, but Mr. A kept postponing the task until he was forced to extend his schooling by a year. His other major interest during this time involved growing and photographing flowers.

Mr. A was born and raised in a large city. His father had been very successful in commercial real estate. His mother, after raising her children, used the substantial real estate holdings she inherited from her father to set up a business to manage them. Most of the money was placed in a trust for the patient and his siblings. His mother had total financial control of the trusts and doled out the proceeds to the children as they needed them. There was no family history of any psychiatric disorders.

The patient described his mother as very loving and caring but to the point of being intrusive and controlling. For example, the mother arranged the initial treatment but then was angry that the psychiatrist had not called her regularly to report on her adult son's progress. She was also critical of various aspects of the treatment as reported to her by her son. The patient's two older siblings had attended prestigious colleges and graduate schools but had returned home to work in the mother's real estate management company. The 30-year-old sister was living in the parents' home. The 35-year-old brother had lived at home for a time but then moved out to a location a few blocks away. There was a younger brother, still in college, who also smoked marijuana excessively. He tried to minimize the patient's problems to the family and tried to protect the patient, who desperately had not wanted to return home. Of note is that none of the children were married, although the two older ones had each had a couple of serious relationships.

The children seemed to regard the mother with affectionate amusement and bemusement. The father was seen as a very caring but undemonstrative man who put much energy into keeping the mother from becoming too upset and encouraged the children to do the same. The children often wanted to provoke the mother for her judgmental, detail-oriented intrusiveness. The father discouraged them but occasionally found their provocations amusing.

The family viewed itself as very close, with strong values oriented toward community service and family loyalty. The family belonged to a religious community but expressed their involvement primarily in social service and social action volunteer work, accompanied by very generous financial contributions.

The patient had been a very successful debater in high school and recalled his development as very positive but provided few details. He tended to place himself in the role of the outsider, an observer of humanity, which he saw as consonant with the role of a writer. He was proud to have bipolar I disorder and tried to regulate his medications so that he would be hypomanic much of the time, seeing this as enhancing his creativity. He viewed his use of marijuana in the same vein. One of the most distressing aspects to him of his depressive episodes was that marijuana no longer created a feeling of well-being but made him feel worse. His current depressive episode involved no neurovegetative symptoms. Rather, he presented as flat, numb, apathetic, ashamed, and anhedonic. He was particularly ashamed of being back in his hometown and of living with his parents.

The patient ostensibly understood and accepted his illness well and had read much about it. However, the family had responded to the information "with proper treatment, bipolars can live normal lives," meaning that the information should be kept secret so that he should be treated normally. Mr. A, on the other hand, was very open with friends at graduate school about his illness and his pride in it, and the creativity he associated with it.

The patient had two longstanding recurrent dreams. One involved him flying. The narrative line varied, but the flying theme recurred. Often, he had other magical powers in his dreams, such as the ability to heal, not be killed by bullets, to save the world or some group of people from mortal danger, and so on. The other recurrent dream was of a hotel lobby. These dreams regularly began with him entering a hotel lobby to meet a group of people, accompanied by a feeling of dread (Table 6.3–1).

ADOLF MEYER

Adolf Meyer (1866–1950) immigrated to the United States after training as a neuropathologist in Switzerland (Fig. 6.3–1). Uninterested in metapsychology, he espoused a commonsense psychobiological approach to understanding mental disorders, emphasizing the interrelationship of symptoms and individual psychological and biologic functioning. His approach to the study of personality was biographical; he attempted to bring psychiatric patients and their treatment out of isolated state hospitals and into communities and was also a strong advocate of social action for mental health. He began his career as a clinically oriented state hospital pathologist, becoming the second head of the Pathological Institute (later, the New York Psychiatric Institute), and initiating its affiliation with the Columbia College of Physicians and Surgeons.

He later became the president of the American Psychiatric Association and had a 32-year tenure as chair of psychiatry at Johns Hopkins. His major social contribution was helping to found the National Committee on Mental Hygiene.

Psychobiology

Despite his background as a neuropathologist, which was common among European psychiatrists of his generation, Meyer strongly opposed the Kraepelinian view of mental illness as having a predetermined course based on phenomenologically identified syndromes. Instead, he believed that individuals' habitual reaction patterns made them more susceptible to specific types of breakdown. Meyer used ***biographical study*** and strongly encouraged psychiatrists to take a thorough, detailed life history of the patient to understand each individual's reaction patterns. He saw development as lifelong, thus viewing the data from adolescence onward as equal in importance compared with data from childhood in understanding the patient's narrative.

Table 6.3–1.
Techniques and Terminology

Meyer	Biographical study approach
	Distributive analysis and distributive synthesis
	Focus on patient assets
	Habit training
Adler	Individual psychology
	Inferiority toward mastery; “inferiority complex”
	Birth order
	Dynamism (future directed, goal oriented)
	Removing blocks to living productively in the real world
	Reframing
	Paradoxical communication
Jung	Complexes
	Archetypes
	Projective/introjective processes
	Persona, shadow, anima, and animus
	Typology (introversion–extroversion, sensation–intuition, thinking–feeling)
Alexander	Psychosomatic medicine
	Corrective emotional experience
	Controversial attempts to shorten/manipulate psychodynamic therapy
Reich	Ego psychology
	Character armor and personality types
	Bioenergetics
Horney	Gender, culture, and personality traits
	Theory of neurosis, character neuroses, and neurotic trends
	Alienation
Fromm	Importance of historical and social context for human behavior
	Marxist view of character types

Sullivan	Therapeutic communities
	Definition of personality and emphasis on interpersonal relationships
	Cognitive modes: prototaxic, parataxic, syntactic
	Social development model
	Interpersonal psychotherapy
Berne	“Strokes” as primary motivators
Chodorow	Feminist psychoanalysis
	Gender identity development
	Parent child relationship and “mothering”

He observed reaction patterns, attempted to predict the conditions under which they might occur, and tested and validated methods for their modification. He acknowledged the contributions of Freud and Jung but believed they were too narrow. A thorough pragmatist, he preferred common sense to metapsychological constructs as the means to understand and deal with psychopathology.

Theories of Personality and Psychopathology

Meyer believed human beings possess a fundamental tendency toward integration, and multiple biologic, social, and psychological forces contribute to personality development. The vulnerable person uses poorly planned, ill-suited means of adaptation. Meyer used the biographical approach as a practical guide to elicit information about personality development, organize the information, and re-evaluate the information obtained under different circumstances. His clinical examination assessed each patient’s life history; physical, neurologic, genetic, and social status; and the relationship between those factors and personality factors. A diagnosis and an individual treatment plan were based on this assessment.

FIGURE 6.3–1. Adolf Meyer. (Courtesy of National Library of Medicine, Bethesda, M.D.)

Treatment

The aim of psychobiological therapy was to help individuals make the best possible adaptation to changing environmental circumstances. It began with the development of a collaborative relationship. Out of this collaborative relationship came ***distributive analysis***, an examination of the factors in patients' lives that contributed to their adjustment or lack thereof, and concluded with ***distributive synthesis***, helping patients to understand themselves and to develop better coping skills. The first step in distributive analysis is the patient's own exposition of the presenting problem. Assets and liabilities are then determined by eliciting the life history in terms of the memories that are immediately available and those that are later fleshed out by reconstructing past experiences.

Treatment is initiated by ***focusing on patients' assets***. It involves psychological, chemical, physical, and environmental measures as needed. In more severe cases, attention is paid first to patients' sleep habits, nutrition, and daily routines because these must be normalized before psychological work can be done. Patients are helped to describe their difficulties in detail. In addition to eliciting complaints or worries, patients are asked what eases or worsens their complaints and what significance they attach to their symptoms and concerns. In doing this, the therapist attempts to use the patient's own language and concepts to communicate suggestions and advice.

Meyer did not pay attention to unconscious mechanisms but focused on patients' functioning in reality. Both present-day and long-term adaptive patterns are considered. Therapeutic sessions proceeded from immediate, obvious problems in the present to longer-term issues and historical data. With guidance, patients investigate their own personality problems, ascertain the origin of their conflicts, and work to develop more useful behavior patterns. Meyer called this ***habit training***, a term he may or may not have borrowed from a behaviorist tradition. When unhealthy adaptive patterns are modified, proper adjustment and personal satisfaction result.

Meyerian Treatment of Mr. A

Meyerians would focus first on the adequate treatment of Mr. A's mental illness. Controlling biologic forces that might disrupt his life is always a primary goal. Because he is still dependent on his parents, they are included in the treatment plan and might be seen separately by a social worker. Both Mr. A and his parents are told that failure to control the symptoms of his mental illness could cost him his life and any satisfaction that he might derive from it. Mr. A's mood swings are stabilized on appropriate medication, and he is observed closely for restoration of normal sleeping and eating habits. With our modern knowledge of neurobiology, the vulnerability of Mr. A's dopamine reward system to addiction would be another biologic force to be dealt with. He is asked to discontinue his habit of marijuana use for this reason, but also because of its tendency to aggravate the symptoms of his mental illness. Random urine screening might be recommended. Mr. A is helped to develop a daily schedule, including time for work, social interaction, and recreation. He is not encouraged to return to school because of his tendency to use school as an escape from responsibility. He is encouraged to develop his photography as a hobby after a period of stable work performance. Later, he would begin the distributive analysis phase of his treatment and would be asked to reflect on the impact of his bipolar I disorder and his avoidance of responsibility in his life. In the distributive synthesis phase of treatment, he is helped to realize that his avoidance of responsibility keeps him from achieving independence and the pleasure of attaining any realistic goal. He is asked to develop a plan for achieving both appropriate independence and for remaining within the family structure. Mr. A's parents are seen periodically for a progress report and to urge them to push Mr. A toward appropriate self-care, toward making use of his own resources. They might be asked to place Mr. A on an allowance and to hold him to it strictly to reinforce his accountability for his actions. Over time, his parents are asked to slowly withdraw financial support and to give Mr. A more independence. They might be asked to place his inheritance in a trust designed so that he could not impulsively misspend it.

ALFRED ADLER

Alfred Adler (1870–1937) was born in Vienna, Austria, and spent most of his life there (Fig. 6.3–2). A general physician, he became one of the original four members of Freud's circle in 1902. However, Adler never accepted the primacy of the libido theory, the sexual origin of neurosis, or the importance of infantile wishes. In 1911, he resigned as president of the Vienna Psychoanalytic Society and continued the development of his own socially and interpersonally focused theory of development. He posited a striving for self-esteem through overcoming a sense of inferiority, which he saw as an inevitable force in the human condition as a result of its extended childhood. He equated psychological health with constructive social engagement, developing a system that he called ***individual psychology***, which is still vigorous in many countries. His major social contribution was the establishment of child guidance centers in Vienna that served as a model for the rest of the world.

Dr Alfred Adler

FIGURE 6.3–2. Alfred Adler (print includes signature). (Bettmann/ Contributor via Getty Images.)

Personality Theory

Adler saw individuals as unique, unified biologic entities whose psychological processes fit together into an individual lifestyle. He also postulated a principle of ***dynamism***, which in every individual is future directed and moves toward a goal. Once the goal is established, the psychic apparatus shapes itself toward the attainment of that goal. Life goals are chosen and are thus subject to change. Changes require modification of memories, dreams, and perceptions to fit the accomplishment of the new goal. Adler also emphasized the interface between individuals and their social environment and the primacy of action in the real world over fantasy. Community mindedness, acceptance of the need to conform to the legitimate demands of society, is an important precept, but Adler also recognized a dialectic that occurs between individuals and their interpersonal environment, each constantly reacting to and shaping the other. Thus, Adler anticipated some of the modifications of Freud's theories introduced by Heinz Hartmann as well as some of Sullivan's thinking.

Normal Personality and Adaptation. The cornerstone of Adler's personality theory is the concept of ***moving from a sense of inferiority to a sense of mastery***. Early in life, everyone has a sense of inferiority resulting from realistic comparison with adults' size and abilities. Moving from this sense of inferiority to a sense of adequacy is the most important motivation in life. Thus, the ideal person strives for superiority and does so through high social engagement; the emotionally handicapped person continues to feel inferior and reinforces that position through a lack of striving and social interest.

There are many obstacles to the development of self-esteem and social interest. Prominent among them are poorly developed or "inferior" organs or systems (such as poor eyesight or poor eye–hand coordination), childhood diseases, pampering, and neglect. Physical disabilities and childhood diseases may promote self-centeredness and loss of social interest; ***birth order*** is another factor. According to Adler, first-born children, having lost their position as the only child, tend not to share and become conservative. Later, children are more open to change and become social activists. The youngest children feel secure because they have *never* been displaced. Adler may have been the first to anticipate recent psychological research on the importance of birth order in human behavior. We acknowledge our indebtedness to him whenever we use the term “inferiority complex.” Adler was also the first theoretician to place mastery and self-esteem at the center of personality development.

Theory of Psychopathology

Adler believed emotional disorders result from “mistaken lifestyles” that are subject to change by will and by self-understanding. Individuals subject to emotional disorders have false ideas about themselves and the world and inappropriate goals that lead them away from constructive social interest. Individuals with a pampered lifestyle, for example, expect and demand support from others, avoid responsibility, and blame others for their failures, but they also feel incompetent and insecure because their well-being depends on pressing others into service. If life poses no challenge, a mistaken lifestyle may have no consequences. When a mistaken lifestyle is ineffective, symptoms develop that protect self-esteem while helping the individual to avoid dealing realistically with the problem being confronted. The difference between neurosis and psychosis is that neurotic individuals maintain social interest but are blocked from life goals by symptoms, whereas psychotic individuals lose social interest and retreat into their own world.

Psychotherapy

Because his theory emphasized the mismatch of lifestyles with the demands of the real world, Adler focused on ***blocks to living productively in the real world***, not on exploring unconscious conflict. His aim was to point out mistaken self-views and mistaken views of the world and then, by mobilizing will, make the needed changes, including a change in life goals.

Therapeutic Process. Starting with three sessions per week and tapering off to one per week, a positive relationship with patients is established and used to lead patients to awareness of their lifestyle, how it is discordant with the demands of social reality, and ways in which it may be reoriented. Instead of striving for goals of no social value that falsely increase self-esteem, patients are pushed to work toward ameliorating their situation. Having become aware of the obstacles they have placed in their own path and the consequences of these self-defeating behaviors, they are now helped to develop constructive interests in themselves and others. As they become less self-engrossed, they find themselves better accepted by others, which reinforces their constructive efforts. People who have dedicated themselves to symbolically defeating others learn to cooperate and advance toward useful goals. Any endeavor in which patients can develop real competence is encouraged, whether social, vocational, artistic, or musical.

Patients are encouraged to address the concrete obstacles to developing a useful lifestyle, including reading instruction for slow readers or contact lenses for people who are self-conscious about their appearance. Early recollections, birth order, dreams, daydreams, and present-day interactions are all used to help patients see the inappropriateness or falseness of their ideas and life goals. Actual life events or memories of events are less important than individuals' reactions to those events or memories. Because memories are likely to be retrospective falsifications justifying an erroneous lifestyle, there is little need to verify them. There is also no need to look for latent content in dreams; they are merely an expression of present-day concerns. Nor is it necessary that therapists' interpretations be correct because they need only to help patients build a useful conception of themselves and the world. This perspective anticipated what has come to be called the *hermeneutic approach* to psychotherapy.

Several of Adler's techniques, including reframing and paradoxical communication, now enjoy wide popularity. ***Reframing*** is viewing the same data from a different point of view. Indecision, for example, was reframed from a product of mixed feelings to a wish to maintain the status quo. Failure to act keeps everything the same, which is the self-fulfilling prophecy of the discouraged person. After this reframing statement, patients are pushed to act constructively. ***Paradoxical communication*** is instructing patients to do the opposite of what the therapist wishes them to do. In dealing with an indecisive person, for example, Adler might caution against doing anything rash. Adler also paid attention to the impact of his patients on their environment and recognized that individuals do much to create their own interpersonal worlds. In response to complaints about being treated unfairly by others, Adler asked patients how they dealt with the people about whom they complained. Above all, Adler treated his patients as rational and as able to learn more productive ways of living. In his emphasis on practical, constructive solutions, misconstrued goals, and misperceived views of the world and the self, Adler also anticipated important elements of cognitive therapy.

Adlerian View of Mr. A

As seen from the Adlerian point of view, Mr. A has developed a mistaken lifestyle and an inappropriate life goal. He maintains himself in fantasy as a writer while failing at the accomplishments that would enable him to become a writer. Thus, he has been attempting to make the normal step of moving from inferiority to mastery in fantasy instead of through realistic achievement. Blocks to his development include pampering and his concomitant denial and abuse of his bipolar I disorder, the latter representing his organ inferiority. He lost the social interest he had earlier in life and became extremely self-centered, unconcerned with others or with the consequences of his actions. He uses drugs and hypomania to avoid the pain of defeat.

An Adlerian therapist encourages Mr. A to develop a realistic self-view. He has a mental illness that requires lifelong treatment. Grandiose dreams and intoxication with drugs cannot substitute for accomplishments in the real world and always lead to defeat. He is asked to set and strive to accomplish small, realistic goals for himself, such as holding a steady job while enjoying photography as a hobby. His mental illness is reframed as a challenge to his creativity and his use of marijuana as an obstacle to mobilizing his creativity. He might initially be encouraged to accept his dependency on his family and later, as he stabilizes, to reduce his dependency on his family as a means to heighten his self-esteem and to make a transition into adulthood. He is encouraged to join the Depression and Bipolar Support Alliance or other available support and educational groups as a means to better understand and accept his illness, to develop a social conscience, and to stimulate his altruism. His dream about flying is interpreted as his wish to achieve mastery; his dream about the hotel lobby is recast as awareness that he has been trying to substitute fantasy for reality.

CARL GUSTAV JUNG

Carl Gustav Jung (1875–1961) was a lifelong resident of Switzerland (Fig. 6.3–3). He trained in psychiatry under Eugen Bleuler at the Burgholzli Mental Hospital in Zurich and was strongly involved with Freud and the psychoanalytic movement from 1906 to 1914 when he resigned as president of the International Psychoanalytic Association. He was an important figure in connecting psychoanalysis with clinical psychiatry. After a period where he experienced psychotic symptoms but made use of them as a spur to creative thinking, which lasted from 1914 to 1918, Jung became an advocate of active introspection as the means to intrapsychic change. This episode and Jung’s subsequent interpretation of it, as well as that of his disciples, form the focus of much of the controversy about him and his ideas. Although Jung rejected Freud’s notion of libido as sexual energy and the Oedipus complex as a universal developmental stage, he believed not only in the unconscious mind but in a shared racial and species unconscious. This interest of Jung’s is the source of another controversy involving Jung’s relationship to the Nazi establishment. Jung, an intuitive introvert, was not interested in the practical aspects of living in the world; his focus was instead on individuation through becoming aware of the unconscious. His theories have particular attractiveness to clinicians and nonclinicians interested in spirituality, creativity, and mystical experience.

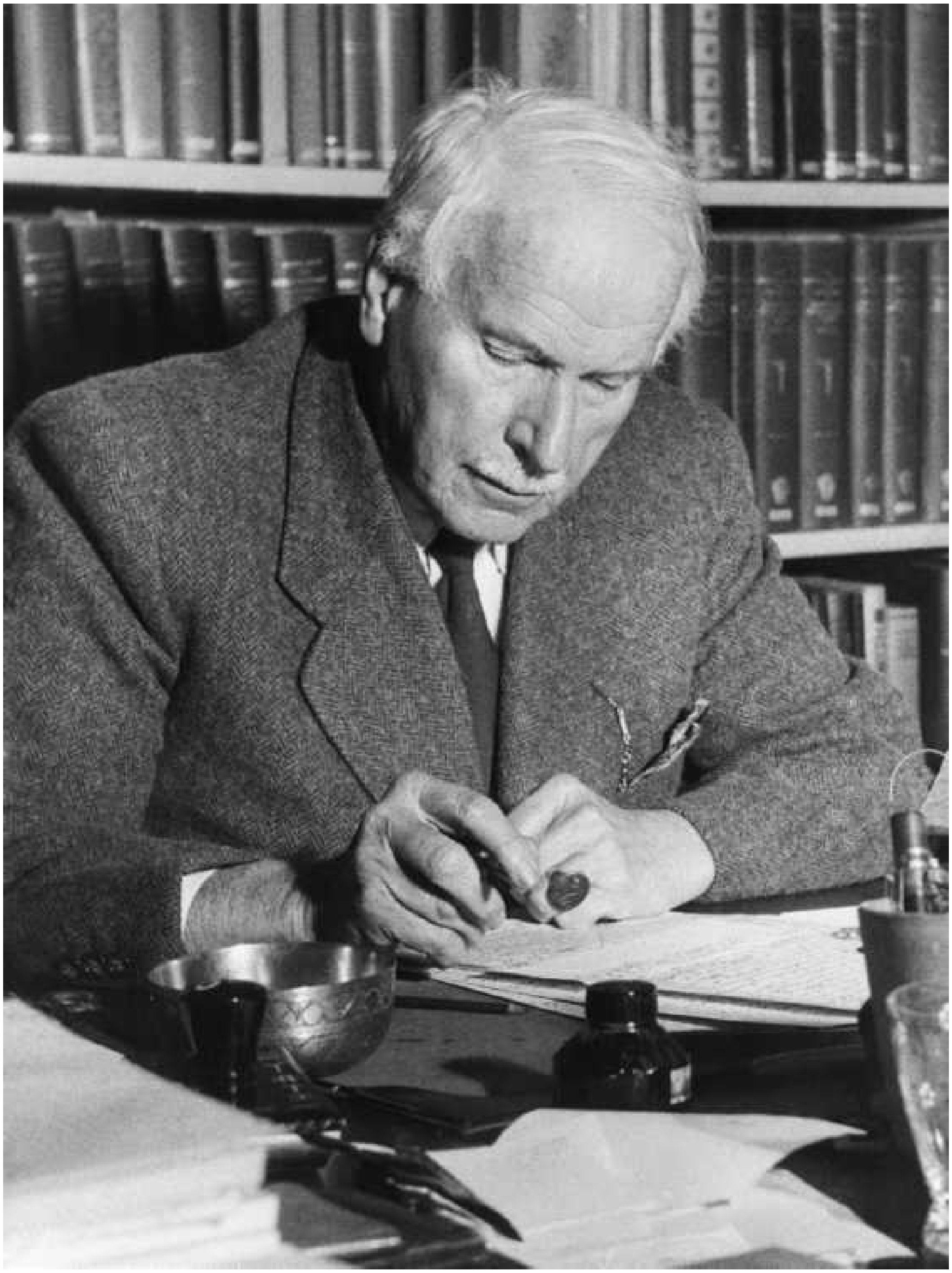

FIGURE 6.3–3. Carl Gustav Jung. (Bettmann/Contributor via Getty Images. The Getty image doesn’t include Jung’s signature in the lower left corner.)

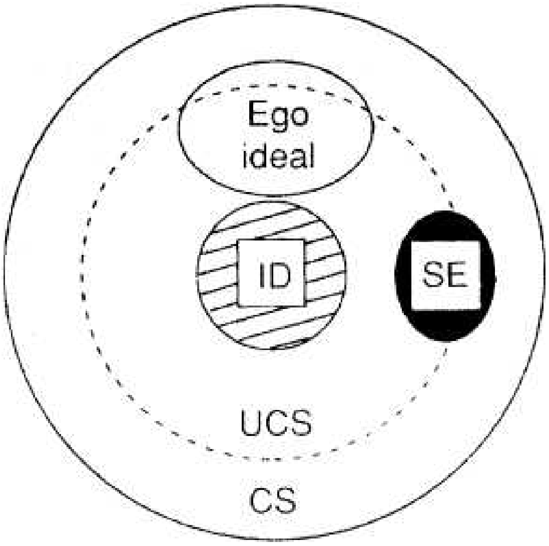

FIGURE 6.3–4. The Freudian topology of the psychic apparatus. CS, conscious; SE, superego; UCS, unconscious.

Personality Theory

Jung developed an elaborate metapsychology that was every bit as detailed and formalized as Freud's. His theories identify most of the important fault lines in classical psychoanalysis. Jung's construct of the psychic apparatus differs from the Freudian structure of the ego, superego, id, and ego ideal (Fig. 6.3–4). Below an outer rim of consciousness is the personal unconscious, which contains the complexes. Contained within the personal unconscious and connected to the complexes are the archetypes, the elements of the self, which in turn connect to the surface of the personality as the ego.

Complexes. ***Complexes*** are groups of unconscious ideas associated with particular emotionally toned events or experiences. Jung inferred them from his early word association studies when he noted that certain words provoked intense reactions or produced less reaction than expected. Complexes are built around genetically determined intrinsic models of the world known as ***archetypes***. Complexes are also reinforced by environmental events and by selective attention or inattention and are thus self-perpetuating. They are endowed with psychic energy from their affective tone: positive, negative, mild, or strong. The more intense the complex is, the greater the emotion, imagery, and tendency to action. Complexes are often stimulated by interactions with others. A father complex can be stimulated by a person who symbolizes a father (such as an older friend) or by a stimulus, such as music or art, that evokes father memories. The complex, formerly dormant in the unconscious, comes to the fore and tends to dominate consciousness and to displace other complexes. This is a very different model of the dynamic unconscious from Freud's, in that the boundary between the conscious and unconscious is far more permeable, and the emphasis is on what in external reality stimulates feelings and thoughts to enter awareness rather than on what forces a complex out of awareness. Similarly, Jung's conceptualization of interpersonal boundaries was far more permeable than Freud's; Jung also placed much more emphasis on nonverbal stimulation (music, art, spiritual life) in this process than did Freud.

Some complexes are more conscious, better developed, and more acceptable to the individual; others are less conscious, poorly developed, and uncomfortable. The latter are projected onto the environment, especially by children, and from this, ***projective*** and ***introjective*** processes evolve. One person may introject and identify with a complex being projected by another person. Thus, therapists may become psychologically infected by their patients. It is also possible to project a complex that is not integrated within oneself onto another person and then develop a relationship with that projected complex. One can envision an interpersonal environment charged with mutual projections available for introjection, thus offering endless potential for mutating perceptions and misperceptions. Thus, for Jung, the psychological boundary between individuals was far less clear than for Freud or Adler. In addition, one can see that Jung was much less concerned with an ultimate, objective, rational reality to be determined and compared with the individual's complexes of projections and introjections. Jung also had a parallel appreciation for the process of projective identification that Melanie Klein was developing at approximately the same time. Working as he did for many years in a long-term sanatorium with very ill patients, Jung clearly became comfortable with blurred interpersonal boundaries. His theory has been criticized for accepting as normative what might be regarded as pathologic models of relationships.

Another important aspect of complexes is their ***bipolarity***, again paralleling Klein's work on splitting. Each complex has a positive and negative pole, such as good father and bad father or rewarding father and punishing father. One pole of a complex can be projected onto another person, who in turn introjects it and acts on it in a relationship. In this way, the theory of complexes is a theory of interpersonal as well as intrapsychic relationships.

In Jungian theory, the ego is also a complex. It serves the same function as the Freudian ego, controlling conscious life and bridging the intrapsychic and external worlds. The other complexes that also make up the psyche may align with or lie in opposition to the ego. Emotionally charged primitive complexes (e.g., the earth mother in a patient ambivalent about motherhood) have a tendency to become autonomous and may, when activated by external stimuli, behave to oppose or control the ego, leading to irrational, impulsive behavior or misperceptions. They appear as images in dreams, as hallucinations, and as separate personalities in dissociative identity disorder. They also appear in séances when mediums bring forth so-called personalities from the dead. For Jung, this phenomenon also explained animism and states of possession.

Jung had a great interest in mystical experience, which Freud saw as the persistence of infantile, magical thinking, primary process, or wish fulfillment. This is another of the clear divisions in their thinking. For Freud and Freudians, the world is a harsh, demanding place that forces one to give up primitive wishes for magical experiences, symbiotic relationships, and intense gratifying emotional moments. For Jung and Jungians, magic is alive and well in the inner world and is to be embraced.

Archetypes. Complexes are connected to structures embedded more deeply in the psychic apparatus, the archetypes. ***Archetypes*** are the inherited capacity to initiate and carry out behaviors typical of all human beings, regardless of race or culture, such as nurturance and aggression. Just as vision cannot develop without visual input during critical physiologic stages, so archetypes require interactional stimulation for their elaboration into complexes. Thus, the human infant’s psyche is not amorphous energy awaiting organization by the environment; it is instead a complex and organized set of potentials whose fulfillment and expression depend on environmental stimuli. There are as many archetypes as there are prototypical human situations. In positing the archetypes based on anthropologic data, Jung was remarkably prescient of the current understanding of the brain’s organization.

The mother complex archetype illustrates the interrelationship of complex and archetype. All humans are born with a poorly formed but relatively clear model of an all-nurturing caretaker, which Jung referred to as the *earth mother archetype.* The mother complex emerges from this based on experiences with mothers or mother surrogates: their attitudes, personalities, and relationship to the particular individual. The mother archetype is found in dreams or fantasies, often as a huge woman or an animal with many breasts. The motif of a many-breasted animal, found in many cultures, is that of unlimited nurturance.

Unconscious. The Jungian unconscious has two layers: the more superficial being the personal unconscious and the deeper layer the collective unconscious. The complexes exist in the personal unconscious, the archetypes in the collective unconscious or *objective psyche.* The *personal unconscious* is the equivalent of the Freudian unconscious, a repository of individual memories that have been repressed. The *collective unconscious* is the residue of what has been learned in humankind's evolution and ancestral past, much as human deoxyribonucleic acid (DNA) is an aggregate of the past. In this portion of the psyche reside instincts, the potential for creativity, and the spiritual heritage of humankind. The potent synthesis of Darwinian natural selection theories with Mendelian genetics, which has had such a profound impact on all of biology, occurred in the 1920s. It is not clear whether Jung was aware of this intellectual thrust, but his understanding of the collective unconscious and of the archetypes is remarkably consonant in form, if not in detail, with the modern understanding of hereditary and developmental neurobiology as applied by cognitive science.

The psyche, like all other living systems, attempts to stay in balance. Jung's term for homeostasis in the relationship of conscious to unconscious life is the *law of compensation*. For any conscious attitude or experience that is overly intense, there is an unconscious compensation. A person experiencing neglect might fantasize or dream about a huge, many-breasted mother. When interpreting dreams, Jung asked himself for what conscious attitude the dream compensated.

Personality Structure. At the center of the conscious personality is the complex called the *ego*. Several universal complexes attend to the ego. The ***persona***, or public personality, mediates between the ego and the real world. The ***shadow***, a reverse image of the persona, contains traits that are unacceptable to the persona, whether positive or negative. A brave persona, for example, has its fearful shadow. The archetype of the shadow is the enemy or feared intruder. The ***anima*** is a residue of all the experiences of woman in a man’s psychic heritage; the ***animus***, the residue of all the experiences of man in a woman’s psychic heritage. The anima or animus connects the ego to the inner world of the psyche and is projected onto others in day-to-day or intimate relationships. When connected with the shadow, a man, for example, might see attributes of woman as undesirable and might experience guilt encountering such qualities in himself.

SELF. The *self* is the archetype of the ego; it is the innate potential for wholeness, an unconscious ordering principle directing overall psychic life that gives rise to the ego, which compromises with and is partly shaped by external reality. In Jungian metapsychology, the unconscious gives rise to integration, order, and individuation. The self appears from the unconscious in dreams, fantasies, and altered states of consciousness to give direction. In the first half of life, the ego attempts to identify with the self and to appropriate the power of the self in the service of the ego’s growth and differentiation. During this time, the ego may become inflated with an unrealistic sense of power. If it is cut off from the self, there may be a sense of alienation and depression.

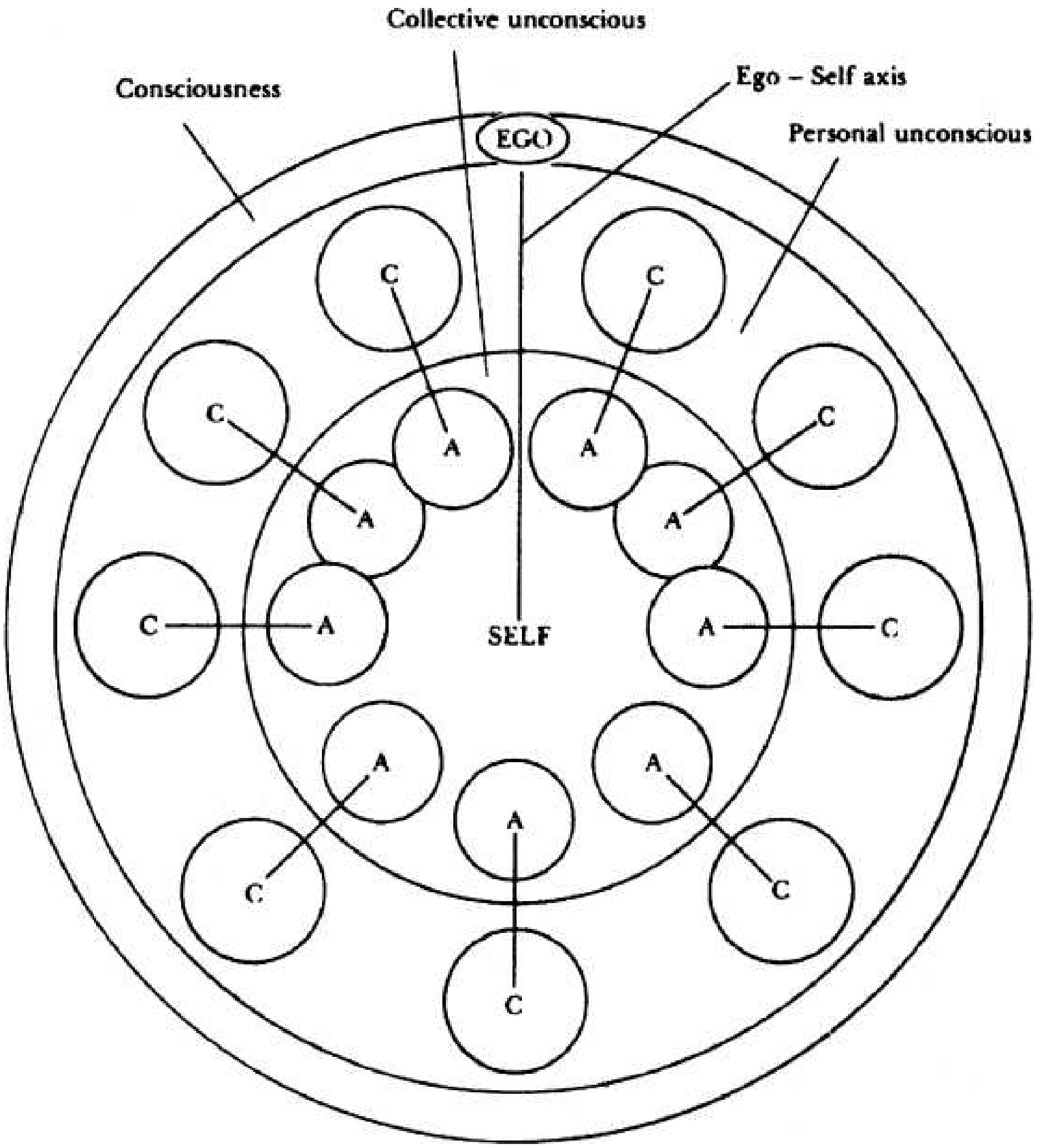
Collective unconscious
Consciousness
Ego – Self axis
Personal unconscious
EGO
C
A
SELF

FIGURE 6.3–5. The Jungian psychic apparatus. A, archetype; C, complex. (Reprinted with permission from Princeton University Press. Stevens A. On Jung, Updated Edition, 1999.)

INDIVIDUATION. In the second half of life, the ego begins to attend more to the self than to the conscious realm of life. Jung called this developmental process *individuation,* the drive for individuals to become unique and to fulfill the spiritual propensities common to all humanity. Often, this process requires withdrawing from earlier identities and conventional definitions of success and seeking new paths. This change often has the paradoxical effect of leading to broader and more mature relationships in addition to greater creativity.

PSYCHOLOGICAL TYPES. Jung's theory of psychological types has three axes (Fig. 6.3–5). The *extroversion–introversion* polarity refers to the two basic types of object relatedness. Extroverts are oriented to others and to the world of consciousness. Their energy flows outward first, then inward. Introverts are oriented to their inner world, their energy flowing first inward and then to outer reality. Introverts might therefore be seen as selfish and unadaptable because they attend first to their inner world and then determine how the outer world can fit them. Extreme extroverts, on the other hand, can seem insensitive to themselves and to the inner lives of others.

The *sensation–intuition* polarity concerns perception. The sensation type comes to understand a situation by assembling the details; the intuitive type grasps the overall situation before attempting to assimilate its parts. The sensation type sees the trees first; the intuitive type sees the forest first.

The *thinking–feeling* polarity deals with information processing and judgment. In the *thinking* mode, data are evaluated according to logical principles. *Feeling*, at the opposite pole, is making judgments through nonlogical processes having to do with values and understanding relationships. In social relationships, the thinking type deals with people according to their social rank or according to the tradition of etiquette; a feeling type deals with others in terms of their present social relationship or perceived emotional state.

Each individual has a preferred mode on each of these three dimensions. By placing an individual on each of the three axes indicated in Figure 6.3–6, each individual can be identified as a type. An extroverted–sensation–thinking type is oriented to the real world, tends to perceive external details, and organizes them into a logical structure. An introverted–intuition–feeling type is self-oriented, grasps situations as a whole, and is sensitive to their emotional implications.

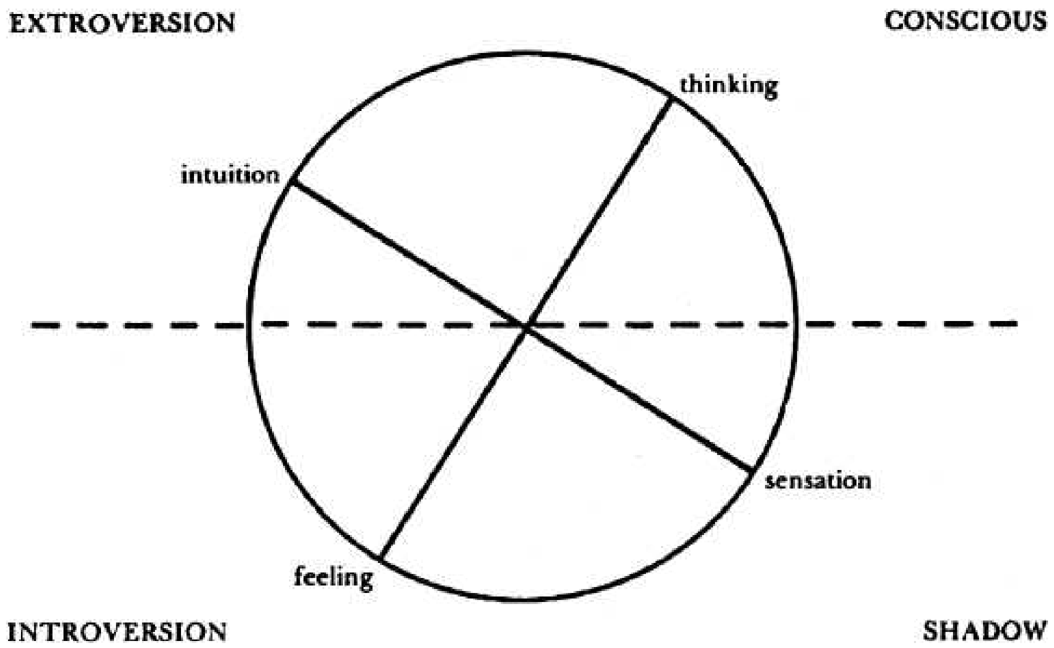

FIGURE 6.3–6. The three personality axes: extroversion–introversion, thinking–feeling, and intuition–sensation. (Reprinted with permission from Princeton University Press. Stevens A. On Jung, Updated Edition, 1999.)

Everyone's psyche contains all of the types. However, each person has a superior set of types that are evolved from early life and that are shaped strongly by temperamental factors. In the second half of life, adults who continue the process of individuation attempt to integrate or broaden and deepen their understanding of their inferior functions. The Myers–Briggs Type Indicator, a simple paper-and-pencil test consisting of approximately 40 questions, reliably places individuals on these dimensions. It has become a very popular instrument among psychology, management consulting, counseling, and self-improvement organizations for quickly identifying the qualities of individual participants. This typology may be the most widely known and accepted aspect of Jung's theories, although it is often used independently of its original context.

Therapeutic Process

Jung suggested that therapy begin with four visits per week and then be spaced out to one or two per week. Present-day Jungian practitioners may work with their analysands once a week, face to face. Jung placed great emphasis on the human relationship between analyst and analysand and noted that both parties change in the course of analysis. He defined *transference* as the patient's attempt to develop psychological rapport with the doctor and held that without rapport and object relationship, the technical operations of the analyst were of no value.

With its emphasis on reintegration, symbolism, and dreams, Jungian analysis seems well suited to help educated people deal with the developmental problems of midlife. Having achieved a professional identity, material success, and established a firm family role, they often begin to ask, “Who is the real me?” “What is most meaningful to me?” and “What is my relationship to humankind and human history?” However, the complicated, often blurred intrapsychic relationships described and the interest in mystical experience with its potential shading into magical thinking mean that Jungian clinicians must be careful about the risk of blurring the therapeutic boundaries. The break between Freud and Jung, which had for decades been ascribed to Freud’s rigidity and authoritarianism, now appears to have had a contribution from Jung: Freud disapproved of a complicated, dependent relationship Jung is believed to have had with a former patient.

Jungian View of Mr. A

Jungians see Mr. A’s symptoms of bipolar I disorder and use of marijuana as having unleashed conflicting archetypes, all attacking the ego complex he had grown up with. This could be seen as an opportunity for Mr. A to claim a fuller definition of the self. Mr. A’s aspirations to be a great writer are taken as potentially healthy, and even the content of his psychotic and substance-induced states is analyzed so as to incorporate their meanings into his ego complex. Mr. A’s previous psychological type is described as extroverted–sensation–thinking—an observer of life but not a full participant. The type emerging in his psychotic and intoxicated states is described as introverted–feeling–intuitive—the self-absorbed, self-experiencing lover of beauty and ideas. The dreams about flying suggest an expression of the hero archetype; the anxiety dream suggests some emergence of his shadow in association with his aspirations to be a writer. His love of flowers and creativity during his periods of psychosis and intoxication suggests that the shadow is connected to his anima.

The task of therapy is to assist Mr. A in accepting these alien parts of himself and integrating them into his ego complex. Relatively less emphasis is placed on education or specific treatment of his bipolar I disorder, except insofar as it prevented a good therapeutic alliance. His fantasies, dreams, and psychotic experiences are investigated, discussed, and analyzed in terms of the psychological functions used and the archetypes expressed. His interpersonal issues with his family are seen primarily as projections by him of other unacceptable archetypes that need identification, analysis, acceptance, and integration. For example, his relationship with his mother is seen as a projection and fear of the earth mother archetype. The overall goal is to harness the unleashed unconscious material to allow Mr. A to pursue his ambitious and creative goals.

FRANZ ALEXANDER

Franz Alexander (1891–1964) was one of the second generation of psychoanalysts. He was born in Budapest and attended medical school there, graduating in 1912 (Fig. 6.3–7). He conducted research at the Institute for Experimental Pathology in bacteriology until World War I, when he practiced clinical microbiology on the Italian front, primarily combating malaria. After the war, he joined the Department of Psychiatry at the University of Budapest Medical School as a brain researcher. This led to an encounter with Freud's work, and, in 1919, he became the first student of the Berlin Psychoanalytic Institute. In 1930, Alexander became a visiting professor of psychoanalysis at the University of Chicago and, in 1932, founded the Chicago Psychoanalytic Institute. He established the institute independent of the psychoanalytic societies, leading the Chicago Institute to be one of the more creative sources of psychoanalytic thought. During the same time, he became interested in what he called psychosomatic illnesses (today referred to as somatic symptoms and disorders), helping to found the journal *Psychosomatic Medicine.* A guiding principle of his work was to make psychoanalysis an integral part of medicine.

FIGURE 6.3–7. Franz Alexander. (Courtesy of Franz Alexander.)

In 1946, Alexander became a professor of psychoanalysis at the University of Southern California, where he continued his work on ***psychosomatic medicine*** and became interested in the interface of learning theory, the psychophysiology of stress, and psychoanalysis.

Theory of Personality

Alexander did not develop a unique overarching theory of personality; his contribution was his application of psychoanalytic thought to pathophysiologic processes. In this, he laid the groundwork for the burgeoning fields of psychosomatic medicine, behavioral medicine, and psychophysiology. Alexander created the basis for the biopsychosocial model and studied the mind and body at a time when American psychiatry, despite Freud's original ideas, had become purely psychological in its orientation. In studying and treating many patients with serious physical illnesses, he was also forced to consider creative modifications of therapeutic techniques.

Alexander and his group began by intensively studying, by means of clinical interviews, patients who had one of seven illnesses that had been identified by general practitioners as having strong psychological components: peptic ulcer disease, ulcerative colitis, essential hypertension, Graves disease, neurodermatitis, rheumatoid arthritis (RA), and bronchial asthma. Out of these clinical studies emerged the specificity hypothesis, which proposed that certain illnesses are the product of a complex interaction of specific constitutional predispositions, specific unconscious conflicts, and specific types of stressors that activate such conflicts. The group then tested these hypotheses in a series of clinical studies of patient populations. The independent variable was usually the ability of skilled clinicians to predict a patient's illness from disguised case reports.

In the field of psychotherapy, Alexander was one of many who sought to shorten the analytical process. He hypothesized that intellectual insight was not the central curative factor in therapy, a radical and revolutionary proposal at the time. Rather, he emphasized the role of ***corrective emotional experience*** or the exposure to past conflicts or emotions in the safety of the therapeutic setting**.** The corrective emotional experience allowed patients to re-experience old conflicts and painful emotions while creating a new outcome within the context of a more positive emotional experience with the analyst. During the session, patients may draw on strength and support from the therapist, which provides an opportunity for processing the emotion in a healthier way. Alexander began to experiment with variations in technique that might facilitate such experiences, including using transference to serve as a proxy parent and respond to patients in ways that correct the pathologic patterns of parental responses they received as children. Consideration of a corrective emotional experience as a core feature to treatment, his most controversial stand, nearly ruptured his relations with the psychoanalytic movement.

Theory of Psychopathology

The seven diseases that Alexander studied (peptic ulcer disease, ulcerative colitis, essential hypertension, Graves disease, neurodermatitis, RA, and bronchial asthma) were studied to identify the single, specific core conflict that, in interaction with constitutional predispositions and in the circumstance of a particular stressor, would activate the disease.

The core conflict identified in peptic ulcer disease was hyperindependence as a defense against unacceptable dependency needs. The stressor that results in an acute attack can be any situation that demands that the individual openly acknowledge or ask that dependency needs be met. Ordinarily, patients use dominance and control to intimidate others into meeting their dependency needs. This particular illness model received confirmatory evidence from a study of army inductees in whom psychological profiles as described by Alexander, in combination with measurements of serum pepsinogen, were successful in predicting the development of duodenal ulcers. It is not clear what the impact of the recent recognition of a specific pathogen means for this particular hypothesis. Perhaps the identified pathogen, *Helicobacter pylori*, is the constitutional factor that had previously been believed to be serum pepsinogen concentration, or perhaps Alexander's hypothesis only applies to some subgroup of patients with duodenal ulcers. More recent research suggests a number of correlates of peptic ulcer in addition to *Helicobacter pylori,* including use of pain-relieving medicine (particularly nonsteroidal anti-inflammatory drugs or NSAIDs), heavy alcohol consumption, use of nicotine products, low socioeconomic status, and different types of stressors such as physical, psychological, and occupational. Evidence supports the impact of stressors on ulcers, but not specifically the effects of unconscious conflicts or defenses.

Alexander's theory of ulcerative colitis also implicated dependency conflicts; however, rage at unmet needs was seen as the defining feature. This rage provokes guilt and the urge to make restitution toward the object of anger by means of gifts of achievements and successes. The model here is clearly the angry child seeking to placate a parent by means of performance. The precipitating event for reactivation of the illness is the perception that the efforts at placation will be unsuccessful. Alexander claimed that this resulted in excess parasympathetic activity, leading to diarrhea. Current research on the etiology of ulcerative colitis remains inconclusive, with some research suggestive of little to no connection between psychiatric diagnoses or stressful life events and disease status and severity. Other research indicates a moderate correlation between depression and Crohn disease, but not ulcerative colitis. However, evidence for Alexander's theory of dependency conflicts causing ulcerative colitis has not been supported or confirmed.

Alexander's hypothesis about essential hypertension focused on inhibited anger and suspiciousness in an outwardly compliant, cooperative individual. A hypertensive patient often goes for long periods with blood pressure under good control and then, seemingly inexplicably, experiences a dramatic increase in blood pressure. Alexander attributed these episodes to incidents when the chronic anger is exacerbated, and intense defenses must be used, causing chronic sympathetic fight-or-flight activation. Several psychophysiological studies have suggested that this theory has some validity, at least in terms of short-term changes in the blood pressure of a subset of people with hypertension. Individuals with labile mood and hypertension appear to fit this model best. This theory is not generalizable to all cases of hypertension and does not account for other factors identified in current literature. Modern-day views of hypertension focus on various causative factors such as genetics, behavior, and psychosocial stress. Evidence supports the impact of chronic psychosocial stress on the risk of developing high blood pressure over time. Factors such as socioeconomic status, racial discrimination, sex, and stress in occupation and relationships are also being studied as potential correlates to the development and exacerbation of hypertension.

The proposed specific conflict for RA postulates rebellion against overprotective parents. A compromise formation in which the conflict is discharged via physical activity, especially sports, works for a time, but the anger eventually is expressed in self-sacrifice designed to control others. Failure of this pattern results in increased ambivalent tension, directly expressed via muscular contractions that lead to joint degeneration. Alexander's specific theory about RA is not supported by current research. However, psychological stress, particularly adverse childhood events and chronic distress, can exacerbate symptoms of RA and increases the risk of developing arthritis.

Alexander proposed that the wheeze of bronchial asthma represented a symbolic cry. The specific conflict, according to him, was the wish for protection versus the fear of envelopment. This conflict leads to sensitization to separation issues, which become the events that provoke the suppressed cry of the asthma attack. In recent years, it has become clear that the population with asthma is far more heterogeneous both psychologically and physiologically (in terms of vulnerability to allergens) than was recognized in Alexander’s time, and that immunologic variants play the primary role in the illness etiology and not unconscious wishes and fears. The role of a vicious physiologic–psychological cycle in which asthma stimulates panic, which in turn triggers pathologic pulmonary psychophysiological responses, has been a focus of more recent research.

Although the role of stress and anxiety in neurodermatitis remains widely accepted by clinicians, the specific conflict proposed by Alexander, that early deprivation leads to wishes for closeness that are opposed by a fear of it, is no longer accepted. Finally, Graves disease (thyrotoxicosis) is no longer widely accepted as a purely somatic disorder. Alexander’s hypothesis was that premature responsibility led to a martyr-like denial of dependency. Although the exact cause remains unknown, genetic and environmental risk factors play a role, and physical and psychological stress can precede the onset of the disease.

Treatment

The specificity hypothesis led Alexander to focus his psychotherapeutic efforts in a way that other analysts of his time did not. He reasoned that if he could help patients to resolve their core-specific conflict without necessarily addressing other parts of their personality structure, the medical illness would improve. Indeed, he published numerous case studies suggesting just this kind of success. In addition, he was among the first to question the value of intellectual insight as to the curative agent in psychotherapy. He proposed that a corrective emotional experience is the central agent of change. A corrective emotional experience involved disconfirmation within the transference relationship of previous assumptions and projections.

Thus, Alexander felt justified in introducing a variety of techniques that would initially induce and heighten the emotional experience of the transference and, subsequently, challenge the underlying unconscious assumptions. These techniques included manipulating the frequency and length of sessions, making direct suggestions about the patient’s life, self-conscious alteration of the therapist’s behavior according to the patient’s conflict, and behavior therapy techniques. In many ways, this was the most controversial aspect of Alexander’s work. Serious questions about the validity of his suppositions and the ethics of his “manipulative stance” were raised. He was impatient with the slow, methodical process of convincing his colleagues; his intellectual energy led him to embark on ever-newer experiments while other analysts were still struggling to digest his previous suggestions. Yet today, few quarrel with the concept that emotional learning is at least as important as intellectual insight in psychotherapeutic success. Alexander's efforts to modify and shorten the analytical process have come closer to the norm in psychiatric practice than classic analysis. Indeed, his emphasis on a specific focal conflict that could be addressed in brief therapy by modifying techniques anticipated the later work of Peter Sifneos, David Malan, Habib Davanloo, James Mann, and Gerald Klerman, and Myrna Weissman, who systematically developed broad-based models of brief psychodynamic psychotherapy.

Ironically, although Alexander's therapeutic innovations seem prescient today, his specificity hypothesis, which was far less controversial, seems simplistic, naive, and forced. He did not have available the sense of how complicated illness causation truly is. The dominant model of the time was infectious disease: One organism, one illness. In addition, the complexity of social phenomena and stressors was unknown in his time. Yet, Alexander was the first to postulate a multicausal etiology for disease: a specific constitutional defect, a specific conflict, and a specific stressor, all necessary for a disease to occur. He was also the first to study mind–body interactions in a systematic way.

Alexander's View of Mr. A

In the absence of a clear somatic disorder, Alexander would have had little to say specifically about Mr. A's dynamics. However, like any good dynamically oriented psychotherapist, he would have readily recognized Mr. A's core conflict about dependence–independence. Mr. A would be understood as rebelling against his mother's nurturing control, although secretly desiring to continue basking in it. He has serendipitously found that illicit drugs and his mental illness can be harnessed such that his mother has to rescue him and take over while he avoids becoming aware of the conflict. Alexander's unique contribution to Mr. A's case would be to propose that he could be treated with 40 or fewer sessions, focusing exclusively on this conflict (and ignoring others such as oedipal components and issues with his father). Further, Alexander's agenda would be to structure the therapeutic relationship to intensify Mr. A's experiencing this conflict. He might do this by taking a somewhat intrusive, directive stance, thus creating the opportunity for a rapidly intense transference around this issue. Alexander might then spread out the sessions to elicit expression of the patient's yearning for that control. The key therapeutic effort is then to respond constructively to attempts by Mr. A to assert himself appropriately and take over control of his life—for example, by learning more about his illness and taking responsibility for his treatment.

WILHELM REICH

Wilhelm Reich (1897–1957) was one of Freud's most controversial disciples; his later years were marred by mental illness (Fig. 6.3–8). Reich fixed on and elaborated Freud's early, but later discarded, view that neurosis results from the damming up of sexual energy. In so doing, Reich laid the groundwork for a psychoanalytic theory of personality. Before Reich, the focus was on symptoms and psychopathology. There was occasional mention of "character" in psychoanalytic work, but there was no focus on personality as an entity unto itself. This shift is of great importance because, as neurobiology has come to explain more and more Axis I mental disorders, it has become clear that one of psychodynamic theory's enduring strengths involves understanding and treating personality traits and disorders. Together with Anna Freud and Hartmann, Reich is regarded as the originator of ***ego psychology***.

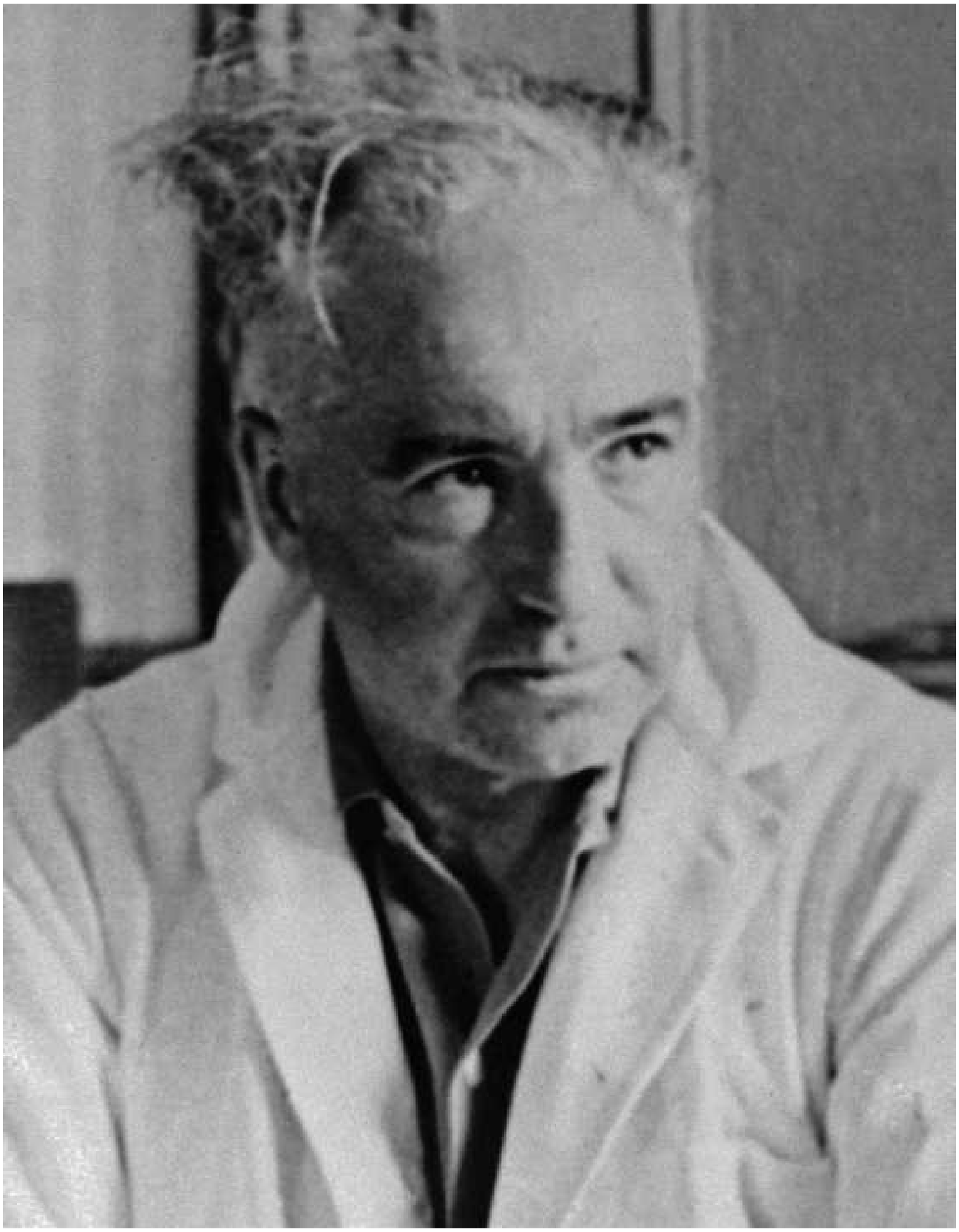

FIGURE 6.3–8. Wilhelm Reich. (Bettmann/Contributor via Getty Images.)

Personality Theory

Reich did not disagree with Freud's notions concerning personality development, including character types based on fixation at specific levels of psychosexual maturation. Reich elaborated on the interpersonal and physical behavior of these personality types. Specific behavioral traits constitute ***character armor*** that defends against internal and external dangers. Character armor is comprised of involuntary, repetitive, ego-syntonic behaviors that prevent the emergence of repressed impulses. For instance, the trait of ingratiation frequently defends against hostile impulses, just as the traits of hostility or self-assertion may defend against wishes to be dependent and passive. These traits manifest physically in the voluntary musculature as characteristic postures (clenched jaw or fist, rigid or bowed back) or in excessive fluidity of movement. Although Reich's ideas of the behavioral and postural components being central features of character armor are no longer accepted, his ideas successfully transformed the focus of psychodynamic thought from individual wishes defended against by individual defenses, expressed in specific transferential forms, into an emphasis on patterns, organization, and intertwining of all of these elements. He also added substantially to the attention paid to nonverbal cues and their implications.

Hysterical Character. The hysterical character has the least body armoring, hence the most lability of function. Body movements tend to be soft, rolling, and sexually suggestive. These individuals are superficial, excitable, flighty, fearful, highly suggestible, and easily disappointed. Their armor helps to defend against easy sexual arousability by flushing out potential sexual stimuli in the environment and then reacting to them with anger.

Compulsive Character. These individuals are tense and restrained, walk stiffly, and sit rigidly. They are overconcerned about orderliness, tend to ruminate, and are indecisive and distrusting. They experience a blockage between their thoughts and feelings. Because they have little access to their feelings, they have little ability to prioritize their actions, make decisions, or sense others' reactions to them. The compulsive character avoids the expression of repressed impulses by rigid overcontrol. Because of this, these individuals are very threatened by trivial changes in routine.

Phallic-Narcissistic Character. Phallic-narcissistic people appear cold, reserved, and prickly. They are outspoken, provocative, and seek positions of power. Frustrated at the genital-exhibitionist stage of development, the men are identified with the penis, and the women with the fantasy of having a penis. The men have strong erective potency but little capacity for intimacy; the women actively dominate men.

Masochistic Character. Masochistic individuals suffer, complain, damage, and deprecate themselves in ways that provoke and torture others. Reich differed from the analytical interpretation that these people enjoy suffering. He believed the opposite—that pleasure is painful for the masochist because of an enormous need, excessive guilt, and the resultant low tolerance for love or pleasure. Suffering allows the masochist to then indulge in a certain amount of self-gratification. Sexual intercourse can be enjoyed, for example, if the partner is inconsiderate or overly domineering.

Treatment

Reich's major contribution was in the realm of treatment. He was the first to recognize the need to deal with character resistances before attempting to recover repressed material, and that interpersonal resistances need to be dealt with before free association is possible. He did this by analyzing patients' character armor—their characteristic behaviors (including tone of voice, posture, and physical movements)—in the analytical setting before proceeding to an analysis of the unconscious. Reich worked face to face with patients and sought to relax their character armor by physical manipulation. This type of therapy, called *vegetotherapy* by Reich, was later practiced by Reich's followers as ***bioenergetics***.

Reichian View of Mr. A

A Reichian analyst might see Mr. A's seeming rebellion against conformity as armor against his fear of being away from his mother's protection and, more importantly, a fear of domination and dependence. The Reichian might point out that the more Mr. A rebels, the more tightly he binds himself to his mother. The therapist might also suggest that the behaviors that seem so much under Mr. A's control are, in fact, masochistic behaviors. His prolongation of his manic states and use of drugs, as well as difficulties with task completion and subsequent praise, help him sustain feelings of need and guilt and avoid those of true pleasure. The "pride" he takes in embracing mental illness may be viewed as deprecation as a way of provoking others and maintaining overall suffering. In addition, the family structure does not allow for the direct provocation of Mr. A's mother. As such, masochistic behaviors may be the only way Mr. A can engage in provocation. Of note, in Reichian analysis, the therapist would specifically attend to nonverbal cues that would reflect this character armor and potentially even work directly with the patient to physically relax their body posture.

KAREN HORNEY

Physician–psychoanalyst Karen Horney (1885–1952), who emphasized the pre-eminence of social and cultural influences on psychosexual development, focused her attention on the differing psychology of men and women and explored the vicissitudes of marital relationships (Fig. 6.3–9). She was one of three women whom Freud trained specifically to have female analysts who would contribute their unique perspective to the psychoanalytic theory about women; the other two were Anna Freud and Helene Deutsch. Horney's view that repression and sublimation of biologic drives are not the primary determinants of personality development led to her removal as an instructor in the New York Psychoanalytic Institute and her founding, in 1941, of the American Psychoanalytic Institute.

Personality Theory

Horney believed that personality development results from the interaction of biologic and psychosocial forces that are unique for each individual. At the core of each personality is an enduring real self, which combines choice, will, responsibility, and identity with spontaneity and aliveness. A natural unfolding process of self-realization leads to the development of human potential in three basic directions: toward others, to express love and trust, against others, to express healthy opposition, and away from others toward self-sufficiency. Horney proposed three main character types, which are based on the predominant mode of relating to others. The *compliant, self-effacing type* results from the defensive operation of clinging to others. These individuals try to curry favor with others, subordinate themselves, and are reluctant to disagree. The *aggressive, expansive type* results from moving against others and relies heavily on power and mastery as a means to achieve security. The *detached, resigned type* results from moving away from others in an attempt to avoid both dependency and conflict. These are very private individuals who, although refusing to compete openly, see themselves as rising above others.

FIGURE 6.3–9. Karen Horney. (Bettmann/Contributor via Getty Images.)

Horney believed that all human beings have basic anxiety, which she saw as the normal response to the infant's helplessness and separateness. Although conditions during childhood may block psychological development, healthy growth is always possible if the internal blockages can be removed. How families respond to a child's anxiety, guided by cultural norms, determines whether individuals spend their lives struggling with basic anxiety or pursuing self-realization. To Horney, attributes of passivity and suffering were not biologically determined or specific to women, as taught by the analysts of her day. She believed that male and female personalities are, in fact, ***culturally determined***.

Horney asserted that differences between men and women were not simply based on anatomical differences and biology but also on differences in role expectations. She suggested that from birth, girls are treated as second-class citizens who are seen as inferior to boys. Expected gender roles encourage women to be dependent on men for “love, wealth, care, and protection.” Societal views of women placed their worth and value in the role of wife and mother alone. Because culture shaped expectations, a woman’s value was assumed to be derived from connection to and relationships with men and children. Women were not seen as valuable in their own right but only as a function of their place in the family and home. Horney wrote, “Women’s efforts to achieve independence and enlargement of her field of activities are continually met with a skepticism which insists such efforts be made only in the face of economic necessity, and that they run counter to their inherent character and natural tendencies. Accordingly, all efforts are said to be without any vital significance for women whose every thought should center upon the male and motherhood.” Horney criticized the androcentric nature of psychoanalysis and the focus on male development and sexuality, suggesting that women’s ability to become pregnant and bear children were underappreciated and neglected. She emphasized the importance of cultural factors in women’s position as inferior and suggested women were not envious of the penis (as was suggested by psychoanalytic theories of the day), but rather the superior position held by men in society. Furthermore, based on her clinical data, she asserted that womb envy in boys was at least as likely as penis envy in girls. Horney’s beliefs about the cultural influence of personality development led to the emergence of feminine psychology and an increased focus on the development of girls and women.

Theory of Neurosis

Horney defined neurosis in both intrapsychic and interpersonal terms. She noted that her patients complained not of the symptomatic neuroses, such as phobias and compulsions, but of unhappiness, blockage, lack of fulfillment in their work, and inability to establish or maintain relationships. She saw these individuals as having a complex system of self-perpetuating defensive patterns against basic anxiety—***character neuroses***.

Safety-seeking children move psychologically in three directions to relieve their anxiety, make life safe and predictable, and achieve satisfaction. They seek affection and approval, become hostile, or withdraw. Children eventually use the coping strategy that best meets their needs, but if only one basic strategy is used, children become limited in their coping repertoire and in their experience of themselves and their world. Their sense of safety is tenuous because there is now danger from within, from suppressed or repressed feelings and impulses. Given continued unfavorable environmental conditions, conflicting feelings are driven into the unconscious, and such children are left with a sense of discomfort, anxiety, apprehension, and an insecure sense of self. At this juncture, their point of reference is externalized; patterns of behavior become more rigid and increase blockages to growth and development. Horney designates these complex, relatively fixed attitudes toward self and others as ***neurotic trends***.

Idealized Image. Individuals who grow up to be neurotic create an ideal image that, if achieved, promises to end their painful feelings and provide self-fulfillment. The idealized image covers over all the contradictions, conceals the defensive nature of their behavior, and restores a sense of wholeness. Energy formerly available for self-realization is now used in efforts to become like the idealized image. For example, an individual who has adopted the strategy of moving toward others and is consequently dependent on others for affection and approval experiences the fear of reasonable self-assertion, striving instead for saintly humility and considerateness of others.

Neurotic people are easily bruised by confrontation with reality and work excessively to prove they are, in fact, their ideal selves. This results in a type of perfectionism that insists on flawless excellence in which “I should” replaces “I want” or “I need.” It also results in the neurotic ambition to be first and fosters a strong drive for revenge against those perceived as having interfered with their efforts to match the ideal self. This aspect of Horney’s theories anticipates some of Heinz Kohut’s ideas on the origins of narcissism.

Claims, “Shoulds,” and Self-Hatred. Despite their frequent self-disparagement, neurotic individuals expect to be treated as though they were their ideal selves. These claims to special treatment, when frustrated, produce anger, righteous indignation, and resentment. The “shoulds,” or self-imposed demands that they live up to their idealized selves, are irrational and unrelated to the realities of daily life. They are projected, experienced as demands made by others, and are also demanded of others. This results in neurotic people being critical of others and very sensitive to criticism themselves. Self-hatred results when the threat arises that neurotic individuals may be unable to achieve their idealized selves. If support is not needed for the idealized self, claims, shoulds, and self-hatred are not such important parts of the psychic apparatus.

Neurotic Pride and the Pride System. Glorifying aspects of the idealized self, *neurotic pride*, substitutes for healthy self-confidence. Thus, when their pride is injured by others, neurotic individuals become enraged and seek to avenge their injury and conceal their self-deception by achieving a vindictive victory over the offending person. Together with supporting claims and shoulds, neurotic pride and self-hatred form a defensive network or pride system that protects the idealized self. Despite the armoring of their defensive network, such individuals are not at peace because they are in inner conflict with the forces that protect them. The conflict between the forces driving toward healthy self-realization and the pride system is the *central inner conflict*.

Alienation. ***Alienation*** from self is one of the most serious consequences of neurotic development. It results from the combination of repeated denial of external reality and the repression of genuine thoughts, feelings, and impulses. As the process of alienation continues, neurotic individuals lose touch with the core of their being and can no longer determine or act on what is right for them. Their feelings may range from uncertainty and confusion to inner deadness and emptiness.

Therapeutic Process

Horney did not regard adult neurotic people as recapitulating childhood experiences and thus did not focus on the recovery of childhood memories; she dealt instead with the self-perpetuating neurotic process. She stressed the importance of dreams in analysis. As unconscious attempts to solve conflicts, dreams can show constructive forces at work that are not yet discernible in patients' conscious thoughts and behavior. Later, she stressed the exploration of the patient–analyst relationship. She was one of the earliest analysts to recognize and make constructive use of her own feelings toward patients. To Horney, psychoanalysis was a cooperative venture that enabled patients to free themselves from their neurotic structures and mobilize themselves toward self-realization. The analyst's responsibility was to assist in liberating patients from *blockages*, the forces that impede healthy growth.

Early in therapy, termed the *disillusioning process*, the two types of blockages are identified and examined. The first group of safety-oriented blockages, *protective blockages*, helps to avoid the anxiety caused by self-awareness. They include silence, lateness, depreciating the analyst, the use of drugs, and even the use of self-accusation as a means to avoid further exploration. *Positive-value blockages* reinforce patients' satisfaction with themselves and support their idealized selves. In the disillusioning process, the analyst identifies both types of blockages, exposing the protective blockages before exposing the blockages that defend the idealized image. Analyzing the positive-value blockages first arouses too much fear. As the central conflict is resolved successfully, patients move into the final phase of treatment: the discovery and use of their real inner self.

Horney's View of Mr. A

From the standpoint of Horney, Mr. A's process of self-realization has been blocked in all three directions. He has not developed the ability to love and to trust; he expresses opposition in an unhealthy way, and he has made self-defeating moves toward independence. He is seen as having developed a detached style of relating to others, having substituted hypomania and the use of drugs for real relatedness. He justifies his illness by taking pride in it. His goal of becoming a writer is part of the development of an ideal self that is a detached observer. He supports his ideal self by having fantasies of involvement with other writers, by convincing his professor that he was able to write a play as his term paper, and by procrastination. He has become more and more isolated from his real self by denying the reality of the fact that his mental illness and cannabis abuse pose a danger to him and that he was failing in school.

A therapist in the Horney tradition begins by pointing out that Mr. A's drug use and his sustaining of manic episodes are blocking his ability to learn about himself and works with Mr. A toward abstinence from substances of abuse and appropriate use of mood-stabilizing agents. The therapist later begins to point out that Mr. A's pride in his manic episodes serves the purpose of sustaining a false self of boundless energy and creativity. Mr. A is encouraged to decide who he really wants to be—a writer in fantasy or a person able to obtain real satisfaction from real accomplishments and real relationships. His dreams are examined and their themes explored. His omnipotent, messianic dreams are interpreted as evidence of his ideal self and viewed in the light of reality. His dreams of being exposed are interpreted as the fear engendered by his ideal self as a means of defending itself against exposure: "If you expose me, you will be embarrassed and humiliated." Mr. A is supported through his fear of humiliation and the pain of realizing that he has been deceiving himself. His self-loathing is interpreted as the activity of the pride system in defending his ideal self. As he begins to relinquish his ideal self, he will begin to discover and mobilize his real inner self in directions that might not have been predictable at the beginning of treatment.

ERICH FROMM

Psychoanalyst Erich Fromm (1900–1980) was often thought of as the archetypical neo-Freudian, the leader among those who emphasized that culture and social setting influence an individual's dynamics as much as instincts do (Fig. 6.3–10). Neither physician nor biologist, Fromm, a native German, received his doctorate in philosophy, sociology, and psychology from the University of Heidelberg in 1922. There he was exposed to a Marxist emphasis on how history shapes societies and how societies, in turn, shape individuals according to economic needs. He was trained as a psychoanalyst at the Berlin Psychoanalytic Institute and then founded, with his wife, Frieda Fromm-Reichmann, the Frankfurt Psychoanalytic Institute. In 1933, he immigrated to the United States and in 1949 moved to Mexico City to found another psychoanalytic institute. In 1974, he moved to Switzerland, where he died in 1980. As much a social critic as a personality theorist, he was later claimed by the existential and humanistic psychoanalysts. Fromm's intellectual agenda was the integration of

Freud’s theory of a dynamic unconscious with Karl Marx’s theory of history and social criticism.

FIGURE 6.3–10. Erich Fromm. (Hulton Archive/Stringer via Getty Image.)

Personality Theory

For Fromm, two central facts dominate human behavior: The inevitability of separateness and the ***historical and social moment*** into which each person is born. He argued that every person yearns to recapture the state of blissful union that existed prenatally. From the moment the baby begins to recognize itself as a separate human being, a titanic struggle begins, pitting the desperate anxiety of loneliness against the urge to fully express and actualize oneself and, ultimately, to transcend the self. Most individuals find the loneliness too painful to bear and suppress their striving for individuation in the service of maintaining the illusion of connectedness. They are socialized by their parents into the roles defined by their society of birth. Fromm actually used the term *symbiosis* years before Margaret Mahler employed it to describe the universal human yearning for fusion, safety, and security while facing aloneness and choosing individuation to lead to freedom and a productive life. However, true freedom is too terrifying for many people, who instead construct a series of illusions that engender the feeling of safety. They create a pseudoself, think pseudothoughts, and experience pseudofeelings in support of these illusions, thereby cutting themselves off from the fullness of their own inner lives. Fromm saw Freud's theory as a special case of his own more general ideas. The illusions that Victorian society offered involved sublimation of sexuality and aggression in the service of social respectability. Social respectability, in turn, provided the illusion of acceptance and security. In other places and at other times, different solutions might be offered. Early in World War II, shortly after his escape from Nazi Germany, Fromm wrote about the willingness of people to give up their freedom to serve an authoritarian society. In the 1950s, he wrote of the pursuit of material acquisitions in the service of postwar productivity, leading to self-satisfied conformity.

To achieve true freedom, Fromm said four basic human needs must be met: *relatedness*, *transcendence*, *identity*, and a *frame of orientation*. *Relatedness* is the need to feel connected to other humans. *Transcendence* refers to rising above basic instincts. *Identity* is the need to feel accepted yet unique. Emphasis on the need for a *frame of orientation* led Fromm late in his career to an exploration of the constructive and destructive roles that religion may play in individual lives.

Theory of Psychopathology

As a social philosopher and critic, Fromm did not really develop a systematic theory of psychopathology. He identified three major mechanisms of retreat from individuation. Some individuals, he said, may seek an authoritarian solution, trying to live through someone or something external to themselves to achieve a sense of adequacy. Others may become destructive, attacking anything that confronts them with their separateness and aloneness. Most individuals develop a conformist attitude, warding off the anxiety of experiencing their own intentionality by accepting socially offered thoughts, roles, and attitudes.

These mechanisms result in four unproductive orientations or ***characters*** typical of modern capitalist society: *receptive*, *exploitative*, *hoarding*, and *marketing*. The *receptive character* often appears to be cooperative and open; however, the primary agenda is to establish a passive relationship with a leader who solves problems magically. *Exploitative characters* are likewise interested in filling themselves up from the outside; however, they aggressively manipulate and usurp whatever reduces their terror, for example, power. *Hoarders* collect, store, and close in on themselves, often being cold and aloof in their efforts to feel secure. *Marketers* treat themselves as a plastic commodity to be manipulated as needed to achieve externally validated success.

Treatment

Fromm wrote nothing at all on the practice of psychotherapy; therefore, what is known is derived from anecdotal reports by those who studied with him or were treated by him. They report his emphasis on a tender and empathic inquiry into the self-deceptions and illusions created by patients in their efforts to ward off the anxiety of separateness and to maintain some sense of connectedness to significant others. He placed great emphasis on the tendency of unloved children to identify intensely with parental values to capture the magical safety they seem to offer. At a time when most psychoanalysts were preoccupied with detailed examinations of the instincts and defenses, Fromm contributed a sense of the range and richness of inner experience that underlies superficial adaptation. He contributed the sense that a new authenticity could be found by those willing to confront the truth about themselves with all its terror of aloneness.

Fromm’s View of Mr. A

In some respects, Fromm’s ideas are uniquely applicable to Mr. A, who is clearly maintaining the illusion of his separateness by conforming to the socially sanctioned role of the rebellious artist. The bipolar I disorder becomes an unexpected weapon in his battle with his terror of aloneness. The dread in the hotel lobby dream is probably the closest this fear comes to consciousness. Fromm would gently probe this dream and Mr. A's self-image. He would point out the chains created by Mr. A's seemingly rebellious independence. The self-destructive quality of the patient's lifestyle would also be confronted and investigated. Ultimately, Mr. A has to experience his own loneliness and face his terror of it to find true freedom, true intimacy, and an authentic self-definition.

HARRY STACK SULLIVAN

Harry Stack Sullivan (1892–1949) is generally acknowledged as the most original and distinctive American-born theorist in dynamic psychiatry (Fig. 6.3–11), and most American psychiatrists make significant use of concepts and approaches he developed. For many years, the primary theoretical dispute within dynamic psychiatry circles was between the classic Freudians and the Sullivanians (or interpersonal psychoanalysts). When psychiatrists use the term *parataxic distortion*, apply the concept of self-esteem, consider the importance of preadolescent peer groups in development, or view a patient's behavior as an interpersonal manipulation, they are applying concepts Sullivan first proposed.

Sullivan graduated from medical school in Chicago in 1917. From 1921 to 1930, he worked with patients with schizophrenia at St. Elizabeth's and then Sheppard and Enoch Pratt Hospitals, where he developed a reputation as a remarkable clinician with an uncanny ability to communicate with floridly psychotic patients. He initiated the first of what are now called ***therapeutic communities***. Later, he entered private practice in New York and eventually returned to the Washington, DC area, where he was involved in clinical, consulting, and teaching activities. In the 1920s and 1930s, he wrote a number of papers on schizophrenia, later collected in *Schizophrenia as a Human Process*. His other books were compiled from his lectures by his students; most were published posthumously, explaining some of the density and seeming disorganization of his written work.

FIGURE 6.3–11. Harry Stack Sullivan. (Courtesy of the New York Academy of Medicine.)

Personality Theory

Sullivan rejected the Kraepelinian dogma of his day that dominated psychiatric thinking about schizophrenia. In searching for alternative understandings of psychosis, he initially turned to Freud but rejected his theories as increasingly rigid and dogmatic. Thus, he developed his own working theory of personality, psychopathology, and therapy.

Sullivan believed language could be misleading and was therefore wary of self-reifying conceptualizations that led to rigid theories. He emphasized the psychiatrist as participant/observer in the clinical situation seeking to keep observations as objective as possible while recognizing the difficulty this presented in dealing with private emotional experience. What can be observed is the social interaction of patients; thus, he defined ***personality*** as the "relatively enduring pattern of interpersonal relations which characterize a human life." From the outset, his focus was very different from the intrapsychic emphasis of psychoanalysis. By approaching psychopathology in this way, he necessarily created a field theory, characterized by temporal and interactive processes, rather than a structural theory.

Sullivan's theory is fundamentally one of needs and anxiety. *Needs* are divided into needs for satisfaction and needs for security. *Anxiety* occurs when fundamental needs are in danger of not being met and is the primary motivator of human behavior. Needs for satisfaction include physical needs (e.g., air, water, food, warmth), and emotional needs include need, especially for human contact and for expressing one's talents and capacities. Because infants are utterly unable to meet their own needs, provision of basic physical needs by caretakers and interpersonal relationships are crucial. Decades before Mahler wrote of a symbiotic stage in infant development, Sullivan spoke of the "empathic linkage" between caretaker and infant and described the complicated interaction of infants communicating tension and anxiety, arousing anxiety in the caretaker, leading to tender responses to the infant's needs. Failure to meet these needs results in loneliness and anxiety.

Sullivan defined *security* as the absence of anxiety. Thus, *needs for security* are defined as the need to avoid, prevent, or reduce anxiety. Needs for satisfaction and security, particularly the lack of such needs being met, can have long-lasting effects on an individual's health and quality of life. Research suggests that social determinants of health, including safety in the home, school, and work environments, engagement in recreational activities, and access to adequate education, health care, and social support (to name only a few) impact the development and exacerbation of mental and physical health disorders. Specifically, deprivation in these and other areas may lead to greater health disparities, limitations to accessing treatment, and increased morbidity and mortality rates. If needs for satisfaction and security are generally met, major negative health effects may not emerge, but anxiety remains inevitable and, as Sullivan theorized, becomes the primary driver for personality development. The *self-system* is defined by Sullivan as the dynamism that is responsible for avoiding or reducing anxiety. Sullivan equated the self, identity, or ego with the individual's developed patterns for avoiding the discomforts that arise from the inability of others to meet one's fundamental needs. It exists, like all else, purely within an interpersonal framework.

The self-system develops a set of mechanisms, called *security operations*, which effect this goal. Security operations function within Sullivan's theory much as defense mechanisms do within psychoanalytic theory. The specific security operations, however, were defined interpersonally, and Sullivan tried to link them closely to actual observation or experience. Some bore the same labels and definitions as Anna Freud's, but Sullivan is best known for three contributions that bore his distinct stamp: apathy, somnolent detachment, and selective inattention. These were drawn from observing the way infants and young children react to painful interactions, such as scolding, with their parents.

The self-system accrues from ever-evolving interpersonal experiences—that is, fulfillment of needs for satisfaction as a result of the empathic linkage with the mother. The most difficult experiences are not necessarily those involving the inability to meet the child's needs but the child's sensing of the caretaker's anxiety in the process of responding to those needs. This arouses anxiety in the child, promotes the need to establish a sense of security, and leads to the evolution of the self-system and the development of security operations. The self-system is divided into three parts. The "good me" is a set of images, experiences, and behaviors associated with an unanxious, tender, empathic, and approving or accepting response from the environment. The "bad me" comes to be associated with ideas, actions, and perceptions that provoke anxiety and disapproval from caretakers. Some situations, however, provoke such intense anxiety that they are entirely disavowed and disowned; they become part of the "not me." Eventually, the empathic linkage becomes unnecessary, and the self-system operates autonomously within the individual, developing ever more subtle and complex ways to manage the person's anxiety.

Developmental Theories. Sullivan had two theories of development, one cognitive, the other social. He postulated three developmental ***cognitive modes*** of experience whose degree of persistence into adulthood is important in understanding psychopathology. The *prototaxic* mode, characteristic of infancy and early childhood, involves a series of disconnected, brief states experienced as totalities with no temporal relationship. In later life, mystical experiences represent persistent prototaxic experiences. *The parataxic* experience begins early in childhood as the self-system begins its more independent functioning. It, too, involves a series of momentary experiences; however, they are now recorded in sequence and with an apparent connection to one another. They may be given symbolic meanings, but rules of logic are absent, and coincidence plays a major role in how the world is perceived. The self-system uses this mode to seek effective anxiety-reducing behaviors and to repeat them, seeking sameness and predictability. Sullivan used this mode to explain transference, slips of the tongue, and paranoid ideation. The *syntactic* mode of experiencing is based on the development of language and consensual validation. The world and the self are perceived within rules of logic, temporal sequencing, external validity, and internal consistency. Thinking about oneself as well as others becomes testable and modifiable based on rigorous analysis of experiences in a variety of different situations. *Maturity* may be defined as an extensive predominance of the syntactic mode of experiencing.

Social development is somewhat based on these evolving cognitive modes. However, disturbed interpersonal relationships may cause persistence of the more primitive (prototaxic or parataxic) ways of experiencing the world. Infancy spans birth to the onset of language and is characterized by the primary need for bodily contact and tenderness. The prototaxic mode predominates, and the primary zones of interaction are oral and, to some extent, anal. Insofar as needs are fulfilled with a minimum of anxiety, the infant experiences euphoria and a sense of well-being. To the extent that some anxiety is commonly present in the caretakers, apathy and somnolent detachment are regularly used as security operations, persisting into adult life as a basic detached and passive stance. If anxiety and inconsistency are severe, intense experiences of dread persist, presenting in later life as the eerie, uncanny, bizarrely disruptive internal states seen in individuals with schizophrenia.

Childhood begins with the onset of usable language, continues until the beginning of school, and is characterized by the child’s focus on the parents as the other from whom praise and acceptance are sought. The primary mode of experience shifts to the parataxic, and the most common zone of interaction is anal. The child needs an approving adult audience. This leads to a variety of learning tasks of language, behavior, self-control, and so on. It can also be observed in a variety of trial-and-error efforts by the child to find what pleases the adult. Gratification leads to an expansive self-system with many facets of life associated with the “good me” and positive self-esteem. Moderate anxiety leads to chronic anxiety, uncertainty, and insecurity. Extreme anxiety results in giving up known successful behavior in favor of self-defeating patterns that fulfill others’ expectations.

The juvenile era covers ages 5 to 8 years. The shift to syntactic cognitive modes begins, and the interpersonal focus spreads to the peer group and outside authority figures. Peers and teachers have the opportunity to approve and accept behavior previously frowned upon within the family (e.g., talking dirty with one’s friends). Interpersonal cooperation, competition, play, and compromise become the gratifying experiences. Juveniles learn to negotiate between their own needs and legitimate social concerns without sacrificing their self-esteem in the process. The risks of excessive anxiety are either too great a need to control and dominate social situations, or they become an internalization of restrictive, prejudicial social attitudes.

Preadolescence, ages 8 to 12 years, marks the child’s movement from peer group cooperation and competition based on roles toward genuine intimacy with what Sullivan called a chum or a best friend. Sullivan saw this phase as a particularly important stage in which the give and take of the friendship could repair and undo distortions that resulted from excessive anxiety at earlier stages. This is the point at which the individual truly moves outside the family and engages in free, creative interaction with another person unfettered by the same dynamics. During this stage, the major shift toward syntactic thinking takes place, although some distortions may persist into adolescence. The preteen years see the initiation of a capacity for attachment, love, and collaboration or their inability to develop these in the face of excessive anxiety. Although sexual exploration may be a part of the friend relationship, Sullivan did not see sexuality as a central element in this developmental phase.

Adolescents, beginning at puberty, are seen to have concerns similar to those of preadolescents, except that sexual desire is added to the interpersonal equation. Thus, the same needs for a relationship persist but shift to the other sex for their outlet, wherein a major opportunity for learning or severe anxiety begins. As the person faces culturally defined stereotyping, many new opportunities for social experimentation may lead to consolidation of self-esteem or self-ridicule. The struggle to integrate lust with intimacy is accomplished by painful trial and error. If this is completed with the self-system relatively intact, the later years of adolescence are an opportunity to expand the syntactic mode to such areas as a consensual view of interpersonal relations, values and ideals, career decisions, and social concerns (Table 6.3–2).

Theory of Psychopathology

Sullivan abhorred diagnostic labeling for being unhelpful, overly restrictive, and dehumanizing. In discussing schizophrenia, he said, “We are all much more simply human than otherwise.” Thus, he sought to understand the fundamental human process within his patients, especially his sickest ones. He saw psychopathology as resulting from excessive anxiety arresting development of the self-system, thereby limiting both opportunities for interpersonal satisfaction and available security operations. He viewed psychiatric patients as struggling to maintain their self-esteem with very limited means. To understand them, the developmental phase at which they operate has to be gauged, and the interpersonal needs they express have to be understood.

Table 6.3–2.
Sullivan’s Social Development Theory

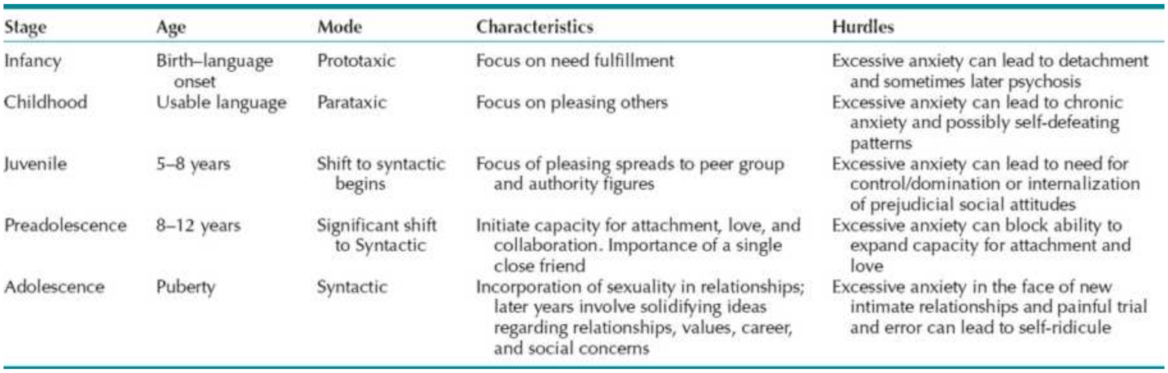

Stage	Age	Mode	Characteristics	Hurdles
Infancy	Birth–language onset	Prototaxic	Focus on need fulfillment	Excessive anxiety can lead to detachment and sometimes later psychosis
Childhood	Usable language	Parataxic	Focus on pleasing others	Excessive anxiety can lead to chronic anxiety and possibly self-defeating patterns
Juvenile	5–8 years	Shift to syntactic begins	Focus of pleasing spreads to peer group and authority figures	Excessive anxiety can lead to need for control/domination or internalization of prejudicial social attitudes
Preadolescence	8–12 years	Significant shift to Syntactic	Initiate capacity for attachment, love, and collaboration. Importance of a single close friend	Excessive anxiety can block ability to expand capacity for attachment and love
Adolescence	Puberty	Syntactic	Incorporation of sexuality in relationships; later years involve solidifying ideas regarding relationships, values, career, and social concerns	Excessive anxiety in the face of new intimate relationships and painful trial and error can lead to self-ridicule

Sullivan believed that several factors could play a role in the particular form that these disturbances might take. The level of anxiety at particular developmental stages can lay the groundwork for a developmental arrest. Basic cognitive capacity might play a role in the choice of security operations relied on or retained. The degree of success achieved interpersonally combined with whatever capacities are used affects later success. Finally, the chance occurrence of stresses encountered during life is deemed a factor. Thus, Sullivan theorized that anyone might develop schizophrenia, even people with relatively successful developmental histories, should their chosen defenses fail dramatically and their life stresses mount in the extreme. However, it was more likely that patients with schizophrenia would be highly vulnerable along all four dimensions, whereas others with greater developmental strengths might become obsessive, hysteroid, schizoid, or paranoid.

Interpersonal Psychotherapy

Sullivan's ***interpersonal psychotherapy*** approach emphasized that the psychiatrist is a participant–observer in all interactions with patients. He thought deeply and extensively about the nuances and opportunities involved in this unique situation. By interacting actively with patients, verbal and nonverbal expressions of recurrent interpersonal patterns become apparent. These observations then inform the therapist's further behavior, thereby creating the opportunity for change. This process occurs over the seconds and months, and years as the psychotherapy unfolds. Sullivan saw this perspective as an antidote to what he perceived as the wrongheaded emphasis on objective neutrality embodied by the "blank screen" model of psychotherapist behavior. He argued that parataxic distortions emerge in all interactions, not only in the classic analytical situation. This differing view of transference and of it being a universal human process was among the core debates for decades between classic analysts and interpersonal analysts.

Sullivan saw therapy as elucidating the patient's interpersonal patterns, exploring their usefulness in the service of the patient's needs, and considering alternative, more favorable possibilities. Thus, he shared the ego psychologists' understanding that even the sickest behavior was the best adaptation available at a given moment to the patient. He emphasized the experiencing of the distortions, the needs, the patterns, and the potential changes within the ongoing interaction with the therapist.

He saw great power in the very entanglement of the therapist with the patient and recognized the ability of a skilled therapist to manage the interpersonal process to reveal patterns and to shape the patient's emotional experience. However, he constantly emphasized and respected the ultimate autonomy of his patients, who could still, in the end, choose not to reshape their approach to the world.

Sullivan viewed psychotherapy as divided into four distinct stages: *inception*, *reconnaissance*, *detailed inquiry*, and *termination*. *Inception* involves the very beginning, often only a part of the first interview, during which the contract and roles are stipulated. *Reconnaissance* might go on for as many as 10 to 15 sessions, during which the therapist identifies the patient's recurring patterns and assesses their adaptive and maladaptive qualities. The *detailed inquiry* is a very lengthy process of exploring the patient's thoughts, feelings, and memories and evaluating and re-evaluating data from earlier stages, seeking to recognize, clarify, and change persistent parataxic distortions. The recurrent patterns are discussed within the context of the person's developmental history, needs, anxieties, failures, and successes. There is often much ongoing interchange between patient and psychiatrist as feelings and perceptions are validated or questioned within the context of mutual emotional interchange within each session. *Termination* is a product of the evolving contract and understanding between the patient and therapist and may reflect either extensive or limited goals. Sullivan emphasized the constant reassessment of goals by the therapist and the power of the ongoing negotiation and renegotiation of the therapeutic contract as a means to reveal and change parataxic distortions. The ultimate goal of psychotherapy is to experience as much as possible within the syntactic mode and to broaden the repertoire of the self-system. To the extent that this is achieved, individuals are in a position to become responsible for their ongoing growth through subsequent interpersonal interactions.

Sullivan's View of Mr. A

Sullivan would see Mr. A as likely arrested in childhood when his fear of displeasing his mother led him to give up healthy self-esteem strivings for independence in favor of a distant yet dependent position. He uses drugs and psychosis as escapes to maintain some degree of self-esteem. His "good me" consists of his debating and his intelligence. His "bad me" is expressed in his rebelliousness, whereas the "not me" seems to encompass issues of closeness, independence, and constructive engagement with others and with life’s tasks. In therapy, the reconnaissance phase is extremely important, identifying the moment-to-moment interactions through which Mr. A’s security operations interfere with his attempts at constructive independence. His bemusement and detachment are identified, as is his escape into prototaxic thought through drugs and noncompliance with medications. His acceptance of his mother’s management of his trust fund is noted as well. A Sullivanian therapist actively interacts, empathically identifying and confronting Mr. A’s ways of avoiding authenticity and constructive interaction. Once identified, their meanings are gently probed, as are the feelings associated with them. The terror of displeasing his mother and its effects on Mr. A’s ongoing interactions with the therapist are addressed. Interactions are examined for their consequences, with the assumption that the outcome has a bearing on the motivation. Finally, Mr. A is encouraged to try out different ways of relating to the therapist as a prelude to restructuring his outside relationships. He is encouraged to think rationally about his circumstances, to seek a good peer group, develop close friendships, and gradually move away from his near-exclusive focus on his immediate family for all of his interpersonal needs.

FIGURE 6.3–12. Eric Berne. (Courtesy of Wide World Photos.)

ERIC BERNE

Eric Berne (1910–1970) was an American original in both style and substance. He worked in the San Francisco area most of his career, breaking with psychoanalysis in the mid-1950s but never becoming antianalytical, as did many of his followers (Fig. 6.3–12). Like many others, he felt the need to develop briefer treatments than were offered at the time.

A group gathered around him that came to be known as the San Francisco Transactional Analysis Seminars. Through weekly discussions of clinical cases and social and political issues, Berne gradually refined his theory. Berne was wry and provocative in his approach to human behavior, contributing much to “pop psychology” in the 1970s and 1980s. Few clinicians now call themselves transactional analysis therapists; nonetheless, Berne’s ideas remain useful in grasping hidden agendas in human interactions.

Personality Theory

For Berne, the primary motivator of all human behavior is the need for “***strokes***”—attention, recognition, and response from others. Early survival depends on adequate physical contact, stimulation, and nurturance. This need remains strong but later becomes more symbolic and interpersonal. Children rapidly learn what works within their families and practice it extensively. This led to one of Berne’s more widely quoted observations: “Negative strokes are better than no strokes at all.” People evolve ways of interacting with their world to obtain regular strokes in whatever way possible and in whatever way they have been taught to define a stroke (e.g., sympathy in response to chronic depression may provide such gratifying attention that the depression cannot be given up). So great is the need for regular stroking that blatantly destructive actions persist in the face of insight, recognition, and enduring psychological or physical pain.

Like Adler and Sullivan, Berne suggested that hidden social needs motivate human behavior to the extent that, with rare exceptions, there is an interpersonal hidden agenda in all human activity. For Berne, the unit of observation was the transaction, the short-term process of individuals interacting with one another. He spent much energy analyzing transactions to try to discover patients’ definitions of strokes and their preferred mechanism for obtaining them. He noted that most people engage in very predictable, stereotyped, repetitive transactions. The content may vary from situation to situation, but the form tends to be quite rigid. He called these transactions ***games***, and his bestseller, *Games People Play*, captured the imagination of the American public in 1964. Some games are harmless, some are socially encouraged, many have destructive elements or at least limit opportunities for more gratifying relationships (intimacy), and some are highly destructive. A common, socially accepted game is cocktail party flirtation, which is ordinarily pleasant and harmless but, depending on the intensity, frequency, and seriousness with which it is played may cause an inability to experience intimacy, disrupted marriages, or even physical harm.

Berne divided the human psyche into three primary parts: *child, parent*, and *adult*, with two of those further subdivided. He called these *ego states*. An *ego state* consists of characteristic body language, voice qualities, verbal productions, and affective experience. The child represents the persistence of childlike experience and expression in all people. It is divided into the *natural child*, the ego state in which the spontaneity, joy, and intuitive perceptiveness of young children persists in all adults; the *adapted child*, the part that is compliant and cooperative; and the *rebellious child*, the repository of that part of each person prone to fight authority, challenge accepted wisdom, and struggle for autonomy. The parent is the residue of internalized parental messages and injunctions. It is divided into two parts: the *critical parent* and the *nurturing parent.* The *critical parent* bears some resemblance to Freud's superego, embodying rules, values, instruction, criticism, and restrictions. The *nurturing parent* is the internalization of a positive, caring experience, the memory of loving interactions.

Mental health is the flexible availability of all ego states with no one predominating. Excess critical parent produces guilt and depression, but insufficient critical parent produces sociopathy. Excess nurturing parent produces narcissistic laziness, whereas insufficient nurturing parent causes an inability to soothe oneself and to maintain self-esteem. Excess adult results in a cold, overly rational person, but insufficient adult leaves individuals unable to balance the various internal forces in their lives. Too much natural child may result in irresponsible behavior, but not enough depletes the ability to experience joy in living. An overabundance of rebellious child results in self-destructive battles with no constructive purpose; however, insufficient rebellious child may result in an overly conformist stance. Excess-adapted child prevents appropriate striving toward autonomy; insufficient adapted child prevents participation in group or hierarchical efforts.

Berne placed great emphasis on the child's ability to intuit parental messages and instructions, especially those communicated nonverbally and unconsciously. In this way, the growing child might encode a conscious verbal instruction in the critical parent (e.g., be strong, be perfect, be smart), whereas internalizing a more powerful, unconscious message in the child (don't grow up, don't leave me, don't surpass me).

Usually, a model for carrying out the instruction is preserved in the adult (be passive, drink alcohol, run around with women, act crazy). Together, these make a ***script***. Berne emphasized the active role of the child in searching for the messages and in accepting them. This was crucial for him because it emphasized the individual's responsibility in deciding to follow the script, although this might have occurred at a very young age. The entire basis for psychological change lies within the person's capacity, having once accepted the script, to later reject it. However, Berne was impressed by the intensity and persistence with which individuals play out the script throughout their lives. Scripts come in all varieties, ranging from the successful to the utterly self-destructive. Thus, for Berne, the business of human living involved carrying out one's script according to the transactions learned as a child.

Theory of Psychopathology

Given this view of human nature, Berne's understanding of psychopathology is based on the adaptiveness of a person's games and script and the capacity for adaptive use of all ego states. Written at a time when American psychiatry rejected phenomenologic diagnosis, Berne's theories are difficult to accept. However, he did describe numerous clinical situations and case histories that are familiar to all psychiatrists, and he analyzed them according to their scripts. He was very impressed with the role of fantasy and fairy tales in children's development and often asked patients what their favorite story was as a child. He then searched for the hidden identifications and messages within the story to help discern the patient's script. Insofar as he developed a nosology, it lay in the scripts encouraged by culturally sanctioned fairy tales. For example, Cinderella is about a young girl waiting to be rescued by a fairy godmother or a charming prince and unwilling to assert herself against unjust authority. Little Red Riding Hood is about a girl who likes to chat with wolves and gets into trouble trying to please everybody. Men, he found, often identified with the wolf or Prince Charming and were unable to regard women as real people.

A separate nosology of games was also developed and cataloged by Berne. Examples from this nosology include cocktail party flirtation (Rapo I), which concludes with the man and woman having experienced the stroke of feeling attractive and attracted, moving on to other conversations, and seeking other people to attract. Rapo II usually ends with a line like, “What kind of girl do you think I am?” and Rapo III ends up with a painful, destructive affair. Another game is “Why Don’t You … Yes, But.” In its mild, socially acceptable form, one person presents a problem, successfully enlisting the advice of many others; however, for every proffered solution, the person has a ready explanation of why it will not work. The stroke (or payoff) is the attention and the feeling of superiority achieved by defeating others. The second-degree version of this game produces a chronic, helpless depression; the third-degree version may lead to the backward of a state hospital.

There was no rigid mapping of his nosology of games and of scripts, and Berne never claimed that either was exhaustive. They were intended as clinical examples to guide the psychotherapist’s thinking in evaluating each patient who presented unique script and transactional issues. Individuals were assessed according to the straightness of their transactions (Were there hidden communication and invitations emanating from different ego states?), the adaptiveness of their scripts, and the predominant ego state and its effects. Thus, Berne might describe a compulsive person as having too much critical parent or a histrionic person as having too much natural child. In the final analysis, however, he saw his catalogs and nosology as poor substitutes for careful, clinical study of individual patients.

Treatment

Having been trained in psychoanalysis, Berne used many psychoanalytic techniques, but he applied them very differently from traditional psychoanalytic methods. Much of his clinical work was done in groups, with the therapist being very active in confronting and interpreting the interpersonal behavior of the patient. He emphasized the initial contract with the patient and the setting of clear, concrete goals. He encouraged therapists to inquire very early on in therapy, “How will we know when our work is finished?” If a clear and specific answer was not forthcoming, he would inform the patient that therapy had not yet begun. The focus would then be on the patient’s difficulty in defining a recognizable endpoint for the work. Berne’s approach was heavily informed by the psychoanalytic concept of resistance, but he applied it in a very different manner. Berne examined any treatment goal defined by patients for evidence that it was a way to continue playing games or pursuing scripts more effectively rather than giving them up entirely.

Like Sullivan, Berne regarded the ongoing interaction between patient and therapist as the key ingredient of psychotherapy. The therapist's job was to interact actively with the patient; recognize the ego states, games, and scripts being enacted; and counter them using confrontation, interpretation, or various other interpersonal maneuvers that were designed to thwart the enactment and to confront the patient with a choice and an opportunity to relate differently. He emphasized simple direct statements using everyday words to stimulate affective experience and interaction. Commonly, interactions began with his inquiry, "What do you want to work on today?" Observers sometimes described his work as individual therapy within a group context because the other patients, who were silent observers, often responded emotionally to the interaction.

Berne placed great emphasis on the role of individual responsibility for one's life and experience, which he believed had been obscured by the emphasis on psychic determinism. He constantly communicated the opportunity for choosing to continue or discontinue the patterns or habits encouraged as a child. He also showed patients how they invite others to behave in various ways, thereby creating their own interpersonal reality. Health lies in recognizing one's option for deciding which invitations to offer to others and which to accept from others, although these invitations may be communicated unconsciously and nonverbally.

Berne's View of Mr. A

Mr. A is clearly playing a very destructive game and living out a particularly unproductive script. The script seems to contain the messages: "Don't ever leave me" and "act crazy." The message "you must be special" may also exist. The games used might be called "I'm out of control," "Helpless," "Artiste," and "I'm better than you are." A transactional analyst places Mr. A in a group and works with him on these games and scripts. Many such therapists have productively combined Berne's theories with Fritz Perls's empty chair technique. Mr. A might be asked to carry on imaginary dialogues with his mother, professor, his illness, marijuana, and admired writers of fiction, with Mr. A assuming both sides of the dialogue. The transactions he uses are elucidated in this way, and he is invited to recreate the transaction more constructively. His internalization of his mother as an excessively critical parent is identified, and the fantasy dialogues are used to enable more internalization of both his mother's and father's nurturing qualities, which contain other script messages such as “grow up” or “be successful.” Similarly, his excess rebellious child is demonstrated, and efforts to bring it under more adult control are made.

FIGURE 6.3–13. Nancy Chodorow. (Courtesy of Nancy J. Chodorow, Ph.D.)

NANCY CHODOROW

Nancy Chodorow, born in New York, in 1944, is a sociologist, psychoanalyst, and educator (Fig. 6.3–13). She earned her PhD in Sociology from Brandeis University and trained in Psychoanalysis at the San Francisco Psychoanalytic Institute. Influenced, in part, by Karen Horney and Melanie Klein, Chodorow developed a theory of psychoanalytic feminism that incorporated traditional aspects of Freudian psychoanalysis, object relations theory, and feminist views, including a focus on the mother–infant relationship and the impact of gender on development. She is known as a leader in the feminist movement and describes herself as a feminist psychoanalyst. Her seminal works, including *The Reproduction of Mothering: Psychoanalysis and the Sociology of Gender*, have influenced contemporary feminist theory and fused traditional psychoanalytic views with modern-day developmental theory. This book was selected as one of the ten most influential books of the past 25 years by Contemporary Sociology.

Chodorow began her academic career teaching Women's Studies at Wellesley College in the mid-1970s. From there, she went on to join faculty in the Sociology Department at the University of California, Santa Cruz, from 1974 to 1986. Since 1986, she has been a professor in the Sociology Department at the University of California, Berkeley, and in 2005 was named Professor Emerita. Chodorow has won numerous awards and has written many books on feminist psychology and sociology.

Theory of Personality and Gender Identity

Nancy Chodorow's theory combined classical psychoanalysis, object relations, sociology, and feminism. Similar to Freudian psychoanalytic theory, her theory of personality development assumed the role of unconscious drives and dynamics and development of gender identity through relation to the parents, particularly the mother. Yet, Chodorow is an object relations theorist, and it is from this point of view that she proposed a theory of personality development and gender identity rooted in feminist ideas. Where traditional psychoanalysis suggested people are motivated primarily by drives, impulses, and the unconscious, object relations theory proposed people are motivated by seeking relationships with other people ("objects" in psychoanalytic terminology). Chodorow focused more on object relations and children's connection to their mothers than on unconscious impulses and conflicts. Her theory also diverged from classical Freudian ideas by focusing more on development during the preoedipal period than the oedipal phase, although she acknowledged the necessity and importance of resolution of the Oedipus complex. Chodorow emphasized attachments to objects were related to unconscious dynamics between mother and child, but she also acknowledged that children internalize normative gender roles and behaviors based on how and in what environments they are raised.

At the heart of Chodorow’s theory of personality are two primary questions: how do individuals develop gender identity? And why do women (and not men) develop a desire to mother? Psychoanalysis and feminist theory suggest individuals’ personalities are made, not born, and biology alone does not adequately explain the development of gender and personality. In her theory of development, Chodorow explored not only the child’s relationship to the mother but also the mother’s relationship to the child. She suggested that the presence, stability, and consistency of an attentive caretaker are more important than the exclusive bond and attachment between biologic mother and child. The development of personalities of boys and girls differs as a result of the mother’s varied approach to raising and ***mothering*** children based on the sex of the child. Mothers see daughters as like themselves and sons as different from them, resulting in a difference in approach to the mothering relationship with male and female children.

Young children—both girls and boys—have a closer bond with their female primary caretaker early in life and therefore begin to internalize the feminine characteristics of the mother. Girls identify with their mothers because they are similar and will later continue internalizing feminine traits. In contrast, boys will separate from the mother to gain autonomy and increased masculinity. Because boys are encouraged to attain masculinity through separation from the mother and her feminine attributes, masculinity is viewed as difficult to attain and therefore idealized and fantasized about for both male and female children. This valuation of masculinity as something to be sought, versus femininity as something to be left behind, contributes to and rests upon the generalized (and societal) evaluation of masculinity and, therefore, maleness or manhood as more valuable than femininity, femaleness, and womanhood. Successful development for female children involves growing in relation to others, where optimal development of male children lies in disconnection and separation from others, particularly the mother.

Through mothering a girl, the mother re-experiences herself as being mothered and feels like her daughter. This connection and closeness make gender identity for girls easy, innate, and more stable, as daughters develop a sense of sameness with their mother and have an immediate and accessible model of the female gender. Girls are less avoidant of revealing their vulnerability and dependence on others compared to males. Girls also develop a sense of connection, closeness, relatedness, and a sense of permeability of boundaries in personality development. Girls' personalities are anchored in relation to others and relationships, which makes them more stable than males.

For boys, gender identity is more difficult, as they begin to separate themselves as being unlike their mother. Boys cannot rely on connectedness and closeness with the mother to establish a sense of gender identity and begin to look for sources of masculinity from which to model, including male caretakers, other male authority figures, and even superheroes. As a result, boys develop personalities that are autonomous, independent, with firmer boundaries than girls, and with less emphasis on relation and connection. Boys must reject attachment and dependency on the mother to find and maintain masculinity, leading to greater challenges and complexities with male gender identification compared to that of females. Because male gender identity is based foundationally on not being something (feminine), it is more fragile and less stable, requiring reconstruction and affirmation along the way in development.

Unfortunately, because masculinity is psychologically constructed as opposing femininity, one way to affirm masculinity is to denounce or condemn femininity, which leads to the derogation and disparagement of girls and women. Boys are taught to be masculine rather than feminine, and a boy's masculinity is focused more consciously than a girl's femininity. Although masculinity may be foundationally less solidified, it is also more valuable and worthy of maintaining compared to femininity.

In sum, early gender differences are rooted in how girls are able to remain identified with their mother, while boys must reject this identification and establish a separate masculine identity. These initial differences are greatly reinforced throughout childhood by domestic and social gender roles. Such gender roles generally encourage girls to assume tasks of domestic activities and caretaking, whereas boys are encouraged to become independent with activities focused outside the home and family. As adults, men's and women's different roles within the family are supported by gender-stereotyped ideals, particularly that women mother and focus on activities within the home and that men are heads of the household and focus on activities outside of the home. Women are seen as being most effective in roles that require connectedness and relation to others, while men may be successful and effective in roles that do not require connection to others. This can also be interpreted as women’s focus and work being related more to emotional experiences than men’s, which perpetuates the expectation that women will be responsible for emotional labor in the home. Personality traits of men and women are oriented toward the differences in the division of domestic and vocational roles associated with each gender. As Chodorow noted, “The sexual division of labor and women’s responsibility for child care are linked to and generate male dominance.” This cycle is perpetuated and supported by gender roles learned early in childhood and by women’s continued mothering of children.

Parenting Roles. Chodorow’s clinical experience and the conventional wisdom at the time suggested that women, but not men, want to mother (provide nurturance and serve as primary caretaker). Chodorow asserted that women choose to mother and are not biologically destined or hormonally driven to do so. She suggested a woman’s anatomical and biologic makeup provide for why she can become pregnant, give birth, and nurse children, but does not account for why she should or would mother children. Research on hormonal differences between men and women does not support physiologic reasons for women’s role as the mother compared to men’s. Although hormonal changes occur that are directly related to pregnancy, parturition, and lactation, these do not explain the continued care provided to children after such hormones dissipate. Chodorow indicated girls and women are socialized to mother and encouraged through family dynamics, sexual division of labor in the home and work environments, and societal expectations to bear and care for children. She used the example of young girls being “wrapped in pink blankets and given baby dolls” and not being “allowed to get dirty” to highlight early socialization toward domestic activity, nurturing behavior, and motherhood.

Chodorow believed that women’s ability to bear and nurse children were not a sufficient explanation for why they then do most of the caring and mothering. She considered other sociologic and economic theories about why women mother but concluded that they do not completely explain why women want to mother. These theories included societal expectations to serve the role that evolution and biology intended, or individual motivation to become economically supported by serving the role of wife and mother and becoming attached to a financially successful man. Instead, Chodorow argued women get the desire to mother by having been mothered by women themselves. She challenged the notion that women are born with maternal instincts or motherliness. She suggested the biologic clock and maternal instincts were shaped by unconscious fantasies and a woman's internal relationship to her own mother and are not necessarily instinctual on their own. Chodorow proposed that becoming a mother should be and is a choice rather than the assumed fate of women and biologic females.

In her 1978 book *The Reproduction of Mothering*, Chodorow argued that mothers are the primary caretakers (not because of biology, but because of the absence of the father, who is presumed to be working and not expected to be involved in parenting). She theorized that if parenting roles were modified and men spent more time in fathering and parenting activities, boys and girls would get equal representation of both male and female parents in the home, and gender identity development would be reshaped, resulting in social change and change in perceptions of gender roles. The relationship between mother and infant forms the basis from which boys and girls develop relational abilities and connections to others. It is from this first relationship that children learn the capacity for potential parenting that may be utilized in adulthood, as, regardless of sex and gender, both women and men are mothered as children. The early relationship to the mother results in both boys and girls developing a capacity for and sense of ability to participate in parent–child relationships and to provide adequate parental care.

Chodorow's theory about mothering and the mother–child relationship emerged at a time of change in culture and the beginning of a change in society's view of women. With feminist beliefs emerging and gaining ground in the 1970s, Chodorow's theory followed from a growing awareness that the view of women's role as primarily mother and homemaker was oppressive. In her updated preface to the second edition of *The Reproduction of Mothering*, written in 1999, Chodorow noted that fathers' increasing involvement in rearing and parenting children over time may have been in response to cultural shifts in the view of men's role within the family, changes in perception of masculinity, a response to feminist views regarding women's role within families, or response of needs associated with families in which both parents work full time. Chodorow noted early on that men's greater involvement in parenting would likely have an impact on gender identity and the cycle of development, gender role expectations, and perpetuation of men's and women's perceived roles.

Psychopathology

Psychopathology can arise as a result of various factors, including the mother–child relationship. As toddlers, children began to experience incrementally more autonomy as a function of increased mobility and locomotion. The mother then becomes responsible for managing how and when children move about. As children develop a greater and more separated sense of self, mothers have to balance managing the child's choices with allowing them to make such choices and movements on their own. This transition period becomes challenging as the mother tries to avoid both overprotection (resulting in extended dependence on the mother) and neglect or deprivation (resulting in premature autonomy without sufficient decision-making abilities). Although Chodorow did not directly connect these early life experiences to later pathology, one could assume a potential impact on adult relational interactions based on such experiences with the mother in childhood.

Regarding the mother–son relationship, the increase in emotional distance between mother and son as the male child becomes more autonomous and seeks masculine attachment figures can go too far. This can have negative effects for the child, including becoming estranged from the ability to experience emotions and participate in intimacy (which often requires some amount of vulnerability). Boys may seek external interests, and men may seek multiple sexual partners to avoid the closeness associated with intimate relationships. In addition, boys and men may have difficulty accessing and expressing their emotions and may even become intolerant or denigrating of people, primarily women, who can express their emotional needs, such as the need for love.

For girls, experiences of a prolonged preoedipal phase can result in abnormal or pathologic psychic development ranging from neuroses to psychosis depending on the resolution of the oedipal phase and their ability to differentiate themselves as separate from the mother. Particularly if the mother has difficulty separating from the child, individuation may not occur, and girls' identity development becomes stifled, resulting in an overdependence on the mother. In adulthood, psychopathology for women can also result from conflict between their own motherhood and careers, other interests, or past life choices.

As mothers, women may seek psychological and emotional fulfillment from relationships with their children, particularly when these needs are not provided for by their partners, which may result in overinvestment in mothering. In such cases, mothers may have difficulty differentiating their lives from that of their children. Overattachment can have negative effects on the relationship between parent and child and the child's ability to separate and differentiate from their mother. Similarly, when children serve as a mother's only close relationship, expectations that emotional needs be met can create conflict and discord between mother and child. In addition, Chodorow presented two case studies of women who postponed or unconsciously sabotaged having children and later developed an extreme sense of loss when child-bearing years had passed, and motherhood was no longer an option.

Treatment

Chodorow is a sociologist and feminist, but her approach to treatment remains psychoanalytic. She emphasized the need for people to harness the power of their emotions and to recreate a life story that aligns the unconscious and conscious experiences. The foundation of treatment is relational, acknowledging the real connection between analyst and patient, as well as the influence of the past on the present. Chodorow encouraged recognition of each patient as unique and emphasized the development of a personal therapeutic relationship that allows for navigation of emotional and fantasied content that arises in treatment. The focus of treatment is on affect and the connection between fantasy and reality. According to Chodorow, psychoanalysis should offer an examination of meaning and the use of feeling to analyze how the unconscious creates stories and influences situations that arise in the present. Case examples presented by Chodorow indicate an emphasis on finding the source and reason for internal conflict related to the self and early relationships or attachment figures.

Chodorow's View of Mr. A

Mr. A is having significant difficulty transitioning from adolescent-like behavior to a more adult lifestyle. He does not take responsibility for his choices, including decisions to avoid treating his bipolar disorder appropriately. Although he fears returning to his hometown and living with his family again, he is not able to make the connection between his actions and the resulting consequences or does not have the ability to make the necessary changes required to avoid such outcomes. Mr. A's (and his siblings) relationship with his overattached and controlling mother likely developed early on between mother and child. His mother, who stayed home to raise the children when they were young, may have been challenged by allowing Mr. A and his siblings to gain autonomy and separation as infants and toddlers. Perhaps Mr. A's mother experienced her own conflicts related to being placed in the role of full-time mothering, especially given his father's absence. Because of a lack of autonomy, Mr. A was unable to individuate and formulate an independent sense of self. In addition, because his father was presumably less involved in childcare than his mother, Mr. A had only one model for the parent–child relationship. As Mr. A and his siblings grew up, their father's involvement in childcare was limited to placating the mother rather than parenting the children or encouraging or introducing change within the dynamics of the family, particularly the mother's relationship with her children. Once the children were grown and the mother returned to working outside the home, Mr. A was expected to begin acting like an adult (complete college, begin a career, and be able to support himself). However, his mother maintained control of his life not only through his finances but also through the management of his mental illness.

Mr. A's refusal to effectively treat his bipolar disorder could be seen as an inability to recognize his needs or an active rebellion against an overattached mother. Mr. A views himself as an outsider and perhaps is proud of his mental illness because it sets him apart (even makes him special) compared to others. His refusal to treat his disorder may be an attempt to separate himself physically and metaphorically from his mother and his family, particularly because no one else in the family has a mental illness. His increased and daily use of marijuana and efforts to maintain a state of hypomania may be attempts to control emotions by only allowing access to enjoyable or pleasurable emotions mimicked by both intoxication and elevated mood states. Chodorow's assertion that boys have greater difficulty experiencing and expressing emotions because of their early distancing from the female primary caretaker may be influencing Mr. A's use of marijuana and hypomania as methods of avoiding emotional states altogether. However, this may be in contrast with his overattachment with his mother. Regarding his dreams, Mr. A's flying dream could be a hope or desire to separate himself from his mother and his family, to create an individualized identity and sense of self. The dream about the hotel could represent dread associated with remaining connected to his family and feeling stifled by this connection.

In looking at the formulations and treatment plans, each of these theorists might propose for Mr. A, significant, even contradictory, interpretations become apparent. Commonalities exist but are expressed in different words and concepts. Each of these great thinkers has contributed to the overall understanding of the psychodynamics of psychopathology and treatment.

FURTHER READINGS

Alexander F, French TM, Pollack GH. *Psychosomatic Specificity*. University of Chicago Press; 1968.

Ansbacher HL, Ansbacher RR, eds. *The Individual Psychology of Alfred Adler: A Systematic Presentation in Selections from His Writings*. Basic Books; 1956.

Balsam R. Reminiscing and reflecting on Nancy Chodorow's the reproduction of mothering. In: Bueskens P, ed. *Nancy Chodorow and the Reproduction of Mothering Forty Years On*. Palgrave; 2021:87–110.

Berne E. *Games People Play*. Grove; 1964.

Berne E. *What Do You Say After You Say Hello? The Psychology of Human Destiny*. Bantam; 1972.

Bridges MR. Activating the corrective emotional experience. *J Clin Psychol*. 2006;62(5):551–568.

Carlson J, Watts RE, Maniacci M. *Adlerian Therapy: Theory and Practice*. American Psychological Association; 2006.

Chodorow NJ. Reflections on the reproduction of mothering: twenty years later. *Stud Gend Sex*. 2000;1(4):337–348.

Chodorow NJ. *The Reproduction of Mothering*, 2nd ed. University of California Press; 1999.

Chodorow NJ. "Too late": ambivalence about motherhood, choice, and time. *J Am Psychoanal Assoc*. 2003;51(4):1181–1198.

Christian C, Safran JD, Muran JC. The corrective emotional experience: a relational perspective and critique. In: Castonguay LG, Hill CE, eds. *Transformation in Psychotherapy: Corrective Experiences Across Cognitive-Behavioral, Humanistic, and Psychodynamic Approaches*. American Psychological Association; 2012:51–67.

Conci M. Erich Fromm, a rediscovered legacy. *Int Forum Psychoanal*. 2000;9:141.

Fisher H, Atzil-Slonim D, Bar-Kalifa E, Rafaeli E, Peri T. Emotional experience and alliance contribute to therapeutic change in psychodynamic therapy. *Psychotherapy (Chic).* 2016;53(1):105–116.

Fromm E. *Escape from Freedom*. Avon Books; 1965.

Gerson MJ. Psychosomatics and psychoanalytic theory: The psychology of ulcerative colitis and Crohn's disease. *Psychoanal Psychol*. 2002;19(2):380–388.

Graves' disease. Overview. Mayoclinic.org. Accessed June 14 2022. https://www.mayoclinic.org/diseases-conditions/graves-disease/symptoms-causes/syc-20356240

Hirsch I. Review of the interpersonal theory of psychiatry. *J Am Psychoanal Assoc*. 2004;52(1):257.

Horney K. *Neurosis and Human Growth*. Norton; 1950.

Jacobs TJ. The corrective emotional experience—its place in current technique. *Psychoanal Inq*. 1990;10:433.

Jung BD. *A Biography*. Little, Brown and Company; 2003.

Jung CG. *Two Essays on Analytic Psychology*. Princeton University Press; 1966.

Liu MY, Li N, Li WA, Khan H. Association between psychosocial stress and hypertension: a systematic review and meta-analysis. *Neurol Res*. 2017;39(6):573–580.

Markowitz JC, Weissman MM. Interpersonal psychotherapy: past, present, and future. *Clin Psychol Psychother*. 2012;19(2):99–105.

Martin K, Chodorow NJ. Fruitful dialectics in feminist thought. In: Stones R, ed. *Key Sociological Thinkers*. 3rd ed. Palgrave; 2017:288–299.

Meyer A. *Psychobiology: A Science of Man*. Charles C Thomas; 1957.

Mitchell S, Black M. *Freud and Beyond: A History of Modern Psychoanalytic Though*. HarperCollins; 1996.

Mosak HH. Adlerian psychotherapy. In: Corsini RJ, Wedding D, eds. *Current Psychotherapies*. 7th instr. ed. Thomson Brooks/Cole; 2005:52.

Mulahy P. *Psychoanalysis and Interpersonal Psychiatry: The Contributions of Harry Stack Sullivan*. Science House; 1970.

Overmier JB, Muriso R. Restoring psychology's role in peptic ulcer. *Appl Psychol Health Well Being*. 2013;5(1):5–27.

Paris BJ. *Karen Horney: A Psychoanalyst's Search for Self Understanding*. Yale University Press; 1994.

Reich W. *Character Analysis*. Fount, Stores & Young; 1949.

Schwetlik SN, Baldock KL, Hill CL, Ferrar K. Chronic stress and arthritis: a scoping review. *Arthritis Care Res (Hoboken)*. 2022;74(6):982–996.

Spruill TM. Chronic psychosocial stress and hypertension. *Curr Hypertens Rep*. 2010;12:10–16.

Stevens A. *On Jung*. Routledge; 1990.

Sullivan HS. *The Interpersonal Theory of Psychiatry*. W.W. Norton; 1953.

6.4 Approaches Derived from Philosophy and Psychology

PAUL T. COSTA, JR., PH.D., AND RALPH L. PIEDMONT, PH.D.

All the sciences trace their beginnings back to philosophy, but psychology and psychiatry have particularly strong links. Among the topics of most concern to ancient philosophers were the properties and limitations of the human mind: The exercise of reason, the influence of passion, and the mystery of madness. Philosophers sought to understand how the mind works and how it might be guided and strengthened. Psychiatrists today have much the same goals.

One of the functions of this section is to raise philosophic consciousness by hinting at some of the fundamental ideas concerning human nature that may challenge the implicit tenets of modern psychiatry. It is, of course, impossible to do justice to the full scope of philosophic thought in the space of this section. Four major schools of personality—psychoanalytic, behaviorist, humanistic, and trait—reflect four different philosophies of human nature and have very different implications for the field of psychiatry. Contemporary trait theories of personality reflect the recognition that all three of the other schools or approaches contain a kernel of truth. Are people driven by their impulses? Some are, some are not. Some are habit-bound, some are heroic; people vary in their emotional, interpersonal, experiential, attitudinal, and motivational tendencies. Trait psychology's emphasis on consistent differences among individuals can be seen as an alternative philosophy of human nature, one better suited to pluralistic and diverse societies.

The remainder of the chapter provides an overview of psychological theories grouped under the traditional rubrics of behavioral, humanistic, and trait approaches. Because of its importance in contemporary personality research, special attention is given to the trait perspective, especially the Five-Factor Model of Personality (FFM) and its accompanying-Factor Theory of the personality system.

PHILOSOPHY

The recognition of individual differences is probably as old as human culture, but differences in personality were long confounded with differences in status, class, or caste. Individuals from a higher status were assumed to be superior human beings, more sensitive, honorable, and wise. They were endowed with these characteristics by education, by bloodline, or—in the view of Indian philosophy—by moral behavior in a previous life.

Plato

The first major account of personality in Western thought was provided by Plato (circa 428–348 BCE). In The Republic, he made extended comparisons between the constitutions of different states and the constitution of the soul. Just as every state must have peasants, artisans, soldiers, and rulers, so too must each individual have appetites (for food, sex, and so on), passions (for honor and advancement), and reason. The relative strength of these three determines character, as well as fitness for a particular place in society. Intelligent and thoughtful people ought to rule, and passionate people should be chosen to defend the state, whereas dull and spiritless individuals, lacking reason and passion, should be given the menial chores of agriculture and industry.

Plato assumed that these psychological characteristics, like physical strength or musical talent, were largely inborn, but he regarded them as a property of the individual, not the individual's social status. Although he assumed that high-status citizens would normally bear children of the greatest potential, he specifically acknowledged that there would be exceptions, and, in his ideal state, children of the lower classes would be promoted, and those of the higher classes would be demoted on the basis of their own merit. Here, personality would be the basis of the social order, not vice versa. His most radical extension of this idea was to argue that women should be given social equality and allowed to become soldiers and rulers if they possessed the necessary mental and physical qualifications.

One of the recurring questions in personality theory has been the relative importance of nature versus nurture. Trait psychologists—particularly those interested in the study of temperament—have frequently pointed to innate differences in personality, whereas behaviorists and psychoanalysts have laid great stress on the formative influences of the environment and early childhood experiences. Ready as he was to acknowledge the importance of inborn potential, Plato was also keenly aware of the influence of education. Much of The Republic is devoted to his views on the effects of physical exercise, mental instruction, and poetry and music on the development of personality. Present-day concerns about the influences of video games and rap music on children follow this tradition.

Like most philosophers, Plato was more concerned with understanding virtue and vice than psychopathology, but, because he considered vice to be the result of weak or corrupted nature rather than free but evil choice, his discussions of flawed character can be viewed as early descriptions of what might now be viewed as personality disorders. Just as there are better and worse forms of government for states, so are there also better and worse configurations of reason, passion, and appetite. Plato described five types corresponding to five forms of government. The ideal of mental health, corresponding to government by wise rulers, is one in which reason holds in check passions and appetites. In the arrogant and ambitious type, there is an excess of passion or pride; in the avaricious type, there is an excess of appetite. However, in both of these types, reason still retains some authority; for example, avaricious individuals can control most of their appetites to indulge their desire for money. In the fourth, self-indulgent type, appetites are undisciplined, and, in the debauched, fifth type (corresponding to a government by despot), reason is completely disregarded, and a clearly psychopathological state is reached: “Thus, when nature or habit or both have combined the traits of drunkenness, lust, and lunacy, then you have the perfect specimen of the despotic man.”

Aristotle

Plato’s vivid but rough typology was succeeded by Aristotle’s (384–322 BCE) detailed analysis of human character in the Nicomachean Ethics. Courage, temperance, generosity, pride, ambition, irascibility, friendliness, boastfulness, and shame were all defined and distinguished. Aristotle attributed pathologic variations on these traits to innate defects or to disease processes: “Among all the excesses of foolishness, cowardice, intemperance, and irritability some are bestial, some diseased. If, for example, someone’s natural character makes him afraid of everything, even the noise of a mouse, he is a coward with a bestial sort of cowardice.” He considered variation within the normal range to be the result of training and habit and thus to be subject to praise or blame.

Aristotle’s basic moral precept is the golden mean: He argued that extreme standing on either end of a trait dimension should be avoided. Thus, stinginess and extravagance are vices, whereas generosity is a virtue; similarly, vanity and humility represent excessive or insufficient self-esteem. This conception continues to influence some modern notions of psychopathology, in which marked deviation from the norm in either direction is considered pathologic. Individuals excessively concerned with social attention may be regarded as having a histrionic personality disorder; those insufficiently concerned with social attachments may have a schizoid disorder.

Aristotle carried conceptual analysis to a level that has seldom been surpassed, distinguishing, for example, between the superficially similar qualities of temperance and self-control: Individuals are temperate if they have healthy and moderate impulses that require little control, whereas individuals have self-control only if they have immoderate appetites that are nevertheless held in check. Such considerations allowed him to form rational taxonomies of traits that anticipate the empirical taxonomies proposed by 20th-century factor analysts.

Shankara and Eastern Philosophy

Indian philosophy is a vast body of work that combines early science and psychology with Hindu theology and its offshoots (including Buddhism). Its foremost thinker (from the perspective of mainstream Hinduism) was Shankara (788–820 CE). In a brief life of 32 years, he restored the authority of the ancient Hindu scriptures, the Veda, after centuries of criticism by heterodox schools, and he articulated a worldview that has influenced Indian thought ever since. Shankara argued that the basis of existence was Brahman (the Absolute, Emerson's Oversoul), which pervades everything but is itself indivisible. The physical universe is not, for Shankara, an illusion but a manifestation of Brahman. What is an illusion is the idea that the individual is part of the material world, when in fact, the Self is nothing other than Brahman. Through study and virtuous actions, it is possible to come to understand this fact, leading to a state of enlightenment in which the individual becomes blissfully detached from life while continuing to participate in it.

Although the philosophical rationales vary immensely, much Eastern thought shares the core ideas and ideals of Shankara's system. In particular, Eastern thought emphasizes renunciation and asceticism, not as a way to mortify the flesh but as a way to liberate the soul. Themes of self-transcendence are also seen in the reverence for all life that is common to Hinduism, Buddhism, and Jainism (and that is increasingly seen in Western concerns with the environment).

Eastern philosophy was much admired by Carl Jung, and it had some influence on psychiatric thinking during the 1950s and 1960s. However, despite recent calls for multiculturalism in psychology, there appears to be relatively little interest in the topic today. Yet Eastern thought provides an entirely different perspective on what the goal of psychiatric treatment ought to be: Not palliation of distress or more effective functioning in social groups, but liberation from the illusions that create distress and dysfunction. This goal and the large body of techniques for achieving it developed over the centuries surely merit more serious attention.

Immanuel Kant and Arthur Schopenhauer

The last great period of the Western philosophical tradition, begun by Plato and Aristotle, was inaugurated at the end of the 18th century by Immanuel Kant (1724–1804), whose critical philosophy forms a transition between purely rational approaches to human nature and the empirical sciences that followed. One of Kant's last works, Anthropology from a Pragmatic Point of View, considered natural and moral variations in character and reintroduced the Roman physician Galen's taxonomy of choleric, phlegmatic, sanguine, and melancholic types to modern psychology.

Sciences are usually distinguished from philosophy by their reliance on empirical data, but it should not be imagined that philosophers made no use of experience. On the contrary, like many of the clinicians who offered psychological theories of personality, philosophers based their ideas heavily on their observations of human nature. A striking instance of this is provided by Arthur Schopenhauer (1788–1860), a follower of Kant, whose dark view of the world as a place of purposeless striving had an extraordinary influence on early personality theorists, including Freud and Jung.

One of the central beliefs of Western thought has been that human happiness or misery is the result of external conditions—what is now called quality of life. However, Schopenhauer's acute observations, reported in The World as Will and Representation, led him to propose "the paradoxical but not absurd hypothesis that in every individual the measure of pain essential to him has been determined once and for all by his nature…. His suffering and well-being would not be determined at all from without, but only by…what is called his temperament."

In support of this view, he noted that wealth and power do not make people happy, "for we come across at least as many cheerful faces among the poor as among the rich." He also argued that the effects of great misfortunes or successes are short-lived and that, for the most part, evaluations of the external causes of the state of mind are illusory attributions: “We often see our pain result only from a definite external [cause] and…believe that, if only this were removed, the greatest contentment would necessarily ensue. But this is a delusion…[The pain] would appear in the form of a hundred little annoyances and worries over things that we now entirely overlook.”

Recent scientific research on psychological well-being has confirmed this account in every detail: Well-being is chiefly a function of enduring personality dispositions; wealth, social class, and other markers of the objective quality of life are virtually unrelated to subjective happiness; and processes of adaptation quickly return individuals to their own characteristic baseline of happiness after favorable or unfavorable life events. Schopenhauer’s observations are also consistent with recent evidence on the heritability and lifelong stability of many mental disorders.

Yet, reason and insight are not enough. On the basis of his own experience and the examples of history, Schopenhauer also concluded that “man inherits his moral nature, his character, his inclinations, his heart from the father, but the degree, quality, and tendency of his intelligence from the mother”—a conclusion not supported by the findings of modern behavior genetics. Psychology broke from philosophy over precisely this need to seek empirical verification of hypotheses, but it carried with its concepts and insights accumulated over two millennia of profound thought about human nature.

Jacques Derrida and Postmodernism

Throughout much of the 20th century, philosophers tried to support empiricism by examining the conceptual foundations of the scientific method and inductive inference. Concepts such as operationalism, theory falsification, and construct validity are products of the philosophy of science that have been widely adopted in scientific practice. The most influential philosophical movement of recent years, however, has often been seen as antiscientific. Deconstructionism and its intellectual descendants have provided a radical critique of conventional scientific thinking that has had a marked impact, for good or ill, on the humanities and social sciences.

Jacques Derrida (1930–2004) was a French philosopher who is usually identified as the founder of deconstructionism. Through analyses of literature, history, and psychoanalysis, he questioned the basic tenets of

Western thought. He drew attention away from the ostensible subject of any writing, pointing to the writing itself as the object of investigation. Following this lead, social constructionists argued that there is no objective reality for science to study; instead, reality is constructed as a product of language and culture.

Derrida has had an enormous impact on the humanities, which is seen in the scholarly journals that abound with references to discourse, text, and narrative. In combination with other influences, such as feminism, social constructionism, and multiculturalism, deconstructionism has led to postmodernism, a stance that is critical of conventional empirical science on several accounts. Postmodernists argue that science is a set of cultural conventions revered in the West but is ultimately no more valid than any other belief system. The evidence that a Darwinian offers in favor of evolution is no more convincing than the evidence that a Fundamentalist offers in favor of creationism. Both are discourses that have their own justification. Postmodernists are also critical of the claim that science is value-neutral. They argue that the enterprise of science is inextricably bound to value judgments, and they have pointed out numerous ways in which science has contributed to the oppression of women, minorities, and members of non-Western cultures.

Many psychoanalysts, humanistic psychologists, and anthropologists have embraced postmodernism. Within psychiatry, one of the pioneers of postmodern thinking is Thomas Szasz, who argued that mental illnesses are not objective medical conditions but questionable value judgments. Kenneth Gergen has urged psychologists to adopt a postmodern perspective to liberate their science from orthodox methods and theories and to make psychology more relevant outside Western culture. Mainstream psychologists, like most psychiatrists, tend to view postmodernism skeptically; indeed, many reject it outright.

However, the criticisms and alternatives offered by postmodernists can be useful to empirical scientists, even if the radical premises are not accepted. Some of the qualitative methods they advocate, such as content analysis, can provide a valuable supplement to quantitative methods. It is useful to be reminded that the constructs of science, such as the categories of the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), are not direct representations of reality but rather imperfect maps always in need of revision. Therapists are well-advised to be sensitive to the cultural background of their patients, and, surely, all scientists need to be mindful of the social consequences of their work. This is particularly true in psychiatry, which has so much power over the lives and well-being of its patients.

BEHAVIORAL AND SOCIAL LEARNING APPROACHES

Theories of Personality

Behaviorism as a school of psychology grew up in reaction to the prevailing mentalistic model, in which introspection was used to determine the contents and operations of consciousness. John B. Watson (1878–1958) proposed that scientific psychology should confine itself to an examination of observable behavior and should explain all human conduct in terms of stimuli and learned responses. Ivan Petrovich Pavlov's (1849–1936) experiments with conditioned responses offered hope that such a science could be successful, and theorists such as Clark L. Hull (1884–1952) provided elaborate mathematical models of learning.

Radical Behaviorism. The most influential behaviorist and, perhaps, the most influential psychologist of the century was B. F. Skinner (1904–1990). Skinner's basic concept was operant conditioning, in which behaviors are viewed as a function of the organism's history of reinforcement. The observation that animals can be taught tricks by giving them rewards and punishments is nothing new. However, behaviorists, such as Skinner, refined and systematized this idea, using elegant experimental designs to tease apart the effects of the amount and schedule of reinforcements, the use of reinforcers and punishers, and the difficulty of the discriminations required. Behaviors could be shaped, maintained, or eliminated by the judicious use of these principles.

Skinner was a radical behaviorist, a purist who denied not only the scientific value but also even the existence of the mind. Furthermore, he avoided any neurophysiologic or psychophysiological theorizing, preferring to study the empty organism. Individual differences were ignored in understanding basic phenomena and were explained in individuals as the result of different histories of reinforcement. Even differences between species were neglected: Skinner believed that the pigeon provided an adequate model for the study of learning in all organisms, and he and his followers were, in fact, able to replicate many of their animal findings by using human subjects. Much of Skinner's success can be attributed to his single-minded pursuit of a highly circumscribed set of variables.

Skinner's view of personality was, predictably, a reductionistic one. As he stated in About Behaviorism, "a self or personality is at best a repertoire of behavior imparted by an organized set of contingencies. The behavior a young person acquires in the bosom of his family composes oneself; the behavior he acquires in, say, the armed services composes another." This position is rejected by humanistic psychologists, who attribute more choice and control to the individual, and by trait psychologists, who see consistencies of behavior that appear to transcend the consistencies of the reinforcing environment. Many personality psychologists have argued that controlled laboratory experimentation is a poor basis for theories of personality because individuals play a large role in selecting and shaping their own environments. Skinner's radical behaviorism was rejected or modified by many later learning theorists who acknowledged the power of conditioning but also recognized differences among species and among individuals within a species.

Social Learning Theory. One of the most distinctive features of human organisms is their use of speech, which makes possible elaborate thinking and planning and complex social interactions. In recent decades, learning theorists have increasingly emphasized social and cognitive processes. Among the most important have been Julian Rotter (1916–2014) and Albert Bandura (1925 to present), both of whom have offered versions of social learning theory.

Rotter's theory proposes that human behavior is guided not only by the actual history of reinforcement but also by plans, goals, and expectations of success. Individuals perform a behavior if they believe it is likely to lead to a valued goal, based on their past experiences in general and in similar situations. Individuals with a history of success are likely to have a generalized expectancy that they can control their lives; they are described as having an internal locus of control. At the opposite extreme are those whose prior efforts have been generally unsuccessful; they come to believe that rewards and punishments are a matter of luck or the arbitrary decisions of powerful others. Such individuals are said to have an external locus of control. Locus of control has been one of the most popular variables in personality research, used in numerous studies that generally are consistent with Rotter's theory.

Bandura's version of social learning theory also acknowledges the importance of internal cognitive processes. Individuals learn not only on the basis of their own experience but also through vicarious reinforcement from the observation of others. Bandura's demonstration of modeling effects in experiments conducted in the 1960s gave scientific legitimacy to the social learning perspective.

Rotter's and Bandura's theories are general theories of behavior, not specific theories of personality. However, for them, personality is something more than a collection of learned behaviors. The total pattern of experience leads to a generalized expectation of reinforcement or to a general sense of self-efficacy that can be considered the central individual difference variable. People are to be characterized primarily on the basis of their beliefs in their own ability to control their lives because these beliefs powerfully determine the effort they make to adapt to their surroundings.

Social-cognitive approaches to personality are among the most influential for current research in personality. These approaches focus on the individual's understanding of themselves and how these self-appraisals shape goals, plans, and behaviors. Because of their origins in social learning theory, they tend to emphasize the role of the environment, pointing out that an individual's sense of self varies from setting to setting. Because of their ties to social theory, they usually explain personality in terms of the effects of social interactions. For example, Hazel Markus (1949 to present) has suggested that individuals have a number of possible selves: Conceptions of what one is or could be, which result from the messages significant others provide. For such theorists, concern has moved beyond the social learning of specific behaviors to the learning of entire identities.

Theories of Psychopathology

Psychoanalytic theories of personality grew out of attempts to understand psychopathology, and there are thus intimate connections between the two. By contrast, behavioral approaches have focused on the general principles by which behavior is acquired and maintained, and psychopathology, where it is considered at all, is usually treated as an area of application. Learning theories have had much more influence on methods of psychotherapy than on theories of psychopathology itself.

Behavioral approaches might suggest two different classes of explanations for psychopathology: Psychopathology might be related to the mechanisms of learning themselves, or it might be considered the result of learned behaviors that are maladaptive or socially unacceptable.

In the 1950s, Hans Eysenck (1916–1997), a psychologist who has figured prominently in both learning and trait schools of personality, proposed that individual differences in the dimension of introversion versus extraversion determined the ease with which individuals could acquire conditioned responses, which in turn determined the form of psychopathology to which they were prone. Extremely introverted individuals, he proposed, were easily conditioned and thus acquired many inhibitions. They were predisposed to the development of depressive, anxious, and obsessive-compulsive disorders. By contrast, extreme extraverts were considered to be resistant to conditioning and were likely to develop hysterical and psychopathic disorders. (In later versions of Eysenck's theory, psychopathic disorders were grouped with psychotic disorders and linked to a different dimension of personality, psychoticism.)

Most behaviorists have viewed the laws of learning as universal processes and considered psychopathology to be the result of normal learning processes. In the 1920s, Irena Shenger-Krestovnika taught a dog to discriminate between a circle and an ellipse as a cue for food. When the ellipse was made increasingly circular, the dog's ability to discriminate between them was taxed, and the dog began to struggle, squeal, and bite. Pavlov dubbed this an experimental neurosis and proposed that human neuroses might have parallel causes.

Probably the most famous attempt to explain psychopathology in learning theory terms was provided by John Dollard (1900–1980) and Neal Miller (1909–2002). Dollard, an anthropologist, and Miller, an experimental psychologist, shared an interest in psychoanalysis. Their goal was to translate psychoanalytic concepts into the more testable terminology of learning theory. Consider, for example, the central psychoanalytic notion of repression. Parents might punish sexual behaviors in the child, and the child might learn to associate the behaviors with pain. By stimulus generalization, even the thought of the behaviors would elicit anxiety, and cognitive processes that blocked those thoughts would lessen anxiety and thus would be reinforced. Eventually, the thoughts would be effectively barred from consciousness.

Many behaviorists who did not share Dollard and Miller’s enthusiasm for psychoanalysis followed their lead in attempting to explain psychopathology in terms of principles of learning. Phobias, in particular, were easily explained as conditioned responses reinforced by avoidant behavior. Similarly, compulsions could be understood as a kind of self-reinforcing behavior: Each time the compulsive act is performed, the anxiety associated with not performing the act is reduced, increasing the probability that the behavior will be repeated.

Social learning theorists have also noted the self-perpetuating nature of some maladaptive behavior. Individuals who lack a strong sense of self-efficacy in social situations may avoid them. As a consequence, they fail to learn the social skills that would enhance their self-confidence. Self-defeating behaviors, which may appear irrational from an outside perspective, are often understandable in terms of the dynamics of learning.

Application of Theory to Therapy

Behaviorists have a rather rudimentary view of personality, seeing it as an assemblage of learned behaviors. They also tend to see psychopathology in superficial terms. Psychological maladjustment is considered to be the result of learned behaviors, which are called symptoms, but there is no underlying disorder of which they are symptomatic. Curing the symptoms cures the disorder. At worst, this position is naive and simplistic, equating the patient’s presenting problem with the real source of difficulty. At best, however, it focuses attention on a specific problem that can be concretely addressed.

A large number of techniques for behavior modification have been used with considerable success in treating symptoms of psychopathology. Joseph Wolpe (1915–1997) developed systematic desensitization as a treatment for phobias. Patients were instructed to relax and were then presented with increasingly vivid cues of the phobic object. Eventually, they were able to face the object itself without anxiety. A more dramatic technique is implosive therapy, in which the individual is confronted directly with the feared object (e.g., a room full of snakes) without an opportunity to escape. Because the object itself is harmless, and because avoidant behavior cannot be performed and is therefore not reinforced, the phobic reaction is swiftly extinguished.

Therapeutic interventions may be based on eliminating the reinforcements that sustain behavior, punishing unwanted behaviors, modeling or shaping more desirable behaviors, and so on. Any variable known to affect the acquisition or extinction of behaviors may provide an opportunity for behavior change, and behavioral techniques have been applied to physiologic responses, as well as voluntary behaviors, through techniques of biofeedback. Behavior therapies have been used extensively in treating phobias, controlling addictive behavior, reducing the self-destructive behavior of autistic children, and improving classroom discipline—for the behaviorist, the distinction between psychopathology and bad behavior is generally unimportant. Behavior therapies are most effective when the problem can be clearly traced to a particular set of behaviors or conditioned responses; they are much less effective in dealing with vague complaints of confusion and distress, although these are frequently the problems the patient presents to the clinician.

HUMANISTIC APPROACHES

Theories of Personality

Psychoanalysis and behaviorism are mechanistic theories that trace human behavior and experience to the gratification of instinctual impulses or to the acquisition of learned responses. Many of the most influential personality theorists have defined themselves in terms of their opposition to these two approaches. Although they vary widely in terms of the explanations they offer of personality, humanistic approaches share a positive evaluation of human nature and emphasize its unique and distinctively human aspects. Personality produces and reflects organization, rationality, consistency, future orientation, planfulness, self-expression, cognitive complexity, and adaptability. Most humanistic theorists prize human reason and freedom of will, the capacity for growth and change, the need for love, and self-transcendence. The positive psychology movement promoted by Martin Seligman (1942 to present) falls within this tradition.

Social learning theories might be seen as a humanized form of behaviorism because they recognize the role of complex symbolization and language in human learning. In a similar way, many now-classic humanistic theories were intended as modifications of psychoanalysis. Indeed, the first major psychoanalytic revisionist was Carl G. Jung (1875–1961), who argued that human beings had spiritual, as well as sexual, needs. Henry Murray (1893–1988) also made major modifications to psychoanalysis in his view of personality. For example, he credited the mature ego with much more autonomy than Freud granted it, and he argued that the individual's sense of morality was not fixed by the superego instilled in childhood but could continue to develop into more rational and altruistic forms. Erich Fromm (1900–1980) and Karen Horney (1885–1952) minimized the instinctual origin of personality development and suggested that culture played a large role in shaping the individual. Erik Erikson (1902–1994) proposed stages of psychosocial development to parallel Freud's psychosexual development, emphasizing such distinctly human characteristics as identity, intimacy, and generativity. He also theorized that personality development continued throughout the life span, giving encouragement to research on aging and personality.

Gordon Allport. Although he is usually classified as a trait psychologist, in his general orientation to theorizing, Gordon Allport (1897–1967) was clearly a humanist. For him, man's behavior is proactive, reflecting internal, self-initiating characteristics more than situational forces. In his view, human personality possesses psychological coherence and momentary (cross-sectional) and long-term (longitudinal) organization. Allport considered personality functioning to be characteristically rational, organized, and influenced by such conscious characteristics as long-range goals, plans of action, and philosophies of life.

Perhaps the most salient and controversial feature of his approach was the extreme emphasis that he put on the uniqueness of the individual personality. Allport viewed the major task of personology (or personality psychology) as the understanding and prediction of the individual case. Personal dispositions must be assessed to grasp the real personality, and this requires intensive study of an individual's past, present, and anticipated future functioning through the use of such techniques as the case history and content analysis of personal documents. In Becoming: Basic Considerations for a Psychology of Personality, Allport championed the view that concepts and laws must be developed to fit the individual case, creating the terms idiographic and morphogenic to symbolize his conviction that “each person is an idiom unto himself, an apparent violation of the syntax of the species.”

Allport was deeply concerned with identifying personality functions, which he discussed under the concept of the proprium, the superordinate concept in his system. Propriate functioning not only organized and integrated actions and experience but also provided the impetus to psychological growth. Allport described the functions of the proprium as a sense of body, self-identity, self-esteem, self-extension, rational coping, self-image, and propriate striving. These propriate functions were vital to personality, and, although they were ongoing, they were by no means considered unchanging. Allport theorized that the propriate functions are modified throughout life, predominantly in the direction of greater differentiation and integration or growth. The development of selfhood, away from the undifferentiated, opportunistic functioning of infancy and early childhood toward propriate functioning and striving, is part of human nature for Allport. The person guides or directs their life by attempting to fulfill their sense of self or proprium. Development continues into adulthood, with increasing signs of maturity and personal lifestyle.

Abraham Maslow. Abraham Maslow (1908–1970) interpreted personality in motivational terms. The individual’s whole life—their perceptions, values, strivings, and goals—is focused on the satisfaction of a set of needs, and the needs themselves are arranged in a universal hierarchy. Maslow’s needs serve an organizing and integrating role in life and are not to be understood as a simple and invariant set of responses to environmental pressures. They organize and create action possibilities and external reality.

At the lowest level of the motivational hierarchy are physiologic needs for food, water, sex, and sleep. The second level consists of safety needs, needs for protection, and security. The third level consists of needs for love and belongingness, and the fourth level consists of needs for self-respect and esteem from others. Above these lie the higher needs—for beauty, truth, justice, and self-actualization, the development of one’s full potential as a unique human being.

Individuals live at the lowest level of motivation which is problematic for them. That is, if needs for food and shelter are not routinely met, they become the overriding concern in life: The hungry individual risks danger and social ostracism to find food. Those who have always been well-fed, however, learn to take the satisfaction of physiologic needs for granted, and their attention is dominated by higher needs. Instead of examining specific behaviors and their reinforcements, Maslow's theory concerns the long-term satisfaction of broad classes of needs and thus gives a much broader depiction of the individual.

Maslow devoted much of his writing to a characterization of higher needs, including self-actualization, a drive to fulfill one's unique potential. He believed that personality psychology had become obsessed with psychopathology and that a corrective emphasis on positive mental health was needed. His biographical studies of such exemplary people as Eleanor Roosevelt and Abraham Lincoln suggested a number of distinctive characteristics of self-actualizers, including an accurate perception of reality, creativity, a need for privacy, and the frequent experience of mystical or peak experiences. As an exception to his general theory of motivation, he also noted that such individuals often skip the lower levels and proceed directly to self-actualization. The most creative artists and musicians never seemed to care about poverty or lack of social acceptance.

George Kelly. One of the most unconventional theories of personality was offered by George Kelly (1905–1967). It is, in some respects, a purely cognitive approach, but one with few ties to traditional learning theory. Instead of seeing them as organisms that are conditioned by their environment, Kelly argues that human beings should be seen as scientists trying to make sense of their world. In The Psychology of Personal Constructs, he states the fundamental postulate of his theory: "A person's processes are psychologically channelized by the ways in which he anticipates events."

The basic unit for understanding personality is the personal construct, a schema for classifying and interpreting experiences. For example, an individual might construe other people in terms of the contrast between strong versus weak. Each new acquaintance would be categorized as a strong or weak person, and subsequent interactions with this person would be guided by the original construal. In the course of experience, it would probably be necessary to reclassify some individuals initially thought to be strong as weak and vice versa; more importantly, some people might act in ways that were neither strong nor weak, leading the individual to develop new constructs (say, friendly vs. hostile) that were more useful in predicting other people's behavior.

This rather abstract and bloodless theory is made relevant to psychopathology by Kelly's ingenious reconstruals of some basic emotional reactions. He defines anxiety as the awareness that one's construct system is inadequate for construing important events. Guilt is the recognition that one's behavior is inconsistent with the ways in which one construes oneself. Hostility is viewed as the attempt to force experience to fit one's existing constructs. Such definitions are remote from common sense and clinical notions of anxiety, guilt, and hostility, but, precisely for that reason, they offer the prospect of novel ways of treating them.

Carl Rogers. Carl Rogers (1902–1987) is probably the most influential humanistic personality theorist. He articulated a formal theory of personality; pioneered a major school of therapy, client-centered therapy; and encouraged rigorous research on his theory and therapy. Rogers held that all organisms tend toward their own actualization—that mental health and personal growth are the natural condition of humankind. Psychopathology is a defensive distortion of this actualization process, and psychotherapy consists of creating conditions in which defense is unnecessary. Given these conditions, patients (or clients, as Rogers called them) essentially cure themselves.

Under ideal conditions, people's needs, desires, and goals emerge naturally as part of self-actualization and are recognized as part of the self; individuals are fully open to experience. In real life, however, one person's needs and desires often conflict with others' needs and desires; in particular, children find themselves in conflict with their parents, who withhold love when the child (from their perspective) misbehaves. Because love is so essential, children internalize these conditions of worth and believe that they are good and worthwhile individuals only when their self is consistent with the ideals imposed by significant others. To maintain their sense of worth, they may distort their experience; this leads to anxiety and self-defeating behavior.

In some respects, Rogers's theory is much like Freud's: Both see psychopathology as the result of defensive distortions and see the ideal state as one in which individuals can accept conflicts and deal with them rationally. Rogers is a humanistic theorist because he assumes that human nature is essentially good and that defenses are ultimately unnecessary. For Freud, the impulses of the id are eternally primitive and selfish, and their full actualization would be socially catastrophic.

Viktor Frankl. An Austrian neurologist and philosopher, Viktor Frankl's (1905–1997) distinctive view of human nature and psychopathology was profoundly shaped by his experience in Nazi concentration camps. There he came to the conclusion that even the most appalling circumstances could be endured if one found a way of making them meaningful. He described his experience in Man's Search for Meaning, a book that has been read by millions around the world.

Frankl was both a humanist and an existentialist. He believed that human beings shared with other animals somatic and psychological dimensions but that humans alone also had a spiritual dimension that confers both freedom and responsibility. People find meaning in their lives through creative and productive work, through an appreciation of the world and others, and by freely adopting positive attitudes even in the face of suffering. Those who fail to find meaning face alienation, despair, and existential neuroses. Traditional societies provided a framework of meaning in religion and shared cultural values; in modern society, people must find their own sources of meaning, and Frankl attributed many social problems, such as drug abuse and suicide, to their failures to do so.

Because of the spiritual dimension, human beings show self-transcendence and self-distancing. The former refers to the capacity to put other values (e.g., the well-being of a loved one) above self-interest. The latter is the ability to take an external perspective, as seen in a sense of humor. These capacities form the basis for therapeutic interventions in Frankl's version of psychotherapy known as logotherapy.

Dan McAdams. In the past three decades, a number of scholars in the humanities and social sciences have turned their attention to the life narrative as a focus of research. These writers argue that the consciousness of self that distinguishes human beings from other animals is not a static list of personal characteristics; it takes the form of a story. In telling their life stories, people explain themselves in the context of their history and their significant relationships. Life narratives are not objective life histories; they are subjective interpretations that give personal meaning to past, present, and future events.

The personality psychologist who is most closely associated with this perspective is Dan McAdams (1954 to present). McAdams postulated that personality could be understood on three levels: Level 1 consists of traits, abstract and enduring tendencies seen in general styles of action and experience. Level 2 is defined by personal concerns, the goals, plans, and strategies that preoccupy the individual at a certain time and place; later, McAdams adopted the term characteristic adaptations to describe a broader conception of this level. Level 3 is the life narrative, the story the person tells to themselves and to others in an effort to give a sense of unity and purpose to life. A common theme in America is the "redemptive self," in which the protagonist faces obstacles but eventually triumphs. Such a narrative appears to inspire a productive and caring life.

Psychiatrists and psychotherapists have, of course, been listening to patients' life stories for many decades, usually seeking clues to the origins of current problems. McAdams's approach is different; he wishes to understand the narratives as stories. Stories can be analyzed in terms of such features as narrative tone (e.g., tragic, comic, and ironic), imagery, theme (what goals the characters pursue), ideologic setting, and nuclear episodes (crises and turning points). From this perspective, key life events are not construed as causes of development but as symbols of identity. A patient's vivid recollection of a childhood humiliation may be important chiefly because it conveys the patient's sense of being a victim to whom fate has always been, and always will be, unfair.

Much of the scholarship on life narratives has been in a humanistic tradition that eschewed empirical rigor: a narrative reduced to a set of numbers is no longer a narrative. McAdams and his colleagues, however, have conducted psychometric studies on such basic issues as the inter-rater reliability and temporal stability of life story variables. These studies form the basis for a scientific evaluation of narrative perspectives on personality.

Theories of Psychopathology

In general, humanistic theories of personality stress positive aspects of human nature and discuss maladjustment in terms of failures and blocks to the full growth and development of the individual. It is of some interest to note that Rogers and Kelly formulated their theories in the context of counseling students—individuals who presumably had relatively minor maladjustments and considerable personality strengths. The applicability of humanistic theories to patients experiencing schizophrenia or dementia is certainly questionable, but the theories have had a profound effect on routine clinical practice, in which clearly diagnosable psychiatric disorders seldom account for all of the patient's problems in living.

Humanistic psychologists differ tremendously in their views on the origins of maladjustment. Rogers pointed to internalized conditions of worth acquired chiefly during childhood. Fromm, who was influenced by Marx, blamed society as a whole for instilling nonproductive orientations in individuals, such as hoarding or marketing orientations. Kelly said little about the origins of maladjustment but thought its essence was an ineffective construct system too rigid to be corrected by experience. Frankl believed some mental disorders were somatic or purely psychological in origin, but he also pointed to spiritual causes for the class of existential disorders.

Application of Theory to Therapy

Some humanistic theories of personality—for example, Allport's—have had little impact on psychotherapy; others, such as Rogers's, have been tremendously influential. It is helpful to recall that, for decades, the dominant form of therapy was psychoanalysis, a process that might require years and in which treatment was focused on dreams, childhood memories, and the ongoing relationship with the therapist (the transference) rather than on the immediate problems of the patient. Many of the standard techniques of contemporary counseling, clinical psychology, and psychiatry rest on a very different set of assumptions about human nature made scientifically respectable by the work of humanistic psychologists.

Brief psychotherapies often consist of opportunities for individuals to express their feelings and rethink their problems in a supportive atmosphere. The therapist may provide advice and guidance or at least may offer new ways in which patients can think about their problems. (Many psychiatrists use this general approach as an adjunct to medication, even in the treatment of serious mental disorders.) This process implicitly assumes that individuals, even those requiring psychotherapy, are basically rational and able, with some help, to solve their own problems; it also assumes that, given the right conditions, they move toward mental health: Patients are seen as scientists and self-actualizers.

Noting that Freud’s “psychoanalytic cure” was based on a conscious understanding of one’s life experiences, McAdams has argued that having a satisfying life narrative is a requirement for complete mental health. A good life story has coherence, credibility, a capacity for growth and change, and a generative orientation toward the world. One goal of psychotherapy, then, may be to help the patient rewrite their life narrative along these lines. Some (e.g., Sayer and colleagues) have claimed that writing about traumatic experiences has beneficial effects on mental and physical health, perhaps because the exercise of writing allows a rethinking of one’s life narrative.

Modifying a patient’s life story is a modest but, perhaps, realistic goal. Psychotherapists cannot easily alter patients’ personality traits, nor can they undo traumatic events of the past. They may, however, be able to help patients make sense of their dispositions and their life histories. Human tragedy can bring great suffering, but, in the hands of an artist, it can also be the source of great beauty. This was one of the major principles of Frankl’s logotherapy, in which patients were invited to examine their attitudes and reconstrue their suffering in ways that made its value apparent. A terminally ill woman who despaired that she would have to leave behind the children that she loved was asked if she would rather have been childless; that question allowed her to appreciate her life and legacy.

The most celebrated technique derived from logotherapy is the paradoxical intention, in which patients are invited to use their capacity for self-distancing to will what they fear or yield to their compulsions. A patient who stutters may be instructed to stutter as violently as possible; this reduces the pressure of trying not to stutter, and that, in turn, makes it possible to speak clearly. A number of behavioral techniques, such as implosive therapy, work on a similar principle of blocking a vicious circle of reinforcing undesirable behavior, but it is often difficult to convince patients to hazard the experiment. Frankl makes use of the characteristically human sense of humor to frame the paradoxical command in ways that motivate the patient.

The humanistic emphasis on freedom and responsibility has often clashed with the psychiatric tradition of regarding mental disorders as diseases. Labeling individuals as schizophrenics or phobics is held by some to be dehumanizing, and critics such as Thomas Szasz have argued that mental disorders are social and ethical judgments, not matters of medical fact. There is, of course, abundant evidence that some mental disorders have a biologic basis, but the criticisms of humanistic psychologists make the point that the disorder occurs in a human being who, in many respects, may be like any other person. There may be normal and abnormal behaviors, experiences, or relationships, but there are not normal and abnormal people.

TRAIT AND FACTOR MODELS

Individual differences are peripheral concerns in many social learning and humanistic theories of personality; they are the central focus of trait theories. The study of variations in human character and temperament goes back at least to Theophrastus (371–287 BCE), a Greek whose Characters depicted 30 different types. The morose type, for example, he described as follows:

A malignant temper sometimes vents itself chiefly in ferocity of language. The man whose tongue is thus at war with all the world, cannot reply to the simplest inquiry except by some such rejoinder as—“Trouble not me with your questions: Nor will he return a civil salutation…. He has no pardon for those who may unwittingly shove or jostle him, or tread upon his toe… He will neither wait for nor stay with anyone long: Nor will he sing, or recite verses, or dance in company. It is a man of this spirit who dares to live without offering supplications to heaven.”

Theories of Personality

The scientific study of individual differences in personality can be traced to Sir Francis Galton (1822–1911) in England, who laid the foundations of psychometrics, and to Gerard Heymans (1857–1930) in the Netherlands, who undertook the first large-scale study of rated personality traits. The first major trait theorist in the United States was Gordon Allport, whose 1937 volume, Personality: A Psychological Interpretation, spelled out the basic issues in trait psychology. He defined a trait as “a neuropsychic structure having the capacity to render many stimuli functionally equivalent, and to initiate and guide equivalent (meaningfully consistent) forms of adaptive and expressive behavior.” In this view, something in the brain of the morose man makes him see even simple questions or greetings as a personal affront, and his sullen attitude is expressed in a variety of social situations.

Allport believed that traits were concrete features of individuals that uniquely described them and that might be understood by a case study of a single individual. A contrasting view is that traits are dimensions of individual difference that can only be discovered by comparing and contrasting different individuals; individuals are then described in terms of their standing on a set of common traits. The two definitions are closely related; people who are more anxious than 99% of the population have, presumably, a neuropsychic structure that makes them so anxious.

Characteristics of Traits. Although there are many different trait theories, there is general agreement on several key features of traits: (1) Traits are tendencies to show consistent patterns of thoughts, feelings, and actions. Behaviors that are specific to a single setting or situation may better be considered habits than traits; some evidence of cross-situational consistency is necessary to infer a trait. However, concrete instances of behavior have many determinants, including learned habits, aroused needs, social contexts, role requirements, and the influence of many different and potentially conflicting traits, so the influence of a specific trait on any particular behavior may be quite modest. It is usually only by viewing behavior across many different situations that a consistent pattern can be detected. (2) Traits are relatively enduring features that characterize the individual. In this respect, they are to be distinguished from transient moods or episodes of mental disorders that affect the individual. The fact that traits are relatively enduring does not mean that they cannot change; traits are not immutable, even if they are durable. (3) Traits are continuously distributed, usually approximating a normal or bell curve. Although it is convenient to speak about introverts and extroverts, in fact, most individuals are ambiverts, showing some of the characteristics of introverts and some of the characteristics of extroverts. There is no consistent evidence of discrete personality types.

From time to time, there has been controversy over the reality of traits, fueled by the fact that human beings easily and readily ascribe traits to others on the basis of little or no information, with correspondingly limited accuracy. These personality ascriptions may be triggered by stereotypes of age or physical appearance or may be quite idiosyncratic. Demonstrations of this fact in laboratory experiments by social psychologists, together with the relatively loose cross-situational consistency of most traits, led a generation of psychologists in the 1970s to conclude that traits were cognitive fictions. Subsequent work—particularly, demonstrations that judges who knew the individual well agreed with one another and with the self-reports of the individual about their standing on a variety of traits—restored faith in the consensual validity of traits. However, the controversy does make the crucial point that some trait ascriptions are more accurate than others and that first impression may be quite misleading. Psychiatrists ought not to assume that their clinical judgments of a patient's personality are correct; validated personality questionnaires and rating forms completed by knowledgeable others may be needed to portray and understand personality accurately.

Personality Structure and Factor Analysis. The most important differences among trait theorists are in the specific traits that they have conceptualized and measured. Jung identified introversion and extraversion as basic personality variables, Bandura emphasized self-efficacy, and Rogers was concerned with openness to experience. Over the years, literally, thousands of scales have been developed to measure traits that psychologists considered important in understanding personality. As early as 1936, Allport pointed to another source for identifying personality traits: The natural language. In a monograph he published with Henry Odbert, he listed some 18,000 terms extracted from an unabridged dictionary that could be used to describe people; some of them were mere evaluations (e.g., swell and awful), but he regarded approximately 4,000 as legitimate trait terms.

The problem for trait psychologists was how to choose a manageable set of traits from among the many possible constructs. It was obvious that trait terms were highly redundant—for example, anxious, worrying, nervous, apprehensive, and fearful reflect similar, if not identical, characteristics—so what was needed was a procedure for identifying major groups of traits that covaried. Factor analysis, a statistical technique that reduces the complexity of a set of correlations among variables, was first used in personality research by J. P. Guilford (1897–1987) and has remained one of its basic tools. The factors or dimensions identified in this process correspond to groups of closely related traits; the set of basic dimensions identified by the factor analysis constitutes a model of the structure of personality traits.

Raymond Cattell (1905–1998) developed one of the first and most influential factor models. He reasoned that, in the course of cultural evolution, any personality trait important in human social interaction would have been noticed and named; the 4,000 trait terms identified by Allport and Odbert could thus be assumed to represent an exhaustive listing of personality characteristics (this has become known as the lexical hypothesis). Cattell grouped synonyms and near-synonyms together to obtain a set of 35 personality variables and asked respondents to rate acquaintances on each of these sets of terms. He intercorrelated the ratings and factored the correlations, identifying 12 factors. Together with four more factors found in research using self-report questionnaires, these became the basis for the 16 Personality Factor Questionnaire (16PF), a self-report instrument that has been widely used in personality research and clinical psychology for more than 50 years.

It was originally hoped that factor analysis would provide an objective solution to the question of personality structure, but for many years there was little agreement among factor analysts. Eysenck believed that Cattell's model could not be replicated and was needlessly complex. He proposed a simple and powerful two-dimensional model that identified extraversion–introversion (E) and neuroticism–emotional stability (N) as superfactors and showed that, if the scales of the 16PF were themselves factored, the two largest factors resembled his E and N. He and his wife, Sybil Eysenck (1927 to present), developed a series of instruments to measure these factors (and, later, a third superfactor called psychoticism) that has also been widely used, particularly in Great Britain. Eysenck's stature as a learning theorist and a critic of psychoanalysis contributed to the importance of these dimensions in psychiatric contexts.

Five-Factor Model. Eysenck's two factors were widely replicated, but they seemed to omit many important characteristics, such as curiosity, aggression, and achievement striving. An alternative solution was offered in 1961 by two U.S. Air Force psychologists, Ernest Tupes and Raymond Christal. They began with the 35 clusters of traits identified by Cattell and obtained ratings on these clusters in eight different samples. They found that a five-factor solution fit the data in all eight samples. Two of the factors resembled Eysenck's E and N, but three other factors were new.

A small group of lexical researchers, including Warren Norman (1930–1998) and Lewis R. Goldberg (1931 to present), continued to study personality structure as represented by natural language trait adjectives, and, after 20 years, they came to the conclusion that the five-factor structure proposed by Tupes and Christal was essentially correct. Renewed interest in this model showed that the five factors could be recovered in analyses of self-reports and observer ratings; in ratings of children, college students, and older adults; and in several different languages, including non–Indo-European languages, such as Chinese, Hebrew, and Filipino. The factors appeared in analyses of trait adjectives, descriptive phrases, and questionnaire scales. Contemporary five-factor researchers differ somewhat in their conceptualizations of the factors and consequently give them somewhat different labels. The terms neuroticism (N), extraversion (E), openness to experience (O), agreeableness (A), and conscientiousness (C) are used here. The five factors and some of the traits, or facets, that define them are given in Table 6.4–1, along with associated adjectives.

Many psychologists were skeptical of the lexical hypothesis that had led to the FFM. These critics believed that personality theory and clinical experience would lead to the identification of important traits for which no lay terms existed, and they continued to offer alternative models. Katharine Briggs (1875–1968) and Isabel Myers (1897–1980) operationalized Jung's psychological functions in the Myers–Briggs Type Indicator, which classifies individuals in terms of the dichotomies of introversion versus extraversion, intuition versus sensing, thinking versus feeling, and judging versus perceiving. Some (e.g., Jerry S. Wiggins, 1931–2006) have argued that the traits that influence social interactions were better represented in a circular order than as a set of factors, and many instruments have been developed to measure this model.

A particularly important system was suggested by Theodore Millon (1928–2014), who was interested in personality traits associated with psychiatric disorders. His reviews of clinical literature led to a theory of personality and psychopathology that specified 11 personality disorders as extreme variants of normal traits. For example, the histrionic personality disorder is supposed to be related to the trait of gregariousness; the schizoid personality disorder is related to the trait of detachment. Millon's theory had a profound impact on the formulation of Axis II in the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III) and in subsequent editions. His instrument, the Millon Clinical Multiaxial Inventory, has been widely used by psychiatrists and clinical psychologists.

The fact that trait theories of personality are usually tied to assessment inventories makes it relatively easy to make empirical comparisons, and, in the last decade, researchers in several countries have undertaken the task of relating different trait models. There is a growing consensus among these researchers that virtually all the traits measured by theory-based personality questionnaires—including those derived from psychodynamic, behavioral, and humanistic theories—are related to one or more of the five factors of Tupes and Christal. However, instruments vary in comprehensiveness. For example, the four scales of the Myers–Briggs Type Indicator correspond to four of the five factors, whereas measures of the interpersonal circumplex represent only two factors (extraversion and agreeableness).

These comparisons of instruments can lead to significant reconceptualizations. For example, Cloninger's Temperament and Character Inventory (TCI) consists of four factors that are intended to assess temperament and three that assess character. However, joint analysis with a measure of the FFM shows that the Harm Avoidance temperament scales and the Self-Directedness character scales are really opposite poles of a single neuroticism factor, whereas the Reward Dependence temperament scales and the Cooperation character scales are both measures of agreeableness. From the perspective of the FFM, the Temperament and Character Inventory's distinction between temperament and character appears not to be supported.

At the broadest level, then, the problem of personality structure appears to have been resolved: Personality is described by five basic factors. This does not mean that personality can be exhaustively described by standing on five dimensions. Most trait psychologists adopt hierarchical models of trait structure. They assume that the broadest factors are composed of more specific traits, which, in turn, are defined by subtraits that correspond to individual items in a questionnaire. In the NEO Personality Inventory 3 (NEO-PI-3), for example, six specific traits or facets are measured for each of the five factors (or domains) of personality. The facet scales are listed in Table 6.4–1; the hierarchical organization is illustrated in Figure 6.4–1.

Assessment on the level of specific facets provides a more detailed and personalized portrait of the individual.

Origin and Development of Personality Traits. Psychoanalytic, learning and humanistic theories have usually offered causal explanations for individual differences. Thus, differences in self-efficacy are supposed to be caused by different histories of success in pursuing goals; variations in openness to experience might be the result of differences in the conditions of worth imposed by parents. The factor analysts who scoured the dictionary for trait terms usually bypassed this issue, being content to describe the personality differences they found in adults. In principle, some traits might be inherited, some might be instilled by parents, and some might be learned from experience; whatever their origin, they could be measured and used to predict important life criteria.

Until recently, it was generally assumed that personality was shaped primarily by a variety of environmental influences, including parental love and discipline, social and economic opportunities, and life experiences through childhood and adolescence. Surprisingly, this assumption has rarely been tested: There has been little prospective longitudinal research documenting links between early childhood experiences and subsequent adult personality. A different research design, using the techniques of behavior genetics, can answer these questions by comparing personality measures in adults with different degrees of genetic and environmental similarity. For example, the similarity between identical twins reared apart can be attributed to genetic influences, whereas the similarity between adopted siblings reared together must be due to shared environmental influences.

Since the early 1980s, behavior genetics studies using many different samples, personality measures, and methods of data analysis have converged on the surprising conclusion that personality traits are, to a considerable extent, heritable, and, to a considerable extent, due to unknown and idiosyncratic causes. The molecular genetics underlying trait heritability has been a topic of great interest in the past few years, but research approaches are still being refined. Early studies claiming to find single genes for major traits are giving way to newer paradigms. These more sophisticated approaches use combinations of traits to define personality styles or patterns that better define the personality phenotypes for genetic linkage and association studies.

Less success has been found in identifying the environmental origins of personality traits. In fact, they are hardly attributable to shared environmental causes at all. Socioeconomic status, family diet, religious training, parental modeling, and all the other influences that children growing up in the same household would normally share seem to have little or no influence on adult personality. This conclusion applies as much to character traits, such as achievement striving and modesty, as to temperament traits, such as anxiety and energy level. This dramatic and counterintuitive finding will doubtless reshape theories of personality. As Sandra Scarr (1936 to present) commented, “psychology [currently] has no adequate theories to account for individual variation in behavior because our theories address the wrong sources of variation.”

Table 6.4–1.

Five Factors of Personality, Defining Traits, and Empirically Associated Adjectives

Factor	Trait	Adjectives
Neuroticism (N)	Anxiety	Anxious, fearful, worrying, tense
	Angry hostility	Irritable, impatient, moody, not gentle
	Depression	Pessimistic, worrying, not contented, moody
	Self-consciousness	Shy, not self-confident, timid, inhibited
	Impulsiveness	Hasty, self-centered, excitable, loud
	Vulnerability	Not confident, not efficient, not clear-thinking, anxious
Extraversion (E)	Warmth	Friendly, warm, sociable, not aloof
	Gregariousness	Sociable, outgoing, talkative, not withdrawn
	Assertiveness	Assertive, forceful, aggressive, confident
	Activity	Energetic, hurried, quick, active
	Excitement seeking	Pleasure-seeking, adventurous, daring, spunky
	Positive emotions	Enthusiastic, humorous, optimistic, jolly
Openness (O)	Fantasy	Dreamy, imaginative, artistic, complicated
	Aesthetics	Artistic, original, inventive, idealistic
	Feelings	Excitable, spontaneous, affectionate, insightful
	Actions	Wide interests, versatile, adventurous, imaginative

	Ideas	Curious, original, insightful, inventive
	Values	Not conservative, not cautious, flirtatious, unconventional
Agreeableness (A)	Trust	Trusting, not suspicious, forgiving, not wary
	Straightforwardness	Not shrewd, not autocratic, not charming, not demanding
	Altruism	Soft-hearted, gentle, generous, kind
	Compliance	Not stubborn, not demanding, not headstrong, not impatient
	Modesty	Not a show-off, not clever, not argumentative, not self-confident
	Tender-mindedness	Sympathetic, soft-hearted, warm, kind
Conscientiousness (C)	Competence	Efficient, thorough, resourceful, intelligent
	Order	Organized, precise, methodical, thorough
	Dutifulness	Thorough, not careless, not distractible, not lazy
	Achievement striving	Ambitious, industrious, enterprising, persistent
	Self-discipline	Energetic, not lazy, organized, not absent-minded
	Deliberation	Not hasty, not impulsive, not careless, not immature

Adjectives are significantly correlated with traits, *N* = 305; *P* < .001.

Adapted from McCrae RR, Costa PT Jr. Discriminant validity of NEO-PI-R facet scales. *Ed Psychol Meas*. 1992;52(1):229. doi: 10.1177/001316449205200.

One new theory that does begin to address the “right” sources of variation is the five-factor theory. This theory attempts to provide a conceptual framework in which the body of research on the FFM can be interpreted. The five-factor theory is a systems theory, represented schematically in Figure 6.4–2. The central components are basic tendencies (including personality traits and intelligence, and other abilities) and characteristic adaptations (such as skills, attitudes, values, and roles). In this model, as suggested by behavioral genetics, the only input to basic tendencies is from biologic bases. However, personality traits and external influences are important in shaping characteristic adaptations. For example, an individual who is temperamentally high in self-consciousness and who repeatedly experienced ridicule as a child might develop an avoidant personality disorder, which is a characteristic (mal)adaptation.

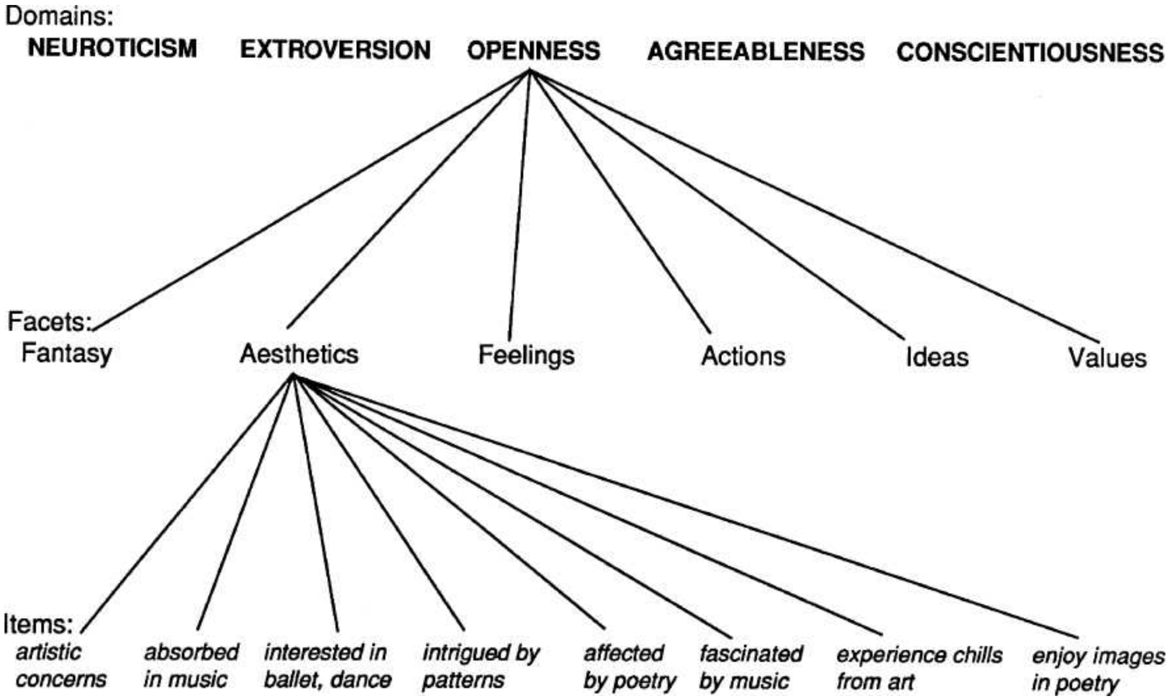
Domains:
NEUROTICISM
EXTROVERSION
OPENNESS
AGREEABLENESS
CONSCIENTIOUSNESS
Facets:
Fantasy
Aesthetics
Feelings
Actions
Ideas
Values
Items:
artistic concerns
absorbed in music
interested in ballet, dance
intrigued by patterns
affected by poetry
fascinated by music
experience chills from art
enjoy images in poetry

FIGURE 6.4–1. An example of hierarchical organization in the Revised NEO Personality Inventory: domains, facets of openness, and items measuring openness to aesthetics.

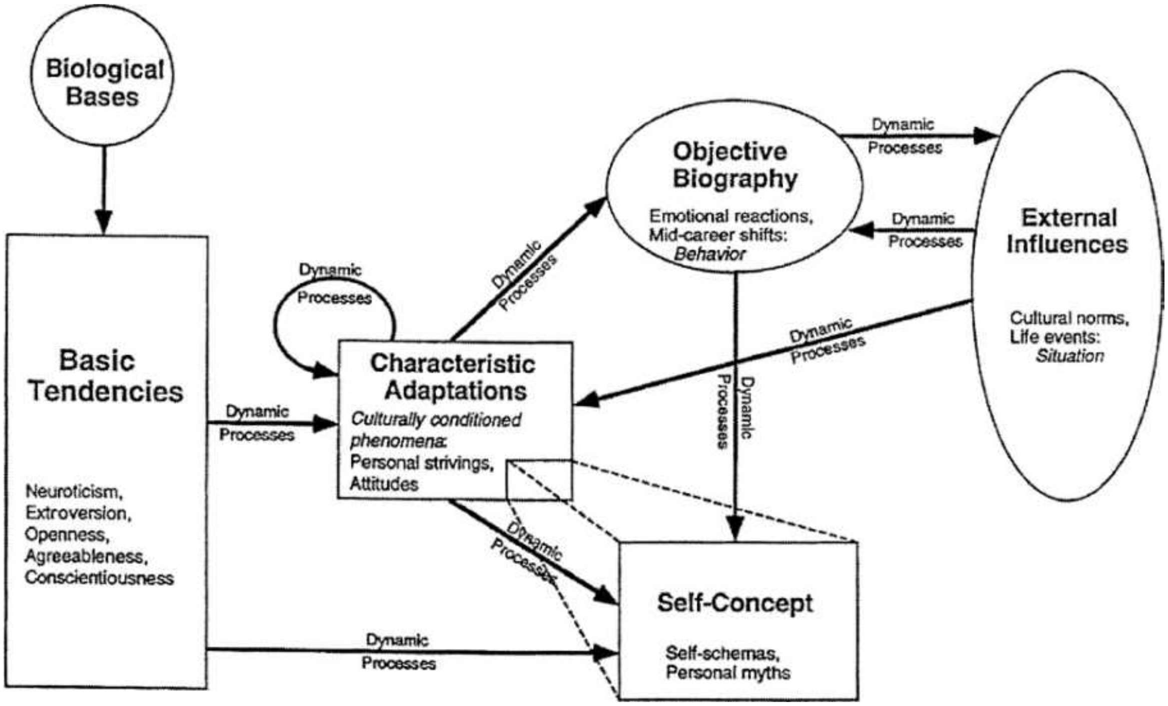
Biological Bases
Basic Tendencies
Neuroticism, Extroversion, Openness, Agreeableness, Conscientiousness
Dynamic Processes
Characteristic Adaptations
Culturally conditioned phenomena: Personal strivings, Attitudes
Objective Biography
Emotional reactions, Mid-career shifts: Behavior
External Influences
Cultural norms, Life events: Situation
Self-Concept
Self-schemas, Personal myths

FIGURE 6.4–2. A representation of the personality system as depicted by the five-factor theory. The ellipses represent peripheral elements, and the rectangles represent central elements. Examples of content are given in each. Arrows indicate the direction of causal influences. (Reprinted with permission of Guilford Publications. McCrae RR, Costa PT Jr. A five-factor theory of personality. In: Pervin LA, John OP, eds. *Handbook of Personality: Theory and Research*. 2nd ed. 1999:139. Adapted with permission from Costa PT Jr, McCrae RR. Set like plaster? Evidence for the stability of adult personality. In: Heatherton TF, Weinberger JL, eds. *Can Personality Change?* American Psychological Association; 1994:24. Copyright © 1994 by American Psychological Association.)

For many years developmental psychologists assumed that personality development ended with adolescence. A 21-year-old is legally an adult and is unlikely to grow much in physical height or intellectual capacity, but recent research has shown consistent changes in personality between college age and middle adulthood. There is a dramatic decline in excitement seeking over that interval, but, in addition, cross-sectional and longitudinal studies of American college students have shown that of the five-factor traits, N, E, and O decline during the decade of the 20s, whereas A and C increase. As a result, 30-year-olds are less emotional and better socialized than 20-year-olds.

There is reason to believe that these changes are not the product of the relatively indulgent American experience of adolescence. The same pattern of age differences is seen in less-affluent countries, such as Croatia, and in non-Western countries, such as South Korea. Such cross-cultural studies suggest that there may be intrinsic maturational processes in personality development, which is consistent with the biologic basis of personality traits hypothesized by the five-factor theory.

There have also been important developments in understanding personality change in individuals older than 30 years of age. Traits are defined as relatively enduring patterns of behavior, but most psychologists assumed that traits would be modified by life experiences. Jung postulated that the process of individuation required each individual to express all of their potentials, and, thus, the young introvert would normally become an extravert in old age, and vice versa. Lay stereotypes of aging held that, as they age, people become cranky, conservative, and depressed. Gerontologists rebutted these myths of aging, arguing that old age was more likely to bring maturity, wisdom, or detachment. Theories of adult development, popularized in Gail Sheehy's best-selling Passages, suggested that personality changed in stages and, in particular, that men (and perhaps women) went through a midlife crisis in their 30s or 40s.

In sharp contrast to all these theories are the findings from a number of independent longitudinal studies of personality in adulthood. These studies present a clear picture of predominant stability of the full range of personality traits. Although there are very gradual changes in the mean levels of traits across adulthood, individuals tend to maintain the same rank order. The 30-year-old, who is outgoing, curious, and hardworking, is likely to become an outgoing, curious, and hardworking 80-year-old. This conclusion has been supported by studies using self-reports and observer ratings of all five factors of personality and appears to apply to men and women. Of course, there are often dramatic changes in personality as a result of dementing disorders in old age, but normal life experience seems to have little impact on personality in individuals older than 30 years of age.

There are distressingly little data on the long-term effects of psychopathology on personality. Individuals in recovery from an episode of major depressive disorder generally score high on measures of the five-factor trait N, and it is sometimes argued that this is a result of the experience of depression. However, the few prospective studies of the first onset of major depression typically show that these individuals scored high on N before the first episode, suggesting that high N may be an enduring feature of people prone to clinical depression.

Theories of Psychopathology

The links between psychopathology and trait psychology are at once intimate and complex. For decades, psychiatrists and clinical psychologists have administered psychological tests such as the Minnesota Multiphasic Personality Inventory (MMPI) and Cattell's measure of personality, the 16PF, to inform the diagnosis of mental disorders; these same tests have been used routinely by trait psychologists to examine relations between personality and such criteria as vocational preference, creative potential, and coping with life stress. The dividing line between normal variations in personality dispositions and psychopathology is frequently unclear.

This poses more of a problem to psychiatric nosology than it does to trait theories of personality. Psychiatric diagnoses are supposed to represent the presence of a discrete disorder, and the DSM-5 specifies the criteria that must be met to confer each diagnosis. This categorical model is appropriate for some disorders but not others. It may be difficult to separate social phobia, for example, from extreme shyness and self-consciousness or a schizoid personality disorder from marked introversion. This is entirely consistent with trait models of personality, which regard individual differences as continuously distributed variables.

Neuroticism and Psychopathology. One theory of psychopathology, therefore, is that some psychiatric disorders reflect extreme standing on personality traits, particularly those related to the five-factor trait N. Among the traits that covary in normal populations to define this factor are predispositions to experience chronic levels of negative affects, such as fear, anger, shame, and sadness. Individuals with high standing on these traits may qualify for a diagnosis of generalized anxiety disorder, borderline personality disorder, social phobia, or dysthymia.

The five-factor trait N may also be considered a risk factor for the development of psychiatric disorders that are not themselves trait-like. Several recent studies have demonstrated that individuals scoring high on measures of N are at increased risk of subsequently receiving a diagnosis of major depressive disorder. Poor control of urges and impulses and excessive concern with physical functioning are also characteristics associated with N, and it might be hypothesized that individuals high in N are predisposed to develop eating disorders and hypochondriasis. N is, in essence, a generalized disposition to experience psychological distress, and individuals seeking psychiatric treatment are almost always distressed. It is therefore not surprising that virtually all clinical populations, from drug abusers to schizophrenics, score high on measures of N.

Historically, the term neurosis or psychoneurosis was coined to identify psychiatric disorders of functional origin that were closely related to anxiety and its control. It survived in the revised third edition of DSM (DSM-III-R) in many secondary labels (dysthymia, e.g., was also called depressive neurosis) but is not used in subsequent editions. Many personality psychologists object to the term neuroticism as a label for a dimension of normal personality because of its suggestion of psychopathology. Many psychiatrists object to the term because it appears tied to an outdated psychiatric nosology. However, the term serves a useful purpose: It is a reminder to clinicians that patients with a wide variety of diagnoses share many features related to chronic psychological distress and is also a reminder to personality psychologists that the difference between normal and abnormal functioning is often only one of degree.

Maladaptive Styles. Shapiro’s (1999) classic book on neurotic styles offered an extremely useful perspective for understanding the differences between normal and abnormal personalities. The clinical value of Shapiro’s approach is its integrative perspective on functioning that recognizes a multiplicity of factors, both adaptive and maladaptive, that operate in peoples’ lives to produce a distinctive style of psychological functioning that provides direction and consistency in behavior. Schapiro’s approach is insightful and rich in clinical detail. While the FFM and FFT were not around when Shapiro developed his ideas, the FFM can nonetheless provide a useful interpretive paradigm for understanding how the expression of various personality trait combinations contribute to maladaptive styles of functioning and impairments in various spheres of life. The NEO Style Graphs can provide insight into some of the ways people act in different aspects of life. These circumplexical style graphs are formed by the intersection of the two factors and are divided into four quadrants (e.g., +/ –; +/ +; –/+; and –/–). There are ten combinations of the five broad personality factors that describe an individual’s styles of Well Being, Defense, Anger Control, Impulse Control, Interests, Interactions, Activity Attitudes, Learning, and Character. Relevant to Shapiro’s neurotic styles, we will consider only the four maladaptive styles of Well-Being (N+ by E–) Defense (N+ by O–), Anger Control (N + by A–), Impulse Control (N+ by C–).

The high N and Low E combination represents the “Gloomy Pessimist” aspect of the Style of Well-Being. Such individuals, because of their high distress proneness and low positive affect or energy, face a dark and dreary life with few things that cheer them and many that cause anguish and distress. Stressful circumstances often lead them to feeling depressed. Even under normal circumstances, they often find life hard and joyless. These are individuals who feel no positive affect and believe that life holds little value or happiness for them. Such individuals may experience a number of existential crises, including a diminished sense of personal worth and value, a lack of personal meaning, and a belief that the future holds no solutions or opportunities for improvement. Further compounding their situation is the lack of a useful social support network to provide emotional care. These individuals are frequently loners who are detached from their environments and are prone to experience deep loneliness. These individuals should not be confused with schizoid personality disorder; such individuals may be aloof and detached from social interactions, but that is their preferred style. The high N and low E person are isolated and detached not because of preference, but rather because they are overwhelmed with feelings of inadequacy and emotional distress and are fearful of social criticism. They may also lack the social tools and skills necessary to effect any positive coping with their situation.

The high N and Low O combination defines the “Maladaptive” type of defense style. These individuals tend to use primitive and ineffective defenses such as repression, denial, and reaction formation. They prefer not to think about disturbing ideas, and they may refuse to acknowledge possible dangers (such as a serious illness). They lack insight into the distressing affects they experience, and because they cannot verbalize their feelings, they may be considered alexithymic. The low O may also create a more authoritarian orientation toward life, believing that there are clear rules and guidelines that define and determine appropriate intra- and interpersonal processes. Securely settled within their own acquired belief structures, such individuals may be unable to adjust or adapt their behaviors in ways that facilitate their engagement with others. Thus, they may easily put off others and develop feelings of alienation and isolation.

Style of anger control considers levels of N and Agreeableness (A). While Extraversion examines our interest in being with others and our level of engagement with them, A relates to the quality of our interactions with others, ranging from a very compassionate, caring, and altruistic orientation to a more antagonistic, selfish, and manipulative style. The presence of high N serves to attenuate the communal, bonding aspects of the highly agreeable person while exacerbating the hostile, confrontational aspects of the antagonistic individual. The style involving high levels of N with low levels of A is termed “Temperamental.” These individuals are easily angered and tend to express anger directly. They may fly into a rage over a minor irritant, and they can seethe with anger for long periods of time. They are deeply involved in themselves and take offense readily, and they often overlook the effects of their anger on others. They may be prone to physical aggression and verbal abuse. Because these individuals may possess feelings of personal inadequacy and have a tenuous hold on their life tasks, they are keenly sensitive to any criticism or negative evaluation of their behavior, seeing them as threats that may expose them as impostors or worse. They respond quickly and harshly to any accusations of personal worth and are willing to physically strike out against any external agency.

The fourth and final maladaptive style relates to the role of Conscientiousness in moderating the expression of negative affect influencing self-regulation, which is the impulse control style. High levels of C are associated with good self-control, including active processes of planning, organizing, and carrying out tasks. Low levels of C are related to a more “here and now” orientation, a desire to take the path of least resistance, and a goal to maximize hedonic pleasure. For those who are low on C, their manner of impulse control is labeled “Undercontrolled.” These individuals are often at the mercy of their own impulses. They find it difficult and distressing to resist any urge or desire, and they lack the self-control to hold their urges in check. As a result, they may act in ways that they know are not in their long-term best interests. They may be particularly susceptible to substance abuse and other health risk behaviors. These individuals are very hedonistically oriented, preferring immediate pleasure to long-range success. These individuals frequently end up disappointing others in their lives because of their inability to live up to their commitments and obligations. Further complicating their relationships with others is underlying selfishness to their strivings. Others are valued and useful primarily to the extent that they can provide for whatever physical and/or emotional needs they may have. Undercontrolled style people are not necessarily manipulative in their manner of relating to others so much as they are interested in fulfilling their personal algorithm of obtaining maximum benefit for minimum effort or investment.

The above style descriptions represent only a fraction of available possibilities. McCrae and Costa (2010) provide more detailed descriptions for every quadrant in the 10 Style Graphs.

Personality Traits and Personality Disorders. DSM-5 provides a general definition of a personality disorder that applies to all 10 of the specific disorders: “A personality disorder is an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual’s culture, is pervasive and inflexible, has an onset in adolescence and early adulthood, is stable in time, and leads to distress or impairment.” (p. 645)

It is reasonable to ask whether these inflexible and maladaptive traits are the same as or different from those encountered in nonpsychiatric populations. Several recent studies on this question have concurred in finding strong and replicable links between scales measuring personality disorders and the five factors in normal and clinical populations.

In the case of N, only high scores are associated with psychiatric impairment, but both poles of the other factors appear to be associated with specific forms of psychopathology. Individuals high in E tend to have histrionic and narcissistic personality disorders; those who are low in E have avoidant and schizoid disorders. High C is associated with obsessive-compulsive personality disorder; low C is associated with antisocial personality disorder. Low A, or antagonism, is characteristic of individuals with paranoid, antisocial, and narcissistic personality disorders. High A is associated with dependent personality disorder.

FFM Descriptions of Personality Disorders. Thomas Widiger and colleagues have linked the personality disorders with FFM traits. In one study, the diagnostic criteria and text descriptions of each personality disorder were coded in terms of the 60 poles of the 30 facets of the FFM (see Table 6.4–1). In other studies, personality disorder researchers and practicing clinicians were asked to describe a prototypic case of each of the ten personality disorders using the negative and positive poles of the 30 facets of the FFM. Table 6.4–2 presents an integrative summary of these FFM descriptions.

These associations show some of the ways in which disorders can be conceptually distinguished. For example, avoidant and schizoid personality disorders are both characterized by low gregariousness, but only the former shows associations with high anxiety, depression, self-consciousness, and vulnerability. Research in several countries has generally supported this mapping of disorders onto personality traits, although the conceptual distinctions highlighted in Table 6.4–2 are usually blurred in real data because of the pervasive comorbidity of personality disorders.

It is a matter of current controversy just how these personality traits are related to the disorders: Are normal personality traits carried to an extreme inherently maladaptive, or do these traits merely predispose individuals to develop disorders under certain circumstances? Some research has also called into question the meaningfulness of the syndromes recognized by DSM-5 by showing that the symptoms used to define the disorders, when separately assessed, do not covary in ways that match the DSM-5 syndromes. Instead, the symptom clusters that do emerge are interpretable in terms of the FFM. As a result of such evidence, proposals have been made to replace the current categorical model with dimensional models that relate problems in adjustment to standing on basic dimensions of personality. This would appear to be a logical extension of trait theories of psychopathology.

Revisions in the DSM-5. More than previous editions, the DSM-5 highlights the importance of individual differences, particularly in personality functioning and personality traits, stating that “the DSM like other medical disease classifications should accommodate ways to introduce dimensional approaches to mental disorders including dimensions that cut across current categories” (p. 5) but further “The DSM-5 Task Force recognized that it is premature scientifically to propose alternative definitions for most disorders.” (p. 13)

After a 12-year process (including research conferences, task force meetings, drafts, and revisions posted on a public website, etc.), the Personality and Personality Disorders Work Group presented an alternative dimensional model, which was strongly endorsed by the APA DSM-5 Task Force. This alternative model is similar to the dimensional FFM of general personality structure. “These five broad domains are maladaptive variants of the five domains of the extensively validated and replicated model known as the “Big Five” or “Five-Factor Model of Personality (FFM)” …” (p. 773).

However, the American Psychiatric Association Board of Trustees overruled the DSM-5 Task Force and retained all 10 DSM-IV-TR personality disorder categories, placing the Alternative Model in Section III—Emerging Measures and Models, hoping that such a move would allow additional research and give clinicians time to become better acclimated to this dimensional approach to diagnosis.

An alternative way to revise the diagnosis of personality disorders has also been offered by Thomas Widiger and colleagues, who have proposed a four-step process. In the first step, domains and facets of personality would be assessed, providing all the benefits of trait assessment described in the following discussion. In the second step, clinicians would identify specific problems in living associated with salient traits. For example, it might be noted that a patient high in self-consciousness experiences intense embarrassment when around strangers. In the third step, the clinician would evaluate the degree of impairment that the problems caused to determine if they are clinically significant. Does the patient’s embarrassment interfere with holding a job or finding a mate? Finally, in an optional fourth step, the clinician could interpret the full personality profile to see if it matched the prototype for an identified personality disorder. For example, if the patient was high in anxiety, depression, and vulnerability, as well as self-consciousness, and low in gregariousness, assertiveness, and excitement seeking, they could be considered to have an avoidant personality pattern (Table 6.4–2). A major advantage of this system is that patients who did not meet full criteria for any recognized personality disorder could still be treated for personality-related problems in living.

Table 6.4–2.
Hypothesized Relations between DSM-5 Personality Disorders and Five-Factor Model Personality Traits

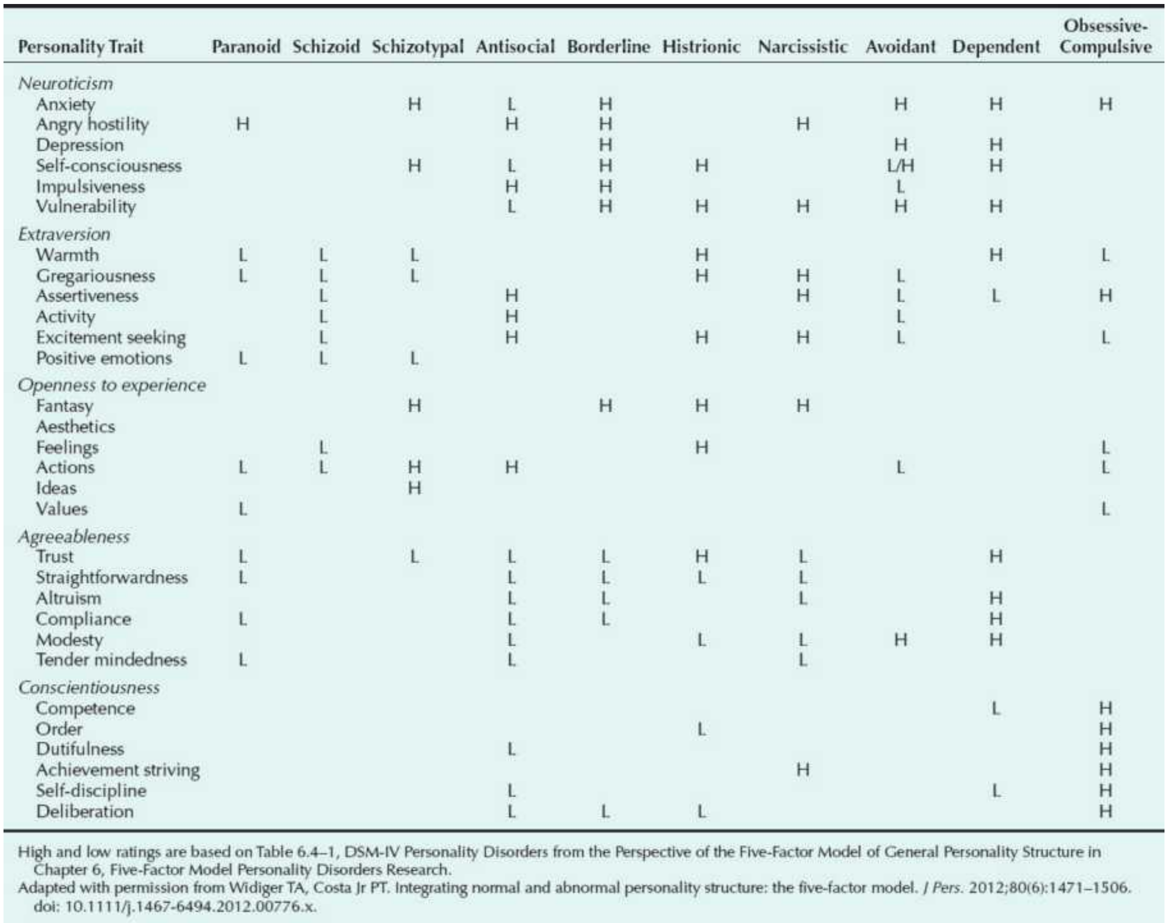

Personality Trait	Paranoid	Schizoid	Schizotypal	Antisocial	Borderline	Histrionic	Narcissistic	Avoidant	Dependent	Obsessive-Compulsive
Neuroticism										
Anxiety			H	L	H			H	H	H
Angry hostility	H			H	H		H			
Depression					H			H	H	
Self-consciousness			H	L	H	H		L/H	H	
Impulsiveness				H	H			L		
Vulnerability				L	H	H	H	H	H	
Extraversion										
Warmth	L	L	L			H			H	L
Gregariousness	L	L	L			H	H	L		
Assertiveness		L		H			H	L	L	H
Activity		L		H				L		
Excitement seeking		L		H		H	H	L		L
Positive emotions	L	L	L							
Openness to experience										
Fantasy			H		H	H	H			
Aesthetics										
Feelings		L				H				L
Actions	L	L	H	H				L		L
Ideas			H							
Values	L									L
Agreeableness										
Trust	L		L	L	L	H	L		H	
Straightforwardness	L			L	L	L	L			
Altruism				L	L		L		H	
Compliance	L			L	L				H	
Modesty				L		L	L	H	H	
Tender mindedness	L			L			L			
Conscientiousness										
Competence									L	H
Order						L				H
Dutifulness				L						H
Achievement striving							H			H
Self-discipline				L					L	H
Deliberation				L	L	L				H

High and low ratings are based on Table 6.4–1, DSM-IV Personality Disorders from the Perspective of the Five-Factor Model of General Personality Structure in Chapter 6, Five-Factor Model Personality Disorders Research.

Adapted with permission from Widiger TA, Costa Jr PT. Integrating normal and abnormal personality structure: the five-factor model. *J Pers.* 2012;80(6):1471–1506. doi: 10.1111/j.1467-6494.2012.00776.x.

Application of Theory to Therapy

Psychodynamic, behavioral, and humanistic theories all specify the causal mechanisms by which psychopathology is created and maintained and thus imply points of intervention. If maladjustment is caused by rigidly internalized conditions of worth, then unconditional positive regard in the therapeutic setting may provide a cure; if the problem is a learned behavior, then altering reinforcements may extinguish it. Trait models do not emphasize causal or developmental explanations and thus may seem to have no implications for psychotherapy. In fact, they have many.

Recent evidence on the substantial heritability of most personality traits clearly indicates that Allport was right: Traits do have some underlying neuropsychic structure. Research on the psychophysiology of traits is a topic of growing interest and parallels the extensive work done on the neurophysiologic basis of many forms of psychopathology. In this broad sense, then, psychopharmacologic approaches to psychotherapy are, in principle, consistent with trait theories of personality: If personality traits and disorders reflect brain processes, drugs that affect the brain may offer useful instruments for psychotherapy. In practice, there is still much to learn. For example, individuals with dysthymia, who score high on measures of trait depression, often do not respond to antidepressant medication.

Personality Assessment. Historically, the chief role of trait psychology in psychotherapy has been in diagnosis and assessment. A number of self-report measures, such as the MMPI and the Millon Clinical Multiaxial Inventory, were designed specifically to measure psychopathology and include primarily items that tap psychiatric symptoms. They can be regarded as measuring psychopathologic traits and states (although they have often been used in college and volunteer samples as measures of personality per se). These instruments are chiefly of value as aids to psychodiagnosis. General personality questionnaires, including the 16PF, the Guilford-Zimmerman Temperament Survey, and the Edwards Personal Preference Schedule, have also been used for decades in clinical psychology and psychiatry as part of a complete psychological assessment. Several measures of the FFM have recently appeared, including the Big Five Inventory and the NEO Personality Inventory, and there has been considerable interest in the clinical application of the FFM. The primary advantage of this model is its comprehensiveness. By assessing traits from all five dimensions, the clinician can efficiently obtain a full portrait of the individual.

Some psychiatrists are concerned that personality scores obtained from acutely depressed patients are invalid. This concern is based on the empirical finding that scores, particularly on scales related to the five-factor trait N, change with remission and on the belief that personality traits cannot change. If the latter belief were true, it would be necessary to conclude that depression distorts self-reports of personality. However, as Figure 6.4–2 suggests, personality traits may change if their biologic basis is altered, and acute major depression has demonstrated effects on brain functioning. Controlled studies and clinical experience suggest that trait scores obtained from acutely depressed patients are indeed informative in most cases.

Personality psychologists and psychometricians have devoted years to the development of self-report inventories, and the usefulness of this approach to assessment is beyond doubt. Self-reports, however, are by no means infallible. Patients may not understand their own personalities or may deliberately misrepresent themselves. Concerns about defensiveness and socially desirable responses have led to the use of projective tests, the development of special validity scales to detect and correct distorted responses, and reliance on the clinical judgment of the psychiatrist. Each of these possible solutions introduces problems of its own, however, and most clinicians rely on multiple sources of information.

One source of information that has been underused in clinical settings is informant ratings. Research in the past decade has confirmed that ratings on standardized instruments by significant others—usually spouses or parents—can provide reliable and valid assessments of personality. These findings appear to hold for clinical, as well as volunteer, samples. Personality ratings may be particularly valuable when patients are incapacitated or are strongly motivated to present an overly favorable or unfavorable picture of themselves. A case example of the clinical interpretation of informant ratings is given below.

Uses of Trait Profiles. Traits are enduring and stable features of the patient's behavior and experience, so assessing a broad range of traits gives the clinician a sense of what the person is like, which can be useful for many purposes beyond the formulation of a diagnosis. A complete personality profile can point to the patient's strengths as well as weaknesses, can help predict the course of therapy, and can aid in the selection of an optimal mode of treatment.

Humanistic theories of personality stress human potentials for altruism, creativity, and commitment and argue that psychotherapy must use these assets. Trait theorists would qualify that stance: They believe that some people are altruistic, creative, or committed, but others are not. Assessing traits allows the clinician to identify and capitalize on the particular strengths that characterize each patient.

Standing on trait dimensions can also give clues to the probable course of therapy. Patients who are low on the five-factor trait A are distrustful and uncooperative, and it may prove difficult to form a therapeutic alliance with them; by contrast, patients who are extremely high on A may be excessively compliant and become dependent on the therapist. The five-factor trait C involves commitment and self-discipline. Patients high in C will probably adhere to treatment recommendations and work hard to solve their problems; those low in C are less persistent and dedicated, and the therapist may have to work to motivate them.

Finally, a few controlled studies and a good deal of clinical experience suggest that personality traits influence the effectiveness of various kinds of treatment interventions. Interpersonal therapies, in which patients are required to speak a great deal about themselves, appear to be most effective for extroverts; pharmacologic management may be superior for introverts. Similarly, individuals who are high in the five-factor trait O benefit from such techniques as guided imagery, whereas those who are low in O prefer biofeedback. Ideally, the choice of therapies should be guided not only by the nature of the disorder but also by the enduring characteristics of the patient.

An in-depth case study incorporating trait data with other assessment perspectives is provided in Jerry Wiggins's volume on Paradigms of Personality Assessment. Madeline G was a flamboyant Native American woman. She had overcome a history of early delinquency to become a lawyer who fought for the rights of her people. She had no children, and her common-law husband left her a year after her personality assessment. This led to a period of depression, and 3 years after her assessment, she still remained alone. She provided self-reports on the NEO-PI-R, and her husband rated her. They agreed that she was extraverted, open to experience, and very low in agreeableness but disagreed on neuroticism and conscientiousness. Figure 6.4–3 plots her husband's ratings of her.

This is a dramatic profile, with a number of extreme scores. From her husband's perspective, she is high in neuroticism, especially anger, hostility, and impulsiveness. Although he sees her as high in most facets of extraversion, he rates her as very low in warmth. Similarly, he describes her as being uniformly low in facets of agreeableness. He acknowledges that she is high in achievement striving but claims she is very low in dutifulness and deliberation.

The four-step process of personality disorder assessment predicts that she would be vulnerable to a specific set of problems in living. Her high angry hostility means that she is likely hypersensitive to criticism and prone to react with rage. Her low warmth suggests that she may have difficulty developing or sustaining personal, intimate relationships. Low agreeableness can lead to rudeness that alienates friends and trouble with the law (as seen in her adolescent delinquency). Because she is very low in self-discipline, she might be expected to have difficulty holding a job, but in fact, she is successful in her career. Although personality traits suggest the possibility of certain problems in living, the clinician must determine whether or not those problems are present in a particular case.

The clinician must also judge whether the problems are severe enough to warrant a diagnosis and treatment. Madeline G did not seek treatment, but both her depression and inability to develop new intimate relationships after her husband left her might have justified a psychiatric intervention.

Whether Madeline G met the criteria for the diagnosis of one of the ten DSM-5 personality disorders cannot be ascertained simply by looking at her personality profile. However, she shares personality traits with patients who have paranoid, schizotypal, antisocial, borderline, histrionic, and narcissistic personality disorders. Conversely, she is unlikely, according to her husband's description, to have schizoid, avoidant, or dependent personality disorders.

Madeline G's personality profile suggests that if she were in psychotherapy, she would find it difficult to form a therapeutic alliance (low agreeableness) and might have some problems making a sustained effort at the work of therapy (low conscientiousness). But she does have a number of strengths, including high assertiveness and achievement striving. Because she is quite open to experience, she would be willing to consider interpretations and suggestions for change made by her therapist.

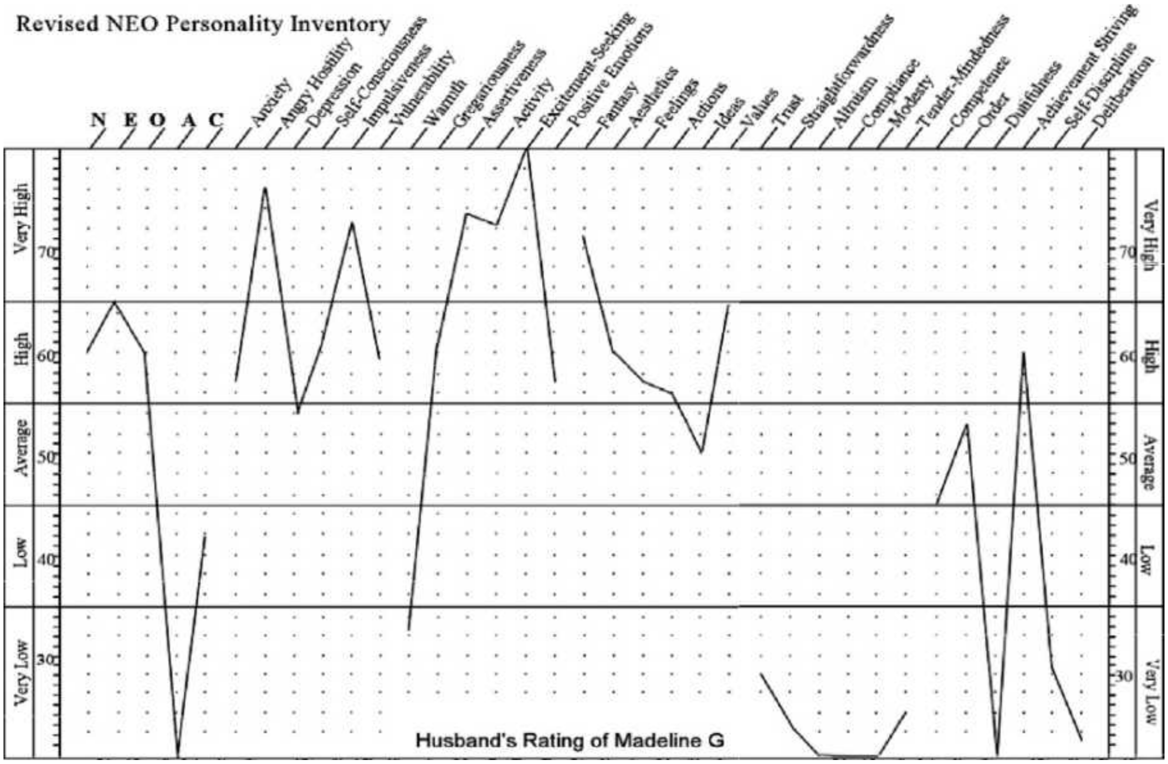

FIGURE 6.4–3. Personality profile of Madeline G as rated by her husband. T-scores (M = 50; SD = 10) comparing her to other adult men and women are plotted. The five factors are given on the left; the 30 facets, grouped by factor, are given toward the right. (NEO-PI-R profile form reproduced by special permission of the publisher, Psychological Assessment Resources, Inc., 16204 North Florida Avenue, Lutz, FL 33549, from the NEO Personality Inventory-Revised by Paul T. Costa, Jr., and Robert R. McCrae. Copyright 1978, 1985, 1989, 1991, 1992 by Psychological Assessment Resources, Inc. [PAR]. Further reproduction is prohibited without the permission of PAR.)

Clinical Applications for the FFM: Observer Ratings. The use of observer ratings is one area where research with the FFM can lead to further advances in clinical work and in identifying new domains. There are several values for using observer ratings of personality. There is no single method that is the gold standard for measuring an individual's personality. There is no one source or method for obtaining a true, unbiased profile of an individual. Rather there are multiple sources of information (e.g., self-report, observer rating, life outcome, etc.), and each provides its own, unique perspective on a person. Also, while each rating perspective has its own source of error across information domains, these errors are independent. Such uncorrelated error ensures that when scores converge across rating sources, one can be confident that the results are not contaminated by method biases. So, when both self-reports and observer ratings agree on the magnitude or level of the trait (e.g., low, average, or high), such convergence gives more confidence to the finding than either source alone. By using multiple sources of information, one is able to triangulate across them in a manner that removes any systematic method bias. When the different ratings do diverge, while providing an interpretive caution flag, such a situation may provide an opportunity for making a deeper evaluation of the client's motivational processes (see Piedmont, 1994; Piedmont, McCrae, Riemann, and Angleitner, 2000 in the references for discussions on observer-rated information).

One area where the use of multiple information sources can be diagnostically useful is with the assessment of couples. Many instruments designed to assess couple dynamics are very static in nature, in that they only evaluate each partner individually (e.g., a marital satisfaction scale) and/or address procedural aspects of the relationship (e.g., the roles expected of each other, such as who pays the bills, whose family is visited on what holidays, e.g., the ENRICH/PREPARE scales, Piedmont and Rodgerson, 2013). Yet, relationships are much more dynamic than what these types of scales capture. Interactions between members of couples are based on each member's own personality characteristics and the motivational goals they are pursuing. Levels of satisfaction are, in part, derived from each partner having a clear set of insights and expectations about their partners that enable them to accurately anticipate their partners' behaviors. Piedmont and Rodgerson have developed the "cross-over analysis" (COA) as a method for directly assessing these dynamic patterns. It involves a couple completing the NEO PI-3 for themselves and then rating the perceived personality of their partner.

This process generates four profiles, two self, and two observer profiles. Piedmont and Rodgerson outlined several different ways these data can be examined. The COA is the central interpretive strategy and requires that the observer rating of one member is compared to the self-report of the other. In this approach, levels of agreement between the two profiles speak directly to the types of issues the rater has with their partner. First, higher levels of agreement are associated with higher levels of relationship satisfaction in the observer; lower levels indicate less well-being with the partner. Second, scales where significant disagreements in scale scores exist (e.g., a difference of 15 or more T-score points) will indicate the types of difficulties the observer is having with the partner.

A couple who presented for marital counseling found the wife to have rated the husband significantly lower on O and C, while the husband rated his wife significantly lower on N and A. The wife found her husband to be emotively inadequate and very selfish. He did not like speaking about feelings, was not very supportive or caring, and seemed to want to be doing his own things. The husband, in turn, saw his wife as very emotionally stable and very antagonistic. In short, she was a real “bitch;” always yelling and criticizing him for being self-centered and unconcerned about her. He found these verbal attacks as an attempt to undercut his masculinity. What was missing in this relationship dynamic was that the husband did not realize how emotionally unstable his wife was. She was very insecure and anxious and needed much succorance and emotional support, something the husband did not recognize. Her yelling and berating of him was not an “attack” so much as an attempt by her to get her husband to respond supportively, to calm her fears, and reassure her. Her acting out was to get emotional help; he did not see the emotional need and rather interpreted these behaviors as an attack. As such, when she began her tirade, he would emotionally and physically withdraw to somewhere else. She saw this behavior as his inability to deal with and manage intimate emotions and to just “run away” from them. By examining the sources of these inaccurate perceptions and reframing the apparent underlying motivations, the couple came away with a different awareness of each other that promoted a more accurate understanding of their emotional needs.

The value of COA is that it provides a space for couples to examine their perceptions of each other in ways that are not judgmental or blameful. The comparison of the different protocols allows for an assessment of the views each partner holds of the other, and the resulting character styles that the rater believes are operating in their partner. Clinicians can then engage with these perceptions that are driving each person’s response style. Piedmont and Rodgerson provided a number of interpretive strategies for viewing discrepancies between the two sets of ratings. Comparing multiple profiles provides a rich amount of insightful information about couples.

Psychotherapy and Personality Change. If psychopathology is an expression or an outcome of personality traits and if, as longitudinal studies demonstrate, personality traits change little over time, how can psychotherapy be effective? The pessimistic answer is that it cannot. Many psychiatric disorders are lifelong or recurrent, and treatment consists of management rather than cure. It is well known that those who have the best prognosis for recovery are those who are initially least impaired. In these cases, the disorder may represent a transient adjustment reaction among individuals who have relatively healthy personality traits. Individuals high in N and low in A and C may seem to be poor candidates for psychotherapy.

However, such a pessimistic conclusion is unwarranted. Longitudinal studies show that people change little in the course of normal experience, but they do not rule out the possibility that direct interventions can make changes. It is unlikely that psychotherapy will make dramatic and lasting changes in basic personality traits, but more modest improvements may be enough to allow the patient to function adequately in daily life. Interventions may be particularly effective in early adulthood when traits are not yet fully developed. Even without changing personality, therapy may help individuals to adapt to their own nature. The chronically anxious person may learn techniques of relaxation; the disagreeable individual may learn social skills that improve interpersonal relationships.

In some sense, most forms of psychotherapy can be seen as learning experiences. Psychoanalysts promote insight into unconscious conflicts; behaviorists help their clients understand the contingencies that reinforce troubling behavior; humanistic psychologists encourage patients to discover their true potential. Trait psychology can also be a source of insight. All human beings are, in some respects, alike, and it is often comforting to learn that one is not alone in one’s suffering. In other respects, however, people are different from one another, and it can also be a therapeutic experience to learn this fact. Helping patients understand their own enduring dispositions can prepare them for some of the problems they face in life.

ACKNOWLEDGMENT

The authors would like to acknowledge the contribution of Robert R. McCrae to previous versions of this chapter.

FURTHER READINGS

Brummett BH, Babyak MA, Williams RB, et al. A putatively functional polymorphism in the HTR2C gene is associated with depressive symptoms in white females reporting significant life stress. *PLoS One*. 2014;9(12):e114451.

Costa PT Jr, McCrae RR. A Five-Factor Model perspective on personality disorders. In: Strack S, ed. *Handbook of Personology and Psychopathology*. Wiley; 2005.

Costa PT Jr, McCrae RR. The NEO Inventories as instruments of psychological theory. In: Widiger TA, ed. *The Oxford Handbook of the Five-Factor Model of Personality*. Oxford University Press; 2015.

Digman JM. Personality structure: emergence of the Five-Factor Model. *Annu Rev Psychol*. 1990;41:417.

Dollard J, Miller NE. *Personality and Psychotherapy: An Analysis in Terms of Learning, Thinking, and Culture*. McGraw-Hill; 1950.

Eysenck HJ, Eysenck M. *Personality and Individual Differences*. Plenum; 1985.

*Feist J, Feist GJ, Roberts TA. *Theories of Personality*. 8th ed. McGraw-Hill; 2012.

Frankl VE. *On the Theory and Therapy of Mental Disorders: An Introduction to Logotherapy and Existential Analysis (JM Dubois, Trans.)*. Brunner-Routledge; 2004.

Herbst JH, Zonderman AB, McCrae RR, Costa PT Jr. Do the dimensions of the Temperament and Character Inventory map a simple genetic architecture? Evidence from molecular genetics and factor analysis. *Am J Psychiat*. 2000;157:1285.

Kelly GA. *The Psychology of Personal Constructs*. Norton; 1955.

*Maddi SR, Costa PT Jr. *Humanism in Personology: Allport, Maslow and Murray*. Aldine; 1972.

Maslow AH. *Motivation and Personality*. Harper & Row; 1954.

McCrae RR. A more nuanced view of reliability: specificity in the trait hierarchy. *Pers Soc Psychol Rev*. 2015;19:97.

McCrae RR, Costa PT, Jr. *NEO Inventories: Professional Manual*. Psychological Assessment Resources; 2010.

*McCrae RR, Costa PT Jr. *Personality in Adulthood: A Five-Factor Theory Perspective*. Guilford; 2003.

*McCrae RR, Harwood TM, Kelly SL. The NEO inventories. In: Harwood TM, Beutler LE, Groth-Marnat G, eds. *Integrative Assessment of Adult Personality*. 3rd ed. *Guilford*; 2011.

McCrae RR, John OP. An introduction to the Five-Factor Model and its applications. *J Pers*. 1992;60:175.

McCrae RR, Terracciano A, Personality Profiles of Cultures Project. Universal features of personality traits from the observer's perspective: data from 50 cultures. *J Pers Soc Psychol*. 2005;88:547–561.

Miller T. The psychotherapeutic utility of the Five-Factor Model of personality: a clinician's experience. *J Pers Assess*. 1991;57:415–433.

Mõttus R, Kandler C, Bleidorn W, Riemann R, McCrae RR. Personality traits below facets: the consensual validity, longitudinal stability, heritability, and utility of personality nuances. *J Pers Soc Psychol*. 2017;112:474–490.

Piedmont RL. *The Revised NEO Personality Inventory: Clinical and Research Applications*. Plenum Press; 1998.

Piedmont RL. Validation of the rater form of the NEO PI-R for college students: Toward a paradigm for measuring personality development. *Assessment*. 1994;1:259–268.

Piedmont RL, McCrae RR, Riemann R, Angleitner A. On the invalidity of validity scales: evidence from self-reports and observer ratings in volunteer samples. *J Pers Soc Psychol*. 2000;78:582–593.

Piedmont RL, Rodgerson TE. Cross-over analysis: using the FFM and NEO PI-R for assessing couples. In: Costa PT, Widiger TA, eds. *Personality Disorders and the Five-Factor Model of Personality.* 3rd ed. American Psychological Association; 2013:375–394.

Piedmont RL, Rodgerson TE. Cross-over analysis: using the FFM and the NEO PI-3 for assessing compatibility and conflict in couples. In: Widiger T, ed. *Oxford University Handbook of the Five-Factor Model*. Oxford University Press; 2017.

Rogers CR. *On Becoming a Person: A Therapist's View of Psychotherapy*. Houghton Mifflin; 1961.

Sayer NA, Noorbaloochi S, Frazier PA, et al. Randomized controlled trial of online expressive writing to address readjustment difficulties among U.S. Afghanistan and Iraq War Veterans. *J Trauma Stress*. 2015;28:381.

Scarr S. Distinctive environments depend on genotypes. *Brain Behav Sci*. 1987;10:38.

Shapiro D. *Neurotic Styles*. Basic Books; 1999.

Skinner BF. *About Behaviorism*. Knopf; 1974.

Vachon DD, Lynam DR, Widiger TA, Miller JD, McCrae RR, Costa PT Jr. Basic traits predict the prevalence of personality disorder across the lifespan: the example of psychopathy. *Psychol Sci*. 2013;24:698.

*Widiger TA, Costa PT Jr, McCrae RR. Diagnosis of personality disorders using the Five-Factor Model. In: Widiger TA, Costa PT Jr, eds. *Personality Disorders and the Five-Factor Model of personality*. 3rd ed. American Psychological Association; 2013.

Widiger TA, Trull TJ. Plate tectonics in the classification of personality disorder: shifting to a dimensional model. *Am Psychol*. 2007;62:71.

Widiger TA, Trull TJ, Clarkin JF, Sanderson C, Costa PT Jr. A description of the DSM-IV personality disorders with the Five-Factor Model of personality. In: Costa PT Jr, Widiger TA, eds. *Personality Disorders and the Five-Factor Model of Personality*. 2nd ed. American Psychological Association; 2002.

Wiggins JS. *Paradigms of Personality Assessment*. Guilford; 2003.

7

Diagnosis and Psychiatry: Examination of the Psychiatric Patient

7.1 Psychiatric Interview, History, and Mental Status Examination of the Adult Patient

Rory P. Houghtalen, M.D., and John S. McIntyre, M.D.

George Engel wrote that the clinical interview in medicine is “Virtually indispensable for the physician–patient interaction, the well-constructed interview truly may be regarded as the most powerful, the most sensitive, and the most versatile instrument available to the physician.” Despite advances in science and technology, this statement remains true today, and perhaps nowhere in medicine is it more salient than in psychiatric practice, a specialty that relies almost entirely on a careful history and examination to develop a diagnosis.

Two primary aims shape the initial psychiatric interview. One is to develop a biopsychosocial database that informs diagnosis, risk assessment, and person (patient)-centered treatment planning. The other is to form a positive alliance that sets the stage for effective treatment that may follow the initial encounter. This chapter will focus on the initial psychiatric interview, but much of the commentary here is relevant to patient contacts in the course of treatment.

The settings in which psychiatric interviews occur include the psychiatric inpatient units, medical-surgical inpatient units, partial hospital programs, emergency rooms, outpatient offices, recovery services, nursing homes, other group homes, and correctional facilities. As the COVID-19 pandemic emerged and persisted, mental health professionals had to confront longstanding hesitancy to utilize distance technologies that have come to make virtual visits with patients a routine option in most mental health services in the United States and other developed countries. We address some specific issues related to virtual interviewing in the *Telepsychiatry* section of this chapter.

The sections of this chapter that follow will cover general principles and the processes of the diagnostic psychiatric interview from the initial request for consultation to the closing of the contact. Content includes commentary about documentation practices, addresses factors that influence the dynamics and quality of communication, and provides commentary on interviewing techniques. The chapter closes with a review of challenging patient scenarios and special topics related to the psychiatric interview.

GENERAL PRINCIPLES

Pre-Interview Contacts and Interactions

Referrals for psychiatric evaluation come from many sources, including patients seeking help, concerned family, other medical providers, schools, workplaces, and other agencies interested in the person's well-being. Some referrals are for reasons other than clinical treatment, for example, forensic evaluations, that require special considerations that this chapter's section *Forensic Interviews* briefly reviews.

The first contact in ambulatory settings usually begins with a phone call during which demographic, insurance information is obtained, and the patient (or other caller) is briefed on what to expect in the initial encounter, including the length of the encounter, fees, and other particulars of the office or service setting. Patients' first impressions influence attitudes and expectations. Typically, nonclinical staff are the first to make contact with new patients. Therefore, it is crucial to help them recognize and appreciate the anxieties that patients bring to a first encounter and offer strategies to help patients feel as much at ease as possible.

Clarifying the purpose of the referral with the referring source to tailor the evaluation is often essential, especially in a consult-liaison setting. Pre-encounter medical, mental health, and medication histories may be facilitated via mailed forms or secure electronic transmission. Patients may initiate retrieving relevant records prior to the first visit to enhance the database for diagnosis and treatment planning.

Many clinicians introduce self-report tools to gather history and identify symptom clusters and patterns. Sometimes, these tools are solicited before the first encounter or filled out before the first contact as part of the waiting room registration process to reduce the chances of missing a significant problem and make the initial interview more efficient. Broadband screening tools, like the Brief Patient Health Questionnaire, gather information about depressive and anxiety symptoms, social distress, traumatic exposure, and for female patients, screen for the possibility of pregnancy and other reproductive health issues. Disorder-based screening tools, like the Patient Health Questionnaire-9 (PHQ-9), aid in identifying patients likely to have a specific diagnosis, in this case, major depressive disorder. Others, such as the Quality-of-Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q), are designed to survey social and emotional functioning. Patients are increasingly familiar with the use of such tools in medical settings. While this data-gathering method has many advantages, and research demonstrates that patients generally view these efforts as a sign of an organized, scientific method, the informed clinician knows that tools generate hypotheses but do not make a diagnosis. Any screening tool produces false-positive and false-negative identifications depending on its sensitivity and specificity and the base rate of symptoms in a population. The clinician who uses screening tools must be prepared to define the true significance of a screening result and balance tools with careful interview and examination.

Setting the Stage for the First Encounter

In contemporary behavioral health practice, the psychiatrist interacts with patients in diverse settings that may be variably suited to the psychiatric interview. With the advent of telemedicine, the clinician and patient may not even be in the same location. Nevertheless, certain principles for staging a successful first interview apply no matter the setting, including attention to optimizing comfort, safety, and privacy. Most patients come to their first encounter experiencing some level of trepidation. Some arrive cajoled or forced by others to have come at all. Anxieties are reduced by early clarification of the nature and purpose of the encounter and defining a mutual agenda before embarking on the interview. To the extent feasible, the patient should experience the clinician focused on them for their time together. To this end, unnecessary distractions such as beepers, text alerts, and phone calls should be minimized. Potential distractions related to note-taking and the electronic medical record are discussed in the *Special Topics* section of this chapter.

Comfort and Safety

For most patients, the emotional distress that brings a person to the initial interview is considerable and there is apprehension about what it means to see a mental health professional. Inadequate knowledge about mental health care and misconceptions about psychiatric treatment arising from stigma can fuel these anxieties. The skilled clinician is aware of these potential issues and interacts to decrease, or at least not increase the distress. An adage some attribute to Maya Angelou (others to Carl W. Buehner) offers some words of wisdom that apply to these first interactions with new patients: “I’ve learned that people will forget what you said, people will forget what you did, but people will never forget how you made them feel.” The patient’s first encounter experience is bound to affect engagement and set the tone for subsequent interviews or treatment that may follow and may affect attitudes about psychiatric and general medical care far into the future.

Not much is accomplished in any human interaction when the participants are physically uncomfortable or excessively anxious about personal and emotional safety in an environment. The clinician controls the environment in some settings, like a private office, but not so much in others, like a public clinic or the lock-up of a county jail. Regardless of where the initial contact occurs, interviewers should do what is within their power to adjust the environment to provide a reasonable level of comfort and assure the safety of both the patient and the clinician.

Engaging with patients known or found to have active psychosis, acute intoxication, or a history of aggression requires prudence and planning. It is wise not to interview patients with these problems in isolated areas. Optimally, areas used for interviewing should have adequate interpersonal space and arrangements that allow both patient and interviewer an easy exit without “going through” one another’s space. Tenuous situations may require modifications to usual interview techniques, including having another staff member present in the room during an encounter or standing by outside the door.

Explicitly discussing safety can reduce uncertainty, tension, and risk. The overtly disturbed patient is often calmed by acknowledging their discomfort and reassurance that the first goal is for the patient and staff to remain safe. Sometimes pointing out the clinician’s own anxieties about safety can level tension. A statement like this may help: “When you swear and speak in such a loud voice, it makes me anxious too. When I am very anxious, it is hard for me to pay attention to what you are saying, and interferes with my ability to help you. Do you think you can tone the language down a bit so that I listen better?” “Is there anything that I can do to help you feel calmer?” Even in a properly staged and managed interview, it may be necessary to shorten or quickly terminate the encounter if the patient escalates or becomes explicitly threatening. Judicious use of a therapeutic “show of force” can actually reduce the patient’s fears of loss of control by demonstrating adequate staff presence and a commitment to keeping everyone safe.

Privacy and Confidentiality

Privacy and confidentiality are essential components of therapeutic relationships. Staging the interview to ensure that others do not overhear the content is important. In some settings, such as a shared inpatient room or a cellblock in a jail, privacy may be difficult to establish. In less-than-optimal environments, judiciously steering the interview away from material that may overexpose the patient is wise.

The level of confidentiality afforded the patient varies depending on the nature of the encounter, making it essential to define the boundaries of confidentiality at the onset so that patients can make informed decisions about the nature and extent of material they are comfortable sharing. The clinician should consider reviewing with the patient situations that may lead one to share information without expressed consent. Examples include contacting informants to evaluate concerns about imminent suicide or violence risk, transferring information in medical or psychiatric emergencies, and mandated reporting of child abuse or neglect. Insurance claims and communications with pharmacies routinely require sharing of information. Minor children need to understand that a parent or guardian must be informed about diagnosis and treatment options to provide surrogate consent. Detailing the boundaries of confidentiality and exceptions in writing may strengthen trust and confidence in the clinical engagement.

Specific consent is first obtained (best in writing or at least carefully documented in the medical record) when information is shared outside of a genuine emergency or a mandated report. In the United States, the federal Health Insurance Portability and Accountability Act (HIPAA) rules define standards for the proper release of protected health information, and state-by-state regulations provide additional guidance for the management of protected health information. The details of this are beyond the scope of this chapter, but the clinician should understand and carefully follow the appropriate procedure for the release of information, understanding that HIPAA usually defers to the state regulation if it is more restrictive.

When it is necessary to share information without specific consent, the nature and extent of disclosure should be limited to that information essential to the task. This discretion conserves confidence to the extent possible, even when the clinician has a responsibility that may conflict with the patient's privacy.

Involvement of Third Parties

Often members of the patient's family, or other supports, attend a first encounter and may expect to be involved in the interview. These persons can be significant sources of collateral information, and ultimately everyone may agree they should be involved in treatment decisions and management. The patient and psychiatrist should discuss whether and how to involve significant others in clinical interviews and the nature and extent of information the patient wishes to share, assuming the patient is competent to assert confidentiality and an overriding factor, such as suicide or violence risk, does not trump the patient's choice. Patients typically appreciate the principle of "no meeting about me without me." Therefore, it is usually best to sequence the involvement of family or important others by first meeting privately with the patient and then, with the patient's agreement, bringing others into the encounter with the patient present.

Engagement and Relationship Building

The term "doctor–patient relationship" is a phrase used to describe that special interpersonal connection that sets the stage for healing interventions and on which current culture confers special status and obligations. Recently, the term "patient–doctor relationship" was coined, reversing the order of the parties to reinforce that the treatment should always be patient-centered. Projecting a friendly and open demeanor, maintaining a respectful and nonjudgmental posture, and demonstrating genuine interest in developing an empathic understanding of the patient's predicament are physician behaviors that build the type of rapport from which the patient–doctor relationship is forged. In the initial interview, the successful development of a trusting relationship allows the patient to feel comfortable enough to share sensitive and private information.

The relationship between patients and physicians will vary depending on their respective personalities and experiences. Equally important are the setting and purpose of the encounter. However, when followed, there are general principles that help ensure that the relationship established is one from which successful treatment can occur. Jerome Frank first examined therapeutic relationships from an anthropologic perspective in 1974. His observations discovered common characteristics that confer on an individual the status of healer regardless of culture. In Westernized cultures, specialized training and experience are factors. In addition to training and expertise, physicians are expected to behave in ways that project caring, instill hope, and maintain focus on the patient's best interest.

In the clinical setting, rapport can be defined as the harmonious responsiveness of the physician and patient to one another. It is vital that the patient experiences the evaluation as a joint effort and that the psychiatrist is genuinely interested in their story. Appreciating how and why the patient thinks and feels requires understanding the patient's life experience and perspectives in some detail and depth. Early in an initial interview, the clinician forms hypotheses about the patient's experience through what the patient says, what affects accompany the story, and, just as importantly, what is not said. As the interview progresses, the details of the patient's symptoms and story unfold, and patterns of thinking and behavior become evident, resulting in greater clarity about the patient's actual lived experience.

Empathy is the capacity to appreciate and understand the experience of the patient at an emotional level, achieved when the interviewer can imagine themselves in the patient's position while at the same time maintaining objectivity. Objectivity differentiates empathy from sympathy. Sympathetic identification with the patient may distort and stunt the development of a mature therapeutic relationship. Consider the experience of interviewing a patient with a borderline personality disorder. Sean Shea has noted, "One feels compelled to say a silent prayer for the poor patient with borderline features who meets a clinician who boldly proclaims, ‘I can feel your pain.’” An awareness and understanding of the dynamics of the patient–physician relationship is vital in helping the physician maintain objectivity while still transmitting concerned understanding and extending a helping hand. Empathic expressions (“*That must have been very difficult for you*” “*I’m beginning to understand how awful that felt.*”) serve to increase rapport. Frequently a nonverbal response (raised eyebrows or leaning toward the patient) or a very brief response (“Wow.”) will be similarly effective.

Ultimately, it must become apparent to the patient that the doctor understands and cares about his experience for effective treatment. In 1927, Francis Peabody wrote, “The secret of care of the patient is caring for the patient.” Caring for the patient is not automatic, and it is frankly difficult at times. James Groves (1978) published a classic paper entitled “Taking Care of the Hateful Patient.” Rael Strous and colleagues updated the paper for contemporary practice in 2006. Groves described patterns of patient behaviors that made them unlikeable to physicians and warned how this disdain could interfere with optimal medical care noting that if physicians become more aware of their feelings and reactions, they have increased opportunities to interact in a manner that can be helpful even under challenging situations.

The development of a healthy patient–physician relationship is reinforced when the physician is perceived as genuine. Being comfortable enough to laugh in response to a humorous comment, admit a mistake, or apologize for an error that inconvenienced the patient, like being late for or missing an appointment, strengthens the therapeutic alliance. It is also essential to be flexible in the interview and responsive to patient initiatives in relationship building. If the patient brings in an item, for example, a photo she wants to show the psychiatrist, it is good to look at it, ask questions, and thank the patient for sharing it. Much can be learned about the family history and underlying relationship dynamics from such a seemingly sidebar moment. The psychiatrist should be mindful that there are no irrelevant moments in the interview room.

At times patients will ask questions about the clinician that range from inquiries about training and experience to more personal matters such as marital status, sexual orientation, or religious beliefs. Depending on the reason for the interview, the nature of the question, and the particular characteristics of the setting and patient, these questions may properly be answered directly or reflected back to the patient to gain insight into the motive for the inquiry. Avoiding excessive sharing of personal information is a good general principle to adhere to, especially if the interview is likely to proceed into a psychotherapeutic treatment that relies on the relative anonymity of the therapist as a central tenet, such as in the case of psychodynamic therapy. On the other hand, the clinician should not be excessively stingy with information for more self-serving reasons than patient-centered at the risk of appearing aloof or callous.

A clinical example may help to illustrate these principles. An inpatient psychiatrist rounds for the first time on a newly admitted patient with major depressive disorder. The following conversation initiated their contact:

Psychiatrist: *Good morning, Mr. Smith. I'm Dr. Jones.*

Patient: (Interrupting) *Before you get started, I have one question for you. Are you Catholic?*

Psychiatrist: *Wow. That's quite a start. May I sit down?*

Patient: *You can sit down, but I'm not going to answer any questions until you tell me if you are Catholic.*

Psychiatrist: *I can see that's very important to you. Can you tell me why that is important?*

Patient: *I'm not going to say anything till you tell me.*

Psychiatrist: *I can tell you something about who I am and why I'm seeing you this morning.*

Patient: *Never mind about that stuff. I want to know if you're Catholic.*

The psychiatrist presses on, attempting to understand the reason for the question. Was there a prior unsatisfying experience with a

Catholic or non-Catholic psychiatrist? Does the patient worry that his religious concerns will not be understood or incorporated into care? The same resistant responses occur with each attempt to reflect on or address Mr. Smith's feelings.

Psychiatrist: *I'd answer the question, but I anticipate answering will lead to more questions about me, and we won't talk about you and why you were hospitalized or how you feel about being here.*
Patient: *No, that's the only question I have. If you answer that question, I'll answer whatever questions you have.*
Psychiatrist: *Okay. I'm Catholic.*
Patient: *Do you go to mass every week?*

There are occasions when, depending on the nature of the planned or ongoing treatment, it can be helpful for the psychiatrist to share some personal information even if the patient does not inquire directly. However, in all cases, the motive for self-revelation should be to strengthen the therapeutic alliance in the patient's service, not to meet the doctor's needs.

Meaning and Communication

An understanding of the process and content of the psychiatric interview would be incomplete without considering the role of the unconscious. From psychodynamic and neurobiologic perspectives, the reality is that most mental activity remains outside of conscious awareness. The psychiatrist and the patient both operate on conscious and unconscious levels and bring their respective personalities, personal histories, and idiosyncrasies to the encounter. The psychiatrist should remain alert for manifestations of unconscious processes in communication and relationship dynamics in themselves and the patient.

Unconscious processes in the patient–doctor relationship reveal themselves through transference and countertransference manifestations and various psychological defense mechanisms.

Transference is the process by which the patient unconsciously and often habitually displaces/transfers onto individuals in his current life those patterns of behavior and emotional reactions that originated with significant figures from earlier in life, often from childhood, like those attached to a parent or authority figure. Similarly, countertransference is the process whereby the physician unconsciously and reciprocally is involved in displacement to the patient. Remaining mindful that such distortions drive affects and behaviors of the patient and the psychiatrist alike reduces the burden of unnecessary emotional distress, unhelpful behaviors, and missed therapeutic opportunities when these processes are not recognized.

The primary care physician (PCP) offered considerable detail about an older woman’s background and symptoms in a letter sent to initiate a referral for psychiatric evaluation. Early in the psychiatrist’s first interview, it became apparent that the patient’s primary concern was the sexual relationship with her husband, which the psychiatrist found puzzling because the PCP was psychologically minded and had mentioned nothing about this in the letter. When asked if she had discussed this issue with her PCP, the patient responded, “Well, it took me awhile to even broach the topic with him because he is young enough to be my son. I tried to bring it up a couple of times, but he looked uncomfortable and changed the subject. I didn’t bring it up again; I didn’t want to bother him.” Subsequently, in a conversation about this patient, the PCP spontaneously commented, “She’s a very nice lady; she reminds me a lot of my mother.”

Other conscious and unconscious processes represented in the so-called “defense mechanisms” of psychodynamic theory shape and sometimes interfere with interviewing. The patient is generally unaware of the impact of these feelings and behaviors, which take many different forms, including tangential references to an issue, slips of the tongue, mannerisms of speech, avoidance of topics, exaggerated emotional responses, intellectualization, generalization, missed appointments, or other acting out behaviors; each of these may be examples of the unconscious at work. For example, phrases such as “to be honest with you” or “to speak frankly” suggest that the speaker is perhaps not being honest or speaking frankly. Interpreting the meaning of statements like these when they occur in initial interviews is tempting but generally ill-advised. It is better to note such manifestations while refraining from premature exploration of defenses until the relationship is secure enough to tolerate this manner of intervention and the patient seems ready to receive it as helpful, though a reflection such as “I would hope you will always be honest with me,” may be appropriate and constructive.

Open-Versus Closed-Ended Questions

An open-ended question identifies and introduces an area of inquiry but provides minimal structure about responding to the question or the desirable response. This style of questioning, in theory, allows the subject to respond spontaneously with as much detail as the subject chooses. It also minimizes the potential that the questioner will influence the answer. An example of an open-ended question is, “What can you tell me about this nervousness you mentioned earlier?” Closed-ended questions invite brief, “Yes” and “No” types of responses and may telegraph the desirable response. An example of a closed-ended question: “You are not nervous all the time, are you?”

The flow of a typical interview moves from open-ended questions that probe a particular topic to closed-ended questions that provide definition and detail when required. Both types of questioning are fruitful to explore topics fully, and each approach has its relative strengths and weaknesses. The balance of open- versus closed-ended questions will vary depending on the patient’s capacity to grasp the invitation of open-ended questions, the nature of the material explored, the patient’s response style (e.g., the patient prone to circumstantial responses may have to be structured more than others) the person’s openness to disclosure and the purpose and time available for the interview. No prescription defines the proper balance.

Person-Centered and Disorder-Based

The initial psychiatric interview has at least three crucial goals: establish a traditional biomedical diagnosis, understand the person's biopsychosocial dimensions, and build rapport. Clarifying the agendas of the clinician and the patient early in the encounter helps to define the boundaries of the interview and plan the best use of time. Starting an interview with a statement like this can help initiate agenda-setting: "We have all or part of an hour together today. My agenda is to understand the problems that bring you to this visit. I also want to learn the background of your life story to understand you as a person. Before we begin, I would like to know what you hope to get out of our meeting today and what questions you may have as we begin."

The individuality of the patient's experience is a central theme. The interviewer elicits the patient's life history in more or less detail subject to the constraints of time and the patient's willingness to share details of experiences and their private mental life. Adolf Meyer's "life-charts" were graphic representations of the material collected in this endeavor and were a core component of the "psychobiological" understanding of illness. The patient's early life experiences, cultural background, family dynamics, education, work history, religious beliefs and practices, hobbies, talents, relationships, traumatic events, and losses each, in concert with genetic and biologic factors, represent the field in which distress arose. An appreciation of the richness of life experiences and their impact on the person is necessary to develop a complex understanding of the patient. Gleaning all of this information in a first encounter may not be practical, but gathering enough to grasp the essence of both the problem and the person is an important goal.

Traditionally, medicine has focused on illness and deficits rather than strengths and assets. A person-centered approach focuses on strengths and assets as well as deficits. During the assessment, it can be helpful to ask the patient, "*Tell me about some of the things you do best.*" or, "*What do you consider your greatest asset?*" A more open-ended question such as, "*Tell me about yourself.*" may elicit information that focuses more on either strengths or deficits depending on several factors, including the patient's mood and self-image. Cultural norms also play an essential role in how much a person is willing to divulge about their inner experience and sense of self.

In addition to being person-centered, it is also vital that the psychiatric interview be disorder-based. Significant advances in medicine have occurred over the past two centuries as psychiatrists began to recognize that specific clusters of signs and symptoms were characteristic of disorders with common epidemiology, features, presentations, course, and response to treatment. Developing a criteria-based nomenclature, as cataloged in the *Diagnostic and Statistical Manual of Mental Disorders,* now in its fifth edition (DSM-5), has furthered psychiatric diagnosis and effective treatment. The DSM has some limitations, of course, but it transformed the esoteric methods of diagnosis that preceded it into a standardized method that improved the reliability of diagnosis and aided research and clinical work. In part, evidence-based psychiatric practice has flourished because of the increased reliability of diagnoses.

The best practice interview integrates the person-centered and the disorder-based approaches, and the interview process weaves back and forth between the two ending in a working diagnosis, differential diagnosis, and treatment plan. Incorporating the patient's needs and goals into the treatment plan is essential. Numerous studies have demonstrated that too often, the patient's goals for treatment (e.g., safe housing) are not the same as the psychiatrist's (e.g., decrease in hallucinations). This dichotomy may originate in an initial interview wherein the focus was not sufficiently person-centered but rather was exclusively or predominantly symptom-based. Even when the interviewer asks explicitly about the patient's goals and aspirations, a patient exposed on numerous occasions to what a medical professional is interested in hearing about may attempt to focus on "acceptable" or "expected" goals rather than expressing their version. Clinicians should explicitly encourage patients to assert their goals and aspirations and express them in their own words.

Time and Number of Sessions

The time devoted to the initial interview will vary considerably by the goal of the encounter and the setting. For example, about an hour is generally allotted in a typical outpatient treatment environment. However, in settings like emergency and hospital inpatient units where patients are often distressed, confused, actively psychotic, or agitated, the time may be foreshortened, and assessment may by default occur over multiple, brief encounters or rely heavily on collateral historians to obtain necessary information.

Table 7.1–1.
Sections of the Psychiatric History and Examination

- Identifying data
- Sources of information
- Chief complaint
- History of present illness (HPI)
- Past psychiatric history (including suicide and violence risk)
- Substance use/abuse/addictive behaviors
- Past medical history
- Family history
- Developmental and social history
- Review of systems
- Mental status examination (MSE)
- Physical assessment
- Formulation
- DSM-5 diagnosis
- Treatment plan

The clinician must accept that any history obtained is never complete or entirely accurate because of the nature of human memory and the distortions of perception that each person involved in an experience makes when recounting it, along with variations that occur with each retelling. In addition, an interview is dynamic, and some aspects of the evaluation are ongoing—new facts emerge so that the history often morphs to a lesser or greater degree depending on various factors, including the patient’s gradual comfort with revealing and the dynamics of the relationship between the patient and doctor.

HISTORY AND EXAMINATION

Process of the Interview

The key elements of the psychiatric interview are the patient’s histories and the mental status examination (MSE). The histories are based primarily on the patient’s subjective report but should be enriched by and compared to available records and the reports of collateral historians when indicated. In addition, it is essential to document the sources of information contributing to the evaluation with commentary on the quality and reliability of the sources. Table 7.1–1 lists the traditional elements of the psychiatric history and examination in the conventional sequence that findings are typically documented. The following sections will further define and discuss these elements.

The MSE is, in essence, the “physical examination” of psychiatry, a series of observations of signs and symptoms that are gathered and documented systematically. This chapter’s section, *Physical Assessment,* provides an overview of physical assessment in the context of psychiatric evaluation. The histories, MSE, and physical assessment define a working diagnosis.

Opening the Interview

After introductions (including clarifying how the patient prefers to be addressed), explaining the initial evaluation process, agenda-setting, and answering immediate questions and concerns, the interview itself begins.

The psychiatrist can begin with an open-ended inquiry, "*Why don't we start by you telling me your understanding of what has led to your being here today.*" Sometimes, the patient is uncertain why a referral to a psychiatrist was necessary or may feel frustrated or angered by the need to see a psychiatrist.

The following kind of conversation is not uncommon:

Mr. Smith: *Dr. Jones said I should see you.*

Doctor: *What's your understanding of why Dr. Jones recommended you see me?*

Mr. Smith: *Well, you're a psychiatrist, right?*

Doctor: *Yes, I am.*

Mr. Smith: *Well, after my tests all came back negative, he kind of implied this nervousness—that it's all in my head and I should see you. I think he's pretty frustrated with me*.

The patient should have an opportunity to express his feelings about the need for the encounter and his anxieties. The psychiatrist has an opportunity to inform the patient about their role as a medical specialist. They can also discuss the nature and importance of collaborative relationships, such as those with primary care providers or other mental health clinicians in patient care.

Chief Complaint

Traditionally, the chief complaint is recorded in the patient's own words, for example, "I have been depressed for months" or "I have a lot of anxiety in public speaking."

The following example illustrates the importance of recording the chief complaint as the patient offered it:

An older adult man presented in a psychiatric emergency room with the chief complaint, “I’m melting away like a snowball.” He had become increasingly depressed over the prior three months. Four weeks before this emergency visit, the PCP increased imipramine (an antidepressant) from 25 to 75 mg and added hydrochlorothiazide (a diuretic) 50 mg because of mild hypertension and slight pedal edema. Over the ensuing month, the patient’s condition deteriorated. In the emergency room, he was noted to have a depressed mood, a hopeless attitude, weakness, significant weight loss, psychomotor retardation, and was described as appearing “depleted.” He also appeared dehydrated. Blood work found elevated BUN and creatinine levels and hypokalemia. Examination of his medication supplies revealed that the medications were dispensed in the wrong bottles. The patient was actually taking 25 mg of imipramine, generally a nontherapeutic dose for depression, and 150 mg of hydrochlorothiazide, generally a supratherapeutic dose for diuresis. He was indeed “melting away like a snowball” from dehydration. Fluid and potassium replacement rapidly improved his general medical status and allowed the antidepressant to be safely titrated toward a therapeutic dose to treat his depression.

After establishing the chief complaint, the interview proceeds into the history of present illness (HPI).

History of Present Illness

The HPI is a chronologic description of the evolution of the symptoms of the current episode. As with any well-developed HPI in medicine, the psychiatric history includes details about the nature, frequency, severity, and pattern of symptoms, including exacerbating and alleviating factors. In addition, stressful life events, physical symptoms, medication changes, and substance use patterns may provide clues to the context of the presentation, the primary diagnosis, and comorbid conditions that can influence treatment choices.

The exploration begins, ideally, with an open-ended invitation like, “Please tell me how this problem began, what the experience has been like, and how it has played out over time.” The interviewer may have to structure the patient through parts of the story. It is crucial to try to answer the question, “Why is the patient presenting now?” (e.g., *“I’m here now because my girlfriend told me if I don’t get help with this nervousness, she will leave me.”*)

The psychiatrist reviews treatments the patient utilized during the current episode. They should attend to response, adherence, and attitudes about the treatment. The common use of alternative and over-the-counter treatments makes it crucial to inquire specifically about the use of these products.

Suicide and violence risk assessment (sometimes termed “lethality” assessment) is a standard element of the HPI and any complete psychiatric evaluation. Patients may not spontaneously announce dangerous ideation or acts for many reasons, including shame and fear that revealing may lead to undesirable consequences, including reporting and confinement. The *Suicide and Violence Risk History* section of this chapter provides additional information about risk assessment.

Rounding out a comprehensive HPI is a review of psychiatric and physical symptoms to detect additional symptoms and comorbid disorders relevant to comprehensive diagnosis, treatment planning, and prognosis. Two examples may illustrate the importance of a review of symptoms. First, failing to conduct a review of psychotic symptoms with the depressed patient could lead to a missed diagnosis of psychotic depression, the treatment of which varies significantly from nonpsychotic depression. Failing to conduct a medical review of symptoms, the clinician could easily miss that a patient with depression and congestive heart failure (CHF) developed shortness of breath after the PCP initiated a tricyclic antidepressant (TCA). Not recognizing that the TCA was exacerbating the CHF, the psychiatrist missed the opportunity to recommend a therapeutic alternative for depression that would steer away from inotropic side effects. Table 7.1–2 outlines major categories of psychiatric symptoms, including mood, anxiety, psychotic symptoms, that are part of a psychiatric review of systems. The *Review of Systems* section of this chapter and the figures that accompany it offer more information about the physical review of symptoms.

Past Psychiatric History

The past psychiatric history explores psychiatric illness prior to the current presentation, including symptoms, course, and treatment. Table 7.1–3 summarizes the items generally covered in a complete past history review. Details of past episodes, including the age of onset, context, nature and duration of episodes, the diagnosis offered, treatment applied and its setting, degree of response, treatment adherence, and attitudes toward treatment, are essential facts to gather. In addition, understanding the details of past treatments will lead to a better understanding of which treatments are viable alternatives and which to avoid.

Most psychiatric disorders have a recurrence pattern, and comorbidity is the rule rather than the exception. Some episodes of illness may have gone undiagnosed, especially milder forms of illness. Establishing the patient's best functional baseline and the extent of recovery between episodes provides a perspective on the impact of the problem(s) on the trajectory of the patient's life. It helps identify elements of a comprehensive treatment plan that optimizes the fullest possible recovery.

Fastidious review of prior medication trials provides valuable information to define reasonable treatment medication options and rule out others. Essential elements of this review include the dose and duration of trials, the nature and degree of effect and side effects, and whether the medication was used alone or in combination with other psychotropics or psychotherapies. Exploring adherence may lead to discovering obstacles such as negative attitudes, side effects, dosing complexities, stigma, social pressures, and financial burdens that affected prior medication trials. Normalizing problems with medication adherence helps to obtain frank responses. Consider these types of questions to invite the patient to acknowledge challenges with adherence: “Most people miss some doses of medication. How many times a week do you miss your medication doses?” “What gets in the way of taking the medication regularly?”

Table 7.1–2.
Psychiatric Review of Systems

Sleep

Sleep phase problems (initial, middle, terminal insomnia), total sleep time, abnormal sleep events

Mood

Depression: Persistent sadness, reduced interest or pleasure in usual activities, tearfulness, reduced or excessive sleep, reduced or increased appetite, weight loss or gain, low energy, reduced concentration, low libido, excessive or inappropriate guilt, psychomotor change (slowing or agitation), negative self-appraisal, helpless and hopeless thinking thoughts of death or suicide. A common mnemonic used to remember the symptoms of major depression is SIGECAPS (Sleep, Interest, Guilt, Energy, Concentration, Appetite, Psychomotor agitation or slowing, Suicidality).

Hypomania/Mania: Elevated, expansive, or irritable mood, decreased need for or inability to sleep, excessive energy, marked increase in goal and pleasure-directed activity, increase amount and pace of speech and thought, grandiosity, heightened libido, impulsivity and/or recklessness in behaviors such as spending and sex

Anxiety

Experience of panic attacks, somatic symptoms of anxiety, phobic, or social avoidance

Psychosis

Experience of hallucinations, delusions, disorganized behavior, speech or thought, negative symptoms

Obsessive-Compulsive

Repetitive intrusive and unwanted thoughts, compulsive behaviors to neutralize anxiety, hoarding behaviors

Trauma

Traumatic exposure; intrusive and avoidance symptoms, negative alterations in cognitions and mood, excessive arousal and reactivity

Behavioral

Substance use, gambling, impulse control problems, disordered eating, repetitive self-harm

---

It is similarly essential to gather details about psychotherapy trials, including the type of psychotherapy (individual, group, couple, or family), the model of psychotherapy (Was it evidence-based for the diagnosis?), and the frequency and duration of the treatment (Was it delivered in a manner that one expects to be effective?).

Table 7.1–3.
Past Psychiatric History

1. Details of prior mental health contacts
   - Sites/Levels of care: School-based, primary care-based, outpatient, emergency, inpatient, partial hospital, recovery programs
   - Nature of prior treatment: Psychotherapy, medications, technologies (e.g., ECT), recovery-based, alternative, and complimentary
   - Adherence and response to treatment
2. Prior acts of self-harm, suicide, aggression, and violence

To ensure the history is complete, the clinician should inquire about less common technologic modalities that the patient may have used, including light therapy, biofeedback, electroconvulsive therapy, and vagal nerve and transcranial magnetic stimulation.

Suicide and Violence Risk History

Developing the risk history is a critical component of a complete psychiatric history. The suicide and violence risk assessment field has matured to understand that risk is a dynamic process influenced by risk-enhancing and risk-reducing factors with origins in genetics, social learning, current environments, and the precise nature of present signs and symptoms. A complete discussion of risk assessment is beyond the scope of this chapter, but it is critical to point out key elements of a risk assessment history. These include details of past events of intentional self-harm and aggression, including the number of events, details of the context of prior dangerous behaviors, the level of intent and seriousness of behaviors, harmful outcomes, and the means used to harm self or others. A full exploration includes nonlethal behaviors that increase the risk of future lethality, including intentional self-injury. In addition to the behavior itself, the psychiatrist should explore the feelings and attitudes about the behavior and the emotional states that accompany or follow it. In addition, the psychiatrist should also try to understand to what degree the patient tried to conceal evidence of these behaviors. The degree of family or community support and religious beliefs that admonish against suicide are potential risk-reducing factors that should be assessed. Firearms account for slightly more than half of the suicide deaths in the United States, making screening for firearm access a routine aspect of psychiatric evaluation.

Additional information about risk assessment is found in other sections of this chapter: the *Mental Status Examination* section under *Risk Assessment,* the *Selected Challenging Interview Scenarios* under the headings *The Patient with Depression, The Potentially Suicidal Patient,* and *The Potentially Violent Patient*.

Substance Use, Abuse, and Addictive Behaviors

Substance use may be the primary reason for a referral and should be thoroughly evaluated. In addition, it is often a critical secondary problem. Substance use disorders can mimic or induce psychiatric syndromes, elevate the risk of suicide and violence, and impact safe medication prescribing. These facts highlight the importance of understanding the patient’s current and past patterns of substance use.

Starting the inquiry about substance use with a normalizing statement can help to reduce defensiveness and elicit a fuller history: “Most people drink alcohol at times, and many have also used a recreational drug at some point. I want to ask you some questions about your past and current use of alcohol and recreational drugs.”

Various tools can be used to aid in gathering the substance use history. Examples include the commonly used CAGE questionnaire, which has been modified to include other drugs (and now called CAGE-AID) though studies have found the operating characteristics of this tool wanting in primary care settings and with Caucasian females, pregnant women, college students, and less severe levels of drinking. The National Institute on Alcohol Abuse and Alcoholism (NIAAA) recommends a procedure for alcohol use inquiry (available at https://www.niaaa.nih.gov/sites/default/files/publications/YouthGuidePocket.pdf) that offers the opportunity for education about safe levels of drinking for those patients at risk in addition to screening for current alcohol use problems using the alcohol use disorders identification test (AUDIT).

For each substance the patient acknowledges using, the interviewer gathers basic information about the age of onset, the pattern of use over time, current frequency, use level, and consequences (physical, mental, social, and legal). These facts help to identify problematic current or past patterns of use. Inquiry about tobacco use and problem gambling should be a part of the addictive behavior inventory.

The interviewer should identify periods of abstinence and determine what helped the patient achieve control of use, including any specialized treatment for substance use with details about the setting, nature of and response to treatment, and peer support engagement such as Alcoholics and Narcotics Anonymous.

Past Medical History

Past medical history is a bit of a misnomer, as this section of the initial interview and its documentation is usually used to review both past and present medical history. The interviewer is interested in obtaining an accounting of major medical disorders to develop a complete history and identify illnesses that could mimic a psychiatric disorder, contribute to the context of the presentation, or factor into treatment planning.

Medical disorders can precipitate a psychiatric disorder, such as an anxiety disorder in a patient newly diagnosed with cancer. They can also mimic a psychiatric disorder (e.g., hypothyroidism presenting “as if" major depression) or be precipitated by treatment of a psychiatric disorder (e.g., metabolic syndrome emerging during exposure to a second-generation antipsychotic medication). Medical disorders can also influence the choice of treatment of a psychiatric disorder (e.g., renal insufficiency and the use of lithium carbonate).

A well-developed past medical history archives current and past major medical disorders, surgeries, hospitalizations, and significant physical trauma, such as head injuries. Neurologic and endocrine disorders are of particular interest in psychiatry because of the significant overlap in symptoms and signs with psychiatric syndromes. For female patients, obtaining a reproductive and menstrual history is essential, as is a careful assessment of the potential for current pregnancy and plans for future pregnancy.

A thorough review of all current medications is essential because some medications used in other fields in medicine can create side effects that mimic psychiatric disorders and identify potential drug-drug interactions before prescribing. Patients have access to various over-the-counter, herbal, and alternative medications. They may not mention the use of these products unless specifically asked. Many over-the-counter and herbal/alternative products can produce psychiatric symptoms or interact with psychotropic medications in ways that range from nuisance side effects to outright hazards. Frank Ayd published results of the so-called "Brown Bag Study" that illustrated just how revealing an inspection of all the medications and products that a patient takes can be, and that serves as a reminder of the value of having the patient bring to the appointment their "brown bag" of bottles for inspection.

Allergies to medications must be identified, including the nature of the allergic response.

Family History

Because many psychiatric illnesses have a genetic predisposition, if not cause, a careful review of family history is essential to the assessment and can aid in diagnosis and establishing the expected prognosis. This area of history identifies family members with histories of known or suspected mental illness, substance use problems, and other behavioral problems, for example, criminality. Patients may not know the diagnosis of a family member but may be able to account for the relative's signs and symptoms from which reasonable hypotheses regarding diagnosis can be made. When the patient offers a diagnosis of a family member without much information to assess the reliability of the claim, the alleged diagnosis should be held skeptically or some attempt made to clarify it. Specific inquiry about family history of completed suicide is vital to suicide risk assessment because this finding elevates the index patient's future risk. In addition, a family history of medical disorders may offer clues to the patient's risk factors. For example, diabetes in first-degree relatives ought to raise a concern about the potential for metabolic syndrome with exposure to second-generation antipsychotics.

Developmental and Social History

Factors such as the reason for the interview and the amount of time available for it will determine how broad and deep the inquiries about developmental and social history can be, though it is crucial in first encounters for the clinician to give patients a sense that one is interested in their personhood and the contexts of their distress, not just their symptoms. Often in the first interview, an outline is obtained that is further developed in subsequent encounters. Table 7.1–4 outlines topics for developmental and social history.

The developmental and social history reviews the stages of the patient's life from gestation to the present to understand the essential exposures, relationships, and events that shaped the person's life story. One is interested in understanding the nature of the person's temperament and character and how the person has achieved developmentally appropriate role functions such as academic progress, work, peer and romantic relationships, and parenting capacity. Gestational and birth history, developmental milestones and early childhood development, the family of origin, cultural identifications, educational, occupational, legal, and military histories are explored. Histories of abuse (emotional, physical, and sexual), neglect (emotional and physical), and specific traumatic exposures are essential for a specific inquiry. The nature of the patient's current social environment is defined, including financial status, housing, and current relationships.

It is often helpful to review the social history chronologically; doing so provides a natural flow to the questions and ensures a complete history. Depending on the interviewer's assessment of the importance of missing information and assessing the credibility of the patient's account, it may be necessary to consult collateral historians to obtain a clear developmental and social history.

Table 7.1–4.
Social and Developmental History

1. Pregnancy and delivery
2. Developmental milestones
3. Educational (including history of special needs, in-school counseling, disciplinary problems)
4. Occupational history
5. Military
6. Legal
7. Relationships/committed relationships
8. Parenting
9. Leisure
10. Traumatic or potentially traumatic exposures (including neglect, physical and sexual abuse)

Review of Systems

As in a general medical interview, the review of systems is intended to capture any current physical signs and symptoms not already identified in the HPI or past medical history (including Table 7.1–2 and is organized by asking sentinel questions about the major systems of the body).

Mental Status Examination

The MSE is the functional equivalent of the physical examination in other areas of medicine. It is a systematic collection of the observations (e.g., signs such as blunt affect or rapid speech) and reported mental experiences (e.g., symptoms such as depressed mood or hallucinations) that produce a picture of the patient's current mental state. The interviewer makes these observations throughout an encounter and ultimately documents the findings together in the MSE section of the evaluation document.

Table 7.1–5 lists the traditional elements of the MSE along with examples of potential abnormal findings. The following sections offer commentary about each item of the MSE and model statements to illustrate how these findings might be documented in a medical record.

Appearance and Behavior

This section of the examination archives findings related to the patient's cooperativeness and behavioral activation levels, manner of relatedness, interaction, and state of grooming and dress.

> The patient appeared her stated age, was dressed eccentrically, and was poorly groomed. She seemed tense, distracted, and difficult to structure but was otherwise cooperative with the interview.

Table 7.1–5.
Elements of the Mental Status Examination with Examples of Abnormal Findings

Appearance and behavior: Odd or eccentric dress, poorly groomed, disheveled, unkempt, difficult to engage, evasive, indifferent, defensive, seductive, hostile

Eye contact: Avoidant, excessively prolonged, intense

Motor activity: Reduced activity, restlessness, agitation, posture and gait abnormalities, praxis problems, mannerisms and other stereotyped behaviors, posturing, tics, tremor, choreic, athetoid, and dyskinetic movements

Mood: Sad, irritable, angry depressed, elevated, expansive, euphoric, elated, dysphoric, anxious, incongruous with affect

Affect: Restricted, constricted, blunt, flat, labile, inappropriate to mood

Speech: Nonspontaneous, mute, rapid or slowed, low or high volume, abnormal pitch, tone or tempo, monotonous, robotic, stilted, excessive or reduced production, paraphasia, word approximations, poverty of content, dysarthria

Thought process/form: Derailment, tangential, circumstantial, flight of ideas, incoherence, blocking, perseveration, clanging

Thought content: Obsession, delusion, magical thinking, overvalued ideas, ideas of reference or influence, persecutory ideas

Perception: Hallucination, misperceptions (e.g., illusion), depersonalization, derealization, *jamais vu*, *déjà vu*

Risk assessment: Suicidal or violent ideation, intent, plan

Cognition: Impairments of orientation, attention, memory, language functions, calculation, visual-spatial capacity, following commands, impairment of intelligence, impairment of abstraction

Insight: Impairment along a continuum

Judgment: Impairment along a continuum

Eye Contact

Humans are keenly aware of eye contact and deduce a good deal about another's state of mind from it. Eye contact and its modulation in the psychiatric interview offer essential clues to the subject's internal emotional state. For example, a depressed patient may make little or no eye contact, an anxious patient's eye contact may be intermittent, whereas a patient with psychosis may stare and maintain intense eye contact. Eye contact is typically regulated to modulate the intensity and prevent inadvertent dominance displays. If a patient is especially intense or overtly paranoid, it is helpful to avoid prolonged eye contact.

> Eye contact was intermittent. When she made eye contact, it was inappropriately prolonged and intense.

Motor Activity

This examination section offers observations about the patient's general level of motor activity, posture, and gait and identifies the presence of abnormal movements such as tics, mannerisms, posturing, choreic, athetoid, and dyskinetic movements. If neurologic screening is conducted, such as testing muscle tone and deep tendon reflexes, these findings are archived here. In addition, many clinicians use this space to document the Abnormal Involuntary Movement Scale (AIMS) examination for tardive dyskinesia.

> Her gait was slowed and unsteady. When sitting, she was restless and rubbed the top of her head repetitively. On AIMS examination, she had mild rolling movements of her tongue at rest that were accentuated when her tongue was extended.

Mood

The terms mood and affect vary in their definition, such that one term often merges with the other. Several authors have recommended combining the two elements into the "emotional expression."

Traditionally, mood is defined as the patient's internal and sustained emotional state. Its experience is subjective, and hence it is best to use the patient's own words when describing mood. *Normal, calm, sad, happy, angry, irritable*, and *anxious* are examples of descriptors patients may offer.

> She said her mood was "calm."

Affect

Affect differs from mood in that it is the outward manifestation of a mood state visible to others. Mood is deduced to some extent from the person's emotional display. Some terms used to describe the quality (or tone) of a patient's affect include *dysphoric, happy, excited, euthymic, irritable, angry, tearful, blunt, and flat.* Speech provides affective information through volume, tone, and word choices. Quantity of affect is a measure of its intensity. For example, two patients both described as having depressed affect can greatly vary when one is mildly depressed and the other is severely depressed. Terms such as restricted, constricted, normal, or labile describe the affect range. Affect that correlates with the patient's stated mood and expressed thoughts is said to be appropriate or congruent to mood or thought content, whereas affect that contrasts sharply is said to be inappropriate with or incongruent to mood or thought content.

> Her affect was labile, ranging from inconsolable tearfulness to wild displays of irritability accompanying hostile speech, and was incongruent with her claim that her mood state was "calm."

Speech

Speech has an expected range of rate and volume and elements of pitch, tone, and tempo that are recognized within a normal range though cultural variation in these characteristics must be considered. It usually flows in a relatively linear manner conveying information efficiently. Strong emotions, such as intense anxiety, can lead to transient speech abnormalities. Certain psychiatric conditions, such as mania and schizophrenia, often have characteristic speech abnormalities that affect the linear flow of speech and its effectiveness in communication. The psychiatric interviewer should be alert for evidence of aphasic speech, which can be mistaken for disordered speech due to a psychiatric condition. Patients with speech disorders are often unaware that the communicative value of their speech is failing and may not show reciprocity in the speech that is characteristic of the interactive nature of spoken communication.

> At times, she was mute. When she did speak, her speech was excessive, difficult to interrupt, and came in rapid, loud bursts that made most of her communication ineffective.

Thought Process/Form

Thought process, also referred to as thought form, is how ideas are connected to convey thoughts to a listener or a reader in the case of written speech. Typically, ideas string together in a relatively linear pattern obeying conventions of grammar and syntax. Disordered thought takes many forms. Some descriptive terms include *circumstantial* (an inordinately circuitous route from question to answer), *tangential* (an answer to a question that veers off from the target of the question, but the connection may still be appreciated or inferred), and *derailment*, also called *loose associations* (thoughts proceed from idea to idea in a manner so oblique that the listener cannot follow the train of thought). Table 7.1–6 contains examples of named thought disorders and their typical characteristics.

Occasionally, her speech was linear and conveyed information well, but for the most part, she was quite disorganized with tangential and circumstantial speech. She was hard to interrupt during periods of flight of ideas and became frustrated when the interviewer tried to clarify what she was trying to say. At one point, she used an apparent neologism. For example, when asked how she was transported to the hospital, she said, “In a convestation, you fool”!

Table 7.1–6.
Examples of Disordered Thought Process/Form

Clanging: Use of the sounds of words rather than their meaning or conventional application to determine the connection between ideas

Circumstantial: Circuitous and overly detailed responses

Derailment (loose associations): Convoluted and random connections between ideas such that communicative value is lost

Flight of ideas: Derailment in the context of rapid speech

Incoherence (word salad): Complete disregard for conventions of word usage, grammar, and syntax resulting in incoherence

Neologism: A word created by the speaker with idiosyncratic meaning

Tangential: Responses that have only partial or remote connection to the original idea

Thought blocking: Loss of the goal of a communication and not being able to return to the topic

Thought Content

Thought content refers to the general themes of the patient’s ideation and is an area of the examination that identifies thought abnormalities such as obsessional thoughts and delusional ideation. The extent to which the examiner gains access to the patient’s private ideational life is determined by factors such as the patient’s comfort with the examiner and interviewing techniques to elicit this content.

Some patients share a broad range of content spontaneously and may even require structure to restrain the field of information they offer, while others perseverate or ruminate on specific thought content and require prodding to move beyond a narrow focus. Some patients guard content carefully. Guarding of thought content has many roots, including embarrassment or shame about the ideation, as is often the case with obsessional thoughts and paranoia that spills over into distrust of the interviewer.

A patient may reveal abnormal thought content freely, for example, complaining vigorously about persecution as a chief concern. At other times, gaining access to abnormal ideation is a delicate dance, especially when the patient is anxious about the consequences of revealing such material. Identification and exploration of delusional ideation often require particular care and respectful inquiry. A delusion is a false belief that the patient holds firmly despite evidence to the contrary and that is outside the normal range of beliefs of the patient’s culture or subculture. Common types of delusions are listed in Table 7.1–7. Delusions may be divided into two main groupings. One grouping distinguishes whether the idea is bizarre or not and the other whether the idea is mood-congruent or incongruent. Bizarre delusions are those that no one in the culture would consider a reasonable belief, whereas nonbizarre delusions are plausible but unlikely to be true. Mood-congruent delusions have themes consistent with the patient’s dominant mood state, for example, a depressed patient’s certainty that he has committed a mortal sin, whereas mood-incongruent delusions have no reference to a mood state, for example, the belief that aliens have implanted a device in one’s tooth through which thought is controlled.

A gentle exploration of the degree of conviction about a delusional belief can test the patient’s capacity to consider the possibility that their conviction is flawed. If one confronts the legitimacy of a delusional idea, this should only be done when there is a firm enough relationship for the patient to tolerate challenge and respect that delusions often serve an explanatory role. It may be hard for the patient to consider the possibility that the idea misrepresents reality, for doing so may require confronting an intolerable alternative truth.

Table 7.1–7.
Examples of Delusional Thought

Bizarre: A delusional belief that members of one's culture consider impossible

Erotomanic: Belief that another person is in love with one, usually someone perceived as having higher status

Grandiose: Belief that one has special powers, influence, or a special relationship with a deity or famous person

Identity: Belief that the identity of oneself or another person has been altered, replaced, replicated, for example, the Capgras syndrome.

Influence: Belief that someone of something outside one's self is influencing one to think, feel, or act in a certain way

Persecution: Belief that one is the subject of persecution, attack, harassment, or is being conspired against

Reference: Belief that an event, person, or object that is unconnected to the person has some special meaning or significance to the person

Somatic: Belief that a physical sensation or physical change is occurring to one's body caused by a source other than a normal biologic or pathophysiologic process

Thought broadcasting: Belief that the individual has thoughts that are produced by someone or something outside oneself

Thought insertion: Belief that the individual has thoughts that are produced by someone or something outside oneself

A delusion is a firmly held, false belief based on an incorrect inference that is unshakeable despite evidence to contradict it. It cannot be simply a cultural belief.

Consider the example of a young man with new onset of schizophrenia who is failing his college coursework. He attributes academic problems to a conspiracy among peers and faculties he is sure are modifying his work to ensure failure because his intellect threatens them. To consider the possible truth that there is something desperately wrong with his mind now interfering with his cognitive capacities and derailing his academic progress may be too much to bear.

When the psychiatrist inquired about his concerns about fellow students and his professors, he said, “Certainly you can see as well as I do that the other students and faculty members are threatened by me. They are colluding to undo me, to push me out of the way. No, I do not think this could be a trick of my mind, and the fact that you would even bring that up makes me wonder if you too are a part of this effort to ruin my academic career.” This appeared to be a delusion with persecutory and grandiose themes.

Perception

Perceptual abnormalities include phenomena like *hallucinations, illusions, depersonalization*, and *derealization*. Table 7.1–8 offers names and definitions for some of the more common types of perceptual disturbances.

Hallucinations are perceptions in the absence of external stimuli to account for them; they may occur in any of the five senses (auditory, visual, gustatory, olfactory, and tactile) and seem as authentic in the person’s experience as an accurate perception. Auditory hallucinations are the most common type of hallucination in nonorganic psychiatric conditions. In North American culture, the finding of nonauditory hallucinations should pique curiosity about an organic cause rather than a mental disorder. While visual hallucinations do present in North American culture (usually accompanying auditory hallucinations), some research in other cultures finds visual hallucinations to be the most common form of hallucinations in schizophrenia. The interviewer should be careful to distinguish between a true hallucination and a misperception of actual sensory stimulus (i.e., an illusion) and other phenomena, like hypnagogic and hypnopompic hallucinations, that are sleep–wake phenomena and a sign of a sleep disorder, not a psychotic illness. Most everyone can point to an example of a misperception or fleeting perception, for example, hearing one's voice called when no one else is home. These phenomena should not be mistaken for a hallucination, much less evidence of psychosis.

Table 7.1–8.
Examples of Perceptual Abnormalities

- Déjà vu: Perception that a present circumstance is the duplicate of an experience that has occurred in the past
- Derealization: Perception that one's surrounding and events are experienced as if the person is detached from them, or that they are distorted, changed, or unreal
- Depersonalization: Perception that one is standing outside oneself as a detached observer to surroundings, experiences, and events that occur
- Hallucination: The experience of a perception that seems genuine yet occurs without an actual external stimulus. The perception may occur in any of the five senses: auditory, gustatory, olfactory, tactile/somatosensory, or visual
- Illusion: Inaccurate perception or interpretation of an actual perception
- Jamais Vu: Perception that a present experience is entirely foreign when it actually should be very familiar

When documenting perceptual abnormalities, the interviewer should provide some description of the patient's experience, the context(s) in which the experience occurs, its frequency and intensity, the degree of the patient's conviction of the reality of the perception, and the degree of discomfort it causes, as well as any steps the patient may have taken to alleviate adverse effects of the experience.

In the case of auditory hallucinations, details of the nature and character of the voices can be critical to diagnosis: Does the patient hear one or several voices, simple statements, or complex sentences? Do the voices engage in a conversation or comment on the patient's thoughts? Generally speaking, the more complex the hallucinations, the more likely they are due to a schizophrenic spectrum condition. Whether the voices urge or command the patient to actions, including suicidal or violent behavior, is crucial to risk assessment.

As with delusions, the character of hallucinations is often divided into those that are mood-congruent (e.g., a depressed patient hearing a voice chiding her for failure and urging her to suicide) and mood incongruent (e.g., a patient with schizophrenia who, despite being quite paranoid hears female voices making warm, sexual comments that he finds arousing and soothing).

> The patient is experiencing hallucinations. She said, "Yes, I do hear a voice that keeps repeating statements like "you have failed as a mother, you will lose your family, you might as well be dead.""

Risk Assessment

Suicidal, violent, and homicidal ideation fall under the category of thought content, but many interviewers document these findings in a separate section of the examination labeled risk assessment. Risk assessment is an essential aspect of any initial clinical psychiatric evaluation.

Developing an in-depth understanding of high-risk ideation requires something more than simply asking the subject, "Are you suicidal or homicidal?" This approach will be an improper use of the closed-end question if one wants to know about this type of ideation. Ideas about suicide and violence are among the most private of thoughts. Patients are often hesitant to divulge ideation for fear of how the revelation will be received and the clinician's actions. Most will take any easy off-ramp to avoid it. A helpful approach begins with a normalizing statement followed by a question. For example, “In my experience, people who are struggling with the kinds of emotions and life problems you have described to me are also having ideas about death, dying, or taking their own life. Has this been the case for you?” A complete inquiry into risk addresses the presence of ideation and the level of intent, plans, and preparations that the patient has considered. Questioning about suicidal and violent ideation generally flows best by moving along the continuum from ideation to action.

Because many patients will withhold specific information about recent suicidal behaviors or suicidal ideation, Shawn Shea recommends a technique called the chronologic assessment of suicidal events (CASE) approach (access at https://suicideassessment.com/the-case-approach/). This technique uses a sequence of specific behavioral questions to determine how close the patient was to a lethal attempt. For example, if a patient had suicidal ideation involving a gun, the psychiatrist might ask, “Is there a gun in the home?” If the answer is “yes,” then the interviewer continues with a series of follow-up questions until a “no” response is obtained: “Have you taken out the gun?” “Did you load the gun?” “Did you point the gun at yourself?” “How long did you point the gun at yourself?” “What stopped you from completing the attempt?” Any developed planning for or practicing suicide or violence should be viewed as a high-risk finding requiring immediate attention. The American Psychiatric Association Practice Guideline for the Assessment and Treatment of Patients with Suicidal Behaviors is one source for more detailed information about best practices in suicide assessment. The recommended interview methods are analogous to the CASE approach just described, highlighting the importance of hallucinations and delusions in promoting suicidal acts.

Insight

Insight is a frontal lobe/executive function providing the capacity to appraise whether one’s thoughts, feelings, behaviors, perceptions, and planned actions are appropriate and realistic, and by the capacity to reflect on how one’s presentation may be perceived and interpreted by others. In everyday clinical practice, the degree of insight is typically rated on a continuum from absent to full; many patients have partial insight. Furthermore, the degree of insight does not track neatly with the severity of the illness. For example, a person with dementia may be painfully aware of cognitive decline, while a person with a mild anxiety disorder may have little or no insight into their overreactions to a fear stimulus. When recording the level of insight, describing examples is more helpful than merely saying “intact.”

> Her insight appeared to be partial. She had some sense that her behavior was being perceived as abnormal but could not consider the possibility that she is ill in some way that requires medical attention despite the evidence for overt psychosis.

Judgment

Judgment is another executive function mediated by the frontal lobe that represents the capacity of the individual to appraise a situation or problem, consider and decide among alternatives, plan and execute a course of action and modify the course of action when necessary based on new inputs. At times, judgment is impaired despite adequate insight. For example, a patient may recognize that he is confused about money management, including the funds in his checking account, but decides to make a large purchase without clarifying his account balance.

Tests of judgment that do not require real-world decision-making generally offer little value to an accurate judgment assessment. Take the example of a patient just arrested for dancing nude down the middle of a busy street while yelling at passing cars in the throes of a manic episode. The patient may answer a question about what to do if one found a stamped envelope on the sidewalk with a socially acceptable response such as, “mail it,” but the clinician would not, therefore, assess judgment as good. For this reason, it is helpful to distinguish between the abstract application of judgment and demonstrable judgment. Most examiners rate judgment on a continuum from very poor to good.

Her social judgment in the context of the interviewer was fair. For the most part, she behaved well in the examination, but she did become loud at one point and called me a “fool.” Recent events in the community, for example, her disrobing in public, demonstrate very poor judgment. She stopped her medication weeks ago and is now unwilling to consider any psychotropic medication.

Cognitive Assessment

During the interview, the clinician will develop a history and make observations that inform the cognitive assessment. Vocabulary level and abstract reasoning capacity are rough surrogates for intellectual capacity. The use of cognitive screening tests can be quite helpful in defining areas of strength or deficit that may not be obvious in a casual conversation without using more specific probes.

Domains of cognitive function essential to the initial assessment include the level of alertness, orientation, attention/concentration, visual-spatial function, memory (registration and recall), calculation, receptive and declarative language functions, fund of knowledge, capacity to abstract, and executive functions including insight and judgment.

The amount of detail required in assessing cognitive function will depend on the purpose of the examination and findings that raise a concern about the possibility of cognitive problems, like a history of learning challenges or an apparent decline from a previous level of cognitive function.

Two bedside tests commonly used to screen for cognitive function are the mini-mental state examination (MMSE) and the montreal cognitive assessment (MOCA). Each requires about 5 to 10 minutes to administer. They are also scored on a 30-point scale providing cutoffs between normal range and levels of impairment. The MOCA supplants the better-known MMSE in many settings because it tests a broader range of functional problems, identifies mild cognitive impairment, and provides administration instructions in many languages. The MOCA is available as an open-source product accessed at www.mocatest.org though it now requires users to complete online training to gain full access to the tool, while the authors of MMSE currently assert copyright protection. Because of these copyright issues, several other screening instruments have been developed, such as the informant questionnaire on cognitive decline in the elderly (IQCODE) and the Saint Louis University mental status exam.

Physical Assessment

The extent of physical assessment will depend on the nature of the referral, the setting, and the history and MSE findings. In the outpatient setting, routine physical examination is not generally a part of the initial consultation, while it is often critical to an emergency department or inpatient setting. Psychiatrists do not usually conduct comprehensive physical examinations but may conduct focused examinations such as neurologic or thyroid examinations. The psychiatrist generally relies on the PCP to conduct the physical examination in the outpatient setting. It is helpful in the initial evaluation to record the date of the most recent physical examination and review recent laboratories that are available. All psychiatrists should be prepared to measure vital signs and conduct other measurements, like waist circumference, indicated when psychotropic medications are used. A home blood pressure cuff is a valuable purchase for patients seen exclusively or predominantly virtually. The AIMS is an example of a screening physical examination conducted at baseline and serially during treatment with antipsychotic medication to monitor abnormal movements like tardive dyskinesia. An in-person visit is often required to conduct the complete exam.

Reviewing and interpreting laboratory results and ordering relevant laboratory tests or special studies, such as neuroimaging, is a critical aspect of many psychiatric evaluations. Psychiatrists lacking confidence in interpreting test results should consult with medical providers that can interpret them and offer opinions about the significance of results in the context of other findings.

Closing the Interview

Inexperienced interviewers are often naive to the importance of the last 5 to 10 minutes of the first encounter and find the closing of encounters challenging. It is helpful to alert the patient to the remaining time: "*We have to end in about 10 minutes.*" This signal allows the patient to plan how to use the time, particularly if they have delayed or failed to address something significant. Directly asking the patient if there is anything more they want to share before the end of the encounter is a valuable technique to reduce the patient's risk of bringing up an important issue "at the door" that cannot be adequately addressed. Sometimes late interview disclosures are too complex to address in the remaining time. If so, this may indicate a second interview to complete the evaluation. It is essential to leave time to share opinions and recommendations and answer questions about the diagnosis and plans.

COMMUNICATION OF OPINIONS AND RECOMMENDATIONS

The Formulation

The psychiatric interview and examination culminate with a case formulation that pulls together the biopsychosocial data into a working hypothesis about diagnosis and recommendations for treatment. The formulation should include a brief summary of the relevant findings from the history and examination, including the psychosocial contexts in which the problem has developed, and comments on the relevant contributions to the presentation of personality function, medical problems, social stress, and other social and cultural factors.

Although the formulation typically comes at the end of an evaluation document, the hypothesis-checking process is a dynamic one that continues from the first contact with the patient to the summary of opinions. Since most psychiatric diagnoses represent syndromes that often have overlapping features and presentations, a differential diagnosis of possibilities other than the working diagnosis is usually found in the formulation. It usually includes plans to obtain other data (history, tests, studies) to sharpen the diagnosis when this is indicated. The DSM-5 is the standard to establish a criteria-based diagnosis in the United States. Some other areas of the world use ICD-11. The two systems crosswalk in coding frameworks.

Finally, the formulation should include a summary of the risk assessment with estimates of acute and long-term risk of suicidal or violent behavior and an opinion about the appropriate level of care that will lead to a safe and successful outcome.

Treatment Recommendations

Consultations initiated by the patient and those requested by other professionals for input or care collaboration will usually include treatment recommendations. However, certain types of evaluations, for example, disability determination, competency assessment, and other forensics-related assessments, may not.

When evaluation produces treatment recommendations, these are typically shared with the patient after the encounter in a manner consistent with the patient’s capacity to receive and process the information and with explicit discussion of matters relevant to informed consent for recommended treatment. Treatment discussions typically involve much psychoeducation about diagnosis, the nature, risks, benefits of recommended treatments, and information addressing stigma and adherence. It is wise to involve significant others in these conversations, especially if there are concerns that the patient may need assistance in processing information and making decisions to ensure that decisions are consistent with the patient’s best interest. Recommendations may include referral to other professionals, or peer supports such as the National Alliance for Mental Illness (NAMI), the Mental Health Association (MHA), or Alcoholics Anonymous (AA). In addition, plans for crisis contacts and supports are typically addressed. These counseling and coordination of care efforts should be documented in the medical record.

TECHNIQUES

This section of the chapter will review several specific techniques that aid in facilitating and expanding information acquisition referred to here as “facilitating” and “expanding” interventions. Also reviewed here are some interventions that may prove to be counterproductive and interfere with patients telling their stories, which can interrupt the development of the therapeutic alliance.

Facilitating Interventions

Facilitating interventions enable the patient to continue sharing their story and help promote a positive patient–doctor relationship (Table 7.1–9). At times, some of these techniques may be combined in a single intervention.

Table 7.1–9.
Facilitating Interventions

Reinforcement
Reflection
Summarizing
Education
Reassurance
Encouragement
Acknowledging emotion
Humor
Nonverbal communication
Silence

Reinforcement. Brief phrases such as “I see,” “Go on,” “Yes,” “Tell me more,” “Hmm,” or “Uh-huh,” all convey the interviewer’s interest in the patient continuing. These phrases should fit naturally into the dialogue.

Patient: *For the past two months, I've been waking up at about 4 AM, and I can't get back to sleep. I feel anxious, like something bad is going to happen. A lot of times, I feel bad all day; it's only about 8 PM when I begin thinking of going to bed that I feel a little bit better.*
Psychiatrist: *I see.*
Patient: *I used to be a good sleeper. This seemed to come out of nowhere. It's a miserable feeling; I can tell you that.*

Although seemingly simplistic, reinforcement interventions are critical in encouraging the patient to continue sharing material. Without these reinforcements, the interview often becomes less productive.

Reflection. By using the patient's words, the psychiatrist indicates that they have heard what the patient is saying and conveys the implicit message that they are interested in hearing more.

Patient: *I don't know what's happening. I don't like going in to work anymore. The other guys at work are really starting to bug me.*
Psychiatrist: *Really starting to bug you.*

This response is not a question. A question with a slight inflection at the end calls for clarification (see below). The reflection is not stated with a challenging or disbelieving tone but rather as a statement of fact: the patient's experience is what it is, and the psychiatrist hears it. Sometimes it is helpful to paraphrase the patient's statement so it does not sound like it is coming from an automaton. However, in the example provided, the interviewer is unclear what the patient meant by "bug me," so changing the words (other than the pronoun) may steer the patient in a different direction.

Summarizing. Periodically during the interview, it is helpful to summarize what the patient offered about a specific topic. This approach provides the opportunity for the patient to clarify or modify the psychiatrist's understanding and possibly add new material.

Psychiatrist: *So, as I understand it, ever since you got a new boss this spring, things began to happen. You began to feel more anxious, and some of the comments of your coworkers really bothered you.*

Patient: *Yeah, and now that I think about it, around that time, my wife started complaining I wasn't fun anymore.*

New material has been introduced. The psychiatrist may decide to explore the previous discussion further and return to the issue concerning the patient's wife later.

Psychiatrist: *I want to hear about how things were between you and your wife, but before we talk about that, is there anything else you can say about how you felt at work?*

Education. At times in the interview, it is helpful for the psychiatrist to educate the patient about the interview process.

Patient: (after considerable hesitation) *There are some problems at home, but I don't know if that's what I'm supposed to be talking about.*

Psychiatrist: *It's helpful to talk about whatever has been bothering you. If I think we're getting off track, I'll let you know. Tell me about the problems at home.*

If this is not the first session and the patient has generally been sharing information, it might be helpful to focus on the hesitation.

Psychiatrist: *It seems difficult for you to mention that. Why do you hesitate to talk about the problems at home?*

This may lead to a discussion of confidentiality, loyalty to family members, or the nonjudgmental stance of the psychiatrist.

Reassurance. It is often appropriate and helpful to provide reassurance to the patient. For example, accurate information about the usual course of an illness can decrease anxiety, encourage the patient to continue to discuss their illness, and strengthen their resolve to continue in treatment.

Patient: *I don’t think I’ll ever feel better.*

Psychiatrist: *I understand how hopeless it feels for you right now. Feelings like that are common in depression, but most people with this type of depression do get better, and I think it’s very likely you will also.*

It is generally inappropriate to reassure the patient when the psychiatrist cannot predict with some confidence what the outcome will be. In these cases, the psychiatrist can assure the patient that they will continue to be available and help in whatever way they can.

Encouragement. It is difficult for many patients to come for a psychiatric evaluation. Often, they are uncertain about what will happen, and receiving encouragement can facilitate their engagement.

Patient: *I’m not doing very well describing this nervousness. I’ve never done this before.*

Psychiatrist: *I think you are doing well in describing the nervousness. As you talk, I’m getting a clearer picture of what it’s been like for you.*

The psychiatrist is careful not to overstate the progress in the interview. The patient is given positive feedback about their efforts, but the secondary message is that although the “picture” is getting “clearer,” more work is needed.

Acknowledgment of Emotion. The interviewer needs to acknowledge the expression of emotion by the patient. This acknowledgment frequently leads to the patient sharing more feelings and being relieved that they can do so. Sometimes a nonverbal action, such as moving a tissue box closer can suffice, or be used as an adjunct behavior to an empathic comment.

Patient: *He was a good friend.*
Psychiatrist: *As you talk about him, you look very sad.*

If the display of the emotion is clear (e.g., patient openly crying), then it is not helpful to comment directly on the expression of the emotion.

Patient: (sobbing) *I really miss him.*
Psychiatrist: *I see that you are crying.*
Patient: *You’re very observant.*

It is better to comment on the associated feelings.

Psychiatrist: *You feel awful without him.*

Humor. At times, patients make humorous comments or tell a joke. It can be beneficial if the psychiatrist smiles, laughs, or even, when appropriate, adds another punch line. This humor sharing can decrease tension and anxiety and reinforce the interviewer’s genuineness. It is vital to be sure that the patient’s comment was indeed meant to be humorous and that the psychiatrist conveys that she is laughing with the patient, not at the patient.

Silence. Careful use of silence can facilitate the progression of the interview. The patient may need time to think about what has been said. They also may need time to experience a feeling that has arisen in the interview. The psychiatrist who cannot bear the anxiety of silence may terminate it prematurely retarding the development of insight or the expression of feeling by the patient. As an experienced interviewer encouraged a new trainee, “Don’t just do something, sit there.”

On the other hand, extended or repeated silences can deaden an interview and lead to a struggle as to which party can outwait the other. If the patient is looking at their watch or looking about the room, then it might be helpful to comment, “It looks like there are other things on your mind.” If the patient has become silent and looks like they are thinking about the subject, then the psychiatrist might ask, “What thoughts or feelings do you have about that?”

Nonverbal Communication

Common facilitating interventions are often nonverbal. Nodding of the head, body posture including leaning toward the patient, body positioning becoming more open, moving the chair closer to the patient, putting down pen and folder, and facial expressions including arching of eyebrows all indicate that the psychiatrist is concerned, listening attentively, and engaged in the interview. Virtual interviewing interferes with some nonverbal techniques. Excessive repetition or exaggeration of nonverbal techniques risks being perceived as a caricature of the psychiatrist nodding their head repeatedly regardless of what is said.

Expanding Interventions

Several interventions can be used to expand the focus of the interview. These techniques (Table 7.1–10) are helpful when the line of discussion has been sufficiently mined, at least for the time being, and the interviewer wants to encourage the patient to talk about other issues. These interventions are most successful when the psychiatrist has established meaningful trust and the patient feels they are nonjudgmental.

Table 7.1–10.
Expanding Interventions

Clarifying
Associations
Leading
Probing
Transitions
Redirecting

Clarifying. Carefully clarifying what the patient has said can lead to unrecognized issues or psychopathology.

> A 62-year-old widow describing how it feels since her husband died 14 months ago repeatedly comments, “Everything is empty inside.” The resident interprets this to mean that her world feels empty without her spouse and makes this interpretation on a few occasions. The patient’s nonverbal cues suggest that she disagrees with this view. The supervisor asks the patient to clarify what she means by “empty inside.” After some avoidance, the patient states that she is indeed empty inside; all her organs are missing; they have “disappeared.”

The resident’s interpretation may have been psychodynamically accurate, but the interpretation interfered with the opportunity to discover a somatic delusion. Correctly identifying what the patient was saying led to exploring other thoughts and uncovered delusions. This vignette of “missing” the delusion exemplifies the interviewer “normalizing” what the patient is saying. The interviewer used secondary process thinking to understand the patient’s words, while the patient used primary process thinking.

Associations. As the patient describes symptoms, other areas that may accompany or relate to a symptom should be explored. For example, hearing about the patient's nausea leads to questions about appetite, bowel habits, weight loss, and eating habits. In addition, experiences that are temporally related may be investigated. For example, when a patient talks about their sleeping pattern, it can be a good opportunity to ask about dreams.

Leading. Often, encouraging the patient to continue their story can be facilitated by asking "what," "when," "where," or "who" questions. "Why" questions are generally not helpful early in an interview.

Patient: *And I said,* "*That's enough.*" (pause)
Psychiatrist: *What happened then?*

Sometimes the psychiatrist may suggest or ask about something not yet introduced by the patient because the psychiatrist surmises it may be relevant. In the courtroom, leading the witness is sure to raise an objection, but in clinical work, it can be helpful as long as the interviewer is not making too far an intuitive leap and is open to reframing the question pending a response by the patient.

Patient: "*…feels like my husband is always telling me what to do, criticizing my driving.*"
Psychiatrist: *You mentioned that earlier. Have there been other relationships where you have also experienced that?*
Patient: *Well…*(pause) *actually my father used to do that to my mother all the time. I really hated that.*

Probing. The interview may point toward an area of conflict, but the patient may minimize or deny any difficulties. Gently encouraging the patient to talk more about this issue may be quite productive.

Psychiatrist: *Tell me more about how things are at home.*
Patient: *Not really a whole lot to tell; everything is cool.*
Psychiatrist: *You mentioned that weekends are difficult.*
Patient: *I didn't mean that difficult.*
Psychiatrist: *OK. How are they different from before?*
Patient: *I don't know what you mean.* (pause) *Well, I guess it's a little different. Now my wife does her things, and I do mine.*

Transitions. Sometimes transitions occur very smoothly. For example, the patient is talking about their major in college, and the psychiatrist asks, "*Did that lead to your work after college?*" On other occasions, the transition means moving to a different interview area, and a bridge statement is helpful.

Psychiatrist: *That gives me a good idea about your nervousness. Perhaps now you can tell me about your health in general.*

Redirecting. A challenging technique for unseasoned interviewers is redirecting the focus of the patient. It can be tricky to move the interview in a different direction when reinforcing the patient telling their story. However, the reality of time constraints and obtaining a broad overview of the patient's life and the current problems make redirection necessary at times. Redirection is also used when the patient changes topics too quickly or persists in offering excessive information about a nonproductive or already well-covered area. For conscious and often unconscious reasons, the patient may avoid certain vital areas and need guidance in approaching these subjects.

Patient: (After beginning to describe some significant issues with her mother) *But enough of that, let me tell you about my job.*

Psychiatrist: *Before we move on to your job, perhaps you can say more about your struggles with your mother. It sounds like that has been very upsetting to you.*

Patient: (After much discussion about her arthritis) *And then six months ago, I saw a new rheumatologist…*

Psychiatrist: *I know your arthritic pains have been a big burden to you, and we can come back to that, but I would also like to hear more about the issues you mentioned with your daughter.*

Sometimes if the patient is wandering and does not respond to an attempt at transitioning, it is necessary to be more directive.

Psychiatrist: *I'm going to interrupt at this point because I am aware of the time we have left, and there are several other areas it would be good to talk about.*

Obstructive Interventions

While supportive and expanding techniques facilitate gathering information and developing a positive patient–doctor relationship, some other interventions are not helpful for either task (Table 7.1–11). Sometimes these interventions are technically correct, but if introduced in an unclear manner, unconnected to content, or poorly timed, they may be experienced as unresponsive to the patient's concerns or feelings.

Closed-Ended Questions. A series of closed-ended questions early in the interview can hinder the patient's story's natural flow and reinforce the patient providing one-word responses or brief answers with little or no elaboration.

Psychiatrist: *Did you have a happy childhood?*
Patient: (Pause) *Yes, I guess so.*
Psychiatrist: *Did you have friends?*
Patient: (Intuitively trying to get the interview pointed in a more productive direction) *Well, I guess it depends on what you mean by a friend.*
Psychiatrist: (Not joining the patient's attempt to enrich the discussion) *Kids you did things with. How many?*

This example illustrates that the patient can be a partner in the interview unless blocked by the psychiatrist.

Many patients, some of whom have previous experiences in therapy, come prepared to talk about even painful matters. Psychiatrists, particularly with supervision, learn from patients and refine their interviewing skills with time.

Compound Questions. Some questions are difficult for patients to respond to because more than one answer is being sought.

Psychiatrist: *How did you feel? What did you do?*
Patient: *I'm not sure what happened.*

Table 7.1–11.
Obstructive Interventions

Excessive use of closed-ended questions
Compound questions
Excessive “Why”? questions
Judgmental questions
Minimizing patient concerns
Premature advice
Premature interpretations
Abrupt transitions
Ambiguous nonverbal communication

Why Questions. Especially early in the psychiatric interview, “why” questions are often nonproductive. Often, seeking an answer to a “why” question is one of the reasons that the patient has sought help.

Patient: *I felt very depressed. I just couldn’t go on.*
Psychiatrist: *Why were you so depressed?*
Patient: *I don’t know. I thought you might know.*

Judgmental Questions or Statements. Judgmental interventions are generally nonproductive for the issue at hand and inhibit the patient from sharing even more private or sensitive material.

Patient: *I felt she wasn’t paying attention, so I yelled at her.*
Psychiatrist: *Don’t you know that yelling only makes matters worse?*

It would be better for the psychiatrist to help the patient reflect on that behavior’s success.

Psychiatrist: *How did that go?*
Patient: *Not good. Then she got upset about the yelling.*

Minimizing Patient's Concerns. In an attempt to reassure a patient, the psychiatrist makes the error of minimizing a concern.

Patient: *I worry I'm going to have a heart attack. My doctor says I have to lose weight.*
Psychiatrist: *I don't think you need to worry. You look pretty healthy, and your father is still alive.*

Rather than being reassured, the patient may feel that the psychiatrist does not understand what he is trying to express and trivializes the experience. Furthermore, the advice is counter to what the PCP has said and confused the patient. Exploring the concern is much more productive; there is likely much more material not yet shared.

Premature Advice. Early advice is often flawed because the interviewer does not yet know all the variables. In addition, it can preempt the patient from arriving at a plan for themselves.

Patient: *My boss is so demanding.*
Psychiatrist: *Why don't you get in early in the morning before he does and give him a list of what will improve everyone's performance?*

Premature Interpretation. A premature interpretation can be counterproductive even if it is accurate, as the patient may respond defensively and feel misunderstood.

Patient: *I was so angry at my neighbor when he said that, and I'm not sure why.*
Psychiatrist: *He reminds you of your brother and the way he was always trying to control you.*
Patient: (angrily) *He doesn't remind me of my brother; he doesn't even look like him.*

Transitions. Some transitions are too abrupt and may interrupt important issues the patient discusses.

Patient: *Ever since my father died, I've been feeling anxious.*
Psychiatrist: *Tell me more about your job.*

Nonverbal Communication. The psychiatrist repeatedly looking at their watch, turning away from the patient, yawning, or refreshing the computer screen convey boredom, disinterest, distraction, or annoyance. Just as reinforcing nonverbal communications can be powerful facilitators of a good interview, these obstructive actions can quickly obstruct an interview and undermine the patient–doctor relationship.

SELECTED CHALLENGING INTERVIEW SCENARIOS

Patients with Psychosis

Patients with psychotic symptoms are often frightened, guarded, and suspicious about the purpose of the interview and the intentions of the clinician. They may have difficulty with reasoning, thinking, or communicating. Hallucinations may cause inattention and distraction. The psychiatrist must adapt to match the capacity and tolerance of the patient. Patients with psychosis may not perceive their behaviors, speech, delusions, or hallucinations as abnormal in any way, making it essential to inquire in a manner that encourages the patient to share their ideational and perceptual experience without seeming judgmental of it.

Toward this end, prefacing direct questioning about psychotic symptoms by making normalizing statements such as “Many people have ideas that others find strange or different,” can increase the patient’s comfort with sharing delusional ideation and similarly for hallucinations by stating that, “Many people have had the experience of hearing a voice or seeing a vision.” Such a statement may help the patient talk more openly.

The psychiatrist should be alert for cues that psychotic processes may be part of the patient’s productions during the interview. It is usually best to ask directly about any such behaviors or comments.

During an initial interview, a 28-year-old man suddenly stood up and turned 360 degrees, sat down, and after a brief pause, resumed talking. When asked, “What just happened?” the patient responded that he “had to unwind.” While exploring this example of concrete, primary process thinking, several other psychotic features were uncovered.

By definition, patients with delusions have fixed false beliefs. With delusions, as with hallucinations, it is crucial to try to explore the symptoms in some detail. Patients are often reluctant to discuss their experience because of past occurrences of being dismissed, ridiculed, or confined when they revealed their thoughts and perceptions. They may ask directly if the clinician believes the delusion. Interviewers are advised not to endorse the false beliefs, but it is rarely helpful to challenge the delusion directly, particularly in the initial examination. It can be helpful to shift the focus of attention back to the patient’s beliefs while acknowledging the need for more information by responding like this: “I believe that what you are experiencing is frightening and I would like to know more about your experiences so I can form an opinion.”

With patients evidencing a high degree of suspiciousness, it is helpful to maintain distance and modulate statements or behaviors that patients may interpret as promoting excessive closeness, such as excessive use of empathic statements or prolonged eye contact. Harry Stack Sullivan recommended that rather than sitting face to face with the paranoid patient, the psychiatrist might sit more side by side (but not too near), “looking out” with the patient. While interviewing the psychotic patient, the interviewer should consider that the patient might be incorporating the clinician into the psychotic process. It can be helpful to ask about this directly, for example, “Are you concerned that I am involved?”

Acute psychosis raises concerns about the potential for violent behaviors that are further reviewed in this section under the heading *The Potentially Violent Patient*.

The Patient with Depression

The depressed patient may have particular difficulty participating in the interview due to psychomotor slowing, cognitive deficits, and lack of motivation to engage owing to the depression. In addition, feelings of hopelessness may contribute to a lack of engagement.

> Patient: “*What’s the use (of talking)? It’s not going to make any difference. Nothing is going to change.*”
>
> Psychiatrist: “*I’m glad that you can share that. It sounds like you feel hopeless about getting better. Hopeless feelings are common in patients who have depression. I think your condition will improve, and you will feel more hopeful.*”

The depressed patient’s responses to questions are often latent and may lack spontaneous elaboration. Depending on the severity of symptoms, the interviewer may need to turn to more direct questioning and abandon an open-ended format. The interview of severely depressed patients can be a laborious process and may require more time than expected to gather even minimally necessary information. Even then, collateral historians may need to fill in gaps in the patient’s history.

The Potentially Suicidal Patient

There are times in initial interviews when it is apparent that one is encountering a patient on the brink of suicide, but in many other instances, patients have suicidal ideation without betraying much, if anything, in the way of clues. For this reason, it is vital to conduct a suicide risk assessment in an initial psychiatric interview. Specific diagnoses and problems heighten suicide risk: major depression, borderline personality disorder, high anxiety levels, acute psychosis, acute intoxication, patients who have become hopeless about their substance use problems, acute grief and loss reactions, and patients with grave or burdensome medical problems. In addition, persons recently arrested and patients who are homicidal may also be suicidal. As discussed in earlier sections of this chapter, normalizing the potential for suicidal ideation is a good way to encourage patients to share their thoughts and feelings. An invitation like this can help to initiate the conversation about suicide: “Many people who describe a life event or the feelings that you have today are also having thoughts about death, dying, or even ending their own life. Has this been the case for you?”

The science of suicide risk assessment is beyond the scope of this chapter, but there are important risk factors that elevate suicide risk to survey as the history is developed. Some relate to demographics (e.g., age, gender, living alone, race and ethnicity, sexual orientation), others to certain factors related to diagnosis and clinical features (see above), and life experiences such as childhood physical and sexual abuse. Family history of suicide, prior attempts at suicide, and acts of deliberate self-harm are important historical risk factors. In addition, active substance use, acute psychosis, high anxiety levels, or insomnia seem to fuel suicide risk. Protective factors include the level of social support, children in the home, pregnancy, and religious beliefs that admonish suicide but are in themselves not entirely reliable guardrails.

The Potentially Violent Patient

Safety for the patient, interviewer, and others present is the primary concern when engaging with hostile or agitated patients. If one knows in advance that an encounter with a hostile or agitated patient will occur, there is an opportunity to plan and marshal supports. As noted elsewhere in the chapter, thoughtful attention to managing personal space and having an attendant during the encounter can reduce risks in these situations. In addition, it is good practice to alert other clinical staff in the area and security if an officer is present.

Reducing excessive stimulation of already hostile or agitated patients is a priority. Interviewers should be aware of their body position and avoid postures, behaviors, eye contact, tone of voice, and choice of words that could be perceived as threatening. Approach the interview in a calm, direct manner and take care not to make bargains or promises to elicit cooperation in the interview, for example, implying that cooperation will avoid an admission. These tactics will only likely escalate agitation and the potential for aggression when the patient discovers the examiner cannot deliver.

Managing the interview requires thoughtful choices about areas to explore and those to leave alone and the phrasing and timing of questions to avoid stirring up unnecessary conflict. Taking breaks, agreeing to leave a topic that causes tension for another time, and attending to basic comforts like hunger and nicotine replacement before pressing on with the interview may reduce tension and build rapport. Avoid unnecessary displays of dominance and give patients choices where there are reasonable options to maintain some semblance of control.

Despite best efforts, some interviews must be terminated when the patient’s agitation escalates. Generally, preceding unpremeditated violence is a period of gradually escalating psychomotor agitation such as pacing, loud speech, and threatening comments. As noted earlier in the chapter, some agitated patients will be calm when the interviewer shares that the patient’s behavior makes the interviewer too anxious to be helpful. However, there is no virtue in pressing on and taking unnecessary risks, especially when the interviewer is frightened. Instead, consider terminating the interview and seeking assistance if the patient’s behaviors escalate and do not readily respond to structure and feedback.

The hostile or agitated patient may be placing others in his community at risk. The most common victims of violence by acutely ill psychiatric patients are those who live or work close to the patient, such as family, residential staff, and coworkers. If the patient can tolerate it, explore feelings and attitudes about those close to him to discover persons who may be at particular risk. If the patient makes specific threats or identifies specific targets, take the threat seriously and carefully define the actual risk.

Like the assessment of suicide risk, a complete discussion of the violence risk assessment is beyond the scope of this chapter. Specific facts are worth noting here. First, research reveals more about risk factors for violence than protective factors. Among the critical risk factors are patient age (youth), lower intelligence, past acts of violence and general suspiciousness, and active psychosis; individuals physically or sexually abused as children are also at greater risk of violence. Interviewing for violence risk follows much the same pattern as that used for suicide risk assessment. The interview generally moves along the continuum of ideation to plan to action exploring the nature and extent of the patient’s ideation, planning, and intent. Like suicide assessment, the clinician seeks to define past instances of violence. This includes assessing the particulars of these episodes and the outcomes to understand the risk level during the interview and identify patients at elevated risk over time.

The Patient Engaging in Deception

For this discussion, deception is defined as the intentional production or exaggeration of symptoms. Psychiatrists train to maintain vigilance for deception while eliciting information, but they are not immune to being deceived. Studies show that even experienced clinicians will fail to recognize malingered mental illness in some settings. Patients engage in deceit for many different reasons. Some are motivated by secondary gains, like work excuses, disability, or to secure desired medications. Others seek primary gains, like the psychologic benefits of assuming a sick role.

A soldier in basic training presents to the military base mental health service with a long list of symptoms. He escalates the drama of his responses each time he perceives that the psychiatrist is not reacting to his productions as he expected. Finally, after describing some elaborate visual hallucinations that do not comport to those typically experienced by psychotic patients and that also failed to impress the psychiatrist, the soldier blurts out, “What’s the matter? You never heard of paranoid schizophrenia?” Later, the psychiatrist learned the soldier visited the base library to read about psychiatric illnesses and schizophrenia. The soldier admitted he was feigning symptoms to avoid an undesirable deployment.

There are no infallible means to detect deceit. However, there are markers in interviews that raise suspicion, including vague and contradictory responses, symptom claims that are inconsistent with functional findings, claims that are far beyond the variations of psychopathology seen in certain conditions or problems, or that are frankly preposterous. View skeptically patients that draw excessive attention to symptoms and those that become irritated with attempts to clarify the nature of complaints. This said, it is essential to remain cognizant that there are unusual presentations, and one should not assume that all that confuses or does not fit typical patterns is, therefore, evidence of deceit. Collateral historians become critically important when there is concern about the credibility of patient reports. Specific psychological tests are designed to detect deceit (e.g., the Word Memory Test), and others have validity scales imbedded in their structure that measure the nature of response style to identify likely exaggeration or suppression of symptoms in reporting (e.g., the Minnesota Multiphasic Personality Inventory [MMPI]).

It is critical to be confident in one’s opinion about deceit before confronting a patient with this conclusion. It is also essential to maintain an empathic approach when discussing findings, ideally helping the patient save face while accepting responsibility.

The following is an example of an exchange with a potentially malingering patient.

> "Would you agree that it is important for you to be truthful with me about your history and symptoms and for me to be truthful with you about my findings and recommendations? Okay, then, to be a good doctor to you, I must share that I believe that you are exaggerating some of the symptoms you have described. These are the reasons why I believe this to be so (clinician explains the reasons). I am concerned that you are trying to present yourself as ill in certain ways that, in the end, could lead to bad outcomes, like applying treatments that could do more harm than good. If I am right about this, let us start over and talk about what is really going on and why you might have tried to present yourself as ill in this way. Then I can see if there are ways I can help you with your predicament that are reasonable and safe."

SPECIAL TOPICS

Note Taking and the Electronic Medical Record

Many psychiatric interviewers take handwritten notes during the interview. With the advent of electronic medical records, some advocate for a method called "contemporaneous documentation." This method involves pausing periodically during the encounter or at the end to create all or part of the consultation document with the patient's participation. Advocates of this method point out that a patient increasingly has direct access to all or parts of their medical record through patient portals. Writing the note with the patient, so to speak, reduces the chances of documenting inaccurate information or writing statements in a manner that the patient might find offensive.

This is an example of how contemporaneous documentation might proceed:

Psychiatrist: “Now you have told me a lot about your current problem, and I’d like to pause now and document what you said to summarize it in the record. I will summarize what I am writing while I type it. You listen and tell me whether I have heard you correctly.”

The process of contemporaneous documentation can result in synergistic efficiency that allows summarizing, clarifying, and correcting information while completing documentation. For those that now offer an hour for the consultation and then do documentation after the encounter, extending the consultation to include time for contemporaneous documentation may improve time economy.

Note-taking and contemporaneous documentation in an electronic record during the interview presents some risks. These activities can interfere with observing nonverbal information, like affects and body language, that are often as or more important than what the patient says. Failing to recognize these nonverbal cues can result in losses of essential data and empathic failures affecting the development of rapport. The activities can also distract the patient and the interviewer if they dominate the interview.

Cultural Variation

One definition of culture is a common heritage, beliefs, and values that set expectations for behaviors, thoughts, and even feelings. Culture influences the patient’s attribution of illness, decisions about when and where to seek help, how distress is articulated or demonstrated to others (including when speaking with medical providers), and perceptions of necessary and acceptable forms of treatment. Often, individuals from a minority population are reluctant to seek help from a physician from the majority group, especially for emotional difficulties.

Just as the array of cultural syndromes and variations is beyond the scope of this chapter, interviewers cannot be aware of all culturally bound syndromes or idioms of distress. However, they should make an effort to learn about variant presentations and syndromes that affect cultural groups commonly encountered in the scope of their practice settings. Reducing the risk of overdiagnosis (e.g., misreading a cultural belief for a delusion) and the possibility of over or undertreatment resulting from inaccurate diagnosis are important reasons cultural awareness is essential.

By remaining humble and curious about the patient's cultural identifications and beliefs and being open and respectful, the psychiatrist increases the likelihood of developing a trusting working relationship and a collaborative and effective plan.

Interviewing with an Interpreter

Translated interviews require forethought and preparation. A professional translator is favored over using a person familiar to the patient whenever possible. Professional translators have training and experience to interact with patients and providers skillfully, the vocabulary skills to translate medical words and concepts, operate under a code of ethics that highlights confidentiality, and understand the importance of verbatim translation and neutrality. The use of nonprofessional interpreters risks a poor-quality translation fraught with all the challenges of agency, skill, and interpersonal dynamics that friends and family bring to the process of interpretation and too often lead to incomplete or frankly inaccurate communication.

It is helpful to speak with the interpreter before the interview to clarify the examination goals and expectations. Most professional interpreters are attuned to the fact that their role is to facilitate, not create communication and understand the importance of verbatim translation and that the interpreter does not intersperse ad-lib questions or edit the patient's responses. If the interpreter does not primarily work with psychiatric patients, it may be helpful to highlight the need for verbatim translation even if the responses are disorganized or contain unusual content. During the interview, the interpreter may slip into giving a sparse interpretation of longer speech productions. If this occurs, the interviewer should pause and ask the interpreter to offer everything the patient said and remind the interpreter of the request for verbatim interpretation.

To avoid the patient connecting more with the interpreter than the interviewer, one should arrange the parties' positions to allow the patient to maintain gaze in the direction of both interviewer and interpreter during the encounter. Interviewers need to be sensitive to the pace of questioning and parse out the questions and comments into smaller segments to allow the interpreter time to facilitate communication. Generally, more than one or two sentences or three complete thoughts is more than anyone, including interpreters, can hold accurately in memory to reproduce a verbatim translation once the interviewer stops speaking.

After the interview, the interpreter who is exceptionally knowledgeable about the patient's cultural background may be able to provide helpful insights and opinions about the nature of the patient's communication, for example, if a formal thought disorder was present and whether ideas the patient expressed were outside the norm for the cultural group.

The functionality of many telehealth applications allows grouping the patient, interviewer, and translator together.

Observed and Recorded Interviews

One-way mirror observation or recording of interviews occur in some settings for clinical, educational, or forensic purposes. Before initiating an observation or recording, the patient's informed consent for these activities must be precisely established. The patient should understand the purpose of observation or recording (e.g., psychotherapy, supervision), how it will be used, and who may have access to it (e.g., an attorney or other third parties in a forensic case). When recordings are made for clinical purposes, it should be clear how long the recording will be archived before being destroyed. If a recording is to be used for educational purposes, there are special informed consent considerations, including the patient's right to withdraw the consent and other matters beyond this chapter's scope. The clinician must ensure that the patient receives the same nature and quality of care in treatment settings regardless of whether consent to observe or record is granted.

Observation or recording of an interview can perturb the clinical interaction in expected and unexpected ways. Therefore, the clinician should remain vigilant for untended adverse effects on engagement and the free flow of information when interviews are observed or recorded.

Forensic Interviews

For this section, a forensic interview is defined as one taking place in a legal context for a purpose other than clinical treatment. For example, the psychiatrist may be retained by an insurance company to conduct an independent medical examination (IME) to offer an opinion in a disputed disability case, or by an attorney representing or prosecuting a defendant in a criminal case to offer opinions about competency to stand trial or criminal responsibility. Even in forensic settings where a psychiatrist is acting in a clinical role, for example, providing treatment in a correctional facility, some of the issues to be discussed here will apply.

Except for those providing clinical care in a forensic setting, forensic consultants do not technically form patient–doctor relationships with subjects, yet should still operate in these settings observing the ethical principles that apply and aspire to maintain the subject's dignity.

The agency or duty of the psychiatrist in the forensic setting is typically owed to a third party, for example, the attorney that retained the psychiatrist. Before proceeding with the interview, the subject should understand the agency of the psychiatrist and the boundaries of confidentiality. For example, when the psychiatrist is retained by the subject's defense attorney, confidential material is protected by the attorney–client privilege unless and until the attorney and client decide to use an opinion, at which point a report or testimony may be provided requiring a waiver of confidentiality to produce it. If the examiner is retained by a prosecuting attorney, an insurance company in an adversarial proceeding, or a judicial order demands a report, the subject waives confidentiality by agreeing to the examination. Some recent court decisions expand the rights of forensic subjects to confidentiality privileges under HIPAA that may require specific, informed consent and considerations about sharing of reports and other products of the evaluation. These matters are beyond the scope of this chapter, but those operating in these environments should educate themselves or consult legal counsel to ensure their practices align with requirements.

The principles of the forensic interview are built on the foundation of sound clinical practices though there are some critical differences. The level of detail required in a forensic consultation is typically greater than that sought in a general psychiatric interview. Forensic interviewing techniques require more detailed probing of the patient’s story and mental state that may sometimes require more clarification and even confrontation than typically utilized in a general psychiatric interview. Such interventions should be conducted in an ethically sound and sensitive manner. Collateral sources of information are critical to forensic examinations. These sources help discover vital elements of the history that the subject may not provide and serve to verify information offered by the subject externally. Forensic interviews are much more likely to be recorded than general psychiatric interviews and require special recording considerations beyond the scope of this chapter, but the subject must know when recording is occurring and understand who may have access to such recordings as part of the consent process. Psychiatrists engaging in forensic interviewing should have a clear sense of the boundaries of their training and experience and recognize the implications of their activities and the opinions they offer. Knowledge of the legal system (including the legal standards that apply to clinical questions), skills in specialized interviewing techniques, the role and proper application of psychological testing, and the vicissitudes of engaging with attorneys and courts are examples of the specialized information and skills that forensic settings demand.

Telepsychiatry

The advent of the COVID-19 pandemic forced mental health clinicians to quickly adopt and adapt to the use of telehealth technologies regardless of pre-existing attitudes about its utility. Previously, these technologies were primarily utilized in underserved areas where the geographic distance from or reliable access to providers made it unrealistic for the patient to come to a live appointment. Now telepsychiatry has emerged as an essential alternative means for access to care, safe and comfortable care delivery in the face of the COVID threat and appears to be here to stay as an alternative to in-person visits. Research and opinion leaders on the topic of telepsychiatry indicate that it is generally well accepted and a much more successful and viable alternative than most of us in the field may have predicted.

One must be cognizant that technology is not pristine enough to protect against some dropout of information usually gleaned in a live interview. Televised observation and communication can degrade nuances of speech, subtleties of the nature of eye contact, emotional display, body language, and other forms of nonverbal communication. Causes of information degradation include the narrow perspective of the camera, the quality of the audio-visual product, and the unexplained effects that humans have on one another in a live interaction that are not entirely replicated in a distance medium. On the other hand, the availability of telehealth has expanded many patients' access to care, and many patients find the medium a more comfortable means to discuss difficult topics and appreciate the logistical advantage that reduces time off work, travel and waiting time, childcare challenges, and the like. Some others do not adapt to the virtual visit, and still, others may not have the right technology or facility with technology to use the medium effectively. Many of us who have utilized the technologies have come to appreciate that we are now essentially doing home visits and learning so much more about the patient's environments than we ever did before. It also makes meetings between the collaterals and the patient possible in logistically challenging cases, for example, the student at college and the parent at home.

The American Telemedicine Association and the American Medical Association are associations that collaborate with other professional organizations to develop evidence-based guidelines promoting quality and safe practices for various telemedicine applications, including telepsychiatry. Clinicians engaging in telemedicine are advised to familiarize themselves with the special considerations that this medium introduces, including medical-legal and reimbursement issues. Interstate use of telehealth is generally forbidden without special waivers or licensing considerations that vary from state to state.

FURTHER READINGS

American Psychiatric Association. *Practice Guideline for the Psychiatric Evaluation of Adults*. 3rd ed. American Psychiatric Association; 2016.

Balint M. *The Doctor, His Patient, and the Illness*. 2nd ed. International Universities Press; 1972.

Bokolo Anthony JNR. Implications of telehealth and digital care solutions during COVID-19 pandemic: a qualitative literature review. *Inform Health Soc Care*. 2021;46:68–83.

Chiles JA, Strosahl KD. *Clinical Manual for Assessment and Treatment of Suicidal Patients*. 2nd ed. American Psychiatric Publishing; 2019.

Cruz M, Roter DL, Cruz RF, et al. Appointment length, psychiatrists' communication behaviors, and medication management appointment adherence. *Psychiatr Serv*. 2013;64:886–892.

Dhalla S, Kopec JA. The CAGE questionnaire for alcohol misuse: a review of reliability and validity studies. *Clin Invest Med*. 2007;30:33–41.

Engel GL. The need for a new medical model: a challenge for biomedicine. *Science*. 1977;196:129–136.

Epstein RM. Whole mind and shared mind in clinical decision-making. *Patient Educ Couns*. 2013;90:200–206.

Folstein MF, Folstein SW, McHugh PR. Mini-mental state: a practical method of grading the cognitive state of patients for the clinician. *J Psychiatry Res*. 1975;12:189–198.

Fong T. Assessment. In: Brady KT, Levin FR, Glanter M, Kleber HD, eds. *Textbook of Substance Abuse Treatment*. 6th ed. American Psychiatric Publishing; 2021.

Frank JD, Frank JB. Persuasion and healing. *A Comparative Study of Psychotherapy*. 3rd ed. The Johns Hopkins University Press; 1991.

Gabbard GO. *Psychodynamic Psychiatry in Clinical Practice*. 5th ed. American Psychiatric Publishing; 2014.

Gaw AC. *Cross Culture, Ethnicity, and Mental Illness*. American Psychiatric Press; 1993.

Gold LH, Frierson RL, eds. *The American Psychiatric Publishing Textbook of Suicide Assessment and Management*. 3rd ed. American Psychiatric Association Publishing, Inc.; 2020.

Groves JE. Taking care of the hateful patient. *N Engl J Med*. 1978;298:883–887.

Guinart D, Marcy P, Hauser M, Dwyer M, Kane JM. Mental health care providers' attitudes toward telepsychiatry: a systemwide, multisite survey during the COVID-19 pandemic. *Psychiatr Serv*. 2021;72:704–707.

Irani JS, Middleton JL, Marfatia R, Omana ET, D'Amico F. The use of electronic health records in the exam room and patient satisfaction: a systematic review. *J Am Board Fam Med*. 2009;22:553–562.

Kashner TM, Rush AJ, Suris A, Biggs MM, Gajewski VL. Impact of structured clinical interview on physicians' practices in community mental health settings. *Psychiatr Serv*. 2003;54:712–718.

MacKinnon RA, Michels R, Buckley PJ. *The Psychiatric Interview in Clinical Practice*. 3rd ed. American Psychiatric Publishing; 2015.

McCabe R, Priebe S. Communication and psychosis: it's good to talk, but how? *Br J Psychiatry*. 2008;192:404–405.

McLean M, Armstrong D. Eliciting patient's concerns: a randomized controlled trial of different approaches by the doctor. *Br J Gen Pract*. 2004;54:663–666.

Peabody F. The care of the patient. *JAMA*. 1927;88:877–882.

Rodgers R, Bender S, eds. *Clinical Assessment of Malingering and Deception*. 4th ed. Guilford Press; 2020.

Sadock BJ, Sadock VA, eds. *Kaplan & Sadock's Comprehensive Textbook of Psychiatry*. 10th ed. Lippincott Williams & Wilkins; 2017.

Shea SC. *Psychiatric Interviewing; The Art of Understanding*. 3rd ed. Elsevier; 2016.

Shore JH. Telepsychiatry. Videoconferencing in the delivery of psychiatric care. *Am J Psychiatry*. 2013;170:256–262.

Strous RD, Ulman AM, Kotler M. The hateful patient revisited: relevance for 21st century medicine. *Eur J Int Med*. 2006;17:387–393.

7.2 The Psychiatric Report and Outline for a Psychiatric Examination

Benjamin J. Sadock, M.D.

The psychiatric history and the mental status are to the psychiatrist what the medical history and physical examination are to the internist. The psychiatric history describes life events from infancy to old age to the extent they can be recalled. It is sometimes called the *anamnesis* (from the Greek meaning “to remember”). The mental status covers how the patient appears and the thoughts and feelings evoked as he or she responds to specific questions asked by the examiner.

The psychiatric report is more than the psychiatric history and mental status. It not only includes those components it also includes a summary of positive and negative findings and an interpretation of the data found, thereby providing an understanding of the case. The report also addresses critical questions: Are future diagnostic studies needed, and, if so, which ones? Is a comprehensive neurologic workup, including an electroencephalogram (EEG) or imaging studies, needed? Are psychological tests indicated? Are psychodynamic factors relevant? The report includes a diagnosis made according to the latest edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM). A prognosis is also discussed in the report, with good and bad prognostic factors listed. The report concludes with a discussion of a treatment plan and makes firm recommendations about the management of the case.

The outline for the psychiatric examination that follows was developed by the author and Harold Kaplan, MD. It has been used by countless numbers of psychiatrists over several generations and has been revised and updated periodically, especially for this edition. It includes a tremendous amount of information about the patient; however, not all need to be obtained depending on circumstances in the case. Beginning clinicians are advised to get as much information as possible; more experienced clinicians can pick and choose among the series of questions they might ask. In all cases, the person is best understood within the context of their life events.

PSYCHIATRIC HISTORY

A. *Identification:* Name; age; marital status; sex; occupation; language, if other than English; race and ethnic origin; nationality; religion; and persons with whom the patient lives.

B. *Chief complaint:* Exactly why the patient came to the psychiatrist, preferably in the patient’s own words; if that information does not come from the patient, note who supplied it.

C. *History of present illness:* Chronologic background and development of the symptoms or behavioral changes that culminated in the patient's seeking assistance; patient's life circumstances at the time of onset; personality when well; how illness has affected life activities and personal relations—changes in personality, interests, mood, attitudes toward others, dress, habits, level of tenseness, irritability, activity, attention, concentration, memory, speech; psychophysiologic symptoms—nature and details of dysfunction; pain—location, intensity, and fluctuation; level of anxiety—generalized and nonspecific (free-floating) or specifically related to particular situations, activities, or objects; how anxieties are handled—avoidance, repetition of a feared situation, or use of drugs or other activities for alleviation.

D. *Past psychiatric and medical history:* (1) Emotional or mental disturbances—extent of incapacity, type of treatment, names of hospitals, length of illness, and effect of treatment; (2) physical disorders—allergies, arthritis, colitis, chronic fatigue, recurrent colds, and skin conditions; (3) medical conditions—customary review of systems, sexually transmitted diseases, alcohol or other substance abuse, or at risk for acquired immunodeficiency syndrome (AIDS); (4) neurologic disorders—headache, craniocerebral trauma, loss of consciousness, seizures, or tumors (5) Vaccinations—smallpox, DPT, flu, Covid.

E. *Family history:* Elicited from the patient and from someone else if possible because quite different descriptions may be given of the same people and events; ethnic, national, and religious traditions; other people in the home, descriptions of them—personality and intelligence—and what has become of them since the patient's childhood; descriptions of different households lived in; present relationships between the patient and those who are in the patient's family; the role of illness in the family; family history of mental illness; where does the patient live—neighborhood and particular residence of the patient; is the home crowded; privacy of family members from each other and from other families; sources of family income and difficulties in obtaining it; public assistance (if any) and attitudes about it; will the patient lose their job or apartment by remaining in the hospital; who is caring for the patient's children.

F. *Personal history:* History of the patient's life from infancy to the present to the extent it can be recalled; gaps in history as spontaneously related by the patient; emotions associated with different life periods (painful, stressful, and conflictual) or with phases of the life cycle.

1. Early childhood to the extent it can be recalled or about which patient was told (through 3 years of age).
    - a. Prenatal history and mother's pregnancy and delivery: Length of pregnancy, spontaneity, and normality of delivery, birth trauma, whether the patient was planned and wanted, birth defects.
    - b. Feeding habits: Breastfed or bottle-fed, eating problems.
    - c. Early development: Maternal deprivation, language development, motor development, signs of unmet needs, sleep pattern, object constancy, stranger anxiety, separation anxiety.
    - d. Toilet training: Age, the attitude of parents, feelings about it.
    - e. Symptoms of behavior problems: Thumb sucking, temper tantrums, tics, head bumping, rocking, night terrors, fears, bed-wetting or bed-soiling, nail-biting.
    - f. Personality and temperament as a child: Shy, restless, overactive, withdrawn, studious, outgoing, timid, athletic, friendly, patterns of play, reactions to siblings.
    - g. Early or recurrent dreams or fantasies.
2. Middle childhood (3 to 11 years of age): Early school history—feelings about going to school, early adjustment, gender identification, conscience development, punishment, social relationships, and attitudes toward siblings and playmates.
3. Later childhood (prepuberty through adolescence).
    - a. Peer relationships: Number and closeness of friends, leader or follower, social popularity, participation in group or gang activities, idealized figures; patterns of aggression, passivity, anxiety, antisocial behavior.
    - b. School history: How far the patient went in school; adjustment to school; relationships with teachers—teacher's pet or

rebellious; favorite studies or interests; particular abilities or assets; extracurricular activities; sports; hobbies; relationships; problems or symptoms in any school period. School history can be extended to college in this section.

- c. Cognitive and motor development: Learning to read and other intellectual and motor skills, minimal cerebral dysfunction, learning disabilities—their management and effects on the child.
- d. Particular adolescent emotional or physical problems: Nightmares, phobias, bed-wetting, running away, delinquency, smoking, drug or alcohol use, weight problems, feeling of inferiority.
- e. Psychosexual history.
  - i. Early curiosity, infantile masturbation, sex play.
  - ii. Acquiring of sexual knowledge, attitude of parents toward sex, sexual abuse.
  - iii. Onset of puberty, feelings about it, kind of preparation, feelings about menstruation, development of secondary sexual characteristics.
  - iv. Adolescent sexual activity: Crushes, parties, dating, petting, masturbation, wet dreams (nocturnal emissions), and attitudes toward them.
  - v. Attitudes toward same and opposite sex: Timid, shy, aggressive, need to impress, seductive, anxious.
  - vi. Sexual practices: Sexual problems, homosexual and heterosexual experiences, paraphilias, promiscuity.
- f. Religious background: Strict, liberal, mixed (possible conflicts), the relationship of background to current religious practices.

4. Adulthood.
   - a. Occupational history: Choice of occupation, training, ambitions, and conflicts; relations with authority, peers, and subordinates; the number of jobs and duration; changes in job status; current job and feelings about it.
   - b. Social activity: Whether the patient has friends; whether he or she is withdrawn or socializing well; social, intellectual, and

physical interests; relationships with same-sex and opposite sex; depth, duration, and quality of human relations.

c. Adult sexuality.
   i. Premarital sexual relationships, age of first coitus, sexual orientation, gender confusion.
   ii. Marital history: Common-law marriages; legal marriages; description of courtship and role played by each partner; age at marriage; family planning and contraception; names and ages of children; attitudes toward raising children; problems of any family members; housing difficulties, if important to the marriage; sexual adjustment; extramarital affairs; areas of agreement and disagreement; management of money; the role of in-laws.
   iii. Sexual symptoms: Anorgasmia, erectile disorder, premature ejaculation, lack of desire.
   iv. Sexual practices: Paraphilias, such as sadism, fetishes, voyeurism; attitude toward fellatio, cunnilingus; coital techniques, frequency.

d. Military history: General adjustment, combat, injuries, referral to psychiatrists, type of discharge, veteran status.

e. Value systems: Whether children are seen as a burden or a joy; whether work is seen as a necessary evil, an avoidable chore, or an opportunity; current attitude about religion; belief in heaven and hell; attitudes toward pregnancy and having children; contraceptive practices and feelings about them.

MENTAL STATUS

Sum total of the examiner's observations and impressions derived from the initial interview.

A. Appearance.

1. Personal identification: May include a brief, nontechnical description of the patient's appearance and behavior as a novelist might write it. Attitude toward examiner can be described here: cooperative, attentive, interested, frank, seductive, defensive, hostile, playful, ingratiating, evasive, or guarded.

    2. Behavior and psychomotor activity: Gait, mannerisms, tics, gestures, twitches, stereotypes, picking, touching examiner, echopraxia, clumsy, agile, limp, rigid, retarded, hyperactive, agitated, combative, or waxy.
    3. General description: Posture, bearing, clothes, grooming, hair, nails; healthy, sickly, angry, frightened, apathetic, perplexed, contemptuous, ill at ease, poised, old looking, young-looking, effeminate, masculine; signs of anxiety—moist hands, perspiring forehead, restlessness, tense posture, strained voice, wide eyes; shifts in the level of anxiety during the interview or with a particular topic; eye contact (50% is normal).

B. *Speech:* Rapid, slow, pressured, hesitant, emotional, monotonous, loud, whispered, slurred, mumbled, stuttering, echolalia, intensity, pitch, ease, spontaneity, productivity, manner, reaction time, vocabulary, prosody.

C. Mood and affect.

    1. Mood (a pervasive and sustained emotion that colors the person's perception of the world): How does patient say he or she feels; depth, intensity, duration, and fluctuations of mood—depressed, despairing, irritable, anxious, terrified, angry, expansive, euphoric, empty, guilty, awed, futile, self-contemptuous, anhedonic, alexithymic. Suicide: ideas, plans, ruminations, or obsessions about, previous attempts.
    2. Affect (the outward expression of the patient's inner experiences): How the examiner evaluates the patient's affects—broad, restricted, blunted, flat, shallow, amount and range of expression; difficulty in initiating, sustaining, or terminating an emotional response; whether the emotional expression is appropriate to the thought content, culture, and setting of the examination; examples should be given if the emotional expression is not appropriate.

D. Thinking and perception.

    1. Form of thinking.
        a. Productivity: Overabundance of ideas, paucity of ideas, flight of ideas, rapid thinking, slow thinking, hesitant thinking; whether the patient speaks spontaneously or only when questions are asked; the stream of thought, quotations from the patient.

    - b. Continuity of thought: Whether the patient’s replies really answer questions and are goal-directed, relevant, or irrelevant; loose associations; lack of cause-and-effect relationships in the patient’s explanations; illogical, tangential, circumstantial, rambling, evasive, persevering statements, blocking or distractibility.
    - c. Language impairments: Impairments that reflect disordered mentation, such as incoherent or incomprehensible speech (word salad), clang associations, neologisms.
  - 2. Content of thinking.
    - a. Preoccupations: About the illness, environmental problems; obsessions, compulsions, phobias; thoughts of harming others, specific antisocial urges or impulses.
  - 3. Thought disturbances.
    - a. Delusions: Content of any delusional system, its organization, the patient’s convictions as to its validity, how it affects their life; persecutory delusions—isolated or associated with pervasive suspiciousness; mood-congruent or mood-incongruent; thought insertion.
    - b. Ideas of reference and ideas of influence: How ideas began, their content, and the meaning that the patient attributes to them.
  - 4. Perceptual disturbances.
    - a. Hallucinations and illusions: Whether the patient hears voices or sees visions; content, sensory system involvement, circumstances of the occurrence; hypnagogic or hypnopompic hallucinations; thought broadcasting or thought insertion, disturbances in smell.
    - b. Depersonalization and derealization: Extreme feelings of detachment from self or from the environment.
  - 5. Dreams and fantasies.
    - a. Dreams: Prominent ones, if the patient will tell them; nightmares.
    - b. Fantasies: Recurrent, favorite, or unshakable daydreams.
- E. Sensorium.

1. Alertness: awareness of the environment, attention span, clouding of consciousness, fluctuations in levels of awareness, somnolence, stupor, lethargy, fugue state, coma.
2. Orientation.
   a. Time: Whether the patient identifies the day or the approximate date and the time of day correctly; if in a hospital, whether the patient knows how long he or she has been there; whether the patient behaves as though oriented to the present.
   b. Place: Whether patient knows where he or she is.
   c. Person: Whether the patient knows who the examiner is and the roles or names of the persons with whom the patient is in contact.
3. Concentration and calculation: Whether the patient can subtract 7 from 100 and keep subtracting 7s; if the patient cannot subtract 7s, whether easier tasks can be accomplished, such as 4 × 9 and 5 × 4; whether the patient can calculate how many nickels are in $1.35; whether anxiety or some disturbance of mood or concentration seems to be responsible for difficulty.
4. Memory: Impairment, efforts made to cope with impairment—denial, confabulation, catastrophic reaction, circumstantiality used to conceal deficit; whether the process of registration, retention, or recollection of material is involved.
   a. Remote memory: Childhood data, important events known to have occurred when the patient was younger or free of illness, personal matters, neutral material.
   b. Recent past memory: Past few months.
   c. Recent memory: Past few days, what did the patient do yesterday and the day before, what did the patient have for breakfast, lunch, and dinner.
   d. Immediate retention and recall: Ability to repeat six figures after the examiner dictates them—first forward, then backward, then after a few minutes' interruptions; other test questions; whether the same questions, if repeated, called forth different answers at different times.
   e. Effect of defect on the patient: Mechanisms the patient has developed to cope with the defect.

5. Fund of knowledge: Level of formal education and self-education; estimate of the patient's intellectual capability and whether the patient is capable of functioning at the level of their basic endowment; counting, calculation, general knowledge; questions should have relevance to the patient's educational and cultural background.
6. Abstract thinking: Disturbances in concept formation; the manner in which the patient conceptualizes or handles their ideas; similarities (e.g., between apples and pears), differences, absurdities; meanings of simple proverbs, such as "a rolling stone gathers no moss"; answers may be concrete (giving specific examples to illustrate the meaning) or overly abstract (giving generalized explanation); appropriateness of answers.
7. Insight: Degree of personal awareness and understanding of illness.
   a. Complete denial of illness.
   b. Slight awareness of being sick and needing help but denying it at the same time.
   c. Awareness of being sick but blaming it on others or external factors, or on medical or unknown organic factors.
   d. Intellectual insight: Admission of illness and recognition that symptoms or failures in social adjustment are due to irrational feelings or disturbances, without applying that knowledge to future experiences.
   e. True emotional insight: Emotional awareness of the motives and feelings within and of the underlying meaning of symptoms; whether the awareness leads to changes in personality and future behavior; openness to new ideas and concepts about self and the important people in the patient's life.
8. Judgment.
   a. Social judgment: Subtle manifestations of behavior that are harmful to the patient and contrary to acceptable behavior in the culture; whether the patient understands the likely outcome of personal behavior and is influenced by that understanding; examples of impairment.
   b. Tests of judgment: The patient's prediction of what he or she would do in imaginary situations; for instance, what the patient

would do with a stamped, addressed letter found in the street.

FURTHER DIAGNOSTIC STUDIES

A. Physical examination.
B. Neurologic examination.
C. Additional psychiatric diagnostic interviews.
D. Interviews with family members, friends, or neighbors by a social worker.
E. Psychological, neurologic, or laboratory tests, as indicated: EEG, CT scan, magnetic resonance imaging (MRI), tests of other medical conditions (e.g., human immunodeficiency virus [HIV]), reading comprehension and writing tests, test for aphasia, projective or objective psychological tests, dexamethasone-suppression test (DST), 24-hour urine test for heavy metal intoxication, urine screen for drugs of abuse, sleep studies in chronic insomnia.

SUMMARY OF FINDINGS

Mental symptoms, medical and laboratory findings, and psychological and neurologic test results, if available, are summarized. Include the medications that the patient has been taking, their dosage, and their duration.

Table 7.2–1.

Common Questions for the Psychiatric History and Mental Status

Topic	Questions	Comments and Helpful Hints
Identifying data: Name, age, sex, marital status, religion, education, address, phone number, occupation, source of referral	Be direct in obtaining identifying data. Request specific answers.	If patient cannot cooperate, get information from family member or friend; if referred by a physician, obtain medical record.
Chief complaint (CC): Brief statement in patient's own words of why patient is in the hospital or is being seen in consultation	Why are you going to see a psychiatrist? What brought you to the hospital? What seems to be the problem?	Record answers verbatim; a bizarre complaint points to psychotic process.
History of present illness (HPI): Development of symptoms from time of onset to present; relation of life events, conflicts, stressors: drugs; change from previous level of functioning	When did you first notice something happening to you? Were you upset about anything when symptoms began? Did they begin suddenly or gradually?	Record in patient's own words as much as possible. Get history of previous hospitalizations and treatment. Sudden onset of symptoms may indicate drug-induced disorder.
Previous psychiatric and medical	Did you ever lose consciousness? Have a	Ascertain extent of illness, treatment,

disorders: Psychiatric disorders; psychosomatic; medical, neurologic illnesses (e.g., craniocerebral trauma, convulsions)	seizure?	medications, outcomes, hospitals, doctors. Determine whether illness serves some additional purpose (secondary gain).
Personal history: Birth and infancy: To the extent known by the patient, ascertain mother’s pregnancy and delivery, planned or unwanted pregnancy, developmental landmarks—standing, walking, talking, temperament	Do you know anything about your birth? If so, from whom? How old was your mother when you were born? Your father?	Older mothers (>35 years) have high risk for Down syndrome baby; older father (>45 years) may contribute damaged sperm-producing deficits including schizophrenia.
Childhood: Feeding habits, toilet training, personality (shy, outgoing), general conduct and behavior, relationship with parents or caregivers and	Toilet training? Bed-wetting? Sex play with peers? What is your first childhood memory?	Separation anxiety and school phobia associated with adult depression; enuresis associated with fire setting. Childhood memories before the age of 3 are

peers, separations, nightmares, bed-wetting, fears		usually imagined, not real.
Adolescence: Peer and authority relationship, school history, grades, emotional problems, drug use, age of puberty	Adolescents may refuse to answer questions, but they should be asked. Adults may distort memories of emotionally charged adolescent experience. Sexual molestation?	Poor school performance is a sensitive indicator of emotional disorder. Schizophrenia begins in late adolescence.
Adulthood: Work history, choice of career, marital history, children, education, finances, military history, religion	Open-ended questions are preferable. Tell me about your marriage. Be nonjudgmental: What role does religion play in your life, if any? What is your sexual preference in a partner?	Depending on CC some areas require more detailed inquiry. Manic patients frequently go into debt or are promiscuous. Overvalued religious ideas associated with paranoid personality disorder.
Sexual history: Sexual development, masturbation, anorgasmia, erectile disorder, premature ejaculation, paraphilia, sexual	Are there or have there been any problems or concerns about your sex life? How did you learn about sex? Has there been any change in your sex drive?	Be nonjudgmental. Asking *when* masturbation began is a better approach than asking *do you* or *did you ever* masturbate.

orientation, general attitudes, and feelings		
Family history: Psychiatric, medical, and genetic illness in mother, father, siblings; age of parents and occupations; if deceased, date and cause; feelings about each family member, finances	Have any members in your family been depressed? Alcoholic? In a mental hospital? In jail? Describe your living conditions. Did you have your own room?	Genetic loading in anxiety, depression, schizophrenia. Get medication history of family (medications effective in family members for similar disorders may be effective in patient).
Mental Status		
General appearance: Note appearance, gait, dress, grooming (neat or unkempt), posture, gestures, facial expressions. Does patient appear older or younger than stated age?	Introduce yourself and direct patient to take a seat. In the hospital, bring your chair to bedside; do not sit on the bed.	Unkempt and disheveled in cognitive disorder; pinpoint pupils in narcotic addiction; withdrawal and stooped posture in depression.
Motoric behavior: Level of activity: Psychomotor agitation or psychomotor	Have you been more active than usual? Less active? You may ask about obvious mannerisms, e.g., “I	Fixed posturing, odd behavior in schizophrenia. Hyperactive with stimulant

retardation—tics, tremors, automatisms, mannerisms, grimacing, stereotypes, negativism, apraxia, echopraxia, waxy flexibility; emotional appearance—anxious, tense, panicky, bewildered, sad, unhappy: voice—faint, loud, hoarse; eye contact	notice that your hand still shakes, can you tell me about that?” Stay aware of smells, e.g., alcoholism/ketoacidosis.	(cocaine) abuse and in mania. Psychomotor retardation in depression; tremors with anxiety or medication side effect (lithium). Eye contact is normally made approximately half the time during the interview. Minimal eye contact in schizophrenia. Scanning of environment in paranoid states.
Attitude during interview: How patient relates to examiner—irritable, aggressive, seductive, guarded, defensive, indifferent, apathetic, cooperative, sarcastic	You may comment about attitude: You seem irritated about something; is that an accurate observation?	Suspiciousness in paranoia; seductive in hysteria; apathetic in conversion disorder (*la belle indifference*); punning (*witzlesucht*) in frontal lobe syndromes.
Mood: Steady or sustained emotional state—	How do you feel? How are your spirits? Do you have thoughts that life is	Suicidal ideas in 25% of depressives;

gloomy, tense, hopeless, ecstatic, resentful, happy, bashful, sad, exultant, elated, euphoric, depressed, apathetic, anhedonic, fearful, suicidal, grandiose, nihilistic	not worth living or that you want to harm yourself? Do you have plans to take your own life? Do you want to die? Has there been a change in your sleep habits?	elation in mania. Early morning awakening in depression; decreased need for sleep in mania.
Affect: Feeling tone associated with idea—labile, blunt, appropriate to content, inappropriate, flat	Observe nonverbal signs of emotion, body movements, facies, rhythm of voice (prosody). Laughing when talking about sad subjects, e.g., death, is inappropriate.	Changes in affect usual with schizophrenia: loss of prosody in cognitive disorder, catatonia. Do not confuse medication adverse effect with flat affect.
Speech: Slow, fast, pressured, garrulous, spontaneous, taciturn, stammering, stuttering, slurring, staccato. Pitch, articulation, aphasia, coprolalia, echolalia, incoherent,	Ask patient to say “Methodist Episcopalian” to test for dysarthria.	Manic patients show pressured speech; paucity of speech in depression; uneven or slurred speech in cognitive disorders.

logorrhea, mute, paucity, stilted		
Perceptual disorders: Hallucinations—olfactory, auditory, haptic (tactile), gustatory, visual; illusions; hypnopompic or hypnagogic experiences; feeling of unreality, *déjà vu, éjà entendu,* macropsia	Do you ever see things or hear voices? Do you have strange experiences as you fall asleep or upon awakening? Has the world changed in any way? Do you have strange smells?	Visual hallucinations suggest schizophrenia. Tactile hallucinations suggest cocainism, delirium tremens (DTs). Olfactory hallucinations common in temporal lobe epilepsy.
Thought content: Delusions—persecutory (paranoid), grandiose, infidelity, somatic, sensory, thought broadcasting, thought insertion, ideas of reference, ideas of unreality, phobias, obsessions, compulsions, ambivalence, autism, dereism, blocking, suicidal or homicidal	Do you feel people want to harm you? Do you have special powers? Is anyone trying to influence you? Do you have strange body sensations? Are there thoughts that you can’t get out of your mind? Do you think about the end of the world? Can people read your mind? Do you ever feel the TV is talking to you? Ask about fantasies and dreams.	Are delusions congruent with mood (grandiose delusions with elated mood) or incongruent? Mood-incongruent delusions point to schizophrenia. Illusions are common in delirium. Thought insertion is characteristic of schizophrenia.

preoccupation, conflicts, nihilistic ideas, hypochondriasis, depersonalization, derealization, flight of ideas, *idé* fixe, magical thinking, neologisms		
Thought process: Goal-directed ideas, loosened associations, illogic, tangential, relevant, circumstantial, rambling, ability to abstract, flight of ideas, clang associations, perseveration	Ask meaning of proverbs to test abstraction, e.g., “People in glass houses should not throw stones.” Concrete answer is, “Glass breaks.” Abstract answers deal with universal themes or moral issues. Ask similarity between bird and butterfly (both alive), bread and cake (both food).	Loose associations point to schizophrenia; flight of ideas, to mania; inability to abstract, to schizophrenia, brain damage.
Sensorium: Level of consciousness—alert, clear, confused, clouded, comatose, stuporous; orientation to time, place, person; cognition	What place is this? What is today’s date? Do you know who I am? Do you know who you are?	Delirium or dementia shows clouded or wandering sensorium. Orientation to person remains intact longer than orientation to time or place.
Memory: Remote	Where were you born?	Patients with

memory (long term): Past several days, months, years	Where did you go to school? Date of marriage? Birthdays of children? What were last week's newspaper headlines?	dementia of the Alzheimer type retain remote memory longer, than recent memory. Gaps in memory may be localized or filled in with confabulatory details. Hypermnesia is seen in paranoid personality.
Recent memory (short term): Recall of events in past day or two	Where were you yesterday? What did you eat at your last meal?	In brain disease, recent memory loss (amnesia) usually occurs before remote memory loss.
Immediate memory (very short term): Laying down of immediate information with ability to quickly recall data	Ask patient to repeat six digits forward, then backward (normal responses). Ask patient to try to remember three nonrelated items; test patient after 5 minutes.	Loss of memory occurs with cognitive, dissociative, or conversion disorder. Anxiety can impair immediate retention and recent memory. Anterograde memory loss (amnesia) occurs after taking certain drugs, e.g.,

		benzodiazepines. Retrograde memory loss occurs after head trauma.
Concentration and calculation: Ability to pay attention; distractibility; ability to do simple math	Ask patient to count from 1 to 20 rapidly; do simple calculations (2 × 3, 4 × 9); do serial 7 test, i.e., subtract 7 from 100 and keep subtracting 7. How many nickels in $1.35?	Rule out medical cause for any defects vs. anxiety or depression (pseudodementia). Make tests congruent with educational level of patient.
Information and intelligence: Use of vocabulary; level of education; fund of knowledge	Distance from New York City to Los Angeles. Name some vegetables. What is the largest river in the United States?	Check educational level to judge results. Rule out mental retardation, borderline intellectual functioning.
Judgment: Ability to understand relations between facts and to draw conclusions; responses in social situations	What is the thing to do if you find an envelope in the street that is sealed, stamped, and addressed?	Impaired in brain disease, schizophrenia, borderline intellectual functioning, intoxication.
Insight level: Realizing that there are physical or mental problems; denial of illness, ascribing blame to	Do you think you have a problem? Do you need treatment? What are your plans for the future?	Impaired in delirium, dementia, frontal lobe syndrome, psychosis, borderline intellectual functioning.

outside factors;
recognizing need
for treatment

DIAGNOSIS

Diagnostic classification is made according to the DSM.

PROGNOSIS

Opinion about the probable future course, extent, and outcome of the disorder; good and bad prognostic factors; specific goals of therapy.

PSYCHODYNAMIC FORMULATION

Causes of the patient’s psychodynamic breakdown—influences in the patient’s life that contributed to the present disorder, environmental, genetic, and personality factors relevant to determining the patient’s symptoms; primary and secondary gains; outline of the major defense mechanism used by the patient.

COMPREHENSIVE TREATMENT PLAN

Modalities of treatment recommended role of medication, inpatient or outpatient treatment, frequency of sessions, probable duration of therapy; type of psychotherapy; individual, group, family, or cognitive-behavioral therapy among others; symptoms or problems to be treated. Initially, treatment must be directed toward any life-threatening situations, such as suicidal risk or risk of danger to others, which require psychiatric hospitalization. Danger to self or others is an acceptable reason (legally and medically) for involuntary hospitalization. In the absence of the need for confinement, a variety of outpatient treatment alternatives are available: Day hospitals, supervised residences, outpatient psychotherapy, or pharmacotherapy. In some cases, treatment planning must attend to vocational and psychosocial skills training and even legal or forensic issues. Comprehensive treatment planning requires a therapeutic team approach using the skills of psychologists, social workers, nurses, activity and occupational therapists, and a variety of other mental health professionals, with referral to self-help groups (e.g., Alcoholics Anonymous [AA]) if needed. If the patient or family members are unwilling to accept the recommendations of treatment and the clinician thinks that the refusal of the recommendations may have serious consequences, then the patient, parent, or guardian should sign a statement to the effect that the recommended treatment was refused.

Finally, clarity of thinking is reflected in the clarity of writing. Avoid run-on sentences. Use the active voice as much as possible. Be precise. When summarizing the mental status, for example, the phrase “patient denies hallucinations and delusions” is not as precise as “patient denies hearing voices or thinking that he is being followed.” The latter indicates the specific question asked and the specific response given. Then, at the conclusion of the report, one would write, “Hallucinations and delusions were not elicited.” This is consistent with the principle stated above: Use the patient’s words—and those of the examiners—as much as possible.

In Table 7.2–1, the reader will find sample questions to be used in eliciting data for the psychiatric history and mental status, along with comments and helpful hints about their clinical significance.

FURTHER READINGS

*Ball C. Evidence-based medicine on the wards: report from an evidence-based minion. *ACP J Club*. 1999;130:15.

Callis R, Polmantier PC, Roeber EC. *A Casebook of Counseling*. Appleton-Century-Crofts; 1955.

Coleman M, Gillberg C. *The Schizophrenias: A Biological Approach to the Schizophrenia Spectrum Disorders*. Springer; 1996.

Cook L, Alegria M. Racial-ethnic disparities in substance abuse treatment: the role of criminal history and socioeconomic status. *Psychiatr Serv*. 2011;62:1273.

Dell’osso L, Shear MK, Carmassi C, Rucci P, Masser JD. Validity and reliability of the structured clinical interview for the trauma and loss spectrum (SCI-TALS). *Clin Pract Epidemol Ment Health*. 2008;4(1):2.

Dwyer J, Shih A. The ethics of tailoring the patient's chart. *Psychiatr Serv*. 1998;49:1309.

Evans J. *Three Men*. Knopf; 1954.

*Fischer CB, Fried AL. Internet-mediated psychological services and the American Psychological Association Ethics Code. *Psychother Theor Res Pract Train*. 2003;40:103.

Gersteneckera A, Masta B. Mild cognitive impairment: a history and the state of current diagnostic criteria. *Int Psychogeriatr*. 2015;27:199.

Gordon W, Perrott J, Dahri K. Accuracy of medication histories: the first step of medication reconciliation. *Ann Pharmacother*. 2008;42(1):144.

Greist H. Computer-based assessment of patients. *J Clin Pharmacol*. 1998;18:359.

Groth-Marnat G. *Handbook of Psychological Assessment*. 4th ed. John Wiley & Sons; 2003.

Jeffords JM. Confidentiality of medical information: protecting privacy in an electronic age. *Prof Psychol Res Pract*. 1999;30:115.

Kellera B, Heddenc T. Effects of physician awareness of symptom-related expectations and mental disorders. *Cortex*. 2015;64:271.

Krauss JB. A matter of privacy. *Arch Pract Nurs*. 2003;17:99.

Lysaker PH, Wickett AM, Campbell K, Buck KD. Movement toward coherence in the psychotherapy of schizophrenia: a method for assessing narrative transformation. *J Nerv Ment Dis*. 2003;191:538.

*Mark D. Effects of physician awareness of symptom-related expectations and mental disorders: a controlled trial. *Arch Fam Med*. 1999;8:135.

Park E, Abraham MK. Altered mental status and endocrine diseases. *Emerg Med Clin N AM*. 2014;32:367.

Pierluissi E, Fischer MA, Campbell AR, Landefeld CS. Discussion of medical errors in morbidity and mortality conferences. *JAMA*. 2003;290:2838.

Shea SC. *Psychiatric Interviewing*. 2nd ed. WB Saunders; 1998.

Wedding D, Corsini RJ. *Case Studies in Psychotherapy*. Peacock; 1989.

Young HM, Gray SL, McCormick WC, Sikma SK, Reinhard S. Types, prevalence and potential clinical significance of medication administration errors in assisted living. *J Am Geriatr Soc*. 2008;56:1199.

Zweig RA, Turkel E. The social adjustment scale-self-report: psychometric properties for older adults. *Psychol Rep*. 2007;101:920.

7.3 Practice Guidelines in Psychiatry

Laura J. Fochtmann, M.D., M.B.I., and John S. McIntyre, M.D.

Over the past half century there have been great advances in the knowledge base for all of medicine including psychiatry. Much has been learned about the etiology and epidemiology of mental illnesses and the efficacy and effectiveness of treatments. Evidence-based medicine is the process of translating those advances into the day-to-day care of patients. However, a large number of studies and reports have documented that effective interventions are not always implemented in a timely fashion, if at all. Compounding this problem, especially in the United States, is the reality that reimbursement for services is not necessarily driven by what has been shown to be optimal care. These issues can be addressed using practice guidelines, which are sets of evidence-based strategies that are aimed at assisting physicians and patients in clinical decision making and improving care.

EARLY GUIDELINE EFFORTS

Throughout the history of medicine there have been recommendations for the care of patients, which have been promulgated orally or in written form. In 1851, 7 years after its founding, the American Association of Medical Superintendents of Hospitals for the Insane (later to become the American Psychiatric Association [APA]) issued 26 propositions concerning “the construction of hospitals for the insane.” Two years later 14 propositions concerning the organization of these facilities were published. These propositions were intended to provide guidance about psychiatric care but did not include documentation of the evidence behind the propositions or the process by which they were developed. Early efforts at promulgating evidence-based guidelines in the United States included the publications of the American Academy of Pediatrics on immunizations in the 1940s. Over the subsequent four decades the number of guidelines began to increase and by the mid-1980s the growth had become exponential.

AMERICAN MEDICAL ASSOCIATION AND GUIDELINES

In the late 1980s the American Medical Association (AMA) convened a group of representatives from 14 medical specialty organizations, including the APA. Also attending on a regular basis were representatives from the Agency for Health Care Policy and Research, Joint Commission on Accreditation of Health Care Organizations, American Hospital Association, National Council on Quality Assurance, and the Health Care Financing Administration. This consortium was initially named the Practice Parameter Partnership and later renamed the Practice Guideline Partnership. In 1989, the Partnership described five recommended attributes of guidelines: (1) be developed by or in conjunction with physician organizations, (2) use reliable methodologies that integrate relevant research findings and clinical expertise, (3) be as comprehensive and specific as possible, (4) be based on current information, (5) be widely disseminated. In 1996, a sixth attribute was added: Guideline development should include outcomes research, goals, and measures. The focus on measures was spurred by the increased interest in assessing whether guidelines were followed and led to the development of the AMA-led Physician Consortium for Performance Improvement (PCPI).

In 1999, the AMA and the American Association of Health Plans collaborated with the Agency for Healthcare Research and Quality (AHRQ) of the U.S. Department of Health and Human Service in establishing the web-based National Guidelines Clearinghouse.

INSTITUTE OF MEDICINE REPORTS

In parallel with the increasing involvement of professional organizations in developing practice guidelines, the Institute of Medicine (IOM) began to discuss and define optimal characteristics of practice guidelines. In 1990, the IOM published a monograph: *Clinical Practice Guidelines—Directions for a New Program*. Incorporating some of the recommendations included in the AMA report, “good guidelines” were identified as having eight attributes as identified in Table 7.3–1.

This IOM report describes practice guidelines as “valid if, when followed, they lead to the health and cost outcomes projected for them, other things being equal.” Practice guidelines are defined as “reliable and reproducible” if a different group of developers, given the same evidence and using the same methods, arrive at the same recommendations and if the guidelines are interpreted and applied consistently by clinicians in similar settings. “Clinical applicability” requires that the population being treated with the guideline be explicitly identified. A guideline has “flexibility” if the generally expected exceptions are identified and “clarity” if terms are well defined and unambiguous language is used. The report indicates that a “multidisciplinary process” may include “[participants from different disciplines] serving on panels that develop guidelines, providing evidence and viewpoints to the panels and reviewing draft guidelines.” The report also notes that practice guidelines should include specific plans about “review and revision” schedules and that there should be “meticulous documentation and description” of the evidence used, the process followed, and the individuals involved in developing guidelines.

Table 7.3–1.
Guideline Attributes

Validity
Reliability/reproducibility
Clinical applicability
Clinical flexibility
Clarity
Multidisciplinary process
Scheduled review
Documentation

Data from Field MJ, Lohr KN, eds. *Clinical Practice Guidelines—Directions for a New Program*. National Academies Press; 1990.

Two decades later, a series of IOM reports expanded on these recommendations for practice guideline development. A 2008 report, *Knowing What Works in Health Care: A Road Map for the Nation*, described the reasons that changes were needed in synthesizing of information and using it for clinical decisions. The report cited information overload and health policy challenges such as unsustainable health care costs and geographic variations in the amount and quality of care. It also noted that available systematic reviews and clinical practice guidelines had not fulfilled their original promise due to a lack of universal development standards, variations in evidence appraisal, and inconsistent handling of conflicts of interest. This report served as the impetus for the Medicare Improvements for Patients and Providers Act of 2008 to request two additional IOM reports, *Finding What Works in Health Care: Standards for Systematic Reviews* and *Clinical Practice Guidelines We Can Trust*.

For systematic reviews, the IOM gives 21 recommendations to promote objectivity, transparency, and reliability in the review process. These recommendations include, but are not limited to, establishing a systematic review team with the appropriate expertise and experience, managing bias and conflict of interest among team members and those providing input into the review, obtaining public input on the review at multiple steps in the process, developing a protocol for the review that is established a priori, and using clear and transparent processes for a comprehensive search and for evidence review and synthesis. In the complementary report on clinical practice guidelines, the IOM gives eight additional recommendations that discuss the importance of transparency and management of conflict of interest, appointing a balanced multidisciplinary guideline development group, using rigorously conducted systematic reviews, providing ratings of the strength of recommendations based on the underlying quality of available evidence, and wording guideline recommendations in precise actionable terms. The IOM also notes that practice guidelines should be subjected to public review and regular updating. Many of these precepts of guideline development were subsequently affirmed by the Council of Medical Specialty Societies in their “Principles for the Development of Specialty Society Clinical Guidelines.”

PRACTICE GUIDELINES IN PSYCHIATRY

One of the first psychiatric associations to publish guidelines was the Royal Australian and New Zealand College of Psychiatrists (RANZCP) beginning in 1981. This project initially had a unique methodology for blending research findings and clinical consensus. A group of research experts developed a guideline for a particular disorder and a group of clinicians developed a guideline for the same disorder and then the two guidelines were integrated into a single document.

Other psychiatric guideline efforts soon followed. In the United States, the APA has been developing guidelines since 1989 (https://www.psychiatry.org/psychiatrists/practice/clinical-practice-guidelines), and the American Academy of Child and Adolescent Psychiatry (AACAP) published its initial Practice Parameters in 1991 (http://www.jaacap.com/content/pracparam). The Veterans Administration also has a robust guideline development process, which includes psychiatric guidelines (http://www.healthquality.va.gov/). Other agencies and professional organizations that have developed guidelines of relevance to psychiatry include, but are not limited to, the American College of Physicians, the American Psychological Association, the Canadian Psychiatric Association, Canadian Network for Mood and Anxiety Treatments, The National Institute for Health and Care Excellence (NICE), the Scottish Intercollegiate Guidelines Network, and The World Federation of Societies of Biological Psychiatry.

APA PRACTICE GUIDELINE PROJECT

History

The APA Practice Guideline project began in 1989. An action paper in the APA Assembly called for the APA to develop practice guidelines to assist psychiatrists in their care of patients. Following approval by the APA Board of Trustees (BOT) and President, a Steering Committee was formed to create a process to develop guidelines and provide ongoing oversight.

Guideline Committees and Writing Groups

Early in the APA project, the decision was made that no industry money would be sought or accepted for the guideline development. In addition, writing group and practice guideline committee members are carefully screened for potential conflicts of interest in order to limit the risk of bias in guideline recommendations. Several APA staff, including the medical editor of the guidelines, have been assigned to the project. With their participation and the volunteer efforts of countless work group members and committee members, the project has continued to flourish.

Based upon the recommendations of the IOM, guidelines are written by a multidisciplinary group of individuals, some of whom have specific expertise on the guideline topic but others of whom have more general expertise in the treatment of psychiatric disorders.

Choosing a Topic

The initial step in developing a guideline is to choose a topic. At the onset of the APA project it was decided to develop guidelines for specific illnesses/disorders as defined by the *Diagnostic and Statistical Manual of Mental Disorders*. Having a criteria-based diagnostic nomenclature was a major asset in starting the project. More recently, some guidelines have addressed more circumscribed topics in patients with a given diagnosis, such as treatment of specific symptoms (e.g., agitation or psychosis in patients with dementia) or use of specific types of treatment (e.g., pharmacotherapy in patients with alcohol use disorder). In addition, some guidelines have focused more broadly on evaluation of individuals who present with psychiatric symptoms or conditions.

Typically, the Committee on Practice Guidelines selects an illness/disorder that represents a significant public health problem (high prevalence and seriousness), is relevant to psychiatric practice, and for which there is a substantial body of evidence to support recommendations.

Identifying the Evidence

As noted above, guidelines must describe specific interventions and document the evidence that supports these recommendations. At the heart of the systematic review process is a detailed literature search, which is clearly delineated in each guideline, identifies keywords, and employs a MEDLINE search as well as specialized databases such as PsycINFO and the Cochrane Library.

Table 7.3–2.
Grading of the Quality of Individual Studies

Population
Intervention
Comparators
Outcomes
Timing
Setting
Study design
Overall risk of study bias
Selection bias
Performance bias
Attrition bias
Detection bias
Reporting bias
Sponsor-related bias
Applicability

Adapted from Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality. 2014. www.effectivehealthcare.ahrq.gov

The citations retrieved by the literature search are screened for inclusion in the systematic review based on the study design and relevance to the guideline topic. A rigorous approach is used such as that outlined by the AHRQ, which has many similarities to the methodology used by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) group. Studies that meet inclusion criteria are assessed for aspects of study design, findings, and quality as shown in Table 7.3–2.

Synthesis of the findings of available research also follows a defined process as outlined in Table 7.3–3.

The overall strength of research evidence is not a measure of the size of the effect of the intervention but rather reflects the degree of confidence that additional studies will show similar conclusions. Factors that enter into this determination include the consistency of findings across studies, the directness of the effect on a specific health outcome, precision of the estimate of the effect, and the risk of bias in available studies.

Table 7.3–3.
Grading of the Overall Supporting Body of Research Evidence

Overall strength of research evidence
Risk of bias
Consistency
Directness
Precision
Dose–response relationship (if applicable)
Magnitude of effect
Confounding factors (including likely direction of effect)
Applicability

Adapted from Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality. 2014. www.effectivehealthcare.ahrq.gov

Developing Guideline Recommendations

The drafting of practice guideline recommendations is a natural outgrowth of the review of the available evidence.

Beginning in 2014, these guideline recommendations are developed by the guideline writing group using an iterative Delphi process with discussion followed by blinded voting. If consensus is not reached, additional rounds of discussing and voting occur.

Based on the recommendations of the IOM, guideline statements are crafted in specific actionable terms that describe the circumstances or patient populations in which an intervention is to occur, the object of the action (if not the identified patient), and the individual carrying out the action (if not the psychiatrist or treating clinician). Whenever possible vague terms (e.g., moderate to severe symptoms) are explicitly defined or eliminated.

In addition, guideline statements are based upon the relative balance of benefits and harms of an intervention. Thus, a recommendation indicates confidence that benefits of an intervention clearly outweigh harms, whereas a suggestion indicates a greater degree of uncertainty, either that the balance of benefits and harms is more difficult to judge or that either the benefits or the harms are unclear.

Some guideline developers have attempted to limit guideline recommendations to those supported by randomized controlled trials (RCTs). However, other guideline developers recognize that additional forms of evidence are often needed in crafting guideline recommendations. For example, many aspects of clinical practice, including obtaining a patient history, have not been and cannot be studied with RCTs. Also, especially in psychiatry, RCTs and other trials often focus on pharmacotherapy or other somatic treatments and are limited in duration (often 6 to 12 weeks) because of funding issues. In addition, significant research challenges exist when studying some treatments, such as psychodynamic psychotherapy and some social therapies. For these reasons, expert clinical consensus is an essential component in the development of a guideline.

In an effort to make the expert consensus process more rigorous, APA has pioneered a “snowball” sampling methodology to identify clinical and research experts based on nomination of their peers. Nominated experts then receive a web-based survey with questions on the guideline topic. These expert survey results do not substitute for rigorous clinical trials, but answers to the expert survey are reviewed by the guideline writing group and can be particularly informative for situations in which clinical trials are not possible.

Format of APA Practice Guideline

Each APA practice guideline published between 1999 and 2014 followed a standardized format that included: (I) Executive Summary, (II) Formulation and Implementation of a Treatment Plan, (III) Specific Clinical Features Influencing the Treatment Plan, (IV) Disease Definition, Epidemiology, and Natural History, and (V) Review and Synthesis of Available Evidence. A discussion of Future Research Needs was also included.

The section of the guideline on formulation and implementation of the treatment plan discussed the use of specific somatic treatments and psychotherapies, but also described other important elements of psychiatric management that were essential for good care. Examples of such activities included: performing a diagnostic evaluation, evaluating the safety of the patient and others, evaluating the level of functional impairments, determining a treatment setting, establishing and maintaining a therapeutic alliance, monitoring the patient’s psychiatric status and safety, addressing comorbid disorders, providing education to patients and families, enhancing treatment adherence, and working with patients to address early signs of relapse.

Beginning in 2014, the format of the APA practice guidelines shifted to a different format, in part aimed at fulfilling the recommendations of the IOM. The new format is centered around specific guideline statements (e.g., recommendations, suggestions) accompanied by ratings of the strength of the recommendation and the strength of the supporting research evidence. A section on potential benefits and harms of the guideline statement is aligned with the GRADE methodology and describes the pluses and minuses of the intervention. This information is useful in understanding why the guideline statement was made and, along with the section on patient preferences, can be useful in shared decision making and, where relevant, informed consent. A section on implementation discusses approaches for integrating the recommendations or suggestions into routine care, including discussion of barriers to integration and ways to overcome them. This section also incorporates a discussion of circumstances in which guideline recommendations may not be appropriate, due to factors such as comorbid physical or psychiatric conditions. It also discusses relevant aspects of psychiatric management and ways in which patient values and preferences may shape recommended treatment approaches. To assure transparency of the guideline development process, the review of the supporting research evidence is included in the guideline in detail. Each guideline statement is also accompanied by a description of other organizations’ guideline recommendations on the same topic. This section is aimed at highlighting areas of consistency among guidelines but also to identify where different organizations may have drawn different conclusions about the evidence. An additional content area for each guideline statement relates to quality measurement. It discusses existing quality measures that are related to the guideline statement and, if existing quality measures do not exist or are insufficient, discusses whether and how to derive a quality measure from the recommendation. This section also discusses whether the guideline recommendation might be incorporated into electronic decision support or used as part of organizational quality initiatives to improve care.

In addition to information provided in relation to each guideline statement, the guideline includes descriptions of the development process, including the systematic review methodology. A section on the rationale of the guideline describes the guideline purpose and the epidemiology, morbidity/mortality, and other public health implications of the disorder or problem. Information on variations in clinical practice may also be discussed, if available and relevant. This section is intended to show clinicians, insurers, and policy makers why the guideline recommendations are important to patients, their families, health professionals, and society.

Review and Approval Process

The writing, review, and editing of a guideline is an iterative process. The number of drafts has evolved over the history of the practice guideline project and current processes focus on allowing a thorough review without untoward delays during which evidence becomes outdated. With the current process, the guideline draft is prepared by the guideline writing group and then reviewed by a large number of individuals, each of whom is asked to identify any potential conflict of interest. More specifically, the draft is distributed for review to at least 50 selected national and international expert reviewers, members of APA components and assembly, district branches, and allied organizations. Allied organizations include other medical specialties (e.g., neurology and family medicine), other mental health specialties (nursing, psychology, social work), and patient and family groups (e.g., National Alliance on Mental Illness [NAMI]). Reviewers are asked to make specific recommendations and, if possible, provide evidence to support their suggestions. Each suggestion is considered by the work group chair, the medical editor of the project, the project manager, and often the chair of the Committee on Practice Guidelines. The final draft is then prepared and is sent to the APA Assembly and the BOT for approval. Hence each guideline carries the authority of the Association, which aids in its acceptance and dissemination to the field.

Dissemination and Implementation Strategies

From the beginning of the project much attention has been directed to dissemination and implementation strategies. After approval, each guideline or a key portion of the guideline is published in the *American Journal of Psychiatry*. Many of the guidelines have also been published in a *Compendium of APA Guidelines* or, more recently, as free-standing publications. Some guidelines have also been formatted as six to eight panel pocket guides, following a model developed by the Veterans Administration, with the goal of improving the accessibility and usability of the guidelines in the day-to-day care of patients.

Each approved guideline is published in its entirety online on the APA website (https://www.psychiatry.org/psychiatrists/practice/clinical-practice-guidelines). Other resources that are available online at the APA site are clinician summaries, patient and family guides, continuing medical education courses, performance in practice (PIP) tools, and educational slide sets.

Updating of Guidelines

Because of the rapid growth of knowledge, keeping guidelines current is one of the major challenges facing this or any other guideline project. APA Practice Guidelines were scheduled for revision every 3 to 5 years. Because it takes about 2 years to develop a guideline, even if this schedule is followed (which has not consistently happened), guidelines may be 7 years old before a revision is published.

The change to a new format for guidelines was, in part, motivated by the need for continual updating of evidence and current recommendations. When guidelines encompassed all aspects of treatment for a given disorder, evidence and recommendations on one treatment might be outdated even when the remainder of the guideline was still current. When guidelines are based on more discrete topics, they can be updated independently of other topics that relate to the same disorder. APA is also exploring approaches to use informatics approaches to facilitate screening of new literature in order to support targeted updating of guidelines.

ADVANTAGES AND LIMITATIONS OF GUIDELINES

Table 7.3–4 identifies a number of potential advantages and limitations of guidelines.

Advantages of Guidelines

The most important benefit, and the major reason for developing and using guidelines, is to improve the quality of care received by patients. A growing body of literature, particularly related to cardiovascular disease, suggests that guideline adherence results in better outcomes. However, the available evidence is less clear in psychiatry, at least in part due to limited guideline adherence. Such studies are difficult to conduct especially when the guideline recommendations are multifaceted. Furthermore, rates of patient participation and adherence can confound interpretation of findings.

Table 7.3–4.
Advantages and Limitations of Guidelines

Advantages of Guidelines
Improve quality of care
Decrease variation in practice
Standardize and facilitate education of professionals
Education of patients and family
Facilitates patient involvement in decision making
Stimulate research
Encourage funding of research
Support reimbursement of evidence-based care
Reduce stigma

Limitations of Guidelines
Gaps in knowledge
Reality of comorbidity
Must change with new knowledge
Cookbook rather than art
Risk of reductionism
Stifle innovation
Concern about increased liability
Cost of development and implementation
Biased by funders of development
Difficult to implement

Geographic variations are well documented for rates of hospitalizations, procedures (such as hysterectomy or coronary artery bypass grafts), medication use, and other aspects of health care utilization. Within psychiatry, one study suggests that rates of death due to depression or schizophrenia also show significant geographic variation. Although this could be biased by differences in ascertainment, it may also reflect variations in care that are associated with variations in morbidity and mortality. Having nationally developed guidelines (if they are followed) can presumably decrease such variation and improve outcomes for patients.

Another major advantage of guidelines is their use in educational programs. Guidelines that include a clear and comprehensive description of the evidence and references are especially helpful. Interestingly, in many training programs it is the residents who push for increased use of the guidelines. Content related to the guidelines is also included in The Psychiatry Residents-in-Training Examination (PRITE) administered by the American College of Psychiatrists (ACP). Use of guidelines for educational purposes extends well beyond residency training into CME and MOC programs. The American Board of Psychiatry and Neurology (ABPN) has designated the APA guidelines as a major source for their MOC examinations and questions related to the guidelines are incorporated into the FOCUS self-assessment programs. Performance-in-practice modules are also available that are linked to guideline content.

Education of patients and families is also facilitated by the existence of guidelines. Many developers of guidelines also prepare a patient and family companion document that parallel the guideline but is written in a manner understandable for a lay audience. Such material also aids in involving the patient in the decision making, a major element in patient-centered care.

As guidelines are written, gaps in our evidence base become painfully clear. An important section in many guideline projects is to identify those issues that need to be further studied. This can provide direction to the field and also can be used to support requests for additional research funding.

To clinicians, decisions by third-party payers about what treatments are reimbursed frequently seem arbitrary or not driven by clinical reasoning. Practice guidelines can be used to identify those treatments that are supported by evidence and should be reimbursed and those treatments that are shown to be ineffective and should not be reimbursed. They can also delineate situations in which a number of options are appropriate and in which patient preferences and values should be taken into consideration and supported.

Guidelines are also valuable in demonstrating that mental illnesses are real and diagnosable, and that specific disorders and symptoms can be ameliorated with specific treatments. This is a crucial role for guidelines because stigma and its effects on patients, families, and treatment continue to be a major challenge. Over the past four decades, recognition of these realities has had a significant impact, directly and indirectly, on the funding of and access to treatment of mental illnesses.

Limitations of Guidelines

In terms of limitations to practice guidelines, gaps in the knowledge base make it challenging to write authoritative guidelines without overstating what the evidence supports. Poorly written guidelines can mislead physicians and patients and discredit the scientific underpinnings of our field.

Guidelines must be regularly revised as new knowledge becomes available. As the medical literature continues to grow at exponential rates and as the number of guidelines in a specialty area increases, this becomes a significant challenge for developers of guidelines. However, if guidelines are not updated regularly, they can be misleading to clinicians, patients, families, and policy makers.

Although treatment-related costs were highlighted by Dr. David Eddy as important considerations in practice guidelines, most guidelines do not incorporate cost-effectiveness data when developing guideline recommendations. Financial data to support such considerations is rarely available, and costs of a specific treatment or service can show wide variations. Translating costs to value (e.g., using quality adjusted life years) adds further complexities and sources of possible bias. Patients also vary in their available resources and in their cost-related preferences for treatment. Despite these challenges, the usefulness of practice guidelines is likely to be enhanced by some consideration of the costs and value of recommended care.

Guidelines that focus on a single illness/disorder are limited by the reality that many patients have multiple illnesses. In psychiatry, at least 50% of patients have more than one psychiatric diagnosis. Physical conditions are also common and can require modifications in treatment or monitoring. Hence, single-illness guidelines must address these realities. Ideally, more guidelines should be developed for comorbid conditions, but this is challenging, as there is limited evidence on the simultaneous treatment of comorbid conditions.

Care is also needed to reduce the number of inconsistent recommendations among guidelines written by different groups. Often these inconsistencies result from differences in the patient population of interest (e.g., primary care vs. specialty practices) or legitimate differences of opinion in interpreting the available evidence. Nevertheless, conflicting recommendations can be confusing for patients and clinicians alike. Transparency in the synthesis of evidence and rationale for recommendations can help to reduce some of these difficulties. Reviewing the recommendations of other organizations' guidelines as part of the guideline development process can also assist in clarifying or resolving discrepancies.

A frequent criticism of guidelines, especially when they were first being introduced, is that they encourage "cookbook" medicine and interfere with the practice of the "art" of medicine by being overly prescriptive. Well-written guidelines can avoid this pitfall by presenting the existing evidence and explicitly stating the role of the physician in using the recommendations to inform decision making about care.

A related issue is concern about guidelines reinforcing reductionism. Guidelines can potentially decrease the richness and breadth of assessment and treatment strategies, and developers of guidelines must be alert to this potential. Another concern is that new treatment approaches will be discouraged, and cutting-edge innovations will not be pursued. To address both of these concerns, guidelines may need to be more inclusive and less dogmatic, always describing the available evidence clearly and noting that other approaches may also constitute proper care for the disorder being considered.

Clinicians have been concerned that the presence of guidelines would lead to an increase in successful malpractice actions, as plaintiff attorneys could point out discrepancies between guideline recommendations and rendered care. Although evidence is limited, it appears that guidelines have been used by defendants as well as by plaintiffs to support their cases without significant increases in the number or severity of successful malpractice claims. To mitigate the possibility for misuse of guidelines in litigation, guideline writers should avoid being overly prescriptive unless the evidence is overwhelming and other choices are not reasonable. It must also be clearly stated that the recommendations are indeed guidelines, not standards of care. Instead, recommendations should be considered by the physician and decisions should be based on the needs of the individual patient, the available diagnostic and treatment options, and other circumstances.

Another limitation of guideline development is cost, which can be considerable. The federal government's Agency for Health Care Policy and Research developed 19 guidelines from 1992 to 1996, but the program was discontinued in part because of the considerable cost. As more scientific literature is available for review and as systematic review standards have become more rigorous, the costs of developing a guideline are even higher than in prior decades. While some of this cost can be offset by publication revenue from the guideline and related materials and income from educational programs, developing and disseminating guidelines is generally not a profitable undertaking. Some non-APA guidelines have been developed with the support of industry, especially pharmaceutical companies. This financial support raises a concern about the scientific integrity of the process and the independence of the recommendations.

Costs of implementation can also play a role in reducing guideline adoption with several studies noting that lack of administrative support and insufficient staff time were significant barriers. More recently, there has been increasing integration of guideline recommendations and related quality measures into electronic health records (EHRs), which can also add to implementation costs for software modification and staff time. In addition to these costs of guideline implementation, opportunity costs are often overlooked. If clinicians are focusing on delivering guideline-concordant care, they may overlook or lack the time to focus on other issues that would be of greater benefit to a particular individual.

Other difficulties in guideline implementation are also important to note. There are numerous studies indicating that often physicians do not use published guidelines. M.D. Cabana identified six primary barriers to guideline use: Lack of awareness, lack of familiarity, lack of agreement, lack of self-efficacy, lack of outcome expectancy, and inertia of previous practice. Some of these barriers can be surmounted through specific interventions. For example, some studies have shown that patient involvement in medical decisions increases the probability of receiving guideline concordant care. Other studies found that physician feedback or multifaceted educational programs including the use of opinion leaders improves the rate of physicians using guidelines. It has also been shown that use of guidelines may be enhanced by incorporating recommendations into EHRs. J.M. Grimshaw concluded that multifaceted interventions targeting different barriers to change in physician behavior are more likely to be effective than single interventions.

THE FUTURE OF PRACTICE GUIDELINES

There are now thousands of guidelines for all areas of medicine, which have been developed by professional associations, insurance companies, health care systems, and regional and national governments. The number of new guidelines will continue to increase dramatically over the next decade and will play an increasing role in the delivery of health care. With the integration of guidelines into EHRs, recommendations can be presented to clinicians at the point of service with more direct effects on clinical decisions. More research is needed on other tools and incentives to increase the use of guidelines. At the international level, greater efforts will be needed to explore the applicability of guidelines among different countries and to adjust recommendations on the basis of cultural differences, health care delivery systems, and available resources. If guidelines are well developed, truly evidence based, and appropriately flexible to meet the needs and wishes of individual patients, their use can significantly improve the quality of care for patients.

FURTHER READINGS

AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care*. 2003;12:18–23.

Ament SMC, de Groot JJA, Maessen JMC, Dirksen CD, van der Weijden T, Kleijnen J. Sustainability of professionals’ adherence to clinical practice guidelines in medical care: a systematic review. *BMJ Open*. 2015;5(12):e008073.

American Psychiatric Association, Standing Committee. Report on the organization of hospitals for the insane. *Am J Insanity*. 1853;67.

Andrews G, Armstrong MS, Brodaty H. A methodology for preparing ‘ideal’ treatment outlines in psychiatry. *Aust N Zeal J Psychiatry*. 1982;16:153.

Boyd CM, Darer J, Boult C, Fried LP, Boult L, Wu AW. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. *JAMA*. 2005;294:716–724.

Cabana MD, Rushton JL, Rush AJ. Implementing practice guidelines for depression: applying a new framework to an old problem. *Gen Hosp Psychiatry*. 2002;24:35–42.

Chassin MR, Brook RH, Park RE, et al. Variations in the use of medical and surgical services by the Medicare population. *N Engl J Med*. 1986;314:285–290.

Eddy DM. Clinical decision making: from theory to practice. Connecting value and costs. Whom do we ask, and what do we ask them? *JAMA*. 1990;264(13):1737–1739.

Gaebel W, Weinmann S, Sartorius N, Rutz W, McIntyre JS. Schizophrenia practice guidelines: international survey and comparison. *Br J Psychiatry*. 2005;187:248–255.

*Gagliardi AR, Alhabib S; Members of Guidelines International Network Implementation Working Group. Trends in guideline implementation: a scoping systematic review. *Implement Sci*. 2015;10:54.

Girlanda F, Fiedler I, Becker T, et al. The evidence-practice gap in specialist mental healthcare: systematic review and meta-analysis of guideline implementation studies. *Br J Psychiatry*. 2017;210(1):24–30.

Guyatt GH, Oxman AD, Vist GE, et al.; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924–926.

Hepner KA, Rowe M, Rost K, et al. The effect of adherence to practice guidelines on depression outcomes. *Ann Intern Med*. 2007;147:320–329.

Institute of Medicine, Committee on Clinical Practice Guidelines. *Guidelines for Clinical Practice: From Development to Use*. National Academy Press; 1990.

*Institute of Medicine. In: Graham R, Mancher M, Wolman DM, et al. eds. *Committee on Standards for Developing Trustworthy Clinical Practice Guidelines; Clinical Practice Guidelines We Can Trust*. National Academies Press (US); 2011.

Institute of Medicine. In: Eden J, Levit L, Berg A, Morton S. *Committee on Standards for Systematic Reviews of Comparative Effectiveness Research. Finding What Works in Health Care: Standards for Systematic Reviews*. National Academies Press (US); 2011.

Institute of Medicine. *Knowing What Works in Health Care: A Roadmap for the Nation*. The National Academies Press; 2007.

Keepers GA, Fochtmann LJ, Anzia JM, et al. The American Psychiatric Association practice guideline for the treatment of patients with schizophrenia. *Am J Psychiatry*. 2020;177(9):868–872.

Krahn M, Naglie G. The next step in guideline development: incorporating patient preferences. *JAMA*. 2008;300(4):436–438.

Meeuwissen JAC, Feenstra TL, Smit F, et al. The cost-utility of stepped-care algorithms according to depression guideline recommendations: Results of a state-transition model analysis. *J Affect Disord*. 2019;242:244–254.

Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Content last reviewed October 2022. Effective Health Care Program, Agency for Healthcare Research and Quality, Rockville, MD. https://effectivehealthcare.ahrq.gov/products/collections/cer-methods-guide

Polednak AP. US regional differences in death rates from depression. *Soc Psychiatry Psychiatr Epidemiol*. 2012;47(12):1977–1983.

Pronovost PJ. Enhancing physicians' use of clinical guidelines. *JAMA*. 2013;310(23):2501–2502.

Reus VI, Fochtmann LJ, Bukstein O, et al. The American Psychiatric Association practice guideline for the pharmacological treatment of patients with alcohol use disorder. *Am J Psychiatry*. 2018;175(1):86–90.

Reus VI, Fochtmann LJ, Eyler AE, et al. The American Psychiatric Association practice guideline on the use of antipsychotics to treat agitation or psychosis in patients with dementia. *Am J Psychiatry*. 2016;173(5):543–546.

*Silverman JJ, Galanter M, Jackson-Triche M, et al. The American Psychiatric Association practice guidelines for the psychiatric evaluation of adults. *Am J Psychiatry*. 2015;172(8):798–802.

Van Duin D, Franx G, Van Wijngaarden B, et al. Bridging the science-to-service gap in schizophrenia care in the Netherlands: the Schizophrenia Quality Improvement Collaborative. *Int J Qual Health Care*. 2013;25(6):626–632.

*Weinmann S, Koesters M, Becker T. Effects of implementation of psychiatric guidelines on provider performance and patient outcome: systematic review. *Acta Psychiatr Scand*. 2007;115(6):420–433.

Yager J, Kunkle R, Fochtmann LJ, et al. Who's your expert? Use of an expert opinion survey to inform development of American Psychiatric Association practice guidelines. *Acad Psychiatry*. 2014;38:376–382.

Zarin DA, Pincus HA, McIntyre JS. Practice guidelines. *Am J Psychiatry*. 1993;150:175.

7.4 Neuropsychological Assessment of Adults

JOSEPH R. SADEK, PH.D., A.B.P.P.-C.N., OWEN GAASEDELEN, PH.D., REX M. SWANDA, PH.D., A.B.P.P.-C.N., AND KATHLEEN Y. HAALAND, PH.D., A.B.P.P.-C.N.

NEUROPSYCHOLOGICAL ASSESSMENT OF ADULTS

Clinical neuropsychology is a specialty in psychology that examines the relationship between behavior and brain functioning in the domains of cognitive, motor, sensory, and emotional functioning. The clinical neuropsychologist integrates the medical and psychosocial history with the reported complaints and the pattern of performance on neuropsychological procedures to determine whether results are consistent with a particular area of brain damage or a particular diagnosis. Although neurologic conditions are often the focus of referrals, the neuropsychological examination has a valuable place in diagnosing and identifying the pattern of behavioral and cognitive abilities that can be associated with various medical, psychological, and psychiatric diagnoses. Neuropsychology also plays a role in medicolegal and forensic applications, such as determination of decisional capacity, disability determination, and litigation over whether deficits are related to an injury.

WHO IS A CLINICAL NEUROPSYCHOLOGIST?

Clinical neuropsychology has well-established guidelines for standards of training and specialized practice in neuropsychology. In 1996, the field of neuropsychology was formally recognized as a distinct specialty in psychology by the American Psychological Association. The 1997 Houston Conference resulted in a policy statement that formally identified training experiences appropriate to developing a specialization in clinical neuropsychology. These training guidelines call for both specialized and generalized clinical training throughout a program of doctoral studies, followed by an internship and postdoctoral residency, leading to eligibility for specialty board certification through the American Board of Professional Psychology (ABPP), which is the parent board of the American Board of Clinical Neuropsychology (ABCN). Efforts are underway to update clinical neuropsychology training to apply competency-based standards put forth by the American Psychological Association.

RELATIONSHIP TO OTHER DISCIPLINES

The unique contributions of neuropsychology can be understood by examining its relationship to the closely allied disciplines of clinical psychology, behavioral neurology, and neuropsychiatry.

Clinical Psychology

Clinical neuropsychology is distinct from general clinical psychology by applying knowledge of brain–behavior relationships to assessing human behavior. Both areas rely on the psychometric foundations of reliability, validity, and normative standards to define behavioral symptoms and complaints objectively. Emotional factors are also the province of both. In evaluating cognition, the clinical psychologist focuses on issues involving general intellectual, academic, and vocational skills compared to clinical neuropsychology’s focus on neurologic factors. Clinical psychologists and clinical neuropsychologists are both involved in treatment, including psychotherapy and cognitive retraining to remediate deficits. Psychoeducation is also provided by both fields in a therapeutic context that discusses the relationship between symptoms and the assessment results and provides recommendations for patients, caregivers, and other health professionals.

Behavioral Neurology and Neuropsychiatry

The medical specialties of behavioral neurology and neuropsychiatry overlap considerably with clinical neuropsychology. However, the neuropsychologist's unique contribution to assessing cognitive deficits is based on a theoretical background shared with cognitive psychology and the neurosciences. It also uses rigorous, psychometrically based methods for developing and applying standardized tests with demographically appropriate normative data whenever possible. In contrast, the behavioral neurologist and neuropsychiatrist are more likely to utilize a mental status examination approach for identifying cognitive deficits. These clinicians typically rely on "internal norms," which are based upon the individual clinician's extensive experience with the examination procedures or pathognomonic signs (e.g., dysfluent speech or visual neglect). The behavioral neurologist is trained to diagnose a broad spectrum of neurologic disorders; the neuropsychiatrist's focus lies in diagnosing neurologic features of psychiatric patients though that focus has broadened in recent years. The neuropsychologist focuses on a formal assessment of various disorders' cognitive and behavioral manifestations that can produce cognitive and behavioral impairment in medical, surgical, neurologic, and psychiatric patients. Ideally, all three work together to provide complementary perspectives.

HISTORICAL INFLUENCES

Clinical neuropsychology has its roots in psychology, neurology, and psychiatry. The groundbreaking work of Broca and Wernicke in the 19th century first suggested that complex functions, such as speech and language comprehension, could be localized in particular areas in the left hemisphere. In 1909, K. Brodmann published a cytoarchitectonic map of the cerebral cortex (Fig. 7.4–1) based on different histologic patterns of cells in various parts of the cortex. Over the years, this representational map has proven valuable in identifying functional differences across distinctive cortical regions. Renewed interest in functional localization grew in the 1960s with Norman Geschwind's work in identifying how connections among different parts of the brain were most important in producing complex behavior. Arthur Benton's Iowa School developed a series of carefully constructed psychometric tests based on concepts originally identified by behavioral neurology to assess specific deficits relative to the normal population. At about the same time, Hans-Lukas Teuber defined the concept of "double dissociation," which is regarded as the strongest evidence for the localization of a particular function. This notion is based on observations of mutually exclusive brain–behavior relationships, such that damage affecting region 1 produces performance deficits in test 1 but not in test 2, while damage affecting region 2 produces performance deficits in test 2 but not in test 1. Aleksander Luria, the Russian neuropsychologist, also developed notions of functional localization but utilized a theoretical framework that associated component cognitive processes with complex skills and their neuroanatomical correlates. In the United States, Ralph Reitan applied the psychometric standards of North America to several instruments for the express purpose of assessing brain damage, resulting in the *Halstead–Reitan Neuropsychological Test Battery (HRNTB).* Current neuropsychological assessment practices attempt to use an efficient hypothesis testing approach, in which a core screening protocol is supplemented by procedures that assess specific areas of functioning. This approach is an outgrowth of the Boston or process approach to neuropsychological assessment combined with the more rigid and time-consuming HRNTB approach.

More recently, the development of increasingly sophisticated imaging techniques in the neurosciences (magnetic resonance imaging [MRI], computed tomography [CT], positron emission tomography [PET], functional MRI with connectivity analyses, and diffusion tensor imaging [DTI]) has brought dramatically enhanced capacity to integrate neuroimaging results with neuropsychological methods for measuring cognitive and emotional processing. Advances in neuroimaging have also influenced advances in cytoarchitectonic mapping. While Brodmann's map still functions as a standard reference for identifying cortical areas, an improved understanding of gray and white matter functional architecture leads to more detailed concepts (Amunts et al., 2020). The advances in brain mapping have not yet translated into similar advances in behavioral assessment.

NEUROANATOMICAL CORRELATES

The early history of neuropsychology was primarily driven by the goal of linking behavioral deficits to specific neuroanatomical areas of damage. Although this early assessment method helped validate neuropsychological tests commonly used today, the localizing function of neuropsychological assessment is now less critical since neuroimaging can easily visualize brain lesions. Increasing knowledge in neurosciences has also led to a more sophisticated view of brain–behavior relationships. Complex cognitive, perceptual, and motor activities are controlled by neural circuits and related small functional units rather than single structures within the brain. When evaluating patients with brain damage, it is necessary to understand these brain–behavior relationships and how disease states compromise their functioning. The following section will briefly review some basic concepts of brain–behavior relationships used routinely by the neuropsychologist to interpret neuropsychological results.

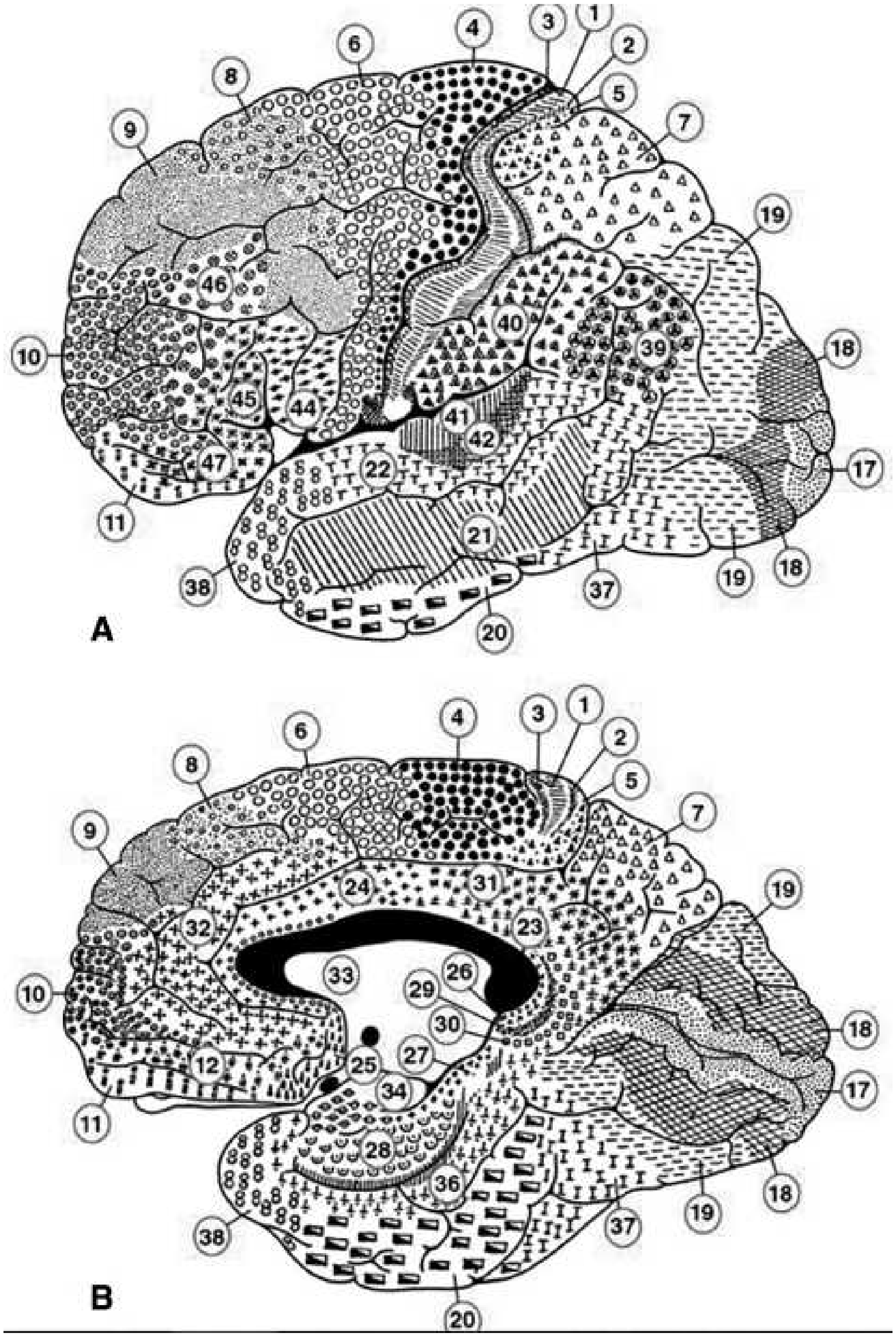
6
4
3
1
2
5
8
7
9
19
46
40
39
10
18
45
44
41
42
47
22
17
11
21
38
37
19
18
20
A
6
4
3
1
2
5
8
7
9
24
31
19
32
23
10
33
29
26
30
18
12
25
27
34
17
11
28
36
18
38
37
19
B
20

FIGURE 7.4–1. Brodmann's cytoarchitectonic mapping of the human brain. **A:** Lateral view. **B:** Medial view. (From Parent A. *Carpenter's Human Neuroanatomy*. 9th ed. Williams & Wilkins; 1996;882.)

Hemispheric Dominance and Intrahemispheric Localization

Many functions are mediated by both the right and left hemispheres. However, critical qualitative differences between the two hemispheres can be demonstrated after lateralized brain injury. Table 7.4–1 lists some cognitive skills linked to right-handed individuals' left or right hemisphere. Although language is the most obvious function controlled mainly by the left hemisphere, even among most left-handed individuals, the left hemisphere is also generally considered dominant for limb praxis (i.e., performing complex movements, such as brushing teeth). In contrast, the right hemisphere plays a more critical role in controlling visuospatial abilities and hemispatial attention, associated with the clinical presentations of constructional apraxia and neglect, respectively. Although lateralized deficits such as these are typically characterized by damage to the right or left hemisphere, it is essential to remember that preserved brain functions can also characterize the patient's performance. In other words, it is the remaining intact brain tissue that drives many behavioral responses following injury to the brain.

Table 7.4–1.
Selected Neuropsychological Deficits Associated with Left or Right Hemisphere Damage

Left Hemisphere	Right Hemisphere
Aphasia	Visuospatial deficits
Right–left disorientation	Impaired visual perception
Finger agnosia	Hemispatial neglect
Dysgraphia (aphasic)	Spatial memory
Dyscalculia (number alexia)	Dyscalculia (spatial)
Constructional apraxia (details)	Constructional apraxia (Gestalt)
Limb apraxia	Dressing apraxia
Verbal memory	Anosognosia

Language Disorders

LESIONS AND THE APHASIAS. Appreciation for the unique role of the "dominant" left hemisphere in controlling language in most right-handed individuals has been validated in many focal lesion, functional neuroimaging, and neurosurgery studies. Although it is rare for right-handed individuals to be right hemisphere dominant for language, it does occur in about 1% of cases. Hemispheric dominance for language in left-handed individuals is less predictable. About two-thirds of left-handed individuals are still left-hemisphere dominant for language, while about 20% each are right hemisphere dominant or bilaterally dominant.

Several classification systems have been developed over the years to describe various language breakdown patterns. A common method takes into account the presence or absence of three key features: (1) fluency, (2) auditory comprehension, and (3) repetition (i.e., intact ability to repeat verbally presented words or phrases). All of the following four types of aphasia have damage within some part of the middle cerebral artery's core distribution on the left hemisphere's lateral surface. Table 7.4–2 details the characteristics of *Broca's, Wernicke's, conduction, and global aphasia.* Each is characterized by different areas of brain damage within the middle cerebral artery territory and a different combination of nonfluent speech, impaired auditory comprehension, and repetition, or a combination of all three problems. In addition, more recent neuroimaging results in stroke patients show that the lesion locations specified in Table 7.4–2 may not be broad enough. For example, the full syndrome of *Broca aphasia,* including *agrammatism* (telegraphic speech), is found only in the presence of more extensive damage (i.e., suprasylvian area from Broca area to the posterior extent of the Sylvian fissure). Wernicke aphasia is associated with damage to the superior and middle temporal gyrus as well as the inferior frontal and parietal cortices, not just the superior temporal gyrus. The comprehension deficits seen with Wernicke aphasia result in impaired ability to comprehend their own language. Therefore, in contrast to patients with Broca aphasia who repeatedly try to correct their speech errors, the patient with *Wernicke aphasia* is typically not aware of their communication problems. This lack of insight is similar to the condition of *anosognosia,* in which patients fail to appreciate their own deficits, and presents a particularly frustrating condition for many family members and caregivers. *Conduction aphasia* produces more subtle, focal repetition deficits that negatively impact daily functioning somewhat less. *Global aphasia* is characterized by impairment in all three language functions due to damage to the core language areas on the lateral surface of the left hemisphere. Importantly, this system cannot neatly classify many aphasic patients because the pattern of deficits does not precisely fit clear, descriptive categories. Detailed language assessment of most aphasic patients typically demonstrates deficits in all three areas, although the degree of deficit among the three areas varies.

Table 7.4–2.
Aphasia Types with Traditional Lesion Loci

Aphasia	Primary Language Impairment	Lesion Location
Broca’s	Production	Posterior inferior frontal gyrus
Wernicke’s	Auditory Comprehension	Posterior superior and middle temporal gyrus
Conduction	Repetition	Arcuate fasciculus
Global	Production, auditory comprehension, and repetition	Broad area of the middle cerebral artery

NEURODEGENERATIVE LANGUAGE DISORDERS. Aphasia can be the initial presenting problem in variants of frontotemporal dementia (FTD), called *primary progressive aphasia*. Age of onset is typically in the sixth decade of life, and prevalence is less common than Alzheimer disease. Neuropsychological testing can help determine whether aphasia is the primary problem and can help determine whether it meets the criteria for the subsyndromes of language variant FTD. Subtypes of language variants are listed in Table 7.4–3.

Limb Apraxia. Limb apraxia and other cognitive-motor skills deficits are more commonly seen with left than right hemisphere damage. However, differences in the incidence of limb apraxia after left or right hemisphere damage is not as significant as with language, suggesting that left hemisphere dominance for disorders of complex movement is not as strong as that for language. Studies demonstrate that deficits in daily functioning are common with limb apraxia and that it negatively affects rehabilitation outcomes. Different types of limb apraxia have been associated with damage to different parts of the brain, especially in the left parietal, temporal and frontal regions. For instance, *ideomotor limb apraxia* is characterized by impaired spatiotemporal execution of complex movements, which results in orientation and timing errors, such as carving a turkey by moving the knife in a “jerky” up and down motion rather than smooth back and forth movements. These deficits have been associated with impairment of the “how” occipitoparietal circuit, which stores movement representations for familiar movements. *Ideational apraxia* might result in using the wrong object to perform a movement, such as attempting to use a toothbrush to eat, and sequencing errors, such as trying to light a candle before striking the match. Both types of problems have been associated with impaired semantic knowledge, characterized by an impaired appreciation of the object’s function and damage to the occipitotemporal lobe “what” system. Assessment of limb praxis can be challenging because aphasia and hemiparesis can influence the diagnosis, and there are relatively few standardized tests of limb apraxia, though a screening assessment that can be easily used at the bedside is now available.

Table 7.4–3.
Progressive Aphasia Types

Subtype	Clinical Features	Common Pathology and Location
Progressive nonfluent Aphasia	▸ Agrammatism ▸ Effortful, halting speech ▸ Impaired complex comprehension	Tau Left inferior frontal gyrus
Semantic dementia	▸ Impaired confrontation naming ▸ Impaired word comprehension ▸ Impaired object knowledge	TDP-43 Left anterior temporal lobe
Logopenic aphasia	▸ Impaired word finding ▸ Impaired repetition ▸ Phonologic paraphasias	AD Plaques and tangles Left temporoparietal junction

AD, Alzheimer disease; TDP-43, transactive response (TAR) DNA binding protein with Mr 43 kDa.

Adapted from Graff-Radford J, Jones DT, Graff-Radford NR. Pathophysiology of language, speech and emotions in neurodegenerative disease. *Parkinsonism Relat Disord*. 2014;20(S1):S49–53. Copyright © 2014, with permission from Elsevier. doi: 10.1016/S1353-8020(13)70014-2.

Visuospatial Disorders. Right hemisphere damage in right-handed individuals is frequently associated with deficits in visuospatial skills. Common assessment techniques include drawings and constructional or spatial assembly tasks. Although spatial problems are not recognized as frequently as language disorders, they can be highly significant from a functional standpoint. For example, they can be associated with getting lost even in familiar places and difficulty using a map to navigate unfamiliar places. Persons with hemispatial neglect (most frequently affecting the left hemispace) might have difficulty reading or reading an analog clock due to neglect of the left hemispace. In addition, persons with hemispatial neglect are at high risk for driving problems and, in extreme cases, might even leave food uneaten on the neglected side of a plate, for example. Spatial disorders might also produce difficulty dressing due to problems differentiating the garment's top, bottom, or left or right. Distinctive qualitative errors in constructing a complex geometric configuration can be seen with right or left hemisphere damage, though the deficits are typically more extreme after right hemisphere damage. For example, a patient with right parietal damage includes isolated design details while failing to convey the interrelationship of the details in the overall design configuration.

In contrast, a patient with left hemisphere damage tends to maintain the design's global framework or Gestalt but lose the details. See Figure 7.4–2 for an example from a common block design test. Therefore, many neuropsychologists emphasize that a neuropsychological understanding of the impairment depends not just on test scores but also on a qualitative description of the type of error. This often allows the impairment to be linked to a specific neuroanatomical region and enables a better understanding of the mechanisms of the deficit for rehabilitation purposes. This qualitative focus on the type of error is similar to the *pathognomonic* approach that behavioral neurologists often use.

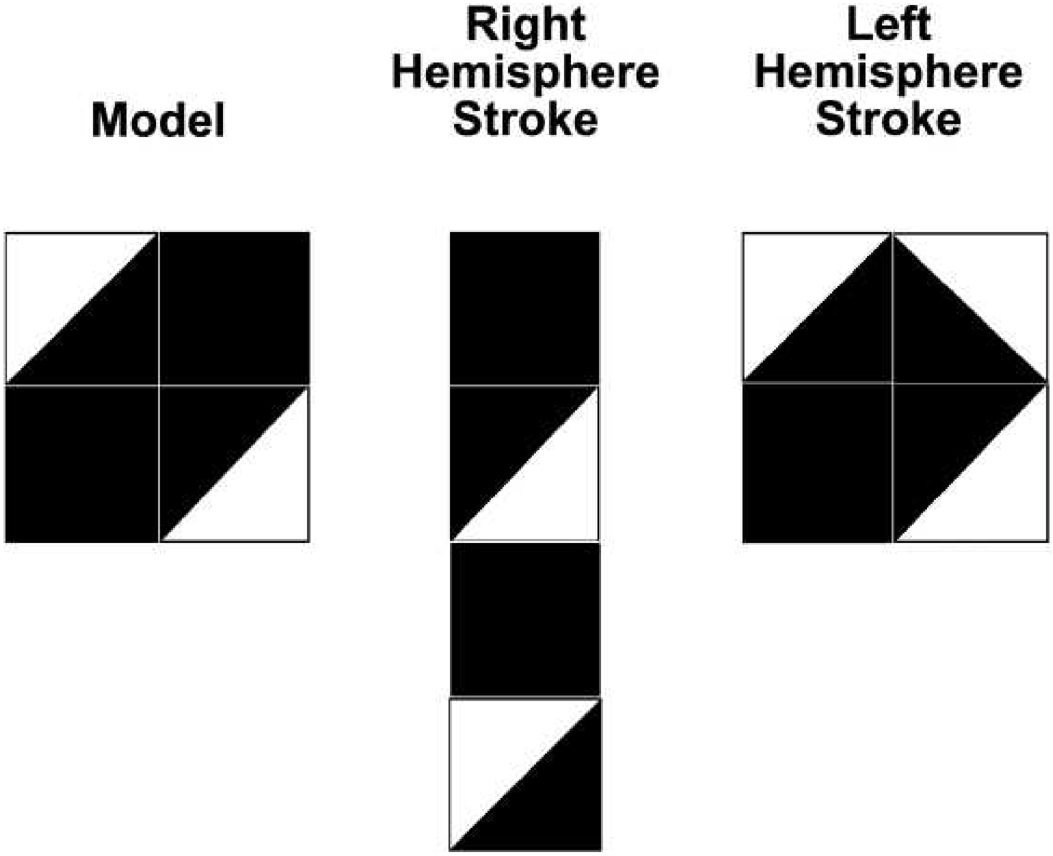

FIGURE 7.4–2. Block Design Examples.

NEGLECT. *Neglect syndromes* are characterized by failure to detect visual or tactile stimuli or move the limb in the contralateral hemispace. They are most commonly associated with right hemisphere damage in the parietal region, but damage to other areas within the cerebral cortex and subcortical areas can also produce this problem. Although neglect syndromes have a similar incidence and may co-occur with visual field cuts or somatosensory deficits, the neglect syndrome is distinct and not explained by any primary motor or sensory problems that may be present. Visual neglect can be assessed with line cancellation and line bisection tasks, in which the paper is placed at the patient's midline, and the patient is asked to either cross out all of the lines on the page or to bisect a single line. One of the most sensitive ways to assess visual neglect is with double simultaneous stimulation, which tests for the tendency to make errors in the contralesional hemispace when competing stimulation is in the ipsilesional hemispace. Neglect syndromes can have devastating functional effects on safety and the ability to live independently and should be considered standard in the evaluation process. For example, when working with a patient with visual neglect, the examiner should stand in the patient's nonneglected (usually right) hemispace to avoid being ignored. It is also essential to position a patient in an inpatient hospital setting so that the patient's "neglected" hemispace is oriented away from the door, thereby facilitating the patient's ability to pay attention to visitors entering the room.

Memory Disorders. Memory complaints constitute the most common referral to neuropsychology. Thorough neuropsychological examination of memory includes the verbal and/or spatial modality. Presentation formats systematically assess different aspects of information processing and storage that form the basis for human memory. Accumulated research indicates that specialized verbal and spatial memory material processing tends to be differentially mediated by the left and right hemispheres, respectively. In addition to interhemispheric differences in functional localization, specific memory problems can be associated with a breakdown at any stage in the information processing model of memory (Fig. 7.4–3). These stages include (1) registration of the material through *attention,* (2) initial processing and encoding of the material within *short-term memory,* (3) consolidation and storage of material in *long-term memory,* and (4) *retrieval* processes, in which material is accessed from long-term memory storage. A significant advantage of neuropsychological assessment is that these various types of memory problems can be specifically identified and described in the examination procedures. Once identified, the specific nature of the deficit can then have important implications for diagnosis, treatment, and prognosis.

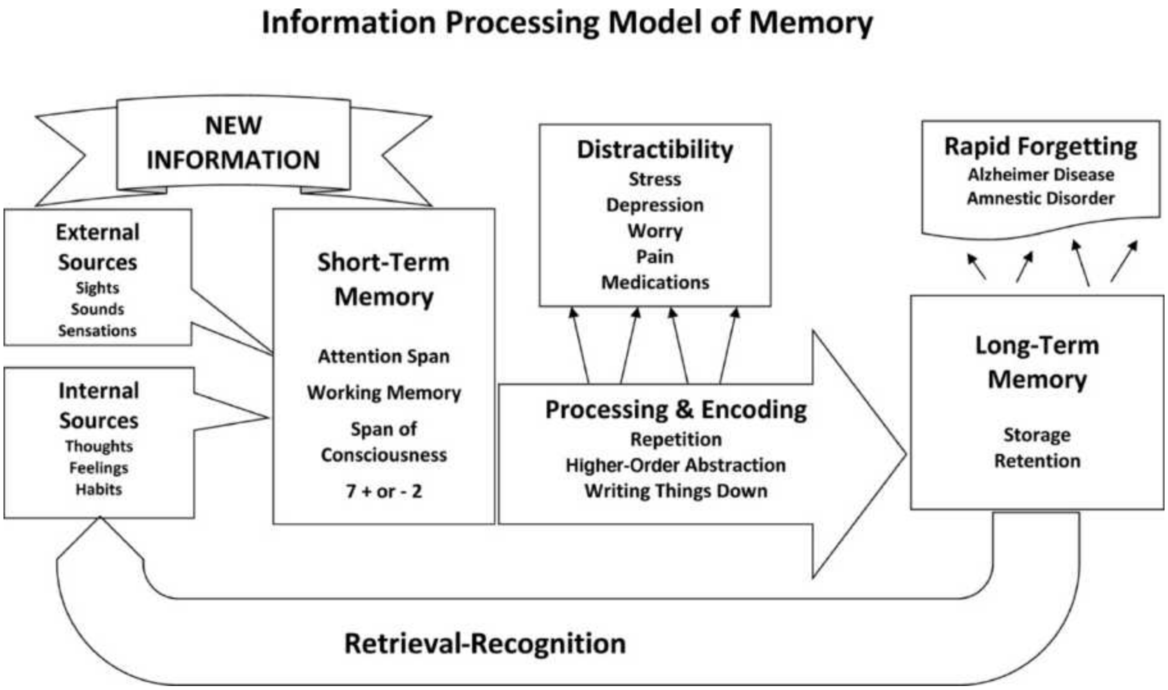
Information Processing Model of Memory
NEW INFORMATION
External Sources
Sights
Sounds
Sensations
Internal Sources
Thoughts
Feelings
Habits
Short-Term Memory
Attention Span
Working Memory
Span of Consciousness
7 + or - 2
Distractibility
Stress
Depression
Worry
Pain
Medications
Processing & Encoding
Repetition
Higher-Order Abstraction
Writing Things Down
Rapid Forgetting
Alzheimer Disease
Amnestic Disorder
Long-Term Memory
Storage
Retention
Retrieval-Recognition

FIGURE 7.4–3. Information Processing Model of Memory.

ENCODING. The initial encoding of new material can be influenced by various factors, including deficits in attention, language, and spatial processing abilities. It is usually measured by the immediate recall of newly learned information (e.g., narrative stories or designs) or by demonstrating the ability to learn new material that has been presented across multiple “learning trials” (e.g., word lists). Attention itself is a relatively fragile cognitive function that can be affected by many factors, including neurologically based disorders (e.g., traumatic brain injury [TBI], delirium) and psychiatric disorders (e.g., depression or anxiety, attention deficit disorder), so it is a crucial aspect of a proper memory assessment.

STORAGE AND RETRIEVAL. Storage of new information (retention) is assessed by determining the percent of the information recalled following a delay interval compared to what was initially recalled. Deficits in the delayed recall can be associated with impaired storage and retention of new information, or it can be due to impaired retrieval, in which case the material has been stored but not readily accessible. The best way of differentiating these problems is to examine *recognition memory,* in which a patient is typically asked to choose from a set of multiple-choice alternatives or discriminate target words from *false-positive* foils. Poor retrieval is the most likely problem if the patient demonstrates accurate recognition but poor recall. However, if recognition and recall are both impaired, the most likely problem is impaired encoding or storage of new information. This distinction is vital because different neuroanatomical systems subserve the functions of retrieval and storage. Impaired storage is more often associated with dysfunction of the medial temporal lobe–diencephalic systems, whereas impaired retrieval can be associated with various structures, especially the frontal lobes.

Executive Function. The prefrontal cortex and its interconnections with the rest of the brain play an important role in *executive functions,* which are essential for planning and organizing complex responses and self-monitoring, inhibiting inappropriate responses, and flexibly switching from one response strategy to another. Damage to the frontal lobes has also been associated with significant personality changes. This was historically exemplified by the famous 19th-century case of Phineas Gage, who became irresponsible, socially inappropriate, and unable to carry out plans after a tamping iron was blown through his frontal lobes. As conceptualized by Lezak, the executive functions include volition (i.e., formulation of a goal, motivation to achieve the goal, and awareness of one’s own ability to achieve the goal), planning, purposive action (response selection and initiation, maintenance, switching, and stopping), and execution, which involves self-monitoring and self-correction as well as control of the spatiotemporal aspects of the response. Although hemispheric differences in the control of executive functions by the frontal lobes have not been as well documented as language and spatial abilities in the parietal and temporal lobes, the Montreal Neuropsychology Group led by Brenda Milner has identified asymmetry in the control of verbal and design fluency.

Jeffrey Cummings proposed a model in which the prefrontal cortex and its interconnections are separated into the dorsolateral, orbitofrontal, and anterior cingulate divisions. In this model, various neurobehavioral functions are linked to corresponding neuroanatomical relationships between cortical and subcortical areas, especially the basal ganglia and thalamus. The *dorsolateral* prefrontal circuit is associated with higher cognitive functions, such as working memory, hypothesis generation, retrieval and initiation of unique words and designs, organization of information (e.g., clustering word lists based on semantic category), and development of alternating and sequential motor programs. The *orbitofrontal* syndrome is characterized by marked personality changes, such as instability, lability, and impulsivity. The *anterior cingulate* circuit is associated with impaired response initiation (i.e., akinetic mutism in the most extreme case) and response inhibition. These patients have difficulty inhibiting prepotent responses on “go–no go” tests and may be perceived by their families to be depressed, even though the patient might be content doing nothing. Yana Suchy has provided a more detailed description of these executive function deficits, including the Dysexecutive, Disorganized, Disinhibited/Initiation/Maintenance/Apathetic, and Socially Inappropriate Syndromes.

Motor Skills. The neuropsychological evaluation commonly includes formal motor skills tests, such as finger tapping speed, grip strength, and fine motor dexterity (e.g., grooved pegboard test). These tests, which have demonstrated validity and reliability, help assess lateralized motor impairment and determine the extent that motor deficits may confound the assessment of cognitive functions. Such motor deficits also have important implications for functioning in daily life and vocational planning.

PSYCHOMETRIC ISSUES

One of the strengths of neuropsychological assessment lies in the detailed quantitative record and normative standards for different areas of cognitive performance, which are the basis for conclusions. The normative database for any given test is significant to consider to ensure that deficits seen on individual test performances cannot be explained by demographic characteristics of the patient, such as age, education, language, or other social considerations. Of course, this approach is also dependent on the care used to develop and validate specific tests. The American Psychological Association has established ethical guidelines for psychological assessment that require tests to have acceptable levels of validity and reliability for the particular clinical application in which they are used. These guidelines serve as the basis for valid judgments about change over time, based on a given individual’s performance over successive examinations. This approach to standardization also ensures that test scores can be regarded as comparable even when administered by different examiners and in different settings. The American Academy of Clinical Neuropsychology has proposed uniform labeling of normed test performance to encourage more consistent descriptions of performances across providers and, therefore, more consumer-friendly reporting of results. Table 7.4–4 contains those labels and test descriptors.

Validity

Validity refers to a test's ability to measure what it purports to measure. Validity can be ascertained by (1) correlating scores from a particular test with other, well-established measures of the same skill, (2) relating test scores to diagnostic classifications or to measures of brain structure and function, or (3) relating test outcomes to aspects of everyday functions that are thought to be dependent on that skill. Historically, neuropsychological tests have been validated by examining their ability to differentiate brain-damaged patients from non–brain-damaged patients. More recently, they have been validated against neuroradiologic or neurosurgical data and diagnostic categories (e.g., patients with multiple sclerosis vs. medical controls). The movement to establish evidence that traditional neuropsychological tests can predict real-world outcomes has resulted in an empirical basis for inferring everyday functional problems from clinical neuropsychological test performance. Even validated measures directly assess everyday abilities, such as managing personal finances.

Table 7.4–4.
Uniform Labels of Neuropsychological Test Performance

Preamble: "In the Results section, the following phrases describe performance within the range of percentile ranks indicated in parentheses. Unless otherwise indicated, these descriptors are based on age-appropriate statistical norms and the aforementioned reservations. When adequate group norms are unavailable, more general terms reflecting behavioral observation and clinical judgment are used (e.g., "normal," "intact," "weak").

Label	Percentile Range
Exceptionally high	(98 or greater)
Above average	(91–97)
High average	(75–90)
Average	(25–74)
Low average	(9–24)
Below average	(2–8)
Exceptionally low	(<2)

Reliability

Reliability refers to the consistency of test scores. Tests are judged by (1) their internal consistency, or the extent to which all items appear to be measuring the same thing, (2) their ability to yield similar scores across multiple test occasions in normal individuals, and (3) the consistency in scores that are obtained when different people administer and score a test. Individual cognitive measures often have lower reliabilities than composite scores derived from several measures, and even within a particular ability domain, some measures exhibit higher reliabilities than others. For example, within the Wechsler Memory Scale-IV (WMS-IV), some individual tests of immediate and delayed memory have relatively low test–retest reliability (e.g., $r = 0.67$ and 0.64, for immediate and delayed visual reproduction, respectively), but the composite index scores for the Immediate Memory Index and Delayed Memory Index have higher test–retest stability ($r = 0.81$ and 0.82, respectively). This underscores the importance of using more than one test to measure each type of ability in a thorough neuropsychological evaluation. Therefore, it is crucial to consider reliability if only one test measure is used.

Predictive Utility

Predictive utility considers the sensitivity (true positives) and specificity (true negatives) of a test. A test that yields positive findings for most persons with the disease and negative outcomes for most persons without disease has high sensitivity and specificity. Published values for sensitivity and specificity are often based on ideal comparisons (e.g., clear cases of Alzheimer disease vs. carefully screened normal controls) and do not permit a clinician to estimate how useful a test will be in a general clinical population. It is essential to estimate in advance the predictive utility of particular tests for particular diagnoses because the predictive value of neuropsychological tests varies considerably for different diagnostic questions. For example, as discussed above in the section on *Neuroanatomical Correlates*, some particular neuropsychological deficits can be linked to injury in particular areas of the brain. However, many other neurologic diagnoses involve multiple brain systems that result in generalized impairment across a battery of tests. In testing for dementia, for example, neuropsychological test results are typically more sensitive than they are specific, so the interpretation must consider other medical and historical data to arrive at a more specific diagnosis. This underscores the importance of interpreting neuropsychological data in the context of the patient’s history and carefully examining the pattern of deficits and abilities with particular attention to demonstrating consistent deficits across different tests of a particular cognitive domain to reach greater accuracy.

Base Rates

Computation of positive and negative predictive values addresses practical questions: Given a positive test outcome, what is the probability of disease? Moreover, given a negative test outcome, what is the probability of not having a disease? Positive and negative predictive values are computed as the ratio of accurate positive findings to total positive outcomes and accurate negative findings to total negative outcomes, respectively. Predictive values are judged relative to estimated base rates of illness within the referral setting. For example, if the base rate of dementia among older adults referred for psychiatric assessment is

50%, then any test that purports to be sensitive to dementia must have greater than 50% predictive value to warrant its application.

Another aspect of the base rate is the frequency with which abnormal neuropsychological findings are obtained in normal individuals. This frequency is a significant consideration in neuropsychology, in which large batteries of multiple tests are usually administered. It is actually more typical than not for most “normal” persons to perform in the impaired range on some measures in an extensive battery of tests. For example, the 2004 manual of normative standards for the Halstead Reitan Battery points out how common it is for non–brain-damaged individuals without psychiatric problems to perform in the “impaired” range. Only 13.2% of healthy individuals completed a battery of 25 neuropsychological tests without any scores in the impaired range. In contrast, the same healthy group had a median value of three scores in the abnormal range, and more than 25% of the “normal” controls performed in the impaired range on six or more scores in the 25-score battery. Given these data, it would be very naïve to conclude that an individual is impaired by simply counting the number of neuropsychological test scores in the “impaired range.”

Instead, these data underscore how important it is to base the neuropsychological conclusions on a *pattern of performance* that is interpreted in the *context of the patient’s history.* The most substantial evidence of brain injury should typically be based on a pattern of impairment in domains consistent with empirical evidence for that condition, instead of impairment on isolated tasks scattered across different cognitive domains.

Selecting Appropriate Normative Data

Considerable research in recent years has been directed toward expanding the normative database for clinical neuropsychological tests concerning critical demographic factors. Normative data are now available for many cognitive measures across a broad *age* range. On average, persons in their early 70s score about a standard deviation lower than those in their 20s, especially on tasks that combine cognitive processing demands with speed requirements across a wide range of cognitive skills. So it is essential to compare a given older individual’s performance on neuropsychological tasks with that of age peers to ensure that any deficits demonstrated are not due to normal age-related changes. *Education* is also known to affect many aspects of cognitive performance strongly.

Gender differences are relatively small on most clinical neuropsychological tests, but there are a few types of tasks, such as motor speed and strength measures, where gender differences can be found.

Cultural factors and language background pose particular challenges in neuropsychological assessment concerning normative data selection. These factors are frequently addressed through clinical judgment and consultation with experts, but current best practices would encourage a neuropsychologist to seek the best possible cultural and language match between examiner and examinee. Although normative standards based on ethnicity have been developed for some groups, it is unlikely that appropriate normative standards will ever be developed for every specific ethnic and demographic group. Even then, ethnicity- and race-based norms may not be the appropriate basis for correction when the true causes of nonneurologic low test scores lie elsewhere, such as acculturation, literacy, native language, and issues related to economic status and geographical location. Much more research needs to be done to identify and quantify the full range of social determinants of test performance.

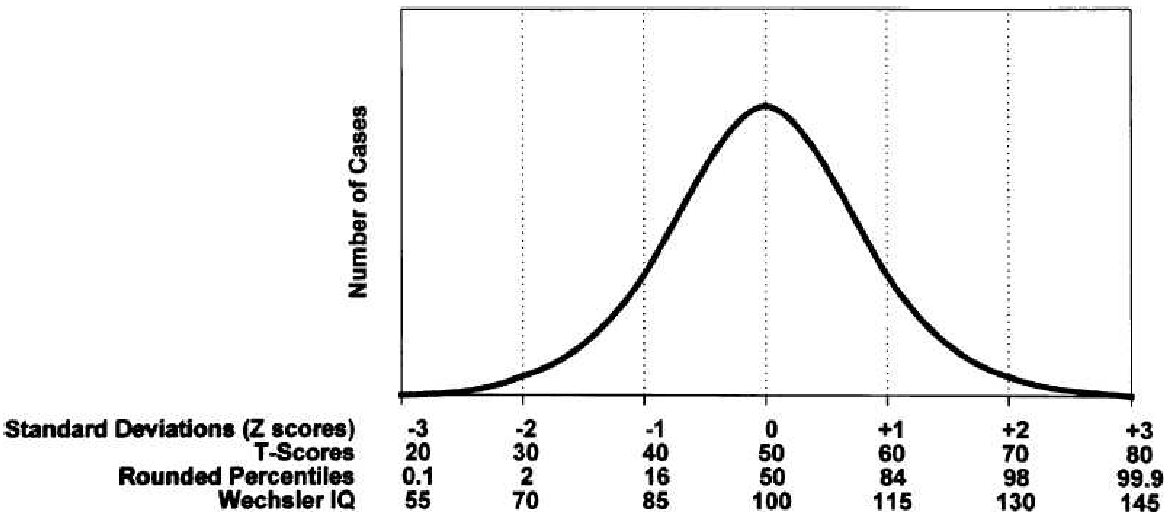

FIGURE 7.4–4. Normal Curve and Standardized Scores.

Estimates of Premorbid Functioning

The degree to which an individual's intellectual and cognitive performance is thought to be impaired is always based on the degree to which the performance differs from that individual's baseline or longstanding level of functioning, usually referred to as *premorbid functioning.* Because prior standardized baseline testing of cognitive abilities is not routinely available for most individuals, *premorbid functioning* is usually estimated. This functioning is based on educational and occupational history and performance on tests that are thought to be minimally affected by brain damage, such as tests of *crystallized intelligence* like vocabulary, general knowledge, or word-reading abilities. Impairment can then be defined as any statistically significant deviation from that individual's estimated level of premorbid ability. As illustrated in the normal *bell curve* that is shown in Figure 7.4–4, statistically significant results on any given set of tests would fall at least one standard deviation above or below the *mean* (or average) performance that is established for that individual, either in comparison to established normative standards or concerning his or her performance across the other tests.

If a 32-year-old Anglo man from the Midwest has only 4 years of education and has worked as a laborer, his premorbid level of functioning might be estimated to lie in the borderline to low average range. This represents the population that performs at least one standard deviation below average, and therefore most test scores for that individual would be expected to hover around that level. In the absence of pathognomonic signs of impairment, such as unilateral neglect or aphasia, it is typically more challenging to identify cognitive signs of cerebral dysfunction in such an individual unless a pattern of test scores falling at least two standard deviations below average emerges that is consistent with the medical history. This example demonstrates the importance of considering the broad context of demographic and cultural factors in the interpretation of neuropsychological evaluation.

Consideration of Diversity and Other Moderating Factors

As discussed above, performance on neuropsychological tests can be influenced by cultural, transient, or situational factors. In addition to gender, race, and ethnicity, notable aspects of individual culture and diversity that have been highlighted in the neuropsychological literature include education and literacy, degree of acculturation, primary language, socioeconomic status, communication style (e.g., direct vs. indirect), familiarity with the testing environment, the cultural conceptualization of intelligence, and immigration status. Practice guidelines from the American Academy of Clinical Neuropsychology call for neuropsychologists to be educated explicitly about cultural issues and have experience in administering and interpreting procedures that are relevant for the patient's cultural background, and refer to or consult with a more qualified examiner when needed. Ideally, patients should be referred to a neuropsychologist who shares a relevant cultural and linguistic background with the examinee and uses appropriate norms to reduce biased interpretation. However, even these steps might not account for more subtle sources of cultural bias. For example, an examinee, whose conceptualization of intelligence emphasizes careful attention to detail over speed, may interpret instructions for a measure of speed and attention very differently than an examinee who views intelligence as speed focused.

When diverse cultural factors appear to outweigh the assumptions on which normative comparisons are based, the neuropsychologist is expected to point out these limitations and acknowledge the conclusions' tentative nature. In such cases, direct observation techniques and intraindividual comparisons, combined with serial testing over time, may need to be prioritized over normative comparison. Careful consideration of cultural factors is also crucial in providing therapeutic and effective feedback. For example, individuals from specific cultural backgrounds may expect feedback that emphasizes the relationship between examiner and examinee, while others may desire greater focus on direct delivery of the findings from the evaluation and their real-life implications.

Transient factors, such as *alcohol and drug use,* can impair attention, memory, and cognitive processing speed, as can many *prescription medications,* especially in the older adults and brain injured. *Chronic pain* and the medications used to treat pain are also essential to consider. *Psychiatric disorders* such as mania, severe depression, or psychosis can undermine cognitive performance, rendering formal testing invalid in severe conditions or reducing scores below expected levels in milder cases. In short, all of these moderating factors combined with clinical observation and input from the patient and collaterals must be considered in interpreting the neuropsychological test results. An experienced neuropsychologist should also be expected to explicitly identify relevant factors that could be expected to undermine or contribute to lower than expected performance, such as impaired vision or hearing loss, acute illness, or recent significant stressors.

GENERAL REFERRAL ISSUES

Referring providers turn to neuropsychology for many reasons that include differential diagnosis, baseline measurement, treatment planning, and opinions regarding causality and decisional capacity. Because many referring providers have limited experience and knowledge of neuropsychology, the neuropsychologist needs to take an active role in refining the specific questions and providing practical information about the limitations of the consultation. An early conversation with a potential referring provider about such issues will foster realistic expectations and ensure that the neuropsychological evaluation is helpful in the patient's care. In general, it is not considered appropriate for a referring provider to request specific tests or procedures because the specific methods of neuropsychological assessment will necessarily reflect the patient's unique presentation of symptoms and complaints, history and development, the perspective of the neuropsychologist, and the referral question.

Level of Functioning

A common referral issue involves documentation of the level of functioning for various purposes, including assessment of functional change, vocational planning, capacity to live independently, manage personal health care, or make decisions, especially in the presence of diagnoses such as dementia, stroke, or head injury.

Differential Diagnosis

Like any other diagnostic procedure, the results of a neuropsychological examination must be interpreted in light of all available information, including the history and any associated medical factors that are documented or reported for the individual. Many neurologic and psychiatric disorders have similar clusters of symptoms in common, with complaints of concentration or memory problems being among the most frequently reported problems. For example, impaired attention and concentration are commonly found among patients who present with a history of mild closed head injury, psychiatric issues (e.g., depression, post-traumatic stress disorder [PTSD], or anxiety), or any of several types of neurologic disorders associated with subcortical involvement (e.g., Parkinson disease, multiple sclerosis, or human immunodeficiency virus [HIV]-related cognitive decline). A pattern of cognitive impairment across several critical areas of functioning does not necessarily provide a sufficient basis for a specific diagnostic conclusion. However, it must be integrated with the medical and relevant psychosocial factors to arrive at a specific differential diagnosis. The following discussion reflects some common clinical syndromes in which psychiatrists and neuropsychologists share an interest and may work together in assessing and treating patients.

Age- or Stress-Related Cognitive Change. Many middle-aged and older adults have concerns about everyday concentration and memory failures, and with heightened public awareness about conditions such as Alzheimer disease, an increasing number of these individuals seek evaluations for these concerns. A neuropsychological evaluation provides a detailed, objective picture of different aspects of memory and attention, which can be helpful in reassuring healthy persons about the integrity of their abilities. In addition, testing can reveal previously undetected mood or anxiety disorders that may contribute to cognitive concerns and provide a basis for suggestions about mnemonic strategies that can improve functioning in everyday life.

Mild Traumatic Brain Injury. TBI can be mild, moderate, or severe. These severity ratings are most commonly characterized by the duration of loss of consciousness or posttraumatic amnesia, as shown in the DSM-5 severity ratings in Table 7.4–5. In addition, most neuropsychological studies of the cognitive consequences associated with TBI have consistently found a strong relationship between the duration of loss of consciousness and cognitive recovery following TBI.

The vast majority of TBI cases referred for neuropsychological consultation involve mild TBI, from which an excellent cognitive recovery can be expected within days to weeks after injury. However, a significant proportion of persons who have suffered a mild TBI report subjective complaints that persist for many months after the injury, such as poor concentration and mood changes, fatigue, photosensitivity, balance problems, tinnitus, headache, or other forms of pain. This set of persistent subjective complaints is sometimes diagnosed as a postconcussional syndrome in the 10th edition of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) or mild neurocognitive disorder in ICD-11. Several methodologically sound studies indicate that premorbid histories of psychological symptoms, including trauma and substance abuse, may be factors to consider. As noted in DSM-5, since co-occurring symptoms, such as depression, irritability, fatigue, headache, photosensitivity, sleep disturbance tend to resolve in the weeks following mild TBI, substantial subsequent deterioration in these areas should trigger consideration of additional diagnoses.

In many cases of mild TBI, there is no clear evidence of brain injury on MRI or other neuroradiologic tests. Therefore, neuropsychological testing often plays a crucial role in determining the extent of objective cognitive deficit and examining the possible role of psychological factors in perpetuating cognitive problems.

Table 7.4–5.
Traumatic Brain Injury Severity Criteria

Criteria	Mild TBI	Moderate TBI	Severe TBI
Structural imaging	Normal	Normal or abnormal	Normal or abnormal
Loss of consciousness (LOC)	0–30 minutes	>30 min and <24 hours	>24 hours
Alteration of consciousness/mental state (AOC)	a moment	up to 24 hours	>24 hours
Post-traumatic amnesia (PTA)	0–1 day	>1 and <7 days	>7 days
Glasgow Coma Scale (best available score in first 24 hours)	13–15	9–12	<9

Understanding the long-term consequences of TBI has improved significantly over the past 20 years. The work of Sureyya Dikmen and her colleagues has been particularly significant in demonstrating a *dose–response* relationship between the severity of the initial injury and the long-term cognitive consequences. For example, when loss of consciousness is behaviorally defined as “time to follow commands” on the Glasgow Coma Scale, those with relatively brief loss of consciousness are indistinguishable from healthy control subjects on cognitive functioning tests after about a year. The preponderance of evidence at this point indicates that persons who meet the criteria for mild TBI can be expected to recover fully in terms of cognitive functioning. However, when people with a history of mild TBI complain of enduring symptoms that exceed expectation, they should be carefully evaluated in order to identify any contributing psychological factors that could offer potential treatment strategies, such as symptoms of depression, anxiety, PTSD, substance abuse, or other psychiatric disorder. The issue of repeated mild TBIs causing a neurobehavioral disorder (chronic traumatic encephalopathy or CTE) has been proposed but not universally accepted.

The neuropsychologist should bear in mind that many patients with mild TBI are involved in litigation, complicating the neuropsychologist's ability to identify the causes for impairment. Although outright malingering is thought to be relatively infrequent and difficult to verify, the possibility of contributing psychological factors due to somatic symptom disorder, depression, PTSD, and substance use should be a significant consideration when potential legal settlements or disability benefits are in question. Several psychological models, such as "symptom misattribution" or "diagnosis threat," might explain some tendency to exaggerate symptoms based on the individual's expectations about how an individual with TBI might be expected to perform. The possibility of contributing psychological factors is significant to consider in the case of mild head injury, when subjective complaints may be disproportionate to the objectively reported circumstances of the injury.

Poststroke Syndromes. After the acute phase of recovery from stroke, patients may have residual deficits affecting cognitive skills or mood. A neuropsychological evaluation can help identify areas of strength, which can be used in planning additional rehabilitation and provide feedback on the functional implications of residual deficits for work or instrumental activities of daily living. Assessment of functional skills can also be helpful to a psychiatrist who is managing mood and behavioral symptoms or dealing with family caregivers.

Detecting Early Dementia. Conditions that warrant neuropsychological assessment for early detection and potential treatment include HIV-related cognitive deficits and normal pressure hydrocephalus. When concerns about a person's memory functioning are expressed by relatives instead of the patient, there is a higher probability of a neurologic basis for the functional problems since the loss of insight is a crucial feature of most neurodegenerative syndromes. Neuropsychological evaluation, combined with a good clinical history and other medical screening tests, can be highly effective in distinguishing early dementia from the mild changes in memory and executive functioning that can be seen with normal aging. A neuropsychological evaluation is beneficial in documenting the rate and pattern of cognitive deterioration over time and differentiating among different forms of dementia. An additional incentive for early diagnosis of dementia is that patients with early dementia may be candidates for memory-enhancing therapies (e.g., acetylcholinesterase inhibitors), and testing can provide an objective means of monitoring treatment efficacy.

Neuropsychological profiles have been used to identify patients who do not meet full criteria for Alzheimer disease but who report or demonstrate memory deficits with normal performance in other cognitive domains and with no or minimal functional complaints. This pattern, termed amnestic *mild cognitive impairment* (MCI), may represent prodromal Alzheimer disease or a transitional state in that such patients show a higher incidence of conversion to Alzheimer disease than individuals with no such cognitive impairment. MCI conversion rates to Alzheimer disease have been reported as 15% annually relative to conversion rates in those without cognitive deficits of about 2% annually. However, these conversion rates vary from 10% to 70% depending on the initial MCI diagnostic criteria used, which emphasizes the difficulties of operationally defining cognitive and functional decline. It is also important to note that the definition of MCI has been revised since its introduction to include nonmemory deficits, especially executive dysfunction, with the introduction of nonamnestic MCI and the potential of conversion to other types of dementia. However, the clinical implications of conversion to dementia are complex. For example, the diagnosis of Alzheimer disease should be addressed cautiously due to the possible negative emotional repercussions of such information, to the fact that not all MCI patients convert to Alzheimer disease, and that current medications produce only mild improvement even in patients diagnosed with Alzheimer disease. The MCI diagnosis should stimulate some consideration for providing affected patients and their families with information for future planning, such as financial management oversight and advanced directives. There is also some evidence that cognitive retraining may improve cognitive performance and subjective cognitive and emotional complaints. In the future, the combination of genetic information, neuroimaging and other biomarkers, and the behavioral evidence of neuropsychological assessment are likely to improve the ability to diagnose dementia at an earlier stage, which will be important as new treatments are developed.

Distinguishing Dementia and Depression. A substantial minority of patients with severe depression exhibit severe generalized impairment of cognitive functioning. In addition to problems with attention and slowing of thought and action, there may be significant forgetfulness and problems with reasoning. A neuropsychological evaluation can help identify a dementia syndrome associated with depression, historically known as pseudodementia, by examining the pattern of cognitive impairment. Mixed presentations are also common, in which symptoms of depression coexist with various forms of cognitive decline and exacerbate the effects of cognitive dysfunction beyond what would be expected from the neurologic impairment alone. Neuropsychological testing, in this case, can be beneficial by providing a baseline for measuring the effect of antidepressants or other therapy in alleviating cognitive and mood symptoms.

Presurgical Assessment

Neuropsychological evaluations before or after neurosurgery are used to assist surgical planning, assess prognosis, and evaluate outcomes. Neurosurgery for intractable seizures in patients with epilepsy usually involves pre- and postsurgical neuropsychological evaluations, in combination with the intracarotid sodium amobarbital procedure (ISAP), also referred to as the Wada test, or functional neuroimaging (also commonly conducted by neuropsychologists.). The ISAP has been considered the “gold standard” for language lateralization, and while testing utilizing fMRI is typically preferred as a less invasive alternative, the ISAP remains prioritized for examinees where atypical language representation is suggested, or fMRI is inconclusive. Tissue resection is planned to maximize seizure control and minimize cognitive morbidity. Medication-resistant Parkinson disease can be treated with deep brain stimulation (DBS), and neuropsychological assessment is conducted primarily to assess for risk of poor cognitive outcomes. Many studies have documented that cognitive impairment prior to DBS placement increases the risk for worse cognitive outcomes. Some brain tumor resection programs and epilepsy surgery programs employ neuropsychologists to conduct assessments during awake surgery to assist in protecting eloquent cortex. Finally, neuropsychologists also commonly perform evaluations prior to organ transplant to identify risks for poor outcomes, including cognitive impairment that may reduce adherence to medical treatments and health maintenance.

Change in Functioning Over Time

Because many neurologic diagnoses carry clear expectations regarding normal recovery rates and decline over time, it is frequently important to reexamine a given patient with a follow-up neuropsychological assessment to document any change after 6 months to a year. For example, neuropsychological evaluation could document the decline in independent functioning associated with a progressive dementia or identify improvement following a stroke or tumor resection. Follow-up examinations also provide an opportunity to objectively examine complaints of longstanding or worsening cognitive sequelae following mild head trauma, even though the current literature indicates that the most significant proportion of recovery of function is likely to occur over the initial 6 months to a year postinjury. Although continuing subtle signs of recovery can continue after that period, failure to improve following the injury—or to worsen of complaints—would suggest the possibility of contributing psychological factors or the existence of a preexisting or coexisting condition, such as substance abuse, dementia, or outright malingering.

Assessment of Decision-Making Capacity

Neuropsychologists are often asked to assist in determining an individual’s capacity to make decisions or manage personal affairs. A neuropsychological evaluation can be helpful in these cases by documenting areas of significant cognitive impairment and identifying areas of strength and well-preserved skills. However, legal opinions about decision-making capacity are seldom based on test scores alone but usually rely heavily on clinical interviews, collateral interviews with family or caregivers, and direct observations (e.g., in-home assessment) of everyday functioning. Appraisal of an individual’s level of insight and capacity to appreciate his or her own limitations is typically the most critical basis for an opinion about decisional capacity. Assessment of insight is based on the consistency between an individual's self-appraisal of their abilities and functioning, compared to objective evidence and observations of how the person functions independently and further informed by objective testing results. If there is a high agreement between the person's observed level of functioning and the person's self-appraisal of their own functioning, that person's level of insight is probably more intact than the person who rigidly insists that they are doing fine and defensively rejects assistance, despite objective evidence of self-neglect in the home, mounting financial calamity in the form of unpaid bills and a dangerous driving record. State statutes generally define standards for decision-making capacity, and, of course, the ultimate determination of competence rests with the presiding judge. However, the neuropsychologist or other health care professional can play a significant role in shaping the judge's ruling by providing evidence to the court in the form of a professional opinion supported by compelling behavioral data that has strong face validity.

As a general rule of thumb, consideration of decision-making capacity must be addressed in the narrowest possible sense to infringe as little as possible on the individual's freedom to represent their interests. The most common issues related to decisional capacity involve the areas of (1) financial and legal matters, (2) health care and medical treatment, and (3) ability to live independently. In addition, some capacity issues involve higher standards, such as the ability to drive and the ability to work or practice in a given profession (e.g., air traffic controller, pilot, surgeon, or financial advisor). In such cases, the neuropsychologist needs to rely on appropriate normative expectations for the type of activity and the patient's demographics. An excellent resource, entitled "Assessment of Older Adults With Diminished Capacity," is jointly published by the American Bar Association and the American Psychological Association, with three versions available online written for use by attorneys, judges, and health care professionals.

Forensic Evaluation

Neuropsychological evaluation of individuals in criminal or civil law matters requires specialized knowledge beyond general expertise in neuropsychology. Neuropsychologists are frequently called upon as experts in matters involving head injury in particular, especially in the case of mild head injury associated with a motor vehicle accident. As a distinct subspecialty, this area of practice requires integration of knowledge of statutes, laws, precedents, and legal procedures and expertise in identifying and describing the impact of an injury or event on cognitive, emotional, and behavioral functioning.

Ethical Considerations

Several ethical considerations are especially unique to the practice of neuropsychology. They include test and test data security, appropriate test measures, and confidentiality. Exposure of test materials to an examinee prior to the examination can render the results invalid and subsequently lead to invalid conclusions and possible harmful outcomes. As such, neuropsychologists must safeguard their testing materials. Since the raw test data (i.e., examinee responses) often contain copyrighted test materials, either in the examinee responses or in the test forms on which the responses are recorded, there are limited instances in which raw data should be released, usually only to another neuropsychologist.

In selecting tests, neuropsychologists must use tests that have sufficient evidence to support the test's intended use. Relatedly, neuropsychologists must use appropriate norms, and if significant differences exist between the examinee and the normative sample to which they are compared, neuropsychologists must account for these differences in their interpretation of the findings. With the increased utilization of assessment via videoconference and telephone, neuropsychologists must prioritize tests for which evidence exists for their utility in such a setting.

Neuropsychologists must explicitly discuss the limits of confidentiality with each examinee. Unlike psychotherapy, the results of neuropsychological evaluations are often shared with other professionals, so examinees should be informed of likely recipients of the final report and the degree of disclosure of sensitive information that may be necessary.

APPROACHES TO NEUROPSYCHOLOGICAL ASSESSMENT

The neuropsychological examination systematically assesses a broad range of cognitive functions and emotional status. Because deficits in cognitive performance can only be interpreted in comparison to a person's longstanding or *premorbid level of functioning,* overall intellectual abilities are typically examined to measure the current level of overall functioning and identify any changes in intellectual functioning. Psychological contributions to performance are also considered concerning personality and coping style, emotional lability, presence of thought disorder, developmental history, and significant past or current stressors. The expertise of the neuropsychologist lies in integrating findings obtained from many diverse sources, including the history, clinical presentation, and several dozen discrete performance scores that make up the neuropsychological data.

The specific neuropsychological assessment techniques used in a particular examination typically reflect the training and preferences of the neuropsychologist, as well as the referral issue. Historically, the actual practice and procedures of clinical neuropsychology developed in two general directions: the battery approach and the hypothesis testing approach. These are discussed next, and the reader can find more detailed descriptions among our references.

Battery Approach

The battery approach, exemplified by the HRNTB or the *Neuropsychological Assessment Battery* (NAB), grew directly out of the psychometric tradition in psychology. This approach typically includes a large variety of tests that measure most cognitive domains and sensory and motor skills. Traditionally, all parts of the test battery are administered regardless of the patient's presenting problem though the NAB has a screening examination that covers all appropriate domains. The battery approach has the advantage of identifying problems that the patient might not have mentioned and that the medical history may not necessarily predict. However, it has the disadvantage of being very time-consuming (i.e., 6- to 8-hour examination for the HRNTB).

Hypothesis Testing Approach

The qualitative hypothesis testing approach is historically best exemplified by the work of Luria and more recently developed as the *Boston Process Approach* by Edith Kaplan and her colleagues. It is characterized by a detailed evaluation of areas of functioning related to the patient's complaints and predicted areas of impairment, with relatively less emphasis on aspects of functioning that are less likely to be impaired. The hypothesis testing approach has been particularly helpful in illuminating the differential roles of the two hemispheres, as discussed above. This approach has the advantage of efficiently honing in on areas of impairment and producing a detailed description of the quantitative and qualitative deficits from a cognitive processing standpoint. However, it has the shortcoming of potentially overlooking unexpected areas of deficit.

Screening Approaches

Many practitioners have moved away from strict battery or hypothesis testing approaches since the 1990s and developed more flexible and efficient screening approaches. These changes in clinical practice are driven by economic factors that make it less practical to evaluate patients for long periods and the use of more efficient, empirically based clinical practice standards that draw on a widening array of well-normed clinical instruments. In this model, the neuropsychologist utilizes a core set of screening procedures as a first step in determining whether a diagnosis can be made with less information or whether additional testing is necessary to identify more subtle problems. Therefore, a screening protocol that efficiently assesses the significant areas of neuropsychological functioning may or may not be followed by more detailed testing in selected areas that might help to understand better the reasons for the deficits demonstrated on the screening evaluation. The *Repeatable Battery for the Assessment of Neuropsychological Status* (RBANS) and the NAB provide examples of standardized screening approaches that have the advantage of uniform normative reference groups. However, many clinicians routinely use tests drawn from many sources.

Mental Status Examinations

In some cases, usually involving very acute or severe cognitive impairment, it is simply not feasible to administer extensive cognitive examination procedures. Instead, the neuropsychologist might rely on bedside mental status examinations or very brief cognitive screening procedures to address referral issues. However, research has shown that, even with brief screening procedures, the systematic use of a structured examination format can significantly increase the accuracy of detecting cognitive impairment. Therefore, several mental status screening procedures are commonly used by various health care providers in general medical and geriatric settings to provide an initial assessment of mental status. Along with a careful history, these screening procedures can serve as an indicator for more precise neuropsychological evaluation, as well as a quick baseline against which to monitor changes in cognitive functioning.

One of the most widely used screening instruments for documenting gross changes in mental status is the Mini-Mental State Examination (MMSE), although its use has declined since it is no longer in the public domain. However, two newer mental status screening approaches that are publicly available may be even more sensitive to cognitive deficits than the MMSE. Both the Montreal Cognitive Assessment (MoCA) and the St. Louis University Mental Status Examination (SLUMS) include measures of verbal fluency and clock drawing that can reflect executive functioning. The MoCA has the additional advantage of assessing abstraction ability, while the SLUMS includes items that assess story recall, mental calculations, and size differentiation. Like the MMSE, all of these mental status screening tools are likely to underestimate the degree of cognitive deficits in well-educated older persons with early Alzheimer disease or in younger adults with focal brain injury but more likely to overestimate the presence of cognitive deficits in persons with little formal education. Therefore, cut-off scores should be adjusted for age and education before concluding that impairment is present. Although mental status examinations can be beneficial in screening for gross signs of cognitive impairment, they do not provide a sufficient foundation for diagnosing specific etiologies of cognitive impairment, and they are not interchangeable with a more detailed neuropsychological evaluation.

Teleneuropsychology

The COVID-19 pandemic forced a seldom-used approach (i.e., telephonic and videoconference evaluations) into the mainstream. Fortunately, it will result in a new and permanent addition to a widespread practice that will benefit patients in remote locations who may not otherwise have access to neuropsychological services. The practice of teleneuropsychology, of course, limits the kinds of tests that can be administered since the medium of telephone or video telehealth introduces variables that can reduce the reliability or validity of tests that were designed for in-person use (e.g., the patient’s video device may display stimuli differently enough from the original that the test becomes invalid). When the telehealth technology falters or fails, the interruption can invalidate tests, especially when the responses involve timing. In addition, there can be a complete loss of connection, and the evaluation may have to be conducted in person. The test environment cannot be controlled with teleneuropsychology, where the examinee is usually in a location separate from the examiner. This dilemma is particularly problematic for forensic evaluations, where threats to validity are even more closely scrutinized. As of the writing of this chapter, there is limited literature on the reliability and validity of teleneuropsychology, but the rapid increase of its use will undoubtedly result in more empirical evidence to support its use. It likely will result in the development and validation of teleneuropsychology tests.

DOMAINS OF FORMAL NEUROPSYCHOLOGICAL ASSESSMENT

The past decade has seen a virtual explosion in the growth of more sophisticated and better-standardized tests and procedures for neuropsychological evaluation. A comprehensive listing of tests and techniques is beyond the scope of this chapter, but standard works that provide excellent reviews of current techniques are found in texts referenced in this chapter. A list of some examples of common neuropsychological tests and techniques is provided in Table 7.4–6.

Table 7.4–6.
Selected Tests of Neuropsychological Functioning

Area of Function	Comment
Intellectual Functioning	
Wechsler Intelligence Scales	Age-stratified normative references; appropriate for adults up to age 89, adolescents, and young children
Test of premorbid functioning	Estimates predisease intellectual abilities based on a test of word reading and demographic factors. Conormed with the WAIS-IV to yield regression-based IQ and other ability estimates
Attention and Concentration	
Digit span	Auditory-verbal measure of a simple span of attention (*Digits Forward)* and cognitive manipulation of increasingly longer strings of digits (*Digits Backward, Sequencing*)
Digit Vigilance Test	Assesses both speed and accuracy of attention using an array of rows of randomized numbers in which a specific digit must be crossed out
Paced Auditory Serial Addition Test (PASAT)	Requires double tracking to add pairs of digits at increasing rates; particularly sensitive to subtle simultaneous processing deficits, especially in head injury
Memory	
Wechsler Memory Scale-IV	Comprehensive set of subtests measuring attention and encoding, retrieval, and recognition of various types of verbal and visual material with both immediate recall and delayed retention; Excellent age-stratified normative comparisons for adults up to age 90 with intellectual data for direct comparison
California Verbal Learning Test	Documents encoding, recognition, and both immediate and 30-minute recall; affords examination of possible learning strategies as well as susceptibility to semantic interference with alternate and short forms available

Fuld Object Memory Evaluation	Selective reminding format requires the patient to identify objects tactually, then assesses the consistency of retrieval and storage as well as ability to benefit from cues; Normative reference group is designed for use with older individuals
Hopkins Verbal Learning Test-Revised	A brief list learning paradigm in which a 12-item word list is recalled across 3 learning trials. Delayed recall after 15 minutes followed by a recognition discrimination trial help isolate memory problems related to encoding, retention, and retrieval
Brief Visuospatial Memory Test-Revised	Serial learning approach used to assess recall and recognition memory for an array of 6 geometric figures; 6 alternate forms
Language	
Boston Diagnostic Aphasia Examination	Comprehensive assessment of expressive and receptive language functions
Boston Naming Test-Revised	Documents word-finding difficulty in a visual confrontation format
Verbal fluency	Measures ability to fluently generate words within semantic categories (e.g., animals) or phonetic categories (e.g., words beginning with “S”)
Token test	Systematically assesses comprehension of complex commands using standard token stimuli that vary in size, shape, and color
Visuospatial–Constructional	
Judgment of line orientation	Ability to judge angles of lines on a page presented in a match-to-sample format
Facial recognition	Assesses matching and discrimination of unfamiliar faces
Clock drawing	Useful screening technique is sensitive to organization and planning as well as constructional ability

Rey–Osterreith Complex Figure Test	Ability to draw and later recall a complex geometric configuration; sensitive visual memory as well as executive deficits in the development of strategies and planning
Motor	
Finger tapping	Standard measure of simple motor speed; particularly useful for documenting lateralized motor impairment
Grooved pegboard	Ability to rapidly place notched pegs in slotted holes; measures fine finger dexterity as well as eye–hand coordination
Grip strength	Standard measure of lateralizing differences in strength
Executive Functions	
Wisconsin Card Sorting Test	Measure of problem-solving efficiency is particularly sensitive to executive deficits of perseveration and impaired ability to generate alternative strategies in response to feedback flexibly
Category Test	This measure of problem-solving ability also examines the ability to benefit from feedback while flexibly generating alternative response strategies; regarded as one of the most sensitive measures of general brain dysfunction in the Halstead–Reitan battery
Trail-Making Test	Requires rapid and efficient integration of attention, visual scanning, and cognitive sequencing
Delis–Kaplan Executive Function System (D-KEFS)	Battery of measures that are sensitive to executive functions
Comprehensive and Screening Batteries	
Neuropsychological Assessment Battery	The NAB is a comprehensive and conormed assessment that offers the combined strengths of both a flexible and fixed battery to allow

	clinicians to focus on specific areas of concern, with 5 domains and 33 tests
Halstead–Reitan Neuropsychological Test Battery (HRNTB)	A compilation of neuropsychological tests designed to evaluate functioning, with versions for children and adults. Demographically corrected norms for an expanded version are available for adults
Repeatable Battery for the assessment of Neuropsychological Status	Brief battery with five domains and two versions designed for repeat testing. Norms for ages 12 to 89, though especially useful in the older adults
Psychological Factors	
Beck Depression Inventory	Brief (5–10 minutes) self-report measure that is sensitive to symptoms of depression. Best for screening depression in adults up to late middle-age, who can be expected to report symptoms frankly. Available in standard (21 four-choice items) or short (13-item) form
Geriatric Depression Scale	30-item self-report screen for symptoms of depression. The yes–no format is less cognitively demanding than other scales
Minnesota Multiphasic Personality Inventory (MMPI)-2-RF and MMPI-3	This self-report instrument is highly useful for documenting a wide range of psychological symptoms and statistically determined validity indicators and information about an individual’s response style. The MMPI-2-RF was a psychometric revision of the previous instrument, while the MMPI-3 was recently published with new norms and new and updated items
Personality Assessment Inventory (PAI)	Self-report instrument that is useful for psychological treatment planning, implementation, and evaluation has empirical support for use in forensic settings. In addition,

	fourth grade reading-level questions make it somewhat easier to complete than the MMPI
Performance Validity	
Test of Memory Malingering (TOMM)	Effectively identifies persons who tend to exaggerate impairment using a forced-choice recognition procedure based on 50 line drawings of common objects
Green Word Memory Test	Identification of poor effort on memory testing, indices that distinguish feigned vs. genuine memory impairment

Interview

The clinical interview provides the single best opportunity to identify the patient’s concerns and questions, elicit a direct description of current complaints from the patient, and understand the context of the patient’s history and current circumstances. Although the patient typically serves as the primary information source, it is vital to seek corroborating information for the patient’s account from an interview with caregivers or family members and thoroughly review relevant records, such as medical and mental health treatment, educational, and employment experiences.

Intellectual Functioning

Assessment of intellectual functioning serves as the cornerstone of the neuropsychological examination. The Wechsler Intelligence Scales have represented the traditional “gold standard” in intellectual assessment for many years, based on carefully developed normative standards. The scope and variety of subtests on which the summary intelligence quotient (IQ) values are based also provide valuable benchmarks against which to compare performance on other tests of specific abilities. The latest revision of this instrument, the *Wechsler Adult Intelligence Scale IV* (WAIS-IV), released in 2008, offers the additional advantage of greatly extended age norms (ages 16 to 89) directly related to normative WMS-IV performances. The Wechsler Intelligence Scales utilize a broad set of complex verbal and visuospatial tasks that can be summarized as Full Scale IQ, with composite scores in the domains of Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed. In the context of a neuropsychological examination, the patient’s performance across the procedures provides a reasonable basis for estimating an individual’s level of premorbid abilities and the current level of intellectual functioning. Most neuropsychologists recognize that the summary Full Scale IQ provides only a ballpark range for characterizing an individual’s general level of functioning. It is usually more appropriate and meaningful to characterize an individual’s level and consistency of functioning (e.g., borderline, low average, average, high average, or superior) across the various domains represented by the composite scores rather than focusing on a specific IQ value.

Careful examination of the individual’s performance across the various subtests within each composite area can provide information regarding the patient’s pattern of strengths and weaknesses and the degree to which these performance characteristics are consistent with the history and performance on other aspects of the neuropsychological examination. For example, tests that measure longstanding, accumulated knowledge, such as vocabulary or fund of general information, provide a basis for estimating an individual’s longstanding (or premorbid) level of intellectual abilities, which in turn can help to gauge the degree to which an individual may have deteriorated.

Attention and Information Processing

Attention underlies performance in virtually all other areas of functioning and should always be considered a potential contributor to impairment on any tests requiring sustained concentration and vigilance or rapid integration of new information. In a broad-based neuropsychological examination, measures of attention and concentration provide a helpful basis for “previewing” the individual’s ability to comprehend, process information, and otherwise engage in the assessment process. *For example, digit span* requires patients to repeat increasingly longer strings of digits as a way of assessing the ability to process relatively simple information. In contrast, *digit span backward* reflects more complex simultaneous processing and cognitive manipulation demands or working memory. More demanding attention and information processing measures have also been developed that are highly sensitive to subtle deficits in rapid simultaneous processing. For example, the *Paced Auditory Serial Addition Test* (PASAT) requires patients to successively sum pairs of numbers at increasingly faster rates of presentation, adding each new digit to the immediately preceding digit in a highly demanding test of working memory. Other procedures for assessing attention and sustained concentration can be found in pencil and paper measures of digit, letter, or symbol cancellation and various versions of computer-assisted measures of attention and vigilance.

Memory

Complaints of memory problems constitute one of the most common reasons for referral to neuropsychology. As described above, the neuropsychologist utilizes an information processing approach to assess memory problems that might involve difficulty with encoding, retrieval, or storage of new information. The (*WMS-IV* is the latest revision of a widely used battery of subtests that utilizes several measures of attention, memory, and new learning ability. The original version, which David Wechsler developed in 1945, was extremely limited and did not include delayed recall. A revised version of the battery was published in 1987, which added delayed memory and included five summary indices: Verbal and Visual Memory, General Memory, Attention/Concentration, and Delayed Memory. The current version, published in 2009 (WMS-IV), has a modified index structure that assesses auditory memory using immediate and delayed recall for a story narrative (Logical Memory I and II) and a set of word pairs (Verbal Paired Associates I and II). Visual Memory is assessed using immediate and delayed memory for the visual detail and spatial location of a set of Designs and Visual Reproduction of designs to make sure that deficits are not due to visuospatial deficits independent of memory. Other Index domains of the WMS-IV include visual working memory, immediate memory, and delayed memory. The WMS-IV offers additional advantages by extending the normative comparisons to age 90 and adding a set of mental status screening procedures (Brief Cognitive Status Exam) for older adults.

A well-standardized example of a verbal list learning technique is the *California Verbal Learning Test 3* (CVLT-3), published in 2017. This procedure employs 5 learning trials of a 16-item word list, which comprises 4 exemplars from each of 4 semantic categories. Aside from measuring several aspects of learning and memory, the CVLT-3 also yields rich qualitative data about the patient’s use of higher-order conceptual strategies and the efficiency of verbal learning. Proactive interference is examined using a single free recall trial of a second list of words. Free and cued recall of the original list after both short and long delays, as well as recognition and forced-choice trials, further examine memory processes of storage, retention, and retrieval. Published norms for this instrument extend from ages 16 to 89 for males and females with education, region, ethnicity, and age norms. This revised task also offers the advantages of an alternative form and a nine-item short form. Because administration time for the standard form typically requires 15 to 25 minutes (not counting a 20-minute delay interval), the short form can be beneficial for examining more severely impaired individuals or individuals with limited formal education. The Rey Auditory Verbal Learning Test (RAVLT) is similar to the CVLT except that the words are not semantically related and the normative standards are not as extensive.

The *Hopkins Verbal Learning Test-Revised* (HVLT-R) is another helpful list-learning task for older or more impaired individuals in which a 12-item word list is presented in 3 learning trials, with follow-up measures of delayed recall after a 15-minute delay interval and a recognition discrimination trial. Its other significant advantage over the CVLT and RAVLT is that there are multiple parallel forms.

Several procedures have been standardized over the years for assessing memory for visual material. These include the *WMS-IV Visual Reproduction*, *Rey–Osterreith Complex Figure Test,* and further discussed as a visual-spatial assessment technique below. In addition, the *Brief Visuospatial Memory Test-Revised* offers six alternate forms and uses a three-trial serial learning approach to assess recall and recognition memory for an array of six geometric figures presented in a 2 × 3 array on a single page.

Language

Assessment of language examines both expressive abilities and comprehension. However, most neuropsychologists screen for language impairment rather than administer an extensive formal language assessment battery, such as the *Boston Diagnostic Aphasia Examination* (BDAE) or the *Western Aphasia Battery* (WAB). Measures of verbal fluency commonly assess expressive language. Specific procedures and normative data for several versions of verbal fluency measures can be found in the *NAB* and the Heaton norms manual. Visual confrontation naming is another commonly assessed aspect of language. The *Boston Naming Test-Revised,* published in 2001, provides a standard 60-item measure, and the NAB offers a shorter alternative measure of visual confrontation naming ability. Standardized comprehension evaluation often employs procedures that systematically assess one-, two-, and three-step commands in a format, such as the *Token Test.* At the same time, the Complex Ideational Material subtest of the *Boston Diagnostic Aphasia Examination* examines an individual's ability to comprehend more complex narrative material through responses to a series of yes–no questions.

Visuospatial Functions

Complex visuospatial abilities can be assessed through an extensive array of procedures found in various batteries with well-developed normative data and stand-alone tests that vary in the quality of the norms. The perceptual reasoning measures of the WAIS-IV include block design, visual puzzles, and matrix reasoning, all of which can help provide an understanding of an individual's ability to perform on tests of visual constructional ability and abstract concept formation. The NAB offers several standardized assessment procedures. It has a spatial module that includes visual discrimination based on a match-to-target paradigm. A design construction task also uses a set of seven geometrically shaped pieces to copy a two-dimensional design and a figure drawing task with immediate recall of a geometric figure. The battery also includes a map reading task that measures visual–spatial skill, spatial/directional orientation, right–left orientation, and visual scanning. Arthur Benton's laboratory developed a set of widely used stand-alone measures, such as *Facial Recognition, Visual Form Discrimination,* and *Judgment of Line Orientation.* Other measures of visual constructional ability examine the person's ability to draw spatial designs or assemble two- or three-dimensional figures (see examples in Fig. 7.4–2). In addition to the significant visuospatial component, these tasks reflect executive planning and organizational abilities contributions. More impaired individuals can be asked to copy simple geometric forms, such as a Greek cross or intersecting pentagons, to examine visuospatial abilities less influenced by planning and organization.

The widely used technique of *clock drawing* provides a surprisingly sensitive measure of planning and organization, especially for older individuals at risk for dementia. While problems involving poor organization, perseveration, and possible neglect may be obvious, more subtle difficulties can also be detected, especially when a patient's performance is evaluated based on premorbid expectations. Like most neuropsychological assessment techniques, it is essential to use even this relatively simple procedure in a standard manner. Most neuropsychologists use the same set of instructions each time, offering the patient a single sheet of unlined paper and saying, "Now I would like you to draw a clock. Put all the numbers on the clock, and set the time to 10 after 11." It is best to encourage the patient to use the whole page and draw a relatively large clock so that subtle evidence of neglect or executive deficits in organization and planning are more pronounced. Although the rich qualitative information that can be gleaned from this task may not be sufficient to stand alone for diagnostic purposes, this measure provides a valuable screening technique—especially for older or more seriously impaired individuals—and can serve as a basis for deciding whether to refer the patient for more detailed neuropsychological examination.

The *Rey–Osterreith Complex Figure Test* is a more demanding constructional task that requires the individual to copy a highly complex geometric configuration and then draw the figure from memory. Important qualitative information about the examinee's problem-solving approach can be extracted by carefully monitoring and documenting the examinee's strategy to draw the figure. It should be noted that apparent visual–spatial constructional difficulties can also be observed in the presence of frontal executive types of problems due to impaired organization and planning or poor self-monitoring.

Sensory and Motor Functions

Double simultaneous stimulation in the visual, tactile, and auditory modalities is a standard component of the HRNTB. It can help assess the integrity of essential sensory functions and neglect if deficits are present on one side only on the simultaneous bilateral trials and not when stimulation is unilateral. *For example, grip strength* and rapid *finger tapping* are commonly used measures of motor strength and speed-sensitive to lateralized brain dysfunction. A related measure, the *grooved pegboard,* assesses the individual's ability to integrate fine finger dexterity, motor speed, and eye–hand coordination efficiently but may be bilaterally impaired following unilateral brain injury due to its high demands for cognitive integration.

Executive Functions

One of the most critical aspects of the neuropsychological examination lies in the assessment of *executive* functions, which play a vital role in the planning and initiation of independent activities, self-monitoring of performance, inhibition of inappropriate responses, switching between tasks, and planning and control of complex motor and problem-solving responses. Although the prefrontal lobes have long been regarded as an essential component in mediating these functions, more recent developments in the neurosciences have also led to an increased appreciation for the essential role that is played by extensive cerebral interconnections between subcortical and cortical regions of the brain.

An underlying executive component is implied in many of the techniques discussed above. For example, as assessed on a serial addition task, complex attention depends on shifting focus and sustaining a cognitive set flexibly. Standard expressive language measures of verbal fluency are highly dependent on the individual's ability to self-activate and maintain a productive flow of responses without losing the cognitive set. Memory functions rely on processing information efficiently, including using higher-order processing strategies to organize incoming information (e.g., clustering words in a memory list based on semantic categories). Memory deficits attributed to retrieval failure in the information processing model of memory are often associated with executive deficits. However, some specific techniques have gained favor over the years as primary measures of executive functions.

The *Wisconsin Card Sorting Test* (WCST) requires an individual to sort a long series of cards with designs according to various principles that can only be gleaned by incorporating feedback from the examiner. This measure, then, assesses the individual's ability to generate and test various solutions, flexibly respond to changing circumstances, and integrate new information with experience to respond as efficiently as possible. The *Category Test* from the Halstead–Reitan Battery is less commonly used to assess the individual's flexibility to incorporate feedback into problem-solving responses. The *Delis-Kaplan Executive Function System* (D-KEFS) is based on several other techniques associated with executive abilities. These include variations on the *Trail-Making Test, Verbal Fluency Test, Design Fluency Test, Color-Word Interference Test, Sorting Test, Twenty Questions Test, Word Context Test, Tower Test,* and *Proverb Test.*

Psychological Factors

A key component of any neuropsychological examination involves considering the degree to which longstanding personality or other psychological factors (including current stressors) might contribute to the patient's presentation. Common techniques for assessing personality and psychological factors include paper and pencil techniques, such as the *Beck Depression Inventory-II, Geriatric Depression Scale,* and *the Minnesota Multiphasic Personality Inventory*, which is now in the 3rd edition; however, the 2nd edition and the *Revised Form* are still commonly used (MMPI-2/RF/3*).* In addition, the *Personality Assessment Inventory* (PAI) is another more extended personality inventory that has become increasingly used by neuropsychologists in recent years.

Performance and Symptom Validity

To ensure the validity of the neuropsychological assessment, the neuropsychologist needs to evaluate performance and symptom validity as a standard part of every assessment. Performance validity refers to the credibility of examinee performance on cognitive tests, while symptom validity refers to the veracity of symptom endorsement.

When the test data raise concerns about validity, the neuropsychologist faces a sensitive diagnostic scenario since it is difficult to differentiate between nonconscious factors that can result in noncredible responding versus intentional exaggeration of symptoms (e.g., factitious disorder or malingering). Researchers have proposed several sets of criteria that help clinicians think about the possibility of malingering by considering discrepancies between test data and the examinee's self-report, in contrast to known patterns of brain functioning, observed behavior, reliable collateral reports, and documented background history. In addition, poor performance on performance validity tests (PVTs) or signs of psychiatric, cognitive, or somatic overreporting on symptom validity tests (SVTs) would also add weight to the possibility of noncredible responding or outright malingering.

Several free-standing PVTs exist, including forced-choice measures such as Green's *Word Memory Test* and the *Test of Memory Malingering.* The forced-choice measures are beneficial at discerning intentionality, as significantly below chance performance provides the most substantial evidence for bona fide malingering. Nonforced choice measures also include *Rey's 15-Item Memory Test* and the *Dot Counting Test.* Still, other performance validity measures can be found embedded within other standard neuropsychological procedures. These PVTs share a foundation based on normative research that indicates that patients with histories of bona fide brain injury or even dementia can be expected to perform at adequate levels on these instruments. Poor performance on a PVT suggests that factors other than brain injury alone should be considered as possible explanations. Other indicators of performance validity are based on the pattern of an individual's performance on standard procedures in a neuropsychological examination. For example, because it is always easier to recognize than to freely recall memory material, comparing an examinee's performance on recognition and recall offers a means for assessing the possibility of poor effort. In general, a patient's level of effort is best evaluated in the context of the broader history and pattern of neuropsychological performance, so it would not make sense to administer PVTs in isolation. An experienced clinical neuropsychologist will avoid equating poor performance on PVTs with "malingering" and instead use these results to think about factors other than brain injury or outright malingering that might influence an examinee's performance.

Commonly used SVTs include those embedded within the more extended personality inventories used by neuropsychologists (i.e., MMPI-2/RF/3 or PAI), but freestanding SVTs also exist, such as the Structured Inventory of Malingering Symptomatology (SIMS). The F-family scales within the MMPI-2/RF/3 encompass scales that assess overreporting of psychiatric, somatic, and cognitive symptoms. In addition to the F-family scales, the MMPI-2/RF/3 also includes the Response Bias Scale (RBS), which was derived to be uniquely indicative of feigned cognitive dysfunction. The MMPI-2/RF/3 also includes a few SVTs indicative of under-reporting of psychiatric symptoms. The PAI includes several SVTs that assess the presence of over and under-reporting of symptoms, including the recently derived Cognitive Bias Scale (CBS), which has been shown to operate similarly to the RBS. Therefore, it can be beneficial for neuropsychologists.

A 32-year-old woman with 13 years of education was seen for disability evaluation, claiming current “trouble remembering things.” Her account of personal history was vague, and she “forgot” information, such as her birth date and mother’s maiden name. Response latencies were extremely long, even for highly familiar information (e.g., count from 1 to 20), she could not repeat more than three digits forward consistently, and on a word list learning procedure, she was not able to correctly recognize more items (only five) than she could freely recall (also five). Despite her otherwise fluent language, she could only generate five examples of animals in 1 minute. On a forced-choice PVT, she performed significantly below chance levels. When she was asked to recall 15 items on a procedure (Rey’s 15-Item Memory Test) that the examiner presented as challenging, although, in reality, it is relatively simple, her performance demonstrated exaggerated errors of commission. On objective personality testing (i.e., MMPI-2-RF), the RBS scale was significantly elevated. The evaluation concluded that current levels of cognitive functioning could not be conclusively established due to overt noncredible performance.

THERAPEUTIC DISCUSSION OF RESULTS

A vital component of the neuropsychological examination process is found in the therapeutic opportunity to discuss the examination results with the patient and family or other caregivers. This meeting can represent a powerful therapeutic opportunity to educate and clarify individual and relationship issues, impacting the identified patient’s functioning. If the patient’s active cooperation in the initial examination has been appropriately enlisted, then the patient will be prepared to invest value and confidence in the examination findings. When discussing the results, it is helpful to review the examination goals with the patient and supportive family or caregivers and clarify the expectations of those present. Typically, these sessions will include information about the patient’s diagnosis, emphasizing the natural course and prognosis as well as compensation and coping strategies for the patient and family. Given the impact of chronic neurologic disease upon the family system and the patient, explicit discussion of these issues is critical in maximizing adjustment to brain injury. It is equally important to discuss the impact of the results on the patient's current living circumstances, future goals, and course of adjustment. It is not unusual for strong emotions and underlying tensions within family relationships to come to light in the context of honest discussion. So the results discussion can represent a significant therapeutic opportunity to model effective communication and problem-solving techniques.

The most challenging discussions usually arise when the objective examination results do not meet the preconceived beliefs of the patient or family members. In such cases, it is helpful to begin the feedback discussion by first asking the patient to describe their impression of their performance to identify the patient's perspective and associated expectations explicitly. For those persons who appear to be strongly invested in confirming a pathologic diagnosis of memory impairment, it may be beneficial to adopt a psychoeducational approach by using materials such as the Information Processing Model of Memory presented in Figure 7.4–3.

FURTHER READINGS

American Bar Association/American Psychological Association Assessment of Capacity in Older Adults Project Working Group. *Assessment of Older Adults with Diminished Capacity: A Handbook for Psychologists*. 2008.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. American Educational Research Association; 2014.

American Psychological Association & American Bar Association. ABA/APA Assessment of Capacity in Older Adults. 2008. https://www.apa.org/pi/aging/programs/assessment

Amunts K, Mohlberg H, Bludau S, Zilles K. Julich-Brain: a 3D probabilistic atlas of the human brain's cytoarchitecture. *Science*. 2020;369(6506):988–992.

Barr WB. Historical trends in neuropsychological assessment. In: Morgan JE, Joseph H, eds. *Textbook of Clinical Neuropsychology*. 2nd ed. Routledge; 2018:3–13.

Benton AL, Sivan AB, des Hamsher K, Varney NS, Spreen O. *Contributions to Neuropsychological Assessment: A Clinical Manual*. 2nd ed. Oxford University Press; 1994.

American Academy of Clinical Neuropsychology, Board of Directors. American Academy of Clinical Neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. *Clin Neuropsychol*. 2007;21(2):209–231.

Bondi MW, Smith GE. Mild cognitive impairment: a concept and diagnostic entity in need of input from neuropsychology. *J Int Neuropsychol Soc*. 2014;20(2):129–134.

Boone KB. *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective*. Guilford Press; 2007.

Cummings JL. Frontal-subcortical circuits and human behavior. *Arch Neurol*. 1993;50(8):873–880.

Dronkers NF, Ogar J. Aphasia: overview. In: Aminoff MJ, Daroff RB, eds. *Encyclopedia of the Neurological Sciences*. Elsevier Inc; 2014:244–247.

Fujii DEM. Developing a cultural context for conducting a neuropsychological evaluation with a culturally diverse client: the ECLECTIC framework. *Clin Neuropsychol*. 2018;32(8):1356–1392.

Gorno-Tempini ML, Hillis AE, Weintraub S, et al. Classification of primary progressive aphasia and its variants. *Neurology*. 2011;76(11):1006–1014.

Graff-Radford J, Jones DT, Graff-Radford NR. Pathophysiology of language, speech and emotions in neurodegenerative disease. *Parkinsonism Relat Disord*. 2014;20(Suppl 1):S49–S53.

Grant I, Adams KM. *Neuropsychological Assessment of Neuropsychiatric and Neuromedical disorders*. 3rd ed. Oxford University Press; 2009.

Haaland KY, Dum RP, Mutha PK, Strick PL, Troster AI. The neuropsychology of movement and movement disorders: neuroanatomical and cognitive considerations. *J Int Neuropsychol Soc*. 2017;23(9-10):768–777.

Hannay HJ. Proceedings of the Houston Conference on specialty education and training in clinical neuropsychology, September 3–7, 1997, University of Houston Hilton and Conference Center. *Arch Clin Neuropsychol*. 1998;13(2):157–158.

Heaton RK, Miller SW, Taylor MJ, Grant I. *Revised Comprehensive Norms for an Expanded Halstead-Reitan Battery: Demographically Adjusted Neuropsychological Norms for African American and Caucasian adults, Professional Manual.* Psychological Assessment Resources; 2004.

Heilman KM, Valenstein E. *Clinical Neuropsychology*. 5th ed. Oxford University Press; 2012.

Lezak MD, Howieson DB, Bigler ED, Tranel D. *Neuropsychological Assessment*. 5th ed. Oxford University Press; 2012.

Luria AR, Haigh B. *Higher Cortical Functions in Man*. 2nd ed. Springer US; 1980.

Miller BL, Cummings JL. *The Human Frontal Lobes: Functions and Disorders*. 2nd ed. Guilford Press; 2007.

Morgan JE, Ricker JH, eds. *Textbook of Clinical Neuropsychology*. 2nd ed. Routledge; 2018.

Strauss E, Sherman EMS, Spreen O, eds. *A Compendium of Neuropsychological Tests*. 3rd ed. Oxford University Press; 2006.

Stucky KJ, Kirkwood MW, Donders J. *Clinical Neuropsychology Study Guide and Board Review*. 2nd ed. Oxford University Press; 2020.

Stuss DT, Knight RT, eds. *Principles of Frontal Lobe Function*. Oxford University Press; 2003.

Suchy Y. *Executive Functioning: A Comprehensive Guide for Clinical Practice*. Oxford University Press; 2015.

Vanbellingen T, Kersten B, Van de Winckel A, et al. A new bedside test of gestures in stroke: the apraxia screen of TULIA (AST). *J Neurol Neurosurg Psychiatry*. 2011;82(4):389–392.

Wechsler D. *WAIS-IV Administration and Scoring Manual*. Psychological Corporation; 2008.

7.5 Personality Assessment of Adults

Russell L. Adams, Ph.D., A.B.P.P-NP., C.L., and David L. Adams, Psy.D., B.C.B.A.-D.

Although the term personality has been defined in various ways over the years, in this chapter, personality is defined as an individual's enduring and pervasive motivations, emotions, interpersonal skills, attitudes, and traits. The systematic empirical measurement of personality traits is the purview of personality assessment. Most personality tests are norm based and have specified reliability and validity. A given test's reliability and validity together with the size and nature of the normative base largely determine its usefulness.

For a given personality test to be accepted, the test developer must demonstrate how reliable the test is and how valid it is. Reliability refers to the consistency with which a given individual would obtain the same or similar score if given the same test or a parallel version of the same test was administered again. Under similar circumstances, the validity of a given test refers to the degree to which the test measures what it purports to measure. In other words, does a personality test measuring depression differentiate effectively between individuals who independently have been determined to have no depression, mild, moderate, or severe depression? A valid depression test should be able to measure changes in depression that may occur over time with effective treatment.

The systematic measurement of such difficult-to-define personality traits as depression and anxiety can be challenging. Particularly difficult is assessing characteristics such as suicide potential, impulsivity, or inability to delay gratification.

Personality tests are used for a variety of purposes. A personality test can screen or identify individuals in need of treatment or intervention. Personality tests can also be used to assist in the differential diagnosis of a given patient, especially when the patient does not fit neatly in one of the DSM-5 categories. Some personality tests can help inform the clinician concerning what treatment might be most effective. Personality tests can also monitor the progress of a given patient in treatment over time.

Personality tests are used extensively in the medical–legal context. For example, is a given patient suffering psychologically from a severe motor vehicle accident? Does that patient have post-traumatic stress disorder (PTSD)? Some personality tests are also used to identify individuals who are over- or underreporting psychological symptomatology.

PURPOSES OF PSYCHOLOGICAL TESTING

Personality testing can be an expensive undertaking. A considerable amount of time is required to administer, score, and interpret psychological test results. Personality testing should not be routinely obtained from all psychiatric patients, especially in today's cost-containment environment. Nevertheless, personality testing can be helpful with selective patients from both a clinical and a cost-benefit analysis perspective.

Assisting in Differential Diagnosis

As any clinician knows, psychiatric diagnosis can be a difficult and, at times, confusing exercise. However, knowing a patient's diagnosis is essential to treatment, as a proper diagnosis can assist in understanding the etiology of the presenting psychiatric problem and the prognosis of the disorder. Consider the diagnostic dilemma in the following example:

> A 49-year-old White man had abruptly resigned his position as an accountant and decided he was going to start an oil exploration business. He had never worked in the oil business and knew nothing about the profession. The patient had received a revelation from an unknown entity through an auditory hallucination. This voice told him he would become quite wealthy in the business if he would simply follow the directions given to him. Around this time, the patient had a marked change in personality. Although his grooming was formerly very neat and appropriate, he became disheveled. He began sleeping about 3 hours a night. He became somewhat agitated and talked loudly to those around him.

The differential diagnosis in this case illustration includes schizophrenia and bipolar disorder. Psychological testing might be helpful in assisting in this differential diagnosis, as well as in formulating a treatment plan.

Aiding in Psychotherapy

Psychological tests can be useful in psychotherapy. The usefulness of these tests can be even more important for short-term, problem-centered therapy where understanding the patient and their problem must be accomplished quickly. Psychological assessment can be used in pretreatment planning, assessing progress once therapy begins, and in evaluating the effectiveness of therapy. Patients need to have objective information about themselves at the time of therapy if they are to go about changing themselves productively. Personality tests, particularly objective tests, allow patients to compare themselves to objective norms and evaluate the extent and magnitude of their problem. Testing also can reveal areas of the patient's life that may be problematic but for which the patient may not have a full appreciation. Information about patients' willingness to reveal information about themselves can also be helpful. In some instances, the test report can be the subject of an early treatment session. Psychological tests may reveal considerable information concerning the patient's inner life, feelings, and images, which may make therapy progress faster. Psychological testing can provide baseline information at the beginning of therapy, and repeat testing can then be used to assess change that occurred during the course of therapy.

Providing Narrow-Band Assessment

Narrow-band personality tests measure a single personality characteristic or a few related characteristics. Broad-band personality tests, on the other hand, are designed to measure a wide spectrum of personality characteristics. A psychiatrist may need answers to specific questions, such as those that arise when assessing the degree of clinical depression, measuring the intensity of the state or trait anxiety, or, possibly, quantifying the amount of a patient's anger. Such quantification can be helpful in measuring severity or in providing a baseline for future assessment.

Psychological tests can often assist with a specific problem area. For example, a young college student may be unsure what to major in. Vocational interest inventories, such as the Strong Vocational Interest Inventory and the Kuder Occupational Interest Survey, may help to better define the student’s vocational interest. The Strong Vocational Interest Inventory compares a given individual’s responses with those of a large group of successful individuals in a variety of occupations. Test results can show that a student’s interests are similar, for example, to those of successful persons in occupations such as social work or public administration and dissimilar to those of engineers or scientists.

PSYCHOMETRIC PROPERTIES OF PERSONALITY ASSESSMENT INSTRUMENTS

The quality of personality tests varies widely. On the one hand, there are well-constructed, empirically validated instruments, and, on the other hand, there are “psychological tests” that one can find in the Sunday supplement of the newspaper or on the internet. Evaluating the usefulness of particular psychological instruments can be challenging, even to the well informed.

Just as the *Physician’s Desk Reference* helps physicians evaluate the usefulness of medications, the *Buros Mental Measurement Yearbook* helps psychologists evaluate the usefulness of personality instruments and other psychological and achievement tests. The *Mental Measurement Yearbook* is a reference source of critical reviews of individual tests. These reviews evaluate the scientific underpinnings of each test by an expert in the field. The reviews generally address such issues as the reliability and validity of the tests, the normative group used in construction, and various empirical studies utilized to check out the reliability and validity of the instrument. The advantages and disadvantages of the test are described. The reviews generally describe situations in which the test can be appropriately used. Practical information concerning the test publisher, length of time of administration, and other considerations are also included. The *Buros Mental Measurement Yearbook* is available in most medical school and university libraries and can be ordered through the book’s publisher.

Normative Sample

To construct a personality test, a representative sample of subjects (normative sample) should be administered the test to establish expected performance. Basic issues, such as the size and representativeness of the sample used to construct the test, must be evaluated. A sample of 30 individuals of a given diagnosis in a single location is seldom appropriate, because there are so many moderating factors impacting results, such as age, education, sex, and racial and ethnic grouping. These factors frequently impact the norms, and appropriate controls must be made for these variables in many cases.

To illustrate this point, the Minnesota Multiphasic Personality Inventory 2 (MMPI-2), a well-constructed instrument, initially tested approximately 2,900 subjects. However, approximately 300 subjects were eliminated because of test invalidity or incompleteness of needed information. The final group was comprised of 2,600 community volunteers, including 1,462 women and 1,138 men. The MMPI-2 normative group was selected using census data as a guide and was broadly representative in terms of region of the country, racial composition, age, and formal education. Despite efforts to control for moderating variables, the sample was slightly overrepresentative in higher education levels. Obviously, the cost of obtaining large census-based normative samples is enormous, and few tests are able to meet these challenging criteria. Nevertheless, many tests that fall short of the national sample (and most do) can be useful in the hands of an experienced clinician who is sensitive to the problem and makes appropriate clinical adjustments in interpreting the results.

Test Characteristics

To be useful, any psychological test must be completed, in its entirety, by the intended test taker. If the questions are offensive or are difficult to understand, then the individual taking the test may not complete all items. These omissions can create problems, especially when normative tables are used to interpret results. To be useful, an instrument should be short enough to complete in a relatively short period of time. The Beck Depression Inventory or the Geriatric Depression Scale can each be easily completed in 5 to 10 minutes. Because of its short length, the test can be given repeatedly to a depressed patient in treatment to monitor changes over time. The MMPI-2, on the other hand, takes approximately 1 to 1½ hours to complete. This test is not appropriate for repeat testing over a short period of time. Most tests are easily scored and compared to normative tables.

Validity Issues

Perhaps the most important characteristic in evaluating the scientific merit of a given personality test is the validity of the instrument. Does the test measure what it purports to measure? If a test is designed to measure depression, does it indeed measure depression? Although validity may seem like a simple issue to address, it can become complex, especially when attempting to measure such characteristics as self-esteem, assertiveness, hostility, or self-control. The test developer must first operationally define what is meant by the personality characteristic in question. A number of different approaches have been used to measure validity, including face validity, logical content validity, construct validity, criterion validity, and factor validity.

Face Validity. *Face validity* refers to the content of the test items themselves. In other words, do the items appear to measure what they purport to measure? For example, does a depression scale ask questions concerning sadness, guilt, sleep disturbance, weight loss, helplessness, hopelessness, or suicidal thoughts? One problem with face validity is that professionals differ in their subjective appraisal of individual items. One psychologist may think a given item does, indeed, measure the concept under consideration, while other professionals may not. To address these differences in definition, the test developers may utilize an approach known as *logical content validity*. Using this approach, the developer will describe in the manual the procedure used to select items for the test that will not depend significantly on the subjective judgment of any one professional. For example, in developing the Taylor Manifest Anxiety Scales, a number of different psychologists were asked to review items from the MMPI that seemed to measure a formal definition of manifest anxiety, and only those items agreed on by this group of psychologists were included. Many creative approaches have been used to demonstrate the logical content validity of various tests.

Criteria and Construct Validity. Although face validity refers to the degree that test items appear on the surface to measure what the instrument, as a whole, purports to measure, *criterion validity* uses data outside the test itself to measure validity. For example, if a test were designed to measure hypochondria, one would expect that a patient with high scores would have more visits to the physician’s office, complain of more physical symptoms, and use prescribed and over-the-counter medications more extensively. These criteria are objective and external to the test itself and can be independently measured. Scores on the hypochondria instrument can be statistically correlated with a number of separate objective criteria, such as those mentioned previously, to determine if the test measures the theoretical construct being measured (*construct validity*).

Concurrent and Predictive Validity. To determine test *concurrent validity*, external measures are obtained at the same time that the test is given to the sample of subjects. Thus, the concurrent validity of the test reveals that, at a given point of time, high scorers on a test may be more likely than low scorers on a test to manifest the behavior reflected in the criteria (e.g., more physician visits or more medication for a hypochondriac patient). Occasionally, however, a test developer is interested in predicting future events. For example, a test designed to predict marital satisfaction among engaged couples may use the couples' future marital satisfaction and marital status as an external criterion. Thus, the external criterion would have to be obtained a number of years down the road for the test to have this predictive validity. This type of validity is referred to as *predictive validity*. Because it is difficult to obtain such information several years later, predictive validity is seldom available in most personality tests, although such data can be valuable when available.

The *discriminant validity* of a test tells whether the test is able to discriminate between known groups of patients at a given time. For example, is a test of depression able to statistically discriminate nondepressed subjects from subjects who suffer from independently diagnosed major depression, adjustment disorder with depressed mood, and dysthymia? Is a measure of depression able to statistically discriminate among mild, moderate, and severe major depressive disorder?

Factor Validity. Factor validity utilizes a multivariate statistical technique known as *factor analysis* to determine if certain major groups of items on a given test empirically cluster together. For example, on a personality test measuring depression, do items concerning vegetative symptoms tend to covary together? Are there other clusters of items on the depression test measuring cognitive problems, mood disturbance, and social withdrawal?

Reliability

To be useful, a test must not only be valid, but it must also be reliable. *Reliability* refers to the degree that a test measures what it purports to measure, consistently. The key word here is *consistently.* For example, a patient with agoraphobia takes a new agoraphobia test today and takes the same test tomorrow. However, on the second testing, the scores change markedly even though the patient had not received any treatment or undergone any change in their underlying agoraphobia. Such a test would not be useful, because it would not consistently measure the trait in question. The test would have poor reliability. There are several means of checking reliability, including test–retest reliability, internal consistency reliability, and parallel form reliability.

Test–Retest Reliability. *Test–retest reliability* is obtained by simply administering the same test on two occasions to a group of subjects and statistically correlating the results. To be useful, the correlation coefficient should be at least 0.80 if the two tests were administered within 2 weeks of each other and if the trait in question is stable.

Internal Consistency Reliability. Another approach to determine *internal consistency reliability* is to divide a given test into two equal parts and statistically correlate the two halves for the test with each other. This technique determines the *split-half reliability* of a test. The first half of the test should be highly correlated with the second half of the test if the test is consistently measuring what it purportedly measures. Alternatively, the odd-numbered items could be correlated with the even-numbered items (*odd–even consistency or reliability*). When a test is split into two parts and the obtained scores correlate with each other, the resulting correlation coefficients statistically underestimate the true reliability of the test because the number of items in the test influences the magnitude of the correlation. Hence, if there were only half that number of items in the original test, then the results would be underestimated. To correct for this underestimation, the *Spearman–Brown formula* is used. Again, a reliability coefficient of 0.80 to 0.85 is needed to demonstrate usefulness in most circumstances. However, the higher the reliability as measured by the correlation coefficient, the better the test instrument. No test is perfectly reliable. The highest possible correlation coefficient is 1.0, and the lowest is 0, assuming no negative correlation.

To determine how internally consistent a given test is, a test developer also may use a statistical technique called *Cronbach's coefficient alpha*. With Cronbach's coefficient alpha statistic, one can determine the average correlation among all the items on the test. The higher the average correlation, the better the reliability.

Parallel Form Reliability. Sometimes, two separate forms of the same test are needed. For example, if the process of taking a test at one point in time would by itself influence a patient's score the second time that they took the same test, then parallel forms of the test are needed. *Parallel forms* of a test measure the same construct but use different items to do so. To ensure that the test does, in fact, measure the same construct, the correlation coefficient between the two parallel forms of the same test is computed. Such parallel form reliability should be at least 0.90 or higher.

Use of Standard Error of Measurement to Assess Reliability. Another way to assess the usefulness of a given test is to examine the test's standard error of measurement (SEM), which should be included in the test's manual. The SEM is a single statistic that is used to estimate what the score of a given patient would be on the test if the patient took the same test again within a short period of time. Consider the following example: A patient takes a given test today and obtains a score of 90 on that test. The SEM on that test is 3. An SEM value of 3 together with an individual's score of 90 on the test means that if that patient were to take that same test again tomorrow, they would score within ± 1 SEM of their original test performance (e.g., ±3 points) 68% of the time. Thus, the patient in question should score between 87 (90 − 3) and 93 (90 + 3) on the second test 68% of the time. The same individual would score ±2 SEM values (e.g., 2 × 3 = 6) on the second test 95% of the time.

In the above example, the patient with an original score of 90 on a given test would score between 84 (90 − 6) and 96 (90 + 6) on the second test again 95% of the time. Other things being equal, the smaller the SEM relative to the standard deviation of the tests, the better.

ADULT PSYCHOLOGICAL TESTS

Objective Personality Tests

Objective personality tests are rather straightforward in approach. Patients are usually asked specific and standard questions in a structured written or oral format. Each patient is typically asked the same question. The data obtained from a given patient are compared to similar data obtained from the normative group. The degree to which the patient deviates from the norm is noted and is used in the interpretive process. The patient's responses are scored according to certain agreed-upon criteria. The obtained scores are then compared with normative tables and often converted to standardized scores or percentiles, or both. The MMPI-2 is an example of an objective personality test.

There are literally hundreds of different objective personality instruments, and discussing all of these instruments is beyond the scope of this section. Table 7.5–1 contains a representative sample of objective personality tests and concise descriptions of each test, along with a brief listing of strengths and weaknesses. Because the MMPI-2 is, by far, the most frequently used objective personality test, it is discussed in considerable detail in this section to give the reader an orientation to objective personality assessment, as well as objective personality development.

This focus on the MMPI-2 is not meant to imply that the MMPI-2 is the only valid and reliable objective personality instrument or that other objective personality tests are not as useful as the MMPI-2 for a particular purpose. There are many useful objective personality instruments.

Minnesota Multiphasic Personality Inventory-2. Two excellent books on the MMPI-2 are James N. Butcher's *The MMPI-2 in Psychological Treatment* and John R. Graham's *MMPI-2 Assessing Personality and Psychopathology*. The following discussion relies heavily on information contained in these two resources. The MMPI-2 consists of 567 true/false questions concerning a wide variety of issues. The patient completes the test using a pencil. The MMPI-2 is relatively easy to administer and score and takes approximately 1 to 1½ hours for most patients to complete. It requires only an eighth-grade reading comprehension. The test provides a number of internal checks for measuring the patient's degree of cooperation. The patient is simply asked to answer true and false to such questions as "I have a good appetite," or "No one seems to understand me."

Scoring of the MMPI-2 involves adding up the number of responses on numerous scales and comparing the results to certain normative information. Interpretation of the MMPI-2 is more straightforward than with many other tests, because of the vast amount of empirical information available concerning the MMPI-2. More than 12,000 articles and books have been published on this test to date.

There are over 115 translations of the MMPI-2. The test is also cost-effective, because relatively little professional time is involved in the administration and scoring. The patient is simply given instructions, and then they complete the test independently in the professional's office. In certain situations, the MMPI-2 can be administered on the computer. A computer can also plot test profiles and compute various supplemental scales. The number of available specialized scales increases the usefulness of the MMPI-2 with various populations (e.g., patients with PTSD, college students who are maladjusted, and prisoners with violent tendencies).

Questions cover a wide range of different areas. When a patient takes the MMPI-2, questions are not grouped in any particular order to aid in interpretation. Various items in the MMPI-2 can be selected, sorted, and analyzed according to various criteria. The most common scales of the MMPI-2, including the validity scales, clinical scales, content scales, supplemental scales, and critical items, are discussed below.

VALIDITY SCALES. The patient can approach a mental status interview in a variety of ways. The patient can be painfully honest and straightforward with the interviewer, revealing their innermost aspirations and fears. The patient may be somewhat distrusting of the interviewer and may try to slant the content of the interview in a certain fashion to manipulate the outcome they want. The patient may be disingenuous or even attempt to be deceitful by malingering. Obviously, the information obtained by psychiatrists or other mental health professionals in an interview may well be colored by the patient's approach to the interview. Some patients are, by their nature, reserved and may use denial successfully during a clinical interview. Other patients may have little insight into their feelings and thoughts. Many times, a psychiatrist may not know how candid or straightforward a patient is during an interview.

When patients are administered personality tests like the MMPI-2, they approach the test in a similar fashion to how they approach an interview.

One advantage of the MMPI-2 is that the test has been devised to assess different test-taking attitudes (e.g., fake bad, fake good, or social desirability responding). Through use of the validity scales, the clinician gleans information that helps in interpreting the remaining scales of the MMPI-2. Is the patient overly defensive? Are they attempting to fake good or fake bad? Does the patient maintain a reasonable balance between self-disclosure and excessive defensiveness?

Lie (L) Scale. This scale consists of 15 items designed to measure general frankness. The items include questions concerning personal characteristics that, although highly desirable in our culture, are, nevertheless, infrequently practiced (e.g., questions such as, "I have never failed to tell the truth," or "Once in a while I think of things too bad to talk about," and "I do not always tell the truth"). Patients who score high on the lie scale may be psychologically naive or defensive. They may want to be seen in a positive light or they may have difficulty admitting common human frailties.

F-Scale. This is one of the longest scales, consisting of 60 items that 10% or fewer of the normative samples responded to in the positive direction. This scale includes such items as follows: "Evil spirits possess me at times," and "Most any time I would rather sit and daydream than do anything else."

The items on the F-Scale are quite diverse in content and may be indicative of severe psychopathology such as hallucinations or delusions. Most people in the normal sample responded to fewer than five of these items. Individuals who obtain exceptionally high scores on this scale are often either extremely pathologic or trying to claim mental illness falsely. As with all personality tests, interpretation depends greatly on the setting in which the test was given, as base rates for different settings and circumstances differ. For example, patients who have been seen for court sanity evaluations have different base rates of symptom magnification than do patients seen for adjustment problems in an outpatient clinic.

Table 7.5–1.
Objective Measures of Personality

Name	Description	Strengths	Weaknesses
Minnesota Multiphasic Personality Inventory-2 (MMPI-2)	567 items; true–false; self-report format; 20 primary scales	Large research base; most known and utilized objective personality assessment instrument; provides considerable clinically useful information while requiring limited clinician time	Required that the patient be able to read items, comprehend instructions, and follow direction; some patients attempt to fabricate responses but validity scales available
Minnesota Multiphasic Personality Inventory-3 (MMPI-3)	335 items; includes 72 new items with 52 scales and 10 validity scales; takes 25–35 minutes to administer	Spanish version available; large national diverse normative sample; based on a long extensive history with the respected MMPI and MMPI-2; can be computer-administered with an empirically based interpretative report	Relatively new test without the strong independent research base enjoyed by the MMPI and MMPI-2; unclear if previous research of MMPI-2 will apply, as the structure of the tests is slightly different than previous versions
Clinical Multiaxial Inventory-III (MCMI-III)	175 items; true–false; self-report format; 25 primary scales	Brief administration time; corresponds well with DSM-IV	High degree of item overlaps in various scales; may overdiagnose pathology in some cases through

			many scale elevations
16 Personality Factor Questionnaire (16 PF)	True–false; self-report format; 16 personality dimensions	Sophisticated psychometric instrument with considerable research conducted on nonclinical populations	Limited usefulness with clinical populations
Personality Assessment Inventory (PAI)	344 items; Likert-type format; self-report; 22 scales	Includes measures of psychopathology, personality dimensions, validity scales, and specific concerns to psychotherapeutic treatment	The inventory is increasingly being utilized, but it does not currently have the strong research basis found in the MMPI-2
Personality Inventory for DSM-5 (PID-5)	220 items; Likert-type format; 25 personality trait facets with 5 broad trait domains	Corresponds to the 25 personality traits outlined in the DSM-5	Additional research warranted
NEO Personality Inventory-Revised (NEO PI-R)	240 items with 3 validity scales	Corresponds to the Five Factor Model for personality	Requires a 6th-grade reading level
California Personality Inventory (CPI)	434 questions - true–false; self-report	Well-accepted method of assessing patients who do	Limited usefulness with clinical populations

	format; 20 "folk" scales of character with 4 classes	not present with major psychopathology	
Jackson Personality Inventory (JPI-R)	300 questions–true–false; self-report format; 15 personality scales	Constructed in accord with sophisticated psychometric techniques; controls for response sets	Unproved usefulness in clinical settings
Edwards Personal Preference Schedule (EPPS)	Forced choice; self-report format	Follows Murray's theory of personology; accounts for social desirability; measures 15 personal needs; often used in vocational counseling	Not widely used clinically because of restricted nature of information obtained
Psychological Screening Inventory (PSI-2)	139 items; true–false; self-report format	Yields 4 Basic Screening Scales, 21 Brief Content Scales, and 6 Misrepresentation Scales; can be used as screening measures concerning the possibility of a person's need for psychological treatment	The scales are short and have correspondingly limited reliability
Eysenck	Utilizes	Useful as a	Scales are short,

Personality Questionnaire (EPQ)	yes/no questions (48 questions in short version; 100 in longer version); self-report format	screening device; test has a theoretical basis with research support; measures personality traits using 4 scales	and items are transparent as to purpose; not recommended for other than a screening device
Adjective Checklist (ACL)	True–false; self-report or informant report	Can be used for self-rating or other rating; measures 37 traits and each of these scales are grouped into 6 Success Factors at work	Scores rarely correlate highly with conventional personality inventories
Comrey Personality Scales (CPS)	True–false; self-report format; 10 scales, including the validity check scales; 180 items	Factor analytic techniques used in test construction	Not widely used; factor analytic interpretation problems
Tennessee Self-Concept Scale (TSCS:2), 2nd edition	100 items; true-false; self-report format; 15 scales	Brief administration time yields considerable information; measures self-concept in 15 scales, which includes 4 validity scales	Brevity is also a disadvantage, lowering reliability and validity; useful as a screening device only

Adapted and revised from Robert W. Butler, Ph.D., and Paul Satz, Ph.D.

Backpage Infrequency Scale. A few patients, particularly those with passive–aggressive tendencies, will start to complete the MMPI-2 but, toward the end of the test, may respond to the items in a random fashion. The patient may simply want to finish the test quickly and leave the testing situation. The Backpage Infrequency (Fb) scale for the MMPI-2 is designed to pick up this tendency.

Variable Response Inconsistency Scale. The Variable Response Inconsistency Scale VRIN consists of 67 pairs of items. Each question of the pair probes for the same or similar information in a slightly different way, for example: (1) “I sleep very poorly at night,” and (2) “I wake up fully refreshed each morning.” If the patient responds “true” to question 1, then they would normally respond “false” to question 2. Failure to do so would reflect inconsistency.

CLINICAL SCALES. The clinical scales are generally considered the meat or substance of MMPI-2 clinical interpretations. The original clinical scales have an empirical base in the sense that the items were selected because they differentiated between specific clinical groups and the normative control group. The control group for the original MMPI, developed in the 1940s, consisted of 1,500 persons who were visitors to the University of Minnesota hospitals and other normal groups. Clinical groups varied in size but, for the most part, contained at least 50 subjects. For example, the schizophrenic scale group was developed with patients with different types of schizophrenia, and the depression scale was developed with a sample of psychiatric patients with depression, including some with manic depressive illness. Table 7.5–2 presents a brief illustration of the original sample groups. Although most clinical scales were developed using a specific clinically defined group, researchers with the MMPI soon learned that a given patient (such as one with schizophrenia) did not simply score high on the schizophrenia scale and low on all other scales. A schizophrenic patient might score high on many scales, but their highest score may well be the schizophrenic scale. If one were to diagnose all patients who scored high on the schizophrenic scales as schizophrenic patients, then a large number of false positives would result. To address this confusion, clinicians began to refer to the clinical scales by a number

(0 to 9) rather than the diagnostic entity that underlies the scale's development.

COMBINING SCALE ELEVATIONS IN THE CODE TYPES. As has been illustrated previously, interpreting single-scale evaluations on the MMPI-2 is problematic; thus, most clinicians use a combination of scale elevations in interpretation. This procedure usually involves using the patient's highest two clinical scales that exceed a *t* score of 65 to provide a configured interpretation that takes into account the relation between the scales. In this approach, for example, a 1-2 or 2-1 code type would be the same. In other words, a patient who scored highest on scale 1 (hypochondriasis [Hs]) and second highest on scale 2 (depression [D]) would be considered to have the same basic profile as an individual who scored highest on scale 2 (D) and second highest on scale 1 (Hs).

A three-point code, likewise, would list the three highest scales, all of which normally would be above a *t* score of 65. These scales also would usually be significantly higher than all of the other clinical scales.

On the MMPI-2, uniform *t* scores are used on clinical scales except in scales 5 (masculinity/femininity) and 0 (social introversion). *Uniform t scores*, briefly stated, are standardized scores in which the percentile equivalent and distribution are matched among the various clinical scales in terms of skewness and kurtosis. Thus, a uniform *t* score of 65 on scale 7 (psychasthenia) is roughly equivalent to a uniform *t* score of 65 on scale 9 (hypomania). The use of uniform *t* score rather than linear *t* scores was one of the improvements of the MMPI-2 over the original MMPI.

Table 7.5–2.
Minnesota Multiphasic Personality Inventory Scale Development

Scale 1—(**Hs: Hypochrondriasis**): The clinical group of this scale consisted of patients who had an obsessive overconcern about their physical health.

Scale 2—(**D: Depression**): This group consisted of a variety of patients with symptomatic depression, including reactive depression and manic depression among others.

Scale 3—(**Hy: Hysteria**): The hysteria group consisted of patients who had some type of sensorimotor deficit for which no organic basis could be found.

Scale 4—(**Pd: Psychopathic Deviancy**): This patient group had continual run-ins with legal authorities, yet they had normal intelligence and no serious neurotic or psychotic symptomatology. No major criminal types were included. These patients exhibited delinquent acts such as stealing, lying, excessive drinking, and sexual promiscuity.

Scale 5—(**Mf: Masculinity–Femininity**): This scale originally was developed to discriminate homosexual invert males. However, there were only a small number of items that statistically differentiated this group from the controls. As a result, other items were included that differentiated men from women in the normative sample.

Scale 6—(**Pa: Paranoia**): Patients in this group consisted of suspicious, mistrusting individuals who often had diagnoses other than paranoia. These persons could have a paranoid personality, paranoid schizophrenia, or an affective disorder. Interestingly, the scale had few false positives, but had a rather large number of false negatives.

Scale 7—(**Pt: Psychasthenia**): This scale was originally developed using a group of patients with psychasthenia. This diagnosis is no longer used today but does resemble a diagnosis of obsessive-compulsive disorder. These patients worry excessively; have major fears, compulsions, and rituals; and possess generalized anxiety and distress.

Scale 8—(**Sc: Schizophrenia**): The patient group used for this scale consisted of various types of schizophrenic patients. Efforts to develop separate scales for different subtypes of schizophrenia were not successful.

Scale 9—(**Ma: Mania**): Because patients who have full-blown mania are not testable, this scale was developed utilizing patients who were in the earlier stages of manic episodes.

Scale 0—(**Si: Social Introversion**): This scale was not developed on a particular diagnostic group. Rather the items were determined by contrasting selected females who scored in the high or low range on another measure of social introversion. Social Introversion–Extraversion scale of the Minnesota TSE (Thinking–Social–Emotional) Inventory.

Data from Graham J. *MMPI–2 Assessing Personality and Psychopathology*. Oxford University Press; 1990.

A review of the interpretation of common two- and three-code types can be found in most standard textbooks on the MMPI, some of which are included in the references. Certain MMPI authorities have used slightly more complex means of interpreting the MMPI-2. This approach emphasizes specific relationships about multiple scales. For example, Harold Gilderstadt and Jan Duker described criteria in their book *A Handbook for Clinical and Actuarial MMPI Interpretation* for what is referred to as a *1,2,3,4-type scale*, as follows:

1. Hs, D, Hy, and Pd greater than *t* score 70 and higher than all other scales.
2. Si less than *t* score of 70.
3. L, F, and K less than *t* score 70 unless two or more scales are greater than 100, in which case F less than *t* score 80.

The problem in using this complex method is that, in many settings, only a relatively small percentage of patients match these criteria. In other words, if one uses this method rigidly, then a substantial number of patients seen in an inpatient unit would not precisely match one of these complex criteria.

CONTENT SCALES FOR MMPI-2. Several efforts have been made to group items according to certain specified dimensions. Using a combination of rational and statistical procedures, Butcher and colleagues developed these 15 narrow-band scales. The 15 scales are internally consistent and relatively independent. Item overlap between scales is minimal. Scales include such content areas as anxiety, anger, cynicism, low self-esteem, fears, obsessiveness, and depression among many others. Norms for both males and females were simply derived using the same uniform *t* scores used in the MMPI-2 clinical scale formation. Reliability data on the scales are acceptable given the relatively few items in some of the content scales.

John R. Graham presents scoring guidelines in his book *MMPI-2 Assessing Personality and Psychopathology*. Validity data for the new scales are limited at the current time. Because efforts were made to develop reasonably homogeneous items, interpretation of these scales consists of examining high scores and low scores. For example, patients with high scores on the low self-esteem scale tend to have poor self-concept, anticipate failure, be overly sensitive to criticism, and have trouble making decisions. Individuals with low scores, on the other hand, are self-confident and decisive. A given content scale such as the anger scale, when considered alone, is very similar to the rapid assessment instruments discussed in a later section of this chapter.

HARRIS & LINGOES CONTENT SCALES. As previously stated, most of the MMPI clinical scales were developed using empirical key approaches. For example, an item was included in the depression scale because that item discriminated patients in the reference group from persons in the normal group. Although this method is scientifically sound, the approach results in the inclusion of subtle items that seemingly may have little to do with depression. For example, items such as, “I sometimes tease animals,” (false) and, “It takes a lot of argument to convince most people of the truth,” (false) are included in the depression scale (D) even though, on the surface, they have little to do with depression. Such items are called *subtle items*. Obvious items are those that have strong face validity, such as, “I wish I could be as happy as others,” (true) and, “I brood a great deal” (true). The depression subscale reflects a heterogeneous item group coming from a variety of aspects of depression, such as subjective depression, psychomotor retardation, mental dullness, etc. The same total score might reflect a considerably different combination of items. Knowing which items in combination actually contribute to the total score can be helpful in interpretation. The Harris & Lingoes scale can help in that process.

The Harris & Lingoes scales were developed logically by grouping items in a given clinical scale into standard scales that were judged to reflect a more homogeneous trait. Thus, the Harris & Lingoes scales for depression reflect separate scales for subjective depression (D1), psychomotor retardation (D2), physical malfunction (D3), mental dullness (D4), and brooding (D5).

The Harris & Lingoes scales can be used for two major purposes: (1) to see which group of items was elevated for a patient scoring high on a given clinical scale and (2) to see why a marginally elevated scale was elevated. For example, one patient who scored marginally high on the depression scale may be demonstrating psychomotor retardation and no subjective depression while another may be just the opposite.

SUPPLEMENTAL SCALES. Given the MMPI contained 567 items, numerous investigators have used the items to develop special purpose or supplemental scales. As early as 1972, more than 450 supplemental scales were developed, and the number has kept growing. Scales have been developed to predict success in psychotherapy (ego strength [Es]), to differentiate alcoholic patients from nonalcoholic patients (MacAndrew Revised [MACR]), and to differentiate assaultive prisoners from others not committing a crime (overcontrolled hostility [Oh]). Many dissertations for doctoral research have been conducted to uncover a new supplemental scale. The success of supplemental scales in accomplishing their stated purpose varies considerably. Many such scales were utilized for only one study, and others, such as the MacAndrew Alcoholism (MAC) scale, have been used in many studies over the years.

COMPUTER SCORING AND INTERPRETATION. A computer program can score the MMPI-2. A typical computer scoring program not only scores individual items into validity, clinical, content, and supplemental scales but also prints out certain positive critical items marked by the patient. These programs save considerable time because the program scores all of the material that may take hours to score by hand. Few clinicians have the time or patience to routinely score the vast number of scales found in the output of a typical computer program. The MMPI-2 computer programs can also be used to interpret the test clinically, or at least give clinical hypotheses concerning the patient. These computer-generated reports are in narrative format. Because the report comes from a computer, one may get the impression that the information is more valid or accurate than a clinical interpretation. Although attempts have been made to “validate” narrative statements made by the computer, efforts have only begun in this regard. One problem with computer interpretation of narrative reports is that the algorithms underlying the narrative reports are proprietary information owned by the company producing the program and are not available to the general user.

MMPI-2-Restructured Form. Yossef S. Ben-Porath and Auke Tellegen have developed a version of the MMPI-2 called the MMPI-2-Restructured Form (RF). Unlike the MMPI-2, which contains 567 items, the RF version contains 338 items; consequently, it takes less time to administer. The MMPI-2-RF has 51 revised empirically validated scales. The test is designed to take 35 to 50 minutes for most patients to complete. The MMPI-2-RF is not a replacement of the well-validated MMPI-2 test but instead provides an alternative instrument. The MMPI-2-RF has the Restructured Clinical (RC) scales at its core. The 51 scales comprising the MMPI-2-RF include eight different validity scales including a new Fake Bad Scale revised (FBS-r) symptom validity scale and three higher-order scales, such as the Thought Dysfunction Scale (THD).

As previously mentioned, the nine RC scales are at its core. The RC scales include an Antisocial Behavior Scale (RC-4) as well as an Ideas of Persecution Scale (RC-6).

In addition to the aforementioned scales, there are 3 different somatic scales, 10 internalizing scales, and 4 externalizing scales. In order to assess interpersonal issues, five different scales are provided, including the family problems scale. There are also two interest scales and five personality psychopathology scales.

Most mental health professionals use the computer for scoring because hand scoring is much too time-consuming and subject to some scorer errors. If the 567 items in the MMPI-2 are administered to a given patient, then a special software program converts the MMPI-2 scores to an MMPI-2-RF record.

The Minnesota Multiphasic Personality Inventory-3. The Minnesota Multiphasic Personality Inventory-3 (MMPI-3), the most recent version of the MMPI, was published in 2020 by the University of Minnesota press. A primary goal of the MMPI-3 was to update the norms to keep up with the changing demographics in the United States. According to the technical manual, the MMPI-3 was normed on 1,620 subjects selected to match the U.S. Census projections for race and ethnicity. The normative individuals (810 women and 810 men) were at least 18 years old. The normative sample was diverse, consisting of 60.3% White, 12.4% Black, 5.1% Asian, 14% Hispanic, 4.5% mixed-race, and 3.7% other racial-ethnic groups. Similarly, the sample matches the current census distribution in the United States for age. A special Spanish translation of the MMPI-3 is also available. This version was derived from a normative sample of U.S. Spanish speakers that included 275 men and 271 women.

In addition to having new norms, the 335-item MMPI-3 includes 72 new items that were not in previous versions and 24 reworded items that were included in previous versions but were updated in more contemporary language. The MMPI-3 has fewer items than the original MMPI and therefore can be administered in 25 to 30 minutes by a computer or 35 to 50 minutes with the paper and pencil version.

The MMPI-3 has 52 scales, 5 of which are completely new and not in previous versions. The new scales were designed to assess Combined Response Inconsistency (CRIN), Eating Concerns (EAT), Compulsivity (CMP), Impulsivity (IMP), and Self-Importance (SFI). The 52 scales include 10 validity scales, 3 higher-order scales, 8 RC scales, 26 Specific Problem Scales, and 5 Psychopathology Scales. The MMPI-3 is normally scored by the computer, and an interpreted report can also be generated.

Although the MMPI-3 is the most recent version, the MMPI-2 is still being published by Pearson and currently is regularly used by many psychologists if not most psychologists. The MMPI-2 has a long history with thousands of independently conducted research investigations supporting its continued use in a wide variety of clinical situations. At this point, psychologists can confidently use either version based on their preference.

The Personality Inventory for DSM-5—Adult. As stated in DSM-5, the diagnosis of personality disorders is a difficult and confusing endeavor. The problems include the fact that a given patient frequently meets criteria for more than one personality disorder. Moreover, the patient may not formally meet the criteria for any single personality disorder, but instead receive a diagnosis of “other specified” or “unspecified personality disorder.” Although this may be the correct diagnosis, it is mostly uninformative in terms of treatment or prognosis.

An alternative model for personality disorders is presented in DSM-5. This alternative model is found in Section III of the DSM-5 under “Emerging Measures and Models.” The current approach to personality disorders still appears in Section II of DSM-5. A complete discussion of this alternative DSM-5 model is beyond the scope of this chapter on personality assessment.

Briefly, the Section III Multidimensional Personality Trait model involves assessing a given patient on five broad domains of personality and 25 traits. This approach uses a formal psychometric rating instrument entitled "Personality Inventory for DSM-5" (PID-5), which is a self-report inventory completed by the patient and/or another similar form completed by a significant other in the person's life Personality Inventory for DSM-5-Informant-Report Form (PID-5-IRF).

This PID-5 consists of 220 questions. The Personality Disorders Workgroup of the DSM-5 developed it in 2012. It consists of a Likert scale with the following rating system: 0 (very false), 1 (somewhat false), 2 (somewhat true), and 3 (very true). The scoring of this instrument provides a raw score on five domains (negative affect, detachment, antagonism, disinhibition, and psychoticism). In addition, the PID-5 provides information on one of 25 personality trait facets including anhedonia, anxiousness, attention seeking, callousness, deceitfulness, depressivity, distractibility, eccentricity, emotional lability, grandiosity, hostility, impulsivity, intimacy avoidance, irresponsibility, manipulativeness, perceptual dysregulation, perseveration, restricted affectivity, rigid perfectionism, risk taking, separation insecurity, submissiveness, suspiciousness, unusual beliefs and experiences, and withdrawal. An average raw score is calculated for each of the five domains and each of the 25 personality trait facets. Higher average raw scores on a particular trait or domain indicate greater dysfunction.

The major advantage of the PID-5 is that it is directly tied to the DSM-5 model for personality disorders. Second, this instrument is a very easy scale for clinicians to administer. Third, it holds much promise for future research investigation.

The available research of the PID-5 is limited at the current time. In addition, the rating reliability of a practicing clinician is not known at the current time. Moreover, the interrater reliability of practicing clinicians using this measure has not yet been established.

There are also some shorter versions of the PID-5, including the PID-5-SF, the PID-5 faceted brief form (PID-5-FBF), and the PID-5-BF. The PID-5-SF version and the PID-5BF only have 100 questions, as compared to the 220 questions in the PID-5. The American Psychiatric Association published the PID-5-BF. The PID-5-BF measures five personality traits including detachment, negative affect, antagonism, disinhibition, and psychoticism, with a version for children between the ages of 11 through 17 years of age and another one for adults. This instrument only has 25 questions. However, there are limited studies demonstrating the reliability and validity of all the shorter versions.

Millon Clinical Multiaxial Inventory-III. Published in 1994, the Millon Clinical Multiaxial Inventory-III (MCMI-III) is an objective psychological test consisting of 175 true–false questions designed to measure psychopathology. The MCMI-III was updated in 2008 to include a new norming sample of 752 individuals with a diversity of mental health conditions. The MCMI-III should only be used with adults over the age of 18 with suspected mental health conditions. This instrument was standardized with psychiatric patients with mental health problems, so it should not be used with nonclinical populations. The MCMI-III is particularly effective with describing personality disorders.

The scales correspond to the 14 personality disorders and 10 clinical scales outlined in the DSM-IV. In addition, this instrument has 42 Grossman Faucet Scales (3 scales for each of the 14 personality disorders). Therefore, this instrument has a large number of scales for a limited number of items. The MCMI-III uses five Correction scales: the Disclosure Scale (X), the Desirability Scale (Y), the Debasement Scale (Z), the Invalidity Scale (V), and the Inconsistency Scale (W). These five correction scales help to determine the validity of the assessment based on the responses of the patient.

The advantages of this instrument are that it is quite easy to administer and only takes approximately 30 to 45 minutes to complete the test. The computerized narrative can be helpful to the clinician. One complaint concerning this test is that it may over diagnose pathology through many scale elevations.

Personality Inventory-Revised. Developed by Costa and McCrae, the NEO Personality Inventory Revised (NEO-PI-R) is a personality assessment that is based upon the Big Five Personality domains (openness to experience, conscientiousness, extraversion, agreeableness, neuroticism). These five personality domains are also known as the Five Factor Model (FFM). This instrument contains 240 items and is designed to take approximately 30 to 45 minutes to complete. The patient must be able to read on the sixth-grade level. The NEO-PI-R was published in 1990, and the NEO-PI-3 is the most up-to-date version of the NEO-PI-R.

There is a brief version of the NEO-PI-R that was revised in 2004 called the NEO-FFI. The NEO-FFI published the most recent update in 2010. The NEO-FFI only has 60 items (consisting of 12 items in each domain), and it takes 10 to 15 minutes to complete.

Personality Assessment Inventory. Another increasingly popular objective personality test is the Personality Assessment Inventory (PAI). This test consists of 344 items that are written at a fourth-grade reading level. This reading level ensures that most patients can complete it without experiencing any reading problems. The PAI takes about 45 to 50 minutes to complete for most patients. The PAI was normed on 1,000 community-dwelling individuals stratified according to gender, race, and age. There are not separate norms for male and female as there are in the MMPI. In addition, data were gathered on 1,246 clinical subjects and 1,051 college students in the normative process. The clinical subjects were drawn from a variety of different clinical settings, including inpatient psychiatric facilities (25%), outpatient psychiatric facilities (35%), correctional institutions (12%), medical settings (2%), and substance abuse treatment programs (15%).

Unlike the MMPI-2 where the patient is asked simply to answer whether a given item is true or false for them, the PAI allows for four alternative responses from the patient. The patient is asked if a given item for them is totally false, slightly true, mainly true, or very true.

Similarly to the MMPI-2, the PAI has four validity scales that measure the degree to which the patient is attempting to present themself in an overly positive or overly negative fashion. In other words, is the patient trying to feign psychopathology (fake bad) or underreport symptoms (fake good)? If the patient is attempting to fake bad or fake good, then the resulting PAI profile would of course likely be invalid and could not be interpreted.

The PAI has 11 clinical scales. These main clinical scales are similar to the MMPI-2 clinical scales and measure such personality issues as somatic concerns, depression, paranoia, borderline features, or alcohol or drug problems. The PAI also has five treatment-related scales that are designed to address such issues as treatment rejection, suicide ideation, or aggression. Two PAI interpersonal scales measure dominance and warmth.

Several strengths of the PAI are that the scales are nonoverlapping, the reading level is low, and four response choices to individual items are given. The MMPI-2 and PAI are not to be used as diagnostic instruments by themselves, but diagnosis determination needs to be based on a clinical interview and historical and medical information as well.

Rapid Personality Assessment Instruments. Rapid personality assessment instruments are narrow-spectrum self-report objective instruments that measure some unidimensional aspect of personality. The instrument is written in clear, simple language and provides face validity. Typically, these short instruments can be completed in less than 15 minutes. The scoring is simple and straightforward, usually just adding up the number of positive responses. On many instruments, there are cutting scores, above which an individual is considered to have a clinically significant problem in that area.

There are literally hundreds of such measures capable of assessing a wide variety of personality states and traits, such as geriatric depression, obsessive-compulsiveness, impulsivity, and state-trait anger. Table 7.5–3 presents common examples of such measures. The overall score typically provides an index of the magnitude of the patient's problems. Most of these tests can be used repeatedly to assess change over time. A number of instruments were developed using relatively small samples and have limited scientific studies documenting their usefulness in a variety of situations.

Some rapid assessment instruments were developed out of the behavioral assessment tradition. One advantage of short personality assessment procedures is high face validity for the domain of the behavior in question. Most instruments have standardized instructions and small normative groups. Performances of patients with the psychiatric problem in question (anger dyscontrol or geriatric depression) are frequently reported in terms of mean performance and standard deviation. One can find a rapid assessment instrument to assess many of the common psychiatric problems seen today. Another advantage of these instruments is the ease of administration and scoring. The major disadvantage is that these instruments differ greatly in their reliability and validity. Normative groups are frequently small and are obtained in one location. Most do not have national samples. Many of these short instruments have few, if any, studies of their construct validity. Most instruments do not systematically assess response styles such as faking bad, faking good, or social desirability responding. Examples of the hundreds of rapid assessments include measures of bulimia, obsessive-compulsive behavior, sexual functioning, fear and panic, and personal beliefs.

Projective Personality Tests

Projective personality tests, in contrast to objective personality instruments, are more indirect and unstructured. Unlike objective tests in which the patient may simply mark “true” or “false” to given questions, the variety of responses to projective personality tests are almost unlimited. Instructions are usually very general in nature, allowing the patient’s fantasies to be expressed. The patient generally does not know how their responses will be scored or analyzed. Consequently, trying to feign the test becomes difficult. Projective tests typically do not usually measure one particular personality characteristic such as “type A personality” (e.g., narrow-band measurement) but instead are designed to assess one’s personality as a whole (e.g., broad-band measurement).

Projective tests often focus on “latent” or unconscious aspects of personality. Obviously, psychologists and others differ in the degree to which they rely on “unconscious” information. In many projective techniques, the patient is simply shown a picture of something and asked to tell what the picture reminds them of. An underlying assumption of projective techniques (projective hypothesis) is that when presented with an ambiguous stimulus, such as an inkblot, for which there are an almost unlimited number of responses, the patient’s responses will reflect fundamental aspects of their personality. The ambiguous stimulus is a sort of screen on which the individual projects their own needs, thoughts, or conflicts. Different persons have different thoughts, needs, and conflicts and, hence, have widely different responses. The responses of a schizophrenic patient often reflect a rather bizarre, idiosyncratic view of the world.

Table 7.5–3.
Rapid Personality Instruments

Name	Description	Strengths	Weaknesses
Geriatric Depression Scale	30 items covering depression in the elderly; given orally or written; does not emphasize the vegetative signs of depression	Has reasonably good internal consistency reliability; good concurrent validity with other measures; can distinguish between normal, mildly depressed, and severely depressed patient	Geriatric patients frequently unwilling to admit depression because of denial; normative data limited
Beck Depression Inventory II	21 self-administered items for assessing depression in persons having at least a sixth-grade education; originally designed for use with trained interviewers, the test is self-administered; can assess the severity of depression; sensitive to changes due to psychotherapy and medication; measures patient	Large number of studies available in the literature establishes both its test–retest reliability and validity; widely used	No validity scales to assess approach to testing; is not specific to depression; patients with other disorders may score high on the test

	change weekly during therapy; assessed depression during specified period of 1 week; alpha coefficients were reported to be in the high 0.80s; effective in discriminating mild, moderate, and severe depression		
State Trait Anxiety Inventory	Anxiety is conceptualized as subjective tension and worry along with heightened arousal of the autonomic nervous system; state anxiety is the amount of anxiety a person has at a given moment in time; trait anxiety refers to the degree of anxiety a person usually experiences; there are 2 separate but similarly worded instruments, one measuring state anxiety (S Anxiety) the other measuring trait anxiety (T Anxiety); both consist of 20 questions	Very large normative sample of more than 5,000 in one study; inventory translated in many languages; excellent internal consistency reliability and concurrent validity	No validity scales; patients who score high on anxiety also frequently score high on depression and other psychological measure instruments
Michigan	Designed to screen for	Most frequently	Extensive

Alcoholism Screening Test (MAST-R)	alcoholism and problem drinking, presumably in people not manifesting suffering from these problems; the revised version has 25 items that can be completed in 10 minutes	used screening instrument for alcoholism; high internal consistency and good test–retest reliability; discriminates controls from other alcoholic problem groups (hospitalized alcoholics, driving under the influence offenders, persons convicted of drinking and disorderly conduct)	norms of population not available; may overdiagnose alcoholism in some cases and produced high false positive rates, particularly in college students; lack of candor on part of respondent can impact usefulness; high face validity makes it easy to fabricate responses
McGill Pain Questionnaire (MPQ)	The purpose of the MPQ is to assess quantitatively the qualitative characteristics of pain; instrument consists of 78 adjectives that the patient marks concerning the sensory, affective,	Used with patients suffering from all types of pain; can be used to assess effectiveness of various pain treatments	Adjectives can be hard for individuals with low educational background and/or low IQ to understand; pain complaints

	and evaluative nature of pain along with other characteristics; a pain rating index and pain intensity score is calculated; there is an MPQ-2 short version available, too (SF-MPQ-2)	(medication, biofeedback, relaxation exercise, surgery); has been used to discriminate between patients who have pain associated with objective physical findings and individuals who have pain without such findings	can be easily malingered

There are literally hundreds of different projective techniques developed—most of which are not widely used today. Table 7.5–4 contains a list of common projective tests together with a description of the tests. The strengths and weaknesses of the projective tests are also presented. This section will focus on more widely used projective instruments such as the Rorschach test, Thematic Apperception Test (TAT), and sentence completion.

The literature on projective techniques is vast, and the history of projective techniques goes back for many years. Arthur Carr pointed out in the fourth edition of the *Comprehensive Textbook of Psychiatry* that Shakespeare had an appreciation of the importance of one's association to ambiguous stimuli. In an effort to feign insanity, Hamlet associated to a cloud formation certain images that presumably reflected his personality.

Rorschach Test. Herman Rorschach, a Swiss psychiatrist, developed the first major use of projective techniques around 1910. The Rorschach test is the most frequently used projective personality instrument (Fig. 7.5–1). The test consists of 10 ambiguous symmetrical inkblots. The inkblot card appears as if a blot of ink was poured onto a piece of paper and folded over—hence, the symmetrical appearance. These 6½- × 9½-inch inkblot cards are the standard stimuli and, by convention, are referred to as Roman numerals I to X.

Table 7.5–4.
Projective Measures of Personality

Name	Description	Strengths	Weaknesses
Rorschach test	10 stimulus cards of inkblots, some colored, others achromatic	Most widely used projective device and certainly the best researched; considerable interpretative data available	Some Rorschach interpretive systems have unproven validity
Thematic Apperception Test (TAT)	20 stimulus cards depicting a number of scenes of varying ambiguity	A widely used method that, in the hands of a well-trained person, provides valuable information	No generally accepted scoring system results in poor consistency in interpretation; time-consuming administration
Sentence completion test	A number of different devices available, all sharing the same format with more similarities than differences	Brief administration time; can be a useful adjunct to clinical interviews if supplied beforehand	Stimuli are obvious as to intent and subject to easy falsification
Holtzman Inkblot Technique (HIT)	Two parallel forms of inkblot cards with 45 cards per form	Only one response is allowed per card, making research less troublesome	Not widely accepted and rarely used; not directly comparable to Rorschach interpretive strategies
Figure drawing	Typically, human	Quick	Interpretive

	forms but can involve houses or other forms	administration	strategies have typically been unsupported by research
Make-a-Picture Story (MAPS)	Similar to TAT; however, stimuli can be manipulated by the patient	Provides idiographic personality information through thematic analysis	Minimal research support; not widely used

Courtesy of Robert W. Butler, Ph.D., and Paul Satz, Ph.D.

Using John E. Exner's Comprehensive System, the examiner, in essence, gives the patient the following instructions: "You'll be given a series of inkblots. Look at each inkblot and tell me what you see." The examiner gives the patient each card and asks, "What might this be?"

Minimal interaction between the examiner and the patient occurs while the Rorschach is administered, which ensures standardization procedures are upheld. The examiner writes down verbatim what the patient says during the above-described "free association" or "response proper" phase. If the patient rotates the card during their response, then the examiner makes the appropriate notation on the test protocol. After the patient has given responses to all 10 cards, an inquiry phase of administration begins. The examiner asks the patient to go through the cards again and help the examiner see the responses they gave. The examiner reads the patient's initial response and asks the patient to point out what they saw and explain what made it look like that to them. An almost unlimited range of responses is possible with the Rorschach test and most projective tests. Six different scoring and interpretation systems for the Rorschach test are available for the clinician; however, Exner's Comprehensive System is the most widely used and accepted. A structural summary of the various measures derived from the patient's response is usually computed.

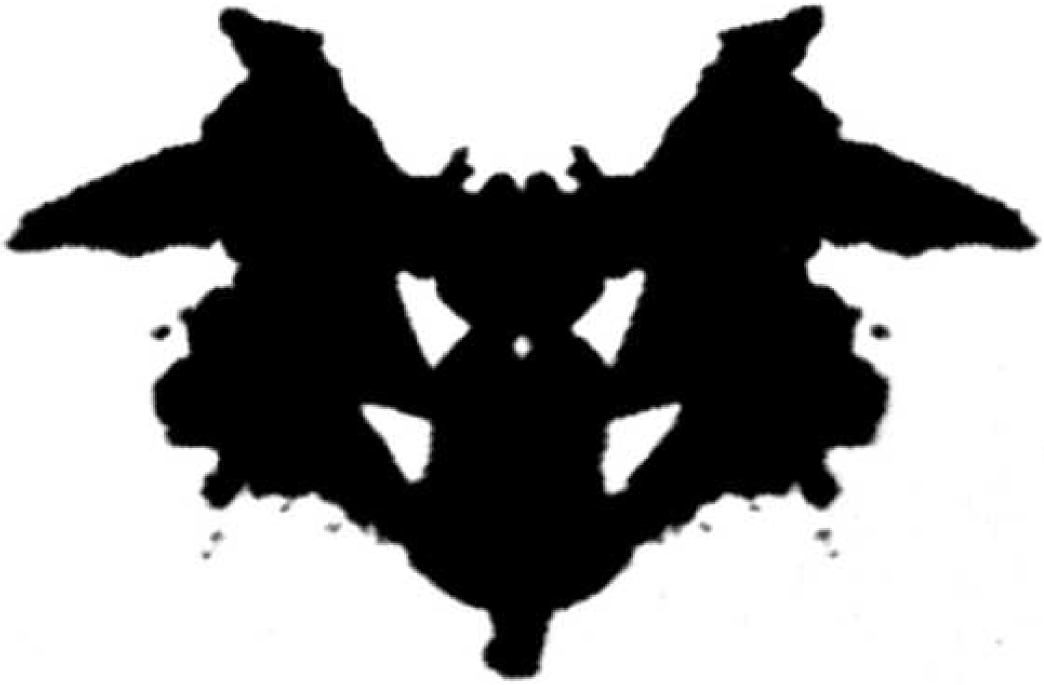

FIGURE 7.5–1. Card I of the Rorschach test. (From Hermann Rorschach, Rorschach®-Test. Copyright © Verlag Hans Huber, Hogrefe Verlag Bern, Switzerland, 1921, 1948, 1994, 2017, with permission.)

Scoring and interpretation of the Rorschach test can be a time-consuming process. Administering and scoring the Rorschach test usually takes about 1½ to 2 hours of time. Typically, examiners score the Rorschach responses according to the *location* on the blot where the patient sees the response. The factors that make patients respond the way that they did are referred to as the *determinants* (shape of blot, color, and shading). The examiner also scores the actual *content* of the patient's response (e.g., animals or human anatomy). The scoring system is elaborate and can result in dozens of different indices being computed. Although some investigators question the reliability of certain measures, many studies do support the usefulness of the Rorschach test as a personality assessment instrument in the hands of a qualified and experienced psychologist.

The limited scientific support of the Rorschach test, however, has resulted in critical reviews by some psychologists. A number of graduate schools of clinical psychology no longer offer formal coursework in the Rorschach test, although many still do.

Table 7.5–5 gives the sample Rorschach responses of a number of patients to Rorschach card I. Notice how different the responses are by various patients to this same card and how the responses reflect the patient's underlying personality.

Table 7.5–6 shows certain cluster variables derived from Exner's Comprehensive Scoring System. The interpretive utility of these cluster variables is also described in the table.

Thematic Apperception Test. Although the Rorschach test is clearly the most frequently used projective personality test, the TAT is probably in second place. Many clinicians will include both the TAT and the Rorschach test in a battery of tests for personality assessment. The TAT consists of a series of 10 black-and-white pictures that depict individuals of both sexes and of different age groups, who are involved in a variety of different activities. For example, on Card 1, a young boy is shown sitting at a table, looking at a violin. Card 2 depicts a farm scene in which a young woman is in the foreground carrying books in her hands. A man is working in the fields nearby, and an older woman is seen in the background. Typically, a patient is shown 10 TAT cards and is asked to make up stories about them. The patient is asked to tell what is going on in the picture, what is going on before the picture was taken, what the individuals in the picture are thinking and feeling, and what is likely to happen in the future.

Table 7.5–5.

Sample Rorschach Responses and Interpretation to Card 1 by Diagnostic Category

Diagnostic Category	**Response Proper**	**Inquiry**	**Interpretation**
Nonpatient	Well, the first impression is of a wolf's head, ears here and eyes.	Overall, the head, nose, eyes, ears, the whole thing.	Common response utilizing the white space as the eyes, responding to the whole blot.
"Schizo" patient	I will say the first thing that comes to mind, it would be a predator on the movie, loss of ink, predator, have you seen that movie?	The whole picture all the way, and I liked the movie, and the reason why is because you said the first thing that comes to mind. I've seen the movie over and over again, ten times. The whole picture, something that looks at me mean (Looks at you mean?) Because it's a person that kills without any feeling, that's why I thought the movie was good.	The response shows difficulty staying focused on the task at hand, tangential thought process, some paranoia and distortion of the form of the card.
Depression	Could be a leaf.	Could be a leaf that has fallen off the tree and started to decay, leafs have jagged edges (You said decayed?) Because if it was a perfect leaf all this would be part of the blot, and here parts of it that have fallen off. That would have been the stem. Also, you have the main vein running down the middle, this is more	The response shows morbid content (decay) that indicates a poor view of self and the world. Leaf responses also weigh on the isolation index.

		prevalent because it’s biggest, if it was on a tree, it would get nutrients to the leaf like our blood system.	
Anxiety	I don’t know, I’d have to think about it. It could be trees.	Like if you are in the forest, stalks you know, when trees bland together they are nondescript. I’m a real outdoor person, I like to go to the forest and spend time there. (Can you help me see the trees?) Just this little Christmas tree here and here, the outline.	The response shows discomfort in committing to a response, and discomfort in beginning the task. Some anxiety is also seen in the rambling personalization used to justify the response.

Courtesy of Dana Stelter, Ph.D., Clinical Psychologist.

Henry Murray developed the TAT in 1943 at the Harvard Psychological Clinic. The stories that the patient makes up concerning the pictures, according to the projective hypothesis, reflect the patient’s own needs, thoughts, feelings, stresses, wishes, desires, and view of the future. According to the theory underlying the test, a patient identifies with a particular individual in the picture. This individual is called the *hero*. The hero is usually close to the age of the patient and frequently of the same sex, although not necessarily so. Theoretically, the patient would attribute their own needs, thoughts, and feelings to this hero. The forces present in the hero’s environment represent the *press* of the story, and the *outcome* is the resolution of the interaction between the hero’s needs and desires and the press of the environment. The TAT, in the hands of a skilled interpreter, can reveal considerable information about the patient’s personality that can then be combined with other information obtained from the interview and other testing.

Table 7.5–6.
Cluster Variables (Exner System) and Basic Interpretive Utility

Cluster Variable	Interpretive Value
Control Stress Tolerance	Can assess a person's stress tolerance, ability to control their behavior, and responses in the environment.
Situation-Related Stress	Can assess the amount of situational stress affecting a person's capacity for control.
Affective Features	Can assess how a person manages emotional stimuli, types of emotional features such as depression and anger, and ability to modulate their affect.
Self-Perception	Can assess how a person views themselves and ability to look at themselves.
Interpersonal Perception	Can assess how a person views others, ways they interact with others, and how well they identify with others.
Information Processing	Can assess how a person organizes information in their world and how they process it.
Cognitive Mediation	Can assess the conventionality of a person's perceptions and degree of reality distortions.
Ideation	Can assess aspects of ideation such as processing style, use of defenses such as fantasy and intellectualization, and if their thoughts proceed in logical, coherent fashion or have some thought slippage.

Courtesy of Dana Stelter, Ph.D., Clinical Psychologist.

By itself, the TAT is not an effective diagnostic instrument, nor was it designed to be. Moreover, as measured by traditional methods of assessing reliability and validity, the TAT lacks the scientific merit of objective instruments. The fact that it has remained reasonably popular today, however, speaks to the instrument's usefulness in clinical practice.

Sentence Completion Test. Although a projective instrument, the sentence completion test is much more direct in soliciting responses from the patient. They are simply presented with a series of incomplete sentences and is asked to complete the sentence with the first response that comes to mind. The following are examples of possible incomplete sentences:

My father seldom …
Most people don’t know that I’m afraid of …
When I was a child, I …
When encountering frustration, I usually…

The purpose of the test is to elicit, in a somewhat indirect manner, information about the patient that cannot be elicited from other measures. Because the patient responds in writing, the examiner’s time is limited. The length of time it takes to complete the sentence completion varies greatly depending on the number of incomplete sentences. Tests can range from less than 10 sentences to greater than 75.

There are many variations of sentence completion tests. Some clinicians have developed their own tests. One form of the test, developed by Julian Rotter, has some established validity and reliability, but most sentence completion tests do not. Special purpose sentence completion tests have been developed to measure different problem areas. For example, there is a sentence completion test that is used with patients who have chronic pain. The sentence completion test is seldom, if ever, used alone but is combined with other appropriate instruments.

Advantages of the sentence completion test are short administration time, ease of administration, variety of instruments, and ease of construction. Disadvantages are lack of reliability and validity studies and ease of fabrication and deception.

Behavioral Assessment

Behavioral assessment involves the direct measurement of a given behavior. Rather than focus primarily on human characteristics, such as repression, ego strength, or self-esteem (vague terms to a behaviorist), strict behavioral measurement concentrates on the direct measurement that can be observed, such as a number of temper tantrums per unit of time, duration and intensity and number of hyperventilation episodes, or the number of cigarettes smoked per 24-hour period.

Although early strict behaviorists would count only behaviors that were observable, a broader definition of behavior has emerged, under which just about anything people do—whether it is overt such as crying, swearing, or handwashing or covert such as feeling and thinking—is considered behavior. To measure covert behavior, the patient is typically the informant. For example, to measure dysfunctional thinking, the patient may simply carry a mechanical counter and press it every time they have a particular target self-disparaging thought during a given unit of time.

Direct Behavioral Measurement. The patient, a family member, or an impartial observer can measure behavior directly. Therapists use direct behavioral measurements to establish baselines of a given undesirable behavior (i.e., violent thoughts that the patient may wish to reduce). Similarly, therapists can measure behavior that the patient wants to increase (time studying or distance walked on a treadmill). Follow-up measures of the same behavior monitor progress and quantify improvement.

Different aspects of a target behavior can be counted or measured, such as frequency of behavior (e.g., number of behavioral occurrences), duration of behavior (length of time behavior occurred), and intensity. Event recording is a behavior measurement technique in which an individual counts instances of behavior. This approach is best used for discrete behaviors that have a beginning and end. For example, one can count the number of instances of nail-biting, the number of times out of bed, the number of occasions one arrives late to work, or the number of dysfunctional thoughts they have concerning an employer during a given 18-hour period.

Interval Recording is another type of behavioral measurement technique that can be used to measure how often a behavior occurs. Sometimes it is not feasible to count each instance of behavior as the behavior may be too frequent. In addition, certain behaviors may not be discrete (with a beginning and end) such as the duration of time an individual stays on task on a work project.

There are three types of interval recording. In momentary time sampling, one records whether or not the behavior occurred at a particular time. For example, every 15 minutes, one records whether or not the behavior is occurring at that particular moment. Momentary time sampling is best use for very high-frequency behaviors. The second type of interval recording is called whole interval recording. In whole interval recording, one divides a time segment into many intervals (e.g., 30 intervals consisting of 15 seconds for each interval). An interval is scored only if the behavior occurred during the entire interval. Whole interval recording is used for behaviors that one wishes to increase, but it tends to underestimate the duration of the behavior. Partial interval recording is set up as the same as whole interval recording, but the interval is scored if the behavior occurs anytime in the interval. This type of interval recording is best used for behaviors to decrease, and this approach tends to overestimate the occurrence of the behavior. Whole and partial interval recording will provide one with the percentage of intervals a behavior occurs (e.g., 15% of intervals).

The advantage of using event recording and interval recording is that it is direct, objective, and easily understood and quantifies the degree of behavioral change. The problem with counting specific behaviors is that many patients have so many different symptoms it would not be possible to quantify them all. Sometimes, the process of counting a target behavior may itself alter the behavior. It can also be used to help increase awareness of behavioral patterns.

Rating Scales

Rating scales have been developed to detect various aspects of personality. On some of these scales, the patient simply rates themself on a given scale dimension such as sadness, 1 = I am not sad, or 7 = I wish I were dead. On other scales, a significant other person is asked to rate the patient, 1 = the individual seldom mentions being sad, or 7 = the individual says they would be better off dead. Some of the more commonly used rating scales, however, are designed to quantify information obtained by the mental health professional during a mental status interview.

Brief Psychiatric Rating Scale. One of the more commonly used rating scales is a Brief Psychiatric Rating Scale (BPRS). To use the BPRS, the psychologist or psychiatrist simply completes a mental status interview with the patient. After completing the interview, the interviewer rates the patient on a series of 18 psychiatric symptoms, such as motor retardation, blunted affect, conceptual disorganization, anxiety, and guilt.

Expanded definitions of each of these terms are provided to the examiner. The interviewer rates each domain on a seven-point Likert scale from “not present,” the lowest rating, to “extremely severe,” the highest rating. An experienced interviewer can complete the ratings in 2 or 3 minutes. The BPRS has been used extensively in drug outcome and other studies. The advantage of the BPRS is that the interrater reliability is reasonably high for a rating scale of this nature. A summary article of over 300 studies utilizing the BPRS found interrater reliability correlative of 0.80 or higher on the total score in the majority of studies. The first author has used this instrument to teach medical students to be observant of a patient’s behavior, and relatively inexperienced medical students can establish reasonably good reliability. Other advantages of the BPRS include the ease and speed of rating and well-defined symptom description.

One disadvantage of this scale is examiner subjectivity. Another disadvantage is that the interviewer can only rate the patient on what they personally observes during the interview. Examiners differ greatly in their style of conducting a mental status interview, and thus, one examiner might ask many questions on interviews in one given area (i.e., somatic concern), while nearly overlooking another area (i.e., suspiciousness). Another examiner might do just the opposite.

Integrated Psychological Assessment to Answer Specific Referral Questions

Normally in the practice of medicine, a physician evaluates a patient’s chief complaints, signs, and symptoms. On the basis of this information, the physician may suspect a given medical problem. In order to further investigate, the physician may order a series of lab tests or other studies to further delineate the problem at hand. For example, a physician may suspect that a patient has a dysfunctional liver and consequently orders liver function tests. Test results are reported back to the physician in terms of quantitative information that helps the treating physician to better understand the nature of the patient’s problems.

In contrast, in requesting a psychological assessment, it is normally better for the psychiatrist or other health professional to ask the psychologist specific referral questions rather than specify particular tests to be administered. Below are listed examples of referral questions:

1. What is the patient’s DSM-5 diagnosis?

2. Is the patient psychotic?
3. Does this patient have schizophrenia or bipolar disorder?
4. Does this patient have an attention deficit disorder or are their presenting problems due to anxiety?
5. Does the patient have the cognitive and personality resources to assist in their own defense in a criminal case?
6. Is this patient feigning a psychiatric disorder or do they, in fact, have a genuine disorder?

In order to answer the referral question, the psychologist will determine which tests are most appropriate given the nature of the referral question(s). There are literally hundreds of different psychological tests that may be utilized to answer various referral questions. After conducting a biopsychosocial assessment, a mental status evaluation, and personality testing, and perhaps an interview with a significant other person in the patient’s life, a psychologist will integrate this information in a comprehensive psychological assessment report in order to answer the psychiatrist’s specific referral question. In this case, the goal is not simply to report back to the psychiatrist a series of isolated scores, but rather the goal is to interpret the scores in relationship to a referral question that the psychiatrist posed. By using this approach, the psychiatrist is most likely to have their referral question answered. In order to answer specific referral questions, the psychologist may utilize personality tests, but they may also utilize cognitive, achievement, and narrowband tests. In some cases, it may be necessary to order specialized neuropsychological testing in order to answer the referral question. When a neuropsychological test is required, the evaluation is normally completed by a neuropsychologist who has special training and experience in brain–behavior relationships.

Psychologists will often have the patient complete a comprehensive battery consisting of different tests to answer the referral question. For example, based on a specific referral question concerning a psychiatric diagnosis, a psychologist may choose a battery of tests that include the MMPI-2, MCMI-3, and the Beck Depression Inventory-2. The psychologist will write a comprehensive report that presents the findings of psychological testing and a summary and recommendations section that addresses the specific referral question.

FURTHER READINGS

American Educational Research Association. *Standards for Educational and Psychological Testing*. AERA; 2014.

Archer RP, Krishnamurthy R. *Essentials of MMPI-A Assessment*. John Wiley; 2001.

Ben-Porath YS. *Interpreting the MMPI-2-RF*. University of Minnesota Press; 2012.

Bornstein RF. Supervising personality assessment. In: Wright AJ, ed. *Essentials of Psychological Assessment Supervision*. John Wiley; 2019.

Butcher JN. *MMPI-2: A Practitioner's Guide*. American Psychological Association; 2005.

*Carducci BJ, Nave CS. *The Wiley Encyclopedia of Personality and Individual Differences: Volume II Measurement and Assessment*. John Wiley & Sons; 2020.

Conway CC, Krueger RF. Personality disorders. In: Anthony MM, Barlow DH, eds. *Handbook of Assessment and Treatment Planning for Psychological Disorders*. 3rd ed. Guilford Press; 2020.

Cooper C. *Individual Differences and Personality*. 4th ed. Routledge; 2021.

Friedman AF, Bolinskey KP, Levak RW, Nichols DF. *Psychological Assessment with the MMPI-2/MMPI-2-RF*. 3rd ed. Routledge; 2014.

Graham JR. *MMPI-2: Assessing Personality and Psychopathology*. 5th ed. Oxford University Press; 2011.

Gregory RJ. *Psychological Testing: History, Principles and Application*. 7th ed. Pearson; 2016.

*Groth-Marnat G, Wright AJ. *Handbook of Psychological Assessment. Hoboken*. 6th ed. John Wiley & Sons; 2016.

Harwood TM, Beutler LE, Groth-Marnat G. *Integrative Assessment of Adult Personality*. 3rd ed. Guilford Press; 2012.

Marvish ME. *Handbook of Psychological Assessment in Primary Care Settings*. 2nd ed. Routledge; 2017.

Megargee EI, Spielberger CD. *Personality Assessment in America: A Retrospective on the Occasion of the Fiftieth Anniversary of the Society for Personality Assessment*. Routledge; 2019.

Melton GB, Petrila J, Poythress NG, et al. *Psychological Evaluation for the Courts: A Handbook for Mental Health Professionals and Lawyers*. 4th ed. Guilford Publications; 2017.

Miller LA, Louler RL. *Foundations of Psychological Testing: A Practical Approach*. 6th ed. Sage; 2020.

Paniagua FA, Yamada AM. *Handbook of Multicultural Mental Health: Assessment and Treatment of Diverse Populations*. 2nd ed. Academic Press; 2013.

Rush AJ Jr, First MB, Blacker D. *Handbook of Psychiatric Measure*. 2nd ed. American Psychiatric Publishing; 2008.

Smith SR, Krishnamurthy R. *Diversity-Sensitive Personality Assessment*. Routledge; 2018.

Verdon B, Azoulay C. *Psychoanalysis and Projective Methods in Personality Assessment*. Hogrefe; 2020.

*Weiner IB, Greene RL. *Handbook of Personality Assessment*. 2nd ed. John Wiley; 2017.

Wright AJ. *Conducting Psychological Assessment: A Guide for Practitioners*. Wiley; 2021.

Wright AJ, Raiford SE. *Essentials of Psychological Tele-Assessment*. John Wiley; 2021.

7.6 Medical Assessment and Laboratory Testing in Psychiatry

JENNIFER L. KRUSE, M.D., AND JOHN O. BROOKS, III, PH.D., M.D.

The goal of this section is to outline laboratory tests, neuroimaging studies, and other assessments applicable to psychiatric practice and the strategies for using them in an effective manner. The role of imaging studies, the electroencephalogram (EEG), various biochemical, hematologic, endocrine, immunologic, and toxicologic assessments are discussed. A logical and systematic approach to the use of these tools by the psychiatrist is vital to achieving the goals of arriving at accurate diagnoses, identifying medical comorbidities, implementing appropriate treatment, and delivering cost-effective care. With respect to the diagnosis or management of medical disease, consultation with colleagues in other specialties is important. Medical specialists also are a useful source of information about test methodology and the proper performance and interpretation of tests. Analogous consultation with appropriate specialists is important when ordering nonlaboratory tests such as the EEG, electrocardiogram (ECG), and neuroimaging.

Medical education often accustoms physicians to ordering tests, and psychiatrists are no exception. Laboratory testing, in particular, is routinely ordered before most psychiatric hospital admissions or transfers, for example. Yet, the result of any test should have the potential to alter patient management. If the test results will not alter management, then the test is unlikely to be relevant to the patient's current care. Despite increasing pressure to deliver cost-effective care, many clinicians still order a large number of screening laboratory and other studies. In reality, ordering "routine" screening laboratory tests very rarely detect abnormalities of any clinical significance. At the same time, in the outpatient setting, testing may not be ordered as frequently as indicated, especially for patients without closer medical outpatient follow-up.

At the time of admission to an inpatient unit, the lack of consensus regarding optimal patient workup may contribute to a tendency to order unnecessary laboratory tests. Aside from underserved populations, there is little evidence that broad laboratory tests improve patient care. Rather, choosing tests based on clinical hypotheses derived from a particular patient's clinical data, thereby minimizing the ordering of unnecessary studies and tests, constitutes a more reasonable approach. Such an approach often uses a narrow selection of routine screening tests. Subsequently, clinicians order nonroutine, additional tests as necessary based on the results of screening tests or particular findings on history or examination. The best screening evaluation includes a good history and physical examination, including a neurologic examination and mental status examination.

Psychiatrists use diagnostic testing to identify possible medical or neurologic illnesses that may cause or contribute to a patient's clinical presentation. Such testing occurs within the context of the history of the present illness, the physical examination, and the review of systems. Other tests are specific to psychiatry. For example, the measurement of blood levels of psychotropic medications can be useful to determine the adequacy of dosing or to monitor compliance. In addition, many tests, such as liver function tests (LFTs), white blood cell (WBC) count, and serum glucose, are performed to monitor the consequences of treatment of psychiatric illness. Laboratory detection of substances of abuse is helpful to identify patients who abuse illicit substances. For patients known to abuse substances, additional testing may be performed to assess the consequences of a patient's substance abuse.

In an outpatient setting, the clinical workup may differ because patients generally present with less acute symptoms and are less likely to have emergent conditions. Generally, a physical examination is impractical in the outpatient setting and, if needed, is often obtained through a patient's primary care provider. Patients who present for outpatient psychiatric care may or may not need initial laboratory testing, but if they do, focused testing should be performed based on suspected intersecting conditions. Neuroimaging studies may be underutilized in the outpatient psychiatric setting but should be ordered through specialist care if it is not possible for a psychiatrist to order them directly.

Many medical illnesses produce psychiatric symptoms or mimic psychiatric illness. Presentations with delirium, anxiety, depression, and hallucinations are nonspecific. Other medical and neurologic causes must first be excluded to be certain that a patient's symptoms are due to a mental illness. This assessment is not performed in a vacuum but occurs in the context of the patient's medical history and physical examination, and laboratory studies supplement the history and examination. The goal of the assessment is to identify causative factors that contribute to the patient's current symptoms. When psychiatric symptoms are found to be due to a general medical condition, initial efforts are focused on the treatment of the underlying condition. For many patients, treatment of the underlying medical condition results in a resolution of psychiatric symptoms. In addition, the diagnostic criteria for many psychiatric conditions, as listed in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), exclude symptoms attributable to a general medical condition or substance abuse.

Aside from using laboratory testing for diagnostic purposes, psychiatrists increasingly use laboratory testing to establish that administration of a particular treatment is safe, both before its initiation and to monitor ongoing adverse effects. For example, baseline laboratory studies may be obtained before the initiation of particular psychotropic medications. Testing also may be used to monitor compliance, measure therapeutic concentrations, and assess toxicity. Many psychotropic medications may have significant effects on physical health. Consequently, a psychiatrist may perform a pretreatment assessment of a patient's general health before initiating treatment. This assessment commonly includes a variety of laboratory tests.

Finally, there is continuing work toward biologic markers, or biomarkers, of major mental illnesses that will assist in the diagnosis and personalization of treatment. Unfortunately, research has not yet identified biomarkers that are ready for use in routine clinical practice.

PHYSICAL HEALTH MONITORING

Monitoring the physical health of psychiatric patients has two goals: To provide appropriate care for existing illnesses and to protect the patient's current health from possible future impairment. Disease prevention should begin with a clear concept of the condition to be avoided. Ideally, in psychiatry, this would focus on common conditions that could be a significant source of morbidity or mortality. It is clear that in psychiatry, a small number of clinical problems underlie a significant number of impairments and premature deaths.

Many psychiatric patients, especially in lower socioeconomic brackets, fail to receive periodic health maintenance examinations. Examinations and laboratory testing at the time of admission to a psychiatric hospital may be a rare opportunity to deliver preventive care. Preventive care can include monitoring for hypertension or providing counseling about healthy eating or smoking cessation.

Lifestyle-related health problems such as obesity, diabetes, hypertension, cardiovascular disease, and sexually transmitted diseases (STDs) are frequent comorbidities in some psychiatric populations. It is helpful to discuss proper nutrition, diet, and exercise as part of the prevention and management of excessive weight gain. The physician also should discuss information about preventable diseases that are more likely to occur in obese individuals and the role of monitoring fasting blood sugars and fasting lipid levels in disease prevention. Sexually transmitted infections are more common in young psychiatric patients than in the general population. In patients with risk factors for human immunodeficiency virus (HIV) infection, it is important to order screening tests for the virus (though screening may require a patient’s permission).

Illnesses such as cardiovascular disease are often associated with aging, but this is not necessarily the case in people with mental illness. For example, among patients with bipolar disorder, comorbid cardiovascular disease leads to greater mortality than suicide and occurs at a younger age than in nonmentally ill counterparts. Patients with risk factors for metabolic syndromes, such as obesity, should be screened for hypertension, dyslipidemia, and glucose levels. All patients who use tobacco should be counseled about smoking cessation with the goal of preventing coronary artery disease, stroke, and other medical comorbidities.

Detecting medical complications not only extends lifespan but also lessens the impact of mental illness and thereby improves quality of life. Treatment of medical conditions lessens the severity of symptoms of affective and psychotic disorders. Early treatment of medical conditions may reduce the likelihood of medical hospitalization, which leads to an increased risk of rehospitalization for self-harm and suicidal acts after discharge for patients with serious mental illness.

DIAGNOSTIC ASSESSMENT

This section reviews studies that may be useful in the diagnostic assessment of patients presenting with psychiatric complaints. Clinicians should perform an examination that is individualized to the specific needs of the patient. There is no consensus about standard laboratory tests for psychiatric patients, and standards of care differ between outpatient and inpatient settings.

In the outpatient setting, laboratory testing is relatively uncommon unless a medical illness is suspected of generating a psychiatric symptom or collection of symptoms. Such suspicions may be supported by a patient's history or physical finding and may indicate laboratory testing, neuroimaging, and follow-up by specialists to treat an underlying disorder. In certain situations, psychiatric care may be beneficial in symptomatic treatment alongside the primary treatment.

When patients are admitted to an inpatient psychiatric unit, many times "screening" laboratory studies are ordered, although there is no evidence that supports the role of laboratory studies in "medical clearance" for admission to psychiatric units. Certainly, such studies are indicated in patients who have a history or suspected history of conditions that may affect hepatic or renal function, but for younger, healthy patients, general laboratory studies do little to advance patient care. Rather, specific studies can be ordered as subsequent information is obtained.

For patients with a suspected or established history of alcohol use or hepatic impairment, LFTs including alanine aminotransferase (ALT) and aspartate aminotransferase (AST) to assess their ability to metabolize prescribed medication are helpful. Patients who have a history of lithium use or suspected renal impairment should have an assessment of renal function, including blood urea nitrogen (BUN) and creatinine (Cr), and a urinalysis to evaluate their capacity to excrete metabolized medication. In patients with a history of treatment-refractory depression or appropriate physical findings, thyroid-stimulating hormone (TSH) may be considered.

Role of History and Physical Examination

A thorough history, including a review of systems, is the basis for a comprehensive patient assessment. The history guides the clinician in the selection of laboratory studies that are relevant for a specific patient. Some psychiatric patients, because of the nature of their illnesses, are not capable of providing sufficiently detailed information. Collateral sources of information, including family members and prior clinicians and their medical records, may be particularly helpful in the assessment of such patients.

The history of present illness includes the chief complaint (or the reason the patient seeks or was brought for assistance), the nature of the current problems, and their current treatments. The past psychiatric history should include information about previous psychiatric problems and a history of prior treatments. It should contain the medications used to treat the patient, the doses prescribed, and the durations of each medication trial. The past psychiatric history should also contain documentation of dates and locations of past psychiatric admissions, the reason for admission, and the treatments provided. If the patient reports an allergy to a specific medication, then the nature of the allergic response should be noted.

The patient’s medical history is an important component of the history. It should include notation of prior injuries and, in particular, head injuries that resulted in the loss of consciousness and other causes of unconsciousness. The patient’s medical history also should note pain conditions, ongoing medical problems, prior hospitalizations, prior surgeries, and a list of the patient’s current medications. Toxic exposures are another important component of medical history. Such exposures are often workplace related.

The social history contains many of the details relevant to the assessment of character pathology, including risk factors for personality disorders, as well as information relevant to the assessment of other psychiatric disorders. Commonly, the social history includes legal history, information about family and other significant relationships, and educational and occupational history. The clinician should take a thorough history of alcohol and substance use. History of the use of substances, illicit and prescribed, should contain information regarding the timing of the initiation of use, the frequency of use, the quantity used, and the date of most recent use for each substance. A thorough review of symptoms should be performed. Unrecognized medical conditions are often detected during the review of systems.

The frequency with which psychiatrists perform physical examinations on their patients is highly variable; it is common for other physicians to perform the physical examination in acute settings before or after the psychiatrist obtains the history. Under certain clinical scenarios, it may be appropriate to postpone the physical examination, such as when a patient is agitated or aggressive. Such deferments should be for as brief a time as possible. As a general rule, it is wise to have a chaperone present during the physical examination.

Physical examinations may provide data helpful in the diagnosis and management of substance abuse. For example, intravenous (IV) drug use may be identified by the presence of “track” marks—small scars following the outline of a vein produced from repeated injections. In chronic users, track marks may be less apparent and hyperpigmented, fibrotic, depressed, round papules on the forearms or legs may be apparent from “skin popping” (subcutaneous injection). Perforation of the nasal septum or chronic nosebleeds may occur following chronic insufflation of cocaine. Opioid intoxication is associated with sedation, dysarthria, slowed cognition, and pinpoint pupils. In contrast, withdrawal from opiates is associated with lacrimation, rhinorrhea, yawning, piloerection, myalgias, pupillary dilation (mydriasis), and bowel discomfort. These symptoms occur 1 to 2 days following discontinuation of methadone and 6 to 12 hours after stopping shorter-acting opiates. Alcohol or sedative-hypnotic withdrawal is associated with autonomic hyperactivity, nausea, tremulousness, diaphoresis, headache, anxiety, and agitation, as well as perceptual disturbances and clouding of sensorium in severe cases.

When evaluating patients with cognitive impairment, the physical examination may also provide etiologic clues when accompanying neurologic signs are identified, such as cogwheel rigidity and tremor associated with Parkinson disease or neurologic deficits suggestive of prior strokes. Standard laboratory studies commonly assessed in patients with neurocognitive disorders include a complete blood count (CBC), serum electrolytes, LFTs, BUN, Cr, thyroid function tests, serum $B_{12,}$ and folate levels Venereal Disease Research Laboratory (VDRL) test, and a urinalysis. Structural neuroimaging studies are often performed, particularly if there are focal neurologic findings. EEG and/or cerebrospinal fluid (CSF) evaluation may be performed in some cases if seizures, central nervous system (CNS) infection, and/or other etiologies requiring such evaluations are suspected. When patients are delirious, the neurologic examination may be complicated by inattention and/or altered level of consciousness. Delirium workup may include a similar laboratory workup, but the appropriate evaluation for a particular delirious patient depends upon historical and clinical features. Urine or blood cultures, chest radiograph, arterial blood gases, chemistry panel, endocrine evaluation, neuroimaging studies, or EEG may be appropriate, among other studies.

As is true for most presenting symptoms, the evaluation of neuropsychiatric symptoms requires consideration of a broad differential diagnosis; the differential diagnosis for a neuropsychiatric presentation may include primary psychiatric disorders, neurologic disorders, or other medical conditions. An important part of the medical and laboratory assessment of psychiatric symptoms is to determine which medical conditions to include in the differential diagnosis based upon the patient's history and examination findings. Further, comorbid medical conditions may complicate the management of psychiatric patients, and laboratory evaluation for a variety of conditions may be important in the course of clinical care, often in collaboration with other specialists.

Serum Electrolyte Evaluation

Serum electrolyte levels may be useful in the initial evaluation of a psychiatric patient, especially if he or she appears obtunded or cognitively slowed for unclear reasons. Older adults who exhibit a more rapid onset of decreased cognitive function or depression may have electrolyte abnormalities. Abnormalities also may occur in response to the administration of psychotropic medications.

Hypokalemia may be present in eating disorder patients who purge or abuse laxatives and in psychogenic vomiting. Diuretic abuse by eating disorder patients also may produce hypokalemia. Low levels of potassium are associated with weakness and fatigue. Characteristic ECG changes occur with hypokalemia and consist of cardiac arrhythmias, U waves, flattened T waves, and ST-segment depression.

Eating disorder patients with anorexia nervosa or bulimia nervosa usually receive a fairly standard set of laboratory studies, including serum electrolytes (particularly potassium and phosphorus), blood glucose, thyroid function tests, liver enzymes, total protein, serum albumin, BUN, Cr, CBC, and ECG. Serum amylase is often assessed in bulimic patients.

Magnesium levels may be low in patients with alcohol use disorders or malnutrition. Low magnesium levels are associated with agitation, confusion, and delirium. If untreated, then convulsions and coma may follow. Additionally, hypomagnesemia increases the risk for arrhythmias such as torsade de pointes.

Low levels of serum phosphorus may be present in eating disorder patients with purging behavior. Phosphorus levels may also be low in anxiety patients who hyperventilate. Hyperparathyroidism may produce low serum phosphorus levels. Elevated serum phosphorus levels are seen in hypoparathyroidism.

Hyponatremia is seen in psychogenic polydipsia and SIADH and in response to certain medications, such as oxcarbazepine and carbamazepine. In older adults, low sodium levels may result from physiologic or dietary alterations and be associated with cognitive slowing or delirium and could be mistaken for dementia or depression. Low serum chloride levels may occur in eating disorder patients who purge by self-induced vomiting.

Serum calcium abnormalities are associated with a variety of behavioral abnormalities. Low serum calcium levels are associated with depression, delirium, and irritability. Elevated levels are associated with depression, psychosis, and weakness. Laxative abuse, common in anorexia binge–purge type, can be associated with hypocalcemia. Hypocalcemia secondary to hypoparathyroidism may occur in patients who have undergone surgery for thyroid disease. Hypercalcemia can be seen in association with hyperparathyroidism, which is sometimes a side effect of treatment with lithium.

Renal Function

Tests of renal function, including BUN and Cr, are serum markers used to estimate glomerular filtration rate (GFR). Other relevant laboratory studies include routine urinalysis and Cr clearance. Renal failure may result in lethargy or delirium. BUN is commonly elevated with dehydration. Cr is the serum marker most useful in clinically estimating GFR.

Routine urinalysis is a common screening test that includes an examination of the general appearance of the sample and measurement of the pH, urine-specific gravity, urine bilirubin, and urine glucose. Elevations in urine glucose can be seen in diabetes. Elevated urine bilirubin occurs in biliary disease. A high urine-specific gravity may occur in diabetes mellitus, in dehydrated states, or from SIADH. A low urine-specific gravity may occur in diabetes insipidus (DI) or from excessive water ingestion. In older adults, sudden changes in behavior or cognition may portend an underlying urinary tract infection, and urinalysis is generally warranted.

Elevated levels of porphobilinogen are found in the urine of symptomatic patients with acute intermittent porphyria. Symptoms of this disease include psychosis, apathy, or depression, along with intermittent abdominal pain, neuropathy, and autonomic dysfunction. If urine porphobilinogen levels are elevated when the patient is symptomatic, then a collection of a 24-hour urine specimen for quantitative assessment of porphobilinogen and aminolevulinic acid is indicated.

Liver Function

LFTs commonly include the serum aminotransferases, alkaline phosphatase, γ-glutamyl transpeptidase, serum bilirubin, and tests of synthetic function, usually the serum albumin concentration and prothrombin time.

Elevations in AST may occur with diseases of the liver, heart, lungs, kidneys, and skeletal muscle. In patients with alcohol-induced liver disease, AST is typically more elevated than ALT. In viral- and drug-induced liver disease, ALT is often elevated. Serum GGT is elevated in hepatobiliary disease, including alcohol-induced liver disease and cirrhosis. Various psychotropic medications, including valproate, may induce transient transaminitis. Dosage of second-generation antipsychotics and some antidepressants must be reduced in patients with impaired hepatic function.

Alkaline phosphatase elevations occur in many diseases, including diseases of the liver, bone, kidney, and thyroid. Levels of alkaline phosphatase may be elevated in response to some psychiatric medications, most notably the phenothiazines.

Serum ammonia levels are often elevated in patients with hepatic encephalopathy. High levels are associated with the delirium of hepatic encephalopathy. Serum ammonia levels also may be elevated in patients undergoing treatment with valproate.

Serum bilirubin is an index of hepatic and bile duct function. Prehepatic, unconjugated, or indirect bilirubin and posthepatic, conjugated, or direct bilirubin are often assessed to help elucidate the origin of the elevation in bilirubin.

The total protein is the measurement of serum albumin and globulins. Many psychiatric medications are protein bound. Low levels of serum total protein result in a high fraction of an ingested drug remaining in the unbound (active) state. High levels of unbound drug increase the active effects and side effects that a patient experiences. Total protein may be low in malnourished states, such as those seen in eating disorder patients. Low total protein and serum albumin in a patient with elevated liver enzymes is an indicator of significant liver disease. Consequently, these levels are often assessed in patients who abuse alcohol.

A relatively higher elevation in serum aminotransferases compared to alkaline phosphatase suggests hepatocellular injury. The opposite pattern is seen with cholestasis. Low albumin levels are seen in chronic conditions such as cirrhosis. Alcoholic liver disease is suggested when the ratio of AST to ALT is 2:1 or greater, although this may occur with steatohepatitis or hepatitis-C–related cirrhosis. When the AST to ALT ratio is 2:1, or greater and the GGT is twice the normal level, the pattern strongly suggests alcoholic liver disease.

Infectious Disease

The diagnostic assessment for patients with psychiatric symptoms may include a workup for infectious illness, depending upon the patient's history and examination findings. Further, comorbid infections requiring clinical attention and referral may be present in psychiatric patients with high-risk behaviors such as IV substance use or high-risk sexual behavior. Below, infectious disease as a differential diagnosis in psychiatric presentations is reviewed. Additionally, infectious diseases that may be a focus of clinical attention in patients with established psychiatric illness are also discussed.

A 49-year-old homeless male with a history of bipolar I disorder was brought by ambulance to the emergency department after a witnessed seizure. He had a fever of 39 °C, and his heart rate was in the 120s. He was severely disorganized and agitated, with unintelligible speech, and appeared to be responding to internal stimuli. He was unable to cooperate with a neurologic examination. Magnetic resonance imaging (MRI) of the head demonstrated T2/FLAIR hyperintensity involving the temporal lobe. Cerebrospinal fluid (CSF) examination revealed an elevated WBC of 121 cells/μL (mostly lymphocytes) and mild protein elevation (57 mg/dL). Herpes simplex virus (HSV) was identified in CSF. Following antiviral treatment for herpes encephalitis, the patient recovered to his baseline.

In this patient's case, despite known bipolar I disorder and severe psychiatric symptoms at the time of presentation, there were indicators of a comorbid medical illness, including seizure, aphasia, and fever. An evaluation for infectious etiologies was required in order to initiate the correct treatment for his condition.

Central Nervous System Infection. Infections of the CNS often present with neuropsychiatric symptoms. These symptoms may range from confusion, inattention, or altered level of consciousness, to severe agitation and psychosis. CNS infection with herpes simplex virus (HSV) is particularly notorious for causing psychiatric disturbances. Similarly, tertiary syphilis can have a number of psychiatric and neurologic manifestations. Other CNS infections may also cause psychiatric symptoms; delirium is especially common. It is unusual for CNS infections to cause neuropsychiatric symptoms in the absence of signs such as fever, nuchal rigidity, seizure, or other neurologic findings. However, there are exceptions, and it is important to know the immune status of a patient. Infected patients who are immunocompromised may be less likely to mount a fever; mental status changes among the immunocompromised host should alert clinicians to a broad differential that includes a variety of infections. Various infectious agents can infect the CNS, including bacteria, viruses, prions, fungi, and parasites. The workup initiated depends upon many factors, including the patient's travel history, occupational and recreational exposures, sexual behaviors, and immune status, all of which may put the patient at risk for particular infections.

A 58-year-old man with schizoaffective disorder, treated with stable doses of haloperidol, clozapine, and lithium, presented to the emergency department with his conservator for evaluation of fever and altered mental status. On evaluation, he was febrile, tachycardic, rigid, and tremulous. Neuroleptic malignant syndrome (NMS) was suspected, and psychotropic medications were discontinued. MRI was unremarkable, but CSF revealed elevated WBC of 54 cells/μL (mostly lymphocytes), elevated protein of 71 mg/dL, and positive serology for West Nile virus. With supportive care and eventual reinitiation of his psychotropic medication regimen, he eventually recovered to baseline, but not without first requiring psychiatric hospitalization for relapse of manic and psychotic symptoms.

Other Infections. In addition to CNS infection, a systemic infection may cause mental status changes. For most patients presenting with delirium, an infection should be included in the differential diagnosis. Patients with sepsis or septic shock commonly develop delirium. For vulnerable patients such as the older adults or cognitively impaired, an infection as mild as cystitis may be sufficient to induce mental status changes. An infectious workup for delirium generally includes a CBC, urinalysis, chest x-ray, and sometimes blood cultures, among other specific evaluations depending upon the individual patient.

In addition to considering infectious disease in the differential diagnosis when appropriate clinical and historical features are present, evaluation for certain infectious diseases is sometimes indicated in psychiatric patients with known risk factors. Below, specific infectious agents that may be relevant in patients with high-risk sexual behavior or IV substance use are discussed. Some evaluations such as HIV screening may be offered in the course of routine clinical care, regardless of patient risk factors; CDC guidelines now recommend opt-out HIV screening for all adolescents and adults and annual HIV screening for patients at high risk for HIV infection.

Intravenous Drug Use and Infection. The IV route is used for many substances of abuse. Most commonly, heroin, amphetamines, and cocaine are used alone or in combination. Because needles may be contaminated, IV drug users are at risk for bacterial endocarditis, hepatitis

B and C, and HIV infection. It has been estimated that over 60% of new cases of hepatitis C occur in individuals with a history of injecting illicit drugs.

The use of contaminated needles or nonsterile injection sites places IV drug users at risk for bacterial infections, including abscesses, bacteremia, and bacterial endocarditis. Findings on physical examination suggestive of endocarditis, possible bacteremia, or abscess necessitate obtaining a CBC to rule out an elevated WBC. Blood cultures should be obtained if the patient is febrile or if findings are suggestive of bacteremia or endocarditis, and internal medicine consultation should be obtained.

Sexually Transmitted Diseases. Some psychiatric illnesses, such as mania and substance abuse, are associated with a higher risk of contracting STDs. STDs include HSV types 1 and 2, Chlamydia, hepatitis B and C viruses, gonorrhea, syphilis, and HIV. Risk factors for STD include contact with sex workers, drug abuse, prior history of STDs, meeting sex partners on the Internet, multiple sex partners, a new sex partner, and being young or unmarried.

Many patients require supportive counseling before and after STD testing. Counseling consists of providing patients with educational material, allowing patients the opportunity to ask questions, and assessing their emotional response to the process. Inquiring about a patient’s motivation for testing often is a helpful way to initiate discussion of other issues. Patients should be informed about the meaning of positive and negative results. Patients who test positive may require referral to a specialist who can discuss the implications of the result and future treatment. Patients with high-risk behaviors and negative test results should be cautioned about the length of time required for seroconversion with HIV testing and the need, when appropriate, for supplemental testing. It is useful to evaluate the patient’s understanding of ways to reduce the risk of contracting or transmitting STDs. Regardless of their test results, all patients can benefit from a discussion of high-risk behaviors. For patients who test positive, in addition to treatment, it is particularly important for clinicians to provide the patient with information about how to minimize the risk of STD transmission. Clinicians should emphasize the need for protected sex with fully informed partners and the avoidance of donating sperm, blood, or organs. Clinicians should warn patients about the risks of transmitting certain STDs associated with sharing razors, toothbrushes, and any other objects that could have blood on them.

Chlamydia. Chlamydia infection can cause pelvic inflammatory disease, ectopic pregnancy, chronic pelvic pain, and infertility. It is also associated with an increased risk of contracting and transmitting HIV. Reinfection is common and is usually due to a lack of partner treatment. Many infections are asymptomatic. Chlamydia screening uses tests based on nucleic acid amplification (NAA). These amplification tests have high sensitivity and specificity for men and women, using urine or cervical or urethral swabs.

Gonorrhea. Gonorrhea infection can cause pelvic inflammatory disease, ectopic pregnancy, chronic pelvic pain, and infertility. It is also associated with an increased risk of contracting and transmitting HIV. Reinfection is common and is usually due to a lack of partner treatment. Many infections are asymptomatic. The diagnosis of a gonococcal infection can be made with NAA techniques, deoxyribonucleic acid (DNA) probes, Gram stain, or culture on Thayer–Martin media. NAA tests have high sensitivity and specificity for men and women, using urine or cervical, vaginal, or urethral swabs. In women, vaginal or cervical swabs have higher sensitivity than urine samples.

Human Immunodeficiency Virus (HIV). The enzyme-linked immunosorbent assay (ELISA) test and the Western blot test are used in the assessment of a patient’s HIV status. Both tests detect the presence of antibodies to HIV. Antibodies to HIV become detectable anywhere from 3 weeks to 3 months after infection. Typically, the ELISA test is used as a screening test, and the Western blot test is used to confirm the result. Other tests are performed to assess the extent of HIV disease burden, treatment response to antiretroviral therapy, and the degree to which the virus has affected the immune system. Measurement of viral load is used to assess the HIV disease burden. Measurement of $CD4^+$ T cell count provides information about the need for prophylaxis against opportunistic infections.

The Centers for Disease Control and Prevention recommend HIV screening for all adolescent and adult patients. Additionally, annual HIV testing is recommended in high-risk populations. High-risk groups include men who have sex with men, individuals who exchange sex for money or substances, IV drug users, individuals who have had sex with one or more partners since their last HIV test, and individuals who have had sex with an HIV-infected person or anyone in the aforementioned high-risk groups. HIV is transmitted via blood and certain body fluids (semen, rectal fluids, vaginal fluids, breast milk) when these fluids come into direct contact with mucous membranes, damaged tissues, or when injected directly into the bloodstream.

Psychiatric and behavioral symptoms associated with HIV vary according to whether patients are in the acute or chronic stage of infection. In the acute stage, HIV may present with headache and malaise, whereas HIV-infected patients may have CNS manifestations as their predominant symptoms even when other symptoms are absent. Behavioral changes in an individual patient may be due to either the direct effect of the virus on the brain or associated conditions such as infection, mass effect, or malnutrition. CNS manifestations of HIV infection include dementia, personality changes, affective symptoms of either mania or depression, and psychosis. HIV infection may be associated with delirium, dementia, mania, and atypical presentations of anxiety, schizophrenia, and depression.

With antiretroviral therapies, HIV-associated dementia is less common today, but milder HIV-associated neurocognitive disorders (HAND) remain prevalent, with approximately half of HIV-infected individuals demonstrating a lower level of cognitive performance than would be predicted when compared to population norms. Depression occurs in up to 58% of people living with HIV. Mania associated with HIV or AIDS also may occur either in the acute infection or in the context of HIV-related dementia.

Syphilis. Nontreponemal tests for syphilis, such as the rapid plasma reagin (RPR) test and the VDRL test, are common screening tests for syphilis. Quantification testing may be performed after a positive RPR or VDRL test. Quantification enables levels to be followed to assess response to treatment. After a positive nontreponemal screening test, a more specific treponemal test, such as the fluorescent treponemal antibody absorption (FTA-ABS) test, is used to confirm the positive result. The FTA-ABS test detects antibodies against *Treponema pallidum* spirochetes. Once positive, even with adequate treatment, specific treponemal tests like the FTA-ABS may remain positive for years. False-positive results may occur; this is true for both treponemal and nontreponemal tests. Pregnant patients may screen falsely positive for syphilis. False-positive nontreponemal tests are also associated with acute febrile illnesses, and sometimes in patients with chronic conditions, especially patients with systemic lupus erythematosus (SLE).

When syphilis infection of the CNS is suspected, a VDRL is performed on CSF. Patients with neurosyphilis also commonly have abnormalities in protein levels, γ-globulins, and cell counts in CSF. Due to the low sensitivity of the VDRL, the FTA-ABS test may also be used to evaluate CSF when neurosyphilis is suspected. It has higher sensitivity but lower specificity.

Viral Hepatitis. Several types of viruses may cause viral hepatitis. Viral hepatitis produces abnormalities in LFTs, including elevation of liver enzymes. Symptoms range from mild flu-like manifestations to rapidly progressive and fatal liver failure. Psychiatric manifestations include depression, anxiety, weakness, and psychosis. Viral hepatitis also may impair the metabolism of psychotropic medications that are metabolized by the liver. Impaired liver metabolism requires an adjustment of the dose of medications metabolized by the liver or consideration of agents that are less affected by alterations in liver metabolism. Viruses causing hepatitis include hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis D virus (HDV), and hepatitis E virus (HEV). HAV and HEV are usually transmitted via the fecal–oral route. HBV and HCV are transmitted via contamination by body fluids; this may include transmission by contact with infected blood or blood products or by sexual contact. HDV is a defective RNA (ribonucleic acid) virus that causes hepatitis only in association with HBV infection.

The WBC is usually normal to low in patients with hepatitis, especially in the preicteric phase. Large atypical lymphocytes occasionally are present. Rarely, aplastic anemia follows an episode of acute hepatitis not caused by any of the known hepatitis viruses. Mild proteinuria is common, and bilirubinuria often precedes the appearance of jaundice. Acholic stools frequently are present during the icteric phase. Strikingly elevated AST or ALT occurs early, followed by elevations of bilirubin and alkaline phosphatase. In a minority of patients, elevations of bilirubin and alkaline phosphatase persist after aminotransferase levels have normalized. Cholestasis may be substantial in acute HAV infection. Marked prolongation of the PT in severe hepatitis correlates with increased mortality.

Chronic hepatitis, characterized by elevated aminotransferase levels for more than 6 months, develops in less than 5% of immunocompetent adults with acute HBV infection. Fifty percent to 85% of all persons with acute HCV infection develop chronic HCV infection. Ultimately, cirrhosis develops in as many as 50% of those with chronic HCV infection and 20% of those with chronic HBV infection; the risk of cirrhosis is even higher in patients coinfected with both viruses or with HIV. Patients with cirrhosis are at risk, with a rate of 0% to 3% per year, of hepatocellular carcinoma. Even in the absence of cirrhosis, patients with chronic HBV infection—particularly those with active viral replication—are at an increased risk.

HEPATITIS B VIRUS. Psychiatric manifestations have been reported during all stages of HBV infection. HBV is transmitted via contamination by body fluids and is usually transmitted by contact with infected blood or blood products, by sexual contact, or from mother to baby at birth. The virus is present in saliva, semen, and vaginal secretions, as well as blood. The incubation period of HBV is 1 to 4 months. Transmission of HBV commonly occurs during sexual contact and via the use of shared implements for IV drug use; it also may occur through tattooing and piercing and as a result of needle-stick injuries. After acute HBV infection, HBV infection persists in 1% to 2% of immunocompetent adults. Persons with chronic HBV infection are at substantial risk of cirrhosis and hepatocellular carcinoma. There is common comorbidity between HBV infection and HIV infection, given the shared risk factors for both conditions.

HBV infection can be detected by measuring viral antigens, antibodies to the virus, and viral DNA. The appearance of HBsAg is the first evidence of infection and persists throughout the clinical illness. The detection of HBsAg establishes active infection with HBV and implies infectivity. When it is present beyond 6 months, infection is considered chronic. Hepatitis B e antigen (HBeAg) is a soluble protein found only in HBsAg-positive serum. HBeAg indicates viral replication and infectivity. Antibody to HBsAg (anti-HBs) appears after clearance of HBsAg and after successful vaccination against hepatitis B. Disappearance of HBsAg and the appearance of anti-HBs signal recovery from HBV infection, noninfectivity, and immunity. Detection of the hepatitis B surface antigen (HBsAg) indicates active or chronic infection. IgM antibody to HBcAg (anti-HBc) in the setting of acute hepatitis indicates a diagnosis of acute HBV infection. It fills the serologic gap in patients who have cleared HBsAg but do not yet have detectable anti-HBs. IgG anti-HBc also appears during acute HBV infection but persists indefinitely, whether the patient recovers or develops chronic hepatitis.

HEPATITIS C VIRUS. HCV infection is often diagnosed in individuals who abuse IV drugs, as these individuals are at high risk for contracting HCV infection via the use of contaminated equipment. Clinical illness is generally mild, usually asymptomatic; it is characterized by waxing and waning aminotransferase elevations and a high rate (50% to 85%) of chronic HCV infection. HCV infection is associated with an increased risk of cirrhosis and hepatocellular carcinoma. Diagnosis of HCV infection is based on immunoassays that detect antibodies to HCV (anti-HCV). These tests are complemented by assays for HCV RNA to confirm active versus past infection. Fortunately, antiviral treatments have up to a 95% success rate in the treatment of HCV after infection.

Endocrine Evaluation

Endocrine disease is of great relevance to psychiatry. Endocrine illness frequently has psychiatric manifestations. Management of psychiatric illness may be complicated by comorbid endocrine disease. For these reasons, screening for endocrine disease often is of relevance to the psychiatrist.

Adrenal Disease. Adrenal disease may have psychiatric manifestations, including depression, anxiety, mania, dementia, psychosis, and delirium. However, patients with adrenal disease rarely come to the attention of psychiatrists. Assessment and management of these patients are best done in conjunction with specialists.

Low plasma levels of cortisol are found in Addison disease. These patients may have symptoms that are also common in psychiatric conditions, including fatigue, anorexia, weight loss, and malaise. Patients may also have memory impairment, confusion, or delirium. Depression or psychosis with hallucinations and delusions may occur.

Elevated levels of cortisol are seen in Cushing syndrome. About half of all patients with Cushing syndrome develop psychiatric symptoms. These symptoms may include lability, irritability, anxiety, panic attacks, depressed mood, euphoria, mania, or paranoia. Cognitive dysfunction may include cognitive slowing and poor short-term memory. Symptoms usually improve when cortisol normalizes. If not, or if symptoms are severe, then psychiatric treatment may be necessary.

Cortisol levels have not been found to be useful in the assessment or management of a primary psychiatric disease. In particular, the dexamethasone-suppression test (DST) remains a research tool in psychiatry that is not used in routine clinical care.

Anabolic Steroid Use. The use of anabolic steroids has been associated with irritability, aggression, depression, and psychosis. Some athletes and bodybuilders may abuse anabolic steroids, but they are also increasingly used to enhance sexual performance, even in older adults. It is important to inquire about steroid use given the availability, even if a patient does not appear to have excess muscular bulk. Urine specimens can be used to screen for these agents. Because so many compounds have been synthesized, a variety of tests may be required to confirm the diagnosis, depending upon the compound that has been used. Consultation with a specialist is advised.

Generally, androgens other than testosterone can be detected by gas chromatography and mass spectroscopy. Exogenous testosterone is indistinguishable from endogenous testosterone. When the use of exogenous testosterone is suspected, the urinary ratio of testosterone to its endogenous epimer, epitestosterone, is assessed. Usually, the ratio ranges from 1:1 to 3:1. When exogenous use occurs, the ratio may be greater than 4:1 and even greater than 6:1. Another method for detection of exogenous testosterone use is to evaluate the ratio of ^{13}C to ^{12}C in urinary metabolites of testosterone. Because synthetic testosterone preparations are derived from plant sterols, their presence, in comparison to the presence of endogenous testosterone alone, produces a lower ratio of ^{13}C to ^{12}C.

Antidiuretic Hormone. Arginine vasopressin (AVP), also called antidiuretic hormone (ADH), is decreased in central DI. DI may be central (due to the pituitary or hypothalamus) or nephrogenic. Nephrogenic DI may be acquired or due to an inherited X-linked condition. Lithium-induced DI is an example of an acquired form of DI. Lithium has been shown to decrease the sensitivity of renal tubules to AVP. Patients with central DI respond to the administration of vasopressin with a decrease in urine output. Secondary central DI may develop in response to head trauma that produces damage in the pituitary or hypothalamus.

About one-fifth of patients taking lithium develop polyuria, and a larger amount may have some degree of impairment in concentrating urine. Chronic treatment with lithium is a common cause of nephrogenic DI. However, there are other causes of polyuria in lithium-treated patients in addition to nephrogenic DI. Primary polydipsia is common and is often associated with the dry mouth associated with many psychiatric medications. When patients chronically treated with lithium develop altered mental status, they may become functionally fluid restricted due to inability to drink to thirst (often with polydipsia and polyuria at baseline). Close monitoring of fluid intake and urine output is required; patients are at high risk of developing hypernatremia and severe dehydration if they have impaired ability to drink to thirst, combined with impaired ability to concentrate urine.

A 73-year-old gentleman with bipolar I disorder had been stably managed on lithium for over 40 years. He presented to the hospital with altered mental status. Pneumonia was diagnosed, and the patient ultimately required intubation due to respiratory distress. While intubated, he developed hypernatremia and acute renal failure (prerenal). A review of his fluid intake and output revealed that he'd had several liters of daily urine output but was not receiving adequate fluid replacement. Following nephrology evaluation of serum and urine osmolality studies, DI was diagnosed. IV fluids were increased and adjusted to provide free water replacement.

Excessive secretion of AVP increases the retention of fluid in the body. This condition is called *SIADH.* Water retention in SIADH causes hyponatremia. SIADH may develop in response to injury to the brain or from medication administration (including phenothiazines, butyrophenones, carbamazepine, oxcarbazepine, as well as SSRIs and many other agents). The hyponatremia associated with this condition may produce delirium. The relationship between AVP and psychogenic polydipsia remains unclear. However, it appears that in a minority of cases with psychogenic polydipsia, AVP levels may be abnormal. SIADH is associated with hyponatremia, low serum osmolality, and high urine osmolality. It should be considered when patients demonstrate hyponatremia and concentrated urine (i.e., high urine osmolality or high specific gravity). This condition should be investigated with the assistance of a specialist.

Human Chorionic Gonadotropin. Human chorionic gonadotropin (hCG) can be assessed in the urine and blood. The urine test for hCG is the basis for the commonly used urine pregnancy test. Routine tests are most accurate when performed 1 to 2 weeks after a missed period. However, there are ultrasensitive urine hCG tests that can accurately detect pregnancy seven days after fertilization. Pregnancy tests often are obtained before initiating certain psychotropic medications, such as carbamazepine and valproic acid, which are associated with congenital anomalies.

Parathyroid Hormone. Parathyroid hormone modulates serum concentrations of calcium and phosphorus. Dysregulation in this hormone and the resulting production of abnormalities in calcium and phosphorus may produce depression or delirium. Chronic lithium treatment may result in hyperparathyroidism and hypercalcemia in a small percentage of patients.

Neuropsychiatric symptoms associated with hyperparathyroidism and hypercalcemia include depression, anxiety, cognitive dulling, lethargy, and psychosis; delirium and stupor may occur as calcium levels rise. Symptoms usually improve with treatment of the underlying condition.

Prolactin. Prolactin levels may become elevated in response to the administration of antipsychotic agents (high potency first-generation and some second-generation) that block dopamine receptors in the pituitary. Dopamine inhibits the release of prolactin; thus, dopamine blockade leads to an increase in prolactin release.

Elevated prolactin levels in premenopausal women are associated with menstrual abnormalities, infertility, alterations in libido, decreased bone density, and galactorrhea. Most premenopausal women who have hyperprolactinemia do not have galactorrhea. Generally, levels greater than 100 ng/mL are associated with infertility and amenorrhea, hot flashes, and vaginal dryness. Levels from 50 to 100 ng/mL may cause amenorrhea or oligomenorrhea. Values of 20 to 50 ng/mL may cause infertility even when there is no abnormality of the menstrual cycle. The normal range for serum prolactin is from 5 to 20 ng/mL.

In men, hyperprolactinemia may result in decreased libido, erectile dysfunction, fatigue, gynecomastia, and infertility. Only rarely does galactorrhea occur. Chronic hyperprolactinemia may result in loss of body hair, loss of muscle mass, and osteoporosis.

Cerebral magnetic resonance imaging (MRI) is not usually performed if the patient is taking an antipsychotic drug known to cause hyperprolactinemia and the magnitude of the prolactin elevation is consistent with drug-induced causes.

Prolactin levels may briefly rise after a seizure. For this reason, prompt measurement of a prolactin level after possible seizure activity may assist in differentiating a seizure from a nonepileptic behavioral spell.

Thyroid Hormone. Disease of the thyroid is associated with many psychiatric manifestations. Thyroid disease is most commonly associated with depression and anxiety but may also give rise to symptoms of panic, dementia, and psychosis. Thyroid studies are often evaluated in patients with depression, as thyroid disease may mimic depression. It is difficult to achieve euthymia if a patient is not euthyroid.

Measurements of TSH typically are used as a screening test for thyroid disease. If normal, then no further assessment is performed. If high, then TSH may be repeated and free T4 added to assess the degree of hypothyroidism. If serum TSH is low, then free T4 and T3 are assessed to determine the degree of hyperthyroidism. Hypothyroidism is characterized by an elevated TSH and a low free T4. Subclinical hypothyroidism consists of a high TSH and a normal free T4. Hyperthyroidism consists of low TSH and high free T4 and T3.

Usually, thyroid function is assessed in psychiatric patients with a history of thyroid disease or those with symptoms suggestive of thyroid illness. Such symptoms may include increased appetite, weight loss, and heat intolerance in hyperthyroidism or weight gain, fatigue, and cold intolerance in hypothyroidism. Alternatively, risk factors for thyroid disease may prompt examination of thyroid function. Risk factors include prior treatment with lithium or iodine-131, a history of irradiation of the neck and chest, or a history of thyroidectomy. Additionally, thyroid function is commonly assessed in patients currently receiving exogenous thyroid hormone replacement to ensure that they are euthyroid.

Interpretation of thyroid function tests may be complex and is often best done with the assistance of a specialist. Interpretation is further complicated by the presence of transient abnormalities of thyroid function in many newly admitted psychiatric patients. These abnormalities normalize over time without treatment and may be related to stress.

Pancreatic Function. The pancreas is both an exocrine and endocrine gland. The endocrine function of the pancreas is to control blood sugar levels through the release of the hormones insulin and glucagon. See the section on metabolic syndrome for a discussion of insulin. The exocrine function of the pancreas is to assist in digestion. Pancreatitis may develop in patients with alcohol use disorders. Additionally, pancreatitis is a rare complication of treatment with some psychotropic medications, including clozapine, risperidone, olanzapine, and valproate. Pancreatitis is usually accompanied by elevations in serum amylase and lipase. Serum amylase levels may be fractionated into salivary and pancreatic components. Elevations in salivary levels are commonly observed in eating disorder patients who purge by self-induced emesis. When a pancreatic disease is suspected, serum lipase is typically measured. Elevations in amylase and lipase suggest pancreatic disease. Patients with eating disorders who purge often have elevated levels of serum amylase and normal levels of serum lipase.

Autoimmune and Inflammatory Disease

Psychiatric symptoms are highly comorbid in autoimmune and inflammatory conditions. In fact, autoimmune and inflammatory conditions sometimes present with psychiatric symptoms and thus should be included in the differential diagnosis for patients with compatible historical and clinical features.

Systemic Lupus Erythematosus. SLE is a chronic inflammatory disorder. Tests for SLE are based on the detection of antibodies formed as part of the disease. Antinuclear antibodies are found in virtually all patients with SLE at some point during the illness. An indirect immunofluorescence assay is used to detect antinuclear antibodies. This test can be positive in a variety of rheumatologic diseases. For this reason, a positive result usually is followed by additional tests, including a test to detect antidouble-stranded DNA antibodies. Anti-dsDNA antibodies, when associated with antinuclear antibodies, are strongly suggestive of a diagnosis of SLE. Anti-dsDNA antibodies often fluctuate with disease activity, may predict flares, and are often followed to monitor disease activity and response to treatment.

Neuropsychiatric manifestations of SLE include depression, cognitive impairment, delirium, mania, and psychosis. About 5% of patients with SLE experience psychosis. Cognitive dysfunction is common and may include apraxia, aphasia, and agnosia, along with impairments in abstract thinking, short- and long-term memory, judgment, and visual–spatial function. Depression is also highly comorbid in patients with SLE. Many of the neuropsychiatric symptoms of SLE are also caused by steroids, which are used to treat SLE exacerbations. It is not uncommon that patients with SLE experience depression when undergoing a course of steroids, but in more severe cases, patients may experience psychosis or steroid-induced mania. If patients require repeated steroid therapy, pretreatment with second-generation antipsychotics may reduce or eliminate steroid-induced psychosis.

Autoimmune or Paraneoplastic Encephalopathy. There is growing recognition and identification of antibodies that target the CNS, resulting in a variety of psychiatric and neurologic signs and symptoms. Though presentations vary, common presenting symptoms include cognitive impairment, mood symptoms, psychosis, seizure, and abnormalities of movement. Infrequently, patients may present with psychiatric symptoms in the absence of neurologic symptoms. These syndromes occur across the lifespan. Psychiatrists should suspect these disorders when patients have atypical presentations of psychiatric illness, especially when there are comorbid neurologic signs or symptoms.

There are currently a number of clinical laboratory evaluations for antibodies associated with paraneoplastic and autoimmune neuropsychiatric disorders. Antibodies may target either intracellular or extracellular antigens. Antibodies to intracellular antigens include anti-Hu, anti-Ma2, and anti-CRMP5, among others. Extracellular (synaptic) antibody targets include N-methyl-D-aspartate (NMDA) receptors, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors, γ-aminobutyric acid (GABA) receptors, and voltage-gated potassium channel complex associated antibodies, that is, LGI1 and Caspr2, among others. Laboratory evaluations for CNS autoantibodies will certainly continue to grow as research continues; reputable laboratories should be consulted for up-to-date testing panels.

When an autoimmune or paraneoplastic encephalopathy is suspected, serum and CSF antibody testing, routine CSF studies, as well as EEG and brain MRI should be considered. These studies serve both to rule out other conditions (such as brain neoplasm) and may also provide supporting evidence for an autoimmune or paraneoplastic cause. Patients with autoimmune or paraneoplastic disorders may have mildly to moderately elevated WBC and protein in CSF. EEG may reveal seizure activity, epileptiform discharges, and/or focal or generalized slowing. MRI is often normal but may reveal signal hyperintensities on FLAIR or T2-weighted imaging sequences. The prompt identification of autoimmune or paraneoplastic neuropsychiatric disorders facilitates the rapid initiation of appropriate treatments.

Anti-NMDA receptor encephalitis, for example, is an autoimmune encephalitis in which a typically young adult female patient initially presents with mood symptoms, agitation, psychosis, or other variable psychiatric manifestations. (Of note, while majority young adult and female, the disorder can present in either sex and has been documented to occur across the lifespan). The rapid evolution of symptoms, often in combination with neurocognitive impairment and neurologic symptoms (e.g., movement abnormalities, seizures), can be worked up with serum and CSF antibody testing and pelvic ultrasound, which may reveal an ovarian teratoma. Treatment involves removal of the tumor, if found, steroid administration, IV immunoglobulin and/or plasma exchange, and sometimes second-line immunotherapies. The majority of patients slowly recover.

An 18-year-old female with no past medical or psychiatric history presented to the emergency department with abrupt onset of mood lability, paranoia, and auditory hallucinations. She was inattentive, difficult to engage, and exhibited echolalia. She had occasional flailing movements of her upper extremities. Collateral history from her father revealed that she had a generalized tonic–clonic seizure 1 day prior to presentation, for which she was evaluated in the emergency department and released home, with an appointment for outpatient neurologic follow-up. At her second emergency department visit in 2 days, EEG revealed epileptiform discharges. MRI was unremarkable. CSF examination revealed mildly elevated WBC (18 cells/μL) and modestly elevated protein (74 mg/dL). Infectious workup was negative. CSF antibody evaluation revealed IgG antibodies to the NMDA receptor, and autoimmune NMDA receptor encephalitis was diagnosed.

Other Autoimmune and Inflammatory Conditions. Autoimmune and inflammatory conditions ranging from multiple sclerosis (MS) to rheumatoid arthritis to inflammatory bowel disease are frequently associated with comorbid psychiatric illness, especially affective disorders. This is likely due to a variety of factors, including the psychosocial impact of chronic illness, treatment with medications that may have psychiatric side effects, as well as potential neurobiologic effects of inflammatory cytokines or direct involvement of the brain (e.g., MS). It is important to identify and treat comorbid psychiatric illnesses that cause suffering, worsen the quality of life, and threaten health outcomes. Similarly, it is important for psychiatrists to identify signs and symptoms of autoimmune and inflammatory conditions in their patients and refer them to specialists for appropriate care. When managing psychiatric patients who develop symptoms concerning autoimmune or inflammatory illness, referral to appropriate specialists who can order the appropriate medical assessment and laboratory evaluation is essential.

Neurologic Disorders

Many neurologic disorders have accompanying psychiatric symptoms. Neurodegenerative disorders, cerebrovascular disease, movement disorders, seizure disorders, and demyelinating disorders are often accompanied by psychiatric symptoms. Imaging studies, EEG, and neuromuscular studies, in addition to laboratory evaluations of blood and CSF, may be part of the diagnostic evaluation of these disorders.

Neoplastic Disorders

Infrequently, psychiatric symptoms may be the result of a primary or metastatic neoplasm involving the CNS. Most often, neurologic symptoms accompany psychiatric symptoms when there is a CNS tumor, but this is not always the case. Neuroimaging studies are the medical assessment of choice to rule out CNS tumors as etiologic for neuropsychiatric symptoms. Some other cancers, such as pancreatic cancer, are commonly accompanied by depression, even before the cancer has been diagnosed. In the setting of psychiatric symptoms accompanied by weight loss and new or concerning somatic symptoms (abdominal pain, jaundice, bowel changes, hemoptysis), a workup for malignancy may be indicated, depending upon patient demographics, risk factors, history, and examination findings. This may include positron emission tomography (PET) studies, as well as other targeted evaluations depending upon presenting signs and symptoms.

Head Trauma

In an acutely altered patient, trauma must be ruled out via history and physical examination, with head imaging supplementing the evaluation when indicated. Head computed tomography (CT) is superior at identifying bony deformities and acute bleeds and is thus generally preferred in the acute evaluation of altered mental status when there is suspicion for head trauma.

Genetic Disorders

Though many psychiatric disorders have been demonstrated to be heritable to a certain extent, classically inherited genetic disorders are not commonly included in the differential diagnosis of psychiatric illness. Occasionally, laboratory evaluation for genetic disorders is indicated. Sometimes there is a known family history of a genetic disorder that prompts evaluation. Other times, compatible clinical features and an astute clinician lead to a diagnosis of a genetic disorder, even in the absence of family history.

A 43-year-old male with a history of recurrent major depression presented with new-onset psychotic symptoms. Neurologic examination revealed chorea, which had reportedly developed over the past 2 years. The patient did not know his family history; he had been adopted as an infant. Genetic testing revealed an abnormally large number of trinucleotide repeats (CAG) in the *huntingtin* gene on chromosome 4. Huntington disease was diagnosed.

Huntington Disease. Huntington disease is an autosomal dominant disorder involving an excess of trinucleotide repeats in the *huntingtin* gene on chromosome 4. Laboratory evaluation for Huntington disease is most commonly indicated when there is a known family history, often in combination with compatible clinical features, or for planning purposes. Huntington disease is also known as Huntington chorea. The basal ganglia are affected, leading to a movement disorder dominated by chorea (involuntary, fidgety movements of the face and extremities). Dystonia, rigidity, and other movement abnormalities often occur as well. Cognitive deterioration and behavioral symptoms accompany the movement disorder, sometimes preceding it.

Wilson Disease. Wilson disease is a rare autosomal recessive disorder that leads to abnormal copper metabolism. The disorder leads to excessive copper deposition in the brain and liver and may result in decreased intellectual functioning, personality changes, psychosis, liver dysfunction, and movement disorder. Symptoms usually present in the second and third decades of life. Laboratory assessment for Wilson disease includes the measurement of serum ceruloplasmin, the transport protein for copper, which is low, and urine copper, measured in a 24-hour specimen, which is elevated.

Other Genetic Disorders. Other genetic disorders may sometimes be suspected, and laboratory evaluations may be ordered when a clinical history and examination is concerning for a genetic etiology. For example, Fragile X syndrome may be suspected in a child with intellectual disability or autism who has characteristic physical features such as large ears and a long face, particularly when there is a compatible family history of a similar disability. Other genetic disorders with associated neuropsychiatric symptoms include Prader–Willi syndrome, adrenoleukodystrophy, and other inborn errors of metabolism, among others. These disorders are rare; when a genetic disorder is suspected, the appropriate laboratory evaluation is determined in consultation with specialists.

Nutritional Deficiency

A number of nutritional deficiencies may have accompanying neuropsychiatric symptoms; clinical history and examination should provide clues regarding whether nutritional deficiency may be contributing to the patient’s neuropsychiatric presentation. Thiamine and folate deficiency are common in patients who abuse alcohol. Thiamine deficiency can lead to Wernicke encephalopathy and Korsakoff syndrome in severe cases. Blood thiamine concentration can be measured to assess for deficiency. However, for patients at known risk for thiamine deficiency, most clinicians simply order thiamine supplementation without first ordering laboratory evaluation of thiamine. Folate and B_{12} deficiency, among other nutritional deficiencies, may be associated with dementia, delirium, psychosis, fatigue, and personality changes. Low folate levels may be found in patients who abuse alcohol or have poor nutrition. Additionally, some medications such as methotrexate and phenytoin can lead to folate deficiency. B_{12} deficiency is usually the result of poor absorption due to pernicious anemia or gastric disease. Folate and B_{12} can be directly measured. The metabolic intermediaries methylmalonic acid and homocysteine may be useful in the assessment of folate and B_{12} deficiency, particularly when vitamin levels are in a borderline range where deficiency cannot be definitely ruled in or out. Many severe nutritional deficiencies can result in neuropsychiatric disturbances, but the aforementioned deficiencies are those most commonly assessed in psychiatric patients.

Intoxication/Toxic Exposure

Laboratory assessment for intoxication or toxic exposures may be indicated under several different circumstances. Routine urine toxicology is seldom beneficial for inpatient psychiatric admissions but can be very helpful if illicit substance use is suspected or if the source of a patient’s symptoms is unclear. For example, abrupt, new-onset paranoia may be due to a primary psychiatric disorder, but it also may be due to intoxication with a number of substances. Toxicology screening may also be helpful in the evaluation and management of substance use disorders. Finally, there may be a concern for environmental exposures or for accidental or intentional overdoses.

Laboratory studies in intoxicated adults may help to identify the cause or consequence of intoxication. When the patient is unable or unwilling to provide information about the source of intoxication, appropriate studies, in addition to toxicology studies, include calcium, serum BUN, Cr, CK, glucose, ketones, electrolytes, LFTs, magnesium, serum osmolality, and a urinalysis. Reproductive-age women should undergo urine pregnancy testing.

Testing for drugs of abuse usually is performed on urine specimens. It also may be performed on specimens of blood, breath (alcohol), hair, saliva, and sweat. Urine screens provide information about recent use of frequently abused drugs such as alcohol, amphetamines, cocaine, marijuana, opioids, and phencyclidine (PCP) along with 3,4-methylenedioxymethamphetamine (MDMA) (ecstasy). Savvy patients may attempt to avoid a positive urine drug screen by providing a sample from another patient, diluting the sample, or modifying it with additives that may interfere with the drug abuse screen. Many substances may produce false positives with urine drug screening tests. When a false positive is suspected, a confirmatory test may be requested.

In cases where it is unclear what agents that the patient may have ingested, it is useful to screen for salicylates and acetaminophen. In the early stages of overdoses with lethal amounts of these medications, there may be few clinical signs. However, if an overdose is detected early, then effective treatments are available.

Comprehensive qualitative toxicology screening is usually performed by liquid and gas chromatography. This may require many hours to perform and is rarely done in routine clinical situations. It is usually performed in patients with unexplained toxicity and an atypical clinical picture. Quantitative toxicology assessments may be useful in managing patients who overdose when combined with clinical assessment and knowledge of when the ingestion occurred. With the exception of alcohol, drug levels for substances of abuse are not usually determined. Instead, only the presence or absence of the drug is determined. There is usually not a meaningful or useful correlation between the level of the drug and clinical behavior.

Drug Abuse. Patients may be unreliable when reporting their substance use history. Substance-induced mental disorders often resemble primary psychiatric disorders. Furthermore, substance use can exacerbate pre-existing mental illnesses. Indications for ordering a drug abuse screen include unexplained behavioral symptoms, a history of illicit drug use or dependence in a new patient evaluation, or a high-risk background (e.g., criminal record, adolescents, and history of prostitution). A drug abuse screen also is frequently used to monitor patient abstinence during treatment of substance use disorders. Such tests can be ordered on a scheduled or random basis. Many clinicians believe random testing may be more accurate in the assessment of abstinence. The tests also may help to motivate the patient.

Other laboratory data may suggest a problem with substance abuse. An increase in the mean corpuscular volume (MCV) is associated with alcohol abuse. Liver enzymes may be increased with alcohol abuse or from hepatitis B or C acquired from IV drug abuse. Serologic testing for hepatitis B or C can confirm that diagnosis.

TESTED SUBSTANCES. Routine tests are available for PCP, cocaine, tetrahydrocannabinol (THC) (*marijuana*), benzodiazepines, methamphetamine, and its metabolite amphetamine, morphine, codeine, methadone, barbiturates, lysergic acid diethylamide (LSD), and MDMA (*ecstasy*). Different laboratories have varying thresholds for the detection of these substances. Some laboratories report a positive finding based only on the results of an initial screening test, whereas others require a second confirmatory test, often by a more sensitive or specific method.

Drug screening tests may have high false-positive rates. This is often due to the interaction of prescribed medication with the test, resulting in false-positive results and a lack of confirmatory testing. False-negative tests are common as well. False-negative results may be due to problems with specimen collection and storage. An unbroken chain of custody of the specimen is important. Patients may attempt to alter a specimen by dilution, by changing the pH (bleach and baking soda), or by using other additives (perfume and salt) that interfere with the test. Allowing a specimen to remain at room temperature too long may result in inaccurate testing.

The length of time that a substance can be detected in the urine is listed in Table 7.6–1. These are approximate times that are influenced by the frequency and duration of antecedent drug use and the presence of illnesses that could alter excretion, such as renal or hepatic disease.

Urine drug screens may detect some opiate metabolites up to 3 to 4 days after use.

Heroin is metabolized to 6-monoacetylmorphine, a metabolite unique to heroin. Because it can be detected with urine testing, it can be used to differentiate heroin from other opioids. False-positive drug screen results for opiate may occur in patients taking quinolones or rifampin. Poppy seeds have also been reported to produce false-positive results. Confirmation testing can identify false positives from quinolones or rifampin but may not be able to detect poppy seed use.

Patients using dronabinol for chemotherapy-induced nausea or HIV wasting syndrome will test positive for THC/marijuana. It is not possible to differentiate dronabinol from marijuana on urine drug screens.

Alcohol. There is no single test or finding on physical examination that is diagnostic for alcohol use disorders. The history of the pattern of alcohol ingestion is most important in making the diagnosis. Laboratory test results and findings on physical examination may help to confirm the diagnosis. Alcohol use disorders may be associated with telangiectasias, hepatomegaly, and evidence of trauma on physical examination. In patients with acute alcohol intoxication, a blood alcohol level (BAL) may be useful. A high BAL in a patient who clinically does not show significant intoxication is consistent with tolerance. The degree of alcohol intoxication can also be assessed using the concentration of alcohol in expired respirations (breathalyzer). BAL is used to legally define intoxication in the determination of whether an individual is driving under the influence. The legal limit in many states is 80 mg/dL. Importantly, the degree of clinical impairment is an individual matter that is a function of a patient’s tolerance. For example, a chronic alcohol user who maintains a daily BAL of 250 mg/dL may show less cognitive impairment than the occasional drinker with a BAL of 120 mg/dL. In general, however, a BAL in the range of 50 to 100 mg/dL is associated with impaired judgment and coordination and levels greater than 100 mg/dL produce ataxia.

Table 7.6–1.
Drugs of Abuse That Can Be Detected in Urine

Drug	Length of Time Detected in Urine
Alcohol	7–12 hours
Amphetamine	48–72 hours
Barbiturate	24 hours (short-acting); 3 weeks (long acting)
Benzodiazepine	3 days
Cocaine	6–8 hours (metabolites 2–4 days)
Codeine	48 hours
Heroin	36–72 hours
Marijuana	2–7 days
Methadone	3 days
Methaqualone	7 days
Morphine	48–72 hours

Alcohol withdrawal is often thought to begin when BAL approaches zero, but it is important to realize that withdrawal can occur at much higher BAL in chronic users of alcohol. For example, if one maintains a daily BAL of 250 mg/dL, withdrawal may begin 150 mg/dL because the degree and rate of change in BAL are associated with symptoms of withdrawal. Alcohol withdrawal presents with an elevation in blood pressure and pulse and a tremor. These symptoms commonly occur when serum alcohol levels are low or rapidly declining. Active intervention with benzodiazepines may be necessary to prevent the condition from progressing to alcohol withdrawal delirium, which is a medical emergency that may be fatal. Patients who consume large amounts of alcohol may require correspondingly large doses of benzodiazepines to treat withdrawal. When BAL is suspected of being low or undetectable, patients with alcohol use disorders should be assessed frequently to prevent the onset of withdrawal symptoms.

Common laboratory studies for the evaluation of a patient with the diagnosis of alcohol abuse or dependence include a CBC (to assess MCV and red cell distribution width [RDW] for a macrocytic profile), serum electrolytes (to obtain levels of magnesium and potassium, which are commonly low), and liver enzymes (to assess GGT, which is sensitive for acute alcohol abuse, along with AST and ALT). Alcoholic liver disease is characterized by elevations in AST and ALT, typically in a ratio of AST to ALT of 2:1 or greater. GGT may be elevated. Carbohydrate-deficient transferrin (CDT) may also be helpful in the identification of chronic heavy alcohol use. Total protein, serum albumin, and PT are assessed as indices of nutritional status and liver function. Folate and thiamine deficiencies are frequently present in patients with alcohol abuse. Urine toxicology screens for other drugs of abuse may be helpful in understanding the full clinical picture.

Environmental Toxins. Specific toxins are associated with a variety of behavioral abnormalities. Exposure to toxins may occur through occupation or hobbies. Aluminum intoxication can cause a dementia-like condition. Aluminum can be detected in the urine or blood. Arsenic intoxication may cause fatigue, loss of consciousness, anemia, and hair loss. Arsenic can be detected in urine, blood, and hair. Manganese intoxication may present with delirium, confusion, and parkinsonian syndrome. Manganese may be detected in urine, blood, and hair. Symptoms of mercury intoxication include apathy, poor memory, lability, headache, and fatigue. Mercury can be detected in urine, blood, and hair. Manifestations of lead intoxication include encephalopathy, irritability, apathy, and anorexia. A blood lead level is used to diagnose lead toxicity occurring over the past few weeks. The free erythrocyte protoporphyrin test is a test for lead toxicity over the past 3 months.

Significant exposure to organic compounds, such as insecticides may produce behavioral abnormalities. Many insecticides have strong anticholinergic effects. Presentation with the cluster of anticholinergic effects such as dry mouth, blurred vision, constipation, and urinary retention typically alerts clinicians to the possibility of toxicity. There are no readily available laboratory tests to detect these compounds. Poison control centers may assist in the identification of appropriate testing facilities.

Volatile Solvent Inhalation. Volatile substances produce vapors that are inhaled for their psychoactive effect. The most commonly abused volatile solvents include gasoline, glue, paint thinner, and correction fluid. The aerosol propellants from cleaning sprays, deodorant sprays, and whipped cream containers may be abused. Nitrites, such as amyl nitrite (“poppers”) and butyl nitrite vials (“rush”), and anesthetic gases, such as chloroform, ether, and nitrous oxide, are also abused.

Chronic abuse of volatile solvents is associated with damage to the brain, liver, kidneys, lung, heart, bone marrow, and blood. Abuse may produce hypoxia or anoxia. Signs of abuse include short-term memory loss, cognitive impairment, slurred and “scanning” speech, and tremor. Cardiac arrhythmias may occur. Exposure to toluene, which is present in many cleaning solutions, paints, and glues, has been associated with loss of clear gray–white matter differentiation and with brain atrophy on MRI scans. Methemoglobinemia has occurred with butyl nitrite abuse. Chronic use of volatile solvents is associated with the production of panic attacks and an organic personality disorder. Chronic use may also produce impairment in working memory and executive cognitive function.

Initial treatment is to prevent access to the drug and to allow clearing of the substance through the respiratory system. When hypoxia occurs, supplemental oxygen administration may be necessary. This will not only treat the hypoxia but will also facilitate the clearance of the inhaled chemical. Antiarrhythmic medication may be needed. Over the long term, patients will benefit from entering sobriety programs to prevent relapse.

Acetaminophen or Salicylate Overdose. Intentional or unintentional overdose with acetaminophen or salicylates requires medical and laboratory assessment. In an altered patient, especially those with past suicide attempts or suicidal ideation, laboratory evaluation for ingestions should be performed.

ACETAMINOPHEN. Acetaminophen is one of the most frequently used agents in intentional drug overdoses and is a common cause of overdose-related deaths. Sequelae of acute overdose most notably include hepatotoxicity and acute hepatic failure, and also acute renal failure in some cases as 4 g of acetaminophen can lead to acute hepatic failure. Chronic abusers of alcohol may be particularly vulnerable to hepatotoxicity following acetaminophen overdose.

After an overdose of acetaminophen, large amounts of toxic metabolites are formed as the primary metabolic pathways of sulfate and glucuronide conjugation become saturated. The toxic metabolite can be inactivated by glutathione. Acetylcysteine (Mucomyst) treatment, which protects the liver by maintaining appropriate concentrations of glutathione, must occur promptly after an overdose to prevent hepatotoxicity. Activated charcoal effectively binds acetaminophen. After an acetaminophen overdose, activated charcoal is often administered with acetylcysteine to prevent hepatotoxicity. The plasma acetaminophen concentration should be determined not less than 4 hours after ingestion of the overdose. If the initial determination shows a high concentration, then a full course of acetylcysteine should be administered. If a low plasma concentration is reported, then initiation or continuation of acetylcysteine treatment is not necessary. LFTs (serum AST, serum ALT, PT, and bilirubin) should be performed at 24-hour intervals for at least 96 hours postingestion if the plasma acetaminophen concentration indicates potential hepatotoxicity. If no abnormalities are detected within 96 hours, then further determinations are not needed.

SALICYLATE TOXICITY. Aspirin (salicylate) is also used for overdose, though it is less common than acetaminophen. Ingestion of sufficient amounts of aspirin may be fatal. Some patients with chronic inflammatory conditions and/or chronic pain may ingest large amounts of salicylates for therapeutic reasons. Ingestion of 10 to 30 g of aspirin may be fatal in adults. Most patients will develop symptoms of toxicity when salicylate levels are greater than 40 mg/dL (2.9 mmol/L). Signs and symptoms of toxicity include acid–base abnormalities, tachypnea, tinnitus, nausea, and vomiting. In cases of severe toxicity, symptoms may include hyperthermia, altered mental status, pulmonary edema, and death.

Treatment includes administration of multiple doses of activated charcoal as well as alkalinization of systemic pH by administering sodium bicarbonate. A salicylate level greater than 7.2 mmol/L is one of several absolute indications for hemodialysis following salicylate overdose. Renal failure, altered mental status, and pulmonary edema are other indications for hemodialysis.

The Role of Brain Imaging in Diagnosis

Brain imaging can be broadly divided into two domains: Structural and functional. Structural imaging provides detailed, noninvasive visualization of the morphology of the brain. Functional imaging provides a visualization of the spatial distribution of specific biochemical processes. Structural imaging includes x-ray, CT, and MRI. Functional imaging includes PET, single-photon emission computed tomography (SPECT), functional MRI (fMRI), and magnetic resonance spectroscopy (MRS). With the limited exception of PET scanning, functional imaging techniques are still research tools that are not yet ready for routine clinical use.

A common clinical dilemma involves whether to obtain an MRI or a CT scan. CT scans usually are less expensive and more commonly available. Although CT scans are a relatively inexpensive screening examination, most clinicians have a hypothesis-based approach to the selection of an imaging modality: The question that they need to be answered drives the selection of the scan. The following subsections provide some guidance regarding the relative advantages of MRI scans and CT scans.

Currently, the use of structural imaging techniques is limited to the identification of nonpsychiatric disease. Scans are performed when a clinician suspects or seeks to rule out a nonpsychiatric illness. Attempts to use structural imaging techniques to subcategorize various psychiatric illnesses have not yielded data that are useful in clinical practice. Consequently, structural imaging techniques are not useful for making psychiatric diagnoses.

Neuroimaging in patients with dementia is associated with some degree of controversy. This is due to the large number of scans that do not have clinically significant findings. Some argue that noncontrast MRI scans or CT scans are useful screening for all dementia patients at the time of initial diagnosis. A contrasting point of view suggests that the probability of significant findings is greater in patients younger than 60 years of age, those with focal findings on the neurologic examination, or recent onset of symptoms.

Findings in patients with dementia may include atrophy, either focal or generalized and white matter lesions. These are nonspecific changes.

Imaging in delirious patients is unlikely to alter management. This is particularly true when there is no history of head trauma or loss of consciousness, and there are no focal findings on the neurologic examination. Imaging may be performed if the delirious patient fails to respond to treatment or has limited responsiveness or alertness. Imaging in delirious patients may reveal underlying factors that may predispose the patient to delirium. However, it is less common that these are etiologic factors.

Magnetic Resonance Imaging. MRI scans are used to distinguish structural brain abnormalities that may be associated with a patient's behavioral changes. These studies provide the clinician with images of anatomical structures viewed from cross-sectional, coronal, or oblique perspectives. MRI scans can detect a large variety of structural abnormalities. The MRI is particularly useful in examining the temporal lobes, the cerebellum, and deep subcortical structures. It is unique in its ability to identify periventricular white matter hyperintensities. MRI scans are useful in examining the patient for particular diseases, such as nonmeningeal neoplasms, vascular malformations, seizure foci, demyelinating disorders, neurodegenerative disorders, and infarctions. Advantages of MRI include the absence of ionizing radiation and the absence of iodine-based contrast agents. Contrast agents are based on gadolinium, a rare earth metal, and usually do not provoke an allergic response.

MRI scans are not always obtained with the same parameters. The nature of the scan is modified in a manner that is most likely to provide data relevant to the clinical hypothesis that generated the request for the scan. For this reason, clinicians should provide the radiologist with clear, concise indications for the scan. Scan parameters include modes referred to as T1 and T2. T1-weighted scans provide detailed visualizations of brain anatomy. Extracellular fluid, such as edema, looks dark. In contrast, T2-weighted scans reveal a white pattern for edema. These scans are useful in examining the patient for white matter disease.

MRI scans are contraindicated when the patient has a pacemaker, aneurysm clips, or ferromagnetic foreign bodies.

Research currently uses a variant of MRI scanning called functional MRI (or fMRI) that is able to measures blood oxygenation levels in the brain while people are performing tasks in the scanner. Such studies have identified potential neural biomarkers of treatment response and the potential to be combined with other measures to personalize treatment for mental illness. Still, other work suggests that neuroimaging may be informative in providing evidence of “bottom-up” or neurophysiologically based methods of defining mental illness, in contrast to current symptom-based approaches.

Computed Tomography. CT scans are used to identify structural brain abnormalities that may contribute to a patient’s behavioral abnormalities. These studies provide the clinician with cross-sectional x-ray images of the brain. CT scans can detect a large variety of structural abnormalities in the cortical and subcortical regions of the brain. CT scans are useful when a clinician is looking for evidence of a stroke, subdural hematoma, tumor, or abscess. These studies also permit the visualization of skull fractures. CT scans are the preferred modality when there is suspicion of a meningeal tumor, calcified lesions, acute subarachnoid or parenchymal hemorrhage, or acute parenchymal infarction. These scans have some resolution limits in the posterior fossa and the inferior surfaces of the temporal lobes owing to artifacts created by the skull. CT scans are not as successful in detecting focal gliosis as MRI scans.

CT scans may be performed with or without contrast media. The purpose of contrast is to enhance the visualization of diseases that alter the blood–brain barrier, such as tumors, strokes, abscesses, and other infections. Contrast agents are administered by IV means. Allergic reactions to contrast agents may occur and can include an anaphylactic reaction with cardiovascular collapse. This type of reaction most commonly occurs in patients with asthma or allergies to seafood.

Indications for a CT scan are not agreed upon in the literature. Less controversial indications include focal findings on the neurologic examination or abnormalities on an EEG. Slightly more controversial indications include impaired cognition on the mental status examination, first episode of psychosis (especially if the presentation is atypical), delirium, late-onset (after 50 years of age) personality disorders, psychosis, or affective illness. Prolonged catatonia is also a possible indication.

The Role of the Electroencephalogram in Diagnosis

The EEG is the assessment of regional cerebral cortical electrical activity. Clinical neuroscience has a long history of using the EEG. The EEG can be used in different ways to study specific brain states or activities by modifications to the technique of data collection or to the data itself. EEG data can be displayed on paper tracings in the manner of conventional EEG recordings. Alternatively, the data can be digitized and transformed to yield color-coded topographic brain maps of regional activity. The collection periods can be prolonged, and the data can be electronically displayed along with video monitoring of the patient to provide telemetry assessments of patients with epilepsy. Telemetry assessments are typically performed in an effort to correlate behavioral abnormalities with brain electrical activity as part of the workup of seizure disorders.

Prolonged periods of EEG recording during sleep, when coupled with a recording of a limited lead ECG and facial muscle activity, result in the sleep EEG or polysomnography. Many clinicians also use the EEG to monitor ECT administration.

Clinicians use the EEG to localize seizure foci and to evaluate delirium. The EEG and its topographical descendants have not found a clear role in the diagnostic assessment of psychiatric patients, although EEG is being investigated as a biomarker of psychiatric disorders. The EEG is usually used in psychiatry to rule out nonpsychiatric disease, such as seizure disorders or delirium, as a cause of psychiatric symptoms. When the differential diagnosis includes strokes, tumors, subdural hematomas, or dementia, the yield is usually higher with imaging tests. Not surprisingly, the yield is the highest in patients with a history of a seizure disorder or a clinical history that is strongly suggestive of a recent seizure or other organic illness. Such clinical features would include a history of altered consciousness, atypical hallucinations (e.g., olfactory), head injury, and automatism. In addition, the EEG is commonly obtained when there is an abnormal CT or MRI. It is important to remember that seizures are a clinical diagnosis; a normal EEG does not rule out the possibility of a seizure disorder.

When a conventional EEG does not demonstrate the dysrhythmia (abnormal bursts of spike, sharp wave, or slow activity) associated with seizures, other measures can be taken to enhance the sensitivity of the evaluation. These measures include sleep deprivation, hyperventilation, and photic stimulation. Detection also can be improved through the use of nasopharyngeal leads, computer-transformed recordings, and evoked potentials (EPs). Sleep deprivation assists in seizure detection by a simple mechanism. Patients are more likely to fall asleep during the EEG after a period of sleep deprivation. It is during the changes in consciousness associated with various stages of sleep that the spike and sharp wave activity associated with seizure disorders is more likely to be detected. Hyperventilation, usually for approximately 3 minutes, may produce slow-wave activity or spike waves. Photic stimulation consists of exposing the patient to a flashing strobe light during the EEG. Abnormal wave activity not synchronized to the flashing light is pathologic. Studies involving the use of nasopharyngeal leads, although uncomfortable for the patients, may increase the yield of positive findings when there is a suspicion of temporal lobe epilepsy.

The EEG can be helpful in patients with altered consciousness by identifying contributing conditions. Nonconvulsive seizures lack motor manifestations but impair consciousness. Nonconvulsive status epilepticus may also cause impairment in consciousness.

Encephalopathies due to underlying metabolic alterations may show diffuse bilateral slowing of the background rhythm along with moderate or high wave amplitude. Specific EEG findings are more commonly seen in some conditions, such as the triphasic waves found in hepatic encephalopathy or periodic sharp wave complexes in Creutzfeldt–Jakob disease.

Evoked Potential. EP testing is the measurement of the EEG response to specific sensory stimulation. The stimulation may be visual, auditory, or somatosensory. During visual EPs, the patient is exposed to flashing lights or a checkerboard pattern. With auditory EP, the patient hears a specific tone. In somatosensory EP, the patient experiences an electrical stimulation to an extremity. These stimuli repeatedly occur while the patient undergoes a routine EEG. Using a computer, the responses to these stimuli are recorded and averaged. The time frame is measured in milliseconds. These tests are useful in neurology and neurosurgery. For example, they assist in the assessment of demyelinating disorders such as MS. In psychiatry, EP testing may help in the differentiation of organic from functional impairments. A classic example is the use of EP testing to evaluate possible functional blindness. The usefulness of these tests in psychiatry is still under investigation.

Polysomnography. Polysomnography is used to assess disorders of sleep by concurrently assessing the EEG, ECG, blood oxygen saturation, respiration, body temperature, electromyogram, and electrooculogram. Polysomnography has demonstrated an increase in the overall amount of rapid eye movement (REM) sleep and a shortened period before the onset of REM sleep (decreased REM latency) in patients with major depression. Polysomnography is useful in diagnosing comorbid sleep disorders so that appropriate treatment can be initiated.

A 57-year-old obese male with recurrent major depressive disorder presents for follow-up evaluation of his depression. He has had a worsening of his mood as well as worsening of fatigue and daytime drowsiness. His wife reports that he snores each night loudly. A sleep study is ordered and completed, and obstructive sleep apnea is diagnosed. The patient's fatigue and daytime drowsiness improve with the use of a continuous positive air pressure (CPAP) machine.

TREATMENT SELECTION AND MONITORING

Medical and laboratory assessments are important not only in the diagnostic assessment of patients with neuropsychiatric symptoms or disorders but are also important in ensuring the safe initiation and monitoring of appropriate treatments. Pretreatment assessments may be necessary to determine the safety of recommended treatments. Many treatments for psychiatric disorders require routine monitoring of drug levels and/or physical health parameters, which may be adversely impacted by treatments. A summary of suggested laboratory assessments for selected psychotropic agents is provided in Table 7.6–2. Finally, suspected complications of psychiatric treatment may also prompt medical and laboratory assessments in psychiatric patients.

Medical Clearance for Electroconvulsive Therapy

Electroconvulsive therapy (ECT) is typically used for patients with severe or treatment-resistant depression, catatonia, psychotic depression, and in some cases of treatment-refractory psychosis. Typical laboratory tests obtained before the administration of ECT include a CBC, serum electrolytes, urinalysis, and LFTs. However, no specific laboratory tests are required in the pre-ECT evaluation. Usually, an ECG is also obtained. A spinal x-ray series is no longer routinely indicated because of the low risk of spinal injury associated with modern administration techniques that use paralyzing agents. A comprehensive medical history and physical examination are useful screening tools to identify possible conditions that could complicate treatment. Patients with a complex or concerning medical history may require a more extensive evaluation prior to initiating ECT; these additional evaluations are usually guided by evaluating consultants with expertise in medically assessing patients' risk level for undergoing procedures under anesthesia.

Electrocardiogram in Treatment Selection and Monitoring

The ECG is a graphical representation of the electrical activity of the heart. Abnormalities in this activity correlate with cardiac pathology. The ECG is most commonly used in psychiatry to assess the side effects of psychotropic medications.

Many psychotropic medications, including antipsychotics and antidepressants, have been associated with prolongation of the QT interval. The corrected QT interval (QTc) is of importance because lengthening of the QTc is associated with subsequent development of a potentially fatal arrhythmia Torsades de pointes. While the QTc interval alone is not a strong predictor of Torsade de pointes, a QTc greater than 450 msec is associated with increased risk in conjunction with other risk factors including older age (>65 and especially >75), female sex, history of arrhythmia, congestive heart failure, cardiac conduction defects, and concurrent use of arrhythmogenic medication. For this reason, clinicians may elect to obtain an ECG before initiation of treatment with agents known to prolong the QT interval, particularly if the patient has other risk factors for prolonged QT interval. Further, bradycardia, hypokalemia or hypomagnesemia, or the concurrent use of other drugs that prolong the QT interval all increase the risk for arrhythmias. When a patient in need of psychotropic medication has a prolonged QT interval, the use of agents with low or no risk for further prolonging the QT interval is recommended when possible. This assessment, and the selection of agents most likely to help but not harm the patient, requires a careful risk/benefit analysis.

Tricyclic or tetracyclic antidepressants (TCAs) are associated with ECG changes. Anticholinergic effects may increase heart rate. Prolongation of the PR, QT, and QRS intervals, along with ST-segment and T-wave abnormalities, may occur. The TCAs can cause or increase pre-existing atrioventricular or bundle branch block. Many clinicians obtain an ECG before beginning a TCA in a patient older than 40 years of age and in any patient with known cardiovascular disease.

Table 7.6–2.
Suggested Laboratory Monitoring for Selected Psychotropic Agents

Psychotropic Agent	Baseline	Routine Monitoring
Atypical antipsychotics	Fasting lipid panel, hemoglobin A1c/fasting blood glucose ECG if patient has risk factors for QT prolongation	At 12 weeks, 1 year, and then annually: fasting lipid panel, fasting blood glucose/hemoglobin A1c
Clozapine	CBC with differential (in addition to baseline labs for atypical antipsychotics)	CBC with differential: weekly for 6 months, every other week for an additional 6 months, and then every 4 weeks, if no abnormalities in WBC or ANC
Lithium	BMP, TSH, CBC, Calcium, pregnancy test ECG if >40 or with known cardiac risk factors	Weekly lithium trough level during dose titration until stable therapeutic drug level is achieved, then every 3–6 months Every 6 months: BMP, TSH, and calcium Therapeutic serum levels: 0.6–1.2 mEq/L (for acute mania: 0.8–1.2 mEq/L)
Valproate	CBC, LFTs, pregnancy test Avoid in women with childbearing potential	Obtain trough valproic acid level 1–2 weeks after initiation and each dosing change, then as clinically indicated. At 6 months and then annually: CBC, LFTs Therapeutic serum levels: 50–125 µg/mL (for acute mania: 85–125 µg/mL)
Carbamazepine	CBC, LFTs, BMP, pregnancy test	Serum carbamazepine levels: obtain weekly levels until stable, then every 3–6 months.

	In patients of Asian ancestry, avoid use or first obtain genetic testing for HLA-B*1502 allele	CBC: 2 weeks after dosage changes, 2 months after initiation, and then every 6 months LFTs and BMP: every 6 months Therapeutic serum levels (in seizure disorders): 4–12 μg/mL
Oxcarbazepine	BMP, pregnancy	Only as clinically indicated
TCAs	ECG if >40 or with known cardiac disease or risk factors	Blood levels as clinically indicated Therapeutic serum levels: Nortriptyline[a]: 50–150 ng/mL Desipramine[a]: >125 ng/mL Imipramine (desipramine + imipramine): >200 ng/mL

[a] frequency of serious side effects may increase over 300 ng/mL.

BMP, Basic metabolic panel; ECG, electrocardiogram; CBC, complete blood count; WBC, white blood cell; ANC, absolute neutrophil count; TSH, thyroid-stimulating; LFT, liver function test.

Lithium therapy can cause benign reversible T-wave changes; rarely, it can impair sinoatrial (SA) node function and can cause heart block. ECGs are often obtained before initiation of treatment with lithium if there are risk factors for coronary artery disease. An ECG is obtained in cases of lithium toxicity or overdose.

In treating patients with eating disorders, it may be reasonable to obtain ECG as patients commonly have low potassium levels that may result in abnormal ECG recordings. As the serum potassium drops below normal, T waves become flattened (or inverted), and U waves may appear.

Metabolic Monitoring

Metabolic syndrome is an important complication that can result from a number of antipsychotics. Consequently, prior to initiating antipsychotic therapy, it is important to assess. Prior to initiating certain psychotropic agents, particularly second-generation antipsychotics, it is important to pursue baseline medical and laboratory assessments, including measurement of weight, blood pressure, fasting glucose or hemoglobin A1c and fasting lipid profile. The risk of metabolic syndrome varies with different second-generation antipsychotic medications, and the choice of medication should be chosen with this in mind.

Many psychotropic drugs, including antidepressants, mood stabilizers, and antipsychotic medications, have weight gain as a potential side effect. Weight gain can lead to obesity, a major risk factor for the development of multiple medical problems such as hypertension, diabetes, cardiovascular disease, cerebrovascular disease, and lipid abnormalities. Weight gain can also contribute to medication nonadherence and poor self-image. Waist circumference measuring greater than 35 in for a woman or 40 in for a man also is of similar concern. Initial intervention to address weight gain typically includes education and initiation of an appropriate diet and exercise program. Discontinuation of the offending medication and substitution, if possible, with a medication less likely to give rise to weight gain also may be an appropriate intervention.

Weight. Obesity is associated with a heightened risk of coronary artery disease, stroke, type II diabetes mellitus, dyslipidemia, hypertension, and osteoarthritis. Treatment with first-generation antipsychotic agents as well as atypical agents is associated with obesity, although there is a significant variation between agents in the propensity to produce obesity. Clozapine and olanzapine are the most likely to produce weight gain. Aripiprazole, ziprasidone, and lurasidone have a substantially lower risk for producing weight gain. All patients with schizophrenia and patients taking atypical antipsychotic medications should have their weight monitored. Likewise, patients with anxiety or mood disorders who take antidepressants or mood stabilizers associated with weight gain should have their weight monitored. Measurement of waist circumference may be used as a risk indicator. Men with a waist larger than 40 in and women with a waist larger than 35 in are at increased health risk. Education about the role of a healthy diet and exercise in preventing weight gain is effective. It is easier to prevent weight gain than it is to lose weight. Many methods proposed to reduce weight are ineffective.

Body Mass Index. BMI values above 25 are associated with an increased risk for the illnesses noted above. Some clinicians advocate regular monitoring of BMI by both patient and clinician for patients with schizophrenia or other patients taking atypical antipsychotic medication and recommend intervention in the form of nutritional counseling, exercise program, or change in antipsychotic medication when BMI increases by one unit.

Monitoring of weight and body mass index (BMI) is important in the management of many patients with psychiatric illness. A variety of psychiatric illnesses can affect appetite and may cause weight loss or weight gain. Measurement of weight may be useful in assessing response to treatment and recovery.

Fasting Blood Sugar. Atypical antipsychotic agents have been associated with abnormalities in serum glucose levels, including the development of diabetes mellitus. Many clinicians monitor their patients who take atypical antipsychotic agents for the development of hyperglycemia by obtaining fasting blood glucose levels and glycosylated hemoglobin levels on a quarterly or semiannual basis. In addition, extremes in serum glucose concentrations have been associated with delirium. Evaluation for diabetes or other abnormalities in glucose metabolism is usually best done by specialists.

Glycosylated Hemoglobin. Glycosylated hemoglobin (hemoglobin A1c) is a correlate of the mean serum glucose concentration over the preceding 8 to 12 weeks. New hemoglobin is formed free of attached glucose. Because red cells are permeable to glucose, the sugar becomes permanently attached to hemoglobin at a concentration proportional to the serum glucose concentration. Thus, hemoglobin A1c is a measure of the mean blood glucose concentration over the prior 8 to 12 weeks. A normal hemoglobin A1c level is below 5.7%. A hemoglobin A1c level of 6.5% or above indicates diabetes mellitus. Levels between 5.7 and 6.4 indicate an increased risk of developing diabetes. For patients with diabetes, a hemoglobin A1c level of 7% or less is a common treatment target.

Lipids. Hyperlipidemia, including elevated cholesterol and triglyceride levels, is associated with cardiovascular and cerebrovascular disease. Certain antipsychotic medications, including clozapine and some of the second-generation antipsychotics, have been associated with hyperlipidemia. Fasting lipid panels should be evaluated prior to initiation of medications associated with dyslipidemia, at 12 weeks into treatment, and annually thereafter.

Psychotropic Medications: Serum Levels and Physical Health Monitoring

Testing to determine blood concentrations of certain psychotropic medications enables the clinician to ascertain whether blood levels of medications are at therapeutic, subtherapeutic, or toxic levels. Psychiatric symptoms are not uncommon when prescribed medications are at toxic levels. Among older adults and patients with severe medical conditions, severe reactions may occur even within typical therapeutic ranges.

Blood levels of psychotropic medications often are obtained when a patient displays mental status changes after a dosage change, after an overdose or suspected overdose, or to assess medication compliance.

Serum concentrations of psychotropic medications are assessed to minimize the risk of toxicity to patients receiving these medications and to ensure the administration of amounts sufficient to produce a therapeutic response. This is particularly true for medications with therapeutic blood levels. Medication levels are often influenced by hepatic metabolism or renal clearance.

Cytochrome P450 enzymes (CYPs) are important in the hepatic metabolism of drugs. Some drug substrates induce or inhibit CYP activity. Induction results in acceleration of drug metabolism and a resultant decrease in the levels of agents metabolized by that particular CYP. Administration of drugs that inhibit CYPs may result in higher drug levels of agents metabolized by the inhibited CYP. Genetic variation also influences CYP activity. Genotypes at a particular CYP, such as CYP2D6, may lead to any of the following phenotypes: ultrarapid metabolizer, extensive metabolizer, intermediate metabolizer, or poor metabolizer. At usual doses of a given medication, individuals with the poor metabolizer phenotype may have high serum levels of CYP2D6 metabolized drugs due to the poor ability of CYP2D6 to metabolize the drug. Inversely, ultrarapid metabolizers may have very low levels of a drug due to rapid metabolism.

Antipsychotic Agents. Serum concentrations are not routinely obtained for most antipsychotic medications. Management usually consists of titration of the prescribed dose to levels that achieve reasonable symptomatic control and acceptable side effects. However, in some cases, blood levels may be helpful. These situations include suspected medication noncompliance and cases of poor response in the context of clinically adequate doses. In the latter scenario, the concern may include rapid hepatic metabolism.

Atypical antipsychotic agents may be associated with the development of metabolic syndrome. Therefore, appropriate laboratory monitoring should occur at regular intervals. At the initiation of treatment with an atypical antipsychotic medication, a clinician should obtain a fasting lipid profile and fasting blood glucose or hemoglobin A1c. This assessment should be repeated in 3 months and annually thereafter. Lipid profile and blood glucose should be repeated sooner if the patient experiences significant weight gain or symptoms suggestive of diabetes. A measurement of weight, BMI, and waist circumference also should be obtained at the initiation of treatment with an antipsychotic agent. These variables should be assessed monthly for the first 3 months of treatment and then at quarterly intervals. Baseline assessment should also include a personal or family history of obesity. Many clinicians will switch to a different antipsychotic agent when a 5% or more weight gain has occurred. Alternatively, if a medication change is not clinically indicated, adjunctive treatment with metformin may be considered.

CLOZAPINE. Clozapine (Clozaril) levels are trough levels determined in the morning before administration of the morning dose of medication. Though a therapeutic range for clozapine has not been established, a level of 100 ng/mL is widely considered to be the minimum therapeutic threshold. Although concentrations between 200 and 700 ng/mL may correlate more with response, nonresponse does occur within this range. Studies suggest that plasma levels of clozapine in the range of 200 to 400 ng/mL may lead to the achievement of maximal clinical efficacy; however, there is great variability among patients in terms of response and side effects at given dosages. One study demonstrated that patients with refractory schizophrenia who responded to clozapine were more likely to have clozapine blood levels of greater than 350 ng/mL. The risk of seizures increases with clozapine dosages of 600 mg/day or greater.

Clozapine is a common cause of leukopenia in psychiatry. A baseline CBC must be determined prior to the initiation of treatment. The baseline WBC count must be at least 3,500/mm^3, and the absolute neutrophil count (ANC) must be at least 2,000/mm^3 for initiation of therapy. Weekly CBCs are required during the first 6 months of treatment with clozapine because of the risk of agranulocytosis. Results must be sent to the dispensing pharmacy for the patient to receive his or her medication. Clozapine should be held for a WBC of less than 3,000/mm^3 or an ANC of less than 1,500/mm^3. For WBC 3,000 to 3,500/mm^3 and/or ANC 1,500 to 2,000/mm^3, clozapine may be continued, but the frequency of WBC and ANC monitoring is increased to twice weekly. If the WBC remains greater than 3,500/mm^3 and the ANC remains above 2,000/mm^3 during the first 6 months of clozapine treatment, then monitoring may be reduced to every other week for the next 6 months. If both the WBC and the ANC remain at appropriate levels after the second 6 months of treatment, then monitoring can be further decreased to every 4 weeks.

After clozapine discontinuation, clinicians usually monitor the CBC for an additional 4 weeks if laboratory values are within normal limits. If clozapine treatment was interrupted or discontinued due to abnormal laboratory values, the degree of severity of leukopenia or granulocytopenia determines whether daily, twice-weekly, or weekly laboratory monitoring is required. Ongoing monitoring is required until the WBC is at least 3,500/mm^3 and the ANC is at least 2,000/mm^3, and for at least 4 weeks from the date of clozapine discontinuation.

When moderate leukopenia or granulocytopenia develops (WBC between 3,000 and 2,000/mm^3 and/or an ANC between 1,500 and 1,000/mm^3), clozapine treatment must be interrupted, but the patient may be retreated with clozapine in the future. Repeat exposure to clozapine may occur once the WBC is at least 3,500/mm^3 and the ANC is at least 2,000/mm^3. These patients require weekly monitoring for a year after the reinstatement of clozapine. Subsequently, the frequency of monitoring may be reduced to every other week for 6 months and then further reduced to every 4 weeks. If granulocytopenia or leukopenia are severe (WBC <2,000/mm^3 and/or ANC <1,000/mm^3), the patient should not be rechallenged with clozapine in the future, regardless of WBC and ANC recovery.

If there is a break in clozapine treatment lasting less than 1 month during the period of weekly monitoring and the WBC consistently has been greater than 3,500/mm^{3,} and the ANC consistently has been greater than 2,000/mm^3, then treatment may resume with weekly monitoring. In addition, the patient will be eligible to reduce the frequency of monitoring to every other week after 6 months of total treatment rather than after 6 months from the time of the break in treatment. However, if the break in clozapine treatment lasts more than 1 month, then the patient must have a continuous 6 months of weekly monitoring without abnormalities before reducing the frequency of monitoring to every other week.

A similar rule applies for patients on clozapine for at least 6 months. If there is a break in clozapine treatment lasting less than 1 month in a patient undergoing CBC monitoring every other week and the WBC has been consistently greater than 3,500/mm^3 and ANC consistently has been greater than 2,000/mm^3, then the patient must undergo weekly monitoring for 6 weeks. Thereafter, the patient may resume every other week monitoring. However, if the break in treatment is longer than 1 month, then the patient must resume weekly monitoring for 6 months.

For a patient on clozapine for more than 1 year, a similar rule applies. If there is a break in clozapine treatment lasting less than 1 month in a patient undergoing monitoring every 4 weeks and WBC consistently has been greater than 3,500/mm^3 and ANC consistently has been greater than 2,000/mm^3, the patient has been taking clozapine for longer than 1 year, and the gap in treatment is for less than 30 days, then they may resume treatment with weekly monitoring for 6 weeks and then return to monthly monitoring. However, if the treatment gap has been longer than 30 days, then monitoring must be weekly for 6 months, then every other week for another 6 months, and then monthly monitoring may resume.

CLOZAPINE-ASSOCIATED MYOCARDITIS OR CARDIOMYOPATHY. Infrequently, clozapine treatment may precipitate myocarditis or cardiomyopathy. If patients taking clozapine develop new cardiac or respiratory symptoms concerning for myocarditis or cardiomyopathy, a medical evaluation should ensue with the assistance of specialists. Medical and laboratory assessment may include ECG, evaluation of cardiac biomarkers such as troponin, and echocardiography. Echocardiography is the visualization of cardiac anatomy by the use of computer-transformed echoes of ultrasound.

Psychiatry was consulted to evaluate a 36-year-old woman with schizophrenia who had been started on clozapine approximately 1 month ago and had been admitted to medicine 1 day ago with fever, tachycardia, and dyspnea. The primary team was concerned about NMS due to her autonomic symptoms. In addition to holding the clozapine, the psychiatrist recommended a cardiology evaluation due to the possibility that the patient's symptoms represented clozapine-induced myocarditis. The patient had nonspecific ECG changes, elevated cardiac troponins, and left ventricular dysfunction on echocardiogram; myocarditis was diagnosed.

NEUROLEPTIC MALIGNANT SYNDROME. Neuroleptic malignant syndrome (NMS) is a rare, potentially fatal consequence of neuroleptic administration. The syndrome consists of altered mental status, generalized rigidity (often severe), hyperthermia, and autonomic instability. Sustained muscle contraction, as well as impaired central mechanisms of thermoregulation, result in peripheral heat generation and muscle breakdown. Muscle breakdown contributes to elevated levels of creatine kinase (CK). Leukocytosis and low serum iron concentration are common. Myoglobinuric acute renal failure may occur due to rhabdomyolysis. Liver enzymes may become elevated. Patients may die from hyperpyrexia, aspiration pneumonia, renal failure, hepatic failure, respiratory arrest, or cardiovascular collapse. Treatment includes discontinuation of the neuroleptic, hydration, and general supportive care. Administration of dantrolene (a muscle relaxant) or bromocriptine (a dopamine agonist) may be considered, along with other treatments that have anecdotal success, including amantadine, benzodiazepines, or ECT. Data is limited with regard to the efficacy of any of these specific therapies.

A typical laboratory workup of NMS includes a CBC, serum electrolytes, BUN, Cr, and CK, as well as evaluation for other serious neurologic or medical conditions that may be mistaken for NMS. A urinalysis, including an assessment of urine myoglobin, is also usually performed. As part of the differential diagnosis, blood and urine cultures are performed as part of a fever workup. Pronounced elevations in the WBC may occur in NMS. WBCs are typically in the range of 10,000 to 40,000/mm^3.

CK elevations are typically more than 1,000 IU/L. Moderate elevations in CK may occur in the earliest stages of the disease. CK level may be normal or near-normal if significant rigidity has not yet developed. Other causes of muscle injury leading to increases in serum CK levels may include repeated intramuscular (IM) injections, prolonged or agitated periods in restraint, or dystonic reactions from neuroleptic administration. Importantly, elevated CK levels in the absence of hyperthermia and rigidity are not indicative of NMS.

Mood Stabilizers

CARBAMAZEPINE. Carbamazepine (Tegretol) may produce changes in the levels of WBCs, platelets, and, under rare circumstances, red blood cells. Aplastic anemia, thrombocytopenia, agranulocytosis, exfoliative dermatitis, and hepatic failure may all occur but are rare. Hyponatremia is common, particularly in the older adults. Pretreatment evaluation typically includes a CBC with differential and platelet count, as well as LFTs and renal function tests. Given the risk for hyponatremia, baseline serum electrolytes may also be obtained, particularly in more vulnerable populations such as the older adults.

The risk of exfoliative dermatitis (e.g., Stevens–Johnson syndrome) is greater in patients with the HLA-B*1502 allele. This allele is more common in patients of Asian descent and is largely absent in individuals of other ethnic backgrounds. Prior to initiating carbamazepine, genetic testing for this allele should be pursued in patients with ancestry, which puts them at a higher likelihood of having the HLA-B*1502 allele. If this allele is identified, that patient should not be treated with carbamazepine. Many clinicians monitor the CBC and LFTs every 2 weeks for the first 2 months of administration. Then, if the laboratory values have been within normal limits, CBC, serum carbamazepine level, and LFTs are monitored every 3 to 6 months. Patients should be cautioned to watch for petechiae, fever or signs of infection, pallor, or undue weakness, as well as signs and symptoms of hepatic injury or dermatologic reactions.

Carbamazepine is teratogenic and may produce a variety of congenital abnormalities, including neural tube defects. A pretreatment urine pregnancy test usually is obtained in women of child-bearing years, and women should be cautioned to use adequate contraception when taking carbamazepine. Carbamazepine may decrease plasma concentrations of hormonal contraceptives.

When used as an anticonvulsant, target blood levels for carbamazepine are 4 to 12 μg/mL. Target blood levels have not been established for mood disorder patients, but most clinicians aim for a level in this range before discontinuation of carbamazepine owing to lack of efficacy. Levels should be obtained before the morning dose and after the patient has taken the same total daily dose for a minimum of 5 days. Serum levels of carbamazepine can help establish an effective dose for an individual patient. Toxic side effects are common with levels greater than 15 μg/mL.

Manifestations of toxicity may include nausea, vomiting, urinary retention, ataxia, confusion, drowsiness, agitation, or nystagmus. Patients who exhibit neurocognitive changes while taking carbamazepine should also be checked for hyponatremia, which is associated with carbamazepine and its analogs oxcarbazepine and eslicarbazepine. At very high levels, symptoms also may include cardiac dysrhythmias, seizures, and respiratory depression.

Carbamazepine levels may be altered by the coadministration of medications that inhibit the metabolism of carbamazepine. Such agents include erythromycin and related antibiotics, valproic acid, cimetidine, propoxyphene, and isoniazid, among other CYP 3A4 inhibitors. Carbamazepine is a powerful CYP 3A4 inducer and thus must be used cautiously and with careful monitoring when combined with other medications metabolized by CYP 3A4. Further, carbamazepine has the ability to lower its own serum level through the process of hepatic enzyme autoinduction. Typically, autoinduction of hepatic enzymes occurs several weeks after initiation of carbamazepine treatment. Many clinicians follow levels over time and adjust the administered dose as necessary to maintain a therapeutic concentration. In malnourished patients or those with low protein due to other medical conditions, it may be advisable to monitor serum albumin levels. A portion of the drug is bound to serum proteins. It is the free (unbound) component that is biologically active. When there are reductions in serum proteins that bind the drug, an adjustment in the dose may be necessary, particularly in patients who show signs of toxicity.

OXCARBAZEPINE. Oxcarbazepine is structurally similar to carbamazepine but has fewer drug interactions and fewer side effects, with the exception of increased frequency of clinically significant hyponatremia. Many clinicians obtain baseline serum electrolytes and repeat this testing as clinically indicated. Therapeutic serum oxcarbazepine levels have not been established. In women of childbearing age, a pretreatment pregnancy test should be evaluated; women should be advised that hormonal contraception may be less effective during oxcarbazepine treatment.

A 58-year-old woman with bipolar I disorder was recently transitioned from carbamazepine to oxcarbazepine due to poor tolerability of carbamazepine. One month after initiation of oxcarbazepine, she and her spouse reported that she was having difficulty with memory and periods of acute confusion. She appeared drowsy in the office and had poor attention. Serum electrolyte evaluation revealed sodium of 118 mEq/L. She was admitted to internal medicine for treatment, and oxcarbazepine was discontinued.

LAMOTRIGINE. Therapeutic serum lamotrigine levels have not been established. Dosing is titrated to clinical response.

LITHIUM. Lithium has a narrow therapeutic index. Consequently, blood levels of lithium must be monitored to achieve therapeutic dosing and avoid toxicity. For acute mania, target serum levels of lithium are in the range of 0.8 to 1.2 mEq/L. Optimum maintenance serum levels may vary from patient to patient, with the goal of achieving a balance between clinical efficacy and tolerability. Many clinicians aim for maintenance levels between 0.6 to 0.8 mEq/L, especially if patients complain of side effects at levels of 0.8 mEq/L or higher. Side effects are dose dependent. Symptoms of toxicity may begin to manifest with serum levels of greater than 1.2 mEq/L and are common with levels greater than 1.4 mEq/L. Older or seriously medically ill patients may show signs of toxicity with levels less than 1.2 mEq/L. With high levels (e.g., >2 mEq/L), consultation with a nephrologist or internist should occur. In patients with high serum lithium levels, dialysis is preferable to reduce lithium levels because prolonged exposure increases the likelihood of irreversible cerebral damage. Symptoms of toxicity include tremors, sedation, and confusion. At higher levels, delirium, seizures, coma, and death may occur.

Steady-state serum lithium levels are achieved after five half-lives, and lithium levels typically reach steady-state 5 days after a change in dosage in patients with normal renal function. Levels are obtained as close to 12 hours after the last dose as possible and are usually measured using samples obtained in the morning before administration of the morning dose (trough level). In cases of suspected lithium toxicity, levels are obtained sooner.

Maintenance levels are checked every 3 to 6 months and sometimes more frequently in patients with compromised renal function, general debilitation, or signs of toxicity. Close monitoring is required for patients who take medications that may alter lithium levels, such as diuretics or certain nonsteroidal anti-inflammatory drugs (NSAIDs).

Baseline pretreatment studies and periodic follow-up tests for lithium should include monitoring of renal and thyroid function. Renal function is monitored using BUN, serum Cr, and a urinalysis. Because lithium commonly affects thyroid function, a baseline screening assessment of thyroid function (TSH) is obtained. These tests are repeated once or twice during the first 6 months of treatment and subsequently every 6 months to every year during ongoing treatment. Clinicians also should obtain thyroid studies in patients taking lithium who develop fatigue or asthenia, cold intolerance, or changes in the skin, voice, or weight (symptoms of impaired thyroid function). Many clinicians measure calcium levels at baseline and every 6 months to 1 year, given the association of lithium treatment with parathyroid dysfunction and hypercalcemia. Also, because lithium commonly produces a mild leukocytosis, a pretreatment WBC is often obtained.

A 48-year-old man with schizoaffective disorder had been stably managed with lithium and risperidone for the past 3 years. He was recently medically hospitalized for evaluation of chest pain. The primary team requested a hematology/oncology evaluation due to concern over his persistently elevated WBC ranging from 12,000 to 16,000/mm^3. Upon medical record review, the consulting psychiatrist noted that prior to the initiation of lithium, his WBC had been within normal limits. The team was advised of this common side effect of lithium for their consideration in formulating the differential diagnosis.

Before the administration of lithium in women of childbearing age, a screening urine pregnancy test should generally be obtained. The use of lithium during the first trimester of pregnancy requires a careful risk/benefit analysis, as there is an increased risk of congenital heart disease with fetal exposure to lithium. Lithium levels should be carefully monitored during pregnancy and after labor. Rapid shifts in fluid volume can give rise to rapid changes in lithium levels during this time.

In patients over age 40 or with known risk factors for cardiovascular disease, a pretreatment ECG is recommended. Many clinicians monitor ECG on a yearly basis; guidelines vary.

VALPROATE. Valproate (valproic acid [Depakene] and divalproex [Depakote]) is used to treat acute mania and for maintenance treatment of bipolar disorder. Therapeutic levels for acute mania are thought to range from 50 to 125 mcg/mL; based upon analysis of clinical trial data, most clinicians aim for a valproate level between 85 and 125 mcg/mL for acute mania. During maintenance treatment, clinicians work individually with patients to achieve a dose that is within a therapeutic range but remains tolerable. Side effects are more common with levels greater than 100 mcg/mL.

Because of the risk of hepatotoxicity and hematologic abnormalities during valproate treatment, baseline LFTs and hematologic measures are advised. Irreversible hepatic failure and agranulocytosis are rare, idiosyncratic effects of valproate; patients should be advised to report immediately for evaluation should they develop concerning symptoms. Hepatic necrosis is rare and is more likely to occur in the very young (younger than 2 years of age) and in individuals taking multiple anticonvulsants. More commonly, valproate may cause an asymptomatic elevation in liver transaminase levels and mild leukopenia or thrombocytopenia. Most clinicians monitor CBC and LFTs every 6 to 12 months during valproate treatment.

Valproate increases the risk of birth defects, including neural tube defects (with rates of 3% to 5%), craniofacial abnormalities, limb malformations, and cardiac defects. A pretreatment urine pregnancy test should be obtained in women of childbearing years. Women should be cautioned to use adequate contraception. Careful consideration of the known teratogenic effects of mood-stabilizing agents is essential when initiating treatment for a woman of childbearing age, regardless of whether or not she is currently pregnant. Given teratogenicity, many clinicians avoid prescribing valproate to any woman who has childbearing potential.

Treatment with valproate may increase serum ammonia levels, even in the absence of LFT abnormalities. It is prudent to obtain an ammonia level in a patient undergoing valproate treatment who presents with altered mental status or lethargy. Acute pancreatitis may also occur but is rare.

Antidepressants

MONOAMINE OXIDASE INHIBITORS. Treatment with monoamine oxidase inhibitors (MAOIs) can cause orthostasis and, rarely, hypertensive crisis as a result of drug–drug or drug–food interactions. Baseline blood pressure measurement should be obtained before the initiation of treatment, and blood pressure should be monitored during treatment as clinically indicated. There are no meaningful blood levels for MAOIs, and direct monitoring of MAOI blood levels is not clinically indicated.

TRICYCLIC AND TETRACYCLIC ANTIDEPRESSANTS. TCAs can affect cardiac conduction. Thus, patients should be screened for a history of cardiac disease or cardiac risk factors that may lead to conduction abnormalities prior to TCA administration. Patients with risk factors for cardiac disease or who are over the age of 40 should have a baseline ECG to rule out existing conduction abnormalities. Heart rate may increase as a result of anticholinergic effects of the medication. In an overdose, a QRS interval of longer than 100 milliseconds may predict the development of seizures and ventricular arrhythmias and is an indication for bicarbonate therapy.

Baseline vital signs with orthostatic measurements should be obtained before the initiation of TCAs, especially in the older adults in whom TCA treatment can significantly increase the risk of hip fracture owing to TCA side effects of orthostasis. Subsequent orthostasis measurements should be obtained to assist with dose titration in this population.

Plasma levels of TCAs may be useful when there is a need to rapidly achieve a therapeutic dose of the medication, when patients need the lowest effective dose (e.g., sensitivity to side effects), when questions exist about a patient’s compliance, and when there is a poor response to a standard dose. Other indications for obtaining plasma TCA levels include concomitant administration of medications that may alter TCA blood levels or the re-emergence of depression in a patient who had an initial therapeutic response to TCA treatment.

TCA levels usually are obtained 12 hours after the administration of a dose. Typically, this is in the morning before administration of the morning dose. Levels are most meaningful when a steady-state has been achieved, at least five half-lives after the last change in dose. Meaningful serum levels have been determined only for nortriptyline (50 to 150 ng/mL), desipramine (>125 ng/mL), and imipramine (total of imipramine and desmethylimipramine [desipramine] >200 ng/mL).

CONCLUSION

An important role of the psychiatric clinician is to maintain vigilance for medical or neurologic conditions that may mimic primary psychiatric disorders and to order or recommend appropriate studies to evaluate for such conditions when indicated. Additionally, the safe provision of psychiatric care often requires evaluating for potential toxicity of certain agents and for the monitoring of particular laboratory studies over a period of time, at regular intervals, depending upon the treatment provided. Whether utilizing laboratory or imaging studies for screening, diagnosis, or treatment monitoring, it is important to judiciously select tests and carefully interpret results, acknowledging that no test has perfect sensitivity, specificity, and/or positive or negative predictive value. It is preferable to select and order tests based on the ability of the test results to change management rather than to routinely order a host of screening tests that may be costly and unlikely to meaningfully impact patient care.

FURTHER READINGS

Allen MH, Hirschfeld RM, Wozniak PJ, Baker JD, Bowden CL. Linear relationship of valproate serum concentration to response and optimal serum levels for acute mania. *Am J Psychiatry*. 2006;163(2):272–275.

*American Diabetes Association; American Psychiatric Association; American Association of Clinical Endocrinologists; North American Association for the Study of Obesity. Consensus development conference on antipsychotic drugs and obesity and diabetes. *J Clin Psychiatry*. 2004;65:267.

*American Psychiatric Association. *APA Practice Guidelines for the Treatment of Psychiatric Disorders: Compendium*. American Psychiatric Association; 2006.

Barnes TRE, Paton C, Cavanagh M-R, Hancock E, Taylor DM; UK Prescribing Observatory for Mental Health. A UK audit screening for the metabolic side effects of antipsychotics in community patients. *Schizophr Bull*. 2007;33:1397.

*Beach SR, Celano CM, Noseworthy PA, Januzzi JL, Huffman JC. QTc prolongation, torsades de pointes, and psychotropic medications. *Psychosomatics*. 2013;54(1):1–13.

Bhatia MS, Munjal S. Prevalence of depression in people living with HIV/AIDS undergoing ART and factors associated with it. *J Clin Diagn Res*. 2014;8(10):WC01–4.

Branson BM, Handsfield HH, Lampe MA, et al. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. *MMWR Recomm Rep*. 2006;55(RR-14):1–17; quiz CE11-14.

Brooks JO 3rd, Chang H-S, Krasnykh O. Metabolic risks in older adults receiving second-generation antipsychotic medication. *Curr Psychiatry Rep*. 2009;11:33–40.

Byerly M, Suppes T, Tran Q-V, Baker RA. Clinical implications of antipsychotic-induced hyperprolactinemia in patients with schizophrenia spectrum or bipolar spectrum disorders: recent developments and current perspectives. *J Clin Psychopharmacol*. 2007;27:639–661.

Carr RB, Shrewsbury K. Hyperammonemia due to valproic acid in the psychiatric setting. *Am J Psychiatry*. 2007;164:1020–1027.

Centers for Disease Control and Prevention (CDC). Testing for HCV infection: an update of guidance for clinicians and laboratorians. *MMWR Morb Mortal Wkly Rep*. 2013;62(18):362–365.

Chennapan K, Mullinax S, Anderson E, et al. Medical screening of mental health patients in the emergency department: a systematic review. *J Emerg Med*. 2018;55(6):799–812.

Clifford DB, Ances BM. HIV-associated neurocognitive disorder. *Lancet Infect Dis*. 2013;13(11):976–986.

Crismon ML, Argo TR, Bendele SD, Suppes TS. *Texas Medication Algorithm Procedural Manual: Bipolar Disorder Algorithms*. Texas Department of State Health Services; 2007.

Dalmau J, Armangué T, Planagumà J, et al. An update on anti-NMDA receptor encephalitis for neurologists and psychiatrists: mechanisms and models. *Lancet Neurol*. 2019;18(11):1045–1057.

Dalmau J, Graus F. Antibody-mediated encephalitis. *N Engl J Med*. 2018;378(9):840–851.

Goldberg RW, Kreyenbuhl JA, Medoff DR, et al. Quality of diabetes care among adults with serious mental illness. *Psychiatr Serv*. 2007;58:536–543.

Graus F, Titulaer MJ, Balu R, et al. A clinical approach to diagnosis of autoimmune encephalitis. *Lancet Neurol*. 2016;15(4):391–404.

Hannuksela M, Liisanantti MK, Nissinen AET, Savolainen MJ. Biochemical markers of alcoholism. *Clin Chem Lab Med*. 2007;45:953–961.

Janiak BD, Atteberry S. Medical clearance of the psychiatric patient in the emergency department. *J Emerg Med*. 2012;43(5):866–870.

Jiang W-L, Cai D-B, Yin F, et al. Adjunctive metformin for antipsychotic-induced dyslipidemia: a meta-analysis of randomized, double-blind, placebo-controlled trials. *Transl Psychiatry*. 2020;10(1):117.

Kaplan EB, Sheiner LB, Boeckmann AJ, et al. The usefulness of preoperative laboratory screening. *JAMA*. 1985;253(24):3576–3581.

LeFevre ML; U.S. Preventive Services Task Force. Screening for hepatitis B virus infection in nonpregnant adolescents and adults: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;161(1):58–66.

*Levenson JL, ed. *Textbook of Psychosomatic Medicine: Psychiatric Care of the Medically Ill*. 2nd ed. American Psychiatric Publishing, Inc.; 2011.

*Marder SR, Essock SM, Miller AL, et al. Physical health monitoring of patients with schizophrenia. *Am J Psychiatry*. 2004;161:1334.

National Collaborating Centre for Mental Health (UK). *Bipolar Disorder: The NICE Guideline on the Assessment and Management of Bipolar Disorder in Adults, Children and Young People in Primary and Secondary Care*. London: The British Psychological Society and The Royal College of Psychiatrists; 2014 Sep.

Newcomer JW, Hennekens CH. Severe mental illness and risk of cardiovascular disease. *JAMA*. 2007;298:1794–1796.

Pfeiffer RF. Wilson's disease. *Semin Neurol*. 2007;27:123.

Smith FA, Wittmann CW, Stern TA. Medical complications of psychiatric treatment. *Crit Care Clin*. 2008;24(4):635–656, vii.

Wu L-T, Pilowsky DJ, Sclenger WE. Inhalant abuse and dependence among adolescents in the United States. *J Am Acad Child Adolesc Psychiatry*. 2004;43:1206–1214.

7.7 Principles and Applications of Quantitative Electroencephalography in Psychiatry

LESLIE S. PRICHEP, PH.D.

DEDICATION

This chapter is dedicated to the legacy of E. Roy John, Ph.D., whose pioneering innovation and insights into brain function are at the foundation of the work presented herein and have significantly advanced the clinical utility of quantitative electrophysiology. Dr. John was a co-author in prior editions of this chapter.

This section is intended to provide a primer for the underlying neurophysiologic theory, basic concepts, and nomenclature of quantitative electroencephalography (qEEG) and an overview of clinical applications of qEEG in neuropsychiatry. This chapter reviews the signal processing methods used to extract quantitative features from the raw EEG, convert them to standard scores, display them as qEEG-based tomographic brain images, and use them as input to classification algorithms, which together reduce the profuse electrical phenomenology of the brain to more objective, anatomically localized, and clinically useful descriptors of abnormal brain function. A clinically oriented section describes statistical methods used to seek distinctive patterns or profiles of qEEG features for patients with specific disorders. This section (1) explains the use of machine learning derived statistical classifier methods for identification of distinctive profiles of abnormality for groups of patients diagnosed using the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) (edition according to time the patients were tested); (2) gives examples of the performance and clinical utility of qEEG-based classification algorithms; (3) explains cluster analytic methods used to identify pathophysiologic heterogeneity (subtypes/phenotypes) within clinically homogeneous cohorts of patients with the same DSM diagnosis; (4) summarizes applications of the performance and clinical utility of qEEG subtyping methods in clinical optimization of treatment decisions; and (5) describes the uses of qEEG source localization in expanding the understanding of underlying pathophysiology of disorders. Finally, research on the neurophysiology of anesthesia and information processing is integrated to review the current understanding of brain mechanisms that bind sensations, emotions, motivation, and memories into perceptual experiences that become the momentary content of conscious awareness, to describe processes that mediate how affect and mood modulate subjective reactions to these experiences, and to explore how dysfunction of these systems can contribute to psychiatric illnesses.

FROM EEG TO qEEG

Background and History

In 1928 the human EEG was first recorded by the German psychiatrist Hans Berger, who believed that its study would reveal information about psychiatric disorders and cognitive functions and revolutionize the practice of psychiatry. However, for over half a century, its major clinical application was in the arena of detection and treatment of seizure disorders. By 1938, Berger's discovery had aroused so much interest among neurologists and psychiatrists that an International Federation of Electroencephalographers agreed upon a standard method for positioning 19 electrodes upon the head, the International 10/20 Electrode Placement System, used to this day. After extensive neuroanatomical studies, this proportional method locates each electrode approximately above a specified functional region or Brodmann cytoarchitectural area of the brain independent of the absolute size of the head, defining positions as percentages of arcs measured on the head. The locations of this system are shown schematically in Figure 7.7–1.

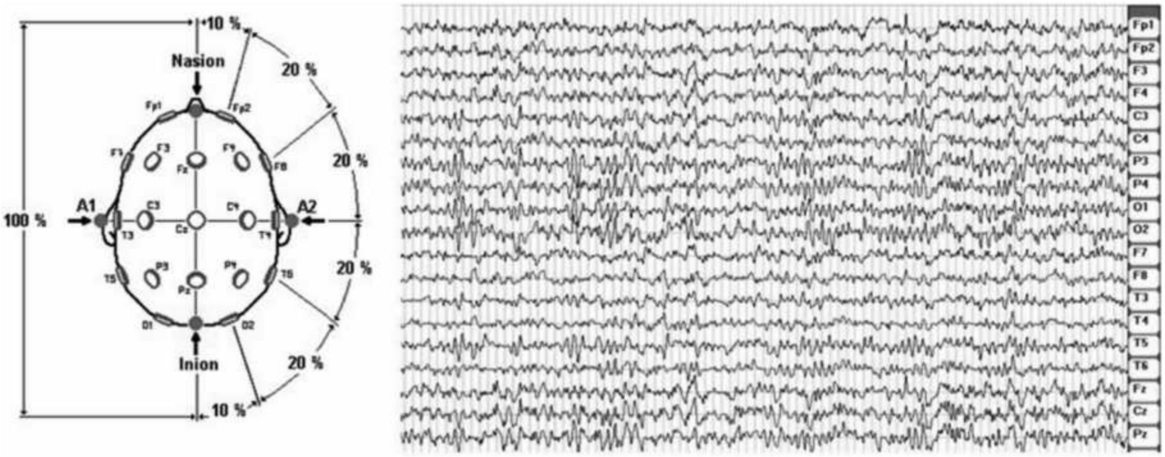

FIGURE 7.7–1. (*Left*) A schematic of the standardized electrode positions of the International 10/20 Electrode Placement System that locates electrodes at distances proportional to 10% or 20% of the circumference and several arcs measured from side to side and from nasion to inion on the head, treated as an ellipsoid is shown. (*Right*) Eight seconds of EEG data from the screen of the computer, showing EEG data from each electrode location referenced to linked ears.

Contrary to Berger’s hopes, insights from the EEG into the underlying causes of psychiatric disorders or how to improve the clinical treatment of psychiatric patients were not readily evident. However, its utility for the detection, diagnosis, and treatment of a wide variety of neurologic disorders became rapidly apparent. Using visual examination of oscilloscopic analog recordings of voltage fluctuations recorded from the scalp, generations of electroencephalographers who were focused primarily on neurologic illnesses visually examined tracings and sought to identify distinctive wave shapes and spatial patterns related to their clinical experience. After the introduction in around 1970 of computer methods to record and analyze the EEG with readily available digital computers, digital electroencephalography (DEEG) began gradually to replace the conventional analog recording instruments, with great advantages of compact and economic digital media replacing voluminous paper recordings.

qEEG was introduced in the next decade. qEEG augmented visual analysis by computation of the power spectrum to analyze the EEG oscillations from each region on the scalp, replacing tracings with an objective set of numbers that quantified the distribution of power as a function of frequency. During the next 35 years, psychiatrists and psychologists began to study the power spectrum, and the relevance of qEEG to the study of psychiatric and neuropsychiatric disorders, as well as of brain functions and psychophysiological processes, became more and more evident. qEEG measures were expressed as numbers that quantified features extracted from recordings at each electrode position, such as the absolute amount of power within specified frequency ranges or "bands," the percentage of the total power that falls in each band (relative power), the synchronization or symmetry of such measures between corresponding or "homologous" regions on the two hemispheres, and the relationships of electrical transactions among regions between and within each hemisphere ("connectivity"). Over the last decade, methods derived from signal processing advances and analytic methods (e.g., from genomics, proteomics, and machine learning [ML]) were also applied to the EEG, yielding further feature descriptors and expanding the underlying processes that could be described. qEEG methods have yielded objective data about brain electrical correlates of information processing, cognitive processes and cognitive impairments of aging and in childhood, and alterations of consciousness such as during anesthesia and sleep, as well as in psychiatric illness and revealed distinctive qEEG effects of various classes of psychotropic drugs. Such research provided important insights into neurophysiologic processes mediating attention, sensation, perception, behavior, and cognition.

Distinctive patterns of qEEG features have been described for a wide variety of developmental, neurologic, and psychiatric disorders, akin to the clinical significance of particular abnormal rhythms and patterns in the ECG. Some of these findings have been reported from the relational database of the Brain Research Laboratories of the Department of Psychiatry at the New York University School of Medicine, one of the largest qEEG quantitative electrophysiologic databases in the world, which contains "raw" unprocessed EEG and event-related potential (ERP) recordings, extracted features and clinical data from over 15,000 examinations of patients and normal functioning individuals, collected between 1980 and 2015. This database includes EEG and multimodal evoked related potential (ERP) recordings from patients with many different diagnostic conditions, including attention-deficit/hyperactivity disorder (ADHD), alcoholism, autism spectrum disorder (ASD), and Asperger syndrome, cognitively impaired older adult patients (with a staging of cognitive deterioration ranging from mild to severe in Alzheimer-type dementia, AD), depression, drug dependence, intractable chronic pain, learning disabilities, obsessive-compulsive disorder (OCD), schizophrenia, stroke, tinnitus, and traumatic brain injury (TBI). Also included are data from patients in coma and persistent vegetative state (PVS), from patients undergoing many different neurosurgical and cardiovascular surgical procedures using a wide variety of anesthetic agents, and from healthy, normally functioning children and adult normally functioning control subjects. These records were stored digitally in their original, unprocessed form, which could be readily retrieved for analysis with new mathematical algorithms as they are developed or to test new theoretical hypotheses and constitute a major scientific resource. Multiple collaborative projects with leading research groups in many countries steadily dynamically increased the scope of this database.

For several of these disorders, profiles of deviant qEEG findings have been found that are either distinctive for patients diagnosed with a particular DSM disorder, predictive of the evolution of the condition, or indicative of differential response to treatment, discussed in more detail below. Certain deviations from this baseline have been found to occur invariantly and reversibly with loss and return of awareness during anesthesia or coma that offer insights into brain mechanisms related to the production and maintenance of consciousness, further discussed later in this section.

NEUROMETRIC qEEG: HOMEOSTATIC REGULATION OF THE EEG FREQUENCY SPECTRUM

A major enhancement of the clinical utility of qEEG derives from the well-confirmed discovery and recognition that the electrical activity of the brain of a person of any age, alert but at rest in a quiet, dimly lit environment, is dynamically regulated by an anatomically extensive, genetically based neurophysiologic homeostatic system. This power spectrum is extremely stable and highly replicable, distinctive for each brain region in a healthy, normally functioning individual, and independent of cultural, ethnic, or racial factors. Therefore, this predictable, known power spectrum can be used as a normative baseline that specifies the expected electrical activity of the healthy, normally functioning human brain across the human life span. The interactions between different brain regions, such as synchronization or coherence (nonrandomly phase-locked oscillations) and power symmetry between corresponding locations within and between the two hemispheres, are similarly regulated.

If the procedures described below are followed, then deviations from these norms can be described as standard scores, providing a reliable and objective metric for abnormalities of brain electrical activity. Further, since all features are expressed in the same units, they can be combined and expressed as multivariate descriptors of brain function. Like the ECG, deviations from the normal pattern of electrical activity are indicative of pathophysiologic processes and may have diagnostic implications and utility. Particular profiles of deviant standard scores have been found to be reliably correlated with certain categories of brain dysfunction or disorders or with positive as well as negative responses to specific treatments or evolution of the illness.

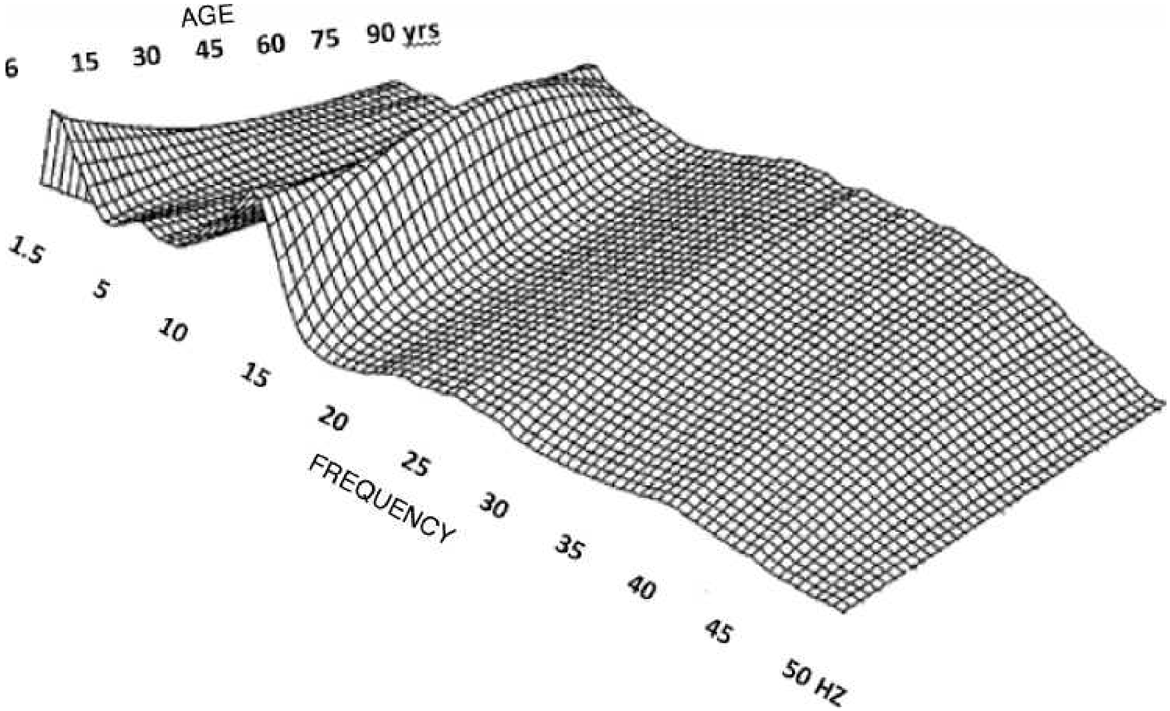

FIGURE 7.7–2. Schematic of the normative (*N* = *456*) EEG "relative power" ground state of the brain, for the 10/20 electrode at the midline parietal region (Pz) location, represented as a three-dimensional surface: *X-axis*, frequencies in 0.39 Hz intervals from 1.5 to 50 Hz; *Y-axis*, age shown in 1-year intervals from 6 to 90 years; and *Z axis*, magnitude of percent (%) power of the EEG at each frequency, 0% to 100%, represented by the height of the surface. Each "slice" of the image from left to right side of the image is the normal frequency composition of the EEG for that age. As you go from front to back of the image you are viewing the composition of frequency stacked across all ages shown. Note also that while shown here in one-year intervals, the actual normative regression equations are for continuous age, so that each individual is compared with norms for their exact age.

For each of the classes of qEEG features, the distribution of electrical measures within different portions of this range has conventionally been evaluated in so-called "wide-frequency bands," namely, the delta (1.5 to 3.5 Hz), theta (3.5 to 7.5 Hz), alpha-1 (7.5 to 10 Hz), alpha-2 (10 to 12.5 Hz), beta (12.5 to 25 Hz), and gamma (30 to 50 Hz) frequency bands. It is noted that the alpha, beta and gamma, and other bands can also be further divided. These seemingly arbitrary definitions of frequency bands, initially made by electroencephalographers on the basis of repeated visual inspection of apparently independent variations of rhythmic oscillations, correspond to the findings of independent sources of variance in the digitized EEG subsequently obtained by computerized principal component analyses. In current qEEG practice with advances in signal processing methodology and instrumentation, there is a steady expansion of the number of electrodes to as many as 256 positions on the head and of the frequency range to examine much faster oscillations as high as 1 kHz (1,000 counts per second) and a shift toward the use of the so-called very narrow band (VNB) (0.39- or 1.0-Hz divisions) spectral analysis, which has enabled the use of source localization and expanded exploration of the underlying pathophysiology.

Normative values of these various qEEG descriptors across a wide age range (6 to 90 years) have been established and repeatedly independently replicated in multiple countries, indicating that they are independent of race or ethnicity. High test–retest reliability has also been published. This regulation reflects activity within and interactions among many brain regions, reflected in local field potentials (LFPs) as well as the spontaneous EEG, mediated by many different chemical neurotransmitters whose synthesis and metabolism are similarly dynamically regulated and define the ground state of the brain. In Figure 7.7–2, the evolution of the EEG ground state for relative power (percent of total power at each frequency) at the midline parietal electrode (Pz) is depicted as a three-dimensional developmental surface; an analogous surface has been computed for the ground state of each electrode in the 10/20 system. While not shown here, normative values have been extended to the first year of life.

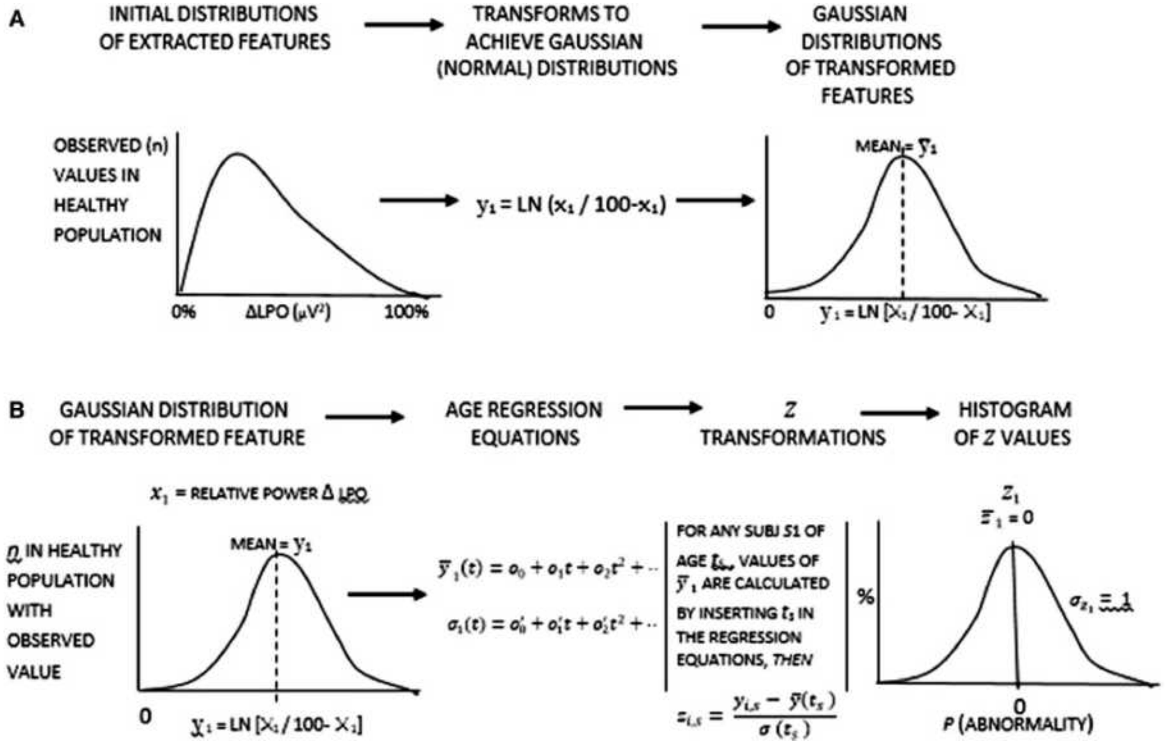
A
INITIAL DISTRIBUTIONS OF EXTRACTED FEATURES
TRANSFORMS TO ACHIEVE GAUSSIAN (NORMAL) DISTRIBUTIONS
GAUSSIAN DISTRIBUTIONS OF TRANSFORMED FEATURES
OBSERVED (n) VALUES IN HEALTHY POPULATION
0%
ΔLPO (μV²)
100%
y₁ = LN (x₁ / 100-x₁)
MEAN = y₁
0
y₁ = LN [X₁ / 100- X₁]
B
GAUSSIAN DISTRIBUTION OF TRANSFORMED FEATURE
AGE REGRESSION EQUATIONS
Z TRANSFORMATIONS
HISTOGRAM OF Z VALUES
x₁ = RELATIVE POWER Δ LPO
n IN HEALTHY POPULATION WITH OBSERVED VALUE
MEAN = y₁
0
y₁ = LN [X₁ / 100- X₁]
ȳ₁(t) = o₀ + o₁t + o₂t² + ··
σ₁(t) = o′₀ + o′₁t + o′₂t² + ··
FOR ANY SUBJ S1 OF AGE t_s, VALUES OF ȳ₁ ARE CALCULATED BY INSERTING t_s IN THE REGRESSION EQUATIONS, THEN
z_{i,s} = (y_{i,s} − ȳ(t_s)) / σ(t_s)
%
Z₁
z̄₁ = 0
σ_{z₁} = 1
0
P (ABNORMALITY)

FIGURE 7.7–3. Steps in “neurometric” processing of univariate EEG features to multivariate QEEG: (Step A, top panel) transformations to achieve Gaussianity (normalization) of the distributions; (Step B, bottom panel) age regression and *Z* transformation to obtain univariate standard scores. Step C, not shown in this figure is the combination of selected univariate *Z*-scores into multivariate estimators (Mahalanobis distances).

The concept of the ground state of the brain being an anatomically extensive, genetically based neurophysiologic homeostatic system is consistent with current hypotheses of the default mode network (DMN). The default network or DMN can be defined as the “brain system,” which is active when individuals are not focused on or receiving input from the external environment. The scientific literature describes the anatomy of the DMN as involving a distributed set of regions, including, in particular, the medial frontal regions, the inferior parietal lobule, the hippocampus, the medial temporal lobe, and the posterior cingulate cortex. The interaction between the regions involved in the DMN has been described as hubs, each including interactions between subsystems that are highly correlated with each other and weakly correlated with the others. Further, just as deviations from the age-expected normal values of the ground state have been demonstrated to be characteristic of different psychiatric disorders, the DMN has been postulated to be abnormal in many psychiatric states, including autism, schizophrenia, ADHD, PTS, and AD, all of which can be thought of as conditions associated with cognitive dysfunction in domains linked to the DMN, supporting the link between the ground state and DMN.

STEPS IN NEUROMETRIC qEEG ANALYSIS

qEEG Features

“Neurometric” qEEG analysis refers to the computer extraction from the visually or computer-edited, artifact-free EEG, recorded from 19 scalp electrodes located by the 10/20 system, of a large number of quantitative “univariate” features that assess different attributes of the electrical activity of the brain, and compared to age expected normal values. The steps presented in Figure 7.7–3 describe in detail the two main steps in neurometric analysis. The first step, seen in Panel A of the figure, shows how variables are transformed to obtain more Gaussian distributions using $\log_{10}$ transforms. The ability of these transforms to improve the Gaussianity of the distributions has been confirmed. Some investigators have more recently suggested different transforms, which may further improve the Gaussianity of the features, but such changes do not result in a significant change in feature values.

The second step is the computation of the *Z* score for the difference between age-predicted normative value (using age-regression) and the value obtained in the patient, which permits the estimation of the probability that such a value might be obtained by chance from a healthy peer’s exact age of the patient. As can be seen in Panel B of the figure, the mean and standard deviation of the normal population for the age of the patient is used to derive the *Z* score of the patient. The use of *Z* scores for all features validates the use of parametric statistical methods for further analysis. This method is standardly used by clinical laboratories to report findings from blood tests.

Conversion of all variables to the common metric of probability enables multivariate combination of variables, with initially different physical dimensions (e.g., uV for absolute power, percentage for relative power, and correlation for coherence), into a multivariate composite feature or brain state vector, to obtain a quantitative estimate of the overall severity of dysfunction within a region, a set of regions, or between regions within a measure. One such multivariate method used in Neurometrics is the Mahalanobis distance. Computation of the Mahalanobis distance removes the contribution of intercorrelations among the set, providing a corrected estimate of the overall abnormality of the combined set. Such multivariate descriptors can be thought of as characteristics of the system.

A particularly informative use of the Mahalanobis distances is in developmental disorders, in which deviations can be characterized as “maturational lags” or “developmental deviations.” Maturational lags are those multivariate patterns of brain function that would be normal if the child was younger. That is, such activity does occur in normal brain function but is not age-appropriate. On the other hand, developmental deviations are multivariate patterns that would not be considered normal regardless of the age of the individual. Such characterization of dysfunction can have a significant impact on the prognosis and treatment of the individual. Unpublished findings using this characterization of abnormalities suggested that those who responded to a pharmacologic intervention were more likely to have been described as having developmental deviations, whereas those with abnormalities best described as maturational lags were likely to become more normal over time without such intervention.

Reporting Results

In analyses of individuals, extracted features include an extensive set of features, each expressed as deviations relative to the age-appropriate normative values. These results can be reported in numerical tables, with highlighting of statistically significant results. For each of the 19 monopolar derivations, with the steps described above, the following measure sets are derived from the frequency spectra: the absolute power (uV2), relative power (percent), and mean frequency (the point above and below which half of the power falls in a band) are computed for the delta (1.5 to 3.5 Hz), theta (3.5 to 7.5 Hz), alpha (7.5 to 12.5 Hz), beta (12.5 to 25 Hz), and gamma (30 to 50 Hz) frequency bands and the total spectral power. Changes in spectral power can reflect changes, including those related to neurochemistry, oxygen flow, glucose metabolism, brain edema, and other conditions. In addition, connectivity between and within hemispheres is derived, including measures of coherence (synchronization), phase, and asymmetry (power ratios) in all frequency bands computed for all electrode pairs within and between hemispheres, reflecting disruption in neuronal transmission between brain regions. These same measures are computed for bipolar derivations of each electrode versus every other electrode within each hemisphere (connectivity). Measures of signal “complexity” including entropy, fractal dimension and others are also computed reflecting disruption of neural networks.

Interpolated topographic *Z*-score images can depict the statistical significance of local deviations relative to age-expected normal values for all univariates in each measure set as well as for a selected set of multivariate measures. Such maps are color-coded using a palette in which different hues represent the probability of the findings at each brain region. Because each *Z* value has a known statistical probability associated with it (e.g., $Z = \pm 1.96$ has $P < .05$), the results of such analyses are objective as well as self-interpreting. Figure 7.7–4 shows a sample of group average qEEG Z images (nose up, left hemisphere on the left) for different populations of psychiatric patients in each row. Each column represents a selected measure set. Significant local deviations from age-expected values can be seen as colored regions on the maps. In the color-coding for significance in this figure, the center of the scale, black, indicates no deviations from expected normative values. Moving up the scale through red to orange and yellow goes toward significant excess of the measure and down through blue to light green indicates a significant deficit of the measure. Looking at the top row on the left of this figure, all maps appear dark to black, showing no deviations from age-expected values in this large group of normal adults. On the other hand, in all other rows, significant excesses and deficits can be seen in all groups, and differences in the patterns of these deviations are also evident between the groups.

It is interesting to note that the progressive magnitude of abnormalities can be seen when comparing the images from the mild cognitive impairment (MCI) and Alzheimer disease (AD) groups (especially in relative power in theta, relative power in beta, and mean frequency of the total spectra, columns two, three, and five of the right panel). Also noted are the similarities between the schizophrenic patients off medication for 90 days and those on medication at the time of the test (bottom two rows in the left panel), suggesting the presence of the medication. The presence of distinctive patterns of qEEG abnormalities and differences between groups will be further discussed in more detail below.

Measures of connectivity have been increasingly demonstrated to be of importance in the description of brain dysfunction in mild traumatic brain injury (mTBI), concussion, postconcussive syndrome, post-traumatic stress disorder (PTSD), autism, and AD. Figure 7.7–5 shows an example of the images for coherence connectivity measures in the theta frequency band in a 64-year-old with early signs of cognitive decline. Each electrode location shows the results for the coherence between the labeled region and each other regions in the labeled band (Delta, Theta, Alpha, Beta, Beta2). The significance of the hypo- or hypercoherence is color-coded.

In addition to the conventional measures described above, the same methods have been extended to describe the complexity of the EEG signal measures of complexity, including chaotic/fractal measures, as well as theory-based measures of entropy and measures of mutual information. The complexity of the EEG signal reflects the disorganization of neural networks. These measures have been shown to be particularly important in the study of anesthetic affects (loss and return of consciousness), TBI, and PTSD and expand the measure of disturbances in the EEG signal to include nonlinear measures beyond the frequency domain.

HOMEOSTATIC SYSTEM GENERATING AND REGULATING THE EEG

The neuroanatomy and neurochemistry of the homeostatic system generating and regulating the EEG are schematized in Figure 7.7–6. The alpha rhythm arises from pacemaker neurons in the thalamus, which have an intrinsic oscillatory rhythm in the EEG alpha frequency range (7.5 to 12.5 Hz) and synchronize similar activity from neurons dispersed throughout the neocortex. When “instructed” by the cortex, the nucleus reticularis (encapsulating much of the thalamus like the skin of a pear) exerts differential GABAergic influences that hyperpolarize these pacemakers, slowing their frequency down to as low as 5 or 6 Hz (cycles per second) to enter the theta range.

When arousing afferent input reaches the midbrain reticular formation, and specific sensory thalamocortical relay (TCR) nuclei of the thalamus, cholinergic outflow from the ascending reticular activating system (ARAS) is sent to the nonspecific nuclei of the thalamic diffuse projection system and to the nucleus reticularis, antagonizing the inhibitory modulation of γ-aminobutyric acid (GABA), depolarizing pacemaker neurons, and allowing them to accelerate their slowed oscillations even into the 12 to 15 Hz range, usually referred to as beta. These interactions can serve to focus attention by modulating the excitability of TCR gates.

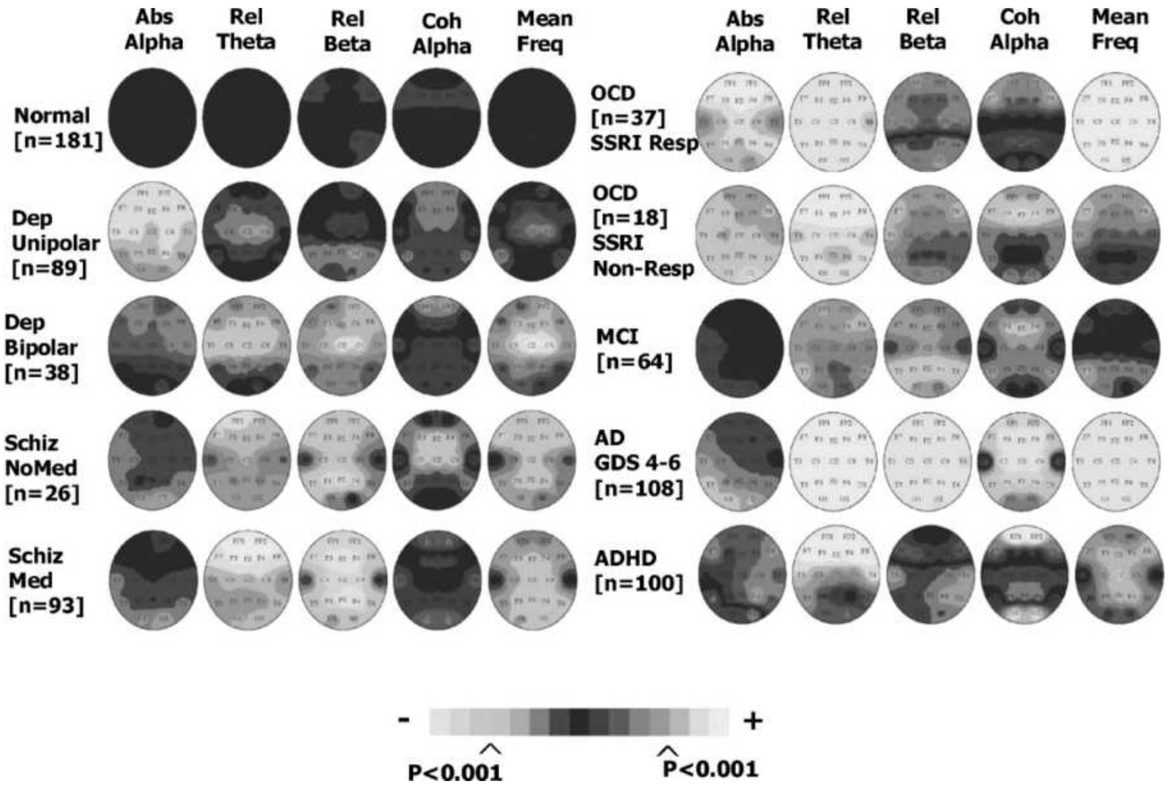
Abs Alpha
Rel Theta
Rel Beta
Coh Alpha
Mean Freq
Normal [n=181]
Dep Unipolar [n=89]
Dep Bipolar [n=38]
Schiz NoMed [n=26]
Schiz Med [n=93]
Abs Alpha
Rel Theta
Rel Beta
Coh Alpha
Mean Freq
OCD [n=37] SSRI Resp
OCD [n=18] SSRI Non-Resp
MCI [n=64]
AD GDS 4-6 [n=108]
ADHD [n=100]
-
+
P<0.001
P<0.001

FIGURE 7.7–4. Group average topographic images of selected neurometric variables (columns) across different clinical populations (rows). Columns, left to right: (1) absolute power (microvolts squared) in the alpha frequency band; (2) relative power (%) in the theta band; (3) relative power (%) in the beta band; (4) interhemispheric coherence (% synchronization) in the alpha band; and (5) mean frequency of total spectra of the EEG (Hz). (*Left*) Row 1, Normal individuals not used to construct the norms (*n* = 181); Row 2, Unipolar patients in depressed state (*n* = 89); Row 3, Bipolar patients in depressed state (*n* = 38); Row 4, Chronic schizophrenic patients, nonmedicated >3 months (*n* = 26); Row 5, Chronic schizophrenic patients, currently on medication (*n* = 93). (*Right*) Row 1, Drug-free baseline data in obsessive-compulsive disorder (OCD) patients who go on to be selective serotonin reuptake inhibitor (SSRI) responders (*n* = 37); Row 2, Drug-free baseline data in OCD patients who go on to be SSRI Nonresponders (*n* = 18); Row 3, mild cognitively impaired (MCI) older adult patients with a GDS staging of 3 (*n* = 64); Row 4, Alzheimer disease (AD), with GDS staging of 4 to 6, (*n* = 108); Row 5, Children with attention-deficit/hyperactivity disorder (ADHD) (*n* = 100). The color scale is proportional to the significance of the deviation from age-expected normal values, with black, no departure; blue to light green, deficits; and red to yellow, excesses. Statistical significance of entry for any group can be estimated by multiplying the value of the indicated hue on the *Z* score color bar by the square root of the group *N*. (See eBook for color image.)

Oscillations in the beta range (12.5 to 25 Hz) and in the gamma range (30 to 50 Hz) arise in the EEG to reflect thalamocortical and corticocortical transactions concerned with information processing and consciousness. Delta rhythms (0.5 to 3.5 Hz) arise from neurons in the cortex and often reflect functional deafferentation or neuropathologic influences such as ischemia or increased intracranial pressure but can be depressed by certain pharmaceutical agents or substance abuse. Hypothetical functional roles and processes attributed to regions in this diagram will be discussed further in the text.

Additional EEG rhythms arise from the activation of the dorsal striatum (sensorimotor integration) and ventral striatum and amygdala (emotional state and affect), interacting with major components of the limbic system for storage and retrieval of memories. Activation of this system results in outflow via the nucleus medialis dorsalis of the thalamus and the anterior and posterior cingulate cortices, of rhythmic theta waves (3.5 to 7.5 Hz) carrying phase-locked gamma oscillations (30 to 50 Hz) reflecting retrieval of memories via the hippocampus. This is discussed in further detail later in this section.

Multivariate features

Rescaling all neurometric features as *Z* scores, providing a common dimension proportional to probability, allows any selected set of qEEG measures to be compressed into a multivariate estimate of the total overall abnormality across the set. With one such method, the Mahalanobis distance (see above), the contribution of intercorrelations among the set is removed, providing a corrected estimate of the overall abnormality of the combined set.

With Mahalanobis distances, results of a qEEG evaluation can be represented quantitatively in a brain signal space. The length of the multivariate vector corresponds to the severity of the deviation of the patient's brain electrical activity from the origin defined by age-appropriate normative values, and its orientation in the physiologic space represents the quality or nature of the pathophysiology. By this representation of patients with different diagnoses and normal individuals as points in the signal space, two questions relevant to diagnosis can be answered quantitatively and will be addressed later in this section: (1) Can stepwise classifier functions (such as discriminant analysis) be derived that identify electrophysiologic biomarkers that separate patients with different diagnoses from each other or from normal individuals? (2) Can cluster analytical methods reveal subtypes/phenotypes with distinctive qEEG profiles but with the same clinical diagnosis, unmasking pathophysiologic heterogeneity in that diagnostic category?

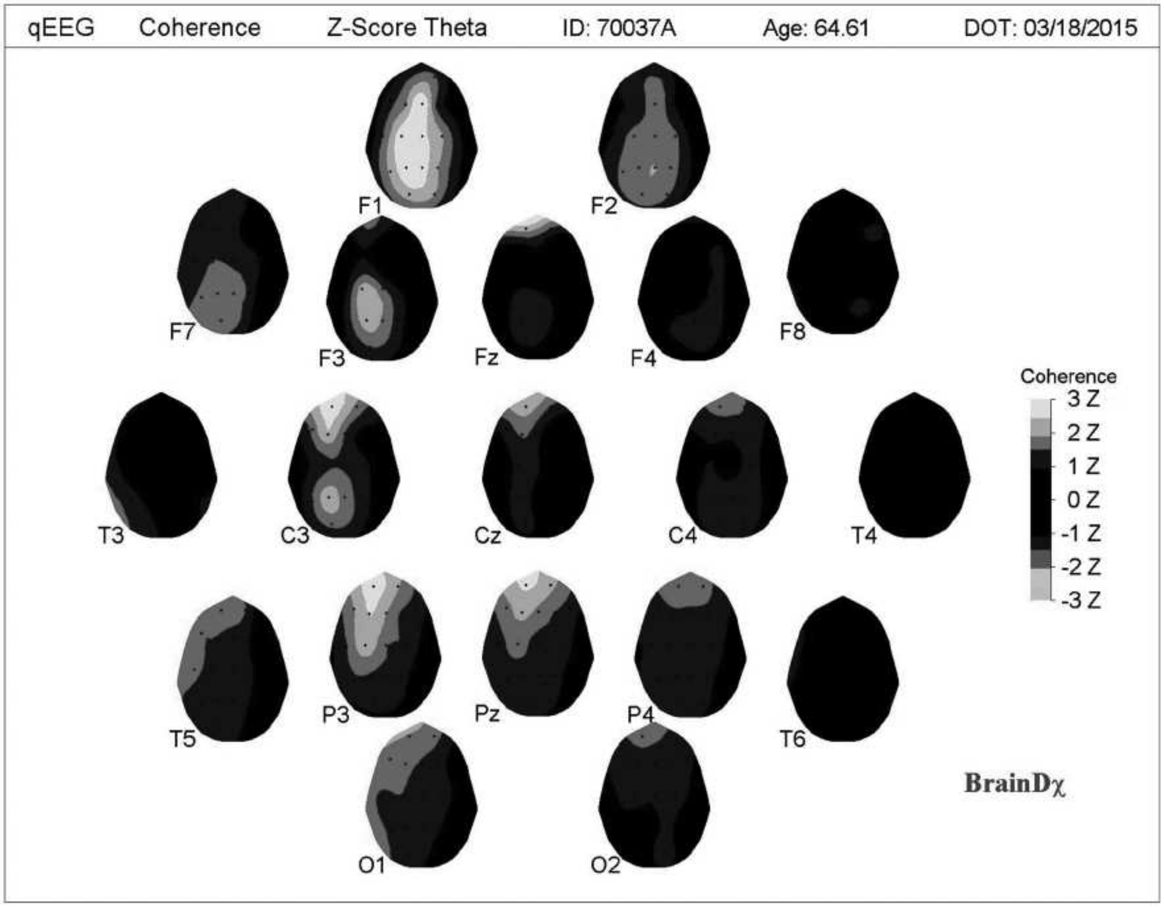
qEEG
Coherence
Z-Score Theta
ID: 70037A
Age: 64.61
DOT: 03/18/2015
F1
F2
F7
F3
Fz
F4
F8
T3
C3
Cz
C4
T4
T5
P3
Pz
P4
T6
O1
O2
Coherence
3 Z
2 Z
1 Z
0 Z
-1 Z
-2 Z
-3 Z
BrainDχ

FIGURE 7.7–5. Each component image (individual head map with nose up and left on the left) shows the results for the coherence between the labeled region and all other regions in the labeled band (in this case the theta frequency band), color-coded by significance level using the Z-scale shown. Regions in black are within normal limits, all colored regions are significant at $P < .05$ with the extremes at $P < .001$. The top left image shows the connectivity between the left frontopolar regions (FP1) and all other regions in the theta band, a band that has been shown in the literature to be correlated with degree of cognitive decline in older adults. It can be seen that there is a significant hypercoherence between these regions and both short and long-distance relationships on the left hemisphere and midline regions. This hypercoherence is also reflected in the coherence between FP1 and the left central (C3), left parietal (P3), and midline parietal (Pz) regions. This hypercoherence between frontal and more distant regions in a patient with early cognitive decline might serve as a baseline for a trajectory of increasing abnormality when tracked longitudinally in this patient. (See eBook for color image.)

qEEG Source Localization Tomographic Brain Images

A powerful expansion of the contribution of qEEG to the understanding of the underlying pathophysiology of neuropsychiatric disorders and the clinical utility of qEEG has been achieved over the past decade by the development and application of source localization analytics. qEEG low-resolution source localization methods perform the computation of an "inverse solution" to localize the mathematically most probable intracranial sources of the scalp-recorded EEG signal. With such methods, voltages or *Z*-scores of qEEG measurements from the scalp can be imaged, providing three-dimensional brain images that have been demonstrated to be coregistered with other neuroimaging methods. The standardized electrode positions of the proportional 10/20 system are placed in geometrical registration with a proportional Probabilistic MRI Brain Atlas, constructed at Montreal Neurological Institute based upon the average magnetic resonance imaging (MRI) anatomy of 300 normal brains. Referring to the computed sources to an average MRI, atlas averts the need to obtain an MRI of the individual patient. The positions of sources at any EEG frequency are depicted upon images of transaxial, sagittal, or coronal slices from the MRI Atlas, with each voxel color-coded in hues that quantify the deviation in each voxel of the computed source from the expected age-appropriate normative value, as in neurometric qEEG. There are two source localization methods that implement voxel norming, Variable Resolution Electromagnetic Tomographic Analysis (VARETA) and standardized Low-Resolution Brain Electrical Tomography (sLORETA); examples using these methods of source localization will be given below (see Figs. 7.7–7 to 7.7–10). The anatomical accuracy of these brain images computed from the qEEG has been repeatedly confirmed by coregistration studies using MRI, functional MRI (fMRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT).

Neurometric qEEG source localization provides a unique, innovative technology for functional brain mapping, since no other functional imaging method, neither fMRI, PET, nor SPECT, provides access to age-appropriate normative voxel data and only electromagnetic data (EEG or magnetoencephalography [MEG]) has the time resolution in milliseconds, allowing more precise identification of neuronal processes. This method provides the further advantage of the ability to obtain accurate functional brain images with economical and portable equipment which can be used at the point of care without the need for movement restraint within the equipment, camera, or imposition of a burden of radiation. It should be noted that advances in this method allow real-time imaging of sources, which has many interesting possibilities for unique ways to study the potential underlying pathophysiology of psychiatric disorders.

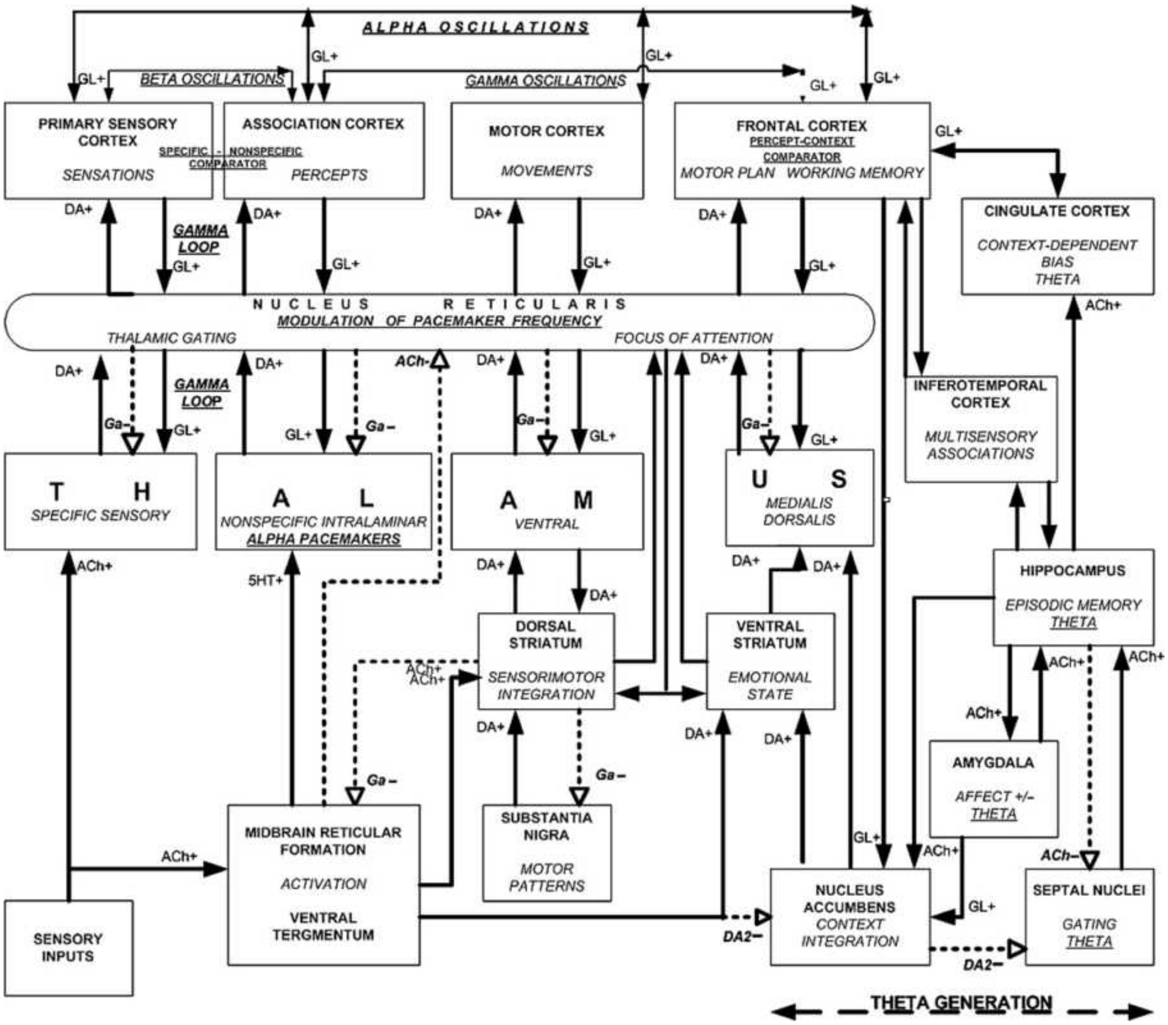
ALPHA OSCILLATIONS
BETA OSCILLATIONS
GAMMA OSCILLATIONS
GL+
PRIMARY SENSORY CORTEX
SENSATIONS
SPECIFIC - NONSPECIFIC COMPARATOR
ASSOCIATION CORTEX
PERCEPTS
MOTOR CORTEX
MOVEMENTS
FRONTAL CORTEX
PERCEPT-CONTEXT COMPARATOR
MOTOR PLAN WORKING MEMORY
CINGULATE CORTEX
CONTEXT-DEPENDENT BIAS
THETA
DA+
GAMMA LOOP
NUCLEUS RETICULARIS
MODULATION OF PACEMAKER FREQUENCY
THALAMIC GATING
FOCUS OF ATTENTION
ACh+
ACh-
Ga-
INFEROTEMPORAL CORTEX
MULTISENSORY ASSOCIATIONS
T H
SPECIFIC SENSORY
A L
NONSPECIFIC INTRALAMINAR
ALPHA PACEMAKERS
A M
VENTRAL
U S
MEDIALIS DORSALIS
HIPPOCAMPUS
EPISODIC MEMORY
THETA
5HT+
DORSAL STRIATUM
SENSORIMOTOR INTEGRATION
VENTRAL STRIATUM
EMOTIONAL STATE
AMYGDALA
AFFECT +/-
THETA
MIDBRAIN RETICULAR FORMATION
ACTIVATION
VENTRAL TERGMENTUM
SUBSTANTIA NIGRA
MOTOR PATTERNS
NUCLEUS ACCUMBENS
CONTEXT INTEGRATION
SEPTAL NUCLEI
GATING
THETA
DA2-
SENSORY INPUTS
THETA GENERATION

FIGURE 7.7–6. Circuits serving a major role as generators of indicated electroencephalogram (EEG) frequency bands, interrelationships, and mediating neurotransmitters proposed to comprise the homeostatic system regulating the EEG ground state of the brain. Salient hypothesized functional roles have been assigned to regions and putative transmitters assigned to relationships indicated by arrows: ACh, acetylcholine; DA/DA2, dopamine; Ga, γ-aminobutyric acid; Gl, glutamine; 5HT, serotonin; "+," excitatory actions; "-," inhibitory actions. A more complete explanation of this figure can be found in the text. In brief, pacemaker neurons in the thalamus generate an intrinsic oscillatory rhythm in the EEG alpha frequency range (7.5 to 12.5 Hz) propagated to the cortex that can be slowed into the theta range (6 to 7 Hz) by GABAergic influences from the nucleus reticularis, tending to close thalamic gates to the cortex. Incoming sensory information is transmitted by an "exogenous system" via thalamocortical relay (TCR) neurons to the cortex and also enters the midbrain, from which cholinergic arousal influences arise to antagonize γ-aminobutyric acid (GABA) and accelerate the EEG rhythms, opening the thalamic gates, and serving to focus attention. Oscillations in the beta range (12.5 to 25 Hz) and in the gamma range (30 to 50 Hz) reflect

> thalamo–cortical and cortico-cortical information processing. Delta rhythms (0.5 to 3.5 Hz) arise from neurons in the cortex and often increase with neuropathology but can be depressed by substance abuse. Afferent input also activates the dorsal striatum (sensorimotor integration) and ventral striatum, and amygdala (emotional state and affect), interacting with major components of the limbic system for storage and retrieval of memories. Readout from this "endogenous system" results in outflow via the nucleus medialis dorsalis of the thalamus and the anterior and posterior cingulate cortices, of rhythmic theta waves (3.5 to 7.5 Hz) carrying phase-locked gamma oscillations (30 to 50 Hz) reflecting retrieval of memories via the hippocampus. Thus, theta rhythms in the EEG can arise from two quite different causes.

Selection of Anatomical Regions of Interest

Since the statistical significance of deviations from expected normative brain electrical activity can be quantitatively encoded for every voxel within a source image, this input can be used to quantify the significance of deviations for specified anatomical regions of interest (ROIs), allowing the evaluation of the maximum abnormalities in the brain regions of an individual patient. This can be done in a variety of ways. One method maps sets of voxels that are within specific anatomical regions (using an atlas such as Talairach), quantifying and displaying the mean, median, or some percent of significant voxels in the region. Another way to select ROIs can be done by search algorithms that can be conceptually understood as methods to systematically remove from the three-dimensional representation all voxels that have an abnormality less than some selected threshold, for example, the $P < .05$ level ($Z = \pm 1.96$). The remaining voxels will lie within some three-dimensional region within the brain, which an algorithm automatically identifies from the coregistered corresponding anatomical neurosurgical stereotaxic atlas. In this way, ROIs are defined by the significance of deviations rather than anatomical localization. A numerical estimate of the degree of severity of the abnormality can be calculated, for example, by integrating the Z-scores of all voxels within the volume of the ROI, thus defined or counting the number of voxels above the threshold. These methods for quantifying abnormalities could be applied to the evaluation of treatment efficacy, whereby the magnitude of the abnormality before and after treatment is compared. Other methods for statistical analysis of source images include statistical probability mapping (SPM) or nonparametric statistical mapping (SnPM) as used in other brain imaging methods, especially fMRI.

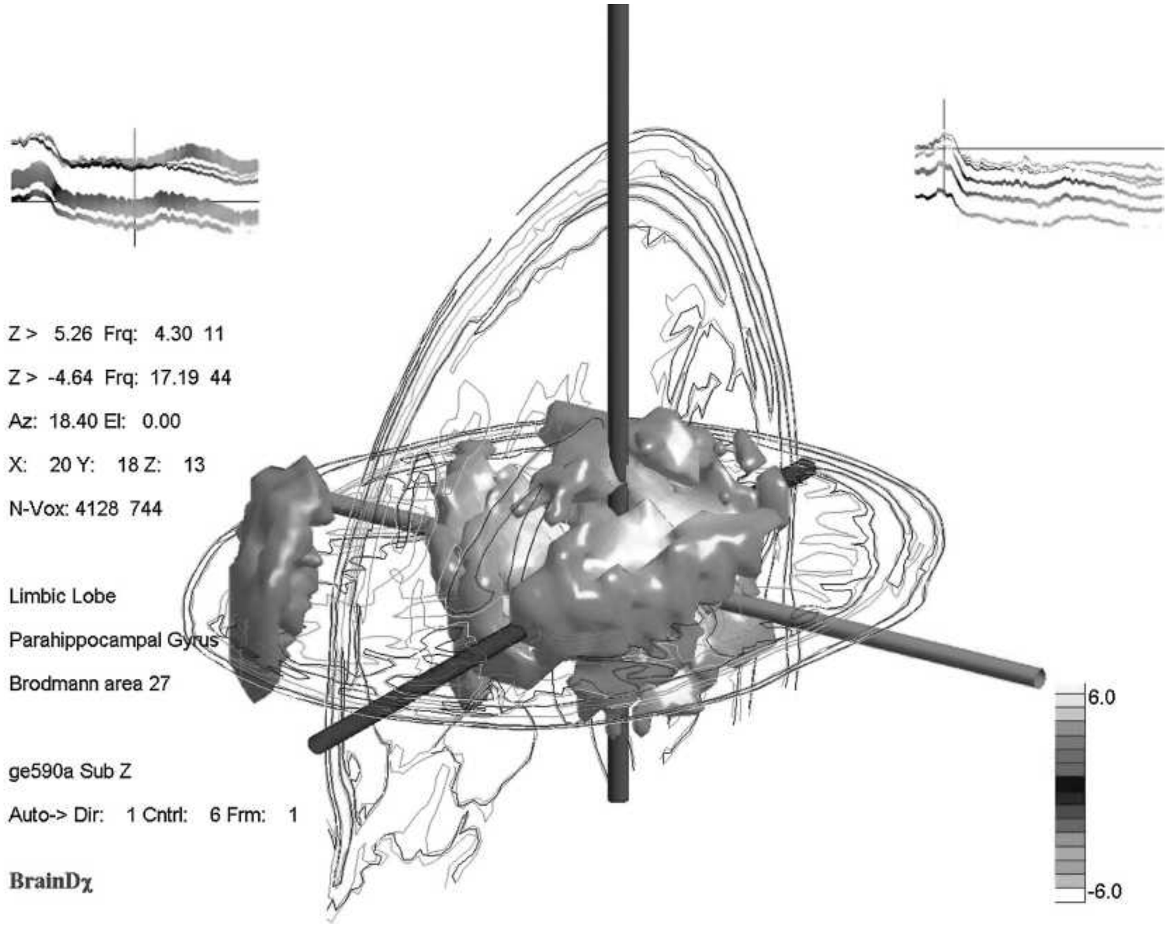
Z> 5.26 Frq: 4.30 11
Z> -4.64 Frq: 17.19 44
Az: 18.40 El: 0.00
X: 20 Y: 18 Z: 13
N-Vox: 4128 744
Limbic Lobe
Parahippocampal Gyrus
Brodmann area 27
ge590a Sub Z
Auto-> Dir: 1 Cntrl: 6 Frm: 1
BrainDχ
6.0
-6.0

FIGURE 7.7–7. Three-dimensional low-resolution electromagnetic tomographic analysis (LORETA) image of a "dissolved brain" of a patient with AD, after removal of all voxels with a *Z* score less than 3.29 ($P < .001$). The maximum theta excess (*dark red shading*, 4.3 Hz) relative to age-expected normal values is in the hippocampal and parahippocampal regions, while the frontal cortex (Brodmann area 27) reveals a minimum beta deficit (*blue*, 17 Hz). (See eBook for color image.)

Figure 7.7–7 presents an image of the brain of a patient with AD after all voxels with a *Z*-score less than 3.29 ($P < .001$) have been removed from the brain volume. It can be seen in this figure that the area of maximum abnormality relative to age-expected normal voxel values is in the hippocampal and parahippocampal regions.

Brain ROIs identified by neurometric source imaging, scaled for degrees of abnormality quantified by deviation from norms or alteration of function relative to some reference state, can be used as an adjunct to diagnosis, the selection of treatments, evaluation of bioavailability, and pharmacodynamic assessment of efficacy or even in drug development research. ROIs thus identified might then be interrogated by a noninvasive assay using Magnetic Resonance spectroscopy (MRS). MRS already has the capability to quantify the concentration of a steadily increasing number of brain metabolites, neurotransmitters, and critical ionic electrolytes. Such multidisciplinary assessments might eventually become routine in evidence-based personalized psychiatric practice.

CLINICAL APPLICATIONS OF qEEG DEPARTURES FROM GROUND STATE OR DEFAULT MODE NETWORK

Psychopathology: qEEG Biomarkers of Diagnoses

Deviations from normative values imply dysregulation (local and network) that may have functional and treatment implications. In numerous peer-reviewed articles, there is evidence of the high proportion of psychiatric patients in any diagnostic category who consistently display significant deviations from the expected normative qEEG values. This is clearly shown in Figure 7.7–4 for large populations of many different diagnostic groups. Many of the articles present evidence that these profiles are sufficiently distinctive not only to indicate the most probable DSM diagnosis for an individual patient but, in some instances, also can be used to predict differential responses to treatments or different evolution of illness; reviews of the abundant relevant literature are available. It should also be noted that the degree of dysfunction, such as the progressive stages of dementia, can be accurately reflected in the qEEG, demonstrating that increasing qEEG abnormality is correlated well with an increasing degree of cognitive impairment. Thus, qEEG can represent a biomarker, be useful as an adjunct to diagnosis and staging, as well as to aid in the optimal selection and objective evaluation of treatment and treatment response. An example of such, the relationship between the degree of cognitive decline and magnitude of qEEG abnormalities, can be seen in Figure 7.7–4, comparing MCI (third row on the right panel) to AD (fourth row on the right panel), especially as seen in increased excess of relative power in the theta band (second column), an exaggerated deficit of relative beta activity (third column), and significant slowing of the mean frequency of the spectra (fourth column).

If the abnormal qEEG features detected in each member of a group of patients with different DSM-diagnosed disorders are represented in multivariate dimensions, then they can be envisaged as swarms of points (features) dispersed in a multidimensional brain electrical signal space. To the extent that particular profiles of qEEG abnormalities (sets of features or biomarkers that covary with the disease state) are in fact distinctive for patients with one specific symptomatic, DSM-based disorder, patients with the same diagnosis would be expected to lie close together within such a swarm in a region of the signal space, while patients with different DSM diagnoses would constitute swarms in different regions. This concept is shown schematically for patients with disease X and patients with disease Y in Figure 7.7–11.

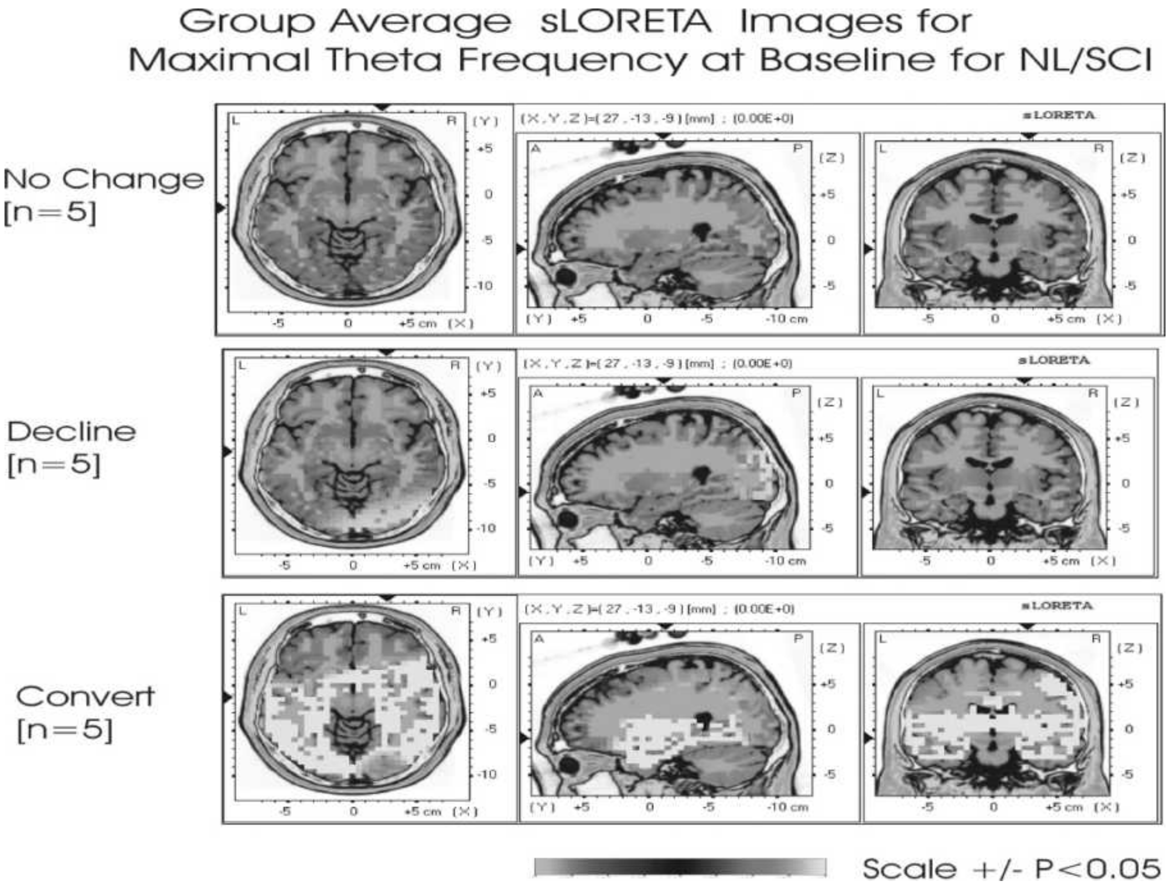
Group Average sLORETA Images for
Maximal Theta Frequency at Baseline for NL/SCI
No Change
[n=5]
Decline
[n=5]
Convert
[n=5]
(X,Y,Z)=(27,-13,-9) [mm] ; (0.00E+0)
sLORETA
L
R
A
P
(Y)
(Z)
(X)
+5
0
-5
-10
-10 cm
+5 cm
Scale +/- P<0.05

FIGURE 7.7–8. Group average low-resolution electromagnetic tomographic analysis (LORETA) images (*N* = 5 in each group) of the regions displaying a maximal excess of theta activity at baseline for three different groups (*N* = 5 in each group) of older adult patients with only subjective complaints of impairment (SCI), all diagnosed as GDS 2, with no demonstrable neurocognitive deficits, who upon long-term follow-up (5 to 7 years) were found to show: Top row, no change; middle row, decline to mild cognitive impairment (MCI); bottom row, conversion to Alzheimer's disease (AD), with GDS 4 to 6 and substantial impairment. The images in each row, from left to right, depict transaxial, sagittal, and coronal slices automatically selected to intersect at the coordinates of the voxel that had the most deviant standard score for excessive theta activity (4.3 Hz). Each voxel is color-coded for the statistical significance of excess theta in standard (z) scores, with hues ranging from red to yellow as significance of abnormality (overactivation) increases. Regions of increasing abnormalities include the hippocampus, parahippocampus, and parietal–temporal cortex, as also indicated in other brain imaging methods. (See eBook for color image.)

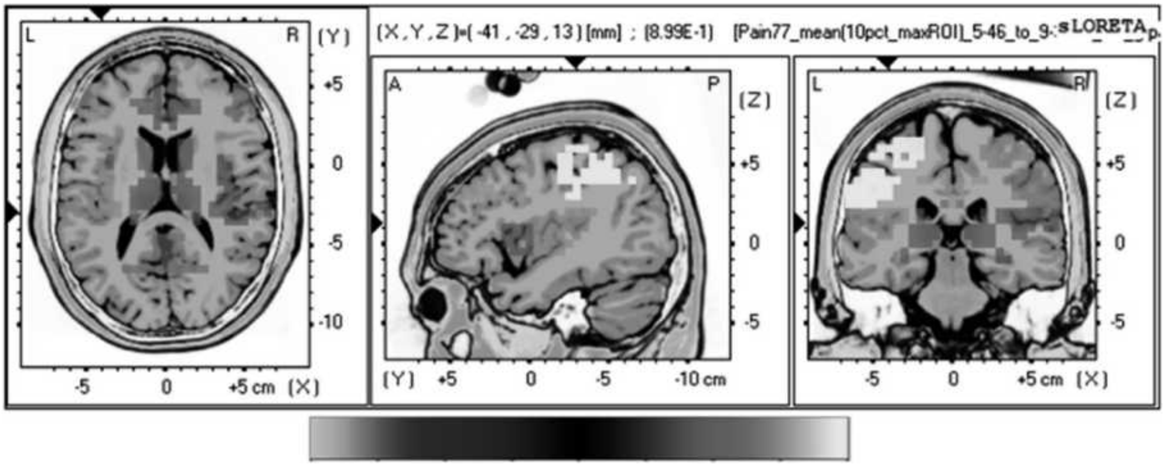
(X,Y,Z)=(-41,-29,13) [mm] ; (8.99E-1) [Pain77_mean(10pct_maxROI)_5-46_to_9-
sLORETA
L
R
(Y)
+5
0
-5
-10
-5
0
+5 cm
(X)
A
P
(Z)
+5
0
-5
(Y)
+5
0
-5
-10 cm
L
R
(Z)
+5
0
-5
-5
0
+5 cm
(X)

FIGURE 7.7–9. Source localization (sLORETA) images for group average voxel z-scores for ROIs in the band 5.46-9.36 Hz for the pain population ($N = 77$), shown for a single axial, sagittal, and coronal MRI level (−41, −29,13). The images are color-coded for significance of difference according to the color scale (below figure), where extremes of the color scale are $P < .01$. Overactivation is shown in red/yellow and deactivation in blue/aqua (none seen in this image). (See eBook for color image.)

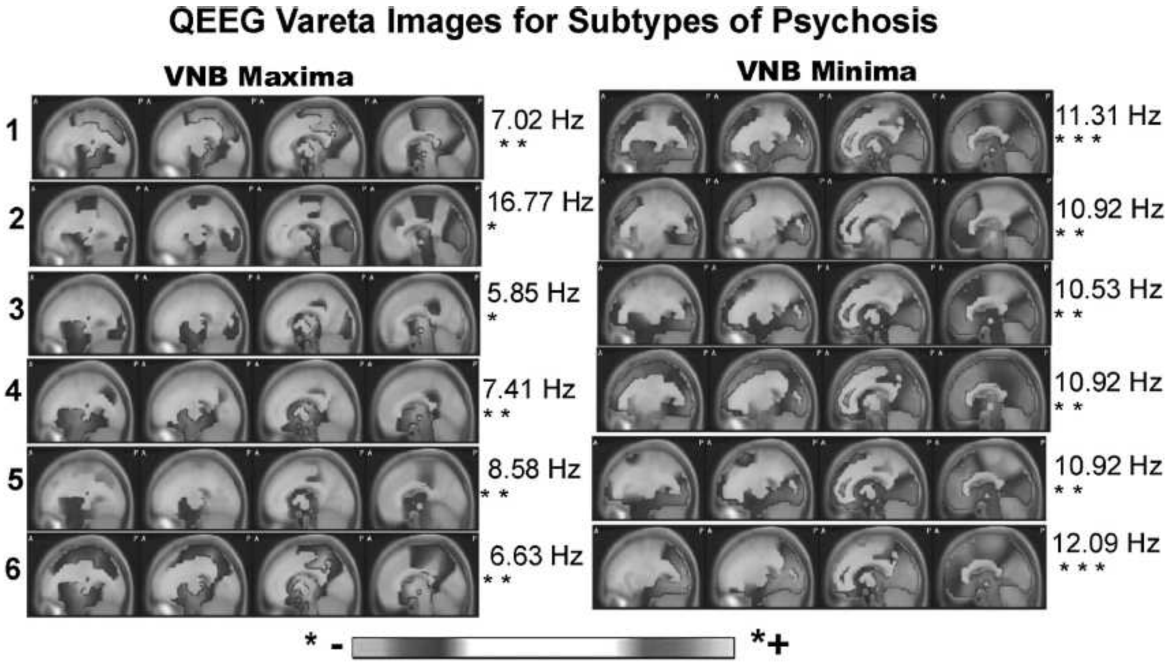

FIGURE 7.7–10. From top to bottom, each row depicts group average quantitative electroencephalography (qEEG) variable resolution electromagnetic tomographic analysis (VARETA) images of selected sagittal slices from the six subtypes (1 to 6) of psychotic patients identified by cluster analysis of a population of 377 psychiatric patients, all of whom were psychotic. Each cluster contained patients who had received a variety of fourth revised edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR)/tenth edition of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) diagnoses, including alcoholism, depression, and chronic medicated, unmedicated, and never-medicated first-episode schizophrenia (see text). (*Left*) Sagittal slices depicting the anatomical regions that had the maximum excessive activity are shown, and (*right*) the corresponding slices with the most deficient activity, with the corresponding most deviant very narrow band (VNB) spectral frequencies of each subtype indicated at the right side of each set of slices, are shown. Note that the most deviant excessive frequencies (Hz) of the six clusters (*left*) cover a wide range (frequency at left of each row), and were found in a diverse set of brain regions, suggesting a variety of putative upregulated neurotransmitters, while the most deficient frequencies were within a very narrow range (within the mid to high alpha band), and were found in a very consistent set of brain regions, suggesting a very small set or possibly even a single downregulated neurotransmitter. The deviation from normative values of every voxel in each slice is color-coded proportional to standard scores, with the appropriate hues corresponding to the color bar that is shown below. Note that the statistical significance of the hues of each voxel can be estimated by multiplying the *z*-score of the corresponding value from the color bar (and the palette range is indicated for each row by: *, ±1.0; **, ±2.0; ***, ±2.5) by the square root of the cluster size *N*, where: Clus 1, *N* = 53;

Clus 2, *N* = 93; Clus 3, *N* = 65; Clus 4, *N* = 24; Clus 5, *N* = 110; Clus 6, *N* = 32. Thus, all voxels colored other than gray in these slices are extremely deviant from normative values, with a minimum *P* < .001. (See eBook for color image.)

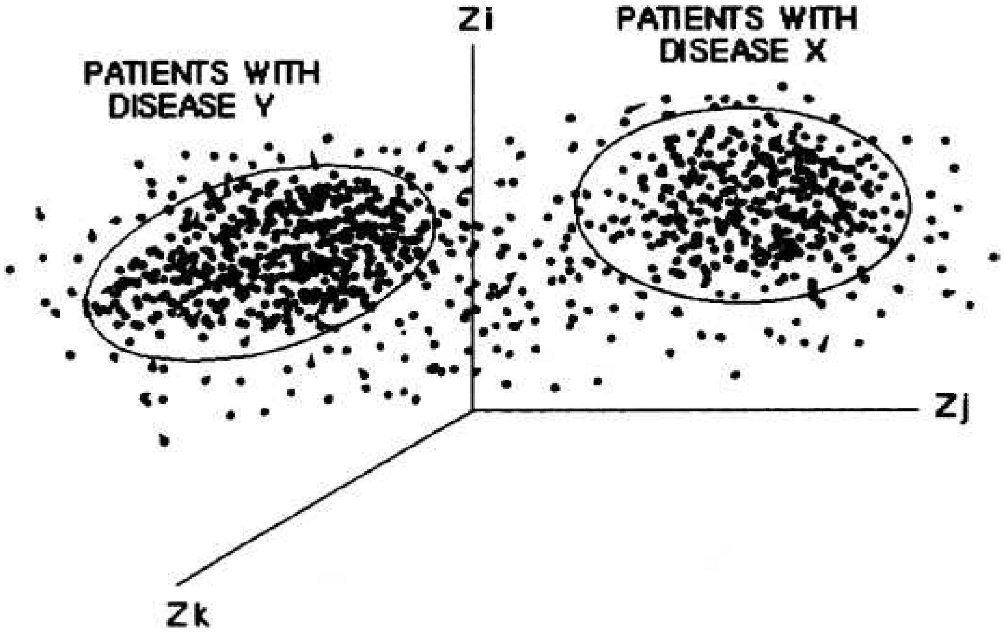

FIGURE 7.7–11. Point swarms representing the hypothetical scatter of brain state features in two different regions of a brain signal space, suggesting that discriminant classification analyses might separate patients diagnosed with Disease X from others diagnosed with Disease Y.

Statistical methods (classifier development) used to achieve separation between groups involve searching the multimodal feature space using classification methodologies and ML to identify distinctive profiles of abnormality. Several such methods exist, including genetic algorithms (GA), logistic regression, support vector machines (SVM), neural networks, and other tools, used to obtain classifier candidates that can optimally separate the two (or more) categories (disease states). For example, evolutionary algorithms (such as the GA) perform a stochastic search involving randomness from one iteration to the next and evaluates a series of candidate solutions, where each new candidate is informed by high-performing predecessors, similar to genetic evolution. The final GA solution identifies a selected set of features for which a weighted combination (when compared to a threshold derived from the receiver operator curve [ROC]) will allow probability estimates of the likelihood of belonging to one of the groups. Such classifier functions can accurately identify individuals with specific psychiatric disorders and distinguish those with one DSM diagnosis from those with a different one or from normal. Since the initial successful application of this method, numerous qEEG discriminant algorithms have been reported with sensitivity, specificity, area under the receiver operating curve (AUC), and positive and negative predictive values equal or superior to the widespread use of such well-established diagnostic imaging tools as, for example, CT for detection of stroke or TBI or MRI/DTI (diffusion tensor imaging) for concussion.

The use of qEEG features to identify an electrophysiologic biomarker of concussion has demonstrated the accuracy of such an approach. The main contributors to a concussion biomarker have been reported to be qEEG connectivity features, reflecting what is understood about the pathophysiology of concussion. Of particular importance is the demonstration of a significant correlation between this biomarker and DTI findings in the concussion population, providing evidence that this qEEG-based biomarker reflects white matter tract dysfunction.

Other studies have supported the relationship between qEEG biomarkers of neuropsychiatric disorders and advanced neuroimaging in different diagnostic groups. Such data suggests the use of qEEG as a surrogate for neuroimaging, especially at the point of care, for tracking of treatment response, and for aiding in a better understanding of the pathophysiology underlying a disorder.

Subtyping: qEEG Heterogeneity within DSM Diagnoses using qEEG Biomarkers

As is apparent from the highly variable efficacy of a given psychotropic agent in patients with the same DSM diagnosis, patients grouped together as displaying symptomatically homogeneous psychopathology may suffer from a variety of pathophysiologic dysfunctions responsive to different treatments or display different evolution of the illness. This variable responsivity suggests that individuals within the same swarm of points in a region of electrophysiologic signal space, representing first-order similarities among a group of patients with the same DSM diagnosis, might actually not be homogeneous. Within the region containing such a swarm or cloud, there might be sectors containing individuals with second-order differences in features that also contribute to their electrophysiologic profile or pattern of disorder. These differences may reflect the existence of subtypes of pathophysiology related to biomarkers that are critical for response to a given treatment, in addition to the first-order influences that produce sufficiently similar global symptoms to warrant the same DSM diagnosis, as shown schematically in Figure 7.7–12. Within the two-point swarms, the figure depicts individual members of the distinct subtypes A and B within the cloud diagnosed with disease Y and of subtypes C and D within the other cloud diagnosed with disease X. It is important to note that although referred to as “features” a feature can be a multivariate composite of more than one feature (e.g., the ratio of two features or a measure of connectivity between regions such as cordance).

In most instances, diagnosis is largely determined by clinical evaluations and DSM symptom-based categorization except for cases that present dual-diagnostic problems. Potentially, the most important and useful clinical applications of qEEG lie not just as an adjunct to diagnosis but in identifying subtypes whose members, while relatively homogeneous in terms of displaying similar symptoms and receiving like diagnoses, are statistically probable to have response-relevant differences (heterogeneous underlying qEEG profile) with respect to response to particular treatments, prognosis of evolution of illness, or prediction of treatment outcome. Investigations using methods of qEEG-based classification or cluster analyses have been useful in unmasking such heterogeneity in a number of disorders, such as (1) negative as well as positive responsiveness to or selection of stimulant therapy in ADHD; (2) response to selective serotonin reuptake inhibitors (SSRIs) in OCD (see Fig. 7.7–4, right panel, rows 1 vs. 2); (3) probability of relapse in crack-cocaine dependence; (4) pathophysiologic profiles of psychotic states or schizophrenia; (5) prediction of future cognitive decline in normal older adult; (6) response to antidepressants among patients presenting in a state of depression (see Fig. 7.7–4 left panel, row 2 vs. 3); (7) the response to nonstimulant or stimulant medication in ADHD patients; and (8) prediction of rapid or prolonged recovery in concussion. Reviews of this literature are available.

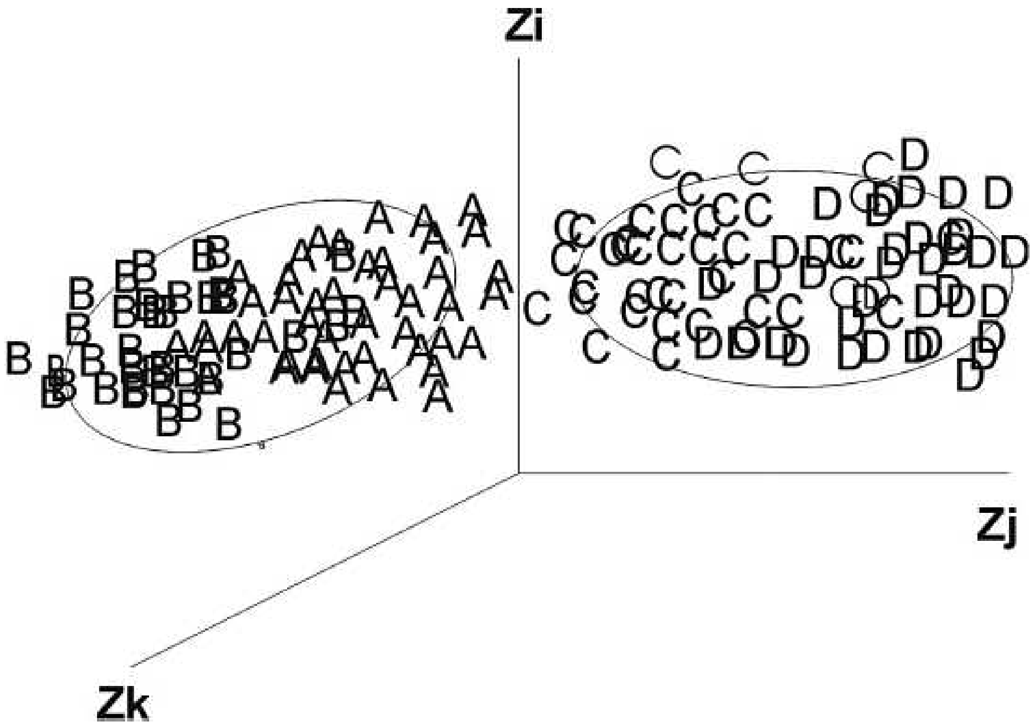

FIGURE 7.7–12. Point swarms representing the hypothetical identification of subtypes of patients in four different regions of the signal space, with different pathophysiologic profiles A and B within those diagnosed with Disease Y and with different profiles C and D within those diagnosed with Disease X.

qEEG Applications in Attention-Deficit/Hyperactivity Disorder (ADHD), Alzheimer Disease (AD), Chronic Pain, and Psychosis. The scientific literature applying qEEG methods to the study of psychiatric and neurologic patients, their diagnosis, underlying pathophysiology, and treatment response is vast and includes numerous studies in ADHD, depression, dementias, drug dependence, learning disabilities, OCD, panic disorder, schizophrenia, traumatic head injury, PTSD, chronic pain, and many other disorders. This subsection provides an overview of the literature in three areas: ADHD, dementia, and chronic pain.

Attention-Deficit/Hyperactivity Disorder

Extensive studies demonstrate the clinical utility of qEEG as an adjunct to the diagnosis, evaluation, and treatment of children, adolescents, and young adults with learning and attention problems. Using selected qEEG features (e.g., frontal theta activity or theta/beta ratio, alpha reactivity, alpha peak frequency [APF], and cordance), high accuracy, sensitivity, and specificity have been achieved in distinguishing normal controls from

ADHD/attention-deficit disorder (ADD) children, children with ADHD/ADD from those with learning disabilities, ADHD/ADD patients with a positive response to specific medication from those with a negative response. Augmenting the qEEG features with behavioral measures further enhanced the sensitivity for predicting long-term treatment response. A review of this expanding literature is available.

We will briefly summarize meta-analyses that research a large number of studies, including diagnostic capabilities, subtype identification, and treatment response in the ADHD population. One such meta-analysis of qEEG studies in ADHD (DSM-IV-TR diagnosis) reported on a total of almost greater than 1,250 patients and concluded that although many studies reported ADHD to be characterized by increased theta/beta ratios (TBR, theta power increase with a beta power decrease), this meta-analysis did not conclude that TBR is a reliable diagnostic trait of ADHD when compared to normal controls. However, results did support TBR as a prognostic value in predicting treatment outcome to stimulant medication and/or neurofeedback (NFB) in a substantial subgroup of ADHD. Additionally, more recent studies report that low frontal APF was predictive of nonresponse to methylphenidate. An additional meta-analysis reviewing NFB in ADHD reported that compared to nonactive control treatments, NF was found to have a more sustained treatment effect for at least 6 months post-treatment. Under DSM-5, ADHD has been expanded to more accurately characterize the disorder in adults as well as children and adolescents. Symptoms are still divided into categories of inattention, hyperactivity, and impulsivity. Several studies using qEEG have explored the ability to identify different patterns of abnormalities reflecting these different symptom profiles and have reported suggestive results. The existence of such electrophysiologic subtypes may be related to the prediction of different treatment responses.

The existence of electrophysiologic subtypes within the ADHD population has been further demonstrated in studies applying cluster analysis of qEEG features within a large population of ADHD patients. Such studies have shown the existence of subtypes within the population with different underlying sources for the different subtypes. While preliminary studies are suggestive of a relationship between subtype membership and treatment response, much work in larger populations remains to be done in this area.

A nonpharmacologic treatment of ADHD patients that is considerably growing is qEEG, and qEEG source, NFB. A meta-analysis including five qEEG NFB studies and a total of approximately 300 patients supports the improvement of some aspects of the ADHD symptom profile following NFB. In addition, preliminary studies have indicated the possibility of selecting NFB protocols based on individual qEEG biomarkers to optimize efficacy. More research is ongoing to support NFB as a first-line, standalone treatment modality and for the identification of the subtypes of ADHD that might most benefit from this treatment modality.

An extensive Bayesian network meta-analysis of the comparative efficacy of pharmacologic and nonpharmacologic treatments (including psychological, alternative medicine, and other treatments) of ADHD children and adolescents reported weak evidence for monotherapies being superior to placebo and that combination therapies (e.g., behavioral therapy in combination with stimulant therapy) were superior to placebo. Overall, these weak results further support that heterogeneity must be taken into account in order to adequately evaluate treatment efficacy.

Another example of the use of qEEG in the ADHD population for a better understanding of the underlying pathophysiology includes studies of alpha reactivity. This measure reflects the magnitude of the differences between alpha activity in the eyes-closed resting condition and that with eyes open with increased external stimulation. These studies suggest that ADHD patients show less alpha reactivity, supporting abnormalities in arousal mechanisms. Application of cluster analysis to the ADHD population revealed the existence of neurophysiologic subtypes and showed a relationship between cluster membership and response to treatment. qEEG source localization (VARETA) applied to individual subtypes demonstrated different underlying pathophysiology. A cluster characterized by excess alpha activity (maximum abnormality at 11 Hz, in the high alpha range) displayed primarily cortical abnormalities, maximal in right parietal cortical regions, whereas VARETA images of a cluster with theta excess (maximum abnormality at 5.4 Hz, in the theta range) revealed primarily temporal cortical and hippocampal abnormalities.

The utility of incorporating qEEG as an adjunct to the evaluation and treatment of ADHD was highlighted in a 2004 report by the American Academy of Pediatrics, which stated that “new brain wave analysis techniques like quantitative EEG will help experts more clearly document the neurologic and behavioral nature of ADHD, paving the way for better understanding and treatment.” Now, more than a decade later, a growing literature supports this thesis, and the advanced methodology implementing qEEG, ML, and subtyping addresses this need directly.

Cognitive Deterioration and Alzheimer Disease

A vast literature spanning more than three decades reports qEEG changes in the pattern of brain electrical activity associated with the severity of cognitive deterioration in older adults. Space does not permit a review of these numerous qEEG studies that consistently demonstrate the importance of the following qEEG features related to the severity and progression of dementia: Increased delta and/or theta power, decreased mean frequency of the total spectra, anteriorization of the alpha peak, decreased frequency of the dominant occipital rhythm, and changes in coherence, synchronization, and connectivity patterns. As additional advanced signal-processing methods are applied to the EEG signal, additional features have been proposed as qEEG biomarkers of disease severity in AD. These include features based on canonical correlation coefficients reflecting synchrony between multiple regions and theta current source density. A meta-analysis of portions of this literature is available.

Sensitivity to the earliest preclinical stage (i.e., the first symptomatic manifestation pre-MCI) of AD, currently described as subjective cognitive impairment (SCI), has also been demonstrated. Multiple publications have used qEEG features to achieve accurate differentiation of early AD from older adult normal controls, discrimination of dementia from depression in older adults, the distinction between early versus late-onset AD, discrimination between AD and vascular dementia, and identification of mild dementia and MCI. Further, studies demonstrate that when qEEG and ML techniques are combined with other neuroimaging methods (e.g., fMRI, MRI/DTI, SPECT), diagnostic accuracy can be increased.

The use of qEEG features observed in initial evaluations for the prediction of cognitive decline of normal older adult patients and those with MCI or early-stage mild dementia to more severe dementia has been demonstrated in multiple studies. In one such longitudinal study, the exquisite sensitivity of qEEG at the earliest stage of SCI, normal older adult subjects with only subjective complaints about memory, but no deficits on neurocognitive testing, were followed longitudinally for 7 years. On the basis of only the qEEG data from the initial qEEG evaluation, future cognitive decline was predicted with over 90% accuracy based upon subsequent clinical staging. Significant abnormalities in qEEG features were seen in the baseline evaluations of those who went on to decline and were absent in those who showed no change over time. The most salient abnormalities were seen in theta power, the mean frequency of theta and of the total spectrum, and intrahemispheric coherence across all bands. With source localization in those who go on to decline, the underlying pathophysiology identified at baseline included excessive theta activity in the hippocampus, amygdala, and temporoparietal cortex, all regions identified by other neuroimaging technology as important in dementia and related to subsequent decline to dementia. Figure 7.7–8 shows the sLORETA source localization images from the initial baseline qEEG data for the abnormal activation seen in the theta frequency band for three groups (N = 5 each group) of SCI subjects. The top row of this figure shows those who, upon long-term follow-up, were found to show no cognitive decline, the middle row shows those who are found to decline to MCI, and the bottom row shows those who convert to dementia. Such data demonstrate the high sensitivity of qEEG to the earliest signs of functional brain impairment prior to the presence of structural abnormalities. It is noteworthy that theta power has been shown to be negatively correlated with regional perfusion, especially in temporoparietal and central regions, and with PET glucose metabolism in temporoparietal and frontal regions, and is positively correlated with hippocampal atrophy seen on MRI.

Chronic Pain and Thalamocortical Dysrhythmia

Chronic pain affects over 35% of the U.S. adult population, increasing greatly in those over the age of 65. Furthermore, chronic pain is reportedly comorbid with depression in the vast majority of the patients, making chronic pain and its effective treatment a problem for both neurology and psychiatry. Contributing to the high prevalence of chronic pain and both the over- and undertreatment of the condition is the subjective standard of care for rating pain, which can lead to inappropriate and suboptimal treatment. The use of qEEG features for the objective quantification of pain has led to an increasing number of studies that have demonstrated the coregistration between functional neuroimaging studies and qEEG studies of the relationship between activation of the brain's "pain matrix" and the experience of pain intensity. Studies have emphasized that despite the high spatial resolution of neuroimaging tools, including fMRI and PET, these tools do not have the temporal resolution of qEEG, which gives them the advantage of better reflecting the dynamic nature of pain perception. Findings of qEEG source localization in pain patients support the potential to derive a quantitative measure of the severity of pain using information extracted from a multivariate descriptor of the abnormal overactivation of regions of the pain matrix. The frequency-specific (theta band) overactivation reported in the regions including the thalamus in chronic pain patients has been demonstrated in a number of studies and supports a model of thalamocortical dysrhythmia (TCD) as the mechanism of the neuropathic pain condition. This hypothesis is further supported by studies in chronic pain patients following successful central lateral nucleus thalamotomy (CLT), which showed a significant diminution in overactivation in the previously overactivated regions.

Using qEEG source localization, significant differences between those chronic pain patients and those without clinically significant depression have been shown. No significant correlation was found between the frequency-specific theta maximum abnormality in the ROIs of the Pain Matrix pain patients and the Beck depression scores in these patients. Thus, although both disorders share underlying sources in the limbic system, results support the existence of distinctive profiles of electrophysiologic abnormalities.

Figure 7.7–9 applies the methods described above to image the ROIs in chronic pain patients. A clear relationship can be seen between those regions with significant dysregulation (overactivation) and those identified using conventional, more costly, time-consuming, and not readily available neuroimaging tools. These regions include the anterior cingulate, anterior and posterior insula, parietal lobule, thalamus, S1, and dorsolateral prefrontal cortex (DLPFC). While consistent with those reported with conventional functional imaging, the use of qEEG source localization also include the mid and posterior cingulate not previously reported, suggesting that the increased temporal resolution of electrophysiologic measures allows more precise identification of the pain pathway.

Psychosis

EEG abnormalities in schizophrenic patients have been the subject of a large number of publications. Consistent descriptions of increased slow activity, especially in anterior and frontal locations, increased fast beta activity, and interhemispheric asymmetries (R > L) appear in this literature. The influence of variations of medications, chronicity of medication, and length of illness may explain some of the inconsistencies between studies. Another problem in the interpretation of these studies is the existence of electrophysiologic heterogeneity within the populations. Some studies have specifically addressed this issue, with illuminating findings that reveal subtypes with the same diagnoses but with different profiles of salient qEEG abnormalities, as well as different responses to various psychotropic medications.

The results of a qEEG subtyping study of a large population of psychotic patients are summarized here. With multivariate cluster analysis of qEEG features for a large number of psychotic patients (N = 390), including medicated, unmedicated, and never-medicated first-episode patients diagnosed with the DSM-IV or ICD-10 with schizophrenia, psychotic depression, alcohol psychosis, and other psychoses, from a multisite international study, six clusters were identified, replicating and extending prior findings in schizophrenic patients. The distribution of patients across the six subtypes found in this study was not uniform, with subtypes containing as few as 24 or as many as 110 members out of the total cohort of 390 cases. Patients bearing a diagnosis of psychotic depression or schizophrenia either on or off medication, as well as first-episode medication-naive patients, were represented in at least four of the six clusters.

All clusters showed power asymmetry in all or almost all bands and the power was greater over the right hemisphere, especially in regions of the frontal lobe. The uniform strong right hemispheric dominance of every frequency band in each cluster would suggest a general impoverishment of language. It has long been reported that left-hemispheric impairment is characteristic of schizophrenia. In particular, abnormalities of the left parahippocampal gyrus, the left medial temporal lobe, the loss of synaptic proteins from the left thalamus, etc., have been repeatedly reported in imaging studies of schizophrenic patients. Almost all clusters also displayed hypofrontality, especially in the dorsolateral and mesial prefrontal cortical regions. These findings are concordant with reports of decreased metabolism in the frontal cortex consistently noted in other functional neuroimaging studies. PET studies have shown decreased frontal but increased basal ganglia activity in schizophrenic patients, also evident in the qEEG findings, and which may reflect neocortical–striatal–

thalamocortical interactions perturbed by dopamine (DA) imbalances in psychotic patients.

This data suggest that the left-sided deficiency generalizes to psychosis, with the right hemisphere becoming dominant because of a lack of interhemispheric suppression, possibly related to the pervasive disturbance of coherence in one or another frequency band between various homologous regions, to the extent that such coherence may reflect the activation of a feedback loop between the two sides of the brain. The power dominance of the right hemisphere, together with the hypocoherence found in this study, may indicate a deficit of homeostatic feedback between homologous regions. The severe hypocoherence in both short-fiber and long-fiber tracts within each hemisphere found in every cluster in this study provides strong support for the presumption that information processing is disturbed in psychotic patients. One possible explanation of the coherence findings is that the short fiber tract hypocoherence reflects a deficit in the fast-spiking (FS) and low threshold spiking (LTS) GABAergic interneurons, modulated by DA and serotonin, which facilitate synchronization of local circuit neurons, while the long fiber tract hypocoherence reflects a deficit in the long-range interneurons which can synchronize neuronal ensembles across very long distances. There have been repeated suggestions that the personality and thought disorder observed in schizophrenics reflect the failure to integrate activity across distributed neural circuits.

Neuropsychological demonstrations of deficits in learning, memory, and executive tasks were interpreted to indicate widespread generalized fronto–limbic dysfunction. Of particular interest in the context of the authors' pervasive findings of disturbed coherence are recent neurobiologic studies of the binding process, which attribute a major role in perception and consciousness to the synchronization of EEG oscillations. Repeated reports have called attention to phase-locked oscillations in the gamma (30 to 100 Hz) frequency band. Phase-locking is a quantitative index of the temporally synchronized interactions in distributed neuronal systems. Decreased functional connectivity has also been reported in first-episode neuroleptic-naive schizophrenic patients.

VARETA source localization images of the brain regions displaying the computed intracerebral sources of the most excessive (hyperactive) and most deficient (hypoactive) activities are shown in Figure 7.7–10. These underlying sources of the qEEG clusters indicate pathophysiologic heterogeneity among patients diagnosed as psychotic, arising from two different causes of dysfunction in six basic subtypes, namely, overexcitability of one set of brain regions suggesting different underlying causes in each subtype, together with underactivation of a second set of brain regions perhaps due to the same cause in each subtype.

These clusters suggest that the optimal management of the psychotic patient will entail separately upregulating one neuroanatomical system while downregulating a different system, although it is conceivable that the distributed hypoactivity reflects inhibitory influences exerted by the hyperactive system. The regions in which homeostatic regulation is most severely disturbed (dysregulation) across all six subtypes, displaying both hyperactivity at one frequency as well as suppression of activity at a different frequency, are the amygdala, caudate, cingulate, DLPFC, hippocampus, and thalamus, suggesting that psychotic symptoms may primarily reflect a dysfunction arising from subcortical dysregulation.

Other studies have used eLORETA to evaluate the EEG Source Functional Connectivity (EEG-SFC) in schizophrenic patients. One such study compared stable medicated schizophrenia patients with healthy controls. Results demonstrated abnormalities in functional connectivity, especially in the theta frequency band where synchronization was increased (hypercoherent) in schizophrenic patients, which was increased in frontal regions with increased duration of the disease. Alpha band connectivity was seen to be decreased (hypocoherent) in frontal regions in schizophrenics and independent of the length of illness. The higher frequencies (beta1, beta2, and gamma) also showed disturbances in connectivity involving frontal regions and parietotemporal regions. The interactions found with chronicity need to be taken into consideration when interpreting such findings, but results lend support to the hypothesis of schizophrenia involving disturbances in connectivity.

qEEG-Based Depth of Anesthesia Monitor: Level of Consciousness

Instruments that use qEEG to provide dependable real-time monitors of the depth of anesthesia during surgery, measuring the level of consciousness, have demonstrated distinctive and quantitative changes from the ground state. Intraoperative studies using such instruments revealed that with loss of consciousness (LOC), there was a marked anteriorization of alpha activity into all regions of the frontal cortex, followed by a shift to a power spectrum dominated by activity in the regions of the power spectrum at the margins between delta and theta activity (2 to 4 Hz), which gradually disappeared as the patients emerged from the levels of surgical anesthesia through deep sedation to arousal. It is of particular interest that gamma coherence between prefrontal and posterior brain regions was invariantly sharply diminished precisely at LOC during anesthesia induced by any anesthetic agent studied and returned to preinduction levels just prior to the return of consciousness. While anterior–posterior coherence similarly dropped sharply at LOC for activity in the delta, theta, alpha, and beta bands, it was noteworthy that coherence for anterior–posterior neural transactions in these frequency ranges did not return to preinduction levels until as late as 10 to 15 minutes following the return of consciousness, as gauged by overt responsivity to verbal commands and speech. This suggests that gamma coherence may be of unique relevance to the maintenance of consciousness per se, which will be further elaborated in a later subsection of this section.

Application of source localization algorithms (VARETA, described above) to the EEG data collected during anesthesia revealed extensive inhibition of the frontal lobes, thalamus, basal ganglia, amygdala, and hippocampus coincident with the LOC. Return of consciousness was always preceded by reversal of the qEEG changes and disinhibition of these structures. Such findings have led to a new neurophysiologic theory of the mechanisms of action of general anesthesia.

The intraoperative studies of the invariant and reversible qEEG changes that accompanied the loss and return of consciousness are compatible with five neurophysiologic steps during the induction of surgical depth of anesthesia with any combination of agents, as follows: Step 1: Depression of the brainstem reduces the influences of the ARAS on the thalamus and cortex. Step 2: Depression of mesolimbic–DLPFC interactions leads to blockade of memory storage. Step 3: Further depression of the ARAS releases its inhibition of the nucleus reticularis of the thalamus, resulting in the closure of thalamic gates (especially in the diffuse projection system) by hyperpolarizing GABA-mediated inhibitory action of the nucleus reticularis (theta increase), thereby blocking (Step 4) thalamo–cortico–thalamo reverberations and perception (gamma decrease), so that (Step 5) parietal–frontal transactions are uncoupled (gamma coherence decreases), blocking cognition, and (Step 6) prefrontal cortex is depressed to reduce awareness (an increase of frontal delta and theta).

Since its appearance, a number of publications have confirmed various aspects of this theory, particularly with respect to the central role of the changes in coherence between anterior–posterior oscillatory activity in the gamma band. More specifically, studies of qEEG symbolic transfer entropy have shown that directional connectivity between anterior and posterior regions is inhibited by anesthesia; that there is a reorganization of hub networks from parietal to frontal regions under anesthesia; and that qEEG directional connectivity before and after LOC show decreases between the fronto–parietal and fronto–occipital regions. In summary, changes in qEEG features with loss and return of consciousness suggest an “anesthetic cascade” involving a blockade of thalamocortical reverberations and uncoupling of normal transmission between frontal and parietal/occipital regions.

THERAPEUTIC CORRECTION OF DEPARTURES FROM THE GROUND STATE

There are a number of clinical areas in which the validity of the proposition that therapeutic maneuvers that restore previously deviant qEEG features to the ground state will have a positive clinical outcome has been demonstrated. These studies suggest the possible utility of serial qEEG evaluations and the presence of electrophysiologic biomarkers as an objective adjunct to the selection of treatment, evaluation of efficacy, and optimization of dosage.

Pharmaco-EEG

Since pioneering work began in the 1960s, qEEG has been used as a tool in the development of new psychopharmaceutical agents. It was early observed, using only a few electrodes, that drugs in a particular class caused a distinctive alteration of the frequency composition of the spontaneous resting EEG. For example, anxiolytics typically cause an increase in power in the beta band, especially in EEGs recorded from electrodes over the prefrontal regions. These observations gradually became systematized so that data were collected from the full 19-channel array of the 10/20 system.

With the advent of qEEG methods, dose-response curves were calculated to correlate the amount of agent, time after administration, and quantitative changes in specific qEEG variables. It was noted that these changes followed typical time courses for different agents, indicating the potential utility of qEEG to study pharmacodynamics and bioavailability. These relationships are now sufficiently well understood for some agents that serial qEEG evaluations can be used to obtain an early indication of whether or not a specific drug and dose will be efficacious for an individual patient, and dose-response curves can be constructed using serial computations of qEEG features and the likelihood of the existence of biomarkers across successive doses.

Since under normal conditions, the qEEG is highly reproducible, in some instances, it is possible to use a simple test–retest paradigm to evaluate whether a specific drug is appropriate for an individual patient, with the two tests separated by a time interval sufficiently long to enable an effective concentration to be established in the brain. For example, in a child with ADHD/ADD who will be a good responder, within 45 minutes after a single 10-mg dose of methylphenidate, pharmacodynamic effects can be seen as a significant decrease of premedication qEEG abnormalities.

The importance of the potential contribution of qEEG to pharmacotherapy and increasing use led to the formation of a committee of thought leaders by the IPEG (International Pharmaco-EEG Society), which has published Guidelines for the Recording and Evaluation of Pharmaco-EEG in man.

Key-Lock Principle

As pharmaco-EEG knowledge accumulated, researchers discerned that certain patterns of qEEG effects were distinctive for certain classes of drugs. It gradually became apparent, at least for some conditions and some drugs, that the so-called “key-lock” principle appeared to provide a guide for the selection of medication. For example, the optimum medication for a patient whose qEEG abnormality was a deviation of the power spectrum from normative baseline with excess theta and deficit of alpha would be the pharmaceutical substance that, when administered to normal individuals, caused the appearance of the opposite power spectral deviation: Decrease of theta and increase of alpha. A body of evidence has been published that provides substantial support for this proposition.

A clinical example of this principle is demonstrated in the relationship between response to SSRIs and qEEG cluster membership shown in a study of OCD patients, all of whom were considered clinically to be candidates for SSRI treatment. Two clusters were found in this study, with differences in qEEG features that can be seen in Figure 7.7–4, where the pretreatment baselines of the two different OCD clusters are shown in the right panel rows 1 (responders) and 2 (nonresponders). One of these OCD clusters displayed an excess of theta activity (Fig. 7.7–4, right panel column 2, second row), while members of the other cluster displayed an excess of alpha power (Fig. 7.7–4, right panel column 1, first row). The effect of SSRIs on the normal EEG is to decrease alpha and increase theta activity. As can be seen in Figure 7.7–4, consistent with what would be predicted by the key-lock theory, the cluster whose baseline (pretreatment) qEEG was characterized by an excess of alpha and deficit of theta activity is the one that contained the SSRI responders (EEG normalized by the medication), and members of the cluster with an excess of theta were nonresponders.

Neurofeedback

NFB is essentially operant conditioning of brain electrical activity, first applied to the treatment of epilepsy in the early 1970s. NFB treatments postulate that: (1) continuous measurement and representation of some selected target correlate of an undesirable activity or physiologic process (autonomic measure, electromyogram [EMG]), ERP component, or composition of the qEEG power spectrum) with provision for visual or auditory feedback upon reduction of the representation or some alternative type of reinforcement to the subject whenever a decrease of the target is detected can be used to help a patient learn how to reregulate some abnormal physiologic process or aspect of brain electrical activity; and (2) that normalizing the target signal will result in a therapeutic benefit. During the last decade, a large number of practitioners have begun to use NFB to treat a wide variety of behavioral disorders and dysfunctions, most widely for ADHD but including such applications as remediation of neurologic deficits resulting from TBI or stroke. Among the several hundred papers reporting the utility of NFB are explorations of its use in addiction, ADHD, especially the suppression of undesired motor activity while remaining attentive, anxiety disorders, autism, autoimmune disorders, chronic fatigue syndrome, conduct disorder, epilepsy, learning disabilities, migraine, mood disorders, and, most recently, performance enhancement of skilled actions. In several of these areas, there have been replicated reports of statistically significant, successful reduction of the targeted symptoms. The basic method of NFB is to record a measure that reflects or is correlated with some physiologic process or condition considered undesirable, pathognomonic, or pathophysiologic, which is being continuously monitored. Some reinforcement is provided whenever the subject/patient succeeds in changing the undesirable process in a direction defined to be therapeutic, increasing or decreasing the measure. The manifestation of the undesirable condition is usually reflected in some stimulus parameter such as the intensity, pitch, or temporal pattern of an auditory stimulus or the size or brightness of a visual image, and the reinforcement is delivered by a change in the parameter, defined as a change of the undesirable process in the desired direction, such as reduction of the intensity of sound or size of a flame.

A variety of measures have been reduced successfully by NFB, including (1) ratios between qEEG measures in different bands (e.g., the “theta/alpha ratio” of the power spectrum recorded at a particular electrode position); (2) measures reflecting the autonomic nervous system, such as anxiety, galvanic skin response (GSR), high heart rate, high blood pressure, or pupillary dilatation; (3) conditions such as anxiety, attention, or obsessive behavior; (4) ERP measures considered putatively to reflect attentional processes, such as mismatch negativity (the occurrence of “different” in a sequence of paired stimuli S1 to S2 that are “same–same” or “same–different”), the amplitude of P300 in the “oddball” paradigm (in which an occasional “rare” event is interspersed in series of “common” events producing an expected late positive component to the rare event in normal cognition). A study in chronic pain demonstrated that the abnormal electrical activity at a particular frequency detected using real-time qEEG source localization within the anterior cingulate could be used for NFB to reduce intractable pain. Examples in the ADHD population are described above. A particularly novel application of NFB has been its use to enhance artistic performance, a domain in which successful intervention is a priori quite unexpected. A series of well-controlled studies sought evidence of operant control of the alpha/theta ratio, with the rationale that such control might induce a meditative state of deep relaxation. This parameter was then used to enhance performance, using criteria established by the Associated Boards of the Royal Schools of Music of England to assess variables of artistic performance, including several in the domains of instrumental competence, musicality, musical understanding, communicative ability, and overall quality. Performances were videotaped and randomized and then evaluated by experienced assessors external to the institution and blinded as to the order of performances and experimental group membership of the students. Highly significant results were obtained and successfully replicated.

An extensive review article, with a survey of evidence that validates NFB in diverse applications but particularly in ADHD and performance enhancement, has recently appeared. Although NFB was initially disparaged by many neuroscientists, neurologists, psychologists, and psychiatrists, the success and extent of this therapeutic activity are indicated by increasingly widespread utilization, and reports of carefully designed scientific studies reported in the peer-reviewed literature suggest the clinical utility of the method.

Important advances in EEG signal processing have enabled the use of real-time qEEG signal measurement, immediate data processing for feature extraction of parameters of interest, and the ability to provide feedback paradigms that use continuous measures of the EEG signal. The goal of such a method is to move the patient toward a more “normal” state, guided by their deviations from age-expected normal values. Such approaches have demonstrated the potential effectiveness of such methodology for future improved clinical utility and personalized applications of NFB.

CONSCIOUSNESS

Definition of Consciousness

Webster’s dictionary defines consciousness as “the quality or state of being aware, especially of something within one’s self.” Webster also defines awareness as “the condition of being conscious of an external object, internal state, or fact.” This tautology reflects the lack of scientific attention to this problem that persisted for much of the last century, while the question of how subjective experience arises from neural tissue was considered not a fit topic for neurobiologic research. With advances during the last decade in functional brain imaging and methods better adapted to the study of systems, neurophysiologic and psychological evidence has been forthcoming that supports a definition of consciousness as: "the subjective experience resulting from the continuous interaction of three influences: (1) incoming information about the external environment and the internal state, melded with (2) information retrieved from a dispersed memory system of stored information about previous working memory or episodic experiences that are the most currently relevant, and (3) state factors modulating both our perceptions of and our reactions to the outer world, by affecting arousal levels, emotional states, and moods."

"Relevant" means those experiences for which the environmental and physiologic state-generated inputs are most evocative because of inherent salience and situational correspondence or their multisensorial similarities. "Modulated" means the biasing of apperception and memory selection due to more global, direct, or indirect influences of humoral factors to alter the excitability of components of both the responsive and the reactive systems.

INFORMATION IS SYNCHRONOUS DEVIATIONS FROM GROUND STATE

If the billions of neurons whose activity is sensed by an EEG electrode were discharging completely at random, then there would be no net voltage detectable by the electrode. Voltage excursions in the resting EEG or ERPs that cause the qEEG power spectrum to deviate from the ground state represent temporal fluctuations in underlying regions of the synchronized activity of a significantly great number of neurons relative to the proportion that would be expected to discharge together solely by chance. Analogously, the images of sensory, perceptual, or cognitive processing detected by fMRI represent the spatially distributed decrease in oxyhemoglobin caused by metabolic requirements of massively synchronized ensembles of brain cells, relative to reference background activity. All such ensembles are intimately interconnected and in constant bidirectional transactions with numerous other ensembles of neurons at varying distances. The homeostatically regulated ground state of the resting EEG power spectrum of every brain region, as well as the remarkable stability of transactions reflected in the coherence and symmetry among these regions within and between the hemispheres, indicates that the spontaneous synchronization of nonrandom firing within and among regional populations of neurons is dynamically restricted and constrained within well-defined limits and regulated by positive and negative feedback processes. The constancy of these quantifiable variables of brain electrical activity over sustained time periods, in spite of the known fluctuations in excitability of individual neurons, indicates that this stability reflects a basic property of the large-scale organization of the central nervous system. Consistency with the concept of the DMN is noted.

Such predictable background brain activity contains no information because the information is that which cannot be predicted. Significantly nonrandom or synchronous brain electrical activity beyond these constrained limits constitutes a deviation from the homeostatically regulated ground state that defines the information content of that perceptual moment and causes the EEG voltage to deviate from the regulated range of random fluctuations. Spatially dispersed nonrandom synchrony is the key signature of significant informational activity within the brain. Information, retrieval of memories, and the generation of the perceptual content of moments of consciousness are encoded as temporal patterns of synchronized nonrandom activity in very large ensembles of neurons distributed spatially in the exogenous and endogenous systems causing dispersed neuronal ensembles to deviate from the ground state. One might speculate that brief excursions around the constrained baseline of the ground state reflect the sporadic thoughts that fleet through the mind of a person resting quietly, which rapidly converge to the ground state unless they are long sustained.

Discontinuity of Consciousness

Subjectively, consciousness appears to be continuous and unitary, that is, seamless. However, psychophysiological studies indicate that perceptual processes within the brain are discontinuous, parsing consciousness into successive “perceptual frames,” or moments of subjective experience. Under controlled experimental conditions, if the interval between two stimuli becomes as short as 80 to 100 milliseconds, then the pair merge into a single event. Although the EEG at every electrode position on the head appears to be an apparently continuous time series of voltages, the topography of the variance of the local deviations from the average voltage over the entire scalp, or global field power (GFP), is also discontinuous. A limited number of topographies actually wax and wane at rather regular intervals that have been called “microstates” or “atoms of thought," found to have an average duration of about 82 ± 2 milliseconds within a normally functioning healthy population of 500 subjects ranging in age from 6 to 90. Similar discontinuities have been found in magnetoencephalographic (MEG) research. Such evidence suggests that the apparent continuity of subjective time is an illusion produced by neurophysiologic processes parsing information into successive chunks about 80 to 100 milliseconds in duration.

On average, four spatial topographies account for 94% of the variance of the scalp EEG, mapping the field in four basic modes: Activation of the right anterior hemisphere, left anterior hemisphere, frontal cortex, and motor cortex. These may, respectively, represent the encoding of spatial information in the environment, verbal encoding of that information, executive interpretation and evaluation of the verbal–spatial representation of the environment in view of relevant memories, especially of the recent past, and the organization of adaptive language and motoric responses. The transitions among modes tend to follow a particular sequence in normal subjects but are found to be different in patients with some psychiatric disorders.

Parallel Multiplexed Information Processing Channels

Although evidence shows that perceptual frames last about 80 to 100 milliseconds, other experiments have established that temporal discriminations can be made at substantially shorter intervals, on the order of tens of milliseconds or less. Thus, the neural receptor systems mediating each sensory modality may be organized into multiple parallel channels, containing perceptual frames offset from one another by a fixed time, like commonly used multiplexing of information. Using averaging techniques time-locked to the presentation of a specific sensory stimulus is equivalent to demultiplexing, enabling the electrical activity reflecting the processing of information in only one of these parallel channels to be extracted as an ERP. It is important to stress that dispersed neuroanatomical structures containing the other parallel channels in the afferent pathways are not only processing that information but are simultaneously engaged in many other neural transactions, continuously active during the time required for the averaging process. As the time-locked activity of one channel converges to a stable ERP wave shape, considered as "*signal,*" the ongoing activity in all of the other channels converges to zero because it is in a random phase relative to the time of stimulus presentation and is considered as "*noise.*" The signal-to-noise ratio in the ERP waveshape extracted by the averaging process is improved proportionally to the square root of the number of samples that have been averaged. If subjective experiential continuity is mediated by consecutive inputs from these hypothesized multiplexed channels, then *consciousness must reflect a global statistical steady state maintained within a spatially extensive brain system that is independent of any individual neurons.*

Construction of Consciousness

The brain is organized to assign an affective and experiential label on every moment of input from the environment that causes a significant nonrandom discharge. When such a discharge is provoked by external or exteroceptive stimuli, it activates a network of neuronal ensembles linking the sensory receptors to subcortical structures in the brainstem and the thalamus, causing volleys of oscillatory activity to be propagated to dispersed regions of the primary sensory cortex and to cortical association areas, providing a fractionated representation of the *present.* These volleys propagate in parallel to structures such as the amygdala, entorhinal cortex, hippocampus, and septum comprising the limbic system, merging with and modulating the relevant activated readout from the *past,* as well as to subcortical structures such as the basal ganglia and hypothalamus reflecting *internal state,* that together mediate emotional valence and motivation as well as motoric organization. A spatially extensive set of nonrandom electrical cortico–limbic–subcorticocortical oscillations links these structures, *binding* them into a dynamic functional network. That which is linked becomes phase-locked and coherent, producing an instant of subjective experience that merges the present with selected aspects of the past: the "*remembered present.*" These coherent discharges define all of the information in the brain at any moment, which must subsume the content of consciousness. Therefore, the content of consciousness logically must arise from this spatially extensive nonrandomness.

The discrete parsing of information into sequential segments in each modality, integrated across modalities during the microstates, is experienced as a series of apparently continuous perceptual frames. The remainder of this chapter attempts to describe how activity in exogenous and endogenous systems is combined to construct the remembered present, first taking advantage of knowledge about the processing of information within only one of the hypothesized channels, a snapshot artificially visualized by the ERP averaging strategy, as described in Figure 7.7–13. The authors consider the EEG to be an envelope integrating simultaneous time-shifted voltage sequences produced by multiple processes similar to this example, occurring in numerous parallel networks offset by the multiplexing time delay as the environment is continuously sampled by the brain.

Neurophysiologic, psychophysiological, and fMRI research, for which reviews are available elsewhere, provides evidence supporting the concept that "mental chronometry" is reflected in the successive peaks of the cortical ERPs during the processing of information. The first positive component (P1) is associated with the arrival of information reporting the sensation of the stimuli, the negative-positive complex N1P2 is associated with attending to the input, the second positive cortical component (P2) is related to the perception of the stimuli, and a later positive component (P3) reflects the cognitive interpretation or "meaning" of the events.

Modulated Exogenous–Endogenous Interaction

As the described construction of consciousness proceeds, each perceptual frame or moment of subjective experience is constructed from interactions between three systems, analogous to a *mental molecule* rather than an atom of thought.

1. An "*exogenous system*" disperses spatiotemporal patterns of activity representing multimodal encoding of the environment to *basal dendrites* of pyramidal neurons throughout the neocortex, modulated by influences of internal state. Discrete attributes are distributed to many different brain regions to be analyzed before they are recombined. Information from different sensory modalities requires disparate times to be dispersed in pathways of different lengths and propagation velocities, so the sustained persistence of the perceptual frame may be essential for information from disparate modalities to be synthesized into one simultaneous multimodal percept.
2. In parallel, an "*endogenous system*" disperses spatiotemporal patterns of input to the *apical dendrites* of the pyramidal neurons representing the readout of memories about the most relevant and salient past experiences. Evidence suggests that an "auto-associative" readout process, based upon multimodal similarities between previous patterns

of activity and the exogenous report of the present, retrieves a facsimile of stored experiences from extensive networks in regions of the limbic system and the basal ganglia, as well as throughout the association regions of the cortex.

3. A “*state, modulation system,*” mediated by interneurons that reflect the state of sympathetic–parasympathetic balance of the autonomic nervous system and endocrine and adrenal activity, act upon large populations of pyramidal neurons at levels of basal or apical dendrites to weigh the relative contributions of the exogenous and endogenous systems, “reality” and “the past,” to the content of consciousness. These interneurons, by nonsynaptic modulation of membrane thresholds as well as synaptic input to pyramidal neurons, have the capability to differentially shift or bias the global thresholds of huge populations of neurons. They can desensitize or hypersensitize the exogenous system to certain kinds of afferent input and increase or diminish the salience of past experiences released from the endogenous systems.

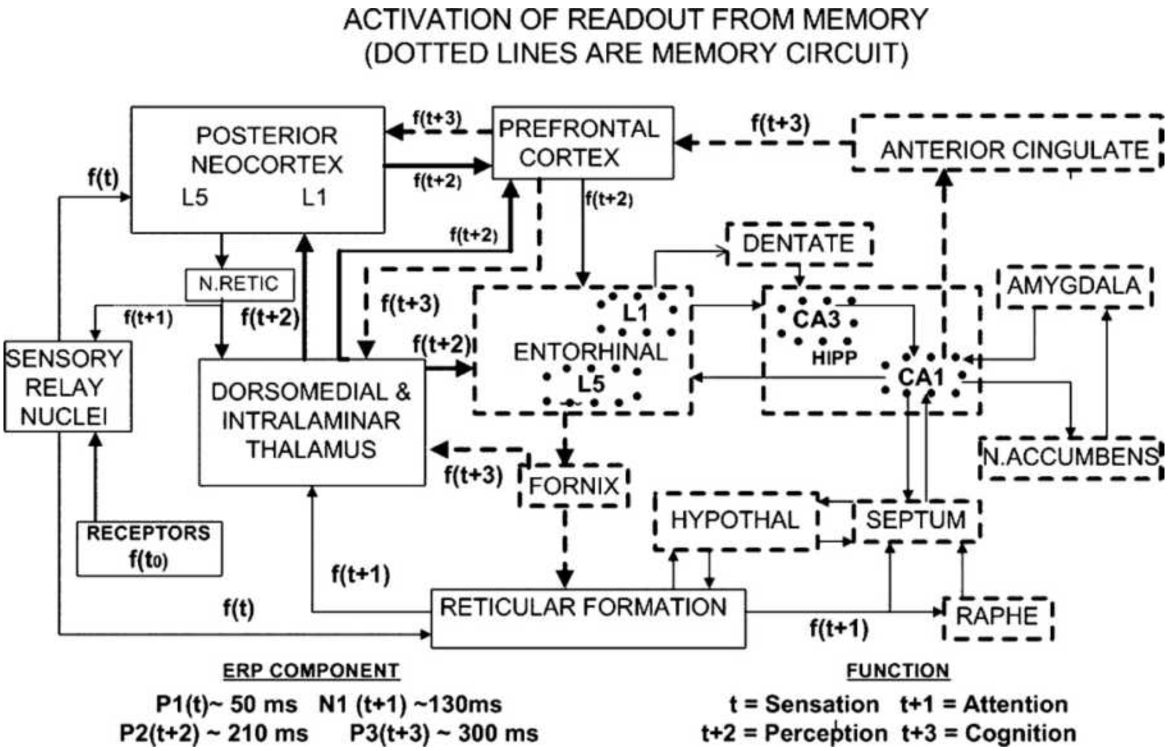
ACTIVATION OF READOUT FROM MEMORY
(DOTTED LINES ARE MEMORY CIRCUIT)
POSTERIOR NEOCORTEX
L5
L1
PREFRONTAL CORTEX
ANTERIOR CINGULATE
f(t+3)
f(t+2)
f(t)
N.RETIC
f(t+1)
SENSORY RELAY NUCLEI
DORSOMEDIAL & INTRALAMINAR THALAMUS
ENTORHINAL
DENTATE
CA3
HIPP
CA1
AMYGDALA
N.ACCUMBENS
FORNIX
HYPOTHAL
SEPTUM
RECEPTORS f(t0)
RETICULAR FORMATION
RAPHE
ERP COMPONENT
P1(t)~ 50 ms N1 (t+1) ~130ms
P2(t+2) ~ 210 ms P3(t+3) ~ 300 ms
FUNCTION
t = Sensation t+1 = Attention
t+2 = Perception t+3 = Cognition

FIGURE 7.7–13. Sequential steps of “mental chronometry” constructing a perceptual frame are postulated to occur simultaneously in numerous parallel processing channels, combining information from the exogenous and endogenous system into a seamless subjective moment of the “remembered present.” Incoming information from sensory receptors is propagated via the thalamocortical relay (TCR) nuclei to basal dendrites of feature-extracting pyramidal neurons dispersed throughout the neocortex in layer 5, producing the first positive (P50) event-related potential (ERP) component (at time *t*, about 50 ms) and simultaneously enters the reticular formation, registering the first elements of sensation. Activated cortical regions instruct the nucleus reticularis to inhibit selected regions of thalamic neurons to enhance the relevant signal-to-noise ratio, producing negative deflections (N150) (at time *t* + 1, about 130 ms) related to attention, while the ascending reticular arousal system (ARAS) sends information to the intralaminar nuclei and the raphe nuclei, hypothalamus, and septum. From the intralaminar nuclei, activity propagates to the apical dendrites of pyramidal cells of the posterior and associative cortical regions, producing the second positive (P200) ERP deflection (at time *t* + 2, about 210 ms) related to perception, and is sent as well from the dorsomedial thalamus to the basal dendrites of the prefrontal cortex and into the entorhinal cortex. Such incoming information from the exogenous system then circulates through the hippocampus and related structures of the limbic system, activating the most relevant past experiences by auto-associative mechanisms. Retrieved memories and their associated valence and context are projected via the entorhinal cortex and the anterior cingulate to the prefrontal cortex and thence to the posterior cortex, producing P300, the third positive ERP component (at time *t* + 3, about 300 ms) related

to cognition and subjective interpretation of the “meaning” of the percept. A late negative component, N400, reflects semantic encoding. Dispersed elements of sensations are linked into perceptions and interpreted cognitively in the context of past experience by binding mechanisms, reflected by the appearance of widespread gamma oscillations among dispersed anatomical regions.

Neuron–Interneuron Interactions

Each interneuron may contact as many as tens of thousands of pyramidal neurons, synchronizing theta and gamma oscillatory neuronal networks engaged in cognitive and emotional processing. Fast spiking (FS) basket cells surround the basal regions, which receive the major inputs from the exogenous system, while LTS basket cells can modulate apical dendrites, which receive the major inputs from the endogenous system and potentially affect mood and recollections relevant to depressive and schizophrenic pathophysiology. Long-range interneurons are “interregional” and facilitate rapid synchronization. Thus, interneuron influences continuously modulate the input and output to pyramidal cells serving as coincidence detectors between elements of the past and the present, as schematized in Figure 7.7–14.

In particular, the release by the locus ceruleus of serotonin and DA can differentially inhibit the release of GABA from FS interneurons to modulate attentional and arousing effects and salience of inputs from the exogenous system onto the basal dendrites of the pyramidal neurons while inhibiting the input of the LTS interneurons to the apical dendrites. Conversely, the release of an adrenal cortical hormone (ACH) and norepinephrine by the nucleus basalis can modulate the excitability of the LTS interneurons and selectively block GABA, disinhibiting thalamic neurons from the inhibitory influences of the nucleus reticularis.

LTS interneurons can display self-induced slow, long-lasting inhibition (SSI) mediated by endocannabinoids, decreasing their inhibitory influences. The resulting heightened excitability of the apical dendrites increases the influences of the endogenous system. Intense activation of the LTS interneurons would lead to SSI, giving an avalanche of positive feedback to the pyramidal neurons due to disinhibition. Imbalance in a specific neurotransmitter might activate a subset of GABAergic interneurons, synchronizing remote ensembles of pyramidal neurons into closely phase-locked rhythmic activity. Such synchronization could enhance or diminish the impact of the environment and experiences upon the exogenous and endogenous systems and their interactions, changing the flavor of the “remembered present.”

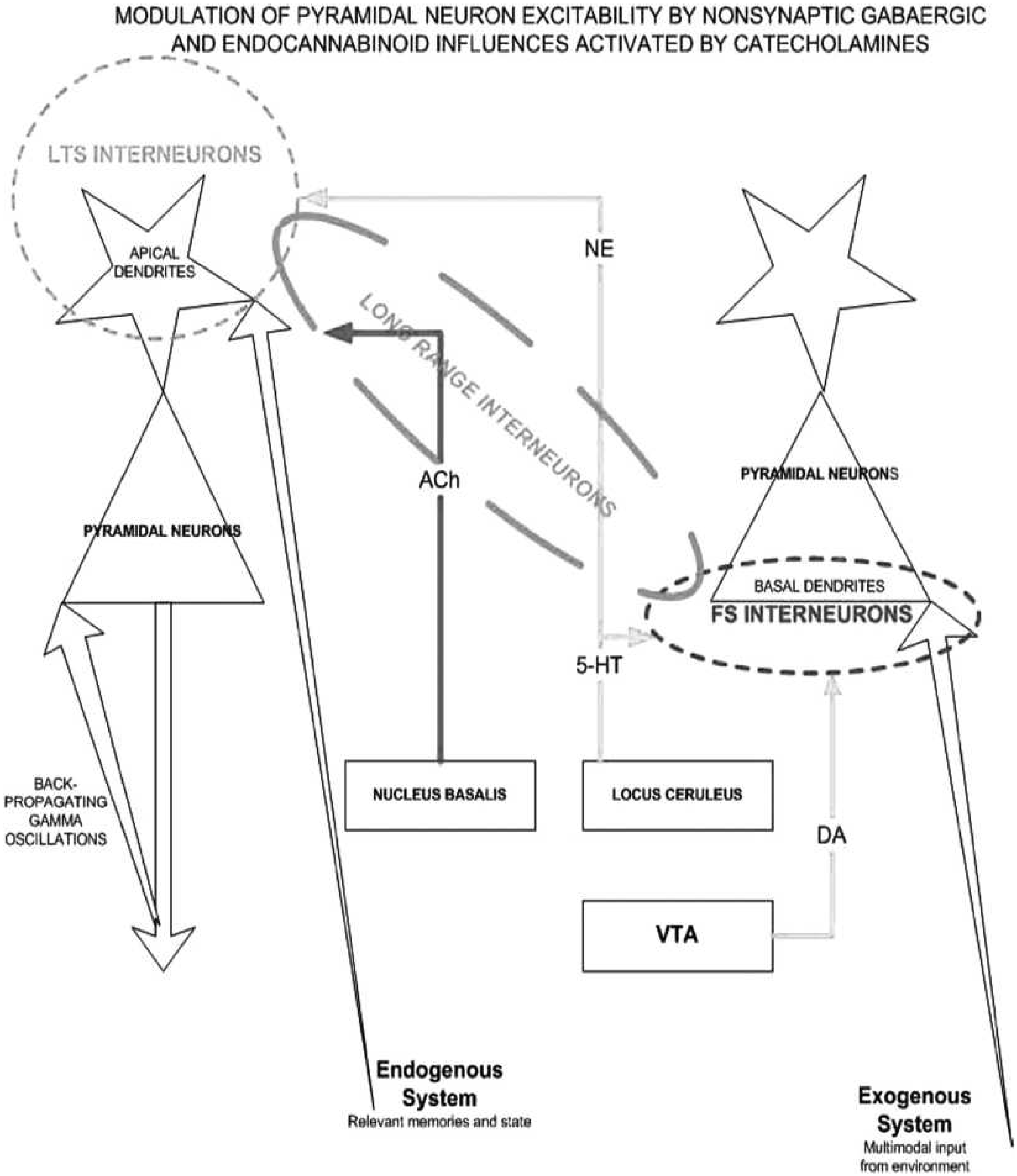
MODULATION OF PYRAMIDAL NEURON EXCITABILITY BY NONSYNAPTIC GABAERGIC AND ENDOCANNABINOID INFLUENCES ACTIVATED BY CATECHOLAMINES
LTS INTERNEURONS
APICAL DENDRITES
NE
LONG RANGE INTERNEURONS
ACh
PYRAMIDAL NEURONS
PYRAMIDAL NEURONS
BASAL DENDRITES
FS INTERNEURONS
5-HT
BACK-PROPAGATING GAMMA OSCILLATIONS
NUCLEUS BASALIS
LOCUS CERULEUS
DA
VTA
Endogenous System
Relevant memories and state
Exogenous System
Multimodal input from environment

FIGURE 7.7–14. Modulation of excitability of basal dendrites sensitive to exogenous influences by fast-spiking (FS) interneurons releasing dopamine (DA) and serotonin (5HT), reflecting the activity of the locus ceruleus and the ventral tegmental area (VTA), and of the excitability of apical dendrites sensitive to endogenous influences by low threshold spiking (LTS) interneurons releasing acetylcholine (Ach), reflecting the activity of nucleus basalis of Meynert and depicting the synchronizing influences on huge, anatomically dispersed ensembles of populations of cells in all categories that can be exerted by long-range interneurons, are depicted.

Coincidence Detection Constructs Fractionated Elements of Perception

Incoming sensory activity is dispersed by the TCR nuclei to the basal dendrites of pyramidal neurons in layers four and five of receiving regions throughout the sensory cortex, specialized to detect specific features of the multimodal focus of attention, as shown by red and blue arrows in the upper panel of Figure 7.7–15. Outputs from an “endogenous system” activated by this information and reflecting recent events, combined with signals reflecting internal state, are sent via the ARAS and the diffuse projection system to apical dendrites of the pyramidal neurons in the association and prefrontal cortices, where they are combined with input from primary receiving areas. This convergence, shown by mauve arrows in the middle panel of Figure 7.7–15, has been shown to be critical for the perception of a sensation to be accomplished.

Binding, Resonance, and Consciousness

The “binding problem” is the question of how attributes of multiple sensory modalities comprising these dispersed fragments of perception are recombined into a holistic perceptual moment of subjective experience, augmented by the associated affect, emotion, and motivational valence. Convergence of similar temporal patterns, arriving via the exogenous inputs upon the basal synapses and via the endogenous system upon the apical synapses of selected neurons in the layers of pyramidal cells dispersed throughout the cortex, results in a coherent corticothalamic discharge of synchronous volleys from this population, with a back-propagated oscillation in the gamma frequency range (~40 Hz). This oscillation causes these regions to depart from the ground state and establishes dispersed fragments of *local negative entropy,* that is, *improbably* synchronized spatiotemporal activity patterns that represent fractionated elements of perception.

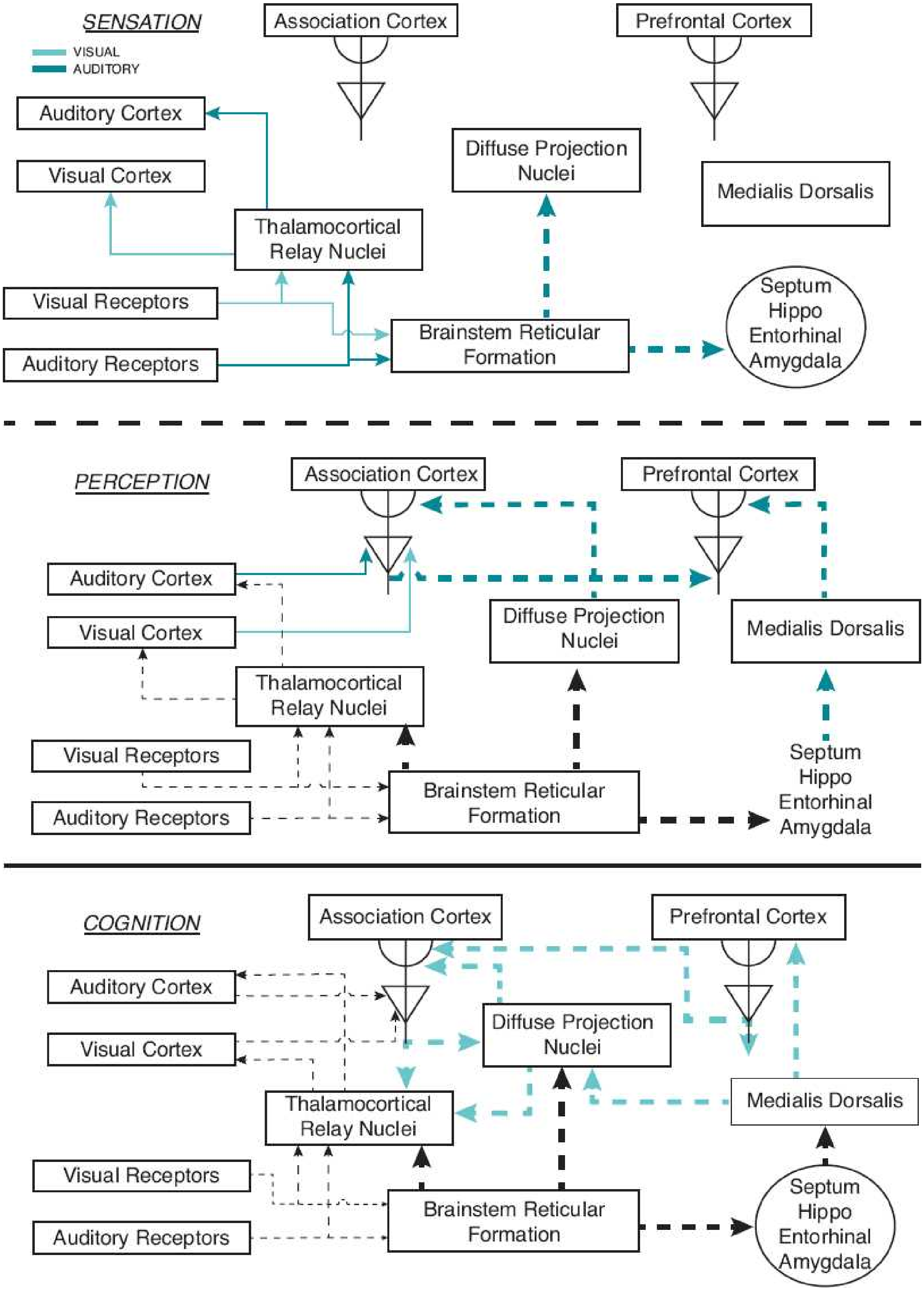
SENSATION
VISUAL
AUDITORY
Association Cortex
Prefrontal Cortex
Auditory Cortex
Visual Cortex
Diffuse Projection Nuclei
Medialis Dorsalis
Thalamocortical Relay Nuclei
Visual Receptors
Auditory Receptors
Brainstem Reticular Formation
Septum Hippo Entorhinal Amygdala
PERCEPTION
Association Cortex
Prefrontal Cortex
Auditory Cortex
Visual Cortex
Diffuse Projection Nuclei
Medialis Dorsalis
Thalamocortical Relay Nuclei
Visual Receptors
Auditory Receptors
Brainstem Reticular Formation
Septum Hippo Entorhinal Amygdala
COGNITION
Association Cortex
Prefrontal Cortex
Auditory Cortex
Visual Cortex
Diffuse Projection Nuclei
Medialis Dorsalis
Thalamocortical Relay Nuclei
Visual Receptors
Auditory Receptors
Brainstem Reticular Formation
Septum Hippo Entorhinal Amygdala

FIGURE 7.7–15. Schema of the stages of construction of conscious awareness and evaluation of percepts as derived from sensations are depicted. (*Top*) Sensations are produced by unimodal information (*red and*

blue) propagating from specific thalamocortical relay nuclei to dispersed sensory-specific cortical receiving areas, with distribution (*mauve*) into nonspecific thalamic and limbic system components of the endogenous system via the reticular formation. (*Middle*) Perception requires the coincidence of exogenous inputs (*red and blue*) to basal dendrites with inputs from the endogenous system (*mauve*) to apical dendrites of pyramidal cells in association and frontal cortical regions. (*Bottom*) Cognition is mediated by the binding together of activated dispersed cortical regions with activated specific and nonspecific thalamic nuclei in a coherent, phase-locked, reverberating gamma oscillation (*green*). (See eBook for color image.)

When enormous numbers of pyramidal neurons dispersed throughout numerous brain regions thus oscillate, reverberation links these regions together by phase-locked transcortical gamma oscillations. At the same time, neuronal ensembles comprising the networks of the endogenous system release memories retrieved by the autoassociative processes, which are reflected by gamma oscillations phase-locked to theta waves propagated to the prefrontal and association areas of the cortex. Cortical pyramidal cells in these regions appear to serve as coincidence detectors that coregister the present with the past. These spatially dispersed oscillations in the same high-frequency range, engaging the cortex, thalamus, subcortex, and limbic system, are physically brought into phase by electronic coupling, perhaps mediated by gap junctions and glial cells, and are thereby bound together. This system of temporally patterned, spatially dispersed but phase-locked, high-frequency oscillations of electrical energy begins to resonate, represented by the green arrows in Figure 7.7–13. Coherent gamma oscillations between prefrontal and parietooccipital areas of the cortex are essential for consciousness, as shown by their invariant reversible disappearance and reappearance with loss and return of consciousness in qEEG studies of anesthesia.

This resonance among spatially dispersed neuronal masses results in a shift of the brain from the homeostatically regulated ground state, which creates an extremely improbable informational field of global negative entropy. Phase-locking and resonance bind the exogenous and endogenous systems together into a continuously updating, globally coherent, anatomically dispersed yet cohesive functional network, whose operation is independent of any dedicated neuron or a specific set of neurons; the posterior regions of the brain (the sensory-specific, association, and motor cortices), dominated by inputs reporting the present state from the exogenous, autonomic, and musculoskeletal systems, are coupled together in global resonance with the anterior regions (the precentral and frontal cortices), dominated by the input of readout from the most relevant endogenous emotional, motivational, and motoric state and memory systems distributed among subcortical structures, including the basal ganglia, thalamus, hypothalamus, and limbic system.

Consciousness is postulated to be a physical property of this organized informational field of global negative entropy dynamically constructed in this spatially extensive resonating system, not a localized process mediated by discrete neuronal transactions. While neuronal activity generates this property, consciousness is not directly a neuronal process but rather a consequence emerging from mass action of enormous numbers of neurons, just as a nuclear explosion is not a property of the fission of an atom. Disturbances in the regulation of these complex transactions among masses of neural tissue are very relevant to the emergence of psychiatric disorders. Such disturbances might variously arise from dysfunction in the exogenous encoding mechanisms, the endogenous auto-associative retrieval mechanisms, or biasing of thresholds due to phasic or tonic fluctuations of the modulatory interneuron meshwork and can be envisaged potentially to generate psychopathological symptomatology. Malfunction of either or all of the exogenous, endogenous, and modulatory systems may distort the subjective experience to produce misperceptions and maladaptive behavior. Not only may each analytical apparatus be defective, but the binding process itself may be askew. The symptoms of many psychiatric disorders may reflect the disorganization of integrative processes dependent upon these binding mechanisms.

CONCLUSIONS

This section has provided an overview of the methods of qEEG and a summary of evidence of its sensitivity and specificity for detecting pathophysiologic conditions in a variety of psychiatric disorders and for identifying treatment-responsive and outcome-related subtypes within disorders. Identification of these distinctive profiles of disorders has been aided by the use of ML-driven classification methods. It has also called attention to the powerful brain imaging capabilities using qEEG source localization, allowing imaging of the mathematically most probable underlying sources of the scalp-recorded data superimposed on a normal MRI template, greatly expanding the potential to contribute to the understanding of the pathophysiology of psychiatric disorders and the clinical utility of qEEG in psychiatric populations. Growing evidence of the correlation between qEEG and other neuroimaging tools underscores the power of qEEG as a more readily available, nonradiation-emitting, cost-effective surrogate of advanced functional neuroimaging, which can be used at the point of care. Space did not permit a review of the contribution of sleep EEG. However, reviews of this literature exist.

This chapter has also presented a theoretical framework that the authors believe is pertinent to an understanding of how the measurements available with qEEG can be used to gain insights into the underlying neurophysiologic factors that may be contributory to the maladaptive behavior and emotional distress suffered by patients with a psychiatric/neuropsychiatric disorder. These symptoms can be thought of as reflecting perturbations of one or another aspect of the system that the authors have proposed to mediate the subjective content of conscious experience. The authors believe that systematic interactions between the symptoms experienced by the patient, the observations of the psychiatrist, and the careful evaluation and interpretation of functional assessments, obtained using readily available qEEG and source localization imaging technology evaluated relative to normative data, can focus attention upon disturbances of equilibrium in the brain. This, in turn, can lead to evidence-guided hypotheses about appropriate most optimal modes of treatment and targets for new and more personalized treatments, with iterated quantitative neurobiologic as well as psychiatric assessment of the appropriateness and efficacy of the treatment.

FURTHER READINGS

Arns M, Vollebregt MA, Palmer D, et al. Electroencephalographic biomarkers as predictors of methylphenidate response in attention-deficit/hyperactivity disorder. *Eur Neuropsychopharmacol*. 2018;28(8):881–891.

Bazarian JJ, Elbin RJ, Casa DJ, et al. *JAMA Network Open*. 2021;4(2):e2037349. doi:10.1001/jamanetworkopen.2020.3734

Bosch-Bayard J, Valdes-Sosa P, Virues-Alba E, et al. 3D Statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA). *Clin Electroencephalogr*. 2001;32:47–61.

Bruckner RL, Andrews-Hann J, Schacter D. The brain’s default network anatomy, function and relevance to disease. *Ann N Y Acad Sci*. 2008;1124:1–38.

Buzsaki G, Schomburg E. What does gamma coherence tell us about inter-regional neural communication? *Nat Neurosci*. 2015;18:484–489.

Chiarenza GA. Quantitative EEG in Childhood Attention Deficit Hyperactivity Disorder and Learning Disabilities. *Clinical EEG and Neuroscience*. 2020;52(2):144–155. doi:10.1177/1550059420962343

Goossens J, Laton J, Van Schependom J, et al. EEG dominant frequency peak differentiates between Alzheimer's disease and frontotemporal lobar degeneration. *J Alzheimers Dis*. 2017;55(1):53–58.

Hughes JR, John ER. Conventional and quantitative electroencephalography in psychiatry. *J Neuropsychiatry Clin Neurosci*. 1999;11:190–208.

Jiao B, Li R, Zhou H, et al. Neural biomarker diagnosis and prediction to mild cognitive impairment and Alzheimer's disease using EEG technology. *Alz Res Therapy.* 2023;15:32. https://doi.org/10.1186/s13195-023-01181-1

Jobert M, Wilson FJ, Ruigt GSF, Brunovsky M, Prichep LS, Drinkenburg WHIM. Guidelines for the recording and evaluation of pharmaco-EEG data in man: International Pharmaco-EEG Society (IPEG). *Neuropsychobiology*. 2012;66(4):201–220.

John ER. Consciousness: from synchronous neuronal discharges to subjective awareness? *Prog Brain Res*. 2005;150:143–171.

John ER, Prichep LS. The anesthetic cascade: how anesthesia suppresses awareness. *Anesthesiology*. 2005;102:447–471.

John ER, Prichep LS, Friedman J, Easton P. Neurometrics: computer-assisted differential diagnosis of brain dysfunctions. *Science*. 1988;293:162–169.

John ER, Prichep LS, Winterer G, Hermann WM, di Michele F. Electrophysiological subtypes of psychotic states. *Acta Psychiatr Scand*. 2007;116:17–35.

Lehmann D, Faber PL, Galderisi S, et al. EEG microstate duration and syntax in acute, medication-naive, first-episode schizophrenia: a multi-center study. *Psychiatry Res*. 2005;138:141–156.

Leuchter AF, Cook IA, Jin Y, Phillips B. The relationship between brain oscillatory activity and therapeutic effectiveness of transcranial magnetic stimulation in the treatment of major depressive disorder. *Front Hum Neurosci*. 2013;7:37.

Llinas R, Urbano FJ, Leznik E, Ramirez RR, van Marle HJF. Rhythmic and dysrhythmic thalamocortical dynamics: GABA systems and the edge effect. *Trends Neurosci*. 2005;28(6):325–333.

Loo SK, Makeig S. Clinical utility of EEG in attention-deficit/hyperactivity disorder: a research update. *Neurotherapeutics*. 2012;9(3):569–587.

Pascual-Marqui RD, Lehmann D, Koukkou M, et al. Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philos Trans A Math Phys Eng Sci*. 2011;369(1952):3768–3784.

Prichep LS, Jacquin A, Filipenko J, et al. Classification of traumatic brain injury severity using informed data reduction in a series of binary classification algorithms. *IEEE Trans Neural Syst Rehabil Eng*. 2012;20(6):806–822.

Prichep LS, John ER, Ferris SH, et al. Prediction of longitudinal cognitive decline in normal elderly with subjective complaints using electrophysiological imaging. *Neurobiol Aging*. 2006;27:471–481.

Prichep LS, Mas F, Hollander E, et al. Quantitative electroencephalographic (qEEG) subtyping of obsessive-compulsive disorder. *Psychiatr Res*. 1993;50:25–32.

Rossini PM, Miraglia F, Vecchio F. Early dementia diagnosis, MCI-to-dementia risk prediction, and the role of machine learning methods for feature extraction from integrated biomarkers, in particular for EEG signal analysis. *Alzheimers Dement*. 2022 Dec;18(12):2699–2706. doi: 10.1002/alz.12645.

Saletu B, Anderer P, Saletu-Zyhlarz GM, Pascual-Marqui RD. EEG mapping and low-resolution brain electromagnetic tomography (LORETA) in diagnoses and therapy of psychiatric disorders: evidence for a key-lock principle. *Clin EEG Neurosci*. 2005;36:108–115.

Thatcher RW, North DM, Biver CJ. LORETA EEG phase reset of the default mode network. *Front Hum Neurosci*. 2014;8:529.

Thibodeau R, Jorgensen RS, Kim S. Depression, anxiety, and resting frontal EEG asymmetry: a meta-analytic review. *J Abnorm Psychol*. 2006;115(4):715–729.

Van Doren J, Arns M, Heinrich H, Vollebregt MA, Strehl U, Loo SK. Sustained effects of neurofeedback in ADHD: a systematic review and meta-analysis. *Eur Child Adolesc Psychiatry*. 2019;28:293–305.

Wan L, Huang H, Schwab N. From eyes-closed to eyes-open: role of cholinergic projections in EC-to-EO alpha reactivity revealed by combining EEG and MRI. *Hum Brain Mapp*. 2019;40(2):566–577.

Wilde EA, Goodrich-Hunsaker NJ, Ware AL, et al. Diffusion tensor imaging indicators of white matter injury are correlated with a multimodal electroencephalography-based biomarker in slow recovering, concussed collegiate athletes. *J Neurotrauma*. 2020;37(19):2093–2101.

Wu W, Zhang Y, Jiang J, et al. An electroencephalographic signature predicts antidepressant response in major depression. *Nature Biotechnol*. 2020;38:439–447.

7.8 Psychiatric Rating Scales

Samuel Justin Sinclair, Ph.D., Mark A. Blais, Psy.D., and Deborah Blacker, M.D., Sc.D.

The term *psychiatric rating scales* encompasses a variety of questionnaires, interviews, checklists, outcome assessments, and other instruments that are available to inform psychiatric practice, research, and administration. Although new instruments and alternative versions of older instruments (especially shorter, self-report, or telephone versions) are developed regularly to improve quality and efficiency, to update diagnostic criteria, and to measure emerging constructs, they tend to be adopted slowly, as evidence of their usefulness accumulates. Older measures slowly fall out of favor as they become outdated and are supplanted by newer measures. Psychiatrists must keep up with major developments in rating scales for several reasons. Most critically, many such scales are useful in psychiatric practice for monitoring patients over time or for providing information that is more comprehensive than what is generally obtained in a routine clinical interview. In addition, healthcare administrators and payors are increasingly requiring standardized assessments to demonstrate the efficacy of psychiatric services or to assess quality of care. Last, but equally important, rating scales are used in research that informs the practice of psychiatry, so familiarity with them provides a deeper understanding of the results of that research and the degree to which it applies to psychiatric practice.

POTENTIAL BENEFITS AND LIMITATIONS OF RATING SCALES IN PSYCHIATRY

In its latest release of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), the American Psychiatric Association (APA) advocates for the use of psychometric tools in clinical practice as a means of improving quality of care at multiple levels. These new “emerging measures” are presented in Section III of the DSM-5, and include tools for both adults and children that are designed to assess an array of clinical domains (e.g., depression, anxiety, mania, etc.), which may then call attention to potential areas of difficulty and guide the clinical evaluation process. Likewise, more in-depth psychometric tools are also included to provide for more comprehensive assessments of specific domains (e.g., depression, mania, substance use), which then can be used to track symptom profiles over time. APA discusses a number of advantages of using these types of assessment tools, including the improvement of clinical precision and decision-making at the point of initial evaluation, a standardized method/metric for evaluating treatment efficacy over time, and finally as a framework for tracking patient-reported outcomes (PROs) more globally.

The key role of rating scales in psychiatry and elsewhere is to standardize the information collected across time and observers. This standardization ensures a consistent and comprehensive evaluation that may aid treatment planning by establishing a diagnosis, assuring a thorough review of symptoms, identifying comorbid conditions, and characterizing other factors affecting treatment response. In addition, the use of a rating scale can establish a baseline for following the progression of an illness over time or in response to specific interventions. This is particularly useful when more than one clinician is involved—for instance, in a hospital, group practice or research setting.

In addition to standardization, most rating scales provide users with empirical information about the measure's performance characteristics. This allows the clinician to know to what extent a given scale produces reproducible results (reliability) and how it compares to more established ways of measuring the same construct (validity).

Rating scales also offer some practical advantages. First, they offer carefully worded, pretested methods for obtaining information that can save valuable physician time—self-administered rating scales can be completed in the waiting room, or a nurse or technician can administer an interview before a session with the physician. In addition, rating scales may make it easier to obtain information about sensitive areas, such as cognitive decline or sexual side effects, in which direct questioning may be experienced as intrusive.

It is important to realize, however, that rating scales are not a panacea. They can provide erroneous measurements due to difficulties in administration, measurement error (or statistical imprecision), or limitations in the underlying construct. In this respect, they are no different than clinical assessments, but may appear to provide more definitive information and thus a spurious sense of security. At the practical level, they take time that might be better devoted to other clinical activities. The critical decision when deciding to use a formal assessment tool in clinical practice is whether, on balance, it contributes useful information in an efficient manner. This decision depends on the specific clinical setting and goal, the practical attributes of the scale, and its psychometric properties.

TYPES OF SCALES AND WHAT THEY MEASURE

Psychometric tools are used in psychiatric research and practice to achieve a variety of goals. They also cover a wide range of areas and use a broad array of procedures and formats.

Measurement Goals

Most psychiatric rating scales fall into one or more of the following categories: Making a diagnosis (e.g., the Structured Clinical Interview for the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* [DSM-5] [SCID] or the Diagnostic Interview Schedule for Children [DISC]); measuring severity and tracking change in specific symptoms (e.g., the Hamilton Anxiety Rating Scale [HAM-A] or the Mini-Mental State Examination [MMSE]), in general functioning (e.g., the Short-Form 36 [SF-36]), or in overall outcome (e.g., the Behavior and Symptom Identification Scale [BASIS-24]); and screening for conditions that may or may not be present (e.g., the CAGE).

Constructs Assessed

Psychiatric practitioners and investigators assess a broad range of areas, referred to as *constructs,* to underscore the fact that they are not simple, direct observations of nature. These include diagnoses, signs and symptoms, severity, functional impairment, quality of life, and many others. Some of these constructs are fairly complex and are divided into two or more domains (e.g., positive and negative symptoms in schizophrenia, or mood and neurovegetative symptoms in major depressive disorder). Many scales yield separate scores, or *subscales,* for each domain. In cases in which these domains are seen as substantially independent, they may be referred to as *dimensions* (e.g., the DSM-5 hybrid dimensional-categorical model proposed in Section III).

Categorical versus Continuous Classification. Some constructs are viewed as *categorical* or classifying, whereas others are seen as *continuous*. Categorical constructs describe the presence or absence of a given attribute (e.g., whether someone meets criteria for depression or not), or the category best suited to a given individual among a finite set of options (e.g., assigning a diagnosis). Continuous measures provide a quantitative assessment along a continuum of intensity, frequency, or severity (e.g., the magnitude or degree of someone's level of depression). In addition to symptom severity and functional status, multidimensional personality traits, cognitive status, social support, and many other attributes are generally measured continuously.

The distinction between categorical and continuous measures is by no means absolute. *Ordinal* classification, which uses a finite, ordered set of categories (e.g., unaffected, mild, moderate, or severe) stands between the two. In addition, a cutoff is frequently used with a continuous or ordinal scale to indicate a threshold for membership in a corresponding category. For instance, individuals with MMSE scores below 24 may be considered to have a dementia, or those with Hamilton Rating Scale for Depression (HAM-D) scores above 8 may be considered to have an episode of major depression.

Measurement Procedures

Rating scales differ in measurement methods. Issues to be considered include format, raters, and sources of information.

Format. Rating scales are available in a variety of formats. Some are simply checklists or guides to observation that help the clinician achieve a standardized rating. Others are self-administered questionnaires or tests. Still others are formal interviews that may be *fully structured* (i.e., specifying the exact wording of questions to be asked) or *semi-structured* (i.e., providing only some specific wording, along with suggestions for additional questions or probes). Whether fully structured or not, instruments may be written such that all questions are always included, or they may have formal skip-out sections to limit administration time.

Individual items also vary in format. Most commonly, scales use questions with yes–no or multiple-choice response options. Often, answers will be graded on a *Likert* scale, an ordinal scale with three to seven points that measures severity, intensity, frequency, or other attributes. Likert scales are most often partially or fully *anchored,* assigning a meaning to each numeric level. The same anchors can apply to all of the items, or the instrument may provide specific anchors for each. Occasionally, questionnaires include open-ended questions, especially at the beginning, which may be used to help establish rapport. In semi-structured or unstructured interviews, this information also serves to guide the rest of the interview and to aid in forming a clinical impression about the patient.

Raters. Some instruments are designed to be administered by doctoral-level clinicians only, whereas others may be administered by psychiatric nurses or social workers with more limited clinical experience. Still other instruments are designed primarily for use by lay raters with little or no experience with psychopathology. In general, more training is required to administer less structured scales. In addition, some scales require extensive training, even for experienced clinicians, to master the appropriate procedures and achieve a good result. Virtually all scales perform better when raters are familiar with their format and specific content.

Source of Information. Instruments also vary in the source of information used to make the ratings. Information may be obtained solely from the patient, who generally knows the most about their condition. In some instruments, some or all of the information may be obtained from a knowledgeable informant. When the construct involves limited insight (e.g., cognitive disorders or mania) or significant social undesirability (e.g., antisocial personality or substance abuse), other informants may be preferable. Informants may also be helpful when the subject has limited ability to recall or report symptoms (e.g., delirium, dementia, or any disorder in young children). Some rating scales also allow or require information to be included from medical records or from patient observation.

ASSESSMENT OF RATING SCALES

In clinical research, rating scales are mandatory to ensure interpretable and potentially generalizable results, and are selected based on coverage of the relevant constructs, expense (based on the nature of the raters, purchase price if any, and necessary training), length and administration time, comprehensibility to the intended audience, and quality of the ratings provided. In clinical practice, one considers these factors and, also, whether a scale would provide more or better information than what would be obtained in ordinary clinical practice or would contribute to the efficiency of obtaining that information. In either case, the assessment of quality is based on *psychometric,* or mind-measuring, properties.

Psychometric Properties

The two principal psychometric properties of a measure are *reliability* and *validity.* Although these words are used almost interchangeably in everyday speech, they are distinct in the context of evaluating rating scales. To be useful, scales should be *reliable,* or consistent and repeatable even if performed by different raters at different times or under different conditions; and they should be *valid,* or accurate in representing the true state of nature.

Relationship Between Reliability and Validity. Establishing a measure's reliability is generally considered primary, as it is difficult to reach valid judgments without first achieving consistency. However, problems with reliability can be overcome to an extent by combining information from several assessments. Unfortunately, improved reliability does not guarantee improved validity, and some efforts to improve reliability may actually limit validity. For example, a personality disorder instrument might focus on overt behaviors rather than inner thoughts and feelings to achieve higher reliability, but at the cost of losing some of the most valid information about personality. Even with clinically trained raters, it is particularly difficult to achieve reliability on items requiring subjective clinical judgment (e.g., feelings evoked in the examiner). Nonetheless, when used by experienced diagnosticians, such items may contribute substantially to valid diagnoses.

Reliability. *Reliability* refers to the consistency or repeatability of ratings and is largely empirical. In the categorical context, it refers to whether there is agreement on the classification of each individual. In the continuous context, it refers to whether there is agreement on the assignment of a given score. It can also be seen as precision (i.e., whether a measure yields a “ballpark” estimate or a finely graded score). An instrument is more likely to be reliable if the instructions and questions are clearly and simply worded, and the format is easy to understand and score. There are three standard ways to assess reliability: *Internal consistency, interrater,* and *test–retest.*

INTERNAL CONSISTENCY. Internal consistency assesses agreement among the individual items in a measure. This provides information about reliability, because each item is viewed as a single measurement of the underlying construct. Thus, the coherence of the items suggests that each is measuring the same thing. Internal consistency is measured most often with *coefficient alpha* (also known as *Cronbach coefficient alpha*), which ranges between 0 and 1—values of 0.70 are typically recommended when evaluating or making decisions about groups of people, whereas more stringent values (usually exceeding 0.85 or 0.90) are suggested when evaluating individuals. However, it should be noted that the internal consistency of a *measure* depends on the internal consistency of the *construct* that the measure purports to assess and will be higher for unidimensional constructs than for those with two or more relatively independent domains. For multidimensional constructs, it is generally preferable to calculate coefficient alpha on individual subscales.

INTERRATER AND TEST–RETEST RELIABILITY. Interrater (also called *interjudge* or *joint*) reliability is a measure of agreement between two or more observers evaluating the same subjects using the same information. Estimates may vary with assessment conditions—for instance, estimates of interrater reliability based on videotaped interviews tend to be higher than those based on interviews conducted by one of the raters. Interrater reliability tends to be higher than *test–retest reliability,* a measure of agreement between evaluations at two points in time in which there may be differences in the information obtained (e.g., associated with differences in interviewer skill, interviewer mood, room conditions, or subject’s attitude). In addition, test–retest evaluations measure reliability only to the extent that the subject’s true condition remains stable in the time interval, which is problematic for many conditions but virtually impossible for rapidly fluctuating conditions like state anxiety. However, the test–retest situation more closely reflects the clinical problems associated with serial evaluations by multiple clinicians, so to the extent that concerns about interval change can be eliminated, it is generally a more useful indicator of reliability in practice.

Interrater reliability and test–retest reliability of continuous constructs are measured with the intraclass correlation coefficient (ICC), whereas those of categorical constructs are measured with the kappa coefficient (κ). A weighted version of κ is available to give greater penalties for large disagreements than for small ones (e.g., between schizophrenia and psychotic depression, as compared to schizophrenia and schizoaffective disorder), or for ordinal measures. Both κ and the ICC are measures of agreement corrected for the agreement expected by chance alone, and both range from 0 to 1. As a rule of thumb, a κ or ICC above 0.8 is considered excellent, those in the 0.7 to 0.8 range are considered good, and those in the 0.6 to 0.7 range are considered fair. However, the degree of reliability required varies with the clinical purpose: Extremely reliable ratings are required to measure small changes in critical values (e.g., to detect small effects in a clinical trial or to decide when to stop a potentially toxic therapy), whereas more modest reliability may be satisfactory for other purposes (e.g., to detect large effects in a clinical trial or to estimate population means or rates).

ISSUES IN INTERPRETING RELIABILITY DATA. When interpreting reliability data, it is important to bear in mind that reliability estimates published in the literature may not generalize to other settings. Factors to consider are the nature of the sample, the training and experience of the raters, and the test conditions. Issues regarding the sample are especially critical. In particular, reliability tends to be higher in samples with high variability in which it is easier to make discriminations among individuals. Thus, for continuous measures, reliability tends to be higher when the sample includes individuals with a wide range of scores; and for categorical measures, reliability tends to be higher when the prevalence of the attribute being measured approaches 50%. Reliability estimates also depend on the fraction of difficult cases (e.g., individuals near a diagnostic threshold or those resistant to being interviewed), as large numbers of these tend to diminish observed reliability.

Validity. *Validity* refers to conformity with truth, or a gold standard that can stand for truth. In the categorical context, it refers to whether an instrument can make correct classifications. In the continuous context, it refers to accuracy, or whether the score assigned can be said to represent the true state of nature. Although reliability is an empirical question, validity is partly theoretical—for many constructs measured in psychiatry, there is no underlying absolute truth. Even so, some measures yield more useful and meaningful data than others do. Validity assessment is generally divided into face and content validity, criterion validity, and construct validity.

FACE AND CONTENT VALIDITY. *Face validity* refers to whether the items appear to assess the construct in question. Although a rating scale may purport to measure a construct of interest, a review of the items may reveal that it embodies a very different conceptualization of the construct. For instance, an insight scale may define *insight* in either psychoanalytic or neurologic terms. However, items with a transparent relationship to the construct may be a disadvantage when measuring socially undesirable traits, such as substance abuse or malingering. *Content validity* is similar to face validity, but differs in that it describes whether the measure provides good and balanced coverage of the construct, and is less focused on whether the items give the appearance of validity. Content validity is often assessed with formal procedures such as expert consensus or factor analysis.

CRITERION VALIDITY. *Criterion validity* (sometimes called *predictive* or *concurrent validity*) refers to whether or not the measure agrees with a gold standard or criterion of accuracy. Suitable gold standards include the long form of an established instrument for a new, shorter version; a clinician-rated measure for a self-report form; and blood or urine tests for measures of drug use. For diagnostic interviews, the generally accepted gold standard is the *L*ongitudinal, *E*xpert, *A*ll *D*ata (LEAD) standard, which incorporates expert clinical evaluation, longitudinal observations, medical records, family history, and any other sources of information.

When continuous measures are compared to a gold standard, a correlation coefficient is the statistic most often reported. For categorical variables, such as diagnoses (or continuous measures with a cutoff), various types of diagnostic efficiency statistics are employed depending on the goal of the study—most commonly *sensitivity* and *specificity*. *Sensitivity* refers to the test's ability to identify true cases, or its *true-*

positive rate. Specificity is the test's accuracy in identifying noncases, or one minus the *false-positive rate.* In general, the more sensitive a test, the less specific it is. If the threshold for diagnosis, for example, is lowered, then more cases will be detected but at the expense of some false positives; if the threshold is raised to decrease the false positives, then true cases will inevitably be missed. The optimal threshold depends on the consequences of false positives and false negatives. Calculation of predictive value positive and negative, which report, respectively, the proportion of positive and negative assessments that are correct *in a given population,* may help to answer this question. Predictive value positive, for instance, can be low even with excellent specificity in a low prevalence population, for instance, when measuring psychotic disorders in the general population. Where sensitivity is low, predictive value negative may be reduced, but less so if prevalence is also low.

CONSTRUCT VALIDITY. When an adequate gold standard is not available—a frequent state of affairs in psychiatry—or when additional validity data are desired, construct validity must be assessed. To accomplish this, one can compare the measure to *external validators,* attributes that bear a well-characterized relationship to the construct under study but are not measured directly by the instrument. External validators used to validate psychiatric diagnostic criteria and the diagnostic instruments that aim to operationalize them include course of illness, family history, and treatment response. For example, when compared to schizophrenia measures, mania measures are expected to identify more individuals with a remitting course, a family history of major mood disorders, and a good response to lithium.

Two special cases of assessing validity using external validators have particular relevance for clinical psychiatry. One is *discriminant validity,* which examines a measure's ability to discriminate between populations that are expected to differ on the construct of interest. For example, does a sociopathy measure correctly separate individuals in jails from those living in the community? Although such discriminations are important in clinical practice, it is important to remember that the true test of a measure is its ability to discriminate at the margins. A study of discriminant validity is more clearly relevant if it includes the types of cases encountered in clinical practice (e.g., psychotic depression versus schizoaffective disorder) rather than more easily discriminated populations (e.g., psychotic depression versus normal). Another special case is *sensitivity to change:* The fact that a measure shows expected changes (e.g., an improvement with an efficacious treatment or a decrement with a progressive disease) can be a strong validator.

When validity is assessed in areas in which there are few established measures and a gold standard or criterion of accuracy cannot be established, the assessment of the validity of the measure is limited by the validity of the construct itself. Nonetheless, by triangulation between a better definition of the construct, better ways to measure it, and better exploration of how it operates in clinical practice and research, the field moves to greater validity over time. Cronbach and Meehl (1955) referred to this process as defining the *nomological net* of constructs to be measured and ways in which they are thought to be interrelated, which then provides a framework for evaluating and establishing construct validity.

SELECTION OF PSYCHIATRIC RATING SCALES

In this section, a group of rating scales used in psychiatric practice and research is presented. The scales are grouped by topic, beginning with general issues, including diagnosis, functioning, symptom severity, and side effects, and then proceeding to specific diagnostic groups, organized according to the sections of the DSM-5. Selections were made based on coverage of major areas and common use in clinical research or current (or potential) use in clinical practice; instruments specifically recommended in DSM-5 were included whenever possible. For each area, there is a brief discussion of measurement issues followed by a description of each instrument, its psychometric properties, and its potential uses. In addition, whenever possible, a brief, clinically useful instrument is provided in each area. References for each measure, organized by topic, are listed in Table 7.8–1. These references include more detailed information about each measure and its psychometric properties and may also provide either the measure itself or instructions regarding how to obtain a copy. Access to instruments via the internet is increasing rapidly, but internet addresses evolve still more rapidly, so they are not included here. Instead, a quick search using Google or a similar search engine is recommended to locate a website that may offer the instrument for purchase or download, an up-to-date list of references, and other information.

Functional Status, Impairment, and General Symptom Severity

This broad area cuts across a variety of diagnoses and is thus useful for grading patients by functional status or overall severity without reference to specific symptomatology. Most of the instruments presented here have a strong mental health focus and often include items on psychiatric symptomatology.

The World Health Organization Disability Assessment Schedule. The World Health Organization Disability Assessment Schedule (WHODAS 2.0) was published in 2010 by the World Health Organization (WHO), and is a 36-item, self-administered measure of disability across six distinct domains including: understanding and communicating, getting around, self-care, getting along with people, life activities (i.e., household, work, and/or school activities), and participation in society. An informant (proxy) version of the tool is also available. Each of the 36 items is rated along a five-point Likert scale (1. None, 2. Mild, 3. Moderate, 4. Severe, 5. Extreme), and then items are summed within the six domains. Scores are linearly transformed to a 0 to 100 metric, where 0 represents no disability and 100 represents maximum disability. Based on DSM-5 field trials of the WHODAS 2.0 in adult populations, the APA has recommended that *average* scores for each domain be used as a means of helping clinicians to think about disability status in terms of the five anchor (scaling) points noted above (no disability, mild, moderate, severe, and extreme disability). This scoring method was found to be reliable, clinically useful, and with promising validity. It also provides clinicians with an alternative method for quantifying level of functioning (or degree of impairment) related to psychiatric symptoms given that DSM-IV's Global Assessment of Functioning (GAF) scale was not retained in DSM-5.

Behavior and Symptom Identification Scale. The Behavior and Symptom Identification Scale (BASIS-32) was developed in the early 1990s to provide a broad but brief overview of psychiatric symptoms and functional status from the patient's point of view for use in assessing the outcome of psychiatric treatment. A revised version, the BASIS-24 (Table 7.8–2), aimed at making the instrument shorter and more broadly applicable across demographic and diagnostic groups and in a wider variety of treatment settings, was released in 2004. The instrument assesses a wide range of areas, including family and work relationships; ability to complete regular tasks at home, work, or school; and symptoms of anxiety, depression, psychosis, and substance abuse. Each item is rated on a five-point scale focused on the degree of difficulty the individual has had in a given area during the preceding week. The BASIS-24 can be completed as a paper-and-pencil test (requiring 5 to 20 minutes), or the questions can be read aloud with the patient selecting the best answer from a laminated card (requiring 15 to 20 minutes). The BASIS-24 is scored by computer and generates an overall score and five subscales: relation to self and others, daily living and role functioning, depression and anxiety, impulsive and addictive behavior, and psychosis. Good reliability and validity have been demonstrated. Its simple administration, brevity, and broad coverage make it well suited to its original task, and it is frequently used at baseline, during treatment, and after treatment to monitor progress. It is capable of providing valid ratings across a wide range of psychiatric impairment but is not generally suitable for individuals with substantial cognitive impairment. It is also not suitable for children under 14 years of age.

Table 7.8–1.
Key References for Measures Included

Rating Scale	Reference
Functional Status, Impairment, and General Symptom Severity	
WHODAS	The World Health Organization (WHO) manual for WHO Disability Assessment Schedule (WHODAS 2.0), 2010, Geneva.
BASIS-24[a]	Eisen SV, Normand SLT, Belanger A, Spiro A, Esch D. The revised Behavior and Symptom Identification Scale (BASIS-24): Reliability and validity. *Med Care.* 2004;42:1230.
SF-36	McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short Form Health Survey (SF-36), II: Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care.* 1993;31:247.
Side Effects	
SAFTEE	Levine J, Schooler NR. SAFTEE: A technique for the systematic assessment of side effects in clinical trials. *Psychopharmacol Bull.* 1986;22:343.
AIMS[a]	Lane RD, Glazer WM, Hansen TE, Berman WH, Kramer SI. Assessment of tardive dyskinesia using the Abnormal Involuntary Movement Scale. *J Nerv Ment Dis.* 1985;173:353.
Simpson-Angus Rating Scale for Extrapyramidal Side Effects	Simpson GM, Angus JWS. A rating scale for extrapyramidal side effects. *Acta Psychiatr Scand*. 1970;46 (Suppl 212):11.
Psychiatric Diagnosis and Screening Measures	
SCID	First MB, Williams JBW, Karg RS, Spitzer RL: Structured Clinical Interview for DSM-5 Disorders, Clinician Version (SCID-5-CV). American Psychiatric Association; 2016.
MINI	Sheehan DV, Lecrubier Y, Harnett-Sheehan K, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar G: The Mini International

	Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview. *J Clin Psychiatry* 1998;59(suppl 20): 22–33.
PHQ	Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a new self-report version of PRIME-MD: The PHQ primary care study—Primary Care Evaluation of Mental Disorder, Patient Health Questionnaire. *JAMA.* 1999;282:1737.
DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measure—Adult	American Psychiatric Association (APA), 2013. http://www.psychiatry.org/practice/dsm/dsm5/online-assessment-measures
DSM-5 Parent/Guardian-Rated Level 1 Cross-Cutting Symptom Measure—Child Ages 6–17	American Psychiatric Association (APA), 2013. http://www.psychiatry.org/practice/dsm/dsm5/online-assessment-measures
DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measure—Child Ages 11–17	American Psychiatric Association (APA), 2013. http://www.psychiatry.org/practice/dsm/dsm5/online-assessment-measures

Broadband Measures of Psychological Functioning

MMPI-2	Butcher JN, Dahlstrom WG, Graham JR, Tellegen A, and Kaemmer B. *MMPI-2: Minnesota Multiphasic Personality Inventory–2: Manual for administration and scoring*. University of Minnesota Press; 1989.
MMPI-3	Ben-Porath Y, Tellegen A. MMPI-3. *Manual for Administration, Scoring, and Interpretation*. Pearson; 2020.
PAI	Morey LC. *Personality Assessment Inventory: Professional Manual* (2nd ed.). Psychological Assessment Resources; 2007.

Spectra: Indices of Psychopathology	Blais MB, Sinclair SJ. *Spectra: Indices of Psychopathology.* Psychological Assessment Resources (PAR); 2018.

Schizophrenia Spectrum and Other Psychotic Disorders

BPRS[a]	Overall JE, Gorham DR. The Brief Psychiatric Rating Scale (BPRS): A comprehensive review. *J Oper Psychiatry.* 1991;148:472.
PANSS	Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987;13:261.
SAPS/SANS	Andreasen NC, Arndt S, Miller D, Flaum M, Nopoulos P. Correlational studies of the Scale for the Assessment of Negative Symptoms and the Scale for the Assessment of Positive Symptoms: An overview and update. *Psychopathology.* 1995;28:7.

Depressive Disorders

HAM-D	Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatr.* 1960;23:56.
MADRS	Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry.* 1979; 134:382.
BDI-II	Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory-II*. Psychological Corporation; 1996.
PHQ-9[a]	Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric Ann.* 2002; 32:509–521.
IDS	Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychol Med.* 1996; 26:477.

Bipolar and Related Disorders

YMRS	Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: Reliability, validity and sensitivity.

	Br J Psychiatry. 1978;133:429.
ASRM[a]	Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman self-rating mania scale, *Biol Psychiatry* 1997;42: 948–955.

Anxiety Disorders

HAM-A	Hamilton M. The assessment of anxiety scales by rating. *Br J Psychol.* 1959;32:50.
GAD-7	Spitzer RL, Kroenke K, Williams JBW, Lowe B. A brief measure for assessing generalized anxiety disorder. *Arch Inern Med*. 2006;166:1092–1097.
The PROMIS Emotional Distress—Anxiety Short-Form	The Patient Reported Outcomes Measurement Information System (PROMIS: http://www.nihpromis.org/). The instrument and scoring / interpretation instructions are also available for use at: http://www.psychiatry.org/practice/dsm/dsm5/online-assessment-measures.
PDSS	Shear MK, Brown TA, Barlow DH, Money R, Sholomskas DE. Multicenter collaborative Panic Disorder Severity Scale. *Am J Psychiatry.* 1997;154:1571.

Trauma and Stressor-Related Disorders

CAPS	Blake DD, Weathers FW, Nagy LN, Kaloupe DG, Klauminzer G. A clinician rating scale for assessing current and lifetime PTSD: The CAPS-1. *Behav Therapist.* 1990;18:187.
PTSD Checklist for DSM-5 (PCL-5)[a]	Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *J Trauma Stress*. 2015;28:489–498.

Obsessive Compulsive and Related Disorders

YBOCS	Goodman WK, Price LH, Rasmussen SA, et al., The Yale-Brown Obsessive Compulsive Scale: I.

	Development, use, and reliability. *Arch Gen Psychiatry.* 1989;46:1006.
RTBS	Goodman WK, Storch E. LEVEL 2—Repetitive Thoughts and Behaviors—Adult (Adapted from the Florida Obsessive-Compulsive Inventory [FOCI] Severity Scale [Part B]; 1994. http://www.psychiatry.org/practice/dsm/dsm5/online-assessment-measures)

Substance Use Disorders

Level 2 Substance Use Module from DSM-5[a]	National Institute on Drug Abuse (NIDA). Adapted from the NIDA-modified ASSIST. http://www.psychiatry.org/practice/dsm/dsm5/online-assessment-measures
CAGE	Ewing JA. Detecting alcoholism: The CAGE questionnaire. *JAMA.* 1984;252:1905.
AUDIT	Saunders JB, Aasland OG, Babor TF, de le Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction.* 1993;88:791.
DAST	Skinner HA. The drug abuse screening test. *Addict Behav.* 1982;7:363.
ASI	McLellan AT, Kushner H, Metzger D, Peters F, Peters R. The fifth edition of the Addiction Severity Index. *J Subst Abuse Treat.* 1992;9:199.

Eating Disorders

EDE	Cooper Z, Fairburn CG. The Eating Disorder Examination: A semistructured interview for the assessment of the specific psychopathology of eating disorders. *Int J Eat Disord.* 1987;6:1.
EDA-5	Sysko R, Glasofer DR, Hildebrandt T, et al. The Eating Disorder Assessment for DSM-5 (EDA-5): Development and validation of a structured

	interview for feeding and eating disorders. *Int J Eat Disord.* 2015;48(5):452–63.
Cognitive Disorders	
MMSE	Folstein MF, Folstein SE, McHugh PR. “Mini-Mental State”: A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12:189.
MOCA[a]	Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment (MoCA©): A Brief Screening Tool For Mild Cognitive Impairment. *J Am Geriatr Soc.* 2005;53:695–699.
CDR	Morris JC. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology.* 1993;43:2412.
NPI	Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA. The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. *Neurology.* 1994;44:2308.
CAM	Inouye SK, VanDyck CH, Alessi CA, et al. Clarifying confusion: The Confusion Assessment Method. A new method for detecting delirium. *Ann Intern Med.* 1990; 113:941–948.
Personality Disorders	
SCID-5-PD	First MB, Williams JBW, Benjamin LS, Spitzer RL. *User’s Guide for the SCID-5-PD (Structured Clinical Interview for DSM-5 Personality Disorder)*. American Psychiatric Association; 2015.
PDQ	Hyler SE, Rieder RO, Williams JBW, Spitzer RL, Hendler J: The Personality Diagnostic Questionnaire: Development and preliminary results. *J Personal Disord.* 1988;2:229.
PID-5	Krueger RF, Derringer J, Markon KE, Watson D, Skodol AE. Initial construction of a maladaptive

	personality trait model and inventory for DSM-5. *Psychol Med.* 2012; 42, 1879–1890.
Childhood Disorders	
CBCL[a]	Achenbach TM. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. University of Vermont; 1991.
DISC	King C, Katz S, Ghaziuddin N, Brand E, Hill E. Diagnosis and assessment of depression and suicidality using the NIMH Diagnostic Interview Schedule for Children (DISC-2.3). *J Abnorm Child Psychol.* 1997;25:173.
Conners Rating Scales	Conners CK, Wells KC, Parker JD, Sitarenios G, Diamond JM. A new self-report scale for assessment of adolescent psychopathology: Factor structure, reliability, validity, and diagnostic sensitivity. *J Abnorm Child Psychol.* 1997;25:487.
Autism Diagnostic Interview—Revised	Lord C, Rutter M, Le Courteur A. Autism Diagnostic Interview Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord.* 1994;24:659.

Note: For full name of scales denoted by these abbreviations, please see relevant section of this chapter.

[a] Full text of scale is included in this chapter.

Table 7.8–2.
BASIS-24 (Behavior and Symptom Identification Scale)

Instructions to Staff: Please fill in the following information completely.

Client ID: _ _ _ _ _ _ _

HCO ID: _ _ _ _ _ _ _

Admission/Intake Date: _ _/_ _/_ _

Time Point:
$_1$☐ Admission/Intake
$_2$☐ Mid-treatment
$_3$☐ Discharge termination
$_4$☐ Post-treatment follow-up

Level of Care:
$_1$☐ Inpatient
$_2$☐ Outpatient
$_3$☐ Partial/day hospital
$_4$☐ Residential

Program Type or Unit: _ _

Instructions to Respondents:

This survey asks about how you are feeling and doing in different areas of life. Please check the box to the left of your answer that best describes yourself during the *PAST WEEK.* Please answer every question. If you are unsure about how to answer, please give the best answer you can.

EXAMPLE: ***During the past week,*** **how much difficulty did you have sleeping?**
$_0$☐ No difficulty
$_1$☐ A little difficulty
$_2$☐ Moderate difficulty
$_3$☐ Quite a bit of difficulty
$_4$☐ Extreme difficulty

During the PAST WEEK, how much difficulty did you have…

1. Managing your day-to-day life?
$_0$☐ No difficulty
$_1$☐ A little difficulty
$_2$☐ Moderate difficulty
$_3$☐ Quite a bit of difficulty
$_4$☐ Extreme difficulty

2. Coping with problems in your life?
$_0$☐ No difficulty
$_1$☐ A little difficulty
$_2$☐ Moderate difficulty
$_3$☐ Quite a bit of difficulty
$_4$☐ Extreme difficulty

3. Concentrating?
$_0$☐ No difficulty
$_1$☐ A little difficulty
$_2$☐ Moderate difficulty
$_3$☐ Quite a bit of difficulty
$_4$☐ Extreme difficulty

During the PAST WEEK, how much of the time did you…

4. Get along with people in your family?
$_0$☐ None of the time
$_1$☐ A little of the time
$_2$☐ Half of the time
$_3$☐ Most of the time
$_4$☐ All of the time

5. Get along with people outside your family?
$_0$☐ None of the time
$_1$☐ A little of the time
$_2$☐ Half of the time
$_3$☐ Most of the time
$_4$☐ All of the time

6. Get along well in social situations?
$_0$☐ None of the time
$_1$☐ A little of the time
$_2$☐ Half of the time
$_3$☐ Most of the time
$_4$☐ All of the time

During the PAST WEEK, how much of the time did you…

7. Feel close to another person?
$_0$☐ None of the time
$_1$☐ A little of the time
$_2$☐ Half of the time
$_3$☐ Most of the time
$_4$☐ All of the time

8. Feel like you had someone to turn to if you needed help?
$_0$☐ None of the time
$_1$☐ A little of the time
$_2$☐ Half of the time
$_3$☐ Most of the time
$_4$☐ All of the time

9. Feel confident in yourself?
$_0$☐ None of the time
$_1$☐ A little of the time
$_2$☐ Half of the time
$_3$☐ Most of the time
$_4$☐ All of the time

During the PAST WEEK, how much of the time did you…

10. Feel sad or depressed?
$_0$☐ None of the time
$_1$☐ A little of the time
$_2$☐ Half of the time
$_3$☐ Most of the time
$_4$☐ All of the time

11. Think about ending your life?
$_0$☐ None of the time
$_1$☐ A little of the time
$_2$☐ Half of the time
$_3$☐ Most of the time
$_4$☐ All of the time

12. Feel nervous?
$_0$☐ None of the time
$_1$☐ A little of the time
$_2$☐ Half of the time
$_3$☐ Most of the time
$_4$☐ All of the time

During the PAST WEEK, how often did you…

13. Have thoughts racing through your head?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

14. Think you had special powers?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

15. Hear voices or see things?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

16. Think people were watching you?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

17. Think people were against you?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

During the PAST WEEK, how often did you…

18. Have mood swings?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

19. Feel short-tempered?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

20. Think about hurting yourself?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

During the PAST WEEK, how often…

21. Did you have an urge to drink alcohol or take street drugs?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

22. Did anyone talk to you about your drinking or drug use?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

23. Did you try to hide your drinking or drug use?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

24. Did you have problems from your drinking or drug use?

$_0$☐ Never
$_1$☐ Rarely
$_2$☐ Sometimes
$_3$☐ Often
$_4$☐ Always

ABOUT YOU

25. How old are you?—

26. What is your sex? $_1$☐ Male
$_2$☐ Female

27. Are you … $_1$☐ Hispanic or Latino
$_2$☐ NOT Hispanic or Latino

28. What is your racial background?

$_1$☐ American Indian or Alaskan native
$_2$☐ Asian
$_3$☐ Black or African-American
$_4$☐ White/Caucasian
$_5$☐ Native Hawaiian or other Pacific Islander
$_6$☐ Multiracial or other (specify)

29. How much school have you completed?

$_1$☐ 8th grade or less
$_2$☐ Some high school
$_3$☐ High school graduate/GED
$_4$☐ Some college
$_5$☐ 4-year college graduate or higher

30. Are you now …

$_1$☐ Married
$_2$☐ Separated
$_3$☐ Divorced
$_4$☐ Widowed
$_5$☐ Never married

31. Outside of your treatment providers, what is your main source of social support?

$_1$☐ Wife, husband, or partner
$_2$☐ Other family (parents, children, relatives)
$_3$☐ Friends/roommates
$_4$☐ Community/church
$_5$☐ Other
$_6$☐ No one

32. Where did you sleep in the past 30 days?

$_1$☐ Apartment or house
$_2$☐ Halfway house/group home/board and care home/residential center/supervised housing
$_3$☐ School or dormitory
$_4$☐ Hospital or detox center
$_5$☐ Nursing home/assisted living
$_6$☐ Shelter/street
$_7$☐ Jail/prison
$_8$☐ Other (fill in)________________

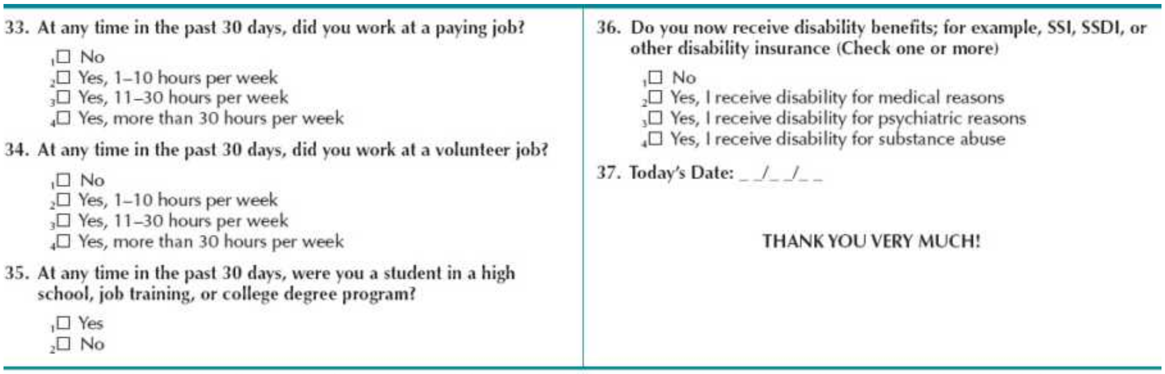

33. At any time in the past 30 days, did you work at a paying job?

$_1$☐ No
$_2$☐ Yes, 1–10 hours per week
$_3$☐ Yes, 11–30 hours per week
$_4$☐ Yes, more than 30 hours per week

34. At any time in the past 30 days, did you work at a volunteer job?

$_1$☐ No
$_2$☐ Yes, 1–10 hours per week
$_3$☐ Yes, 11–30 hours per week
$_4$☐ Yes, more than 30 hours per week

35. At any time in the past 30 days, were you a student in a high school, job training, or college degree program?

$_1$☐ Yes
$_2$☐ No

36. Do you now receive disability benefits; for example, SSI, SSDI, or other disability insurance (Check one or more)

$_1$☐ No
$_2$☐ Yes, I receive disability for medical reasons
$_3$☐ Yes, I receive disability for psychiatric reasons
$_4$☐ Yes, I receive disability for substance abuse

37. Today's Date: _ _/_ _/_ _

THANK YOU VERY MUCH!

Copyright McLean Hospital, 2001. Reprinted with permission of McLean Hospital, Belmont, MA; 2002–2021. *Note:* A license is required to administer the BASIS only through McLean eBASIS.

Short-Form 36. The SF-36 was developed in 1993 as a general measure of health status from the patient's point of view that would be independent of specific diseases and equally applicable to medical and psychiatric patients and to the general population. It is part of a broader family of instruments designed for this purpose. The SF-36 consists of 36 items, a mixture of yes–no questions and short Likert scales. It is focused on the individual's functional status as it relates to physical problems, pain, and emotional difficulties over the last 4 weeks. An acute version, focused on the past week, is also available, as are 8- and 12-item versions; the 4-week, 36-item version is the most widely used. The items are used to rate eight separate subscales: physical function, physical role function, pain, general health, vitality, social function, emotional role function, and mental health. Computer programs are available for administration and scoring. It is designed as a self-report instrument but can easily be read by an interviewer in person or over the phone. The instrument can be completed in 10 to 15 minutes. It has been subjected to extensive psychometric testing in clinical and general population samples, with extensive norms available. There is demonstrated high internal consistency for the eight subscales and good test–retest reliability over a 2-week period. Validity has been demonstrated by correlation with other functional scales and with various severity-of-illness and burden-of-care scales. The SF-36 is in wide use around the world, especially in health services research, and the instrument has special utility for comparing health outcomes across disparate diagnostic groups. It can also be used in clinical and health administration practice to track outcomes for individual patients or health systems and offers the benefit of patient-centered assessment of treatment and its impact.

Side Effects

The instruments described below are used to detect and quantify side effects from medications. One is a general interview that is applicable to all medications. The remaining two (the Abnormal Involuntary Movement Scale [AIMS] and the Simpson-Angus Rating Scale for Extrapyramidal Side Effects) are focused on the motor symptoms commonly observed with antipsychotics and, for this reason, include a brief focused physical examination, as well as questions posed to the patient.

Systematic Assessment of Treatment-Emergent Events. The Systematic Assessment of Treatment-Emergent Events (SAFTEE) is a systematic tool developed in 1986 to assess side effects in clinical trials. It has two versions, a general inquiry form (SAFTEE-GI) and a much more detailed specific inquiry form (SAFTEE-SI) that incorporates a formal review of symptoms. The SAFTEE-GI is rated based on three open-ended questions regarding new symptoms experienced in the past week: (1) Have you had any physical or health problems during the past week? (2) Have you noticed any changes in your physical appearance during the past week? (3) Have you cut down on the things you usually do because of not feeling physically well during the past week? To avoid bias on the part of the examiner or patient, the questions are focused on the novelty of the symptoms rather than on any relationship to the drug. The SAFTEE form provides specified terms (listed by organ system) for reporting the answers to the three questions and suggests specific ways to follow up on any positive answers regarding (1) the duration, temporal pattern, severity, and associated functional impairment for each; (2) any contributory factors other than the drug (e.g., other medication changes or intercurrent illness) and any hypothesized relationship of the drug to the event (e.g., dose response or noted in other subjects); and (3) any observed changes in behavior and appearance (e.g., shortness of breath or rash). The SAFTEE-GI takes 5 to 10 minutes to administer if there are few symptoms but could potentially take much longer. The SAFTEE-GI has fair to good reliability in use (somewhat better for the presence of symptoms than for detailed ratings such as severity). The SAFTEE-GI has good validity based on a comparison with similar measures and on its ability to detect expected side effects. The SAFTEE-GI is more sensitive to side effects than is typical clinical practice, and the SAFTEE-SI is more sensitive still. Although the SAFTEE-SI is only suitable for use in research settings, the

SAFTEE-GI could be used in clinical practice when a new drug is administered as a complement to specific questions about expected side effects of a particular drug.

Abnormal Involuntary Movement Scale. The AIMS (Table 7.8–3) is a clinical examination and rating scale developed in the 1970s to measure dyskinetic symptoms in patients taking antipsychotic drugs. The AIMS has 12 items, each of which is rated on an item-specific five-point severity scale ranging from 0 to 4. Total scores are not generally reported. Instead, changes in global severity and individual areas can be monitored over time. Ten items cover the movements themselves, divided in sections rating global severity and those related to specific body regions; two items concern dental factors that can complicate the diagnosis of dyskinesia. In the presence of extended neuroleptic exposure and the absence of other conditions causing dyskinesia, mild dyskinetic movements in two areas or moderate movements in one area suggest a diagnosis of tardive dyskinesia. The AIMS was developed for clinician raters, but lay raters can be trained to use it. It can be completed in less than 10 minutes. Excellent reliability has been demonstrated, especially for experienced raters, and the instrument appears valid. The AIMS is considered standard clinical practice for patients receiving long-term neuroleptic drugs in many clinical settings and is essential in clinical practice and research, both for monitoring patients for the development of tardive dyskinesia and for tracking changes in tardive dyskinesia over time.

Table 7.8–3.
Abnormal Involuntary Movement Scale (AIMS) Examination Procedure

Patient identification: ________________ Date: ________________

Rated by: ________________

Either before or after completing the examination procedure, observe the patient unobtrusively at rest (e.g., in waiting room).

The chair to be used in this examination should be a hard, firm one without arms.

After observing the patient, they may be rated on a scale of 0 (none), 1 (minimal), 2 (mild), 3 (moderate), and 4 (severe) according to the severity of symptoms.

Ask the patient whether there is anything in their mouth (i.e., gum, candy, etc.) and if there is to remove it.

Ask patient about the current condition of their teeth. Ask patient if they wear dentures. Do teeth or dentures bother patient now?

Ask patient whether they notice any movement in mouth, face, hands, or feet. If yes, ask to describe and to what extent they currently bother patient or interfere with their activities.

0 1 2 3 4	Have patient sit in chair with hands on knees, legs slightly apart, and feet flat on floor. (Look at entire body for movements while in this position.)
0 1 2 3 4	Ask patient to sit with hands hanging unsupported. If male, between legs; if female and wearing a dress, hanging over knees. (Observe hands and other body areas.)
0 1 2 3 4	Ask patient to open mouth. (Observe tongue at rest within mouth.) Do this twice.
0 1 2 3 4	Ask patient to protrude tongue. (Observe abnormalities of tongue movement.) Do this twice.
0 1 2 3 4	Ask the patient to tap thumb, with each finger, as rapidly as possible for 10–15 seconds; separately with right hand, then with left hand. (Observe facial and leg movements.)
0 1 2 3 4	Flex and extend patient's left and right arms. (One at a time.)
0 1 2 3 4	Ask patient to stand up. (Observe in profile. Observe all body areas again, hips included.)
0 1 2 3 4	Ask patient to extend both arms outstretched in front with palms down. (Observe trunk, legs, and mouth.)[a]
0 1 2 3 4	Have patient walk a few paces, turn and walk back to chair. (Observe hands and gait.) Do this twice.[a]

---

[a] Activated movements.

Simpson-Angus Rating Scale for Extrapyramidal Side Effects. The Simpson-Angus Rating Scale for Extrapyramidal Side Effects was developed to monitor the effects of antipsychotic drugs. It has ten items, each of which is rated on an item-specific five-point severity scale ranging from 0 to 4. Scores are reported as the mean on all ten items, with 0.3 considered the upper limits of normal. It is strongly focused on parkinsonian symptoms, particularly rigidity, but includes one akathisia item. It is designed for clinician use but can be administered by trained lay raters and takes approximately 10 minutes to administer. Good reliability has been reported, and validity is supported by the correlation of scores with antipsychotic drug dose. The scale is useful in a wide variety of clinical and research settings to monitor parkinsonian side effects and the impact of interventions to treat these effects.

Psychiatric Diagnosis and Screening

Instruments assessing psychiatric diagnosis are central to psychiatric research and may be useful in clinical practice as well. However, they tend to be rather long, especially when administered to individuals reporting many symptoms, potentially requiring many follow-up questions. When such instruments are evaluated, it is important to be sure that they implement current diagnostic criteria and cover the diagnostic areas of interest. For instance, few cover personality disorders (with the exception in some cases of antisocial personality), and not all cover disorders that typically onset in childhood.

Structured Clinical Interview for DSM. The SCID was initially developed in the early 1990s to provide a standardized DSM-III-R Axis I diagnosis based on an efficient but thorough clinical evaluation and is sometimes referred to as the SCID-I to distinguish it from the SCID-II, a similar measure for personality disorders. It was subsequently updated for the DSM-IV in 1994 and DSM-5 in 2015. The semi-structured diagnostic interview begins with a section on demographic information and clinical background. Then there are eight diagnostic modules focused on different diagnostic groups: mood, psychotic, substance abuse, anxiety, obsessive-compulsive and related disorders, feeding and eating disorders, externalizing disorders, and trauma and stressor-related disorders. Both required and optional probes are provided, and skip outs are suggested where no further questioning is warranted. All available information, including that from hospital records, informants, and patient observation, should be used to rate the SCID. The SCID is designed to be administered by experienced clinicians and is generally not recommended for use by lay interviewers. In addition, formal training in the SCID is required, and training books and videos are available to facilitate this. For individuals without symptoms, the interview takes approximately 1 hour to complete but may take up to 3 hours for individuals with extensive symptomatology. Although the primary focus is research with psychiatric patients, a nonpatient version (with no reference to a chief complaint) and a more clinical version (without as much detailed subtyping) are also available. Reliability data on the SCID suggest that it performs better on more severe disorders (e.g., bipolar disorder) or alcohol withdrawal than on milder ones (e.g., persistent depressive disorder [dysthymia]). Validity data are limited, as the SCID is more often used as the gold standard to evaluate other instruments. It is considered the standard interview to verify diagnosis in clinical trials and is extensively used in other forms of psychiatric research. Although its length precludes its use in routine clinical practice, the SCID can sometimes be useful to assure a systematic evaluation in psychiatric patients—for instance, on admission to an inpatient unit or at intake into an outpatient clinic. It is also used in forensic practice to assure a formal and reproducible examination.

The MINI International Neuropsychiatric Interview (MINI). The MINI International Neuropsychiatric Interview (MINI) is a structured diagnostic interview that was originally developed in 1990 by researchers in the United States and Europe for purposes of assessing DSM-III-R and ICD-10 psychiatric conditions. It has since been updated several times to evaluate for DSM-IV and DSM-5, and has been translated and validated in over 70 languages. It assesses the 17 most common psychiatric disorders in mental health—selected based on prevalence rates of 0.5% or greater in the general population. Similar to the SCID, the MINI uses a branching tree logic interview structure in order to quickly determine whether symptom criteria are met. In addition to the standard MINI, there are multiple other versions available for use in research settings, including for general screening, psychotic disorders studies, attention-deficit hyperactivity studies, eating disorders studies, tobacco use disorder studies, suicidality disorders, and tracking. There is also a MINI KID version that assesses the 30 most common psychiatric disorders in pediatric mental health—along with briefer screening versions organized by domain. The MINI is a proprietary instrument requiring a licensing agreement to use, and training is recommended. Psychometric studies of the MINI have indicated they yield similar reliability and validity coefficients as the SCID, but can be administered in a much shorter period of time.

Patient Health Questionnaire. The Patient Health Questionnaire (PHQ) is a self-administered screening instrument designed primarily for primary care. It was developed in 1999 based on the earlier Primary Care Evaluation of Mental Disorders (PRIME-MD). It focuses on more common disorders likely to be seen in primary care settings: major depressive disorder, panic disorder, somatoform disorder, binge-eating disorder, alcohol abuse, and subthreshold depression and anxiety. It has 54 questions and takes 5 to 10 minutes to complete and thus can easily be completed in the waiting room. The primary care provider then reviews the positive responses and derives diagnoses in 1 to 5 minutes from a review of the patient's answers and an accompanying checklist. Training in the use of the instrument or careful review of the instruction manual is recommended. Reliability appears to be fair to good, better for more severe diagnoses. Validity judged against psychiatrist evaluations is quite good. The PHQ appears to be useful for primary care settings but lacks appropriate coverage for the range of disorders seen in psychiatric settings. Shorter versions, the PHQ-9 and PHQ-2, have been developed explicitly for primary care screening for depression (see below).

The American Psychiatric Association's DSM-5 Cross-Cutting Symptom Measures. With the release of the DSM-5, the APA has proposed several screening measures that can be used at the point of initial evaluation in order to guide the process of diagnosis and treatment, or to track a range of symptoms that cut across diagnostic boundaries. These instruments are referred to as the DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measures, and self-administered forms are available for adults and children (aged 11 to 17). There is also a parent/guardian rated form for children (aged 6 to 17). The adult form contains 23 items tapping 13 domains (Depression, Anger, Mania, Anxiety, Somatic Symptoms, Suicidal Ideation, Psychosis, Sleep Problems, Memory, Repetitive Thoughts and Behaviors, Dissociation, Personality Functioning, and Substance Use); whereas the Child form contains 25 items tapping 12 domains (Somatic Symptoms, Sleep Problems, Inattention, Depression, Anger, Irritability, Mania, Anxiety, Psychosis, Repetitive Thoughts and Behaviors, Substance Use, and Suicidal Ideation). All items are rated along the same five-point Likert scale (None, Slight, Mild, Moderate, Severe), and the highest scores are recorded for each domain. These measures were developed using a combination of items drawn from existing validated measures (e.g., the two depression items were taken from the PHQ 2-item screening measure for depression) and novel items developed by the DSM-5 working group (e.g., the two personality functioning items and one dissociation item). To date, these tools do not include a specific norm-based scoring procedure. Rather, the current convention is to recommend further evaluation (e.g., via more detailed Level 2 scales [available in DSM-5-TR]) for any score of “2” (mild) or greater on any scale (or 1 or greater for suicidal ideation, substance use, and psychosis).

Broadband Measures of Psychological Functioning

The field of psychometrics has evolved considerably over the last 50 years, and has given rise to a number of broadband measures of psychological functioning that have been shown to be psychometrically robust and strong predictors of important clinical outcomes. Among the most well-known are the Minnesota Multiphasic Personality Inventory (MMPI), the Personality Assessment Inventory (PAI), and the more recent Spectra: Indices of Psychopathology. These types of measures allow clinicians to evaluate a wide array of important clinical domains (which may be particularly useful at the point of initial evaluation or in cases where diagnosis may be complex and/or ambiguous) while also assessing other important factors such as test validity (i.e., inconsistent and/or idiosyncratic responding), response style (e.g., whether someone may be presenting in an overly positive or negative manner), interpersonal style, and suicide and violence risk. The strengths of these measures include their comprehensive scope and more systematic evaluation of an array of clinically relevant domains. At the same time, such instruments may also be time-consuming and overly taxing for the patient, and thus careful consideration should be given a priori as to whether and under what conditions this type of assessment would prove most useful. For example, these assessments may be inappropriate in acute settings, where patients may lack the cognitive and emotional resources to complete such a long instrument, and effort may be better spent in administering more targeted (and briefer) assessments instead.

Over the last decade, there has also emerged an accumulating body of research revealing the presence of multiple higher-order domains of psychopathology, which may explain the covariance (i.e., comorbidity) often observed across many types of psychiatric disorders. These higher-order dimensions have commonly been referred to as: Internalizing (unipolar depression, generalized anxiety, social phobia, panic disorder, posttraumatic stress); Externalizing (alcohol abuse, drug abuse, impulsivity, aggression, and adult antisocial behavior); and Reality-Impairing pathology (e.g., psychotic experiences, paranoid ideation, grandiose ideation, and manic activation). Using this *hierarchical-dimensional model* of psychopathology as a framework, there have also recently emerged a number of broadband measures which evaluate these constructs at multiple levels—including the Spectra: Indices of Psychopathology.

The Minnesota Multiphasic Personality Inventory. The standard MMPI-2 is a 567-item, self-administered instrument that assesses a broad range of psychopathology, while also providing indicators of test validity, response style, and interpersonal functioning. It takes roughly 45 to 60 minutes to complete, and uses a standard computer scoring program that produces a profile of scores which clinicians are then able to interpret relative to established population norms. MMPI items are rated along the same True/False scale, which map onto 10 clinical scales (Hypochondriasis, Depression, Hysteria, Psychopathic Deviate, Masculinity/Femininity, Paranoia, Psychasthenia, Schizophrenia, Hypomania, Social Introversion). Likewise, the MMPI also includes a number of validity scales evaluating for response consistency, idiosyncratic responding, and negative and positive response styles. A wide range of other scales have also been developed *post-hoc,* and are included in the current scoring programs available from the test publisher. The original MMPI was initially developed in the early 1940s and restandardized in 1989 (MMPI-2), and there now exist several different versions of the tool, including one for adolescents (MMPI-A) and a somewhat shorter, reformatted version (MMPI-2-RF) based on more modern psychometric methods. The MMPI is among the most widely researched broadband measures of psychological functioning, and its validity and reliability have been well-established across a wide array of clinical and nonclinical settings.

Most recently, the MMPI-3 was published in 2020. A revision of the MMPI-2-RF, the MMPI-3 was developed specifically for purposes of updating item content and coverage, and providing new normative benchmarks for interpretation—with a strong emphasis on establishing sound psychometric properties. At 335 items, the instrument takes about 25 to 30 minutes to complete (per the test publisher), and can be administered electronically or with paper and pencil. Normative data is available both for a nationally representative (census-matched) sample of English-speaking adults (age 18+) living in the United States (*N* = 1,620), as well as a sample of 550 U.S. Spanish speakers. The MMPI-3 has also been translated into Spanish, French, and Canadian versions.

The Personality Assessment Inventory. The PAI was published in 1991, and is a 344-item, multi-scale, self-report assessment tool that assesses a wide range of psychopathology, not just personality. The instrument underwent multiple rounds of rigorous psychometric testing in order to maximize scaling properties (e.g., internal consistency reliability, item-to-scale convergence and divergence, etc.) and has been shown to possess strong construct validity. All 344 items are rated by the patient along the same four point Likert scale (False, Slightly True, Mainly True, Very True), and are then summed (within scales) and converted to T scores relative to a general U.S. population sample. Another advantage of the PAI is that scores are also plotted relative to a large clinical sample, which allows clinicians to make additional interpretations about profile elevations relative to other psychiatric samples. The PAI includes four validity indices assessing inconsistent (INC) and idiosyncratic (INF) responding, as well as negative (NIM) and positive response styles (PIM). It also measures eleven clinical constructs (scales) including: Somatic Complaints (SOM), Anxiety (ANX), Anxiety-Related Disorders (ARD), Depression (DEP), Mania (MAN), Paranoia (PAR), Schizophrenia (SCZ), Borderline Features (BOR), Antisocial Features (ANT), Alcohol Problems (ALC), and Drug Problems (DRG). Finally, it contains five scales evaluating suicide risk and factors that may impact the course of treatment (e.g., treatment amenability, availability of social supports, etc.), as well interpersonal style (i.e., dominance and warmth). The PAI has been well validated across an array of clinical and nonclinical settings, and is frequently used in clinical and nonclinical (e.g., fertility clinics, employee evaluations) settings as a more comprehensive screening tool.

Table 7.8–4.
The Spectra: Indices of Psychopathology

Scale	No. of Items	Construct	Dimension/Spectrum
Depression (DEP)	6	Depressive affect and ideation	Internalizing
Anxiety (ANX)	8	Worry and fear of losing control	Internalizing
Social Anxiety (SOC)	5	Social worry and avoidance behavior	Internalizing
Post-Traumatic Stress (PTS)	6	Intrusive memories, physical reactivity, and avoidance	Internalizing
Alcohol Problems (ALC)	6	Negative effects of alcohol use	Externalizing
Drug Problems (DRG)	6	Negative effects of drug use	Externalizing
Severe Aggression (AGG)	6	Explosive anger and violence	Externalizing
Antisocial (ANTI)	8	Impulsivity, deceitfulness, rule breaking	Externalizing
Psychosis (PSY)	6	Hallucinations and delusions	Reality Impairing
Paranoid Ideation (PAR)	6	Persecutory thoughts	Reality Impairing
Manic Activation (MAN)	5	Mental and physical acceleration	Reality Impairing
Grandiose Ideation (GRA)	6	Elevated sense of self worth	Reality Impairing
Cognitive Concerns	5	Perceived cognitive problems	Supplemental

(COG)			
Psychosocial Functioning (PF)	8	Perceived effective coping	Supplemental
Suicidal Ideation (SUI)	6	Suicide-related ideation	Supplemental

Reproduced with permission from Blais MA, Sinclair SJ. Introduction to the SPECTRA: Indices of Psychopathology: An assessment inventory aligned with the hierarchical-dimensional model of psychopathology [white paper]. PAR; 2019.

The Spectra: Indices of Psychopathology (SPECTRA). The SPECTRA is a brief, self-administered multi-scale measure of psychopathology and functioning that was published in 2018 (Table 7.8–4). Guided by recent advances in hierarchical-dimensional models of psychopathology, it is comprised of 96 items which assess 12 lower-order clinical constructs: Depression, Anxiety, Social Anxiety, PTSD, Alcohol Problems, Drug Problems, Antisocial behavior, Severe Aggression, Psychosis, Paranoid Ideation, Manic Activation, and Grandiose Ideation. These clinical constructs were selected *a priori* based on multivariate research over the last several decades that has demonstrated their strong associations with three higher-order spectra (Internalizing [INT], Externalizing [EXT], and Reality Impairing [RI]) which explain the observed relationships (covariation) among common DSM disorders. These three spectra aggregate to constitute the General Psychopathology Index (GPI), which provides for an overall marker of psychiatric complexity or burden (also referred to as the *p-factor* in clinical research). The SPECTRA can be administered in paper and pencil format or electronically, and completed in 15 to 20 minutes. Norm-based scores are calculated for the clinical scales, spectra domains, and overall GPI—and these are based on a U.S. census-matched standardization sample of *N* = 1,064 adults. Research examining the psychometric properties of the SPECTRA have been quite promising, with potential application in outpatient, inpatient, and forensic contexts.

The SPECTRA is unlike other existing broadband measures of psychopathology (e.g., the MMPI and PAI) in several important ways. First, it has a new underlying hierarchical framework for psychopathology based on recent research, and provides quantitative measures that are interpretable at various levels (i.e., the specific clinical syndromes, the higher-order spectra, and the more global burden of illness) (Figure 7.8–1). This has implications not only for diagnostic efficiency, but also for treatment planning (where clinical disorders may be phenotypically similar or dissimilar) and clinical prediction (particularly in cases where complexity and burden are high). Second, whereas historically the DSM and ICD have traditionally conceptualized psychiatric disorders as discrete/categorical entities, which in turn influenced how classic broadband measures (e.g., MMPI, PAI) are interpreted (using specific cut-scores to denote clinical significance), the SPECTRA employs a dimensional approach to interpretation where scores up and down the continuum can be used to make clinical inferences irrespective of specific cut-scores. Third, in addition to the core clinical domains noted above, the SPECTRA also contains brief scales assessing suicidal ideation, general cognitive concerns, and adaptive psychological functioning. These ancillary scales may be helpful in assessing risk, screening for subjective cognitive complaints, and identifying psychological strengths (e.g., agency, interpersonal supports)—which may be relevant to treatment planning. Finally, on a more practical level, the SPECTRA also provides an efficient assessment of common psychopathology in a significantly shorter period of time as compared to the MMPI-3 and PAI—both of which are roughly 70% longer.

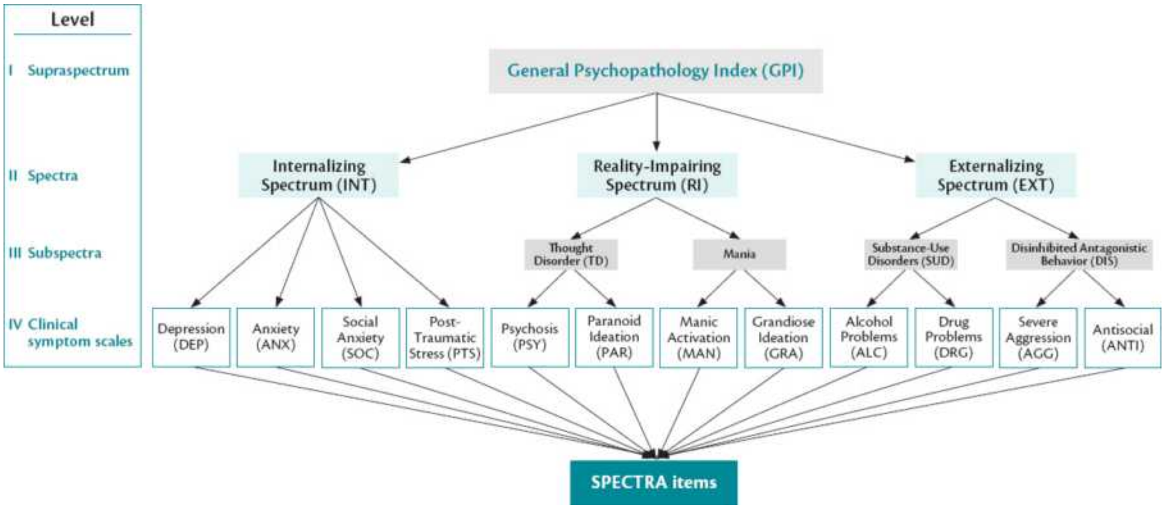
Level
I Supraspectrum
II Spectra
III Subspectra
IV Clinical symptom scales
General Psychopathology Index (GPI)
Internalizing Spectrum (INT)
Reality-Impairing Spectrum (RI)
Externalizing Spectrum (EXT)
Thought Disorder (TD)
Mania
Substance-Use Disorders (SUD)
Disinhibited Antagonistic Behavior (DIS)
Depression (DEP)
Anxiety (ANX)
Social Anxiety (SOC)
Post-Traumatic Stress (PTS)
Psychosis (PSY)
Paranoid Ideation (PAR)
Manic Activation (MAN)
Grandiose Ideation (GRA)
Alcohol Problems (ALC)
Drug Problems (DRG)
Severe Aggression (AGG)
Antisocial (ANTI)
SPECTRA items

FIGURE 7.8–1. The spectra: indices of psychopathology. (Reproduced with permission from Blais MA, Sinclair SJ. Introduction to the SPECTRA: Indices of Psychopathology: An assessment inventory aligned with the hierarchical-dimensional model of psychopathology [white paper]. PAR; 2019.)

Schizophrenia Spectrum and Other Psychotic Disorders

A variety of instruments are used for patients with psychotic disorders. Those reported here are symptom severity measures. A developing consensus suggests that the distinction between positive (e.g., hallucinations and delusions) and negative symptoms (e.g., emotional blunting, social apathy, etc.) in schizophrenia is worthwhile, and more recently developed instruments implement this distinction. Because patients with psychotic disorders often lack insight and are sometimes agitated, observation of the patient is required in addition to direct questioning. Thus, most instruments in this domain must be administered by psychiatrists or others with clinical training.

Brief Psychiatric Rating Scale. The BPRS (Table 7.8–5) was developed in the late 1960s as a short scale for measuring the severity of psychiatric symptomatology. It was developed primarily to assess change in psychotic inpatients and covers a broad range of areas, including thought disturbance, emotional withdrawal and retardation, anxiety and depression, and hostility and suspiciousness. Its 18 items are rated on a seven-point item-specific Likert scale from 0 to 6, with the total score ranging from 0 to 108 (in some scoring systems, the lowest level for each item is 1, so the range is 18 to 126). Because the ratings include observations, as well as patient reports of symptoms, the BPRS can be used to rate patients with very severe impairment. It is intended for use by experienced clinicians and can be administered in 30 minutes or less, including patient interview and observation. Reliability of the BPRS is good to excellent when raters are experienced, but this is difficult to achieve without substantial training; a semi-structured interview and guide has been developed to increase reliability. Validity is also good as measured by correlations with other measures of symptom severity, especially those assessing schizophrenia symptomatology. The BPRS has been used extensively for decades as an outcome measure in treatment studies of schizophrenia; it functions well as a measure of change in this context and offers the advantage of comparability with earlier trials. However, it has been largely supplanted in more recent clinical trials by the newer measures described below. In addition, given its focus on psychosis and associated symptoms, it is only suitable for patients with fairly significant impairment. Its use in clinical practice is less well supported, in part because considerable training is required to achieve the necessary reliability.

Positive and Negative Syndrome Scale. The Positive and Negative Syndrome Scale (PANSS) was developed in the late 1980s to remedy perceived deficits in the BPRS in the assessment of positive and negative symptoms of schizophrenia and other psychotic disorders by adding additional items and providing careful anchors for each. The PANSS includes 30 items on three subscales: seven items covering positive symptoms, seven covering negative symptoms, and 16 covering general psychopathology (e.g., guilt, uncooperativeness). Each item is scored on a seven-point, item-specific Likert scale ranging from 1 to 7; thus, the positive and negative subscales each range from 7 to 49, and the general psychopathology scale ranges from 16 to 112. The PANSS requires a clinician rater because considerable probing and clinical judgment are required. A semi-structured interview guide is available. The ratings can be completed in 30 to 40 minutes. Reliability for each scale has been shown to be fairly high, with excellent internal consistency and interrater reliability. Validity also appears good based on correlation with other symptom severity measures and factor analytic validation of the subscales. The PANSS has become the standard tool for assessing clinical outcome in treatment studies of schizophrenia and other psychotic disorders and has been shown to be easy to administer reliably and sensitive to change with treatment. Its high reliability and good coverage of both positive and negative symptoms make it excellent for this purpose. It may also be useful for tracking severity in clinical practice, and its’ clear anchors make it easy to use in this setting.

Table 7.8–5.
Brief Psychiatric Rating Scale

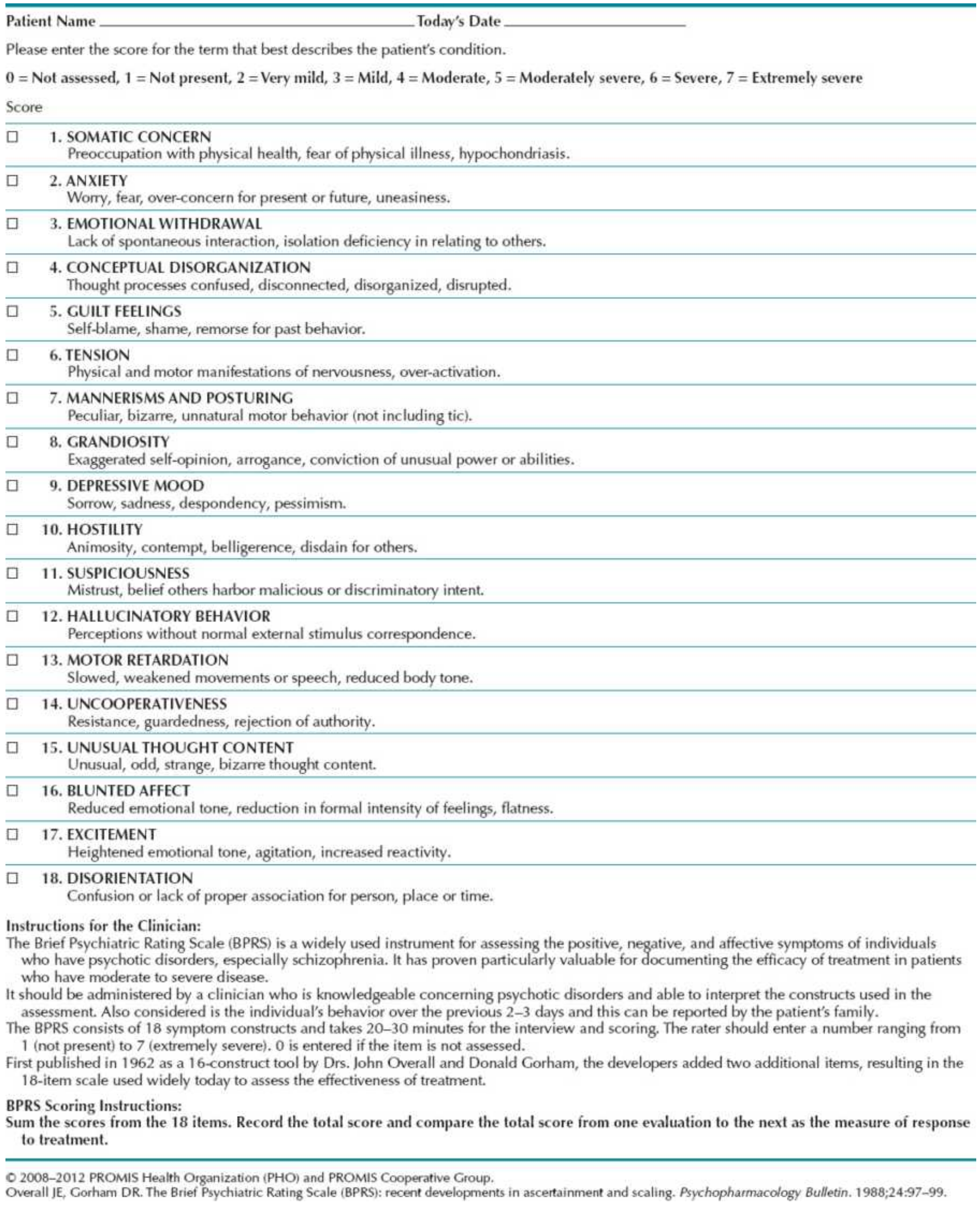

Patient Name ______________________ Today's Date ______________

Please enter the score for the term that best describes the patient's condition.

0 = Not assessed, 1 = Not present, 2 = Very mild, 3 = Mild, 4 = Moderate, 5 = Moderately severe, 6 = Severe, 7 = Extremely severe

Score

☐ 1. SOMATIC CONCERN
Preoccupation with physical health, fear of physical illness, hypochondriasis.

☐ 2. ANXIETY
Worry, fear, over-concern for present or future, uneasiness.

☐ 3. EMOTIONAL WITHDRAWAL
Lack of spontaneous interaction, isolation deficiency in relating to others.

☐ 4. CONCEPTUAL DISORGANIZATION
Thought processes confused, disconnected, disorganized, disrupted.

☐ 5. GUILT FEELINGS
Self-blame, shame, remorse for past behavior.

☐ 6. TENSION
Physical and motor manifestations of nervousness, over-activation.

☐ 7. MANNERISMS AND POSTURING
Peculiar, bizarre, unnatural motor behavior (not including tic).

☐ 8. GRANDIOSITY
Exaggerated self-opinion, arrogance, conviction of unusual power or abilities.

☐ 9. DEPRESSIVE MOOD
Sorrow, sadness, despondency, pessimism.

☐ 10. HOSTILITY
Animosity, contempt, belligerence, disdain for others.

☐ 11. SUSPICIOUSNESS
Mistrust, belief others harbor malicious or discriminatory intent.

☐ 12. HALLUCINATORY BEHAVIOR
Perceptions without normal external stimulus correspondence.

☐ 13. MOTOR RETARDATION
Slowed, weakened movements or speech, reduced body tone.

☐ 14. UNCOOPERATIVENESS
Resistance, guardedness, rejection of authority.

☐ 15. UNUSUAL THOUGHT CONTENT
Unusual, odd, strange, bizarre thought content.

☐ 16. BLUNTED AFFECT
Reduced emotional tone, reduction in formal intensity of feelings, flatness.

☐ 17. EXCITEMENT
Heightened emotional tone, agitation, increased reactivity.

☐ 18. DISORIENTATION
Confusion or lack of proper association for person, place or time.

Instructions for the Clinician:

The Brief Psychiatric Rating Scale (BPRS) is a widely used instrument for assessing the positive, negative, and affective symptoms of individuals who have psychotic disorders, especially schizophrenia. It has proven particularly valuable for documenting the efficacy of treatment in patients who have moderate to severe disease.

It should be administered by a clinician who is knowledgeable concerning psychotic disorders and able to interpret the constructs used in the assessment. Also considered is the individual's behavior over the previous 2–3 days and this can be reported by the patient's family.

The BPRS consists of 18 symptom constructs and takes 20–30 minutes for the interview and scoring. The rater should enter a number ranging from 1 (not present) to 7 (extremely severe). 0 is entered if the item is not assessed.

First published in 1962 as a 16-construct tool by Drs. John Overall and Donald Gorham, the developers added two additional items, resulting in the 18-item scale used widely today to assess the effectiveness of treatment.

BPRS Scoring Instructions:

Sum the scores from the 18 items. Record the total score and compare the total score from one evaluation to the next as the measure of response to treatment.

© 2008–2012 PROMIS Health Organization (PHO) and PROMIS Cooperative Group.

Overall JE, Gorham DR. The Brief Psychiatric Rating Scale (BPRS): recent developments in ascertainment and scaling. *Psychopharmacology Bulletin*. 1988;24:97–99.

Scale for the Assessment of Positive Symptoms and Scale for the Assessment of Negative Symptoms. The Scale for the Assessment of Positive Symptoms (SAPS) and Scale for the Assessment of Negative

Symptoms (SANS) were designed to provide a detailed assessment of positive and negative symptoms of schizophrenia and may be used separately or in tandem. The domains assessed include hallucinations, delusions, bizarre behavior, and thought disorder for the SAPS and affective flattening, poverty of speech, apathy, anhedonia, and inattentiveness for the SANS. Each instrument consists of 30 fully anchored items each scored from 0 to 5; thus, the total score ranges from 0 to 150 for each. Each must be rated by an experienced clinician and requires approximately 30 minutes to complete. Good to excellent interrater reliability has been demonstrated for trained interviewers, and each scale has high internal consistency as well. Validity is supported by correlation with other symptom severity instruments. The SANS and SAPS are principally used to monitor treatment effects in clinical research and have also been used to help characterize positive and negative symptoms in studies of schizophrenia phenomenology. The SAPS has been to some extent supplanted by the PANSS, but the SANS still has a place in studies aimed at negative symptoms. The instruments are longer and more difficult to learn to use than the PANSS and thus generally too detailed for use in clinical practice. However, the comprehensive characterization of symptomatology provided might be useful as part of a formal evaluation.

Depressive Disorders

Rating depression depends, to a substantial extent, on subjective assessment of mood states, so interviews and self-report instruments are both common. Because depression is common in the general population and involves significant morbidity and even mortality, screening instruments—especially those using a self-report format—are potentially quite useful in primary care and community settings. A number of interviewer and self-administered instruments are described below.

Hamilton Rating Scale for Depression. The HAM-D was developed in the early 1960s to monitor the severity of major depression, with a focus on somatic symptomatology. The version in most common use has 17 items, although versions with different numbers of items (including the 24-item version) have been used in many studies as well. The 17-item version does not include some of the symptoms of depression in the DSM-III and its successors, most notably the so-called reverse neurovegetative signs (increased sleep, increased appetite, and psychomotor retardation).

Items on the HAM-D are scored from 0 to 2 or from 0 to 4, with total score ranging from 0 to 50. Scores of 7 or less may be considered normal; 8 to 13, mild; 14 to 18, moderate; 19 to 22, severe; and 23 and above, very severe. The HAM-D was designed for clinician raters but has been used by trained lay administrators as well. Ratings are completed by the examiner based on patient interview and observations. A structured interview guide has been developed to improve reliability. The ratings can be completed in 15 to 20 minutes. Reliability is good to excellent, particularly when the structured interview version is used. Validity appears good based on correlation with other depression symptom measures. The HAM-D has been used extensively to evaluate change in response to pharmacologic and other interventions and thus offers the advantage of comparability across a broad range of treatment trials. It is more problematic in the older adults and the medically ill, in whom the presence of somatic symptoms may not be indicative of major depression.

Montgomery-Asberg Depression Rating Scale. The Montgomery-Asberg Depression Rating Scale (MADRS) was developed in 1979 explicitly for clinical trials, and thus was aimed at maximizing sensitivity to change in established depression. It can be administered by a psychiatrist, other physician, or psychiatric nurse, and a structured interview guide and self-report version are also available. It includes 10 items on seven-point, partially anchored, Likert scales. It can be administered in 10 to 15 minutes and is easy to use with minimal training. Very good reliability and validity have been demonstrated. Because of its sensitivity to change, the MADRS is frequently used in clinical trials of antidepressants and may also have a role in assessing the impact of treatment in clinical practice.

Beck Depression Inventory. The Beck Depression Inventory (BDI) was developed in the early 1960s to rate depression severity, with a focus on behavioral and cognitive dimensions of depression. The current version (the BDI-II) has added more coverage of somatic symptoms to be compatible with more current versions of the DSM and covers the most recent 2 weeks. Earlier versions are focused on the past week or even shorter intervals, which may be preferable for monitoring treatment response. The BDI includes 21 self-report items, each of which has four statements describing increasing levels of severity, and the total score ranges from 0 to 84. Scores of 0 to 9 are considered minimal; 10 to 16, mild; 17 to 29, moderate; and 30 to 63, severe. The scale can be completed in 5 to 10 minutes. Internal consistency has been high in numerous studies. Test–retest reliability is not consistently high, but this may reflect changes in underlying symptoms. Validity is supported by correlation with other depression measures. The principal use of the BDI is as an outcome measure in clinical trials of interventions for major depression, including psychotherapeutic interventions. Because it is a self-report instrument, it is sometimes used to screen for major depression. However, in trials of various cutoffs for a diagnosis of major depression, the best seemed to be 9, which had only fair sensitivity, at a cost of considerable nonspecificity, suggesting that the instrument has limited use for screening. The instrument's strength lies in measuring the depth of depression and in its comprehensive coverage of the cognitive dimension of depression.

The Patient Health Questionnaire-9-Item Screener for Depression. The Patient Health Questionnaire-9 (PHQ-9) is a nine-item, self-administered screening measure that was developed in 1999 (as part of the larger PHQ) as a brief tool for assessing depressive symptomatology in primary care settings (Table 7.8–6). The PHQ-9 items are rated along a four-point Likert scale (0. Not at all, 1. Several Days, 2. More than half of the days, and 3. Nearly every day) and then summed, with total scores ranging from 0 to 27. Although some have suggested cut-scores of 10 (or between 8 and 11) for a diagnosis of Major Depressive Disorder, the APA has also provided a system for interpreting different PHQ-9 score ranges in terms of severity: None = 0 to 4; Mild Depression = 5 to 9; Moderate Depression = 10–14; Moderately Severe Depression = 15 to 19; and Severe Depression = 20 to 27. The PHQ-9 in particular has generally performed well in terms of its underlying psychometric properties, and studies evaluating validity have also been promising. The APA has recommended the PHQ-9 as a brief screening tool for evaluating and tracking depressive symptomatology over time in DSM-5. A still shorter version, the PHQ-2, is also available as a depression screener, with more common application in primary care settings.

Table 7.8–6. The PHQ-9

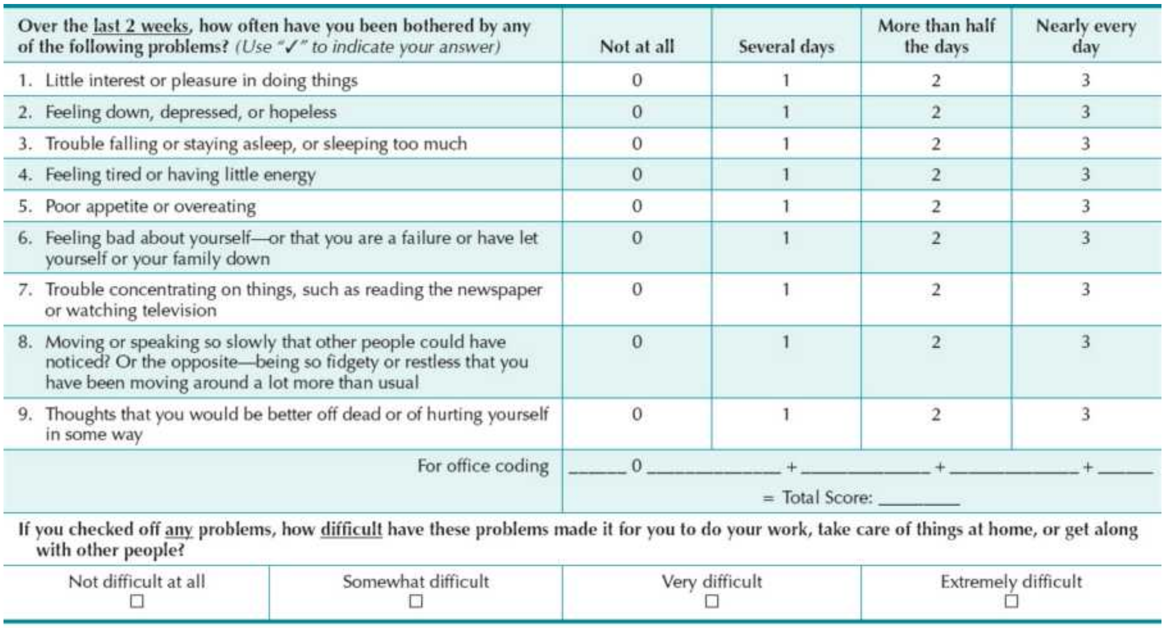

Over the last 2 weeks, how often have you been bothered by any of the following problems? *(Use "✓" to indicate your answer)*	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself—or that you are a failure or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed? Or the opposite—being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9. Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3
For office coding	______ 0 ______	+ ______	+ ______	+ ______
	= Total Score: ________			

If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all ☐	Somewhat difficult ☐	Very difficult ☐	Extremely difficult ☐

Developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke and colleagues, with an educational grant from Pfizer Inc.

Inventory of Depressive Symptomatology and Quick Inventory of Depressive Symptomatology. The Inventory of Depressive Symptomatology (IDS) was developed in 1982 to reflect a broader conceptualization of major depression than the HAM-D and other earlier instruments and to provide a comprehensive assessment of depressed inpatients and outpatients. It has better coverage of the full nature and severity range of DSM-5 symptoms of depression. A shorter version, the 16-item Quick Inventory of Depressive Symptomatology (QUIDS), was released in 2000. Both clinician-rated (IDS-C and QUIDS-C) and self-report (IDS-SR and QUIDS-SR) versions of each instrument are available. The clinician versions come with structured interview guides to improve the reliability of the ratings. Each item has four to five scaled multiple-choice answers, ranging from no difficulty to significant difficulty in that area. The IDS-SR requires 15 to 20 minutes, and the IDS-C 20 to 30 minutes; the corresponding versions of the QUIDS can be completed in half the time. The IDS has been subjected to rigorous reliability and validity assessment. It shows good reliability in a variety of paradigms and excellent validity, as judged by comparison with other depression scales. It has also been shown to have good validity in the classification of depressive subtypes and to be more sensitive to the milder end of the spectrum than older depression instruments. The QUIDS also has very good reliability and validity but does not attempt to measure subtypes. Both instruments have considerable utility in clinical research and clinical practice for the evaluation of depression and assessment of its response to treatment.

Bipolar and Related Disorders

In evaluating mania, as for psychotic disorders, limited insight and agitation may hinder accurate symptom reporting, so clinician ratings including observational data are generally required. Likewise, whether someone is actively experiencing symptoms of mania/hypomania at the point of evaluation may also affect whether and how current and past manic symptoms are endorsed on psychometric tools, and thus a careful retrospective analysis via clinical interview is also recommended to more systematically rule this out. A patient who accurately reports *no* symptoms of mania at the time of evaluation may in fact have a history of mania in the past and meet criteria for Bipolar Disorder—reflecting the episodic nature of these symptoms. A number of instruments that assess for mania/hypomania are summarized below.

Young Mania Rating Scale. The Young Mania Rating Scale (YMRS) is a checklist developed in the late 1970s to provide a brief but thorough evaluation of the severity of mania that could be used to monitor treatment response or detect relapse. It consists of a checklist of 11 items rated either from 0 to 4 (seven items) or from 0 to 8 (four items). Each item has five item-specific anchors. The total score ranges from 0 to 60. Ratings include clinical observation, so it must be rated by a clinician, but reliable ratings have been obtained by nurses on inpatient units and beginning psychiatric residents. Reliability is good based on interrater reliability and internal consistency studies. Validity also appears to be good based on limited data on correlation with other mania measures. The YMRS is useful for evaluating response to treatment in clinical research, and it has been found to be sensitive to change in this setting. It might also be used to assess treatment response or to monitor for relapse in treated or untreated patients, although extensive experience with this use has not been reported.

The Altman Self-Rating Mania Scale (ASRM). The Altman Self-Rating Mania Scale (ASRM) was initially developed in the late 1990s by Altman and colleagues, and consists of five items which assess the presence and/or severity of manic symptoms (Table 7.8–7). Items are rated along unique five-point Likert scales, where patients are asked to rate their symptoms with respect to the last 7 days. In cases where patients may be too impaired to complete the tool, an informant version is also available. Total scores range from 5 to 25, where higher scores are suggestive of greater pathology. Scores of 6 or higher have been shown to be associated with a higher probability of a manic or hypomanic condition, and may signal the need for further evaluation and/or treatment. The ASRM has been shown to possess strong psychometric properties (including test–retest and internal consistency reliability), as well good convergent and discriminant validity (e.g., discriminating patient groups who are manic from other kinds of psychiatric disorders). The ASRM has also been found to be sensitive to change as a function of pharmacotherapy. The APA has recommended the ASRM as a brief measure for assessing mania/hypomania in the Emerging Measures section of the DSM-5.

Anxiety Disorders

The anxiety disorders addressed by the measures below include panic disorder and generalized anxiety disorder. When anxiety measures are examined, it is important to be aware that there have been changes over time in how anxiety disorders are defined. For example, Post Traumatic Stress Disorder (PTSD) and Obsessive-Compulsive Disorders are no longer included in the DSM-5's Anxiety Disorders section, and generalized anxiety disorder has shifted somewhat to focus more on the affective (worry, fear) dimension of the condition. Thus, older measures have somewhat less relevance for diagnostic purposes, although they may identify symptoms causing considerable distress. Whether reported during an interview or on a self-report rating scale, virtually all measures in this domain, like the measures of depression discussed above, depend on subjective descriptions of inner states.

Hamilton Anxiety Rating Scale. The HAM-A was developed in the late 1950s to assess anxiety symptoms, both somatic and cognitive. Because the conceptualization of anxiety has changed considerably, the HAM-A provides limited coverage of the “worry” required for a DSM-5 diagnosis of generalized anxiety disorder and does not include the episodic anxiety found in panic disorder. There are 14 items, each of which is rated from 0 to 4 on an unanchored severity scale, with the total score ranging from 0 to 56. A score of 14 has been suggested as the threshold for clinically significant anxiety, but scores of 5 or less are typical in individuals in the community. The scale is designed to be administered by a clinician, and formal training or the use of a structured interview guide is required to achieve high reliability. A computer-administered version is also available. Reliability is fairly good based on internal consistency, interrater, and test–retest studies. However, given the lack of specific anchors, reliability should not be assumed to be high across different users in the absence of formal training. Validity appears good based on correlation with other anxiety scales but is limited by the relative lack of coverage of domains critical to the modern understanding of anxiety disorders. Even so, the HAM-A has been used extensively to monitor treatment response in clinical trials of generalized anxiety disorder and may also be useful for this purpose in clinical settings.

The Generalized Anxiety Disorder-7. The Generalized Anxiety Disorder-7 (GAD-7) is a seven-item, self-administered measure that was published in 2006 as a general screening measure for symptoms of generalized anxiety. It consists of seven items that inquire about general worry and fear symptoms in the last 2 weeks, all of which are rated along a four-point Likert scale (0. Not at all, 1. Several Days, 2. More than half the days, and 3. Nearly every day). Scores range from 0 to 21 (where higher scores indicate greater levels of symptomatology), and cut-scores of 5, 10, and 15 have been suggested for mild, moderate, and severe GAD, respectively. The GAD-7 has been found to possess strong psychometric properties, and its use in clinical practice has increased in recent years. That being said, although it remains a useful tool for evaluating more generalized symptoms of anxiety, it may not provide the same type of correspondence to a DSM-5-TR diagnosis given differences in timeframe (the GAD-7 asks about symptoms in the last 2 weeks, whereas DSM-5 requires that symptoms be present more days than not for 6 months) and DSM-5’s greater emphasis on the affective dimension of worry/fear.

The Patient-Reported Outcomes Measurement Information System Emotional Distress—Anxiety Short-Form. The Patient-Reported Outcomes Measurement Information System (PROMIS) Emotional Distress—Anxiety Short-Form is a seven-item measure of generalized symptoms of anxiety and worry that was developed by the larger PROMIS. PROMIS was launched in 2004 by the National Institute of Health (NIH) in an effort to develop psychometrically sound measures of patient-reported measures of physical health, mental well-being, and social functioning. Items from the PROMIS Emotional Distress—Anxiety Short-Form are rated along the same five-point Likert scale and then summed, with total scores ranging from 7 to 35. These raw (sum) scores are then converted into T scores using a conversion table provided by PROMIS, which allows for immediate comparison relative to the general population. There are measures available for both adults (18+) and children (6 to 17). The PROMIS Emotional Distress—Anxiety Short-Form has demonstrated strong preliminary reliability and validity and was recommended for use in assessing anxiety in DSM-5, although given its more recent development, additional research is needed to substantiate this research base before widespread adoption.

Panic Disorder Severity Scale. The Panic Disorder Severity Scale (PDSS) was developed in the 1990s as a brief rating scale for the severity of panic disorder. It was based on the Yale-Brown Obsessive-Compulsive Scale (YBOCS) and has seven items, each of which is rated on an item-specific, five-point Likert scale. The seven items address frequency of attacks, distress associated with attacks, anticipatory anxiety, phobic avoidance, and impairment. The items are scored from 0 to 4, and the total score ranges from 0 to 28. The instrument was designed for use by clinicians, but a self-report version is also available. Reliability is excellent based on interrater studies, but in keeping with the small number of items and multiple dimensions, internal consistency is limited. Validity is supported by correlations with other anxiety measures, both at the total and item level; lack of correlation with the HAM-D; and, more recently, by brain imaging studies. Growing experience with the PDSS suggests that it is sensitive to change with treatment and is useful as a change measure in clinical trials or other outcome studies for panic disorder, as well as for monitoring panic disorder in clinical practice.

Table 7.8–7.
The Altman Self Rating Mania Scale (ASRM)

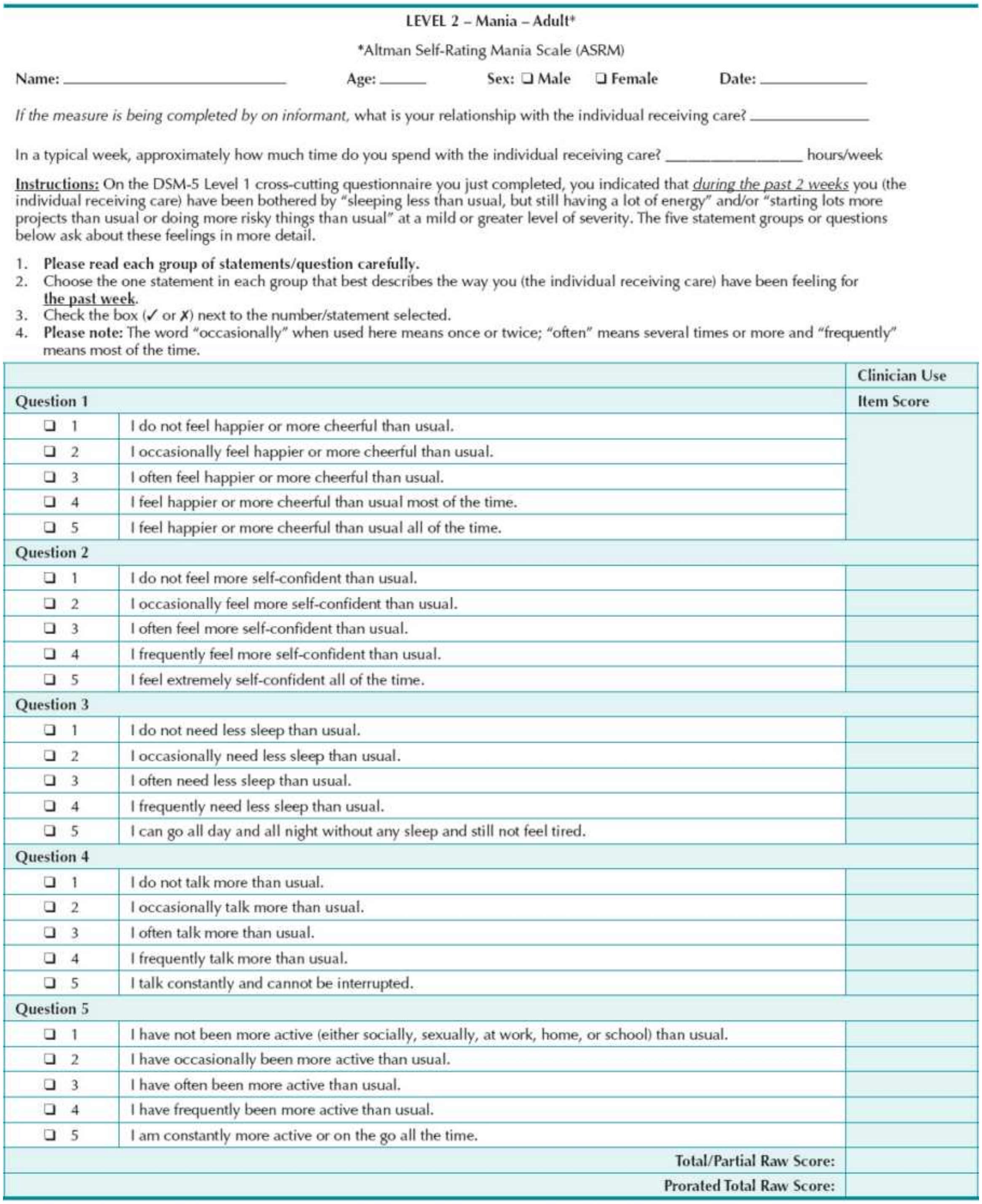

LEVEL 2 – Mania – Adult*

*Altman Self-Rating Mania Scale (ASRM)

Name: ______________________________ **Age:** ______ **Sex:** ☐ Male ☐ Female **Date:** ______________

If the measure is being completed by an informant, what is your relationship with the individual receiving care? ________________

In a typical week, approximately how much time do you spend with the individual receiving care? __________________ hours/week

Instructions: On the DSM-5 Level 1 cross-cutting questionnaire you just completed, you indicated that *during the past 2 weeks* you (the individual receiving care) have been bothered by "sleeping less than usual, but still having a lot of energy" and/or "starting lots more projects than usual or doing more risky things than usual" at a mild or greater level of severity. The five statement groups or questions below ask about these feelings in more detail.

1. **Please read each group of statements/question carefully.**
2. Choose the one statement in each group that best describes the way you (the individual receiving care) have been feeling for **the past week**.
3. Check the box (✓ or ✗) next to the number/statement selected.
4. **Please note:** The word "occasionally" when used here means once or twice; "often" means several times or more and "frequently" means most of the time.

		Clinician Use
Question 1		**Item Score**
☐ 1	I do not feel happier or more cheerful than usual.	
☐ 2	I occasionally feel happier or more cheerful than usual.	
☐ 3	I often feel happier or more cheerful than usual.	
☐ 4	I feel happier or more cheerful than usual most of the time.	
☐ 5	I feel happier or more cheerful than usual all of the time.	
Question 2		
☐ 1	I do not feel more self-confident than usual.	
☐ 2	I occasionally feel more self-confident than usual.	
☐ 3	I often feel more self-confident than usual.	
☐ 4	I frequently feel more self-confident than usual.	
☐ 5	I feel extremely self-confident all of the time.	
Question 3		
☐ 1	I do not need less sleep than usual.	
☐ 2	I occasionally need less sleep than usual.	
☐ 3	I often need less sleep than usual.	
☐ 4	I frequently need less sleep than usual.	
☐ 5	I can go all day and all night without any sleep and still not feel tired.	
Question 4		
☐ 1	I do not talk more than usual.	
☐ 2	I occasionally talk more than usual.	
☐ 3	I often talk more than usual.	
☐ 4	I frequently talk more than usual.	
☐ 5	I talk constantly and cannot be interrupted.	
Question 5		
☐ 1	I have not been more active (either socially, sexually, at work, home, or school) than usual.	
☐ 2	I have occasionally been more active than usual.	
☐ 3	I have often been more active than usual.	
☐ 4	I have frequently been more active than usual.	
☐ 5	I am constantly more active or on the go all the time.	
	Total/Partial Raw Score:	
	Prorated Total Raw Score:	

Reprinted from Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman Self-Rating Mania Scale. *Biological Psychiatry.* 1997;42(10):948–955. Copyright © 1997, with permission from Elsevier. doi: 10.1016/S0006-3223(96)00548-3.

Trauma and Stressor-Related Disorders

With the reorganization of DSM-5, Trauma and Stressor-Related Disorders were moved out of the Anxiety Disorders section and given their own distinct section—which now focuses on an array of conceptually similar disorders in which exposure to a stressful or traumatic event is a necessary criterion. These include Reactive Attachment Disorder, Disinhibited Social Engagement Disorder, PTSD, Acute Stress Disorder, and Adjustment Disorders. PTSD remains the most commonly studied of these disorders, and a PTSD instrument is described here.

Clinician-Administered PTSD Scale. The Clinician-Administered PTSD Scale (CAPS) was developed in 1990 to operationalize the DSM-III-R diagnosis of PTSD and can easily be modified to obtain a diagnosis by DSM-5 criteria. The interview includes 17 items required to make the diagnosis, covering all four criteria: (1) the event itself, (2) reexperiencing of the event, (3) avoidance, and (4) increased arousal. For each of the 17 items, the interviewer rates both the frequency and the intensity on a five-point scale ranging from 0 to 4. The diagnosis requires evidence of a traumatic event, one symptom of reexperiencing, three of avoidance, and two of arousal (typically, an item is counted if frequency is at least 1 and intensity is at least 2). The items can also be used to generate a total PTSD severity score obtained by summing the frequency and intensity scales for each item. The CAPS also includes several global rating scales for the impact of PTSD symptomatology on social and occupational functioning, for general severity, for recent changes, and for the validity of the patient's report. The CAPS must be administered by a trained clinician and requires 45 to 60 minutes to complete, with follow-up examinations somewhat briefer. It has demonstrated reliability and validity in multiple settings and multiple languages, and performs well in the research setting for diagnosis and severity assessment, but is generally too lengthy for use in clinical practice.

The PTSD Checklist for DSM-5. The PTSD Checklist for DSM-5 (PCL-5) is a 20-item self-report instrument that assesses DSM-5 symptoms of PTSD (Table 7.8–8). While structured clinical interviews such as the SCID or CAPS (described above) are the gold standard for making a formal diagnosis of PTSD, the PCL-5 may be useful in a number of contexts—whether for symptom monitoring, screening, or in making provisional diagnoses. The PCL-5 takes roughly 5 to 10 minutes to complete, and is scored by summing the 20 items. Cut scores between 31 and 33 have been suggested as indicative of “probable PTSD.” Of note, the PCL-5 items are organized according to DSM-5 symptom clusters for PTSD (e.g., items 1 to 5 measure PTSD Cluster B symptoms, items 6 to 7 measure PTSD Cluster C symptoms, etc.). As a result, the developers suggest clinicians may use the PCL-5 to make provisional PTSD diagnoses based on the number of symptoms endorsed within these discrete domains. Finally, multiple versions of the PCL-5 are available, including civilian and military forms, and those that are trauma-specific. The psychometric properties of the PCL-5 have been shown to be robust, which may explain its popularity in contexts where symptom tracking is a priority.

Obsessive Compulsive and Related Disorders

Like the Trauma and Stressor-Related Disorders, Obsessive Compulsive Disorder and Related Disorders were also removed from the Anxiety Disorders section and given their own section in DSM-5. This new family of disorders now includes: Obsessive Compulsive Disorder, Body Dysmorphic Disorder, Hoarding Disorder, Trichotillomania, and Excoriation, among others. However, OCD remains the most common and researched of these disorders, and several rating instruments for OCD are summarized below.

Yale-Brown Obsessive-Compulsive Scale. The YBOCS was developed in the late 1980s to measure the severity of symptoms in OCD. It has 10 items rated based on a semi-structured interview. The first five items concern obsessions: The amount of time that they consume, the degree to which they interfere with normal functioning, the distress that they cause, the patient’s attempts to resist them, and the patient’s ability to control them. The remaining five items ask parallel questions about compulsions. Each item has a set of item-specific anchors scored from 0 to 4, so total scores for obsessions and compulsions each range from 0 to 20, and overall total score ranges from 0 to 40. Typical scores for patients with OCD are in the 16 to 30 range, and a threshold of 16 is typically used for inclusion in drug trials. The semi-structured interview and ratings can be completed in 15 minutes or less. A self-administered version can be completed in 10 to 15 minutes. Computerized and telephone use have also been found to provide acceptable ratings. Before the first use of the YBOCS, an associated 64-item checklist is administered to provide a more detailed assessment of the specific content of the patient’s obsessions and delusions. Reliability studies of the YBOCS show good internal consistency, interrater reliability, and test–retest reliability over a 1-week interval. The YBOCS has become the standard instrument for assessing OCD severity and is used in virtually every drug trial. It may also be used clinically to monitor treatment response.

The Repetitive Thoughts and Behaviors Scale. The Repetitive Thoughts and Behaviors Scale was adapted from the Florida Obsessive Compulsive Inventory (FOCI), published in 1994, and consists of five items rated by patients along item-specific five-point (0 to 4) Likert scales that are summed for a total score ranging from 0 to 20—with higher scores suggestive of more intense repetitive thoughts/behaviors. A cut-score of 8 has been suggested as a threshold for more intensive inquiry into potential obsessions/compulsions, and clinicians have also been encouraged to average the five items as a means of obtaining an overall estimate of severity (where 0 = mild; 1 = moderate; 2 = severe; and 3 = extreme). The Repetitive Thoughts and Behaviors scale has shown good initial reliability and validity, but, as with other DSM-5 emerging measures, future research will be helpful in further elucidating the scale's properties and clinical application.

Substance-Related and Addictive Disorders

DSM-5 now organizes substance-related disorders in a way that combines previous distinctions of "abuse" and "dependence" along a common continuum from mild to severe, and focuses on 10 separate classes of substances as separate use disorders: alcohol; caffeine; cannabis; hallucinogens; inhalants; opioids; sedatives, hypnotics, and anxiolytics; stimulants; tobacco; and other (unknown) substances. These disorders, particularly those involving alcohol, are common and debilitating in the general population, so screening instruments are particularly helpful. Because these behaviors are socially undesirable, underreporting of symptoms is a significant problem; thus, the validity of all substance use measures is limited by the honesty of the patient. Validation against drug tests or other measures is of great value, particularly when working with patients who have known substance abuse.

Table 7.8–8.
The PTSD Checklist for DSM-5 (PCL-5)

PCL-5 with Criterion A

Instructions: This questionnaire asks about problems you may have had after a very stressful experience involving actual or threatened death, serious injury, or sexual violence. It could be something that happened to you directly, something you witnessed, or something you learned happened to a close family member or close friend. Some examples are a serious accident; fire; disaster such as a hurricane, tornado, or earthquake; physical or sexual attack or abuse; war; homicide; or suicide.

First, please answer a few questions about your worst event, which for this questionnaire means the event that currently bothers you the most. This could be one of the examples above or some other very stressful experience. Also, it could be a single event (for example, a car crash) or multiple similar events (for example, multiple stressful events in a war-zone or repeated sexual abuse).

Briefly identify the worst event (if you feel comfortable doing so):

How long ago did it happen? ________________________________ (please estimate if you are not sure)

Did it involve actual or threatened death, serious injury, or sexual violence?
_____ Yes
_____ No

How did you experience it?
_____ It happened to me directly
_____ I witnessed it
_____ I learned about it happening to a close family member or close friend
_____ I was repeatedly exposed to details about it as part of my job (for example, paramedic, police, military, or other first responder)
_____ Other, please describe

If the event involved the death of a close family member or close friend, was it due to some kind of accident or violence, or was it due to natural causes?
_____ Accident or violence
_____ Natural causes
_____ Not applicable (the event did not involve the death of a close family member or close friend)

Second, below is a list of problems that people sometimes have in response to a very stressful experience. Keeping your worst event in mind, please read each problem carefully and then circle one of the numbers to the right to indicate how much you have been bothered by that problem in the past month.

In the past month, how much were you bothered by:	Not at all	A little bit	Moderately	Quite a bit	Extremely
1. Repeated, disturbing, and unwanted memories of the stressful experience?	0	1	2	3	4
2. Repeated, disturbing dreams of the stressful experience?	0	1	2	3	4
3. Suddenly feeling or acting as if the stressful experience were actually happening again (as if you were actually back there reliving it)?	0	1	2	3	4
4. Feeling very upset when something reminded you of the stressful experience?	0	1	2	3	4
5. Having strong physical reactions when something reminded you of the stressful experience (for example, heart pounding, trouble breathing, sweating)?	0	1	2	3	4
6. Avoiding memories, thoughts, or feelings related to the stressful experience?	0	1	2	3	4
7. Avoiding external reminders of the stressful experience (for example, people, places, conversations, activities, objects, or situations)?	0	1	2	3	4
8. Trouble remembering important parts of the stressful experience?	0	1	2	3	4
9. Having strong negative beliefs about yourself, other people, or the world (for example, having thoughts such as: I am bad, there is something seriously wrong with me, no one can be trusted, the world is completely dangerous)?	0	1	2	3	4
10. Blaming yourself or someone else for the stressful experience or what happened after it?	0	1	2	3	4
11. Having strong negative feelings such as fear, horror, anger, guilt, or shame?	0	1	2	3	4
12. Loss of interest in activities that you used to enjoy?	0	1	2	3	4
13. Feeling distant or cut off from other people?	0	1	2	3	4
14. Trouble experiencing positive feelings (for example, being unable to feel happiness or have feelings for people close to you)?	0	1	2	3	4
15. Irritable behavior, angry outbursts, or acting aggressively?	0	1	2	3	4
16. Taking too many risks or doing things that could cause you harm?	0	1	2	3	4
17. Being "superalert" or watchful or on guard?	0	1	2	3	4
18. Feeling jumpy or easily startled?	0	1	2	3	4
19. Having difficulty concentrating?	0	1	2	3	4
20. Trouble falling or staying asleep?	0	1	2	3	4

From Weathers FW, Litz BT, Keane TM, et al. *The PTSD Checklist for DSM-5 (PCL-5)—Extended Criterion A [Measurement instrument]*. National Center for PTSD; 2013. https://www.ptsd.va.gov

The Level 2 Substance Abuse Module from the DSM-5. The Level 2 Substance Abuse Module from the DSM-5 is a modified version of the National Institute on Drug Abuse's ASSIST evaluation system and was developed as a brief screener to augment the clinical interview when evaluating for potential substance misuse (Table 7.8–9). The 10-item measure is completed either by the patient or informant prior to meeting with a clinician, takes roughly 5 to 10 minutes, and asks about how frequently the patient has used various classes of substances in the last two weeks on a five-point Likert scale (0, not at all; 1, 1 or 2 days; 2, several days; 3, more than half the days; 4, nearly every day). Clinicians are advised to interpret each item separately given that each assesses use of a different substance. Although no formal scoring program exists, higher scores are indicative of greater complexity and severity of substance use, and suggest the need for further evaluation and treatment. The APA has recommended this measure in the DSM-5's Emerging Measures section as one method for evaluating and tracking substance misuse over time, although validity and reliability data is somewhat limited given its more recent release.

Table 7.8–9.
The Level 2 Substance Use Module from DSM-5

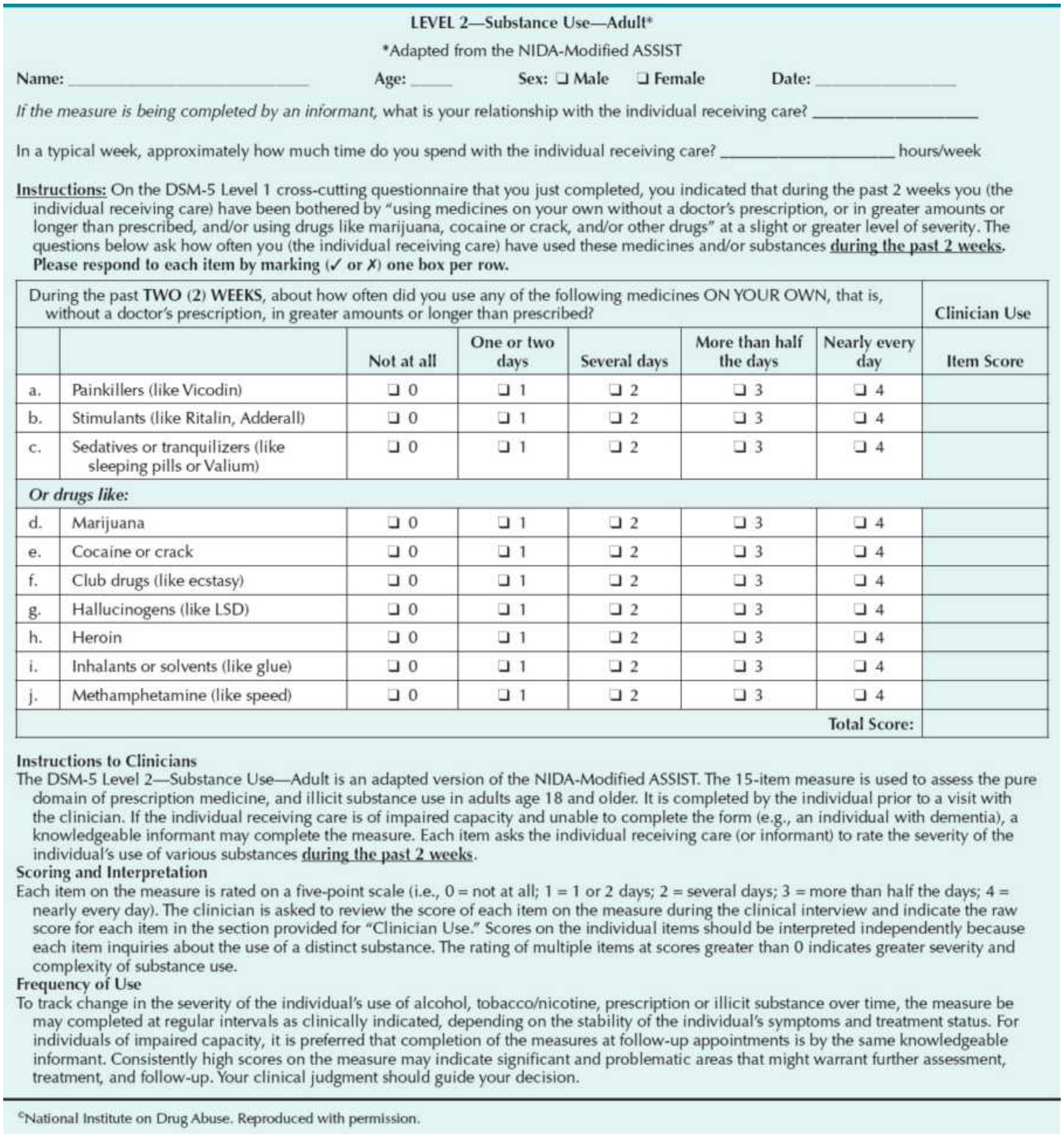

LEVEL 2—Substance Use—Adult*

*Adapted from the NIDA-Modified ASSIST

Name: ____________________ **Age:** _____ **Sex:** ☐ **Male** ☐ **Female** **Date:** ______________

If the measure is being completed by an informant, what is your relationship with the individual receiving care? ______________

In a typical week, approximately how much time do you spend with the individual receiving care? ______________ hours/week

Instructions: On the DSM-5 Level 1 cross-cutting questionnaire that you just completed, you indicated that during the past 2 weeks you (the individual receiving care) have been bothered by "using medicines on your own without a doctor's prescription, or in greater amounts or longer than prescribed, and/or using drugs like marijuana, cocaine or crack, and/or other drugs" at a slight or greater level of severity. The questions below ask how often you (the individual receiving care) have used these medicines and/or substances **during the past 2 weeks**. **Please respond to each item by marking (✓ or ✗) one box per row.**

During the past **TWO (2) WEEKS**, about how often did you use any of the following medicines ON YOUR OWN, that is, without a doctor's prescription, in greater amounts or longer than prescribed?							Clinician Use
		Not at all	One or two days	Several days	More than half the days	Nearly every day	Item Score
a.	Painkillers (like Vicodin)	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	
b.	Stimulants (like Ritalin, Adderall)	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	
c.	Sedatives or tranquilizers (like sleeping pills or Valium)	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	
Or drugs like:							
d.	Marijuana	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	
e.	Cocaine or crack	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	
f.	Club drugs (like ecstasy)	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	
g.	Hallucinogens (like LSD)	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	
h.	Heroin	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	
i.	Inhalants or solvents (like glue)	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	
j.	Methamphetamine (like speed)	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	
						Total Score:	

Instructions to Clinicians

The DSM-5 Level 2—Substance Use—Adult is an adapted version of the NIDA-Modified ASSIST. The 15-item measure is used to assess the pure domain of prescription medicine, and illicit substance use in adults age 18 and older. It is completed by the individual prior to a visit with the clinician. If the individual receiving care is of impaired capacity and unable to complete the form (e.g., an individual with dementia), a knowledgeable informant may complete the measure. Each item asks the individual receiving care (or informant) to rate the severity of the individual's use of various substances **during the past 2 weeks**.

Scoring and Interpretation

Each item on the measure is rated on a five-point scale (i.e., 0 = not at all; 1 = 1 or 2 days; 2 = several days; 3 = more than half the days; 4 = nearly every day). The clinician is asked to review the score of each item on the measure during the clinical interview and indicate the raw score for each item in the section provided for "Clinician Use." Scores on the individual items should be interpreted independently because each item inquiries about the use of a distinct substance. The rating of multiple items at scores greater than 0 indicates greater severity and complexity of substance use.

Frequency of Use

To track change in the severity of the individual's use of alcohol, tobacco/nicotine, prescription or illicit substance over time, the measure be may completed at regular intervals as clinically indicated, depending on the stability of the individual's symptoms and treatment status. For individuals of impaired capacity, it is preferred that completion of the measures at follow-up appointments is by the same knowledgeable informant. Consistently high scores on the measure may indicate significant and problematic areas that might warrant further assessment, treatment, and follow-up. Your clinical judgment should guide your decision.

©National Institute on Drug Abuse. Reproduced with permission.

CAGE. The CAGE was developed in the mid-1970s to serve as a very brief screen for significant alcohol problems in a variety of settings, which could then be followed up by clinical inquiry. CAGE is an acronym for the four questions that comprise the instrument: (1) Have you ever felt you should *c*ut down on your drinking? (2) Have people *a*nnoyed you by criticizing your drinking? (3) Have you ever felt bad or *g*uilty about your drinking? (4) Have you ever had a drink first thing in the morning to steady your nerves or to get rid of a hangover (*e*ye-opener)? Each “yes” answer is scored as 1, and these are summed to generate a total score. Scores of 1 or more warrant follow-up, and scores of 2 or more strongly suggest significant alcohol problems. The instrument can be administered in a minute or less, either orally or on paper. Reliability has not been formally assessed. Validity has been assessed against a clinical diagnosis of alcohol abuse or dependence, and these four questions perform surprisingly well. Using a threshold score of 1, the CAGE achieves excellent sensitivity and fair to good specificity. A threshold of 2 provides still greater specificity but at the cost of a fall in sensitivity. The CAGE performs well as an extremely brief screening instrument for use in primary care or in psychiatric practice focused on problems unrelated to alcohol. However, it has limited ability to pick up early indicators of problem drinking that might be the focus of preventive efforts.

Alcohol Use Disorders Identification Test. The Alcohol Use Disorders Identification Test (AUDIT) was developed by the WHO in the late 1980s as a brief screening instrument designed for the early detection of hazardous (i.e., involving the risk of harm) and harmful (i.e., involving the presence of harm) alcohol use in a variety of settings. It focuses on both the past year and current drinking. It includes a 10-item core screening instrument covering alcohol consumption, drinking behaviors, and alcohol-related problems. Each item is rated using item-specific anchors scored from 0 to 4 and summed for a total score from 0 to 40. The AUDIT can be administered and scored in less than 5 minutes and does not require professional training. The AUDIT also offers a clinical screening procedure involving a physical examination and blood tests, which adds no more than 5 to 10 minutes to a routine medical examination. Reliability of the AUDIT appears good based on internal consistency data. Validity judged against a clinical diagnosis of alcoholism is also good: Using a threshold score of 8 is quite sensitive, at least in men, but somewhat nonspecific; less stringent cutoffs may perform better for women and adolescents. A score of 10 is more specific but at a cost in specificity. Validity also appears good based on correlation with other alcohol self-report measures and risk factors for alcoholism. The AUDIT provides an excellent brief screen for alcohol problems and is particularly good for detecting problem drinking at a fairly mild stage. However, its focus on early detection of hazardous and harmful drinking makes it less suitable to use as a diagnostic instrument.

Drug Abuse Screening Test. The Drug Abuse Screening Test (DAST) was developed in the early 1980s to serve as a screening and assessment instrument for drug abuse. The DAST is an adaptation of the Michigan Alcohol Screening Test (MAST) used to screen for alcoholism. It focuses on lifetime drug use, so it is not designed to measure changes over time. The current version of the DAST has 20 yes–no items, and can be given orally or as a paper-and-pencil questionnaire. An earlier version had 28 items, so the 20-item version is sometimes called the *Brief DAST.* The positive items can be summed to form a 20-point scale. The DAST can be administered and scored in less than 10 minutes. Reliability, based on internal consistency, is very good. Validity, based on ability to detect drug abuse disorder, also appears high, with excellent sensitivity and fairly good specificity using a threshold score of 5. The DAST is useful as a screening tool for drug abuse problems in patients with other mental disorders, particularly alcohol abuse. It also provides an overview of problem severity that may be useful to guide treatment choices. However, because it takes a lifetime approach, it is not useful as a change measure in clinical practice or clinical trials.

Addiction Severity Index. The Addiction Severity Index (ASI) was developed in the early 1980s to serve as a quantitative measure of symptoms and functional impairment due to alcohol or drug disorders. It covers demographics, alcohol use, drug use, psychiatric status, medical status, employment, legal status, and family and social issues. Frequency, duration, and severity are assessed. It has 142 items in varying formats, including yes–no, multiple-choice, and scaled items. It includes both subjective and objective items reported by the patient and observations made by the interviewer. In each area, the ASI yields the rater's global assessment of severity, along with a computer score on a 0 to 1 scale. The 142-item version includes information on the past 30 days and lifetime status, but a shorter version is available for use at follow-up. The instrument is designed for clinician administration but has been used successfully by trained lay raters. Training is recommended, and both manuals and formal training programs are available. A computerized version is also available. The standard ASI requires 45 to 75 minutes to complete, but the follow-up version can be completed in 15 to 20 minutes.

Very good to excellent reliability has been demonstrated for the overall composite score, with somewhat lower reliability for severity ratings in each area. Validity has also been demonstrated based on correlation with other measures and discrimination of patient from nonpatient populations. Normative data are available for a range of populations of alcohol and drug abusers, including alcohol clinic patients, drug abusers, homeless persons, and prisoners. The principal use of the ASI is as an aid to treatment planning and the assessment of treatment outcome in clinical and law enforcement settings. It is relatively time-consuming to administer but performs well for this purpose. It has also been used in clinical research as a sensitive indicator of baseline severity and change over time, which allows comparison between clinical research and clinical practice.

Eating Disorders

Although feeding and eating disorders share a number of psychological and behavioral features, the DSM-5 has expanded the framework for conceptualizing and diagnosing these disorders based on differences in classification, clinical course, and treatment outcome, and as a means of reducing the frequency of the Eating Disorder NOS diagnosis. DSM-5 includes the following: Pica, Rumination Disorder, avoidant/restrictive food intake, disorder, anorexia nervosa, bulimia nervosa, and binge-eating disorder. A wide variety of instruments, particularly self-report scales, are available. Because of the secrecy that may surround dieting, bingeing, purging, and other symptoms, validation against other indicators (e.g., body weight for anorexia or dental examination for bulimia) may be very helpful. Such validation is particularly critical for patients with anorexia, who may lack insight into their difficulties.

Eating Disorders Examination. The Eating Disorders Examination (EDE) was developed in 1987 as the first interviewer-based comprehensive assessment of eating disorders, including diagnosis, severity, and an assessment of subthreshold symptoms, and a 2014 revision (version 17.0) was developed to be compatible with DSM-5. A self-report version (the EDE-Q), as well as an interview for children, has since been developed. The EDE focuses on symptoms during the preceding 4 weeks, although longer-term questions are included to assess diagnostic criteria for eating disorders. Each item on the EDE has a required probe with suggested follow-up questions to judge severity or frequency or both, which are then rated on a seven-point Likert scale. For the self-report version, subjects are asked to make similar ratings of frequency or severity. The instrument provides both global severity ratings and ratings on four subscales: restraint, eating concern, weight concern, and shape concern. The interview, which must be administered by a trained clinician, requires 30 to 60 minutes to complete, whereas the self-report version is somewhat faster. Reliability and validity data for both the EDE and EDE-Q are excellent, although the EDE-Q may have greater sensitivity for binge-eating disorder. The EDE performs well in both the diagnosis and the detailed assessment of eating disorders in the research context. It also has the sensitivity to change required for use in clinical trials or monitoring of individual therapy. Even in the research setting, however, the EDE is fairly lengthy for repeated use, and the EDE-Q may be preferable for some purposes. Although the EDE is too lengthy for routine clinical practice, the EDE or EDE-Q might be helpful in providing a comprehensive assessment of a patient with a suspected eating disorder, particularly during an evaluation visit or on entry into an inpatient facility.

The Eating Disorders Assessment-5 (EDA-5). The Eating Disorders Assessment-5 (EDA-5), developed in 2014, is a semi-structured interview that was designed to assist clinicians in evaluating for DSM-5 Feeding and Eating Disorders in children and adults. The tool is administered by clinicians with expertise in the evaluation of eating disorders; there is also a computerized/internet version. The EDA-5 provides a series of probes to inquire about symptoms corresponding to DSM-5 criteria for each of the Feeding and Eating disorders. Clinicians ask follow-up questions as needed, and rate each criterion using all other available clinical information (e.g., behavioral observations, information from family, medical records, objective measures of height/weight, etc.). The EDA-5 employs an algorithm whereby subsequent questions are administered based on responses to previous questions, and thus the number of items administered varies across patients. The instrument takes roughly 20 to 30 minutes to complete. Although the measure appears promising for DSM-5 feeding and eating disorders, validity and reliability information remains limited given its recent release.

Neurocognitive Disorders

This subsection reviews measures of Major and Mild neurocognitive disorders, which primarily affect the older adults, where they are more commonly known as dementia and MCI; in younger patients, such as those with acquired cognitive disorders resulting from head injury or HIV, the term neurocognitive disorder is often preferred. A wide variety of measures are available, and the focus here on those commonly used among older patients. Most involve cognitive testing and provide objective, quantifiable data. However, scores vary by ability and by educational level in the general population, so these instruments are often more informative when the patient’s own baseline scores are known. Other measures focus on functional status, which can be assessed based on a comparison with a description of the subject’s baseline function; these types of measures generally require a knowledgeable informant and, thus, may be more cumbersome to administer but tend to be less subject to educational biases. A third type of measure focuses on the associated behavioral symptoms that are frequently seen in demented patients. Last, this subsection includes a measure of delirium, an acute or subacute confusional state resulting from a general medical condition.

Mini-Mental State Examination. The MMSE is a 30-point cognitive test developed in the mid-1970s to provide a bedside assessment of a broad array of cognitive function, including orientation, attention, memory, construction, and language. It can be administered in less than 10 minutes by a busy doctor or a technician and scored rapidly by hand. The MMSE has been extensively studied and shows excellent reliability when raters refer to consistent scoring rules. Validity appears good based on correlations with a wide variety of more comprehensive measures of mental functioning and clinicopathologic correlations. Although the MMSE is used commonly to screen for dementia, in both clinical practice and epidemiologic or clinical research, it is not optimal for this purpose, and even worse as a screen for MCI: it has limited memory items and no executive function items and thus tends to miss the areas most likely to be affected early in the course of disease. If the MMSE is used for screening, then a cutoff of 24 for identifying cases of dementia has been suggested, but it is probably more accurate to use age- and education-adjusted norms to interpret the results, and higher cutoffs are better for detecting MCI. For patients with extensive education, who may score 30 out of 30 despite clear evidence of functional decline, a more difficult cognitive test, full neuropsychological battery, or clinical interview about functional change may be required to detect dementia. The other principal use of the MMSE is in following the progression of dementia over time in the clinic or in clinical trials. As a rule of thumb, patients with mild dementia tend to score from 20 to 24, moderate from 11 to 19, and severe from 0 to 10. However, it should be noted that these figures do not take the educational differences noted above into account. In addition, the MMSE does not do as well tracking progression of dementia at later stages; scores of 10 or less are difficult to obtain, as many patients become untestable.

The Montreal Cognitive Assessment. The Montreal Cognitive Assessment (MoCA) is a brief, 30-point cognitive screening measure that was developed in the late 1990s to address many of the shortcomings of the MMSE, and to assist in detecting MCI and early Alzheimer dementia (AD) (Table 7.8–10). The MoCA evaluates a number of cognitive domains including memory (immediate and short delayed recall), attention and working memory, visuospatial abilities, executive functioning, language, and orientation. MoCA scores range from 0 to 30 (with higher scores indicating better functioning), and a cut-score of 26 has been suggested as differentiating normal from abnormal scores. It can be administered by a busy physician or technician in as little as 5 to 10 minutes. However, as of September 2019 the test developer now requires professionals to undergo formal training prior to administering the tool with patients. Initial validation research indicates that sensitivity and specificity of the MoCA in detecting patients with MCI and AD have generally been promising and superior to the MMSE, although research in other settings has been somewhat mixed. As noted above with the MMSE, the MoCA may provide for a useful bedside evaluation and/or briefer assessment of cognitive functioning in settings where more in-depth testing is challenging, although more comprehensive neuropsychological testing may ultimately be necessary in making a differential diagnosis and/or providing a more nuanced picture of someone’s level of functioning. This is particularly true in highly educated populations, where cognitive decline may be more difficult to detect due to ceiling effects in screening measures like the MMSE and MoCA.

Table 7.8–10.
The Montreal Cognitive Assessment (MoCA)

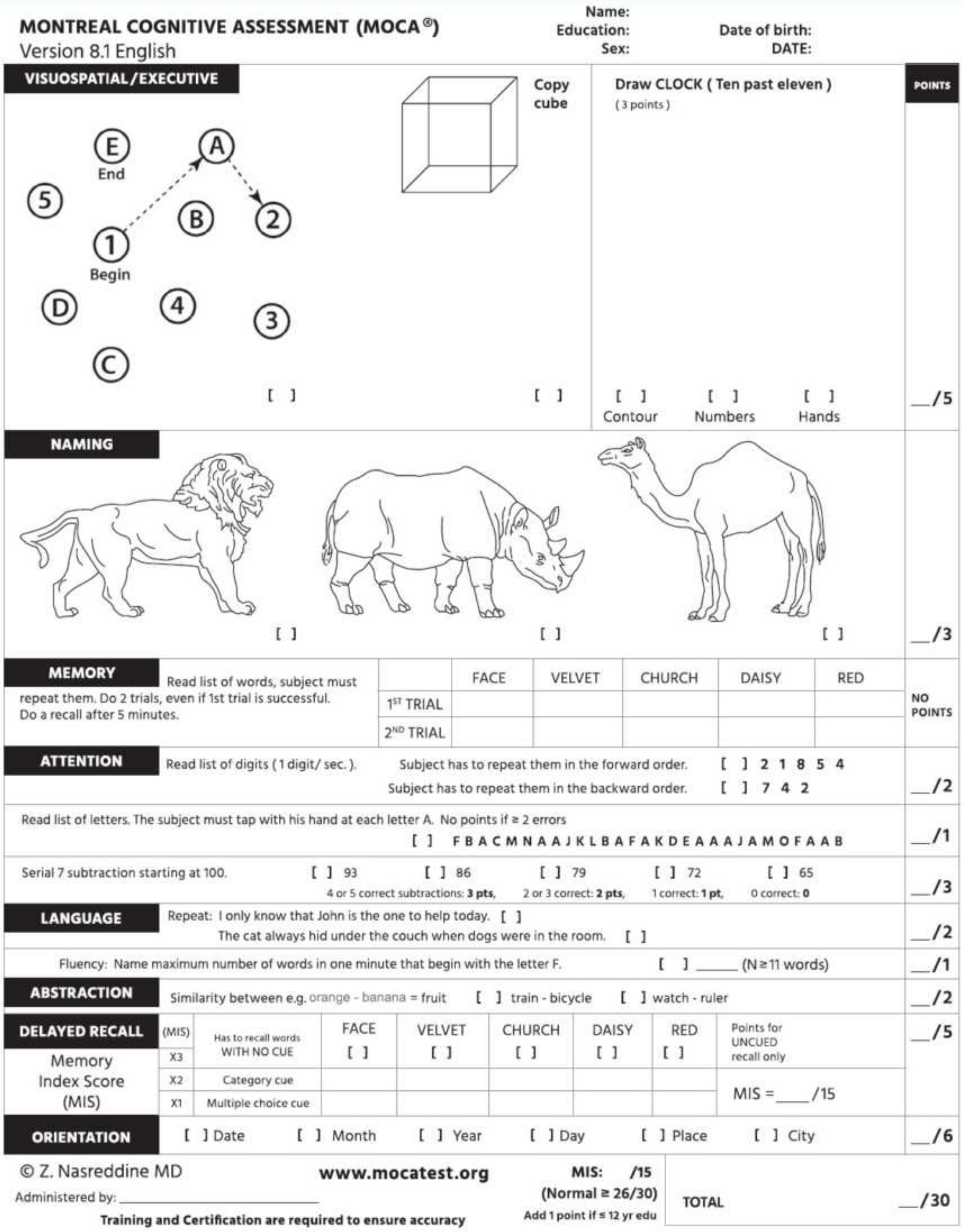
E
End
A
5
B
2
1
Begin
D
4
3
C

MONTREAL COGNITIVE ASSESSMENT (MOCA®)
Version 8.1 English

Name:
Education:
Sex:
Date of birth:
DATE:

VISUOSPATIAL / EXECUTIVE

[ ]

Copy cube [ ]

Draw CLOCK (Ten past eleven) (3 points)
[ ] Contour [ ] Numbers [ ] Hands

POINTS __/5

NAMING

[ ] [ ] [ ] __/3

MEMORY Read list of words, subject must repeat them. Do 2 trials, even if 1st trial is successful. Do a recall after 5 minutes.

	FACE	VELVET	CHURCH	DAISY	RED
1ST TRIAL					
2ND TRIAL					

NO POINTS

ATTENTION Read list of digits (1 digit/ sec.).
Subject has to repeat them in the forward order. [ ] 2 1 8 5 4
Subject has to repeat them in the backward order. [ ] 7 4 2
__/2

Read list of letters. The subject must tap with his hand at each letter A. No points if ≥ 2 errors
[ ] F B A C M N A A J K L B A F A K D E A A A J A M O F A A B
__/1

Serial 7 subtraction starting at 100. [ ] 93 [ ] 86 [ ] 79 [ ] 72 [ ] 65
4 or 5 correct subtractions: **3 pts**, 2 or 3 correct: **2 pts**, 1 correct: **1 pt**, 0 correct: **0**
__/3

LANGUAGE Repeat: I only know that John is the one to help today. [ ]
The cat always hid under the couch when dogs were in the room. [ ]
__/2

Fluency: Name maximum number of words in one minute that begin with the letter F. [ ] _____ (N ≥ 11 words)
__/1

ABSTRACTION Similarity between e.g. orange - banana = fruit [ ] train - bicycle [ ] watch - ruler
__/2

DELAYED RECALL

Memory Index Score (MIS)	(MIS)		FACE	VELVET	CHURCH	DAISY	RED	
	X3	Has to recall words WITH NO CUE	[ ]	[ ]	[ ]	[ ]	[ ]	Points for UNCUED recall only
	X2	Category cue						MIS = ____/15
	X1	Multiple choice cue						

__/5

ORIENTATION [ ] Date [ ] Month [ ] Year [ ] Day [ ] Place [ ] City
__/6

© Z. Nasreddine MD
Administered by: ________________________
Training and Certification are required to ensure accuracy

www.mocatest.org

MIS: /15
(Normal ≥ 26/30)
Add 1 point if ≤ 12 yr edu

TOTAL __/30

Copyright Z. Nasreddine MD. Reproduced with permission. It is mandatory to follow the online MoCA© Training and Certification Program to administer and score the MoCA©. Copies are available at www.mocatest.org.

Clinical Dementia Rating Scale. The Clinical Dementia Rating Scale (CDR) is a standardized staging system for dementia based on all available information about the patient. It was developed in 1992 and in 1993 was revised slightly in the context of a standard assessment protocol. Patients are rated in each of six domains (memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care) on a scale from 0 to 3 (0 = no dementia, 0.5 = questionable dementia [approximately equivalent to MCI], 1 = mild dementia, 2 = moderate dementia, 3 = severe dementia), with detailed descriptions to anchor the ratings in each area. The ratings in each domain are typically combined into an overall CDR rating on the same scale following a weighting system that favors memory. However, they can also be summed to generate a CDR Sum of Boxes score. The staging system can be used on its own or based on the CDR protocol, a semi-structured interview with the patient and an informant, plus brief bedside cognitive testing. Ratings can be made relatively quickly by a clinician who is very familiar with the patient, but the full protocol requires 1 hour or more to administer. CDR ratings of 4 (profound dementia) and 5 (terminal dementia) are based on overall status. Whether based on the CDR protocol or on clinical experience with the patient, the CDR must be rated by a trained clinician. Additional training in the use of the scale and protocol are strongly recommended. The reported reliability of CDR scores based on the full protocol is high, but achieving this level of reliability requires extensive training. Validity of the CDR has also been shown to be high. A somewhat shorter interview and informant questionnaire are also available. Although the CDR *protocol* is clearly intended for research, the CDR *scale* is of potential use as a stage marker for clinical communication and charting decline over time. It is also helpful for understanding the dementia research literature, as the CDR and CDR—Sum of Boxes (SOB) are increasingly used as stage markers for clinical trials and other research. However, the CDR's lack of detail makes it relatively insensitive to change over short periods (the average annual decline is approximately one half-stage per year), and the CDR Sum of Boxes is only modestly better. In addition, the CDR was developed based on the typical course of Alzheimer disease, and tends to perform less well in patients with other dementing illnesses.

Neuropsychiatric Inventory. The Neuropsychiatric Inventory (NPI) was developed in the mid-1990s to assess a wide range of behavioral symptoms that are often seen in Alzheimer disease and other dementing disorders. The current version rates 12 areas: delusions, hallucinations, dysphoria, anxiety, agitation/aggression, euphoria, disinhibition, irritability/lability, apathy, aberrant motor behavior, nocturnal disturbances, appetite, and eating. The standard NPI is an interview with a caregiver or other informant that can be performed by a clinician or trained lay interviewer and requires 15 to 20 minutes to complete. There is also a nursing home interview version, the NPI-NH, and a self-report questionnaire, the NPI-Q. For each area, the NPI asks whether a symptom is present and, if so, assesses frequency, severity, and associated caregiver distress. The instrument has demonstrated reliability and validity and is useful to screen for problem behaviors in both clinical and research settings. Because of the detailed frequency and severity ratings, it is also useful to monitor change with treatment.

Confusion Assessment Method. The Confusion Assessment Method (CAM) was developed in 1990 as a standardized method for identifying delirium. It is typically administered by a clinical interviewer (or lay interviewer with specialized training), and takes roughly 5 to 10 minutes to complete. Full- and short-forms are also available. The CAM short-form specifically assesses for the presence/absence of the following in making a determination about possible delirium: (1) mental status altered from baseline (acute onset or fluctuating); (2) inattention; (3) disorganized thinking; and (4) altered level of consciousness. Delirium is identified when both conditions 1 and 2, and either 3 or 4 (or both) are met. The CAM long form assesses the same four domains above, as well as the following: (5) disorientation; (6) memory impairment; (7) perceptual disturbance; (8) psychomotor agitation and retardation; and (9) altered sleep–wake cycle. There has been a considerable amount of research on the CAM over the last few decades in support of its reliability and validity, and the tool has since been translated into 14 languages.

Personality Disorders and Personality Traits

Personality may be conceptualized categorically as personality disorders or dimensionally as personality traits, which may be viewed as normal or pathologic. The focus here is on personality disorders and the maladaptive traits generally viewed as their milder forms. Although the most recent DSM-5-TR retained the diagnostic system used in DSM-IV for diagnosing personality disorders, an alternative hybrid DSM-5-TR model for personality disorders is also included in Section III. The APA noted that it would include both options to ensure continuity with current and prior clinical and research practice, while also providing a structure for investigating a newer model designed to address many of the shortcomings in the current system. The DSM-5-TR continues to define 10 personality disorders in three clusters. Patients tend not to fall neatly into these personality categories; instead, most patients who meet criteria for one personality disorder also meet criteria for at least one other, particularly within the same cluster. This and other limitations in the validity of the constructs themselves make it difficult to achieve validity in personality measures. Personality measures include both interviews and self-report instruments. Self-report measures are appealing in that they require less time and may appear less threatening to the patient. Because many of the symptoms suggesting personality problems are socially undesirable and because patients' insight tends to be limited, clinician-administered instruments, which allow for probing and patient observation, may provide more accurate data.

Structured Clinical Interview for DSM Personality Disorders. The Structured Clinical Interview for DSM-5 Personality Disorders (SCID-5-PD) is the counterpart of the SCID for making DSM diagnoses of personality disorders. The initial version was developed for the DSM-III-R in the mid-1990s, and the current version makes diagnoses according to the DSM-5. The SCID-5-PD is organized by disorder and includes all ten DSM personality disorders, as well as Other Specified Personality Disorder. Although the DSM-5 retained the same classification system for diagnosing personality disorders as DSM-IV, the SCID-5-PD was updated to more optimally capture these diagnostic criteria and a dimensional scoring approach was also included—in keeping with important advances in field that have called into question categorical approaches to diagnosis. In addition to the primary interview structure, which includes core questions about each of the 10 personality disorders, there is also an optional SCID-5-SPQ that serves as a brief (20-minute) self-report screening tool that can be used to reduce the time of the SCID-5-PD interview. The SCID-5-PD must be administered by doctoral-level clinicians, and training is also required. A computerized administration and scoring program is available. Reliability is good for the presence or absence of any disorder but only fair for specific personality disorders; the reliability of dimensional assessment is somewhat better. Validity is somewhat harder to determine. Agreement with clinician assessment tends to be modest, but given the instrument's comprehensiveness and strict adherence to DSM criteria, the SCID-I5-PD may actually be more valid than clinician ratings. The SCID-II is most useful to provide a standardized, comprehensive assessment of personality disorders, whether in research, forensic, or clinical settings.

Personality Disorder Questionnaire. The Personality Disorder Questionnaire (PDQ) was developed in the late 1980s as a simple self-report questionnaire designed to provide categorical and dimensional assessment of DSM-III-R personality disorders and was subsequently revised for the DSM-IV (and retained by DSM-5) and renamed the PDQ-4. Another alternative version allows for ratings within the last few weeks and is designed to serve as a change measure. The PDQ-4 includes 85 yes–no items, designed primarily to assess the diagnostic criteria for DSM-IV/DSM-5 personality disorders. Within the 85 items, two validity scales are embedded to identify underreporting, lying, and inattention. There is also a brief clinician-administered Clinical Significance Scale to address the impact of any personality disorder identified by the self-report PDQ. The PDQ can provide categorical diagnoses, a scaled score for each, or an overall index of personality disturbance based on the sum of all of the diagnostic criteria. Overall scores range from 0 to 79; patients with personality disorders generally score above 30, psychotherapy outpatients without such disorders tend to score in the 20 to 30 range, and normal controls tend to score below 20. The PDQ can be completed in less than 30 minutes. Computerized administration and scoring are available. Reliability is fair to good for dimensional assessment and quite variable for categorical assessment, with good reliability for obsessive-compulsive and antisocial personality disorders and inadequate reliability for many disorders. Validity judged against semi-structured clinician-administered interviews is also variable. The PDQ, like other self-report instruments, tends to over-diagnose personality disorders, with many false positives and few false negatives. Its brevity, good sensitivity, and poor specificity make it most useful as a screening device, with patients screening positive being followed up with a detailed clinical evaluation or administration of a semi-structured interview.

The Personality Inventory for DSM-5 and Personality Inventory for DSM-5 Brief Form. The Personality Inventory for DSM-5 (PID-5) was developed in 2012 to measure the five higher-order domains (or dimensions) that are included in the alternative Section III of DSM-5, and have been shown to account for a significant amount of the variance observed across the personality disorders: Negative Affect, Detachment, Antagonism, Disinhibition, and Psychoticism. These five domains encompass twenty-five lower order "facets," which themselves represent different features of the higher-order construct. For example, the higher-order dimension of Negative Affect may include lower-order features of anxiety, depressivity, affective instability, and anhedonia. The PID-5 is a 220-item measure that yields scores at both the domain and facet levels, while the Personality Inventory for DSM-5 Brief Form (PID-5-BF) is a briefer 25-item measure designed to yield scores at the domain level only. Initial psychometric testing of these measures has been promising, and in many ways these tools reflect more recent conceptualizations of personality functioning—where the combination and severity of these traits are seen to have more predictive power than a simple criteria-based approach alone. Regardless, these instruments were published only in 2012, so further research is needed to better understand their contribution and clinical utility.

Childhood Disorders

A wide variety of instruments are available to assess mental disorders in children. Despite this rich array of instruments, however, the evaluation of children remains difficult for several reasons. First, the child psychiatric nosology is at an earlier stage of development, and construct validity is often problematic. Multiple changes in diagnostic criteria from DSM-III to DSM-III-R to DSM-IV to DSM-5 complicate the choice of measures. Second, because children change markedly with age, it is virtually impossible to design a measure that covers children of all ages. Last, because children, particularly young children, have limited ability to report their symptoms, other informants are necessary. This often creates problems because there are frequent disagreements among child, parent, and teacher reports of symptoms, and the optimal way to combine information is unclear.

Child Behavior Checklist. The Child Behavior Checklist (CBCL) is a family of self-rated instruments that survey a broad range of difficulties encountered in children from preschool through adolescence. One version of the CBCL, designed for completion by parents of children between 6 and 18 years of age, is shown in Table 7.8–11. Another version is available for parents of children between 1.5 and 5 years of age. The Youth Self-Report is completed by children between 11 and 18 years of age, and the Teacher Report Form is completed by teachers of school-age children. The scale includes not only problem behaviors, but also academic and social strengths. Each version includes approximately 100 items scored on a three-point Likert scale. Scoring can be done by hand or computer, and normative data are available for each of the three subscales: Problem behaviors, academic functioning, and adaptive behaviors. A computerized version is also available. The CBCL does not generate diagnoses but, instead, suggests cutoff scores for problems in the "clinical range." Parent, teacher, and child versions all show high reliability on the problem subscale, but the three informants frequently do not agree with one another. The CBCL may be useful in clinical settings as an adjunct to clinical evaluation, as it provides a good overall view of symptomatology and may also be used to track change over time. It is used frequently for similar purposes in research involving children and, thus, can be compared to clinical experience. The instrument does not, however, provide diagnostic information, and its length limits its efficiency for tracking purposes.

Diagnostic Interview Schedule for Children. The DISC was originally developed in the early 1990s as a fully structured diagnostic interview for making DSM-III diagnoses in children. It has since been revised for the DSM-III-R, DSM-IV, and more recently DSM-5. The current DISC-5 covers a broad range of DSM-5 diagnoses, both current and lifetime. It has nearly 3,000 questions but is structured with a series of stem questions that serve as gateways to each diagnostic area, with the remainder of each section skipped if the subject answers no. Subjects who enter each section have very few skips, so complete diagnostic and symptom scale information can be obtained. Child, parent, and teacher versions are available. Computer programs are available to implement diagnostic criteria and generate severity scales based on each version or to combine parent and child information. A typical DISC interview may take more than

1 hour for a child, plus an additional hour for a parent. However, because of the stem question structure, the actual time varies widely with the number of symptoms endorsed. The DISC was designed for lay interviewers. It is fairly complicated to administer, and formal training programs are highly recommended. Reliability of the DISC is only fair to good and generally better for the combined child and parent interview. Validity judged against a clinical interview by a child psychiatrist is also fair to good—better for some diagnoses and better for the combined interview. The DISC is well tolerated by parents and children and can be used to supplement a clinical interview to ensure comprehensive diagnostic coverage. Because of its inflexibility, some clinicians find it uncomfortable to use, and its length makes it less than optimal for use in clinical practice. However, it is used frequently in a variety of research settings.

Table 7.8–11.
Child Behavior Checklist for Ages 6 to 18

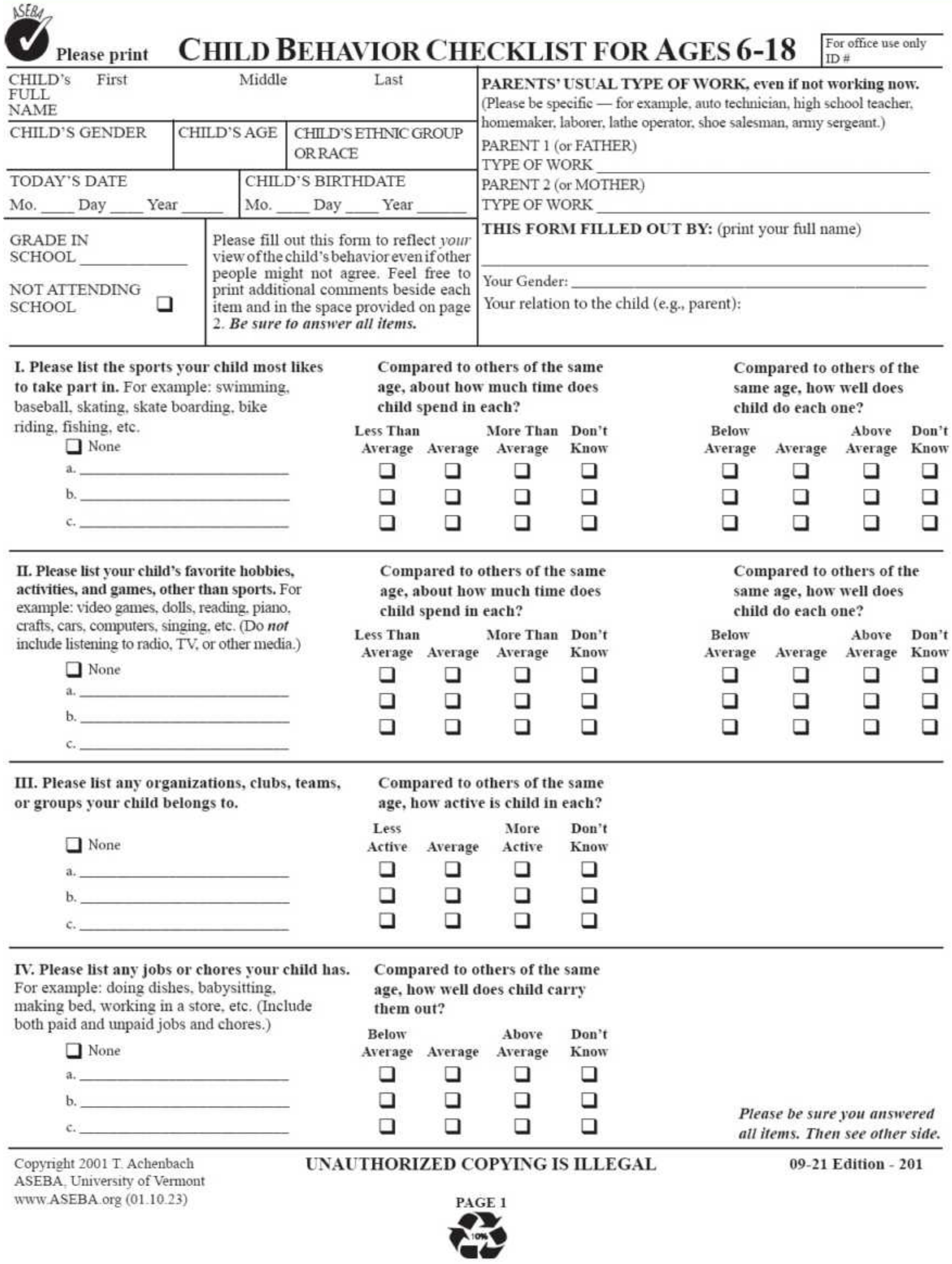

ASEBA

Please print **CHILD BEHAVIOR CHECKLIST FOR AGES 6-18**

For office use only
ID #

CHILD's FULL NAME First Middle Last

CHILD'S GENDER

CHILD'S AGE

CHILD'S ETHNIC GROUP OR RACE

TODAY'S DATE
Mo. ____ Day ____ Year _____

CHILD'S BIRTHDATE
Mo. ____ Day ____ Year _____

GRADE IN SCHOOL ____________

NOT ATTENDING SCHOOL ☐

Please fill out this form to reflect *your* view of the child's behavior even if other people might not agree. Feel free to print additional comments beside each item and in the space provided on page 2. ***Be sure to answer all items.***

PARENTS' USUAL TYPE OF WORK, even if not working now. (Please be specific — for example, auto technician, high school teacher, homemaker, laborer, lathe operator, shoe salesman, army sergeant.)

PARENT 1 (or FATHER)
TYPE OF WORK ________________________________

PARENT 2 (or MOTHER)
TYPE OF WORK ________________________________

THIS FORM FILLED OUT BY: (print your full name)

____

Your Gender: ________________________________

Your relation to the child (e.g., parent):

I. Please list the sports your child most likes to take part in. For example: swimming, baseball, skating, skate boarding, bike riding, fishing, etc.

☐ None

	Compared to others of the same age, about how much time does child spend in each?				**Compared to others of the same age, how well does child do each one?**			
	Less Than Average	**Average**	**More Than Average**	**Don't Know**	**Below Average**	**Average**	**Above Average**	**Don't Know**
a. ____________	☐	☐	☐	☐	☐	☐	☐	☐
b. ____________	☐	☐	☐	☐	☐	☐	☐	☐
c. ____________	☐	☐	☐	☐	☐	☐	☐	☐

II. Please list your child's favorite hobbies, activities, and games, other than sports. For example: video games, dolls, reading, piano, crafts, cars, computers, singing, etc. (Do *not* include listening to radio, TV, or other media.)

☐ None

	Compared to others of the same age, about how much time does child spend in each?				**Compared to others of the same age, how well does child do each one?**			
	Less Than Average	**Average**	**More Than Average**	**Don't Know**	**Below Average**	**Average**	**Above Average**	**Don't Know**
a. ____________	☐	☐	☐	☐	☐	☐	☐	☐
b. ____________	☐	☐	☐	☐	☐	☐	☐	☐
c. ____________	☐	☐	☐	☐	☐	☐	☐	☐

III. Please list any organizations, clubs, teams, or groups your child belongs to.

☐ None

	Compared to others of the same age, how active is child in each?			
	Less Active	**Average**	**More Active**	**Don't Know**
a. ____________	☐	☐	☐	☐
b. ____________	☐	☐	☐	☐
c. ____________	☐	☐	☐	☐

IV. Please list any jobs or chores your child has. For example: doing dishes, babysitting, making bed, working in a store, etc. (Include both paid and unpaid jobs and chores.)

☐ None

	Compared to others of the same age, how well does child carry them out?			
	Below Average	**Average**	**Above Average**	**Don't Know**
a. ____________	☐	☐	☐	☐
b. ____________	☐	☐	☐	☐
c. ____________	☐	☐	☐	☐

Please be sure you answered all items. Then see other side.

Copyright 2001 T. Achenbach
ASEBA, University of Vermont
www.ASEBA.org (01.10.23)

UNAUTHORIZED COPYING IS ILLEGAL

09-21 Edition - 201

PAGE 1

Please print. Be sure to answer all items.

V. 1. About how many close friends does your child have? (Do *not* include brothers & sisters)

☐ None ☐ 1 ☐ 2 or 3 ☐ 4 or more

2. About how many times a week does your child do things with any friends outside of regular school hours? (Do *not* include brothers & sisters)

☐ Less than 1 ☐ 1 or 2 ☐ 3 or more

VI. Compared to others of similar age, how well does your child:

	Worse	Average	Better	
a. Get along with brothers & sisters?	☐	☐	☐	☐ Has no brothers or sisters
b. Get along with other kids?	☐	☐	☐	
c. Behave with parents?	☐	☐	☐	
d. Play and work alone?	☐	☐	☐	

VII. 1. Performance in academic subjects. ☐ **Does not attend school because** ____________________

Check a box for each subject that child takes

Other academic subjects–for example: computer courses, foreign language, business. Do *not* include gym, shop, driver's ed., or other nonacademic subjects.

	Failing	Below Average	Average	Above Average
a. Reading, English, or Language Arts	☐	☐	☐	☐
b. History or Social Studies	☐	☐	☐	☐
c. Arithmetic or Math	☐	☐	☐	☐
d. Science	☐	☐	☐	☐
e. ______________________	☐	☐	☐	☐
f. ______________________	☐	☐	☐	☐
g. ______________________	☐	☐	☐	☐

2. Does your child receive special education or remedial services or attend a special class or special school?

☐ No ☐ Yes — kind of services, class, or school:

3. Has your child repeated any grades? ☐ No ☐ Yes — grades and reasons:

4. Has your child had any academic or other problems in school? ☐ No ☐ Yes — please describe:

When did these problems start?

Have these problems ended? ☐ No ☐ Yes — when?

Does your child have any illness or disability (either physical or mental)? ☐ No ☐ Yes — please describe:

What concerns you most about your child?

Please describe the best things about your child.

Please be sure you answered all items.

Below is a list of items that describe children and youths. For each item that describes your child ***now or within the past 6 months***, please circle the *2* if the item is ***very true or often true*** of your child. Circle the *1* if the item is ***somewhat or sometimes true*** of your child. If the item is ***not true*** of your child, circle the ***0***. Please answer all items as well as you can, even if some do not seem to apply to your child.

0 = Not True (as far as you know) **1 = Somewhat or Sometimes True** **2 = Very True or Often True**

0	1	2	1. Acts too young for age
0	1	2	2. Drinks alcohol without parents' approval (describe): ________________ ________________
0	1	2	3. Argues a lot
0	1	2	4. Fails to finish things they start
0	1	2	5. There is very little they enjoy
0	1	2	6. Bowel movements outside toilet
0	1	2	7. Bragging, boasting
0	1	2	8. Can't concentrate, can't pay attention for long
0	1	2	9. Can't get mind off certain thoughts; obsessions (describe): ________________ ________________
0	1	2	10. Can't sit still, restless, or hyperactive
0	1	2	11. Clings to adults or too dependent
0	1	2	12. Complains of loneliness
0	1	2	13. Confused or seems to be in a fog
0	1	2	14. Cries a lot
0	1	2	15. Cruel to animals
0	1	2	16. Cruelty, bullying, or meanness to others
0	1	2	17. Daydreams or gets lost in thoughts
0	1	2	18. Deliberately harms self or attempts suicide
0	1	2	19. Demands a lot of attention
0	1	2	20. Destroys own things
0	1	2	21. Destroys things belonging to family or others
0	1	2	22. Disobedient at home
0	1	2	23. Disobedient at school
0	1	2	24. Doesn't eat well
0	1	2	25. Doesn't get along with other kids
0	1	2	26. Doesn't seem to feel guilty after misbehaving
0	1	2	27. Easily jealous
0	1	2	28. Breaks rules at home, school, or elsewhere
0	1	2	29. Fears certain animals, situations, or places, other than school (describe): ________________ ________________
0	1	2	30. Fears going to school
0	1	2	31. Fears they might think or do something bad
0	1	2	32. Feels they have to be perfect
0	1	2	33. Feels or complains that no one loves them
0	1	2	34. Feels others are out to get them
0	1	2	35. Feels worthless or inferior
0	1	2	36. Gets hurt a lot, accident-prone
0	1	2	37. Gets in many fights
0	1	2	38. Gets teased a lot
0	1	2	39. Hangs around with others who get in trouble
0	1	2	40. Hears sounds or voices that aren't there (describe): ________________ ________________
0	1	2	41. Impulsive or acts without thinking
0	1	2	42. Would rather be alone than with others
0	1	2	43. Lying or cheating
0	1	2	44. Bites fingernails
0	1	2	45. Nervous, highstrung, or tense
0	1	2	46. Nervous movements or twitching (describe): ________________ ________________
0	1	2	47. Nightmares
0	1	2	48. Not liked by other kids
0	1	2	49. Constipated, doesn't move bowels
0	1	2	50. Too fearful or anxious
0	1	2	51. Feels dizzy or lightheaded
0	1	2	52. Feels too guilty
0	1	2	53. Overeating
0	1	2	54. Overtired without good reason
0	1	2	55. Overweight
			56. Physical problems ***without known medical cause:***
0	1	2	a. Aches or pains (***not*** stomach or headaches)
0	1	2	b. Headaches
0	1	2	c. Nausea, feels sick
0	1	2	d. Problems with eyes (***not*** if corrected by glasses) (describe): ________________
0	1	2	e. Rashes or other skin problems
0	1	2	f. Stomachaches
0	1	2	g. Vomiting, throwing up
0	1	2	h. Other (describe): ________________ ________________

Please be sure you answered all items. Then see other side.

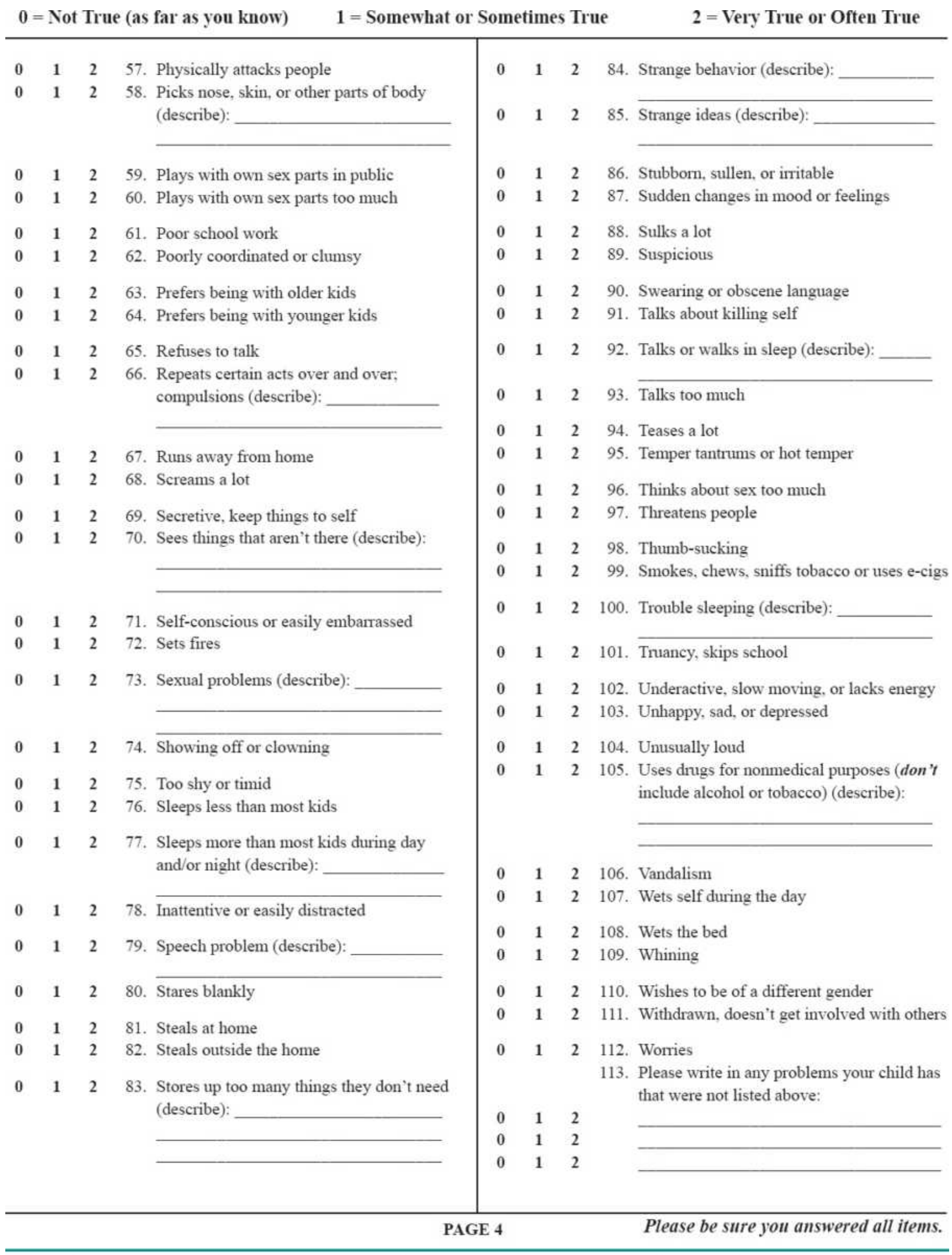

0 = Not True (as far as you know) 1 = Somewhat or Sometimes True 2 = Very True or Often True

0 1 2 57. Physically attacks people
0 1 2 58. Picks nose, skin, or other parts of body (describe): ____________________

0 1 2 59. Plays with own sex parts in public
0 1 2 60. Plays with own sex parts too much

0 1 2 61. Poor school work
0 1 2 62. Poorly coordinated or clumsy

0 1 2 63. Prefers being with older kids
0 1 2 64. Prefers being with younger kids

0 1 2 65. Refuses to talk
0 1 2 66. Repeats certain acts over and over; compulsions (describe): ____________________

0 1 2 67. Runs away from home
0 1 2 68. Screams a lot

0 1 2 69. Secretive, keep things to self
0 1 2 70. Sees things that aren't there (describe): ____________________

0 1 2 71. Self-conscious or easily embarrassed
0 1 2 72. Sets fires

0 1 2 73. Sexual problems (describe): ____________________

0 1 2 74. Showing off or clowning

0 1 2 75. Too shy or timid
0 1 2 76. Sleeps less than most kids

0 1 2 77. Sleeps more than most kids during day and/or night (describe): ____________________

0 1 2 78. Inattentive or easily distracted

0 1 2 79. Speech problem (describe): ____________________

0 1 2 80. Stares blankly

0 1 2 81. Steals at home
0 1 2 82. Steals outside the home

0 1 2 83. Stores up too many things they don't need (describe): ____________________

0 1 2 84. Strange behavior (describe): ____________________

0 1 2 85. Strange ideas (describe): ____________________

0 1 2 86. Stubborn, sullen, or irritable
0 1 2 87. Sudden changes in mood or feelings

0 1 2 88. Sulks a lot
0 1 2 89. Suspicious

0 1 2 90. Swearing or obscene language
0 1 2 91. Talks about killing self

0 1 2 92. Talks or walks in sleep (describe): ____________________

0 1 2 93. Talks too much

0 1 2 94. Teases a lot
0 1 2 95. Temper tantrums or hot temper

0 1 2 96. Thinks about sex too much
0 1 2 97. Threatens people

0 1 2 98. Thumb-sucking
0 1 2 99. Smokes, chews, sniffs tobacco or uses e-cigs

0 1 2 100. Trouble sleeping (describe): ____________________

0 1 2 101. Truancy, skips school

0 1 2 102. Underactive, slow moving, or lacks energy
0 1 2 103. Unhappy, sad, or depressed

0 1 2 104. Unusually loud
0 1 2 105. Uses drugs for nonmedical purposes (*don't* include alcohol or tobacco) (describe): ____________________

0 1 2 106. Vandalism
0 1 2 107. Wets self during the day

0 1 2 108. Wets the bed
0 1 2 109. Whining

0 1 2 110. Wishes to be of a different gender
0 1 2 111. Withdrawn, doesn't get involved with others

0 1 2 112. Worries
113. Please write in any problems your child has that were not listed above:
0 1 2 ____________________
0 1 2 ____________________
0 1 2 ____________________

PAGE 4

Please be sure you answered all items.

From Achenbach TM, University of Vermont, 1 S. Prospect St., Burlington, VT 05401. © 2001 by T. M. Achenbach, reproduced with permission. This sample copy is not to be reproduced.

Conners Rating Scales. The Conners Rating Scales are a family of instruments designed to measure a range of childhood and adolescent psychopathology but are most commonly used in the assessment of ADHD. There are teacher, parent, and self-report (for adolescents) versions and both short (as few as 10 items) and long (as many as 80 items, with multiple subscales) forms. Extensive normative data drawn from an ethnically diverse population are available for each sex across a broad age range. Even the longer forms can be completed in 15 to 20 minutes, and scoring can be accomplished rapidly. Training for raters is not required. Reliability data are excellent for the Conners Rating Scales. However, teacher and parent versions tend to show poor agreement. Validity data suggest that the Conners Rating Scales are excellent at discriminating between ADHD patients and normal controls. It has more difficulty separating ADHD from other disruptive behavioral disorders, such as conduct disorder, but this may be partly related to the genuine clinical difficulties separating these syndromes. Newer versions of the Conners Rating Scales have been developed that aim to improve these discriminations but have not yet been subjected to extensive testing. The principal uses of the Conners Rating Scales are in screening for ADHD in school or clinic populations and following changes in symptom severity over time; sensitivity to change in response to specific therapies has been demonstrated for most versions of the Conners Rating Scales.

Autism Diagnostic Interview—Revised. The Autism Diagnostic Interview (ADI) was developed in 1989 as a clinical assessment of autism and related disorders. The revised ADI-R was developed in 2003 with an aim to providing a shorter instrument with better ability to discriminate autism from other developmental disorders. The instrument has 93 items, is designed for individuals with a mental age greater than 18 months, and covers 3 broad areas, consistent with the DSM-5-TR criteria for autism: language and communication; reciprocal social interactions; and restricted, repetitive, and stereotyped behaviors and interests. There are three versions: one for lifetime diagnosis, one for current diagnosis, and one for patients under 4 focused on considering an initial diagnosis of autism. It must be administered by a clinician trained in its use and takes about 90 minutes to complete. When clinicians are properly trained, it has good to excellent reliability and validity but performs poorly in the setting of severe developmental disabilities. It is generally intended for the research setting, when a thorough assessment of autism is required but may have use in clinical practice as well.

FURTHER READINGS

Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman self-rating mania scale. *Biol Psychiatry*. 1997;42:948–955.

Ben-Porath Y, Tellegen A. *MMPI-3: Manual for Administration, Scoring, and Interpretation*. Pearson; 2020.

Blais MA, Sinclair SJ. *Introduction to the Spectra: Indices of Psychopathology: An Assessment Inventory Aligned With the Hierarchical-Dimensional Model of Psychopathology*. Psychological Assessment Resources (PAR); 2020.

Blais MA, Sinclair SJ. *Spectra: Indices of Psychopathology*. Psychological Assessment Resources (PAR); 2018.

Blais MA, Sinclair SJ, Richardson LA, Massey C, Stein MB. External correlates of the SPECTRA: Indices of psychopathology (SPECTRA) in a clinical sample. *Clin Psychol Psychother*. 2021;28(4):929–938.

Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The Posttraumatic Stress Disorder Checklist for *DSM-5* (PCL-5): Development and initial psychometric evaluation. *J Trauma Stress*. 2015;28:489–498.

Buka SL, Monuteaux M, Earls F. Epidemiology of child and adolescent psychiatric disorders. In: Tsuang MT, Tohen M, eds. *Textbook in Psychiatric Epidemiology*. 2nd ed. Wiley-Liss; 2002.

Burt T, Sederer L, Isgak WW, eds. *Outcome Measurement in Psychiatry: A Critical Review*. American Psychiatric Press; 2002.

Butcher JN, Dahlstrom WG, Graham JR, Tellegen A, Kaemmer B. *MMPI-2: Minnesota Multiphasic Personality Inventory–2: Manual for administration and scoring*. University of Minnesota Press; 1989.

Carmines E, Zeller R. *Reliability and Validity Assessment*. Sage Publications; 1979.

Hulley SB, Cummings SR. Planning the measurements: precision and accuracy. In: Hulley SB, Cummings SR, eds. *Designing Clinical Research: An Epidemiologic Approach*. Williams and Wilkins; 1987.

Kendell R, Jablensky A. Distinguishing between the validity and utility of psychiatric diagnoses. *Am J Psychiatry*. 2003;160:4–12.

Kessler RC. The categorical versus dimensional assessment controversy in the sociology of mental illness. *J Health Soc Behav*. 2002;43:171–188.

Krueger RF. The structure of common mental disorders. *Arch Gen Psychiatry*. 1999;56:921–926.

Krueger RF, Derringer J, Markon KE, Watson D, Skodol AE. Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychol Med*. 2012;42:1879–1890.

McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires*. Oxford University Press; 2006.

Morey LC. *Personality Assessment Inventory: Professional Manual*. 2nd ed. Psychological Assessment Resources; 2007.

Murphy J. Symptom scales and diagnostic schedules in adult psychiatry. In: Tsuang MT, Tohen M, eds. *Textbook in Psychiatric Epidemiology*. 2nd ed. Wiley-Liss; 2002.

Murphy JM, Berwick DM, Weinstein MC, Borus JF, Budman SH. Performance of screening and diagnostic tests: application of receiver operating characteristic (ROC) analysis. *Arch Gen Psychiatry*. 1987;44:550–555.

Myers K, Winters NC. Ten-year review of rating scales. I: Overview of scale functioning, psychometric properties, and selection. *J Am Acad Child Adolesc Psychiatry*. 2002;41:114–122.

Okereke OI, Copeland M, Hyman BT, Wanggaard T, Albert MS, Blacker D. The Structured Interview & Scoring Tool-Massachusetts Alzheimer's Disease Research Center (SIST-M): development, reliability, and cross-sectional validation of a brief structured clinical dementia rating interview. *Arch Neurol*. 2011;68(3):343–350.

Okereke OI, Pantoja-Galicia N, Copeland M, et al. The SIST-M: predictive validity of a brief structured clinical dementia rating interview. *Alzheimer Dis Assoc Disord*. 2012;26(3):225–31.

Regier DA. Dimensional approaches to psychiatric classification: refining the research agenda for DSM-V: an introduction. *Int J Methods Psychiatr Res*. 2007;16(Suppl 1):S1–S5.

Regier DA, Kaelber CT, Rae DS, et al. Limitations of diagnostic criteria and assessment instruments for mental disorders: implications for research and policy. *Arch Gen Psychiatry*. 1998;55:109–115.

Robins LN. Epidemiology: reflections on testing the validity of psychiatric interviews. *Arch Gen Psychiatry*. 1985;42:918–924.

Robins E, Guze SB. Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *Am J Psychiatry*. 1990;126:983–987.

Rush J, First MB, Blacker D, eds. *Handbook of Psychiatric Measures*. 2nd ed. American Psychiatric Press; 2007.

Schutte NS, Malouff JM. *Sourcebook of Adult Assessment Strategies*. Plenum Press; 1995.

Sheehan DV, Lecrubier Y, Harnett-Sheehan K, et al. The Mini International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview. *J Clin Psychiatry*. 1998;59(suppl 20):22–33.

Shrout PE, Spitzer RL, Fleiss JL. Quantification of agreement in psychiatric diagnosis revisited. *Arch Gen Psychiatry*. 1998;744:172–177.

Sinclair SJ, McRitchie A, DeFilippo S, et al. The spectra: Indices of psychopathology: construct validity and sensitivity to change in an inpatient psychiatric sample. *J Psychopathol Behav Assess*. 2022;44(1):286–296.

Spitzer RL. Psychiatric diagnosis: are clinicians still necessary? *Compr Psychiatry*. 1983;24:399–411.

Spitzer RL, Fleiss JL. A re-analysis of the reliability of psychiatric diagnosis. *Br J Psychiatry*. 1974;125:341–347.

Sudman S, Bradburn NM, Schwarz N. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass; 1996.

Winokur G, Zimmerman M, Cadoret R. Cause the Bible tells me so. *Arch Gen Psychiatry*. 1988;45:683–684.

Zachar P, Kendler KS. Psychiatric disorders: a conceptual taxonomy. *Am J Psychiatry*. 2007;164:557–565.

Zarin DA, Earls F. Diagnostic decision making in psychiatry. *Am J Psychiatry*. 1993;150:197–206.

7.9 Electronic Media in Psychiatry

ZEBULON TAINTOR, M.D.

This section is shaped by two forces: technology and the coronavirus (COVID-19) pandemic. Technology has developed in many directions and penetrated almost everywhere, offering diverse, unexpected choices. Concurrently, the pandemic has limited our choices, forcing us to rely on technology, ranging from mRNA to communication and its pitfalls. This subsection considers the use of remote communications with patients and among people and developments in health information technology because of the applications of microprocessors in both areas.

DEFINITIONS AND LEGISLATIVE HISTORY

The prefix "*tele*" denoting transmission over a distance, is used to form *telegraph*, involving the transmission of text, now involving digital code, telephone, which adds sounds, and television, which adds images. *Telemedicine* and *telepsychiatry* are, respectively, the practice of medicine and the medical specialty of psychiatry using television. The prefix "*video*" denotes images, increasingly transmitted as still or moving. *Virtual reality* surrounds a person (using wraparound goggles or a helmet) with a simulated environment. *Holographic projections* intrude three-dimensional images into the real environment and, like virtual reality, can be used in therapy. *E-therapy* is therapy done by electronic communication, often with attachments. Apps are computer programs that perform specific tasks (as contrasted with operating systems in which apps are housed). Most are developed for electronic devices, increasingly connected handheld computers (*smartphones*) that can handle text, sound, and video. *Interactive* apps use two-way communication between user and program. *Online* means connected to the World Wide Web (internet), providing immediate availability for a response, while *asynchronous* connections involve delays between responses. *Unmet online therapy* denotes therapy in which therapist and patient have not met. *Face-to-face* (often abbreviated to *F2F*) therapy involves meeting in person, not necessarily touching, but being able to do so without electronics involved. *Hybrid therapy* combines any of the above, usually with F2F and something else.

Other definitions have emerged with electronic storage and retrieval of medical records. The *Health Information Portability and Accountability Act (HIPAA)*, amended several times since its passage in 1996, defines most terms for electronic records in health care. *Protected Health Information (PHI)* is any health information linked to a specific individual. Privacy rights guarantee that PHI is under a person's control and may not be released without patient consent except as required by law, for example, suspected child abuse. The Center for Medical Services (CMS) (update July 2, 2021) defines HIPAA-*covered entities* to include health plans, clearinghouses, and certain health care providers as follows: *Health Plans:* health insurance companies, HMOs, or health maintenance organizations, employer-sponsored health plans, Government programs that pay for health care, like Medicare, Medicaid, and military and veterans' health programs; *Clearinghouses:* organizations that process nonstandard health information to conform to standards for data content or format, or vice versa, on behalf of other organizations; *Providers:* those who submit HIPAA transactions, like claims, **electronically** are covered, including, but not limited to doctors, clinics, psychologists, dentists, chiropractors, nursing homes, and pharmacies. *Business Associates* are engaged by a covered entity to help carry out its health care activities and functions. There must be a written business associate contract or another arrangement that establishes specifically what the business associate has been engaged to do and requires the business associate to comply with HIPAA.

Examples of Business Associates include the following:

- A third-party administrator that assists a health plan with claims processing
- A consultant that performs utilization reviews for a hospital
- A health care clearinghouse that translates a claim from a nonstandard format into a standard transaction on behalf of a health care provider and forwards the processed transaction to a payer
- An independent medical transcriptionist that provides transcription services to a physician
- Also, a covered health care provider, health plan, or health care clearinghouse can be a business associate of another covered entity

Insurance reimbursement for services provided obtains patient consent for sharing PHI, shared as needed among covered entities and business associates.

Entities not directly linked to reimbursement (e.g., a *Regional Health Information Organization [RHIO]* set up to promote communication among clinicians) require consent according to state laws governing privacy that give patients the right to dictate with whom or what their PHI may be shared.

Although the American Psychiatric Association and other privacy advocates press for *granularity* (enabling the item-by-item retention or release of any item of PHI to any specific person or agency, e.g., deciding one hospital but not another in an RHIO has access to one's PHI, or that pregnancy data be withheld), consent often is a blanket opt in or opt out.

Self-pay patients have complete control over their PHI. HIPAA provides blanket privacy protection for *psychotherapy notes.* The PHI of individuals who receive drug and alcohol abuse treatment in federally funded programs is subject to additional privacy protections under 42 USC § 290dd-2 and 42 CFR § 2.11 (Part 2). HIPAA does not state who owns PHI, only that it must be protected. Each covered entity must have a privacy officer, have written privacy policies and practices available on request, and keep track of PHI disclosures. Individuals may complain to clearinghouses: organizations of the Department of Health and Human Services (HHS) of suspected privacy violations.

The HITECH (Health Information Technology for Economic and Clinical Health) Act of 2009 was an economic stimulus for electronic health records (EHRs) (only 10% of hospitals had them), but it strengthened and sharpened HIPAA. The HITECH Act contains four subtitles (A to D). Subtitle A concerns the promotion of health information technology and is split into two parts. Part 1 is concerned with improving health care quality, safety, and efficiency. Part 2 is concerned with applying and using health information technology standards and reports. Subtitle B covers testing of health information technology, Subtitle C covers grants and loans funding, and Subtitle D covers privacy and security of electronic health information. Subtitle D is also split into two parts. Part 1 is concerned with improving privacy and security of Health information technology (HIT) and PHI, and part 2 covers the relationship between the HITECH Act and other laws. Compliance with the requirements of the HITECH Act became enforceable on November 30, 2009, 12 months following the Act being signed into law. The HITECH requirements were incorporated into HIPAA in the Final Omnibus Rule, which brought HIPAA and HITECH together into the same legislation. The HIPAA Omnibus Final Rule was published on January 25, 2013, and had a compliance date of September 23, 2013.

The HHS was given over $25 billion to achieve its goals. The HHS used some of that budget to fund the *Meaningful Use* program, incentivizing care providers to adopt certified EHRs by offering monetary incentives. Certified EHRs have been certified as meeting defined standards by an authorized testing and certification body. Certified EHRs had to be used in a meaningful way, such as for issuing electronic prescriptions and exchanging electronic health information to improve the quality of care. The financial incentives were significant and increased with each year of the program, and new requirements were introduced at each of the three stages of the Meaningful Use program. The failure to meet the requirements of each stage resulted in a financial penalty (reduction of reimbursements for Medicare and Medicaid). In order to qualify for federal funds, care providers not only had to adopt EHRs but also demonstrate meaningful use of certified EHRs. They had to demonstrate that they had achieved the minimum core objectives in each stage in addition to a set number of menu objectives. It was also necessary to demonstrate compliance with the HIPAA Security and Privacy Rules by conducting risk assessments. Loss of protection for PHI is termed a *data breach*, **all of which** now must be reported to HHS. All reported breaches are investigated, and fines may be imposed from $100 to $50,000/violation (maximum $1.5 million/year). Breaches involving more than 500 people must be reported as soon as known, with penalties of more than a 60-day delay. These are listed (https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf), which showed 842 between October 17, 2019, and October 5, 2021.

The movement to EHRs resulted in the marketing of many systems, which differ in the type of data collected and the software design. Many were clinical add-ons to billing/administrative systems (*Practice Management Software [PMS]*). Efforts at quality control resulted in the Certification Commission for Health Information Technology (CCHIT), founded in 2004. A determinedly collegial process involving many stakeholders in many committees, workgroups set standards for privacy and compliance, security networking (connecting electronic devices within an organization), interoperability (enabling uploading PHI to, and use via, other electronic devices), and the *Continuity of Care Health Record (CCHR)*, what one physician can expect another to tell about a patient. It agreed on sets of items required in various general and specialty-specific EHRs. For several years, CCHIT was to EHRs what the Joint Commission on the Accreditation of Health Care Organizations (JCAHO) is to hospitals and clinics. However, protests about the fees charged for what was seen as too lugubrious a process for certifications resulted in looser standards and competing certifying organizations. The Office of the National Coordinator authorized six (now four) authorized and testing bodies for EHRs and EHR modules (https://www.healthit.gov/topic/certification-ehrs/certification-process), currently 915, mostly modules (https://chpl.healthit.gov/#/search). Users bewail the apparent lack of usability as a certification criterion, despite lip service to the same.

Under Administrative Simplification (a title seen by many as ironic), HHS set documentation standards and rules for Privacy Security, Transactions, and Code Sets (with CPT codes set by the American Medical Association). The CMS, part of HHS, sets documentation standards for Medicare and, somewhat, for Medicaid. As these were seen as burdensome, the Office of the National Coordinator for Health Information Technology (ONCHIT) proclaimed in 2017 a priority of “Patients Over Paperwork.”

HIPAA and HITECH strengthened patient rights to their medical records. Some EHRs spin-off *Personal Health Records (PHRs)* for their patients. Others allow access to lab data, prescription records, visit summaries, etc. The movement is to share more data with patients, especially as they pick up errors. On April 5, 2021, federal rules implemented the bipartisan 21st Century Cures Act Open Notes, specifying that eight types of clinical notes are among electronic information that must not be blocked and must be made available free of charge to patients. To meet the interests of some patients, the rules allow specified exceptions (www.opennotes.org/onc-federal-rule). While simply opening the entire EHR is temptingly little work, some problems arise, such as reading a pathology report of metastatic cancer before one's physician has seen it or feeling hurt or insulted (https://www.nytimes.com/2021/09/30/well/live/medical-records-misunderstanding.html), so extra work is needed to screen the notes patients may see.

COMMUNICATION EVOLUTION

Written Words: From Letter Writing to Instant Messaging

An email basically is a letter, the most ancient form of communication across distance. People use letters to develop and give a particular form to a relationship. Various published correspondences often include the writers' agreement that the letters were more critical than actual meetings. Many examples show the process of psychotherapy in traditional letter writing. The first "email" device was the telegraph, first used between cities in 1844. Telegraphic style is usually preferred over formalities in email and texting, especially for asynchronous communication of attachments such as test results or directions. As cost is insignificant, use has increased exponentially, as recommended in 2001 by the Institute of Medicine's Committee on Quality of Health Care in America: Crossing the Quality Chasm: A New Health Care System for the 21st Century. Almost all email is retrievable and discoverable in litigation; clinicians should anticipate that emails may be regarded as part of the patient record.

Websites offering information, especially with a marketing component, often encourage emails. In addition, professional and self-help organizations have developed apps for patients to use as needed, for example, coping with anxiety, suicidal thoughts, urges to drink, or detecting depression.

Spoken Words: Telephone

Telephone calls have been used to help people feel better since the invention of the telephone in 1885. They are immediate (if answered, which is less frequent as Robo and other unwanted calls increase) and can convey nonverbal data through tone of voice, pauses, speed and rhythm of speech, and inflection. When costs were charged by time and distance, calls were brief, but now unlimited calling easily accommodates therapy sessions. In addition, while analog calls were not retrievable unless recorded, digital calls can be retrieved and recreated, although with varying degrees of difficulty.

Crisis hotlines are widely used for emergencies and are accepted as a necessary component of full service. There is extensive experience with telephone therapy in suicide prevention centers and other hotlines. Hotlines are growing to handle specific problems, such as urges to be violent. Answerers are always available. Caller and answerer have a cold start, not having talked previously or knowing anything about one another. Listening to call recordings leaves little doubt that the services provide meaningful help. However, suicide rates are not lowered significantly. It is assumed that the most seriously suicidal kill themselves rather than call. In general, callers use hotlines once or twice and are encouraged to go to services where they are seen and helped but use crisis hotlines and emergency rooms for months or years. Usually, the ongoing therapy is recognized, and a treatment plan is generated.

The use of the telephone to transmit images became widely available in 1964 to 1966 with the advent of Xerox's facsimile machines, which added images to long-distance relationships. Videophones initially were unpopular; the video function is now part of every smartphone but infrequently used by those concerned about how they appear. Smartphone video showing what someone else is doing is very frequent, while showing oneself is much less so, especially in conversation. People in the same room who could talk directly to each other sometimes text back and forth.

Spoken Word Recording

Both patients and therapists have recorded psychotherapy sessions since audiotape became available in the 1940s, especially as recording device storage and playback became more accessible and less expensive. State laws vary as to what permission must be obtained from the other party(ies) in a conversation before recording by telephone or in person, but therapists typically recognize that trust is paramount and get specific permission. Both patients and therapists use recordings between sessions to enhance efficacy in focusing on issues, interpretations, and what is happening otherwise.

Images: Television

Although television was invented before World War II, its use in therapy awaited recording media, initially videotape. Milton Berger demonstrated the use of videotaping group and other therapy sessions to help patients see themselves as others saw them, particularly how their body language conveyed unintended messages. As with audio recordings, a chance to step back into an observer role offers fresh insights, enhanced with facial expressions and body language. Televised sessions with patients are commonly used in teaching, available as DVDs, flash memory sticks, and storage on the internet.

Smartphones

The development of the smartphone, starting in 2000 as personal digital assistants (PDAs) added cellular telephony, combining mobile telephone and computing functions into one unit. The smartphone is truly revolutionary in many ways, potentially bringing every smartphone user in the world into contact with the internet and each other. Its many innovations included the primacy of a person’s finger as the input device. Use has extended to more than 3 billion individual users worldwide, likely growing several hundred million more in the next few years. All components have been improved, and the 5G network promises faster speeds. Smartphones have been likened to prosthetic brains with enhanced memories, as data can be retrieved easily from multiple sources. They enable a wide array of activities, almost immediately and readily available.

Accidents have increased as people on their phones do not see where they are going. Handheld phones are prohibited for drivers, but walking accidents have increased. Beneficial and detrimental effects on mental health have been reported, with excessive use seen as dependency and addiction, despite the lack of diagnostic criteria distinguishing between use and abuse, as is also the case for internet use (see below). Young people are seen as particularly vulnerable, and guidelines for use by children and families have been developed, including those from the American Academy of Pediatrics. Concerns about the device itself have centered on screen time, microwave proximity and duration, and sleep. As time spent looking at and manipulating a screen is time that could be spent otherwise, consequences stem from whatever activity is missed, such as work, education, or face-to-face socialization. Sleep consequences ensue from staying up with backlit blue light devices, somewhat ameliorated by programs that change screen color to warmer hues. Possible effects of microwave radiation, other than the 2011 report of increased activity in brain regions near smartphones, have not been shown, although investigations continue.

"There's an app for that" hints at the many millions of apps that have been developed. Notable worldwide, as of August 2021, are Instagram (photo sharing) with more than a billion (500+ million daily) users, WhatsApp (messaging, including video calls) with more than 2 billion monthly users generating more than 100 billion messages daily, and their owner, Facebook (2.85 billion monthly users). Tracking the number of downloads from app stores (e.g., Instagram, more than 6 billion) no longer matters as smartphones are increasingly sold with these apps already installed (e.g., almost all smartphones in Brazil). YouTube (videos, mostly noninteractive) is owned by Google and has about 2 billion monthly users. Specific apps (e.g., detecting depression by sensing cell phone movement) have small user bases and struggle to become known. Categories of mental health apps include programs for learning coping skills, therapy, stress relief, meditation, fun, depression, Black, indigenous [and] people of color (BIPOC), bipolar disorder, symptom tracking, attention deficit hyperactivity disorder (ADHD), and post-traumatic stress disorder (PTSD). Many apps have been developed by treatment systems, such as the Veterans Administration in the United States (https://www.va.gov/WHOLEHEALTH/veteran-resources/MobileApps-OnlineTools.asp), which also offers online tools such as podcasts, and the National Health System in the United Kingdom (https://www.nhs.uk/apps-library), especially for substance abuse and smoking cessation. The American Psychiatric Association and other professional groups have developed standards for evaluating apps and promoting user reviews.

While the efficacy of some apps has been established, most mental health apps (estimated total in October 2021: between 10,000 and 20,000) have been marketed without supporting data, resting on face value, and anecdotal evidence. Driving factors of 24/7 availability, low cost, and improvement with monitoring and process (rather than outcome) fuel more and more apps. Developers of medical and mental health apps often are frustrated by low uptake and use, envying apps that suddenly become widely used, for example, Tok-tok (Doujin in China) that, in 4 years, had 2 billion downloads worldwide (315 million as TikTok) and a third of social media users, a haven for short, fun videos. Harmful effects of some medical/mental health apps are not studied but occasionally spotted, for example, a suicide prevention app that was reported to increase suicidal thinking.

Telemedicine/Telepsychiatry

The first recorded use of telemedicine was in 1906 (ECG recordings by telephone). Digital images are routinely transmitted anywhere by radiologists. Telemedicine has been used for specialty consultations for decades. The costs for equipment and transmission became negligible as Zoom, GoToWebinAir, Cisco Webex, and other videoconferencing platforms offered free time. Increasing use requires payment, especially for HIPAA-compatible platforms. The COVID-19 pandemic severely limited in-person medical visits, so the use of videoconferencing skyrocketed. Telehealth claim lines rose 2,817% from December 2019 to December 2020, according to a March 2021 report that revealed how the COVID-19 pandemic has changed how health care is delivered and consumed. Compared to the same month in 2019, the volume of telehealth claim lines rose from 0.22% to 6.5%. Multiple policies and procedures were suspended, including medical licensing and liability laws, varying by state. The CMS provided federal waivers for Medicare and Medicaid, generally adopted by health insured. As the pandemic emergency wanes, some policy suspensions are being allowed to expire or are being reinstituted. At the same time, some changes may become permanent, for example, moving toward medical licensure across states or telemedical licenses that are separate from being physically present to see patients within a state. Florida now provides such a telemedical license. Inevitably, some data are lost off camera, and production values often are inconsistent, with too little attention to light, sound, or image centering. Data reduction is less of an issue for radiology, pathology, and, to a lesser extent, dermatology.

The first recorded use of telepsychiatry was in 1957. Telepsychiatry has benefitted by being seen as part of telemedicine, with established policies and procedures. Psychiatrists have been pleasantly surprised by most patients readily accepting telepsychiatry, especially happy with more scheduling flexibility and no travel time. The American Psychiatric Association offers a helpful toolkit (https://www.psychiatry.org/psychiatrists/practice/telepsychiatry/toolkit/) and collaborated in 2018 with the American Telehealth Association to offer Best Practices in Videoconferencing-Based Telemental Health (available at https://www.psychiatry.org/psychiatrists/practice/telepsychiatry/toolkit/best-practices)

Remote Monitoring. Patients with chronic illnesses treated according to significant biomarkers such as blood sugar for diabetes or blood pressure for hypertension are easily monitored either by harvesting the readings they take on themselves or through body-attached sensors that feed directly into their caretakers' inboxes.

> *In behavioral care, remote patient monitoring devices can help monitor sleep, diet, exercise, mood, and more…the government has been very willing to support and encourage increased utilization alongside telehealth. The efficacy of such tools is so high that Medicare has reimbursed for an increasing number of new RPM-related CPT codes for the last several years.* (With permission from Maheu MM. What is remote patient monitoring? Telehealth News, July 23, 2021. Available at: https://blog.telehealth.org/remote-patient-monitoring-2/)

Pending legislation would increase support. Companies offer remote vital signs and other monitoring during and between sessions and integration into the EHR. Biofeedback and neurofeedback remote monitoring has been reported for several problems with significant biomarkers, for example, anger management.

Reimbursement. Telemedicine reimbursement began in 1999 via Medicare. Coverage was expanded with the Telehealth and Improvement and Modernization Act of 2002. Some changes included the requirement that the "physician or provider" at a distant site be paid the same (consistent with parity laws) as if the services were provided face to face, the elimination of fee sharing, and the expansion of Current Procedural Terminology (CPT) codes eligible for telepsychiatry reimbursement. The

American Telemedicine Association, founded in 1993 (www.americantelemed.org), offers institutional memberships (currently about 400). It refers individual practitioners wishing to start using telemedicine to the National Consortium of Telehealth Resource Centers (12 regional, 2 national, https://telehealthresourcecenter.org/), funded by the Health Resources and Services Administration of HHS to increase rural health services. The American Medical Association, the American Psychiatric Association, and other medical organizations provide toolkits and other assistance to practitioners, while their committees promote the use and favorable policies.

Many studies have shown telepsychiatry equal to in-person visits for efficacy, quality of care, and patient satisfaction. Findings are robust across multiple patient groups and settings while lowering cost and increasing access and convenience. Prior to the pandemic, reimbursement claims typically fell short of budgeted amounts for Medicare and Medicaid, set to encourage use. Telemedicine had ups (especially favorable studies) and downs. In 2015, both Texas and Maryland significantly restricted telemedicine. The Association of Telehealth Service Providers and the Telemedicine Information Exchange shut down. Telemedicine showed some growth in the private sector, but the pandemic brought about exponential growth. Suddenly, there was no choice but to add telemedicine to a comfortable in-person practice. That practice had been locked down. Telemedicine was the only thing to do.

With the pandemic, use skyrocketed (85% to 90% practicing some form of electronic communication with their patients (36% of all patients) as of October 2021, and receiving payments is less of a problem than usual, especially as insurers were not having to pay for many elective procedures that were not being done.

Highly competent enthusiasts did many positive studies. However, practical experience with Zoom fatigue or the limitations of online schooling has resulted in a consensus that televisits often leave something to be desired. They seem to be less helpful and definitive than in-person visits, so hybrid models alternating the two have developed.

Government Support. Successive administrations have supported more access to health and the internet, especially in rural areas. Federal support began in 966 with $42 million from the National Library of Medicine for telemedicine projects. Funding increased fitfully over the years. The 2018 Farm Bill, passed December 12, 2018, included a grant program for distance learning and telemedicine (a 20% set aside whatever rural hospital funding was to be done) and $82 million/year for 2019 through 2023. With the pandemic, the Coronavirus Aid, Relief, and Economic Security (CARES) Act became law on March 20, 2020, establishing the Federal Communications Commission (FCC) COVID-19 Telehealth Program with initial funding of $200 million and additional funding for broadband initiatives in rural and low-income areas. The Consolidated Appropriations Act of 2021, passed December 21, 2020, added $250 million (for providers) to the FCC program, $3.2 billion to ensure that low-income consumers can access internet services at home. In addition, eligible telehealth individuals under Medicare can utilize telehealth to diagnose, treat, or evaluate mental health disorders without geographic restrictions. This new rule will become permanent after the Public Health Emergency ends. This new law also permits beneficiaries to receive telehealth services from their home for mental health diagnosis, treatment, or evaluation (in addition to substance use disorder treatment, which was previously allowed). The new law will apply to most telemental health services, including counseling, psychotherapy, and psychiatric evaluations.

There are critical stipulations regarding the scope of eligible telemental health services. Specifically, eligible patients must have an existing in-person relationship with a provider, as defined by at least one in-person visit within 6 months prior to the telehealth encounter and subsequent periods as determined by the Secretary. However, as written, the bill's language appears to require the in-person visit regardless of where the patient is located (e.g., home, clinic, doctor's office) if that location is newly eligible under the provision (e.g., a doctor's office in an urban area). This in-person requirement would not apply to any eligible location under the previous rules (e.g., eligible originating sites in rural areas).

Some states and some localities appropriated additional amounts to enhance telemedical services.

Licensure Issues. If the patients and physician are in the same state, the state medical license provides the practice framework, but being in different states has led to alternate licenses. The Commonwealth Fund summarized state accommodations to support telemedicine through the pandemic: "Twenty-two states changed laws or policies during the pandemic to require more robust insurance coverage of telemedicine." They used legislation, executive orders, and other agency actions, such as bulletins, notices, and executive orders from the Department of Insurance or a similar agency. Most of these states pursued changes via administrative action. The use of executive authority allowed states to move relatively quickly during the crisis, though it has meant that the new telemedicine coverage requirements are temporary. For example, seven governors included specific telemedicine coverage requirements in executive orders, binding during the public health emergency but expiring after that. In addition, some states used bulletins, notices, or executive orders from the Department of Insurance or a similar agency to enhance coverage. Administrative action also moved through various other vehicles (such as bulletins, notices, and executive orders from the Department of Insurance or similar agency) in states where sufficient statutory authority existed or an executive order created new authority without detailing specific requirements.

Eight states passed new legislation. Although the legislative process takes more time, it is necessary for permanent changes. Five states that did not previously require coverage of telemedicine services changed their policies during the pandemic, bringing the total number of states that currently require coverage to 40. In addition to securing this broad coverage obligation, states took other steps to enhance coverage and ensure access to health services while in-person care was limited during the pandemic.

However, waivers expired or were rolled back (e.g., Johns Hopkins stopped telemedicine for patients living in Virginia). By September 1, 2021, the Alliance for Connected Care (https://connectwithcare.org/) showed waivers in use in only 17 states.

Malpractice. Additional risks related to telemedicine include cases based on prescribing medicines across state lines without having conducted an in-person examination, failure to recognize when telemedicine is not adequate for the patient's illness, violating state laws, data breaches, diagnostic errors, informed consent, lack of policies and protocols, and others. Malpractice involved not following usual and customary procedures, inevitably ambiguous, shifting, and arguable during a pandemic. However, a Lexis-Nexis review of 551 court decisions (not cases settled otherwise) in October 2018 showed no recorded cases involving direct-to-consumer telemedicine practice; the same methodology covering January 1950 to January 2021 found no cases related to skin cancer diagnosis. The Public Health Emergency declaration covering the pandemic led federal and state governments to shield health care professionals from medical liability related to Covid care. Limited exceptions apply for gross negligence, criminal misconduct, and providing care while intoxicated. These protections preempted state and local laws inconsistent with the CARES Act. However, state laws that provide greater liability protections are not preempted.

A list of liability protections for in-state and out-of-state physicians, including volunteers, relicensed or recently licensed medical professionals, services provided through telehealth, and services consistent with expanded scopes of practice are available from the National Governors Association (https://www.nga.org/coronavirus/#states). In March 2020, HHS issued a limited waiver of HIPAA Sanctions and Penalties for the duration of the emergency (https://www.hhs.gov/sites/default/files/hipaa-and-covid-19-limited-hipaa-waiver-bulletin-508.pdf). Insurers vary in charging additional malpractice premiums when a practice is expanded to telepsychiatry (https://www.psychiatry.org/psychiatrists/practice/telepsychiatry/toolkit/telepsychiatry-malpractice-issues). Psychiatrists rarely are sued for malpractice. It is too early to tell if more use of telepsychiatry will make a difference.

A cost-effective, desirable future for medical malpractice litigation can be achieved by emulating the systems found in Denmark, Finland, New Zealand, Norway, and Sweden. It is accepted that adverse outcomes occur from medical treatment, so compensation is provided quickly to those who need it (similar to worker’s compensation). The economic incentive having been removed, clinicians convene to look at the treatment provided and determine fault, if any.

Cost Benefit/Effectiveness/Outcomes. The telemedicine efficacy literature has grown steadily with hardly any negative studies.

Recent applications have shown the efficacy of using telepsychiatry for:

- Buprenorphine maintenance
- Managing complex cases in primary care
- Factors associated with use and satisfaction in rural Virginia
- Use of a program to assist self-managed medical abortion
- Depressive symptoms in comorbid cardiac or diabetic patients
- A computerized adaptive scale for suicidal youth

A June 2021 Cambridge Review of the effectiveness of telepsychiatry surveyed three databases, Cochrane, PubMed, and PsychInfo. They identified 961 papers (English only), reducing those to 321 using exclusion criteria and removal of duplicates. Using thematic analysis, the authors found five themes that occurred across all papers in relation to the effectiveness of telepsychiatry.

- **Patient and clinician satisfaction**—there is consistently high patient satisfaction with telepsychiatry but lower clinician satisfaction, often resulting from cynicism and a lack of familiarity. Clinician satisfaction increases when clinicians trial telepsychiatry and become more positive about its uses.
- **Diagnostic reliability**—telepsychiatry was found to have high levels of interrater reliability equivalent to face-to-face consultations for common disorders, including mood and psychotic disorders, substance misuse, and dementia. It was also found to have high levels of diagnostic reliability across age groups.
- **Outcomes**—telepsychiatry has been found to reduce symptoms of common psychiatric disorders and improve quality of life in various environments, including emergency departments (EDs), inpatient units, and prisons. Telepsychiatry increases access to specialized services resulting in quicker access to treatment and reduced admissions.
- **Technology**—without adequate internet connectivity, clinicians cannot conduct an appropriate mental state examination, and the therapeutic relationship becomes challenging. Inadequate technology can impact the effectiveness of telepsychiatry among those who are socioeconomically disadvantaged and may not have access to appropriate technology.
- **Professional guidance**—there is a concerning lack of guidance around the use of telepsychiatry. There is a lack of standardization without clear protocols, and clinicians are unwilling to integrate telepsychiatry into their practice. The main concerns raised are confidentiality, consent, the appropriateness of specific patient groups, and emergencies.
- **Cost-effectiveness** is established by offsetting travel time and expense, as equipment and compliance costs have shrunk.

Internet

The internet was established in 1969 at the University of Southern California to link computers for national defense issues. Now it is accepted as a public necessity, like clean water or reliable electricity, with similar concerns about cleanliness, reliability, and access, but less regulation, policing, and enforcement. Psychiatrists and their patients use it as below.

Information on Disease, Diagnosis, Treatment, and Therapists. The internet has become a predominant source of data, with countless websites offering pages of information on health care and psychiatric problems. A Google search for “diabetes” yields “about 579 million pages.” In recent years, the diabetes searcher, despite having access to authoritative sources, has been inundated with popups advertisements for diabetes products. Seeking information about an illness or symptom may link to a symptom questionnaire and a suggested diagnosis, the more specific, the less trustworthy. Even so, some apparently reputable sites may actually be purveyors of false information. The pandemic underscores how disinformation brings money and infamy (both desired by wellness product sellers) to some purveyors and death to some believers. The antivaccine movement fosters disproven links to autism. Patients present with a mixture of information, disinformation, and questions. Visits were sometimes lengthened to answer questions; however, many physicians have learned what websites to recommend for reliable information. Often, physicians find that better-informed patients form better partnerships and tend to cooperate with treatment better. However, internet users can find almost anything they want and may refuse treatment based on overblown side-effect descriptions or demand a treatment based on a success story that does not apply to their circumstances. Hundreds of psychiatrists operate websites, setting forth facts about themselves and their treatments, which often are the path for patients to enter treatment. Many commercially operated websites may violate privacy policies with intrusive, unwanted follow-ups to those who contact them. Physicians are rated by sites that gather ratings and reviews; a malicious patient can be damaging. Periodically professional meetings include workshops on how to improve one’s web presence.

Support and Self-Help Groups. Support groups have evolved by diagnosis for most illnesses. Others deal with symptoms. These can lead to groups meeting in person or remotely, providing support, knowledge, and advice of varying helpfulness. Mental health organizations, such as the National Alliance on Mental Illness (NAMI), Mental Health America, and their local chapters offer hotlines and referrals for support and therapy. Numerous sites offer to find Alcoholics Anonymous (AA) meetings by zip code. Whatever the problem, you can probably find a group of people trying to deal with that problem. Some groups are hybrid, occasionally meeting in person and frequently staying in touch electronically. As voluntary associations are generally unregulated and discussions can veer off into unwanted directions, they sometimes ask a professional to be "present" or review or comment on posts (a "moderated support group"). Several hundred mental health professionals moderate such groups, carefully wording terms and conditions of participation. For example, a Yahoo search for "schizophrenia support groups" yields 2,670,000 sites.

Self-help groups are unmoderated support groups that emphasize traditional self-help mechanisms such as following steps to wellness. These groups offer low-cost, usually available support. Some follow the AA's 12 steps and offer sponsors. Members often agree to have instant messaging to reach anyone online. Email technology lends itself to mailing lists because addresses can be added to a message with ease and without costs, and emails typically include a "reply to all" option that makes it easy to respond at once to everyone who received the initial message. Variations in format allow user-friendly "chat rooms" where an email user can sign in and participate in a conversation with users who have signed in for that particular session, which typically lasts minutes or hours. However, asynchronous communications make an extension to days and weeks possible. Still other groups, such as Psychiatric Survivors (see https://www.psychiatricsurvivors.net/, but also https://mindfreedom.org, and https://www.madinamerica.com/2012/08/the-history-and-future-of-our-psychiatric-survivor-movement/) use their websites to oppose specific treatments, especially anything coercive, for example, hospitalization or involuntary medication. Other groups favor such treatments, tracing stigma to **untreated** mental illness. All support groups favor recovery and empowerment. "Recovery" does not mean "cure," but being able to move on with one's life while dealing with whatever remains of one's illness (possibly a disability, for which protocols and treatment are established). Empowerment means taking charge of one's illness and treatment, exercising choice, and protecting civil rights. Anything that smacks of being told what to do or what or how to think is objectionable. Coercive treatment is anathema, and medication is, at the least, suspect.

Online Advice, Diagnosis, and Therapy. It is estimated that thousands of mental health professionals (mostly counselors, social workers, and psychologists) are willing to pick up advice seekers online and work with them only by email. Many nonprofessionals or sued professionals justify this practice because those being helped would not otherwise come to therapy for various reasons, including geographic isolation, loss of privacy, and disability. Initially supported by venture capital from 1997 to 2000, but waning after 2001 with the appreciation of its limitations, it picked up with increased internet use. Many websites recruited therapists who provided diverse therapies before the pandemic, but there has been an exponential increase since. A Bing search for “online therapy” delivers 118,000,000 sites. Many recruit therapists who take insurance (e.g., Talkspace), and many sites offer medication online, including controlled substances (e.g., try.getcerebral.com sends the user to rx.getcerebral.com). Recent research topics include internet-delivered versus. in-person CBT for youths with obsessive-compulsive disorder (OCD) (followed by in-person visits for nonresponders) or a Cambridge database meta-analysis of CT for depression. However, research is not directed at exposing what may be a waste of time and money or harmful. It is generally agreed that mental health is worsening during the pandemic, yet therapists and prescribers of all types are scarce. As it seems a need is being met, issues of standards for good practice, quality, and enforcement will have to wait for more settled times.

Meanwhile, therapy sites are rated through user reviews, much like restaurants. Professional associations have not come forward with ethical standards. For example, the Clinical Social Work Federation, opposed in 2001 to unmet therapy, no longer mentions “online” or “unmet” in its Code of Ethics.

Prescription/Illicit Drug Sales. Physicians increasingly use electronic prescribing, mandated in some form, at least for controlled substances, in 25 states, with effective dates, mainly January 1, 2022, in another nine. All states now have prescription drug monitoring programs for controlled substances. Electronic prescribing is generally accepted, with a continuing objection related to lack of comparison shopping and privacy. Health insurers increasingly require patients to use favored online pharmacies, often avoiding contact with pharmacists. Some online pharmacies put prospective customers in touch with prescribers who will prescribe based on online questionnaires without further investigation. The Federation of State Medical Licensing Boards has barred using online questionnaires as the sole basis for prescribing medication. The Controlled Substances Act and the Food, Drug, and Cosmetic Act and been used to prosecute pharmacies that do not bother with prescriptions, reporting, and shutting down thousands of websites. Despite these efforts, the sale of medications over the internet without valid prescriptions has been a persistent problem. Pharmacies have moved offshore to countries with lax enforcement. Mail from China has been used to send fentanyl to the United States and other countries despite increased Chinese enforcement since 2019 (https://www.theguardian.com/world/2019/apr/01/china-fentanyl-substances-crackdown-us-opioids-crisis). A Yahoo search, “Where can I buy prescription drugs online?” yields 19,700,000 sites. A review of the 591 FDA press releases on enforcement of the past 3 years yields six cases related to internet prescription drug sales. The largest (4,256) website crackdown was in 1996. The FDA’s most wanted fugitives data have not been updated since March 2020.

The Washington Post’s expose on the distribution of 76 billion opiate tablets used Drug Enforcement Administration (DEA) data that the DEA had gathered but not released until required to do so under freedom of information (FOI) law. Tellingly, the pills had gone to many areas with few prescribers. The FBI’s Operation Disruptor announced in February 2020 that it had made 18,000 illegal drug purchases on the “*dark web*,” sites with encrypted identities. “Infamous examples of Dark Web sites include the Silk Road and its offspring, such as Dream Market. The Silk Road was a website for the buying and selling of recreational drugs, and a lot more scary things besides” (https://www.techadvisor.com/how-to/internet/dark-web). A Bing search in October 2021, “where can I buy opiates online?” nonetheless provides promising possibilities. Counterfeit, sometimes lethal, medications continue to be sold as legitimate medications. A September 2021 DEA warning pointed to fentanyl’s role in the 96,000 drug overdose deaths in 1920, as 26% of pills sold as something else (e.g., Adderall) contained lethal amounts. The internet’s role in getting get-high drugs to those who want them is paramount.

Legitimate prescription drug sales may have diminished as pharmacy chains engage in aggressive refill policies, more challenging to stop than start, but a small legitimate traffic persists in some psychotropic medications not available in the United States. Knowledgeable physicians typically can recommend safe sites. Some patients who do not have or do not want to use drug insurance coverage (e.g., wanting to stay out of as many databases as possible) have found their medicines available on the internet cost less than the copay they would have had to pay for using their insurance. As more medicines are made over the counter, they may be less expensive than prescription versions—some pharmacists in some jurisdictions are still prohibited from telling patients so.

Gratification: Sex, Games, and Addiction

PERSONAL EXPERIMENTATION/VALIDATION, SUCH AS DIFFERENT IDENTITIES OR REINFORCING BELIEFS. Pornography remains the top source of gratification from the internet, estimated to constitute 4% to 30% of total content. A Google search “pornography sites” yields 49,500,000 possibilities, but only 42,000 are recognized as active. For April 2020, the world’s top porn site, XVideos, amassed 2.92 billion views among 382 million unique visitors who averaged 12.5 minutes per visit. Almost any man with an internet connection has watched at least once, but Pornhub, the second most popular site, reports that 0.2% of men spent more than 2 hours a week on the site in 2019. It has not posted 2020 statistics, but use increased by 8 billion visits in 2019 from 2018, indicating that 2020 visits were higher. Fantasies related to any sexual appetite can be found, the purpose being sexual arousal and masturbation. Pornography remains controversial, seen by some as a victimless outlet for urges that might otherwise be hurtful, objected to by feminists who see women treated as objects. Excessive use (defined by the patient or significant other) brings many voluntarily to treatment.

The internet is used for dating apps, another exponentially increasing way of connecting since pandemic lockdowns left few choices. Some apps promote recreational sex and extramarital affairs. In the October 2021 rating of top sites, Adult Friend Finder (78 million members) and Ashley Madison (40 million) led the more specialized sites. Problems related to false presentations, exploitations, and crimes are covered well in the media as cautionary. Couples have been known to meet on the internet, fall in love, and arrange to marry, only to find they should have verified one another's sex and gender identities. The movie "You've Got Mail" showed the use of email as a way of expressing kinder, nobler feelings than Tom Hanks' character could do in real life as a cutthroat businessman. However, studies on email show that communications are more likely to be bitter, angry ("flaming"), suspicious, or paranoid than what would have been said in a face-to-face meeting. Easy, impulsive, quick responses play a role. Also, there are fewer cues (facial expressions or nonverbal communication) to facilitate empathic civil discourse.

The internet can offer not only anonymity but also a choice of other identities, such as Second Life (http://www.secondlife.com) which describes itself as "a 3-D world entirely created by its Residents". (Since its founding in 2003, 98,944 regions have been created by 68, 687,941 of them, of whom 45,297 were online during a check on October 12, 2021.) They create and chat using voice and text chat, buying and selling with Linden dollars, convertible to U.S. dollar (as of October 7, 2021, 320 Linden Dollars equal $1). Other alternative life sites include Mozilla Hubs (free, open source), VRChat (free), Tower Unite (paid), and Furcadia (free). Alternate lifers tend to be more colorful, less restrained, or criminal as the sites allow crime and deceit without penalty. Psychiatrists become involved according to whatever problems result.

Although "internet addiction" is a widely used lay and professional term, it was not included in DSM-5 in 2014. Most apparent "addicts" have been found to decrease excessive use and pursue specific knowledge and communication goals rather than surf aimlessly. Internet gaming was included as warranting further study. Its status was updated by the American Psychiatric Association advice to the public on its website (https://www.psychiatry.org/patients-families/internet-gaming):

> *Whether internet gaming should be classified as an addiction/mental disorder is the subject of much debate and a growing body of research. There is neurological research showing similarities in changes in the brain between video gaming and addictive substances.*

> *A study published in the American Journal of Psychiatry in March 2017 sought to examine the validity and reliability of the criteria for internet gaming disorder, compare it to research on gambling addiction and problem gaming, and estimate its impact on physical, social, and mental health. The study found that among those who played games, most did not report any symptoms of internet gaming disorder and the percentage of people that might qualify for internet gaming disorder is extremely small.*

> *The research involved several studies of adults in the U.S., United Kingdom, Canada, and Germany. They found that more than 86 percent of young adults ages 18 to 24 and more than 65 percent of all adults had recently played online games. The percentages of men and women who recently played were roughly equal. However, the research is mixed on whether those who met the criteria for internet gaming disorder had poorer emotional, physical and mental health than those who did not meet the criteria.*

> *The researchers found that 0.3 to 1.0 percent of the general population might qualify for a potential diagnosis of internet gaming disorder. The authors suggest an important distinction between passionate engagement (someone enthusiastic and focused on gaming) and pathology (someone with an illness/addiction). Whether the person is distressed with his/her gaming may be the key factor distinguishing the two.* (Reprinted with permission from the Psychiatry.org, “Internet Gaming,” James Sherer, section “Research is Ongoing” [Copyright © 2023]. American Psychiatric Association. All Rights Reserved.)

However, according to WHO:

> *Gaming disorder, with its online and offline variants, has been included in the 11th edition of the International Classification of Diseases (ICD-11) as a clinically recognizable and clinically significant syndrome, when the pattern of gaming behaviour is of such a nature and intensity that it results in marked distress or significant impairment in personal, family, social, educational or occupational functioning.* (Reproduced from World Health Organization: Inclusion of "gaming disorder" in ICD-11, © 2018. https://www.who.int/news/item/14-09-2018-inclusion-of-gaming-disorder-in-icd-11)

ICD-11 was released in mid-2018. Health concerns associated with gaming behavior are not limited to gaming disorder but also include other aspects of health (e.g., insufficient physical activity, unhealthy diet, problems with eyesight or hearing, musculoskeletal problems, sleep deprivation, aggressive behavior, and depression) and psychosocial functioning. Studies suggest that gaming disorder affects only a small proportion of people who engage in digital- or video-gaming activities. However, people who partake in gaming should be alert to the amount of time they spend on gaming activities, mainly when it is to the exclusion of other daily activities, as well as to any changes in their physical or psychological health and social functioning that could be attributed to their pattern of gaming behavior.

SOCIAL MEDIA. In October 2021, "app engagement" emerged as key to getting apps downloaded and used as Facebook was alleged to choose growth over beneficence and concealed data showing the harmful effects of apps. Some Instagram users, especially teenage girls, felt worse about their appearance, while other surveys reported increases in suicidal ideation, insecurity, and fear of ostracism if quitting use. Anxiety and depression, already widespread (20% of Americans thought to have anxiety disorders), heightened during the Covid pandemic as anger (often displaced) and frustration, fueled by disinformation and political polarization. App engagement research starts with the initial screen and leads to a user clicking on links that lead deeper into the subject. Engagement metrics, extensively researched, are bolstered by artificial intelligence to generate algorithms that promote more engagement ("engagement-based ranking"), often leading to salacious, extreme, attractive, but harmful sites that are meant to keep the user online longer. Facebook and its apps were alleged to have chosen links that attracted users because of sensationalistic falsities. Several studies have shown that falsities spread faster on social media than truth, the briefer, the faster, as on Twitter (X) (daily use in 2021: 199 million, 69 million in the United States). Facebook was termed “morally bankrupt” for not doing more to spread the truth and stop the spread of disinformation. It responded by pointing out that the mountains of leaked data were based on user surveys rather than controlled studies and calling for government regulation of social media. However, recent polling, such as that of 1,638 users (ages 16+) reported on October 8, 2021, showed averages of 20% to 25% feeling more anxious, depressed, alone, left out, and angry after Facebook and Instagram use, wanting to stop using but “can’t stop.” Those who cannot stop probably define their inability to do so as addiction. Aside from gaming, pornography (cybersex), and social media (cyber-relational), overshopping for goods is a significant problem, for which several measurement scales are helpful, especially for convincing patients they have problems. About 30% of those who describe themselves as addicted reported using the internet to escape negative feelings because it was always available at a low cost. This percentage is slightly lower than a recent poll of social media users who described themselves as unpleasantly addicted to social media but unable to stop.

Government regulation is likely, to focus on the alleged algorithms, disinformation, and spread of hate and violence. For example, ISIS online recruiting was very effective, using algorithms that picked up potential recruits and led them through a series of websites.

Internet Treatment Issues. The dependency/addictive component of the issues described above is similar to those of other addictions: withdrawal to less impaired use, often periods of abstinence. Initial treatment may involve one of the thousands of online help websites, but treatment success will depend on engagement with more regarding offline activities. Those who describe being unable to stop using the internet despite negative or joyless feelings may have an obsessive-compulsive dimension amenable to OCD treatments. Issues of disinformation, identity, hate, polarization, and conspiracy involvement are complex; findings from terrorism involvement and disengagement are helpful. Support groups and psychotherapy are beneficial.

Cyberbullying of adolescents is a particular problem, often unreported (http://www.bullyingstatistics.org). Over half of adolescents and teens have been bullied online, and about the same number have engaged in cyberbullying. More than one in three young people have experienced cyberthreats online. Over 25% of adolescents and teens have been bullied repeatedly through their phones or the internet. Over half of young people do not tell their parents when cyberbullying occurs.

There are increasing reports of social media-induced sociogenic illnesses. For example, during the pandemic, neurologists began to encounter patients, particularly teenage girls, with unusual Tourette-like tics, which appeared to be caused by watching TikTok videos of people reporting to have Tourette syndrome.

Virtual Reality

Virtual reality as therapy was pioneered in the early 1990s by Max and Sarah North and was used for various phobias. It usually functions as a type of desensitization therapy. Tips for more effective practice are available (see the article by Krzystanek and colleagues in the references). Several journals (e.g., Virtual Reality, Frontiers of Virtual Reality, International Journal of Virtual Reality) report generally positive research. Adverse outcomes, infrequent, include “cybersickness” and transient identity and reality loss.

MEDICAL RECORDS

Medical recording accompanied the developments of writing and medical care in ancient civilizations. The earliest known medical record is the Edwin Smith Surgical Papyrus, dated to between 5,000 and 4,500 years ago, a treatise that includes 48 typical cases, mainly head injury, with histories ranging from head trauma to infection. Although format and content have greatly expanded over the years, the medical record mainly describes past events recorded on paper. Carole Siegel and Susan Fisher’s unique study of the psychiatric record can be summarized as generally applying to records. Practitioners with access to patients go to the patient, whereas those who cannot get to patients have to use the record. However, many records are illegible, and finding data can be complex.

Early efforts to convert to electronic formats were inspired by the ease of retrieval and legibility but suffered from overexpectation, high cost, nonstandardization of data elements, and other key factors that impeded accessible communication among physicians. Despite a lack of political support, some institutions developed excellent records, but there was little spread. The current effort began in 2004 with federal initiatives: The ONCHIT and an HHS report laying out a 10-year plan for federal, state, and local leadership. It has resulted in multiple public–private partnerships, including the Health Information Technology Standards Panel to define elements and standards and the CCHIT. electronic medical records (EMRs) are expected to improve patient care by reducing medical errors and clinical decision support. They are also expected to reduce care costs by reducing overutilization and promoting continuity of care as clinicians communicate individually or in groups through RHIOs and HIEs. For example, the Veterans Administration's excellent EHR makes it possible to retrieve a patient's record (with permission) anywhere, including all diagnoses, medications, and treatment. This ability has proven to be of great benefit after local records are destroyed, such as happened due to Hurricane Katrina.

Problems in EHRs

Despite the legislation, funding, and universal consensus promoting EHRs, they are cited in every survey of physician burnout as one of the two leading causes of physician burnout (EHRs were marketed to ease the other, prior authorization). Both issues are related to documentation demands imposed from outside the doctor–patient relationship. Both issues have been promoted to improve patient care and reduce costs, but numerous studies show increased costs related to EHRs and prior authorizations. Physicians have been shown to spend as much or more time with their computers as with patients. Workarounds to shorten tasks lead to the causes of EHR malpractice listed below and create sloppiness. Errors can be made in the flash of a click on the wrong item in a table and not picked up, a particular type of typo.

The sources of EMR frustration and bloat include the following:

- Insurance carriers, for profit (mostly) and not for profit, that set documentation requirements conduct audits and withhold payments for services already rendered, sometimes demanding the return of payments already made.

- Accreditors and remedies for deficiencies. Typically, a record problem found on review results in zealous attempts to ensure that every progress note addresses the deficiency, so the record bloats with each added deficiency remedy. If inspectors and accreditors are not satisfied, reimbursements and institutional reputation will suffer.
- The lack of software usability tests led to sloppily designed and cobbled-together systems. Examples are programs requiring users to enter data via one browser but use a different browser for drug ordering.
- Payment incentives, especially for evaluation and management (E/M) coding justification.

The Social Security Act 1862 (a)(1)(A), effective January 1, 1991, defined medical necessity as the basis for reimbursement. Concerns developed over the years that doctors were being paid more for **doing** (e.g., a surgical procedure) than **thinking** (e.g., taking a history, examining, formulating a treatment plan, prescribing, and organizing care for an obese, homebound patient with arthritis, chronic obstructive pulmonary disease, and diabetes). Evaluating and management codes, ranging from simple to complex to reflect levels of thought, were published in 1995 and 1997, well before EHRs. Documentation in the EHR is reviewed to ascertain whether E/M codes are justified. The more complex codes bring greater reimbursement, hence the more documentation. The CMS Meaningful Use requirements (https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/EHR_Medicare_Stg1_BegGuide.pdf) set 13 core objectives, 5 menu objectives, and 9 quality measures, some of which require EHR content that can be described as incomplete and must be expanded. The 2-year anniversary of “Patients over Paperwork was celebrated in 2019: On “October 29, 2019, CMS celebrated the second anniversary of Patients over Paperwork, which began with an RFI that yielded more than 3,000 data points on more than 1,100 different issues. To date, Patients over Paperwork has yielded estimated savings of $6.6 billion to the medical community with a reduction of approximately 42 million hours of burden through 2021.” (https://www.cms.gov/files/document/november-2019-patients-over-paperwork-newsletter.pdf, which describes some achievements unrelated to reducing burden, such as payment changes). The need for more burden relief prompted the 21st Century Cures Act provision (Section 13103, amending HITECH) that led to *The Strategy on Reducing Regulatory and Administrative Burden Relating to the Use of Health IT and EHRs*, final in February 2020 after stakeholder comment (1,400 pages) on the 2019 draft (https://www.healthit.gov/sites/default/files/page/2020-02/BurdenReport_0.pdf). Four work groups, consisting of HHS workers, made recommendations ("consider…improve") for clinical documentation, HIT usability and the user experience, EHR reporting, and public health reporting. Implementation will require specific recommendations, with the change expected in 3 to 5 years. The strategy does not suggest incentives for brevity or finding methods for reimbursement apart from clinical notes. A resolution to separate clinical notes from reimbursement is being considered by the Medical Society of the State of New York.

Interoperability

Exchange of patient HIT is a rare issue that has virtually universal political support. Under HIPAA, the HITECH of 2009, and the Affordable Care Act (ACA) of 2010, voluntary associations of covered entities were set up and variably funded: RHIOs, Community Health Information Exchanges (CHINs), and Health Information Exchanges (HIEs). These are viewed as the primary hope for increasing quality and safety and decreasing overutilization and fraud, becoming the Nationwide Health Information Network (NHIN), reported positively (https://pubmed.ncbi.nlm.nih.gov/30234636/) as having "implemented secure exchange of health records through utilization of the Internet." The NHIN has greatly assisted in achieving the goals of the HITEC Act by promoting the adoption of Meaningful Use. Epic introduced a Health Information Exchange platform, Care Everywhere, facilitating Health Information Exchange availability. The purpose of this research was to determine the impact of NHIN and Epic Care's Care Everywhere on health care to determine whether their use in the ED has increased. The study relied on a literature review. Twenty-eight sources were referenced. With the NHIN implementation, repeated visits were decreased, visit times became faster, and charges were lower. The ED reported significant benefits in sharing clinical information. The NHIN implementation throughout the ED has increased the quality of health care; duplicated tests and drug usage were determined, and a reduction of the ED length of stay was also achieved. However, the NHIN is **not** a central database containing PHI that can be looked up from an emergency room trying to help a patient. Instead, the NHIN is a set of standards and policies. RHIOs exist in most states, statewide or local, currently adequately funded but not immune from organizational problems. New York consolidated its RHIOs down to eight. California's RHIO shut down in 2010; instead, the California Trusted Exchange Network is used for PHI exchanges. RHIO requests for funding from users have been strongly resisted, not seeing the point of funding what they see as unnecessary bureaucracy. Apart from RHIOs and HIEs, providers who know they are taking care of a particular patient can exchange information according to traditional consent methods (as a covered entity if transmitted electronically). Software incompatibility has been a significant barrier. Although CMS has delayed some payer-to-payer interoperability enforcement, two rules took effect in July 1, 2021, explained in FAQs posted in September 15, 2021 (https://www.cms.gov/Regulations-and-Guidance/Guidance/Interoperability/index).

Malpractice

Some studies show that EMR use reduces risk because the record is entirely legible and retrievable, using software that prompts completeness and consistency. However, law firms seeking cases argue that EHRs increase the risk for lawsuits (https://www.jcreiterlaw.com/posts/8-reasons-your-electronic-medical-records-may-result-in-a-medical-malpractice-lawsuit/). Various design flaws, outdated information, errors in voice recognition, and other transcribing errors can put practitioners at risk. User error, such as inappropriate copying and pasting, can also lead to medical negligence. For example, doctors may erroneously copy and paste information from the last patient visit into the record while leaving out what has changed. Similarly, autopopulate functions can add erroneous information. Simple typos can also have disastrous results if they result in a wrong diagnosis or procedure. Finally, the record can harm the doctor–patient relations, cutting down on direct interactions (most of us are familiar with the doctor who seems more interested in their screen than us), or causing burnout and exhaustion.

The Medical Society of the State of New York suggested simply using clinical notes solely to communicate with other clinicians, arguing that clinicians become a party to fraud when the EHR is used for billing in their name, that they are required to report examinations barely done, or attest to yes-or-no suicidality when the situation is ambiguous.

Privacy and Security

EHR regulations carry a substantial documentation burden, both for an overall plan for protecting PHI and documenting informed consent, with additional requirements that each individual practice document how it will protect PHI in electronic media. The more secondary sources of PHI exist, the more difficult it is to determine where confidentiality may have been lost. Such secondary sources can be helpful to individuals seeking data on a particular patient. Such people may be criminals, unethical private investigators, and others. They proceed by hacking into databases and retrieving PHI that might bear on competence, custody, divorce, and other adversarial situations. They, rather than organizations, may be interested in the content of specific emails. Such individuals have broken into offices to get individual patient records in the past. Computer break-ins can be simple, depending on the relative skills of the hacker versus the burglar, and may leave no trace. Although current guidelines for the transmission and storage of PHI recommend password protection and encryption, it would be unwise to expect that such safeguards would prevail against a specific, determined, individual attack. Penalties for criminal acts related to medical records have existed for years with variable enforcement. The risk of violation of a competently protected personal medical record (PMR) probably is less than that of a paper record, whereas a mass of EHRs may be subject to attack by a group seeking money rather than specific information on one individual.

Despite almost weekly reports of confidentiality lapses involving thousands of people, complaints about the mere few hundred criminal complaints and just six convictions for privacy violations under the HIPAA from the first in 2004 through April 2015 resulted in the HITECH breach reporting, and penalties cited above, summarized by HHS (https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/data/enforcement-highlights/index.html) Since the compliance date of the Privacy Rule in April 2003, OCR (Office of Civil Rights) has received over 272,923 HIPAA complaints and has initiated over 1,101 compliance reviews. Ninety eight percent of these cases have been resolved (268,469). In addition, OCR has investigated and resolved over 29,081 cases by requiring changes in privacy practices and corrective actions by providing technical assistance to HIPAA-covered entities and their business associates. Corrective actions obtained by OCR from these entities have resulted in systemic change, affecting all the individuals they serve. OCR has successfully enforced the HIPAA Rules by applying corrective measures in all cases where an investigation indicates noncompliance by the covered entity or their business associate. OCR has settled or imposed a civil money penalty in 100 cases, resulting in a total dollar amount of $130,980,482.00. OCR has investigated complaints against various entities, including national pharmacy chains, major medical centers, group health plans, hospital chains, and small provider offices.

In another 13,294 cases, the investigations found that no violation had occurred. In addition, in 49,553 cases, OCR has intervened early and provided technical assistance to HIPAA-covered entities, their business associates, and individuals exercising their rights under the Privacy Rule without the need for an investigation.

In the rest of the completed cases (176,541), OCR determined that the complaint did not present an eligible case for enforcement. These include cases in which:

- OCR lacks jurisdiction under HIPAA. For example, in cases alleging a violation by an entity not covered by HIPAA
- The complaint is untimely or withdrawn by the filer
- The activity described does not violate the HIPAA Rules. For example, in cases where the covered entity has disclosed PHI in circumstances in which the Privacy Rule permits such a disclosure

From the compliance date to the present, the compliance issues most often alleged in complaints are compiled cumulatively, in order of frequency:

- Impermissible uses and disclosures of PHI
- Lack of safeguards of PHI
- Lack of patient access to their PHI
- Lack of administrative safeguards of electronic PHI; and
- Use or disclosure of more than the minimum necessary PHI

The most common types of covered entities that have been alleged to have committed violations are, in order of frequency: general hospitals;

private practices and physicians; outpatient facilities; pharmacies; and community health centers.

OCR refers to the Department of Justice (DOJ) for criminal investigation appropriate cases involving the knowing disclosure or obtaining PHI in violation of the rules. As of the date of this summary, OCR made 1,182 such referrals to DOJ.

Breaches Increase

HHS is required to report breaches annually to Congress, which shows increases in the numbers of people whose data were breached: Per the HIPAA Journal: “More large healthcare data breaches were reported in 2020 than in any other year since the HITECH Act called for the U.S. Department of Health and Human Services’ Office for Civil Rights to start publishing healthcare data breach figures on its website.” In 2020, health care data breaches of 500 or more records were reported at a rate of more than 1.76 per day. In addition, 2020 saw 642 significant data breaches reported by health care providers, health plans, health care clearinghouses, and business associates of those entities—25% more than 2019, which was also a record-breaking year. More than twice the number of data breaches are now being reported than 6 years ago and three times the number of data breaches in 2010.

Key Takeaways

- Twenty-five percent year-over-year increase in health care data breaches
- Health care data breaches have doubled since 2014
- 642 health care data breaches of 500 or more records were reported in 2020
- 1.76 data breaches of 500 or more health care records were reported each day in 2020
- 2020 saw more than 29 million health care records breached

One breach involved more than 10 million records, and 63 saw more than 100 K records breached. Hacking/IT incidents accounted for 67% of data breaches and 92% of breached records; 3,705 data breaches of 500 or more records have been reported since October 2009; 78 million health care records have been breached since October 2009.

Ransomware

Rather than breaching PHI and selling the data, criminals have found that hacking into a system and paralyzing it until a ransom is paid is very profitable. Ransomware payments reached over $400 million globally in 2020 and topped $81 million in the first quarter of 2021 (https://www.whitehouse.gov/briefing-room/statements-releases/2021/10/13/fact-sheet-ongoing-public-u-s-efforts-to-counter-ransomware). The health care sector, which was already stretched and stressed by the pandemic, continued to be heavily targeted in 2020 with at least 560 facilities being impacted in 80 separate incidents (an attack on a health system can impact multiple facilities) (https://blog.emsisoft.com). These data do not include lost lives or revenue.

CLINICAL ISSUES

These include symptoms, diagnoses, and defense mechanisms. Although email and telepsychiatry share some of the same indications, the mode of initial interaction may vary. With email, the patient and therapist generally have met on their own, either face to face or online. Telepsychiatry presently consists mainly of consultations, with referrals arranged by someone where the patient is located. Availability of consultant expertise usually is the key indication. Home health care monitoring can be beneficial for situations where a home visit would have been used. Both may be useful to overcome limitations of access, distance, and disability. Both are valuable adjuncts to therapy in situations where the patient and therapist meet face to face but want more frequent contact. Its format lends itself, for example, to the sort of written daily homework expected in cognitive behavioral therapy (CBT). It may be helpful as a way of getting patients into treatment that they otherwise would not seek, but that should then progress to more substantial involvement via telepsychiatry or face-to-face contact. As webcams have come into wider personal use, some patients have found telepsychiatry to be more convenient and of lower cost than travel to a therapist’s office.

Virtual Reality

Experiments showing efficacy in various forms of situational anxiety are convincing. Most phobias and fear-provoking situations can be simulated in virtual reality therapy (VRT). Investigations of VRT's usefulness in OCD, attention deficit disorder (ADD), and PTSD are underway.

Limitations

The principal limitations of email are related to the data not being transmitted or available. There are potential problems in identifying both the patient and the therapist. Investigations of those offering online services found that many did not give their professional credentials or were incomplete (e.g., "MS" in a context that implies that the degree is in counseling when it may not be). Similarly, potential clients often give themselves fictional names. Although this is seen as a way of preserving anonymity, without which help might not be sought, therapy cannot progress far without more disclosure. Not all deception is deliberate. As noted earlier, patients may feel that they have a different identity on the internet, leading to their portraying problems consistent with that identity rather than their identity in real life. Although controlled studies show that patients divulge facts more rapidly and entirely to computers, most studies were done with substance abusers before the internet and the rise of confidentiality concerns. Computer-based assessments have not replaced live interviews because communicating facts does not convey how people think or feel.

On the other hand, the disinhibition of emotion noted in unstructured emails is an indication for therapy where defenses must be overcome and a limitation if the feeling is impulsive or marginally related to the issues for which therapy is sought. These same limitations apply to support and self-help groups. Therapist limitations must also be noted. Extensive use of email is not recommended for therapists who have not mastered expressive writing, typing, use of computers and the internet, encryption, security of electronic data, and how to deal with computer glitches. Therapists must be aware of how they may seem online, be able to correct miscommunications in both directions and be able to shift modalities (e.g., from online to face-to-face communication) when indicated.

Telepsychiatry, Online Therapy

A person may get very good at e-therapy and working on the internet without this result generalizing to the physical world. So-called “cyborg theory” has been invoked to describe the mechanisms whereby the animal–man boundary blurs, along with the man–computer boundary.

Reports have described predators taking advantage of victims met on the internet. Online therapists report referring patients for face-to-face therapy, but they often view this as a desirable progression rather than a complication. Chat room addicts can take over self-help groups. Patients and therapists may present false identities. Predators can pick up vulnerable people and develop private correspondences that lead to disaster.

Despite the widespread use and general acceptance of telemedicine, about two-thirds of patients who had it reported in two polls in July and September 2021 (https://media.npr.org/assets/img/2021/10/08/national-report-101221-final.pdf, 42% used, 82% satisfied, but 64% would have preferred in person) that they preferred in-person visits to physicians, despite travel time and cost and waiting room issues. In addition, several studies show that more oxytocin (promoting bonding, trust, affection) is produced when people meet in person. Limitations reported ranged from for e-therapy (https://nymag.com/press/2021/03/on-the-cover-of-new-york-the-lunacy-of-app-based-therapy.html) to telesurgery (https://jamanetwork.com/journals/jamasurgery/fullarticle/2778018), while new insurance requiring the first nonemergent visit be virtual (https://kffhealthnews.org/news/article/virtual-first-insurance-plans-telemedicine/) has appeared.

Contraindications

Both email and telepsychiatry are less desirable when the patient and psychiatrist can meet face to face. However, the patient may continue using a support or self-help group. Psychiatrists should always be sensitive to the possibility that VRT is contraindicated for patients with serious medical problems and patients taking medications with significant physiologic or psychological effects. Flickering images, especially at 8 to 12 Hz, may produce symptoms ranging from headaches to epileptic seizures, so no frame replacement speeds that approach these frequencies should be used.

Email Journaling

A patient moved to another country and put his daily journal on Google Drive. The psychiatrist reviews the entries every week and makes notes daily where appropriate. The patient, previously repeatedly hospitalized for psychoses, became unimpaired.

Telepsychiatry

- A second-opinion consultant elicited new symptoms, substantiated a different diagnosis, and recommended treatment that proved to be effective.
- A well-known psychiatrist performed what he considered an excellent interview but missed that the patient was increasingly resentful about the picture-in-picture showing her. She motioned to have the local crew on-site cover it, which they did, with a sheet of paper covering part of the consultant's image. None of this was conveyed to the consultant, who was now somewhat walled off. The patient continued to smile while the consultant suggested that her previous head trauma could benefit from a neurologic workup and reduce her need for 25 mg of diazepam a day, but observers in the room noted that her shoulders squared and her posture conveyed belligerence. The consultant's first clue that the interview had gone badly was at the end when the patient responded to his asking if she had any questions. She glared and said, "Yeah, what do I do the next time I'm feeling suicidal and don't want to call anyone?"

Virtual Reality

- A patient with social phobia was treated in her psychiatrist’s office with a virtual reality program that provided a series of social stimuli, adding and subtracting as she coped successfully with each. She realized she had a particular problem with older men and developed skills in dealing with them.
- A similar patient was never able to accept that virtual reality had anything to do with his problems and instead benefited from medication, which was decreased as he developed new skills in handling real-life challenges.

Electronic Health Records

Danielle Ofri describes making mistakes from being rushed, having multiple records open simultaneously, typos with consequences, miscommunications with patients, how EHRs negatively affect mindsets and slow down patient care. (https://danielleofri.com/tag/ehr/)

ETHICAL ISSUES

Ethical issues in electronic communication have been explored, and consensus statements are available for the individual practitioner. While the general principles of medical ethics apply, special consideration is required to explain where data are stored—how to be accurate without undue alarming. As pandemic restrictions ease, there will be more choices about using telepsychiatry or not. A review of the ethical challenges in conducting telepsychiatric assessments that might lead to involuntary detention in Britain urges they be avoided as much as possible, returning to previous prohibitions. Ethical issues will also depend on theoretical considerations and the practicality of maintaining confidentiality. The ethics of distributive justice have been used to show that, despite the current trend of improving quality, lowering costs, and more widespread use of electronic media, not all those in need are reached. A therapist referring patients to an online support group will face an added ethical burden to maintain contact with the group and have some sense of what is occurring, given that such groups can change rapidly. Ethical issues around identity and disclosure will continue as patients want to know what is stored and where and how such information can be accessed.

In all cases, the therapist must explain the procedure in detail, how data are stored, safeguarded, and transmitted, and the limitations and potential complications of the treatment. Informed consent must be obtained and must now include warnings that a determined hacker can get into almost any website or communication given enough time, funding, and expertise. Blockchain technology, currently being explored for medical records, may prove an exception, but only if the chain of custody as data are moved remains secure. Therefore, it is wise to record the consent both in writing and electronically.

For EMRs, the ethical dilemmas of data entry issues include entering enough to meet documentation efforts while still trying to make the record clinically useful. “After the Storm” (Siddhartha Mukherjee [MD], Topic Coronavirus Chronicles, New Yorker, May 4, 2020, p. 29) reports clinicians becoming aware of clotting as part of the COVID-19 clinical picture and finding the EMR could not be searched. Details from the clinical encounter have been used to ask for an increased payment, resulting in more details being specified to get more reimbursement, for example, Medicare chart reviews yielding an additional $6.75 billion charges, while simultaneously bloating the EHR beyond any useful clinical legibility (Billions in Estimated Medicare Payments From Chart Reviews Raise Concerns. OEI-03-17-00470, 12/10/190 (https://oig.hhs.gov/oei/reports/oei-03-17-00470.asp). EHRs have been used to perpetrate other financial frauds despite apparently involved clinicians being unaware of the fraud (Kaiser Health News, 12/23/19: “Electronic Health Records Creating A ‘New Era’ of Health Care Fraud”). The records also contain unseen risks to patient safety (KHN 3/31/19: “Death by 1,000 Clicks: Where Electronic Health Records Went Wrong”). Physicians must also be careful not to enter data that might increase the marketability of breached health care records, currently (October 1, 2021, https://www.imprivata.com/blog/healthcare-data-new-prize-hackers) up to $250/record.

Clinicians are often given templates and language that include everything necessary to get reimbursement, often stating that more was done than actually was the case, for example, a neurologic examination or detailed mental status. How much to delete becomes an ethical dilemma.

THEORETICAL ISSUES

Theoretical issues may relate to (1) the use of electronic media per se, (2) the selection of data to be recorded, (3) acceptable and reimbursable medical practice, (4) identity and the internet, (5) treatment of pathologic internet use, (6) uses of HIT, (7) privacy and security, (8) development of an interconnected information society, (9) HIT and health care change, and (10) possible therapeutic uses of games.

Electronic Media

Microprocessors are the only major commodity that continues to do more and more for less and less cost. Moore’s law dictates that the number of microprocessors that can be fit on a chip will double every 2 years has held for 50 years. As of 2019: “Moore’s Law is still valid and in good condition. It did not die, did not slow down, and is not even sick… transistors can reach to 0.1 nanometers” (https://xiaomidna.com/news/industrial-scientific/moores-law-is-still-valid-transistor-will-achieve-0-1nm). Individuals are surrounded by devices using microprocessors that can be linked by varied communication modes that have become faster and cheaper while carrying more data. Sensors in everyday devices can be hacked. In 2014, a refrigerator was sending fundamentalist Christian messages to its owners. We live in a world of interconnected smart things. Many of these devices have multiple uses, including internet access, and simultaneously converge in multiple uses while being differentiated according to user preference. Technology is evolving toward the constant use of mobile devices. Individuals often are unaware of using microprocessors and the internet as auxiliary thinking devices, but the differences in memory retrieval from one’s brain versus the internet are immense. Calculators and spreadsheets outstrip merely human efforts. People vary widely in the amount and frequency of media use for communication, but there is a general increase as communication becomes cheaper and faster. New communication formats continue to arise.

Selection of Data to be Captured

Any electronic data storage involves not capturing other data, but selection must now contend with too much available data, such as websites, social media postings, and bloated EHRs. What is significant? Conversely, anything less than in-person interaction with a patient means some data will be lost, even with the capacity to magnify images and move cameras around. Who controls the camera makes for crucial differences. Whereas some data reduction is desirable, as one's attention inevitably focuses on one thing or another, some choices may miss the key points. E-therapy captures only words without tone of voice, inflections, and the social cues that voices carry. Telephones do not capture facial expressions. Telepsychiatry may focus on the smiling face of an apparently open, cooperative person while missing the squared-off shoulders and rigid posture of controlled hostility. The use of multiple cameras and observers may increase data input, but increased equipment expenses and observer time still do not close the gap between being there and dealing with the reality of the situation. The decision to use electronic media depends on whether the advantages of such use outweigh the disadvantages of data reduction.

Acceptable and Reimbursable Medical Practice

Acceptability is determined mainly by professional associations based on evidence. The Centers for Medicare and Medicaid determine reimbursability in the United States for Medicare and Medicaid Services (CMS), setting standards most commercial insurers accept. Clinicians experience CMS reimbursement as easily obtained, albeit sometimes so low that clinicians choose not to participate, more often in Medicaid. In general, for-profit insurers pay more but may make reimbursement difficult, supposedly in the name of quality, but clinicians suspect saving money is the main motive. Reimbursement disputes often involve E/M coding and documentation thereof. EHRs may progress quickly to being acceptable, reimbursable, and required in some settings, but so far, many physicians see them as user-unfriendly and overly reliant on templates and documentation requirements for billing and reimbursement. Many high-volume, high documentation requirement settings use medical scribes, numbering more than 100,000 in 2021, estimated to be 154,000 by 2028. The use of medical scribes will be an inverse measure of EHR bloat. Psychiatrists' interest in using EMRs varies according to the practice setting, the volume of patient visits, and patient and psychiatrist concerns about privacy. Avoiding documentation, electronic data storage, and other problems cited above leads to not accepting electronic billing or taking health insurance reimbursement.

Increasing the EHR burden despite efforts to reduce it poses whether the EHR can be all things to all stakeholders. Alternatively, the clinical note would be simply for brief communication with other clinicians, not used for reimbursement. Such EHR systems in other countries draw few complaints.

Identity and the Internet

Increasingly, people prefer to work with computers and disclose more facts to impersonal electronic media than they do face to face with live interviewers. Detailed personal information often is made available to everyone on social websites such as Instagram, Facebook or Twitter (currently rebranding to “X”) by individuals who are shy and reserved in person. Several dozen studies show that patients give more detailed and accurate substance abuse histories in computer interviews than live interviewers, although feelings about the substances used are not conveyed well. This finding has led to the widespread practice of having patients fill out forms and screening instruments before seeing a physician in person or while developing PHRs. Electronic media for communication avoids dealing with reactions from real people one is facing.

Virtual reality practitioners note that each person brings their background into a virtual reality experience. They describe a “virtual presence,” which increases as patients spend more time in the virtual environment. Doing so involves giving up some sense of presence in the real world. Although VRT was initially patterned after systematic desensitization as practiced without electronic assistance, some of its practitioners now speculate that it is oriented more toward neurophysiologic information processing theory and the accelerated integrative information processing paradigm presented by Frances Shapiro (https://www.emdria.org/public-resources/emdr-therapy-beginnings-francine-shapiro/). Patients in VRT typically would not communicate back to their therapists in the physical world but instead seemed to be reliving their previous anxiety-provoking experiences. Other therapists point to the flight simulator analogy as one in which the evaluation instrument also remedies skill deficits. Political polarization and radicalization have become societal concerns while internet use has become widespread. Many seek validation rather than information, reinforcing their views rather than new opinions or data.

Treatment of Pathologic Internet Use

Online therapy or anything else online cannot deal adequately with an addiction to being online. Remedying flight from the real world for whatever reason must involve reinvolvement in the real world, including the experience of treatment, as described as follows.

Breaks from the internet are recommended as a public health mechanism, including checking one's email only a few times/day.

Uses of HIT. Physicians will have to adjust to having dynamic medical records that tap automatically into various scientific databases. Clinical decision support is available in real time as decisions are made, not in the form of excessive warnings that guidelines are being violated, but in a positive, efficient, and supportive way. Many time-consuming transactions are performed automatically. Safety and quality can be achieved with EMRs, but cost savings depend on EHRs. Research to date shows costs increasing in the United States due to EHR use. Smart cards, presently used in France and Taiwan, offer a technology that transmits data as a service is given, records the service, and updates the patient EHR. However, these systems do without RHIOs and HIEs. The record is accessible only when a patient is seen and has put the smart card in the physician's card reader. Public health and public policy will be much better informed by having almost instantaneous aggregate patient-deidentified data on patient care.

Privacy and Security. HIT proponents argue that some sacrifices must be made for the common good of informing policy, public health, and research, whereas privacy advocates argue that patients should give affirmative consent to sharing every piece of personal data. It is unclear how much these opposing views will conflict in specific cases or how conflicts will be resolved. The rejection of British government proposals by an overwhelming majority of primary care physicians and their patients suggests that HIT proponents can become overly enthusiastic and lose a sense of public preferences. Only competent auditing can help to reduce the fraud that consumes about 10% to 15% of American health care costs. It is unclear whether clinician organizations will accomplish this or leave it to the insurers with whom there is already a power struggle.

"Privacy is over," runs one argument in the debate, while others argue it can be maintained. The reality is that different data matter more or less to different people. For example, some will post their entire genomes but not access their bank accounts.

Development of an Interconnected Information Society. Extracorporeal processing has created a new culture with alternate realities and identities, affecting communication, the problems patients bring to psychiatrists, and what psychiatrists can do for them. A generation that has grown up with microprocessors and the internet has incorporated them into their way of life. They use all manner of devices with ease and comfort and are driven to take up new ones. They are more concerned with communication than privacy and spend much time in touch with each other. Indicators of time spent sitting indoors show that Americans are spending less time outdoors and are facing an epidemic of obesity. The prevailing cultural model of for-profit medical care has eroded privacy in favor of marketing. Every patient can be viewed as a potential customer. Others outside the for-profit world offer a top-down consensus that many problems in U.S. health care can be eased through the use of EHRs, especially electronic prescribing programs that check for safety against other medications and patient data. Although such efforts have sputtered in the past, the current drive generally benefits from the continuing price decline and quality improvement of electronics and better planning, funding, and standard setting.

Revelations in October 2021 that social media, especially Facebook and Twitter (X), contribute to polarization, divisiveness, anxiety, depression, misinformation, and disinformation have led to theories about how best to achieve a more harmonious interconnected society. In India, the news feed has been especially problematic (https://www.washingtonpost.com/technology/2021/10/24/india-facebook-misinformation-hate-speech/).

HIT and Health Care Change. So far, HIT use has not brought the transparency and public policy consensus required for health care reform. Before agreeing on HIT, all sides agreed that costs were high, primary care doctors were underpaid, and all citizens should have medical bills covered. However, these three and other viability indicators have worsened despite intense fact-finding and discussion. The United States has allowed health insurance profiteering, which is regarded elsewhere as being the result of having a market rather than a coherent system. Pay-for-performance programs have been initiated, offering small bonuses compared to other countries. It is essential not to make promises for HIT that will lead to disappointment. However, the extreme portability of electronic records and the global reach of telemedicine may result in an outsourcing competition that will raise a host of issues.

Possible Therapeutic Uses of Games. While internet gaming can be so intense and time-consuming as to be on the threshold of being a treatable diagnosis, some studies have shown that playing computer games is therapeutic. Test instruments, such as the Wisconsin Card Sorting test, have been turned into games, with the idea that repeated use will improve performance—an example of the assessment test becoming the rehabilitation tool. Unfortunately, results have been mixed and somewhat disappointing. However, games can be linked to neural circuitry improvements (as was shown for alcoholism in 2015 by Weirs et al.). This study shows an interaction of internal, biologic processing circuits synchronizing with external, silicon-based circuitry.

Techniques

Becoming adept in using electronic devices is eased by the awareness that some cognitive and communication functions can be delegated to microprocessors and the internet. Formerly, information was hard to obtain; now, the problem is sifting through too much of it. Former Czech President Václav Havel posed the new problem as knowing more but being less wise. Clinicians should always be aware that people using electronic media make choices in the use of words and images, often involving an individual sense of self, effectiveness, and identity that may or may not be clinically significant. Clinicians should also choose consciously what data not to capture. For example, HIPAA protects psychotherapy notes, but emails and telepsychiatry sessions may mix material entitled to such protection with other data for which less sequestration is possible. The use of electronic media is not a new form of therapy but a different way of carrying out a variety of therapies, some of which lend themselves better than others to the use of electronic media. First, however, there are some general considerations, which are described as follows.

RESEARCH AND EVALUATION

There is little funding for research on electronic media. Government and professional associations have decided to favor use, as described earlier. Equipment manufacturers are preoccupied with producing new models for the general population and remaining price competitive. There is no special equipment for anything described here. EMR vendors sponsor exhibits with little research, but government and the American Medical Records Institute provide some funding.

Problematic Internet Use

In a telephone survey of 2,513 adults, 69% were regular internet users, 6% felt that their relationships suffered as a result of excessive internet use, and 8% used the internet as a way to escape problems or relieve negative mood. Other potential markers for problematic use included preoccupation with offline internet, concealing nonessential use, failed attempts to cut back use, and frequently using the internet longer than intended. Interviews of 216 college students in Taiwan found 87 internet addicts, with associated major depressive disorder (MDD), ADHD, dysthymic disorder, and social phobia, with most addicts being male students with MDD or ADHD. Only seven men and eight women had no comorbidity.

Telepsychiatry

Twenty-eight inmates of New York prisons receiving services from their on-site treatment team and from visiting consultants in person or via telepsychiatry preferred the former for most services. None preferred telepsychiatry to face-to-face communication, but telepsychiatry was accepted and tolerated, especially for contact with experts who would not come into the prison. Diagnostic reliability using the Structured Clinical Interview for DSM Disorders (SCID) in 53 Native American veterans was comparable in person and with telepsychiatry except for past-year substance use and abuse/dependence combined. Telepsychiatric consultations and brief follow-up were rated as producing outcomes equivalent to face-to-face communication in 495 Canadian patients referred by family physicians, with a 10% cost saving. A comparative study of 119 U.S. veterans found no differences in outcome, patient satisfaction and adherence, and cost. Other studies have generated controversies about costs in rural areas. Video feedback is helpful in several studies in which it was used in the context of an ongoing face-to-face relationship.

There are no comparisons between online therapy and face-to-face communication because it is generally accepted that the former excludes body language, facial expressions, eye movement observation, tone of voice, and others. Studies confirm the commonsense view that evaluative feedback (not in therapy) is given more effectively in face-to-face encounters than by computer. Anecdotal evidence supports the widespread view that online availability and email contact can be helpful between meetings in therapeutic sessions.

Published reports on telemedicine are generally brief, emphasizing how to do it and minimizing any differences between the telemedicine experience and face-to-face communication. Evaluative data describe satisfied patients and savings on travel time and costs. Although it is understood that some data are excluded, there has been no systematic investigation of how missing data can influence the outcome. Despite more than five decades of use, a Cochrane review in 2010 (http://www.cochrane.org/CD002098/EPOC_telemedicine-versus-face-to-face-patient-care-effects-on-professional-practice-and-health-care-outcomes) comparing telemedicine with face-to-face communication found only variable and inconclusive results. Subsequent reviews have shown benefits in specific areas such as health monitoring of high-risk patients.

As with most forms of therapy and new treatments, there have been phases of zealotry, exaggeration of possible benefits, some disillusionment, and loss of interest. Annual meeting programs of the American Psychiatric Association, American Psychological Association, and World Psychiatric Association show decreasing reports on the use of electronic media from a peak in 1999, increasing during the pandemic, and less reporting as telepsychiatry and electronic communication are accepted as facts of everyday life. Those who pay for public and private care have adopted the fundamental decision to use more telepsychiatry. In addition to reimbursement through Medicare and Medicaid, the Office for the Advancement of Telehealth of the Health Resources and Services Administration (part of the Department of Health and Human Services), among many funding sources, provides grants and contracts and publishes a directory of supported projects. The WHO promotes telemedicine annually in its regional advisory committee meetings, especially for developing countries. The growth in teleconferencing reflects the evaluation of those who pay for it. Rather than comparisons with face-to-face meetings, telepsychiatry should be compared to no or inexpert service. The Journal of Telemedicine and Telecare (http://jtt.sagepub.com) was started in 1995 and is published monthly. It mainly describes the extension of telemedicine into remote settings and its use in various diagnoses, such as dementia. It includes papers on information infrastructure and the situation in developing countries but no controlled studies.

Virtual Reality Therapy. VRT has been evaluated positively versus standard desensitization for phobias, especially fear of flying, using standard outcome measures. Wiederhold and Wiederhold summarized numerous positive studies. However, no studies demonstrate its superiority, although it is likely that clinicians who invest in the equipment and master the technique will eventually attract patients who have not improved with standard desensitization. A few studies dealing with other diagnoses such as schizophrenia, body image distortion, and eating disorders include new diagnostic methodologies and showed positive outcomes persisting for up to 1 year after treatment.

Electronic Medical/Health Records. Studies of EMRs and EHRs have described the use of databases to get accurate information about patients (e.g., arrests, medications, diagnoses), including refusals and concerns (among those agreeing) of patients to have their paper records entered into electronic media. In general medicine, physicians report a decline in productivity while adapting to EMRs, followed by practicing medicine differently with higher quality and safety and noticing a decline in paperwork and wasted communication time. Only 10% of physicians using electronic records abandon them to return to paper records. Many reports of electronic records are being harvested for previously uneconomic investigations (e.g., substance abuse–related mortality in 169,051 veterans over 7 years).

In Boston, a cardiologist at Massachusetts General Hospital described their frustration on Twitter (X): “Why are nearly all notes in Epic … basically ‘useless’ to understand what’s happening to a patient during a hospital course?” Another doctor’s reply says, “Because notes are used to bill, determine the level of service, and document it rather than their intended purpose, which was to convey our observations, assessment, and plan. Our important work has been co-opted by billing.”

The promise of bringing medical record keeping into the digital age was to maintain a live record of a live patient, enabling clinicians to track patient care across hospital systems and over time. Instead, we have been saddled with systems that cut into patient care (clinicians typically spend an hour feeding documentation for every hour they spend with patients) and, often, are too fragmented to allow a patient's file to follow her from one medical center to another. The EMR, as a colleague of mine put it, is "electronic in the same way your grandfather's radio is electronic." The energized, improvisatory role of medical Twitter (X) inevitably draws attention to what our bulky, billion-dollar systems should have been providing—to the cost, in dollars and lives, of the rapid clinical learning that we have foregone.

It is hardly news that our EMR systems have failed medicine. However, an executive order from New York State, issued at the end of March by Governor Andrew Cuomo, amounted to a grim epitaph: "Health care providers are relieved of record keeping requirements to the extent necessary for health care providers to perform tasks as may be necessary to respond to the COVID-19 outbreak … Any person acting reasonably and in good faith under this provision shall be afforded absolute immunity from liability. A system designed to expedite and improve the delivery of health care was officially recognized as an obstacle."

FURTHER READINGS

Al-Rawi A. The convergence of social media and other communication technologies in the promotion of illicit and controlled drugs. *J Public Health (Oxf)*. 2022;44(1):e153–e160.

Biolcati R, Mancini G, Pupi V, Mugheddu V. Facebook addiction: onset predictors. *J Clin Med*. 2018;7(6):118.

Bisso E, Signorelli MS, Milazzo M, et al. Immersive virtual reality applications in schizophrenia spectrum therapy: a systematic review. *Int J Environ Res Public Health*. 2020;17(17):6111.

Bresnick, J. Are universal EHRs key to healthcare value, trust, and AI adoption? Published September 24, 2018. Accessed August 8, 2023. https://healthitanalytics.com/news/are-universal-ehrs-key-to-healthcare-value-trust-and-ai-adoption

Bulkes NZ, Davis K, Kay B, Riemann BC. Comparing efficacy of telehealth to in-person mental health care in intensive-treatment-seeking adults. *J Psychiat Res*. 2022;145:347–352.

Carl E, Stein AT, Levihn-Coon A, et al. Virtual reality exposure therapy for anxiety and related disorders: a meta-analysis of randomized controlled trials. *J Anxiety Disord*. 2019;61:27–36.

Dell'Osso B, Di Bernardo I, Vismara M, et al. Managing problematic usage of the internet and related disorders in an era of diagnostic transition: an updated review. *Clin Pract Epidemiol Ment Health*. 2021;17:61–74.

Friedman SH, Martinez RP. Boundaries, professionalism, and malpractice in psychiatry. *Focus (Am Psychiatr Publ)*. 2019;17:365–371.

Gerry LM. 22 mental health apps for stress, anxiety, and more. 2021, https://www.verywellmind.com/best-mental-health-apps-4588479

Hunt MG, Marx R, Lipson C, Young J. No more FOMO: limiting social media decreases loneliness and depression. *J Social Clinical Psychology.* 2018;37:751.

Jairoun AA, Al-Hemyari SS, Abdulla NM, et al. Online medication purchasing during the Covid-19 pandemic: a pilot study from the United Arab Emirates. *J Pharm Policy Pract*. 2021;14:38.

Joshi KG. Metadata, malpractice claims, and making changes to the EHR. *Current Psychiatry*. 2021;20(3):e1–e3.

Kariotis T, Ball MP, Tzovaras BG, et al. Emerging health data platforms: from individual control to collective data governance. *Data Policy*. 2020;2:e13.

Krzystanek M, Surma S, Stokrocka M, et al. Tips for effective implementation of virtual reality exposure therapy in phobias-a systematic review. *Front Psychiatry.* 2021;12:737351.

Lester M, Boateng S, Studeny J, Coustasse A. Personal health records: beneficial or burdensome for patients and healthcare providers? *Perspect Health Inf Manag*. 2016;13(Spring):1–12.

Li E, Clarke J, Neves AL, Ashrafian H, Darzi A. Electronic health records, interoperability and patient safety in health systems of high-income countries: a systematic review protocol. *BMJ Open*. 2021;11:e044941.

Lo B, Kemp J, Cullen C, Tajirian T, Jankowicz D, Strudwick G. Electronic health record-related burnout among clinicians: practical recommendations for Canadian healthcare organizations. *Healthc Q*. 2020;23(3):54–62.

Luo C, Sangerb N, Singhala N, et al. A comparison of electronically-delivered and face to face cognitive behavioural therapies in depressive disorders: a systematic review and meta-analysis. *EClinicalMedicine*. 2020:24;100442.

Madden JM, Lakoma MD, Rusinak D, Lu CY, Soumerai SB. Missing clinical and behavioral health data in a large electronic health record (EHR) system. *J Am Med Inform Assoc*. 2016;23(6):1143–1149.

Maheu MM. What is remote patient monitoring? Published July 23, 2021. https://telehealth.org/remote-patient-monitoring-2/

Morris ME. *Left to Our Own Devices*. The MIT Press; 2018.

Payne TH, Lovis C, Gutteridge C, et al. Status of health information exchange: a comparison of six countries. *J Glob Health*. 2019;9(2):020427.

Peccoralo LA, Kaplan CA, Pietrzak RH, Charney DS, Ripp JA. The impact of time spent on the electronic health record after work and of clerical work on burnout among clinical faculty. *J Am Med Inform Assoc*. 2021;28(5):938–947.

Pharmacy Reviews. Online Pharmacies Reviews & Ratings & Scam Reports. 2021 (current): http://www.pharmaciesreview.com/scam-reviews.html

Restrepo A, Scheininger T, Clucas J, et al: Problematic internet use in children and adolescents: associations with psychiatric disorders and impairment. *BMC Psychiatry*. 2020;20:252.

Rothstein MA, Talbott MK. Compelled disclosures of health records: updated estimates. *J Law Med Ethics*. 2017;45:149–155.

Ruble AE, Romanowicz M, Bhatt-Mackin S, Topor D, Murray A. Teaching the fundamentals of remote psychotherapy to psychiatry residents in the COVID-19 pandemic. *Acad Psychiatry*. 2021;45:629–635.

Shayevitz C, Breitinger S, Lerario MP, et al. Implementation of a centralized telepsychiatry consult service in a multi-hospital metropolitan health care system: challenges and opportunities. *J Acad Consult Liaison Psychiatry*. 2021;62:193–200.

Siegel C, Fisher SK. *Psychiatric Records in Mental Health Care*. Brunner Mazel; 1981.

Sujan M, Scott P, Cresswell K. Digital health and patient safety: technology is not a magic wand. *Health Informatics J*. 2019;26(4):2295–2299.

Wechsler LR, Adusumalli S, Deleener ME, Huffenberger AM, Kruse G, Hanson CW. Reflections on a health system's telemedicine marathon. *Telemed Rep*. 2020;1(1):2–7.

Yellowlees P, Shore JH. *Telepsychiatry and Health Technologies: A Guide for Mental Health Professionals*. American Psychiatric Press; 2018.

7.10 Cultural Competency in the Organization and Delivery of Psychiatric Care

CAROLE E. SIEGEL, PH.D., HELENA B. HANSEN, M.D., PH.D., AND JONATHAN MICHEL METZL, M.D., PH.D.

Cultural competency (CC) is a notion borne from considerations of nondiscrimination that has been applied in various contexts, including national and international affairs, education, the workplace, and health care. In all these venues, it speaks of productive ways to achieve equitable ends in situations where a mainstream or majority group–sometimes a power group–interacts with “other” cultural groups–usually minority or possibly disadvantaged groups–who have values and views and sometimes languages that are appreciably distinct from the “mainstream.” The specific contextual issues and the positions taken can substantially differ from venue to venue. For example, while debates can ensue as to whether or not immigrant children in the United States should be educated in their native languages and in specially designed educational courses, there is no analogous debate in health care. In the United States, Western medicine is predominantly delivered in English, but there is wide agreement that language accommodations are required for persons from cultural groups who have limited English proficiency.

Since 2014, several events of notoriety have occurred in which law enforcement agents were involved in adverse community situations, and Black lives were lost. A collective response to this which has had a substantial national impact is the “Black Lives Matter” movement which aims to eliminate the occurrence of such events and to raise public awareness of innate biases in the population and prejudicial treatment of Blacks, as these factors may have played contributory roles. Organizations both private and public used by diverse groups, including educational and medical settings, have responded and actively put in place activities to ensure that their settings are open and respectful of its diverse community members. At NYU Langone Hospital, an Office of Diversity Affairs was established within which activities to fulfill the goals of diversity and inclusion were put in place. There is a roadmap and timeline for promoting an inclusive and equitable culture of medicine. This includes hiring and placement at all institutional levels and particularly in leadership positions, a diverse faculty and staff, as well as the establishment of a greater array of diversity training activities in the School of Medicine. Medical education now includes a curriculum on systemic racism and its impact on health outcomes. Many of the activities that have been suggested are the same as those endorsed under organizational CC that aim to accommodate diversity and promote equitable care.

Here, we explicate the notion of CC, providing background on mental health disparities, some formal government responses through the years, and some proposed solutions for organizations, their programs, and their staff to improve their CC in the delivery of care. In part II, structural competency is introduced, which addresses the societal influences and social determinants of health that can cause disparities in treatment and care outcomes. Specific recommendations for clinical care that integrate social awareness of these factors are exposited.

Suggestions are made to move the discussion of CC in health care–with a focus on mental health care–beyond a rehearsal of compliance issues and the rhetoric of political correctness to a description and justification of the CC activities that can be used to promote patient-centered care. While it is critical to apply these activities in all health areas, it is particularly crucial in mental health care, where diagnoses are made and many treatments delivered strictly based on verbal communication. Operationalizing the activities of CC has remained challenging, but the landscape of the ‘80s, when these notions were first introduced, has considerably changed based on the broad recognition that for CC to move forward, greater specificity was required on how to do this. Recent societal events of negative import reinforce the need to have specific activities in place that will have an impact.

DEFINITIONS: CULTURAL GROUP; CULTURAL COMPETENCY

A cultural group has shared activities, ideas, and traditions, which are reinforced by its members. The group is a dynamic entity as the views and values of its members can change over time and are influenced by where people live as well. Thus Chinese and Chinese American cultures in 2014 differ from each other, and each group differs from itself pre- and post-Chinese economic reforms. Federal policies for health care to date have been framed to ensure nondiscrimination only for minority racial/ethnic groups, but CC is intended to apply to any cultural group that encounters limitations in service access or in participation in services. While clearly including underserved racial/ethnic groups, it also encompasses groups disadvantaged in health care such as lesbian, gay, bisexual, and transgendered communities, the born Deaf, and even rural communities, each having distinct lifestyles. Anthropologists have recently questioned the value of culture in understanding a persons' behaviors, as monolithic constructs fail to account for variations in expression and experiences within cultures. The notion of a cultural identity attempts to address this deficit by recognizing that persons belong to multiple cultural groups, and the intersectionality of these groups is what shapes a worldview. Cultural groups are defined by more than race, ethnicity, country of origin and go beyond to encompass socially defined groups for which membership impacts daily living, such as groups defined by socioeconomic status, professional roles, gender, and stage of life roles. Thus a Black, unmarried, inner-city male will differ in world view from a Black father who lives in suburbia. CC training materials for clinicians take cognizance that persons' views of their cultural identity are fundamental to understanding how they make sense of the world and hence how they respond to and interact with a health care provider.

Many definitions of CC in health care have been proposed, of which the most commonly cited in mental health care is that of Terry Cross. He defines CC as "a set of congruent behaviors, attitudes, and policies that come together in a system, agency, or among professionals and enables that system, agency, or those professionals to work effectively in crosscultural situations." This definition first appeared in a monograph on ways to work with severely and emotionally disturbed minority youth written for the Georgetown University Child Development Center (now the National Center for Cultural Competency), an early promoter of CC in the delivery of children's services. A less proscriptive definition defines Cultural Competence in an individual or organization as having the capacity to function effectively within the context of the cultural beliefs, behaviors, and needs presented by consumers and their communities. While these definitions are all-encompassing, to better understand their implications, it is helpful to parse them into the distinct ways they apply to organizations and to caregivers.

When considering healthcare organizations that broadly include structural entities such as departments of mental health, managed care companies, hospitals, agencies, and programs, CC is referential to the policies and procedures that should be endorsed and promoted throughout an organization to enable caregivers and other staff under the purview of the organization to work effectively and efficiently with different cultures and within multicultural situations to provide high-quality services. On the clinical level, CC refers to the desirable personal attributes of clinicians encompassing their attitudes (affective domain), knowledge (cognitive domain), skills, and actions (behavioral domain) that enable them to work effectively and efficiently across cultures and in multicultural situations to produce clinically and culturally appropriate service outcomes. This latter definition addresses clinician values, overt and covert biases, as well as the knowledge and skills guiding the ways to act. There has been notable clinical resistance to the concept of CC as many believe that it is already addressed or implied in clinical competency or feel it is already adequately addressed when a clinically competent clinician practices patient-centered care. The Institute of Medicine defines clinical competency as the ability to use clinical skills and knowledge to identify each consumer's unique health state and diagnosis, individual risks and benefits of potential interventions, and personal values and expectations. In working with persons with severe mental illness, some curricula endorse core competencies to address recovery, rehabilitation, and empowerment. The argument goes, if these notions are fully operant, CC will be embedded in clinical competency. But for now, CC is not treated as a distinct construct in clinical training or adequately represented in its curriculum, so that CC needs to be viewed as a distinct competency.

It is well documented that clinicians' misunderstandings of cultural sensibilities can lead to failed treatment alliances. Many psychiatric training programs around the country have instituted CC didactics and curriculum that explicate one or more of the attributes of CC identified by Derald Sue and others of attitudes, knowledge, and skills. For example, the New York University psychiatric CC curriculum has included three modules, two of which address attitudes (and awareness): Perceiving the Other and Perceiving Self. In the former, residents learn how sociocultural influences impact clinical presentations, and in the latter, residents experience their own cultural backgrounds and examine their influence on the doctor–patient relationship. Action skills are taught through a third practicum module where residents practice conceptualizing culture in the doctor–patient encounter using a culturally interpretive clinical approach. Other models have been proposed for guiding clinical competency, and one such other is discussed later.

STRUCTURAL COMPETENCY: ADDRESSING INSTITUTIONAL DRIVERS OF INEQUALITY

Jonathan Metzl and Helena Hansen have introduced a construct called structural competency that addresses the institutional and policy-level (i.e., structural) drivers of disparities in treatment and health outcomes. These drivers shape the social determinants of mental health in populations, as well as mental health care delivery systems and the practice environments of clinicians who deliver care. The premise of structural competency is that health inequity results from and reflects institutionally embedded inequalities, including systemic racism, that sicken marginalized groups of people through such social determinants as unstable housing, unsafe polluted neighborhoods, lack of access to fresh food, unemployment, substandard education and high levels of police surveillance and incarceration. Clinical practitioners are traditionally trained in individual patient-level interventions, but as health experts they can, when taught the relevant skills, work with institutional stakeholders to reorient policies and practices toward health promotion. Clinical practitioners are in a privileged position not only to redesign clinical services to address structural barriers to care (such as by providing transportation and social services) but also to collaborate with community organizations, nonhealth sector agencies such as housing assistance, education, and legal aid, as well as policymakers, to promote health through community-based programs. Structural competency frameworks also emphasize the mental health effects of particular health policies, and increasingly clinicians can help to develop and implement policy solutions to some of the most pressing social and political issues of our time, including gun violence and mass incarceration.

A growing number of medical schools are developing structural competency curricula and community practicums (see structuralcompetency.org). They provide classroom and community-engaged experiential learning based on three fundamental principles: (1) understanding patients' experiences in terms of their structural drivers; (2) intervening to address those drivers at institutional levels (through community organizations, nonhealth agencies, and public policy); and (3) practicing structural humility, which means seeking out multidisciplinary collaborators with other types of expertise, including community organizing, and having patience with the slow pace of structural change. Some psychiatry residency programs have worked with faculty members who are social scientists and community organizers to create structural competency trainings that specifically address barriers mental health.

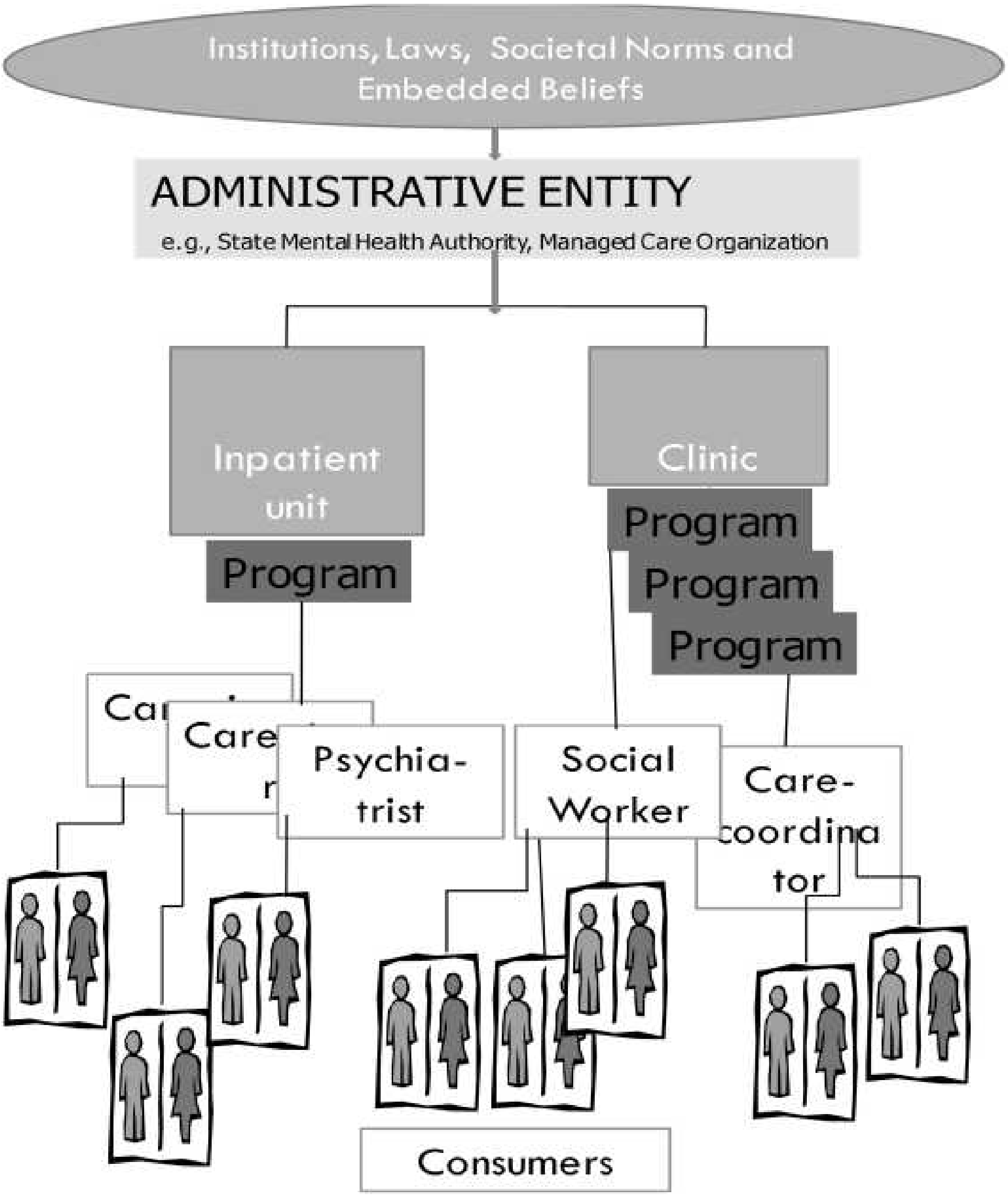
Institutions, Laws, Societal Norms and Embedded Beliefs
ADMINISTRATIVE ENTITY
e.g., State Mental Health Authority, Managed Care Organization
Inpatient unit
Clinic
Program
Program
Program
Program
Car
Care
r
Psychia-trist
Social Worker
Care-coordinator
Consumers

FIGURE 7.10–1. Hierarchical levels of health care system.

Figure 7.10–1 displays the hierarchical levels of a health care system in which CC activities need to be introduced. Structural competency is added as a meta-level to the Terry Cross view of CC as structural competency impacts operations and behaviors at all of the displayed levels.

WHY CC IS IMPORTANT: DIVERSITY, DISPARITIES

The Percent non-White of the U.S. Population is Growing

According to the most recently published U.S. Census figures (2021), 40.7% of the U.S. population is non-White. 18.9% report themselves as Hispanic/Latino, 12.6% as Black/African American, and 5.9% as Asian. The Hispanic/Latino group in the United States is even larger as many in this group have entered the country as illegal immigrants and are not counted officially in any of the U.S. Census counts. Nevertheless, many of these persons will avail themselves of emergency rooms in hospitals and use hospital clinics when confronted with mental health problems. Racial/ethnic groups that require special attention from psychiatric practitioners comprise an ever-increasing proportion of all persons served. For example, in NYS in its public mental health system (which excludes private practice practitioners), in 2013, 24.3% were Black, non-Hispanic, 23.3% were Hispanics, and 1.9% were Asian American.

Continuing Evidence of Disparities

In 1978, the first report of the President's Commission on Mental Health explicitly recognized critical disparities among racial and ethnic minorities. President Jimmy Carter and especially his wife Rosalynn became strong advocates for rectifying these disparities. A series of reports originating with the Western Interstate Commission on Higher Education commissioned by the Substance Abuse and Mental Health Services Administration focused on the four broadly defined racial/ethnic groups identified as underserved and underrepresented in managed mental health care—**Americans of African Descent, Asian and Pacific Islanders, Latinos, and Native American Populations—and compiled into a comprehensive volume a great depth and breadth of knowledge about the mental health needs of these cultural groups.** The Surgeon General's 2001 Supplemental Report reported on the mental health care of these major underserved racial/ethnic groups finding disparities in availability and access to needed services. The report endorsed CC in care delivery as a key approach for the reduction of these disparities, although there was no elaboration on how to do this.

There is strong evidence of sustained disparities in mental health. Researchers continue to find disparities for racial/ethnic minority groups in service accessibility, service utilization, as well as inpatient satisfaction. Asians remain less likely to access care than other groups. There are also still notable disparities in the reported prevalence of mental health conditions among groups, with reasons for this at best only speculative. Several authors have linked prevalence differences to a misunderstanding of cultural presentations of symptoms resulting in both over and underdiagnoses. Papers still speculate on reasons for higher rates of diagnoses of psychosis in Black adults, ADHD in Black children, and lower rates of diagnoses of depression in Latinos. Mechanisms of disparities are often linked to social determinants, for example, poverty, and to structural factors such as federal and economic health policy, operations of the health care system, provider factors, and environmental contexts. Residential segregation of services and providers and provider shortages may also account for access and use disparities. For example, for Native Americans, in 2018, it was noted that the Indian Health Service (HIS) remains understaffed and underfunded. A review of disparities across the country based on some recent studies conducted from 2010 to 2015 is presented in Table 7.10–1.

While it is hard to evaluate whether disparity rates have been increasing or decreasing over time, an effort has gone into producing services likely to eliminate disparities for cultural groups. This includes culturally adapted assessment instruments, norming these instruments for racial/ethnic groups and training psychiatrists on how to make more culturally sensitive diagnoses, adapting evidence-based services to better fit the styles and mores of cultural groups, and introducing alternate services more acceptable to certain cultures, such as those that incorporate religion and spirituality into a therapeutic approach. A recent effort to increase the success of therapeutic alliances between persons with disorders and their therapists directly encourages peers to present to their caregivers the aspects of their cultural identity that can influence their receptivity and the success of the care they are being prescribed. It is described below.

CULTURAL AWARENESS OF CAREGIVERS

Growing Recognition of the Role of Cultural Competency in Care Delivery

Some authors locate the roots of a concern with cultural awareness in clinical training in the 1950s to 1960s postwar social sciences. A group of medical sociologists and anthropologists, Talcott Parsons among them, studied the professional acculturation and socialization of doctors during clinical training. What had been taken for granted as primarily technical education was revealed to be a rich cultural milieu, replete with rituals, key symbols, rites of initiation, and social dramas, which led researchers and clinicians to a growing awareness of the “culture of medicine” as opposed to the culture of patients. There is a consensus, however, that the term “cultural competence” was not widely used in medical journals or clinical training until the 1980s when the Association of American Medical Colleges and the Accreditation Council of Graduate Medical Education called upon medical schools and residency programs to train clinicians to address their patients’ “culture,” which largely referred to their patients’ ethnic and racial identity. The term “cultural competency” and the concern with outcome disparities came from the field of education, which had used the term with regard to addressing educational attainment disparities by ethnicity and race.

In psychiatry, this awareness of cultural differences between practitioners and patients as relevant to clinical practice was codified in the DSM IV, published in 1999. Psychiatrist-anthropologist Arthur Kleinman led a special committee of the DSM IV Task Force devoted to cross-cultural psychiatry that introduced new diagnostic frames and clinical tools to the DSM IV appendix. These included “culture-bound syndromes” and the Cultural Formulation, a guide to interviewing and diagnosing patients using questions that took the patients’ cultural understanding of their problem into account, using Kleinman’s Explanatory Model. This contrasted with the symptom checklists of the DSM III, which had aspired to use universally applicable symptom descriptions, descriptions that soon proved to be interpreted differently across ethnic and language groups, and which some practitioners argued led to misdiagnosis. At about the same time, Derald and Stanley Sue were advocating for the cultural competence of clinicians, particularly in their work in counseling and psychotherapy. They emphasized the need for a greater understanding of psychiatric symptoms that could be misunderstood across the cultural divide between Western medicine and Chinese patients. They emphasized the need for addressing cultural beliefs, attitudes, knowledge, and skills and developing new counseling competency standards when working with multicultural populations.

By 2000, as the DSM IV Cultural Formulation gained visibility, and as increasing numbers of medical schools adopted some form of training in cross-cultural patient care, competing models of training in cultural issues emerged. The first might be called cultural expertise, in reference to the goal of having medical trainees gain knowledge of the beliefs and behaviors of their patients' cultural groups. In practical terms, this led medical schools to emphasize language translation services and, in some cases, "cultural consultants" who are members of minority ethnic groups employed by medical centers to assist healthcare providers in communicating with their patients. The second approach might be called cultural sensitivity, in which the goal is not to master bodies of knowledge about the cultural groups from which patients come, but rather to understand the broader psychosocial context of each patient, to ask each patient's guidance in learning what is relevant to them, and to heighten the cultural self-awareness of clinical practitioners so that they can take their own biases into account in their exchanges with patients. These two approaches, cultural expertise and cultural sensitivity, are not mutually exclusive, but many medical training programs emphasize only one.

Expanding on the cultural sensitivity approach, the term "Cultural Humility" was introduced by pediatricians, who offered "humility" instead of "competency" because it conveys the attitude a doctor would need to adopt in order to learn from the unique circumstances and thinking of each patient. The practical elements they proposed were (1) self-evaluation of the health care provider to root out cultural assumptions that might reinforce their stereotypes of and discrimination against patients, (2) empowering patients to have a voice in their own decision making; and (3) forming partnerships with communities to advocate for more locally responsive health care.

Table 7.10–1.
Selected Disparity Literature Review: 2010 to 2014

Authors (year)	Title	Objective	Sample	Conclusion
Service Accessibility and Utilization				
Wu P et al. (2010)	Mental health service use among suicidal adolescents: findings from a U.S. National Community Survey	Assess patterns of mental health service use among adolescents who had attempted suicide and examine service at individual and community level	877 adolescents (12–17) who attempted suicide in the past 12 months	▶ Mental health needs of minority, lower-income families are unmet ▶ Larger race/ethnic disparities in use of in and outpatient services than in school-based services ▶ School-based services more accessible to suicidal adolescents who may experience barriers in receiving standard care
Alegria M et al. (2011)	Disparities in treatment for substance use disorders (SUD) and co-occurring disorders for ethnic/racial minority youth	Review literature on racial/ethnic disparities, focusing on SUD services and comorbid behavioral health disorders in children and adolescents	N/A	▶ Compared with non-Hispanic-Whites, Black adolescents with SUD reported receiving less specialty and informal care ▶ Latinos reported receiving less informal care ▶ Mechanisms of disparities identified in federal and economic health policy, operations of health care system, provider factors, environmental context, patient factors, etc.
Cummings J et al. (2011)	Racial/Ethnic differences in mental health services use among adolescents with major depression	Examines difference in mental health services use in NH-Whites, Black, Hispanic and Asian adolescents experiencing major depression	7,704 adolescents (12–17) with major depression in past year–nationally representative	▶ Black Hispanic and Asian adolescents significantly less likely than non-Hispanic Whites to receive prescriptions medication, to receive treatment for depression from specialist or provider, and receive outpatient treatment, after adjusting for family income and insurance status
Akincigil et al. (2012)	Racial and ethnic disparities in depression care in community-dwelling older adults in the United States	Investigated racial/ethnic disparities in diagnosis and treatment among community-dwelling older adults	Medicare Current Beneficiary Survey, n = 33,708	▶ Blacks significantly less likely than non-Hispanic Whites to receive depression diagnosis from provider and if diagnosed less likely to be treated
Cook B et al. (2013)	Assessing racial/ethnic disparities in treatment across episodes of mental health care	Investigating race and ethnic disparities in decisions to start and terminate care and decisions during treatment	Whites, Blacks, and Latinos with probable mental illness	▶ Blacks and Latinos initiated care less frequently ▶ Black and Latino episodes were shorter ▶ Black and Latino had fewer psychotropic drug fills
Kim G et al. (2011)	Limited English proficiency (LEP) as a barrier to mental health service use: a study of Latino and Asian immigrants with psychiatric disorders	Examine the effect of LEP on mental health service among immigrant Latino and Asian adults with psychiatric disorders	372 immigrant adults (18–86) with any instrument-determined mood, anxiety, and SUD	▶ LEP of Latinos but not of Asians significantly reduced odds of mental health service use ▶ Asians less likely to access care
Jimenez D et al. (2012)	Disparities in mental health service use of racial and ethnic minority older adults	Apply the IOM definition of health disparities, to measure disparities in different episodes of mental health care	1,658 adults >60 with likely mental health care need	▶ Treatment initiation is significantly lower and the adequacy of treatment is significantly poorer for Black and Latino older adults than White counterparts ▶ Latinos had significantly longer treatment duration, more visits, and higher expenditures
Merikangas K et al. (2012)	Service utilization for lifetime mental disorders in U.S. adolescents: results of the National Comorbidity Survey–Adolescent Supplement (NCS-A)	Examine rates and sociodemographic correlates of lifetime mental health service use by severity, type, and number of mental health disorders	6,483 nationally representative adolescents (13–18)	▶ Black and Non-Hispanic Black were less likely to receive mood and anxiety disorder services than White counterparts
Siegel C et al. (2013)	Access to and use of noninpatient services in New York state among racial-ethnic groups	Report on differences in use of noninpatient services in Hispanics, Asians, and non-Hispanic Blacks and Whites using NYS public mental health system	NYS OMH 2009 Patient Characteristic Survey	▶ Higher treated prevalence for Latinos and Blacks compared to Whites ▶ Lower prevalence rates for Asians ▶ Lower treatment intensity among Black youth with disruptive disorders, Hispanic adults with anxiety, and Asians <65 with depression compared to White counterparts ▶ In upstate metropolitan areas, Black youth and Hispanic adults had lower treatment intensity than Whites ▶ In NYC, Whites >65 had higher treatment intensity than other groups

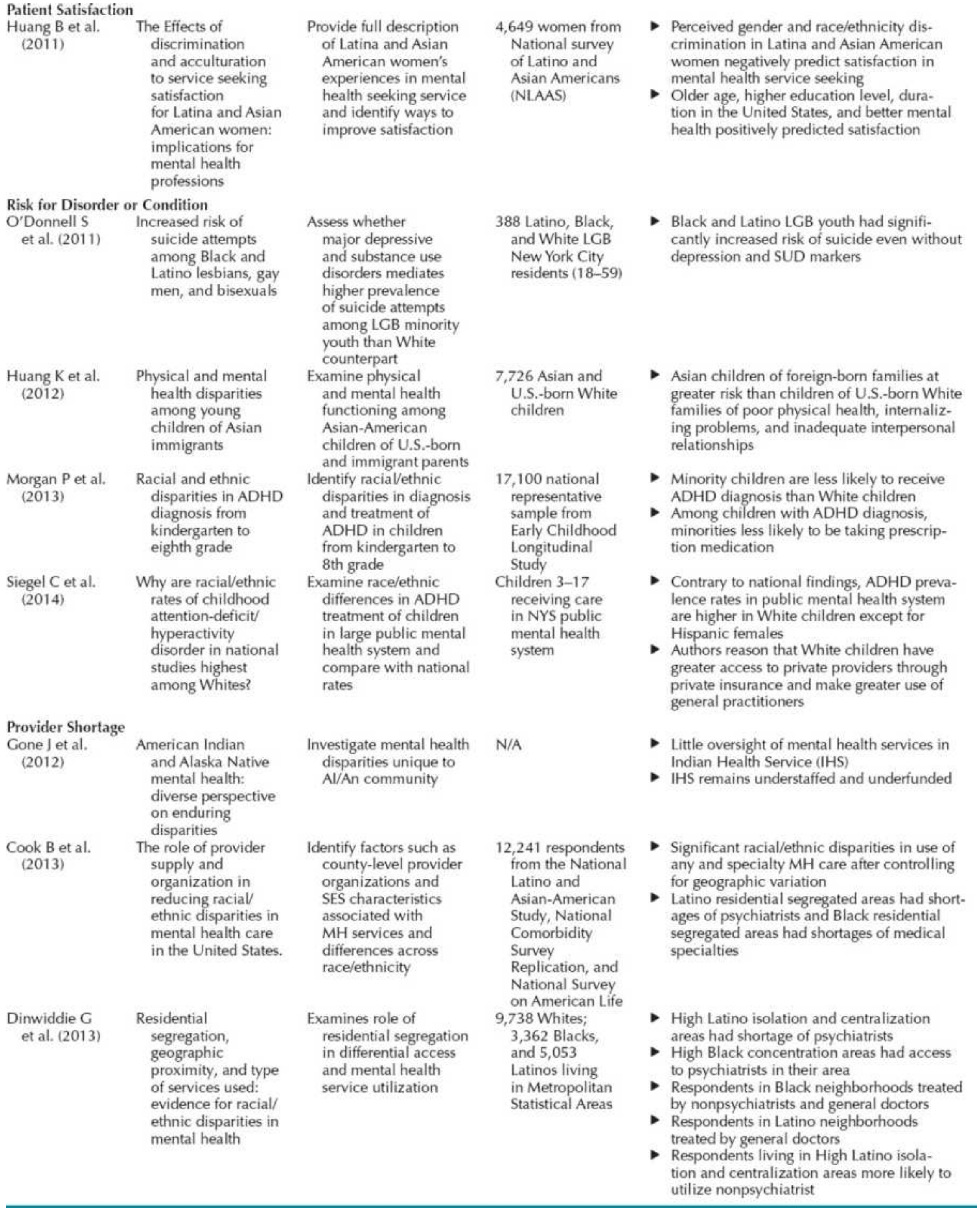

Patient Satisfaction				
Huang B et al. (2011)	The Effects of discrimination and acculturation to service seeking satisfaction for Latina and Asian American women: implications for mental health professions	Provide full description of Latina and Asian American women's experiences in mental health seeking service and identify ways to improve satisfaction	4,649 women from National survey of Latino and Asian Americans (NLAAS)	▶ Perceived gender and race/ethnicity discrimination in Latina and Asian American women negatively predict satisfaction in mental health service seeking ▶ Older age, higher education level, duration in the United States, and better mental health positively predicted satisfaction
Risk for Disorder or Condition				
O'Donnell S et al. (2011)	Increased risk of suicide attempts among Black and Latino lesbians, gay men, and bisexuals	Assess whether major depressive and substance use disorders mediates higher prevalence of suicide attempts among LGB minority youth than White counterpart	388 Latino, Black, and White LGB New York City residents (18–59)	▶ Black and Latino LGB youth had significantly increased risk of suicide even without depression and SUD markers
Huang K et al. (2012)	Physical and mental health disparities among young children of Asian immigrants	Examine physical and mental health functioning among Asian-American children of U.S.-born and immigrant parents	7,726 Asian and U.S.-born White children	▶ Asian children of foreign-born families at greater risk than children of U.S.-born White families of poor physical health, internalizing problems, and inadequate interpersonal relationships
Morgan P et al. (2013)	Racial and ethnic disparities in ADHD diagnosis from kindergarten to eighth grade	Identify racial/ethnic disparities in diagnosis and treatment of ADHD in children from kindergarten to 8th grade	17,100 national representative sample from Early Childhood Longitudinal Study	▶ Minority children are less likely to receive ADHD diagnosis than White children ▶ Among children with ADHD diagnosis, minorities less likely to be taking prescription medication
Siegel C et al. (2014)	Why are racial/ethnic rates of childhood attention-deficit/hyperactivity disorder in national studies highest among Whites?	Examine race/ethnic differences in ADHD treatment of children in large public mental health system and compare with national rates	Children 3–17 receiving care in NYS public mental health system	▶ Contrary to national findings, ADHD prevalence rates in public mental health system are higher in White children except for Hispanic females ▶ Authors reason that White children have greater access to private providers through private insurance and make greater use of general practitioners
Provider Shortage				
Gone J et al. (2012)	American Indian and Alaska Native mental health: diverse perspective on enduring disparities	Investigate mental health disparities unique to AI/An community	N/A	▶ Little oversight of mental health services in Indian Health Service (IHS) ▶ IHS remains understaffed and underfunded
Cook B et al. (2013)	The role of provider supply and organization in reducing racial/ethnic disparities in mental health care in the United States.	Identify factors such as county-level provider organizations and SES characteristics associated with MH services and differences across race/ethnicity	12,241 respondents from the National Latino and Asian-American Study, National Comorbidity Survey Replication, and National Survey on American Life	▶ Significant racial/ethnic disparities in use of any and specialty MH care after controlling for geographic variation ▶ Latino residential segregated areas had shortages of psychiatrists and Black residential segregated areas had shortages of medical specialties
Dinwiddie G et al. (2013)	Residential segregation, geographic proximity, and type of services used: evidence for racial/ethnic disparities in mental health	Examines role of residential segregation in differential access and mental health service utilization	9,738 Whites; 3,362 Blacks, and 5,053 Latinos living in Metropolitan Statistical Areas	▶ High Latino isolation and centralization areas had shortage of psychiatrists ▶ High Black concentration areas had access to psychiatrists in their area ▶ Respondents in Black neighborhoods treated by nonpsychiatrists and general doctors ▶ Respondents in Latino neighborhoods treated by general doctors ▶ Respondents living in High Latino isolation and centralization areas more likely to utilize nonpsychiatrist

Another expansion on the idea of provider self-evaluation was introduced by psychiatrist-social scientists Elizabeth Bromley and Joel Breslow in an approach to clinical education that develops "Critical Thinking Skills." They propose heightening trainees' self-awareness of the culture of medicine by teaching them the history of medicine and sociocultural research on medicine. They argue that this prepares trainees for leadership in medicine through understanding their own role in the larger institutions that are constraining patients' care and options. For example, trainees learn from this material that the U.S. health care system locates disease in the individual, rather than the community or the environment, which explains its focus on "compliance/adherence" rather than on the patient's milieu, their resources, and supports. Another example, clinicians are trained to tell diabetics to limit their intake of sugar and processed foods, but not to ask if they live near grocery stores with fresh foods or if they know how to cook.

Cultural Activation of Consumers to Provide Cultural Information to Their Caregivers

Carole Siegel, Lenora Reid-Rose, and Gary Haugland have posited that since the health activation of consumers has been documented to improve health outcomes, "cultural activation" of consumers might do so as well. Cultural activation was framed as a consumer's recognition of the importance of providing cultural information to providers about their cultural affiliations, challenges to receiving care, views about and attitudes toward behavioral health and general medical health care, and the ability to transmit such information. In conjunction with several consumer groups, they developed prompts for consumers to use to help them culturally represent themselves, thus promoting shared decision making as part of the therapeutic alliance and acting as a form of consumer empowerment (Table 7.10–2).

The prompts help consumers transmit to providers cultural information about their identities, their views of their mental health challenges, and the supports and barriers that they may encounter. They draw on the domains covered in the *DSM-5* Cultural Formulation Interview and on the "three C's" of Kleinman's explanatory model of illness: call, cause, and course. YouTube videos made by the Substance Abuse Mental Health Administration, are available that explain and support the use of these prompts. It is not intended to replace the Cultural Formulation Interview but rather to serve as a portable aid for consumers to transmit culturally relevant information to caregivers whenever the consumer feels it is needed to do so.

GOVERNMENT RESPONSE: STANDARDS, REGULATIONS, AND LEGAL REQUIREMENTS

The civil rights movement of the sixties heightened concerns about discrimination in health care. Title VI of the Civil Rights Act of 1964 states that “no person in the United States shall, on the grounds of race, color or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance.” A primary focus at the federal level has been to remove language barriers that can prevent access to both general and mental health care. The Americans with Disabilities Act of 1990 mandated accessibility to facilities and programs for persons with disabilities, and this included psychiatric conditions. It also provided an impetus for introducing principles of CC into federally funded mental health programs. In 2000, President Bill Clinton signed Executive Order 13166 “Improving Access to Services for Persons with Limited English Proficiency,” which requires federal agencies and those receiving federal funding to identify service needs of persons with LEP and implement systems to ensure persons have meaningful access to services without unduly burdening the mission of the agency. Guidance issued by the Department of Justice spells out compliance standards to be followed to ensure programs and activities provided in English are accessible to persons with LEP. Ensuring that language is no barrier to health care, this action addresses discrimination on the basis of national origin, which was prohibited by Title VI. In 2000, the U.S. Department of Health and Human Services, through the Office of Minority Health, published National Standards for Culturally and Linguistically Appropriate Services in Health Care referred to as the CLAS standards. However, while provisions for culturally competent care and administrative and organizational supports were encouraged, only the provisions for linguistic assistance were made a mandated requirement. Neither the metrics to measure compliance to these standards nor enforcement guidelines accompanied the standards, although these came at a later date. The CLAS standards were revised in 2013 and augmented with an extensive set of guidelines for an organization to employ these standards (see Table 7.10–3). The new CLAS standards are organized into overarching themes. Communication and language access; engagement, and continuous improvement and accountability encompass the original standards, and a new standard, governance, leadership, and workforce, is included to ensure that CLAS standards are integrated and implemented throughout an organization.

Table 7.10–2.
Cultural Activation Prompts for Consumers

Cultural Activation Prompt	Explanation
1. How would you like to be called 2. The kind of person you would feel good about getting care from 3. Your cultural identity 4. Any negative or traumatic experiences you have had because of your cultural identity.	These first prompts are suggested to be used when a consumer and provider first meet and get to know one another. These prompts help create the foundation of a respectful, honest, and open relationship. These set the stage for future conversations about cultural identity. They give providers a baseline understanding of how consumers view themselves from a cultural perspective. The fourth prompt can begin a discussion on how the consumers may have experienced events in relation to those outside their cultural identity groups.
5. The names you and people from your cultural groups give for why you are seeking care. 6. What you and people from your	Prompts 5–8 are also suggested to be used when a consumer and provider first meet and get to know one another. These prompts give consumers a way to articulate the importance of their cultural norms. Finally, they can begin a conversation to help clarify the nature of their challenges in a cultural context.

cultural groups think may be the cause(s) of your health challenges	
7. Why people from your cultural groups think you need care	
8. What people from your cultural groups feel about you seeking care	
9. Who from your cultural groups might help you while you are in care	Prompts 9–11 are suggested to be used when a care plan is being made. The prompts help consumers articulate their preferences from a strengths perspective. They can be used to identify social supports that may not typically be considered.
10. Kinds of different care you have had that were helpful	
11. Care that you have had that	

	was not helpful	
12.	Things you like to do where you live, worship, shop, socialize, etc. with people from your cultural groups	This final prompt is suggested to be used when planning for recovery in the community. This prompt helps identify activities that relate to their cultural identity that can help consumers maintain good outcomes while tapping on cultural community strengths. Additionally, it can begin or continue a conversation on building additional social supports in the community.

The CLAS standards have been tied to the review criteria of the Joint Commission (formerly Joint Commission Accreditation of Hospitals/Healthcare Organizations) for the accreditation of hospitals and other health care organizations that have been expanded to include more explicit cultural and linguistic elements. The Joint Commission is essentially the gatekeeper for hospitals to obtain federal funds for inpatient care and has focused its reviews on an organization’s safety and quality of care standards. Accreditation of hospitals and medical schools requires adherence to these standards. Additionally, the National Committee for Quality Assurance has standards for cultural competence that health care organizations and managed care plans must address as part of their accreditation process.

Some recent government-mandated changes in insurance coverage for health care and the way in which individual insurance plans pay for mental health services need to be considered to examine whether they have influenced the receipt of mental health care by racial/ethnic groups. The Affordable Care Act, passed in 2010, has been shown to have increased access to care for lower socioeconomic groups, which include many racial/ethnic persons most positively affecting those living in states in which there was an expansion for Medicaid recipients who could enroll because of more liberal criteria. Effects on outcome have not yet been well studied, particularly on racial/ethnic groups, and outcome evidence that exists is mixed to date, but access has been documented to have improved. The Mental Health Parity and Addiction Equity Act of 2008 require large-group employer insurance plans to cover mental health services at the same level as medical and surgical services if they offered them in order to make mental health service payments more equitable across illness categories; the effects of this on racial/ethnic minorities would thus only apply to those in the workforce and has not yet been studied.

BEYOND THE RHETORIC: DEFINING THE ACTUAL ACTIVITIES OF CC

On a clinical level, there is skepticism as to whether or not CC can be defined precisely enough to operationalize its aspects into clinical training and best practices. On an organizational level, many feel that CC is achieved once language accommodations are made, such as having interpreters available when needed and having key forms in translated versions. Hospital signs in multiple languages are now commonplace. While certainly necessary, these steps are not sufficient. But substantial progress has been made in isolating specific activities and defining and developing tools and products to promote activities of CC. We begin by making clear that CC activities need to occur on the multiple levels of a care organization in order to achieve quality outcomes.

Table 7.10–3.

National Standards for Culturally and Linguistically Appropriate Services in Health Care (HHS, Office of Minority Health, 2000)

Principal Standard:

1. Provide effective, equitable, understandable, and respectful quality care and services that are responsive to diverse cultural health beliefs and practices, preferred languages, health literacy, and other communication needs.

Governance, Leadership, and Workforce:

2. Advance and sustain organizational governance and leadership that promotes CLAS and health equity through policy, practices, and allocated resources.
3. Recruit, promote, and support culturally and linguistically diverse governance, leadership, and workforce that are responsive to the population in the service area.
4. Educate and train governance, leadership, and workforce in culturally and linguistically appropriate policies and practices on an ongoing basis.

Communication and Language Assistance:

5. Offer language assistance to individuals who have limited English proficiency and/or other communication needs, at no cost to them, to facilitate timely access to all health care and services.
6. Inform all individuals of the availability of language assistance services clearly and in their preferred language, verbally and in writing.
7. Ensure the competence of individuals providing language assistance, recognizing that the use of untrained individuals and/or minors as interpreters should be avoided.
8. Provide easy-to-understand print and multimedia materials and signage in the languages commonly used by the populations in the service area.

Engagement, Continuous Improvement, and Accountability:

9. Establish culturally and linguistically appropriate goals, policies, and management accountability, and infuse them throughout the organization’s planning and operations.
10. Conduct ongoing assessments of the organization’s CLAS-related activities and integrate CLAS-related measures into measurement

and continuous quality improvement activities.

11. Collect and maintain accurate and reliable demographic data to monitor and evaluate the impact of CLAS on health equity and outcomes and to inform service delivery.
12. Conduct regular assessments of community health assets and needs and use the results to plan and implement services that respond to the cultural and linguistic diversity of populations in the service area.
13. Partner with the community to design, implement, and evaluate policies, practices, and services to ensure cultural and linguistic appropriateness.
14. Create conflict and grievance resolution processes that are culturally and linguistically appropriate to identify, prevent, and resolve conflicts or complaints.
15. Communicate the organization's progress in implementing and sustaining CLAS to all stakeholders, constituents, and the general public.

From U.S. Department of Health & Human Services. Office of Minority Health. National CLAS Standards. https://thinkculturalhealth.hhs.gov/clas/standards

Figure 7.10–2 indicates that the policies and procedures of an organization impact the kind of staff employed and the limits placed on their therapeutic work. The composition of the clinical team, the kind of supervision that team members receive, and the approaches taken to coordinate care define the human resource boundaries of the program. The program's work is conducted within a palpable milieu that patients encounter and react to. Clinical work uses particular clinical tools and service delivery modalities that are constrained by reimbursement issues that impact service choices and the duration of delivery. The care that is ultimately delivered by clinicians is subject to their competencies.

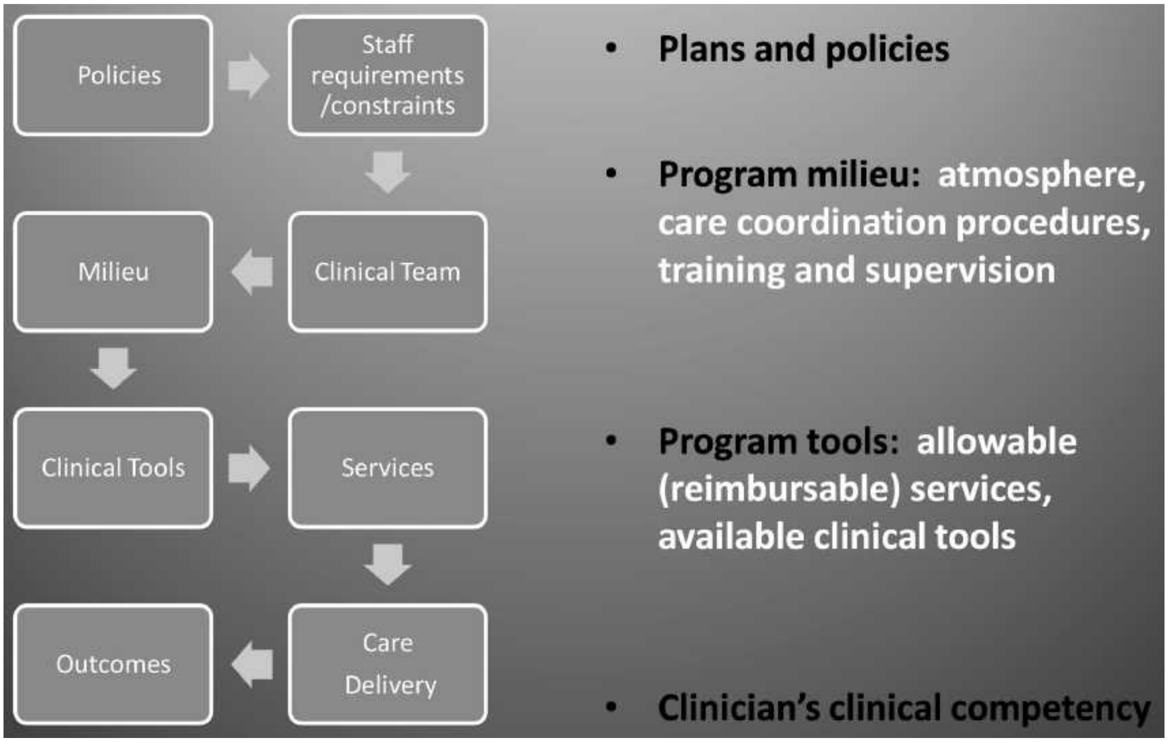

- **Plans and policies**
- **Program milieu:** **atmosphere, care coordination procedures, training and supervision**
- **Program tools:** **allowable (reimbursable) services, available clinical tools**
- **Clinician's clinical competency**

FIGURE 7.10–2. Factors in a care organization impacting care outcomes.

Figure 7.10–3 illustrates how CC needs to guide these activities. The policies of an organization that works with multicultural populations need to address CC deliberately. The starting point is the incorporation of CC into a mission statement of the organization, allocation by the organization of a budget to enable a person/group to be assigned to spearhead CC, and the development of an organizational plan for CC. The plan requires community, provider, and client input. The desirability of staff diversity needs to be addressed in hiring policies, and training activities need to be organized for all levels of the organization. The clinical team needs to address cultural strengths and challenges when managing a client's care and coordinating services. The milieu of the program needs to be welcoming to cultural groups and accommodate their special needs, for example, by having convenient hours of operation, policies on privacy, and child-friendly areas. When delivering services, cultural assessments need to be made in order to appropriately diagnose the client and develop a service plan that will act to engage and retain the client in services. Services need to be chosen that are culturally acceptable to a client, and if sufficient numbers warrant it, services specifically chosen or adapted for cultural groups. In service delivery, clinical competency needs to be augmented with a requirement for CC as the behaviors of clinicians, based on their attitudes and skills, can impact outcomes.

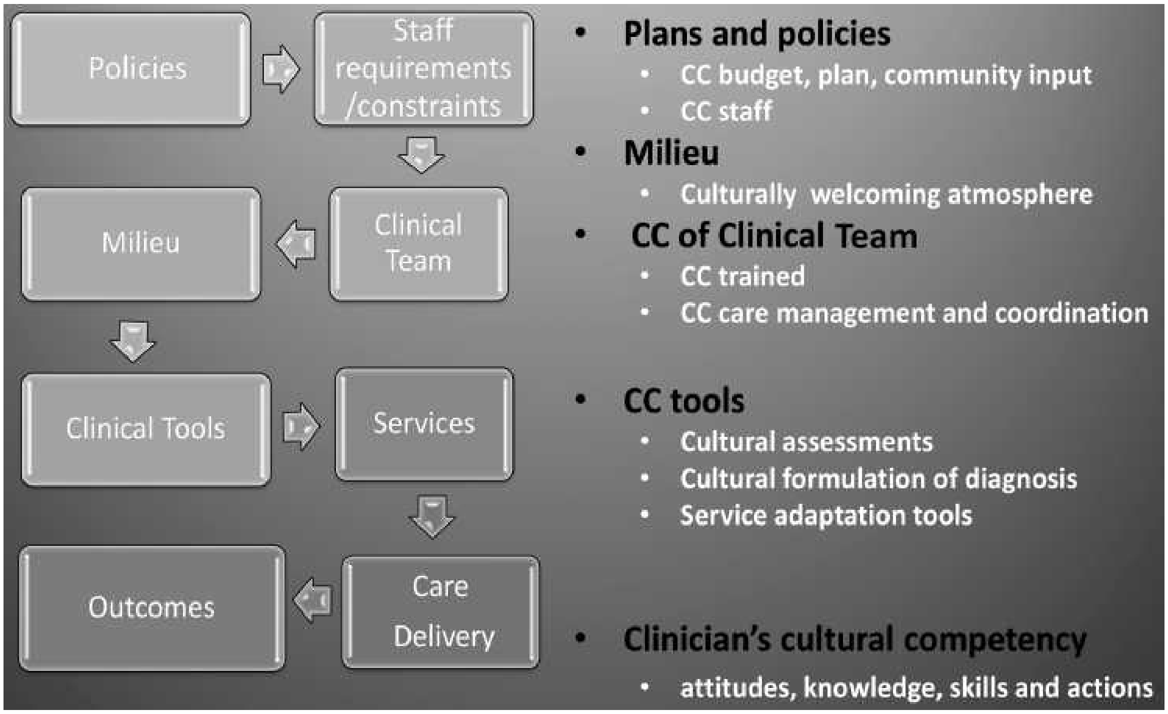
Policies
Staff requirements /constraints
Milieu
Clinical Team
Clinical Tools
Services
Outcomes
Care Delivery
• Plans and policies
• CC budget, plan, community input
• CC staff
• Milieu
• Culturally welcoming atmosphere
• CC of Clinical Team
• CC trained
• CC care management and coordination
• CC tools
• Cultural assessments
• Cultural formulation of diagnosis
• Service adaptation tools
• Clinician's cultural competency
• attitudes, knowledge, skills and actions

FIGURE 7.10–3. Cultural competency activities impacting care outcomes.

Specific Clinical Activities

Psychiatrists play a key role in the engagement and retention of persons from cultural groups in care. They often are the gatekeepers for entry into specific types of services, determine specific treatments and serve as arbiters in treatment team meetings. In the clinical hierarchy, they have a distinct power role, thus impacting other clinicians as well as clients. CC is, however, not a concept discussed in most training programs and not yet a focus of clinical care, particularly in programs that strongly endorse medical models as explanatory of illness and symptoms and medications as the primary modalities of psychiatric care. Being clinically competent and practicing patient-centered care is endorsed and could include CC, but to date is only addressed in a hand full of training programs. But buy-in to CC by psychiatrists is critical to good care, as, in order to engage persons from cultural groups in care and to have them remain in care, they need to trust their psychiatrists. Lack of trust and feelings of stigma can impact engagement in services and adherence to treatment regimens. Trust can only be engendered if clinicians are cognizant of cultural views on mental illness that can impact treatment.

Cultural Competency Curricula for Clinicians

There are several curricula that address ways to enhance the CC of psychiatrists, counseling practitioners, psychologists, and social workers. A comprehensive review of such training guides has been conducted by researchers at the Canadian Centre for Addiction and Mental Health Centre de Toxicomanie et di Santé Mentale. They found consistency across guides in the use of the Sue and colleagues framework of CC encompassing personal level awareness, knowledge, and skills. Awareness modules address understanding and conciliating a client's and one's own biases, beliefs, attitudes, and world views. Antiracist and antioppression conceptual models may be invoked. Knowledge attainment addresses understanding both the generic cultural factors that can impact care, for example, social determinants of illness, level of acculturation, immigration status, as well as specific cultural factors such as a cultural view of illness causation and expectations for recovery. Skill component modules in the more practical guides suggest approaches for engagement and problem-solving based on cultural views of clients, as well as on the tangibles of working with interpreters and seeking consultation from clients' natural caregivers such as clergy and traditional healers. Case studies and scenarios are often introduced here. In addition, several guides also address clinician-client power relationships and even system–client interactions. In the former, sensitivity to issues of power differentials that can engender either trust or mistrust are addressed. In the latter, the curricula cover areas overlapping with recent ideas on structural factors impacting relationships.

We review in greater detail one curriculum developed by psychologists Deborah Altschul and Jeff Brooks-Harris for the broad swath of counseling practitioners. The multicultural skills training model (MST) is a practice-based method of CC that aims to enable providers to adapt their current practices to meet the needs of culturally diverse populations. It is applicable across a broad array of service modalities, including psychotherapy, psychoeducation, and skills training. It also draws upon the Sue and Sue tripartite seminal model of training offering practical approaches for understanding, in addition to the cultural views and biases of clients, one's own views and suggests ways to conciliate clinician-patient views. It addresses the need for practitioners to integrate into their therapeutic approach ways to empower clients to deal with the high likelihood of impacting social determinants of health such as housing needs.

MST is arranged as six clusters of culturally based activities to help better match the individuals' needs to therapeutic care. To illustrate, the module Becoming Aware of Your Own Multiple Worldviews, a prerequisite to working with multicultural clients, recommends confronting your own personal culture-based worldview as it influences views of mental health and the etiology of mental illness and appreciating your own multiple cultural identities and their impact on your thinking and treatment decisions. When working with clients, Creating a Culturally-Sensitive Working Relationship includes knowing the cultural norms and customs of a cultural group as they manifest in a particular individual and determining the style of collaboration that is most comfortable for the consumer. Understanding the Individual's Worldviews and Facilitating Cultural Identity Development parallels the work of the first caregiver module for the client and suggests exploring with clients the historical backgrounds of their cultures, as well as the current placement of the client in that culture. There is a focus on understanding cultural norms about help-seeking, problem attribution, and expected outcomes. On the client's part, it is helping the client to value their own cultural background but also to become open to other worldviews as well. Open discourse is suggested to help clients deal with experiences that they may have had that they perceived as racist or biased in social and work interchanges. In Responding to Societal Structures and Value, counselors are encouraged to work with consumers to confront their feelings of being undervalued because of cultural identity and social status. This could be based on having received slurs, felt racial discrimination, or having felt undervalued as transmitted to others in ambient societal messages. Empowering consumers through social activation and encouraging them to make greater use of cultural resources such as religious affiliations is recommended. A final cluster Utilizing Complementary Roles beyond Psychotherapy suggests ways to incorporate nontraditional healing methods of their cultures into the therapeutic work.

Clinician Self-Assessment of Cultural Competency

There are several practitioner-level CC self-assessment scales. While self-assessments can be pointless because of their subjective structure, if appropriately used, they can serve further as educational tools. One with established psychometrics that is widely used is the Multicultural Counseling Inventory. It addresses the three domains of awareness, knowledge, and skills. It is comprised of 40 items on which the provider rates his/her self-perceived cultural competence across the subscales Multicultural Counseling Skills (Sample Item: “When working with minority clients, I am able to quickly recognize and recover from cultural mistakes or misunderstandings”), Multicultural Awareness (Sample Item: “I am involved in advocacy efforts against institutional barriers in mental health services for minority clients”), Multicultural Counseling Relationship (Sample Item: “When working with minority clients, I find that differences between my worldviews and those of the clients impede the counseling process”), and Multicultural Counseling Knowledge (Sample Item: “When working with minority clients, I make referrals or seek consultations based on the client’s minority identity development).

Domains of Organizational and Program Activities

Except for that provided by psychiatrists or other professionals working entirely independently, most mental health care is organized as programs, bundles of available services, and treatment strategies geared toward ameliorating psychiatric distress. Programs, in turn, are components of mental health care agencies, centrally administered entities that provide policy and procedural guidelines for the programs under their purview and obtain and allocate resources. From an organizational or program perspective, CC activities should take place in the domains illustrated in Figure 7.10–4. Needs assessments refer to gaining an understanding of the diverse cultural groups using a care system, their possible service gaps, and specific service needs. Information exchange refers to mechanisms for receiving and transmitting information on community needs to persons responsible for planning and delivering services. Reflexively, deliberate efforts are required for communities to receive information on services available and ways to access these services. CC tools need to be disseminated to caregivers to improve their ability to provide CC services. Services that are offered need to be appropriate for community cultural groups and modifications to existing services made and/or new services introduced. Language accommodations may need to be made. Organizations need to promote and sustain recruitment, hiring, and retention procedures of culturally competent bilingual and bicultural staff, as well as to educate and train existing staff in CC. Linkages to other systems need to take into account the cultural needs of clients. Outcomes related to disparities in access and quality of care to cultural groups, desirable cultural group outcomes, and satisfaction with services need to be distinctly monitored. These activities are expected to improve outcomes and need to be monitored to see if they are taking place and the impact they may have. They need to be endorsed through the policies and procedures of the organization.

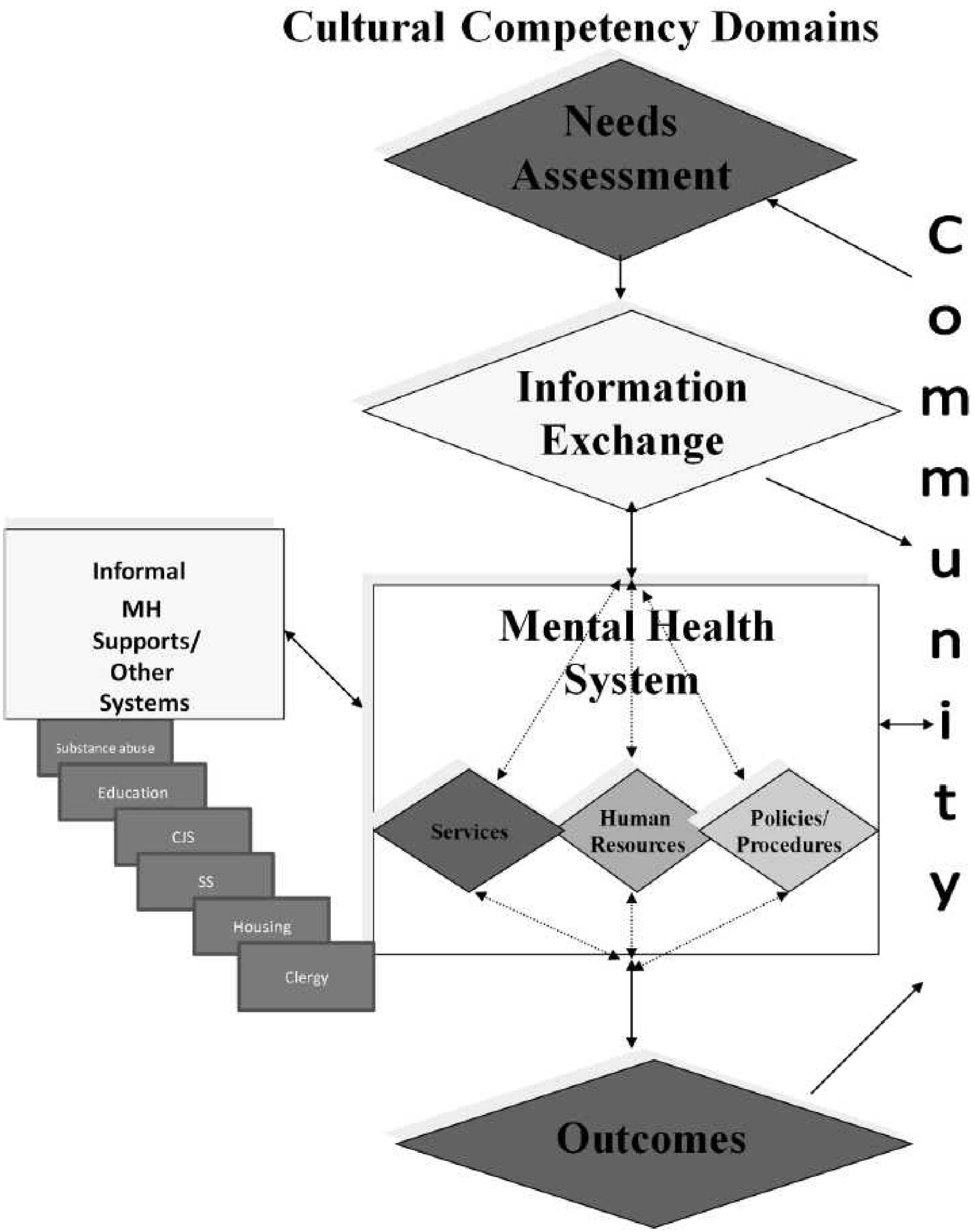
Cultural Competency Domains
Needs Assessment
Information Exchange
Mental Health System
Services
Human Resources
Policies/ Procedures
Outcomes
Informal MH Supports/ Other Systems
Substance abuse
Education
CJS
SS
Housing
Clergy
Community

FIGURE 7.10–4. Cultural competency domains.

Both an organizational-level instrument and a program-level instrument to assess CC have been developed by the Nathan Kline Institute Center of Excellence in Culturally Competent Mental Health. The guiding principles were to have objective measures that demonstrated the importance of the concept of CC, for which it was feasible to collect appropriate data that were reliable and accurate. While clinicians can be assessed on their self-perceived attitudes and beliefs, organizational and program-level items were selected to be verifiable.

The organizational level scale items were selected based on a review of recommendations, standards, and guidelines. A 27-site study succeeded in establishing the predictive validity of the several language accessibility items of the scale, showing they correlated with better engagement and retention. The program level scale was based on an ethnographic study of three long-standing programs that defined cultural groups in which deliberate and culturally grounded steps had been taken to make them acceptable. Key and common elements were observed across these programs and used to define domains of CC. It is informative to review these instruments as they were designed in form not only for assessment purposes but intended to serve as roadmaps for introducing CC activities into organizations and programs. Each criterion is scored with an ordered scale of whether or not an activity is accomplished, reflecting an increasingly complex degree of activity related to CC. A score indicates that the listed activity and all those below it are carried out. We use the scale items to step through specific activities that should be undertaken to promote CC.

Organizational Activities

The first set of criteria is meant to ensure that a procedure for setting up a CC presence in an organization is established. Fundamental is that an organization has an articulated commitment to CC evident in its mission statement. By creating a position with an assigned budget and having this person/group part of the governance of the organization, there are tangible assurances that CC activities can take hold. Strategies for obtaining information about and working with members of diverse cultural groups include establishing a committee that includes members of the largest cultural groups in the community and empowering that group to review current CC work and communicate its concerns and ideas directly with management. The organization needs to identify which cultural groups are being served, what their language needs are, and which members of the community are not being seen in order to target its outreach efforts. Particular data items need to be collected for all clients in a consistent manner so that they can be aggregated to assess the cultures and language needs of the population being served by the agency. Information on the target community allows the agency to tailor its outreach and services to the needs of its cultural groups.

Several criteria relate to staff. Hiring, recruiting, and retaining staff that is from prevalent cultural groups or who have experience working with these groups is often noted as ideal. This diversity should extend beyond clinical staff to include administrative and supervisory staff as well. Organizations can have policies in place that seek culturally representative employees and strategies to identify suitable candidates. Staff training, if related to the work of staff, might be an effective way to enhance the CC of all, both members of cultural and mainstream groups.

Literacy level and English proficiency can be barriers to accessing services. Several criteria address language accommodation. This could be accomplished through hiring bilingual staff or, if not able to do so, having interpreter services available for all service users from the point of the first contact with the organization. Additionally, organizations need to recognize that key forms and documents need to be translated into commonly encountered languages in order to work effectively with persons having Limited English Proficiency. Furthermore, service descriptions and educational materials also need to be translated, and for persons with limited reading skills, alternate formats can be introduced (e.g., novellas, videos, and items in simplified English). Finally, services that do not appear to be effective for persons from certain cultural groups may need to be adapted for use with that cultural group, or new services introduced that have been developed specifically for prevalent cultural groups (Table 7.10–4).

Program Activities

The program-level CC assessment scale identifies activities that are more directly associated with delivering care than organizational activities. The criteria address strategies for engaging consumers from cultural groups into a program, measures program capacity for meeting the language needs of clients, including gaining an understanding of and working with communication styles of clients. The appropriateness of the services offered to the cultural groups served, the use of support groups including peers, family, and kin, and utilizing cultural community resources to sustain a consumer's recovery are also measured. The scale examines program outcome goals and prods a program to include in outcome measurement cultural groups' understanding of and desirability for different modes of recovery. Finally, keeping in mind the health care reform goal of population health, community outreach to cultural groups is assessed.

Engagement in a program means that the client will receive a sufficient dose of the program so that the maximum benefit of the services can be realized. If appropriately engaged, it is likely that participants will adhere to treatment regimens and complete the work of a program. Programs need to understand the cultural needs of their clients to engage them. A welcoming milieu in the program is essential as well. The scale measures whether assessments go beyond the collection of standard demographic information such as age, gender, race/ethnicity, and preferred language to include more detailed information on cultural group identity, level of acculturation, and religious/spiritual beliefs. Primary to engagement is addressing language needs and ensuring effective communication with clients through an understanding of and adjustment to their possibly different styles of communication. Examining ways to increase trust in the staff and addressing the experience of stigma from the perspectives of different cultures also increase engagement.

The scale examines whether services have been culturally modified or new culture-specific services have been offered. Certain service modalities may not be suitable for cultural groups. This is true even for some highly endorsed evidence-based practices. Exploring the possibility of making modifications to services to make them more culturally relevant can be accomplished with the input from members of the cultural groups and further ensures a connectedness of the program with the cultural groups it serves. For example, group therapy that requires sharing of personal information with strangers is not an acceptable mode of care for older adult Chinese populations. However, a mahjong group run by an experienced therapist can combine the positive aspects of both a culturally enjoyable experience and a profitable therapeutic encounter.

Utilizing various support groups available to the client can enhance the possibility of positive outcomes. The scale examines the extent to which the program includes peer staff from the cultural group. These are both paid employees and volunteers who have mental health experience. Staff from the same cultural groups of clients enrolled in the program who have mental health experience can facilitate engagement of cultural clients in the program by enhancing consumer trust in the program and providing role models of recovery. They can also provide other program staff with information on cultural views impacting mental health treatment and acceptable treatment strategies. Family and social networks–as identified by the consumer–can play a large role in the lives of members of cultural groups, particularly in immigrant groups. Some consumers may rely heavily on particular family members for their supports and for maintaining their ties to their communities, while others may rely more heavily on close friends and members of their social network. The extent to which these persons participate in assessment, care planning, and providing ongoing supports is assessed. The scale also measures the extent to which a program identifies community resources that are acceptable and responsive to cultural values and preferences of consumers as referrals to promote healthy community living.

The scale identifies three ways to measure outcomes, a critical aspect of any assessment of health care programs. Service delivery outcomes such as patterns of use and program completion rates, while not measuring the quality of care, should not differ among cultural groups and act as basic markers of disparities. Satisfaction surveys need to be formulated with members of cultural groups as their desirable recovery outcomes may differ from those of clinically comparable consumers from other groups because of different beliefs, relationships, and resources to achieve goals. Finally, a program’s outreach efforts are measured. In order to increase the enrollment of cultural groups into programs that might be of benefit to them, outreach efforts should be undertaken that include identifying these groups in the community and informing them about what the program offers to people from these cultures (Table 7.10–5).

Table 7.10–4.
Organizational CCAS Criteria and Their Rationale

Criterion	Description	Rationale
1. **Agency's Commitment to Cultural Competence**	The agency (or its parent organization [PO]) has a management-level person responsible for CC and a dedicated budget for CC activities, a CC plan, and procedures for updating the CC plan.	A management-level person who has primary responsibility for CC ensures that CC will be addressed. Without a dedicated budget, only limited activities can be conducted. A written plan concretizes the agency's commitment to CC. Review and updating ensure that there is a feedback loop and that corrective actions can be taken.
2. **Assessment of Service Needs**	The agency obtains current data on its service users and its target community that enable identification of their cultures and language needs	Particular data items need to be collected for all clients in a consistent manner so that they can be aggregated to assess the cultures and language needs of the population being served by the agency. Information on the target community allows the agency to tailor its outreach and services to the needs of its cultural groups.
3. **Cultural Input into Agency Activities**	The agency (or its PO) has a CC Committee or other group that addresses cultural issues and has participation from cultural groups of the target community	A committee dedicated to CC will enhance the likelihood that activities appropriate to the cultures served are introduced and carried out. The committee can go under many names or be part of the functions of existing boards, advisory groups, and committees.

		Input is sought from representatives of the most prevalent cultural group of the target community. Having more than one representative from a cultural group makes active participation more likely.
4. Integration of CC Committee or Other Group with Responsibility for CC Within Agency	The CC Committee or other group with responsibility for CC is integrated within the agency evidenced by the following activities: ▶ Reviews services/programs with respect to CC issues at the agency ▶ Reports to Quality Assurance/Quality Improvement program of the agency/PO ▶ Participates in planning and implementation of services at the agency ▶ Directly transmits recommendations to the executive level of agency/PO	The extent to which the functions of the CC Committee are reported and used in the agency provides a measure of the likelihood of change with respect to CC. Four key committee functions are expected to take place, but they may be introduced at different stages.
5. CC Staff: Training Activities	The agency (or PO) offers to staff educational	Training and educating staff in CC enhances the likelihood of the delivery appropriate

	activities in which cultural issues are addressed and requires staff to have an adequate amount of specific training on CC	services to cultural groups. Ideally, educational activities should take place every year, and all staff members who have face-to-face contact with and provide direct clinical care to agency clients receive CC training. Administrative staff should also be knowledgeable about CC.
6. **CC Staff: Recruitment, Hiring, and Retention of Staff from/or Experienced with the Most Prevalent Cultural Groups of Service Users**	The agency is committed to hiring and retaining CC staff who are from or who have had experience working with the most prevalent cultural groups of its service users	Having direct service, supervisory and administrative staff with relevant experience with the most prevalent cultural groups enhances the likelihood of the acceptability and use of CC practices. Hiring and retaining professional staff members who are from the cultures of service users provides positive role models for clients of the agency and affords additional opportunities to increase knowledge about the cultures.
7. **Language Capacity: Interpreters**	The agency (or PO) accommodates persons who have limited English proficiency (LEP) by using interpreter services or bilingual staff	Language needs of persons with limited English proficiency must come to the attention of the agency at the point of first contact in order to identify their needs and link them to the appropriate services. Once a person has agreed to

		receive the services offered by the agency, language needs must continue to be addressed. Ideally, interpreters are formally trained and certified or are bilingual staff members who have received CC training.
8. Language Capacity: Bilingual Staff	The agency has staff who speak the language of the most prevalent cultural group of service users with members who have LEP	Having bilingual staff members who can work directly with persons with language needs is ideal–and likely to be cost-effective as well.
9. Language Capacity: Key Forms	The agency has translated versions of key documents and forms in the language of the most prevalent cultural groups of its service users	Legal documents and critical care information must be understandable to persons with limited English proficiency to ensure that such information is provided equally to all service users.
10. Language Capacity: Service Descriptions and Educational Materials	The agency accommodates persons with LEP by translating service descriptions and accommodates persons with limited reading skills (LRS) by providing service descriptions or educational materials in formats	Using different formats and media for descriptions of available services and for educational materials ensures that literacy level is not a barrier to accessing services. Service-related materials translated into the most prevalent non-English language will be especially helpful to overcome treatment barriers for

	they can understand.	service users and family members.
11. Assessment and Adaptation of Services	The agency (or PO) reviews and adapts or introduces services suitable to the most prevalent cultural groups of service users	A comprehensive CC-based assessment of services may result in the adaptation of existing services and the introduction of new services to better serve members of the most prevalent cultural groups.

Table 7.10–5.
Program Level CCAS

Criterion	Description	Rationale
1. Cultural Assessment of Consumer	Program obtains a detailed cultural assessment of each consumer that is additional to collecting standard (age, gender, race/ethnicity; preferred language) demographic information	Information that provides a view of beliefs, values, and world views of the consumer that can be used to improve diagnosing and the therapeutic alliance of staff with cultural groups. Information is recommended to be collected at the initial clinical assessment but can be updated throughout the course of the program. Cultural information should be made available to all clinical staff and care workers.
2. Engagement: Spoken language capacity	Program has procedures and staff to facilitate working with consumers/families with limited English proficiency or with consumers who are Deaf	If staff can communicate with consumers/families in their preferred language, the therapeutic encounter will be enhanced.
3. Engagement: Communication styles	All levels of program staff communicate with persons from cultural groups in ways that are respectful to them	Barriers between staff and consumers are reduced when modes of communication are respectful, acceptable, and familiar to the consumer.
4. Engagement: Trust building	Program intentionally and actively works to build and maintain the trust	Retention and engagement in the program are likely to increase if consumers feel their personal information

	of consumers and family members from cultural groups	is held confidential and that staff care about them by being responsive to their values and preferences as they relate to the program.
5. Engagement: Culturally framed stigma reduction	Program intentionally conducts activities to reduce stigma of mental illness as experienced by a cultural group	Stigma can impede engagement in services. Persons from cultural groups who might benefit from mental health services often delay seeking treatment or drop out of services because of the cultural group’s views of mental illness, including their attribution of its causes, its impact on other family members, and their social standing in their communities.
6. Culturally friendly service delivery environment	Program is conducted in physical spaces that are made appealing and familiar to the cultural groups served	Features of surroundings that are familiar, welcoming to families, and celebratory of a cultural group enhance consumer and family feelings of comfort and a sense of connectedness while participating in the program.
7. Culturally modified or new services	Program has strategies for making their services responsive to values and preferences of persons from	Services need to be reviewed and modified for applicability and their responsiveness to the values and preferences of the cultural group. New services may need to be developed.

	cultural groups with the aim of improving their effectiveness	
8. Inclusion in program of staff from cultural groups who have mental health experiences (consumer peers)	Program includes staff, both paid and volunteers, who have mental health experience and are from the cultural groups served by the program	Staff from the same cultural groups of clients enrolled in the program who have mental health experience can facilitate engagement of cultural clients in the program by enhancing consumer trust in the program and providing role models of recovery. They can also provide other program staff with information on cultural views impacting mental health treatment and acceptable treatment strategies.
9. Family member involvement	Program engages persons that are identified by the consumer from their family, including from social networks, to take part in care planning and providing ongoing support	Family and social networks can play a large role in the lives of members of cultural groups, particularly immigrant groups. Some consumers may rely heavily on particular family members for their support and for maintaining their ties to their communities, while others may rely more heavily on close friends and members of their social network.
10. Community resources	Program utilizes community	Consumers from cultural groups benefit from

	resources acceptable and responsive to cultural values and preferences of consumers to promote health outcomes	interventions that involve the use of community resources.
11. Satisfaction with Program	Program conducts activities to measure consumer and family member satisfaction with the services provided	Program improvement for cultural groups needs to be based on consumer and family member appraisals of the program.
12. Service Delivery Outcomes	The program examines disparities in service delivery outcomes related to engagement and retention in the program	Service delivery outcomes for cultural groups should not differ from those of clinically comparable consumers from other groups. Consumers from cultural groups may have particular difficulties in engagement in and retention of services. Differences in service delivery measures on engagement and retention could reflect these difficulties. These need to be identified and monitored by the program.
13. Recovery Outcomes	The program examines disparities in service process outcomes and	Recovery outcomes for cultural groups may differ from those of clinically comparable consumers from other groups because

	identifies and monitors recovery outcomes that are desired by consumers from cultural groups	of different beliefs, relationships, tools, and resources to achieve goals. These need to be identified and monitored by the program.
14. Program Outreach	Program identifies cultural groups in the community/service area who underutilize services and might benefit from the program and conducts outreach to increase their access	To increase the enrollment of cultural groups into programs that might be of benefit to them, outreach efforts should be undertaken that include identifying these groups and informing them about what the program can offer.

IN CONCLUSION

The mental health field has grappled with the complexities of psychiatric disabilities in an increasingly multicultural population. Clinicians need to be aware of and respond to more and more inclusive federal legislation focusing on nondiscrimination and the national standards for the promotion of CC in hiring practices and health care delivery. Current health care reform efforts are resulting in new models of integrative care that address Donald Berwick’s triple aims for the improvement of population health, higher quality care, and cost reduction. Clinicians can work from the ground up to incorporate CC into the new models providing an opportunity to better fulfill the triple aims. Population health requires knowing and understanding the needs of diverse cultural groups; higher quality care requires CC of organizations, programs, and clinicians. Cost reductions will follow if care is more appropriately targeted to cultural groups resulting in reduced use of expensive and inappropriate services. Systemic changes, as alluded to in organizational CC and structural competency, provide an outline for the necessary systemic changes. Clinicians, including psychiatrists and other mental health professionals, are the critical lynchpin in successful treatment encounters, so along with systemic change, clinician changes in attitudes and behaviors may act to shift the entire system toward the goals of CC.

FURTHER READINGS

Alegria M, Carson NJ, Goncalves M, Keefe K. Disparities in treatment for substance use disorders and co-occurring disorders for ethnic/racial minority youth. *J Am Acad Child Adolesc Psychiatry*. 2011;50(1):22–31.

Anderson LM, Scrimshaw SC, Fullilove MT, Fielding JE, Normand J. Culturally competent healthcare systems: a systematic review. *Am J Prev Med*. 2003;24(3):68–79.

Bromley E, Braslow J. Teaching critical thinking in psychiatric training: a role for the social sciences. *Am J Psychiatry*. 2008;165(11):1396–1401.

Center for Mental Health Services (US). *Cultural Competence Standards in Managed Care Mental Health Services: Four Underserved/Underrepresented Racial/Ethnic Groups*. Center for Mental Health Services, Substance Abuse and Mental Health Services Administration, US Department of Health and Human Services; 2000.

Clinton W. Executive order 13166: improving access to services for persons with limited English proficiency. *Fed Regist*. 2000;65(159):50121–50122.

Cross T, Bazron B, Dennis K, Isaacs M. *Towards A Culturally Competent System of Care*. Vol. 1. Georgetown University Child Development Center. Child and Adolescent Service System Program Technical Assistance Center; 1989.

Gone JP, Trimble JE. American Indian and Alaska native mental health: diverse perspectives on enduring disparities. *Annu Rev Clin Psychol*. 2012;8:131–160.

Griffen K, Evans L, Bor J. The affordable care act reduced socioeconomic disparities in health care access. *Health Aff*. 2017;10.1377/hlthaff.2017.0083. Online ahead of print.

Gutnick D, Siegel C, Laska E, et al. Making the cut: depression screening in urban general hospital clinics for culturally diverse Latino populations. *Gen Hosp Psychiatry*. 2017;45:85–90

Joint Commission. Advancing effective communication, cultural competence, and patient- and family-centered care: a roadmap for hospitals. Joint Commission; 2010.

Kleinman A. *Patients and Healers in the Context of Culture: An Exploration of the Borderland between Anthropology, Medicine, and Psychiatry*. University of California Press Berkeley; 1980.

Lee H, Porell FW. The effect of the affordable care act medicaid expansion on disparities in access to care and health status. *Med Care Res Rev*. 2020;77(5):461–473.

Metzl JM, Hansen H. Structural competency: theorizing a new medical engagement with stigma and inequality. *Soc Sci Med*. 2014;103:126–133.

Miranda J, McGuire T, Williams D, Wang P. Mental health in the context of health disparities. *Am J Psychiatry*. 2008;165(9):1102–1108.

Satcher D. *Mental Health: Culture, Race, And Ethnicity—A Supplement To Mental Health: A Report Of The Surgeon General*. Office of the Surgeon General (US), Center for Mental Health Services (US), National Institute of Mental Health (US); 2001.

Siegel C, Davis-Chambers E, Haugland G, Bank R, Aponte C, McCombs H. Performance measures of cultural competency in mental health organizations. *Adm Policy Ment Health*. 2000;28(2):91–106.

Siegel C, Haugland G, Chambers ED. Performance measures and their benchmarks for assessing organizational cultural competency in behavioral health care service delivery. *Adm Policy Ment Health*. 2003;31(2):141–170.

Siegel CE, Haugland G, Laska EM, et al. The Nathan Kline Institute cultural competency assessment scale: psychometrics and implications for disparity reduction. *Adm Policy Ment Health*. 2011;38(2):120–130.

Siegel C, Haugland G, Reid-Rose L, Hopper K. Components of cultural competence in three mental health programs. *Psychiatr Serv*. 2011;62(6):626–631.

Siegel C, Haugland G, Schore R. The interface of cultural competency and evidence-based practices. In: Drake RE, Merrens MR, Lynde DW, eds. *Evidence-Based Mental Health Practice: A Textbook*. W. W. Norton & Company; 2005.

Siegel CE, Reid-Rose L, Joseph AM, Hernandez JC, Haugland G. Cultural activation of consumers. *Psychiatr Serv*. 2016;67(2):153–155.

Siegel CE, Wanderling J, Haugland G, Laska EM, Case BG. Access to and use of non-inpatient services in New York state among racial-ethnic groups. *Psychiatr Serv*. 2013;64(2):156–164.

Sodowsky GR, Taffe RC, Gutkin TB, Wise SL. Development of the multicultural counseling inventory: a self-report measure of multicultural competencies. *J Couns Psychol*. 1994;41(2):137.

Sue DW, Sue D. *Counseling the Culturally Diverse: Theory and Practice*. John Wiley & Sons; 2012.

Tervalon M, Murray-Garcia J. Cultural humility versus cultural competence: a critical distinction in defining physician training outcomes in multicultural education. *J Health Care Poor Underserved*. 1998;9(2):117–125.

Tirado MD. *Tools for Monitoring Cultural Competence in Health Care*. Latino Coalition for a Healthy California; 1996.

Tucker CM, Herman KC, Pedersen TR, Higley B, Montrichard M, Ivery P. Cultural sensitivity in physician-patient relationships: perspectives of an ethnically diverse sample of low-income primary care patients. *Med Care*. 2003;41(7):859–870.

US Department of Health and Human Services. *HHS Action Plan to Reduce Racial and Ethnic Health Disparities: A Nation Free of Disparities in Health and Health Care*. US Department of Health and Human Services; 2011.

7.11 Medical Error

NORA C. OBERFIELD, M.D.

Doctors and the medical systems within which they practice strive for perfection in patient care. Nevertheless, as humans and technology are imperfect, medical errors occur in spite of this dedication to excellence. A landmark 1991 Harvard Medical Practice Study estimated that as many as 1.3 million injuries occurred nationwide to patients receiving hospital care. In 1999, the IOM warned the public that between 44,000 and 98,000 Americans may die each year as a result of medical errors. In 2013, in the *Journal of Patient Safety*, John T. James gave updated estimates of deaths associated with medical error during hospitalization based on studies published between 2008 and 2011. He estimated that more than 400,000 premature deaths occurred in hospitals per year related to preventable patient harm. That number is far greater than the estimate previously reported by the IOM, suggesting that there continues to be much room for improvement in this area. In 2016, the *British Medical Journal* noted medical error as the third leading cause of death in the United States. In 2019, the World Health Organization (WHO) reported that there is a 1 in 300 chance of a patient being harmed while receiving medical care. Too often, inherent bias related to race, ethnicity, gender, sexuality, and socioeconomic status also impacts who is more likely to experience a medical error. In 2018, in the *International Journal for Equity in Health*, researchers concluded that women and African-Americans were affected more than other groups.

Although a number of congressional actions have led to legislation requiring mandatory national or state reporting by hospitals of medical errors, reports of such documentation have been scarce. Issues to contend with are the need to protect doctors and hospitals from frivolous lawsuits resulting from public disclosure and failure to maintain confidential information about patients. While more hospitals are engaging in patient safety initiatives, a 2021 *Journal of Patient Safety* study found that these interventions have only been minimally effective in medical error risk reduction.

The topic of medical error has been complicated by the use of confusing and often conflicting terminology. The IOM separates medical errors into two main categories: errors of execution and errors of planning. An error of execution occurs when there has been a “failure to complete a planned action as intended,” and an error of planning occurs when there has been “the use of a wrong plan to achieve an aim.” Thus, by definition, medical errors require that there be a component of negligence. When they do occur, there is a desire for attribution to try to understand why the errors have happened. Did the error occur because the physician departed from an accepted standard of treatment? Was the error avoidable, or the result of poor training? Or was it unintentional—an accident—and better considered a close call or a borderline case?

It is important to distinguish medical errors from adverse events. The latter are related to a particular disease process and are most often predictable or, if not predictable, are known risks. Medical errors usually lead to adverse events and may harm the patient temporarily or permanently, may increase the time for recovery from disease, or may result in patient death. Adverse events may occur in the absence of medical errors, and medical errors may not be associated with an adverse event.

At Johns Hopkins University, the term medical mistake is used instead of medical error and is defined as follows:

> A medical mistake is a commission or an omission with potentially negative consequences for the patient that would have been judged wrong by skilled and knowledgeable peers at the time it occurred, independent of whether there were any negative consequences. This definition excludes the natural history of disease that does not respond to treatment and the foreseeable complications of a correctly performed procedure, as well as cases in which there is reasonable disagreement over whether a mistake occurred.

Types of Error in General Medicine

The current literature highlights five main types of medical errors: errors of commission, errors of omission, errors of communication, errors of context, and diagnostic errors.

1. An error of commission occurs when something done by a member of the medical team leads to patient harm either because it was the wrong thing to do or because it was the correct thing to do, but it was not performed correctly.
2. An error of omission occurs when something that was clearly needed for the patient’s healing process was not done.
3. An error of communication occurs when information is not conveyed effectively, either between at least two providers of care or between a provider of care and the patient.
4. A contextual error occurs when a member of the health care team does not recognize social factors in the patient’s life, which could

create barriers to the implementation of effective postdischarge treatment.
5. Diagnostic errors occur when a wrong diagnosis is made, which could have sequelae such as inappropriate or delayed treatment.

Errors may also be categorized as systematic or individual. Systematic errors are caused by flaws in the hospital system or in the transfer of information from one unit of the hospital or from one group of doctors to another. In general, no one person is to blame. An example would be the unavailability or inability to retrieve a patient's medical records or a mismatched blood transfusion. The failure is in the health care delivery system as a whole rather than in the individual.

Individual errors are attributed to a particular physician or to a specific member of the treatment team. Examples include prescribing the wrong medicine or operating on the wrong leg. The physician bears the responsibility, even if the error is made by a person under his or her supervision. Research at John Hopkins Medical Center involving an examination of 25 years of U.S. malpractice claim payouts from 1986 to 2010 indicated that diagnostic errors were the most common type of medical error. A 2020 report from AHRQ suggests that diagnostic errors remain prominent at this time. While the predominant focus in recent years has been on reducing medication and surgical errors, the importance of improving the error rate in the diagnostic area is equally important.

Types of Error Specific to Psychiatry

While there is increased discussion of medical error generally, errors occurring in patients with primary psychiatric diagnoses are an area often excluded from the literature. S.W. Cullen and colleagues separated errors in inpatient psychiatry units into four categories: treatment errors, diagnostic errors, preventive errors, and other errors.

1. Treatment errors are defined as "…failures in the formulation or delivery of an appropriate intervention, therapy, or procedure."
2. Diagnostic errors are not only inappropriate psychiatric diagnoses but also mistaking a physical illness for a psychiatric condition or vice versa.
3. Preventive errors were the most common and have been defined as "any failure to offer indicated prophylactic treatment, as well as

inadequate or inappropriate monitoring of an initiated treatment plan." Preventive errors include inadequate or disruption of safety protocols "…such as fall precautions, locked unit doors, break-away bars in closets, or 15-minute checks for patients with suicidal ideations."

4. The "other errors" category included problematic transfers and discharges.

In the Cullen study mentioned above, it was found that nonpsychiatric staff members felt ill-prepared to deal with referrals of patients with a history of psychiatric treatment or hospitalization or of patients who were in a state of emotional distress. There was often an inadequate or cursory medical examination in patients who were disruptive or otherwise uncooperative. In order to reduce the error rate among such patients, it will be necessary to decrease stigma about mental illness and educate all health care providers about the complicated presentation of some psychiatric patients with active symptoms such as psychosis. In addition, compliance issues may be a factor in the management of psychiatric patients with a medical illness which may cause difficulties implementing safe and appropriate discharge plans.

Causes of Error in General Medicine

As mentioned above, the causes of medical errors are multifactorial. In general medicine, they include medication errors, transitions of care errors, errors stemming from the use of EMRs, errors related to the challenges of treating a growing geriatric population, and errors related to job overload, fatigue, and physician burnout, among others discussed below. In most instances, medical errors correlate with inexperience. For example, first-year house officers make more errors than second- and third-year residents. Junior attending physicians make more errors than senior attending physicians. Clearly, the aphorism of experience being the best teacher seems to apply in these instances, but experience or the lack thereof is not the only factor.

Medication. According to statistics from the IOM, medication errors in the United States contribute to 1 out of every 131 outpatient deaths and 1 out of every 854 inpatient deaths. The U.S. Food and Drug Administration (USFDA) reports that over 770,000 patients are injured each year due to medication errors. In France, a 2007 study on adverse events in inpatient healthcare noted that 50% of serious adverse events were caused by medication errors. At the Mayo Clinic, medication errors are separated into three categories: medication factors, patient factors, and health care professional factors. Medication factors include medications with names that sound similar or medications with low therapeutic indexes. Patient factors include metabolic challenges related to poor hepatic or renal function, patient cognitive status, and polypharmacy. And, lastly, health care professional factors include poor handwriting and the use of abbreviations in prescriptions and progress notes. A 2019 study out of Sao Paulo, Brazil, indicates that electronic medical record usage in the emergency room setting led to a decrease in medical errors. Several studies found that interruptions during rounds and pressure to complete rounds were also potential contributors to medication and other errors.

Transitions of Care. It is important to note the role that transitions of care can play in a medication error. One study found that nearly 50% of patients had at least one medication error when discharged from inpatient units.

Working Hours. Concerns about resident overwork, which may contribute to medical error, generated a new set of duty-hour regulations recommended by The Accreditation Council for Graduate Medical Education (ACGME) that took effect in July 2011. Those recommendations limited the maximum shift length for first-year residents (also known as interns) to 16 hours a day. Although the particulars of duty hours are discussed further in the section *Regulating Working Hours of Interns and Residents,* they are mentioned here in order to examine how this change affected medical error. The change in duty hours was enacted by the ACGME based on research that demonstrated that long duty hours harm both patients and residents. Decreased shift lengths for interns led to a decrease in the incidence of serious medical errors in intensive care units. Longitudinal studies of duty hour restriction, however, did not show that house staff used that time to sleep more hours than previously, nor was there a lessened incidence of burnout. Working shorter shifts may have had the unintended consequence of increasing stress by requiring residents to complete the same amount of work but in less total time. Increased work compression has been associated with poorer clinical performance and decreased satisfaction among residents.

HANDOFF ERROR. Shorter shifts have also led to more handoffs (when hospital staff who are going off shift communicate information about their patients to the hospital staff coming on shift), and there is a growing literature indicating an elevated rate of medical errors with more handoffs. Different hospitals perform handoffs in different ways, and even individual departments within the same hospital perform handoffs differently. While some hospitals may use a handoff system combining verbal and paper sign-out, other hospitals use an electronic document that is continuously updated. But there is room for error in these frequent transactions. For example, one study found that there were 24 unique handoff mnemonics used in one hospital. In addition, many doctors have never been given formal training in handoff communication. With the rise of errors attributed to handoffs, there is a growing call for a need for a systematic algorithm to be followed by all departments in all hospitals. Efforts are currently being implemented in some hospitals across the United States, focusing on training both medical students and residents to improve handoff communication in order to decrease medical errors.

Electronic Medical Records. The use of EMRs has generally been met with a favorable response, particularly in terms of how electronic documentation decreases medical errors related to illegible handwriting in progress notes and prescriptions. However, some physicians are concerned that EMR may lead to an increase in medical errors because the EMR has replaced face-to-face or phone communications.

James Bernat was among the first to point out the ethical and quality problems associated with the use of EMR. He pointed out the hazards of

> …copying and pasting data obtained from other clinicians, authorship ambiguities, inadvertent inclusion of unobtained data in templated notes, misleading history, and physical exams, failure to review prepopulated data, inadequate discharge summaries, impairments to patient-physician communication, and the transformation of the purpose of the medical record to billing documentation.

In addition, many clinicians have commented on the fact that EMR is making patient care less about the patient and more about record keeping. One study found that medical interns spent only 12% of their time with patients and 40% of their time at computers. Patient encounters have become experiences where the expectation of eye contact between doctor and patient can no longer be guaranteed, with the physician often having to fill out automated checkboxes on the computer screen while speaking with the patient. Paying less direct attention to the patient could lead to medical error by missing pertinent details of the history or missing subtle nonverbal communication signs, which could be diagnostic clues. Physicians should meet the patient before entering information into the computer.

Population Demographics

In 2021, there were 55.9 million people in the United States who were 65 years old or older. Estimates suggest that by 2030, there will be roughly 73 million people in this age category. In the United Kingdom, there were 5.9 million people listed as 75 years old or older in 2020, with an expected rise to 8.7 million by 2033. Through improvements in scientific understanding of disease pathology, people are living longer than they did previously, often due to taking medications that slow or halt the progression of several comorbid medical conditions. Given the growing size of older adult populations internationally, in order to reduce medical errors, there must be an increased understanding of the types of medical errors that occur in the older adult population. Polypharmacy, taking several medications daily, can be a significant contributor to medical error, given that medications may have negative drug–drug interactions, and increased care must be taken to administer particular medications at specific times of day to avoid such interactions. In one study from the United Kingdom examining 13 long-term residential homes, polypharmacy is evident, with the average resident receiving seven or more medications. Age-related metabolic changes in how the body responds to medications must also be taken into account, with a need for greater education on renal and hepatic dosing of medications.

Job Overload, Fatigue, and Physician Burnout. Job overload, in which doctors are required to care for too many patients, is a major contributor to medical error. In the hospital setting, interns are assigned to too many patients with too few supervisors. In the outpatient setting, too many patients are cared for by too few doctors. Private practitioners are under pressure from managed care companies who often reward doctors financially for productivity, which translates into a high caseload of patients that leads to overwork. Similarly, some managed care and insurance companies reward doctors who order few laboratory tests, thus lowering cost; however, this may result in misdiagnosis, especially in complex cases with atypical presentations. Fatigue and burnout interfere with doctors' judgment and lead to mistakes. While studies report rates of burnout in medical students from 31% to 49.6%, among medical residents, this percentage increases, with some specialties reporting rates as high as 76% (internal medicine). Among practicing physicians, the rates vary by specialty, with emergency medicine physicians reporting burnout at a rate of 51%, while psychiatrists are on the lower end of the spectrum, reporting burnout at 33%. In 2018, ACGME introduced several wellness initiatives into medical education and residency training programs in an attempt to decrease burnout.

Causes of Error in Psychiatry

As within general medicine, the causes of error in psychiatry are multifactorial. Areas to examine include the specifics of the inpatient psychiatric care setting, medication errors, suicide, and restraint and seclusion practices.

Inpatient Psychiatry Unit Characteristics. One of the most striking ways in which inpatient psychiatry units are distinct from most other inpatient units in hospitals is that they are locked units. While the primary purpose of this locked structure is to maintain the safety of the patients and those around them, the fact that the units are locked has the potential to make emergency treatment logistically more challenging. Additionally, again with safety in mind, most inpatient psychiatry units do not allow for any sharps on the unit (with the exception of very small needles used for blood draws, which are often kept off the unit until needed). This means that if a patient is at a level of medical distress requiring the use of IVs, delays in getting the patient to a secure environment, which allows for IV access, are inevitable. These delays may lead to patients not receiving emergent care in the most appropriate time frames, which could lead to an increased error rate in this patient population. Overall, more medication errors occur in hospitalized psychiatric patients than in psychiatric outpatients.

Medication. Most mistakes in psychiatry occur in the area of medication. A wrong prescription may be written, the wrong medication for a disorder may be prescribed, or the wrong dosage of a drug may be administered. A common cause of medication error is poor handwriting by physicians, which is then misread by a pharmacist or nurse. A potential remedy is to insist that all prescriptions be typed and transmitted to pharmacies via computer (electronic prescribing has been increasingly adopted by most medical practices throughout the United States). Similar-sounding drug names have also been to blame for medication mistakes, for example, citalopram (Celexa) and celecoxib (Celebrex), buspirone (BuSpar), and bupropion (Wellbutrin) (Table 7.11–1). Drugs with similar phonemes lend themselves to miscommunication in writing or in speech.

Table 7.11–1.
Examples of Drug Name Confusion

Psychoactive Medication	Confused With
Metadate CD (methylphenidate)	Methadone (Dolophine)
Levoxine[a] (levothyroxine) (thyroid preparation)	Lanoxin (digoxin)
Serzone (nefazodone)	Seroquel (quetiapine)
Lamictal (lamotrigine)	Lamisil (terbinafine) (for nail infection)
Zyprexa (olanzapine)	Zyrtec (cetirizine) (for allergies)
Celexa (citalopram)	Celebrex (celecoxib) (for arthritis)

[a] Name changed to Levoxyl after U.S. Food and Drug Administration report in 1994.

Wrong drug administration, that is, giving a drug to a patient that was not the drug prescribed for that patient or giving the correct drug but at the wrong dose, is the most common type of medication error. Wrong drug administration occurs more frequently on units with either fewer registered nurses or two or more patients with the same (or similar) name staying on the same unit. Most prescription errors in psychiatry are with antipsychotic medications followed by anxiolytics and antidepressants. Increasingly, studies are showing that the use of computerized medication orders decreases errors related to drugs. There is a growing consensus that the current dispensing of drugs as needed (the so-called PRN medications) ought to be re-examined. One recent study found an increased risk of error by 15% each time a PRN medication was given on medication rounds. As-needed directions are often ambiguous and do not always indicate when or how often, or at what dosages they are to be used. While doctors are placing the orders, nurses are often on the front lines of decision-making as to when and how these medications should be dispensed, and in some instances, are not trained sufficiently to thoroughly assess how these medications may interact with other medications the patient is taking. Therefore, in an effort to decrease medication error rates, PRN orders should always be written with clear identification of indication, frequency, and dosage.

Suicide. Unlike most physicians, psychiatrists rarely have to deal with the death of patients. Exceptions may be made for geriatric psychiatrists who deal with the aged or for consultation-liaison psychiatrists who deal with the seriously medically ill; however, when death occurs in those instances, it is often expected and part of the disease progress.

Suicide is unexpected. It is also viewed by many persons as preventable. On the basis of that expectation, family members may sue for malpractice after the event. Suicide is not always preventable, however. Some persons who have a strong desire to kill themselves will do so despite the many advances that have been made in the treatment and management of suicidal patients. When mistakes occur, they are usually the result of the psychiatrist having underestimated the suicidal risk or if the psychiatrist has appraised it accurately, the nurses and attendants not taking the necessary steps to ensure the patient's safety (e.g., constant observation).

Litigation is more common for suicides that occur in a hospital than for those that occur in an outpatient setting, presumably because one-to-one coverage is the standard of care for the actively suicidal hospitalized patient. Of all the suicides that occur in the United States, approximately 5% occur during hospitalization. The most common methods of suicide in the hospitalized patient are jumping or hanging. In most cases, safety protocols such as one-to-one monitoring or 15-minute checks were not followed assiduously.

Courts have not penalized psychiatrists for honest errors of judgment in clinical practice. One court stated: "The accurate prediction of dangerous behavior, and particularly suicide and homicide, are almost never possible. Thus, an error of prediction or even of judgment does not necessarily establish negligence." In spite of this, however, psychiatrists are expected to evaluate patients properly for suicide risk and to establish a rational treatment plan to ensure that the risk of a patient successfully killing themself is minimized as much as possible.

The effect of having a patient commit suicide can have profound effects on the treating psychiatrist. There is an initial sense of disbelief and shock upon being informed of the event, often followed by feelings of guilt. Common questions asked are: Did I overlook anything? How could this have been prevented? What should I have done differently? Some psychiatrists are so traumatized that they give up practicing altogether or never again work with patients who are suicidal. More commonly, however, they seek to talk about the experience with other clinicians, especially clinicians who have gone through a similar experience of losing a patient to suicide. This process can be very salutary, allowing the psychiatrist to realize that they are not alone, that others have gone through the same experience, and as in all branches of medicine, including psychiatry, not all illnesses are curable and some suicides, even when care and treatment are exceptional, can neither be predicted nor prevented.

Restraint and Seclusion

Errors in the area of restraint result from not using these devices correctly; this results in physical injury or, in some cases, death by strangulation or cardiac arrest. Patients in seclusion rooms may injure themselves by banging their heads into the walls or exhaust themselves as a result of agitated behavior that can lead to dehydration and death. In both situations, constant vigilance is necessary. In 2006, the Centers for Medicare and Medicaid Services (CMS) mandated more training for health care workers in the area of restraint and seclusion in an effort to promote safer practices and greater acknowledgment of patients' rights. They later released a revised set of guidelines related to restraint and seclusion, which included "face-to-face examination by a physician, nurse, or physician's assistant within 1 hour of placement in restraints or seclusion." Some hospitals have initiated policies in which patients in restraints and seclusion must be on one-to-one constant observation with vitals checks every hour, with physician review for the necessity of restraints every 4 hours. Seclusion rooms must be designed to be environmentally protective, for example, padded walls, ventilation. In New York State nursing homes, efforts have been made to reduce the use of physical restraints, with a decrease from 6.14% in 2011 to 1.86% in 2014.

PREVENTING MEDICAL ERROR

Practice Guidelines

Practice guidelines have been developed by the APA for a variety of disorders. These include delirium, dementia, substance use disorder (SUD), depressive disorder, panic disorder, borderline personality disorder, suicide, and schizophrenia. Guidelines for other psychiatric disorders will be forthcoming. According to the APA, the term *practice guideline* refers to a “set of patient care strategies developed to assist physicians in clinical decisions.” They are careful to point out that guidelines are not meant to be standards of care:

The APA practice guidelines are not intended to be construed or to serve as a standard of medical care. Standards of medical care are determined on the basis of all clinical data available for an individual case and are subject to change as scientific knowledge, technology advance, and patterns evolve. These parameters of practice should be considered guidelines only. Adherence to them will not ensure a successful outcome in every case, nor should they be construed as including all proper methods of care or excluding other acceptable methods of care aimed at the same results. The ultimate judgment regarding a particular clinical procedure or treatment plan must be made by the psychiatrist in light of the clinical data presented by the patient and the diagnosis and treatment options available.

Although few studies have been done, common sense indicates that guidelines will improve the quality of care and diminish medical errors in psychiatry. They are works in progress, however. As new methods of treatment develop, they have to be revised and updated, lest they become fossilized and irrelevant. Other risks are their possible misuse by third parties such as managed care companies and health maintenance organizations (HMOs). Many guidelines are prepared by third-party payers and may be compromised by conflicts of interest of cost versus quality of care. Some HMOs pay bonuses to their doctors based on high scores on how closely they adhere to a set of guidelines in the treatment of a specific disorder. This may influence psychiatrists in the choice of therapy or the length of the patient’s hospital stay if they are not in keeping with the set of guidelines being used.

There is no indication that there has been an increase in malpractice suits against psychiatrists since their introduction; however, published guidelines are often brought forward in litigation as standards of care. Psychiatrists and other physicians should not be intimidated by this practice. Psychiatrists may legally and ethically deviate from guidelines; however, the decision to do so and its rationale should be documented in the chart.

As mentioned previously, the APA guidelines do not include all proper methods of care for a particular disorder, nor do they allow for ambiguity, controversy, or dissent. For this, the practitioner must turn to a textbook, such as the *Comprehensive Textbook of Psychiatry,* in which such differences are discussed.

Electronic Medical Records, Medication Reconciliation, and Electronic Prescribing

The EMR is fast becoming an integral part of most hospital and outpatient clinic settings, with several notable advantages over paper records. According to Dr. Michael Houston, these advantages include "…the ease with which information can be shared among providers, thereby improving the efficiency of health care delivery by reducing redundancy, simplifying billing, and improving quality, and safety." EMR became more commonplace following the 2009 Health Information Technology (HIT) for Economic and Clinical Health (HITECH) Act, part of the American Recovery and Reinvestment Act of 2009, as well as the 2010 Patient Protection and Affordable Care Act, national legislation which included bonuses for medical sites participating in EMR.

The advent of EMR led to improvements in medication reconciliation, defined by the Institute for Healthcare Improvement as "a process of identifying the most accurate list of all medications a patient is taking—including name, dosage, frequency, and route—and using this list to provide correct medications for patients anywhere within the health care system." One study polling patients who were about to be discharged from inpatient hospitalizations noted that 27.9% of the patients were unable to list their discharge medications, and even a smaller percentage had an understanding of what their medications were for. Acknowledging that medication-related errors are frequent contributors to medical error, the Joint Commission declared medication reconciliation a 2005 National Patient Safety Goal. Notably, 2021 NPSGs include decreasing wrong-site surgeries, misidentification of patients, and medication errors. Stephanie Mueller and colleagues have found medication reconciliation to work most effectively when two particular conditions are met: close communication between physicians and pharmacists and ensuring that patients at higher risk of medication error are more closely monitored. This latter category includes the elderly, individuals taking multiple medications, and individuals with multiple medical comorbidities.

In an effort to eliminate medication errors occurring due to illegible handwriting, most medical centers throughout the United States now use electronic prescribing (frequently referred to as e-prescribing). This practice allows prescriptions to be electronically transmitted directly to pharmacies. EMR systems are now used by most hospitals. However, different hospitals often utilize different EMRs, which can pose challenges in sharing patient information when needed. This lack of ease of patient information sharing can lead to more medical errors. On the other hand, however, for hospital systems that share EMRs, facilitated information sharing can theoretically lead to less medical error. Further research on this topic is needed since universal EMR use has only been in practice since 2016.

Blame and Medical Error Reporting Structures

The Royal Australian and New Zealand College of Psychiatrists raised an interesting argument when they suggested that for improvements in medical error to occur, modern medical culture must undergo a significant frameshift. In particular, they recommended moving away from a model of individual blame toward one of acknowledging systems failures. They found that “a “no blame” culture is essential to voluntary reporting, dispassionate analysis, and correction of systems failures.” Although the blame is diffused in this model, accountability is not. Every member of the medical community has a responsibility to their colleagues and their patients to take part in telling the story of what went wrong as the first step in determining how to ensure this error is not repeated. Assurances of confidentiality and ease of reporting are the most pertinent factors leading individuals to freely report errors. In one study, it was found that clinicians and other members of the hospital staff were most likely to tell a member of the medical error report team about an error in person (53%), followed by phoning in the error (35%), using the online data reporting system (6%), submitting a written form (4%), and lastly, by email (2%).

Regulating Working Hours of Interns and Residents and Improving Handoffs

Teaching hospitals often rely on interns and residents to perform services such as phlebotomies and intravenous (IV) therapy and to serve as messengers and transporters, tasks that are more appropriately performed by ancillary personnel. In addition, house staffs are often required to work long hours, and sleep deprivation can impair their judgment and clinical skills. The average resident works more than 80 hours per week, with a shift of 30 hours every third or fourth night. As mentioned above, a limit on the number of hours interns and residents may work was set forth by the CMS. Their work rules were (1) residents were limited to no more than 12 consecutive hours per assignment in emergency services, (2) residents could not work more than 80 hours per week over a 4-week period, and could not be scheduled to work more than 24 consecutive hours, and (3) scheduled rotations had to be separated by not less than 8 hours of nonworking time, provided for each week. Nonworking time is time away from training and patient care activities. In 2000, 8% of residency training programs nationwide were cited by the Accreditation Council for Graduate Medical Education for violating the CMS work hour requirements.

In response to the death of a patient at a New York hospital in 1989, the New York state legislature enacted the Bell Commission Report, which implied that resident work hours contributed in part to that accidental death. The state responded by establishing rules for resident work hours, including 80-hour week limits and 24-hour consecutive call limits.

The impact of these limits on the continuity and quality of care remains to be seen. Legitimate concerns have been raised about shorter shifts leading to more handoffs, which could lead to more opportunities for error.

The Joint Commission has reported that “Communication and handoff failures are among the root causes in nearly two-thirds of “sentinel events,” which are serious, often fatal, preventable adverse events in hospitals.” Acknowledging the need to address handoff failures with a greater scope, the ACGME now mandates that all “…residency training programs provide formal instruction about patient handoffs and that faculty monitor handoffs and ensure adequate handoff skills through direct observation.” Unfortunately, despite this mandate, many residency training programs still do not consistently train their residents in how to perform proper handoffs. In one recent study, roughly half of the residents polled believed that critical patient information was lost during resident shift changes. From June 2010 through February 2014, a comprehensive handoff curriculum, known by the mnemonic I-PASS (Illness severity, Patient summary, Action list, Situation awareness and contingency plans, and Synthesis by receiver) was piloted in nine different pediatric residency programs throughout the United States. In developing this curriculum, it became evident that the key ingredients for an effective handoff include: dedicated time for verbal exchange in a HIPAA-compliant location, the active use of templates for accessing and recording information, priority given to discussing ill patients, and highlighting action items. Use of the I-PASS curriculum was associated with a 23% relative reduction in the medical error rates across all nine study sites, as well as a 30% relative reduction in the rate of preventable adverse events. Importantly, the study found that the use of a comprehensive handoff curriculum did not elongate the handoff process in any way, nor did it cause residents to spend less time with patients. A 2020 review of I-PASS utilization in *Pediatric Quality & Safety* has shown I-PASS to be effective in reducing medical errors.

In an effort to eliminate mistakes that might occur-related to medical student involvement in patient care, the CMS defined medical student responsibilities as follows: (1) Students can take histories and perform physical examinations with the approval of the patient's attending physicians; (2) they may write in the patient's chart, but all entries must be countersigned by the attending physicians; and (3) medical or surgical procedures may be performed if they are under direct inpatient supervision.

Many medical educators believe that the current CMS rules do not go far enough and that interns and residents are not used properly by many teaching hospitals. It is not uncommon for a resident in cardiac surgery to assist with an operation for 14 hours and then to stay on duty for an additional 10 hours. Similarly, a pediatric resident may be in the emergency room for 24 hours without sleep. Even though their hours fall within the CMS guidelines, they are not conducive to the resident's education in view of the inevitable fatigue when a person is without sleep for 24 hours. For residents with families, especially those who are parents, current work schedules disrupt relationship harmony and child-rearing. This added stress interferes with optimal functioning. It has been suggested that chronic sleep deprivation may put residents at increased risk for motor vehicle crashes, medication errors, and decreased mental clarity. There are no easy answers to these problems, but the current situation clearly requires resolution.

Emergency physicians commonly experience sleep deprivation because of the need to work shifts during the evening and late-night hours. The negative effects of this problem are compounded by job stress and traditional methods of scheduling work shifts. Sleep deprivation may be reduced by schedules designed to lessen interference with normal sleep patterns and circadian rhythms. Pharmacologic treatments for sleep deprivation exist in the form of alertness-enhancing agents such as caffeine. Sleep-promoting agents may also help treat the problem by helping physicians to sleep during daytime hours. Minimizing sleep deprivation may help prevent job burnout and prolong the length of an emergency physician’s career.

Coping with Medical Errors and Adverse Events

Doctors have a long tradition of examining adverse events and of learning from their mistakes. The case presentation (also called the *clinicopathologic conference* [CPC] or morbidity and mortality conference [MMC]) is a group exercise attended by clinicians, researchers, educators, and students who examine all aspects of patient diagnosis, treatment, and outcome. Ideally, CPCs/MMCs do not attempt to cast blame on a practitioner; however, if negligence has occurred, then it becomes evident in the open discussion of those in attendance. A 2012 study examining how MMCs are used to improve the quality of medical care found that a lack of structure in the approach to MMCs frequently led to few clear action plans with definitive timelines. While CPCs/MMCs also could be a useful time to address the emotional impact of having made a medical error, there appears to be less comfort spending time on this topic. As the perception of having made a medical error can lead residents to experience significant personal distress and decreased empathy, it is perhaps time to incorporate elements of self-care and coping strategies as part of CPCs/MMCs. A 2020 article in the *Western Journal of Emergency Medicine* suggests that changing the name from CPCs/MMCs to a case-based error reduction conference (CBERC) may eliminate the propensity for the “blame game” and lead to actual systemic improvements in medical error. Most hospitals also have risk management teams whose duty is to identify potential or actual instances of malpractice by physicians in the facility and to approach the patient and discuss the situation openly and honestly. Lastly, offices of quality assurance function to assess and prevent risks to patients in a continuous effort to eliminate adverse effects, as described in Table 7.11–2.

Ethics of Disclosure. There is an ethical duty on physicians to disclose medical errors, accidents, injury, or bad results to their patients who have an absolute right to understand what occurred during their course of treatment. An explanation is mandatory, and an apology is appropriate if an error was made. The American Academy of Pediatrics believes, “An atmosphere of openness and honesty leads to a culture of quality and safety.” If, as a result of a disclosure, legal proceedings ensue, then the physician has the opportunity to offer an explanation—but appropriate compensation for harm suffered as a result of malpractice is a patient’s legal right.

Table 7.11–2.
Medical Documentation Requirements for Patient Adverse Events

These guidelines apply to instances of *adverse patient events,* which are defined as follows: "*A significant adverse patient event is an unplanned or unexpected occurrence that results in patient death, harms the patient, temporarily or permanently, or significantly increases the planned length of hospitalization.*" These guidelines are applicable, however, to all medical documentation and should be universally followed.

It is most important that the responsible attending physician (or covering attending) write a note as soon as possible after an adverse event, certainly within 24 hours. This note should clearly document their active participation in the care of the patient. Subsequently, an attending note must be written daily, for as long as appropriate to the patient's status. This is a supplement to the standard Medical Board documentation requirement, especially crucial in the first few days after an adverse event occurs.

All notes must be dated, timed, and signed.

The purpose of medical documentation is to promote patient care and safety, not to cast blame on an individual for an adverse patient event. When describing a patient adverse event, extreme care must be exercised to avoid implied blame for such an event by another caregiver. The note should clearly document the attending physician's participation in the care of the patient, and the discussion should be an objective analysis of the events that transpired, an assessment of the patient status, and a plan for subsequent management. An open disclosure approach to adverse event documentation must be followed.

Courtesy of NYU Langone Health.

Patient–Doctor Relationship. A good patient–doctor relationship is characterized by a sense of trust, respect, and honesty between the two parties. The better the patient–doctor relationship, the less chance for misunderstanding leading to litigation. Studies have shown that when a medical error or adverse event occurs within the context of a good patient-doctor relationship, litigation is rare. In its absence, however, litigation is likely in as much as 70% of cases of medical error, especially if a major mistake was made, regardless of whether or not death ensues. In surveys of patients, almost 100% desire that doctors report and discuss medical errors with them. Acknowledging a medical error, minor or major, may actually reduce the risk of a malpractice action.

Therapeutic Privilege. Therapeutic privilege is the physician's right to withhold information from a patient if, in their opinion, such disclosure would cause irreparable harm to the patient. Some patients do not wish to know the nature of their illness, and that right needs to be respected. Others may be so prone to anxiety or panic that even a hint of medical error (especially if no adverse outcome occurred) would be potentially dangerous to the patient's overall sense of well-being.

Duty to Disclose. There is a patient–consumer movement supported by some members of the legal profession that would require physicians by law to report medical errors. As mentioned previously, disclosure is currently an ethical, not a legal, obligation. Proponents of duty to disclose argue that self-policing by the medical profession is inefficient, and a legal duty to disclose malpractice would provide a remedy. Proponents of such laws would also require that they apply not only to the person committing the error but also to those who witness the error, for example, nurses and members of the treatment team.

Opponents of the duty to disclose law argue that it would harm the patient–doctor relationship and, in particular, weaken the trust and teamwork among health care professionals who would be required to be informants on one another. Adopting a rule of disclosure also suggests to the public that doctors, as a group, are not to be trusted. Opponents of the law also note that quality assurance and risk management programs that bring negligent acts to light are already in place and that further enforcement is unnecessary. Most experts take the position that physicians are the ones who ultimately must decide to disclose errors based on their professional and ethical beliefs.

Barriers to Disclosure. Some suggest that, rather than focusing on whether disclosure should be an ethical or legal obligation, the greater question to examine is why there is often a fear of speaking up when a medical error has been committed. A 2013 study out of Minnesota reports, "Medical students and residents may be concerned about retaliation in the form of poor grades and evaluations. Confronting an attending physician can be intimidating, and…the “hidden curriculum” (a set of influences that function at the level of the organizational structure and culture) may encourage…students and residents to remain silent.” Work in this area is still ongoing, as evidenced by a 2019 study in *Academic Medicine* which highlighted that medical students experience distress and are unsure of what to do when they encounter or make medical errors. They are still not getting adequate training in error reporting and emotional support when doing so. A medical culture that embraces nonjudgmental systems of error reporting may be critical to decreasing medical errors from occurring in the first place.

Role of Apology. It is becoming standard practice for physicians to apologize to patients when disclosing unanticipated outcomes that fall within the category of medical error. For an apology to be successful, however, there must be a real sense of regret communicated to the patient or family. It is not sufficient to say “I am sorry” pro forma. There must be an empathic communication of concern and personal responsibility. Apologies have generally not been admitted in evidence in malpractice suits because it would have a chilling effect on the entire process. Juries could interpret an apology as an admission of guilt and be less objective about facts in the case that might otherwise be exculpatory.

Regulation of Privacy, Delivery, and Quality of Health Information Distribution in the Age of Electronic Medical Records

Health Insurance Portability and Accountability Act. The HIPAA was passed in 1996 to address the medical delivery system’s mounting complexity and its rising dependence on electronic communication. The act orders that the federal Department of HHS develop rules protecting the transmission and confidentiality of patient information, and all units under HIPAA must comply with such rules.

Two rules were finalized in February 2003: the Transaction Rule and the Privacy Rule. The Transaction Rule facilitates transferring health information effectively and efficiently by means of regulations created by the HHS that established a uniform set of formats, code sets, and data requirements. The Privacy Rule, administered by the Office of Civil Rights (OCR) at HHS, protects the confidentiality of patient information. This means that a patient’s medical information belongs to the patient and that the patient has the right to access it. Psychotherapy notes previously were the exception to this rule, which were deemed as property of the psychotherapist who wrote them. However, the 21st Century Cures Act, enacted on April 5th, 2021, requires that all notes are available to all patients in an effort to increase transparency. Hopefully, patients knowing more about their medical history will also lead them to take a more active role in their medical care and will consequently lead to a decrease in medical error. Individual hospitals are responding to this new ruling at an individual pace.

Table 7.11–3.
Transaction Rule Code Sets

- Health care information: The Transaction Rule defines standards and establishes code sets and forms to be used for electronic transactions that involve the following health care information:
  - Claims or equivalent encounter information
  - Eligibility inquiries
  - Referral certification and authorization
  - Claims status inquiries
  - Enrollment and disenrollment information
  - Payment and remittance advice
  - Health plan premium payments
  - Coordination of benefits
- Code sets: Under the Transaction Rule, the following code sets are required for filing claims with Medicare:
  - Procedure codes
  - American Medical Association Current Procedural Terminology codes
  - Healthcare Common Procedure Coding System codes
  - Diagnosis codes
  - *International Classification of Disease* codes
  - Drugs and biologic
  - National Drug Codes
  - Dental codes
  - Code on dental procedures
  - Nomenclature for dental services

Adapted from Jaffe E. HIPAA Basics for psychiatrists. *Psychiatr Pract Manag Care*. 2002;8:15.

In 2002, the Transaction Rule was implemented, meaning that all entities covered under HIPAA must comply. The health care information involved and code sets are listed in Table 7.11–3. In 2003, the Privacy Rule was executed. Under this rule, patients have new statutory rights (Table 7.11–4).

Under the Privacy Rule, there are certain guidelines by which every practice must abide:

1. Every practice must establish written privacy procedures. These include administrative, physical, and technical safeguards that establish who has access to the patient's information, how this information is used within the facility, and when the information will and will not be disclosed to others.
2. Every practice must take steps to make sure that its business associates protect the privacy of medical records and other health information.
3. Every practice must train employees to comply with the rule.
4. Every practice must have a designated person to serve as a privacy officer. If it is an individual practice or private practice, this person can be the physician.
5. Every practice must establish complaint procedures for patients who wish to ask or to complain about the privacy of their records.

The OCR at HHS is responsible for making sure that the Privacy Rule is enforced. This enforcement is conducted by the government through a *complaint-driven* system in which the OCR responds to complaints made by patients concerning confidentiality violations or denials of access to records, all of which are covered under HIPAA. In such cases, OCR may follow up and audit compliance.

In 2009, the HITECH Act was signed into law as a subsection of the American Recovery and Reinvestment Act. The purpose of the HITECH Act was to "…promote the adoption and meaningful use of HIT" through incentive payments from Medicare and Medicaid for technology use. Examples of "meaningful use" include e-prescribing, completing electronic medication reconciliation, and updating electronic records on vaccination status. This act also led to strengthened breach notification requirements. In 2013, the final HIPAA Omnibus Rule was enacted, which increased penalties for HIPAA noncompliance to as much as $1.5 million per violation.

Table 7.11–4.
Patient's Rights under the Privacy Rule

Physicians must give the patient a written notice of their privacy rights, the privacy policies of the practice, and how patient information is used, kept, and disclosed. A written acknowledgment should be taken from the patient verifying that they have seen such notice.
Patients should be able to obtain copies of their medical records and request revisions to those records within a stated amount of time (usually 30 days). Patients do not have the right to see psychotherapy notes.
Physicians must provide the patient with a history of most disclosures of their medical history on request. There are some exceptions. The APA Committee on Confidentiality has developed a model document for this requirement.
Physicians must obtain authorization from the patient for disclosure of information other than for treatment, payment, and health care operations (these three are considered to be routine uses, for which consent is not required). The APA Committee on Confidentiality has developed a model document for this requirement.
Patients may request another means of communication of their protected information (i.e., request that the physician contact them at a specific phone number or address).
Physicians cannot generally limit treatment to obtaining patient authorization for disclosure of their information for nonroutine uses.
Patients have the right to complain about Privacy Rule violations to the physician, their health plan, or to the Secretary of HHS.

APA, American Psychiatric Association; HHS, Department of Health and Human Services.

Adapted from Jaffe E. HIPAA Basics for psychiatrists. *Psychiatr Pract Manag Care*. 2002;8:15.

The APA has created a HIPAA educational packet, which can be obtained on the APA website (http://www.psych.org/). On the site, there are also recommendations for enabling physicians to comply with HIPAA.

In 2010, the IOM formed the Committee on Patient Safety and HIT to gain a greater understanding of the impact of HIT on patient safety. In 2012, this committee submitted a report, *Health IT and Patient Safety:*

Building Safer Systems for Better Care, in which a disturbing trend was noted, which may negatively impact this goal. Specifically, it appears that some vendors—contracted by medical facilities to create EMR platforms—may be restricting the release of information regarding safety issues by citing contractual confidentiality clauses. This may, however, be leading to difficulties in determining where the problems currently are in HIT, and therefore, limits room for improvement.

Hospital Standards and Performance. A group of agencies, such as the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) and the Liaison Committee on Medical Education (LCME), influence the standards of hospital care and performance. In addition, hospitals must comply with government regulations (city and state health rules). The JCAHO inspects hospitals at least once every 39 months and is also responsible for determining the requirements for hospital accreditation. Hospital reimbursements from Medicare and Medicaid are contingent on meeting these standards, but the accreditation is done on a voluntary basis. The LCME and the Liaison Committee on Graduate Education are charged with accrediting medical schools and residency training programs, respectively. The two accrediting committees review education and training programs every 4 years; the procedure is voluntary.

The JCAHO is also involved in preventing medical errors and developed guidelines to help surgical teams standardize safety systems. The guidelines, called “The Universal Protocol for Preventing Wrong Site, Wrong Procedure, Wrong Person Surgery,” address the verification of documents, marking operative sites, and other issues related to patient safety.

The JCAHO issued a list of dangerous, unapproved abbreviations, acronyms, and symbols that accredited organizations must discontinue using as part of their National Patient Safety Goals requirements. The list of unapproved abbreviations pertains to all handwritten and electronic, patient-specific communication (Table 7.11–5).

The Institute for Safe Medication Practices (ISMP) released an updated List of Confused Drug Names in Feb 2019. To avoid medication errors, ISMP recommends strategies such as: “…using both the brand and generic names on prescriptions and labels; including the purpose of the medication on prescriptions; configuring computer selection screens to prevent look-alike names from appearing consecutively, and changing the appearance of look-alike product names to draw attention to their dissimilarities." The ISMP, along with the FDA, endorse the use of "tall man" (mixed case) letters as visual aids to avoid medication error (Table 7.11–6). The FDA has begun working with drug companies to select names for new products that differ significantly from those already on the market.

Table 7.11–5.
Confusing Medical Notations

Prohibited Abbreviation	Potential Problem	Preferred Term
U (for unit)	Read as a zero (0) or a four (4), causing a 10-fold overdose or greater (4 U seen as "40" or 4 u seen as "44").	Write "unit."
IU (for international unit)	Misread as "IV" (intravenous) or 10.	Write "international unit."
Q.D. and Q.O.D. (Latin for once daily and every other day)	Mistaken for each other. Also, the period after the Q can be mistaken for an I and the O can be mistaken for an I.	Write "daily" or "every other day."
Trailing zero (X.0 mg) and lack of leading zero (.X mg)	Decimal point is missed.	Never write a zero by itself after a decimal point (X mg) and always use a zero before a decimal point (0.X mg).
MS, MSO_4, and $MgSO_4$	Confused for one another. Can mean morphine sulfate or magnesium sulfate.	Write "morphine sulfate" or "magnesium sulfate."

Table 7.11–6.
List of Confused Drug Names Illustrating Tall Man Lettering Practices[a]

Drug Name	Confused Drug Name
MetaDATE CD (methylphenidate)	methaDONE (Dolophine)
LEVOTHYroxine	LANoxin (digoxin)
Serzone (nefazodone)	SEROquel (quetiapine)
LaMICtal (lamotrigine)	LamISIL (terbinafine) (for nail infection)
ZyPREXA (olanzapine)	ZyrTEC (cetirizine) (for allergies)
CeleXA (citalopram)	CeleBREX (celecoxib) (for arthritis)

[a] Note how the same drugs from Table 7.11–1 are listed, but using tall man lettering format that promotes more accurate drug identification.

Adapted with permission, Copyright 2019, Institute for Safe Medication Practices. www.ismp.org. 5200 Butler Pike, Plymouth Meeting, PA 19462. This Material is protected by copyright laws and may not in whole, in part, or by reference be used in any advertising or promotional material, or to compete with ISMP.

Hospital Utilization Review. This in-house evaluation process was created to make sure that institutions provide efficient, quality health care that meets patients' needs. The members of the utilization review committee consist of hospital administrators, physicians, and nurses. The committee reviews each patient's chart within a specified number of days of admission. The appropriateness of the admission, treatment strategies, and the length of the hospital stay is reviewed to facilitate the patient's discharge. Through this process, the utilization review committee determines whether a particular admission was really indicated and whether the hospital stay was longer than necessary. A hospital must conduct utilization reviews to be eligible for JCAHO accreditation.

Professional Standards Review Organization. The Professional Standards Review Organization (PSRO) was set up by the federal government to review and monitor care received by patients whose care is paid for with government funds. PSROs are made up of physicians elected by local medical societies that have been established by local medical associations and serve several functions: They attempt to ensure high-quality care, control costs, determine maximum lengths of stay by patients in hospitals, conduct utilization reviews, and censure physicians who do not adhere to established guidelines. A PSRO may conduct a medical audit to evaluate the quality of care retrospectively by carefully examining charts.

Peer Review Organization. In the early 1980s, the peer review organization (PRO) replaced PSROs as the federal review organization for hospitals receiving Medicare funds. To promote compliance with federal guidelines for health and hospital care, PROs conduct independent utilization reviews and quality-of-care studies, validate diagnose-related group (DRG) assignments, and review hospital admissions and readmissions.

Federally mandated and funded, the PROs had greater authority than the PSROs. PROs were able to impose sanctions on hospitals for inadequate care; they could even recommend the termination of federal funding to hospitals that consistently violated federal standards. In addition, PROs could adjust or refuse payment for health services that they considered unnecessary.

Quality Improvement Organization. In May of 2002, quality improvement organizations (QIOs) replaced PROs. Funded by the federal government through the CMS, the QIOs are state-based organizations with three main aims: to provide safer and more effective care for patients, to improve the overall health of patients, and to make healthcare more affordable. IPRO, New York’s QIO, released their Quality Improvement Report for New York State focused on the years 2011 to 2014. One goal which was achieved and highlighted in the report related to IPRO’s success in helping medical practices integrate HIT for meaningful use and improved care coordination. In 2019, CMS awarded IPRO a 4-year contract to assist smaller hospitals with a number of initiatives, including patient safety.

Centers for Medicare and Medicaid Services. In an attempt to improve patient safety and reduce complications, Medicare stopped paying for some hospital-acquired conditions. These include errors such as an object left in the patient during surgery, blood incompatibility, decubiti, and air embolism, among others. In some states, such as New York, Medicaid has stopped paying for “never events,” once described by the former New York State Health Commissioner, Richard Gaines, as “…a list of things that, frankly, should never happen." The New York State Never events list includes medication errors, patient disability from the use of restraints or bedrails, and surgery performed on the wrong body part. CMS believes that financial disincentives will influence hospitals to provide more aggressive quality reviews and be better able to address issues effectively. Critics of the program claim that some errors are really adverse events citing bed sores as an example of a condition that cannot always be avoided, even with the most highly skilled nursing care.

FURTHER READINGS

Agency for Healthcare Research and Quality Website. *20 Tips To Help Prevent Medical Errors: Patient Fact Sheet*. Agency for Healthcare Research and Quality; 2020. https://www.ahrq.gov/questions/resources/20-tips.html

Allen S, Caton C, Cluver J, Mainous AG 3rd, Clyburn B. Targeting improvements in patient safety at a large academic center: an institutional handoff curriculum for graduate medical education. *Acad Med*. 2014;89(10):1366–1369.

Aspden P, Wolcott J, Bootman JL, Cronenwett LR. *Institute of Medicine Committee on Identifying and Preventing Medical Errors. Preventing Medical Errors*. National Academies Press; 2007.

Bal G, Sellier E, Tchouda SD, François P. Improving quality of care and patient safety through morbidity and mortality conferences. *J Healthcare Qual*. 2014;36(1):29–36.

Bari A, Khan RA, Rathore AW. Medical errors; causes, consequences, emotional response and resulting behavioral change. *Pak J Med Sci*. 2016;32(3):523–528.

*Bell SK, Mann KJ, Truog R, Lantos JD. Should we tell parents when we've made an error? *Pediatrics*. 2015;135(1):159–163.

*Bernat JL. Ethical and quality pitfall in electronic health records. *Neurology*. 2013;80(11):1057–1061.

Blazin LJ, Sitthi-Amorn J, Hoffman JM, Burlison JD. Improving patient handoffs and transitions through adaptation and implementation of I-PASS across multiple handoff settings. *Pediatr Qual Saf*. 2020;5(4):e323.

Chathampally Y, Cooper B, Wood DB, Tudor G, Gottlieb M. Evolving from morbidity and mortality to a case-based error reduction conference: evidence-based best practices from the Council of Emergency Medicine Residency Directors. *West J Emerg Med*. 2020;21(6):231–241.

Connolly W, Li B, Conroy R, Hickey A, Williams DJ, Rafter N. National and institutional trends in adverse events over time: a systematic review and meta-analysis of longitudinal retrospective patient record review studies. *J Patient Saf*. 2021;17(2):141–148.

Cottney A, Innes J. Medication-administration errors in an urban mental health hospital: a direct observation study. *Intern J Ment Heal Nurs*. 2015;24(1):65–74.

*Cullen SW, Nath SB, Marcus SC. Toward understanding errors in inpatient psychiatry: a qualitative inquiry. *Psychiatr Q*. 2010;81:197–205.

DeRienzo CM, Frush K, Barfield ME, et al. Handoffs in the era of duty hours reform: a focused review and strategy to address changes in the accreditation council for graduate medical education common program requirements. *Acad Med*. 2012;87(4):403–410.

Heaton HA, Campbell RL, Thompson KM, Sadosty AT. In support of the medical apology: the nonlegal arguments. *J Emer Med*. 2016;51(5):605–609.

Houston M. The psychiatric medical record, HIPAA, and the use of electronic medical records. *Child Adol Psych Clin N Am*. 2010;19:107–114.

Institute of Medicine. *Health IT and Patient Safety: Building Safer Systems for Better Care*. The National Academies Press; 2012.

IPRO website. Accessed May 30, 2021. https://ipro.org/

*James JT. A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf*. 2013;9(3):122–128.

Kass JS, Rose RV. Medical malpractice reform—historical approaches, alternative models, and communication and resolution programs. *AMA J Ethics*. 2016;18(3):299–310.

Klein CA. Cloudy confidentiality: clinical and legal implications of cloud computing in health care. *J Am Acad Psych Law*. 2011;39:571–578.

Kohn LT, Corrigan J, Donaldson MS. *Institute of Medicine Committee on Quality of Health Care in America. To Err is Human: Building a Safer Health System*. National Academies Press; 2000.

Mueller SK, Sponsler KC, Kripalani S, Schnipper JL. Hospital-based medication reconciliation practices: a systematic review. *Arch Intern Med*. 2012;172(14):1957–1969.

Murray TL, Calhoun M, Philipsen NC. Privacy, confidentiality, HIPAA, and HITECH: implications for the health care practitioner. *J Nurs Prac*. 2011;7(9):747–752.

O'Connor N, Kotze B, Wright M. Blame and accountability 1: understanding blame and blame pathologies. *Australas Psychiatry*. 2011;19:113–118.

Ober KP, Applegate WB. The electronic health record. Are we the tools of our tools? *Pharos Alpha Omega Alpha Honor Med Soc*. 2015;78:8–14.

ONC's Cures Act Final Rule – The Office of the National Coordinator for Health Information Technology. Accessed July 31, 2023. https://www.healthit.gov/topic/oncs-cures-act-final-rule

Piccardi C, Detollenaere J, Bussche PV, Willems S. Social disparities in patient safety in primary care: a systematic review. *Int J Equity Health*. 2018;17:114.

Procyshyn RM, Barr AM, Brickell T, Honer WG. Medication errors in psychiatry: a comprehensive review. *CNS Drugs*. 2010;24(7):595–609.

Ryder HF, Huntington JT, West A, Ogrinc G. What do I do when something goes wrong? Teaching medical students to identify, understand, and engage in reporting medical errors. *Acad Med*. 2019;94(12):1910–1915.

Soerensen AL, Lisby M, Nielsen LP, Poulsen BK, Mainz J. The medication process in a psychiatric hospital: are errors a potential threat to patient safety? *Risk Manag Healthc Policy*. 2013;6:23–31.

Starmer AJ, Spector ND, Srivastava R, et al. Changes in medical errors after implementation of a handoff program. *N Engl J Med*. 2014;371(19):1803–1812.

Szczepura A, Wild D, Nelson S. Medication administration errors for older people in long-term residential care. *BMC Geri*. 2011;11(82):1–10.

Thomas L, Cordonnier-Jourdin C, Benhamou-Jantelet G, Diviné C, Le Louët H. Medication errors management process in hospital: a 6-month pilot study. *Fundam Clin Pharmacol*. 2011;25:768–775.

Topazian RJ, Hook CC, Mueller PS. Duty to speak up in the health care setting: a professionalism and ethics analysis. *Minn Med*. 2013;96(11):40–43.

Vaidotas M, Yokota PKO, Negrini NMM, et al. Medication errors in emergency departments: is electronic medical record an effective barrier? *Einstein (Sao Paulo)*. 2019;17(4):eGS4282.

*Wittich CM, Burkle CM, Lanier WL. Medication errors: an overview for clinicians. *Mayo Clin Proc*. 2014;89(8):1116–1125.

8

Clinical Manifestations of Psychiatric Disorders

ANU A. MATORIN, M.D., ANDRES A. AVELLANEDA OJEDA, M.D., ASIM A. SHAH, M.D., AND PEDRO RUIZ, M.D.

Manifestations of psychiatric disorders express themselves as variations of normal functioning from subtle to obvious, making it confounding for the clinician to integrate data into an accurate diagnostic picture. The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) has placed a strong focus on integration of symptom data, including patients' self-assessments of symptom severity with a goal to improve patients' overall satisfaction and treatment outcomes.

The DSM-5 approach aims to provide an even more thorough evaluation and comprehension of the patient's interpersonal world, work role, family life, and culture than is typical in general medical practice. This dimensional model will introduce a greater amount of information for the clinician to assess, evaluate, and measure, including the severity of psychiatric conditions. Despite the introduction of DSM-5 a few years ago, the lack of scales to screen and assess psychiatry symptoms or conditions continues to be the same as 20 years ago and is therefore primarily driven by the clinician's individual clinical assessment.

The nature and expression of psychiatric signs and symptoms are profoundly altered by the patient’s strengths, coping capacities, psychological defenses, and situational context so that the clinical picture ultimately represents a balance between psychopathology and psychological strengths. It is important to integrate substance use and culture background as part of initial presentation and subsequent treatment planning and follow-up. The clinical symptomatology must also be understood as part of the individual’s attempts to cope with the biologic, psychological, and environmental challenges that they face. When attempts to cope overwhelm the individual’s capacity to respond, less adequate, more disorganized, and ineffective thoughts and behaviors emerge, and these comprise the impairments that present in clinical situations.

In addition, the coronavirus pandemic has led to the introduction of Covid infections with implications in different systems including neurosensory symptoms, long-term effects after being infected, as well as social isolation, depression, and anxiety with phobias, fears, and worries about getting infected with Covid-related illness.

NEED FOR A COMPREHENSIVE CLINICAL PERSPECTIVE

DSM-5, like other medical disease classifications, uses multidimensional approaches to mental disorders, as the previous too-rigid categorical system was unable to capture clinical experience or important scientific observations. The goal of this approach is to capture a more accurate description of patient presentations and for the diagnosis to be more valid.

The DSM-5 dimensional model addresses the longstanding concerns that clinicians often tend to limit their fields of vision and appreciate only part of the available information; the clinician’s theoretical orientation and other personal and cultural factors also limit what is perceived. Research has demonstrated that clinicians tend to perceive primarily those signs and symptoms that are most in accord with their theoretical points of view and with the tools that they have available to treat psychiatric disorders, a phenomenon known as concept-driven perception. Such differences may lead one clinician to see a major mood disorder to be treated with medication, whereas another sees a pervasive personality problem with depressed mood to be treated with psychotherapy and to use different technical terms to label roughly the same phenomena. Psychodynamic-oriented psychiatrists might see psychomotor retardation, whereas a neuropsychiatrist sees bradykinesia; a psychodynamicist might see depressed affect and muted speech, whereas a neuropsychiatrist sees masklike facies and aprosodic speech.

The intermittent nature of many psychiatric signs and symptoms; the potential unreliability, selective recall, and false remembering of patients and others in reporting symptoms and events; differing interpretations of elicited information or observations; and subjective theoretically driven biases that influence the clinician's perception of signs and symptoms all contribute to potential errors in data collection. To help guard against misinformation and simplistic understandings and formulations, whenever possible, complete assessment of a psychiatric patient requires consultation with family, friends, coworkers, and other professional observers to enrich the history and to provide supplemental observations of the patient over time. Finally, via the DSM-5 dimensional assessment model, the comprehensive gathering of clinical symptoms will be the foundation for greater standardization, integration, and monitoring of treatment planning and outcomes.

THE CLINICAL CHALLENGES

DSM-5 development involved physicians, psychologists, social workers, nurses, counselors, epidemiologists, statisticians, neuroscientists, and neuropsychologists. This version even had feedback from patients, families, lawyers, consumer organizations, and advocacy groups. However, one of the costs of this increasing reliability has, in many instances, been the narrowing of the field of clinical vision. That is, clinicians who rely predominantly on structured interviews and checklists may become somewhat closed-minded and risk ignoring clinical phenomena that are very important but that may not be part of the structured interview framework or mental set. An additional related difficulty can be seen in the use of nonclinical research staff that are trained to interview individuals using structured interviews (as in large epidemiologic studies). In these instances, the need for interrater reliability is often achieved at the expense of a more nuance-based appreciation of the different meanings of subjective sensations, especially when the interviewer has no clinical experience with which to interpret the subject's descriptions.

However, in spite of these practical challenges, a thorough assessment of the clinical history and description of the psychopathology and detailed account of the patient's subjective experiences are important for the following reasons:

1. Significant diagnostic distinctions are made primarily on the basis of the historical information and elicited phenomenology. The more detailed, complete, and correct the diagnosis, the more rational and precise the treatment planning and the more reliable the prognosis. Consider, for example, the importance of accurately distinguishing between neuroleptic-induced akathisia versus anxiety symptoms related to psychotic thinking. On the basis of which of these diagnoses the clinician selects, the therapeutic strategy might be diametrically opposite.
2. The clinician's capacity to fully hear and communicate a comprehensive understanding of the patient's internal experiences helps to diminish the patient's sense of isolation so characteristic of many of these disorders and fosters the growth of a therapeutic alliance, increasing the likelihood of treatment adherence.

The Reliability Dilemma

As mentioned, for a given clinical manifestation of a psychiatric disturbance, different clinicians may arrive at different conclusions regarding the same clinical symptoms in the same patient. This reliability dilemma may confound an objective clinical assessment. Many factors contribute to this phenomenon. For example, some discrepancies may be due to differences in the patient's status or in information imparted by the patient from examination to examination, in the clinician's definitions of the symptoms or signs in question, and in perceiving and interpreting the patient's responses to general presentation or questions within the interview. A substantial amount of information variance in clinical practice reflects the impossibility of asking all questions within the time available for clinical evaluation. As an example, in one study, the rate of sexual side effects increased from 14% when depending on spontaneous patient report to 58% when the interviewer directly asked about these side effects.

A 52-year-old schoolteacher with recurrent major depression presented to her treating psychiatrist with her fiancé of 6 months. As part of the visit, the patient requested if she could stop her antidepressant medication, even though she had multiple severe relapses in the past. With her fiancé's encouragement, the patient disclosed arousal difficulties and anorgasmia after she started the medication. The patient had disclosed many painful memories related to past trauma to her psychiatrist, but she reported feeling shy and embarrassed to mention the sexual side effects. Upon further questioning, the patient reluctantly disclosed that she was terrified she would lose her fiancé if she continued taking her current antidepressant and did not know how to address this issue impacting her personal life.

Even when simply responding to direct questions about symptoms, patients may respond differently depending on the interviewer's manner, how the questions are asked, their personal sense of trust or safety, whether they have answered these questions before, the amount of cuing that may signal the "desired" response, their fatigue, or a host of other variables.

Most clinicians still rely heavily on their own clinical intuition and subjective responses to patients as part of a diagnostic assessment.

However, these clinical inferences, whether accurate or not, are often based on nonconscious assumptions, comparisons with other patients not well remembered, or distortions based on the clinician's own personal experiences. When the bases for these intuitions can be identified and described clearly, they may prove to be reliable and valid. However, intuitions are often wrong—simple trust in intuition alone is not sufficient. Thus, a clinician's sense that a patient is angry and potentially violent may result from the patient's subtle (but verifiable) body language and tone of voice, or it may represent a countertransference distortion that is not prompted by any observable patient behavior. At times, clinicians often too quickly label behaviors as inappropriate when they do not appreciate and understand contextual or cultural considerations. Appropriateness depends heavily on context, and definitions of what is proper in a given context may also be highly subjective. Appropriate behavior (or dress) in some parts of California may be inappropriate in Boston. Furthermore, the rapid change and growth of the current multicultural society places cultural understanding and context in psychiatric history taking and assessment in the forefront. Even manifestations of low intensity of emotional expression leading to a clinical description of "constricted affect" may reflect cultural norms or a psychopathological state.

Ambiguity in the Assessment of Psychiatric Signs and Symptoms

Since, at the present time there are no clear-cut diagnostic, radiologic, or laboratory tests to make a psychiatric diagnosis, virtually all psychiatric symptoms can be ambiguous or nonspecific and are usually seen in many different disorders. Depressed mood, for example, occurs in a wide variety of diagnostic groups, including major depressive disorder, schizophrenia, some personality disorders, mood disorders secondary to medical or substance-related states, and so on. Even the so-called first-rank symptoms of schizophrenia described by Kurt Schneider are diagnostically nonspecific. They are seen with some frequency in otherwise classic depressive and bipolar disorders. Apathy offers another good example of an important nonspecific phenomenon. Although apathy is often part of depression, research has shown this symptom to be a clinically distinct syndrome. Apathy is marked by lack of spontaneity, initiation, and emotionality as well as lack of activity and interest in friends, family, and hobbies. Apathy and the associated symptom of amotivation may exist in their own right without the presence of significant depression, as in initial phases of abstinence from cocaine and in other syndromes characterized by reduction in dopaminergic tone, such as frontotemporal dementias, Parkinson disease, or progressive supranuclear palsy.

In general medicine, symptoms not recognized as part of a clearly defined syndrome are often described as being of unknown origin. Thus, a fever that cannot be ascribed to a known disorder, such as pneumonia, is described as a fever of unknown origin. Given the nonspecific nature of psychiatric symptoms, it seems wise to use similar conventions, referring to hallucinations of unknown origin or depressed mood of unknown origin when a symptom cannot be clearly linked to a well-described syndrome. In DSM-5, the “not otherwise specified” (NOS) is replaced by *other specified disorder* and *unspecified disorder.* In the first category, the clinician must specify why the diagnosis doesn’t fall under any specific category within a diagnostic class. If the clinician chooses not to specify the reason, then they will check unspecified class. Even though individual signs and symptoms may be organized into syndromes and disorders, they often have courses of their own. Thus, in the appearance or the resolution of a disorder, certain associated signs and symptoms may appear very early or may persist after all the others have waned. For example, in the restricting form of anorexia nervosa, excessive exercise is often the first symptom to appear and the last to abate even after dieting has stopped. In some cases, certain signs and symptoms that are commonly associated with a given disorder may not appear. Each sign and symptom may have its own pattern and variable response to treatment. In the treatment of schizophrenia, for example, some patients experience rapid resolution of hallucinations but have persistent delusions without ever having any other thinking disturbances, whereas others may have no residual hallucinations or delusions but still have prominent thinking disturbances.

Predisposition and Vulnerability in Psychiatric Symptomatology: Genetic and Intrauterine Considerations

Genetic factors are key in determining the expression of most if not all psychiatric disorders. Prominent among these are dementias of the Alzheimer type, schizophrenia, unipolar and bipolar mood disorders, anxiety disorders, alcohol use disorder, and some personality traits. For virtually all of these disorders and traits, what is largely unknown at this point is the nature of the inherited vulnerabilities.

Intrauterine factors contribute to many psychiatric disorders. For example, maternal starvation and influenza infections during the second trimester of pregnancy have been implicated in the pathogenesis of schizophrenia. Maternal smoking and low birth weight may be risk factors in the pathogenesis of attention-deficit disorders in children. Maternal alcoholism may lead to fetal alcohol syndrome, a major cause of developmental disability.

Fundamental Neurobiologic Factors

Investigational studies have shown that shortly after birth infants vary widely in temperament. For example, differences have been noted in spontaneous activity levels and thresholds, intensity, and duration of their reactions to external stimuli; the regularity or irregularity of certain biologic rhythms such as sleep; tendencies to approach or withdraw from new stimuli; the speed and degree of adaptation; attention span and distractibility; the persistence of behavior; and qualities of mood. On the basis of such early behaviors, children may be described as having easy or difficult temperaments, quick or slow to warm up. Nevertheless, these temperamental qualities correlate somewhat with behavioral problems, at least through early childhood.

Aside from temperament, other persistent normal variations in personality development seem to be constitutionally related and may influence subsequent resilience or vulnerability. Traits, such as introversion, extroversion, and neuroticism, appear to be relatively enduring and stable personality dimensions. Other temperamental qualities that endure include novelty seeking, being relatively open to new experiences, and stick-to-it-iveness. Subtypes of intelligence, such as those related to conceptual, mathematical, musical, kinesthetic, and interpersonal abilities, have been postulated as having separate genetic determinants and patterns of development. The type A and B personality patterns, hardy and resilient personalities, high-strung, sensitive, fussy, irritable, and pessimistic characteristics have all been described as generally lifelong qualities that originate in early childhood. Even dimensions of character, referring to concepts about the self in relation to others that develop over time through social learning and maturation of interpersonal behavior, have been shown to have moderate heritability. These include such qualities as self-directedness, the ability to engage cooperatively with others, and the capacity to “transcend” the self by developing a sense of one’s place or purpose in the larger social context.

Other characteristics relevant to psychopathology but not diagnostically specific may also be inherited. These include psychosis proneness, cognitive styles (such as obsessionality or detail orientation), and emotional reactivity (which is part of the dimension of neuroticism). The relationship between the inheritance of these traits and the patterns of transmission of specific psychiatric disorders is unknown.

Impact of Physical Illness

All of the metabolic, toxic, infectious, and other causes of physical illness produce increased vulnerability to psychiatric disturbance. Studies have shown higher use of psychiatric services by those who are physically ill and higher than expected prevalence of physical disease among the psychiatrically impaired.

Some children with obsessive-compulsive disorder (OCD) and tic disorders with symptoms that appear during the prepubertal period and whose course of illness becomes episodic have been found to have pediatric autoimmune neuropsychiatric disorders associated with streptococcal (group A β-hemolytic streptococcus) (PANDAS) infections. Accompanying symptoms during episodes of exacerbation are emotional lability, separation anxiety, nighttime fears and bedtime rituals, cognitive deficits, oppositional behaviors, and motoric hyperactivity. In PANDAS, patients' flare-ups of behavioral problems are commonly associated with documented group A β-hemolytic streptococcus infections or symptoms of pharyngitis and upper respiratory infections.

Human immunodeficiency virus (HIV) infection leading to acquired immunodeficiency syndrome (AIDS) vividly illustrates the multiple and complex ways in which stressors can lead to psychiatric disturbances. These patients' psychiatric symptoms may represent central nervous system (CNS) changes that are the direct effects of the virus on the CNS producing changes in cognition, personality, and mood, expectable psychological adjustment responses of the patients in response to an overwhelming life-threatening disorder, or the emergence of latent or quiescent primary psychiatric problems provoked by the psychological stress of the viral illness. Adding to the complexity, some individuals with cluster B personality traits may engage in excessive risk-taking behaviors, increasing the likelihood of viral exposure; they may then be at higher risk for developing psychiatric symptoms or disorders in response to being HIV positive because of their pre-existing cluster B traits.

A 31-year-old accountant presented to the ER with report of “waves of anxiety,” including “panic feelings,” palpitations, and dizziness. He reported he had started a new job and was anxious and stressed out. Although he had no prior medical history, he reported new-onset headaches, nausea and “chest wall pain concomitant with ‘waves of anxiety’.” The psychiatrist on duty obtained a medical consult. The patient was eventually diagnosed with a pheochromocytoma.

Effect of Life Events

The impacts of life events, particularly those that are negative and outside of a given individual’s control, have complex effects on the development and persistence of psychiatric symptomatology. In general, the likelihood of developing psychiatric manifestations is increased in those who have a pre-existing psychiatric disorder. After exposure to the same negative stressors, such as a serious accident or act of violence, individuals who previously had anxiety disorders are more prone than those without such histories to subsequently develop symptoms of posttraumatic stress disorder (PTSD). Other factors, such as gender, may also predict risk of developing PTSD, with women more likely to develop the disorder given the same exposure to stress. Although individual responses vary widely, truly catastrophic events, such as incarceration in a concentration camp, cause enduring psychiatric disturbances in a high percentage of survivors regardless of whether they had prior psychiatric problems. Similarly, the stress-related consequences of combat also vary widely, so that some heavily combat-exposed veterans develop long-lasting PTSD, whereas others develop very few persistent symptoms. The death of a parent or spouse, divorce, and major physical injury affect some people profoundly and others hardly at all in the long run. Significant stressors are likely to be more traumatic during early development rather than later or at certain critical developmental periods, compared with other times. For example, the loss of a parent at a very young age is likely to be more traumatic and have more profound and lasting effects than the loss of a parent as an adult.

The combined impact of negative life events and poor emotional and practical social support is important in predicting the emergence of at least some psychiatric disturbances. People who are ordinarily very competent in all role functions may fall apart completely when a supportive spouse who has bolstered them and taken care of many of their needs suddenly dies. Patients presenting with a major depressive episode have experienced more uncontrollable actual and threatened losses such as the death of a spouse in the year before onset.

A 65-year-old woman presented to her primary care physician with her daughter after her neighbors told her daughter that they had noticed that she was more withdrawn and staying in her house for days without checking the mail or watering her favorite flower garden. The patient reported fatigue, “heavy leaden” arms and body, poor appetite, and poor sleep. The patient’s husband of almost 40 years had suddenly passed away 1 year ago. The patient plaintively stated, “We had made so many plans in retirement, and now they are all gone.”

Nevertheless, not all psychiatric disturbance is attributable to easily identified, provoking negative life events; indeed, some major negative life events that at first glance appear to have preceded the onset of a serious psychiatric disturbance may, in fact, have occurred only after the psychiatric disturbance actually began. For example, someone who attributes the onset of depression to having been fired from a job several months previously may already have been functioning suboptimally at that time and may have been fired as a consequence of a depression-induced decline in role function.

Certain environmental features can counter the effects of environmental stressors and protect against psychiatric disturbances. Stable families and friends, good financial circumstances, and supportive religious and social communities offer some protection. Research has shown that individuals with psychiatric disturbances have fewer social supports than normal controls. This may be due to friends' and relatives' withdrawal from deviant behaviors or to the disturbed individual's withdrawal from deleterious family and social relationships. In contrast, physically ill people have more social supports than others, perhaps reflecting their ability to recruit help in times of need. Of course, the quality, as well as the quantity, of social supports is important. As has been demonstrated in schizophrenia and mood disorders, for example, negative relationships even in close families may have deleterious effects, both in developing and in sustaining psychiatric disturbance.

The negative impact of a physiologic or environmental stressor is closely related to its personal meaning. For example, the loss of a spouse who has been chronically demented, disabled, and burdensome ordinarily has a very different impact than the loss of a vital, supportive, loving spouse.

Finally, a relevant study has demonstrated the interaction of genetic vulnerabilities with adverse life events. In a study involving several hundred young adults followed prospectively, genetic vulnerability to depression was conferred by certain polymorphisms of the serotonin receptor. Over a period of several years, individuals with one or two copies of the short allele of the 5-hydroxytryptamine T promoter polymorphism exhibited more depressive symptoms, diagnosable depression, and suicidality in relation to stressful life events involving finances, housing, employment, and relationships than individuals homozygous for the long allele. This research provides strong evidence for gene–environment interactions, in which a person's responses to environmental insults are moderated by their genetic makeup.

DISTINGUISHING FEATURES OF PSYCHIATRIC SIGNS AND SYMPTOMS

Circumstance

Psychiatric signs and symptoms are usually not static entities; depending on the circumstances, they often vary in intensity or even in their existence. The depressed mood of a melancholic depression may persist regardless of external situations, whereas the depressed mood in milder, reactive depression may vanish completely during certain situations—including a psychiatric interview—only to reappear at other times. Signs and symptoms that occur only in specific settings or with certain internal states are referred to as state dependent. For example, certain hallucinations or memories may be present only during states of drug or alcohol intoxication; in some patients, hives may erupt as a psychophysiological response only during states of anger. Interpersonal context is also important. Some people become violent only when involved in sadomasochistic relationships or in certain group settings such as adolescent gangs. In gangs, social pressures for conformity and expectations for aggressive behavior may provoke pathologic behaviors that might otherwise never be expressed by gang members individually.

Cultural Context

Psychiatric signs and symptoms cannot be assessed independent of an individual's background and culture. Many phenomena often considered to be symptoms of psychiatric disorders may not be experienced as psychiatric problems by patients. Hearing an angel's voice may represent a manifestation of a psychotic disorder, yet the patient may vigorously dispute that the experience is a psychopathological symptom. In mainstream American culture, for instance, any experience of thoughts being broadcast out loud is considered pathologic. These signs and symptoms must all be considered in context: exactly what constitutes normal varies from culture to culture and from situation to situation. As mentioned earlier, behavior or subjective experience that may be defined as symptomatic in one context may be perfectly acceptable and within normal bounds in another. A phenomenon should be considered abnormal only if it seems deviant within the patient's specific culture after its full physiologic and environmental context is taken into account and if it causes personal or interpersonal impairment. Too often, phenomena prematurely mislabeled as psychopathology turn out to be perfectly understandable and nonpathologic once the whole situation is appraised. Conversely, some examiners are loath to label certain phenomena as psychopathological even when they clearly are, for fear of stigmatizing the patient.

Within cultures, most interpersonal interactions are carefully regulated by tight sets of rules and controls and constrained by reasonably well-defined sets of expectations and acceptable limits.

An 18-year-old woman was brought to the emergency center by her youth minister. The youth minister stated that his youth group had returned from a week-long religious retreat. The young woman had been looking forward to the retreat for months. At the retreat, she volunteered to lead the singing of religious songs, had the “gift of tongues,” and spoke in tongues to lead the meeting. The youth group was entranced by her fervor and religious dedication. However, for the last 3 days, she had stopped sleeping, was found kneeling alone at night in the wet ground, had started “fasting” for repentance. Expressing concern for this change in her behavior, the youth minister had contacted her parents and immediately brought her to the nearest hospital.

Even slight deviations from these acceptable limits are quickly perceived by laypeople, as well as professionals, because behavioral deviances are often experienced as threats. Deviations in amplitude, duration, and intensity can occur in facial expressions, gestures, postures, vocalizations, language, and other expressions of emotion and thought. A small increase in the rate of speech, an intrusion into one person’s conversation by another who does not allow proper pauses, a gesture that comes just a bit too close to a face, an excessively rigid or distant stance, or a gaze that is too staring or too avoidant each signals social insensitivity and alerts the observer to deviant behavior.

Sign and Symptom Classifications

The classification of clinical signs and symptoms in psychiatry has been accomplished using several different schemas: state versus trait, primary versus secondary, and form versus content. The state versus trait distinction refers to whether the sign or symptom is an enduring characteristic of the person ("traits") or time-limited phenomena associated with specific Axis I disorders, which are usually state phenomena. While DSM-5 does not use a multi axial system for diagnosis, and while personality disorders are still included as they were in DSM-IV-TR, an alternative model in DSM-5 has been included as Section III in the new DSM-5. As in clinical practice distinctions are often fuzzy, and comorbidity among the so-called categorically distinct disorders is often the rule rather than the exception, DSM-5 uses a single axial approach.

DSM-5 has changed the classification of schizophrenia by deleting all subtypes, but has kept all other types, like schizoaffective in the diagnosis. In addition, a dimensional view of psychopathology fits much recent data better than the categorical view that is inherent in the DSM-IV-TR. That is the basis of DSM-5 changes for personality disorders also. For instance, personality disorders fit poorly into a categorical scheme, and the frequent "comorbidity" of these disorders likely reflects the descriptive overlap rather than the patient having two distinct disorders. Similarly, in the DSM-IV-TR, dysthymia and major depression are seen as two different mood disorders, while recent studies indicate that they are more likely manifestations of one disorder that differs in course and intensity. DSM-5 not only deleted the diagnosis of dysthymia, but also combined it with major depression, as persistent depressive disorder, with partial or full remission, early or late onset, and mild, moderate, or severe as just a few of its specifiers.

Impairments and Adversities

In addition to the traditional signs and symptoms of psychiatric disorders, recent focus has been on the impairments and adversities that psychiatric signs, symptoms, and disorders create in affecting specific role functions and causing social and economic burdens for the patient and others. These problems and impairments often cut across traditional sets of signs and symptoms of which categorical diagnoses are comprised, affecting, for example, basic abilities to care for oneself and one's family, marital functioning, child rearing, wage earning, school performance, and social behavior. They constitute the issues with which patients and families contend, and they need to appear on the problem lists that treatment plans and specific interventions target. Studies reveal, for example, that the impairments imposed by major depression are considerable with regard to physical functioning, role limitations, and social functioning. Furthermore, the relationship between symptoms and disorders on one hand and functional impairments on the other is not always straightforward. For example, in both bipolar and unipolar mood disorders, many patients recover symptomatically from episodes but achieve premorbid psychosocial function either months later or not at all. In one review of previously published studies on major depression, there was a consistent lag time between symptomatic and functional recovery. Whether this disparity between symptomatic and functional recovery reflects subtle residual symptoms, unrecognized cognitive disturbances, personality difficulties, or a combination of factors is unknown.

A 30-year-old woman with postpartum depression was reluctant to return to work as a fashion buyer. She felt she just "could not return to work this fast," in spite of marked improvement in her sleep, mood, energy level, and remission of her crying spells. She was treated with cognitive behavior therapy and an antidepressant, which she felt helped her very much. Yet, she felt "paralyzed" to perform in a job that she competitively had sought out prior to her pregnancy.

THINKING DISTURBANCES

Normal Thinking

Thinking is defined as the mental activity and processes used to imagine, appraise, evaluate, forecast, plan, create, and will. Most thought involves complex rules that are probably best currently approximated (albeit inexactly) by fuzzy-logic decision-making algorithms that use neural net technology, increasingly applied by scientists and engineers in vague situations in which all-or-none, black-or-white thinking does not apply but in which multifaceted, contradictory, and competing possibilities and biases are the rule. Most of what is known about thinking derives from the study of language as the product (and reflection) of thought, yet a great deal of thinking takes place preverbally and nonverbally. Thinking occurs in images, music, and kinesthetic sensations and in symbols other than linguistic ones. Attempts to transmit preverbal and nonverbal thought using only words are frustrating and unsatisfactory. Creative artists have considerable difficulty describing the inner states of tension and inchoate awareness from which ideas are distilled.

Ordinary thought is far from logical. Streams of conscious thought are intruded on by competing thoughts and associations and by outside stimuli, and attention is easily distracted. Ordinary conversation is marked by recurring asides, interruptions, delays, and the loss of ideas. Decisions are often made on the basis of very few cues and inadequate evidence: people jump to conclusions. Beliefs are zealously held that are not supported by evidence. Thinking in stereotypes is more common than thinking in logical categories; from an evolutionary perspective, thinking in stereotypes and by approximation has probably been more adaptive than thinking in strictly defined categories. This tendency helps account for clinicians' tendencies to make diagnoses by approximation and intuition based on prototypes and to feel less comfortable using formal lists of criteria found in statistical manuals such as the DSM-IV-TR. Consistent with this notion, a recent study found that more than 60% of patients being treated for personality disorders by clinicians had diagnosable personality disorders when strict DSM-IV-TR criteria were applied, and the most common personality disorder diagnosed based on DSM-IV-TR was NOS. As mentioned previously, in DSM-5, the NOS is replaced by *other specified personality disorder* and *unspecified personality disorder* to minimize the vague and confusing diagnosis of NOS.

Considerable variation exists among individuals regarding predominant cognitive styles, and an individual's style of thinking also shifts considerably from time to time. Cognitive style refers to one's predominant manner of information processing and decision making; the particular biases and distortions thinking processes make by means of augmenting, elaborating, or minimizing incoming information; and the extent to which people use careful and deliberate logic versus intuition versus thoughtless, anxiety-induced impulsivity to guide decision making. In some individuals, a particular cognitive style may come to dominate that person's repertoire so completely as to interfere with the flexible, adaptive responses required to deal with the usual variety of daily needs. An obsessional style of thinking is marked by attention to detail and hypervigilance concerning the possible implications of a particular thought or event. This may take the form of preoccupation with strict adherence to established rules, values, or beliefs. An obsessional style may be highly adaptive in certain situations, as in professions requiring meticulous detail such as librarians, computer programmers, and surgeons. However, excessively rigid obsessionality may be maladaptive, as when someone scrupulously sticks to the rules even when such adherence is self-destructive and short sighted. A hysterical style of thinking is characterized by global, diffuse, impressionistic, emotionally laden evaluations of situations in which lack of attention is given to details and nuances. This style is poorly adaptive to detail-oriented work but may be useful in the arts, certain aspects of marketing and sales, and some social situations.

A 35-year-old pharmaceutical sales manager had difficulty in managing work responsibilities in her new job. She reported spending long hours focusing on minute details of each task herself. She struggled to delegate responsibilities to her staff or to prioritize her workload for fear of missing any small detail. She was unable to meet deadlines for job reports or assigned tasks. Frustrated, she eventually asked for a job transfer.

Types of Thinking

Because of the different ways in which both normal and abnormal thinking expresses itself, differences that are apparent to even a casual observer, attempts have been made to subtype thinking by the extent to which logical versus nonlogical thought is used. Although less commonly used than before, Sigmund Freud's division of thought into primary and secondary process thinking provides a useful descriptive classification.

Primary Process. Primary process thinking, the more primitive type, is typically seen in dreams but is also prominent in young children and in psychotic states. This type of thinking disregards logic, permits contradictions to exist simultaneously, disregards the linear notion of time, and is dominated by wish and fantasy. It uses symbol, metaphor, imagery, condensation, displacement, and concretism in its organization, creating the jumbled and incoherent style of thinking characteristic of dreams. Primary process thinking represents what has been loosely and metaphorically called right brain thinking, associated with visual images and creative thought.

Secondary Process. Secondary process thinking is characterized by logic. In contrast to primary process thinking, the secondary process uses linear notions of time, clearly delineated abstract categories, and deductive rules of logic. The abilities to think abstractly and to think in detail about future plans are characteristic of secondary process thinking. Normal secondary process thinking is also characterized by predictability, coherence, and redundancy. Words, vocal inflections, and gestures provide important contextual cues and create a sense of overall coherence to the communication. Ideas follow one another in a sequence that is understandable to the listener.

A non-Freudian typology of thought divides thinking into three types: fantasy thinking, imaginative thinking, and rational or conceptual thinking. Fantasy thinking allows the person to escape from, or deny, reality and can be seen in normal as well as pathologic thinking. Everyone occasionally uses fantasy thinking when daydreaming. Some dissociative and psychotic phenomena illustrate the most pathologic manifestations of fantasy thinking. Imaginative thinking merges fantasy and memory to generate plans for the future. Rational or conceptual thinking uses logic to solve problems.

Regardless of how one categorizes thought, people can fluidly shift from linear/secondary process/rational thought to fantasy/primary process/nonlogical thought, as in the free associative method used in psychoanalysis. During this process, individuals willfully surrender the controls that maintain secondary process thinking and switch to the less controlled modes of primary process thinking in which thoughts are loosely associated by emotional associations or based on peripheral, concrete, coincidental, loosely similar, or trivial aspects of a thought. In addition, the fact that increases in primary process thinking can be induced in normal people under experimental conditions or with fatigue suggests that more primitive thought processes, such as those seen in psychosis, are usually inhibited by higher-order processes and that their appearance may be release phenomena—that is, nonlinear or psychotic thinking may indicate the functional absence of those overriding control systems that ordinarily sift, evaluate, and regulate the form and flow of thought before it reaches consciousness.

Flow and Form Disturbances

Because the underlying processes that govern thought are not understood, current systems for classifying thought abnormalities are primarily descriptive. Conventional classification separates form and flow from the content of thought, yet many types of abnormal thinking include both form and content abnormalities. Thus, although delusions are usually classified as thought content disturbances, they are also marked by formal abnormalities, such as rigidity and imperviousness of thought to external influence or to information that clearly contradicts the delusional idea.

Although formal thought disorder typically refers to marked abnormalities in the form and flow or connectivity of thought, some clinicians use the term broadly to include any psychotic cognitive sign or symptom.

Normal variations in the flow and form of thought are considerable. For some people, thinking appears to be effortless, rapid and productive, exhibiting linear, goal-directed thoughts and creativity, with digressions and occasional leaps but always controlled and comprehensible. For others, thinking is a difficult exercise, a slow, painstaking process with low output, compared with other people, or “scattered,” with difficulty staying with a topic or finishing a single thought. Most people experience admixtures of these extremes. Disturbances in the flow and form of thought occur with regard to rate, continuity, control, and complexity.

Thinking can be unusually slow or accelerated. Slowed (or retarded) thought, such as noted in depression, is typically goal directed but characterized by little initiative or planning. Patients experiencing retarded thought often describe feeling that even simple thought requires monumental effort, as if molasses were cluttering their thinking. These difficulties are expressed as slowness in decision making and as long latency of response, increased pause times when speech is initiated and during speech. Thought blocking, seen in schizophrenia, is experienced as the snapping off or as a sudden break in a train of thought, as if a wall suddenly comes down, interrupting thinking (and speaking) in midsentence. To an outside observer, without further explanation from the patient, thought blocking may appear identical to thought withdrawal, a disturbance in the control of thought in which the patient feels as if some alien force has intentionally withdrawn the thoughts from consciousness. The patient's further description and explanation of the inner experience is necessary to distinguish these two symptoms.

> A 27-year-old man was brought to the emergency room by ambulance after being found at a gas station making bizarre hand signals at passing motorists. The man spoke in a slow, halting manner. During the interview, he would complete his sentence partially, then he would stare blankly with an odd smile. Eventually, he would complete the remainder of the sentence as if he had never stopped midway. He reluctantly disclosed that his thoughts were being programmed and controlled by powerful computer signals.

Accelerated rates of thinking, typically accompanied by fast talking, can be seen as a normal variant. Rapid rates of speech, influenced heavily by cultural and situational factors, only sometimes reflect truly rapid thought. Pressure of speech—speech that is rapid, excessive, and typically loud—is characteristic of mania (or hypomania), stimulant intoxication, and, occasionally, anxiety. Flight of ideas occurs when the flow of thought increases to the point at which the train of thought switches direction frequently and rapidly. The associative links between conceptual topics during flight of ideas are comprehensible to the listener, although not without considerable effort at times. Listening to a flight of ideas that is not overwhelmingly fast can be both a dizzying and enjoyable experience for the listener, as exemplified by the successful performance style of certain contemporary comedians.

Continuity

Disturbances in the continuity of thought may take several forms. In circumstantiality, the flow of thought includes many digressive turns and associations, often including a great deal of unnecessary detail. Transcripts of circumstantial thought or speech are marked by multiple commas, subclauses, and parenthetic asides. Nonetheless, in circumstantial thought or speech, the speaker eventually returns to the point that was initially intended without having to be prompted by the listener.

In contrast, in tangentiality, the person's thought wanders further and further away from the intended point, without ever returning, so that the person may not even remember what the original point was supposed to be. Tangentiality is a mild form of derailment in which there is a breakdown in associations. Loose associations exemplify more severe derailment, in which the flows of ideas are no longer comprehensible to the listener because the individual thoughts seem to have no logical relation to one another. Loose associations are classically a hallmark feature of schizophrenia. In extreme cases, the associations of phrases and even individual words are incomprehensible, and syntax—the rules of grammar by which phrases are organized into sentences and words into phrases—may be disrupted. Word salad describes the stringing together of words that seem to have no logical association, and verbigeration describes the disappearance of understandable speech, replaced by strings of incoherent utterances.

Clang association refers to a sequence of thoughts stimulated by the sound of a preceding word. For example, a manic patient said, "I'll kill with a drill or a pill—God, I'm ill—what swill." In echolalia, the patient repeats a sentence just uttered by the examiner. Repetition of only the last uttered word or phrase is called palilalia, a symptom found most often in chronic schizophrenia.

Perseveration and stereotypy are two other associative abnormalities in which the flow of thought or speech appears to get stuck. In perseveration, a sentence or phrase is repeated, sometimes several times over, after it is no longer relevant. Perseveration is commonly seen in delirium and other medical, substance related, and/or CNS disorders. Stereotypy refers to the constant repetition of a phrase or a behavior in many different settings, irrespective of context.

Disturbances in the control of thought include delusional passivity experiences and obsessional thinking. In delusional thought passivity, patients experience their own thoughts as being under the control of other forces. Thought passivity may take several forms: in thought insertion, thoughts are experienced as having been placed within the patient’s mind from the outside; in thought withdrawal, thoughts are whisked out of the mind; in thought broadcasting, patients experience their thoughts as escaping their minds to be heard by others.

A 20-year-old man presented for a psychiatric evaluation with his mother, noted to be wearing an orange hat with wires protruding through the cap. The patient had recently started wearing unique head gear. The patient stated that he had an elaborate plan to wear a different color cap with a feather or wires every day. Animatedly, the patient stated that unknown individuals were pulling thoughts out of his mind and broadcasting his business to everyone.

These experiences are often combined with specific delusions of control, seemingly to explain the passivity experiences. Several of these phenomena were included by Schneider among the so-called first-rank symptoms of schizophrenia. Today, these symptoms are viewed more broadly as nonspecific psychotic symptoms, more likely to be seen in schizophrenia but not pathognomonic of the disorder.

A 36-year-old man with schizophrenia had placed silver duct tape tightly around every window in his home. He was convinced that his neighbors had a device connected to their TV satellite dish that could control his thoughts and actions. He calmly explained that the duct tape interfered with the satellite signal. He never ventured out of his home without headphones on so that no one could hear his thoughts or place bad thoughts in his head to control him.

Obsessional thinking is stereotyped, repetitive, persistent thinking that is recognized as one's own thoughts. Obsessional patients do not experience their thoughts as being controlled by outside forces. Nonetheless, they experience only partial control over the obsessional thoughts. They can, with great effort, stop thinking the obsessional thoughts but cannot prevent them from recurring. Thus, characteristic of obsessions are the subjective experience of compulsion and the resistance to it. In classic obsessional thinking, insight is retained, and, as bizarre as some obsessions are, patients know that these thoughts are irrational and their own. However, more recent studies have revealed that insight into obsessional thinking is more variable than had been previously believed, at times becoming delusional. At times, obsessions may be pervasive enough to dominate the patient's consciousness. Obsessions may be simple, a sequence of words, or elaborate, such as enumerating the possible consequences of a past behavior and elaborating a cascading sequence of typically catastrophic events. Typical obsessional themes in OCD involve preoccupations with dirt and contamination, fear of harming others, symmetry, and those related to health and appearance.

A 33-year-old man would leave 2 hours plus early for every appointment. He had an obsession to park his vehicle exactly in-between two white lines, and it had reached the point that he would park his car, come out of the vehicle, and measure the distance on each side with a measuring tape to keep it equal. This would take him over an hour; therefore, he had to leave home couple of hours before each appointment, making him incapable to function and hold any job.

Obsessional thoughts are usually seen in conjunction with compulsive behaviors, which are rituals linked to the obsessions that are typically constructed to undo the effects of the thought. Thus, contamination obsessions are linked to cleaning rituals, fears of harming others lead to checking rituals, and so forth.

A 36-year-old woman was found standing in front of the grocery store counting backward rapidly in multiples of 10 with a chanting rhythmic quality for several minutes. Relieved, she then entered the store convinced that her chanting rituals would ward off mites that would otherwise infest her body. She spent several hours bathing and chanting numerically due to obsessions of contamination and infestation.

The most prominent disturbance of thinking complexity is an impaired capacity to think abstractly. Abstract thinking is the ability to assume a mental set, to keep simultaneously in mind all of the aspects of a complex situation, to move from feature to feature as indicated by the situation, and to abstract common properties. Complex thinking also concerns the ability to simultaneously consider many different, vague, and subtle aspects of situations; to appreciate differing and contradictory points of view; and to integrate these multiple dimensions to form opinions that are differentiated and nuanced. Normal individuals vary greatly in their abilities to engage in abstract thinking—geniuses in mathematics and theoretical physics leave most mortals far behind. Concrete thinking is a disturbance in the ability to form abstract concepts, generally illustrated by literal mindedness and the inability to abstract the commonality of members of a group, for example, the fact that a flea and a tree are similar in that they are both living things. Concrete thinkers seem unable to free themselves from the literal or superficial meanings of words. Concrete thinkers may be more prone to prejudice and stereotypical thinking and more likely to manifest unidimensional or all-or-none reactions to complex situations. Concrete thinking can be seen in individuals with lower intelligence, organic mental disorders, and schizophrenia. Schizophrenic patients may also exhibit highly selective disturbances of abstraction.

Thought Content

The normal content of thought, the buzzing, booming stream of consciousness that constitutes the stuff of everyday life, comprises awareness, concerns, beliefs, preoccupations, wishes, and fantasies occurring with various degrees of clarity, vividness, differentiation, imagination, and strength. Normal thought is often illogic, containing many beliefs and prejudices that may be clearly contradictory but are nevertheless held with passion and conviction.

Belief systems are the scaffolding of thought, chains of impressions, and expectations around which plans and behaviors are organized. Belief systems may be attitudinal, setting general expectations and biases about the world that inform how incoming information is processed; examples are optimism, pessimism, and paranoia. Some beliefs are evanescent, whereas others are pervasive, tenacious, enduring, and influential. The enduring belief systems are associated with behaviors consistent with the belief, at times dominating interpersonal relationships and lifestyles. Some beliefs are unique and private, whereas many are shared by others.

A 48-year-old man had been deeply pessimistic since childhood. Although he had no vegetative features of depression, he believed and behaved as if the world was an awful place dominated by dishonest and selfish individuals. Having any goals seemed foolish to him, given the nature of the world. He expressed these beliefs regularly to his children, to the dismay of his wife, who did not share his belief system.

Imaginative fantasy is an important component of normal thought. The vivid, eidetic imaginations of young children can produce vivid fantasies in which children become fully immersed, almost as if in hypnotic states. During latency, many children develop imaginary companions as playmates. In later years, imaginative thinking in which previously separate streams of thought playfully interact with one another to produce new ideas may be the essence of the creative reverie. Artists, writers, and creative scientists may retain access to these forms of thinking more readily than others. Meditative states of mind may facilitate the emergence of imaginative insights. Such thinking may also occur in dreams. Intrusive reveries are normal and common components of the usual adult stream of consciousness. During periods of specific deprivation, such as starvation or sexual deprivation, elaborate wish-fulfilling daydreams frequently occur.

Ideas are the contents of the stream of thought. Those that are consistent with one's sense of self, compatible with the individual's self-image, are called ego-syntonic. Other thoughts that conflict with one's central values are called ego-alien or ego-dystonic. An ego-dystonic impulse to kill someone, inconsistent with one's predominant value systems, may generate a counteractive ego-syntonic thought such as “you really don't mean it.”

Disturbances in Thought Contents. Abnormal beliefs and convictions form the core of thought content disturbances. Considerations of abnormality regarding beliefs and convictions must take the person's culture into account. Beliefs that may seem abnormal in one culture or subculture may be commonly accepted in another. For example, religious hallucinations, attributed to psychological or biologic factors by contemporary Western societies, are routinely attributed to religious and spiritual causes by many other cultures. With regard to intensity of conviction, distorted beliefs range on a continuum from overvalued ideas to the determined, unshakable belief that is characteristic of fixed delusions. Abnormal beliefs and delusions are, in most circumstances, diagnostically nonspecific. Delusions are commonly seen in mania, depression, schizoaffective disorder, delirium, dementia, and substance-related syndromes, as well as in schizophrenia and delusional disorders.

Overvalued ideas are unreasonable and sustained abnormal beliefs that are held beyond the bounds of reason. Patients with overvalued ideas have little or no insight into the fact that their ideas are very unlikely to be valid; however, the ideas themselves are not as patently unbelievable as most delusions. The distorted body images of body dysmorphic disorder exemplify overvalued ideas. Morbid jealousy and preoccupation with a spouse's possible infidelity may constitute an overvalued idea if no real evidence has ever existed to warrant suspicion.

A 42-year-old man who had a brief relationship with a woman was unable to accept the fact that she no longer wanted to see him. He ruminated about every encounter that they had and interpreted small gestures in the past as indicating her undying love for him. His infatuation led him to follow her repeatedly to work and school, and he pursued her relentlessly, to the point that she brought charges against him for stalking.

Ideas of reference are false personalized interpretations of actual events in which individuals believe that occurrences or remarks refer specifically to them when in fact they do not. Ideas of reference may be less firmly held than delusional beliefs.

A 20-year-old college student was brought to the student health clinic because she was pacing outside her dormitory building in an agitated manner. She stated that she could not sleep for several days because when she heard a cell phone ring, it was a signal that other students were talking about her and laughing at her.

Delusions. Delusions are fixed, false beliefs, strongly held and immutable in the face of refuting evidence, that are not consonant with the person's educational, social, and cultural background. Thus, delusional thoughts can only be understood or evaluated with at least some knowledge of patients' interpersonal worlds, such as their involvements with religious or political groups. One of the mind's functions is to generate beliefs, which may include myths and meaning systems. Shared beliefs provide an individual with a sense of personal and group identity and with ways of understanding reality. They are most noticeable when shared untestable beliefs form the basis for group cohesion, as in religions and cults. Some groups adhere to their cherished beliefs despite the abundance of plausible contrary evidence. Individuals often hold on to their primary beliefs as matters of faith (i.e., alternative, nonrefutable bases for understanding); they may hold these beliefs even in the face of grave personal threat. The strong faith with which religious, political, and nationalistic convictions are held, even at the cost of death, shows the power that these beliefs can have on behavior. Potential mental health advantages of religious beliefs have been demonstrated in epidemiologic studies showing that those with a sense of personal devotion report fewer depressive symptoms.

Subjectively, delusions are indistinguishable from everyday beliefs. Therefore, the subjective experience of a delusion is no different from the subjective experience of believing that the earth is round or that one's spouse is the same person that one married on their wedding day. Because of the identical experience of delusions and other strongly held beliefs, it is generally impossible to argue a patient out of a delusional belief. The content of delusions is highly influenced by culture. Centuries ago, delusions of persecution often concerned persecution by the devil and had religious connotations; persecutory delusions today often take on contemporary technologic, political, and social perspectives.

A 25-year-old man diagnosed with schizophrenia adamantly believes that he has discovered a new program that will transform the internet. He explains this transforming discovery is due to his ability to receive messages directly from his computer. He would not elaborate further on the program because of his belief that other people would steal the program from his brain or that his thoughts would be broadcast out loud, destroying his secret discovery.

Although delusions are diagnostically nonspecific, some types of delusions are more prevalent in one disorder than another. For example, although delusions of control and delusional percepts are often seen in schizophrenia, they also occur, albeit less frequently, in psychotic mood disorders. Similarly, classic mood-congruent delusions, with grandiose themes seen in mania or delusions of poverty characteristic of depression, may also be seen in schizophrenia.

Table 8–1 lists some characteristics by which delusions have been classified. Simple delusions contain relatively few elements, whereas complex delusions may contain extensive elaborations of people, spirits, motives, and situations.

A 32-year-old man is convinced his neighbor is videotaping his movements, monitoring his cell phone activity, and has hacked his WiFi. The patient called the police but later was convinced the police were bought off by his neighbor. Subsequently, the patient installed an elaborate camera system, taped aluminum foil over his windows, and installed a new satellite dish to protect himself against his neighbor's intrusions.

Systematized delusions are usually restricted or circumscribed to well-delineated areas and are ordinarily associated with a clear sensorium and absence of hallucinations. They are often isolated from other aspects of behavior. In contrast, nonsystematized delusions usually extend into many areas of life, and new data—new people and situations—are constantly incorporated to further support the presence of the delusion. The patient usually has concurrent mental confusion, hallucinations, and some affective lability. Whereas the patient with a closed systematized delusional system may go about life relatively unperturbed, the patient with a nonsystematized delusion frequently has poor social functioning and often behaves in response to the delusional beliefs.

Table 8–1.
Characteristics of Delusions

Simple vs. complex
Complete vs. partial
Systematized vs. nonsystematized
Primary (autochthonous) vs. secondary
Persecutory vs. nonpersecutory
How they affect behavior

Complete delusions are those held utterly without doubt. In contrast, partial delusions are those in which the patient entertains doubts about the delusional beliefs. Such doubts may be seen during the slow development of a delusion, as the delusion is gradually given up, or intermittently throughout its course.

During an acute schizophrenic episode, a 23-year-old man was completely convinced that the subtle pattern of superficial veins on his legs was evidence of “astral domination” by extraterrestrial beings. After 1 week of antipsychotic medication, he believed the veins were caused by others with special powers but was less sure. Two weeks later, he denied believing that the patterns on his legs were other than superficial veins and dismissed his previous belief with a shrug.

An autochthonous delusion is one that takes form in an instant, without identifiable preceding events, as if full awareness suddenly bursts forth in an unexpected flash of insight like a bolt from the blue. These delusions may be quite elaborate.

After a two-year period of gradual academic and interpersonal decline, a 21-year-old man had the sudden conviction that certain songs played on the radio used his voice in the role of lead singer. He could not explain why this would be so nor why this belief emerged suddenly when it did.

Aside from the autochthonous types, three other types of delusions have been described as primary. Delusional percept refers to the experience of interpreting a normal perception with a delusional meaning, one that has enormous personal significance to the patient. Delusional atmosphere or delusional mood is a state of perplexity, a sense that something uncanny or odd is going on that involves the patient but in unspecified ways. Ordinary events may take on heightened significance, but the delusional interpretations are fleeting, although the uncanny feeling stays. Typically, after a period, full-blown delusions develop, replacing the delusional mood. Delusional memory is the memory of an event that is clearly delusional. As an example, a patient “remembered” that his fourth-grade teacher slipped lysergic acid diethylamide (LSD) into his apple juice; this memory served to explain his psychotic disorder. The elaboration of false memories and their subsequent fixed belief may assume delusional proportions.

A young schizophrenic woman attended a “trauma group” with her roommate and gradually came to believe that she had been repeatedly sexually assaulted by her father from the time she was in the crib. What started out as vague dreamlike images gradually coalesced into a series of “sensed” memories that then took on specific visual images of her father’s fingers penetrating her and of his looking down at her in the crib leering. Her parents were horrified by these accusations, and there was not a shred of evidence to corroborate her increasingly venomous accusations.

Patients vary considerably in the extent to which they take action in response to delusional thoughts. Just as patients can experience delusions of their thoughts being controlled (thought passivity), they may similarly experience their feelings, behaviors, and will as controlled by outside forces. These delusions of control (or passivity experiences) occasionally, albeit uncommonly, result in dramatic self-destructive or aggressive behaviors, as illustrated by the murderer who called himself Son of Sam. This psychotic killer murdered a series of people in New York and claimed that he was the powerless agent of a force that required him to commit the acts. To defend themselves and others against delusional anticipated events, some patients may take bold and occasionally destructive actions.

A 68-year-old patient was brought to the psychiatry emergency center by police after her frightened neighbors heard gunshots coming from her home. The patient stated she was being persecuted by her neighbor's son and had bought a gun to protect herself. She was convinced the neighbor's son switched identities and had "2700 aliases" that he would use to steal from her. The patient observed the son returning home late and night and acknowledged she was shooting at him to protect her life.

Table 8–2 lists some classic types of delusions. Although less common than those involving paranoia, grandiosity, and influence, delusions of misidentification are prominently reported because of their inherently intriguing nature. In Capgras syndrome, the patient believes that someone close to them has been replaced by an exact double. In Frégoli phenomenon, strangers are identified as familiar people in the patient's life. In the delusion of doubles, patients believe that another person has been physically transformed into themselves.

A 23-year-old young woman, recently diagnosed with schizophrenia, was brought to the emergency center by ambulance after she called 911 screaming, "Give me back my mother, I know what you've done." In the hospital, she accused her mother of being a double, in spite of the mother's attempts to show her old family photographs and her birth certificate.

Olfactory delusions that one emanates a foul odor are common in social anxiety syndromes, in which individuals are particularly concerned about potentially embarrassing themselves and others.

Delusions are not only seen in isolated individuals. Shared delusions may occur in couples (folie à deux) and in families (folie à famille). Many psychiatrists consider group delusions to be present in some cults, as well, but exactly where the cutoff points occur between delusions and other zealous beliefs held by larger, more traditional and well-organized religious, political, and other groups is arguable.

Table 8–2.
Some Classic Types of Delusions

Delusions of persecution
Delusions of grandeur
Delusions of influence
Delusion of having sinned
Nihilistic delusions
Somatic delusions
Delusion of doubles (doppelganger)
Delusional jealousy (Othello syndrome)
Delusional mood
Delusional perception
Delusional memory
Delusions of erotic attachment (Clérambault syndrome)
Delusions of replacement of significant others (Capgras syndrome)
Delusions of disguise (Frégoli phenomenon)
Shared delusions (folie à deux, folie à trios, folie à famille)

> A husband and wife were brought by police for a psychiatric evaluation. They both wore a large colorful talisman made of metal on their head which had a particular significance to them. The reason for the talisman, they concurred, was to ward off cell phone signals which could control their actions and project their thoughts to others. The husband was a dominant personality who had originated this fixed delusion. His wife, a mild-mannered, passive person, seemed to have accepted her husband’s interpretation of events without challenging them at all.

DISTURBANCES OF JUDGMENT

Judgment involves a complex and diverse group of mental functions that includes analytical thinking, social and ethical action tendencies, and depth of understanding or insight. Analytical thinking includes the capacity to discriminate and to weigh the pros and cons of potential alternative actions. Social and ethical action tendencies are closely related to culture and upbringing. The evidence for genetic factors in antisocial personality disorder (defined primarily by judgments that lead to criminal behaviors) points to the additional role of constitutional factors. Insight may reflect intelligence, learning, cognitive style, and the capacity to integrate intellectual knowledge with emotional awareness.

Impairments of judgment occur in many psychiatric disturbances. Anxiety states, intoxications, fatigue, and even group pressures may cause temporary impairments of judgment in otherwise normal individuals. In mood disorders, judgment may be impaired by either an exaggerated evaluation of risk or failure in depression or, conversely, of inadequate appreciation of risk or danger in mania. Brain damage caused from medical conditions, including persistent substance-related conditions and psychotic disorders, may chronically impair any aspect of judgment in any person, regardless of premorbid character. Poor role models and deviant social backgrounds may lead to social and ethical action tendencies quite different from those of the examiner. Thus, someone raised in a criminal environment may have superb analytical judgment and self-awareness, which are, however, put to illegal use.

A 48-year-old attorney missed multiple visits with his established psychiatrist and subsequently ran out of his mood stabilizer, in part due to increased travel related to a prominent out of town legal case. His wife of 20 years became increasingly concerned as he erratically began flying first class, spending $40,000 on airfare, and not disclosing his whereabouts. After not communicating with his wife for 2 days, he was found by security at an out-of-town television station claiming he had been hired as their national legal correspondent to provide legal commentary.

Judgment may be impaired in one dimension and spared in others. Individuals may retain sound ethical judgment when their analytical capacities fail or may retain excellent analytical abilities for nonpersonal matters, although lacking insight into personal situations or behaviors.

A 42-year-old chief financial officer of an oil and gas company was admitted for inpatient psychiatric hospitalization. His wife stated that he had been working on a major financial transaction for his company late into the night and not slept for several days. She then found out that he had also been ordering expensive watches and jewelry from the internet, spent $35,000 on these impulsive purchases, and appeared excessively euphoric. During the psychiatric interview and mental status exam, the patient proudly counted backwards very fast during serial 7’s and said, “I can give you three things that are similar between an orange and an apple.”

Thus, some people who can provide socially appropriate responses to traditional mental status examination questions, such as what one would do in a movie theater if fire broke out, might at the same time be incapable of accurately assessing crucial clinical or more personal matters specifically related to one’s capacity to provide informed consent, such as the pros and cons of receiving treatment; regarding judgments necessary to provide oneself with food, clothing, and shelter; or insight into one's state of health or illness.

The term insight, usually in the context of self-awareness, has been used in a variety of ways. Basic insight refers to a superficial awareness of one's situation. In evaluating insight into one's psychiatric condition, basic insight allows an individual to acknowledge the presence of an illness. A deeper level of insight is operating when the patient has an intellectual appreciation of what is going on (e.g., "I have hallucinations and delusions, and my doctors have told me that I have schizophrenia and must take medication"). Still deeper levels of insight reflect more complete cognitive and emotional appreciation of a situation (e.g., "I realize that I have schizophrenia, that it impairs my judgment and social function at times, and that I will have to take medications if I am to minimize my symptoms and try to make the most of my life. I feel profoundly disappointed about this affliction because it prevents me from achieving some of the goals I've always wished for. Nevertheless, I have to do my best to get over my disappointment and hurt feelings so that I can get whatever I can out of life.")

Of course, different depths of insight as self-awareness can be evaluated in many other situations, such as physical illness, quality and nature of relationships, an appreciation of strengths and weaknesses in professional situations, and so forth. In formal studies of insight using standardized instruments, lack of insight correlates with poor outcome in schizophrenia and bipolar disorder, medication noncompliance, and suicidality. Improvement of psychosis does not necessarily correlate with improved insight. Impaired insight may be associated with frontal lobe abnormalities. Insight may be as seriously impaired in mania as in schizophrenia and, contrary to earlier beliefs, may be lacking in OCD.

Judgment may be impaired by several factors, including cognitive clouding (as in disturbances of consciousness, for example, intoxication, so that one's usual analytical abilities are impaired), self-deception, and impulsivity.

Self-deception refers to the almost universal tendency to hide certain issues about the external world or about oneself from various levels of awareness. Self-deception functions as a coping strategy, fostering or maintaining comfortable perspectives about the world and avoiding confrontation with issues and realities that inevitably stir up painful conflicts or the need for difficult actions, thereby preserving emotional calm. In addition, studies suggest that self-deception enables people to act and to be perceived as more convincing in the service of particular goals, as in romantic relationships or business dealings. Therefore, although "kidding ourselves" may sometimes reflect impaired judgment, it may at times also yield certain important strategic advantages.

Impulsive judgment describes a tendency to avoid taking the time and thought to fully understand and integrate all of the facts and levels of awareness required for optimal decision making. Impulsive judgment may occur only with certain issues or situations (such as how one picks investments), signal an impaired state (such as intoxication), or reflect a pervasive character trait.

Rapidly made judgments, and even the so-called snap judgments, may not be maladaptively impulsive, even when they involve very important areas of life. Rapid decisions can be very accurate, highly adaptive, and even lifesaving, especially if made against a background of great experience, wisdom, and forethought concerning the area requiring the decision.

Because, in clinical practice, the terms insight and judgment are often applied to individuals' awareness and decision making about their psychiatric status, complex motivational states that incorporate insight and judgment related to how one is dealing with one's problems, the so-called stages of readiness for change, bear mention at this point. Initially described in relation to substance abuse, including alcoholism and smoking, these stages have received considerable attention and form an important aspect of clinical assessment across other diagnostic categories such as eating disorders. Several stages have been described: (1) precontemplation, the person expresses no intention to change (may be in denial); (2) contemplation, the person acknowledges a problem and states an intention to change within several months but not right away; (3) preparation, the person intends to do something about changing in the near future and may have already made some false starts; (4) action, the individual has engaged in making sustained behavior changes; (5) maintenance, the individual has been engaged in changed behavior for more than 6 months; and (6) termination, the individual has succeeded in the change and is unlikely to return to the original behavior.

DISTURBANCES OF PERCEPTION

Normal perception first requires that the individual be capable of receiving information as sensations. The data must then be organized to make them meaningful and comprehensible such as distinguishing figure from ground or focusing attention selectively on some part of the sensory field. The organized entities are called percepts. In states of sensory deficit, such as blindness, deafness, and anesthesia, perception is impaired, but perception is still possible because individuals generally perceive information about an object through several sensory modalities concurrently. The intensity of sensation and perception is affected by vigilance and attention. Highly focused attention, as in intense concentration or hypnosis, may result in unusually acute sensation and perception—hyperesthesia, hyperacusis, or extraordinary visual acuity. Focused attention may also result in the inability to sense or perceive: deep anesthesia and negative hallucinations induced by hypnosis are simply induced failures to perceive what exists in the world.

Humans usually operate in an “average expectable environment” in which certain types and levels of sensory input are expected and for which the nervous system is primed. Excessive or inadequate stimulation in any sensory modality, levels of input that are extraordinarily intense, or the presentation of novel stimuli that are entirely different from anything previously experienced by the individual can provoke distorted perceptions in most normal people. For example, total sensory deprivation produced in carefully controlled artificial environments may elicit visual and auditory illusions and hallucinations.

Individuals generally exhibit selective perception of the world, depending on what is salient at the moment and on individual memories, emotions, fantasies, and values. Pregnant women are more likely to perceive babies around them than are people who are not as preoccupied with childbearing.

The intensity of perceptions depends on individual sensitivities as well as on mood, anxiety, and substance use. Unmedicated patients with schizophrenia have deficits in olfactory acuity. Depressed patients often describe that colors look faded and that the world looks washed out or gray, even though their capacity to recognize specific colors is unchanged. Similarly, mania is often characterized by heightened perceptions, hyperesthesia. When extreme, these intense perceptions are uncomfortable. Hyperesthesia can also be seen during benzodiazepine withdrawal, hallucinogen use, and, occasionally, as part of an epileptic aura.

The intensity of perception may vary with cognitive style and other psychological and neurologic factors. Some individuals tend to be augmenters and others minimizers of bodily experiences. Chronic pain and some hypochondriacal syndromes may occur more commonly among somatic augmenters.

Selective deficits in the perception of emotions may occur. Emotional aprosodies have been described in which patients with specific neurologic deficits or depression are selectively unable to recognize the expression of facial emotion. These have been linked by positron emission tomography (PET) scan to blunted activity in the right prefrontal cortex and insula.

Illusions

Perceptual distortions in estimating size, shape, and spatial relations are common even in the absence of psychiatric disorders, especially when one is fatigued or excessively aroused. Illusions are misinterpretations of real sensory stimuli, as when a child in a dark bedroom at night sees monsters emanating from shadows on the walls. Pareidolia are playful and whimsical voluntary illusions that can be seen when one looks at ambiguously defined or evanescent images such as clouds or flames in a fireplace. The onset and termination of these perceptions is under voluntary control. Trailing, another visual illusion, is the perception that an object moving steadily in space is followed by temporally distinct after-images of itself. The effect is that of a series of stroboscopic photos. This phenomenon may occur with fatigue and is typically seen with marijuana and mescaline intoxication, during withdrawal from SSRIs, or, less commonly, in association with nefazodone.

A 20-year-old male presented to the emergency room after he was brought by police because he was behaving bizarrely in the streets. He denied previous psychiatry history. He reported attending a party with some friends last night where someone gave him some drug. On examination, he reported distortions of objects and persons, stating that “police officers look like giants,” but there were no other significant findings in the mental status examination. These types of illusions are classic examples of visual or optical illusions, which are the most common type of illusions.

Hallucinations

Hallucinations are perceptions that occur in the absence of corresponding sensory stimuli. Phenomenologically, hallucinations are ordinarily subjectively indistinguishable from normal perceptions. Hallucinations are often experienced as being private so that others are not able to see or hear the same perceptions. The patient’s explanation for this is typically delusional. Hallucinations can affect any sensory system and sometimes occur in several concurrently. When perception is altered, combinations of illusions and hallucinations, and often delusions as well, are frequently experienced together. In some studies, 90% of patients with hallucinations also have delusions, and approximately 35% of patients with delusions also have hallucinations. Children and early adolescents, however, are more likely to have hallucinations in the absence of delusions. Approximately 20% of patients have mixed sensory hallucinations (mostly auditory and visual) that may accompany functional, as well as organic, conditions. A given external stimulus may evoke very different perceptual distortions in different people.

Three scientists floated in sensory deprivation tanks for long periods. One experienced a few illusions and no hallucinations; the second had many illusions and a few faint auditory and visual hallucinations; and the third had vivid, dramatic, and complex visual and auditory hallucinations.

Hallucinations are experienced by many normal people under unusual conditions. It has been estimated that between 10 and 27% of the general population has experienced memorable hallucinations, most commonly visual hallucinations. The large majority of self-reported hallucinations in community studies, particularly auditory hallucinations, have been associated with depressive and substance use disorders rather than frank psychotic disorders.

A 76-year-old woman reported "hearing and seeing" her deceased husband calling her name, and she wanted to join him in heaven. She had become very depressed after her husband of 51 years passed away 6 months ago. Friends reported that she was more withdrawn, isolated, and stopped going to her place of worship.

Hypnagogic and hypnopompic hallucinations are common, predominantly visual hallucinations that occur during the moments immediately preceding falling asleep and during the transition from sleep to wakefulness, respectively. Hypnagogic and hypnopompic hallucinations both occur in normal people and are also characteristic symptoms of narcolepsy. In acute bereavement, up to 50% of grieving spouses have reported hallucinating the voice or presence of the deceased, and after amputations, phantom limb hallucinations are common. Patients who become visually impaired often develop complex hallucinations (i.e., visual hallucinations with preserved insight) with preserved cognitive status, the so-called Charles Bonnet syndrome.

A 77-year-old woman presented to her psychiatrist with mild anxiety symptoms. During the past year she had experienced severe loss in her visual acuity due to age-related macular degeneration. She reported experiencing unexpected vivid hallucinations and described seeing complex shapes of geometric design and animal figures. She had awareness that these "visions" were not real, and her cognitive screen, including the Mini-Mental Status Examination, was normal.

A parallel phenomenon is the emergence of hallucination, including musical hallucinations in individuals with acquired deafness. These observations suggest a "supersensitivity deprivation" hypothesis that, when deprived of important and anticipated perceptual stimuli, the mental apparatus may overinterpret any sensory stimulation as evidence of the presence of the needed objects.

A perceptual release theory suggests that hallucinations emerge from the combined presence of intense states of internal arousal and diminished sensory input (including poor attention and poor capacity to sort out relevant from irrelevant input). Thus, diminished input from the environment (as in sensory deprivation) or reduced capacity to attend to and take in the input (as in delirious states) heightens the likelihood that internal sensations, images, and thoughts are interpreted as originating in the outside environment.

Hallucinations vary according to sensory modality, degree of complexity of the hallucinated experience, the levels of conviction about their reality, the clarity of their contents, the location of their sources of origin, the degree of volitional control over them, and the degree to which the hallucination influences the person's behavior.

Auditory hallucinations range in complexity from hearing unstructured sounds, such as whirring noises or muffled whispers, to ongoing multiperson discussions about the patient. The simple auditory hallucinations are more commonly associated with organic psychoses such as delirium, complex partial seizures, and toxic and metabolic encephalopathies. Auditory hallucinations are classically associated with schizophrenia (seen in 60% to 90% of patients) but are also frequently seen in psychotic mood disorders. Twenty percent of manic patients and less than 10% of depressed patients experience auditory hallucinations.

Three types of auditory hallucinations commonly associated with schizophrenia (which, however, are also seen less commonly in patients with psychotic depressions and mania) are (1) audible thoughts described as hallucinated voices that speak aloud what the patient is thinking, (2) voices that give a running commentary on the patient's actions, and (3) hearing two or more voices arguing with each other, often about the patient, who is referred to in the third person.

A 37-year-old man with chronic schizophrenia reported hearing multiple voices, both male and female, commenting on his daily activities. He described the male voice as "critical" and commenting in a negative way about his grooming, clothing, and appearance. The female voice was calming and soothing. At times, he stated he could "block the voices out by putting on loud music." Other times he stated that "they would take over," leading him to feeling despondent and hopeless. He stated during these "terrible times" he considered jumping off a bridge to end the "voices torturing me."

While auditory hallucinations in schizophrenia are frequently mood neutral, hallucinations in patients with mood disorders are characteristically consistent with their mood. In psychotic depression, the voices may be unrelievedly critical and sadistic, whereas in mania the voices often refer to the patient's specialness.

A 36-year-old woman with bipolar disorder described auditory hallucinations correlating with episodes of severe mania or depression. When she was depressed, she reported hearing voices that were demeaning and called her “fat, ugly, and worthless.” The voices were so unremitting that she became suicidal. However, during an intense euphoric manic episode, she described hearing joyful angels singing and declaring her as the next savior sent to earth by God.

Command hallucinations order patients to do things. Often, the commands are benign reminders about everyday tasks: “Pick up your shoes” or “Clean off the table.” However, the voices may also be frightening or dangerous, commanding acts of violence toward the self or others such as, “Jump off the roof; you’re not worth anything,” or, “Pick up the knife and kill your mother.” These voices vary in insistence and persistence, and patients differ in their capacities to ignore these commands. Patients with marked passivity may be helpless in the face of command hallucinations and may feel impelled to carry out the orders. Even though one study did not find command hallucinations to be associated with a higher risk of harm to the patient or others, the presence of command hallucinations and the patient’s ability to resist must be assessed carefully.

A 46-year-old woman with schizophrenia heard command hallucinations daily for the last 20 years. When her symptoms were relatively quiescent, the voices were benign, and the commands referred to daily behaviors (such as “say thank you”). Under stress, her psychosis exacerbated, and the voices became louder and more insistent and commanded more complex and dangerous behaviors, once precipitating a self-stabbing when the voices demanded it.

Visual hallucinations occur in a wide variety of neurologic and psychiatric disorders, including toxic disturbances, drug withdrawal syndromes, focal CNS lesions, migraine headaches, blindness, schizophrenia, and psychotic mood disorders. Although visual hallucinations are generally assumed to characteristically reflect organic disorders, they are seen in one-fourth to one-half of schizophrenic patients, often—but not always—in conjunction with auditory hallucinations.

Visual hallucinations range from simple and elemental, in which hallucinations consist of flashes of light or geometrical figures, to elaborate visions such as a flock of angels.

Stimulation of one sensory modality sometimes evokes perceptual distortions in another. Marijuana and mescaline intoxication, for example, have been associated with synesthesia, an experience in which sensory modalities seem fused. This is also a normal experience for many people. Music may be experienced visually, the sound fusing with visual illusions; a tactile sensation may be experienced as a color (e.g., a hot surface may “feel red”).

In certain religious subcultures, visual hallucinations may be experienced as normal. For example, in some fundamentalist Pentecostal churches, worshipers may experience religious visions that they perceive as being from God.

A 62-year-old Native American man was raised on an Indian reservation and was deeply connected to his heritage. He openly acknowledged that the spirit of his elders and ancestors surrounded him. He reported that he could feel and sense their presence around him no matter what he was doing, whether he was working or resting. The presence of these spirits was a guiding force that he felt was calming to him.

Autoscopic hallucinations are hallucinations of one’s own physical self. Such hallucinations may stimulate the delusion that one has a double (doppelganger). Reports of near-death, out-of-body experiences in which individuals see themselves rising to the ceiling and looking down at themselves in a hospital bed may be autoscopic hallucinations. In Lilliputian hallucinations, the individual sees figures in very reduced size, such as midgets or dwarfs. They may be related to the perceptual distortions of macropsia and micropsia, respectively, the perceptions of objects as much bigger or smaller than they actually are.

Haptic hallucinations involve touch. Simple haptic hallucinations, such as the feeling that bugs are crawling over one's skin (formication), are common in alcohol withdrawal syndromes and in cocaine intoxication. When unkempt and physically neglectful patients complain of these sensations, they may be due to the presence of real physical stimuli, such as lice. Some tactile hallucinations—having intercourse with God, for example—are highly suggestive of schizophrenia but may also occur in tertiary syphilis and other conditions and may, in fact, be stimulated by local genital irritation. Olfactory and gustatory hallucinations, involving smell and taste, respectively, have most often been associated with organic brain disease, particularly with the uncinate fits of complex partial seizures. Olfactory hallucinations may also be seen in psychotic depression, typically as odors of decay, rotting, or death.

A 32-year-old woman with schizophrenia described that, over the 15 years of her illness, she had experienced a wide array of hallucinations. She initially noticed celestial voices, auditory hallucinations that included angelic music and comforting words. However, over the years, in conjunction with a pessimistic turn of mind, the voices became more demonic and threatening. At the same time, she began to experience somatic hallucinations consisting of sharp, intermittent electric currents being run through her skin and genitals, which, she was convinced, were the work of evil neighbors who were intent on driving her out of her apartment. With medication, she was partly able to perceive these experiences as hallucinatory; however, most times, particularly when she neglected to take medications, she never doubted their authenticity.

The term pseudohallucination has been used in two ways. First, pseudohallucination refers to perceptions experienced as coming from within the mind (i.e., not at the boundary or outside the mind). Using this definition, loud voices that are alien, ascribed to other beings, but which the patient knows are actually within the mind rather than out in space, are pseudohallucinations. The term has also been used to describe hallucinatory experiences whose validity the patient doubts. A better term for this second phenomenon is partial hallucination, analogous to partial delusion. Functional hallucinations are rare hallucinations that occur only in connection with a specific external perception, for example, in the presence of a sound, such as running water, a color, or a particular place. However, unlike illusions, the hallucinated sounds are not elaborations of the perception but are simply triggered only in that specific context.

Ictal hallucinations, occurring as part of seizure activity, are typically brief, lasting only seconds to minutes, and stereotyped. They may be simple images, such as flashes of light, or elaborate ones, such as visual recollections of past experiences. While the hallucinations are being experienced, the patient ordinarily experiences altered consciousness or a twilight sleep.

Migrainous hallucinations are reported by approximately 50% of patients with migraine. Most are simple visual hallucinations of geometrical patterns, but fully formed visual hallucinations, sometimes with micropsia and macropsia, may also occur. This complex has been called the Alice in Wonderland syndrome after Lewis Carroll’s descriptions of the world in *Through the Looking Glass,* which mirrored some of his own migrainous experiences. In turn, these phenomena closely resemble visual hallucinations induced by psychedelic drugs such as mescaline.

A flashback is an intense visual re-experience of highly charged past events, which are often replays of hallucinations. They are typically associated with heavy use of hallucinogens, such as LSD and mescaline, and often occur months after the last drug ingestion. The images may be simple or complex geometrical patterns, or they may consist of previously experienced elaborate drug-induced hallucinations. Flashback phenomena may be state dependent. For example, visual hallucinations initially experienced with hallucinogens are more likely to be subsequently experienced as flashbacks when the subject is smoking marijuana. In PTSD, some complex, intrusive flashback-like images may attain a hallucinatory vividness. Images often include horrifying memories of traumatic events that may force themselves repeatedly into consciousness until they are acknowledged and worked through.

A 33-year-old military combat veteran had left the military with commendation after serving multiple active duty combat tours. He returned home after picking up party supplies for a pool party he was hosting with his new wife at their home. Upon returning from the store and entering the home, he began to loudly scream, “Get down, get down, active shooter, active shooter here.” He was profusely sweating, appeared dazed, and was vividly calling out military commands. Later, his wife found out that her husband smelled the intense smell of ammonia and burned hair when he entered the home, which triggered the flashback memory and vivid re-living of “burning hair and burning bodies.” The wife said she had tried a new home hair permanent treatment which had gone awry.

Hallucinosis is a state of active hallucination occurring in someone who is alert and well oriented. This condition is seen most often in alcoholic withdrawal, but it may also occur during acute intoxications and other drug-mediated states.

A 21-year-old man used LSD daily for 3 days, after which the visual hallucinations that first occurred while under the influence of the drug waxed and waned. He described pareidolias with distracting images of animals when he looked at traffic lights, the sky, or billboards. Trailers were common after he turned his head. Finally, he saw moving colors when he stared at any object for more than a transient time. He assumed that these phenomena represented a lingering effect of the LSD and was not frightened by them. The hallucinations disappeared within 1 week.

Body Image Distortion

Body image includes both perceptual and ideational components and may reflect primarily perceptual distortions or combinations of disturbed perception and self-appraisal, or both.

Body image disturbances can occur as normal responses to abrupt changes in the body (e.g., after amputation), in brain disease, and in psychiatric disorders. Phantom limb phenomena are classic body image problems in which an amputated limb is still felt to be present. The sensation may diminish gradually over time; the phantom feels as if it is receding into the stump.

A 34-year-old man presented to his VA hospital mental health clinic appointment for persistent insomnia and anxiety. While on his second tour of active combat duty, he sustained a traumatic injury requiring amputation of his left leg. He reluctantly disclosed that he could feel his amputated leg "going in and out" and thought he was "going crazy." He described many sensations, sometimes that the area felt warm and tingling, other times, he felt throbbing, itching, and sharp electrical piercing feelings.

Agnosias, lack of awareness of some parts of the body, may accompany brain damage, most often of the nondominant parietal lobe. Patients with obvious motor or sensory deficits may deny that any deficit exists at all (anosognosia), or the denial may be limited to one-half of the body (hemiagnosia), usually the left side. In hemidepersonalization syndromes, a less common disorder (hemiasomatognosia), patients feel that one of their limbs is missing, again usually on the left side. Body image distortions in which a limb feels too heavy (hyperschemazia) or weightless (hyposchemazia) can occur as a consequence of neurologic conditions such as infarction of the parietal lobe. In duplication phenomena, patients feel as if part or all of them has doubled (e.g., they have two heads or two bodies). These rare phenomena may occur in schizophrenia, complex partial seizures, and migraine.

Dysmorphophobia refers to conditions in which patients distortedly perceive and intensely dislike the shape of a particular body part. As such, these symptoms are misnamed because there is no true phobic component such as fear or avoidant behavior. Fine lines exist between perceptual distortions and realistic but unhappy appraisals of one’s body, given the high social values placed on physical appearance. Dysmorphophobia may occur in the context of some personality disorders or as an isolated disorder called body dysmorphic disorder. In some ways, dysmorphophobia resembles an overvalued idea. Patients may develop dysmorphophobias in relation to any body part; common concerns are hair, breasts, penis size, the shape of the nose, or the shape of the entire body. For some, changing the body part, as in rhinoplasty for those who do not like their nose, seems to effect a lasting positive change in body image, with patients becoming happier with themselves and feeling more attractive for years or lifetimes. In severe dysmorphophobia, patients may undergo multiple plastic surgeries and feel dissatisfied with every result. At times, the condition forms part of a larger and more pervasive syndrome, such as anorexia nervosa.

An attractive 19-year-old woman was preoccupied and perplexed by the fact that she perceived her eyelids as being unequal in size and shape and by the fact that the cleft in her chin was too deep. She spent hours each day in front of the mirror, trying to pose in ways that, from her perspective, did not expose these facial blemishes. She sought consultation from a plastic surgeon, who referred her for psychiatric consultation. After 2 months of treatment with an SSRI, her preoccupations with appearance diminished considerably, and she was able to turn her attention to other issues in her life.

Hypochondriacal complaints also combine perceptual and ideational distortions. Selective hypervigilance to bodily sensations may result in a higher likelihood of perceptions of unpleasant and potentially pathologic body experiences among the “worried well,” hypochondriacal populations, patients with somatization disorder (Briquet syndrome), and some patients with panic disorder.

Body image distortions may, at times, be severe or bizarre. Some psychotic patients, either with schizophrenia or depression, develop somatic delusions. In depression, this often expresses itself as a delusion that part of or the entire body is rotting or filled with cancer. Some culture-bound syndromes in non-Western cultures express themselves with body image distortions such as koro, in which the man fears that his penis is shrinking into his abdomen.

DISTURBANCES OF CONSCIOUSNESS

Consciousness can be defined as subjective awareness of the self and environment. Biologists increasingly believe that a continuum of consciousness exists, extending from lower animals through *Homo sapiens.* However, consciousness is subject to conflicting definitions and conceptualizations, and exactly where consciousness begins remains unclear. Philosophers agree that it is the subjectivity of experience, the so-called qualia of consciousness, that clearly distinguishes living consciousness from at least this generation's best versions of self-regulating automata, elegant computers, or robots. All current attempts to even approach an understanding of consciousness are very unsatisfying, and consciousness remains unexplained and as yet unexplainable from a scientific point of view. Consciousness has been viewed as an emergent property of complex biologic nervous systems, as a poorly understood general property of an even more mysterious and complex universe, or as a phenomenon to be understood only in religious and spiritual terms. One of the best analyses to date of possible relationships between the construction of a brain and the possibility of consciousness has been set forth by Gerald Edelman, who believes that reflective consciousness cannot occur until complex higher-order brain systems evolve whose major functions are to monitor the experiences, activities, and results of activities of those lower-order brain systems that deal directly with appraising and responding to the external and internal environments. Such higher-order metasystems require the presence of memory so that current and immediate impressions can be checked and compared against past experiences. These metasystems may use a variety of sensor mechanisms to detect and signal their sensations or perceptions of various events. Some of these sensors may correspond to feeling states, and some may correspond initially to preverbal thoughtlike mechanisms that contain the capacity to develop and recognize abstract categories and, ultimately, conceptual language-based thought.

Clinically, consciousness can be considered from both qualitative as well as quantitative viewpoints. Qualitatively, consciousness does not seem to be an all-or-none phenomenon. Rather, conscious experiences may gradually and physically shift in focus, intensity, and clarity; altered states of consciousness may occur in which some aspects of consciousness, such as sensation, perception, memory, orientation, and judgment, are enhanced or impaired relative to other aspects. Quantitatively, crude divisions can be made between states depending on the relative presence, impairment, or total absence of consciousness. Even within the single individual, consciousness is not a unitary phenomenon. Multiple streams of thought, operating at multiple levels of preconsciousnesses, appear to exist in all people almost all of the time, with various elements in these coexisting streams constantly shifting into higher or lower levels of conscious awareness. In pathologic states, even more remarkable properties of consciousness are seen, for example, the existence of coconsciousness in humans who have had commissurotomies and of seemingly multiple discrete consciousnesses in patients with dissociative identity disorders.

Experiments involving patients with commissurotomies of the corpus callosum have shown the existence of two virtually separate systems of consciousness that seem to operate side by side. For example, when in the course of an experiment, the picture of a nude woman was flashed only to the right brain (the left visual field) of a commissurotomized patient, the subject verbally denied being aware of anything unusual (i.e., the left brain—the verbal brain—was unaware). But, at the same time, he started to squirm and blush, blurting out, “Oh, you have some machine!” Similarly, when a cup was presented to the right brain (left visual field) only, the patient denied seeing anything (left brain was unaware, and language output of the left brain indicated no awareness), but he was able to pick out the cup from an assortment of objects with his left hand (right brain control). That literal splitting of verbal awareness from visual–spatial awareness in the brain produces behavior that is at least superficially similar to that of patients who deny being consciously upset by an event but who react with strong visceral responses. Although this formulation is simplistic, the separate consciousness for logical–verbal and for visual–spatial awareness demonstrated in split brain experiments may be crude analogs for more highly differentiated and discrete types of awareness and modes of information processing. Furthermore, the very fact that there are separate and, to some extent, competing modes of consciousness may increase the likelihood of psychological distress, because the various modes are capable of yielding internally conflicting views of reality.

Psychological and Physiologic Factors

In ordinary states of alert consciousness, individuals are able to deploy adequate amounts of attention to their surroundings and to reflective thought. Normal people vary enormously in their ability to pay careful attention in different settings without being distracted; individual variations may reflect temperamental and cognitive style differences as well as physiologic shifts within the individual. Many functions of attentive consciousness, including attention, planning, and the capacity to appropriately switch between mental tasks, the so-called distractor-resistant memory, have been linked to the activity of specific neurons in area 46 of the prefrontal cortex.

A sense of increased consciousness with heightened alertness, awareness, and sharper thinking may be experienced in states of highly aroused emotional states such as threat, sexual attraction, falling in love, or other high-stakes events such as hunting among primitives, sporting competitions, or performing in front of an important audience. High levels of arousal do not necessarily guarantee effective attention because optimal consciousness depends on optimal arousal. Too little arousal due to illness or fatigue may result in insufficient stimulation and mental lethargy, diminishing the sense of alertness and attentiveness, whereas too much arousal may result in hyperintense alertness but distractibility and scattered attention.

Consciousness involves, among other things, the experience of a continuous sense of self and of the environment, existing coherently in time and space. The experience of time and its passage may be altered by shifts in the level of awareness and by emotional states such as boredom, concentration, pain, and discomfort. The experience of time and space may be altered by hypnosis, marijuana, psychoactive and psychedelic drugs, and other events directly affecting brain physiology.

Disturbances of Orientation

Orientation refers to one's awareness of time, place, and person. Accurate orientation requires the integrity of attention, perception, memory, and ideation. Impairments occur primarily in organic mental disorders (i.e., structural and toxic metabolic brain abnormalities) and occasionally in dissociative and psychotic states.

A 75-year-old woman arrived at her primary care physician's clinic. She was recently diagnosed with Alzheimer dementia. The staff recognized her, and she was initially polite and calm with staff. However, the patient soon appeared confused and started getting irritable. She began asking the staff where she was, why she was at the clinic, and was adamant that she didn't have any appointment scheduled for that day.

Normal individuals vary tremendously in their attention to the details of time and in the extent to which their bodies automatically keep time. Some people have reliable built-in clocks by which they can awaken themselves at precise times or accurately gauge the passage of time with uncanny accuracy, even in the absence of external cues—in a psychotherapy session, for example. Others have difficulty making judgments about time and may develop pathologic lateness or habitually schedule more activities than could ever be accomplished in the available time. Poor time judgments may be seen in a variety of psychiatric disorders, such as ADHD, or as an independent problem. Benign disorientation to time is common. After a few days in a hospital bed, most people do not know exactly what the day or date is because they are not attending to or receiving their usual cues.

Pathologic time disorientation can be mild or severe, with inaccuracies of estimation ranging from days to years. The dates reported by disoriented individuals may have personal significance such as those of important births, marriages, or deaths.

Because spatial cues are generally more available and obvious than temporal cues, disorientation to place often signifies a greater degree of cognitive impairment than disorientation to time and, therefore, rarely occurs in the absence of time disorientation. Disoriented people may know, more or less, the type of place that they are in without knowing the specific place—patients may recognize that they are in a hospital without being able to name the hospital.

Disorientation to person, a lack of awareness of one's own identity, is typically seen only in advanced dementias such as primary degenerative dementia of the Alzheimer type or in dissociative states. In organically induced postconcussion amnesia, transient global amnesia, and, presumably, psychogenic fugue states, knowledge of one's own identity may disappear, and a person may remain unidentified for an indefinite period until the memory for self-returns.

Disturbances in the Level of Consciousness

Levels of consciousness (i.e., alertness, awareness, and attentiveness) may be pathologically increased or decreased. Such changes are diagnostically nonspecific and can occur in many different disorders. When levels of arousal and alertness are mildly elevated, as in hypomania or with the ingestion of small amounts of psychostimulants, subjective experiences are typically positive. In these situations, the person experiences intense alertness, prolonged concentrating ability, and hyperesthesias in which perceptual vividness is heightened: colors are brighter, sounds are sharper, and touch is more intense than usual. With further increases in arousal and consciousness as seen in mania, more severe intoxications with amphetamines and cocaine, and catatonic excitement, attention fragments. Heightened alertness transforms into hypervigilance and paranoia, and hyperesthesias become unpleasant.

Diminished levels of consciousness can be described on a continuum. Clouding of consciousness is marked by diminished awareness of sensory cues and diminished attentiveness to the environment and to the self. Secondary process thinking is most notably compromised, and more primary process thinking emerges into consciousness. In this state, one's ability to appreciate subtleties and to think in a nuanced manner is diminished and is replaced by more dichotomous all-or-none, stereotypical thinking. The level of consciousness may fluctuate rapidly in relation to the internal physiologic state or to the degree of external stimulation. In alterations of consciousness, confusion may occur with disorientation to time, place, or person. The patient is usually highly distractible and unable to pay sustained attention to a single stimulus.

Torpor is a condition in which the patient is drowsy, falls asleep easily, and shows a narrowed range of perception and slowed thinking. Stupor is a state of diminished consciousness in which the patient remains mute and still, although the eyes are open and may follow external objects. In the most extreme impairment of consciousness, coma, there is no evidence of mental activity at all. The patient essentially appears to be functioning on a decorticate or decerebrate level. In akinetic mutism, or coma vigil, patients with profound brainstem lesions appear to be awake with their eyes open, but there is in fact no evidence of consciousness. The commonly used Glasgow Coma Scale incorporates these dimensions.

Delirium, the acute confusional state, is usually characterized by a relatively abrupt onset and short duration of clouded, reduced, and fragmented attention; impaired memory and learning; perceptual and cognitive abnormalities, such as hallucinations and delusions; disrupted sleep; and other autonomic dysfunction. The level of consciousness may be consistently diminished or may fluctuate. The electroencephalogram (EEG) usually shows diffuse slowing. Typical motor abnormalities include an increase in general restlessness, fine and course tremors, and myoclonic jerks. Autonomic disturbances commonly include tachycardia, fever, elevated blood pressure, diaphoresis, and pupillary dilation. The causes of delirium are legion, including systemic medical disorders, such as metabolic imbalances or infections, intracranial disorders due to traumatic, structural, and electrical causes, and drug intoxication and withdrawal states.

A 45-year-old secretary was brought by ambulance as she appeared confused at work, trying to grab objects that were not there. Her coworkers reported that she has had problems with anxiety and was taking medications for that reason. In the emergency room, she continued to appear confused and reported that she had run out of alprazolam several days earlier. Review of the controlled substances prescription site confirmed that she had been receiving high doses of alprazolam on regular basis.

A 59-year-old truck driver at work collapsed while complaining of severe chest pain. He was rushed to the hospital, and it was determined that he needed emergency coronary bypass surgery. His surgery was successful, and he was slated for discharge the next morning. That night, the resident physician was urgently called to the floor by the nurse taking care of the patient, who stated that the patient was sweating profusely, confused, agitated, scratching his arms, and shouting that “spiders were crawling all over him.”

Attentional difficulties are manifested by impairments in the person's ability to deploy, focus, and sustain attention. Some attentional difficulties first appear in early childhood as developmental problems of uncertain cause and are described as attention-deficit disorders (inattentive or hyperactive types). Secondary attention-deficit difficulties may appear de novo in adulthood due to a variety of exogenous agents, psychiatric disorders, and late-life developmental and degenerative factors.

In narcolepsy, characterized by sudden lapses into sleep, one's usual ability to stay alert and maintain consciousness is impaired. At times, the onset of profound sleepiness is gradual, accompanied by hypnagogic phenomena, in which dreamlike images invade consciousness, and at other times the shift in consciousness appears to be almost instantaneous.

A 24-year-old recently engaged woman presented with her fiancé for evaluation. Her fiancé reported that, while shopping for her engagement ring, she was very excited and had an "attack" of being weak in her knees and falling down which lasted almost 1 minute. She reported fully being aware of her surroundings and hearing the clerk and her fiancé calling her name. During the evaluation, she acknowledged significant daytime sleepiness and that her mother suffered from narcolepsy.

1Altered States of Consciousness

Consciousness may also be qualitatively changed, with the production of altered states. Drugs such as scopolamine, with strong central anticholinergic properties, some seizures, and, on occasion, other conditions associated with delirium can induce twilight states, dreamlike states of wakeful consciousness in which attention is poor, an admixture of primary and secondary process thinking appears, and patients fade in and out of alertness. Dreamlike experiences intrude into the stream of conversation. Emotional outbursts or violent acts may occur during twilight states.

Mystical states of consciousness may occur in normal and pathologic conditions. Intense meditation and peak or epiphanic experiences, reported by more than 10% of normal individuals in community surveys, may produce a sense that the self dissolves or expands, that the self fuses mystically with the cosmos, that time stops, and that universal meaning becomes clear. These perceptions may be accompanied by a sense of rejuvenation and renewed personal identity, ineffability, intense emotionality, and concurrent perceptual changes. Such experiences do not ordinarily last more than a few minutes. Many people have achieved these states through the use of psychedelic agents such as mescaline and LSD. Reports of a white light at the end of a tunnel, described by individuals using psychedelics and in near-death experiences, have been linked to specific neurophysiologic pathways believed to be stimulated under these conditions.

A 43-year-old female accountant with no prior history of spiritual or religious practice sustained a severe head injury in an automobile accident that left her comatose for approximately 6 months. When she awoke, she reported that at some point during her coma, she estimated in the third month, she experienced a profound experience, which included a glorious union with God and knowing that she would only awaken to return to life. After recovery, she became profoundly spiritual, started attending church, reading the Bible, and demonstrated a beatific glow, which she attributed to the experience she had during the coma.

Although hypnosis lacks a consensually accepted definition, its hallmarks are selective attention, suggestibility, and dissociation. Most, but not all, people can be hypnotized to some degree. Up to 90% of people are capable of achieving a light trance, whereas 10% to 20% are capable of entering a deep trance and exhibiting remarkable hypnotic phenomena. Hypnosis occurs when the subject is in a state of heightened, not diminished, attention. EEG studies have shown hypnotized subjects to be fully awake and alert. The heightened concentration probably accounts for the unusual levels of sensory and motor performance often seen under hypnosis and self-hypnosis.

Hypnotic phenomena include hypnotically induced hallucinations (including negative hallucinations in which the subject selectively does not perceive sights, sounds, or other stimuli), anesthesia, sustained motor behaviors and acts of strength ordinarily beyond the individual's capacity, and distortions of memory (both hypermnesia and amnesia). Several phenomena that reveal the multiple nature of consciousness, for example, coconsciousness, are also demonstrable. Experiments have shown that even when a subject in deep trance has achieved profound hypnotic anesthesia and can, for example, keep a hand submerged in ice water for longer periods of time than usual, part of the hypnotized subject's consciousness continues to register exactly how painful the experience actually is and can signal the researcher about the pain, by finger movements, without the subject having any conscious awareness or disturbance. This phenomenon, called the hidden observer, has also been seen in postsurgical patients who, under hypnotic trance after surgery, have been able to accurately recall conversations in the operating room that occurred while they were under general anesthesia. Dissociative and psychosomatic phenomena have also been induced with hypnosis. With posthypnotic suggestion, for example, subjects may carry out complex actions without any hint that they are actually doing so because they were previously instructed to act that way under hypnosis. When asked why they are carrying out these activities, such subjects usually make up various reasons, although seemingly unaware of the real reasons for their actions. It has been suggested, not entirely facetiously, that many normal daily activities are conducted in a trancelike posthypnotic state, and although these activities are attributed to conscious intention, they may in fact be carried out due to previous suggestion. Advertisers know this well. Urticaria (hives) can be hypnotically induced and hypnotically made to disappear. Plantar warts have been successfully treated with hypnosis, and, in these conditions, diminished blood supplies to their bases have been demonstrated. It has recently been appreciated that yoga masters can exert remarkable control over basic bodily functions through self-hypnosis. As yet, little is known of the full extent to which heightened concentration may influence physiologic regulation.

Suggestibility

Pathologic suggestibility may be seen in several clinical conditions. Automatic obedience has been described in echolalia (the automatic repetition of a sentence or phrase just uttered by another person), echopraxia (the automatic mimicking of a movement performed by another person), and waxy flexibility (maintaining, for a prolonged period of time, a posture in which one is placed), symptoms common in catatonic states. In situations of group delusions, and sometimes in cults, passive individuals adopt the delusional beliefs of stronger ones. In epidemic hysteria, as described among young women at the Salem witch trials in Arthur Miller's *The Crucible,* distorted and even delusional perceptions and beliefs may sweep over a group that has been highly aroused by a charismatic leader.

Autosuggestibility can be seen in the constructions of false memories in which an individual progressively comes to believe that something that never happened in fact occurred. Such false memories may be held with such great conviction that they are indistinguishable from the memories of real events. Because of this, memories recovered during therapy (in which there is often great pressure to "remember" certain events) cannot be taken at face value without corroboration from other sources. Various types and degrees of self-deception may be more common in individuals who are more suggestible.

Dissociative Phenomena

Dissociation refers to the splitting off from one another of what are ordinarily closely connected behaviors, thoughts, or feelings. Dissociative states are those in which there is a disturbance or alteration in the normally integrated functions of identity, memory, or consciousness and include trances, fugues, blackouts, multiple personalities (dissociative identity disorder), and dissociative frenzies. Although dissociative states are ordinarily believed to be functional in nature, arising as an adaptive defense in individuals subjected to a great deal of trauma, particularly at early ages, they occur regularly with a variety of neurologic disorders, particularly those with partial complex seizures. In one series, one-third of patients with complex partial seizures had dissociative phenomena, including multiple personality. In these patients, the dissociative phenomena were not related to the seizure activity but to interictal alterations.

As in posthypnotic amnesia, elaborate activities can occur in dissociative states for which the subject has no conscious memory. This amnesia is functional in nature and may be reversed by hypnosis or drug-facilitated disinhibition, for example, with amobarbital (sodium amytal) infusion. In many of the functional dissociative states, amnestic episodes may occur for years or decades before the patient seeks medical or psychiatric attention. Blackouts are periods of amnesia in alcoholism, other intoxications, or after head trauma. An alcoholic blackout period may last for hours to days, after which the person has no recollection of what transpired, although other observers attest to the fact that, during this period, the individual carried out multiple complicated behaviors. Although memory of the blackout is lost to the predominant consciousness, during subsequent re-intoxication, memories of events occurring during the previous blackout may be reawakened. This phenomenon, known as state-dependent memory, occurs in many other conditions as well, signifying that one’s ability to retrieve specific memories may be highly influenced by specific physiologic alterations due to external intoxicants or other unusual physiologic states.

A 43-year-old man with a 20-year history of sustained alcohol abuse was unable, during a period of sobriety, to recall for hospital staff the names, phone numbers, and addresses of his close friends and associates. Several weeks later, when he returned to the emergency department in an intoxicated state, he was able to recall all of those details in a reasonably straightforward manner.

Psychogenic fugue is characterized by prolonged periods in which individuals carry out very complex activities without having any recollection for their previous lives, identities, or even names. They often travel away from customary locales and assume entirely new identities. By definition, psychogenic fugue cannot be due to a neurologic disorder. In comparison, the discontinuity of experience in psychogenic amnesia is typically more circumscribed and does not involve assuming an entirely new identity. The dissociated memories and affects often reveal themselves in disguised form such as nightmares, intrusive visual images, and conversion symptoms. Typically, in psychogenic amnesia, an individual may not be able to recollect what transpired during a specific period, for example, before the age of 9 or 10 years in the context of a traumatic childhood; during catastrophic events, such as traumatic, gruesome combat; or during less momentous events that a person prefers to forget to preserve self-esteem by denying shameful, immoral, or illegal activities.

A 28-year-old infantryman could not recollect the events of the combat situation in which he had been wounded and several of his comrades killed. Several months afterward, he participated in a series of amobarbital-facilitated interviews, during which the events of combat came back to him. He recalled becoming panicky during the firefight, turning, and running away while several other soldiers in his squad, some of whom were killed, yelled to try to get him to stay in place and fight. After the interviews, he cried inconsolably and blamed himself, somewhat irrationally, for their deaths, wondering if his remaining in place and fighting might have saved their lives.

Dissociative Identity Disorder. Dissociative identity disorder, previously known as multiple personality disorder, is a chronic dissociative state in which two or more separate ongoing identities or personalities alternative in consciousness. It usually occurs in people who, as young children, were severely and repeatedly brutalized. The number of identities is variable, with some cases reporting 25 or more identities. The development of dissociated alter personalities is believed to be a last-ditch, primitive psychological defense against inescapable and unbearable traumatic situations. The personalities may be of different ages and even different sexes. Typically, the presenting identity is dysphoric, anxious, and constricted, may have headaches and periods of blackout or amnesia, and is not aware of the other personalities. A second identity is typically vivacious and uninhibited.

A 34-year-old female had recently moved to a new city to start a new job. She presented for an intake for chronic depression and anxiety. She was meticulously groomed, seemed shy and soft-spoken, and was wearing very conservative clothing. During the interview, she suddenly became labile, angry, and loudly said, “Stop talking like that.” Then, shortly, she quietly put both her feet on the couch, and started rocking back and forth, speaking in a childlike voice.

Another identity may know all about the other personalities and has a wise perspective on the life events leading to the problems and of possible solutions. A classic case is described in the popular book and film, *The Three Faces of Eve.*

In the so-called channeling, dissociated, complex part personalities are produced in trance states in which fictitious past lives or spirit lives are created.

Ganser Syndrome. In Ganser syndrome, the patient responds to questions by giving approximate or patently ridiculous answers; for example, in response to the question, “What sound does a dog make?” the patient answers, “Moo.” Additional features of the syndrome include alterations in consciousness, hallucinations (or pseudohallucinations), conversion phenomena, and amnesia for the episode during which these symptoms are manifested. This syndrome has most commonly been reported in prisoners and is generally believed to be a dissociative state, although organic features may be contributing factors.

A 23-year-old man was arrested for auto theft. He had a long history of brushes with the law, including the juvenile justice system, for fire-setting, shoplifting, and physical fights. While in jail, he stated he was confused, could not remember what led to his arrest, appeared distracted at times, pleasant and smiling at other times, and would answer simple questions such as 1 + 1 as equal to 3.

Depersonalization and Derealization. Depersonalization refers to an alteration in one's experience and awareness of the self, leading to feelings of being unreal or detached from one's own body or of feeling like an automaton; it is often accompanied by the complaint that the individual lacks all feelings or sensory experiences. Those experiencing depersonalization frequently fear that they are going crazy. Because of this fear, patients often endure depersonalization experiences for long periods before describing them to a mental health professional. Depersonalization is also characterized by frequent internal nonaudible dialogues between the participating self and the observing self but with full awareness that both parties are the same person (a feature that distinguishes it from hallucinations). Mild sensory distortions—but not hallucinations—are commonly associated with the experience. Depersonalization is seen in a variety of neurologic and psychiatric disorders and is common in complex partial seizures. It may occur in the context of depression, anxiety disorders, or certain personality disorders, or it may occur as an entity by itself. In derealization, individuals feel themselves to be real but feel that the world around them has suddenly become unreal. Derealization often, but not always, accompanies depersonalization. Transient episodes of depersonalization and derealization occur frequently in normal people, particularly during states of fatigue, sleep deprivation, or stressful situations such as bereavement, learning of a terminal diagnosis, or sudden awareness that one is about to be in an inescapable vehicle collision.

A 42-year-old school bus driver was unable to see a large object in the road in a severe thunderstorm and had an unavoidable collision. Although she was able to swerve the school bus at the last minute, the bus sideswiped a vehicle before coming to a screeching halt. When the police arrived, she could only say that she felt that she was “floating outside her body” and that the “world did not seem real.”

DISTURBANCES OF THE SELF

At the most basic level, the key components of self-awareness are the reality and integrity of the self (that I am one person), the continuity of self (that I am the same person now that I was in the past and that I will be in the future), the boundaries of self (that I can distinguish between myself and the rest of the world as not-self), and activity of self (that it is I who is thinking, doing, and feeling). Additional components of a sense of self include body image and various self-evaluations, including self-esteem and ego-ideal (ideal self). Body image is an individual’s mental representation of their own body. Self-esteem is believed to reflect how one measures up to the desired self-image. To the extent that what one sees in oneself approximates what one would like to be, self-esteem is positive. Ego-ideals are fantasies of the optimum person one could ever wish to be. Any of these qualities may be disturbed in psychiatric disorders.

Within each individual is a group of social selves comprised of the roles and identities that a person assumes and that are evoked in various contexts. The presenting “self” varies depending on the people with whom one interacts, such as parents, romantic partner, child, friend, or employer, and depending on what role one assumes, such as child, parent, colleague, or lover. Accompanying each of these “selves” are various levels of objective and subjective self-awareness and self-understanding.

Disturbances in Sense of Self

Disturbances of the basic elements of self-awareness may be seen in a variety of disorders. Discontinuity phenomena are characteristic of dissociative states such as psychogenic amnesias and psychogenic fugue. Depersonalization reflects a mild disturbance in the awareness of self as the agent of activity. More severe disturbance is characteristic of the psychotic passivity phenomena seen in schizophrenia. Boundary disturbances may be considered characteristic of all psychotic states regardless of diagnosis.

Disorders of self-integrity are characteristic of both dissociative identity disorders and severe borderline personality disorders in which a person's self-concept and expression of this concept to others is erratic, leading to a sense of unstable identity.

False self describes a persona or a faulty and limited superficial aspect of the personality that an individual builds up as a mechanism for adapting to a hostile world—to please, control, or negotiate with others and with him- or herself. However, the false self does not incorporate, integrate, or validate important fundamental needs, wants, values, and beliefs. Through self-deception and denial, the individual may consciously believe that this "self" constitutes their entire being. However, the false self is a relatively fragile construction that has usually warded off and denied fundamental strivings, which may include needs for autonomy; acting with integrity; expressing certain desires, beliefs, or talents; or other unacknowledged aspects of the self. When these warded-off needs finally break through and demand expression at various points in development, the defective false self may collapse, leading to a period of distress and identity confusion, which individuals sometimes describe as a "nervous breakdown."

Patients with pseudologia fantastica and the impostor syndrome demonstrate extreme examples of inconsistency in the sense of self. In pseudologia fantastica, patients compulsively spin out webs of lies, ordinarily self-aggrandizing ones, and also appear to be trying very hard to deceive themselves into believing that they are true. In the impostor syndrome, such fantasies are acted out by liars and impostors who seem to fervently wish that these fantasies were their reality, as if they cannot accept themselves and would be overwhelmingly ashamed to be known for who they actually are. The impostor compulsively adopts the identities of others and may, for example, show up properly attired at diplomatic functions and society galas and interact with the other guests under the assumed identity. Some famous impostors have repeatedly insinuated themselves into inner circles of high society and government.

Transsexualism is a syndrome characterized by the feeling that one was born into a body of the wrong sex and marked by the desire, starting at an early age, to be a person of the opposite sex. Male to female transsexualism is reported most often. Both psychodynamic and biologic theories have been advanced to explain these unusual phenomena. Of note, studies following patients who undergo successful male to female transgender surgery have not revealed higher rates of associated psychopathology than in comparison groups.

Self-esteem is a measure of one’s self-appraisal in relation to one’s values and ego-ideals. Negative self-esteem is characteristic of depressive disorders, many personality disorders, and situational failures. Superficially inflated self-esteem may be seen in mania (or hypomania) or, in a fluctuating manner, with narcissistic and other personality disorders.

Although some individuals regard their ego-ideals as unattainable and are content to live as imperfect human beings, others strive to approximate their ideals. People who feel driven to achieve unattainable ideal goals or to become unrealistically perfect ideal selves are likely to be chronically dysphoric and have poor self-esteem because their attempts to become their ego-ideals are doomed to failure.

DISTURBANCES OF MEMORY

Memory is not a unitary phenomenon. Capacities to remember vary for the different senses and perceptions. One person may have prodigious musical memory, with the capacity to remember and reproduce whole musical pieces after one hearing but be incapable of remembering people's names or telephone numbers. Exceptionally detailed verbal memories have been associated with obsessional cognitive styles. When individuals with extraordinary memories complain of memory loss, ordinary memory tests may be inadequate to detect their deficits, as their relative memory loss may have reduced their capacities to a point within the range of most normal people.

Memory functions have been divided into three stages: registration, retention, and recall. Registration (or acquisition) refers to the capacity to add new material to memory. The material may be sensory, perceptual, or conceptual and may come from the environment or from within the person. For new material to be acquired, the person must attend to the information presented, and it must then be registered through the appropriate sensory channels and then be processed or cortically organized. Retention is the ability to hold memories in storage. Large numbers of neurons are believed to be involved in the storage of a specific memory, and it is believed that reverberating circuits are formed in which memory traces are held by means of changes in proteins or synaptic connectivity, or both. Recall is the capacity to return previously stored memories to consciousness.

Newly registered material is transferred incrementally from immediate to short-term memory to long-term memory. Immediate memory lasts for 15 to 20 seconds, short-term memory (or recent memory) for several minutes up to 2 days (the time involved in new learning and its early consolidation), and long-term (or remote) memory for longer periods. Different physiologic processes mediate each of these stages of memory. Because of this, processes that affect immediate or short-term memory often spare long-term memory. The processes by which memories are transferred from short- to long-term stores are unknown.

Cognitive scientists now refer to short-term memory as working memory, the system that briefly stores and processes information needed for planning and reasoning. Recent studies suggest that the working memory system consists of at least two short-term memory buffers, one for verbal and another for visual memories, plus a central executive that manipulates and coordinates information stored in the two buffers for problem solving, planning, and organizing activities. Parallel processing systems involving specific areas of the prefrontal cortex and other brain areas appear to operate separately with respect to various processes concerned with working memory. For example, separate prefrontal areas appear to be involved in working memory functions concerned with object identity and spatial locations.

Other studies suggest that different types of memories are stored and retrieved by different brain systems, so there is at least a dual memory system. The first system, sometimes called a conditioned-emotional system, system for implicit memory, perceptual memory, or nondeclarative memory, is present from birth, operational through life, and addressable by situational, sensory, or affective cues. Past experiences are expressed through images, behaviors, or emotions. These memories need not involve any conscious memories of a past experience. Conditioned fear responses represent examples of memories elicited in this system. The second system, sometimes called narrative–biographical memory, or explicit memory, reflective memory, or declarative memory, emerges during the preschool years and includes information significant to the self. Memories are addressable through intentional retrieval efforts, apart from the original learning conditions. They are identified as representing personally experienced events and compose the individual’s life history, approximately equivalent to memory with consciousness or memory with awareness. Clinical studies suggest that, in at least some amnesias, implicit and explicit memory functions may be dissociated.

Disturbances in memory occur through the interruption of registration, retention, or recall.

Disturbances in Registration

Registration and short-term memory retention are usually impaired in disorders that affect vigilance and attention, such as head trauma, delirium, intoxication, psychosis, spontaneous or induced seizures, anxiety, depression, and fatigue. A variety of other metabolic and structural brain disturbances can affect short-term memory as well, particularly lesions affecting the mammillary bodies, hippocampus, fornix, and closely associated areas. Patients with impaired attention and concentration who are able to demonstrate immediate recall may not be able to retain or recollect these items from short-term memory. Benzodiazepine use has been associated with working memory difficulties, especially in the older adults. Some short-acting, high-potency benzodiazepines used as sleeping pills may be particularly troublesome in this regard.

Disturbances in Retention

The retention of memories is impaired in posttraumatic amnesia as well as in a number of cognitive disorders, such as dementia of the Alzheimer type and Wernicke–Korsakoff syndrome. The latter, which ordinarily results from chronic thiamine deficiency seen with alcoholism, is associated with pathologic alterations in the mammillary bodies and thalamus.

Disturbances in Recall

Disturbances in recall can occur even when memories have been registered and are in storage. At times, inability to recall may signify that the memory traces themselves have disappeared and are no longer retrievable. However, difficulties in recall can occur separately, as in the everyday event of forgetting the name of a person or object, only to spontaneously remember it hours or days later. In normal forgetting, more remote events are less well remembered than recent ones, and important events are most vividly retained in memory. Some demented patients may lose memories for all events occurring after a specific date or event, as if the slate has been wiped clean, but retain earlier memories. Some individuals may progressively erase memories so that they recall only earlier and earlier events. Specific types of memory functions may be subject to selectively different control mechanisms in the brain. For example, recent animal research suggests that endocannabinoids may facilitate the extinction of aversive memories, but not their acquisition or consolidation, through their selective inhibitory effects on local inhibitory networks in the amygdala.

A 21-year-old pregnant woman experiences severe PTSD symptoms related to the trauma of witnessing the father of her baby commit suicide in front of her by shooting himself. She has been self-medicating with two bottles of wine daily, combined with “handlebars” (benzodiazepine bars), which she reports allow her to “forget” and not dwell on the past while getting about her usual day-to-day responsibilities.

Under usual conditions, forgotten events can be recalled with prompting, associative memories, or other forms of stimulation such as hypnosis. As described earlier, state-dependent memories are recall failures, reversed by reinstituting the context in which the memory was originally formed.

Amnesias are syndromes in which short- and long-term memory are impaired within a state of normal consciousness. Thus, memory disturbances in delirium should, strictly speaking, not be considered amnestic syndromes. Anterograde amnesia is the inability to register or learn new information (and therefore to form new memories) from a specific event onward; it typically follows head trauma, states of cerebral physiologic imbalance, or drug effects. Patients who receive electroconvulsive therapy (ECT) frequently have anterograde amnesias during the course of the treatments; the amnesia gradually fades over numbers of weeks. Retrograde amnesia is an impairment in recalling memories that were established before a traumatic event, extending backward in time for variable periods. As memory is regained, the more remote memories usually return first. A patient originally amnestic for the 3-month period before an accident may ultimately be left with amnesia for events only a day or an hour just before the accident. In organically caused retrograde amnesias, remote memories are usually intact, although amnesia may exist for more recent events. This contrasts with psychogenic (functional) amnesia, in which the periods of forgotten events may be more spotty or selective.

Hypermnesia, unusually detailed and vivid memory, may occur in gifted people, in association with obsessive-compulsive and paranoid personality traits, and in hypnotic trances. Intrusive memories may occur in PTSD, signaling failure of the mechanisms that usually keep unwanted memories and information out of working memory.

A 28-year-old firefighter and his partner were the first to respond to a five-alarm fire at a popular local business. The firefighters were valiant in their efforts, however, one employee and the firefighter's partner perished. Immediately thereafter, the firefighter reported inability to fall asleep, vivid dreams of the event, sweating, feeling his heart beating fast, smelling smoke, and hearing his partner's voice calling for help.

Although many forgotten memories can be recalled in hypnotic trance, retrospective falsification and distortion may also occur under hypnosis. (Memories recalled under hypnosis usually are not accepted as evidence in court.) Retrospective falsification of memory, the development of false memories, is called paramnesia, also known as *fausse reconnaissance.* Confabulation is another common form of paramnesia in which the patient fills in memory gaps with inaccurate information. The responses given to questions by patients who confabulate may reflect past experiences or bizarre, fantastic stories. Confabulation correlates poorly with memory deficit and is believed to reflect frontal lobe dysfunction and a failure of self-monitoring. Confabulation is prominent in certain alcohol amnestic syndromes, such as Wernicke–Korsakoff syndrome, as well as other disorders of the mammillary bodies, thalamus, or frontal lobes.

Déjà vu is the sense that one has previously seen or experienced what is transpiring for the first time; it is a false impression that the current stream of consciousness has previously been recorded in memory. Related phenomena are *déjà entendu,* a sense that one has previously heard what is actually being heard for the first time, and *déjà pensé,* a feeling that one has at an earlier time known or understood what is being thought for the first time. Experiences of *jamais vu, jamais entendu,* and *jamais pensé* involve feelings that one has never seen, heard, or thought (respectively) things that, in fact, one has. These phenomena are all common in everyday life but may increase in states of fatigue or intoxication and in association with complex partial seizures or other psychopathological states.

A 28-year-old music teacher drove 12 hours with minimal breaks in order to attend a music festival for the first time. Upon entering the music hall, he reported suddenly feeling sure that he has been to this music hall before, sat in the same seat, and listened to the same music.

Dementia is a syndrome in which the essential feature is an acquired impairment of short- and long-term memory with associated impairments of abstract thinking, judgment, personality changes, and other cortical disturbances. The symptoms always involve more than one sphere of function. In later stages, demented patients may become helpless, too confused to use a stove, and incapable of remembering the names of close relatives. They may wander into dangerous situations, oblivious of their surroundings. Dementias are caused by a variety of pathogenic processes, some of which are reversible, such as hypothyroidism and subdural hematoma, whereas others are irreversible, such as dementia of the Alzheimer type and multi-infarct dementia. Although the characteristic cognitive disturbances seen in severe major depressive episodes are usually called pseudodementias, many neuropsychiatrists believe that profound cognitive dysfunction meeting criteria for dementia associated with depression should properly be labeled as a reversible dementia syndrome. The presence of pseudodementia, however, is predictive of the development of Alzheimer disease (but not in all cases), indicating that the depression was a prodrome of the primary cognitive disorder or that the presence of pseudodementia marks those with an underlying progressive cognitive disorder.

LANGUAGE DISORDERS

Communication difficulties may be due to disorders of thinking as previously described, abnormal speech patterns in mood disturbances and schizophrenia, or primary speech fluency disorders, such as stuttering and stammering, disorders of the articulation and speech apparatus, and CNS disturbances involved in hearing and speech generation (aphasias).

Manic patients typically exhibit pressured speech, in which the speed of the word stream is accelerated. If severe, then the speech may be garbled, imprecise, and difficult to understand. Patients with psychomotor retardation speak slowly and monotonously and have a long speech latency in response to questions. Schizophrenic patients may exhibit a variety of speech abnormalities, including poverty of speech and poverty of content of speech (in which the amount of speaking is normal, but understanding the central message is difficult or impossible). Schizophrenic patients may also be difficult to understand because of the dysarthric effect of antipsychotic medication.

Allusory speech is vague, imprecise, and hard to comprehend because too few cues and details are provided for the listener. Such speech may be heard from some patients with schizophrenia or certain personality disorders, or even normal individuals who wish to convey a sense of mystery by just being suggestive, whose suspiciousness causes them to be reluctant to spell things out clearly, or who believe that the listener is more aware of their private codes, meanings, and allusions than is the case.

Stuttering and stammering (ordinarily synonymous) refer to disturbances in the rhythm and fluency of speech due to blocking, convulsive repetition, or prolongation of sounds. This disorder affects males two to three times as often as females, and there is a high rate of familial transmission.

Aphasias, impairments of language produced by brain dysfunction, are ordinarily described as being fluent and nonfluent. In fluent aphasias, which generally reflect dysfunction in the left temporal and parietal area, patients have a normal or even elevated verbal output, sometimes with logorrhea, but they ignore the social conventions of conversation. They produce many well-articulated phrases with normal prosody, but there is little informational content. The fluent aphasias are further divided according to the extent of comprehension by the patient and the ability of the patient to repeat what the examiner says. The principal fluent aphasias are Wernicke's aphasia, conduction aphasia, anomic aphasia, and transcortical sensory aphasia.

Nonfluent aphasias are characterized by slow and poor verbal output, difficulty with spontaneous speech, omission of grammatical connecting words, and poor prosody. Patients may produce one-word replies or very short phrases. Brain lesions that cause nonfluent aphasias typically tend to occur in the anterior left hemisphere. The principal nonfluent aphasias are Broca's aphasia, transcortical motor aphasia, global aphasia, and the mixed transcortical aphasias.

In aprosodias, the nonverbal aspects of speech, the melody, pauses, timing, stress, accent, and intonation are impaired. Damage to the right prefrontal region has been associated with expressive aprosodias, and damage to the right temporal region and insula has been associated with receptive aprosodias.

DISORDERS OF THE WILL

Central to the sense of self is the concept of will or volition. Psychologically, will is linked to the concepts of intentionality and of transforming awareness and knowledge into initiating action, as the bridge between desire and action. For individuals to manifest normal will, they must be aware and feel desires—and these desires must arise from within themselves. Concepts related to will that may become the focus of clinical attention when disturbed include motivation and decision making (i.e., the capacity to make choices).

Pathologically heightened will, seen primarily in manic states, is characterized by excessively intense desires and an overly facile capacity to make decisions, with complex questions being decided on in an instant. With heightened psychological energy, these individuals can start new courses of action with astonishing rapidity. Closer examination of these actions in more extreme cases, however, reveals that they share much in common with decreased will in that the intense desires and quick decisions often reflect impulsiveness, which can be considered an escape from true willing and decision making, rather than enduring desires or thoughtful decision making.

The term abulia has been used to describe the loss, lack, or impairment of the power to will or to execute what is in mind. Abulic individuals show a diminished sense of motive or desire and impairment in making the transition from motive and desire to execution of action. Deficiencies in will may be seen in a variety of psychiatric disorders and at the end of life, when patients have surrendered their will to live and are simply waiting to die. In schizophrenia, a diminished sense of will can be seen in passivity phenomena, as well as in other negative (or deficit) symptoms that may affect thoughts, feelings, and behaviors. These include lack of drive, impersistence at tasks, and a general inner flatness. Depressed patients also describe volitional disturbances in general apathy and anhedonia. Patients who chronically inhale solvents (e.g., glue, gasoline, or toluene), smoke marijuana very heavily, and chronically use hallucinogens have a characteristic amotivational syndrome. The extent to which this lack of motivation results from or contributes to the chronic substance abuse is a matter of debate.

In OCD, both the obsessional thoughts and the compulsive rituals are experienced as ego dystonic and not consonant with the patient’s conscious desires and will. Similarly, although patients with anorexia nervosa initially have the conscious experience of willing and controlling their intake of food, during the course of the disorder the sense of willfulness is replaced by one of passivity of being subjugated by obsessional thoughts and compulsive behaviors that assume control of the eating behavior.

A 28-year-old woman with anorexia nervosa and OCD reports struggling with grocery shopping, which routinely takes her 4 to 5 hours to complete. She reports that she feels compelled to research each item, particularly the fat and sugar content. She has a “calorie calculator” running in her mind, and for each item she purchases, she feels pressure to determine what else she will eat with it so that she does not exceed her calorie limit for each meal. She reports intrusive thoughts that she is fat and needs to keep counting her daily caloric intake. These intrusive thoughts are experienced in parallel to but separate from any other thinking activities in which she happened to be engaged.

Disturbances of volition are among the more common complaints of patients with personality disturbances who request psychotherapy. Individuals with dependent personalities are characterized by difficulties in making decisions by themselves and often engage in courses of action contrary to their own desires. Similarly, individuals with passive–aggressive personalities obscure their own desires by being excessively involved in the demands made on them by others. Their courses of action do not reflect their own decisions so much as the thwarting of others’ desires. People with compulsive personalities use inflexible rules, thereby precluding courses of action based on independent evaluation, individual desires, and decisions. In other situations, they are indecisive, sometimes making impulsive decisions at the last minute when forced to decide. Finally, many individuals seek treatment because of self-designated disturbances of will: they do not know what they want, cannot make choices among several options, or procrastinate excessively. Often, these problems may mask other fears—of wanting, commitment, taking initiative, hard work, success, making a mistake, being criticized, angering others, and all the consequences related to such actions.

DISTURBANCES OF MOOD

Defining, describing, understanding, and categorizing moods have long been among the most important and difficult tasks in psychiatry. The language of feelings is filled with terms that seem to have mostly idiosyncratic meanings, as patients, phenomenologists, and psychiatrists all struggle both to describe inner emotions and to correlate them with external behavior. Even basic terms, such as mood, affect, emotion, and feelings, lack universal definition. The most common convention, which is used here, defines mood as a sustained or prevailing subjective feeling tone or range of tones. Affect is the moment-to-moment feeling state, sometimes rapidly shifting in response to a variety of thoughts and situations, that the clinician can observe. Emotions have been defined as moods and affects that are connected to specific ideas or to the physical concomitants of moods and affects. Feelings are the most poorly defined of all, leading Karl Jaspers to ultimately describe them as everything for which there is no other name. In common parlance, and often professionally as well, these words are sometimes used interchangeably. The term mood disorders, adopted for the DSM-III-R and DSM-IV-TR, replaced “affective disorders” to describe the same group of psychiatric syndromes, and in DSM-5 it has been sub-divided into bipolar and related disorders and depressive disorders.

Moods, affects, and emotions can be described by a number of important qualities: intensity (shallow to deep), range (broad to narrow [or flat]), stability (rigid to labile), reactivity to external events (none too much), periodicity (periodic to aperiodic), congruence with thought content (congruent [or appropriate] to incongruent), speed of resolution (rapid to slow), and viscosity (evanescent to persistent). The individual’s lifelong predominant mood is one component of temperament. Thus, for example, one may be described as having a calm, buoyant, irritable, depressive, anxious, or sensitive temperament.

Moods, affects, and emotions serve as internal and external signal systems. They signal the state of the individual to others and often elicit necessary help and support from the environment. A baby’s face communicates its state of need, tension, or contentment, thereby recruiting appropriate maternal interventions. As adults, much of the most important interpersonal communications are transmitted nonverbally through cues that signal the observer about our moods. Positive words communicated by a scowling or sullen face lead listeners to perceive an angry message regardless of the spoken words. Moods also have an infectious quality and serve as important ways of influencing others. Thus, when people act cheerfully toward others, they, in turn, are more likely to feel cheerful and to reciprocate that cheerfulness.

Internally, moods, affects, and emotions let individuals know how well or how poorly they are doing, allowing them, for instance, to gauge the distance between actual self-appraisal and desired self-expectations. For example, individuals who desire to master important goals and feel that they have a reasonably good chance of doing so ordinarily experience pleasant emotional states in relation to these goals. If something intervenes to prevent them from reaching these goals so that there is an insurmountable gap between their desires and the likelihood of success, then they may feel hopeless. In addition to serving as signal systems, emotional states of nonspecific tension, arousal, or anger usually imply that some action is necessary to secure their discharge or release.

Emotional states and their expression are regulated by biologic, psychological, and cultural influences. For example, emotional or affective lability, characterized by rapidly shifting emotions that seem unattached to the situation, typically occurs premenstrually in some women, with varying periodicity in cyclothymic individuals and in those with cluster B personality disorders, and in relation to need states such as hunger, sleep deprivation, and sexual frustration. Mood shifts have also been described with environment-related physiologic influences such as seasonal changes in light. Psychological regulation of emotions may be related to specific coping mechanisms and the ability to self-soothe, which are developmentally determined. Conscious and preconscious psychological mechanisms, including varieties of self-talk, may help calm or inflame emotions. Cultural factors significantly regulate emotional expression. Although the facial expressions for basic emotions are similar in all cultures studied, the range and style of emotional expression permitted in relation to specific contexts varies greatly from culture to culture and from family to family. Some cultures and families are stiff lipped and inhibit the open expression of emotion, whereas others encourage emotional display. Marked differences exist among cultures in the emotional expression of acute grief, fear, pain, and affection.

Depression

The term depression has been used variously to describe an emotional state, a syndrome, and a group of specific disorders. When seen as part of a syndrome or disorder, depression has autonomic, visceral, emotional, perceptual, cognitive, and behavioral manifestations. As a nonpathologic, ubiquitous mood state lasting hours to days, but sometimes longer, feelings of depression are synonymous with feeling sad, blue, down in the dumps, unhappy, and miserable. Depressed mood is common and appropriate after a disappointment or loss. For most people, innate psychological resilience, alternative coping options, and supportive social networks help to alleviate these brief depressive states and prevent them from becoming chronic. Some individuals have a chronically depressed mood, tend to view the world as a difficult place filled with obstacles and burdens, see themselves as victimized, and lack hope for the future. The extent to which constitutional, developmental, and ongoing aversive life events contribute to this pervasive world view is unknown. People who, in early life, were deprived and traumatized may be less resilient and more prone to chronic depressive features than others. Repeated failures and the impact of unrelenting, uncontrollable, and unpredictable negative life events may set the stage for learned helplessness in humans just as they do in animals. A subset of chronically depressed individuals may also have temperamental, biologically driven depression, often seen in conjunction with strong genetic loading for severe mood disorders.

Some depressive states are normal and common reactions to major unwelcome and undesirable life events. Normal bereavement best exemplifies this. In bereavement after major losses, such as the death of a parent, spouse, or child, people experience sadness, pining, and yearning but do not ordinarily have the feelings of guilt, unworthiness, and self-reproach that characterize depressive disorders. Feelings of helplessness and hopelessness may be temporarily present in bereavement, but they ordinarily pass with time. In uncomplicated cases, the process of bereavement takes 3 to 6 months in the acute phase and up to 1 year for complete resolution. Bereaved people are more likely to feel physically ill and seek general health care than at other times, and older widowers are more likely to die than age-matched nonbereaved controls. Pathologic or prolonged grief reactions, such as bereavements that last more than 1 year, may be seen when the surviving spouse was excessively dependent on the deceased and is unable to obtain emotional and practical (e.g., financial) support elsewhere or when the survivor is unable to grieve fully because of markedly ambivalent feelings toward the deceased. Distinguishing between pathologic or prolonged grief and grief triggering a depressive episode may not always be possible.

A 33-year-old woman was referred for a psychiatric evaluation by her primary care provider after losing her baby a few hours after birth secondary to anencephaly. The patient reported that she is struggling with low energy, guilt, low appetite, sleep problems, and difficulty concentrating. Her family reports that she has made statements that she wants to join her baby in death as she has no other purpose to live.

A 52-year-old Hispanic woman was referred to see a psychiatrist by her longstanding family physician who had been treating her for multiple medical problems. Eight months ago, her husband of approximately 25 years passed away after a lengthy illness. She subsequently moved to a different town to live with her daughter and her family. The family became increasingly concerned that she had lost 30 lb, was withdrawn, and was continuously chanting on her rosary beads and “begging to die” in order to join her husband in heaven and take care of him.

The inadequate expression of grief due to incomplete bereavement is believed to be pathogenic in many subsequent psychiatric disorders. For example, impulsive acting out among adolescents who previously lost a parent is often assumed to be due to unresolved grief.

A variety of medical disorders may cause depressive syndromes. Most common among these are endocrine abnormalities, such as hypothyroidism and hyperparathyroidism, and CNS disorders, such as cerebrovascular diseases and Parkinson’s disease. Depressions are more common in strokes affecting left anterior lesions than other locations. Some medications, especially antihypertensive agents affecting adrenergic tone, such as reserpine and possibly β-blockers, may also trigger depressions. The importance of a genetic diathesis in these iatrogenic depressions is not yet known. Depressive syndromes and disorders in general, however, are unquestionably familial and are likely to have genetic contributions, especially in depressions associated with bipolar disorder.

Cognitive features of depression are prominent. Characterizing the exact nature of the memory impairment using standardized tests has been difficult. The cognitive tasks requiring sustained effort and elaborate cognitive processing may be more disrupted in depression than are those tasks that can be accomplished more automatically. The so-called cognitive triad of depression consists of pervasive cognitive schema related to feelings of worthlessness, helplessness, and hopelessness—expectations that no one and nothing can, or is likely to, help, either now or in the future; in other words, “I’m not OK, the world is not OK, and it’s never going to get any better.”

Suicidal phenomena are of particular concern. Suicide is common in severe depressive disorders. Depressed patients compose the largest diagnostic group of all completed suicides. However, suicide occurs at high rates in many other conditions as well, notably substance use disorders, schizophrenia, and severe personality disorders. Suicide may occur in these conditions with or without a diagnosable comorbid depressive disorder. Depressed patients with comorbid alcohol use disorder may be at particularly high risk for suicide.

Although consistent, useful, validated predictors of suicide do not exist, certain demographic features are associated with higher risk. These include being white, male, and older and living alone. In the psychiatric history, the single most important factor is that of past suicide attempts. A history of violent behavior may also predict suicide. Murderers have a very high suicide rate, especially those who murder family members during episodes of domestic violence. Among clinical signs, hopelessness, anhedonia, and severe anxiety may predict increased suicide risk. Serious physical illness in association with other risk factors, such as depression, may place a patient at higher risk. A genetic predisposition toward suicidal behavior cuts across diagnostic lines and plays a role in suicide risk. This may reflect a tendency toward impulsive behavior, correlating with low CNS levels of 5-hydroxyindoleacetic acid (5-HIAA), the major metabolite of serotonin.

Suicidal gestures and various acts of self-mutilation are also common among impulsive, dependent, and self-hating depressed people, for whom they serve as tension-releasing behaviors and as cries for help that may enlist desired social support. Because suicidal gestures have been associated with an increased risk for subsequently completed suicide, they should not be taken lightly. Such self-harm behaviors have been shown to persist in patients with borderline personality disorder well into the sixth decade of life. Contrary to earlier thinking, they do not appear to “burn out” with age. Subintentional suicide may result when suicidal gestures go awry or when reckless behavior, such as taking unnecessary risks in combat or driving while drunk, prove fatal.

Elated Moods

Elated moods include euphoria, elation, exaltation, and ecstasy. They are marked by feelings of well-being and expansiveness, optimism, capability, pleasure, and grace. Such moods are normally experienced when life is going very well, when long-sought-after goals are achieved, and in states of love, religious fervor, and spiritual transcendence. Peak experiences and experiences of mystic fusion are often accompanied by feelings of exaltation and ecstasy. Sexual pleasure and some chemically mediated states of altered consciousness may also induce these feelings.

Abnormal elated moods are primarily seen as part of manic states and from the effects of certain medications and street drugs. When subtle, as in hypomania, the mood can be ebullient and brimming with self-confidence but with occasional irritability. Other characteristic symptoms of hypomania are increased energy, decreased need for sleep, rapidly flowing thoughts, excessive talking, inflated self-esteem with a demanding nature toward others, and diminished judgment. Mania is a more extreme state in which judgment and sleep are impaired to the point of marked functional disruption. As the mania exacerbates, irritability and anger increase, alternating rapidly with a brittle expansiveness. Cognitions become increasingly disorganized. Psychotic symptoms, usually involving themes of grandiosity or specialness, occur in 50% or more of manic patients. With increasing escalation of the manic state, thinking becomes very fragmented, psychotic symptoms are more prominent, and the syndrome may appear indistinguishable from acute schizophrenia. These three manic states—hypomania, mania, and the psychotic, fragmented manic state—have been referred to as stage I, II, and III mania, respectively. Manic states occur in bipolar disorder, unspecified bipolar disorder, and cyclothymia and as a secondary mania caused by a variety of physical and toxic conditions. Such secondary manias may follow specific cerebral insults, accompany systemic disorders, or occur after ingestion of some drugs, including amphetamines, antidepressants, bromocriptine, decongestants, and corticosteroids, among others. Mania is the second most common neuropsychiatric disturbance induced by steroids, occurring in 30% to 35% of patients who develop steroid-induced behavioral disorders. Up to 12% of patients treated with levodopa and bromocriptine for parkinsonism develop mania. Right hemispheric brain lesions are specifically associated with secondary mania.

Hyperthymic personality refers to personality characteristics similar to hypomania, although DSM-5 has no mention of this in its personality section. These include unusual energy, ebullience, confidence, intensity, and so forth but without either the episodic course or the functional impairment of hypomania. Whether hyperthymic personality should be considered a bipolar spectrum disorder or refers simply to high-energy individuals is still unclear.

ANXIETY DISORDERS

Like depression, the term anxiety refers to a number of different entities: a normal transient feeling, often with adaptive functions; a symptom seen in a wide variety of disorders; and a group of disorders in which the symptom of anxiety forms a dominant element. As a transient, disagreeable emotional state, anxiety may be adaptive, signaling anticipated or impending threat and motivating necessary action. In contrast to fear, the emotional state that exists when a source of threat is precise and well known, anxiety occurs when the threat is not well defined.

Patients often find it difficult to describe feelings of anxiety precisely; at its core, however, anxiety is characterized by intense negative affect, associated with an undefined threat to one’s physical or psychological self. Patients use terms such as tense, panicky, terrified, jittery, nervous, wound up, apprehensive, and worried to describe their sensations. Anxiety is additionally characterized by somatic, cognitive, behavioral, and perceptual symptoms. The somatic symptoms of anxiety are legion and often dominate; a partial list includes twitching, tremors, hot and cold flashes, sweating, palpitations, chest tightness, difficulty swallowing, nausea, diarrhea, dry mouth, and decreased libido. Cognitively, anxiety is characterized by hypervigilance, poor concentration, subjective confusion, fears of losing control or of going crazy, and catastrophic thinking. Behavioral symptoms include fearful expressions, withdrawal, irritability, immobility, and hyperventilation. Perceptual disturbances, including depersonalization, derealization, and hyperesthesia (especially hyperacusis), are also common.

Trait anxiety refers to a lifelong pattern of anxiety as a feature of temperament. Individuals with trait anxiety are skittish, hypersensitive to stimuli, psychophysiologically more reactive than others, and exhibit catastrophic thinking. In contrast, state anxiety refers to episodes of anxiety that are tightly bound to specific situations and do not persist after the provoking situation has abated. Free-floating anxiety is characterized by a persistently anxious mood in which the cause is unknown and in which large numbers of diverse thoughts and events all seem to trigger and compound the anxiety. In contrast, situational anxiety occurs only in relation to specific occasions or external stimuli, as in phobias.

Anxiety symptoms can result from numerous physical conditions as well as from other psychiatric disorders. Many endocrine, autoimmune, metabolic, and toxic disorders, as well as medication side effects, are known to generate anxiety. The psychiatrist must differentiate the response of the patient to an underlying condition (e.g., secondary anxiety) from symptoms generated by the primary disorder itself. In psychiatric populations, anxiety symptoms are prevalent among patients with psychoses, organic mental disorders, depression, and substance use disorders, as well as the specific anxiety disorders. In patients with schizophrenia, anxiety must be differentiated from akathisia, a common and often overlooked syndrome of subjective restlessness, anxiety, and agitation resulting from antipsychotic medication. The coexistence of anxiety symptoms and depression in major depressive disorder is substantial; anxiety symptoms, such as anxious mood and irritability, are seen in the majority of depressed patients. In addition, one-half to two-thirds of patients with panic disorder experience a major depressive episode during their lifetime. Medication and drug effects—from acute use, side effects, or as part of withdrawal phenomena—are also common causes of anxiety. Many patients with severe anxiety become dependent on anxiolytic drugs, including benzodiazepines, other sedatives, and alcohol, for symptom relief. During attempts to discontinue these substances, or sometimes during their ongoing use, confusing admixtures of anxiety symptoms, medication effects, and withdrawal symptoms may occur.

A 45-year-old female has been home bound since the start of the COVID pandemic. She is terrified that she will contract COVID-19, so she doesn't leave the house. If family visit, she requires that everyone wear gloves and face shields, and she will not shake hands with or hug her family members when they visit. She even cleans her plates, utensils, and glasses with sanitizer after they leave. She sprays the same alcohol-based sanitizer on her mail and only reads it 3 days after she receives it. When she receives grocery by home delivery, she washes every item with soap and water. If she receives any other packages by mail, she leaves them in her garage for 3 days before opening them.

Although all of the anxiety symptoms caused by drug use are also seen in primary anxiety disorders, perceptual disturbances, such as depersonalization and hyperesthesia, may be more common in sedative-hypnotic withdrawal syndromes than in primary anxiety disorders.

From a psychological point of view, anxiety may signal conflict between opposing desires, wishes, or beliefs on the one hand and major disequilibria generated by negative life events on the other hand. Role strains, conflicts among the major social roles that form a person's identity—spouse, parent, child, wage earner, professional, or community member—are common sources of anxiety. The more important the conflict and the less obvious the resolution, the greater the associated anxiety. For example, anxiety symptoms may first emerge when an individual is confronted with an unavoidable, unhappy choice, such as between sustaining a marriage or accepting a career advancement requiring a major move that is unacceptable to the spouse. At times, these conflicts may escape conscious awareness: someone may feel anxious but not know why. Anxiety syndromes frequently result from a combination of several factors. A person in a work conflict facing an important deadline may try to alleviate initial anxiety symptoms by overwork, then ingest caffeine or amphetamines to keep alert, then become exhausted and fatigued, and ultimately use alcohol excessively to calm down, with each of these elements contributing separately to an anxiety state.

Certain developmental life situations are associated with anxiety. Stranger anxiety develops when infants 6 to 8 months of age begin to recognize the difference between mother and others. When children first go to school, mild anxiety symptoms are common; if the anxiety is excessive, then separation anxiety or school phobia may result. During adult life, anxiety often centers around issues of mastery and accomplishment, in both personal and work life. Performance anxiety, or stage fright, is a specific type of pathologic anxiety in which anxiety escalates to panic when public performance is required. In later life, the deterioration of one’s body may engender anxiety related to feelings of helplessness and death anxiety.

Panic attack is a circumscribed episode of severe state anxiety lasting minutes to hours, with symptoms escalating in a crescendo pattern. The subjective experience is one of utter terror, with fear that one will die, go crazy, or lose control, accompanied by many of the somatic symptoms of anxiety mentioned above, including severe chest pains, marked shortness of breath, and exhausting fatigue. Individual isolated panic attacks are common, with up to 30% of the general population experiencing at least one attack each year. Panic attacks occur more regularly, and typically more severely, as part of panic disorder or in association with other anxiety disorders. In classic panic disorder, the attacks are spontaneous; that is, they are not triggered by a specific, predictable environment. In contrast, panic attacks associated with other anxiety disorders are triggered by specific situations such as in social anxiety disorder by social situations or in specific phobia by confronting the feared object or situation. Patients with other psychiatric disorders may experience limited-symptom panic attacks, with episodes characterized by less intense anxiety and by fewer and milder physical symptoms, such as isolated paresthesias or difficulty breathing. These limited-symptom attacks may represent aborted full-blown panic episodes that are not further exacerbated by secondary psychological reactions to the initial symptoms.

> A 32-year-old woman who works as a teacher in an elementary school is having episodes of shortness of breath associated with increased anxiety. She feels as if she is going to die and has to leave her classroom several times during work for that reason. Her primary care provider completed an extensive medical work up that was negative. She began to self-medicate with cannabis, which initially helped to decrease the frequency and intensity of her episodes, but after a few months, they began to recur. Eventually, her concerned husband brought her in for a psychiatric evaluation.

Although American psychiatry has segregated panic attacks from other forms of anxiety, assuming categorical, phenomenologic, and biologic differences, these distinctions are far from universally accepted. Much of European psychiatry views panic as simply an extreme form of anxiety to be understood as part of a continuum of intensity. DSM-5 made it simpler by deleting the association of agoraphobia from panic, and making agoraphobia a separate diagnosis.

Phobias are irrational fears. In an effort to reduce the intense anxiety attached to phobic objects and situations, patients do their best to avoid the feared stimuli. Thus, phobias consist of both the fears and the avoidance components. The fear itself may include all of the symptoms of extreme anxiety, up to and including panic. In specific phobias, persistent, irrational fears are provoked by specific stimuli. Table 8–3 lists some illustrative phobias. Common specific phobias include fear of dirt, excreta, snakes, spiders, heights, and blood (also termed blood-injury phobias).

Behavioral, psychodynamic, and biologic theories have all been advanced as causes of phobias. Some well-known phobias, such as fear of animals, may result either from early traumatic events (developing along the paradigm of classic Pavlovian conditioning) or from displacements of early psychodynamic conflicts. Genetic influences may also play a role in the development of phobias. For example, some individuals with blood-injury phobias, which strongly cluster among biologic relatives, may be genetically predisposed by vagal responses to certain stimuli. Animal models also indicate possible biologic vulnerability. Some monkeys that have never previously been exposed to snakes become panicky when placed in the presence of a snake. Because such fear responses obviously have adaptive value, it has been suggested that some human phobic responses also represent exaggerations of adaptive behaviors shaped by evolutionary biology. Complex phobias, more elaborate than specific phobias, involve fears related to a broader range of situations. Agoraphobia, the best known, means fear of the marketplace, symbolizing a fear of open spaces. Agoraphobia can occur in a variety of settings, however, and although it was previously associated with panic disorder, it can occur even in the absence of a panic disorder diagnosis.

Table 8–3.
Specific Phobias

Phobia	Definition
Acrophobia	Fear of heights
Agoraphobia	Fear of open spaces
Alektorophobia	Fear of chickens
Amathophobia	Fear of dust
Amaxophobia	Fear of riding in a car
Apiphobia	Fear of bees
Arachibutyrophobia	Fear of getting peanut butter stuck on the roof of the mouth
Astrapophobia	Fear of lightning
Aviophobia	Fear of flying
Blennophobia	Fear of slime
Claustrophobia	Fear of enclosed spaces
Coronaphobia	Fear of coronavirus
Cynophobia	Fear of dogs
Decidophobia	Fear of making decisions
Didaskaleinophobia	Fear of going to school
Electrophobia	Fear of electricity
Ephebiphobia	Fear of teenagers
Eremophobia	Fear of being alone
Gamophobia	Fear of marriage
Gatophobia	Fear of cats
Gephyrophobia	Fear of crossing bridges
Gynophobia	Fear of women
Hydrophobia	Fear of water
Kakorrhaphiophobia	Fear of failure
Katagelophobia	Fear or ridicule

Keraunophobia	Fear of thunder
Musophobia	Fear of mice
Nyctophobia	Fear of night
Ochlophobia	Fear of crowds
Odynophobia	Fear of pain
Ophidiophobia	Fear of snakes
Pnigerophobia	Fear of smothering
Pyrophobia	Fear of fire
Scholionophobia	Fear of school
Sciophobia	Fear of shadows
Spheksophobia	Fear of wasps
Technophobia	Fear of technology
Thalassophobia	Fear of the ocean
Triskaidekaphobia	Fear of the number 13
Tropophobia	Fear of moving or making changes

In social anxiety disorder (social phobia), patients become overwhelmingly anxious and fear situations in which they may be observed. In the limited type, only a few specific situations evoke the fear, such as speaking in public or using a public lavatory. In the general type, broad-based fears of social situations globally hamper the person’s interpersonal life.

SOMATIC DISTURBANCES IN PSYCHIATRIC ILLNESS

Nearly all psychiatric disorders manifest some change in basic physiologic function. These alterations in somatic function may be quite nonspecific, but the severity of the disruption in normal function may provide clues to the amount of physiologic malfunction seen in the primary underlying psychiatric illness. Although frequently nonspecific in nature, the severity of these somatic signs and symptoms provides markers as to the amount of biologic disruption seen in the disorders that cause them. Furthermore, somatic symptoms can also cause exacerbations of some disorders. If untreated, then these processes can create destructive feedback loops in which the disorder causes symptoms, which then exacerbates the disorder, which causes increases in the symptoms, and so on. As examples, the insomnia of manic or hypomanic states, if untreated, causes a marked worsening of the mania; similarly, the weight loss of anorexia nervosa causes starvation effects, such as a preoccupation with food, thus exacerbating one of the hallmark features of the underlying disorder.

Sleep Disturbances

Considering the fact that about one-third of each day’s activities are spent sleeping, it is not surprising that psychiatric disturbances are frequently manifested by an alteration in normal sleep. These alterations may be in the timing, the amount, and the quality of sleep as well as the presence of abnormal events occurring while sleeping. In addition, the alterations may be associated with the primary psychiatric disturbance, medications used in psychiatric treatment, or the use of other prescribed agents or substance use.

Insomnia is usually defined by its subjective component as the sensation of not sleeping well or enough. Most, but not all, patients complaining of insomnia demonstrate some sleep abnormality if examined in a sleep laboratory. Insomnia is a common, often chronic, symptom or sign of many different psychiatric disorders, including substance use disorders, depression, generalized anxiety disorder, panic, mania (in which the diminished sleep does not always provoke a complaint), and acute schizophrenia. It may also occur as a consequence of aging or as a symptom or disorder not associated with other psychopathology. Insomnia may also result from the ingestion of substances that alter the normal sleep–wake cycle, including alcohol or stimulants, and by the discontinuation of sedative-hypnotics. Although much attention is often paid to distinguishing patterns of insomnia, such as difficulty falling asleep versus middle or terminal insomnia (early morning awakening), or linking specific patterns to a specific disorder (e.g., melancholic depression with terminal insomnia), the clinical usefulness of these distinctions is sometimes unclear.

A 32-year-old computer engineer described herself as a “worry wart.” She reported insomnia “my whole life,” including initial insomnia and multiple awakenings throughout the night. She reported her mind could not relax as she tried to sleep and found herself worrying about “everything that happened during the day.” Inevitably, she would wake up in the middle of the night to “make a list of things I forgot, or review possible mistakes I made at work.”

Hypersomnia, characterized by either excessive nighttime sleep or excessive sleepiness during the day, is less common than insomnia. It, too, however, may reflect a number of different pathologic states. Some depressed patients, especially those with a history of mania or hypomania, may exhibit hypersomnia. Hypersomnia may also be seen during stimulant withdrawal, with excessive use of sedatives or tranquilizers, or in conjunction with a variety of medical disorders. In narcolepsy, the patient has sudden attacks of irresistible sleepiness, a symptom that may be part of a broader syndrome that includes cataplexy (sudden attacks of generalized muscle weakness leading to physical collapse in the presence of alert consciousness), sleep paralysis (waking from sleep with a sensation of being totally paralyzed that may persist for minutes), and hypnagogic hallucinations (vivid visual hallucinations that occur at the point of falling asleep). Narcoleptic attacks are often precipitated by unusual states of arousal (e.g., cataplexy may immediately follow unrestrained laughter or orgasm). Daytime sleepiness may reflect sleep apnea. In this disorder, typically middle-aged patients demonstrate severe snoring—often first reported by their bed partners—and periods when breathing stops. The condition results from soft palate abnormalities that cause intermittent airway obstruction throughout the night; patients awake repeatedly to find themselves gasping for air. Associated daytime fatigue is common in sleep apnea. Periodic hypersomnia also occurs in Kleine–Levin syndrome, a condition typically affecting young men in which periods of sleepiness alternative with confusional states, ravenous hunger, and protracted sexual activity. Intervals of days, weeks, or months may pass between these episodes.

Somnambulism, or sleepwalking, and sleep terror disorder (night terror) are two sleep disorders characterized, respectively, by aimless wandering with incomplete arousal and by acute anxiety and physiologic arousal without awakening. Although both disorders typically begin in childhood, sleepwalking may be also initially precipitated by some psychotropic medications.

A 54-year-old man with complex PTSD was treated with a sedative hypnotic by his primary care physician. During this period, his wife noted that he often went to the refrigerator in the middle of the night while sound asleep and ate indiscriminately. When she woke him during these episodes, he appeared dazed and had no awareness of his purposive behaviors. To the best of his recollection, confirmed by his wife, he had no history of somnambulistic phenomena before treatment with the sedative hypnotic agent.

Nightmares are a common complaint, often associated with traumatic events, anxiety disorders, and mood disorders, but not uncommon as an occasional event in otherwise healthy individuals. Vivid dreams and nightmares may also be a medication side effect.

A 32-year-old woman was treated for a mild depression with an SSRI. Her sleep pattern before treatment had been characterized by hypersomnia, sleeping between 10 and 12 hours nightly. Soon after the initiation of the antidepressant, she noticed that her sleep was not only shorter and more fragmented, but that she had remarkably intense dreams that she remembered vividly. She made clear that these were not nightmares, just dreams with intense colors and clarity, which she rather enjoyed.

Sensory symptoms during sleep, typically described by patients as peculiar feelings in their legs causing an irresistible need to move around, are characteristic of restless legs syndrome. The motor abnormality of repetitive myoclonic jerking of the legs, awakening both patients and their partners, is known as nocturnal myoclonus.

Appetite and Weight Disturbances

A significant change in appetite and weight can be seen in both medical and psychiatric illness. Anorexia may occur in the latter stages of chronic medical illness and is frequently seen in depression, grief, and some anxiety disorders. In addition, anorexia may be secondary to alterations in taste sensation, as a function of psychiatric disturbance or medication side effect. In eating disorders, such as anorexia nervosa, patients may resist hunger to restrict food intake to achieve a physiologically unrealistic low weight. Hyperphagia (increased appetite) occurs in some depressed patients, both with and without a history of mania or hypomania. Binge eating, of up to several thousand calories per episode, may occur as an attempt to self-soothe and emotionally self-regulate during times of increased tension and anxiety and as a key feature of bulimia nervosa or of binge-eating disorder. Increased appetite may be seen as a side effect of many psychotropics or, albeit rarely, in some hypothalamic disorders or in bilateral temporal lobe dysfunction such as Klüver–Bucy syndrome, in which it occurs in association with emotional placidity, hypersexuality, hyperorality, and other related symptoms.

A 45-year-old man with schizoaffective disorder had been intermittently erratic with his follow-up visits and had requested several telephone refills over the last several months. He was instructed by his psychiatrist to come in to see him before additional refills could be called in to his pharmacy. Subsequently, he presented for an office visit with his psychiatrist. In the interview it became apparent that the patient had gained a significant amount of weight, approximately 40 lb, in the last several months. He admitted to having a “hunger” for sweets and fatty foods and acknowledged that he had a marked increase in his appetite after he was switched to an atypical antipsychotic.

Energy Disturbances

Normal energy levels vary considerably among people. Some people fatigue easily and are perceived by themselves and others as having “weak constitutions,” whereas others appear to have almost boundless energy and much less need for sleep.

Fatigue is a common nonspecific symptom that occurs in both medical and psychiatric disorders. It is also frequently seen as an unexplained complaint in primary care practices; in one study, 24% of patients complaining of fatigue received no medical or psychiatric diagnosis. Historically, fatigue not due to another disorder, typically in association with “nervousness,” has been described by terms such as asthenia, neurocirculatory asthenia, neurasthenia, and psychasthenia. Consistent with this tradition, many fatigued patients, having been labeled depressed or neurotic by their physicians, are referred to psychiatrists after routine workup has ruled out anemia, hypothyroidism, sleep apnea, and other frequent somatic causes.

Recently, those patients with primary complaints of tiredness have been most commonly diagnosed as having chronic fatigue syndrome

(CFS), a disorder characterized by fatigue lasting months to years, typically beginning soon after a viral syndrome. In addition to the fatigue, CFS is characterized by myalgias and cognitive changes such as forgetfulness and poor concentration. Although controversy still exists over the extent to which cases of CFS represent discrete post-viral diagnostic syndromes, mislabeled cases of depression, or modern versions of psychasthenia, evidence continues to mount to suggest that these syndromes are discrete postinfectious entities and are not simply variants of or disguised mood or anxiety disorders.

A highly accomplished, energetic 40-year-old woman was referred for psychiatric consultation after her physicians were unable to offer a definitive physiologic diagnosis despite extensive medical workup after the acute onset of profound fatigue occurring in the wake of a mild viral illness. This fatigue caused her to be totally incapacitated and bedridden for many months and left her feeling quite helpless and distraught. It was exacerbated by even small amounts of alcohol. Two independent psychiatric evaluations concluded that the patient had no psychiatric disorder. The patient developed this syndrome after jogging near Lake Tahoe. Some epidemiologic studies found a series of other cases of profound fatigue originating in this particular geographical area after vigorous exercise. Various antidepressants seemed to provide partial relief, but tolerance to their positive effects developed. The fatigue slowly improved over many years; 7 years after the original symptoms, the patient continued to experience waxing and waning fatigue but at a lesser degree than originally.

Disturbances in Sexual Drive

As with energy, variation in the normal range of sexual drives is common. Some individuals are naturally lusty, whereas others have limited sexual desire. Diminished sexual drive with impotence or decreased libido is seen in a wide variety of neurologic, metabolic, and other somatic conditions. Among neurologic disorders, complex partial seizures are commonly associated with hyposexuality, occurring in 50% of patients. Psychiatric disorders known for diminished sexual drive include depressive disorders, schizophrenia, and substance use disorders. Diminished libido, erectile dysfunction, and anorgasmia are also common sequelae of many psychotropic agents, especially those with a strong serotonergic action.

A depressed 36-year-old man had an excellent antidepressant response to an SSRI but developed decreased sexual interest and anorgasmia with relative preservation of erectile function. When the sexual side effects did not diminish after 4 weeks, his psychiatrist tried a number of potential remedies. Although sildenafil improved his erections somewhat, his libido and orgasmic function continued to be markedly impaired. He and his psychiatrist agreed to switch to a different antidepressant class that was not associated with sexual side effects.

Increased sexual activity may be seen in some neurologic, psychiatric, and drug-induced disorders. Manic patients frequently exhibit hypersexual interests and behaviors to an unusual degree, compared with their euthymic interests and behaviors. Hypersexuality is occasionally seen in conjunction with epileptic syndromes or in patients who have had diencephalic injuries.

Altered sexuality, including fetishism, sadomasochism, pedophilia, and other paraphilias, may be seen as isolated psychiatric syndromes. In individuals whose previous sexual behaviors were within the bounds of social propriety for their groups, inappropriate sexual behaviors may signal early brain disease or psychosis.

Appearance

Studies show that clinicians often formulate an initial psychiatric diagnosis within 30 seconds of seeing a patient. Although approximately one-half of such initial impressions prove to be incorrect, the remainder, however, are validated by psychiatric history and mental status examination, revealing just how much information is communicated by appearance and body language.

Among the physical disorders whose appearances suggest coexistent psychiatric conditions are acromegaly, Cushing syndrome, Down syndrome, systemic lupus erythematosus, fetal alcohol syndrome, Klinefelter syndrome, and Wilson disease, to name a few. The general appearance of the skin may suggest the presence of occult psychiatric problems. The general condition and flush of the skin may reveal hypervascularity and ruddiness suggestive of alcoholism, abscesses indicative of hypodermic needle abuse, weathering and wasting indicative of self-neglect and malnutrition, or even certain tattoos or patterns of tattoos that may have psychological significance to the patient. Healed scars on the wrists and arms may suggest a pattern of self-mutilation. Patchy baldness, especially in conjunction with torn or infected cuticles, indicates trichotillomania, a syndrome of compulsive hair pulling. Psychophysiological symptoms reflecting psychiatric disturbance include urticarial reactions and neurodermatitis, the latter resulting, in part, from self-excoriation, destructive scratching secondary to compulsions, and unrelenting sensations of discomfort.

A 23-year-old woman who had recovered from a severe episode of anorexia nervosa as an adolescent had recurrent depression and persistent personality disturbances characterized by odd and eccentric ideas, shyness, and avoidance. She became enchanted with gothic themes, spent hours on the Internet in chat rooms featuring dark and morbid poetry, and started to have herself tattooed. Within 1 year, she had tattooed every inch of her face, eyelids, lips, earlobes, upper extremities, scalp, and torso with complex, intricate, gothic designs. She acknowledged that tattooing her skin in such an extreme manner served to mark her as deviant and to warn people away from her.

Examination of the head and neck may reveal exophthalmos or puffy eyelids, suggesting thyroid disease, marked pupillary dilation with anxiety or stimulant abuse, miosis with narcotic use, abnormal pupillary pigments in Wilson disease, salivary gland enlargement in bulimia nervosa, or necrosis of the nasal septum in cocaine use disorders, among other signs. Frequent sighing is a common respiratory sign in depression. Simple sighing must be distinguished from respiratory dyskinesia in psychotic patients who have been treated with neuroleptics. The latter may occur as an acute dyskinesia due to antipsychotic medication, or it may be a late manifestation and component of tardive dyskinesia.

Deviant appearance is quickly perceived by laypeople and professionals and may contribute to the frightened and stigmatizing social withdrawal by strangers, as well as acquaintances, so often experienced by psychiatric patients. Akathisia, dystonic movements, and parkinsonian shuffling gait in patients taking neuroleptics (especially the older, conventional antipsychotics), as well as the dilapidated and unkempt appearance of some psychiatric patients, can immediately signal psychiatric patient status to observers. The term *Diogenes syndrome* has been used to describe old people who have filthy personal appearance and demonstrate severe self-neglect about which they have no shame.

DISTURBANCES OF BEHAVIOR

Aggression, Hostility, Impulsiveness, and Violence

The spectrum of aggressive emotions and behaviors is characterized by heightened vigilance in response to a sense of threat and enhanced readiness to attack. Physiologic tone may be geared for a fight. Assertiveness, the adaptive aspect of these emotions, includes sensing that something needs to be done and feeling willing and competent to take constructive action. The manner and extent to which aggressive emotions can be expressed varies from society to society and from situation to situation. These emotions are among the most carefully regulated because of their potential destructiveness. Acts of aggression are on a continuum, beginning with irritability, progressing to verbal threats and intimidation, and extending to physical bullying and assault to homicide, sometimes including acts of calculated violence and sadism.

Irritability is an unpleasant feeling state characterized by inner unease. In contrast to anger, irritability does not lessen after an outburst. Often, others are more aware of an individual's irritability than is the person themself. It is diagnostically nonspecific, seen in a variety of physiologic states—psychotic, anxiety, and mood disorders—and as a lifelong temperamental quality. Hunger, sleepiness, sexual frustration, and pain are among the physiologic triggers commonly associated with irritability.

Individual differences in the tendency toward experiencing and expressing anger and violence are biologic, developmental, and cultural in origin. Some infants are irritable from birth. Subtle early birth injuries and brain anoxia may increase the susceptibility of some people to be violent. Furthermore, studies of EEG patterns in violent people show increased abnormalities, especially in those with repeated violence and violence with little or no obvious motive. Soft neurologic signs are also seen in violent criminals. Biochemically, low cerebrospinal fluid (CSF) 5-HIAA has been associated with a variety of impulsive behaviors, such as violent crimes, recurrent fire setting, and violent suicide attempts.

Consistent with the hypothesis that an inverse relationship exists between central serotonergic system function and impulsive, aggressive behavior, a few double-blind, placebo-controlled studies in patients with cluster B personality disorders have demonstrated that enhancing central serotonergic function by using SSRIs reduces irritability and impulsive aggressive behavior.

A 32-year-old man came for psychiatric treatment after a breakup with his fiancée of 5 years related to his anger outbursts, in which he would curse, yell, slap, and throw things at her. Once, during a rage attack, he slashed the tires of his fiancée's car. Another time, he impulsively took an overdose of pills that he found at home. He responded to treatment with an SSRI, which improved his mood lability and anger outbursts.

The pathologic childhood triad of bed-wetting past the age of 6 years, setting fires, and torturing animals has been associated with subsequent violent behavior in adults. Interpersonally, studies show that violence-prone individuals require more personal space around their physical person than do others. Violent individuals feel threatened when approached too closely, particularly from the rear.

Psychological and social contributions are also strong. Violence in families breeds violence, and battered children often grow up to be battering adults. Cultural norms for the expression of violence differ considerably. In some socioeconomic and ethnic groups, violent gangs organize the energies of many adolescent youth. For some, violent behavior is an adolescent socialization pattern necessary to prove one's manhood or womanhood. Like other social organizations, violent gangs have detailed rules that inhibit and govern the expression of violence. Some unpredictable and unsocialized violent people are too violent to be contained even in gangs.

Aggressive and violent behavior is diagnostically nonspecific. Violence in schizophrenia may occur as a consequence of paranoid delusions, in response to command auditory hallucinations, or secondary to passivity experiences. Manic patients and those in mixed states may be violent, often in response to minimal provocation. Violent behavior commonly occurs in patients with antisocial and borderline personalities (in the latter often self-directed as well as other-directed). Violent behavior may occur in epilepsy, although rarely during true ictal periods; in frontal lobe syndromes as a “release phenomenon;” and in association with abused substances, particularly disinhibiting sedatives, such as alcohol, or stimulants, such as amphetamines and cocaine, which increase irritability, aggressiveness, and paranoia.

Impulsive violence may be provoked by a number of stimuli and situations. Alcohol is perhaps the most common disinhibitor of violence. Intrafamilial violence, the most common setting for homicide, is frequently related to alcohol intoxication. In episodic dyscontrol and intermittent explosive disorder, violent behaviors typically erupt after a person has ingested alcohol, a phenomenon known as pathologic intoxication. In these often ferocious outbursts, the individual may confront or provoke any potential target for violence, including total strangers and police, but significant others, spouses, and parents are frequent victims. Patients with episodic dyscontrol commonly have histories of violent sexual behavior, including rape and, often while intoxicated, speeding and reckless driving, sometimes chasing down, stopping, and attacking other motorists who they believe “get in their way.”

An attractive 40-year-old woman had been highly sought after by rich and powerful men, several of whom she had married and several of whom divorced her because of her uncontrollable rages. She described herself as becoming “insanely jealous” in a flash whenever she saw her husband glancing at other attractive women, and her behavior flared out of control, particularly when she had been drinking alcohol. On one occasion, she bit off one husband’s ear. She attacked another with pots and knives, nearly killing him. She recognized how serious her problems were and sought psychotherapy, medication (SSRIs were helpful), and anger management training to help her control her emotions and behaviors.

Temper Tantrums

Immature individuals with persistent personality problems may not develop mechanisms to inhibit temper tantrums that they displayed as children. Particularly if childhood tantrums produced the desired result, then learned tantrum behaviors may persist into adult life. Although such individuals may be pleasant and sociable when life is going well, they lack the capacity to tolerate frustration and are easily provoked by threats to self-esteem and self-image and by not having their own way. When frustrated or threatened, they may act like bullies, glare, snarl, yell, shout, intimidate, pout and sulk, and sometimes be physically violent.

Displaced Rage

When circumstances prevent the expression of rage directly against those people or institutions provoking frustration, other outlets for aggression are often found. Acts of violence that are either calculated or wanton may result. Cruelty to animals and fire setting may persist as adult forms of destructive behavior. Rape, an act of control, intimidation, terror, and humiliation, may also displace frustrations that are not expressed more adaptively.

Sadism may occur with or without explicit sexual gratification. Calculated cruelty conducted seemingly without anger or emotional arousal may reflect inadequate development of social morality or individual conscience, as in the conduct of torturers and some cold-blooded murderers. In some societies and under specific circumstances at certain times in history, such activity has been socially sanctioned, suggesting at least that some people lack inborn inhibitions against cruelty or violence.

A 50-year-old man with antisocial personality disorder was seen in a psychiatric emergency room for violent behavior. During the course of the assessment, he described how he had been emotionally and physically abused by a sadistic father who whipped him repeatedly until he bled and who chained him to a tree in the backyard. This man grew up feeling callous. He enjoyed intimidating and physically hurting others, as he often did in prison. With a laugh and smirk, he reflected on how he turned out to be "just like my old man."

Self-Mutilation

For a variety of reasons, in many different cultures and in many different disorders, people commit acts of violence against themselves, ranging from body piercing to cutting and burning to autoamputation. Psychotic patients may perform extremely self-destructive acts short of actual suicide that often have symbolic import, such as enucleating their eyes or castrating themselves. Patients with borderline personality disorders or borderline traits may cut themselves repeatedly with broken glass or razor blades or burn themselves with cigarettes on arms, legs, breasts, or other body parts. Patients typically deny that these acts are meant to be suicidal but describe the need to feel external pain to mirror internal suffering, to release tension, or to counteract dissociative-like numbness. Recently, self-cutting has become more common in teenagers, especially girls, as a tension-relieving mechanism.

A 20-year-old college student was taken to the emergency room by her friends, who were alarmed by the recent transverse cuts on her arms. None of the cuts were deep enough to require sutures, but, even when they had healed, they left obvious small red scars on her skin. In explaining her behavior, the student described that when she was upset, such as after a fight with her boyfriend or a negative telephone call with her parents, she became agitated, and cutting herself consistently relieved the agitation and anxious feelings.

Trichotillomania is a syndrome of compulsive hair pulling, resulting in bald patches. It is often associated with other self-mutilatory behavior, such as picking the face, nails, or cuticles to the point of infection and bleeding. Trichotillomania may sometimes be related to OCD.

Children with Lesch–Nyhan syndrome, a developmental disability syndrome caused by a congenital metabolic abnormality, bite and pick at themselves so compulsively as to do themselves great harm and routinely require restraint. Occasionally, patients with Tourette syndrome demonstrate compulsive self-harming behavior.

Other Disturbances of Feelings

Diminished levels of emotional intensity may be seen in anxiety disorders, mood disorders, and schizophrenia. Mild emotional flattening with blunted ability to feel joy is common in dysthymia. Some patients with narcissistic and borderline personality disorders complain of inner emptiness and pervasive boredom and ennui without demonstrating diminished affect in interviews. Similarly, patients with prominent depersonalization describe numbed emotions. Pathologic levels of blunt or flattened affect, indicating markedly diminished affective expression in relation to specific thought content, may be seen in chronic schizophrenia (as part of the deficit syndrome), some organic mental syndromes, and severe depressions and as a side effect to many medications, including

SSRI antidepressants, antipsychotic medications, or excessive tranquilizer use. Although the term blunted affect is not classically used to describe the affective flatness of severe depression, it is not always easy to distinguish between schizophrenic and depressive flatness on phenomenologic grounds. Anhedonia, the lack of pleasurable feelings from activities that ordinarily provide pleasure, is also seen as part of severe depressions or schizophrenia. Chronically psychotic patients often exhibit emotional deterioration in which affective experience and expression are entirely unrelated to thought content. Inappropriate affect is the incongruency of affective expression and thought content. The patient may display loud and raucous laughter or giggling in relation to bland or sad thoughts or may show grief without apparent reason. Inappropriate affect sometimes indicates that the thoughts have private meanings for the patients; the emotional expression might make better sense if the private meanings were understood. Inappropriate affect must be distinguished from affective expressions that may actually be appropriate in a given subculture or ethnic group that is unfamiliar to the observer and from defensive affect, such as the nervous laughter used to alleviate tension or ward off crying. Affective (or mood) lability is characterized by rapid emotional shifts, often within seconds to minutes. It is commonly seen during hypomanic states, late luteal phase dysphoric disorder (premenstrual syndrome), postpartum blues, other states of physiologic instability, and in cluster B personality disorders.

A 51-year-old man with lifelong dysthymic disorder was treated with an SSRI. Although he had an excellent antidepressant response, 2 months after beginning treatment, he noted an affective flatness. He described not becoming as easily upset as before treatment (which both he and his wife interpreted positively), but he also believed his sense of joy, in being with his children or in nature, was markedly blunted. He enjoyed positive events but with a marked diminution of affective intensity.

Alexithymia is difficulty identifying, describing, and differentiating feelings or distinguishing between feelings and physical sensations. Alexithymic individuals often have constricted imaginations and fantasies, are preoccupied with objects and events in the outside world, and have little private, personal internal life. When distressed, the patients are simply aware of not feeling well and usually complain of somatic symptoms, leading to frustrating interactions with their physicians who cannot find physical causes for the presenting physical complaints. Some view alexithymia as a condition in which affect is communicated through somatic language.

With regard to other feeling states that may contribute to pathologic conditions, recent writers have focused on the role of "shame" as a potentially important emotional mediator of specific types of narcissistic injury that may provoke states of depression, rage, PTSD, and suicide, among others.

DISTURBANCES OF INTERPERSONAL RELATIONSHIPS

Normal interpersonal relationships include relationships with parents, children, spouses, siblings, extended family members, friends, colleagues, coworkers, and members of the larger community. These relationships ordinarily help provide for the satisfaction of basic drives, for affiliative needs, and for finding purpose and meaning in life. Through stable and satisfying relationships, human needs are met for intimacy, including love, sex, and affection, to be cared for and nurtured, to provide care, learn, play, relax, dominate, and be productive through mutual effort. Interpersonal relationships are carefully regulated by means of interpersonal signs and signals. The extent to which deviance from these patterns is tolerated in a given relationship varies from behavior to behavior, relationship to relationship, family to family, and culture to culture.

Disturbances in interpersonal relationships may be viewed as characteristics attributable to a single person or as characteristics of an interpersonal system. Individual disturbances are considered to be undesirable or maladaptive personality traits. When these traits are present to a significant extent and interfere with social functioning or cause distress, they may compose a personality disorder. Disturbances of interpersonal relationships have also been described at a system level (e.g., as dyadic and family patterns of system disturbance).

Personality Traits and Disorders

Personality, variably defined, is the characteristic pattern of an individual's attitudes, behaviors, beliefs, feelings, thoughts, and values—the sum of a person's emotional, cognitive, and interpersonal attributes. Personality traits are the prominent and characteristic features of an individual's personality and do not imply psychopathology. Aspects of personality are present from early life, and personality traits are relatively stable from adolescence onward, consistent across different environments, and recognizable by friends and acquaintances. The term personality disorder should be reserved for those consistent patterns of thought, feeling, and behavior that are maladaptive. Personality disturbances manifest primarily in interpersonal contexts and in this way may be viewed as interpersonal behavior disorders.

The determinants of personality are multiple and varied—innate and early biologic, developmental, and environmental factors inside and outside the home. Through learning and the environment, temperamental factors (genetic or constitutional) are shaped into character.

The dimensional approach to personality and personality pathology characterizes individuals along a continuum of traits. Five dimensions of temperament have been described that appear to be somewhat independent and to have strong genetic contributions: neuroticism (highly emotional, reactive, and thin skinned, contrasting with emotional stability), extroversion (contrasting with introversion), openness (contrasting with discomfort with novel experiences), agreeableness (contrasting with contrariness), and conscientiousness (contrasting with fickleness). These temperamental attributes may have implications for the course of psychotherapies that cut across diagnostic categories.

Another dimension of personality not adequately dealt with in the DSM-IV-TR and now DSM-5 concerns moral behaviors such as honesty and integrity. The extent to which individuals behave honestly and with integrity differs considerably across individuals and in different situations. Deception and lying are common behaviors that occur in benign forms (e.g., in “white lies”) and in pathologic forms, psychiatrically important in antisocial and sociopathic disorders, pathologic liars, and malingerers. Deception and lying may be difficult to assess clinically in the absence of additional informants. Studies of nonhuman primates indicate that, at least among chimpanzees, deception (equivalent to lying and dishonesty) is relatively common and, in some situations, adaptive.

Another personality typology characterizes personality along three dimensions related to temperamental characteristics presumed to be strongly influenced genetically: harm avoidance, novelty seeking, and reward dependence. High scores on the three dimensions characterize inhibition and pessimism, impulsive and exploratory behavior, and dependency and sentimentality, respectively. Different personality types can be described according to patterns of scores on the three dimensions. For example, antisocial personalities are characterized by high novelty seeking, low harm avoidance, and low reward dependence, whereas dependent personalities have low novelty seeking, high harm avoidance, and high reward dependence.

The DSM-IV-TR, by contrast, uses a categorical approach, yet both the large overlap among the DSM-IV-TR personality disorders and the clustering of these personality disorders into three broad groups imply a lack of clear boundaries to the currently defined categories. The three DSM-IV-TR clusters describe odd or eccentric types (cluster A); dramatic, emotional, and erratic types (cluster B); and anxious and fearful types (cluster C). While the DSM-5 task force recommended an alternative model and approach to personality disorders, the APA Board of Trustees decided to keep current nomenclature to preserve continuity. The alternative approach is included as Section III in DSM-5. In this model a greater emphasis on personality functioning and trait-based criteria increases the stability and empirical basis of the disorders.

The odd or eccentric group includes paranoid, schizoid, and schizotypal personality disorders. Patients with these personality disorders have the core traits of being interpersonally distant and emotionally constricted. Paranoid personalities are quick to feel slighted and jealous, carry grudges, and expect to be exploited and harmed by others. Schizoid personalities lack friendships or close relationships with others and are indifferent to praise or criticism by others. Schizotypal personalities display odd beliefs, engage in odd and eccentric gestures and practices, and exhibit odd speech.

The dramatic, emotional, and erratic group includes borderline, histrionic, narcissistic, and antisocial personality disorders. Patients with these personality disorders characteristically have chaotic lives, emotions, and relationships. Borderline personalities are impulsive, unpredictable, angry, temperamental, unstable in relationships, compulsively interpersonal, and self-damaging with regard to sex, money, and substance use. Histrionic personalities are attention seeking, exhibitionistic, seductive, self-indulgent, exhibit exaggerated expressions of emotions, and overconcerned with physical appearance. Narcissistic personalities tend to be hypersensitive to criticism, exploitative of others, and egocentric with an inflated sense of self-importance; they feel entitled to special treatment and demand constant attention. Antisocial personalities are described almost exclusively by behavioral rather than affective or relational terms. They are truant, lie, steal, start fights, break rules, unable to sustain work or school, and shirk everyday responsibilities.

The anxious and fearful group includes avoidant, dependent, and obsessive-compulsive personality disorders. Patients with these disorders are characterized by constricting behaviors that serve to limit risks. As examples, avoidant people avoid relationships, dependent personalities avoid being responsible for decisions, and obsessive-compulsive people use rigid rules that preclude new behaviors. Avoidant personalities are hypersensitive to rejection and are reluctant to enter close relationships in spite of strong desires for affection. Dependent personalities show excessive reliance on others to make major life decisions, stay trapped in abusive relationships for fear of being alone, have difficulty initiating projects on their own, and constantly seek reassurance and praise. Obsessive-compulsive personalities exhibit restricted expressions of warmth, tenderness, and generosity, and exhibit stubbornness with a need to be right and to control decisions; indecisive at times, they often use overly rigid application of rules and morals to the point of being inflexible.

A characteristic personality disturbance seen with frontal lobe damage is referred to as organic personality disorder in the tenth edition of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) and personality change due to a general medical condition in the DSM-IV-TR. In DSM-5 it is included as personality change due to another medical condition. Its features include irritability, inappropriate jocularity with euphoria, inappropriate socially disinhibited behavior, and impulsiveness. Other patients with damage to different areas of the frontal lobe, in contrast, exhibit apathy and indifference.

Interpersonal Systems

Couples and families have been studied as systems in their own right, and many qualities of these systems have been identified as being clinically important. A scheme for categorizing relational disorders has been proposed for future editions of the DSM, but, as yet, no single generally accepted typology of family psychopathology or interactional types has been established. However, elements of marital discord and harmony have been operationalized in several standard marital inventories. Characteristics of couples and families that have received the most attention include the rules of communication, such as those governing the directness or indirectness with which disagreement and conflict are addressed; the manner (organized or chaotic) in which communications are conducted; taboo topics and secrets about which no one can openly communicate; the nature and degree of emotional expression, including affection and anger; the cohesiveness, loyalty, and compatibility of members; the nature of the members' shared identities on the one hand and their autonomous development and separateness on the other; the extent to which members treat one another respectfully or take one another for granted and use one another; the distribution of power and decision making among members; the maintenance of generational boundaries (e.g., age-appropriate performance of life roles); and the members' orientation toward and concurrence and disagreement about important values involving moral, religious, intellectual, cultural, financial, occupational, and childrearing issues, as well as aspirations, health practices, leisure activities, and other belief systems.

A characteristic family environment, called high expressed emotion, has been identified that defines a relapse-prone family environment in which one individual has schizophrenia, bipolar disorder, anorexia nervosa, or major depressive disorder. This interactional pattern includes demeaning, intense personal criticism ("You are rotten and lazy") and emotional overinvolvement with the identified patient. Aspects of overinvolvement can be measured by quantifying the numbers of hours of face-to-face contact and by the extent to which relatives categorically assert how the patients feel without ever bothering to ask the patients. Despite these descriptive generalizations, the specific elements of high expressed emotion families that make patients vulnerable to relapse are still obscure.

Couple and family system difficulties are most likely to erupt during predictable stressful events in the normal family life cycle, such as during the newlywed period; pregnancy and childbearing; difficult or contentious childrearing; difficulties with parents, in-laws, and other extended family; insurmountable and unanticipated financial or career problems; serious illness or death of a child or relative; the children's adolescence; departure of children from the home; infidelity; and separation.

Interpersonal attachment styles, based on cognitive schemas that have been linked to earlier repeated experiences with caregivers, influence how individuals perceive and act within interpersonal relationships. Specific types that have been described include insecure attachments (several subtypes include preoccupied, fearful, and anxious-ambivalent attachment), avoidant attachment (including angry-dismissive and withdrawn attachment), enmeshed attachment, unresolved attachment (in which significant losses have not been dealt with), and secure attachment styles. Specific types and subtypes have now been linked to certain types of health behaviors and clinical outcomes. For example, diabetic patients with dismissive attachment styles, a style found in approximately 25% of the general population and characterized by low trust of others and excessive self-reliance, show a decreased ability to collaborate with providers and have poorer glycemic control than patients with secure attachment styles. In a study at a primary care clinic, patients with insecure attachment styles reported more unaccounted for medical symptoms than others. Among these, patients with preoccupied attachment had the highest primary care costs and usefulness, whereas patients with fearful attachment had the lowest, reflecting their respective tendencies to overuse or underuse medical services.

Interpersonal Disturbances in Illness Behavior

Abnormal illness behavior (dysnosognosia) is a persistently pathologic mode of experiencing, evaluating, and responding to one’s own health status despite lucid and accurate appraisal and management options provided by a health professional. These behaviors can be considered as interpersonal disorders between patients and health care professionals. Central to all of these behaviors is the adoption of the sick role by the patient, who then engages in characteristic interactions with health care providers, which typically leave both the provider and the patient dissatisfied. Patients with abnormal illness behavior typically seek repeated medical evaluations from a multitude of physicians, often undergoing a series of expensive laboratory tests. At times, the level of complaints provokes unnecessary invasive laboratory examinations or surgeries, which, in turn, thereby place the patient at genuine medical risk.

Abnormal illness behaviors may be unconscious or conscious. Unconscious abnormal illness behaviors are those in which the patient believes the symptoms reflect some genuine illness. These behaviors may occur in somatic symptom disorder, illness anxiety disorder, and conversion disorder (functional neurologic symptom disorder). Abnormal illness behaviors in which patients act sick when they are fully aware that they are not include malingering (in which external incentives—usually financial—are the motivating factors) and factitious disorder with physical or psychological symptoms (Munchausen syndrome). In Munchausen syndrome, patients repeatedly and compulsively present themselves for medical care with feigned or self-induced illness. These self-induced conditions may be so serious as to cause ultimately death: some patients inject themselves with feces to cause systemic infections that then warrant hospitalization and intensive care. When the self-induced nature of the illnesses is discovered, medical staff often become outraged at these patients. The patients rarely accept or cooperate with psychiatric care, so few have been adequately studied. Most do not appear to be psychotic, but there seems to be a disturbance in personality structure. In Munchausen syndrome by proxy, a caregiver, usually a parent, induces illness in a child.

DISTURBANCES IN MOTOR ACTIVITY

Motor behavior is normally finely coordinated, purposeful, and adaptive, and necessary activities are usually carried out efficiently. In psychiatric disturbances, motor abnormalities can involve generalized overactivity or underactivity or manifest in a wide range of specific disorders of movement. Many motor disturbances are seen in psychiatric disorders. Some form part of the core symptoms of the disorders; some occur in disorders that, by their nature, bridge neurology and psychiatry (such as Tourette syndrome); others are acute or chronic medication side effects.

Overactivity

Restlessness and agitation are diffuse increases in body movement, usually noted as fidgeting, rapid and rhythmic leg or hand tapping, and jerky start-and-stop movements of the entire body, accompanied by inner tension. Restlessness accompanies psychiatric conditions of high emotional arousal or confusion, such as toxic states, deliria, mania, agitated depressive disorders, and anxiety disorders, as well as many medical disorders such as hyperthyroidism. In some depressive states, agitation is often accompanied by pacing and hand wringing.

Generalized overactivity, in which patients seem to have increased physical energy, is distinguished from agitation by its lack of inner tension and by more purposeful movements. It is commonly seen in mania, hypomania, and anorexia nervosa and as part of ADHD and in response to stimulating drugs and medicines.

A 25-year-old extraordinarily restless man in a psychiatric emergency room, whose glances darted incessantly around the room, paced agitatedly, picked at his skin, shifted constantly on his feet, and jiggled his arms and legs. His speech was rapid, and his mood was irritable. He had been using intravenous (IV) methamphetamine on a “run” for more than 1 week and was brought to the emergency room by relatives who were concerned that he was starting to get paranoid.

In catatonic excitement, much less common now than in the pre-neuroleptic era, patients exhibit disorganized and overactive behaviors, including frantic jumping, thrashing of limbs, and seemingly senseless menacing or attacking behaviors. Such excitement is seen in mania, periodic catatonia, catatonic forms of schizophrenia, and some culture-bound syndromes such as amok. Confusional excitement is a state of restlessness and generalized purposeless activity seen in ictal states, some acute intoxications, and deliria.

Decreased Motor Activity

Global reductions in motor activity—motor retardation—are seen in a variety of physical disorders, such as hypothyroidism, Addison disease, some infectious and postinfectious conditions, including CFS and post-polio syndrome, and other fatiguing conditions, as well as in some organic mental disorders, intoxications, schizophrenias, and depressive disorders. Poverty of movement (akinesia, or more properly, hypokinesia) may occur in schizophrenia and as a neuroleptic side effect. Changes in the voice frequently accompany the reduced motor activity in schizophrenia and depression, with normal inflection replaced by monotonous tone and prolonged speech latency. In stuporous states, patients remain immobile, although their eyes are open, and they are apparently awake.

Conversion reactions are functional, nonphysiologic, psychogenic impairments in sensory or motor functions. Common motor forms include various paralyses and pareses, including limb paralyses, ataxias, and aphonias. Patients with astasia-abasia have marked unsteadiness of gait. Sensory conversion reactions include blindness, deafness, anesthesia, and analgesia. Some hyperesthesias and pain syndromes may also originate as conversion symptoms.

An 18-year-old new army recruit from a rural area was brought to the emergency department at an Army post during his period of basic training because his gait had become very unsteady. He was stumbling into walls and seemed unable to maintain an erect posture. He found it necessary to use crutches to ambulate. There was no history of substance use. On examination, the soldier was found to be naïve, tremulous, anxious, and homesick. He had just graduated from high school and had never been away from his family before and missed them deeply. When he was reassured that he would be issued a pass to see his family, his demeanor changed, and his gait instantly improved. To the amazement of his bunkmates and drill sergeant, he was able to leave his crutches at the door. The diagnosis was astasia-abasia.

Catatonic Activity

Catatonia refers to a broad group of movement abnormalities usually associated with schizophrenia but also found in other disorders such as mania, depression, many neurologic disorders (especially those involving the basal ganglia, limbic system, diencephalon, and frontal lobes), systemic metabolic disorders, toxic drug states, and periodic catatonia. Catatonic stupor and excitement have already been noted. Stereotypies are repetitious, bizarre, seemingly non–goal-directed, complex organized gestures or postures that are believed to have private meanings to the patient. Examples include continuously and repeatedly crossing oneself or blessing others in a religious gesture, waving in a stylized manner, and making profane gestures. The stereotypical behaviors commonly seen in autistic children (constant spinning or rocking, often called “stimming”) may provide self-soothing, steady sensory input that helps the patients reduce the degree to which they are disturbed by the ordinarily unpredictable and uncontrollable stimulation coming from the environment. Bizarre posturing may also be seen in catatonia. One chronic schizophrenic patient routinely stood for hours on one leg with his arms in the air like a crane. In echopraxia, the patient imitates the examiner’s movements and in echolalia imitates speech, as if in mimicry. Some catatonic patients exhibit waxy flexibility, maintaining unusual postures in which they have been posed for prolonged periods of time. Negativism may take the form of refusing to behave in a prescribed manner or resisting passive movement.

Mutism

Mutism may result from a variety of peripheral muscle and CNS conditions and from functional disorders. Mutism may occur in profound depression, catatonic states, and conversion reactions. Elective mutism is occasionally seen in acute adjustment disorders and some personality disturbances.

MOTOR DISTURBANCES AND MOVEMENT DISORDERS

Tremor

Tremors, involuntary oscillating movements of the limbs or head, may occur at rest or with movement. Physiologic tremors, which are minimal at rest and increase with activity, are characterized by small amplitudes and high frequencies. They are characteristic of anxiety, fatigue, and toxic or metabolic disorders, such as caffeinism or hyperthyroidism, and are commonly seen in patients taking a number of different psychiatric medications, including lithium, valproate, and stimulating antidepressants. Coarse tremors, with larger amplitudes and lower frequencies, are seen in Parkinson disease and cerebellar disease. Asterixis is a large-amplitude, flapping tremor of the hands seen in hepatic disease. Parkinsonian symptoms and signs may be seen in psychiatric disorders, particularly in patients taking antipsychotic medications. Symptoms include akinesias with a marked decrease in normally spontaneous fidgeting, stiff gait with diminished arm swing, pill-rolling non-intention tremors (which seem to be less common in drug-induced parkinsonism, compared with the idiopathic type), expressionless soft and monotonous speech, micrographical handwriting, and cogwheel rigidity.

Blepharospasm

Blepharospasm is a rapid and violent repetitive, spasmodic movement of the eyelids. These movements are often a side effect of antipsychotic or other medications but are also common in a variety of neurologic disorders, including Meige syndrome and Tourette syndrome.

Tics

Tics are rapid, repetitive, often spasmodic jerking involuntary movements that serve no apparent purpose. The person may try to disguise or hide the tic in a seemingly purposive movement, and the movement may ultimately be shaped into a mannerism. Tics are the central feature of tic disorders, are associated with other disorders, and may occur as a consequence of stimulant use. Tourette disorder is characterized by a chronic shifting array of motor and vocal tics. The tics may include grunts, coughs, clicks, or sniffs, whereas motor symptoms may include eye blinking, tongue protrusions, facial grimacing, hopping, and twitches. Complex tics may merge into complex compulsive behaviors such as squatting, deep knee bends, and retracing steps. Coprolalia, characterized by sudden verbal outbursts of obscenities, occurs in less than one-third of Tourette patients. Mental coprolalia is an associated feature in which obscene words or phrases suddenly intrude into consciousness in an ego-dystonic manner. Obsessive-compulsive symptoms, as well as attention-deficit symptoms, are also common in Tourette syndrome.

A 21-year-old young man with Tourette syndrome presented to his psychiatrist's office after missing his last two appointments and being off his medications for 1 month. He had recently started a new job delivering pizza. He reported he could no longer control his compulsion to count stairs, tap his fingers, and check his car tire pressure multiple times. During the appointment, while talking, he began blinking his eyes, making guttural noises to clear his throat, followed by a coughing sound, nose picking, and then loudly squealed in a manner that sounded like a cat's meow.

DRUG-INDUCED MOVEMENT DISORDERS

Dystonic Movements

Although dystonic movements are seen in many neurologic disorders, in psychiatric patients, they are almost always secondary to the use of antipsychotic medications. Dystonic reactions consist of intermittent or sustained muscle spasms, typically of the head or neck. Common varieties include tongue spasms causing dysarthria, torticollis (neck spasm), and oculogyric crisis in which there is a forced upward gaze. Opisthotonus (spasms of paraspinal muscles leading to an arched posture) is seen less often. These reactions are most common in young males and typically occur soon after beginning or increasing the dose of a conventional antipsychotic medication.

Akathisia

Akathisia is a syndrome of motor restlessness seen predominantly in the context of antipsychotic and some antidepressant medication use. It has subjective, as well as motor, components. Subjectively, patients experience muscle tension, difficulty finding a comfortable body position, and inability to stop moving; they feel as though they are "jumping out of their skin." Objectively, akathisia classically manifests by rocking from foot to foot while standing, frequently crossing and uncrossing the legs when seated, and pacing. Sleep may be disturbed because of physical discomfort. Subjective components of akathisia may be difficult to distinguish from anxiety caused by the primary disorder (typically schizophrenia). Rarely, the restlessness and inner agitation become sufficiently uncomfortable to provoke acts of violence. In pseudoakathisia, objective signs of akathisia are present, but the patient denies feeling restless.

A 45-year-old man with OCD and Tourette syndrome being treated with fluvoxamine, 200 mg per day, inadvertently took 400 mg per day for 3 days. On the third day, he started to feel increasingly agitated, his legs felt restless and jittery, and he started to fidget uncharacteristically. He also noted frequent sighs in his breathing, not his usual respiratory pattern. He recalled feeling these sensations a decade earlier when a previous psychiatrist treated his Tourette syndrome with haloperidol and that his psychiatrist had diagnosed akathisia. After discontinuing the fluvoxamine for 4 days, these symptoms all abated.

Tardive Dyskinesia

Tardive dyskinesia is a movement disorder that occurs only in the context of antipsychotic medication use, occasionally after many months but more commonly after years. The abnormal movements may persist with or without continued medication use or may diminish or disappear over time. The dyskinetic movements occur at rest and can usually be temporarily suppressed volitionally or by purposeful action, distraction, or sleep. The movements are varied. In the most common type, which affects the face, especially the mouth and lips, tongue thrusting, chewing movements, lip smacking, and eye blinking are seen. Another common type is characterized by choreoathetoid movements, such as writhing finger motions. In the less common but more severe truncal dyskinesias, the torso moves in thrusting motions, and respiratory dyskinesia is characterized by grunting and irregular breathing patterns. Other tardive (late) syndromes include tardive akathisia and tardive dystonia, in which the abnormal movements emerge late in treatment or upon medication discontinuation.

Neuroleptic Malignant Syndrome

Neuroleptic malignant syndrome (NMS), a potentially fatal complication of antipsychotic medication, is characterized by muscle rigidity, fever, diaphoresis, delirium, mutism, and blood pressure abnormalities. Some view NMS as the most severe end of a spectrum that starts with antipsychotic-induced parkinsonism and progresses to extrapyramidal syndrome with fever and then to fulminant NMS.

Rabbit Syndrome

This uncommon drug-induced extrapyramidal syndrome is often misdiagnosed as tardive dyskinesia. It most closely resembles a limited expression of a parkinsonian tremor. Patients make rapid chewing movements similar to those made by rabbits, ordinarily faster and more regular than the orofacial tic of tardive dyskinesia. The tongue is spared.

Serotonin Syndrome

Serotonin syndrome is a disorder caused by excessive serotonergic input in the CNS, probably within the hypothalamus. The most common cause of serotonin syndrome is the combination of two or more medications with serotonin-enhancing properties, usually by different mechanisms. It is characterized by restlessness, myoclonus, hyperreflexia, diaphoresis, shivering, tremor, autonomic changes, including fever, and mental status changes, such as confusion.

MOTOR DISTURBANCES OF SCHIZOPHRENIA

Many of the abnormal movements ascribed to tardive dyskinesia and other antipsychotic-induced extrapyramidal syndromes have been described in chronically psychotic patients before the introduction of antipsychotic medications. In one series of 100 patients, the large majority of whom were diagnosed as schizophrenic, a review of medical records before 1955 revealed that abnormal purposive movements were found in 83%, mannerisms and tics in 71%, abnormal eye movements in 27%, abnormal postures or facial movements in 42%, and gait abnormalities in 10%. These findings suggest that many patients with schizophrenia have neurologic symptoms not due to medications and that severe psychiatric disorders may have a neurologic component as well.

OTHER MOVEMENT DISTURBANCES

Gait disturbances in patients with psychiatric disorders include a variety of neurogenic gaits consistent with brain disease, intoxications, and medication side effects. These include the festinating gait of parkinsonism, spastic and ataxic gaits of neurologic disease and psychiatric medications, waddling and reeling gaits associated with intoxications, and the nonphysiologic gait disturbances seen in astasia-abasia, a form of conversion disorder. Gait mannerisms include clowning, prancing, and military gaits.

Bruxism, chronic jaw clenching, may occur involuntarily during tension states, as an isolated occurrence during delta sleep, in which it has sometimes been associated with benzodiazepine or alcohol use, or in association with SSRI use. In severe cases, serious damage to dental enamel and temporomandibular joint pain may occur.

Myoclonus, characterized by focal muscle jerking, can be caused in psychiatric patients by certain medications, such as SSRIs or monoamine oxidase inhibitors (MAOIs). Myoclonic jerks may be difficult to distinguish from tics, but the latter often represent larger muscle groups and more highly organized motor patterns.

Myoclonus may be seen at rest but is more obvious during motor activity.

Seizure-Like Behaviors

In addition to the generalized, petit mal, and complex partial seizures seen in some psychiatric patients, a number of nonepileptic seizure-like behaviors must be distinguished. Breath-holding spells, generally innocuous, impulsive, and tantrum-like phenomena, usually occur in small children who hold their breath during moments of oppositional rage and who may faint as a result. Jerking or twitching motor movements may occur. Temper tantrums in young children may look like seizures, especially to the uninformed observer. The children may lie on the floor, screaming and kicking, and do not respond to the environment. Conversion seizures (hysterical seizures or pseudoseizures) must be differentiated from genuine epileptic seizures. Patients retain consciousness, lack abnormal reflexes, and are not incontinent. However, because so many conversion seizures occur in patients who have genuine epilepsy and who know a good deal about the condition, the differential diagnosis is sometimes difficult.

Compulsive Behaviors

Compulsive behaviors may occur in relation to everyday activities, such as gambling, sexual conquest, shopping, and watching television, or in relation to substances such as alcohol, cocaine, narcotics, and food. Other compulsions involve reckless risk-taking behaviors that provide stimulation and dispel dysphoric moods. Sexual compulsive perversions, such as exhibitionism and sadomasochism, may serve similar purposes. Compulsions are seen in a variety of psychotic and nonpsychotic psychiatric disorders. The cravings that underlie compulsive behaviors are strong motivating forces, and the compulsive behaviors may regulate emotions. Unknown similarities may underlie all compulsive and addictive mechanisms. Current controversies surround the relationships of compulsive, impulsive, and addictive behaviors. Some authorities have proposed that all three should be subsumed under the rubric of a so-called reward deficiency syndrome, resulting from various combinations of gene polymorphisms governing dopamine and opioid mechanisms and including serotonin, cannabinoid, and other transmitter systems as well. From this perspective, the compulsive, impulsive, and addictive behaviors all increase the amount of dopamine and potentially other transmitters in specific brain areas. Studies in the past decade have shown that, in various subject groups, the Taq 1 a1 allele of the DRD2 gene is associated with alcoholism, drug abuse, smoking, obesity, compulsive gambling, and several personality traits. Several other candidate genes are also under investigation. Studies have also shown that certain impulse-like behaviors may be associated with very specific genetic polymorphisms. In studies of patients with severe obesity, binge-eating behavior was seen only in those individuals who had specific mutations in genes controlling the expression of melanocortin 4 receptors.

In OCD, the compulsions are ritualized, repetitive behaviors that are performed with the goal of satisfying, neutralizing, and undoing obsessional thoughts. Although intended to decrease anxiety, rituals are never more than transiently successful. The most common compulsions involve checking to make certain that gas jets and faucets have been turned off and that windows and doors are locked, hand washing, repeating certain phrases, counting objects, and placing objects in a prescribed order.

A 47-year-old woman had always been fussy about order and cleanliness but, since the birth of her child 14 years before, had become increasingly obsessed by cleanliness. She spent 3 hours daily ritualistically cleaning the apartment, using three to four rolls of paper towels daily. If a family member left even a fingerprint on a table, she became agitated and spent much time recleaning the surface and admonishing the family member.

FUTURE PERSPECTIVES

Particularly in the last decade and in the near future, there is hope that new research and advances in the neurosciences will yield an even greater understanding about psychiatric illness. However, clinical acumen in the practice of psychiatry is still as relevant today as during the origins of psychiatry. A firm conceptual grasp of the clinical descriptors of psychopathology is an important part of a clinician's armamentarium. As yet, nothing has replaced the understanding of psychiatric signs and symptoms integrated in the context of the diagnostic interview and mental status examination in formulating an accurate differential diagnosis and ultimately a comprehensive management plan. Even as psychiatrists move toward further improving and modernizing the categorical diagnostic systems vis-à-vis the DSM-5 diagnostic manual, clinicians will continue to strive to hone and refine their understanding of the clues provided in the clinical manifestations of psychiatric disorders in order to best care for their patients.

ACKNOWLEDGMENTS

The authors wish to acknowledge the work of Dr. Joel Yager and Dr. Michael J. Gitlin who wrote this chapter in the eighth edition of the *Comprehensive Textbook of Psychiatry,* part of whose work was adapted for this chapter.

FURTHER READINGS

Alarcón RD. Culture, cultural factors and psychiatric diagnosis: review and projections. *World Psychiatry*. 2009;8(3):131–139.

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Bakker SC, Hoogendoorn ML, Hendricks J, et al. The PIP5K2A and RGS4 genes are differentially associated with deficit and non-deficit schizophrenia. *Genes Brain Behav*. 2007;6:113–119.

Bechara A, Berridge KC, Bickel WK, Morón JS, Williams SB, Stein JS. A neurobehavioral approach to addiction: implications for the opioid epidemic and the psychology of addiction. *Psychol Sci Public Interest*. 2019;20(2):96–127.

Becker JB, Chartoff E. Sex differences in neural mechanisms mediating reward and addiction. *Neuropsychopharmacology*. 2019;44(1):166–183.

Brebion G, David AS, Jones HM, Ohlsen R, Pilowsky LS. Temporal context discrimination in patients with schizophrenia: associations with auditory hallucinations and negative symptoms. *Neuropsychologia*. 2007;45:817–823.

Chamberlain SR, Sahakian BJ. The neuropsychiatry of impulsivity. *Curr Opin Psychiatry*. 2007;20:255–261.

Choong C, Hunter MD, Woodruff PW. Auditory hallucinations in those populations that do not suffer from schizophrenia. *Curr Psychiatry Rep*. 2007;9:206–212.

Feusner JD, Phillips KA, Stein DJ. Olfactory reference syndrome: issues for DSM-V. *Depress Anxiety*. 2010;27(6):592–599.

Frank GK. Recent advances in neuroimaging to model eating disorder neurobiology. *Curr Psychiatry Rep*. 2015;17(4):559.

Girard TA, Martius DL, Cheyne JA. Mental representation of space: insights from an oblique distribution of hallucinations. *Neuropsychologia*. 2007;45:1257–1269.

Grano N, Keltikangas-Jarvinen L, Kouvonen A, et al. Impulsivity as a predictor of newly diagnosed depression. *Scand J Psychol*. 2007;48:173–179.

Habtewold TD, Rodijk LH, Liemburg EJ, et al. A systematic review and narrative synthesis of data-driven studies in schizophrenia symptoms and cognitive deficits. *Transl Psychiatry*. 2020;10:244.

Hawkins M, Sockalingam S, Bonato S, et al. A rapid review of the pathoetiology, presentation, and management of delirium in adults with COVID-19. *J Psychosom Res*. 2021;141:110350.

Hawton K, Harriss L. How often does deliberate self-harm occur relative to each suicide? A study of variations by gender and age. *Suicide Life Threat Behav*. 2008;38:650–660.

Karg K, Burmeister M, Shedden K, Sen S. The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch Gen Psychiatry*. 2011;68(5):444–454.

Keng A, Brown EE, Rostas A, et al. Effectively caring for Individuals with behavioral and psychological symptoms of dementia during the COVID-19 pandemic. *Front Psychiatry*. 2020;11:573367

Kounali D, Zammit S, Wiles N, et al. Common versus psychopathology-specific risk factors for psychotic experiences and depression during adolescence. *Psychol Med*. 2014;44(12):2557–2566.

Kreek MJ, Levran O, Reed B, Schlussman SD, Zhou Y, Butelman ER. Opiate addiction and cocaine addiction: underlying molecular neurobiology and genetics. *J Clin Invest*. 2012;122(10):3387–3393.

Lewis-Fernández R, Aggarwal NK. Culture and psychiatric diagnosis. *Adv Psychosom Med*. 2013;33:15–30.

Mann JJ, Currier DM. Stress, genetics, and epigenetic effects on the neurobiology of suicidal behavior and depression. *Eur Psychiatry*. 2010;25(5):268–271.

Nenadić I, Voss A, Besteher B, Langbein K, Gaser K. Brain structure and symptom dimensions in borderline personality disorder. *Eur Psychiatry*. 2020;63(1):e9.

Povroznik J, Ozga JE, Vonder Haar C, Engler-Chiurazzi EB. Executive (dys)function after stroke: special considerations for behavioral pharmacology. *Behav Pharmacol*. 2018;29(7):638–653.

Riemann D, Kloepfer C, Berger M. Functional and structural brain alterations in insomnia: implications for pathophysiology. *Eur J Neurosci*. 2009;29:1754–1760.

Riper H, Andersson G, Hunter SB, de Wit J, Berking M, Cuijpers P. Treatment of comorbid alcohol use disorders and depression with cognitive-behavioural therapy and motivational interviewing: a meta-analysis. *Addiction*. 2014;109(3):394–406.

Roy D, Ghosh R, Dubey S, Dubey MJ, Benito-León J, Ray BK. Neurological and neuropsychiatric impacts of COVID-19 pandemic. *Can J Neurol Sci*. 2021;48:9–24.

Van't Wout M, Aleman A, Bermond B, Kahn RS. No words for feelings: alexithymia in schizophrenia patients and first-degree relatives. *Compr Psychiatry*. 2007;48:27–33.

Wagner G, Koch K, Schachtzabel C, Schultz CC, Sauer H, Schlösser RG. Structural brain alterations in patients with major depressive disorder and high risk for suicide: evidence for a distinct neurobiological entity? *Neuroimage*. 2011;54(2):1607–1614.

Walder DJ, Faraone SV, Glatt SJ, Tsuang MT, Seidman LJ. Genetic liability, prenatal health, stress and family environment: risk factors in the Harvard Adolescent Family High Risk for Schizophrenia Study. *Schizophr Res*. 2014;157(1–3):142–148.

Weintraub D. Neuropsychiatric symptoms in Parkinson disease and dementia with lewy bodies: what geriatric psychiatry can learn. *Am J Geriatr Psychiatry*. 2013;21(6): 497–500.

Yamamoto V, Bolanos JF, Fiallos J, et al. COVID-19: review of a 21st century pandemic from etiology to neuro-psychiatric implications. *J Alzheimers Dis*. 2020;77(2):459–504.

Zanarini MC, Frankenburg FR, Reich DB, Silk KR, Hudson JI, McSweeney LB. The subsyndromal phenomenology of borderline personality disorder: a 10 year follow-up study. *Am J Psychiatry*. 2007;164(6):929–935.

9

Classification in Psychiatry

9.1 Present and Future of Classification Systems for Mental Disorders

Javier I. Escobar, M.D., M.Sc., and Humberto Marin, M.D.

This chapter will first discuss the importance of classification systems in the field of medicine, with particular emphasis on mental disorders. It will then provide highlights on the history of psychiatric diagnosis. Next, it will focus on historical, diagnostic, practical, and developmental issues related to current classifications with a particular focus on the Diagnostic and Statistical Manual (DSM) of the American Psychiatric Association (APA), the current classification for mental disorders in North America. Some general information will also be provided on the International Classification of Diseases (ICD) of the World Health Organization (WHO), including a few highlights on the recently released ICD-11.

DSM-5, the latest version of the DSM released in 2013, will be discussed in detail, highlighting major changes from previous editions, its advantages, disadvantages, and shortcomings, as well as practical aspects for its clinical use. Recent updates (from 2014 onward) to the DSM-5 will also be briefly outlined.

Finally, this chapter will present some new developments, particularly the use of dimensional diagnoses, with emphasis on the Research

Domain Criteria (RDoC) and the potential of these new developments for clinical and research applications.

CLASSIFICATION IN MEDICINE

Classification systems are essential in all sciences. In the case of the medical field, the main functions for a diagnostic classification are communication and prediction. “Diagnosis is prognosis” was the eloquent articulation of two pioneers of psychiatric diagnosis, Donald W. Goodwin and Samuel Guze, both working at Washington University in St. Louis in the early 1970s.

In the field of psychiatry, a critical and perennial problem in developing valid classifications is the fact that for most mental disorders, there are no beacons or reliable markers that would allow the unequivocal identification and systematic separation of disorders. Despite this absence, the functions of psychiatric classifications are still quite relevant, even if they are currently present only as tentative charts or maps, or more descriptively, “blurry lighthouses” that help guide practitioners in the wide, stormy sea of clinical practice. No matter how primitive they may be, these classification systems bring practitioners at least a modicum of security and reassurance for dealing with the complex issues of human behavior, thought processes, cognition, and other subjective phenomena, as the specialty of psychiatry tries to emulate or at best imitate what other medical specialties do in this regard.

In the absence of objective, reliable, measurable markers for specific mental disorders, there are only a few feasible ways for developing classifications of mental illness. The first and most traditional example has been the authority-based method used for classifying mental illness until the mid-20th century. In this tradition, the descriptions of clinical disorders originally made or endorsed by highly visible authorities in the field were gradually accepted until they became common currency.

The strategies that followed the early models added new components, incorporating a consensual, collegiate mechanism, not very different from that of legislative bodies in a democratic state. A key objective of these new efforts was to fill many of the existing gaps in scientific evidence. This process progressively engaged more and more individuals in the discussion. Remarkably, in the end, this strategy got almost everyone on the planet to utilize the same language when discussing mental illness.

However, as the field of medicine becomes more and more evidence based, a tension has surfaced between the need to adopt the growing body of scientific evidence and the wish to satisfy the needs and interests of a large number of constituent groups, including researchers, clinicians, patients, advocates, public and private health institutions, pharmaceutical companies, data managers, and others. In search of a more thorough inclusion of potential clinical syndromes and seeking additional specificity, the psychiatric specialty has kept building diagnosis after diagnosis, disorder after disorder, hoping perhaps that this would lead to more precise treatment interventions and allow a better prediction of outcomes. However, as it will be seen ahead, the proliferation of disorders supported by rather slim and fragile scientific evidence has become worrisome.

As virtually all of those practicing in the mental health field today were trained as part of the DSM tradition, the system is generally taken for granted, used mechanistically, with little thought about its origins and evolution, benefits, and limitations. Therefore, background information on the origins, evolution, strengths, and weaknesses of the DSM will be provided below.

BRIEF GLOSSARY

First, it is important to define the key terms and concepts to be used herein:

- A *syndrome* is a cluster of symptoms that can result from different disease processes. For example, cough and fever can result from bacterial, fungal, or viral infections, or autoimmunity, all processes implying different interventions and outcomes.
- A *disorder* is something of medical concern in a patient, an abnormality, injury, disability, or aberration. This term is often used instead of the term *disease,* which implies a known pathologic process, used in the case of medical conditions for which the causes are known, and the pathophysiology well established.
- A *diagnosis* is simply the opinion, by someone with expertise in the matter, that a given disorder is present (or absent) in a patient. *Diagnosis* is also used to name the procedure to decide whether or not a certain disease or disorder is present (or absent). A *diagnosis* can be “*categorical*” (the condition is either present or absent) or “*dimensional*” (the condition represents a certain point in a continuum).

- *Diagnostic classification* is the listing of diagnoses grouped by relatedness. Examples include infectious, autoimmune diseases, cancer, or injuries in medicine, or anxiety, affective or psychotic disorders in psychiatry.
- *Diagnostic criteria* are the rules that need to be followed for making a diagnosis. Once made, a diagnosis is *reliable* if the same conclusion about its existence can be reached by at least two independent observers (inter-rater reliability) or if the patient receives the same diagnosis when examined more than once within a reasonable time period (test–retest reliability). A diagnosis is *valid* if it picks out a “real” entity based on etiology and pathophysiology.
- Diagnostic reliability and validity: It is important to emphasize that *reliability* and *validity* are separate dimensions. In the case of mental disorders, the fact that different observers are able to recognize a set of symptoms when they appear together (reliability) does not necessarily mean that they correspond to a real entity (validity). In fact, the emphasis on diagnostic reliability may have limited the progress in reaching diagnostic validity, especially in the case of disorders with a complex or fluctuating clinical course. Clearly, validity remains elusive for a great majority of psychiatric conditions.

DEFINITION OF MENTAL DISORDER (DSM-5)

As defined in the DSM-5 Manual, “A mental disorder is a syndrome characterized by clinically significant disturbance in an individual’s cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological or developmental processes underlying mental functioning. Mental disorders are usually associated with significant distress or disability in social, occupational, or other important activities. An expectable or culturally approved response to a common stressor or loss, such as the death of a loved one, is not a mental disorder. Socially deviant behavior (e.g., political, religious, or sexual) and conflicts that are primarily between the individual and society are not mental disorders unless the deviance or conflict results from a dysfunction in the individual, as described above.” The above definition may be difficult to apply to the totality of conditions listed in the manual, and it may be hard to assure or convince practitioners that each disorder included in DSM-5 truly represents a “specific dysfunction in processes underlying mental functioning.”

BRIEF HISTORY OF PSYCHIATRIC DIAGNOSIS

While the concept of mental illness dates back at least as far as the Greek civilization or even further back (as suggested by the finding of cranium trepanations in prehistoric times), there has always been significant interest and effort placed on the concept of mental disease or infirmity from ancient times to current day, as healers, physicians, and thinkers from different centuries described mental conditions that are recognizable by today's standards.

Regarding the diagnostic controversy between categories and dimensions, it is interesting to add a historical note to highlight that the Greek philosophers might have started this debate. Indeed, the platonic ideal of classification consisted of an exhaustive, mutually exclusive categorical system, while the Aristotelian approach conceived dimensions, described as overlapping entities, separated from normality by arbitrary thresholds.

Many of the old, historical, or classical designations for mental disorders are still in use today, at least in common parlance, and until very recently, some of them were formally utilized in psychiatric nomenclatures as it was the case with hysteria, hypochondria, mania, and melancholia, the four classic mental syndromes of yesteryear according to Michel Foucault's influential book of the early 1960s.

Systematic, "official" classifications and nomenclatures for mental disorders are rather recent undertakings that convey additional features to the process of naming mental conditions. These systems transform this process into a social enterprise, with implications in many domains of life (e.g., fostering scientific dialogue and progress, public discussion and education, policy decisions, reimbursement for services, data management, treatment, and others).

The introduction, by the middle of the last century, of classification systems for mental disorders that were comprehensive and practical enough to gain wide acceptance was a critical factor contributing to the development of modern psychiatry as a branch of the medical sciences, allowing the specialty to go beyond a simple "humanistic" endeavor. Thus, homage should be paid to the people from multiple disciplines that contributed to the creation, implementation, and, perhaps more importantly, to the reform of older classification systems, in order to make sure that they would remain relevant up to this day and age.

CLASSICAL EUROPEAN PSYCHOPATHOLOGY

In the 19th century and the beginning of the 20th century, key figures of classical European psychopathology were Emil Kraepelin, with his descriptive psychopathologic approach, Carl Wernicke, the main proponent of a biologic approach, and Karl Jaspers, a philosopher who applied his phenomenologic approach to psychiatric assessment and diagnosis. During that same period, North American psychiatry followed the lead of European psychiatry, focusing mainly on the Kraepelinian line. Toward the end of this chapter, these historical approaches will be briefly contrasted with current developments in psychiatric classification, highlighting the need for a richer assessment of psychopathology.

THE U.S. CLASSIFICATION SYSTEMS LEADING TO DSM-5

The intent to record and catalog mental illness in the United States started with the 1840 census, a process that, besides counting people, also incorporated scrutiny of physical illnesses as well as a category for “idiocy/insanity.” In the 1880 census, the categories “mania, melancholia, monomania, paresis, dementia, dipsomania, and epilepsy” were added, and these categories were quickly adopted by the “Association of Medical Superintendents of American Institutions for the Insane,” a society formed in 1844, which in 1892 changed its name to “American Medico-Psychological Association,” and that in 1921 would become the “American Psychiatric Association.”

In 1917, together with the National Commission on Mental Hygiene, the society that was the precursor of the APA, developed the “Statistical Manual for the Use of Institutions for the Insane,” and also, with support from the New York Academy of Medicine, the same APA precursor developed the “Mental Subsection of the United States General Medical Guide,” the “Standard Classified Nomenclature of Disease,” known simply as the “Standard.” During World War II, U.S. psychiatrists working in the armed forces were involved in the selection and treatment of soldiers, and the Office of the Surgeon General issued a special classification called “Medical 203,” which was adopted with modifications by different branches of the armed forces and the Veterans Administration. These different versions of “Medical 203” continued to be used, along with the

“Standard,” and this contributed to the “confusion” highlighted in the foreword to the first edition of the DSM that was introduced in 1952. In a “back to the past” fashion, the controversies related to psychiatric nomenclatures reached a level of “confusion of tongues” in 1948, similar to the one that had plagued the medical nomenclatures in the 1920s. At least three psychiatric nomenclatures were in general use then, the “Standard,” and the ones developed by the Armed Forces and the Veterans Administration. All these classification systems were quite disparate and did not fall in line with the International Statistical Classification, so that some of the agencies were in the uncomfortable position of using one nomenclature for clinical use, a second one for disability ratings, and a third one, the international classification, for statistical work. In addition, practically every teaching center in the country had made modifications to the “Standard” for its own use, and assorted modifications of the Armed Forces nomenclature had been introduced into many clinics and hospitals by psychiatrists returning from military duty.

INTERNATIONAL CLASSIFICATION OF DISEASES

The ICD was the descendant of the “Bertillon Classification of Causes of Death,” introduced by the French physician Jacques Bertillon in 1893 at the congress of the International Statistical Institute in Chicago. Multiple countries adopted this classification, including those in North America, and in 1900, an international conference revised what was then called the “International Classification of Causes of Death” and agreed to hold revisions of the system every 10 years. In 1948, the then recently created WHO, an agency of the United Nations, assumed responsibility for the ICD. The sixth revision, published in 1949, involved significant changes, as it included morbidity in addition to mortality, so that the title was changed to “International Statistical Classification of Diseases, Injuries, and Causes of Death.” It is important to point out that the sixth revision of the ICD was the first one to include a section on mental disorders. The ICD is now in its 11th official edition, and this was released only very recently. Both the DSM and the ICD sections on mental disorders have been working in a collaborative fashion for several decades, and the two systems have become very similar. Therefore, many of the general comments that follow on the DSM would also apply to the mental disorders section of the ICD.

EVOLUTION OF MODERN PSYCHIATRIC NOSOLOGY IN THE UNITED STATES

In debating the merits and limitations of the DSM (and obviously, the mental disorders section of the ICD system as well), it is important to remember the historical context operating at the time when these systems originated and how the classifications have evolved (or regressed) with the passage of time. The first edition of the DSM was published in 1952, and the first ICD, including a mental nosology, the ICD-6, was published in 1949. This happened long before the emergence of "evidence-based" medicine and psychiatry, at a time when mental nosology in the United States was in a relatively dormant state. Indeed, those were the 1940s and 1950s, the decades of the clinical and theoretical hegemony of both the psychoanalytic theory of Freud and his followers and the biopsychosocial approach of Adolf Meyer. For both of these perspectives, mental illness was a reaction of the mind to outside forces (the family environment for psychoanalysis and the psychological, social, and biologic realms for Meyer and his followers). Thus, in clinical practice, diagnosis made little practical difference as many practitioners would adhere to the theoretical postulates proposed by these dominant systems. Thus, before the middle of the 20th century, mental health clinicians practicing in North America learned their trade and applied it to their clinical practice, with little regard to classifications, research evidence, theories of disease, or potential treatments. A few enclaves in the United States, particularly in the mid-Atlantic area (St. Louis, Baltimore, New York), maintained at least part of the European tradition (again, mainly the Kraepelinian one) and were also influenced by the medical model approach and philosophical realism (positivism, empiricism) more than the phenomenologic model (Jaspers).

It was only in the 1950s that evidence-based psychiatry began to emerge with the development of the first sets of psychotropic drugs backed by systematic research, such as lithium, phenothiazines, tricyclic antidepressants, mono amino oxidase (MAO) inhibitors, and benzodiazepines. Evidence-based psychotherapies such as cognitive behavioral therapy (CBT) would have to wait much longer (until the late 1970s) to be confirmed as "effective" by well-controlled clinical trials.

This was the environment in which the first edition of the DSM (without numeral) was produced, in 1952, by a group of seven people commissioned by the APA. The 1952 DSM lists 108 disorders in a descriptive section that takes only 32 small pages of text. However, a significant number of these items correspond to neurologic and other medical disorders that have since then migrated to other specialties. Thus, the actual number of mental disorders in the first DSM was in the two-digit range, a number that has grown to almost 400 in DSM-IV and DSM-5 (Fig. 9.1–1). This explosion in the number and detail of psychiatric disorders that started with the DSM-III will be discussed in more detail ahead in this chapter.

It is important to observe that despite the dominance of the DSM in the United States since the introduction of its first edition in 1952, other methodologies and proposals for psychiatric classifications have been periodically offered, such as the St. Louis criteria and the Research Diagnostic Criteria (RDC). These differ from the DSM in that they placed a major emphasis on available scientific evidence. Some of these proposals have, in turn, influenced the DSM, as will be seen later.

A pivotal U.S./U.K. diagnostic study, demonstrating that training and collaboration across the Atlantic would make compatible the differing diagnostic practices of North American and British psychiatrists, and next, the International Pilot Study of Schizophrenia (IPSS), showing that concordant schizophrenia syndromes could be reliably identified in nine countries of the world, had both an important impact on modern diagnostic developments. The Present State Examination (PSE) developed at Maudsley hospital under the leadership of John Wing, a psychiatrist adhering to Jaspers' phenomenologic approach, provided excellent symptomatic descriptions of the major psychiatric syndromes, and to this date, remains a unique and valuable instrument to assess psychopathology. Its rich descriptions of symptoms and signs make it still ideal for teaching psychiatrists in training about the range of psychopathology.

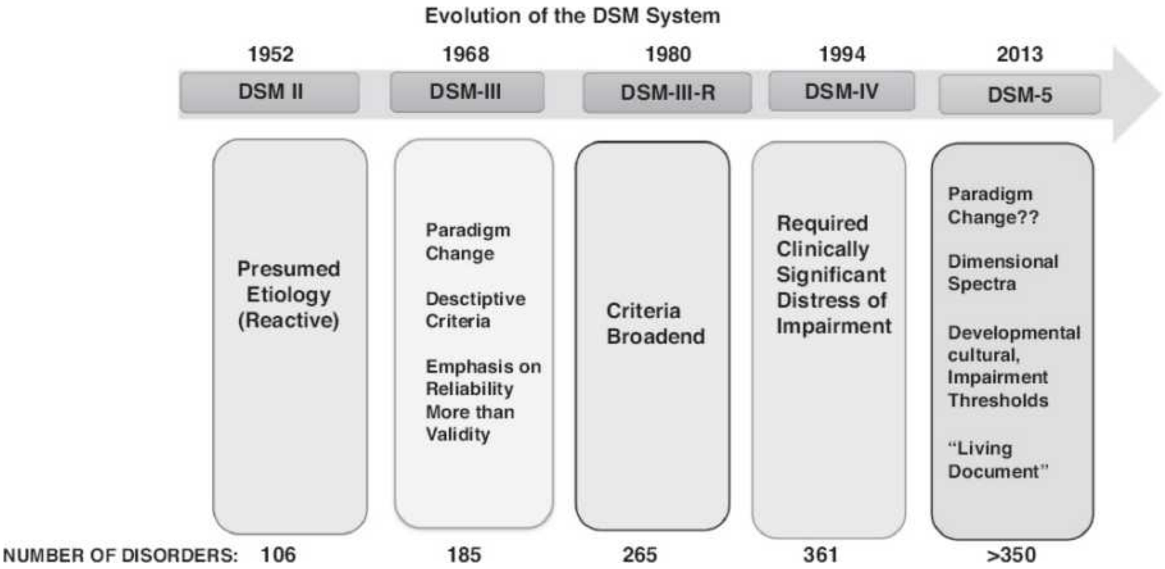
Evolution of the DSM System
1952
DSM II
1968
DSM-III
1980
DSM-III-R
1994
DSM-IV
2013
DSM-5
Presumed Etiology (Reactive)
Paradigm Change
Desctiptive Criteria
Emphasis on Reliability More than Validity
Criteria Broadend
Required Clinically Significant Distress of Impairment
Paradigm Change??
Dimensional Spectra
Developmental cultural, Impairment Thresholds
"Living Document"
NUMBER OF DISORDERS: 106
185
265
361
>350

FIGURE 9.1–1. The DSM system. Key issues and number of categories.

MODERN PSYCHIATRIC CLASSIFICATIONS IN NORTH AMERICA

The St. Louis Group

Modern psychiatric classifications in North America were given a "jump-start" in the 1960s, when a group of investigators at Washington University and Renard Hospital in St. Louis, Missouri (one of the "neo-Kraepelinian" enclaves mentioned above), led by Eli Robins and Samuel Guze, started using a process inclusive of successive steps, such as clinical description, laboratory studies, delimitation from other disorders, follow-up studies, and family history, for validating psychiatric syndromes customarily diagnosed in clinical practice. This led to a list of 12 disorders that were deemed "valid" at that time, as they went through the various validation steps listed above. These "valid" entities included hysteria (somatization disorder), panic disorder (anxiety neurosis), obsessive-compulsive disorder (OCD), phobic disorder, affective disorders (depression, mania), schizophrenic disorder, alcoholism, drug dependence, sociopathy (antisocial personality), brain syndrome, anorexia nervosa, and sexual problems. While the parsimonious Eli Robins and Samuel Guze were not rushing to publicize their strategy quite yet, an ambitious psychiatric resident, the late John P. Feighner, convinced his mentors to prepare a paper for publication. The paper, published in the Archives of General Psychiatry in 1972, has become one of the most quoted papers in the history of psychiatry. Curiously, Goodwin and Guze never mentioned Feighner's paper in any of the editions of their classical textbook "Psychiatric Diagnosis."

Research Diagnostic Criteria

Next, in the 1970s, came the RDC, an effort led by Robert Spitzer at Columbia University, another U.S. enclave where classical psychopathology had a niche. The "neo-Kraepelinian" RDC, largely based on the St. Louis criteria, set the stage for a "paradigm change" in the official APA classification, as Robert Spitzer was named to lead the APA task force on DSM-III. Spitzer would later state: "With its intellectual roots in St. Louis instead of Vienna, and with its intellectual inspiration drawn from Kraepelin, not Freud, the task force was viewed from the outset as unsympathetic to the interests of those whose theory and practice derived from the psychoanalytic tradition." From that point on, almost everyone practicing today knows the rest of the story. This chapter will return to the more recent iterations on the development of the DSM in the sections ahead.

Heuristic Formulations for Psychiatric Diagnosis

In the late 20th century, in Europe, the "psychological dysfunction" model proposed by the British scientist Sir Aubrey Lewis theorized that constructs such as "phobias," "obsessions," or "thought disorder" were probably linked to biologic mechanisms. Lewis also supported the concept of single depression (rather than the melancholic/reactive dichotomy), as a continuum, with "melancholia" in one extreme and "neurasthenia" in the other. Lewis was a key figure in the ICD-8 classification of mental disorders, the first one containing a glossary. According to Fulford and Sartorius, while the ICD-8 and the glossary were based on a report to the WHO in the early 1960s by British psychiatrist Erwin Stengel (which in turn was based on a logical empiricist analysis of psychiatric classification developed by an American philosopher of science, Carl Hempel), it was Lewis who initiated the key move from the theoretical (etiologic) to the descriptive (symptom-based) classification of mental disorders.

Similarly, DSM-IV-R was significantly influenced by the theories of Jerome Wakefield, another philosopher of science, at the time based at Rutgers University in New Brunswick, New Jersey. Wakefield proposed a "harmful dysfunction analysis" as a theoretical backup for psychiatric classifications. According to this theory, there is a subjective element that implies that the condition is “harmful” or “undesirable” to the individual and an objective element that assumes that the condition is due to a “malfunctioning internal mechanism.” Together with Allan V. Horwitz, a social scientist, also at Rutgers, Wakefield published in 2007 a book entitled “The Loss of Sadness: How Psychiatry Transformed Normal Sorrow into Depressive Disorder.” The authors argue in the book that sadness is a normal human emotion that appears in front of losses and misfortunes, and they criticize the almost exclusive biologic focus of psychiatry in interpreting and managing this heterogeneous disorder. To bring into play the contribution of the social sciences, these authors proposed, following the path of DSM-III and IV, the use of bereavement as an exclusionary criterion for depression, adding an additional clause to incorporate and expand relevant contextual factors that would rule out depressive disorder. These would include not only bereavement but also other major life stressors or losses such as marital dissolution, loss of a valued job, and other life events. Little they knew (as will be described ahead under depression) that just a couple of years later, DSM-5 would also strike at the bereavement exclusion.

Dominique Murphy, another philosopher interested in psychiatric diagnosis (who also spent time at Rutgers University), in his book “*Psychiatry in the Scientific Image*,” criticizes the “atheoretical” view of the later editions of the DSM and advocates for a “causal” nosology. While acknowledging the limitations of current theories about causes of mental disorders, Murphy believes it is still possible to move forward even if the theory is not completely worked out. This may become the philosophical or theoretical underpinning for new diagnostic initiatives such as the RDoC to be discussed ahead.

Classifications in Medicine

Despite their pretension of “scientific objectivity,” classification systems in the applied sciences (especially the health sciences) are not created in a social vacuum by completely neutral scientists or practitioners. In the end, these systems represent essentially a consensus or compromise between the needs, beliefs, and expectations of different stakeholders operating at a particular time and place. The complexity of the deliberations is compounded by the fact that the interests and convictions of the various stakeholders may coincide, disagree, or be indifferent to those of the others and may also change over time. As mentioned early in this chapter, diagnostic criteria for diseases in other areas of medicine can often be confirmed with objective, verifiable procedures akin to a “gold standard.” However, there are numerous medical conditions, particularly those entities termed “functional” (chronic fatigue, fibromyalgia, irritable bowel, and many others), for which objective, verifiable, and measurable “gold standards” are not available and whose “validity” is as questionable as that of many mental disorders. Several other “physical” conditions, such as asthma, chronic obstructive pulmonary disease (COPD), and diabetes, among others, continue to present significant challenges for their objective assessment and proper management. For example, criteria for diagnosing COPD include first a set of “key indicators” such as “dyspnea,” “chronic cough and sputum production,” “wheezing and chest tightness,” “history of exposure to risk factors,” and “family history.” Some of these indicators (e.g., dyspnea) are quite subjective, may be hard to assess or detect, particularly in the early stages of the disease, and can be shared by a multitude of conditions. “Objective” assessments such as physical examination or chest x-rays convey limited information in early COPD, and results from “the confirmatory test,” spirometry, the true “gold standard,” are often difficult to interpret unambiguously. Even for disorders with clear, measurable markers such as diabetes, diagnostic thresholds change or fluctuate as classifications are updated. For decades, the diagnosis of diabetes has been based on glucose criteria, either the fasting plasma glucose (FPG) or the 75-g oral glucose tolerance test (OGTT). In 1997, the “expert committee on the diagnosis of diabetes” (the equivalent group to the DSM task force) revised the criteria on the basis of retinopathy assessed with new technologies and recommended lowering the cutoff point of FPG from 140 mg/dL to 126 mg/dL. Thus, even in the case of objective, measurable markers, frequent inspection and revision of criteria seem warranted. To further complicate matters, some authors insist that if the “gold standard” for medical diagnosis is defined as the “measure or precise equivalent that is verifiable,” autopsy and biopsy remain the ultimate “gold standards” in the field. This is the main reason why the “consensus of experts” continues to be frequently used as the “gold standard” in medicine, as many medical conditions are not consistently associated with an essential lesion that is universally accepted. Thus, in many instances, quoting from the neurologic textbook authored by A. Lerner in 2010, "diagnosis in medicine remains a deductive exercise."

Of course, the situation of psychiatric diagnosis is much more problematic, as the "symptoms" for most psychiatric syndromes consist of verbal reports and probes obtained directly from the patient, and the "signs" are the observed behaviors, detected through the traditional methods of anamnesis and mental status examination. In most instances, the current method for defining mental disorders involves a "willing suspension of disbelief," as is accepted, without conclusive proof, that the set of symptoms ("the syndrome") originate from a common source or that they are at least part of some type of unitary functional unit, as in the case of anxiety, depression, mania, psychoses, or dementia.

Advantages of the DSM System

The scientific/collegiate/consensus mechanism adopted by the APA (especially for the more recent versions of the DSM) was based on focused discussion by groups of experts under the umbrella of the organization. Since the APA, in general, has had good social and scientific standing, this mechanism has been a good practical solution throughout the years, both from the scientific and political points of view. The initial impact of clearly and succinctly defining key syndromes such as schizophrenia, panic anxiety, melancholic depression, manic-depressive illness, or dementia, backed by solid historical descriptions in the scientific literature, observations on heritability, and other features, supported the unitary character of the key syndromes. These were the clinical entities acknowledged by the St. Louis Group in the 1960s, whose existence and validity received a powerful confirmation with the therapeutic advances of the 1950s and 1960s, showing that there were pharmacologic agents that specifically targeted each of these syndromes and that apparently, "treatment-response" to a certain agent was a "marker" for a certain clinical condition. This was, of course, an illusion that gradually dissipated with the passage of time, following the accrual of evidence from randomized clinical trials and cumulative clinical experience.

Shortcomings of the DSM System

The most important intrinsic weaknesses attributed to the DSM include the following:

- The proliferation of new disorders in each new version of the DSM classification with only limited support in scientific evidence
- The problem of the arbitrary thresholds, for example, number of symptoms or duration in time required for a diagnosis
- The frequent need to use nonspecific diagnostic codes such as the “not otherwise specified” (NOS).
- The problem of comorbidity, as a significant number of patients qualify for a high number of DSM diagnoses simultaneously.

All these problems are intimately related and will be briefly discussed below.

As mentioned before, only a handful of diagnoses listed in the first two versions of the DSM have survived in recent editions, while the total number of diagnoses approaches 400 in DSM-5.

The main reasons for this explosion in the number of psychiatric diagnoses include the following:

- Breaking down or fragmenting well-accepted diagnostic entities, such as for example, “anxiety,” into large numbers of “subtypes” of decreasing validity. In order to create these new disorders, the detail of each diagnostic definition keeps growing, incorporating complex and arbitrary thresholds. With the increase in detail and restrictiveness, the clinical representativeness of each definition tends to diminish, and trivial disorders may be reified.
- The persistence and the creation in each DSM edition of “mental disorders” defined on the basis of their discordance with culturally sanctioned rules of behavior (such as personality disorders and a number of sexual disorders).
- The proliferation, in the absence of scientific evidence, of disorders defined by circumstances external to the individual, such as the addictive disorders, separated according to individual substances, behavioral addictions, or phobic disorders.
- The need felt by clinical practitioners to use vague, ambiguous codes, such as the NOS categories in DSM-IV, may arise to a large extent from the conflict between the increasing detail and restrictiveness of the DSM classification and the heterogeneity of clinical populations regarding number and assortment of symptoms, their duration or severity, and sequencing or fluctuation over time. Thus, the more “precise” and restrictive the definition of a mental disorder becomes, the lower the proportion of clinical patients that would meet the criteria

for that specific disorder and the larger the number of patients that may have to be allocated into nonspecific categories.

- Unfortunately, this increased detail of the definitions on paper is generally interpreted by administrators and data managers as equivalent to improved diagnostic ability in clinical practice, which is clearly not the case. An emerging reality is that insurers and data managers may end up rejecting nonspecific categories and demand the formulation of precise diagnoses at the risk of losing reimbursement. Moreover, arbitrary thresholds such as the 6-month duration rule may exclude many cases that would go on later to develop full-blown syndromes.

Increasing Complexity of the DSM Process

The first DSM was prepared by the “Committee on Nomenclature and Statistics of the American Psychiatric Association” and published in 1952. This first DSM listed 7 people as members of the “APA Committee on Nomenclature and Statistics” that produced the manual, and this remained unchanged for DSM-II (10 people, including 2 who were “consultants” and 1 “representing the council”). However, the explosion in the size and complexity of the DSM system starts with DSM-III. The number, functions, and size of committees increased exponentially, and the decision-making process was organized in various layers. The DSM-III included a “Task Force on Nomenclature and Statistics” with 19 members, 12 Advisory Committees (each targeting specific types of disorders) with up to 18 people each, as well as other committees such as the “Multiaxial Diagnosis Committee,” the “Glossary of Technical Terms Committee,” the “Assembly Liaison Task Force,” and the “Board of Trustees Ad Hoc Committee” plus numerous consultants. DSM-III, with its operational structure and descriptive criteria presented in a “Chinese menu” format, quickly became the predominant paradigm at a time of limited treatment resources, insufficient research, and scarce evidence about causality. An enhancement of communication was the paramount initial consideration at the time, and DSM-III seemed quite successful in accomplishing this goal, as attested by its worldwide use. Responding to the publication of DSM-III in 1984, Gerald Klerman, an influential psychiatrist, also a “neo-Kraepelinian,” stated categorically, “the problem of diagnostic reliability has been solved.”

The ensuing DSM-III-R simply broadened the criteria, and DSM-IV added a requirement for “Clinical Significant Distress or Impairment.”

DSM-5

The DSM-5 task force was officially convened in August 2007. It originally consisted of 26 members, 20 of them U.S. academic psychiatrists with high visibility in different areas of the specialty, one representative for psychiatry residents, one representative for patients, two psychiatrists representing APA governance (board and assembly), and one psychiatrist representing the World Psychiatric Association. As in the case of DSM-III and IV, the DSM-5 task force was the steering body for the whole process. The first step was to name several workgroups with responsibility for specific diagnostic areas, each led by a member of the task force. Each of these workgroups had, on average, about 10 members, as well as long lists of consultants. For the ensuing 5-year period (2008 to 2013), the task force and workgroups frequently met both in person and via teleconference, reviewing the evidence base related to current diagnoses, as well as that for potential new ones. The initial goal was to come up with a new system that capitalized on the gains of the “decade of the brain” with the tremendous progress made in basic neuroscience and the enthusiasm generated by the sequencing of the human genome. Because the risk for key disorders such as schizophrenia, bipolar, and major depressive disorder has a major familial component, it was hoped that by the time of the new DSM edition, specific genetic markers such as polymorphisms or structural abnormalities would have been identified, which would “tag” certain disorders and help to dissect them into meaningful phenotypes. Another expectation was that the increased research using objective tools such as structural and functional neuroimaging would yield reliable “markers” for several psychiatric disorders. Also, the initial hope was that DSM-5 would become a second “paradigm shift” for psychiatric diagnosis (DSM-III has been viewed as the first) by incorporating a dimensional approach as an alternative to a categorical one, considering new models of classifying psychiatric syndromes (e.g., brain “circuitry”), and hopefully decreasing rather than increasing the number of diagnostic categories in the manual. Upon gathering the task force and the workgroups and reviewing the available evidence, the unbridled enthusiasm on finding “markers” from neuroimaging or genetic studies and the possible use of novel strategies such as “brain circuitry” to classify mental disorders quickly dissipated and “reality” would soon set on the horizon. It thus became clear that the main task of DSM-5 should be an effort to enhance the current descriptive, operational approach by eliminating dubious syndromes, improving wordings and definitions, examining the evidence for new diagnostic entities, and the practical utility of a dimensional approach. As part of this process, the task force carefully examined previous versions of the DSM, particularly DSM-IV, to identify key areas in need of revision or elimination. There was awareness on the fact that in defining mental disorders, previous versions of DSM had not described or mapped natural phenomena accurately, implying—using the traditional analogy of carving a turkey—that the diagnostic system “did not carve nature at the joints.” It became clear that the multiple categories with doubtful boundaries forced clinical practitioners (at least the most scrupulous ones) to overutilize the “NOS” or “NOS” diagnoses. This was confirmed by a survey of general use, showing that the most common diagnosis made by psychiatrists in North America was “Depressive Disorder NOS.” The discussions also acknowledged that while DSM-III (and also DSM-IV) responded to a strong desire for inter-rater reliability by developing operational criteria, they arbitrarily treated every disorder as a category that was discontinuous from “normal” and from other disorders, thus avoiding dimensional approaches for descriptions of disorders. It was also felt that when applied rigidly, DSM-IV criteria appeared to limit research and therapeutic developments. Also, important components of key syndromes, such as cognitive problems in schizophrenia, were not emphasized in the previous DSM criteria and were largely ignored until National Institute of Mental Health (NIMH) took the initiative with the MATRICS project.

At about the same time as DSM-5 developments, the revision of extant categories of mental disorders was also being proposed for the ICD by leading psychiatrists from outside the United States. For example, the British psychiatrist David Goldberg highlighted that the three main problems of DSM-IV and ICD-10 classifications were the high rates of “comorbidity” produced by current rules, the increased use of nonspecific (NOS or NEC) diagnoses by practicing clinicians, and the fact that each edition was longer and more complex than the one preceding it.

The Process of Approving/Disapproving Changes or New Diagnoses in DSM-5

In contrast with the simple structure of the first DSM, in which one committee made all the decisions, DSM-5 had to deal with several levels or layers for clearance and approval of the new proposals. The first premise was that any changes to DSM-IV or proposals for new diagnoses had to be evidence based. Thus, on top of the already complex structure of the DSM-III and DSM-IV, a consulting body charged with veto power was added. Instead of going directly to the task force, committee recommendations went to an “independent” body called the “Scientific Advisory Committee” (SAC), created by APA leadership as an important step to review the scientific justification for proposed changes and assure the rigor and consistency of the DSM review process. This committee functioned outside the DSM organizational structure and reported, in an advisory role, to the APA President and Board of Trustees overseeing the DSM process. A second committee, called the “Clinical Public Health Committee” or CPHC, also named by APA leadership outside DSM’s organizational structure, reviewed proposals keeping in mind potential practical or public-health implications of proposed changes and new diagnoses and also made specific recommendations. The CPHC used a large list of advisors from academia and clinical practice as “peer reviewers.” The SAC and CPHC recommendations then went to the full task force for final debate and a decision on whether or not new diagnoses and significant changes in existing diagnoses should be implemented. These final recommendations were then forwarded to the APA Board of Trustees for their review. The Board had the final say about changes, revisions, and new diagnoses that were printed in the DSM-5 manual. The number of people involved in the DSM-5 process was far higher than that of previous editions. More importantly, about 30% of workgroup members were international colleagues from countries outside the United States.

DSM-5 Field Trials

These trials specifically examined the reliability of DSM-5 diagnostic changes (e.g., depression, post-traumatic stress disorder [PTSD], schizophrenia) and new diagnoses (e.g., hoarding disorder, somatic symptom disorder). The trials were rigorously designed and controlled. People with specific disorders were matched to people without the disorders and examined independently by two clinicians who did not know the patient or their diagnosis on two separate occasions. These trials took place at several academic centers in the United States and included both children and adult populations. As the major purpose of these trials was to assess reliability, a set of expectations was formed based on revisions of the literature on current reliability estimates in the medical field. The key measure of reliability was the intraclass kappa coefficient, a measure of how well a first diagnosis predicts a second independent one. A literature review by Helena Chmura Kramer, PhD and collaborators in 2002 had noted that $\kappa > 0.8$ would be an “extraordinary” result; κ 0.6 to 0.8, an “excellent” result; κ 0.4 to 0.6, a “good” result; κ 0.2 to 0.4, an “acceptable” result, and $\kappa < 0.2$ or lower, an “unacceptable” result. This review found “excellent” to “good” reliabilities with such procedures as coronary arteriography, x-rays, inspection of conjunctives for anemia, and liver scans. Clinical assessments such as the “Lasègue sign” and estimations of reliability to judge “quality of medical care” tended to have only an “acceptable” level of reliability, while procedures such as pelvic examinations in the emergency room (ER) had “unacceptable” levels of reliability.

Table 9.1–1.
Reliability of Major DSM-5 Diagnoses

Disorders	Kappa
PTSD	0.67
Neurocognitive disorder	0.78
Somatic symptom disorder	0.61
Hoarding disorder	0.59
Bipolar	0.56
Schizophrenia	0.46
Major depression	0.27
Mixed anxiety depression	−0.004

PTSD, Post-traumatic Stress Disorder.

Results of the DSM-5 field trials showed “excellent” agreement for major neurocognitive disorder, PTSD, somatic symptom disorders, autism spectrum disorder (ASD), and most of the self-rated “cross-cutting” symptom-based measures. Schizophrenia and bipolar disorder showed “good” reliability. Major depression had a disappointingly low level of reliability, and subtypes such as “anxious depression” were completely unreliable. These field trials were much more rigorous than those of DSM-III and DSM-IV and gave confidence in the robustness of most of the changes and the few proposed new diagnoses that were tested (Tables 9.1–1 and 9.1–2).

Key Changes in DSM-5

The DSM-5 Manual. The Official Manual of the fifth edition of the DSM has 947 pages. The manual is divided into three main sections.

Table 9.1–2.
Reliability of Dimensional Diagnoses and Crosscutting Symptom Measures

	Kappa
Dimensional Measures	
Autism spectrum disorder	0.69
New (alternate) personality disorders	0.41
Crosscutting Symptom Measures	
Psychosis	0.80
Depression	0.79
Anxiety	0.75
Somatic symptoms	0.73

Section I includes “the DSM-5 basics” such as introduction, use of the manual, and a cautionary statement about forensic use of the manual.

Section II, the key section of the manual, includes all the diagnostic criteria and codes for the officially approved diagnoses.

Section III includes new diagnoses not ready for “prime time,” an alternate category of personality disorders, instrumental tools for dimensional measurements, as well as an instrument for assessing functional disability (WHODAS).

There is also an “Appendix” that includes highlights of changes from DSM-IV to DSM-5, a glossary of technical terms, cultural concepts of distress, and an alphabetical listing of DSM-5 diagnoses and ICD 9-CM and ICD-10 codes.

Deletion of the Multiaxial System. The five-axis frame of DSM-III and IV was eliminated in DSM-5. This makes DSM-5 closer to the structure of the ICD diagnostic system that has traditionally used a single axis. However, regarding the number of diagnoses or their quality, this reduction in the number of axes may have been inconsequential, as the personality disorders, one of the most questioned diagnostic groups, which had their own axis, Axis II, in DSM-IV, was kept unchanged and became one of the disorders listed in the main section (section II) of the Manual.

Dimensional Approach. In the introduction to DSM-5, it was stated that on the basis of early DSM-5/ICD-11 collaborations in diagnostic developments, clustering of disorders into "externalizing" versus "internalizing" perspectives represented an empirically supported framework. While strongly insinuated early in the process, a true dimensional perspective did not actually surface in DSM-5. Perhaps, the only noncategorical approach in DSM-5 is the "crosscutting" symptom dimensions (depression, anxiety, somatic symptoms, and others) that tend to be distributed across many diagnoses. These are based on self-rated measures, and a majority are taken from the Patient-Reported Outcome Measurement Information System (PROMIS), developed under the sponsorship of NIH. These are dynamic tools used to measure health outcomes from the patient's perspective. They include symptom dimensions for pain, fatigue, physical functions, depression, anxiety, social function, etc., and are available for children and adults. These symptom dimensions are included in section III of the manual, with the hope that clinicians may include them in the clinical assessments.

Structure and Grouping of Disorders. Major changes in DSM-5 included the removal of PTSD and acute stress disorders from the anxiety disorders category in DSM-IV and their placement in a new stand-alone chapter called Trauma and Stress-Related Disorders (TSRD). OCD was also moved out of the anxiety disorders chapter and also placed in a new chapter called Obsessive-Compulsive and Related Disorders. This group also includes "body dysmorphic disorder," "trichotillomania" (hair-pulling disorder), a new diagnosis “hoarding disorder,” as well as excoriation or “skin-picking” disorder.

New Disorders. Key new disorders that were included in the main section (section II) of DSM-5 were “Autism Spectrum Disorder,” “Somatic Symptom Disorders,” “Hoarding Disorder” (a diagnosis of major public health significance, as every department of health in the country has had to deal with the hoarding issue, either of animals or things), “Skin-Picking Disorder,” “Intellectual Developmental Disorder” (formerly Mental Retardation), “Disruptive Mood Dysregulation Disorder,” “Mild Cognitive Impairment,” “Insomnia Disorder,” “Nightmare Disorder,” and “Suicide Disorder.” New Disorders, judged not to be ready for “prime time,” were included in section III of DSM-5. These include the Alternate Category of Personality Disorders, Persistent Complex Bereavement Disorder, Caffeine Use Disorder, Internet Gaming Disorder, Neurobehavioral Disorder Associated with Prenatal Alcohol Exposure, Suicidal Behavior Disorder, and Nonsuicidal Self-Injury.

Neurodevelopmental Disorders. Mental Retardation was renamed Intellectual Disability (Intellectual Developmental Disorder). ASD replaces Autism, Asperger Syndrome, and Pervasive Developmental Disorder NOS in DSM-IV.

Neurocognitive Disorders. These disorders were previously referred to in DSM-IV as “dementia, delirium, amnestic, and other cognitive disorders.” In DSM-5, they are headed by “delirium.” A new category in DSM-5 is “mild neurocognitive disorder,” added as a lesser form of cognitive deficit, to contrast it with “major cognitive disorder” and the “major–minor” dichotomy is used for the remaining disorders in this category, including “major or minor neurocognitive disorder due to Alzheimer disease,” “major or mild frontotemporal neurocognitive disorder,” “major or mild neurocognitive disorder with Lewy bodies.”

Schizophrenia Spectrum and Other Psychotic Disorders. Included under this heading are “schizotypal personality disorder,” “delusional disorder,” “schizophreniform disorder,” “schizophrenia,” “schizoaffective disorder,” “substance/medication-induced psychotic disorder,” “psychotic disorder due to another medical condition,” and “catatonia,” including several subtypes. In the case of DSM-5 schizophrenia, all subtypes were eliminated, and criteria for the disorders were slightly modified. This included removal of special consideration for “bizarre” delusions and “special” hallucinations under criterion A (characteristic symptoms). Rewording of negative symptoms in efforts to provide more clarity (“affective flattening” was changed to “restricted affect”; “alogia or avolition” was changed to “avolition/asociality”). Also, in order to meet criterion A for the disorder, at least one of two required symptoms must be delusions, hallucinations, or disorganized speech.

Bipolar and Related Disorders. This group includes “bipolar I disorder,” “bipolar II disorder,” “cyclothymic disorder,” “substance/medication-induced bipolar and related disorders,” “bipolar and related disorder due to another medical condition,” “other specified bipolar and related disorders,” and “unspecified bipolar and related disorder.” In the case of bipolar disorder, the inclusion of “increased energy/activity” as a criterion is a symptom of mania/hypomania. Also, the addition of “mixed features” specifiers for mania, hypomania, and depression (see below). The complete list of “specifiers” for bipolar and related disorders include “anxious distress,” “manic or hypomanic episode with mixed features,” “depressive episode with mixed features,” “with rapid cycling,” “with melancholic features,” “with atypical features,” “with psychotic features,” “with catatonia,” and “with seasonal pattern.”

Depressive Disorders. Included here is “Disruptive Mood Dysregulation Disorder,” a new disorder intended to decrease the “excess” of diagnoses of bipolar disorder observed in children and adolescents. “Major Depressive Disorder” remains virtually identical to the DSM-IV category, except for the elimination of the bereavement exclusion (this was done, despite the statement under the definition of mental disorder in the introduction to DSM-5—see our discussion regarding the definition for a mental disorder earlier in this chapter). “Persistent Depressive Disorder (Dysthymia)”; “Premenstrual Dysphoric Disorder”; “Substance/Medication-Induced Depressive Disorder”; “Depressive Disorder Due to Another Medical Condition”; “Other Specified Depressive Disorder”; and “Unspecified Depressive Disorder.” Also, there is a list of “specifiers,” including a new one, “with anxious distress,” in addition to “with mixed features,” “with melancholic features,” and “with atypical features.”

Anxiety Disorders. With the transfer of PTSD and OCDto stand-alone groupings, the disorders remaining in this category are “separation anxiety disorder,” “selective mutism,” “specific phobia” (with several subtypes), “social anxiety disorder,” “panic disorder,” “agoraphobia,” “generalized anxiety disorder,” “anxiety disorder due to another medical condition,” “other specified anxiety disorder,” and “unspecified anxiety disorder.”

Trauma and Stressor-Related Disorders. “Post-traumatic Stress Disorder (PTSD)” and “Acute Stress Disorder” were moved out of the anxiety disorders group (where they were in DSM-IV) to this new, stand-alone category. The criteria for PTSD were carefully revised. In DSM-IV, PTSD included three clusters of symptoms: (1) reexperiencing fear memories (intrusive memories), (2) avoidance of trauma reminders (amnesia, withdrawal, avoidance of situational reminders), and (3) hyperarousal symptoms (sleep disturbance, heightened arousal response). In DSM-5, the definition of a “traumatic event” was clarified. One extra criterion, D3, was added (persistent negative emotional state). Criteria C now focuses on “avoidance,” while criteria D focuses on negative cognitions and mood. Language and descriptions were enhanced for precision. Basically, DSM-5 has expanded the definition of PTSD beyond the “fear” construct, adding another cluster of symptoms to acknowledge negative alterations in cognition and mood (pervasive negative emotional states, negative appraisals about oneself or the world, and excessive blame). In total, DSM-5 increased PTSD symptoms to 20, up from 17 in DSM-IV. Potential combinations of symptoms got increased more than 10-fold. According to some experts, this makes the disorder more heterogeneous and may complicate the research agenda. “Adjustment Disorders” are also listed under TSRD together with “Reactive Attachment Disorder” and a new condition, “Disinhibited Social Engagement Disorder.” The latter emphasizes the recognition of symptoms associated with neglect during childhood. Regarding the effects of these changes in clinical practice, experts have stated, “there is potential for improved assessment of stressor-related conditions in children.”

Substance Use Disorders. Substance abuse and substance dependence were consolidated into a single disorder called “substance use disorder.” A continuum or dimension was created that includes “mild,” “moderate,” or “severe” substance use. One of the DSM-IV abuse criteria was removed (“legal consequences such as multiple arrests”), and a new criterion, “craving,” was added.

Somatic Symptom and Related Disorders. This new category replaced the somatoform disorders category from DSM-IV. All diagnoses that had been subsumed under the somatoform disorders category in DSM-IV were merged into a common category, significantly revised, or placed in another section of the diagnostic manual. The DSM-IV requirement for symptoms to be “medically unexplained” was removed from DSM-5. Somatization Disorder, Pain Disorder, and Undifferentiated Somatoform Disorder were all merged into a new diagnosis called “Somatic Symptom Disorder.” Hypochondriasis became “Illness Anxiety Disorder.” “Conversion Disorder” was significantly revised, and the term “Functional Neurological Symptom Syndrome” was added to the title of this disorder in parentheses. “Psychological Factors Affecting Other Medical Conditions,” “Factitious Disorder,” “Other Specified Somatic Symptom and Related Disorder,” and “Unspecified Somatic Symptom and Related Disorders” were the new disorders adopted in the new category of somatic symptom disorders. “Body Dysmorphic Disorder” was moved to the new obsessive-compulsive and related disorders group. In terms of specific criteria, Somatic Symptom Disorder requires one or more somatic symptoms to be present, but unlike DSM-III and IV, in DSM-5, the emphasis is placed on the behavioral and psychological elements of the syndrome, such as preoccupation and concern with the physical symptoms. It is also possible for Somatic Symptom Disorder to have special features (e.g., predominantly pain symptoms) or different severity dimensions, depending on the number of A and B criteria present beyond the threshold for diagnosis.

Childhood Disorders. A single spectrum diagnosis, ASD, replaces “Autistic Disorder,” “Asperger Disorder,” and “Pervasive Developmental Disorder not otherwise specified.” ASD is now differentiated from social communication disorder by the presence of the restrictive and repetitive behavior criterion. Addition of “specifiers” for all neurodevelopmental disorders “associated with known medical or genetic condition or environmental factors.” As stated by Michael Rutter, a leading expert in this area, “it was necessary to abandon the DSM-IV classification because it had been ignored or was sidestepped by clinicians and researchers.”

Personality Disorders. These remained unchanged in DSM-5, as the original DSM-IV personality disorders were all retained in section II of the DSM-5. However, an “alternate” category of Personality Disorders has been included in section III. This consists of a trait-based diagnostic approach that separates functional impairment (A criteria) and the expression of characteristic traits (B criteria) for the six personality disorders that were tested in the field trials. A more flexible trait-specified profile can now be used for patient presentations that do not fit well the six subtypes. A similar strategy was adopted in ICD-11.

Sleep–Wake Disorders. This group includes a large number of categories and subcategories, starting with Insomnia Disorder, Hypersomnolence Disorder, Narcolepsy (various subtypes), as way as a long list of “Breathing-Related Sleep Disorders” (Obstructive Sleep Apnea–Hypopnea, Central Sleep Apnea, Idiopathic Central Sleep Apnea, Cheyne–Stokes Breathing, and other subtypes) and “Parasomnias” (including nonrapid eye movement sleep arousal disorders, Nightmare Disorder, Rapid Eye Movement Sleep Behavior Disorder, Restless Leg Syndrome, and other subtypes).

Other Conditions that May Be a Focus of Clinical Attention. This includes a long list of headings and subheadings, a majority tagged with “V” codes. The main headings are “Relational Problems,” “Abuse and Neglect,” “Child Physical Abuse,” “Child Sexual Abuse,” “Child neglect,” “Child Psychological Abuse,” “Adult Maltreatment and Neglect Problems,” “Educational and Occupational Problems,” and several others adding to a total of 133 separate codes, making up about one-third of the total categories in DSM-5.

Specific Areas of Concern in DSM-5

Proliferation of Mental Disorders. The most common public criticism of later editions of the DSM manual has to do with the growth in both the number of mental disorders and the scope of their definitions. The expressed or tacit implication of this is that it implies a growing and unjustified medicalization of multiple areas of daily life. As mentioned above, the number of disorders included in the first edition of the manual (DSM-I) has been multiplied at least fourfold in the subsequent editions. Reviewing the various editions of the DSM, it is obvious the growth in the number of disorders does not reflect the inclusion of many new disorders, but that it is due mainly to the subdivision or “splitting” of traditional disorders such as anxiety, affective, addictive, and psychotic disorders into countless subtypes (Table 9.1–3). The numbers speak eloquently: DSM-5 lists 438 diagnostic codes, while DSM-IV had 361, and DSM-III 265. However, the number of specific categories in DSM-5 is smaller than those in DSM-IV.

Table 9.1–3.
The Increase in the Number of Diagnoses between the 1952 DSM I and DSM-5: The Case of Anxiety Disorders

DSM I 1952	Includes anxiety reaction and six related “reactions,” dissociative, conversion, phobic, obsessive-compulsive, depressive, and psychoneurotic. **(Total** = **6 Disorders)**
DSM-II	Eliminates the qualification of “reaction” and includes all of the anxiety-related diagnoses from DSM I, and some related new ones, such as hypochondriacal, neurasthenia, or depersonalization, in the general category of “neuroses.” **(Total** = **8 Disorders)**
DSM-III	The category anxiety disorders appears for the first time, including a total of 10 disorders in four groups: ▶ Phobic disorders (including agoraphobia with panic attacks, agoraphobia without panic attacks, social phobia, simple phobia) ▶ Anxiety states (including panic disorder, generalized anxiety disorder, obsessive-compulsive disorder) ▶ Post-traumatic stress disorder (PTSD; including acute PTSD, chronic PTSD) ▶ Atypical anxiety disorder To the above disorders, two categories should be added as they were included in DSM-II among the anxiety neuroses. They are: ▶ Somatoform disorders (including somatization, conversion, psychogenic pain, hypochondriasis, and atypical) ▶ Dissociative disorders (including psychogenic amnesia, psychogenic fugue, multiple personality, depersonalization, and atypical dissociative) In addition, several ambiguously defined “adjustment disorders” either expressly or implicitly are merged into the anxiety disorders category. **(Total** = **21 Disorders)**
DSM-IV-R	The category anxiety disorders includes, besides generalized anxiety disorder, eight specific anxiety disorders, and two unspecific anxiety disorders; to these, the following added: five specific and two unspecific somatoform disorders, and

	four specific and one unspecific dissociative disorders and again, the adjustment disorders **(Total** = **23 Disorders)**
DSM-5	Anxiety disorders include 10 specific and 2 unspecific disorders (the first 2 were previously accepted only for children): ▶ Separation anxiety disorder ▶ Selective mutism ▶ Specific phobia ▶ Social anxiety disorder (social phobia) ▶ Panic disorder ▶ Panic attack specifier ▶ Agoraphobia ▶ Generalized anxiety disorder ▶ Substance/medication-induced anxiety disorder ▶ Anxiety disorder due to another medical condition ▶ Other specified anxiety disorder ▶ Unspecified anxiety disorder Obsessive-compulsive gets its own group with seven specific disorders and two unspecific categories. Trauma gets its own group (trauma- and stressor-related disorders) which includes reactive attachment disorder and disinhibited social engagement disorder (previously used only in children), PTSD, acute stress disorder, adjustment disorders (including five specific and one unspecific disorders), and two unspecific categories. Somatic symptom and related disorders loses pain disorder and body dysmorphic disorder (which migrates to obsessive-compulsive) but still lists five specific disorders and two unspecific categories. Dissociative includes three specific disorders and two unspecific categories. **(Total** = **More than 45 disorders)**

Personality Disorders. From the perspective of medicine, the concept of “personality” has always been a rather vague sociocultural construct, hard to define even in psychological terms, and thus quite alien to the medical sciences. Indeed, most of the “personality disorders” are value judgments regarding eccentricities, peculiarities, or extremes in the behavior of an individual when they are compared with culturally predominant norms. The personality disorders that remain in DSM-IV and DSM-5 (there were no changes in the latest edition) largely represent such “value judgments,” often designated using critical, condemnatory, or pejorative language. Since DSM-III, personality disorders have remained virtually untouched except for the elimination of the multiaxial system in DSM-5 and the proposal for an “alternate category” with dimensional rather than categorical overtones that was relegated to section III of the DSM-5. This alternative category is being evaluated and shows promise.

Sexual Disorders. In parallel with the personality disorders, the inclusion of certain “sexual disorders” in a section originally headed by homosexuality has also been a point of contention within the DSM. “Sexual deviations” were included among the personality disorders in the first DSM and remained there in DSM-II. From DSM-III on, sexual disorders were listed under a separate group and have continued to proliferate in each new edition of the manual. Thus, DSM-5 added several new disorders, a number of them among the paraphilias, a particularly problematic set of disorders whose definition may even conflict with the definition of a mental disorder as included in section I of the DSM-5.

The Media Storm Surrounding DSM-5. Neither the task force and workgroups nor the APA, in general, did presage the avalanche of criticisms, complaints, passions, and lobbying unleashed during the DSM-5 process, starting soon upon release of the initial plans to revamp DSM-IV and continuing until the release of the printed manual. Nothing like this had happened (at least not to this extent) in the case of DSM-III, DSM-IV, or even ICD-10. A notable observation was the broad news media coverage that went far beyond the United States and Canadian boundaries, a clear proof of the national and international interest generated by the DSM classification and psychiatric diagnosis in general.

Global Issues Related to Diagnosis

Although the DSM system is primarily designed for use in North America, it is also used in many other countries of the world, particularly for research purposes. Starting with DSM-5, international input was actively sought, both in the preparatory phase and also as part of the workgroups. Critiques related to DSM psychiatric diagnoses include their "globalization." For example, attention deficit hyperactivity disorder (ADHD), a disorder first recognized in the United States in the 1960s, spread to the United Kingdom, Germany, France, Italy, and Brazil, possibly influenced by the transnational pharmaceutical industry, the role of the internet, and activism of advocacy groups. A relevant example is the one provided by Professor Sing Lin from Hong Kong, a member of the somatic disorders work group in DSM-5. It relates to what has happened with the diagnosis of Anorexia Nervosa in Hong Kong. In the 1980s and early 1990s, the disorder was very rarely diagnosed, as patients only had somatic complaints ("bloated stomachs"). In 1994, a teenage anorexic girl collapsed and died in a busy downtown street in Hong Kong. This quickly caught the attention of the media, who become aware of the U.S. diagnostic manuals, and reported that anorexia nervosa in Hong Kong was the same disorder as that in the United States and Europe. Obviously, the disorder began to appear with increasing frequency, the major complaints of sufferers shifting from somatic complaints to "fat phobia."

Dealing with Conflict of Interest

A most salient public criticism of DSM-IV and DSM-5 has been the potential "conflict of interest" of those involved in the process of revising and preparing the diagnostic manuals. While having ties with the pharmaceutical industry is the most frequently mentioned conflict of interest, there are several other potential conflicts, such as ties to patient advocacy groups, health insurers, organizational memberships, and other groups that may have a vested interest in the definition of mental disorders. These latter conflicts are rarely addressed. The inclusion or elimination of a mental disorder or a change toward a more restrictive or comprehensive definition of a disorder, all affect the size of the population covered (most commonly, increases it), and with it, the work opportunities for professionals, the potential income for drug companies, and the potential profit for public or private health insurers all grow larger. Also, given the financial benefits accorded by several diagnoses such as the coverage/subsidies for educating autistic children in several states, the worker compensation laws, and the veterans administration service-connected disabilities, it is not surprising that the redefinition of disorders such as autism in DSM-5 led to a heated debate not only in mental health journals but also in the press and mass media, with organizations like

“Autism Speaks” publishing the names and affiliations of members of the work force so that they could be lobbied. Another vested interest implicit in the DSM diagnostic process is obviously that of the APA, the owner of the copyrights for DSM. The APA has a large stake in the commercial success of a manual that has gradually become one of its main sources of income. Creating expert committees completely devoid of conflicts of interest seems quite unrealistic, and it may be an impossible task. However, at a minimum, a “transparency policy” for full disclosure of all possible conflicts of interest should be ensured, and fortunately, it has become the standard.

Changes in Medical Practice

Another factor that may play a role in the proliferation of disorders is the increased access of populations to health information in the age of the internet, resulting in a growing medicalization of daily life. The increased access to treatment occurs all across medicine, and the tendency of people to look for treatment for conditions that were previously nondiagnosable or viewed as mild or even trivial is now a worldwide phenomenon. The thresholds for treatment for elevated cholesterol, blood glucose, or blood pressure seem to be continuously getting lower (e.g., the uniform 100 threshold rule). A growing number of individuals seek the genetic or pharmacologic answers for obesity, a condition that some years ago was not even considered to be a disease. In addition, many individuals currently seek treatment not because of any defect or illness, but for cosmetic or hedonic reasons, not restricted to cosmetic surgery, such as the growing number of males with normal sexual function who ask for prescriptions of phosphodiesterase inhibitors in order to have more frequent sex (the so-called “recreational use”), or patients with no attention disorders who ask for prescriptions of stimulants to gain an edge at school or work.

Alternatives to Traditional Classifications

The DSM-5 task force discussed a number of alternatives to the DSM-IV classification. One initial model was that proposed by Dennis Charney, a biologically oriented psychiatrist, then at NIMH, that suggested a multiaxial futuristic diagnosis, in which Axis I represented Genotype, Axis II, Neurobiologic Phenotype, Axis III, Behavioral Phenotype, Axis IV, Environmental Modifiers or Precipitants, and Axis V, Treatment Targets and Therapeutic Response. While this model appeared quite promising and scholarly, the scientific evidence and the needed “markers” remain elusive.

Brain Circuitry. Progress in basic neuroscience during the decade of the brain has been unraveling, with good precision, brain circuits related to specific behaviors or emotions. Emerging techniques such as optogenetics nurture the “circuit dynamics of psychiatric disease,” receiving ample coverage in medical journals (JAMA) and the news media (The New Yorker). While most of this basic research is in the preclinical phase (animal models), the circuit-level perspective makes it possible to generate hypotheses on how the effect of multiple genes may lead to specific psychiatric syndromes. Some of the main circuits now under scrutiny include the following:

1. A “frontothalamic-striatal circuitry” that may be involved in cognitive control of behavior and thought to be of relevance to schizophrenia, ADHD, addictions, and impulse control disorders.
2. A “fear circuitry,” with neuronal connections based on the amygdala, involved in conditioned fear, of possible relevance to anxiety and depressive disorders.
3. A “reward circuitry” consisting of mesotelencephalic dopaminergic projections that may be relevant to addictions, impulse control disorders, schizophrenia, as well as the hedonic component of mood disorders.
4. A “regulation of mood states” circuitry involving the medial prefrontal cortex, amygdala, and hypothalamus, and possibly relevant to depression, mania, and mood disorders in general.

A key question posed during the development of DSM-5 was whether or not at that time, the specialty (psychiatry) and practicing clinicians (psychiatrists, psychologists, social workers, and others) were ready for classifications based on hypothetical pathophysiology, lacking solid, consistent “markers” and the quick answer was “not quite yet.”

NIMH Research Domain Criteria. The RDoC, launched in 2009, was created in response to the growing awareness of challenges arising from the use of categorical diagnostic systems in research on mental illnesses. The main purpose was to create an alternative system for investigating mental disorders by focusing on disruptions in normal adaptive functions rather than traditional symptom-based diagnoses. The development of the initial version of the RDoC framework was the collaborative effort of over 200 leading scientists and was influenced by decades of behavioral and neuroscience data.

The RDoC matrix was conceived as a tool for use by researchers to help them structure their study designs around behavioral or cognitive concepts.

The current version of RDoC contains six overarching systems or domains that influence human functioning. These are:

- Negative Valence Systems
- Positive Valence Systems
- Cognitive Systems
- Social Processes
- Arousal and Regulatory Systems
- Sensorimotor domain

BRIEF HISTORY OF RDOC. This alternative diagnostic approach was developed by Bruce Cuthbert and Thomas Insel at NIMH, and curiously, it first came into play in the NIMH Director's blog, coinciding with the final release of DSM-5 in 2013. It quickly spread to the news media that "NIMH had disapproved DSM-5," at the same time that it (NIMH) urged (or mandated?) as a condition for future funding, the use of RDoC, their alternative and complex strategy for classification that at the time appeared quite incipient indeed. As described in the NIMH website, RDoC is conceived as a dimensional system (reflecting, brain circuit-level measurements, behavioral activity, etc.) spanning the range from normal to abnormal. As it happens with dimensions in medicine such as hypertension or levels of cholesterol in plasma, this approach has to define cutoff points for the definition and extent of psychopathology (e.g., mild, moderate, and severe). It is of interest to observe that the NIMH document acknowledges that "to the extent that DSM-5 introduces dimensions in addition to classes, the crosswalks to RDoC dimensions may be enhanced."

However, RDoC does not adhere to current DSM categories, as its intent is "to generate classifications stemming from basic behavioral neuroscience." Rather than starting with an illness definition and seeking its neurobiologic underpinnings, RDoC begins with brain–behavior–brain relationships and links them to clinical phenomena based on several units of analysis such as neuroimaging, physiologic activity, behavior, and self-reports of symptoms. One of the key goals stated for RDoC is to allow researchers to choose independent variables from among several different units of analysis. The RDoC framework is explicitly agnostic with respect to current definitions of disorders. For instance, depression as a clinical syndrome has been related to abnormal activity in the amygdala, anterior cingulate cortex, nucleus accumbens, and multiple monoamine systems, while also strongly comorbid with multiple anxiety disorders, eating disorders, etc. The idea is that studying the individual mechanisms may lead to a better understanding of current disorders, or perhaps new and novel definitions of disorders, but in either case, improved information about treatment choices.

The RDoC research framework is described as a matrix whose rows correspond to specified dimensions of function, called "constructs," that represent the fundamental unit of analysis in this system. The hope is that these functional behavioral dimensions and related genes and circuits would be continually refined as science advances. The expectation is that most research studies will focus on a single "construct" or compare two separate "constructs" on relevant measures. "Constructs" are grouped into "domains of functioning." These include "negative and positive valence systems," "cognitive systems," "systems for social processes," and "arousal/regulatory systems." The various columns in the matrix represent different classes of variables (or units of analysis) used to study the domains/constructs. These variables include genes, molecules, cells, neural circuits, physiologic elements (e.g., cortisol, heart rate, startle reflex), behaviors, and patient self-reports. Neuronal or brain circuits represent the core of these variables. The inclusion of a "construct" in this system depends on the specification that a particular brain circuit or brain area implements that dimension of behavior. It is also acknowledged in the document that given the complexity of the brain and behavior, there can be various degrees of ambiguity in this regard.

"Constructs" included in the matrix are based on current literature reviews that have provided a neurobehavioral research base for each of the entries. The document also emphasizes that "these particular domains and constructs are simply starting points that are not definitive or set in concrete. It is expected that these will change, dynamically, with input from the field, and as future research is conducted.” A major aim of RDoC is to create a framework for grouping participants in research studies in order to create foundational research literature that informs future versions of nosologies based upon genetics and behavioral neuroscience. RDoC is not intended for clinical diagnosis at the current time. In the future, research supported by RDoC could inform diagnostic approaches using new laboratory procedures, behavioral assessments, and novel instruments to provide enhanced treatment and prevention interventions. It is also hoped that RDoC will support the enhanced development of new pharmacologic and psychosocial interventions based upon neurobiologic and behavioral mechanisms. The intent of the RDoC is to accelerate the pace of new discoveries by fostering research that translates findings from basic science into new treatments addressing fundamental mechanisms that cut across current diagnostic categories.

In a release, following the DSM-5/RDoC controversy, NIMH Director Thomas Insel argued that the RDoC project is a move forward in “precision medicine.” He stated that “it is possible that specific mutations in relevant brain circuits can be found that explain some cases of schizophrenia, bipolar disorder, or autism, just as mutations in the tumor explain cancer, but that precision medicine for mental disorders will not come from a single genomic glitch. Rather, like many other areas of medicine, many genes each contribute only a small amount of vulnerability as part of an overall risk profile that includes life experiences, neurodevelopment, and social and cultural factors. RDoC assumes that many kinds of data are needed to reach precision, a situation akin to triangulating in order to find your position on a map. These data will draw from many sources, including symptoms, genotype, physiology, cognitive assessment, family dynamics, environmental exposures, and cultural background.”

However, to be clear, this dimensional approach represents a dramatic shift in the classification of psychiatric disorders. It theorizes those behavioral consequences of brain functioning, such as for example, working memory or reactivity of the amygdala, as measured during tests such as functional magnetic resonance imaging (fMRI), would define clinical characteristics that reflect with greater precision the biology of the illness, possibly leading to more specific therapeutic targets compared to the broad overlapping categories of the DSM.

Despite the authority backing it, the initial reception of this new system was mixed among leading researchers, many remaining skeptical about the RDoC approach. According to some of the experts, the existing evidence makes the RDoC approach questionable. Two well-known psychiatry researchers argued that “patients with ADHD and patients with psychosis may show similar abnormalities in frontal brain function just as patients with PTSD and patients with psychosis may show similar overactivation of the amygdala in fMRI images, but the underlying reasons for these superficial similarities are different, and so are the treatments. The treatment of frontal lobe dysfunction in ADHD (e.g., stimulants) will make psychosis worse, just as the treatment for amygdala overactivation in psychosis (antidopaminergic drugs) is likely to have adverse effects in patients with PTSD.”

The DSM-5/RDoC controversy was at least temporarily settled after the NIMH director and the President of the APA produced a joint statement stating that “all medical disciplines advance through research progress in characterizing diseases and disorders. DSM-5 and RDoC represent complementary, not competing, frameworks for this goal.” That statement also highlighted that “for practicing clinicians, DSM-5 represents the best available evidence at this time.”

RDOC RECENT UPDATES. RDoC continues to be used in research sponsored by NIMH, and the institute has recently released some updates to its RDoC framework. The first of these updates, taking place in mid-2018, reorganized the positive valence domain. Since the time of its creation, RDoC was envisioned as a dynamic and continually developing framework. Now, the RDoC unit has unveiled changes to the RDoC framework—a reorganization of content within the positive valence domain. This domain includes systems primarily responsible for responses to positive motivational situations or contexts, such as reward seeking, consummatory behavior, and reward or habit learning. The revisions were spearheaded by the RDoC Matrix (CMAT) Workgroup—a group made up of members of the Nationally Advisory Mental Health Council, scientific experts, and NIMH staff. The main alterations in the positive valence domain include changing some of the constructs listed in the original version to subconstructs in the new version (e.g., Initial Responsiveness to Reward) and the addition of new constructs (e.g., Probabilistic and Reinforcement Learning). The reorganization is intended to make the constructs more straightforward and to align them more closely with computational and animal model perspectives, as well as data from human neuroimaging studies. The updates to the Positive Valence domain are the newest example of RDoC as a dynamic research framework—one that evolves to incorporate the latest findings from the research community and provide exemplars for future research.

A second update, dated January 2019, involved the addition of a sensorimotor domain to the RDoC Framework. The aim of this new sensorimotor domain is to help foster earlier and more precise identification of the role of motor systems disruptions in psychopathology and aid in the development of more effective treatments for people who are affected by these disruptions.

The new sensorimotor domain consists of four constructs (Motor Action, Agency and Ownership, Habit, and Innate Motor Patterns), the first of which includes five subconstructs. Detailed definitions of these domains and constructs can be found on the RDoC page of the NIMH website.

This change to the framework was again spearheaded by the CMAT Workgroup. The addition of the new domain is intended to encourage a systematic approach to the study of motor function in psychopathology, which will hopefully lead to better identification of, and treatment for these disruptions.

The addition of the sensorimotor domain is touted as the latest example of RDoC as a dynamic research framework—one that evolves to incorporate the latest findings from the research community and provide exemplars for future research.

The RDoC group has also provided details on specific constructs/subconstructs under each domain and listed well-referenced specific tasks (or tests) to assess each of them. This should be very helpful for researchers using the RDoC. For example, for assessing the “acute threat” construct of the negative valence system, a number of potential tasks have been detailed (Trier Social Stress Test, Behavioral Approach Test, Cold Pressor Test, CO2 challenge, Stranger Tests and Fear Conditioning Tasks).

A brief example of domains, constructs/subconstructs, and recommended tasks is presented in Tables 9.1–4 and 9.1–5.

ICD-11 and Dimensional Classification. The recently published ICD-11 criteria for mental disorders is also considering a move toward a dimensional classification. ICD-11's workgroup argues, as also did DSM-5, that mental disorders can be best conceptualized as representing several interacting and underlying dimensions and not as discrete categories. They add that dimensional classification places a stronger focus on current clinical presentations and is more compatible with an approach to prognosis or recovery rather than viewing a given diagnosis as representative of a type of person or being stable over time.

ICD 11 has proposed a dimensional category for psychotic disorders, replacing ICD 10 subtypes of schizophrenia (paranoid, hebephrenic, catatonic) with a set of symptom ratings that correspond to the person's current functioning on several dimensions. These dimensional ratings summarize specific problems that may be the target of specific interventions, and that can improve or remit over time.

ICD 11 has outlined a dimensional classification of personality disorders very similar to what DSM-5 initially proposed for the main body of the manual and which unfortunately ended up relegated to section III of the DSM-5 manual as "Alternative DSM-5 Model for Personality Disorders." This section of the manual includes "emerging measures and models" as well as "conditions for further study."

Table 9.1–4.
RDoC Systems for Social Processes: Recommended Task Paradigms

Constructs/Subconstructs	Tasks
1. Affiliation and attachment	
1.1 Rejection sensitivity	Cyberball
1.2 Social motivation	One-armed bandit test
2. Social communication	
2.1 Reception of facial communication	ER-40-Penn emotion recognition test Gaze cueing
2.2 Production of facial communication	None
2.3 Nonfacial communication (merged reception and production)	TASIT 1
3. Perception and understanding of self	
3.1 Agency	None
3.2 Self-knowledge	Self-referential memory paradigm
4. Perception and understanding of others	
4.1 Animacy perception	Point light displays of biologic motion
4.2 Action perception	How part of how/why task
4.3 Understanding mental status	Hinting task

Table 9.1–5.
RDoC Cognitive Systems: Recommended Task Paradigms

Constructs/Subconstructs	Tasks
1. Attention 1.1 Overt/covert 1.2 Capacity and interference control 1.3 Vigilance	Spatial and nonspatial cuing tasks Attention networks task (ANTS) Visual search paradigm Attentional blink during rapid serial visual presentation Dual task paradigms Tasks with “catch” trials (change detection working memory, perceptual threshold effects)
2. Perception 2.1 Visual	Mind-wandering tasks Contrast–contrast task Jittered orientation visual integration task (JOVI)
3. Declarative memory	Relational and item-specific encoding task (RISE) Mnemonic similarity test
4. Cognitive control 4.1 Goal selection, updating, representation, and maintenance 4.2 Response selection, response inhibition/suppression 4.3 Performance monitoring	Continuous performance tests (AX and DPX) Preparing to overcome prepotency task (POP) Go/no-go tasks Stop signals tasks Flanker task versions Simon task versions Stroop task versions
5. Working memory 5.1 Active maintenance 5.2 Flexible updating 5.3 Limited capacity 5.4 Interference control	Match to sample Sternberg tasks Change detection Continuous performance tests (AX and DPX) NBack tasks Self-ordered pointing Change detection

Nback tasks
Sternberg tasks

DSM-5, ICD-11, and Dimensional Diagnoses in Clinical Practice. A key objective in publishing DSM-5 was to improve upon previous versions of the manual by revising categories and rigorously assessing (and proving) their reliability. A secondary goal of DSM-5 was to incorporate new strategies (dimensional approach) and provide potential new disorders in need of further evaluation and testing. The idea was to create an open, “living” system that could be easily updated as new evidence accrued. DSM-5 is unlikely to change clinical practice, and many practicing psychiatrists largely remain unaffected or indifferent in front of debates, as recently highlighted by psychiatrist-correspondent Sally Satel in a leading editorial. Satel adds, however, that DSM-5 diagnoses will “become passports to insurance coverage, keys to special education services, and tickets to disability benefits.” Thus, it seems quite relevant for practicing clinicians to get to know and use DSM-5 criteria properly. Despite the shortcomings detailed above, the DSM-5 manual provides an excellent description of most of the key syndromes encountered in clinical practice.

Therefore, while psychiatrists wait for future diagnostic developments, the proper, “informed” use of DSM-5 will guide clinical practitioners in assessing, understanding, and cataloging complex mental health phenomena. The expectation—assuming the manual is truly a “living document”—is that experiences accruing from its clinical use may lead to the reinforcement, revision, or even the elimination of some diagnostic labels.

Brief Updates on DSM-5 Manual and Diagnoses. The original idea behind DSM-5 was to make it a “living document” that could be updated as needed, based on available clinical and scientific evidence.

In 2017, the APA published a supplement titled “DSM-5 Update, Supplement to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition” that included coding changes related to ICD-10 as well as specific updates on DSM-5 criteria and text, including changes and amendments for several disorders.

The updates to criteria and text included in the supplement include the following disorders:

- Brief psychotic disorder (change to specifiers)
- Bipolar I and II disorders (change in criteria for manic and hypomanic episodes; change in text for severity specifiers and seasonal pattern)
- Specifiers for bipolar and related disorders
- Adjustment disorders (change to specifiers)
- ASD (change to criterion A)
- Delusional disorder (change in subtypes section of text)
- Persistent depressive disorder (change to differential diagnosis section of the text)
- Specifiers for depressive disorders (changes in text definitions; change to the definition of “with mixed features”)
- Disruptive mood dysregulation disorder (change to the text of section on development and course)
- Generalized anxiety disorder (change to differential diagnosis section)
- Acute stress disorders (change to differential diagnosis section)
- Somatic symptom disorder (change to differential diagnosis section of text)
- Alternative model for personality disorders (AMPD) (change to criterion B1 for OCD)
- Depressive episodes with short-duration hypomania (change in criterion A)

Next, another update titled “Supplement to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition” was published in October 2018.

It included minor new changes in the following disorders (it also reiterated the changes proposed in the updates for 2015 and 2016):

- Psychotic Disorder (Coding Note)
- Bipolar Disorders: “With Psychotic Features” Specifiers
- Substance/Medication-Induced Bipolar and Related Disorder (Coding Note)
- Specifiers for Bipolar and Related Disorders: (“With Psychotic Features” Specifiers)
- Substance/Medication-Induced Depressive Disorder: (Coding Note)
- Other Specified Depressive Disorder; Unspecified Depressive Disorder
- Substance/Medication-Induced Anxiety Disorder: (Coding Note)
- Other Specified Anxiety Disorder; Unspecified Anxiety Disorder

- Substance/Medication-Induced Obsessive-Compulsive and Related Disorder: (Coding Note)
- Acute Stress Disorder: (Criterion A)
- Substance/Medication-Induced Sleep Disorder: (Coding Note)
- Substance/Medication-Induced Sexual Dysfunction: (Coding Note)
- Substance/Medication-Induced Major or Mild Neurocognitive Disorder: (Coding Note)
- Diagnoses associated with substance class: adjustment to neurocognitive disorders column
- No Diagnosis or Condition

In February 2020, the most recent update entitled "Supplement to Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition Avoidant/Restrictive Food Intake Disorder Diagnostic Criteria F50.82" was issued. This supplement highlights minor changes in the text of criterion "A" for Avoidant/Restrictive Food Intake Disorder.

Finally, in March 2022, the APA published the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5 TR). This revision updated all diagnostic criteria and ICD-10-CM codes since DSM-5's publication in 2013. Key new elements are the addition of Prolonged Grief Disorder (PGD) to Section II, considerations on the impact of racism/discrimination on mental disorders and new codes to flag and monitor suicidal behavior.

As can be inferred from these updates, most of the action has taken place 1 or 2 years after the official publication of DSM-5 (2014 and 2015), possibly reflecting a transient "post-DSM-5 enthusiasm" in refining criteria. Since that time, there have been relatively minor changes, and updates and these have mainly concerned ICD-10 codes, as well as a few minor changes and "tweaks" in the criteria and text for a handful of disorders.

Current Research on DSM-5 Diagnostic Categories. That current research is actively testing and probing DSM-5 diagnoses can be briefly illustrated through the brief presentation of a couple of recent papers, one focused on the AMPD, the other on PTSD criteria.

The first paper focusing on AMPD provides a brief overview of the dimensionally oriented AMPD that was included in section III of DSM-5. This paper highlights the fact that since its publication in 2013, the AMPD has inspired a substantial number of studies, underlining AMPD's generative influence on the field. These studies illustrate the reliability and validity of the constructs delineated in the AMPD and underline the empirical challenges to the conceptual clarity of the AMPD, such as evidence of substantial correlations between indices of personality functioning (criterion A in the AMPD) and maladaptive personality traits (criterion B in the AMPD). The key future directions pointed by the researchers would include first a practical component, linking the AMPD research efforts to improving the lives of persons who suffer from PDs, and then surmounting the various challenges germane to the evolution of the DSM itself.

The second paper focusing on PTSD aims at answering the question of how well PTSD can be assessed and compared, spanning the transition between DSM-IV and DSM-5, using different PTSD Checklist (PCL) instruments. This diagnostic study included 1,921 individuals and showed substantial to excellent agreement when individual items, “probable PTSD” and “sum scores” were compared between the previous PCL version and that of *DSM-5*. Thus, the two PCL instruments showed nearly identical associations with comorbid conditions. These results provide support for the transition from the PCL-C to the PCL-5 without losing the ability to monitor trends and associated comorbidities.

We certainly hope that more systematic research on updating/improving current diagnostic systems continues to be implemented in the near future and that it leads to additional improvements in the extant classifications.

Closing Reflection: DSM-5, ICD-11, and RDoC: What Can We Learn from the History of Psychiatric Diagnosis? Karl Wernicke believed that all mental disorders were brain disorders. This belief appears to be the main context for new diagnostic approaches such as the RDoC.

Emile Kraepelin’s main premise was that “all psychiatric disorders should follow the medical model of disease.” As it has been articulated above in this chapter, starting with DSM-III, American nosology became uniquely “Kraepelinian,” with an emphasis on descriptive psychopathology, following the medical model approach.

Karl Jaspers, who published his “Allgemeine Psychopathologie” (General Psychopathology) in 1913, maintained that “psychiatry is a hybrid discipline that must combine the methods of the natural and the social sciences.” This “middle ground” approach that incorporates a “social context” had great influence in the psychiatric traditions of the United Kingdom but not in the United States.

A leading proponent of this middle-ground approach was Professor John Wing, at Maudsley Hospital in London, following the path of Sir Aubrey Lewis and other pioneers. Wing developed the PSE, a semistructured instrument rich in clinical phenomenology that provides detailed symptomatic and syndromic descriptions with elaborate probes, thus requiring significant training and clinical expertise. Previous editions of the PSE were successfully used in two pivotal international studies on psychiatric diagnosis in the 1960s and 1970s, the “United States/United Kingdom Diagnostic Project” (U.S./U.K. study) and the “International Pilot Study of Schizophrenia (IPSS),” both coordinated by the WHO. The U.S./U.K. study demonstrated the extent to which a diagnosis of “schizophrenia” in the United States was broader and less definable compared to its use in England and demonstrated that through training and the use of a structured instrument (the PSE in this instance), wide differences in diagnoses of schizophrenia across the Atlantic would virtually disappear. The IPSS showed that a “concordant” group of patients with schizophrenia could be identified in all countries by trained clinicians using the same diagnostic strategy (PSE). Incidentally, a new version of the PSE, the Structured Clinical Assessment in Neuropsychiatry–Present State Examination (SCAN-PSE), continues to be used by a few teams of investigators around the globe, including a few sites in the United States. This approach emphasizes the need to gather rich clinical data (clinical phenomenology) that takes into account social and environmental influences in addition to biology. The assumption is that biologic systems interact closely with the environment and that social influences can generate or amplify biologic impairments. Unfortunately, this “middle ground” approach to psychiatric diagnosis has been abandoned by modern classifications. Moreover, clinical and epidemiologic studies incorporate questionnaires that are often administered by lay interviewers and contain only brief or abridged instruments to gauge psychopathology. This leads to rather dry, impoverished clinical assessments.

The next paradigm change for psychiatric classification (and it is doubtful for now that RDoC truly represents such a breakthrough) will necessarily have to wait for a much clearer understanding of the matrix of brain structures, such as neuronal circuits, the “omics” repertoire (genomics, proteomics, metabolomics), and a more profound understanding on how social experiences affect and influence such systems. In the interim, perhaps a return to Jasper's phenomenologic approach, with its rich and meaningful analysis of symptom presentations, should be welcomed by many scholars, practicing clinicians, and researchers, as a valuable step in this process. This may help the careful delineation of various pathophysiologic hypotheses that would enhance diagnostic validity by providing etiologic knowledge, a promising alternative to the "reliable" but imprecise classifications based on symptoms and subjective reports.

FURTHER READINGS

American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2014;37(Supp 1):S8–S16.

American Psychiatric Association, *DSM-5 Update, Supplements to the Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2017; 2018 and 2020 updates.

Cosgrove L, Krimsky S. A comparison of DSM-IV and DSM-5 panel members' financial associations with industry: a pernicious problem persists. *PLoS Med*. 2012;9(3):e1001190.

Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med*. 2013;11:126.

Deisseroth K, Etkin A, Malenka RC. Optogenetics and the circuit dynamics of psychiatric disease. *JAMA*. 2015;313(20):2019–2020.

de Leon J. DSM-5 and Research Domain Criteria: 100 years after Jaspers' general psychopathology. *Am J Psychiatry*. 2014;171(5):492–494.

Feighner JP, Robins E, Guze SB, Woodruff RA Jr, Winokur G, Munoz R. Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry*. 1972;26:57–63.

Foucault M, Folie et déraison. Histoire de la folie à l'âge Classique, Plon, Paris 1961.

Global initiative for chronic obstructive lung disease: Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease (2023 report), 2222-2023 global initiative for chronic obstructive lung disease, inc. https://goldcopd.org/

Goldberg D. Plato versus Aristotle: categorical and dimensional models for common mental disorders. *Compr Psychiatry*. 2000;41(2 Suppl 1):8–13.

Goldberg D. Should major classifications of mental disorders be revised? *Br J Psychiatry*. 2010;196:255–256.

Goodwin DW, Guze SB. *Psychiatric Diagnosis*. 1996.

Horwitz AV, Wakefield JC. *The Loss of Sadness*. 2007.

Hyman SE. The diagnosis of mental disorders: The problem of reification. *Annu Rev Clin Psychol*. 2010;6:155–179.

Jaspers K, General Psychopathology, The University of Chicago Press; First Edition, January 1,1963, ISBN-10: 0226394875

Johnson RA, Barrett MS, Sisti DA. The ethical boundaries of patient and advocate influence on DSM-5. *Harv Rev Psychiatry*. 2013;21(6):334–344.

Kendler KS. A history of the DSM-5 scientific review committee. *Psychol Med*. 2013;43:1793–1800.

Klerman GL, Vaillant GE, Spitzer RL, Michels R. A debate on DSM-III. *Am J Psychiatry*. 1984;141(4):539–553. doi: 10.1176/ajp.141.4.539.

Kraemer HC, Periyakoil VS, Noda A. Kappa coefficients in medical research. *Stat Med*. 2002;21:2109–2129.

Kramer HC, Kupfer DJ, Narrow WE, Clarke DE, Regier DA. Moving toward DSM-5: The field trials. *Am J Psychiatry*. 2010;167(10):1158–1160.

Krueger RF, Hobbs KA. An overview of the DSM-5 alternative model of personality disorders. *Psychopathology*. 2020;53:126–132.

Leard Mann CA, McMaster HS, Warner S, et al; Millennium Cohort Study Team. Comparison of posttraumatic stress disorder checklist instruments from Diagnostic and Statistical Manual of Mental Disorders, fourth edition vs fifth edition in a large cohort of US military service members and veterans. *JAMA Netw Open*. 2021;4(4):e218072. doi:10.1001/jamanetworkopen.2021.8072

Lerner AJ. *Diagnostic Criteria in Neurology*. Humana Press; 2006.

Murphy D. *Psychiatry in the Scientific Image*. MIT Press; 2006.

Nemeroff CB, Weinberger D, Rutter M, et al. DSM-5: a collection of psychiatric views on the changes, controversies, and future directions. *BMC Med*. 2013;11:202.

Satel SL. Why the fuss over the DSM-5? The New York Times. 2013.

Wing JK, Sartorius N, Ustun TB, World Health Organization & WHO SCAN Advisory Committee. Diagnosis and clinical measurement in psychiatry: a reference manual for SCAN / edited by J. K. Wing, N. Sartorius and T. B. Ustun, Cambridge: Cambridge University, 1998. https://apps.who.int/iris/handle/10665/42426

9.2 The Classification of Mental Disorders in the International Classification of Diseases (ICD-11)

JAVIER I. ESCOBAR, M.D., M.SC., AND CARLOS LÓPEZ-JARAMILLO, M.D., PH.D.

The International Classification of Diseases (ICD) of the World Health Organization (WHO) is the classification system utilized across the world for clinical, epidemiologic, and administrative purposes. It names and classifies conditions in the areas of health and disease and contributes to the implementation of health policies. The ICD is used by physicians, nurses, and other health professionals, as well as administrators, scientists, researchers, and health insurance companies, among others. The systematic use of this classification allows the examination of the incidence and prevalence of diseases and other health problems worldwide. The proper and systematic use of ICD diagnostic codes allows the creation of reliable registries, which provide a solid base for assessing morbidity and mortality statistics globally. It, therefore, makes it possible to have a global perspective on health and health-related issues.

To American psychiatrists, ICD is of utmost importance as ICD-10 is currently the official system for coding and billing purposes in the United States. Fortunately, the DSM-5 developers took this into account, and in the fifth revision published in 2013, each DSM-5 diagnosis has corresponding ICD-10 Clinical Modification (CM) codes. In the 1970s, the U.S. government decided that ICD-9 codes were not specific enough for clinical use in the United States and developed the CM version. In DSM-5, ICD-10 codes are headed by the letter F. For example, the corresponding ICD-10 codes for anorexia nervosa are F50.01 for the restrictive type and F50.02 for the binge-eating/purging type. The fact that ICD-10 codes are embedded in DSM-5 facilitates the task of shifting from ICD-9 to ICD-10. However, as will be seen ahead, this does not necessarily mean that there is a perfect diagnostic correspondence between DSM-5 and ICD-10 or now ICD-11.

The WHO recently released the 11th version of the ICD classification (ICD-11) for mental disorders. While the official release of this classification was expected for 2015, the process was delayed. Because of this, the information related to ICD-11 is quite fresh, and the broad use of the system is yet to be fully realized. WHO has been conducting training programs worldwide to familiarize people with this latest version of the international classification.

The goal of this chapter is to focus primarily on the recently released ICD-11 classification for mental disorders. The chapter will first provide a brief outline of the history and development of the ICD system leading to ICD-10. Next, it will describe the process of revising ICD-10 and developing ICD-11. It will briefly discuss the main goals set forward, the obstacles faced, and the mechanisms utilized to sort these out. The main body of the chapter will be a detailed description of the categories for mental disorders included in ICD-11, based on the review of authoritative references and other WHO sources, including international lectures and courses focusing on key diagnostic areas of this new classification of mental disorders.

Acknowledging the international input on DSM-5 and the ongoing consultation between the American Psychiatric Association (APA) and WHO, a general expectation was that ICD-11 would resemble DSM-5 to a larger extent than ICD-10 resembled DSM-IV. However, as will be seen ahead, significant differences remain between the DSM-5 and ICD-11 systems. The controversies surrounding the new section on mental disorders of the ICD-11, the similarities and differences with DSM-5, and the complex task of harmonizing these classifications will be analyzed in more detail ahead.

Excellent, detailed descriptions of the ICD and the whole WHO Family of International Classifications (WHO-FIC) can be found elsewhere. Our intent here is to provide only a brief description of the ICD classification as well as a brief outline of the relevant WHO-FIC categories. This brief introduction is intended to highlight the most salient features of the ICD classification to serve as a general background to frame discussions on the ICD system.

THE INTERNATIONAL CLASSIFICATION OF DISEASES

The ICD is the evolving core of a set (or a “family” as it has been called) of disease and health-related classifications. It was the product of a commendable international effort intended to improve the analyses, first of mortality statistics and later those of morbidity, for public health purposes. With time, the actuarial and epidemiologic focus of ICD classification would be extended, allowing its use as an instrument to aid diagnostic and assessment processes in medicine in efforts to improve clinical practice, contribute to research, and facilitate administrative tasks.

The ICD within the WHO Family of International Classifications

The WHO-FIC is a set of classifications that are used to describe and produce information about different aspects of health, disease, and health systems around the world. The strong endorsement that WHO-FIC has received from the United Nations reflects its alignment with the core values of this international organism. The core function of the WHO-FIC is to aid the production of reliable statistical information and assist statistical systems at local, national, and international levels. These statistics would, in turn, be employed in various ways by different kinds of stakeholders to provide input for actions intended to improve the health status and health care of populations around the world. In order to achieve these goals, the WHO-FIC uses a standardized language with the main goal of improving communication between groups and individuals working in the field of health across the world.

The WHO-FIC has been subdivided into *reference*, *derived*, and *related* classifications.

Reference classifications: These deal with basic parameters of health such as death, disease, functioning, disability, health, and health interventions. They can be used at both individual and population levels. The reference classifications of the WHO-FIC include the following:

- The ICD
- The International Classification of Functioning, Disability, and Health (ICF 2) classifies functioning and disability
- The International Classification of Health Interventions (ICHI)

Derived classifications: As the name suggests, these are systems derived from reference classifications developed for more specialized purposes. They are usually created by picking up and rearranging various elements from reference classifications that are complemented by adding up new items (concepts, descriptions, etc.) to the original structure.

The most recently developed *derived classifications* in the WHO-FIC include the following:

- The ICD-11 for Mental, Behavioral, or Neurodevelopmental Disorders
- The ICD-11 for Mental and Behavioral Disorders, Clinical Descriptions and Diagnostic Guidelines
- International Classification of Diseases for Oncology, Third Edition (ICD-O-3)
- Application of the International Classification of Diseases to Neurology (ICD-10-NA)
- Application of the International Classification of Diseases to Dentistry and Stomatology, Third Edition (ICD-DA)

Related classifications: These contain descriptions of other aspects of the health experience and the health system that are deemed relevant for the purposes of health care. They may come from other sectors and/or working groups of the WHO (e.g., Classification of External Causes of Injury or ICECI, the Anatomical Therapeutic Chemical Classification System with Defined Daily Doses or ATC/DDD). Also, these can come from other international organizations (e.g., the International Classification of Primary Care, Second edition, or ICPC-2, the International Classification for Nursing Practice, or ICNP, etc.).

Table 9.2–1.
Types of Classifications of WHO Family International Classification

1. **Reference Classifications**
    - International Classification of Diseases (ICD)
    - International Classification of Functioning, Disability and Health (ICF)
    - International Classification of Health Interventions (ICHI)
2. **Derived Classifications**
    - International Classification of Diseases for Oncology, Third Edition (ICD-O-3)
    - The ICD-11 for Mental, Behavioral or Neurodevelopmental Disorders
    - The ICD-11 for Mental and Behavioral Disorders Diagnostic Criteria for Research
    - Application of the International Classification of Diseases to Neurology (ICD-10-NA)
    - Application of the International Classification of Diseases to Dentistry and Stomatology, Third Edition (ICD-DA)
3. **Related Classifications**
    - International Classification of Primary Care, Second Edition (ICPC-2)
    - International Classification of External Causes of Injury (ICECI)
    - Technical aids for persons with disabilities—Classification and terminology (ISO9999)
    - The Anatomical Therapeutic Chemical Classification System with Defined Daily Doses (ATC/DDD)
    - International Classification for Nursing Practice (ICNP)
    - Verbal autopsy standards: ascertaining and attributing causes of death

Table 9.2–1 shows the general structure of the types of classifications used by the WHO-FIC.

Development of the International Classification of Diseases

To classify is to cluster and allocate a set of objects (in this case, disorders) according to logical rules that guide the clustering activity. These rules specify what type of criteria would determine that a given disorder belongs to a particular category (e.g., the presence or absence of delusions and/or hallucinations; a longer duration vs. a brief duration of symptoms, etc.). Also, the relationships between the members of a category must be specified. Usually, an important purpose of the classification of diseases is to bring information to the users about the health problems included within its categories. The information conveyed by the classification will then be used according to the needs of the user (e.g., diagnosis, research, policy, etc.). Thus, classifications are crucial for generating certain types of information that will be stored and then used for comparison purposes, exchanges, analyses, disease monitoring, and of course, policy decisions about health, disease, and health care.

This practical aim of classifications has led to the assertions that classifications are not *true* or *false*, but only *useful* or *not useful,* depending on their purpose. Thus, a classification of diseases can be very useful for clinical diagnosis if it contains diagnostic features. Then it can be used in epidemiologic research on a particular health problem if the diagnosis of the disease (the purpose of the classification) is accurate enough to permit the measuring of prevalence or incidence of the disorder in populations, and so on. However, a diagnostic classification can be less helpful or even useless when utilized for other purposes. The pre-eminence of practical utility over other features of classification such as truthfulness, ease of use, heuristic value, methodologic and/or scientific rigor, social consequences, etc., have been debated for a long time.

As mentioned before, one of the most important uses of disease classifications is to improve communication across a heterogeneous group of users, including clinicians, researchers, or statisticians, but also between organizations and institutions that manage international or global information on diseases and health-related matters. If communication is understood as being an exchange of information between users, it presupposes a common language. The organized collection of certain terms and meanings and a clear description of the rules for their use are the functional characteristics of contemporary international classifications.

Developers of classifications must make decisions about which terms, concepts, and descriptions are to be included or excluded. They must outweigh the pros and cons of these selections. They have to justify a certain degree of standardization of language (and thinking) in efforts to improve communication at the expense of suppressing the natural variability inherent to human experiences of health and disease. Although some of these decisions are supported by empirical research, often, developers and reviewers of nosologies have to resort to nonempirical solutions (e.g., practical reasons).

It must also be remembered that users can come from different disciplines, pertain to different organizations, and thus have different knowledge backgrounds, goals, and interests. The heterogeneity of users is not a trivial matter. One of the main challenges for a multipurpose classification (i.e., a classification that will be used to provide epidemiologic information to public health authorities but at the same time must help clinicians to make accurate diagnoses) is to offer to an ample variety of users the information they need without modifying the basic structure of their practices. Otherwise, a classification should be created and tailored for each group of users, something that is not deemed practically feasible and could lead to communication problems.

A Chronology of ICD Development. It can be said that classifying diseases according to a set of specific principles (e.g., symptoms) has been the rule since the beginning of medicine. The objects, principles, types, methods, and purposes of classification have been changing according to the historical period, and change has been driven by diverse factors such as scientific, philosophical, practical, cultural, political, economic, and legal issues. The mid-19th century, more specifically the year 1853, is generally identified as the one when interest arose to start cataloging the causes of death that could be used in “all countries” as part of the first International Statistical Congress. Two years later, William Farr presented a five-group classification of diseases: epidemic diseases, constitutional diseases, local diseases (arranged according to their anatomical site), developmental diseases, and diseases that are a direct consequence of violence. Revisions using Farr’s principles were made in 1865, 1874, 1880, and 1886. Principles of classification were changed in 1893, when Alphonse Bertillon, a French biometrics researcher, submitted to the newly created International Statistical Institute (ISI) a comprehensive classification of causes of death based on a synthesis of English, German, and Swiss classifications. Causes of death were divided into two main groups: those due to generalized diseases and those localized to a particular organ or anatomical site. Bertillon's classification proved successful and was adopted by several countries, and the ISI officially recommended that this classification (the International List of Causes of Death [ILCD]) should be adopted by all those offices that were not yet using it. Later in 1929 and 1938, the revisions of the ILCD were made by commissions composed of representatives of both the ISI and the Health Organization of the League of Nations. In 1946, the aim of the international efforts on classification was expanded beyond death causes, and the WHO was commissioned to create an International List of Causes of Morbidity. After a meeting in 1948, the work on the revision of the classification was from then on carried out by the WHO with the support of national committees on vital and health statistics. The constitution of the WHO includes the development of the classification as one of its functions. The 7th revision conference was held in Paris; the 8th, 9th, and 10th revision conferences were held in Geneva in 1965, 1975, and 1989.

It seems clear that the most distinctive characteristic of the history of the ICD enterprise has been to foster international participation in efforts to create a classification of causes of death and disease that would be both useful and used by countries all around the world.

The ICD-9

The ninth revision (1975) deserves at least a brief mention here because important changes at different levels of the classification process were implemented during this revision. This marked the beginning of the ICD "family," as several additional documents were introduced, including "the classification of procedures in medicine" and the "classification of impairments and handicaps."

Also, the ICD was adapted to the needs of primary care as well as certain medical specialties (e.g., oncology).

Recommendations and norms were issued about such matters as lay reporting of disease, mortality coding rules, and the selection of a single cause for death statistics. Also, the well-known convention of using a dagger (†) to mark the underlying disease and an asterisk (*) to mark its manifestations was issued during this process. Perhaps one of the most important single events for classification in psychiatry was ICD-9's inclusion of a glossary and brief descriptions of the categories included in the fifth chapter (mental disorders), a procedure that was not part of the other chapters. The intent of the glossary and the clinical descriptions was to assist clinicians in making the diagnosis.

The 10th Revision of the International Classification of Diseases

Mental disorders were included in Chapter V of ICD-10 as “Mental and Behavioral Disorders.”

Approved in 1992, the ICD-10 was divided into three volumes. In volume 1, “the tabular list,” all medical conditions included in ICD-10 were listed. The listing was organized alphanumerically, ranging from A00 to Z99. In this 10th version of the ICD, diagnostic codes do not go beyond the fourth character (e.g., F31.0). Each character corresponds to the hierarchical level of the classification. For example, in the code F321, the “F” corresponded to Mental and Behavioral Disorders; “F3” corresponded to Mood (Affective) Disorders, the “F32” to Depressive Episode, and “F32.1” to Moderate Depressive Episode.

Table 9.2–2 shows the list of chapters of ICD-11.

The ICD-10 Version for Use in Primary Care (ICD-10 PHC). An important side product of ICD-10 was the PHC version that focused on mental disorders commonly seen in primary care. The development of this version involved significant communications and exchanges of information between general practitioners and psychiatrists. It included a list of criteria and guidelines for 28 mental disorders commonly seen in primary care.

The Research Version of ICD-10. Consistent with the idea of a multipurpose classification for multiple kinds of users, the ICD-FIC included a version for research called “Diagnostic Criteria for Research” (ICD-10 DCR). Compared to previous versions, this one was less descriptive but included more precise wording, more restrictive definitions, a description of inclusion and exclusion criteria, and definitions of fixed symptom thresholds.

Table 9.2–2.
List of Chapters of ICD-11

01	Certain infectious or parasitic diseases
02	Neoplasms
03	Diseases of the blood or blood-forming organs
04	Diseases of the immune system
05	Endocrine, nutritional or metabolic diseases
06	**Mental, behavioral or neurodevelopmental disorders**
07	Sleep-wake disorders
08	Diseases of the nervous system
09	Diseases of the visual system
10	Diseases of the ear or mastoid process
11	Diseases of the circulatory system
12	Diseases of the respiratory system
13	Diseases of the digestive system
14	Diseases of the skin
15	Diseases of the musculoskeletal system or connective tissue
16	Diseases of the genitourinary system
17	Conditions related to sexual health
18	Pregnancy, childbirth or the puerperium
19	Certain conditions originating in the perinatal period
20	Developmental anomalies
21	Symptoms, signs or clinical findings, not elsewhere classified
22	Injury, poisoning or certain other consequences of external causes
23	External causes of morbidity or mortality
24	Factors influencing health status or contact with health services
25	Codes for special purposes
26	Supplementary Chapter Traditional Medicine Conditions – Module I

COMPARING ICD AND DSM CLASSIFICATIONS

The existence of two main systems of classification that have widespread international use requires an understanding of their differences and similarities. Comparisons between the two systems have proven difficult, as they represent different conceptualizations of purposes and methods for classifying mental disorders. It is often forgotten that the specific aims that developers of each particular classification system had in mind will heavily influence the methods for its development and the final written product. Therefore, it is not surprising to find out that the ICD and DSM classifications have several differences, some more substantial than others. These differences do not imply that one of the classifications is better than the other. Choosing one over the other is essentially a value judgment, and henceforth relative. For example, a classification X can be considered better than classification Y if clinical utility is the point of reference for judgment. However, classification Y could be considered better than X if it is decided that research utility is the main criterion for judging superiority.

The most important differences between ICD and DSM stem from the goals of the umbrella organizations that created them. Thus, as an international tool, ICD's key aim is to provide clinically useful descriptions for users pertaining to different national and cultural backgrounds across the world without negatively affecting other important features like reliability, validity, and clinical utility. As a national classification in the United States, DSM is built to make diagnostic recommendations of interest primarily to members of a national professional organization (the APA). The ICD is a system for illness reporting, whose use is mandatory in the countries that agreed to use it. National classifications (DSM) do not require governmental approval but are under the authority of the professional associations (in this case, the APA) that regulate their construction, revision, and release.

ICD has to undergo a series of steps prior to submission and approval. The approval comes first from reference centers, and then it is submitted to the health ministers of the countries members of the WHO that will discuss its acceptance. This international process involves a high level of consultation and negotiation, in which the views of different countries (and their health traditions) have to be understood, respected, and taken into account. The international character and orientation of ICD is a different process from the international dissemination and acceptance of the DSM system. ICD is primarily concerned with incorporating conceptual and guiding principles for clinical practice globally, while DSM has other priorities such as the use of the classification for research purposes. Making different versions available for different needs and users means that the ICD has to offer clinical descriptions that are useful and practical for users from different professional backgrounds (not only psychiatrists or physicians) without negatively affecting other important features like its international character and reliability. Consequently, different versions of the ICD have been made available for psychiatrists, primary care practitioners, and researchers. National classifications like the DSM tend to be “prescriptive” in the sense that they regulate what terms can or cannot be used to report diagnoses. More recently, the DSM system also developed a version for use in primary care. International classifications take a different approach, and therefore, rather than dictating or prescribing, they aim at providing useful, practical categories that health professionals can use in their daily clinical work.

A frequent example of the practical advantages of the ICD system over the DSM is the use of a particular approach to facilitate sorting out the frequent presentation of both anxiety and depressive symptoms in a given patient when there is not a clear predominance of either clinical presentation. For this type of situation, the ICD has provided, for some time, a category called “mixed anxiety and depression,” mainly because practicing clinicians view this as an important issue and have insisted on having it available as part of the classification.

The DSM system, on the contrary, has not accepted such a category, although it has been acknowledged that it merits further study. This difference in strategy between ICD and DSM has been framed as the “bottom-up” versus the “top-to-bottom” approaches to classification.

The use of clinical descriptors instead of diagnostic (“operational”) criteria is undoubtedly one of the most important differences between ICD and DSM, reflecting concepts and values underlying the ICD enterprise. ICD developers decided to include brief clinical descriptions—a narrative account of the main features of the disorders—to help diagnosticians. DSM developers took the approach of articulating a list of criteria accompanied by instructions about the number and the combinations of criteria needed to make a diagnosis. Both approaches have strengths and limitations, mostly related to the specific purposes for using the classification (e.g., clinical work vs. research).

If one accepts that both classifications have their strengths and limitations, then some categories may be better portrayed in DSM and some in ICD. Since both systems have been created with different aims in mind, the acceptability and satisfaction of users is an important additional factor to consider when comparing these two diagnostic traditions.

Table 9.2–3.
Summary of Major Differences between DSM and ICD

DSM	ICD
▶ National classification of the United States	▶ Official world classification
▶ Focus on psychiatrists and psychologists	▶ Designed to be used by and useful for the different health professionals
▶ Developed to fulfill the particular information needs of the U.S. health system	▶ Created with the idea of being useful for different health systems around the world
▶ Operational criteria	▶ Clinical descriptions
▶ Property of the APA	▶ Free of charge and open access
▶ U.S. Anglophone perspective	▶ Multilingual and multicultural
▶ Advantages for research	▶ Advantages for clinical use
▶ Different formats, one version	▶ Multiple versions and formats for multiple users

In general, practicing clinicians consider ICD to be clinically useful and many scholars and clinicians, many of them practicing in the United States, view it as more appealing than the DSM system for clinical use.

In the United States, DSM is most frequently used by psychiatrists, less so by psychologists, a trend that might be changing. Surveys of international clinicians indicate that they tend to use DSM more than ICD, possibly because DSM is what they were primarily taught during their psychiatric/psychological training. Moreover, researchers and other international scholars appear to have adopted DSM as their preferred diagnostic system. Researchers seem to feel more comfortable with DSM because of the “operational” criteria, which allows for a more distinct recording and sorting out the clinical features of the disorders.

Table 9.2–3 shows a summary of the differences between DSM and ICD.

Specific Differences between ICD and DSM

Empirical Comparisons: ICD-11 versus DSM-5. Although there is a general impression that ICD-11 and DSM-5 have more similarities than differences, empirical data suggests that concordance between the two systems is not as good as expected.

In an important comparison published in 2009, Michael First reported that out of the 176 criteria sets he examined, only one disorder (transient tic disorder) had identical DSM-5 and ICD-11 definitions. Twenty-two percent of the nonidentical sets had definitional differences that were conceptually based, while the other 78% had nonconceptual-based differences. Several other studies have also reported significant differences between DSM-5 and ICD-11.

MENTAL DISORDERS IN ICD-11

In Chapter 6 of ICD-11, mental disorders were renamed as "Mental, Behavioral or Neurodevelopmental Disorders." The term "neurodevelopmental" has a long history but was not included in previous editions of the ICD or the DSM. The term is used to describe disorders of early onset that affect cognitive and social functions, display important sex differences, and have a chronic course leading to impairment and lasting into adulthood.

Main issues highlighted herein:

1. A general overview of mental disorders in ICD-11.
2. Changes in specific mental disorders between ICD-10 and ICD-11.
3. General issues and problems encountered in the process of updating ICD-10, and a brief discussion of how these were sorted out.
4. The process of "harmonizing" ICD-11 and DSM-5.

The Process of Developing ICD-11 from ICD-10

Several key features deemed essential for developing solid, useful, and practical classifications were incorporated as "conceptual signposts" in the development of ICD-11.

These included the following:

- The system originates and expands from a set of basic points and agreements between mental health professionals.
- It requires simplicity and understandability to allow its easy use by those dealing with common mental disorders.
- It should be attractive to users as a tool of information exchange.

- It must be easily translatable into any already established national classifications.
- It should be conservative and theoretically familiar in order to be acceptable to diverse users with different levels of knowledge, disciplinary backgrounds, and interests.
- Changes should be implemented only when there is sufficient knowledge available to justify innovations.
- Because of its international character, the languages in which it will be translated must be kept in mind.
- Successive revisions must retain continuity with the previous versions.

Why Change ICD-10? No extant classification is considered “good enough,” either from a taxonomic point of view or from the perspective of yielding reliable, practical diagnoses for clinical work and/or research.

Some of these difficulties have been linked to irregularities in terms, concepts, and descriptions leading to limitations in the diagnostic information contained in the descriptions of disorders.

The environmental components and the context in which classifications are used undergo profound changes with the passage of time. Therefore, major reasons for reviewing ICD-10 included efforts to improve the consistency and uniformity of the diagnostic information provided by the clinical descriptions and to adapt to new demands from users and conditions of use. It is important in this context to generate new research data about mental disorders that is backed up by scientific developments (evidence based) and that satisfies new demands from clinical users about utility and ease of use in various political, social, and economic environments. It must also be user-friendly, adapt to new technologic environments, and be amenable to frequent updates.

Improving Classification Systems without Losing What Has Been Gained. A revision process needs to keep a balance between conservatism and innovation. In other words, it should improve the classification as much as possible without losing any of the good elements present in it. The process of revising ICD-10 has been compared to the release of new “software.” Besides offering new features, it also has to be compatible with previous versions. Similar to what was proposed for DSM-5 (“living document”), ICD-11 was expected to experience transformations if and when new scientific information became available.

Maintaining the Basic Purposes While Increasing the Diversity of Stakeholders. The three basic purposes of the classification (clinical, public health, and research use) have to be maintained. The main stakeholders of ICD-11 include the governments of the WHO member countries, health care professionals (not only mental health professionals or psychiatrists), and researchers. However, patients and "users" of mental health services have also been included among the stakeholders, although the mechanisms for their participation have not been yet clearly articulated. Other important stakeholders are legal systems, health policymakers, and payers of health care services.

Clinical Utility: The Core Principle Guiding the 11th Revision of ICD. As stated before, clinical utility backed up by research evidence was the guiding principle for ICD-11. Michael First and collaborators noted that "clinical utility deserved to be an organizing priority for the revision so long as it did not sacrifice validity as established by the best available science." However, owing to the scarcity of data coming from "best available science" that can be used to inform classification, utility, reliability, and some equivalents of validity became the core guiding principles of ICD-11.

The development and refinement of classifications in psychiatry is a reminder that the discourse on science, scientific rigor, objectivity, and so on must be approached cautiously because very often, the practical needs and social pressures are the drivers of diagnostic decisions. This is not a negative issue per se. What is undesirable is the "reification effect" of classifications in the absence of acceptable reliability and validity.

WHO TAG-MND, GSPN, and CDDG. The WHO's Topic Advisory Group (TAG-MND), originally created in 2007, had the main objective of drafting proposals for the classification of mental disorders in ICD-11.

The TAG-MND, in turn, created several subgroups (e.g., epidemiologic group, groups working on harmonization ICD-11/DSM-5, mental disorders in primary care, health problems related to substance misuse, childhood mental disorders, intellectual disability (ID), and personality disorders).

The WHO also created a coordinating group to establish the Global Scientific Partnership Network (GSPN), involving experts representing different disciplines and traditions in psychiatry. Its role is to help in the gathering and analysis of research data relevant to classification, the design of field trials, and the assessment of the translation process, among its principal tasks.

Active consultation (with individuals, institutions, and governments) and a thorough literature review relevant to the classification of mental disorders were also implemented as part of this process. Technical preparation and final ICD-11 products were assigned to a WHO special unit.

These complex tasks turned into a long and tortuous process marked by delays and postponement of deadlines.

For mental health professionals, the “Clinical Descriptions and Diagnostic Guidelines (CDDG) for ICD-11 Mental and Behavioral Disorders” were prepared. Informally known as the “blue book” during the tenure of ICD-10, CDDG was intended primarily for clinical, educational, and service use.

A particularly important feature of the ICD-11 CDDG is the description of the essential features of each disorder (those that clinicians would reasonably expect to find in all cases of the disorder). This highlights essential features that may resemble diagnostic criteria but generally avoid cutoffs or other precise requirements (such as symptom counts, duration of symptoms, etc.).

This approach was intended to conform to the way clinicians actually make diagnoses, flexibly exercising clinical judgment, and enhancing clinical utility by considering and encouraging descriptions of cultural variations in clinical presentations and other contextual and health system–related factors relevant to clinical practice. This approach was consistent with expectations of clinicians elicited early in the process and was confirmed by field studies.

A lifespan approach was also adopted in ICD-11, thus ending the separation of disorders by the age of onset.

Cultural Issues in ICD-11. An important new feature in ICD-11 is the inclusion of relevant information related to cultural background. This was based on expert advice and systematic reviews of the literature. Each diagnostic area in ICD-11 includes a description of possible cultural issues operating in each instance.

A good example to illustrate this new dimension is the cultural outline for assessing panic disorder as included in the ICD-11 manual.

CULTURE-RELATED FEATURES

- The symptom presentation of panic attacks may vary across cultures, influenced by cultural attributions about their etiology. For example, individuals of Cambodian origin may emphasize panic symptoms attributed to dysregulation of *khyâl*, a wind-like substance in traditional Cambodian ethnophysiology (e.g., dizziness, tinnitus, neck soreness).
- There are several notable cultural concepts of distress related to panic disorder, which link panic, fear, or anxiety to etiologic attributions regarding specific social and environmental influences. Examples include attributions related to interpersonal conflict (e.g., *ataque de nervios* among Latin American people), exertion or orthostasis (*khyâl cap* among Cambodians), and atmospheric wind (*trúng gió* among Vietnamese individuals). These cultural labels may be applied to symptom presentations other than panic (e.g., anger paroxysms, in the case of *ataque de nervios*), but they often constitute panic episodes or presentations with partial phenomenologic overlap with panic attacks.
- Clarifying cultural attributions and the context of the experience of symptoms can inform whether panic attacks should be considered unexpected, as must be the case in panic disorder. For example, panic attacks may involve specific foci of apprehension that are better accounted for by another disorder (e.g., social situations in social anxiety disorder). Moreover, the cultural linkage of the apprehension focus with specific exposures (e.g., wind or cold and *trúng gió* panic attacks) may suggest that acute anxiety is expected when considered within the individual's cultural framework.

Definition of Mental Disorders in ICD-11

ICD-11 defines mental, behavioral, and neurodevelopmental disorders as "syndromes characterized by clinically significant disturbance in an individual's cognition, emotional regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes that underlie mental and behavioral functioning. These disturbances are usually associated with distress or impairment in personal, family, social, educational, occupational, or other important areas of functioning."

ICD-11's Prototypes, Criteria, and Clinical Descriptors. The question of which is the best approach to portray symptoms and other clinical features of mental disorders in such a way that they are easy to identify and can be readily used by clinicians was carefully considered during the process of revising.

Thus, as part of the ICD-11 approach, clinicians provided input on whether or not the diagnostic features outlined in the clinical description matched those of “real” patients. The process emphasized the relevance of clinical and phenomenologic information about the disorders, including essential, typical, or pathognomonic features, differential diagnosis, the boundary with normality, typical course, as well as developmental and cultural-specific features.

The procedure called “prototype matching” consists of clinicians estimating to what extent the patient’s clinical presentation matches a paragraph-length description of the disorder. The clinician has to estimate the extent of matching using a 5-point scale, from 1 (“little or no match”) to 5 (“very good match”). The expected result will consist both of a dimensional diagnosis (capturing the extent to which the patient has the disorder) and a categorical one because higher ratings (4 and 5) correspond to the presence of the disorder. A rating of 3 indicates “subthreshold” or “clinically significant features.” Despite its limitations, the “prototype matching” procedure was considered an important development, and there was consensus to keep refining this approach in the near future.

Dimensions and Categories in ICD-11. The use of the term “dimensional” has become fashionable in psychiatry in recent times, following the debate and controversies between DSM-5 proponents and those advocating “pure” dimensional models such as the Research Domain Criteria (RDoC) proposed by the NIMH. The potential complexity of this approach and the focus on assessing specific constructs is illustrated in Tables 9.2–4 and 9.2–5.

DSM-5 considered dimensions, at least in the initial phases of its development, and provided potential dimensions that were included in section III of the DSM-5 manual. Dimensional approaches are viewed as novel, as they may free psychiatric classifications from the conceptual, empirical, and practical restraints of operational categories.

While its emphasis is placed on categories, ICD-11 also considers dimensions for use in a number of clinical scenarios where these may be feasible and also potentially helpful.

Table 9.2–4.
RDoC Negative Valence Systems Recommended Task Paradigms

Construct/Subconstructs	Tasks
1 – Acute Threat	Subjective Unit of Discomfort Score (SUDS) Fear Survey Schedule
2 – Potential Threat	Intolerance of Uncertainty Scale Behavioral Inhibition Scale Fear of Negative Evaluation Scale Anxiety Sensitivity Index Life Events and Difficulties Schedule (LEDS)
3 – Sustained Threat	Youth Life Stress Interview Childhood Trauma Questionnaire LEDS difficulties Traumatic Events Screening Inventory (TESI) Risky Families Stress and Adversity Inventory (STRAIN)
4 – Loss (Analog of Response to Loss)	LEDS (social experience of loss and potential threat) STRAIN
5 – Frustrative Nonreward	Frustrative Nonreward Responsiveness Subscale Questionnaire of Daily Frustrations

Table 9.2–5.
RDoC Positive Valence Systems Recommended Task Paradigms

Construct/Subconstructs	Tasks
1 – Reward Responsiveness	
1.1 Initial Response to Reward	Simple Guessing Task
1.2 Reward Anticipation	Monetary Incentive Delay Task
1.3 Reward Satiation	Fixed Ratio Satiation Schedule
2 – Reward Learning	
2.1 Habit 2.2 Probabilistic and Reinforced Learning 2.3 Reward Prediction Error	Devaluation Task Habit Task Habit Learning Task Probabilistic Reward Task Pavlovian Conditioning Drifting Double Bandit Probabilistic Stimulus Selection Task
3 – Reward Valuation	Rutledge Passive Lottery Task Drifting Double Bandit
3.1 Reward (Probability) 3.2 Delay 3.3 Effort	Probabilistic Choice Task Willingness to Pay Task Delay Discounting Task Effort Expenditure for Reward Task

Dimensional Perspectives in ICD-11. The dimensional potential of the ICD-11 is most clearly realized in the classification of personality disorders.

In nonspecialist settings, the dimensional rating of severity for ICD-11 personality disorders offered greater simplicity and increased clinical utility compared to the ICD-10 classification of specific personality disorder categories. It allows improved differentiation of patients who need more complex interventions as compared to those requiring simpler treatments. It also facilitates tracking clinical changes over time.

The dimensional perspective eliminates both the excessive comorbidity of personality disorders and the unspecified personality disorder diagnoses. It also provides a base for doing research on the various dimensions and potential interventions for the various manifestations of these disorders.

A set of dimensional qualifiers has been introduced to describe the symptomatic manifestations of schizophrenia and other primary psychotic disorders. Rather than focusing on diagnostic subtypes, the dimensional classification focuses on relevant aspects of the current clinical presentation in ways that are much more consistent with recovery-based psychiatric rehabilitation approaches.

These dimensional approaches to personality disorders and symptomatic manifestations of primary psychotic disorders are described in more detail under the respective sections ahead in this paper.

“Specifiers” and “Qualifiers” in ICD-11. ICD does not use coded “specifiers” because of the limits of the coding system. Therefore, in instances where certain features of a disorder are deemed important and need to be highlighted, noncoded “qualifiers” can be added (see, e.g., schizophrenia below).

Chapter 6 in ICD-11: Mental and Behavioral Disorders

The ICD-11 chapter on mental disorders is Chapter 6, thus being located in a different section as it was the case with ICD-10.

Table 9.2–6 shows a list of blocks of Chapter 6, Mental and Behavioral Disorders of ICD-11.

Table 9.2–6.
List of Blocks of Chapter 6 Mental and Behavioral Disorders of ICD-11

- Neurodevelopmental disorders
- Schizophrenia or other primary psychotic disorders
- Catatonia
- Mood disorders
- Anxiety or fear-related disorders
- Obsessive-compulsive or related disorders
- Disorders specifically associated with stress
- Dissociative disorders
- Feeding or eating disorders
- Elimination disorders
- Disorders of bodily distress or bodily experience
- Disorders due to substance use or addictive behaviors
- Impulse control disorders
- Disruptive behavior or dissocial disorders
- Personality disorders and related traits
- Paraphilic disorders
- Factitious disorders
- Neurocognitive disorders
- Mental or behavioral disorders associated with pregnancy, childbirth or the puerperium
- 6E40 Psychological or behavioral factors affecting disorders or diseases classified elsewhere
- Secondary mental or behavioral syndromes associated with disorders or diseases classified elsewhere
- 07 Sleep-wake disorders
- Sexual dysfunctions
- Gender incongruence
- 6E8Y Other specified mental, behavioral or neurodevelopmental disorders
- 6E8Z Mental, behavioral or neurodevelopmental disorders, unspecified

General Structure of the Diagnostic Categories and Clinical Descriptions in ICD-11

In order to improve uniformity in the ICD-11 descriptions, the working groups used a template known as the “content form.” This template standardized the general structure of the information included for each disorder. The variability and unevenness of content and format in ICD-10 are thought to reflect the work of different groups of experts who work independently, generally with little overall coordination and without adequate standardization of the different sections.

For ICD-11, the “content form” includes items arranged in a special manner and guided by rules of application that yield technical, clinical, and administrative information. Also, a standard format for clinical descriptions and diagnostic guidelines is prepared.

Indeed, a number of experts have insisted for some time that diagnostic guidelines should become more realistic from a cognitive perspective, that is, using a structure that truly represents the way clinicians think, practice, and use their clinical judgment.

Table 9.2–7 illustrates the standard format for ICD-11 Clinical Descriptions and Diagnostic Guidelines.

ICD-11 Field Trials

A series of field trials, both clinic based and internet based, are conducted, focusing on clinical utility and reliability. The web-based field studies use standardized case materials in order to test whether or not the clinical guidelines for ICD-11 diagnoses produce more consistent diagnostic behavior than ICD-10. The methodologic rigor and predictive power of the vignette’s methodology is a key strength in this process. Vignettes are designed to produce a behavioral response (diagnosis) by clinicians, which in turn help assess the usefulness of the classification. The Global Clinical Practice Network (GCPN) was established to allow mental health and primary care professionals from all over the world to participate directly in the development of the ICD-11 CDDG through internet-based field studies.

Table 9.2–7.
Standard Format for ICD-11 Clinical Descriptions and Diagnostic Guidelines

Category Name

Brief Definition: containing a summary statement (100–125 words) of the common essential features of the disorder

Inclusion Terms

Exclusion Terms

Essential (Required) Features: relatively explicit guidance regarding features needed to confidently make the diagnosis

Boundary with Normality (Threshold): specifying those aspects of the disorder that are indicative of its pathologic character

Boundary with Other Disorders (Differential Diagnosis): this section provides guidance to the clinician about how to make this differentiation

Coded Qualifiers/Subtypes

Course Features: information on temporal patterns like age of onset, persistence, duration, likely progression over time, etc.

Associated Clinical Presentations: information about features that are important and help to recognize variations in presentation but also highlights areas where clinical intervention might be important

Culture-Related feature: brief information regarding cultural considerations for diagnosis

Developmental Presentations

Gender-Related Features

The GCPN includes about 15,000 clinicians from 155 countries. Most members are physicians, predominantly psychiatrists, and 30% are psychologists.

All WHO global regions are represented in proportions that largely correspond to the number of mental health professionals in each region, with the largest proportions coming from Asia, Europe, and the Americas (approximately equally divided between the United States and Canada on the one hand and Latin America on the other). The ICD-11 field studies program also represents an area of major innovation. This program of work includes novel methodologies for studying the clinical utility of the draft diagnostic guidelines, including their accuracy and consistency of application by clinicians as compared to ICD-10, as well as the specific elements responsible for any observed discrepancies. A key strength of the research program is that most studies are conducted in a given time frame, allowing results to provide a basis for revision of the guidelines in efforts to address any observed weaknesses. Global participation is a defining characteristic of the ICD-11 CDDG field studies program (World Psychiatry 18: 1-February 2019).

Several GCPN studies have been completed to date, most focusing on comparisons of the proposed ICD-11 diagnostic guidelines with ICD-10 guidelines in terms of accuracy and consistency of clinicians' diagnostic formulations, using evaluations of standardized case material to test key differences. Other studies have examined gradients for diagnostic qualifiers and how clinicians actually use classifications. GCPN studies have been conducted in Chinese, English, French, Japanese, Russian, and Spanish and have included an examination of results by region and language to identify potential difficulties in global or cultural applicability as well as problems in translation.

Clinic-based studies have also been conducted through a network of international field study centers to evaluate the clinical utility and practical value of the proposed ICD-11 diagnostic guidelines in natural conditions in the settings in which they intend to be used. These studies also evaluate the reliability of diagnoses that account for the greatest proportion of disease burden and mental health services utilization. International field studies are located in 14 countries across all WHO global regions, and patient interviews for the studies are conducted in the local language of each country.

The ICD-11-Primary Health Care (ICD-11-PHC)

Similar to what was done in the case of ICD-10, the 11th edition of ICD also has revised the classification for use in primary care settings, the ICD-11-Primary Health Care (ICD-11-PHC). The working group had anticipated that although many of the disorders included in the ICD-10-PHC would be retained, some more "radical" proposals were made, pending field test results.

These proposals included recommendations such as:

- Abandon the distinction between anxiety and mood disorders and create a category called "dysphoric disorders." This will allow clinicians to diagnose depression, anxiety, or the combination of both as "anxious depression."

- Anxiety disorder would be largely equivalent to generalized anxiety disorder (GAD), but without the longitudinal criteria duration, to facilitate detection and intervention by general health professionals.
- A new category of "bodily distress disorders" will include bodily distress syndrome (formerly unexplained somatic complaints), health preoccupation, and conversion disorder.

Specific Disorders in ICD-11

Schizophrenia or Other Primary Psychotic Disorders. This grouping includes the following disorders:

6A20 Schizophrenia
6A21 Schizoaffective disorder
6A22 Schizotypal disorder
6A23 Acute and transient psychotic disorder (ATPD)
6A24 Delusional disorder
6A2Y Other specified primary psychotic disorder

Note that the name of the category was changed from "schizophrenia, schizotypal and delusional disorders" in ICD-10 to "schizophrenia or other primary psychotic disorders" in ICD-11.

Symptomatic profile and duration of illness were the main basis for categorization.

Schizophrenia, schizoaffective disorder, and ATPD were the categories that underwent the most significant changes.

SCHIZOPHRENIA. The specificity and diagnostic value of "first-rank symptoms" has been challenged and thus, it was "de-emphasized" in ICD-11. The diagnosis of schizophrenia requires the presence of two out of eight possible symptoms, one of which must be a "core symptom" (delusions, thought insertion/withdrawal, hallucinations, and thought disorder). The duration requirement is the same as it was in ICD-10 (i.e., 4 weeks). If a patient presents with the required symptomatic profile but does not fulfill the duration criteria, the disorder would be classified as "unspecified primary psychotic disorder."

As in DSM-5, traditional clinical subtypes (paranoid, catatonic, and others) were dropped because of their lack of longitudinal stability and prognostic value. Also, research shows that these subtypes are seldom used in clinical practice.

In general, the ICD-11 category of schizophrenia looks quite similar and compatible with the DSM-5 category except for the issue of “functional impairment.” The ICD’s general position of not including functional impairment as part of the definition of a disease or a disorder will be continued in ICD-11. Consequently, it will not be part of the diagnostic guidelines for schizophrenia, although for ICD, generally, “functional impairment” is still an important clinical “qualifier.” In DSM-5, “functional impairment” is a key criterion of clinical significance and is thus necessary to diagnose the disorder.

“Qualifiers” in ICD-11 are equivalent to “specifiers” in DSM-5. They constitute an important resource in the process of defining and characterizing the clinical conditions of patients with schizophrenia and schizoaffective disorder. This seems particularly useful for treatment formulation, planning, and implementation. There are “qualifiers” for symptoms, course of illness, cognitive function, and functional impairment.

- Symptom qualifiers: These include the presence of positive, negative, depressive, manic, or psychomotor symptoms.
- Course qualifiers: First/only episode versus multiple episodes, acute versus chronic, full remission versus partial remission, acute versus insidious onset.
- Cognitive qualifiers: These qualifiers help to focus attention on cognitive impairment for therapeutic and rehabilitation purposes.
- Functional impairment qualifier: This is consistent with the WHO’s International Classification of Functioning, Disability, and Health (ICF).

SCHIZOAFFECTIVE DISORDER. Schizoaffective disorder continues to pose difficulties in the new nomenclatures because it is not clear how diagnostic characterizations can improve the validity and reliability of this problematic construct for clinical and/or research purposes. In ICD-11, this diagnosis is only made when a patient fulfills criteria for both schizophrenia and a mood disorder of at least moderate/severe degree of severity. The total duration of both components should be of 4 weeks. Apparently, the issue about the need and/or utility of longitudinal criteria for diagnosis has not yet been resolved.

SCHIZOTYPAL DISORDER. Schizotypal disorder is established as a functioning pattern for at least 2 years. Patients are characterized by eccentricities in behavior, appearance, speech, cognitive and perceptual distortions, and reduced ability to establish interpersonal relationships. In ICD-11, this diagnosis is made when the unusual behavior does not meet requirements or does not have sufficient intensity or duration to be diagnosed as schizophrenia.

ACUTE AND TRANSIENT PSYCHOTIC DISORDER. The European perspective on ATPD is different from that of DSM-5, and the former perspective has always been the main influence for the ICD classification of these disorders. In DSM-5, these disorders tend to be viewed more pragmatically as being clinically similar to schizophrenia except for duration. The European tradition views these disorders as a heterogeneous group of psychoses that are different from schizophrenia, both in terms of clinical presentation and pathophysiology. The main clinical features include sudden onset, brief duration, and, most importantly, a polymorphic clinical presentation. Focusing on simplicity and ease of use, it is likely that the "acute polymorphic psychotic disorder without symptoms of schizophrenia" (F23.0) in ICD-10 will be the clinical template for ATPD in ICD-11. Duration up to 3 months will be allowed, as 2 to 4 months is the modal duration for ATPD and other psychoses of sudden onset. When an "acute polymorphic disorder with features of schizophrenia" lasts less than 4 weeks, the diagnosis will be of "unspecified primary psychotic disorder." If the duration is more than 4 weeks, then a diagnosis of schizophrenia must be made. "Schizophreniform disorder" will not be introduced as a category in ICD-11.

Mood Disorders. Mood Episode Descriptions:

Depressive episode
Manic episode
Mixed episode
Hypomanic episode
Mood disorder descriptions

Bipolar and Related Disorders:

6A60 Bipolar type I disorder
6A61 Bipolar type II disorder
6A62 Cyclothymic disorder
6A6Y Other specified bipolar or related disorders

Depressive Disorders:

6A70 Single episode depressive disorder
6A71 Recurrent depressive disorder
6A72 Dysthymic disorder
6A73 Mixed depressive and anxiety disorder
GA34.41 6AZY 6A8Y
Premenstrual dysphoric disorder
Other specified depressive disorders
Other specified mood disorders

In the last few decades, the field of mood disorders has been the scenario of profound changes in conceptualization, and there were many discussions related to their classification and diagnosis in ICD-11. How to improve the detection of bipolar disorders, specifically, how to diminish delayed diagnosis, underdiagnosis, and overdiagnosis, was at the center of active work in this area. Also, there have been attempts to approach the difficult question of how to detect bipolar disorder in front of a first "unipolar" depressive episode. Hypomania has also been at the center of discussions due to the inherent problems of deciding on its presence or absence and the difficulties for retrospective detection. The bipolar distinction between types I and II (and other possible subtypes) was part of many debates. Finally, there was the major issue of what to do (nosologically speaking) with those clinical presentations that include both manic and depressive features (formerly called mixed states in DSM) and which were dropped from DSM-5 and conceptualized as "specifiers." However, the presence of a symptomatic mix in the case of unipolar depression created a problem, as it permitted the "intrusion" of a bipolar presentation as part of a unipolar mood disorder. Passionate debates on this topic have been fueled by the concern that the use of antidepressants in patients with bipolar disorder worsens the clinical course of these patients. There were several articles and editorials in the literature devoted to what "ICD-11 should do" with various mood disorders.

Some of the key ideas from these debates included the following:

- Although the intention is to achieve a higher rapprochement between ICD-11 and DSM-5, there are several important points in which divergence can be expected.
- The general structure of the category appearing in the ICD-11 offers a single diagnostic category, “mood disorders,” subdivided into bipolar disorders and depressive disorders.
- The adoption of dimensional measurements and/or modeling of mood disorders seems to be out of consideration for the future.
- Insofar as the decisions made for ICD-11, some of the consultants involved in the review process appeared to question the scientific validity of the concept of “spectrum” and rejected its inclusion in the classification.
- There were proposals directed toward changes in syndromes such as melancholia, psychotic and catatonic depression, cyclothymia, dysthymia, rapid cycling, etc. However, at present, these proposals were not incorporated into the written document.
- The controversial category “mixed depression and anxiety disorder” is retained because of its clinical utility, which has been reinforced through clinicians’ surveys and field trials.
- Key concepts that were carefully considered by ICD-11 workgroups include bereavement and the diagnosis of major depression, the nosologic status of mixed states, and the presence of hypomanic symptoms in major depression.

THE ISSUE OF BEREAVEMENT AS EXCLUSION CRITERIA. An important debate followed the removal of bereavement as an exclusion criterion for diagnosing a major depressive episode in DSM-5. The position of the ICD-11 developers is that bereavement should continue to be an exclusion criterion for the diagnosis of depression. For example, a key leader of WPA has stated that this change in DSM-5 “is likely to be perceived by the general public as a further step in psychiatry’s attempt to pathologize normal human processes.” In several articles on the subject, some of the authors involved in the ICD-10 revision process declared that “this could be one of the points of disagreement that make it difficult to harmonize the two classifications” and have argued that there is no scientific evidence at present to justify the removal of grief. Indeed, this opinion is also shared by a number of U.S. experts such as Alan A. Horwitz and Jerome C. Wakefield in their book “The Loss of Sadness.”

COURSE SPECIFIERS AND MIXED STATES. One of the most-commented novelties in the mood disorders category in DSM-5 refers to the disappearance of the category of “mixed episode” in bipolar disorders and its new conception as a “specifier.” The ICD workgroup members questioned the diagnostic usefulness of this change and have recommended keeping the criteria for a mixed episode unchanged in ICD-11. There was also little support for the inclusion of “depression with hypomanic symptoms” in ICD-11.

Anxiety or Fear-Related Disorders

6B00 Generalized anxiety disorder
6B01 Panic disorder
6B02 Agoraphobia
6B03 Specific phobia
6B04 Social anxiety disorder
6B05 Separation anxiety disorder
6B06 Selective mutism
6B0Y Other specified anxiety or fear-related disorders

GENERALIZED ANXIETY DISORDER. GAD is one of the diagnostic categories with significant reliability/validity problems, and it is considered a residual category because it is very frequently found along with other disorders. In spite of this, GAD retains clinical relevance and is thought to have a significant impact on health and quality of life. Recommendations for the new GAD diagnosis in ICD-11 included the following:

- A need to find better ways to emphasize “worry about everyday matters” as a core feature of GAD. “Apprehensive expectation,” a concept related to worry, but having several unique features, was also recommended as a core feature for the disorder.
- It was also thought that the presence of “excessive worry” (a construct difficult to operationalize and not always recognized by patients) and its frequency or duration should help define the boundaries between “excessive” and “normal” worrying.

Restlessness, tension, and gastrointestinal symptoms were emphasized as the most frequent and most important somatic complaints in patients with GAD.

- The status of GAD as a “residual category” is being addressed by allowing the diagnosis to be made in the presence of other disorders in which worrying is not the central feature.

PANIC DISORDER AND AGORAPHOBIA. The main changes in ICD-11 included the separation of panic disorder from agoraphobia, as DSM-5 had also done. While this may increase the level of comorbidity, the separation of these two disorders is consistent with the results of research and field studies. An important issue about anxiety disorders is the importance of contextual and culturally mediated influences on clinical presentations such as the contents of worry, the importance attached to worrying in certain cultures, the places, and situations where agoraphobia can occur, the fear of offending others versus being criticized by others and other issues.

SPECIFIC PHOBIAS AND SOCIAL PHOBIA. The subtyping of specific and social phobias in DSM is an important difference with ICD. ICD workgroup members argue that naming specific phobias has no clear clinical utility.

Obsessive-Compulsive or Related Disorders

6B20 Obsessive-compulsive disorder (OCD)
6B21 Body dysmorphic disorder
6B22 Olfactory reference disorder
6B23 Hypochondriasis (health anxiety disorder)
6B24 Hoarding disorder
6B25 Body-focused repetitive behavior disorders
6B25.0 Trichotillomania (hair pulling disorder)
6B25.1 Excoriation (skin picking) disorder

6B2Y Other specified obsessive-compulsive or related disorders
While ICD-11 was being developed, the main issue with respect to OCD was whether or not ICD-11 would follow DSM-5 in creating a separate category for OCD and related disorders. In the end, this was the outcome, with the creation of a separate category grouping obsessive-compulsive

and related disorders. The distinction between “urge” and “impulse” was also debated, in particular its cross-cultural aspects. ICD-11 has a qualifier for insight, separating it into “fair to good” or “poor to absent,” although the clinical utility in identifying a broad spectrum of insight was questioned, as clinicians may be unable to reliably distinguish “poor” insight from “good/fair” or “delusional” insight. ICD-11 did not include a tic specifier as done by DSM-5.

Disorders Specifically Associated with Stress

6B40 Post-traumatic stress disorder (PTSD)
6B41 Complex post-traumatic stress disorder
6B42 Prolonged grief disorder
6B43 Adjustment disorder
6B44 Reactive attachment disorder
6B45 Disinhibited social engagement disorder
6B4Y Other specified disorders specifically associated with stress

QE84 Acute stress reaction

Post-Traumatic Stress Disorder. Changes in the description and classification of PTSD in ICD-11 include the following:

- The creation of a separate diagnostic category for stress-related disorders.
- The discussions about the quality of the traumatic/distressing event related to PTSD and other clinical problems in this new category.
- Increased attention to the distinction between PTSD and normal “adaptive fear reactions” to ongoing trauma (e.g., continuing conflict, forced migration, and natural disasters).
- Inclusion of “complex PTSD.”
- Identifying acute stress reaction as a normal reaction to an abnormal event.
- Emphasizing the advantages of ICD-11 over DSM-5 due to its simplicity, clinical utility, and feasibility in “low-resource and humanitarian settings.”

Intellectual Disability/Mental Retardation. The working group on mental retardation (MR) published several papers on the rationale and process for replacing the ICD-10 conceptualization of MR with the concept of intellectual developmental disorders (IDD) in ICD-11.

For the last two decades, there have been very important discussions on how to properly name, define, and assess IDD. In general, the term “intellectual disability” (ID) has widely replaced MR for policy, administrative, and legislative purposes in many industrialized countries and in an increasing number of low- and middle-income countries. However, the question as to whether IDD represents a disability or a health condition remains unanswered. There was a consensus in the working group about the need to reorganize IDD as part of the larger grouping of neurodevelopmental disorders. The term “intellectual” was favored because, in most countries, it is widely used and is broadly acceptable in many contexts. The term “developmental” refers to the period during which the brain and its functions are developing and thus can be impaired. The term “developmental” also implies a process and a lifespan perspective and emphasizes the dynamic nature of IDD. The proposed definition of IDD is “a group of developmental conditions characterized by significant impairment of cognitive functions, which are associated with limitations of learning, adaptive behavior, and skills.” The main descriptors of IDD include the following:

- Marked impairment of core cognitive functions that can have different patterns. This impairment affects the development of knowledge, reasoning, and symbolic representation, according to the level expected for age peers and the cultural and community environment.
- Problems with verbal comprehension, perceptual reasoning, working memory, and processing speed.
- The cognitive impairment is associated with difficulties in learning, academic, and practical knowledge.
- Problems with adaptive behavior with limitations in relevant conceptual, social, and practical skills.
- Difficulties in managing behavior, emotions, interpersonal relationships, and motivation.
- IDD is a life span condition requiring consideration of developmental stages and life transitions.

It is important to signal that this approach distinguishes between IDD (a clinical meta-syndrome) and ID (the functioning/disability counterpart), conditions that have different scientific, social, and policy applications.

Alcohol-Related Disorders. ICD-11 includes the following alcohol-related disorders: acute alcohol intoxication, harmful use, dependence syndrome, withdrawal, mental disorders, and alcohol-induced disorders arising from the use of alcohol. The new coding for alcohol intoxication defines it as a transient state following alcohol consumption resulting in disturbances in the level of consciousness, cognition, perception, affect, behavior, or other psychophysiologic functions and responses. Typical symptoms include mood instability, false judgment, impaired social or professional functioning, and improper sexual or aggressive behavior. Harmful use of alcohol is defined as a pattern of alcohol use that is causing damage to health following repetitive episodes of intoxication, regular intake of large quantities of alcohol, or harmful use of alcohol. The damage may be physical or mental, including violence and self-harmful acts of bodily damage requiring medical intervention. Alcohol dependence core features would include the following:

- Difficulties in controlling substance-taking behavior are often in combination with a strong desire or sense of compulsion to take alcohol.
- Alcohol use plays a dominant role in the individual’s life.
- Physiologic symptoms: tolerance and characteristic physical withdrawal state or syndrome (not just hangover).

Personality Disorders and Related Traits

6D10 Personality disorder (general diagnostic requirements)
6D10.0 Mild personality disorder
6D10.1 Moderate personality disorder
6D10.2 Severe personality disorder

QE50.7 Personality difficulty
Trait Domain Qualifiers:

6D11.0 Negative affectivity

6D31.1 Detachment

The conceptualization and classification of personality disorders in psychiatry and by psychiatrists is one of the areas of persistent debate among mental health clinicians and researchers. In general, there is a profound dissatisfaction with the psychiatric approach to this problem. Tyrer (2015) summarizes what ICD-11 did in this regard:

- Abolish type-specific categories of personality disorder, apart from the presence of personality disorder itself.
- Define different degrees of severity as a helpful tool to decide on the individual's personality functioning at the time of the assessment.
- Five domains are used as qualifiers: (1) negative affective features (tendency to manifest a broad range of distressing emotions); (2) dissocial features (disregard for social obligations, conventions, and the rights and feelings of others); (3) features of disinhibition (persistent tendency to act impulsively); (4) anankastic features (narrow focus on the control of one's own and others' behavior to ensure that it conforms to the "individual particularistic ideal"); and (5) features of detachment (emotional and interpersonal distance).

Table 9.2–8 shows the category names and essential features of personality disorders in ICD-11.

Feeding and Eating Disorders

6B80 Anorexia nervosa
6B81 Bulimia nervosa
6B82 Binge eating disorder
6B83 Avoidant-restrictive food intake disorder
6B84 Pica
6B85 Rumination–regurgitation disorder

6B8Y Other specified feeding or eating disorders

This is another area of controversy and dissatisfaction for clinicians and researchers. In efforts to improve the state of affairs, a number of proposals have been made as summarized below:

- Feeding and eating disorders were merged into a single grouping.
- In the case of anorexia nervosa, dropping some requirements (e.g., amenorrhea) and broadening the category by expanding other criteria (including developmentally and culturally relevant presentations). A severity qualifier "with dangerously low body weight" was included to distinguish the most severe cases of anorexia nervosa.
- Bulimia nervosa was expanded to include subjective binge eating.

- Binge-eating disorder was included as a specific category. Its definition includes subjective/objective binge eating without regular compensatory behaviors.

Impulse Control Disorders

6C70 Pyromania
6C71 Kleptomania
6C72 Compulsive sexual behavior disorder
6C73 Intermittent explosive disorder
6C7Y Other specified impulse control disorders

Disorders due to Addictive Behaviors

6C50 Gambling disorder
6C51 Gaming disorder
6C5Y Other specified disorders due to addictive behaviors

GAMING DISORDER. Gaming disorder establishes a new behavior pattern leading to the persistence of online or offline gaming. In ICD-11, this diagnosis is made when a patient has impaired control over gaming behavior, and this takes priority over other interests in life and daily activities. This behavior must be manifested for a prolonged period of at least 12 months.

ICD-11 AND DSM-5: SOME CONTROVERSIES

Psychiatric classification is undoubtedly one of the most polemic issues surrounding psychiatry as a medical specialty. The controversies about classification across the two systems can be summarized as follows:

- Epistemologic, for example, what kind of knowledge is gained from classifications, and what are their knowledge limits?
- Metaphysical and ontologic, for example, what is a mental disorder? Are mental disorders natural kinds or constructs?
- Ethical, for example, what are the consequences of using a classification system in terms of the good that can be achieved for patients, society, and psychiatry?

- Practical (usefulness). The usual debate is if classification systems serve the purposes for which they were created, that is, improving communication among clinicians, helping clinical practice, research, and administrative/legal issues.
- Methodologic. What is the best way to classify mental disorders, and how this way can be carried out?
- Sociologic, for example, what are the politics behind classification of mental disorders? How do clinical and political uses of classifications relate to each other? What relationships exist between the classification of mental disorders and vested interests like those of pharmaceutical companies? What is the sociology of the groups developing classifications?
- Legal, for example, how do classifications influence the legal system? How do the needs of legal systems influence psychiatric classification?
- Social, for example, do classification systems promote (or defuse) stigma?
- Personal, for example, do classifications create self-stigma in the patients receiving aid from legal or governmental systems based on a psychiatric diagnosis?

Table 9.2–8.
Category Names and Essential Features of the Proposal for Personality Disorders in ICD-11

Panel 2: Proposed category names and essential features of personality disorders in International Classification of Diseases (ICD)-11

Personality disorder

- A pervasive disturbance in how an individual experiences and thinks about the self, others, and the world, manifested in maladaptive patterns of cognition, emotional experience, emotional expression, and behaviour.
- The maladaptive patterns are relatively inflexible and are associated with significant problems in psychosocial functioning that are particularly evident in interpersonal relationships.
- The disturbance is manifest across a range of personal and social situations (ie, is not limited to specific relationships or situations).
- The disturbance is relatively stable over time and is of long duration. Most commonly, personality disorder has its first manifestations in childhood and is clearly evident in adolescence.

Late onset qualifier

- If the disturbance has its origin in adulthood, the qualifier for “late onset” may be added. The “late onset” qualifier should be used for cases in which, by history, there is no evidence of personality disorder or its early manifestations prior to age 25 years.

Mild personality disorder

- There are notable problems in many interpersonal relationships and the performance of expected occupational and social roles, but some relationships are maintained and/or some roles carried out.
- Examples: Able to maintain, and has some interest in maintaining, a few friends. Intermittent or frequent, minor conflicts with peers, co-workers and/or supervisors or, alternatively, exhibits withdrawn, isolative behaviour but, in either case, is capable of sustaining and willing to sustain employment, given appropriate employment opportunities. Has meaningful relationships with some family members but typically avoids or has conflict with others. Mild personality disorder is typically not associated with substantial harm to self or others.

Moderate personality disorder

- There are marked problems in most interpersonal relationships and in the performance of expected occupational and social roles across a wide range of situations that are sufficiently extensive that most are compromised to some degree.
- Examples: Able to maintain very few friends or has little interest in maintaining friendships. Regular conflict with peers, coworkers and/or supervisors or marked withdrawal and isolative behaviour that interferes with the ability to function constructively at work or with others. May exhibit little interest in and/or efforts toward sustained employment when appropriate employment opportunities are available. May have a history of frequently changing employment as a result. Has conflicted, or a marked absence of, relationships with many family members. Moderate personality disorder often is associated with a past history and future expectation of harm to self or others, but not to a degree that causes long-term damage or has endangered life.
- Examples: Recurrent suicidal ideation or suicide attempts without clear expectation of death, recurrent episodes of self-harm without clear lethality, recurrent hostile and confrontational behaviour, or occasional violent episodes that involve only minor destruction of property (eg, breaking things) or interpersonal aggression such as pushing, shoving, or slapping that is not sufficient to cause lasting injury to others.

Severe personality disorder

- There are severe problems in interpersonal functioning affecting all areas of life. The individual’s general social dysfunction is profound and the ability and/or willingness to perform expected occupational and social roles is absent or severely compromised.
- Examples: Has no friends but may have some associates. Unwilling or unable to sustain regular work due to lack of interest or effort, interpersonal difficulties, or inappropriate behaviour (eg, irresponsibility, fits of temper, insubordination), even when appropriate employment opportunities are available. Conflict with or withdrawal from peers and coworkers. Family relationships are absent (despite having living relatives) or marred by significant conflict. Severe personality disorder usually is associated with a past history and future expectation of severe harm to self or others that has caused long-term damage or has endangered life.

- Examples: Suicide attempts with a clear expectation of death, episodes of self-harm that permanently injure, disfigure or deform the individual, episodes of serious property destruction such as burning down someone’s house in anger, or episodes of violence sufficient to cause lasting injury to others.

Reprinted from Tyrer P, Reed GM, Crawford MJ. Classification, assessment, prevalence, and effect of personality disorder. *The Lancet.* 2015;385(9969):717–726. Copyright © 2015 World Health Organization, with permission of Elsevier. doi: 10.1016/S0140-6736(14)61995-4.

The existence of two main classification systems that have important similarities but equally important differences adds up to these complex controversies. The launching of the DSM-5 has revitalized several discussions both at theoretical and practical levels, one of the most important being the issue of how much ICD and DSM should resemble each other in terms of a theory of mental disorders, aims, scope, methodology, and content, among others. This section will focus on the debates about the issue of harmonization because some think that both classifications must resemble each other as much as possible, but at the same time, both are needed, while others, on the contrary, maintain that there must be a certain distance and/or independency between both, and finally, there are authors who think that at the end one of them should replace the other.

Harmonization

The issue of how much DSM and ICD should resemble each other has recently been discussed under the label of “harmonization.” Despite the aesthetic and rhetorical appeal that this term conveys, the concept needs to be carefully defined and articulated. The Oxford Advanced Learner’s Dictionary defines “harmonize” as follows:

- “To harmonize with something” = “if two or more things harmonize with each other or one thing harmonizes with the other, the things go well together and produce an attractive result.”
- “Harmonize” = “something to make systems or rules similar in different countries or organizations.”

Therefore, at least from the perspective of the current English language, harmonization of the two systems is not simply to combine them but to make them similar and/or equivalent. Because of the international and multicultural character of ICD, the meaning of harmonization was also checked in leading Spanish, French, Italian, and German dictionaries. This reveals semantic differences that should be kept in mind for the harmonization enterprise. In Spanish, for example, emphasis is made on the fact that the parts or objects to be harmonized should not reject each other. In French, the definition emphasizes agreement between elements and the importance of the goal to be achieved. Thus, in all of these languages, harmonization does not simply mean unification but some sort of “shared understanding” and complementarity. However, harmonization is also a technical term used for the main goal of reaching a high degree of similarity between international systems of rules, laws, definitions, or specific criteria.

The Direction of ICD-DSM Harmonization

One of the most sensitive issues about this harmonization project is whether ICD should resemble DSM or DSM resemble ICD. Probably, in some instances, the harmonization should go in one direction (e.g., DSM to ICD), while in others, the opposite would be the case (e.g., from DSM to ICD). The necessary comparison between both systems and the decision that one would prevail over the other adds sensitivity to the process. For example, if it is decided that anxiety disorders are better approached by ICD than DSM, DSM should make an effort to resemble ICD in regard to this group of disorders. Conversely, if the criteria for mood disorders in DSM are deemed better, thus, ICD should adopt them. No matter how logical this sounds, problems and debates will arise among proponents or detractors of either system.

Based on the above considerations, criteria for deciding the direction of harmonization should take into account:

- Methodologic issues: Which of the two systems can ensure that its rationale and criteria are fewer subjects to undesirable and avoidable biases?
- Empirical support: Which of the two systems has been tested by means of well-designed studies (including field trials) whose results support the usefulness, validity, or reliability of the diagnostic categories?
- Ease of use: Which system is easier for clinical use and well liked by busy clinicians?

- Participation of different stakeholders: Which system has duly incorporated the opinions and/or feedback of patients, families, clinicians, and health providers?
- Transnational sensitivity and precision: Which classification is better suited for use across different countries, cultures, and languages?
- Cross-cultural sensitivity and precision: Which classification can be better adapted to different cultural environments?
- Development and advance potential: Which classification has a structure that allows rapid incorporation of changes or scientific advances, especially now that both aspire to be “living documents”?

Clearly, these matters cannot be resolved using only empirical data, and in the end, harmonization becomes a political issue. Here, negotiation and the ability to reach agreements are at the core of the harmonization projects. Thus, both empirical and nonempirical issues have to be addressed in the process of harmonization.

“Harmonization” in the Psychiatric Literature

To gain some insight on this issue, the authors reviewed papers addressing directly (in their title, abstract, and/or keywords) the issue of “harmonization” of the two diagnostic systems. Most of the articles merely state the general importance of “harmonization” but do not provide additional details or recommendations. The emphasis on harmonizing clearly stems from the perceived differences across the two systems that may interfere with communication among clinicians and international research efforts. Clinicians, researchers, and developers of both classifications have all called for harmonization. For most of these parties, harmonization entails diminishing and, if possible, eliminating significant differences between ICD and DSM. Also, there is a general consensus on the importance of harmonization, even in the case of authors who are critical about DSM and/or ICD.

The WHO addressed this issue directly by creating a “DSM-ICD Harmonization Coordinating Group” that also included members of the DSM-5 Task Force and the International Advisory Group revising ICD-11. The charge for this group was “to facilitate the achievement of the highest possible extent of uniformity and harmonization between ICD-11 mental and behavioral disorders and DSM-5 psychiatric disorders and their diagnostic criteria.” Also, it was recommended that “The WHO and APA should make all attempts to ensure that in their core versions, the category names, glossary descriptions, and criteria are identical (author's emphasis) for ICD and DSM."

The major tasks of this group were as follows:

- To identify effective mechanisms to share information on the developmental processes of the two diagnostic systems.
- To find strategies to narrow the differences between the two diagnostic systems.
- To coordinate efforts toward harmonization of the two diagnostic systems by making specific recommendations to the respective decision-making bodies.

To provide a forum in which difficult issues around harmonization can be discussed and common approaches identified.

As articulated before, harmonization and unification are different goals, although they could be part of a sequence, starting with harmonization, and once empirical data is available and conceptual agreements and political consensus reached, unification could follow. Clearly, there are system differences that are more difficult to harmonize, for example, conceptual and ideologic issues, but others, for example, operational ones, can be approached in a straightforward fashion. The process should start by listing and describing the main differences between both classification systems. A second step would entail a detailed analysis of the nature of these differences, for example, which of them are operational, which of them are terminologic, and which of them are conceptual. A third step would be to tackle first terminologic and operational differences and later on focus on the conceptual ones.

Michael First, a North American psychiatrist who has been engaged in the ICD process, proposed three basic harmonization strategies:

- To select one of the two systems as the starting point for a future single and unified classification.
- To create a new common set of DSM-5 and ICD-11 criteria by selecting the best elements from each system in terms of reliability, validity, and clinical utility.
- A third strategy would be to pick the definitions and criteria for a disorder from either one of the two systems and identify it as the "harmonized" DSM-5 and ICD-11 definition for each disorder. For a particular disorder, the decision to choose either the harmonized DSM-

IV or ICD-10 criteria set could be based on several mutually agreed steps, including empirical evidence of superior validity, reliability, or clinical utility.

Harmonization is an important task but remains a very difficult one, especially because of the differences in the origins, purposes, scope, concepts, and ideology underlying both classifications.

SUMMARY AND CONCLUSIONS

- The new ICD-11 section on mental disorders is finally out.
- The ICD-11 appears to remain faithful to the original principles and purposes underlying this international classification.
- It seems clear that changes included in the ICD-11 were mostly driven by clinical utility.
- Harmonization between ICD and DSM was one of the original goals articulated by the developers and potential users of each of these systems. However, significant differences across the two systems still remain.
- Due to the recency of the release of ICD-11, more time will be needed to analyze the impact that this system may have on clinical practice.

FURTHER READINGS

Bornstein RF. Toward a firmer foundation for ICD-11: on the conceptualization and assessment of personality pathology. *Personal Ment Health*. 2016;10(2):123–126.

Brewin CR, Cloitre M, Hyland P, et al. A review of current evidence regarding the ICD-11 proposals for diagnosing PTSD and complex PTSD. *Clin Psychol Rev*. 2017;58:1–15.

Cooper JE, Sartorius N. *A Companion to the Classification of Mental Disorders*. Oxford University Press; 2013.

Evans SC, Burke JD, Roberts MC, et al. Irritability in child and adolescent psychopathology: an integrative review for ICD-11. *Clin Psychol Rev*. 2017;53:29–45.

First MB. Harmonization of ICD-11 and DSM-V: opportunities and challenges. *Br J Psychiatry*. 2009;195(5):382–390.

First MB, Reed GM, Hyman SE, Saxena S. The development of the ICD-11 clinical descriptions and diagnostic guidelines for mental and behavioral disorders. *World Psychiatry*. 2015;14(1):82–90.

Gaebel W, Jessen F, Kanba S. Neurocognitive disorders in ICD-11: the debate and its outcome. *World Psychiatry*. 2018;17(2):229–230.

Goldberg DP, Lam TP, Minhas F, et al. Primary care physicians' use of the proposed classification of common mental disorders for ICD-11. *Fam Pract*. 2017;34:574–580.

Gureje O, Reed GM. Bodily distress disorder in ICD-11: problems and prospects. *World Psychiatry*. 2016;15(3):291–292.

Horowitz AV, Wakefield JC. *The Loss of Sadness*. Oxford University Press; 2007.

Hyland P, Shevlin M, McNally S, Murphy J, Hansen M, Elklit A. Exploring differences between the ICD-11 and DSM-5 models of PTSD: does it matter which model is used? *J Anxiety Disord*. 2016;37:48–53.

International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders. A conceptual framework for the revision of the ICD-10 classification of mental and behavioural disorders. *World Psychiatry*. 2011;10(2):86–92.

Keeley JW, Reed GM, Roberts MC, et al. Disorders specifically associated with stress: a case-controlled field study for ICD-11 mental and behavioral disorders. *Int J Clin Health Psychol*. 2016;16(2):109–127.

Keeley JW, Reed GM, Roberts MC, et al. Developing a science of clinical utility in diagnostic classification systems: field study strategies for ICD-11 mental and behavioral disorders. *Am Psychol*. 2016;71(1):3–16.

Kogan CS, Stein DJ, Maj M, First MB, Emmelkamp PMG, Reed GM. The classification of anxiety and fear-related disorders in the ICD-11. *Depress Anxiety*. 2016;33(12):1141–1154.

Kraus SW, Krueger RB, Briken P, et al. Compulsive sexual behavior disorder in the ICD-11. *World Psychiatry*. 2018;17:109–110.

Krueger RB, Reed GM, First MB, Marais A, Kismodi E, Briken P. Proposals for paraphilic disorders in the International Classification of Diseases and Related Health Problems, Eleventh Revision (ICD-11). *Arch Sex Behav*. 2017;46(59):1529–1545.

Marras A, Fineberg N, Pallanti S. Obsessive compulsive and related disorders: comparing DSM-5 and ICD-11. *CNS Spectr*. 2016;21(4):324–333.

Mataix-Cols D, Fernandez de la Cruz L. Hoarding disorder has finally arrived, but many challenges lie ahead. *World Psychiatry*. 2018;17(2):224–225.

Poznyak V, Reed GM, Medina-Mora ME. Aligning the ICD-11 classification of disorders due to substance use with global service needs. *Epidemiol Psychiatr Sci.* 2018;27(3):212–218.

Reed GM. Progress in developing a classification of personality disorders for ICD-11. *World Psychiatry*. 2018;17(2):227–229.

Reed GM, Drescher J, Krueger R, et al. Disorders related to sexuality and gender identity in the ICD-11: revising the ICD-10 classification based on current scientific evidence, best clinical practices, and human rights considerations. *World Psychiatry*. 2016;15(3):205–221.

Reed GM, First MB, Kogan CS, et al. Innovations and changes in the ICD-11 classification of mental, behavioural and neurodevelopmental disorders. *World Psychiatry*. 2019;18(1):3–19.

Saunders JB, Hao W, Long J, et al. Gaming disorder: its delineation as an important condition for diagnosis, management, and prevention. *J Behav Addict.* 2017;6(3):271–279.

Stein DJ, Billieux J, Bowden-Jones H, et al. Balancing validity, utility and public health considerations in disorders due to addictive behaviours. *World Psychiatry*. 2018;17(3):363–364.

Stein DJ, Kogan CS, Atmaca M, et al. The classification of obsessive-compulsive and related disorders in the ICD-11. *J Affect Disord*. 2016;190:663–674.

Stein DJ, Szatmari P, Gaebel W, et al. Mental, behavioral and neurodevelopmental disorders in the ICD-11: an international perspective on key changes and controversies. *BMC Med.* 2020;18(1):21.

Tyrer P, Reed GM, Crawford MJ. Classification, assessment, prevalence, and effect of personality disorder. *Lancet*. 2015;385(9969):717–726.

Ziebold C, Goldberg DP, Reed GM, et al. Dimensional analysis of depressive, anxious and somatic symptoms presented by primary care patients and their relationship with ICD-11 PHC proposed diagnoses. *Psychol Med*. 2019;49(5):764–771.

10

Neurocognitive Disorders

10.1 Cognitive Disorders: Introduction

IQBAL AHMED, M.D., F.R.C.PSYCH.

Cognitive domains include complex attention, learning, memory, language, perceptual-motor function, executive function, and social cognition as defined in DSM-5. These domains are interdependent in their effects on each other and impact personal, occupational, and social functioning. Disorders that primarily affect one or more of these domains are considered cognitive disorders in DSM-5. While these disorders are mostly associated with older adults, they are not uncommon in younger adults as well, especially those that are related to traumatic brain injury, infections, etc. These disorders may have complex presentations, often with the prominence of other psychiatric symptoms such as mood, anxiety, psychotic and behavioral symptoms, multiple etiologies of the cognitive disorders, presence of other premorbid psychiatric disorders, and medical comorbidity. This can lead to diagnostic and treatment challenges.

Delirium is particularly problematic in this regard. Delirium is often superimposed on an existing, although often unrecognized, dementia and possible multiple risk factors such as multiple neurologic and medical comorbidities and the use of deliriogenic medications. Drs Solai, Sujata, Gebara, and Fabian review the burden generated by delirium in medical settings. Delirium is ubiquitous in acute medical settings, making the older adults and, in particular, those with other cognitive disorders vulnerable. Nevertheless, early recognition and correct differential diagnosis of delirium remain a challenge due to the variable and fluctuating nature of the presenting signs and symptoms. This can be a particular challenge with hypoactive delirium, which sometimes gets misdiagnosed as depression. Screening for delirium, including the use of standardized instruments, is therefore critical, especially in patients with dementia, medically ill hospitalized older adults, and all patients in critical care settings for early diagnosis and intervention. The development of effective interventions for delirium has been slow to emerge, perhaps because of its multifactorial nature. Recent work, which indicates the effectiveness of multicomponent behavioral and supportive interventions in the prevention of delirium, provides some hope and a model for reengineering acute hospital care for at-risk patients. Research also suggests that antipsychotic drugs, which for a long time have been used extensively in the treatment of delirium, do not seem to contribute to any beneficial outcomes, except possibly for the treatment of severe agitation and aggression. Certain dementias, such as dementia with Lewy bodies or late stages of Alzheimer disease, may have chronic clinical presentations virtually indistinguishable from delirium except for temporal onset and the lack of an identifiable acute source.

Julie Graziane and Iqbal Ahmed review recent advances in the current understanding of the neurodegenerative dementias. In addition to the advances in imaging and genetics, great strides have been made in understanding Alzheimer disease, for example, identifying the mechanisms by which accumulation of amyloid beta protein can interfere with cognitive function, even prior to the emergence of the defining neuropathologic lesions, plaques, and tangles. This elaboration of the “amyloid hypothesis” of Alzheimer disease has led to efforts to develop a number of antiamyloid therapies. However, trials to treat patients with dementia due to Alzheimer disease with these agents shown limited efficacy. However the Food and Drug Admininstration (FDA) granted expedited approval to two anti-amyloid agents aducanumab and lecanemab. With an increased understanding of the mechanisms of dementia, the clinical imperative has been to develop approaches for early case identification. Currently, there are efforts to study the role of risk factors for the development of Alzheimer dementia, such as traumatic brain injury, vascular risk factors such as diabetes and hypertension, and inflammation. Research, predominantly epidemiologic, has focused on preventive measures such as physical and cognitive exercises, diets such as the Mediterranean-DASH (dietary approaches to stop hypertension) Intervention for Neurodegenerative Delay diet and the Mediterranean diet, increased social engagement, and improved mood and sleep, among others.

The boundaries between types of dementia can be diffuse. The most common neuropathologic presentations associated with dementia reveal mixtures of Alzheimer disease and vascular and Lewy body pathologies. Pure syndromes are relatively less common, although often the dementia is ascribed to one of the coexisting pathologies, most often Alzheimer disease. Strategies regarding how to understand or reconcile the potential contributions of multiple pathologies to disease onset and progression in the clinic are needed. Some of these are now emerging as the result of the approval of several agents for in vivo imaging of amyloid and tau protein deposits in relation to the evolving symptoms of Alzheimer disease. Tau pathology, while present in Alzheimer disease, also extends to define another set of neurodegenerative dementias that include different forms of frontotemporal dementia (FTD), corticobasal degeneration, and progressive supranuclear palsy. Non-tau, ubiquitin-positive inclusions in the brain are the other major pathologic forms of FTD.

Finally, advances in the genetics of FTD have been quite rapid. Twenty percent of FTD is due to a genetic mutation, and 40% of FTD cases have a family history of dementia, psychiatric disease, or motor symptoms with no clear pattern of inheritance. The three most common mutations are in microtubule-associated protein tau (MAPT), progranulin (GRN), and Chromosome 9 Open Reading Frame 72 (C9orf72). Not surprisingly, given the frequent initial presentation of progressive alterations in behavior, psychiatric syndromes are often initially suspected in individuals with FTDs, with up to 50% of individuals with the behavioral variant of FTD first being diagnosed with a primary psychiatric disorder. The Neuropsychiatric International Consortium for Frontotemporal Dementia (NIC-FTD) developed guidelines to assist with distinguishing FTD from primary psychiatric disorders.

Limbic predominant age-related transactive response DNA-binding protein 43 (TDP-43) encephalopathy, referred to as LATE, is a TDP-43 proteinopathy with or without hippocampal sclerosis pathology is a newly described dementia that typically occurs in patients over the age of 80 as a progressive amnestic syndrome, similar to dementia of the Alzheimer type but with a more gradual decline in cognition.

The course of nearly all subjects developing progressive dementia is complicated by the onset of one or more distinct behavioral syndromes, including anxiety, depression, sleep problems, psychosis, and aggression. These symptoms can be as distressing and disabling as the primary cognitive disorder and are a prominent reason for institutional care. Unfortunately, given the prevalence of these behavioral changes, efforts at the development of new treatments targeting behavioral disruption in dementia have been very limited but are now emerging as a new focus for investigation. So far, pharmacologic interventions have shown to have limited effectiveness and carry significant risk of morbidity and mortality. However, behavioral treatments such as the describe, investigate, create, and evaluate (DICE) method have shown promise. By incorporating the emergence of psychotic symptoms in older adults, even in advance of the development of dementia, into an epidemiologic and neurobiologic framework, new approaches for the diagnosis and treatment of psychosis and agitation are being developed.

Daniel Varon and Carmen Andreescu review the increasing knowledge about mild cognitive impairment, a clinical presentation that often indicates the presence of a neurodegenerative disorder. When MCI presents as an isolated amnestic symptom in late life, it appears to predict the subsequent onset of Alzheimer disease. The main differentiation between MCI and a major neurocognitive disorder due to Alzheimer disease resides in the lack of functional impairment in MCI. Although other conditions, including other neurodegenerative syndromes, focal brain lesions, metabolic disturbances, and alcohol use, may lead to isolated amnestic disorders, in practice, MCI due to underlying Alzheimer disease is more common. Recognition that neurodegeneration may precede clinical symptoms by 10 years or more has further led to the proposal of stages for preclinical Alzheimer disease, characterized by the presence of biomarkers of amyloid and tau accumulation and neurodegeneration in the absence of overt cognitive and behavioral symptoms.

The recognition of a long, preclinical stage of neurodegeneration also indicates that in some older adults who have neurodegenerative pathology and cognitive impairments, the cognitive symptoms might actually result from a comorbid condition independent of or additive with the degenerative lesions. One of the possible treatable conditions leading to an amnestic syndrome is thiamine deficiency presenting as delirium and other symptoms (Wernicke syndrome), secondary to nutritional deficiency in alcohol use disorders, malignancies, malnutrition, and hyperemesis. Treatment with high-dose intravenous thiamine may prevent the development of an irreversible amnestic disorder (Korsakoff syndrome).

Robert Boland and Iqbal Ahmed in the chapter on Cognitive Disorders due to Another Medical Condition or Substance discuss additional medical and substance-induced causes of dementia and amnestic disorders that form the differential diagnoses for many of the conditions reviewed in the earlier chapters. These include neurologic disorders such as traumatic brain injury, strokes, demyelinating disorders and tumors, infectious disorders such as HIV and syphilis, endocrine disorders such as thyroid disorders, nutritional disorders such as thiamine and vitamin B12 deficiencies, and autoimmune disorders such as anti-N-methyl-D-aspartate

(NMDA) encephalitis among others. Substances implicated in the development of cognitive symptoms include prescribed drugs, such as drugs with anticholinergic properties, anxiolytics, sedative-hypnotic agents, and opiates, as well as abused drugs, such as inhalants and stimulants. Optimum care for disorders described in this section requires close collaboration with other medical, neurologic, and rehabilitation specialists.

FURTHER READINGS

Calvin CM, Conroy MC, Moore SF, Kuzma E, Littlejohns TJ. Association of multimorbidity, disease clusters, and modification by genetic factors with risk of dementia. *JAMA Netw Open*. 2022;5(9):e2232124.

Dubois B, Villain N, Frisoni GB, et al. Clinical diagnosis of Alzheimer’s disease: recommendations of the International Working Group. *Lancet Neurol*. 2021;20(6):484–496.

Ducharme S, Dols A, Laforce R, et al. Recommendations to distinguish behavioral variant frontotemporal dementia from psychiatric disorders. *Brain*. 2020;143(6):1632–1650.

Fong TG, Hshieh TT, Tabloski PA, et al. Identifying delirium in persons with moderate or severe dementia: review of challenges and an illustrative approach. *Am J Geriatr Psychiatry*. 2022;30(10):1067–1078.

Ismail Z, Creese B, Aarsland D, et al. Psychosis in Alzheimer disease—mechanisms, genetics and therapeutic opportunities. *Nat Rev Neurol*. 2022;18(3):131–144.

Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*. 2020;396(10248):413–446.

Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain*. 2019;142(6):1503–1527.

Schneider JA, Arvanitakis Z, Bang W, Bennett DA. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology*. 2007;69:2197–2204.

Younes K, Miller BL. Frontotemporal dementia: neuropathology, genetics, neuroimaging, and treatments. *Psychiatr Clin North Am*. 2020;43(2):331–344.

10.2 Delirium

MANDAYAM SUJATA, M.D., MARIE ANNE GEBARA, M.D., TANYA FABIAN, PHARM.D., PH.D., AND LALITHKUMAR SOLAI, M.D.

Delirium is a not-so-uncommon clinical syndrome seen in older adults, particularly with increasing age. Delirium, as defined by the Diagnostic and Statistical Manual of Mental Disorders (DSM)-5-Text revision (TR), is characterized by an acute decline in both the level of awareness and cognition with impairment in attention. The challenge with delirium is that it can mimic other clinical syndromes by presenting as perceptual disturbance, abnormal psychomotor activity, mania and sleep cycle impairment. Due to this and its poor recognition, delirium can significantly complicate patient outcomes.

It is reported that less than 2% of community-dwelling older adults suffer from delirium. But it increases significantly on admission to hospitals ranging in rates between 14% and 24%, which further increases during hospitalization to anywhere between 6% and 56% depending on the medical issue leading to the hospitalization. Unfortunately, despite being such a common occurrence in hospitals, delirium is often underrecognized/unrecognized by medical professionals. It has been estimated that 40% to 60% of the patients (depending on the treatment setting) are misdiagnosed. In most cases, the development of delirium is multifactorial, including stress or trauma to the central nervous system (CNS), drug toxicity or withdrawal, and metabolic disorders arising from organ failure. Despite the numerous etiologies that have been noted, the underlying pathophysiology of delirium is still not clearly understood.

DEFINITION

The term delirium originates from the Latin verb *deliro*—to be crazy, which is taken from de + lira, a furrow (i.e., to go out of the furrow or veer off the track). The core symptoms of delirium include a *disturbance of attention and awareness* (changed from the word consciousness in DSM-IV-TR) that develops rapidly, usually hours to days and tends to fluctuate during the day. It is a common clinical syndrome that typically heralds an increase in morbidity and mortality. Based on a good review of symptoms, there is usually a common cause (but not always) that could explain the change observed in the cognition. When a specific general medical condition has been identified as the cause of the delirium, it is diagnosed as delirium due to a general medical condition, such as delirium due to pneumonia. At times the symptoms of delirium could be secondary to drug intoxication (e.g., cocaine or alcohol) or due to drug withdrawal (e.g., alcohol or benzodiazepine). But most delirious episodes happen in the context of multiple factors, such as dehydration, pneumonia, and pain in a patient who is abusing alcohol. These deliriums are defined as delirium secondary to multiple etiologies. International Classification of Diseases (ICD-10) classifies delirium as either due to (1) alcohol or other psychoactive substances or (2) not induced by alcohol or other psychoactive substances. For consistency reasons, this chapter will not focus on delirium induced by alcohol or other psychoactive substances as it will be covered elsewhere. ICD-10 describes delirium (like DSM-5) as an etiologically nonspecific organic cerebral syndrome with features *impairing consciousness and attention, perception, thinking, and memory*. But it also includes other signs and symptoms as part of the diagnostic criteria for delirium. These include (but are not limited to) the following: psychomotor disturbance, disturbance of the sleep–wake cycle, and emotional disturbance (such as irritability).

Typically, due to the misdiagnosis, patients with delirium tend to remain in the hospital longer and are more commonly discharged to long-term care facilities. Behavioral manifestations of delirium may interfere with treatment compliance and are often precipitants for psychiatric consultation. A consultant psychiatrist should provide and/or advocate for the appropriate treatment beyond simple medical expedience, including prevention of future episodes.

HISTORY

The earliest known references to delirium in medical literature are found in the writings of Hippocrates.

Some 2,400 years ago, in his Book of Epidemics, Hippocrates described the following case:

> Erasinus, who lived near the Canal of Bootes, was seized with fever after supper; passed the night in an agitated state. During the first day, he was quiet, but in pain at night. On the second day, symptoms all exacerbated; at night (mad). On the third day, he was in a painful condition; great incoherence. On the fourth day, he was in a most uncomfortable state; had no sound sleep at night but dreaming and talking; then all the appearances worse, of a formidable and alarming character; fear, impatience. On the morning of the fifth day, he was composed, and quite coherent, but long before noon was furiously mad, so that he could not constrain himself; extremities cold, and somewhat livid; urine without sediment; died about sunset.

Even before the term “delirium” was coined, Hippocrates observed many of the syndrome’s essential features. Tradition holds that the word delirium was first used in the formal medical context by Celsus in the first century AD. However, Celsus used the term delirium to describe a spectrum of mental disorders ranging from general insanity to acute transient states of mental disturbance, including phrenitis, lethargus, hysteria, melancholia, and mania.

Aretaeus of Cappadocia, a contemporary of Celsus, was perhaps the first to classify mental disorders into acute and chronic categories. Aretaeus identified both phrenitis and lethargus as acute manifestations of the disease. The ancient Greek terms phrenitis and lethargus have given rise to the modern English words frenzy and lethargy, respectively. This observation is significant as it introduces the concept that delirium is both an acute syndrome and that it consists of different variants or subtypes. This distinction by Aretaeus may represent the first recorded description of both the hyper- and hypoactive motor elements of delirium.

Several centuries later, Philip Barrough advanced and further clarified the concept of delirium in his 1583 textbook *The Method of*

Physick. Barrough proposed that delirium constituted a derangement of some combination of three main internal senses: imagination, cognition, and memory. This concept of deranged senses was elaborated by Thomas Willis in his 1672 treatise De Anima Brutorum, which established that delirium was, in fact, a syndrome with a specific set of symptoms and not a disease.

In the 18th century, Erasmus Darwin and John Hunter made significant contributions to the evolving theory of delirium. Darwin was the first to compare delirium with the dream state, noting that both states constituted an interruption of “voluntary power” and a suspension of the ability to attend to one’s external environment. Darwin’s notion of the interrelation of delirium and dreams was expanded even further by John Hunter, who defined delirium as “a cessation of consciousness” of one’s own existence.

Later in the same century, James Sims was the first to argue that delirium was distinctly different from general insanity and that it constituted an “alienation of the mind.” Sims also distinguished two variants of delirium, which he referred to as low or raving, that correspond almost perfectly with the modern concepts of hypo- and hyperactive motor subtypes of delirium.

The 19th century saw critical advances in the development of the concept of delirium. A medical dictionary published by Rees in 1818 argued strongly for clarification and unification of the concept of delirium within the medical community. He even went so far as to scold other medical writers who persisted in using the term delirium in an ambiguous sense. Rees built on the work of his predecessor Sims, insisting that delirium be recognized as a distinct condition with two specific variants and a unique etiology that was localized to the brain.

This is the sense in which delirium is commonly understood: but by some writers, delirium is the term employed to denote the whole class of mental derangements, including the different forms of mania, melancholy, and febrile delirium.

Arguably, the most important contribution to the development of the concept of delirium took place in the 20th century when George Engel and John Romano demonstrated that delirium was due to a reduction in the metabolic activity of the brain. Through the use of electroencephalograms (EEGs), Engel and Romano illustrated that decreased background activity corresponded proportionately with a decrease in cognition, memory, and attention. Interestingly, these findings of reduced metabolic activity were observed in all subtypes of delirium, regardless of the level of psychomotor activity.

In introducing a reliable diagnostic tool and demonstrating the underlying mechanism of this mental disorder, Engel and Romano finally confirmed the observations of countless astute clinicians. The history of delirium stretches back over two millennia, yet it is only in the latter part of the past century that psychiatrists have managed to unify its concept and elucidate its mechanisms.

COMPARATIVE NOSOLOGY

Since delirium is usually due to multifactorial causes, there are many synonyms for delirium that have been introduced and adopted by medical practitioners. Various specialties have used different terms, such as encephalopathy and acute brain syndrome. Unfortunately, these multiple terms falsely suggest an element of variation that further complicates the recognition and management of delirium within and across medical specialties. In psychiatry, the term acute confusional state has been used as an equivalent designation for delirium, despite the ambiguity this additional term creates.

In general medicine and many of the medical subspecialties, the terms encephalitis and encephalopathy prevail in both the medical literature and clinical practice. Encephalitis refers to an inflammation of the brain parenchyma, as distinct from the meninges, which is usually associated with an infectious state. The term encephalopathy is less specific and refers to any disorder of the brain parenchyma yet is most often associated with organ failure and metabolic disorders. These examples illustrate the diversity of terms that are employed to identify delirium, and many other examples abound.

Although many of the aforementioned terms accurately describe individual elements of delirium, they fail to present a comprehensive summary of the patient. To be accurately diagnosed and managed, all aspects of the delirium, from etiology to resultant symptoms, must be included in the assessment. This comprehensive evaluation requires a consistent nosologic system. Great strides have been made in the establishment of such a system since the American Psychiatric Association's first attempts to clarify the concept of delirium in the third edition of the DSM. The 10th edition of the ICD classification of delirium (F05) now has many similarities with DSM-5.

The confusion surrounding the use and definition of the term delirium is further compounded by the existence of clinical subtypes referred to as the hyperactive, hypoactive, and mixed states of psychomotor activity. It should be noted that these designations are strictly clinical and may represent different manifestations of the same pathophysiologic state, or one type or another may be more commonly seen depending on the underlying etiology. Misconceptions regarding all these terms are a significant concern as they may result in a failure to correctly diagnose and manage the delirious patient.

ETIOLOGIES

Delirium in older adults is often multifactorial and can occur in a multitude of settings. Up to half of hospitalized older adults develop delirium. There are different causes of delirium that can affect older adults, including the emergence of delirium (from anesthesia) postoperatively, intensive care unit (ICU) delirium, delirium at the end of life, as well as delirium superimposed on dementia.

The multifactorial model for delirium has been widely accepted. It highlights the complex relationship between predisposing and precipitating factors. Predisposing factors are baseline vulnerability factors, and precipitating factors initiate or lead to the development of delirium. Individuals with less predisposing factors require more stressors to induce delirium, while highly vulnerable patients require much less noxious insults to develop delirium. Current management approaches to delirium focus primarily on the precipitating factors. Managing predisposing factors of delirium becomes essential in preventing delirium and decreasing the risk of future episodes of delirium and the morbidity/mortality associated with it.

Baseline or predisposing factors that increase the risk of developing delirium include age and dementia. Reduced functional reserve, whether physical or cognitive, results in a poorer response to stress. It is postulated that age results in physiologic brain changes leading to diminished cognitive reserve. Cognitive impairment and dementia are significant risk factors for the development of delirium postoperatively. In fact, there is a bidirectional relationship between delirium and cognitive decline. Other predisposing risk factors include a history of delirium, functional impairment, vision and hearing impairment, depression, and alcohol abuse.

In hospitalized patients, precipitating factors include physical restraints, malnutrition, the addition of more than three medications, use of bladder catheters, and iatrogenic events such as hospital-acquired infections or medication-related complications and electrolyte abnormalities. Medications that can precipitate delirium include sedative–hypnotics such as benzodiazepines, propofol and barbiturates, anticholinergic agents, analgesics (such as opioids), histaminic agents, steroids, and dopamine agonists. Disruption of circadian rhythms and insomnia predispose the development of postoperative delirium. Pain is considered a stressor and directly increases the risk of postoperative delirium in older adults.

End-of-Life Delirium

End-of-life delirium is also known as terminal restlessness/delirium. End-of-life delirium makes treatment of terminally ill patients particularly challenging. It will most often manifest in the last 24 to 48 hours of life but can develop weeks before death. The prevalence of delirium in palliative care inpatient units can be as high as 88% in patients with advanced cancer. This is likely due to the accumulation of multiple contributing factors (such as multiorgan failure, opioid use, dehydration, etc.). Delirium can be reversible in almost half of these episodes when precipitating factors include dehydration or psychoactive medication use. However, in cases of refractory or irreversible delirium, palliative sedation is indicated at the end of life. End-of-life delirium can be particularly distressing to patients and, more so, their families. As such, it is a common reason for admission to inpatient hospice. The precipitating factors for terminally ill patients are similar to other factors in other populations, including anticholinergic medications, anxiolytics, opioids, steroids, infection such as pneumonia or urinary tract infection (UTI), electrolyte or fluid imbalance, as well as poor pain management and constipation.

ICU Delirium

The ICU setting has some of the highest rates of delirium across all health care settings, particularly in patients on mechanical ventilation. The predisposing and precipitating factors for ICU delirium are similar to delirium in other settings, with the addition of mechanical ventilation. The predisposing factors such as age, cognitive disorder, frailty, alcohol or drug use, and severe underlying disease are all significant factors. Precipitating factors include metabolic disorder, in addition to sepsis, hypotension, inadequate pain management, sleep disorders, and complicated surgery. Similar to other settings, drugs that can precipitate delirium include benzodiazepines, opioids, anticholinergic drugs, and steroids. In addition, deep sedation (compared to light or interrupted sedation) is a risk factor. Morphine and meperidine have long half-lives and tend to accumulate, making short-acting opioids a preferred choice. The ABCDE bundle approach has been recommended for ICU delirium. It consists of assessing pain, daily trials of spontaneous awakening and spontaneous breathing trials, choosing shallow sedation if possible, monitoring delirium, early mobility and exercise, and family engagement and empowerment. Hypoactive delirium appears to be more common in the ICU than hyperactive or mixed delirium.

Postoperative Delirium

Postoperative delirium most often occurs within the first 2 days and can occur in up to half of all older adults following a hip fracture or cardiac surgery. Delirium that develops later is often associated with postoperative complications or withdrawal from alcohol or other medications. Looking back at the multifactorial model for delirium, anesthesia and surgical procedures are significant precipitating factors.

Coronary Artery Bypass Graft

The incidence of delirium after coronary artery bypass graft (CABG) procedure has ranged from 3% to 35%. Several factors play a role in the development of delirium in patients with post-CABG. A systematic review that examined data across 97 studies and pooled 60,479 patients who underwent CABG identified several nonmodifiable risk factors for delirium. Significant risk factors include (1) preoperative cognitive impairment, depression, stroke history, and higher European System for Cardiac Operative Risk Evaluation (EuroSCORE) score, (2) intraoperative increase in intubation time, and (3) postoperative presence of arrhythmia and increased days in the ICU. On the other hand, higher preoperative cognitive performance was protective for delirium.

Hip Replacement

Delirium is a common adverse event in patients undergoing hip replacement, with estimates in the range of 15% to 60%. Several studies have indicated that post–hip replacement delirium has been associated with increased morbidity and mortality. A systematic review that pooled 5,364 patients with hip fracture showed a 24% incidence rate of delirium after hip fracture repair. The risk factors for developing delirium included preoperative cognitive impairment (odds ratio [OR] 3.21, 95% confidence interval [CI] 2.26–4.56), advanced age (standardized mean difference 0.50, 95% CI 0.33–0.67), living in an institution (OR 2.94; 95% CI 1.65–5.23), heart failure (OR 2.46; 95% CI 1.72–3.53), total hip arthroplasty (OR 2.21; 95% CI 1.16–4.22), multiple comorbidities (OR 1.37; 95% CI 1.12–1.68), and morphine usage (OR 3.01; 95% CI 1.30–6.94). Female patients were less likely to develop delirium after hip surgery (OR 0.83; 95% CI 0.70–0.98).

Substance Intoxication

Substance intoxication and substance use withdrawal can both present with delirium. These will be covered in detail elsewhere. However, since the number of older adults in the United States with substance use disorder has been projected to almost double in the next 20 years, they will be briefly covered here.

The diagnosis of substance-induced delirium is made when a medication or other chemical agent is thought to be etiologically related to the condition. The responsible substance may be determined from clinical history, physical examination, and laboratory tests, such as a toxicology screen. The goal of treatment is to remove the offending agent. Familiarity with the common toxicologic syndromes (also known as toxidromes) is useful, as it may guide the investigation to include screening for additional agents. Awareness of the local prevalence of various drugs of abuse can also give some indication of the etiology of the syndrome. Clinicians need to keep current with drugs of abuse, newly approved drugs, herbal preparations, and teas as potential causes of delirium.

Intoxication with a variety of drugs of abuse is known to cause delirium. Examples include cocaine, PCP, heroin, alcohol, nitrous oxide, amphetamine and its derivatives (e.g., speed and ecstasy), 3,4-methylenedioxymethamphetamine (MDMA), and marijuana. Flunitrazepam (Rohypnol) and γ-hydroxybutyrate (GHB) are other common choices of recreational drug users with the potential for delirium.

Substance Withdrawal–Induced Delirium. In addition to substance intoxication, acute change in mental status can herald a substance withdrawal delirium. The diagnosis may be supported by autonomic changes or the detection of a substance or its metabolite in the urine or serum. When reviewing the history and chart of a patient presenting with delirium, a thorough review of medications and substances should be conducted to include medications that are prescribed and over the counter as well as the use of illicit substances and any recent changes. This is especially true when a patient is admitted emergently into the hospital and is unable to give a reliable history.

Alcohol Withdrawal. Alcohol withdrawal is associated with the classic agitated withdrawal delirium, also known as delirium tremens. This syndrome, synonymous with severe alcohol withdrawal, presents with delirium, autonomic hyperactivity, and frequent visual and tactile hallucinations and carries a significant risk of seizures and death if untreated. Patients admitted to the hospital may present several days into admission with alcohol withdrawal and may be at risk of delirium tremens. Histories of complicated alcohol withdrawal, medical comorbidity, and elevated blood pressure are associated with an increased risk of delirium tremens. In addition, benzodiazepine withdrawal may appear similar to alcohol withdrawal and may also be accompanied by seizures. The onset of symptoms is dependent on the half-life of the particular benzodiazepine.

Opiate Withdrawal. Opiate withdrawal presents with severe flu-like syndromes, gastrointestinal (GI) cramping, diarrhea, diaphoresis, autonomic hyperactivity, and craving. Delirium may also occur with opioid withdrawal and has been reported when patients are switched from transdermal fentanyl (Duragesic) to morphine (Duramorph). The presentation may occur because of unknown prior use of opiates in the context of abuse or prescription or methadone (Dolophine) maintenance.

DIAGNOSIS

Delirium, despite its common presentation in older adults, is frequently missed in various clinical settings. Fluctuation over hours in the presentation of symptoms further complicates the identification of delirium. Some studies have indicated that up to two-thirds of delirium cases have been missed by clinicians. Obtaining a good history from the patient’s caretaker is an essential step in diagnosing delirium. In fact, neglecting this step is the leading reason for missing the diagnosis of delirium. If no history can be obtained, the patient should be assumed to be delirious until proven otherwise. Older patients should be aroused during rounds and evaluated daily for the hypoactive form of delirium, which is subtler and often overlooked.

Approaches to Diagnosis

When clinicians search for the underlying cause of delirium, they need to be aware of the possibility of occult or atypical presentations of many diseases in the older adults, including myocardial infarction, infection, and respiratory failure, like delirium, which is often the sole manifestation of serious underlying disease. During the current COVID pandemic, reports indicate that delirium may be a symptom of COVID in older patients. All preadmission and current medications, including longstanding ones should be reviewed for indication and risk/benefit. These can contribute to delirium and should be reevaluated. If changes in long-term medications are appropriate, the hospital represents a safe and ideal venue for making these changes. A complete medical history must be meticulously obtained to detect occult alcohol or benzodiazepine use, which can contribute to delirium.

Rating Scales in Delirium

Delirium is a diagnosis that is detected through a good review of signs and symptoms and bedside cognitive assessment. But certain instruments can help or assist in standardizing the approach to the diagnosis. Every older hospitalized patient should undergo brief but formal cognitive testing with the use of instruments such as the Mini-Mental State Examination (MMSE) and the Confusion Assessment Method (CAM). CAM primarily focuses on four major clinical features of delirium, including acute onset, inattention with fluctuations, disorganized thinking, and altered level of consciousness. The full CAM also looks for evidence of disorientation, memory impairment, psychomotor changes, altered sleep–wake cycle changes, and perceptual disturbances. CAM has been validated in a number of settings, such as inpatient units, emergency rooms, and the ICUs. In mechanically ventilated older patients, the instrument evidenced high rates of validity when compared to expert clinicians using DSM-IV-TR criteria with 95% sensitivity and 88% specificity. The administration time is usually about 2 minutes.

PATHOLOGY AND LABORATORY EXAMINATION

The utility of extensive routine laboratory testing in delirium has not been demonstrated. In many instances, the cause of delirium is obvious and often predictable, particularly when a good review of systems is conducted with the patient and or caregiver. In most patients, only testing what is necessary to confirm the diagnosis might be performed. In other situations, delirium may be the initial manifestation (or the only sign) of an underlying illness. The clinician must decide whether the known acute illness (such as pneumonia or hyponatremia) is sufficient to cause delirium, often a difficult decision. Some investigators recommend extensive enough lab or radiographic testing to exclude the common causes of delirium. A chest roentgenogram should be considered because pneumonia is one of the most common causes of delirium in older patients, and typical symptoms may be absent. In the current pandemic, testing for COVID is seen as a standard of care in such patients. An electrocardiogram (ECG) may be helpful in detecting cardiac ischemia. If hypoxia is suspected, pulse oximetry could be done, especially in postoperative patients. Arterial blood gases are most useful in patients with suspected acid–base disturbances, hypercapnia, and severe hypoxia.

Blood Chemistry, Urinalysis, and Cerebrospinal Fluid Examination

When routine testing is considered, then it should be individualized and limited initially to basic blood chemistry studies that assess for electrolyte, glucose, calcium, renal, and hepatic abnormalities and complete blood cell counts that assess for anemia or leukocytosis (indications of possible infection). Additional tests should be considered as directed by findings from the history and physical examination or if the cause remains elusive after the initial assessment.

Results of urinalysis that suggest the presence of infection should be followed with cultures and sensitivity studies. Caution should be used when ordering urinalysis in a nursing home setting patients as there is a big initiative for antibiotic stewardship and proper identification of true UTI with fever, burning or pain during micturition, increased WBCs with neutrophils. This has to be balanced with the knowledge that these patients may have compromised immunity and may not evidence all the classic signs and symptoms of UTI except for the sudden onset of mental status change or delirium. Determination of drug levels and toxicology screens may be necessary for selected patients.

Most hospitalized older patients with fever and delirium, however, have primary causes outside the CNS and do not require routine cerebrospinal fluid (CSF) studies. A CSF examination is important in the patient with delirium and fever in whom encephalitis or meningitis is suspected. Infection, dehydration, and metabolic disturbances accounted for 97.5% of cases of fever and delirium in one study, which found only one case each of bacterial meningitis and aseptic meningitis. One other important piece of information to remember while investigating the causes of delirium is that many times there may be multiple causes of delirium than just one factor.

Electroencephalography and Neuroimaging

Again, delirium is purely a clinical diagnosis. When that is in doubt, an EEG may be useful to clarify the diagnostic issues; however, caution must be used as EEG has a false-negative rate of 17% and a false-positive rate of 22%. Most patients with delirium will demonstrate diffuse slow-wave or low-voltage activity. The EEG may be useful in differentiating delirium from psychiatric disorders and identifying certain diagnoses, including hepatic encephalopathy, uremia, CNS infections, and seizures. In some patients, the EEG findings may indicate a focal rather than a global impairment and may thus suggest a localized cerebral event.

Neuroimaging as a routine test in delirium has a very low yield. When it is used, it is preferable to go with computed tomography (CT) head before considering magnetic resonance imaging (MRI) or functional studies. Neuroimaging studies should be reserved for patients with new focal neurologic signs, those with (a) history or signs of head trauma, (b) fever and acute changes in mental status in whom encephalitis is suspected, or (c) no other identifiable cause of the delirium. Neuroimaging may also be considered when the history cannot be obtained or the neurologic examination cannot be completed (e.g., when the patient is combative) so as not to miss uncommon conditions that are life-threatening but treatable such as subarachnoid hemorrhage and encephalitis. In older patients with a history of falling, neuroimaging such as CT head may be important for ruling out subdural hematoma (even if a head injury was not reported). If signs of focal neurologic impairment are present, CT or MRI of the head is indicated. Functional imaging studies such as blood flow, single-photon emission computerized tomography (SPECT), or positron emission tomography (PET) have yielded important findings on the pathophysiologic features of delirium; however, their role in the evaluation of delirium is unclear.

DIFFERENTIAL DIAGNOSIS

As mentioned earlier in this section, a good clinical history, physical examination, and laboratory studies are helpful in distinguishing delirium from other causes of global cognitive impairment, such as dementia, depression, and functional psychosis. Dementia (covered elsewhere within this section) has a gradual onset, persists for more than 1 month, is usually progressive, and is not associated with reduced alertness until terminal stages. Dementia, however, predisposes to delirium. Clinicians have to be very alert in picking up on delirium through the sudden deterioration of cognitive or functional ability in a patient with dementia. A delirium that is superimposed on a dementia process seen for the first time may be difficult to distinguish from a natural progression of the cognitive impairment or an atypical dementia. Considerable overlap exists between the two conditions, and the diagnosis of dementia should be deferred until the delirium clears and an appropriate evaluation has been done after several months have passed.

Hypoactive delirium can easily be mistaken for depression. Social withdrawal, limited speech, apathy, and poor attention with impaired cognitive testing could occur in both clinical situations. But typically, in depression, there is no change in the consciousness level, and there is a lack of fluctuation of the clinical presentation as seen in delirium. Acute psychosis can mimic a hyperactive delirium, but hallucinations tend to be typically auditory, delusions are more systematic and well formed, and the symptom fluctuation so characteristic of delirium is not noted. EEG findings are usually normal in subjects with psychosis or depression in comparison with the diffuse slow-wave activity seen in patients with delirium. Similarly, anxiety, a state of profound apprehension with increased motor activity and autonomic arousal, can resemble hyperactive delirium. Once again, there is a lack of change in the level of consciousness, and there is limited fluctuation in the presentation. Anxiety is common among hospitalized patients, particularly postoperatively and in ICUs, where fears of pain, disability, and death can be exacerbated by sensory impairment, sleep deprivation, current pain, and medications. It is imperative to clarify the diagnosis of anxiety versus delirium as treatment of one may adversely affect the other. For example, an agitated patient with hepatic encephalopathy can experience worsening if treated incorrectly for anxiety with benzodiazepines.

Diagnosing Delirium in Special Populations

Delirium in Patients with Dementia. Dementia is the biggest risk factor for delirium, and it has been reported that close to two-thirds of the cases of delirium occur in patients with dementia. There is a strong relationship between dementia and delirium that is not yet fully understood. The underlying vulnerability of the brain in patients with dementia may predispose them to the development of delirium as a result of insults related to acute medical illnesses, medications, or environmental perturbations. Some Studies suggest that delirium could persist for months or even years in some patients. The presence of such types of persistent delirium blurs the boundaries between delirium and dementia further. Some argue that delirium and dementia may not be independent disorders but could represent points along the continuum of cognitive disorders. This is substantiated by similar changes seen in cerebral metabolism (decrease), cholinergic deficiency, and certain inflammatory changes between these two disorders.

The presence of baseline dementia in a patient makes it challenging to detect delirium. Studies have shown that the severity of dementia can also influence the prevalence of delirium, meaning the more severe dementia, the higher the prevalence of delirium.

One important question that always comes up when treating an older patient with delirium is “Does delirium contribute to dementia”? The simple answer is that delirium does not cause the changes seen in dementia but does contribute to the worsening of the overall status of a patient with dementia. In the past, it was thought that if the underlying causes of delirium were treated well, then the subject’s premorbid cognitive and functional status could be restored. Several recent findings, such as long-term cognitive decline in patients with delirium seen in an epidemiologic study and some of the irreversible causes of delirium that result in neuronal damage indicate otherwise lately.

Families typically note that patients never return to baseline after an episode of delirium. This has been well noted in older patients with dementia. Delirium tends to worsen the trajectory of cognitive decline in dementia patients, leading to an increased rate of hospitalizations, institutionalization, and also increases the mortality risk by at least twofold in one study.

Delirium in Patients with Serious Mental Illness. Since delirium can present with psychosis and disturbance of sleep issues, it can confound the diagnosis when it presents in patients with schizophrenia and bipolar disorder. Only 48% of delirious patients were actually recognized as having delirium at the time it occurred. Patients with these disorders are at higher risk for delirium due to their comorbid medical issues and medications (many anticholinergic). In one study on the prevalence of delirium in psychiatric inpatients, the overall incidence of delirium was 14.6%. Delirium was most common among patients with schizophrenia and bipolar disorder; patients with bipolar disorder had the highest incidence (35.5%). Older age was the only significant demographic risk variable. The hospital stays of patients with delirium were 62.1% longer than those of patients without delirium.

Delirium in Patients with Terminal Illness. Delirium can complicate the clinical presentation in terminally ill patients. As the goals of care change in these patients, there is an urgency to better manage pain with the use of pain medications, treating infections, and avoiding hospitalizations. In one study, 228 participants (49.9%) among 457 terminally ill inpatients were screened. The prevalence of delirium was 46.9% (n = 107). Of these, the most common subtype was hypoactive (68.2%). The mortality rate of inpatients with delirium (77.6%) was higher than those without delirium (50.9%). The overall detection rate by any member of the palliative care team was 44.9%. The detection rate of the hypoactive subtype was only 20.5%, which was significantly lower than that of the hyperactive/mixed subtypes. Therapy for delirium was prescribed in 42.1%, with haloperidol being the most common medication. It has been estimated that approximately 50% of cases of delirium occurring in patients receiving specialist palliative care are potentially reversible.

Delirium in Children. The clinical features of delirium are not well known in children as in adults. One study in more than 100 consecutive psychiatric consultations over 4 years showed 84 patients aged 6 months to 18 years were identified with delirium. These patients had a longer length of stay and higher mortality (20%). Overall, the presentation and course of delirium were similar in adults (except for psychosis and disorientation), and the DSM criteria were found applicable in the pediatric population.

In another study, children had more sleep–wake disturbance, fluctuating symptoms, impaired attention, irritability, agitation, affective lability, and confusion as compared to adults. Adults had more impaired memory, depressed mood, speech disturbance, delusions, and paranoia. Impaired alertness, apathy, anxiety, disorientation, and hallucination occurrence were similar between both groups. These may represent true differences in the presentation of delirium across the life cycle or may be attributable to inconsistent methodologies. Prospective studies are needed to resolve these questions.

The classic onset of delirium is rapid—over a few hours or days—and the symptoms can be highly variable and intermittent. Various team members and families may witness completely different behaviors within a brief time span. A patient can go from being drowsy to being wakeful to then being hypervigilant within a few hours. Delirious patients can present with manic-like symptoms (poor sleep, pressured speech, high energy, agitation), depression-like symptoms (withdrawn, isolative, poor motivation, low energy), anxiety-like symptoms (anxious, poor sleep, restlessness, agitation), sleep disorder–like symptoms (reversed sleep–wake cycle, insomnia, hypersomnolence), and schizophrenic-like symptoms (hallucinations, delusions, paranoia, thought disorganization). As a result, the early detection diagnosis of delirium may be overlooked. Alternatively, caregivers might recognize that the patient is confused but fail to appreciate the significance of this change in condition, particularly when they seem to be less confused at times. Consequently, the problem may not be addressed until further deterioration ensues.

A cardinal feature of delirium is reduced alertness, manifested by difficulty maintaining or shifting attention. The delirious patient cannot ignore irrelevant stimuli and is easily distractible. Since many of the cognitive functions depend on attention, they tend to perform poorly in many cognitive tasks. The impaired memory seen in delirium usually represents an inability to register recent information because of inattentiveness, but retrieval of older information can also be disturbed. Higher integrative functions are similarly affected; the result is a reduced ability to plan, solve problems, or guide actions. Due to this, a typical new-onset delirious patient could be much worse cognitively than a patient suffering from dementia for a few years. These disturbances in attention often increase toward evening, a characteristic known as sundowning, which may result from fatigue and reduced sensory input. The confused patient cannot plan cognitive functions appropriately. Consequently, subjects demonstrate incoherent speech and a diminished ability to sequence actions properly (e.g., rising from a bed or walking), which can lead to injury. Typically, sleep becomes brief and fragmented, and fatigue tends to aggravate the symptoms of delirium. Subjects often demonstrate a reversal of sleep patterns, with daytime napping and nocturnal wakefulness.

Several clinical subtypes of delirium have been described. Commonly, delirium is classified on the basis of the level of psychomotor activity. Although certain types of delirium may frequently occur in patients with particular disease states, they are neither exclusive to nor diagnostic of specific underlying medical conditions. Similarly, the manifestation of delirium cannot be fully predicted by the presence of a particular etiologic toxin or illness. The hyperactive (increased psychomotor activity) variant of delirium is most commonly recognized and tends to be readily apparent even to the casual observer. It is often associated with the adverse effects of anticholinergic drugs, drug intoxication and withdrawal states. Characteristically, patients may exhibit agitation, psychosis, and mood lability and may refuse to cooperate with medical care, may demonstrate disruptive behaviors (such as shouting or resisting), or may sustain injuries from falling, combativeness, or pulling out catheters and intravascular lines.

The hypoactive (decreased psychomotor activity) variant of delirium is more common than hyperactive delirium in older patients, but it is less frequently recognized. Often it may be misattributed to depression or dismissed as a transient, insignificant problem. Metabolic conditions such as hepatic or renal encephalopathy are frequently associated with hypoactive delirium, but other causes are also common. Patients with hypoactive delirium may appear sluggish and lethargic as well as confused. The absence of disruptive, bizarre, and injurious behaviors contributes to the lack of recognition. In some studies, it has been noted that up to two-thirds of the patients with delirium could be unrecognized or underrecognized. This quiet state of withdrawal, apathy, and clouded inattention differs from the drowsiness seen in persons dozing or sleeping in that these latter subjects can be quickly aroused to (and remain in) normal consciousness by mild stimuli. In hypoactive delirium, however, strong stimuli (e.g., vigorous shaking or shouting) are often needed for arousal, which is incomplete and transient at best.

Because of the multiple etiologic factors, the fluctuating course, and the individual medical comorbidities, many patients who experience delirium have a mixture of both hypo- and hyperactive variants. Such patients are thought by some researchers to be at the highest risk of substantial morbidity and mortality. In one recent study conducted on critically ill patients, mixed type was most common (54.9%), followed by hypoactive delirium (43.5%) and purely hyperactive delirium (1.6%). In a post–hip surgery population, 47% were the hyperactive type of delirium, and both hypoactive and mixed types were each seen in 26% of the group. Similar rates were reported in a patient group admitted to an inpatient geriatric ward.

Hallucinations, misperceptions, illusions, and delusions are reported to occur in at least 40% of cases of delirium and can accompany both hypo- and hyperactive subtypes. Hallucinations are usually visual, ranging from dream-like experiences to terrifying visions, and may, in fact, represent “sleep without muscle atonia.” Auditory hallucinations, as well as those involving taste and smell, also occur, although much less frequently. Delusions are often paranoid or persecutory (e.g., suspicion of poisoned food or fear of intended harm by nursing staff).

Although these are some of the most distressing symptoms, the presence or absence of perceptual disturbances is not helpful in determining the cause or level of severity of delirium. Their presence, however, does confirm the diagnosis of an acute delirious state. Of importance, these sensory features tend to be less common in older than in young patients. Rather, somatic features such as urinary incontinence, gait impairment, tremor, and language disorders (including receptive and expressive aphasias) tend to predominate.

COURSE AND PROGNOSIS

Symptoms of delirium usually last about 7 days in most patients, but there is much variability, with slow resolution of symptoms contributing to persistent symptoms of delirium at 6 to 8 weeks for severely ill patients. In one study, it was reported that 80% of all delirious patients recovered by 7 to 10 days, but 5% of them do end up having the episode beyond 1 month. Symptom resolution is frequently incomplete by hospital discharge, with as many as 15% of patients remaining symptomatic of delirium at 6 months. There have been numerous methodologic limitations of some studies examining long-term follow-up. The presence of dementia and the level of medical comorbidity were not always controlled for, and the sample sizes were frequently small. In general, studies suggest that the increased mortality risk associated with delirium was maintained at 12, 24, and 36 months, with a risk ratio of at least 2 at all time points. In addition, at 24 months, the increased risk of cognitive and functional impairment remained. One study followed older patients discharged from an emergency room who were delirious. The mortality outcome at 6 months postdischarge for delirious patients not identified by the emergency room treating team was three times higher than the delirious patients who were identified and treated in the emergency room. The mortality rate was similar for the treated delirious patients as the nondelirious patients.

Early recognition of delirium is hypothesized to improve prognosis by allowing clinicians to address underlying medical conditions and substances earlier in the course. Surprisingly, preventive measures to improve prognosis have demonstrated mixed results. This may be due to insufficient exposure to staff, poor retention of the information presented, or other patient and clinician factors. The outcome measure used in these trials included time to delirium resolution, increase in independence in activities of daily living, decreased length of stay, and increased rate of discharge to the community rather than institutional settings. For example, one positive study demonstrated that the presence of a liaison psychiatrist in an orthopedic unit reduced the frequency of delirium and decreased the length of stay for the patients in the intervention group. Another study, which looked at whether a geriatric consultation followed by the involvement of a liaison nurse with the consultant and primary team, failed to demonstrate that the intervention had any benefit. An educational intervention to physicians and nurses by means of distribution of guidelines for recognizing delirium and formal educational in-services about the guidelines failed to improve outcomes in acute care hospitals in England.

Recent reports of multidisciplinary educational interventions have shown some benefit in decreasing the new onset of delirium and the length of stay in delirious patients in medical units. In another report, in an academic medical center in the United States, a multicomponent interventional strategy failed to reduce the rate of delirium, placement into nursing homes, need for home care visitation, readmission rates into the hospital, and self-rated health measures, and failed to improve functional status. Adherence to nonpharmacologic interventions for delirium was directly related to patient outcomes in this multicomponent intervention in the United States.

PREVENTION

Preventing delirium is the most effective strategy for reducing its frequency and complications. Successful prevention strategies include recognition of predisposing factors and early intervention to address or remove precipitating factors. The Hospital Elder Life Program (HELP), a model of care designed by S.K. Inouye provides a comprehensive patient-centered approach to early identification of delirium risk factors and targeted interventions to address cognitive impairment, sleep deprivation, immobility, dehydration, and vision or hearing impairment. HELP has been implemented in over 200 hospitals worldwide and has shown to be both effective and cost-effective in reducing delirium and improving outcomes for hospitalized elders. Similarly, the Yale Delirium Prevention Trial demonstrated the effectiveness of nonpharmacologic interventions to reduce the risk of developing delirium, including orientation and therapeutic activities for cognitive impairment, methods to improve communication, use of eyeglasses and hearing aids to reduce vision and hearing impairment, and employment of nonpharmacologic interventions to prevent sleep deprivation, volume depletion, and immobility while minimizing the use of psychoactive drugs.

While current evidence does not support the routine use of medications to prevent delirium, efforts to minimize or avoid the use of potentially inappropriate medications in older adults are an effective prevention strategy. Medications that have the potential for inducing or worsening delirium include anticholinergic agents, antipsychotics, sedative–hypnotics including benzodiazepines and nonbenzodiazepine “Z drugs,” corticosteroids, H2-receptor antagonists, and opioid analgesics.

TREATMENT

Nonpharmacologic Treatment

Early identification and treatment of underlying medical conditions and/or contributing medications remain the first line in the management of delirium along with multimodal nonpharmacologic approaches. These may include modifying the patient’s surroundings to maximize the safety and calmness of the environment and providing reassurance to decrease fear and agitation associated with delirium. Environmental considerations may enhance the patient’s ability to appropriately interpret their surroundings; for example, rooms with a window view will provide orientation, and patients with eyeglasses and hearing aids should have access to these devices when appropriate. Paper and pencil should be provided to improve communication when the patient cannot respond verbally because of intubation. Furthermore, avoid extremes of sensory input such as stimulus overload in a loud and busy hospital unit or placing two affected patients in the same hospital room. Placing patients near a nursing station to facilitate close observation and maintenance of patient comfort and adequate pain control are also important in managing delirium. However, determining the appropriate level of analgesia in a patient with delirium can be challenging, and the use of certain medications can worsen delirium. Physical activity should be initiated as soon as possible in order to minimize the adverse effects of immobility (e.g., pressure sores), avoid deconditioning, and enhance orientation. Normal sleep–wake cycles can be promoted by the use of environmental cues such as windows and clocks. Adequate nutrition is important for recovery from delirium and the underlying illness, as well as for providing useful orientation. Interruptions of sleep should be minimized when possible, and brief, judicious use of sedative agents may be appropriate in some cases.

Psychosocial support can be provided by family and staff. Family members and staff can provide reassurance and reorientation to reduce anxiety and agitation. Familiar items from home may be useful for some patients. The availability of staff or family members to provide constant observation can minimize the need for physical restraints. When possible, continuity in staff can help decrease patient fears and paranoia because of unfamiliar faces. Conversations within earshot of a patient with delirium can be misinterpreted and produce a worsening of the patient’s symptoms. Therefore, education of family and staff and modification of behavior can be helpful.

Pharmacologic Treatment

At present, there are no U.S. Food and Drug Administration (FDA)-approved medications for the treatment of delirium. Few controlled trials have been conducted to determine which agents are most beneficial in managing the symptoms of delirium; however, two general classes of medications—antipsychotics and benzodiazepines—are most frequently utilized in clinical practice, depending on the etiology of delirium. The use of certain psychoactive medications can worsen delirium and, therefore, should be reserved for the management of behaviors associated with delirium that pose a safety risk for the patient and others.

Antipsychotics. Despite a lack of evidence demonstrating the efficacy of antipsychotics in the treatment of delirium or positive impact on clinical outcomes, low-dose antipsychotics may be used off-label when agitation or psychotic symptoms are refractory to nonpharmacologic approaches and impact the safety of the patient and staff, as well as proper care of the patient. Given safety concerns associated with antipsychotics in older patients, antipsychotics should be given at the lowest effective dose and only when the benefit outweighs the risk. Furthermore, these agents should be tapered and discontinued upon symptom resolution.

Haloperidol is commonly used off-label to treat delirium despite weak evidence to support its efficacy or impact on the duration of delirium. This potent agent may be preferred over other antipsychotics because it has fewer anticholinergic and hypotensive side effects. However, the increased potency of haloperidol is associated with an increased frequency of extrapyramidal side effects, including dystonic reactions, akathisia, tardive dyskinesia, malignant catatonia, and cardiovascular effects. Therefore, the patient should be monitored closely not only for effectiveness but also for adverse events.

Haloperidol is commonly administered orally and intramuscularly in various clinical settings. The use of intravenous haloperidol (which is also not FDA approved for the treatment of delirium or any other indication) is most commonly seen in critical care settings and is associated with an increased risk of QTc prolongation. ECG monitoring is recommended when haloperidol is administered via the intravenous route. Clinical observations have noted a possible decrease in the serious side effects in comparison with the oral and intramuscular routes. Other older antipsychotic drugs, including thioridazine, droperidol, and chlorpromazine, have been found to be equally effective; however, because of a higher incidence of anticholinergic side effects (which may worsen delirium), haloperidol is preferred.

Several recent studies have evaluated the use of atypical antipsychotics in the management of delirium and have shown some promise in terms of both efficacy and safety, particularly in postoperative patients. Similar to low-dose haloperidol, evidence supports the use of low-dose risperidone, olanzapine, and quetiapine in controlling the aggression associated with delirium. Although atypical antipsychotics are relatively less likely to cause extrapyramidal side effects, they are not without adverse effects. The FDA has issued a safety class warning regarding the use of all antipsychotics in patients with dementia as they are associated with a higher incidence of mortality.

Benzodiazepines. Benzodiazepines have been used historically in the management of delirium primarily to sedate agitated patients with hyperactive delirium. Current evidence suggests that benzodiazepines can increase both the risk and duration of delirium, particularly in older adults. Therefore, the use of benzodiazepines should be reserved for the management of agitation associated with sedative–hypnotic withdrawal (e.g., alcohol, benzodiazepines, barbiturates, meprobamate, and carisoprodol) and the treatment of delirium tremens. In addition, patients experiencing alcohol withdrawal with suspected Wernicke encephalopathy should be given high-dose thiamine via the intramuscular or intravenous route as oral absorption is very poor to prevent progression to Korsakoff syndrome.

Melatonin and Ramelteon. A 2014 study got a lot of attention as it showed significant prevention of delirium in acute care patients receiving 8 mg of ramelteon (3% vs. 32% in the placebo group) at night. Studies since then have looked at melatonin and ramelteon in delirium prevention and treatment. Four of the five trials studying the effect of melatonin analogs in preventing delirium reported a beneficial effect, but study heterogeneity limited any broad recommendations. Further studies are necessary to assess the benefits of melatonin and ramelteon in both the prevention and treatment of delirium.

Finally, limited evidence suggests that dexmedetomidine may be effective in the management of delirium in the ICU setting; however, results have been mixed. Larger, well-designed trials are warranted to more clearly define the place in therapy for these agents.

Electroconvulsive Therapy. Electroconvulsive therapy (ECT) is reserved as a treatment of last resort for patients with delirium and severe agitation who are not responsive to pharmacotherapy. ECT is usually given en bloc or daily for several days, sometimes with multiple treatments per day. ECT use in delirium should be monitored closely as ECT can cause delirium on its own.

Treatment of Specific Etiologies of Delirium

Anticholinergic Intoxication. As previously stated, psychoactive medications can contribute to an increased risk of developing delirium, especially those medications with high anticholinergic properties like benzodiazepines, sedative–hypnotics, antipsychotics, and opioid analgesics in addition to many over-the-counter medications. It is important to note that there can be a cumulative effect of the anticholinergic load from multiple medications. Therefore, the preferred course of action is rapid identification and removal of the offending agent.

Anticholinergic poisoning almost always results in delirium and is often accompanied by physical agitation and visual hallucinations. Physical signs resulting from antimuscarinic action include widely dilated, poorly reactive pupils; warm, dry skin; dry mouth; fever; tachycardia; elevated blood pressure; constipation; and urinary retention. The use of cholinesterase inhibitors, such as physostigmine, has been shown to reduce the severity of delirium but requires repeated dosing because of a short half-life. In a comparison study of physostigmine and benzodiazepines for anticholinergic poisoning, physostigmine was effective, whereas the benzodiazepine was ineffective and was associated with a higher complication rate. Cholinesterase inhibitors approved by the FDA for Alzheimer disease (donepezil, rivastigmine, and galantamine) have not been well studied in the treatment of delirium and are not without safety concerns. Thus, these agents are not recommended as a therapeutic modality in the management of delirium.

Substance Intoxication. Substance cessation is the first essential step in the management of substance-induced delirium. The primary focus in managing a patient with substance intoxication is supportive, ensuring that the patient does not have significant respiratory depression or cardiovascular abnormalities. For benzodiazepine ingestions, the benzodiazepine receptor antagonist flumazenil may be helpful; however, repeated administrations might be required. In mixed ingestions that include benzodiazepines and tricyclic antidepressants or carbamazepine, flumazenil may precipitate cardiac dysrhythmias or seizures. Opioid intoxication may cause respiratory depression in addition to delirium. Reversal with naloxone or naltrexone may be considered with adequate support and monitoring of the cardiovascular and respiratory status. Naloxone acts through competitive binding at opioid receptors and results in the reversal of receptor-mediated actions of opioids and, therefore, can precipitate acute withdrawal. Naloxone may be administered intramuscularly, intravenously, or intranasally. This agent has a short half-life and may require repeated administration depending on the half-life of the ingested opioid. There are many other agents that are used in the context of emergency care as antidotes to substances ingested.

Substance Withdrawal. Benzodiazepines are used widely for the treatment of alcohol withdrawal, with the goals of reducing the severity of withdrawal, preventing delirium, and reducing the incidence of seizures. A systematic review identified 57 relevant randomized controlled trials with a total of 4,275 participants. The quality of the included studies, many of which date from the 1970s and 1980s, was fair, and most were small. Only 9 studies clearly concealed the allocation of patients to treatment or control groups, and only 21 adequately described the randomization procedures, although all but 3 were double blinded.

The evidence-based guideline from the American Society of Addiction Medicine was recently updated and still recommends benzodiazepines as a first-line agent for the treatment of alcohol withdrawal. The guideline notes that although agents with a longer duration of action may provide fewer breakthrough symptoms, those with a shorter duration of action, such as lorazepam (Ativan), may be preferred when there is concern about prolonged sedation (e.g., in patients with significant comorbidities or liver disease).

End-of-Life Delirium. Not all causes of delirium are reversible, and realistic treatment expectations should be set after a discussion with the patient and caregivers. When delirium occurs in the context of a terminal illness, issues about advanced directives and the existence of a health care proxy become more significant. This scenario emphasizes the importance of early development of advance directives for health care decision making while a person has the capacity to communicate the wishes regarding the extent of aggressive diagnostic tests at life’s end. The focus may change from an aggressive search for the etiology of the delirium to one of palliation, comfort, and assistance with dying.

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (5th ed.). 2013.

By the 2019 American Geriatrics Society Beers Criteria® Update Expert Panel. American Geriatrics Society 2019 updated AGS Beers Criteria® for potentially inappropriate medication use in older adults. *J Am Geriatr Soc*. 2019;67(4):674–694.

de la Cruz M, Fan J, Yennu S, et al. The frequency of missed delirium in patients referred to palliative care in comprehensive cancer center. *Support Care Cancer*. 2015;23(8):2427–2433.

Devlin JW, Skrobik Y, Gélinas C, et al. Clinical practice guidelines for the prevention and management of pain, agitation/sedation, delirium, immobility, and sleep disruption in adult patients in the ICU. *Crit Care Med*. 2018;46(9):e825–e873. doi:10.1097/CCM.0000000000003299

Egberts A, Moreno-Gonzalez R, Alan H, Ziere G, Mattace-Raso FUS. Anticholinergic drug burden and delirium: a systematic review. *J Am Med Dir Assoc*. 2021;22(1):65–73.e4.

Girard TD, Exline MC, Carson SS, et al; MIND-USA Investigators. Haloperidol and ziprasidone for treatment of delirium in critical illness. *N Engl J Med*. 2018;379(26):2506–2516.

Greaves, D., Psaltis PJ, Davis DHJ, et al. Risk factors for delirium and cognitive decline following coronary artery bypass grafting surgery: a systematic review and meta-analysis. *J Am Heart Assoc*. 2020;9(22):e017275.

Han Y, Wu J, Qin Z, et al. Melatonin and its analogues for the prevention of postoperative delirium: a systematic review and meta-analysis. *J Pineal Res*. 2020;68:e12644.

Hatta K, Kishi Y, Wada K, et al; DELIRIA-J Group. Preventive effects of ramelteon on delirium: a randomized placebo-controlled trial. *JAMA Psychiatry*. 2014;71(4):397–403.

Hosker CM, Bennett MI. Delirium and agitation at the end of life. *BMJ*, 2016;353:i3085.

Inouye SK, Bogardus ST Jr, Baker DI, Leo-Summers L, Cooney LM Jr. The hospital elder life program: a model of care to prevent cognitive and functional decline in older hospitalized patients. Hospital Elder Life Program. *J Am Geriatr Soc*. 2000;48(12):1697–1706.

Inouye SK, Bogardus ST Jr, Charpentier PA, et al. A multicomponent intervention to prevent delirium in hospitalized older patients. *N Engl J Med*. 1999;340:669–676.

Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. *Lancet*. 2014;383(9920):911–922.

Lawlor PG, Gagnon B, Mancini IL, et al. Occurrence, causes, and outcome of delirium in patients with advanced cancer: a prospective study. *Arch Intern Med*. 2000;160(6):786–794.

Lipowski ZJ. Delirium: how its concept has developed. *Int Psychogeriatr*. 1991;3(2):115–120.

Morandi A, Brummel NE, Ely EW. Sedation, delirium and mechanical ventilation: the 'ABCDE' approach. *Curr Opin Crit Care*. 2011;17(1):43–49.

Neufeld KJ, Needham DM, Oh ES, et al. Antipsychotics for the Prevention and Treatment of Delirium. Comparative Effectiveness Review No. 219. (Prepared by the Johns Hopkins University Evidence-based Practice Center under Contract No. 290-2015-00006-I-2.) AHRQ Publication No. 19-EHC019-EF. Rockville, MD: Agency for Healthcare Research and Quality; September 2019. Posted final reports are located on the Effective Health Care Program search page. DOI: https://doi.org/10.23970/AHRQEPCCER219.

Nikooie R, Neufeld KJ, Oh ES, et al. Antipsychotics for treating delirium in hospitalized adults: a systematic review. *Ann Intern Med*. 2019;171(7):485–495.

Nishikimi M, Numaguchi A, Takahashi K, et al. Effect of administration of ramelteon, a melatonin receptor agonist, on the duration of stay in the ICU: a single-center randomized placebo-controlled trial. *Crit Care Med*. 2018;46(7):1099–1105.

Oh ES, Leoutsakos JM, Rosenberg PB, et al. Effects of ramelteon on the prevention of postoperative delirium in older patients undergoing orthopedic surgery: the RECOVER Randomized Controlled Trial. *Am J Geriatr Psychiatry*. 2021;29(1):90–100.

Oh ES, Fong TG, Hshieh TT, Inouye SK. Delirium in older persons: advances in diagnosis and treatment. *JAMA*. 2017;318(12):1161–1174.

Oh ES, Needham DM, Nikooie R, et al. Antipsychotics for preventing delirium in hospitalized adults: a systematic review. *Ann Intern Med*. 2019;171(7):474–484.

Pandharipande PP, Ely EW. Humanizing the treatment of hyperactive delirium in the last days of life. *JAMA*. 2017;318(11):1014–1015.

Rivière J, van der Mast RC, Vandenberghe J, Van Den Eede F. Efficacy and tolerability of atypical antipsychotics in the treatment of delirium: a systematic review of the literature. *Psychosomatics*. 2019;60(1):18–26.

Schubert M, Schürch R, Boettger S, et al. A hospital-wide evaluation of delirium prevalence and outcomes in acute care patients—a cohort study. *BMC Health Serv Res*. 2018;18(1):550.

Sieber F, Neufeld KJ, Gottschalk A, et al. Depth of sedation as an interventional target to reduce postoperative delirium: mortality and functional outcomes of the Strategy to Reduce the Incidence of Postoperative Delirium in Elderly Patients randomised clinical trial. *Br J Anaesth*. 2019;122(4):480–489.

Vaurio LE, Sands LP, Wang Y, Mullen EA, Leung JM. Postoperative delirium: the importance of pain and pain management. *Anesth Analg*. 2006;102(4):1267–1273.

Vlisides P, Avidan M. Recent advances in preventing and managing postoperative delirium. *F1000Res*. 2019;8:F1000 Faculty Rev-607.

World Health Organization (WHO). *The ICD-10 Classification of Mental and Behavioural Disorders*. World Health Organization; 1993.

Yang Y, Zhao X, Dong T, Yang Z, Zhang Q, Zhang Y. Risk factors for postoperative delirium following hip fracture repair in elderly patients: a systematic review and meta-analysis. *Aging Clin Exp Res*. 2017;29(2):115–126.

10.3 Major Neurocognitive Disorders (Dementia)

JULIE GRAZIANE, M.D., AND IQBAL AHMED, M.D., F.R.C.PSYCH.

This chapter will provide an overview of the clinical concept of major neurocognitive disorders that are caused by neurodegenerative disorders and review the common features of these major neurocognitive disorders. The most common subtypes of the major neurocognitive disorders will then be presented, and specific features will be discussed in detail. This chapter will discuss neurodegenerative neurocognitive disorders not caused by general medical conditions or substances: Alzheimer disease, frontotemporal dementia (FTD), dementia with Lewy bodies (DLB), dementia due to Parkinson disease (PDD), corticobasal degeneration (CBD), progressive supranuclear palsy (PSP), dementia due to Huntington disease, the prion diseases, and limbic predominant age-related TDP-43 encephalopathy (LATE). Please refer to the relevant chapters in this section for a discussion of vascular dementia, dementia in human immunodeficiency virus (HIV), and substance-induced dementia.

DEFINITION

Dementia refers to a disease process marked by progressive cognitive impairment in clear consciousness. The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) describes dementia, now subsumed under the term major neurocognitive disorder, as significant cognitive impairment in one or more of the domains of complex attention, executive function, learning and memory, language, perceptual-motor ability, and social cognition. These deficits represent a decline from a previous level of functioning, so dementia does not refer to low intellectual functioning or intellectual disability which are developmental and static conditions. This decline is often noted by the individual affected, a family member or other caretaker, or the clinician, and should be demonstrated on standardized neuropsychological testing, and if that is not possible, another measurable clinical assessment. These cognitive deficits interfere with independent functioning in daily activities and cannot occur exclusively in the context of a delirium or be better explained by another mental disorder. In DSM-5, major neurocognitive disorder is distinguished from mild neurocognitive disorder by the amount of cognitive decline. Mild neurocognitive disorder requires modest cognitive decline, as well as a lack of interference in activities required for independent living, although greater effort or compensatory mechanisms may be necessary to maintain this independence. The etiologies of major neurocognitive disorders are diverse and include degenerative diseases, general medical illnesses and infection, those that are substance-induced, and combinations of the above.

HISTORY

The word dementia derives from the Latin word *dementatus*, meaning out of one's mind, and, as such, was potentially applicable to any state of psychopathology. Celsus probably first used the term *dementia* in the first-century CE, although one of the first attempts to describe an etiology beyond old age was in the 4th-century CE by Oribasius, physician to Emperor Julian. Oribasius wrote of a disease of cerebral atrophy that caused loss of intellectual capacity and weakness of movement. It was not until the 19th century that the distinction between cognitive impairment due to dementia was separated from that caused by mental illness by Jean Etienne Dominique Esquirol, as described in his classic work *Mental Maladies: A Treatise on Insanity.* Esquirol identified three varieties of dementia: Acute, chronic, and senile. Senile dementia was defined as a "cerebral affection . . . characterized by a weakening of the sensibility, understanding, and will" and also by marked impairments in memory, reasoning, and attention. Esquirol also described hallucinations, delusions, aggressive behavior, and motor impairments in patients with dementia.

In 1845, Wilhelm Griesinger was the first to describe senile dementia as a disease of the cerebral arteries. However, it was one of Griesinger's students, Emil Kraepelin, who later narrowed the scope of dementia by differentiating senile dementia from psychoses with cerebral arteriosclerosis, which later came to be known as dementia praecox and, finally, as schizophrenia. Also, in the late 19th century, reports of plaques and atrophy in the brains of patients with memory deficits and mental confusion began to emerge. In 1907, Alois Alzheimer was the first to identify specific histopathologic changes associated with progressive degenerative dementia. He described two cases of dementia (in 1907 and 1911) characterized by symptoms of aphasia, apraxia, agnosia, and the histopathologic finding of neurofibrillary tangles and milar foci (plaques) that distinguished it from dementia associated with cerebral arteriosclerosis. Today, efforts continue to focus on elucidating the etiology of dementia and to further delineate the associated pathologic and physiologic abnormalities observed across the range of clinical presentations with specific emphasis on genetics, prevention, and treatment.

COMPARATIVE NOSOLOGY

The conceptualization of dementia has evolved through successive editions of the DSM. In DSM-III, dementia was categorized as a syndrome and as a disorder. Organic brain syndrome included dementia and distinguished dementia from cognitive impairment syndromes caused by general medical conditions or substances. Dementias were also listed separately as specific disorders and included primary degenerative dementia of the Alzheimer type, multi-infarct dementia, dementia due to alcoholism, and dementia not otherwise specified. In DSM-IV, specific criteria for dementia due to Alzheimer disease, vascular disease, specific medical conditions, and substance-induced conditions were presented. The DSM-IV-TR further divided dementia into two more categories, adding dementia due to multiple etiologies, where each individual etiology is listed separately, and dementia not otherwise specified, which encompassed dementia that did not fall into one of these categories or where the etiology had not been established.

The DSM-5 utilizes the terms major neurocognitive disorder and mild neurocognitive disorder, which broadly correspond with the commonly used terms dementia and mild cognitive impairment (MCI), respectively. The term neurocognitive disorder is broader than “dementia”; it extends beyond late-life to include neurocognitive impairments at all ages, including, for example, those caused by traumatic brain injury (TBI) or HIV infection. The addition of mild neurocognitive disorder is a change from DSM-IV. Another change is that the diagnosis of major neurocognitive disorder can, rarely, be based on substantial impairment in even a single cognitive domain, if sufficient to interfere with independence in everyday activities (e.g., profound isolated memory or language deficit caused by a strategically located stroke or head trauma). Major and mild neurocognitive disorders are the basic clinical syndromes, with a major neurocognitive disorder requiring significant cognitive decline in one or more neurocognitive domains. DSM-5 provides definitions and examples for the different cognitive domains, which include complex attention, executive function, learning and memory, language, perceptual-motor, and social cognition. With DSM-5, significant impairment in any of these six domains can qualify for a diagnosis of a major neurocognitive disorder, as opposed to earlier versions of the DSM when memory impairment was a required criterion for a diagnosis of dementia. Assessment of these domains and related functional activities are essential in the conceptualization of an individual's dementia process and related impairments in the daily life of the individual. DSM-5 provides further diagnostic criteria for the etiologic subtyping of the syndromes, for example, with criteria for neurocognitive disorders due to Alzheimer disease (where memory is typically impaired) and frontotemporal lobar degeneration (where speech, behavior, and executive function are typically impaired), with further criteria provided in accordance with current expert criteria for those diseases. Specifiers exist for the presence of behavioral symptoms, as well as for the level of severity in major neurocognitive disorder, with mild severity indicating difficulties in instrumental activities of daily living, moderate severity indicating difficulties with activities of daily living, and severe for those who are fully dependent.

The etiology-specific criteria within the DSM-5 are fairly compatible with the criteria set forth by many of the major expert workgroups and consensus guidelines, including the National Institute on Aging-Alzheimer Association Workgroup Criteria for Alzheimer disease, the International Work Group on Frontotemporal Dementia, and Pick Disease criteria for Frontotemporal Dementia, the Consortium on Dementia with Lewy Bodies International Workshop Criteria for DLB, the Movement Disorder Society Task Force criteria for Dementia Associated with Parkinson Disease, the Movement Disorder Society Criteria for PSP, and an International Consortium's Criteria for Corticobasal Degeneration. Since the publication of DSM-5, a consensus working group has also formed to describe the clinical entity and neuropathologic criteria for the diagnosis of LATE, which will be further described later in this chapter. As more clinical data and biomarkers for specific neurocognitive disorders are identified, these guidelines are updated. Since DSM-5 was published, there have been updated expert workgroups and consensus guidelines for the diagnosis of Alzheimer disease, DLB, and PSP. As a result, the DSM-5 criteria may be slightly different from that of a more recently updated expert workgroups.

Another classification system for dementias is offered in the *International Statistical Classification of Diseases and Related Health Problems*, 10th revision (ICD-10). This still uses a syndromic approach and offers criteria for general dementia and specifically for four types, Alzheimer disease, vascular dementia, dementia in other diseases classified elsewhere, and unspecified dementia. These criteria require a decline in activities of daily living.

EPIDEMIOLOGY

Worldwide, it is estimated that about 50 million individuals have dementia. With the aging of the population, this number is projected to increase to 152 million by the year 2050. A significantly higher number of women worldwide have dementia compared to men, with an estimated 27 million women versus 16.8 million men. While women do have a longer life span than men, this discrepancy is not based solely on this, as the age-standardized global prevalence of dementia in women is about 1.17 times that of men.

In the United States, an estimated 6.2 million individuals age 65 years or older were living with Alzheimer dementia in the year 2021. Prevalence rates of Alzheimer dementia increase with age, and range from an estimated 5.3% of individuals age 65 to 74, to 13.8% of individuals between ages 75 to 84, to 34.6% of individuals older than the age of 85. The prevalence of all dementias in individuals over the age of 65 in the United States is estimated to be about 11%. In the United States, the prevalence of dementia in individuals older than 65 is highest in African American individuals (13.8%), followed by Hispanic individuals (12.2%), non-Hispanic white individuals (10.3%), American Indian and Alaska Natives (9.1%), and lowest in Asian and Pacific Islanders (8.4%).

The prevalence differs depending on the population. The prevalence rates reported above apply to the general population. However, in a primary care population, the prevalence rates are higher and higher still in the general hospital setting. In the skilled nursing facility population, rates are closer to 50% to 60%.

Prevalence rates and anticipated changes in these are variable depending on the world region being studied. About 60% of individuals with dementia live in low or middle-income countries. A 2013 systematic review and meta-analysis estimates slightly higher prevalence rates for those over the age of 60 in Latin America (8.48%) and the Caribbean (8.12%) and lower rates for the four sub-Sahara African regions (2.07% to 4%). Latin America, North Africa, and the Middle East, as well as China, India, South Asia, and the Western Pacific, are expected to have a rapid increase in the prevalence of dementia. The lower rates of dementia, as well as a smaller expected rise in prevalence in the sub-Sahara regions, may be due to the differing projections of aging there, secondary to that region's higher rates of child mortality and HIV. A 2017 review of 14 population-based studies across Western Europe, the United States, Japan, and Nigeria showed declining or stable prevalence and incidence of dementia, particularly in Western countries. The Global Burden of Disease Study 2016 estimated age-standardized prevalence of dementia to be highest in Turkey, followed by Brazil.

Risk and Protective Factors

The leading risk factor for dementia is older age.

Beyond age, there are a number of modifiable risk factors of dementia, some of which can begin to be addressed in early life, and others in mid- and then others in later life. The 2020 report of the *Lancet* Commission estimates that these modifiable risk factors may contribute to about 40% of dementias worldwide. Addressing these risk factors as early as possible allows for the potential enhancement and then maintenance of cognitive reserve as well as the reduction of neuropathologic change.

The level of education an individual achieves is the first of these modifiable risk factors. The brain has its greatest neuroplasticity in adolescence, and so emphasis should be placed on advocacy for strong primary and secondary education. Less education, as well as illiteracy, have been associated with a higher risk of dementia. More years of formal education may subsequently confer some protection against dementia, with some studies showing that every additional year of education decreases dementia risk. Education after adolescence may be more helpful in maintaining cognitive functioning and in the promotion of cognitively stimulating activities, which may aid in the maintenance of cognitive reserve.

Vascular risk factors increase the risk not only for vascular dementia but also for Alzheimer disease and dementia in general. Hypertension, particularly long-term hypertension starting in midlife, increases the risk for dementia, with this risk increasing if this persists into late life. Midlife hypertension has been associated with reductions in brain volume and increased white matter intensities. It is recommended that individuals maintain a systolic blood pressure of 130 mm Hg or less in midlife. It should be noted that the U.S. and Puerto Rico Systolic Blood Pressure Intervention Trial (SPRINT) showed fewer dementia diagnoses, MCI diagnoses, and combined MCI or dementia diagnoses in individuals in the intensive treatment arm (systolic blood pressure of <120) versus those in standard treatment (<140). Lower blood pressure in late life, however, may actually be a sign of illness. High cholesterol is also a risk factor for dementia. Cardiovascular disease, particularly when combined with peripheral vascular disease, also confers greater risk. Cerebrovascular disease is also a risk, not just for vascular dementia but also for Alzheimer disease. Congestive heart failure and atrial fibrillation are also risk factors.

Diabetes has been associated with an increased risk of dementia. This risk increases with the severity and duration of diabetes. Evidence is mixed as to whether treatment for diabetes mitigates this risk or not. Obesity is also associated with increased dementia risk and is another modifiable risk factor that should be addressed as early as possible.

Early-life depression has been associated with an increased risk of Alzheimer disease, and the long-term depressive burden has also been shown to be associated with an increased risk of cognitive decline, dementia, and MCI. The relationship between late-life depression and dementia is more nuanced. Late-life depression is associated with increased dementia, but at least some of this risk is due to late-life depressive symptoms occurring as part of the prodrome or early signs of dementia. It remains unclear whether antidepressant use helps to mitigate the risk of depression.

Sleep disturbance has been associated with an increased risk of dementia. A 2018 meta-analysis showed associations between insomnia and a heightened risk of Alzheimer disease, and sleep-disordered breathing was associated with increased risk of all-cause dementia, Alzheimer disease, and vascular dementia.

There is an increased risk of dementia, and more specifically, Alzheimer disease in individuals with TBI. This risk further increases when the TBI is associated with loss of consciousness and with more severe injury.

Excessive alcohol consumption is associated with dementia, while mild to moderate alcohol intake may actually be protective. There is also an association between alcohol use disorders and early-onset dementia. Individuals should avoid drinking more than 21 units of alcohol per week. Individuals who smoke are also at higher risk of developing dementia than nonsmokers. While there is an inverse relationship between smoking and the development of Parkinson disease, given the numerous other health consequences of smoking, individuals should be encouraged to quit smoking whenever possible.

Anticholinergic medications, particularly those used for depression (mainly amitriptyline, dosulepin, and paroxetine), antiparkinsonism, and urologic reasons (mainly oxybutynin and tolterodine), have been associated with an increased incidence of dementia. This association is present even for these medications prescribed 15 to 20 years prior to the diagnosis of dementia. Gastrointestinal medications with a high anticholinergic burden have not been associated with a dementia diagnosis. Selective serotonin reuptake inhibitors (SSRIs) with a lesser anticholinergic burden have been associated with a higher dementia incidence, but this association has only been seen for prescriptions that were closer to the time of diagnosis, suggesting a possible reverse causation effect.

The relationship between physical activity and the risk of development of dementia is complex, particularly given physical activity’s role in the management of diabetes, hypertension, and obesity, but evidence suggests that physical inactivity increases the risk of dementia. Individuals, particularly those who are overweight, who engage in more leisure-time physical activity in midlife as well as increased leisure-time physical activity from mid- to late-life have been shown to have lower dementia risk. More long-term trials are needed, particularly in regard to the specific types of exercise and the duration needed to mitigate this risk.

Exposure to air pollution, particularly high nitrogen dioxide and fine ambient particulate matter from traffic exhaust and residential wood burning are associated with a heightened risk of dementia. This environmental risk may affect those from more socioeconomically disadvantaged groups.

Reduction in social isolation in later life has also been shown to provide some protection against the development of dementia. This may help in the maintenance of cognitive reserve.

Hearing impairment has been associated with increased temporal lobe volume loss, including in the hippocampus and entorhinal cortex, and has also been associated with a heightened risk of dementia. Notably, the use of hearing aids helps to mitigate this risk. As such, the use of sensory assistive devices when needed, as well as promoting protection from excessive noise exposure earlier in life, are important in the reduction of this risk.

There is increasing evidence that higher adherence to a Mediterranean diet is associated with less cognitive decline and a reduced risk of developing Alzheimer disease.

While there has been some research looking at protective factors in the development of dementia, clinical trials of putative protective factors including vitamin E, folate, vitamin B_6, and vitamin B_{12} to decrease the incidence of dementia have been negative. There has also been mixed epidemiologic evidence regarding statins decreasing the risk of developing dementia; however, a 2009 Cochrane review found no effect from the use of statins in late life on the prevention of Alzheimer disease or dementia.

Many of the risk factors described above are considered to be modifiable, providing opportunities to be addressed years and sometimes decades before the typical onset of dementia. The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability provided older individuals at risk for dementia multimodal interventions including management of vascular and metabolic risk factors, nutrition recommendations, physical exercise, cognitive training, and social activities; these interventions were associated with a reduction in the risk of cognitive decline, as well as improved executive functioning and processing speed. In addition, a retrospective cohort study of 196,000 individuals showed a healthier lifestyle, which was defined as no smoking, regular physical activity, a healthy diet, and moderate alcohol consumption was associated with a lower risk of development of dementia in individuals with high genetic risk compared to those with unfavorable lifestyle factors.

ETIOLOGY

There are many different etiologies for major neurocognitive disorders. Neurodegenerative disorders account for the vast majority of dementia cases. The most common is Alzheimer disease, which primarily involves neurodegeneration of the cholinergic system, among others. Alzheimer disease, alone or in combination with other etiologies, accounts for about 70% to 80% of dementias. On neuropathologic examination of individuals with dementia, the majority of diagnoses are of mixed pathology, with Alzheimer disease comorbid with vascular infarcts or cortical Lewy body disease. In one neuropathology study of individuals diagnosed with probable Alzheimer disease, on autopsy, only 42.4% were found to have pure Alzheimer disease pathology, and more than 45% of individuals were found to have mixed pathology. DLB is characterized by the same neuropathologic changes as Alzheimer disease, with neurofibrillary tangles and plaques, plus Lewy bodies in cortical and brain stem regions, and accounts for about 15% to 35% of dementias. Vascular disease, both small vessel and large vessel disease affecting cortical or subcortical regions accounts for about 5% to 20% of dementias. Cognitive deficits depend on which brain regions are affected or disconnected. The FTDs involve neurodegeneration and neuropathologic changes in frontal lobe and temporal lobe regions. In early-onset dementias (those occurring before the age of 65), frontotemporal lobe degeneration or vascular disease is the second most common etiology, following Alzheimer disease. Other neurodegenerative disorders affecting primarily the basal ganglia also cause dementia, specifically Parkinson disease, CBD, PSP, and Huntington disease. These account for a smaller proportion of dementia cases. Finally, prion diseases (e.g., Creutzfeldt–Jakob disease [CJD], Gerstmann–Sträussler-Scheinker disease [GSS]) include both spontaneous and transmissible forms and can cause rapidly progressive dementias.

Other neurologic disorders that can cause impairment in multiple cognitive domains include multiple sclerosis, brain tumors, normal-pressure hydrocephalus, subdural hematoma, and TBI. Other medical conditions can also cause dementia, such as endocrinopathies (e.g., hypothyroidism), vitamin B_{12} deficiency, uremia, autoimmune disease (systemic lupus erythematosus), and others. Infectious diseases such as

HIV and syphilis can cause dementia; HIV infection will be discussed in another chapter. See Table 10.3–1 for a list of etiologies of dementia.

DIAGNOSIS AND CLINICAL FEATURES

Clinical Features

The hallmark feature of dementia is cognitive impairment. Patients must demonstrate impairment in one or more cognitive domains (see Table 10.3–2 for DSM-5 criteria). The onset, duration, course, presentation, and associated symptoms vary depending on the type of dementia. A general overview is provided here, while a more detailed discussion of the clinical features of each type of dementia will be presented in each individual section.

Table 10.3–1.
Etiologies of Dementia

Neurodegenerative	Alzheimer Disease
	Dementia with Lewy bodies
	Frontotemporal dementia
	Parkinson disease
	Huntington disease
	Corticobasal degeneration
	Progressive supranuclear palsy
	Prion diseases (Creutzfeldt–Jakob disease, bovine spongiform encephalitis, Gerstmann–Straüssler syndrome, fatal insomnia, variably protease-sensitive prionopathy)
	Limbic predominant age-related TDP-43 encephalopathy (LATE)
Vascular	Infarction
	Binswanger disease
	Hemodynamic insufficiency
Neurologic disease	Multiple sclerosis
	Normal-pressure hydrocephalus
	Brain tumor (primary or metastatic)
Endocrine	Hypothyroidism
	Hypercalcemia
	Hypoglycemia
Nutritional	Vitamin B_{12} deficiency
	Thiamine deficiency
	Niacin deficiency
Infectious	Human immunodeficiency disease
	Neurosyphilis
	Cryptococcus
Metabolic	Hepatic insufficiency

	Renal insufficiency
	Wilson disease
	Metachromatic leukodystrophy
	Neuroacanthosis
Traumatic	Subdural hematoma
	Chronic traumatic encephalopathy/dementia pugilistica
Exposure	Alcohol
	Heavy metals
	Irradiation
	Anticholinergic medications
	Carbon monoxide

Table 10.3–2.
DSM-5 Criteria for Major Neurocognitive Disorder

A. Significant cognitive decline from a previous level of functioning in one or more cognitive domains, as evidenced by:
    1. Concern from the individual, the clinician or another knowledgeable informant that there has been significant cognitive decline
    2. Substantial impairment in cognitive performance, ideally seen on standardized neuropsychological testing, or if that is not available, another clinical assessment

B. Cognitive deficits interfere in independence in daily activities
C. The deficits do not occur exclusively during the course of a delirium
D. The deficits cannot be explained by another mental disorder

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:679–680.

Cognitive Impairment. Memory impairment refers to the inability to learn new information or to recall previously learned information. In general, short-term memory is impaired before long-term memory. Short-term memory impairment is often manifest as forgetting recent events and conversations, repetitive questions, the repetitive retelling of stories, forgetting the date, forgetting appointments, misplacing objects, losing valuables, and forgetting that food is cooking on the stove. Patients are unable to learn new information or new skills and may become overwhelmed, disoriented, or confused in new environments (i.e., traveling or moving). Later in the illness progression, long-term memory will become impaired, and patients will no longer recall their personal past history and will forget previously highly learned material. Procedural memory for previously learned skills may be retained longer.

Language impairment can manifest as difficulties in receptive or expressive language. Aphasia may be evident as word-finding difficulties resulting in reduced fluency of productive speech, word substitutions, or mispronouncing words. Patients may have difficulty remembering the names of distant family members or friends, or acquaintances they see infrequently. As language impairment progresses, patients may forget the names of close family and friends. They often have difficulty maintaining conversations or may use generalized, vague terms to try to maintain the conversation; eventually, they stop initiating conversation. Individuals may appear more withdrawn and disengaged as their ability to engage with others through conversation becomes more impaired. They may develop paraphasic errors and have poverty of content with empty speech. Eventually, speech may become incomprehensible and unintelligible.

Perceptual-motor impairments include difficulties in visual perception and visuoconstruction, as well as apraxia and agnosia. There may be difficulties in navigating in familiar environments, and individuals may have more trouble during times of dusk, when lighting changes may affect perception. Apraxia, or impaired ability to carry out motor activities despite intact motor function, may manifest in the earlier stages as difficulty using familiar tools or machines (i.e., microwave oven, washing machine, lawn mower, drill) or trouble performing previously acquired skills (i.e., knitting, woodworking). In later stages, there may be dressing, bathing, or feeding apraxias that contribute to impairment in the ability to perform activities of daily living. Agnosia, or failure to recognize or identify objects despite intact sensory function, may be reflected in the inability to recognize familiar objects, familiar faces, or in later stages, one’s own reflection in the mirror. It also manifests as a lack of insight into one’s own impairment and unrealistic assessment of one’s abilities.

Individuals with impairments in complex attention typically have difficulty when there are competing stimuli, such as multiple conversations or ambient sound from the television or radio. They are easily distracted and have trouble registering new information; thinking may also appear to take longer.

Impairment in executive functioning refers to disturbances in planning, organizing, sequencing, and abstracting. Impairment in this area is reflected in difficulty performing complex tasks or problem-solving. This may manifest as trouble preparing a meal, managing medications, or managing finances. Judgment is often impaired, and bad decisions are made because of deficits in executive functioning combined with a lack of recognition of difficulties.

Social cognition includes the ability to recognize emotions and the ability to consider another individual's mental state. Impairments in social cognition may initially be manifested as difficulty recognizing social cues or emotions in others, apathy, disinhibition, and changes in levels of extraversion or introversion. Family members of the individual may describe a change in personality. As these impairments progress, behavioral changes become more overt, with individuals making sexually or politically inappropriate comments and showing disinhibited and intrusive behaviors that lack awareness of another's personal space.

Table 10.3–3.
Basic and Instrumental Activities of Daily Living

Basic Activities of Daily Living	Instrumental Activities of Daily Living
Ambulation	Transportation
Dressing	Using the telephone
Grooming	Meal preparation
Bathing	Medication management
Feeding	Financial management
Toileting	Housekeeping
	Laundry
	Shopping

Functional Impairment. According to DSM-5, each of the cognitive deficits must result in a significant impairment in functioning that represents a decline from a previous level of functioning. If patients are still working, there must be difficulties with job performance. Difficulties in managing finances or medications may also be evident. Initially, there may be difficulty performing instrumental activities of daily living and eventually basic activities of daily living (see Table 10.3–3 for a description of these). There may also be difficulties in social functioning. This may manifest as disengagement from usual activities and interests or trouble maintaining social relationships and social roles.

Screening

The role of dementia screening is dependent on the medical setting and the history presented by the patient or concerned family members. The U.S. Preventative Services Task Force does not recommend for or against screening for dementia in the primary care setting. However, in 2011 the Center for Medicare and Medicaid Services began to require screening for cognitive impairment during the Annual Wellness Visit. Given the high prevalence of dementia, especially with advancing age, screening for dementia in high-risk groups may offer great clinical benefit. Dementia is associated with higher rates of delirium, adverse drug reactions, and falls so recognizing dementia earlier in its course might raise awareness and target prevention efforts to try to avoid these potential complications. In the psychiatric setting, screening for cognitive impairment in older adults is essential given dementia's relationship to depression. Also, some of the earliest signs of certain dementias may be behavioral or personality changes that may lead individuals to initially present for psychiatric evaluation and treatment. Early recognition would also allow for initiating potentially disease-modifying agents, when they become available, earlier in the course of illness that may potentially delay progression to more advanced stages.

Evaluation

A complete evaluation for dementia begins with a thorough history from both the subject and a reliable informant. Given the nature of memory loss in dementia, an accurate history cannot be expected to be obtained from the patient alone. Relevant aspects of the history include the onset of cognitive impairment (insidious or sudden), course (gradual or step-wise, progressive or episodic, or fluctuating), and duration of the impairment. An assessment of each cognitive domain is critical. For memory, enquire about short-term, long-term, and remote memory. For language, enquire about word-finding difficulties and remembering names of family members and friends. For complex attention, enquire about difficulties maintaining attention when there are competing stimuli; for example, an individual may have difficulty following conversation in large groups of people or when the television or radio is also on. For perceptual-motor abilities and praxis, enquire about difficulties in navigating previously familiar environments, use of familiar tools or machines, maintenance of previously acquired skills, and dressing or feeding apraxias. For agnosia, enquire about recognition of familiar objects and insight into their condition and limitations. Enquiring into executive function involves assessing one's ability to perform complex tasks or solve problems, such as the ability to manage finances or medications. For social cognition, changes in behavior or personality may be subtle or more overt, depending on the severity of the dementia.

Evaluation of Cognitive Domains

The degree of functional impairment should be ascertained. Assessment of executive functioning can be helpful in evaluating functional status (see Table 10.3–4). Asking about driving, management of finances, medication management, and activity planning can provide some information as to an individual's day-to-day functioning level. If there is suspicion of greater levels of impairment, enquiring about the ability to perform basic and instrumental activities of daily living is important (see Table 10.3–3). In some settings, evaluation with a structured measure, such as the Kohlman Evaluation of Living Skills (KELS) test, may be available to assist in gauging an individual's safety in the home.

In addition to the cognitive symptoms and related decline in functioning, patients should be evaluated for the behavioral and neuropsychiatric manifestations of dementia; these are described in more detail below. The Neuropsychiatric Inventory (NPI) is a useful tool that was developed to assess psychopathology in dementia patients and evaluates 12 neuropsychiatric disturbances: delusions, hallucinations, agitation, dysphoria, anxiety, apathy, irritability, euphoria, disinhibition, aberrant motor behavior, night-time behavior disturbance, and appetite and eating abnormalities.

Medical and neurologic history is relevant, especially any changes that may have a temporal relationship to the onset or worsening of cognitive dysfunction. Cerebrovascular events can coincide with the onset or worsening of vascular dementia or dementia of any etiology. Certain medical illnesses can contribute to the onset of dementia, and these may even be reversible causes of dementia that, when treated, can result in improvement in cognitive symptoms and functioning. Vascular risk factors, history of head trauma, and HIV risk factors may be relevant.

Psychiatric history is also relevant as certain psychiatric conditions may overlap with or predispose to dementia. Depression can cause cognitive impairment that may mimic dementia and may be associated with an increased risk for Alzheimer disease. Patients with Down syndrome will develop Alzheimer disease at an early age. Alcohol and substance abuse history should be ascertained. Family history of dementia, early-onset dementia, Huntington disease, or other neurologic diseases is important.

Several aspects of the social history are particularly relevant. Education level is important as higher education may be a protective factor, and level of education is relevant in interpreting findings on cognitive testing as normative values are dependent on age and education. Exposure to certain chemical agents may also be relevant as some chemical exposures may be associated with particular types of dementia.

Table 10.3–4.
Evaluation of Cognitive Domains

Complex attention	Does the individual have difficulty paying attention? Is the individual easily distracted in environments where there are competing stimuli? Does the individual have difficulty retaining newly presented information? Are routine tasks taking longer than usual?
Executive function	Does the individual have more difficulty with multitasking? Is the individual still able to manage their own finances or medications? Is more assistance required to plan out activities or instrumental activities of daily living? Is more assistance required to make decisions? Is the individual having more trouble in large social situations or finding them less enjoyable?
Learning and memory	Is the individual having difficulty recalling recent events and activities? Is the individual repeating himself or herself more often? Is the individual needing to rely more on lists or reminders? Is the individual misplacing or losing items? Does the individual need more reminders to attend to a particular task?
Language	Is there word-finding difficulty? Is the individual making more grammatical errors in conversation? Is the individual forgetting people’s names? Is the individual making more generalized, vague statements in response to questions? Is the individual using more generalized pronouns as substitutes for particular names of items?
Perceptual motor	Are they getting lost when traveling or in familiar environments? Are they having more difficulty using previously familiar tools and appliances? Are they having difficulty driving a car?
Social	Has the individual or family members noticed a personality

cognition	change? Does the individual appear to be less empathic or more disinhibited in their speech or behavior? Is the individual making sexually, politically, or religiously inappropriate comments? Does the individual show a lack of awareness of another’s interpersonal space?

Data from Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (Copyright © 2013). American Psychiatric Association.

A complete physical and neurologic examination is necessary. The physical examination is targeted toward medical conditions that may cause dementia. The neurologic examination is necessary to identify focal neurologic signs that may point to specific brain lesions or pathology.

The mental status examination is useful to assess not only the pattern of cognitive impairment, which may help with the diagnosis of the type of dementia, but also self-care abilities (appearance and grooming), behavior (degree of engagement in the interview, demeanor, inappropriate behavior), affect (apathy, depression, anxiety), speech (language function, fluency, comprehension), perception (hallucinations, paranoia, delusions), and insight and judgment.

Bedside cognitive testing is performed to characterize the pattern and extent of cognitive impairment. Two of the more commonly known bedside cognitive exams, the mini-mental status exam and the Montreal cognitive assessment (MoCA), are no longer in the public domain and have been monetized by their authors. Publically available tests include the mini-cog, the St. Louis mental status exam (SLUMS), the Kokmen short test of mental status (STMS), and the Rowland Universal dementia assessment scale (RUDAS). Addenbrooke cognitive examination (ACE), now in its third version, is another screening test that also examines attention, memory, language, orientation, and visual perceptual and visuospatial skills. There is a shorter version of this scale, known as the M-ACE, which has been developed and validated; this exam is scored out of 30 points and can be given in under 5 minutes. While bedside cognitive evaluation can be helpful when a clinician is suspicious of a mild or major neurocognitive disorder, consideration should be given for referral to a neuropsychologist for detailed cognitive testing on one of several standardized test batteries used commonly in the older adults with dementia. These more detailed tests may be beneficial in more clearly delineating areas of impaired function and in providing a baseline against which illness progression can be measured.

Severity Measures. There are several widely employed scales to assess the severity of dementia in terms of the level of cognitive and functional impairment. The Clinical dementia rating scale (CDR) measures several domains of function, including memory, orientation, judgment and problem-solving, community affairs, home and hobbies, and personal care. Another staging scale is the functional assessment staging (FAST) which has been developed to track the course of Alzheimer disease and is also useful for other dementias. Staging is based on the ability to perform certain skills or functions and is dependent on baseline skills, acquired deficits, and the social environment. Therefore, if a patient is not performing a certain skill because they never did or because their current environment does not demand it, it is not counted as a deficit. Patients with a CDR score of 0.5 or a FAST stage 2 to 3 have mild deficits in memory and sometimes other areas but have only mild or questionable functional difficulties. This generally corresponds with a diagnosis of MCI or mild dementia. Patients with a CDR score of 1 and a FAST stage 4 have some difficulty with instrumental activities of daily living, but they can still remain fairly independent with some assistance and prompting which corresponds to “mild” dementia. Patients with a CDR score of 2 and FAST stage 5 or 6 have impairment with more basic aspects of maintaining a home and themselves and will require some assistance with instrumental activities of daily living. Patients with a CDR score of 3 or FAST stage 6 or 7 have “severe” dementia and require considerable to total assistance with personal care and activities of daily living. In the terminal phase, patients develop contractures and are bed-bound, and require total care.

Neuropsychiatric Manifestations

Neither the DSM-5 nor ICD-10 requires neuropsychiatric or behavioral symptoms for the diagnosis of dementia. However, these symptoms commonly co-occur with dementia and are often the cause of the greatest caregiver distress.

Neuropsychiatric symptoms or behavioral disturbance are common in many types of dementia, although prevalence rates vary widely depending on the population and dementia subtype studied. It is estimated that almost 90% of patients with Alzheimer disease will have some type of neuropsychiatric symptoms during the course of the disease. The prevalence of behavioral disturbance increases with advancing dementia. Behavioral disturbances include mood changes, anxiety, personality changes, psychosis, sleep disturbances, agitation/aggression, and other behaviors, each of which is described below.

Mood Changes. Depressive symptoms are the most common neuropsychiatric symptom in dementia, followed closely by apathy and anxiety. Major depression is common early in the course of dementia, and prevalence rates are estimated at about 10%. Depressive symptoms are even more common, occurring in about half of patients with Alzheimer disease, and may be an early manifestation of the disease. These may be a consequence of aspects of neurodegeneration, and depressive symptoms correlate with greater temporal lobe and cingulate gyrus hypometabolism. In addition, depressive symptoms, emotional lability, and irritability may be related to psychological reactions to the awareness of losing one’s memory and anticipation of increasing functional impairment. Suicidal ideation and behavior may also occur through the course of dementia but are especially concerning in mildly impaired individuals who have insight into their deficits and still have the executive capacity to formulate and carry out a plan. Depression tends to decrease in prevalence with advanced dementia, so treatment for early depression may be withdrawn as the illness progresses. Elation and even a frank manic episode can also occur in the context of dementia but is rare.

Anxiety. Anxiety is fairly common throughout the course of dementia and is estimated to occur in about 60% of patients. Anxiety is often manifest as fear of being alone, and patients will search for their caregivers so as not to be alone. Some patients exhibit “catastrophic reactions” when confronted with their cognitive limitations or when facing changes in routine or environment.

Personality Changes. A range of personality changes may be evident. FTD often presents with disinhibition and impulsivity, which often predate obvious cognitive deficits. Examples include making inappropriate jokes, being overly familiar with strangers, disregarding social norms, inappropriate sexual remarks or behaviors, impulsive buying, or succumbing to salespeople or exploitative schemes. Alternatively, an apathy syndrome with amotivation and withdrawal may also be associated with various types of dementia, including Alzheimer disease and vascular dementia. In some cases, pre-existing personality traits may become stronger or exaggerated during the course of a dementing illness.

Psychosis. Psychotic symptoms in dementia generally occur in the middle stages of illness and often co-occur with behavioral disturbance. A common early psychotic symptom is paranoia and the belief that belongings have been stolen. This may develop as a result of poor short-term memory and misplacing objects. The inability to find a lost or misplaced object is interpreted in a paranoid way with the conviction that the object must have been stolen. This paranoia can become more pervasive as the illness progresses. In later stages, patients may have hallucinations in any modality, but most commonly, visual hallucinations. They may see deceased relatives. The hallucinations are often congruent with delusions of the same theme. People often believe they are still working or engaging in previous activities. Psychotic symptoms can sometimes lead to agitation, aggression, and behavioral disturbance as patients act out on their hallucinations, paranoia, or delusional thinking.

Sleep Disturbance. With advancing dementia, altered sleep–wake cycles can result in disrupted and fragmented sleep in about half of patients. Patients may have phase-shifted sleep, going to bed late, and sleeping late in the morning. They may also take frequent naps, which may be more prevalent in less stimulating environments where there is less to keep them engaged and awake. Sleep disturbance has been linked to activation of the inflammatory response and β-amyloid deposition.

Agitation and Aggression. Agitation and aggression become more prominent with increasing levels of dementia severity. There is a significant spectrum of behaviors that fall under the label of “agitation.” These behaviors are not always aggressive or directed toward a target. Agitation may also include nondirected behaviors such as continually pacing back and forth or purposeless hyperactivity such as continuously rummaging through drawers. Agitation and aggression can come in both physical and verbal forms, with some individuals hitting or striking out, while others may swear, make threats, or have persistent disruptive but nonthreatening vocalizations. The behaviors can be environmentally or task-dependent, such as with personal care, or may be more generalized. These behaviors often lead to acute hospitalization, placement in higher levels of care, cause significant caregiver burden, and even cause disability in caregivers due to injury.

Other Behavioral Disturbances. As cognitive impairment progresses, some individuals become at risk of wandering. This can be particularly dangerous in situations of inclement weather and for those individuals with altered sleep–wake cycles who still reside at home with a caregiver, as wandering may occur at nighttime. Disinhibited social behavior may also be seen, with individuals making inappropriate comments or gestures in public places.

LABORATORY EXAMINATION AND PATHOLOGY

The following are descriptions of the different studies available to aid in the diagnosis of dementia. The decision to order these to aid in the identification of a specific etiology of dementia is guided by the history and clinical examination, as well as the differential diagnosis.

Laboratory Studies

The purpose of obtaining routine laboratory studies in the evaluation of dementia is to rule out other general medical causes of dementia, including potentially reversible causes (see Table 10.3–5). These include complete chemistry panel, complete blood count, hepatic function, thyroid-stimulating hormone, vitamin B_{12}, and folate. Other laboratory studies that may be indicated depending on the history might include venereal disease research laboratory (VDRL) for syphilis, erythrocyte sedimentation rate, and HIV.

Structural Neuroimaging

Structural neuroimaging studies are obtained to rule out abnormalities in the brain that can cause or contribute to dementia, including normal pressure hydrocephalus, subdural hematoma, brain tumor or metastases, or cerebrovascular events. In general, a computed tomography (CT) scan is sufficient. However, magnetic resonance imaging (MRI) is preferred if looking for small or deep brain infarcts, as this imaging modality is better for visualizing vascular changes. Atrophy may be seen in patterns suggestive of specific dementia causes (see the review of individual causes of dementia, in the section on Major Neurocognitive Disorder Subtypes, below). An MRI with gadolinium contrast is useful if there is a high index of suspicion for a tumor.

Table 10.3–5.
Laboratory Evaluation of Dementia

Assessments	Rationale
Labs: Complete blood count, serum electrolytes, renal and hepatic function, glucose, albumin and protein, vitamin B_{12} and folate, rapid plasma reagin (syphilis), thyroid-stimulating hormone, urinalysis	Rule out correctable or contributory causes of dementia
Structural imaging: Computed tomography or magnetic resonance imaging	Rule out infarcts, mass lesions, tumors, and hydrocephalus
Molecular and functional imaging: PET-amyloid imaging, tau-PET imaging, Dopamine transporter (DaT) scan, SPECT, fluorodeoxyglucose-PET	Identifies certain pathologic biomarkers (Aβ or tau accumulation) or patterns of cerebral blood flow and/or metabolism that may be characteristic of specific neurodegenerative disorders
Cerebrospinal fluid	Detect altered levels of amyloid-beta and tau protein in Alzheimer disease, or elevated 14-3-3 protein and neuron-specific enolase in Creutzfeldt–Jakob disease

Molecular Neuroimaging

In an effort to identify biomarkers of the differing dementia processes, much work has gone into the development of molecular and functional imaging. During the progression of a particular neurodegenerative process, these markers may have differing utility. Positron emission tomography (PET) amyloid imaging can be used to identify beta-amyloid (Aβ) accumulation. The absence of amyloid deposits indicates the dementia is not likely to be due to Alzheimer disease. Tau-specific PET imaging allows for the in vivo detection of hyperphosphorylated tau aggregates. These aggregates can be seen in Alzheimer disease, PSP, CBD, frontotemporal lobar degeneration, and chronic traumatic encephalopathy. The main tracer currently available, flortaucipir, shows the greatest affinity for the hyperphosphorylated tau seen in Alzheimer disease. In addition, the degree of tau accumulation seen on PET imaging appears to correlate better with the degree of cognitive impairment compared to amyloid accumulation. Single-photon emission computed tomography (SPECT) or fluorodeoxyglucose (FDG)-PET can be used to demonstrate blood flow or metabolic rate in various brain regions to try to elucidate the etiology of dementia. For example, in Alzheimer disease, there is a characteristic pattern of decreased uptake in the posterior cingulate, temporal, and parietal cortex. In vascular dementia, there will be patchy areas of decreased uptake correlating to areas of hypoperfusion or infarct. In FTD, there will be a pattern of diminished uptake in primarily frontal as well as anterior temporal lobe regions, and in DLB, there may be a pattern of occipital hypoperfusion. Dopamine transporter imaging uses ligands specific for the dopamine transporter to examine the nigrostriatal dopamine system and is abnormal in Parkinson disease, DLB, multisystem atrophy, and PSP.

CSF Studies

CSF biomarkers are reflective of rates of production coupled with clearance of specific proteins at a specific point in time. These studies may be used to measure beta-amyloid protein, tau protein, and phospho-tau, all of which may be altered in Alzheimer disease. In Alzheimer disease, there is a characteristic pattern of decreased beta-amyloid with increased phosphorylated and total tau. Phosphorylated-tau is only elevated in Alzheimer disease, whereas increases in total tau are more reflective of more generalized neuronal damage at a specific point in time, such as after a TBI or a stroke. CSF studies in sporadic CJD may show elevations in 14-3-3, total tau, or neuron-specific enolase.

Electroencephalogram

Electroencephalograms (EEGs) may have a role in dementia evaluation if there is a concern about a seizure disorder or delirium. In CJD, EEG can show periodic sharp wave complexes. However, EEG is not necessary for the routine evaluation of dementia.

Neuropsychological Testing

Neuropsychological testing may be useful in elucidating the pattern of cognitive deficits in cases where the presentation is not typical of any one type of dementia. In Alzheimer disease, there is a pattern of early short-term memory and language impairment followed by the development of executive dysfunction and then visuospatial dysfunction. LATE presents with an amnestic syndrome similar to that seen in Alzheimer disease. In FTD, memory is relatively spared, while either language impairments or executive dysfunction predominate. In vascular dementia, there is a pattern of patchy deficits depending on where vascular lesions have occurred in the brain. Neuropsychological testing may also be helpful in identifying when cognitive impairment is due to depression or a factitious disorder.

Neuropathology

In individuals whose dementia does not arise from a previously characterized familial autosomal dominant mutation, a definitive diagnosis of dementia type can only be made with a neuropathologic investigation of the brain through either brain biopsy or autopsy. There is usually not a clinical indication for brain biopsy, so postmortem evaluation of the brain is the only way to provide a definitive diagnosis of dementia type.

DIFFERENTIAL DIAGNOSIS

In the evaluation of cognitive impairment, the psychiatrist must first determine that this is not occurring in the sole presence of a delirium and that the cognitive impairment is not the result of a mild neurocognitive disorder. Medical conditions that can cause cognitive impairment and medication-induced impairment should be ruled out. Other mental disorders, such as a schizophrenia spectrum or psychotic disorder, a depressive disorder, a substance-related disorder, and a factitious disorder, must be ruled out as well. Once these causes have been ruled out, the focus switches to identifying the diagnosis of the specific underlying neurodegenerative disorders. These are described briefly below.

Delirium

Delirium can be differentiated from dementia by the presence of an altered level of consciousness with the cognitive impairment, whereas dementia by definition is in clear consciousness. In delirium, the onset is generally sudden and acute and represents a significant change from baseline. There is typically a fluctuating level of alertness. Attention is the hallmark cognitive domain that is impaired. Delirium can occur in patients with dementia, and in fact, dementia is a risk factor for the development of delirium. Dementia accounts for the baseline cognitive decline that is punctuated by an acute worsening related to the delirium. Delirium is usually precipitated by one or more, and usually a combination of, medical factors or medications, or substances.

Mild Neurocognitive Disorder

Mild neurocognitive disorder, more commonly known as MCI, is impaired cognition in one or more domains, but to a lesser degree than in dementia. There may be mild impairments in the more complex aspects of daily functioning, or an individual may need to rely on more compensatory strategies; however, there is no significant impairment on activities of daily living. Because the pathophysiologic process of neurodegenerative dementias begins years before there is evidence of cognitive or functional impairment, MCI may represent a prodromal state of these disorders. However, not all individuals with MCI will go on to develop dementia. Research to recognize and quantify biomarkers to identify those both with and without MCI who will go on to develop dementia is underway.

Medical Conditions

General medical conditions that can cause or exacerbate cognitive impairment must be ruled out. Treatable factors that can cause or contribute to dementia include hypothyroidism, hypercalcemia, hypoglycemia, vitamin B_{12}, thiamine or niacin deficiency, renal or hepatic impairment, or autoimmune disease (i.e., lupus erythematosus, vasculitis, neurosarcoidosis, Hashimoto encephalopathy; see Table 10.3–1).

Medications

Many medications, alone or in combination, can cause cognitive impairment that may mimic dementia. Medications with anticholinergic properties are major culprits (see Table 10.3–1).

Schizophrenia Spectrum and Other Psychotic Disorders

Schizophrenia is associated with multiple cognitive deficits. These cognitive deficits may emerge prior to the onset of the psychotic syndrome in young adulthood and often progress during the early years of subsequent illness but are relatively constant thereafter. Some individuals with schizophrenia will develop a progressive deterioration of cognitive functions beginning in late life, which may represent a superimposed neurodegenerative process.

Depressive Disorders

Cognitive impairment frequently co-occurs with depressive episodes in late life. The impairment is substantial (though less so than in dementia) and is present in about 50% of depressed older adults, which is more frequent than in younger patients. Cognitive impairment in the context of depression was formerly termed “pseudodementia.” In the past, it was widely believed that cognitive impairment mimicked the symptoms of dementia but reversed upon resolution of the depressive episode. Numerous studies have now shown that when cognitive impairment co-occurs with depression in late life, it generally persists into depression remission, particularly in the memory and executive function domains. Thus, in many cases, both depression and cognitive impairment in the older adults likely have shared neurobiologic underpinnings. Moreover, depression seems to increase the risk for future dementia (both Alzheimer disease and vascular dementia), although it is not known whether depressed mood represents prodromal dementia or is etiologically related to future progressive decline. The varied findings in depression suggest that there are likely multiple phenotypes and pathways with different long-term trajectories and cognitive outcomes.

Alcohol Use Disorder

Substance use disorders, particulary alcohol use disorder, can cause neurocognitive disorders. Cognitive impairment is usually worse when patients are still actively using alcohol or other substances, and there may be some improvement with abstinence.

Factitious Disorder

Malingering or factitious disorder may present with dementia-type symptoms. The cognitive deficits are inconsistent over time and generally uncharacteristic of the pattern and course of dementia. There is some secondary gain from appearing cognitively impaired.

Neurodegenerative Disorders

Once the diagnosis of dementia, or major neurocognitive disorder, is identified and the above etiologies have been ruled out, the differential diagnosis shifts to identify the most probable specific etiology of the neurodegenerative process. Information and data obtained from the history, examination, neuropsychological testing, and imaging studies aid in the identification of the specific etiology and are summarized for each disorder in the individual sections below. It is also possible for an individual to have more than one neurodegenerative process occurring at the same time (e.g., both Alzheimer disease and Lewy body disease); this can blur the diagnostic picture, and unfortunately, a definitive diagnosis cannot be made until autopsy.

COURSE AND PROGNOSIS

The course and prognosis of dementia are variable depending on the etiology.

In general, onset is insidious, although abrupt onset due to a cerebrovascular or traumatic event is possible. Duration is generally years, although it may vary from 6 months in the case of CJD to 15 years with Alzheimer disease. In neurodegenerative dementias, there is a progressive cognitive decline and progressive functional impairment that may be of variable rate and pattern depending on the etiology and type. Dementia eventually leads to death unless another disease process supersedes it.

Identifying the correctable causes of cognitive impairment such as vitamin B_{12} deficiency, hypothyroidism, or normal-pressure hydrocephalus can result in some improvement in symptoms. Also, modifying vascular risk factors can ameliorate the course of dementia, even of nonvascular etiologies.

Gradual functional decline renders patients increasingly more dependent over the course of the illness. They slowly lose the ability to function independently. They lose the ability to perform instrumental activities of daily living such as shopping, meal preparation, managing finances, and managing medications. Then they have difficulty bathing and maintaining hygiene and dressing and feeding themselves. They require more and more prompting and then assistance with these tasks as the disease progresses, and they will eventually require total care.

Dementia increases the risk for complications from intercurrent medical illness, including delirium. Any medical illness can precipitate a rapid decline in cognition and functioning and markedly increase the level of care required. Dementia is also associated with an increased risk of falls and adverse reactions to medications.

MANAGEMENT

Management of dementia requires an individualized, multimodal approach not only targeting cognitive deficits but also addressing neuropsychiatric symptoms and overall level of functioning as well as caregiver issues, legal issues, safety, community support, and levels of care. Treatment modalities and strategies will vary with the stage and type of dementia and must evolve over time to address changing issues. Educating patients and their families about the course of illness and providing individualized treatment plans that ensure safety and adequate levels of care are essential.

General principles of dementia management will be provided here. The American Psychiatric Association Treatment Guidelines recommend that patients with neuropsychiatric symptoms of dementia be assessed for the type, frequency, severity, pattern, and timing of symptoms, as well as for potentially modifiable contributing factors, such as pain. For those patients with agitation or psychosis, it is recommended that response to treatment be assessed with a quantitative measure, such as the Brief Psychiatric Rating Scale or the NPI Questionnaire. The evidence and rationale for specific treatments for each specific type of dementia will be discussed in that section.

Nonpharmacologic Management

Psychological, behavioral, and psychosocial interventions which are specifically tailored to the individual with dementia are an important component of the management of the patient with dementia of any type. Educating patients and families about dementia and the course and what to expect at each stage is important for general treatment and anticipatory planning. Maintaining safety, ensuring adequate supervision and support, connecting patients and their families with community supports and resources, providing support for caregivers, and encouraging legal and financial planning all play an important role in the management of the patient with dementia.

Psychotherapeutic Approaches. Psychosocial interventions are an important component of the management of a patient with dementia of any type. While these interventions are important, there is no conclusive data on what type of patient or disease process to best use them for. Evidence supporting many of these therapies is limited as few have been subjected to randomized, double-blind, controlled trials, and many of the reported studies are often limited by small sample size, poorly characterized patient populations, incompletely described interventions, poorly defined outcomes, or short duration of follow-up.

Psychotherapy. Psychotherapy may be useful in earlier stages of dementia before short-term memory is too impaired. A 2014 Cochrane review of six studies looking at the effects of therapy on depression and anxiety in individuals with dementia and MCI demonstrated a reduction of depressive symptoms; however, there was no effect on self-rated or caregiver-rated anxiety. The psychotherapeutic approaches in these studies ranged from therapeutic counseling to cognitive behavioral therapy to brief interpersonal psychodynamic therapy. In the latter, emphasis was placed on improving interpersonal relationships and included the caregiver for a brief portion of each session.

Reminiscence therapy has also been studied in dementia. This therapy allows patients to recall and relive past life events, stimulating memory and mood within the context of their life history. This therapy has resulted in modest short-term gains in mood, behavior, and cognition in “confused” older adults.

Behavioral Management. Behavioral management has an important role in the management of dementia. Behavioral treatments that identify antecedents and consequences of problem behaviors and then effect changes in the environment to alter the behaviors have been shown to be beneficial in reducing disruptive behaviors in small trials and case studies and are widely employed in clinical practice. For example, scheduled toileting can reduce the frequency of urinary incontinence. Behavior management training programs for caregivers in the community or in residential or nursing homes have shown modest reductions in problem behaviors and, in some studies, improvement in mood, cognition, and physical role function. Limited follow-up data suggest that the benefits do not persist beyond the intervention.

Cognitive Training. Cognitive training places a focus on the guided practice of tasks, with the goal of improving orientation, memory, attention, and problem-solving. A 2019 Cochrane review examining 33 studies found cognitive training to be associated with small to moderate positive effects on global cognition and verbal semantic fluency when compared to a control intervention, but not to alternative treatments, such as occupational therapy, mindfulness, or reminiscence therapy. Future trials are needed to establish whether these effects are retained in the long term.

Stimulation-Oriented Therapies. Stimulation-oriented therapies aimed at enhancing pleasurable activities have some support from clinical trials. Examples include recreational therapy (i.e., crafts, games), art therapies, pet therapy, multisensory stimulation, simulated presence, aromatherapy, and exercise. Results with most of these modalities are mixed, with some studies demonstrating improvements in mood and behavior and reductions in agitation and problem behaviors and others showing no significant benefit. Despite the lack of convincing evidence for these modalities, their inclusion in a multimodal treatment plan for patients with dementia is humane and makes sense.

Psychoeducation. Providing education to patients and their families about dementia, its course, consequences, and treatment is important in helping them to plan for future disability. While anticipating the upcoming stages can be difficult, it affords patients the opportunity to make arrangements while they are still able to be involved in the decision-making process. There are many excellent resources, including books and community agencies, available to help guide families through the caregiving process.

Safety Awareness and Management

HOME SAFETY. Another important aspect of dementia management is maintaining safety in the community and at home. Patients with dementia are at greater risk for accidents. Forgetting to turn off the stove while cooking can lead to fires. Individuals may have more difficulty preparing meals, so assistance for this from family or supplemental meal programs may be helpful. Forgetting to lock doors can increase the risk of burglary or home invasion. As dementia becomes more severe, patients are at higher risk of wandering, so door alarms or identification bracelets may be necessary. Dementia patients are more vulnerable to predatory scams through door-to-door, phone, or mail solicitation. Assessing gait and risk of falls as well as a patient's insight into what to do during an emergency is also important. Addressing threats to patients' safety is necessary to protect patients from inadvertently harming themselves or others.

MEDICATION MANAGEMENT. As patients have more difficulty with memory and executive functioning, it may be more difficult for them to manage their own medications, particularly if they are required to take multiple medications or medications multiple times a day. The assistance of a family member is often necessary, and at times may be needed during the MCI stage.

DRIVING. Driving safety is an important issue. Dementia patients are at greater risk for motor vehicle accidents. Patients should be advised to limit their driving in the mild stages of dementia and to stop driving at the first signs of impaired driving ability or accidents or when they have reached the moderate stage of illness. Referral for driving evaluations is recommended to assess driving skills when in doubt or to confirm to patients their limitations. Many states have mandatory reporting laws. Driving restrictions can be very difficult for patients to accept due to the loss of independence. Families often need to get involved in enforcing these restrictions, removing keys, or disabling or removing vehicles.

Legal and Financial Planning. Patients and their families should be encouraged to pursue legal planning such as living wills, power of attorney, and guardianship for when they no longer have the capacity to make decisions for themselves. Careful financial planning for in-home or facility-based care is often needed and is best accomplished in advance.

Caregiver Support. Caregiver burden associated with caring for a loved one with dementia is great. About two-thirds of dementia caregivers are women, and almost a third are age 65 or older. Over half of caregivers are caring for a parent or in-law with dementia, and about 10% are caring for a spouse with dementia. In 2020, caregivers provided an estimated 15.3 billion hours of unpaid help or 26.3 hours of care per caregiver per week. As individuals with dementia decline, they become more dependent and require help with more activities of daily living. Caregiving is physically as well as emotionally demanding because the patient has less ability to communicate, eventually may not recognize the caregiver, and frequently exhibits inappropriate and difficult behaviors. Fifty-nine percent of caregivers report high to very high emotional stress from caregiving. Caregivers are at risk of physical injury by the patient during periods of providing direct assistance with activities of daily living or due to agitation or acting on delusional beliefs. Behavioral disturbance in patients is a major source of caregiver burden and also increases the risk for caregiver depression. Finally, caregivers also frequently become isolated from their social supports as patients can no longer engage in their usual social functions. The lack of social support is another risk factor for developing depression.

Caregiver stress and depression can also impair the caregiver’s ability to provide adequate and appropriate care and is a risk factor for abuse and neglect of the demented patient. Therefore, treatment of the demented patient must also include addressing the needs of, and when appropriate, treating the caregiver. Interventions shown to reduce caregiver burden include psychoeducation about dementia and comprehensive support, including individual and family counseling and support groups. Problem-solving therapy for caregivers of individuals with dementia and MCI has recently been shown to be effective in reducing depression and anxiety. Multidisciplinary and individualized approaches particularly aimed at managing problem symptoms and behaviors may be especially beneficial.

Abuse and Neglect. Patients with dementia are at greater risk for elder abuse and neglect. The strains of caregiving are a risk factor for caregivers to abuse those in their care. In addition, due to poor memory and inability to recall or solve problems, the individual with dementia may be less likely to report abuse, or the veracity of their reports may be questioned.

There are several instruments used to screen for elder mistreatment, but none has been widely accepted. In situations where abuse or neglect of the demented patient is suspected, further evaluation is indicated. Reporting is mandated in some states when elder mistreatment is suspected so that a complete investigation can be conducted and interventions performed when necessary.

End-of-Life Issues. Dementia is a terminal illness. Alzheimer disease is listed as the sixth leading cause of death in the United States and as the fifth leading cause of death worldwide.

While patients often die of intercurrent illnesses before the end stage of dementia, many patients succumb in the end stage of dementia, often due to pneumonia or other infections. Complications of end-stage dementia include loss of appetite, inability to feed oneself, and swallowing difficulties, making it difficult to ensure adequate nutrition. When dementia progresses to this stage, families will be faced with decisions regarding many life-prolonging interventions, including placement of feeding tubes and treatment of emergent infections. Ideally, these issues will have been addressed with the patient while in early illness stages or via a preexisting advance directive, so the patient’s wishes can be used to guide the decision-making process.

Palliative care, and eventually hospice, are important treatment aspects to consider, as it allows the focus of treatment to be on the relief of symptoms, and in doing so help to prevent both under treatment of symptoms such as pain, as well as overtreatment of other symptoms with possibly aggressive means. Symptoms palliative care can help focus on include pain, eating and drinking problems, and breathlessness. The incorporation of palliative care may also be helpful in the reduction of caregiver burden.

A main challenge in the integration of palliative care and hospice services in the treatment of dementia is identifying the appropriate timing, particularly as dementia is often not acknowledged as a terminal illness.

The National Hospice and Palliative Care Organization notes that a patient is likely to die within six months if their FAST is at stage 7 and there are one or more listed medical or nutritional complications; stage 7 is defined as dependence in all activities of daily living, occasional incontinence of bowel and bladder, and speech limited to five words. The advanced dementia prognostic tool (ADEPT) was developed with the goal of estimating 6-month survival in nursing home residents with advanced dementia and appears to be more useful as a continuous measure. This index weighs the following factors in determining a risk score: length of stay in the nursing home, age, male sex, shortness of breath, presence of at least one pressure ulcer at stage 2 or higher, activities of daily living, length of time spent in bed, poor oral intake, bowel incontinence, BMI of less than 18.5, recent weight loss, and congestive heart failure.

Pharmacologic Treatments

Pharmacologic treatments available for dementia are limited to symptomatic relief. To date, with the exception of interventions to reduce the risk of recurrent stroke in vascular dementia, there are no treatments available that can alter the disease course or produce substantial or long-lived improvements in symptomatology.

Treatment of Neuropsychiatric Manifestations

Treating the neuropsychiatric manifestations is an important part of the management of the patient with dementia. The neuropsychiatric manifestations, particularly agitation, behavioral disturbances, and psychosis, contribute substantially to caregiver stress and often prompt transitions to higher levels of care. These neuropsychiatric symptoms can occur with all etiologies of dementia. It is estimated that 97% of individuals with dementia will experience a noncognitive neuropsychiatric symptom as part of their dementia. Adequate control of these symptoms can markedly improve the quality of life of the patient and caregiver alike.

Psychosis and Behavioral Disturbance. Psychotic symptoms in dementia include hallucinations, delusions, and paranoia which are often associated with agitation or aggression. The term agitation as it relates to individuals with cognitive impairment or dementia covers a broad array of behaviors and has been defined by the International Psychogeriatric Association as behavior consistent with emotional distress, manifesting itself as excessive motor activity, verbal aggression, or physical aggression, which causes excess disability and is not solely attributable to another cause. Motor activity may include pacing, rocking, gesturing, pointing fingers, restlessness, or performing repetitious mannerisms. Verbal aggression can include yelling, screaming, shouting, or the use of profanity. Physical aggression can be directed toward a target or not and may include behaviors like grabbing, shoving, hitting, pushing, kicking, scratching, biting, or destroying property. Agitation can occur both with and without aggression. Agitation is present in many outpatients with dementia and maybe even more common in those residing in nursing facilities.

NONPHARMACOLOGIC MANAGEMENT. Nonpharmacologic treatments of neuropsychiatric symptoms are first-line, with the goals of nonpharmacologic management being prevention of the problematic behavior, symptom relief, and reduction of caregiver stress. The DICE approach has been identified as a nonpharmacologic approach to the management of these symptoms and involves the provider, the caregiver, and the individual with dementia if possible. The first step is to *describe* the problematic behavior in the context of its occurrence. The second step requires the provider to *investigate* potential underlying and modifiable causes of the behavior, such as problematic medications, pain, infection, change in routine, etc. The third step in this approach is to then *create* and implement a treatment plan for the behavior; this step often requires collaboration between the different parties involved and should take into account the patient, the caregiver, and the environment. Finally, the last step of this approach is to *evaluate* whether the treatment plan was instituted and the effectiveness of this.

The creation of a treatment plan often involves caregiver education, enhancing communication between the caregiver and the individual with dementia, and helping the caregiver to identify simplified yet meaningful tasks and routines for the individual with dementia. When a target behavior is associated with a particular activity, either modifying the activity or avoiding it if possible, can eliminate the behavior. Simplifying activities or modifying expectations to the cognitive and skill level of the patient can eliminate frustration. Modifying the environment to compensate for the patient’s deficits can eliminate triggers for problem behaviors. Ensuring adequate sleep, nutrition, and stimulating activities can also help reduce agitation.

When the behavior poses a threat to the patient's or caregiver's safety, more restrictive measures must be considered. In a nursing home or hospital setting, one-on-one supervision may be required. The use of physical restraints should be avoided if possible as they may increase the risk of falls and mortality.

When nonpharmacologic measures are ineffective, or safety is an issue, pharmacologic approaches should be considered; however, risks associated with these medications may outweigh the potential benefit. As such, it is essential that a discussion between the clinician and the patient's surrogate decision-maker occur when pharmacologic treatments are being considered.

PHARMACOLOGIC TREATMENTS

Selective Serotonin Reuptake Inhibitors. There have been several studies of SSRIs in the management of behavioral disturbance. As the tolerability of these medications may be greater than that of the antipsychotics, they may help in the management of agitation, especially in patients in whom the risk-benefit ratio for the use of antipsychotic medication is not favorable. A 2011 Cochrane review of nine trials examining the effects of antidepressants and trazodone on agitation found there to be limited evidence to support their use. The Citalopram for Agitation in Alzheimer disease study found that patients who were treated with citalopram compared to placebo were less likely to show delusions, anxiety, irritability, or lability, and caregiver distress was also decreased. However, treatment was also associated with a worsening of cognition and prolongation of the QTc interval.

Antipsychotics. The second-generation antipsychotics risperidone (Risperdal), olanzapine (Zyprexa), quetiapine (Seroquel), and aripiprazole (Abilify) have been studied in placebo and active comparator-controlled trials of psychosis and agitation in dementia in both nursing home and community outpatient settings. Studies of risperidone, olanzapine, and aripiprazole have shown modest but clinically significant improvements in measures of agitation, aggression, psychosis, and behavioral disturbances. These agents are indicated as first-line treatments in patients with dementia when agitated, aggressive, or psychotic behaviors are present and either the patient's function is impaired, or they are endangering themselves or others.

A meta-analysis of these studies demonstrates greater benefit for symptoms of agitation than psychosis. There is also a significant placebo response rate in these studies suggesting the importance of nonpharmacologic measures in the treatment of agitation and psychosis. The largest comparison study of three different second-generation antipsychotics—risperidone, quetiapine, and olanzapine—in patients with Alzheimer disease and psychosis or agitation treated for up to 36 weeks, the CATIE-AD trial, demonstrated no differences between any of the treatment groups and placebo in the primary outcome measure, time to discontinuation of treatment for any reason. However, time to discontinuation for lack of efficacy favored olanzapine and risperidone, while time to discontinuation due to adverse events or intolerability favored quetiapine and placebo.

First-generation antipsychotics have also been extensively studied for the treatment of psychosis and agitation in dementia. Two meta-analyses of these trials showed clinically significant improvement with several first-generation antipsychotics of 59% and 64% compared to 41% and 38% with placebo, respectively. There were no significant differences between agents. However, adverse effects were common, and drop-out rates were high.

Clozapine (Clozaril) has been shown to be useful in controlling psychotic symptoms in patients with Parkinson disease and DLB due to the absence of extrapyramidal side effects. There is no currently available data to support the efficacy of ziprasidone (Geodon), iloperidone (Fanapt), lurasidone (Latuda), asenapine (Saphris), and paliperidone (Invega) in treating behavioral disturbances in dementia. In 2023, the FDA approved brexpiprazole (Rexulti) specifically for the treatment of agitation associated with dementia due to Alzheimer disease; this is the first treatment to be approved for this indication.

Antipsychotic medications are associated with serious side effects and morbidity, particularly in the older adults. The first-generation antipsychotics, particularly the high potency agents such as haloperidol, are associated with a higher risk of tardive dyskinesia, neuroleptic malignant syndrome, akathisia, and parkinsonism. The low-potency agents, such as chlorpromazine, are associated with central and peripheral anticholinergic effects, sedation, and postural hypotension.

Second-generation antipsychotics are associated with an increased risk for metabolic syndrome, including weight gain, diabetes mellitus, and hyperlipidemia, but can also cause sedation, fatigue, and extrapyramidal side effects as well.

Both first- and second-generation antipsychotics have been associated with increased risk of cerebrovascular events and mortality, with first-generation medications having a slightly higher risk. In discussing the risks and benefits of these medications, it is important to recognize the above risks, as well as the risks of untreated or undertreated psychosis and agitation, which could lead to inadvertent harm to self or others, restraint use, poor quality of life, and placement in a higher level of care. Therefore, the use of antipsychotic medications for the treatment of psychosis and agitation must be carefully considered after nonpharmacologic measures have been implemented, weighing the evidence for the potential benefit for target symptoms, against the risks of the medications themselves and the risks of alternative approaches to maintaining safety.

After 4 weeks of treatment with antipsychotic medications, if there is no clinical improvement, then the medication should be tapered and withdrawn. If an antipsychotic trial is beneficial, there should be a discussion with the patient and surrogate decision-maker about the benefits of the medication, potential harm, and previous attempts at a taper, with consideration being given to the recommendation to taper and withdraw the medication within four months of initiation. The patient should then be assessed at least monthly for the reemergence of symptoms for at least 4 months after discontinuation. Long-acting antipsychotic injectable medications should not be utilized unless it is also indicated for a co-occurring chronic psychotic disorder.

Benzodiazepines. Benzodiazepines have been compared to placebo as well as to antipsychotics for the treatment of behavioral disturbances. In general, they are more effective than placebo but less effective than antipsychotics in controlling behavior symptoms. However, they may be associated with paradoxical effects, disinhibition, worsening cognition, and falls. Cumulative toxicity tends to be greatest with long half-life agents, which should be avoided. At present, the use of benzodiazepines is largely limited to short-term control of acute agitated behavior with short-acting agents such as lorazepam (Ativan) or oxazepam (Serax).

Anticonvulsants. Data on the use of carbamazepine (Tegretol) is favorable. It has been shown to have modest benefits for agitation in short-term trials; however, its risk of side effects and drug-drug interactions are limiting. Oxcarbazepine and valproate (Depakene) have not been shown to be efficacious in randomized controlled trials. A 2019 review of 15 case series and reports suggest a possible benefit of gabapentin and pregabalin for the treatment of behavioral symptoms, but no randomized controlled trials have been performed. The newer anticonvulsants have not been studied for behavioral disturbances in dementia.

Others. There is some evidence to suggest that trazodone may be helpful in decreasing irritability, anxiety and restlessness, and affective disturbance, especially in patients with FTD. Acetylcholinesterase inhibitors and memantine have been shown to have little benefit in the treatment of agitation and psychosis. Pimavanserin, which is a selective 5-HT$_{2A}$ receptor inverse agonist and antagonist, demonstrated potential efficacy in the treatment of psychotic symptoms in Alzheimer dementia in a randomized placebo-controlled study, with there being a larger treatment effect shown in patients with more severe psychotic symptoms at baseline. The benefit of treatment compared to placebo was significant at 6 weeks but not at 12 weeks. A Cochrane review showed potential benefits of methylphenidate in the improvement of apathy in dementia, as well as possible improvement of cognition and instrumental activities of daily living. There is some evidence to support the use of prazosin for the treatment of behavioral symptoms in patients with agitation or aggression, and there is currently an ongoing multicenter trial evaluating the use of prazosin in adults with Alzheimer dementia. Treatment with the combination dextromethorphan-quinidine resulted in a significant decrease in agitation and aggression in one randomized placebo-controlled trial in patients with probable Alzheimer dementia. A 2020 systematic review and meta-analysis examining the use of cannabinoids, including tetrahydrocannabinol, dronabinol, and nabilone, for the treatment of neuropsychiatric symptoms was supportive of the efficacy of these agents and found them to be well tolerated. Minimal evidence supports the use of beta-blockers or lithium (Eskalith), and both classes of agents can have significant adverse events that preclude recommending them for use in a frail older adults population. There is minimal evidence that buspirone (Buspar) may be efficacious in the treatment of agitation or anxiety in dementia.

Depression. Several antidepressants have been studied for the treatment of depressive symptoms in patients with Alzheimer disease. SSRIs have been the most studied, as there are clinical concerns regarding the tolerability and risk profile of tricyclics and monoamine oxidase inhibitors. A 2018 Cochrane review was not supportive of the efficacy of antidepressants for the treatment of depression in dementia. Similarly, a 2011 review of seven studies with 330 participants found that while two studies found antidepressants to have a benefit over placebo, overall, efficacy was only suggested and not confirmed. There is no data clearly favoring any specific antidepressant; however, because of its very long half-life in the older adults, fluoxetine (Prozac) is not generally recommended. There are no controlled trials of electroconvulsive therapy in the depression of dementia.

Nonpharmacologic interventions may be efficacious in the treatment of depression in dementia and include cognitive stimulation, exercise, massage and touch therapy, multidisciplinary care, occupational therapy, and reminiscence therapy.

Sleep Disturbance. Sleep disturbance is common in dementia, with individuals having decreased nocturnal sleep time, increased sleep fragmentation, and increased daytime sleepiness. A dual approach of enhancing sleep at nighttime through improved sleep hygiene and decreasing nocturnal interruptions, like light and sound, coupled with increasing daytime wakefulness through social and physical activity is important. One small study looking at the effects of trazodone showed improvements in nocturnal sleep time. Melatonin and ramelteon have not been found to be efficacious. There have been no randomized controlled trials of benzodiazepines or the nonbenzodiazepine sedative-hypnotics for the older adults with dementia, but because of their potential for cognitive and gait impairments, they are not recommended.

HEALTH DISPARITIES IN DEMENTIA

There are significant health and healthcare disparities present in dementia diagnosis, risk, and quality of care patients and their caregivers receive. The higher prevalence of dementia in African American and Hispanic populations magnifies these inequities. These disparities begin to evolve decades before the diagnosis of dementia is made, and many are related to the modifiable risk factors described above, such as the correlation between lower education and higher dementia risk. Minority and lower socioeconomic communities may be more likely to live in areas with higher exposure to pollution, further exacerbating their risk of dementia during their life course. Additional risks of hypertension, diabetes, and obesity, all carry with them their own contributing social and economic determinants, which then serve to further exacerbate the risk of dementia.

Experiences with discrimination when receiving medical care are common for people of color; half of Black Americans and more than a third of Native, Asian, and Hispanic Americans report that they have experienced healthcare discrimination, with many of these individuals citing discrimination as a barrier to receiving dementia care in the future. When caregivers are surveyed, Black, Hispanic, Asian, and Native Americans have all reported feeling like they are not listened to, that they are treated with less courtesy and respect, and that providers have acted like they were not smart in higher numbers when compared to White Americans. There is a strong desire among people of color to find providers who are culturally competent, but there is little confidence in their ability to do so.

The lack of trust in medical research among many individuals of color, as well as differing knowledge about the symptoms of dementia compared to normal aging, will only serve to worsen these inequities going forward. An awareness of these barriers, as well as a better understanding of the impact of many of these modifiable risks, provides opportunities to advocate for systemic change going forward.

COVID-19 AND DEMENTIA

The COVID-19 pandemic has disproportionately affected older individuals. Individuals with dementia are particularly vulnerable as most rely on other people for at least some if not significant assistance in IADLs and ADLs. An individual’s cognitive impairment may also make it difficult to adhere to mitigation measures. Many individuals with dementia live in nursing homes or assisted livings, and as such, were at increased risk of contracting COVID-19 from the communal nature of their residence. The recommendation for social/physical distancing and limitations placed on visitation, while important in mitigating disease spread, brought with them significant psychological ramifications, including increased loneliness as well as neuropsychiatric symptoms for many individuals with dementia, both in the community and long-term care settings. In 2020, there were approximately 16% more deaths than expected from Alzheimer disease. In death certificates listing COVID-19 as the primary cause of death, 4% listed Alzheimer disease as one of the multiple causes of death, and 11% listed vascular or unspecified dementia as an additional cause. This discrepancy was even more notable in individuals over the age of 85 who died from COVID-19, where Alzheimer was listed in 8% of death certificates, and vascular or unspecified dementia was listed in 20%. The COVID-19 pandemic also placed increasing stress among caregivers of individuals with dementia, as community supports temporarily closed and opportunities for respite became limited.

MAJOR NEUROCOGNITIVE DISORDER SUBTYPES

Alzheimer Disease

Definition. Dementia of the Alzheimer type is the most common form of dementia. The onset and steady progression of Alzheimer disease tend to follow a characteristic pattern, with diagnosis typically made on the basis of clinical history. However, autopsy studies have shown that up to 15% to 30% of clinically diagnosed Alzheimer dementias do not actually show pathologic changes indicative of Alzheimer disease. With the emerging knowledge in the field of biomarkers, diagnostic clarity is gradually shifting to a future where there are both clinical symptoms and objective pathologic changes needed for a diagnosis of Alzheimer dementia. While there is a better understanding of Alzheimer dementia than many other forms of dementia, there is still much to understand to be able to provide preventative strategies and treatments that can alter the disease course. There have been tremendous advances in the last decade in the understanding of the pathophysiology and genetics of Alzheimer disease. However, treatments are limited to symptomatic relief and modest delays of progression. Given the high prevalence in older age and advancing age of the population, Alzheimer disease is a growing problem affecting the lives of not only the sufferers but also their families.

History. Alois Alzheimer first described the illness in 1906 when he reported the case of a 51-year-old woman with a progressive cognitive disorder marked by memory loss, confusion, topographical disorientation, aphasia, and prominent neuropsychiatric manifestations of persecutory delusions, misidentifications, and behavioral disturbances. At her death 4 years later, the autopsy revealed profound atrophy, amyloid plaques, and neurofibrillary tangles.

Comparative Nosology. According to the DSM-5, a diagnosis of probable major neurocognitive disorder due to Alzheimer disease is made when the criteria for major neurocognitive disorder are met, and there is either evidence of a causative genetic mutation or the individual shows impairment in memory and learning, in addition to one other cognitive domain. There must also be a steadily progressive decline, and there is no evidence of another etiology of the cognitive impairment. Diagnosis is further classified into subtypes of with or without behavioral disturbance.

Specific criteria for Alzheimer disease were also defined by the National Institute of Neurological and Communicative Disorders and Stroke—Alzheimer disease and Related Disorders Association Work Group (NINCDS-ADRDA) in 1984. These diagnostic recommendations were revised in 2011 by the National Institute on Aging-Alzheimer’s Association (NIA-AA) workgroups. The 2011 guidelines identified three stages of Alzheimer disease: the preclinical stages of Alzheimer disease, MCI due to Alzheimer disease, and dementia due to Alzheimer disease. For a diagnosis of probable Alzheimer dementia, an individual must first meet the criteria for dementia described above; additionally, the symptoms should be of insidious onset, and there must be a report of worsening cognition. The diagnosis of probable Alzheimer disease has additional descriptors for the incorporation of documented decline or when it is diagnosed in a carrier of a causative Alzheimer disease genetic mutation. A diagnosis of possible Alzheimer disease dementia can be made when there is an atypical course, or an individual is showing signs of a dementia with mixed etiologies. Finally, the 2011 NIA-AA guidelines described a third category of diagnosis for when the criteria for probable Alzheimer disease dementia is met, and there is biomarker evidence supporting the diagnosis.

Since 2011, additional biomarkers have become recognized, and it became important to identify a more biologically based approach to the diagnosis of Alzheimer disease, particularly for observational and interventional studies. Utilizing a combination of cognitive staging and a biomarker profile, the NIA-AA updated its guidelines in 2018, with the main goal being to define and stage Alzheimer disease across the entire spectrum, ultimately helping to shift the diagnosis from a clinical set of symptoms to one defined by its pathologic process. While this framework is currently only intended for research purposes, it does provide an approach to think about Alzheimer disease as a continuum and allows for the incorporation of new biomarkers as they become available. It should be noted, however, that relying merely on biomarkers for diagnosis could lead to problems with a predictive accuracy of the diagnoses of MCI and dementia. Those individuals who are biomarker positive for Alzheimer disease but are cognitively unimpaired should only be considered at risk for progression to dementia, as the diagnosis of Alzheimer dementia should be based on both clinical and biologic factors. In addition, sole reliance on biomarkers for a specific dementia at one point in time has the potential to exclude additional etiologies of dementia that may also be present. Biomarkers will be discussed more in-depth below.

Epidemiology. 6.2 million Americans are estimated to have Alzheimer dementia in the year 2021. With the aging of the population, projections estimate that by the year 2060, 13.8 million people over the age of 65 will have Alzheimer disease, with 6.7 million of those individuals being over the age of 85.

The prevalence of the disease increases with age, with 5.3% of individuals age 65 to 74 estimated to have Alzheimer dementia, 13.8% of individuals between the ages of 75 to 84, and 34.6% of individuals over the age of 85 estimated to have Alzheimer dementia. More women than men in the United States have Alzheimer dementia, with 12% of women over age 65 compared to 6% of men. One of the main reasons for this disparity is the longer life span of women on average compared to men, but other possible contributing causes include differences in biology, as well as environmental and geographical or cultural factors. Older Black and Hispanic Americans are more likely than older White Americans to have Alzheimer or other dementias, with 18.6% of Black Americans and 14% of Hispanic Americans over the age of 65 having Alzheimer dementia, compared to 10% of White Americans. This difference is largely believed to be due to differences in medical comorbidities, such as cardiovascular disease and diabetes, as well as health and socioeconomic inequities. Due to projected increases in the age of the population of the United States, each state is expected to experience an increase of prevalence of at least 6.7%, with the West and Southeast portions of the country expected to have the largest increases.

Currently, these prevalence estimates are based on symptomatic clinical diagnosis, but as the field moves toward the incorporation of diagnostic biomarkers, the prevalence of Alzheimer dementia may actually be lower, as 10% to 30% of individuals who were thought to have Alzheimer dementia did not show Alzheimer related neuropathologic changes on autopsy. Conversely, the prevalence of those with Alzheimer disease may rise in the future, as biomarker-based diagnoses become more common, as 30% to 40% of cognitively unimpaired individuals show Alzheimer disease-related pathology on autopsy or have abnormal amyloid biomarkers.

Etiology. The etiology of Alzheimer disease is multifactorial. Mutations or copy number variations within several genes that result in the cleavage of amyloid precursor protein (APP) into Aβ invariably lead to early-onset Alzheimer disease within affected families. In contrast, late-onset Alzheimer disease is a complex disorder in which risk is influenced by multiple genes, environmental factors, and by their interactions.

GENETICS

Early-Onset Alzheimer Disease. Mutations at three genetic loci associated with early-onset Alzheimer disease have been identified. The first identified Alzheimer disease gene was the APP gene located on chromosome 21. Over 30 different mutations in APP that can cause Alzheimer disease have been identified in fewer than 100 families worldwide. More recently, families with duplications of APP, leaving the mutation carrier with a total of three copies of the gene on their two chromosomes, have been reported. This is similar to the increased risk of Alzheimer disease associated with trisomy 21. A second Alzheimer disease locus was found on chromosome 14, presenilin-1 (PS1). More than 180 mutations in the PS1 locus have now been reported in several hundred families throughout the world. The onset of symptoms for those with PS1 mutations is often before age 50, though later ages are possible. Like the APP mutations, PS1 mutations appear to act as autosomal dominant traits with nearly complete penetrance (i.e., inheriting a single copy of the mutated gene is sufficient to cause disease in all individuals). A third Alzheimer disease gene has been localized to chromosome 1 and termed presenilin-2 (PS2). These mutations appear to be rare; only a few families of different ethnicity have been identified with PS2 mutations. Although the PS2 mutations are autosomal dominant, they may not be fully penetrant. PS2 mutations generally lead to the onset of Alzheimer disease symptoms before age 65. There has been conflicting information on the proportion of early-onset cases attributable to mutations in APP, PS1, and PS2. Some reports have suggested that PS1 mutations may account for most early-onset familial Alzheimer disease cases. Others have reported that a more realistic estimate is 20% or fewer of such cases, with PS2 and APP accounting for 1% and 5%, respectively. For the remainder of individuals with early-onset familial Alzheimer disease, it is unclear whether other causative single-gene mutations will be identified or if these will prove to be multigenic in origin.

Late-Onset Alzheimer Disease. The first gene to have an established relationship to late-onset Alzheimer disease was apolipoprotein E, with an increased risk of Alzheimer disease found in individuals carrying the epsilon 4 allele (APOE4) in both familial and sporadic cases. The risk of Alzheimer disease increases, and the mean age of onset of Alzheimer disease is earlier, as the number of epsilon 4 alleles an individual carries increases from 0 to 2. A single copy of the APOE4 allele confers an approximate two- to four-fold increased risk while two APOE4 alleles increase the risk by approximately four- to eightfold. The magnitude of the risk associated with APOE4 may vary by age, and ethnicity, with the risk, attenuated in older individuals and in African American and Hispanic individuals (relative to Caucasians). In contrast, the risk conferred by APOE4 may be increased in individuals of Japanese ethnicity. It is generally estimated that APOE4 accounts for up to 50% of the genetic contribution to late-onset Alzheimer disease. Nevertheless, carrying one or even two APOE4 alleles is neither necessary nor sufficient for the development of Alzheimer disease. It has been estimated that for those who are disease-free at age 65, at most 50% of APOE4 homozygotes will develop Alzheimer disease within their lifetimes, thus limiting the use of APOE4 genotype as a predictive test. Alternatively, when APOE4 testing is added to the evaluation of individuals clinically diagnosed with cognitive symptoms consistent with dementia, it tends to lower the sensitivity but enhance the specificity of the diagnosis of Alzheimer disease.

Recent genome-wide association studies have identified 40 known susceptibility loci in Alzheimer disease in European populations. Implicated genes thus far include *APOE, CR1, BIN1, TREM2, CLU, SORL1, ADAM10, ABCA7, CD33, SPI1, PILRA, HLA-DRB1, CD2AP, PT2KB, PICALM, CASS4, SPI1, MS4A4A,* and *MS4A6A*. Most Alzheimer disease loci do not yet have functional variants or genes associated with them. In addition, genomic research is still needed in more diverse populations.

AMYLOID-β PROTEIN AND HYPERPHOSPHORYLATION OF MICROTUBULE-ASSOCIATED PROTEIN τ. The identification of causative mutations in early-onset familial Alzheimer disease has been a key piece of evidence indicating the causal role of Aβ protein in Alzheimer disease. As noted above, extra copies of APP via mutation or trisomy of chromosome 21 result in Alzheimer disease. It is noteworthy that, as assessed in animal and in vitro systems, the expression of the identified mutations in APP, PS1, and PS2 that lead to early-onset familial Alzheimer disease all result in overproduction and/or increased aggregation of Aβ. These genetic findings have also provided the key to understanding how APP is metabolized in the brain into Aβ. In brief, APP protein is cleaved by three enzymes, α-, β-, and γ-secretase. Cleavage by β-secretase then γ-secretase generates Aβ, while cleavage by α-secretase yields a nontoxic peptide.

More recently, the pathway from the overproduction of Aβ to the development of cognitive deficits has been substantially elaborated. It has been appreciated for a number of years that the strongest correlate of cognitive impairment in individuals with Alzheimer disease is loss of synapses across cortical regions. Substantial evidence now indicates that self-aggregation of Aβ into soluble low-n oligomers is a primary source of synaptotoxicity in Alzheimer disease. Soluble Aβ oligomers exist in equilibrium with fibrillar deposits of Aβ in plaques. Plaques themselves form substantial space-occupying lesions in the cortex of individuals with Alzheimer disease and serve as a site for inflammatory responses. Both these mechanical and inflammatory processes may further contribute to synapse loss and neuronal death.

In addition to the direct effects of Aβ on synapse loss, Aβ may lead to synapse loss and neuronal death via other downstream effectors. In a healthy brain, tau is a highly soluble and unfolded protein. Mainly found in axons, it is involved in the formation of tubulin into microtubules. Aβ contributes to the hyperphosphorylation of microtubule-associate protein τ (MAPT), which subsequently enhances its aggregation into neurofibrillary tangles. This consequently causes synapse loss and neuronal death independent of the effects of Aβ. Aggregation of α-synuclein also occurs in some individuals with Alzheimer disease and can contribute independently to cognitive impairments. At present, however, whether the aggregation of α-synuclein is enhanced by a mechanism related to Aβ is not known.

As the pathologic cascade in Alzheimer disease progresses, other brain changes may become evident. For example, numerous neurochemical abnormalities have been demonstrated on autopsy and with functional neuroimaging in Alzheimer disease in later disease stages. These include deficits in acetylcholine (ACh), norepinephrine, serotonin, dopamine, γ-aminobutyric acid (GABA), glutamate, corticotrophin-releasing hormone, and somatostatin. The most robustly described neurochemical deficits are in the cholinergic system. The activity of choline acetyltransferase, the enzyme responsible for the final step of ACh synthesis, is substantially reduced in patients with Alzheimer disease. This has been confirmed in the autopsy of patients with at least a moderate stage of the illness but not in patients with mild illness. This is particularly evident in the nucleus basalis of Meynert and several neocortical regions. This finding has led to the successful development of efficacious therapies directed at the enhancement of cholinergic function.

As excitatory synapses onto dendritic spines are particularly affected by Aβ-induced synaptic loss, it is not surprising that reductions in glutamate receptors, which are predominantly located at postsynaptic sites on dendritic spines, have been described in Alzheimer disease. The relationship between Aβ-induced synapse loss and glutamate receptors is complex. Aβ appears to destabilize dendritic spines, resulting in retraction and synapse loss via modulation of the function of both α-amino-3-hydroxy-5-methyl-4-isoxazoleproprionic acid (AMPA) and *N*-methyl-ᴅ-aspartate (NMDA) classes of postsynaptic glutamate receptors. It is therefore not surprising that memantine (Namenda), a low-affinity NMDA antagonist that functions to modulate the activation of NMDA receptors, has been shown to be of benefit in Alzheimer disease.

RISK AND PROTECTIVE FACTORS. In addition to the above-discussed risk and protective factors for dementia as a whole, there are some additional factors more specific to Alzheimer dementia which will be reviewed below.

Age remains the greatest risk factor for developing Alzheimer disease.

Epidemiologic studies have suggested that people with less education are at increased risk of developing Alzheimer disease. Functional neuroimaging findings of patients with Alzheimer disease indicate that people with higher educational attainment or higher premorbid intellectual functioning but similar levels of dementia severity have more severe findings on neuroimaging.

TBI can increase the rate of cognitive decline in the older adults and is associated with an increased risk of developing dementia, and in particular, Alzheimer disease. There is evidence that head trauma can lead to overexpression of APP, increase inflammatory mediators, and cause Aβ deposition in the brain. The effects of traumatic injury on Aβ may be mediated by the APOE genotype.

Estrogen’s role in the development of Alzheimer disease remains unclear. On a neuronal level, estrogen increases the density of a synaptic population lost in Alzheimer disease, excitatory synapses onto dendritic spines, in both hippocampus and cortex. It also may interact with APP processing to increase brain Aβ concentrations. However, prospective studies of estrogen treatment in patients with Alzheimer disease have yielded little evidence of cognitive enhancement or reduced cognitive decline. Similarly, results from the Women’s Health Initiative indicated no prevention of cognitive decline with estrogen replacement therapy in women aged 65 years or older and evidence of increased incidence of dementia in those women receiving combined estrogen/progesterone replacement therapy. The argument has been made that the timing of the hormone replacement (i.e., more effective during the perimenopausal or very early postmenopausal periods) may be the critical variable in this; however, current recommendations are that hormone therapy is not to be used in the prevention of dementia.

Several lines of evidence suggest a role of the immune system in Alzheimer disease. Neuropathologic studies demonstrate that the brains of patients with Alzheimer disease have increased concentrations of acute phase reactants, cytokines, and complement protein when compared to age-matched controls. A review evaluating the role of NSAIDs in the prevention of dementia showed an association in observational studies between NSAID use and reduced dementia risk; however, the one randomized controlled trial that was included showed no effect of NSAID use on the prevention of dementia. Results from one randomized controlled trial indicate that timing of NSAID use may be important, with NSAID use possibly having an adverse effect on individuals in later stages of Alzheimer disease pathogenesis but who have not yet clinically progressed to dementia. The use of NSAIDs for the prevention of Alzheimer disease is most likely precluded by offsetting risks of bleeding and cardiac toxicity.

Oxidative stress has been postulated to contribute to Alzheimer disease. For example, Aβ can damage and kill neurons by inducing oxidative stress, as can the microglial inflammatory reaction surrounding plaques. Increased markers of oxidative stress have been found in the brains of subjects with MCI, suggesting that these processes may be altered relatively early in the disease course. Thus, antioxidants have been evaluated for evidence that they are neuroprotective. However, a 2011 Cochrane review examining the effects of the antioxidant vitamin E in three studies found no convincing evidence of its beneficial effects on the treatment of Alzheimer disease or MCI. Studies have also not provided sufficient evidence to support treatment with the antioxidant ginkgo biloba.

Diagnosis and Clinical Features. As described above, Alzheimer dementia has a gradual and insidious onset, and there must be two cognitive domains affected, as well as functional impairment. The most common presentation is an amnestic one, in which the primary deficits are in the learning and memory domain. There are also nonamnestic presentations, where the prominent impairment could be in the language, visuospatial, or executive functioning domains. Individuals will have increasing difficulty functioning in their daily activities as the disease progresses.

Biomarkers. As described above, the National Institute on Aging and Alzheimer’s Association described a biomarker profile for the diagnosis of Alzheimer disease. Biomarkers are evidence of pathologic change happening in vivo. Biomarkers in Alzheimer disease include those found in CSF and imaging studies. In Alzheimer disease, biomarkers can be split into three different groupings.

The first group is indicative of aggregated Aβ or an associated pathologic state and includes cortical amyloid PET ligand binding or low CSF $A\beta_{42}$ or $A\beta_{42}/A\beta_{40}$ ratio. Several agents for PET-amyloid imaging have recently been approved for use. The absence of tracer retention during PET-amyloid imaging indicates that dementia is unlikely to be due to Alzheimer disease (Fig. 10.3–1). The second grouping is indicative of aggregated tau or neurofibrillary tangles or an associated pathologic state and includes CSF phosphorylated tau, or cortical tau PET ligand binding. The third set of biomarkers is indicative of neurodegeneration or neuronal injury and is not necessarily specific to Alzheimer disease and can be evident in other non-Alzheimer disease dementias; these biomarkers include CSF total tau, anatomic MRI, and FDG PET. Quantification of change in global (whole brain and ventricles) and regional (entorhinal cortex, hippocampus, corpus callosum) volumes in patients with Alzheimer disease frequently demonstrate group differences in comparison to cognitively intact older adults patients, with MCI patients intermediate (see Fig. 10.3–2 for an example MRI in a patient with Alzheimer disease). Currently, SPECT and FDG-PET are not established as stand-alone diagnostic tests. However, the addition of SPECT or FDG-PET can increase the diagnostic accuracy of an expert evaluation and can be particularly useful in differentiating individuals with Alzheimer disease from those with an FTD.

Course and Prognosis. The disease typically causes a gradual decline initially in short-term memory and language function. Initially, there is MCI, which may not be recognized by family members or be clinically relevant as there is no significant functional impairment. Gradually these impairments in short-term memory and language worsen until they start to interfere with the ability to perform instrumental activities of daily living.

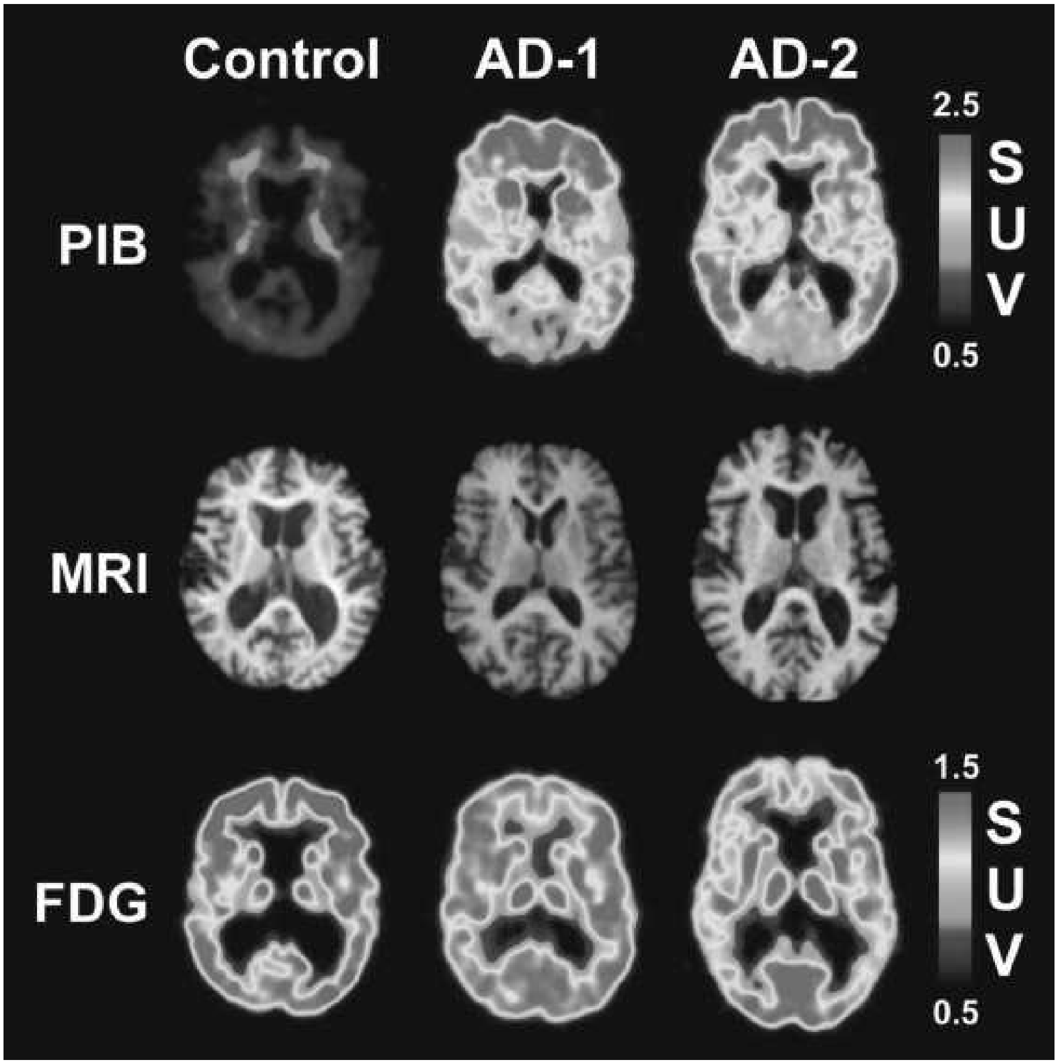

FIGURE 10.3–1. Pittsburgh Compound B (PIB) binding in normal aging and Alzheimer disease (AD). Comparison of PIB images (*top row*), magnetic resonance images (MRIs) (*center row*), and fluorodeoxyglucose (FDG) images (*bottom row*) from a cognitively normal subject (71-year-old patient; Mini-Mental State Examination [MMSE] score of 30) and two mild AD cases. AD-1 (*center column*) was a 68-year-old patient, had an MMSE score of 26, showed marked PIB accumulation compared to controls but a normal FDG scan. AD-2 (*right column*) was a 69-year-old patient, had an MMSE score of 21, also showed marked PIB accumulation, and an FDG scan typical for AD showing temporoparietal and frontal hypometabolism. Scale bars indicate standardized uptake values (SUVs) for PIB and FDG. Note that the dynamic range of PIB is twice that of FDG. (Courtesy University of Pittsburgh Amyloid Imaging Group.)

Typically, several years pass between the onset of early symptoms and recognition of deficits and presentation to medical attention. There is often a precipitating factor that alerts families to the cognitive impairment and prompts them to seek medical attention. The precipitating event can be an intercurrent medical illness that causes delirium, at which time families will look back and, in retrospect, recognize early symptoms of decline in the past. There may be a loss, such as the death of a spouse, that unmasks deficits in functions that had previously been performed for the patient by the deceased. Sometimes symptoms are identified during a change of residence, after which the affected individuals are required to learn their new environment and are unable to meet these demands, calling their families' attention to their impairment.

After the initial presentation, a gradual progressive decline in cognitive function ensues and extends eventually to all cognitive domains, including motor functioning (apraxia) and perception (agnosia). The rate of decline is not necessarily steady as there may be periods of plateau followed by periods of accelerated decline.

As the illness progresses, there will be greater functional impairment and greater dependency needs. The attendant loss of independence may be very difficult for the individual with Alzheimer dementia to accept; depressive symptoms may be evident if their level of insight is not fully impaired. Alternatively, if there is agnosia and a lack of insight into their deficits and impairment, the acceptance of additional help for their functional impairments can be quite difficult. Safety issues become of great concern, especially such things as leaving on the stove and burning food or pots or causing a fire, or driving and the increased risk of traffic accidents. Patients may forget to eat or take their medications, thus becoming malnourished and exacerbating their concurrent medical problems. The level of care and supervision required will increase over time as the illness progresses. Initially, daily checking-in may suffice. Later, help several hours per day, particularly around meals, medications, bathing, and dressing, is often required. As the disease progresses, delusions, hallucinations, and agitation become more prevalent. Eventually, total care and constant supervision for assurance of safety are needed.

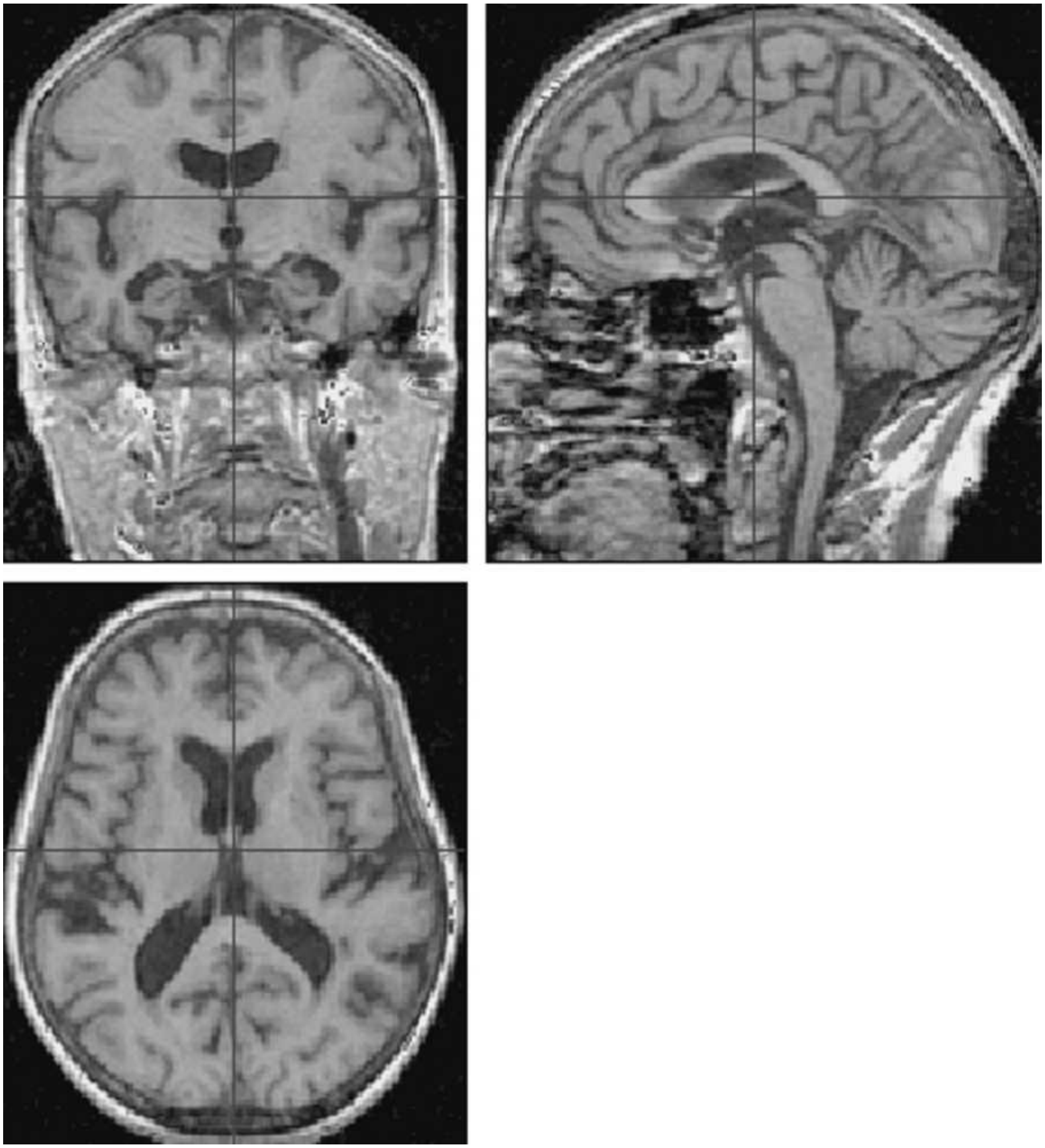

FIGURE 10.3–2. Magnetic resonance image of an individual with Alzheimer disease. Three plane views of the brain of a patient with probable Alzheimer disease at the time of their entry into the University of Pittsburgh Alzheimer Disease Research Center. Note there is significant atrophy in the hippocampus and related structures in the mesial temporal lobe (seen in the upper left image) and cortical atrophy in the frontal and temporal neocortex.

In the final stages of Alzheimer disease, parkinsonian symptoms may be present, mobility and gait are impaired, and the risk of fall is great. Eventually, patients will be bedridden. Swallowing impairments are common. Patients will often succumb to pneumonia or other infections. Palliative care is helpful in the end-stage to provide comfort measures and support for the family to ensure a peaceful and dignified death.

Neuropathology. Neuropathologic changes in the brain begin years before a dementia diagnosis is made. Macroscopically, the brains of early Alzheimer disease patients may appear grossly normal or with atrophy confined to the hippocampus and often most easily recognized by the increased volume of the inferior horn of the lateral ventricles. As the disease progresses, widened sulci throughout the cortex and increased ventricular size are seen. Atrophy in the posterior temporal, parietal, and frontal lobes is often the most prominent.

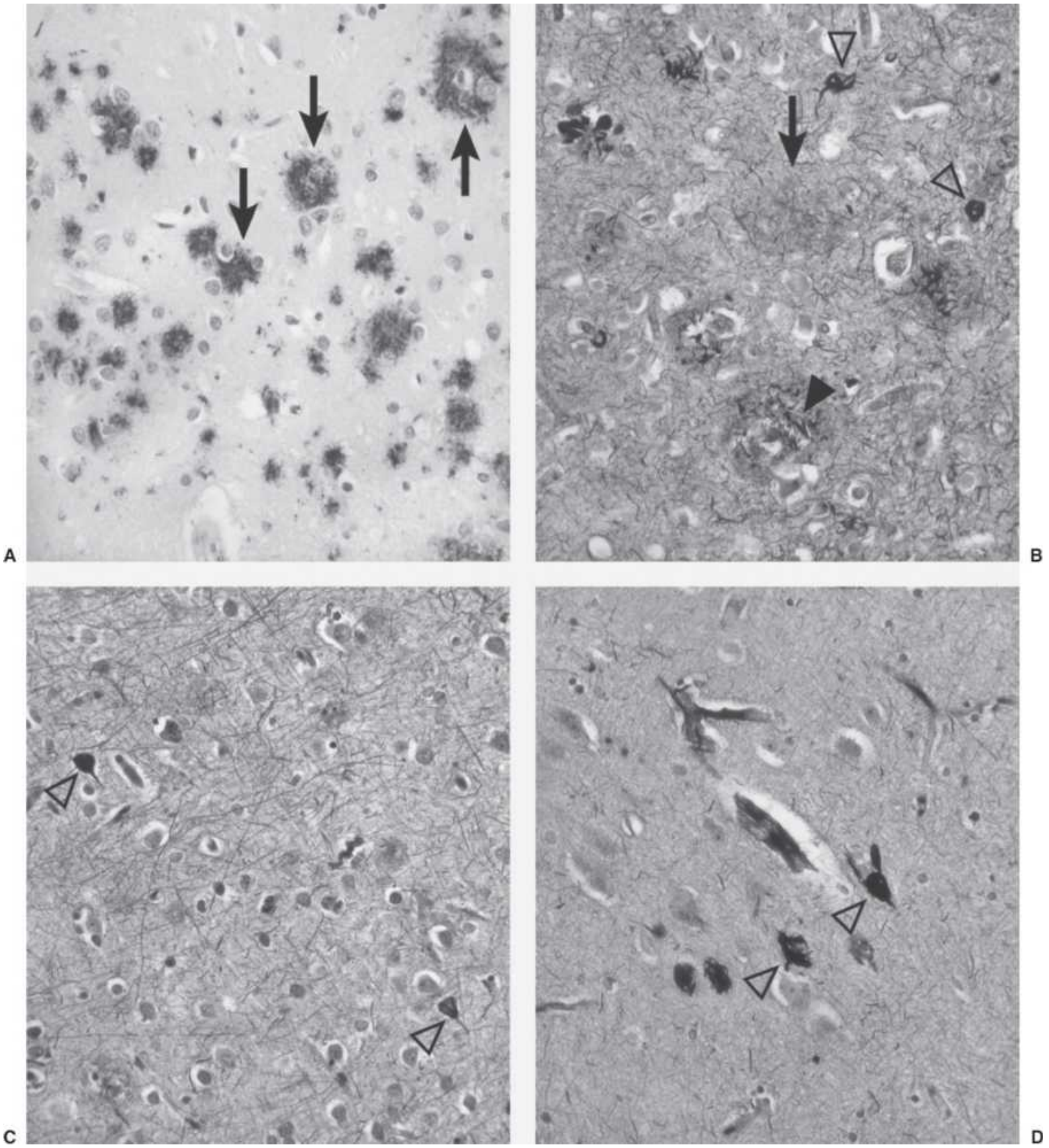

FIGURE 10.3–3. Photomicrographs of Alzheimer disease neuropathology. **(A)** Deposition of insoluble fibrillar Aβ into plaques begins in the neocortex, labeled here using an antibody against Aβ and appearing as reddish-brown deposits (*arrows*). **(B)** Bielchowsky stain of neocortex from an individual who died in advanced stages of Alzheimer disease (Braak stage VI). The Aβ plaques appear as dark brown in this preparation (*arrows*) and can be seen to be associated to varying degrees with dystrophic neuronal processes (*arrowheads*) in which insoluble microtubule-associated protein τ (MAPT) aggregates appear as black deposits. This neurofibrillary pathology also appears extensively throughout the neuropil, and several neurofibrillary tangles can be seen (*open arrowheads*). **(C)** Bielchowsky stain of neocortex from an individual who died in a less advanced disease stage (Braak stage

IV). Though some neurofibrillary tangles are still evident (*open arrowheads*), the degree of neurofibrillary pathology in the neuropil is substantially diminished. **(D)** Isolated neurofibrillary tangles (*open arrowheads*) in the entorhinal cortex that are seen in normal aging (Bielchowsky stain). Notice the lack of Aβ plaques and limited neuropil involvement. (All images obtained at 200× magnification, and provided courtesy of Dr. Ronald L. Hamilton, Department of Pathology, Division of Neuropathology, University of Pittsburgh School of Medicine.)

Microscopically, the characteristic lesions of Alzheimer disease are amyloid plaques and neurofibrillary tangles. Amyloid plaques are the defining lesion of Alzheimer disease, composed of insoluble deposits of fibrillar Aβ (Fig. 10.3–3). Surrounding this central core are dystrophic neurites, microglial cells, and reactive astrocytes. Amyloid plaques are first found throughout the neocortex, followed by the entorhinal cortex and hippocampus, then subcortical regions.

The second major histopathologic feature is the neurofibrillary tangle. Neurofibrillary tangles are intracellular inclusion bodies, which contain paired helical filaments composed of aggregates of hyperphosphorylated MAPT (see Fig. 10.3–3). Neurofibrillary tangles are first found in the entorhinal cortex and spread from there to the hippocampus and lateral temporal lobe, then more widely through the neocortex. The presence of neurofibrillary tangles is not pathognomonic for Alzheimer disease, as they are the principal lesion in MAPT-related FTDs. Additional pathologic features of Alzheimer disease include abnormal aggregation of α-synuclein, which may be detectable in as many as one-half of individuals with Alzheimer disease. These aggregates most commonly are found as thread-like deposits or as intraneuronal inclusions (Lewy bodies) in the amygdala and transentorhinal cortex, though Lewy bodies may also be present in the neocortex. Other neuropathologic abnormalities described in this disease include neuronal loss, most markedly in the entorhinal cortex and hippocampus. Later in the disease, loss of cholinergic neurons in the nucleus basalis, as well as loss of brain stem noradrenergic and serotonergic neurons, occurs. Granulovacuolar degeneration and the presence of Hirano bodies, most commonly affecting pyramidal neurons in the hippocampus, may be found.

The major pathologic protein aggregates found in Alzheimer disease and the overlap with other neurodegenerative dementias are summarized in Table 10.3–6.

Management. Nonpharmacologic management should be utilized throughout the course of illness. The below sections will discuss pharmacologic measures specific for the management of Alzheimer dementia.

CHOLINESTERASE INHIBITORS. The first class of medications approved by the U.S. Food and Drug Administration (FDA) for the treatment of Alzheimer disease was the cholinesterase inhibitors: Tacrine (Cognex), donepezil (Aricept), rivastigmine (Exelon), and galantamine (Reminyl). These agents are reversible inhibitors of the enzyme acetylcholinesterase, which degrades Ach in the synaptic cleft, thereby increasing the availability of intrasynaptic Ach.

These agents have all been shown in multiple randomized, double-blind, placebo-controlled trials to slow cognitive decline in mild to moderate Alzheimer disease. These medications have been associated with a modest decrease in long-term cognitive impairment over time and have also been associated with reduced mortality risk; galantamine was also associated with a reduced risk of developing severe dementia in a matched cohort study. Patients who have a good response to cholinesterase inhibitor treatment can expect to stay in their homes for 3 to 8 months longer. There are limited trials in severe Alzheimer dementia. Donepezil is the only cholinesterase inhibitor FDA-approved for the treatment of severe Alzheimer dementia. There are no data on whether or how to switch from one agent to another.

Common side effects are related to cholinergic excess and include nausea and vomiting, decreased appetite and weight, increased gastric acid secretion, muscle cramps, bradycardia with syncope, and sleep disturbances. They may induce or exacerbate urinary obstruction, worsen asthma or chronic obstructive pulmonary disease, and cause seizures. Tacrine can cause reversible hepatotoxicity; because of this common risk and the four times per day dosing, tacrine is now rarely used.

Cholinesterase inhibitors should be used with caution in patients with gastric ulcer disease, sick sinus syndrome or conduction defects, asthma or chronic obstructive pulmonary disease, or cerebrovascular disease. The only contraindication is hypersensitivity. The tolerability of these different agents is variable, so if patients do not tolerate one agent, an alternative should be tried. If gastrointestinal side effects are particularly problematic, individuals can be switched to transdermal patch preparations (see Table 10.3–7 for dosing guidelines).

Patients should be titrated gradually to the maximally tolerated dosage as there is a dose-dependent response. If there is a continued rapid decline despite the medication, it can be discontinued. However, if there is a stabilization of symptoms or slowing of decline, which is the desired response, the medication should be continued. The duration of treatment is unclear. Placebo-controlled studies demonstrate efficacy for up to 6 months to 2 years, depending on the agent. Discontinuation of cholinesterase inhibitors has been associated with marked deterioration and regression of cognition to levels of the placebo control group. There is some suggestion that after cholinesterase treatment has been interrupted for 6 weeks or more before being restarted, patients do not return to the cognitive level they were at before the medication was discontinued. Case reports have described intense hallucinations, difficulty sleeping and concentrating, agitation, changes in mood, and delirium with abrupt discontinuation of donepezil.

MEMANTINE. Memantine is a noncompetitive NMDA receptor antagonist that may modulate excitatory signaling and have neuroprotective actions.

Memantine has been studied in mild, moderate, and severe stages of Alzheimer disease. It is FDA-approved for use only in moderate to severe Alzheimer disease and has demonstrated efficacy for both cognitive and functional improvement over placebo. It has been studied as monotherapy and in addition to stable doses of donepezil and is effective and well-tolerated when added to the cholinesterase inhibitor. There are no data for use longer than 6 months duration. It is generally well tolerated with drop-out rates comparable to placebo. Infrequent adverse events include confusion, dizziness, headache, sedation, agitation, falls, and constipation (see Table 10.3–7 for dosing guidelines).

Table 10.3–6.
Pathologic Proteins in the Neurodegenerative Dementias

	Pathologic Protein			
Dementia Type	**Amyloid-Beta**	**Tau**	**Alpha-Synuclein**	**TDP-43**
FTD due to *C9ORF72*, *GRN*, *VCP*, *TARDBP*, and *FUS* mutations	No	No	No	Yes
FTD—Pick disease and due to *MAPT* mutations	No	Yes	No	No
Corticobasal Degeneration	No	Yes	No	No
Progressive Supranuclear palsy	No	Yes	No	No
Alzheimer disease	Yes	Yes	Present in up to 60% of cases	Present in 19–57% of cases
Dementia with Lewy bodies	Yes	Yes	Yes	No
Parkinson disease dementia	Yes	Yes	Yes	No
Limbic predominant age-related TDP-43 encephalopathy (LATE)	No	No	No	Yes

Table 10.3–7.
Pharmacologic Treatments for Cognitive Impairment

Class	Agent	Starting Dose	Titration	Effective Dose	Maximum Dose
Cholinesterase inhibitor	Donepezil	5 mg	5 mg after 4–6 weeks	5–10 mg	23 mg
	Galantamine	4 mg b.i.d.	4 mg b.i.d. every 4 weeks	16–24 mg daily divided b.i.d.	24 mg daily divided b.i.d.
	Galantamine extended release	8 mg daily	8 mg every 4 weeks	16–24 mg daily	24 mg daily
	Rivastigmine	1.5 mg b.i.d.	1.5 mg b.i.d. every 2 weeks	3–12 mg daily divided b.i.d.	12 mg daily divided b.i.d.
	Rivastigmine transdermal patch	4.6 mg/24 hours daily	Increase to 9.5 mg/24 hours daily after 4 weeks	4.6–9.5 mg/24 hours daily	9.5 mg/24 hours daily
NMDA-receptor antagonist	Memantine	5 mg	5 mg weekly divided b.i.d.	20 mg divided b.i.d.	20 mg divided b.i.d.
	Memantine extended release (Namenda-XR)	7 mg	7 mg weekly	28 mg	28 mg
Combination treatment	Memantine ER and Donepezil IR (Namzaric)	7/10 mg	7 mg weekly (convenience pack formulation)		28/10 mg

NMDA, *N*-methyl-D-aspartate.

MONOCLONAL ANTIBODY TREATMENTS. Aducanumab (Aduhelm) is an amyloid beta-directed antibody, and was approved by the FDA in 2021 for the treatment of Alzheimer’s disease through their Accelerated Approval pathway. This pathway allows for the approval of treatments for serious or life threatening illness where there is the expectation of clinical benefit despite some uncertainty and when the therapy may provide an advantage over existing treatments. This approval was based on a surrogate endpoint of the reduction of beta amyloid plaque, as there was conflicting data in two phase 3 clinical trials surrounding the reduction of clinical decline. The treatment will now need a post-approval trial demonstrating clinical benefit. This is the first disease modifying therapy approved by the FDA for Alzheimer’s disease, and is the first medication approved by the FDA for Alzheimer’s disease since 2003.

Intravenous lecanemab (Leqembi) was also granted accelerated approval for the treatment of Alzheimer disease in early 2023, with a recent phase 3 trial demonstrating a reduction in brain amyloid levels and less cognitive decline. It is recommended that it should be used in patients with mild cognitive impairment or mild dementia, the population in which treatment was studied in clinical trials. Due to concerns about developing amyloid-related imaging abnormalities or ARIA, a type of brain swelling or bleeding brain, imaging monitoring is recommended for both these agents.

OTHER MEDICATIONS. Other medications, including vitamin E, ginkgo biloba, NSAIDs, hormone replacement therapy, selegiline, ergoloids, and antiamyloid therapies, have been studied for the prevention and/or treatment of cognitive decline in Alzheimer disease, but placebo-controlled trials have not shown sufficient efficacy.

Frontotemporal Dementia

Dementia associated with atrophy of the frontal and temporal lobes encompasses both sporadic and familial diseases. The hallmark of FTD is a progressive change in personality and behavior, with variable degrees of language and other cognitive impairment. Especially when it presents in midlife, it can be a significant clinical challenge to distinguish FTD from primary psychiatric disorders. In late-life presentations, it can be difficult to distinguish from a frontal-predominant presentation of Alzheimer disease.

Definition. FTD, as used here, is meant to identify a group of clinical syndromes characterized by progressive circumscribed atrophy of frontal and temporal lobe cortices, identified as frontotemporal lobar degeneration. FTD includes two broadly accepted subgroups: behavioral-variant FTD (bv-FTD); and language-predominant FTD (including progressive nonfluent aphasia and semantic dementia). The spectrum of presentations of frontotemporal lobar degeneration also includes the motor syndromes of CBD, PSP, and motor neuron disease, as there are both clinical and pathologic overlaps among these.

History. Arnold Pick, in the late 19th and early 20th century, provided the initial descriptions of the clinical syndrome of a dementia with associated circumscribed atrophy of the frontal and temporal lobes. Alzheimer subsequently described associated neuropathologic abnormalities, including the presence of intraneuronal inclusions and ballooned neurons. The inclusions were later named Pick bodies, and the syndrome of dementia with frontotemporal atrophy and Pick bodies was named Pick disease.

Comparative Nosology. Consensus criteria for the diagnosis of FTD were first published in 1998. This group recognized that the prototypic clinical presentation of *FTD* is disturbances of personality and behavior or impairments of language (primary progressive aphasia or semantic dementia).

The diagnostic criteria for the bv-FTD were updated most recently in 2011 by the International Behavioral Variant FTD Consortium. The updated criteria are meant to provide greater sensitivity and more flexibility in the diagnosis of FTD in its earlier stages. For a diagnosis of possible bv-FTD, three of six behavioral and cognitive symptoms must be demonstrated (behavioral disinhibition; apathy or inertia; loss of sympathy or empathy; perseverative, stereotyped or compulsive/ritualistic behavior; hyperorality or dietary changes; and executive or generation deficits with relative sparing of memory and visuospatial functions). A diagnosis of probable bv-FTD requires the criteria for possible bv-FTD, as well as significant functional decline and imaging consistent with bv-FTD. In 2011, an international workgroup published criteria for the language-predominant forms of FTD: progressive nonfluent aphasia (also known as nonfluent/agrammatic), semantic dementia, and added a third variant, logopenic. Descriptions of these variants can be found later in this section.

The DSM-5 includes both a behavioral variant and a language variant in its description of major frontotemporal neurocognitive disorder. The behavioral variant is characterized by a decline in social cognition or executive functioning, as well as behavioral disinhibition, apathy or inertia, loss of sympathy or empathy, perseverative, stereotyped or compulsive behavior, and hyperorality. The language variant is characterized by a prominent decline in language abilities, with difficulties seen in speech production, word-finding, object naming, grammar, or word comprehension. In both variants, there is an insidious onset, with gradual progression and a relative sparing of learning, memory, and perceptual-motor functioning.

Epidemiology. In early age of onset dementia, the prevalence and incidence of FTD have been reported to be similar to that of early-onset Alzheimer dementia. FTD is the third most common form of dementia, following Alzheimer dementia and DLB, and is the second most common form in patients younger than 65. Incidence rates have been reported to be 3.5 per 100,000 person-years in a 45 to 64 years age group and 3.3 per 100,000 person-years in a 50 to 59 years old age group. In population-based studies conducted in the United Kingdom and the Netherlands, prevalences of FTD of 3.6 to 15 cases per 100,000 individuals in mid- to late-life have been reported. Age at presentation is most commonly in the sixth decade, but this can range from the third to the ninth decade. The data on the role of gender is mixed. A 2005 study examining the demographic characteristics of FTD did find a male predominance in the behavioral variant and semantic dementia and a female predominance in progressive nonfluent aphasia.

Etiology. FTD occurs in both sporadic and familial forms. A family history is present in up to approximately one-third to one-half of cases, typically with an autosomal dominant pattern of transmission. The etiology of the sporadic forms is currently unknown.

GENETICS. The most common genetic mutations in familial FTD include those in the microtubule-associated protein tau (*MAPT*), progranulin (*PGRN*), and chromosome 9 open reading frame 72 (*C9orf72*) expansion mutations. FTD attributed to *MAPT* mutations tends to present as the behavioral variant and at younger ages compared to FTDs attributable to other mutations. FTD attributed to *PGRN* mutations also tends to present as the behavioral variant, and neuropsychiatric symptoms are common. About half of patients with *PGRN*-associated FTD have extrapyramidal symptoms, and early language involvement is more common in *PGRN*-associated FTD compared to FTDs attributable to other mutations. The *C9ORF72* mutation, which is a hexanucleotide expansion, typically presents as the behavioral variant-FTD, motor neuron disease, or a combination of the two. Apathy, disinhibition, loss of empathy, and executive dysfunction are common. This mutation is the most common cause of ALS-FTD, and is associated with transactive response DNA-binding protein (TDP-43). Approximately 26% of familial and 5% of sporadic FTD cases may carry *C9ORF72* expansions. Mutations in *GRN* found on chromosome 17 that result in loss of function may account for up to 5% to 10% of FTD cases. Mutations in vasolin-containing protein (*VCP*) present with common symptoms of behavioral change, executive dysfunction, and aphasia, and mutations in transactive DNA-binding protein (*TARDBP*) are associated with both behavioral and semantic subtypes and can present with and without motor neuron disease. Genetic testing should be considered if one first-degree relative is affected with mutations in *GRN*, *MAPT*, or *C9orf72*, and testing for *C9orf72* should be considered for individuals with behavioral variant FTD or in individuals with late-onset psychiatric disorders and one first-degree relative with FTD or ALS.

Diagnosis and Clinical Features. The temporal evolution of symptoms and their neuroanatomical correlates are important in the diagnostic clarification of the differing variants.

In the bv-FTD, apathy, disinhibition, and loss of self and other awareness are early signs. Family members and loved ones may be concerned that the individual is experiencing a midlife crisis of some sort, and often a psychiatric disorder is suspected as the underlying etiology of this change in behavior early on.

Behavioral disinhibition can be characterized as socially inappropriate behavior or a loss of manners or decorum. An individual may begin to make impulsive, rash, or careless actions. Apathy or inertia can also be seen. There may also be an early loss of sympathy or empathy, with an individual displaying diminished reactivity to other’s needs or feelings, or diminished social interest, interrelatedness, or personal warmth. Individuals may be less supportive of a spouse’s health problems. Perseverative, stereotyped, compulsive, or ritualistic behaviors can also be seen. These may range from simple repetitive movements to more complex or compulsive behaviors, like collecting specific objects or hoarding. Stereotypy of speech, as well as mental rigidity, can also be seen. Individuals may display hyperorality or dietary changes, including altered food preference, such as eating foods that are just one color or binge eating. Neuropsychological testing reveals executive function deficits, with relative sparing of memory and visuospatial functioning. Individuals with the behavioral variant often have very limited insight into their alterations of behavior.

About 20% of individuals with the behavioral variant also show signs of parkinsonism, with an akinetic rigid form present in about 60% of these individuals and tremor predominant present in 40%. Five percent to ten percent of individuals with FTD have co-occurring motor neuron disease as well, with symptoms including muscle wasting, weakness, and fasciculations. The motor neuron disease may occur prior to, during, or after a diagnosis of FTD has been made. These individuals have ubiquitin and TDP-43 positive inclusions in the lower motor neurons of the spinal cord and the granular cells of the hippocampus. FTD and motor neuron disease are felt to be part of the same clinical spectrum of TDP-43 pathology.

Three different language variants have been described.

Core features of the semantic variant, also known as semantic dementia, include impaired confrontation naming and impaired single-word comprehension. Other diagnostic features include impaired object knowledge, surface dyslexia (difficulty with recognition of irregular sounding words, like yacht), or dysgraphia. Individuals may substitute categories for words (“food” for “apple”) and eventually may refer to all nouns as “things.” As the disease progresses, speech becomes empty of content. Repetition and speech production are usually spared. Those individuals with the semantic variant but who have right-sided predominant atrophy may show some early behavioral features, including emotional detachment or lack of empathy, dysfunction in emotional processing, compulsions, or may also have impairments in facial recognition.

The nonfluent/agrammatic variant, also known as progressive nonfluent aphasia, requires one of the two core features for a diagnosis: agrammatism in language production or effortful, halting speech not consistent with apraxia of speech. Speech is often described as slow and labored. Speech sound errors are made, and there is abnormal prosody. There may also be impaired comprehension of syntactically complex sentences. Single-word comprehension and object knowledge are spared. Individuals with progressive nonfluent aphasia tend to have more insight into their deficits and maintain social norms; if atrophy progresses to the right frontal regions, behavioral symptoms are likely to follow.

Core features of the logopenic variant include impaired single-word retrieval in spontaneous speech and naming and impaired repetition of sentences and phrases. Speech is slow due to word-finding problems. There may be phonologic errors in speech and naming. Single-word comprehension, object knowledge, and motor speech are spared, and there is a lack of agrammatic errors, with preservation of articulation and prosody. As some of these individuals go on to show amyloid pathology on autopsy, this variant may actually be a less common variant of Alzheimer disease, particularly if these individuals show a longitudinally stable logopenic pattern.

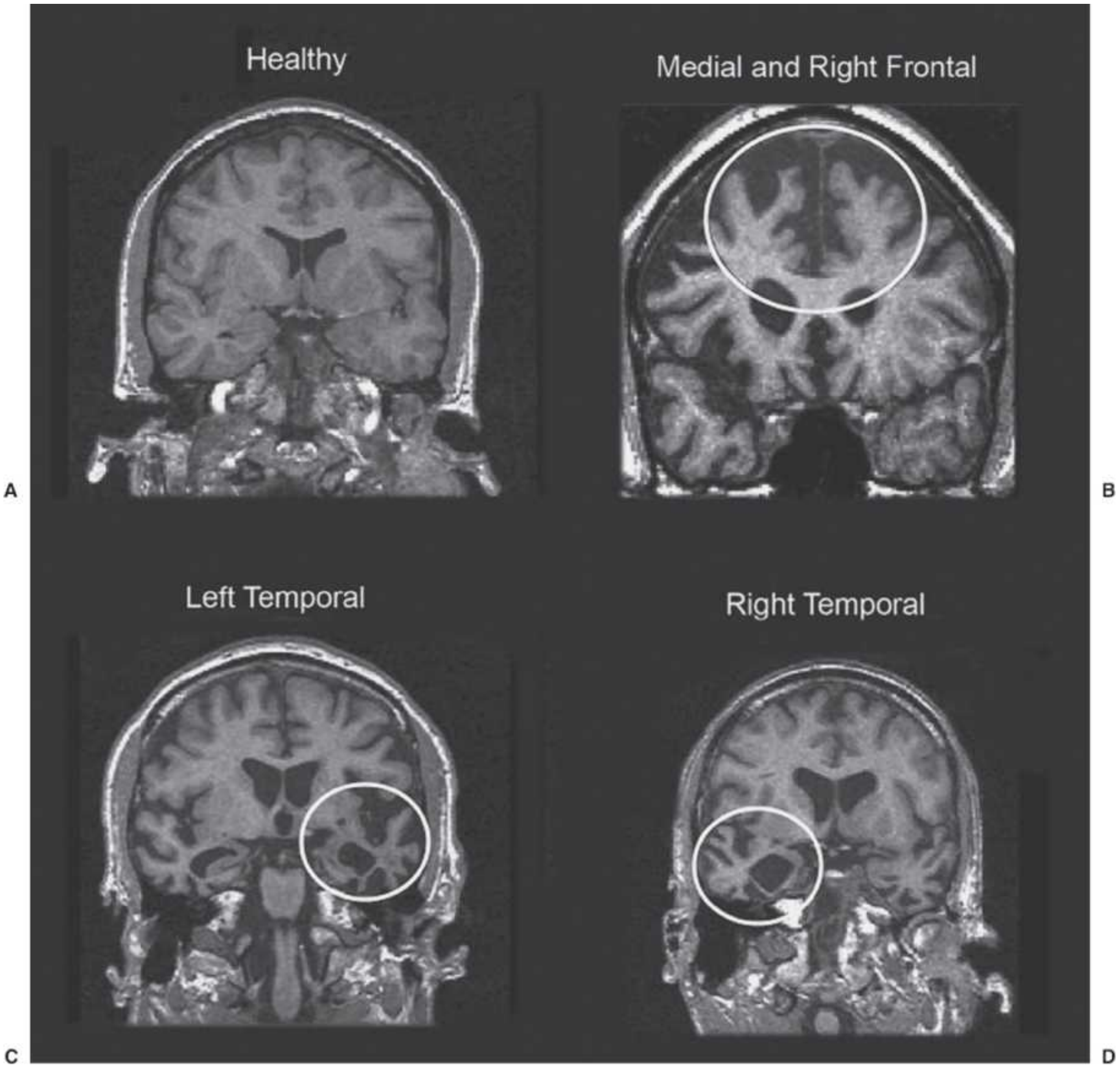

FIGURE 10.3–4. Patterns of atrophy (frontotemporal lobar degeneration [FTLD]) in patients with frontotemporal dementia. T_1-weighted coronal magnetic resonance images of **(A)** a healthy age-matched control participant, **(B)** a patient with frontal-variant FTLD, **(C)** a patient with temporal-variant FTLD showing predominantly left-sided atrophy, and **(D)** a patient with temporal-variant FTLD showing predominantly right-sided atrophy. (Images provided by Indre Viskontas, Ph.D., and Bruce Miller, M.D., the Memory and Aging Center, University of California San Francisco.)

Biomarkers. Currently, there are no established biomarkers for the diagnosis of FTD. In practice, the use of MRI with T_1 and FLAIR sequences, including coronal cuts, is helpful to establish the presence of circumscribed frontal and/or temporal atrophy (Fig. 10.3–4).

In the behavioral variant of FTD, there is early degeneration of the paralimbic structures of the ventromedial prefrontal cortex, anterior cingulate cortex, and anterior insula on neuroanatomical imaging. MRI may demonstrate volume loss in the prefrontal cortex, anterior temporal regions, insula, anterior cingulate, striatum, and thalamus. In individuals where there is diagnostic uncertainty and without clear frontotemporal atrophy on CT or MRI imaging, SPECT or FDG-PET can be considered, with preference given to FDG-PET imaging if possible. PET or SPECT imaging would reveal hypoperfusion or hypometabolism in the frontal or anterior temporal regions in the behavioral variant, with specific changes evident in the uncinated fasciculus, genu of the corpus callosum, and cingulum, superior and inferior longitudinal fasciculi, and inferior fronto-occipital fasciculus.

In primary progressive aphasia, atrophic changes evident on MRI begin asymmetrically, occurring primarily in the left hemisphere. Imaging may demonstrate predominant anterior temporal lobe atrophy or predominant anterior temporal hypoperfusion/hypometabolism on SPECT or PET. In the semantic variant, there is decreased activity in the left superior longitudinal fasciculus, corpus callosum, left cingulum, left orbitofrontal, inferior frontal, anterior temporal, and inferior white matter regions. In the nonfluent variant, functional neuroimaging findings may include decreased activity in the left uncinated fasciculus, inferior longitudinal fasciculus tracts, corpus callosum, and cingulum, with sparing of the occipital white matter, brainstem, and cerebellum. In the logopenic variant, MRI, SPECT, or PET scans show predominant left posterior perisylvian or parietal atrophy, hypoperfusion, or hypometabolism.

In CSF studies, one would expect CSF amyloid-β_{42} to be normal, while CSF tau levels may be normal to increased due to the neurodegeneration occurring.

Differential Diagnosis. Not surprisingly, given the frequent initial presentation of progressive alterations in behavior, psychiatric syndromes are often initially suspected in individuals with FTDs, with up to 50% of individuals with the behavioral variant of FTD first being diagnosed with a primary psychiatric disorder, most commonly depression. Individuals with FTD caused by mutations in *C9ORF72* can often present with prodromal psychiatric symptoms, including symptoms suspicious of bipolar disorder, obsessive-compulsive disorder, and schizophrenia. In GRN mutations, visual hallucinations and delusions develop in up to a quarter of patients at some point in their illness. The Neuropsychiatric International Consortium for Frontotemporal Dementia (NIC-FTD) developed guidelines to assist with distinguishing bv-FTD from primary psychiatric disorders. In addition to a detailed history of symptoms, the NIC-FTD supports the use of FTD-specific scales, such as the Frontal Behavioral Inventory, the Stereotypy Rating scale, the DAPHNE, and the Frontotemporal Dementia versus Primary Psychiatric Disorder Checklist. A bedside exam should look for the presence of parkinsonism, upper motor neuron signs, and tests of executive function, all of which may be helpful in distinguishing between bv-FTD and a primary psychiatric disorder. Social cognition is typically more impaired in bv-FTD. Atrophy in the frontal or anterior temporal lobes makes a diagnosis of bv-FTD more likely. Serum or CSF neurofilament light chain markers, which are indicative of neurodegeneration, may be helpful in the differentiation of bv-FTD from primary psychiatric disorders as well.

Differential diagnosis may also include the other causes of dementia, commonly frontal predominant presentations of Alzheimer disease, particularly if the logopenic variant of primary progressive aphasia is diagnosed, or the combination of cognitive impairment and parkinsonism found in DLB. The absence of amyloid binding is helpful in ruling out AD-related pathology. The presence of parkinsonism requires that one also consider the diagnoses of CBD and PSP. For example, the additional symptoms of progressive parkinsonism, axial rigidity, and supranuclear gaze palsy would indicate an FTD with PSP. Motor symptoms of asymmetric rigidity and apraxia may signal an associated CBD. Both PSP and CBD will be discussed in further detail below.

Course and Prognosis. The course of FTD is one of progressive deterioration. With progression, the differing initial presentations described above frequently merge, and ultimately loss of other cognitive functions may occur. Motor symptoms can precede, emerge concurrently, or follow the onset of the changes in behavior, personality, and language. On average, survival is eight years from symptom onset and 4 to 5 years from diagnosis. Those individuals with FTD and motor neuron disease have a significantly shorter survival time compared to those without motor neuron disease. Overall, individuals with FTD show more rapid cognitive and functional decline than those individuals with Alzheimer disease.

Neuropathology. The pathologic examination has been essential in revealing the underlying processes resulting in FTD. Macroscopically, FTD is characterized by focal atrophy of the frontal cortex, the temporal cortex, or both. Atrophy is usually symmetric, though asymmetry is sometimes present. The Consortium for Frontotemporal Lobar Degeneration has elaborated guidelines for the neuropathologic assessment of cases with frontotemporal lobar degeneration. A systematic survey of frontal, temporal, and parietal cortices and the hippocampus with histologic stains and immunocytochemistry using antibodies directed against τ, ubiquitin, or P62 (sequestosome-1) and α-synuclein is recommended to identify lesions associated with the frontotemporal lobar degenerations and to exclude other neurodegenerative dementias. Further evaluation of additional brain regions, including the basal ganglia, midbrain, brain stem, and spinal cord, with the above staining approaches, immunocytochemistry for TDP-43 and for τ isoforms, and molecular genetic identification of mutations may also be necessary to achieve specific diagnoses.

Using these approaches, the frontotemporal lobar degenerations can be divided into two major groups, one with τ-positive inclusions, another with τ-negative and ubiquitin/P62-positive inclusions. These two presentations do not account for the behavioral versus language-predominant presentations of FTD, which are presumed to relate more to the location of pathology. The associated motor syndromes, however, do demonstrate selective associations with these pathologic groupings, as do genetic etiologies.

The frontotemporal lobar degenerations with τ-positive inclusions include classically described Pick disease, CBD, PSP, and dementia due to MAPT mutations. These subtypes can, to some extent, be distinguished by the pattern of lesion location and morphology, as revealed through the use of antibodies that nonselectively recognize phosphorylated τ. However, distinguishing among subtypes may be more reliably achieved through the use of antibodies selective for τ isoforms.

The second major group is the frontotemporal lobar degenerations with τ-negative ubiquitin/P62-positive inclusions. The most frequent pathologic presentation in this group and the most frequent pathologic presentation among all cases of frontotemporal lobar degenerations is the presence of intracellular inclusions positive for TDP-43. It is now recognized that many cases formerly classified as dementia lacking distinctive histologic features are TDP-43 positive and that if dementia lacking distinctive histologic features exists, it is quite rare. FTDs with TDP-43 lesions can occur with or without associated motor neuron disease. FTDs due to *GRN* and *VCP* mutations are TDP-43 positive, while those due to *CHMP2B* mutations are ubiquitin/P62 positive and TDP-43 negative. FTD due to *C9ORF72* presents with both TDP-43 positive and ubiquitin/P62 positive/TDP-43 negative inclusions, the latter most often found in the cerebellum.

Management. Similar to other dementias, nonpharmacologic management should be integrated throughout the course of the dementia process and are the first-line for the treatment of neuropsychiatric symptoms. Currently, there are no treatments for the cognitive deficits associated with FTD and no treatments to prevent the progression of the underlying pathologies. Data does not support the use of cholinesterase inhibitors or memantine, and there is some concern that these medications can worsen behavioral symptoms. Symptomatic use of trazodone, second-generation antipsychotics, SSRIs, and anticonvulsants may help agitation, disinhibited, and aggressive behavior. Typically, SSRIs are first-line pharmacotherapy treatments for FTD, with three small open-label studies providing some evidence for the use of SSRIs, although these are not FDA-approved for the treatment of FTD. Second-generation antipsychotic medications should be used with caution and close observation, as individuals with FTD may be more sensitive to extrapyramidal side effects. Stimulants may be helpful with risk-taking behaviors and disinhibition, but data is limited.

Corticobasal Degeneration

CBD can present as a syndrome with both clinical and pathologic heterogeneity, with the clinical syndrome being more reflective of the neuroanatomical distribution of pathology than a particular histology.

Definition. The clinical syndrome of CBD has core features of progressive asymmetric rigidity and apraxia, as well as cortical and basal ganglia dysfunction.

Comparative Nosology. This clinical entity was first identified in 1967 in three individuals with a progressive asymmetrical akinetic-rigid syndrome and apraxia. It was later renamed CBD in the early 1990s.

The term corticobasal syndrome (CBS) is sometimes used to describe the clinical features seen, reserving the term CBD for the histopathologic diagnosis. Other names for this syndrome that have been used in the past include corticonigral degeneration and corticobasal ganglionic degeneration. In 2013, an international consortium published diagnostic criteria for the diagnosis of probable CBD and a broader set of criteria for possible CBD. Possible CBD includes phenotypes that may be consistent with CBD but which may also overlap with other tauopathies, such as PSP or FTD.

Epidemiology. The age of onset has been reported to range from 45 to 77.

Etiology. CBD is, for the most part, a sporadic disorder. The h1 haplotype of the MAPT gene has been associated with the development of this disease.

Diagnosis and Clinical Features. For a diagnosis of probable sporadic CBD, there must be insidious onset with gradual progression for at least one year, no family history of the disease, and an age of onset of at least 50 years of age. For a diagnosis of the broader possible CBD, there is no minimum in terms of age of onset, and family history and a genetic mutation affecting τ are both permitted. There are four major phenotypes of CBD described; these include CBS, frontal–behavioral–spatial syndrome, nonfluent/agrammatic primary progressive aphasia, and PSP syndrome. There is also increasing evidence for a predominant cognitive phenotype as well.

CBS is one of the more common presentations of CBD. Clinical features of probable CBS include asymmetric presentation of limb rigidity or akinesia, limb dystonia, and limb myoclonus, as well as orobuccal or limb apraxia, cortical sensory deficit, and alien limb phenomenon. Possible CBS shares similar clinical features, but they may be symmetric.

PSP syndrome is only considered a phenotype of possible CBD; clinical features of this include one CBS symptom and three of the following symptoms: axial or limb rigidity or akinesia, postural instability or falls, urinary incontinence, behavioral changes, supranuclear vertical gaze palsy or decreased velocity of vertical saccades. Tremor is a common presenting symptom in CBD, but in contrast to tremor in Parkinson disease, tremor in CBD is minimal at rest and amplified with action. Furthermore, symptoms are not responsive to L-dopa therapy. Neuropsychological testing may reveal impairments in attention and concentration, executive function, language, and visuospatial function; learning and memory are often spared. Aphasia and apraxia of speech may also be evident, as can depression and apathy. Features of an FTD with behavioral disturbance can also be seen, with executive dysfunction, behavioral or personality changes, or visuospatial deficits present in the frontal behavioral spatial syndrome phenotype. Dementia is common in pathologically confirmed cases of CBD.

Biomarkers. There are no CSF or serum biomarkers available for diagnosis. On structural imaging, asymmetrical cortical atrophy can be seen, particularly in the frontoparietal regions; atrophy is more prominent in the hemisphere contralateral to the side of the symptoms. There may also be asymmetrical atrophy in the basal ganglia and cerebral peduncles. SPECT and PET can have asymmetrical hypoperfusion in the parietal and frontal cortex and basal ganglia. In dopamine transporter imaging, there is an asymmetric reduction in tracer uptake in the nigrostriatum, with greater reductions seen contralateral to the more clinically affected side.

Differential Diagnosis. The presence of parkinsonism on examination causes one to include the other Parkinson plus syndromes in the differential. There is a significant overlap between CBD and PSP; those individuals with the CBD phenotype of PSP syndrome are more likely to have PSP, but only autopsy provides a definitive diagnosis. Individuals with CBD may have underlying PSP pathology, and those with the classical symptoms of PSP may have CBD pathology. Differences in cognitive impairments may be helpful in distinguishing PSP and CBD. In PSP, there is predominant axial rigidity, as opposed to the progressive asymmetric rigidity seen in CBD. In contrast to Parkinson disease, the tremor typically seen in CBD is minimal at rest and amplified with action, and the parkinsonian symptoms are typically not responsive to dopaminergic medications. Visual hallucinations should cause one to consider DLB. FTD should also be considered. Finally, the focal nature of deficits in CBD should also lead to the consideration of vascular dementia, but one would expect to see infarcts on imaging in this case.

Course and Prognosis. Symptoms are usually asymmetric at first, but most often, the rigidity and apraxia begin in an upper limb and then spread about two years into the disease course to either the ipsilateral lower limb or the contralateral upper limb. Median survival time has been reported to be about 6 to 7 years.

Neuropathology. On neuropathologic examination, there is focal cortical neuronal loss, typically in the frontal, parietal, or temporal lobes, as well as neuronal loss in the substantia nigra. The pathologic features of CBD include Gallyas/tau-positive astrocytic plaques and threads, particularly in the superior frontal gyrus, the superior parietal gyrus, pre- and postcentral gyri, and the striatum. Tau-positive oligodendroglial coiled bodies may be evident. Ballooned neurons are considered to be a hallmark feature but are not required for a diagnosis.

Management. Symptomatic treatments of the motor symptoms include botulinum toxin for focal dystonia, propranolol or primidone for tremor, and clonazepam for myoclonus, as the parkinsonism present is not typically responsive to L-dopa therapy. The importance of physical, occupational, and speech therapies has been emphasized.

Progressive Supranuclear Palsy

Definition. PSP is a gradually progressive disorder with progressive supranuclear gaze palsy, postural instability and falls, and parkinsonism.

Comparative Nosology. In 1963, a case series on PSP was published; initially described as “heterogeneous system degeneration,” this report detailed eight patients who had impairments in ocular gaze, spasticity of facial musculature with dysarthria, and at times dysphagia, nuchal rigidity, and dementia. This was officially termed PSP in 1964.

Formal diagnostic criteria were initially proposed in 1987. In 1996, the National Institute for Neurological Disorders and Stroke and the Society for PSP (NINDS-SPSP) refined the diagnostic criteria. At the time, core features included age of onset over the age of 40, with vertical supranuclear palsy or both vertical supranuclear palsy and prominent postural instability with falls in the first year of symptoms; this presentation is referred to as Richardson syndrome. However, this combination of symptoms failed to capture many of the other phenotypes presented earlier in the disease process. As such, this criteria was most recently updated in 2017 by the International Movement Disorder Society, with additional emphasis placed on the domains of akinesia and cognitive dysfunction, with the goal of promoting earlier diagnosis. For a diagnosis of PSP to be made, there must be sporadic occurrence of illness, occur in an individual age 40 years or older, and there must be a gradual progression of PSP-related symptoms, which fall into the four functional domains of oculomotor dysfunction, postural instability, akinesia, and cognitive dysfunction.

Epidemiology. The average annual incidence rate of PSP for ages 50 to 99 has been estimated to be 5.3 per 100,000 person-years. The incidence rate rises with increasing age, with an estimated incidence ranging from 1.7 for ages 50 to 59 to 14.7 person-years for ages 80 to 99. Incidence has been noted to be slightly higher in men. However, given previous diagnostic constraints, and when accounting for other clinically variant PSP phenotypes, the true prevalence is likely higher.

Etiology. PSP is considered to mostly be a sporadic disease. The H1 haplotype of the MAPT gene has been associated with PSP; inheritance of two copies of the H1 haplotype is considered to be a genetic risk factor. A large genome-wide association study identified *STX6*, *EIF2AK3,* and *MOBP* as additional risk alleles. Areas on chromosome 11 containing the DNA damage-binding protein 2 (DDB2) and the lysosomal acid phosphatase 2 (ACP2) have also been identified as areas of interest.

Diagnosis and Clinical Features. The median age of onset is 63. Diagnosis has typically been made about 3 to 4 years after symptom onset. The 2017 Movement Disorder Society criteria describe four functional domains of symptoms that may be evident, with each domain including symptoms that carry with them a variable degree of certainty pertaining to the diagnosis. Each of these domains is described as follows, with their respective symptoms in decreasing order of certainty as they pertain to the diagnosis of PSP. Oculomotor dysfunction may include vertical supranuclear gaze palsy, slow velocity of vertical saccades, and frequent macro square wave jerks or “eyelid opening apraxia.” Postural instability is characterized by repeated unprovoked falls within 3 years, a tendency to fall on the pull-test within 3 years, or the individual taking more than two steps backward on the pull-test within 3 years. An individual meeting the criteria for akinesia may show progressive gait freezing within 3 years, parkinsonism which is akinetic-rigid, predominantly axial and levodopa resistant, or parkinsonism with tremor and/or asymmetric predominance of limbs and/or levodopa responsive. Cognitive dysfunction may manifest as a speech or language disorder such as which is seen in the nonfluent variant of primary progressive aphasia or progressive apraxia of speech, or it could manifest as frontal cognitive or behavioral presentations, which could include apathy, slowed thinking, dysexecutive syndrome, reduced verbal fluency, and impulsivity, or CBS. Supportive features in the diagnosis may include levodopa resistance, hypokinetic, spastic dysarthria, dysphagia, and photophobia. Taking various combinations of the above symptoms into account, there are multiple clinical phenotypes of PSP, which include PSP with Richardson syndrome, PSP with predominant parkinsonism, PSP with progressive gait freezing, PSP-CBS, PSP with predominant speech or language disorder, PSP with predominant frontal presentation and PSP with predominant cerebellar ataxia. Most patients eventually progress to show the constellation of symptoms evident in Richardson syndrome. Individuals who meet the criteria for PSP-CBS have at least one cortical sign (apraxia, cortical sensory deficit, alien limb phenomenon) and one movement disorder sign (limb rigidity, akinesia, or myoclonus) which can be asymmetric or symmetric.

Biomarkers. Imaging findings may include predominant midbrain atrophy and atrophy of the superior cerebellar peduncles on anatomical MRI. On MRI, atrophy of the midbrain tegmentum, sometimes referred to as the morning glory sign because of its concave appearance, and the hummingbird sign, in which the atrophy of the rostral midbrain as seen on midsagittal images, looks like a hummingbird has high specificity for PSP but lower sensitivity. On SPECT and PET imaging, there is frontosubcortical hypoperfusion and hypometabolism; these changes are usually symmetric. There may be hypometabolism predominant in the midbrain relative to the pons on [^{18}]FDG-PET imaging, as well as postsynaptic striatal dopaminergic degeneration on [^{123}]IBZM-SPECT or [^{18}F]-DMFP-PET.

Neurofilament light chain concentrations may be elevated in both CSF and plasma, with higher baseline plasma concentrations associated with faster progression of symptoms.

Differential Diagnosis. Given the variable presentations of PSP, the differential can be quite broad. Imaging can be useful in patients with sudden onset or step-wise progression of the disease to help exclude stroke, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) or severe cerebral amyloid angiopathy. CSF biomarkers which are present in Alzheimer disease, as well as positive β-amyloid PET imaging, may be helpful in excluding PSP. The CBS phenotype is a rare phenotype of PSP, and those who meet the criteria for this fall under the category of possible PSP, but probable 4-repeat tauopathy, as both CBD and PSP, are an accumulation of the 4R tau isoform. In younger patients, it is recommended that Wilson disease, Niemann–Pick disease, hypoparathyroidism, neuroacanthocytosis, and neurosyphilis be excluded based on appropriate laboratory work-up. Whipple disease should also be considered in younger patients who present with atypical neurologic features in addition to gastrointestinal symptoms, arthralgias, and fever. In patients whose illness appears to be rapidly progressive, prion disease and paraneoplastic encephalitis should be considered and ruled out. Given the progressive parkinsonism present in PSP, Parkinson disease should also be considered on the differential. FTD and CBD should also be considered, given the symptom and neuropathologic overlap.

Course and Prognosis. Survival time after symptom onset is about 5 to 9 years.

Neuropathology. Neuropathologic exam is the gold standard for diagnosis. On neuropathologic examination, there are neurofibrillary tangles and neuropil threads composed of tau proteins in the basal ganglia and brainstem. The presence of tau-positive astrocytes in these areas is supportive of the diagnosis. There may also be neuronal loss, neurofibrillary tangles, tufted astrocytes, and oligodendroglial coiled bodies.

Management. Similar to CBD, there is a poor response to L-dopa. Physical, occupational, and speech therapy may be helpful for symptomatic management throughout the course of the disease.

Dementia with Lewy Bodies

Introduction. DLB is now considered to be the second most common cause of dementia after Alzheimer disease in the older adults. Operationalized diagnostic criteria for DLB have been defined and revised over the past decade. DLB shares some clinical and pathologic features with both Alzheimer disease and Parkinson disease. As a result, complete discrimination among these syndromes, even at autopsy, may not be possible.

Definition. DLB is neurodegenerative dementia characterized by a progressive cognitive decline of sufficient magnitude to interfere with normal social or occupational function. Dementia must be present for a diagnosis of DLB to be made. Memory impairment need not be prominent or persistent early in the course of the illness. Deficits of attention and frontal executive skills, and visuospatial ability may be more prominent early in the course. The core features of DLB include (1) fluctuating cognition with pronounced variations in attention and alertness, (2) recurrent, complex visual hallucinations, which are typically well-formed and detailed, (3) spontaneous motor features of parkinsonism, and (4) REM sleep behavior disorder. Supportive clinical features include severe sensitivity to antipsychotic agents, repeated falls and syncope, transient unexplained loss of consciousness, severe autonomic dysfunction, hypersomnia, hyposomnia, hallucinations in other modalities, systematized delusions, apathy, anxiety, and depression. Indicative biomarkers include reduced dopamine transporter uptake in the basal ganglia demonstrated by SPECT or PET, abnormal (low uptake) ^{123}iodine-MIBG myocardial scintigraphy, and polysomnographic confirmation of REM sleep without atonia. Supportive biomarkers include relative preservation of medial temporal lobe structures on CT/MRI scan, generalized low uptake on SPECT/PET perfusion/metabolism scan with reduced occipital activity plus/minus the cingulate island sign on FDG-PET imaging, and prominent posterior slow-wave activity on EEG with periodic fluctuations in the pre-alpha/theta range.

History. Lewy bodies, the intracytoplasmic, spherical, eosinophilic neuronal inclusion bodies, identified originally in subcortical nuclei as one of the hallmarks of idiopathic Parkinson disease, have also been identified in the brain stems and cortex of older adults with dementia. Lewy body disease was first described in 1984 by Kosaka, who distinguished "diffuse" Lewy body disease, where Lewy bodies are present throughout cortical regions, from "brainstem-predominant" Lewy body disease, where Lewy bodies are present primarily in brain stem nuclei, and a "transitional type," which is intermediate between the two. Clinical and neuropathologic diagnostic criteria were published by the Consortium on Dementia with Lewy bodies in 1996 and revised in 2005.

Comparative Nosology. Several historical terms have been used to describe what is now referred to in general as DLB. These terms reflect the overlap clinically and pathologically with Alzheimer disease and Parkinson disease and include Lewy body variant of Alzheimer disease, Alzheimer disease with Parkinson disease changes, dementia associated with cortical Lewy bodies, diffuse Lewy body disease, and senile dementia of the Lewy body type.

The first consensus clinical criteria for DLB were published in 1996. These criteria defined the core clinical features and those clinical features that were supportive of the diagnosis. The presence of one core feature was required for a diagnosis of possible DLB and at least two core features for probable DLB. The distinction between DLB and Parkinson disease with dementia was operationalized by requiring evidence of dementia within one year of the motor symptoms for a diagnosis of DLB. The criteria did not exclude patients who simultaneously met clinical and/or pathologic criteria for a diagnosis of Alzheimer disease. The criteria for DLB were revised in 2005, with efforts made to clarify the overlap of DLB with Parkinson disease and Alzheimer disease.

Though the 1-year rule for differentiating DLB and Parkinson disease with dementia was recommended in research use, clinically, the criteria were clarified to indicate that DLB should be diagnosed when dementia precedes or is concurrent with parkinsonian symptoms and should not be diagnosed when dementia arises in the "context of well-established Parkinson disease." It was also noted that any such distinction, however, may be seen as arbitrarily dividing a continuum of presentations due to underlying Lewy body disease.

Criteria for the diagnosis of DLB were most recently revised in 2017 by the Dementia with Lewy Bodies Consortium Fourth Consensus and places emphasis on both clinical features and diagnostic biomarkers. Clinical features are considered core or supportive, while biomarkers are considered indicative or supportive. For a diagnosis of probable DLB, two or more clinical features should be present with or without the presence of indicative biomarkers, or only one core clinical feature is present, but with one or more indicative biomarkers. For a diagnosis of possible DLB, only one core clinical feature of DLB is present, with no indicative biomarker evidence, or one or more indicative biomarkers are present, but there are no core clinical features.

The DSM-5 incorporated the 2005 criteria in the diagnosis of major neurocognitive disorder with Lewy bodies. For a diagnosis of major neurocognitive disorder with Lewy bodies, an individual must have two core features or one suggestive feature with one or more core features; if an individual only has one core feature or one or more suggestive features, then a diagnosis of possible neurocognitive disorder with Lewy bodies is made.

Epidemiology. The prevalence of DLB is dependent on whether clinical or pathologic criteria are used and what patient population is studied. As a consequence of these confounds and the unclear clinical and neuropathologic boundaries between DLB, Alzheimer disease, and Parkinson disease dementia in early diagnostic criteria, there remains uncertainty regarding the incidence or prevalence of DLB. The prevalence of DLB in populations over 65 has been found to range from 0.1% to 5%. Clinical studies suggest that DLB likely accounts for 15% to 35% of all dementia patients. One population-based study calculated the incidence of DLB to be 3.5 per 100,000 person-years, with a peak incidence in the 70- to 79-year age group. Neuropathologic studies conducted to date are best understood as describing the frequency of Lewy body pathology in community samples, where it is typically the third most common chronic neuropathologic change after lesions of Alzheimer disease and vascular lesions, and often in combination with both of those.

Etiology. Lewy bodies are comprised of abnormal fibrillar deposits of α-synuclein. Normally, α-synuclein is a protein located in presynaptic axon terminals and is presumed to have a role in synaptic function. Mutations and copy number variations in the gene encoding α-synuclein, *SNCA*, result in disorders with Lewy body pathology, including DLB and Parkinson disease. Mutations in β-synuclein, *SNCB*, and *GBA* resulting in DLB have also been described. Mutations in other genes (*LRRK2, EIF4G1, VPS35, PARK2, PINK1*, and *PARK7*) are associated with autosomal dominant or recessive Parkinson disease. Their relationship with DLB, however, is not clear. The APOE ε4 allele has also been associated with an increased risk of DLB. The causes of most cases of sporadic DLB remain unknown.

Diagnosis and Clinical Features. The essential feature of DLB is a progressive cognitive decline of sufficient magnitude to interfere with normal social or occupational function. Attention, executive functioning, and visuospatial processing deficits are more likely to be evident earlier than deficits in memory or naming. Memory impairment may become more evident with the progression of the illness.

CORE CLINICAL FEATURES. The four core features of DLB are (1) fluctuating cognition with pronounced variation in attention and alertness, (2) recurrent visual hallucinations that are typically well-formed and detailed, (3) spontaneous motor features of parkinsonism, and (4) REM sleep behavior disorder. REM sleep disorder was added as a core clinical feature in the DLB Consortium’s updated 2017 guidelines, and as such, was not considered one at the time of DSM-5 publication.

Fluctuation in cognitive function and pronounced variation in attention and alertness is common in DLB and often occur earlier in DLB compared to other dementias. Fluctuation may occur rapidly in minutes to hours or slower with weekly or monthly variation in the level of attention or alertness. Patients may exhibit daytime drowsiness and lethargy with lengthy naps (hypersomnia), followed by periods of confusion upon awakening. They may have reduced awareness of the environment, manifest as staring into space. They may even appear to “switch off” or go blank at times. There may be episodes of disorganized speech. Then there may be periods of marked lucidity during which patients appear to spontaneously remit. Global performance may fluctuate with cognition, attention, and alertness, with periods of pronounced impairment in functioning alternating with periods of normal or near-normal functioning.

Visual hallucinations occur early in the course of DLB, whereas, in Alzheimer disease, hallucinations are generally a later complication. The visual hallucinations are usually recurrent, well-formed, detailed, and tend to persist. Visual hallucinations occur in up to 80% of individuals with DLB and often feature people, children, or animals. Seeing inanimate objects or writing on the walls and ceiling is not infrequent. Patients may have delusional explanations of the hallucinations, or they may have insight into their unreality. Emotional responses to hallucinatory experiences can range from indifference to amusement to fear. Visual hallucinations may be more common during periods of reduced consciousness and exacerbated by visual impairment. Patients with visual hallucinations demonstrate more profound visuoperceptual dysfunction compared to those without hallucinations. Visual hallucinations correlate with a higher density of Lewy bodies in the anterior and inferior temporal lobe and amygdala, as well as reduced occipital perfusion in the primary and secondary visual cortex.

Spontaneous motor features of parkinsonism are another core feature. Symptoms of parkinsonism include bradykinesia, rest tremor, or rigidity; only one of these symptoms is needed to fulfill these criteria. If dementia onset precedes or is concurrent with the onset of parkinsonian symptoms, a diagnosis of DLB is warranted. The most common extrapyramidal symptoms noted in DLB are rigidity and bradykinesia, followed by hypophonic speech, masked facies, stooped posture, and a slow and shuffling gait. Postural instability is also common. The parkinsonism seen in DLB tends to include greater postural instability, gait impairment, and facial immobility compared to those patients with Parkinson disease without dementia. Resting tremor may occur but is less common in DLB than in Parkinson disease.

REM sleep behavior disorder is now considered a core feature of DLB and is characterized by vivid and often frightening dreams during REM sleep accompanied by lack of muscle atonia allowing for “acting out of dreams.” Patients respond to vivid visual images with vocalizations, thrashing about, arising from bed, and often moving about violently. Bed partners are therefore at risk of injury. Patients usually have little or no recall of their dreams or these episodes. The diagnosis can be confirmed by polysomnography. REM sleep behavior disorder may predate the emergence of parkinsonism or dementia by many years; that is, REM sleep behavior disorder may be a manifestation of an evolving synucleinopathy.

SUPPORTIVE CLINICAL FEATURES. Severe neuroleptic sensitivity is present in about half of patients with DLB. They will demonstrate acute onset or exacerbation of parkinsonism and impaired consciousness with even a low dose of a first- or second-generation antipsychotic agent. Even though some patients may tolerate antipsychotics, this does not preclude a diagnosis of DLB. Sensitivity to antipsychotics has profound treatment implications in DLB. Visual hallucinations, hallucinations in other modalities, and delusions are common, but the standard treatment for these psychotic symptoms, namely antipsychotic medications, can cause severe adverse effects. Therefore, in DLB, antipsychotics should generally be avoided in the treatment of psychotic symptoms.

In addition to visual hallucinations, patients may also have hallucinations in other modalities, as well as a sense of presence and visual illusions. Systematized delusions may arise to provide an explanation for the hallucinations or other perceptual disturbances. They are often complex and bizarre as compared to the less well-developed delusions in Alzheimer disease. The most common delusions in DLB are persecution and theft, phantom boarders, television characters in the room, spousal infidelity, and Capgras syndrome (delusional misidentification syndrome).

Repeated falls are common in DLB, although they are not limited to this subtype of dementia. In DLB, falls are generally not associated with focal neurologic signs and symptoms as would be present in balance disorders or transient ischemic attacks. Syncope may be associated with brain stem and autonomic nervous system Lewy body pathology. Transient loss of consciousness may represent an extreme form of fluctuation in attention and alertness.

Hypersomnia and hyposomnia have also been added as supportive clinical features. Hypersomnia is defined as excessive daytime sleepiness, and hyposomnia tends to occur earlier in DLB than in dementia of the Alzheimer type.

Additional clinical features that are supportive of a diagnosis of DLB are severe autonomic dysfunction, resulting in orthostatic hypotension, constipation or urinary incontinence, and depression, apathy, and anxiety.

Biomarkers. Functional imaging of the dopamine transporter system with SPECT or PET may be used to assess the nigrostriatal dopaminergic system. One would expect to see reduced dopamine transporter uptake in the basal ganglia on a DaT (dopamine transporter) scan. This is one of three indicative biomarkers of DLB.

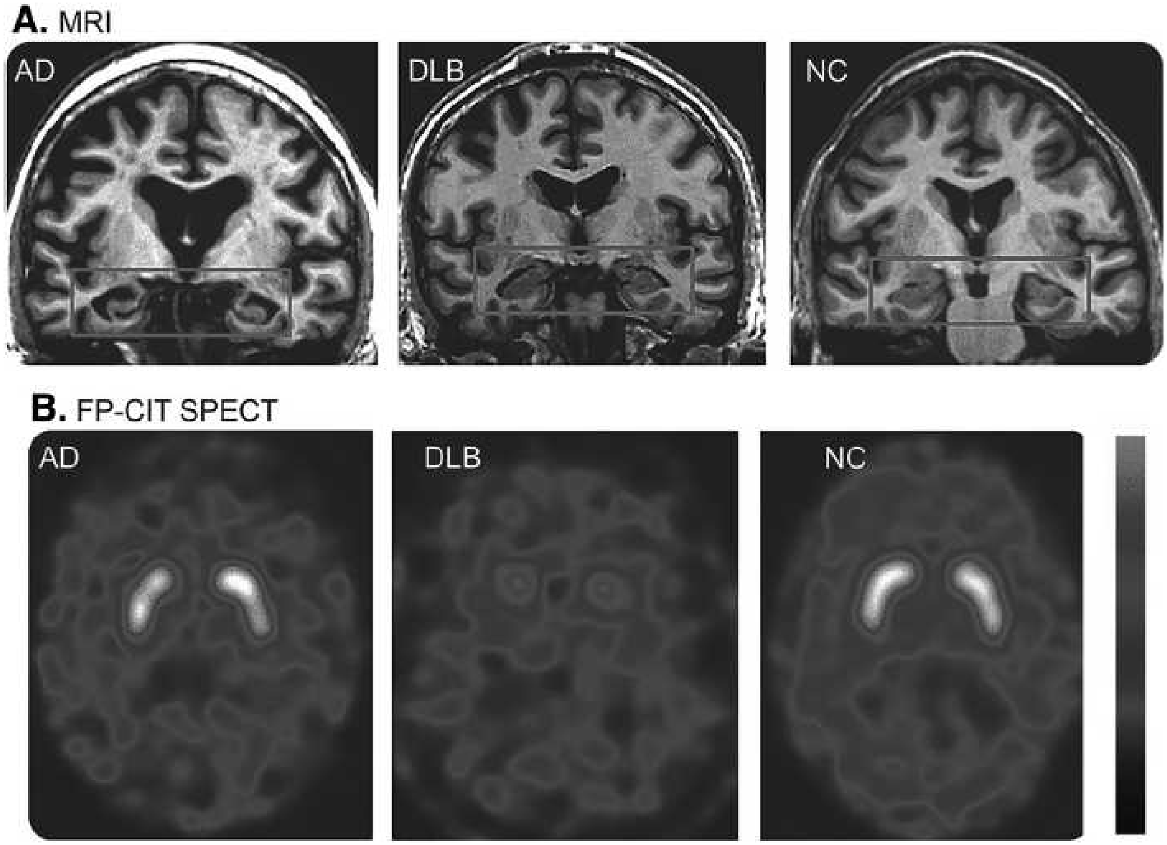

FIGURE 10.3–5. Coronal T1-weighted MRI and ^{123}iodine FP-CIT SPECT images in Alzheimer disease (AD), dementia with Lewy bodies (DLB), and normal controls (NC): **(A)** On the MRI, note the relative preservation of medial temporal lobe volume (rectangles) in DLB, which is similar to NC, whereas atrophy is obvious in AD. **(B)** On the FP-CIT SPECT images, note the minimal uptake in DLB, which is restricted to the caudate (period or full-stop appearance) compared to the robust uptake in the caudate and putamen in AD and NC (comma appearance). (Reproduced with permission from Dr. Val Lowe, Mayo Clinic, Rochester, MN.)

Reduced uptake on metaiodobenzylguanidine myocardial scintigraphy is also considered to be an indicative biomarker of DLB. Postganglionic sympathetic cardiac innervation is reduced in DLB; however, there are additional medical comorbidities as well as medications that may also produce reduced uptake of this.

The third indicative biomarker of DLB is polysomnographic evidence of REM sleep without atonia.

Supportive biomarkers of DLB include relative preservation of medial temporal lobe structures on CT or MRI scan and generalized low uptake on SPECT/PET perfusion scans with reduced occipital activity, with possible posterior cingulate island sign present on FDG-PET (see Fig. 10.3–5). The third supportive biomarker is prominent slow-wave activity on EEG with periodic fluctuations in the prealpha/theta range.

Differential Diagnosis. DLB overlaps with Alzheimer disease both in terms of clinical presentation and neuropathologic changes. The distinction between DLB and Alzheimer disease is complex. The clinical symptoms are overlapping, and no specific cutoff based on the number or pattern of core, suggestive, and supportive features present is definitive, though the presence of at least two of the core features in the early disease course (fluctuating attention, prominent visual hallucinations, and parkinsonism) indicates DLB with high specificity. However, even patients who present with a typical insidious amnestic syndrome characteristic of Alzheimer disease may have Lewy body pathology at autopsy. There may be overlap with the two conditions not being mutually exclusive and patients meeting criteria for both.

There is also considerable clinical and pathologic overlap and comorbidity between DLB and Parkinson disease. The primary diagnosis should be made based on the temporal relationship of the onset of cognitive and motor symptoms. DLB should be diagnosed when dementia occurs with or before parkinsonism. Parkinson disease with dementia should be diagnosed when dementia develops in a patient with well-established Parkinson disease. Dementia develops in at least 30% of patients with Parkinson disease, and perhaps even up to 78% when patients are followed over 8 years. In cases where there is not a clear temporal relationship between the motor and cognitive symptoms, the term *Lewy body disease* may be useful. For research purposes, the “1-year rule” should be used between the onset of parkinsonism and dementia to distinguish between these two entities. While the distinction is useful clinically, both diagnoses likely to represent a common spectrum of Lewy body pathology with a common pathophysiology of abnormalities in α-synuclein metabolism.

Multiple system atrophy (MSA), which also has α-synuclein involvement, may also be considered in the differential. Clinical symptoms of MSA include parkinsonism, cerebellar ataxia, pyramidal symptoms, and autonomic dysfunction. Autonomic dysfunction is necessary for a diagnosis of MSA and includes orthostatic hypotension, erectile dysfunction in males, constipation, urinary symptoms (frequency, urgency, incontinence, and incomplete bladder emptying), and decreased sweating. Dementia is much less common in MSA compared to the other neurodegenerative illnesses with parkinsonism, and according to the second consensus statement, it is a nonsupporting feature of the diagnosis. However, there is recent evidence that certain cognitive domains are affected, with frontal executive dysfunction being the most common and memory and visuospatial dysfunction also showing impairments.

Distinguishing between delirium and DLB can be challenging. Delirium and DLB share several common symptoms. The fluctuating level of alertness and consciousness in DLB is also a hallmark of delirium. Visual hallucinations are also common in both conditions, although they may be less well-formed and less detailed in delirium. DLB, like other dementias, may also increase the risk for a superimposed delirium due to medications, medical illness, and after hospitalization for any reason. Delirium, however, should typically have an acute or subacute onset and a duration limited to the period of exposure to the underlying causal agent. The contribution of medications or other comorbid medical conditions to worsening cognitive, hallucinatory, motor, or behavioral symptoms should be investigated.

Vascular dementia should also be considered in the differential diagnosis. A history of stroke, as evident by focal neurologic signs or brain imaging, makes the diagnosis of DLB less likely. However, white matter lesions in the periventricular and deep white matter regions, microvascular changes, and lacunes may be present in up to 30% of autopsy-confirmed DLB cases, suggesting that a diagnosis of both DLB and cerebrovascular disease may be appropriate in many cases.

CJD may be considered in patients who develop myoclonus, which is also characteristic of a rapidly progressive form of DLB.

Course and Prognosis. DLB has an insidious and gradual onset, with the mean age at onset being 75 years old with a range from 50 to 80 years. There is a slight male predominance. DLB usually presents with one of the core features. It may have a more rapidly progressing course than Alzheimer disease or vascular dementia. Survival time is usually about 10 years, although it may be as little as 1 to 2 years. The course is often marked by psychosis and behavioral disturbances, which makes management more complicated.

Neuropathology. Lewy bodies are the only essential finding in the pathologic diagnosis of DLB. Lewy bodies are spherical intracytoplasmic eosinophilic neuronal inclusion bodies (see Fig. 10.3–6). “Classic” Lewy bodies are inclusions with a hyaline core and pale halo and are typically seen in the brain stem nuclei, substantia nigra, and locus coeruleus. Cortical Lewy bodies are less-well-defined spherical inclusions. Lewy bodies are composed predominantly of fibrillar deposits of α-synuclein and can also include neurofilament proteins and ubiquitin. Though Lewy bodies can be detected with hematoxylin and eosin stain and antiubiquitin antibodies, they are most sensitively and specifically labeled with antibodies against α-synuclein. In addition to forming intracytoplasmic inclusions, aggregates of α-synuclein occur in neuronal processes called Lewy neurites. The number of cortical Lewy bodies correlates significantly with cognitive impairment in DLB. There are also other pathologic changes that may be seen in DLB but that are not required for diagnosis. Alzheimer disease pathology, particularly amyloid-containing plaques, is also a common feature in most cases of DLB. In contrast, the presence of cortical neurofibrillary tangles indicates that DLB is less likely. Microvacuolation or spongiform changes may also occur in some cases of DLB.

Management. Evidenced-based treatments for DLB remain limited. However, early diagnosis and detection are essential to allow for proper management of symptoms as well as to reduce harm from potentially prescribed antidopaminergic and anticholinergic medications.

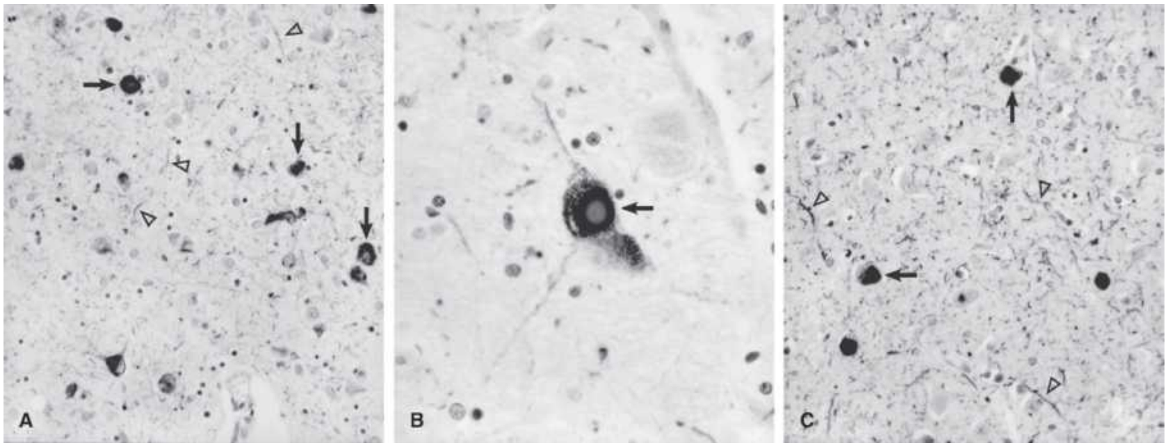

FIGURE 10.3–6. Photomicrographs of Lewy body pathology. **(A)** Abnormal accumulation of α-synuclein aggregates demonstrated by immunocytochemistry in the amygdala of a subject with dementia. Lewy bodies appear as dense intracellular inclusions (*arrows*), but staining of neuronal processes can be seen throughout the neuropil (*arrowheads*). In individuals in whom Lewy body pathology occurs concurrently with Alzheimer disease, the amygdala is often the only region affected. **(B)** Classic appearance of a Lewy body (*arrow*) in a large pigmented neuron of the substantia nigra. **(C)** Lewy body pathology in the neocortex. Both Lewy bodies (*arrows*) and substantial labeling of neuronal processes in the neuropil (*arrowheads*) are evident. (Magnification for [A] and [B] 200×, for [C] 400×. All images provided courtesy of Dr. Ronald L. Hamilton, Department of Pathology, Division of Neuropathology, University of Pittsburgh School of Medicine.)

NONPHARMACOLOGIC MANAGEMENT. Nonpharmacologic interventions have not been systematically studied in DLB. Strategies to increase levels of arousal and attention through social interaction and environmental novelty may reduce the visual hallucinations and fluctuations in cognition and function. There is preliminary evidence for the role of exercise, cognitive training, and caregiver education to help with the management of behavioral disturbances, particularly agitation and psychosis.

PHARMACOLOGIC MANAGEMENT

Cholinesterase Inhibitors. The cholinesterase inhibitors are the mainstay of treatment for DLB. The strongest evidence exists for donepezil and rivastigmine, with a 2015 meta-analysis showing these medications to have beneficial effects on cognition, hallucinations, delusions, and activities of daily living in individuals with DLB. However, adverse events were more common, particularly with rivastigmine, and included nausea, vomiting, anorexia, and somnolence. The rivastigmine patch may mitigate some of the gastrointestinal side effects associated with the oral formulation. Data for the usage of galantamine is limited to an open-label trial.

Memantine. Data in regard to the usage of memantine remains mixed. A handful of studies have shown that memantine is associated with improved quality of life and improvements in global clinical status, goal attainment, caregiver burden, attention, and REM sleep behavior.

Antipsychotics. First-generation antipsychotics are generally not well tolerated in DLB and should be avoided. There is little data to guide the use of second-generation antipsychotics, for which DLB patients may also have a low tolerance. A placebo-controlled study of olanzapine in psychosis in Alzheimer disease looked at a subgroup of patients retrospectively diagnosed with DLB. These patients demonstrated a decrease in positive symptoms of psychosis without worsening of parkinsonism or cognition with olanzapine. However, concern remains about the ability to tolerate antipsychotics in individuals with DLB. A small open-label study of quetiapine for psychosis and agitation in DLB found a clinically significant reduction in symptoms in about half the patients, but several had to discontinue treatment due to somnolence and orthostatic hypotension. In a multicenter randomized placebo-controlled study of 23 individuals with DLB, quetiapine was well tolerated and did not worsen parkinsonism, but there were no significant differences in efficacy between quetiapine and placebo. Risperidone did not appear to be well tolerated in a randomized controlled trial comparing this to citalopram, with individuals with DLB in both treatment groups having high discontinuation rates. Although there are no controlled trials involving the use of clozapine in the treatment of neuropsychiatric symptoms of DLB, given this medication’s utility in treating psychosis in Parkinson disease, there may be some clinical value in its treatment of psychotic symptoms in DLB as well. More data is needed regarding the use of pimavanserin, a recently approved atypical antipsychotic for the treatment of Parkinson disease psychosis, for the treatment of psychotic symptoms in DLB.

Antidepressants. As discussed above, citalopram does not appear to be beneficial or well-tolerated in individuals with DLB. There are no other controlled trials of SSRIs or selective and norepinephrine reuptake inhibitors (SNRIs). Antidepressants with anticholinergic properties should be avoided.

Benzodiazepines. Clonazepam may have a benefit for the treatment of REM sleep behavior disorder, but like all long half-life, benzodiazepines must be used cautiously in the older adults.

Melatonin and Ramelteon. There is increasing evidence for the potential use of melatonin in REM sleep behavior disorder, but prospective clinical trials are still necessary. There are no studies on the use of melatonin in DLB, but a case series of four patients with DLB utilizing ramelteon, a melatonin agonist, showed a reduction in hallucinations and REM sleep behavior disorder episodes.

Zonisamide. A phase 2 placebo-controlled randomized trial examining the use of zonisamide as an adjunct to levodopa showed improvement of parkinsonism in patients with DLB without worsening of cognitive functioning or psychiatric symptoms. There was an increased risk of adverse effects with increasing dosage, mainly being weight loss and decreased appetite.

Antiparkinsonian Medications. Antiparkinsonian medications, including L-dopa and direct dopamine agonists, may be tried in DLB. Responses are variable but generally less robust than in Parkinson disease. About a third of patients may experience improvement. However, these patients are also susceptible to psychotic symptoms. Given their increased potential for neuropsychiatric complications, initial doses should be small, with gradual increases while monitoring behavioral symptoms. In addition to worsening psychosis, additional potential side effects include postural hypotension, sedation, nausea, and vomiting. Anticholinergic medications are generally not recommended because of the associated cognitive side effects.

Dementia in Parkinson Disease

Cognitive impairment in Parkinson disease is common, with almost a quarter of newly diagnosed individuals displaying some impaired performance.

Definition. In 2007, the Movement Disorder Society task force published criteria to guide the diagnosis of dementia associated with Parkinson disease. For an individual to be diagnosed with probable Parkinson disease dementia, there must be a dementia syndrome that developed within the context of Parkinson disease. Similar to other dementia syndromes described in this chapter, there must be impairment in more than one cognitive domain, this must be a decline from a previous level of functioning, and there are impairments in daily life independent of those impairments due to the motor or autonomic symptoms. Cognition is impaired in two out of the following four domains (attention, executive function, visuospatial function, and free memory recall), and there must also be the presence of at least one behavioral symptom.

History. Parkinson disease was first written about in 1817 by James Parkinson. In a treatise on the "shaking palsy," he described six cases he had observed. In 1888, the French neurologist, Jean-Martin Charcot, named this clinical entity after Parkinson.

Comparative Nosology. In the DSM-5, for a diagnosis of major neurocognitive disorder due to Parkinson disease, the criteria for major neurocognitive disorder must be met, the impairment occurs in the setting of well-established Parkinson disease, there is an insidious onset and gradual progression, and the impairment cannot be attributable to another medical condition or mental disorder.

Epidemiology. The prevalence of Parkinson disease increases with age. About a third of newly diagnosed patients with Parkinson disease may demonstrate some MCI. Up to 80% of individuals with Parkinson disease will develop dementia, with roughly 10% of individuals with Parkinson disease developing dementia each year. On average, Parkinson disease dementia is diagnosed 8 to 10 years from the onset of Parkinson disease, but there is a wide range. The age of onset of Parkinson disease is between the sixth and ninth decades; increased duration of illness is a risk factor for the development of dementia.

Etiology. Risk factors for the development of dementia in Parkinson disease include age, more severe parkinsonism, and MCI at initial diagnosis. While APOE ε4 is a known risk factor of Alzheimer disease and has been associated with DLB, evidence is mixed as to whether this plays a role in the development of Parkinson disease dementia. The hereditary *SCNA* and *GBA* forms of Parkinson disease have been associated with dementia.

Diagnosis and Clinical Features. As this clinical entity occurs within Parkinson disease, the cognitive impairment should occur at least 1 year after symptoms of Parkinson disease were first noted. On neuropsychiatric examination, impairments in attention, executive function, visuospatial function, language, and memory may be seen. Impaired mental speed may be evident. On memory testing, recognition is better than free recall. Regarding language impairment, word-finding difficulties and impaired comprehension of complex sentences may be seen. Associated neuropsychiatric symptoms that may be present include apathy, depression, anxiety, hallucinations, delusions, and personality changes. Hallucinations have been reported in 45% to 65% of individuals with PDD, and delusions in 25% to 30%. Similar to DLB, individuals may have REM sleep behavior disorder and excessive daytime sleepiness may also be seen.

Biomarkers. MRI imaging may show thinning in the frontal and anterior cingulate regions. Functional MRI may demonstrate reduced synchronization of frontal regions, and PET imaging may demonstrate hypometabolism in parietal and temporal cortices, as well as in visual areas and the frontal lobes. Dopamine transporter imaging is abnormal in Parkinson disease; however, this abnormality does not differentiate Parkinson disease from Parkinson disease with dementia. Cardiac metaiodobenzylguanidine scintigraphy is reduced in Parkinson disease; of note, this test is normal in MSA and PSP. On EEG, those with dementia in Parkinson disease have slower baseline EEG activity than those without cognitive impairment.

Differential Diagnosis. The 1-year rule can be useful in distinguishing Parkinson disease dementia from DLB. The presence of parkinsonian symptoms and cognitive impairment should also cause one to think about other Parkinson plus syndromes, including CBD, PSP, and MSA. The tremor in CBD differs from that typically seen in Parkinson disease in that it is typically minimal at rest and amplified with action, and CBD is often associated with an asymmetric rigidity. The vertical gaze impairment as well the prominent axial rigidity in PSP can aid in distinguishing this from

Parkinson disease; however, the gaze impairment can be absent in up to 50% of individuals with PSP. The parkinsonism in both CBD and PSP is typically not responsive to dopaminergic medications. In MSA, individuals may display bradykinesia, rigidity, postural instability, and postural tremor; the cerebellar and autonomic dysfunction may be helpful in distinguishing this from Parkinson disease. Of note, in multisystem atrophy, up to 30% of individuals may show some response to dopaminergic medications at some point during the disease process.

Course and Prognosis. The course of dementia in Parkinson disease is progressive over time, although individuals may have some plateaus, as well as fluctuations from day to day. Hallucinations may occur, as can other psychiatric symptoms, particularly apathy. Dementia has been associated with increased functional impairment, nursing home placement, and decreased survival in Parkinson disease.

Neuropathology. Parkinson disease is associated with neuronal loss in the substantia nigra. Lewy bodies, made up of α-synuclein, are present in the cell bodies of the surviving neurons. α-Synuclein pathology is more extensive in Parkinson disease dementia compared to Parkinson disease without dementia, with more global cortical and limbic involvement. Individuals may also have β-amyloid plaque burden as well as the presence of tau neurofibrillary tangles.

Management. Evidence suggests that acetylcholinesterase inhibitors can improve cognition, attention, executive function, and global status and may have beneficial effects on overall activities of daily living; rivastigmine is an approved treatment for Parkinson disease dementia. Memantine has not been shown to be beneficial. Should psychotic symptoms be problematic, attempts to reduce or discontinue antiparkinsonian medications, balancing psychotic and motor symptoms are warranted, with consideration of reduction in L-dopa as a last resort. Pimavanserin, which is a novel 5-HT_{2A} inverse antagonist, is an approved medication for the treatment of delusions and hallucinations in Parkinson disease psychosis. Antipsychotic medication should be used very cautiously. Clozapine is the only antipsychotic medication that has been deemed clinically useful for the treatment of psychosis in Parkinson disease by the Movement Disorder Society Task Force on Evidence-Based Medicine

Review of Treatments for Parkinson disease. There is insufficient evidence for the use of quetiapine, and olanzapine is not felt to be useful given its significant worsening of motor symptoms.

Dementia in Huntington Disease

Definition. Huntington disease is a rare disorder that is characterized by motor, cognitive, and psychiatric symptoms. The motor symptoms include chorea, or involuntary dance-like movements, and ataxia; the disease inevitably leads to dementia.

History. The clinical entity of Huntington disease was first described by George Huntington in late 1872. In 1993, the mutant gene was identified.

Comparative Nosology. For an individual to be diagnosed with the DSM-5 diagnosis of major neurocognitive disorder due to Huntington disease, there must be clinically established Huntington disease, or risk of this based on family history or genetic testing, in addition to meeting the criteria for major neurocognitive disorder, and there is a gradual and insidious onset.

Epidemiology. The worldwide prevalence of Huntington disease has been reported to be 2.71 per 100,000, with a higher prevalence seen in Europe, North America, and Australia.

Etiology. Huntington disease is an autosomal-dominant inherited disease transmitted through a triplet CAG repeat mutation located on chromosome 4 in the *HTT* gene. This mutation results in a polyglutamate expansion in the Huntington protein. Expression of the abnormal protein ultimately results in neuronal cell death, primarily in the caudate nucleus and frontal lobes. Family history is important in the diagnosis of this disease, and it can be confirmed by genetic testing for the CAG repeat expansion.

Diagnosis and Clinical Features. In addition to the chorea and ataxia, other motor symptoms include bradykinesia, dystonia, and ideomotor apraxia. Both cognitive and psychiatric symptoms can precede motor symptoms. Psychiatric symptoms include personality changes, depression, blunting of affect, irritability, psychosis, and aggressive–

compulsive behavior, with depression being the most common psychiatric symptom. The dementia is often described as a subcortical dementia, with deficits in processing speed, attention, planning, and sequencing being more common. Memory deficits can be in both declarative and procedural domains. Dysarthria and dysphagia are present in more severe stages.

Biomarkers. Atrophy of the caudate is evident on structural imaging, and striatal atrophy may be seen on imaging before a clinical diagnosis of Huntington disease is made.

Differential Diagnosis. Given the autosomal dominant pattern, obtaining a family history is important. If family history is unknown or if genetic testing is negative, CJD (discussed below) should be considered in the differential, given the ataxia and myoclonus that may be present. The differential for chorea includes both other hereditary and acquired causes.

Course and Prognosis. The age of onset of Huntington disease is typically between 20 and 65 years of age, with the disease itself lasting for about 20 years. Huntington disease is inevitably fatal.

Management. Antidepressants and sedatives can be used for associated mood disorders. Tetrabenazine is an FDA-approved treatment for chorea, but it is recommended that this medication be avoided in patients with uncontrolled depression and a history of suicidality, making this a challenge as psychiatric symptoms are particularly prevalent in Huntington disease. Deutetrabenazine is also FDA-approved for the treatment of chorea; this medication allows for less frequent daily dosing, and data suggests a more favorable adverse effect profile. Dopamine-depleting agents may be used for the treatment of chorea, but there is weaker evidence for the use of amantadine and riluzole compared to tetrabenazine. Antipsychotic medications may be used for the treatment of chorea, in addition to the treatment of agitation and psychosis. The more potent antipsychotic medications may also be beneficial for the chorea, but patients should be evaluated for potentially worsening parkinsonism as well if these are used. Cholinesterase inhibitors have not been shown to have benefits in the treatment of cognitive impairment.

Prion Diseases

Definition. The prion disorders encompass genetically induced, sporadically occurring, and acquired diseases, all of which can be attributed to a misfolded prion protein, which ultimately leads to a further cascade of misfolded prion proteins. The genetic prion diseases include familial CJD, fatal familial insomnia and Gersttmann–Straussler–Scheinker syndrome (GSS). The sporadically occurring prion diseases include sporadic CJD, sporadic fatal insomnia, and variably protease-sensitive prionopathy. The acquired prion illness include iatrogenic CJD, variant CJD, and kuru.

History. CJD was first termed in 1922 and was used to describe a clinical entity of rapidly progressive myoclonus, ataxia, and dementia, described in earlier case reports. In the 1950s, kuru was discovered in Papua New Guinea and was characterized by truncal ataxia, tremor, and quick progression to dementia and death. Kuru was eventually found to share similar histopathology and transmission properties with CJD. Later it was determined that these infectious agents were proteins, and the term prion was created in 1982 to characterize these proteinaceous infectious particles. In the late 1990s, there was a series of variant CJD cases in the United Kingdom and France, which resulted from the transmission to humans of bovine spongiform encephalopathy, commonly known as mad cow disease, through consumption of contaminated meat products.

Comparative Nosology. The DSM criteria for major neurocognitive disorder due to prion disease include an insidious onset with rapid progression of impairment, and motor features of prion disease or biomarker evidence must be present, and the presentation cannot be attributable to another medical or mental disorder.

Epidemiology. The incidence of sporadic CJD worldwide is 1 per 10^6 and familial CJD is 1 per 10^7 per year. In sporadic CJD, the peak age of onset is 67 years, while the familial and variant form typically occurs in younger individuals. About 15% of prion diseases are genetic. Fatal insomnia usually presents between the ages of 35 and 55.

Etiology. The prion diseases can be sporadic, inherited, and also horizontally transmissible, with about 15% of cases being caused by specific mutations of the prion protein gene. The gene for the prion protein is found on chromosome 20. Underlying these diseases is a misfolding of the prion protein, which occurs via a spontaneous or genetically induced mutation. Once this mutation occurs, the pathologic form of the prion propagates. CJD and fatal insomnia are both sporadic and familial. GSS is a familial disease. Iatrogenic CJD has been associated with medical or surgical equipment contaminated with the pathologic prion protein. Transmission through ingestion has also been demonstrated, as seen with Kuru and the practice of ritual endocannibalism in the Foré linguistic group in Papua New Guinea, and also variant CJD through the ingestion of contaminated meat products.

Diagnosis and Clinical Features. Sporadic and familial CJD initially present with memory impairment and confusion, which quickly progress to a cortical dementia-like picture with ataxia; other symptoms may include myoclonus, pyramidal/extrapyramidal dysfunction, visual dysfunction, cerebellar dysfunction, and akinetic mutism. Iatrogenic forms of CJD typically show more cerebellar ataxia at symptom onset. GSS first presents with ataxia, and then dementia develops. Variant CJD often presents with psychiatric symptoms, usually apathy or depression, and painful sensory symptoms in distal extremities; the illness progresses to eventually include symptoms of ataxia, abnormal movements, and cognitive decline. Often GSS presents initially with ataxia or dysarthria and then progresses to dementia; other symptoms that have been reported include memory impairment, spastic paresis, motor impairment, and behavioral disturbances; both age of onset and duration of illness are more variable in comparison to the other prion diseases. In fatal insomnia, there is insomnia, which may be present for months until other symptoms present themselves; additional symptoms may include visual changes, ataxia, myoclonus, and psychiatric symptoms, such as visual hallucinations, personality change, depression, and disinhibition; cognitive impairment and dementia do not appear until much later in the course. Both GSS and Kuru tend to have less cognitive impairment. Kuru typically presents with a rapid course, with symptoms including progressive ataxia, dysarthria, dysphagia, tremor, and motor dysfunction. Both familial and sporadic fatal insomnia presents with sleep abnormalities, neuropsychiatric symptoms, and gait abnormalities. Variably protease-sensitive prionopathy tends to have a longer disease duration, with presentations being more variable depending on the genetic polymorphisms involved.

Biomarkers. MRI, EEG, and CSF studies may be helpful in making the diagnosis of a prion disease. On MRI, there are high signal abnormalities in the caudate nucleus and/or putamen on diffusion-weighted imaging (DWI) and fluid-attenuated inversion recovery (FLAIR) in CJD. In sporadic CJD, DWI shows restricted diffusion in the cortex, caudate, and/or putamen, as well as abnormal FLAIR hyperintensities in the cortical ribbon. DWI in variant CJD characteristically shows the pulvinar sign. Hypometabolism in the thalamus and cingulate region on PET scan points toward fatal insomnia.

Elevations in 14-3-3, total tau, or neuron-specific enolase in CSF studies have been associated with sporadic forms of CJD. Real-time quake-inducing conversion (RT-QuIC) amplification of CSF prion protein may be helpful in the diagnosis of sporadic CJD.

EEG may be useful in the diagnosis of prion diseases. In sporadic and familial CJD, EEG shows bilateral periodic discharges, with EEG in CJD showing 1 to 2 Hz periodic sharp-wave (often biphasic or triphasic) complexes later in the disease. Iatrogenic CJD has a diffuse slow-wave pattern. Slow waves can also be evident in GSS and fatal familial insomnia.

As of yet, there are no prodromal biomarkers available for the identification of potential disease.

Differential Diagnosis. Given the progressive cognitive impairment and motor features of the prion diseases, the differential is quite broad and includes many of the neurodegenerative conditions that have been discussed in the chapter, including Alzheimer disease, the varying frontotemporal lobar dementias, DLB, CBD, and Huntington disease. Autoimmune and antibody-mediated disorders should be considered on the differential, as well as encephalitis, infectious, and neoplastic causes. The spinocerebellar ataxias, as well as Hashimoto encephalopathy, can also have similar symptom presentations.

Course and Prognosis. Sporadic CJD is most commonly diagnosed in the seventh or eighth decade. The course is rapidly progressive, with the prognosis somewhere between 4 and 6 months. Diagnosis is typically not made until the third month after symptom onset when symptoms are usually quite severe already. The familial form of the disease has a more extended course, typically of 1 to 5 years. The duration of GSS is typically 2 to 6 years, and for fatal insomnia, it is 1 to 2 years.

Neuropathology. The hallmark clinical feature of sporadic and familial CJD is spongiform degeneration of the cortical gray matter; reactive gliosis is often evident as well. Pathology is usually focused on the neocortex, subiculum of the hippocampus, caudate, putamen, and thalamus. In GSS, there are extensive prion protein plaques with less spongiform degeneration. There is focal thalamic gliosis and minimal spongiform degeneration seen in fatal insomnia. The neuropathology seen in variant CJD includes prion protein plaques surrounded by a halo of vacuolar change with diffuse spongiosis. Neuronal loss and gliosis are seen focally in the thalamus and the inferior olives in fatal insomnia. Spongiform changes and gliosis is seen in variably protease-sensitive prionopathy.

Management. At this time, treatment is limited to symptomatic therapies, including antiepileptic medications for seizure control, clonazepam for myoclonus, and small doses of antipsychotic medications for distressing psychiatric symptoms, including hallucinations and delusions.

Limbic Predominant Age-Related TDP-43 Encephalopathy

Definition. LATE, referred to as LATE, is a TDP-43 proteinopathy with or without hippocampal sclerosis pathology. This is a relatively new diagnosis, with consensus-based recommendations on diagnosis and staging released in 2019.

Epidemiology. LATE typically occurs in patients over the age of 80, who are often characterized as the oldest-old. Rates may be higher in females simply because they have a longer life expectancy compared to males. LATE neuropathologic changes are evident in over 20% of individuals over the age of 80, with this prevalence increasing with increasing age; however, these may not be the sole neuropathologic changes present.

Genetics. TDP-43 is encoded by the *TARDBP* gene. Gene variants on *GRN* on chromosome 17q, *TMEM106B* on chromosome 7p, *ABCC9* on chromosome 3q, and *APOE* on chromosome 19q are all suspected of playing a role in LATE neuropathologic changes. *KCNMB2* is also suspected of playing a role.

Diagnosis and Clinical Features. LATE presents as a progressive amnestic syndrome, similar to dementia of the Alzheimer type. Individuals with only LATE neuropathologic changes have a more gradual decline, as opposed to those with LATE neuropathologic changes in combination with Alzheimer disease neuropathologic changes. The course typically begins with impairment in episodic memory, then expands to include other cognitive domains, and ultimately affects an individual's independent functional status.

Biomarkers. There are no CSF, serum, or neuroimaging biomarkers yet available. Neuroanatomical imaging may show atrophy in the mesial temporal lobe, inferior frontal, anterior temporal lobes, and the insular cortices. On MRI, hippocampal atrophy is greater in LATE than Alzheimer disease, particularly if hippocampal sclerosis is present.

Differential Diagnosis. Given the similar symptom profile, dementia of the Alzheimer type should be considered in the differential. One autopsy study of 25 individuals with LATE neuropathologic change and no pathologic changes associated with Alzheimer disease or Lewy body disease showed that 67% had received a diagnosis of dementia of the Alzheimer type, and 12.5% had received a diagnosis of vascular dementia. LATE shares some genetic similarities to FTD, and TDP-43 proteinopathy is seen in both FTD-TDP as well as amyotrophic lateral sclerosis. The older age of presentation in LATE should help to distinguish this from FTD-TDP, and the clinical symptoms should distinguish this from ALS.

Course and Prognosis. There are varied reports as to whether those with both LATE neuropathologic change and Alzheimer disease neuropathologic change together have greater cognitive impairment and a steeper decline compared to those with "pure" LATE. The presence of both Alzheimer disease and LATE neuropathologic changes together have been associated with lower total neuropsychiatric symptom burden compared to dementia of the Alzheimer type alone. Those individuals over the age of 90 appear to have a slower decline compared to those individuals 89 years old or less.

Neuropathology. Similar to most dementias, an autopsy provides a definitive diagnosis. 97% of patients with LATE neuropathologic change on autopsy initially presented with cognitive symptoms.

The presence of TDP-43 proteinopathy is necessary to make a diagnosis of LATE-neuropathologic changes. TDP-43 protein inclusion bodies can be found in the neuronal cytoplasm, and abnormal TDP-43 accumulation can be present in nuclei and neurites of neurons, as well as in oligodendroglia and astrocytes. The distribution of neuropathology starts in the amygdala, then progresses to the hippocampus, and then the middle frontal gyrus.

Hippocampal sclerosis may or may not be present and is not required to make a diagnosis of LATE, with one autopsy study showing that hippocampal sclerosis was evident in 42% of individuals who had LATE neuropathologic change.

FURTHER READINGS

Alzheimer’s Association. 2021 Alzheimer’s disease facts and figures. *Alzheimers Dement*. 2021;17(3):327–406.

Armstrong MS, Litvan I, Lang AE, et al. Criteria for the diagnosis of corticobasal degeneration. *Neurology*. 2013;80(5):496–503.

Baldwin KJ, Correll CM. Prion disease. *Semin Neurol*. 2019;39(4):428–439.

Coyle-Gilchrist ITS, Dick KM, Patterson K, et al. Prevalence, characteristics, and survival of frontotemporal lobar degeneration syndromes. *Neurology*. 2016;86(18):1736–1743.

Cummings J, Mintzer J, Brodaty H, et al. Agitation in cognitive disorders: International Psychogeriatric Association provisional consensus clinical and research definition. *Int Psychogeriatr*. 2015;27(1):7–17.

Davies SJC, Burhan AM, Kim D, et al. Sequential drug treatment algorithm for agitation and aggression in Alzheimer’s and mixed dementia. *J Psychopharmacol*. 2018;32(5):509–523.

Dubois B, Villain N, Frisoni GB, et al. Clinical diagnosis of Alzheimer’s disease: recommendations of the International Working Group. *Lancet Neurol*. 2021;20(6):484–496.

Ducharme S, Dols A, Laforce R, et al. Recommendations to distinguish behavioral variant frontotemporal dementia from psychiatric disorders. *Brain*. 2020;143(6):1632–1650.

Eisenmann Y, Golla H, Schmidt H, et al. Palliative care in advanced dementia. *Front Psychiatry*. 2020;11:699.

Emre M, Aarsland D, Brown R, et al. Clinical diagnostic criteria for dementia associated with Parkinson’s disease. *Mov Disord*. 2007;22(12):1689–1707.

Fleisher AS, Pontecorvo MJ, Devous MD Sr, et al. Positron emission tomography imaging with [18F] flortaucipir and postmortem assessment of Alzheimer disease neuropathologic changes. *JAMA Neurology*. 2020;77(7):829–839.

GBD 2016 Dementia Collaborators. Global, regional, and national burden of Alzheimer’s disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol*. 2019;18(1):88–106.

Höglinger GU, Respondek G, Stamelou M, et al. Clinical diagnosis of progressive supranuclear palsy: the movement disorder society criteria. *Mov Disord*. 2017;32(6):853–864.

Hsieh S, McGrory S, Leslie F, et al. The Mini-Addenbrooke's Cognitive Examination: a new assessment tool for dementia. *Dement Geriatr Cogn Disord*. 2015;39(1-2):1–11.

Jack CR Jr, Bennett DA, Blennow K, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14(4):535–562.

Kales HC, Gitlin LN, Lyketsos CG; Detroit Expert Panel on the Assessment and Management of Neuropsychiatric Symptoms of Dementia. Management of neuropsychiatric symptoms of dementia in clinical settings: recommendations from a multidisciplinary expert panel. *J Am Geriatr Soc*. 2014;62(4):762–769.

Karageorgiou E, Miller BL. Frontotemporal lobar degeneration: a clinical approach. *Semin Neurol*. 2014;34(2):189–201.

Leonpacher AK, Peters ME, Drye LT, et al. Effects of citalopram on neuropsychiatric symptoms in Alzheimer's Dementia: evidence from the CitAD study. *Am J Psychiatry*. 2016;173(5):473–480.

Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the *Lancet* Commission. *Lancet*. 2020;396 (10248):413–446.

McKeith IG, Boeve BF, Dickson D, et al. Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB Consortium. *Neurology*. 2017;89(1):88–100.

Mitchell SL, Miller SC, Teno JM, et al. Prediction of 6-month survival of nursing home residents with advanced dementia using ADEPT vs hospice eligibility guidelines. *JAMA*. 2010;304(17):1929–1935.

Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain*. 2019;142(6):1503–1527.

Rabins PV, Rovner BW, Rummans T, Schneider LS, Tariot PN. Guideline watch (October 2014): practice guideline for the treatment of patients with Alzheimer's disease and other dementias. *Focus (Am Psychiatr Publ)*. 2017;15(1):110–128.

Reus VI, Fochtmann LJ, Eyler AE, et al. *The American Psychiatric Association Practice Guideline on the Use of Antipsychotics to Treat Agitation or Psychosis in Patients with Dementia*. American Psychiatric Association; 2016. Accessed March 29, 2021. https://psychiatryonline-org.ezaccess.libraries.psu.edu

Steinberg M, Shao H, Zandi P, et al. Point and 5-year period prevalence of neuropsychiatric symptoms in dementia: the Cache County Study. *Int J Geriatr Psychiatry*. 2008;23(2):170–177.

Stinton C, McKeith I, Taylor J, et al. Pharmacological management of Lewy Body Dementia: a systematic review and meta-analysis. *Am J Psychiatry*. 2015;172:731–742.

Younes K, Miller BL. Frontotemporal dementia: neuropathology, genetics, neuroimaging, and treatments. *Psychiatry Clin N Am*. 2020;43(2):331–344.

Young JJ, Lavakumar M, Tampi D, et al. Frontotemporal dementia: latest evidence and clinical implications. *Ther Adv Psychopharmacol*. 2018;8(1):33–48.

Watt JA, Goodarzi Z, Veroniki AA, et al. Comparative efficacy of interventions for reducing symptoms of depression in people with dementia: systematic review and network meta-analysis. *BMJ*. 2021;372:n532.

Wyant KJ, Ridder AJ, Dayalu P. Huntington's disease- update on treatments. *Curr Neurol Neurosci Rep*. 2017;17(4):33.

10.4 Mild Cognitive Impairment and Amnestic Disorders

Daniel Varon, M.D., and Carmen Andreescu, M.D.

In the past, amnestic disorders referred specifically to “pure” memory deficits, most often defined in the context of substance abuse/dependence or in the context of a general medical condition. However, 20 years ago, a new concept was developed: *mild cognitive impairment* (MCI). As defined by the Report of the International Working Group on MCI, it requires the presence of cognitive decline not warranting the diagnosis of dementia and/or evidence of decline over time on objective cognitive tasks but with preserved basic activities of daily living. Thus, MCI differs from classic amnestic disorders both by its etiology and by the domains of its clinical presentation.

Table 10.4–1.
DSM-V Diagnostic Criteria for Mild Neurocognitive Disorder

A. Evidence of modest cognitive decline from a previous level of performance in one or more cognitive domains (complex attention, executive function, learning and memory, language, perceptual-motor, or social cognition) based on:
  1. Concern of the individual, a knowledgeable informant, or the clinician that there has been a mild decline in cognitive function; and
  2. A modest impairment in cognitive performance, preferably documented by standardized neuropsychological testing or, in its absence, another quantified clinical assessment.

B. The cognitive deficits do not interfere with the capacity for independence in everyday activities (i.e., complex instrumental activities of daily living such as paying bills or managing medication are preserved, but greater effort, compensatory strategies, or accommodation may be required).

C. The cognitive deficits do not occur exclusively in the context of a delirium.

D. The cognitive deficits are not better explained by another mental disorder (e.g., major depressive disorder, schizophrenia).

Specify whether due to: Alzheimer disease, frontotemporal lobar degeneration, Lewy body disease, vascular disease, traumatic brain injury, substance/medication use, HIV infection, prion disease, Parkinson disease, Huntington disease, another medical condition, multiple etiologies, or unspecified.

Coding note: For mild neurocognitive disorder due to any of the medical etiologies listed above, code **331.83 (G31.84).** Do *not* use additional codes for the presumed etiologic medical conditions. For substance/medication-induced mild neurocognitive disorder, code based on the type of substance; see “Substance/Medication-Induced Major or Mild Neurocognitive Disorder.” For unspecified mild neurocognitive disorder, code **799.59 (R41.9).**

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:680, 681.

In the revised fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR), MCI was hosted by various provisions (cognitive disorder not otherwise specified [NOS], amnestic disorder NOS). In DSM V, the term mild neurocognitive disorder embraces cognitive disorders NOS and MCI (Table 10.4–1). According to DSM V, amnestic disorder can be classified under Major Neurocognitive Disorders if the severity of the dysfunction is enough to grant this diagnosis.

MILD COGNITIVE IMPAIRMENT

Definition

Although the term MCI has been in use since 1982 as a descriptor of stage 3 on the Global Deterioration Scale, it was later developed by the Mayo Clinic Group as a diagnostic category designed to fill the gap between cognitive changes associated with aging and cognitive impairment suggestive of dementia. The criteria proposed by the Mayo Clinic Alzheimer's Disease Research Center (MCADRC criteria) include (1) memory complaint, preferably confirmed by an informant; (2) objective memory impairment for age and education; (3) preserved general cognitive function; (4) intact activities of daily living; and (5) not demented (Table 10.4–2).

Table 10.4–2.
Mild Cognitive Impairment Original Criteria

1. Memory complaint, preferably qualified by an informant
2. Memory impairment for age and education
3. Preserved general cognitive function
4. Intact activities of daily living
5. Not demented

In 2003, the concept was refined (Table 10.4–3) and expanded to include amnestic and nonamnestic subtypes, which in turn can be subdivided into single- and multiple-domain subcategories. (Fig. 10.4–1). These divisions can be used when trying to approximate the etiology of the MCI syndrome. In 2011, a workgroup assembled by the National Institute on Aging (NIA) and the Alzheimer's Association (AA) delineated specific criteria for the diagnosis of MCI due to Alzheimer disease (AD) (Table 10.4–4), and subsequently, criteria for MCI due to Parkinson disease were formulated as well. Similarly, the DSM-V criteria for mild neurocognitive disorder include specifiers for multiple etiologies (Table 10.4–1). As the availability of biomarkers increases, criteria for each etiology will become progressively more specific.

Historical Perspective

The imprecise border between normal aging-related cognitive decline and dementia-related cognitive impairment has been described for several decades. Thus, in 1962, Kral introduced the terms "Benign Senescent Forgetfulness" (forgetfulness for less important facts and awareness of problems) and "Malignant Senescent Forgetfulness" (memory problems for recent events and lack of awareness). In 1986, the National Institute of Mental Health (NIMH) recommended the term "Age-Associated Memory Impairment" for age-related normal memory changes. However, the validity of this concept was disputed due to diagnostic tests based on the performances of younger adults (Table 10.4–5).

In 1994, the International Psychogeriatrics Association presented the concept of "Age-Associated Cognitive Decline," which described cognitive deficits, including but not limited to memory impairment in the absence of dementia or other affecting cognitive conditions.

Table 10.4–3.
Recommendations for the General Criteria for MCI

General Criteria for MCI
Not normal, not demented (does not meet DSM IV-TR/ICD-10 criteria for a dementia syndrome)
Cognitive decline
Self and/or informant report and impairment on objective cognitive tasks
And/or
Evidence of decline over time on objective cognitive tasks
Preserved basic activities of daily living/minimal impairment in complex instrumental functions

MCI, mild cognitive impairment; ICD-10, *International Classification of Diseases*, 10th edition.

Reprinted with permission from Winblad B, Palmer K, Kivipelto M, Jelic V, et al. Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med.* 2004;256(3):256–240. doi:10.1111/j.1365-2796.2004.01380.x.

“Cognitive Impairment No Dementia” was introduced in 1997 by the Canadian Study of Health and Aging to describe the presence of nondemented cognitive impairment regardless of the underlying process (neurologic, psychiatric, medical). It is thus a more inclusive concept (encompasses those with static encephalopathies who are not necessarily having a progressive worsening cognitive decline).

Several other classifications, including age-consistent memory impairment and late-life forgetfulness, are defined on the bases of performance on various cognitive tests.

The concept of MCI is currently defined in DSM-V as Mild Neurocognitive Disorder. Just as functional impairment is an exclusion criterion for MCI, in Mild Neurocognitive Disorder, cognitive deficits should not interfere with the capacity for independence in everyday activities. Developments in identifying biologic markers for MCI will likely solidify the conceptualization and, hopefully, the treatment of prodromal dementia.

Epidemiology and Etiology of MCI

Various studies on the prevalence of MCI are presented in Table 10.4–6. The recognition that AD pathology may exist in the brain long before the presence of clinical symptoms led to the focus on preclinical stages, with the purpose of characterizing initial impairments that are associated with an increased risk of progression to AD.

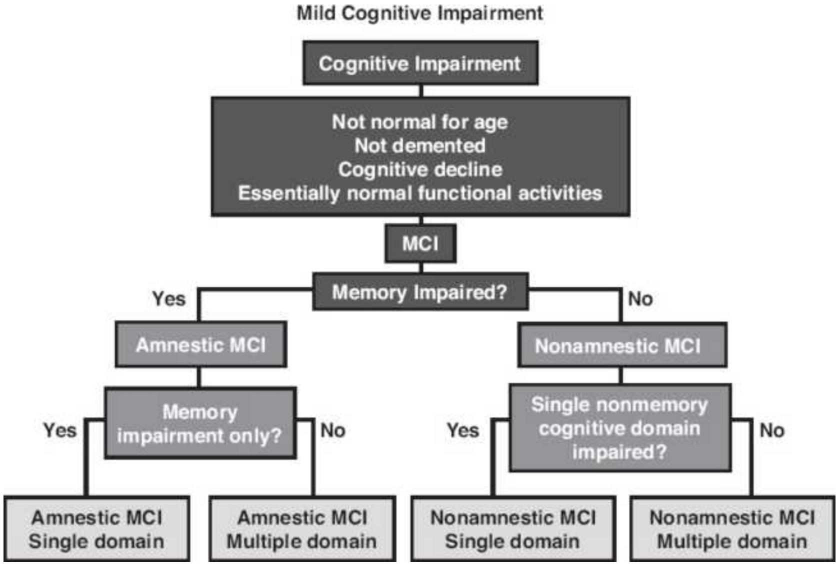
Mild Cognitive Impairment
Cognitive Impairment
Not normal for age
Not demented
Cognitive decline
Essentially normal functional activities
MCI
Memory Impaired?
Yes
No
Amnestic MCI
Nonamnestic MCI
Memory impairment only?
Yes
No
Single nonmemory cognitive domain impaired?
Yes
No
Amnestic MCI Single domain
Amnestic MCI Multiple domain
Nonamnestic MCI Single domain
Nonamnestic MCI Multiple domain

FIGURE 10.4–1. The current algorithm for diagnosing mild cognitive impairment (MCI) and its subtypes. (Reprinted with permission from Petersen: Mild cognitive impairment. *Continuum.* 2016;22(2):404–418. © 2016 American Academy of Neurology. doi: 10.1212/CON.0000000000000313; originally modified with permission from Petersen RC: Mild cognitive impairment. *Continuum*. 2004;10(1):9–28. © 2004 American Academy of Neurology. doi: 10.1212/01.CON.0000293545.39683.cc.)

Table 10.4–4.
Criteria for the Diagnosis of MCI NIA-AA Workgroup

A. **Core Clinical Criteria:**
   1. Concern regarding change in cognition
   2. Impairment in one or more neurocognitive domains
   3. Preservation of independence in functional abilities
   4. No evidence of dementia should be noted

B. **Research Criteria Incorporating Biomarkers:**

Diagnostic category…………………………biomarker probability of………………………Aβ……………
……………………………AD etiology…………………………… (PET or CSF)……………………(tau, FD
sMRI)…………………… MCI-core clinical criteria………………Uninformative………………
Conflicting………………………Conflicting………………………
Indeterminant………………Indeterminant……………………… Untested……………………Untested
intermediate
likelihood………………………intermediate………………………positive…………………………untest
Untested……………………………positive MCI due to AD-high
likelihood………………………Highest………………………Positive……………………………Positive
AD……………………………Lowest……………………………Negative………………………Negative

MCI, mild cognitive impairment; NIA, National Institute on Aging; AA, Alzheimer Association.

Reprinted with permission from Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dementia.* 2011;7(3):270–279. doi:10.1016/j.jalz.2011.03.008.

The clinical expression of MCI can be viewed as a result of the interaction between several risk factors and several protective factors. The most significant risk factors are related to the different types of neurodegeneration witnessed in dementias. These are clinically expressed in different subtypes of MCI (e.g., amnestic MCI [a-MCI], single domain, is considered to be related to underlying AD pathology; a-MCI, multiple domains, is considered to be related to either AD or vascular dementia type of pathology; while nonamnestic MCI [na-MCI] has been related more closely to dementia with Lewy bodies (DLB) or frontotemporal neurodegeneration). Other risk factors include the apolipoprotein E 4 (APOE4) allele status and cerebrovascular events in the form of either cerebrovascular accident or lacunar disease. The role of chronic exposure to high levels of cortisol, as seen in late-life depression, is also hypothesized to increase the risk for cognitive impairment through hippocampal volume reduction. The notion of "brain reserve" suggests that effects of brain size and neuron density may be protective against dementia despite the presence of neurodegeneration (a larger number of neurons and a bigger brain volume would protect against clinical manifestations of AD despite the presence of neurodegeneration) (Fig. 10.4–2).

MCI Clinical Presentation

The clinical picture of MCI is a function of the criteria used to define it. The MCADRC criteria emphasize memory impairment but do not specify how it should be evaluated. The second criterion (objective memory impairment for age and education) is often interpreted as an objective deficit of memory or another cognitive domain that is more than 1.5 standard deviations below the mean for individuals of similar age and education. The inherent heterogeneity is driven by the novelty of the concept at the time, and the lack of consensus between experts was recognized by the Current Concepts in MCI Conference, which classified MCI into different subgroups, each associated with an increased risk of developing a type of dementia. Thus, a-MCI usually progresses to AD, multiple domains of MCI (md-MCI) may progress to either AD or another type of dementia, while MCI-single nonamnestic domain develops predominantly into non-Alzheimer dementia-like frontotemporal dementia or DLB. With the introduction of the NIA-AA criteria for MCI due to AD, two sets of criteria were outlined: The Core Clinical Criteria meant to be used broadly, in any setting and without requiring any highly specialized tests or procedures and The Clinical Research Criteria, which incorporates the use of biomarkers, intended to be used only in research, academic centers, and clinical trials (Table 10.4–4). Biomarkers are not included in the DSM-V criteria for mild neurocognitive disorder due to AD. However, both the NIA-AA working group and DSM V task force acknowledge these biomarkers will most likely make their way into clinical practice once they are adequately validated and widely available.

The correlation between a clinical presentation of MCI and dementia outcome varies. The annual "conversion" rate was 12% in the Mayo MCI sample (compared with 1% to 2% for controls), while the numbers vary from 8% to 100% depending on the setting (community vs. clinical) and the study design (prospective vs. retrospective). Thus, in the Monongahela Valley Independent Elderly Survey (a community-based study), although 10% to 17% of MCI subjects progressed to dementia, 33% to 56% became non-MCI, and half of those returned to normal (Table 10.4–7).

Table 10.4–5.
Terms Related to Mild Cognitive Impairment

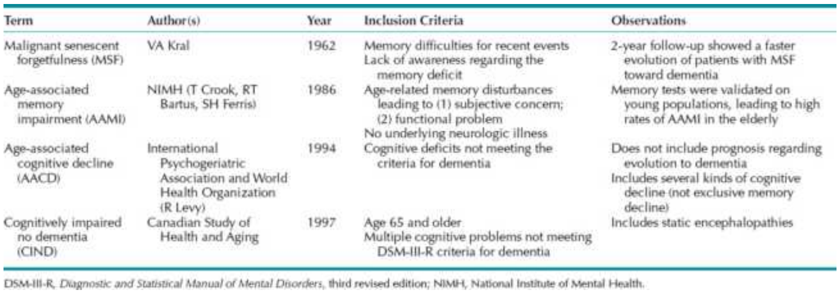

Term	Author(s)	Year	Inclusion Criteria	Observations
Malignant senescent forgetfulness (MSF)	VA Kral	1962	Memory difficulties for recent events Lack of awareness regarding the memory deficit	2-year follow-up showed a faster evolution of patients with MSF toward dementia
Age-associated memory impairment (AAMI)	NIMH (T Crook, RT Bartus, SH Ferris)	1986	Age-related memory disturbances leading to (1) subjective concern; (2) functional problem No underlying neurologic illness	Memory tests were validated on young populations, leading to high rates of AAMI in the elderly
Age-associated cognitive decline (AACD)	International Psychogeriatric Association and World Health Organization (R Levy)	1994	Cognitive deficits not meeting the criteria for dementia	Does not include prognosis regarding evolution to dementia Includes several kinds of cognitive decline (not exclusive memory decline)
Cognitively impaired no dementia (CIND)	Canadian Study of Health and Aging	1997	Age 65 and older Multiple cognitive problems not meeting DSM-III-R criteria for dementia	Includes static encephalopathies

DSM-III-R, *Diagnostic and Statistical Manual of Mental Disorders*, third revised edition; NIMH, National Institute of Mental Health.

Table 10.4–6.
Prevalence of MCI in Different Studies

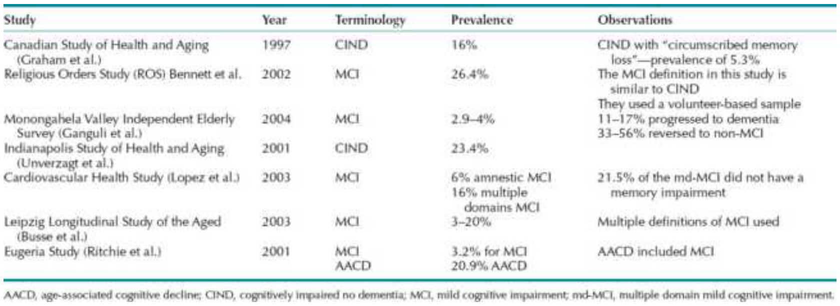

Study	Year	Terminology	Prevalence	Observations
Canadian Study of Health and Aging (Graham et al.)	1997	CIND	16%	CIND with "circumscribed memory loss"—prevalence of 5.3%
Religious Orders Study (ROS) Bennett et al.	2002	MCI	26.4%	The MCI definition in this study is similar to CIND They used a volunteer-based sample
Monongahela Valley Independent Elderly Survey (Ganguli et al.)	2004	MCI	2.9–4%	11–17% progressed to dementia 33–56% reversed to non-MCI
Indianapolis Study of Health and Aging (Unverzagt et al.)	2001	CIND	23.4%	
Cardiovascular Health Study (Lopez et al.)	2003	MCI	6% amnestic MCI 16% multiple domains MCI	21.5% of the md-MCI did not have a memory impairment
Leipzig Longitudinal Study of the Aged (Busse et al.)	2003	MCI	3–20%	Multiple definitions of MCI used
Eugeria Study (Ritchie et al.)	2001	MCI AACD	3.2% for MCI 20.9% AACD	AACD included MCI

AACD, age-associated cognitive decline; CIND, cognitively impaired no dementia; MCI, mild cognitive impairment; md-MCI, multiple domain mild cognitive impairment.

Neuropsychiatric symptoms can appear along with or prior to the cognitive deficits seen in MCI. The concept of mild behavioral impairment (MBI) has been described in the past and has gained more relevance in recent years. This construct describes the emergence of neuropsychiatric symptoms, which can be precursors to cognitive decline and eventually dementia. The description and detection of MBI have been operationalized by the International Society to Advance Alzheimer's Research and Treatment—Alzheimer's Association (ISTAART-AA) criteria. Five subcategories have been proposed, including motivation, affect, impulse control, social appropriateness, and perception/thought content. Rating scales to capture these criteria have been proposed, such as the mild behavioral impairment checklist (MBI-C).

MCI Assessment

Neuropsychological Assessment. Most experts agree that earlier deficits are noted in episodic (vs. semantic) memory. There is no consensus among experts with regard to which memory tests and which cutoffs to use. There is a lack of norms, test scores do not have normal distributions, and test performance is influenced by multiple demographic characteristics. Several experts have proposed that a scale like the delayed recall task from the consortium to establish a registry for AD might be useful in detecting AD in the earliest stages. Brief mental status instruments (like mini-mental state examination) are relatively insensitive for the detection of memory problems in MCI. The Montreal Cognitive

Assessment (MoCA) has been used as a more adequate way to screen patients with MCI. It has been validated multiple times in different pathologies, populations, and languages. It includes more adequate testing for frontal lobe dysfunction when compared to the Mini-Mental State Examination (MMSE).

The NIA-AA workgroup suggests tests of episodic memory such as the Free and Cued Selective Reminding Test, the Rey Auditory Verbal Learning Test, the California Verbal Learning Test, or the Logical Memory I and II of the Wechsler Memory Scale-Revised can be used. Tests to assess other cognitive domains are also mentioned, including the Trail Making Test for executive function, the Boston Naming Test, letter and category fluency for language, figure copying for spatial skills, and digit span forward for attention.

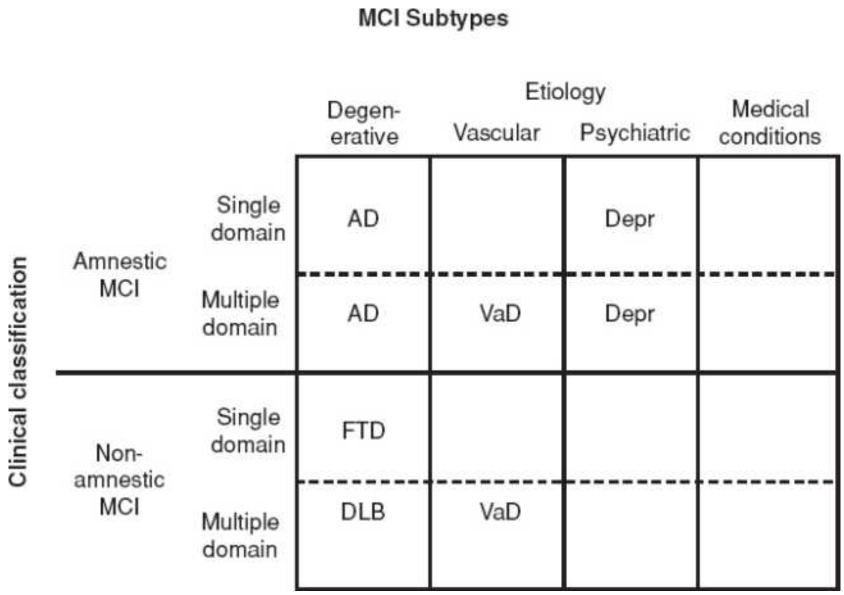

MCI Subtypes

Clinical classification		Etiology			
		Degenerative	Vascular	Psychiatric	Medical conditions
Amnestic MCI	Single domain	AD		Depr	
	Multiple domain	AD	VaD	Depr	
Non-amnestic MCI	Single domain	FTD			
	Multiple domain	DLB	VaD		

FIGURE 10.4–2. Outcome of clinical phenotypes of mild cognitive impairment (MCI) according to presumed etiology. AD, Alzheimer disease; Depr, depression; VaD, vascular dementia; FRD, frontotemporal dementia; DLB, dementia with Lewy Bodies. (Adapted from Petersen RC. Aging, mild cognitive impairment, and Alzheimer's disease. *Neurol Clin*. 2000;18(4):789–805.)

Table 10.4–7.
Conversion Rates from MCI to Alzheimer Disease in Different Studies

Study	Conversion Rate	Observations
Petersen et al. (1999)	12% per year	Clinical sample *N* = 76
Ganguli et al. (2004)	10–17% every 2 years	Community sample
Monongahela Valley Independent Elderly Survey	27% at 10 years	10-year follow-up Fourfold increased risk for progression to Alzheimer disease in MCI group compared with controls
Boston study-Fisk et al. (2005)	18.7% over 3 years	Community volunteers
St. Louis study-Morris et al. (2001)	100% over 9.5 years	Clinical sample
Eugeria study-Ritchie et al. (2001)	11.3% over 3 years	Primary care-based clinical sample
Larrieu et al. (2002)	8.3% per year	Community-based France
QUID Study (Personnes Agees)		40% reverted to normal after 2 years

MCI, mild cognitive impairment.

Biomarkers. Several markers of progression from MCI to AD have been studied in the past two decades. The most well-established biomarkers have been divided into two categories, those that measure amyloid deposition and those that relate to neurodegeneration. Low levels of $A\beta_{42}$ (the 42 amino acid form of β-amyloid) and increased uptake of amyloid positron emission tomography (PET) tracers are associated with fibrillar Aβ deposits. Increased levels of CSF total tau (t-tau) and phosphorylated tau (p-tau), hypometabolism on fluorodeoxyglucose (FDG) PET, and atrophy on MRI are measures of neurodegeneration. The ratio of CSF Aβ levels to t-tau and phosphotau-231 (ptau-231) appears to be the best biomarker in CSF. Locating alterations in the expression of proteins involved in the pathogenetic pathways of AD (proteomic approach) is another approach used to help early detection of AD. Several proteins (cystatin C, β-2 microglobulin, and BEGF polypeptides) have been detected through specific techniques, and currently, there are a number of proteins from both CSF and blood that are implicated in AD pathology. New techniques to accurately identify markers such as p-tau181 and p-tau217 in blood have become available in recent years and will likely allow faster screening and diagnosis of patients at the early stages of the disease.

In 2018, the NIA-AA proposed a new system to define AD more adequately, with the purpose of shifting from a syndromal to a biologic construct. Biomarkers associated with AD were grouped into one of three categories, Aβ deposition, pathologic tau, or neurodegeneration [AT(N)]. Within this framework, living persons can be characterized as positive or negative for all three of these categories independent of their clinical stage. Patients with evidence of an amyloid biomarker (A+) combined with present or absent biomarkers for Tau (T+ or T−), and neurodegeneration (N+ or N−) would be classified within the Alzheimer continuum. Patients with no presence of amyloid (A−) but evidence of tau (T+), neurodegeneration (N+), or both would be classified as having non-AD pathologic change (Table 10.4–8). This framework was recommended for use in research only, with the goal of thoroughly examining its validity. Once this is accomplished, it could eventually make its way to general clinical practice.

Genetics. As MCI is regarded as the prodromal stage for several disorders (AD, frontotemporal, Lewy Body disease, or vascular dementia), different genes are probably related to MCI. Three genes involved in rare autosomal dominant forms of AD have been identified: the amyloid precursor protein (*APP*) gene, presenilin-1 (*PSEN1*), and presenilin-2 (*PSEN2*). Screening for each of these mutations will have very limited value for the diagnosis of MCI in the general population unless a clear family history is present. Multiple other genes have been associated with sporadic AD. Among those, APOE has the greatest effect on subjects with MCI. Longitudinal studies have demonstrated that MCI subjects who carry a copy of the E4 allele are at higher risk of progressing to dementia and demonstrate a greater degree of whole brain and hippocampal atrophy. However, testing for APOE genotype is not currently recommended since 75% of people with one copy of the E4 allele remain free of AD, and up to 50% of people with AD do not carry the E4 allele. As the etiology of MCI is heterogeneous, it is likely that a very large number of different genes underlie the pathology of MCI.

Neuroimaging

Advances in neuroimaging studies aim to develop measures allowing the differentiation between MCI and healthy aging as well as within MCI between subjects who will convert to AD or will remain stable over time. This objective has led to a worldwide effort that started in the United States with the Alzheimer's Disease Neuroimaging Initiative (ADNI) and expanded to multiple other countries, including Japan, China, Taiwan, Argentina, and Australia. These studies include a large number of subjects with subjective cognitive impairment, MCI, and mild dementia who are being followed over time. Subsequently, a similar project was developed for Parkinson disease, the Parkinson's Progression Markers Initiative (PPMI).

Table 10.4–8.
Biomarker Profiles and Categories

AT(N)	Biomarker Category	
A-**T-(N)**-	Normal AD biomarkers	
A+**T-(N)**-	Alzheimer Pathologic change	Alzheimer continuum
A+**T+(N)**-	AD	
A+T+**(N)**+	AD	
A+**T-(N)**+	Alzheimer and concomitant suspected non-Alzheimer pathologic change	
A-T+**(N)**-	Non-AD pathologic change	
A-T-(N)+	Non-AD pathologic change	
A-T+**(N)**+	Non-AD pathologic change	

AD, Alzheimer disease; A, aggregated Aβ—cerebrospinal fluid Aβ 42, or Aβ 42/Aβ 40 ratio, amyloid PET; T, aggregated tau—cerebrospinal fluid phosphorylated tau, Tau PET; (N), neurodegeneration or neuronal injury—anatomic MRI, FDG PET, cerebrospinal fluid total tau.

Reprinted with permission from Jack CR Jr, Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's Dementia.* 2018;14(4):535–562. doi:10.1016/j.jalz.2018.02.018.

Structural volumetric studies of MCI subjects showed early changes in the medial temporal structures, reflecting neuronal atrophy, decreased synaptic density, and overall neuronal loss seen in postmortem studies. Volume loss in the hippocampus and entorhinal cortex has been described in MCI. Atrophy of the hippocampal formation and entorhinal cortex measured using volumetric methods and visual rating scales have been reported to predict the rate of progression from MCI to AD. Three-dimensional modeling techniques have localized shape alteration and specific regions of atrophy within the hippocampus. Other methods like tensor-based morphometry allow tracking brain changes in detail, quantifying tissue growth or atrophy throughout the brain, and indicating the local rate at which tissue is being lost. Other innovations in neuroimaging include MR relaxometry, imaging of iron deposition, and diffusion tensor imaging which have been used to assess homeostasis of iron in the gray matter and disruption of myelin in the white matter, both of which correlate with incipient pathologic changes in subjects with AD.

Perhaps the most important development has been the advent of PET tracer compounds that visualize amyloid plaques and neurofibrillary tangles. Pittsburgh compound B (carbon-11-PIB) opened the door to imaging amyloid deposition in vivo, but limitations due to cost and half-life of the radiotracer restricted its use to research. The development of fluorinated radiotracers addressed some of these concerns and allowed the approval of these methods by the U.S. Food and Drug Administration (FDA) for clinical use. To date, three compounds have been approved, and guidelines for their use in clinical practice have been outlined (Table 10.4–9). Subsequently, tracers to visualize neurofibrillary tangles have been developed as well and are being studied but are not yet available for clinical use. Specific tracers that visualize amyloid and tangle deposition have allowed investigators to evaluate the pathologic process in vivo and are being used to monitor progression from asymptomatic stages to MCI, dementia due to AD. However, the burden of β-amyloid plaques does not always correlate with the clinical stages, as some MCI subjects can present with a minimal burden similar to healthy controls, while others have an amyloid burden comparable to AD subjects (Fig. 10.4–2) suggesting other factors may be at play. Thus, the combination of several markers further increases the accuracy of the prediction and will probably become the norm as a more specific outline of the disease course is established (Fig. 10.4–3).

MCI Diagnostic Differential

The Cognitive Continuum. The cognitive continuum describes the subtle pathway from age-related cognitive decline to MCI to dementia. Per this model, there is an overlap at both ends of MCI, which indicates that it can be quite challenging to identify the transition points (Fig. 10.4–4). In practice, differentiating MCI from age-related cognitive decline relies mainly on neuropsychological testing, showing a cognitive decline more severe for age and less education. The main differentiation between MCI and a major neurocognitive disorder due to AD resides in the lack of functional impairment in MCI. An attempt to correlate the cognitive changes within this continuum with the underlying pathologic processes has been underway for several years.

Table 10.4–9.
Criteria for the Use of Amyloid PET Imaging

Appropriate use would be considered in an individual with the following characteristics:

1. A cognitive complaint with objectively confirmed impairment
2. AD as a possible diagnosis, but the diagnosis is uncertain after a comprehensive evaluation by a dementia expert
3.

When knowledge of the presence or absence of Aβ pathology is expected to increase diagnostic certainty and alter management

  a. Patients with persistent or progressive unexplained MCI
  b. Patients satisfying core clinical criteria for possible AD because of unclear clinical presentation, either an atypical clinical course or an etiologically mixed presentation
  c. Patients with progressive dementia and atypically early age of onset (usually defined as 65 years or less in age)

Amyloid imaging is inappropriate in the following situations:

1. Patient with core clinical criteria for probable AD with a typical age of onset
2. To determine dementia severity
3. Based solely on a positive family history of dementia or the presence of APOEε4
4. Patients with a clinical complaint that is unconfirmed on clinical examination
5. In lieu of genotyping for suspected autosomal mutation carriers
6. In asymptomatic individuals
7. Nonmedical use (e.g., legal, insurance coverage, or employment screening)

Aβ, amyloid-beta; AD, Alzheimer disease; APOE, apolipoprotein E; MCI, mild cognitive impairment.

Reprinted with permission from Johnson KA, Minoshima S, Bohnen NI, et al. Appropriate use criteria for amyloid PET: a report of the Amyloid Imaging Task Force, The Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association. *Alzheimer's and Dementia.* 2013;9(1):e1–e16. © 2013 by the Alzheimer's Association, and the Society for Nuclear Medicine and Molecular Imaging. doi:10.1016/j.jalz.2013.01.002.

Information gathered from longitudinal studies suggests different pathways during MCI can lead to a similar presentation and pathologic findings when progression to dementia has occurred. In the case of AD, an initial model presented in 2009 suggested the major AD biomarkers became abnormal in a temporally ordered manner with amyloid accumulation (CSF and PET amyloid measures) occurring early, followed by CSF tau, FDG PET, and structural MRI occurring later and leading to clinical changes. The MCI syndrome was thought to become evident once a large portion of amyloid had been accumulated and only minimal neurodegeneration had occurred. A more recent model suggests biomarkers do not occur in sequence but rather become simultaneously more abnormal. Within this model, β-amyloid is considered necessary but not sufficient to produce the AD syndrome, and while its accumulation appears to loosely correlate with the rate of cognitive decline, neurodegeneration biomarkers appear closely associated with a decline in both typical and atypical forms of the disease. The time lag between biomarker evidence and the appearance of clinical symptoms would be moderated by factors such as cognitive reserve and comorbid disorders.

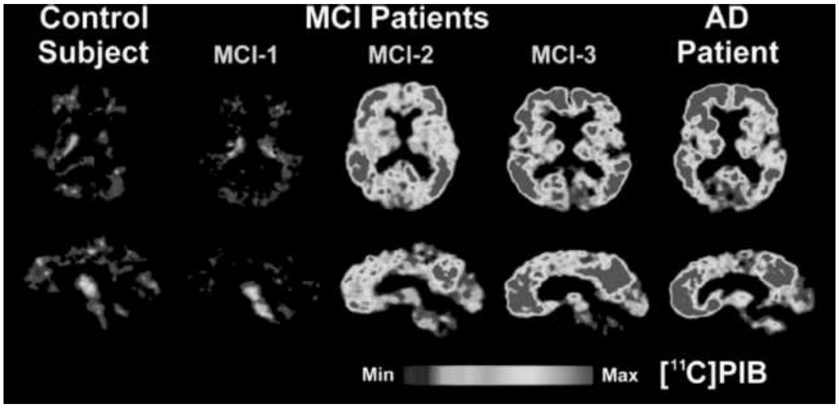

FIGURE 10.4–3. Positron emission tomography images were obtained with the amyloid-imaging agent Pittsburgh Compound-B ([carbon-11]-PIB) in normal control (*far left*), three different patients with mild cognitive impairment (MCI; *center images*), and a mild Alzheimer disease (AD) patient (*far right*). Some MCI patients have control-like levels of amyloid, some have Alzheimer disease–like levels of amyloid, and some have intermediate levels. (Courtesy of William E. Klunk, M.D., Ph.D., University of Pittsburgh, Department of Psychiatry, Pittsburgh, PA. All rights retained.)

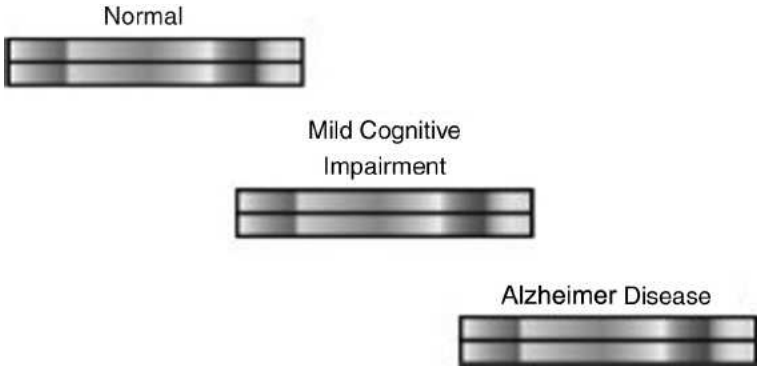

FIGURE 10.4–4. Cognitive continuum showing the overlap in the boundary between normal aging and mild cognitive impairment and Alzheimer disease. (Adapted with permission from Petersen RC. Normal aging, mild cognitive impairment, and early Alzheimer's disease. *Neurologist.* 1995;1:326–344.)

Longitudinal data from ADNI and Mayo Clinic has demonstrated some puzzling findings. Patients with MCI followed over time can be subdivided according to their positivity in biomarkers. Some will have only amyloid deposition, others only evidence of neurodegeneration, a third group both amyloid deposition and neurodegeneration and lastly, a fourth group with no evidence of either one. Subjects with evidence of both amyloid deposition and neurodegeneration appear to progress at the fastest rate toward dementia, while subjects with no evidence of either one progress very slowly, if at all. However, those with only amyloid deposition do not progress as quickly as those with only neurodegeneration. In fact, those with neurodegeneration progress almost as quickly as the ones with both amyloid and neurodegeneration in the Mayo sample. These findings speared the evolution of the concept, SNAP (Suspected Non-Alzheimer Pathology), suggesting that some subjects with MCI could have an entity that resembles AD but behaves differently at the pathologic level. In 2019, a consensus workgroup proposed the term limbic-predominant age-related TDP-43 encephalopathy (LATE). LATE is intended to describe a tauopathy where the distribution of TDP-43 is distinct from the one seen in FTLD, is often associated with hippocampal sclerosis, clinically presents with an amnestic dementia syndrome similar to AD, is commonly seen in patients over the age of 80, and could explain the findings seen in some of the subjects with MCI who had been designated as having SNAP. According to the working group's proposal, the term LATE is intended to encompass earlier terms used to designate TDP-43 proteinopathies which may be associated with cognitive impairment such as hippocampal sclerosis, hippocampal sclerosis of aging, hippocampal sclerosis dementia, cerebral age-related TDP-43 with sclerosis, and TDP-43 pathologies in the elderly. In the Rush Memory and Aging Project and Religious

Orders Study, half the participants had TDP-43 pathology, and approximately 37% had both TDP-43 pathology and AD. This data and other studies seem to suggest mixed pathologies are often the norm in aging brains.

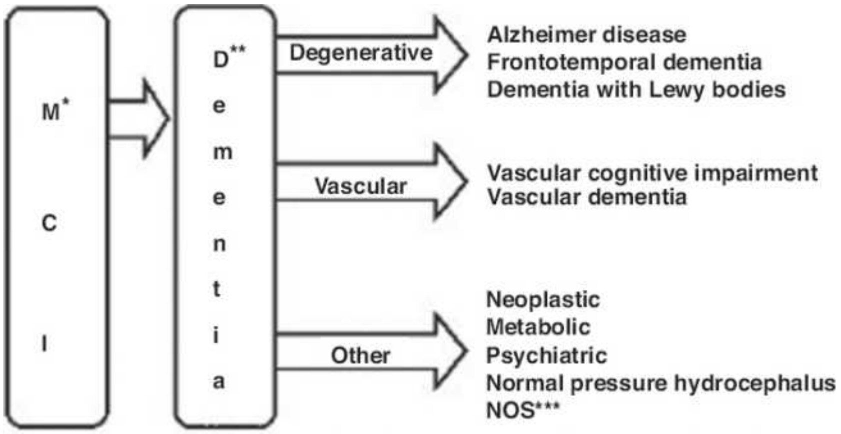
M* C I
D** e m e n t i a
Degenerative
Alzheimer disease
Frontotemporal dementia
Dementia with Lewy bodies
Vascular
Vascular cognitive impairment
Vascular dementia
Other
Neoplastic
Metabolic
Psychiatric
Normal pressure hydrocephalus
NOS***

* Impairment in any cognitive domain but of insufficient severity to affect function

** Impairment in any two or more cognitive domains and sufficiently severe to affect function

*** Does not fit a known dementia syndrome

FIGURE 10.4–5. Overview of mild cognitive impairment preceding the diagnosis of dementia. NOS, not otherwise specified. (Reprinted with permission from Petersen RC, O'Brien J. Mild Cognitive Impairment Should Be Considered for DSM-V. *J Geriatr Psychiatr Neurol.* 2006;19(3):147–154. doi:10.1177/0891988706291085; originally adapted with permission from Petersen RC. Conceptual overview. In: Petersen RC. *Mild Cognitive Impairment: Aging to Alzheimer's Disease*. Oxford University Press. Copyright © 2003 by Mayo Foundation for Medical Education and Research.)

The integration of data generated from following subjects who are asymptomatic but have the presence of biomarkers, people with subjective complaints who do not meet the criteria for MCI and very early MCI subjects will allow a much better characterization of all the possible pathologic pathways within the cognitive continuum and their clinical correlations.

Figure 10.4–5 outlines the various etiologic pathways associated with an initial MCI clinical presentation.

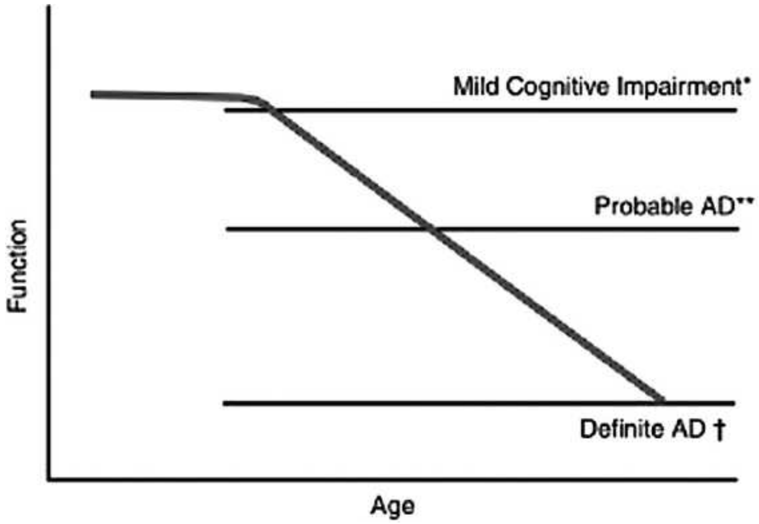

FIGURE 10.4–6. Hypothetical change in function as an individual develops Alzheimer disease. (Reprinted from Petersen RC. Aging, mild cognitive impairment, and Alzheimer's disease. *Neurologic Clinics.* 18(4):789–805. Copyright © 2000 W. B. Saunders Company. doi: 10.1016/ s0733-8619(05)70226-7.)

Course and Prognosis

The typical rate at which MCI patients progress to AD is 10% to 15% per year and is associated with progressive loss of function (Fig. 10.4–6). However, there are several studies indicating that the diagnosis is not stable in both directions; patients can either convert to AD or revert back to normal. This variability in the course is related to the heterogeneous source of the subjects (clinical vs. community) as well as to the heterogeneous definition criteria used by different studies. a-MCI has been associated with increased mortality (greater in multiple-domains a-MCI than in single-domain a-MCI) (hazard ratio 1.7) when compared with reference subjects.

Epidemiologic studies that followed over time large numbers of community-dwelling subjects, which eventually developed dementia, consistently demonstrate a slow change in cognitive measures with a sharp increase in the rate of decline seen a few years before the clinical diagnosis of MCI or dementia is made. The point in time when this inflection occurs varies from study to study. In some, the rate of cognitive decline increases sharply 5 to 6 years before the AD diagnosis, while in others, the change is seen only 3 years before. Subjects diagnosed with a-MCI appear to present with cognitive changes earlier and decline more rapidly when compared to na-MCI subjects.

The clinical manifestations and rate of decline in MCI are likely not only dependent on the underlying neuropathology but also on a host of factors that can either exacerbate or ameliorate the biologic impact of the underlying disease. Some of these factors include but are not limited to the expression of specific genes, the age of disease onset, baseline brain and cognitive reserve, comorbid medical and psychiatric illness, cerebrovascular burden, and lifestyle characteristics (e.g., diet, physical and cognitive activity). The modulation of some of these factors has been considered a potential way to decrease the likelihood of progression from MCI to dementia.

MCI Treatment

There are no FDA-approved treatments for MCI at this time. MCI treatment involves adequate screening and diagnosis. Ideally, MCI treatment would also include improvement of memory loss, together with the prevention of further cognitive decline to dementia. Cognitive training programs have been reported as mildly beneficial for compensating memory difficulties in MCI. Controlling for vascular risk factors (high blood pressure, hypercholesterolemia, diabetes mellitus) may be a preventive method for those MCI cases underlying vascular pathology. Currently, sensitive tools (imaging techniques or biomarkers) are not available for MCI screening in the general population. In the primary care setting, clinicians should maintain a high suspicion for subjective cognitive complaints and should corroborate these complaints with collateral information whenever possible. Also, identifying reversible causes of cognitive impairment (hypothyroidism, vitamin B_{12} deficiency, medication-induced cognitive impairment, depression) can further benefit some of the prodromal dementia MCI cases.

Currently, there is no evidence for the long-term efficacy of pharmacotherapies in reversing MCI. Several epidemiologic studies indicated a reduced risk of dementia in persons taking antihypertensive medications, cholesterol-lowering drugs, antioxidants, and anti-inflammatory and estrogen therapy, but no randomized controlled trials verify these data. With regard to cognitive enhancers, so far, multiple trials designed for a-MCI have reported mostly negative results (Table 10.4–10). Most of these studies were confronted with several problems: (1) obtaining homogeneous samples and identifying potential beneficiaries of treatment; (2) treating a wider population, which led to large percentages of negative responses and problematic side effects; (3) translating the MCI construct into multiple cultures and languages and using AD diagnosis as the primary outcome, given the variability of this diagnosis in different countries. Clinical trials using immunologic methods including active immunization, monoclonal antibodies, and intravenous immunoglobulin (IV IG) therapy have been used in an attempt to reduce the amyloid burden and, in turn, slow or stop disease progression in patients with mild or moderate AD. Many compounds have proven effective in reducing amyloid deposits. However, this effect has not translated so far into clinical improvement. In October 2019, aducanumab, a monoclonal antibody directed at the N-terminus of the amyloid β-peptide, showed modest but significant clinical and functional benefits in a post hoc analysis of two phase 3 trials. Aducanumab was approved to treat patients with MCI and mild dementia due to AD in June 2021, although several experts have questioned the extent of the clinical benefits and the risk-benefit profile, mostly due to the potential for inflammatory brain changes that can be seen on imaging, including vasogenic edema, sulcal effusion, or microhemorrhages. These changes are known as amyloid-related imaging abnormalities (ARIA). Other agents are currently being studied during symptomatic phases of the disease (gantenerumab, donanemab, BAN2401, ALZ-801, plasma exchange). However, many researchers are putting their efforts into treating patients with biologic markers of disease before they develop symptoms. Both subjects with autosomal dominant inherited forms of AD and asymptomatic elderly subjects without genetic predisposition but positive biomarkers are being studied. If positive, these trials could lead the way in treating neurodegenerative disorders during the preclinical stages of disease, preventing people at risk from ever becoming symptomatic.

FUTURE DIRECTIONS

Advances in MCI detection remain paramount for early detection and treatment of AD, as experts agree that disease-modifying treatments for AD will focus on cognitively intact individuals at increased risk. The field of identifying sensitive and specific biomarkers (biologic and neuroimaging markers) will continue to witness exponential development in the coming years.

Table 10.4–10. Treatment Trials for MCI

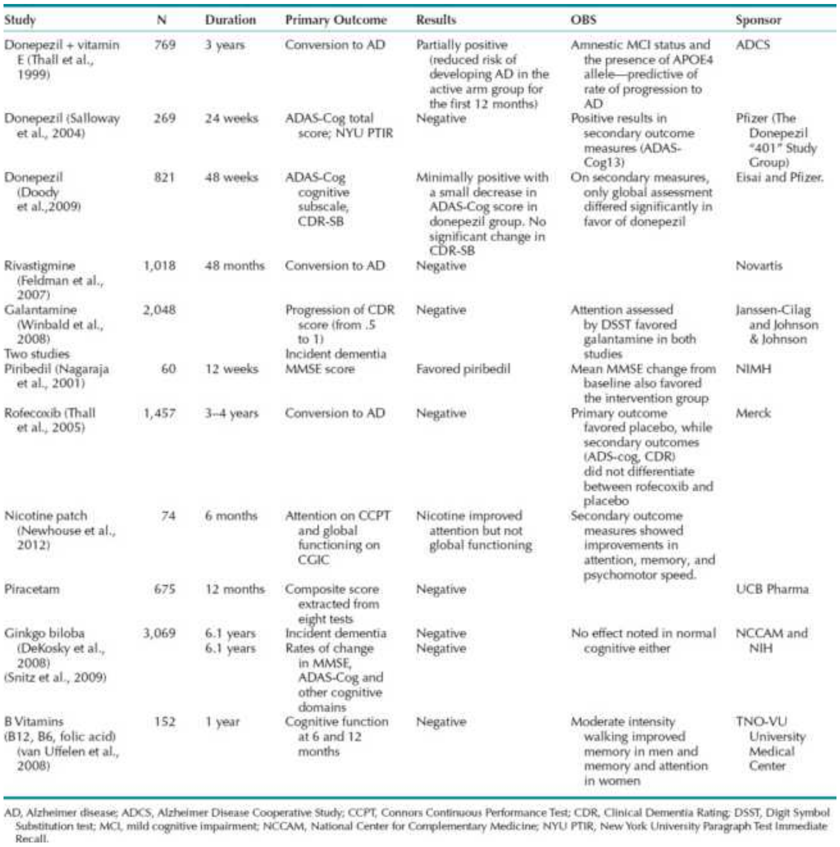

Study	N	Duration	Primary Outcome	Results	OBS	Sponsor
Donepezil + vitamin E (Thall et al., 1999)	769	3 years	Conversion to AD	Partially positive (reduced risk of developing AD in the active arm group for the first 12 months)	Amnestic MCI status and the presence of APOE4 allele—predictive of rate of progression to AD	ADCS
Donepezil (Salloway et al., 2004)	269	24 weeks	ADAS-Cog total score; NYU PTIR	Negative	Positive results in secondary outcome measures (ADAS-Cog13)	Pfizer (The Donepezil "401" Study Group)
Donepezil (Doody et al.,2009)	821	48 weeks	ADAS-Cog cognitive subscale, CDR-SB	Minimally positive with a small decrease in ADAS-Cog score in donepezil group. No significant change in CDR-SB	On secondary measures, only global assessment differed significantly in favor of donepezil	Eisai and Pfizer.
Rivastigmine (Feldman et al., 2007)	1,018	48 months	Conversion to AD	Negative		Novartis
Galantamine (Winbald et al., 2008) Two studies	2,048		Progression of CDR score (from .5 to 1) Incident dementia	Negative	Attention assessed by DSST favored galantamine in both studies	Janssen-Cilag and Johnson & Johnson
Piribedil (Nagaraja et al., 2001)	60	12 weeks	MMSE score	Favored piribedil	Mean MMSE change from baseline also favored the intervention group	NIMH
Rofecoxib (Thall et al., 2005)	1,457	3–4 years	Conversion to AD	Negative	Primary outcome favored placebo, while secondary outcomes (ADS-cog, CDR) did not differentiate between rofecoxib and placebo	Merck
Nicotine patch (Newhouse et al., 2012)	74	6 months	Attention on CCPT and global functioning on CGIC	Nicotine improved attention but not global functioning	Secondary outcome measures showed improvements in attention, memory, and psychomotor speed.	
Piracetam	675	12 months	Composite score extracted from eight tests	Negative		UCB Pharma
Ginkgo biloba (DeKosky et al., 2008) (Snitz et al., 2009)	3,069	6.1 years 6.1 years	Incident dementia Rates of change in MMSE, ADAS-Cog and other cognitive domains	Negative Negative	No effect noted in normal cognitive either	NCCAM and NIH
B Vitamins (B12, B6, folic acid) (van Uffelen et al., 2008)	152	1 year	Cognitive function at 6 and 12 months	Negative	Moderate intensity walking improved memory in men and memory and attention in women	TNO-VU University Medical Center

AD, Alzheimer disease; ADCS, Alzheimer Disease Cooperative Study; CCPT, Connors Continuous Performance Test; CDR, Clinical Dementia Rating; DSST, Digit Symbol Substitution test; MCI, mild cognitive impairment; NCCAM, National Center for Complementary Medicine; NYU PTIR, New York University Paragraph Test Immediate Recall.

AMNESTIC DISORDERS

Definition

The DSM-IV-TR would provide four diagnostic entities under the category of amnestic disorder: (1) amnestic disorder due to cerebral or systemic medical condition, (2) substance-induced amnestic disorder, (3) amnestic disorder due to unknown etiology, and (4) amnestic disorder NOS. DSM V criteria for mild and major neurocognitive disorder include all these categories (Table 10.4–1).

In the 11th edition of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-11), amnestic disorder is defined as a severe memory impairment relative to the individual's age and general level of intellectual functioning, not attributable to substance intoxication or substance withdrawal, and presumed to be attributable to an underlying neurologic condition (trauma, infection, tumor, or other disease process affecting specific areas of the brain) or to chronic use of specific substances or medications.

Epidemiology

As discussed in previous editions, data regarding the overall prevalence of amnestic disorders is lacking. The prevalence of alcohol-induced amnestic disorder varies from 0.2% to 4%, while studies of the incidence of transient global amnesia report 5.2 to 10 cases per 100,000 per year.

Table 10.4–11.
Causes of Amnestic Disorders

Primary Brain Conditions	Notes
Seizure Disorders	
Traumatic brain injury	The fastest-growing cause of amnestic disorder in the United States
Cerebral tumors	Mainly localized around the third ventricle
Stroke	In the posterior cerebral artery territory
Paraneoplastic syndromes (limbic encephalitis)	Focal inflammation of the temporal lobes
Infections (herpes simplex encephalitis)	Preferentially attacks the temporal lobes and produces a hemorrhagic necrosis
Metabolic (hypoxia)	Several causes (respiratory, cardiovascular, metabolic compromise or carbon monoxide intoxication) attacking the sensitive CA1 hippocampal region
Transient global amnesia	Profound anterograde and retrograde amnesia improving over the course of several days
ECT	Retrograde amnesia that usually remits after days/weeks
Systemic Conditions	
Alcohol-induced amnestic syndrome (Korsakoff syndrome)	
Thiamine deficiency	
Hypoglycemia	
Substance-induced amnestic disorder (excluding alcohol) sedative–hypnotics	

ECT, electroconvulsive therapy.

Etiology

The causes of amnestic disorders are listed in Table 10.4–11. The neuroanatomical structures involved in the etiology of amnestic memory have been described inferentially by matching various clinical syndromes with lesions in specific brain areas. These areas include the mamillary bodies, the hippocampus, the amygdala, the dorsomedial, and the midline thalamic nuclei. Bilateral lesions are considered the main culprit, but there are studies reporting unilateral (left regions being more often involved) damage as the cause of amnestic disorders. Some experts differentiate between hippocampal amnesia and diencephalic amnesia, although both manifest with disproportionate impairment of anterograde memory and time-related impairment of retrograde memory (worse closer to illness's onset) while maintaining an intact implicit memory, immediate recall, and attention.

Symptoms

Memory impairment is the core symptom of all amnestic disorders. It includes both anterograde amnesia (inability to learn new information) and retrograde amnesia (inability to recall previously learned information). Short-term memory can also be impaired to a variable degree. Typically, the long-term retrograde memory impairment is temporally graded (more remote memories are better preserved). Immediate recall is usually not affected. Attention and implicit learning (which includes procedural learning as well as other forms of learning without explicit awareness) are also intact.

Associated symptoms include confabulations, changes in personality (apathy, lack of initiative, impulsivity), as well as neurologic symptoms corresponding to the underlying illness (stroke/multiple sclerosis/alcohol-induced amnestic syndrome). The confabulations are more frequent in diencephalic amnesia (prototypical Korsakoff syndrome) than in hippocampal amnesia.

The following vignette presents the case of Clive Wearing, who developed severe anterograde amnesia following herpes encephalitis. His wife, Deborah Wearing, described his condition in her 2005 book *Forever Today*:

> His ability to perceive what he saw and heard was unimpaired. But he did not seem to be able to retain any impression of anything for more than a blink. Indeed, if he did blink, his eyelids parted to reveal a new scene. The view before the blink was utterly forgotten. Each blink, each glance away and back, brought him an entirely new view. I tried to imagine how it was for him…Something akin to a film with bad continuity, the glass half empty, then full, the cigarette suddenly longer, the actor's hair now tousled, now smooth. But this was real life, a room changing in ways that were physically impossible…It was as if every waking moment was the first waking moment. Clive was under the constant impression that he had just emerged from unconsciousness because he had no evidence in his own mind of ever being awake before…"I haven't heard anything, seen anything, touched anything, smelled anything," he would say. "It's like being dead."

Korsakoff Syndrome. In 1889, Sergei Korsakoff described a syndrome of polyneuritis, anterograde amnesia, and confabulations in subjects with chronic alcohol use. He did not, however, differentiate between the acute stage of the illness (the acute encephalopathy had been described 20 years earlier by Carl Wernicke) and the chronic memory impairment. That connection between the acute and the chronic condition was made at the beginning of the 20th century by Karl Bonhoeffer. Later, in the eighth edition of his classic textbook, Emil Kraepelin distinguished between alcohol-related Korsakoff syndrome and Korsakoff syndrome related to other nonalcoholic processes (e.g., neurosyphilis). The etiologic correlation between Korsakoff syndrome and thiamine deficiency caused by poor nutrition associated with chronic alcohol use was described in the early 1930s. Maurice Victor described the corresponding diencephalic lesion in the 1960s. Korsakoff syndrome usually develops secondary to Wernicke encephalopathy, and frequently they are described together (Wernicke–Korsakoff syndrome) (for a description of Wernicke encephalopathy clinical presentation, see chapter on delirium). However, some cases of Korsakoff syndrome develop in the absence of the initial stage of acute Wernicke encephalopathy. Throughout the years, the terms Korsakoff psychosis, Korsakoff dementia, and Korsakoff syndrome have been used interchangeably to describe the clinical presentation of alcohol-induced amnestic disorder.

The overall rates of both Wernicke encephalopathy and Korsakoff syndrome are as low as 0.4% in France (although they have one of the highest per capita consumption of alcohol) and as high as 2% in the United States, Scandinavia, and Australia. It is usually seen in subjects over age 40 (as it follows alcohol use of 20 or more years). Beyond malnutrition, alcohol use induces thiamine deficiency through folate deficiency (which reduces thiamine absorption from the gastrointestinal tract) and through direct damage to the intestinal mucosa.

Primary thiamine deficiency is caused by general malnutrition and other conditions like anorexia nervosa, acquired immune deficiency syndrome (AIDS), hyperemesis gravidarum, thyrotoxicosis, metastatic cancer, hypomagnesemia, long-term dialysis, and congestive heart failure treated long term with diuretics (renal tubular reabsorption of filtered thiamine is exacerbated by furosemide).

Thiamine pyrophosphate (TPP), the active form of thiamine, is involved in three enzymatic reactions essential for the metabolism of glucose and the production of several neurotransmitters, including acetylcholine.

Numerous mechanisms have been suggested to explain the selective loss of neurons found with thiamine deficiency. These include cerebral energy dysfunction, breakdown of the blood–brain barrier, altered glutamate neurotransmission caused by extracellular glutamate and its impaired transport function, together with *N*-methyl-D-aspartate (NMDA) receptor-mediated excitotoxicity, the accumulation of amyloid precursor-like protein, increased free radical production; increased expression of superoxide dismutase consequent to increased microglial response, induction of nitric oxide, and oxidative stress. Changes in a number of central nervous system (CNS) neurotransmitter systems have been reported in thiamine deficiency, suggesting that thiamine is important for normal neurotransmitter function. Acetylcholine, γ-aminobutyric acid (GABA), glutamate, and aspartate are produced primarily through the oxidative metabolism of glucose. Serotonin metabolism is also influenced by thiamine deficiency. Cerebrospinal fluid (CSF) levels of 5-hydroxyindoleacetic acid (5-HIAA) were reduced in patients with Wernicke–Korsakoff syndrome. Recent work suggests that thiamine may inhibit serotonin uptake, providing further evidence that thiamine may be acting as a neurotransmitter/neuromodulator as well as a cofactor in brain metabolism and that its deficiency may be associated with behavioral and mood changes. The anatomical lesions described in association with thiamine deficiency include the medial dorsal thalamic nuclei, the mammillary bodies, the fornix, and the cerebellum. The anterior thalamic nucleus has also been associated with thiamine deficiency; it receives input from the mammillary complex and has a major reciprocal connection with the subiculum–presubiculum portion of the hippocampal complex and with the cingulate gyrus. Lesions to these regions (anterior thalamic nuclei, mamillary complex) are related to impaired explicit memory.

Clinically, subjects with Korsakoff syndrome, while maintaining an overall intact intelligence, have a striking inability to form new memories (anterograde amnesia). The retrograde amnesia is less severe, although present to a certain degree (usually for events that occurred shortly before the onset of the illness). Immediate recall and implicit learning (e.g., learning procedures) are preserved. Individuals with Korsakoff syndrome classically present with confabulation, which is explained by most experts as an attempt to cover the memory deficit.

Assessment

Neuropsychological testing usually shows a discrepancy between the performance on memory scales and the performance on a general cognitive scale (a discrepancy between memory performance and overall intelligence).

Laboratory tests show evidence of chronic alcohol use, malnutrition, vitamin deficiency, or systemic illness. Neuroimaging tests are helpful in differentiating between several causes of the amnestic syndrome (mammillary bodies atrophy in Korsakoff syndrome, localized strategic strokes, anoxic lesions of the hippocampal structures, T2-weighted hyperintensities in bilateral temporal poles in herpes encephalitis).

Differential

In contrast to delirium, amnestic disorders evolve on a background of clear consciousness and intact attention, with no perceptual disturbances or mood lability. The amnestic syndrome has no impairment in abstract thinking, judgment, or executive functions. Cognitive impairment in dementia extends to semantic memory, language, and praxis.

Psychogenic amnesia can be differentiated by its presentation (onset after traumatic event, no evidence of substance/general medical condition, amnesia for personal identity, which is conserved in amnestic disorder, and for a circumscribed event, but preserved memory for new events and preserved ability for learning) and its course (abrupt onset and resolution, no residual impairment).

Course and Prognosis

About 25% of patients with Korsakoff syndrome recover; the rest continue to suffer from various degrees of memory impairment. A significant minority needs permanent placement. Most amnestic syndromes have a stable course. A few would recover (such as transient global amnesia or postelectroconvulsive therapy amnesia) or improve (such as amnestic disorder secondary to traumatic brain injury).

Treatment

Treatment with high doses of thiamine (1 to 2 g intravenous daily) is likely to partially or completely reverse Wernicke encephalopathy, but its benefits in the chronic stages (Korsakoff syndrome) are not established. There are case reports but no controlled studies regarding the benefits of donepezil (Aricept), rivastigmine (Exelon), and memantine (Namenda) in treating Korsakoff syndrome. For the other types of amnestic syndrome, the treatment is tailored accordingly to the etiology (acyclovir for herpes encephalitis, hyperbaric oxygen for carbon monoxide poisoning). However, once the amnesic deficit is installed, there are very few pharmacologic options. Most amnestic syndromes have a stable and some even a reversible course (transient global amnesia). Cognitive rehabilitation programs can help improve the patient’s level of functioning, but so far, none have indicated benefits regarding restoring/improving memory in the impaired domain.

FURTHER READINGS

Aggarwal NT, Wilson RS, Beck TL, Bienias JL, Berry-Kravis E, Bennett DA. The apolipoprotein E epsilon4 allele and incident Alzheimer’s disease in persons with mild cognitive impairment. *Neurocase*. 2005;11(1):3–7.

Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer’s disease: recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. *Alzheimers Dement*. 2011;7(3):270–279.

Ganguli M, Dodge HH, Shen C, DeKosky ST. Mild cognitive impairment, amnestic type: an epidemiologic study. *Neurology*. 2004;63(1):115–121.

Ganguli M. Mild cognitive impairment and the 7 uses of epidemiology. *Alzheimer Dis Assoc Disord*. 2006;20(3 Suppl 2):S52–S57.

Gauthier S, Reisberg B, Zaudig M, et al; International Psychogeriatric Association Expert Conference on mild cognitive impairment. Mild cognitive impairment. *Lancet*. 2006;367(9518):1262–1270.

Haas RH. Thiamin and the brain. *Annu Rev Nutr*. 1988;8:483–515.

Harding A, Halliday G, Caine D, Kril J. Degeneration of anterior thalamic nuclei differentiates alcoholics with amnesia. *Brain*. 2000;123(Pt 1):141–154.

Harper C, Fornes P, Duyckaerts C, Lecomte D, Hauw JJ. An international perspective on the prevalence of the Wernicke-Korsakoff syndrome. *Metab Brain Dis*. 1995;10(1):17–24.

Heister D, Brewer JB, Magda S, Blennow K, McEvoy LK; Alzheimer’s Disease Neuroimaging Initiative. Predicting MCI outcome with clinically available MRI and CSF biomarkers. *Neurology*. 2011;77(17):1619–1628.

Jack CR Jr, Bennett DA, Blennow K, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer’s disease. *Alzheimers Dement*. 2018;14(4):535–562.

Jack CR Jr, Knopman DS, Jagust WJ, et al. Tracking pathophysiological processes in Alzheimer’s disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol*. 2013;12(2):207–216.

Jack CR Jr, Wiste HJ, Weigand SD, et al. Amyloid-first and neurodegeneration-first profiles characterize incident amyloid PET positivity. *Neurology*. 2013;81(20):1732–1740.

Jelic V, Kivipelto M, Winblad B. Clinical trials in mild cognitive impairment: lessons for the future. *J Neurol Neurosurg Psychiatry*. 2006;77(4):429–438.

Katzman R, Terry R, DeTeresa R, et al. Clinical, pathological, and neurochemical changes in dementia: a subgroup with preserved mental status and numerous neocortical plaques. *Ann Neurol*. 1988;23(2):138–144.

Kirshner HS. Mild cognitive impairment: to treat or not to treat. *Curr Neurol Neurosci Rep*. 2005;5:455–457.

Klunk WE, Engler H, Nordberg A, et al. Imaging brain amyloid in Alzheimer’s disease with Pittsburgh Compound-B. *Ann Neurol*. 2004;55(3):306–319.

Kruse M, Navarro D, Desjardins P, Butterworth RF. Increased brain endothelial nitric oxide synthase expression in thiamine deficiency: relationship to selective vulnerability. *Neurochem Int*. 2004;45(1):49–56.

Larrieu S, Letenneur L, Orgogozo JM, et al. Incidence and outcome of mild cognitive impairment in a population-based prospective cohort. *Neurology*. 2002;59(10):1594–1599.

Lopez OL, Jagust WJ, Dulberg C, et al. Risk factors for mild cognitive impairment in the Cardiovascular Health Study Cognition Study: part 2. *Arch Neurol*. 2003;60(10):1394–1399.

Loy CT, Schofield PR, Turner AM, Kwok JBJ. Genetics of dementia. *Lancet*. 2014;383(9919):828–840.

Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain*. 2019;142(6):1503–1527.

Petersen RC, Aisen P, Boeve BF, et al. Mild cognitive impairment due to Alzheimer disease in the community. *Ann Neurol*. 2013;74(2):199–208.

Petersen RC, Doody R, Kurz A, et al. Current concepts in mild cognitive impairment. *Arch Neurol*. 2001;58(12):1985–1992.

Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med*. 2004;256(3): 183–194.

Rademakers R, Baker M, Gass J, et al. Phenotypic variability associated with progranulin haploinsufficiency in patients with the common 1477C→T (Arg493X) mutation: an international initiative. *Lancet Neurol*. 2007;6(10):857–868.

Smith G, Ivnik RJ, Petersen RC, Malec JF, Kokmen E, Tangalos E. Age-associated memory impairment diagnoses: problems of reliability and concerns for terminology. *Psychol Aging*. 1991;6(4):551–558.

Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer’s disease: recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. *Alzheimers Dement*. 2011;7(3):280–292.

Thompson PM, Hayashi KM, Dutton RA, et al. Tracking Alzheimer’s disease. *Ann N Y Acad Sci*. 2007;1097:183–214.

Wenisch E, Cantegreil-Kallen I, De Rotrou J, et al. Cognitive stimulation intervention for elders with mild cognitive impairment compared with normal aged subjects: preliminary results. *Aging Clin Exp Res*. 2007;19(4):316–322.

Winblad B, Palmer K, Kivipelto M, et al. Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med*. 2004;256(3):240–246.

10.5 Cognitive Disorders Due to Another Medical Condition or Substance

ROBERT J. BOLAND, M.D., AND IQBAL AHMED, M.D., F.R.C.PSYCH.

Most cognitive disorders, particularly the dementias, are caused by the disorders already discussed in this section. However, many other medical and neurologic disorders may cause cognitive deficits, and we discuss some of the most common in this section. The DSM-5-TR diagnosis of "Major or Mild Neurocognitive Disorder Due to Another Medical Condition" covers many of these causes. However, the DSM also includes several diagnoses for causes not usually included as types of dementia, such as traumatic brain injury (TBI) or HIV infection. Here we concentrate on cognitive disorders due to other medical conditions, substances, multiple etiologies, and cases when the cause is unknown.

NOSOLOGY

DSM-5-TR includes the following diagnosis:

- *Major or Mild Neurocognitive Disorder due to Another Medical Condition*
- *Substance/Medication-Induced Major or Mild Neurocognitive Disorder*
- *Major or Mild Neurocognitive Disorder Due to Multiple Etiologies*

They also include the residual category of *Unspecified Neurocognitive Disorder*.

For each of the diagnoses (see Tables 10.5–1 to 10.5–5), there are three primary criteria:

1. The patient meets the general criteria for a major or mild neurocognitive disorder (Table 10.5–1)
2. There is evidence for the cause
3. Another cause does not better explain the noncognitive deficit.

ICD-11 lists four specific dementias:

- *Alzheimer disease*
- *Progressive focal atrophies*
- *Lewy body disease*
- *Frontotemporal lobar degeneration*

ICD-11 groups the remaining causes under two diagnoses:

- *Other specified disorders with a neurocognitive impairment as a major feature*
- *Disorders with neurocognitive impairment as a major feature*

The difference between these two diagnoses rests on whether the cause of the impairment is known. These diagnoses are considered residual categories and not defined in any detail.

ETIOLOGY

Many medical conditions and substances can cause cognitive dysfunction. A common strategy for organizing these is to group these conditions based on their primary mechanisms of action, such as neurologic, other medical, and substance-induced causes.

DIAGNOSTIC AND CLINICAL FEATURES

Like any medical or psychiatric illness, diagnosing a cognitive or mental disorder due to another medical condition depends on comprehensive history, detailed physical examination, and appropriate laboratory and imaging studies. For the history, specific details of the mental symptoms must be obtained, including presenting features, onset and course, and associated symptoms. More than any other part of the assessment, the description of the presenting features and the timing of onset is the key to the correct diagnosis. We should also explore past medical, psychiatric, and family history, as this can increase the pretest probability of a specific diagnosis and help dictate diagnostic testing. Clinicians should also comprehensively review all prescription and nonprescription medications, vitamins, dietary and herbal supplements, drugs of abuse, and environmental toxin exposures.

Physical examination should include, but is not limited to, cognitive assessment, vital signs, and a comprehensive neurologic examination. Table 10.5–6 outlines several frequently used general cognition tests and specific cognitive domains.

Neuropsychological testing refers to administering a comprehensive battery of multiple tasks evaluating a range of cognitive domains. Examination of specific organ systems is indicated by history and presenting symptoms. Specific findings on the physical examination may narrow the list of potential causative medical conditions (Table 10.5–7). An array of laboratory, brain imaging, and additional diagnostic studies may help diagnose the causative condition (Table 10.5–8). However, performing all possible tests on all patients would be inappropriate and a waste of medical resources. The choice of diagnostic studies should be guided by history, presenting symptoms, and physical examination findings. Table 10.5–9 summarizes the general principles for diagnosing disorders due to another medical condition or substance/medication-induced disorders.

Table 10.5–1.
DSM-5-TR Criteria for Major or Mild Neurocognitive Disorder

Major Neurocognitive Disorder

Diagnostic Criteria

A. Evidence of significant cognitive decline from a previous level of performance in one or more cognitive domains (complex attention, executive function, learning and memory, language, perceptual-motor, or social cognition) based on:
   1. Concern of the individual, a knowledgeable informant, or the clinician that there has been a significant decline in cognitive function; and
   2. A substantial impairment in cognitive performance, preferably documented by standardized neuropsychological testing or, in its absence, another quantified clinical assessment.

B. The cognitive deficits interfere with independence in everyday activities (i.e., at a minimum, requiring assistance with complex instrumental activities of daily living such as paying bills or managing medications).

C. The cognitive deficits do not occur exclusively in the context of a delirium.

D. The cognitive deficits are not better explained by another mental disorder (e.g., major depressive disorder, schizophrenia).

Specify whether due to:

Note: Each subtype listed has specific diagnostic criteria and corresponding text, which follow the general discussion of major and mild neurocognitive disorders.

- ▶ Alzheimer disease
- ▶ Frontotemporal degeneration
- ▶ Lewy body disease
- ▶ Vascular disease
- ▶ Traumatic brain injury
- ▶ Substance/medication use
- ▶ HIV infection
- ▶ Prion disease
- ▶ Parkinson disease
- ▶ Huntington disease
- ▶ Another medical condition
- ▶ Multiple etiologies
- ▶ Unspecified etiology

Coding note: Code based on medical or substance etiology. In most cases of major neurocognitive disorder, there is need for an additional code for the etiologic medical condition, which must immediately precede the diagnostic code for major neurocognitive disorder, as noted in the coding table on pp. 682–683.

Specify (see coding table for details):

- ▶ **Without behavioral disturbance:** If the cognitive disturbance is not accompanied by any clinically significant behavioral disturbance.
- ▶ **With behavioral disturbance** *(specify disturbance):* If the cognitive disturbance is accompanied by a clinically significant behavioral disturbance (e.g., psychotic symptoms, mood disturbance, agitation, apathy, or other behavioral symptoms). **Coding note:** Use additional code(s) to indicate clinically significant psychiatric symptoms due to the same medical condition causing the major neurocognitive disorder (e.g., **F06.2** psychotic disorder due to Alzheimer disease, with delusions; **F06.32** depressive disorder due to Parkinson disease, with major depressive–like episode). **Note:** Mental disorders due to another medical condition are included with disorders with which they share phenomenology.

Specify current severity:

- ▶ **Mild:** Difficulties with instrumental activities of daily living (e.g., housework, managing money).
- ▶ **Moderate:** Difficulties with basic activities of daily living (e.g., feeding, dressing).
- ▶ **Severe:** Fully dependent.

Coding and Recording Procedures

The following are examples of coding and recording major neurocognitive disorders due to an etiologic subtype *(for more information, see coding table on pp. 682–683 and coding notes in the specific* *diagnostic criteria for each major and mild neurocognitive disorder subtype)*:

- **Major neurocognitive disorder due to probable Alzheimer disease, without behavioral disturbance, mild:** G30.9 Alzheimer disease, F02.80 major neurocognitive disorder due to probable Alzheimer disease, without behavioral disturbance, mild.
- **Major neurocognitive disorder due to traumatic brain injury, with behavioral disturbance, moderate:** S06.2X9S diffuse traumatic brain injury with loss of consciousness of unspecified duration, sequela; F02.81 major neurocognitive disorder due to traumatic brain injury, with behavioral disturbance, moderate; F06.34 bipolar and related disorder due to traumatic brain injury, with mixed features.

Mild Neurocognitive Disorder

Diagnostic Criteria

A. Evidence of modest cognitive decline from a previous level of performance in one or more cognitive domains (complex attention, executive function, learning and memory, language, perceptual-motor, or social cognition) based on:
   1. Concern of the individual, a knowledgeable informant, or the clinician that there has been a mild decline in cognitive function; and
   2. A modest impairment in cognitive performance, preferably documented by standardized neuropsychological testing or, in its absence, another quantified clinical assessment.

B. The cognitive deficits do not interfere with capacity for independence in everyday activities (i.e., complex instrumental activities of daily living such as paying bills or managing medications are preserved, but greater effort, compensatory strategies, or accommodation may be required).

C. The cognitive deficits do not occur exclusively in the context of a delirium.

D. The cognitive deficits are not better explained by another mental disorder (e.g., major depressive disorder, schizophrenia).

Specify whether due to:

Note: Each subtype listed has specific diagnostic criteria and corresponding text, which follow the general discussion of major and mild neurocognitive disorders.

- Alzheimer disease
- Frontotemporal degeneration
- Lewy body disease
- Vascular disease
- Traumatic brain injury
- Substance/medication use
- HIV infection
- Prion disease
- Parkinson disease
- Huntington disease
- Another medical condition
- Multiple etiologies
- Unspecified etiology

Coding note: For mild neurocognitive disorder due to any of the medical etiologies listed above, code **G31.84.** Do *not* use additional codes for the presumed etiologic medical conditions. For substance/medication-induced mild neurocognitive disorder, code based on type of substance. For unspecified mild neurocognitive disorder, code **R41.9.**

Specify (behavioral disturbance cannot be coded but should still be recorded):

- **Without behavioral disturbance:** If the cognitive disturbance is not accompanied by any clinically significant behavioral disturbance.
- **With behavioral disturbance** *(specify disturbance):* If the cognitive disturbance is accompanied by a clinically significant behavioral disturbance (e.g., psychotic symptoms, mood disturbance, agitation, apathy, or other behavioral symptoms). **Coding note:** Use additional code(s) to indicate clinically significant psychiatric symptoms due to the same medical condition causing the mild neurocognitive disorder (e.g., **F06.2** psychotic disorder due to traumatic brain injury, with delusions; **F06.32** depressive disorder due to HIV disease, with major depressive–like

episode). **Note:** Mental disorders due to another medical condition are included with disorders with which they share phenomenology.

Coding and Recording Procedures

The following are examples of coding and recording mild neurocognitive disorders due to an etiologic subtype *(for more information, see coding table on pp. 682–683 and coding notes in the specific diagnostic criteria for each major and mild neurocognitive disorder subtype)*:

- G31.84 Mild neurocognitive disorder due to Alzheimer disease, without behavioral disturbance.
- G31.84 Mild neurocognitive disorder due to traumatic brain injury, with behavioral disturbance; F06.34 bipolar and related disorder due to traumatic brain injury, with mixed features.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:679–681.

Table 10.5–2.
Major or Mild Neurocognitive Disorder Due to Another Medical Condition

Diagnostic Criteria

A. The criteria are met for major or mild neurocognitive disorder.

B. There is evidence from the history, physical examination, or laboratory findings that the neurocognitive disorder is the pathophysiologic consequence of another medical condition (e.g., multiple sclerosis).

C. The cognitive deficits are not better explained by another mental disorder (e.g., major depressive disorder) or another specific neurocognitive disorder (e.g., major neurocognitive disorder due to Alzheimer disease).

Coding note (see coding table on pp. 682–683):

For major neurocognitive disorder due to another medical condition, with behavioral disturbance, code first the other medical condition, followed by the major neurocognitive disorder due to another medical condition, with behavioral disturbance (e.g., G35 multiple sclerosis, F02.81 major neurocognitive disorder due to multiple sclerosis, with behavioral disturbance). For major neurocognitive disorder due to another medical condition, without behavioral disturbance, code first the other medical condition, followed by the major neurocognitive disorder due to another medical condition, without behavioral disturbance (e.g., G35 multiple sclerosis, F02.80 major neurocognitive disorder due to multiple sclerosis, without behavioral disturbance). *Note:* The severity specifiers “mild,” “moderate,” and “severe” cannot be coded for major neurocognitive disorder but should still be recorded.

For mild neurocognitive disorder due to another medical condition, code G31.84. (*Note*: Do *not* use the additional code for the other medical condition. “With behavioral disturbance” and “without behavioral disturbance” cannot be coded but should still be recorded.)

For major or mild neurocognitive disorder due to another medical condition: Use additional code(s) to indicate clinically significant psychiatric symptoms due to another medical condition (e.g., F06.32 depressive disorder due to multiple sclerosis, with major depressive–like episode).

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:729, 730.

COURSE AND PROGNOSIS

The course and prognosis of neurocognitive and mental disorders due to other medical conditions and substance/medication depend on the specific medical condition and its response to the treatment or the specific substance/medication involved. The course can range from remission of all symptoms to progressive decline, ultimately leading to death. When the neurocognitive or mental disorder has been of brief duration, and the medical condition is easily identified and responds to treatment, the neurocognitive or mental disorder may improve. However, when the underlying causative condition is long standing or progressive, the causal substance cannot be discontinued, or its effects are permanent (e.g., previous environmental toxin exposure), a favorable outcome is less likely even when we can treat the causative medical condition (e.g., neurosyphilis).

Table 10.5–3.
Substance/Medication-Induced Major or Mild Neurocognitive Disorder

Diagnostic Criteria

A. The criteria are met for major or mild neurocognitive disorder.
B. The neurocognitive impairments do not occur exclusively during the course of a delirium and persist beyond the usual duration of intoxication and acute withdrawal.
C. The involved substance or medication and duration and extent of use are capable of producing the neurocognitive impairment.
D. The temporal course of the neurocognitive deficits is consistent with the timing of substance or medication use and abstinence (e.g., the deficits remain stable or improve after a period of abstinence).
E. The neurocognitive disorder is not attributable to another medical condition or is not better explained by another mental disorder.

Coding note (see also coding table on pp. 682–683): The ICD-10-CM codes for the [specific substance/medication]-induced neurocognitive disorders are indicated in the table below. Note that the ICD-10-CM code depends on whether or not there is a comorbid substance use disorder present for the same class of substance. In any case, an additional separate diagnosis of a substance use disorder is not given.

Substance-induced major neurocognitive disorder: If a mild substance use disorder is comorbid with the substance-induced major neurocognitive disorder, the 4th position character is “1,” and the clinician should record “mild [substance] use disorder” before the substance-induced major neurocognitive disorder (e.g., “mild inhalant use disorder with inhalant-induced major neurocognitive disorder”). For alcohol and sedative, hypnotic, or anxiolytic substances, a mild substance use disorder is insufficient to cause a substance-induced major neurocognitive disorder; thus, there are no available ICD-10-CM codes for this combination. If a moderate or severe substance use disorder is comorbid with the substance-induced major neurocognitive disorder, the 4th position character is “2,” and the clinician should record “moderate [substance] use disorder” or “severe [substance] use disorder,” depending on the severity of the comorbid substance use disorder. If there is no comorbid substance use disorder, then the 4th position character is “9,” and the clinician should record only the substance-induced major neurocognitive disorder.

Substance-induced mild neurocognitive disorder: If a mild substance use disorder is comorbid with the substance-induced mild neurocognitive disorder, the 4th position character is “1,” and the clinician should record “mild [substance] use disorder” before the substance-induced mild neurocognitive disorder (e.g., “mild cocaine use disorder with cocaine-induced mild neurocognitive disorder”). If a moderate or severe substance use disorder is comorbid with the substance-induced mild neurocognitive disorder, the 4th position character is “2,” and the clinician should record “moderate [substance] use disorder” or “severe [substance] use disorder,” depending on the severity of the comorbid substance use disorder. If there is no comorbid substance use disorder, then the 4th position character is “9,” and the clinician should record only the substance-induced mild neurocognitive disorder.

The severity specifiers “mild,” “moderate,” and “severe” (for major neurocognitive disorder) and the accompanying symptom specifiers “with behavioral disturbance” and “without behavioral disturbance” (for major or mild neurocognitive disorder) cannot be coded but should still be recorded.

Specify if:

- Persistent: Neurocognitive impairment continues to be significant after an extended period of abstinence.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022: 712–713.

Table 10.5–4.
Major or Mild Neurocognitive Disorder Due to Multiple Etiologies

Diagnostic Criteria

A. The criteria are met for major or mild neurocognitive disorder.

B. There is evidence from the history, physical examination, or laboratory findings that the neurocognitive disorder is the pathophysiologic consequence of more than one etiologic process, excluding substances (e.g., neurocognitive disorder due to Alzheimer disease with subsequent development of vascular neurocognitive disorder).

  - *Note*: Refer to the diagnostic criteria for the various neurocognitive disorders due to specific medical conditions for guidance on establishing the particular etiologies.

C. The cognitive deficits are not better explained by another mental disorder and do not occur exclusively during the course of a delirium.

Coding note (see coding table on pp. 682–683):

For major neurocognitive disorder due to multiple etiologies, code first all of the etiologic medical conditions (with the exception of vascular disease, which is not coded), followed by either F02.81 for major neurocognitive disorder due to multiple etiologies, with behavioral disturbance; or F02.80 for major neurocognitive disorder due to multiple etiologies, without behavioral disturbance. If vascular disease is among the multiple etiologic medical conditions, code next either F01.51 for major vascular neurocognitive disorder, with behavioral disturbance; or F01.50 for major vascular neurocognitive disorder, without behavioral disturbance. *Note:* The severity specifiers “mild,” “moderate,” and “severe” cannot be coded for major neurocognitive disorder but should still be recorded.

For example, for a presentation of major neurocognitive disorder, moderate, with a behavioral disturbance, that is judged to be due to Alzheimer disease, vascular disease, and HIV infection, and in which heavy chronic alcohol use is judged to be a contributing factor, code the following: G30.9 Alzheimer disease, B20 HIV infection; F02.81 major neurocognitive disorder due to Alzheimer disease and HIV infection, moderate, with behavioral disturbance; F01.51 major vascular neurocognitive disorder, moderate, with behavioral disturbance; and F10.27 severe alcohol use disorder with alcohol-induced major neurocognitive disorder, moderate, nonamnestic-confabulatory type.

For mild neurocognitive disorder due to multiple etiologies, code G31.84. (Note: Do *not* use the additional codes for the etiologies. “With behavioral disturbance” and “without behavioral disturbance” cannot be coded but should still be recorded.)

For major or mild neurocognitive disorder due to multiple etiologies: Use additional code(s) to indicate clinically significant psychiatric symptoms due to the various etiologies (e.g., F06.2 psychotic disorder due to Alzheimer disease, with delusions; F06.31 depressive disorder due to cerebrovascular disease, with depressive features).

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:731.

Table 10.5–5.
Unspecified Neurocognitive Disorder

(R41.9)

This category applies to presentations in which symptoms characteristic of a neurocognitive disorder that cause clinically significant distress or impairment in social, occupational, or other important areas of functioning predominate but do not meet the full criteria for any of the disorders in the neurocognitive disorders diagnostic class. The unspecified neurocognitive disorder category is used in situations in which the precise etiology cannot be determined with sufficient certainty to make an etiologic attribution.

Coding note: For unspecified major or mild neurocognitive disorder, code R41.9. (*Note*: Do *not* use additional codes for any presumed etiologic medical conditions. “With behavioral disturbance” and “without behavioral disturbance” cannot be coded but should still be recorded.)

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:732.

Table 10.5–6.
Commonly Used Cognitive Assessment Tools

Cognitive Domain	Examples of Specific Cognitive Tests
Global cognition	Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), Mini-Cog, Dementia Rating Scale, Saint Louis University Mental Status Examination, Abbreviated Mental Test score (AMTS), Addenbrooke Cognitive Assessment (ACE) and Mini-ACE
Learning and memory	California Verbal Learning Test, Rey–Osterrieth Complex Figure Test
Visuospatial/construction	Clock Drawing Test, Block Design Test
Language	Boston Naming Test, Controlled Word Association (Fluency)
Executive function	Trail Making Test B, Stroop Test
Attention	Digit Span-Forward, Continuous Performance Tests
Psychomotor speed	Digit Symbol Substitution Test, Trail Making Test A

Table 10.5–7.
Specific Neurologic Signs and Symptoms and Associated Medical Conditions

Neurologic Sign or Symptom	Possible Related Medical Conditions
Change in consciousness	Infection, hepatic encephalopathy, substance-induced hypoglycemia (diabetes), thiamine deficiency
Headache	Meningitis, tumor, giant cell arteritis
Tremor	Parkinson disease, alcohol abuse, human immunodeficiency virus–associated dementia, Wilson disease
Gait disturbance	Normal pressure hydrocephalus, stroke, Parkinson disease
Meningeal signs	Meningitis
Involuntary movements	Huntington disease, Parkinson disease, Parkinson plus syndromes such as Lewy body dementia, autoimmune encephalitis
Pathologic reflexes	Creutzfeldt–Jakob disease, hypothyroidism
Motor deficit	Stroke, amyotrophic lateral sclerosis, Parkinson and Parkinson plus syndromes, Huntington disease
Sensory deficit	Stroke, B_{12} deficiency
Gaze palsy	Thiamine deficiency, alcohol abuse
Cranial nerve defect	Stroke
Language disturbance	Stroke, Alzheimer disease, neurosyphilis

TREATMENT

The mainstay of treatment is the correction or amelioration of the underlying medical condition or discontinuation of the offending substance or exposure. Adjuvant treatments include psychopharmacologic interventions and psychotherapeutic interventions.

Table 10.5–8.
Diagnostic Tests for Cognitive and Mental Disorders Due to a General Medical Condition

Laboratory studies
Complete blood count[a]
Electrolytes[a]
Fasting glucose[a]
Blood urea nitrogen and creatinine[a]
Calcium, magnesium, and phosphorus[a]
Liver function tests[a]
Thyroid function tests[a]
Rapid plasma reagin/Venereal Disease Research Laboratory
Vitamin B_{12} and folate[a]
Erythrocyte sedimentation rate
Arterial blood gases
Toxicology
Urinalysis with or without culture/sensitivity
Human immunodeficiency virus
Ceruloplasmin
Catecholamines
Porphobilinogens
Specific drug level (e.g., medications and drugs of abuse)
Specific toxins (e.g., lead, mercury)
Brain imaging
Structural brain imaging: computed tomography or magnetic resonance imaging[a]
Functional magnetic resonance imaging
Positron emission tomography
Single photon emission computed tomography
Additional tests
Electroencephalography
Cerebrospinal fluid for biomarkers, infectious agents
Chest radiograph
Electrocardiogram

---

[a] Indicates tests routinely performed when evaluating mental disorders; all other tests are performed if the history or physical examination suggests a need.

Table 10.5–9.

General Principles for Diagnosing Disorders Due to Another Medical Condition or Substance/Medication-Induced Disorders

Disorders Due to Another Medication Condition:

- The symptoms of the psychiatric disturbance may be indistinguishable from that of a primary psychiatric disorder.
- Symptoms cannot be better accounted for by a primary psychiatric disorder or reflect an adjustment disorder due to the stress of having a medical illness.
- The disturbance does not occur exclusively during delirium.
- A temporal link between the course of the psychiatric disturbance and the onset, exacerbation, or remission of the medical disorder supports a causal link.
- Treatment of the underlying medical disorder improves the neurocognitive symptoms.
- Evidence from the medical literature supports a link between psychiatric disturbance and medical illness.
- Evidence of atypical features (e.g., late onset, lack of prior episodes, and lack of family history) support a link between the psychiatric disturbance and the medical condition.
- A targeted workup aims to “rule in” a specific medical diagnosis suggested by the patient’s somatic symptoms.

Substance/Medication-Induced Disorders:

- The symptoms of the psychiatric disturbance may be indistinguishable from that of a primary psychiatric disorder.
- Symptoms cannot be better accounted for by a primary psychiatric disorder or psychiatric disorder due to another medical condition.
- The disturbance does not occur exclusively during delirium.
- The symptoms of the psychiatric disturbance arise in the context of initiating or changing the dose of a medication or during intoxication with or withdrawal from a substance and are more than those usually associated with the medication’s use or the substance’s intoxication or withdrawal state and warrant independent clinical attention.
- The absence of psychiatric symptoms supports a link between the psychiatric disturbance and the substance/medication prior to use of the substance/medication, resolution of symptoms within 4 weeks of last substance/medication use, lack of prior similar episodes that occurred independently of substance/medication and the presence of psychiatric symptoms consistent with the extent of substance/medication use.
- Evidence from the medical literature supports a link between psychiatric disturbance and substance/medication.

Clinicians typically use psychopharmacologic interventions to treat associated neuropsychiatric symptoms, including mood or behavioral symptoms, especially when correction of the causal medical condition or withdrawal of the causal substance does not result in remission. However, given the medically compromised nature of the patient, the clinician should be mindful of the possibility of altered pharmacokinetics (e.g., decreased clearance of drugs) and potential drug–drug interactions. Perhaps most important is the frequent lack of empirical evidence supporting psychopharmacologic interventions for these conditions—often, we are treating by analogy—treating a disorder with a given medication because it looks similar to a disorder for which that medication is effective. Although sometimes this is a reasonable strategy, we should remain skeptical when we treat outside the bounds of studied indications and be prepared to discontinue ineffective medications.

Psychotherapeutic interventions in neurocognitive and mental disorders due to medical conditions include behavior approaches used in other neurocognitive disorders, including Alzheimer disease. Other specific interventions promote life-coping strategies, educate patients about illness courses, and bolster supportive networks with family and friends. Detoxification, medication use for cravings and relapse prevention, and a comprehensive rehabilitation program are frequently necessary for mental disorders due to substance use, but relapse rates are high.

SPECIFIC DISORDERS

Major or Mild Neurocognitive Disorder Due to Another Medical Condition

Definition. We should reserve the DSM-5 category of neurocognitive disorder due to another medical condition for neurocognitive disorders that are not among the more specific diagnostic categories. Although current evidence is limited, many potential mechanisms exist for the association between specific medical conditions and neurocognitive disorder. For example, direct neurotoxicity and disruption in cholinergic neurotransmission may be common pathways.

For this diagnosis, the evidence must suggest that the neurocognitive disorder is the pathophysiologic consequence of another medical condition. In addition, DSM-5 requires that another neurocognitive disorder (e.g., Alzheimer disease) should not better explain the cognitive deficits. A temporal relationship between the onset of the medical disorder and the onset of the cognitive impairment is strong evidence of an association, as is a reduction or stabilization of the cognitive impairment upon treatment for the causative medical condition. Thus, this is a diagnosis of inclusion, not exclusion, and we should not use it when the cause of the cognitive deficit is unknown (for that, Unspecified Neurocognitive Disorder is more appropriate).

Epidemiology. Given the many possible causes of major or mild neurocognitive disorder due to another medical condition, it is difficult to give a meaningful prevalence of this disorder.

Etiology. Table 10.5–10 lists some of the common causes of cognitive decline due to secondary causes.

Etiology is dependent on the underlying causative medical condition. One example within this category is Wilson disease, an autosomal recessive disorder caused by an enzymatic defect in copper metabolism, resulting in a toxic buildup of copper in the brain. The clinical course includes both cognitive, other neuropsychiatric, and motor symptoms (e.g., tremor).

Structural brain lesions (e.g., tumor, subdural hematoma, and normal pressure hydrocephalus) can cause cognitive impairment by compression of specific neuronal tissue, disruption of the blood–brain barrier, or impedance of cerebral blood flow. In addition, these lesions cause specific cognitive deficits based on the location of the lesion. Normal pressure hydrocephalus classically presents with cognitive impairment, urinary incontinence, and gait abnormality.

Specific metabolic abnormalities include organ dysfunction (e.g., hypothyroidism, diabetes mellitus, hypercalcemia, heart failure, and kidney disease) and vitamin deficiencies. Heart disease can result in cerebral hypoperfusion, and hyper- and hypoglycemia have disrupted cholinergic systems. Kidney disease can cause a buildup of toxic metabolites.

Table 10.5–10.
Medical Conditions and Substances/Medications Causing Neurocognitive Disorders and Other Mental Disorders

Neurologic

Degenerative diseases (Alzheimer disease, corticobasal degeneration, frontal–temporal dementia [Pick disease], Huntington disease, Lewy body disease, Parkinson disease, prion diseases, progressive supranuclear palsy)

Demyelinating diseases (multiple sclerosis)

Acquired injuries (cerebrovascular disease, chronic traumatic encephalopathy [e.g., dementia pugilistica], traumatic brain injury)

Space occupying lesions (abscess, normal pressure hydrocephalus, subdural hematoma, tumor)

Other Medical Disorders

Metabolic disorders (acid–base imbalance, metabolic brain diseases, disorders of calcium metabolism, DNA repair-deficiency disorders, glucose metabolism disorders, hyperlactatemia, iron metabolism disorders, lipid metabolism disorders, malabsorption syndromes, metabolic syndrome X, inborn error of metabolism, mitochondrial diseases, phosphorus metabolism disorders, porphyrias, proteostasis, deficiencies, metabolic skin diseases, wasting syndrome, water-electrolyte imbalance)

Anemias (aplastic anemia, iron deficiency anemia, sickle cell anemia, thalassemia, vitamin deficiency anemia)

Diseases targeting specific organs (Wilson disease/hepatolenticular degeneration, end-stage organ dysfunction)

Endocrine disorders (adrenal, pancreases, thyroid)

Nutritional disorders (folate deficiency, niacin deficiency, vitamin B1 deficiency, vitamin B12 deficiency

Infectious/inflammatory disorders (human immunodeficiency virus [HIV], slow viruses, e.g., Creutzfeldt–Jakob disease, neurosyphilis, Lyme disease

Autoimmune disorders (giant cell arteritis, N-methyl-D-aspartate Receptor (NMDAR) antibody encephalitis, systemic lupus erythematosus)

Substances

Medications (anticholinergic, anticonvulsants, antihypertensive, anxiolytics/sedative–hypnotics, corticosteroids, immunosuppressants, opioids, stimulants, sympathomimetics)

Substances of abuse[a] (alcohol, cannabis, cocaine, hallucinogens, inhalants)

Toxins (carbon monoxide, heavy metals, lead, mercury)

[a] Many medications can also be abused.

The classic vitamin deficiency associated with cognitive impairment is Wernicke–Korsakoff syndrome, which is due to neurologic damage secondary to B vitamin deficiency, and in particular, deficiency of thiamine (vitamin B_1). While frequently associated with chronic alcohol misuse, Wernicke–Korsakoff syndrome can occur in any clinical situation that is caused by or causes malnutrition (e.g., Crohn disease, chemotherapy). The estimated prevalence of Wernicke encephalopathy in the general population is 2%. Thiamine is involved in glucose and lipid metabolism as well as neurotransmitter metabolism. Wernicke encephalopathy (or Wernicke disease) is classically characterized by ophthalmoplegia, ataxia, and confusion, although this classic presentation is uncommon and a range of neurologic symptoms are possible. If a patient receives adequate parenteral thiamine and other nutritional support, Wernicke encephalopathy is reversible. When Wernicke encephalopathy occurs in the setting of chronic alcohol misuse, most patients progress to Korsakoff syndrome. Dense amnesia with confabulation characterizes Korsakoff syndrome. Recovery from the cognitive impairment of Korsakoff syndrome is less common and less complete. Based on brain imaging, patients with Wernicke–Korsakoff syndrome typically demonstrate atrophy of mammillary bodies and thalamus.

Inflammatory and infectious states (e.g., neurosyphilis, chronic meningitis, and systemic lupus erythematosus [SLE]) may cause cognitive dysfunction, and the mechanisms may vary. For example, SLE may impair the blood–brain barrier and cause microglial activation. Many infections can affect the brain. Most recently, the cognitive effects of SARS-CoV-2 have become apparent. However, the exact mechanism is not understood, and both direct neuroinvasion (e.g., via the olfactory nerve or through an ocular route) and transmission via the blood (by penetrating the blood-brain barrier) are possible.

Similarly, specific medications (e.g., anticholinergics), substances (e.g., alcohol), and toxins (e.g., lead) can cause cognitive impairment and dementia, also through varied mechanisms.

Damage or disruption of neurotransmission in areas involved in memory (e.g., hippocampus) can result in persistent amnesia. For example, specific infectious diseases such as herpes simplex can result in encephalitis that primarily targets the medial temporal lobes of the cortex, including the hippocampus. Similarly, anti-NMDA receptor (NMDAR) encephalitis can cause cognitive dysfunction in addition to other neuropsychiatric symptoms. Their effect appears to be mediated through autoantibodies against the GluN1 subunit of the NMDA receptor. This subunit is essential for memory consolidation in the hippocampus.

Diagnosis and Clinical Features. According to DSM-5, the development of new cognitive impairment independent of delirium or another neurocognitive disorder is the hallmark of this diagnosis. In addition, a temporal and causative link between the cognitive impairment and the medical condition or substance must exist (Table 10.5–6).

AG is a 76-year-old woman with congestive heart failure, systolic dysfunction, hypertension, coronary artery disease and atrial fibrillation, osteoarthritis, hypothyroidism, and diabetes mellitus. She had been hospitalized repeatedly over the past year for acute heart failure decompensation resulting in hypotension, dyspnea, and leg edema. She admitted to poor compliance with her medications, contributing to the decompensation of her heart failure. She said that she could not remember to take them and began to have problems with her memory about 1 year ago when she was first hospitalized for congestive heart failure. She was also noncompliant with dietary and fluid restrictions. Prescribed medications included lisinopril, furosemide, metoprolol, aspirin, glipizide, acetaminophen, and levothyroxine.

On initial examination, congestive heart failure seemed well compensated without pulmonary edema and only mild peripheral edema. AG's heart rate was 68, and her blood pressure was 95/45. Given her difficulties with memory, she underwent cognitive testing. AG scored 26/30 on the Montreal Cognitive Assessment, with difficulty on serial sevens. Attention was otherwise intact, but her clock drawing indicated poor planning and inability to indicate the correct time. She also displayed difficulty with problem solving and abstract reasoning. Brain imaging revealed mild small vessel disease but no infarcts. Thyroid-stimulating hormone level was normal, and the only lab abnormalities were a creatinine of 1.5 and glycosylated hemoglobin of 7.9.

She was again hospitalized with hypoxia and pulmonary edema 2 weeks later. She was hypotensive and required pressor support to manage her decompensated heart failure. AG was diagnosed with major neurocognitive disorder due to congestive heart failure with cerebral hypoperfusion without behavioral disturbance.

Laboratory Evaluation. Patients should receive a thorough workup to identify the underlying causative medical condition and ruling out any more common neurocognitive disorders (e.g., cerebrovascular disease). Given the extent of medical conditions that can cause cognitive impairment and dementia, we should tailor the testing to the history and the presenting symptoms. For example, we may order an erythrocyte sedimentation rate (ESR) if an inflammatory disorder is suspected, although ESR is a nonspecific test. Copper and ceruloplasmin levels can be measured if Wilson disease is suspected. A detailed history of the proband and the family is essential to guide this workup. Initial workups usually include structural brain imaging, and functional brain imaging may help differentiate neurocognitive syndromes in specific cases. Lumbar puncture is not routine, but advances in understanding specific CSF markers may help the differential diagnosis, especially if Creutzfeldt–Jakob disease (CJD), neurosyphilis, or other infectious etiology is suspected. Toxicology tests may be appropriate in select cases. Neuropsychological testing can aid in the diagnosis.

Differential Diagnosis. Ruling out delirium and all other neurocognitive disorders is the most critical part of the differential diagnosis.

Course and Prognosis. Onset may be sudden or slowly progressive. The course is variable depending on etiology. Recovery may occur over the short- and long-term depending on the severity of the initial injury and the persistence of the underlying medical condition.

Treatment. Treatment of the underlying medical condition is the most important therapy. For example, prevention of cognitive impairment is possible in individuals at risk of thiamine deficiency (e.g., secondary to long-standing alcohol misuse or severe malnutrition) through the administration of vitamin $B_{1,}$ particularly in the intravenous form, as this will prevent the development of Wernicke–Korsakoff syndrome. In addition, memory retraining or using memory aids may have some benefits.

Substance/Medication-Induced Mild/Major Neurocognitive Disorder

Definition. Patients commonly have cognitive impairment when intoxicated or during the acute withdrawal phase from substances and some medication. However, the categories of major or mild substance/medication-induced neurocognitive disorder only apply when the cognitive impairment persists beyond the typical intoxication or withdrawal period. DSM-5 also clarifies that the suspected substance/medication must have a known link to persistent cognitive impairment and that the onset of the cognitive impairment must be temporally associated with exposure to the substance/medication.

Epidemiology. The prevalence of substance/medical-induced neurocognitive disorder is unknown. However, risk factors for developing this disorder include older age, a more extended period of exposure to the substance or medication, poor nutrition, liver disease, and cerebrovascular disease.

Etiology. Persistent cognitive impairment likely reflects the slow recovery of underlying brain function. However, for some patients, the neurocognitive disorder is irreversible. For example, chronic alcohol abuse resulting in thiamine deficiency can cause irreversible Korsakoff syndrome due to permanent damage to the mammillary bodies and thalamus. Table 10.5–10 lists some substances that may cause cognitive deficits. Given the many new drugs, both legal and illicit, such lists are never comprehensive, and the purpose here is to give some typical examples of frequently seen drugs and drug classes.

Diagnosis and Clinical Features. Chronic users of sedative, hypnotic, or anxiolytic drugs can develop an neurocognitive disorder, typically mild and primarily consisting of impairment in memory. However, studies demonstrate that chronic users continue to demonstrate impairment in multiple cognitive domains, especially verbal memory and processing speed, even after 6 months of abstinence.

Chronic cocaine, methamphetamine, opioids, and phencyclidine users risk developing a mild persistent neurocognitive disorder. For methamphetamine, neurocognitive disorder primarily involves impairment in learning, memory, and executive functions. Some evidence suggests that prolonged abstinence (i.e., 14 months) from methamphetamine is associated with recovery in some but not all cognitive domains. Methamphetamine and cocaine use increases the risk of stroke, which can result in additional persistent cognitive impairment.

Chronic inhalation of volatile hydrocarbons, for example, inhalant abuse, can produce impairment in memory, attention, visuospatial and executive functions, and motor abilities. Studies have demonstrated that abstinent individuals can recover over 2 years. However, permanent disability may occur in those individuals with severe abuse or those who had abused leaded gasoline products and developed concomitant lead encephalopathy with cerebellar damage.

As discussed above, chronic alcohol use can lead to confusion (Wernicke encephalopathy) and progress to a dense amnestic state (Korsakoff syndrome). Other substances that can cause amnestic syndrome are cannabis, lead, carbon monoxide, and some insecticides.

Other medications associated with cognitive impairment include steroids, chemotherapy agents, nucleoside reverse transcriptase inhibitors (NRTIs), L-Dopa and dopamine agonists, select antiepilepsy drugs (e.g., topiramate), and medications with anticholinergic properties. Cognitive deficits induced by medications typically resolve upon discontinuation of the offending drug. Corticosteroid-induced cognitive deficits typically affect declarative (verbal) and working memory. This is true for acute and chronic treatment, and the pattern suggests that the hippocampus and dorsolateral prefrontal cortex are sensitive to the adverse effects of steroids.

Pathology and Laboratory Examination. Other than toxicology studies to help identify the causative substance/medication, there are no other pathognomonic lab findings or imaging findings for major or mild substance/medication-induced neurocognitive disorder. Therefore, the patient should receive a thorough workup to identify potential medical causes and rule out any more common neurocognitive disorders.

Differential Diagnosis. The differential includes delirium and all other neurocognitive disorders. It is important to remember that persons who misuse substances are more prone to other events associated with cognitive impairment, such as TBI and certain infections, for example, HIV. They may also be more likely to have comorbid psychiatric diagnoses, such as posttraumatic stress disorder (PTSD) and mood disorders, including cognitive dysfunction. In addition, some substances of abuse, such as cocaine and methamphetamine, are associated with an increased risk of stroke, which can also produce cognitive impairment.

Course and Prognosis. For many substances and medications, abstinence leads to improved cognitive function. However, the time course of recovery is highly variable and depends on the substance/medication used, the severity and duration of use, and the presence of other risk factors for cognitive impairment.

Treatment. For neurocognitive disorder, the treatment of choice is the cessation of that substance or medication due to substances and medications. Unfortunately, this may not be possible for some medications. Often, nutritional and other medical support can help improve cognition. For example, prevention of cognitive impairment is possible in individuals at risk for thiamine deficiency (e.g., secondary to long-standing alcohol misuse) through the administration of vitamin B_1, as this will prevent the development of Wernicke–Korsakoff syndrome. Cases of substance misuse will require a comprehensive treatment program, including medication-assisted treatment (MAT), after stabilization.

Major or Mild Neurocognitive Disorder Due to Multiple Etiologies

DSM-5-TR includes this diagnosis for cases where multiple diseases or injuries are likely to have contributed to the cognitive disorder. For example, it is not unusual to see evidence of brain degeneration coexisting with vascular changes, or patients with Alzheimer disease could also have HIV. Similarly, patients with a neurodegenerative disorder may also drink heavily.

Unspecified Neurocognitive Disorder

DSM-5-TR also includes a *Neurocognitive Disorder For Cases With Unknown Etiology*. The analogous ICD-11 diagnosis is *Disorders With Neurocognitive Impairment As A Major Feature, Unspecified*. In cases of this diagnosis, the clinician should pursue continued efforts to identify a cause, which may become more apparent as the disorder progresses.

ACKNOWLEDGMENT

This chapter represents an update and revision of previous chapters in this series, most recently by Jennifer Clare Prins, M.D, and Ellen M. Whyte, M.D. from the last edition. We are indebted to Drs Prins and Whyte for their work and retain much of their prose in this current edition.

Further Readings

Agrawal S, Kumar S, Ingole V, et al. Does anemia affects cognitive functions in neurologically intact adult patients: two year cross sectional study at rural tertiary care hospital. *J Family Med Prim Care*. 2019;8(9):3005–3008.

doi:10.4103/jfmpc.jfmpc_599_19

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR*. 5th ed. American Psychiatric Association Publishing; 2022.

By the 2019 American Geriatrics Society Beers Criteria® Update Expert Panel. American Geriatrics Society 2019 updated AGS Beers Criteria® for potentially inappropriate medication use in older adults. *J Am Geriatr Soc*. 2019;67(4):674–694. doi:10.1111/jgs.15767

Chen Y, Yang W, Chen F, Cui L. COVID-19 and cognitive impairment: neuroinvasive and blood-brain barrier dysfunction. *J Neuroinflammation*. 2022;19(1):222. doi:10.1186/s12974-022-02579-8

Dalmau J, Armangué T, Planagumà J, et al. An update on Anti-NMDA receptor encephalitis for neurologists and psychiatrists: mechanisms and models. *Lancet Neurol*. 2019;18(11):1045–1057. doi:10.1016/S1474-4422(19)30244-3

Liu YF, Pan L, Feng M. Structural and functional brain alterations in cushing's disease: a narrative review. *Front Neuroendocrinol*. 2022;67:101033. doi:10.1016/j.yfrne.2022.101033

Isenberg-Grzeda E, Kutner HE, Nicolson SE. Wernicke-Korsakoff-syndrome: under-recognized and under-treated. *Psychosomatics*. 2012;53(6):507–516. doi:10.1016/j.psym.2012.04.008

Judd LL, Schettler PJ, Brown ES, et al. Adverse consequences of glucocorticoid medication: psychological, cognitive, and behavioral effects. *Am J Psychiatry*. 2014;171(10):1045–1051. doi:10.1176/appi.ajp.2014.13091264

Mizrachi M, Anderson E, Carroll KR, Tehrani N, Volpe BT, Diamond B. Cognitive dysfunction in SLE: an understudied clinical manifestation. *J Autoimmun*. 2022;132:102911. doi:10.1016/j.jaut.2022.102911

Wang Y, Lv J, He J, Wen G, Wu X. Mechanism of psychoactive substance-induced cognitive disorders: does tau protein play a role? *Front Biosci (Landmark Ed)*. 2022;27:(1):6. doi:10.31083/j.fbl2701006

World Health Organization (WHO). International classification of diseases, eleventh revision (ICD-11). 2019. https://icd.who.int/browse11

11

Substance-Related Disorders

11.1 Substance Use Disorders: Introduction

JOJI SUZUKI, M.D.

The opioid crisis in the United States, considered by many to be one of the most significant public health disasters of our times, has now entered the third decade. The crisis began in the 1990s with the rise in the use of prescription opioid analgesics to manage both acute and chronic pain, partly in response to the call to make pain the “fifth vital sign.” The widespread nonmedical use of these opioids grew dramatically, leading to the dramatic increase in the development of opioid use disorder among those misusing these drugs, which directly contributed to the rise in both intentional and unintentional fatal opioid overdoses. The ever-worsening crisis has now claimed the lives of over half a million individuals. During and after the COVID-19 pandemic, the United States experienced the most fatal period of drug-related overdoses ever recorded in a 12-month period. Instead of prescription opioids, we are now battling heroin and synthetic opioids, notably fentanyl, along with cocaine and methamphetamine. Through all of this, the morbidity and mortality associated with alcohol have continued to climb. Of great concern to mental health practitioners, this increase in substance-related suffering is occurring while the rate of suicides and suicide-related behaviors have also climbed dramatically. Some estimate that up to 30% of all overdose deaths are, in fact, suicides. Despite all the efforts in the past decade to combat the opioid crisis, these data suggest we are not winning this war, but instead, we are falling behind. Therefore, this is a call to all mental health practitioners that we must do more to tackle this crisis, given that patients with psychiatric disorders are disproportionally affected by substance use disorders (SUDs).

As in the previous edition, this section provides a series of chapters addressing SUDs for each of the major classes of misused substances as described in the DSM-5-TR's chapter on Substance-Related and Addictive Disorders. The selection of chapter topics is generally based on the categorization of substances as presented in that section, with the exception of anabolic–androgenic steroids. Given the cultural interest and concern with anabolic steroid use, the topic remains here even though there is not a separate steroid use disorder section recognized in the DSM-5-TR. In addition, even though gambling disorder is now in the SUD section, it is not addressed in this section of this textbook and is addressed separately. As one considers the broader categorization of medical disorders that have a core feature that defines them by dysregulation of choice (an inability by the person to control choice in a healthy and reasonable manner), then the inclusion of gambling and drug use disorders is a highly appropriate move on the part of the field. It seems likely that the underlying phenomenology and biology of disorders characterized by this dysregulation of choice share common features, logic, and assumptions that can help psychiatrists and others in the mental health field to better understand these disorders, their etiology, and their treatment. Other disorders that also have a fundamental feature of choice dysregulation, such as eating disorders and sexual disorders, should also be considered in this categorization (in future iterations of DSM), as their biology and social circumstances can have features that overlap with the SUDs as well as gambling disorder.

The remainder of this introduction will briefly review some general topics related to SUDs—highlighting the diagnosis of these disorders, the extent of use, and public policy issues. Details of specific disorders, including their treatment, can be found in the subsequent chapters.

The Definition of Substance Use Disorders

SUD is best understood as a chronic biopsychosocial disease, similar to hypertension or diabetes, manifesting as a behavioral syndrome characterized by repeated use of a substance, inability to stop or cut back, accompanied by strong subjective urges, physiologic dependence, and accumulating harm. Sometimes this can be referred to as the "3 Cs" of addiction—loss of control, craving, and negative consequences. The DSM-5-TR criteria indeed fall into these categories.

A significant change with DSM-5-TR from DSM-IV is that there no longer are separate diagnoses of "substance abuse" and "substance dependence." DSM-5-TR merges the criteria for these two diagnoses, unifying them under the term "use disorder" but adding a criterion for craving and removing the criterion for legal problems. These changes have resulted in essentially 11 criteria used to make a diagnosis of an SUD, although it is worth noting that the specific criteria can vary slightly from one class to the next. In addition, instead of conceptualizing SUD as a categorical diagnosis (i.e., either you do or do not have an SUD), the modifiers indicate the severity as mild (2 to 3 criteria), moderate (4 to 5 criteria), or severe (6 or more criteria). Clinical programs, insurers, and researchers often use the cutoff of moderate or severe as the necessary severity for study protocols and the authorization of more intensive levels of treatment services. Additional modifiers are also available, for example, "on maintenance therapy" for opioid use disorders if the patient is being treated with a medication such as methadone or buprenorphine.

While clinicians are most familiar with the DSM-5 criteria, other criteria and coding for diagnoses are used in clinical practice and for billing purposes—most notably, the ICD-11 coding. ICD-11 differs from DSM-5 in several ways, including the inclusion of problematic use categories of "harmful use" and "dependence." ICD-11 also has more drug use categories—for example, separating cocaine from other "stimulants, including amphetamines, methamphetamine, or methcathinone."

Finally, while not diagnostic, the field of addiction research and clinical care continues to use various forms of quantity and frequency of substance use as a metric for determining risk of use, treatment needs, and study outcomes. Such work has primarily focused on alcohol, the identification of standard drinks of alcohol, and daily numbers of such drinks that increase the risk of adverse consequences to use. This work has helped to move the alcohol field forward in its use of the concept of heavy drinking days (HDDs), which can be a useful marker in studies of treatment outcomes. With the growing liberalization of cannabis, there is also an increasing need to measure the quantity and frequency of cannabis use, given the large variety of forms of cannabis that are used widely today. While there are limitations to extending these concepts to other drugs, the overall concept of quantity and frequency of use may have utility for both clinical screening and research purposes, and this is a potential area of development in the coming years.

The Extent of Substance Use Disorders

In the United States, annual surveys of substance use are conducted under the sponsorship of a federal agency, the Substance Abuse and Mental Health Services Administration (SAMHSA). These surveys date back to the 1970s and for many years were titled the "National Household Survey on Drug Abuse," although the name of these reports was changed to the "National Survey on Drug Use and Health" (NSDUH) as of 2002. Subsequent chapters will address the prevalence of specific types of substances used; however, some overall findings will be noted here based on the 2020 NSDUH results. While the focus here in this brief review is on NSDUH data, other surveys in the United States provide complementary data that can be useful for the reader interested in gaining greater knowledge on drug abuse epidemiology. These include the Monitoring the Future (MTF) study of students and young adults, the Treatment Episode Data Set (TEDS), which examines patterns of substance abuse treatment admissions, the National Survey of Substance Abuse Treatment Services (N-SSATS), which assesses treatment programs, the Drug Abuse Warning Network (DAWN), which assesses emergency department admissions related to substances, and other large studies of drug abuse such as the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC).

In 2020, 162.5 million persons in the United States 12 years and older reported past month use (i.e., defined as “current use”) of tobacco, alcohol, or an illicit drug. Alcohol remained the most commonly used substance in the past month, with 138.5 million persons reporting current use. Among those 12 and older, 22.2% (61.6 million) reported a past month binge of alcohol, defined as drinking on the same occasion five or more drinks for men and four or more drinks for women. The data, therefore, suggest that close to half of those who consume alcohol have had a past month binge. Furthermore, 12.8% of those who drink alcohol and 28.8% of those who had a binge episode meet criteria for heavy drinking, defined as five or more episodes of binge drinking within a 30-day period.

Second to alcohol, the most commonly used substances are tobacco (51.7 million persons who meet criteria for current use) and vaping products (10.4 million persons who meet criteria for current use). Prior to 2020, NSDUH did not include questions about vaping, highlighting how rapidly new substance use can emerge; indeed, even though tobacco use has gone down overall, electronic cigarette use has approached epidemic proportions in the United States.

The most commonly used illicit drug is marijuana, with 32.8 million persons with current use, followed by prescription analgesic misuse (2.5 million), prescription sedative misuse (2.2 million), cocaine (1.8 million), hallucinogens (1.8 million), methamphetamine (1.7 million), and prescription stimulant misuse (1.5 million). While these findings demonstrate that there is substantial experimentation with illicit drugs, past month use of any illicit drug, cannabis, and nonmedical use of psychotherapeutics generally represents less than 20% of those who have ever used an illicit drug. While this should by no means minimize the extent of drug use, it does demonstrate that many people try a drug but do not go on to regular or problematic use.

Other national surveys are conducted by the European Union and Australia; the former conducts these under the auspices of the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), and the latter is conducted by the Australian Institute of Health and Welfare (AIHW). In general, findings from Australia, Europe, and the United States are similar (e.g., that there are high rates of cannabis use as the most commonly used illicit drug, that rates of prescription drug use have been a growing concern, and that tobacco and alcohol are more extensively used than illicit drugs). When making cross-regional comparisons (which are not easily done for a variety of reasons, such as the tools that are used), it is notable that the United States reports on drug use for persons 12 and older, the European Union for persons 15 and older, and Australia uses a cutoff of 14 and older.

Finally, while the number of persons or rates of use capture one aspect of substance use, the impact of this use can be lost in such numbers. The Global Burden of Disease study looked at a variety of medical conditions worldwide and found that two of the top five leading causes of disease were substance abuse related—tobacco (fourth) and alcohol use (fifth). Furthermore, according to this survey, illicit drug use in developed parts of the world was the eighth leading cause of disease. While the extent of use is one measure of impact, this evidence of disease burden is a telling measure of the significant effect that drug use has on individuals, their families, and societies.

Treatment

As alluded to already, SUDs are chronic diseases, which do not allow the disease to be cured, as can be done with acute illnesses such as an infection or a fracture. A consequence of understanding SUD as a chronic disease is that the process of recovery will likely involve multiple episodes of remissions and recurrences over time, just like diabetes or obesity. This requires that treatment should not simply aim at reducing or stopping substance use (i.e., abstinence) but also aim to improve the quality of life and increase the chances of remaining in recovery by addressing the underlying factors that contribute to substance use.

One of the critical messages consistent across the substances listed in this section is that addiction treatment is effective. This efficacy impacts drug use itself, as well as showing an impact on other aspects of lifestyle and health. For alcohol, opioid, and tobacco use disorders, there are numerous FDA-approved pharmacotherapies. For opioid use disorder specifically, the use of medications such as buprenorphine and methadone are life-saving treatments and are considered effective even if prescribed alone without specific psychosocial or behavioral treatments. There are numerous behavioral and psychosocial interventions with a robust evidence base supporting their efficacy. Of great interest is the growing evidence base in support of ketamine and psychedelics in treating SUD, with several randomized trials showing efficacy for alcohol, tobacco, and cocaine use disorders. However, more research on efficacy and safety is needed before they become accepted as legitimate treatment options for SUD.

Finally, the growing body of evidence now points to peer support as well as formal 12-step groups as being effective adjuncts to professional treatment. Indeed, most states now have a mechanism in place to train individuals in recovery to become “peer recovery coaches” (also referred to by other names, such as “peer recovery specialist,” “peer navigators,” or “recovery coach”). These individuals use their “lived experience” of sustained recovery to help, support, and mentor individuals in recovery and often provide services outside the scope of traditional clinical treatments. One of the distinguishing features of these peers is their explicit support for multiple pathways to recovery. For example, recovery coaches can support those who do not want to be a member of alcoholics anonymous (AA), or even those who do not want or are not ready to be completely abstinent.

Public Policy Issues Related to Substance Use Disorders

One of the most dramatic changes in recent years related to drug policy is the rapidly evolving move of cannabis from an illicit drug to a potential therapeutic (“medical marijuana”) or even to outright legalization for use by adults. This changing legal and social landscape has been rapidly occurring in the United States and other countries and varies considerably at the state level, with some states maintaining cannabis as an illegal drug and others allowing varying degrees of use. The medical use of cannabis has been promoted with little of the traditional evidence base that would provide the foundation for determining the efficacy and safety of a new medication, while the greater availability of legal use for recreational purposes has gained fairly widespread acceptance by the general public despite the known harms associated with recreational use of cannabis.

Other policy issues related to addiction treatment are the removal of the training requirement for obtaining the certification to prescribe buprenorphine, the approval of safe-consumption sites in New York City, and the growing evidence base now pointing to the potential for psychedelic drugs (i.e., psilocybin) to be used therapeutically for SUD. The emergence of psychedelics, most of them schedule 1 drugs, as potential treatment means there may be changes in their scheduling.

SUMMARY

The current drug crisis we are witnessing evolved initially from an opioid crisis but now includes all classes of drugs, such as alcohol and stimulants. This evolving crisis is also occurring in the context of a mental health crisis, made worse by the COVID-19 pandemic. This is, therefore, a call to action for all mental health providers, especially psychiatrists. SUDs remain highly prevalent disorders that exact a social as well as a personal burden on the individual, disproportionately affecting individuals with psychiatric disorders. This topic area spans aspects of public policy, neurobiology, ethics, and behavior, and addictions will likely continue to be a major topic area of concern for years to come, as they have been for thousands of years in the past. Of note, the growing body of knowledge about the biologic and environmental aspects of these disorders and how these aspects impact experimentation and problematic or pathologic use have helped to bring SUDs under a medical umbrella and moved forward prevention and treatment efforts. However, much more work needs to be done in understanding the treatment of these disorders. The following chapters in this section provide an overview of these disorders, their neurobiology and social circumstances to use, the behavioral aspects of reinforcement, the extent of substance use, and the treatment of these disorders.

FURTHER READINGS

Agboola S, McNeill A, Coleman T, Leonardi Bee J. A systematic review of the effectiveness of smoking relapse prevention interventions for abstinent smokers. *Addiction*. 2010;105:1362–1380.

Amato L, Minozzi S, Davoli M, Vecchi S, Ferri M, Mayet S. Psychosocial combined with agonist maintenance treatments versus agonist maintenance treatments alone for treatment of opioid dependence. *Cochrane Database Syst Rev*. 2004;(4):CD004147.

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.)

Australian Institute of Health and Welfare. National Drug Strategy Household Survey detailed report 2013. Drug statistics series no. 28. Cat. no. PHE 183. Canberra: AIHW. Accessed October 22, 2016. http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129549848

Ball JC, Ross A. *The Effectiveness of Methadone Maintenance Treatment*. Springer-Verlag; 1991.

Bogenschutz MP, Johnson MW. Classic hallucinogens in the treatment of addictions. *Prog Neuropsychopharmacol Biol Psychiatry*. 2016;64:250–258.

Dutra L, Stathopoulou G, Basden SL, Leyro TM, Powers MB, Otto MW. A meta-analytic review of psychosocial interventions for substance use disorders. *Am J*

Psychiatry. 2008;165(2):179–187.

Eddie D, Hoffman L, Vilsaint C, et al. Lived experience in new models of care for substance use disorder: a systematic review of peer recovery support services and recovery coaching. *Front Psychol*. 2019;10:1052. doi:10.3389/fpsyg.2019.01052

EMCDDA. *2012 Annual Report on the State of the Drugs Problem in Europe*. European Monitoring Centre for Drugs and Drug Addiction; 2012. http://www.emcdda.europa.eu/publications/annual-report/2012

Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJ; Comparative Risk Assessment Collaborating Group. Selected major risk factors and global and regional burden of disease. *Lancet*. 2002;360:1347–1360.

Falk D, Wang XQ, Liu L, et al. Percentage of subjects with no heavy drinking days: evaluation as an efficacy endpoint for alcohol clinical trials. *Alcohol Clin Exp Res*. 2010;34:2022–2034.

Gowing LR, Ali RL, White JM. Buprenorphine for the management of opioid withdrawal. *Cochrane Database Syst Rev*. 2002;(2):CD002025.

Hartman RL, Brown TL, Milavetz G, et al. Cannabis effects on driving lateral control with and without alcohol. *Drug Alcohol Depend*. 2015;154:25–37.

Mattick RP, Breen C, Kimber J, Davoli M. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. *Cochrane Database Syst Rev*. 2014;(2):CD002207.

Mattick RP, Breen C, Kimber J, Davoli M. Methadone maintenance therapy versus no opioid replacement therapy for opioid dependence. *Cochrane Database Syst Rev*. 2009;(3):CD002209.

National Institute on Alcohol Abuse and Alcoholism (NIAAA). *Helping Patients Who Drink Too Much: A Clinician's Guide*. National Institute on Alcohol Abuse and Alcoholism; 2007. NIH Publication No. 07–3769.

Rehm J, Taylor B, Room R. Global burden of disease from alcohol, illicit drugs and tobacco. *Drug Alcohol Rev*. 2006;25:503–513.

Schwartz RP, Highfield DA, Jaffe JH, et al. A randomized controlled trial of interim methadone maintenance. *Arch Gen Psychiatry*. 2006;63(1):102–109.

Stitzer ML, Vandrey R. Contingency management: utility in the treatment of drug abuse disorders. *Clin Pharmacol Ther*. 2008;83:644–647.

Strain EC, Stitzer ML, Liebson IA, Bigelow GE. Buprenorphine versus methadone in the treatment of opioid dependence: self-reports, urinalysis, and addiction severity index. *J Clin Psychopharmacol*. 1996;16:58–67.

Strain EC, Stitzer ML, Liebson IA, Bigelow GE. Methadone dose and treatment outcome. *Drug Alcohol Depend*. 1993;33:105–117.

Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: Results from the 2020 National Survey on Drug Use and Health (HHS Publication No. PEP21-07-01-003, NSDUH Series H-56). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2021. https://www.samhsa.gov/data

Vandrey R, Raber JC, Raber ME, Douglass B, Miller C, Bonn-Miller MO. Cannabinoid dose and label accuracy in edible medical cannabis products. *JAMA*. 2015;313(24):2491–2493.

Volkow ND, Swanson JM, Evins AE, et al. Effects of cannabis use on human behavior, including cognition, motivation, and psychosis: a review. *JAMA Psychiatry*. 2016;73(3):292–297.

11.2 Alcohol-Related Disorders

Marc A. Schuckit, M.D.

This chapter reviews conditions that affect 20% to 30% of middle-class men and about 15% of women seeking medical care. The typical individual with these diagnoses has a job, lives with their family, and is relatively high functioning, but is likely to die a decade or more prematurely. Almost half will develop temporary, but clinically significant, anxiety or depression, and they carry an enhanced risk for serious medical problems along with a heightened potential for adverse reactions to medications. The condition that is the focus of this chapter costs the United States almost $250 billion a year with worldwide 134 million disability-adjusted life years lost. Despite these characteristics, most physicians have received few hours of training regarding recognizing and treating these disorders, and many physicians, including psychiatrists, feel inadequate in making the diagnosis or treating these alcohol-related conditions.

Alcohol use disorders (AUDs) have special relevance to psychiatry. Alcohol is a potent drug that causes both acute and chronic changes in almost all neurochemical systems, with the result that heavy drinking can produce serious *temporary* psychological symptoms including depression, anxiety, and psychoses. Several pre-existing psychiatric diagnoses, including antisocial personality, bipolar, and schizophrenic disorders, increase the risk for later

AUDs. In light of the importance of alcohol-related conditions to psychiatric practice, this chapter offers a broad and clinically focused overview of recent developments in the field of AUDs. The following case facilitates the discussion of alcohol-related problems.

> JP is a 47-year-old ophthalmologist referred by his internist for psychiatric evaluation. The presenting complaints included depressive and anxiety symptoms, concerns by his wife that he was spending more time alone in his study in the evenings, and that there were times he appeared to not think clearly. The differential diagnosis included a major depressive episode, generalized anxiety disorder, and the early stages of a dementing process resembling Alzheimer disease. The problems had become more prominent in the past year for this man who had graduated from both college and medical school with distinction and was still married and respected by his wife of 28 years and their young adult children. It soon became apparent that heavy drinking was a major contributor to the observed problems and that, in many ways, JP was a typical middle-class person with an AUD.

JP will be used as an example of several major lessons in this chapter.

DEFINITION AND COMPARATIVE NOSOLOGY

Alcohol Use Disorders

Diagnoses in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) are intended to identify clinically significant conditions through criteria that are both reliable and valid. The two major current psychiatric diagnostic systems (DSM-5 and the *International Classification of Diseases* eleventh edition [ICD-11] developed by the World Health Organization) agree that an AUD diagnosis requires repeated impairments in multiple life areas as a result of alcohol use. In the DSM-5 criteria published in 2013, an AUD is diagnosed as the repeated presence of at least 2 of 11 major areas of life impairments related to alcohol that cluster together in a syndrome over approximately the same 12-month period. The criteria are very similar to those listed for abuse and dependence in DSM-IV, with the exception that one DSM-IV item was dropped (legal problems) and one new item was added (craving).

In ICD-11, the diagnostic guidelines for *dependence* (items that are labeled as general guidelines, not criteria) require at least two of three central features that were present over 12 or more months. The first of the three elements is impaired control over alcohol and includes craving to use the drug. The second guideline requires that alcohol became an increasing priority in life that took precedence over other activities, interests, or life interests such that alcohol use continued despite repeated problems associated with its use. The third member of the triad relates to the development of neuroadaptation to alcohol as judged by tolerance, withdrawal symptoms, and/or the use of alcohol to prevent or relieve withdrawal. ICD-11 also lists guidelines for a second syndrome, *a harmful pattern of use of alcohol,* which is viewed as a “subdependence” label indicating alcohol use that damaged a person’s own physical or mental health or contributed to harm to others. The latter could include physical harm or a mental disorder directly related to the first person’s own harmful pattern of alcohol use.

Several approaches have been tested as potential alternative AUD criteria and as efforts to divide AUDs into clinically meaningful subgroups. For example, a more severe early onset of an AUD, accompanied by criminality and problems with other drugs, has been labeled as *type 2* or *type B alcoholism.* This approach reflects the fact that most early-onset medical and psychiatric disorders have a more severe course. However, much of the prognostic significance of type 2 or type B alcoholism rests with concomitant conduct disorder and antisocial personality disorder (ASPD) in early-onset substance use disorders (SUDs). As mentioned elsewhere in this chapter, few advantages have been found for diagnostic items related to drinking patterns to a family history of AUDs.

Severity and Remission

DSM-5 asks clinicians to record the prior year severity of an AUD as *mild* (2 to 3 criteria endorsed), *moderate* (with 4 to 5 criteria met), or *severe* (with 6 or more of the 11 criteria), while ICD-11 has no formal notation of severity. Remission criteria in DSM-5 are based on the fact that the highest risk for relapse is in the first 3 to 12 months of recovery. Thus, early remission occurs if none of the 11 AUD criteria (other than craving) were experienced during the prior 3 months, and the notation of sustained remission indicates that no such problems were experienced for at least 12 months.

Although JP did not fit the erroneous stereotype of a person with an AUD as homeless and jobless, he was fairly typical of the middle-class person with this disorder. The history revealed that alcohol problems had become prominent in his mid-20s, with an age of onset of the AUD (i.e., when two or more of the AUD criteria occurred within the approximate same 12 months) of about 28 years. Problems included tolerance; spending a great deal of time drinking (when JP retired to his office at home and spent most of the evening drinking); and repeatedly exceeding preset amounts and periods of time spent drinking (e.g., setting a limit of two glasses of wine after dinner but consuming two bottles). Consistent with his relatively young age and general good health, JP had never experienced a withdrawal syndrome.

A Brief History of Alcohol and AUDs

Beverage alcohol (ethanol) is a product of fermenting fruit or grain and was probably around before humans walked the earth. Evidence of human consumption of these beverages dates back at least 12,000 years, and alcohol was used as part of ceremonies in Babylon, Greece, and the Roman Empire over 5,000 years ago when this drug was also used for pleasure, nutrition, and for its medicinal properties.

Repeated drunkenness and associated problems have been noted in almost all civilizations, and related stories are seen in the Bible. Most early consumption involved beer and wine, but distillation of beverages with higher alcohol contents, such as whiskey, dates back to about 1400 AD in Ireland. Laws actually encouraging the consumption of spirits as a way to raise tax revenues appeared in England in the 1600s.

The increased use of ethanol and the enhanced prevalence of drunkenness by the early 1800s became a public health concern and a threat to maintaining a reliable workforce, especially in the context of the industrial revolution. Later in the 1800s, England tried to reverse course and control alcohol use, and a prominent movement toward total abstinence grew in strength in both Europe and America. Following World War I, there were calls for prohibition of alcohol consumption that culminated in national drinking bans, including the 18th amendment to the U.S. constitution (the Volstead Act) in 1920. However, by this point in history, alcohol had become so ingrained in most societies that prohibition was difficult to enforce, and beverage alcohol became legal again in the United States in 1933.

Following the end of Prohibition, alcohol-related problems quickly escalated. The number of people affected and the total costs have been much higher than figures associated with any other drug.

EPIDEMIOLOGY

Prevalence of Drinking, Related Problems, and AUDs

About 90% of Americans have consumed alcohol at some time in their lives, with most beginning in their middle teens (Table 11.2–1). By the end of high school, 80% of students have imbibed alcohol, and more than 60% have been intoxicated. During any year, 70% of individuals were drinkers, with a ratio of the proportions of men versus women who consumed alcohol of approximately 1.3 men to 1.0 women. The highest drinking frequency and quantity usually occur from the late teens to the mid-20s. In 2019, 30% of men and

22% of women aged 18 and older reported binge, or heavy episodic drinking, defined as an intake on the same occasion of five or more standard drinks for men and four or more for women at least once in the prior month. The rates of drinkers, including those with heavier intake, change over time as a reflection of cohort effects and major life events. A recent example of the latter is the increase in the proportion of drinkers and their frequency of alcohol intake during the coronavirus epidemic.

Table 11.2–1.
Alcohol Epidemiology

Condition	Population (%)
Ever had a drink	90
Current drinker	60–70
Temporary problems	40+
Alcohol use disorder (AUD) lifetime	Male: 15+ Female: 10+
AUDs in psychiatric patients	30

The percentages listed here are approximations generated across studies.

Men and women with higher education and income are the ones most likely to imbibe, and, among religious denominations, Jews have among the highest proportions who consume alcohol but also have among the lowest rates of AUDs. Other ethnicities, such as the Irish, have higher rates of severe alcohol problems, but are also likely to have higher rates of abstention. Very high levels of alcohol problems are found among most, but not all, Native American and Inuit tribes, with estimates that more than 60% of men and women in some tribes have at some time had an AUD. In the United States, in 2018, the World Health Organization reported that the average person aged 15 or older consumed about 2.4 gallons of absolute alcohol in a year, a figure considerably less than the five gallons of absolute ethanol yearly intake in the 1700s. In 2010, per capita consumption ranged from 4.6 gallons in Belarus to 0.05 gallons in Saudi Arabia.

Almost half of drinkers experience temporary alcohol-related problems. One survey noted that the impaired memory for events occurring during a heavy drinking evening (i.e., an alcohol-related blackout) was reported by almost half of drinking men and women, including about 15% who had three or more such experiences in any year. Among relatively well-educated middle-class men, about 30% admitted to having driven while impaired by alcohol, and about 20% had missed school or work because of a hangover or a desire to “party.”

The lifetime risk for AUDs is approximately 15% for men and 10% for women, with 1-year prevalence rates of about 6%, figures that are applicable across most socioeconomic and educational levels. Rates among individuals coming to a hospital or visiting a physician’s office are almost double that. The peak ages of onset of an AUD are from the early 20s to about age 40. Despite multiple difficulties, most people with AUDs have jobs, families, and relatively high levels of functioning, and, thus, few fit the stereotype of a homeless, unemployed person. It is important to note that between 2001 and 2013, AUDs increased by about 10% overall, including growth of almost 85% in women, and rates increased between 90% and 100% across individuals aged 45 to 64 and those over 65, respectively. At the same time, rates of drinking among adolescents and young adults might have slightly decreased.

The course of AUDs is as predictable as most medical and psychiatric conditions, with an early onset likely to be associated with greater severity. Therefore, if an AUD begins in the midteens, it is often associated with conduct disorder, concomitant drug difficulties, and antisocial behaviors that are likely to predate the alcohol problems. At the other extreme, an onset after age 40 tends to be associated with less severe social difficulties and more subtle AUD signs and symptoms, but a greater likelihood of medical problems.

Comorbid Conditions

As discussed in more detail later in this chapter, psychiatric syndromes likely to occur at increased rates in heavy-drinking individuals fall into two general categories. The first group includes *alcohol-induced* depressive episodes, anxiety conditions, and psychoses. These resemble independent depressive (e.g., major depressive disorder), anxiety (e.g., panic disorder), and psychotic conditions (e.g., schizophrenia) but are only observed during severe persistent intoxication or withdrawal and are likely to clear spontaneously within about a month of abstinence. The second group involves psychiatric disorders that occur *independently* of alcohol intake but are associated with an increased risk for AUDs, including schizophrenia, bipolar disorder, ASPD, and post-traumatic stress disorder (PTSD).

PHARMACOLOGY AND ALCOHOL'S EFFECTS

Properties and Metabolism of Alcohol

Ethanol (beverage alcohol) is a simple molecule that is well absorbed through the mucosal lining of the digestive tract in the mouth, esophagus, and stomach. The most prominent area of uptake is in the proximal small intestine, which is also the site of absorption of many B vitamins. Ethanol rapidly enters the bloodstream and, reflecting its high solubility in water, is distributed throughout the body.

Wine, beer, and distilled spirits (e.g., whiskey, gin, and vodka) differ in their content of components other than alcohol. These *congeners*, responsible for much of the taste of the beverage, include methanol, butanol, aldehydes, phenols, tannins, lead, cobalt, and iron.

A standard alcoholic drink in the United States is defined as containing 10 to 12 g of ethanol, which is roughly the amount in 12 oz of beer (containing ~3.6% ethanol), 5 oz of table wine (~12% ethanol), and about 1.5 oz of 80-proof spirits (~40% ethanol). For an average 70-kg (155-lb) person with average body fat, one drink is likely to raise the blood alcohol level by approximately 15 to 20 mg/dL (the same as 0.015 to 0.020 g/dL), with roughly the same amount metabolized in about an hour. Faster absorption from the digestive tract occurs on an empty stomach and if alcohol is taken as a carbonated beverage.

Between 2% and 10% of the alcohol is excreted unchanged through the lungs, kidneys, or sweat, some is broken down in the lining of the esophagus and stomach (especially in men), but most is metabolized in the liver through several pathways that result in the production of acetaldehyde. Most of the process occurs through the actions of alcohol dehydrogenase (ADH) in the cell cytosol, although at high blood–alcohol levels, some alcohol is also broken down in the microsomes of the liver's smooth endoplasmic reticulum (the microsomal ethanol oxidizing system [MEOS]). The ADH process is the usual rate-limiting metabolic step, occurring relatively slowly because of the liver's need to handle the produced hydrogen ions through the actions of a cofactor that is in relatively short supply, nicotinamide adenine dinucleotide (NAD).

Acetaldehyde is destroyed by aldehyde dehydrogenase (ALDH) both in liver cell cytosol and mitochondria. This occurs rapidly so that the average person has low levels of this substance after drinking. This is fortunate because, while low levels of acetaldehyde can be both stimulating and reinforcing, at high levels, this substance can release histamines and catecholamines that contribute to rapid changes in blood pressure along with nausea and vomiting. Variants of ALDH and ADH are related to a higher reaction to alcohol and lower risk for AUDs, especially in East Asian (e.g., Japanese, Chinese, and Korean) men and women.

Alcohol as a Depressant Drug

Acutely, all intoxicants pleasurably change how a person feels, thus enhancing the drive to continue to take the drug despite potential consequences. Such drugs have divergent effects that are best understood by placing them into categories based on their most prominent effects at the usual doses taken. In this scheme, depressants are drugs that produce somnolence and decreased neuronal activity but are not powerful in attenuating pain. Thus, alcohol, benzodiazepines, barbiturates, and related drugs are considered central nervous system depressants and produce similar intoxications, are potentially lethal in overdose (especially when multiple depressants are taken together), have cross-tolerance with other depressants, and can produce physical dependence with similar withdrawal syndromes.

Neurochemical Effects of Ethanol

Alcohol has prominent effects on almost all neurochemical systems, with opposite actions during intoxication versus withdrawal. Perhaps the most potent effects are on γ-*aminobutyric acid* (GABA) complexes in the brain, especially on the GABA-A receptor ($GABA_A$), contributing to the sedating, sleep-inducing, anticonvulsant, and muscle-relaxing properties of alcohol. Ethanol also impacts glutamate-gated ionophoric receptors, especially those that bind *N-methyl-D-aspartate* (NMDA), with dampened stimulatory effects during intoxication and heightened activity during withdrawal. Alcohol acutely increases *dopamine* and its metabolites, and chronic drinking changes dopamine receptor numbers and sensitivity, with related effects on intoxication and craving, in part through changes in the pleasure centers in the ventral tegmental area of the brain. Consuming beverage alcohol also increases *serotonin* in the synapses and upregulates serotonin receptors. Lower brain levels of serotonin may be associated with a less intense response to alcohol and with consuming more alcohol per occasion. Finally, in this brief review, alcohol acutely enhances the functioning of the *opioid-related* brain systems and impacts *adenosine*, *acetylcholine*, and *cannabinoid* 1 (CB1) receptors.

Tolerance

With repeated administration, larger doses of alcohol are required to produce the same effects. This phenomenon of *tolerance* also produces an ability to tolerate higher doses of the substance through at least three processes. In *behavioral tolerance,* a person learns through practice how to perform tasks effectively while experiencing the effects of alcohol. *Pharmacokinetic tolerance* involves adaptations of alcohol-metabolizing systems to rid the body of alcohol more rapidly. Most importantly, *pharmacodynamic* or *cellular tolerance* reflects an adaptation of the nervous system so that it can function despite high blood alcohol concentrations by resisting the actions of alcohol on the cell.

Once tolerance has developed to one depressant, an individual is likely to demonstrate a similar reaction to a second drug of that class (*cross-tolerance*). Therefore, a person who has been drinking heavily, has tolerance to alcohol, and then stops drinking can be expected in the short run to require a higher dose of other depressants such as benzodiazepines for sleep induction. However, if individuals take two depressant drugs concomitantly, the effect of one drug can magnify, or potentiate, the effects of the other, with a possible lethal outcome.

FT is a 73-year-old emeritus professor at a small liberal arts college in the Midwest. He had no persistent alcohol-related problems prior to his early 60s and used to consume one or two standard drinks (scotch) before dinner. At age 62, he moved to a retirement community in Arizona, where he had more time to play golf, interact with his wife, and socialize with his new friends. In the context of a lack of work obligations and the freedom to increase his hours of socializing, FT slowly increased his alcohol intake. Subsequently, the number of standard drinks required to produce a mild buzz increased from two to four during an evening, and his daily consumption increased to five or six standard drinks most nights.

Blackout

A *blackout* is a memory impairment (*anterograde amnesia*) for the period when a person was drinking heavily but remained awake. This phenomenon is the result of the ability of high doses of any brain depressant (e.g., alcohol or benzodiazepines) to interfere with the acquisition and solidification of memory. As many as 50% of drinkers have at some time had a blackout, and repeated experiences of this effect of high doses of alcohol is a warning sign that an AUD may be present.

Sleep Impairment

Alcohol can help a person fall asleep, but after two or more drinks, the sleep pattern can be impaired. A heavy drinker is likely to awaken after several hours and have problems going back to sleep. Alcohol suppresses rapid eye movement (REM) sleep, inhibits stage 4 sleep late in the night, and is associated with frequent alternations between sleep stages (*sleep fragmentation*), with more intense and disturbing dreams late in the night as the blood alcohol levels fall. More intense forms of these problems are seen in individuals with AUDs, for whom sleep stages might not return to normal until three or more months of abstinence. More sleep problems following abstinence are associated with a greater risk for relapse.

Cerebellar Degeneration

Characterized by unsteady gait, impaired standing steadiness, and mild nystagmus, cerebellar degeneration results from combinations of effects of ethanol, acetaldehyde, and vitamin deficiencies. Treatment requires total abstinence and vitamin supplementation, although complete recovery is not usual.

Other Effects on the Central Nervous System

A thiamine deficiency, especially in the context of a pre-existing vulnerability such as a transketolase deficiency, can present as any of several neurologic syndromes now known as alcohol-induced major neurocognitive disorder, amnestic confabulatory type, persistent. Using the more widely accepted prior terminology, this can include a sixth cranial nerve palsy (Wernicke) and a severe anterograde amnesia that is out of proportion to the general level of confusion (Korsakoff). These rare conditions are discussed in greater depth in the section “Diagnosis and Clinical Features.”

Effects on the Rest of the Body

One to two drinks per day of any alcoholic beverage might decrease the risks for myocardial infarction and thrombotic stroke, probably through decreasing platelet aggregation and enhancing high-density lipoprotein (HDL) cholesterol levels. An additional potential cardioprotective action, unique to red wine, may occur through antioxidant flavonoids or the inhibition of a vasoconstrictor (endothelin-1). Low doses of alcohol may also decrease the risks for some dementias, peripheral arterial disease, and gallstones. However, any amount of alcohol may be harmful to a developing fetus, and the risks of even one drink per day are prohibitive for individuals in recovery from a moderate or severe AUD, people taking medications that may adversely interact with alcohol, and persons with any of several medical or psychiatric syndromes. Consuming more than three drinks a day increases low-density lipoprotein (LDL) cholesterol, triglycerides, and blood pressure, thus increasing the risk of cardiac disorders and cancers.

Peripheral Neuropathy. Approximately 10% of people with AUDs develop a deterioration of functioning of nerves in the hands and feet, called *peripheral neuropathy.* The symptoms include numbness of the hands and feet, often bilateral, frequently accompanied by tingling and paresthesias. Although the condition is usually relatively mild and often improves with abstinence, in some individuals, the pain and numbness may be permanent.

Gastrointestinal Problems. The gastrointestinal (GI) system can be severely affected by heavy drinking, including an acute and potentially severe inflammation of the esophagus or stomach, often accompanied by vomiting and bleeding. If vomiting occurs in the presence of dilated esophageal veins, as seen with cirrhosis, it can induce a potentially lethal blood loss.

The liver and the pancreas are also especially vulnerable to alcohol. In the liver, frequent heavy drinking results in the accumulation of fats and proteins in the cells, producing a reversible swelling of the liver, often described as a *fatty liver.* Inflammation of liver cells accompanied by increases in liver function tests is called *alcoholic hepatitis*, which can lead to deposition of excessive amounts of hyalin and collagen near blood vessels in the liver, an early stage of *cirrhosis*, with the latter seen in approximately 15% of patients with AUDs. As damage progresses and liver scar tissue increases, the normal flow of blood through the liver is impaired, dilated veins (or varices) develop from the increased abdominal venous pressure, and fluid seeps from the liver capsule, accumulating in the abdomen as *ascites*. Subsequent liver failure can cause a cognitive impairment known as hepatic encephalopathy.

Perhaps 10% of people with AUDs develop an inflammation of the pancreas or *acute pancreatitis*. This can lead to a chronic irreversible condition with associated impaired sugar metabolism (a form of diabetes).

Cardiovascular Problems. Heavy alcohol intake and associated withdrawal symptoms can increase blood pressure and elevate both LDL cholesterol and triglycerides, enhancing the risk for myocardial infarction and thrombosis. At high doses, alcohol is a striated-muscle toxin with a resulting deterioration in the heart muscle that manifests as beating irregularities and signs of heart failure (*alcoholic cardiomyopathy*). The risks for cardiovascular disease are likely to be observed at even lower alcohol doses in East Asian individuals who have the mutation in ALDH described above in the section on alcohol metabolism. Thus, the leading cause of early deaths associated with AUDs is heart disease.

Blood-Producing Systems. Consumption of four to eight drinks or more per day decreases the production of white blood cells and impairs the ability of those cells to migrate to sites of infection. Such drinking can also affect the stem cells that produce the red blood components, increasing the average size of red cells (the mean corpuscular volume [MCV]), and can impair the production of blood platelets.

Cancer. AUDs are associated with high rates of most cancers, especially tumors of the head, neck, esophagus, stomach, liver, colon, lungs, and breast. The association with cancer probably reflects alcohol-related immune system suppression and the direct effects of ethanol and acetaldehyde on mucosal membranes. The heightened rates of malignant tumors remain significant even when the impact of smoking and poor nutrition is considered, and cancer is the second cause of premature death in individuals with AUDs.

Fetal Alcohol Effects

Alcohol and acetaldehyde have deleterious effects on the developing fetus. Both substances cross the placenta and, in high enough doses, can produce fetal death and spontaneous abortion. At least 5% of surviving infants of heavy-drinking mothers evidence mixtures of a *fetal alcohol syndrome* that can include low IQ, a small head, low birth weight, facial abnormalities (e.g., flat bridge of the nose, absent philtrum, and epicanthal eye folds), an atrial septal heart defect, and syndactyly. More subtle offspring behavioral effects of drinking during pregnancy have been reported with even lower alcohol doses. None of these problems are likely to be reversible, and most cognitive defects remain throughout life. Additional less obvious problems are part of a range of *fetal alcohol spectrum disorders* (FASDs), including mild to moderate cognitive difficulties (e.g., impaired executive functioning in attention and planning, problems in letter and word fluency, difficulties with impulse control, etc.), and may carry an enhanced risk for SUDs.

Other Problems

High doses of alcohol adversely affect almost all body systems, cutting a decade or more off the lifespan even for people with higher socioeconomic and educational status. Additional body effects related to AUDs include testicular atrophy; bone fractures; a host of eye effects, including cataracts; dental difficulties; skeletal muscle wasting; and greater risk for accidents.

ETIOLOGY

Many factors affect the decision to drink, the development of temporary alcohol-related difficulties, and the development of AUDs. The initiation of drinking depends largely on social, religious, and personality characteristics, although genetic influences linked to impulsivity and risk-taking also contribute. However, the factors that influence the decision to drink or those that contribute to temporary problems might be a bit different from those that add to the risk for AUDs.

Genetic factors are responsible for about 50% to 60% of the proportion of risk for AUDs and operate through several intermediate characteristics that interact with the environment. While each risk component is discussed separately in this chapter, these divisions are heuristic, as it is the combination of psychological, sociocultural, biologic, and other factors that are responsible for the development of severe, repetitive alcohol-related life problems.

Psychological and Sociocultural Theories

Plentiful data support a relationship between *stress* and drinking behaviors, but the strength of the interaction and the directions of the correlations are complex. Retrospective studies focusing on prominent adverse life events often report that higher stress is associated with heavier drinking and related problems. However, prospective studies often find less impact of stress during any one day on alcohol use that same day or in the subsequent 48 hours. In addition, these relationships are bidirectional, with drinking on day 1 also associated with more stress on day 2. At higher doses and especially at falling blood alcohol levels, objective measures of tension in muscles indicate that muscle tension can actually increase. Thus, tension-reducing effects of alcohol may be modest and are most obvious in light-to-moderate drinkers, and stress may play a moderate role in causing AUDs.

A person's *expectations* of the effects of alcohol also play a role in patterns of alcohol use. Positive expectations of the impact of excessive drinking on social interactions, alleviating sadness and anxiety, and sexual performance are associated with higher probabilities of heavier drinking. Similarly, the more a person expects that alcohol will interfere with relationships, sexual performance, and other negative outcomes, the less the risk for heavy alcohol intake.

A third characteristic closely linked to attitudes toward alcohol and subsequent drinking patterns is the *perceived pattern of drinking among peers,* where individuals tend to overestimate the alcohol consumption of friends. Thus, they exaggerate the amount of drinking that is acceptable, healthy, and safe, and may be more likely to engage in heavier drinking themselves.

Additional psychological and sociocultural factors related to heavy drinking include the neighborhood in which one lives and the ease of availability of alcohol, a person's income and the money available to buy alcohol, and their education and religious beliefs. Religious beliefs and practices affect alcohol intake through moderating the effects of other characteristics, with less likelihood that poverty, for example, predicts heavier drinking among individuals with closer religious affiliations.

Sociocultural theories are sometimes based on extrapolations from social groups that have high and low rates of AUDs. For example, Jews tend to introduce young children to modest drinking in religious rituals, eschew drunkenness, and also have low rates of AUDs. However, other groups have similar practices, such as French Catholics, but have higher rates of AUDs. Recent studies have found little evidence that introducing children to modest drinking in the home impacts the risks for heavy drinking and alcohol problems. So, cross-cultural comparisons are useful in generating theories, but it can be difficult to establish key cultural practices contributing to drinking-related disorders.

Psychodynamic Theories

Classic psychoanalytical theory hypothesizes that at least some people with AUDs may have become fixated at the oral stage of development and use alcohol to relieve their frustrations by taking the substance by mouth. However, hypotheses regarding arrested phases of psychosexual development, although heuristically useful, have had little effect on the usual treatment or prevention approaches for AUDs and are not the focus of extensive ongoing research.

Similarly, there is little support for an “addictive personality” that might be associated with lack of control of intake of most substances and some foods. Although pathologic scores on personality tests are often seen during intoxication, withdrawal, and early recovery, many of these characteristics do not predate AUDs, and most disappear with abstinence.

Background on Biologic Contributors

The genetic contribution to the AUD risk involves many different genetically influenced characteristics, and there is no known single alcoholism gene. Using a parallel example from medicine, myocardial infarctions are genetically influenced, but in some families, the genes relate to high blood pressure; in others, the heightened risk is related to genes contributing to high triglycerides; yet others develop heart attacks through a genetically influenced vulnerability toward cardiomyopathy. Each genetic mechanism is moderated by environmental characteristics such as diet, exercise, and smoking. The following sections highlight data supporting the importance of genetic influences in AUDs, and describe biologic mediators (Table 11.2–2).

Importance of Genetic Influences in AUDs

Three lines of evidence support genetic influences in AUDs. First, there is a three- to fourfold increased risk for severe alcohol problems in close relatives of individuals with AUDs, with rates increasing with a greater number of such relatives, a greater severity of their AUDs, and a closer genetic relationship. Second, the rate of similarity for severe alcohol-related problems is higher in identical twins of persons with an AUD with whom they share 100% of their genes than in fraternal twins who share only 50%. Third, the risks for AUDs in the offspring of AUD parents are still high even if they were adopted away and knew nothing of the problems in their biologic family.

Table 11.2–2.
Data Supporting Genetic Influences in AUDs

Close family members of individuals with an AUD have a fourfold increased risk.

The identical twin of a person with an AUD is at higher AUD risk than is a fraternal twin.

Adopted-away children of individuals with AUDs have a fourfold increased AUD risk, even if not raised by parents with alcohol problems.

Biologic Mediators of the AUD Risk

As shown in Table 11.2–3, a series of independent characteristics impact the alcoholism risk.

Alcohol Metabolizing Enzymes. Variations in genes on chromosome 12 that control ALDH and genes on chromosome 4 related to ADH can decrease the alcoholism risk. The most potent polymorphism occurs in the mitochondrial low Km ALDH2 gene reflecting the *ALDH2*2* genotype that is seen in about 40% of Japanese, Chinese, and Korean individuals. An *ALDH2*2*, *2*2* homozygote inherits a disulfiram (or Antabuse)-like aversive reaction to alcohol because their enzyme cannot efficiently metabolize acetaldehyde, with resulting nausea, vomiting, and unstable blood pressure after one or two drinks and an AUD risk close to zero. Heterozygotes for the ALDH mutation (e.g., with *ALDH2*2, 2*1* alleles) have a mild to modest facial flush after drinking, a higher heart rate, and a more intense (although not necessarily aversive) response to alcohol that contributes to a modest decreased risk for AUDs. Variations in ADH genes (e.g., ADH1B and ADH1C) are more prevalent among Asian, Black, and Jewish individuals who experience a slightly increased rate of breakdown of alcohol with a possible increase in acetaldehyde and a modest protective effect for AUDs. These gene variations do not affect the risk for other SUDs.

Impulsivity and Disinhibition. A second genetically influenced mechanism increases the risk for both AUDs and SUDs through genes that affect impulsivity, sensation seeking, and disinhibition. These characteristics are more prevalent in individuals with AUDs, are observed in a substantial minority of their children, and reflect the action of a range of genes, including those coding for GABA_A and CHRM2 receptors. Extreme impulsivity and disinhibition characterize ASPD, where the risk for AUDs and SUDs is about 80%.

Table 11.2–3.
Some Genetically Influenced Characteristics that Affect the AUD Risk

Alcohol-metabolizing enzymes
Impulsivity and disinhibition
Some psychiatric disorders (e.g., schizophrenia and bipolar disorders)
A low level of response (low LR) to alcohol

Independent Psychiatric Disorders. A third genetically influenced mechanism that increases both AUD and SUD risks operates through genes that enhance the vulnerability to several psychiatric disorders. In addition to ASPD, two genetically influenced psychiatric conditions closely tied to AUDs and SUDs are schizophrenia and bipolar disorders. These relationships might reflect the poor judgment associated with these major psychiatric disorders and the impulsivity associated with mania. The link might also relate to an overlap of genes contributing to the predisposition toward substance-related problems and those that impact on schizophrenia and bipolar disorder. Some of the relationship might also occur as an individual attempts to moderate their psychiatric symptoms or the side effects of medications through alcohol or illicit drugs.

The Level of Response (or Sensitivity) to Alcohol. A fourth genetically influenced risk factor for AUDs (but not SUDs) involves how a person responds to alcohol. This includes enhanced risks for people who experience greater stimulation at rising blood alcohol levels and those who demonstrate a low level of response (LR) at peak and falling blood levels regarding feelings of intoxication and sedating effects of alcohol. The most thoroughly studied of these characteristics is the low LR, where the need for higher doses to achieve alcohol's effects contributes to larger quantities consumed per occasion. This subsequently affects the choice of heavy-drinking friends, enhances positive expectations of the effects of heavy drinking, and is associated with the use of alcohol to deal with stress. Several genes that may contribute to the low LR have been identified, and functional magnetic resonance imaging (fMRI) studies have highlighted a possible cognitive inefficiency that might make it more difficult for some people to recognize the effects of low alcohol doses. The data indicating that this characteristic is genetically influenced and the identification of some environmental mediators of the effect of a low LR on heavy drinking are being used to develop programs to prevent heavy drinking and AUDs in individuals with low LRs before they develop.

DIAGNOSIS AND CLINICAL FEATURES

AUDs are among the most common of the serious life-threatening behavioral or psychiatric syndromes, and, unfortunately, these repetitive patterns of problems associated with heavier drinking are often not obvious to the clinician. The average man or woman with an AUD does not fit the public stereotype as these individuals are likely to be neatly dressed, show no signs of severe alcohol withdrawal, and have a job and a family, but present with a variety of physical conditions or temporary but potentially severe psychiatric complaints. Thus, the clinician must gather a history of alcohol-related life problems from every patient and, if possible, an additional informant and try to determine whether alcohol has contributed to the presenting complaint. Table 11.2–4 lists the alcohol-related disorders in DSM-5.

Diagnosing Alcohol-Related Disorders

This generic term includes AUDs (previously abuse and dependence in DSM-IV), as well as alcohol intoxication, withdrawal, and alcohol-induced conditions.

Table 11.2–4.
DSM-5 Alcohol-Related Disorders and Corresponding ICD-11 Disorders

DSM-5	ICD-11 Corollary
Alcohol use disorders	**Alcohol dependence** **Harmful pattern of alcohol use**
Alcohol-induced disorders	
Alcohol intoxication	**Alcohol intoxication**
Alcohol intoxication delirium	Alcohol-induced delirium
Alcohol withdrawal	**Alcohol withdrawal**
Specify if: With perceptual disturbances	
Alcohol withdrawal delirium	**Alcohol-induced delirium**
Alcohol-induced persisting neurocognitive disorder	
Alcohol-induced major neurocognitive disorder, amnestic type, persisting	
Alcohol-induced psychotic disorder	**Alcohol-induced psychotic disorder**
Specify if: With onset during intoxication/with onset during withdrawal	
Alcohol-induced depressive disorder	**Alcohol-induced mood disorder**
Specify if: With onset during intoxication/with onset during withdrawal	
Alcohol-induced anxiety disorder	**Alcohol-induced anxiety disorder**
Specify if: With onset during intoxication/with onset during withdrawal	
Alcohol-induced bipolar disorder	**No ICD-11 equivalent**

Specify if: With onset during intoxication/with onset during withdrawal	
Alcohol-induced sexual dysfunction	**Alcohol-induced sexual dysfunction**
Specify if: With onset during intoxication or during withdrawal	
Alcohol-induced sleep disorder	**Alcohol-induced sleep–wake disorder**
Specify if: With onset during intoxication/with onset during withdrawal	
Unspecified alcohol-related disorder	

Data from DSM-5, *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition and *ICD-11, International Classification of Diseases*, 11th edition.

Alcohol Use Disorders

The key to a diagnosis of an AUD is heavy and repetitive use of alcohol to the point of developing recurrent problems. Quantities and frequencies of alcohol intake, while useful, are by themselves not sufficient to establish the diagnosis because of the large variation in the amounts of alcohol required for problems across men versus women, larger versus smaller individuals, and older versus younger people. Repetitive problems indicate that a person keeps using alcohol despite problems, which, in turn, predicts the probability that a return to drinking is likely to be associated with the re-establishment of problems.

As noted above, there are 11 elements in the DSM-5 definition of AUDs. Ten of these items are the same as those listed in DSM-III-R and DSM-IV. The 11th item, craving, describes a strong drive to drink that occurs despite the problems caused by alcohol.

JP's referring clinician was concerned that little progress was being made with the patient's mood swings and somatic complaints. Therapy, primarily using a cognitive and behavioral approach, had been instituted 6 months previously for the depressive symptoms and irritability that were present most days for the prior year. JP had recently undergone a routine physical examination at work and was noted to have modest hypertension (blood pressure of 145/95) along with blood test results that included an MCV of 92.5 μm^3 and a γ-glutamyltransferase (GGT) level of 53 U/L.

On interview, the patient admitted that his wife had been complaining for several years about his drinking pattern. She had emphasized that in the mornings after drinking, he was more irritable than usual and complained about his restless sleep. Dr P had recognized potential alcohol problems and tried on several occasions to cut back on his drinking. These efforts usually resulted in abstinence of between 2 and 4 weeks, followed by several months of setting limits on his intake, but these always gave way to increased drinking during celebrations, business trips, vacations, or to "deal with" stress. Once ad-lib drinking began, he had to increase the amount consumed per day to achieve the desired effects.

Alcohol Intoxication

The DSM-5 criteria for alcohol intoxication require recent ingestion of ethanol, associated impairment, and at least one of six signs or symptoms. In ICD-11, intoxication is a temporary condition after consuming alcohol that is associated with impaired cognition, coordination, mood, and/or level of consciousness.

As a conservative approach to identifying blood levels that are likely to have major effects on driving abilities, the legal definition of intoxication in the United States requires a blood concentration of 80 mg of ethanol per dL of blood (mg/dL), which is the same as 0.08 g/dL. Many other countries define legal intoxication regarding driving as 40 mg/dL, and the limits in Poland are 20 mg/dL. For most people, blood levels as low as 20 to 30 mg/dL are associated with some behavioral changes, a slowing in motor performance, and a decrease in the ability to think clearly, as shown in Table 11.2–5. Blood concentrations between 80 and 200 mg/dL produce more severe problems, including severely impaired coordination (ataxia), lability of mood, and progressively greater levels of cognitive deterioration. Blood alcohol concentrations of 200 to 300 mg/dL produce severe slurring of speech and en bloc *anterograde amnesia* (*alcoholic blackouts*) regarding much of the evening. Further increases in blood levels result in the first level of anesthesia, and the nontolerant person who reaches 400 mg/dL risks respiratory failure, coma, and death.

Table 11.2–5.
Impairment Likely to Be Observed at Different Blood Alcohol Concentrations

Level	Likely Impairment
20–30 mg/dL (0.02–0.03 g/dL)	Slowed motor performance and decreased thinking ability
30–80 mg/dL	Increases in motor and cognitive problems
80–200 mg/dL	Increases in incoordination and judgment errors Mood lability Deterioration in cognition
200–300 mg/dL	Nystagmus, marked slurring of speech, and alcoholic blackouts
>300 mg/dL	Impaired vital signs and possible death

Table 11.2–6.
Key Elements of DSM-5 Criteria for Alcohol Withdrawal[a]

A. Recent decreasing or stopping regular heavy drinking
B. Accompanied by the development of at least two of the following:
   1. Increased pulse, blood pressure, and/or sweating
   2. Tremor of the hands
   3. Problems sleeping
   4. Gastrointestinal upset
   5. Hallucinations
   6. Inability to sit still
   7. Nervousness
   8. Seizures

[a] A short version of the criteria is offered here. See DSM-5 for complete criteria.

Alcohol Withdrawal

If a person has been drinking heavily over a prolonged period, a rapid decrease in blood alcohol levels can produce a *withdrawal syndrome* characterized by symptoms that are primarily the opposite of what was initially experienced with intoxication. Due to the effects of tolerance described earlier, with chronic exposure to a CNS depressant, the brain adapts by downregulating the inhibitory systems while also upregulating the excitatory system. When the depressant is suddenly removed, the excess excitatory system becomes manifest as withdrawal. Here, withdrawal includes a coarse tremor of the hands, insomnia, anxiety, and signs of autonomic hyperactivity that can include increased blood pressure, heart rate, body temperature, and respiratory rate, as described in Table 11.2–6. In ICD-11, the description of alcohol withdrawal is similar.

Withdrawal phenomena usually begin within about 8 hours of abstinence, peak on the second or third day, and diminish by the fourth or fifth day. The symptoms persist in a milder form for as long as 3 to 6 months as part of a protracted withdrawal syndrome, the symptoms of which might contribute to relapse.

FT, the retired professor discussed elsewhere in this chapter, went on to experience alcohol withdrawal delirium. He was in relatively good health when he entered the hospital for an elective hernia repair. The brief history in the chart offered no comments regarding his drinking pattern and made no mention of his GGT value of 55 U/L and MCV of 93.5 μm^3. Eight hours postsurgery, the nursing staff noted an increase in his pulse to 110, a blood pressure of 150/100, prominent diaphoresis, and a bilateral hand tremor just before the patient had a grand mal seizure. He awoke agitated and disoriented. Re-evaluation of the history documented an AUD with a consumption of approximately six standard drinks per night. Over the following 4 days, the patient's autonomic nervous system dysfunction decreased, and his cognitive impairment disappeared.

Alcohol Withdrawal Delirium

Less than 5% of alcohol intoxications and withdrawals are accompanied by severe cognitive impairments known as alcohol intoxication delirium and alcohol withdrawal delirium. These conditions require that the patient meet criteria for both delirium and DSM-5 alcohol intoxication or withdrawal.

Mr T is a 64-year-old lawyer with a 35-year history of an AUD. Alcohol-related problems included tolerance, spending a great deal of time using alcohol (despite a relatively successful career), repeatedly giving up important business and family events because he was too intoxicated or hung over, and significant interference with his relationship with his wife. She complained about his sarcastic wit and lack of care about her interests while drinking heavily.

Mr T had been able to stop drinking on multiple occasions in the past, at which times he experienced a tremor, anxiety, and problems sleeping for several days. He typically self-medicated these symptoms with five or more 10-mg capsules of chlordiazepoxide per day, using pills that had been prescribed for him by his general practitioner in response to his complaints of anxiety and insomnia. Recently, Mr T was diagnosed with moderate hypertension and adult-onset diabetes, and his family practitioner urged him to stop drinking. In recognition of his age and his associated medical problems, it was believed that the current withdrawal would be best treated in a medical setting.

An examination confirmed the withdrawal syndrome along with a mildly decreased hematocrit and platelet count. The evaluation, carried out about 12 hours after his most recent drink, revealed the smell of alcohol on his breath and a blood alcohol level of 50 mg/dL. The patient demonstrated a prominent tremor of both hands along with a pulse of 110 beats per minute, a respiratory rate of 25 breaths per minute, and a blood pressure of 150/96. Mr T complained of feeling agitated and noted that while he was very tired, he was unable to sleep but was otherwise alert and oriented. These symptoms improved greatly by day 5 of abstinence, with mild levels of autonomic overactivity and insomnia gradually disappearing over the next 2 months.

Alcohol-Induced Psychiatric Disorders

Alcohol-induced psychiatric disorders are clinically meaningful (e.g., interfere with functioning and/or cause severe distress) psychiatric syndromes (e.g., depression) that resemble independent psychiatric disorders (e.g., major depressive disorders) and are in excess of symptoms usually associated with alcohol intoxication or withdrawal. They develop during, or soon after, intoxication or withdrawal from alcohol and are likely to clear within a month of abstinence. Note that independent psychiatric disorders are described in the section below dealing with the differential diagnoses for individuals presenting with an AUD.

For men and women presenting with psychiatric symptoms (e.g., anxiety, depression, or psychoses) with concomitant alcohol-related problems, the first step in establishing a diagnosis of an alcohol-induced disorder is to obtain a chronologic history of the conditions. The focus should be on full psychiatric disorders, not subsyndromal sadness or nervousness, because the full syndromes have the clearest prognostic implications. In this process, a timeline can be useful by noting: (1) the age of onset of alcohol problems severe and repetitive enough for a diagnosis of an AUD; (2) periods of abstinence of a month or more since the AUD onset to determine if a full psychiatric syndrome ever occurred outside the context of problematic alcohol use; and (3) the ages when the patient met criteria for the major psychiatric disorder. If a review of the timeline reveals no evidence that the additional psychiatric syndromes either clearly antedated the heavy alcohol intake or persisted for a month or more after abstinence, an AUD is the major disorder. The additional psychiatric syndromes are likely to be temporary conditions that occurred during alcohol intoxication or withdrawal (i.e., are alcohol induced). DSM-5 lists substance-induced conditions within the clinically relevant sections (e.g., within depressive, anxiety, and psychotic disorders sections). Independent psychiatric conditions in individuals with AUDs are discussed elsewhere in this chapter, including in the section on Differential Diagnoses.

Alcohol-Induced Depressive Disorder. In the context of heavy and repetitive intake of any brain depressant, symptoms of severe depression are common and may be labeled as an *alcohol-induced depressive disorder* (see Table 11.2–4). In ICD-11, these conditions are labeled as alcohol-induced mood disorders. This condition includes the same symptoms observed in major depressive disorders, but the alcohol-induced intense sadness markedly improves within several days to 1 month of abstinence without antidepressant medications. Eighty percent of people with AUDs report intense depressions, including about 40% who were depressed for ≥2 weeks at a time, but only about 15% of such men and women have ever had a depression that meets the criteria for a major depressive disorder when they have not been drinking heavily. Although an alcohol-induced bipolar diagnosis is listed in DSM-5, there is no convincing evidence that heavy drinking can mimic a manic episode.

A consultation was requested on a 42-year-old woman with an AUD who complained of persisting severe depressive symptoms despite days of abstinence. In the initial stage of the interview, she stated that she had “always been depressed” and believed that she “drank to cope with the depressive symptoms.” Her current complaint included a prominent sadness that had persisted for several weeks, difficulties concentrating, initial and terminal insomnia, and a feeling of hopelessness and guilt. In an effort to distinguish between an alcohol-induced mood disorder and an independent major depressive episode, a timeline-based history was obtained. This focused on the age of onset of a DSM-5 AUD, periods of abstinence that extended for at least a month since the onset of the AUD (periods where independent depressions might have been observed), and the ages of occurrence of clear major depressive episodes lasting 2 or more weeks. Despite this patient’s original complaints, there was no evidence of a major depressive episode before her mid-20s when her AUD began, and during a 1-year period of abstinence related to the gestation and neonatal period of her son, she was euthymic. A provisional diagnosis of an alcohol-induced mood disorder was made. The patient was offered education, reassurance, and cognitive behavior therapy (CBT) to help her deal with the depressive symptoms, but no antidepressant medications were prescribed. The depressive symptoms remained at their original intensity for several additional days and then began to improve. By 3 weeks of abstinence, the patient no longer met criteria for a major depressive episode, although she demonstrated mild mood swings for several additional weeks. This case is a fairly typical example of an alcohol-induced depressive disorder in an individual with an AUD.

Alcohol-Induced Anxiety Disorders. Anxiety symptoms fulfilling the diagnostic criteria for alcohol-induced anxiety disorders in DSM-5 and ICD11 are common in the context of acute and protracted alcohol withdrawal. Almost 80% of people with AUDs report prominent anxiety during an acute withdrawal episode, and their complaints can be intense enough for the clinician to consider diagnosing a panic disorder, generalized anxiety disorders, social phobia, and other anxiety conditions. However, when psychological or physiologic symptoms of anxiety are observed in people with AUDs only in the context of heavy drinking or within the first month or so of abstinence, the symptoms are likely to diminish and subsequently disappear with time alone. Only a few pre-existing anxiety or stress disorders (e.g., panic disorder, social phobia, and PTSD) have been clearly shown to modestly increase the risk for later AUDs.

A 48-year-old woman was referred for evaluation and treatment of her recent onset of panic attacks. These episodes occurred two to three times per week over the past 6 months, with each lasting typically between 10 and 20 minutes. Panic symptoms occurred independently of levels of stress and could not be explained by the side effects of current medications or by medical conditions. The workup included an evaluation of her laboratory test values, which revealed a carbohydrate-deficient transferrin (CDT) level of 3.5%, a uric acid level of 7.1 mg, and a GGT value of 47. All other blood tests were within normal limits.

The atypical late age of onset of panic attacks, along with the blood results, encouraged the clinician to probe further regarding the pattern of alcohol-related life problems with both the patient and, separately, her spouse. This documented a history of an AUD with an onset at approximately 35 years of age, with no evidence of a panic disorder before that time. The patient denied having repetitive panic attacks beyond a month of abstinence during her frequent periods of nondrinking that often lasted for 3 or 4 months. A working diagnosis of an AUD with an alcohol-induced anxiety disorder characterized by panic attacks was made, and the patient was encouraged to abstain. She was offered CBT to help decrease feelings of panic and anxiety and was appropriately treated for moderate withdrawal symptoms. Over the subsequent 3 weeks after a taper of benzodiazepines used for the withdrawal, the panic symptoms diminished in intensity and subsequently disappeared.

Alcohol-Induced Psychotic Disorders. About 3% of people with AUDs experience these conditions, described in DSM-5 and ICD 11 as auditory hallucinations and/or paranoid delusions in a clear sensorium that develop in the context of heavy drinking or withdrawal, typically referred to as alcoholic hallucinosis, as coined by Eugene Bleuler. This rate is lower than psychotic conditions noted during intake of high doses of cannabis or stimulants (i.e., all forms of cocaine and amphetamines). Many clinicians exclude from this label psychotic symptoms that only occur during the course of a delirium or dementia. The symptoms of alcohol-induced psychoses resemble those seen in schizophrenia, but when the psychotic features develop only in the context of alcohol problems, they are likely to clear spontaneously within a few days to a month of abstinence. It is estimated that <5% of these psychoses occur in individuals who later develop schizophrenia.

Alcohol-Induced Major Neurocognitive Disorders, Amnestic Confabulatory Type, Persistent (Wernicke–Korsakoff Disorder)

One of the more dramatic alcohol-related CNS syndromes is the rare DSM-5 diagnosis of what in the past was referred to as an alcohol-induced persisting amnestic disorder, which is the result of a relatively severe CNS deficiency of thiamine. Similar criteria are offered in ICD-10 as an amnesic syndrome. Some people are at higher risk than others for this condition because of a genetically influenced transketolase deficiency. The syndrome has been historically subdivided into (1) Wernicke encephalopathy, with prominent ataxia and palsy of the sixth cranial nerve, a condition that tends to reverse fairly rapidly with vitamin supplementation; and (2) Korsakoff syndrome, which is permanent in at least a partial form in about 50% of the people affected. Korsakoff syndrome is characterized by a pronounced anterograde and retrograde amnesia along with potential impairment in visuospatial, abstract, and other types of learning. In most cases, the level of loss of recent memory is out of proportion to the global level of cognitive impairment. The patients who recover respond to 500 mg of IV thiamine, two to three times a day for 3 to 5 days, or 100 mg of oral thiamine a day that is usually administered for many months. High-dose IV thiamine is currently considered superior to PO supplementation in preventing and treating Wernicke, because the high serum concentrations are needed to drive thiamine across the blood–brain barrier.

Alcohol-Induced Major Neurocognitive Disorders (Persisting Dementias)

A poorly studied, heterogeneous long-term cognitive problem that can develop in the course of AUDs was called *alcohol-induced persisting dementia* in DSM-IV and ICD-10. Global decreases in intellectual functioning, cognitive abilities, and memory are observed, but recent memory difficulties are no more prominent than other elements of this condition. Brain functioning tends to improve with abstinence, but perhaps half of patients have long-term and even permanent disabilities. More than 50% of these patients have increased brain ventricle sizes and shrinkage of the cerebral sulci, although these changes are usually reversible with abstinence.

IDENTIFICATION IN CLINICAL SETTINGS AND RELATED LABORATORY TESTS

To establish an AUD diagnosis, a history of alcohol-related life problems must be obtained from the patient and, when available, an additional reliable informant. Because any patient, regardless of their education and income, might have an AUD, everyone should be screened for histories of alcohol-related accidents, interpersonal difficulties, problems at work, and encounters with the law. When a problem is apparent, the clinician should ask more about the situation and about the complaints voiced by others. If an AUD is probable, the diagnostic criteria and drinking history should be reviewed with the patient.

Several simple questionnaires can help screen for problems. The Alcohol Use Disorders Identification Test (AUDIT, Table 11.2–7) uses 10 items to review the pattern of life problems related to alcohol. More simple instruments, such as the CAGE ([need to] *C*ut down [on drinking], *A*nnoyance when asked about drinking, *G*uilt [about drinking], [need for] *E*ye-openers the morning after heavy drinking), are too short to be optimally sensitive or specific. Although the AUDIT is useful, such questionnaires do not diagnose an AUD but only highlight individuals who are especially appropriate for a more detailed clinical interview.

Identifying heavier drinkers vulnerable to AUDs is facilitated by the several blood tests noted in Table 11.2–8. These *state markers of heavy drinking* reflect physiologic alterations likely to occur after ingesting about five or more drinks a day over several weeks. One marker with a sensitivity and specificity of about 60% is a level of at least 35 U/L of GGT, an amino acid–transporting enzyme that is temporarily induced by repetitive heavy drinking. Enzyme levels are likely to return to normal after 2 to 4 weeks of abstinence, after which further increases over 20% are useful in identifying a return to drinking. Similar results are seen for the deglycosylated form of the protein transferrin, known as CDT. A value over 3% has sensitivities and specificities of 60% for the consumption of five or more drinks per day for 1 to 2 weeks. With a biologic half-life of about 16 days, this test is also useful in monitoring abstinence. The combination of GGT and CDT is better than either test alone for identification of heavy drinking and monitoring abstinence.

Table 11.2–7.

Key Elements of the Alcohol Use Disorders Identification Test (AUDIT)

Item	5-Point Scale (Least to Most)
1. How often do you have a drink containing alcohol?	Never (score 0) to 4+ per week (score 4)
2. How many drinks containing alcohol do you have on a typical day?	1 or 2 (0) to 10+ (4)
3. How often do you have six or more drinks on one occasion?	Never (0) to daily or almost daily (4)
4. How often during the last year have you found that you were not able to stop drinking once you had started?	Never (0) to daily or almost daily (4)
5. How often during the last year have you failed to do what was normally expected from you because of drinking?	Never (0) to daily or almost daily (4)
6. How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session?	Never (0) to daily or almost daily (4)
7. How often during the last year have you had a feeling of guilt or remorse after drinking?	Never (0) to daily or almost daily (4)
8. How often during the last year have you been unable to remember what happened the night before because you had been drinking?	Never (0) to daily or almost daily (4)

9. Have you or someone else been injured as a result of your drinking?	No (0) to yes, during the last year (4)
10. Has a relative, friend, doctor, or other health worker been concerned about your drinking or suggested that you should cut down?	No (0) to yes, during the last year (4)

The AUDIT is scored by simply summing the values associated with the endorsed response.

Adapted with permission from Babor TF, Higgins-Biddle JC, Saunders JB, Monterio MG. *AUDIT: The Alcohol Use Disorders Identification Test Guidelines for Use in Primary Care*, 2nd ed. Geneva, World Health Organization (WHO); 2001. https://www.who.int/publications/i/item/WHO-MSD-MSB-01.6a, accessed 3/22/23. WHO does not endorse any specific companies, products or services.

An MCV over 91 μm^3 has a sensitivity of about 50% in identifying heavy drinking, but the 120-day lifespan of red cells does not allow MCV to be useful as an indicator of a return to drinking. Other indicators that can be helpful in identifying patients who are regularly consuming heavy doses of alcohol include high normal values of uric acid (e.g., >6.4 mg/dL) and elevations in the usual liver function tests. Useful physical findings include modest elevations in blood pressure, frequent bruising, cancer of the head, neck, or upper digestive tract, an enlarged liver, evidence of cirrhosis, and pancreatitis.

Table 11.2–8.
State Markers of Heavy Drinking Useful in Screening for AUDs

Test	Relevant Range of Results
γ-glutamyltransferase (GGT)	>35.0 U/L
Carbohydrate-deficient transferrin (CDT)	>3.0%
Mean corpuscular volume (MCV)	$>91.0\ \mu\text{m}^3$
Uric acid	>6.4 mg/dL for men >5.0 mg/dL for women
Serum glutamic oxaloacetic transaminase (aspartate aminotransferase)	>45.0 IU/L
Serum glutamic pyruvic transaminase (alanine aminotransferase)	>45.0 IU/L
Ethylglucuronide (EtG)	>100 ng/mL
Phosphatidylethanol (PEth)	>200 ng/mL

Ethylglucuronide (EtG), a glucuronide metabolite of ethanol, has utility in monitoring abstinence because EtG is present for several days in urine after any consumption of alcohol, hence making this suitable for identifying a recent relapse. Levels above 100 ng/mL are generally considered a positive result, although caution in interpretation is warranted because false positives can result from using alcohol-containing hand sanitizers or mouthwash.

Finally, phosphatidylethanol (PEth) is another metabolite produced only in the presence of alcohol that remains elevated for 1 to 2 weeks following a heavy drinking episode. Testing for PEth requires a blood sample, and levels above 200 ng/mL are generally considered a cut-off for heavy drinking, while negative results suggest abstinence.

DIFFERENTIAL DIAGNOSIS

Once a pattern of alcohol-related life problems has been identified, the diagnosis of an AUD may be fairly obvious. However, the differential diagnosis is more challenging when a patient presents with psychiatric symptoms that may indicate an *alcohol-induced psychiatric condition,* as described above in the section on diagnostic and clinical features. It is equally important to identify *independent psychiatric disorders* that predate the AUD and/or continue after at least 4 abstinent weeks. This section offers more detail regarding independent comorbid conditions seen with AUDs.

Antisocial Personality Disorder

ASPD in DSM-5 is based on severe antisocial behaviors beginning before age 15 as conduct disorder and continuing into adulthood. People with ASPD are likely to be impulsive, sometimes violent, and risk-takers who do not easily learn from mistakes. About 80% of people with this disorder develop severe alcohol problems and/or SUDs; they are likely to discontinue treatment prematurely and have relatively poor prognoses in AUD treatment, where they represent about 5% of women and 15% of men. Symptoms of other personality disorders are often observed during intoxication and withdrawal, but other than ASPD and borderline personality disorders, few have been documented to predate the AUD. Thus, it is advisable to defer a diagnosis of additional personality disorders until at least a month after abstinence unless the syndrome clearly antedated the AUD. The treatment of AUDs in patients with independent ASPD involves the same general approaches as for alcohol-related disorders overall, with common sense variations to deal with the interpersonal problems.

Independent Schizophrenia

A second independent disorder associated with elevated risks for both AUDs and SUDs is schizophrenia. Typically characterized by a slow onset of paranoid delusions and auditory hallucinations in a clear sensorium and usually beginning in the late teens to the 20s, schizophrenia is likely to be severe, long lasting, and debilitating. Possibly reflecting a lack of long-term psychiatric treatment facilities, people with schizophrenia are sometimes homeless, a situation that, along with impaired judgment associated with their psychoses, may contribute to alcohol- and drug-related problems. There may also be overlap in genetic mechanisms (e.g., related to dopamine) that contribute to AUDs and psychotic disorders. Heavy drinking is likely to undercut the effectiveness of antipsychotic medications, increase mood swings, intensify psychoses, and contribute to a higher probability of repeatedly revolving into and out of inpatient care. AUDs are observed in about 40% of individuals with schizophrenia.

A 34-year-old unemployed, divorced man entered treatment for an AUD. His history revealed seven hospitalizations since age 23 related to schizophrenia and alcohol use. By age 23, the patient had experienced auditory hallucinations and paranoid delusions for the prior year but had not yet demonstrated severe enough problems related to alcohol or other drugs to fulfill criteria for an AUD or SUD. A hospitalization at age 28 was followed by 6 months of abstinence while living in a recovery home, during which time the patient continued to have auditory hallucinations and paranoid delusions without insight. The current hospitalization involved the treatment of alcohol withdrawal while re-evaluating and stabilizing the antipsychotic medications for this man with an independent schizophrenia.

Independent Depressive and Bipolar I Disorders

A third group of conditions that are part of the differential diagnoses of AUDs involves mood-related diagnoses. Substance-induced major depressive disorders are discussed more extensively above under alcohol-induced psychiatric disorders. However, when a major depressive episode occurs before the onset of an AUD or during a period of abstinence, it needs to be treated as an independent condition while the clinician also addresses the alcohol-related problems. While some studies indicate that independent major depressions do not increase the risk for later AUDs, pre-existing bipolar I disorder is more closely linked with the AUD risk. This may reflect heavy drinking that began during mania-related hyperexcitability, impulsiveness, and associated poor judgment. As with schizophrenia, there also may be overlap in genetic mechanisms contributing to both bipolar and substance-related disorders. Bipolar II, on the other hand, is difficult to evaluate in patients with AUDs, because their intoxication, withdrawal, and reactions to life problems can mimic hypomania. Bipolar II labels should be reserved for AUD patients with clear hypomanic episodes antedating the AUD or directly observed by the clinician after at least a month of abstinence.

Independent Major Anxiety, Trauma, and Stress-Related Disorders

There is at least a modest association between the risk for AUDs and independent panic disorder, social phobia, and PTSD. Regarding the latter, a delay of many months between the stressful event and the development of PTSD might indicate an impact of alcohol withdrawal or intoxication-related depression or anxiety on subclinical trauma-related symptoms. In these instances, the PTSD might markedly improve with continued abstinence, perhaps falling below a diagnostic threshold for PTSD. In other individuals, however, PTSD was clearly present before the AUD or remained severe during abstinence, in which case, once the acute alcohol-related problems are addressed, the patient should consider longer-term CBT or other evidence-based psychotherapies, as well as medication-based therapies appropriate for PTSD.

Other Disorders

Attention deficit hyperactivity disorder (ADHD) with a childhood onset involves learning problems, difficulties focusing attention, and problems sitting still and following through on tasks. Because these children can be disruptive at home and school, and because many are treated with stimulant medications such as methylphenidate, it is possible that they might carry a heightened future risk for AUDs and SUDs. However, studies to date do not uniformly support a close tie between ADHD and later substance-related problems unless the child also met criteria for conduct disorder.

There are also complex relationships between AUDs and SUDs. Men and women with ASPD have a high risk for multiple SUDs, including AUDs. It is also probable that individuals with addiction to opioids and stimulants (cocaine and amphetamines) have an increased risk for AUDs by turning to alcohol to moderate the intoxication or side effects from their preferred drug. However, while many individuals with AUDs have used other substances, most do not meet criteria for an SUD. An exception is tobacco use disorders, which are elevated among individuals with AUDs. This finding may relate to attempts to use tobacco or nicotine-containing products to moderate some of the effects of high doses of alcohol or withdrawal, or a possible genetic relationship between nicotine- and alcohol-related syndromes.

Course and Prognosis of AUDs

Most people with AUDs, regardless of their backgrounds, exhibit general similarities in the time course and prevalence of alcohol-related life difficulties. The differences that do exist mostly reflect characteristics of individuals in society in general and may not indicate unique aspects of their AUD. For example, women in the general population and among those with AUDs have higher rates of independent depressive episodes and are less likely to be violent, have a full-time job, or have police records compared to comparable men. However, some sex differences do exist (e.g., women are likely to begin drinking a few years later than men and may demonstrate a faster escalation of AUD symptoms and faster onset of medical complications—i.e., telescoping effect). Similarly, older individuals in the general population and with AUDs are more likely to have medical problems, take multiple medications, and have fewer extensive social support systems compared to younger individuals.

Early Course

Patients with ASPD who develop an AUD are likely to have an early onset of drinking, intoxication, and alcohol-related problems, but that scenario is not applicable to the other 80% of men and 95% of women with AUDs. As shown in Table 11.2–9, the usual person with an AUD has their first drink (other than taking a sip from a parent’s glass) between the ages of 13 and 15 years, the first intoxication usually occurs at 15 or 16, and the first minor alcohol-related problem is usually observed in the late teens, milestones similar to what is expected for people in the general population. More severe alcohol-related difficulties begin to cluster together in the early 20s to the middle 30s.

Table 11.2–9.
Clinical Course of AUDs[a]

Age at first drink[b,c]	13–15 years
Age at first intoxication[b,c]	15–17 years
Age at first problem[b]	16–22 years
Age at onset of an AUD	23–40 years
Age at death	60 years
Fluctuating course of abstention, temporary control, alcohol problems	NA
Spontaneous remission in 20%	NA

AUD, Alcohol use disorder.

NA indicates that age of onset does not apply to this item.

[a] The data presented here are based on DSM-IV abuse and dependence, and it is likely that mild AUD (2 or 3 of 11 criteria endorsed) is similar to DSM-IV abuse, while moderate or severe AUD (4+ criteria endorsed) is more similar to DSM-IV dependence.

[b] Same as general population.

[c] While these ages reflect the usual course of alcohol use in individuals without conduct disorder or antisocial personality disorder, an earlier drinking or intoxication is associated with an increased AUD risk.

Later Course

Once a moderate to severe AUD has developed, unless the person abstains, there are likely to be repeated future periods of drinking problems that alternate with periods of abstinence and *temporary* controlled drinking. The usual person with an AUD uses the temporary cessation of drinking problems to convince themselves that alcohol is not really a cause for concern. Those periods of abstinence, lasting days to months, are usually followed by times when drinking rules are established and at least temporarily followed. For example, the person might avoid spirits (ignoring that a glass of beer or wine and a shot of whiskey have similar amounts of alcohol) or limit drinking to certain times of the day and under specified conditions. This period of temporary control usually leads to an escalation of alcohol intake, followed by new problems that precipitate new abstinent periods, and the cycle begins again.

An important development in at least 20% of people with AUDs is a spontaneous remission. Here, without recent treatment or self-help group participation, the person spontaneously stops drinking and remains abstinent for extended periods. Such outcomes are most likely for those with fewer alcohol-related problems, more social supports, jobs, greater life stability, and no concomitant ASPD or SUD diagnoses.

The fluctuating course and high rate of spontaneous remission that are typical of the course of AUDs are major reasons why evaluations of potential treatments for these conditions require double-blind controlled studies. Individuals with AUDs are likely to come for treatment at a point when their alcohol problems are the most severe. However, at any time of follow-up, they are likely to be functioning better than at treatment entry. This change reflects both temporal fluctuations in their drinking pattern over time as well as spontaneous and perhaps permanent remissions.

However, if drinking continues, the person with AUDs potentially decreases their lifespan by at least 10 years as a result of the alcohol-related pathologies described earlier in this chapter. This early death also reflects an almost 10% risk for suicide, usually during a depressive episode.

Prognosis

As few as 10% to 20% of people with AUDs ever receive treatment for their disorder. However, once they complete a program, a majority do well and develop long-term, often permanent, abstinence. Good outcomes are most likely in patients who do not have ASPD and in individuals who have a job to go back to and close family contacts who offer social support. These attributes predict about a 60% chance for >1 year of abstinence, with 1-year rates associated with a good chance for continued abstinence over an extended period. Individuals with independent major psychiatric conditions, such as ASPD, schizophrenia, and bipolar I disorder, are likely to run the course of their independent psychiatric illness as well.

TREATMENT

The core of treatment for AUDs involves maximizing motivation for abstinence, helping the person restructure a life without alcohol, and minimizing relapse to substance-using behaviors. This CBT-based approach is similar to efforts appropriate for any long-term disorder that requires changes in lifestyle, such as diabetes or hypertension. Much of the challenge for clinicians rests with recognizing the high prevalence of AUDs, how often these conditions involve temporary symptoms of other psychiatric syndromes, and how to use clinical clues, physical findings, and laboratory tests to identify an AUD.

Three general steps are involved in treating the person with an AUD: intervention; detoxification (if needed); and rehabilitation, including relapse prevention. These approaches can begin after life-threatening medical conditions are addressed and psychiatric emergencies treated. Thus, suicidal plans, psychotic features without insight, and alcohol withdrawal delirium require inpatient hospitalizations until the patient is stabilized enough to begin alcohol rehabilitation.

Intervention

The goal is to use the principles of motivational interviewing and brief interventions to help patients recognize the choices they have and the adverse consequences likely to occur if they do not stop drinking. A discussion of the presenting complaint (e.g., insomnia or depression) is a useful way to both show empathy and enhance motivation for change. The emphasis is then placed on discussing how alcohol either created or contributed to these problems and reassuring the patient that life changes, potentially including abstinence, are possible. If the person does not respond to the first brief intervention or motivational interview, the same nonjudgmental but persistent approach can be used each time an alcohol-related problem is identified. Many individuals with AUDs take time to realize how alcohol contributed to problems and to be ready and willing to change.

For the physician, the intervention often begins with a “brief intervention,” a cost-effective approach that can be used in the clinician’s office. It is summarized by the acronym FRAMES: giving Feedback on alcohol risks; emphasizing a person’s Responsibility for change; offering Advice; giving a Menu of options to consider; employing Empathy; and emphasizing Self-efficacy, or the ability to change. The goal is to establish an alliance with patients by demonstrating an understanding of their viewpoint while encouraging them to think through consequences associated with alcohol and the way that changing behaviors might yield benefits. During these discussions, it is important to recognize the patient’s ambivalence toward abstinence and to monitor the person’s readiness to change. Resistance is best handled through discussion and problem solving rather than direct confrontation.

Two 15-minute sessions, along with a workbook available from the National Institute on Alcohol Abuse and Alcoholism (pubs.niaaa.nih.com), can be used to review drinking problems and help the patient understand how life cues contribute to the drinking pattern. Between two visits separated by about 1 month, patients keep notes of their drinking patterns. The second session reviews the same issues and is followed by several short telephone interviews.

Detoxification

Most people with an AUD have relatively mild, if any, withdrawal symptoms when they stop drinking. If the patient is in relatively good health and has a good social support system, the withdrawal syndrome associated with alcohol usually resembles a mild flu or bad hangover. The essential first step in detoxification is a thorough physical examination. In the absence of a serious medical disorder or a comorbid SUD, severe alcohol withdrawal is unlikely. The second step is to offer rest, nutrition, and oral multivitamins, including thiamine.

Mild or Moderate Withdrawal. Alcohol withdrawal develops because the brain, adapted to the presence of this drug, downregulates the inhibitory systems and cannot function adequately in the absence of alcohol. Many patients, especially young and healthy individuals, have no clinically significant withdrawal symptoms and can go directly into the rehabilitation phase of treatment. Others will benefit from medications, giving enough of a brain depressant on the first day to diminish symptoms and then weaning the patient off the medication over the course of acute withdrawal, usually over about 5 days. If the symptoms are mild and the person is in good health, this can be done as an outpatient with daily brief visits. Any depressant can work, but most clinicians choose benzodiazepines for their relatively low cost and higher safety. Adequate treatment can be given with either short-acting drugs, such as lorazepam, or long-acting medications such as chlordiazepoxide or diazepam.

An example is administering 25 mg of chlordiazepoxide by mouth three or four times a day on the first day, with a notation to skip a dose if the patient is sleepy. Up to two additional 25-mg doses during the first day can be used if tremor or autonomic dysfunction increases. The first day’s dose can be decreased by about 20% each subsequent day, with a resulting need for no medication after 4 or 5 days. However, the prior day’s dose should be restored if withdrawal symptoms increase and further decreases in doses delayed a day or two. When using a long-acting agent, such as chlordiazepoxide, the clinician must avoid producing excessive sedation; if the patient is sleepy, the next scheduled dose should be omitted. When using a short-acting drug, such as lorazepam, the patient must not miss any dose because rapid changes in benzodiazepine concentrations can precipitate a seizure. Other depressants, including anticonvulsants (e.g., gabapentin) can also be effective but have added dangers and costs compared to benzodiazepines. Some clinicians recommend β-adrenergic receptor antagonists (e.g., propranolol) or α-adrenergic receptor agonists (e.g., clonidine), although these medications do not decrease the risk of seizures or delirium and can mask the intensity of some withdrawal signs that are used to monitor withdrawal severity.

Alternative to this fixed-dose schedule, another option, especially for those with mild to moderate withdrawal, is to adopt a symptom-triggered approach. The gold-standard tool to measure alcohol withdrawal severity is the Clinical Institute Withdrawal Assessment for Alcohol-Revised (CIWA), which can be administered to arrive at a score. Above a certain cut-off, typically 8, medications are administered but withheld if scoring below this cut-off. Studies have shown that a symptom-triggered approach can still treat withdrawal adequately but reduces the duration of treatment and the overall need for medications without increasing the risk of progression to severe withdrawal.

Severe Withdrawal. For the less than 3% of AUD patients with extreme autonomic dysfunction, agitation, and confusion (alcoholic withdrawal delirium or delirium tremens [DTs]), there is no perfect treatment. DTs represent a medical emergency and will usually require treatment in intensive care. The first important step is to determine why such severe withdrawal symptoms occurred, especially whether there is a concomitant medical problem that needs immediate treatment. The withdrawal symptoms in DTs can then be minimized with benzodiazepines (high doses are sometimes required), along with the adjunctive use of an antipsychotic (e.g., haloperidol [Haldol]) to control severe agitation and hallucinations. If the patient does not respond to benzodiazepines, a clinician experienced in treating withdrawal delirium in an ICU can consider inducing a state of arousable deep sleep with propofol or the α-2-receptor agonist dexmedetomidine. Here, higher doses are used on the first or second day to control behavior, and the patient is then typically weaned off medications by about the fifth day.

Another 1% of patients in alcohol withdrawal have a single grand mal convulsion, with a peak incidence on day 2. This requires a neurologic evaluation, but in the absence of an independent seizure disorder, a single seizure does not indicate using anticonvulsant drugs.

Protracted Withdrawal. Finally, regarding withdrawal, mild symptoms of anxiety, insomnia, and autonomic overactivity are likely to continue for 2 to 6 months following acute withdrawal. Although no pharmacologic treatment for this syndrome has been identified, a medication discussed below for the rehabilitation phase, acamprosate, might diminish some of these symptoms. It is important to warn patients that some sleep problems or feelings of nervousness might remain after acute withdrawal and consider efforts to help them feel more comfortable because these protracted withdrawal symptoms enhance the probability of relapse.

Rehabilitation

Rehabilitation includes three major components: (1) increasing and maintaining high levels of motivation for change; (2) helping patients readjust to a lifestyle that maximizes functioning and decreases the risk for drinking; and (3) relapse prevention. These steps often begin in the context of acute and protracted withdrawal syndromes and life crises.

No single major life event or psychiatric disorder is a unique cause of AUDs. The effects of earlier causes are likely to have been joined by reinforcing effects of the drug, especially regarding relief of hangovers and mild withdrawal symptoms, and the years of a stressful lifestyle in the context of heavy drinking. This is true even though many people with AUDs believe that the unique cause was depression, anxiety, difficult life events, prior stress, or pain syndromes. However, the alcohol was as likely to have contributed to the mood disorder, accident, or life stress as the other way around.

The same general treatment approach is used in inpatient and outpatient settings. The selection of the more expensive, but more intense and effective, inpatient mode often depends on evidence of additional severe medical or psychiatric syndromes, severe stresses that would interfere with treatment, the absence of nearby outpatient groups, and/or the patient's nonresponse to outpatient care. The treatment process in either setting usually involves 2 to 4 weeks of care as an intensive period of help, followed by at least 3 to 6 months of less frequent outpatient care. The latter uses a combination of individual and group counseling, avoiding psychotropic medications unless needed for independent disorders, and involvement in self-help groups such as Alcoholics Anonymous (AA).

Counseling. Counseling in the first several weeks to months of recovery focuses on day-to-day life issues to help patients maintain a high level of motivation for abstinence and to enhance their functioning. Psychotherapy techniques that provoke anxiety or that require deep insights may not be beneficial during the early phases of treatment, and, at least theoretically, the added anxiety may impair the ability to abstain.

One goal of these CBT- and motivation-based counseling sessions is to help the patient learn ways to cope with life stresses. The clinician can use modeling and role-playing of situations likely to be encountered while working toward abstinence, along with encouraging patients to practice these skills between sessions. At the same time, individuals work to identify areas of problems in day-to-day functioning, paying special attention to how they react to these challenges and the impact that substance use might have on the outcomes.

To optimize motivation, treatment sessions explore the consequences of drinking, the likely future course of alcohol-related life problems, and the improvement that can be expected with abstinence. Counseling also deals with how to build a lifestyle free of alcohol. Discussions cover the need for a sober peer group, a plan for social and recreational events without drinking, and approaches for re-establishing communication with family members and friends. Whether in an inpatient or an outpatient setting, individual or group counseling is usually offered about three times a week for the first 2 to 4 weeks, followed by less intense efforts, perhaps once a week, for the subsequent 3 to 6 months.

Relapse Prevention. The third component of rehabilitation, relapse prevention, begins with identifying situations with high risks for relapse. The counselor helps patients develop ways to identify and cope with craving and with challenging events or emotional states that make drinking more likely. It is also important to remind patients that short-term relapses are not an excuse for returning to regular drinking. Recovery is a process of trial and error, and slips can be clues to help identify high-risk situations and to develop more appropriate coping techniques.

Importance of the Family. AUDs affect many people in the patient’s life, and an important aspect of recovery involves helping family and friends understand more about AUDs and how rehabilitation is an ongoing process. Counseling and support for spouses, offspring, and friends can help them work to rebuild key relationships and learn how to avoid protecting the patient from the consequences of drinking, a step likely to undercut the patient’s motivation to abstain.

Treatment of Independent Psychiatric Disorders. Regarding some of the independent psychiatric disorders discussed under Differential Diagnoses above, clinicians must treat the patient who has bipolar I disorder with appropriate psychotherapy and lithium and/or valproic acid; use relevant psychological and behavioral techniques for the patient with ASPD; and offer long-term antipsychotic medications to the patient with schizophrenia. The goal is to minimize the symptoms of the independent psychiatric disorder in the hope that a greater level of life stability will be associated with a better prognosis for the patient’s alcohol problems. Initial treatment should work to control mania or severe depression along with treatment of withdrawal, and alcohol rehabilitation should be considered once the independent disorder is under control. Especially for more severe psychiatric symptoms, concurrent treatment of the AUD and the comorbid psychiatric disorders are recommended, while sequential treatment may be reasonable if the symptoms are less severe.

Possible Medications for the AUD. Patients with AUDs who do not have an independent psychiatric disorder have no clear need for antidepressants or antianxiety drugs after withdrawal in the treatment of their AUDs. Moderate anxiety and insomnia that linger as part of a protracted abstinence syndrome should be treated with CBT, education, and reassurance. Medications, including benzodiazepines, for anxiety and insomnia are likely to lose effectiveness as tolerance develops, and patients sometimes increase the dose and have subsequent problems related to the prescribed drug.

Similarly, although low levels of sadness and mood swings can linger for months, controlled clinical trials indicate little benefit to prescribing antidepressant medications or lithium for the average person with an AUD who has no independent psychiatric disorder. The substance-induced mood disorder is likely to clear before the medications can take effect, and patients who resume drinking while on the medications face potential dangers of alcohol–drug interactions. With little evidence that such medications are routinely effective, the dangers outweigh potential benefits from their routine use.

As is true for most clinical disorders, the use of medications for the treatment of AUDs requires remembering that all treatments have both benefits and risks. Double-blind, controlled studies are important to establish an acceptable ratio of assets to dangers while controlling for the impact of spontaneous remission and fluctuations in patients’ conditions that occur over time. The random variation in the course of AUDs, along with the probability that patients are most likely to seek treatment when their problems are most intense, means that the only way to be confident that a medication contributed to a good outcome is through controls for spontaneous remissions and fluctuating intensities of drinking problems by also observing outcomes in an appropriate comparison group.

Data from double-blind trials support modest effects for two medications offered in the context of CBT. The first drug is acamprosate, an analog of the amino acid neurotransmitter taurine with a structure that resembles GABA. This medication antagonizes neuronal overactivity related to the excitatory neurotransmitter glutamate, at least in part, by acting as an antagonist to NMDA receptors. Thus, acamprosate may diminish mild anxiety, mood swings, and other sleep difficulties associated with the protracted withdrawal syndrome observed after the first 4 to 5 days of abstinence. Using about 2,000 mg divided into three doses per day, the majority of studies report 15% to 20% better drinking outcomes with active drug compared to placebo, with relatively mild side effects of GI problems such as diarrhea. Although there are no clear guidelines, clinicians often prescribe this drug for 6 months, during which time it is hoped an alcohol-free lifestyle can be developed.

The second promising medication is the long-acting opioid antagonist naltrexone (taken orally or in an injectable form). This agent has been marketed for many years as a treatment for maintaining opioid abstinence and is effective for rehabilitation with highly motivated individuals with opioid use disorder because this opioid blocker drug attenuates an opioid high. Naltrexone works by blocking opioid receptors in the brain with associated decreased activity in the ventral tegmental reward area of the brain and is hypothesized to decrease the rewarding effects of a drink and to diminish craving—interrupting the positive reinforcement from alcohol. The majority of double-blind, placebo-controlled trials supports the superiority of naltrexone with about 15% to 20% better drinking outcomes than for placebo. The side effects for 50 mg per day include mild GI upset and lethargy. Several studies indicated that individuals with a gene variation related to the Asn40Asp location in the gene for the mu-opioid receptor (OPRM1) have better AUD outcomes, but recent studies have reported contradictory results. There is also a 380-mg, once-per-month naltrexone injection, which might be particularly useful for noncompliant patients. By avoiding the liver's first pass effect, the injectable formulation may carry less risk of causing hepatotoxicity, which is one of the side effects of naltrexone that limits its use.

The combination of acamprosate (666 mg three times a day) plus naltrexone (50 mg per day or three times a week) may be more effective than either drug alone, while not increasing the side effect profile. However, not all studies agree, and further research is needed.

A third drug of possible interest is the alcohol-sensitizing agent disulfiram, which is usually prescribed at 250 mg per day, the highest dose most patients will be likely to tolerate. The goal is to warn patients that drinking alcohol while taking disulfiram precipitates nausea, vomiting, and changes in blood pressure, and the medication is best used when taken daily while being observed by someone like a relative or pharmacist. However, few data convincingly prove that disulfiram is more effective than placebo, probably because patients on placebo are given the same caveat. Many clinicians have stopped prescribing this agent, partly because of the dangers associated with the drug itself, including mood swings, rare instances of psychosis, a possible increase in peripheral neuropathies, and a rare but potentially fatal hepatitis. Moreover, patients with pre-existing heart disease, stroke, diabetes, and other conditions cannot be given disulfiram because an alcohol reaction could be fatal.

Several additional medications are worth brief mention but have only limited data to support their use. First, while not approved for the treatment of AUDs, the anticonvulsant topiramate has the usual GABA-boosting activity associated with antiseizure treatments along with a glutamate receptor blockade and effects on dopamine. Perhaps related to these combined effects, several studies using topiramate (300 mg per day) in AUD treatment reported improvements in drinking patterns, but studies also noted side effects of mental slowing, decreased concentration, memory impairment, and paresthesias that are difficult for many patients to tolerate. Similar results and side effect profiles, as well as abuse liabilities, have been reported for another anticonvulsant, gabapentin, which has been evaluated at between 900 and 1,800 mg/day, but has been reported to have a relatively high dropout rate. Still, there is a growing evidence base for the role of gabapentin in addressing both mild to moderate withdrawal and supporting ongoing abstinence. The misuse liability of gabapentin, notably among those with opioid use disorder, remains a concern and warrants careful consideration. Third, some data indicate that 4 micrograms per kilogram body weight of the serotonin-3 receptor antagonist ondansetron might be beneficial in AUD treatment, particularly in early-onset AUDs with comorbid drug dependence and criminality. Additional medications that have been evaluated for the treatment of AUDs include several selective serotonin reuptake inhibitors (SSRIs, e.g., sertraline up to 200 mg per day), with some evidence that patients with the SS genotype in the promoter region of the serotonin transporter gene may be more likely to respond; the GABA-B receptor agonist baclofen; and some antipsychotic medications, although these medications have not consistently proven to be better than placebo in controlled trials and are not recommended for routine use.

Self-Help Groups. Clinicians should consider the potential importance of “12-step-based” self-help groups such as AA. Similar to what occurs in CBT and relapse prevention, AA participants are encouraged to change the way they view their drinking while building a lifestyle that optimizes sober friends and uses the support of family members. Help from AA is available 24 hours a day, and participants learn they can still enjoy social functions without drinking, while observing models of recovery through the accomplishments of sober members of the group.

Clinicians play important roles by encouraging AA attendance while helping patients select the most appropriate group. Some meetings are comprised only of men or women, and others are mixed; some are mostly attended by blue-collar men and women, whereas others are mostly for professionals; some groups place great emphasis on religion, and others are more eclectic. Patients with independent psychiatric disorders may need additional education about AA. The clinician should remind them that some members of AA may not understand their special need for medications and should suggest ways to cope with inappropriate suggestions that important medications be stopped. Although difficult to evaluate using double-blind controls, most studies indicate that participation in AA is associated with improved outcomes.

Following the family intervention and at the insistence of his wife and partners, JP agreed to enter treatment. He and his family opted for a more intensive inpatient-based alcohol and drug treatment program of 4 weeks duration. The physical examination on entrance to treatment revealed no major medical problems. Vital signs after 3 days of sobriety were within normal limits except for a slight increase in heart rate (80 beats per minute) and blood pressure (145/90). To treat a mild withdrawal syndrome, 25 mg chlordiazepoxide up to three times per day as needed was prescribed for systolic blood pressure over 140/90, tremor, or other signs of autonomic dysfunction. JP received two doses on his first day of hospitalization, one dose on his second day, and no further benzodiazepines were required.

The core of therapy involved group counseling twice per day, emphasizing the patient's responsibility for the decision to abstain, discussing problems associated with continued substance use, and describing steps that can be taken to help rebuild one's relationship with family, colleagues, and friends. JP cooperated well with treatment, and the discharge planning included a discussion of the possible medications during the rehabilitation phase. After consulting with the patient and the family, both acamprosate (666 mg by mouth three times a day) and naltrexone (50 mg by mouth daily) were prescribed, with the goal of continuing the medications for approximately 6 months.

In addition, couples counseling was carried out with JP and his wife to optimize her understanding of the recovery process and offer her the opportunity of dealing with questions and feelings of frustration, anger, and anxiety that developed during JP's drinking years. The subsequent aftercare outpatient group met twice a week. Steps were taken to help JP identify an AA group with which he might be most comfortable in terms of shared background with members as well as time and location. An Al-Anon group was identified for his wife as well.

JP's course involved two brief relapses in the first 6 months of treatment, each involving a return to alcohol lasting less than 2 days. The rapid recovery of abstinence was likely to be related to his prior relapse prevention training as well as his return to both AA and aftercare treatment. After the first 6 months, no further relapses were observed over the next year. Currently, JP continues with one or two AA meetings per week and has become involved in outreach from AA, visiting a local hospital and helping to lead an AA group for individuals in early recovery.

FURTHER READINGS

Anthenelli RM, McKenna BS, Smith TL, Schuckit MA. Relationship between level of response to alcohol and acute tolerance. *Alcohol Clin Exp Res.* 2021;45(7):1504–1513.

*Compton WM, Dawson DA, Goldstein RB, Grant BF. Crosswalk between DSM-IV dependence and DSM-5 substance use disorders for opioids, cannabis, cocaine and alcohol. *Drug Alcohol Depend*. 2013;132:387.

Donoghue K, Elzerbi C, Saunders R, Whittington C, Pilling S, Drummond C. The efficacy of acamprosate and naltrexone in the treatment of alcohol dependence, Europe versus the rest of the world: a meta-analysis. *Addiction*. 2015;110:920–930.

Edwards AC, Kendler KS. Twin study of the relationship between adolescent attention-deficit/hyperactivity disorder and adult alcohol dependence. *J Stud Alcohol Drugs*. 2012;73:185–194.

*Garbutt JC, Greenblatt AM, West SL, et al. Clinical and biological moderators of response to naltrexone in alcohol dependence: a systematic review of the evidence. *Addiction*. 2014;109:1274–1284.

Grant BF, Chou SP, Saha TD, et al. Prevalence of 12-month alcohol use, high-risk drinking, and DSM-IV alcohol use disorder in the United States, 2001–2002 to 2012–2013: results from the National Epidemiologic Survey on alcohol and related conditions. *JAMA Psychiatry*. 2017;74(9):911-923.

*Hasin DS, O’Brien CP, Auriacombe M, et al. DSM-5 criteria for substance use disorders: recommendations and rationale. *Am J Psychiatry*. 2013;170:834–851.

Hasler BP, Martin CS, Wood DS, Rosario B, Clark DB. A longitudinal study of insomnia and other sleep complaints in adolescents with and without alcohol use disorders. *Alcohol Clin Exp Res*. 2014;38:2225–2233.

Humphreys K, Barreto NB, Alessi SM, et al. Impact of 12-step mutual help groups on drug use disorder patients across six clinical trials. *Drug Alcohol Depend*. 2020;215:108213.

Jin S, Cao Q, Yang F, et al. Brain ethanol metabolism by astrocytic ALDH2 drives the behavioral effects of ethanol intoxication. *Nat Metab*. 2021;3(3):337–351.

Kandaswamy R, Allegrini A, Plomin R, von Stumm S. Predictive validity of genome-wide polygenic scores for alcohol use from adolescence to young adulthood. *Drug Alcohol Depend*. 2021;219:108480.

Kendler KS, Ohlsson H, Sundquist J, Sundquist K. Prediction of onset of substance-induced psychotic disorder and its progression to schizophrenia in a Swedish National sample. *Am J Psychiatry*. 2019;176(9):711–719.

Kranzler HR, Soyka M. Diagnosis and pharmcotherapy of alcohol use disorder: a review. *JAMA*. 2018;320(8):815–824.

Lees B, Mewton L, Jacobus J, et al. Association of prenatal alcohol exposure with psychological, behavioral, and neurodevelopmental outcomes in children from the adolescent brain cognitive development study. *Am J Psychiatry*. 2020;177(11);1060–1072.

Luk JW, Liang T, Wall TL. Gene-by-environment interactions on alcohol use among Asian American college freshmen. *J Stud Alcohol Drugs*. 2017;78(4):531–539.

Mann K, Lemenager T, Hoffman S, et al; PREDICT study team. Results of a double-blind placebo controlled pharmacotherapy trial in alcoholism conducted in Germany and comparison with the US COMBINE Study. *Addict Biol*. 2013;18:937–946.

NIAAA. Alcohol facts and statistics. 2021. Accessed May 25, 2021. https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/alcohol-facts-and-statistics

Pollard MS, Tucker JS, Green HD Jr. Changes in adult alcohol use and consequences during the COVID-19 pandemic in the US. *JAMA Netw Open*. 2020;3(9):e2022942.

Pridemore WA. The impact of hazardous drinking on suicide among working-age Russian males: an individual-level analysis. *Addiction*. 2013;108:1933–1941.

Prom-Wormley EC, Ebejer J, Dick DM, Bowers MS. The genetic epidemiology of substance use disorder: a review. *Drug Alcohol Depend*. 2017;180:241–259.

Reed MB, Woodruff SI, DeMers G, et al. Results of a randomized trial of screening, brief intervention, and referral to treatment (SBIRT) to reduce alcohol misuse among acitve-duty military personnel. *J Stud Alcohol Drugs*. 2021;82(2):269–278.

Rehm J, Heilig M, Gual A. ICD-11 for alcohol use disorders: not a convincing answer to the challenges. *Alcohol Clin Exp Res*. 2019;43(11):2296–2300.

Reilly MT, Noronha A, Goldman D, Koob GF. Genetic studies of alcohol dependence in the context of the addiction cycle. *Neuropharmacology*. 2017;122:3–21.

Roos CR, Maisto SA, Witkiewitz K. Coping mediates the effects of cognitive-behavioral therapy for alcohol use disorder among out-patient clients in Project MATCH when dependence severity is high. *Addiction*. 2017;112(9):1547–1557.

Rosenstrom T, Torvik FA, Ystrom E, et al. Prediction of alcohol use disorder using personality disorder traits: a twin study. *Addiction*. 2018;113(1):15–24.

Schuckit MA, Smith TL, Clausen P, et al. The low level of response to alcohol-based heavy drinking prevention program: one-year follow-up. *J Stud Alcohol Drugs*. 2016;77(1):25–37.

Schuckit MA, Smith TL, Kalmijn JA. The patterns of drug and alcohol use and associated problems over 30 years in 397 men. *Alcohol Clin Exp Res*. 2014;38:227–234.

Schuckit MA. Recognition and management of withdrawal delirium (delirium tremens). *N Eng J Med*. 2014;371:2109–2113.

Studer J, Gmel G, Bertholet N, Marmet S, Daeppen JB. Alcohol-induces blackouts at age 20 predict the incidence, maintenance and severity of alcohol dependence at age 25: a prospective study in a sample of young Swiss men. *Addiction*. 2019;114(9):1556–1566.

Yip SW, Doherty J, Wakeley J, et al. Reduced subjective response to acute ethanol administration among young men with a broad bipolar phenotype. *Neuropsychopharmacology*. 2012;37:1808–1815.

11.3 Stimulant-Related Disorders

ROCCO A. IANNUCCI, M.D., AND ROGER D. WEISS, M.D.

Stimulants in their natural forms, such as the plants *Erythroxylon coca*, *Ephedra sinica*, and *Catha edulis*, have been used by humans for millennia. As a class of substances, they share the acute effect of increasing physiologic and mental arousal through dopaminergic pathways. They include amphetamine-type substances (those with a substituted phenylethylamine structure), structurally dissimilar agents with similar effects (such as methylphenidate), and cocaine. In the United States, cycles of widespread stimulant misuse and associated problems have occurred for more than 100 years. Stimulant use and stimulant use disorders present a staggering burden worldwide, with an estimated 27 million people misusing some form of stimulant in 2019. In the United States, stimulants are responsible for a disproportionately high number of emergency department visits and overdoses, underscoring the potential dangerousness of misusing these substances. In recent years, overdoses due to stimulants have risen to alarming levels and have exacerbated the opioid overdose crisis. More than ever, developing and disseminating effective treatment for people with stimulant use disorders remains a pressing public health need.

DEFINITIONS

The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) describes 10 stimulant-related diagnoses, which are further categorized into stimulant use disorder and stimulant-induced disorders. Stimulant use disorder is defined in DSM-5 by a cluster of cognitive, behavioral, and physiologic symptoms indicating continued use of stimulants despite significant problems related to or caused by their use. In the 10th revision of the *International Classification of Diseases* (ICD-10), the corresponding amphetamine- and cocaine-related diagnoses are defined by cognitive, behavioral, and physiologic phenomena reflecting persistent use of these stimulants despite harmful consequences and to the exclusion of other activities and obligations. The hallmark of a severe stimulant use disorder is a pattern of compulsive substance use and loss of control over use that interferes with other aspects of life and health. It may involve the development of tolerance and/or withdrawal symptoms, but neither is necessary or sufficient to meet diagnostic criteria for stimulant use disorder. Stimulants share physiologic and subjective effects and a direct impact on dopaminergic function. Other substances with stimulating properties but markedly different mechanisms of action, such as nicotine and caffeine, are classified elsewhere.

Stimulant tolerance is defined in DSM-5 as the need for an increased amount of amphetamine, cocaine, or other stimulants to achieve the same effect, or a diminished response to continued use of a given amount. Tolerance may occur as a symptom of a stimulant use disorder, but it is not included as a disorder itself. On the other hand, stimulant withdrawal is identified as a stimulant-induced disorder in DSM-5. It is defined as a constellation of symptoms opposite to the effects of stimulant intoxication, including dysphoric mood and characteristic physiologic changes. Another important aspect of stimulant use disorders is craving for the drug. Individuals with stimulant use disorders describe powerful urges to use the substance. These may arise *de novo* or may be triggered by access to stimulants, environmental cues, or seemingly unrelated stresses.

HISTORY

Cocaine

Cocaine in all its forms is a product of the leaves of the plant *Erythroxylon coca*. This plant is indigenous to the high mountain ranges of South America, where coca leaves have been chewed for at least 15 centuries. With European expansion into the Western hemisphere, the Spanish conquistadores first came into contact with coca use during the 16th century. While at first prohibiting the use of coca leaves, the Spanish subsequently used control over coca distribution to rule and exploit local inhabitants, including paying workers in coca leaves. The Spaniards found in this practice the additional economic benefit of energizing the colonized Indian laborers so they could work long hours in gold and silver mines with little rest.

In turn, the Spanish brought coca leaves back to Europe. However, worldwide use of cocaine products did not begin until the active constituent of coca leaves was extracted, first by Friedrich Gaedecke in 1855 and subsequently by Albert Niemann, who coined the term cocaine in 1859. Early experimentation with cocaine peaked around the turn of that century, with luminaries including Sigmund Freud expounding the virtues of the drug. In the United States, cocaine made its way into medicinal tonics and even soft drinks such as Coca-Cola, of which it remained an ingredient until 1903. In the early part of the 20th century, increased awareness of exaggerated and often patently false claims by the manufacturers of cocaine-containing medicinal products raised public concern about spreading cocaine use. At the same time, tales of cocaine addiction began to circulate, and newspaper reports linked the drug with crime. In 1910, President William Taft proclaimed cocaine a national threat. Shortly thereafter, the Harrison Narcotic Act of 1914 was passed, banning the nonmedical use of cocaine and highly regulating its medicinal use. Notwithstanding that cocaine is a stimulant rather than a narcotic, the effect of the Harrison Act was to severely curtail cocaine use, essentially driving it “underground.” It remained there, the purview of relatively small communities of people such as jazz musicians and the wealthy and privileged until the 1970s.

The reasons for the widespread use of cocaine after 1970 remain somewhat unclear. At the time, cocaine was largely perceived as a nonaddictive, short-acting, “soft” drug. Furthermore, attitudes among people who had experimented with other drugs in the 1960s left them open to trying a substance then seen as relatively safe. Increased production in the 1980s drove down the price of the drug, and cocaine use spread to groups of people who previously could not afford it, reaching epidemic proportions in the United States in the early-to-mid 1980s. Indeed, between 1976 and 1985, cocaine-related medical emergencies rose to nine times their previous level, and cocaine-related deaths increased 11-fold. The mid-1980s also brought about the advent of an inexpensive, ready-to-smoke cocaine alkaloid called “crack,” the use of which was highly reinforcing due to its inhaled route of administration.

Estimates of active cocaine use peaked in 1985 at 5.7 million, or 3% of the U.S. population. After that point, cocaine use began to fall dramatically, with a 20% drop in use between 1986 and 1987 alone and a 66% decrease by 1992. While the precise factors leading to decreased cocaine use after its peak in the 1980s remain unclear, it coincided with a number of possibly related events. In part due to a far-reaching awareness campaign, along with well-publicized celebrity deaths attributed to cocaine, the perceived danger of cocaine use began to increase. The previously glamorous position of cocaine changed as the drug became associated with addiction, overdose deaths, crime, fetal injury, and the stigma of marginalized groups. At the same time, enforcement of antidrug laws became more aggressive, with increased arrests and penalties for cocaine-related offenses.

It is controversial whether punitive legal approaches have had any beneficial effect on rates of drug use and addiction. What is more clear are the substantial harms of the so-called “war on drugs” experienced by people of color. The Congressional Anti-Drug Abuse Act of 1986 established sentencing standards that disproportionately affected Black Americans who were more likely to use “crack” cocaine while relatively sparing the typically White, wealthier people who tended to use powdered cocaine. Under this law, carrying 100 g of “crack” cocaine was treated as high-level trafficking, while the same treatment was reserved only for those who carried more than 5,000 g of powdered cocaine. Strikingly, this law established sentencing policies under which the possession of 5 g of “crack” cocaine carried the same minimum sentence as 500 g of powdered cocaine without any fact-based rationale. This and similar laws contributed to a persistent inequity: while African Americans are not more likely to use drugs than White Americans, they are more likely to be arrested, convicted, and imprisoned for drug use.

Amphetamine-Type Stimulants

Like cocaine, the use of amphetamine-type stimulants dates back many centuries. Ma-Huang, a form of the stimulant ephedrine, has been used as a Chinese herbal remedy for thousands of years. In 1887, the active ingredient ephedrine was isolated by Nagayoshi Nagai. At the same time, amphetamine was synthesized in Germany by Lazar Edeleanu. Methamphetamine was first synthesized in Japan in 1918. Therapeutic applications of amphetamines in Western civilizations began in the 1930s with the use of amphetamine (Benzedrine) as a bronchodilator. At that time, the cognitive and emotional effects of amphetamines were recognized, including reinforcing properties such as euphoria. Indeed, amphetamines were widely used during World War II to counteract soldiers’ fatigue. Following the war, the first amphetamine epidemic took place in Japan, where surplus military supplies of methamphetamine led to widespread use.

An increasingly abundant supply of amphetamines in the decades that followed was associated with peaking rates of misuse in the United States during the 1960s. At the end of the decade, regulatory agencies responded with efforts to decrease the manufacture of amphetamines to more closely match medicinal needs. During this period, amphetamines were reclassified as Schedule II medications, reflecting closer government oversight and regulation than in previous decades. Despite these efforts, American misuse of methamphetamine increased dramatically beginning in the mid-1990s as illicit manufacturing sources took over. Methamphetamine use remains a significant problem worldwide, and levels are on the rise in many regions of the United States. While vulnerable to misuse, prescription amphetamines remain useful for therapeutic purposes, including the treatment of attention deficit hyperactivity disorder (ADHD) and some sleep disorders.

COMPARATIVE NOSOLOGY

In diagnosing stimulant-related disorders, clinicians are guided by two sets of clinical criteria. These are delineated in the American Psychiatric Association's DSM and in the World Health Organization (WHO)'s ICD. As is the case for other substance use disorders (SUDs), the advent of DSM-5 heralded a consolidation of previously separate stimulant abuse and dependence diagnoses into one category: stimulant use disorder, with level of severity specified on a spectrum of mild, moderate, or severe. Reflecting the high degree of overlap in drug effects, patterns of use, and symptoms, the previously distinct diagnostic entities for cocaine- and amphetamine-related disorders were combined into one comprehensive category, stimulant-related disorders.

The ICD-11 classification system maintains distinct disorders for different classes of stimulants: cocaine, synthetic cathinones, and stimulants including amphetamines, methamphetamine, or methcathinone. Disorders of stimulant use are identified as primary diagnoses, while conditions caused by stimulant use are considered associated diagnoses. Stimulant dependence (with the particular stimulant class identified) is the most severe stimulant use diagnosis, generally corresponding to moderate-to-severe DSM-5 stimulant use disorder. It is diagnosed when a person experiences symptoms in two of three categories: difficulty with or inability to control use, stimulant use taking priority over other activities, or physiologic dependence including withdrawal, tolerance, or use of a stimulant to prevent withdrawal. The other ICD-11 stimulant use diagnoses consist of harmful patterns of use of a stimulant and episodes of harmful stimulant use. These diagnoses are appropriate when criteria for stimulant dependence are not met, but stimulant use is detrimental to the mental or physical health of a person or of others interacting with that person. Stimulant use that increases the risk of health or other problems is categorized with problems associated with health behaviors, using the term hazardous use of stimulant. Key DSM-5 stimulant-related diagnoses and corresponding ICD-11 diagnoses (including conditions caused by stimulant use) are listed in Table 11.3–1.

Table 11.3–1.
DSM-5 and ICD-11 Stimulant-Related Disorders

DSM-5	ICD-11 (specify stimulant category[a])
Stimulant use disorder (mild, moderate, severe)	Stimulant dependence Harmful pattern of use of stimulant Episode of harmful use of stimulant
Stimulant intoxication	Stimulant intoxication
Stimulant withdrawal	Stimulant withdrawal
Stimulant-induced psychotic disorder	Stimulant-induced psychotic disorder
Stimulant-induced bipolar disorder Stimulant-induced depressive disorder	Stimulant-induced mood disorder
Stimulant-induced anxiety disorder Stimulant-induced obsessive-compulsive disorder Stimulant-induced sleep disorder Stimulant-induced sexual dysfunction Stimulant intoxication delirium	Stimulant-induced anxiety disorder Stimulant-induced obsessive-compulsive or related disorder Stimulant-induced impulse control disorder Stimulant-induced delirium hazardous use of stimulant

[a] ICD-11 Stimulant Categories:

1. Cocaine
2. Stimulants including amphetamines, methamphetamine, or methcathinone
3. Synthetic cathinones

EPIDEMIOLOGY

Lifetime cocaine use among Americans in 2019 was estimated at over 40 million people, nearly 2 million of whom had used it in the past month. The number of Americans who have used cocaine in the past year has remained relatively stable from 2015 to 2019, at 5.5 million, representing 2.0% of the population. While fluctuating somewhat year to year, estimates of methamphetamine and prescription stimulant misuse also remained relatively stable from 2016 to 2019. In 2019, past-year use of methamphetamine was estimated at 0.7% of Americans (or 2.0 million people). Past-year nonmedical use of prescription stimulants in 2019 was estimated at 1.8% of Americans (or 4.9 million people). Over 1 million of these Americans had used methamphetamine, and 1.5 million had used prescription stimulants in the past month.

The worldwide burden of psychostimulant use speaks to an even more significant problem. The United Nations Office on Drugs and Crime (UNODC) World Drug Reports estimated that in 2021, roughly 36 million people worldwide had used amphetamine-type stimulants and 22 million people had used cocaine in the past year. Afghanistan and Iraq were highlighted in the report as areas with increasing methamphetamine use. The number of individuals using a substance for the first time in the past year offers a useful measure of potential trends in drug use. Initiation of cocaine use has remained relatively stable since 2013, with 671,000 using the drug for the first time in 2019. The average age at onset of cocaine use was 20.4 years. New onset of nonmedical use of prescription stimulants increased from 603,000 in 2013 to 901,000 in 2019. Likewise, new use of methamphetamines increased from 144,000 in 2013 to 184,000 in 2019.

The National Drug Early Warning System (NDEWS) follows regional trends in selected “sentinel” communities throughout the United States. The NDEWS 2020 report described an increase in methamphetamine-related overdose deaths in multiple sites, particularly in the Western United States, as well as in Florida, Philadelphia, and Maine. Cocaine-related overdoses rose in New York City and other regions. Cocaine was also frequently present in fentanyl overdoses, possibly reflecting the presence of fentanyl in cocaine products or the combined use of opioids and cocaine. Data from the National Center for Health Statistics “CDC Wonder” database underscores this concerning trend. Nationally, overdoses due to amphetamines skyrocketed from 547 in 1999 to 16,167 in 2019. Deaths related to amphetamines increased whether or not fentanyl was also present. Overdoses on cocaine increased from 3,822 in 1999 to 15,883 in 2019, with much of that accounted for by deaths also involving fentanyl. Furthermore, cocaine overdose deaths have disproportionately occurred in non-Hispanic Black people, among whom cocaine caused more deaths than did opioids in the period from 2000 to 2015.

The National Household Survey on Drug Use and Health (NSDUH) is one of the primary sources of knowledge of patterns and prevalence of drug use among people 12 and older in the United States. NSDUH information on the prevalence and distribution of stimulant use has been available since 1991, and measures of SUDs have been included in the questionnaire since 2000. Males are more likely to use cocaine and methamphetamine and are somewhat more likely to misuse prescription stimulants than are females. Analogous data are not yet collected on gender nonconforming people. Cocaine is used across ethnicities but inhaled “crack” cocaine use is somewhat more common among people identifying as Black or African American. Methamphetamine use has historically been more common among men who have sex with men (MSM). In this group, stimulants are sometimes used to enhance sexual activity and can be associated with risky sexual practices and sexually transmitted diseases. The annual prevalence of cocaine-type stimulant use disorder in the United States in 2019 was estimated at 0.4% of the population or 1.0 million Americans. The age group most heavily affected was 18 to 25 year olds, who remain twice as likely as the general population to have met criteria for cocaine use disorder in the past 12 months. A total of 1.0 million (0.4%) Americans met criteria for methamphetamine use disorder and 550,000 (0.2%) for prescription stimulant use disorder in 2019; again, the 18- to 25-year-old age group was the most heavily affected. The Treatment Episode Data Set (TEDS) is a component of the Drug and Alcohol

Services Information System (DASIS), a national monitoring network of drug and alcohol treatment in the United States. TEDS provides information on the numbers and characteristics of patients admitted for treatment of SUDs. Data from this survey indicate that admissions for primary treatment of cocaine use disorders decreased from 11.6% to 5.6% of admissions for SUD treatment from 2008 to 2018. In contrast, admissions for amphetamine-type stimulants increased from 6.3% to 10.5% of total admissions in that same time period. As a group, stimulants accounted for 16% of SUD admissions in 2018, following alcohol (29.4%) and opiates (33.6%). The Drug Abuse Warning Network (DAWN) survey of emergency department treatment presents a somewhat more alarming picture. The dangers of stimulant misuse are reflected in the disproportionately high number of emergency department visits caused by this class of substances. Preliminary data from the 2020 DAWN initiative indicate that, as a group, stimulants were the primary reason for 44.8% of substance-related emergency room visits, surpassing those caused by marijuana (22.0%), alcohol (20.9%), and heroin (17.9%). Methamphetamine was the most commonly implicated substance, accounting for 33.7% of substance-related episodes.

“Bath Salts”

Catha edulis, or “Khat,” is a plant indigenous to Eastern Africa and Southern Arabia. Psychoactive chemicals released from the plant include cathinone and cathine, compounds with stimulant properties. Khat is used like chewing tobacco, held between the cheek and gums after chewing. NIDA estimates that as many as 10 million people worldwide use Khat, including many in Eastern Africa and the Middle East, as well as an unknown number of immigrants to the United States.

In the mid-2000s, synthetic cathinones began to appear, marketed as “legal high” chemicals, the most common being mephedrone, methylone, and methylenedioxypyrovalerone (MDPV). They were packaged under labels including “screen cleaner,” “jewelry cleaner,” and most notoriously, “bath salts.” They are generally produced as powders marked “not for human consumption” and can be taken by oral, intranasal, inhaled (smoked), or intravenous routes. These compounds are substituted cathinones that have been chemically modified to avoid regulatory control. There are a large number of specific compounds in the class, and a given package often contains a mix of agents, further increasing risk of toxicity. Recently, synthetic cathinones have been found as adulterants or substitutes for methylenedioxymethamphetamine (MDMA) in products sold as “Molly.” In general, substituted cathinones increase synaptic catecholamine levels by inhibiting the dopamine, serotonin, and norepinephrine reuptake transporters in a fashion similar to cocaine. Most also facilitate catecholamine release, similar to methamphetamine. Animal models indicate strong reinforcing properties of these compounds, which are considered to have high addictive potential.

Drug screens for methamphetamine and PCP may cross-react with some substituted cathinones, but this is not a reliable test. While specific assays for some cathinones are commercially available, gas chromatography remains the most reliable means of detecting synthetic cathinones.

MDMA is another closely related substituted amphetamine with similar pharmacologic properties.

ETIOLOGY

Drug use disorders are believed to result from the complex interplay of cultural, environmental, familial, and neurobiologic influences. Environmental factors such as accessibility and perceived danger or stigma related to use may influence whether an individual initially tries psychostimulants or other drugs. For individuals with stimulant use disorders, environmental precipitants may increase cortisol and other stress-related hormones, increasing the risk for relapse or persistent use of drugs even when there is a desire to stop. Genetic factors may both increase the likelihood of initially using illicit substances and confer genetic vulnerability for developing SUDs.

Environmental influences can impact epigenetic factors, regulating the expression of genes in ways that affect risk of developing stimulant use disorders.

Mechanism of Action

Neurotransmitter Systems

DOPAMINE. It is well established that stimulants directly produce changes in dopaminergic neurotransmission. In vitro animal studies using radiolabeled cocaine or cocaine analogs demonstrate high amounts of binding in dopamine-rich brain regions, including the caudate-putamen and the ventral tegmental area (VTA). Moderate levels of cocaine binding also occur in the substantia nigra, amygdala, hypothalamus, and locus coeruleus. In humans, cocaine rapidly permeates the striatum, then quickly redistributes out of it with a half-life of about 20 minutes. The time course of self-reported "high" closely follows this pattern of rapid uptake and clearance. Likewise, amphetamines have profound effects on brain reward regions, including the mesolimbic regions (VTA and nucleus accumbens). Furthermore, methamphetamine is distributed even more diffusely and persists longer in the brain than does cocaine.

Dopaminergic projections from the VTA to the nucleus accumbens are most strongly identified with drug-related reinforcement. Cocaine administration causes transient increases in extracellular dopamine levels in these regions by binding to the presynaptic dopamine transporter (DAT), thereby inhibiting dopamine reuptake. Amphetamines not only inhibit dopamine reuptake but are also believed to reverse the function of the DAT. By this mechanism and through promoting the release of dopamine stored in vesicles, amphetamines potently increase synaptic dopamine levels. Transient changes in dopamine levels play a primary role in the reinforcing properties of all psychostimulants. Subjective euphoria appears to be a function of both DAT occupancy (with a threshold of 50% occupancy to be detectable) and rapidity of onset, with the strongest responses occurring when dopaminergic neurotransmission is increasing dynamically.

OTHER NEUROTRANSMITTER SYSTEMS. While it is widely accepted that dopamine plays a primary role in the effects of stimulants, more complex models are required to completely understand drug-related behaviors and disorders. By binding to monoamine transporters, psychostimulants can inhibit reuptake of serotonin and norepinephrine as well as dopamine. Other neurotransmitters implicated in stimulant use disorders include glutamate and the inhibitory neurotransmitter GABA. Glutaminergic projections to the nucleus accumbens modulate regional dopamine neurotransmission, and glutamate antagonists can block the behavioral effects of stimulants in laboratory animals. Similarly, GABA affects dopamine neurons in the nucleus accumbens and VTA. Reduced glutamate levels during abstinence that follows psychostimulant exposure lead to the upregulation of AMPA receptors in the nucleus accumbens. Such changes are associated with drug-related behaviors in lab animals, where glutamate antagonists and modifiers can block the reinstatement of drug-seeking behavior.

Learning and Memory

Positive Reinforcement Mechanisms. Stimulant use disorders and related behaviors can be understood as forms of behavioral conditioning. *Operant conditioning* is a type of learning in which the likelihood of a given behavior is controlled by its consequences. *Reinforcers* are those stimuli that increase the likelihood of the behavior in question. As such, stimulants and other substances of abuse operate as powerful reinforcers, increasing the likelihood of behaviors that result in opportunities to obtain drugs (i.e., “drug-seeking behaviors”).

Principles of operant conditioning explain why animals can learn to self-administer reinforcing substances like cocaine. Other drug-associated responses can be accounted for by *respondent conditioning*. In this form of learning, an unconditioned stimulus (one that produces a response without training) can be paired with a stimulus that normally does not produce a response. Pairing the two stimuli for a period of time produces a conditioned response, in which the animal responds to the initially neutral stimulus as though the unconditioned stimulus were also present. In this way, drug-associated cues such as paraphernalia, people, or places associated with use, and even times of day, can elicit urges to use stimulants. This phenomenon may also account for cocaine-dependent individuals who experience a subjective “high” while preparing to take drugs, as the practice of obtaining or handling drugs leads to conditioned euphoric responses.

During the early stages of a stimulant use disorder, drug use is reinforced by increased dopamine transmission from the VTA via the medial forebrain bundle to the nucleus accumbens. This fits a classic model of drug addiction based on positive reinforcement, in which cues associated with drug use gain emotional valence through learned associations involving brain regions such as the hippocampus and limbic structures (in particular, the amygdala). Abnormalities in these brain areas, which are involved in essential pathways of learning and memory, are associated with development of stimulant use disorders and other SUDs. Such dysfunction may explain persistent use of psychostimulants to the exclusion of other, natural reinforcers such as food. Conventional reinforcers produce a dopamine response in the VTA and amygdala that attenuates over time. In contrast, the response to stimulants does not diminish in this way, and some laboratory animals will self-administer cocaine in preference to food and water to the point of death.

Respondent conditioning explains the activation of drug cravings by stimuli associated with drug use. Cravings that are induced by priming doses of stimulants involve a brain circuit projecting from the VTA to the medial prefrontal cortex and the nucleus accumbens. Cravings that are induced by associated cues (“people, places, and things”) activate this circuit under influence of the basolateral amygdala and glutamatergic projections in limbic and prefrontal regions. It is also widely recognized that stress can elicit drug craving. This is mediated by expression of corticotropin-releasing factor in limbic areas that include the amygdala, modulating the activity of dopaminergic projections from VTA to nucleus accumbens.

Negative Reinforcement Mechanisms. In addition to reward mechanisms described above, a pattern of sustained stimulant use through negative reinforcement (avoidance of the aversive withdrawal/abstinence state) may augment and even supplant positive reinforcement. This is mediated by what may, in simple terms, be thought of as a “dopamine-deficient state” caused by dopamine receptor downregulation, decreased striatal dopamine release, and neurotoxic effects on dopaminergic neuron systems. Other neurotransmitters such as glutamate also show altered function with chronic stimulant misuse. This likely contributes to states of dysphoria and drug craving that occur with cessation of use. Indeed, it has been demonstrated that abstinent methamphetamine users who experience persistently low levels of striatal dopamine activity may be at increased risk for relapse. Lower availability of D2 and D3 receptors in methamphetamine users has also been posited to be associated with increased impulsivity and “steep temporal discounting,” the tendency to disproportionately value small, immediate reinforcers over delayed rewards of greater magnitude. Negative reinforcement could also play a role earlier in the course of the illness, in which attenuation of negative feelings, in addition to the pleasurable effects of the drug, maybe a strong contributor to the ongoing use of the substance.

Compulsive Use. In addition to stimulant use that is compelled by positive and negative reinforcement, stimulant use disorders are marked by what is sometimes termed “compulsive” or “habitual” use. This involves apparently automatic, perseverative drug use even in the face of severe consequences and long after many patients describe loss of previous euphorigenic drug effects. In recent years, preclinical and basic research have elucidated the complexity of stimulant use and its impact on the mammalian brain. Beyond a straightforward model of dopamine stimulation and resulting hypofunction, it is now clear that there are many downstream and alternate consequences of stimulant misuse. Changes in the following pathways are associated with chronic, compulsive stimulant use.

CAMP RESPONSE ELEMENT–BINDING PROTEIN. Dopamine receptors act through second-messenger systems that alter levels of cyclic adenosine monophosphate (cAMP). Dynamic interplay between dopamine receptor subtypes leads to alterations in cAMP, which in turn can increase activity of protein kinase A (PKA). PKA levels in the nucleus accumbens are increased with chronic stimulant administration. Elevations in PKA lead to increased activity of cAMP response element–binding protein (CREB), itself implicated in cocaine-related behaviors, including place preference and sensitization.

GENE TRANSCRIPTION. Cocaine and methamphetamine affect gene transcription through transcription factors such as ΔFosB, whose downstream targets include genes that influence synaptic plasticity, and CREB, among others. CREB signaling has been implicated in a host of cellular processes believed to underlie SUDs, including pathways involving the “antireward” gene dynorphin, cell adhesion, synaptic plasticity, and brain-derived neurotrophic factor (BDNF, itself a candidate for impacting cellular mechanisms of addiction). The CREB pathway is also believed to affect function of glutamate, potassium, and GABA receptors that contribute to long-term effects of stimulant misuse.

EPIGENETICS. While various definitions of epigenetics have been put forward, it generally describes the processes by which minor deoxyribonucleic acid (DNA) modifications (such as methylation) and the physical winding of DNA around histones influence whether genes are transcribed. It is clear that these processes are altered in SUDs and that such changes contribute to the process of addiction. Though not proven, there is some evidence suggesting that epigenetic information created by events such as parental trauma, substance use, or other adversity can be passed on to offspring, where it may influence the likelihood of SUDs in future generations.

MICRORNA. Gene expression is also regulated at the level of translation, the process by which ribonucleic acid (RNA) is used to code proteins. Small, regulatory RNA strands called microRNAs are known to function as inhibitors of translation by interacting with transcripts, leading to suppression or degradation of coding RNA. Stimulants alter the production of microRNAs, which themselves regulate BDNF, FosB, and CREB. A microRNA transcript called MiR-212, for instance, has been implicated in the alteration of gene expression by cocaine.

OXIDATIVE STRESS, INFLAMMATION, AND NEUROTOXICITY. Methamphetamine has been shown to promote brain inflammatory responses by activating astrocytes and microglia, and by leading to the production of inflammatory cytokines. Preclinical studies suggest that anti-inflammatory drugs can attenuate behavior, reflecting drug reinforcement in rodents. Methamphetamine also has a neurotoxic effect due to the production of highly reactive chemicals, including peroxide and free radicals. It may also lead to excitotoxicity by causing excess activity of glutamate. These effects likely contribute to executive and cognitive dysfunction, stimulant-related psychosis, and an increased likelihood of Parkinson disease in those using methamphetamine.

NEUROENDOCRINE DYSREGULATION AND STRESS. The role of environmental stress in exacerbating or precipitating episodes of mental illness has long been recognized. Individuals with SUDs may experience relapse or increased symptoms in the presence of psychosocial stressors. Animal models of stress and addiction demonstrate that stress-inducing situations (e.g., a shock to the foot) or administration of corticosteroids can increase the likelihood of drug self-administration. Conversely, antagonists of corticotrophin-releasing hormone (CRH) or glucocorticoids decrease stress-induced drug self-administration in laboratory animals by blocking

CRH activity in the VTA. Drug withdrawal itself is associated with increases in ACTH, corticosterone, and CRH in the amygdala. These changes are associated with the anxiety and craving often seen in early abstinence from stimulants.

COGNITIVE/EXECUTIVE DYSFUNCTION. Habitual drug use involves a pattern of executive dysfunction, poor impulse control, and cognitive inflexibility resulting in difficulty attributing appropriate reward values to stimuli, changing behavior in response to environmental feedback, and inhibiting impulsive or habitual behavior. Chronic stimulant use is associated with anatomic, metabolic, and vascular changes in areas of the brain responsible for such higher-order cognition. Structurally, cocaine and methamphetamine use disorders are associated with reduced gray matter density in the frontal cortex. Other findings include gray matter deficits in limbic areas and low hippocampal volume. Conversely, enlarged basal ganglia (areas important for neurotransmission of dopamine) have been demonstrated with stimulant use disorders. A dose-dependent decrease in gray matter volume of the orbitofrontal cortex is associated with increased compulsivity of drug use. Abnormal connectivity in prefrontal regions is evident in both individuals with stimulant use disorders and their unaffected siblings, suggesting a condition that pre-exists drug exposure. Diminished connectivity between the anterior cingulate cortex (an area active during inhibition of impulses) and the amygdala is associated with an increased risk of relapse in those with cocaine use disorders. Stimulant users manifest decreased blood flow in the prefrontal cortex, with frequent users showing the largest deficits. Decreased metabolic activity seen in frontal regions may persist for months after cessation of use. Metabolic dysfunction in frontal regions, including the orbitofrontal cortex, may be related to the development of compulsive aspects of drug addiction. Individuals with stimulant use disorders show deficits in verbal, spatial, and recognition memory, functions that are generally attributed to the hippocampus. Other prominent deficits include impaired attention, decreased cognitive flexibility, impaired decision making, impulsivity, and poor delayed reward discounting. This pattern of neurobiologic vulnerabilities clearly contributes to persistence of maladaptive drug use, even among those for whom substance use causes great suffering.

Routes of Administration and Metabolic Factors

The pharmacokinetic properties of stimulants are strongly influenced by how they are taken. In particular, the time it takes a drug to initially appear in the blood and to reach its peak serum concentration is strongly related to its mode of administration. Smoking and intravenous use of cocaine produce the most rapid pharmacologic and subjective onset of action, as well as faster termination of subjective effects. Cocaine-induced euphoria is more strongly correlated with rate of increase in serum levels than with absolute amount of cocaine ingested. As such, the use of cocaine by these rapid routes is more likely to be associated with progression to compulsive use and cocaine use disorder. Similarly, injecting or smoking methamphetamine is associated with much faster penetration of the blood–brain barrier, on the order of seconds, in comparison to use by oral or intranasal routes. Injecting methamphetamine is typically associated with more severe SUDs and with poorer prognosis.

Cocaine is rapidly metabolized by liver and serum esterases, with a variable half-life on the order of 50 to 90 minutes. In humans, cocaine is catabolized by hydrolysis of its ester groups to benzoylecgonine and ecgonine methyl ester in approximately equal proportions. Neither has biologic activity in humans. In the presence of alcohol, cocaine is converted by liver esterases into ethyl cocaine, or cocaethylene. Cocaethylene produces cocaine-like pharmacologic and subjective effects, has a longer half-life, and may be more toxic than cocaine. Only about 1% to 5% of cocaine is excreted unchanged in the urine, with the rest excreted as metabolites. Urinary benzoylecgonine can generally be detected for 2 to 3 days after a binge, although extremely heavy users may have measurable urinary benzoylecgonine levels as long as 10 to 14 days after last use.

The half-life of amphetamine in adults is on the order of 13 to 14 hours. Amphetamine undergoes oxidation by liver enzymes (primarily cytochrome P450 2D6) into the active metabolites 4-hydroxyamphetamine and norephedrine and the inactive metabolite alpha-hydroxy-amphetamine. Further oxidation of active metabolites produces the inactive metabolite 4-hydroxy-norephedrine. About 30% of amphetamine is excreted unchanged in the urine, with higher rates of excretion with more acidic urine. The rest is excreted as metabolites. Amphetamine is typically detectable in urine for 3 to 5 days after binge use.

Genetic Factors

Twin studies have allowed an estimate of the genetic component of cocaine use disorder. Estimates from these studies indicate that stimulant use disorders are among the most heritable of psychiatric conditions, with the familial component of cocaine use disorder placed in the range of 60% to 80%. Similarly, a recent study estimated the heritability of amphetamine-type stimulant use disorder at 68%. Some of the apparent genetic susceptibility to cocaine and other stimulant use disorders may be explained by inheritance of traits associated with drug use. In particular, heritable personality characteristics such as a high degree of novelty seeking, stress reactivity, and sensation seeking have been linked with initiation of drug use. Both individuals with stimulant use disorders and their unaffected siblings show deficits in executive function, high levels of trait anxiety, and poor response control. Those with stimulant use disorders show more problems across these domains than do their siblings, suggesting both a genetic state predisposing to drug addiction and exacerbation of deficits with persistent stimulant use.

Given the polygenic nature of drug addiction and the contribution of environmental factors such as availability of drugs, identifying particular genes directly involved in the etiology of stimulant use disorders is challenging. Furthermore, associations between stimulant use disorders and particular genetic loci in one population may not apply to another. It is also likely that some genetic patterns are associated with both SUDs and other psychiatric conditions more generally as risk factors, rather than conferring risk for a specific SUD via a straightforward pattern of inheritance. For instance, there are shared genetic risk factors for schizophrenia and some SUDs. Likewise, the gene for the alpha-5 nicotinic receptor subunit (CHRNA5) is implicated in both tobacco use disorder and in cocaine use disorders. The complex nature of genetic pathways related to SUDs is underscored by another study, showing that the same allele encoding a subunit of the nicotinic acetylcholine receptor (CHRNA5) is associated with increased risk of tobacco use disorder, while it appears protective for cocaine use disorders.

Table 11.3–2 lists some known genes associated with stimulant use disorder. Rather than presenting a comprehensive, cohesive genetic pathway, this list illustrates the complex picture that is gradually beginning to emerge, including transcription factors, genes governing epigenetic control mechanisms, monoamine receptors and transporters, and receptors for other neurotransmitters such as GABA, glutamate, opioids, cannabinoids, and nicotine.

DIAGNOSIS AND CLINICAL FEATURES

Stimulant Use Disorder

Stimulant use disorder is characterized by a maladaptive pattern of psychostimulant use accompanied by multiple drug-related problems in different domains. Stimulant use disorder is diagnosed when at least 2 of 11 possible criteria are met. As with other substances, tolerance and withdrawal are neither necessary nor sufficient for a diagnosis of stimulant use disorder. Criteria include (1) stimulant use in greater quantities or durations than intended, (2) persistent desire to curtail use, (3) significant time spent engaging in drug-related behavior, (4) craving for the desired stimulant, (5) recurrent use to the exclusion of other obligations, (6) persistent use despite social or interpersonal problems caused by use, (7) curtailing other important activities due to stimulant use, (8) use when it is physically hazardous, (9) use in spite of related physical or psychological harm it causes, (10) tolerance, and (11) withdrawal.

The category of substance that is misused (amphetamine type, cocaine, or other/unspecified) is named. The severity of stimulant use disorder is graded as mild (two to three symptoms), moderate (four to five symptoms), or severe (six or more symptoms). Early remission is defined by the absence of all symptoms for between 3 and 12 months and sustained remission by the absence of all symptoms for more than 12 months. The additional specifier “in a controlled environment” is used when the individual is in an environment where access to stimulants is limited.

Table 11.3–2.
Genes Associated with Stimulant Use Disorders

Genes Associated with Cocaine Use Disorder	**Genes Associated with Amphetamine Use Disorder**
SLC6A4 *Serotonin transporter*	FAAH *Fatty acid amide hydrolase*
CHRNB3, CHRNA3, CHRNA5 *Nicotine receptor subunits*	BDNF *Brain derived neurotrophic factor*
HRH3 *Histamine receptor*	MOCOS *Molybdenum cofactor sulfurase*
BCHE *Butyrylcholinesterase*	GABBR1 *GABA receptor*
DRD2, DRD3 *Dopamine receptors*	GRIA3 *AMPA receptor subunit*
FOSB/ JUN *Transcription factor*	GAD1, GAD2 *Glutamic acid decarboxylase*
ARC *Regulator of synaptic plasticity*	A2A *Adenosine receptor*
KCNJ9/GIRK3 *Potassium channel*	PROKR2 *Prokineticin receptor*
NR4A2 *Transcription factor*	GDNF *Glial cell line-derived neurotrophic factor*
MECP2 *Transcriptional repressor*	AKT1 *Protein kinase*
NDUFB9 *Mitochondrial subunit*	ARRB2 *Beta-arrestin*
C1qL2 *Regulator of glutamatergic connectivity*	COMT *Catecholamine-metabolizing enzyme*
KCTD20 *Component of potassium channel*	GTSP1 *Glutathione-S-transferase*
STK38 *Serine/threonine kinase*	OPRM1 *Mu-opiate receptor*
CANCA1C *Calcium channel*	NPYR1 *Neuropeptide 1 receptor*
OR3A1, OR3A2, OR1L4 *Olfactory receptor*	D4DR *Dopamine receptor*
SYT 1 *Regulator of synaptic vesicle exocytosis*	GABRA1, GABRG2 *GABA receptors*
LINC01411 *Non–protein-coding RNA*	P2D6 *Cytochrome*
TMEM5 *Transmembrane protein*	GLYT1 *Glycine transporter*
TRAK2 *Protein associated with GABA-A receptor*	PDYN *Prodynorphin*
SYNGR1 *Synaptogyrin*	PICK1 *Protein kinase C–associated protein*
	SLC22A3 *Cation transporter*

HTR2B *Serotonin receptor*
NR3C1 *Glucocorticoid receptor*
ADRA1A, ASDRA2A *Alpha adrenoceptors*
RGS17 *Regulator of receptor signaling via g-proteins*
5HT3A, 5HT3B, HTR2A *Serotonin receptors*
SLC6A4 *Serotonin transporter*
NCS-1 *Dopamine receptor–interacting protein*
OPRD1 *Delta-opioid receptor*
PDYN *Prodynorphin*
CNR 1 *Cannabinoid receptor*
PER2 *Circadian regulator*

While DSM-5 criteria for stimulant use disorder are identical to those for alcohol and opioids, patterns of use may be quite different. Although some individuals with stimulant use disorders are daily users who experience tolerance and withdrawal, many others abstain intermittently between periods of heavy use, either to recover from the effects of the drug or as a result of depleted finances that prevent access to the desired drug. During periods of use, frequency of administration varies based on the duration of drug effects. For instance, cocaine is generally taken more frequently than is methamphetamine, reflecting cocaine's shorter half-life.

Stimulant Intoxication

Stimulant intoxication is marked by the appearance of characteristic symptoms shortly after the use of an amphetamine-type substance, cocaine, or other stimulants. It may occur within minutes after ingestion of the drug. The duration of intoxication is typically briefer for cocaine than other stimulants, due to its shorter half-life. Characteristic symptoms are divided into psychological or behavioral and physiologic effects. Psychological symptoms may include changes in mood state, social interactions, or judgment. Physical symptoms include adrenergic effects (changes in heart rate, blood pressure, thermoregulation, psychomotor changes, pupillary dilation), gastrointestinal (GI) disturbances, motor changes, cardiopulmonary symptoms, and neurologic problems. Both a significant psychological disturbance and two or more physical symptoms are required for diagnosis. Severe intoxication can be associated with evidence of toxicity, which may include grand mal seizures, cardiac arrhythmia, hyperpyrexia, and death. The presence of perceptual disturbances is specified when drug-induced hallucinations or delusions occur during intoxication. If such perceptual disturbances are present and there is no insight that they are induced by the substance, the diagnosis of stimulant-induced psychotic disorder should be considered instead.

In susceptible individuals, paranoia tends to worsen during the course of a binge and to appear earlier in the course of subsequent binges as use continues—sometimes referred to as reverse tolerance or sensitization. Psychotic symptoms are common with both cocaine and methamphetamine. They may prompt attempts to "self-medicate" with alcohol or opioids. While cocaine users who experience delusions may complain of symptoms persisting beyond peak plasma cocaine levels, the duration is generally brief, lasting hours in most cases. Typically, paranoia consists of worries that police, drug dealers, or others are pursuing or planning to attack the cocaine user. Perceptual disturbances are usually limited to misinterpretations of existing stimuli (e.g., mistaking a car alarm for a police siren). Psychotic symptoms associated with methamphetamine use may be more difficult to distinguish from primary psychotic disorders. Symptoms caused by methamphetamine generally remit within a week of cessation of use. Persistent symptoms suggest the presence of an underlying psychotic disorder.

Stimulant Withdrawal

Stimulant withdrawal is common after periods of prolonged or intense use and is marked by dysphoric mood accompanied by two or more characteristic physiologic symptoms: fatigue, vivid and unpleasant dreams, and disturbances of sleep, appetite, and psychomotor activity. Bradycardia, drug cravings, anhedonia, and other depressive symptoms frequently occur.

There is substantial symptomatic overlap between characteristic physiologic symptoms of stimulant withdrawal and the neurovegetative symptoms of depression. At times, depressive symptoms would meet criteria for a major depressive episode were they not the direct consequence of drug use. In such cases, the diagnosis of stimulant-induced depressive disorder should be considered. Suicidal ideation or behavior can also occur in association with stimulant withdrawal and other stimulant-related disorders.

Stimulant-Induced Psychotic Disorder

Brief periods of stimulant-induced paranoia are common when people use stimulants heavily, occurring in 50% to 70% of individuals with cocaine use disorders and nearly 30% of those who use methamphetamine. While paranoia is common with heavy use of cocaine or methamphetamine, the diagnosis of stimulant-induced psychotic disorder is reserved for cases in which symptoms exceed the quality or severity characteristic of stimulant intoxication.

Increased risk of cocaine-induced paranoia is linked to heavy exposure to the drug, as measured by duration and quantity of use. Increased risk is also associated with routes of administration other than intranasal and with co-occurring ASPD or cannabis use disorder. Likewise, methamphetamine-related psychotic symptoms are more likely with heavy use. Psychotic symptoms are also associated with an increased risk of violence among individuals who use methamphetamine. People with ADHD appear at increased risk for either cocaine-induced or methamphetamine-induced psychotic symptoms. There may also be a genetic predisposition to stimulant-induced paranoia. Studies have identified increased levels of cocaine-induced paranoia or methamphetamine-induced psychosis in association with variations in several genes, as well as with alterations in epigenetic and microRNA regulation of gene expression.

For some, a more prolonged psychotic reaction may occur, and a subset of individuals who experience a stimulant-induced psychotic disorder will eventually receive a formal diagnosis for a psychotic disorder such as bipolar disorder or schizophrenia. As such, ongoing assessment and monitoring by a mental health professional may be warranted in such individuals.

Other Stimulant-Induced Disorders

Mood changes and anxiety are common with stimulant use, intoxication, and withdrawal. Stimulant-induced depressive, bipolar, or anxiety disorders are diagnosed when the severity of these symptoms warrants treatment and exceeds that usually seen with intoxication or withdrawal. Stimulant-induced mood disorders can resemble depressive, manic, hypomanic, or mixed episodes. Symptoms in the manic spectrum are usually present in the setting of intoxication, while depressive symptoms often occur during withdrawal. Suicidal ideation or behavior can also occur in association with stimulant withdrawal or stimulant-induced mood disorders. Compulsive symptoms such as stereotyped behavior and skin picking are also associated with stimulant intoxication.

Stimulant-induced mood or anxiety disorders are distinguished from mood and anxiety symptoms expected during the typical course of intoxication or withdrawal by their intensity, onset, and time course. While neurovegetative symptoms suggesting a major depressive episode may be present, full criteria for major depressive disorder or hypomania need not be met. Similarly, the diagnostic criteria for stimulant-induced anxiety or obsessive-compulsive disorder require only the presence of prominent anxiety, obsessions, compulsions, or panic attacks in the setting of recent use of a psychostimulant and out of proportion with typical signs of intoxication.

Other stimulant-induced disorders include stimulant-induced sexual dysfunction, sleep disorder, and unspecified stimulant-related disorder. Stimulants such as cocaine or methamphetamine are at times used for their purported enhancement of sexual function. Chronic use of stimulants is associated with sexual dysfunction either during use or within 1 month of use. Psychostimulants may also produce a sleep disorder, typically characterized by periods of insomnia during intoxication and hypersomnia during withdrawal. Insomnia during stimulant withdrawal is also commonly described. Sleep cycles may be shifted when binges extend into the night or when withdrawal-induced hypersomnia leads to daytime sleeping.

Unspecified Stimulant-Related Disorder

The category of unspecified stimulant-related disorder is used for disorders that do not meet the full criteria for a specific stimulant-related disorder but that nonetheless cause significant impairment or distress.

COMORBIDITY

Studies of the prevalence of non–substance-related psychiatric disorders among individuals with stimulant use disorders have consistently found elevated levels of psychiatric disorders in this population. Results from the community-based Epidemiologic Catchment Area (ECA) Study demonstrated elevated rates of non–substance-related mental disorders in those with cocaine use disorders. In this study, the estimated lifetime prevalence rate for psychiatric disorders was 76% or 11 times that found in the general population; nearly 85% of those with cocaine use disorders also had co-occurring alcohol abuse or dependence. Mood disorders, anxiety disorders, post-traumatic stress disorder (PTSD), ADHD, and psychotic disorders occur at elevated rates in people with stimulant use disorders. These high levels of co-occurring non–substance-related psychiatric disorders are not specific to stimulant use disorders, being observed among people with SUDs more generally.

Misuse of nonstimulants and co-occurring other SUDs are also common in people with stimulant use disorders, with alcohol and cannabis occurring at particularly high rates and opioid use on the rise in some geographic regions.

TOXICITY AND MEDICAL COMPLICATIONS

Stimulant use can cause medical complications both as a result of its direct toxic effects and because of complications related to methods of preparation and administration. Stimulants are potent sympathomimetic drugs, causing physiologic effects including vasoconstriction, increased heart rate, and elevated blood pressure through adrenergic stimulation. The combination of increased myocardial demand with constriction of coronary blood vessels can precipitate angina pectoris and even myocardial infarction. Methamphetamine can cause cardiomyopathy and cardiogenic shock. Acute hypertension associated with stimulant use can also predispose to arterial hemorrhage, such as hemorrhagic stroke or aortic dissection. Vasoconstriction can involve damage to other end organ, including the brain, kidneys, or intestines. With intranasal use, decreased blood flow to the nasal mucosa frequently causes nasal symptoms (e.g., rhinorrhea, nosebleeds) and mucosal ulceration. A combination of vasoconstriction and bruxism associated with the use of methamphetamine causes markedly poor dentition, or “meth mouth.” Grand mal seizures are common with stimulant overdose and may occur with routine use as well. Furthermore, individuals who misuse stimulants regularly may develop neuropsychological deficits, including problems with attention, learning, and memory. In many cases, these abnormalities persist during periods of abstinence. Furthermore, long-term misuse of amphetamines is associated with an increased risk of Parkinson disease later in life.

Stimulant misuse can also precipitate behaviors that place an individual at increased risk of harm by violence or accidental injury. Unsafe sexual behaviors may lead to transmission of HIV or other sexually transmitted diseases. Driving while intoxicated and other hazardous behaviors may lead to accidental physical harm. Finally, the associated illegal behaviors involved in obtaining cocaine or methamphetamine increase the likelihood of being a victim of violence.

Pathology and Laboratory Examinations

The physical and laboratory assessment of patients with stimulant-related disorders focuses on clarifying the diagnosis, excluding other possible diagnoses, and assessing for co-occurring problems or related complications. Metabolites of cocaine and amphetamine can be detected in blood, hair, sweat, saliva, and urine. In clinical practice, urine and blood testing are most useful. Blood testing can be helpful to assess for acute intoxication when very recent ingestion is suspected. Urine testing is often used as a means of monitoring patients for relapses and detecting concurrent use of other illicit substances.

Relative to other substances of abuse, the short half-life of cocaine renders it more difficult to discover occasional use with urine drug screening. On the other hand, with heavy use of cocaine, its breakdown products accumulate in fat tissue. As they gradually diffuse out of this body compartment, cocaine metabolites may be detectable for up to 2 weeks after cessation in some extremely heavy users. Methamphetamine is generally detectable for longer periods of time than is cocaine, on the order of 3 to 5 days after brief use. Synthetic cathinones present more of a challenge, as many are not detectable by routine drug screening. Gas chromatography may be necessary when cathinone ingestion is strongly suspected.

Other examinations are guided by historical or physical evidence of stimulant-related complications. Seizures or other acute neurologic deficits may prompt assessment by brain computed tomography (CT) or magnetic resonance imaging (MRI) to detect physical abnormalities such as hemorrhagic or ischemic strokes. The presence of chest pain or other cardiovascular abnormalities may prompt assessment with an electrocardiogram (ECG) or an assay for serum cardiac enzymes to evaluate for coronary ischemia. Similarly, respiratory symptoms may indicate need for a chest x-ray to rule out pneumonitis. Likewise, signs of infection in the setting of injection use may require blood cultures to exclude bacteremia or endocarditis. A history of any drug use by injection places individuals at risk for hepatitis (type B or C) and HIV; testing for these co-occurring medical disorders is thus often indicated.

As described above, chronic stimulant misuse produces changes in levels of stress-related hormones, including CRH and cortisol. Stimulant use can also elevate levels of prolactin, luteinizing hormone, and follicle-stimulating hormone. However, there is little clinical utility to assessing circulating hormone levels in individuals with stimulant use disorders, and such assessments generally are not part of routine care.

DIFFERENTIAL DIAGNOSIS

There is substantial overlap between symptoms of stimulant intoxication or withdrawal and other psychiatric diagnoses. Symptoms suggestive of a manic episode or of a psychotic or anxiety disorder are common during stimulant intoxication. Likewise, stimulant withdrawal frequently involves symptoms of depression and anxiety. Symptoms that are more severe than usually observed suggest a stimulant-induced psychiatric disorder rather than simple intoxication or withdrawal. Typically, stimulant-induced symptoms will attenuate or remit over several weeks with cessation of use. Symptoms that persist following a 1-month period of abstinence or that precede initiation of stimulant use suggest an independent psychiatric disorder.

Cycles of binge use followed by variable periods of abstinence may lead to mood instability; periods of decreased need for sleep, agitation, and other symptoms suggestive of mania may alternate with periods of low energy, dysphoria, and hypersomnia. When a patient presents with such a symptom pattern, which resembles bipolar spectrum illnesses, a careful history and laboratory evaluation to exclude use of cocaine or other psychostimulants is necessary for accurate diagnosis.

To some extent, the quality of symptoms may provide clues regarding whether they are manifestations of stimulant toxicity or a co-occurring psychiatric disorder. For instance, psychotic symptoms caused by cocaine use generally involve nonbizarre delusions or misperceptions. Tactile hallucinations (“cocaine bugs”) also suggest cocaine-related psychosis.

History and laboratory testing can also distinguish stimulant-related symptoms from those induced by other psychoactive substances. In particular, phencyclidine (PCP) and stimulant intoxication share common features.

COURSE AND PROGNOSIS

Severity of stimulant use disorder has been among the most consistently observed predictors of unfavorable outcomes. Surrogate markers of more severe stimulant use disorders, such as heavy baseline use, positive urine toxicology screen prior to treatment entry, withdrawal symptoms, and injection drug use, are all negative prognosticators. In a recent study, fMRI patterns of functional connectivity in the brain predicted treatment response, particularly when baseline cocaine use was taken into account. Other factors associated with better or worse outcomes are presented in Table 11.3–3.

Co-occurring mental illness may also present particular challenges to patients who struggle with stimulant use disorders. Among individuals with severe mental illness treated in a contingency management program, better prognosis was associated with lower levels of psychiatric severity, negative drug screens at onset of treatment, and use of treatment services prior to entering the study. Depression appears to denote a worse prognosis for cocaine use disorder, though some results have contradicted this conclusion. A study demonstrated that anhedonia accounted for the apparently worse outcomes associated with depression more than depression or neurovegetative symptoms themselves. While PTSD may not predict more cocaine use in treatment, it is associated with more persistent impairment overall and with worse outcomes on psychosocial measures related to drug use. Two personality characteristics, impulsivity and novelty seeking, are associated with early treatment dropout, as is cognitive impairment.

Table 11.3–3.
Prognostic Factors for Stimulant Use Disorders

Associated with Worse Prognosis	Associated with Better Prognosis
Higher baseline use	Self-help involvement, sponsorship
Injection use	Greater social support
High cortisol levels	Involvement in treatment
Cognitive dysfunction	Employment
Tobacco use disorder	Extended residential treatment (cocaine)
Lower education level	Favorable brain connectivity patterns on fMRI
Other drug use	Higher self-efficacy
Anhedonia	
Impulsivity	

TREATMENT

Intensity of Treatment

Often, the treatment of alcohol, benzodiazepine, and opioid use disorder may be initiated in hospital or detoxification units, where severe or complicated withdrawal can be closely monitored and managed. In contrast, the treatment of individuals with stimulant-related disorders is often begun in outpatient settings, in part because the withdrawal syndrome rarely requires specific medical intervention. This presents an obstacle to many patients struggling to break free of compulsive stimulant use in home environments that are not conducive to remission.

In addition to clinical considerations, treatment setting and intensity may be influenced by a host of other factors such as the availability of treatment resources, financial considerations, and patient preference. Factors that may prompt initiation of treatment in an inpatient or residential setting include medical or psychiatric conditions independently warranting intensive treatment, a lack of response at less intensive levels of care, or the need to remove an individual from environmental influences that may sustain drug use.

The severity of co-occurring problems must also be taken into account when selecting a treatment setting. Patients with persistent psychotic or severe depressive symptoms involving suicidal ideation may require inpatient treatment to stabilize their co-occurring condition before outpatient care can be initiated. Inpatient treatment may also be appropriate for patients with multiple drug dependences to provide safe detoxification from alcohol, sedatives, or opioids. Finally, some patients may require medical stabilization for complications of stimulant use such as systemic infection, cerebrovascular accidents, or myocardial infarction and arrhythmia.

Treatment of Intoxication and Withdrawal

There are no reliably effective specific pharmacologic treatments for cocaine or amphetamine intoxication or withdrawal, and supportive care is the mainstay of treatment for these conditions. While self-limited, stimulant intoxication can be complicated by toxic reactions, including paranoia, agitation, seizures, hypertension and its sequelae, and end-organ ischemia. Agitation or paranoia may call for intensive monitoring in a controlled setting and short-term treatment with benzodiazepines or antipsychotic medications. However, most patients will recover spontaneously in a matter of hours to days without the need for pharmacologic therapy. Unlike withdrawal syndromes described for alcohol, benzodiazepines, and barbiturates, stimulant withdrawal is generally milder, self-limited, and not typically associated with delirium or other medical complications.

Psychosocial Treatments

Psychosocial treatments remain the cornerstone of treatment for stimulant use disorders, particularly given the lack of a clearly effective medication to treat these disorders. Three behavioral models (contingency management, cognitive behavioral therapy [CBT], and twelve-step facilitation [TSF]) have received the strongest empirical grounding for the treatment of stimulant use disorders. Most study of 12-step approaches has focused on treatment of cocaine use disorders.

Contingency Management. Contingency management is a treatment firmly grounded in basic behavioral principles. In this approach, stimulant use disorders are conceptualized as the result of the drugs' powerful reinforcing effects, which supplant other natural reinforcers and aversive conditions that might otherwise halt drug use. In contingency management, competing reinforcers are created and applied to increase the likelihood of desired, nondrug behaviors. In most cases, the behaviors that are reinforced include treatment participation and the provision of drug-free urine samples during testing. Most commonly, contingency management procedures are combined with other active treatments such as comprehensive behavioral therapy programs and/or pharmacotherapy for co-occurring opioid use disorders.

An intermittent reinforcement schedule is commonly employed, in which an element of chance is introduced so that rewards are provided on a variable schedule. In one version of this method, study participants who engage in targeted nondrug behaviors (such as providing drug-free urine samples) are provided chances to draw for prizes that vary in value. This strategy has the advantage of providing robust reinforcement of desired behaviors and decreasing the overall cost of prizes (to an average cost of $200 or less per patient for 12 weeks of treatment). Beneficial results of such an approach have been demonstrated for patients with co-occurring cocaine and opioid use disorders in methadone-maintenance programs as well as for patients with cocaine use disorders treated in community settings. Positive effects of intermittent, abstinence-based reinforcement include increased likelihood of drug abstinence and improved treatment retention. Contingency management has also shown efficacy in other populations with cocaine use disorders, including those with co-occurring ASPD, pregnant women, people with HIV, those exposed to tuberculosis, and those with high levels of psychiatric symptoms. Clinical trials have confirmed similar benefits of contingency management in the treatment of methamphetamine use disorders, including in MSM and in those with co-occurring psychiatric disorders.

A recent, large network meta-analysis of psychosocial treatments for stimulant use disorders concluded that contingency management was consistently superior to treatment as usual on the outcomes of abstinence and retention in treatment. Further, contingency management in conjunction with community reinforcement (providing social rewards and support for treatment-related goals) was superior to other treatments assessed, including cognitive behavioral and 12-step–based approaches.

Despite the many trials consistently demonstrating efficacy of contingency management in treatment of people with stimulant use disorders, most patients are unable to access this potentially beneficial treatment, as it is simply not available in many communities. The reasons for this are complex and multifactorial, including a moralistic desire not to “pay patients to get better,” overestimating the actually low cost of the treatment, and logistic and regulatory difficulties of implementation. The end result is that an important tool for helping people with disorders for which too few beneficial options exist is vastly underutilized, to the detriment of patients.

Cognitive Behavioral Therapy. CBT for SUDs is predicated on the theory that maladaptive thought and behavior patterns sustain drug use in individuals with SUDs. CBT theoretically interrupts such patterns by helping patients to identify thoughts, feelings, and events that lead to drug use. CBT techniques for treating drug use disorders generally include providing psychoeducation and practice on skills such as avoiding or coping with high-risk situations, monitoring, and challenging drug-related thoughts and beliefs, and scheduling non–drug-related activities. Often, CBT treatments include written homework and other practice assignments between scheduled sessions. Homework adherence is associated with improved outcomes.

The efficacy of CBT in the treatment of cocaine and amphetamine use disorders has been demonstrated in several controlled trials. These trials include a computer-based adaptation for treatment of people with cocaine use disorders, as well as modifications for people of gender and sexual minority backgrounds. As described above, there is some evidence that CBT may be a less effective approach for stimulant use disorder treatment than contingency management, though further study including head-to-head comparisons is needed to definitively establish this. On the other hand, CBT offers the advantage of being much more widely available to patients. It is also a transdiagnostic approach, offering an evidence-based approach to treat the many co-occurring psychiatric and other SUDs that often accompany stimulant use disorders.

The Matrix model is a particularly well-studied example of a treatment approach that uses CBT as a core component. It is a manualized treatment employing elements of several evidence-based approaches, including CBT, family therapy, motivational interviewing, and other behavioral techniques. It has demonstrated efficacy in improving treatment participation and retention as well as in increasing the proportion of methamphetamine-free urine samples provided by patients with methamphetamine use disorders.

Other Psychosocial Treatment Approaches. Treatment that is based on the principles of 12-step recovery is common in clinical practice and has been studied in several well-designed trials. The 12-step model serves as the basis for self-help (often called mutual-help) groups such as AA, Narcotics Anonymous, Cocaine Anonymous, and Crystal Meth Anonymous. In addition to these peer-led groups, treatments based on 12-step principles are used by professionals in treating SUDs; the two most widely studied forms of professional treatment based on 12-step principles are individual drug counseling (IDC) and TSF; both encourage patients to attend 12-step meetings and to accept the disease model of addiction that serves as a core principle of 12-step philosophy. Both treatments have shown efficacy in clinical trials of the treatment of cocaine use disorders, particularly for people without co-occurring psychiatric illness. The body of literature regarding the impact of self-help groups in those with cocaine-related disorders is much smaller than that for alcohol-related disorders. There is also very little empirical data regarding the efficacy of 12-step and self-help approaches for amphetamine-type stimulant use disorders.

Motivational interviewing is useful for the treatment of SUDs in general, though it has been studied to a more limited extent for the treatment of stimulant use disorders. Exercise appears to benefit people with methamphetamine use disorders who do not use daily, but it should be considered a potentially useful component of therapy rather than a stand-alone treatment.

Pharmacologic Agents

To date, there are no FDA-approved medications to treat stimulant use disorders. Many agents that have shown promise in preclinical studies, open-label trials, or small double-blind studies have then had mixed or disappointing results in larger clinical trials. Nonetheless, accumulating evidence has recently allowed application of meta-analytic techniques to uncover approaches that show the most promise as future treatments. The following discussion describes potential areas for further research, though none of the medications described can yet be unequivocally recommended as a treatment.

Knowledge of the effects that cocaine and amphetamines have on brain dopamine has led to the investigation of many medications that modulate dopaminergic neurotransmission, including direct and indirect agonists as well as dopamine antagonists. Trials of direct dopamine agonists (e.g., amantadine, bromocriptine, L-dopa/carbidopa, pramipexole, pergolide) have largely been negative. Methylphenidate, which increases presynaptic dopamine release, has not shown consistent benefit for stimulant-dependent patients in the absence of co-occurring ADHD. The atypical psychostimulant modafinil has also shown little benefit.

A recent meta-analysis of psychostimulant treatment for people with stimulant use disorders reviewed the effect of treatment with amphetamines, methylphenidate, and modafinil on substance-related outcomes. Conclusions were that stimulant treatment likely showed benefit for abstinence-based outcomes, irrespective of the presence of co-occurring ADHD. These outcomes were largely driven by the impact of relatively high-dose treatment (>60 mg daily) with amphetamines for cocaine use disorders. The use of stimulants did not appear to improve retention in treatment. Concerns regarding the potential misuse or diversion of amphetamine-type stimulants remain an obstacle to widespread adoption as a treatment for people with stimulant use disorders.

Analogous to the use of the opioid antagonist naltrexone in the treatment of opioid dependence, dopamine antagonists have been assessed as potential medications for stimulant use disorders. Antipsychotics appear to be useful for the treatment of psychotic symptoms in people with stimulant use disorders, with some evidence that risperidone may be better tolerated than aripiprazole. However, antipsychotics do not show evidence of improving substance-related outcomes.

In addition to their effects on dopaminergic brain pathways, cocaine and amphetamines act as inhibitors of presynaptic serotonin reuptake. Serotonergic agents, including antidepressants, appear to blunt some of cocaine’s reinforcing properties in animal models. However, clinical trials of predominantly serotonergic agents (serotonin-specific reuptake inhibitors, buspirone [BuSpar], and venlafaxine [Effexor]) have been disappointing. Mirtazapine, an antidepressant with novel mechanisms including serotonergic effects, reduced both methamphetamine-positive urine samples and risky sexual behaviors in methamphetamine-use-disordered MSM. Bupropion, an antidepressant believed to have a primarily dopaminergic mechanism, improved methamphetamine use among those with lighter use at baseline.

While mood stabilizers as a class do not show evidence of improving substance-related outcomes in people with stimulant use disorders, topiramate has appeared beneficial in some clinical trials. In a 13-week clinical trial for cocaine use disorder, topiramate at a 200-mg-daily dose was associated with increased rates of abstinence. In a subsequent trial, 300-mg-daily topiramate in conjunction with CBT was associated with increases in number of abstinent days and the likelihood of cocaine-free weeks. In a trial involving people with methamphetamine use disorder, topiramate did not increase abstinence rates overall but was associated with decreased methamphetamine use in those who were abstinent at the start of the trial. In yet another study, the combination of mixed amphetamine salts and topiramate was associated with an increased rate of continuous abstinence over a 3-week period in comparison to placebo. A similar finding was seen in a study of topiramate for individuals with co-occurring cocaine and alcohol use disorders (AUDs). Despite these encouraging findings, three negative trials of the use of topiramate to treat cocaine use disorders have been published: two involving patients with co-occurring opioid use disorders and one with individuals using “crack” cocaine.

Disulfiram has long been used in the treatment of AUD because it inhibits acetaldehyde dehydrogenase, an enzyme involved in the metabolism of ethanol. This leads to unpleasant physical symptoms if ethanol is ingested. Initial interest in disulfiram for the treatment of cocaine use disorders was stimulated by the frequent co-occurrence of AUDs with cocaine use disorders. Early studies suggested that disulfiram had an indirect effect on cocaine use, mediated by its effects on alcohol use. However, disulfiram also appears to have a more direct pharmacologic effect on the metabolism of dopamine and cocaine itself; it acts as an inhibitor of dopamine beta-hydroxylase, thereby slowing the breakdown of synaptic dopamine and increasing dopamine levels. Disulfiram has also been shown to increase serum cocaine levels severalfold in pretreated persons. Results from several studies suggest that disulfiram may be of benefit to some people with cocaine use disorders, though many do not respond.

Pharmacogenetic studies have described several genes associated with differential responses to disulfiram for treatment of cocaine use disorder, including the alpha-1-adrenoreceptor gene (ADRA1A), ankyrin repeat and kinase domain-containing 1 (ANKK1), the D2 dopamine receptor (D2DR), the methylation enzyme encoded by MTHFR/C677T, variants of the serotonin transporter (5-HTTLPR), dopamine-beta-hydroxylase (DβH), and enzymes involved in serotonin synthesis (TPH2). Notably, allelic differences in DβH are also associated with differential response to doxazosin treatment of cocaine use disorder. Alleles producing normal or increased DβH activity are associated with disulfiram response, whereas alleles producing low activity are associated with doxazosin response.

In an effort to bolster therapeutic effects, medications thought to have complementary mechanisms of action have been combined. As described above, combination treatment with amphetamines and topiramate may be beneficial for treatment of cocaine use disorder. Building on earlier separate studies showing that naltrexone or bupropion might be beneficial for treatment of methamphetamine use disorder in those with lower baseline use, a study combining the two was recently completed. This trial involved subjects with moderate-to-severe methamphetamine use disorder who had used methamphetamine on 18 of the preceding 30 days. Compared to placebo, those treated with a combination of long-acting injected naltrexone and oral bupropion had higher rates of treatment response (measured as abstinence at consecutive time points). In another study of combined medications, buprenorphine along with long-acting injected naltrexone improved cocaine-related outcomes.

Investigational Treatments. A novel approach being studied for the treatment of stimulant use disorders is through immunization, using a therapeutic vaccine to induce antibodies to psychostimulants. Vaccines for cocaine and for methamphetamine are in different stages of development and testing. Creating an effective vaccine against either methamphetamine or cocaine is a challenge, given the small size of the compounds (too small to create an immune response independent of immunogenic carriers). It is hoped that more potent vehicles for delivering vaccines will yield better results in the future. Electrical stimulation of the brain, either through transcranial direct current stimulation (tDCS) or transcranial magnetic stimulation (TMS), is being explored as a possible treatment approach for stimulant use disorders. There is some evidence of potential benefit, with most studies assessing measures of executive functioning, craving, and impulse inhibition as target outcomes. Given the alterations in gene expression produced by heavy stimulant use, epigenetic modifiers have also been proposed as a new treatment paradigm. As knowledge of the genetic, cellular, and neurobiologic mechanisms of drug addiction are elucidated, it can be hoped that more effective biologic treatments for stimulant use disorder will someday be available to patients.

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders. 5th ed.* American Psychiatric Association; 2013.

Asharani PV, Hombali A, Seow E, Ong WJ, Tan JH, Subramaniam M. Non-pharmacological interventions for methamphetamine use disorder: a systematic review. *Drug Alcohol Depend*. 2020;212:108060.

Bali P, Kenny PJ. Transcriptional mechanisms of drug addiction. *Dialogues Clin Neurosci*. 2019;21(4):379–387.

Center for Disease Control and Prevention, National Center for Health Statistics. Drug overdose death rates. Accessed March 1, 2021. https://www.drugabuse.gov/drug-topics/trends-statistics/overdose-death-rates

Coffin PO, Santos G-M, Hern J, et al. Effects of mirtazapine for methamphetamine use disorder among cisgender men and transgender women who have sex with men: a placebo-controlled randomized clinical trial. *JAMA Psychiatry*. 2020;77:246–255.

Crits-Christoph P, Siqueland L, Blaine J, et al. Psychosocial treatments for cocaine dependence: National Institute on Drug Abuse Collaborative Cocaine Treatment Study. *Arch Gen Psychiatry*. 1999;56:493–502.

Crits-Christoph P, Wadden S, Gaines A, et al. Symptoms of anhedonia, not depression, predict the outcome of treatment of cocaine dependence. *J Subst Abuse Treat*. 2018;92:46–50.

Crunelle CL, Veltman DJ, Booij J, Emmerik-van Oortmerssen K, van den Brink W. Substrates of neuropsychological functioning in stimulant dependence: a review of functional neuroimaging research. *Brain Behav*. 2012;2(4):499–523.

De Crescenzo F, Ciabattini M, D'Alò GL, et al. Comparative efficacy and acceptability of psychosocial interventions for individuals with cocaine and amphetamine addiction: a systematic review and network meta-analysis. *PLoS Med*. 2018;15(12):e1002715.

Huggett SB, Stallings MC. Genetic architecture and molecular neuropathology of human cocaine addiction. *J Neurosci*. 2020;40(27):5300–5313.

Kampman KM. The treatment of cocaine use disorder. *Sci Adv*. 2019;5:eaax1532.

Limanaqi F, Gambardella S, Biagioni F, Busceti CL, Fornai F. Epigenetic effects induced by methamphetamine and methamphetamine-dependent oxidative stress. *Oxid Med Cell Longev*. 2018;2018:492453.

McKay JR, Van Horn D, Rennert L, Drapkin M, Ivey M, Koppenhaver J. Factors in sustained recovery from cocaine dependence. *J Subst Abuse Treat*. 2013;45(2):163–172.

Paillet-Loilier M, Cesbron A, Le Boisselier R, Bourgine J, Debruyne D. Emerging drugs of abuse: current perspectives on substituted cathinones. *Subst Abuse Rehabil*. 2014;5:37–52.

Regier DA, Farmer ME, Rae DS, et al. Comorbidity of mental disorders with alcohol and other drug abuse. Results from the epidemiologic catchment area (ECA) study. *JAMA*. 1990;264(19):2511–2518.

Roncero C, Daigre C, Grau-López L, et al. An international perspective and review of cocaine-induced psychosis: a call to action. *Subst Abus*. 2014;35(3):321–327.

Ronsley C, Nolan S, Knight R, et al. Treatment of stimulant use disorder: a systematic review of reviews. *PLoS ONE*. 2020;15(6):e0234809.

Salo R, Flower K, Kielstein A, Leamon MH, Nordahl TE, Galloway GP. Psychiatric comorbidity in methamphetamine dependence. *Psychiatry Res*. 2011;186(2-3):356–361.

Substance Abuse and Mental Health Services Administration. (2020). *Key substance use and mental health indicators in the United States: Results from the 2019 National Survey on Drug Use and Health* (HHS Publication No. PEP20-07-01-001, NSDUH Series H-55). Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/.

Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. *Treatment Episode Data Set (TEDS): 2018. Admissions to and discharges from publicly funded substance use treatment*. Substance Abuse and Mental Health Services Administration; 2020.

Tardelli VS, Bisaga A, Arcadepani FB, Gerra G, Levin FR, Fidalgo TM. Prescription psychostimulants for the treatment of stimulant use disorder: a systematic review and meta-analysis. *Psychopharmacology (Berl)*. 2020;237:2233–2255.

Trivedi MH, Walker R, Ling W, et al. Bupropion and naltrexone in methamphetamine use disorder. *N Engl J Med*. 2021;384:140–153.

United Nations Office on Drugs and Crime. *World Drug Report*. United Nations Publication, 2023

Vagins DJ, McCurdy J. Cracks in the system: twenty years of the unjust federal crack cocaine law. *American Civil Liberties Union*. 2006. Accessed March 1, 2021. https://www.aclu.org/other/cracks-system-20-years-unjust-federal-crack-cocaine-law

Volkow ND, Boyle M. Neuroscience of addiction: relevance to prevention and treatment. *Am J Psychiatry*. 2018;175(8):729–740.

Wang GJ, Smith L, Volkow ND, et al. Decreased dopamine activity predicts relapse in methamphetamine abusers. *Mol Psychiatry*. 2012;17(9):918–925.

Weiss RD, Mirin SM, Bartel RL. *Cocaine*. 2nd ed. American Psychiatric Press; 1994.

Weiss RD, Griffin ML, Gallop RJ, et al. The effect of 12-step self-help group attendance and participation on drug use outcomes among cocaine-dependent patients. *Drug Alcohol Depend*. 2005;77:177–184.

World Health Organization. (2023, January 24). ICD-11 International Classification of Diseases 11th Revision: The global standard for diagnostic health information. Https://icd.who.int/en

Yip SW, Scheinost D, Potenza MN, Carroll KM. Connectome-based prediction of cocaine abstinence. *Am J Psychiatry*. 2019;176(2):156–164.

Zhang X, Nielsen DA, Domingo CB, Shorter DI, Nielsen EM, Kosten TR. Pharmacogenetics of dopamine beta-hydroxylase in cocaine dependence therapy with doxazosin. *Addict Biol*. 2019;24(3):531–538.

11.4 Caffeine-Related Disorders

LAURA M. JULIANO, PH.D., AND MARY MARGARET SWEENEY, PH.D.

Caffeine (1,3,7-trimethylxanthine) is the most widely used mood-altering drug in the world. Caffeine is found naturally in more than 60 species of plants (e.g., coffee, tea, yerba mate, cacao) and belongs to the methylxanthine class of alkaloids, which also includes theobromine and theophylline. Caffeine is also added by manufacturers to various types of beverages, foods, dietary supplements, and medications. Approximately 90% of adults in the United States consume caffeine. Caffeine directly antagonizes the endogenous neuromodulator adenosine, as well as indirectly alters dopamine neurotransmission. Consequently, it affects various neurobiologic and physiologic systems and produces significant psychological effects. At typical dietary doses, caffeine is not associated with any life-threatening illnesses. However, caffeine use can be contraindicated for some medical conditions and medications, and discrete psychiatric symptoms and disorders can result from its use. The ubiquitous habitual consumption of caffeine and its integration into daily customs and routines can lead to an underestimation of the role that caffeine may play in caffeine users' daily experiences and can make recognizing caffeine-associated disorders particularly challenging. Clinicians need to be familiar with the various sources of caffeine and the psychiatric symptoms and disorders that can result from its use.

DEFINITION

Caffeine use is associated with five discrete syndromes in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5-TR) (Table 11.4–1). Four of these syndromes—caffeine intoxication, caffeine withdrawal, caffeine-induced anxiety disorder, and caffeine-induced sleep disorder—are formal diagnoses. A fifth syndrome—caffeine use disorder—is included as a condition for further study. The ICD-10 and ICD-11 also recognize caffeine-related disorders, including but not limited to caffeine intoxication, caffeine withdrawal, caffeine dependence (ICD-10 only), caffeine-induced anxiety disorder, and a harmful pattern of use of caffeine.

Table 11.4–1.

Caffeine-Related Disorders in the DSM-5-TR, ICD-10, and ICD-11

	DSM-5-TR	ICD-10	ICD-11
Caffeine intoxication	305.90	F15.0	6C48.2
Caffeine withdrawal	292.0	F15.3	6C48.3
Caffeine-induced anxiety disorder	292.89	F15.180 (with use disorder mild) F15.280 (with use disorder moderate/severe) F15.980 (without use disorder)	6C48.4
Caffeine-induced sleep disorder	292.85	F15.182 (with use disorder mild) F15.282 (with use disorder moderate/severe) F15.982 (without use disorder)	6C48.Z
Caffeine use disorder	Condition for further study	F15.2	6C48.Z[a]
Harmful use	N/A	F15.1	6C48.0 (episode) 6C48.1 (pattern)
Hazardous use	N/A	N/A	QE11.5
Disorders due to use of caffeine-unspecified	292.9	F15.9	6C48.Z
Other specified disorders due to use of caffeine	N/A	F15.8	6C48.Y

[a] No longer a discrete diagnosis in ICD-11.

HISTORY

Caffeine-containing foods and beverages have been consumed for thousands of years. Tea has been cultivated and consumed in China for more than 2,000 years, and coffee cultivation spread from Ethiopia to Arabia in the 15th century and subsequently became a popular beverage in Arabic cultures during the 16th century. Coffee spread to Europe in the 17th century (along with tea, tobacco, and chocolate). Medical literature during this time described coffee as a useful beverage for a wide variety of conditions, although it is notable that coffee use to induce sobriety was a particularly attractive feature. However, considerable controversy about whether the use of coffee was beneficial or detrimental to health accompanied its expansion into Western cultures, a controversy that is mirrored in contemporary concerns over the use of caffeine. The widespread use of coffee during the 17th and 18th centuries is reflected in the estimate that there were between 2,000 and 3,000 coffeehouses in London by the early part of the 17th century (about one coffeehouse for every 200 to 300 people). Perhaps the most famous of the coffeehouses in London was operated by Edward Lloyd. Like most coffeehouses, Lloyd’s was a center of business, especially for insurance agents, and by the end of the 18th century, Lloyd’s became the well-known Lloyd’s of London insurance company. Coffeehouses also served as a place for intellectual, political, and social discourse. In England, tea consumption eventually supplanted coffee use, with the shift in beverage preference occurring in the first half of the 18th century.

In the United States, a shift from tea to coffee use occurred in 1773 when colonists protesting British taxes threw cargoes of tea overboard in the Boston harbor—the Boston Tea Party. The repercussions of this event continue to the present day; the United States is now the major consumer of coffee in the world. Coffee, a major source of caffeine, is integral to the economic activity of several countries. For example, only crude oil produces more foreign exchange earnings than coffee for developing countries. Coffee is the major agricultural import of the United States and is second only in value to oil among all imports. This economic and trade activity underscores the extent to which caffeine is now available and accepted virtually everywhere in the world.

In the late 19th century, caffeine began appearing in various soft drink beverages. While a fraction of caffeine contained in cola soft drink beverages is derived from the kola nut, most caffeine in soft drinks is added—reportedly as a flavoring agent. However, there is little scientific evidence that caffeine functions as a flavoring agent at levels found in soft drinks. In 1997, Red Bull caffeinated energy drink was first introduced in the United States, with hundreds of other brands to follow. Energy drinks and energy shots now represent a multibillion-dollar market in the United States. These products are banned or regulated in some countries due to concerns about their high caffeine levels. In the United States, energy drinks and energy shots can be marketed as dietary supplements, which presently have looser regulations regarding caffeine content and labeling than soft drinks and drugs. In the early 2000s, caffeinated alcoholic beverages became available to consumers. By late 2010, amidst growing popularity of these beverages among young people, these products were removed from the market by manufacturers due to pressure from the U.S. Food and Drug Administration (FDA), who deemed caffeine added to alcoholic malt beverages an "unsafe food additive." At the present time, caffeine is being added to a wide variety of beverages, foods, dietary supplements, and medications. While the FDA considers how to better define the regulatory boundaries pertaining to caffeine, it is important for health care providers and consumers to remain cognizant of the potential sources of caffeine and its potential adverse effects.

COMPARATIVE NOSOLOGY

The DSM-III (1980) was the first to include caffeine (and tobacco) use disorders. At that time, only caffeine intoxication was included as a discrete syndrome with specific diagnostic criteria. Caffeine intoxication has remained in all subsequent editions of the DSM, with no substantive changes in the diagnostic criteria. Caffeine-induced anxiety disorder and caffeine-induced sleep disorder first appeared in the DSM-IV (1994). Caffeine withdrawal was included in the DSM-IV and DSM-IV-TR as a research diagnosis and was included as a formal diagnosis in the DSM-5 and DSM-5-TR with some modifications to the criteria based on a large body of empirical data. Caffeine use disorder was added as a condition for further study in the DSM-5 and DSM-5-TR, as there is growing evidence that some individuals have a profile of symptoms consistent with problematic caffeine use and may benefit from treatment.

The ICD-10, published by the World Health Organization, contains caffeine-related diagnoses under the category of mental and behavioral disorders due to use of other stimulants, including caffeine (F15), that include but are not limited to caffeine intoxication, harmful use, dependence syndrome, and withdrawal state.

The ICD-11 includes a category of disorders due to the use of caffeine (6C48), which includes caffeine intoxication, harmful use, caffeine withdrawal, and caffeine-induced anxiety disorder. Hazardous caffeine use is a new entry defined by caffeine use that increases the risk of harmful physical or mental health consequences warranting advice from health professionals but has not yet caused physical or mental harm to the user or others. Caffeine dependence syndrome is no longer explicitly included in the ICD-11 as a discrete diagnosis. Individuals with disordered caffeine use may display a harmful pattern of use of caffeine, which is described as having caused “clinically significant harm to a person’s physical or mental health” over a period of at least 12 months if use is episodic or for at least 1 month if use is continuous. However, this diagnosis does not recognize other hallmark features of a caffeine use disorder, including unsuccessful attempts or persistent desire to cut down caffeine or caffeine withdrawal. The ICD-11 also has a category of other unspecified disorders due to use of caffeine, which may be used to diagnose problematic caffeine consumption.

EPIDEMIOLOGY

Caffeine is consumed via beverages (e.g., coffee, tea, yerba mate, soft drinks, energy drinks, or workout supplements), foods (e.g., chocolate, coffee yogurt), dietary supplements, weight loss products, and prescription and over-the-counter medications. Approximately 70% of all soft drinks consumed in the United States have added caffeine, including many types of noncola soft drinks (e.g., orange soda, root beer). During the past two decades, caffeine-containing energy drinks, energy shots, and energy drink mixers (concentrated liquids or powders) have been made widely available to consumers. These products are capable of providing large doses of caffeine, as these products often contain significantly more caffeine per volume than soft drinks or at higher concentrations than recommended. Furthermore, some manufacturers add caffeine to foods that do not naturally contain caffeine (e.g., candy, potato chips, protein bars).

Determining an individual's level of caffeine exposure can be challenging because of the wide variety of products that contain caffeine, large differences in serving sizes, variability in caffeine content across products of the same type, and unknown or undisclosed caffeine amounts in some products. For example, a 12-oz cup of coffee would be estimated to contain approximately 200 mg of caffeine on average but may contain anywhere from 40 mg to 420 mg of natural caffeine depending on the variety, brew strength, and preparation method. A 12-oz caffeinated soft drink can vary from 19 to 69 mg of caffeine, and energy drinks/workout drinks can vary from 25 to 400 mg of caffeine for a can or bottle. There is also wide variability in caffeine levels across foods, medications, and dietary supplements/weight loss products. Table 11.4–2 shows the caffeine content of selected caffeine-containing products.

The most comprehensive epidemiologic data on caffeine use is derived from two main sources. The National Health and Nutrition Examination Survey (NHANES) is given annually to a representative sample of children and adults in the United States and collects information on health and nutrition including all consumed foods and beverages based on 24-hour recall. The Kantar Worldpanel

Beverage Consumption Panel includes a representative sample of the U.S. population who complete 7-day web-based diaries of beverage consumption. These surveys consistently find a high rate of regular caffeine use among children and adults, with approximately 85% of the population aged 2 years and older consuming at least one caffeinated beverage per day. It is notable that more than 60% of children aged 2 to 5 consume caffeine daily. The prevalence of caffeine use tends to increase with age, with approximately 90% of adult men and women consuming caffeine daily.

More than 95% of all caffeine ingestion comes from beverages. Coffee is the primary source of caffeine among adults in the United States, followed by tea and soft drinks. More than half of adults consume coffee every day, consuming an average of about 3 cups per day. Soft drinks are the primary source of caffeine among children and adolescents, followed by tea and coffee. A review of studies from 1999 to 2011 found increasing use of coffee and caffeinated energy drinks among children and adolescents, with the frequency of energy drink use varying considerably across studies. More recent studies suggest that energy drink use has continued to increase over time, with the largest increases among young adults. A review of 2003 to 2016 NHANES data from more than 33,000 respondents aged 12 to 59 found that the prevalence of energy drink consumption significantly increased over time across all age groups, and per capita consumption increased among young adults. The report also concluded that energy drinks were a major source of caffeine among those who consume them. A nationally representative survey of adolescents conducted in 2014 found that nearly two-thirds had reported ever trying energy drinks, and 41% had used energy drinks in the prior 3 months. It is important to note that sales of energy drinks more than doubled from 2014 to 2021.

Table 11.4–2.
Caffeine Content of Common Foods and Medications

Product	Serving Size (Volume or Weight)	Typical Caffeine Content (mg)	Range (mg)
Beverages			
Coffee			
Brewed/drip	12 oz	200	107–420
Instant	12 oz	140	40–260
Espresso	1 oz	70	60–95
Decaffeinated	12 oz	8	0–20
Starbucks drip	12 oz	260	
Starbucks cappuccino	12 oz	75	
Starbucks espresso	0.75 oz	75	
Starbucks bottled frappuccino	13.7 oz	110	
Starbucks decaffeinated	12 oz	20	
Tea			
Brewed	6 oz	40	30–90
Instant	6 oz	30	10–35
Canned or bottled	12 oz	20	8–32
Soft drinks			
Typical caffeinated soft drink	12 oz	40	22–69
Mountain Dew/Diet Mt. Dew	12 oz	54	
Diet Coke	12 oz	46	

Sunkist/Diet Sunkist	12 oz	19	
Dr. Pepper/Diet Dr. Pepper	12 oz	41	
Pepsi	12 oz	38	
Diet Pepsi	12 oz	36	
Pepsi Zero Sugar	12 oz	69	
Coke Classic	12 oz	34	
Barq’s Root Beer	12 oz	22	
A & W Root Beer	12 oz	0	
7UP/Zero Sugar 7UP	12 oz	0	
Fanta Orange	12 oz	0	
Sprite/Sprite Zero Sugar	12 oz	0	
A&W Cream Soda/Zero Sugar Cream soda	12 oz	0	
Cocoa/hot chocolate	6 oz	7	2–10
Chocolate milk	6 oz	4	2–7
Caffeinated water			
Typical amount	16.9 oz	60	60–200
Water Joe	23.7 oz	85	
Cafeina	16.9 oz	100	
Energy drinks/workout drinks			
Typical amount	varies	varies	25–400
Kill Cliff Energy X Recovery	12 oz	25	
Mountain Dew AMP	16 oz	142	

Red Bull	8.4 oz	80
Full Throttle	16 oz	160
Monster	16 oz	160
Rockstar	16 oz	160
Celsius energy drink	12 oz	200
Spike Shooter	12 oz	300
Redline Xtreme	8 oz	316
Wired-X-344	16 oz	344
Rage inferno	24 oz	375
Hyde Xtreme Pre Workout	12 oz	400
Energy shots		
Typical	varies	varies
5-hour energy	1.93 oz	200
5-hour energy extra strength	1.93 oz	230
7 eleven energy shots (various flavors)	2 oz	260
Energy Drink Mixers		
Pure Kick Energy Drink Mix Packets	0.10 oz	80
Vital 4U Liquid Energy	0.5 oz	155
Foods		
Chocolate		
Hershey’s Chocolate Bar	1.55 oz	9
Hershey’s Special	1.45 oz	20

Dark			
Kit Kat Wafer Bar	1.5 oz	6	
Reese’s Peanut Butter Cups	1.6 oz	4	
Miscellaneous foods			
Starbucks Classic Coffee Ice Cream	4 oz	30	
Dannon Coffee Yogurt	5.3 oz	30	
Jolt Caffeinated Gum	1 stick	45	
Powerbar Tangerine Powergel	41 g	50	
Military Energy Gum	1 stick	100	
Prescription Medications			
Headache/migraine/pain			
Fiorinal	2 capsules	80	
Fioricet/Esgic/many others	2 tablets	80	
Cafergot	2 tablets	200	
Norgesic	2 tablets	60	
Over-the-Counter Medications			
Stimulants			
Typical	1 tablet	100 or 200	100–200
Vivarin	1 tablet	200	
No-Doz	1 tablet	200	
Ultra Pep-Back	1 tablet	200	
Analgesics			
Goody’s Headache	1 powder packet	32.5	

Powder			
BC Original	1 powder packet	65	
BC Arthritis	1 powder packet	65	
Anacin Advanced Headache	2 tablets	64	
Excedrin Extra Strength	2 tablets	130	
Menstrual pain relief/diuretics			
Diurex Water Pills	2 tablets	100	
Midol Menstrual Complete	2 caplets	120	
Pamprin Max	2 caplets	130	
Dietary Supplements/Weight Loss Products			
Typical	1 or 2 tablets	varies	50–300
Dexatrim Max	1 caplet	55	
Hydroxycut Weight Loss Formula	2 caplets	200	
Leptopril	2 capsules	220	
Stacker 2	1 capsule	200	
Stacker 3	1 capsule	250	
Twinlab Ripped Fuel	2 capsules	220	
Swarm Extreme Energizer	1 capsule	300	
Xenadrine EFX	2 capsules	200	

Caffeine values for all brand name products were obtained directly from product labels, the manufacturer’s website, or customer service department when available.

Most caffeine is consumed in the morning, and exposure progressively decreases throughout the day. However, it is more common for caffeinated tea, soft drink, and energy drink consumption to be distributed throughout the day than coffee. Furthermore, adolescents tend to spread caffeine consumption across a day more so than adults. Weekday and weekend levels of caffeine use appear comparable across all age groups.

Determining the absolute level of caffeine exposure in the general population is challenging and estimates tend to vary widely depending on the study sample and methodology. There also tends to be a considerable lag in the time from data collection to publication of data summaries. A report was commissioned by the FDA that summarized the U.S. population's consumption of caffeine between 2003 and 2008. The report was based on the NHANES and NPD Group's Food Consumption surveys, which collected self-reported food consumption using daily diaries for 14 consecutive days from 2,000 households deemed representative of the population. The mean daily caffeine intake among adults (over age 22) was estimated to be 300 mg. Among coffee drinkers, the mean daily caffeine intake was estimated to be 375 mg. Another NHANES-based study estimated the caffeine intake of the U.S. population (>18 years) from 2001 to 2010. Total per capita consumption was 186 mg per day, and among caffeine consumers was 211 mg per day. Fourteen percent of the sample consumed more than 400 mg of caffeine per day. Across various studies, men tend to consume more caffeine than women. Non-Hispanic White individuals tend to consume more caffeine than non-Hispanic Black individuals, with Hispanic individuals falling in between.

ETIOLOGY

In the United States, it is not uncommon for the first exposure to caffeine, usually as coffee or soft drinks, to occur during childhood. Onset of regular use of soft drinks occurs at an earlier age than regular use of coffee or tea. Continued caffeine consumption may be influenced by several different factors, such as the pharmacologic effects of caffeine, caffeine's reinforcing effects, genetic predispositions, caffeine withdrawal, characteristics of the consumer (e.g., age, ethnicity), and social norms. About 14% of former consumers of caffeine report stopping caffeine consumption in response to health concerns or unpleasant side effects.

Pharmacology and Effects in the Central Nervous System

Caffeine is well absorbed from the gastrointestinal (GI) tract, with peak plasma concentrations typically occurring within 30 to 60 minutes after ingestion. Caffeine is readily distributed throughout all body fluids and is metabolized by the liver. More than 90% of caffeine is metabolized by the cytochrome P450, isozyme 1A2 (CYP1A2) system. Hence, caffeine is frequently used as a metabolic probe to assess CYP1A2 activity. Caffeine has three primary active metabolites—paraxanthine, theophylline, and theobromine—that also need to be considered in the overall biologic effects of caffeine.

The half-life of caffeine is 4 to 6 hours on average, with large variability across individuals. Caffeine metabolism is affected by various factors, including genetics, age, pregnancy, oral contraceptives, smoking, medications, diet, and diseases. The variability in CYP1A2 activity can be largely explained by a genetic polymorphism involving an A to C substitution. Carriers of the C allele (C/C or C/A) metabolize caffeine more slowly than an A/A genotype. Various other gene polymorphisms are also associated with caffeine metabolism (e.g., CYP1A1, CYP2A6, NAT-2, XO). Although caffeine half-life does not differ between younger and older adults, caffeine half-life is markedly increased in newborns (half-life of 80 to 100 hours) whose liver enzyme system does not completely develop until about 6 months of age.

Caffeine clearance slows down considerably during the second and third trimesters of pregnancy with more than a 50% reduction of CYP1A2 activity by the 24th week of pregnancy. Caffeine's half-life nearly doubles among women using oral contraceptives. There are various other medications that increase or decrease caffeine clearance. Caffeine's half-life increases when taking clozapine, fluconazole, methoxsalen, some quinolone antibiotics, cardiovascular drugs including verapamil, propranolol, and warfarin, and bronchodilators including theophylline. Grapefruit juice and liver disease also slow caffeine clearance. Cigarette smoking nearly doubles the speed of caffeine elimination due to the induction of liver enzymes by cigarette smoke (but not nicotine). Caffeine's half-life is decreased by medications including cimetidine, omeprazole, and fluvoxamine.

Caffeine exerts effects throughout the body, including bronchodilation (hence the therapeutic application of caffeine and the related methylxanthine theophylline in the treatment of asthma and apnea of prematurity); modest increases in blood pressure (which are attenuated in caffeine-tolerant individuals); increased production of urine; increases in gastric acid secretion; and increases in plasma epinephrine, norepinephrine, renin, and free fatty acids. Centrally, caffeine affects turnover or levels of various neurotransmitters and functions as a central nervous system (CNS) stimulant.

Numerous investigations have examined the effects of caffeine in the CNS to understand the neurobiologic basis for caffeine's effects. The principal cellular site of action of caffeine is the adenosine receptor, where caffeine functions as an antagonist at both A_{2A} and A_1 receptors. Adenosine serves an important modulatory role in many brain processes and functions (e.g., reduces spontaneous electric activity in the brain, dilates blood vessels in the brain). As an adenosine antagonist, many of caffeine's effects are opposite to those produced by adenosine (e.g., CNS stimulation, constricts blood vessels in the brain). Figure 11.4–1 shows the similarities in the molecular structures of caffeine and adenosine. There is evidence that adenosine receptors form functional relationships, or heteromerize, with other types of receptors, including dopamine and glutamate receptors. For example, some of the behavioral stimulant properties of caffeine are believed to be due to an increase in dopamine activity due to antagonism of adenosine at the A_{2A} to D_2 receptor heteromer. It has also been reported that caffeine increases dopamine release in the shell of the nucleus accumbens, an effect believed to be a neuropharmacologic mechanism underlying the addictive potential of drugs. Other mechanisms proposed as mediating caffeine's effects have included the inhibition of phosphodiesterases and effects on intracellular calcium, although these effects may only occur at doses of caffeine that exceed typical human consumption. However, it is not known if these other mechanisms may mediate some of the clinical effects produced when caffeine concentrations are elevated, such as may be the case during caffeine intoxication.

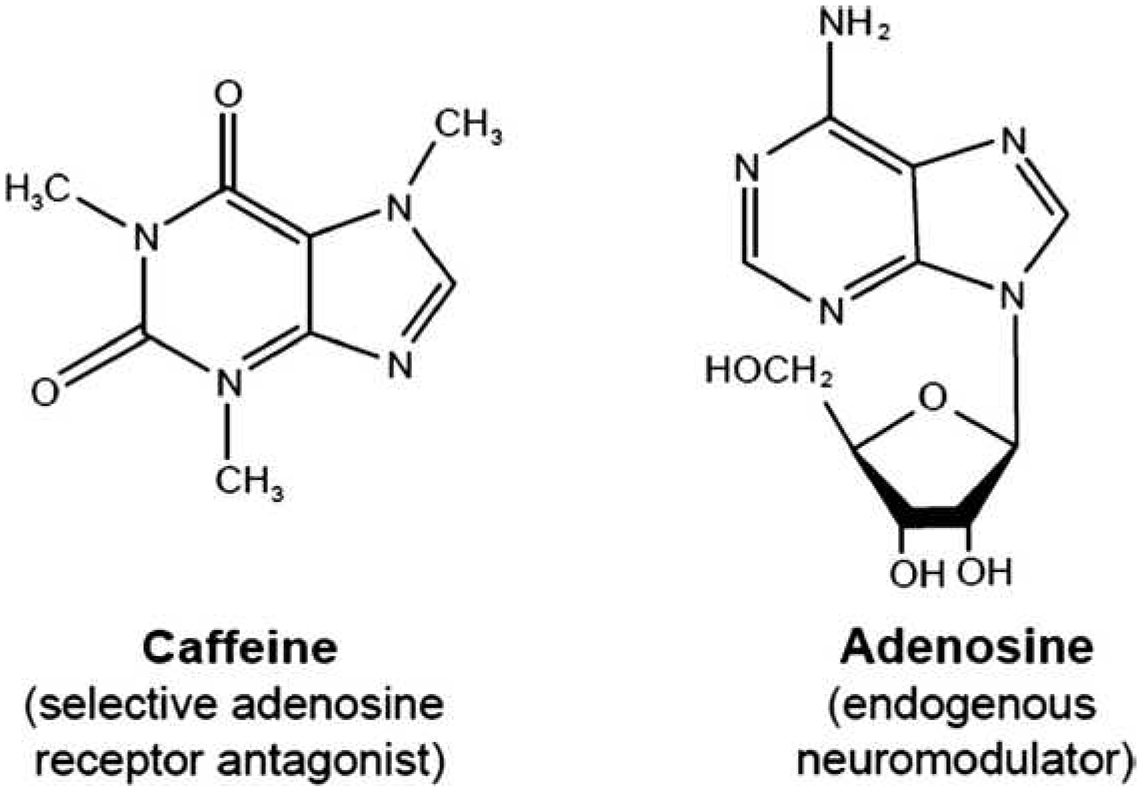
O
CH3
H3C
N
N
O
N
N
CH3
Caffeine
(selective adenosine
receptor antagonist)
NH2
N
N
N
N
HOCH2
O
OH OH
Adenosine
(endogenous
neuromodulator)

FIGURE 11.4–1. The chemical structures of caffeine (1,3,7-trimethylxanthine) and adenosine. Caffeine is structurally similar to adenosine, an endogenous neuromodulator. Most of the physiologic effects of caffeine, including the central nervous system stimulant effects, are believed to be mediated through adenosine receptor antagonism.

Subjective Effects, Performance Effects, and Reinforcement

Subjective reactions to caffeine are dose dependent and may vary depending on individual differences in sensitivity and caffeine tolerance. In general, single low-to-moderate doses of caffeine (i.e., 20 to 200 mg) produce pleasurable subjective effects. Studies have shown that such doses of caffeine result in increased ratings on measures such as well-being, alertness, energy, concentration, sociability, and motivation to work. In addition, these doses of caffeine produce decreases in ratings of feeling sleepy or tired. Higher doses of caffeine (e.g., ≥400 mg) are more likely to produce negative subjective effects such as anxiety, nervousness, and jitteriness; however, some individuals may report such negative effects at lower doses (e.g., 200 mg).

Numerous studies have examined the effects of caffeine on human performance. Recent research suggests that some of the variability in performance effects across individuals may be associated with single nucleotide polymorphisms (SNPs) on the CYP12A gene. There is compelling evidence that caffeine improves performance that has been degraded by fatigue, sleep deprivation, prolonged vigilance, or caffeine abstinence. Compared with placebo, caffeine has been shown to improve reaction time, tapping speed, sustained attention (or vigilance), and perhaps memory. A rich literature on the effects of caffeine on exercise performance suggests that, relative to placebo, caffeine (in the range of 3 to 6 mg/kg) reliably enhances performance during long-term (30 to 60 minutes) aerobic exercise, reduces ratings of perceived exhaustion, and improves power output or speed. Some studies, but not all, have also demonstrated caffeine-induced enhancement during short-term, high-intensity exercise and other aerobic and anaerobic sport-specific actions. The most reliable and strongest effects of caffeine on exercise performance are observed during aerobic exercise when caffeine (3 to 6 mg/kg) is consumed approximately 60 minutes prior. High doses of caffeine (e.g., 9 mg/kg) produce no additional benefit and result in negative side effects.

Although animal studies have generally found it difficult to demonstrate that caffeine functions as a reinforcer, well-controlled studies with humans have shown that people choose caffeine over placebo when given the choice under controlled double-blind experimental conditions. In habitual users, the reinforcing effects of caffeine are potentiated by its ability to suppress low-grade withdrawal symptoms after overnight abstinence. Thus, the positive subjective effects of caffeine and its ability to ward off the negative effects of caffeine withdrawal contribute to the regular use of caffeine. Studies have also shown that repeatedly pairing caffeine with a novel-flavored drink produces increased preference ratings for the flavor, while pairing placebo with a novel-flavored drink decreases preference ratings for the flavor among abstinent habitual caffeine consumers. These findings suggest that caffeine may play an important role in the development of preferences for specific types or brands of beverages.

Genetics and Caffeine Use

Investigations comparing coffee or caffeine use in monozygotic versus dizygotic twins show higher concordance rates for monozygotic twins for total caffeine consumption, heavy use, caffeine tolerance, caffeine withdrawal, caffeine intoxication, and coffee and tea intake, with heritability ranging from 34% and 77%. Multivariate structural equation modeling of caffeine use, cigarette smoking, and alcohol use suggests that a common genetic factor—polysubstance use—underlines use of these three substances. An analysis of monozygotic and dizygotic male and female twin pairs demonstrated that problematic use of caffeine and nicotine was substantially influenced by genetic factors unique to these licit drugs and distinct from genetic factors found to be common among illicit drugs. Furthermore, problematic use of caffeine was shown to have the greatest substance-specific genetic variance among the five drug dependencies that were evaluated.

Over the past 20 years, there have been numerous studies examining associations between candidate genes associated with the pharmacokinetics and pharmacodynamics of caffeine and patterns of caffeine consumption and caffeine-related effects. Most research has focused on polymorphisms in the adenosine A_{2A}

receptor gene (ADORA2A), the CYP12A enzyme gene that differentiates fast metabolizers and slow metabolizers of caffeine, and the aryl hydrocarbon receptor gene (AHR), which influences the transcription of CYP12A. ADORA2A polymorphisms have been shown to be associated with various outcomes including caffeine consumption, caffeine sensitivity, and caffeine's effects on psychomotor vigilance, anxiety, startle, sleep, coronary responses, postprandial glucose levels, and blood pressure. Adenosine A_{2A} receptor knockout mice show a reduced preference for caffeine compared to controls. Polymorphisms in the CYP1A2 and ACH receptor genes are most reliably associated with variability in caffeine consumption. Associations between the "fast metabolizer" CYP1A2 variant and better performance outcomes have been observed but are inconsistent across studies. The "slow metabolizer" CYP1A2 variant has been shown to be associated with increased risk for coffee-associated hypertension and myocardial infarction. In addition to the associations described above, recent research has identified dozens of other SNPs and possible associations with caffeine use and effects. However, a consistent pattern of findings across these types of studies has yet to be observed.

Age, Sex, and Race

Population-based surveys reveal that daily caffeine exposure increases with age up until middle age and then begins to slightly decline. Some studies show that men use more caffeine than women, but this effect disappears when accounting for differences in body weight (i.e., mg/kg), employment status, and other covariates. Among adults in the United States, non-Hispanic White individuals typically consume more caffeine than African American, Hispanic, and Asian individuals.

Special Populations

Cigarette smokers consume higher levels of caffeine than nonsmokers. This observation may reflect a common genetic vulnerability to caffeine use and cigarette smoking. It may also be due to faster caffeine elimination in cigarette smokers because of cigarette smoke stimulating the CYP1A2 enzyme system. Individuals who consume coffee or energy drinks consume more caffeine daily than those who do not. Preclinical and clinical studies indicate that regular caffeine use can potentiate the reinforcing effects of nicotine. Heavy use of alcohol, alcohol use disorders, and other substance use disorders (SUDs) are also associated with heavy use and problematic use of caffeine. One study found that among pregnant women, a family history of alcohol use disorders was associated with heavy use of caffeine both before and during pregnancy, despite physician advice to cease caffeine use during pregnancy. Another study also showed that a greater proportion of heavy caffeine consumers use benzodiazepines. High daily caffeine consumption has also been found among individuals with a variety of psychiatric disorders, particularly among those with schizophrenia. One twin study found significant positive relationships between caffeine use, caffeine toxicity and problematic caffeine use, and lifetime history of all seven psychiatric disorders that were assessed, including major depression, generalized anxiety disorder, panic disorder, alcohol use disorder, antisocial behavior, and cannabis and cocaine abuse. Interestingly, the authors concluded that shared familial factors accounted for the relationships between caffeine use and the psychiatric disorders, rather than a direct link. Some studies show that caffeine use is positively associated with novelty seeking or sensation seeking as well as legal and management occupations.

CAFFEINE INTOXICATION

Caffeine intoxication has long been recognized as a discrete syndrome associated with the use of a significant amount of caffeine. DSM-5-TR diagnostic guidelines state that the diagnosis is dependent on recent consumption well in excess of 250 mg of caffeine per day, but the syndrome is usually associated with higher doses (e.g., more than 1 g according to the ICD-11). Intoxication can be the result of an acute ingestion of a large amount of caffeine by a person who has not regularly consumed caffeine. Alternatively, caffeine intoxication can occur in a person who chronically consumes large amounts of caffeine, which produces a more complicated clinical picture. Interestingly, while intoxication of some psychoactive substances, such as alcohol, is sought out by some people, it appears that most individuals do not actively seek repeated episodes of caffeine intoxication (Table 11.4–3).

Epidemiology

There is an important need for data on the incidence and prevalence of caffeine intoxication in the general population. To date, most studies investigating caffeine intoxication symptoms have involved selected populations (e.g., psychiatric inpatients, college students) using ambiguous or incomplete criteria. Notably, one random-digit telephone survey among 162 caffeine users in Vermont found that two-thirds of respondents reported at least one caffeine intoxication symptom and 7% met the DSM criteria for caffeine intoxication in the past year. One study that assessed more than 3,600 twins found that 29% reported having experienced feeling ill or shaky or jittery after consuming caffeinated beverages. Psychiatric patients may have higher rates of caffeine intoxication than healthy controls. Children and adolescents may be at increased risk for caffeine intoxication because of low body weight, lack of tolerance, and lack of knowledge about the caffeine content of readily available products (e.g., energy drinks). Reports from the Poison Control Center's National Poison Data System provide some information about the rate of adverse effects from caffeine. For example, in the category of single-substance exposures (i.e., not involving caffeine contained in dietary supplements or energy drinks), there were 3,031 cases involving caffeine in 2019. Twenty-seven percent of these exposures were treated in a health care facility. Of all cases, 37% involved individuals aged 5 years and younger, and 57% involved individuals aged 19 years and younger. In 2019, there were 1,724 cases involving nonalcoholic caffeinated energy drinks, 596 cases involving unknown energy drinks, and 232 involving energy products. Seventeen percent of these cases were treated in a health care facility, and nearly 60% involved individuals aged 5 years and younger. The American Association of Poison Control Centers has issued an alert for energy drinks, stating that they can be dangerous for children or adolescents. The American Academy of Pediatrics recommends that children and adolescents do not consume energy drinks.

Table 11.4–3.
Caffeine Intoxication Symptoms

DSM-5-TR Symptoms
Restlessness
Nervousness
Excitement
Insomnia
Flushed face
Diuresis
Gastrointestinal disturbance
Muscle twitching
Rambling thoughts and speech
Tachycardia or cardiac arrhythmia
Inexhaustibility
Psychomotor agitation
Additional Symptoms
Fever
Irritability
Tremors
Sensory disturbances
Tachypnea
Headaches
Nausea/vomiting
Seizures
Dizziness
Delirium

Diagnosis and Clinical Features

Signs and symptoms of caffeine intoxication include restlessness, nervousness, excitement, insomnia, flushed face, diuresis, GI disturbance, muscle twitching, rambling thoughts and speech, tachycardia or cardiac arrhythmia, inexhaustibility, and psychomotor agitation. To receive a DSM-5-TR diagnosis, patients must present with five or more of the symptoms along with clinically meaningful distress or functional impairment. In addition, patients with caffeine intoxication can have fever, irritability, tremors, sensory disturbances, tachypnea, headaches, vomiting, and seizures. A series of case reports based on Poison Control Center data found that a particular energy drink containing 250 mg of caffeine was repeatedly implicated in caffeine intoxication incidents. The most common symptoms reported were hypertension, jitteriness, agitation, tremors, nausea, vomiting, and dizziness. Finally, although rare, it may be possible for a patient to develop delirium from ingesting an acute, extremely high dose of caffeine. Cases of intentional and unintentional fatal overdose from caffeine are rare but have been documented, usually from the ingestion of caffeine in the form of pills or powder.

Pathology and Laboratory Examination

Blood serum levels greater than 100 μg/mL may be lethal, although there have been reports of individuals surviving much greater caffeine serum levels with medical intervention. However, it has also been noted that serum caffeine levels may be a poor indicator of toxicity due to the variable relationship between blood concentration and toxic effects. Urine or blood drug screens can also be used to rule out the presence of other drugs in the body.

Differential Diagnosis

There are various disorders that should be included in the differential diagnosis of caffeine intoxication. These include other substance-related disorders, such as intoxication from other stimulants (e.g., cocaine or amphetamines), and withdrawal from drugs such as sedative–hypnotics (e.g., benzodiazepines), alcohol, or nicotine. In addition, other psychiatric disorders (e.g., anxiety disorders, mania, sleep disturbances), somatic disorders (e.g., arrhythmia, hyperthyroidism), and medication side effects (e.g., akathisia) should be considered in the differential diagnosis of caffeine intoxication. The final diagnosis of caffeine intoxication depends on confirming the ingestion of a significant quantity of caffeine prior to the onset of symptoms.

Course and Prognosis

Caffeine has a relatively short half-life (4 to 6 hours), and thus, caffeine intoxication usually resolves quickly and often with no significant sequelae. Some substances including oral contraceptives may significantly increase the half-life of caffeine and thus, increase the risk of caffeine intoxication. Patients with caffeine intoxication generally have a good prognosis, and fatal incidents are rare. The lethal dose of caffeine is generally considered to be about 10 g or the equivalent of 80 to 100 cups of brewed coffee.

Treatment

The first step in treating a patient presenting with symptoms suggesting caffeine intoxication is to confirm the diagnosis. It is important to consider all possible sources of caffeine when considering a diagnosis of caffeine intoxication, including medications, dietary supplements, energy drinks, shots, powders, and beverages that may not be initially recognized as containing caffeine. Because caffeine has a relatively short half-life, short-term management of the patient can be supportive while the syndrome resolves spontaneously. For the patient who habitually consumes large amounts of caffeine and repeatedly experiences episodes of caffeine intoxication, it may be helpful to assist the patient to recognize their total daily caffeine consumption using daily diaries, teach the patient about the adverse effects associated with caffeine, and provide treatment assistance (see Table 11.4–4). In the case of significant caffeine overdose, the patient should receive intensive monitoring, symptomatic treatment (e.g., for tachycardia), stomach aspiration, and assessment of serum caffeine level. Caffeine overdose has been successfully treated with hemodialysis or hemoperfusion.

CAFFEINE-INDUCED ANXIETY DISORDER

In addition to anxiety symptoms that are a feature of caffeine intoxication, caffeine consumption can precipitate an anxiety disorder. A caffeine-induced anxiety disorder can be panic disorder, generalized anxiety disorder, social phobia, or obsessive-compulsive disorder, although the patient does not need to fulfill all criteria for one of these disorders to quality for a diagnosis of caffeine-induced anxiety disorder. There has been no research examining this specific DSM-defined caffeine-related disorder, although there have been numerous investigations examining the anxiogenic effects of caffeine and case reports of caffeine-related anxiety disorders. In general, patients with anxiety disorders or anxiety sensitivity report greater anxiety than control subjects after consuming caffeine. Even among individuals who do not have an anxiety disorder, a sufficiently high dose of caffeine can produce panic attacks and increased anxiety ratings. First-degree relatives of patients with panic disorder also show an anxiogenic response to a high dose of caffeine. Polymorphisms in the *ADORA2A* gene are associated with anxiogenic responses to caffeine. Some studies have found that individuals with anxiety disorders, such as panic disorder, use lower levels of caffeine than healthy controls. However, some studies have failed to show a relationship between anxiety disorders and caffeine use, while others have found a relationship between greater caffeine use and anxiety disorders. It seems reasonable to conclude that some but not all individuals with anxiety disorders will spontaneously limit caffeine exposure, which may contribute to ambiguity within epidemiologic data associating caffeine consumption and anxiety disorders.

Table 11.4–4.
Caffeine Modification Guidelines

1. *Education*
   Patients should be educated about potential sources of caffeine. Provide patients with a list of common caffeinated products (c.f. Table 11.4–2). It may be helpful to point out that caffeine is present in some noncola soft drinks, foods, and over-the-counter medications and to emphasize the wide variability in caffeine levels within and across product types.
2. *Self-monitoring*
   Caffeine consumption should be self-monitored for 1–2 weeks to determine a baseline level. Calculate daily caffeine exposure in milligrams, considering the caffeine content of specific products, the serving sizes, and the number of servings. If a self-monitoring period is not feasible, treatment providers can determine a rough estimate of total caffeine consumption via patient report. Patients should continue to self-monitor during the caffeine reduction phase of treatment.
3. *Determine a caffeine modification goal*
   Some individuals may be interested in eliminating caffeine entirely while others may wish to reduce their caffeine consumption. Patients should be advised to set a goal of consuming no more than 50 mg per day or to use caffeine only occasionally (e.g., no more than 2 consecutive days) if they want to avoid withdrawal effects when they go without caffeine.
4. *Determine a dose reduction schedule*
   A gradual reduction schedule should help to minimize or prevent caffeine withdrawal symptoms. A reasonable decrease would be 25% of the baseline dose every week or so until the caffeine reduction or cessation goal is achieved. Reduction strategies can be based on caffeine dose in milligrams or number of beverages/servings, as long as all sources of caffeine in the diet have been considered. Caffeinated beverages can either be gradually omitted or can be mixed with decaffeinated beverages with the ratio of caffeinated product decreasing over time.
5. *Offer treatment advice and follow-up*
   Patients may benefit from behavior modification techniques that

are effective in the treatment of other substance use disorders (e.g., nicotine). Such techniques may include self-monitoring, coping response training, reinforcement for abstinence, identifying barriers to change, social support, and framing withdrawal as a temporary inconvenience. Follow-up with patients to check on their progress.

The diagnosis of caffeine-induced anxiety disorder depends on linking the use of caffeine to the anxiety symptoms of concern. For a patient with a suspected caffeine-induced anxiety disorder, a trial of caffeine abstinence may aid in clarifying the diagnosis. Caffeine-induced anxiety disorder should be diagnosed in patients with caffeine intoxication only if the anxiety disturbance is in excess of that which would be expected from intoxication. Other diagnoses to consider include caffeine intoxication, a primary anxiety disorder, and an anxiety disorder due to a general medical condition.

Mr B was a 28-year-old single African American male graduate student who was in good health and had no history of previous psychiatric evaluation or treatment. He took no medications, did not smoke cigarettes or consume alcohol, and had no current or past history of illicit drug use. His chief complaint was that he had begun feeling mounting “anxiety” when working in the laboratory where he was pursuing his graduate studies. His work had been progressing well, he felt his relationship with his advisor was good and supportive, and he could not identify any problems with staff or peers that might explain his anxiety. He had been working long hours but found the work interesting and had recently had his first paper accepted for publication.

Despite these successes, he reported feeling a “crescendoing anxiety” as his day would progress. He noted that by the afternoon, he would be experiencing palpitations, bursts of his heart racing, tremors in his hands, and an overall feeling of “being on the edge.” He also noted a nervous energy in the afternoons. These experiences occurred daily and seemed confined to the laboratory (although he admitted he was in the laboratory every day of the week).

When reviewing Mr B’s caffeine intake, it was found that he was consuming large amounts of coffee. Staff made a large urn of caffeinated coffee each morning, and Mr B routinely started with a large mug of coffee. Over the course of the morning, he would consume three to four mugs of coffee (the equivalent of about six to eight 6-oz cups of coffee), and continued this level of use throughout the afternoon. He occasionally had a single can of caffeinated soft drink and used no other forms of caffeine regularly. Mr B estimated he drank a total of six to eight or more mugs of coffee per day (which was estimated to be at least 1,200 mg of caffeine per day). Once pointed out to him, he realized this level of caffeine consumption was considerably higher than at any other time in his life. He admitted he liked the taste of coffee and felt a burst of energy in the morning when he drank coffee that helped him start his day.

Mr B and his physician developed a plan to decrease his caffeine use by tapering off caffeine. (Details of such a tapering schedule can be found in the section on treatment below.) Mr B was successful in decreasing his caffeine use and had good resolution of his anxiety symptoms once his daily caffeine use had been markedly decreased.

Treatment

There are no studies on the treatment of caffeine-induced anxiety disorder. However, one study reported significant improvements in anxiety symptoms among patients seeking treatment at an anxiety disorders clinic who had been instructed to cease caffeine use for 1 week. Notably, some patients required no additional treatment and had successful discontinuation of anxiolytic medications. In cases of suspected caffeine-induced anxiety disorder, clinicians may want to avoid the use of anxiolytic medications until caffeine use has been eliminated. Clinicians should also be aware that some patients may be hesitant about eliminating caffeine and may express skepticism about the role of caffeine in their anxiety symptoms. In such instances, clinicians should try to encourage patients to take an empirical approach and attempt a caffeine abstinence trial. Gradual reduction is recommended to avoid the aversive effects of caffeine withdrawal (see “Caffeine Withdrawal” section below). General guidelines for caffeine modification can be seen in Table 11.4–4.

CAFFEINE-INDUCED SLEEP DISORDER

Caffeine-induced sleep disorder is characterized by a prominent sleep disturbance (e.g., insomnia, hypersomnia, parasomnia, or mixed) that is etiologically related to caffeine use. Caffeine use is most frequently associated with insomnia. Caffeine-induced sleep disorder should be diagnosed in patients with caffeine intoxication only if the sleep disturbance is in excess of that which would be expected from intoxication. Other diagnoses that should be considered in the differential diagnosis of caffeine-induced sleep disorder include caffeine intoxication and caffeine withdrawal, sleep disturbances due to another mental disorder, a primary sleep disorder, and a sleep disorder due to a general medical condition.

Although there are no studies directly examining DSM-defined caffeine-induced sleep disorder, there has been a great deal of research on the relationship between caffeine use and sleep. The closer caffeine is taken to bedtime, the more likely it is to produce disruptive effects on sleep. It has been shown that consuming 200 mg of caffeine before bedtime (the equivalent of about 12 oz of brewed coffee) delays sleep onset and reduces sleep efficiency and total sleep time. Other studies have shown that 200 mg of caffeine taken early in the day can disrupt that evening's sleep.

Sleep disturbances secondary to caffeine use are more likely in people who are not regular consumers of caffeine. Regular caffeine consumers develop some tolerance to the sleep-disrupting effects of caffeine; however, because this tolerance is incomplete, they still may be vulnerable to caffeine-related sleep problems. Genetic factors may also underlie individual differences in the sleep-disruptive effects of caffeine. A polymorphism in the *ADORA2A* gene has been identified as being associated with differential responses to caffeine's effects on both subjective and objective measures of sleep.

It should be noted that caffeine can help to prevent unwanted sleep (e.g., such as when driving long distances) and can have beneficial effects when a person is sleep deprived. Studies have shown that sleep deprivation is associated with decrements in performance on a variety of tasks. Caffeine can reduce these performance decrements but does not completely reverse them.

Treatment

There are no studies on the treatment of caffeine-induced sleep disorder. In cases of suspected caffeine-induced sleep disorder, clinicians may want to discourage the use of sleep medications until caffeine use has been eliminated. Although patients may be reluctant to eliminate caffeine, clinicians should try to encourage patients to attempt a caffeine abstinence trial to rule out an etiologic role of caffeine in sleep problems (see Table 11.4–4).

CAFFEINE WITHDRAWAL

It is well documented that chronic caffeine consumption produces physical dependence, which manifests as a characteristic withdrawal syndrome when caffeine consumption is abruptly discontinued or substantially reduced. Caffeine withdrawal is a diagnosis in DSM-5-TR, ICD-10, and ICD-11.

Epidemiology

Studies have examined the incidence of caffeine withdrawal under conditions in which it is experimentally induced. In one study of community volunteers whose average daily caffeine consumption was like that of the general population, 52% of the subjects developed a moderate or severe headache when caffeine withdrawal was experimentally induced under double-blind conditions. Other studies have found that half or more subjects experience symptoms of caffeine withdrawal under double-blind acute abstinence. Symptoms may be detected sooner and be rated as more severe under naturalistic conditions in which individuals are aware that they have missed their usual caffeine dose.

In the United States, it is estimated that more than 85% of the adult population are regular consumers of caffeine, suggesting that there are large numbers of people who potentially could experience caffeine withdrawal. However, there have been relatively few attempts to conduct large-scale population-based assessments of caffeine withdrawal. Studies most often involve convenience samples and have reported caffeine withdrawal rates ranging anywhere from 8% to 58%. In one population-based survey, 71% of individuals who tried to stop using caffeine reported experiencing DSM-IV-TR—defined withdrawal symptoms, and 24% reported having headache plus other symptoms that interfered with performance. In a sample of individuals seeking treatment for problematic caffeine use, 96% reported experiencing caffeine withdrawal symptoms upon abstinence. Headache was the most frequently endorsed criterion (i.e., 89%), and 43% of the sample reported that caffeine withdrawal caused functional impairment (e.g., missing work). Caffeine withdrawal has been documented among individuals fasting for medical procedures or religious observances (e.g., fasting during Ramadan, giving up coffee for Lent). Caffeine withdrawal has also been documented in children and adolescents and in neonates born to mothers who consumed large amounts of caffeine during pregnancy. Those who chronically use caffeine-containing analgesics may also be at increased risk of experiencing caffeine withdrawal upon abstinence.

Comorbidity

According to DSM-5-TR, caffeine withdrawal may be associated with major depressive disorder, panic disorder, general anxiety disorder, antisocial personality disorder, alcohol use disorder, and cannabis and cocaine use.

Diagnosis and Clinical Features

A review of the research literature on caffeine withdrawal summarized caffeine withdrawal signs and symptoms reported in 66 studies (Table 11.4–5). Symptoms that were deemed to have the strongest empirical basis included headache, fatigue/drowsiness, difficulty concentrating, mood disturbances (e.g., dysphoric mood, depressed mood, or irritability), and flu-like symptoms (e.g., nausea, vomiting, and muscle pain/stiffness). These five symptom categories are those that appear in the DSM-5-TR. In the DSM-5-TR, a diagnosis is made when a patient experiences symptoms in three or more of these symptom categories along with clinically significant distress or impairment after abrupt cessation or a reduction in caffeine use. Patients can experience withdrawal symptoms even without a concurrent headache, suggesting that the other features of caffeine withdrawal are not simply secondary to the headache. Anxiety was a proposed symptom in the DSM-IV-TR, but anxiety was omitted from DSM-5 and DSM-5-TR because it is rarely observed in double-blind studies of caffeine withdrawal. Anxiety may be present under naturalistic conditions in which individuals know that they have missed a usual caffeine dose and anticipate withdrawal symptoms. Other withdrawal signs and symptoms that have been reported in the scientific literature are listed in Table 11.4–5. It is important to note that patients may be unaware of their physical dependence on caffeine and under conditions of caffeine abstinence may misattribute symptoms of caffeine withdrawal to other causes (e.g., illness such as influenza or sleep deprivation).

Table 11.4–5. Empirically Validated Signs and Symptoms Resulting from Caffeine Abstinence

Strong Evidence	Suggestive Evidence
Headache	Decreased desire to socialize
Tiredness/fatigue	Unmotivated for work
Decreased energy/activeness	Decreased self-confidence
Decreased alertness/attentiveness	Yawning
Drowsiness/sleepiness	Heavy feelings in arms and legs
Decreased contentedness/well-being	Craving/strong desire to use
Depressed mood	Analgesic use
Difficulty concentrating	Increased nighttime sleep quality/duration
Irritability	Impaired behavioral and cognitive performance
Muzzy/foggy/not clearheaded	Decreased motor activity
Flu-like symptoms	Increased cerebral blood flow
Nausea/vomiting	Electroencephalogram (EEG) changes
Muscle pain/stiffness	

Data from Juliano LM, Griffiths RR. A critical review of caffeine withdrawal: Empirical validation of symptoms and signs, incidence, severity, and associated features. *Psychopharmacology.* 2004;176:1–29. doi:10.1007/s00213-004-2000-x.

Caffeine Withdrawal Headache. Headache is a hallmark symptom of caffeine withdrawal. It is usually described as a generalized, throbbing headache that is sensitive to movement and worsened by exercise and the Valsalva maneuver. Headache typically begins 12 to 24 hours after last ingesting caffeine, although the onset can occur up to 40 hours after the last dose. It usually resolves in 2 to 4 days, although there have been reports of a protracted headache up to 21 days after caffeine cessation. Caffeine consumption usually rapidly (30 to 60 minutes) alleviates a caffeine withdrawal headache.

Headache is a frequent postoperative symptom. Patients undergoing operative procedures are sometimes required to abstain from all oral intakes prior to the procedure, including their usual daily caffeine. Studies indicate that caffeine withdrawal plays an etiologic role in postoperative headache. Greater caffeine use is associated with a greater risk of postoperative headache in a dose-related fashion. Furthermore, caffeine users given caffeine on the day of the surgical procedure have lower rates of postoperative headache than those given placebo.

Caffeine Withdrawal and Caffeine Dose. Higher daily doses of caffeine are associated with greater probability and greater severity of withdrawal. However, it should be noted that caffeine withdrawal can occur after ingestion of relatively low dosages of caffeine, such as 100 mg per day (the equivalent of about 6 oz of brewed coffee). Caffeine withdrawal can also develop after as few as 3 days of exposure to moderate doses of caffeine (e.g., 300 mg per day). Ingestion of low doses of caffeine (e.g., 25 mg) may lessen the severity of caffeine withdrawal among those who typically consume moderate doses of caffeine.

Severity of Caffeine Withdrawal. Caffeine withdrawal severity can vary from mild (e.g., slight fatigue) to severe (e.g., the worst headache ever experienced). A review of caffeine withdrawal studies found that the incidence of clinically significant impairment in functioning (e.g., unable to work) during caffeine withdrawal ranged from 10% to 55% depending on the study, with a median rate of 13%.

Differential Diagnosis

Several disorders should be considered in the differential diagnosis of caffeine withdrawal. Caffeine withdrawal can mimic migraine and other headache disorders, viral illnesses, sinus conditions, tension, other drug withdrawal states (e.g., from amphetamines, cocaine), and medication side effects. The final determination of caffeine withdrawal should rest on an evaluation of the pattern and amount of caffeine consumption, the time interval between the last ingestion of caffeine and the onset of symptoms, and the clinical features presented by the patient. Attenuation of symptoms by a dose of caffeine may also be useful in clarifying the diagnosis.

Course and Prognosis

Caffeine withdrawal typically begins 12 to 24 hours after the last dose of caffeine and usually resolves in 2 to 9 days. Most symptoms peak within the first 48 hours of caffeine abstinence. However, it should be noted that there can be considerable variability across individuals in the manifestations, severity, and time course of caffeine withdrawal. Furthermore, it has been shown that the same person can experience different symptoms and different degrees of severity of symptoms across different episodes of caffeine withdrawal.

Treatment

There has been little work done in systematically investigating the optimal treatment for caffeine withdrawal. The symptoms of caffeine withdrawal typically resolve within 30 to 60 minutes of caffeine ingestion. A person with a provisional diagnosis of caffeine withdrawal who continues to be symptomatic after caffeine consumption has been terminated for more than 2 weeks should be carefully re-evaluated. Caffeine withdrawal headaches may also respond to aspirin or other analgesics. If caffeine withdrawal occurs because a person is attempting to cease caffeine use, the clinician may want to advise the patient to taper caffeine consumption over time to lessen withdrawal severity. Caffeine modification strategies are outlined in Table 11.4–4.

Ms E was a 32-year-old single White woman employed full time at a local factory. She occasionally used nonsteroidal anti-inflammatory drugs but was taking no regular prescription medications. She had a history of alcohol use disorder, in remission for 9 years, and was otherwise in good health. She first began consuming caffeine when she started college, and her current beverage of choice was coffee. She typically drank four to five mugs of coffee each day, and preferred to drink it without cream, milk, or sugar. She estimated that 5 minutes elapsed between the time she got up in the morning and the time she had her first cup of coffee; her roommate made a pot before Ms E got up, and Ms E immediately poured a mug when she got out of bed. She spaced her mugs over the course of the day, with her last mug either after lunch or with dinner. Physicians had recommended she cut down or stop her coffee use because of complaints of mild indigestion, but she had been unable to do so. Her roommate had also complained about her coffee use at times. Ms E routinely drank hot coffee in her car and had spilled it and burned herself on one occasion. When she had stopped caffeine abruptly, Ms E experienced marked irritability; poor concentration; and a severe, generalized headache. When asked to rate the severity of the headache, she replied that “on a scale of 1 to 10 it’s a 12.” She also had muscle aches, low energy, lethargy, and a craving to drink a mug of coffee. On the day that she had stopped coffee use abruptly, she left work 2 hours early because of problems with concentrating on the job and went to bed several hours earlier than usual. She then returned to her usual pattern of coffee use.

CAFFEINE USE DISORDER

Caffeine use disorder is defined as a cluster of physiologic, behavioral, and psychological symptoms indicative of problematic caffeine use along with difficulty controlling use despite a desire to do so. Several research studies have identified adults and adolescents with problematic caffeine use characteristic of caffeine use disorder. In the DSM-5-TR, caffeine use disorder is included as a condition for further study. The DSM-5 workgroup acknowledged the existing body of evidence supporting caffeine use disorder as a clinically significant condition, but also sought to generate additional research to determine its reliability, validity, and prevalence. They also recognized a particular need to associate a diagnosis of caffeine use disorder with clinical distress and functional impairment. The provisional diagnostic framework proposed parallels other SUD diagnoses with a few key exceptions. An SUD diagnosis (referred to in prior editions of the DSM as “substance abuse” or “substance dependence”) requires fulfilling at least two of a constellation of diagnostic criteria that are linked by the theme of problematic use of the drug. The DSM-5-TR-proposed criteria for caffeine use disorder contain a much more restrictive set of criteria than the DSM-5-TR SUD criteria that apply to other drugs (see “Diagnosis and Clinical Features”). Following the inclusion of caffeine use disorder as a diagnosis for further study in DSM-5, additional research has associated caffeine use disorder with distress and functional impairment, and clinical trials have evaluated effective behavioral interventions. It will be important for future research to further characterize the features and prevalence of caffeine use disorder, its clinical significance and prognosis, associated features, and effective treatment strategies to inform the next DSM and clinical practice.

Epidemiology

Caffeine use is common throughout the world. In the United States, more than 85% of persons aged 2 years and older consume caffeine on a given day. Daily caffeine consumption among adult caffeine consumers has been estimated to be upward of 200 mg per day, with approximately 14% consuming more than 400 mg of caffeine per day (the equivalent of about four 6-oz cups of brewed coffee). Given the substantial number of people who consume caffeine regularly and their use of daily dosages that produce mood-altering effects, even a very low prevalence rate of caffeine use disorder could represent a nontrivial number of people. A growing number of studies have examined the prevalence of caffeine use disorder in general population samples. For example, a recent online study among adult caffeine consumers in the United States found that 8% of participants met DSM-5-TR-proposed criteria for caffeine use disorder (see "Diagnosis and Clinical Features"). Those who met these criteria reported significantly greater caffeine-related distress and were more likely to report functional impairment due to withdrawal or a physical or psychological problem caused by caffeine relative to those who did not meet the proposed criteria. Other research studies among general samples of adults in the United States, Italy, and Hungary estimated the prevalence of caffeine use disorder (as defined by the three key diagnostic criteria proposed by DSM-5-TR or similar criteria from prior editions) range from 9% to 13%. Studies among persons seeking treatment for problematic caffeine use have found that 72% to 84% of participants meet the key diagnostic criteria found in the DSM-5-TR. The prevalence of problematic caffeine use may be higher among special populations including heavy caffeine consumers, college students, psychiatric patients, and persons with a history smoking, alcohol, or other drug use. Additional data among nationally representative samples and special populations of interest, such as psychiatric patients, will provide further information about the prevalence of caffeine use disorder and associated distress.

Comorbidity

Conditions associated with high levels of caffeine consumption are discussed in the caffeine intoxication section above. Published reports among persons meeting criteria for caffeine use disorder provide support for common comorbidities of alcohol use disorder, cigarette smoking, mood and anxiety disorders, and sleep disorders. For example, in a detailed evaluation of 16 adults with a DSM-IV—defined substance dependence disorder *when applied to caffeine*, about two-thirds had a past history of a psychiatric disorder in remission, most commonly SUDs followed by mood or anxiety disorders and eating disorders. Multiple studies have observed a clustering of problematic caffeine use, cigarette smoking, and a history of alcohol use disorder in a subset of subjects, consistent with the known shared genetic vulnerability of these substances. A survey study among U.S. adults found that those with DSM-5-TR—defined caffeine use disorder had greater scores on a general measure of depression, anxiety, and stress, and reported greater sleep problems relative to those who did not meet criteria. Survey respondents who met criteria for caffeine use disorder named anxiety and insomnia as the two most common psychological and physical problems caused or exacerbated by caffeine. The association of broader features of caffeine use disorder, such as the development of tolerance and withdrawal, with multiple psychiatric disorders (e.g., major depression, generalized anxiety disorder, panic disorder) and SUDs (alcohol, cannabis, cocaine) in a large sample of twins provides further evidence of shared vulnerability between caffeine-related problems and other psychiatric conditions.

Diagnosis and Clinical Features

ICD-10 previously recognized caffeine dependence as a diagnosis, but ICD-11 does not include a specific SUD related to caffeine. Nevertheless, caffeine use disorder could be characterized with ICD-11 as a “disorder due to use of caffeine” with a “harmful pattern of caffeine use,” or “other unspecified disorders due to use of caffeine.” As stated previously, caffeine use disorder is listed as condition for further study in DSM-5-TR, and symptoms characteristic of caffeine use disorder could be categorized as unspecified caffeine-related disorder according to DSM-5-TR. Caffeine use disorder as proposed by DSM-5-TR includes nine criteria; however, endorsement of the following three specific criteria is necessary for a diagnosis: (1) a persistent desire or unsuccessful efforts to cut down or control caffeine use; (2) continued caffeine use despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by caffeine; and (3) withdrawal, as manifested by either: (a) the characteristic withdrawal syndrome for caffeine, or (b) caffeine (or a closely related substance) being taken to relieve or avoid withdrawal symptoms. The restrictive criteria for caffeine use disorder put forward by DSM-5-TR can be compared with a DSM-5-TR diagnosis of SUD for drugs other than caffeine, which requires the endorsement of only 2 of 11 possible criteria. The aim of using more restrictive criteria for caffeine use disorder is to limit the diagnosis to the most clinically important cases of problematic caffeine use. The nine possible criteria for caffeine use disorder overlap with other SUD criteria, except for two features that DSM-5-TR excluded for caffeine, specifically: (a) important social, occupational, or recreational activities given up or reduced because of substance use or (b) recurrent use of substance in situations in which it is physically hazardous. Because the criteria for caffeine use disorder are still evolving, more information is needed about the extent to which the proposed criteria and closely related criteria for SUD appropriately characterize those with caffeine-related distress or functional impairment. Initial item response analysis of the DSM-proposed diagnostic criteria for caffeine use disorder found that craving and use of more caffeine/use over a longer period than intended were commonly endorsed by study participants but were less informative with respect to caffeine use disorder severity. By contrast, failure to fulfill major role obligations and persistent or recurrent social or interpersonal problems resulting from caffeine use were less commonly endorsed but were characteristic of high caffeine use disorder severity. Individuals who consume higher daily amounts of caffeine may be more likely to meet criteria for caffeine use disorder, but caffeine use disorder has been associated with a wide range of daily caffeine use and primary sources of caffeine. For example, one study involving individuals seeking treatment for caffeine use (84% diagnosed with caffeine use disorder) observed caffeine use ranging from 200 to 2,667 mg daily. Persons meeting criteria for caffeine use disorder have been found to be consumers of coffee, tea, soft drinks, energy drinks/shots, or a combination of sources.

Differential Diagnosis

Habitual use of caffeine is so commonplace that many clinicians do not recognize caffeine as a substance capable of producing problematic outcomes. Because of this, caffeine consumption is not routinely assessed, and a clinical syndrome of caffeine use disorder may be easy to overlook. Other SUDs and related syndromes should be included in the differential diagnosis of caffeine use disorder, such as intoxication or withdrawal related to other stimulant drugs. Symptoms of caffeine use disorder overlap with diagnoses of DSM-defined caffeine intoxication and caffeine withdrawal, which should also be considered in the differential diagnosis. Because anxiety and sleep disturbance have been commonly reported as problems caused or exacerbated by caffeine use by those who meet diagnostic criteria for caffeine use disorder, the presence of underlying anxiety disorders and sleep disorders should also be evaluated.

Course and Prognosis

The natural course and prognosis for patients with a diagnosis of caffeine use disorder is not known. Continued use of caffeine despite repeated efforts to discontinue caffeine use is one of the key characteristics of caffeine use disorder. Although clinical trials implementing gradual caffeine reduction programs have shown success in sustained reductions in caffeine use (see “Treatment”), relapse to caffeine use after treatment has also been observed. This suggests that caffeine use disorder may be characterized as a chronic relapsing disorder like use disorders associated with other drugs.

Treatment

Current recommendations for addressing caffeine use disorder include caffeine reduction to mitigate the effects of physical or psychological problems caused or worsened by continued caffeine use. Gradual caffeine reduction over a period of weeks is recommended to avoid the potential aversive effects of withdrawal that can occur with sudden caffeine cessation, especially among heavy consumers. Although clinicians may recommend that patients modify caffeine use for various conditions (e.g., anxiety, insomnia, tachycardia, pregnancy), they may not realize that caffeine modification may be challenging for some patients. In fact, individuals diagnosed with caffeine use disorder have reported being unable to follow their doctors' recommendations to modify caffeine use. Thus, it should not be assumed that clinician suggestion to modify caffeine use will be sufficient for successful behavior change. Patients may not be aware of all sources of caffeine in their diet and patient education may be necessary. Effective strategies for caffeine reduction may include education about the effects of caffeine and common sources of caffeine, self-monitoring of daily use, coping response training, social support, and reinforcement for decreased use. SUD treatment approaches such as relapse prevention and motivational interviewing could also be applied to the treatment of caffeine use disorder. Two randomized clinical trials among persons seeking treatment for caffeine use evaluated the effectiveness of a manualized brief intervention for caffeine reduction over a period of 6 weeks. One study included a 1-hour counselor-guided session to introduce the treatment manual, caffeine diary, weekly caffeine goals, and cognitive behavioral strategies for caffeine reduction, whereas the other study presented the treatment manual to the participant very briefly (i.e., less than 5 minutes) and the participant implemented caffeine reduction independently. The treatment manual provided to the participants consisted of facts and information about caffeine, strategies for caffeine reduction based on the principles of relapse prevention, a caffeine diary, and instructions on calculating weekly caffeine goals during caffeine reduction. Both studies observed sustained reductions in caffeine use following treatment relative to a waitlist control group. Although the success of the self-implemented manualized intervention for caffeine reduction is promising for practical application, there may well be persons with more complex presentations of caffeine use disorder with psychiatric and medical comorbidities that would benefit from additional psychosocial and medical support. These initial studies have provided evidence that caffeine reduction improves DSM-defined caffeine-related distress; however, more information is needed to evaluate the potential benefits of caffeine reduction on common comorbid symptoms such as anxiety and sleep, as well as to evaluate the efficacy of the intervention in special populations. Table 11.4–4 provides suggested guidelines for caffeine use modification.

Ms G is a 38-year-old female who is married with four children. She works part time as a dental assistant. Ms G presented to the primary care clinic with flu-like symptoms. She complained of nausea, heavy feelings in her arms and legs, muscle aches, and extreme lethargy. She also said she had a pounding headache that was one of the worse headaches she had ever experienced. Ms G said that her symptoms began around 2 PM the day prior and she left work early because she could no longer function. She called her mother to help care for her school-aged children as her husband was away on a business trip. Upon examination, Ms G's heart rate and blood pressure were within normal limits and she was afebrile. An assessment of Ms G's caffeine use revealed that she consumes three cups of black drip coffee each morning. Ms G reported that she had not had caffeine the last two mornings because she had given it up for Lent. Ms G said that she was unaware of her physical dependence on caffeine because she hadn't gone more than a day without caffeine since her last pregnancy 9 years ago. Ms G was diagnosed with caffeine withdrawal syndrome. Ms G chose to drink a small cup of coffee and her symptoms resolved after approximately 45 minutes.

Ms S is a 45-year-old Caucasian female who presented for assistance with problematic caffeine consumption, or what she described as her “caffeine addiction.” Ms S is married and has two elementary school—aged children and works full time as a nurse for a large research hospital. She takes no prescription medications, does not smoke, and has no history of psychiatric problems. She reports occasionally drinking alcohol on the weekends. She reports having tried marijuana a few times in college and has no other history of illicit drug use. She began regularly consuming caffeinated beverages while in college, and her consumption increased to its present levels during nursing school. Her primary source of caffeine comes from diet cola, which she consumes daily. She reports drinking coffee or tea only on occasion (e.g., when dining with friends). Ms S has her first soft drink within 5 minutes of waking up in the morning and then continues drinking diet colas throughout the day until around 6 to 8 PM in the evening. She typically consumes a total of eight 20-oz diet cola bottles each day, the equivalent of about 600 mg of caffeine. She and her husband have argued about her soft drink use in the past, especially during her pregnancies. However, she continued to consume diet colas during both of her pregnancies. Ms S described having attempted to give up diet cola many times, with her most recent attempt a few years prior. She said that after going cold turkey on the first day of Lent, she experienced excruciating headaches, extreme fatigue, irritability, and an inability to function properly at work. She said that she returned to drinking diet colas after 3 days of “misery.” Ms S reports that her motivation for ceasing her diet cola consumption at this time is health related. She is having increasing difficulties with urinary incontinence and, furthermore, feels that her diet cola consumption is generally not healthy. She is also presently complaining of sleep difficulties and anxiety that she attributes to the ailing health of her mother, although she acknowledged that her caffeine use could also be contributing to these symptoms. Ms S reported that she was interested in completely eliminating diet cola from her diet rather than reducing her consumption. Ms S and the therapist developed a schedule to gradually reduce her soft drink consumption over a period of 4 weeks and completely eliminating soft drinks on week 5 (i.e., six bottles to four bottles to two bottles to one bottle to zero bottles). The therapist provided a booklet that described the effects of caffeine and caffeine withdrawal, and that contained tips for how to cope with temptations to consume caffeine and caffeine withdrawal symptoms should they occur. Ms S was advised to substitute noncaffeinated drinks during the times that she normally consumed diet colas, and she identified carbonated water as a desirable alternative. A follow-up appointment was scheduled for 5 weeks later. At the 5-week follow-up, Ms S reported that she had reduced her consumption to two 12-oz diet colas per day. At the 6-month follow-up, Ms S reported having "been off" diet colas for the past 3 months. She reported having headaches for the first few days after abstaining from caffeine but that they were tolerable, especially if she took ibuprofen. Ms S reported that her problems with urinary incontinence and poor sleep had both improved, and she felt confident that she could continue to refrain from diet cola consumption.

Caffeine Use and Nonpsychiatric Illnesses. Numerous studies examining the relationship between caffeine use and physical illness have not identified any significant health risk from nonreversible pathologic consequences of caffeine use, such as cancer, heart disease, or human reproduction. However, caffeine use is often considered to be contraindicated for various conditions, including generalized anxiety disorder, panic disorder, primary insomnia, gastroesophageal reflux, urinary incontinence, and pregnancy. In addition, the modest ability of caffeine to increase blood pressure and the documented cholesterol-elevating compounds in unfiltered coffee have raised the issue of the relationship of caffeine and coffee use to cardiovascular disease. Finally, there may be a modest association between higher daily caffeine use in women and delayed conception, higher rates of miscarriage and stillbirth, and lower birth weight. However, there are studies that have not found such associations, and effects when found are usually with relatively high daily doses of caffeine (e.g., the equivalent of five cups of brewed coffee per day). For a woman who is considering pregnancy, especially if there is some difficulty in conceiving, it may be useful to counsel elimination of caffeine use. Similarly, for a woman who becomes pregnant and has moderate-to-high daily caffeine consumption, a discussion about decreasing her daily caffeine use may be warranted. Caffeine may possibly offer protective effects against some diseases, most notably Parkinson disease.

FURTHER READINGS

Ágoston C, Urbán R, Richman MJ, Demetrovics Z. Caffeine use disorder: an item-response theory analysis of proposed DSM-5 criteria. *Addict Behav*. 2018;81:109–116.

Clark I, Landolt HP. Coffee, caffeine, and sleep: a systematic review of epidemiological studies and randomized controlled trials. *Sleep Med Rev*. 2017;31:70–78.

Evans SM, Griffiths RR. Caffeine withdrawal: a parametric analysis of caffeine dosing conditions. *J Pharmacol Exp Ther*. 1999;289(1):285–294.

Evatt DP, Juliano LM, Griffiths RR. A brief manualized treatment for problematic caffeine use: a randomized control trial. *J Consult Clin Psychol*. 2016;84(2):113–121.

Ferré S. Mechanisms of the psychostimulant effects of caffeine: implications for substance use disorders. *Psychopharmacology (Berl)*. 2016;233(10):1963–1979.

Fulgoni VL 3rd, Keast DR, Lieberman HR. Trends in intake and sources of caffeine in the diets of US adults: 2001–2010. *Am J Clin Nutr*. 2015;101(5):1081–1087.

Fulton JL, Dinas PC, Carrillo AE, Edsall JR, Ryan EJ, Ryan EJ. Impact of genetic variability on physiological responses to caffeine in humans: a systematic review. *Nutrients*. 2018;10(10):1373.

Hughes JR, Oliveto AH, Liguori A, Carpenter J, Howard T. Endorsement of DSM-IV dependence criteria among caffeine users. *Drug Alcohol Depend*. 1998;52(2):99–107.

Griffiths RR, Evans SM, Heishman SJ, et al. Low-dose caffeine physical dependence in humans. *J Pharmacol Exp Ther*. 1990;255(3):1123–1132.

Guest NS, VanDusseldorp TA, Nelson MT, et al. International society of sports nutrition position stand: caffeine and exercise performance. *J Int Soc Sports*

Nutr. 2021;18(1):1.

Gummin DD, Mowry JB, Beuhler MC, et al. 2019 annual report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 37th annual report. *Clin Toxicol (Phila)*. 2020;58(12):1360–1541.

Institute of Medicine. *Caffeine in Food and Dietary Supplements: Examining Safety: Workshop Summary*. National Academies Press; 2014.

Irwin C, Khalesi S, Desbrow B, McCartney D. Effects of acute caffeine consumption following sleep loss on cognitive, physical, occupational and driving performance: a systematic review and meta-analysis. *Neurosci Biobehav Rev*. 2020;108:877–888.

James JE. Maternal caffeine consumption and pregnancy outcomes: a narrative review with implications for advice to mothers and mothers-to-be. *BMJ Evid Based Med*. 2021;26(3):114–115.

Jones AW. Review of caffeine-related fatalities along with postmortem blood concentrations in 51 poisoning deaths. *J Anal Toxicol*. 2017;41(3):167–172.

James JE. *Understanding Caffeine: A Biobehavioral Analysis*. Sage; 1997.

Juliano LM, Griffiths RR. A critical review of caffeine withdrawal: empirical validation of symptoms and signs, incidence, severity, and associated features. *Psychopharmacology (Berl)*. 2004;176:(1):1–29.

Juliano LM, Huntley ED, Harrell PT, Westerman AT. Development of the caffeine withdrawal symptom questionnaire: caffeine withdrawal symptoms cluster into 7 factors. *Drug Alcohol Depend*. 2012;124(3):229–234.

Lieberman HR, Agarwal S, Fulgoni VL 3rd. Daily patterns of caffeine intake and the association of intake with multiple sociodemographic and lifestyle factors in US adults based on the NHANES 2007–2012 surveys. *J Acad Nutr Diet*. 2019;119(1):106–114.

McLellan TM, Caldwell JA, Lieberman HR. A review of caffeine's effects on cognitive, physical and occupational performance. *Neurosci Biobehav Rev*. 2016;71:294–312.

Meredith SE, Juliano LM, Hughes JR, Griffiths RR. Caffeine use disorder: a comprehensive review and research agenda. *J Caffeine Res*. 2013;3:114–130.

Mitchell DC, Knight CA, Hockenberry J, Teplansky R, Hartman TJ. Beverage caffeine intakes in the U.S. *Food Chem Toxicol*. 2014;63:136–142.

Musgrave IF, Farrington RL, Hoban C, Byard RW. Caffeine toxicity in forensic practice: possible effects and under-appreciated sources. *Forensic Sci Med Pathol*. 2016;12(3):299–303.

Nawrot P, Jordan S, Eastwood J, Rotstein J, Hugenholtz A, Feeley M. Effects of caffeine on human health. *Food Addit Contam*. 2003;20(1):1–30.

Nehlig A. Interindividual differences in caffeine metabolism and factors driving caffeine consumption. *Pharmacol Rev*. 2018;70:384–411.

Pendergrast M. *Uncommon Grounds: The History of Coffee and How It Transformed Our World*. Basic Books; 2019.

Reissig CJ, Strain EC, Griffiths RR. Caffeinated energy drinks–a growing problem. *Drug Alcohol Depend*. 2009;99(1):1–10.

Strain EC, Mumford GK, Silverman K, Griffiths RR. Caffeine dependence syndrome. Evidence from case histories and experimental evaluations. *JAMA*. 1994;272:1043–1048.

Sweeney MM, Juliano LM, Ferré S, Griffiths RR. The pharmacology of caffeine. In: Miller SC, Fiellin DA, Rosenthal RN, Saitz R, American Society of Addiction Medicine, eds. *The ASAM Principles of Addiction Medicine*. 6th ed. Wolters Kluwer; 2018.

Sweeney MM, Meredith SE, Juliano LM, Evatt DP, Griffiths RR. A randomized controlled trial of a manual-only treatment for reduction and cessation of problematic caffeine use. *Drug Alcohol Depend*. 2019;195:45–51.

Sweeney MM, Weaver DC, Vincent KB, Arria AM, Griffiths RR. Prevalence and correlates of caffeine use disorder symptoms among a United States sample. *J Caffeine Adenosine Res*. 2020;10(1):4–11.

Temple JL. Review: trends, safety, and recommendations for caffeine use in children and adolescents. *J Am Acad Child Adolesc Psychiatry.* 2019;58(1):36–45.

Temple JL, Bernard C, Lipshultz SE, Czachor JD, Westphal JA, Mestre MA. The safety of ingested caffeine: a comprehensive review. *Front Psychiatry*. 2017;8:80.

Wikoff D, Welsh BT, Henderson R, et al. Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children. *Food Chem Toxicol.* 2017;109:585–648.

Willson C. The clinical toxicology of caffeine: a review and case study. *Toxicol Rep*. 2018;5:1140–1152.

11.5 Cannabis-Related Disorders

Janni Leung, Ph.D., Wayne D. Hall, Ph.D., and Louisa Degenhardt, Ph.D.

Cannabis is the most widely used illicit drug in the world, with an estimated 192 million past-year users worldwide in 2018 (3.9% of the global adult population). The global prevalence of cannabis use disorder is 0.3%, which is 23.8 million people. Over the past 50 years, cannabis use has become part of youth culture in many developed societies, with first use occurring in the mid- to late teens.

Cannabis is the fourth most commonly used psychoactive drug among adults in the United States, after caffeine, alcohol, and nicotine.

HISTORY

Cannabis preparations are obtained from the plant *Cannabis sativa*, which has been used in China, India, and the Middle East for approximately 8,000 years, primarily for its fiber and secondarily for its medicinal properties. Cannabis was introduced to Europe in the early 19th century by Napoleon's army returning from Egypt, and later in the same century to Britain for medical use by an Irish surgeon on his return from India. Cannabis was most often used in Europe for fiber and, to a lesser extent, for therapeutic purposes. It was first used recreationally by Parisian bohemians in the late 19th century.

Recreational cannabis use was introduced to the United States in the 1930s from Mexico and spread via jazz musicians to cities in the Northeastern United States. Its use was banned in the United States in 1937 and most other developed countries in 1961 under the UN Single Convention on Narcotic Drugs. Recreational cannabis use among bohemians in the United States in the 1940s and 1950s spread to the U.S. youth population in the late 1960s. Use increased dramatically in the United States during the 1970s before peaking in the early 1980s. It then declined until the early 1990s and stabilized at a lower level before beginning to increase among adults in the 2010s.

In the 1990s and 2000s, citizen-initiated referenda legalized the use of cannabis for medical purposes in around half the states of the United States. Recreational cannabis use was legalized by state referenda in Colorado and Washington State in 2012, Alaska and Oregon in 2014, and since then, in 15 states. To date, these policy changes have had minimal effects on cannabis use and cannabis-related harm in the United States, but this may change as cannabis is fully commercialized if legalization becomes national U.S. policy.

CANNABIS PREPARATIONS

The cannabis plant occurs in male and female forms. The female plant contains the highest concentrations of more than 100 cannabinoids unique to the plant. Delta-9-tetrahydrocannabinol (THC) is the cannabinoid that is primarily responsible for the psychoactive effects of cannabis. It is found in the resin that covers the flowering tops and upper leaves of the female plant. Most of the other cannabinoids are either inactive or only weakly active, although some, such as cannabidiol (CBD), may modulate the effects of THC and may have therapeutic uses.

Until recently in the United States, the most common cannabis preparations were marijuana, hashish, and hash oil. Marijuana is prepared from the dried flowering tops and leaves of the plant. Its potency depends on the growing conditions, the genetic characteristics of the plant, the ratio of THC to other cannabinoids, and the part of the plant that is used. The flowering tops have the highest THC concentration, with much lower concentrations in the leaves, stems, and seeds. Varieties of cannabis cultivated for hemp fiber usually contain very low levels of THC. Cannabis plants may be grown to maximize their THC production by the sinsemilla method (“without seed”), in which only female plants are grown together.

Cannabis concentrates or extracts are increasingly commonly used preparations. For example, butane hash oil (BHO), commonly known as honey oil or dabs, is prepared using butane pushed through a container of cannabis trimmings to extract cannabinoids into a honey-like substance. Cannabis resin is prepared by pressing cannabis with heat. THC vape juice and oil are prepared by extracting THC oil from cannabis buds and then adding it to a base liquid.

Cannabis edibles have also increased in popularity. Cannabis may be ground and then baked in the oven or heated with water and oil to make a butter that can be used to prepare sweets. Edibles may also be prepared by adding concentrates to a dish when cooking. Cookies and brownies, gummies, sparkling drinks, tea and coffee, chocolates, and popcorn are some commercial edible products that have appeared since U.S. state-level jurisdictions began enacting legalization of medical and nonmedical cannabis use within their borders.

The average concentration of THC in marijuana sold in the United States has increased in recent years, from 0.5% in 1970 to 9% in 2008 and 17% in 2017. The sinsemilla variety may contain as much as over 20% THC. Hashish or hash consists of dried cannabis resin and may contain 2% to 8% THC. Hash oil is obtained by extracting THC from hashish (or marijuana) in oil, and it may contain between 15% to 20% THC. Cannabis extracts and concentrates obtained by chemical processes may contain over 70% THC.

METHODS OF USE

Cannabis has traditionally been smoked as marijuana in hand-rolled cigarettes, or joints, which may include tobacco to assist burning. The use of a water pipe, or bong, is also a popular way to smoke all cannabis preparations, probably because it delivers a larger bolus dose of THC and maximizes the delivery of a given amount of cannabis. Cannabis can also be administered using a vaporizer to avoid the respiratory risks from inhaling the smoke from a cannabis joint.

Hashish may be mixed with tobacco and smoked as a joint or in a pipe, with or without tobacco. Because hash oil is extremely potent, a few drops may be applied to a cigarette or joint, or the oil may be heated and the vapor inhaled. Cannabis users typically inhale the smoke or vapor deeply and hold their breath to maximize the absorption of THC by the lungs.

The oral route of administration may also be used by eating edible cannabis products. In experimental studies, THC is often swallowed in gelatine capsules containing THC dissolved in sesame oil. In India, cannabis may be consumed in the form of *bhang*, a tea brewed from the leaves and stems of the plant. THC is not water soluble and so is not readily injected. Almost all cannabis users in the United States smoked cannabis in joints or used bongs or vaporizers until the 2000s because the chemistry and pharmacology of cannabis made smoking or inhaling the most efficient way to obtain THC. Since the legalization of cannabis, first for medical and later recreational use, the vaporization of cannabis flower and concentrates and the use of cannabis edibles has increased.

COMPARATIVE NOSOLOGY

Cannabis use disorder and problematic cannabis use have been topics of considerable research, codified in the most recent editions of the two major diagnostic classification systems. The 11th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-11), released in 2018, includes a diagnosis of cannabis dependence characterized by a strong internal drive to use cannabis manifested by impaired ability to control use, increasing priority given to use over other activities, and persistence of use despite harm or negative consequences. The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), released in 2013, includes the category of cannabis use disorders by severity. In 2022, a text revision was released (DSM-5-TR) with no changes for cannabis use disorders. ICD-11 includes a category of *harmful use* that attempts to capture problematic cannabis use that does not satisfy criteria for dependence but nonetheless causes harm. The concordance between ICD-11 and DSM-5 is good for comparisons of dependence and severe cannabis use disorder but poor for comparisons of harmful and mild cannabis use disorder.

EPIDEMIOLOGY OF CANNABIS USE

Prevalence

In the United States, two major annual surveys of illicit drug use have been conducted since the early 1970s. These are the Monitoring the Future nationwide survey of high school seniors, college students, and young adults conducted annually since 1975; and the National Household Survey on Drug Use and Health

(NHSDUH), which has surveyed household samples of adults annually throughout the United States since 1972.

In 2020, 46.4% of a U.S. national sample of persons aged 12 years and older reported that they had used cannabis at some time in their lives, 21% had done so in the past year, and 11% in the past month (Fig 11.5–1). Lifetime use increased from 16% among those aged 12 to 17 years to 52% in those aged 18 to 25 years, 59% among those aged 26 to 34, and decreased with age to 32% among those aged 65 or older. Rates of discontinuation of use were high: 55% of adults who had used cannabis at some time in their lives had not done so in the past year and 75% had not done so in the past month. Past-month cannabis use was more common among men (14%) than women (9%) and most common among those aged 18 to 25 years (23%).

The NHSDUH series from 1974 to 2000 show that rates of past-month cannabis use increased throughout the 1970s, peaked in 1979, and declined steadily throughout the 1980s to reach their lowest level in 1992 before increasing again after 1995. Between 2007 and 2014, cannabis use declined in the past year and past month; between 2016 and 2019, cannabis in the past-year use increased from 18% to 21%.

The Monitoring the Future project has estimated the prevalence of cannabis use among secondary school students, college students, and young adults since 1975. These surveys have revealed long waves of cannabis use among U.S. adolescents since 1975. Among 18 year olds, lifetime prevalence peaked at 65% in 1980, before falling to nearly half by the early 1990s, but increasing again after that time to 42% in 2006. Prevalence reduced a little through 2007 to 2008, with some rebounds, and then remained relatively stable at around 45% from 2017 to 2020.

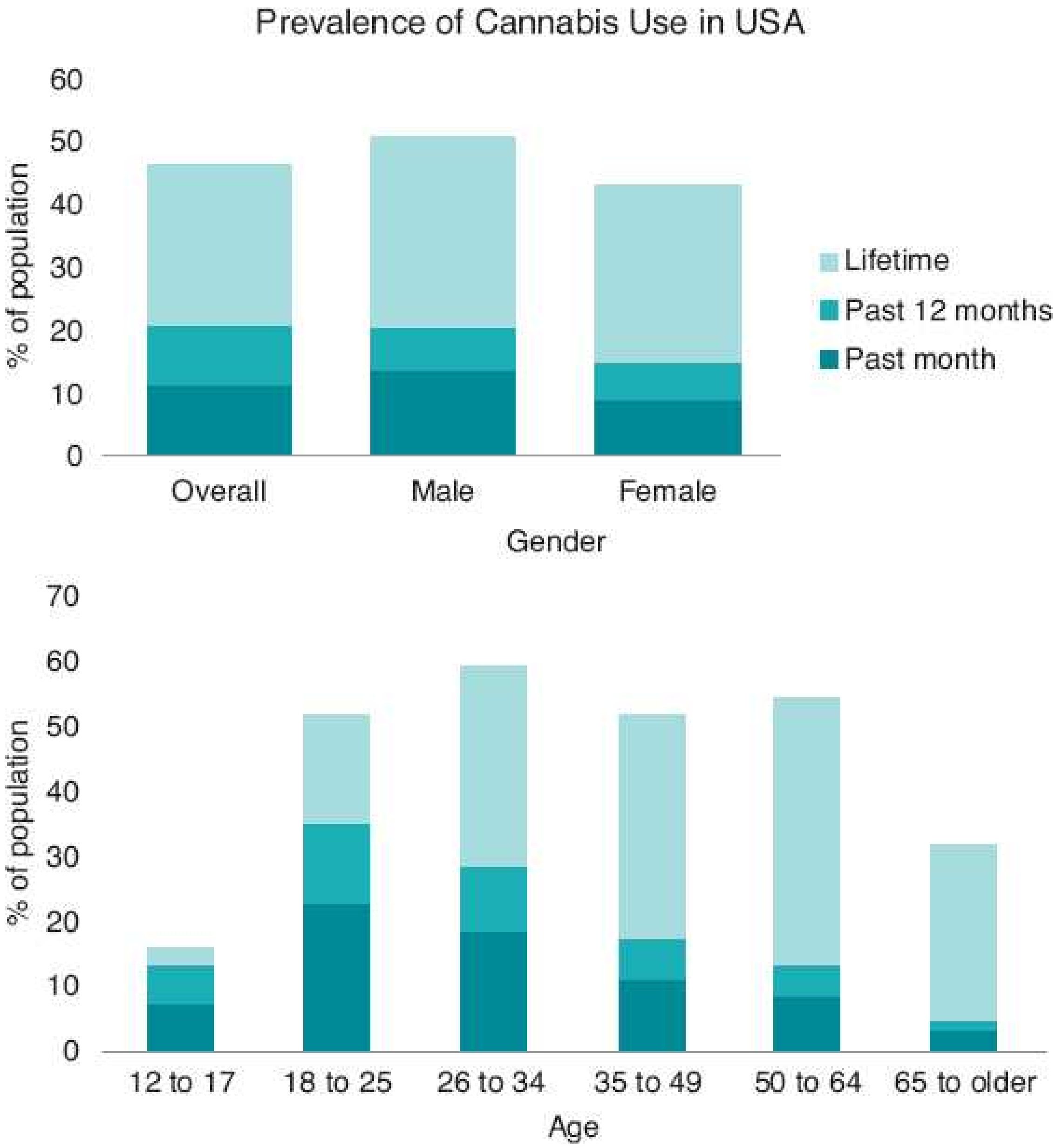

FIGURE 11.5–1. Prevalence of cannabis use by gender and age according to the U.S. National Household Survey on Drug Use and Health (2019). Note: Numbers are per 100 per year. (From SAMHSA. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2019 National Survey on Drug Use and Health*. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2020.)

The use of cannabis in the past year among 12th graders peaked at 51% in 1979 and fell to 22% by 1992. It increased to 32% in 2006 and has been around 35% in recent years. Experience with synthetic

THC (K2/Spice) was uncommon, with 2.4% of 12th graders reporting its use in the past year. Marijuana vaping in the past year has increased rapidly from 10% in 2017 to 22% in 2020. The discontinuation of cannabis use has declined between 1979 to 2020, with fewer of those who had used cannabis 10 or more times ceasing their use by age 18. Among those who have ever tried cannabis, 16% used it on a daily basis in 2020. Most of those who ceased use had not had a great deal of experience with cannabis. Although most users of other illicit drugs also had used cannabis, trends in the use of other illicit drugs were independent of the trends in cannabis use.

Attitudes and Beliefs Toward Cannabis

Changes in young people's beliefs about the risks of cannabis use seem to be responsible for both the reduction in use between 1979 and 1991 and the rise in its use after 1992. There is a strong negative correlation over time between the rates of cannabis use and the perceived risks of using it and perceived disapproval of use by peers. Between 1979 and 1992, a marked increase in perceived risk and a smaller increase in personal disapproval preceded a large decline in use. Perceived risk was relatively steady through the 1990s. Since the 2000s, there have been decreases in the perception that cannabis use is risky among both adolescents and adults. There has also been a reduction in the disapproval of regular cannabis use.

The changes in the perceived risks and benefits of cannabis use have been associated with a steep increase in support for cannabis legalization between 1990 and 2016. Younger males were the most supportive of legalization, while older adults aged 50 and over were the least supportive in the beginning, but the proportion of supporters has steadily increased.

Legalization of cannabis so far has had little impact on the prevalence of uptake in young people. However, people who were already using cannabis have used it more often after legalization, probably in response to increased marketing of cannabis products, greater access, and large reductions in cannabis retail prices.

Patterns of Use

Cannabis use is common among young people in many high-income countries. Many young adults in Europe, the United States, Canada, Australia, and New Zealand have tried cannabis. Most stop using in their 20s based on available data, such as those in the United States. Very few engage in daily cannabis use over a period of years. In the United States and Australia, about 10% of those who ever use cannabis become daily users, and another 20% to 30% use weekly. The pattern of use in developed countries differs from that in traditional cannabis-using countries, such as Egypt and India, where recreational cannabis use is uncommon and heavy cannabis use is largely confined to socially marginalized groups in the population.

Only a small proportion of cannabis users use the drug for several years or more. Daily or near-daily use over a period of years is the pattern with the greatest risk of producing adverse health and psychological consequences. Daily cannabis users are more likely to be male and less well educated; they are also more likely to regularly use alcohol and to have experimented with other illicit drugs such as amphetamines and other psychostimulants, hallucinogens, sedatives, and opioids.

Correlates of Cannabis Use

Age. First use of cannabis typically begins in the mid- to late teenage years, and the heaviest use typically occurs in the early 20s. Rates of cannabis use remain relatively high during the early 20s and decline during the late 20s. The majority of young adults who experiment with cannabis have done so by age 18, and in the United States, rates of use have, until recently, declined steadily from the mid-20s into the early 30s. There is some evidence that in recent U.S. cohorts, regular cannabis use is more likely to continue into the 30s. In recent surveys, cannabis use has also become more common among older adults aged 65 and older, perhaps in part because of increased use for medical reasons.

Gender. Rates of recreational cannabis use in the user's lifetime, the past year, and the past week are consistently higher among males than females. The same is true for daily use and long-term daily cannabis use. Cannabis use for medicinal reasons is common in both genders, with females more likely to use medicinal cannabis for headaches or migraines, nausea or vomiting, problems sleeping, and anxiety symptoms.

Income. In the United States, cannabis use has been positively correlated with income and paid work in adolescence. However, heavy cannabis use impacts cognitive function and capacity to work. Adolescents who use cannabis regularly have lower educational attainment, limiting their choice of occupation and reducing their income later in life. More recent evidence suggests that the liberalization of cannabis use has shifted regular use toward groups with lower incomes. Cannabis users with lower incomes are more likely to use daily or near daily, be less in control of their use, and spend a larger proportion of their income on cannabis.

Socioeconomic Status. The relationship between cannabis use and socioeconomic status has been weak. Higher rates of cannabis use are sometimes found among individuals from lower socioeconomic status, but in the past two decades, there has been no consistent relationship between parental education and cannabis use among 12th-grade students in the United States. Those whose parents used cannabis are more likely to use cannabis themselves. Variations in the relationship between parental socioeconomic status and cannabis use in their children may be explained by changes in use among different income groups across different periods and cohorts. Since 2000, regular cannabis use has become concentrated among less well-educated Americans without college educations, particularly among blue-collar males.

Ethnicity. In the Monitoring the Future surveys, adolescent student samples from several years have to be combined to make reliable comparisons of prevalence rates between the three largest ethnic groups. Previously, these have shown that White students had higher rates of cannabis use than African American students. Rates of use by Hispanic students generally fall in between. In recent years, the differences have narrowed as cannabis use has increased among African American and Hispanic students. White students are more likely to have tried synthetic cannabis than African American students, but the highest rates were seen among Hispanic 12th graders. There has been a considerable increase in support for cannabis legalization among all three racial groups during the 2000s, in students and adults, with White and African Americans more likely to support cannabis legalization than Hispanics.

Availability. As a general rule, the more available a drug is, the more often it is likely to be used. This has been true for alcohol consumption: the larger the number of licensed outlets and the longer the hours of trading, the higher the levels of community alcohol consumption and alcohol-related problems. There is limited evidence that this is also the case for cannabis use. Self-reports from surveys on how easy it is to obtain cannabis have shown very little change over long periods of time during which rates of cannabis use have risen and fallen in the United States. The recent legalization of retail cannabis stores and declining cannabis prices have increased availability, but, to date, this has largely increased the frequency of use among current users rather than increased cannabis use in the population.

PHARMACOLOGY OF CANNABINOIDS

Laboratory research with animals and humans has demonstrated that the primary psychoactive constituent in cannabis is THC. THC and other cannabinoid agonists produce their effects by interacting with an endogenous cannabinoid signaling system. There are at least two types of cannabinoid receptors (CB1 and CB2). The CB1 receptor is found primarily in the brain and mediates the psychological and behavioral effects of THC. The CB2 receptor, located mostly outside the central nervous system, is associated with the immune system and appears to modulate inflammatory responses. CB1 and CB2 receptors belong to a large group of G-protein–coupled receptors, membrane proteins that are involved in modulating chemical signaling between nerve cells via neurotransmitters.

CB1 cannabinoid receptors are found in brain regions known to be involved in mood, motor control, memory formation, regulation of food intake, pain modulation, sleep, and immune and reproductive functions. The fact that cannabis disrupts short-term memory is consistent with the abundance of CB1 receptors in the hippocampus, the brain region most closely associated with memory. The CB1 and CB2 receptors respond to endogenous cannabinoid ligands that occur naturally in the brain. These ligands include anandamide and 2-arachidonyl-glycerol (2AG).

Some active compounds found in cannabis have effects that are independent of cannabinoid receptors and endogenous cannabinoids. For example, both THC and CBD can be neuroprotective through their antioxidative activity. Other likely examples of receptor-independent cannabinoid activity are modulation of activity in membrane-bound enzymes (such as adenosine triphosphatase [ATPase]), arachidonic acid release, and perturbation of membrane lipids. An important caveat is that concentrations of THC or CBD used in studies of cellular effects are generally much higher than the concentrations of THC or CBD that would be achieved in humans who smoke herbal cannabis, although this may change with the increased use of high-potency cannabis products.

A typical joint of between 0.5 and 1.0 g of cannabis contains between 5 and 150 mg of THC. Twenty percent to 70% of the THC is found in the smoke that reaches the lungs; the rest is lost in sidestream smoke. Only 5% to 24% of the THC in a joint reaches the bloodstream when cannabis is smoked. Two to 3 mg of THC will produce a brief high in an occasional user, and a single joint may provide enough THC for two or three such users. A heavy cannabis smoker may use five or more joints per day.

Different methods of using cannabis produce differing absorption, metabolism, and excretion of THC. When cannabis is smoked or vaporized, THC is absorbed from the lungs into the bloodstream within minutes. It is first metabolized in the lungs and then in the liver. The metabolite 9-carboxy-THC can be detected in blood within minutes of smoking. Peak blood levels of THC are usually reached within 10 minutes of the first puff and may happen before the end of the smoking session. They rapidly decline to about 5% to 10% of their initial level within an hour as THC is metabolized and distributed to fatty tissues, including the brain.

When swallowed, THC takes 1 to 3 hours to enter the bloodstream, delaying the onset of its psychoactive effects. The metabolite, 11-hydroxy-THC, which is 20% more potent than THC, is found in high concentrations after cannabis is swallowed.

THC and its metabolites are highly fat soluble, so they may remain in the fatty tissues of the body for long periods of time. THC and its metabolites may accumulate in the body of chronic users because of their slow rate of clearance. THC may be detected in the blood for several days, and traces may persist for several weeks; it may be stored in body fat for more than 28 days.

It has been argued that the medical literature may underestimate the adverse health effects of cannabis because it is based on the use of less potent forms of cannabis than those that have become available in the past several decades. The average potency of cannabis has increased along with an earlier age of first use and more regular use of more potent forms of cannabis. The effects of these changes on risks of dependence and acute harms from cannabis use are not yet well characterized.

DIAGNOSTIC AND CLINICAL FEATURES

Cannabis Dependence

For much of the 1960s and 1970s, cannabis was not regarded as a drug of dependence because it was not believed to produce tolerance or a withdrawal syndrome. The adoption of a broader concept of dependence in the late 1970s and early 1980s reduced the importance of tolerance and withdrawal in defining dependence and placed more emphasis on the compulsion to use, a narrowing of the drug-using repertoire, rapid reinstatement of dependent use after abstinence, and the high salience of drug use in the user's life. Since the 1980s, a cannabis dependence syndrome has been identified as a common experience in regular cannabis users.

Cannabis Use Disorder in DSM-5 or DSM-5-TR. Changes in the DSM-5 or DSM-5-TR criteria from the fourth edition version for cannabis and other drug use disorders raise questions about the relevance of epidemiologic studies that used earlier DSM diagnostic criteria. In earlier versions of the DSM, cannabis abuse and dependence were separate diagnostic categories within the cannabis use disorder category. In DSM-5 and DSM-5-TR, cannabis use disorders are diagnosed consistently without separate criteria for abuse or dependence. Studies that compared how the criteria affect prevalence showed that more people are classified as having moderate to severe cannabis use disorders when using the DSM-5 criteria than the DSM-IV criteria, and a slightly different group is diagnosed.

The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) study estimated that in 2001 to 2002, 1.5% of U.S. adults met DSM-IV criteria for past-year cannabis use disorders. In 2012 to 2013, the NESARC-III study found that 2.5% of adults in the United States met DSM-5 criteria in the past year. Yearly population surveys conducted by the National Surveys on Drug Use and Health study estimated that, between 2002 to 2017, the past-year prevalence of cannabis use disorders remained stable at around 1.5% based on DSM-IV criteria, while it increased slightly from 2.1% to 2.6% based on DSM-5 criteria.

The Risk of Cannabis Use Disorder. People who use cannabis daily or near daily over weeks to months are most likely to become dependent. Longitudinal studies that followed young people into adult life found that 1 in 10 of those who have ever used cannabis developed dependence later in life. The risk increased to 1 in 3 among those who used cannabis regularly.

Research is underway to provide risks of cannabis use disorders defined in DSM-5 or DSM-5-TR. A recent survey found that one in three adults who have used cannabis had met the criteria for a cannabis use disorder at some time in their lives. Future longitudinal studies can use DSM-5-TR to assess the risk of cannabis use disorder among people who have used cannabis by the severity of the disorder.

As a rule, the earlier the age of first use, the more often cannabis has been used, and the longer it has been used, the higher the risk of dependence. Other factors that predict a higher risk of dependence include poor academic achievement, conduct disorder in childhood and adolescence, nonconformity and rebelliousness, personal distress and maladjustment, poor parental relationships, parental history of drug and alcohol problems, and heavier cannabis use.

Clinical Populations. Only a minority of people with a cannabis use problem seek treatment. In 2017, cannabis was the primary drug of concern for 12.4% of clients admitted to U.S. addiction treatment agencies. There were 244,117 treatment admissions with cannabis as the primary substance in 2017, compared to 240,271 admissions in 2015. There has also been an increase in the number seeking treatment for cannabis problems in the Netherlands, where the personal use of cannabis was decriminalized in the mid-1970s. This suggests that increased treatment seeking in the United States is not solely due to legal coercion. Common symptoms among people seeking help to cease cannabis use include an inability to stop using cannabis, feeling bad about their cannabis use, procrastinating, a loss of self-confidence, memory loss, and withdrawal symptoms when they stop using.

Intoxication

Most young people report using cannabis to experience a “high,” that is, mild euphoria, relaxation, and perceptual alterations, including time distortion, and the intensification of ordinary experiences, such as eating, watching films, listening to music, and engaging in sex. When used in a social setting, the high may be accompanied by infectious laughter, talkativeness, and increased sociability.

Cognitive changes include impaired short-term memory and attention, making it easy for the user to become lost in pleasant reverie and more difficult to engage in sustained goal-directed mental or physical activity. Motor skills, reaction time, motor coordination, and many forms of skilled psychomotor activity are impaired while the user is intoxicated, increasing the risks of accidents if users drive an automobile.

Intoxication Delirium

Psychotic symptoms, such as delusions and hallucinations, may occur at very high doses of THC and in lower doses in individuals with a vulnerability to psychosis, for example, a first-degree affected relative.

High doses of THC have been reported to produce visual and auditory hallucinations, delusional ideas, and thought disorders in normal volunteers. In cannabis-using cultures, such as India, a “cannabis psychosis” has been reported among heavy users of cannabis. In these cases, psychotic symptoms were reportedly preceded by unusually heavy cannabis use and remitted after a period of abstinence.

The existence of a cannabis psychosis in Western cultures is still a matter for debate. In its favor are case series of cannabis psychoses and a small number of controlled studies that compare the characteristics of cannabis psychoses with those of psychoses in individuals who were not using cannabis at the time of hospital admission. Critics of the hypothesis emphasize the fallibility of clinical judgments about etiology, the poorly specified criteria for diagnosing cannabis psychoses, the dearth of controlled studies, and the striking variations in the clinical features of "cannabis psychoses." Nevertheless, there is robust evidence base for cannabis-induced psychotic disorder, with cannabis being a substance that carries an elevated risk of causing this reaction compared to other substances.

Cannabis and Schizophrenia

Schizophrenia and regular cannabis use are associated in community surveys and clinical populations. A substantial number of longitudinal studies also suggest that regular cannabis use can precipitate schizophrenia in vulnerable individuals and exacerbate its symptoms in those who have already developed the disorder. For example, a follow-up of Swedish conscripts found that cannabis use reported in adolescence predicted an increased likelihood of receiving a diagnosis of schizophrenia over a 35-year follow-up. The Swedish findings have been supported by prospective studies in Australia, France, Germany, the Netherlands, and New Zealand that have all found associations between self-reported cannabis use in young adults and later reports of psychotic symptoms. A meta-analysis of these studies provided strong support for the hypothesis that regular cannabis use, especially early-onset cannabis use and use of cannabis with a high THC content, can precipitate schizophrenia.

Meta-analyses of these studies have found that those who had ever used cannabis had 1.4 times higher odds of psychotic symptoms. The odds increased 2.1 times among regular cannabis users and were nearly four times higher in persons who used more potent forms of cannabis. Young people with psychoses or psychotic symptoms who use cannabis have an average age of first episode 2 years earlier than those who did not use cannabis. People with a first episode of psychosis who stop using cannabis have better clinical outcomes than those who persist in using it.

These findings suggest that cannabis use precipitates schizophreniform disorders in vulnerable individuals. It is less certain whether cannabis use can produce psychotic disorders that would not otherwise have occurred. Most of the 274 Swedish conscripts who developed schizophrenia had not used cannabis, and only 7% of cases of schizophrenia could be attributed to cannabis use. Trends in the *treated* incidence of schizophrenia, particularly of early onset, acute cases, during the 1970s and 1980s, when cannabis use increased among young adults in the United States and Australia, have been unclear. Recent studies in Britain and Switzerland have suggested that the incidence of psychoses may have increased in recent birth cohorts with high rates of cannabis use.

The hypothesis that cannabis use is a *contributory* cause of schizophrenia (one factor among others that increase the likelihood of developing the disorder) is more plausible than the alternative hypotheses that (a) persons with schizophrenia use cannabis as a form of self-medication or that (b) cannabis use is associated with other drug use (e.g., amphetamines) that are linked to schizophrenia or (c) of an increased genetic vulnerability to schizophrenia.

Epidemiologic studies have tested the self-medication hypothesis by measuring the temporal sequence between cannabis use and psychosis and only conducting analyses on data from those participants who reported no symptoms before the use of cannabis. Other studies have specifically recruited participants who reported no history of psychotic symptoms or statistically controlled for a prior history of symptoms of mental disorders. These studies report that cannabis use usually preceded the onset of schizophrenia symptoms, suggesting that cannabis use increases the risk of psychosis rather than psychosis increasing cannabis use. Prospective studies of outcomes in people with psychoses who used cannabis before diagnosis and discontinued their use afterward are inconsistent with the self-medication hypothesis. In systematic reviews, people with psychoses who discontinue cannabis use have much lower rates of relapse and fewer positive symptoms than those who continue to use cannabis.

A variety of studies have assessed genetic explanations of the association between cannabis use and schizophrenia. These include studies of the size of the association in cohorts of persons with varying degrees of genetic relationship (e.g., twins, parents, siblings, cousins, and unrelated), Mendelian randomization studies, and studies that have adjusted the association for polygenic risk scores for cannabis use and psychosis. These studies suggest that there are shared genetic risks for both cannabis use and psychosis and that emerging psychosis increases the risks of cannabis use, but there is also consistent evidence that cannabis use plays a small contributory causal role in the development of psychosis.

A contributory causal role for cannabis use is biologically plausible. Psychotic disorders involve disturbances in the dopamine neurotransmitter systems: Drugs that increase dopamine release produce psychotic symptoms when given in large doses (e.g., methamphetamine), and neuroleptic drugs that reduce psychotic symptoms also reduce dopamine levels. THC appears to increase dopamine release; hence it can potentially produce psychotic symptoms. Most directly, double-blind provocation studies indicate that THC produces dose-related increases in psychotic symptoms in persons who do and do not have schizophrenia.

Cannabis and Anxiety

Some cannabis users report acutely increased anxiety, panic, and a fear of going mad. These experiences are most often reported by naive users who are unfamiliar with the effects of the drug, including some patients given THC for medicinal reasons. More experienced recreational users may also report these effects if they use more potent forms of cannabis than usual or if they use the oral route, in which case they may find it difficult to titrate their cannabis dose and experience more pronounced effects of much longer duration than usual.

There is mixed evidence on the relationship between cannabis use and anxiety. Some epidemiologic studies have found associations between cannabis use and specific anxiety disorders, but in some studies, the relationship was no longer significant after common risk factors were taken into account. Some people may use cannabis in an effort to cope with feelings of anxiety, but the regular use of cannabis for this reason may worsen their anxiety symptoms and may add cannabis dependence to their problems, undermining their ability to cope.

Changes in Heart Rate and Blood Pressure

The most immediate effect of smoking cannabis is a 20% to 50% increase in heart rate that occurs within a few minutes to a quarter of an hour of use. Changes in blood pressure can also occur that depend on posture: blood pressure is increased while the person is sitting and decreases on standing. A sudden change from lying down to standing up may produce postural hypotension and feelings of light-headedness and faintness. These are often the earliest indications of cannabis effects in naive users.

The cardiovascular risks of cannabis smoking are probably highest in older adults who have a higher underlying risk of cardiovascular disease. Younger adults with undiagnosed cardiovascular disease who use higher doses of THC may also be at risk. A French study, for example, found that 9.5% of 200 cannabis-related hospitalizations in the Toulouse area between January 2004 and December 2007 were for cardiovascular disorders. This included several cases of myocardial infarction and a fatal stroke in young adults who had no other risk factors for these disorders. An analysis of data between 1990 to 2014 from the U.S. National Vital Statistics System suggests that rates of cardiac deaths had increased in states that had passed medical cannabis laws, especially in males, and in states with a more lax approach to cannabis dispensing.

Cannabinoid Hyperemesis Syndrome

Cannabinoid hyperemesis syndrome (CHS) is a syndrome of cyclical vomiting, nausea, and abdominal pain reported by chronic heavy cannabis users. The symptoms are temporarily relieved by taking a hot shower or bath and ceasing cannabis use. The number of presentations of this syndrome to emergency departments increased at a large hospital in Aurora, Colorado, after cannabis legalization. Rates have also increased nationally across all hospitals in the United States, but it is unclear how much of the increase is due to increased cannabis use and how much to increased detection of cannabis use among cases of a cyclical vomiting syndrome.

Withdrawal and Tolerance

Regular use of THC leads to the development of tolerance to its behavioral and physiologic effects and withdrawal if use ceases. Cannabis withdrawal symptoms typically begin within 24 to 48 hours after cessation of use and may last for 1 to 2 weeks. There are typically no serious risks for individuals going through withdrawal if they are otherwise healthy. The relatively long half-life and complex metabolism of cannabis may explain the fact that cannabis withdrawal syndrome is less intense than that for alcohol or opiates. Animal studies suggest that animals repeatedly dosed with THC show withdrawal symptoms such as scratching, face rubbing, and shakes when given a cannabinoid antagonist.

The DSM-5 or DSM5-TR diagnostic criterion for cannabis withdrawal is met when three or more of the following signs develop within a week of cessation: irritability, anger, or aggression; nervousness or anxiety; sleep difficulties (such as insomnia or disturbing dreams); decreased appetite or weight loss; restlessness; depressed mood; and at least one physical symptom causing severe discomfort from abdominal pain, shakiness or tremors, sweating, fever, chills, or headache. A diagnosis should cause clinically severe distress or impairment in social or occupational settings or other important areas of functioning. Studies of people who have used cannabis regularly and who had to stop using (e.g., due to being in the hospital) report decreased mood and appetite and increased irritability, anxiety, physical tension, and physical symptoms. PET scan studies in humans showed that decreased CB1 cannabinoid receptors were associated with more severe withdrawal symptoms. Administration of CB1 agonists, such as dronabinol/nabilone or nabiximols can reduce withdrawal symptoms. Double-blind trials of a cannabis extract, Sativex, relieved withdrawal symptoms in cannabis-dependent persons undergoing treatment for cannabis dependence.

Disorders Not Otherwise Specified

Amotivational Syndrome. There are case study reports of an “amotivational syndrome” in chronic heavy cannabis users that is characterized by a lack of motivation and interest in activities other than using cannabis. Controlled field studies in the 1970s did not support such a syndrome, although their value was limited by small sample sizes and the restricted sociodemographic characteristics of the participants. Laboratory studies of sustained heavy daily use did not support the syndrome, although these typically involved short periods of drug use in healthy youth, and they made minimal demands on volunteers. Current empirical research does not support such a specific syndrome. The behavior reported among heavy, chronic cannabis users is better described as an effect of chronic daily intoxication that is common among dependent cannabis users.

Cognitive Impairment. The fact that cannabis use acutely impairs cognitive functioning has raised concerns that chronic cannabis use may chronically impair cognition, with particular concerns for youth development and cannabis-impaired driving. There is clinical, electrophysiologic, and epidemiologic evidence that long-term cannabis use produces impairment in the higher cognitive functions of memory, attention, organization, and integration of complex information. Deficits in verbal learning, memory, and attention in regular cannabis users have generally but not always been related to the duration and frequency of cannabis use, the age of initiation, and the estimated cumulative dose of THC received. Impairment is clearest while users are intoxicated. There is some recovery after abstinence, but it is uncertain if cognitive function fully recovers after the cessation of cannabis use.

A longitudinal study in a Dunedin birth cohort suggested that sustained heavy cannabis use over several decades produces declines in cognitive performance that may not be wholly reversible. This study assessed changes in IQ between ages 13 (before cannabis was used) and 38 in 1,037 New Zealanders born in 1972 to 1973. It found that early and persistent cannabis users, who used throughout adulthood, had an average decline in levels of IQ by 8 points greater than those who had not used cannabis at all and cannabis users who had not used cannabis in this sustained way. Nevertheless, other similar longitudinal studies fail to show any drop in IQ among heavy users; therefore, more research is needed to better understand the long-term impact of cannabis use on cognition.

Effects on Adolescent Development. Longitudinal studies of adolescents indicate that chronic heavy cannabis use may adversely affect adolescent psychosocial development. Interpretation of this evidence is complicated by the fact that many of the risk factors for these adverse developmental outcomes precede cannabis use and also make it more likely that a young person will become a regular user of cannabis. These outcomes include minor delinquency, other substance use, poor educational performance, nonconformity, and poor social adjustment.

Among U.S. adolescents in the 1970s and 1980s, the typical sequence of initiation into drug use was that alcohol and tobacco use preceded cannabis use, which, in turn, preceded the use of hallucinogens, amphetamine, heroin, and cocaine. In general, the earlier the age of first use of any drug and the greater the involvement with the drug, the more likely a young person will use the next drug in the sequence. The specific order in which particular drugs are used depends on the availability, marketing, and levels of use among adults in a given society at the time.

The role of cannabis in this sequence is debated. The hypothesis that cannabis use has a pharmacologic effect that increases the risk of using later drugs in the sequence is not the most compelling explanation. More plausible hypotheses are that the sequence of drug involvement reflects a combination of (a) the early recruitment to drug use of nonconforming and deviant adolescents who are more likely to use alcohol, tobacco, and illicit drugs; (b) a shared genetic vulnerability to become dependent on alcohol, cannabis, and tobacco; and (c) the socialization of cannabis users in an illicit drug-using subculture that increases the exposure, opportunity, and encouragement to use other illicit drugs. Recent prospective studies that have controlled for many of these factors and studies of twin pairs discordant for cannabis use have reduced the magnitude but failed to eliminate the apparent “gateway effect” of cannabis. Such studies may not be able to disentangle the complicated relationships to produce conclusive answers on the pharmacologic effects of cannabis. This leaves open the possibility of some form of pharmacologic sensitization, in addition to social selection in illicit drug use, which, however, is unlikely to be a cannabis-specific feature.

In adolescents, regular cannabis use is associated with a higher risk of failing to complete high school and job instability in young adulthood. The complication in interpreting this finding is that young people who regularly use cannabis had lower academic aspirations and poorer school performance *before* they started using it. When taking these differences into account, the relationship between cannabis use and poor educational performance persists but is more modest in size. Nonetheless, the effects of cannabis and other drug use on educational performance and underachievement limit the choice of adult occupation, level of income, choice of mate, and quality of adult life.

There is some evidence that heavy cannabis use has adverse effects on family formation and increases involvement in drug-related crime. However, the strong associations found in cross-sectional survey data are more modest in longitudinal studies that control for associations between cannabis use and other variables that predict these adverse outcomes. The relationship between cannabis use and crime may also change with the legality of cannabis use and sales.

Flashbacks. There are case reports of flashbacks in which users re-experience symptoms of cannabis intoxication days or weeks after last using cannabis. It has been difficult to decide whether these are (a) rare events that coincide with cannabis use; (b) the effects of other drugs that are often taken together with cannabis; (c) rare consequences of cannabis use that only occur at doses that are much higher than those used recreationally; (d) the result of unusual forms of personal vulnerability; or (e) the results of interactions between cannabis and other drugs. Such recurrence of perceptual symptoms is well documented among those using hallucinogens, which is often precipitated by the use of cannabis.

ADVERSE EFFECTS OF CANNABIS USE

Psychomotor Effects and Driving

Cannabis intoxication impairs a wide range of cognitive and behavioral functions that are involved in safely driving automobiles or operating machinery. These effects are generally larger, more consistent, and more persistent in tasks that require sustained attention. Animal studies showed that THC impairs balance, movement, and mobility. Recreational doses of THC produce similar impairments in human performance on laboratory tests and standardized driving courses to blood alcohol concentrations (BACs) of between 0.07% to 0.10%. Studies of the effect of cannabis use on driving performance on driving courses and on the road have found more modest impairments, probably because cannabis-impaired drivers drive more slowly, and take fewer risks, than alcohol-impaired drinkers. Regular cannabis users may be more aware of their psychomotor impairment than alcohol users.

Epidemiologic studies have found that self-reported cannabis use was associated with an approximately twofold relative risk of hospitalization for accidental injury, which has persisted after controlling for alcohol and tobacco use in men. Studies that have measured THC in blood have found a dose–response relationship between THC in blood and the risk of accidents. The combination of THC and alcohol produces more impairment and increased accident risk. The size of the risk of accidents among cannabis users is more modest than that for alcohol, but it is still of potential public health significance given the high prevalence of cannabis use among young males.

There is mixed evidence on the impacts of cannabis legalization on motor vehicle accidents. Overall, rates of motor vehicle accidents have declined in the United States since the early 2000s. Roadside testing and surveys in California did not detect increases in the prevalence of drivers with THC in oral fluid after cannabis was decriminalized in 2010. On the other hand, investigations of the Fatality Analysis Reporting System have shown an increase in the prevalence of drivers testing positive for cannabinoids in blood/urine among those involved in a fatal motor vehicle accident. It is unclear how much of this increase is explained by increased testing for cannabis use.

The Cardiovascular System

The acute toxicity of cannabis and cannabinoids is very low. There are no cases of fatal cannabis poisoning in the human medical literature. Animal studies indicate that the dose of THC required to produce 50% mortality in rodents is extremely high by comparison with the equivalent doses of other pharmaceutical and recreational drugs. The lethal dose also increases as one moves up the phylogenetic tree, suggesting that the lethal dose in humans could not be achieved by smoking or swallowing cannabis.

In healthy young cannabis users, the cardiovascular effects of cannabis use are likely to be of minimal clinical significance because tolerance develops to the effects of THC, and young, healthy hearts will only be mildly stressed. These changes in heart rate and blood pressure may be less benign in older adults with hypertension, cerebrovascular disease, and coronary atherosclerosis. Several epidemiologic studies suggest that cannabis use may provoke a myocardial infarction in middle-aged adults with underlying heart disease. The public health implications of these effects may increase as the cohort of continuing cannabis users from the late 1970s and early 1980s enters old age, the period of maximum risk for atherosclerosis in the heart, brain, and peripheral blood vessels.

There are recent case reports of serious cardiovascular events in young males who are very heavy daily cannabis smokers. Hospital data between 1990 and 2014 showed increased rates of cardiac-related deaths following the passage of medical cannabis policies, especially in older men, and in states with more a lax approach to cannabis dispensing. Epidemiologic studies are needed to evaluate the impact of more potent cannabis products on the cardiovascular system, given their increased availability on legal cannabis markets in some states in the United States.

Given the growing popularity of edibles, however, the possibility for inadvertent exposure to large doses becomes highly problematic, particularly for young children who may accidentally consume these products. Indeed, there are numerous case reports of severe cannabis exposures requiring medical treatment among pediatric patients following legalization of medical and recreational cannabis.

The Respiratory System

Regular cannabis smoking impairs large airway function and causes symptoms of chronic bronchitis such as coughing, sputum, and wheezing. It is possible that chronic cannabis smoking increases the risks of respiratory cancers because tobacco and cannabis smoke contain similar carcinogenic substances and tobacco smoke has adverse effects on the respiratory system. Case-control studies of aerodigestive tract cancers and cannabis smoking have produced conflicting findings, and their results are often unclear because most of the cannabis smokers in these studies also smoked tobacco. A recent meta-analysis of these studies concluded that there was at most a weak association between regular cannabis smoking and lung cancer. The increasing trend of vaping instead of smoking a combustible cannabis cigarette probably means lower respiratory harm from cannabis use, but the long-term impacts of vaping are yet to be studied.

Cellular Effects and Cancers

There is very weak evidence that THC alters cell metabolism and deoxyribonucleic acid (DNA) synthesis in the test tube. There is stronger evidence that cannabis *smoke* produces mutations in cells in the test tube and in live animals and hence is a potential cause of cancers of the lung and upper aerodigestive tract.

Aerodigestive tract cancers have been reported among young adults who have been daily cannabis users. A systematic review and meta-analysis concluded that the relationships between cannabis exposure and cancers of the lung and upper aerodigestive tract, and also head and neck cancers, were not statistically significant after adjusting for tobacco and alcohol use. These observations were based on studies with few cannabis-only smokers. Future investigation is needed given the changing patterns of cannabis use in the population. There is weak evidence for an increased risk of cancers among children born to women who smoked cannabis during pregnancy. Three studies of very different types of cancer have reported an association with maternal cannabis use. However, none of these studies was designed to study the role of cannabis use in these cancers, so replication is required. There have not been any increases in the rates of these cancers that parallel increased rates of cannabis use over the past three decades.

There is more consistent evidence of an increased risk of testicular cancer among cannabis users. A case-control study that found an increased risk of nonseminoma testicular cancer has been replicated in two case-control studies. These found a doubling of risk for nonseminoma testicular tumors and found that the risk increased with earlier initiation and more frequent use. The effect is also biologically plausible since cannabinoid receptors are found in the male reproductive system.

Immunologic Effects

Cannabinoids impair cell-mediated and humoral immunity in rodents and reduce resistance to bacterial and viral infections in animals. Cannabinoid receptors are expressed in cells of the immune system in animals and humans, although their clinical significance is unclear. Cannabis smoke also impairs the functioning of alveolar macrophages, the first line of the body’s immune defense system in the lungs. The clinical relevance of these findings is uncertain because the doses that produce these effects have been high, and extrapolation to the doses used by humans is complicated by the fact that tolerance may develop to these effects.

The limited experimental and clinical evidence in humans suggests that the adverse effects seen in animals are not replicated in humans. There is no conclusive evidence that cannabinoids impair immune system function in humans, as measured by T lymphocytes, B lymphocytes, macrophages, or immunoglobulin levels. There is suggestive evidence that THC impairs T lymphocyte responses to mitogens and allogeneic lymphocytes. The clinical and biologic significance of any immunologic effects of cannabis use is uncertain because there is no robust epidemiologic evidence of increased or decreased rates of infectious disease among chronic heavy cannabis users.

Reproductive Effects

Chronic administration of THC disrupts the male and female reproductive systems in animals. It reduces testosterone secretion and sperm production, motility, and viability in males, and disrupts the ovulatory cycle in females. There are inconsistent findings on whether cannabis use has these effects on humans.

Use During Pregnancy

Cannabis use during pregnancy produces babies with smaller birth weights, possibly because of a shorter gestation, and probably by the same mechanism as cigarette smoking. There is no clear evidence that cannabis use during pregnancy increases the risk of birth defects due to the fetus' exposure to cannabis in the uterus. Infants exposed to cannabis in the uterus may show transient behavioral and developmental effects during the first few months after birth. These effects are small by comparison with those caused by tobacco use during pregnancy, and they have not been observed in all studies. In previous studies of women who used cannabis during pregnancy, co-use of other substances such as tobacco and alcohol was common. We currently have little data to single out the effects of cannabis on birth outcomes. Future studies may provide answers, given that prenatal prevalence of tobacco and alcohol use has decreased, but the use of cannabis has not, especially in women with nausea and vomiting, along with increased marketing of cannabis to manage morning sickness.

LABORATORY EXAMINATIONS

Cannabinoids can be detected in the head hair, pubic hair, urine, sweat, saliva, and blood of people who use cannabis. They are stored in the fat cells of the body and so can remain in the body for a more extended period than other drugs, and THC may be detectable in urine for up to 11 weeks after use. Some research has suggested that higher concentrations of cannabinoids may be found in pubic than head hair. THC may be detected in saliva and sweat, but the concentration in these fluids tends to be lower than in urine, and in some cases, may not be detected in these fluids.

Cannabinoid levels in the blood vary between individuals and depend on the dose received and the individual's history of cannabis use. Blood levels of THC typically ranged between 0 and 500 ng/mL after flower cannabis use and up to over 1,000 ng/mL after using 70% to 90% THC concentrates that can be purchased from dispensaries. The level depends on the potency of the cannabis and the time since use. Levels of THC in blood above 10 to 15 ng/mL indicate recent use, although it is difficult to be precise about how recent. A more precise estimate of time since last use is provided by the ratio of THC to 9-carboxy-THC. When the blood concentrations of THC and 9-carboxy-THC are similar, cannabis has been used in the past 20 to 40 minutes, and hence suggests a high probability of intoxication. These relationships are less clear in regular users.

Cannabis intoxication impairs skills required to drive a motor vehicle, so it would be desirable to have a measure of cannabis intoxication similar to the breath test for alcohol intoxication. The major challenge is the lack of a simple relationship between THC metabolites in blood and psychomotor impairment, because THC can be detected after intoxication effects have passed. Saliva tests for THC may provide evidence of recent use that can be confirmed by analyses of THC levels in the blood. Consensus is yet to be reached on a level of THC in the blood that provides evidence of impairment analogous to the use of BACs of 0.05% or 0.08%. The American Automobile Association for Traffic Safety position is that we do not currently have scientific evidence to support a quantitative THC threshold for cannabis-impaired driving. The U.S. states of Colorado and Washington have defined driving with blood THC greater than 5 ng/mL as evidence of cannabis-impaired driving that must be supported by evidence of behavioral impairment.

With repeated frequent use of cannabis, THC accumulates in fatty tissues in the human body, where it may remain for considerable periods. The significance of this storage for users is unclear. It would be a serious concern if THC were a highly toxic substance that was physiologically active while stored in body fat. THC is not, however, a highly toxic substance, and it is inactive while stored in fat.

More recently, a study utilizing near-infrared spectroscopy was able to reasonably identify impairment from cannabis by examining signal in the prefrontal cortex. Given the challenges in using blood levels to infer impairment, such alternative methods may become important adjuncts to toxicology testing.

TREATMENT

Cannabis Use Disorder and Withdrawal

Cannabis use disorder and cannabis withdrawal symptoms can hinder quitting. In patients treated for cannabis use disorders, 54% of outpatients and 87% of inpatients report clinically severe symptoms of withdrawal. Randomized controlled trials have compared different ways to assist cannabis-dependent persons to stop using cannabis. Most people with cannabis use disorders, who are otherwise healthy, benefit from supportive care.

Behavioral interventions to help with cannabis withdrawal include psychoeducation and coping skills training. Information is provided to the patient about what they should expect regarding symptoms and time course, and ways to manage symptoms (e.g., exercise or hot baths to manage irritability, avoiding excessive caffeine to address restlessness, consuming nutritious food to counter decreased appetite, and keeping in mind that symptoms are temporary).

Psychosocial interventions for adults with cannabis use disorders include cognitive behavioral therapy (CBT), motivational enhancement therapy (MET) including brief MET (bMET), contingency management, social support counseling, drug education counseling, relapse prevention, mindfulness meditation, and mutual-help groups, based on the 12-step approaches (Marijuana Anonymous). CBT, MET, relapse prevention, and contingency management have received more research attention; findings showed that they are more effective than placebo and treatment as usual. Studies have reported that combining CBT and MET, as well as contingency management results in better outcomes.

For adolescents seeking treatment for cannabis use disorders, psychosocial interventions include individual-, group-, or family-based interventions. In addition to CBT, MET, and contingency management for the individual adolescent, studies have shown positive outcomes from family therapy.

The type of intervention used may depend on time, access, personnel, and health service system costs. Contingency management is costly with more provider burden, and the approach may be resisted by some providers. CBT typically involves more sessions over a longer period than MET. Brief interventions, telephone-based interventions, and online interventions are lower in costs, but research has not yet established their effectiveness for these emerging modes of interventions.

There is an absence of approved pharmacologic treatments for cannabis use disorders. A few studies have found that CB1 agonists (e.g., dronabinol or Sativex) can provide relief for people with higher severity of withdrawal symptoms but do not improve rates of long-term abstinence. Zolpidem and other benzodiazepines (nitrazepam) have been used to manage sleep problems. Studies of medications to manage mood, sleep, or craving have produced mixed results. Systematic reviews found limited support for the effectiveness of selective serotonin reuptake inhibitors (SSRIs), the antidepressant bupropion, the anxiolytic buspirone, and the selective noradrenaline reuptake inhibitor atomoxetine. Of note, a phase 2 trial for CBD did show improved cannabis use outcomes compared to placebo, raising the possibility for the therapeutic potential of CBD.

Therapeutic Effects of Cannabinoids

When taken for medicinal purposes, cannabinoids are primarily used to relieve symptoms rather than treat diseases. They may be used, for example, to relieve nausea and vomiting caused by cancer chemotherapy and muscle spasticity and chronic neuropathic pain in neurologic disorders such as multiple sclerosis (MS). The following cannabinoid-containing products have been authorized for medical use:

- Dronabinol: oral capsules containing synthetic THC for (1) anorexia in patients with acquired immune deficiency syndrome (AIDS) and (2) nausea and vomiting associated with cancer chemotherapy.
- Nabilone: oral capsules containing a synthetic cannabinoid similar to THC for the same indications as dronabinol.
- Nabiximols: an orolingual preparation containing approximately equal quantities of THC and CBD cannabis extracts approved for the treatment of muscle spasticity resulting from MS.
- CBD: an oral solution for the treatment of seizures associated with Lennox–Gastaut syndrome or Dravet syndrome in children.

Evidence from controlled clinical trials indicates that cannabinoids relieve the symptoms of some illnesses when used as adjunctive treatments and usually only after a patient has failed to respond to recommended treatments.

Chronic Noncancer Pain. Cannabis and cannabinoid products are increasingly used to manage chronic noncancer pain. A systematic review of existing evidence found limited effectiveness of cannabinoids on neuropathic pain, fibromyalgia, rheumatoid arthritis, and chronic noncancer pain in general. There is some evidence that cannabinoids were more effective than placebo in pain management, but positive effects were more commonly observed in 1-day studies, with smaller or no effects in studies with 3 months or longer follow-up. Numbers needed to treat to benefit and to harm analyses showed that cannabis was only modestly superior to placebo.

Nausea and Vomiting. Controlled clinical trials have compared the antiemetic effects of oral THC with either a placebo or an antiemetic drug in patients with nausea and vomiting related to cancer chemotherapy. Cannabinoids were more effective than placebo in reducing nausea and vomiting in adults receiving chemotherapy, but patients were more likely to withdraw due to adverse side effects. Systematic reviews have concluded that THC and other cannabinoid agonists had similar effectiveness to older antiemetic drugs in treating chemotherapy-induced nausea and vomiting. There is a lack of research comparing cannabinoids with newer antiemetic drugs such as ondansetron.

Wasting Syndrome and Appetite Stimulation in HIV/AIDS. THC was approved in the United States in 1999 for use as an appetite stimulant in patients with AIDS-related wasting syndrome based on a few small clinical trial cases that a systematic review found provided weak evidence for its effectiveness. A meta-analysis found no difference between cannabinoids and placebo in treating nausea or vomiting in HIV patients but did find increased appetite and side effects on mental health. The advent of highly effective antiretroviral therapy means that very few HIV-infected persons now develop AIDS and require appetite stimulation.

Muscle Spasticity and Multiple Sclerosis. Muscle spasticity is the increased resistance to passive stretching of muscles. Involuntary contractions may occur, which can be painful and debilitating. About 90% of MS patients develop muscle spasticity, as do a substantial proportion of patients with spinal cord injuries. Systematic reviews of trials of the effects of pharmaceutical cannabinoids on these MS symptoms have found that they are modestly effective in reducing symptoms of pain and muscle spasticity and produce mild to moderate adverse events.

Movement Disorders. Movement disorders are caused by abnormalities in areas of the brain connected to the motor cortex that produces abnormal skeletal muscle movements in the face, limbs, and trunk. The movement disorders most often mentioned as candidates for medical use of cannabinoids are dystonia, Huntington disease, Parkinson disease, and Tourette syndrome. Two small trials found limited evidence that cannabis is useful for treating any of these disorders. A review of the evidence concluded that pharmaceutical THC did not improve symptoms of Tourette syndrome and increased rates of adverse events.

Epilepsy. There are controlled trials indicating that CBD can reduce the frequency of seizures in some genetic epilepsy syndromes in children. A systematic review of trials of CBD as an adjunctive treatment for treatment-resistant epilepsy in children showed that it was more effective than placebo at reducing seizure frequencies and improving quality of life, but it increased risks of adverse events. CBD (which has no psychoactive effects) is not a controlled substance, and it has been approved by the U.S. Food and Drug Administration and medicines agencies in other countries as an adjunctive treatment for these types of epilepsy (Epidiolex).

Glaucoma. Glaucoma is a condition that damages the optic nerve via high intraocular pressure (IOP) and can produce blindness if untreated. Cannabis has been used as a potential treatment for reducing pressure in the eye. A systematic review to search for good quality evidence on medicinal cannabis for the treatment of glaucoma identified one trial conducted with six participants. The trial’s findings did not support that cannabinoids (THC, CBD, and CBD spray) were effective in treating glaucoma. There are reports from lower quality studies that cannabis and THC taken orally, by topical eye drops, or intravenously reduce IOP by 25%, but the effect lasts only 3 to 4 hours, which would not help to prevent blindness.

Other Uses for Medicinal Purposes. Patient advocacy groups have campaigned for the right to use cannabinoids to treat a variety of conditions for which there is limited evidence of efficacy. These include anxiety, post-traumatic stress, depression, sleep, neurologic conditions, and inflammatory bowel diseases. For many of these conditions, there is limited or no robust scientific evidence to support their effectiveness, and in some instances, the evidence has found adverse effects from their use. Strong levels of advocacy for cannabinoids used for conditions that do not yet have an empirical basis has often stemmed from personal perceptions, case reports, and anecdotal evidence often uncritically communicated through traditional and social media.

CONCLUSION

The definition of cannabis-related disorders has evolved over time. The prevalence of cannabis use in youth and adults has fluctuated across years with changing social attitudes and policies. A 2020 U.S. population survey showed that 46% of persons had ever used cannabis and 21% had done so in the past year. Cannabis use by vaping has been an increasing trend in U.S. adolescents. Approximately 1 in 10 people who use cannabis develop cannabis dependence later in life, and the risk increases to 1 in 3 among those who use cannabis daily or near daily.

With the liberalization of cannabis use in the United States, there has been a boom in new cannabis products offering higher THC concentrations at lower prices promoted via social media. THC interacts with the brain to affect mood, motor control, cognition, memory, pain, and the immune system. The use of cannabis, especially high THC products, has been linked to negative outcomes, such as psychosis, cardiovascular events, hyperemesis syndrome, respiratory problems, cognitive impairment, and a range of withdrawal symptoms.

Only a minority of people with cannabis use problems seek treatment. Trials have found positive treatment outcomes from behavioral interventions and support, and psychosocial interventions such as CBT, MET, and contingency management. No pharmacologic treatments have been approved for cannabis use disorders. There has been increasing attention on the therapeutic use of cannabinoids to relieve symptoms of medical conditions such as nausea and vomiting, chronic pain, MS, and epilepsy. With the population's evolving relationship with cannabis, research is warranted on more effective treatments for cannabis use disorders and more rigorous evaluations of the use of cannabinoids to treat psychological and physical health conditions.

FURTHER READINGS

Black N, Stockings E, Campbell G, et al. Cannabinoids for the treatment of mental disorders and symptoms of mental disorders: a systematic review and meta-analysis. *Lancet Psychiatry*. 2019;6(12):995–1010.

Carliner H, Brown QL, Sarvet AL, Hasin DS. Cannabis use, attitudes, and legal status in the U.S.: a review. *Prev Med*. 2017;104:13–23.

Caulkins JP, Kilmer B, Kleiman MAR. *Marijuana Legalization: What Everyone Needs to Know*. Oxford University Press; 2016.

Chandra S, Radwan MM, Majumdar CG, Church JC, Freeman TP, ElSohly MA. New trends in cannabis potency in USA and Europe during the last decade (2008–2017). *Eur Arch Psychiatry Clin Neurosci*. 2019;269(1):5–15.

Compton WM, Han B, Jones CM, Blanco C. Cannabis use disorders among adults in the United States during a time of increasing use of cannabis. *Drug Alcohol Depend*. 2019;204:107468.

Connor JP, Stjepanović D, Le Foll B, Hoch E, Budney AJ, Hall WD. Cannabis use and cannabis use disorder. *Nat Rev Dis Primers*. 2021;7(1):16.

Degenhardt L, Bharat C, Bruno R, et al; WHO World Mental Health Survey Collaborators. Concordance between the diagnostic guidelines for alcohol and cannabis use disorders in the draft ICD-11 and other classification systems: analysis of data from the WHO's World Mental Health Surveys. *Addiction*. 2019;114(3):534–552.

Freeman TP, Craft S, Wilson J, et al. Changes in delta-9-tetrahydrocannabinol (THC) and cannabidiol (CBD) concentrations in cannabis over time: systematic review and meta-analysis. *Addiction*. 2021;116(5):1000–1010.

Gage SH, Jones HJ, Burgess S, et al. Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study. *Psychol Med*. 2017;47(5):971–980.

Ghasemiesfe M, Barrow B, Leonard S, Keyhani S, Korenstein D. Association between marijuana use and risk of cancer: a systematic review and meta-analysis. *JAMA Netw Open*. 2019;2(11):e1916318.

Gobbi G, Atkin T, Zytynski T, et al. Association of cannabis use in adolescence and risk of depression, anxiety, and suicidality in young adulthood: a systematic review and meta-analysis. *JAMA Psychiatry*. 2019;76(4):426–434.

Goodman S, Wadsworth E, Leos-Toro C, Hammond D; International Cannabis Policy Study Team. Prevalence and forms of cannabis use in legal vs. illegal recreational cannabis markets. *Int J Drug Policy*. 2020;76:102658.

Hall W, Leung J, Lynskey M. The effects of cannabis use on the development of adolescents and young adults. *Ann Rev Dev Psychol*. 2020;2:461–483.

Hall W, Stjepanović D, Caulkins J, et al. Public health implications of legalising the production and sale of cannabis for medicinal and recreational use. *Lancet*. 2019;394(10208):1580–1590.

Hammond D, Goodman S, Wadsworth E, Rynard V, Boudreau C, Hall W. Evaluating the impacts of cannabis legalization: the International Cannabis Policy Study. *Int J Drug Policy*. 2020;77:102698.

Han BH, Palamar JJ. Trends in cannabis use among older adults in the United States, 2015–2018. *JAMA Intern Med*. 2020;180(4):609–611.

Hasin DS. US epidemiology of cannabis use and associated problems. *Neuropsychopharmacology*. 2018;43(1):195–212.

Johnston, L. D., Miech, R. A., O'Malley, P. M., Bachman, J. G., Schulenberg, J. E., & Patrick, M. E. (2020). Monitoring the Future national survey results on drug use 1975-2019: Overview, key findings on adolescent drug use. Ann Arbor: Institute for Social Research, University of Michigan.

Leung J, Chan GCK, Hides L, Hall WD. What is the prevalence and risk of cannabis use disorders among people who use cannabis? a systematic review and meta-analysis. *Addict Behav*. 2020;109:106479.

Leung J, Chiu V, Chan GCK, Stjepanović D, Hall WD. What have been the public health impacts of cannabis legalisation in the USA? A review of evidence on adverse and beneficial effects. *Curr Addict Rep*. 2019;6:418–428.

Lorenzetti V, Hoch E, Hall W. Adolescent cannabis use, cognition, brain health and educational outcomes: a review of the evidence. *Eur Neuropsychopharmacol*. 2020;36:169–180.

Marconi A, Di Forti M, Murray RM, Lewis CM, Vassos E. Meta-analysis of the association between the level of cannabis use and risk of psychosis. *Schizophr Bull*. 2016;42(5):1262–1269.

Murray RM, Hall W. Will legalization and commercialization of cannabis use increase the incidence and prevalence of psychosis? *JAMA Psychiatry*. 2020;77(8):777–778.

NASEM. *The Health Effects of Cannabis and Cannabinoids: The Current State of Evidence and Recommendations for Research*. National Academies Press; 2017. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK423845

SAMHSA. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2019 National Survey on Drug Use and Health.* Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health

Services Administration; 2020. https://www.samhsa.gov/data/report/2019-nsduh-annual-national-report

Stockings E, Campbell G, Hall WD, et al. Cannabis and cannabinoids for the treatment of people with chronic noncancer pain conditions: a systematic review and meta-analysis of controlled and observational studies. *Pain*. 2018;159(10);1932–1954.

Stockings E, Zagic D, Campbell G, et al. Evidence for cannabis and cannabinoids for epilepsy: a systematic review of controlled and observational evidence. *J Neurol Neurosurg Psychiatry*. 2018;89(7):741.

UNODC. *World Drug Report 2020: Booklet 2: Drug Use and Health Consequences*. United Nations; 2020. https://wdr.unodc.org/wdr2020/field/WDR20_Booklet_2.pdf

Young-Wolff KC, Sarovar V, Tucker LY, et al. Trends in marijuana use among pregnant women with and without nausea and vomiting in pregnancy, 2009–2016. *Drug Alcohol Depend*, 2019;196:66–70.

Zimmermann M, Chong AK, Vechiu C, Papa A. Modifiable risk and protective factors for anxiety disorders among adults: a systematic review. *Psychiatry Res*. 2020;285:112705.

11.6 Hallucinogen-Related Disorders

BRIAN S. BARNETT, M.D., AND JOJI SUZUKI, M.D.

Many substances can, if consumed in sufficient quantity, dramatically change one's perception, cognition, mood, affect, social relatedness, sense of meaning, and sense of self. However, hallucinogens are unique in their ability to profoundly alter these aspects of consciousness without simultaneously inducing delirium. They are broadly conceived in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR), with the hallucinogen class of substances encompassing a wide range of sources (naturally occurring, semisynthetic, or fully synthetic), structures, mechanisms of action, effects, durations of action, and potential for harm/therapeutic benefit. Nevertheless, the more commonly encountered hallucinogens include 5-methoxy-N, N-dimethyltryptamine (5-MeO-DMT), dimethyltryptamine (DMT), lysergic acid diethylamide (LSD), mescaline, psilocybin, and other serotonin 2A receptor agonists; entactogens such as 3,4-methylenedioxymethamphetamine (MDMA); dissociative arylcyclohexamines such as phencyclidine (PCP) and ketamine; and kappa-opioid agonists such as salvinorin A.

The utility of grouping such a wide array of compounds into a single class, as well use of the term *hallucinogen* to describe them, are subjects of debate. Though hallucinogen remains the official term for these compounds, as per the DSM-5-TR, it is an imprecise descriptor given both its narrow focus on the perceptual aspects of the complex alterations of consciousness these compounds produce and the fact that many hallucinogens actually produce illusions much more commonly than strict hallucinations. Alternatively, sometimes these compounds have been referred to as *club drugs* given their propensity, particularly for MDMA or ketamine, to be used in club settings. However, the term *psychedelic*, derived from combining the Greek terms psykhē, or "mind," and dēloun, or "make visible, reveal," has been proposed to describe hallucinogens in a broad sense and, in particular, classic hallucinogens—those hallucinogens with effects primarily mediated by agonism or partial agonism at the 5HT2A receptor. This term emphasizes the profound effects on sense of self occasioned by classic hallucinogens and has already displaced hallucinogen within common parlance. PCP and related arylcyclohexamines such as ketamine have historically been referred to as *dissociatives* or *dissociative anesthetics* but are classified as hallucinogens in the DSM-5-TR and are increasingly placed under the umbrella of psychedelics. The terms *empathogen* and *entactogen* have been proposed for 3,4-methylenedioxymethamphetamine (MDMA) and related compounds whose effects are dominated by feelings of emotional closeness to others. The term *entheogen* has also been used to refer to hallucinogens when the spiritual or religious aspect of the experience produced is of primary importance.

The DSM-5-TR recognizes two categories of hallucinogens: PCP and related arylcyclohexamines, and "other hallucinogens," which comprises several distinct and sometimes overlapping classes of hallucinogens. Because of the great diversity of drugs labeled hallucinogen, each of the main classes of hallucinogens will be discussed briefly before we present the hallucinogen-related disorders. Characteristics of hallucinogen classes are summarized in Table 11.6–1. Photographs of commonly used synthetic hallucinogens and sources of commonly used naturally occurring hallucinogens are featured in Figure 11.6–1.

Psychiatry and hallucinogens share a complicated, still-evolving relationship that has been the subject of renewed interest in recent years. Considered primarily as substances of misuse by the field for much of the last half-century, some hallucinogens, such as LSD, were once employed as adjuvants to psychotherapy, while ketamine and esketamine have recently joined the psychiatric armamentarium. Still more, such as MDMA and psilocybin, may soon join or rejoin it if clinical trials continue to demonstrate promising therapeutic results. Should the clinical applications of hallucinogens become more widespread, one concern is that harms from nonmedical use, rarely encountered by clinicians now, may be seen with increasing frequency in years to come. Therefore, improved training for clinicians on prevention and treatment of adverse effects secondary to hallucinogen use, as well as expanded research into identifying predictors of and treatments for these effects, may prove instrumental for minimizing hallucinogens' societal harms should they come into broader clinical use.

HISTORICAL OVERVIEW

Some hallucinogens, such as ayahuasca and peyote, have been employed ritualistically for spiritual and medicinal purposes by indigenous peoples for millennia. In the New World, hallucinogen use among these groups was suppressed by colonial conquerors for centuries. Today, use of some hallucinogens by indigenous groups for spiritual purposes, such as peyote use by members of the Native American Church within the United States, is now legally protected. However, possession of fungi- and plant-based hallucinogens remains illegal across the world, though a growing movement to decriminalize possession is currently underway.

Hallucinogens first came to mainstream attention in Western culture in the 1950s, following the discovery of LSD's psychoactive effects in 1943 by Swiss chemist Albert Hoffmann of Sandoz Laboratories. In subsequent decades, enthusiasm developed regarding potential psychiatric applications of LSD and other serotonergic hallucinogens. LSD was initially employed by psychiatrists to produce a "model psychosis" capable of aiding clinicians in better understanding psychotic disorders. Prior to LSD's discovery, no other compounds at the submilligram dose were known to produce profound alterations in cognition and perception, leading to speculations that LSD may resemble an endogenous ligand responsible for psychosis and other mental illnesses. Mescaline's potential for producing a model psychosis had been explored to a lesser degree starting in the 1920s, though the compound did not gain the mainstream scientific attention of LSD. The utility of employing LSD as a potential tool for inducing psychosis was eventually determined to be limited since true hallucinations are more frequently produced by other compounds, such as stimulants. Subsequently, scientific interest in hallucinogens in the 1950s and 1960s gradually shifted to potential therapeutic applications.

During those decades, psychiatrists and psychologists administered LSD to thousands of patients. Some studies from that period demonstrated promising findings for LSD in the treatment of alcohol use disorder, opioid use disorder, and cancer-related psychological distress. However, interpretation of study findings from this period is complicated by the fact that studies were often poorly designed, and results were sometimes misrepresented. Research into LSD's therapeutic potential at that time was also sullied by unethical studies conducted by the U.S. military, the Central Intelligence Agency, and academic investigators.

Nonmedical use of serotonergic hallucinogens became widespread among young people in the 1960s, with hallucinogen intoxication leading to severe negative reactions becoming a common problem encountered by clinicians. In a 1972 cross-sectional study of 2,411 U.S. adults, while only 4.6% of all respondents reported lifetime LSD use, that number rose to 22.0% of 18 to 21 year olds and 18.2% of 18 to 25 year olds.

Due to concerns about potential harms associated with nonmedical hallucinogen use, some supported by data and some not, such as "bad trips," psychotic sequelae, suicide, and accidental deaths, by the early 1970s, possession of most serotonergic hallucinogens in the United States and many other countries was declared illegal. Research into serotonergic hallucinogens' therapeutic potential largely ceased for the next several decades due to funding and administrative barriers, though it has resumed in recent years.

Along with serotonergic hallucinogens, arylcyclohexamines are important hallucinogens from a public health perspective. Their history begins with the synthesis of PCP in 1926 and its initial use as an anesthetic in the late 1950s. Once on the market, PCP was found to induce postoperative dysphoria and psychosis in many patients, and use in humans was discontinued in 1965. That same-year nonmedical use of PCP started, and associations between PCP intoxication with bizarre and violent behavior quickly drew attention. PCP eventually came to be considered the most serious drug problem in the United States. However, nonmedical use peaked in the late 1970s. In 1979, it was estimated that 14.5% of persons aged 18 to 25 years had used PCP. In 1988, the rate was 4.4% for those between ages 18 and 25. 12.8% of high-school seniors reported use in 1979, which decreased to about 3% from 1987 to 1995. Emergency department (ED) visits related to PCP increased fivefold between 1974 and 1979 and then subsequently declined. Today, PCP is encountered by clinicians in only a few, primarily urban, areas of the United States.

Ketamine, another arylcyclohexamine, was discovered in 1962 and approved by the U.S. Food and Drug Administration (FDA) as an anesthetic in 1970. Though less potent than PCP and having a considerably shorter duration of action, ketamine’s side-effect profile proved superior to PCP’s, though some patients still experience dissociative and psychotic effects. Ketamine became popular in the rave scene in the 1980s, particularly in locations in Asia, such as Hong Kong, where ketamine has been the drug of choice for those attending raves. Ketamine was not federally scheduled as a controlled substance in the United States until 1999 when it was added to Schedule 3. Nonmedical use of ketamine is relatively uncommon in the United States currently, though it is increasing and associated with serious harms such as ulcerative cystitis in heavy users. Medical use of ketamine remains widespread for a variety of indications, including anesthesia and treatment of acute agitation. More recently, ketamine has gained interest as a rapid treatment for depression, culminating in the approval of esketamine by the FDA for treatment-resistant depression in 2019 and major depressive disorder with acute suicidal ideation or behavior in 2020.

Table 11.6–1.
Classes of Hallucinogens

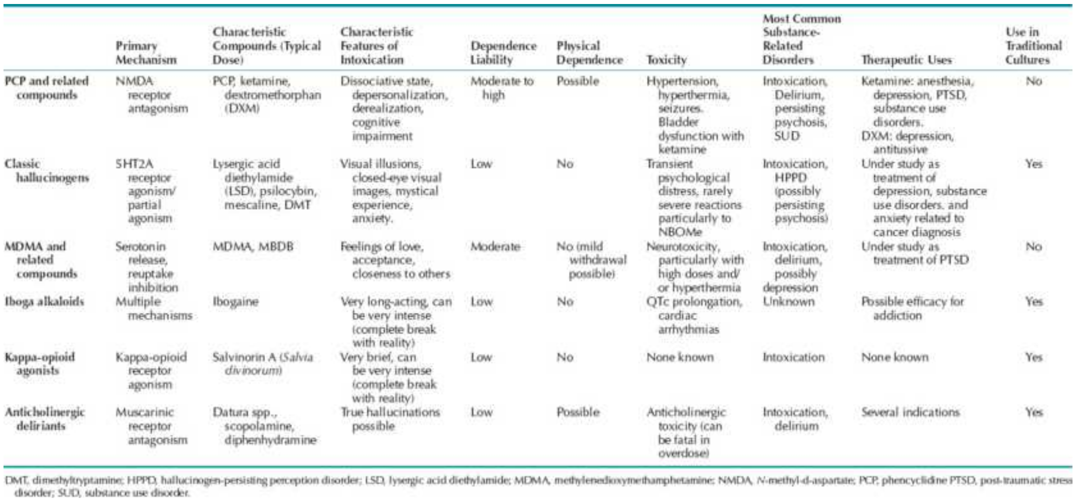

	Primary Mechanism	Characteristic Compounds (Typical Dose)	Characteristic Features of Intoxication	Dependence Liability	Physical Dependence	Toxicity	Most Common Substance-Related Disorders	Therapeutic Uses	Use in Traditional Cultures
PCP and related compounds	NMDA receptor antagonism	PCP, ketamine, dextromethorphan (DXM)	Dissociative state, depersonalization, derealization, cognitive impairment	Moderate to high	Possible	Hypertension, hyperthermia, seizures. Bladder dysfunction with ketamine	Intoxication, Delirium, persisting psychosis, SUD	Ketamine: anesthesia, depression, PTSD, substance use disorders. DXM: depression, antitussive	No
Classic hallucinogens	5HT2A receptor agonism/ partial agonism	Lysergic acid diethylamide (LSD), psilocybin, mescaline, DMT	Visual illusions, closed-eye visual images, mystical experience, anxiety.	Low	No	Transient psychological distress, rarely severe reactions particularly to NBOMe	Intoxication, HPPD (possibly persisting psychosis)	Under study as treatment of depression, substance use disorders. and anxiety related to cancer diagnosis	Yes
MDMA and related compounds	Serotonin release, reuptake inhibition	MDMA, MBDB	Feelings of love, acceptance, closeness to others	Moderate	No (mild withdrawal possible)	Neurotoxicity, particularly with high doses and/ or hyperthermia	Intoxication, delirium, possibly depression	Under study as treatment of PTSD	No
Iboga alkaloids	Multiple mechanisms	Ibogaine	Very long-acting, can be very intense (complete break with reality)	Low	No	QTc prolongation, cardiac arrhythmias	Unknown	Possible efficacy for addiction	Yes
Kappa-opioid agonists	Kappa-opioid receptor agonism	Salvinorin A (*Salvia divinorum*)	Very brief, can be very intense (complete break with reality)	Low	No	None known	Intoxication	None known	Yes
Anticholinergic deliriants	Muscarinic receptor antagonism	Datura spp., scopolamine, diphenhydramine	True hallucinations possible	Low	Possible	Anticholinergic toxicity (can be fatal in overdose)	Intoxication, delirium	Several indications	Yes

DMT, dimethyltryptamine; HPPD, hallucinogen-persisting perception disorder; LSD, lysergic acid diethylamide; MDMA, methylenedioxymethamphetamine; NMDA, *N*-methyl-d-aspartate; PCP, phencyclidine PTSD, post-traumatic stress disorder; SUD, substance use disorder.

A. 3,4-Methylenedioxymethamphetamine (MDMA) pills

B. Lysergic acid (LSD) tabs, with one tab divided into microdoses

C. *Echinopsis pachanoi* (San Pedro Cactus), a source of mescaline

D. *Banisteriopsis caapi*, a component of the hallucinogenic brew ayahuasca

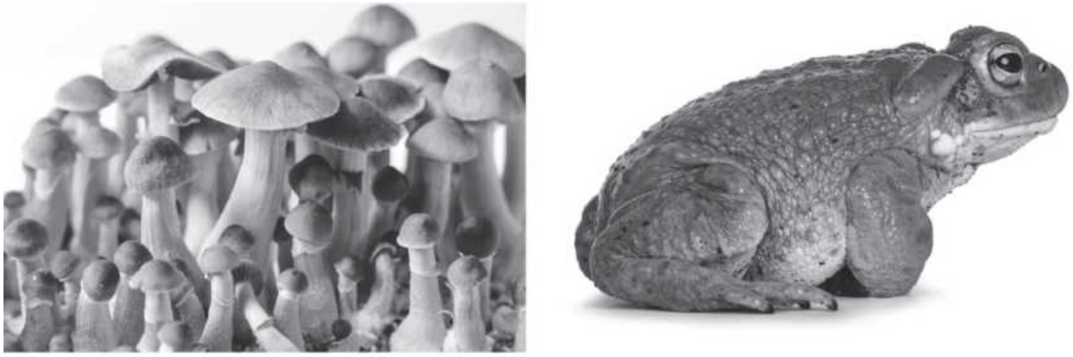

E. *Psilocybe cubensis*, psilocybin-containing mushrooms

F. *Incilius alvarius* (Sonoran Desert toad), a source of 5-MeO-DMT (5-methoxy-N,N-dimethyltryptamine)

FIGURE 11.6–1. Photographs of commonly used synthetic hallucinogens, as well as animal, fungal, and plant sources of commonly used naturally occurring hallucinogens. (Photos from iStock.com/Nynke van Holten; portokalis; eskymaks; yurok; Chris world; and Yarygin.)

Other naturally occurring hallucinogens such as the kappa-opioid receptor agonists salvinorin A have drawn attention in recent years, and a variety of novel synthetic hallucinogens such as 3-methoxyphencyclidine (3-MeO-PCP) and 25I-NBOMe are currently being disseminated by clandestine chemists via the internet, presenting new challenges for clinicians and society alike.

EPIDEMIOLOGIC OVERVIEW OF HALLUCINOGEN USE

According to the 2019 National Survey on Drug Use and Health (NSDUH), 16% of Americans aged 12 years and older have used a hallucinogen at least once, while 2.2% and 0.7% have used one in the last year and month, respectively. This makes hallucinogens the fourth most commonly used class of illicit drugs in the United States, following cannabis, opioids, and stimulants. Rates of past-year/past-month use are highest in the 18- to 25-year-old age group (7.23/1.95%) and lowest in the >25-year-old group (1.41/0.5%). Peak age for lifetime use is age 26 to 34 (23.46%), while for past-year and past-month use, the peak is age 21 to 25 at 7.8% and 2.1%, respectively. Among U.S. 12th graders in the 2020 Monitoring the Future (MTF) study, 7.5%, 5.3%, and 1.8% reported lifetime, past-year, and past-month use of hallucinogens, respectively.

While rates of hallucinogen use had been fairly stable since 2002, it appears that lifetime hallucinogen use is now rising. In the 2017 MTF study, 1.9% of 8th graders, 4.2% of 10th graders, and 6.7% of 12th graders reported lifetime hallucinogen use compared, respectively, to 3.0%, 4.8%, and 7.5% in 2020. According to the NSDUH, among people aged 12 years and older, the lifetime rate of hallucinogen use rose from 15.5% in 2017 to 16.0% in 2019. From 2015 to 2019, past-year use of LSD, the most commonly used hallucinogen, rose from 0.55% to 0.87% in the United States.

In 2019, as with most substances, more men (19.7%) had used a hallucinogen in their life than women (12.6%). Lifetime use was most common among White (19.3%), American Indian/Alaskan Native (23.4%), and individuals of more than one race (25.7%). Rates of hallucinogen use were somewhat higher in the Western United States and urban counties. Notably, since 2006, significant increases have been seen in the use of DMT, AMT, 5-MeO-DIPT, and *Salvia divinorum*. Table 11.6–2 summarizes additional information on hallucinogen use estimates from the NSDUH, as well as estimates for 12th graders from the MTF survey.

In a 2019 study, alcohol, cannabis, cocaine, stimulant, and club drug/other drug use disorders were associated with past-year hallucinogen use, while alcohol, cannabis, cocaine, heroin, nicotine, opioid, stimulant, and club drug/other drug use disorders were associated with lifetime hallucinogen use. Cocaine, nicotine, and club drug/other drug use disorders were associated with lifetime hallucinogen use disorder. Regarding psychiatric disorders, having any personality disorder was associated with past-year and lifetime hallucinogen use, while personality disorder and post-traumatic stress disorder (PTSD) were associated with lifetime hallucinogen use disorder.

Regarding potential mental health sequelae from hallucinogen use, population-level studies have identified no associations between lifetime serotonergic hallucinogen use and negative mental health outcomes during the previous year, including depression, anxiety, psychosis, suicidal thoughts, suicidal plans, and suicide attempts. Nevertheless, these studies do not imply that individuals using these substances may not be harmed; therefore, caution is warranted in interpreting their findings.

Given increasing media attention on positive findings from clinical trials of hallucinogen-assisted therapies for mental health conditions and substance use disorders (SUDs), clinicians could face new challenges in coming years if patients seek to self-treat with hallucinogens. Though its prevalence is unknown, hallucinogenic microdosing, the act of taking serotonergic hallucinogens at doses that do not alter consciousness in the hopes of improving emotional well-being or enhancing cognition and creativity, has gained increasing media attention in recent years. However, recent placebo-controlled trials of microdosing, though few in number, have failed to identify benefits for patients, and a 2021 prospective study of individuals microdosing in the community found that positive expectations predicted improved mental health outcomes, suggesting a strong role for expectancy in perceived benefit.

HALLUCINOGEN SUBCLASSES

NMDA Receptor Antagonists

The DSM-5-TR recognizes a class of PCP-type substances explicitly including PCP, ketamine, cyclohexamine, and dizocilpine. The primary effects of these drugs are mediated by noncompetitive *N*-methyl-D-aspartate (NMDA) receptor antagonism. Other drugs sharing this mechanism of action include dextromethorphan (DXM) and nitrous oxide (the latter usually categorized as an inhalant because it is a gas). PCP and ketamine bind to PCP binding sites on the NMDA receptor and block the ionophore channel, decreasing influx of calcium and sodium ions. The highest densities of PCP receptors are found in the dentate gyrus, hippocampus, and anterior forebrain. Downstream effects of inhibition of NMDA receptor–mediated glutamatergic neurotransmission include a secondary increase in glutamate release in the medial prefrontal cortex and an increase in pyramidal cell firing, probably due to blockade of NMDA receptors on γ-Aminobutyric acid (GABA)-ergic inhibitory interneurons. In rats, noncompetitive NMDA receptor antagonists also increase extracellular prefrontal and mesolimbic dopamine levels, probably by stimulating glutamate release in the ventral tegmental area and dorsal raphe nuclei. The effects on dopamine release are thought to mediate the reinforcing effects of these drugs, as well as hyperactivity and stereotypic movements. At very high doses, PCP blocks monoamine transporters, but ketamine does not—possibly accounting for greater stimulation and reward with PCP than ketamine. At even higher doses, PCP blocks sodium and potassium channels, which may contribute to the incidence of seizures following PCP overdose. In animal models, behavioral effects of NMDA antagonists are attenuated by enhancers of NMDA receptor–mediated neurotransmission, such as glycine and D-serine. Some effects are also attenuated by D2 and 5HT2A receptor-blocking agents.

Effects of PCP and Related Drugs. Sometimes termed *dissociatives* or *dissociative anesthetics*, a prominent feature of intoxication with these compounds is a sense of separation of mind from body, accompanied by anesthesia with preserved airway reflexes. These drugs can also produce hallucinogenic effects, including visual illusions or true hallucinations, other sensory changes, altered thought processes, labile or inappropriate affect, anxiety, and paranoia or other delusional ideation. PCP intoxication, in particular, can be associated with unpredictable and potentially violent behavior. It is also associated with nystagmus, ataxia, hypertension, tachycardia, and muscular rigidity.

Table 11.6–2. Data on Prevalence of Hallucinogen Use and Hallucinogen Use Disorder

		NSDUH 2019,[a] Age 12 and Up					MTF 2020,[b] 12th Grade		
Substance	Demographic	Lifetime Use	Past-Year Use	Past-Month Use	Past-Year Abuse	Past-Year Dependence	Lifetime Use	Past-Year Use	Past-Month Use
Any hallucinogen		16.0	2.2	0.7	0.08	0.05	7.5	5.3	1.8
	Male	19.7	2.87	0.92	0.11	0.07			
	Female	12.6	1.49	0.47	0.05	0.04			
	12–17 years	2.5	1.81	0.62	0.13	0.1			
	18–25 years	16.6	7.23	1.95	0.27	0.2			
	>25 years	17.5	1.41	0.5	0.04	0.02			
	Hispanic/Latino	10.8	2.1	0.5					
	Not Hispanic/Latino	17.1	2.2	0.7					
	White	19.3	2.3	0.8					
	Black/African American	7.3	1.4	0.5					
	American Indian/ Alaskan Native	23.4	3.9	1.9					
	Native Hawaiian and Other Pacific Islanders	11.5	2.1	0					
	Asian	6.5	1.6	0.8					
	Two or more races	25.7	4.5	1.3					
LSD		10.0	0.9	0.2			5.9	3.9	1.4
	Male	13.0	1.24	0.3					
	Female	7.3	0.53	0.13					
MDMA		7.3	0.9	0.2			3.6	1.8	0.8
	Male	8.4							
	Female	6.1							
PCP		2.0	0.03	0					
	Male	2.7	0.04	0.1					
	Female	1.3	0.01	0					
Mescaline		2.8							
	Male	4.0							
	Female	1.6							
Peyote		2.1							
	Male	3.1							
	Female	1.1							
Psilocybin		9.1							
	Male	12.2							
	Female	6.1							
Cough or cold medicine		4.1	0.76	0.29				3.2	
	Male	5.0	0.86	0.38					
	Female	3.2	0.67	0.21					
Ketamine		1.4	0.18	0.08				1.3	
	Male	1.8	0.26	0.12					
	Female	1.0	0.11	0.04					
N,N-Dimethyltryptamine (DMT), α-Methyltryptamine (AMT), 5-Methoxy-N, N-diisopropyltryptamine (5-MeO-DIPT)		1.2	0.19	0.06					
	Male	1.8	0.26	0.08					
	Female	0.7	0.13	0.04					
Salvia divinorum		1.8	0.05	0.01				0.7	
	Male	2.8	0.07	0.01					
	Female	0.9	0.03	0.01					

LSD, lysergic acid diethylamide; MDMA, methylenedioxymethamphetamine; PCP, phencyclidine.

[a] Center for Behavioral Health Statistics and Quality. (2020). Results from the 2019 National Survey on Drug Use and Health: Detailed tables. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/ on 09/19/2021.

[b] Johnston LD, Miech RA, O'Malley PM, Bachman JG, Schulenberg JE, Patrick ME. (2021). Monitoring the Future national survey results on drug use 1975–2020: overview, key findings on adolescent drug use. Ann Arbor: Institute for Social Research, University of Michigan. Retrieved from http://monitoringthefuture.org/data/data.html on 09/19/2021.

The duration of acute effects depends on the specific drug, dose, and route of administration. Serum half-life of PCP in humans is highly variable, averaging about 20 hours. Ketamine has a considerably shorter half-life, about 2 hours. PCP is highly lipid soluble and accumulates in fat. Metabolism is primarily hepatic for both drugs.

PCP and ketamine produce compulsive use in animal models. Monkeys given free access to PCP will self-administer to maintain near-constant intoxication. Patterns of compulsive use of PCP and ketamine are seen in humans as well.

PCP and related drugs also lower the threshold for brain self-stimulation reward, a characteristic shared with most other addictive drugs. Pronounced tolerance has been demonstrated with PCP in animals and ketamine in humans. A withdrawal syndrome is seen in primates, including vocalizations, bruxism, oculomotor hyperactivity, diarrhea, piloerection, somnolence, tremor, and seizures.

NMDA Receptor Antagonist Model of Psychosis. The use of PCP intoxication as a model for psychosis was proposed in 1959. Although PCP is no longer administered to humans due to its toxicity, ketamine is still employed for investigating the role of altered glutamatergic neurotransmission in the pathogenesis of psychosis. The NMDA antagonist model of psychosis more closely mimics symptoms of schizophrenia than other drug models (e.g., amphetamine or classic hallucinogen). Both PCP and ketamine induce symptoms of thought disorder, including withdrawal, autism, negativism, inability to maintain cognitive set, and abnormal responses to projective testing. Catatonia may also occur. Disturbances in attention, perception, and symbolic thinking are also very similar to those seen in schizophrenia. Both drugs inhibit prepulse inhibition, replicating a well-established neurobiologic marker of schizophrenia. Well-formed hallucinations and delusions are less common and may be absent even in the presence of severe thought disorder. A single relatively low dose of PCP can trigger a full-blown psychotic episode in patients with schizophrenia.

Serotonergic Hallucinogens

There are several distinct and overlapping groups of naturally occurring and synthetic hallucinogens that act primarily by affecting serotonergic pathways but vary greatly in structure, effects, and mechanisms of action. The “other hallucinogen” class defined by DSM-5-TR primarily includes these drugs, though it also contains a variety of other types of hallucinogens.

Classic Hallucinogens. The classic hallucinogens are those drugs whose effects are primarily mediated by agonism or partial agonism at the 5HT2A receptor. The prototypical drug of this class is LSD. Other representatives of this large group of drugs are psilocin (the primary active ingredient in hallucinogenic mushrooms) and its prodrug psilocybin, mescaline (the primary active ingredient of hallucinogenic cacti such as peyote and the San Pedro cactus), and DMT. Structurally, these drugs fall into three main categories. The tryptamine hallucinogens, which include psilocybin/psilocin and DMT, are structurally related to serotonin (5-hydroxytryptamine), containing the indole nucleus and various substituents affecting receptor affinities and drug metabolism. The ergoline class, which includes LSD, consists of the indole structure within a much more complicated four-ring molecule. LSD was originally derived from ergot fungus that infects grain and causes ergotism or St. Anthony's fire, outbreaks of which frequently occurred in Europe in the Middle Ages. Other ergolines are found in members of the morning glory family, including *Rivea corymbosa*, *Ipomoea violacea*, and *Argyreia nervosa*. The third group of classic hallucinogens, which includes mescaline and 2,5-Dimethoxy-4-methylamphetamine (DOM), is the phenylalkylamines. While the phenethylamine structure is the basis for amphetamines, phenylalkylamines can have effects similar to LSD, making them likely to be mediated primarily via 5HT2A receptor agonism.

The classic hallucinogens have long provided evidence implicating the serotonergic system in the pathophysiology of psychotic disorders. In the early 1920s, similarities were first described between mescaline intoxication and symptoms of schizophrenia. When Albert Hoffman accidentally discovered the psychoactive effects of LSD in 1943, he described frightening psychotic-like symptoms, including thoughts that he was possessed by a demon and that his neighbor was a witch. This prompted investigation into the use of LSD and other classical hallucinogens as models for understanding the biologic basis of psychosis. In fact, when LSD was first marketed by Sandoz in the 1940s, psychiatrists and medical trainees were encouraged to take LSD in order to gain insight into the schizophrenic experience. Some researchers speculated that schizophrenia might be caused by excess production of an endogenous psychotogenic compound. DMT drew particular attention, with studies showing it produced psychotic-like positive symptoms in normal volunteers and other studies reporting differences in urine and plasma DMT levels between patients with schizophrenia and controls. Although these findings were not consistently replicated, it has been shown that DMT is produced in the pineal gland of rats.

CLASSIC HALLUCINOGEN EFFECTS. The effects of all classic hallucinogens are similar, differing principally in time course, which in turn depends on the particular substance, dose, and route of administration. An important feature of all classic hallucinogens is that the psychological content and emotional tone of the experience are unpredictable but are thought to be strongly influenced not only by the dose and particular drug taken but also by the mental state, psychological preparation, and intention of the person taking the drug (mental "set"), and the environment in which the effects are experienced ("setting"), including the broader cultural context. For example, one early LSD study found that the severity of adverse psychological reactions varied based on how study staff interacted with participants, with less severe adverse reactions occurring when staff were instructed to be warmer and the study setting was more relaxed. Support for the role of culture's impact on the classic hallucinogen experience comes from observational surveys indicating that the frequency of "bad trips" associated with classic hallucinogens decreased from the mid-1960s to 1975 as conflict over these compounds became less prominent within societal discourse.

Although classic hallucinogens are safe in carefully screened and prepared study participants, in controlled doses, and under ideal conditions, all have some inherent risks. These risks are greatly magnified in uncontrolled settings. People at high risk of adverse outcomes (e.g., people with psychotic disorders or a family history of psychosis) may choose to use these drugs, and preparation and support may be lacking. In addition, under typical conditions of uncontrolled use, the identity and dose of the substance are rarely known with certainty, so overdose or unexpected effects can easily happen. Risks are multiplied when new substances are introduced to the illicit market due to ignorance as to the potency and effects of the compounds as well as to the inherent risks of the drugs.

Somatic effects of classic hallucinogens are typically mild and can include anorexia, chills, tremor, unsteadiness, nausea, vomiting, xerostomia, paresthesias, mydriasis, insomnia, and blurred vision. Pulse and blood pressure may become mildly to moderately elevated. Significant physiologic toxicity is not seen in typically used dosages. Sensory effects are often prominent, including alteration in perception of shape, size, and color and the illusion of movement. With eyes closed, vivid imagery is often perceived, ranging from elementary geometric or fractal patterns to vivid representational images of various kinds. Other sense modalities may be involved, and synesthesia can occur. True hallucinations are uncommon. Affective alterations are extremely variable and can change rapidly and frequently during a single episode of use. Anxiety is common, as are strong cathartic emotional experiences, often related to past or current life experiences. Feelings of bliss, joy, and peace may also occur. Depersonalization, derealization, and paranoia can occur. Distortions in perception of time can occur, most commonly resulting in time seeming to pass slowly. The content of the subjective experience may be dominated by personal psychological concerns (conflicts, relationships, grief, and loss), symbolic representations of archetypal themes (heroic quests, interactions with mythologic figures), or spiritual matters. Particularly in high doses, mystical-like experiences may occur, typically including ineffable feelings of unity and transcendence of space and time, which are experienced as deeply positive, sacred, and meaningful.

An important exception regarding the effects of classic hallucinogens is that of the NBOMe class of hallucinogens, often sold as LSD due to their high potency and similar subjective and physiologic effects. NBOMes, the N-benzyl-methoxy analogs to the 2C-x family of phenethylamine hallucinogens, were initially developed as highly specific 5HT2A receptor agonists for neuroimaging research. Unlike LSD, however, NBOMes have been implicated in numerous hospitalizations for severe complications, including seizures and severe toxidrome requiring prolonged ICU stays, as well as fatalities.

Tolerance to LSD develops very rapidly, with significant decreases in the effect of subsequent doses after a single dose. This effect may be due to downregulation or internalization of 5HT2A receptors and is completely reversed within a week of abstinence. DMT may be an exception to this effect, as marked tolerance is not observed. Cross-tolerance exists among LSD, mescaline, and psilocybin, further evidence of a shared mechanism of action.

DIFFERENCES AMONG CLASSIC HALLUCINOGENS. Classic hallucinogens differ principally in potency, duration of action, and traditional patterns of use. LSD is one of the most potent drugs known, with noticeable effects beginning at approximately 25 μg. Because of its high potency, LSD is often distributed on small squares of blotter paper, which are held in the mouth to allow buccal absorption prior to swallowing. Effect onset usually occurs within 60 minutes, with peak effects around 3 to 4 hours and total duration of action of 8 to 10 hours. A typical street “hit” (dose) of LSD is currently approximately 50 μg, much lower than the doses of 100 to 250 μg that were typical in the 1960s. This may partially account for the lower incidence of LSD-associated adverse reactions in recent years.

Mescaline is among the least potent hallucinogens, with a dose of 300 to 400 mg estimated to be roughly equivalent to 100 μg of LSD. Mescaline’s onset of action is somewhat slower than LSD’s, and the duration of action is slightly longer (10 to 12 hours). Mescaline is primarily used in the form of peyote or other mescaline-containing cacti, though purified or synthetic mescaline is sometimes found. Humans have used peyote for at least 5,000 years, and it is currently used sacramentally within the Native American Church and by tribes in northern Mexico, including the Huichol. The San Pedro cactus of South America has a long history of ritualistic use as well. Mescaline, at least when consumed as cactus, is associated with more nausea and vomiting than LSD or psilocybin. Its low potency also results in a lower index of safety than other classic hallucinogens. Autonomic hyperstimulation can become problematic in doses above 500 mg.

Psilocybin is intermediate in potency between LSD and mescaline, with a dose of 10 to 15 mg being roughly equivalent to 100 μg of LSD. Psilocybin’s effects are shorter (4 to 6 hours) than those of LSD and mescaline. It is almost always consumed in the form of mushrooms, and mild-to-moderate headache is common for up to 24 hours following psilocybin use.

When taken orally by itself, DMT is not psychoactive due to rapid metabolism by monoamine oxidase A (MAO-A). Therefore, it is normally smoked, or occasionally injected or insufflated. In smoked doses of around 25 mg, DMT produces a brief (15 minutes) and extremely intense experience that can involve complete loss of contact with reality, intense visuals, loss of ordinary sense of reality and self, mystical experience, and/or the experience of contact with other worlds or beings. Marked tolerance does not seem to develop to DMT, but frequent use is unusual in any case because of the intense and sometimes overwhelming nature of the experience. Ingestion of an MAO-A inhibitor, either before or concurrently with DMT, can render the compound orally psychoactive. Ayahuasca, an herbal preparation traditionally used in the western Amazon basin, typically contains both monoamine oxidase inhibitors (MAOIs) and DMT. In the typical brew, the MAOIs are harmala alkaloids (harmine, harmaline, and tetrahydroharmine) that come from the *Banisteriopsis caapi* vine, and the DMT comes from the leaves of *Psychotria viridis*. The brew has effects similar to those of classic hallucinogens, although it should be noted that the harmala alkaloids are probably psychoactive to some degree, and they, not the DMT carrier, are traditionally considered the defining and essential ingredient to ayahuasca. Ayahuasca has been used in shamanic traditions in South America for hundreds of years and is used as a sacrament in several syncretic Brazilian religions that developed in the 20th century. More recently, it has been used with increasing frequency in North America and Europe for therapeutic purposes or spiritual exploration. The effects of ayahuasca typically last about 4 hours and are generally similar to those of classic hallucinogens. Nausea and vomiting are common, probably due to the harmala alkaloids, and purging is believed by some to be a valuable aspect of the experience.

MDMA and Related Compounds (Entactogens). A large number of phenylalkylamines, although structurally similar to classic hallucinogens such as mescaline and DOB, have effects quite distinct from those of classic hallucinogens. MDMA, 3,4-methylenedioxy-*N*-ethylamphetamine (MDEA), and numerous other compounds have subjective effects characterized by feelings of closeness and loving acceptance toward others, with less pronounced alterations in sense perception or thought process common to the classic hallucinogen experience. To varying degrees, these drugs have stimulant effects as well. Because of the effects on feelings toward others, they have been termed “entactogens” or “empathogens.” MDMA has attracted interest as a potential adjunct to psychotherapy because it appears to increase tolerance of negative affect and painful memories (e.g., in PTSDs), possibly allowing them to be dealt with more directly and effectively. Unlike the classic hallucinogens, these drugs are moderately reinforcing and therefore have greater potential for compulsive use. MDMA is taken up presynaptically by the serotonin transporter and induces release of serotonin from presynaptic stores. It also acts as a reuptake inhibitor and, particularly at high doses, directly stimulates 5HT2A and other serotonin receptors. Overdose can occur easily because MDMA inhibits its own metabolism by CYP 2D6 and CYP 3A4, resulting in blood levels that increase ever more steeply as the dose increases. Given that illicit MDMA is often sold with contaminants or adulterants or contains other substances with no MDMA at all, overdoses or other complications can inadvertently occur from drug–drug interactions such as serotonin syndrome, hyperthermia, or seizures.

ENTACTOGEN EFFECTS. Somatic effects of MDMA include moderate increases in blood pressure and pulse and an increase in body temperature. Although the latter averages less than ½ degree Fahrenheit in clinical research settings, it can be much greater when the drug is used in all-night dance parties (raves), particularly if the user is dehydrated. Trismus and bruxism are common side effects. In high doses, muscle stiffness and rhabdomyolysis can occur. In overdose, seizures and cardiorespiratory collapse can occur. Psychological effects include a heightened sense of closeness to others and a deep sense of well-being or acceptance. The emotional tone tends to be very positive and maybe euphoric at times. However, strong releases of negative emotion can also occur, as well as anxiety and paranoia, although much less commonly than with classic hallucinogens.

Other Serotonergic Hallucinogens. Hundreds of serotonergic hallucinogens have been characterized, and new ones are synthesized continually. Many of these “research chemicals” find their way to the streets, not always accurately labeled or dosed. 2C-B and related compounds (e.g., 2C-E, 2C-I) appear, based on self-reports of users, to have effects that combine aspects of classic hallucinogen and entactogen effects. Other drugs are reported to have unique effects that have not been adequately characterized. Recently, a number of highly potent hallucinogens with novel structures have begun to be marketed that have strong LSD-like effects but considerably greater toxicity. For example, Bromo-DragonFLY, an agonist at 5HT2A, 2B, and 2C receptors, is active in doses under 1 mg, with effects lasting for several days. As indicated earlier, NBOMes are highly potent full agonists at the 5HT2A receptor (LSD is a partial agonist), which are also psychoactive in doses under 1 mg. These drugs appear to be much more toxic than classic hallucinogens and have caused a number of fatalities, often also producing seizures.

Iboga Alkaloids

Ibogaine is a hallucinogenic alkaloid found in several plants, most notably in the root of the *Tabernanthe iboga*. It is used for medicinal and spiritual purposes by the indigenous people of Western Africa and has hallucinogenic effects in doses of 400 to 1,500 mg, producing an intense and often unpleasant experience that can last 4 to 6 hours, with lingering effects felt for up to 2 to 3 days. Although ibogaine acts at 5HT2 and 5HT3 receptors, it also inhibits the serotonin transporter and acts at NMDA, mu-, kappa-, and sigma-opioid, and muscarinic and nicotinic cholinergic receptors. Ibogaine attracted the interest of addiction researchers when reports emerged of patients with opioid use disorder losing the desire to use heroin after taking ibogaine. Although initial preclinical and clinical results in the treatment of opioid withdrawal were encouraging, and preclinical work suggested that ibogaine may have therapeutic effects across several classes of drugs, research on ibogaine has been plagued by toxicity concerns, particularly QTc prolongation and sudden death due to cardiac arrhythmias. Currently, there is interest in developing less toxic–related molecules, such as 18-methoxy-coronaridine (18-MC). Although ibogaine is unlikely to be encountered as a substance of misuse, it continues to be used in some treatment centers outside the United States and within underground treatment settings in the United States.

Kappa-Opioid Agonists

Salvia divinorum is a species of sage plant, the leaves of which are used traditionally by the Mazatecs in Oaxaca, Mexico. When purified, the main active component, Salvinorin A, is an extremely potent hallucinogen, active in doses as small as 4.5 μg/kg when smoked. Effects are intense but brief, with a time course comparable to smoked DMT, peaking within 2 minutes and dissipating within about 15 minutes. Effects not only overlap with those of classic hallucinogens but also include dissociative effects and impaired recall of the experience. Users report that the effects are distinct from those of other drug classes, producing intense experiences that may range from pleasant to highly distressing. Injuries from falls may occur due to incoordination, and the experiences can be psychologically unsettling or even traumatic. United States Drug Enforcement Administration (DEA). Many analogs have now been synthesized. The most likely dangers from Salvinorin A appear to be the possibilities of psychologically damaging experiences and unintentional self-harm during the intoxication state. Users of high doses have been reported to become agitated and perform violent, purposeless actions, for example, running into walls.

Anticholinergic Deliriants

Anticholinergic alkaloids from plants of the nightshade family have been used in shamanic practice and witchcraft for millennia. Although one could argue that the hallucinatory states produced by these drugs should be classified as anticholinergic delirium, the DSM-5-TR explicitly lists jimsonweed as an ethnobotanical hallucinogen. There is some logic to this since plants like jimsonweed, henbane, and belladonna have been used specifically for their hallucinogenic properties. These solanaceous plants contain antimuscarinic alkaloids, principally atropine and scopolamine. Synthetic anticholinergics such as benztropine and diphenhydramine (an H1 inverse agonist with strong antimuscarinic effects) can produce similar effects in high doses. Anticholinergic effects include dry mouth, mydriasis, blurred vision, increased heart rate, decreased sweating, hyperthermia, seizures, and, occasionally, death. These drugs can produce true hallucinations and other features of delirium, including disorientation, agitation, grossly impaired cognition and memory, and fluctuating mental status.

Differences between and Similarities to Other Substances of Misuse

Despite the great diversity among the classes of hallucinogenic drugs, there is significant overlap in the effects of most of these drugs and differences from other major classes of psychoactive substances. While the DSM-5-TR states that “all drugs that are taken in excess have in common the ability to directly activate the brain reward systems,” this is not true of classic hallucinogens, ibogaine, or kappa-agonists. Their effects are neither reliably reinforcing in animal models nor reliably pleasurable to humans. Thus, it is generally accepted that these drugs do not produce compulsive use or true addiction, even though they can certainly be the object of a SUD. In general, people use these drugs because they value the effects for reasons other than primary reward, for example, because they find the experience meaningful, informative, interesting, or transformative. Some clinical trials have also indicated that classical hallucinogens may actually be helpful in treating SUDs.

The nonaddictive nature of classic hallucinogens and their possible SUD treatment potential has received only limited study. One potential explanation for these attributes centers around classic hallucinogens’ 5-HT2C receptor agonism. In preclinical models, this type of agonism has been shown to oppose the reinforcing potential of several types of addictive compounds, likely via modulation of dopaminergic activity in the ventral tegmental—nucleus accumbens (NAc) reward pathway.

On the other hand, NMDA antagonists, such as ketamine and serotonergic hallucinogens with entactogen and/or stimulant properties, such as MDMA, do activate the primary reward pathways. The resulting experience is therefore reinforcing to many users. Thus, the usual addictive processes can also motivate continued use of these substances.

COMPARATIVE NOSOLOGY

Table 11.6–3 summarizes the relationships among hallucinogen-related disorders in DSM-5-TR and ICD-11. While DSM-5-TR includes PCP and related compounds as a category of hallucinogens, ICD-11 includes PCP and related substances under “dissociative drugs.” Neither classification system recognizes a hallucinogen withdrawal syndrome. Hallucinogen intoxication symptoms listed in ICD-11 are similar to those in DSM-5-TR, but not exactly the same: in ICD-11, specific behavioral, perceptual, and physical changes are listed (corresponding to criteria B, C, and D in DSM-5-TR). Finally, while hallucinogen-persisting perception disorder (HPPD) is listed in DSM-5-TR, ICD-11 codes HPPD under “other specified disorders due to use of hallucinogens.”

PHENCYCLIDINE AND OTHER HALLUCINOGEN INTOXICATION

Definition

The DSM-5-TR criteria for PCP intoxication and other hallucinogen intoxication are similar but differ in some specifics. The diagnosis of intoxication for both classes of drugs requires recent use of a relevant substance, clinically significant psychological or behavioral disturbance, and the lack of a better explanation for the symptoms (e.g., another mental disorder). Although there is overlap in the symptoms of intoxication for PCP and other hallucinogens, as defined in DSM-5-TR, there are significant differences. Criterion B of PCP intoxication requires disruptive behavior changes, which are more frequent and prominent with PCP use than with other types of hallucinogens. The criteria for other hallucinogen intoxication mention behavioral disturbances in criterion B, but the criteria emphasize problematic psychological changes (criterion B) and perceptual changes (criterion C). Both diagnoses require two or more specific signs and symptoms (criterion C of PCP intoxication and criterion D of other hallucinogen intoxication). These signs and symptoms have no overlap. Signs of PCP intoxication include horizontal, vertical, or rotatory nystagmus, hypertension or tachycardia, numbness or diminished responsiveness to pain, ataxia, dysarthria, muscle rigidity, seizures or coma, and hyperacusis. Rotary nystagmus is a highly specific sign of NMDA antagonist intoxication. For other hallucinogen intoxication, the signs and symptoms include pupillary dilation, tachycardia, sweating, palpitations, blurring of vision, tremors, and incoordination. These signs and symptoms are characteristic of classic hallucinogens and MDMA but may not be seen with all types of “other” hallucinogens.

Table 11.6–3.
DSM-5-TR Hallucinogen-Related Disorders and ICD-11 Billing Codes

DSM-5-TR Disorder	ICD-11
Phencyclidine (PCP) use disorder	6C4D.1
Mild	6C4D.2
Moderate	6C4D.1
Severe	
Other hallucinogen use disorder	6C49.1
Mild	6C49.2
Moderate	6C49.2
Severe	
Phencyclidine intoxication	6C4D.3
Other hallucinogen intoxication	6C49.3
Hallucinogen-persisting perception disorder	6C49.Y
Other phencyclidine-induced disorders[a]	6C4D.5
Phencyclidine-induced psychotic disorder	6C4D.6
Phencyclidine-induced bipolar disorder	6C4D.6
Phencyclidine-induced depressive disorder	6C4D.6
Phencyclidine-induced anxiety disorder	6C4D.4
Phencyclidine-induced intoxication delirium	
Other hallucinogen-induced disorders[a]	6C49.5
Other hallucinogen-induced psychotic disorder	6C49.6
Other hallucinogen-induced bipolar disorder	6C49.6
Other hallucinogen-induced depressive disorder	6C49.6
Other hallucinogen-induced anxiety disorder	6C49.4
Other hallucinogen-induced intoxication delirium	
Unspecified phencyclidine-related disorder	6C4D.Z
Unspecified hallucinogen-related disorder	6C49.Z

[a] Onset is specified as during intoxication, not during withdrawal. These disorders are diagnosed instead of intoxication only when the symptoms are sufficiently severe to warrant independent attention.

Epidemiology

Classic hallucinogen intoxication is now encountered much less frequently by clinicians than in the 1960s and 1970s. In addition to decreases in overall frequency of use since then, social learning and changes in patterns of use likely partially account for decreased adverse effects of intoxication. The typical dose of illicit LSD currently taken now (25 to 50 μg) is smaller than the 100 to 250 μg doses more commonly used in the 1970s. In addition, today, many users have much more information about the possible effects of various doses of hallucinogens and the importance of set and setting in minimizing negative effects.

There are no data providing accurate estimates of the incidence of hallucinogen intoxication. The rate of DSM-5-TR hallucinogen intoxication is necessarily lower than the rate of use because the former requires significant adverse behavioral or psychological consequences. Data from the Drug Abuse Warning Network (DAWN) provide estimates of drug-related ED visits that provide a relevant measure of cases of drug use that come to medical attention, although the ED visits are not necessarily directly related to the use of the drug. Table 11.6–4 shows the frequency of drug-related ED visits for several classes of hallucinogens based on 2011 data. PCP is the hallucinogen most commonly associated with ED visits, with 24.2 ED visits per 100,000 population. MDMA is associated with 7.2 visits per 100,000 population, and all other hallucinogens are much lower. These numbers are low relative to other substances: the corresponding numbers for other drug classes are stimulants 51.3, marijuana 146.2, heroin 82.7, cocaine 162.1, and alcohol 112.0. Rates are substantially higher for males than for females for each class of hallucinogen. Interestingly, rates are higher in the under-21 age group for the serotonergic hallucinogens but are higher in the over-20 age group for PCP and ketamine, suggesting that perhaps harmful patterns of use are more persistent for the NMDA antagonist drugs than for the serotonergic drugs, consistent with their greater dependence liability and the greater intrinsic danger of their acute effects. In 2004, DXM was associated with 5.7 ED visits per 100,000 population, with higher rates in lower age groups (14.7/100,000 among those aged 12 to 20 years). New data from Project Dawn have not been published since 2011. However, a preliminary report using April 2019 to October 2020 data indicates that of 1,275,925 ED visits in 49 participating hospitals, approximately 400 visits (0.03%) involved LSD, approximately 300 involved PCP (0.02%), and approximately 200 involved other hallucinogens (0.01%). For comparison, approximately 17,800 visits (1.40%) involved stimulants, and approximately 11,200 involved marijuana (0.88%).

Table 11.6–4.
DAWN Data 2011: ED Visits per 100,000 Population

	Total	Male	Female	Under 21	Over 20
PCP	24.2	33.9	14.9	6.5	33.1
Ketamine	0.5	0.7	—	—	0.4
MDMA	7.2	10.2	4.3	11.6	5.5
LSD	1.5	2.6	0.5	2.7	1.1
Other hallucinogen (including mushrooms)	2.6	4.4	0.8	5.3	1.5

Substance Abuse and Mental Health Services Administration, *Drug Abuse Warning Network, 2011: National Estimates of Drug-Related Emergency Department Visits.* HHS Publication No. (SMA) 13–4760, DAWN Series D-39. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2013.

Etiology

Diagnosable hallucinogen intoxication is due to an interaction between the biologic effects of the drug and individual and environmental risk factors. As noted above, risks vary from drug to drug and are greater for higher drug doses, for biologically vulnerable and/or unprepared individuals, and with use in stressful environments lacking support.

Diagnosis and Clinical Features

Hallucinogen intoxication presents as acutely altered mental status and behavioral changes following drug ingestion. Patients seeking clinical care are often agitated and fearful and may worry they are “losing their minds” or experiencing panic or paranoia. Occasionally, users become acutely depressed, which can lead to suicidal behavior during or soon after intoxication. In other cases, patients may present in an expansive, grandiose, or euphoric state resembling mania. Adverse reactions are more common in inexperienced users, with unexpectedly large doses, or under stressful circumstances, but can occur at any time. Recent drug ingestion is critical to diagnosis, but confirmation of this may be difficult to ascertain. It is important to carefully assess for a history of other drug ingestion or exposure and psychiatric and somatic illness and to perform an appropriate physical examination, including neurologic evaluation. All classes of hallucinogens can produce acute anxiety, agitation, inappropriate and labile affect, visual illusions and hallucinations, and paranoia or other delusions. Toxicology testing is recommended to confirm or rule out ingestion of commonly used substances, and if available, testing for the broad range of hallucinogens should be conducted to aid in the diagnostic clarification. While identifying the specific substance ingested may not change the clinical management for most patients, some hallucinogens can cause a prolonged toxidrome lasting days or more, in which case analytic confirmation of the hallucinogen can aid in the management as well as patient education following recovery.

PCP and Related Substances. PCP and ketamine may be taken by any route, but PCP is more commonly smoked, while ketamine is frequently injected, snorted, or orally ingested. A typical street dose is about 5 mg for PCP and 100 to 200 mg for ketamine. PCP is manufactured only illicitly, while ketamine is derived both from diverted licit material and illicit manufacture. DXM is usually taken in the form of over-the-counter cough syrup or tablets. These drugs produce a dissociative state with a greater degree of cognitive and neurologic impairment (including the signs listed in the criteria for PCP intoxication) than that produced by classic hallucinogens. Symptoms become more severe with higher doses and may warrant the diagnosis of PCP or ketamine-induced delirium if disturbances of attention and cognition are prominent. With PCP intoxication, typical features include impaired cognition, nystagmus, hypertension, tachycardia, dysarthria, hyperacusis, and increased deep tendon reflexes. Thought disorder is prominent, but hallucinations may also be present. Behavioral symptoms include disinhibition, agitation, panic, rage, aggression, and exaggerated reactions to environmental stimuli.

Medical complications are much more common than with classic hallucinogens. Hyperexcitability, hypertension, or hyperthermia may indicate significant toxicity constituting a medical emergency. Hypertensive crisis can lead to cerebral hemorrhage. Severe hyperthermia can occur. At higher doses of PCP, symptoms of dissociative anesthesia emerge, with numbness and diminished responsiveness to pain. At still higher doses, other receptor effects become prominent, as PCP interacts with monoamine uptake pumps, sigma receptors, sodium and potassium channels, nicotinic and muscarinic acetylcholine receptors, and mu-opioid receptors.

Unpredictable, agitated, and violent behavior occurs more commonly with PCP than with other hallucinogens. Formal thought disorder, delusions, and frank hallucinations can occur, as well as confusion, impairment of motor control, impaired speech, and catatonic-like symptoms. Overdoses of PCP and related compounds can have serious medical sequelae, including seizures, hyperthermia, hypertensive crisis, cerebral hemorrhage, respiratory arrest, coma, and death.

Ketamine intoxication is generally less malignant than PCP intoxication. Medical complications are less common due to the lower potency and shorter half-life of ketamine and because ketamine has fewer significant actions at sites other than the NMDA receptor. Presenting complaints include anxiety, chest pain, and palpitations. Tachycardia and hypertension are common but usually much less severe than when seen with PCP intoxication. Rotary nystagmus is possible but uncommon. Death from ketamine intoxication is very rare, except when ketamine is used with other drugs.

PCP or ketamine intoxication is often a sign of moderate or severe SUD, so that should be diagnosed and treated if present.

Classic Hallucinogens. Although patients with classic hallucinogen intoxication can show a wide range of symptoms similar to those of PCP intoxication, physiologic effects are generally mild and unlikely to require specific intervention. Patients with classic hallucinogen intoxication are more likely to remain fully oriented and less likely to be confused or show severe cognitive, verbal, or motor impairment. Anxiety and panic are sometimes observed. Marked mydriasis is a prominent feature of classic hallucinogen intoxication. Mild-to-moderate elevation in blood pressure and pulse may be observed. Overdoses of LSD and psilocybin are not known to cause death.

3,4-Methylenedioxymethamphetamine. MDMA intoxication does not usually cause thought disorder, paranoia, disorganized behavior, or aggression. It usually comes to attention in cases of overdose, in which case symptoms are essentially those of serotonin syndrome: hyperthermia, rigidity, hyperreflexia, seizures, rhabdomyolysis, renal failure, etc. Symptomatic intoxication is more likely to occur in the context of high dose, dehydration, and/or hyperthermia. Overdose can easily occur with MDMA because of the relatively low index of toxicity and the nonlinear pharmacokinetics of MDMA. MDMA inhibits its own metabolism, with the result that a relatively small increment in dose can lead to a much greater increase in blood levels and effects. In high doses, MDMA can cause classic hallucinogen-like effects. Delirium may be present, in which case it would be diagnosed instead of intoxication.

Other Serotonergic Hallucinogens. The effects of many of these drugs have not been adequately characterized. The effects of Bromo-DragonFLY and NBOMe analogs are reported to be similar to those of LSD, but serious adverse effects appear to be much more common, including hypertension, tachycardia, seizures, hyperpyrexia, elevated creatine kinase, vasoconstriction leading to tissue necrosis, and deaths, especially after users mistakenly took very large doses. The high potency combined with a low index of safety makes these drugs much more dangerous than classic hallucinogens.

Ibogaine. Ibogaine intoxication resembles classic hallucinogen intoxication but may include a greater degree of dissociation and confusion, similar to PCP and related drugs. While the acute effects can last 4 to 6 hours, many users report lingering effects that persist for several days.

Salvia. The effects of *Salvia* intoxication are similar to those of PCP or ketamine intoxication, but the duration (15 minutes or less) makes it unlikely that they will be observed in a clinical setting.

Anticholinergics. Anticholinergic drugs produce true hallucinations in the context of delirium, as described in the section on hallucinogen-induced delirium below. Peripheral signs may be less prominent with agents that are relatively potent centrally (e.g., scopolamine).

Pathology and Laboratory Examination

Urine toxicology may be useful for establishing the acute ingestion of hallucinogens and/or other drugs, although hallucinogens other than PCP and MDMA will not be detected with routine urine drug screens. False-negative screens for PCP are common. Toxicologic screening for other drugs of misuse is useful since more than one drug may have been ingested. Serum, urine, or gastric content samples can be sent to analytic laboratories for specific mass spectrographic analysis, but initial diagnosis and treatment must be from clinical assessment. The dose delivered from plant material is always uncertain.

MDMA intoxication usually comes to the attention of clinicians in cases of overdose, in which case patients may have rhabdomyolysis (elevated creatine phosphokinase), disseminated intravascular coagulation, hyperthermia, and acute renal failure. Deaths have occurred secondary to cerebral edema and seizures secondary to hyponatremia. Hyponatremia occurs much more frequently in women than in men. Cardiac arrhythmias and cardiovascular instability are also possible. Depending on drug–drug interactions, serotonin syndrome may also occur.

Differential Diagnosis

The most important immediate task in the evaluation of cases resembling hallucinogen intoxication is to rule out or treat any medical and psychiatric emergencies present. Since the symptoms of hallucinogen intoxication can overlap with those of psychosis and delirium, it is critical to distinguish hallucinogen intoxication from these disorders. Hallucinogen-induced psychosis may be diagnosed if the symptoms of psychosis dominate the clinical presentation and are beyond (in severity or duration) those typically associated with intoxication. Particularly for PCP, diagnostic boundaries between intoxication, delirium, and psychosis are blurry, and the patient’s mental state can fluctuate.

Intoxication with some drugs (e.g., stimulants) and withdrawal from others (e.g., alcohol and sedative–hypnotics) can also cause symptoms mimicking those of hallucinogen intoxication. The principal diagnostic tools include a history, if available, urine toxicology, additional laboratory tests to rule out metabolic derangements, and physical, neurologic, and mental status examination, with observation over time. In most cases, acute hallucinogen intoxication will resolve within hours. Except for very long-lasting drugs (DOM, ibogaine) or psychedelics prone to causing prolonged toxidromes (NBOMe), symptoms that last longer than 12 hours may be due to other causes or to hallucinogen-induced psychotic disorder. PCP intoxication can resemble the effects of other toxic drug ingestions or deliria. The combination of nystagmus (especially rotary), ataxia, and hypertension is relatively specific for PCP intoxication.

Course and Prognosis

Uncomplicated hallucinogen intoxication normally resolves when the acute effects of the drug wear off within 12 hours, except for very long-acting hallucinogens such as DOM or ibogaine. Barring medical complications, complete recovery from PCP intoxication within 24 to 72 hours is usual. Uncommonly, hallucinogen-induced disorders may follow hallucinogen intoxication (see below).

Treatment

PCP (or PCP-Like Hallucinogens). Treatment of acute ingestion is supportive, aimed at minimizing medical emergencies (seizures, hypertension, hyperthermia, and rhabdomyolysis) and behavioral consequences. PCP levels and symptoms may fluctuate over many hours (especially with oral ingestion). Ionized PCP may be trapped in the stomach. Activated charcoal may be administered if oral administration is suspected. Sensory input should be minimized. Physical restraint can prevent violent behavior but can also increase agitation and the risk of rhabdomyolysis. Agitation, seizures, and hypertension should be treated primarily with benzodiazepines. Antipsychotics may be useful in cases of relatively mild intoxication, but in more severe cases, they can lower the seizure threshold, cause dystonia, and exacerbate hyperthermia and anticholinergic symptoms. Hyperthermia should be treated with standard cooling measures. Rhabdomyolysis should be treated with intravenous fluids and diuretics. Alkalization of the urine is controversial because there is a theoretical risk of causing increased reabsorption of PCP.

Classic Hallucinogens. Except in cases of delirium or overdose, treatment of acute ingestion of a classic hallucinogen is conservative. Reassurance and a calm and safe environment are the most useful interventions. Anxious or paranoid patients may benefit from reminders that they are under the temporary influence of a drug. Some degree of insight is often preserved. If patients are agitated or in severe distress, oral or parenteral benzodiazepines are first-line treatment and are usually sufficient. Antipsychotics should be used only if behavior can not otherwise be managed. Antipsychotics with 5HT2A antagonist effects are preferred. D2 blockers such as haloperidol may worsen symptoms. Restraints should only be used if there is immediate danger of self-harm or assaultive behavior. Patients should be monitored closely, and their mental status should be assessed at frequent intervals due to potentially rapid fluctuations. Support from a sober family member or close friend can be helpful following the resolution of the acute intoxication state. Patients who have had very disturbing or traumatic experiences may benefit from short-term therapy to help integrate the experience and to identify and treat any psychological sequelae. Patients should be reminded that severe reactions could reoccur with future ingestion of hallucinogens. For those experiencing a severe toxidrome from NBOMe or psychedelics such as Bromo-DragonFLY, known to cause toxic reactions, aggressive sedation with benzodiazepines and fluid resuscitation are the mainstays of treatment. If severe serotonin syndrome is suspected, cyproheptadine may be considered in severe cases.

3,4-Methylenedioxymethamphetamine. MDMA intoxication often comes to medical attention in cases of overdose or medical complications. Such patients should be evaluated emergently, and ICU admission may be required. Cardiac monitoring and frequent monitoring of vital signs and mental status should be conducted. Activated charcoal should be administered if recent ingestion is suspected. Cooling measures should be undertaken in cases of hyperthermia. Dantrolene is sometimes used to treat hyperthermia due to MDMA ingestion, but its efficacy is not proven. Seizures, hypertension, and agitation are treated with benzodiazepines. In cases of rhabdomyolysis, intravenous fluids, alkalization of the urine, and furosemide are recommended.

PHENCYCLIDINE USE DISORDER AND OTHER HALLUCINOGEN USE DISORDER

Definition

The DSM-5-TR criteria for PCP use disorder and other hallucinogen use disorder are identical to those for all other classes of substances, with the exception that the criterion for withdrawal does not apply: there is no recognized syndrome of withdrawal in humans for hallucinogens, although there is evidence of clinically meaningful withdrawal following heavy MDMA use and evidence of PCP withdrawal in animal models. As with other SUDs, the SUD severity is specified as mild (2 to 3 criteria), moderate (4 to 5 symptoms), or severe (6 or more criteria).

Epidemiology

Problematic use of hallucinogens has not been described in cultures where hallucinogens are traditionally employed for religious, healing, or divinatory purposes, possibly due to cultural protections against misuse arising from long-established ceremonial group use. With the sudden increase in hallucinogen use in the 1960s in Western cultures, where there were no established cultural rituals around hallucinogen use, problematic patterns of use emerged and were the focus of considerable concern, in part because hallucinogen (particularly LSD) use was intertwined with the cultural upheaval of those times. Similar problematic patterns arose when MDMA, initially used for therapeutic purposes, became associated with raves in the 1980s.

Today, hallucinogen use disorders are uncommon. The 2019 NSDUH found past-year rates of DSM-IV hallucinogen abuse and dependence of 0.08% and 0.05%, respectively. Hallucinogen abuse was most prevalent among individuals aged 18 to 20 years (0.33% and 0.26%, respectively). In a 2019 study using data from 2011 to 2012, hallucinogen use disorder was similarly uncommon, with past-year and lifetime prevalence of 0.05% and 0.60%, respectively. 6.4% of lifetime hallucinogen users in that study reported lifetime hallucinogen use disorder, while 8.1% of past-year users reported past-year hallucinogen use disorder. Of patients meeting criteria for hallucinogen use disorder, past-year prevalence of mild, moderate, and severe forms was 79.9%, 13.1%, and 7.0%, respectively, and lifetime prevalence was 66.8%, 18.5%, and 14.6%, respectively. Unfortunately, most epidemiologic surveys do not distinguish between the various classes of hallucinogens, meaning little is known about rates of dependence for most individual hallucinogens.

According to a 2006 study, the most common features of hallucinogen dependence were emotional difficulties secondary to hallucinogen use, difficulty controlling use, salience of use, and tolerance. That study found that among individuals 12 to 21 years old initiating hallucinogen use in the previous 2 years, 2% exhibited hallucinogen dependence, and another 10% appeared to be at risk for developing dependence or already in the early stages of dependence. Use of particular hallucinogenic compounds appeared to increase the risk of rapid transition to hallucinogen dependence, with MDMA demonstrating more risk than LSD. In a 2007 study, of 2,035 people who had used any hallucinogen for the first time in the last 24 months, 56 (2.8%) had developed dependence according to DSM-IV criteria. Individuals who first used hallucinogens in early adolescence were at increased risk of developing dependence during this early time frame, while the risk was lower for Hispanic people. Increased risk of developing dependence was also associated with use of mescaline, MDMA, and PCP. In a 2008 study examining characteristics of past-year MDMA users compared to past-year users of other hallucinogens (including LSD, PCP, mescaline, peyote, and psilocybin) using NSDUH data from 2005, about 1.4% of the sample had used one or more hallucinogens in the past year. Overall prevalence of hallucinogen use disorder was 0.11%. Both groups were approximately two-thirds male, and two-thirds of both were aged 25 years or younger. Users of other hallucinogens were more likely to be White than were users of MDMA. Mean age of first use was 18 to 19 years in both groups. Most users in both groups used infrequently, with 34% of MDMA users and 22% of other hallucinogen users having 12 or more occasions of use in the past year. Among the past-year users, 4.9% of both groups met DSM-IV criteria for abuse, but 3.6% of the MDMA users versus 2.2% of the other hallucinogen users met criteria for dependence, suggesting that MDMA use may confer greater risk of moderate to severe hallucinogen use disorder than use of other hallucinogens. Frequent hallucinogen use and past-year marijuana use disorder were associated with hallucinogen use disorder among past-year hallucinogen users. In a 2017 study examining all major SUDs except for tobacco use disorder, hallucinogen use disorder was the most likely to occur in people with multiple SUDs, with alcohol use disorder being the least likely.

Etiology

As with other drugs, hallucinogen use disorder likely arises out of multiple interacting biologic, psychological, and social factors. Hallucinogen use disorders differ from other classes of SUD in that classic incentive reward-driven mechanism does not play a central role except for PCP and related compounds and MDMA-type drugs. PCP is avidly self-administered in animal models, and human users frequently show patterns of compulsive use, relapse, etc. Classic hallucinogens are not strongly reinforcing but rather aversive in animal models. Hallucinogens can cause euphoria, but this effect is not reliable, and dysphoric or mixed experiences are common. A withdrawal syndrome is not recognized for any of the hallucinogens. Although some degree of physical dependence is possible with NMDA receptor antagonist hallucinogens and MDMA-like compounds, it is unlikely that continued use is driven by negative reinforcement through relief of withdrawal symptoms. Strong cravings also do not appear to be common in hallucinogen users. Exposure, other drug use, expectancies, and perceived meaningfulness of the experience probably play more important roles.

Diagnosis and Clinical Features

Although the diagnostic criteria for hallucinogen use disorder are identical to those of any SUD, the criteria perform somewhat differently for hallucinogens than for other substances. It seems probable that there would be differences between the various types of hallucinogens as well. In general, hallucinogen use disorders appear to be associated with fewer adverse consequences and lower frequency of use than other SUDs. The two criteria most linked to loss of control (inability to cut down or control use and using more than intended) are by far the most difficult items (last to be endorsed) in item response theory studies of hallucinogen use disorder criteria. This suggests that people often continue to use hallucinogens because they enjoy or value the effects rather than due to having lost the ability to regulate use. There is no recognized withdrawal syndrome for the classic hallucinogens, although physical dependence to arylcyclohexamines and MDMA can occur.

True physical dependence does not occur with the classic serotonergic hallucinogens. The SUD criteria most commonly endorsed are spending a lot of time using or recovering from the substance and tolerance. The tolerance criterion is of questionable validity for classic hallucinogens because of the very rapid and reversible development of tolerance. Marked tachyphylaxis lasting no more than a few days can develop following a single dose of a classic hallucinogen such as LSD. This phenomenon does not imply tolerance in the sense of the DSM-5-TR SUD criterion.

Pathology and Laboratory Examination

There are few physical signs or symptoms of hallucinogen use disorders, independent of those of intoxication, though chronic ketamine use is associated with disorders of the GI (gastritis, cholestasis) and urinary tracts (ulcerative cystitis, impaired renal function). MDMA use may be associated with decreased numbers of serotonin transporters and 5HT2A receptors on positron emission tomography (PET) scan, but this is not used diagnostically. Urine, blood, or hair analysis can be used to detect ingestion of hallucinogens. The window of detection is often short, though PCP is detectable in urine for up to 8 days after use. Aside from PCP, other hallucinogens are not included in routine tests.

Neurotoxicity. MDMA is neurotoxic in animal models, and there is evidence of adverse effects on the brains of people who have used the drug extensively. There is fairly consistent evidence of reductions in serotonin transporters and 5HT2A receptors in MDMA users. Neurocognitive deficits in memory and inhibitory control have been found in some studies, although most of these studies are confounded by other drug use. Based on animal data, it seems likely that the neurotoxicity of MDMA may be dose related and mediated in part by hyperthermia. Hyponatremia also could play a role. No adverse cognitive effects have been observed in clinical studies of MDMA in which controlled doses (up to 150 mg) are given on a few occasions. In clinical research settings, the maximum temperature increase is about ½ degree Celsius.

Deficits in memory and executive function have been observed in heavy users of PCP and ketamine. Vacuolization is seen in high doses in animal models with

PCP and ketamine, possibly contributing to cognitive impairment in association with chronic use in humans. In rodents and primates, the most sensitive brain regions are the posterior cingulate and retrosplenial cortices.

There is no evidence of significant neurotoxicity with classic hallucinogens. Well-controlled trials with large samples have not yet explored this, but studies of long-term ritualistic users of peyote and ayahuasca have revealed no neurocognitive impairments.

Differential Diagnosis

Use of hallucinogens frequently occurs in the context of other substance use. In such cases, it is important to be clear about which symptoms are caused by which substance in order to arrive at the correct SUD diagnosis/es.

Course and Prognosis

Hallucinogen use disorders tend to be short and self-limited. Prevalence rates drop to near zero in people over the age of 50.

Treatment

There is little evidence to guide treatment of hallucinogen use disorders. Since there are no specific treatments or programs for treatment of hallucinogen use disorders, treatment-seeking patients with these disorders are treated within addiction treatment programs designed for other SUDs or multiple addictions. The efficacy of psychosocial treatments that have empirical support for the treatment of other SUDs is unknown for hallucinogen use disorders.

HALLUCINOGEN-PERSISTING PERCEPTION DISORDER

Definition

With HPPD, users re-experience one or more perceptual effects similar to those experienced during previous hallucinogen use, which cause clinically significant distress or impairments in functioning, and are not due to a general medical condition (i.e., migraine with aura, temporal or occipital lobe epilepsy, and temporal lobe infectious illness), better ascribed to another mental condition (i.e., schizophrenia, dementia, and delirium), or due to hypnopompic hallucinations. The perceptual symptoms mentioned as examples in the diagnostic criteria are visual (e.g., geometric hallucinations, image “trails” behind moving objects), but perceptual abnormalities from other sensory modalities and re-experiencing of other aspects of hallucinogen-induced altered states have been reported. The symptoms can occur spontaneously or be triggered by stress, anxiety, exercise, or use of another drug (i.e., cannabis). They can occur episodically or take the form of constant perceptual disturbances without an identifiable trigger. The term "flashback" has sometimes been used synonymously with HPPD but is distinguished from HPPD in connoting brief persisting perceptual effects, similar to those experienced during recent hallucinogen use, that do not necessarily cause clinical distress or impairment and can even be pleasurable. Unlike flashbacks, HPPD tends to be recurrent with a continuous or intermittent/paroxysmal pattern of symptoms that is either a slowly reversible or chronic permanent condition experienced as distressing. Today, these two related phenomena are sometimes referred to as HPPD Type I (i.e., flashbacks that are brief) and Type II (i.e., more persistent and disabling HPPD).

History

Persisting perceptual abnormalities ("flashbacks") following use of hallucinogens were first described over 100 years ago with mescaline. Reports next arose in the 1950s and 1960s, mostly in relation to LSD ingestion. These described prolonged changes in normal perception similar to the perceptual effects produced by hallucinogens, persisting subacutely (weeks to months) and chronically after hallucinogen use. It should be noted that in addition to the classic hallucinogens, some cases of persisting perceptual disturbances attributed to the use of MDMA, ketamine, and cannabis have been reported.

Epidemiology

Although the exact prevalence of HPPD is unknown, it is thought to be rare given the relatively few cases reported out of the tens of millions of doses of hallucinogens used over the last 60 years. The DSM-5-TR reports a prevalence of HPPD of approximately 4.2% among individuals who use hallucinogens. However, a population study using NSDUH data from 2001 to 2004 found no association between lifetime use of psychedelics and past-year use of LSD with past-year symptoms of HPPD-type visual phenomena. HPPD appears to be most prevalent in males, with symptoms developing in late adolescence. It is significantly less common in research settings with careful screening and preparation and has most commonly been associated with street-level LSD.

Etiology

Individual or family history of anxiety disorders and the presence of pre-existing tinnitus, visual "floaters," and concentration impairments may indicate vulnerability to HPPD. The pathophysiology of HPPD has yet to be conclusively elucidated. However, given the predominance of visual symptoms, their typology (i.e., geometric hallucinations, trailing, and positive afterimages), and that patients typically do not believe the visual phenomena are real, lower levels of the visual pathway in the brain (including the visual cortex) have been implicated. However, multiple brain regions may be affected, and the thalamus may be a particular area of interest since nonvisual perceptual stimuli (i.e., auditory and sensory) can also be abnormal in patients with HPPD. Abraham and Aldridge initially proposed that damage to lower-level visual system inhibitory cortical interneurons (which express 5HT2A receptors and release GABA upon activation) were primarily responsible for the mechanism of action of HPPD and could account for visual disturbances by altering the signal-to-noise ratio of incoming visual stimuli. A broader conceptualization of the pathophysiology of HPPD includes a misbalance or dysregulation between excitatory versus inhibitory input in the visual system where both increases in excitation and/or decreases in inhibition could result in the variety of visual disturbances seen in HPPD.

Diagnosis and Clinical Features

The most common perceptual distortions (illusions or hallucinations) in HPPD are in the visual system (i.e., geometric hallucinations, flashes of color, intensification of colors, trails of images of moving objects, positive afterimages, halos around objects, false perceptions of movement in the peripheral visual fields, micropsia, and macropsia), although auditory and sensory distortions have been reported as well. In addition to visual illusions and hallucinations, patients commonly experience depersonalization and derealization symptoms. It is important to note that HPPD is not a psychotic illness since patients are aware that the illusions and hallucinations are not real. However, symptoms may be disabling to those with Type II HPPD, and comorbid anxiety, depression, and suicidal ideation have been described.

Differential Diagnosis

Alternative explanations for unusual perceptual experiences such as visual epilepsy, migraine, delirium, dementia, schizophrenia, hypnopompic hallucinations, PTSD, and other hallucinogen-induced disorders must be considered before diagnosing HPPD. HPPD is easily distinguished from psychotic disorders because symptoms are sensory (primarily visual), delusions are absent, and insight, reality testing, and thought process are intact.

Course and Prognosis

The course of HPPD is variable, with a more benign course for Type I as opposed to a more persistent course for Type II. Symptoms among about half of patients with HPPD cease within 5 years, though they persist for decades in some cases. Symptoms can be worsened by a variety of stimulant drugs, marijuana use, fatigue, exercise, and infections.

Treatment

Supportive psychotherapy and cognitive behavioral treatment are warranted to help individuals cope with and adapt to their symptoms. Reassurance that the symptoms are due to HPPD, as opposed to an underlying psychotic disorder, may help patients cope. There is no established evidence-based pharmacologic algorithm to treat HPPD, with most data coming from case reports, case series, and open-label trials. Benzodiazepines (e.g., alprazolam, clonazepam) have been shown to reduce some symptoms without leading to disease remission. Their efficacy may be related to their action at the above-mentioned cortical serotonergic inhibitory interneurons with GABAergic outputs. Along these lines, there is some evidence that phenytoin (which, in addition to inhibiting sodium channels, also increases CNS GABA levels and potentiates GABA-mediated postsynaptic inhibition) decreases symptoms. Clonidine in an open-label trial was also shown to alleviate some of the symptoms, suggesting a possible role for sympathetic nervous system activation in the pathophysiology of HPPD. There is some evidence supporting the use of the opioid receptor antagonist naltrexone and typical antipsychotics (haloperidol, trifluoperazine, and perphenazine), although the adverse side effect profile of typical antipsychotics limits their use. The data on the utility of selective serotonin reuptake inhibitors (SSRIs) to ameliorate symptoms are mixed, with some data suggesting efficacy and some suggesting a worsening of symptoms. Risperidone (and possibly the atypical antipsychotics in general) should be avoided as there is evidence that risperidone worsens HPPD symptoms. More controlled trials are needed to establish better pharmacologic treatments for HPPD, which could also help further elucidate HPPD's pathophysiology. Given the absence of proven pharmacologic treatment, avoidance of triggers, especially cannabis, is an important recommendation for patients. Finally, some patients have found that wearing sunglasses, especially indoors, helps ameliorate the visual disturbances.

PCP- AND OTHER HALLUCINOGEN-INDUCED PSYCHOTIC DISORDER

Definition

DSM-5-TR defines hallucinogen-induced psychosis as a syndrome meeting criteria for a psychotic episode having its onset during or shortly following use of a hallucinogen, which persists beyond the usual period of action of the drug, and is not better accounted for by another psychotic disorder or other psychiatric, neurologic, or medical disorder. The disorder must be sufficiently severe to cause significant impairment and/or distress.

History

Phencyclidine. Although acute intoxication with PCP can closely resemble a psychotic state, persisting psychosis following PCP use was eventually recognized in 1974. These episodes closely resembled severe schizophrenic episodes, often included violent or aggressive behavior, and frequently persisted for weeks despite aggressive antipsychotic treatment before resolving completely. Although the prevalence and attention to this phenomenon have receded as the prevalence of PCP use has decreased, such cases are still seen fairly frequently in areas with prevalent PCP use.

Other Hallucinogens. After LSD became widely used as a recreational drug in the 1960s, it was increasingly recognized that its use was sometimes associated with persisting psychosis, a common example being first use by a teenager or young adult (during the age of typical onset of schizophrenia) who developed first-episode psychosis and went on to develop a chronic course of schizophrenia. However, it has never been conclusively shown that classic hallucinogens can produce a persistent psychosis in someone without an underlying psychotic disorder or diathesis (see “Epidemiology” section below).

Epidemiology

Phencyclidine. PCP psychosis is a well-known presentation in hospitals, but its prevalence is unknown. Increased delusional ideation has been reported among frequent users of ketamine. Ketamine and DXM have been reported to cause persisting psychotic episodes, but much less frequently than PCP. Risk factors also are not well known. Patients with pre-existing psychotic disorders are very sensitive to the psychotogenic effects of PCP, but such individuals would not generally receive the diagnosis of PCP-induced psychosis.

Other Hallucinogens. Reports of new cases of schizophreniform psychoses and related disorders following hallucinogen exposure began to emerge as hallucinogen use became widespread in the United States. Estimates of the prevalence of LSD-induced psychotic episodes as assessed by early psychedelic researchers and clinicians (many administering LSD to psychiatric inpatients) were reported as follows in two major studies of the topic: 0.8/1,000 research volunteers (a single case out of 1,250, where the volunteer was the identical twin of an individual with schizophrenia), and 1.8/1,000 psychiatric patients (7 cases out of approximately 3,850 patients); 0/170 research volunteers and 9/1,000 among psychiatric patients (37 out of 4,300 patients).

A 2018 longitudinal study of 6,788 patients with substance-induced psychoses, including 114 with hallucinogen-induced psychoses, followed for up to 20 years, found that 28% of those with a hallucinogen-induced psychosis eventually converted to schizophrenia or bipolar disorder. Notably, patients who experienced psychoses induced by marijuana, amphetamines, and alcohol had higher risks of converting to schizophrenia than patients who experienced hallucinogen-induced psychoses. Rates of conversion to bipolar disorder were higher for patients who experienced psychoses induced by sedatives, marijuana, and alcohol than those who experienced hallucinogen-induced psychoses. Being 16 to 25 years of age at the time of a substance-induced psychosis and self-harming after the episode were associated with higher risk of converting to either schizophrenia or bipolar disorder.

Though classic hallucinogen use can cause transient psychotic-like positive symptoms in normal volunteers and provoke sustained psychosis in vulnerable people with psychotic spectrum illnesses, there is little evidence linking classic hallucinogen use to prolonged psychosis in individuals without a psychotic diathesis. Cross-sectional population-level studies have found no significant associations between psychedelic use and increased rates of any psychiatric symptoms, including psychosis. The prevalence of schizophrenia has also remained relatively stable since first being measured, even though use of classic hallucinogens has fluctuated widely.

Rather than producing de novo chronic psychotic disorders, it appears most likely that hallucinogens precipitate psychosis in people who would inevitably develop it later or that previously unrecognized psychotic disorders may become apparent amid hallucinogen use. In a 2019 longitudinal study of 7,606 individuals with substance-induced psychoses, including 197 with hallucinogen-induced psychosis, familial risk scores for nonaffective psychosis predicted progression from substance-induced psychosis to schizophrenia. This points to hallucinogen and other substance-induced psychoses likely being drug-precipitated disorders in highly vulnerable individuals. Consistent with this hypothesis, in contemporary clinical trials investigating classical hallucinogens, in which participants with personal or family histories of psychotic disorders or bipolar disorder are excluded, there have been no reported cases of prolonged psychosis.

Etiology

Phencyclidine. Although PCP and ketamine intoxication (a transitory state) have been used extensively as a model of psychosis, etiology and pathogenesis of PCP-induced psychosis are not well studied. In animal models, a single dose of MK801 (an NMDA receptor antagonist with particularly strong psychotogenic effects) has been used to model first-episode psychosis. A single injection can cause persisting changes in hippocampal glutamatergic neurotransmission, resulting in loss of synaptic plasticity and deficits in learning and cognitive function, although other behavioral effects gradually resolve.

Other Hallucinogens. In many cases, the onset of psychosis following hallucinogen ingestion may be due to the expression of underlying diatheses for schizophrenia or bipolar disorder. There is no established etiology for other hallucinogens inducing persisting psychosis, but persistent dysregulation of serotonergic systems is possible. Also, because classic hallucinogens cause downstream effects on glutamatergic neurotransmission similar to those produced by PCP and ketamine, it is possible that persistent changes in glutamatergic systems could play a role in some cases of persisting psychosis induced by classic hallucinogens.

Diagnosis and Clinical Features

A history of recent hallucinogen use is necessary to establish the diagnosis. However, it may be difficult to confirm or rule out recent hallucinogen ingestion. PCP, in particular, is often misrepresented as other drugs. Patients may deny drug use, and drug screens will often be negative.

Phencyclidine. PCP-induced psychosis is often phenomenologically indistinguishable from schizophrenia. Thought disorder and negative symptoms may be prominent, and delusions and hallucinations are not always present.

Other Hallucinogens. Patients with hallucinogen-induced psychosis are younger at onset, have more visual hallucinations and depression, and have a greater prevalence of a family history of mood disorder than non–drug-using patients with schizophrenia. They also have more mood swings, euphoria, grandiosity, hyper-religiosity, and multimodal hallucinations, and more prominent positive than negative symptoms. Thus, the psychosis tends to have elements of mania. If mood symptoms are prominent, the diagnosis of a hallucinogen-induced mood disorder should be considered.

Pathology and Laboratory Examination

Urine drug screens may be useful to confirm recent ingestion of hallucinogens, but routine tests are available only for PCP and MDMA. Physical examination and laboratory findings will be normal following resolution of acute intoxication or delirium and any medical sequelae.

Differential Diagnosis

Hallucinogen-induced psychosis usually resembles schizophrenia but sometimes has manic features as well (in which case the diagnosis of hallucinogen-induced mania should be considered). The symptoms of hallucinogen-induced psychosis may be indistinguishable from those of a primary psychotic disorder. Sudden onset following recent history of hallucinogen use strongly suggests the diagnosis. Signs and symptoms of hallucinogen intoxication overlap substantially with those of hallucinogen-induced psychosis, so the diagnosis of hallucinogen-induced psychosis should be used only if the symptoms significantly exceed the expected duration and/or intensity of the intoxication state. The diagnosis of hallucinogen-induced delirium should be used if symptoms occur solely in the course of a delirium. If symptoms persist beyond 4 to 8 weeks, the patient likely has a primary psychotic disorder that was exacerbated or precipitated by hallucinogen use.

Course and Prognosis

Hallucinogen-induced psychoses follow a variable course. At one extreme are cases that resemble intoxication with the substance, resolving rapidly and completely. In other cases, symptoms persist for weeks despite aggressive treatment. This pattern is relatively common with PCP-induced psychosis that has been observed to persist for up to 6 weeks but eventually resolves completely. Finally, there are cases in which the psychosis becomes chronic. In these cases, a diagnosis of primary psychotic disorder is appropriate, most commonly schizophrenia or bipolar disorder.

Treatment

Hallucinogen-induced psychoses are treated much like any other acute psychosis. Hospitalization may be necessary to ensure safety. Since the course of illness is unpredictable, treatment should aim to prevent irreversible consequences of illness and preserve relationships and functional capacity. Chronic psychotic disorders with onset during hallucinogen use are treated like any other case of first-break psychosis, except that particular care should be taken to minimize future harms associated with substance use. Even if symptoms resolve and the patient is not experiencing any further psychotic symptoms, given that a substantial minority of patients will eventually receive a formal psychotic disorder diagnosis, ongoing periodic monitoring and assessment by a mental health provider for psychotic symptoms is strongly recommended, particularly for younger patients.

Phencyclidine. Antipsychotic medications are frequently used to treat PCP-induced psychosis, but data on their efficacy are limited. Benzodiazepines can be used to control agitation. In animal models, atypical antipsychotics are more effective than typical antipsychotics in reversing the effects of PCP. Electroconvulsive therapy (ECT) may be necessary for patients with antipsychotic-resistant PCP-induced psychosis. The use of drugs that modulate NMDA receptor–mediated neurotransmission, such as lamotrigine, glycine, D-serine, and D-cycloserine, as well as group II metabotropic glutamate receptor agonists such as the compound LY354470, has been proposed based on preclinical data. However, there are no clinical trials of these agents in PCP- or ketamine-induced psychosis.

Other Hallucinogens. Treatment of other hallucinogen-induced psychosis does not differ substantially from the treatment of other psychoses. In addition to antipsychotic medications, lithium, anticonvulsants, ECT, benzodiazepines, and antidepressants have been used with reported success, but controlled trials are lacking.

PHENCYCLIDINE- AND OTHER HALLUCINOGEN-INDUCED BIPOLAR DISORDER

The diagnosis of substance/medication-induced bipolar and related disorder is used if the patient develops a manic, hypomanic, or mixed episode during or soon after hallucinogen intoxication. Hallucinogen effects on mood are common but usually transient. Mood may fluctuate rapidly. Symptoms resembling those of major depression, mania, or hypomania may appear but rarely persist beyond intoxication. Although this syndrome is not well characterized, manic features are sometimes seen following hallucinogen intoxication, particularly in association with psychosis. Since there are no specific data on treatment of this disorder, treatment should ensure safety, and clinicians should consider medications that are typically effective for similar symptoms. As with most substance-induced disorders, symptoms are likely to remit within days to weeks with cessation of the causative substance. If symptoms persist beyond 4 to 8 weeks, a primary bipolar disorder diagnosis should be considered.

PHENCYCLIDINE- AND OTHER HALLUCINOGEN-INDUCED DEPRESSIVE DISORDER

The diagnosis of substance/medication-induced depressive disorder is used if the patient develops a depressive episode during or soon after hallucinogen intoxication. Depressive symptoms can develop during dysphoric hallucinogen experiences (“bad trips”) and usually resolve with the acute effects of the drug. Evaluation and management of suicide risk are imperative since suicides have occurred under these conditions. Hallucinogen use can also exacerbate a pre-existing mood disorder.

PHENCYCLIDINE- AND OTHER HALLUCINOGEN-INDUCED ANXIETY DISORDER

The diagnosis of substance/medication-induced anxiety disorder is used if the patient develops an anxiety episode during or soon after hallucinogen intoxication. Clinically significant anxiety (distinct from paranoia or agitation) is rarely reported with PCP or ketamine intoxication but is common during intoxication with *Salvia* and many of the classic hallucinogens. Full-blown panic attacks can occur. These symptoms are self-limited but may be very distressing. Patients usually respond well to reassurance and being in a calm and safe environment. Oral or parenteral benzodiazepines may be useful in extreme and/or persistent cases. There is little or no evidence for persisting hallucinogen-induced anxiety disorders.

PHENCYCLIDINE- AND OTHER HALLUCINOGEN-INDUCED INTOXICATION DELIRIUM

The diagnosis of substance intoxication delirium should be used instead of substance intoxication if disturbances in attention and cognition predominate and are severe enough to warrant clinical attention. Because such symptoms are part of the intoxication syndrome for PCP and related substances, the diagnosis of delirium depends primarily on the severity of the neurocognitive symptoms. This is also true for anticholinergic agents, whose hallucinogenic effects are usually considered to be symptoms of anticholinergic delirium, with features typically including disorientation, agitation, grossly impaired cognition and memory, and fluctuating mental status. For most other hallucinogens, delirium is uncommon but may occur in cases of severe intoxication or may represent a response to serious medical consequences of intoxication (e.g., in cases of overdose of MDMA or other nonclassic serotonergic hallucinogens). Severe toxidromes characterized by prolonged delirium have been reported with NBOMe analog ingestion.

Management of substance-induced delirium is primarily focused on correcting any dangerous medical situations and keeping the patient safe while they recover from the effects of the substance. Aggressive sedation, airway protection, and fluid resuscitations are central to managing substance-induced delirium. Activated charcoal may be administered in cases of ingestion. Specific issues in the treatment of delirium secondary to PCP and related hallucinogens are discussed in the section on PCP and related substance intoxication above.

Treatment of anticholinergic delirium involves supportive care and administration of benzodiazepines. Antipsychotics should be avoided. Use of physostigmine is controversial and should be reserved for cases where it is clear that anticholinergics and no other substances have been ingested.

UNSPECIFIED PHENCYCLIDINE-RELATED DISORDER AND UNSPECIFIED HALLUCINOGEN-RELATED DISORDER

This residual category is used to describe situations in which symptoms characteristic of a PCP- or other hallucinogen-related disorder predominate and cause clinically significant distress or impairment but do not meet full criteria for any specific disorder.

Further Readings

Abraham HD, Aldridge AM, Gogia P. The psychopharmacology of hallucinogens. *Neuropsychopharmacology*. 1996;14(4):285–298.

Alper KR. Ibogaine: a review. *Alkaloids Chem Biol*. 2001;56:1–38.

Bokor G, Anderson PD. Ketamine: an update on its abuse. *J Pharm Pract*. 2014;27(6):582–586.

Canal CE, Murnane KS. The serotonin 5-HT2C receptor and the non-addictive nature of classic hallucinogens. *J Psychopharmacol*. 2017;31(1):127–143.

Cohen S. Lysergic acid diethylamide: side effects and complications. *J Nerv Ment Dis*. 1960;130:30–40.

Drug Abuse Warning Network, 2011. Selected tables of national estimates of drug-related emergency department visits. *Center for Behavioral Health Statistics and Quality.* SAMHSA; 2013.

Fantegrossi WE, Murnane KS, Reissig CJ. The behavioral pharmacology of hallucinogens. *Biochem Pharmacol*. 2008;75(1):17–33.

Griffiths RR, Johnson MW, Richards WA, Richards BD, McCann U, Jesse R. Psilocybin occasioned mystical-type experiences: immediate and persisting dose-related effects. *Psychopharmacology (Berl)*. 2011;218(4):649–665.

Halberstadt AL. Recent advances in the neuropsychopharmacology of serotonergic hallucinogens. *Behav Brain Res*. 2015;277:99–120.

Halpern JH, Lerner AG, Passie T. A review of hallucinogen persisting perception disorder (HPPD) and an exploratory study of subjects claiming symptoms of HPPD. *Curr Top Behav Neurosci*. 2018;36:333–360.

Halpern JH, Pope HG Jr. Do hallucinogens cause residual neuropsychological toxicity? *Drug Alcohol Depend*. 1999;53(3):247–256.

Hartogsohn I. *American Trip: Set, Setting, and the Psychedelic Experience in the Twentieth Century*. MIT Press; 2020.

Johansen PØ, Krebs TS. Psychedelics not linked to mental health problems or suicidal behavior: a population study. *J Psychopharmacol*. 2015;29(3):270–279.

Johnson M, Richards W, Griffiths R. Human hallucinogen research: guidelines for safety. *J Psychopharmacol*. 2008;22(6):603–620.

Johnson MW, MacLean KA, Reissig CJ, Prisinzano TE, Griffiths RR. Human psychopharmacology and dose-effects of salvinorin A, a kappa opioid agonist hallucinogen present in the plant Salvia divinorum. *Drug Alcohol Depend*. 2011;115(1–2):150–155.

Kaertner LS, Steinborn MB, Kettner H, et al. Positive expectations predict improved mental-health outcomes linked to psychedelic microdosing. *Sci Rep*. 2021;11(1):1941.

Kendler KS, Ohlsson H, Sundquist J, Sundquist K. Prediction of onset of substance-induced psychotic disorder and its progression to schizophrenia in a Swedish National Sample. *Am J Psychiatry*. 2019;176(9):711–719.

Kerridge BT, Saha TD, Smith S, et al. Dimensionality of hallucinogen and inhalant/solvent abuse and dependence criteria: implications for the Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition. *Addict Behav*. 2011;36(9):912–918.

McCabe SE, West BT, Jutkiewicz EM, Boyd CJ. Multiple DSM-5 substance use disorders: a national study of US adults. *Hum Psychopharmacol*. 2017;32(5):10.

Morgan CJ, Muetzelfeldt L, Curran HV. Consequences of chronic ketamine self-administration upon neurocognitive function and psychological wellbeing: a 1-year longitudinal study. *Addiction*. 2010;105(1):121–133.

Nichols DE. Hallucinogens. *Pharmacol Ther*. 2004;101(2):131–181.

Nichols DE, Walter H. The history of psychedelics in psychiatry. *Pharmacopsychiatry*. 2021;54(4):151–166.

Response Analysis Corporation. *Drug Experience, Attitudes and Related Behavior Among Adolescents and Adults: Detailed Tabulations, Part 2c. Experience Data. A Nationwide Study for the National Commission on Marihuana and Drug Abuse*. Response Analysis Corporation; 1973.

Shalit N, Rehm J, Lev-Ran S. Epidemiology of hallucinogen use in the U.S. results from the National epidemiologic survey on alcohol and related conditions III. *Addict Behav*. 2019;89:35–43.

Stone AL, O’Brien MS, De La Torre A, Anthony JC. Who is becoming hallucinogen dependent soon after hallucinogen use starts? *Drug Alcohol Depend*. 2007;87(2–3):153–163.

Stone AL, Storr CL, Anthony JC. Evidence for a hallucinogen dependence syndrome developing soon after onset of hallucinogen use during adolescence. *Int J Methods Psychiatr Res*. 2006;15(3):116–130.

Studerus E, Kometer M, Hasler F, Vollenweider FX. Acute, subacute and long-term subjective effects of psilocybin in healthy humans: a pooled analysis of experimental studies. *J Psychopharmacol*. 2011;25(11):1434–1452.

Substance Abuse and Mental Health Services Administration. *Results from the 2019 National Survey on Drug Use and Health: Summary of National Findings*. Substance Abuse and Mental Health Services Administration; 2020.

Vollenweider FX, Kometer M. The neurobiology of psychedelic drugs: implications for the treatment of mood disorders. *Nat Rev Neurosci*. 2010;11(9):642–651.

Wu LT, Ringwalt CL, Mannelli P, Patkar AA. Hallucinogen use disorders among adult users of MDMA and other hallucinogens. *Am J Addict*. 2008;17(5):354–363.

Wu LT, Ringwalt CL, Weiss RD, Blazer DG. Hallucinogen-related disorders in a national sample of adolescents: the influence of ecstasy/MDMA use. *Drug Alcohol Depend*. 2009;104(1–2):156–166.

11.7 Inhalant-Related Disorders

Cameron J. Davidson, Ph.D., D.W. Svenson, M.A., and Scott E. Bowen, Ph.D.

Inhalants are an assorted group of compounds that readily vaporize at room temperature and are intentionally inhaled for their psychoactive effects. These volatile compounds are used in many household products as solvents, propellants, markers, thinners, cleaning products, and fuels. Some of these volatile compounds are also commonly used for their solvent properties, which allow for the dissolution of solids into a liquid form without damaging the structure of the solid. The most common example of this is the use of toluene in glues. When a glue, a liquid mixture of solids and the solvent toluene, is pushed from its tube, the toluene readily vaporizes into the air leaving only the solid behind to set or harden. The same principle is also used in markers and paint. Conversely, to remove a hardened glue, a solvent (containing toluene or some similar compound) may be used to resuspend the solid in a liquid form and allow for removal. The term "*inhalants*" encapsulates a wide range of volatile hydrocarbons, such as toluene, hexane, trichloroethylene, trichloroethane (TCE), xylene, and gasoline, among many others. These compounds and others are grouped together under the term inhalants because of their route of administration, lack of combustion or vaporization, and their shared pharmacologic properties.

Many categorization strategies have been used to organize inhalants, such as product type or form, chemical properties, or shared pharmacologic impacts from a shared pattern of misuse. At present, however, data are too scant to support a more refined classification system than that in the *Diagnostic and Statistical Manual of Mental Disorders,* Fifth edition (DSM-5). Due to the loose inclusion criteria to this category, it is easier to distinguish specific compounds that are not to be included in the category of inhalants. For example, the DSM-5 excludes combusted substances (e.g., tobacco, marijuana, crack cocaine), anesthetic gasses (e.g., nitrous oxide), and nitrites (amyl, butyl, and isobutyl), instead classifying these into specific disorders (e.g., tobacco, marijuana, and crack cocaine) or as other (or unknown) substance-related disorders. However, some disorders associated with the excluded compounds are briefly discussed at the end of this chapter.

Individuals, especially young adolescents (12 to 17 years old), inhale the vapors from these products for the intoxicating effects (see Table 11.7–1). Lifetime inhalant use for eighth graders in 2019 was reported by the Monitoring the Future (MTF) study to be 9.5%, which falls in fourth place of most used intoxicants for this age range only behind alcohol (24.5%), marijuana (15.2%), and cigarettes (10.0%). According to the Substance Abuse and Mental Health Services Administration (SAMHSA), the proportion aged 12 years and older reporting use in the past year is generally low but has been rising. The overall rate reporting "use in the last year" for 2016 was ~0.6%, but rose to ~0.8% in 2019, representing an approximate 30% increase in the usage rates over this time frame. With increasing age, prevalence of past-year inhalant use declines and few people ever meet criteria for inhalant use disorder (IUD). While ~3% of those aged 12 to 17 years reported use (~743,000 people), 1.7% of those aged 18 to 25 years (~577,000 people) reported inhalant use in the last year which represents approximately seven and four times the rate of those aged 26 years and above (0.4%; ~822,000 people). Inhalants are associated with multiple problems, including conduct disorder, noninhalant substance use disorders (SUDs), mood, anxiety, and personality disorders, suicidality, history of physical/sexual abuse and neglect, and victimization. For some adolescents, the early time-limited use of inhalants may be a signal of a lifelong problem with risk-taking propensity, novelty seeking, impulsivity, and externalizing behaviors. A comparatively smaller subpopulation chronically uses inhalants, which have been associated with multiple sequelae, including major and minor behavioral problems and toxicity-related organ pathology.

Table 11.7–1.
Commonly Misused Inhalants

Group	Characteristics	Commercial Products
Alkyl nitrites	A nitrite group	"Poppers" Room odorizers
Nitrous oxide	Inorganic anesthetic gas/propellant	Whipped dairy cream chargers, nitrous oxide tanks
Other anesthetics	Gases or volatile liquids	Anesthesia bottles (ether, Halothane, Isoflurane, Desflurane)
Propellants–fuels	Flammable gases that are liquefied under pressure	Cigarette lighters, cooking fuels, hairsprays, central air conditioning units
Solvents	Liquids that dissolve other substances without any change in chemical composition	Paints, paint thinners, glues, inks, lacquers, varnish removers, degreaser agents, spot removers, motor fuels

Adapted from Bowen SE, Cruz SL. Inhalants: addiction and toxic effects in the human. In: Madras B, Kuhar M, eds. *The Effects of Drug Abuse on the Human Nervous System*, Vol. 2, 553–569. Academic Press. Copyright © 2014, with permission from Elsevier. doi:10.1016/B978-0-12-418679-8.00018-6.

DEFINITION

The section on inhalant-related disorders in the DSM-5 includes four relevant categories: inhalant intoxication, inhalant use disorder, inhalant-induced disorders, and unspecified inhalant-related disorder. IUD is characterized by clinically significant impairment and/or distress due to a problematic pattern of volatile hydrocarbon inhalation. Inhalant-induced disorders is a broad category and includes disorders that result from the toxic effects of inhaled substances (e.g., inhalant intoxication, inhalant intoxication delirium, and inhalant-induced psychotic, depressive, anxiety, and major and mild neurocognitive disorder). The specific inhalant-induced disorders will be discussed later in this chapter. An unspecified inhalant-related disorder diagnosis is more of an "umbrella term" reserved for persons presenting with signs of an inhalant-related disorder who experience clinically significant distress, but who do not meet full diagnostic criteria for any specific inhalant-related disorder or any other DSM-5 substance-related and addictive disorder.

COMPARATIVE NOSOLOGY

The DSM-5 contains the diagnostic criteria and description of specific subtypes of this disorder. These subtypes include IUD, inhalant intoxication, and unspecified inhalant-related disorder and are found in the inhalant-related disorders section of the substance-related and addictive disorders chapter of the DSM-5. Related but associated impacts such as inhalant-induced depressive disorder are included and can be found in the substance/medication-induced depressive disorders section.

The 11th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-11) refers to inhalants as "volatile solvents" and describes their use as "characterized by the pattern and consequences of volatile inhalant use" (6C4B). The use can be classified as a harmful pattern of volatile inhalants that is episodic (6C4B.10), continuous (6C4B.11), or unspecified (6C4B.1Z). The ICD-11 includes a diagnosis of "inhalant withdrawal," whereas the DSM-5 doesn't discuss inhalant withdrawal symptomatology due to a "lack of evidence of significant withdrawal documented in humans similar to that of phencyclidine and/or hallucinogens." The ICD-11 expanded upon the withdrawal diagnosis (6C4B.4) and provided the definition as "a clinically significant cluster of symptoms, behaviors and/or physiologic features, varying in degree of severity and duration, that occurs upon cessation or reduction of use of volatile inhalants in individuals who have developed volatile inhalant dependence or have used volatile inhalants for a prolonged period or in large amounts." Some of the specific signs are defined by insomnia, anxiety, irritability, dysphoric mood, shakiness, perspiration, nausea, and transient illusions. More evidence of a significant but minor or short-lived withdrawal period in humans may lead to changes in the next iteration of the DSM to include diagnostic criteria or description of withdrawal symptomatology.

EPIDEMIOLOGY

The impact on public health from inhalant misuse may be quantified by the prevalence of inhalant use, DSM-5-defined IUD, inhalant-related visits to hospital emergency departments (EDs), inhalant-related admissions to drug treatment programs, and deaths associated with inhalant use. Each of these factors will be discussed, outlined, and explored below.

Prevalence of Inhalant Use

The U.S. government sponsors three large nationally representative studies of drug use. The MTF survey assesses 8th, 10th, and 12th graders and the Youth Risk Behavior Survey (YRBS) assesses 9th to 12th graders, whereas residents 12 years of age and older, including adults, living in U.S. households are assessed in the National Survey on Drug Use and Health (NSDUH). Because of differences in sampling and interview methods, the MTF survey and YRBS often show higher prevalence rates of adolescent inhalant use compared to the NSDUH. All three surveys tend to demonstrate similar patterns and trends over time and provide important data related to population statistics and prevalence of drug use generally. More U.S. eighth graders report use of inhalants than any other illicit drug except marijuana (13.4%). In 2019, about 1 in 10 U.S. eighth graders reported inhalant use even once in their lifetime (~9.5), and about half who reported lifetime use also reported use in the last year (~4.7). A little less than half of those reporting use in the last year reported use in the past 30 days (~2.1%). This is in contrast to about 1 in 100 who reported cocaine use (~1.2%), with ~0.7% reporting use in the past year and only ~0.3% reporting use in the past month. The rates of inhalant use in the past year among youths (12 to 17 years old: ~3%) correspond to an approximate population of ~743,000; compared to young adults (18 to 25 years old: ~1.7%) at ~577,000; and adults (25 years old and above; 0.4%) at ~822,000 people. While the rates in the 12 to 17 and 18 to 25 age ranges increased from the previous 3 years, the rate of those 25+ years old has remained relatively stable. The reported increases in the younger two categories resulted in all persons 12 years or older reporting use in the previous year rising from ~0.6% in 2018 to ~0.8% in 2019 (an increase of ~400,000 people). Taken together, these values represent a small but not insignificant portion of ~2.1 million people in the United States in the previous year and ~730,000 initiates of inhalant use, which are greater numbers than cocaine, prescription sedative misuse, methamphetamine, and heroin use. Although the rates of initiation are relatively high, the noncontinuation rate among 12th graders is also very high at ~65%. Considering that rates of inhalant misuse and disorder are higher in other typically underdeveloped nations with many homeless young people, the worldwide number would be much greater.

Between 2005 and 2019, lifetime and annual inhalant use prevalence rates in the MTF survey steadily declined; however, there was also a decline in the perceived harmfulness (e.g., “great risk in the use of inhalants once or twice”). This decline in perceived harmfulness among 8th graders since 2012 represents a reversal of the trend for previous years' surveys to report an increase in this belief and declines among 10th graders since 2014. This pattern of increase and subsequent decline may be linked to the discontinuation of a governmental campaign that took place in the mid 1990s to the early 2000s to inform the public of the dangers of inhalant use. This campaign was subsequently terminated after observing a the intended effect of decline in inhalant use. This pattern has been termed "generational forgetting" and results in a situation where the dangers of a drug were discussed and addressed but over the years the perceptions about its dangers are unknown and these drugs are "rediscovered." The declines in perception of risk potentially indicate that an increase in use may be reported on future reports.

The prevalence of inhalant use among adolescents is generally similar between males and females, however, recently the rates of female lifetime use have been higher in 8th (11.3%) and 10th (7.5%) grades than that of males (7.6% and 6.2%) but not in the 12th grade (5.8% male vs. 4.8% female: MTF). The most recent YRBS for 2019 showed that females tended to use at higher rates than males (6.9% vs. 5.7%). There were also differences among sexual identities, with those identifying as heterosexual (straight) having rates around 5% and those reporting "gay or lesbian" or "bisexual" reporting much higher rates of 11% and 15.8%, respectively. American Indian or Alaska natives tend to have higher rates (15.1%) compared to Caucasian (White) (6.3%), Black or African American (7.2%), Hispanic or Latino (6.6%), or multiple races (7.2%). Lowest of all is that of Asian individuals (1.9%). Between 1995 and 2013, a significant linear decline in the prevalence of lifetime inhalant use was noted. More current iterations of the YRBS have shown decreases from 2011 to 2019 with no change between 2017 and 2019 (~ 6.3% report using at least once). Inhalant use prevalence rates varied significantly across the states and large urban school districts participating in the YRBS. Continuation rates can be seen when looking at the past-month prevalence rate compared to the lifetime rate, which are disparate for inhalants, indicating high discontinuation rates. Inhalant use occurs mainly in younger age groups, with the prevalence of past-year inhalant use falling from a peak of around 3.5% to 4.5% at age 14 to less than 1% for all adult inhalant users. Most new users do not endorse regular inhalant use, and more than half report only 1 to 3 days of use in the prior 12 months.

The prevalence of inhalant problems in small, isolated groups or communities and in high-risk populations, including juvenile and criminal justice system–involved persons, may deviate from larger national trends. For example, certain Native American communities experience a high prevalence of these problems, with higher odds of initiation (~2 × comparable Caucasian (White) youth). However, rates of use may vary across communities. For example, in one large recent general population study of two Native American communities assessing persons aged 11 to 17 years who identified as American Indian or Caucasian

(White), odds ratios were different by tribe for males with Upper Great Lakes having an increased likelihood than their South West peers, whereas females did not differ. These indicate that geographic or ethnic enclaves may have differential rates of initiation and require focused interventions for the specific populations.

Studies of juvenile probationers, incarcerated youth in short- and long-term residential treatment, and youth and adults treated for behavioral problems or other noninhalant substance dependencies have reported lifetime inhalant use prevalence rates ranging from 18% to nearly 40%. This may indicate that inhalant use may be more common among those who have behavioral issues, especially incarcerated youths; however, this is a correlation among a restricted sample, thus, causality cannot be inferred without further investigation.

In countries other than the United States, rates vary from country to country but are high among unhoused children. These children may live on the street in loose-knit juvenile gangs where inhalant use is socially acceptable in their peer group and used to cope or dissociate. Among such children, inhalant use can be very common with prevalence rates reportedly exceeding 70% in some cities. In addition, there may be cultural differences in the specific products inhaled due to their availability or standards set for specific countries.

Prevalence of Inhalant Use Disorder

Prevalence of inhalant use *per se* should not be confused with the prevalence of IUD as defined by the DSM-5, and most users never progress to the formal diagnosis of IUD. Although new surveys using these criteria have not yet been conducted, information about prevalence can be determined using previous iterations of the DSM. Given the changing nature of the demographics as well as declining use generally for inhalants, newer studies need to be undertaken using DSM-5 criteria. Nationally representative surveys suggest that IUD is relatively rare. For example, a national household survey assessed DSM-IV-TR-defined past-year IUDs. From this survey, about 0.1% of the U.S. population (12 years of age or older) met criteria for an IUD in any given 12-month period. As might be predicted from information available regarding inhalant use patterns, the prevalence of past-year DSM-IV-TR inhalant abuse and dependence is somewhat higher (i.e., 0.4%) among youth 12 to 17 years of age. In 2012, the household survey noted that an estimated 34,000 persons in the United States met DSM-IV-TR criteria for inhalant dependence disorder.

Another survey of about 8,000 randomly sampled Americans (15 to 54 years of age) found that 0.3% had been dependent on inhalants in their lifetimes (as defined by DSM-III-R); again, rates of inhalant dependence (0.6%) were higher among younger age groups (15 to 24 years of age). In that study, most people who had tried inhalants had not become dependent on them; only about 1 in 27 respondents who reported trying inhalants met criteria for lifetime dependence (1 in 13 of individuals aged 15 to 24 years); in contrast, about 1 in 3 individuals reporting nicotine use met criteria for lifetime dependence on that substance. Although prevalence of inhalant use is similar across genders in most studies, inhalant dependence was about four times more likely among males, compared with females.

A nationally representative study of over 43,000 respondents 18 years of age or older found similarly low past-year (0.02%) and lifetime (0.3%) prevalence rates of IUDs (DSM-IV-TR-defined), with most of those individuals meeting criteria for abuse; inhalant abuse and inhalant dependence prevalence rates were the lowest among the drug categories reported (i.e., alcohol, sedatives, tranquilizers, opioids, amphetamines, hallucinogens, cannabis, and cocaine). In that study about one in five lifetime inhalant users met criteria for a lifetime IUD, although prevalence of lifetime inhalant use was relatively low (1.7%) compared to most other studies. Even though women are more likely to use inhalants, lifetime IUDs were about five times more prevalent among men.

Limited data are currently available that help to identify which inhalant users are likely to progress to formal inhalant-related disorders. One national cross-sectional survey found that living in a large metropolitan area, previous use of mental health services, number of delinquent behaviors, multiple other drug use disorders, and number of different inhalants used were associated with a DSM-IV-TR diagnosis of inhalant abuse among past-year inhalant users. Use of mental health services, history of foster care placement, multiple other drug use disorders, weekly use of inhalants, and early age at first inhalant use were associated with inhalant dependence among past-year inhalant users.

Several studies of high-risk populations have identified high rates of IUDs among inhalant users. One statewide survey of incarcerated youth reported that 47% of inhalant users met criteria for lifetime DSM-IV-TR inhalant abuse or dependence. Similarly, a study of 162 young adult inhalant users in St. Louis, Missouri, reported that 47% met criteria for DSM-IV-TR inhalant abuse or dependence. Another large study of youth in treatment for substance use and behavioral disorders reported that 18.3% of inhalant users met criteria for DSM-IV-TR–defined IUDs. Thus, while IUD may be uncommon in the general population, high levels of inhalant use and IUD may prevail in correctional and institutional settings serving at-risk youth and adults. New investigations should be undertaken using the DSM-5 to explore how the changes in usage rates relate to diagnosis of IUD, as both have changed from the previous reports.

Prevalence of Inhalant-Related Emergency Department Visits

The U.S. government monitors drug-related ED visits in hundreds of non-Federal, short-stay, general medical, and surgical hospitals around the country.

In 2011, among the nearly 2.5 million drug-related ED visits, 51% which involved illicit drugs, inhalants (including nitrous oxide and nitrites and excluding accidental exposures) were implicated in an estimated 10,032 visits (0.8% of ED visits involving illicit drugs); in contrast, there were 505,244 cocaine- and 455,668 marijuana-related ED visits in 2011. Of the ED visits for inhalants, only 23.4% also involved alcohol, which is among the lowest illicit drugs with only heroin (18.6%) and methamphetamine/amphetamine (16.2%) showing lower coinvolvement rates. These data are now about 10 years or older, however, and represent the last iteration of the Drug Abuse Warning Network (DAWN) study observing ED admittance for misused substances. Newer studies should be conducted as the rates of use, as well as the proportions of the sexes and races, have changed over the last 10 years.

Prevalence of Inhalant-Related Treatment Admissions

The U.S. government collected data on 1,935,541 admissions to U.S. addiction treatment programs during 2018. The number of admissions with inhalants listed as the primary substance of misuse was 829, constituting 0.04% of treatment admissions that year. Approximately 60% of inhalant admissions were male and 40% were female. Caucasian (White) accounted for 79% of admissions, African Americans for 7.2%, American Indians/Alaska Natives for 4.0%, Asians/Hawaiians for 0.6%, and “other” races for 9.2%. Approximately 80% of all admissions were for persons aged 20 years or younger. In general, these findings indicate that, although inhalant use is generally considered a problem of childhood or adolescence, most persons treated for primary IUDs are late adolescence to early adulthood.

Prevalence of Inhalant-Related Deaths

The repercussions of abusing inhalants can tragically result in fatalities. Inhalant misuse has been reported to result in death from the direct toxic effects of the chemicals, as well as accidental injury, aspiration, suffocation, flash burns, and many other causes. Inhalant misuse is thought to have direct effects on the heart, increasing myocardial cell depolarization and reducing conductivity, which predisposes the individual to arrhythmias, especially when paired with increased blood concentrations of epinephrine. This has sometimes been termed *sudden sniffing death or sudden sniffer’s death*.

Within the United States, there are several sources available to help in determining the prevalence of mortalities associated with inhalants. First, medical examiner reports are available from numerous jurisdictions and metropolitan areas and some states. Despite this, only four statewide investigations (e.g., North Carolina, Texas, Virginia, Washington) of inhalant-related deaths have been conducted in the United States, and some of these data are a decade or older now. A review of 39 inhalant-related fatalities in Virginia between 1987 and 1996 found that deaths were most frequently caused by gas fuels including butane and propane (46%), chlorofluorocarbons (26%), and chlorinated hydrocarbons and alkylbenzenes (21%). The fatalities were mostly male (95%) and Caucasian (White) (95%), with 70% of the deceased being 22 years old or younger (median age = 19). An investigation of 144 inhalant-related deaths in Texas between 1988 and 1998 similarly found that most decedents were male (92%), Caucasian (White) (81%), and relatively young (mean age = 25.6 years). In 35% of cases, the cause of death was inhalation of Freon, followed in frequency by inhalation of chlorinated hydrocarbons (implicated in 25% of deaths), and alkylbenzenes (e.g., toluene) identified in 17% of deaths. A review of 56 inhalant-related deaths in Washington state between 2003 and 2012 revealed that most fatalities were male (88%), occurred in young adults (age 20 or older), and were Caucasian (White) (93%). While 12 different chemicals were associated with the deaths, difluoroethane was reported on 30 death certificates (54%), and gas duster products were named as the difluoroethane source for 12 deaths (no source was reported for the remaining 18 difluoroethane deaths). An examination of 30 inhalant-related deaths in North Carolina between 2000 and 2008 reported that most deaths were males (81%) and Caucasian (White) (93%). However, decedents were somewhat older compared to earlier studies (mean and median ages were 31.2 and 28.0, respectively). Deaths due to “air duster” misuse (i.e., 1,1-difluoroethane) accounted for nearly one-third of all cases, and this compound was the second most identified inhalant following toluene. Finally, the Florida State Medical Examiners Commission identified 54 inhalant-related deaths in 2018 and reported that 81% of these cases involved halogenated agents like 1,1-difluoroethane. In 2017, 68 persons in Florida were identified as having inhalant-related deaths, 54 due to 1,1-difluoroethane use. Taken together, these medical examiner reports suggest that within the United States, adult Caucasian (White) males are at the greatest risk for inhalant-related fatalities. The North Carolina study cited above also noted that most decedents were heavy polydrug users with significant health and mental health impairments.

A second source of information on the prevalence of mortalities associated with inhalants is available from calls to national poison control centers within the United States. The most comprehensive investigation examined all inhalant-related reports made to 60 poison control centers between 1983 and 2008. The highest rate of cases occurred among youth aged 12 to 17 years, peaking at age 14 with nearly three-quarters of all cases being boys. The 35,453 inhalant cases reported involved more than 3,400 different products. The most frequently identified inhalants were propellants, gasoline, and paint. Propellant-related cases (e.g., “air dusters”) increased significantly over the study period, particularly since 2003, and constituted the largest-single category of inhalant-related cases. The three inhalants most associated with fatal outcomes were butane, propane, and air fresheners.

In the United States, there are still no national estimates of inhalant-related mortality, and the available data likely contain numerous biases. For example, in many instances, poison control centers may never be contacted regarding an inhalant-related death, and the sometimes nonspecific findings following an inhalant-related death may make proper identification of the cause of death difficult.

In the United Kingdom, a systematic study of inhalant-related mortality (i.e., 2,093 deaths) between 1983 and 2007 concluded that the number of deaths during this interval had declined substantially, particularly among women, and that the mean age at death had increased by more than a decade in males and females. Particularly notable decreases in inhalant-related deaths were observed for children following a 1992 federal advertising campaign that warned parents that their children might be misusing inhalants. The number of inhalant-related deaths peaked in the early 1990s and declined to their lowest in the mid-2000s. In 2005, 45 inhalant-related deaths (9 of which were suicides) were reported. This was the lowest number in the prior 20 years, and this reduction may relate to a parent-focused advertising campaign and/or a ban on the sale of cigarette lighter refills to those under the age of 18. The percentage of glue-related inhalant deaths declined during this 25-year interval, whereas the percentage of deaths due to gas fuels (e.g., butane and propane) increased significantly. In the United Kingdom, butane was implicated in 80% of such deaths. In 4 of 10 of these deaths, there was no evidence of previous inhalant use, and most deaths occurred at home. Inhalant-related deaths were predominantly male. Among 10 to 14 year olds, there were more inhalant-related deaths than deaths associated with other drug misuse (as defined by the Drug Misuse Act of 1971, including cocaine, cannabis, LSD, and opioids, among others).

A 15-year-old Caucasian (White) boy was admitted to an adolescent day treatment program. His biologic father was identified as a drug dealer, was abusive to his biologic mother, and went to prison for robbery; his parents separated when he was only a few months old, and the patient did not recall ever meeting his father. His mother remarried when he was 12. Reportedly, his stepfather would "get smashed," hit his mother, get very close to the patient's face, and "yell at [him] for no reason." His mother was in the process of divorcing his stepfather. At age 10, the patient began using marijuana, reporting that it helped him "get away" and reduce his anger, and his use rapidly progressed to using "all day every day." He then tried alcohol, cocaine, and prescription medications (i.e., benzodiazepines and opioids).

The patient reported a history of delinquent behaviors. These behaviors included stealing from his mother's purse, stealing her prescription medicines, pawning his stepfather's possessions, truancy, and initiating fights. In sixth grade, he received a drug paraphernalia ticket from the police, although this charge was later reduced to minor in possession of alcohol. He was later charged with accessory to aggravated robbery. The patient was placed on probation, and because of his history of substance involvement, was admitted to a short-term outpatient, mainly group-based treatment program. He was discharged from that program after a few sessions for an inability to remain drug free and suspected use of his mother's prescription pain medication.

After serving 3 weeks in a juvenile detention center, he was admitted to a day-treatment program. A thorough intake evaluation revealed a history of cannabis, cocaine, alcohol, and prescription drug use disorders, no other mental health concerns, and no relevant medical issues. The patient did well in treatment, attending regularly, receiving above-average scores for his behavior, completing his weekly school contracts and all his initial treatment-related assignments, submitting three times a week to observed urine samples (tested for alcohol, opiates, benzodiazepines, PCP, cocaine, cannabis, and amphetamines), and attending individual and family therapy. Reports from all observers indicated the patient was making progress in treatment. The patient expressed pride in being drug free and staying away from negative peers and began discussing future-oriented prosocial goals. After 1 month in treatment, the patient was progressed to higher privileges, recognizing his treatment progress.

The following week, the client was found by his mother submerged in his bathtub. He had been "huffing" (i.e., inhaling) a common spray used to remove dust from computers (i.e., "air duster," containing 1,1-difluoroethane) and other surfaces. Although attempts were made to revive him, he later died. This case demonstrates several important points about inhalant use: an association with other substance use and conduct disorder, adolescent onset, and the risk of inhalant-related mortality.

Comorbidity

Inhalant users have a high prevalence of multiple comorbid issues/diagnoses. For example, in nationally representative studies of the U.S. population, respondents reporting any inhalant use were significantly more likely to have injected drugs; injection, in turn, is a risk factor for infection with human immunodeficiency virus (HIV) and hepatitis B and C viruses (HBV/HCV). Inhalant use, especially among persons also reporting cannabis use, is a marker for the elevated risk of other SUDs, and inhalant users are more likely to report delinquent behaviors and meet criteria for conduct disorder than inhalant nonusers. In one nationally representative study of adults, for those reporting inhalant use before age 18, more than 40% met criteria for antisocial personality disorder. In that same study, inhalant users showed high rates of mood (48%) and anxiety (36%) disorders, especially among women. A study of 664 nationally representative adult inhalant users found that 40% met criteria for 3 to 5 DSM-IV-TR SUDs, whereas nearly one-quarter (i.e., 22%) met criteria for 6 to 11 such disorders. Virtually all (94%) adult inhalant users met DSM-IV-TR criteria for an alcohol and/or noninhalant drug use disorder. Studies of nationally representative and juvenile justice samples support a link between inhalant use/IUDs and suicidality. Several studies also support an association between inhalant use and a history of child abuse and/or neglect victimization. Although multiple explanations are possible, the etiology underlying this comorbidity has not been well studied; exploratory analyses suggest that the link between early inhalant use and later injection drug use may be explained by a general predisposition to early-onset drug use. In addition to further exploring this link, newer studies should investigate using the DSM-5 criteria as currently little information exists about this relevant, and not insignificant, clinical population, even though the DSM-5 was published/updated almost a decade ago.

Summary

Trying inhalants at least once is comparatively common around puberty and into adolescence, but continued frequent inhalant use is uncommon, and the prevalence of at least monthly use decreases with increasing age. Discontinuation rates by young adulthood are higher for the inhalants than for any other psychoactive drug class. Accordingly, the limited data suggest that IUD is rare in the general population. Related to this, inhalant-related problems are the primary contributor to relatively few addiction treatment admissions. Inhalant-related problems account for a small proportion of drug-related ED contacts and are only implicated in a small proportion of all drug-related deaths. One warning, however, is that elevated rates of IUDs are often characteristic of incarcerated and institutionalized youth and adult populations, and these populations are typically not reflected in epidemiologic studies. Similarly, homeless youth and adults, persons living on the street or transiting between different geographic areas, and other marginalized and "hidden" populations, may show higher levels of inhalant use and IUD than are reflected in epidemiologic studies of general and clinical populations. Inhalant users also report numerous barriers to treatment utilization and are commonly coping with multiple comorbid psychiatric disorders and SUDs and impoverished social circumstances that may make it difficult for them to access health care.

The available data suggest that most adolescents who survive a brief experimentation with inhalants and who do not have conduct disorder soon abstain from inhalants and avoid severe pathologic outcomes. However, among adolescents with conduct disorder, perhaps half develop adult antisocial personality disorder. Inhalant problems in these youths often herald serious alcohol and polysubstance use in adulthood, and a small percentage of these youths become chronic, deteriorated, inhalant-dependent adults.

PHARMACOLOGY AND TOXICOLOGY

National surveys regularly report that the inhalants most often misused by U.S. citizens are toluene-containing products (e.g., glues and thinners), gasoline, lighter fluid (which usually contains butane), spray paints, aerosols, degreasers, and correction fluids (see Table 11.7–2). Self-administration of these inhalants results in rapid uptake in the lungs, saturation of the blood, and distribution to the brain. The paraphernalia needed for use is easily attainable such as a bag, rag, worn clothing, hands/nails, cotton balls, inverted aerosols, balloons, or just inhaling directly from the container itself. If using a rag or clothing, these items are sprayed or soaked with the substance and then placed over the mouth and/or nose to inhale the vapors. The act of inhaling these vapors has been termed as "huffing" or "to huff." If using a bag or container, sometimes referred to as "bagging," the mouth is placed at the opening of the bag to inhale the fumes. More dangerously, the bag could be placed over the head to inhale, which can easily cause an accidental suffocation. Finally, the substance could also be directly administered via a spray into the mouth or nose, referred to as "dusting," although a towel or other materials may be used to limit the direct damage to lungs.

This misuse typically involves 15 to 20 deep breaths, which results in rapid intoxication. Once intoxication is achieved, behavioral effects have been reported to persist from 15 to 60 minutes and can be sustained for hours by either perpetual or repeated bouts of inhalation. Rebreathing the exhaled air may enhance the effect if the inhalant is exhaled into a bag. While we cannot say with certainty the exact solvent concentration inhaled for various compounds, estimates for one of the most misused inhalants, toluene, vary from 5,000 to 15,000 ppm. This range is several orders of magnitude above the industrial maximum of 100 ppm time-weighted averaged for an 8-hour workday. Even these concentrations for those durations can produce minor temporary performance decrements. Tolerance to the intoxicating effects appears to develop, with individuals increasing the number of tubes of glue they consume over repeated days of use.

Table 11.7–2. Pharmacology and Putative Mechanism of Action (MOA) of Commonly Misused Inhalants

Group	Putative MOA	Example Structure	Physiologic Effects
Alkyl nitrites	Vasodilation, nitric oxide production	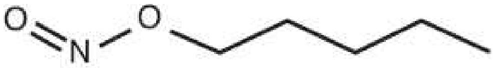 Amyl nitrite	Vasodilation, smooth muscle relaxation
Nitrous oxide	B2 and NMDA antagonist	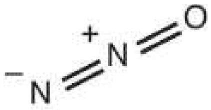 Nitrous oxide	CNS depression, hilarity, analgesia, anesthesia
Other anesthetics	NMDA blockade, GABAergic PAM	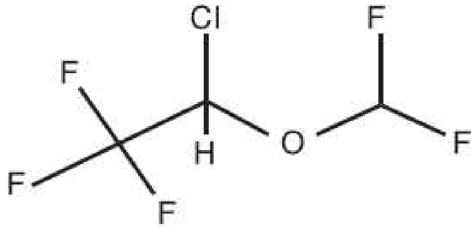 Isoflurane	Relaxation, numbness analgesia, anesthesia, respiratory depression
Propellants–fuels	Unknown	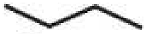 Butane	Relaxation, hallucinations, illusions
Solvents	NMDA antagonist, GABAergic PAM	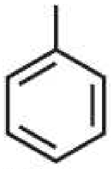 Toluene	Initial disinhibition followed by a more prolonged depression. Illusions, hallucinations.

Structure images were obtained from National Center for Biotechnology Information. PubChem Compound Summary for CID 10026, Amyl nitrite. https://pubchem.ncbi.nlm.nih.gov/compound/Amyl-nitrite; PubChem Compound Summary for CID 948, Nitrous oxide. https://pubchem.ncbi.nlm.nih.gov/compound/Nitrous-Oxide; PubChem Compound Summary for CID 3763, Isoflurane. https://pubchem.ncbi.nlm.nih.gov/compound/Isoflurane; PubChem Compound Summary for CID 7843, Butane. https://pubchem.ncbi.nlm.nih.gov/compound/Butane; PubChem Compound Summary for CID 1140, Toluene. https://pubchem.ncbi.nlm.nih.gov/compound/Toluene. Accessed March 22, 2023.

Toxicokinetic studies for inhalants like toluene show that inhaling these compounds results in very rapid absorption through the lungs that produces a rapid rise in blood concentrations. Toluene and other inhalants can also be absorbed through the skin and the GI tract, which increases total body burden. Both animal and human studies have shown a positive correlation between toluene levels in alveolar air and blood levels. In addition, a recent *in vitro* study showed that certain aromatic hydrocarbons, including toluene, may change the surfactibility of alveolar spaces, which may correspond to the vapor pressure of each solvent when suspended in air. This finding may suggest that not only do solvents impact lung function, but they also exert physiologic changes in the lung that correspond to the solvent's vapor pressure. Toluene is absorbed by both red blood cells, where it binds to the heme protein, and plasma, which enhances toluene transportation to all areas of the body. Given that toluene is a relatively nonpolar, lipophilic compound, it is found concentrated more in tissues that are high in lipids and well vascularized. Due to these properties, as well as toluene being a small molecule, it readily crosses the blood–brain barrier and is found in higher concentrations in myelinated areas. One study in which humans were exposed continuously to 80 ppm of toluene for a 4-hour duration found that toluene blood levels peaked (6 to 7 μmol/L) at approximately 2 hours. Autopsy data from a 16-year-old adolescent boy who died after inhaling glue show that blood and brain toluene concentrations were reported at 20.6 μg/mL and 297 μg/mg, respectively. In another case, a man died after collapsing while painting and being inadvertently exposed to toluene. Toluene levels in his blood and brain were reported at 48 and 80 μg/g, respectively. When consumed with ethanol, toluene blood levels can increase significantly, which can dramatically increase levels of behavioral and physiologic toxicity.

Autoradiography studies have seen elevated levels of radioactivity in body fat and the central nervous system (CNS), including the spinal cord and white matter of the brain, in mice exposed to toluene inhalation. Both toluene and its metabolites leave the body relatively quickly; however, toluene has been shown to remain in adipose tissue for up to a few days. Both human and rodent studies have reported that approximately 80% of toluene is absorbed by the body when inhaled, with increased physical activity increasing the amount of toluene absorbed. Toluene that is not absorbed is later expelled unchanged in breath and urine, with the liver metabolizing the toluene that is absorbed first into benzyl alcohol and is excreted from the body primarily as hippuric acid in urine. While various demographic, genetic, and physiologic factors influence the rate of toluene elimination from the body, toluene concentrations in breath begin decreasing within minutes of exposure, while concentrations in the blood reduce at a much slower rate, taking anywhere from 4 to 10 hours before being undetectable. Elimination of hippuric acid in urine takes longer, and anything greater than a 1-g hippurate per gram creatinine ratio has been suggested as a marker for inhalant misuse involving toluene. However, occupational exposure to toluene and some food preparations may also lead to hippurate/creatinine ratios greater than 1.

Understanding of the molecular mechanisms and targets of inhalants has improved within recent decades because of the increase of *in vitro*/*in vivo* studies of these compounds. Unfortunately, the number of inhalants investigated remains scarce, with toluene receiving the most attention. Toluene, like all other addictive substances, alters dopamine levels in the brain. Several studies have shown that toluene exposure either increases or decreases dopamine in the striatum, which may depend on concentration and duration of use. In addition, transient increases in neuronal firing activity have been observed within the ventral tegmental area (VTA). Toluene exposure also produces marked increases in dopamine levels within the prefrontal cortex and other areas related to reward circuitry. Repeated toluene exposure increases the amount of dopamine within the caudate and nucleus accumbens. Accumbal dopaminergic activity may be modulated by activity within the medial prefrontal cortex (mPFC).

While the neurochemical mechanisms responsible for the increased dopaminergic activity are unclear, evidence from studies conducted on midbrain dopamine neurons suggests that dopaminergic alterations may be dependent on *N*-methyl-ᴅ-aspartate (NMDA) activations. In line with this sentiment, both the GABAergic and glutamatergic systems have been explored extensively. Toluene influences both ligand-gated and voltage-gated ion channels at concentrations similar to ranges that are inhaled during episodes of voluntary exposure. Although toluene has a simple molecular structure (i.e., a monosubstituted benzene ring with a methyl group), it appears to be rather selective in its neurochemical actions. Studies have proven that toluene will concentration-dependently inhibit the *N*-methyl-ᴅ-aspartate receptor (NMDAR), a specific type of ionotropic glutamate receptor associated with excitation. Moreover, toluene is thought to be most selective for the GluN2B receptor subtype. Conversely, while toluene's putative mechanism of action (MOA) is often recognized as inhibition of the NMDA receptor, it appears to have the curious effect of increasing glutamatergic activity in neurons in the mPFC. Toluene has also been shown to affect non-NMDA glutamatergic sites. For example, toluene has been proven to increase the amplitude of excitatory postsynaptic currents in α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors (AMPARs), a non-NMDA ionotropic glutamate receptor, which suggests an increase in AMPAR expression on the postsynaptic neuron. Increasing NMDA or glycine concentrations does not overcome toluene's inhibitory effects, suggesting that toluene's effects at this receptor site are noncompetitive. Repeated exposure to toluene reportedly produces an upregulation of NMDARs within certain areas of the brain. Other known inhalants, such as 1,1,1-TCE and several alkylbenzenes (i.e., m-xylene, ethylbenzene, and propylbenzene), have been reported to inhibit NMDA-induced currents in both a concentration- and subunit-dependent manner. Thus, at the acute level, NMDAR inhibition seems to be a common MOA for several misused inhalants, with continued exposure altering the expression of these glutamatergic receptor subtypes.

Numerous *in vitro* reports have shown that at the acute level, toluene, TCE, and trichloroethylene increase α1 $GABA_A$ and α1 glycine receptor function via positive allosteric modulation. Conversely, continuous toluene use inhibits spontaneous GABA currents in hippocampal neurons, as well as increasing $GABA_A$ α1 subunit levels in the mPFC while decreasing them in the substantia nigra. This differential GABAergic activity is suggestive of the development of tolerance to the enhancement of $GABA_A$ receptor function following acute toluene inhalation. In contrast, flurothyl, a nonmisused proconvulsant solvent, inhibits $GABA_A$ receptors and acts as a noncompetitive antagonist at $GABA_A$ receptors with no effect on α1 glycine receptors. These results suggest that acute behavioral inhibitory or excitatory effects of inhalants associate well with *in vitro* $GABA_A$ receptor activation or inhibition.

Toluene's receptor selectivity has also been reported for nicotinic acetylcholine receptors (nAChRs), where toluene inhibits α4β2 receptors to a greater extent than α7 nAchRs. Like other sites of action shared among inhalants, nAChRs may serve as another potential shared target. Another widely misused solvent, perchloroethylene, has been shown to be even more potent than toluene in inhibiting the types of nAChRs. However, it is unclear whether inhibition of nAChRs is indeed a common MOA for other misused inhalants.

Toluene also alters P2X adenosine triphosphate (ATP)-gated channel activity. Toluene promotes currents through P2X3 receptors but decreases the function of P2X2 and P2X4 receptors within the ATP-gated channels. Finally, several *in vitro* studies have also shown that toluene, TCE, and trichloroethylene differentially affect 5-hydroxytyptamine (5-HT) receptor function, with the $5\text{-}HT_3$ receptor subtype being a common target. Taken as a whole, findings from these reports indicate that misused inhalants have distinct sites of action on individual channel subunits, and the differences in regional and anatomical subunit expression appear to be important determinants of inhalant effects.

While the behavioral effects of inhalants vary by magnitude and duration depending on the inhalant in question, research in animals examining the effects of exposure to levels of inhalants suggests that many inhalants have similar pharmacologic profiles to CNS depressants like ethanol, barbiturates, and other depressants. At concentrations comparable to those seen in misuse situations, behavioral effects transition from motor excitation to motor dysfunction, ataxia, loss of righting reflex, and finally narcosis or anesthesia. Continuous exposure to higher concentrations of inhalants may result in coma, seizures, and ultimately death by respiratory depression. Many inhalants appear to have depressant or inhibitory effects by nature. However, not all misused inhalants show these characteristics. For instance, despite certain alkanes showing CNS depressant properties, several inhaled isoparaffins, branched alkane compounds, have been shown to produce concentration-dependent increases in locomotor activity, and isoamyl nitrite has been shown to produce concentration-dependent changes with lower doses showing increased locomotor activity and higher doses showing no effect, while other inhalants, such as ethyl and butyl-acetate, produce the canonical decrease in locomotor activity. While toluene and other related compounds are depressants by nature, they yield intriguing findings when paired with other inhalants. For example, repeated exposure to toluene and TCE produces sensitization to the motor excitation effects of these inhalants, and repeated toluene exposure interacts with cocaine to increase cocaine's motor effects, suggesting a similar neuropharmacologic MOA.

While clinical studies seldom report seizures from inhalant misuse, rodent models have demonstrated both convulsant and anticonvulsant effects as a consequence of misuse pattern exposures. Sustained or repeated exposure to certain inhalants yields handling-induced convulsions in mice upon cessation of exposure. In one study, mice repeatedly exposed to TCE over multiple days developed handling-induced seizures, as well as an increased sensitivity to pentylenetetrazol-induced convulsions upon cessation of TCE exposure. When these mice were exposed to TCE or toluene vapors, or other depressants (i.e., ethanol, midazolam, or pentobarbital), the number and intensity of seizures decreased. Hence, physical dependence develops to inhalants, and they display cross-tolerance with other inhalants as well as CNS depressants, which may suggest a common mechanism.

Other volatile organic compounds, like flurothyl, a $GABA_A$ antagonist, appear to produce proconvulsant activity. While the mechanisms of these convulsant actions are mostly unknown, high concentrations of any inhalant might have nonspecific and global effects on CNS reactivity. Conversely, the anticonvulsant effects of several inhalants at low concentrations have also been reported. When given as a pretreatment, a single toluene injection has been shown to protect against convulsions brought on by several chemical agents including bicuculline, nicotine, NMDA, and picrotoxin but not strychnine. Protective effects have also been demonstrated for inhaled toluene against NMDA-induced seizures. These anticonvulsant effects may be associated with toluene's potentiating effects on $GABA_A$ receptors or its ability to inhibit NMDA and nicotine receptors, and/or to increase dopamine release. It should, however, be noted that the concentration needed to yield these effects, and the immediate dangers associated with inhalant use, may preclude inhalants, such as toluene, from having any recognized medicinal value as an anticonvulsant. While the authors speculate the MOA of this convulsant and anticonvulsant activity to involve NMDA, GABA, or both, these findings require further mechanistic exploration to better parse out the role of each transmitter system. Given that seizure episodes involve excitation of the CNS, often because of glutamatergic or GABAergic activity, it follows that these targets may be responsible for the convulsant and anticonvulsant effects of inhalants. Further, the differential activity observed may be explained in terms of dependence and rebound activity that occurs after binge-pattern use of inhalants. In other words, repeated use followed by cessation of use may result in a rebounding spike in glutamatergic activity or, on the other hand, a sharp decrease in GABAergic activity, thereby causing convulsions in the user.

Because many inhalants act as positive allosteric modulators at the $GABA_A$ receptor site, it is no surprise that animal research has also proven that acute and repeated inhalant exposure produces anxiolytic, narcotic, and sensory effects, as well as alterations in learning and memory. In the elevated plus maze (EPM), toluene and TCE increased the number of entries and time spent in the open arms, comparable to diazepam, which suggests that these two inhalants may have anxiolytic-like effects. In addition, abstinence following toluene exposure has been shown to precipitate a rebound in anxiety-like behavior, and re-exposure to toluene reduces this increase in preclinical models. This finding is noteworthy as anxiolytic drugs characteristically show rebound increases in anxiety upon cessation of continuous use. Further support can be obtained from the defensive burying paradigm, where toluene and other inhalants attenuated the latency to bury an increasingly aversive stimulus. Conversely, flurothyl may have anxiogenic-like effects in that it attenuated both the number of open-arm entries and the time spent in open arms in the EPM paradigm, while apparently having no effect in the defensive burying paradigm at concentrations up to 800 ppm, which is near the seizure threshold for this vapor. Whether other inhalants share this anxiolytic property with toluene and whether tolerance or sensitization occurs with continuous use remains to be determined, albeit there may be challenges with assessing this effect in a preclinical model.

The reinforcing properties of drugs, like inhalants, have been repeatedly demonstrated in animal models. In one study, animals were trained to self-administer toluene and TCE intravenously. In addition to self-administration, many conditioned place preference (CPP) studies with toluene have shown that animals will spend more time on the toluene-paired side of a chamber. Comparable findings have been found for mixtures of other solvents, such as toluene, benzene, ethyl acetate, and methyl chloride. Interestingly, one recent study demonstrated that γ-vinyl GABA, a selective GABA-transaminase inhibitor, blocks the previously established CPP with toluene. Other studies have expanded on these rewarding effects by showing that animals trained to respond on one lever after a saline injection and on another lever after a CNS depressant (ethanol or pentobarbital injection) will respond on the lever paired with the depressant, implying that the subjective effect of toluene may be depressant like. Other animal studies using this assay have shown that toluene produces partial substitution for PCP and amphetamine, suggesting that inhalants may share subjective effects with a wider range of drugs than was previously expected.

In addition to effecting behavioral changes via pharmacologic mechanisms in the mesocorticolimbic system, short- and long-term exposure to high concentrations of volatile organic compounds has been shown to produce intracellular oxidative stress and lipid peroxidation in various organs throughout the body, such as the brain and liver. Ordinarily, a balanced level of redox behavior (i.e., oxidation and reduction) is a sign of healthy cellular function, but a shift in the balance between oxidation and reduction may produce cytotoxic effects and pathologic alterations in various organs throughout the body. Both oxidative stress and lipid peroxidation involve reactive oxygen species (ROS) taking electrons from various components within the cell. For example, lipid peroxidation yields alterations in the phospholipid bilayer of the cell membrane by taking electrons from lipid proteins. This pathway may serve as one potential source of toluene’s cytotoxic effects. Oxidative stress may serve as a potential pathway through which toluene precipitates long-term cellular and tissue damage and pathogenesis. When compared to its acute pharmacologic effects, toluene’s long-term pathogenic effects are not as well documented.

Toluene-induced oxidative stress has been linked to DNA fragmentation and apoptosis. While the literature suggests toluene is altering intracellular levels of antioxidants, such as glutathione peroxidase (GPx), glutathione (GSH), and superoxide dismutase (SOD), the particular pathway is still unclear for, perhaps, many reasons: (1) we cannot say to what extent which ROS are generated by toluene, which may be owing to the transient nature of ROS, thereby making them difficult to observe and measure; (2) toluene very clearly produces apoptosis and DNA fragmentation, but is this the direct consequence of elevated intracellular oxidative stress and lipid peroxidation or a separate mechanism?; (3) like ethanol, toluene is metabolized into a reactive aldehyde (benzaldehyde), which, in a manner similar to ethanol-induced acetaldehyde production, may produce higher levels of oxidative stress than toluene itself; and (4) which compartments in the cell appear to be the greatest source of toluene-induced oxidative stress? To the latter point, toluene is metabolized at various stages in different compartments of the cell. While the particular pathway and components of toluene-induced oxidative stress have yet to be fully elucidated, the authors surmise that the pathologic damage seen in various organs throughout the body is owing to the inhalant’s impact on oxidative stress.

Organ Pathology

Inhalant use greatly increases the potential for injury in mild to severe cases and, in some instances, death. Given that several inhalants have MOAs comparable to those of other drugs involving modulation of GABA and/or NMDA, higher concentrations of the inhalant may result in moderate to severe CNS depression. These effects may be amplified when used in combination with other CNS depressants such as ethanol. In addition to CNS effects, many inhalants have been shown to sensitize the heart to endogenous catecholamines, such that any abrupt physical effort or unanticipated excitement may lead to ventricular arrhythmias and even death. The commonly employed practice of putting inhalants into plastic bags and then placing the bag over the user's head may result in unconsciousness and, potentially, death from suffocation. The user may die from aspiration of either the compound or vomit. Moreover, fatalities have also been reported as a direct result of the inhalant. These include anoxia, asphyxiation, severe injury to the lungs, or cardiac dysfunction. Gas fuels such as butane have been shown to produce laryngeal edema as well as laryngospasm. The severe behavioral and physical deficits associated with inhalant intoxication have resulted in physical trauma (e.g., falling and car accidents) that resulted in fatalities. On a similar note, death has also been reported from fire-related injuries from inhalant combustion, as some inhaled compounds are heated before inhalation or a cigarette was used near where the inhalant was being misused. In addition to the acute and chronic toxicologic effects precipitated from an inhalant's pharmacologic actions in the body, this class of substances has also been shown to produce a host of pathogenic effects in various organs throughout the body.

Chronic inhalant use has been linked to renal and hepatic failure, as well as liver tumors, which can also lead to death. These effects may be the consequence of long-term damage associated with lipid peroxidation and oxidative stress, given that both phenomena are linked to a host of pathologic effects, such as apoptosis, ferroptosis, and mitochondrial dysfunction. In line with this thought, one study found that male Wistar rats exposed to 1,500 ppm of toluene for 4 hours over the course of 7 days showed higher levels of oxidative stress and apoptosis in liver and kidney cells, which may suggest that repeated exposure to toluene produces distortions in cellular redox signaling and, eventually, cell death. Further, given the bioavailability of toluene in lipid-rich regions of the body, it may be that the greater the lipid content, the more susceptible the organ is to insult. Inhalants, such as toluene, have been shown to accumulate in higher concentrations in the liver and kidneys. For example, a pharmacokinetic study demonstrated higher levels of toluene accumulation in the liver, kidney, lung, and fat than other areas, such as the heart and spleen, 30 minutes postinjection in a rodent model. While the particular cause of inhalant-induced tissue damage has not been shown in humans, the combination of preclinical models, sequalae in long-term users, and the postmortem presence of tissue damage in lipid-rich areas may, again, suggest that cytotoxicity and tissue damage are the consequence of oxidative stress and lipid peroxidation. While identifying a specific mechanism poses a challenge, it is certain that it is the inhalant that is causing the damage in humans, given what has been observed in clinical literature.

Long-term toluene misuse has been shown to produce ototoxic effects, such as tinnitus and hearing loss. Inhalant misuse has been shown to produce damage to the cardiovascular system, as well as cardiovascular sensitivity to endogenous catecholamines. A few studies have reported renal failure with inhalant misuse because of tubular necrosis. In addition, liver damage, from centrilobular tumors and necrosis, has also been documented. However, it is worth noting that reports of irreversible damage and toxicologic effects in certain organs, such as the liver and kidney, have been inconsistent. For example, some studies report changes in serum hepatic enzymes, whereas others report no changes. Prolonged use of inhalants such as isobutyl nitrite and naphthalene can impede bone marrow production, which can lead to anemia and leukemia. Repeated inhalation of isopropyl nitrite or toluene can produce visual sequalae, such as nystagmus and vision loss. Pulmonary symptoms have also been documented with a plethora of inhalants resulting in chronic coughing, chest pain, shortness of breath, wheezing, and pneumonia. Additional harmful effects include rhabdomyolysis (kidney damage from muscle degeneration) and GI perturbations, such as nausea, pain, and vomiting. The specificity of these effects remains the subject of contention. However, the brain has been shown to be particularly vulnerable to the toxicologic effects at various concentrations, ranging from low to high.

Chronic inhalant misuse can have profound consequences for nervous system functionality, and, especially, the brain, which is high in lipid content. In addition, longer durations and larger magnitudes of inhalant misuse are positively correlated with neuroimaging and neurobehavioral abnormalities. For example, one magnetic resonance imaging (MRI) study showed that young adult chronic users display neurologic abnormalities, which positively correlated with higher ratings on the Halstead–Reitan Battery, an indicator of neuropsychological impairment. This finding suggests that toluene-induced psychological impairment may correspond to increases in white matter abnormalities. Significant cognitive and neurologic deficits have been observed in these individuals and include personality changes, hallucinations, paranoid psychosis, memory deficits, and deficits in information processing, decreased cognitive performance, concentration deficits, speech impairments, impaired judgment, problems with impulsivity, abnormal electroencephalogram (EEG) activity, temporal lobe epilepsy, convulsions, tremors, as well as ambulatory abnormalities. Several imaging and neuropathologic techniques have shown brain abnormalities in those who use inhalants, such as demyelination, shrinkage of brain tissue, and white matter damage (leukoencephalopathy). MRI has shown significant atrophy of cerebral, cerebellar, and brain stem tissue, diffuse CNS demyelination, widening of sulci, and ventricular enlargement in long-term users of toluene. As previously suggested, the greatest observations of atrophy were in the lipid-rich white matter areas of the brain, with several cases displaying a loss of differentiation between white matter and grey matter areas. Additional reports using T2-weighted MRI showed that persistent toluene misuse leads to demyelination, as well as significant axonal damage and loss. Numerous rodent studies have corroborated these findings with chronic exposure to higher concentrations of toluene producing deterioration of the frontal cortex, the hippocampus, and brainstem in addition to impairments to dendritic outgrowth and white matter damage. Repeated long-term toluene exposure has been shown to produce neuronal damage to the hippocampus. Finally, several single-photon emission computed tomography (SPECT) studies have shown that chronic inhalant misusers show significant decreases in regional blood flow to the bilateral prefrontal cortices as well as perfusion irregularities in other brain areas. Additional investigations of individuals who have sustained long-term exposure to occupational levels of toluene have shown abnormal cerebral blood flow, as well as atrophy to a number of brain areas.

Given a dearth in the literature, it is challenging to determine whether individuals can reverse neurophysiologic damage and restore cognitive functions after chronic inhalant misuse. As previously described, many of the neuroimaging studies would suggest that any neurobehavioral deficits observed in these individuals would be irreversible. Indeed, several investigations examining former long-term users of inhalants failed to detect any notable improvements in MRI measures of brain pathology or neuropsychological functioning after 5 to 18 months of abstinence. The lack of improvement reported in these studies may be due, in part, to the timing of measurement, because a few studies have reported that neurobehavioral measurements taken 2 years after users discontinued petrol sniffing were significantly improved compared to earlier measurements. While full recovery was not reached, those individuals with the most impairment showed the greatest improvements. Moreover, a small section of the literature has shown that prevention of damage may be achieved through adjunctive therapy with various pharmaceutical agents, such as minocycline (an NF-KB inhibitor); however, this has not been established in the clinical literature.

Maternal Use and Fetal Solvent Syndrome

With prevalence rates for inhalant use proving to be nearly equivalent for adolescent males and females, prenatal solvent exposure poses a unique challenge for childbearing females. Indeed, a recent review of drug misuse among females of reproductive age featured toluene-based solvents in addition to other commonly used psychoactive substances during pregnancy. An increasing number of cases of fetal developmental disorders and malformations in the offspring of females who misuse inhalants during pregnancy has been observed in the medical literature. Many of these clinical embryopathy and teratology reports argue that these infants display symptoms of a *fetal solvent syndrome (FSS),* so called because of the similarities to infants gestationally exposed to ethanol who showed signs of *Fetal Alcohol Spectrum Disorder (FASD)*. Females who misuse inhalants during their pregnancies gave birth prematurely to infants with developmental delays. Further, some of these infants displayed physical abnormalities such as a reduction in head size, hypoplastic midfaces and jaws, lower set ears, shortened palpebral fissures, thin upper lips, and deep-set eyes. Additional anomalies reported were downturned corners of the mouth, irregular hair patterns on the scalp, small fingernails with blunted fingertips, abnormal palmer creases, and urinary tract deformities. In certain cases, fetal death was reported in women with histories of inhalant misuse where maternal exposure to the inhalant was extremely high during pregnancy. Follow-up investigations with these FSS infants showed continued developmental delays and deficits such as moderate growth retardation, cerebellar dysfunction, hyperactivity, language impairments, and elevated levels of aggression. It is unclear whether these delays and deficits last into adulthood, as there are no documented clinical therapeutics or remedies for the observed effects of FSS exposure, nor are there follow-up studies of these affected children as they grow into adults. The accompanying stigmatization and discrimination associated with drug use during pregnancy may prevent many inhalant-misusing women from properly accessing needed health care services during pregnancy. To further support findings in human populations, preclinical studies that mimic the high-dose binge-type inhalant use pattern during pregnancy provide additional evidence of an FSS. These animal investigations have reported that the high toluene levels that are observed in maternal tissues and organs are also found in the placenta and fetal tissue, including the brain. These high concentrations of toluene in animals produce a myriad of adverse outcomes, such as physical malformations, growth retardation, delayed ossification of skeletal structures, deficits in coordination and forelimb strength, impairment in memory recall tasks, and changes in amphetamine-induced locomotion. Whether these delays and deficits last into adulthood or whether it is possible to prevent or mitigate these delays and deficits is still unclear.

Patterns of Inhalant Use

Qualitative research has shown that inhalant users report three main use scenarios: (1) using on rare occasions to have fun, (2) using because alternative substance(s) could not be used (because of no access or urine drug monitoring), and (3) using in day-long binges. Recent efforts have been made to classify different subtypes of inhalant users. In one study, some inhalant users used many different inhalants, had elevated levels of psychopathology (i.e., depression, anxiety, externalizing behaviors, and substance use), suffered serious family problems, had high levels of affective instability, and used inhalants to help cope with those feelings or problems. Other inhalant users had decreased levels of these problems, many of those individuals had only used one type of inhalant, and these individuals were hypothesized to be “experimenters.” A third group used inhalants because they were easy to obtain. Another taxonomic investigation identified three subtypes of adolescent inhalant users based on the social contexts in which they used inhalants. Youth in the class that most tended to use inhalants with friends to celebrate and have a good time and who had difficulties in refusing invitations to engage in inhalant use, or who socialized with groups of youth who used inhalants, had the highest levels of substance use and related problems, psychiatric distress, and psychiatric symptoms, and tended to be more fearless and impulsive, than inhalant users whose use was less dependent on peer and group contextual factors.

Development of Inhalant-Related Use Disorders

Inhalant-related disorders have a multifactorial etiology. First, availability is important in determining the prevalence of drug use disorders. Inhalants are cheap, available in several forms in most households, easily concealed, legal to possess, and simple to consume. In addition to these factors, legislation can vary from state to state, with most states having enacted laws attempting to curb inhalant misuse, but the effectiveness of that legislation needs to be studied. Second, inhalant use is directly rewarding pharmacologically and may be indirectly rewarding though social effects. Several animal models have demonstrated the rewarding properties of these substances (discussed above), and adolescents usually gather in small groups to use inhalants. Use of inhalants facilitates entry into the group, socially reinforcing the use. Third, due to the quick intoxication profile, inhalant users often can evade detection or punishment by parents or authorities. Fourth, inexpensive inhalants may be one of the few exciting and novel experiences available to youths in impoverished and/or isolated communities. This parameter may help to explain the high prevalence of inhalant use on some U.S. and Canadian Native American reservations and among Aboriginal Australians.

In addition to extrinsic factors, intrinsic factors also contribute to inhalant-related problems. Persons with adolescent conduct disorder or antisocial personality disorder are prone to taking extreme risks, and these disorders have an increased comorbidity with IUD. Several studies suggest an association of inhalant use and conduct problems. Among youths in grades 7 through 12, inhalant users (compared with others who used no drugs or who used only cannabis or alcohol) had many characteristics suggesting conduct disorder. They accepted cheating more readily, admitted to more stealing, perceived less objection to drug use from their families, liked school less, and reported more sadness, tension, anger, and a feeling of being blamed by others. In addition, school surveys showed that inhalant users were more likely to be involved with other drugs. Similarly, among youths referred to court-mandated education for minor alcohol offenses, those who also had used inhalants reported fewer school honors and more expulsions, truancy, academic failures, criminal offenses, run away behavior, and associations with troubled peers, as well as many more drug and alcohol problems. More of these individuals also had family history of alcohol- or drug-related problems, specifically mothers or siblings. Studies of juvenile probationers, incarcerated youth, and other samples of delinquent and antisocial adolescents report lifetime inhalant use prevalence rates between 30% and 40%. One recent investigation of a large national sample of adult inhalant users concluded that persons with IUD had significantly elevated levels of antisocial conduct, including a variety of early-onset and interpersonally violent behaviors. Some families are burdened by antisocial personality disorder and substance dependence in the adults and by conduct disorder and SUDs (sometimes including IUD) in the adolescent children, and there is growing evidence that genetic factors play a role in these familial disorders. Less respect for parents, less familism (i.e., personal values related to family), less parental monitoring, and experiences of trauma within the family unit have all been related to increased risk of initiation of inhalant use. Even more than other drug users, inhalant users report histories of childhood abuse and/or neglect.

DIAGNOSIS AND CLINICAL FEATURES

Inhalant Intoxication

Inhalant intoxication is a clinical psychological disorder that occurs during, or immediately following, the inhalation and uptake of volatile hydrocarbons into the bloodstream. Hydrocarbon inhalation may be intentional (in the instance of misuse) or unintentional (in the instance of occupational exposures). Inhalation of toxic solvents from common products such as glues or paints, which often contain toluene and/or admixtures of other volatile solvents, can result in a constellation of adverse psychological and behavioral impairments, such as dizziness, incoordination, slurred speech, unsteady gait, lethargy, tremors, depressed reflexes, psychomotor retardation, generalized muscle weakness, blurred vision, diplopia, subjective feelings of euphoria, nystagmus, and stupor or coma. Two or more of these signs/symptoms occurring in parallel, or in close temporal contiguity, with inhalation of a hydrocarbon should render a DSM-5 diagnosis of inhalant intoxication. Other reported concomitant factors or pathologic sequelae of inhalant intoxication include vomiting, aspiration of vomitus, bronchospasm, chest pain, cardiac arrhythmias or arrest, accidental burns, seizures, illusions, auditory or visual hallucinations, delusions, perceptions of altered body size, and suffocation in a plastic bag placed over the mouth and nose. The DSM-5 and ICD-11 require that an assessment for IUD coincide with a diagnosis of inhalant intoxication (6C4B.3). Individuals meeting the criteria for inhalant intoxication can have a clinically relevant pattern of use differentiated as mild (6C4B.3XS5W), moderate (6C4B.3XS0T), or severe (6C4B.3XS25). If no comorbid IUD is identified in an individual meeting the criteria for an "episode of harmful use of volatile inhalants," the code is 6C4B.0.

If known, the specific inhalant (e.g., toluene) should be indicated in the DSM-5 diagnosis (e.g., butane intoxication instead of inhalant intoxication). However, many inhaled products are mixtures of psychoactive hydrocarbons, and it is therefore often difficult to identify the specific agent(s) leading to intoxication. Immediate factors that should arouse suspicion for inhalant intoxication include the identification of chemical odors on the breath and/or clothing of an intoxicated individual or the discovery of products that can be inhaled to produce psychoactive effects in the individual's possession. General indicators of increased risk of inhalant intoxication include signs of intoxication in a youth between ages 10 and 17 and in persons who appear intoxicated but who test negative on commonly administered toxicologic screens for substance use (which typically do not test for inhalant use).

No nationally representative estimates are available for the annual or lifetime prevalence of inhalant intoxication. However, it is assumed that inhalant intoxication rates are comparable to those of inhalant use, in general. If it is assumed that most inhalant users experience intoxication at least once in their inhalant-using experiences, then lifetime inhalant intoxication may be fairly common, especially among youth and in diverse high-risk populations (juvenile and criminal justice, rural, etc.). The same can be assumed for differences in race and gender proportions reported herein.

Laboratory Examination

Random, intermittent evaluation of biologic samples has little value for verifying self-reported abstinence in currently nonintoxicated patients with IUD, because inhalants have a small detection window in the body and tertiary metabolites can be attributed to other sources. However, patients displaying behavioral signs of inhalant intoxication do, in most cases, produce levels of inhalants within detection range in urine and breath. Confirming their presence may later aid in clinical and forensic investigations. Currently, the ratio of hippuric acid (a toluene metabolite) to creatinine in urine is one of the most frequently used methods for detecting toluene in the system. It is advised that urine samples be collected in glass chromatography vials containing little or no air space, tightly sealed, and refrigerated until ready for analysis of the ratio of hippuric acid to creatinine (a proxy for toluene concentration). Breath samples can also be collected in specially designed glass traps or in Tedlar (not Mylar) bags designated for solvent containment and should be assessed for the presence of inhalants within a few hours following collection.

Differential Diagnosis

Differentiation from intoxication with other substances, particularly those due to sedative (e.g., ethanol, benzodiazepine, and barbiturate) use, is aided by an identified history of inhalant use, the detection of inhalant odors or paraphernalia (e.g., solvent-soaked rags, bags, etc.), and substance residues on the skin or clothing, perioral or perinasal "glue-sniffer's rash" from contact with organic solvents, and toxicologic examination of body fluids. Similar to other substances of misuse, polysubstance use is commonly observed among inhalant users, and concurrent intoxication with additional substances may be assessed by history and toxicologic examinations. As noted above, intoxication in persons who test negative on conventional drug screens or who are members of high-risk groups for inhalant use also raises suspicion for inhalant intoxication. In addition, parental, familial, or peer reports of the patient's prior inhalant use and/or known affiliation with inhalant users increases the likelihood of an inhalant-related intoxication. Despite evidence of inhalant intoxication, in unconscious or comatose patients, other explanations (e.g., closed head injury) must be considered. Dextrose 50% for injection (50 g) and naloxone 2 mg intravenously may help rule out incapacitation or coma originating from diabetes or opioid use, respectively. If delirium develops at any point during intoxication with inhalants, the diagnosis is likely inhalant intoxication delirium, rather than inhalant intoxication. If the patient displays signs of depressed mood, anxiety, or psychosis during inhalant intoxication, and if those symptoms manifest in a manner severe enough to warrant independent clinical attention, the diagnosis should be inhalant-induced depressive disorder, inhalant-induced anxiety disorder, or inhalant-induced psychotic disorder, respectively.

Course and Prognosis

The onset of intoxication is nearly instantaneous following solvent inhalation, given that these compounds are rapidly absorbed across pulmonary membranes in alveolar spaces and are rapidly distributed to the brain and other areas of the body high in lipid content. Inhalants are quickly metabolized and excreted from the body through respiration or urine, and inhalant intoxication usually lasts a short duration of a few hours or less, unless the user is intermittently or continuously administering the solvent. Unless trauma, hypoxia, cardiac arrest, chemical burns, or other problems ensue, there are likely to be no lasting or irreversible effects from one or a few intoxication sessions, except that each administration of these drugs produces reinforcing effects, which may increase the likelihood of further use. Prolonged and repeated use may cause persisting harmful effects.

Treatment

Inhalant intoxication, similar to ethanol intoxication, typically requires no medical attention and resolves over a short duration. However, in extreme instances, certain effects resulting from intoxication, such as coma, bronchospasm, laryngospasm, cardiac arrhythmias, trauma, or chemical burns, require medical treatment. Otherwise, caring for an intoxicated individual primarily involves offering reassurance, quiet support, and careful attention to vital signs and level of consciousness. Some evidence suggests that inhalant users who experience physical agitation during episodes of intoxication may be prone to cardiac arrhythmias or arrest. So, surrounding individuals should take care to ensure a calm and reassuring environment when they suspect that someone is under the influence of an inhalant. Sedative drugs (e.g., benzodiazepines), given their similar pharmacologic profile, are contraindicated with inhalant intoxication, since they may potentiate the effects of the inhalant. Given that inhalants have short half-lives, inhalant intoxication usually improves markedly after only 30 minutes of abstinence, unless other potentiating drugs were consumed around the same time as the inhalant. Following resolution of inhalant intoxication, a careful evaluation is necessary with proper intervention or referral for IUD, other SUDs, conduct disorder, antisocial personality disorder, and possible comorbid mood, anxiety, neurocognitive, and other psychiatric conditions.

INHALANT USE DISORDER

In persons showing intermittent changes accordant with substance intoxication, together with smells of organic solvents, inhalation paraphernalia, or the occasionally present perioral or perinasal papular glue-sniffers’ rash, a diagnosis of IUD should be considered. The DSM-5 IUD criteria require recipients of the diagnosis to have shown a problematic pattern of use (which implies recurrent use) and to display related clinically significant functional impairments and/or distress. To receive an IUD diagnosis, patients must meet at least 2 of 10 diagnostic criteria. IUD criteria assess whether the patient has used inhalants over a longer period and/or in larger amounts than intended; whether there has been an ongoing desire or actual attempts to quit or reduce inhalant use that have proved unsuccessful; and whether there are indications of craving for, tolerance to, or recurrent physically hazardous inhalant use. Other diagnostic criteria assess for inhalant-related social role impairments, use of inhalants despite social, interpersonal, physical, or psychological problems that are caused or worsened by inhalant use, and increased salience of inhalant use reflected in growing amounts of time spent using or recovering from the effects of inhalants. Several specifiers may accompany an IUD diagnosis. Individuals who meet two or three diagnostic criteria are diagnosed with mild IUD, while those meeting four or five criteria receive diagnoses of moderate IUD, and those meeting six or more criteria are diagnosed with severe IUD. Diagnostic specifiers can also be used to indicate whether the person with IUD is in early or sustained remission. If a specific inhalant is identified, this agent should be specified in the diagnosis (e.g., toluene use disorder instead of IUD), but inhalants are often mixtures of psychoactive chemicals, and it is often difficult to identify a specific inhalant. Thus, the diagnosis is often the generic “IUD.”

Although the DSM-5 provides no diagnosis for a characteristic inhalant withdrawal syndrome, and inhalant-related withdrawal signs and symptoms are not included in the diagnostic criteria for IUD, several studies have identified inhalant withdrawal signs/symptoms in some inhalant users; inhalant users with withdrawal signs/symptoms reportedly experience mild restlessness, inattentiveness, anxiety, and other physical symptoms. One national study of persons with a DSM-IV-TR disorder of inhalant dependence found that nearly 48% showed three or more clinically significant inhalant withdrawal symptoms and called for the inclusion of these symptoms in the DSM-5. Thus, despite the absence of a separate inhalant withdrawal diagnosis, patients’ complaints of withdrawal symptoms probably should be counted toward a diagnosis of IUD. ICD-11 uses the category of “volatile solvent dependence” (6C4B.2) to describe a similar disorder. These can be further classified as current use (6C4B.20), early full remission (6C4B.21), sustained partial remission (6C4B.22), sustained full remission (6C4B.23), or unspecified (6C4B.2Z).

Several studies of the DSM-IV-TR inhalant abuse and inhalant dependence diagnostic criteria specifically, and multiple studies of DSM-IV-TR substance abuse and substance dependence disorder criteria, indicated that SUDs were most properly conceptualized as unidimensional disorders of varying levels of severity. Hence, the DSM-5 includes only one “IUD” diagnosis but allows specification of severity based on the number of diagnostic criteria a patient meets.

Laboratory Examination

Therapy for SUDs often uses repeated tests of biologic samples to confirm patients' reports of abstinence or use. However, with inhalants, such tests may be difficult to interpret. First, these volatile compounds have a brief sojourn in the body and may be detected in urine for only a few hours after use. Second, even if the compounds occur in urine, they may volatilize out of samples during transfer or storage. Third, although hippuric acid, a toluene metabolite, can be detected longer than toluene, hippuric acid also may be produced from foods, raising a question of false-positive findings. Fourth, inhalants may bind to, or pass through, the plastic of urine cups or breath collection bags, reducing concentrations, and making the compounds undetectable. Thus, the most careful monitoring of inhalant use would involve frequent urine samples (e.g., two or three per week) at random times, collected in tightly sealed glass containers with little or no air space, and carefully refrigerated until analysis. Analyses for inhalants usually require gas chromatography, followed by flame ionization, electron capture, or mass spectroscopic procedures. Simultaneous analyses of the ratio of hippuric acid to creatinine may reveal recent toluene use. However, even under ideal conditions, the short half-life of inhalants makes inhalant monitoring much less valuable clinically than monitoring for many other substances. Similarly, breath samples may be collected in specially designed glass traps or in Tedlar (not Mylar) bags. The half-life of inhalants in alveolar air is a matter of minutes, and so breath samples may be useless for monitoring treatment progress in patients who show no current signs of intoxication. Moreover, the authors have examined costs in their region for monitoring tests, as a proportion of the per-patient, per-day reimbursement rates that must cover *all* costs of intensive day treatment. Each urine test for hippuric acid (a metabolite used to test for abuse toluene exposure) uses up about 1.8 days' reimbursement and takes about a week to complete. Thus, frequent biologic monitoring for inhalants is not very practical.

Differential Diagnosis

Several conditions should be considered in the differential diagnosis. First, most adolescents who experiment with inhalants stop spontaneously after one or a few episodes of use and never meet criteria for IUD. Although such use is dangerous and occasionally fatal (often on the first try), the many spontaneous resolutions support recommendations of minimalist, nonalarmist interventions for adolescent users who do not meet criteria for diagnoses of IUD; parents also should be involved. Firm advice to remain abstinent, coupled with education of parents on recognizing future problems, may be sufficient. Second, polysubstance use is common in adolescent patients, and dependence on drugs other than inhalants is proven through history, physical findings, and toxicologic screens. Such disorders may exist in addition to, or instead of, IUD. Third, uncontrolled and impulsive behavior during repeated episodes of inhalant intoxication may mimic aspects of, or be comorbid with, conduct disorder or antisocial personality disorder. Antisocial behavior before the initiation of IUD, or during periods of abstinence from inhalants, suggests the presence of these disorders.

The DSM-5 includes the following conditions for consideration in the differential diagnosis of IUD: unintentional (e.g., occupational) exposure to inhalants, intentional inhalant misuse that does not meet IUD criteria, intentional inhalant intoxication in a patient who does not meet IUD criteria, inhalant-induced psychiatric disorder (e.g., inhalant-induced depression) in a patient who does not meet criteria for IUD, and a number of additional substance-related (particularly sedative-related) and medical disorders that may present similarly to inhalant misuse and its consequences.

Course and Prognosis

The high prevalence of inhalant use in high school surveys, and its relatively low prevalence in adulthood, led one expert to state that inhalant use “should be regarded as a passing phase or fad.” Most inhalant users do not progress to IUD, but some studies indicate that inhalant use is associated with increased risk of future diagnoses of a broad range of psychiatric disorders including antisocial personality disorder and other SUDs. Among adult substance-dependent patients, a history of inhalant use was associated with significantly enhanced risk of the following: antisocial personality disorder, depression, anxiety disorders including social phobia, polysubstance use, and injection drug use. In fact, those who had been inhalant users within a group of adults with opioid (heroin) use disorder appeared to be a “marginal group with particularly unfavorable developmental conditions and a specific course of addiction.”

Most adolescents who progress from inhalant use to IUD eventually shift to use of, and often dependence on, other drugs, but some also continue inhalant use into adulthood. Such chronic patients may use inhalants for extended periods each day for many years; they often show moderate criminal activity, weight loss, significant medical disease, slow and slurred speech, and impaired attention and memory. In a national sample of inhalant users, those with an alcohol use disorder (AUD) and/or mood/anxiety disorder antedating their inhalant use were most likely to transition to IUD, typically within 1 year of inhalant use initiation. With frequent inhalant use, tolerance to the effects occurs as well as potential for mild withdrawal signs symptoms including sleep disturbance, irritability, shakiness, sweating, fleeting illusions, and nausea. One observer reported tachycardia, delusions, and hallucinations during withdrawal.

Medical problems in chronic users may include (1) muscle weakness, sometimes with myoglobinuria and rhabdomyolysis; (2) GI problems, such as abdominal pain, nausea, vomiting, or hematemesis; (3) renal dysfunction, often with severe electrolyte imbalance; (4) cardiomyopathy; (5) hepatotoxicity; (6) pulmonary disorders (e.g., pulmonary hypertension, increased airway resistance, and acute respiratory distress); and (7) hematopoietic disorders (including elevated carboxyhemoglobin levels, methemoglobinemia, hemolytic anemia, aplastic anemia, and even acute myelocytic leukemia). Neurologic problems include (1) headache, (2) paresthesias with peripheral neuropathy, (3) reversible cerebellar signs or cerebellar degeneration, (4) radiologic abnormalities of widened sulci and basal cisterns, and (5) dementia (e.g., lead encephalopathy from leaded gasoline or white matter dementia from toluene). There have been significant associations between longer duration of inhalant use and greater likelihood of experiencing tuberculosis, bronchitis, asthma, sinusitis, tinnitus, positive HIV/AIDS status, sexually transmitted disease infections, anxiety, and depression. As reviewed above, maternal inhalant use during gestation has been associated with an FSS similar to fetal alcohol syndrome.

Treatment

A Cochrane Collaboration systematic review published in 2010 showed no randomized or controlled clinical trials of IUD treatments. A systematic review of psychological interventions for inhalant misuse arrived at similar conclusions. Almost all evaluations of psychosocial treatments for IUDs were case series with poor methodologic quality. However, one study of Turkish adolescents assessed a brief cognitive behavioral intervention for inhalant misuse and observed significantly greater rates of continuous abstinence from inhalants at 1-year follow-up for treated youth compared to controls (i.e., 58% vs. 38%).

The lack of evidence-based treatment strategies for IUD is problematic for several reasons. Recent surveys of mental health practitioners show that they do not feel prepared to work with inhalant misusers, although they indicate that this patient group needs particularly intensive and specialized rehabilitative therapy. Moreover, research has documented high incidence rates of health services utilization and many perceived barriers to treatment among inhalant misusers. Some clinicians have called for specialty IUD treatment programs (although there are no data supporting this recommendation) and/or clinics and extended periods of treatment given the myriad psychiatric, neurocognitive, and health disorders shown by many, if not most, persons with an IUD. Given the small prevalence rates of IUDs, the development and implementation of such programs may prove challenging. Specialized programs should target vulnerable communities where IUDs are more common.

Case studies of pharmacologic interventions have assessed the use of drug therapy to reduce craving (e.g., aripiprazole, lamotrigine), inhalant-related withdrawal symptoms (e.g., baclofen), neuropsychiatric symptoms caused by, or cooccurring with, inhalant misuse, such as paranoia, psychosis, depression, and anxiety (e.g., carbamazepine, haloperidol, and risperidone), and behavioral symptoms resulting from acute or chronic inhalant misuse such as increased locomotor activity (e.g., vigabatrin, mGluR 2/3 agonist LY 379268). In addition, practitioners may have some success with administering benzodiazepine drugs to patients with acute withdrawal symptoms, given the similarities between inhalants and ethanol (another drug whose acute withdrawal symptoms are often treated with benzodiazepines). Although many of these case studies show potentially promising findings, the literature is simply too sparse and lacking in rigor to support a codified system of treatment guidelines for pharmacologic treatment of persons with inhalant-related disorders.

The scientific literature relating to psychosocial interventions for IUD is equally undeveloped, although holistic, indigenous-led residential treatments, family-based, activity-oriented, and broadly focused biopsychosocial interventions (with aftercare) have been recommended considering the variety and number of impairments typically evidenced by inhalant users. In addition, appropriate medical care is needed for the pathologic sequelae of inhalant use. Given that risk of IUDs is greatest in youth, comprehensive treatment is needed for adolescent patients who progress from experimentation to inhalant-related disorders. Most of these youths have comorbid conduct disorder and are at high risk of serious adverse outcomes. One authority suggests that a comprehensive treatment plan should include eight factors: detoxification; a peer advocate system; assessment of physical, cognitive, and neurologic deficits; building on existing strengths; developing new strengths; therapists trained in inhalant misuse; attention to personal and family issues; and aid in returning to the community.

Many inhalant-using youths are treated in various outpatient, day treatment, and nonhospital residential programs. However, treatment of these youths, many of whom are dependent on multiple substances, is not usually tailored specifically to drug classification (e.g., inhalants), but instead targets substance use and conduct problems more broadly. Referrals are often from social service and juvenile justice agencies, which pay for the treatment. Treatment in these programs begins with a detailed interview, addressing substance use and assessing for SUDs for each drug category in the DSM-5. Most patients meet the criteria for multiple SUDs. Interviews also assess diagnoses of disorders commonly found to be concomitant in this group: conduct disorder, attention deficit hyperactivity disorder (ADHD), major and persistent depressive disorders, and PTSD. Interviews also address experiences of abuse or neglect, which are common in these patients. Group and individual therapies are behaviorally oriented, offering immediate rewards for progress toward actionable goals in treatment and punishments for lapses to earlier behaviors.

Patients may attend on-site schools with special education teachers, together with planned recreational activities. The patients' parental families, often very chaotic, can be engaged in modifications of structural family therapy or multisystemic therapy, both of which have strong empirical support. Participation in 12-step programs is often needed. Treatment interventions may be coordinated closely with interventions by community social workers and probation officers. Medications are typically not prescribed for IUDs per se, but a psychiatrist often prescribes antidepressants for depression, disulfiram for comorbid AUD, or atomoxetine or bupropion for ADHD (stimulant medications are often avoided in youth with SUDs). Progress is checked with urine and breath samples analyzed for alcohol and other drugs at intake and during treatment. However, frequent monitoring of breath for inhalants, or of urine for hippuric acid, is too expensive for most public treatment programs. Patients who do not abstain in day treatment may be transferred to residential care. Treatment usually lasts 3 to 12 months. Termination is considered successful if the youth has practiced a plan to stay abstinent; is showing fewer antisocial behaviors; has a plan to continue any needed psychiatric treatment (e.g., treatment of comorbid depression); has a plan to live in a supportive, drug-free environment; is interacting with family in a more productive way; shows greater attendance at work or school; and is associating with drug-free, nondelinquent peers. In many cases, of course, these goals are only partially accomplished. As with all other recommended treatments of IUD, controlled studies of long-term outcomes are lacking. Additional therapies may be required for those demonstrating long-term symptoms associated with inhalant-induced damage.

While the long-term damage associated with persistent inhalant use (especially those with chronic encephalopathy) may pose difficulties with treatment interventions, a number of potential preventative and mitigating measures may prove useful in the future. In preclinical models, researchers have realized some success in reversing previous damage or preventing further damage in animals given repeated administrations of toluene. For example, as noted in a previous section, a preclinical study on mice found that apoptotic effects in the mPFC incurred by toluene exposure can be reversed by administering minocycline. Additional evidence suggests that compounds that promote aldehyde dehydrogenase production (e.g., metyrapone) may be useful in reducing the amount of a given aldehyde in the system. In the instance of toluene, benzaldehyde is of particular concern, as this compound, like acetaldehyde in the instance of ethanol consumption, may be responsible for precipitating oxidative stress within the cell. Little research has been done to corroborate this finding in toluene administration at either the preclinical or clinical level. One final method that has also received little attention is the administration of antioxidants in recovering, or even current, inhalant users. Antioxidants may help to restore intracellular levels of GSH and other defense systems to aid in the breakdown of superoxide production, which may in turn help to prevent further damage done. The long-term impact of inhalant use results in brain damage for a small, but not insignificant, portion of inhalant users. Development of both effective preventative and mitigative treatment strategies is imperative for clinical practice and will require collaboration with preclinical research.

INHALANT-INDUCED PSYCHIATRIC DISORDERS

Inhalant-Induced Delirium

DSM-5 provides a diagnostic category for inhalant intoxication delirium, which is a disturbance of consciousness and a change in cognition that is a direct physiologic consequence of intoxication with inhalants. Inhalant intoxication delirium should be diagnosed instead of inhalant intoxication if disturbances in attention and other cognitive functions (e.g., memory, language, and perception) predominate in the clinical picture and are severe enough to warrant clinical attention. In DSM-5 and ICD-11, all persons diagnosed with inhalant intoxication delirium must be assessed for a comorbid IUD. Persons with inhalant intoxication delirium are assigned a code of 6C4B.5, which should be associated with a causative condition (6C4B.0, 6C4B.1, 6C4B.2) and may be associated with volatile inhalant intoxication (6C4B.03) or volatile inhalant withdrawal (6C4B.4). The course and treatment of inhalant intoxication delirium are like those of inhalant intoxication, but the additional confusion requires special attention to patient and surrounding individual’s safety. If the delirium results in severe behavioral disturbances, short-term treatment with a dopamine receptor antagonist (e.g., haloperidol) may be necessary. Benzodiazepines should be avoided because of the possibility of adding to the patient’s respiratory depression.

Inhalant-Induced Psychotic Disorder

The essential features of inhalant-induced psychotic disorder are hallucinations or delusions judged by laboratory tests, history, or physical examination to be due to the direct physiologic effect of inhalant substances. This diagnosis applies to patients who meet criteria for inhalant intoxication but who also have psychotic symptoms in excess of those usually associated with inhalant intoxication. The psychotic symptoms must cause clinically significant functional impairment and/or distress and be severe enough to warrant independent clinical attention. This diagnosis is not made in the presence of inhalant intoxication delirium.

Consistent with DSM-5 and the ICD-11, persons diagnosed with inhalant-induced psychotic disorder are assessed for a comorbid IUD. The diagnosis of inhalant-induced psychotic disorder (6C4B.6) should be made with a causative condition (6C4B.0, 6C4B.1, 6C4B.2) and may be associated with volatile inhalant intoxication (6C4B.03) or volatile inhalant withdrawal (6C4B.4).

The course and treatment of inhalant-induced psychotic disorder are like those of inhalant intoxication. The disorder is brief, lasting a few hours to (at most) a very few weeks beyond the intoxication. Clinical evidence suggests that tetraethyl lead is a potent contributor to psychotic symptoms in countries still using leaded gasoline. Vigorous treatment of such life-threatening complications as respiratory or cardiac arrest, together with conservative management of the intoxication itself, is proper. Confusion, panic, and psychosis mandate special attention to patient and surrounding individuals' safety. Severe agitation may require cautious control with haloperidol (5 mg per 70 kg intramuscularly). Sedative drugs, including benzodiazepines, may potentiate and worsen inhalant intoxication.

In a randomized clinical trial, 40 hospitalized psychotic patients with DSM-III-R diagnoses of inhalant dependence and inhalant-induced organic mental disorder volunteered for treatment with either haloperidol (up to 40 mg per day orally) or carbamazepine (up to 1,600 mg per day orally) for up to 5 weeks. Both groups could receive intramuscular haloperidol for acute agitation. The antipsychotic response of the two groups was similar, but there were significantly fewer extrapyramidal reactions in the carbamazepine group. However, without a placebo-control group, these data cannot support either medication being better (or worse) than no medication. One case study of a 25-year-old inhalant-dependent Caucasian man found that 0.5 mg of risperidone daily for 4 weeks, following by a dosage increase to 1 mg daily, was effective in significantly reducing paranoia symptoms and craving for, and use of, inhalants. It was not clear, however, to what extent the psychotic symptoms the patient displayed were due to an independent disorder or inhalant induced. Controversy continues as to whether inhalants produce persisting psychotic states in the absence of delirium or neurocognitive disorder.

Inhalant-Induced Mood Disorder

The essential feature of inhalant-induced mood disorder is a prominent and persistent disturbance in mood characterized by depressed mood and/or anhedonia judged to be due to the direct physiologic effect of inhalant substances. Such mood symptoms sometimes develop during intoxication with inhalants, so this diagnosis applies to patients who meet criteria for inhalant intoxication but who also have depressive symptoms in excess of those usually associated with inhalant intoxication. The depressive symptoms must be severe enough to warrant independent clinical attention and associated with distress, functional impairment, or both. This diagnosis is not made in the presence of inhalant intoxication delirium. Per DSM-5 and ICD-11, individuals presenting with inhalant-induced depressive disorder are also assessed for a comorbid IUD. Persons with volatile inhalant-induced mood disorder are coded as 6C4B.70, and this should be made with a causative condition (6C4B.0, 6C4B.1, 6C4B.2) stated and may be associated with volatile inhalant intoxication (6C4B.03) or volatile inhalant withdrawal (6C4B.4).

The course and treatment of inhalant-induced depressive disorder are like those of inhalant intoxication. Inhalant-induced depressive disorder is brief, lasting a few hours to (at most) a very few weeks beyond the intoxication. Although antidepressant drugs are seldom proper for these relatively brief disorders, the risk of suicide requires a careful history and may require psychosocial treatment. Suicide has been implicated in about 40% of both inhalant-related medical examiners' death reports and inhalant-related visits to hospital EDs.

Inhalant-Induced Anxiety Disorder

The essential features of inhalant-induced anxiety disorder are panic attacks or other prominent anxiety symptoms judged to be due to the direct physiologic effect of inhalant substances. Such anxiety symptoms sometimes develop during intoxication with inhalants; thus, this diagnosis applies to patients who meet criteria for inhalant intoxication but who also have anxiety symptoms that predominate in the clinical presentation and that are in excess of those usually associated with inhalant intoxication. The anxiety symptoms must cause distress and/or functional impairments severe enough to warrant independent clinical attention. This diagnosis is not made in the presence of inhalant intoxication delirium. In ICD-11, acute intoxication due to the use of volatile solvents includes patients with marked anxiety symptoms arising during the intoxication.

The DSM-5 and ICD-11 diagnostic protocols require that all patients with inhalant-induced anxiety disorders be assessed for the presence and severity of comorbid IUD. Persons with a volatile inhalant-induced anxiety disorder are coded as 6C4B.71 and should be linked to the causing condition (e.g., 6C4B.0, 6C4B.1 or 6C4B.2) and may also be associated with other codes (6C4B.3 or 6C4B.4). The course and treatment of inhalant-induced anxiety disorder are like those of inhalant intoxication. Sedative and antianxiety drugs are contraindicated since they worsen inhalant intoxication, which precipitates inhalant-induced anxiety disorder.

Inhalant-Induced Major and Mild Neurocognitive Disorder

Before going into great detail, it is worth noting that the DSM-5 diagnostic criteria do not make a distinction between the specific substances of misuse when diagnosing substance-induced major and mild neurocognitive disorder. In addition, the diagnosis does not account for the unique effects of each substance outlined in its diagnostic criteria. Studies of inhalant-related neurocognitive disorder are riddled with equivocal findings, beset by numerous methodologic problems, including a narrowed focus on adolescent users with fewer lifetime exposures when the adverse consequences of inhalant use may manifest later in adulthood. Moreover, several investigations of adjudicated adolescent inhalant- and noninhalant-using polysubstance users observed greater deficits in executive function, processing speed, behavior, language, and memory in the inhalant-using group. In a series of studies conducted in Australia, adolescent inhalant users were found to have a variety of neurocognitive impairments relative to control when they were compared to cannabis/other drug users and to community controls. However, these inhalant users did not display interindividual variability among themselves across many measures. Clinical and some research evidence indicate that some inhalant-using adults develop inhalant-induced neurocognitive disorder. For example, in a study mentioned earlier, among toluene users (mean age 29 years) studied with MRI, the neuropsychological impairments correlated strongly with the severity of cerebral white matter abnormalities (e.g., poor white/gray matter distinction), and those abnormalities appear to be caused by inhalants.

The predominant feature of major neurocognitive disorder is a marked decline in level of performance in one or more cognitive domains such that the impacted individual experiences difficulties in the performance of instrumental activities of daily living. The cognitive difficulties can include impaired executive functioning (i.e., planning, organizing, sequencing, and abstracting), deficits in learning and memory, and difficulties with speech, social cognition, and perceptual-motor abilities. The symptoms must represent a noteworthy decrement from earlier functioning, not occur exclusively during delirium, and persist beyond the typical duration of inhalant intoxication. Most impacted persons will meet criteria for a moderate/severe IUD. The ICD-11 category of dementia in other specified diseases classified elsewhere includes inhalant-induced persisting dementia.

Inhalant-induced mild neurocognitive disorder is discerned from inhalant-induced major neurocognitive disorder primarily on the basis that the impairment(s) identified represent(s) a “modest” rather than a “significant” decrement in level of performance of one or more cognitive capacities, and the identified deficits do not interfere with the complex instrumental activities of daily living. In both major and mild inhalant-induced neurocognitive disorder, it is anticipated that abstinence from inhalants (concomitantly with abstinence from other neurotoxic agents) is associated with a stabilization or increment of neurocognitive impairments (i.e., not progressive decline). Both the DSM-5 and ICD-11 require that any patient meeting the criteria for a major or mild inhalant-induced neurocognitive disorder be evaluated for a comorbid IUD. Individuals with an inhalant-induced mild neurocognitive disorder may be identified by the code 6C4B.Y which is “other specified disorders due to the use of volatile inhalants” and should be made with a causative condition (6C4B.0, 6C4B.1, 6C4B.2) and may be associated with volatile inhalant intoxication (6C4B.03) or volatile inhalant withdrawal (6C4B.4).

Differential Diagnosis. Almost all patients who meet the criteria for an inhalant-induced neurocognitive disorder have a moderate to severe IUD, and many are alcohol dependent, which can also cause dementia. Many of these patients suffer from other conditions that may by themselves, or in a compounding manner, result in neurocognitive difficulties. These difficulties include a history of head injury and certain drug-induced infections, such as HIV. Such difficulties pose clinical challenges that require careful consideration. Even with clear evidence of prolonged inhalant use, a full evaluation of the myriad causes of neurocognitive dysfunction is required to make a proper diagnosis.

Course and Prognosis. Few of these patients have been studied prospectively. However, a recent 15-year follow-up of 77 previous inhalant users with no chronic encephalopathy showed significant improvements in standard neurologic, cognitive, and oculomotor tests compared to healthy participants. A subset of 17 users who had previously developed chronic encephalopathy demonstrated fewer improvements in the 15-year follow-up. Long-term abstinence may yield promising outcomes. Cessation in use may allow for recovery of brain function in a currently unknown proportion of inhalant users. Given current findings, some inhalant users may develop irreversible neurologic deficits from chronic encephalopathy following prolonged inhalant use. These impairments may cause inhalant users to have difficulties with executive functions (e.g., response inhibition), which may in turn increase the likelihood of relapse, complicating treatment, and each relapse may yield compounding cerebral toxicity. The sooner the user ceases use, the better the outcome and the less likely the user is to develop serious encephalopathy.

Treatment. There is no distinguished treatment of inhalant-induced neurocognitive disorder. Street outreach and comprehensive social service support measures have been offered to severely impacted inhalant-dependent homeless adults. Patients may require extensive support within their families or in foster or domiciliary care.

ADDITIONAL INHALANT CLASSIFICATIONS

Unspecified Inhalant-Related Disorder

The diagnosis of unspecified inhalant-related disorder is reserved for inhalant-related clinical presentations associated with clinically significant distress and/or functional impairment that do not meet full diagnostic criteria for any specific inhalant-related disorder.

Nitrous Oxide–Related Disorders

The DSM-5 includes nitrous oxide (N_2O)-related disorders among other (or unknown) substance-related disorders because of differences between nitrous oxide and other inhalants in their modes of action and associated problems. Still, many investigations categorize and label nitrous oxide and nitrites as “inhalants,” inconsistent with the DSM-5, and very few studies examine nitrous oxide and nitrites separately from other “inhalants.” Epidemiologic studies suggest that while toluene-containing products may frequently be utilized by adolescents, among adult past-year “inhalant” users, 68% had used nitrous oxide or whippets, making this the most commonly used “inhalant” in that group.

Nitrous oxide was introduced in 1844 by an American Dentist, Dr Horace Wells, for clinical practice, and it is still a widely used inorganic anesthetic gas. Nitrous oxide is also sold in small cartridges (known as “whippets”) as a propellant to dispense whipped cream. These cartridges are easily purchased through internet retailers and, not surprisingly, nitrous oxide misuse seems to appear most commonly among the health care and food service workers who use these preparations. A recent NSDUH estimated that between 90,000 to 122,000, 12- to 17-year-old Americans had inhaled “nitrous oxide” or “whippets’” at least once in their lifetime between 2018 and 2019 unassociated with a medical procedure. In one recent and relatively large study of conduct-disordered youth, nearly 16% reported lifetime nitrous oxide use.

Prior diagnosis with a psychiatric disorder and temperamental fearlessness were significant predictors of nitrous oxide use. Youth who had used volatile solvents and nitrous oxide had significantly greater psychiatric and SUD comorbidity compared to users of volatile solvents alone, nitrous oxide alone, or neither nitrous oxide nor volatile solvents. Nitrous oxide has a rapid onset and offset of action and is mostly excreted in the breath with little or no biotransformation. Nitrous oxide has been shown to reinforce self-administration behavior in animals. Although there is some evidence that nitrous oxide influences opioid systems, the subjective effects of nitrous oxide are not reduced by the opioid antagonist, naloxone.

In studies of persons without SUDs, some prefer inhaling nitrous oxide more than placebo, but others do not. In such studies, the drug produces feelings described as “drunk,” “dreamy,” “coasting or spaced out,” and “pleasant bodily sensations.” Individuals reporting more of these feelings are more likely to choose nitrous oxide over placebo. When compared in the laboratory to oxygen, nitrous oxide produces significantly higher ratings of euphoria but no significant increases in subjective sleepiness, distinguishing it from the CNS depressant effects of most inhalants. Nitrous oxide–exposed subjects do as well as oxygen-exposed subjects on tasks of coordination and reaction time but perform significantly worse on some measures such as tasks of logical reasoning and digit symbol substitution, where subjects replace numbers with symbols according to a digit-symbol code.

A recent meta-analysis of 85 studies reported that the most frequent clinical manifestations of nitrous oxide use included paresthesia, unsteady gait, and weakness. Chronic use of nitrous oxide can also produce severe outcomes with neurologic or hematologic disorders reported in almost all published cases, with more than half of the reports occurring in recreational settings. Diffuse polyneuropathy and myelopathy with extensive, although sometimes reversible, neurologic symptoms mimicking those of vitamin B_{12}–related pernicious anemia have been reported. Active vitamin B_{12} requires reduced cobalt, but nitrous oxide irreversibly oxidizes cobalt, suppressing the activity of an important enzyme, methionine synthase. This highlights the need to evaluate whether correcting B_{12} deficiency would prevent nitrous oxide–related toxicity.

A 35-year-old male dentist with no history of other substance use complained of problems with nitrous oxide misuse for 10 years. This use had begun as experimentation with what he had considered a harmless substance. However, his level of use increased over several years, eventually becoming almost daily for months at a time. He felt a craving for nitrous before sessions of use. Then, using the gas while alone in his office, he immediately felt numbness, a change in his temperature and heart rate, and an alleviation of depressed feelings. “Things would go through my mind. Time was erased.” He sometimes fell asleep. Sessions might last a few minutes, or up to 8 hours; they ended when the craving and euphoria ended. He had often tried to stop or cut down, sometimes consulting professionals about the problem.

Cases in the literature strongly suggest that nitrous oxide intoxication, misuse, and dependent use do occur. Considering the drug’s brief duration of action and users’ intermittent inhalation patterns, nitrous oxide appears unlikely to produce a clinically significant withdrawal syndrome. It is not clear, however, whether nitrous oxide produces other substance-related disorders. There are no clinically practical urine or breath tests for the presence of nitrous oxide. Misusers of the drug who develop neurologic symptoms may have low serum concentrations of vitamin B_{12}, and remission of neurologic symptoms has followed combined administration of B_{12} and folate, together with abstinence from nitrous oxide. In the absence of controlled studies of treatment of nitrous oxide use disorders, general principles for treating other SUDs should guide the treatment of these patients.

Amyl, Butyl, and Isobutyl Nitrite–Related Disorders

The DSM-5 includes amyl, butyl, and isobutyl nitrite–related disorders among "Other (or Unknown) Substance-Related Disorders." In a 2018 national household survey in the United States, among adult past-year inhalant users, amyl nitrite/poppers/rush were the second most used "inhalant" (>6,000,000) behind nitrous oxide. In comparison, nitrites were the 11th most used "inhalant" among adolescents in that same survey. While prevalence of nitrite use is relatively low in the general population, in some populations (i.e., urban HIV-positive gay and bisexual men), high rates of nitrite/popper use have been shown, and nitrite use has been associated with risky sexual behaviors.

The vapors of the volatile liquid amyl nitrite produce vasodilation and smooth muscle relaxation. Amyl nitrite, once supplied in easily "popped" glass vials, found wide use as acute inhalation therapy for angina pectoris between 1867 and ~1980. Sublingual nitroglycerin tablets have now superseded that use. Amyl nitrite and its close relative butyl nitrite enjoyed a flurry of use as recreational drugs during the 1970s and 1980s. Reports from that time suggest that the use was among adolescents, users of other drugs, and homosexual men. The latter group especially reported using the drugs to enhance orgasm, and these nitrites also were considered to produce a high, a feeling of wild abandon, or an altered state of consciousness. A recent study of antisocial youth found that 83% of nitrite users reported that they experienced intoxication when they inhaled nitrites. Nonmedical users called amyl nitrite vials "*poppers,*" and the availability of those vials declined with their attenuating medical use. Butyl nitrite and related compounds were then increasingly marketed as room deodorizers under suggestive names such as Rush or Locker room. These were intentional ways to skirt laws blocking their abuse like consumption however the target demographic of inhalant abusers remained the same.

In one study of patients who had misused these drugs, most experienced dizziness, light-headedness, cardiac palpitations, blurred vision, and a feeling of warmth immediately after inhalation. Others complained of immediate headache, nasal burning, nausea, cough, dyspnea, or syncope. Nearly half found the experience not at all pleasant; the others rated it only fair to good. Some experts believe the drug does not have primary reinforcing effects mediated in the CNS, but that certain pleasurable sensations associated with peripheral vasodilation lead to its continued use.

A few case reports document severe, but nonfatal, methemoglobinemia after nitrite inhalation by persons genetically deficient in methemoglobin reductase.

There also are reports of rapidly fatal methemoglobinemia after oral ingestion (not inhalation) of these compounds by genetically normal persons. Some epidemiologic studies further suggested that among homosexual men, nitrite use might increase the risk of HIV infection or of developing Kaposi sarcoma after HIV infection. However, those observations were heavily confounded since nitrite users tended to have more partners and to pursue unsafe sexual practices more frequently.

Considered together, the lack of a compelling instantaneous high, the immediate unpleasant feelings, and the longer-term risks from amyl and butyl nitrite seem to contribute to a low prevalence of use. In 2000 to 2001, among Americans 12 to 17 years of age living in households, 1.5% said that they ever had used “amyl nitrite, ‘poppers,’ locker room odorizers, or ‘Rush.’” Recent MTF survey findings and the results of a large study of conduct-disordered youth also found that less than 2% of 12 to 17 year olds had tried inhaling nitrites.

Some persons with sustained industrial exposures to nitrites experience withdrawal headaches on weekends or vacations, but it seems unlikely that intermittent, brief exposures to these short-acting drugs would produce clinically significant withdrawal in recreational misusers of nitrites. There are no studies for the treatment of amyl or butyl nitrite–related disorders, and so such treatment should follow general principles for the treatment of other substance-related disorders.

FUTURE DIRECTIONS

Although considerable progress has been made in better understanding the pharmacology of inhalants and the epidemiology of inhalant use and IUDs, there are several critical areas that should be the focus of future investigations. First, there is limited empirical research to guide the treatment of adolescents or adults with inhalant-related disorders; such work is vital to clinicians working with such populations. Second, addiction treatment often requires regular objective monitoring for substance use. The currently available methods for monitoring inhalant use are cost prohibitive, are specific to individual inhalants (do not cover the broad spectrum of possible inhalants), generally can only detect very recent use (i.e., testing for specific inhalants), or may lack specificity (i.e., testing for inhalant metabolites such as hippuric acid). Development of sensitive and specific monitoring methods, which are cost-effective, would represent a major contribution. Third, individuals with inhalant-use disorders are very likely to have other mental health problems. Little research thus far has attempted to better understand the reasons underlying that comorbidity (directionality of causality or a common shared risk). Given that inhalant use is common and that a single use may be fatally toxic, it is important to support continued efforts to develop empirically tested, effective inhalant prevention programs that may be broadly and cost-effectively implemented. Taxonomic research identifying subtypes of inhalant users could provide data that would facilitate matching of inhalant users to specific treatment interventions. These classifications should focus on sex differences as the reasons for, patterns of, and impacts from inhalant use. Studies of the short- and long-term consequences of inhalant use and longitudinal investigations of the natural history of IUDs are needed. Ethnographic investigations of cultural and subcultural factors relevant to inhalant use could prove informative as to the reasons why and how inhalants are misused among geographic subpopulations. Psychometric research is needed to examine the reliability, validity, and latent structure of DSM-5 inhalant use and inhalant-induced disorders diagnostic criteria in different populations of inhalant users. Health services research is needed to better understand the health care utilization patterns of persons with inhalant-related disorders. Additional investigation of the phenomenology of inhalant intoxication across different inhalants and mixtures of inhalants would provide additional insight into the reinforcing properties of inhalants. Last, exploration of the impact that initial inhalant use has on subsequent drug and psychiatric trajectories is warranted.

Further Readings

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Agency for Toxic Substances and Disease Registry (ATSDR). *Toxicological Profile for Toluene*. Agency for Toxic Substances and Disease Registry, U.S. Department of Health and Human Services, U.S. Public Health Service; 2017. https://www.atsdr.cdc.gov/toxprofiles/tp56.pdf

Bowen SE, Balster RL. A comparison of the acute behavioral effects of inhaled amyl, ethyl, and butyl acetate in mice. *Fundam Appl Toxicol*. 1997;35(2):189–196.

Bowen SE, Cruz SL. Inhalants: addiction and toxic effects in the human. In: Madras B, Kuhar M, eds. *The Effects of Drug Abuse on the Human Nervous System*. Vol. 2. Academic Press; 2014:553–569.

Bowen SE, Daniel J, Balster RL. Deaths associated with inhalant abuse in Virginia from 1987 to 1996. *Drug Alcohol Depend*. 1999;53(3):239–245.

Bowen SE, Hannigan JH, Davidson CJ, Callan SP. Abstinence following toluene exposure increases anxiety-like behavior in mice. *Neurotoxicol Teratol*. 2018;65:42–50.

Cairney S, O'Connor N, Dingwall KM, et al. A prospective study of neurocognitive changes 15 years after chronic inhalant abuse. *Addiction*. 2013;108(6):1107–1114.

Carreon-Garcidueñas M, Godinez-Hernàdez D, Alvarado-Gómez N, Ortega-Varela LF, Cervantes-Durán C, Gauthereau-Torres MY. Participation of voltage-gated sodium and calcium channels in the acute cardiac effects of toluene. *Toxicol Mech Methods*. 2018;28(9):670–677.

Centers for Disease Control and Prevention. Youth risk behavior survey data. 2019. Accessed March 12, 2021. https://www.cdc.gov/healthyyouth/data/yrbs/index.htm

Crossin R, Arunogiri S. Harms associated with inhalant misuse in adolescent females – a review of the pre-clinical and clinical evidence. *Drug Alcohol Depend*. 2020;216:108232.

Cruz SL. Inhalant misuse management. The experience in Mexico and a literature review. *J Subst Use*. 2018;23(5):485–491.

Cruz SL, Armenta-Reséndiz M, Carranza-Aguilar CJ, Galván EJ. Minocycline prevents neuronal hyperexcitability and neuroinflammation in medial prefrontal cortex, as well as memory impairment caused by repeated toluene inhalation in adolescent rats. *Toxicol Appl Pharmacol*. 2020;395:114980.

Cruz SL, Rivera-García MT, Woodward JJ. Review of toluene action: clinical evidence, animal studies and molecular targets. *J Drug Alcohol Res*. 2014;3:235840.

Filley CM, Halliday W, Kleinschmidt-DeMasters BK. The effects of toluene on the central nervous system. *J Neuropathol Exp Neurol*. 2004;63(1):1–12.

Ford JB, Sutter ME, Owen KP, Albertson TE. Volatile substance misuse: an updated review of toxicity and treatment. *Clin Rev Allergy Immunol*. 2014;46(1):19–33.

Hall MT, Edwards JD, Howard MO. Accidental deaths due to inhalant misuse in North Carolina: 2000–2008. *Subst Use Misuse*. 2010;45(9):1330–1339.

Konghom S, Verachai V, Srisurapanont M, et al. et al. Treatment for inhalant dependence and abuse. *Cochrane Database Syst Rev*. 2010;(12):CD007537.

Marín-Navarrete R, Toledo-Fernández A, Villalobos-Gallegos L, Pérez-López A, Medina-Mora ME. Neuropsychiatric characterization of individuals with inhalant use disorder and polysubstance use according to latent profiles of executive functioning. *Drug Alcohol Depend*. 2018;190:104–111.

Maxwell JC. Deaths related to the inhalation of volatile substances in Texas: 1988–1998. *Am J Drug Alcohol Abuse*. 2001;27(4):689–697.

Miech RA, Johnston LD, O'Malley PM, Bachman JG, Schulenberg JG, Patrick ME. *Monitoring the Future National Survey Results on Drug Use, 1975–2019*. Vol I. Secondary school students. Institute for Social Research, The University of Michigan; 2020. http://monitoringthefuture.org/pubs/monographs/mtf-vol1_2019.pdf

Mustonen A, Niemelä S, McGrath J J, et al. et al. Adolescent inhalant use and psychosis risk—a prospective longitudinal study. *Schizophr Res*. 2018;201:360–366.

Ossiander EM. Volatile substance misuse deaths in Washington State, 2003–2012. *Am J Drug Alcohol Abuse*. 2015;41(1):30–34.

Oussalah A, Julien M, Levy J, et al.et al. Global burden related to nitrous oxide exposure in medical and recreational settings: a systematic review and individual patient data meta-analysis. *J Clin Med*. 2019;8(4):551.

Stanley LR, Swaim RC. Initiation of alcohol, marijuana, and inhalant use by American-Indian and white youth living on or near reservations. *Drug Alcohol Depend*. 2015;155:90–96.

Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. *Treatment Episode Data Set (TEDS): 2018. Admissions to and Discharges from Publicly Funded Substance Use Treatment*. Substance Abuse and Mental Health Services Administration; 2020. https://www.samhsa.gov

Substance Abuse and Mental Health Services Administration. *Drug Abuse Warning Network, 2011: National Estimates of Drug-Related Emergency Department Visits*. HHS Publication No. (SMA) 13–4760, DAWN Series D-39. Substance Abuse and Mental Health Services Administration; 2013. https://www.samhsa.gov

Substance Abuse and Mental Health Services Administration. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2019 National Survey on Drug Use and Health* (HHS Publication No. PEP20-07-01-001, NSDUH Series H-55). Center for Behavioral Health Statistics and quality: Substance Abuse and Mental Health Services Administration; 2020. https://www.samhsa.gov

Woodward JJ, Beckley J. Effects of the abused inhalant toluene on the mesolimbic dopamine system. *J Drug Alcohol Res*. 2014;3:235838.

World Health Organization. International classification of diseases for mortality and morbidity statistics (11th Revision). 2018. https://icd.who.int/browse11/l-m/en

11.8 Tobacco-Related Disorders

SMITA DAS, M.D., PH.D., M.P.H., BRIAN ISAKSON, PH.D., AND DOUGLAS M. ZIEDONIS, M.D., M.P.H.

Tobacco use disorder (TUD) is the most prevalent, deadly, and costly of substance use disorders (SUDs). The prevalence of addiction to nicotine, the primary psychoactive substance in tobacco, surpasses all other SUDs. Tobacco currently remains unrivaled as the leading cause of premature death and preventable morbidity in the United States and globally with an incidence of one in five deaths in the United States being attributed to cigarette smoking. TUDs are the leading major public health issue and uniquely affect psychiatric patients.

TUD has created tremendous health disparities for patients with mental illness or SUDs, and individuals with multiple adverse childhood events (ACEs) are also at high risk for TUD. The literature on the evidence on why mental health and addiction treatment settings must address tobacco use and TUDs is very clear, including the fact that 44% of all cigarettes in the United States are consumed by smokers with mental illness and SUDs. Among patients in mental health and addiction treatment settings, 50% to 95% are smokers. This is in stark contrast to the general trend of decreasing smoking rates in the United States at large (14% use cigarettes, with an additional 4.5% using e-cigarettes). The most common causes of morbidity and mortality among individuals with a psychiatric disorder or SUD, such as lung cancer, diabetes, and cardiovascular disease, are related to tobacco, and tobacco accounts for the disturbing statistic that people with mental illness die on average 25 years earlier than the general population. In addition to the negative physical health consequences, there is a probable independent association between suicide and current smoking and longer lifetime smoking duration. In a recent analysis, for example, longer lifetime smoking (≥41 vs. ≤10 years) was associated with higher odds of suicide (odds ratio [OR] = 2.26, 95% confidence interval [CI] = 1.30–3.93).

Unfortunately, TUD has also been one of the most often ignored SUDs in mental health and addiction treatment settings, particularly by psychiatrists; the need to address tobacco in these settings is clear, and clinical practice guidelines delineate expectations to do so. Historically, mental health and addiction treatment settings not only neglected to address tobacco use and TUDs, but there was a minimization and rationalization that supported use. In a survey by the American Association of Medical Colleges (2007) of more than 3,000 physicians, 47% of psychiatrists felt patients had more immediate problems to address and 22% reported that cessation heightens other symptoms. This survey marked a change of improved attitudes for the majority; however, these attitudes of rationalizing and not addressing the topic isolated psychiatrists compared to all other physician specialties surveyed. These assumptions are not supported by the literature.

While psychiatric patients may have a worse tobacco withdrawal experience during the first 2 weeks, their long-term mental status is not worsened and likely improves with their physical health. In fact, accruing evidence demonstrates significant improvement in mental health and quality of life as a result of tobacco cessation. A recent systematic review and meta-analysis conducted by Taylor et al. (2014) assessed 26 smoking cessation studies and found a significant decrease in anxiety, depression, and stress, as well as improvements in overall mood and quality of life among quitters compared to nonquitters (samples not limited to mental health populations). Tobacco cessation also enhances the likelihood of successful abstinence and sobriety from other addictive substances among individuals with an SUD, as continued tobacco use is actually found to hinder recovery from other addictions. While they have worse withdrawal as a result of heavier dependence, people with psychiatric comorbidities are able to quit with overall stable mental health and much better physical health, quality of life, finances, and often lower dosages of psychiatric medication.

Psychiatrists have an opportunity to help improve overall well-being and the lifespans of their patients by evaluating and treating TUD using evidence-based approaches in psychiatric treatment settings, including psychosocial treatments, pharmacotherapies, community resources, and carbon monoxide (CO) monitors. Seven U.S. Food and Drug Administration (FDA)-approved pharmacologic agents are available to help individuals with TUD, including five nicotine replacement therapies

(NRTs) and two non-NRT pills (bupropion and varenicline). All seven of these medications can be effectively used in psychiatric patients; three NRTs are available over the counter (OTC) or by prescription (patch, gum, lozenge), two NRTs are only by prescription (nicotine inhaler and nasal spray), while the two non-NRTs are by prescription only. The integration of these medications with TUD-specific psychosocial interventions, community resources, and providing guidance and appropriate monitoring can improve quit rates by 50%. Treating TUD will not only help improve a patient's well-being, but these services can be reimbursed by completing "meaningful use" assessments of tobacco, performing CO assessments, and performing TUD treatments. Ultimately, treating TUD in people with mental illness has multiple benefits at the individual, provider, and system levels. There are also evidence-based psychosocial interventions that should be integrated, including through referrals to state quit lines and other community resources that are available by phone, internet, and through in-person meetings.

HISTORY

Tobacco use in the New World dates back to at least 600 AD and was introduced into European culture in the 16th century. Early on, most tobacco use was via pipes, smokeless tobacco, cigars, or in traditional ceremonial activities. The cigarette became popular beginning in the early 1900s with the invention of the cigarette-making machine, the match, and the use of acidifying agents to permit nicotine to enter the lower respiratory tract, where it could be rapidly absorbed into the arterial circulation. Cigarette use grew dramatically in the first half of the 20th century, initially through distribution for free to the military in World War I and II and next with targeting women as well through targeted marketing. The first reports of the association of smoking and disease began in the 1950s and culminated with the 1964 U.S. *Surgeon General's Report on Smoking and Health*. The tobacco industry sought to fight the clear risk of cigarettes by marketing "light" or "low tar" cigarettes, which in reality are no less harmful than other cigarettes and are often just smoked more efficiently. With the tobacco industry's efforts in marketing, tobacco use in the United States did not decline significantly until the 1970s and 1980s owing to tobacco-control activities and taxation. In the 1990s and 2000s, tobacco use in the United States declined a little, and that decline may be largely due to environmental smoke laws after the Environmental Protection Agency declared secondhand smoke dangerous as well. Of note, the tobacco industry also tried to suppress these findings. Although tobacco use in many Western nations has declined, use in developing countries is substantially increasing in part due to relaxed policy enforcement and increased marketing practices. Recent attempts by the FDA and the World Health Organization (WHO) to regulate and discourage tobacco use are important but have been only partially successful.

The tobacco industry targeted individuals with mental illness and/or addiction in numerous ways, a likely cause for the high rates of TUD in this population, as described by Prochaska and the evidence documented in The Legacy Tobacco Documents Library (http://legacy.library.ucsf.edu). For example, the tobacco industry attempted to validate the "self-medication hypothesis," that people with schizophrenia needed to smoke to manage their psychiatric symptoms. R.J. Reynolds in a Legacy Library record even specifically studied what they considered the "downscale market profile," people who are homeless or mentally ill, and documents from them note the population is "more impressionable to marketing/advertising … they're more susceptible. They're less formed intellectually … more malleable." Research promoting smoking among those with mental illness was widely distributed in meetings, publications, and through lobbying/policy efforts. Advertising was directed at people with mental illness, implying that tobacco was part of psychiatric illness (i.e., advertisement headline reading, "Schizophrenic. For New Merit, having two sides is just normal behavior"). Also worrisome is the intimate history the tobacco industry had with psychiatric institutions, supplying them with free, low-cost, or tax-free cigarettes. While much of this activity took place from the 1950s to 1970s, even in the 1990s, mental health centers were still communicating with companies like R.J. Reynolds for cigarette donations. Tobacco companies also lobbied to allow for smoking in psychiatric facilities after the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) mandated smoke-free hospitals.

Beyond the health consequences of tobacco use, the 1988 U.S. Surgeon General's report, *The Health Consequences of Smoking—Nicotine Addiction*, comprehensively reviewed the evidence and concluded that smoking was a form of substance dependence. The most recent reports in 2014 and 2020, "The Health Consequences of Smoking

—50 Years of Progress" and "Smoking Cessation: A Report of the Surgeon General" outline the dangerous history of tobacco and the strong effects of tobacco control policies in the United States. Nonetheless, tobacco remains the leading preventable cause of death in America.

COMPARATIVE NOSOLOGY AND DIAGNOSTIC FEATURES

Although tobacco dependence prevalence has been high over much of the last century, it was only in the 1980s that the term nicotine dependence was first accepted and included (though not coded) in the *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition (DSM-III-TR) and the ninth edition of the *International Classification of Diseases* (ICD-9) as an SUD. In the 1970s, tobacco corporations sought to influence the American Psychiatric Association (APA) task force revising the DSM; these efforts may have led to the narrow definition of tobacco dependence initially, for example, requiring someone to have seriously tried to quit in order to be diagnosed with TUD. The 1988 U.S. Surgeon General's report, *The Health Consequences of Smoking—Nicotine Addiction* comprehensively reviewed the evidence and concluded that smoking was a form of substance dependence. Also in DSM-III, *abstinence-induced nicotine withdrawal syndrome* was first classified. Nicotine *dependence* and *withdrawal* were the two defined nicotine-related disorders in the revised fourth edition of the DSM-IV-TR, which distinguished nicotine from other substances.

Contrary to other substances, nicotine *abuse* was omitted as a diagnosis in DSM-IV due to the strong propensity for use to transition to what was defined as dependence (when tolerance and withdrawal criteria are met). Nicotine abuse was also confined to significant psychosocial problems but not physical problems, and the former is rare with nicotine use. (Physical problems, however, are present even with nondaily smokers.) The 10th edition of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) included *harmful use*, a category that is similar to abuse but includes continued use that causes physical problems (which was warranted given the high rates of use despite numerous health consequences). ICD-11, effective as of 2022, also includes harmful patterns of use, nicotine dependence, intoxication, and withdrawal among other nicotine-related diagnoses, relying less on tobacco terminology.

May 2013 marked the release of the current DSM-5, which now includes among the 10 substance-related and addictive disorders the classification of tobacco-related disorders. Throughout this chapter, prior DSM-IV *nicotine dependence* terminology has been replaced with the current DSM-5 classification of *TUD*. The DSM-5 diagnostic criteria will be discussed in more detail in this chapter in the assessment section related to establishing a TUD diagnosis.

EPIDEMIOLOGY

In the United States, 14% of the adult population smokes cigarettes, with TUD being the most prevalent of all substance-related and addictive disorders. The prevalence of smoking in the United States has significantly decreased—it was just shy of 50%—over the past five decades. Unfortunately, some groups still disproportionately use tobacco, such as 50% to 85% of those with mental illness/SUDs, with lower income, and with lower education. Females have slightly lower rates of tobacco use (15.2% in women vs. 26.2% in men). Lower tobacco use rates are observed in those older than 65 years of age (11.4%) likely as a result of premature death. Worldwide, 80% of the 1.3 billion tobacco users worldwide live in low- and middle-income countries.

Ninety percent of smokers tried their first cigarette by age 18 and 98% by age 26. Sixty-nine percent of current smokers want to quit smoking completely, and 52.4% of smokers in the United States have made a quit attempt for >1 day in the past year. While 6.2% have sustained cessation, less than half of smokers (48.3%) who saw a health professional in the past year reported receiving advice to quit. In those who try to quit, use of counseling and/or cessation medications was 31.7%.

In terms of the diagnosis of nicotine dependence, approximately 28% of people in the United States develop nicotine dependence at some point in their lives, making it the most prevalent psychiatric disorder. Approximately 50% of current adult daily smokers are nicotine dependent. Nicotine withdrawal occurs in approximately 50% of smokers who try to quit.

Prevalence of Other Traditional and Emerging Tobacco Products

As a reference to describe other products, a description of a cigarette is warranted. A cigarette is a cylinder of tobacco rolled in thin paper for smoking. Nicotine, a naturally occurring pesticide, is the primary psychoactive and addictive chemical in tobacco, and hence cigarettes. Other chemicals are added to the shredded tobacco mixtures such as humectants (i.e., propylene glycol) to keep the tobacco moist, flavoring products (e.g., menthol), enhancers, sugars, and chemicals, such as ammonia, which boost the addictive properties of cigarettes, especially when burned. A cigarette contains 6 to 11 mg of nicotine, allowing for 1 to 2 mg of absorption, and smoking a pack per day equals 20 to 40 mg of nicotine absorbed per day.

Use of other forms of tobacco products, including older products (e.g., cigars), in addition to new emerging products, such as smokeless tobacco (e.g., snus, hookahs, and dissolvables) and other alternatives, such as electronic nicotine delivery systems (ENDS or e-cigarettes), are gaining popularity today. National estimates based on the 2020 National Health Information Survey (NHIS) reported every day or someday use of the following among adults: Any tobacco product, 20.8%; any combustible product, 16.7%; cigarettes, 14%; cigars/cigarillos/filtered little cigar, 3.6%; regular pipe/water pipe/hookah, 1%; e-cigarettes, 4.5%; smokeless tobacco, 2.4%; and ≥2 tobacco products, 3.9%. Notably, e-cigarette use has been on the rise, and these products are becoming more appealing and accessible among children and adolescents. There are also valid concerns around these products serving as gateways to other tobacco products. Patients should be educated on the addictive nature of emerging forms of tobacco as well as possible misconceptions regarding safety of these products, which are not FDA approved.

Adverse health effects of cigars and cigarillos are comparable to those associated with cigarette use, including cancers (oral, esophageal, and lung) and tobacco-related illnesses that are significantly correlated with level of smoke exposure. Cigars are arguably more harmful than cigarettes as they contain higher concentrations of known carcinogens, toxins, and tar. Whether a cigar user inhales or not, they are still absorbing nicotine; when inhaled, nicotine is delivered more quickly through the lungs, whereas if not inhaled, nicotine is absorbed at a slower rate through the tissue in the mouth. Cigars are offered in a wide range of sizes, thus the amount of nicotine that one might contain can vary largely. Cigars typically contain around 100 to 200 mg of nicotine but can contain as much as 444 mg of nicotine. It can be approximated that in proportion to the total nicotine amount and yield of cigarettes, that nicotine exposure from cigars can range anywhere from 9 to 68 mg. The smallest version of cigars is equivalent to the standard size of a cigarette, called a little cigar (containing about 1 g of tobacco—slightly more than the amount in a cigarette). A marginally larger and narrower version is called a cigarillo (containing about 3 g of tobacco).

Electronic cigarettes, also referred to as e-cigarettes or ENDS, have increasingly been linked to marketing statements that promote them as safer alternatives to tobacco, though there is a dearth of clinical research, including nicotine pharmacology studies, in this area to support this claim and meta-analyses indicating an association between use in adolescents and subsequent cigarette use (Khouja, 2021). Electronic cigarettes still contain, in addition to the addictive substance nicotine, toxic chemicals such as propylene glycol (an irritant in antifreeze). As supported by recent clinical studies, the nicotine exposure provided by e-cigarettes, compared to traditional cigarettes, varies according to the experience level of the user, where blood levels of nicotine and its metabolite, cotinine, tend to be similar to that delivered by cigarettes for current e-cigarette smokers and less concentrated for inexperienced users. The amount of nicotine in the e-liquid ranges in strengths of 6 to 24 mg/mL in a cartridge, though there is variation among different e-cigarette brands and brand models in overall nicotine content, as well as in the amount that is actually delivered to the user. The flavored nature of the substance can add appeal to vulnerable populations (youth) and also enhances the addictive potential. In addition, the products are newly regulated by the FDA (in 2016) and the regulation is limited, lending to surges of advertising and widespread availability in convenience stores, vape shops, etc. Emerging research about ENDS is located later in the chapter, including cessation potential and lung injury.

A *water pipe*, also referred to as hookah or shisha, is a device with origins in India. Water pipes date back to the 1700s and are used for smoking tobacco by delivering cooled charcoal tobacco smoke that passes through water before inhaled. Levels of CO, nicotine, tars, and heavy metal exposure from this product can often be even higher than what traditional cigarettes produce, contrary to the common belief that water pipe smoking has less health consequences than cigarette smoking. Variation in nicotine exposure is influenced by a variety of determinants, including the temperature at which the tobacco is burned, type of tobacco and water pipe product, and individual patterns of smoking behaviors.

Snus, originally from Sweden, is a form of smokeless tobacco that is encapsulated in a porous pouch that the user stores in between the cheek and the gum; it does not require the need to spit, as is the case with chewing tobacco/dip. Though it has been suggested to be associated with less significant harmful effects than cigarettes if entirely replacing cigarette smoking, it can be a gateway to other tobacco products and may decrease the likelihood of successful smoking cessation attempts. In Sweden, snus use has exceeded smoking, and the country has seen a dramatic decline in lung cancer rates; however, many former smokers are now dependent on snus, and there is evidence of an increased risk of pancreatic cancer.

Chewing tobacco and snuff are fairly similar smokeless products, with the former being a more finely ground form of tobacco and the latter provided in the form of loose leaves. Both are stored in the mouth and then spit out after use. Snuff is encapsulated in a porous pouch, whereas chew or finely ground tobacco is directly placed as a wad between the gum and the cheek, to allow nicotine to be absorbed by the mouth tissue. Typically, the nicotine content of these products can range from 3.6 to 4.5 mg. Compared to the 7 seconds it takes for the brain to respond to nicotine from inhalation of cigarette smoke, the same effect can be delayed to up to 30 minutes for chew or snuff users, though it is continuously delivered for up to 60 minutes after use (whereas nicotine delivery ceases immediately after the last puff of a cigarette). In terms of nicotine content, one dip or pouch of these forms of “spit tobacco” is equivalent to five cigarettes, whereas one can, containing 144 mg of nicotine, equates to smoking 4 packs of cigarettes. Smokeless tobacco users suffer the same systemic adverse effects as smokers due to tars and tobacco additives and have higher rates of oral, prostate, pancreatic, and cervical cancers compared with nonsmokers. Snuff users have a 50-fold increased risk of cancer of the cheek and gum. Smokeless tobacco is associated with numerous dental, periodontal, and oral soft tissue problems, including gingival recession, periodontal attachment loss, tooth staining, halitosis, and leukoplakia (a precursor to oral cancer).

Dissolvables are smokeless products that melt in the mouth, and similar to snus, are promoted as forms of tobacco not subject to prohibited use in smoke-free locations. They are provided in forms that mimic candy, such as lozenges, orbs, pellets, strips, or sticks, where the juices are swallowed and are often sweet or minty in flavor. Dissolvables are still addictive in nature and are contributing to poly-nicotine use in individuals who are trying to quit cigarettes (as often these products are marketed as quit aids).

ETIOLOGY

Nicotine and acetylcholine bind to and activate nicotinic acetylcholine receptors. Originally, these receptors were thought to be confined to the ganglia and neuromuscular junction but have now been found in several areas of the CNS. The two types of ligand-gated receptors that appear to be involved in the reinforcing properties of nicotine are the α4/β2 and α7 subtypes. These receptors are unusual in that they quickly desensitize. This phenomenon, plus the fact that repeated use of nicotine increases, not decreases, the number of receptors, suggests that nicotine may actually act as much as an antagonist as an agonist at this receptor. Emerging data indicate that distinct nicotinic receptors including those that contain the β4 subunit are involved in nicotine withdrawal symptoms.

The reinforcing effects of nicotine appear to be modulated by dopamine and glutamate release due to nicotinic receptors on dopamine cells in the ventral tegmental area and the core and shell of the nucleus accumbens, whereas withdrawal symptoms upon nicotine cessation are modulated by glutamate and corticotropin-releasing factor (CRF) release in the habenula-interpeduncular axis. The mesocorticolimbic system neural circuits are now seen as a common pathway for addictive drugs.

There are multiple contributors to tobacco dependence. Nicotine dependence is partly due to its multiple pharmacologic effects and the cigarette delivery device. Nicotine is best absorbed in the oral mucosa, nose, skin, and lung (more so than the stomach due to acidity). It also is able to enter the bloodstream without being processed by the liver (first-pass effects). Nicotine via cigarettes is rapidly absorbed directly into the arterial circulation and reaches the CNS in less than 15 seconds. Peak behavioral and cardiovascular effects occur within a few minutes. Nicotine is metabolized via the liver and has a half-life of 2 hours. The rapid effects of nicotine and the rapid decline lead to withdrawal symptoms that can form rather quickly in smokers.

Binding of nicotine to receptors leads to the subjective gratifying effects (through modulation of dopamine, glutamate, GABA, norepinephrine, and other systems) in the mesocorticolimbic system. The release of these neurotransmitters (via nicotine receptors) leads to the experience of reinforcing effects of smoking such as pleasure, arousal, and subjective lessening of anxiety. Many people with dependence subjectively feel as though nicotine is required to cope with life demands. Over time, continued use may be to suppress withdrawal. Tolerance also develops (part of dependence), as shown in several studies, and is established quickly and increases over the course of a day. Compared to nonsmokers, smokers often report less of the pleasurable effects of nicotine, such as arousal, vigor, and a subjective experience often referred to as "buzz," as a result of this tolerance.

Nicotine's different use patterns related to other drugs of misuse—that is, not striving for intoxication—may also contribute to nicotine dependence. Lack of intoxication allows the smoker to imbibe large amounts of nicotine on a daily basis without producing an obvious behavioral interference in daily activities, which, in turn, increases the chance of physical or psychological dependence. Legal, and relatively inexpensive compared to other drugs, cigarettes can be obtained with few social skills or resources. Being legal, punishment for smoking only occurs from social approbation.

The ideal reinforcer is one that occurs rapidly, frequently, and consistently, with minimal effort, and is unlikely to be punished. Nicotine via cigarettes has all these attributes. Inhalation (smoking) is the most rapid method to deliver a drug to the brain. It is also one of the most, if not the most, frequently reinforced behaviors. Taking 10 puffs per cigarette, times 20 cigarettes per day, times 365 days per year, times 20 years, results in millions of drug self-administrations. Because cigarettes are manufactured, they deliver precise doses of drug. Of course, the tobacco industry has long been invested in making cigarettes more addictive via design changes, paper choice, droplet sizes, additives (such as propylene glycol to keep the tobacco moist, or menthol flavoring), sugars, and chemicals, such as ammonia, which boost the addictive properties of cigarettes, especially when burned. The combination of multiple drug effects, lack of intoxication, and method of delivery makes nicotine users highly susceptible to developing dependence. In addition to the primary reinforcement described above, tobacco is highly subject to secondary reinforcement (classical conditioning), where conditioned stimuli (seeing a cigarette, smelling it, or being in situations where one usually smokes) promote the urge to smoke.

Although not widely known, initiation and cessation of tobacco use are as heritable as alcohol dependence or abuse. Some of the genetic effects are shared with alcohol, and some are specific to tobacco. Specific genes may affect responses of the body and the brain to nicotine (metabolism of nicotine, receptor sensitivity, and the levels of neurotransmitters). An example is variants in the *CYP2A6* gene that influence the rate at which a smoker clears nicotine; factors like this also contribute to an individual's susceptibility to tobacco addiction and response to treatments. Similarly, adolescent initiation of smoking can affect the trajectory to dependence (more than 90% of smokers begin as teenagers), social environment, peer modeling, and family environment. In addition, multiple ACEs are highly predictive of tobacco use and are an opportunity for prevention and early intervention.

Children are more likely to start smoking due to ACEs, and as well as those who have a high need to conform, low academic performance, rebelliousness, depressive symptoms, and poor self-esteem. Peer and family influences are paramount. Attention deficit, conduct, and alcohol or drug use disorders increase the risk of initiation and maintenance of smoking. Adolescent initiation of smoking can affect the trajectory to dependence (more than 90% of smokers begin as teenagers), along with social environment, peer modeling, and family environment.

DIAGNOSIS AND CLINICAL FEATURES

Diagnosis in SUDs has evolved to not include abuse versus dependence and rather be framed as levels of severity. To meet diagnostic criteria for SUDs including TUD, at least two of these 11 conditions must occur within a 12-month period in the DSM-5 (these are a blending of abuse and dependence criteria): (1) tobacco use often persists at a higher volume and over a longer duration of time than intended; (2) repeated desire or unsuccessful attempts to quit, cut down, or control tobacco use; (3) a significant amount of time is spent engaging in activities related to acquiring or using tobacco products; (4) craving, or an intense desire, to use tobacco; (5) inability to fulfill major role obligations at work, school, or home due to persistent use; (6) continued use despite continued or recurrent problematic social or interpersonal consequences, occasioned or worsened by the effects of tobacco (e.g., arguments with others about tobacco use); (7) important social, recreational, and work-related activities are given up or limited due to tobacco use; (8) tobacco use persists in situations where physically unsafe (e.g., smoking in bed); (9) tobacco use is continued despite knowledge of the probable worsening or contributing effects of tobacco use on persistent or recurrent physical or psychological problems; (10) withdrawal: nicotine associated withdrawal symptoms occur with stopping use; (11) tolerance: an increasingly larger dose of nicotine is needed to obtain the desired effect.

Abstinence threshold is defined as presenting none of the TUD criteria, with the exception of craving, for 3 or more months, with a cutoff of 12 months for early remission, and a threshold of 12 or more months for sustained abstinence. Severity can be assessed by the DSM-5 tobacco-related disorder classification and ICD-11 classification of dependence relies on just mild criteria being met. In DSM-5, mild severity is represented by the presence of two to three symptoms, moderate is four to five symptoms, and severe is six or more symptoms.

Tobacco withdrawal follows use for at least many weeks and either cessation or reduction in use resulting in anxiety; depressed mood; irritability, frustration, or anger; insomnia; difficulty concentrating; restlessness; and increased appetite. To meet withdrawal criteria, at least four of the seven signs or symptoms listed above must apply within 24 hours of abrupt cessation or volume reduction. Contrary to DSM-IV classification of withdrawal, DSM-5 omitted reduced heart rate or weight gain symptoms. In support of this change, research suggests that these variables have no predictive value on smoking cessation outcomes. The insomnia appears to be specific to increased awakenings and intense dreaming. Craving is common, and increased coughing, constipation, mouth ulcers, and decreased performance on vigilance tasks can occur. The syndrome is typically most severe with abstinence from cigarettes, intermediate with abstinence from smokeless tobacco, and least severe with abstinence from nicotine replacement products. Most withdrawal symptoms peak at 1 to 3 days and can last up to 2 to 3 weeks. In addition, craving can persist for 6 months or more.

Tobacco use and TUD are two to three times more prevalent among patients with anxiety, attention deficit, mood, substance use, and other psychiatric disorders. Conversely, these psychiatric disorders are two to three times more common among current smokers than among people who have never smoked or are ex-smokers. There are several possible reasons for this association; for example, shared genetic influences on smoking and psychiatric disorders, subjective improvement by nicotine of a psychiatric disorder, modeling other psychiatric patients, and boredom. Depressive symptoms, substance use, and attention deficit predict the likelihood of starting to smoke, the probability of becoming dependent, and the inability to stop smoking. Smokers with a current or past history of most psychiatric disorders are more dependent, have more withdrawal on cessation, and have a lower rate of smoking cessation.

DIFFERENTIAL DIAGNOSIS

Many of the symptoms of tobacco withdrawal can mimic, exacerbate, or mask the symptoms of psychiatric disorders or the adverse effects of psychiatric medications; for example, akathisia, anxiety, depression, irritability, insomnia, and weight gain. Individuals who do quit smoking may have increased psychiatric medication side effects or have caffeine intoxication due to the metabolic changes in the P450 1A2 isoenzyme with stopping smoking. Stopping smoking and maintaining the same caffeine intake will increase the availability of caffeine and its intoxication effects. A differential diagnosis for change in mental status or increased medication side effects should include investigating whether there was a change in smoking status. Nicotine intoxication (rare) causes abdominal pain, dizziness, headaches, nausea, pallor, palpitations, sweating, vomiting, and weakness; treatment is supportive and does not usually involve medications. Children are at risk of nicotine intoxication if they are able to access and consume nicotine medications or other nicotine products, including e-cigarettes.

COURSE AND PROGNOSIS

Twenty percent of all mortality in the United States is due to smoking, making it the leading cause of death in the United States, with a third of those deaths related to cardiovascular/cerebrovascular causes, another third related to cancer, and about a fifth related to respiratory diseases. About 10% of smoking-related deaths are from secondhand smoke exposure in nonsmokers and approximately 45% of smokers die of a smoking-related disease. In total, there are more than 480,000 deaths per year in the United States, which includes 41,000 due to secondhand smoke. Smoking is a major risk factor for lung cancer and chronic obstructive pulmonary disease (COPD), accounting for more than 75% of all deaths from these disorders. Other common smoking-related diseases include low–birth weight offspring, perinatal complications, other cancers (e.g., throat, bladder, and pancreas), and ulcers. Additional, less obvious health effects of tobacco demonstrated in research including impaired wound healing, diabetes, impotence, and osteoporosis. Secondhand smoke increases the risk of cancer and heart disease in spouses and the incidence of respiratory and ear problems in children.

With reference to cancer, the 2014 Surgeon General's report, *The Health Consequences of Smoking—50 Years of Progress*, reinforced there is sufficient evidence to infer a causal relationship between smoking and cancers of the lung, larynx, oral cavity, pharynx, esophagus, pancreas, bladder, kidney, cervix, stomach, and acute myeloid leukemia. Cigarette smoke has carcinogens that can act as direct carcinogens, cocarcinogens, or tumor promoters/initiators; smoke, free radicals, and other components can cause irritation, inflammation, cell proliferation/hyperplasia, oxidative stress and damage, and decreased organ function. Lung cancer, the leading cause of cancer deaths in the United States, is predominantly attributable to cigarette smoking, and demonstrates a clear dose response where the risk is associated with the number of cigarettes smoked per day and years smoked—this dose response is evident in other cancers as well.

Aside from cancer, cigarette toxicity in the lungs leads to several other health effects: loss of cilia, hyperplasia, metaplasia, mucus plugging of small airways, destruction of alveoli, and a reduced number of small arteries. Other lung/respiratory sequelae of TUD include COPD (emphysema and chronic bronchitis; the third leading cause of death in the United States, with 90% of cases attributable to cigarette smoking). Unfortunately, COPD is less noticeable in early stages of the disease process since small airways are affected first and are less clinically prominent. This often leads to later notice of effects on the large airways (emphysema) and irreversible progression of the disease. Other respiratory effects include increased risk of respiratory infection, such as pneumonia, in part due to cilia damage and triggering of asthma attacks.

Heart disease is the leading cause of death in the United States, and 30% of heart disease mortality is caused by smoking, where smokers have two to four times the risk of coronary heart disease, and women are more affected. The end effects of cardiovascular disease include myocardial infarctions, strokes, aortic aneurysm, and the impacts of peripheral vascular disease. Acutely, nicotine causes increases in systolic blood pressure, heart rate, and cardiac output while causing vasoconstriction; another acute effect of smoking is that CO reduces oxygen availability to the heart. Chronically, cigarette smoke increases inflammation, platelet activation/thrombosis, and endothelial dysfunction that can lead to atherosclerosis. Later in the inflammation process, smoking destabilizes plaques and promotes oxidation of low-density lipoproteins; the result is reduced myocardial blood flow, reduced oxygen, and artery occlusion. Nicotine is also arrhythmogenic (it increases serum catecholamines) and this, combined with carboxyhemoglobin lowering the threshold for ventricular fibrillation, can lead to sudden cardiac death, accounting for half of all deaths from coronary heart disease. Quitting smoking reduces the risk of sudden death immediately and in the long term. The risk of myocardial infarction is dose dependent, but as few as one to four cigarettes a day raises the risk of dying from ischemic heart disease. Also, when it comes to quitting smoking and heart disease, the risk is reduced by about half after just the first year of cessation and falls to that of “never-smokers” after 15 years of abstinence.

The course and prognosis of the health effects of TUD are worse with comorbid disorders and exposures. There is an additive effect of environmental carcinogens with smoking that worsens cancer outcomes, for example, with radiation and asbestos exposure. Similarly with oral, laryngeal, and esophageal cancers, alcohol use can worsen cancer outcomes, possibly by chemical interaction; for example, the risk of larynx cancer is 75% higher in people who use tobacco and alcohol versus those exposed to either substance alone. Finally, due to the deleterious effects on the cardiac and respiratory systems, smokers are less able to be physically active. This increases the risk of many other comorbidities.

Environmental tobacco smoke: Secondhand smoke impacts the prognosis of TUDs and the likelihood of quitting when living with other smokers. There are an estimated 7,357 chemical compounds in cigarette smoke, which include respiratory irritants (acrolein and acetaldehyde), cardiovascular damaging chemicals (cyanide, arsenic, and the cresols), and carcinogens (other metals, *N*-nitrosamines, and polycyclic aromatic hydrocarbons); over 70 of these chemicals are carcinogenic. Often smokers cohabitate with other smokers, especially in psychiatric group or extended living situations. However, even separating smokers from nonsmokers within the same air space, cleaning the air, opening windows, and ventilating buildings do not eliminate secondhand smoke exposure. Therefore, secondhand smoke is very relevant in psychiatry. Secondhand smoke exposure causes nearly 34,000 heart disease deaths annually among adult nonsmokers in the United States. Secondhand smoke is a cause of cardiovascular disease and lung cancer (increase in risk is 25% to 30%). There is also an association between exposure to secondhand smoke and risk of stroke.

Non-nicotine chemicals in tobacco smoke activate cytochrome P450 liver cytochrome (CYP1A2), thereby decreasing the levels of several medications. As a result, smoking cessation increases the concentrations of these medications. Many of these medications are psychiatric medications, and often the increase can be clinically significant in psychiatric patients; for example, haloperidol, clozapine, and fluvoxamine concentrations increase 30% to 40% with abstinence. Other examples are more antipsychotics (olanzapine and fluphenazine), antidepressants (amitriptyline, nortriptyline, imipramine, clomipramine, and trazodone), benzodiazepines, and caffeine. Therefore, it is important to monitor psychiatric medication doses in smokers where quantity smoked is changing. With a relapse to smoking in a patient on clozapine, for example, clozapine blood levels can be reduced, and psychiatric symptoms may re-emerge and can be life-threatening. Similarly, smoking cessation in someone maintaining steady caffeine consumption can incur caffeine “intoxication,” which will feel like severe nicotine withdrawal, including agitation and restlessness. It is also useful to motivate patients to quit smoking with the potential that they may decrease psychiatric medication doses.

Another special population includes the pregnant smoker: The health of a pregnancy and outcomes in children born to smokers or born to environments with smoke exposure should be addressed. Smoking can even make it harder for a woman to become pregnant and, once pregnant, can increase the risk of preterm (early) delivery, stillbirth (death of the baby before birth), low birth weight, sudden infant death syndrome (SIDS or crib death), ectopic pregnancy, and orofacial clefts in infants.

There may also be an association that children born to smoking mothers can also have increased risks of mental retardation, learning problems, and ADHD. Fathers are not immune from risks: Smoking can affect sperm and sexual function, which can reduce fertility and also increase risks for birth defects and miscarriage. The use of medications for smoking cessation during pregnancy is debatable. Whether the harm from smoking during pregnancy is due to nicotine, CO, or other chemicals in cigarette smoke is unclear. The concentrations of nicotine from medications are much lower than those from smoking, plus no CO and tar exposure occurs. Often, medications are reserved for pregnant smokers who have failed at cessation without medication.

TREATMENT

Given the widespread evidence base of the harmful effects of TUD, a great deal of research has been done on treatment and allows for the effective use of evidence-based treatments. Several websites and organizations offer information on TUD treatment, including free booklets. There are excellent guidelines for assessing and treating TUDs published in 2008 by the U.S. Public Health Service (USPHS) and for the treatment of SUDs in 2006 by the APA. Also, most state health departments have telephone lines to assist smokers in quitting and have websites with information for providers and patients (Table 11.8–1).

All patients should be assessed for (1) smoking status and (2) motivation to quit (Table 11.8–2). A useful approach to TUD treatment is the 5As model: *Ask, Advise, Assess, Assist,* and *Arrange* follow-up or the minimum standard of AAR: *Ask, Advise,* and *Refer*. Given the overwhelming prevalence and morbidity/mortality of TUD in psychiatric settings, all patients should be *asked* about current and past tobacco use, type of tobacco used, and frequency of use; TUD should be listed in their problem list and treatment plan with strategies for either increasing their motivation or recommendations for how to quit, perhaps including psychosocial treatment, education, community resource referral, and/or medication options and guidance. In addition to asking about tobacco use, assess for DSM-5 criteria. All patients using tobacco should be *advised* to quit with a personalized message, for example, including their specific health risks. With psychiatric patients, it is often important to state one's belief that the patient has the ability to stop. For some, the above advice has no immediate effect. However, the cumulative effect of repeating the advice at later visits and the receipt of advice from others can often motivate a quit attempt. Helping smokers who are not ready to quit to reduce the number of cigarettes smoked daily (aided by nicotine replacement medications) can motivate them to try to quit in the near future. The clinician should *assess* for desire to quit. Motivational interviewing/enhancement can be employed even if a patient is not ready to quit. Common motivators to quit are health concerns, effects of smoking on others, and social pressure. Common barriers to cessation are withdrawal, fear of failure, and fear of weight gain. Next, smokers need *assistance*, which can include setting a quit date, preparing for the quit date, need for medication or telephone or in-person therapy, weight gain concerns, high-risk situations, making cigarettes unavailable, discussions with significant others (especially if they too are smokers), and so forth. Finally, the fifth A is for *arranging* follow-up since TUD is an addictive disorder and relapse is likely. Follow-up can include working with the psychiatric provider or using other community resources, such as state quit lines (1-800-QUIT-NOW) (see Table 11.8–1). Because relapse is often rapid, the first follow-up phone call or visit should be 2 to 3 days after the quit date.

Table 11.8–1
Selected Community Resources for Tobacco Use Disorder Treatment

Internet Resources

https://www.hhs.gov/surgeongeneral/reports-and-publications/tobacco/index.html	U.S. Public Health Service guidelines
http://www.cancer.org	American Cancer Society
http://www.cdc.gov/tobacco	Centers for Disease Control and Prevention
http://www.lung.org	American Lung Association
http://www.nci.nih.gov	National Cancer Institute
https://www.srnt.org/page/Treatment	Society for Research on Nicotine and Tobacco
www.attud.org	Association for Treatment of Tobacco Use and Dependence
www.makesmokinghistory.org	Make Smoking History
www.quitnet.com	Quit Net
www.nicotine-anonymous.org	Nicotine Anonymous

Quit-Lines and Live-Chat Support

1-800-QUIT-NOW	National Network of Tobacco Cessation Quit-lines
1-855-QUIT-VET	U.S. Department of Veterans Affairs Quit-line
https://livehelp.cancer.gov/	National Cancer Institute Live Help Online Chat

Mobile Apps

www.text2quit.com	Text2Quit
www.smokefree.gov/smokefreetxt/	Smokefree TXT
www.livestrong.com/quit-smoking-app/	My Quit Coach
www.becomeanex.com	Become An Ex

http://teen.smokefree.gov/sftapps.aspx	Smoke-Free Teen
http://women.smokefree.gov	Smoke-Free Women

Other Local Resources to Explore in Your Area

- Nicotine Anonymous (www.nicotine-anonymous.org)
- Contact Employee Assistance Program (EAP). They may offer free help
- Contact insurance company to ask for help. Some assistance may be free or partially reimbursable, including counseling and NRT medications

Table 11.8–2.
5 A’s Model of Nicotine Cessation Intervention Recommended by the U.S. Public Health Service

1. **Ask About Tobacco Use:** Systematically screen (and document) current and past nicotine use at every visit
    - Signs of more highly dependent smokers: Smoking within the first 30 minutes after awakening and/or over 20 cigarettes per day
    - Document applicable DSM-5 criteria, withdrawal symptoms, prior quit attempts, types of treatment, social support for quitting, medical conditions
    - Can use a CO meter to assess cigarette use: 15–50 ppm = smoker; 0–5 ppm in 2–3 days of quitting
    - Ask about all forms of nicotine use (including e-cigarettes)
2. Advise to Quit (or reinforce people who have quit) with a clear, strong, and personalized message
    - “Your other health diagnoses, such as asthma, are worsened by smoking and quitting is one of the most important things you can do for your health”
    - “People with psychiatric diagnoses suffer more from smoking-related illnesses and have unique challenges to quitting that we can overcome”
3. Assess for Desire to Quit
    - “How does the idea of quitting smoking sit with you?”
4. Assist Patients

    Strategies for Lower Motivated Patients:
    - For those in precontemplation or contemplation, listen to reasons for not being in preparation
    - Try to increase motivation through CO meter scores, $$ spent, other disorders, increasing freedom, options in relationships, more job opportunities, effect of tobacco smoke on their loved ones (family, friends, pets), longer life, and fewer sick days
    - Do Pros/Cons or “Importance” or “Confidence” ruler
    - Consider the 5 Rs **relevance, risks, rewards, roadblocks,** and **repetition** (revisit the conversation in the future).

    Strategies for Higher Motivated Patients:
    - STAR Quit Plan: Set a quit date, Tell family/friends, Anticipate challenges such as withdrawal and triggers, and Remove nicotine products from environment
    - Discuss specific strategies for common triggers
    - Discuss withdrawal course (worst within 2 weeks)
    - Discuss and plan for slips and relapses

- Discuss alcohol, stress, weight gain specifically
- When possible, provide prescription for one or combination of 7 FDA medications for TUD (noting that some insurances cover over-the-counter medications when ordered by a physician)
- Encourage some form of counseling support (Quit lines, Nicotine Anonymous, counseling)
- Use shared decision making to increase motivation and buy-in from patient

5. **Arrange Follow-Up**
   - Set a date for a follow-up visit and repeat all of the A's in follow-up
   - Document and integrate into assessment/plan
   - Engagement with state's tobacco quit line (1-800-QUIT-NOW)

Psychosocial Therapies and Community Resources

Psychosocial interventions increase the chance of quitting by 40% to 80% based on a recent Cochrane review analysis and can be varied based on motivational level for quitting. Less-motivated-to-quit individuals can benefit from motivational interviewing/motivational enhancement therapy with personalized feedback. Feedback that has more impact includes CO meter scores, cost of cigarettes per year for them, health problems that they already have that are worsened or caused by ongoing tobacco use, and effect on family and significant others. Educational programs such as learning about healthy living provide information that can help enhance motivation. If patients worry about weight gain with quitting, exercise and nutrition to help them begin to lose weight before quitting can be a helpful strategy.

Behavior therapy, including relapse prevention, is the most widely accepted and well-proven psychological therapy for treating TUDs. Behavior therapy consists of several techniques, three of which are supported by good evidence. *Skills training* and *relapse prevention* identify high-risk situations and plan and practice behavioral or cognitive coping skills for these situations. *Stimulus control* refers to eliminating cues for smoking in the environment. This approach also requires preparation such as removing smoking triggers and working with loved ones to plan for support.

Mindfulness-based interventions are being utilized to help individuals. Mindfulness approaches help individuals stay present in the moment and recognize cravings as temporary thoughts and feelings that will dissipate over time and which they can notice and accept without needing to act on those thoughts or feelings. Mindfulness also includes foundational attitudes such as patience, trust, acceptance, gratitude, and openness to new ways, which are all aligned with SUD recovery. Mindful self-compassion adds to mindfulness the concepts of shared humanity (they are not alone, others are trying to quit, and it can be difficult at times) and kindness to self (similarly how they might be kind to others suffering in a similar manner). Mindfulness is also helpful for stress reduction in general.

Twelve-Step Facilitation is a standard evidence-based treatment in most addiction treatment; however, there are no studies of Nicotine Anonymous (NicA) in the literature. The strength of attending meetings can be in the social support and a plan for recovery. There are national phone meetings for easy access for anyone, as well as local meetings, but those are less numerous than Alcoholics Anonymous.

There are increasing studies on incentives or contingency management also helpful for smoking cessation, but implementation in real-world settings is limited by financial and regulatory constraints.

There are an increasing number of other community resources. The most widely used are quit lines that provide individual coaching over the phone and sometimes with free NRT. 1-800-QUIT-NOW works in all states and has moderate certainty evidence in Cochrane reviews. There are online internet websites/apps for those who prefer computer support (i.e., www.becomeanex.org and www.ffsonline.org). Other resources are clinic-run smoking cessation groups, peer specialists, peer websites (www.njchoices.org); a training module for peer specialists (www.Rxforchange.org); and information for club houses (www.genesisclubhouse.org). Other therapies, such as acupuncture, hospitalization, psychedelics, and hypnosis, have been studied, but overall, the results are equivocal and yet to be proven effective on a larger scale. Many of these resources are becoming increasingly more accessible and many are available at no cost. Medicare, Medicaid, and other insurance companies are also now supporting reimbursement of many psychosocial treatments, with more and more options being considered.

Pharmacologic/Interventional Therapies

The USPHS and the APA guidelines recommend that all smokers be offered medication to aid in smoking cessation. Recent reviews and meta-analyses suggest that medications may be more effective than psychotherapy approaches alone, and certainly combinations are best. The seven FDA-approved medications include five NRTs, bupropion, and varenicline. NRT patch, gum, and lozenge are available OTC, while NRT inhaler, NRT nasal spray, varenicline, and bupropion are by prescription only in the United States. All the medications approved by the FDA for smoking cessation double the quit rate (similar to results obtained with behavior therapy) and appear equally effective (Table 11.8–3). The one exception may be varenicline, which has been shown superior to some other medications. It can improve quit rates by two to three times compared to treatment as usual. Still though, NRTs can improve quit rates by 50% to 60% compared to placebo. The medications produce few side effects; less than 5% of patients stop medication due to adverse effects (except for nicotine nasal spray). While some medications have prescribing considerations for pregnant women or people with heart conditions, it is certainly the case that continued tobacco use is very harmful for these groups.

History and patient input are essential to choosing medications: cost can be a consideration, as well as previous attempts with a medication, medical comorbidities, severity of dependence/withdrawal, and breakthrough symptoms. Typically, combined patch and another NRT or monotherapy with varenicline are first tried. Combinations of medication have greater efficacy. Updated FDA guidelines on NRTs allow for use while smoking for risk reduction and promote using more than one medication (as shown in practice and research for many years).

Nicotine Replacement Therapies. Because nicotine undergoes first-pass metabolism by the liver, nongastrointestinal routes have had to be used to deliver nicotine. NRTs are effective, presumably because they reduce nicotine withdrawal without the immediate reinforcing effect of rapidly absorbed nicotine in tobacco smoke. Often NRT is used in inpatient or other clean air settings. Most replacement therapies use a short period of maintenance (6 to 12 weeks), often followed by a gradual reduction period (6 to 24 weeks). However, longer-term use may be necessary for some. Most NRTs should be used with caution in cardiovascular patients (specifically those who are 2 weeks postmyocardial infarction, have serious arrhythmias, or have unstable angina). Finally, to optimize oral absorption of nicotine in the gum, lozenge, or inhaler, minimize acidic beverage (coffee, juice, soda) consumption 15 minutes prior to and during use.

Nicotine patches (OTC), applied daily to a clean and hairless location, provide long-acting nicotine administration. Typically, patients use a 21-mg patch, although 7- and 14-mg patches can be used for lighter smokers (10 to 15 cigarettes per day). Side effects include local skin reactions (minimized with creams and by rotating patch sites) and insomnia/vivid dreams (minimized by removing the patch prior to bed).

Nicotine gum (OTC) may be used alone or in combination with other medications and is generally utilized hourly in initial phases of cessation. The 2-mg gum is recommended for those smoking less than 25 cigarettes per day, while the 4 mg is generally used for heavier smokers (but dose can vary by preference). Proper usage education is important for efficacy and to decrease adverse effects: patients should “chew” a few crunches until a peppery or tingly flavor emerges and then “park” between the cheek and the gum until the tingle goes away. This will help maximize nicotine absorption (for about 30 minutes) and reduce transfer to the stomach, which can irritate the gastrointestinal tract. The nicotine lozenge (OTC) is easier to take consistently and will achieve 25% to 27% higher plasma concentrations over time compared to the gum (allow the lozenge to dissolve, do not chew it). Patients who smoke their first cigarette within 30 minutes of awakening should use the 4-mg lozenge, while others might consider the 2 mg dose, but again preference should be considered. Both the gum and lozenges provide relief of immediate withdrawal, especially in high-risk situations.

Prescription NRT in the United States includes the inhaler and spray. The inhaler is intended for oral mucosal absorption like the gum and lozenge and provides the smoker with the hand-to-mouth routine of smoking. The nicotine nasal spray provides the fastest absorption of the NRT products and has the highest potential for dependency, as well as more perceived complications of use such as rhinitis.

Non-Nicotine Prescription Medications. Bupropion, which affects dopamine, norepinephrine, and nicotinic–cholinergic receptors, can help decrease cravings and withdrawal symptoms. Dosages of 300 mg per day reliably double quit rates in smokers with and without a history of depression. Combining bupropion and a nicotine medication has greater efficacy. Contraindications to use of bupropion include a seizure disorder (the risk of seizures in those appropriately screened is less than 1 in 1,000), current/past bulimia or anorexia nervosa, recent/concurrent monoamine oxidase inhibitor (MAOI) use, or other bupropion use. More common side effects are insomnia and dry mouth. This medication can be started 1 to 2 weeks before a quit date and is often used for up to 6 months postquit.

Varenicline is a partial agonist at the α4β2 neuronal nicotinic acetylcholine receptor; it both relieves craving and withdrawal and, unlike other medications, reduces the reinforcing effects of nicotine by blocking dopaminergic stimulation responsible for smoking reinforcement/reward. While there were concerns about the FDA-issued black box warnings for neuropsychiatric adverse events, more recent research of varenicline in psychiatric patients has found that psychiatric symptoms were not worsened, and there is no evidence of an increased suicide risk. This was especially demonstrated in the 2016 EAGLES trial, which included 8,144 participants across 140 centers in 16 countries; Anthenelli et al. demonstrated no significant increase in neuropsychiatric adverse events attributable to varenicline or bupropion relative to nicotine patch or placebo. Furthermore, the study demonstrated that varenicline was more effective than placebo, nicotine patch, and bupropion in helping smokers achieve abstinence, and bupropion and nicotine patch were more effective than placebo. A subsequent 2016 FDA review did not support the original neuropsychiatric warning which has since been removed.

Table 11.8–3.
Pharmacotherapies for Tobacco Use Disorders

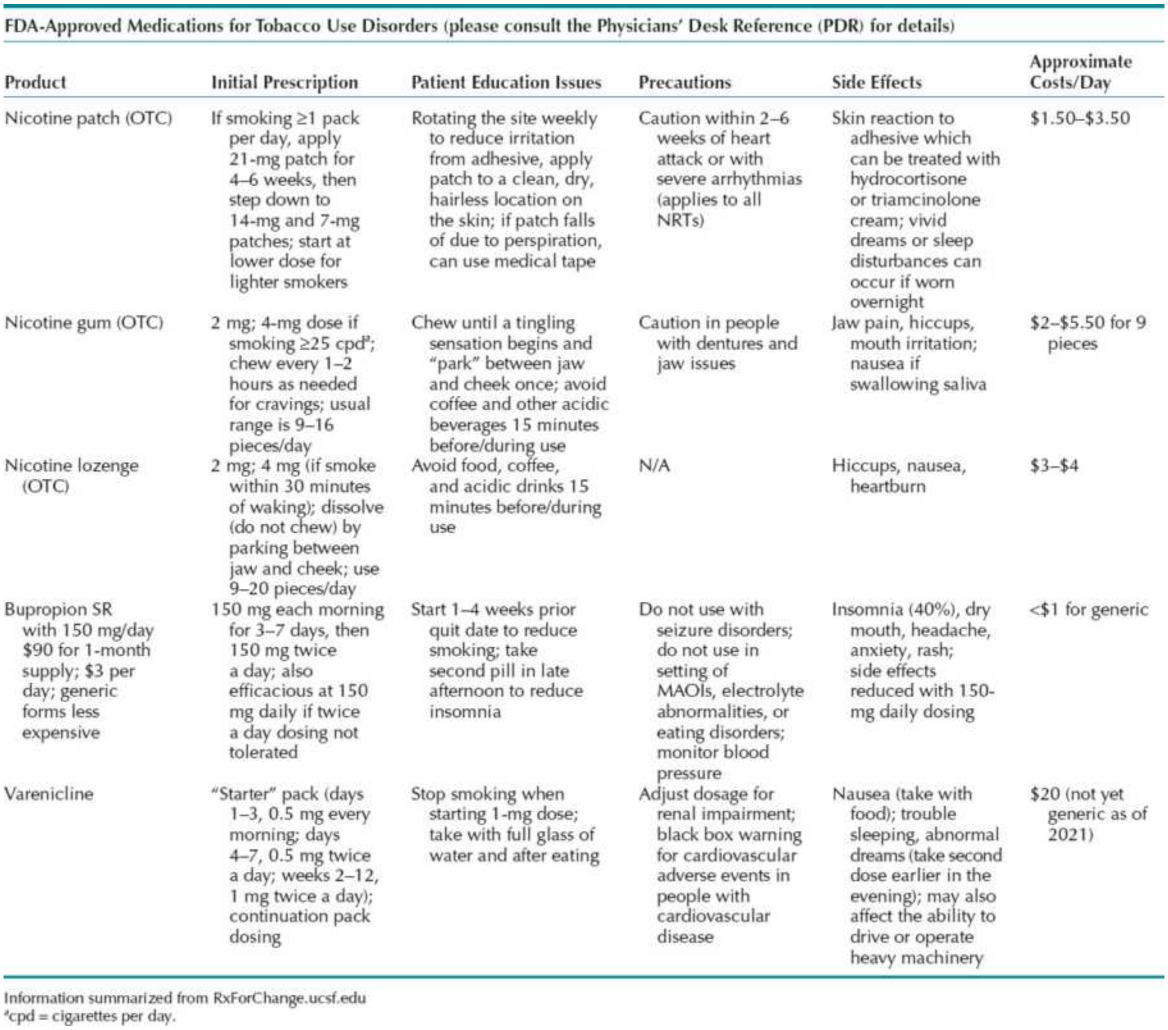

FDA-Approved Medications for Tobacco Use Disorders (please consult the Physicians' Desk Reference (PDR) for details)

Product	Initial Prescription	Patient Education Issues	Precautions	Side Effects	Approximate Costs/Day
Nicotine patch (OTC)	If smoking ≥1 pack per day, apply 21-mg patch for 4–6 weeks, then step down to 14-mg and 7-mg patches; start at lower dose for lighter smokers	Rotating the site weekly to reduce irritation from adhesive, apply patch to a clean, dry, hairless location on the skin; if patch falls of due to perspiration, can use medical tape	Caution within 2–6 weeks of heart attack or with severe arrhythmias (applies to all NRTs)	Skin reaction to adhesive which can be treated with hydrocortisone or triamcinolone cream; vivid dreams or sleep disturbances can occur if worn overnight	$1.50–$3.50
Nicotine gum (OTC)	2 mg; 4-mg dose if smoking ≥25 cpd[a]; chew every 1–2 hours as needed for cravings; usual range is 9–16 pieces/day	Chew until a tingling sensation begins and "park" between jaw and cheek once; avoid coffee and other acidic beverages 15 minutes before/during use	Caution in people with dentures and jaw issues	Jaw pain, hiccups, mouth irritation; nausea if swallowing saliva	$2–$5.50 for 9 pieces
Nicotine lozenge (OTC)	2 mg; 4 mg (if smoke within 30 minutes of waking); dissolve (do not chew) by parking between jaw and cheek; use 9–20 pieces/day	Avoid food, coffee, and acidic drinks 15 minutes before/during use	N/A	Hiccups, nausea, heartburn	$3–$4
Bupropion SR with 150 mg/day $90 for 1-month supply; $3 per day; generic forms less expensive	150 mg each morning for 3–7 days, then 150 mg twice a day; also efficacious at 150 mg daily if twice a day dosing not tolerated	Start 1–4 weeks prior quit date to reduce smoking; take second pill in late afternoon to reduce insomnia	Do not use with seizure disorders; do not use in setting of MAOIs, electrolyte abnormalities, or eating disorders; monitor blood pressure	Insomnia (40%), dry mouth, headache, anxiety, rash; side effects reduced with 150-mg daily dosing	<$1 for generic
Varenicline	"Starter" pack (days 1–3, 0.5 mg every morning; days 4–7, 0.5 mg twice a day; weeks 2–12, 1 mg twice a day); continuation pack dosing	Stop smoking when starting 1-mg dose; take with full glass of water and after eating	Adjust dosage for renal impairment; black box warning for cardiovascular adverse events in people with cardiovascular disease	Nausea (take with food); trouble sleeping, abnormal dreams (take second dose earlier in the evening); may also affect the ability to drive or operate heavy machinery	$20 (not yet generic as of 2021)

Information summarized from RxForChange.ucsf.edu

[a] cpd = cigarettes per day.

Transcranial Magnetic Stimulation. The FDA has approved transcranial magnetic stimulation (TMS) for major depressive disorders in 2008, and in 2020 TMS was approved for use in smoking cessation in adults. This is a novel approach with more community-based experience needed to understand how it is being adapted and used in treatment programs.

Combined Psychosocial and Pharmacologic Therapy

Combining medication and behavior therapy increases quit rates over either therapy alone. Also, combination medications have higher efficacy than monotherapy. Combination examples include using nicotine patch and short-acting NRT, bupropion and NRT, and varenicline and bupropion or varenicline and NRT. The USPHS Clinical Practice Guidelines recommend, contingent on the absence of medical contraindication, the use of at least one of the seven FDA-approved front-line agents for smoking cessation in conjunction with counseling to increase the likelihood of a successful quit attempt. Patients should be encouraged to use both pharmacologic and psychosocial treatment approaches; however, medications can be effective in the absence of psychosocial therapy; thus, the latter should not be a requirement for the former.

Ms. H is a 45-year-old patient with schizophrenia who smokes 35 cigarettes per day. She began her cigarette use at approximately 20 years of age during the prodromal stages of her first psychotic break. During the first 20 years of treatment, no psychiatrist or physician advised her to stop smoking, largely because they believed that she could not stop.

At 43 years of age, her primary care physician recommended smoking cessation, especially with her history of asthma and heart disease. The patient attempted to stop on her own, but this lasted only 48 hours, partly because her housemates and friends smoked. During a routine medication check, her psychiatrist *asked* about her smoking history and *advised* that she stop smoking, and the patient described her prior attempts. The psychiatrist *assessed* for motivation, using a 1 to 10 scale of motivation to quit and discussing where Ms. H was on the scale. He also had her breathe into a CO meter to demonstrate decreased oxygen capacity of her lungs and how that affects her asthma and heart disease. The psychiatrist and the patient discussed ways to avoid smokers and had the patient announce her intent to quit and request her friends to try not to smoke around her and to offer encouragement for her attempt to quit. The psychiatrist also noted that Ms. H became irritable, slightly depressed, and restless and had insomnia during prior cessation attempts and thus recommended medications for *assistance*. The patient chose to use nicotine patch plus nicotine gum as needed.

The psychiatrist had the patient call 2 days after her quit attempt as had been *arranged* at the appointment. At this point, the patient stated that the patch and gum were helping. One week later, the patient returned after having relapsed back to smoking. The psychiatrist praised the patient for not smoking for 4 days. After some motivation enhancement, Ms. H was still reluctant to try again immediately.

At the next routine visit 2 months later, the psychiatrist *asked* about smoking again, as it was in the assessment and treatment plan. After *advice* and *assessment*, the psychiatrist and patient listed several activities that the patient could do to avoid being around friends who smoked, phoned the patient’s boyfriend to ask him to assist the patient in stopping, asked the nurses of the inpatient ward to call the patient to encourage her, plus they *arranged* for the patient to be in a support group for the next 4 weeks. This time the psychiatrist prescribed the medication varenicline. The patient was followed with 15-minute visits for each of the first 3 weeks. She had two “slips” but did not go back to smoking and remained an ex-smoker.

Ms. H’s psychiatrist was correct in using pragmatic plans to help the patient overcome specific problems, in following the patient with short visits or phone calls, and in recommending medication. The total amount of time spent with the patient on smoking was approximately 3 hours, and there was reimbursement; more importantly, the intervention was important to decreasing her morbidity and mortality and was unlikely to be given by other care providers.

Tobacco Use Disorder Treatment of Psychiatric Patients and Other Special Populations

Psychiatric patients who are trying to stop smoking should be encouraged to undertake psychosocial and pharmacologic therapies. These patients experience greater withdrawal symptoms and have a poorer ability to stop as a result of greater smoking levels and more efficient smoking (inhaling more deeply), both resulting in higher levels of nicotine. In addition, these patients should be followed more regularly to detect any possible exacerbation of their psychiatric problem; often the perceived exacerbation results from withdrawal from nicotine and can be managed with NRT. Research demonstrates that smokers with mental illness do not have worsening of clinical symptoms of unipolar depression, bipolar disorder, post-traumatic stress disorder, or schizophrenia with tobacco treatment. Patients on smoke-free psychiatric wards can have withdrawal symptoms that require treatment, and pharmacotherapy is effective for psychiatric populations, including severe mental illness and SUDs. Treating TUDs in psychiatric patients not only reduces illness and lengthens life but can also present other benefits such as more discretionary income, less discrimination, and more options for relationships and employment.

People with SUDs should also be encouraged to quit smoking. Current evidence suggests that, for the large majority of (but not all) patients, stopping smoking does not threaten sobriety. Some newer studies suggest that quitting smoking can increase the likelihood of maintaining sobriety from other drugs.

Nonabstinence Goals

Given the highly addictive nature of tobacco, some tobacco users may not be able to or want to try to fully quit tobacco. The idea of using less to have less harm or to use an e-cigarette or another nicotine product is raised by some tobacco users. Harm reduction (using less tobacco) has been a method of gradual reduction for lower motivated individuals; however, this strategy often comes with compensatory smoking (i.e., smoking each cigarette more intensely to obtain the necessary nicotine), which appears to minimize any potential benefit of these strategies. Although some have feared that offering or agreeing to the patient's plan for reduction of smoking as a treatment goal would undermine motivation to quit, reduction in tobacco use appears to increase the chance of later quitting and may help motivate some tobacco users initially not willing to quit to try later to quit. Thus, agreeing to monitor a reduction method may not improve health but may lead to cessation by continuing to engage and work with the patient on the ultimate goal of cessation.

SPECIAL POPULATIONS

Expectant Smokers

Smoking and secondhand smoke in pregnancy or in women wanting to get pregnant are associated with multiple increased risks including difficulty becoming pregnant, risks of miscarriage, risk of SIDS, stillbirth, prematurity, low-birth weight, perinatal morbidity such as tissue damage (particularly in the lung and brain), and mortality. Furthermore, children of smoking mothers are twice as likely to become smokers themselves. Addressing smoking in pregnant women and women who wish to be pregnant can have a major impact, as this is often a life event that motivates half of affected smokers to quit. USPHS guidelines recommend that clinicians ask all pregnant women about tobacco use, advise tobacco cessation at all gestational ages, and provide behavioral interventions (motivational interviewing, cognitive behavioral therapy, 5As, quit lines) for those who smoke. The same should be done postdelivery, as over half of women relapse after delivery. The USPHS has concluded that current evidence is insufficient to assess the balance of benefits and harms of nicotine replacement products or other pharmaceuticals for tobacco cessation during pregnancy. The American College of Obstetricians and

Gynecologists (ACOG) advises that NRT should be considered only after a detailed discussion with the patient of the known risks of continued smoking, the possible risks of nicotine replacement therapy, and need for close supervision. Further, they note that if nicotine replacement therapy is used, it should be with the clear resolve of the patient to quit smoking. A systematic review and meta-analysis to assess the safety of bupropion and varenicline in pregnancy found no evidence that either of these treatments might be harmful in pregnancy and no strong evidence to suggest safety, but available evidence is of poor quality. The ACOG guidelines recommend physicians counsel women about the risks of smoking and the benefits of cessation and discuss the resources available, and if medications are prescribed, a discussion of risks/benefits.

Adolescent Smokers

Adolescents are another vulnerable population. It is proposed that nicotine acts as a gateway drug on the brain via multiple pathways (smoking tobacco, passive tobacco smoke, or e-cigarettes). Earlier use of e-cigarettes is related to initiation of cigarette use. Therefore, adolescence is an important time to screen and treat. The Public Health Service guidelines initially recommend adolescent smokers be provided with counseling interventions to aid them in quitting smoking as this can double quit rates; however, absolute abstinence rates of those who received counseling are low (11.6% quit rate at 6 months). Medications for smoking cessation are not FDA approved for use in children/adolescents. It is noted that although NRT has been shown to be safe for adolescents to use, it is not currently recommended as a component of pediatric tobacco use interventions in the guidelines. More recently the USPHS recommended that primary care clinicians provide interventions, including education or brief counseling, to prevent initiation of tobacco use among school-aged children and adolescents, *but* that the current evidence is insufficient to assess the balance of benefits and harms of primary care-feasible interventions for the cessation of tobacco use among school-aged children and adolescents. As part of routine psychiatry intakes, substance use history, including tobacco and vaping history, is clinically indicated. Vaping is especially important to teens as, since 2014, e-cigarettes have been the most commonly used tobacco among U.S. middle and high school students.

EMERGING AREAS OF RESEARCH

While cigarettes have existed for decades, there are continued developments in research and health effects, particularly with respect to ENDS.

E-Cigarettes for Smoking Cessation

There is still insufficient evidence for using e-cigarettes as consumer products for smoking cessation. However, there is emerging evidence that, in RCTs, monitored free e-cigarettes as part of a therapeutic intervention may be associated with increased smoking cessation according to a meta-analysis. A separate meta-analysis found that odds of quitting cigarettes were 28% lower in those who used e-cigarettes compared with those who did not use e-cigarettes.

E-Cigarette, or Vaping, Product Use-Associated Lung Injury

E-cigarette, or vaping, product use-associated lung injury, also known as EVALI, was reported in 2,398 people between February 2019 and January 2020. 82% of patients hospitalized with EVALI reported THC use with strong linkage to vitamin E acetate. The Centers for Disease Control and Prevention (CDC) recommended that people not use THC-containing e-cigarette or vaping products, especially those acquired from informal sources such as friends, family members, or from in-person or online dealers, and that vitamin E acetate should not be added to any e-cigarette or vaping products. These cases have reduced and are something we should continue to be mindful of. While vitamin E may have been the cause of these cases, EVALI brings up the concern of using vaping devices with substances other than nicotine. For example, THC vaping more than doubled among adolescents between 2017 and 2019. E-cigarette flavors are also of concern as they can increase both the likelihood of dependence and more adverse health effects.

Emerging Treatment Approaches Using Technology

In addition to traditional methods of cessation assistance described earlier, new technologies beyond just quit lines have emerged. These include text-to-quit programs, apps, and digital therapies. Cochrane Database reviews show promising evidence on text messaging support. Internet-based treatments show moderate efficacy, but aside from increasing access, may not be better than other active smoking treatments. Data on real-time video counseling is very limited and more studies need to be done. Unfortunately, there is no strong evidence for use of “apps” currently, as well; however, a number of apps such as the Veteran’s Affair’s Apps are promising. Another interesting area of research is with nicotine vaccines, of which none are currently licensed for public use, but there are a number in development.

PREVENTION AND POLICY INITIATIVES

Policy initiatives to reduce smoking prevalence have focused on tobacco taxation, smoking restrictions, restriction of tobacco advertising, counter advertising via mass media, attacks on the tobacco industry as immoral, restriction of sales to minors, and warnings on cigarette packages. Among these, the effect of taxes is especially large and replicable and influences initiation among youth.

A major policy advancement in many psychiatric settings has to do with tobacco-free public buildings and spaces. In 1993, JCAHO (now called The Joint Commission) banned smoking in hospitals. In addition to reducing access to smoking, thereby compelling a smoker into a quit attempt or at least reduction in cigarettes smoked, environmental tobacco smoke is reduced, and tobacco-related cues are eliminated, as well.

Given that TUDs have their onset typically in adolescence, general school-based tobacco prevention programs have been a logical approach to consider. There may be opportunities to consider high-risk adolescent groups, including those with psychiatric disorders being treated in mental health settings. Tobacco use often precedes mental illness and/or other SUDs. Strategies in school-based settings have focused on preventing the initiation of tobacco or nicotine-related products (such as e-cigarettes) through methods such as education (e.g., about risks of smoking), denormalizing smoking (e.g., correcting overestimates of prevalence of smoking), skills training (e.g., refusal training), and advocacy (e.g., students demonstrate again tobacco industry). Among these, only denormalizing and skills training have good empirical support; however, even with these interventions, the long-term effect of school-based programs is not large.

Policies to Address Nicotine and Health Equity

It is essential to consider tobacco-related health disparities to understand tobacco use in diverse populations and in implementing tobacco cessation interventions. The CDC (2015) published a document on best practices in health equity in tobacco prevention and control to support health care agencies and government agencies in ensuring equity in their interventions. The CDC defines tobacco-related health disparities as, “Differences that exist among population groups with regard to key tobacco-related indicators, including patterns, prevention, and treatment of tobacco use; the risk, incidence, morbidity, mortality, and burden of tobacco-related illness; and capacity, infrastructure, and access to resources; and secondhand smoke exposure” (CDC, 2014).

Tobacco use has declined significantly, in general, in the United States since the first surgeon general’s report on smoking in 1964. However, specific demographic groups continue to have higher rates of tobacco use and tobacco-related health conditions than the general population. These groups include rural populations, individuals of low socioeconomic status, members of the LGBTQ community, individuals with mental health and substance abuse disorders, American Indian/Native Alaskans, and individuals with high school or lower level of educational attainment.

The same factors that influence health inequity across health conditions influence inequities in tobacco use and cessation, such as access to cessation services in some communities or inequities in health insurance covering cessation services. A lack of cohesive policies that promote healthy communities create disparities for vulnerable populations. In addition, as advertising laws have changed for tobacco products, the tobacco industry has targeted advertising at specific groups where advertising is still allowed (e.g., promotion of sales in convenience stores; sponsoring social and cultural events). Other relevant health inequities in smoking use and cessation include discrimination and a general higher level of stress in groups facing health inequities, such as African Americans.

Considerations for Interventions. Health equity should be a goal of tobacco cessation interventions beyond general health promotion activities. For example, studies have shown that increasing the price of tobacco products reduces use in low-income populations. Health inequities could be reduced by updating advertising policies and laws by reducing exposure to advertising and promotions in convenience stores and by enacting stricter laws where people are exposed to secondhand smoke. Policies can be enacted to reduce the density of sale of tobacco products in communities by restricting sale of products in pharmacies, near schools, etc. Finally, increasing access to health insurance and the availability of cessation services covered by health insurance would improve health inequities, particularly in rural and other workforce shortage communities. For example, tobacco screening and cessation interventions should be incorporated into training programs and in all health care services.

When developing and implementing tobacco cessation programs, it is critical to understand cultural factors in use of tobacco and factors that promote or deter cessation. For example, tobacco has been used by many Native American tribes in their ceremonies. However, the use of commercial tobacco is high in Native American populations, and health care providers should understand the cultural relevance of ceremonial use while promoting cessation of commercial products that are much more harmful and addictive. It is important to include community members and cultural experts in developing and implementing cessation interventions to ensure they are culturally relevant to have the greatest impact. These partners can help incorporate language, values, beliefs, health practices, communication styles, and literacy levels of communities when developing culturally relevant interventions. These partners can also help prevent stereotypes or harmful messages that may not be considered by health care providers.

FURTHER READINGS

2008 PHS Guideline Update Panel, Liaisons, and Staff. Treating tobacco use and dependence: 2008 update U.S. Public Health Service Clinical Practice Guideline executive summary. *Respir Care*. 2008;53(9):1217.

American Association of Medical Colleges. *Physician Behavior and Practice Patterns Related to Smoking Cessation, Summary Report*. Association of American Medical Colleges; 2007.

American Cancer Society. *How are Cigars Different From Cigarettes? [Internet]*. American Cancer Society; 2020. [cited 2015 February 17]. https://www.cancer.org/cancer/risk-prevention/tobacco/is-any-type-of-smoking-safe.html

American College of Obstetricians and Gynecologists. Tobacco and nicotine cessation during pregnancy: ACOG Committee Opinion Number 807. *Obstet Gynecol*. 2020;135:e221–e229.

Anthenelli RM, Benowitz NL, West R, et al. Neuropsychiatric safety and efficacy of varenicline, bupropion, and nicotine patch in smokers with and without psychiatric disorders (EAGLES): a double-blind, randomised, placebo-controlled clinical trial. *Lancet*. 2016;387:2507–2520.

Antin TMJ, Annechino R, Hunt G, Lipperman-Kreda S, Young M. The gendered experience of smoking stigma: implications for tobacco control. *Crit Public Health*. 2016;27:443–454.

Apollonio D, Philipps R, Bero L. Interventions for tobacco use cessation in people in treatment for or recovery from substance use disorders. *Cochrane Database Syst Rev*. 2016;11(11):CD010274.

Assari S. Association of educational attainment and race/ethnicity with exposure to tobacco advertisement among US young adults. *JAMA Netw Open*. 2020:3:e1919393.

Balbuena L, Tempier R. Independent association of chronic smoking and abstinence with suicide. *Psychiatr Serv*. 2015;66(2):186.

Benowitz NL. Neurobiology of nicotine addiction: implications for smoking cessation treatment. *Am J Med*. 2008;121:S3.

Brewer JA, Elwafi HM, Davis JH. Craving to quit: psychological models and neurobiological mechanisms of mindfulness training as treatment for addictions. *Psychol Addict Behav*. 2013;27(2):366.

Buetner-Schmidt K, Miller DR, Maack B. Disparities in rural tobacco use, smoke-free policies, and tobacco taxes. *West J Nurs Res*. 2019;41:1184–1202.

Cabrera-Nguyen EP, Cavazos-Rehg P, Krauss M, Kim Y, Emery S. Awareness and use of dissolvable tobacco products in the United States. *Nicotine Tob Res*. 2016;18(5):857–863.

Caceres BA, Brody A, Luscombe RE, et al. A systematic review of cardiovascular disease in sexual minorities. *Am J Public Health*. 2017:107:e13–e21.

Cahill K, Lindson-Hawley N, Thomas KH, Fanshawe TR, Lancaster T. Nicotine receptor partial agonists for smoking cessation. *Cochrane Database Syst Rev*. 2016;2016(5):CD006103.

Cancer Trends Progress Report. National Cancer Institute, NIH, DHHS, Bethesda, MD, February 2019. https://progressreport.cancer.gov

Centers for Disease Control and Prevention (CDC). *Best Practices for Comprehensive Tobacco Control Programs—2014*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014. https://www.cdc.gov/tobacco/stateandcommunity/guides/pdfs/2014/comprehensive.pdf

Centers for Disease Control and Prevention (CDC). *Best Practices User Guide: Health Equity in Tobacco Prevention and Control*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2015. https://www.cdc.gov/tobacco/stateandcommunity/best-practices-health-equity/pdfs/bp-health-equity.pdf

Centers for Disease Control and Prevention (CDC). Quitting smoking among adults—United States, 2001–2010. *MMWR Morb Mortal Wkly Rep*. 2011;60(44):1513.

Centers for Disease Control and Prevention (CDC). Tobacco use and quitting among individuals with behavioral health conditions. 2020. Accessed June 7, 2021.

https://www.cdc.gov/tobacco/disparities/mental-illness-substance-use/index.htm

Chelladurai Y, Singh S. Varenicline and cardiovascular adverse events: a perspective review. *Ther Adv Drug Saf*. 2014;5(4):167.

Chou SP, Goldstein RB, Smith SM, et al. The epidemiology of DSM-5 nicotine use disorder: results from the national epidemiologic survey on alcohol and related conditions-III. *J Clin Psychiatry*. 2016;77(10):1404–1412.

Cobos-Campos R, de Lafuente AS, Apiñaniz A, Parraza N, Llanos IP, Orive G. Effectiveness of mobile applications to quit smoking: systematic review and meta-analysis. *Tob Prev Cessat*. 2020;6:62.

Cornelius ME, Wang TW, Jamal A, Loretan CG, Neff LJ. *Tobacco Product Use Among Adults – United States, 2019 [Internet]*. Centers for Disease Control and Prevention. Morbidity and Mortality Weekly Report (MMWR). 2020:69. https://www.cdc.gov/mmwr/volumes/69/wr/mm6946a4.htm?s_cid=mm6946a4_w

Drope J, Liber AC, Cahn Z, et al. Who's still smoking? Disparities in adult cigarette smoking prevalence in the United States. *CA Cancer J Clin*. 2018;68:106–115.

Dwyer-Lindgren L, Mokdad AH, Srebotnjak T, Flaxman AD, Hansen GM, Murray CJL. Cigarette smoking prevalence in US counties: 1996–2012. *Popul Health Metr*. 2014;12:5.

Glasser AM, Roberts ME. Retailer density reduction approaches to tobacco control: a review. *Health Place*. 2021;67:102342.

Hartmann-Boyce J, Cahill K, Hatsukami D, Cornuz J. Nicotine vaccines for smoking cessation. *Cochrane Database Syst Rev*. 2012;(8):CD007072.

Hartmann-Boyce J, Chepkin SC, Ye W, Bullen C, Lancaster T. Nicotine replacement therapy versus control for smoking cessation. *Cochrane Database Syst Rev*. 2018;5(5):CD000146.

Hitsman B, Moss TG, Montoya ID, George TP. Treatment of tobacco dependence in mental health and addictive disorders. *Can J Psychiatry*. 2009;54(6):368.

Kalkhoran S, Glantz SA. E-cigarettes and smoking cessation in real-world and clinical settings: a systematic review and meta-analysis. *Lancet Respir Med*. 2016;4(2):116–128.

Kandel DB, Kandel ER. A molecular basis for nicotine as a gateway drug. *N Engl J Med*. 2014;371(21):2038–2039.

Kendzor DE, Businelle MS, Reitzel LR, et al. Everyday discrimination is associated with nicotine dependence among African American, Latino, and White smokers. *Nicotine Tob Res*. 2014;16:633–640.

Khouja JN, Suddell SF, Peters SE, Taylor AE, Munafò MR. Is e-cigarette use in non-smoking young adults associated with later smoking? A systematic review and meta-analysis. *Tob Control*. 2020;30(1):8–15.

Koh H, Gracia JN, Alvarez ME. Culturally and linguistically appropriate services—advancing health with CLAS. *N Engl J Med*. 2014;371(3):198–201.

Krishnasamy VP, Hallowell BD, Ko JY, et al.; Lung Injury Response Epidemiology/Surveillance Task Force. Update: characteristics of a nationwide outbreak of e-cigarette, or vaping, product use-associated lung injury—United States, August 2019–January 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(3):90–94.

Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. *Cochrane Database Syst Rev*. 2002;3(3):CD001292.

Lasser K, Boyd JW, Woolhandler S, Himmelstein DU, McCormick D, Bor DH. Smoking and mental illness: a population-based prevalence study. *JAMA*. 2000;284.

Lawn S, Pols R. Smoking bans in psychiatric inpatient settings? A review of the research. *Aust N Z J Psychiatry*. 2005;39(10):866.

Leonardi-Bee J, Jere ML, Britton J. Exposure to parental and sibling smoking and the risk of smoking uptake in childhood and adolescence: a systematic review and meta-analysis. *Thorax*. 2011;66(10):847–855.

Lipari RN, Van Horn SL. *Smoking and mental illness among adults in the United States*. The CBHSQ Report: March 30, 2017. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, Rockville, MD; 2017.

Matkin W, Ordóñez-Mena JM, Hartmann-Boyce J. Telephone counselling for smoking cessation. *Cochrane Database Syst Rev*. 2019;5(5):CD002850.

McAndrew F, Thompson J, Fellows L, et al. Infant Feeding Survey 2010. The NHS HSE Info Centre; 2012. Accessed July 06, 2017. https://digital.nhs.uk/catalogue/PUB08694

McMillen R, Maduka J, Winickoff J. Use of emerging tobacco products in the United States. *J Environ Public Health*. 2012;2012:989474.

Miech RA, Patrick ME, O'Malley PM, Johnston LD, Bachman JG. Trends in reported marijuana vaping among US adolescents, 2017–2019. *JAMA*. 2020;323(5):475–476.

Neuman MD, Bitton A, Glantz SA. Tobacco industry influence on the definition of tobacco related disorders by the American Psychiatric Association. *Tob Control*. 2005;14(5):328.

Notley C, Gentry S, Livingstone-Banks J, Bauld L, Perera R, Hartmann-Boyce J. Incentives for smoking cessation. *Cochrane Database Syst Rev*. 2019;7(7):CD004307.

Peckham E, Brabyn S, Cook L, Tew G, Gilbody S. Smoking cessation in severe mental ill health: what works? An updated systematic review and meta-analysis. *BMC Psychiatry*. 2017;17(1):252.

Prochaska JJ, Delucchi K, Hall SM. A meta-analysis of smoking cessation interventions with individuals in substance abuse treatment or recovery. *J Consult Clin Psychol*. 2004;72(6):1144.

Prochaska JJ, Hall SM, Bero LA. Tobacco use among individuals with schizophrenia: what role has the tobacco industry played? *Schizophr Bull*. 2008;34:555.

Roberts E, Evins AE, McNeill A, Robson D. Efficacy and tolerability of pharmacotherapy for smoking cessation in adults with serious mental illness: a systematic review and network meta-analysis. *Addiction*. 2016;111:599–612.

Rodgman A, Perfetti TA. *The Chemical Components of Tobacco and Tobacco Smoke*. CRC Press, Taylor & Francis Group; 2009.

Rogers ES, Smelson DA, Gillespie CC, et al. Telephone smoking-cessation counseling for smokers in mental health clinics: a patient-randomized controlled trial. *Am J Prev Med*. 2016;50:518–527.

Schroeder MJ, Hoffman AC. Electronic cigarettes and nicotine clinical pharmacology. *Tob Control*. 2014;23:30.

Secades-Villa R, González-Roz A, García-Pérez Á, Becoña E. Psychological, pharmacological, and combined smoking cessation interventions for smokers with current depression: a systematic review and meta-analysis. *PLoS One*. 2017;12(12):e0188849.

Slopen N, Dutra LM, Williams DR, et al. Psychosocial stressors and cigarette smoking among African American adults in midlife. *Nicotine Tob Res*. 2012;14:1161–1169.

Taylor GMJ, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. *Cochrane Database Syst Rev*. 2017;9(9):CD007078.

Taylor G, McNeill A, Girling A, Farley A, Lindson-Hawley N, Aveyard P. Change in mental health after smoking cessation: systematic review and meta-analysis. *BMJ*. 2014;348:1151.

Thurgood SL, McNeill A, Clark-Carter D, Brose LS. A systematic review of smoking cessation interventions for adults in substance abuse treatment or recovery. *Nicotine Tob Res*. 2016;18:993–1001.

Turner E, Jones M, Vaz LR, Coleman T. Systematic review and meta-analysis to assess the safety of bupropion and varenicline in pregnancy. *Nicotine Tob Res*. 2019;21(8):1001–1010.

Tzelepis F, Paul CL, Williams CM, et al. Real-time video counselling for smoking cessation. *Cochrane Database Syst Rev*. 2019;2019(10):CD012659.

Unger JB, Soto C, Baezconde-Garbanati L. Perceptions of ceremonial and nonceremonial uses of tobacco by American-Indian adolescents in California. *J Adolesc Health*. 2006;38:443.e9–443.e16.

U.S. Department of Health and Human Services (USDHHS). *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*. Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014.

U.S. Department of Health and Human Services, Office of Minority Health. *National Standards for Culturally and Linguistically Appropriate Services in Health and Health Care: A Blueprint for Advancing and Sustaining CLAS Policy and Practice*. US Dept of Health and Human Services, Office of Minority Health; 2013.

U.S. Preventive Services Task Force; Owens DK, Davidson KW, Krist AH, et al. Primary care interventions for prevention and cessation of tobacco use in children and adolescents: US Preventive Services Task Force recommendation statement. *JAMA*. 2020;323(16):1590–1598.

E-Cigarettes, Vapes, and other Electronic Nicotine Delivery Systems (ENDS). https://www.fda.gov/tobaccoproducts/labeling/productsingredientscomponents/ucm456610.htm

Wang RJ, Bhadriraju S, Glantz SA. E-cigarette use and adult cigarette smoking cessation: a meta-analysis. *Am J Public Health*. 2021;111(2):230–246.

Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. *Cochrane Database Syst Rev*. 2019;10(10):CD006611.

Williams JM, Ziedonis DM, Vreeland B, et al. A wellness approach to addressing tobacco in mental health settings: learning about healthy living. *Am J Psychiatr Rehabil*. 2009;12:352.

World Health Organization (WHO). Western Pacific Regional Office (WPRO), 2020-last update, Tobacco Fact Sheet [Homepage of World Health Organization, Western Pacific Region], [Online]. 2020. https://www.who.int/news-room/fact-sheets/detail/tobacco

Young JR, Galla JT, Appelbaum LG. Transcranial magnetic stimulation treatment for smoking cessation: an introduction for primary care clinicians. *Am J Med*. 2021;134(11):1339–1343.

Young-Wolff KC, Kline-Simon AH, Das S, Mordecai DJ, Miller-Rosales C, Weisner C. Smoking trends among adults with behavioral health conditions in integrated health care: a retrospective cohort study. *Psychiatr Serv*. 2016;67(9):996–1003.

Zevin S, Benowitz NL. Drug interactions with tobacco smoking. *Clin Pharmacokinet*. 1999;36(6):425.

Zhao-Shea R, Liu L, Pang X, Gardner PD, Tapper AR. Activation of GABAergic neurons in the interpeduncular nucleus triggers physical nicotine withdrawal symptoms. *Curr Biol*. 2013;23(23):2327.

Ziedonis D, Hitsman B, Beckham JC, et al. Tobacco use and cessation in psychiatric disorders: National Institute of Mental Health Report. *Nicotine Tob Res*. 2008;10(12):1691.

Ziedonis DM, Williams JM, Steinberg MB, Foulds J. Addressing tobacco addiction in office-based management of psychiatric disorders: practical considerations. *Prim Psychiatry*. 2006;13(2):51.

11.9 Opioid Use and Related Disorders: From Neuroscience to Treatment

SEAN X. LUO, M.D., PH.D., AND ADAM BISAGA, M.D.

The term **opioid** describes a class of psychoactive compounds, both naturally occurring and chemically synthesized, that are related to *opiates*: alkaloid compounds found as natural products in the opium poppy plant, *Papaver somniferum*. Opioids are among the oldest known drugs worldwide: opium poppy use predates recorded history. Opioids can be used medically for pain relief but are also commonly misused by means of nonprescribed use of pharmacologic opioids or illicitly obtained heroin or synthetic opioids such as fentanyl. Opioid use disorders (OUDs) affect a wide range of individuals around the world, and the United Nations Office on Drugs and Crime (UNODC) estimates that around 1.5% of the global population in 2019 had nonmedical use of any opioid-derived compounds, representing an increase from around 0.5% in 2009. In the United States, OUD has dramatically increased in the past few decades and now constitutes a large public health problem affecting the young and the old, the affluent and the poor, and professionals and the unemployed. At the same time, there have been significant advances in knowledge about opioids and their effect on the central nervous system (CNS). Practical and effective treatment strategies that target specific brain abnormalities have emerged. There is now a consensus in the medical community that OUD is a chronic, relapsing disorder involving multiple brain circuits and is amenable to pharmacologic and psychosocial interventions.

HISTORICAL BACKGROUND

The Greek word "opos," which means juice, suggests people may have first discovered the liquid from the unripe poppy seed capsule when it is notched. The written record of the early history of opium use and misuse is ambiguous. The preparation described by Homer in *The Odyssey,* given by Helen to Telemachus and his friends to help them forget their grief over Odysseus' absence, was attributed to Homer's imagination by Theophrastus (300 BC). Sumerians cultivated poppies and isolated opium at the end of the third millennium BC Arab traders brought opium to India and China around the 8th century AD, and between the 10th and 13th centuries, opium traveled from Asia Minor to Europe. Thomas Sydenham noted in 1680 that "Among the remedies which it has pleased Almighty God to give man to relieve his sufferings, none is so universal and efficacious as opium." On the other hand, written records of drug misuse and tolerance can be found in Turkey, Egypt, Germany, and England, and the problematic practice of smoking opium reached its apogee in China in the mid-17th century after tobacco smoking was banned. Efforts by the Chinese government to criminalize opium use conflicted with British and French commercial imperatives, leading to the Opium Wars in the 19th century, which greatly influenced the postcolonial era geopolitical landscape.

Sertürner used advanced chemical methods in 1806 to isolate the active ingredient in opium, naming it morphine after Morpheus, the god of dreams. Codeine was isolated from opium a few years later. Over the next century, these purified compounds gradually replaced crude opium for medicinal purposes, although nonmedical use of opium, especially through smoking, still persisted in many parts of the world. The first semisynthetic opium derivative—diacetylmorphine or heroin—was introduced into medicine in 1898 and marketed as an effective cough suppressant, and at the time, thought to have very little misuse liability. The first purely synthetic morphine-like opioid, meperidine, was serendipitously discovered in 1939 in a search for atropine replacement. Methadone was discovered in the 1940s and introduced to medical practice in the subsequent decades.

The recognition of opioid withdrawal and addiction by the medical community, especially in the United States, generally began in the 19th century. It was noted at the time that morphine addiction was common among White middle-class women who were given it frequently for numerous medical symptoms such as neurasthenia and cough, as well as in the postpartum period. A variety of remedies were used to treat opioid withdrawal and “cure” morphine addiction. However, these were often more dangerous than untreated withdrawal and were ineffective in preventing relapse. A variety of other opioid-related issues became prominent in the early 20th century: prevalent opium smoking by Chinese immigrants working on railroads, intravenous (IV) use of opioids with the newly introduced hypodermic needle, an increase in illegal activities surrounding drug use, and the importation of opioids from other countries. A variety of medications on the market included opioids components surreptitiously and were mostly unregulated. As drug use was gradually linked to crime, the U.S. Federal Government first passed the Pure Food and Drug Act of 1906, which required accurate labeling of component contents. The subsequent Harrison Narcotics Tax Act of 1914 demarcated a starting point for federal regulation and taxation of the production, importation, and distribution of opiates and Coca products. The Harrison Act had a profound impact on persons exhibiting physical and psychological symptoms of tolerance and withdrawal because it was interpreted by the Treasury Department to exclude the provision of using opioids to treat these symptoms for medical purposes. Clinics that had been established specifically to provide therapeutic morphine to patients with opioid use problems were shut down by the early 1920s. Doctors were urged to avoid treating patients with these symptoms, and some were prosecuted for doing so. Detoxification efforts were plagued by frequent relapses, and many sought alternative access to opioids. As medical prescription of opioids decreased, illicit traffic increased substantially, providing mostly morphine and heroin to individuals with addiction. Users were frequently arrested for the possession and sale of opioids and incarcerated.

In the 1930s, the U.S. Public Health Service established two federal hospitals—at Lexington, Kentucky, and Fort Worth, Texas—to deal with the growing number of federal prisoners who were also addicted to opioids. These hospitals provided long-term residential treatment for prisoners and some voluntary patients, and a significant amount of early addiction research was conducted there. Follow-up studies of these prisoners found that relapse rates were high even after long periods of abstinence and treatment. In light of the lack of effective treatment strategies and increasing criminal activities surrounding drug use, worldwide penalties for the illicit use of opioids increased after World War II. The Single Convention on Narcotic Drugs of 1961 established an international treaty to prohibit the production and supply of specific drugs ratified by most member countries. As part of World Health Organization (WHO), the United Nations Commission on Narcotic Drugs was empowered to add, remove, and transfer drugs among the treaty's four schedules of controlled substances. This treaty also required the criminalization of cultivation, production, distribution, possession, and trafficking of scheduled opioids. Governmental regulation of opioids through criminal law became a worldwide phenomenon. In the United States, the Controlled Substance Act of 1970 served as the national implementing legislation for the Single Convention on Narcotic Drugs, and heroin was categorized as a Schedule I drug, the most stringent category of regulation (indicating no valid medical use). Other opioids, such as morphine, oxycodone, and methadone, were designated Schedule II drugs.

Since the 1960s, it has been increasingly recognized that legal regulation is an insufficient intervention both for individuals with problematic opioid use and for the community as a whole. Attention has been shifting toward the development of better treatment programs for opioid-dependent individuals. The Synanon organization was founded in California in 1958 and constituted an early model of a therapeutic community. The 12-step programs, such as Narcotics Anonymous (NA), which grew out of Alcoholics Anonymous (AA), first emerged in Los Angeles in the early 1950s. Since the late 1960s, federal and state agencies have increasingly focused on treatment research. Methadone maintenance began as a research program at Rockefeller University in

1964. This effective treatment was delivered in the setting of outpatient methadone maintenance clinics, a model that subsequently expanded and became widely adopted throughout the United States and many parts of the world. Problematic opioid use became a condition amenable to medical treatment by physicians and allied health care professionals.

Naltrexone was synthesized in the 1960s and later used to treat opioid addiction; it was approved by the U.S. Food and Drug Administration (FDA) in 1984 as an oral tablet and in 2006 as a long-acting injection. Buprenorphine was initially formulated in the United Kingdom in 1978 as an injection to treat severe pain, and a sublingual formulation was released in 1982, providing a convenient treatment for OUDs introduced in the 1990s. Long-acting preparations of buprenorphine became later available for the treatment of OUD, including an implant released in 2016 and an injection released in 2019. As the availability of residential addiction treatment decreased significantly in the 1990s, the U.S. Congress passed the Drug Abuse Treatment Act of 2000, which permitted physicians with qualifying training to prescribe approved Schedule III, IV, or V medications (such as buprenorphine) for the treatment of addiction or detoxification outside of a methadone clinic setting. Since this law was passed, office-based treatment for OUD with buprenorphine has expanded substantially as the number of individuals presenting for treatment increased with the growing epidemic of misuse of prescription opioids in the United States. Comprehensive Addiction and Recovery Act of 2016 (CARA) increased the treatment limit from 30 cases to 100 cases per year, and SUPPORT for Patients and Communities Act of 2018 (SUPPORT Act) allowed expansion to 275 patients per physician after 1 year of 100 cases. Moreover, the CARA act allowed additional prescribers, qualifying nurse practitioners and physician assistants, to prescribe buprenorphine for the treatment of OUD.

COMPARATIVE NOSOLOGY

The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* divides opioid and related disorders into five main diagnoses: OUD, opioid intoxication, opioid withdrawal, other opioid-induced disorders, and unspecified opioid-related disorder.

Opioid use disorder is diagnosed for users of opioids who meet at least 2 of the 11 criteria in the previous 12-month period. These criteria have physiologic, behavioral, and cognitive components and have been organized into four categories: impaired control, social impairment, risky use, and pharmacologic changes. This disorder primarily reflects a pattern of compulsive, prolonged self-administration of opioid substances with no legitimate medical purpose or at doses greatly exceeding the amount needed for a medical condition, with continuing use of opioids despite significant substance-related problems. The OUD diagnosis is further specified by a set of additional modifiers: (1) the duration of remission: either *early remission*, where none of the criteria are met for more than 3 months but less than 12 months, or *sustained remission*, where none of the criteria have been met for more than 12 months; (2) whether the individual is *on maintenance therapy*: if the individual is maintained on a medication specifically for the treatment of OUD; and (3) whether the individual is in a *controlled environment*: where access to opioids is restricted. The diagnosis is also coded based on symptom severity. *Mild* OUD denotes the presence of two to three symptoms, *moderate*, four to five symptoms, and *severe*, six or more symptoms.

Opioid intoxication is diagnosed when an individual who recently used an opioid exhibits clinically significant problematic behavioral or psychological changes, such as dysphoria, impaired judgment, or psychomotor agitation or retardation that developed during, or shortly after, opioid use. These individuals also develop symptoms such as drowsiness or coma, slurred speech, and impaired attention and memory. *Opioid withdrawal* is diagnosed when an individual ceases or reduces opioid use or receives an opioid antagonist after a period of opioid use. Three of the nine withdrawal symptoms must develop and cause clinically significant distress for this diagnosis to apply. In addition, DSM-5 also contains unspecified opioid-related disorder to denote clinically significant opioid-related disorders that do not meet any criteria as above, as well as substance/medication-induced depressive disorder, substance/medication-induced psychotic disorder, substance/medication-induced bipolar and related disorder, and substance/medication-induced anxiety disorder, which describe a set of psychotic, depressive, bipolar-related or anxiety-related symptoms that occur soon after exposure or withdrawal from opioids (or other substances).

The DSM-5, released in 2013, represents a significant revision and update of criteria from the DSM-IV-TR and ICD-10 (WHO: International Statistical Classification of Diseases and Related Health Problems 10th revision) definitions of opioid-related disorders. Substance use disorder in DSM-5 combines the DSM-IV categories of substance abuse and dependence into a single disorder measured on a continuum from mild to severe, as it has been increasingly observed that these conditions are not dichotomous entities. In DSM-5, each substance is now addressed as a separate use disorder, but the overarching set of criteria is the same. Whereas DSM-IV only requires 1 symptom to diagnose a substance abuse disorder, DSM-5 requires at least 2 symptoms from a list of 11. Drug craving is added to the list, and substance-related legal problems are eliminated because legal contexts differ across cultures and DSM-5 aims to provide a set of criteria that can be applied internationally.

There is a high level of agreement between the DSM-5 and ICD-11 and both have as their central feature an emphasis on the compulsive aspect of drug-using behavior. However, ICD-10 continues to make the distinction between abuse (harmful use category) and dependence. In DSM-5, the code for OUD, mild now corresponds to the ICD-10 definition of harmful use of opioids (F11.10), while the DSM-5 criteria for OUD, moderate and severe, correspond to the ICD-10 code for opioid dependence syndrome (F11.20). It is important to note that within both classification systems, physical dependence, as evidenced by tolerance or withdrawal, is a criterion that is neither necessary nor sufficient for a diagnosis of either OUD or opioid dependence. One of the rationales for the new DSM-5 terminology of OUD, rather than opioid dependence, is to avoid the confusion of symptoms of physical dependence versus the syndrome of drug dependence, as well as avoiding the potentially stigmatizing label of *addiction*, which is commonly used in the lay literature to describe more severe forms of a substance use disorder but also has many other meanings. A more neutral label of a “use disorder” was selected, and the diagnosis of *addiction* does not appear in either DSM-5 or ICD-10, though it remains in the title of the DSM-5 “Substance-Related and Addictive Disorders” chapter.

ICD-10 differs from DSM-5 in that the concept of harmful use is limited to physical and mental health (e.g., hepatitis, overdose, and skin abscess) and specifically excludes social impairments with the reasoning: “Harmful patterns of use are often criticized by others and frequently associated with adverse social consequences of various kinds. The fact that a pattern of use or a particular substance is disapproved of by another person or by the culture, or may have led to socially negative consequences such as arrest or marital arguments, is not in itself evidence of harmful use.”

EPIDEMIOLOGY

Extensive epidemiologic surveys from both the United States and around the world track trends in opioid use and misuse. In the United States, there are several national and public surveillance systems that provide useful information on different aspects of opioid use and addiction over time: the annual National Survey on Drug Use and Health (NSDUH) estimates the prevalence of substance use, abuse, and dependence within community-dwelling persons aged 12 years and older; the Monitoring the Future (MTF) survey reports on the prevalence of substance use as well as attitudes about drug use among a national sample of 8th-, 10th-, and 12th-grade students, and the Treatment Episode Dataset (TEDS) collects demographic and drug use data from persons of all ages admitted to substance abuse treatment facilities receiving state or federal funding (with some exceptions). Internationally, the World Drug Report provides an annual review of the major developments in drug markets for the various drug categories, from production to consumption. These publicly available data sources, combined with new methods in data mining and visualization, provide an easier way for researchers and public health workers to respond more rapidly to emerging trends in opioid misuse. In this section, heroin and other opioids are discussed separately, as large epidemiologic differences have been observed for these two categories of substances.

Prevalence of Heroin Use and Dependence

In 2019, the NSDUH reported the past-year prevalence of heroin use in the United States among persons aged 12 years or older to be approximately 745,000 people or 0.3% of the population, and the overall diagnosis of OUD affected 438,000 people or 0.2%. A reported decline in the survey was noted from 2015 to 2019, particularly with a decline in young adults aged 18 to 25 years, from 1.5% in 2015 to 0.7% in 2019. It is important to note that a substantial portion of the surveyed participants endorsed symptoms of misuse, such as using medication that is not prescribed or for good drug effect, but only a minority of these individuals receive a diagnosis of OUD. Nevertheless, the prevalence of heroin use and opioid misuse, in general, does appear to have peaked between 2011 and 2015 (Fig. 11.9–1A,B). The overall past-year opioid misuse peaked in 2015 at 8.7% and decreased to 5.3% in 2019, and pain reliever misuse constitutes the majority of individuals who use nonprescribed opioids (96.6%). The 2020 to 2021 coronavirus (SARS-CoV-2) pandemic may affect the overall trend of this decrease, and future data will shed light also on changes in the prevalence of high-potency opioids in the general population—something that has not been specifically surveyed so far by NSDUH.

For adolescents (8th, 10th, and 12th grade), the MTF survey showed some stabilization of the prevalence of opioid use. The lifetime prevalence of heroin use in 8th graders was 0.5% in 2020 and 0.3% and 0.4%, respectively, for 10th and 12th graders. For narcotics other than opioids, lifetime prevalence has decreased from 6.8% in 2017 to 5.3% in 2019 and 2020 in 12th graders. Past year reported use also decreased from 4.2% in 2017 to 2.1% in 2020. Extended-release (XR) oxycodone use in the past year was reported to be 0.9% in 8th graders in 2020, 1% in 10th graders, and 2.4% in 12th graders, with prevalence not substantially changed in the past 5 years.

According to a TEDS report (last available for 2017), opioids now constitute the largest fraction of inpatient admissions aged 12 years and older in public substance use disorder treatment settings and have overtaken alcohol and marijuana since 2015. Opioids constituted 34% of all admissions in 2017, with heroin at 26.6%, the largest subgroup. Alcohol constituted 29.5%, and cannabis 12.5%. Importantly, TEDS documented an overall increase in medication-assisted opioid therapy admission for heroin, from 93,108 out of 292,265 (31.8%) in 2007 to 210,853 out of 533,394 (39.4%), but the majority of admissions were still not with planned medication-assisted therapy. There is a predominance of male gender for patients receiving medication-assisted opioid therapy (64.5%), and no ethnic disparity was observed, with White (non-Hispanic) at 60.6% of admissions, Black (non-Hispanic) at 17.5%, and Hispanic of any origin at 14.2% (see Fig. 11.9–1).

Trends in Overdose

According to the Centers for Disease Control (CDC), almost 450,000 people have died in the United States from overdoses involving any opioid, including prescription and illicit opioids, from 1999 to 2019. The overdose deaths have been historically associated with three waves: the first wave began with increased prescription opioids, and starting in 2010, there was a rapid increase in overdose death involving heroin. Since 2013, significant increases in overdose deaths involved synthetic and illicitly manufactured high-potency opioids, such as fentanyl. While the overdose stabilized somewhat in 2017 to 2018 (Fig. 11.9–2), more recently there has been a resurgence of overdose associated with these high-potency opioids as well as overdose that involve synthetic opioids combined with methamphetamine or cocaine.

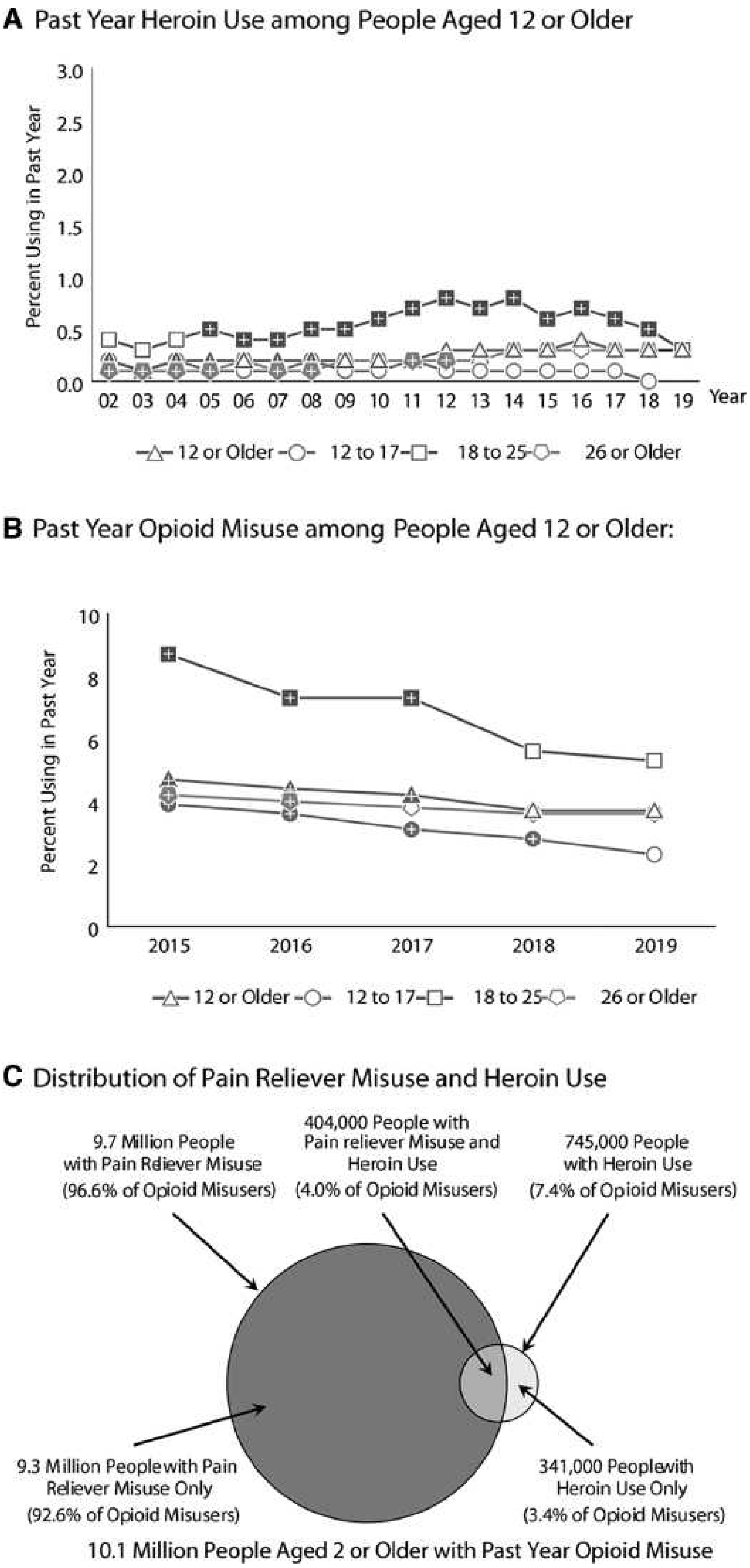

A Past Year Heroin Use among People Aged 12 or Older

B Past Year Opioid Misuse among People Aged 12 or Older:

C Distribution of Pain Reliever Misuse and Heroin Use

FIGURE 11.9–1. Epidemiology of opioid use in the United States. **A:** Past year use of heroin in several age groups showing a peaking of use and most common use in the age group 18 to 25 at around 1% in 2014, with a gradual

decrease in most recent survey years. **B:** A corresponding decrease in overall opioid misuse (nonprescribed) has been seen in all age groups. **C:** Most of the opioid misuse in the United States involves prescription pain relievers rather than heroin. (Adapted from U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration. 2019 National Survey of Drug Use and Health [NSDUH] Releases.)

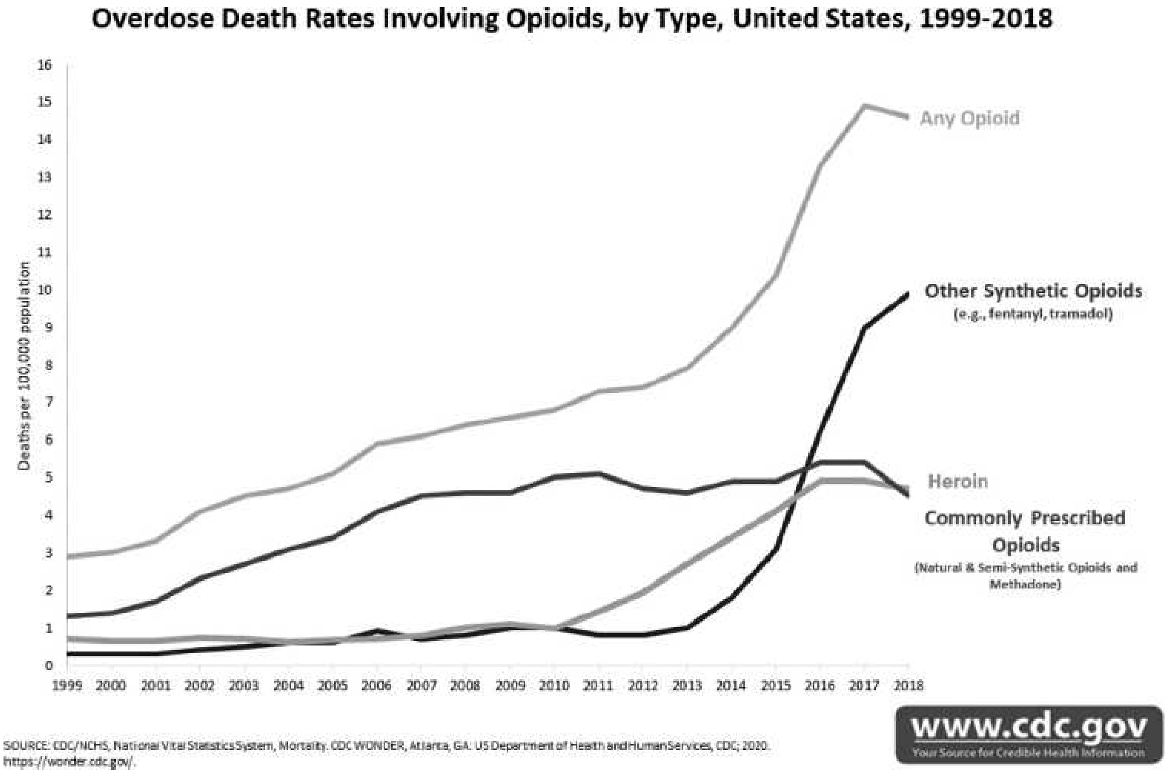

SOURCE: CDC/NCHS, National Vital Statistics System, Mortality. CDC WONDER, Atlanta, GA: US Department of Health and Human Services, CDC; 2020. https://wonder.cdc.gov/.

FIGURE 11.9–2. The increase and stabilization of opioid overdose deaths in the United States since 1999. The overdose deaths arrived in three waves, the initial rise in commonly prescribed opioids prior to 2005, then an increase in heroin between 2005 and 2014, and most recently, a rapid rise of newer and higher-potency synthetic opioids since 2013. (See eBook for color images.) (Adapted from CDC/NCHS National Vital Statistics System, Mortality. CDC WONDER, Atlanta, GA: US Department of Health and Human Services, CDC; 2020. https://wonder.cdc.gov/)

Worldwide Trends

The WHO continues to monitor drug use trends worldwide with the World Drug Report, last published in 2020. In 2018, 57.8 million people globally were estimated to have used opioids in the past year, a figure that includes those who had used opiates (30.4 million) and those who had misused pharmaceutical opioids. This corresponds to a past-year prevalence of opioid use of 1.2% of the global population aged 15 to 64 years.

Nonmedical use of synthetic opioids continues to fuel a worldwide public health crisis. In West, Central, and North Africa, there has been a large increase in tramadol use, and in North America, the prevalence of fentanyl use has increased. The change of OUD affecting disability-adjusted life years increased dramatically (28%) from 2008 to 2017. The market for opiates in Central Asia and the Russian Federation has changed as well, with the number of people in treatment for opioid-related problems in Russia decreasing sharply. In particular, the worldwide COVID-19 pandemic of 2020 caused drug shortages at the retail level, including reports of heroin shortages in Europe, Southwest Asia, and North America. Pharmacologic opioid painkillers also had significant reported shortages. People with OUD and who use prescription and nonprescription opioids have a high risk of comorbidities such as chronic obstructive pulmonary disease and potentially increased risk for complications and mortality from COVID-19. The overall pattern is that while opioid use is highest in North America, Australia, and New Zealand, much of this opioid use is pharmaceutical painkillers. Heroin and opium use are more common in the Near East and South Asia (see Fig. 11.9–3).

ETIOLOGY AND PATHOGENESIS

OUD is a complex phenomenon that involves multiple etiologic factors at different levels of analysis: genetic and epigenetic influences, synaptic plasticity, alterations in neural circuits, deficits in coping skills, abnormal patterns of thoughts and behavior, and dysfunction in family and social surroundings. The abnormal behavior characteristic of addiction represents the manifestation of a common final pathway and a heterogeneous set of underlying causal factors. These factors influence each other in space and in time in complex and nonlinear ways. For example, peer pressure (a social factor) may commonly precipitate the initiation of heroin use, but differences in neural circuits, likely generated, in part, by differences in genetic variations, eventually yields a difference in drug-related experiences, specifically the balance between positive effects (e.g., euphoric) versus negative effects (e.g., nausea, vomiting) which determine further drug-taking behavior. The interaction between the effects of repeatedly self-administered substance and the resulting experiences subsequently causes synaptic plasticity, either reinforcing/strengthening or extinguishing/weakening selected brain circuitries, which in turn affect an individual's behavior—either continuing with a drug-using peer group or a withdrawal from the group.

Initiation of opioid use is dependent on social and cultural context. When opioids are injected or inhaled, the rapid onset of action in the brain induces a *rush* or *flash*, a sudden, brief but intensely pleasurable sensation lasting 1 to 2 minutes. This euphoric sensation subsequently leads individuals to seek repetition of the drug taking (reinforcement). Repeated drug taking and euphoria cause opioid tolerance; changes in neural circuitry make it harder to achieve the same intensity of euphoria with an equal amount of opioids. An abrupt cessation of use by an individual who has been using repeatedly over an extended period produces opioid withdrawal, a set of physical symptoms such as nausea, vomiting, and intense craving signifying the presence of physiologic dependence. The emergence of withdrawal on stopping use and euphoria with continuing use is the basis of a long-term learning process that assigns salience to external and internal cues and experiences associated with positive and negative drug effects (conditioning) and alters the behavioral repertoire of the addicted individual in the form of increased impulsivity or decreased response to social rewards.

Neurobiology of Opioid Use Disorders

Opioids and Opioid Receptors. A variety of opioid compounds exist (Fig. 11.9–4). *Morphine*, *codeine,* and *oxycodone* are commonly prescribed medications for pain relief but can also be diverted for nonmedical use. *Heroin* or diacetylmorphine is the opioid that is most commonly misused by injection; common street names include dope, horse, smack, and tar. Opium extracted from the poppy plant, which is often smoked, is more commonly misused in the Middle East and Asia. More recently, synthetic analogs of fentanyl and nonfentanyl compounds (AH-7921, MT-45, U-47700) with opioidergic activity have become more widely available. Most commonly encountered opioids act at one of the three main opioid receptor systems (mu, kappa, delta). The term *endorphin* refers to a subclass of endogenous opioids, peptides produced naturally by the human body, including dynorphins, enkephalins, endorphins, endomorphins, and nociceptin.

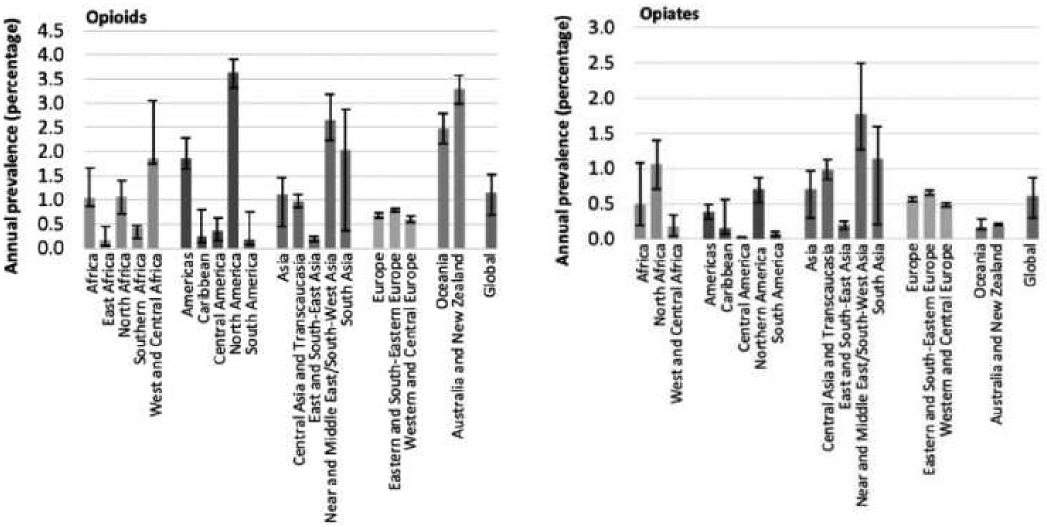

MAP 1 Most commonly used opioid, 2018 or latest available data

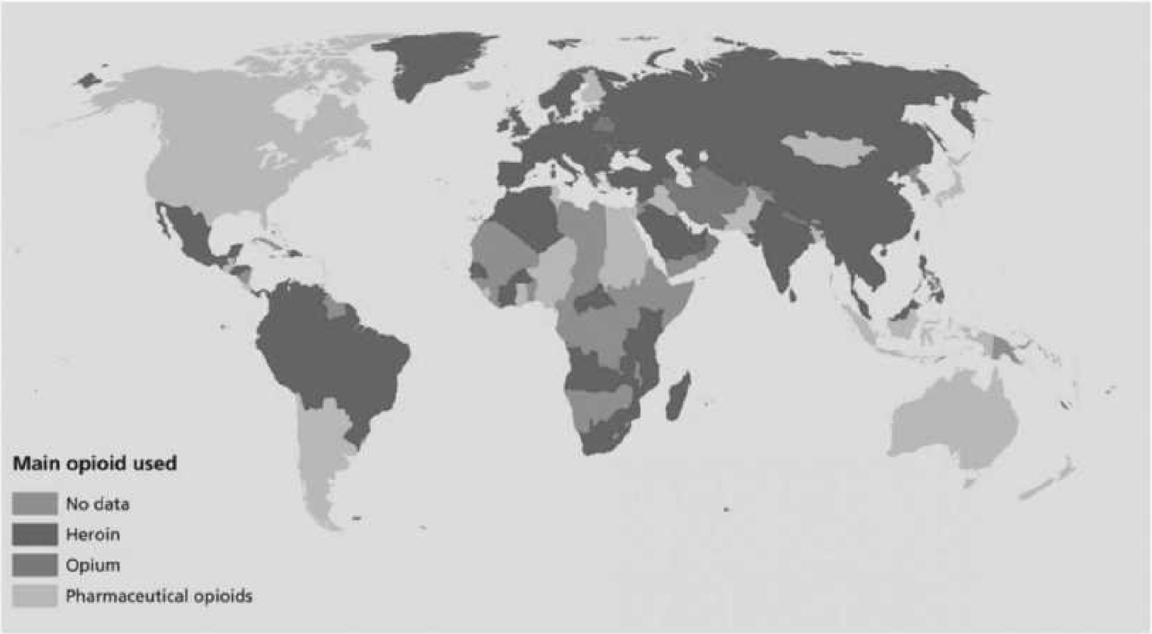

Source: UNODC, responses to the annual report questionnaire.

Note: Information is based primarily on the reported prevalence of opioid use and, when that was not available, on the ranking or data on treatment of opioid use reported in the annual report questionnaire.

The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the United Nations. Dashed lines represent undetermined boundaries. The final status of Jammu and Kashmir has not yet been agreed upon by the parties. The final boundary between the Republic of Sudan and the Republic of South Sudan has not yet been determined. A dispute exists between the Governments of Argentina and the United Kingdom of Great Britain and Northern Ireland concerning sovereignty over the Falkland Islands (Malvinas).

FIGURE 11.9–3. Worldwide distribution of opioid use. Upper two panels show the prevalence of use of opioids (all synthetic and naturally occurring opioid compounds) and opiates (naturally occurring compounds such as heroin, morphine, and codeine). **Map 1** shows the predominant form of opioid use in a country. While North America has the highest prevalence of opioid use, most of the opioid used was pharmaceutical opioids. (See eBook for color images.) (Reprinted from UNODC. *Book 3: Drug Market Trends: Cannabis Opioids. World Drug Report 2022*. United Nations Publication; 2022. https://www.unodc.org/res/wdr2022/MS/WDR22_Booklet_3.pdf)

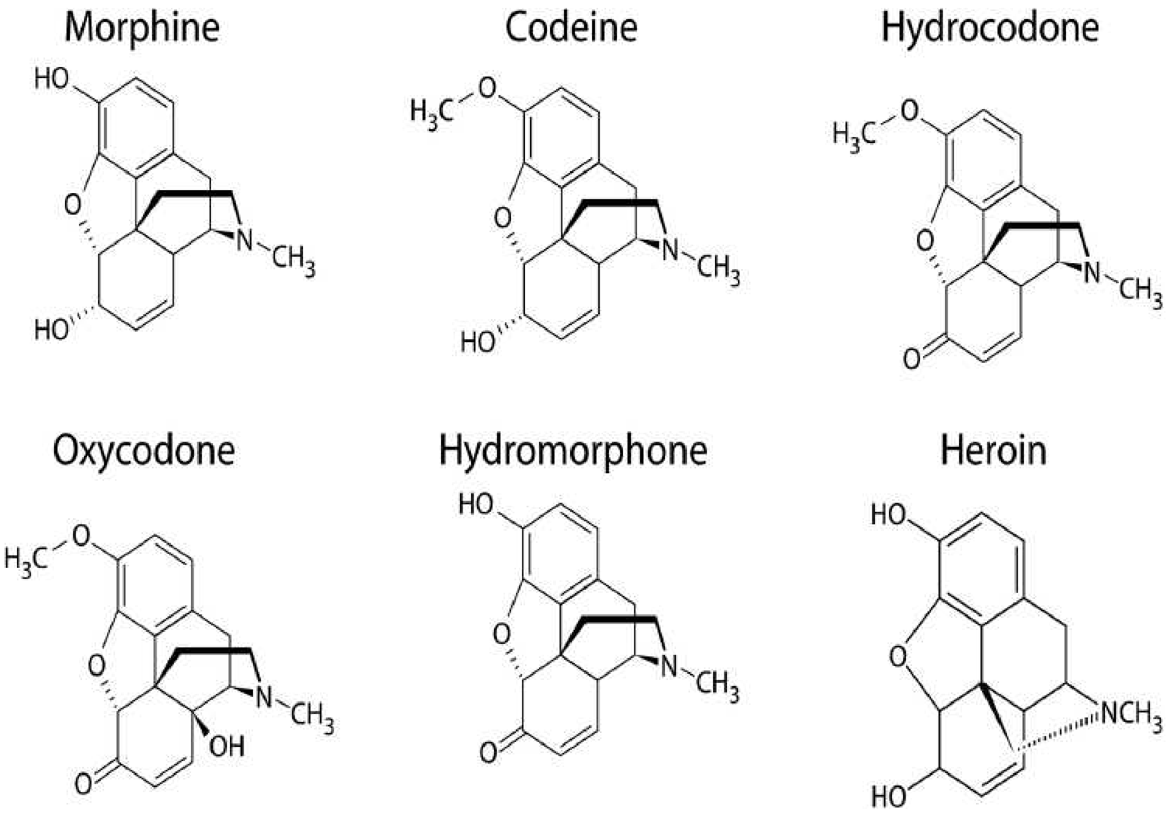

Morphine

Codeine

Hydrocodone

Oxycodone

Hydromorphone

Heroin

FIGURE 11.9–4. Chemical structures of common medicinal and illicit opioids. They share a three-ring phenanthrene structure as well as a methylamine group. All opioids are chemically related to morphine, the parent compound, and bind to the opioid receptors. Hydromorphone, hydrocodone, oxycodone, and heroin are semisynthetic. These compounds all have analgesic effects and can potentially be substances of abuse, though they have differing pharmacokinetic and pharmacodynamic properties.

Opioid receptors are a group of G protein–coupled receptors (GPCRs) with opioids as ligands and respond to a family of endogenous opioids (endorphins). The opioid receptors are distributed widely in the brain, the spinal cord, and digestive tract. There are four currently known types of opioid receptors: delta (δ), kappa (κ), mu (μ) and nociceptin receptor (NOP). All four types of receptors have seven transmembrane domains (TM1–7). Amino acid residues at the same positions in TM3, TM6, and TM7 interact with ligands to produce downstream effects—the “message” region of the ligands. Another region—the “address” region—produces opioid selectivity and occupies one of two different areas of the binding pocket, depending on the type of opioid (see Fig. 11.9–5).

The downstream activity of opioid receptors results in two known signal transduction pathways: classical and alternative. The classical signal transduction relates to the opioid receptor by inducing G proteins to interact with an inward rectifying potassium channel, Kir3. This process causes cellular hyperpolarization and inhibits neural activity. All four opioid receptors also cause a reduction in Ca^{2+} content both intracellularly and in synaptic vesicles and synaptosomes, with compensatory upregulation of vesicular Ca^{2+} content during the development of opiate tolerance. In addition, μ-, δ-, and κ-opioid receptors inhibit adenylyl cyclase activity and inhibit the cAMP-dependent Ca^{2+} influx. More recently, activation of opioid receptors has been shown to induce β-arrestin recruitment and binding in the alternative transduction pathway. Arrestin molecules bind to phosphorylated opioid receptors to regulate their traffic, and this interaction plays a role in the development of tolerance. Phosphorylated arrestin–bound opioid receptor complex also recruits alternate signal transduction cascades, including Mitogen-activated protein (MAP) kinases, inducing numerous other complex downstream molecular events.

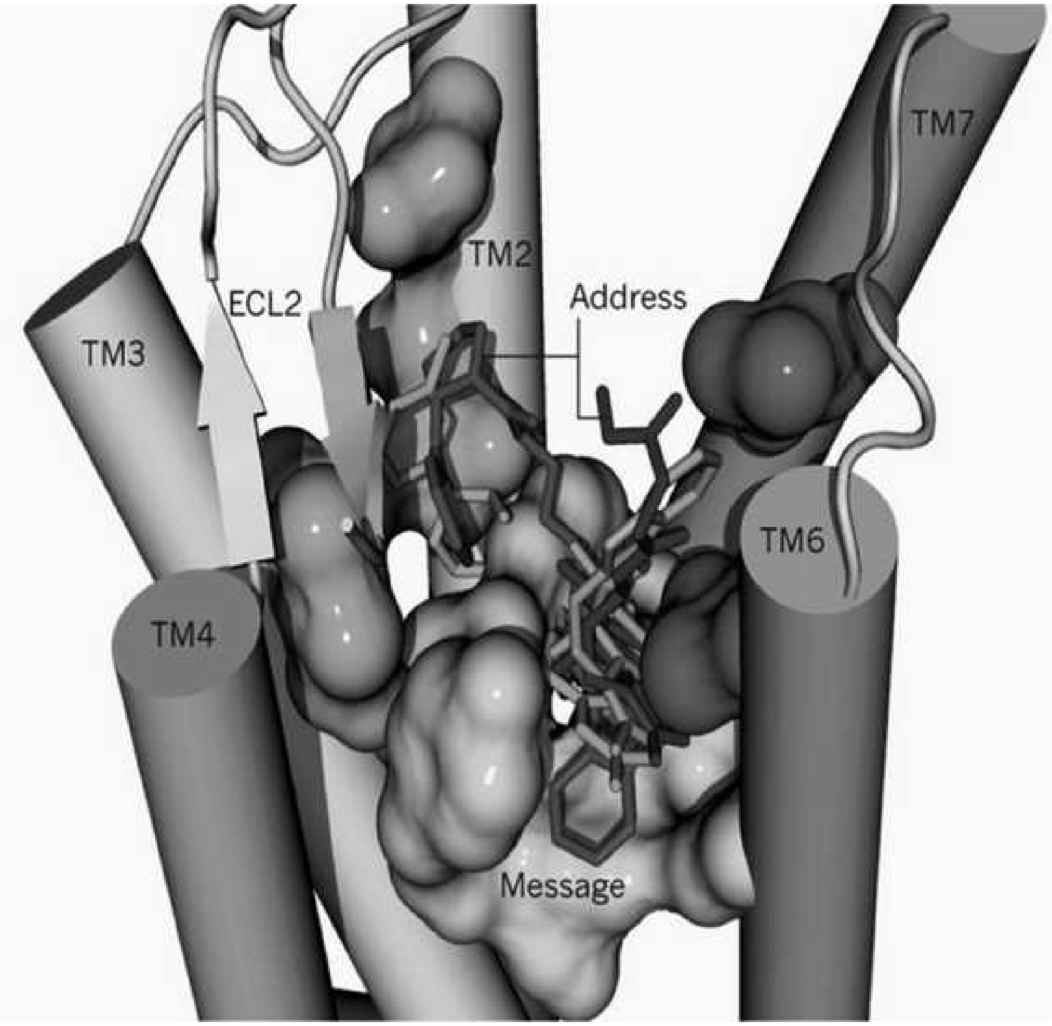

FIGURE 11.9–5. Common structural elements of the opioid receptors. Several features are shared by all four opioid receptor types: seven transmembrane helices (TM1–7), and five shown (*grey cylinders*). ECL2 is a β-hairpin loop region; the arrows represent β-sheets. Ligands are depicted as stick representations inside the binding pocket. The cyan surface indicates the amino acid residues from TM3, TM6, and TM7 that interact with the ligands' "message" regions. The magenta surfaces indicate the residues from TM6 and/or TM7 that interact with the "address" region—responsible for opioid selectivity. (Reprinted with permission from Springer Nature. Filizola M, Devi L. How opioid drugs bind to receptors. *Nature.* 2012;485(7398):314–317. doi: 10.1038/485314a.)

The molecular and cellular mechanisms mediating the development of tolerance to morphine continue to be debated. One hypothesis proposes that tolerance is a result of downregulation of opioid receptors following chronic agonist exposure. More recent *in vivo* studies show that downregulation does not occur consistently with each and every agonist.

Newer evidence suggests that the receptors may be desensitized and uncoupled from downstream signaling pathways and that chronic exposure changes levels of second messengers, such as cAMP, and upregulation of adenylyl cyclase.

Opioid Receptor Pharmacology. Opioids can interact with opioid receptors in several ways (see Fig. 11.9–5). An opioid *agonist* is a molecule that binds to an opioid receptor and induces downstream signal transduction events. Many opioids produce actions similar to that of morphine (a prototypical μ-opioid receptor agonist), and some bind to various receptor subtypes in a pattern that is distinct from that of morphine, producing a large, diverse array of possible patterns of opioid agonism. Molecules that bind to opioid receptors but initiate no signal transduction action are opioid *antagonists* at those receptors. Naloxone or naltrexone acts as μ-opioid receptor antagonists. When both antagonist and agonist are present, they compete against each other at the binding site. *Partial agonists* such as buprenorphine bind to opioid receptors but only produce a limited effect in intensity or maximal effects compared to a full agonist.

Agonist effect at μ-opioid receptors is the main pharmacologic effect associated with OUDs. The direct physiologic action of μ-agonists include analgesia, respiratory depression, miosis, euphoria, drowsiness, decreased ability to concentrate, changes in endocrine and other functions regulated by the hypothalamus, and increased tone of smooth muscle in the gastrointestinal (GI) tract. Agonists at the κ-opioid receptor often produce dysphoria and analgesia but no significant pupillary change. Heroin, morphine, hydromorphone, oxycodone, and fentanyl are all μ-opioid receptor agonists, and they differ from each other primarily in metabolism, pharmacokinetics, and relative affinities at other opioid receptors, which may produce considerable differences in their subjective effects. For instance, heroin (diacetylmorphine) is more potent and more lipid soluble than morphine, crossing the blood–brain barrier more rapidly and producing a more rapid onset of subjective effects. Patients are often able to discriminate heroin from oxycodone and from fentanyl.

Neural Circuits in Opioid Use Disorders. A variety of animal models have provided a pathway to dissect the neural circuits that mediate the effect of opioid molecules on the opioid receptors to affect animal behavior. The focus of experimental work has been converging on three aspects of opioid-associated behavior: self-administration, remodeling occurring with chronic exposure, and adaptation and relapse.

CIRCUITS IN OPIOID SELF-ADMINISTRATION. The μ-opioid receptor mediates the reinforcing effects of heroin and morphine. Blockade of the μ-opioid receptor, but not other opioid receptors, attenuates opiate self-administration. Knockout mice with no μ-opioid receptor expression in the brain do not readily associate contextual cues when these cues are paired with opiate administration (conditioning). Opioid drug conditioning is mediated by neuronal hyperpolarization and neuronal inhibition in multiple structures, including those that provide major afferent input to the ventral tegmental area (VTA); that is, nucleus accumbens (NAc), ventral pallidum, prefrontal cortex (PFC), amygdala, mediodorsal thalamus, and the pedunculopontine tegmental nucleus. This interconnected network is also known as the cortical–pallidal–striatal circuit and is characterized by a high density of opiate receptors with reciprocal connections. Site-specific blockade of μ-opioid receptor in the VTA or NAc attenuates heroin self-administration. Damage to the NAc can abolish opioid self-administration. These findings support the hypothesis that this circuitry mediates the rewarding effects of these agents.

Dopamine appears to play a central role in mediating the rewarding effect of opioids in this circuit as μ-opioid receptor agonists activate VTA dopaminergic neurons and enhance dopamine release in the NAc following heroin self-administration. Electrophysiologic studies support the idea that overlapping and heterogeneous subpopulations of NAc and medial prefrontal cortex (mPFC) neurons are activated during opiate self-administration and mediate its reinforcing effects. There is also evidence that a diverse group of neurons releasing GABA and glutamate, as well as the cannabinoid systems, are involved in mediating the reinforcing effects of opiates.

CIRCUITS INVOLVED IN CHRONIC OPIOID EXPOSURE. Cessation of chronic opioid exposure alters the activity of mesolimbic dopaminergic neurons, and the subsequent decreased dopamine neurotransmission may be related to anhedonia and dysphoria during the early phase of opioid withdrawal. The network level reconfigurations observed in chronic opioid users remain unclear, but dopamine receptor agonists can attenuate some signs of withdrawal in animal models. Chronic opioid use also alters the activity of NAc medium spiny neurons that are primary targets of dopamine neurons. These neurons are involved in a diverse number of transmitter release functions, including GABA and the opioid peptides dynorphin and enkephalin, as well as substance P, and project to the ventral pallidum and VTA. Alterations in these output neurons may also contribute to the aversive consequences of opiate withdrawal.

Opioid receptor antagonists, such as naloxone, are hypothesized to function primarily in blocking the μ-opioid receptor in NAc and other regions, including the amygdala, VTA, and locus coeruleus. Selective δ-opioid receptor antagonists can precipitate affective but not somatic components of withdrawal syndrome in animal models. Dynorphin is the endogenous ligand for the κ-opioid receptor, and agonists at these receptors produce dysphoria and negative affect mediated in part by the activation of NAc opioid receptors and inhibition of dopamine release. Increased dynorphin expression after abstinence is related to opiate-induced upregulation of the cAMP pathway and the induction of the gene transcription regulator, cAMP response element-binding protein (CREB). Long-term abstinence from opiates in a previously dependent animal often exhibits a phenomenon of hypersensitivity in dopaminergic neurons—with more dopamine release subsequently with the same dose of drugs. This is possibly related to behavioral sensitization to various drugs of misuse as well as playing a role in the reinstatement of compulsive drug-seeking behavior.

Chronic opioid use also induces circuit reconfiguration in the basolateral amygdala, locus coeruleus, and the bed nucleus of the stria terminalis, involving a variety of neurotransmitters such as glutamate and norepinephrine. Each of these adaptations may be involved in expression of the aversive effects of the opiate-withdrawal syndrome.

CIRCUITS MEDIATING RELAPSE. A large number of diffuse neural circuits appear to be engaged in reinstatement of opioid use in animal models of relapse. Factors involved in relapse in humans, such as stress, renewed contact with the drug, or drug cues and increased drug availability, reliably reinstate drug-seeking behavior in laboratory animals. The mesolimbic dopamine system plays a major role in mediating the priming effects of opiates in drug-seeking behaviors. Intra-NAc infusion of amphetamine, which stimulates dopamine release, can reinstate heroin seeking, while dopamine D2 receptor antagonists block heroin-induced reinstatement of drug seeking. Glutamatergic neurons may also be involved in reinstatement. Stress can reliably reinstate heroin seeking in rats, potentially mediated through corticotropin-releasing factor (CRF) releasing neurons, and may result from a distinct neural circuit outside of the mesolimbic dopamine system. Cue-induced relapse has been relatively poorly understood. Dopamine plays a less central role in this phenomenon, and PFC glutamate α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor plasticity is important in cue-induced relapse to heroin seeking.

Human Neuroimaging of Opioid Use Disorders

Key evidence supporting the hypothesis that drug addiction is a chronic disease of the brain comes from brain imaging studies examining a number of critical neural circuits. These studies provide a link between animal models of drug self-administration and human clinical studies. The existing evidence suggests that across all substances, there is a general circuit that is engaged in drug self-administration independent of the specific substance. There are then distinct substance-dependent circuits that mediate reinforcing effects of a specific substance. In opioids, the complex opioid receptor systems throughout the brain interact with the dopamine system in generating and maintaining the complex clinical phenomenon of OUDs. The main brain imaging techniques are structural magnetic resonance imaging (MRI), functional MRI, resting-state MRI, MR spectroscopy (MRS), positron emission tomography (PET), and single photon emission computed tomography (SPECT). Each of these methods can reveal different aspects of brain structure and function.

Converging brain imaging data suggest a distributed neural network of brain regions regulated by the μ-opioid receptor system to mediate pain, reward, motivation, and integration of emotional value. These regions include the anterior cingulate and PFC, the anterior and posterior insular cortex, orbitofrontal cortex, ventral striatum, ventral pallidum, thalamus, hypothalamus, amygdala, periaqueductal gray (PAG), and others. The κ-opioid receptor system is involved in suppressing the antinociceptive actions of opioids at the μ-opioid receptor sites and is the most widely expressed receptor system in the brain. Multiple studies report a push-and-pull antagonism between the μ- and κ-receptor systems. The μ-opioid receptor system is also associated with the positive reinforcing effect of other misused drugs, including cocaine and alcohol, in mediating their reward-related behaviors, and increasing binding in μ-opioid receptor is correlated with self-reported craving in cocaine.

PET studies show that dopamine is implicated in OUDs in a number of different ways. As the central neurotransmitter that mediates reward, dopamine D2/3 receptor availability is reduced in subjects with heroin dependence at baseline, which is an effect seen similarly in those dependent on cocaine, alcohol, tobacco, and methamphetamine. Presynaptic dopamine release is also blunted in individuals with heroin dependence (Fig. 11.9–6). There might also be a correlation between dopamine release and the subjective effects of heroin. These results suggest that the dopamine system may be involved in regulating a common underlying circuit in producing a general addiction phenomenon independent of the specific substance of misuse.

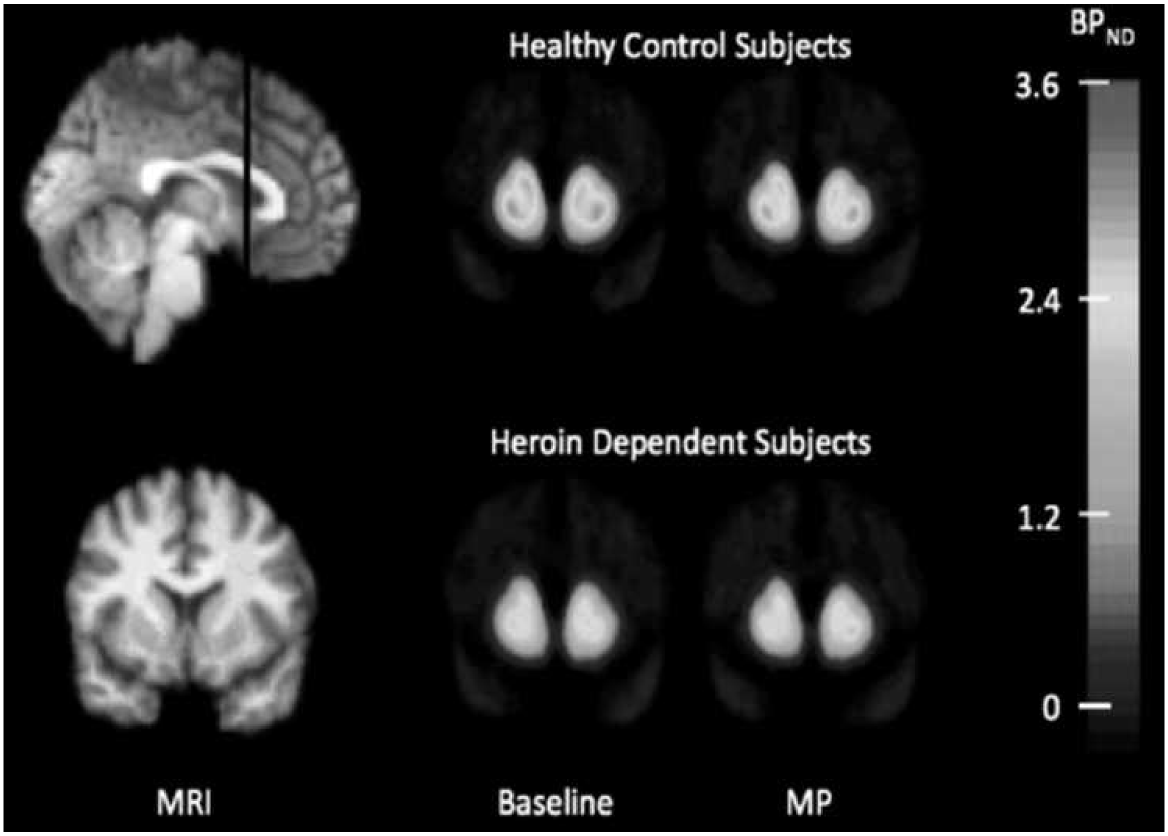

FIGURE 11.9–6. Deficit in dopamine neurotransmission in heroin-dependent subjects. [11C]raclopride is a dopamine radioligand whose binding potential (BPND) maps are averaged across all subjects. Left column: representative magnetic resonance imaging (MRI) from one subject showing the level of the coronal positron emission tomography slices (**bottom**). The *black line* in the sagittal view (**top**) shows the level of the coronal slices. Center and right columns: healthy control subjects (**top**) and heroin-dependent subjects (**bottom**) at baseline (**center**) and after methylphenidate (MP) (**right**). (Reprinted from Martinez D, Saccone PA, Liu F, et al. Deficits in dopamine D2 receptors and presynaptic dopamine in heroin dependence: commonalities and differences with other types of addiction. Biological psychiatry. *Biologic Psychiatry.* 2012;71(3):192–198. Copyright © 2012, with permission from Society of Biological Psychiatry. doi: 10.1016/j.biopsych.2011.08.024.)

Task-dependent fMRI studies in individuals with opiate dependence suggest that a decreased activity of a cingulofrontoparietal–cerebellar neural circuit may be involved in opioid use, with the region dorsal anterior cingulate cortex (dACC) in particular as being associated with performance deficits on tasks requiring behavioral control and error estimation. MRS suggests reduced concentrations of N-acetyl aspartate (NAA) and glutamate/glutamine within the dACC. Resting-state fMRI shows a significant decline in connectivity of the NAc to a number of areas, including laterobasal amygdala, hippocampus, parahippocampus, and the orbital frontal and midcingulate cortices.

While current neuroimaging evidence strongly suggests that there is an underlying circuit-level abnormality that mediates the initiation and maintenance of problematic opioid use, the precise mechanisms, both at the level of network activities of cortical and striatal neurons and their transmitter release, continue to be unclear. There is also no current neuroimaging method for predicting and prognosticating outcomes for OUDs. These important questions represent active areas of research.

Molecular Genetics and Genomics of Opioid Use

The genetic contribution to the development and maintenance of opioid dependence is significant. Twin studies estimate that around 50% to 60% of the liability for heroin addiction is genetic. There is also evidence that genetic vulnerability is made up of common factors that increase the likelihood of developing any drug addiction (16%) and factors that increase the likelihood of developing opioid addiction specifically (38%).

Linkage studies for opioid dependence show clustered linkage on chromosome 17, as well as linkage on a region of chromosome 14 overlying the neurexin 3 gene (*NRXN3*). They also identified a male-specific linkage peak on chromosome 10q and potentially a signal chromosome 4q. Functional polymorphism in *OPRM1* gene (Asn40 Asp), which encodes the μ-opioid receptor, and *OPRD1* (encoding δ-receptor) and *OPRK1* (encoding kappa receptor) have been extensively studied in haplotype-specific case-controlled designs, but results have been inconsistent. A potentially robust positive association between *OPRD1* variants and opioid dependence has been reported but requires replication.

Several genome-wide association studies (GWAS) have now been completed to investigate the genetics of OUD. The largest GWAS of OUD risk so far published (Gelenter et al, 2014) cumulatively included >12,000 subjects from several eastern U.S. recruitment sites and additional subjects accessed via publicly availably (dbGAP application) data from the Study of Addiction: Genetics and Environment (SAGE). The total sample included 6,877 European American and 5,432 African American subjects. The results identified several SNPs that mapped to KCNG2, which encodes a potassium voltage-gated ion channel. Another study isolated association signals of SNPs that mapped to the cornichon family

AMPA receptor auxiliary protein 3 gene (*CNIH3*), a glutamate receptor–associated regulatory protein encoded on chromosome 1. One study looking at genome-wide SNP data from genotyping microarrays to identify copy number variations (CNVs) that investigated their association with OUD included 5,152 subjects and showed OUD cases had fewer CNVs than the controls, in particular an intergenic region on 18q12.3, which has also been implicated in schizophrenia risk. These studies continue to be of small size by GWAS standards and are constrained by other technical limitations, but taken together, suggest that the molecular genetic basis of OUD arises from the same common underlying genetic loci that affect broad neurodevelopmental phenotypes that affect a distributed cortical expression network.

Psychosocial Factors

Genetic and neural circuits eventually contribute to addiction behavior through their interaction with complex environmental and psychosocial factors. Social attitudes, peer pressure, individual temperament, and drug availability are all determinants of the emergence of problematic opioid use. There is less social stigma associated with prescription opioids relative to heroin, and prescription opioids are believed to be safer in comparison to heroin and are more widely available. As individuals using prescription opioids develop a number of functional impairments, their opioid use often transitions to heroin and the IV route of administration due to lower cost and increasing tolerance for slower-acting agents. The presence of a significant number of opioid users in a neighborhood, town, or city also can create a subculture of opioid experimentation. In the United States, areas in which the prevalence of OUD is high also often have associated high crime rates, unemployment rates, and poorly performing school systems. While correlation does not equal causation, these factors can all be conceptualized as contributing to reduced resistance to opioid use and may worsen the prognosis and treatment efficacy.

Epidemiologic studies show that people who have OUDs place more value on independence and less value on academic achievement. Patients with heroin use disorders have more novelty-seeking, risk-taking, and impulsivity traits and show significant signs of delinquency before their first experimentation with opioids. A significant proportion of heroin users meet the criteria for antisocial personality disorder.

Most urban residents who use heroin come from single-parent families, and poor parental functioning has been consistently reported as a risk factor for opioid addiction. Family risk factors involve disruption of family structure and disrupted parental relationships. Cross-generational alliances between the individual who uses drugs and one parent against another parental figure are common. Dysfunctional family equilibrium can threaten the recovery process, and an assessment of family dynamics is an important component in the assessment and treatment of someone with OUD.

Integrative Approach

Across different levels of analyses, the etiology of OUD is now conceptualized as a complex interaction of genes, neural circuits, and environment. While the underlying mechanistic details remain subjects of active research, several consensus hypotheses have emerged from a large amount of converging experimental data. First, it appears that there is a large component of risk of developing OUD that derives from genetic influences, and these genetic variations change the pattern of biochemical signaling pathways involved in both opioid receptor and non–opioid receptor systems. These genetic factors of risk interact with environmental risk factors, which derive from sensitization during critical periods, traumatic experiences, and adverse social and economic conditions. The combination of genetic and environmental factors interacting in complex ways eventually leads to a heterogeneous set of neural circuit level abnormalities involving multiple cortical circuits and transmitter systems, likely involving abnormal release and reinforcement of dopamine neurotransmission. These systems subsequently influence both behavior and the rest of the body through their influence on the stress system (pituitary-adrenal system). The initiation of opioid use subsequently induces neuroadaptations in systems regulating reward, motivation and drive, memory, and behavioral control. Changes in other neural systems lead to the development of tolerance and physiologic dependence. As a result, affected individuals develop a range of problematic behavior: not only maintaining drug use but also relapsing following cessation of use. These behaviors lead to criminality, poverty, and a deficit in social support, which in turn strengthens drug-seeking and -using behavior.

This complex, multifactorial etiologic model stipulates that abnormalities in brain function are behind the etiology and maintenance of OUD. It is important that patients and their families recognize the significance but also the limitations of this model. OUD is a chronic, relapsing–remitting behavioral brain disorder that is not a character flaw or moral weakness. On the other hand, the central role of neural circuit-level abnormalities does not imply that patients have no control over the choices they make, particularly choices related to their own treatment and recovery. The brain is an incredibly plastic organ, and neural circuits can be remodeled in response to environmental changes, new behaviors, and administration of medications. Effective treatment for opioid addiction should therefore target all of these etiologic factors in a multilevel fashion.

DIAGNOSIS AND CLINICAL FEATURES

Opioid Intoxication and Overdose

The DSM-5 criteria for opioid intoxication are broadly inclusive, and in clinical situations, opioid intoxication symptoms can vary in severity and presentation. Initial euphoria is usually followed by apathy, dysphoria, psychomotor agitation or retardation, and impaired judgment. Pupillary constriction, drowsiness, and slurred speech are common. Other physical signs include cyanosis, cold, clammy skin, and decreased body temperature. There may also be x-ray evidence of pulmonary edema and frothing at the mouth. Blood pressure is decreased but only falls dramatically with severe anoxia. Cardiac arrhythmias have also been reported and may be related to anoxia or to the presence of quinine as an adulterant in the opioid. A marked tolerance usually develops over time to the subjective intoxicating effects of opioids. A summary of sample diagnostic criteria in DSM-5 is included in Table 11.9–1.

On occasion, intoxication is followed by an overdose, a state characterized by unresponsiveness to external stimulation (coma). Signs of opioid overdose include coma with pinpoint but reactive pupils and respiratory depression or apnea. In severe cases, severely depressed respiration can develop, which can lead to anoxia and large, nonreactive pupils. Overdose due to fentanyl can also include respiratory muscle rigidity and laryngospasm (“wooden chest” syndrome). Overdose death can result from hypoxic–ischemic injury to the brain and cardiovascular system, as well as sudden cardiac death from arrhythmia due to electrolyte abnormalities.

Opioid Tolerance

Tolerance to opioids' intoxicating and hedonic effects is a common reason for escalation of use, seen as an increase in the frequency of use (up to several times per day) and the amount used on each occasion. However, even chronic heroin users experience brief periods of euphoria each time they receive the drug. Ongoing opioid use is, therefore, not solely used to avoid withdrawal. Because of tolerance to physical symptoms of opioids, such as respiratory depression and sedation, opioid effects may not be readily detected in chronic users by even the careful observer, which makes the detection of opioids based on simple observation difficult. High tolerance levels, observed in patients maintained on higher doses of methadone, can be beneficial in minimizing the effects (blocking) of illicit opioids and is one of the mechanisms for the therapeutic effects of methadone and, to a lower extent, buprenorphine.

Table 11.9–1.
Summarized DSM-5 Diagnostic Criteria for Opioid Intoxication

DSM-5 recognizes individuals as experiencing opioid intoxication with the following typical history, signs, and symptoms.

- Behavioral or psychological changes (e.g., euphoria followed by apathy, dysphoria, psychomotor agitation or retardation, impaired judgment).
- Pupillary constriction.
- Drowsiness or coma.
- Slurred speech.
- Impairment in attention or memory.

Opioid Withdrawal

Opioid withdrawal develops after the cessation of or sudden reduction in opioid use that has been heavy and prolonged. The withdrawal syndrome also occurs when an opioid partial agonist such as buprenorphine or an opioid antagonist such as naloxone or naltrexone is administered to a patient who is chronically taking an opioid agonist ("precipitated withdrawal"). Many withdrawal symptoms are the physiologic opposite of intoxication symptoms. Patients undergoing withdrawal complain of anxiety, irritability, restlessness, an "achy feeling" in the legs and back, and increased sensitivity to pain. Few objective signs may be present in milder withdrawal. In more severe withdrawal, patients present with nausea or vomiting, abdominal cramps and diarrhea, severe muscle aches, lacrimation and rhinorrhea, yawning, pupillary dilation, and increased sweating and chills. Fever, severe insomnia, confusion, and delirium can occur in severe withdrawal. A summary of sample diagnostic criteria in DSM-5 is included in Table 11.9–2.

Opioid withdrawal can be either *spontaneous* (stopping daily use) or *precipitated* (due to the administration of an opioid antagonist or partial agonist). Spontaneous withdrawal starts 6 to 8 hours after the last use of the short-acting opioids (heroin, morphine, oxycodone, hydrocodone), 24 hours after the last use of longer-acting opioids such as methadone or buprenorphine, and even later in some cases of nonpharmaceutical fentanyl use. Withdrawal severity peaks around 36 to 48 hours and resolves after 5 to 7 days for short-acting agents and after 2 or more weeks in the case of methadone. Administration of an antagonist precipitates opioid withdrawal within minutes after parenteral administration (naloxone) and 2 to 3 hours after oral or injectable administration (naltrexone). Opioid withdrawal precipitated by naltrexone is usually more severe than spontaneous withdrawal but resolves faster, generally within 1 day.

Table 11.9–2.
Summary DSM-5 Diagnostic Criteria for Opioid Withdrawal

DSM-5 Recognizes individuals as experiencing opioid withdrawal with the following typical history, signs, and symptoms.

History:

- Cessation of (or reduction in) opioid use.
- Administration of an opioid antagonist after a period of opioid use.

Signs and Symptoms:

- Dysphoric mood.
- Nausea or vomiting.
- Muscle aches.
- Lacrimation or rhinorrhea.
- Pupillary dilation, piloerection, or sweating.
- Diarrhea.
- Yawning.
- Fever.
- Insomnia.

A period of continuous receptor activation is required before spontaneous withdrawal is clinically significant, but a precipitated withdrawal can occur even after a few days of taking an opioid. There is considerable variability between individuals in presenting symptoms during opioid withdrawal. The severity of opioid withdrawal also varies across individuals and in time, and depends on the pharmacologic characteristics of the opioid, such as its potency, degree of opioid agonism (e.g., full versus partial agonist), dose, the duration of use, and the rate at which the opioid is removed from the receptors (e.g., a slow taper versus abrupt cessation). In general, the higher the potency, agonist efficacy, and dose, the longer the duration of opioid use, and the faster the removal of opioids from the receptor, the more intense the withdrawal syndrome. Standardized instruments have been developed to quantify the severity of opioid withdrawal symptoms in clinical practice. The Clinical Opiate Withdrawal Scale (COWS) is the most common and utilizes resting pulse rate, sweating, restlessness, pupil size, pain, rhinorrhea, GI symptoms, tremor, yawning, anxiety, and piloerection as anchor points (see Table 11.9–3).

Opioid withdrawal symptoms can occur after abrupt discontinuation of maintenance treatment with buprenorphine, a partial opioid receptor agonist, but such symptoms are generally less severe than in the case of cessation of a full agonist. Some patients report significant insomnia, irritability, anxiety, and depression. These symptoms tend to resolve within 7 to 10 days. Drug effects from larger doses of buprenorphine (16 to 32 mg or more sublingually) can persist for several days due to its long half-life, and withdrawal symptoms can be delayed for 96 hours or more. This feature of buprenorphine has led to its use on an intermittent basis to prevent withdrawal (e.g., dosing every other day).

Following the resolution of an acute opioid withdrawal syndrome, there also appears to be a period of physiologic abnormality that lasts for weeks to months, referred to as *protracted or postacute withdrawal syndrome*. It is characterized by depressed mood and anhedonia, anxiety, irritability, labile mood, stress intolerance, insomnia, and recurrent craving and urges to use opioids. This protracted syndrome may result from alterations and remodeling of the neural circuitry associated with withdrawal, such as changes in the serotonergic and dopaminergic systems. This syndrome, while not formally included in the DSM-5, is likely a significant contributor to relapse to opioid use.

Opioid-Induced Mental Disorders

DSM-5 specifies a category of substance-induced disorders for a number of different substances that includes a variety of opioid-induced mental disorders (psychotic, bipolar, depressive, anxiety, obsessive-compulsive, sleep, sexual, delirium, and neurocognitive disorders). Chronic opioid use can produce several psychiatric syndromes, and opioid-using patients can also have comorbid psychiatric diagnoses that precede or are independent of opioid use. The important differential considerations, therefore, rely on a detailed history of chronicity and order of emergence for these symptoms. The onset of opioid-induced psychiatric syndromes needs to be temporally connected with the substance use (within 1 month), as opposed to occurring prior to the initiation of substance use or during a prolonged period of abstinence, and opioids need to be capable of producing the specific syndrome. Clinically, it is common to see some degree of depressed mood during and for several weeks after opioid withdrawal. Opioid-induced mood disorder should not be diagnosed during and immediately after opioid withdrawal unless the severity of mood disturbance exceeds what is normally encountered or persists for more than a few weeks and is of sufficient intensity to warrant independent clinical attention (e.g., presenting with profound psychomotor retardation or suicidality). Opioid-induced psychotic symptoms are rare, although they have been clinically observed, especially manifesting as derealization and occasionally hallucinations. Opioid-induced sleep disorder and opioid-induced sexual dysfunction are common, and both hypersomnia and insomnia are common complaints of patients who either have OUD or are maintained on opioids for therapeutic purposes. Common sexual dysfunctions for chronic opioid users include erectile dysfunction and orgasmic difficulties. While the presentation and consequences of opioid-induced mental disorders can be identical as in the case of the independent mental disorder, most opioid-induced disorders are likely to improve within days or weeks of abstinence from the abused opioid.

Table 11.9–3.
Clinical Opioid Withdrawal Scale (COWS)

Resting Pulse Rate: (record beats per minute)

Measured after patient is sitting or lying for 1 minute

0 pulse rate 80 or below
1 pulse rate 81–100
2 pulse rate 101–120
4 pulse rate greater than 120

Sweating: *Over past ½ hour not accounted for by room temperature or patient activity*

0 no report of chills or flushing
1 subjective report of chills or flushing
2 flushed or observable moistness on face
3 beads of sweat on brow or face
4 sweat streaming off face

Restlessness: *Observation during assessment*

0 able to sit still
1 reports difficulty sitting still, but is able to do so
3 frequent shifting or extraneous movements of legs/arms
5 Unable to sit still for more than a few seconds

Pupil Size

0 pupils pinned or normal size for room light
1 pupils possibly larger than normal for room light
2 pupils moderately dilated
5 pupils so dilated that only the rim of the iris is visible

Bone or Joint Aches: *If patient was having pain previously, only the additional component attributed to opiates withdrawal is scored*

0 not present
1 mild diffuse discomfort
2 patient reports severe diffuse aching of joints/muscles
4 patient is rubbing joints or muscles and is unable to sit still because of discomfort

Runny Nose or Tearing: *Not accounted for by cold symptoms or allergies*

0 not present
1 nasal stuffiness or unusually moist eyes

2 nose running or tearing
4 nose constantly running or tears streaming down cheeks

GI Upset: *Over last ½ hour*
0 no GI symptoms
1 stomach cramps
2 nausea or loose stool
3 vomiting or diarrhea
5 Multiple episodes of diarrhea or vomiting

Tremor: *Observation of outstretched hands*
0 No tremor
1 tremor can be felt, but not observed
2 slight tremor observable
4 gross tremor or muscle twitching

Yawning: *Observation during assessment*
0 no yawning
1 yawning once or twice during assessment
2 yawning three or more times during assessment
4 yawning several times/minute

Anxiety or Irritability
0 none
1 patient reports increasing irritability or anxiousness
2 patient obviously irritable anxious
4 patient so irritable or anxious that participation in the assessment is difficult

Gooseflesh Skin
0 skin is smooth
3 piloerrection of skin can be felt or hairs standing up on arms
5 prominent piloerrection

Score:
5–12 = **mild;**
13–24 = **moderate;**
25–36 = **moderately severe;**
more than 36 = **severe withdrawal**

From Wesson DR, Ling W. The Clinical Opiate Withdrawal Scale (COWS). *J Psychoactive Drugs.* 2003;35(2):253–259, reprinted with permission of Taylor & Francis Ltd, http://www.tandfonline.com. doi: 10.1080/02791072.2003.10400007.

Opioid Use Disorder

DSM-5 OUD diagnosis spans a spectrum of signs, symptoms, and problematic behaviors (Table 11.9–4). These include problems related to opioid use (social/interpersonal, neglect in major roles, physical and psychological problems), abnormal behavior (inability to control the amount or to stop use, spending a lot of time on use at the expense of other recreational activities), physiologic changes (tolerance and withdrawal), and craving. Two or more criteria are set as a threshold for the diagnosis, and three levels of severity are specified (mild, moderate, and severe). The diagnostic criteria are primarily assessed during a detailed clinical interview, while repeated toxicology tests (e.g., urine or saliva) can be used as corroborating evidence of chronic use. No single criterion is given special weight: the presence of tolerance and withdrawal are neither necessary nor sufficient for the diagnosis. The broad emphasis for these criteria is to specify the degree of impairment in social and occupational function due to a pattern of problematic opioid use rather than on the timing, quantity, type, or route of use.

Pathology and Laboratory Examination

There is currently no biomarker for the diagnosis of OUDs. The most commonly used urine toxicology tests can detect recent opioid use. Urination is a major route for drug elimination, and urine testing is a convenient alternative to testing blood. Drug metabolites are first detectable in urine 3 to 6 hours after use. Morphine, heroin, codeine, and hydrocodone can be detected in urine for 1 to 3 days after use. Heroin metabolite (6-acetylmorphine) can be detected in urine for up to 8 hours after use, after which heroin can be detected as morphine (its metabolite). The ingestion of modest amounts of poppy seeds (of the type and amount used in baking), which contain opium alkaloids, may also produce positive results on tests with high sensitivity. Synthetic and semisynthetic opioids such as methadone, fentanyl, buprenorphine, tramadol, and oxycodone require specific tests as these do not share metabolites with natural opioids (morphine and codeine). Methadone can be detected in urine for up to 4 days after a single-use episode and for longer with repeated use.

Table 11.9–4.
Summarized DSM-5 Criteria for Opioid Use Disorder

DSM-5 recognizes opioid use disorder as a pattern of problematic opioid use leading to significant clinical impairment within a 12-month period, with at least 2 of the following elements in history, symptoms, and signs:

- Opioids taken in larger amounts or over a longer period.
- Persistent desire to cut down.
- A great deal of time is spent to obtain, use or recover from opiates.
- Craving.
- Failure to fulfill major role obligations at work, school, or home.
- Persistent or recurrent social or interpersonal problems
- Important social, occupational, or recreational activities.
- Opioid use that is physically hazardous.
- Continued opioid use despite physical or psychological problems.
- Tolerance.
- Withdrawal.

Other toxicology methods involve testing saliva or hair. Samples of oral fluids can detect recent opioid use with approximately the same sensitivity as urine testing, though they may have a shorter detection window, especially with fat-soluble drugs such as THC. Moreover, drugs can be detected in saliva immediately after smoking or injecting and within 1 to 2 hours after swallowing. The main advantage of saliva testing is that sample collection can be fully observed, which minimizes the risk of sample adulteration (substitution, dilution, and additives) encountered with unobserved urine testing. Analysis of hair clippings can provide information and offers an average measure of use over the preceding 2 to 3 months. Hair testing is a better indicator of chronic use than urine, but drug use within the previous week will not be detected in hair.

Patients with a history of IV opioid use should also be tested for hepatitis (B and C) and human immunodeficiency virus (HIV) infections. Sharing straws for intranasal (IN) opioid use is a risk factor for hepatitis C. At the time of initial evaluation, it is also important to assess liver functions, complete blood count, and basic metabolic panel, as opioid-using patients tend to have poor nutritional status, and acute opioid intoxication can cause metabolic abnormalities and leukocytosis. Chest x-rays may show evidence of tuberculosis or pulmonary fibrosis in persons using injection materials contaminated with microcrystalline talc or cotton particulates. Methadone may prolong the QT interval on EKG; therefore EKG can be useful in detecting individuals with prolonged QT.

Physical examination is often normal though signs of intoxication or withdrawal may be present. IN use of heroin may cause nasal irritation and necrosis. IV use can lead to needle marks throughout the body because injection sites are concealed (e.g., soles of the feet). Infections (e.g., thrombophlebitis, skin abscess), lymphatic obstruction, and edemas are not uncommon. Subcutaneous and muscle tissue hardening can result from repeated intramuscular (IM) injections of meperidine (seen among health professionals). Skin lesions consisting of scars, skin atrophy, and hyperpigmentation suggest chronic use. Endocarditis may produce fever and heart murmurs. Signs of liver abnormalities, such as stigmata and ascites, should also be documented. Mental status examinations often show abnormalities in mood, affect, and cognition.

Differential Diagnosis

Diagnosing is usually straightforward in clinical situations when an individual requests treatment and is open about the duration and frequency of use and the nature of associated problems. However, in earlier stages of the disease, collateral information from family and peers confirmed with toxicology results is necessary, as individuals tend to underrecognize or diminish the existence and impact of drug-related problems. Patients with a long history of addiction who relapsed may try to hide the history of illness when trying to obtain opioids in the medical setting—engaging the family or reaching out to previous medical providers is important for an accurate diagnosis and an adequate treatment plan.

OUD can be misdiagnosed in patients with chronic pain or intermittent acute pain syndromes (e.g., sickle cell disease) who exhibit a range of problematic behavior such as arguing with medical staff, requesting higher doses and more potent opioids, and seeking medications from multiple doctors. Ongoing and poorly controlled pain may elicit these behaviors, a phenomenon sometimes referred to as *pseudoaddiction*. Careful longitudinal observation, outside informants, and urine toxicology testing, alongside continued appropriate attempts at pain relief and treatment of co-occurring medical and psychiatric problems, are required in generating the differential diagnosis. An empathic, compassionate, and nonargumentative approach to working with such patients is essential, and a physician–patient contract specifying the treatment plan and procedures to follow is helpful.

In most cases, *opioid withdrawal* is easy to diagnose, as its physiologic signs are very typical. Diagnosing withdrawal in the early phases, when patients mostly report subjective symptoms of various aches, pains, and anxiety, may be more difficult, and further monitoring may be necessary. Poor response to treatment of withdrawal with opioid agonists such as methadone suggests withdrawal from other substances. Administering buprenorphine to an opioid-dependent individual who does not have clear signs of opioid withdrawal can worsen the withdrawal (precipitated withdrawal). The presence of severe withdrawal (autonomic instability, confusion, or seizures) without classic opioid-withdrawal signs (e.g., yawning, vomiting, diarrhea) raises the strong possibility of withdrawal from other substances such as alcohol or sedative–hypnotics. Toxicology evaluation, which includes assays for synthetic opioids, is essential in all cases presenting with drug-withdrawal symptoms.

Similarly, *opioid intoxication* must be differentiated from mixed intoxications in which opioids only play a minor role. In general, a failure to respond to repeated doses of naloxone suggests that intoxication is caused by a nonopioid. Older adult patients may respond to opioids with dysphoria and confusion in *atypical opioid intoxications* rather than opioid-induced delirium or opioid-induced psychotic disorders, which are rare.

COURSE AND PROGNOSIS

The clinical course for OUDs is highly variable. Some persons apparently can use opioids occasionally (e.g., several times a month) over periods of months or years without meeting the criteria for an OUD. This is the more common course for patients who use prescription opioids but is also seen in those using heroin. Data from national surveys suggest that approximately one-third of recreational users develop moderate to severe OUD in their lifetimes, and the rates are lower in people who use prescription opioids recreationally. Even though most recreational users do not go on to develop a use disorder, there is no reliable way to predict among first-time users on an individual level who will progress to OUD. The presence of a co-occurring psychiatric or substance use disorder and IV use of heroin are major risk factors. Patients who meet criteria for an

OUD at some point in their lives can achieve significant periods of abstinence, often without formal treatment. Most commonly, affected individuals will experience a chronic and recurring course of illness.

Patterns of Use and Relapses and a Long-Term Prognosis

OUD typically has a chronic course, though the progress and remission through this course in a particular individual is highly variable. Opioid use commonly begins with single doses used intermittently (e.g., weekend use) with longer periods of abstinence, either self-imposed or involuntary in the form of imprisonment or hospitalization, lasting from a few weeks to years. Gradually, there is a transition to daily use and an increase in the frequency and, later, the quantity of use episodes. Multiple uses of short-acting opioids throughout the day are common, often with a first dose in the morning preceding all other daily activities to eliminate or avoid early withdrawal symptoms.

Periods of self-detoxification and voluntarily imposed abstinence are typical but often followed by a relapse, especially under stress. Stress, depression, and life crises (especially arguments and interpersonal losses) are commonly associated with relapse. Some persons can maintain a lower level of use early on following relapse, but more commonly, the frequency of use, tolerance, physiologic dependence, and drug-related problems rapidly escalate, usually at a faster pace than during early periods of the disorder. The most common reasons for seeking abstinence are accumulating occupational, family, financial, and legal problems, feelings of fatigue and exhaustion with activities needed to maintain daily use, and inevitable periods of withdrawal. For others, emergence of severe depression or a legal mandate is the primary reason they want to stop using and seek out treatment.

Treatment allows individuals to initiate and maintain long-term abstinence. One of the major goals of treatment is to avoid accidental overdose and favorably modify the course of the disorder by minimizing the impact of risk factors on relapse (relapse prevention). In most cases, more than one treatment episode is necessary to achieve long-term remission, however, among individuals who have a good treatment response, multiple treatment attempts can have a cumulative effect, and the duration of each drug-using episode tends to shorten. Many patients who have had a positive experience in treatment will be able to recognize high-risk situations and protect themselves from relapse or will be able to return to treatment after the first few episodes of drug use (lapse) and minimize the risk that lapse would progress to a full-blown relapse.

Repeated relapse, although common, is not an inevitable course of OUD. Cohort studies show that 10% to 40% of interviewed opioid users are abstinent at long-term follow-up. Treatment research conducted throughout the past four decades shows that a reasonable proportion of opioid users enter long-term remission. Sustaining ongoing treatment and monitoring, longer lengths of stay in treatment, and self-motivation have been identified as the most important reasons for long-term recovery. Other protective factors include practicing religion, spirituality, employment, and family involvement. Positive outcomes, such as greater participation in prosocial activities, including employment and improved mental health, are both causes and effects of long-term recovery. Many patients who develop an OUD do not enter treatment and tend to have a less severe form and course of the disorder and fewer co-existing medical and psychiatric problems as compared to those seen in treatment programs.

For a large proportion of patients, OUD is a life-long condition with severe health and social consequences. A longitudinal study of 581 patients with moderate to severe OUD who were placed under the California Civil Commitment Program in the early 1960s showed that during the 33-year follow-up period, 49% died, a death rate more than 50 times greater than age-matched controls. Most of the deaths involved drug-related causes or violence. Among the survivors, many had serious medical and psychiatric problems and were involved in the criminal justice system. At the time of the follow-up interview, 20% were using heroin, 10% refused testing, 4% were incarcerated, and 56% were currently abstinent, but 40% had used heroin within the past year. More than 60% of subjects reported long periods of abstinence (>1 year). A review of long-term follow-up studies found abstinence rates between 10% and 20% for both drug-free and methadone maintenance treatment. In more recent studies, such as the AIDS Linked to the Intravenous Experience (ALIVE) Study, conducted over the 12-year study period, four patterns were noted in patients with a pattern of injection drug use: 29% of participants remained persistent drug injectors, 20% ceased injection, 14% relapsed once, and 37% had multiple transitions.

Legitimate work, lack of criminal activity or other drug use, solid family relationships, and good psychological adjustment are favorable indicators for good long-term prognosis. Severity of comorbid psychiatric conditions at the beginning of treatment usually conveys a bad prognosis. Over shorter follow-up periods, methadone maintenance, therapeutic communities, and drug-free programs were significantly superior to detoxification-only programs.

Opioids and Crime

The relationship between drug use and crime is complex. Heroin use correlates with increased levels of criminal activity, especially income-generating crime. In outpatient opioid maintenance treatment, crime reduction is often one of the main outcomes of interest. Observational studies have shown that methadone maintenance treatment reduces crime in persons using heroin. Retention in treatment is consistently shown to be an important factor in mediating this reduction. Patients in continuous treatment had the greatest reductions in criminal activity, while those who drop out after several treatment episodes have the highest rates of criminal convictions. On the other hand, there is a smaller association between crime and illicit use of other opioids and other causal factors of crime, such as poverty, lack of social services, and discrimination.

Morbidity and Mortality

Chronic use of pure opioids, including methadone and buprenorphine, in nontoxic doses, has relatively few known side effects as tolerance develops to most of the acute effects. The cognitive impairment seen with long-term alcohol and sedative use is not generally found with long-term use of opioids. The only organ toxicity appears to be endocrinopathy with lower levels of testosterone and other sex hormones and resulting physiologic changes (loss of libido, impotence, amenorrhea). Occasionally, patients report chronic constipation and sleep disturbances. There is, however, toxicity associated with frequently found contaminants in the heroin supply, especially found in individuals who inject opioids.

Despite the relative safety of chronic opioids, the life expectancy of those with OUD, especially those using heroin, is markedly reduced. Death occurs at a young age, with an average of 44 years of potential life lost for each fatality. The rate of mortality among those with OUD is 6.5 times the rate expected in the general population, with a higher rate for younger persons. The overall standardized morality rate in the United

States is about 1.0% to 1.5% per person per year, 0.8% to 1.2% in a cohort of people treated for opioid dependence in Australia, and 2% to 3% in British clinics providing unadulterated injectable drugs to younger patients. Overdose is the most frequent cause of death, followed by unintentional injuries (accidents and violence) and suicide. Recent reports suggest that ongoing medication treatment with buprenorphine is associated with lower all-cause mortality, emergency room admissions, and inpatient hospitalizations. Death related to liver disease, particularly viral hepatitis, is very common, while HIV-related mortality has been steadily decreasing thanks to HIV prevention efforts and harm reduction programs such as needle exchange.

Overdose. Even though overdose is the most common cause of death in individuals with OUD, most overdoses are nonlethal. Only 1 out of 25 to 50 overdoses results in death; however, harm associated with nonlethal overdoses can be significant because even brief periods of complete anoxia can irreversibly damage CNS, and there is an accumulating effect of multiple overdoses. Most cases of overdose occur in older, long-term users, many of them using via injection, and many of those individuals have one or more nonlethal overdoses before a lethal one. It is not well understood why some individuals are at increased risk of overdose. A majority of overdoses involve multiple CNS depressants, most commonly alcohol and benzodiazepines, but also other opioids such as methadone. The risk of overdose appears to be much higher in those using potent synthetic opioids such as fentanyl analogs as compared to those using heroin or prescription opioids.

Medical Comorbidity. Most of the medical complications seen in chronic users of opioids, particularly users of heroin, are due not to heroin but to medical complications related to exposure to other substances used concurrently (e.g., impurities and adulterants in heroin and acetaminophen in prescription painkillers), harmful drug administration practices (e.g., nonsterile injecting), and high-risk behaviors associated with life experiences (e.g., violence or unprotected sex). Infections of the skin and subcutaneous tissue (abscesses, cellulitis), thrombophlebitis, endocarditis, pneumonia, and hepatitis are frequently seen medical complications. Acute infections are the major reason for inpatient admission of people who inject drugs.

Viral hepatitis, particularly hepatitis C (HCV), is the leading infectious cause of death in the United States and the leading cause of liver transplantation and liver cancer. In the United States, up to 80% of HCV cases are seen among persons who inject drugs, as are half of all new cases of infection. Abnormal liver function is seen in two-thirds of those who use heroin, and it may persist after the cessation of injections. High-risk sexual and drug-use behaviors are the most common mode of transmission of hepatitis and HIV. Despite education efforts and other prevention measures, the frequency of these behaviors remains high in the population of those who inject drugs. All patients who are in treatment should be offered vaccination for hepatitis B.

Acquired immune deficiency syndrome (AIDS) remains an important factor in mortality among IV opioid-using patients. HIV seroprevalence has a wide geographic variability and is lower among those who have continuously been in drug treatment. The incidence of tuberculosis is also higher among heroin users than in the general population and has been linked to the compromised immune system of these individuals regardless of whether or not they have HIV. Persons who inject drugs should have access to measures that reduce the spread of hepatitis and HIV, such as clean injecting equipment, condoms, and access to medical care.

Other infections related to IV substance use and sharing of needles include malaria, tetanus, osteomyelitis, and syphilis. Endocarditis related to staphylococcus but also pseudomonas or candida may result in serious complications. Endocarditis can cause abscesses in the brain or spinal cord, meningitis, diffuse vasculitis, and subarachnoid hemorrhage from rupture of an infected aneurysm. Various contaminants (e.g., talc) entering the bloodstream can cause pulmonary emboli, which can eventually result in angiothrombotic pulmonary hypertension and right ventricular failure. Staphylococcal pneumonitis may also be related to septic emboli. Poor IV injecting technique can cause chronic edema of extremities. Subcutaneous or intracutaneous injection (known as “skin popping”) may cause widespread ulceration and disfigurement.

Patients surviving overdose may experience long-term sequelae of anoxic brain injury, such as transverse myelitis, amblyopia, plexitis, peripheral neuropathy, parkinsonian syndrome, movement disorder, intellectual impairment, and personality changes. Pathologic changes in muscles and degeneration of peripheral nerves have also been seen. Synthetic opioids can affect cardiac conduction by lengthening the QT interval, which can cause serious arrhythmias, such as torsade des points. Methadone has a lower margin of safety as compared to buprenorphine, with prolonged QT interval (≥500 msec) seen in approximately 2% of patients treated with methadone.

Chronic pain is commonly seen within the opioid-dependent population. Pain is often cited as the reason for initiating use of prescription opioids and heroin, and chronic and usually post-traumatic pain is often reported by patients who use opioids.

Psychiatric Comorbidity. Comorbid psychiatric conditions are the rule rather than the exception in OUD. The type and severity of additional psychiatric diagnoses can influence the course of the disorder and treatment. The first and second waves of the U.S.-based National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), collected between 2001 and 2006, found that 69% of men and 74% of women with OUD in their lifetimes met criteria for an additional mood, anxiety, or another substance use disorder at some point in their lives. Some of the most common diagnoses included mood disorders, antisocial personality disorder, anxiety disorders, and alcohol use disorder (see Table 11.9–5). Findings from the Australian Treatment Outcome Study (ATOS) also show a high prevalence of mental disorders among opioid users, among those both who are in treatment and those who are not. In the ATOS, the most prevalent comorbid psychiatric conditions were major depression, post-traumatic stress disorder (PTSD), and antisocial personality.

Among the NESARC cohort, women were twice as likely as men to have a lifetime mood or anxiety disorder. Women were twice as likely as men to have paranoid personality disorder but half as likely as men to have antisocial personality disorder. Women were less likely than men to meet criteria for abuse of cocaine and marijuana and less likely than men to be dependent on alcohol. Women in the ATOS cohort also were more likely to have depression or PTSD.

The long-term outcome for OUD among persons with additional major psychiatric disorders, such as schizophrenia or bipolar I disorder, is not generally known. However, a study that specifically examined the efficacy of additional psychiatric care for dual-diagnosis subjects in an outpatient specialty clinic found that, although they initially had significantly more frequent illicit opioid use than patients without comorbid psychiatric conditions, they had higher retention rates in treatment and, after 6 months of treatment had comparable success rates to the single-diagnosis group.

TREATMENT

The most effective treatment for OUD involves a combination of medication to normalize brain functioning, psychosocial interventions aiming to develop new skills and reduce the risk of relapse, and community-based supportive interventions to integrate the patient with their community. The primary goals of treatment include (1) reduction or cessation of illicit opioid use to initiate abstinence, (2) prevention of relapse to regular use in abstinent individuals, (3) minimizing the medical, psychiatric, and social complications of drug use, especially the risk of overdose and death, and (4) improving the medical, psychological, and social function and the quality of life. Specific treatment goals should be individualized for each patient and include their input throughout the treatment episode.

Table 11.9–5. Substance and Non–Substance-Related Psychiatric Disorders in Opioid Users

Mood, anxiety, and personality disorders among individuals with lifetime opioid use disorders in NESARC by gender.

(*N* = 355)	Men (%) (*N* = 223)	Women (%) (*N* = 578)	Total (%)	Odds Ratio (95% Confidence Interval)
Heroin/other opioids				
Abuse	75.6	67.5	73	0.67 (0.43, 1.04)
Dependence	24.4	32.5	27	1.49 (0.96, 2.31)
Cocaine				
Abuse	37.6	17	30.9	0.34 (0.20, 0.58)***
Dependence	17.4	20.5	18.4	1.22 (0.73, 2.03)
Marijuana				
Abuse	56.2	39.3	50.7	0.51 (0.33, 0.78)**
Dependence	19.4	19.9	19.6	1.03 (0.62, 1.72)
Alcohol				
Abuse	22.6	20.4	21.8	0.88 (0.55, 1.41)
Dependence	67.1	54.5	63	0.59 (0.40, 0.86)**
Sedatives				
Abuse	37.7	23.3	33	0.50 (0.30, 0.85)*
Dependence	9.1	12.9	10.3	1.48 (0.77, 2.87)
Amphetamine				
Abuse	35.9	28	33.3	0.70 (0.45, 1.08)
Dependence	13.3	18.5	15	1.47 (0.80, 2.72)
Hallucinogen				
Abuse	37.4	27.5	34.2	0.63 (0.40, 1.01)
Dependence	9	9.5	9.1	1.06 (0.51, 2.20)
Tranquilizers				
Abuse	37.7	22.7	32.9	0.49 (0.30, 0.78)**
Dependence	8.7	12.4	9.9	1.49 (0.80, 2.81)
Inhalants				
Abuse	13	6.7	10.9	0.48 (0.24, 0.95)*
Dependence	1.7	2.1	1.9	1.22 (0.30, 4.99)
Any other Substance				
Abuse	23.1	21.6	22.6	0.92 (0.60, 1.41)
Dependence	72.4	62.3	69.1	0.63 (0.43, 0.91)*
Any mood or anxiety disorder	68.5	73.8	70.3	1.30 (0.84, 2.00)
Any mood disorder	54.1	71.2	59.7	2.10 (1.35, 3.26)**
Major depression	45.7	63.5	51.5	2.07 (1.38, 3.12)**
Dysthymia	15.3	25.3	18.6	1.88 (1.20, 2.93)**
Manic disorder	16.7	27.1	20.1	1.86 (1.17, 2.94)*
Hypomanic disorder	9	10.9	9.6	1.24 (0.61, 2.50)
Any anxiety disorder	33.6	50.1	39	1.98 (1.35, 2.92)***
Panic disorder w/o agoraphobia	13.2	18.3	14.9	1.46 (0.88, 2.45)
Panic disorder w/ agoraphobia	4.1	6.6	4.9	1.65 (0.63, 4.33)
Social phobia	9.3	19.8	12.7	2.42 (1.36, 4.30)**
Specific phobia	12.9	30.1	18.5	2.91 (1.69, 5.02)***
Generalized anxiety disorder	9.7	17.7	12.3	2.01 (1.13, 3.56)*
Any personality disorder	49.2	50.7	49.7	1.06 (0.74, 1.52)
Paranoid	12.9	22.2	15.9	1.93 (1.17, 3.21)*
Avoidant	9.5	12.4	10.4	1.35 (0.71, 2.57)
Dependent	3	2.3	2.8	0.77 (0.25, 2.37)
Obsessive compulsive	16	23	18.3	1.56 (0.95, 2.57)
Schizoid	11.5	12.8	11.9	1.13 (0.61, 2.11)
Histrionic	9.8	8.5	9.4	0.86 (0.42, 1.76)
Antisocial	34.1	22.4	30.3	0.56 (0.36, 0.87)*

Adapted from Grella CE, Karno MP, Warda US, et al. Gender and comorbidity among individuals with opioid use disorders in the NESARC study. *Addict Behav.* 2009;34(6–7):498–504. Copyright © 2009, with permission from Elsevier. doi: 10.1016/j.addbeh.2009.01.002.

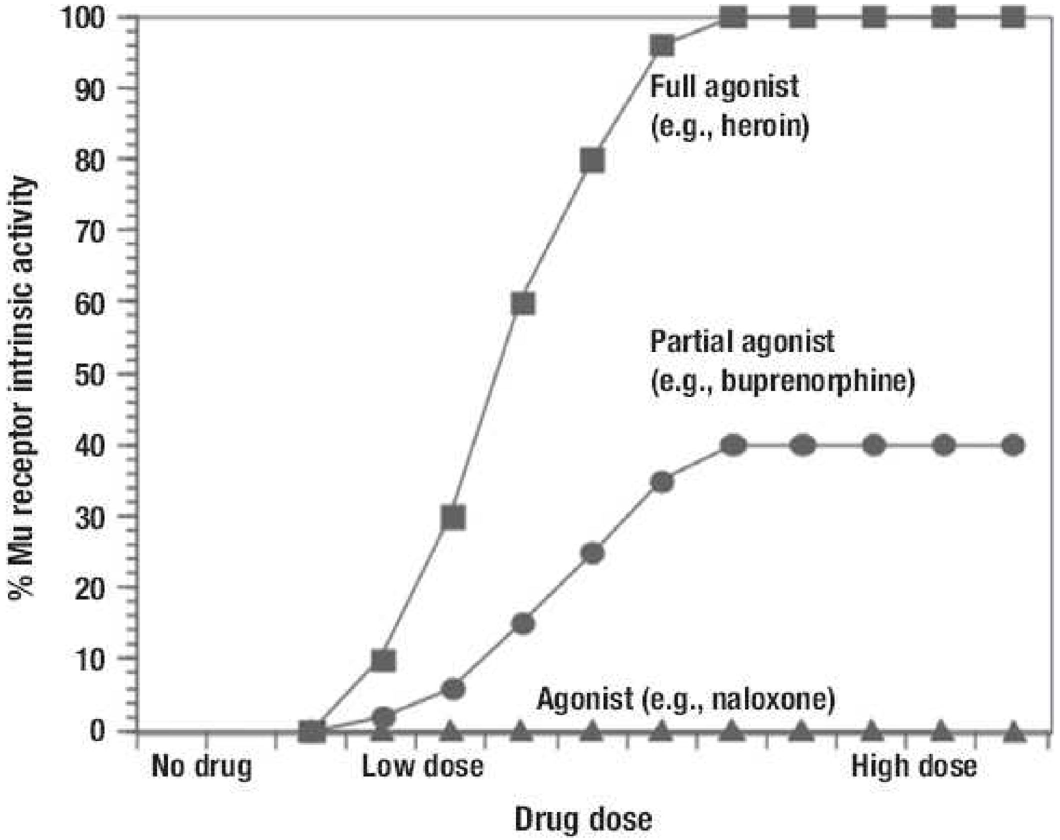

FIGURE 11.9–7. Schematic of pharmacologic action of different classes of agents affecting μ-opioid receptors. A full agonist, such as heroin, induces all of the capacity for intrinsic activity of the opioid receptor. A partial agonist, such as buprenorphine, only allows for half of the activity to proceed, blocking other full agonists from achieving the full biologic effect. Antagonists, such as naloxone and naltrexone, eliminate the potential for any other agent to bind to the receptor and activate the receptor.

Clinical Pharmacology of Medications Used in Treatment of Opioid Use Disorders

The primary pharmacologic treatment strategy, broadly known as medications for opioid use disorder (MOUD) treatment, aims at engaging the μ-opioid receptor. The three classes of medications used in treatment are full agonists, partial agonists, and antagonists, each of which exerts a different pharmacologic effect at the receptor (Fig. 11.9–7). Opioid full agonists, which include methadone, morphine, heroin, and oxycodone, bind to the receptor and activate it to the maximum capacity, with the pharmacologic and clinical effect that is directly correlated with the dose. Methadone is used for detoxification and as a maintenance agent.

The opioid partial agonist buprenorphine strongly binds to the receptor but activates it only partially; at most, the effect is approximately 50% of the maximum effect produced by the full agonist. The agonist effect of buprenorphine is directly correlated with the dose at the lower range of doses, but as the dose increases, the pharmacologic and clinical effect increases very little, and it reaches a ceiling level beyond which no further increase in agonist effect is possible. Functionally, buprenorphine acts primarily as an agonist (producing receptor activation). However, buprenorphine has a high affinity (binding strength) to the receptor, and therefore it can function as an antagonist (blocker), preventing agonists with lower receptor affinity from activating the receptor (i.e., can block the effects of heroin as a therapeutic mechanism). Buprenorphine can also act as a functional antagonist if given to individuals who are physically dependent, with agonists occupying receptor sites. In such a situation, buprenorphine will abruptly displace the agonist from the receptor, but since it will have only a partial effect at the receptor, it will cause a sudden drop in the level of agonist effect (decreased receptor activation), resulting in a precipitated withdrawal. Buprenorphine is used to assist detoxification and as a maintenance agent.

Opioid antagonists (e.g., naloxone, nalmefene, and naltrexone) bind to the receptor but do not exert any activity. These medications function as antagonists—they compete for binding at the receptor with agonists and partial agonists, thereby preventing the agonist from binding to the receptor and exerting any effect (receptor blocking effect as a therapeutic mechanism). In an individual physically dependent on opioids, antagonists given in sufficient doses will displace the remaining agonist from the receptor, suddenly decreasing receptor activation, which leads to precipitation of opioid withdrawal. Naloxone is used in the treatment of overdose; it can rapidly reverse opioid receptor activation, such as following a large dose of heroin or fentanyl. However, higher doses of naloxone can precipitate opioid withdrawal. Naltrexone is a long-acting agent used following detoxification as a strategy to prevent relapse and re-establishment of physiologic dependence.

The last group of medications used in the treatment of OUD includes medications that target various signs and symptoms emerging during acute and protracted opioid withdrawal (e.g., clonidine, lofexidine).

Treatment of Opioid Overdose

Opioid overdose or overdose on mixed opioid–sedatives generally presents with unconsciousness and respiratory depression and constitutes a medical emergency. As most overdoses that come to medical attention are nonlethal, immediate treatment is essential to minimize the impact of anoxia and prevent irreversible injury to the CNS and other organs. Immediate interventions include airway protection and support of breathing in case of very low respiratory rate and use of supplemental oxygen attached to a bag-valve mask. Naloxone, an opioid antagonist, should be administered parenterally as soon as possible to reverse opioid overdose, preferably via the IV or the IM route.

Naloxone is given at the dose of 0.2 to 0.4 mg to a patient in a coma to aid in the maintenance of respiration. Higher initial doses (1 to 2 mg) should be given to patients with apnea or in cardiorespiratory arrest. Naloxone has a relatively short half-life (60 to 90 minutes) and must be repeatedly administered or given as an IV infusion in patients who have overdosed on an opioid with a longer half-life (e.g., XR oxycodone); otherwise, the patient can fall back into a coma after the effect of naloxone wears off. To minimize the risk of precipitated withdrawal, naloxone should be titrated gradually, and its fast elimination allows very accurate dosing without the risk of an unnecessarily prolonged effect.

Naloxone can also be administered subcutaneously or intramuscularly if it is difficult to establish IV access. IN administration can also be used as it is a popular form of naloxone available in take-home programs, acceptable to lay-people responders, but the IN dose has a slightly slower onset of action, and the dose might need to be higher to provide response comparable to the IM doses. Signs of improvement—increase in respiratory rate and pupillary dilation—should occur within 2 to 3 minutes following IV administration. If no response is seen during this time, higher (1 to 2 mg) repeated doses should be given at 2- to 3-minute intervals while ventilation and oxygenation are maintained. There is limited evidence suggesting that overdose following fentanyl administration may require higher doses of naloxone administered much sooner to successfully reverse overdose as compared to heroin overdose.

The primary goal of naloxone administration is to restore adequate ventilation (above 10 respirations/minute) and oxygenation (above 92%) rather than full alertness. Administering higher doses of naloxone can precipitate opioid withdrawal and may cause difficulties with controlling agitation. An acute lung injury (with hypoxia, rales, and mouth frothing) can be seen in rapid overdose reversals in patients with marked hypoventilation. If there is no improvement after 10 mg of naloxone is given, alternative diagnoses must be considered. Most overdoses include other sedatives such as alcohol and benzodiazepines, and therefore strategies aimed at managing toxicity from other substances should also be considered if the response to naloxone is insufficient.

The most important strategy to reduce frequency of overdoes is the medication-based treatment of OUD. Most individuals who overdose on heroin are not in treatment at the time of overdose. Those who enter treatment significantly reduce the use of opioids and other CNS depressants and have a significantly lower risk of overdose.

Treatment of Opioid Withdrawal (Detoxification)

Opioid withdrawal is usually treated in the setting of detoxification, which continues to be a major medical treatment modality used in working with opioid-dependent individuals. The primary goals of detoxification are (1) to provide safe removal of the opioid and the physiologic dependence from the organism and prevent life-threatening complications that might happen if opioid withdrawal is left untreated, (2) to address and stabilize the patient’s physical and psychological health, and (3) to provide psychoeducation about long-term treatment approaches and offer a pathway to access and initiate such treatment.

It is important to note that detoxification by itself is never a sufficient treatment modality. Detoxification that is not followed by a relapse-prevention treatment has a very high rate of early relapse, as much as 90%, and the risk of opioid overdose is much greater in individuals who were detoxified as the detoxification also removes the physiologic tolerance to opioids which protects against overdose. Therefore, “detoxification” has no clinical utility in the absence of a plan for medication-based maintenance treatment and has no evidence for improving outcomes unless serving as a bridge to long-term treatment with medications. Detoxification is necessary before starting relapse-prevention treatment with the opioid antagonist naltrexone, but it is not needed before starting maintenance treatment with the opioid agonist methadone and is needed only partially before starting maintenance treatment with buprenorphine (transition from full to partial opioid receptor activation).

Improved detoxification methods have evolved over the past 150 years, reflecting a more humanitarian view of addicted individuals and a better understanding of addiction as a medical disorder. Medication-assisted detoxification usually occurs under medical supervision in an inpatient or outpatient treatment setting. The inpatient setting is reserved for patients with a history of severe or complicated withdrawal, those expected to have severe withdrawal based on the baseline pattern of substance use (e.g., patients with mixed opioid and sedative–hypnotic withdrawal), and those presenting with co-existing medical and psychiatric problems that will require close monitoring and stabilization. For many individuals, inpatient detoxification may not be a realistic option because of occupational or domestic responsibilities and lack of adequate medical insurance to pay for intensive hospital treatment, whereas outpatient detoxification and relapse prevention treatment initiation allow more individuals to benefit from treatment.

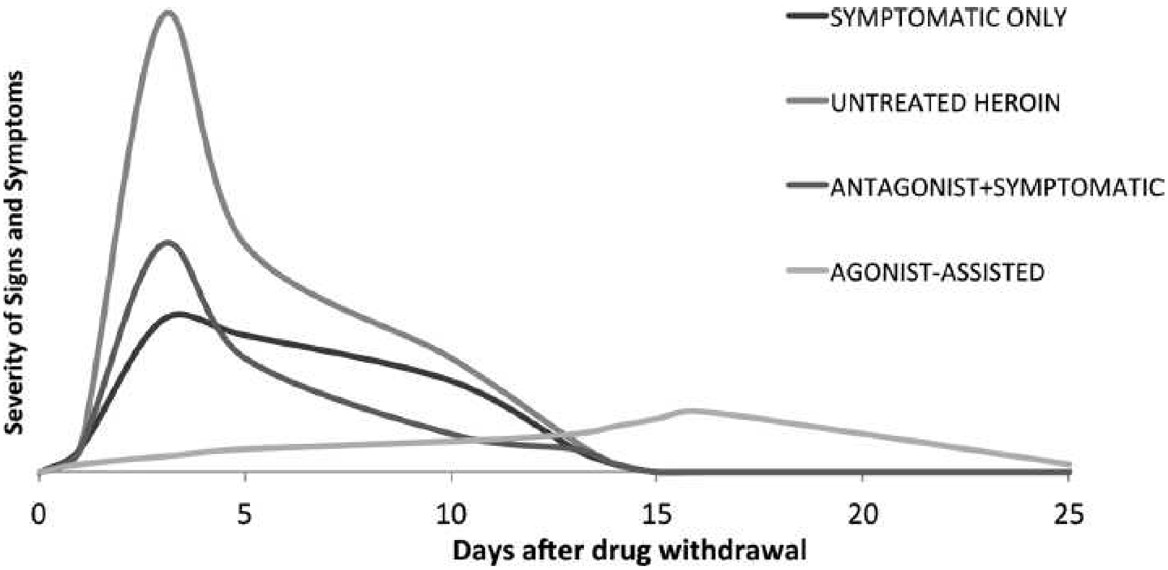

FIGURE 11.9–8. Trajectory of opioid withdrawal symptom severity under different treatment strategies. Patients who are untreated (*orange*) initially experience prolonged and intense withdrawal symptoms. Symptomatic treatment can improve these early withdrawal symptoms. Agonist-assisted treatment can significantly attenuate early withdrawal symptoms but may induce a prolonged state of mild withdrawal after several weeks. Antagonist-assisted withdrawal with symptomatic treatment will shorten the duration of withdrawal though the initial withdrawal severity may be greater. (See eBook for color images.)

Three main detoxification strategies are currently in clinical use: (1) symptomatic treatment only with nonopioid medication, (2) agonist-assisted detoxification, and (3) antagonist-assisted detoxification with symptomatic treatment. Each of these regimens is expected to produce a different dynamic of the onset, severity, and duration of withdrawal symptoms (Fig. 11.9–8). The main difference is that in agonist-assisted detoxification, withdrawal symptoms tend to be more severe toward the later phase of treatment when the agonist is tapered off and may continue after treatment completion, while in symptomatic treatment only and antagonist-assisted detoxification, withdrawal symptoms are more severe early in treatment and are usually minimal later in treatment.

Symptomatic Treatment of Opioid Withdrawal. Opioid withdrawal is characterized by a wide range of signs and symptoms affecting multiple organ systems. The severity of withdrawal can be significantly alleviated with medications directed at each component of withdrawal. This so-called "symptomatic treatment" or treatment with ancillary (adjunctive) medications is frequently utilized, particularly in situations where opioids are not used in detoxification because of legal restrictions or concerns about medication diversion. Ancillary medications are also a mainstay of treatment during antagonist-assisted detoxification and can also be used during agonist-assisted detoxification to alleviate residual withdrawal symptoms after the agonist is tapered off. Similarly, symptomatic treatment can be offered to individuals attempting initiation of treatment with buprenorphine to help with withdrawal emerging during the initial agonist-washout period.

The following classes of medications are used to target specific syndromes emerging during opioid withdrawal:

1. Autonomic (sympathetic) arousal can be decreased using α-2 adrenergic agonists. Clonidine is the most widely used symptomatic-treatment medication, effective in relieving a wide range of autonomic withdrawal symptoms (sweating, restlessness, tremor, rhinorrhea). Sedation is a particularly welcomed effect of clonidine; however, other aspects of withdrawal (such as muscle/bone pain and insomnia) are not helped by clonidine. Clonidine is given orally, starting at doses of 0.1 to 0.3 mg three to four times a day or given as a 0.1 to 0.2 mg/day patch. In outpatient settings, doses tend to be lower due to the potential for postural hypotension that is more difficult to monitor.

Hypotensive risks of clonidine are most pertinent in the situation of fluid loss (vomiting/diarrhea). Maintaining aggressive oral hydration with electrolyte fluids, best used as a preventive intervention, is helpful in increasing the tolerability of clonidine. Lofexidine is another FDA-approved adrenergic agent with an effectiveness profile similar to clonidine, with a risk of hypotension that appears to be lower compared to clonidine.

2. Anxiety and restlessness are particularly troublesome aspects of opioid withdrawal and often the reason why some patients prematurely withdraw from treatment. Clonidine may only partially help with anxiety, and therefore anxiolytic/sedating agents that work through a different mechanism may be very helpful, especially when given in combination with clonidine. Benzodiazepines are the most commonly used anxiolytics because they have a wide margin of safety and are available as several agents with varying duration of action, allowing great flexibility in use. Clonazepam (0.5 to 1.0 mg), chlordiazepoxide (25 to 50 mg), or oxazepam (15 to 30 mg) can be given two to four times per day to prevent and alleviate anxiety/agitation. Benzodiazepines are especially suited for use in the inpatient setting as there is concern about the potential for benzodiazepines to be misused or unsafe (if taken in combination with an opioid or alcohol) during outpatient detoxification. Other classes of sedative agents that can be used include antihistamines such as diphenhydramine or hydroxyzine.
3. Insomnia, which at times can be severe, is seen in most patients undergoing opioid withdrawal, and targeting insomnia can greatly improve the tolerability of detoxification though rarely can it be completely treated. Specific hypnotic agents used include zolpidem (5 to 10 mg) or zopiclone (7.5 to 15 mg), but sedating antidepressants (e.g., trazodone 50 to 150 mg) or atypical neuroleptics (e.g., quetiapine 50 to 200 mg) can also be used.
4. Other symptoms of withdrawal can also be treated symptomatically. Musculoskeletal pain can be alleviated using nonsteroidal anti-inflammatory drugs (ibuprofen or ketorolac) or antispasmodic agents (cyclobenzaprine or methocarbamol). GI distress responds well to antiemetic agents (prochlorperazine, ondansetron), and diarrhea can be treated with loperamide, or octreotide in severe cases.

Agonist-Assisted Detoxification. Two agonist medications are approved in the United States for use in the treatment of opioid withdrawal: methadone and buprenorphine. Both methadone and buprenorphine are effective in treating signs and symptoms of opioid withdrawal and have a long elimination half-life that makes them good candidates for detoxification. Agonist-assisted detoxification involves the substitution of the short-acting agonist at the receptor site (usually heroin or oxycodone) with a long-acting agonist, followed by a gradual reduction in the dose of the agonist to minimize the severity of associated withdrawal. The severity of associated withdrawal is directly proportional to the rapidity of dose reduction; however, it tends to be relatively greater in the final stages of agonist discontinuation, especially during methadone taper (necessitating extended taper), and persists after the last dose is given.

Methadone is a full opioid agonist and can be given immediately after the cessation of opioids to preclude withdrawal. However, buprenorphine is a partial opioid agonist, and it will precipitate withdrawal if given to a physically dependent patient who does not show signs of withdrawal; thus, a waiting period to allow for the spontaneous withdrawal to emerge is necessary before the first buprenorphine dose can be given. Therefore, it may be more difficult for patients to initiate detoxification with buprenorphine as compared to methadone. On the other hand, the use of buprenorphine has fewer restrictions, a better safety profile, and a lower severity of terminal withdrawal after administration of the final dose, making it a more appropriate agent in most treatment settings.

METHADONE. For patients who have been using heroin or illicit prescription opioids, the initial dosage of methadone is usually up to 30 mg orally. If objective signs of withdrawal persist after the first dose of methadone, additional doses can be given after approximately 2 to 4 hours. As a general rule, initial stabilization does not require more than 40 mg of methadone during the first 24 hours, but it can be higher in the inpatient setting if clear signs of moderate or severe withdrawal are present after the initial 6 to 10 hours of the observation period. For example, individuals with easy access to large doses of high-potency opioids, such as health care providers, may require higher initial doses. After 24 to 48 hours of dose stabilization, the patient’s methadone dose can be gradually reduced by 10% to 20% or by 5 mg per day. Inpatient detoxification with methadone is usually completed within 1 week.

In case of detoxification in an outpatient setting, the first-day dose should not exceed 40 mg. Over the next few days, the dose may be raised (generally by no more than 10-mg increments every 2 to 3 days) until the patient has no discernible withdrawal symptoms. In most cases, the dose of 40 to 60 mg per day is sufficient to prevent any opioid withdrawal; however, this dose may not be adequate to eliminate craving. In case of continuing opioid use, the dose of methadone needs to be increased further, and the attempt at detoxification needs to be postponed with the goal changed to stabilization (no withdrawal symptoms, craving, or opioid use). In general, a period of dose stabilization, at minimum 4 weeks in duration, is necessary prior to a slow dose reduction schedule. Studies of outpatient methadone withdrawal show that gradual dose reductions (e.g., 3% per week) have a greater likelihood of success versus more rapid dose reductions (e.g., 10% per week). Dose reductions are generally well tolerated, but as the dose is reduced to below 20 to 30 mg per day, there is often increasing withdrawal distress with each dose change and increased risk of illicit opioid use. Another factor that interferes with the final discontinuation is the negative expectancy effect, where the patient expects an increase in withdrawal symptoms as the medication is stopped and experiences significant distress. Conducting single-blind detoxification, where the patient agrees to be blinded to the actual dose and rate of taper, can improve tolerability and success. Outpatient detoxification using an opioid agonist can usually be accomplished in 1 to 2 months, but slower reductions, with periods of stable dose before subsequent dose tapers, can be more effective in patients who have difficulty tolerating the process.

BUPRENORPHINE. While methadone has been safely and effectively used for detoxification for almost 50 years, buprenorphine has become an alternative option following FDA approval of sublingual buprenorphine products indicated for the treatment of opioid withdrawal. The sublingual buprenorphine is marketed in two formulations: a *monotherapy product* containing buprenorphine alone and a *combination therapy product* that contains buprenorphine and naloxone. The monotherapy product is usually used during buprenorphine-assisted detoxification, though a combination product can also be used.

Detoxification using buprenorphine can be safely accomplished in the inpatient setting under direct monitoring. The first dose of buprenorphine should be given after a period of agonist washout when the patient experiences at least moderate severity of withdrawal. It may take 12 to 24 hours, or even longer, for the withdrawal syndrome to emerge, and it is preferred that objective signs are present (yawning, rhinorrhea, mydriasis, chills). Some patients are unable to tolerate the initial agonist washout and may benefit from medications providing symptomatic relief (e.g., clonidine, hypnotics) or transition to treatment with methadone.

The first dose of buprenorphine typically given is 1 to 2 mg to minimize the risk of worsening withdrawal, though it may be greater (4 mg) in case of severe withdrawal. Subsequent 2 mg doses are given every 2 to 4 hours until most withdrawal symptoms are eliminated. Usually, the total dose needed to effectively suppress withdrawal on the first day is 8 to 10 mg. Buprenorphine taper is usually started on the second day, with the daily dose decreased by 2 mg every day, followed by 1 mg decreases during the last 2 days of taper. The complete taper can usually be accomplished in 5 to 7 days, and the long elimination half-life of buprenorphine (31 to 35 hours) is very useful in extending and gradually reducing the medication effect after it is discontinued.

Buprenorphine-assisted detoxification can also be accomplished in the outpatient setting, and the strategy is generally similar to methadone-assisted detoxification. Ideally, the first day’s dosing should occur under medical supervision to minimize the risk of buprenorphine-precipitated withdrawal. Patients are instructed to arrive in withdrawal, which typically starts 12 hours after the last dose of a short-acting opioid or 24 hours or more after a long-acting opioid like methadone. The first dose of buprenorphine is usually 1 to 2 mg, and if no precipitated withdrawal is observed, subsequent 2-mg doses are given at the clinic every 2 hours, with additional doses given to take at home until the withdrawal symptoms are eliminated. If withdrawal worsens after the first buprenorphine dose, it is best to wait for 4 to 6 hours before giving the next buprenorphine dose, while the precipitated withdrawal can be treated symptomatically with medications such as clonidine. The total daily dose needed on the first day is usually 8 to 12 mg. Subsequent daily dose increases of 2 to 4 mg are recommended if needed to relieve opioid withdrawal and craving and initiate opioid abstinence. Once a period of stabilization is established, slow dose reductions can begin, usually in 2-mg dose increments. Because of the prolonged elimination half-life of buprenorphine, dose reductions are best introduced at weekly intervals, as the patient may not experience the effect of dose reduction until several days later. As is the case with methadone, slower dosage reduction schedules are usually better tolerated and less likely to prompt relapse. There appear to be less severe withdrawal symptoms after completion of buprenorphine taper versus methadone taper, and buprenorphine-assisted detoxification becomes a more widely used intervention, especially since the use of buprenorphine for that purpose does not require a special license that is needed for methadone.

Antagonist-Assisted Detoxification. Although agonist taper and symptomatic treatment can effectively minimize the distress of detoxification, these strategies do not shorten and may extend the duration of detoxification to 7 to 10 days, and many individuals drop out of treatment before completion. Several methods have been developed to shorten the duration of detoxification and to use it as a starting point for antagonist-based treatment to prevent postdetoxification relapse, the procedure known as "naltrexone induction." To accomplish it, opioid antagonist such as naltrexone is introduced during detoxification, and the emerging withdrawal symptoms are treated symptomatically (e.g., clonidine and clonazepam combination). To minimize the severity of naltrexone-precipitated withdrawal, naltrexone is introduced 2 to 3 days after the last dose of opioid agonist in low-dosage (e.g., 1 to 3 mg) and quickly titrated upward over the ensuing 2 to 4 days up to the full blocking dose of naltrexone (25 to 50 mg oral tablet or an injection of XR preparation). As a variant to this procedure, 1 day of buprenorphine (4 to 8 mg) can be used as a step-down from a full agonist before the initiation of naltrexone.

Precipitated withdrawal can be quite severe if the initial dose of naltrexone is large and given too early (before completing detoxification), and it could be dangerous due to severe fluid loss, hypotension, or delirium. Therefore, early naltrexone initiation requires close monitoring, preferably on an inpatient basis, and is best used in patients without medical problems that may worsen during withdrawal (e.g., asthma, diabetes, heart disease).

Detoxification under heavy sedation or general anesthesia, so-called ultrarapid procedures, has been advocated in an effort to initiate naltrexone immediately after cessation of opioids and to shorten the duration of the procedure (resolution of physiologic dependence) to 1 to 2 days. While these methods may increase the likelihood of successful induction onto naltrexone compared to more gradual and longer procedures, these methods require a highly experienced anesthesiology team and are expensive. Moreover, these methods can produce life-threatening, though rare, adverse events, including aspiration pneumonia, pulmonary edema, renal failure, and ketoacidosis, complications that are never seen in the course of traditional detoxification methods. Therefore, ultrarapid methods should not be suggested as an alternative to other methods of detoxification; however, in highly selected cases, there might be a benefit to using these methods to initiate relapse-prevention treatment with long-acting opioid antagonists.

Relapse After Detoxification. Extensive evidence shows that detoxification—resolution of opioid withdrawal after the cessation of opioid use—is by itself not adequate to treat OUD. In the absence of relapse-prevention treatment, most patients will relapse within the first few months, and such relapse can be very dangerous, as detoxified individuals do not have physiologic tolerance to protect them from acute opioid toxicity. Therefore, detoxification can only be seen as the first step of engagement in treatment, while the primary treatment goal is the prevention of relapse to illicit opioid use.

Relapse-prevention treatment strategies that can be implemented after the cessation of opioid use generally include a combination of pharmacologic treatment with antagonist or agonist agents and psychosocial interventions. The outcomes of treatment that includes only psychosocial approaches, even if delivered in a high-quality and optimized manner, are inferior to treatments that include pharmacologic support. The exception might be a long-term residential treatment (e.g., therapeutic community) where individuals have no ready access to opioids while undergoing intensive treatment and rehabilitation and are monitored for protracted withdrawal symptoms (e.g., insomnia, low energy, anxiety, and drug craving), emerging during early abstinence. Nevertheless, the strongest evidence points to the benefits of added pharmacologic support to minimize the risk of relapse during the initial abstinence from opioids.

Treatment of Opioid Use Disorder

The ultimate goals of treatment vary not only for patients but also for providers. There are several perspectives as to what constitutes the primary goal of treatment for OUD, ranging from sustained, life-long abstinence from all psychoactive substances to the minimization of harm to the individual that continues using opioids. For the medical community, treatment goals usually include protection against overdose and death, as well as the reduction in symptoms of the disorder, more specifically, cessation of illicit opioid use and the elimination of craving and withdrawal with resulting improvement in physical and psychological health. Often, complete abstinence early in treatment is an unattainable goal, and more realistic immediate goals are treatment engagement, reduction of use amount or frequency, and the elimination of particularly risky drug-use practices. For others, the primary goal of treatment is the achievement of recovery, often defined as living in abstinence from all psychoactive substances, including psychoactive medications, supported by changes in many aspects of life (social, recreational, and vocational). Unfortunately, sometimes the goal of recovery is seen as incompatible with the use of medications, particularly opioid agonists. However, these two approaches do not have to be exclusionary and can be complementary on a time continuum, with more immediate medical goals as a first step leading to long-term lifestyle changes.

Psychosocial Interventions. A broad range of psychosocial interventions has been used and evaluated as treatment strategies for opioid dependence. Psychological interventions commonly include unstructured, supportive psychotherapy and drug counseling but may also include various psychological techniques such as cognitive behavioral therapy (CBT) or contingency management (CM). Social interventions include assistance with daily needs (food, clothing, housing), employment counseling, assistance with access to health care, and engagement in social support networks.

CBT is a particularly valuable approach to the treatment of OUD with strong evidence for its effectiveness. CBT is a structured and goal-oriented approach based on the assumption that drug taking is a learned behavior. During treatment, patients are introduced to and practice (learn) new skills and cognitive strategies to replace the dysfunctional behaviors and cognitions and, through these changes, support abstinence. Clinical trials evaluating psychosocial interventions have shown that regular counseling reduces illicit drug use, and the quality of counseling and rapport between patient and counselor significantly impact outcome measures. In patients with higher levels of psychiatric symptoms like depression, CBT or supportive psychotherapy produced superior results compared to drug counseling alone. CBT can also be effective in reducing HIV risk behaviors associated with IV drug use.

A CM approach is often used as a part of treatment that focuses on promoting new behaviors that are competing with drug use. CM involves giving patients tangible rewards to reinforce positive behaviors such as abstinence, treatment attendance, and compliance with medication. Monitored urine collection and testing with immediate consequences are necessary for CM effectiveness. Patients do not receive cash payments for abstinence, but rather they receive vouchers to purchase goods and services consistent with their treatment goals. Toxicology also provides an objective index of treatment progress, and test results can be discussed in therapy sessions in order to promote better understanding of situations that lead to continued drug use. Both CBT and CM can work in synergy with medication treatment as these can be focused on medication adherence.

NA is a self-help group of abstinent drug abusers modeled on the 12-step principles of AA. This approach emphasizes two strategies to maintain abstinence: spirituality (faith and hope for recovery) and pragmatism (belief in doing “whatever works” to avoid using drugs again). Such groups now exist in most large cities and online, providing useful group support. Although the evidence supporting the usefulness of such groups has been very limited, observational studies show their effectiveness, and therefore, patients receiving medications should be encouraged to participate in the self-help groups. Because 12-step groups emphasize abstinence from all psychoactive substances, patients treated with opioid agonists may find themselves being ostracized in traditional NA groups. Alternative approaches were developed for patients on agonist medications, including Methadone Anonymous (MA), a variant of a 12-step program specifically for patients treated with methadone. Additional approaches, such as family and network therapy, can improve compliance with medication and can enhance outcomes in opioid addiction treatment.

Pharmacologic Interventions. In the United States, four medications are currently approved by the U.S. FDA for the treatment of OUD (MOUD); methadone (used since the mid-1960s and approved in 1972), oral naltrexone (used since the 1970s and approved in 1984), buprenorphine (approved in 2002), injectable, long-acting naltrexone (approved in 2010), and several newer XR buprenorphine formulations (2017/2021). LAAM (levacetylmethadol), a long-acting agent similar to methadone that was in use since its approval in 1993 but was discontinued in Europe and the United States in the early 2000s due to limited uptake (in the United States), and cardiac toxicity concerns (in Europe) (Table 11.9–6).

Several large, multisite, pragmatic randomized clinical trials have now been completed by the National Drug Treatment Clinical Trials Network (NIDA-CTN), consistently demonstrating the effectiveness and treatment response characteristics of pharmacologic interventions. Patients assigned to buprenorphine and methadone had a comparable rate of treatment success (avoiding relapse) in an outpatient setting, between 10% and 30% at 12 weeks. Patients who presented with prescription OUD rather than heroin use disorder also substantially benefitted from buprenorphine maintenance treatment, compared with a detoxification and taper program. Patients assigned to treatment with daily buprenorphine versus monthly XR-naltrexone have shown higher rates of inpatient induction onto buprenorphine. However, patients discharged to outpatient care on buprenorphine or XR-naltrexone had comparable rates of success (avoiding relapse), between 44% and 48%, at 24 weeks.

METHADONE. In the United States, there are over 400,000 patients in outpatient treatment programs receiving methadone (as of 2021). The original approach pioneered in the 1960s emphasized using methadone dosage sufficient to suppress opioid drug craving and producing a cross-tolerance blockade of other opioids. Doses of methadone were in the 80 to 120 mg per day range, and patients were encouraged to remain on methadone indefinitely, working toward recovery. Subsequent treatment programs used lower doses of methadone (less than 50 mg per day) with fewer ancillary services as the methadone maintenance treatment in support of recovery was seen not as a goal in itself but rather a transitional stage before the eventual detoxification. Lower doses of methadone were often adequate for eliminating withdrawal and retaining patients in treatment, but often these doses only partially suppressed craving and drug-seeking behavior and did not produce adequate cross-tolerance to block large doses of heroin. Empirically, patients maintained on lower doses of methadone are far more likely to discontinue treatment, and those who do so are unlikely to maintain abstinence from illicit opioids.

Table 11.9–6.
Medications for Opioid Use Disorder (MOUD)

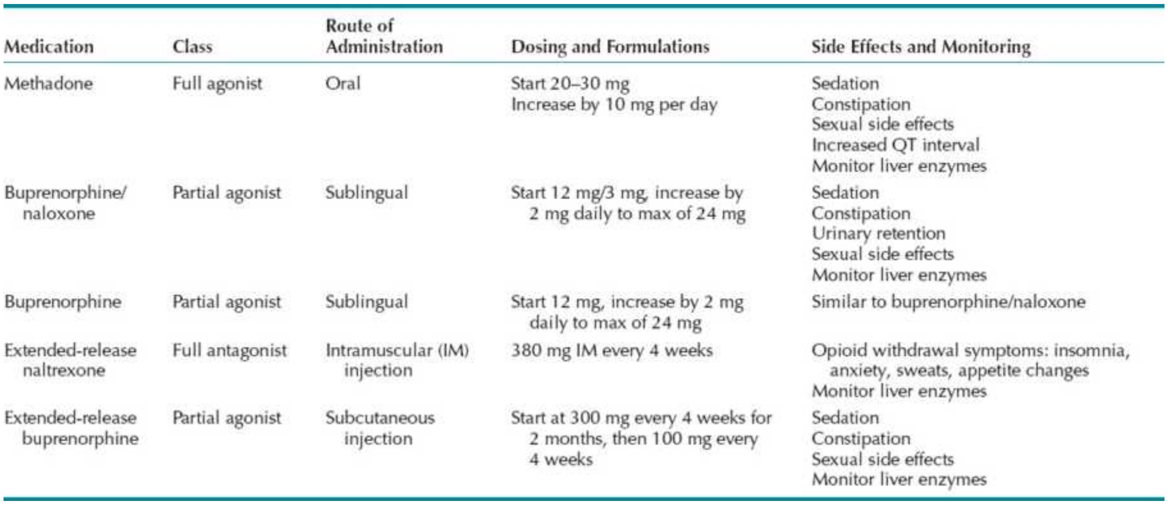

Medication	Class	Route of Administration	Dosing and Formulations	Side Effects and Monitoring
Methadone	Full agonist	Oral	Start 20–30 mg Increase by 10 mg per day	Sedation Constipation Sexual side effects Increased QT interval Monitor liver enzymes
Buprenorphine/ naloxone	Partial agonist	Sublingual	Start 12 mg/3 mg, increase by 2 mg daily to max of 24 mg	Sedation Constipation Urinary retention Sexual side effects Monitor liver enzymes
Buprenorphine	Partial agonist	Sublingual	Start 12 mg, increase by 2 mg daily to max of 24 mg	Similar to buprenorphine/naloxone
Extended-release naltrexone	Full antagonist	Intramuscular (IM) injection	380 mg IM every 4 weeks	Opioid withdrawal symptoms: insomnia, anxiety, sweats, appetite changes Monitor liver enzymes
Extended-release buprenorphine	Partial agonist	Subcutaneous injection	Start at 300 mg every 4 weeks for 2 months, then 100 mg every 4 weeks	Sedation Constipation Sexual side effects Monitor liver enzymes

Meta-analytical studies show methadone's clear superiority over treatments without medication with regard to improving treatment retention and reducing heroin use. The majority of patients treated in methadone programs also show decreased nonopioid drug use, criminal behavior, symptoms of depression, and increased gainful employment. The effectiveness of opioid maintenance treatment is influenced by the dose and the quality of the additional services provided. The optimal dose is generally in the 80 to 120 mg/day range, with some patients requiring higher doses to achieve an optimal clinical response. Effective individual drug counseling can significantly augment treatment outcomes in terms of decreasing illicit opioid and cocaine use, decreasing needle sharing and crime, and increasing psychological well-being.

Among patients who initiate methadone maintenance, there is an increase in mortality due to respiratory depression (due to ongoing illicit opioid use); however, after that time, there is a reduction in mortality that remains as long as the patient is maintained on the medication. The risk of overdose is again increased after patients stop taking methadone. Methadone is relatively safe as a long-term maintenance agent. The main safety concern is the risk of excessive sedation with a rapid dose escalation and where other sedating agents, prescribed or illicit, are used in combination with methadone.

Among the most common side effects with methadone are constipation, excessive sweating, decreased libido, and sexual dysfunction. Opioids can reduce plasma levels of testosterone and follicle-stimulating hormone (FSH). Sleep abnormalities (insomnia and nightmares) and altered electroencephalogram (EEG) sleep patterns are frequently found during the first months of methadone treatment and return to baseline, but complaints of sleep abnormalities may persist.

Clinicians should be aware that a variety of medications could cause a change in plasma levels of methadone by inducing or inhibiting hepatic enzymes that metabolize methadone (CYP3A4). For example, adding rifampin, phenytoin, carbamazepine, darunavir, or efavirenz to the methadone regimen may cause a decrease in methadone levels with resulting opioid withdrawal, while adding fluconazole/ketoconazole, fluvoxamine, erythromycin, or delavirdine may produce an increase in methadone levels, causing excessive sedation. Methadone appears to increase the QTc interval in some patients, an effect that can increase the risk of a cardiac conduction abnormality (torsade de pointes). This effect is dose related and is more commonly seen with very high doses of methadone. However, other risk factors for QTc prolongation may increase the risk for cardiac complications in patients treated with methadone, including coadministration with other medications increasing QT interval, such as amifampridine, metronidazole, thioridazine, or ziprasidone.

In the United States, methadone can be dispensed only at an outpatient opioid treatment program (OTP) certified by the Substance Abuse and Mental Health Services Administration (SAMHSA) and registered with the Drug Enforcement Administration (DEA) or to a hospitalized patient who is admitted for a medical or psychiatric reason other than to treat OUD. SAMHSA-certified OTP facilities provide daily doses of methadone under direct supervision until the patient is deemed stable enough to receive take-home doses. The decision to dispense methadone from the OTP, and to be taken unsupervised at home, should balance the benefits of reduced frequency of clinic visits in stable patients with the risks of misuse (e.g., injection) and diversion of methadone to the illicit drug market.

BUPRENORPHINE. Buprenorphine is a high-affinity partial μ-opioid receptor agonist, κ-receptor antagonist, and ORL-1 agonist. At lower doses of up to 16 mg, buprenorphine has a dose-dependent opiate-like effect, but at higher doses, there is progressively less increase in opiate-like effects up to a point where a maximum effect is reached. This ceiling to the μ-opioid agonist effects of buprenorphine limits adverse effects such as respiratory depression, which improves its safety profile. Buprenorphine has poor oral bioavailability, and therefore it is usually administered transmucosally (sublingually as a tablet or film and as a buccal film). To decrease the potential for misuse of buprenorphine as an outpatient maintenance agent, pharmaceutical formulations contain buprenorphine mixed with naloxone in a four-to-one dose ratio. Because naloxone has poor sublingual bioavailability, it exerts minimal effect when a medication is taken sublingually as prescribed. However, if a buprenorphine–naloxone product is dissolved and injected by a person physically dependent on opioids such as heroin, naloxone exerts a maximal effect, blocking some of the effects of buprenorphine and, in certain cases, precipitating withdrawal.

Buprenorphine is highly lipid soluble and has a long elimination half-life (24 to 48 hours), allowing for once-daily dosing and even less frequent dosing with higher doses. Agonist effects of buprenorphine prevent withdrawal symptoms from emerging. Because buprenorphine has a higher receptor affinity than most other opioids, it effectively blocks the effects of agonists by a direct occupation of the receptor sites (unlike methadone, which produces a blocking effect primarily via the mechanism of cross-tolerance, i.e., maintaining high levels of tolerance). Buprenorphine has been shown to be effective in the treatment of opioid dependence in an office-based setting, and buprenorphine 12 mg is equally effective in promoting treatment retention and suppression of heroin use as 50 to 60 mg of methadone. There is preliminary evidence that buprenorphine doses above 16 mg per day may be more effective than lower doses.

To avoid precipitation of opioid withdrawal during initiation of buprenorphine, patients should present with at least mild severity withdrawal symptoms with gradual buprenorphine dose titration during the first day, similarly to the protocol used for buprenorphine-assisted detoxification (see above). Unlike methadone, there is little concern about overdose during treatment initiation with buprenorphine. However, combining buprenorphine with benzodiazepines or other sedative–hypnotics can produce excessive sedation, and there are case reports of fatalities associated with parenteral misuse of buprenorphine along with a benzodiazepine. Common side effects of buprenorphine include constipation, nausea, headache, upset stomach, excessive sweating, somnolence, and decreased interest in sex. Most of these side effects are mild in severity and decrease over time.

Similarly to methadone, buprenorphine is primarily metabolized by the CYP3A4, and medications that block this enzyme, such as atazanavir/ritonavir, may produce excessive sleepiness while enzyme inducers, such as rifampin, may precipitate opiate withdrawal. The clinical impact of pharmacokinetic interactions with buprenorphine is generally less significant than with methadone; however, dose adjustment may be necessary to achieve optimal clinical response and minimize risks for adverse effects. There is no evidence that buprenorphine produces significant cognitive impairment or electrocardiogram (ECG) QTc prolongation.

In the United States, treatment providers can treat OUD using buprenorphine in a manner similar to treatment with other controlled substances (unlike methadone which cannot be prescribed for that purpose, it can only be dispensed in the OTP setting). The law passed in 2000 (the Drug Addiction Treatment Act) allowed office-based opioid treatment (OBOT) of opioid dependence with schedule III, IV, or V medications that were approved for an opioid dependence indication, with buprenorphine being the only opioid meeting these criteria. In order to prescribe buprenorphine, physicians must apply for permission (known as an X-waiver) granted by the DEA. As of 2021, waivered physicians are qualified to treat up to 30 patients in the first year of their waiver. Providers who wish to treat a larger number of patients need to complete a training course and attest to the provision of psychosocial treatment, after which they can request permission to increase the patient cap to 100, and even up to 275, provided they meet additional criteria. Buprenorphine can also be dispensed in a manner similar to methadone in the OTP without the DEA waiver or restriction on the number of patients treated by a prescribing physician.

Several formularies of XR buprenorphine have been approved by the U.S. FDA. In 2016, FDA approved a polymeric rod implant of buprenorphine that slowly releases low doses of buprenorphine over the course of 6 months before needing replacement. However, the product was discontinued due to difficulties in commercialization as the product was indicated for stable patients treated with low-dose buprenorphine, implant insertion and removal required a small in-office incisional procedure, which was burdensome, and the storage and management of controlled substance were also difficult for many providers. In 2017, the FDA approved a once-monthly subcutaneous injection formulation of buprenorphine, which achieves a constant, therapeutic blood concentration of buprenorphine.

Buprenorphine sustained release injection (Sublocade) is usually initiated after the patient is stabilized on the daily doses of sublingual buprenorphine preparation. The initial 300-mg dose of SC buprenorphine is followed by another 300-mg dose given a month later. After that maintenance dose of 100 mg is administered monthly with the buprenorphine blood level produced by this regimen roughly corresponding to 24 mg of daily sublingual buprenorphine (approximately 3 ng/mL). The maintenance can continue with monthly doses of 300 mg, which produces higher blood levels and appears to be more beneficial for patients who inject opioids.

NALTREXONE. The use of opioid antagonists to treat opioid dependence is based on the idea that after detoxification, relapse can be prevented using an opioid receptor antagonist, which directly blocks heroin from exerting euphoric effects. Individuals who resume opioid use while maintained on the antagonist usually stop using after a few attempts, and with continuing abstinence, drug-seeking behavior gradually stops (behavioral extinction mechanism). In addition to providing blockade of opioid effects in case of use, patients taking naltrexone experience less craving in the presence of opioid-related cues and are more likely to maintain full abstinence. The effectiveness of antagonist-based treatment is contingent on full adherence to the medication regimen.

Naltrexone is a high-affinity and long-acting μ-opioid receptor antagonist that provides a dose-dependent blockade of heroin and synthetic opioids effects. Naltrexone may be an attractive treatment option to some patients, as in contrast to methadone or buprenorphine, it is not a controlled substance, so it can be prescribed without a special license, it has no misuse or diversion potential, and there is no withdrawal distress on missing the dose or stopping treatment as patients treated with naltrexone are not physiologically dependent on opioids.

Naltrexone can be useful in selected patients such as (1) those who had limited or no treatment success with agonists (e.g., continued to experience craving, frequently used illicit opioids, or had intolerable side effects), (2) patients with high motivation for abstinence from all opioids such as those very active in an NA 12-step program, and (3) patients who have been successful in agonist therapy and wish to discontinue it but would like to be protected against relapse with medication. However, naltrexone remains a less popular treatment option in countries where agonists are widely available due to the need to be fully detoxified before treatment can be started, poor medication adherence in many patients, and protracted withdrawal symptoms that may persist in the first few weeks of treatment with naltrexone.

Naltrexone is available as an oral 50-mg tablet that can be taken daily or three times a week (e.g., 100 mg on Mondays and Wednesdays and 150 mg on Fridays) to provide stable blood levels to block substantial doses of heroin. Naltrexone is also available in XR preparations that can maintain sufficient blocking serum levels of naltrexone. One such XR preparation available and approved by the FDA is a monthly IM injection (Vivitrol). There are also implantable preparations of naltrexone with a much longer duration of action (3 to 6 months) available in other countries; however, only one of these implants has full regulatory approval of the national drug monitoring agencies (an implant approved in Russia).

The effectiveness of treatment with injectable XR-naltrexone has more than double the success rate of treatment with oral naltrexone, and it became a first-line treatment when available. Oral treatment can still be effective in the context of involving a family member or a professional to provide ongoing support and monitoring of medication administration, which is critical at the beginning of treatment. Alternatively, adherence with oral naltrexone may be improved by incorporating a CM behavioral approach into the treatment.

Naltrexone can be readily initiated in patients who have undergone opioid withdrawal in the inpatient and residential setting for more than a week after completing the opioid taper. Initiating naltrexone maintenance is more challenging in an outpatient setting because of the need for an extended opioid-free period during which patients continue to experience mild severity opioid withdrawal. To determine that a therapeutic dose of naltrexone can be administered without precipitating withdrawal, a urine drug test negative for all opioids is necessary to confirm opioid abstinence. In addition, a test dose of an antagonist (challenge), either with naloxone (0.8 to 1.2 mg IM) or low-dose naltrexone (6 to 12.5 mg), can be administered to determine that there is no residual physiologic dependence and to confirm that the patient will not experience protracted withdrawal following the administration of the full dose (380 mg of XR-naltrexone or 50 mg oral naltrexone).

Because naltrexone is not metabolized by the P450 cytochromes, interactions with other medications are not of clinical concern. Naltrexone is contraindicated in patients taking any opioids, including partial and mixed agonists such as buprenorphine and nalbuphine, as it will precipitate severe withdrawal in such patients.

OTHER MEDICATION-BASED APPROACHES. There are a considerable number of patients who are unable to benefit from treatment with methadone, buprenorphine, or naltrexone, and alternative approaches to provide these individuals with effective treatment have been explored. This includes a maintenance treatment with slow-release oral morphine and with a medically prescribed pharmaceutical-grade heroin.

Oral morphine has been used in individuals who are unable to tolerate or do not find the effects of methadone useful as being different from the effects of heroin. In those patients, morphine can have positive effects on mood and sleep and eliminate craving for heroin in contrast to methadone. It may also minimize the use of other medications (e.g., sedatives) often seen when such patients are treated with methadone. Maintenance treatment with a once-daily slow-release preparation of morphine is available in several European countries. However, because of the short duration of action, many such patients continue to need methadone to prevent withdrawal symptoms overnight, and the use of morphine in a treatment setting is complicated by the difficulties in assessing heroin use to measure outcomes. At the present time, there is not a sufficient number of experimental studies to draw generalizable conclusions.

Prescription heroin as a treatment strategy for heroin dependence has been available in the United Kingdom since 1926, and in the past 30 years, this approach has been piloted and researched in other European countries and Canada. Heroin maintenance is usually offered as a “last resort” treatment, and there is a broad consensus that it should be offered to only a very selective group of patients with chronic illness who were offered and failed other approaches, including a high-dose methadone maintenance treatment. Recent studies focused on the effectiveness of supervised prescription injectable heroin in participants that failed treatment with methadone and where prescribed heroin was used in combination with oral methadone. In such studies, patients treated with heroin have better treatment retention and reduced use of illicit drugs as compared to patients treated with methadone, but heroin-treated patients had a greater risk for adverse events. At the present time, there is an insufficient number of experimental studies to draw conclusions about the practice of heroin treatment in the community.

Treatment of Special Populations

There is a complex relationship between the criminal justice system, patients with substance abuse disorders, and drug-dependence treatment systems. Treatment of patients involved with the criminal justice system includes efforts directed at persons who are incarcerated, and interventions are designed for persons under legal restraints but still residing in the community. Specialized interventions have been developed for this patient population. Drug courts, which now operate within each state in the United States, combine close criminal justice supervision with long-term drug abuse treatment that typically requires drug testing of parolees.

Illicit opioid use during pregnancy is associated with adverse effects for both the pregnant woman and the fetus though it is not associated with any specific teratogenic effects. There is a higher risk for infections, preeclampsia, miscarriage, premature rupture of membranes, and premature labor, as well as stillbirths, premature infants, and infants with low birth weight and neonatal opioid withdrawal (also called neonatal abstinence syndrome). Infants who survive may be infected with HIV and other diseases as a result of maternal high-risk behavior and are at risk for sudden infant death syndrome (SIDS). Most of these risks can be substantially reduced with methadone or buprenorphine maintenance treatment, with fewer data available studying the impact of treatment with naltrexone.

Health professionals often come to treatment when their drug use is discovered by colleagues or by drug enforcement agencies. Under supervision, most (as much as 85%) of these patients recover and are able to return to practice. Returning to specialties with easy access to opioids can pose a serious hazard to continued abstinence. Treatment with XR naltrexone, which is easy to supervise, may be particularly useful in those patients.

CONCLUSION

OUD is a chronic, relapsing–remitting illness that has grown to become a sustained, global public health challenge. In the past decade, it has affected every demographic and socioeconomic group and caused significant morbidity and mortality through overdose and other medical consequences. There are three approved medication treatments for OUD: methadone, buprenorphine (both oral and XR injection formulations), and XR naltrexone. All three medications have been shown to be effective for reducing relapse and medical consequences. Patients also benefit from psychotherapy that targets behavioral change and motivational enhancement. In those who are engaged with treatment, sustained recovery can be achieved.

FURTHER READINGS

Amass L, Pukeleviciene V, Subata E, et al. A prospective, randomized, multicenter acceptability and safety study of direct buprenorphine/naloxone induction in heroin-dependent individuals. *Addiction*. 2012;107(1):142–151.

American Psychiatric Association. Practice guideline for the treatment of patients with substance use disorders, second edition. *Am J Psychiatry*. 2007;164(Suppl):4.

Bukten A, Skurtveit S, Gossop M, et al. Engagement with opioid maintenance treatment and reductions in crime: a longitudinal national cohort study. *Addiction*. 2012;107(2):393–399.

Cornish R, Macleod J, Strang J, Vickerman P, Hickman M. Risk of death during and after opiate substitution treatment in primary care: prospective observational study in UK General Practice Research Database. *BMJ*. 2010;341:c5475.

Darke S, Mills KL, Ross J, Teesson M. Rates and correlates of mortality amongst heroin users: findings from the Australian Treatment Outcome Study (ATOS), 2001–2009. *Drug Alcohol Depend*. 2011;115(3):190–195.

Ferri M, Davoli M, Perucci CA. Heroin maintenance for chronic heroin-dependent individuals. *Cochrane Database Syst Rev*. 2011;2011(12):CD003410.

Filizola M, Devi LA. Structural biology: how opioid drugs bind to receptors. *Nature*. 2012;485(7398):314–317.

Gelernter J, Kranzler HR, Sherva R, et al. Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. *Biol*

Psychiatry. 2014;76(1):66–74.

Goldstein RZ, Volkow ND. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat Rev Neurosci*. 2011;12(11):652–669.

Gowing L, Farrell MF, Bornemann R, Sullivan LE, Ali R. Oral substitution treatment of injecting opioid users for prevention of HIV infection. *Cochrane Database Syst Rev*. 2011;(8):CD004145.

Grella CE, Karno MP, Warda US, Niv N, Moore AA. Gender and comorbidity among individuals with opioid use disorders in the NESARC study. *Addict Behav*. 2009;34(6-7):498–504.

Havnes I, Bukten A, Gossop M, Waal H, Stangeland P, Clausen T. Reductions in convictions for violent crime during opioid maintenance treatment: a longitudinal national cohort study. *Drug Alcohol Depend*. 2012;124(3):307–310.

Hser YI, Hoffman V, Grella CE, Anglin MD. A 33-year follow-up of narcotics addicts. *Arch Gen Psychiatry*. 2001;58(5):503–508.

Jones HE, Heil SH, Baewert A, et al. Buprenorphine treatment of opioid-dependent pregnant women: a comprehensive review. *Addiction*. 2012;107(Suppl 1):5–27.

Jones HE, Kaltenbach K, Heil SH, et al. Neonatal abstinence syndrome after methadone or buprenorphine exposure. *N Engl J Med*. 2010;363(24):2320–2331.

Kakko J, Svanborg KD, Kreek MJ, Heilig M. 1-year retention and social function after buprenorphine-assisted relapse prevention treatment for heroin dependence in Sweden: a randomised, placebo-controlled trial. *Lancet*. 2003;361(9358):662–668.

Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *Am J Psychiatry*. 2003;160(4):687–695.

Koob GF, Volkow ND. Neurocircuitry of addiction. *Neuropsychopharmacology*. 2010;35(1):217–238.

Kreek MJ, Nielsen DA, Butelman ER, LaForge KS. Genetic influences on impulsivity, risk taking, stress responsivity and vulnerability to drug abuse and addiction. *Nat Neurosci*. 2005;8(11):1450–1457.

Krupitsky E, Nunes EV, Ling W, Illeperuma A, Gastfriend DR, Silverman BL. Injectable extended-release naltrexone for opioid dependence: a double-blind, placebo-controlled, multicentre randomised trial. *Lancet*. 2011;377(9776):1506–1513.

Kunøe N, Lobmaier P, Vederhus JK, et al. Naltrexone implants after in-patient treatment for opioid dependence: randomised controlled trial. *Br J Psychiatry*. 2009;194(6):541–546.

MacArthur GJ, Minozzi S, Martin N, et al. Opiate substitution treatment and HIV transmission in people who inject drugs: systematic review and meta-analysis. *BMJ*. 2012;345:e5945.

Mattick RP, Kimber J, Breen C, Davoli M. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. *Cochrane Database Syst Rev*. 2014;(2):CD002207.

McLellan AT, Arndt IO, Metzger DS, Woody GE, O'Brien CP. The effects of psychosocial services in substance abuse treatment. *JAMA*. 1993;269(15):1953–1959.

McNicholas L. *Clinical Guidelines for the Use of Buprenorphine in the Treatment of Opioid Addiction: A Treatment Improvement Protocol TIP 40*. Substance Abuse and

Mental Health Services Administration; 2004.

SAMHSA. *Results from the 2012 National Survey on Drug Use and Health: Summary of National Findings*. Substance Abuse and Mental Health Services Administration; 2013.

Saxon AJ, Ling W, Hillhouse M, et al. Buprenorphine/naloxone and methadone effects on laboratory indices of liver health: a randomized trial. *Drug Alcohol Depend*. 2013;128(1-2):71–76.

Sigmon SC, Bisaga A, Nunes EV, O'Connor PG, Kosten T, Woody G. Opioid detoxification and naltrexone induction strategies: recommendations for clinical practice. *Am J Drug Alcohol Abuse*. 2012;38(3):187–199.

Strain EC, Bigelow GE, Liebson IA, Stitzer ML. Moderate- vs high-dose methadone in the treatment of opioid dependence: a randomized trial. *JAMA*. 1999;281(11):1000–1005.

Strain EC, Stitzer ML. *The Treatment of Opioid Dependence*. Johns Hopkins University Press; 2006.

Teesson M, Mills K, Ross J, Darke S, Williamson A, Havard A. The impact of treatment on 3 years' outcome for heroin dependence: findings from the Australian Treatment Outcome Study (ATOS). *Addiction*. 2008;103(1):80–88.

UNODC. United Nations Office on Drugs and Crime. World Drug Report 2014. New York, United Nations; 2014.

Weiss RD, Potter JS, Fiellin DA, et al. Adjunctive counseling during brief and extended buprenorphine-naloxone treatment for prescription opioid dependence: a 2-phase randomized controlled trial. *Arch Gen Psychiatry*. 2011;68(12):1238–1246.

WHO. *World Health Organization Guidelines for the Psychosocially Assisted Pharmacological Treatment of Opioid Dependence*. WHO Press; 2009.

Woody GE, Poole SA, Subramaniam G, et al. Extended vs short-term buprenorphine-naloxone for treatment of opioid-addicted youth: a randomized trial. *JAMA*. 2008;300(17):2003–2011.

11.10 Sedative-, Hypnotic-, or Anxiolytic-Related Disorders

JOHN J. MARIANI, M.D.

DEFINITION

This chapter will discuss the epidemiology, neurophysiology, clinical characteristics, and management of sedative-, hypnotic-, and anxiolytic-related disorders. Sedative, hypnotic, and anxiolytic medications constitute a broad group of agents that are classified by their clinical effects, with most medications in this grouping having γ-aminobutyric acid (GABA) agonist or GABAergic transmission facilitation effects. In this chapter, these medications will be collectively referred to as "sedative–hypnotics," with the understanding that within this class of medications, distinctions as to clinical effects are, to some degree, artificial. From both a public health and clinical management perspective, the benzodiazepines are the most important sedative–hypnotic agents because of their common use and wide availability, both by prescription and from nonprescription sources. Benzodiazepines and other sedative–hypnotic agents are widely prescribed because of their efficacy in the treatment of anxiety, insomnia, and other therapeutic indications. Like other prescribed controlled substances, most individuals will be able to use sedative–hypnotic agents without developing a pattern of problematic or compulsive use. However, for a minority of individuals exposed to sedative–hypnotic agents, a pattern of use meeting criteria for a substance use disorder may develop. These individuals need specialized assessment and care to avoid the potential behavioral and medical consequences of sedative–hypnotic use disorder.

Studies of sedative, hypnotic, or anxiolytic use disorders are limited by the operational definition of substance use disorders set forth in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (5th ed. Text rev.; DSM-5 TR). Sedative–hypnotic agents, and related anxiolytic medications, have the potential to induce tolerance, and with regular use, individuals may be at risk of a discontinuation syndrome or frank withdrawal state. However, drug addiction is a sociobehavioral phenomenon, albeit with significant pharmacologic origins and medical consequences. There is a tendency to conflate the neuroadaptive changes that develop in response to chronic sedative–hypnotic administration (physiologic dependence) with the concept of a substance use disorder or addiction. Many individuals who have difficulty tapering or discontinuing prescribed sedative–hypnotic agents because of neuroadaptive changes may be considered "addicted" by themselves or clinicians but do not meet full criteria for a substance use disorder unless other symptoms are present.

The addictive potential of a substance is usually inferred from data on illicit trafficking; population surveys of nonprescription use; audits of prescribing practices; postmarketing studies; monitoring of diversion from physician's offices, pharmacies, or manufacturing sites; emergency room treatment; and overdose deaths. Pharmacologic characteristics also predict misuse liability: Drug-induced euphoria and the development of tolerance, an abstinence syndrome, or cross-tolerance with known addictive substances may all predict the risk of misuse or addictive potential. Finally, a number of laboratory methods for determining drug preferences in human and animal models have been developed to determine misuse liability. Sedative–hypnotic medications that have general GABAergic effects tend to have the potential for misuse and addiction. In general, because of their risk of misuse and addiction, sedative–hypnotic medications are classified as controlled substances. This risk is related to, but distinct from, the tendency for GABAergic medications to induce pharmacologic tolerance and be predisposed to developing withdrawal. An individual who developed tolerance and withdrawal while prescribed a sedative–hypnotic agent in the context of legitimate medical treatment, and displayed no other symptoms, would not meet criteria for a substance use disorder.

Various terms are used to refer to patterns for the use of controlled substances. *Misuse* refers to the use of a drug in a manner that is not consistent with generally accepted medical practice or social and legal custom (e.g., use without a valid prescription or use to deliberately produce intoxication, pleasure, or a high). *Recreational use* refers to the use of drugs solely for their hedonic value, but this term lacks specificity and has an inappropriate harmless connotation. Mild substance use disorder in DSM-5 TR requires two to three symptoms from a list of 11. The WHO 11th revision of the *International Classification of Diseases and Health Problems* (ICD-11) retains the term “substance dependence” to indicate “addiction.” The term *physiologic dependence* is applied more narrowly in the psychopharmacologic context, “referring solely to the development of withdrawal symptoms on cessation of drug use.”

The *abstinence or discontinuation syndrome* that develops with the discontinuation of sedative–hypnotic agents has specific characteristics referable to the autonomic nervous system (e.g., tremor, sweating, tachycardia, and startle response) and must be clearly distinguished from recurrence of the features of the underlying disease for which the drug was originally prescribed. It can be reversed or attenuated by readministering the discontinued drug or administering a drug with cross-tolerance. A *withdrawal syndrome* is associated with many sedative–hypnotic agents, which in some cases can be associated with severe medical sequelae. In this chapter, abstinence syndrome will refer to subjective symptoms that cannot be objectively measured (e.g., anxiety, insomnia, and craving), and withdrawal syndrome will refer to a more severe neurophysiologic disturbance where symptoms can be objectively measured (e.g., elevated pulse or blood pressure) or observed (e.g., tremor).

Substances

The drugs discussed in this section are referred to as *anxiolytic* or *sedative–hypnotic drugs*. The terminology for this group of drugs is not clearly established. Their sedative or calming effects are on a continuum with their hypnotic or sleep-inducing effects. Anxiolytic drugs cover a wide spectrum of pharmacologic agents, and the terminology implies more specificity than actually exists. The broad group of drugs that historically has been included in that class exhibits considerable variation in clinical use, toxicity, risk for addiction, and potential for diversion and misuse. Sedative–hypnotic agents are classified in this section as follows: (1) benzodiazepines and functionally related drugs, (2) barbiturates, and (3) miscellaneous sedative–hypnotic drugs. In the practice of psychiatry, the drugs that are most important clinically are the benzodiazepines.

ETIOLOGY AND NEUROPHARMACOLOGY

Although most of the drugs discussed in this chapter have their primary effects on the γ-aminobutyric acid type A receptor ($GABA_AR$) complex, there are subtle differences in their mechanisms of action at this site. Furthermore, neuroadaptive changes that occur with prolonged drug exposure affect not only the $GABA_AR$ but other neurotransmitters, the hypothalamic–pituitary axis, and calcium-modulated systems. GABA is the major inhibitory neurotransmitter in the brain. While historically understudied, there is developing evidence that long-range GABAergic input from the ventral tegmental area (VTA) to the nucleus accumbens is involved in reward and reinforcement behavior.

$GABA_ARs$ are a family of ligand-gated ion channels responsible for inhibitory regulation of the CNS. Benzodiazepines bind directly to a specific modulatory site on $GABA_ARs$, which enhances the effect of the inhibitory neurotransmitter but does not open the chloride channel in the absence of GABA. Chronic exposure to benzodiazepines leads to alterations of GABAergic neurotransmission, which are manifested as the symptoms of tolerance, dependence, and withdrawal. The mechanism of these changes is unclear and may involve alterations in the expression of individual $GABA_AR$ subtypes, but it is known to be linked to the dose and duration of benzodiazepine use. While the direct effects of benzodiazepines are GABA related, the withdrawal syndrome may, in part, be mediated by calcium channel and glutamatergic mechanisms. In addition, neuroendocrine pathways, including regulation of the hypothalamic–pituitary–adrenocortical axis by CRF, may be involved in the mechanism of benzodiazepine dependence.

Activation of $GABA_ARs$ by agonists or positive modulators leads to an influx of chloride ions into neurons, which then produce hyperpolarization, thereby decreasing the influence of excitatory input. $GABA_ARs$ exist as protein structures that consist of five subunits. In the brain, these structures commonly contain two α-, two β-, and one γ-subunit. Each subunit constitutes a unique sequence of amino acids, which are differentiated into numbered subtypes based on minor variations in the amino acid sequence. For example, the α-subunit is thought to have six subtypes, designated α1 to 6. Receptors containing α1, α2, α3, and α5 subunits are sensitive to the effects of benzodiazepines. In contrast, the presence of α4 or α6 subunits leads to a lack of sensitivity to benzodiazepines. Benzodiazepines bind at the interface of the γ2 and α-subunits.

Agonists that interact with the benzodiazepine receptor produce alterations in the structure of the $GABA_A$Rs, which can enhance the affinity of these receptors for GABA and direct receptor agonists, and are often referred to as positive modulators to differentiate them from direct agonists, such as GABA and muscimol. The benzodiazepine-induced allosteric changes in the $GABA_A$R complex result in an increase in the frequency of opening of the chloride channel located in the center of the $GABA_A$R complex.

Barbiturates share some similarities in their actions with those of benzodiazepines. At low concentrations, barbiturates act at a receptor site other than the GABA binding site on the $GABA_A$R complex to positively modulate the actions of GABA via an allosteric mechanism. However, at higher concentrations, barbiturates act as $GABA_A$R agonists, directly increasing the activity of these receptors. $GABA_A$Rs that contain the β2 or β3 subunit show greater sensitivity to the effects of pentobarbital than receptors with the β1 subunit. The type of α-subunit may be a factor in determining the extent to which a barbiturate may act as a direct GABA agonist. Pentobarbital shows the greatest efficacy as a GABA receptor agonist in receptor complexes that contain the α6 subunit, which is a benzodiazepine-insensitive subunit.

Barbiturates also may have actions on neurons that involve mechanisms other than stimulation of $GABA_A$R activity. At least in the hippocampus, pentobarbital may inhibit the activity of Dl-α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptors. These receptors are involved in the mediation of some of the effects of the excitatory neurotransmitter glutamate. Barbiturates also may inhibit the release of neurotransmitters by blocking voltage-sensitive calcium channels.

Diazepam and other classical benzodiazepines appear to bind with equal affinity to benzodiazepine-sensitive receptors with different subunit composition. In contrast, the so-called nonbenzodiazepine hypnotics, such as the imidazopyridine, zolpidem, the pyrazolopyrimidine, zaleplon, and the cyclopyrrolone, eszopiclone, act more selectively on $GABA_A$Rs that contain the α1 subunit (often referred to as $GABA_{A1}$ receptors). These differences in selectivity may be of significance because the nature of the subunits may be related to the specific clinical effects of the agonist. $GABA_A$Rs that contain α1 subunits appear to mediate the sedative–hypnotic effects of benzodiazepines. This subunit has also been implicated in playing a key role in the production of the ataxic effects of both benzodiazepines and zolpidem. The antianxiety effects of benzodiazepine may be mediated by receptors containing the α2 and possibly the α3 subunit. The α2 and α3 subunits also may be involved in the mediation of the muscle relaxant effects of benzodiazepines.

Tolerance develops to many of the actions of benzodiazepines. Tolerance first develops to the sedative effects of these drugs. This may then be followed by tolerance to benzodiazepine-induced amnesia and to the anticonvulsant effects of these drugs. In addition to tolerance, benzodiazepines may produce a withdrawal syndrome upon abrupt discontinuation.

The development of tolerance to the effects of benzodiazepines may involve mechanisms that include the uncoupling of the allosteric interaction that occurs between benzodiazepines and GABA binding sites and the internalization into neurons of the $GABA_A$R complex. Tolerance to benzodiazepines also may involve alterations in the subunit composition of $GABA_A$Rs. Chronic exposure to classical benzodiazepine compounds such as diazepam may lead to an alteration in the gene expression for the mRNAs that are responsible for the synthesis of specific $GABA_A$ subunits. These changes may include a decrease in the mRNAs for cortical and hippocampal α1 subunits. Within the hippocampus, binding to hippocampal α5-GABA receptors is decreased in animals that are tolerant of the sedative effects of diazepam. The tolerance to the motor activity depressant effects of diazepam was not observed in animals missing the gene for the α5 subunit.

Chronic exposure to the α1 selective drugs zolpidem and zaleplon results in less extensive changes in the expression of $GABA_A$R units than is observed following prolonged treatment with classical benzodiazepines. The repeated administration of either zolpidem or zaleplon may not decrease mRNAs encoding for the α1 and other $GABA_A$R subunits. Following short-term withdrawal from these agents, however, an elevation in α4 mRNA may occur, and changes in the expression of mRNA for other $GABA_A$R subunits are similar to that seen with withdrawal from diazepam, including a decrease in the mRNA for the α1 subunit.

The primary mechanisms underlying withdrawal symptoms following the discontinuation of chronic benzodiazepine treatment are related to changes in the GABA system: $GABA_A$ subunit expression, $GABA_AR$ downregulation, or uncoupling of benzodiazepine receptors. Changes in the glutamatergic systems have also been shown to be involved in withdrawal. In several brain regions, including the frontal cortex and hippocampus, withdrawal symptoms occur in association with increased expression of the AMPA receptor GluR1 subunit. Electrophysiologic evidence also indicates that the activity of AMPA receptors is increased during withdrawal from benzodiazepines. In addition to changes in AMPA receptor function, chronic exposure to benzodiazepines may result in upregulation of L-type calcium channels. Such upregulation occurs in association with physical dependence on a variety of abused substances, including ethanol and nicotine.

Tolerance to barbiturates may involve both an increase in the rate of metabolism of these compounds and several neuroadaptive responses. Barbiturates such as pentobarbital and phenobarbital can increase the activity of cytochrome P450 enzymes, including the CYP2B and CYP3A enzymes. This effect can lead to enhanced barbiturate metabolism that, in turn, results in a reduction in barbiturate plasma concentrations and pharmacokinetic tolerance. The prolonged administration of barbiturates may produce some alterations in the expression of different $GABA_AR$ subunits. The expression of the α6 $GABA_AR$ subunit, for example, is increased in pentobarbital-tolerant animals. Chronic phenobarbital administration may result in an increase in NMDA receptors in the cortex during withdrawal. This upregulation of NMDA receptors may explain the high level of CNS excitability, which is a feature of withdrawal from barbiturates.

MISUSE LIABILITY

Surveys of prescribing practices, patient-initiated dosage changes, and recreational or nonmedical benzodiazepine use, along with reports from emergency rooms, medical examiners, and law enforcement agencies, provide a complex picture of the public health implications of the benzodiazepines. The vast majority of medical and psychiatric patients use benzodiazepines appropriately, although rates of misuse by patients dependent on alcohol or other drugs may be higher than by individuals with anxiety or insomnia without a history of substance use problems. Furthermore, long-term use is high among patients with psychiatric disorders and the elderly, with the latter group being sensitive to drug toxicity. Past-year benzodiazepine or opioid misuse is associated with suicidal ideation in older adults; 25.4% of individuals with both past-year benzodiazepine and opioid misuse report suicidal ideation.

One approach to assessing misuse liability is by measuring the reinforcing properties of the drug. If its pharmacologic effects increase behavior (i.e., self-administration of a dose or the work required to permit self-administration), then the drug is a positive reinforcer and has misuse liability. Animal and human studies have assessed the reinforcing properties of benzodiazepines.

$GABA_A$ Receptor Subtypes and Addiction Risk

Benzodiazepines are positive allosteric modulators of the $GABA_A$Rs. Recent work on the neural basis of the addictive properties of benzodiazepines suggests that receptors that contain the $GABA_A$R α1 subunit are important for triggering drug-evoked synaptic plasticity at VTA synapses, which is an important first step underlying drug reinforcement. Benzodiazepines have a stronger direct effect on GABA neurons than on dopamine neurons. GABA neurons are more strongly hyperpolarized in the presence of benzodiazepines and no longer inhibit dopamine neurons, resulting in an increase in dopamine transmission. This disinhibition of VTA dopamine release may be a mechanism for the development of benzodiazepine addiction, as well as potential benzodiazepine facilitation of addiction to other substances.

Animal Studies

Animal studies have consistently demonstrated that benzodiazepines have minimal reinforcing effects. Although a small number of studies show that animals self-administer benzodiazepines more frequently than other drugs, they are significantly less potent than other drugs of abuse (e.g., cocaine and other CNS stimulants) or even many other sedative–hypnotic drugs, including most barbiturates. Studies that use the conditioned place preference paradigm have also demonstrated that benzodiazepines have minimal reinforcing effects. In some studies, diazepam has produced place preference among laboratory rats, but to a lesser extent than classic drugs of abuse, such as morphine or amphetamine.

Animal studies using schedule-induced polydipsia (the use of intermittent food administration to the experimental animal, which promotes ingestion of fluids, including drug solutions) indicate that midazolam is ingested more often than water, but self-administration of chlordiazepoxide and flurazepam does not differ significantly from that of water.

Another animal model predicting misuse liability consists of drug infusion in response to activity by the animal, such as lever pressing (Fig. 11.10–1). Drug infusion may allow continuous or intermittent drug availability, and rates of administration of the drug are compared to those of the vehicle, other sedative–hypnotic drugs, and standard drugs of abuse, such as cocaine. Studies with continuous drug access suggest that triazolam, diazepam, midazolam, clobazam, and chlordiazepoxide have reinforcing effects; studies with intermittent drug availability show that diazepam, midazolam, triazolam, alprazolam, lorazepam, chlordiazepoxide, and bromazepam maintain responses at a higher level than vehicle, indicating abuse liability. With the possible exception of triazolam and midazolam, in studies using this model, the benzodiazepines consistently demonstrate lower misuse liability than barbiturates other than phenobarbital, which also has relatively low misuse liability. The nonbenzodiazepine hypnotics that act on the $GABA_1$ receptor, such as zolpidem, zaleplon, and eszopiclone, are similar to the benzodiazepines in animal models of misuse. Partial benzodiazepine receptor agonists, such as abecarnil and bretazenil, have low misuse liability in these models. Prior exposure to ethanol or sedative–hypnotic drugs may increase the reinforcing effects of benzodiazepines.

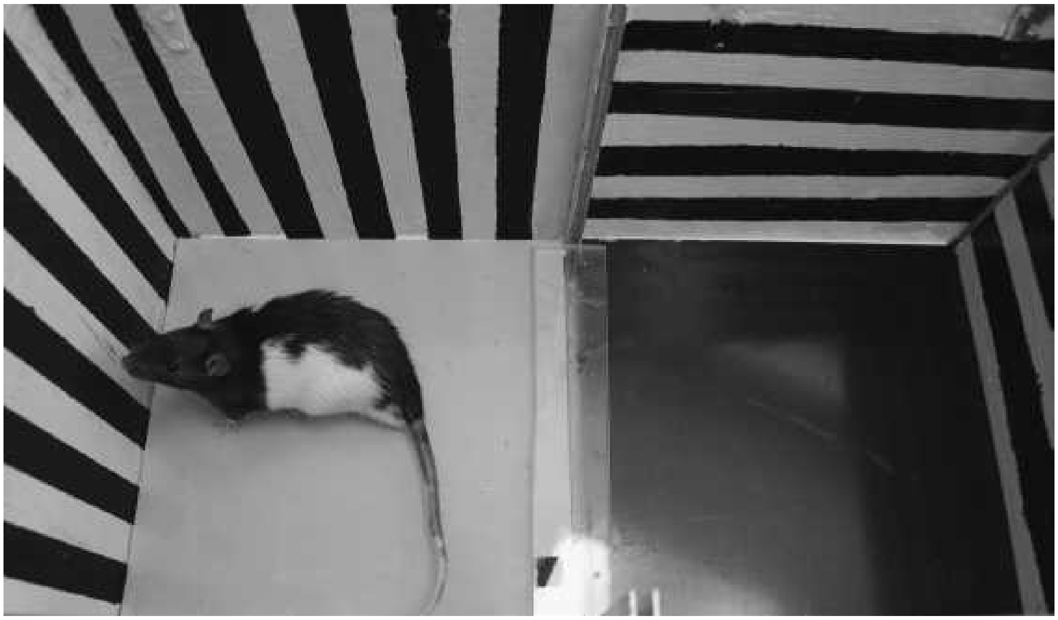

FIGURE 11.10–1. In this conditioned place preference experiment, the animal spends more time in the chamber compartment in which a reinforcing drug had been repeatedly administered (*the white floor*) than in compartment he associates with saline administration (*the black floor*). (Courtesy of Conan Kornetsky, Ph.D.)

Animal studies have demonstrated that a withdrawal syndrome occurs on abrupt discontinuation of benzodiazepines and related sedative–hypnotics. The animal models using the social interaction test and social conflict paradigms, along with isolation-induced ultrasonic vocalization recordings, indicate that abrupt discontinuation of benzodiazepine tranquilizers may induce numerous anxiogenic responses, including decreased social interaction and increased aggressive behaviors. Studies that use a competitive benzodiazepine receptor antagonist, such as flumazenil, to precipitate withdrawal have also demonstrated the development of a withdrawal syndrome using electrophysiologic measures.

Human Studies

The assessment of misuse liability in human laboratory studies relies on variations in two predictive models: self-administration of benzodiazepines by experimental subjects with a history of substance use problems and subjective responses that correlate with high misuse potential. In the first paradigm, one type of study allows subjects with a history of sedative–hypnotic use disorder to self-administer orally (i.e., sample) different color-coded drugs and placebo under double-blind conditions in a simulated social environment. During subsequent sessions, they are permitted to self-administer the drug that they like the best. Some studies require that the subjects perform a task (e.g., riding a stationary bicycle) to earn doses. The misuse liability of different drugs is assessed by how often they are chosen for self-administration or by how much work a subject does to earn a dose. Scales measuring drug effect are also administered. The paradigm has the advantage of directly observing the behavior involved with ingesting the drug, albeit under somewhat artificial circumstances. It works best for individuals with a history of a sedative–hypnotic use disorder in remission without a co-occurring psychiatric disorder. It is probably safe to infer that those subjects are administering the drug for its hedonic value. The reason for self-administration by other subjects may not be clear. Do patients with anxiety disorder self-administer the drugs because they are effective anxiolytics or because they induce euphoria? Similarly, if subjects are in withdrawal from sedative–hypnotic drugs or ethanol, do they choose the agent that alleviates their withdrawal symptoms most effectively or the drug that produces the greatest euphoria? An additional problem with this design is that it exposes subjects to multiple doses of potentially addictive drugs.

An alternative strategy is to administer only a single dose of the test drugs and then to compare subjective responses associated with misuse liability. A set of scales, which differs slightly among laboratories, usually consists of the Morphine Benzedrine Group subscale of the Addiction Research Center Inventory (ARCI-MBG); self-reports of sedation, liking, and intensity of drug effect; monetary street value of the drug; and similarity to reference drugs of abuse. Studies demonstrate that estimates of misuse liability using this paradigm are consistent with assessments based on the self-administration model. One disadvantage of the single-dose paradigm is that it measures only subjective effects, not behavior. It also has some of the same limitations as the self-administration paradigm. Does the subject like the drug because it has a therapeutic effect or because it is intoxicating? Both designs attempt to circumvent the problem by including an adequate battery of scales that measure several different dimensions of drug effect. Dose equivalency and homogeneity of the subjects (especially with respect to a personal or family history of psychoactive substance abuse and psychiatric illness) are critical factors in the study design and in the interpretation of data from both models.

The predictive models in humans clearly establish that benzodiazepines occupy an intermediate position on the spectrum of misuse liability of sedative–hypnotic drugs. Methaqualone and barbiturates generally produce higher ARCI-MBG and drug-liking scores than do benzodiazepines. Buspirone, a nonbenzodiazepine anxiolytic agent (i.e., it is essentially devoid of benzodiazepine agonist activities), has little or no misuse liability in those models. With respect to differences among the benzodiazepines, the weight of the evidence suggests that diazepam has greater misuse liability than halazepam, oxazepam, chlordiazepoxide, or clorazepate. Diazepam, lorazepam, alprazolam, and triazolam exhibit similar profiles of misuse potential using those paradigms. The clinical relevance of the findings has not been adequately studied.

Flunitrazepam, a benzodiazepine agonist, is not marketed in the United States but has entered the country through illegal sources in Central and South America. Reports of misuse and resulting health consequences have been reported. The drug has been implicated in sexual assaults, in which flunitrazepam is surreptitiously added to an alcoholic beverage consumed by the victim to reduce judgment, inhibition, and physical ability to resist sexual activity. Memory for the event may also be impaired by the combination of alcohol and flunitrazepam. The weight of the evidence supports the safety and efficacy of flunitrazepam in clinical populations; however, individuals with alcohol or drug dependence commonly misuse it in combination with alcohol, illicit drugs, or both.

Pharmacologic characteristics, especially pharmacokinetics, of benzodiazepines and related compounds influence misuse liability. Some drugs (e.g., diazepam) may appear to have high misuse liability because of rapid absorption from the GI tract after oral doses. However, pharmacodynamic differences among the drugs are not well studied. For example, little is known about differences in misuse liability between full and partial agonists or among drugs that have preferential effects on specific subtypes of the benzodiazepine receptor. Zolpidem, for example, is an imidazopyridine hypnotic agent with a rapid onset of action and a short elimination half-life. When the drug was first marketed in the 1980s, the manufacturer presented data indicating a low misuse liability for zolpidem, but subsequent investigations show otherwise. Although in vitro evidence indicates that zolpidem binds preferentially to the benzodiazepine receptor with a high-affinity ratio for the α1 subunit, it is unlikely that this characteristic would provide lower potential for misuse;

rather, selective binding more likely explains the lack of anticonvulsant and muscle relaxant effects of the drug. Also supporting the similarity between zolpidem and classic benzodiazepines is the fact that zolpidem and triazolam depress energy metabolism in the same areas of the brain. With respect to subjective, psychomotor, and memory effects, there are no significant differences between a single 10-mg dose of zolpidem and a single 0.25-mg dose of triazolam. High doses of triazolam (0.75 mg) are identified more often as similar in effect to barbiturates, benzodiazepine, or alcohol than are high doses of zolpidem (45 mg). In the United States, New Zealand, and some European countries, zolpidem and zopiclone (another nonbenzodiazepine hypnotic) have been shown to have reduced efficacy after a few weeks of treatment, requiring increases in doses, which may contribute to a higher risk of misuse and physiologic dependence. In some cases, zopiclone dosages have reached levels as high as 82.5 mg per day after an initial prescribed dosage of 3.75 to 15.0 mg per day. Dosages of zolpidem have reached 800 mg per day after an initial prescribed dose of 5 to 20 mg per day. Abrupt discontinuation of zolpidem or zopiclone after months of use has resulted in severe abstinence syndromes, which may include seizures. Eszopiclone has been studied in 6-month trials of insomnia and did not show tolerance but was associated with a mild withdrawal syndrome, consisting of rebound insomnia and anxiety in a very small percentage (2% or less) of patients. Eszopiclone at doses of 6 and 12 mg produced euphoric effects comparable to 20 mg of diazepam in previous sedative–hypnotic misusers. Studies examining the misuse liability of 25, 50, and 75 mg of zaleplon in subjects with a history of sedative–hypnotic use disorder indicated misuse potential similar to benzodiazepines and benzodiazepine-like hypnotics. A withdrawal syndrome similar to benzodiazepines, including seizures, should be expected in patients who have had long-term, high-dose treatment with zaleplon. In general, clinical experience with zolpidem, zopiclone, eszopiclone, and zaleplon (nonbenzodiazepine hypnotics or z-drugs) has been favorable; however, they produce subjective effects associated with misuse in experienced drug users, and caution should be exercised when they are prescribed in addiction-prone individuals. Although the nonbenzodiazepine hypnotics (z-drugs) are similar to benzodiazepines in human misuse liability measures, the actual rates of misuse may be lower, according to some authorities.

Clinical experience and research findings have identified patient characteristics that influence misuse liability. For example, when individuals with a history of alcohol use disorder and their adult children are given a single dose of alprazolam or diazepam, they have greater increases on misuse potential scales than do nonalcoholic controls without a family history of alcoholism. The differences in reinforcing effects are not attributable to pharmacokinetic differences as measured by peripheral plasma concentrations; however, pharmacodynamic differences have been noted using functional MRI and EEG measures. Studies have found that patients with prior alcohol or substance use disorders exhibit a clear preference for benzodiazepines in self-report models. Enhanced mood states after benzodiazepine challenges have also been found in moderate drinkers, although the confound of a family history of substance use problems has not been adequately controlled in all studies. Other findings indicate that moderate drinkers self-administer higher doses of diazepam than do light drinkers. Whether these results are drug-specific or apply to the class of benzodiazepines as a whole is not known. Patients with a family history of substance abuse have reported a benzodiazepine preference in some studies. A survey of 5,000 male college students found that 23% of those with a positive family history of alcohol use problems reported using benzodiazepines (vs. 0.9% of those without such a family history).

Anxiety disorder patients with a history of alcohol or other substance use disorders have been commonly thought to be at a higher risk of misusing benzodiazepines. A study conducted by the Harvard/Brown Anxiety Disorders Research Program (HARP) examined total daily dose, as-needed (PRN) use, and continued use of benzodiazepine in patients with and without a history of alcohol dependence. That study found that prior alcohol use disorder was not a strong predictor for benzodiazepine misuse.

Experimental paradigms do not provide evidence that individuals with anxiety disorders are at greater risk of benzodiazepine misuse when compared to nonanxious subjects. The fact that some anxious subjects choose diazepam over placebo in self-administration experiments underscores the limitations of the study design rather than indicating misuse potential. In such studies, it is impossible to determine if subjects choose the drug to relieve psychic distress or to experience its pleasurable effects. In clinical experience, the risk of benzodiazepine misuse in patients with anxiety disorders is low.

EPIDEMIOLOGY

The lifetime risk of sedative–hypnotic use disorders is approximately 1%. Benzodiazepine treatment admissions are more likely than other treatment admissions to receive detoxification and inpatient treatment. Risk factors for benzodiazepine use disorder include participation in substance use disorder self-help groups, younger age, longer duration of benzodiazepine use, higher dose of benzodiazepine use, lower level of education, nonnative cultural origin, and outpatient treatment of substance use disorders.

According to the U.S. SAMHSA's 2020 National Survey on Drug Use and Health (NSDUH), among people aged 12 years or older, 2.2% misused tranquilizers or sedatives in the past year, with the percentage highest among young adults aged 18 to 25 years (3.7%). Among people aged 12 years or older in 2020, 1.7% misused prescription benzodiazepines in the past year, with the rate highest among young adults aged 18 to 25 years (3.3%). The 2020 NSDUH reported that in 2020 approximately 0.4% of people aged 12 years or older (1.2 million individuals) in the United States met criteria for a sedative–hypnotic use disorder, with the highest rate among young adults aged 18 to 25 years (0.7% or 235,000 people).

The 2021 Drug Abuse Warning Network (DAWN) reports that alprazolam (a benzodiazepine) was the 10th most common cause for drug-related emergency department visits. The 2020 Treatment Episode Data Set (TEDS) reports that approximately 1% of admissions for substance use disorder treatment for individuals 12 years and older were primarily for benzodiazepines, which has been consistent at that rate since 2010. Also, according to the 2020 TEDS, alprazolam was the most commonly cited sedative–hypnotic as a reason (primary, secondary, or tertiary) for treatment admission, with an overall rate of 1.8%, and 1% citing as the primary reason for admission. There are disparities in benzodiazepine misuse among sexual minorities, with bisexual females demonstrating higher odds of risky use of concurrent opioid and benzodiazepine use than other groups.

According to the U.S. Centers for Disease Control and Prevention, benzodiazepine overdose deaths increased from 2017 to 2020. In 2020, 92.7% of benzodiazepine-involved deaths also involved opioids, and 66.7% involved fentanyl analogs.

DIAGNOSIS AND CLINICAL FEATURES

Diagnosis

DSM-5 TR lists a number of sedative-, hypnotic-, or anxiolytic-related disorders but contains specific diagnostic criteria only for sedative, hypnotic, or anxiolytic intoxication and sedative, hypnotic, or anxiolytic withdrawal. Other sedative-, hypnotic-, or anxiolytic-related disorders have their diagnostic criteria outlined in the DSM-5 TR sections that are specific for the major symptom (e.g., sedative-, hypnotic-, or anxiolytic-induced psychotic disorder).

Substance Use Disorder. Sedative, hypnotic, or anxiolytic use disorders are diagnosed according to the general criteria in DSM-5 TR for substance use disorders.

Intoxication. DSM-5 TR contains a single set of diagnostic criteria for intoxication by any sedative, hypnotic, or anxiolytic substance. Although the intoxication syndromes induced by all those drugs are similar, subtle clinical differences are observable, especially with intoxications that involve low doses. The diagnosis of intoxication by one of that class of substances is best confirmed by obtaining a blood sample for substance screening.

Withdrawal. DSM-5 TR contains a single set of diagnostic criteria for withdrawal from any sedative, hypnotic, or anxiolytic substance. The clinician can specify *with perceptual disturbances* if illusions, altered perceptions, or hallucinations are present but are accompanied by intact reality testing. Benzodiazepines are associated with a withdrawal syndrome, and withdrawal from barbiturates can be life-threatening. Withdrawal from benzodiazepines can also result in serious medical complications, such as seizures.

Delirium. DSM-5 TR allows the diagnosis of sedative, hypnotic, or anxiolytic intoxication delirium and sedative, hypnotic, or anxiolytic withdrawal delirium. Delirium that is indistinguishable from delirium tremens associated with alcohol withdrawal is more commonly seen with barbiturate withdrawal than with benzodiazepine withdrawal. Delirium associated with intoxication can be seen with barbiturates or benzodiazepines if the dosages are high enough.

Neurocognitive Disorders. DSM-5 TR allows diagnosis of sedative-, hypnotic-, or anxiolytic-induced neurocognitive disorders. The diagnosis of the disorder is complex because it is difficult to know whether a neurocognitive disorder is due to the substance use itself or to the associated features of the substance use.

Psychotic Disorders. The psychotic symptoms of barbiturate withdrawal can be indistinguishable from those of alcohol withdrawal delirium. Agitation, delusions, and hallucinations are usually visual, but sometimes tactile or auditory features develop after approximately 1 week of abstinence. Psychotic symptoms associated with intoxication or withdrawal are much more common with barbiturates than with benzodiazepines and are diagnosed as sedative-, hypnotic-, or anxiolytic-induced psychotic disorders. The clinician can further specify whether delusions or hallucinations are the predominant symptoms.

Other Disorders. Sedative, hypnotic, and anxiolytic use has also been associated with mood disorders, anxiety disorders, sleep disorders, and sexual dysfunctions. When none of those diagnostic categories is appropriate for a person with a sedative-, hypnotic-, or anxiolytic-use disorder, the appropriate diagnosis is sedative-, hypnotic-, or anxiolytic-related disorder not otherwise specified (NOS).

Clinical Features and Patterns of Misuse

Sedative, hypnotic, and anxiolytic drugs can nearly all be taken orally, occasionally to achieve a time-limited specific effect or regularly to obtain a constant, usually mild, intoxication. The physical dangers of injection include transmission of HIV, cellulitis, vascular complications from accidental injection into an artery, infections, and allergic reactions to contaminants. IV use is associated with rapid and profound tolerance, dependence, and a severe withdrawal syndrome. One survey reported that 28% of individuals addicted to heroin injected benzodiazepines, and these patients had higher rates of polydrug use, HIV risk-taking behavior, criminal activity, psychological distress, health problems, and overdose than those who did not inject benzodiazepines.

Benzodiazepines. Since the introduction of chlordiazepoxide in 1960, benzodiazepines have become the primary drugs used to treat anxiety and insomnia, largely replacing barbiturates and other sedative–hypnotic agents. Zolpidem, an imidazopyridine hypnotic drug, is chemically distinct from the benzodiazepines but has similar clinical effects and acts at the $GABA_{A1}$ benzodiazepine receptor complex. At present, it has the largest market share of hypnotic prescriptions in the United States; other hypnotics acting at the $GABA_{A1}$ receptor, such as zaleplon and eszopiclone, also have effects very much like the classical benzodiazepines. The benzodiazepines have lower misuse liability than most barbiturates, pose a much lower risk when taken in overdose, and have fewer interactions with other drugs. The advantages of the benzodiazepines must be weighed against the risk of misuse and physiologic dependence.

INTOXICATION. Benzodiazepine intoxication can be associated with behavioral disinhibition, potentially resulting in hostile or aggressive behavior. The effect is perhaps most common when benzodiazepines are taken in combination with alcohol. Benzodiazepine intoxication is associated with less respiratory depression than barbiturate intoxication.

WITHDRAWAL SYNDROME

Signs and Symptoms. The signs and symptoms of the benzodiazepine discontinuation syndrome (Table 11.10–1) have been classified as major or minor, like those of the alcohol withdrawal syndrome. According to that classification, minor symptoms include anxiety, insomnia, and nightmares. Major symptoms (which are extremely rare) include grand mal seizures, psychosis, hyperpyrexia, and death.

The discontinuance syndrome may also be divided into symptoms of rebound, recurrence, and withdrawal. *Rebound symptoms* are those for which the benzodiazepine was originally prescribed that return in a more severe form than they had before treatment. They have a rapid onset after termination of therapy and a brief duration. *Recurrence* refers to the return of the original symptoms at or below their original intensity. The pattern and course of these symptoms reflect the anxiety disorder for which treatment was originally instituted.

Discontinuation symptoms are described with similar terms for benzodiazepines and SSRIs. Withdrawal reactions to SSRIs appear to be similar to those for benzodiazepines; referring to these reactions as part of a dependence syndrome in the case of benzodiazepines, but not SSRIs, may not be rational.

Withdrawal symptoms (see Table 11.10–1) are loosely categorized into four types: (1) disturbances of mood and cognition, (2) disturbances of sleep, (3) physical signs and symptoms, and (4) perceptual disturbances. Mood and cognitive symptoms are anxiety, apprehension, dysphoria, irritability, obsessive ruminations, and paranoia. Sleep disturbances include insomnia, altered sleep–wake cycle, and daytime drowsiness. Somatic symptoms are agitation, tachycardia, palpitations, motor restlessness, muscle tension, tremor myoclonus, nausea, coryza, diaphoresis, lethargy, muscle and joint pain, hyperreflexia, ataxia, tinnitus, and seizures. Perceptual disturbances include hyperacusis, depersonalization, blurred vision, illusions, and hallucinations.

Table 11.10–1.
Signs and Symptoms of the Benzodiazepine Discontinuation Syndrome

The following signs and symptoms may be seen when benzodiazepine therapy is discontinued; they reflect the return of the original anxiety symptoms (recurrence), worsening of the original anxiety symptoms (rebound), or emergence of new symptoms (true withdrawal):

- *Disturbances of mood and cognition*
  Anxiety, apprehension, dysphoria, pessimism, irritability, obsessive rumination, and paranoid ideation
- *Disturbances of sleep*
  Insomnia, altered sleep–wake cycle, and daytime drowsiness
- *Physical signs and symptoms*
  Tachycardia, elevated blood pressure, hyperreflexia, muscle tension, agitation/motor restlessness, tremor, myoclonus, muscle and joint pain, nausea, coryza, diaphoresis, ataxia, tinnitus, and grand mal seizures
- *Perceptual disturbances*
  Hyperacusis, depersonalization, blurred vision, illusions, and hallucinations

The temporal sequence of symptom development is not well established, but on the abrupt cessation of benzodiazepines with short elimination half-lives, symptoms may appear within 24 hours and peak at 48 hours. Symptoms arising from abrupt discontinuance of benzodiazepines with long half-lives may not peak until 2 weeks later. Although some investigators suggest that a subgroup of patients have withdrawal syndromes that last for many months, no medical or scientific evidence validates the existence of such a syndrome. Prolonged symptoms are almost certainly attributable to the recurrence of the original anxiety or progression of the anxiety disorder itself.

Risk Factors. Factors influencing the development of the discontinuance or withdrawal syndrome are listed in Table 11.10–2. Studies on humans and animals have attempted to identify risk factors for the development of physiologic dependence. Animal studies demonstrate that the severity of withdrawal is greater with higher doses and longer periods of drug administration, but that has not been as consistently demonstrated in clinical populations. When patients are stratified into high- and low-dose groups, the withdrawal syndrome is seen to be more severe in the high-dose group. However, two recent studies, one using a daily 21-mg dose of diazepam equivalent as the cutoff separating high- and low-dose groups and the other using three groups based on diazepam daily dose (less than 6 mg, between 6 and 10 mg, and more than 10 mg), indicate no significant differences in withdrawal symptoms between groups. One possible explanation for the failure to demonstrate a consistent relation between dose and withdrawal symptoms is that subjects in clinical studies are being treated for an anxiety disorder, and when the benzodiazepine is discontinued, the re-emergence of the original anxiety symptoms may be confused with withdrawal symptoms. Furthermore, the drug dosage is probably higher in patients with the most severe disease, complicating any relationship between dosage and the intensity of the discontinuance syndrome.

According to one study, withdrawal symptoms are related to treatment duration when the length of the treatment is less than 8 months but not when it is 1 year or longer. The findings of another group indicate that patients taking benzodiazepines for more than 5 years have more withdrawal symptoms than those taking them for less than 5 years.

Mild withdrawal symptoms may occur with abrupt discontinuation of therapeutic doses after 4 weeks of benzodiazepine treatment. There is a risk of rebound insomnia after a few days to 1 week of treatment with benzodiazepine hypnotic drugs with short elimination half-lives. Benzodiazepine hypnotics with longer elimination half-lives are less likely to induce rebound insomnia on abrupt discontinuation. The likelihood of a serious withdrawal syndrome increases as treatment continues, and many authorities see 4 months of treatment at therapeutic doses as a critical point in the development of clinically significant physiologic dependence. That does not imply, however, that 4 months is an upper limit for treatment duration.

Table 11.10–2.
Factors Influencing the Development of the Benzodiazepine Discontinuation Syndrome

Dosage of benzodiazepine
Duration of benzodiazepine treatment
Rate of drug taper
Psychopathology

The risk of withdrawal is also influenced by the rate at which the drugs are discontinued. Gradual tapering of the benzodiazepines with short elimination half-lives is associated with fewer withdrawal symptoms than their abrupt discontinuation. Clinical research shows less distinct differences between gradual and abrupt discontinuation of benzodiazepines with long half-lives because these drugs have a self-tapering action or because the period of tapering and observation in many studies is insufficient.

If benzodiazepines are abruptly discontinued, the withdrawal syndrome related to short–half-life agents appears earlier and may be more intense than that with long–half-life drugs. Differences in intensity have not been proven because most studies do not monitor withdrawal symptoms for 2 weeks, which may be the time of peak symptoms with some drugs with long half-lives. Furthermore, any difference in withdrawal severity between short- and long–half-life agents is not entirely supported by animal studies. There is no difference in symptom severity of the withdrawal syndrome after animals are treated with midazolam or chlordiazepoxide, for example, because essentially all benzodiazepines, regardless of their pharmacokinetic properties in humans, are eliminated rapidly in small animal species.

The development of physiologic dependence differs depending on whether it results from benzodiazepine partial agonists or full agonists. In animal models, the partial agonists bretazenil and abecarnil are associated with fewer withdrawal symptoms than are classic agonists such as diazepam or clorazepate (the prodrug for desmethyldiazepam). Clinical studies in anxious patients, however, indicate that abrupt termination of abecarnil is associated with withdrawal symptoms in some patients.

Eszopiclone has demonstrated only mild withdrawal symptoms with discontinuation of therapeutic doses after 6 months of treatment. Supratherapeutic doses, however, lead to an abstinence syndrome that is indistinguishable from that produced by classical benzodiazepines. Clinical experience with eszopiclone, zolpidem, and zaleplon is limited compared to benzodiazepine hypnotics, although many clinicians feel that rebound insomnia, upon discontinuation, may be mild and of a shorter duration than with classical benzodiazepines; however, sufficient studies are lacking.

Some animal studies indicate that periodic administration of the benzodiazepine antagonist flumazenil during the chronic administration of lorazepam, diazepam, triazolam, or clobazam may attenuate the withdrawal syndrome.

Personality traits may be a risk factor for the development of the benzodiazepine withdrawal syndrome. Withdrawal severity is greater in patients with higher scores on the dependence scale of the Minnesota Multiphasic Personality Inventory 2 (MMPI-2), high prewithdrawal levels of anxiety and depression, lower educational level, and passive-dependent personality disorder.

Patient attitudes and expectations toward benzodiazepine cessation and withdrawal are important clinical variables, with high-dose benzodiazepine users reporting frequent unsuccessful attempts to discontinue and having a perception that withdrawal was a difficult, complicated, and unpredictable process.

OVERDOSE. The benzodiazepines, in contrast to the barbiturates and the barbiturate-like substances, have a large margin of safety when taken in overdoses, a feature that contributed significantly to their rapid acceptance. The ratio of lethal to effective dose is approximately 200 to 1 or higher because of the minimal respiratory depression associated with the benzodiazepines. Even when grossly excessive amounts (more than 2 g) are taken in suicide attempts, the symptoms include only drowsiness, lethargy, ataxia, some confusion, and mild depression of the user’s vital signs. A much more serious condition prevails when benzodiazepines are taken in overdose in combination with other sedative–hypnotic substances, such as alcohol. In such cases, small doses of benzodiazepines can cause death. The availability of flumazenil, a specific benzodiazepine antagonist, has reduced the lethality of the benzodiazepines because flumazenil can be used in emergency rooms to reverse the effects of the benzodiazepines.

Barbiturates. Since the advent of the benzodiazepines, barbiturate use has been limited in modern medicine. Phenobarbital is still prescribed as an anticonvulsant and as a sedative, especially for children. It is also a common component of many combination products and reduces the stimulating effects of sympathomimetic agents. Butalbital is an intermediate-acting barbiturate found in a widely used combination product that also contains acetaminophen and caffeine (Fioricet) and is approved for the treatment of muscle contraction headaches. Given the limited availability of barbiturates, it is uncommon, at least in the United States, for addiction units to treat patients dependent on barbiturates other than butalbital and phenobarbital. The barbiturates marketed currently in the United States are listed in Table 11.10–3.

A major disadvantage of the use of the barbiturates is the development of pharmacokinetic and pharmacodynamic tolerance. *Tolerance* is defined as reduced drug response as a result of decreased drug concentration at the site of action, usually the result of increased drug metabolism (pharmacokinetic) or of cellular adaptive changes with unchanged or higher drug concentrations at the site of action (pharmacodynamic). Pharmacodynamic tolerance begins after acute doses and continues to develop over weeks to months. Tolerance to the mood-altering and sedative effects develops to a greater extent than does tolerance to the lethal effects, increasing the risk of accidental overdose.

INTOXICATION. When barbiturates and barbiturate-like substances are taken in relatively low doses, the clinical syndrome of intoxication is indistinguishable from that associated with alcohol intoxication. The symptoms include sluggishness, incoordination, difficulty in thinking, poor memory, slowness of speech and comprehension, faulty judgment, disinhibition of sexual and aggressive impulses, a narrowed range of attention, emotional lability, and exaggeration of basic personality traits. The sluggishness usually resolves after a few hours, but the impaired judgment, distorted mood, and impaired motor skills may remain for 12 to 24 hours, depending primarily on the half-life of the misused substance. Other potential symptoms are hostility, argumentativeness, moroseness, and occasionally, paranoid and suicidal ideation. The neurologic effects include nystagmus, diplopia, strabismus, ataxic gait, positive Romberg sign, hypotonia, and decreased superficial reflexes.

Table 11.10–3.
Barbiturates and Other Sedative–Hypnotic Agents

Generic Name	Trade Name	Dose (mg)
Amobarbital	Amytal	100
Aprobarbital	Alurate	40
Butabarbital	Butisol	100
Butalbital	Many combination products (e.g., Fioricet)	100
Pentobarbital	Nembutal	100
Secobarbital	Seconal	100
Chloral hydrate	Aquachloral Supprettes	500
Ethchlorvynol	Placidyl	500
Glutethimide	Doriden	250
Meprobamate	Miltown	400

Note: The substitution technique for sedative–hypnotic withdrawal requires calculation of equivalent doses of phenobarbital to replace the sedative–hypnotic agent that the patient is taking. The above are doses of various sedative–hypnotic agents for which a 30-mg dose of phenobarbital should provide adequate coverage of a withdrawal syndrome. Daily doses of phenobarbital should rarely exceed 600 mg using this protocol.

DEPENDENCE AND WITHDRAWAL. Physiologic dependence may develop after a daily dose of 400 mg of pentobarbital for 3 months; abrupt discontinuation results in paroxysmal abnormalities on the EEG in approximately 30% of patients. At a daily dose of 600 mg of pentobarbital for 1 to 2 months, a withdrawal syndrome characterized by anxiety, insomnia, anorexia, tremor, and EEG changes occurs in approximately half of patients, and 10% may have a single seizure. At higher dosages of 800 to 2,200 mg per day for several weeks to months, abrupt discontinuation leads to minor symptoms, which include apprehension and uneasiness, insomnia, muscular weakness, twitches, coarse tremors, myoclonic jerks, postural faintness and orthostatic hypotension, anorexia, vomiting, and EEG changes, within 24 hours of the last dose. Minor symptoms may persist for as long as 2 weeks. At high doses, major symptoms develop. As many as 75% of patients may have grand mal seizures on the second or third day after withdrawal, and two-thirds have more than one seizure. The interictal EEG shows four spike–wave discharges per second. Two-thirds of these patients develop delirium between the third and eighth days of withdrawal, which is sometimes accompanied by hypothermia, which may be fatal. Disorientation, visual hallucinations, and frightening dreams may precede the onset of full delirium. Delirium may be exceedingly difficult to reverse, even with large doses of a barbiturate; thus, clinicians should never wait for the appearance of withdrawal symptoms before instituting therapy. The duration of the withdrawal syndrome is between 3 and 14 days; most end by day 8.

OVERDOSE. Barbiturates are lethal when taken in overdose because they induce respiratory depression. In addition to intentional suicide attempts, accidental or unintentional overdoses are common. Barbiturates in home medicine cabinets are a common cause of fatal drug overdoses in children. As with benzodiazepines, the lethal effects of barbiturates are additive to those of other sedative–hypnotic drugs, including alcohol and benzodiazepines. Barbiturate overdose is characterized by induction of coma, respiratory arrest, cardiovascular failure, and death.

The lethal dose varies with the route of administration and the degree of tolerance for the substance after a history of long-term abuse. For the most commonly misused barbiturates, the ratio of lethal to effective dose ranges between 3 to 1 and 30 to 1. Dependent users often take an average daily dose of 1.5 g of a short-acting barbiturate, and some have been reported to take as much as 2.5 g per day for months. The lethal dose is not much greater for the long-term user than it is for the neophyte. Tolerance develops quickly to the point at which withdrawal in a hospital becomes necessary to prevent accidental death from overdose.

Miscellaneous Sedative–Hypnotic Drugs. The older sedative–hypnotic drugs are similar in effect to the barbiturates, and they lead to tolerance and physiologic dependence. The withdrawal syndrome is similar to the barbiturate withdrawal syndrome, and the protocols described for barbiturates should be used for safe withdrawal of patients dependent on those drugs.

MEPROBAMATE AND CARISOPRODOL. Meprobamate, a carbamate derivative, has weak efficacy as an antianxiety agent. At clinical doses in humans, it has minimal muscle-relaxant effects, but it may have a mild analgesic effect in musculoskeletal pain and may potentiate analgesics. Typical daily doses are between 1,200 and 1,600 mg, with a maximum of 2,400 mg, in three or four divided daily doses. A discontinuance syndrome occurs after several weeks of treatment with a daily dose of 2,400 mg, and mild symptoms may be seen with long-term therapy at doses of 1,600 mg daily. Onset of the discontinuance syndrome occurs within 12 to 48 hours after abrupt discontinuation and lasts for an additional 12 to 48 hours. Seizures may be common in withdrawal from meprobamate, and reports published in the 1970s suggest that serious withdrawal symptoms are more common with meprobamate than with barbiturates. Fatalities from overdose have been reported after doses as low as 12 g. The manufacturer offers the following guidelines for blood concentration: 5 to 20 μg/mL is the normal range with recommended doses; 30 to 100 μg/mL is associated with mild-to-moderate overdose, with patients in a stupor or light coma; 100 to 200 μg/mL is associated with serious overdose, deeper coma, and fatalities; concentrations greater than 200 μg/mL are associated with more fatalities than survivals. Meprobamate induces microsomal enzymes and may exacerbate intermittent porphyria.

Carisoprodol is a commonly prescribed muscle relaxant that is metabolized to meprobamate, hydroxyl meprobamate, and hydroxyl carisoprodol and has been associated with abuse, dependence, and impairment of driving ability.

γ-HYDROXYBUTYRATE. γ-Hydroxybutyrate (GHB) was used for decades in Europe with a limited number of reported adverse effects and misuse. However, beginning in the late 1980s, when clinical use of GHB as a dietary supplement, growth promoter, and mild sedative became more prevalent, public health officials began inquiring about its safety and efficacy. By the early 1990s, concern that GHB was developing into a popular drug of misuse and evidence that GHB was being used as a recreational club drug led to FDA removal of GHB from the market. The FDA has now approved GHB (sodium oxybate) as a treatment for excessive daytime sleepiness and cataplexy in patients with narcolepsy. Also, it is still possible to obtain products containing GHB and its precursors through the internet.

GHB is a naturally occurring short-chained fatty acid naturally distributed throughout the CNS. It is synthesized in the brain from GABA via deamination to succinic semialdehyde (SSA) by GABA aminotransferase. Other precursors of GHB, such as 1,4-butanediol and γ-butyrolactone, may have important roles as well. GHB activity is terminated by synaptic reuptake and subsequent metabolism of SSA to succinate, which is further degraded to carbon dioxide and water. Although the half-life of GHB is reported as 30 to 50 minutes, its pharmacokinetics are nonlinear, so that high doses may take considerably longer to clear from the body.

The pharmacologic actions of GHB are mediated through specific GHB receptors that are G-protein coupled. GHB may also bind to GABA type B ($GABA_B$) receptors, although some investigators believe that the affinity is so low that this may not be of physiologic importance. Other data suggest that weak $GABA_B$ agonist activity has been demonstrated in studies that have shown that $GABA_B$ antagonists block GHB effects and that the $GABA_B$ agonist baclofen produces effects similar to GHB. Euphoric effects of GHB may be mediated by enhanced dopamine release in the striatum and mesolimbic system, although the relationship of dopaminergic effects to misuse liability is not well understood. Abrupt discontinuation of chronic high doses has been associated with anxiety, restlessness, insomnia, nausea, lethargy, tachycardia, sweating, tremors, muscle cramps, and psychosis. However, 6-month open-label clinical trials of therapeutic doses of sodium oxybate have not shown tolerance. The discontinuation effects of sodium oxybate have not been systematically studied in humans, although in rare cases, abrupt discontinuation may be associated with anxiety and insomnia.

Determining the risk of GHB administration in humans is difficult due to insufficient laboratory testing of GHB intoxication, limited availability of data reporting blood and tissue levels of hospitalized patients, or co-ingestion of alcohol or other drugs in clinical cases. GHB, in combination with other CNS depressants, especially ethanol, creates amplified CNS depression and has been found to result in acute intoxication, often requiring medical attention. Similar to benzodiazepines and barbiturates, when GHB and alcohol are combined, memory impairment, confusion, increased sleep, and unconsciousness are likely to occur. Other adverse effects include mild hypothermia, dizziness, nausea, vomiting, weakness, loss of peripheral vision, agitation, hallucinations, decreased respiratory effort, and coma.

CHLORAL HYDRATE. Chloral hydrate is used as a hypnotic in doses of 0.5 to 2.0 g. After oral administration, it is rapidly transformed to trichloroethanol, the metabolite responsible for its pharmacologic activity. It is also effective in blocking experimentally induced seizures; however, anticonvulsive and hypnotic doses are similar in humans, making the benzodiazepines and barbiturates better choices for anticonvulsive medications.

At high doses, chloral hydrate may produce respiratory depression, hypotension, gastric necrosis, depressed cardiac contractility, and a shortened refractory period. Even at therapeutic doses, it is associated with gastric distress and flatulence. Somnambulism, disorientation, and paranoid ideation may occur. Tolerance and physiologic dependence develop with long-term use, and the withdrawal syndrome is similar to the barbiturate abstinence syndrome. Several drug interactions occur with chloral hydrate, including displacement of oral anticoagulants and acidic drugs from protein binding sites by the metabolite trichloroacetic acid. When combined with furosemide, flushing, tachycardia, hypotension or hypertension, and diaphoresis have been reported. It potentiates the effects of ethanol, and it should not be used in patients with intermittent porphyria. Fatalities from overdose may occur with as little as 4 g, although patients have survived after taking 30 g.

Table 11.10–4. Guidelines for the Treatment of Benzodiazepine Discontinuance Syndrome

- Evaluate and treat concomitant medical and psychiatric conditions
- Obtain drug history and urine and blood samples for drug and ethanol assay
- Determine required dose of benzodiazepine or barbiturate for stabilization, guided by history, clinical presentation, drug-ethanol assay, and (in some cases) challenge dose
- Detoxification from supratherapeutic dosages:
  - Hospitalize if there are medical or psychiatric indications, poor social supports, or polysubstance dependence or if the patient is unreliable
  - Some clinicians recommend switching to a longer-acting benzodiazepine for withdrawal (e.g., diazepam [Valium], clonazepam [Klonopin]); others recommend stabilizing on the drug that the patient was taking or on phenobarbital
  - After stabilization, reduce dosage by 30% on the second or third day and evaluate the response, keeping in mind that symptoms that occur after decreases in benzodiazepines with short elimination half-lives (e.g., lorazepam [Ativan]) appear sooner than with those with longer elimination half-lives (e.g., diazepam)
  - Reduce dosage further by 10–25% every few days, if tolerated
  - Use adjunctive medications if necessary—carbamazepine (Tegretol), gabapentin (Neurontin), β-adrenergic receptor antagonists, divalproex (Depakote), clonidine (Catapres), and sedative antidepressants have been used, but their efficacy in the treatment of the benzodiazepine abstinence syndrome has not been established
- Detoxification from therapeutic dosages:
  - Initiate 10–25% dose reduction and evaluate response
  - Dose, duration of therapy, and severity of anxiety influence the rate of taper and the need for adjunctive medications
  - Most patients taking therapeutic doses have uncomplicated discontinuation
  - Psychological interventions may assist patients in detoxification from benzodiazepines and in the long-term management of anxiety

TREATMENT

Withdrawal

Benzodiazepines. Guidelines for the treatment of benzodiazepine discontinuance syndrome and the clinical management of withdrawal are presented in Table 11.10–4. The most common clinical situations that require medically supervised withdrawal from the benzodiazepines involve patients with anxiety disorders or chronic insomnia who have been maintained on therapeutic doses of benzodiazepines for several months or years, patients taking supratherapeutic doses of benzodiazepines for treatment-resistant types of anxiety or because of inappropriate dosage escalation, and patients who misuse benzodiazepines as part of a mixed misuse pattern (e.g., those who use benzodiazepines to self-medicate alcohol withdrawal or to end a cocaine run).

In planning a rational withdrawal, clinicians should remember the factors that influence the development of physiologic dependence. The risk of withdrawal is greatest in patients taking high doses over long periods. The onset of withdrawal and its peak severity are earlier with agents with short half-lives, and if they are not tapered, withdrawal from them may be more severe than that from agents with long elimination half-lives, although the last point has not been definitively established. Concurrent use of other sedative–hypnotic agents (e.g., barbiturates or ethanol) may alter the time course and severity of the benzodiazepine withdrawal syndrome.

TAPERING. The basic principle underlying safe withdrawal is gradual tapering of the benzodiazepine dosage. In cases of detoxification from therapeutic doses, the daily dose is known, and the tapering protocol is easily calculated. Anyone taking a benzodiazepine for 2 weeks or longer should be tapered from the drug. Initially, the dose is reduced by approximately 10% to 25%, and the patient is observed for any signs and symptoms of withdrawal. Withdrawal signs appear earlier for drugs with short half-lives than for drugs with long half-lives. Subsequent reductions depend on how the patient responds to initial changes and require individualization based on careful observation of the patient’s condition after each dose change.

Lower doses of benzodiazepines and no previous withdrawal attempts predict successful discontinuation of benzodiazepines in patients with benzodiazepine use disorder. In individuals with benzodiazepine use disorder, clinically significant dose reductions are associated with improvements in quality of life measures. Treatment options for benzodiazepine use disorder are limited, with no approach clearly effective.

In some cases, patients taking high therapeutic doses of a benzodiazepine for a year or more may require several months or longer to stop the medication completely, not primarily for pharmacologic reasons but so they can learn alternative strategies for coping with anxiety. Most patients taking midrange therapeutic doses for a shorter time tolerate weekly reductions ranging from 10% to 25%. Most authorities agree that the last phases of drug discontinuation are the most difficult for the patient, and it may be necessary to slow the taper rate or to stop the taper entirely. The adjunctive use of nonbenzodiazepine medications may be useful in these circumstances.

Detoxification from supratherapeutic dosages of benzodiazepines may require a slightly different approach. Most clinicians hospitalize such patients because of the greater medical risks associated with supratherapeutic dosage withdrawal. Patients may be tapered using the benzodiazepine that they have been taking or switched to a drug with a long elimination half-life, such as diazepam or clonazepam, using the dose equivalencies listed in Table 11.10–5 to determine the initial daily dose. The estimated equivalent dose is then administered in divided doses on day 1 to ensure that an accurate history and appropriate equivalency are established. Some clinicians stabilize patients on that dose for 2 to 3 days; others prefer to reduce the dose by 30% on the second day, followed by 5% to 10% daily reductions.

Research studying the management of discontinuation of therapeutic benzodiazepine treatment has generally supported the approach of a slow, gradual taper. For patients with benzodiazepine use disorder, a consensus exists that gradual taper is preferred to abrupt discontinuation. Small clinical trials have tested propranolol, buspirone, and progesterone without any clear evidence of efficacy. Hydroxyzine can reduce anxiety symptoms associated with a benzodiazepine taper. Carbamazepine has shown promise as an adjunct to the gradual taper of benzodiazepine. Trazodone and valproate were associated with improved abstinence rates but had no effect on withdrawal symptoms. Cyamemazine, an anxiolytic antipsychotic agent, has shown promise as a treatment for benzodiazepine withdrawal syndrome. The use of a slow infusion of the benzodiazepine antagonist flumazenil has been shown to be feasible but carries a high risk of seizures. No clearly effective pharmacotherapy for the treatment of benzodiazepine use disorder has been identified. The use of carbamazepine or divalproex in high-dosage withdrawal often permits even larger daily reductions in the benzodiazepine dose. The efficacy of other anticonvulsants, such as levetiracetam, gabapentin, and topiramate, has not been established, but these agents are sometimes used by experienced clinicians. Pregabalin shows promise as a treatment for tapering long-term benzodiazepine use. Most high-dose detoxification can be completed in 2 weeks or less using this protocol. For some patients, the dose cannot be reduced that rapidly, and slower, longer tapers are sometimes necessary instead. The efficacy of anticonvulsants for the treatment of the benzodiazepine discontinuance syndrome has not been established, although clinical experience to date is encouraging.

Table 11.10–5. Approximate Therapeutic Equivalent Doses of Benzodiazepines

Generic Name	Trade Name	Dose (mg)
Alprazolam	Xanax	1
Chlordiazepoxide	Librium	25
Clonazepam	Klonopin	0.5–1.0
Clorazepate	Tranxene	15
Diazepam	Valium	10
Estazolam	ProSom	1
Flurazepam	Dalmane	30
Lorazepam	Ativan	2
Oxazepam	Serax	30
Temazepam	Restoril	20
Triazolam	Halcion	.25
Quazepam	Doral	15
Zolpidem	Ambien	10
Zaleplon	Sonata	10

Some authorities have proposed agonist substitution for high-dose benzodiazepine users using a model successful for treating opioid use disorder. While research has not yet thoroughly examined this approach, the concept of using a long-acting benzodiazepine as an agonist replacement strategy is intriguing.

When the physician knows that the patient has been taking supratherapeutic doses but cannot determine the exact dose, 20 mg of diazepam may be given to estimate tolerance. The dose should be repeated every 2 hours until mild sedation occurs. The total dose required to induce mild sedation is then considered the initial dose, and the detoxification proceeds as described previously.

If the primary drug that the patient is taking is not going to be used in the withdrawal protocol, clonazepam or lorazepam may be the best drug to use when detoxifying patients who have been taking supratherapeutic dosages of triazolobenzodiazepines (e.g., triazolam). It is not clear whether higher doses of diazepam would have alleviated the alprazolam withdrawal syndrome in the published case reports. Some authorities recommend the use of phenobarbital in such cases. Withdrawal from the nonbenzodiazepine hypnotics zolpidem, zaleplon, and eszopiclone is usually treated with benzodiazepines, although case reports describe gradual taper of the hypnotic as also being effective.

Despite the risk of benzodiazepine use problems among methadone maintenance treatment patients, no clearly effective strategy for managing this clinical problem has been developed. As for the specific circumstance of managing methadone maintenance treatment patients with a benzodiazepine use disorder, limited research has been conducted. Potential treatment interventions that have been investigated, including benzodiazepine maintenance, contingency reinforcement, and slow outpatient detoxification, have not been accepted into clinical practice. The optimal strategy for the treatment of benzodiazepine use disorders in methadone maintenance treatment patients remains unknown.

DETOXIFICATION FROM MULTIPLE DRUGS OF MISUSE. Polysubstance misuse may present special problems for safe withdrawal. The most common polysubstance misuse pattern influencing medical management of detoxification from benzodiazepines is the concurrent use of alcohol, opioids, and cocaine. Combination with alcohol alters the time course and increases the severity of the sedative withdrawal syndrome. Adequate dosages of a benzodiazepine usually suffice for safe withdrawal; however, in rare instances, even large dosages of diazepam can be ineffective, and switching to a barbiturate becomes necessary, with phenobarbital being the preferred agent.

When supratherapeutic doses of benzodiazepines are misused with large amounts of opioids, it is best to discontinue the former drug first and then stabilize the patient on methadone or buprenorphine. When the period of risk of major symptoms of sedative–hypnotic withdrawal passes, opioid withdrawal may begin. Addiction to low or moderate dosages of benzodiazepines or opioids often permits simultaneous withdrawal from both classes of drugs.

Benzodiazepines are often used to terminate cocaine or other psychostimulant run or to decrease anxiety induced by cocaine. Benzodiazepines used intermittently in that manner usually do not require detoxification. The chronic use of high doses of benzodiazepines with cocaine is less common but does require medically supervised benzodiazepine tapering. The cocaine abstinence syndrome is characterized by depression, irritability, lethargy, lack of motivation, hypersomnolence, confusion, and drug craving. The clinical presentation of the sedative–hypnotic withdrawal syndrome may be altered by recent cocaine use and is best monitored by carefully measuring vital signs (e.g., increases in pulse, temperature, and blood pressure).

Some clinicians prefer to use phenobarbital for the detoxification of patients with mixed-drug misuse. A dose of phenobarbital equivalent to the benzodiazepine and other sedatives the patient is taking is calculated (see Tables 11.10–3 and 11.10–5) or, preferably, the dose is determined from a challenge test, and the patient is stabilized on that amount in divided daily doses for the period of peak risk of withdrawal. The maximum daily dose of phenobarbital is 600 mg. After stabilization, the phenobarbital dose is reduced by 30 to 60 mg every 2 or 3 days.

ADJUNCTIVE MEDICATIONS. Several adjunctive medications are used to treat the benzodiazepine withdrawal syndrome, but in no case has their efficacy been consistently supported in controlled trials, nor is there evidence that any specific adjunctive medication is superior to another.

Carbamazepine, given in doses that achieve the therapeutic concentrations for treatment of seizures, may partially attenuate the withdrawal syndrome from benzodiazepines. Initial dosages are 200 mg twice a day, although some clinicians recommend a single dose of 400 mg at bedtime to take advantage of the sedative effect. If initial doses are well tolerated, carbamazepine is increased to 600 mg on the second or third day. Typical total daily doses range from 400 to 800 mg. After several days of treatment, the benzodiazepine dosage is tapered. Despite reports that tapering can be completed within a week, even from supratherapeutic dosages, some patients require a slow taper that lasts several weeks. Carbamazepine can be tapered quickly 2 to 4 weeks after the last benzodiazepine dose. Carbamazepine may not be successful in withdrawing patients with panic disorder because it is ineffective in blocking panic attacks, but it has shown some success even in these patients, as long as a slow benzodiazepine taper is used. The efficacy of other anticonvulsants is undetermined, but clinical experience suggests that valproate may reduce withdrawal symptoms as well. Enteric-coated tablets of divalproex are used in starting dosages of 500 to 1,000 mg divided into two or three daily doses. Dosages are increased to achieve serum concentrations of 50 to 120 μg/mL. Some clinicians prefer to use daily loading doses of 20 mg/kg. For patients who consume alcohol and benzodiazepines, valproate should be avoided when liver enzyme activities are three times their normal values. Gabapentin and pregabalin may also reduce benzodiazepine withdrawal symptoms, but experience is limited.

The α-adrenergic receptor antagonists may also attenuate withdrawal symptoms, although meta-analyses have challenged their efficacy. Propranolol in doses of 60 to 120 mg may reduce the symptoms of benzodiazepine withdrawal, although tachycardia and elevated blood pressure are the primary symptoms affected, whereas the subjective sense of discomfort and dysphoria persists. The α-adrenergic receptor antagonists do not have prophylactic effects against seizures and should not be used in monotherapy for withdrawal. Nevertheless, in some cases, propranolol or atenolol, 50 to 100 mg, may be helpful adjuncts to withdrawal treatment.

Clonidine, 0.1 mg twice a day to 0.2 mg three times a day or as a patch, has been used in the treatment of low-dose benzodiazepine withdrawal with variable results. Clonidine does not provide protection against withdrawal seizures. It may be effective only when used before withdrawal symptoms develop.

Although antidepressants have not been rigorously studied in the treatment of benzodiazepine withdrawal, many experienced clinicians prescribe them. Agents with well-established antipanic and sedative effects may be useful adjunctive medications for patients whose anxiety recurs during taper or after drug discontinuation. Imipramine, amitriptyline, doxepin, nefazodone, trazodone, SSRIs, and MAOIs may all be useful in that role. Melatonin may be a useful treatment for insomnia during withdrawal, although clinical studies have been contradictory.

A potential approach under study is the administration of flumazenil, a benzodiazepine receptor antagonist. While flumazenil is typically used as a treatment for benzodiazepine overdose, it has been studied as a method for benzodiazepine detoxification when administered by a slow, gradual infusion.

PSYCHOLOGICAL TREATMENTS. Specific psychological treatments for withdrawal from anxiolytics and hypnotics include brief advice, various types of cognitive behavioral therapy, and medically managed withdrawal. Cognitive behavioral strategies for anxiety management have been applied to patients withdrawing from benzodiazepines with mixed results. Patients who meet the criteria for generalized anxiety disorder are poor candidates for relaxation training. Cognitive behavioral therapy appears particularly effective for patients withdrawing from hypnotics, although one study found structured medical management as effective as self-help cognitive behavioral therapy (which incorporates education, planned sage reductions, engaging patients in their taper schedule, and medical support).

Many practitioners believe that successful withdrawal requires transmitting to the patient a sense of control over the withdrawal symptoms. In support of this belief, placebo substitution is associated with fewer withdrawal symptoms after drug discontinuation than is withdrawal without placebo. Apparently, the belief that one is taking an effective medication is enough to reduce symptoms.

One way to foster a sense of control over withdrawal symptoms is to link the symptoms of anxiety to environmental intrapsychic stressors. Cognitive restructuring teaches patients about the withdrawal syndrome and helps them identify and relabel withdrawal symptoms as anxiety, thus permitting the implementation of adaptive coping strategies, which include systematic desensitization, in vivo graded exposure, and group problem solving. Using a diary to record mood states and their precipitants sometimes helps cognitive restructuring. A diary may also alert the clinician to maladaptive responses to benzodiazepine withdrawal, such as increased alcohol consumption.

The concept of *cognitive coping* should be introduced to patients, and examples of self-statements (talking to oneself) that are applicable to withdrawal and anxiety management should be discussed and provided as flash cards. They include such statements as “I feel nervous now, but it won’t last,” “I feel bad, but it’s not worse than when I had the flu,” or “I feel uncomfortable, but nothing bad is going to happen to me.” Self-statements should be short, simple, and relevant to the patient.

Patients are encouraged to be active participants in developing their withdrawal schedules. Clinicians should be flexible about the rate of taper and encourage the patient to join in the withdrawal process, especially those who are withdrawing from therapeutic doses. It is often useful to request that patients choose which doses should be decreased or eliminated according to their perceived needs. For some patients, setting a maximum permissible daily dose may be necessary.

LONG-TERM OUTCOME. Few adequately designed follow-up studies have been carried out on patients who terminated benzodiazepine treatment. Many studies suffer from sample selection bias, diagnostic heterogeneity of the sample, and failure to assess environmental factors or to control for concurrent psychotherapy or psychoactive medication; however, despite those shortcomings, several interacting findings have emerged.

In one study of patients who had discontinued benzodiazepines from 10 months to 3.5 years earlier, 70% were no longer taking benzodiazepines and had few or no symptoms, 22% were not taking benzodiazepines but required other psychotropic medications, and 8% had returned to benzodiazepine use. Another study assessed patients who were evaluated for a discontinuation program 3 to 6 years earlier and found that 45% were not using benzodiazepines, although some were using other psychoactive medications. Of those who successfully completed the tapering program, 73% were benzodiazepine free at follow-up, compared to 39% of patients who entered but did not complete the program and 14% of patients who did not participate in the program. Another study of patients treated for benzodiazepine dependence found that 56% of benzodiazepine-free (no use during the preceding month) patients had moderate-to-severe anxiety, and 38% had moderate-to-severe depression at 1- to 5-year follow-up. A different study found that although 66% of patients were not taking benzodiazepines at follow-up, 75% had taken them at some time during the 5 years after entering a discontinuation program.

The follow-up studies suggest that most patients can successfully discontinue taking benzodiazepines. They also indicate that some patients experience recurrent anxiety that requires alternative psychoactive medication or reinstitution of benzodiazepine therapy. However, a subgroup of patients who discontinue benzodiazepines actually shows improvement in anxiety over pretapering levels, which suggests that those who continue to have significant symptoms of anxiety or depression while taking a benzodiazepine should probably undergo a drug-free trial period. The value of a structured program for discontinuation and the substitution of medications with low misuse liability is supported in several studies. There appears to be no relationship between long-term outcome (i.e., not using benzodiazepines at follow-up) and the duration of treatment or the specific benzodiazepine that was discontinued. Some evidence suggests that low doses of benzodiazepines predict successful discontinuation, although other studies have found weak or no correlation between dose and return to benzodiazepine use.

Barbiturates. There are three common methods of establishing the barbiturate dosage required for safe withdrawal: (1) administration of a test dose of pentobarbital to determine tolerance, (2) calculation of required doses based on estimated equivalencies, and (3) administration of phenobarbital loading doses.

In the first method, the initial step is to determine tolerance. Once intoxication has subsided, but before withdrawal symptoms have developed, an oral dose of 200 mg of pentobarbital is administered on an empty stomach, and the effects are observed after 1 hour to determine the stabilization dose (Table 11.10–6). If no changes are observed, the test is repeated 3 hours later using 300 mg. If there is still no response, the total requirement may exceed 1,600 mg per day. The calculated daily dose is divided and given every 4 to 6 hours for a 2- to 3-day stabilization period. Daily dose reductions are usually 10% of the stabilization dose, but withdrawal regimens must be individualized. Phenobarbital may be substituted for pentobarbital for stabilization and withdrawal at one-third of the dose.

Table 11.10–6.
Pentobarbital (Nembutal) Test Dose Procedure for Barbiturate Withdrawal

Symptoms after Test Dose of 200 mg of Oral Pentobarbital	Estimated 24-Hour Oral Pentobarbital Dose (mg)	Estimated 24-Hour Oral phenobarbital Dose (mg)
Level I: Asleep, but arousal is possible; withdrawal symptoms not likely	0	0
Level II: Mild sedation; patient may have slurred speech, ataxia, and nystagmus	500–600	150–200
Level III: Patient is comfortable; no evidence of sedation; may have nystagmus	800	250
Level IV: No drug effect	1,000–1,200	300–600

Adapted from Ciraulo DA, Shader RI. *Clinical Manual of Chemical Dependence*. American Psychiatric Press; 1991 and data from Ewing JA, Bakewell WE. Diagnosis and management of depressant drug dependence. *Am J Psychiatry*. 1967;123:909.

The second method of barbiturate detoxification requires calculating the equivalent hypnotic dose of phenobarbital based on the dosage of barbiturate or other sedatives the patient reports taking, also taking into consideration any alcohol consumed. A dose of 30 mg of phenobarbital is substituted for an equivalent hypnotic dose of the sedative. The patient is stabilized for the period of peak vulnerability to withdrawal, 2 to 3 days for sedative–hypnotic agents with short half-lives, and reductions of 30 to 60 mg of phenobarbital are made every 2 or 3 days. Clinicians using this method report that only rarely are doses of phenobarbital greater than 600 mg required. The procedure is not recommended because published equivalencies are only approximations, and patients vary greatly in their metabolism and pharmacodynamic response to the same dose of the sedative. Furthermore, animal studies indicate that cross-tolerance among the various anxiolytic and sedative–hypnotic agents is not complete. Furthermore, the method relies on the patient’s history of drug use, which is often unreliable. Plasma concentrations can establish which sedatives have been ingested and may indicate the degree of tolerance when clinical correlations are made. Estimation of tolerance (and thus determination of dosage requirements) is best made after a test dose or a loading dose is administered. Tables 11.10–3 and 11.10–5 list approximate equivalencies of selected barbiturates, benzodiazepines, and sedatives.

The final method for medical withdrawal from barbiturates involves using loading doses of phenobarbital. According to the original protocol, 120-mg doses are administered every 1 to 2 hours until three of five signs are present—nystagmus, drowsiness, ataxia, dysarthria, or emotional lability—or, if patients are in withdrawal, until abstinence symptoms abate. No additional drug is administered, and because of its long half-life, the drug self-tapers. Patients are assessed before each dose. Some clinicians have modified the protocol to use doses of 60 mg every 1 to 2 hours and a gradual taper rather than abrupt discontinuation. The originators of the protocol reported a mean loading dose of 1,440 mg, with hourly doses sometimes required for 15 to 20 hours. They have also administered phenobarbital IV at 0.3 mg/kg per minute, with close medical supervision, for medically ill patients. A loading dose procedure for withdrawal from butalbital-containing headache remedies has also been described. When patients are known to be taking less than 10 tablets (i.e., Fioricet) per day, reductions of 1 tablet every 2 to 3 days may be successful. However, since patients do not always reliably report drug intake, an alternative strategy is to use a loading dose of 120 mg of phenobarbital hourly, guided by a rating scale to document effects on nystagmus, dysarthria, ataxia, and drowsiness. One study found that the mean number of phenobarbital doses required was 9.7 (range 7 to 14), and withdrawal was completed in less than 24 hours.

Overdose

The treatment of overdose of the general class of substances involves gastric lavage, activated charcoal, and careful monitoring of vital signs and CNS activity. Overdose patients who come to attention while awake should be kept from slipping into unconsciousness. Activated charcoal should be administered to delay gastric absorption. If the patient is comatose, the clinician must establish an IV fluid line, monitor the patient's vital signs, insert an endotracheal tube to maintain a patent airway, and provide mechanical ventilation, if necessary. Hospitalization of a comatose patient in an intensive care unit is usually required during the early stages of recovery from such overdoses.

FURTHER READINGS

Al-Hasani R, Gowrishankar R, Schmitz GP, et al. Ventral tegmental area GABAergic inhibition of cholinergic interneurons in the ventral nucleus accumbens shell promotes reward reinforcement. *Nat Neurosci*. 2021;24(10):1414–1428.

Allison C, Pratt JA. Neuroadaptive processes in GABAergic and glutamatergic systems in benzodiazepine dependence. *Psychopharmacol Ther*. 2003;98:171.

American Psychiatric Association. *Diagnostic and Statistical Manual Of Mental Disorders*. 5th ed. American Psychiatric Association; 2022. https://doi.org/10.1176/appi.books.9780890425787

Bruce SE, Vasile RG, Goisman RM, et al. Are benzodiazepines still the medication of choice for patients with panic disorder with or without agoraphobia? *Am J Psychiatry*. 2003;160:1432.

de las Cuevas C, Sanz E, de la Fuente J. Benzodiazepines: more “behavioural” addiction than dependence. *Psychopharmacology (Berl)*. 2003;167:297.

*Denis C, Fatséas M, Lavie E, Auriacombe M. Pharmacological interventions for benzodiazepine mono-dependence management in outpatient settings. *Cochrane Database Syst Rev*. 2006;3:CD005194.

Druid H, Holmgren P, Ahlner J. Flunitrazepam and evaluation of use, abuse, and toxicity. *Forensic Sci Int*. 2001;122:136.

Faccini M, Leone R, Opri S, et al. Slow subcutaneous infusion of flumazenil for the treatment of long-term, high-dose benzodiazepine users: a review of 214 cases. *J Psychopharmacol*. 2016;30(10):1047–1053.

Hood S, O’Neil G, Hulse G. The role of flumazenil in the treatment of benzodiazepine dependence: physiological and psychological profiles. *J Psychopharmacol*. 2009;23(4):401–409.

Huang B, Dawson DA, Stinson FS, et al. Prevalence, correlates, and comorbidity of nonmedical prescription drug use and drug use disorders in the United States: results of the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry*. 2006;67(7):1062–1073.

*Lee SC, Klein-Schwartz W, Doyon S, Welsh C. Comparison of toxicity associated with nonmedical use of benzodiazepines with buprenorphine or methadone. *Drug Alcohol Depend*. 2014;138:118–123.

Liebrenz M, Gehring MT, Buadze A, Caflisch C. High-dose benzodiazepine dependence: a qualitative study of patients’ perception on cessation and withdrawal. *BMC Psychiatry*. 2015;15:116.

Liu S, O’Donnell J, Gladden RM, McGlone L, Chowdhury F. Trends in nonfatal and fatal overdoses involving benzodiazepines—38 states and the district of Columbia, 2019–2020. *MMWR Morb Mortal Wkly Rep*. 2021;70:1136–1141.

MacDonald T, Gallo AT, Basso-Hulse G, Bennett KS, Hulse GK. A double-blind randomised crossover trial of low-dose flumazenil for benzodiazepine withdrawal: a proof of concept. *Drug Alcohol Depend*. 2022;236:109501.

Martínez-Hernández E, Sandoval A, González-Ramírez R, Zoidis G, Felix R. Inhibition of recombinant N-type and native high voltage-gated neuronal Ca^{2+} channels by AdGABA: mechanism of action studies. *Toxicol Appl Pharmacol*. 2011;250(3):270–277.

*Möhler H. The rise of a new GABA pharmacology. *Neuropharmacology*. 2011;60(7–8):1042–1049.

Morin CM, Bastien C, Guay B, Radouco-Thomas M, Leblanc J, Vallières A. Randomized clinical trial of supervised tapering and cognitive behavior therapy to facilitate benzodiazepine discontinuation in older adults with chronic insomnia. *Am J Psychiatry*. 2004:161(2):332–342.

Mumford GK, Evans SM, Fleishaker JC, Griffiths RR. Alprazolam absorption kinetics affects abuse liability. *Clin Pharmacol Ther*. 1995;57:356.

Nardi AE, Freire RC, Valença AM, et al. Tapering clonazepam in patients with panic disorder after at least 3 years of treatment. *J Clin Psychopharmacol*. 2010;30(3):290–293.

Nordfjærn T. A population-based cohort study of anxiety, depression, sleep and alcohol outcomes among benzodiazepine and z-hypnotic users. *Addict Behav*. 2012;37(10):1151–1157.

*Otto MW, McHugh RK, Simon NM, Farach FJ, Worthington JJ, Pollack MH. Efficacy of CBT for benzodiazepine discontinuation in patients with panic disorder: further evaluation. *Behav Res Ther*. 2010;48(8):720–727.

Posternak MA, Mueller TI. Assessing the risks and benefits of benzodiazepines for anxiety disorders in patients with a history of substance abuse or dependence. *Am J Addict*. 2001;10:48.

Schepis TS, Simoni-Wastila L, McCabe SE. Prescription opioid and benzodiazepine misuse is associated with suicidal ideation in older adults. *Int J Geriatr Psychiatry*. 2019;34(1):122–129.

Song J, Shen G, Greenfield LJ Jr, Tietz EI. Benzodiazepine withdrawal-induced glutamatergic plasticity involves up-regulation of GluR1-containing alpha-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptors in Hippocampal CA1 neurons. *J Pharmacol Exp Ther*. 2007;322(2):569–581.

Specka M, Bonnet U, Heilmann M, Schifano F, Scherbaum N. Longitudinal patterns of benzodiazepine consumption in a German cohort of methadone maintenance treatment patients. *Hum Psychopharmacol*. 2011;26:404–411.

Struble CA, Thomas K, Stenersen MR, et al. Sexual minority disparities in opioid and benzodiazepine misuse among adults with opioid use disorder. *Am J Addict*. 2022;31(3):200–209.

Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. *Treatment Episode Data Set (TEDS): 2020. Admissions to and Discharges from Publicly Funded Substance Use Treatment Facilities*. Substance Abuse and Mental Health Services Administration; 2022.

Substance Abuse and Mental Health Services Administration. *Key Substance Use and Mental Health Indicators in the United States: Results From the 2020 National Survey on Drug Use and Health* (HHS Publication No. PEP21-07-01-003, NSDUH Series H-56). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2021. https://www.samhsa.gov/data

Substance Abuse and Mental Health Services Administration. *Preliminary Findings from Drug-Related Emergency Department Visits, 2021; Drug Abuse Warning Network* (HHS Publication No. PEP22-07-03-001). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2022. https://www.samhsa.gov/data

Tan KR, Brown M, Labouèbe G, et al. Neural bases for addictive properties of benzodiazepines. *Nature*. 2010;463(7282):769–774.

*Vorma H, Naukkarinen H, Sarna S, Kuoppsalami K. Long-term outcome after benzodiazepine withdrawal treatment in subjects with complicated dependence. *Drug Alcohol Abuse*. 2003;70:309.

Vorma H, Naukkarinen HH, Sarna SJ, Kuoppasalmi KI. Predictors of benzodiazepine discontinuation in subjects manifesting complicated dependence. *Subst Use Misuse*. 2005;40(4):499–510.

Vorma H, Naukkarinen H, Sarna S, Kuoppasalmi K. Symptom severity and quality of life after benzodiazepine withdrawal treatment in participants with complicated dependence. *Addict Behav*. 2004;29(6):1059–1065.

World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*. 11th ed. World Health Organization; 2019. https://icd.who.int

Xiang K, Tietz EI. Benzodiazepine-induced hippocampal CA1 neuron alpha-amino-3-hydroxy-5-methylisoxasole-4-propionic acid (AMPA) receptor plasticity linked to severity of withdrawal anxiety: differential role of voltage-gated calcium channels and N-methyl-D-aspartic acid receptors. *Behav Pharmacol*. 2007;18(5–6):447–460.

11.11 Anabolic-Androgenic Steroid-Related Disorders

Harrison G. Pope, Jr., M.D., and Gen Kanayama, M.D., Ph.D.

The use of anabolic-androgenic steroids (AAS) has now become a major form of substance use disorder worldwide, with tens of millions of individuals who have used these drugs to improve athletic performance or, more commonly, to simply enhance personal appearance. However, AAS use is a recent phenomenon, having become common in the general population only within the last three or four decades. Because widespread AAS misuse is so new, psychiatric research findings on the epidemiology, features, and treatment of this condition remain limited in comparison to the volume of research on other forms of substance use disorders. In particular, there are still only a modest number of published clinical studies where AAS users have been recruited and evaluated in person. Thus, many findings in this field remain tentative, and erroneous beliefs about AAS use are still common among both lay individuals and mental health professionals.

Table 11.11–1.
Examples of Commonly Used Anabolic Steroids

Compounds usually administered orally

- Fluoxymesterone (Halotestin, Android-F, Ultrandren)
- Methandienone (formerly called methandrostenolone; Dianabol)
- Methyltestosterone (Android, Testred, Virilon, Metandren)
- Oxandrolone (Anavar)
- Oxymetholone (Anadrol, Hemogenin)
- Mesterolone (Mestoranum, Proviron)
- Stanozolol (Winstrol)

Compounds usually administered intramuscularly

- Nandrolone decanoate (Deca-Durabolin)
- Nandrolone phenylpropionate (Durabolin)
- Drostanolone propionate (Masteron)
- Methenolone enanthate (Primobolan Depot)
- Boldenone undecylenate (Equipoise[a])
- Stanozolol (Winstrol-V[a])
- Testosterone esters blends (Sustanon, Sten)
- Testosterone cypionate
- Testosterone enanthate (Delatestryl)
- Testosterone propionate (Testoviron, Androlan)
- Testosterone undecanoate (Andriol, Restandol)
- Trenbolone acetate (Finajet, Finaplix[a])
- Trenbolone hexahydrobenzylcarbonate (Parabolan)

Note: Many of the brand names listed above are foreign but are included because of the widespread illicit use of foreign steroid preparations in the United States.

[a] Veterinary compound.

The AAS are a family of hormones that includes the natural male hormone testosterone, together with numerous derivatives of testosterone that have been synthesized over the last 80 years (Table 11.11–1). All AAS produce both *anabolic* (muscle-building) and *androgenic* (masculinizing) effects. Importantly, AAS should be distinguished from other families of steroid hormones such as *corticosteroids,* which include the natural hormone cortisol, together with its synthetic derivatives such as prednisone. Corticosteroids have no anabolic properties and hence little misuse potential. Often, when the generic term "steroids" is used, individuals confuse these two families of hormones and may think that the "steroids" administered by doctors for shoulder or back pain are the same "steroids" used illicitly by weightlifters to gain muscle. To avoid this confusion, the term "AAS," rather than "steroids" is used throughout this chapter to indicate that the authors are referring to testosterone and its synthetic relatives.

HISTORY

Some of the highlights in the history of AAS are summarized in Figure 11.11–1. Testosterone was first synthesized in 1935 in Nazi Germany, and over the next few years, numerous synthetic analogs of testosterone were developed, creating the family of AAS. Both the anabolic and androgenic actions of these drugs were quickly recognized. In the late 1930s and early 1940s, a number of psychiatric reports suggested that testosterone and other AAS might possess antidepressant properties. These hormones were believed to be especially effective for the "male climacteric" and were prescribed in Europe and in the United States for depressed middle-aged men. With the advent of electroconvulsive therapy (ECT) and standard antidepressant medications, however, the use of AAS for psychiatric purposes rapidly waned. Then, starting in the 1990s, interest in the therapeutic effects of AAS in psychiatry was renewed, partially as a result of the recognition that testosterone produced antidepressant effects in men with HIV infection. Also, by the 2000s, various transdermal testosterone gels were developed and marketed, eliminating the need to administer testosterone by cumbersome intramuscular injections. Thus, the medical use of testosterone for its potential antidepressant properties, or simply (and more controversially) as an "antiaging" treatment, has increased markedly in the United States over the last two decades.

In parallel with the evolution of medical applications for AAS, the use of these drugs for muscular development and athletic performance also began to evolve soon after World War II. Elite athletes quickly discovered that these drugs could create muscle gains greater than could be achieved by any natural individual, and by the 1950s, AAS use had already entered competitive sports, especially those requiring muscle mass and strength, such as bodybuilding, powerlifting, American football, and field events such as the shot put. The first widely documented use of AAS in international competition was by the Russians at the world weightlifting championships in Vienna in 1954. By the 1960s and 1970s, AAS were used extensively in the Olympics and other major athletic events. Perhaps the most dramatic example of AAS use was in the German Democratic Republic, which developed an extensive state-sponsored doping program, which allowed a country with only 18 million citizens to win more than 150 Olympic gold medals between 1960 and 1990.

Although the power of AAS was widely recognized by athletes, the medical community was embarrassingly slow to acknowledge these effects. During the 1960s and 1970s, numerous placebo-controlled studies attempted to assess whether AAS were effective for gaining muscle mass. Many of these studies produced negative results, almost certainly because they employed dosages and durations of AAS administration far below those used by actual athletes in the field. On the basis of these often-flawed studies, many reputable textbooks and reviews in pharmacology and endocrinology concluded that AAS were ineffective for athletic purposes. Athletes who had actually used these drugs tended to hold the medical profession in contempt for this display of ignorance. Evidence suggests that even to this day, AAS users often distrust medical professionals on the subject and consequently discount medical warnings about the possible physical and psychiatric dangers of these agents. For example, in one recent study of 43 AAS users, 56% reported that they had never disclosed their AAS use to any physician that they had ever seen, and 40% reported that they would trust information on AAS from their local drug dealer at least as much as information from any physician.

AAS use remained largely confined to athletes until the 1980s but then began to break out of the athletic world and onto the street. Underground handbooks on how to use AAS began to appear, providing detailed information about the various AAS available on the illicit market and instructions on how to use and self-inject them. Also, during this era, Western culture showed a growing emphasis on male muscularity. Images of muscular male bodies increasingly appeared in Western magazines, advertising, television, and Hollywood movies. Even children’s action toys, such as “G.I. Joe,” grew steadily more muscular from the 1970s onward. These trends helped to propel a rapid rise in the prevalence of illicit AAS use among ordinary rank-and-file men lifting weights in gymnasiums, who often had no athletic aspirations at all but simply wanted to look bigger. Starting in the late 1980s, these trends began to be documented by epidemiologic studies of AAS use among students, using anonymous questionnaires. Several of these studies noted that a majority of students were using AAS not for athletic purposes, but only for personal appearance. More recent studies of adult AAS users have repeatedly reinforced this impression, finding that a great majority of AAS users have not used these drugs for any competitive athletic activity at all. Nevertheless, many people still erroneously believe that AAS use is primarily a problem of athletes.

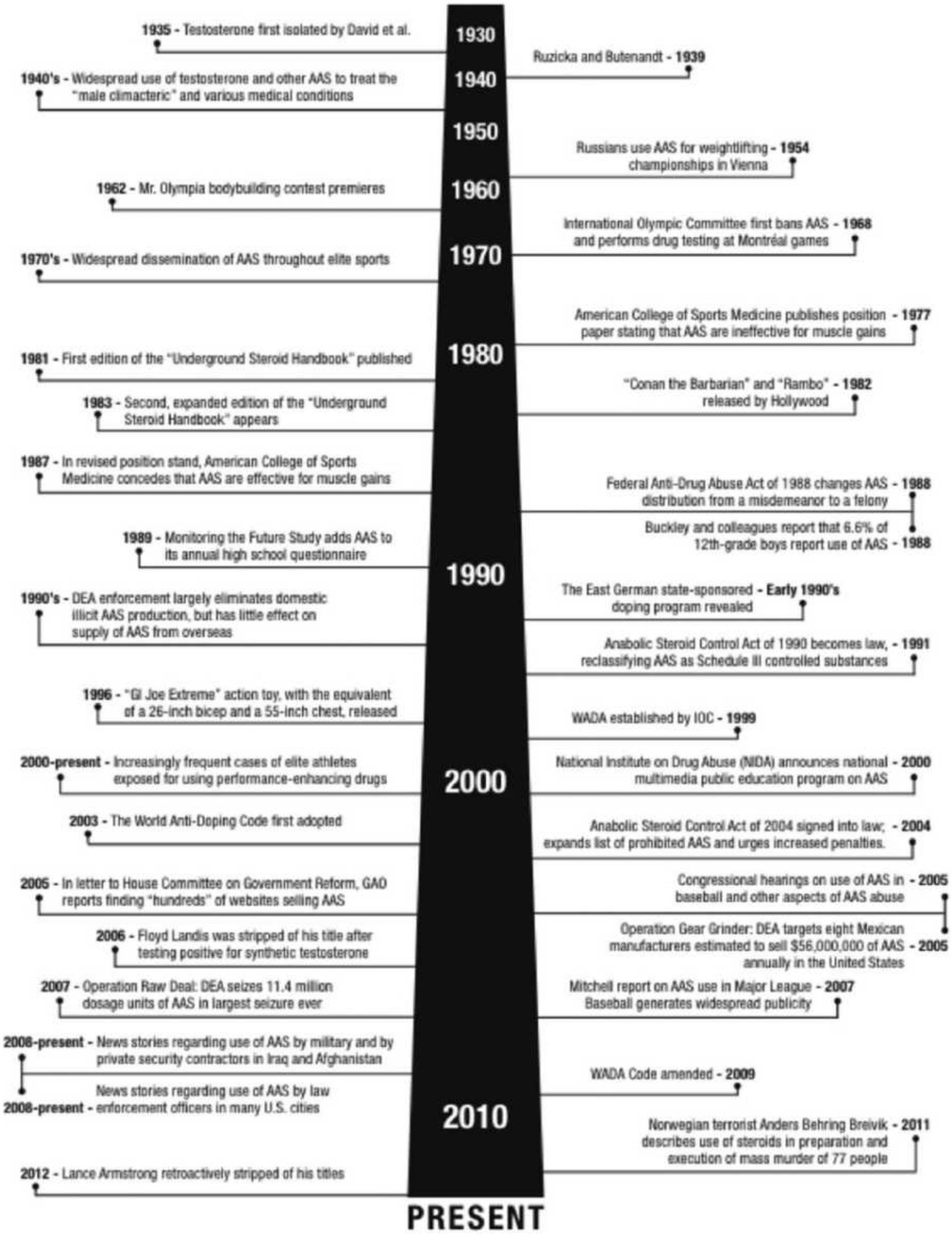
1930
1940
1950
1960
1970
1980
1990
2000
2010
PRESENT
1935 - Testosterone first isolated by David et al.
Ruzicka and Butenandt - 1939
1940's - Widespread use of testosterone and other AAS to treat the "male climacteric" and various medical conditions
Russians use AAS for weightlifting championships in Vienna - 1954
1962 - Mr. Olympia bodybuilding contest premieres
International Olympic Committee first bans AAS and performs drug testing at Montréal games - 1968
1970's - Widespread dissemination of AAS throughout elite sports
American College of Sports Medicine publishes position paper stating that AAS are ineffective for muscle gains - 1977
1981 - First edition of the "Underground Steroid Handbook" published
"Conan the Barbarian" and "Rambo" released by Hollywood - 1982
1983 - Second, expanded edition of the "Underground Steroid Handbook" appears
1987 - In revised position stand, American College of Sports Medicine concedes that AAS are effective for muscle gains
Federal Anti-Drug Abuse Act of 1988 changes AAS distribution from a misdemeanor to a felony - 1988
Buckley and colleagues report that 6.6% of 12th-grade boys report use of AAS - 1988
1989 - Monitoring the Future Study adds AAS to its annual high school questionnaire
The East German state-sponsored doping program revealed - Early 1990's
1990's - DEA enforcement largely eliminates domestic illicit AAS production, but has little effect on supply of AAS from overseas
Anabolic Steroid Control Act of 1990 becomes law, reclassifying AAS as Schedule III controlled substances - 1991
1996 - "GI Joe Extreme" action toy, with the equivalent of a 26-inch bicep and a 55-inch chest, released
WADA established by IOC - 1999
2000-present - Increasingly frequent cases of elite athletes exposed for using performance-enhancing drugs
National Institute on Drug Abuse (NIDA) announces national multimedia public education program on AAS - 2000
2003 - The World Anti-Doping Code first adopted
Anabolic Steroid Control Act of 2004 signed into law; expands list of prohibited AAS and urges increased penalties. - 2004
2005 - In letter to House Committee on Government Reform, GAO reports finding "hundreds" of websites selling AAS
Congressional hearings on use of AAS in baseball and other aspects of AAS abuse - 2005
Operation Gear Grinder: DEA targets eight Mexican manufacturers estimated to sell $56,000,000 of AAS annually in the United States - 2005
2006 - Floyd Landis was stripped of his title after testing positive for synthetic testosterone
2007 - Operation Raw Deal: DEA seizes 11.4 million dosage units of AAS in largest seizure ever
Mitchell report on AAS use in Major League Baseball generates widespread publicity - 2007
2008-present - News stories regarding use of AAS by military and by private security contractors in Iraq and Afghanistan
2008-present - News stories regarding use of AAS by law enforcement officers in many U.S. cities
WADA Code amended - 2009
Norwegian terrorist Anders Behring Breivik describes use of steroids in preparation and execution of mass murder of 77 people - 2011
2012 - Lance Armstrong retroactively stripped of his titles

FIGURE 11.11–1. A historic timeline of the evolution of image- and performance-enhancing drug use. (Adapted from Kanayama G, Hudson JI, Pope HG Jr. Long-term psychiatric and medical effects of anabolic-androgenic steroids: A looming public health concern? *Drug Alcohol Depend.* 2008;98(1–

2):1–12. Copyright © 2008, with permission from Elsevier. doi: 10.1016/j.drugalcdep.2008.05.004.)

With the growing illicit use of AAS in the 1980s, preliminary reports also began to document adverse psychiatric and medical effects associated with using high-dose AAS. These reports included cases of hypomanic and manic syndromes associated with AAS use and depressive syndromes associated with AAS withdrawal. These findings led to increasing concern about the public health consequences of using AAS, and in 1991, the U.S. Congress passed the Steroid Trafficking Act, adding AAS to the list of Schedule III controlled substances and placing jurisdiction over these substances in the hands of the Drug Enforcement Administration (DEA). In the wake of this legislation, it quickly became difficult for individuals who misuse AAS to find doctors willing to prescribe these drugs, and production of AAS by American pharmaceutical companies declined rapidly. Thus, on the illicit market, the supply of genuine AAS diminished, and a larger number of counterfeit AAS, with no pharmacologic effects, began to appear. Underground guides to use AAS began to include sections with photographs of bottles and syringes to help users to distinguish genuine preparations from counterfeits. However, despite the existence of counterfeits, an abundant supply of genuine AAS continues to be available because, unlike drugs of misuse such as cocaine and marijuana, AAS may be purchased legally without a prescription in many countries and can be sent legally by mail to the United States (e.g., to post office boxes). Thus, with the rise of the internet, many AAS users can receive these drugs by mail from abroad, usually without interdiction.

Although the general public is quite familiar with many forms of illicit substance misuse, most laypersons and arguably even most physicians are surprisingly unfamiliar with the effects of AAS or the extent of their use. In particular, most people are unaware that, without massive doses of AAS, it is virtually impossible to achieve the combination of low body fat and high muscularity exhibited by many competitive bodybuilders and professional athletes. Since athletes and their promoters have large personal and financial motivations to deny such drug use, the public has been hoodwinked for decades about the magnitude of the AAS problem. However, recent years have seen mounting scandals and revelations about AAS use by elite athletes in the Olympics and other competitions, as well as by prominent sports professionals, movie and television actors, and law enforcement personnel. Public awareness has also increased in response to reports of premature deaths, including suicide, in athletes known or suspected to have taken AAS.

COMPARATIVE NOSOLOGY

Psychiatric syndromes associated with use of AAS are cited in the fifth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) under the general category of “other substance-related use disorders.” Within this category are subcategories of other substance use disorder, other substance intoxication, other substance withdrawal, other substance-induced disorders, and unspecified other substance-related disorder. Similarly, the 11th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-11) classifies AAS-related disorders under the category of “Disorders due to use of other specified psychoactive substances, including medications.” Unfortunately, these various terms do not apply easily to AAS use disorders, because AAS produce little “intoxication” in the manner of classical drugs of misuse and usually (but not always) lack major psychoactive effects. Therefore, in the text below, the authors will attempt to parallel the standard terminology of DSM-5 for other drugs of misuse, while recognizing that AAS deviate from this terminology in a number of important respects.

EPIDEMIOLOGY

The epidemiology of AAS use has long been associated with persistent erroneous beliefs, including particularly the misconceptions that (1) AAS are used primarily by athletes, (2) AAS users typically start use in adolescence, and (3) girls and women often use AAS.

It has been very difficult to estimate the lifetime prevalence of illicit AAS use in the United States because most available data rely on anonymous surveys of students, particularly students of high school age. There are two serious problems with the use of these data. First, the median age of onset of AAS use, both in the United States and other countries, is approximately 23 years—an older age of onset than for almost any other major category of illicit drugs. In pooled data from five American studies, collectively evaluating more than 2,000 AAS users in which age of onset was ascertained, only 5.9% of users had initiated AAS prior to the age of 17, and only 39.3% prior to age 21. Thus, studies of high-school students greatly underestimate the number of individuals who will ultimately use AAS in their lifetimes, because the individuals being surveyed have only barely begun to enter the age of risk for initiating AAS use.

Second, anonymous surveys are vulnerable to serious errors in the opposite direction as a result of false-positive responses to the survey questions about "steroids." When students are asked a typical survey question, such as whether they have used "steroid pills or shots without a doctor's prescription," they may incorrectly respond that they have used "steroids" when in fact they have taken only corticosteroids or have purchased nutritional supplements over the counter that they mistakenly believed were "steroids." Many students fail to realize that genuine AAS are available only on the black market in the United States and most other Western countries. Thus, they may often respond positively to a "steroid" question when in reality they have never used AAS at all. A recent analysis of anonymous survey data in the United States has suggested that virtually all high school girls claiming "steroid" use on questionnaires represent false positives and that the true use prevalence of AAS use in girls may be as low as 0.1%. This finding is hardly surprising when it is considered that girls rarely desire to become extremely muscular and are also vulnerable to the androgenic effects of AAS, such as beard growth, deepening of the voice, and masculinization of secondary sexual characteristics. Similarly, anonymous survey data for high-school boys also need to be corrected sharply downward.

After allowing for these two sources of error, and digesting all available American survey data, a recent analysis has suggested that the true prevalence of AAS use in high-school girls is close to 0%, and the prevalence of AAS use in high-school boys is about 1%. However, the same analysis suggested that by adulthood, a far larger proportion of American men have used AAS, with an estimate of 2.9% to 4.0% of American men who have tried AAS at some time in their lives. However, the corresponding prevalence for adult women is likely extremely low; in five available studies since the year 2000 that have recruited AAS users without regard to gender, the mean proportion of female users, calculated using a random effects model, was 1.8%—a male–female ratio of more than 50 to 1. Moreover, the authors are aware of only one study worldwide in the last 25 years in which a group of investigators recruited a cohort of actual female AAS users in person, and even in this study, the investigators located only 25 such women over a 2-year interval in three major American cities. Thus, from a public health standpoint, the problems of AAS use are largely confined to males.

AAS users often use other *ergogenic* (performance-enhancing) drugs to gain muscle, lose fat, improve aerobic capacity, or to lose water for bodybuilding competitions. These drugs include thyroid hormones, such as liothyronine and levothyroxine; human growth hormone and its daughter compound, somatomedin-C (also called insulin-like growth factor-1 [IGF-1]); insulin; sympathomimetic agents such as amphetamines, ephedrine, and pseudoephedrine; β-adrenergic receptor agonists such as clenbuterol; the amino acid derivative gamma-hydroxybutyrate (GHB); and various laxatives and diuretics.

AAS users also frequently display other substance use disorders involving classical drugs of misuse, such as cannabis, cocaine, and opioids. One interview study of 44 AAS users found that 44% had a history of misuse of or dependence upon at least one other drug or alcohol. Another study of 62 users found that 65% reported a history of misuse or dependence upon at least one drug other than alcohol. Similarly, among a large sample of 10,552 college students in 2001, a diagnosis of alcohol dependence in the past year was over seven times more likely among past-year AAS users than nonusers of AAS after adjusting for age, gender, race, intercollegiate athletic status, grade-point average, and geographical region. Several European studies have also documented an association between AAS use and misuse of multiple other classical drugs.

Opioid misuse and dependence appear to be particularly associated with AAS use in the United States. One study found that 21 (9%) of 227 sequential males addicted to heroin at one facility had apparently been introduced to opioids through AAS; they reported that they had first learned about opioids from fellow AAS users in the gym and had subsequently first purchased opioids from the same individual who had sold them AAS. Another study of 88 men admitted to a hospital unit for treatment of opioid dependence found that 22 (25%) had used AAS at some time in their lives. Interestingly, most of these men reported that AAS were the first drugs that they had learned to use by injection. Collectively, these data suggest that AAS use is often part of a pattern of polydrug use and that AAS have increasingly become “street” drugs, rather than drugs used purely for athletic performance.

PHARMACOLOGY

Chemistry

All steroid drugs—including AAS, estrogens, and corticosteroids—are synthesized in vivo from cholesterol and resemble cholesterol in their chemical structure (Fig. 11.11–2). Testosterone has a four-ring chemical structure containing 19 carbon atoms (Fig. 11.11–2). Most synthetic derivatives of testosterone are produced by one or more of the following modifications: (1) forming an ester at the C-17-β-hydroxyl group, (2) adding an alkyl group to the C-17-α position, or (3) altering the structure at a different carbon site.

Aqueous testosterone suspension is essentially inactive when ingested by the oral route because of first-pass metabolism in the liver and is only briefly active when injected because it is rapidly absorbed and metabolized. Synthetic AAS overcome these problems. Testosterone esters (e.g., testosterone cypionate and testosterone enanthate) confer the advantage of slower absorption than testosterone when injected, and their effects last from 2 to 4 weeks. Testosterone esters are hydrolyzed to testosterone after they are absorbed, so these agents can be monitored by blood concentrations. Testosterone is metabolized in part to dihydrotestosterone, which is up to 10 times more potent than testosterone as an androgen, and in part to the estrogen estradiol, which accounts for some of the adverse effects of AAS seen in men, such as gynecomastia. Testosterone is also metabolized to neurosteroids in the brain, especially 3-α-androstanediol, which may account for some of its psychoactive effects. The effects of testosterone depend both on the balance of metabolic enzymes and on the concentration of androgen receptors in a particular tissue. The C-17-alkylated synthetic AAS make oral administration possible, but once-daily use is generally required.

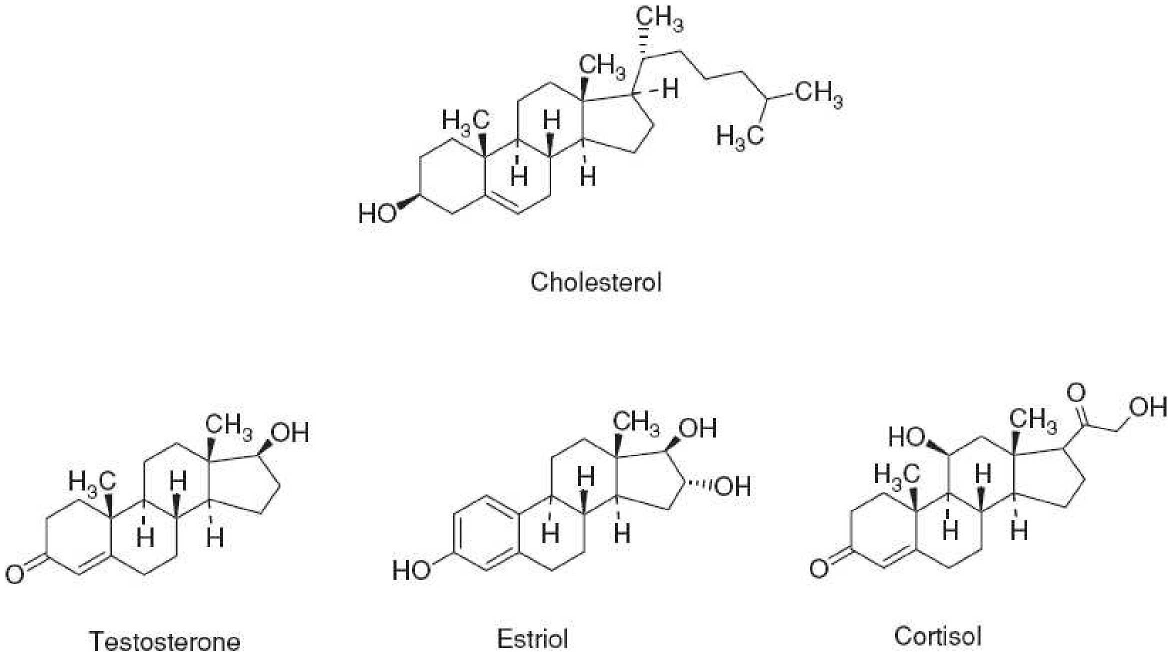

FIGURE 11.11–2. Chemical structure of cholesterol and three hormones derived from cholesterol: testosterone (an anabolic–androgenic steroid), estriol (an estrogenic steroid), and cortisol (a corticosteroid).

Pharmacodynamics

AAS bind androgen receptors in the cytoplasm of cells. The androgen-receptor complex is then translocated to the cell nucleus where it augments gene transcription and ultimately new protein synthesis. The anabolic effects desired by illicit users result especially from new protein synthesis in muscle tissue, which likely results from a genomic action of AAS. The mechanism of action for psychoactive effects is poorly understood but may involve both genomic and nongenomic actions of AAS as well as actions on both neurons and glial cells. Nongenomic actions may include binding to membrane-associated androgen receptors, including allosteric sites on γ-aminobutyric acid (GABA) receptors. Androgen metabolites such as estradiol and neurosteroids can also have psychoactive effects. The neurosteroid metabolite of testosterone, 3-α-androstanediol, for example, is a positive allosteric modulator of GABA receptors, thereby potentiating the action of GABA.

A number of neural structures and neurotransmitters have been implicated in the psychiatric effects of AAS. Most of this work comes from animal studies, demonstrating that parts of the limbic system (including the amygdala, hippocampus, and nucleus accumbens) and the hypothalamus are all target sites for the actions of AAS. In addition, androgen receptors are known to be expressed in the cerebral cortex. Animal studies also suggest that AAS-induced increases in vasopressin, substance P, and glutamate, as well as AAS-induced decreases in brain serotonin and 5-hydroxyindoleacetic acid (5-HIAA), may be associated with aggressive behavior. Forebrain GABAergic mechanisms may also be involved in AAS-induced aggression and anxiety. Finally, AAS may exert their reinforcing effects through their actions on GABA, opioid, and dopamine systems, as will be discussed below.

In humans, a recent experimental study reported significantly higher levels of 5-HIAA in the cerebrospinal fluid of men treated with methyltestosterone as compared to men receiving placebo. Moreover, 5-HIAA levels correlated significantly with certain AAS-associated effects, including increased energy, diminished sleep, and sexual arousal.

Normal testosterone plasma concentrations for men range from 300 to 1,000 ng/dL. In general, 200 mg of testosterone cypionate taken every 2 weeks will restore physiologic testosterone concentrations in a hypogonadal male. A eugonadal male who initiates physiologic dosages of testosterone will have no net gain in testosterone concentrations, because exogenously administered AAS will shut down endogenous testosterone production via feedback inhibition of the hypothalamic–pituitary–testicular (HPT) axis. Consequently, illicit users take dosages 10 to 100 times greater than therapeutic doses in order to achieve the desired effects. Doses in this range are most easily achieved by taking combinations of oral and injected AAS, which illicit AAS users often do. Transdermal testosterone gels or patches are generally inadequate to produce massively supraphysiologic AAS levels.

Effects on Specific Organs and Systems

During human development, the first effects of endogenous androgens are to induce male sexual differentiation in the fetus. Accumulating evidence suggests that fetal and perinatal androgens cause sexual dimorphism of the brain (at least in nonhuman mammals) as well as differentiation of the sexual organs. During puberty, testosterone levels increase in boys to produce male secondary sex characteristics. The phallus, prostate, and seminal vesicles all grow and develop, and testosterone and gonadotropins are involved in sperm production. Androgens cause the larynx to enlarge, the vocal cords to lengthen and thicken, and the voice to deepen. A male pattern of body hair develops, including hair on the face and chest. Sebaceous glands grow and increase their secretions, predisposing the adolescent to acne. Bones and muscles also grow during puberty as a result of androgen stimulation. Exogenous administration of AAS to a prepubertal male will also induce all of these effects.

In adults, physiologic levels of testosterone continue to be important for spermatogenesis, although abnormally high levels will suppress spermatogenesis by inhibiting the release of gonadotropins, which are also required for spermatogenesis. Exogenous AAS increase muscle growth, promote mineralization of bones, and stimulate erythropoiesis by increasing erythropoietin levels and by direct action on the bone marrow.

Medical Uses

AAS are primarily indicated for testosterone deficiency (male hypogonadism), hereditary angioedema (a congenital skin disorder), and some uncommon forms of anemia caused by bone marrow or renal failure. In women, they are given, although not as first-choice agents, for metastatic breast cancer, osteoporosis, endometriosis, and adjunctive treatment of menopausal symptoms. In men, they have been used experimentally as a male contraceptive and for treating depressive disorders and sexual disorders in eugonadal men. Although some data suggest that testosterone may be useful as an augmentation strategy in men with major depressive disorder refractory to antidepressant medication, findings in this area remain conflicting. More recently, testosterone has been used to treat wasting syndromes associated with AIDS, where it may also have antidepressant effects.

Adverse Medical Effects

The most common acute adverse medical effects of AAS include acne (which occasionally may be severe), gynecomastia (caused by aromatization of AAS to estrogenic hormones), striae (caused by overly rapid muscle growth beneath the skin), and acceleration of male pattern baldness. These features, if observed in an unusually muscular individual, may alert the clinician to surreptitious AAS use.

Although the *acute* adverse medical effects of AAS are usually modest, the *long-term* dangers of these drugs are substantial and remain inadequately studied. One reason for the dearth of research is the fact that widespread illicit AAS use did not appear until the 1980s, as mentioned above. Thus, even most of the older AAS users—those who first tried AAS as youths in the 1980s and 1990s—are only now reaching middle age and entering the age of risk for various adverse medical effects.

Perhaps the best-documented long-term adverse effects of AAS are on the cardiovascular system. A rapidly evolving literature has shown that AAS exposure can cause dilated cardiomyopathy, characterized by both systolic dysfunction (as evidenced by a reduced ventricular ejection fraction) and by diastolic dysfunction (as manifested by reduced diastolic tissue velocity). Numerous studies have now shown that the effects of weightlifting alone cannot explain these phenomena, in that non–AAS-using weightlifters display few abnormalities on these indices. Tentative data suggest that AAS-induced cardiomyopathy is at least partially reversible, with normalization of systolic and diastolic function occurring in a majority of AAS users within months after stopping the drugs. However, impaired diastolic tissue velocity may be at least partly attributable to fibrotic changes caused by AAS-induced cellular injury, and hence may be potentially irreversible.

Another serious cardiovascular effect of AAS is accelerated atherosclerosis. These atherosclerotic changes appear largely attributable to the fact that AAS increase low-density lipoprotein and decrease high-density lipoprotein cholesterol. Dyslipidemia of this type is particularly severe with use of 17-α-alkylated AAS that are administered orally. Atherosclerotic disease may cause premature death from myocardial or cerebral infarction among occasional AAS users even in their 20s or 30s; at least a dozen such case reports have been published in the last 20 years. One imaging study of 14 professional weightlifters with long-term AAS use found markedly higher coronary artery calcium scores than would be expected for men of comparable age. Another study performed coronary computed tomography angiography in 86 AAS users aged 34 to 54 years, all of whom reported at least 2 years of lifetime exposure to AAS, and 54 age-matched similar male weightlifters reporting no AAS use. On measures of the number of diseased coronary artery segments, the worst degree of stenosis, and total plaque volume, AAS users showed greater pathology than nonusers, and the degree of pathology was strongly associated with total duration of lifetime AAS exposure. The potential clinical consequences of such changes are suggested by a study from Finland that examined the mortality of 62 male powerlifters who placed first to fifth in Finnish championships during 1977 to 1982 (all highly likely to have used AAS) and 1,094 population controls. Eight (12.9%) of the powerlifters but only 34 (3.1%) of controls had died—a ratio of 4.6—and 3 of the 8 deaths among powerlifters were ascribed to myocardial infarction.

AAS have also been implicated in causing hypertension, cardiac conduction abnormalities, and thrombotic events in both human and animal studies, although the frequency and severity of these phenomena in AAS users remain poorly understood. Chronic use of high doses of AAS may also frequently cause polycythemia, with some users developing a hematocrit of 55% to 60%, and sometimes higher than 60%. Again, however, the magnitude and frequency of clinical morbidity associated with AAS-induced polycythemia are uncertain.

AAS may also exert substantial neuroendocrine toxicity because AAS suppress HPT function. Thus, male users will often become hypogonadal after stopping a course of AAS use. Although HPT function usually recovers spontaneously within a few weeks to a few months, especially in younger individuals with a shorter history of AAS use, hypogonadism may persist for more than a year after discontinuing AAS and appears to be partially or even completely irreversible in some cases. Several different mechanisms may contribute to AAS-induced hypogonadism in different men. Some appear to show hypogonadotropic hypogonadism, due to incomplete recovery of gonadotrope function. Several reports have described successful treatment of such cases with gonadotropin-releasing hormone agonists such as triptorelin, selective estrogen receptor modulators such as clomiphene, or agents mimicking pituitary hormonal function, such as human chorionic gonadotropin (HCG). However, these approaches sometimes fail. For example, some patients may respond to clomiphene with normalization of pituitary hormone levels, but then fail to normalize testosterone levels, suggesting possibly irreversible damage to Leydig cells. Recent reports have also described former AAS users who showed persistent erectile dysfunction and loss of libido even when treated with physiologic levels of exogenous testosterone—possibly reflecting some type of end-organ resistance.

One recent study compared 19 former long-term AAS users and 36 non–AAS-using weightlifters using measures of psychopathology, neuroendocrine function, and erectile function. Five of the former AAS users displayed markedly reduced testosterone levels of less than 200 ng/dL despite having discontinued AAS 3 to 26 months earlier. The former users also displayed markedly and significantly lower mean scores on ratings of sexual desire. Five former users had experienced episodes of major depressive disorder during AAS withdrawal. Another study compared 33 former AAS users who had last used AAS a median of 2.5 years earlier as compared to 30 similar weightlifters with no history of AAS use. Among the former AAS users, nine (27.3%) displayed plasma total testosterone levels below the normal reference limit of 12.1 nanomoles per liter (350 nanograms per deciliter) as compared to none of the control participants.

Also of concern is the effect of prolonged AAS use on tissues sensitive to androgens, such as the prostate. AAS may contribute to prostatic hypertrophy, and two case reports have documented prostate cancer among AAS users who were still in their 40s. However, both of these reports appeared more than 25 years ago; the apparent lack of reports since 1991 argues against a major association of AAS use with prostatic neoplasia.

Among the musculoskeletal effects of AAS is an increased risk of tendon rupture. One recent study found that 15 (17%) of 88 long-term AAS users had ruptured an upper body tendon (biceps, triceps, or pectoralis) at some time in their lives as opposed to 0 cases among 50 age-matched non–AAS-using weightlifters. Notably, in this and other studies, most of the AAS users did not rupture tendons while doing heavy weightlifting, but instead as a result of lifting objects, falling, or other daily activities. The increased risk of tendon rupture in AAS users appears at least partially attributable to the fact that these men develop muscles that become too strong relative to their tendons. However, some animal data suggest that AAS may additionally interfere with the architecture of the tendons themselves by creating collagen dysplasia, causing tendons to become stiffer and less flexible, with an increased crimp angle and earlier liability to failure.

Rarely, AAS may cause hepatotoxicity, with the effects including peliosis hepatis (an accumulation of blood-filled cysts in the liver), and the development of both benign and malignant hepatic tumors. These hepatotoxic effects occur almost exclusively with use of orally active 17-α-alkylated AAS. Importantly, the true incidence of AAS-induced hepatotoxicity is often overestimated, because rhabdomyolysis from weightlifting can increase transaminase levels, and this finding is often incorrectly interpreted as evidence of hepatotoxicity. In reality, true AAS-induced serious hepatotoxicity is very rare.

Rhabdomyolysis from heavy weightlifting may also possibly lead to renal failure in occasional AAS users, and one study has described a series of cases of focal segmental glomerulonephritis in long-term AAS users. Another study found estimated glomerular filtration rates of less than 90 mL/min/1.73 m^2 in 41 (48%) of 85 men aged 34 to 54 years currently using AAS as compared to 11 (21%) of 52 similar weightlifters with no history of AAS use. There is some possibility that AAS is very rarely associated with renal neoplasms.

Several recent studies have raised the alarming possibility that high doses of AAS may display neurotoxicity, thus possibly leading to cognitive deficits. For example, in several laboratory studies, supraphysiologic levels of AAS—within the range likely attainable by human AAS abusers—led to accelerated apoptotic effects on neuronal cells. Following up on these findings, one recent study has assessed cognitive function in long-term AAS users versus a comparison of non–AAS-using weightlifters. The study found significant deficits on two tests of visuospatial memory in the AAS users, and the severity of these deficits was significantly associated with total lifetime AAS exposure.

Another group, comparing 84 current or former AAS users with 69 nonusing comparison weightlifters found deficits in processing speed, working memory, and problem solving among AAS users as compared to controls even after adjustment for a number of potentially confounding variables. In another evaluation of the same cohort of users and nonusers employing magnetic resonance imaging of the brain, AAS users were found to display a thinner cortex together with smaller anatomical volumes including total grey matter and cerebral cortical volume. On functional magnetic resonance imaging, the AAS users also displayed reduced functional connectivity between key regions, including decreased connectivity between the amygdala and the default mode network, as well as a decrease in connectivity between the dorsal attention network and a frontal node that included the anterior cingulate cortex and superior and inferior frontal gyri.

In the rare females who use AAS, these drugs may induce shrinkage of breast tissue, irregular menses (diminution or cessation), and masculinization (clitoral hypertrophy, hirsutism, and deepened voice). Masculinizing effects in women may be irreversible, although this issue has not been well studied. Misuse of AAS by children has led to concerns that AAS-induced premature closure of bony epiphyses could cause shortened stature. However, as discussed above, use of true AAS by young teenagers is rare, and the authors have never personally seen an actual case of short stature unequivocally caused by AAS use. Other uncommon adverse effects include edema of the extremities due to water retention, exacerbation of tic disorders, and sleep apnea.

LABORATORY FINDINGS

As mentioned above, the use of AAS can be associated with elevations in liver function tests such as bilirubin, lactate dehydrogenase (LDH), alkaline phosphatase, aspartate aminotransferase (AST or SGOT), and alanine aminotransferase (ALT or SGPT). However, elevations in LDH, ALT, and AST, as well as the muscle enzyme creatine phosphokinase (CPK), are common in weight-training athletes, even in the absence of AAS use, as a result of muscle trauma from intensive weightlifting. The enzyme γ-glutamyl transaminase (GGT), which is present exclusively in liver and not in muscle, may be more useful to distinguish frank liver toxicity from mere muscle trauma in users of AAS.

As also discussed above, an altered cholesterol profile (decreased levels of high-density lipoprotein cholesterol and increased levels of low-density lipoprotein cholesterol) is frequently seen, especially with oral administration of the C-17-alkylated AAS. Total cholesterol and triglyceride concentrations may also be elevated. Increases in the hematocrit and hemoglobin levels may also occur.

Endocrine Tests

AAS inhibit the release of luteinizing hormone (LH) and follicle-stimulating hormone (FSH), resulting in lower serum levels. Serum testosterone and estradiol levels vary with the particular agent and timing of the sample. For example, the use of supraphysiologic doses of testosterone esters will increase both testosterone and estradiol levels, whereas the exclusive use of most other AAS will decrease these hormonal levels. Similarly, testosterone and estradiol levels are decreased during AAS withdrawal because of the lingering inhibition of the HPT axis, as discussed above. Semen analysis may reveal decreased sperm count and motility as well as abnormal sperm morphology.

ETIOLOGY

The major reason for taking illicit AAS is to enhance either athletic performance or physical appearance. The dramatic effects of AAS on muscle growth are illustrated in Figure 11.11–3, which compares a “natural” bodybuilder who has never used these drugs with a bodybuilder of identical height and body fat who has used AAS extensively. Although the anabolic properties of AAS represent the prime motivation for use, psychoactive effects may also contribute to the persistent and dependent use of AAS. Anecdotally, some AAS users report feelings of power, aggressiveness, and euphoria, which become associated with and can reinforce AAS taking.

As mentioned earlier, AAS use occurs almost entirely in males, despite persistent misconceptions that females commonly use these drugs. AAS use is widespread in Western cultures, with Scandinavian countries showing perhaps the highest prevalence, followed by the United States, British Commonwealth countries, and Brazil. Moderate rates of AAS use have been documented in studies in continental Europe and the Middle East, but AAS use is rare in Eastern Asian countries such as China, Korea, and Japan. It has been speculated that in these latter societies, masculinity is defined in the Confucian tradition, as someone who has intellect, integrity, and force of character—but not hypertrophied muscles. For example, far Eastern traditions have few equivalents to muscular Western statues depicting gods of Greek and Roman mythology, or to the depictions of the Norse gods or the Vikings. Thus, East Asian men may see little reward in attaining high levels of muscularity and hence are rarely attracted to AAS.

Two well-documented risk factors for AAS use among Western men are conduct disorder and body image disorders. One study examined 233 experienced American weightlifters aged 18 to 40 years, comprising 102 AAS users and 131 nonusers. In this study, men reporting a history of adolescent conduct disorder were significantly more likely to have subsequently developed AAS use. In the same study, men reporting higher levels of adolescent concern about muscularity showed a strikingly higher risk of starting AAS use than those reporting lower levels. Several other studies have shown a strong association between muscle dysmorphia—a form of body dysmorphic disorder characterized by preoccupations that one is not muscular enough—and use of AAS.

Interestingly, this literature suggests that muscle dysmorphia is a risk factor for initiation of AAS use, but paradoxically may worsen after initiating AAS use, perhaps because the individual becomes more involved in the AAS-using culture and becomes increasingly focused on body muscularity.

ANABOLIC STEROID DEPENDENCE

The possibility of AAS dependence was mentioned in the scientific literature as early as 1980. Three case reports appeared in the medical literature between 1988 and 1990, which described young men who initially took AAS to enhance their weightlifting and bodybuilding activity, and later used the drugs to combat depression when they tried to stop using them. The men reported feeling addicted and sought professional help to stop using illicitly obtained AAS because they were unable to accomplish discontinuation on their own.

Over the last three decades, nine studies, to the authors' knowledge, have utilized criteria for substance dependence adapted from DSM-III-R or DSM-IV to diagnose AAS dependence in 363 of 1,247 AAS users—a prevalence of 32.5%. Notably only two (0.6%) of the AAS-dependent individuals in these studies were women. One group has proposed formal diagnostic criteria for AAS dependence based on DSM-IV criteria for substance dependence. These criteria include symptoms of tolerance and withdrawal; prolonged use of AAS and difficulty in discontinuing AAS (e.g., because of anxiety about losing muscle mass); spending large amounts of time involved in AAS-related activities (weightlifting, attending to particular diets for gaining muscle mass, and associating with other AAS users); and continued use of AAS despite adverse effects, such as gynecomastia, sexual dysfunction, or evidence of cardiovascular toxicity.

Another group has suggested that it may be too simplistic to diagnose AAS dependence alone using traditional substance dependence criteria, because AAS use typically occurs as part of a larger complex of behaviors, including the use of other appearance- and performance-enhancing drugs, compulsive patterns of dieting and exercise, and behaviors associated with body image disturbance.

Etiology

AAS dependence may develop via at least three different etiologic pathways (Fig. 11.11–4). First, individuals with body-image disorders may become dependent upon AAS because they develop intense anxiety that they will lose muscularity if they stop taking these drugs. Such individuals may finish a course (a so-called “cycle”) of AAS use, but then start a new course of AAS more quickly than planned because of fears that they are beginning to get smaller. Eventually, they may come to use AAS continuously, with no off-drug periods, as a result of this anxiety. Second, as discussed above, men will often develop hypogonadism after stopping a course of AAS, with attendant loss of libido and erectile function, and possible symptoms of major depression. In response to these dysphoric symptoms, they may be quick to resume AAS use in order to restore normal mood and sexual function, again potentially leading to continuous or near-continuous AAS use as a result of this neuroendocrine “addiction.”

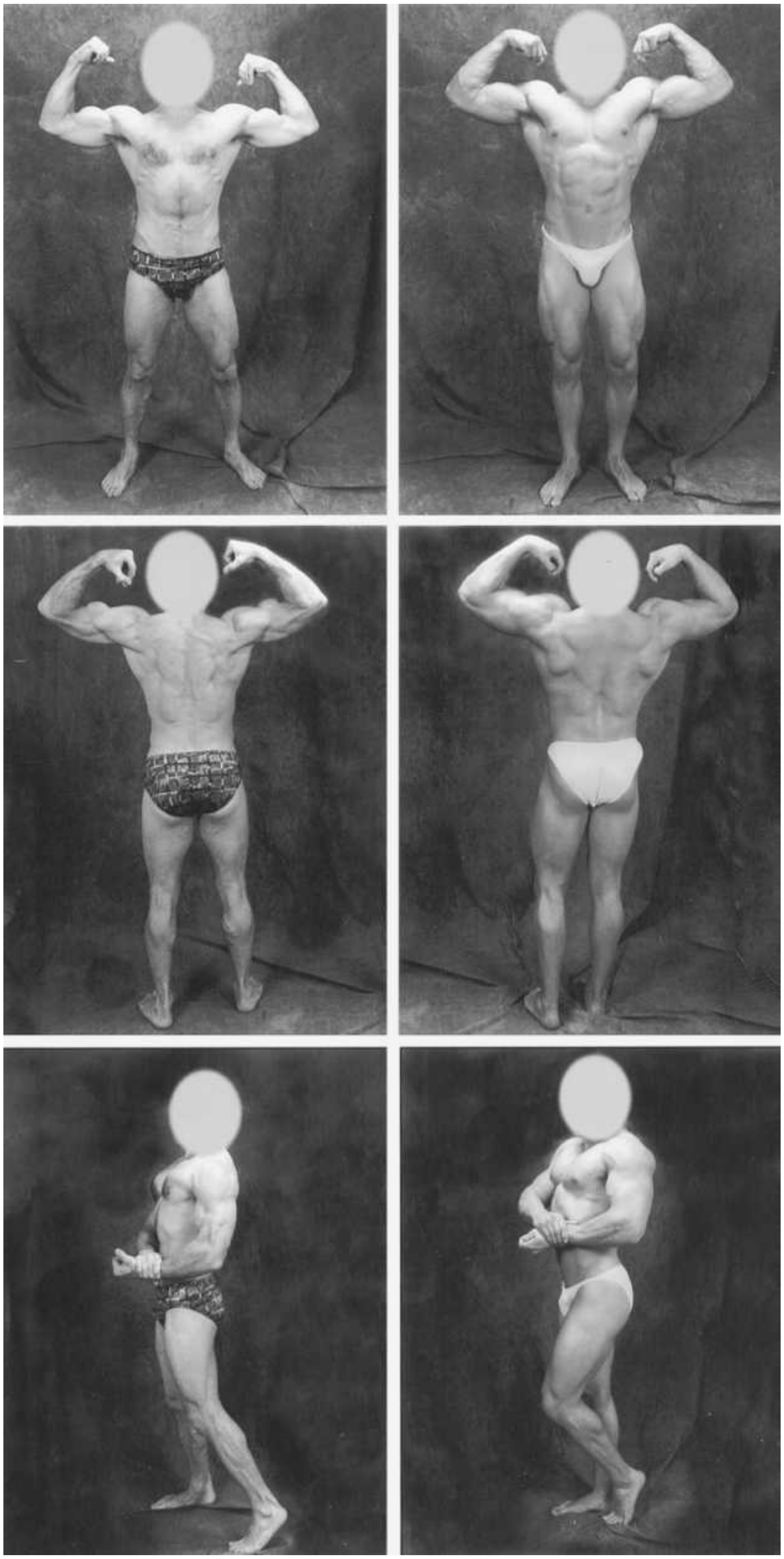

FIGURE 11.11–3. Physical effects of anabolic steroid use. These photographs compare a “natural” bodybuilder who has never used anabolic steroids (**left**) with a man who used large doses of anabolic steroids over several years

(**right**). Both men are 67 inches tall and have 7% body fat. The man on the **left** weighs 170 lb and represents approximately the maximum degree of muscularity obtainable without drugs. His fat-free mass index is 25.4 kg/m^2 (by the formula of Kouri EM, Pope HG Jr, Katz DL, Oliva PS. Fat-free mass index in users and nonusers of anabolic–androgenic steroids. *Clin J Sport Med*. 1995;5:223). The man on the **right** weighs 213 lb and has a fat-free mass index of 31.7 kg/m^2. Note that the muscle hypertrophy from steroid use is particularly marked in the upper body in the pectoralis, deltoid, trapezius, and biceps muscles. Any man significantly more muscular than the man on the left has almost certainly abused anabolic steroids.

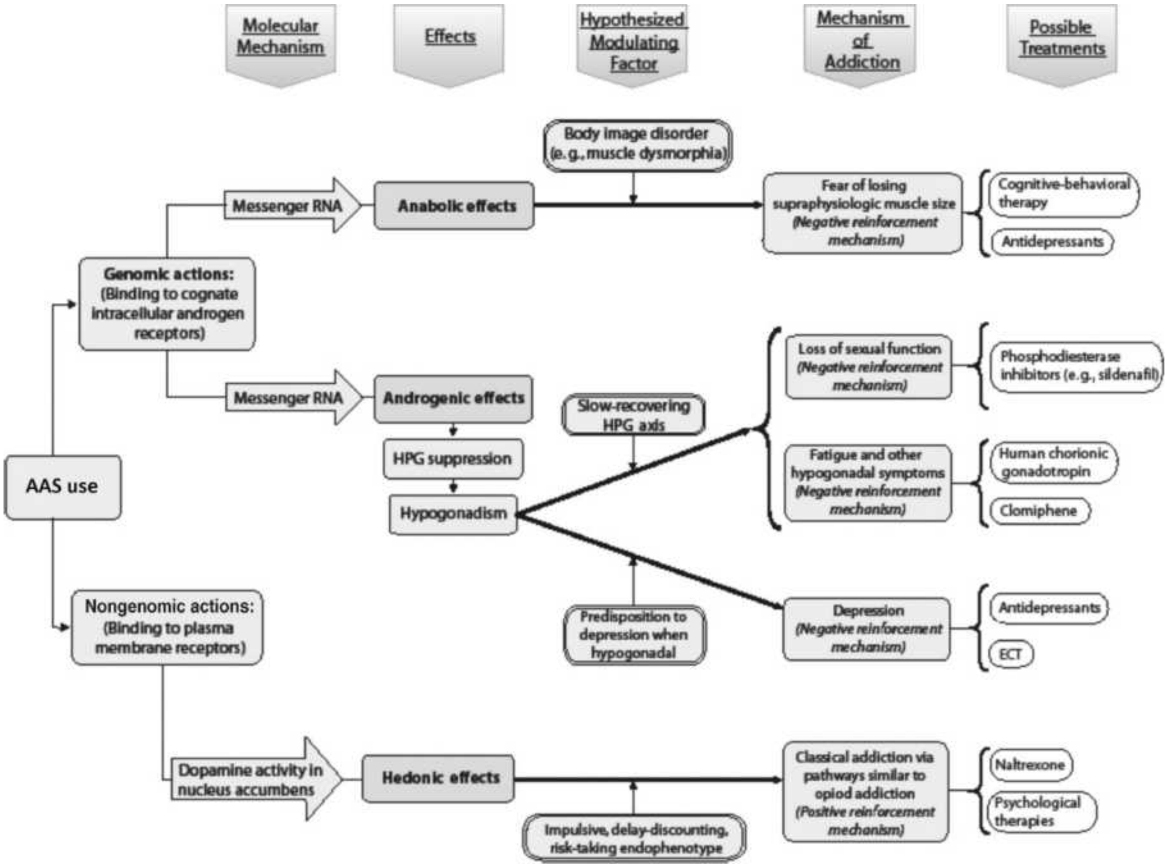
Molecular Mechanism
Effects
Hypothesized Modulating Factor
Mechanism of Addiction
Possible Treatments
Body image disorder (e.g., muscle dysmorphia)
Messenger RNA
Anabolic effects
Fear of losing supraphysiologic muscle size (Negative reinforcement mechanism)
Cognitive-behavioral therapy
Antidepressants
Genomic actions: (Binding to cognate intracellular androgen receptors)
Loss of sexual function (Negative reinforcement mechanism)
Phosphodiesterase inhibitors (e.g., sildenafil)
Messenger RNA
Androgenic effects
Slow-recovering HPG axis
HPG suppression
Fatigue and other hypogonadal symptoms (Negative reinforcement mechanism)
Human chorionic gonadotropin
AAS use
Hypogonadism
Clomiphene
Nongenomic actions: (Binding to plasma membrane receptors)
Predisposition to depression when hypogonadal
Depression (Negative reinforcement mechanism)
Antidepressants
ECT
Dopamine activity in nucleus accumbens
Hedonic effects
Classical addiction via pathways similar to opiod addiction (Positive reinforcement mechanism)
Naltrexone
Psychological therapies
Impulsive, delay-discounting, risk-taking endophenotype

FIGURE 11.11–4. A model showing three hypothesized pathways by which anabolic steroid dependence may develop, together with potential treatment strategies to address each of these pathways. HPG axis, hypothalamic–pituitary–gonadal axis; ECT, electroconvulsive therapy. (Reprinted from Kanayama G, Brower KJ, Wood RI, Hudson JI, Pope HG Jr. Treatment of anabolic-androgenic steroid dependence: emerging evidence and its implications. *Drug Alcohol Depend.* 2010;109(1–3):6–13. Copyright © 2010, with permission from Elsevier. doi: 10.1016/j.drugalcdep.2010.01.011.)

Third, AAS have been shown to possess hedonic effects, as evidenced by animal studies. For example, rats and mice will exhibit conditioned place preference to AAS, preferring to go to an environment where they have previously received AAS. Also, male hamsters will self-administer testosterone even to the point of death, and this self-administration can be blocked by treatment with opioid antagonists. These data suggest that, in addition to their genomic actions described above (binding to cognate intracellular androgen receptors), AAS also likely possess nongenomic effects mediated by binding to plasma membrane receptors. These nongenomic effects may perhaps increase dopaminergic activity in the nucleus accumbens, leading to hedonic and reinforcing properties. The degree of these hedonic effects in humans is difficult to quantitate, although there is no doubt that some AAS users anecdotally describe feelings of increased self-confidence, invincibility, and sometimes euphoria. Thus, it seems likely that hedonic effects combine together with the previous two pathways to contribute to the evolution of AAS dependence in some individuals.

In another recently proposed model of AAS dependence, the roles of body-image disturbance and neuroendocrine factors are similarly acknowledged. This model postulates that AAS use and exercise initially produce desirable secondary gains (e.g., muscle gains and decreased body fat), but these behaviors put pressure on the homeostatic system as a result of increased metabolic/musculoskeletal allostatic load. This combines with increased hormonal allostatic load to contribute to physiologic addiction, characterized in part by use of ancillary substances to reduce the side effects of allostatic load.

Course and Prognosis

No longitudinal studies of dependent AAS users have been published to date, in part because AAS-dependent individuals rarely seek clinical treatment. One study estimated that AAS users accounted for less than 0.1% of patients entering treatment programs for substance use disorders. The low number of documented treatment cases is striking when it is considered that among the 2.9 to 4.0 million American men who have used AAS, perhaps 1 million have experienced AAS dependence.

Treatment

Few descriptions and no controlled trials have been published of treatment for AAS-related disorders. Given the lack of published treatment studies, perhaps the best guidelines for approaching AAS treatment are those suggested in Figure 11.11–4, which summarizes each of the three proposed pathways to AAS dependence discussed above. First, body image concerns, such as muscle dysmorphia, may improve with cognitive behavioral therapies. Although the authors are aware of no formal trials of cognitive behavioral therapy in muscle dysmorphia per se, a growing literature suggests that cognitive behavioral approaches are effective in other forms of body dysmorphic disorder, and thus may also be helpful in muscle dysmorphia. Individuals able to address cognitive distortions about their muscularity may be more successful in managing anxiety about loss of muscle mass from discontinuing AAS.

Another modality of demonstrated efficacy for body dysmorphic disorder is treatment with serotonergic antidepressants. Although the authors are not aware of any studies with these drugs in muscle dysmorphia per se, it appears likely that they would show efficacy in such individuals and might be of value in AAS users showing prominent and refractory obsessions about body image.

Second, it is imperative that patients be assessed and treated for AAS-induced hypogonadism. Men who experience loss of libido, erectile dysfunction, and possible major depression during AAS withdrawal will be strongly tempted to resume AAS use if the symptoms are not appropriately treated by an experienced endocrinologist. Typically, such patients will require initial treatment with physiologic doses of exogenous testosterone after they stop using supraphysiologic doses of AAS. Then, to restore HPT function, treatment typically involves use of clomiphene, which stimulates the pituitary to resume production of LH and FSH. Simultaneously, patients can be treated with HCG to stimulate the testis to resume production of testosterone and spermatozoa. Then exogenous testosterone can gradually be tapered as the patient's own HPT axis begins to reawaken. Patients may also require treatment with phosphodiesterase inhibitors such as sildenafil to treat erectile dysfunction. Importantly, as noted earlier, hypogonadism may be prolonged and refractory in some patients, requiring persistent and aggressive treatment by an endocrinologist. In addition, some individuals will develop major depression during AAS withdrawal and require treatment with antidepressant medications or even ECT (see below).

Finally, given the evidence for the hedonic effects of AAS, individuals may experience craving and other withdrawal effects comparable to those experienced by individuals attempting to achieve abstinence from classical drugs. It follows that dependent AAS users may benefit from psychological treatments of established efficacy in the treatment of dependence in other substance use disorders such as opioid use disorder. Such treatments might include, for example, contingency management, behavioral couples therapy or behavioral family counseling, and supportive–expressive therapy. Couples therapy may be particularly indicated in cases where women are reporting distress about living with an AAS-dependent male partner.

ANABOLIC STEROID–INDUCED MOOD DISORDERS

Hypomanic or Manic Episodes

The effects of AAS on mood are highly variable, with a majority of individuals showing only modest mood changes or no mood changes at all, even when taking markedly supraphysiologic doses of AAS. However, some individuals show substantial irritability, aggressiveness, hypomania, and frank mania associated with AAS use, and these psychiatric effects represent a significant public health issue associated with these drugs. Although AAS users have long recognized that these drugs could cause syndromes of anger and irritability, these syndromes were little recognized in the scientific literature until the late 1980s and 1990s. Since then, numerous observational field studies have described individuals who exhibited prominent hypomanic or even manic symptoms during AAS use. In a recent review of 12 such studies, 11 reported at least some such effects in association with AAS use, with an apparent association between the dosage of AAS used and the frequency of psychiatric effects. Specifically, in individuals using the equivalent of 300 mg of testosterone a week or less, psychiatric effects appear rare; in individuals taking intermediate dosages, between 300 and 1,000 mg of testosterone equivalent per week, mood syndromes appear more frequent, albeit still uncommon; and in individuals taking the equivalent of more than 1,000 mg a week, mood syndromes become quite common and are occasionally severe.

The association of AAS dosage with psychiatric effects is illustrated by examining individual studies. At the low end of the dosage range, for example, one study of 53 men examined individuals who were using a mean dose of only 318 mg of testosterone or equivalent a week. The men in this study exhibited essentially no psychiatric changes during use of

AAS. By comparison, another study examined 71 men and 6 women who were using a mean dose estimated as somewhat more than 500 mg a week. The authors of this study noted that one (1.3%) of the subjects developed mania and six (7.8%) developed hypomania. One subject (1.3%) described major depressive disorder during AAS use, and five (6.5%) subjects reported major depressive disorder in association with AAS withdrawal. Three (3.9%) of the 77 users in this study reported a suicide attempt during a period of withdrawal from AAS use. At the high end of the dosage range, a third study examined 39 men and 2 women who had used a mean of 750 mg per week. In this study, five (12%) of the subjects reported a manic syndrome and five (12%) reported psychotic symptoms at some time in association with AAS use. Interestingly, among the eight subjects reporting manic or psychotic symptoms or both in this study, the average weekly dose was about 900 mg.

Observational field studies, however, are subject to a number of methodologic limitations common to retrospective studies of this nature. First, subjects may be vulnerable to recall bias because they are being asked to describe psychiatric symptoms that may have occurred well in the past. Second, subjects are generally taking drugs of unknown potency or authenticity obtained on the illicit market. Thus, assumptions regarding the nature and dosage of the drugs are necessarily crude. Third, selection bias may influence which subjects present for interviews in observational studies: subjects who have experienced prominent psychopathology in association with AAS use might be either more likely or less likely to participate. Fourth, expectational effects and influences from the subculture of the gymnasium may greatly influence subjects' reactions to AAS. For example, an individual lifting weights regularly in a "hardcore" gymnasium might be primed to expect increases in his irritability and aggression with AAS use, even if such changes do not actually occur.

Laboratory studies using AAS in normal volunteers avoid most of these limitations but are subject to problems of their own. Specifically, most laboratory studies, including endocrinologic investigations, physiologic studies, and studies using AAS to treat various medical conditions, have used very modest doses of testosterone or other AAS, rarely exceeding the equivalent of 300 mg of testosterone a week. Furthermore, such studies have rarely attempted to measure psychiatric effects systematically. Not surprisingly, these studies have generally failed to note prominent psychiatric changes in their subjects. As a result, some such studies have erroneously concluded that AAS lack psychiatric effects because they have inappropriately extrapolated from modest doses used under laboratory conditions and have failed to consider that illicit AAS users use far larger doses in the field. This error is similar to that of many earlier negative studies of the effects of AAS on muscle growth, again because the studies typically tested doses far lower than those actually employed by illicit AAS users.

We are aware of only four laboratory studies in which normal male volunteers were given dosages of at least 500 mg a week of testosterone or equivalent under placebo-controlled, double-blind conditions and where psychiatric measures were reported. These studies represent the only controlled investigations that have used doses of AAS more closely approximating those used by actual individuals using AAS—although even these laboratory studies still do not approach the massive doses of AAS employed by some actual illicit users. Of 109 men receiving such dosages in these four studies combined, 5 exhibited prominent hypomanic or manic syndromes during AAS exposure. These subjects displayed levels of symptomatology that might be expected to seriously impair normal social or occupational functioning; one study subject, while receiving methyltestosterone 240 mg per day under blinded conditions, even requested that he be placed in seclusion on the ward where the study was conducted. Notably, however, the great majority of the men in these four studies exhibited few behavioral changes with high dosages of AAS. Thus, it appears that most individuals taking AAS, even at fairly high dosages, will experience few psychiatric effects, whereas an occasional individual may develop a severe manic or hypomanic reaction. At present, the mechanism of these reactions is unknown; they may represent an idiosyncratic phenomenon unrelated to premorbid psychiatric function.

Figure 11.11–5 shows changes in ratings of manic symptoms in three subjects who received testosterone cypionate in dosages up to 600 mg a week versus inert placebo, each for periods of 6 weeks in a double-blind, crossover design. Subjects varied widely in their response to testosterone. The subject on the left side of the figure exhibited virtually no change in manic ratings during either the testosterone or placebo periods; the subject in the middle displayed a modest rise in manic symptoms in association with testosterone, but not sufficient to significantly affect his social or occupational functioning. By contrast, the subject on the right displayed a marked manic reaction, approaching the range of patients hospitalized for the treatment of manic episodes, requiring the investigators to discontinue testosterone administration 1 week earlier than planned. The symptoms of mania resolved promptly within 2 weeks following the discontinuation of testosterone.

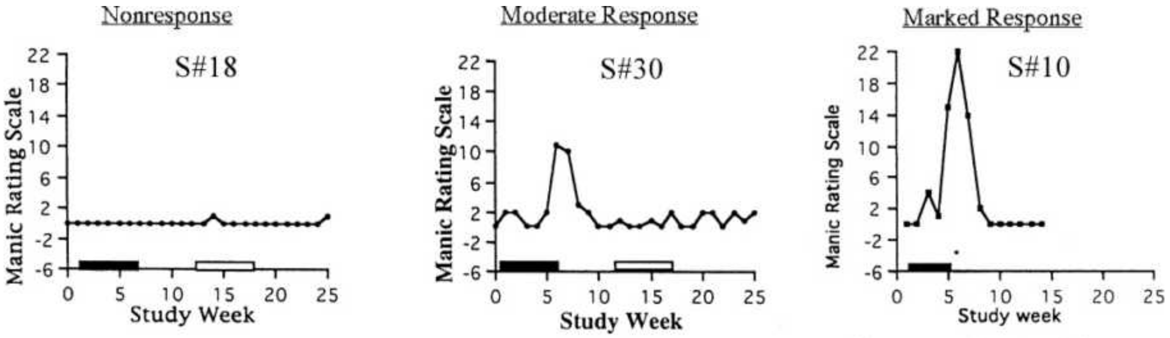

* Testosterone discontinued after week 5 due to manic symptoms

FIGURE 11.11–5. Examples of nonresponse, moderate response, and marked response to testosterone versus placebo on the Young Manic Rating Scale. The *black bar* represents the period of testosterone cypionate (Depo-Testosterone) administration in each subject. Testosterone was administered in gradually increasing doses (150 mg per week for the first 2 weeks, 300 mg per week for the next 2 weeks, and 600 mg per week for the final 2 weeks of the period). The *white bar* represents the period of placebo administration. A 6-week washout period was interposed between the two treatment periods to minimize carryover effects. (Adapted with permission from Pope HG, Kouri EM, Hudson JI. Effects of supraphysiologic doses of testosterone on mood and aggression in normal men: a randomized controlled trial. *Arch Gen Psychiatry.* 2000;57(2):133–140. doi:10.1001/archpsyc.57.2.133. Copyright © 2000 American Medical Association. All rights reserved.)

It is difficult to extrapolate from these findings to estimate the frequency of prominent hypomanic or manic reactions among AAS users in the field. Illicit users may employ higher dosages than could ethically be administered in the laboratory; such users may take several different AAS together (“stacking”); they may use alcohol or other drugs while taking AAS; and, of course, they do not screen themselves with the same care as would be the case in a laboratory study. Finally, some users may become more sensitized to the psychiatric effects of AAS after using them for multiple courses over time; since most of the subjects in available laboratory studies were AAS-naive prior to the treatment, they may not be representative of actual AAS users in the field. For all these reasons, the crude rate of approximately 5% estimated above may represent a lower bound for the true prevalence of AAS-induced hypomanic and manic syndromes in the overall population of illicit users.

Major Depressive Episodes

AAS-induced depressive syndromes have generally not been reported in laboratory studies but have been documented in field studies. Several reports of completed suicides associated with AAS withdrawal have also appeared. Usually, it appears that a brief and self-limited syndrome of depression occurs upon AAS withdrawal, probably as a result of the depression of the HPT axis following exogenous AAS administration. Such syndromes might be expected to occur more often in the field than in laboratory settings because men typically take AAS for longer intervals than would occur in the laboratory, resulting in more pronounced neuroendocrine depression and shrinkage of the testes.

Not all depressive syndromes of this nature, however, are self-limited. Some reports have described episodes of depression that continued for months after discontinuing AAS and which required treatment with fluoxetine, tricyclic antidepressant agents, or ECT. Again, however, given the uncontrolled and retrospective nature of these observations, it is difficult to judge the true frequency of such syndromes in the field.

Mixed Episodes

Although the authors are not aware of specific published descriptions of mixed affective episodes associated with AAS use, we have witnessed many such cases in our anecdotal experience. Mixed episodes appear to be particularly frequent in individuals who have developed AAS dependence and are typically using AAS regularly with only brief drug-free intervals. Prolonged and relatively continuous AAS use of this type typically produces profound HPT suppression so that the individual becomes markedly hypogonadal and often seriously depressed promptly upon stopping AAS. Therefore, he will often quickly resume AAS to self-treat the depression, only to experience a recurrence of depression when he next decreases the dose or attempts to discontinue the drugs. In the authors’ experience, such individuals sometimes develop “brittle” mood disorders, with a mixture of manic and depressive symptoms, or sharp swings between manic and depressive symptoms, often associated with prominent impulsivity, sometimes leading to suicide attempts. In some instances, such a suicide attempt may be the first event that brings the AAS user to the attention of clinicians.

Violence

A possible serious consequence of AAS-induced mood disorders, especially manic, hypomanic, or mixed episodes, may be violent or even homicidal behavior. Several published reports have anecdotally described individuals with no apparent history of psychiatric disorder, no criminal record, and no history of violence, who committed violent crimes, including murder, while under the influence of AAS. In a number of cases, AAS use has been cited in criminal trials as a possible mitigating factor in the defense of such individuals. One study of 133 sequential, newly incarcerated male felons found that 3 (2.3%) reported committing violent crimes while under the influence of AAS. Another recent paper mentions six cases of AAS-induced criminal behavior seen by the authors, including three homicides and three violent assaults. The authors reported that the men in all cases apparently experienced psychotic symptoms and that the mental status of all six perpetrators cleared within 2 months after discontinuing AAS; all reported a clear memory of the violent act and of their delusional thinking at the time. Forensic researchers in Sweden have described several similar cases of AAS users, many with no prior history of criminality or violence, who committed violent offenses, including homicide, in association with AAS use.

Finally, another study suggests that violence associated with AAS is frequently directed at women. This study compared 23 male AAS users with 14 male nonusers, using the Conflict Tactics Scales to assess users' relationships with their girlfriends or spouses. Significant differences emerged on several of these scales between users on drugs and nonusers and between users on drugs and the same users when off drugs.

A Swedish forensic research group has recently used epidemiologic techniques to assess possible links between AAS use and violence. For example, these investigators found that deceased AAS users were significantly more likely to have died from homicide or suicide than a comparison group of deceased amphetamine and/or heroin users. These investigators also found that deceased AAS users, during the last 3 years of their lives, were more likely to have been sentenced for violent crimes but less likely to have been sentenced for crimes against property than deceased individuals who tested positive for other drugs but negative for AAS. Another study from the same group showed that prisoners testing positive for AAS were significantly less likely to have been convicted of a crime against property but significantly more likely to have been convicted of a weapons offense than AAS-negative prisoners (using an analysis that excluded cases referred from substance abuse centers). Although these findings do not permit causal inferences in specific cases, they appear consistent with the findings described in previous paragraphs.

A recent review describing the forensic aspects of AAS-induced violence or criminality has noted that there is good scientific consensus that AAS are capable of producing severe reactions of this type in a small minority of individual users but has noted that it is more difficult to establish causality in a specific individual case. To do so, one must be able to establish that the individual was actually using AAS at the time of the violent or criminal acts and, more importantly, that he did not display any comparable behavior prior to AAS use or following AAS cessation. The authors also noted anecdotally that men with AAS-induced violent or criminal behavior displayed a number of stereotypic features, including a sense of invincibility, obsessional features (e.g., being unable to dismiss an angry preoccupation), and an element of anosognosia (being unaware that one is behaving abnormally).

ANABOLIC STEROID–INDUCED PSYCHOTIC DISORDER

Psychotic symptoms are rare in association with AAS use but have been described in a few cases, primarily in individuals who were using the equivalent of more than 1,000 mg of testosterone a week. Usually, these symptoms have consisted of grandiose or paranoid delusions, generally occurring in the context of a manic episode, although occasionally occurring in the absence of a frank manic syndrome. In most cases, the reported psychotic symptoms have disappeared promptly (within a few weeks) after the discontinuation of the offending agent, although temporary treatment with antipsychotic agents was sometimes required. As with manic reactions to AAS, the mechanism for these seemingly idiosyncratic psychotic reactions is unknown. It is interesting to note, however, that the chemically related family of hormones, corticosteroids, also produces idiosyncratic manic and psychotic symptoms in occasional individuals while creating few psychiatric effects in the majority of patients. The mechanism of action of these idiosyncratic effects with corticosteroids is also unknown.

ANABOLIC STEROID–RELATED DISORDER NOT OTHERWISE SPECIFIED

Anxiety Disorders

Occasional anecdotal reports have described decreases in symptoms of anxiety disorders, such as panic disorder and social phobia, during AAS use and exacerbations of these disorders during AAS withdrawal. Conversely, some individuals report increases in anxiety symptoms while taking AAS. However, no systematic data are available on these entities.

Associated Substance Abuse or Dependence

As mentioned earlier, AAS use may be associated with the use of other ergogenic drugs, especially among the minority of AAS users who are competitive athletes. Among the larger number of noncompetitive AAS users, extensive use of illicit drugs is common, as described in detail above. Indeed, individuals with a history of AAS use may first come to clinical attention when they require treatment for dependence upon alcohol or illicit drugs such as opiates, cocaine, or cannabis. For example, in one recent study of 223 men admitted to an addiction treatment unit, 29 were found to have a past history of AAS use, but this history was documented in physicians’ admission evaluations in only 4 of the cases. Nineteen of the 29 men were interviewed in detail, and 5 were found to exhibit a prior history of AAS dependence. As this finding illustrates, AAS use is often undetected by clinicians, despite its potential importance in causing the various psychiatric syndromes described in the above sections.

Mr. A is a 26-year-old, single White man. He is 69 inches tall and weighs 204 pounds with a body fat of 11%. He reports that he began lifting weights at age 17, at which time he weighed 155 pounds. About 3 years after beginning his weightlifting, he began taking AAS, which he obtained through a friend at his gymnasium. His first "cycle" (course) of AAS, lasting for 9 weeks, involved methandienone, 30 mg orally a day, and testosterone cypionate, 600 mg intramuscularly a week. During these 9 weeks, he gained 20 pounds of muscle mass. He was so pleased with these results that he took five further cycles of AAS over the course of the next 6 years. During his most ambitious cycle, approximately 1 year ago, he used testosterone cypionate, 600 mg per week; nandrolone decanoate, 400 mg a week; stanozolol, 50 mg a day; and oxandrolone, 10 mg a day.

During each of the cycles, Mr. A noted euphoria, irritability, and grandiose feelings. These symptoms were most prominent during his most recent cycle, during which he felt "invincible." During this cycle, he also noted a decreased need for sleep, racing thoughts, and a tendency to spend excessive amounts of money. For example, he impulsively purchased a $2,700 stereo system when he really could not afford to spend more than $500. He also became uncharacteristically irritable with his girlfriend and on one occasion, put his fist through the side window of her car during an argument—an act inconsistent with his normally mild-mannered personality. After this cycle of AAS ended, he became mildly depressed for about 2 months.

Mr. A smoked marijuana almost daily during his last 2 years of high school and continues to smoke at least twice a week. He has also used hallucinogens on about 50 occasions and cocaine on about 200 occasions over the last 8 years. He has also used a number of drugs to lose weight in preparation for bodybuilding contests. These include ephedrine, amphetamine, triiodothyronine, and thyroxine. Recently, he has also begun to use the opioid agonist–antagonist nalbuphine intravenously to treat muscle aches from weightlifting. He also uses oral opioids, such as controlled-release oxycodone, at least once a week. He uses oral opioids sometimes to treat muscle aches, but often simply to get high. He reports that use of nalbuphine and other opioids is widespread among other AAS users of his acquaintance.

Mr. A exhibits characteristic features of muscle dysmorphia. He checks his appearance dozens of times a day in mirrors or when he sees his reflection in a store window or even in the back of a spoon. He becomes anxious if he misses even 1 day of working out at the gym and acknowledges that his preoccupation with weightlifting has cost him both social and occupational opportunities. Although he has a 48-inch chest and 19-inch biceps, he has frequently declined invitations to go to the beach or a swimming pool for fear that he would look too small when seen in a bathing suit. He is anxious because he has lost some weight since the end of his previous cycle of AAS and is eager to resume another cycle of AAS in the near future.

Mr. B is a 37-year-old personal trainer and bodybuilder. He describes a normal childhood and adolescence, although he was periodically mildly depressed. By the age of 15, he began to drink alcohol regularly, and at age 18 used cocaine several days per week for about a year. He describes "mood swings" starting in his late adolescence, but these were not severe enough to compromise his social or academic functioning. He obtained a master's degree in physical therapy from a local community college at age 21.

Starting in his late teens, Mr. B became "obsessed" with bodybuilding and his physique. At age 21, he began his first 16-week "cycle" of injectable testosterone cypionate up to 600 mg per week, plus oral stanozolol up to 25 mg per day. After stopping the cycle, he became mildly depressed, with fatigue, decreased sex drive, and a lack of enthusiasm for going to the gym. He was also distressed to notice that he regained some body fat and lost some muscle during the weeks after stopping the cycle.

Four months later, he started a second cycle of AAS with a larger "stack" of drugs, including injectable testosterone cypionate, up to 600 mg per week; injectable nandrolone decanoate 400 mg per week; and oral oxymetholone, 50 mg per day. Although Mr. B had intended to stop the cycle after 16 weeks, he was so gratified by his muscle gains that he decided to extend the cycle to 30 weeks. Upon stopping, he again experienced depression, despite injecting illicit HCG in an attempt to prevent hypogonadism.

Mr. B continued to take cycles of AAS regularly for the next 5 years; he estimates that he was on the drugs "about two-thirds of the time" from ages 22 to 27. During his cycles, he felt "high" and very sexually active; between the cycles, he felt increasingly depressed and eager to resume taking AAS. At the age of 27, Mr. B got married, and his wife was eager to have a child. An endocrinologist found that Mr. B was sterile as a result of the suppression of sperm cell production due to his AAS use. To please his wife, Mr. B stopped AAS entirely and managed to remain off of AAS for more than 2 years. His sperm count finally rose to the normal range 7 months after stopping AAS; soon after, his wife became pregnant. The couple had a son, followed a year and a half later by a daughter. Mr. B used virtually no illicit drugs, and only occasional alcohol, during this period.

Shortly after his wife became pregnant with the second child, Mr. B resumed using AAS despite his wife's objections. Once back on AAS, he felt self-confident and proud of his muscle gains; he became sexually promiscuous. He reports that his promiscuity contributed to his divorce from his wife 2 years later. Each time upon stopping AAS, he became markedly depressed; he was started on fluoxetine during one such period of AAS withdrawal at the age of 31. Fluoxetine was initially quite helpful, but it did not deter him from returning to a cycle of AAS within a few months.

Over the next 6 years, Mr. B continued to take AAS about "80% of the time," by his account—but from ages 35 to 37, he used AAS continuously, remaining on a modest dose of testosterone cypionate even during the spaces between the cycles. He described increasingly "brittle" mood swings, with periods of euphoria, irritability, and increased self-confidence alternating with sudden dips into depression with suicidal ideation. At times, Mr. B described a mixture of affective symptoms where he would experience euphoria, irritability, and suicidal thoughts almost simultaneously. Over the course of these years, his physician tried several other antidepressants (citalopram, venlafaxine, and duloxetine), plus several concomitant medications for mood stabilization (lithium carbonate, valproate, and lamotrigine), with inconsistent benefit—partially because Mr. B would often become noncompliant. Also, Mr. B began drinking alcohol almost daily and began to use increasing amounts of controlled-release oxycodone to self-treat his dysphoric moods. Mr. B also became increasingly impulsive and on one occasion, took a substantial overdose of citalopram, necessitating a 3-day hospitalization in a psychiatric unit.

Following this hospitalization, Mr. B became more consistently compliant with a regimen of duloxetine and lithium carbonate. He continued to be highly focused on weightlifting and his body appearance. However, he still used oxycodone frequently and continued to complain of sharp "dips" of severe depression. During one of these "dips," shortly after tapering down from high-dose AAS to his usual "maintenance" level of testosterone cypionate, he made a second, nearly successful suicide attempt with a combination of illicitly obtained oral opiates. He was again hospitalized, treated with ECT, and responded well. A cardiac evaluation during this hospitalization, using Doppler echocardiography, revealed evidence of dilated cardiomyopathy judged likely attributable to AAS.

Mr. B has resolved not to resume taking AAS, but 6 months after hospital discharge, he is already contemplating going back on a low dose of AAS because he is distressed by his deteriorating physique.

Mr. B's case exemplifies several AAS-related disorders, including apparently AAS-induced bipolar disorder, AAS withdrawal, and AAS dependence. In addition to physiologic withdrawal, Mr. B met several other diagnostic criteria for AAS dependence: taking more AAS than originally intended, resuming use after several attempts to stop completely, requiring increasing doses of AAS (tolerance), and continuing use despite knowledge of adverse consequences such as divorce. Another important aspect of this case was his concurrent use of alcohol, opioids, and cocaine. All four classes of substances share a final common pathway of dopamine augmentation; in addition, alcohol, opioids, and AAS have effects on the endogenous opioid system; and alcohol and AAS both also act as GABA receptor modulators. Taking these drugs in combination typifies the loss of control commonly seen with addiction but also may reflect an attempt to enhance the rewarding effects of AAS (dopamine and opioid systems) and analgesic aspects of AAS (opioid system) while minimizing some of the hyperactivating effects of AAS (via the antianxiety/sedative effects of GABA and opioid systems).

FURTHER READINGS

Baggish AL, Weiner RB, Kanayama G, et al. Cardiovascular toxicity of illicit anabolic-androgenic steroid use. *Circulation*. 2017;135:1991–2002.

*Bhasin S, Storer TW, Berman N, et al. The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. *N Engl J Med*. 1996;335:1–7.

Bjørnebekk A, Walhovd KB, Jørstad ML, Due-Tønnessen P, Hullstein IR, Fjell AM. Structural brain imaging of long-term anabolic-androgenic steroid users and nonusing weightlifters. *Biol Psychiatry*. 2017;82:294–302.

*Brower KJ. Anabolic steroid abuse and dependence in clinical practice. *Phys Sportsmed*. 2009;37:131–140.

Carrillo M, Ricci LA, Melloni RH. Glutamate-vasopressin interactions and the neurobiology of anabolic steroid-induced offensive aggression. *Neuroscience*. 2011;185:85–96.

Estrada M, Varshney A, Ehrlich BE. Elevated testosterone induces apoptosis in neuronal cells. *J Biol Chem*. 2006;281(35):25492–25501.

*Frye CA. Some rewarding effects of androgens may be mediated by actions of its 5alpha-reduced metabolite 3alpha-androstanediol. *Pharmacol Biochem Behav*. 2007;86:354–367.

Givens ML, Deuster PA, Kupchak BR. CHAMP symposium on androgens, anabolic steroids, and related substances: what we know and what we need to know. *Mil Med*. 2016;181(7):680–686.

Graham MR, Ryan P, Baker JS, et al. Counterfeiting in performance- and image-enhancing drugs. *Drug Test Anal*. 2009;1(3):135–142.

Gruber AJ, Pope HG Jr. Psychiatric and medical effects of anabolic-androgenic steroid use in women. *Psychother Psychosom*. 2000;69:19–26.

Hartgens F, Kuipers H. Effects of androgenic-anabolic steroids in athletes. *Sports Med*. 2004;34(8):513–554.

Hildebrandt T, Lai JK, Langenbucher JW, Schneider M, Yehuda R, Pfaff DW. The diagnostic dilemma of pathological appearance and performance enhancing drug use. *Drug Alcohol Depend*. 2011;114(1):1–11.

Ip EJ, Barnett MJ, Tenerowicz MJ, Perry PJ. The Anabolic 500 survey: characteristics of male users versus nonusers of anabolic-androgenic steroids for strength training. *Pharmacotherapy*. 2011;31(8):757–766.

Kanayama G, Brower KJ, Wood RI, Hudson JI, Pope HG Jr. Treatment of anabolic-androgenic steroid dependence: emerging evidence and its implications. *Drug Alcohol Depend*. 2010;109:6–13.

Kanayama G, Hudson JI, Pope HG Jr. Culture, psychosomatics and substance abuse: the example of body image drugs. *Psychother Psychosom*. 2012;81(2):73–78.

Kanayama G, Kean J, Hudson JI, Pope HG Jr. Cognitive deficits in long-term anabolic-androgenic steroid users. *Drug Alcohol Depend*. 2013;130:208–214.

Kaufman MJ, Kanayama G, Hudson JI, Pope HG Jr. Supraphysiologic-dose anabolic-androgenic steroid use: a risk factor for dementia? *Neurosci Biobehav Rev*. 2019;100:180–207.

Klotz F, Garle M, Granath F, Thiblin I. Criminality among individuals testing positive for the presence of anabolic androgenic steroids. *Arch Gen Psychiatry*. 2006;63:1274–1279.

Kouri EM, Pope HG Jr, Katz DL, Oliva P. Fat-free mass index in users and nonusers of anabolic-androgenic steroids. *Clin J Sport Med*. 1995;5:223–228.

Pärssinen M, Seppälä T. Steroid use and long-term health risks in former athletes. *Sports Med*. 2002;32:83–94.

Pope HG Jr, Kanayama G, Hudson JI. Risk factors for illicit anabolic-androgenic steroid use in male weightlifters: a cross-sectional cohort study. *Biol Psychiatry*. 2012;71:254–261.

*Pope HG Jr, Kouri EM, Hudson JI. The effects of supraphysiologic doses of testosterone on mood and aggression in normal men: a randomized controlled trial. *Arch Gen Psychiatry*. 2000;57:133–140.

*Pope HG Jr, Wood RI, Rogol A, Nyberg F, Bowers L, Bhasin S. Adverse health consequences of performance-enhancing drugs: an Endocrine Society scientific statement. *Endocr Rev*. 2014;35:341–375.

Rahnema CD, Crosnoe LE, Kim ED. Designer steroids—over-the-counter supplements and their androgenic component: review of an increasing problem. *Andrology*.

2015;3(2):150–155.

Rasmussen JJ, Selmer C, Østergren PB, et al. Former abusers of anabolic-androgenic steroid exhibit decreased testosterone levels and hypogonadal symptoms years after cessation: a case-control study. *PLoS One*. 2016;11(8):e0161208.

Rubinow DR, Schmidt PJ. Androgens, brain, and behavior. *Am J Psychiatry*. 1996;153(8): 974–984.

Skarberg K, Nyberg F, Engstrom I. Multisubstance use as a feature of addiction to anabolic-androgenic steroids. *Eur Addict Res*. 2009;15(2):99–106.

Su TP, Pagliaro M, Schmidt PJ, Pickar D, Wolkowitz O, Rubinow DR. Neuropsychiatric effects of anabolic steroids in male normal volunteers. *JAMA*. 1993;269:2760–2764.

*Thiblin I, Petersson A. Pharmacoepidemiology of anabolic androgenic steroids: a review. *Fundam Clin Pharmacol*. 2005;19(1):27–44.

*Wood RI. Anabolic-androgenic steroid dependence? Insights from animals and humans. *Front Neuroendocrinol*. 2008;29(4):490–506.

12

Schizophrenia and Other Psychotic Disorders

12.1 Introduction and Overview

Carol A. Tamminga, M.D.

Schizophrenia is arguably one of the most puzzling yet disabling of all brain diseases, with its severe and persistent psychotic manifestations accompanied by variable cognitive dysfunction and profound psychosocial impairment. Its disease onset, at least of the psychotic manifestations, occurs at an unusual time of life in late adolescence/early adulthood. Its symptoms are often undetectable in early life, are at their worst during child-bearing years, and often improve, to a degree, with aging. People with psychotic illnesses, in general, suffer greatly throughout life and medical need worldwide continues to be very high. Fortunately, in the pages of this section, the broad aspects of this condition are described and discussed, as far as they are known, by experts in the field.

Substantial progress in understanding and treating schizophrenia will depend on discovering the neural mechanisms of the elements of the condition, one of those being the mechanism for psychosis. This understanding is necessary for developing drug targets, novel therapeutics, and biologically based disease definitions. It is hard to conceive of the exact neural mechanisms whose dysfunction could interfere so extensively with such broad cognitive functions as does psychosis. This section lays out current perspectives of schizophrenia presentation, treatment, and biology that include much of what is known about its disease manifestations today and brain dysfunctions in the illness. What provides clear hope of progress for the future is not only the quality of the tools available to study mammalian brain function but, in particular, the wealth of basic neurobiology available and the extent to which this understanding of basic brain mechanisms translates to human diseases. In this section, the curious reader will find the facts of psychosis presentation and the sum of the manifestations and biomarkers that any model needs to consider in formulating a disease model. Of course, blatant heterogeneity exists in the different psychotic illnesses, with nearly as much heterogeneity within a diagnosis as between, along with a lack of distinguishing biologic characteristics. Several scientists represented in this section speculate on the use of alternative strategies for classification, attempting to increase biologic homogeneity within disease-like clusters to advantage mechanistic discovery.

It is always poignant to review the long history of psychotic illness, including its most severe variant, schizophrenia. Descriptions of schizophrenia can be found in works of literature from the earliest human times. Schizophrenia-like symptoms are described in individuals labeled as seers and prophets, as well as witches and devils. The perspective of schizophrenia being a human brain disease did not develop until the 19th century. Then, “humane” treatment was prescribed for “insanity,” an approach that advocated care, protection, and human understanding for those afflicted. It was not until the middle of the 20th century that antipsychotic medications became widely available. Now in the early 21st century, two, maybe three, generations of antipsychotic medications were developed, several known risk genes discovered, an evolving anatomy discussed, and several regional human brain-based models for symptoms manifestations of cognitive dysfunction (prefrontal cortex) and psychosis (hippocampus). Meanwhile, the focus of inquiry into the illness has moved from conventional diagnostic criteria to dimensions of dysfunction, and a fundamental understanding of the neural basis of complex functional brain diseases has developed. Therefore, it might be optimistically predicted that it will eventually be possible to describe the genes, circuits, cellular and molecular mechanisms, and rational treatments for psychosis manifestations over the next several decades.

The articulation of a modern biologic hypothesis for schizophrenia began with dopamine. The pivotal observation that drugs that diminish dopamine-mediated neurotransmission in the brain reduce psychotic symptoms was made only in the mid-20th century. The first antipsychotic drug (chlorpromazine) was discovered serendipitously but with a strong clinical rationale and astute observation. The experimental result that implicated the action of chlorpromazine connected to dopamine receptor blockade was a hypothesis that fueled decades of drug discovery and was the basis for a related etiologic hypothesis that dopamine was indeed the core pathophysiology of schizophrenia. This speculative extension of a well-supported treatment hypothesis has been harder to prove and establish as the basis for schizophrenia pathophysiology, although ideas are still being tested. Although data have consistently supported the dopamine mechanism of antipsychotic drug action, support for the dopamine hypothesis of pathophysiology has been inconsistent.

At the same time, understanding the pharmacology and basic biology of human cognition and its dysfunctions has evolved into a very intricate picture, one unlikely to be explained simply as overactivity of dopamine. A component of cognition like declarative memory, which has been intensively studied since Henry Molaison (known in the clinical literature as H. M.) lost declarative memory function, is understood now to be served by alterations in synaptic strength, postsynaptic calcium interactions with receptor trafficking, and epigenetic remodeling in hippocampus and the memory neocortex. An alteration in this process of memory physiology could be the basis of a model of psychosis. Many speculations have emerged that multiple manifestations of psychotic illness may be associated with distinct pathophysiologies, for example, one for psychosis and another for cognitive dysfunction. Most would agree that psychosis itself will have multiple etiologies, much like “fever.” Anyone interested in the pathophysiologic basis of psychiatric conditions will have to understand normal brain processes already being articulated within neuroscience laboratories worldwide. Full attention will have to be directed to these discoveries in basic neuroscience to understand modern articulations of brain-based pathophysiologies for psychosis and cognitive dysfunction.

In clinical research on schizophrenia, there has been a willingness to abandon conventional diagnostic categories, experimentally and hypothetically, to examine alternate views of pathophysiology, etiology, and treatment. This has focused on the pathophysiology of *dimensions* of dysfunction for postulating mechanisms and exploring treatments. Not that the dimension will be the final explanatory element, but that it will allow a resorting among similar disease manifestations to gain biologic homogeneity. Dimensions of normal cerebral function have been proposed, connected with known neural systems, a dysfunction or perversion that would generate psychopathology. Specifically, the independence of symptom domains such as psychosis, cognitive dysfunction, affective disturbance, and negative symptoms is now being explored in a research domain criteria (RDoC) framework. The implication of this formulation, which is now being explored, is that each of these composite symptom complexes will have its own etiology, mechanism, and selective treatment and can be treated independently and individually. Personalized medicine, as well as new disease formulations, are likely to spring from these approaches.

The field of schizophrenia practice and research needs to understand some obvious areas. First, effective diagnostic and treatment algorithms for the early detection and treatment of psychotic young adults need to be established. Effective medications are necessary, even when optimized across each individual person, but they are by no means sufficient. Psychotic illness needs broad, diverse, and comprehensive treatment, especially at its inception or when it threatens to break through. This kind of comprehensive treatment extends beyond antipsychotic medication to include cognitive remediation approaches, psychosocial treatments, behavioral therapies, family training, and community involvement. The newest study in the treatment of early psychosis, the Recovery After an Initial Schizophrenia Episode (RAISE) study, found that implementing all of these treatments early on in disease manifestations generates better outcomes sooner in young people. Treating a psychotic illness such as schizophrenia “piecemeal” generates only partial gains and falls short of the success that comes with a comprehensive plan allowing potentiated gains.

It would certainly be useful if more were known about the mechanism of action of the effective drugs used for psychosis, all of which are dopamine receptor antagonists. What actually do these antidopaminergic drugs do in the brain that quiet the storms of psychotic symptoms? This would be an approach to discovering disease pathophysiology. One of the single issues that affect psychotic illnesses most adversely is the lack of disease understanding. What the field needs is a simple, direct, and concrete presentation of the mechanism(s) of psychosis to combat stigma, direct treatment development, and on which to base behavioral and cognitive treatment approaches.

What if there were biologic formulations of psychotic diseases that were confirmed and demonstrated? How different would that be for the serious mental illnesses that psychiatrists treat? To answer this question, medical conditions that changed after full disease understanding could be examined. For example, there was once a disease called “dropsy” in the late 19th century, which was the presentation of morbid edema that progressed rapidly and caused extreme morbidity and finally death. Its determinants and treatments have been progressively understood and more precisely treated by medical discoveries throughout the 20th and 21st centuries. Dropsy can be explained by a plethora of cardiac, pulmonary, renal, and metabolic conditions, some genetic, others developmental, and others metabolic or structural. Its association with specific organ failures focused attention on a component of the dropsy cases (all phenomenologically quite similar, if not identical) on dysfunctions of the particular organ systems. Some treatments are specific for treating the diseased organ, and some medications are specific for the edema. Now edema rarely becomes morbid anymore. And our increased understanding of *prophylaxis* directed toward cardiac, metabolic, and diseases has improved health and helped avoid dropsy-like conditions. These kinds of disease structures and their determinants need to be uncovered in cognitive and affective brain diseases so that the same kind of understanding can be gained for psychosis brain health.

FURTHER READINGS

Carlsson A, Lindquist L. Effect of chlorpromazine or haloperidol on formation of 3-methoxytyramine and normetanephrine in mouse brain. *Acta Pharmacologica et Toxicologica*. 1963;20:140–145.

Clementz BA, Sweeney JA, Hamm JP, et al. Identification of distinct psychosis biotypes using brain-based biomarkers. *Am J Psychiatry*. 2016;173:373–384.

Hyman SE. Revolution stalled. *Sci Transl Med*. 2012;4(155):155cm11.

Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010;167:748–751.

Lewis DA, Curley AA, Glausier JR, Volk DW. Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. *Trends Neurosci*. 2012;35:57–67.

Li W, Ghose S, Gleason K, et al. Synaptic proteins in the Hippocampus indicative of increased neuronal activity in CA3 in schizophrenia. *Am J Psychiatry*. 2015;172:373–382.
Tamminga CA. Development in somatic treatments: introduction. *Am J Psychiatry*. 1994;151:216–219.

12.2 Phenomenology of Schizophrenia

STEPHEN F. LEWIS, B.S., M.D., RODRIGO ESCALONA, M.D., AND SAMUEL J. KEITH, M.D.

Schizophrenia continues to be defined and diagnosed as a psychiatric disorder solely based on its clinical phenomenology. Efforts to include the use of biomarkers to make the diagnosis and predict treatment response or course of illness have not been successful thus far. The concept of schizophrenia as a distinct clinical entity is a recent historical development. However, detailed descriptions of psychotic disorders can be found through centuries of written history, with early Greek physicians documenting cases that resemble schizophrenia in its current form. Psychosis only emerged as a medical condition worthy of scientific study and treatment during the 18th century in Europe. At that time, the concept of psychosis included a large undifferentiated group of disorders, including psychotic symptoms secondary to medical conditions–general paresis and syphilitic insanity, as well as primary or “functional” psychoses such as affective psychosis and schizophrenia. It would take at least another half century for a more consistent and systematic classification of psychotic disorders to appear and set the basis for modern psychiatric diagnoses.

DIAGNOSTIC CRITERIA

History

Emil Kraepelin (1856 to 1926) pioneered the psychiatric nosology that would set the foundation for the development of modern psychiatric diagnostic systems. He first separated, from the broad group of psychosis, the organic psychoses such as syphilitic insanity and general paresis, and then most importantly, he described two major groups of primary psychotic disorders: The *manic depressive psychosis* and *dementia praecox* (dementia of the young, a term coined earlier by Morel). He used the long-term prognosis and course of illness as principal features to define the two distinct categories. The assumption at that time was that all affective psychosis had a nondeteriorating course as opposed to dementia praecox, which was deemed as deteriorating and irreversible. He believed that dementia praecox was a loss of the inner unity of the activities of intellect, emotion, and volition. He included in the concept of *dementia praecox* a broad group of disorders known separately at that time, such as paranoia, catatonia, and hebephrenia, and defined them as subtypes of the same illness.

In 1911, the term *schizophrenia,* as the disease is known today, was introduced by Eugen Bleuler (1857 to 1939). He proposed the name to denote a "splitting" of psychic functions, which he considered to be the core feature of the illness. He was less concerned with the course of illness than Kraepelin had been and much more concerned with the understanding of the psychological mechanisms underlying the disease process. His description of the illness included primary and secondary symptoms. His four primary symptoms ("the Four As") were *abnormal associations, autistic behavior and thinking, abnormal affect,* and *ambivalence.* Bleuler considered the loss of harmonic coordination among thought, emotion, and behavior to be the hallmark of the illness. For Bleuler hallucinations, delusions, social withdrawal, and diminished drive were secondary manifestations of the illness that depended on the adaptive capacity of the individual and environmental circumstances. Thus, according to Bleuler, the most obvious and striking manifestations of schizophrenia were only "accessory symptoms." In addition, he saw the symptoms of schizophrenia as a continuum with normal behavior, a concept that led to major diagnostic controversies among psychiatrists around the world for decades, until the recent formalization of diagnostic criteria.

Kurt Schneider (1887 to 1967) in his classification of thought disorders attempted to make the diagnosis of schizophrenia more reliable by identifying a group of symptoms that were the most characteristic of the illness. His so-called first-rank symptoms included various forms of hallucinations, thought withdrawal, thought insertion, thought broadcasting, delusional perceptions, and experiencing feelings and actions as made or influenced by external agents (Table 12.2–1). The presence of one of these symptoms, in the absence of intoxication, brain injury, or clear affective illness, was sometimes taken as sufficient for making the diagnosis of schizophrenia.

The different conceptualizations of the illness and diagnostic criteria used in clinical practice and research led to wide disparities in the rates of diagnosis of schizophrenia among clinicians and researchers around the world. This problem ignited a search for agreements and conventions to facilitate communication and define more formal and reliable diagnostic criteria giving birth to modern nosology.

Modern Diagnostic Criteria

In the United States, the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) of the American Psychiatric Association (APA), in its second edition (DSM-II), was heavily influenced by Bleulerian criteria to make the diagnosis of schizophrenia. The predominance of psychological and psychosocial theories in the field at the time, such as the theories of Adolph Mayer, Harry Stack Sullivan, and the Freudian followers, were fertile ground for the adoption of a view of schizophrenia according to Bleuler. The consequences of adopting such concepts led to broad and arbitrarily applied criteria to diagnose schizophrenia, sometimes making the rate of diagnosis rather high. Many conditions now classified elsewhere were included in the diagnosis of schizophrenia according to the DSM-II. As a consequence, studies performed during the 1960s and 1970s in the United States and in Europe showed that the reliability of diagnosis locally or across countries was very low. In the United States, the more inclusive diagnostic criteria for schizophrenia led to rates that were much higher than in Europe, where many more patients were instead diagnosed with affective psychosis. Through the middle of the 20th century, the rate of schizophrenia as an admission diagnosis for psychiatric facilities in London stayed similar over a 40-year period, whereas the diagnosis of schizophrenia appeared to be increasing as an admission diagnosis in New York City, with no clear explanation for the differences. The United States–United Kingdom Cross National Project study of hospitals in London and New York showed that approximately 60% of the admissions in New York State hospitals were diagnosed with schizophrenia, whereas in London hospitals, admissions with the diagnosis of schizophrenia were about half that rate. When standardized diagnostic interviews were applied in these settings, the rates of schizophrenia in New York decreased to levels that matched the rates in London. These findings triggered the need to define more standardized diagnostic criteria on both sides of the Atlantic. A later groundbreaking study that propelled the adoption of explicit diagnostic criteria was the International Pilot Study of Schizophrenia (IPSS) sponsored by the World Health Organization (WHO). One thousand two hundred patients with psychotic disorders from nine countries were interviewed using a structured diagnostic interview, the present state examination (PSE). This study showed that schizophrenia manifested itself in similar ways across very diverse cultures, with similar rates of hallucinations, delusions, first-rank symptoms, and affective symptoms among people with schizophrenia in developing and developed countries. The findings from this study reaffirmed the belief that useful and uniform formal criteria could be developed and would contribute to better communication among clinicians and researchers alike.

Table 12.2–1.
First-Rank Symptoms of Kurt Schneider

Audible thoughts	Auditory hallucinations of a person's voice being spoken aloud
Voices arguing or discussing	Auditory hallucinations of two or more voices arguing or discussing, usually about the person experiencing the hallucination
Voices commenting on patient's actions	Auditory hallucinations commenting on a person's behaviors
Somatic passivity	Tactile or visceral hallucinations that are imposed by some external agent; can be combinations of different somatic hallucinations
Thought withdrawal	Sensation of thoughts being actively removed from a person's mind
Thought insertion	Thoughts inserted into a person's mind by some external agent
Thought broadcasting	The sense that a person's thoughts are experienced as real phenomena by others—the thoughts are made audible or may be experienced by others through telepathy
Made feelings	Feelings that are not a person's own are imposed on that person by an external agent
Made impulses or drives	An impulse for action is imposed on a person by some external agent
Made volitional acts	A person's actions are from and are controlled by an external agent; the person is a passive participant in the action
Delusional perception	A perception that has a unique and idiosyncratic meaning for a person, which leads to an immediate delusional interpretation

The changes between DSM-II to DSM-III in the diagnostic criteria of schizophrenia represent the biggest conceptual departure from previous diagnostic classification systems and set the foundation of current nomenclature. This was a major break with decades of American and European nosologic tradition. The result was a much narrower definition of the disorder. This major conceptual change was possible due to the efforts published in 1972 by the Washington University psychiatry research group in St. Louis, Missouri, that introduced a diagnostic classification system aiming at greater validity and reliability of diagnoses. These criteria, known as the Feighner criteria after the first author of the initial report, were rigorous in trying to define homogeneous diagnostic groups. The principles underlying this classification system were (1) clinical description, (2) laboratory study, (3) delimitation from other disorders, (4) follow-up studies, and (5) family studies. The next major step toward current diagnostic criteria was the introduction of the research diagnostic criteria (RDC) in 1978. This effort borrowed many concepts from the Feighner criteria and the Schneiderian tradition. The RDC emphasizes the temporal length of observable psychopathology and differentiation based on the presence or absence of mood symptoms. The main difference between the Feighner and the RDC criteria was the reduction of the duration of active symptom phase from 1 month in the Feighner criteria to 2 weeks in the RDC. The duration of illness was kept at 6 months to reach diagnosis. In addition to the definition of formal criteria for schizophrenia, a structured diagnostic interview was developed (Schedule for Affective Disorders and Schizophrenia, or SADS). This kind of structured diagnostic interview became the precursor to several instruments currently in use in research.

The DSM-III, published in 1980, for the first time defined schizophrenia much more narrowly, placed more emphasis on Schneiderian concepts, and included a multiaxial diagnostic system. Dimensions such as secondary diagnosis, impact of functioning, and relevant medical illness and severity were now included in the formal diagnosis. In addition, clear exclusion criteria were described, and the maximum age of onset at 45 years was eliminated. The principles used to guide the development of DSM-III were also used for the development of DSM-IV and DSM-5. Among the many important features considered to make the diagnosis of schizophrenia, there was a new emphasis on using course of illness in addition to cross-sectional symptoms in an effort to improve reliability and validity of the diagnosis. A requirement for symptoms of at least 6 months' duration was added as a requirement for the diagnosis of schizophrenia.

Current Diagnostic Criteria

The current versions of the DSM and *International Classification of Diseases* (ICD) diagnostic systems are the DSM-5 TR and the ICD-11. In spite of some differences, both systems have a great deal of concordance in terms of the definition of schizophrenia. Both systems reflect a clear narrowing of the diagnosis of schizophrenia compared to previous diagnostic practices, especially in the United States. One of the major differences between them is the longitudinal course criterion. The DSM-5 TR requires a minimum duration of the illness of 6 months (with at least 1 month of active symptoms) as opposed to only 1-month total duration of illness in the ICD-11. Also, the DSM-5 TR requires deterioration in social and occupational function, in contrast to the ICD-11, which does not have such a requirement for the diagnosis of schizophrenia. The DSM-5 criteria for schizophrenia, unchanged in DSM-5 TR, were released in 2013 and included some important changes while retaining the core structure of DSM-IV. The symptom threshold for the diagnosis was raised from needing one core symptom in DSM-IV-TR to at least two symptoms. Core symptoms continue to be delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms. Both DSM-5 TR and ICD-11 eliminated the special diagnostic importance for first-rank Schneiderian symptoms such as running commentary hallucinations, two or more voices conversing with each other, and having bizarre delusions. Delusions are now treated equally toward the diagnosis of schizophrenia regardless of their content. Also a major departure from previous editions of the DSM and ICD was the elimination of all subtypes of schizophrenia due to lack of data supporting their stability over time and any clinical or research predictive value.

Diagnostic Criteria: Functioning. In addition to the cross-sectional criteria, DSM-5 TR requires the presence of significant deterioration in one or more major areas of functioning, such as work, interpersonal relations, or self-care. In contrast, ICD-11 does not require functional deterioration for a diagnosis of schizophrenia and in fact, states that deterioration is not inevitable.

Diagnostic Criteria: Duration. The DSM-5 TR criteria require a total duration of 6 months. Within this 6-month period, there must be at least 1 month of active-phase symptoms (overt psychotic symptoms). A shorter duration of the active phase is allowed only if successful treatment is instituted. The rest of the 6-month period may include continuing psychotic symptoms, prodromal symptoms preceding clear-cut psychosis, or residual symptoms after the resolution of the psychotic symptoms. Residual symptoms are defined as attenuated forms of psychotic or disorganization symptoms, such as odd beliefs, magical thinking, ideas of reference, odd perceptual experiences, peculiar or concrete thinking, vague speech, or odd behavior. Negative symptoms may also be included as residual symptoms. If the 6-month duration criterion is not met, but the other criteria are present, the diagnosis of schizophreniform disorder should be made instead.

In contrast, ICD-11 requires only a minimum duration of 1 month. This time does not include the prodromal phase and does not consider treatment as a variable to shorten the duration of the acute phase. If this duration requirement is not met, the diagnosis of other primary psychotic disorders should be made instead.

Diagnostic Criteria: Exclusions. The DSM-5 TR criteria require that mood and schizoaffective disorder be ruled out to make the diagnosis of schizophrenia. The patient should not meet criteria for a manic or depressive episode during the psychotic phase, or if there is a concomitant mood episode with active phase symptoms, the duration of the mood episodes should be brief relative to the duration of the psychotic illness. This includes residual symptoms. In contrast, according to ICD-11, a patient with a depressive or manic episode can still meet criteria for schizophrenia, if criteria for schizophrenia were met before the onset of the mood episode.

DSM-5 TR excludes a diagnosis of schizophrenia if the symptoms are the result of a direct physiologic effect of a substance, a drug of abuse, or a medication, as well as when the symptoms are due to a neurologic or medical condition. ICD-11 includes very similar exclusion criteria and now classifies psychotic syndromes with a known cause as secondary psychotic syndromes and place them under their respective medical or psychiatric categories.

Diagnostic Subtypes

The diagnostic subtypes of schizophrenia were dropped in DSM-5 TR and ICD-11. Schizophrenia subtypes have been controversial constructs and their clinical utility has been consistently questioned. They have a weak relationship to biologic variables, poor long-term stability, and poor predictive value. The paranoid subtype appears to have greater stability than the other subtypes and was used more often in clinical practice along with the undifferentiated type. Instead, in order to capture the rich and diverse clinical presentation of schizophrenia with heterogeneous symptoms and signs that vary over time, the DSM-5 included in its assessment measures section, the Clinician-Rated Dimensions of Psychosis Symptom Severity, an assessment measure that allows quantifying the five core dimensions that define schizophrenia spectrum disorder (hallucinations, delusions, disorganized speech, abnormal psychomotor behavior, and negative symptoms) plus impaired cognition, depression, and mania. Each dimension is rated on a 0 (not present) to 4 (present and severe) scale and can be used for treatment planning, outcomes, and longitudinal follow-up. Catatonia was designated as a separate category in DSM-5 and remains a separate category in DSM-5 TR, that can now be linked to multiple disorders such as schizophrenia, schizoaffective disorder, and mood disorders. ICD-11 now has a similar approach to characterized individual symptoms over time.

SYMPTOMS OF SCHIZOPHRENIA

The descriptive power of Kraepelin dementia praecox and the explanatory power of Bleuler schizophrenia took hold in psychiatry during the first half of the 20th century. By the middle of the century, varying conceptual orientations among psychiatrists led to widely differing applications of the diagnosis of schizophrenia. Schizophrenia was broadly applied, with the diagnosis given to both chronically hospitalized patients with constant hallucinations and bizarre behaviors and independent working people with problems of histrionic behaviors and difficulty making empathic connections with their therapists. With the introduction of new medical therapies to treat the symptoms of depression and psychosis in the 1950s, it became essential to define which symptoms were associated with which illnesses, so that the right treatments could be applied to the right illnesses and treatment responses could be determined.

To apply statistics to the severity of symptoms and treatment responses in psychiatric illnesses required that both symptoms and responses be quantified, and the tool for quantification arose with the generation of rating scales for psychiatric disorders. The first broadly used symptom scale was the Brief Psychiatric Rating Scale (BPRS) from Overall and Gorham, published in 1962. This scale was derived from earlier scales and from the evaluation of hundreds of patients and assessed aspects of psychopathology that were familiar to most clinicians. It was short (its first form consisting of just 16 items), easy to administer, and reliable. Initially intended to provide efficient assessment of treatment response for patients in inpatient settings, its utility led to wide use in measuring treatment responses of severely ill patients across clinical settings. The BPRS was the standard tool for measuring treatment response in psychotic disorders for decades, and it remains widely used in research and as an aid to clinicians treating chronically and severely ill patients.

At about the same time the BPRS was being developed, John Wing and his colleagues were trying to address the poor reliability of psychiatric diagnoses with creation of the PSE. The PSE is a much longer instrument than the BPRS and evaluates 140 signs and symptoms and requires experienced clinicians for reliable administration. Algorithms to facilitate computer-based interpretation of the large bodies of data generated were developed in parallel, with the aim of using the PSE to identify groups of patients united by similar symptom patterns. Despite the time involved in administering the PSE and the large bodies of data created when administered to groups, the PSE was used in both the New York–London project and in the IPSS, and its use helped define the parameters of schizophrenia. Like the BPRS, the PSE was intended to be an instrument applicable across a broad range of psychiatric disorders, but it has its greatest impact on studies of schizophrenia.

Reliable rating scales of symptoms and psychopathology, such as the PSE and BPRS, were essential to the progress made in understanding the treatment and course of schizophrenia and finding the biologic and social correlates of the illness. The authors of the BPRS and the PSE, and the authors of the many rating scales that followed, were all motivated to address particular questions about illness and treatment, and none of them intended to constrain the clinical interview of patients suffering from schizophrenia. But the reliability and utility of these scales and the diagnostic criteria derived from them have nonetheless had the effect of focusing researchers' and clinicians' attention on the questions asked by the scales and diverting attention from the stories patients have to tell. Kenneth Kendler has reviewed the development of the concept of schizophrenia at length and found that early descriptions included a broad number of symptoms, and more so particular signs, that have been lost in the diagnostic criteria of schizophrenia and so also lost in teaching and routine assessment. The scales, like the brief and reliable contemporary diagnostic criteria psychiatrists, apply, have become the definition of illness in the mind of many mental health clinicians, and this view limits inquiry about the individual experience of patients.

An early application of the PSE was in the WHO's IPSS. The results of the IPSS showed that populations of patients sampled in a range of rural and urban sites around the world all had similar profiles of symptom types, although individual participants displayed widely varying clinical presentations. Across sites the most common symptoms were lack of insight and an inadequate description of problems, suspiciousness, unwillingness to cooperate, ideas of reference, flat affect, persecutory delusions, delusional mood, poor rapport, a range of auditory hallucinations including verbal hallucinations, thought alienation, and gloomy thoughts. Subsequent follow-up studies, other studies conducted across countries, and studies of different cultures within individual countries confirmed that when studied as groups there was a similar overall distribution of symptoms within populations.

By the 1970s the body of data produced through the use of the PSE in large samples led to new efforts to find correlates between the symptoms of the disease and functional outcomes. Strauss, Carpenter, and Bartko made an early attempt to find patterns in the expression of the symptoms of schizophrenia. In their studies, the PSE-rated symptoms of schizophrenia clustered into groups of positive and negative symptoms and into a group of symptoms associated with "disorders of relating." An advantage of the PSE is complexity of patients' experience captured by its number range of symptoms and signs, but as the utility of rating scales for clinical practices, and especially for pharmacology studies became apparent, the BPRS, and later the Positive and Negative Syndrome Scale (PANSS), were adopted as reliable options that were easier to administer and interpret.

By the mid-1980s factor analytic techniques were broadly applied to the assessment of the symptoms of schizophrenia, and separate investigators found that three dimensions or subsyndromes of schizophrenia could be derived from rating scales. These dimensions were psychotic symptoms (hallucinations and delusions), negative symptoms (e.g., social withdrawal, blunting of affect, interest, and motivation), and disorganization symptoms (disorganization of thought and behavior). A similar clustering of symptoms was found to exist in studies of populations from around the world. Later studies showed that the negative syndrome especially and to a lesser extent the disorganization syndrome showed temporal stability, and that each subsyndrome had its own distinct relationship to demographic factors, cognition, and heritability.

It has become clear that the number of subsyndromes within schizophrenia is defined in part by the number of signs and symptoms measured, so that more thorough ratings with larger numbers of symptoms lead to finding more subsyndromes. Scales or combinations of scales that include more symptoms lead to the reliable derivation of psychotic, negative, disorganization, depression and anxiety, and agitation dimensions. There have been many thousands of patients enrolled in studies using the PANSS conducted across continents and cultures that reliably find these same five factors. The meaning of an individual patient's symptom might vary across cultures but the structure of symptoms within the illness seems remarkably consistent. For the most commonly used rating scales, age, symptom severity, duration of illness, and short-term medication withdrawal do not seem to affect the finding of separate subsyndromes.

This dissection of the syndrome of schizophrenia into subsyndromes will multiply as more and more elaborate assessments of signs, symptoms, and history are completed. As an example, a study that detailed premorbid and prodrome histories along with routine assessments of clinical symptoms found separate factors for educational achievement and premorbid function with unique heritability. More detailed assessment of disorganization symptoms finds subfactors for motor and verbal disorganization, and when a larger number of mood and anxiety questions are asked, mood and anxiety may be split into separate factors. However, symptoms in each subfactor are most highly correlated with the other symptoms in the same class. At this time, the inclusion of additional factors, or the division of the five PANSS factors into separate subfactors, is primarily important for investigators with a narrow focus on therapeutics or neurobiology applied to a particular subfactor—the five-factor model is a valid and useful way to organize the clinical symptoms that are assessed to make decisions about patient’s care.

Psychotic Symptoms

The psychotic symptoms of schizophrenia define schizophrenia for the general public and for many clinicians. They are the obvious and dramatic symptoms of the illness, and for most people, they are what is fascinating and frightening about schizophrenia. Within psychiatry the interest in delusions and hallucinations has overshadowed investigation and treatment of the other facets of the psychopathology of schizophrenia. This emphasis on psychotic symptoms was reinforced with the break from the impressionistic diagnostic scheme of DSM-II to a symptom-based diagnosis in DSM-III, when the diagnosis came to depend in large part on clinicians’ ability to recognize the psychotic symptoms considered characteristic of schizophrenia.

It is now clear that there is no particular sign or symptom, not even the most carefully specified kind of hallucination or delusion, which by itself justifies a diagnosis of schizophrenia. It is also clear that a focus on psychotic symptoms to the exclusion of the other signs and symptoms of schizophrenia will ignore significant sources of morbidity and likely exacerbate other forms of psychopathology and disability. Nonetheless, psychotic symptoms, because of their unique nature and easy identification, aid assignment of diagnoses and serve as markers of treatment response, and these symptoms will remain at the core of concepts about the nature of schizophrenia.

The use of the term psychotic in relation to symptoms should be restricted to hallucinations and delusions. In the past, symptoms of inappropriate affect, purposeless activity, and disorganization of speech and behavior were also considered psychotic symptoms, but it is clear these are better considered in the separate disorganization dimension. Psychotic symptoms are also not synonymous with the term “positive symptoms,” which has traditionally included symptoms such as incoherence and loosening of associations, or motor restlessness, properly assigned to disorganization and agitation factors. Similarly, the symptoms in the original PANSS “Positive” subscale are a composite of psychotic, disorganization, and agitation factors. There is no reason to use the original PANSS positive subscale.

The responsiveness of psychotic symptoms to antipsychotic medications has led to unforeseen costs in attempts to effectively treat schizophrenia. The focus on psychotic symptoms of schizophrenia that arose with the adoption of the diagnostic scheme of Kurt Schneider and its emphasis on “first-rank” symptoms of psychosis (Table 12.2–1) became prominent in the United States at about the same time the demonstrable benefits of antipsychotic medications became clear to most practicing psychiatrists. This evolved into a seemingly reasonable treatment approach that focused on maximizing antipsychotic doses to minimize or eliminate the psychotic symptoms that were judged to define the illness. The costs of this approach were not apparent to most clinicians, and only in the 1970s and 1980s were there psychiatrists who suggested that high-dose medication treatments might impose burdens such as secondary negative symptoms, dyskinesias, and increased rates of medical complications that were not always offset by the benefits achieved with decreased psychosis.

Hallucinations. The earliest descriptions of schizophrenia did not share the contemporary emphasis on psychotic symptoms, but there was certainly clear awareness of the burden hallucinations impose. Bleuler considered hallucinations accessory symptoms but in his *Textbook of Psychiatry* (1924), he evoked the disruption that hallucinations must impose on the inner life of those afflicted with schizophrenia.

> The patients hear blowing, roaring, humming, rattling, shooting, thundering, music, crying and laughing, but above all whispering, speaking, calling; they see individual things, landscapes, animals, people and all sorts of impossible figures; they smell and taste all sorts of pleasant and unpleasant things; they touch things, animals and people and are struck by rain drops, fire, and bullets, they feel all the tortures and perhaps all the pleasant things also that our bodily sensations can transmit.

Although an exclusive or predominant focus on the psychotic symptoms of schizophrenia is not appropriate, hallucinations do hold a special place in the experience of people with schizophrenia. Perceptions that can cause people to “feel all the tortures and perhaps all the pleasant things” will certainly influence behavior, and what seems incomprehensible in schizophrenia may become understandable when a patient’s experiences with hallucinations are revealed. Hallucinations also drive some of the understanding of illness that patients, their families, and their physicians share or do not share. Families and health care providers may not have the overwhelming experience of distorted perception that people with schizophrenia often have, but families and providers have dreams and vivid perceptions, and these experiences can help them imagine something of what happens with patients. Alternatively, families may chide patients for listening to people who are not there, and physicians will hurry through cursory encounters with patients who seem to pay more attention to voices of intangible people than they do to the doctor in front of them.

Many of Kurt Schneider's first-rank symptoms, proposed to specify a diagnosis schizophrenia—hearing one's thoughts spoken aloud, voices referring to oneself, a running commentary on the patient's thoughts or behaviors, a sense of being the recipient of externally imposed sensations—are hallucinations and were initially considered to make a diagnosis of schizophrenia likely in the absence of identifiable brain pathology. In fact, hallucinations are common in other psychiatric illnesses, and their presence appears to be an indicator of overall illness severity across mood, anxiety, personality, and eating disorders. The first-rank symptoms have been found to vary from a minority of patients to very high rates in both schizophrenia and other psychiatric disorders, depending on the population being studied. Although there is no particular unique feature of hallucinations in schizophrenia there is typically an earlier age of onset and greater chronicity than found in other psychiatric illnesses.

Hallucinations and other psychotic symptoms are not rare in the general population. A meta-analysis put the rate of "psychosis-like experiences" in the general population slightly greater than 7%, and in a large nonclinical sample comprising populations from around the world hallucinations were about four times more common than delusions. In the general population symptoms of anxiety are particularly associated with hallucinations, and auditory hallucinations predominate. Not surprisingly substance abuse is also associated with all kinds of hallucinations, but a particular association between substance abuse and auditory hallucinations is not found. Conceiving hallucinations as a discrete manifestation of psychopathology that establishes the presence of a mental illness might be a convenient assumption that provides security to diagnosticians' but does not in fact reflect the nature of the experience. Younger age, living in an urban environment, low income, less education, unemployment, female gender, and being single are all associated with increased rates of hallucinations. Most of these demographic factors seem similar to those of people who will develop schizophrenia. Also similar is the finding that in the general population hallucinations predict the presence of delusions, and that acute stressors increase the likelihood of hallucinations. It has been clear for decades that hallucinations do not confirm or exclude schizophrenia versus any other psychiatric diagnosis, and the presence of hallucinations does not by itself confirm psychiatric illness.

Since the time of Karl Jaspers there has been a view that hallucinations that seem to arise from outside the body are "true" hallucinations, while hallucinations that seem to arise inside the body are pseudo-hallucinations; that this is a distortion of Jaspers' actual discussion has not affected views. There is a similar view among some that good insight into hallucinations means that the perceptions are not really hallucinations. Neither of these distinctions, nor the volume, number of voices heard, or content of the hallucinations, has diagnostic significance or demographic correlates. Many patients will describe auditory hallucinations that arise both inside and outside their heads or will develop insight into the nature of their hallucinations only after decades of full belief in their external reality. When considering patients, these distinctions should not be used to make a diagnosis or generate a treatment plan.

There is more information about the nature of psychosis than any other symptoms of schizophrenia, but even in the study of hallucinations, there are large gaps in the current knowledge. Most investigations focus on the presence or absence of a symptom, and when there is a more detailed inquiry, it is typically limited to questions about frequency. There is literature dealing with questions of when hallucinations arise, what provokes them, how patients cope with them, or what other perceptions or thoughts are associated with hallucinations, but the answers to these questions arise from a surprisingly limited number of studies.

Studies that examine the prevalence of hallucinations across cultures come up with similar rank orders, though the exact prevalence of each type of hallucination in each setting varies. Auditory hallucinations are the most common type, followed by visual hallucinations, and tactile (or *haptic*), olfactory, and gustatory hallucinations are less common. Visceral and other deep tissue hallucinations (*cenesthetic* hallucinations) can provoke odd descriptions that are inconsistent with the internal sensory apparatus, with complaints of burning in the brain or movement of the thoracic organs to new positions, and are likely the least common type of hallucination in schizophrenia. There has been an emphasis in clinical assessment and research in schizophrenia on auditory hallucinations that have obscured the frequent experience of hallucinations across multiple sensory modalities. A study of a large and well-characterized population in the Netherlands found that about two-thirds of patients had the experience of hallucinations in more than one sensory modality, and 25% had experiences in three or more senses.

AUDITORY HALLUCINATIONS. In schizophrenia, auditory hallucinations are clearly the most common type of hallucination. This may not always have been the case; a century ago, visual hallucinations were more common than auditory hallucinations in London, and visual hallucinations have been decreasing in frequency, and auditory hallucinations have been increasing in frequency over the past century in Vienna. Auditory hallucinations are relatively less common, and visual hallucinations are relatively more common in African, Near Eastern, and some Asian populations than in Western societies. Nonetheless, evidence from the IPSS and other sources confirms that the most prevalent type of hallucination and the most common symptom of psychosis in schizophrenia is auditory hallucinations.

In keeping with their prevalence, auditory hallucinations are the best-characterized of the psychotic symptoms. The IPSS studies suggest that internationally, more than 70% of people with schizophrenia have auditory hallucinations, and that percentage is probably higher in industrialized societies. As with many symptoms, careful questioning in prospective studies is associated with a higher rate of reporting—for auditory hallucinations leading to a prevalence of 98% in one report.

Patients generally believe that their hallucinations are real manifestations of someone, somewhere, talking to them or transmitting a voice or voices. Patients frequently attempt to test the reality of their voices, but patients' tests often rely on attempts to confirm that the voices truly know what the patient is thinking, doing, or planning. Since the voices inevitably do manage to know such things, their reality is proved to the patient. Even when voices fail their experiments (e.g., failing to accurately predict what will be served for supper) a delusional explanation may be generated that confirms the reality of an external cause. Ideas of reference and other delusions further reinforce these interpretations of the hallucinations. Although there have been decades of speculation about why the content of auditory hallucinations is so often threatening or demeaning, there is at present no clear explanation for the generally negative content of the auditory hallucinations.

Verbal auditory hallucinations, both in words and in prosody, are generally derogatory or threatening and often quite profane. Patients will hear expletives, threats, demeaning personal comments, and accusations of vile thoughts or behaviors; this kind of negative content comes more commonly from male voices. Patients are often distressed and frightened, or angered by these experiences, and can have a complicated sense of guilt, depression, or belligerence in response. Patients typically feel less control over the experience of negative hallucinations than they do over benign or positive hallucinations, and severity of delusions and negative symptoms can further inhibit coping with hallucinations. In cultures that have socially accepted explanations for the experience of auditory hallucinations, the affective tone appears more benign, and there is less distress associated with the experience of hallucinations.

Emotional experiences, often sadness or apprehension, often precede hallucinations, but patients also often experience somatic symptoms of anxiety, fear, or anger, before the recurrence of hallucinations. Social stressors, physical illness, insomnia, and chronic pain can all increase the frequency of hallucinations and the intrusiveness and dysphoria associated with the experience. Hallucinations can be unremitting and last from waking until bedtime or can come in irregular, infrequent periods of seconds or minutes a few times a month. Although it is not a rule, hallucinations are generally less frequent and shorter in duration in the morning and become progressively more frequent and more extended in duration through the day until peaking before retiring for the night. It is common that people with schizophrenia have some period during their illness when their auditory hallucinations are present for the majority of the day.

Voices are the most common kind of hallucination and tend to be clear and understood, but unintelligible sounds of whispers or distant conversations are not rare. Patients can also hear footsteps, telephones ringing, the sounds of pebbles hitting a window, helicopters, buzzing, clicking, and musical phenomena that range from single notes present throughout the day to complex passages of orchestral music. Bleuler described "*extracampine* hallucinations" in which the perception was received by the patient outside the sensory apparatus—for example, a sound is appreciated through the knee.

When it is voices that are heard, single words are probably the most common. Although there is an enormous range in the number of voices that are heard by different patients, the mean number is estimated to be 3. The gender of the speaker can usually be identified. The person afflicted often hears voices that are interpreted as coming from a person who has more social status than the patient. Younger patients will tend to hear somewhat younger voices, while older patients hear somewhat older voices. The person afflicted by them often reliably assesses the location, loudness, and clarity of the voices, and in general, persistent auditory hallucinations will eventually lead the patient to settle on identities for their voices.

With the passage of time, most patients will experience a decrease in the frequency of their hallucinations or a change in the nature of their hallucinatory experience. Where hallucinations are initially very upsetting, provoking anger or panic, after the first stages of illness have subsided patients may generate coping strategies through a process of trial and error over a period of years and can come to mitigate the effects of hallucinations on their lives. Simple strategies such as yelling at the voices can sometimes decrease or stop the experience, though clinicians and families may see this as a distressing sign of deterioration. Less disturbing to people around the patient, singing, humming, or listening to music can block the experience, and sometimes shifts in posture or talking to someone else can help. Unfortunately for those afflicted the only available regular contact with the voices of other people may be online or by television, and for some, these can reliably exacerbate hallucinations. After years or decades, patients often start to question the content of their hallucinations and if their questions do not lead to insight, at least they sometimes provoke awareness that the threats or predictions rarely come true. Some patients can gain limited control of their hallucinations and can shift the prosody or content of the discussion, and there is a sizable minority of patients who can regularly start and stop their hallucinations at least some of the time. Religious themes are common in auditory hallucinations, and while they commonly start with a focus on the sins of the patient, with time there can be a conversion to a view of the patient having a special relationship with their god or gods, or that the individual might have some special message to spread or task to perform.

As a person gets used to the presence of their hallucinatory voice or voices, they can become companions, and given the frequent social isolation that is experienced in schizophrenia, this companionship can be welcome. The voices can take on the role of a guide through daily life for the person with schizophrenia, often in middle age or later so that the hallucinations will tell a person what to buy at the grocery store, how to structure their day, sometimes even how to respond in social interactions. The voices might even take on a moral role for the patient, suggesting the emotional consequences of particular behaviors or acting as a conscience that spurs generosity or remorse in response to the patient's actions. Although some people with schizophrenia will, with time, develop a good insight into the nature of their hallucinations, despite ongoing delusions or negative symptoms, other patients with predominately positive symptoms or residual schizophrenia will go through long periods with few obvious symptoms and no hallucinations but retain a full delusional interpretation of their past hallucinatory experiences.

COMMAND AUDITORY VERBAL HALLUCINATIONS. Sensational violent crimes committed by people with severe mental illnesses have led to an increased popular and clinical focus on the phenomenon of command verbal auditory hallucinations. Command auditory verbal hallucinations are very common, present at some phase of the illness for at least one-third and maybe as many as three-quarters of patients with schizophrenia; there are estimates that the lifetime prevalence of command hallucinations approaches the prevalence of auditory hallucinations as a whole. The emphasis on command hallucinations in clinical assessments of risk is unfortunate, because other clinical symptoms and features of the illness history that contribute more to the risk of violence are neglected once a patient's experience, or lack of experience, with command hallucinations is elicited in the clinical interview. There is a consensus that patients are more likely to comply with command hallucinations if (1) the voice is believed to be a real communication from someone else, especially if there is an identity for that person, (2) voices have benevolent intentions toward the patient or the action described will help the patient, (3) the voice has some omnipotence or other power greater than the patient, (4) there are adverse consequences to the patient for not complying, and (5) the command is for a nonviolent action. Patients with few coping strategies and a sense that the hallucinations are beyond their control are more likely to act on the commands. Trait anger and impulsivity also increase the risks of compliance. A patient's delusional structure in relation to command hallucinations is important–without some belief that a command should be followed, there is little reason for someone to act on the command.

VISUAL HALLUCINATIONS. Visual hallucinations are less common than auditory hallucinations in schizophrenia, but they are not rare. Estimates of the prevalence of visual hallucinations do not generally exceed 55%, and an estimate of about 40% of people with schizophrenia having visual hallucinations at some point in their illness appears consistent with the available evidence. As is the case with other psychiatric symptoms, careful investigation can sometimes elicit a history of visual hallucinations when casual questioning fails to detect their presence. Up to two-thirds of those who have auditory hallucinations will also have visual hallucinations, but it is relatively uncommon for someone with schizophrenia to have visual hallucinations in the absence of auditory hallucinations. Visual hallucinations have been suggested to predict a more severe illness, and there are some data to support this contention. It may be that patients with predominately psychotic symptoms are less likely to have visual hallucinations than patients with significant primary negative symptoms or disorganizations symptoms.

There is a broad range of visual experiences that afflict people with schizophrenia. The most commonly reported visual hallucinations are formed images of animate objects, people, or parts of people (especially heads and faces), religious images, animals, and fantastic creatures that may be similar to images in movies or television. Inanimate objects are less common, but flashes of light, streaks of color, or grids or lines in the visual field may also be described. Images can be in color or black or white, and hallucinations in only shades of white have been described. Visual hallucinations are generally more discrete and more limited in duration than auditory hallucinations, but there are patients who will experience visual hallucinations throughout the course of a day. Some particularly unfortunate people will have fluctuating periods of dramatic distortions of the world itself; the ground beneath someone will seem to be at angles that are inconsistent with the patient’s proprioceptive awareness, sidewalks will briefly appear to point into the air, and walls will appear or disappear in buildings. Lilliputian hallucinations—hallucinations of small people, frequently busy attending to some task—are generally considered a sign of an encephalopathy but they are sometimes present in schizophrenia. In vision, the continuum with illusions (distortions of objects seen in the environment) is more obvious than it is with auditory perceptions. Shadows that seem like people or animals but reveal themselves as only shadows with closer inspection, objects that change size or shape, outlines of objects that seem to float just outside the object itself, and objects that have acquired some transparency are all kinds of visual distortions that are reported.

The content of auditory and visual hallucinations is often dependent on the culture of the person experiencing hallucinations, but there seem to be some culturally appropriate visual hallucinations in non-Western societies that have only parallels to religious-themed hallucinations in the West. Even in cases of unequivocal schizophrenia, in non-Western societies, there can be a social acceptance that the hallucinatory experience is an extension of a commonly shared reality. It seems plausible, though not demonstrated, that cultural acceptability might influence the sensory mode of hallucinations experienced (e.g., visual instead of auditory), as well as the content of the hallucinations.

OLFACTORY, GUSTATORY, TACTILE, AND VISCERAL HALLUCINATIONS. Tactile hallucinations are present in a range of 15% to 25% of people with schizophrenia, with no clear pattern of variation across cultures. Particular tactile hallucinations, such as the sense of bugs crawling on or under the skin (*formication*), are found in a variety of mental illnesses and encephalopathies. Like auditory and visual hallucinations, the tactile hallucinations of schizophrenia can take on a broad variety of forms. Feelings of being touched, burned, and cut are common, and there are surprising numbers of patients who will feel as though they are subject to a kind of sexual assault, inflicted on their genitals or anus by unseen (and sometimes seen) entities. Less localized sensations of electric shocks or tearing or stabbing in the limbs and torso also occur. Patients with tactile hallucinations can also have a physical sense that somehow other people or magical beings enter into and exit their bodies, and while they can communicate that this sense is very distressing, articulating what it actually feels like may be very difficult.

Olfactory and gustatory hallucinations tend to be reported by a small minority of patients, with olfactory hallucinations the more common of the two types. As with other kinds of hallucinations the experience tends to be unpleasant for most patients, with the smells of rotting meat, garbage, and feces common, and the taste of blood or metal frequently described. Often tastes and smells will go together in an unpleasant combination. Visceral hallucinations are rarely noted in contemporary studies of hallucinations; whether this is due to a change in the nature of schizophrenia or a failure to fully investigate the experience of our patients is not known.

Delusions. Like hallucinations, delusions hold a special place in both the medical and popular perception of schizophrenia, with a view that delusions are a discrete category of thought that is generated and maintained through unique mechanisms and consisting of ideas that are inaccessible to normal reason. A standard definition of delusions is frequently reduced to calling them “fixed false beliefs.” In the glossary for DSM-5, delusions are defined as “A false belief based on incorrect inference about external reality that is firmly held despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary” and “not ordinarily accepted by other members of the person’s culture or subculture.” This emphasis on falsity and imperviousness to argument owes its origin in large part to the influence of the German psychiatrist and philosopher Karl Jaspers. In *General Psychopathology,* Jaspers wrote that delusions were “held with an extraordinary conviction, with an incomparable, subjective certainty,” that they were “impervious to other experiences and counter-argument,” and that their content is “impossible.” This description makes a delusion sound like something entirely unlike ideas generally held and understood, which we like to believe are arrived at through considered investigation and reasoned thought. Jaspers further argued that “delusions proper” arose from a unique kind of knowing or experiencing the world—not from the exercise of faulty judgment about personal experience but from a novel sort of experience that intrudes itself into the mind, a “direct experience of meaning.” This exists on no continuum with the experience of the general population, who derive meaning for their experiences from context, history, emotion, cognitive style, and state of mind. However, the variability in the type and character of delusions presented by patients has confronted the notion that delusions arise de novo, immutable, and impervious to reason or experience. To deal with this incompatibility between observation and theory, the awkward idea of “delusion-like-ideas” has been introduced.

Like hallucinations, delusions are not experiences that happen only to the mentally ill. When a representative sample of the German population was asked about their beliefs with a questionnaire, a quarter reported that they had special abilities or powers, and a third endorsed a belief in telepathy. Experiences as diverse as childhood trauma and sleep deprivation can provoke delusions in nonclinical populations, and as is the case in people with schizophrenia, the content is often paranoid. Reflection on daily life shows that many people clinicians know outside of clinical settings, and clinicians themselves frequently hold ideas that have no grounding in a shared external reality.

Furthermore, the delusions that are held by persons with schizophrenia are frequently quite flexible, and while argumentative confrontation accomplishes little, life circumstances or shifting attention can change delusions in dramatic ways. There are patients who will cling tenaciously to the idea that they have been persecuted by agents of a foreign government or who have had chips implanted in their brains by aliens from space, but every day around the world, other patients will check out fantastic ideas with family, friends, nursing staff, and doctors, and happily accept the assurance that their odd concerns are unfounded.

When delusions are examined as they exist in schizophrenia, there is a clear continuum with the overvalued ideas everyone is prone to. Delusions can vary from fleeting concerns about personal meanings taken from the way a television reporter intoned a particular word to a preoccupation with ideas about how giant corporations and governments are organized to persecute the individual with schizophrenia. The delusions found in schizophrenia frequently are different from the overinvested ideas that almost everyone has, but the differences do not seem to arise from a novel process that is found only in schizophrenia. Contrary to Jaspers’ formulation, delusions in schizophrenia often develop from normal cognitions, and the path by which delusions are developed often seems quite clear to clinicians when the circumstances and state of mind of the patient are described. Delusions encountered in clinical settings can be presented as impressions that quickly abate, and it is a common experience for clinicians to inquire about a delusional concern previously reported by a patient, only to find that the patient has dismissed their old idea, as though it is of no consequence, and now has new preoccupations. Instead of arising from a new kind of perception, delusions often arise from social stressors, and threats to self-esteem seem particularly prone to provoke persecutory or grandiose delusions in response.

What does seem to distinguish delusions in schizophrenia is analogous to what distinguishes schizophrenia itself: the frequency, severity of the impact of the symptom, and the pattern in relation to other symptoms, instead of any unique or pathognomonic content or structure. The number of delusional beliefs, the strength of preoccupation with delusions, and the amount of distress they cause for patients all seem to distinguish delusions in schizophrenia. Although delusions in schizophrenia are not always “firmly sustained,” there is in general a stronger conviction that the delusional belief is accurate despite contradictory views from others. Delusions are common in many psychiatric syndromes but tend to be held more strongly, preoccupy more time, and are more pervasive in their influence on patient cognition than is found in other psychiatric illnesses. Most people with schizophrenia will have delusions, and the percentage of people who will experience clinically significant delusions may exceed 70%. In Martin Harrow’s long-term follow-up studies, 25% of patients had delusions at all assessments over a 20-year period. As is the case with hallucinations, many patients will moderate the effects of delusions on their lives, often by titrating amounts of social contact. This increasing engagement can decrease their preoccupation, although when the delusions center on concerns of persecution this relief may not be available. For most patients, the presence of one delusion predicts the presence of other delusions.

CONTENT OF DELUSIONS. It is an unfortunate feature of delusions in schizophrenia that the focus of the delusion is so often based on fears of persecution or threats to a person’s safety. People with schizophrenia are more likely than people in the general population to fear that people look at them in an unusual or odd way, to believe that they are being watched or being harmed, or that their feelings and actions are not under their control. In schizophrenia, most delusions will fall into some broad patterns of content. The most common delusions in schizophrenia are delusions of persecution, which are estimated to lead to about half of help-seeking encounters, and are endorsed by up to 80% of inpatients. These persecutory fears can range from transient doubts about the intentions of strangers around them to profound convictions that charitable organizations or family members are plotting to have the patient tortured or killed. These may arise from misinterpretation of social interactions or social events or from misattribution of intent on the part of others.

A sizable literature has arisen around the formation of persecutory delusions. Patients with paranoia tend to have a worried thinking style, sleep disturbance, and anomalous internal experiences. They have increased interpersonal sensitivity and negative beliefs about themselves and are influenced by life experiences and social circumstances to view the world as threatening. Living in urban environments and co-occurring substance abuse seem to exacerbate the problem.

Other highly prevalent delusions, such as delusions of bodily control and of having one's thoughts audible to others, are also delusions that can lead to persistent dysphoria. *Delusions of reference* refer to delusions in which neutral stimuli or events or people take on a particular significance in reference to the person who experiences the delusions; in the DSM-5 glossary, they are "a delusion in which events, objects, or other persons in one's immediate environment have a particular and unusual significance." There is usually a persecutory slant to these delusions, but there may also be self-aggrandizing or religious meaning derived from them. In DSM-5 TR, they are considered distinct from *ideas of reference,* "feeling that causal incidents and external events have a particular and unusual meaning that is specific to the person." As already discussed, delusional ideas lie on a continuum with normal beliefs, so the distinction between delusions of reference and ideas of reference seems of little value in clinical settings and is a source of unreliability in assessments for research purposes. As important, "delusions of reference" and "ideas of reference" are frequently treated as synonyms by clinicians. Other common delusions encountered in schizophrenia, whose names do not clearly define their meaning, include *thought broadcasting* and *thought insertion.* These refer to a delusion that the afflicted person's thoughts are audible to others and a delusion that certain thoughts are not the person's own but are inserted into the deluded person's mind, respectively. Also common are grandiose ideas of personal power or significance and religious themes, which can be persecutory as often as grandiose in their content. There are a series of eponymous delusions, most commonly referring to delusions that focus on questions of the identity of the deluded person or those around them or referring to a particular impossible state of existence. When first identified, each of these eponymous delusions was supposed to define a particular diagnostic association with schizophrenia, but since they can be found in almost any medical or psychiatric condition where delusions arise, there is little value in knowing each particular name.

A type of delusion that holds special significance in schizophrenia is the concept of "bizarre delusions." In DSM-III and DSM-IV, the presence of bizarre delusions was deemed important enough that its existence, in isolation from other characteristic symptoms, was sufficient to meet Criteria A for schizophrenia. Like the idea that auditory verbal hallucinations of voices conversing or commenting, the significance of these delusions arose from the tradition of Schneider's first-rank symptoms. Bizarre delusions are those that are considered implausible by people who share the patient's culture, and this is generally taken to mean something that is judged physically impossible. There have been many criticisms of the concept of bizarre delusions, the most prominent being that when assessing presence or absence, the interrater reliability varies from fair to very low, even among trained raters. Bizarre delusions are not a universal finding and are not pathognomonic of any disorder, but they are more common in schizophrenia than in any other psychiatric illness. However, when included in diagnostic criteria, they do not affect the reliability of the diagnosis of schizophrenia versus other illnesses with psychosis, and the special status of bizarre delusions was dropped in DSM-5.

Although the most common theme in the delusions found in schizophrenia is persecutory, it might be that the most common inciting event for the formation of a delusion is the need to explain the experience of auditory hallucinations. The validity of the hallucinatory experience seems self-evident to most patients, but there is frequently considerable difficulty generating any clearly communicated idea about how the sounds of someone else's voice is perceived by the patient and the patient alone. A common resolution is for the patient to invoke a fantastic explanation about science and technology and conspiracies; in less technologically focused societies, magic and witchcraft and religion are the most common explanations. In both settings, the explanations fall into culturally validated explanations for otherwise incomprehensible phenomena. Complex and persistent hallucinations are associated with complex and systematized delusions, where contradictions and conflicts with the reality of a patient's situation have ready answers in the form of ever more elaborate and internally consistent delusions. In these circumstances, the power of a delusional idea can be impressive and sufficient to make patients discount widely shared knowledge possessed for most of their lives. Systematized delusions are most common in middle age, and this may be a reflection of how much time it takes after the onset of schizophrenia to accommodate a new reality of the illness. Other hallucinatory experiences and other severe symptoms of schizophrenia can also provoke delusional interpretations from patients. Olfactory and gustatory hallucinations in particular may predict more severe delusions.

As noted earlier, careful application of detailed assessments can lead to hallucinations and delusions being considered separate subfactors in the general class of psychotic symptoms. Outside of research settings, the separation seems to have little utility. Hallucinations and delusions together remain the most treatable cluster of symptoms in schizophrenia, and the severity and change in one symptom class are highly correlated with change in the other.

Potentially less distressing themes, such as grandiose and religious content in delusions, are experienced by half or fewer people with schizophrenia. It seems that awareness of the frequent occurrence of grandiose delusions in schizophrenia is being lost in clinical settings; like the tendency in the past to diagnose anyone experiencing hallucinations with schizophrenia, it has become common to diagnose everyone with grandiose delusions with bipolar disorder, disregarding other diagnostic discriminators. It can be generally expected that delusions of persecution or religious content will have the most effect on a patient’s behavior, although the response of each patient to each thought is dependent on that person and that situation.

DELUSIONS AND COGNITION. There does appear to be a route to delusions that can be understood outside the idiosyncratic experience of the patient who experiences a “direct experience of meaning.” Although the ultimate cognitive deficits underlying delusions remain the subject of conjecture, it is clear that contrary to Jasper’s position, there are reasoning errors that accompany delusion formation in schizophrenia. When tested, people with schizophrenia who have delusions will make decisions based on less information than people with schizophrenia, and based on less information than people in normal populations. Despite being poorly supported, these judgments are held in greater confidence than the decisions of control groups. There is a cognitive stance that externalizes causal attribution for events and judges people and neutral stimuli as hostile or threatening. It is not the case that people with schizophrenia and delusions are incapable of making reasoned judgments because when patients are instructed to hold on to their decision-making until all information is presented, they tend to make the same decisions as controls. This “rush to judgment” is a trait characteristic that does not disappear when delusions remit, and in schizophrenia, it may for a common cause for the formation of delusions.

Negative Symptoms

Negative symptoms remain under-recognized and underappreciated as a cause of disability in schizophrenia. It could be claimed that negative symptoms are the most important clinical symptoms in schizophrenia because the severity of negative symptoms predicts short-term and long-term disability and quality of life, better than the severity of psychotic or disorganization symptoms. Negative symptoms appear to be the most significant predictor of social function and are more stable across time than positive and disorganization symptoms. Negative symptoms “represent a loss or diminution of normal functions” and stand in contrast to positive symptoms where there are perceptions and cognitions and behaviors added to normal mental functions.

In the original conceptions of schizophrenia, negative symptoms held a more prominent place. In *Dementia Praecox and Paraphrenia,* Kraepelin remarked on “emotional dullness” and related it to “the loss of interest, the loss of inner sympathy, with the giving way of those emotional main-springs which move us to exert our mental powers, to accomplish our tasks, to follow trains of thought.” For Bleuler “‘emotional deterioration’ stands at the forefront of the clinical picture.” Kurt Schneider appreciated the emotional withdrawal and lack of empathy that clinicians frequently experienced in dealing with patients with schizophrenia but did not include these disturbances in his characteristic symptoms of schizophrenia. Following Schneider, the emphasis on reducing or eliminating hallucinations and delusions as the primary goal in the treatment of schizophrenia led to the under-recognition and undertreatment of negative symptoms, and that lack of recognition and treatment persists to the present.

The publication of DSM-III did little to change the emphasis on positive symptoms. The only acknowledgment of negative symptoms came in a reference to blunted or flat affect, which could elevate a formal thought disorder to status of a symptom when they occurred together. But the accumulating research of John Strauss and William Carpenter, T. J. Crow, Nancy Andreasen, and others through the 1970s and 1980s led to an acknowledgment of the role of negative symptoms in schizophrenia, and flat affect was added to the diagnostic criteria for schizophrenia in DSM-III-R. In the DSM-5 TR diagnostic criteria, negative symptoms manifest in "diminished emotional expression or avolition" to constitute one of the characteristic signs or symptoms necessary for fulfilling Criteria A for the diagnosis of schizophrenia. Adding negative symptoms to the diagnostic criteria in the DSM has not increased the incidence of schizophrenia, an indirect confirmation that these symptoms have been present but unrecognized throughout the period of modern diagnostic criteria.

The two-syndrome concept as formulated by T. J. Crow was especially important in spurring research into the nature of negative symptoms. Crow's formulation of two separate dimensions of psychopathology, with a treatment-responsive set of positive symptoms and a less-responsive and potentially irreversible set of negative symptoms tied to demonstrable neuropathology, led to the generation of testable hypotheses and research efforts that continue to be active to this day. These investigations were facilitated by the creation of reliable and valid rating scales that extended a focus on negative, as well as positive, symptoms, with the Scale for the Assessment of Negative Symptoms (SANS) and the Scale for the Assessment of Positive Symptoms (SAPS) of Nancy Andreasen, and later the PANSS as the most important of these new tools. It would be shown that the groupings of symptoms into negative and positive subsyndromes in the SAPS, SANS, and PANSS, did not reflect the reality of the five different subsyndromes—psychotic, disorganization, negative, mood and anxiety, and agitation—but this does not diminish the creative efforts that led to these scales or importance of these scales for research. In fact, it was only through careful analysis of the structure of symptoms in these scales that a more accurate characterization of the phenomenology of schizophrenia was possible.

Negative symptoms do not exist only in schizophrenia or only in psychiatric illness. A full range of negative symptoms are seen in people who have had brain injuries, and mild expression of these symptoms is seen in 5% to 10% of the population who do not meet criteria for a psychiatric illness. A significant negative syndrome does seem to be much more common in schizophrenia than in other psychiatric diagnoses, including in comparison with mood disorders, and in psychiatric illnesses the pattern of multiple enduring negative symptoms not due to medication or social effects appears most often in schizophrenia. In chronic populations about half the patients will have at least one overt negative symptom. In Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE), 40% of patients had significant negative symptoms, and of those, only one-half had significant positive symptoms. Although our diagnostic schemes are tied firmly to psychotic and disorganization symptoms, negative symptoms may better distinguish schizophrenia from other psychiatric syndromes, and it appears that negative symptoms are most reliably associated with outcome in schizophrenia.

There is a range of demographic, treatment, and symptom variables associated with the presence of significant negative symptoms. These include male gender, poor premorbid education and vocational function, earlier onset of illness, longer duration of untreated illness, a small or absent social network, and a lower likelihood of romantic relationships. Negative symptoms can, in some patients, predate the prodrome, and presence at onset predicts persistence. There is a reciprocal relationship with affective symptoms, where more prominent affective symptoms are associated with a milder severity of negative symptoms.

It is possible to conceive that any function of the mind might be lost, and it seems that the range of negative symptoms could be vast. In patient populations, the number of negative symptoms is fairly circumscribed. Many functions of the kind assessed with cognitive testing are impaired in schizophrenia, but except for deficits in recognition of emotion, these functions are generally not included in the negative symptoms. There is a modest but consistent correlation between negative symptoms and cognition, but neither is well explained by the other, and they should be considered separate domains of psychopathology. Importantly, the kinds of function lost in negative symptoms of schizophrenia have been consistent in studies conducted at various times in various sites around the world.

There is a weak but consistent relationship between levels of negative symptoms and levels of disorganization symptoms in populations with both kinds of symptoms. In both groups of symptoms, there is a greater increase in familial risk than is seen for psychotic symptoms. There is a greater concordance in monozygotic twins for negative versus psychotic symptoms, and there is an overall increase in negative symptoms in the relatives of patients with negative symptoms. Negative symptoms have also been found to correlate with measures of neuropathology, especially those measures of frontal lobe function or the volume of cerebral tissue.

Table 12.2–2.
Symptoms Assessed in the Brief Negative Symptom Scale

1. Intensity pleasure during activities
2. Frequency pleasurable activities
3. Intensity expected pleasure future activities
4. Lack of normal distress
5. Asociality: Behavior
6. Asociality: Inernal experience
7. Avolition: Behavior
8. Avolition: Internal experience
9. Facial expression
10. Vocal expression
11. Expressive gestures
12. Quantity of speech
13. Spontaneous elaboration

Categories of Negative Symptoms. The MATRICS consensus conference on negative symptoms, convened under the auspices of the National Institutes of Mental Health (NIMH) in 2005, suggested that there are five general categories of negative symptoms: avolition, anhedonia, social withdrawal (asociality), affective blunting, and alogia. New rating scales were developed to more carefully assess these categories, in the Clinical Assessment Interview for Negative Symptoms (CAINS) and the Brief Negative Symptom Scale (BNSS) (Table 12.2–2); the BNSS was designed for both research and clinical settings, and might be useful for clinicians who deal with large numbers of patients with psychotic disorders.

Multiple exploratory factor analytic studies have found negative symptoms to be subdivided into two main groups–avolition symptoms, with avolition, anhedonia, and asociality, and the affective blunting symptoms, including affective blunting and alogia. These two groups are the basis for the inclusion of avolition and affective blunting in DSM-5 TR as the representative negative symptoms in Criteria A. The first of these subdivisions—the avolition group—is assessed with interview, while affective blunting and alogia are signs observed by the clinician. This division of negative symptoms into disturbances of volition and displayed social communication, while conceptually useful for clinicians, has been shown by Gregory Strauss and colleagues to be less accurate than considering each of the five categories of negative symptoms as discrete; at present, this difference is most important for investigators examining the underlying biology of negative symptoms.

The most common negative symptoms are *avolition* and *anhedonia.* Avolition is the loss of will or drive (in neurology sometimes referred to as *abulia*). In psychiatry, the loss of will or drive is sometimes considered to be a manifestation of anhedonia, but in fact, these refer to separate kinds of pathology. Avolition is similar to apathy, and these are considered closely related, with avolition identifying a deficit in the ability to initiate action and apathy a loss of concern for an idea or task. The definition of avolition in the DSM-5 glossary is “an inability to initiate and persist in goal-directed activities” and specifies subsequent deficits in work, intellectual pursuits, and self-care as resulting from avolition. Avolition seems associated with deficits in grooming and hygiene, and it seriously impairs educational and vocational progress. Although often overlooked, this loss of will or drive is the most significant driver of the disability associated with negative symptoms.

Anhedonia is a loss of the ability to find or derive pleasure from experiences, activities, or relationships, and is likely the most persistent of the negative symptoms. In DSM-5, anhedonia is “lack of enjoyment from, engagement in, or energy for life’s experiences; deficits in the capacity to feel pleasure and take interest in things.” Estimates of the prevalence of anhedonia vary widely, but probably half of patients experience considerable anhedonia. Although present in depressive disorders, when anhedonia is present as part of a negative syndrome it is not a manifestation of depression. Evidence available at the time of the last edition of this text suggested that there was not a deficit in the ability to feel pleasure or reward at the moment, but instead, an inability to anticipate future hedonic capacities that were experienced by patients. Recent studies have convincingly shown that people with schizophrenia have a generally diminished hedonic experience, but there is no differential deficit in anticipated versus current hedonic capacity.

Social withdrawal is sometimes referred to as “passive or apathetic social withdrawal” and includes indifference to social relationships and decreases in the drive to socialize. This includes a decrease in the desire for relationships with family or friends, a decrease in sexual interests, constricted relationships with family and friends, and an inability to feel intimacy or closeness. Patients with social withdrawal have deficits in *theory of mind,* which refers to the ability to understand how another person might be thinking when there is common knowledge of their current circumstance.

Affective *blunting* is a reduction or inability to express emotion and is an important predictor of functional impairment in schizophrenia. The blunting of expression includes deficits in production, facial expression, prosody, and gestures, and there is a link to social cognition through a parallel deficit in understanding these social signals. In the DSM-5 glossary, it is termed “blunted affect” and defined as a “reduction in the intensity of emotional expression,” and “flat affect” the “absence or near absence of affective expression.” It is more common in men, in people with an early onset illness, and in people with poor premorbid function, and it predicts lower scores on quality-of-life measures.

Alogia is a decrease in verbal communication, or “poverty of speech,” and it is found in up to 25% of people with schizophrenia. Initially conceptualized as both the loss of production and a paucity of content with a normal volume of words, it is only the loss of production of speech that is a negative symptom. The lack of speech production is considered to result from a decreased rate of verbal cognition, and alogia is conceived of as a “negative thought disorder.” This loss of production includes an increased latency to response, short verbal responses, extended pauses, and a paucity or complete lack of spontaneous production. In DSM-5, it is “impoverishment in thinking that is inferred from observing speech and language behavior. There may be brief and concrete replies to questions and restriction in the amount of spontaneous speech.” A paucity of meaningful content in the presence of a normal amount of speech was considered a component of alogia but is actually an expression of a disorganization of thought and properly included with the disorganization symptoms. Affective blunting and alogia are two measures of communication, and some authors have labeled these symptoms as “disorders of relating.”

Course and Associated Features of Negative Symptoms. Although the prototype of a patient with negative symptoms is someone with years of illness, residing in a group home, and passive and indifferent to the events around them, negative symptoms can precede the onset of illness. Anhedonia and social isolation can predict schizophrenia in high-risk populations. These symptoms may be obscured by the drama around fear, agitation, and psychosis in first-break patients, but a broader range of negative symptoms can be found early in overt illness, with most patients having some negative symptoms. Anhedonia is most common, present in about half of patients at onset, and avolition and affective blunting are also common. If recognizable *akinesia* (obvious difficulty in initiating motor activity) is present at the first admission there is an increased risk of subsequent significant negative symptoms. All of the negative symptoms can progress during the first years of illness and alogia in particular can increase from relatively low rates at onset. About 25% of patients will have significant negative symptoms within 2 years of onset of the illness. There is no course of illness that any particular patient can be expected to follow, and for many, there is a variable course for their negative symptoms, with significant expression at some points, and subsiding of symptoms at other times. But prominent negative symptoms at onset are better predictors of the future course of illness than are levels of psychotic or disorganization symptoms.

Negative symptoms arise from multiple sources, including iatrogenic causes. Long-term institutional care or subsistence-level existence on the streets and in shelters removes normal, expectable pleasures and reinforcing activities and will narrow the number of relationships someone with schizophrenia can have. People with schizophrenia may lose the capacity to experience drives and social relationships and reward simply because circumstances eliminate them. A frequent and frequently unrecognized cause of negative symptoms is treatment with antipsychotic medications. Medications for the treatment of schizophrenia can cause a phenocopy of primary negative symptoms that can be distinguished only with a full history of illness or empiric trials of medication adjustment. Reports of excessive doses of antipsychotic medications causing blunting of affect, diminution of motor activity, impairments in grooming and hygiene, and anhedonia have existed since the introduction of neuroleptics, but for many patients overtreatment with medications remains the rule and not the exception. Extrapyramidal symptoms should not be a proxy measure for the presence of medication-induced negative symptoms, as extrapyramidal rating scales may remain stable even when dose changes sufficient to reduce negative symptoms have been made.

Negative symptoms can also arise from the presence of other psychiatric symptoms. Positive symptoms are a common cause of apparent apathy, social withdrawal, and alogia. Persecutory delusions can engender fear of human contact leading to social withdrawal, and prominent hallucinations can reinforce these fears. Frequent and intrusive hallucinations can lead directly to withdrawal and inaction, strictly through their effects on attention. Depression is common in schizophrenia, and with inattention or perfunctory clinical interviews it can be overlooked, and the appearance of apathy and affective blunting are then attributed to the disease processes intrinsic to schizophrenia instead of a treatable mood disorder. Secondary causes of negative symptoms, with the exception of medication effects, will often remit within a year of their onset, and sometimes the passage of time can help distinguish primary from secondary negative symptoms; symptoms that do not diminish over the course of a year and do not remit with manipulation of the medication regimen should lead to consideration of primary negative symptoms.

Negative symptoms are frequently part of the clinical picture associated with exacerbations of symptoms. As is the case with depression in schizophrenia, an increase in negative symptoms may parallel an increase in psychotic symptoms, and these symptoms may remit when the psychotic symptoms remit. In these cases, use of antipsychotic medications would be expected to decrease the severity of negative symptoms. This, in fact, is found in studies of medication treatment of recently admitted patients where antipsychotic medication does reliably reduce the severity of negative symptoms.

Table 12.2–3.
The Schedule for the Deficit Syndrome

The Schedule for the Deficit Syndrome requires at least 2 negative symptoms be present in the preceding year. The symptoms are not better attributed to other symptoms of schizophrenia, or to symptoms of another psychiatric illness, and are not due to effects of a medication.

Signs and symptoms assessed in the Schedule for the Deficit Syndrome:

1. Restricted affect
2. Diminished emotional range
3. Poverty of speech
4. Curbing of interests
5. Diminished sense of purpose
6. Diminished social drive

Primary Negative Symptoms. Primary negative symptoms are those negative symptoms that appear intrinsic to the disease process. Primary negative symptoms are the least variable of the symptoms of schizophrenia, so that a person with prominent primary negative symptoms is likely to continue to have significant manifestation of these symptoms throughout the course of their illness. This persistence of negative symptoms predicts future disability. According to Bleuler, “It has been known since the early years of modern psychiatry that an “acute curable” psychosis became “chronic” when the affects began to disappear.” When there are multiple negative symptoms and a careful history suggests that these are enduring and not due to secondary causes, the deficit syndrome (Table 12.2–3) is established.

People with schizophrenia and deficit syndrome have a very poor prognosis regarding the possibility of a full functional recovery. They will tend to have a lower clinician-rated quality of life, more time in a hospital, an increased frequency of symptomatic periods, and more social and physical anhedonia. They will generally have lower intelligence quotients, less education, and worse work histories, and they are less likely to be married and more likely to be men than people with schizophrenia who do not have the deficit syndrome. There is a more insidious onset, with a progression of negative symptoms, and more disorganization of behavior, and the negative symptoms in a person with the deficit syndrome are likely to respond poorly to medication treatment or social skills training. Not all symptoms and prognostic factors are worse in people with the deficit syndrome; they have no more delusions or hallucinations than nondeficit groups, they are less likely to have severe delusions around social themes and tend to have less severe suspiciousness, they are less likely to endorse prominent dysphoria, and they are less likely to have suicidal ideation or suicide attempts than nondeficit populations. Negative symptoms are associated with impairments in insight, and patients with significant negative symptoms rate themselves higher on subjective assessments of quality of life, despite lower scores on scales of function. There are profound implications for treatment and function for a patient with the diagnosis of the deficit syndrome, but reliable assessment has been found to depend on special training and the use of the Schedule for the Deficit Syndrome Scale.

The deficit syndrome affects a minority of people with schizophrenia, but it is a sizable minority. The estimated rates of the deficit syndrome in schizophrenia vary widely, with rates that are lower early in course of illness and progress for the first 5 years. A meta-analysis of 23 studies calculated a rate of almost 30%. Demographic associations for the deficit syndrome include an increase in summer births, reversing the usual pattern of risk for schizophrenia. It has been proposed and is generally accepted that the deficit syndrome is the result of a separate disease process within schizophrenia, and is not just the manifestation of particularly severe illness.

The deficit syndrome is clearly of great importance in clinical populations, but in usual treatment settings, time constraints and insufficient resources might make it impossible to differentiate long-standing and problematic negative symptom from true deficit symptoms. For many patients, a regular sequenced trial of medications and medication doses, attendance in psychosocial rehabilitation programs, or changes in living situations, which all might help tease out a secondary versus primary cause for negative symptoms, is not possible. So the distinction between deficit and nondeficit symptoms can be difficult for the practicing clinician. Robert Buchanan suggested the idea of “persistent negative symptoms” where possible secondary negative symptoms fail to respond to usual remediation methods. The persistent negative symptom criteria requires the presence of a diagnosis of schizophrenia and at least two negative symptoms, for at least 12 months, present during periods of clinical stability and not due to psychiatric comorbidities or medication effects. The potential for extrapyramidal symptoms to cause negative symptoms is emphasized. If widely adopted, these simple-to-apply criteria might help focus increasing clinical resources on what is currently an under-recognized problem.

Disorganization

The disorganization syndrome has been part of the conception of schizophrenia since Kraepelin included hebephrenia in dementia praecox, but the behaviors, experiences, and associated pathology have been subject to less investigation than the psychotic and negative symptoms. Consequently, the boundaries of the disorganization syndrome are less clear than those of the psychosis and negative syndromes. However, the disorganization syndrome might be the most heritable of the subsyndromes of schizophrenia. Family and twin studies show an association between degrees of relatedness and the presence of prominent disorganization. This evidence parallels the finding that the old classification hebephrenia is the most heritable of the traditional subtypes of schizophrenia.

The disorganization symptoms certainly include the formal thought disorders, bizarre and catatonic behaviors, and inappropriate affect. Factor analytic studies also consistently find that impairment of attention and poverty of content of speech fit best in the disorganization symptoms. Disorganization symptoms can affect almost any kind of behavior and disrupt almost any form of thought. The disruptions of behavior can be marked, with wildly inappropriate expressions of affect including odd or exaggerated gestures, singsong or childlike changes in prosody, silly or inappropriately bright affects, or alternatively, grimacing and wide-eyed expressions of surprise or anger. This disruption of affect can be socially disabling and unfortunately seems to respond more poorly to treatment than other symptoms in the disorganization subsyndrome. During acute episodes, approximately 20% of patients exhibit an inappropriate affect. The motor manifestations of disorganization may change with time; first-episode patients have only a modest relationship between disorganized motor behavior and disorganization of speech but there are strong correlations in chronic populations.

Disorganization can also include erratic behaviors such as broad movements of arms or the trunk as the patient ambulates, jerking the head or the neck, or posturing that can resemble aggressive stances. Hygiene is frequently poor, and as a group those with disorganized schizophrenia are more indifferent to routine activities of daily living, including health care and health care maintenance, than are other patients with schizophrenia. Attire can be simply unusual or inappropriate (e.g., adults squeezing into adolescents' clothing, multiple layers of clothes during hot summer days, or light-weight attire in the middle of winter), or sometimes dramatically eccentric.

Prominent disorganization symptoms at first episode, especially with an early onset of illness, are associated with a decrease in general levels of function. Although disorganization symptoms follow a variable course for most patients, there are a small number of patients who have multiple persistent disorganization symptoms throughout the course of their illness, and in some cases, these symptoms can be progressive. This group constitutes the traditional hebephrenic subtype of schizophrenia and is less prevalent than the group of patients with enduring primary negative symptoms or the deficit syndrome. These people are also markedly and persistently disabled and should be the subject of a greater research focus.

Motor Symptoms. Motor behaviors can include subtle repetitive hand movements or broad, complex, and purposeless movements that involve the limbs and the trunk. These movements can occur when speaking and seem to be the gestural equivalent of a neologism or can involve multiple movements of a limb or limbs, as though the person were performing some complex but undecipherable manual task. Patients may sometimes mimic the motor behaviors of others, which is referred to as *echopraxia.* Rocking when sitting down or standing is common, as is patients hugging themselves, wringing their hands, or toying with their clothing, hair, or small objects around them. More complex motor behaviors, odd but apparently purposeful, have been described. There can be repetitive directed behaviors such as gestures and hugging others, or patients might have periods of appearing to read materials over and over–even if the text is upside down, or stare for protracted periods into the mirror. An early onset of prominent motor symptoms has been associated with significant thought disorder and may predict greater disability and a deteriorating course.

Symptoms of catatonia are included in the disorganization symptom, although it is important to note that catatonia is more common in psychotic mood disorders and in cases with brain injury and encephalopathies than schizophrenia. This is made explicit in DSM-5 TR, where catatonia can be a diagnostic specifier for other mental illnesses. The DSM-5 defines catatonia as “a marked decrease in reactivity to the environment” and its essential feature as a “psychomotor disturbance that may involve decreased motor activity, decreased engagement during interview or physical examination, or excessive and peculiar motor activity.” In its appearance, it is as though central motor programs are engaged without direction from frontal areas engaged in higher-level planning. The diagnostic criteria require at least three of 12 motor signs: stupor, catalepsy, waxy flexibility, mutism, negativism, posturing, mannerism, stereotypy, agitation, grimacing, echolalia, and echopraxia.

There is a long history of description of catatonic symptoms, and many general classes have been identified. *Catatonic excitement,* with constant purposeless motor activity, might be the most common type seen in psychiatric settings in the Western world; this is sometimes referred to as “agitated catatonia.” Catatonic *rigidity* is similar but refers to patients holding simple, fixed, rigid postures. DSM-5 clarifies some definitions—*negativism* refers to lack of response or automatic resistance to instructions or stimuli, catatonic *posturing* refers to spontaneous and active maintenance of a posture, and *mannerisms* are an “odd, circumstantial caricature” of normal motor behaviors. Some people with *catatonic stupor* can be obviously awake but immobile without rigidity and will sit or lay in these postures until moved. *Catalepsy* is holding a posture held against gravity and is differentiated from waxy flexibility or *cerea flexibilitas,* which is in DSM-5 a tendency to pose modest resistance to repositioning of limbs; in the past, catalepsy and waxy flexibility were often used as synonyms. *Echopraxia* and *echolalia* retain their familiar meanings of unconscious mimicking action or speech, respectively.

Catatonia is less prevalent in the Western world than it was before the introduction of convulsive therapies and antipsychotic medications. In old photographs or films of state hospitals, odd postures and gestures, grimacing, or patients immobile and slumped in chairs are ubiquitous. In the first edition of this textbook, this description was given:

> Until the middle 1930s, mental hospitals were filled with stuporous catatonics, many of whom would lie motionless for weeks or months, unresponsive to almost every stimulus,

> their eyelids flickering, saliva drooling from their mouths. Since they would not take any nourishment, they had to be fed by stomach tube twice a day as long as they remained stuporous.

Despite its decreasing prevalence in schizophrenia, some signs of catatonia are present in a large minority of patients during some phase of their illness. One estimate puts catatonia at 8% lifetime prevalence in schizophrenia, and DSM-5 suggests up to a 35% rate in inpatient settings. Unlike other markers of severity of illness in schizophrenia, the prevalence is essentially equal across genders. The risk may increase with increasing maternal age, and there is an increased risk of suicide in the absence of an increase in the prevalence of mood disorders.

The experience of catatonia is often difficult to elicit from patients after an episode has resolved. Patients often have little recall or will describe disjointed and impressionistic memories, like those of patients who have recovered from a prolonged delirium. Recollections of motivation or understanding the cognitive sequence that led to their behaviors are particularly hard to obtain. Other patients have recovered to report an awareness of activity outside of their control, without an ability to respond to others or organize any meaningful communication.

In the developing world, catatonia and other marked disturbances of behavior are more common among those with schizophrenia, and this may be particularly true in Africa. It may be that even the infrequent or inconsistent treatment with antipsychotic medications that patients are exposed to in industrialized societies has a prophylactic benefit for catatonia.

Thought Disorder. Thought disorder is the most studied form of the disorganization symptoms. It is referred to as “formal thought disorder,” or “conceptual disorganization,” in studies that examine cognition or subsyndromes in schizophrenia. It has been said that the thought disorder of schizophrenia expressed in speech is easy to recognize but hard to define, but that has not limited attempts to define it. As a primary care physician uses the fundoscopic examination as a way to visualize the vasculature of the central nervous system, in psychiatry speech is examined as a means to understand the central processes of a patient’s verbal cognition. “Thought disorder” here refers to disorganization of the form of thought and not content. An older use of the term “thought disorder” included the phenomena of delusions and sometimes hallucinations, but this is confusing and ignores the clear differences in the relationships between symptoms that have become apparent over the past 35 years. Thought disorder should be taken to mean formal thought disorders or a disorder of verbal cognition.

In its milder manifestations, thought disorders do not preclude essential communication. Speech is frequently stilted or vague, and sentences may be incomplete. Speech may be persistently vague and have a *paucity of content*, or *tangential,* so that the associative chain moves obliquely off-topic. With the progression of time, or an increased severity of illness, speech can further deteriorate, and inappropriate content can intrude. *Loosening of associations* refers to changes in topic without identifiable links between the subjects and indicates a lack of internal structure for thought expressed as speech. The DSM-5 glossary distinguishes *derailment* as a shift in topics between clauses, where *incoherence* is a shift within clauses. Patients with prominent thought disorders have more confabulation and engage in more symbolic speech. In its most severe form, patients can present with *neologisms* (novel words, and in schizophrenia, taken to indicate that the word has an idiosyncratic or nonsense meaning), *echolalia* (in DSM-5 “pathologic, parrot-like, and apparently senseless repetition [echoing] of a word or phrase”), and incomprehensible speech without meaning or logical internal connections —*word salad.*

There is a recurrent or persistent identifiable thought disorder identified in 20% to 50% of subjects across a variety of studies, and this number goes up to approximately 75% of patients when they present with first episode or acute exacerbations of their illness. Of those subjects who do present with a thought disorder, many will have only mild or moderate disruption, and this can respond to treatment. The possibility that there is a secondary thought disorder, similar to secondary negative symptoms, due to distraction or anxiety that arises during psychotic exacerbations has been poorly explored but may account for some of the improvement seen following these episodes.

There are disturbances of thought or language found in many psychiatric disorders, but the thought disorder in schizophrenia is distinctive. When compared to subjects with bipolar disorders, siblings, and controls, those with schizophrenia are found to be more impaired in comprehension, attention, semantic organization, and fluency and complexity of speech. Prominent thought disorder at illness onset predicts a worse prognosis. It is associated with an earlier onset of illness, increased risk of hospitalization, decreased functional outcomes, higher rates of disability, and more inappropriate social behaviors. Verbal production tends to be less complex in patients with thought disorder. Thought disorder that does not respond well to treatment also predicts a worse course. In one study, only 20% of those subjects with a severe thought disorder at a posthospital follow-up regained a high level of function at a subsequent follow-up visit. There is a modest but consistent association with some measures of formal cognition, including deficits in abstracting ability, mental flexibility, planning, and both response initiation and inhibition. The thought disorders found in schizophrenia tend to be more stable in their severity than hallucinations and delusions, and when prominent, they seem less likely to attenuate in middle age than psychotic symptoms.

Depression and Anxiety

Depression. Most people with schizophrenia will experience significant depression and anxiety during the course of their illness, and this frequency has provoked a discussion about whether these symptoms are comorbidities, or if they should be considered as intrinsic parts of schizophrenia-like psychotic, disorganization, and negative symptoms. Depression is likely not a unitary phenomenon in schizophrenia, and it has been proposed that for some patients it is intrinsic to schizophrenia and for some, it is a reaction to the illness. People with schizophrenia frequently have the experience of simultaneous positive and negative emotional responses to experiences that are viewed as positive by others, and this plausibly initiates and maintains depression. Awareness of depression in schizophrenia is not universal; clinicians and families may not notice depression and anxiety in patients with schizophrenia, or maybe too distracted by patients' psychotic and disorganization symptoms to notice. Anxiety and depression contribute to the disability associated with schizophrenia, and anxiety and depression symptoms can worsen the other core symptoms of schizophrenia. Clinicians should attempt to rule in or rule out these conditions so that if present they can be properly treated.

Depression can be confused with negative symptoms and vice versa, though there is almost no concordance between negative symptoms and depression in schizophrenia. In DSM-5 the diagnosis of major depression is neither specifically excluded nor endorsed, though the reality of depression in schizophrenia is acknowledged in the presence of a depression dimension in the Clinician Rated Dimensions of Psychosis Symptom Severity. Despite the uncertain role for depression symptoms in diagnostic schemes for schizophrenia, when patients are asked about the experience of depression they are usually fully capable of acknowledging its presence. In fact, patients are more likely to endorse their past experiences of depression than their past psychotic experiences. Depression clearly adds to the burden of the illness and impairs functional recovery, is associated with increased relapses, and decreased adherence to treatment. However, the presence of significant depression might also be an indicator of the absence of the deficit syndrome and the long-term disability associated with primary negative symptoms.

Depression is a common symptom in the early course of illness, with one-half or more of all patients experiencing a significant depressed mood during the prodrome and first break. In cross-sectional studies, the prevalence of major depression is fairly low, with rates of around 10%. When symptoms are assessed, instead of a complete depressive syndrome, patients followed long-term after first hospitalization reveal rates of some depressive symptoms of up to 75% or 80%. In one group of patients followed from their first episode, follow-up visits showed depression present in about one-quarter of visits when the subjects were also psychotic; in almost all cases, this depression resolved in parallel with resolution of the psychotic exacerbation. Estimates of the lifetime rates of depression vary widely and are often based on point prevalence studies. Studies with prolonged ascertainment periods find significant depression or a full depressive syndrome in up to 80% of patients.

The presentation of depression in schizophrenia may be masked by the frequent focus in clinical settings on the vegetative symptoms of depression. Disturbances of sleep, appetite, energy, and concentration can be so common in schizophrenia that they become useless in screening for depression in these patients. Also, antipsychotic medications can produce a syndrome that appears identical to major depression, and an increasing emphasis on mood symptoms, poor self-esteem, guilt, and nihilism may be necessary to differentiate a primary mood state from medication effects. The Calgary Depression Scale (Table 12.2–4) emphasizes the psychic symptoms of depression, has good reliability, and has the very clinically useful attributes of reliably distinguishing depression from negative symptoms and detecting mood changes in schizophrenia across the varying phases of illness. This scale should be the first choice for monitoring depression and treatment of depression in schizophrenia.

Table 12.2–4.
Symptoms Assessed in the Calgary Depression Scale for Schizophrenia

1. Depression
2. Hopelessness
3. Self-depreciation
4. Guilty ideas of reference
5. Pathologic guilt
6. Morning depression
7. Early wakening
8. Suicide
9. Observed depression

There has been a common focus on postpsychotic depression in studies of mood disorders in schizophrenia. Postpsychotic depression has been attributed to demoralization and increasing insight following resolution of exacerbations of psychotic symptoms, with patients' increasing insight revealing to them the severity of their illness and the likelihood of a future disability and relapse. Although periods of depression that follow acute exacerbations of symptoms are periods of increased risk of suicide, there is little else to distinguish them from depression that arises during periods of relative stability. At any time in the illness, the combination of depression symptoms and increased insight predicts an increasing sense of distress among patients.

Anxiety. Systematic reviews and meta-analyses of existing studies confirm a high frequency of anxiety disorders in schizophrenia. The presentation of frequent anxiety complaints by patients with schizophrenia is well-known to most clinicians, but the frequency of full syndromes fulfilling diagnostic criteria for specific disorders is probably not as well-known. Anxiety is not surprisingly associated with increases in psychotic symptoms and disorganization of speech and an increase in service utilization. Anxiety disorder diagnoses are not excluded in schizophrenia by DSM-5 criteria, though there is a suggestion that specific phobias are sometimes better explained by delusions, and if generalized worry occurs only during the course of schizophrenia generalized anxiety disorder should not be diagnosed.

Social phobia, panic attacks and panic disorder, and generalized anxiety disorder are common in schizophrenia. Studies of the prevalence of anxiety disorders in schizophrenia find social phobia to be the most common anxiety disorder, with a prevalence of 15%. Life histories of patients with schizophrenia find frequent premorbid social anxiety that continues through the prodrome and first stages of illness and adds considerably to the social, educational, and vocational deficits associated with the illness. There is evidence of increased suicidality, decreased functional recovery, and decreased self-esteem in patients with social phobia, but there is a relative paucity of research about social phobia in schizophrenia as compared to studies of panic.

Panic attacks are more common than panic disorder, and in some patients, the frequency of panic attacks peaks at around the onset of psychotic symptoms. Although there is a wide range reported for the prevalence of panic attacks in schizophrenia, estimates of 30% to 45% of patients experiencing panic at some point in their illness are reasonable. There is also a broad range of estimates of the frequency of panic disorder in schizophrenia, but a recent meta-analysis finds a rate of 11%, as compared to rates of 2% or 3% in the general population. In schizophrenia, panic seems especially related to psychotic symptoms, and there might be feed-forward cycles of persecutory delusions and related hallucinations precipitating panic, which then can worsen psychosis. Unaffected relatives of patients with schizophrenia and panic have increased risks of panic, raising the possibility of a neurobiologic link between panic and schizophrenia.

Generalized anxiety disorder and agoraphobia are also present at higher rates in people with schizophrenia, and in the case of agoraphobia, the rates are markedly higher. There is no increase in the rates of specific phobias, and the rates in schizophrenia might be lower when compared to rates in the general population found in the National Comorbidity Study.

Trauma

There is a high prevalence of post-traumatic stress disorder (PTSD) in schizophrenia, but the relationship between the two diagnoses makes it difficult to tease out causation versus shared risks and increased vulnerability. Adverse events including but not limited to traumatic stress during childhood and adolescence increase the risk of developing schizophrenia, but it is also true that the difficult life situations for many people with schizophrenia put them at an increased risk of experiencing traumatic stressors. It has become clear that psychotic experiences themselves can be traumatic stresses and lead to a diagnosis of PTSD. The circumstances surrounding treatment, especially in the first episode, including forced medication and involuntary treatment, can also act as traumatic stressors. Inpatients and outpatients in public mental health facilities in the Eastern United States were found to have a PTSD rate of about 30% in schizophrenia and related diagnoses in male and female patients; the rates of physical and sexual assault experienced by patients of both genders in this study were very high. A meta-analysis found a rate of PTSD in schizophrenia of 12%, which is higher than the prevalence in the general population. In the Recovery After an Initial Schizophrenia Episode (RAISE) study, which by design was limited to patients with first-episode psychotic disorders, there was a very high rate of trauma found in patients, but rates of actual PTSD were about 5%. Symptoms of PTSD that persist but do not meet full criteria for the diagnosis have been found in up to 80% of patients. The comorbidities that attend PTSD are already increased in schizophrenia, and the combination of PTSD and schizophrenia further amplifies their severity and effect. The core symptoms of schizophrenia are more severe in patients with comorbid PTSD, depression is more common, suicidal ideation and suicide attempts are increased, patients require more care (including requiring more hospitalizations), and it is likely that substance abuse rates are higher in this population.

Obsessions and Compulsions

Possibly because of the relatively well-defined neural circuitry in obsessive-compulsive disorder (OCD) and the overlap in these circuits with some of those implicated in schizophrenia, there has been informed speculation about potential overlap in the underlying pathology of schizophrenia and OCD. It does appear that the prevalence of obsessive-compulsive symptoms (OCS) and OCD in schizophrenia exceed the prevalence in the general population, with estimates of about 25% of people with schizophrenia having comorbid symptoms of OCD, and around 10% qualifying for the diagnosis of OCD. An analysis of CATIE data found about 5% of the patients enrolled had OCD. Ascertainment of OCD can be difficult, with a special difficulty separating delusions and ruminative concern about these delusions from true obsessions.

Comorbid OCD was once considered a good prognostic factor in schizophrenia, but this is clearly not the case. OCS are associated with an earlier age of onset, and OCS or OCD often precedes the onset of psychosis. The presence of OCS or OCD is associated with more severe psychotic symptoms, more depression, more disability, increased hospitalizations, and an increased risk for suicide attempts. As with panic, there is an increased risk for OCD in the relatives of patients with schizophrenia and OCD as compared to relatives with schizophrenia alone. A considerable body of literature has been developed that suggests that for a subset of patients with schizophrenia a real "schizo-obsessive" form of the illness might exist.

Agitation and Hostility

The clinical features of agitation, irritability or frank hostility, uncooperativeness, and impulsivity are familiar to those who deal with clinical populations with schizophrenia, and these symptoms segregate together independent from the influence of culture. These features are associated with increased severity of psychotic symptoms, rehospitalization, and use of high doses of medications. The symptoms are found more commonly in male patients and in those with substantial cognitive impairments. Comorbidities and treatment variables can increase the risk for agitation, with substance abuse particularly prominent. There are other treatment-specific risks, including poor medication adherence, exacerbations of psychosis, akathisia, and other extrapyramidal symptoms (including tardive dyskinesia), that increase the risk of agitation and hostility. Unfortunately, beyond this, little progress has been made in understanding agitation in schizophrenia. The link between agitation and mania has not been adequately explored; intuitively it seems that this might be a subsyndrome, such as psychosis, in which some of the genetic overlaps between schizophrenia and bipolar disorders are expressed. The evidence base that would clarify whether agitation and mania are variable manifestations of a similar underlying process or analogous manifestations of unrelated causes does not exist. Over the last decade, the predominant focus in research into agitation has concerned its links to violence.

Violence. The topic of violence in schizophrenia is contentious. In the mind of the general public schizophrenia is sometimes synonymous with unreasoned violence. For people with schizophrenia, and for their families and advocates, the real problem is that people with schizophrenia are far more likely to be the victims of violence than the perpetrators. This is true, and although there are studies with negative findings, it also remains true that a small minority of people with schizophrenia will commit a disproportionate amount of violence and even murder. Recent systematic reviews and meta-analyses have led to some downward revision in the estimated rates of violence among those with schizophrenia, especially in the absence of psychiatric comorbidities, but the finding of an increased rate remains. Large cohort studies and smaller diagnosis-specific studies, conducted in a variety of countries that have varying baseline rates of violence, show that the presence of a diagnosis of schizophrenia increases the risks of committing violence or being arrested.

No single variable associated with schizophrenia accounts for the increased rates of violence. A series of factors increase risks, though these factors are so prevalent in the population that the predictive power of any one feature is very small. Some demographic and premorbid features are associated with violence, including gender, with men committing the majority of violent acts. Some people with schizophrenia will exhibit a premorbid conduct disorder in childhood and will move into more serious antisocial behaviors as they age. This group is particularly likely to engage in severe violence and to calculate and plan their violent acts. It has been suggested that most of the violence perpetrated by people with schizophrenia arises in those with premorbid conduct disorder or histories of delinquency and premorbid aggression. Paul Appelbaum has asserted that “only a small proportion of violence appears to be directly due” to the symptoms of mental illness. At present, there are no figures for what proportion of the risk of violence is due to antisocial personality disorder, and how much can be attributed to schizophrenia, but clearly, this is not solely a problem attributable to a psychotic disorder.

Recent careful examination of populations with schizophrenia has made it clear that acute stressors, most prominently violent victimization including abuse during childhood or adolescence, increases rates of violence in adults with schizophrenia. A study of a large population in Sweden found that victimization and other acute stressors increased the risks of aggression. This included violent victimization but also stressors like parental bereavement, traumatic brain injuries, and accidental injuries. The effects were clearly more pronounced in people with schizophrenia than in the general population or people with bipolar disorder. Hopefully, the effect was most prominent in the week after the stressor, and the risk was briskly attenuated after that.

Some factors intrinsic to schizophrenia also increase the risks of violence. Poor impulse control and cognitive deficits can predate the onset of the illness but can also arise with the expression of the illness. Insight is variable in schizophrenia but those with significant impairments have an increased risk of violence; prominent disorganization symptoms can also increase risks. Not surprisingly, threat delusions, especially those that provoke negative affects, increase risks of violence, and the deficits in interpreting social and emotional communication common in schizophrenia also increase risks. As reviewed in the discussion of auditory hallucinations, the congruence of the content of hallucinations with a delusional structure that focuses on threat and the benefit of compliance increases the likelihood of following the commands; command hallucinations without this delusional structure pose less threat. Delusions in the absence of hallucinations are often sufficient to provoke violence. For many people with schizophrenia, the persecutory delusions that make them believe their lives are at risk are significant contributors to their violent behaviors, and for those suffering from the delusions violence is an act of self-defense. The influence of paranoid delusions on an increased rate of violence persists even when the results are controlled for the effects of demographic factors and substance abuse. In particular, systematized persecutory delusions that suggest that the person is a particular target, with co-occurring delusions of control, may increase the rates of violence. In this group, the increase in the rate of violence perpetrated by women is greater than the increase in rates among men, and in inpatient psychiatric units the rates of assault by men and women may be equivalent. It has been suggested that confusion, frequently seen in emergency and inpatient settings, contributes to a significant percentage of assaults in these settings. Acute exacerbations of illness are usually times of a higher burden of all symptoms, which also increases the risk of violence.

A broad range of comorbidities also increases the risk of violence. The effects of substance abuse, in particular, the effects of alcohol abuse or dependence parallel the effects of substance abuse in the general population. A widely reported meta-analysis found an odds ratio for violence of almost 9 with comorbid schizophrenia and substance abuse diagnoses, and an odds ratio of 2 with schizophrenia alone. Reporting at the time emphasized the dramatic increase in risk with substance abuse, but the study also clearly showed that there is an increase without the presence of substance abuse, and it poses a clinical and societal problem. Frequently co-occurring with substance abuse, nonadherence with medication treatment is associated with increased rates of violence in people with schizophrenia. Suicide threats and attempts, and neurologic comorbidities, each increases rates of violence.

In chronically ill and disabled patients, there are often less severe acts of violence, frequently directed against family, friends, caregivers, and health care providers. This may be the most common type of assault in treatment settings and frequently occurs as medical staff members attempt to modify patient behavior. These incidents seem more likely related to the effects of impoverished and constricted lives, where there is difficulty for patients to gain experience modulating affect and impulses. Impulsivity and poorly moderated affect can lead to transient displays of aggression, and the effects of substance abuse frequently exacerbate these.

People with schizophrenia are overrepresented in the numbers of people who commit homicide. One estimate puts the percentage of homicide offenders with schizophrenia at greater than 6%; a common estimate is that the risk of homicide is 10 to 20 times higher in schizophrenia. In any location, the risks reflect the community rate of homicide, so that areas with low homicide rates see smaller rates of people with schizophrenia, and the rates rise in areas with higher community rates. The MacArthur study found that violence was more likely to be directed at family members than people outside the family or strangers. Other lines of evidence suggest that there is a considerable risk of stranger homicide by people with schizophrenia. The large majority of these individuals were men; many were in the early stages of their illness with a recent exacerbation of their illness and had a recent change in the quality of their delusions. Homelessness, a lack of psychiatric treatment, conduct disorder, or a history of antisocial behaviors are also associated with stranger homicide.

Suicide

Suicide accounts for a good part of the excess mortality that is usually found in schizophrenia. A commonly cited figure for suicide rate in schizophrenia was 10%, but there is little contemporary evidence to support this figure. Meta-analyses suggest that 5% of people with schizophrenia will commit suicide, with increased risk early in the course of illness–including during the prodrome, in patients with frequent relapses, and at around the time of admission or immediately following discharge from a psychiatric facility. Also, periods of increased psychosis, especially when paranoid delusions are prominent, preserved or emerging insight, and a fear of mental deterioration increase risk. Suicide attempts and suicidal ideation are much more common, with rates of attempts reported to approach 50%. Nonspecific factors such as youth, depression, hopelessness, motor restlessness and agitation, impulsivity, substance abuse, and anxiety, increase the risk for suicide in schizophrenia, as they do in other psychiatric illnesses. People with schizophrenia who attempt suicide tend to use more lethal means in their attempts, have a strong intent to die, and make multiple attempts. The rates of suicide in schizophrenia may be increasing; a review of asylum records from around the turn of the 19th to 20th century found a rate of suicide of about one-twentieth of that in the late 20th century. As noted previously, patients with significant and persistent primary negative symptoms seem to have a lower risk for suicide.

Insight

Lack of insight was the most common symptom of schizophrenia in the IPSS studies of symptoms and course of schizophrenia, and subsequent studies across a wide variety of populations find rates of 50% to 80% of patients with some degree of impairment of insight. Impaired insight has long been recognized as associated with poor social and functional outcomes, lowered quality of life, decreased treatment adherence, increased levels of psychopathology, and recurrent exacerbations of illness and hospitalization. It is clear that better understanding of insight has the potential to generate treatment strategies that significantly improve outcomes—though there is evidence that better insight is associated with increased suicidality, neuroticism and depression, and self-reported distress. For decades, the primary investigators of insight into schizophrenia came from the philosophic tradition of phenomenology, where the aim is to understand as completely as possible the subjective experience of a person; in schizophrenia, the major focus has been on the psychotic symptoms and the nature of the perception and understanding in psychosis. The functional consequences of impaired insight or the neurobiology behind the impairments are less easily appreciated in this approach.

A lack of self-reflection is frequent in schizophrenia, and as is seen in delusions there can be an undue certainty in judgments that are based on insufficient evidence. Splits in insight are often observed in patients, and it is common for chronically ill patients to find someone who accepts that treatment helps to function and prevents worsening of symptoms but who deny illness and the idea that psychotic symptoms are mental phenomena. Poor insight is correlated with negative symptoms, inappropriate affect and thought disorder, and delusions, although the strength of the correlations is fairly weak. Over the last 20 years there has been a broad surge of interest in insight into schizophrenia, continuing the phenomenologic approach but also driven in part by broad emphasis on understanding the neuropsychological underpinnings of schizophrenia, attributable to an emphasis on outcomes over symptoms and also due to the interest of cognitive therapists.

Insight is not a unitary phenomenon and impaired insight is not due to any single underlying process. There is a continuum of insight, from complete understanding of illness and effects of illness on function and production of symptoms, to understanding that auditory hallucinations are associated with illness but nonetheless feeling that the voices are by some external agent, to total denial of illness and symptoms and dysfunction. Cognitive mechanisms certainly play a part, and there are relationships between cognition as measured with neuropsychological testing and the severity of impairment of insight, but results are inconsistent and taken together the relationships are modest. A separate approach addresses “clinical insight” and focuses on those aspects of insight that effect care or are observed in clinical settings. One useful scheme for conceptualizing insight in schizophrenia suggests four dimensions: awareness of illness, awareness of symptoms and attribution of symptoms to illness, awareness of the need for treatment, and awareness of the consequences of the illness. There is separately a recent focus on “cognitive insight,” which actually refers to the metacognitive processes that underlie the ability to reflect on conscious cognition and the certainty underlying cognitive processes. Presumably, there will soon be an elaboration of cognitive therapy techniques to address this aspect of insight.

Illustrative Case

A 37-year-old man presents for assessment. He has a chief complaint of “don’t like my medicine,” though a review of his medication profile shows only irregular refills of the atypical antipsychotic agents and serotonin reuptake inhibitors he has been prescribed over the last year. There are notes by two other psychiatrists in the electronic record, but this record includes only a handful of notes and a pattern of infrequent visits and missed appointments over the last 4 years.

As the appointment proceeds, he explains that he feels as though his current medication “makes me too tired” and “makes me fat” but also acknowledges that he has not been taking the medication “for a while” and that current complaints of an altered sleep–wake cycle, daytime fatigue, increased appetite, and lack of self-initiated hobbies or social activities cannot be attributed to medication effects. He denies depressed mood, but blunted responses of affect and little investment in ideas or hopes for his future are observed. He discloses regular but brief verbal auditory hallucinations of a male and female voice that are familiar to him but he does not know the identity of either voice. He feels as though the voices are entities controlled by some unknown group “who have been watching me for a long time” with cameras, microphones, and airplanes. He feels “they used to want to get me” and do physical harm, but for years they have been focused just on monitoring his whereabouts and “what I’m thinking.” Mild perioral and lingual dyskinetic movements and occasional grimacing are evident, but he is also seen to shift to awkward postures and there are occasional complex repetitive movements of his fingers that are judged to be mannerisms. Hygiene is adequate, but he has not recently shaved, and his hair and clothing are unkempt. Speech is very terse and frequently lacking in pertinent detail, with occasional allusions to ideas that are not clear and cannot be explained. He is aware of the clinical setting, his living situation, and the month and the year but cannot recall the day of the month or the week. After some extensive prodding, he agrees to work with his uncle to get his antipsychotic medications refilled and to follow-up within a month.

On his return visit, there is a dramatic decrease in the rate of ambulation, few spontaneous directed movements, and he has a flat, slack-jawed appearance. There is a clear low-frequency tremor in his hands, more prominent in the right than in the left, and significant stiffness in the elbows bilaterally, but the oral dyskinesias are now absent. Speech production is generally reduced to one- or two-word answers, but he does say that his auditory hallucinations are gone. His uncle accompanies him and confirms good adherence to medication treatment, but although the patient is much slower, his temper seems worse. The patient says that he feels “real bad” and wants to discontinue his medications and wishes he could “go away,” which he indicates means to escape the experience of existence, but he does not want to die or kill himself. He nods assent when asked whether he is depressed. He is prescribed a lower dose of antipsychotic medication and as-needed anticholinergic medications, with follow-up in 2 weeks.

In the interval, copies of his old paper-based chart are retrieved, and there is a more thorough review of his care. He was first seen in a child clinic for presumed depression at 16 years of age, with significant social withdrawal and declining academic performance, along with decreased participation in sports and music classes, disrupted sleep–wake cycle, and frequent irritability. His developmental history indicated no serious medical illnesses, but he tended to be isolative and never more than an adequate student. Social and educational performance deteriorated as his parents went through periods of unemployment and the family frequently moved to stay with various relatives for periods of months to several years. He was treated with antidepressant medications and brief individual and family therapy. His school performance improved slightly, but he remained withdrawn from former friends and interests. After 15 months, he was discharged from care.

He was next seen in the hospital system at 20 years of age. He had graduated from high school but then held a series of entry-level jobs for very short periods and was frequently fired after an inability to get to work on time or inadequate job performance. His parents suspected substance abuse but never found any evidence of illicit drug use. After a half year of getting hired and fired from a series of jobs, he stopped looking for jobs and stayed at home. He was occasionally incomprehensible when explaining his online interests to his parents and became more and more cryptic and more and more irritable with them while his daily rhythms and relationships fragmented. His return to the mental health system was precipitated by an arrest, and then transfer from jail, after he threatened his parents with some broken glass while arguing that they were “helping to poison me.”

During his first admission, he explained that his parents were trying to poison him so that they could steal his ideas; he knew this based on information from people who were simultaneously plotting with his parents and trying to protect him, who communicated with him through radio waves. He was very agitated, and there were also infrequent but odd motor movements in gait, posture, and gesture. He had a slow initial response to a high dose of an atypical antipsychotic agent, and the slow pace of his response prompted an extensive medical workup by his team. The only significant findings were of mild obesity, a modest elevation in fasting glucose, and a magnetic resonance image of the head that showed lateral ventricles judged to be “minimally enlarged for patient’s age.” After 18 days in the inpatient unit, he was discharged from the clinic.

Because of resistance on the part of the patient, follow-up was erratic through most of his twenties. He was seen in several of the other community hospitals in the same city and treated with a variety of antipsychotic, mood-stabilizing, and antidepressant medications. Several years after his first admission and after a period of medication nonadherence and increasing psychosis, he was transferred to an inpatient unit from an intensive care unit, following admission for injuries he suffered from stabbing himself in the neck. Through his twenties, there were two other suicide attempts, the onset of alcohol abuse and heavy cigarette smoking, an inability to find employment, and variable periods of treatment adherence and nonadherence. During one stretch of almost 2 years in his late twenties, he went to a day hospital program and then to a psychosocial rehabilitation program where he made friendly acquaintances with whom he would sometimes go out to eat or go to a movie. Despite these activities, he still spent most of his time at home, watching television, eating, and smoking cigarettes. For some reason he could not explain, he discontinued the medication he had been taking, became progressively more psychotic, complained of recurrent depressed mood, and had delusions of persecution reinforced by the content of his auditory hallucinations. He was again threatening his parents and had his last psychiatric admission. After discharge from this hospital stay, his parents were unwilling to take him back home, and he moved in with his uncle. He has been treated in the same system since the last hospitalization, with tobacco use, obesity, and hyperlipidemia as comorbidities.

At subsequent follow-up visits, a low dose of an antipsychotic is found that causes minimal stiffness and no other perceptible extrapyramidal symptoms. The patient and his uncle are educated on sleep hygiene, and his sleep–wake cycle stabilizes. The depressed mood and wish to not exist expressed at the second visit quickly diminish in parallel with the resolution of his extrapyramidal symptoms. His auditory hallucinations returned, but with discussion, it becomes clear that he usually finds the content of the voices helpful, and that they provide companionship that he likes; he is reluctant to consider trials of other medications because it might silence his voices. His dyskinetic movements return, but the odd postures and gestures are alleviated. Despite urging from family and his psychiatrist, his paucity of external activities and lack of exercise only slowly improve, though his better mood allows him to go out to eat or see movies with a variety of family members. Efforts to return him to a psychosocial rehabilitation program are ongoing.

RESEARCH IMPLICATIONS

Since the last edition of this text, there has been a general adoption of either established factor analytic structures (e.g., the consensus five-factor model of Wallwork and colleagues), or incorporation of study-specific factor analysis in the methodology of studies of symptom associations and novel therapeutics. There remains a relative paucity of research investigation into the nature of some symptoms, particularly in the disorganization and agitation factors, that lags behind efforts in psychosis and negative symptoms. The nomenclature assigned to factors has become progressively more confused as the number of investigators increases. In many cases, factors appear to be labeled with one name or another in order to lay claim to the symptoms as supporting a particular hypothesis. Disorganization symptoms have been called “cognitive” symptoms and general cognition labeled “neurocognition”; but now “neurocognition” has been applied to the disorganization factor. The excitement factor is variably hostility, agitation or “manic” though none of the symptoms included in the factor is specific to mania–and the symptom of grandiosity generally loads on the psychosis factor. As noted earlier, there is a continued use of “positive symptoms” in an undefined fashion–sometimes as the PANSS Positive Factor, sometimes for agitation and confusion in conjunction with psychotic symptoms, and sometimes as a synonym for psychotic symptoms. For clinicians who look to research to enhance their understanding of the illnesses they treat, the nomenclature can make the literature so confusing as to be useless.

In the years immediately before the preparation of this chapter, novel analytic strategies including confirmatory factor analysis, network analysis, and machine learning techniques, have been applied to the assessment of relationships between symptoms and to the relationships of symptoms and other markers of disease. There is not a sufficient number of such studies to make any kind of prediction about the importance of any particular method, but these investigations open the possibility of more detailed dissection of the structure of schizophrenia and schizophrenia’s relationship to other mental illnesses. The opportunities to use phenomenologic approaches to better define the relationship between schizophrenia and bipolar disorders, especially in psychosis and agitation symptoms, and to carry out parallel investigations into concordant (or discordant) nature of negative and disorganization symptoms in schizophrenia and the symptoms of autism spectrum illness have not been exhausted.

FURTHER READINGS

Addington D, Addington J, Maticka-Tyndale E. Assessing depression in schizophrenia: the Calgary Depression Scale. *Br J Psychiatry Suppl*. 1993;22:39–44.

Barch DM, Bustillo J, Gaebel W, et al. Logic and justification for dimensional assessment of symptoms and related clinical phenomena in psychosis: relevance to DSM-5. *Schizophr Res*. 2013;150:15–20.

Berrios GE. Phenomenology, psychopathology and Jaspers: a conceptual history. *Hist Psychiatry*. 1992;3:303–327.

Bleuler E. *Dementia Praecox or the Group of Schizophrenias*. International Universities Press; 1952.

Bleuler E. *Textbook of Psychiatry*. Macmillan; 1924.

Buchanan RW. Persistent negative symptoms in schizophrenia: an overview. *Schizophr Bull*. 2007;33:1013–1022.

Buswell G, Haime Z, Lloyd-Evans B, Billings J. A systematic review of PTSD to the experience of psychosis: prevalence and associated factors. *BMC Psychiatry*. 2021;21:9.

Carpenter WT Jr, Strauss JS, Bartko JS. The diagnosis and understanding of schizophrenia. Part I. Use of signs and symptoms for the identification of schizophrenic patients. *Schizophr Bull*. 1974;11:37–49.

Cassidy RM, Yang F, Kapczinski F, Passos IC. Risk factors for suicidality in patients with schizophrenia: a systematic review, meta-analysis, and meta-regression of 96 studies. *Schizophr Bull*. 2018;44:787–797.

Crow TJ. The two syndrome concept: origins and current status. *Schizophr Bull*. 1985;11:471–486.

Edwards CJ, Cella M, Emsley R, Tarrier N, Wykes THM. Exploring the relationship between the anticipation and experience of pleasure in people with schizophrenia: an experience sampling study. *Schizophr Res*. 2018;202:72–79.

Fazel S, Gulati G, Linsell L, Geddes JR, Grann M. Schizophrenia and violence: systematic review and meta-analysis. *PLoS Med*. 2009;6:e1000120.

Gaebel W, Kerst A, Stricker J. Classification and diagnosis of schizophrenia or other primary psychotic disorders: changes from ICD-10 to ICD-11 and implementation in clinical practice. *Psychiatria Danubiana*. 2020;32:320–324.

Garrett M, Silva R. Auditory hallucinations, source monitoring, and the belief that "voices" are real. *Schizophr Bull*. 2003;29:445–457.

Harrow M, Jobe TH. How frequent is chronic multiyear delusional activity and recovery in schizophrenia: a 20-year multi–follow-up. *Schizophr Bull*. 2010;36:192–204.

Jaspers K. *General Psychopathology*. Hoenig J, Hamilton MW, trans-eds. University of Chicago Press; 1968.

Kay SR, Opler LA, Lindenmayer JP. Reliability and validity of the positive and negative syndrome scale for schizophrenics. *Psychiatry Res*. 1988;23:99–110.

Kirkpatrick K, Buchanan RW, Ross DE, Carpenter WT Jr. A separate disease within the syndrome of schizophrenia. *Arch Gen Psychiatry*. 2001;58:165–171.

Kraepelin E. *Dementia Praecox and Paraphrenia*. Livingstone; 1919.

Lim A, Hoek HW, Deen ML, et al. Prevalence and classification of hallucinations in multiple sensory modalities in schizophrenia spectrum disorders. *Schizophr Res*. 2016;176:493–499.

Nayani TH, David AS. The auditory hallucination: a phenomenological survey. *Psychol Med*. 1996;26:177–189.

Overall JE, Gorham DR. The brief psychiatric rating scale. *Psychol Rep*. 1962;10:799.

Rabinowitz J, Berardo CG, Bugarski-Kirola D, Marder S. Association of prominent positive and prominent negative symptoms and functional health, well-being, healthcare-related quality of life and family burden: a CATIE analysis. *Schizophr Res*. 2013;150:339–342.

Rund BR. A review of factors associated with severe violence in schizophrenia. *Nord J Psychiatry*. 2018;72:561–571.

Schneider K. *Clinical Psychopathology*. Grune & Stratton; 1959.

Spitzer RL, Endicott J, Robins E. *Research Diagnostic Criteria*. Biometrics Research Division, New York State Psychiatric Institute; 1975.

Strauss JS, Carpenter WT Jr, Bartko JJ. The diagnosis and understanding of schizophrenia. Part III. Speculations on the processes that underlie schizophrenic symptoms and signs. *Schizophr Bull*. 1974;1:61–69.

Strauss GP, Esfahlani FZ, Galderisi S, et al. Network analysis reveals the latent structure of negative symptoms in schizophrenia. *Schizophr Bull*. 2019;45:1033–1041.
Wallwork RS, Fortgang R, Hashimoto R, Weinberger DR, Dickinson D. Searching for a consensus five-factor model of the Positive and Negative Syndrome Scale. *Schizophr Res*. 2012;137:246–250.

12.3 Global Burden of Schizophrenia

Daniel Mamah, M.D., M.P.E.

The global burden of disease (GBD) study was initiated by the World Health Organization (WHO) in the early 1990s, in collaboration with the World Bank and the Harvard School of Public Health. GBD 1990, as it was termed, produced prevalence estimates from 107 diseases and injuries over eight world regions and introduced for the first time a new integrated unit of disease burden measurement, the disability-adjusted life year (DALY). One DALY corresponds to one year of healthy life lost in a given population due to the combined effects of disability and premature mortality resulting from disease. Thus, for any given disease:

$$\text{DALY} = \text{YLL} + \text{YLD}$$

Where YLL refers to the years of life lost to mortality and YLD is the years of (healthy) life lost to disability.

Updated GBD studies were conducted by the WHO over the next two decades. Since 2010, subsequent iterations of the GBD study have been done through the Institute of Health Metrics and Evaluation (IHME) based at the University of Washington, expanding the number of disorders included and using an improved methodology. The GBD 2019 study estimated metrics due to 369 diseases and injuries for 204 countries and territories. Worldwide, several psychiatric conditions, including self-harm, depressive disorders, anxiety disorders, conduct disorder, and drug use disorders, were among the top 20 conditions accounting for total DALYs across all ages combined. Among 25 49-year-olds, schizophrenia ranked 22nd in global DALYs, accounting for 1.5% of the total. The findings on schizophrenia reflect the burden of morbidity and disability that the disorder inflicts on individuals, families, and communities worldwide and underestimates the burden of premature mortality, which is difficult to ascertain. Global DALYs have remained largely stable from 1990 to 2019, and with correcting for population growth and aging, there has been a noticeable improvement in general health. Global disease burden has also continued to shift away from communicable to noncommunicable diseases and injury, and from premature death to years lived with disability. For most diseases, changes in rates of death and disability have been relatively stable or slow; however, regarding behavioral disorders, it is notable that death from drug use disorders has risen significantly since the last decade. Over half of the global deaths from a drug overdose in 2019 occurred in the United States. The prevalence of depressive or other mental disorders has appeared largely stable over the last three decades; however, self-harm had a significant decline of 28.4% during this time period.

This chapter reviews evidence from past and present research, bearing on the individual components that jointly make up the aggregate burden of disease and disability associated with schizophrenia—the variations in its incidence and prevalence, course and outcome, comorbidity and mortality, and socioeconomic costs. In evaluating the evidence, it is important to bear in mind that schizophrenia still remains a broad clinical syndrome (rather than a “disease”) and that the neurobiology and genetics underlying its manifestations are not yet fully understood.

However, research into the genetic underpinnings of schizophrenia recently reached a new milestone with the genome-wide association study (GWAS) of the largest sample to date of patients and healthy controls (36,989 cases and 113,075 controls), which identified 108 associated genomic loci significantly associated with schizophrenia risk. The study, by an international consortium of about 80 research groups, found robust associations of the disorder with the dopamine receptor DRD2, as well as with molecules involved in glutamatergic neurotransmission, suggesting plausible pathogenetic pathways. A surprising novel discovery was the strongest association of schizophrenia risk with the major histocompatibility complex on chromosome 6, pointing to an important, yet poorly understood, role of the immune system in the pathophysiology of the disorder. However, these findings still explain a relatively small fraction (~30%) of the estimated heritability of schizophrenia and can be seen as the prelude to the next-generation whole-genome sequencing studies. Such studies will have the potential to identify rare genomic variants (mutations present in less than 2% to 5% of the population) that may contribute to familial and individual risk of the disorder.

The clinical concept underlying the present diagnostic classifications of schizophrenia does not indicate whether one or several pathologic processes are involved or whether the causes of schizophrenia are the same in all cases. These gaps in the biologic characterization of schizophrenia reflect its complexity, as well as the currently incomplete understanding of the basic neurobiology underpinning key psychological functions that are impaired in schizophrenia, such as memory, multimodal sensory integration, self-monitoring, and intentionality of behavior.

This chapter builds on text from earlier editions of this textbook, authored by Assen Jablensky. Throughout the chapter, the terms *developing countries* and *developed countries* are used to refer to a broad (and somewhat fluid) grouping of societies still characterized by traditional community structures, beliefs and values, and a relatively recent transition to industrialization and large-scale urbanization, and societies that have undergone the transition to “modernity” much earlier, with all the consequences for community life, family support patterns, and systems of care for the sick and disabled.

SCHIZOPHRENIA BEFORE AND AFTER KRAEPELIN

Since its early formulation in 1896 by Emil Kraepelin (1856 to 1926), the concept of *dementia praecox,* later renamed schizophrenia by Eugen Bleuler (1857 to 1939), has been a hotly disputed diagnostic entity, raising, among other things, the question of its applicability to different populations and cultures. Although the Western medical literature before the 18th century contains unequivocal descriptions of depressive, manic, hysterical, and organic-dementing clinical pictures, it provides hardly any explicit reference to what later became known as schizophrenia. A plausible early description of schizophrenia can be attributed to Philippe Pinel (1745 to 1826), who described “frequent or continual appearance of unconnected ideas and shallow dissociated emotions; disorganized movements and extravagant actions…a kind of automatic existence.” However, for most of the 19th century, such occasional descriptions of schizophrenic disorders were completely overshadowed by the all-embracing notions of “melancholia” and “mania.” There are at least two possible explanations for the underestimation of schizophrenic disorders. One, less likely, is that until the second half of the 19th century, schizophrenia indeed was a rare disease and that the significant increase of the share of schizophrenia in treated psychiatric morbidity did not occur until the last decades of the 19th century. Statistics on patients admitted to the University Psychiatric Clinic in Munich in 1908—some of them diagnosed personally by Kraepelin—indicate that as late as 10 years after the establishment of the diagnostic category of *dementia praecox,* the frequency of this diagnosis among all first admissions was as low as 9.1% in men and 7.3% in women. However, little was known at that time about the prevalence of untreated cases in the community. A second possible explanation is that schizophrenia did exist and was frequent, but for most of the 19th century, it was assimilated with other syndromes, “mania” and “melancholia” in particular. There was no phenomenologic description of schizophrenia as such because the concept did not exist, although all the symptoms and signs that were to be integrated later into the diagnostic category had been present all the time. Social factors likely also played a role. Diminishing tolerance in growing urban communities to bizarre behavior and deviance, and the commitment of increasing numbers of “insane” persons to the mental hospital where “scientific” diagnosis and treatment were believed to be practiced, may have facilitated the nosographic description of new clinical entities. According to statistics from Austria, the community census of “insane persons” rose slightly from 0.94 per to 1.35 per 1,000 population between 1873 and 1905, while the proportion of such persons institutionalized in hospitals during the same period increased from 32.5% to 61.9%. Such unprecedented concentration of psychotic persons in medical institutions undoubtedly made it easier for clinicians to identify syndromes, subtypes, and other varieties of mental disorders.

Comparative Psychiatry and Ethnopsychiatry

During the same period, there was a growing interest in the manifestations and frequency of mental disorders, primarily psychoses, in non-Western cultures. British, Dutch, and French colonial administrations imported into their overseas dependencies the asylum model of care for the mentally ill and built mental hospitals for patients with intractable chronic illness, in particular, including “dangerous” psychotics. In 1903, Kraepelin traveled to Java and spent several weeks at the Buitenzorg (now Bogor) Hospital. He came to the conclusion that the basic forms of *dementia praecox* and *manic-depressive insanity* in the Javanese were essentially the same as those in Europe, though “racial characteristics, religion, and customs” modified their clinical manifestations. Kraepelin anticipated “rich rewards” from the capacity of cross-cultural studies to “throw light on the causes of mental disorder” and proposed *comparative psychiatry* as the systematic study of mental disorders and personality traits across cultures. However, for decades, the epidemiology of schizophrenia and other psychotic disorders in non-European cultures remained restricted to mental hospital statistics or rudimentary field surveys undertaken by colonial public health administrators. His conclusion was that “true” schizophrenia and depression were rare among indigenous peoples and that chronically ill psychotic people were well tolerated in the community. The ideologic framework in which such early *ethnopsychiatric* observations were embedded varied between benign paternalism and overt racism. In a 1953 monograph, John Carothers (1903 to 1989) claimed that the paucity of structured delusional contents and the lack of systematization of delusions in “the African” could be explained by congenital underdevelopment of the frontal lobes of the brain. Similar assumptions led other authors to conclude that depression was rare in sub-Saharan Africa or in Asia because of the lack of Judaeo-Christian cultural values that made the experience of guilt possible.

The sketchy and often distorted picture of the epidemiology of psychoses in developing countries began to change in the postcolonial era when locally born psychiatrists, educated in the West, entered practice and teaching in their countries. Although trained in the colonial metropolis, they were keen to understand the nature of mental disorders in their own cultures and to introduce culturally appropriate alternatives to the colonial mental hospital, such as the model “psychiatric village” in Ibadan, Nigeria, founded by Thomas A. Lambo (1923 to 2004), the first Western-trained psychiatrist in Africa. In Asia, Pow Meng Yap (1921 to 1971) systematically charted the so-called *culture-bound syndromes,* highlighting their differentiation from schizophrenia. Epidemiologic surveys providing data on the incidence and prevalence of schizophrenia were carried out in Mauritius, Nigeria, India, Taiwan, and mainland China. Since the 1960s, such studies expanded to include a variety of populations in

Africa, Asia, and South America, about which little had been known previously.

The comparative study of schizophrenia across populations and cultures gained in scope and momentum with the research program initiated by WHO in the 1960s. Two multicenter collaborative studies, the International Pilot Study of Schizophrenia (IPSS) and the Ten-Country Study (Determinants of Outcome of Severe Mental Disorders) generated cross-sectional and follow-up data on more than 3,000 cases of schizophrenia in 16 geographical areas in 12 countries in Africa, the Americas, Asia, and Europe. These studies introduced standardized assessment instruments, research procedures, and explicit diagnostic criteria in data collection at community and hospital facilities by local psychiatrists and mental health workers who had been trained to use such research tools and participated in their development. The WHO studies enabled for the first time direct comparisons of the population rates, psychopathology, and burden of disease associated with schizophrenia across various cultures.

DIAGNOSTIC COMPARABILITY OF THE SCHIZOPHRENIA PHENOTYPE ACROSS POPULATIONS

There is, as yet, no validated biomarker that could assist in the diagnosis of schizophrenia. Neuroimaging and EEG can only help in ruling out other psychosis secondary to other medical conditions, such as head injury, seizures, or mass-occupying lesions. The paradigm-transformative DSM-III and subsequently ICD-10, DSM-IV, and DSM-5 are descendants of the Kraepelinian nosology and research work at Washington University developing criterion-based diagnostic criteria for mental disorders, determined using specific clinical description and exclusion from other disorders, as well as outcome and family studies. However, there is little evidence that most currently diagnosed mental disorders, including psychoses, are separated by natural boundaries. There is a growing understanding that the syndromal spectrum of schizophrenia is the end-point for heterogeneous gene networks, pathophysiologic pathways, and environmental modifiers. Increasing experimentation with research-oriented models and cross-disciplinary methodologies linked to the emerging Research Domain Criteria (RDoC) may eventually reshape diagnostic classification, but this is a long-term project.

Despite the wide acceptance of the DSM and ICD criteria that promote reliable diagnostic identification, the question of whether cases diagnosed as schizophrenia in different populations and cultures are phenotypically homologous is of critical importance, considering that the biologic basis of the disorder is largely unknown and no objective diagnostic test is currently available. Phenotypic comparability implies that the constellation of symptoms is coherent and replicable and that course, outcome, and response to treatment are broadly predictable.

In the past, the belief was widespread that schizophrenia was a “Western” disease and had no counterpart in indigenous populations untouched by modern technology and lifestyles. That this belief was mistaken is demonstrated by the fact that no human community has yet been found to be free of schizophrenia, diagnosable by commonly accepted criteria. Although no single symptom is *pathognomonic* of the disorder, the overall clinical presentation of schizophrenia is similar across cultures. Acutely ill patients in different cultural settings describe the same characteristic symptoms, such as hallucinatory voices commenting in the third person on their thoughts and actions, thoughts being broadcast or taken away, or their surroundings being imbued with special meaning. Negative symptoms, such as psychomotor poverty, social withdrawal, and amotivation, are common, irrespective of the cultural setting. The conclusion that patients diagnosed with schizophrenia in different cultures suffer from the same disorder is further supported by similar age- and sex-specific distribution of the onset of symptoms, which tends to peak in early adulthood. Similarities being present despite the variety of social norms and beliefs about illness across cultures suggests some commonalities in the underlying neurobiology. Nevertheless, variations exist that may affect the recognition and treatment of schizophrenia. For example, Nigerian psychiatrists have described in patients with schizophrenia a characteristic symptom complex consisting of anxiety, depression, vague hypochondriacal symptoms, bizarre magicomystical ideas, episodic twilight or confusional states, atypical depersonalization, emotional liability, and retrospective falsification of memory based on hallucinations or delusions. Certain variants of the schizophrenia syndrome, such as an acute onset form and a catatonic subtype, appear to be more common in traditional rural communities. Research conducted within the WHO program of cross-cultural schizophrenia studies has confirmed the existence of population differences in the mode of presentation of the disorder. For instance, acute onset was found to characterize in 40.3% and marked catatonic symptoms in 10.3% of cases assessed at research centers in India and Nigeria, compared to 10.9% and 1.2% at European and North American centers.

Differentiation from Other Disorders

A common clinical problem in developing countries is the differentiation of schizophrenia from psychoses due to infectious or parasitic diseases. African trypanosomiasis often results in a symptomatic psychosis that has a slow, insidious onset and may mimic schizophrenia. Since a variety of infectious, parasitic, and nutritional diseases are endemic in the developing world, it has been suggested that a large proportion of the cases of schizophrenia in those populations may, in fact, be symptomatic psychoses associated with tropical communicable diseases. Another brain pathology known to contribute to psychoses with schizophrenia-like features is temporal lobe epilepsy. The association between epilepsy and a chronic, interictal psychosis that is difficult to distinguish from schizophrenia has been well documented clinically. Studies in Europe and North America suggest a nearly 10-fold increase in the risk of schizophrenia-like psychosis among people suffering from epilepsy. Considering that epilepsy is highly prevalent in many regions, undiagnosed and untreated epilepsy may account for a proportion of the cases of schizophrenia-like psychoses. It is unlikely, however, that either communicable diseases or organic brain disorders such as epilepsy account for more than a small fraction of the cases presenting as schizophrenia. Among some 5,000 individuals with psychotic illnesses who were screened in India and Nigeria for inclusion in the WHO Ten-Country Study, only 11.7% were excluded on the grounds of having a physical disease that might explain their psychotic symptoms. Thus, problems of differential diagnosis in respect of organic brain diseases may arise in individual cases, but the majority of schizophrenic disorders in the Third World cannot be attributed to somatic disease or gross brain pathology. Considering all available evidence, schizophrenic disorders in non-Western populations can be reliably distinguished from acute transient psychoses, culture-bound syndromes, and probably affective disorders, although the boundary with the latter has not been adequately explored, and symptomatic overlap may exist. Family morbidity data for developing countries are still scarce, but where such information is available, it suggests that genetic factors appear to contribute to the transmission of schizophrenia in the same way as in developed countries.

VARIATIONS IN THE INCIDENCE AND PREVALENCE OF SCHIZOPHRENIA

No population or culture has been identified to date in which schizophrenia does not occur. The rates of its occurrence are broadly comparable, provided that the populations being compared are large enough to minimize the effects of small-area variation. This certainly does not imply that the incidence of the disorder is uniform across all populations but rather that the range of variation is modest in comparison with other complex or multifactorial diseases, such as noninsulin-dependent diabetes, ischemic heart disease, or multiple sclerosis, in which 10- to 30-fold differences among populations are not uncommon.

Prevalence

The prevalence of a disease is the estimated number of cases per 1,000 persons at risk present in a population at a given time (point prevalence) or over a defined period (period prevalence). Since schizophrenia cases in prolonged remission are likely to be missed in point prevalence surveys, it is useful to estimate the lifetime prevalence by supplementing the assessment of the present mental state with data about past episodes of the disorder. An overview of selected general population prevalence studies of schizophrenia and related disorders is presented in Table 12.3–1. These studies differ in many aspects of methodology but are generally characterized by a high intensity of case-finding. A systematic review of 188 studies in 46 countries, published between 1965 and 2002, estimated the median value for point prevalence at 4.6 per 1,000 persons, period prevalence at 3.0 per 1,000 persons, and lifetime prevalence at 7.2 per 1,000. The median combined prevalence estimates were also lower in developing nations (2.6 per 1,000) compared to developed countries (3.3 per 1,000) or emerging economy countries (4.7 per 1,000) in this review. Estimates from the GBD 2016 study, which used a Bayesian meta-regression tool to derive prevalence estimates for 195 countries and territories, reported a global age-standardized point prevalence of schizophrenia to be 0.28%, which did not vary widely across countries. However, China showed the highest age-standardized prevalence estimates of schizophrenia (0.42%), and the Netherlands is the highest in Europe (0.36%). The lowest mean prevalence rates were found across Africa and the Middle East (0.16% to 0.22%). Estimates, however, are less reliable considering the very limited high-quality epidemiologic data from low- and middle-income countries.

The 2010 Australian National Survey of psychosis covered a population of 1,464,923 people aged 18 to 64 years, which was about 10% of all Australians in that age range. The study employed a two-phase design: a psychosis screen of all users of mental health services was followed by a random selection of 1,825 people who were interviewed using a standardized schedule. In addition, two cognitive tests were employed, and all participants provided blood samples for a screen of metabolic abnormalities. The study found the 1-month prevalence of schizophrenia to be 3.10 cases per 1,000 population at risk and the estimated lifetime morbid risk at 3.45 per 1,000. The mean cognitive ability of the participants was 1.6 SD below the population average, but higher scores on the cognitive tests were significantly associated with better independent functioning. A striking finding was the high prevalence of the metabolic syndrome, affecting 60.8% of participants (65.9% were current smokers; 47.4% were obese; 32.4% were sedentary; 50.8% had a lifetime history of cannabis dependence; and 49.8% were chronic alcohol users). These alarming findings pointed to an urgent need for physical health interventions, cognitive remediation therapies, and workforce planning focusing on supported employment.

Incidence

The incidence rate is the estimated annual number of first-onset cases in a defined population per 1,000 persons. Incidence is of greater interest for risk-factor epidemiology than prevalence since it is interpretable in terms of the probability of disease occurrence in a given population and is temporally closer to the antecedents of the disease or its precipitating factors. However, its estimation depends critically on the capacity to determine the point of the disorder reliably—a requirement that is difficult to meet for schizophrenia due to its often long prodromal period and the fuzzy boundary between premorbid state and psychosis onset. Since objective biomarkers of the disease process are still lacking, onset is usually defined as the point when clinical manifestations become recognizable and can be diagnosed according to specified criteria. The first hospital admission, which has been used as a proxy for disease onset in many studies, is not a robust indicator because of the variable time lag that may exist in treatment facilities between the earliest appearance of symptoms and hospitalization (and many cases do not get hospitalized). A better approximation is provided by the time point of the person's *first contact* with any psychiatric, general medical, or alternative aid agency for reasons likely related to emerging psychosis. Such estimates are less reliable in developing countries with limited access to mental health care. Clearly, both first-admission and first-contact studies share the limitation of measuring "treated" incidence and risk omitting cases that do not present for assessment or treatment. This limitation can be overcome by periodically repeated door-to-door surveys of the same population or by longitudinal cohort studies (though both are difficult to mount for reasons of cost and logistics).

A systematic review of incidence data from some 160 studies from 33 countries, published between 1965 and 2001, yielded a median value of 0.15 and mean value of 0.24 per 1,000, with a fivefold range of the rates and a tendency for recent studies to report lower rates. A more recent systematic review and meta-analysis of studies published between 2002 and 2017 found a median incidence of schizophrenia to be 0.22 per 1,000, with high variation across studies. The same study indicated a pooled incidence of 0.27 per 1,000 for psychotic disorders. A significant portion of the variation highlighted can be attributed to methodologic differences and confounders, including the type of study sample and design, coverage of case findings (hospital-based, field surveys, case registers, birth cohorts), diagnostic practices, sample size, statistical power, and methods of data analysis. For example, birth cohort studies and case registers tend to produce higher rates than cross-sectional community surveys and hospital admission studies. Larger studies often estimate higher incidence rates compared to smaller studies. Furthermore, first-contact studies often lead to lower incidence rates as they include standardized diagnostic assessments, which can reduce the number of false positive results. Notwithstanding such limitations, real variation is undoubtedly present in schizophrenia (as in any human disease). The interesting question concerns the extent and sources of this variation and whether its study adds any clues to the understanding of schizophrenia etiology.

Table 12.3–1. Selected General Population Prevalence Studies of Schizophrenia

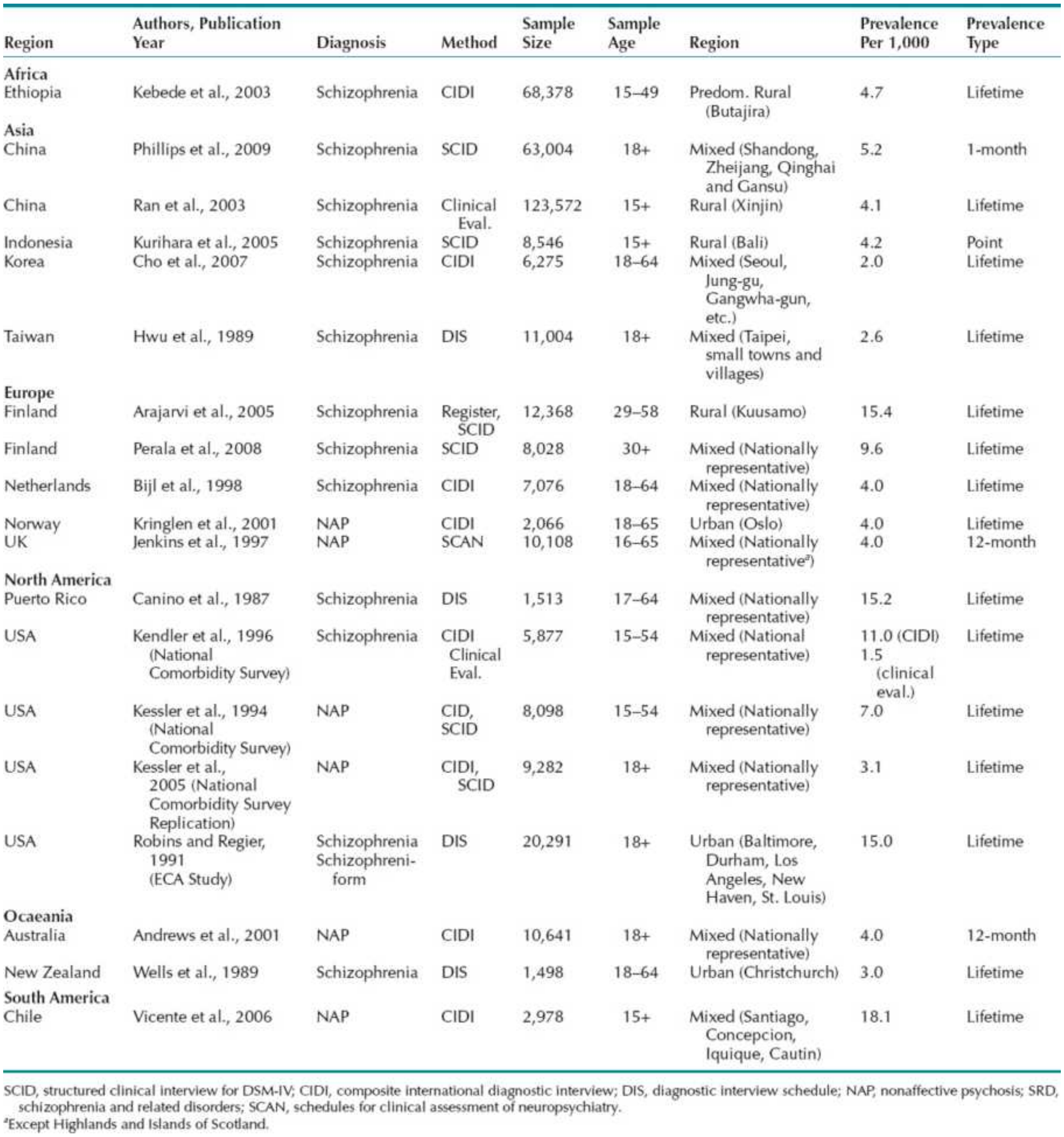

Region	Authors, Publication Year	Diagnosis	Method	Sample Size	Sample Age	Region	Prevalence Per 1,000	Prevalence Type
Africa								
Ethiopia	Kebede et al., 2003	Schizophrenia	CIDI	68,378	15–49	Predom. Rural (Butajira)	4.7	Lifetime
Asia								
China	Phillips et al., 2009	Schizophrenia	SCID	63,004	18+	Mixed (Shandong, Zheijang, Qinghai and Gansu)	5.2	1-month
China	Ran et al., 2003	Schizophrenia	Clinical Eval.	123,572	15+	Rural (Xinjin)	4.1	Lifetime
Indonesia	Kurihara et al., 2005	Schizophrenia	SCID	8,546	15+	Rural (Bali)	4.2	Point
Korea	Cho et al., 2007	Schizophrenia	CIDI	6,275	18–64	Mixed (Seoul, Jung-gu, Gangwha-gun, etc.)	2.0	Lifetime
Taiwan	Hwu et al., 1989	Schizophrenia	DIS	11,004	18+	Mixed (Taipei, small towns and villages)	2.6	Lifetime
Europe								
Finland	Arajarvi et al., 2005	Schizophrenia	Register, SCID	12,368	29–58	Rural (Kuusamo)	15.4	Lifetime
Finland	Perala et al., 2008	Schizophrenia	SCID	8,028	30+	Mixed (Nationally representative)	9.6	Lifetime
Netherlands	Bijl et al., 1998	Schizophrenia	CIDI	7,076	18–64	Mixed (Nationally representative)	4.0	Lifetime
Norway	Kringlen et al., 2001	NAP	CIDI	2,066	18–65	Urban (Oslo)	4.0	Lifetime
UK	Jenkins et al., 1997	NAP	SCAN	10,108	16–65	Mixed (Nationally representative[a])	4.0	12-month
North America								
Puerto Rico	Canino et al., 1987	Schizophrenia	DIS	1,513	17–64	Mixed (Nationally representative)	15.2	Lifetime
USA	Kendler et al., 1996 (National Comorbidity Survey)	Schizophrenia	CIDI Clinical Eval.	5,877	15–54	Mixed (National representative)	11.0 (CIDI) 1.5 (clinical eval.)	Lifetime
USA	Kessler et al., 1994 (National Comorbidity Survey)	NAP	CID, SCID	8,098	15–54	Mixed (Nationally representative)	7.0	Lifetime
USA	Kessler et al., 2005 (National Comorbidity Survey Replication)	NAP	CIDI, SCID	9,282	18+	Mixed (Nationally representative)	3.1	Lifetime
USA	Robins and Regier, 1991 (ECA Study)	Schizophrenia Schizophreniform	DIS	20,291	18+	Urban (Baltimore, Durham, Los Angeles, New Haven, St. Louis)	15.0	Lifetime
Ocaeania								
Australia	Andrews et al., 2001	NAP	CIDI	10,641	18+	Mixed (Nationally representative)	4.0	12-month
New Zealand	Wells et al., 1989	Schizophrenia	DIS	1,498	18–64	Urban (Christchurch)	3.0	Lifetime
South America								
Chile	Vicente et al., 2006	NAP	CIDI	2,978	15+	Mixed (Santiago, Concepcion, Iquique, Cautin)	18.1	Lifetime

SCID, structured clinical interview for DSM-IV; CIDI, composite international diagnostic interview; DIS, diagnostic interview schedule; NAP, nonaffective psychosis; SRD, schizophrenia and related disorders; SCAN, schedules for clinical assessment of neuropsychiatry.

[a] Except Highlands and Islands of Scotland.

Table 12.3–2.

Age-Standardized Schizophrenia Incidence and DALY Rates from the Global Burden of Disease Study of 1990 and 2017

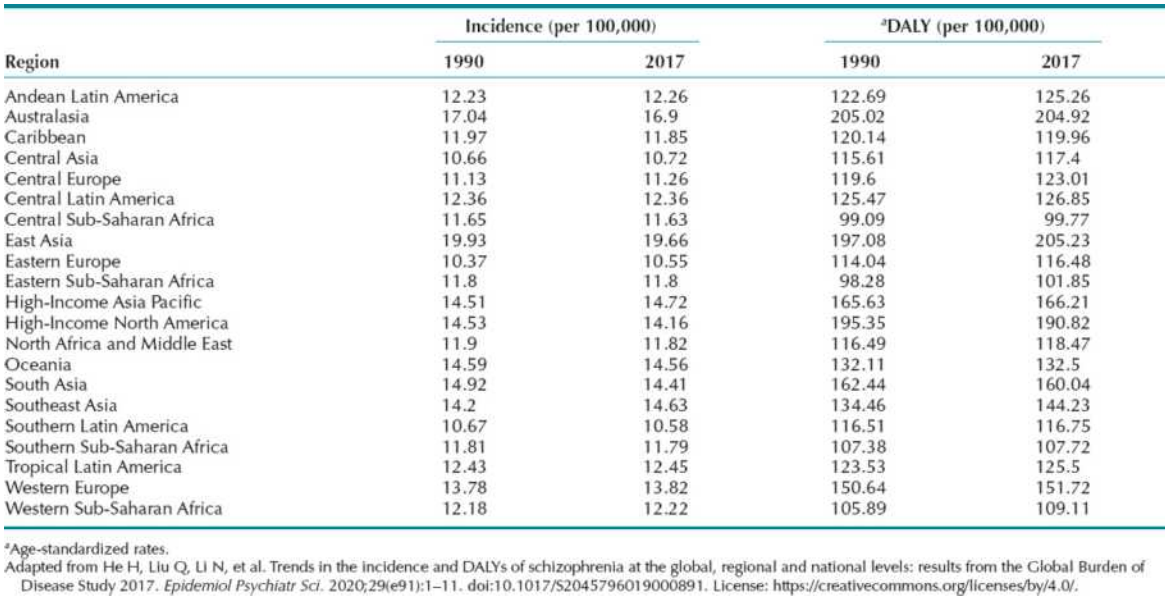

Region	Incidence (per 100,000)		[a]DALY (per 100,000)	
	1990	2017	1990	2017
Andean Latin America	12.23	12.26	122.69	125.26
Australasia	17.04	16.9	205.02	204.92
Caribbean	11.97	11.85	120.14	119.96
Central Asia	10.66	10.72	115.61	117.4
Central Europe	11.13	11.26	119.6	123.01
Central Latin America	12.36	12.36	125.47	126.85
Central Sub-Saharan Africa	11.65	11.63	99.09	99.77
East Asia	19.93	19.66	197.08	205.23
Eastern Europe	10.37	10.55	114.04	116.48
Eastern Sub-Saharan Africa	11.8	11.8	98.28	101.85
High-Income Asia Pacific	14.51	14.72	165.63	166.21
High-Income North America	14.53	14.16	195.35	190.82
North Africa and Middle East	11.9	11.82	116.49	118.47
Oceania	14.59	14.56	132.11	132.5
South Asia	14.92	14.41	162.44	160.04
Southeast Asia	14.2	14.63	134.46	144.23
Southern Latin America	10.67	10.58	116.51	116.75
Southern Sub-Saharan Africa	11.81	11.79	107.38	107.72
Tropical Latin America	12.43	12.45	123.53	125.5
Western Europe	13.78	13.82	150.64	151.72
Western Sub-Saharan Africa	12.18	12.22	105.89	109.11

[a] Age-standardized rates.

Adapted from He H, Liu Q, Li N, et al. Trends in the incidence and DALYs of schizophrenia at the global, regional and national levels: results from the Global Burden of Disease Study 2017. *Epidemiol Psychiatr Sci.* 2020;29(e91):1–11. doi:10.1017/S2045796019000891. License: https://creativecommons.org/licenses/by/4.0/.

Table 12.3–2 presents the age-standardized incidence rates (ASIR) and accompanying age-standardized DALY rates (ASDR) across world regions in 1990 and 2017, based on the GBD 2017 study. The global ASIR (per 100,000) decreased only slightly from 14.98 in 1990 to 14.39 in 2017, while the ASDR (per 100,000) remained largely stable from 157.01 to 156.45. ASIR estimates were lowest in Eastern Europe and sub-Saharan Africa and highest in East Asia. ASDR estimates were lowest in sub-Saharan Africa and highest in East Asia. Figures 12.3–1 and 12.3–2 show the ASIR and ASDR global estimates, respectively, based on the GBD 2017 study. The ASIR and ASDR were highest in China, the Netherlands, and Australia.

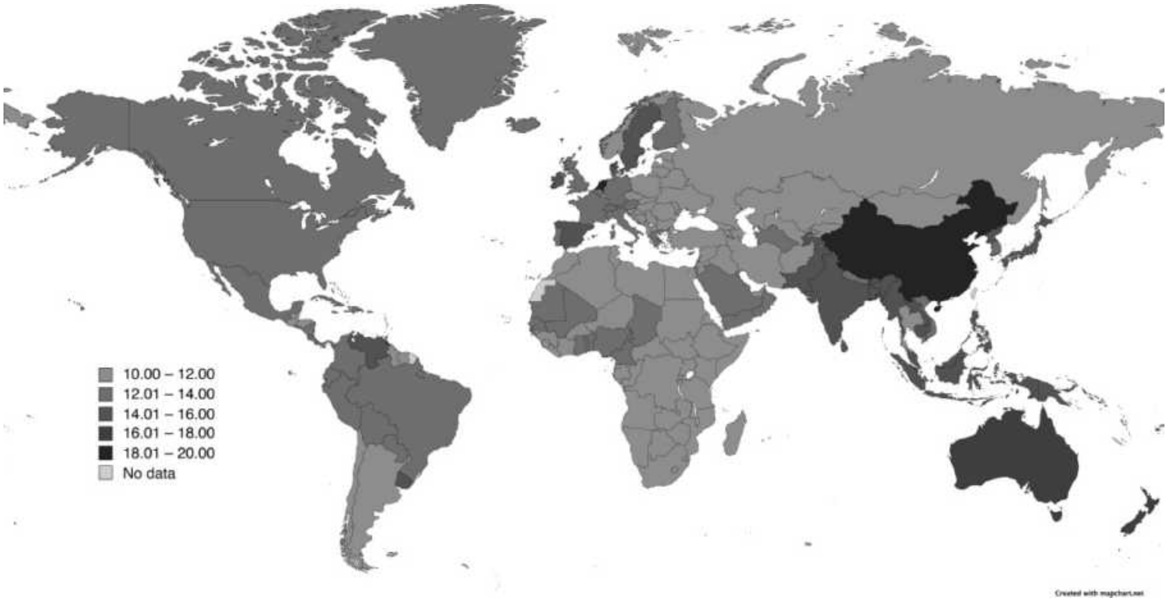
10.00 – 12.00
12.01 – 14.00
14.01 – 16.00
16.01 – 18.00
18.01 – 20.00
No data
Created with mapchart.net

FIGURE 12.3–1. Age-standardized schizophrenia incidence rates (ASIR) globally in 2017, estimated from the GBD 2017 study. (Data from He H, Liu Q, Li N, et al. Trends in the incidence and DALYs of schizophrenia at the global, regional and national levels: results from the Global Burden of Disease Study 2017. *Epidemiol Psychiatr Sci.* 2020;29(e91):1–11. doi:10.1017/S2045796019000891.)

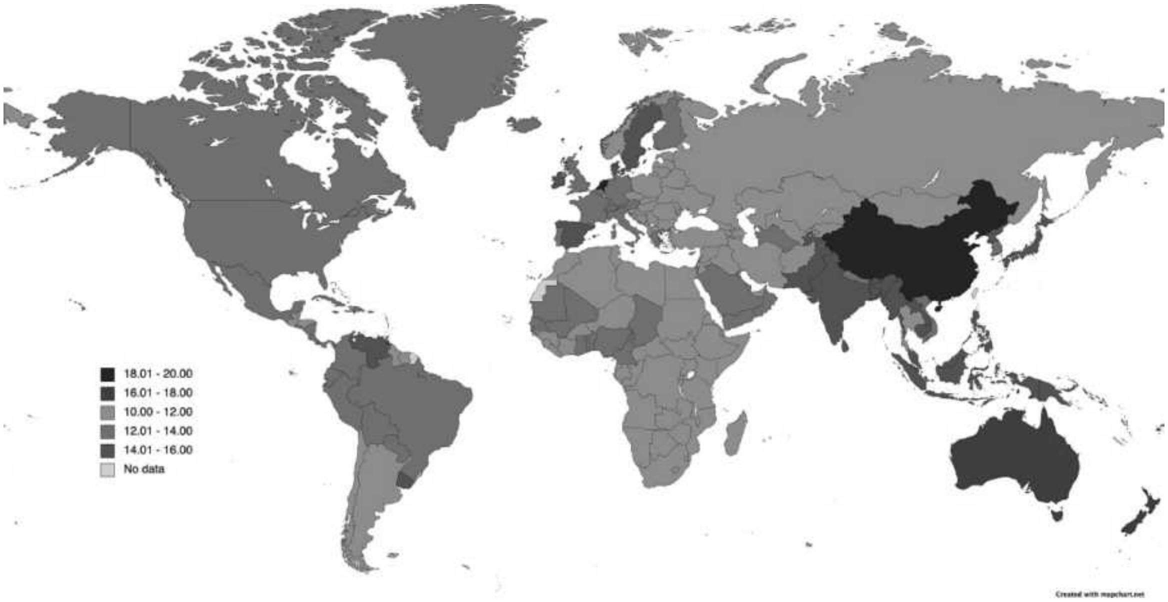
18.01 - 20.00
16.01 - 18.00
10.00 - 12.00
12.01 - 14.00
14.01 - 16.00
No data
Created with mapchart.net

FIGURE 12.3–2. Age-standardized schizophrenia DALY rates (ASDR) globally in 2017, estimated from the GBD 2017 study. (Data from He H, Liu Q, Li N, et al. Trends in the incidence and DALYs of schizophrenia at the global, regional and national levels: results from the Global Burden of Disease Study 2017. *Epidemiol Psychiatr Sci.* 2020;29(e91):1–11. doi:10.1017/S2045796019000891.)

The WHO Ten-Country Study

Inferring a more general perspective on the comparative incidence or prevalence of schizophrenia in different populations from individual studies conducted at different times and with different methods or from pooled data from multiple studies is beset with difficulties. The WHO Ten-Country Study, published in 1992, remains to date the only investigation that has applied a uniform design and common research tools to generate directly comparable incidence data for different populations. Incidence counts in the study were based on first-in-lifetime contacts with any aid agency within defined areas (including traditional healers in developing countries), which were monitored over a 2-year period. Potential cases and key informants were interviewed by clinicians using standardized instruments, and the timing of onset was reliably ascertained for the majority of the patients. In 86% of the 1,022 patients, the onset of diagnostic symptoms of schizophrenia was within the year preceding their first contact, and, therefore, the first-contact incidence rate was adopted as a reasonable approximation to the "true" onset rate. Two definitions of *caseness,* differing in the degree of sensitivity and specificity, were used: a "broad" clinical definition comprising ICD-9 schizophrenia and paranoid psychoses and a more restrictive definition of "nuclear" schizophrenia with Schneiderian *first-rank symptoms* of "nuclear" schizophrenia. Intersite reliability of case classification was ensured by applying a computerized diagnostic algorithm. The rates for eight of the catchment areas in which a nearly complete case finding was achieved are shown in Table 12.3–3. The differences between the area rates for "broadly" defined schizophrenia (0.16 to 0.42 per 1,000) were significant ($P < .001$), but those for "nuclear" schizophrenia were not, suggesting that the frequency of the latter subgroup varies less across different populations. No differences were found between the "broad" and "nuclear" cases with regard to age at onset or 2-year course and outcome. Therefore, it is unlikely that "nuclear" and "broad" schizophrenia identify different clinical disorders. Several replications of the WHO design, including its research procedures and instruments, have been carried out with very similar findings in India, the Caribbean, and the United Kingdom.

Table 12.3–3.
Incidence Rates per 1,000 Population, Ages 15 to 54 Years, for a “Broad” and a “Narrow” Case Definition of Schizophrenia (WHO Ten-Country Study, 1992)

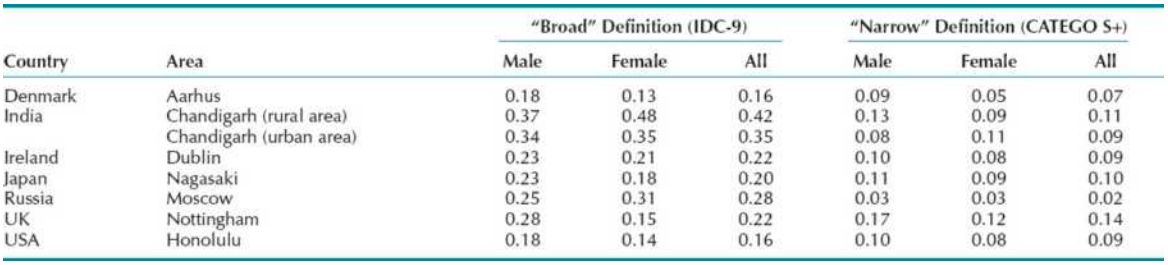

		“Broad” Definition (IDC-9)			“Narrow” Definition (CATEGO S+)		
Country	Area	Male	Female	All	Male	Female	All
Denmark	Aarhus	0.18	0.13	0.16	0.09	0.05	0.07
India	Chandigarh (rural area)	0.37	0.48	0.42	0.13	0.09	0.11
	Chandigarh (urban area)	0.34	0.35	0.35	0.08	0.11	0.09
Ireland	Dublin	0.23	0.21	0.22	0.10	0.08	0.09
Japan	Nagasaki	0.23	0.18	0.20	0.11	0.09	0.10
Russia	Moscow	0.25	0.31	0.28	0.03	0.03	0.02
UK	Nottingham	0.28	0.15	0.22	0.17	0.12	0.14
USA	Honolulu	0.18	0.14	0.16	0.10	0.08	0.09

Small Area Variations

Since schizophrenia is a low-incidence disorder, variations in its prevalence are more prominent in comparisons of rates obtained from small geographical areas and communities. In a study of an ethnically and socioeconomically homogeneous rural region in Ireland, the overall prevalence of schizophrenia at 3.9 per 1,000 was well within the modal range, but analysis by small district electoral divisions revealed significant variation in rates, ranging from 0.0 to 29.4 per 1,000. Such variation stands in contrast to the more similar rates typically found in studies of large urban areas or at a national level. Local variation can be attributed to several factors, for example, spatial clustering of cases due to shared genetic vulnerability within extended pedigrees, differential mobility and mortality, and, possibly, differential exposure to risk factors influencing intrauterine growth and early neurodevelopment. Such effects may give rise to outlier pockets of high or low incidence and prevalence that tend to cancel each other in larger population agglomerations.

Migrant Populations

Multiple observational studies have found that migration is a significant factor in increasing the risk of developing psychotic disorders. Schizophrenia risk in migrants is typically found to be 2 to 4 times that of the nonimmigrant population, but rates up to 10-fold have also been reported in some groups. A meta-analysis of studies of first- and second-generation immigrants (SGI) (largely in Europe, but also in Israel, Australia, and North America) found that a mean weighted incident rate ratios (IRR) for schizophrenia of 2.3 for first-generation immigrants (FGI) and 2.1 for SGI. The mean-weighted IRR for black FGI was higher (4.0) than for white FGI (1.8) or another ethnicity FGI (2.0). Black SGI also had IRR rates that were higher (5.4) than that of black FGI, a finding that was not seen among white or other immigrant groups. The persistence of schizophrenia risk in the second generation of immigrants suggests that postmigration factors play a significant role in increasing risk.

The highest IRRs of immigrants have been reported in studies in the United Kingdom, the Netherlands, and Scandinavian countries. An exceptionally high incidence rate of schizophrenia has been found in the African–Caribbean population in the United Kingdom, with a ninefold increase, compared to the native white British population in the 2006 AESOP study, compared to an IRR of 5.8 in black Africans, and 1.4 in South Asians. Similarly, a nearly fourfold excess over the general population rate has been reported for the Dutch Antillean, Moroccan, and Surinamese immigrants in the Netherlands. The causes of the phenomenon remain obscure. Incidence studies do not indicate any excess schizophrenia morbidity in the indigenous populations from which migrants are recruited. Explanations in terms of biologic risk factors have found little support. Misdiagnosis of psychotic disorders in immigrant groups has been debated but is unlikely to fully account for the large disparities in schizophrenia rates found across groups in multiple studies. With an absence of evidence from source countries of elevated schizophrenia rates, the social environment is believed to be the key factor underlying high rates in immigrants. Adverse social experiences involving being marginalized or perceptions of discrimination could be uniquely stressful, particularly when paired with limited social or family support. This is further suggested by findings from several studies showing that the incidence of schizophrenia in a group increases as they represent a smaller share of the host population.

Ethnic Minorities

Almost all empirical studies of psychotic disorder or schizophrenia have described disproportionate effects of race on the diagnosis. In a systematic review and meta-analysis of papers between 2002 and 2017, ethnic minorities had an increased risk of all psychotic disorders, which included nonaffective psychotic disorders. Race has also been related to diagnosis after using standardized diagnostic criteria and structured interviews.

Most empirical studies involving ethnicity or race and psychotic disorders have been conducted in the United States. These studies also include more indigenous ethnic minority populations and fewer immigrants than those conducted in other countries. In clinical settings, African Americans are disproportionately diagnosed with schizophrenia, typically 3 to 5 times compared to White Americans, and when admitted to state hospitals, are up to five times as likely to receive a schizophrenia diagnosis. They are also more likely to be diagnosed as having a psychotic disorder than Latino Americans. A meta-analysis of racial comparison studies in schizophrenia diagnosis found that African Americans had the disorder at a rate 2.4 times higher than White Americans, and odds ratios were slightly higher for studies using unstructured (2.4) than for those using structured (1.8) diagnostic assessments. African Americans and Latin American children and adolescents who present to psychiatric emergency rooms are also twice as likely to receive a diagnosis of a psychotic disorder than White Americans. When lifetime psychotic symptoms were assessed, its prevalence was 15% in African Americans, 13% in Latin Americans, 9% in Asian Americans, and 9% in White Americans. In addition, Blacks are associated with higher doses of antipsychotics, more frequent prescription of long-acting injections, more frequent use of first- compared to second-generation antipsychotics, lower quality of life, and worse outcomes after the first psychotic episode.

There is also a strong effect of ethnicity on diagnosis in other international studies. Black inpatients in Portugal are more likely to be diagnosed with schizophrenia and psychosis than white inpatients, and Black citizens in Sweden receive a psychotic disorder at an increased rate compared to those of other ethnic backgrounds. In the Netherlands, black and minority ethnic groups and whites from other western groups are over twice as likely to receive mandated hospital admissions than Dutch natives.

There are several potential reasons for disproportionately increased diagnosis of schizophrenia in African Americans, though no single empirically proven explanation appears to be sufficient. Unconscious clinical bias based on stereotypes can result in misdiagnosing African Americans at higher rates of schizophrenia, even among people of the same race as the patient. Clinicians may associate symptoms in African Americans with disruptive and socially deviant behaviors to a greater degree, influencing diagnostic choices. They may also use inappropriate cultural context or misinterpret normative guardedness as paranoid symptoms. In addition, a pattern of less access to healthcare, more distrust in mental health professionals, the higher social stigma of mental illness, and culture-specific methods of dealing with distress among many ethnic minority groups could lead to patients being more symptomatic when they present to treatment settings. Increased rates of schizophrenia in African Americans compared to lower rates in native Africans living in their continent is suggestive of the role of underlying environmental risk factors. Urbanicity is related to the increased incidence of schizophrenia and may be a mediating factor for race relationship to diagnosis, considering ethnic minorities tend to live in urban areas. Migrant studies conducted internationally have attributed perceived discrimination, social marginalization, and low social capital as factors driving factors toward increased psychotic disorder diagnosis. Such social factors are likely also relevant to ethnic minorities such as African Americans and may drive increased schizophrenia diagnosis rates. Differences in substance use rates across ethnic groups have also been hypothesized to contribute to diagnostic disparities. Data from the 2005 to 2013 National Surveys of Drug Use and Health involving 340,456 U.S. adults found that cannabis use disorder (CUD) rates were higher among native Americans (4.2%) and Blacks (2.4%) compared to Whites (1.3%). While specific substance use disparities may contribute to higher schizophrenia rates among African Americans, it is unlikely to be the primary factor. Black FGIs to Europe typically have low substance use rates, despite a relatively high prevalence of psychotic disorders.

Genetic Isolates

Isolated populations provide another source of incidence and prevalence variation. Isolates are characterized by common descent from a small number of *founder* ancestors, a degree of endogamy or inbreeding, and a restricted admixture of immigrants, due to geographical or cultural seclusion over multiple generations, often ranging over hundreds of years. Such populations may vary considerably in size but are likely to be less heterogeneous with regard to genetic makeup and environmental exposures than the outbred (panmictic) populations constituting the world’s majority, in which theoretically all individuals are potentially mating partners. The so-called young isolates comprise up to 20 to 30 generations and typically have arisen following drastic population size reductions (bottlenecks) due to wars, famine, epidemics, religious persecution, or other cataclysms. Subsequent population expansion results in a more uniform genetic background, including wider intervals of linkage disequilibrium, a more uniform environment and lifestyle, and a significantly higher or lower prevalence of certain diseases, including psychiatric disorders. If coupled with the availability of genealogic memory or records, such isolates present unique opportunities for genetic linkage and association studies of Mendelian (monogenic) diseases and, hopefully, of complex traits such as schizophrenia, bipolar disorder, and other psychiatric syndromes.

A number of isolated populations in different parts of the world, including Finland, Iceland, and northern Sweden; the Pima Indians; the Bedouins; the people of the Central Valley of Costa Rica, several areas in Quebec, as well as religious communities, such as the Old Order Amish, the Hutterites, and the Mennonites, have been studied by epidemiologists and geneticists with a view to identifying large pedigrees, informative for diseases ranging from asthma and diabetes to schizophrenia and bipolar disorder. High rates of psychoses (two to three times the national or regional rate) have been reported for population isolates in northern Sweden and several areas in Finland. Although the whole population of Finland shares some features of an old isolate (approximately 2,000 years), the northern and eastern regions of Finland have been settled relatively recently (in the 16th and 17th centuries), and one particular subregion with a current population of 18,000 has been founded by 40 families at the end of the 17th century, that is only 12 generations back. Genetic–epidemiologic studies in this isolate estimate the lifetime risk of schizophrenia at 2.2%, compared to 1.2% for the whole of Finland. A case register-based study of a birth cohort (14,817 individuals) from this region established a lifetime prevalence of 1.5% for schizophrenia spectrum psychotic disorders.

Daghestan in the Northern Caucasus (Russian Federation) is a region that has been inhabited for more than 3,000 years by some 26 small ethnic groups constituting together at least five genetically distinct populations, varying considerably in their morbidity patterns. Possibly, the highest-ever lifetime prevalence estimate for schizophrenia (4.9%) has been reported from one such highland subisolate (3,000 members).

The population of the Palau islands (Micronesia), currently 20,470 people, has been geographically and ethnically isolated from other Pacific populations for nearly 2,000 years. A genetic epidemiologic study of treated cases estimated the lifetime prevalence of schizophrenia at 2.77% in males and 1.99% in females, that is, high in excess of the modal risk of about 1% usually reported for large outbred populations. All of the 160 Palau cases were concentrated in 59 families, each traceable to a single common founder, with 11 of them having 5 to 11 affected members each.

At the other extreme, the lowest known prevalence rate of schizophrenia in any population (and a very low rate of bipolar disorder) has been found among the Hutterites in South Dakota, a Protestant sect of European descent (founded in the 16th century) whose members live since the 1870s in closely knit, endogamous rural communities in Manitoba (Canada) and South Dakota (United States). According to preserved pedigree records, all of the present 45,000 Hutterites are descendants of fewer than 90 ancestors who lived in the 18th and early 19th centuries. Reduced genetic heterogeneity and communal lifestyle with minimum variation in environmental exposures make this population an attractive laboratory for a variety of disease studies, including psychiatric disorders. An early epidemiologic study in which the entire population of several Hutterite communities was screened resulted in a schizophrenia lifetime prevalence of only 1.1 per 1,000. Subsequent reanalysis of the data using DSM-III-R criteria, and a repeat survey, replicated the original finding. Both genetic (low frequency of psychosis-predisposing alleles) and lifestyle factors (protective community support) have been proposed as an explanation for the unusually low rate of psychosis. Negative selection for individuals with schizoid traits who fail to adjust to the communal lifestyle and eventually outmigrate without leaving progeny has also been suggested but not definitively proven.

VARIATION IN THE COURSE AND OUTCOME OF SCHIZOPHRENIA

Systematic investigations into the course and outcome of schizophrenia were initiated by Kraepelin, who believed that the natural history of the disorder could provide a provisional validation of the disease concept until final verification by brain pathology and understanding of etiology. However, contrary to Kraepelin postulate of “mental weakness” as the ultimate outcome of *dementia praecox,* it is precisely the long-term course of schizophrenia that reveals the greatest extent of variation across populations and cultures. The course and outcome of schizophrenia exhibit striking heterogeneity and variability, both across and within populations. Patients with similar clinical and diagnostic characteristics at baseline assessment develop a broad spectrum of outcomes ranging from stable clinical and social recovery after a single psychotic episode to chronic unremitting psychosis and severe impairment. Long-term follow-up studies lend credibility to the conclusion that a high proportion (more than 30%) of patients meeting the diagnostic criteria for schizophrenia may achieve relatively favorable outcomes.

Untreated Schizophrenia

Since the great majority of schizophrenic patients in developed countries are today receiving pharmacologic treatment, one could assume that current research does not reflect the "natural" course and outcome of untreated schizophrenia. However, several studies in an urban area in Scotland and in remote rural areas in India and China have succeeded in detecting and assessing *never hospitalized* patients with schizophrenia. About half of the Scottish patients had been prescribed antipsychotic medications by their general practitioners, while the Indian and Chinese patients were virtually untreated. In all three settings, the outcomes of these samples were heterogeneous, but, except for a larger proportion of Chinese patients having marked psychotic symptoms, they did not differ much from the outcomes in the treated groups. In a historical study of 70 Swedish patients with first admissions in 1925, lifetime records were retrieved and rediagnosed in accordance with DSM-III. None of these patients had received antipsychotics. The final outcome was rated as "good" in 33%, as "profoundly deteriorated" in 43%, and as "intermediate" in 24%.

Secular Trends in the Outcome of Schizophrenia

A long-term perspective on the course of schizophrenia over successive generations is provided by a meta-analysis of 320 outcome studies on schizophrenia or *dementia praecox* published between 1895 and 1992, which comprise a total of 51,800 subjects. Overall, about 40% of the patients have been described as "improved" after an average length of follow-up of 5.6 years. There was a significant increase in the rate of improvement from 1956 to 1985 compared to 1895 to 1955, clearly related to the introduction of antipsychotic medications, but a secular trend toward better outcomes with every successive decade had been present for much longer. Coupled with the virtual disappearance of the most malignant or “catastrophic” forms of schizophrenia that tended to result in a profound “defect state” after a single psychotic episode or in death (*lethal catatonia*), these observations suggest that a transition to a less deteriorating course of schizophrenia had been taking place prior to the advent of modern pharmacologic treatment. Among the factors that might explain this shift are improvements in general care, progressive changes in attitudes and hospital regimes that occurred in a number of institutions on both sides of the Atlantic in the 1930s and 1940s, and the expectation that psychosocial measures such as psychotherapy or rehabilitation could result in a cure or remission of the disorder.

Concurrently, with the above trends in Europe and North America, reports from developing countries, usually based on small clinical samples, pointed to a less disabling course and a high rate of recovery from schizophrenic psychoses in traditional societies such as Mauritius and Sri Lanka. Good outcome cases included patients who would be expected to have a poor outcome if “Western” prognostic criteria had been applied. However, selection bias could not be excluded, as these studies were based on hospital admissions only. In addition, clinical improvement could have been confounded with the social adjustment many patients achieve in a comparatively undemanding environment. Thus, the room was left for doubts concerning the validity of the early observations of a better prognosis of schizophrenia in non-Western settings.

Table 12.3–4.

Two-Year Course and Outcome Features of 1,070 Patients With Schizophrenia in the WHO Ten-Country Study

Course and Outcome Descriptor	**Percent Patients in Developing Countries**[a] **(*n* = 467)**	**Percent Patients in Developed Countries**[b] **(*n* = 603)**
Remitting, complete remissions	62.7	36.8
Continuous or episodic, no complete remission	35.7	60.9
Psychotic <5% of the follow-up	18.4	18.7
Psychotic >75% of the follow-up	15.1	20.2
No complete remission during follow-up	24.1	57.2
Complete remission for >75% of the follow-up	38.3	22.3
On antipsychotic medication >75% of the follow-up	15.9	60.8
No antipsychotic medication during follow-up	5.9	2.5
Hospitalized for >75% of follow-up	0.3	2.3
Never hospitalized during follow-up	55.5	8.1
Impaired social functioning throughout follow-up	15.7	41.6
Unimpaired social functioning >75% of	42.9	31.6

follow-up

---

[a] Study areas in Colombia, India, and Nigeria.

[b] Study areas in Czech Republic, Denmark, Ireland, Japan, Russia, the United Kingdom, and the United States.

Many of the methodologic issues attending such research were later addressed by the WHO multicenter studies that employed standardized assessment, recruitment of homogeneous, incidence-based samples, and more refined, dimensional measures of course and outcome. These studies included a total of 2,736 patients in 16 countries diagnosed with schizophrenia according to strict criteria.

Course and Outcome in the WHO Studies

The earliest WHO investigation, the IPSS, found that higher proportions of patients in India, Colombia, and Nigeria had better outcomes on most dimensions than patients in the developed countries at 2- and 5-year follow-up assessments. However, the IPSS was not free of bias since patients were recruited from hospitals, where admission policies could have led to the overinclusion of chronic cases in the developed countries and of recent onset, acute cases in the developing countries.

Such confounding factors were largely eliminated in the subsequent WHO Ten-Country Study, in which epidemiologic cohorts of first-episode cases were uniformly assessed upon their first contact with community or hospital services. Diagnostic misclassification due to the potential admixture of transient psychotic episodes of good prognosis or of psychoses due to acute physical illness was excluded, as such cases were carefully screened out. The follow-up of these incidence cohorts provided ample confirmation of the finding that the outcome of schizophrenia was generally better in developing than in developed countries (Table 12.3–4). The better course and outcome in the developing country areas could not be attributed to any particular clinical subtype of the disorder. The main outcome difference across the study areas was in the average length of symptom-free remissions. Analysis of the data revealed that the better overall pattern of course and the less disabling outcome in the study areas in developing countries was mainly attributable to a significantly greater percentage of patients remaining in a remission of symptoms over longer periods after an acute psychotic episode, rather than to milder or shorter psychotic episodes. Statistically, this pattern was significantly predicted by the setting (developing versus developed country), acute onset, being married or cohabiting with a partner, and having access to a supportive network (family or friends). The length of remission was unrelated to antipsychotic treatment, which generally was administered for much shorter periods of time to patients in developing countries.

Table 12.3–5.
Long-Term (15- and 25-Year) Outcome in Patient Cohorts Assessed in the International Study of Schizophrenia (ISoS, 2007)

Outcome Variable	Incidence Cohorts (*n* = 1,171, Including 15-year Follow-up of the WHO 10-Country Cohort)	Prevalence Cohorts (*n* = 462, Including 25-year Follow-up of 373 Cases from the WHO IPSS)
Recovered at follow-up (Bleuler criteria)	48.1	53.5
Not psychotic in the past 2 years	42.8	40.8
GAF-S >60	54.0	56.7
Working most of past 2 years	56.8	73.9
GAF-D >60	50.7	60.3
SMR (range)	0.00–5.71[a]	1.04–8.88[b]

GAF-S, Global Assessment of Functioning–Symptoms Scale; GAF-D, Global Assessment of Functioning–Disability Scale; SMR, standard mortality ratio.

[a] Rochester (0.000), Moscow (1.41), Chandigarh urban (1.88), Prague (2.53), Chandigarh rural (3.02), Honolulu (3.13), Nottingham (3.31), Dublin (4.10), and Nagasaki (5.71).

[b] Sofia (1.04), Cali (1.31), Madras (1.90), Agra (1.86), Beijing (2.97), Prague IPSS (3.84), Mannheim (5.55), Hong Kong (5.76), and Groningen (8.88).

A subsequent long-term follow-up investigation, the International Study of Schizophrenia (ISoS), involving 18 research centers in 14 countries, traced 75% of cases assessed in the earlier WHO studies referred to previously, as well as additional cohorts from China, Hong Kong, and India. Follow-up data were collected on a total of 1,633 cases (surviving or dead), and 890 patients were reinterviewed at either 15- or 25-year follow-up since their first assessment. Key findings from this study are presented in Table 12.3–5.

Several general conclusions can be drawn from the WHO studies. The frequencies of both relapses and remissions tend to increase over time. However, regardless of the increased relapse rate, the cumulative proportion of follow-up time during which patients have psychotic symptoms (as a percentage of the total follow-up time) tends to remain stable or even decrease. At the end of the 5-year follow-up, 57% of the patients had experienced a total of fewer than 9 months of active psychosis; only 22% had been psychotic for 45 to 60 months. At 15-year and 25-year follow-up, 43% and 41%, respectively, have been free of active psychotic symptoms for the past 2 years. The levels of social impairment assessed at 2 years changed little during the subsequent years of follow-up. Overall, most of the observed changes in the clinical state and social functioning of patients between the 2-year follow-up and the 5-year follow-up were toward improvement rather than deterioration. This was congruent with the findings at 15-year and 25-year follow-up. Independently of the WHO studies, similar findings of better outcomes of schizophrenia in developing countries have been reported by other investigators.

Factors Influencing the Variations in Outcome

The differences in outcome between patients in developing countries and patients in developed countries remain insufficiently understood but are unlikely to be reducible to a single variable or to a small number of explanatory variables. The variation in course and outcome may result from additive or interactive effects of genetic differences between populations; from lower incidence in traditional societies of the type of chronic stress to which people with schizophrenia are particularly vulnerable; or from higher odds of an individual–environment fit in traditional communities that minimize social isolation. While genetic differences undoubtedly exist (e.g., in the allelic structure of genes coding for key neurotransmitter receptors), nothing more specific can at present be said as to their role in the course and outcome of schizophrenia. Worse schizophrenia outcomes in migrants and ethnic minorities compared to that observed in their source countries suggests that genetic differences may be less determinative than environmental influences. A strong modifying influence of the psychosocial environment on the course of schizophrenia is plausible, considering the differences between developing and developed countries with regard to social support systems, kinship networks, and beliefs about mental disease. As regards psychosocial stress, the WHO studies found that the index of *expressed emotion* (EE), a short-range predictor of psychotic relapse, was as effective in Indian families as in European and North American families, but that high-EE families were much rarer in India than in Denmark or the United Kingdom. If this finding could be replicated in other settings in developing countries, the relative rarity of at least one type of pathogenic stress in the daily environment of schizophrenic individuals in such communities could be demonstrated. However, EE may not be the only type of stress to which schizophrenic individuals respond with psychotic exacerbation. Ethnopsychiatrist Henry Murphy (1915 to 1987) proposed four criteria for schizophrenia-evoking stress: (i) a situation demanding action or decision; (ii) complexity or ambiguity of the information supplied to deal with the task; (iii) unless resolved, the situation demanding action or decision persists; (iv) the person has no “escape route” available. Each component of this model may occur at different frequencies in traditional and industrialized societies, a proposition that should be testable epidemiologically or experimentally. For example, important differences exist between traditional cultures and Western societies regarding the perception of the sick role and the beliefs and practices related to mental illness. Shared magical–mystic explanations of mental illness and traditional healing practices, still prevalent in traditional rural communities, certainly provide no cure for schizophrenia, but they may play a role in lowering the barriers to recovery and reintegration into the community. Substance use likely also plays a role in influencing outcomes, particularly cannabis use, which has been associated with triggering psychosis. According to the GBD study, the United States has the highest age-standardized prevalence of cannabis, opioid, and cocaine use disorders. CUDs are next highest in Australia and Western Europe, while Eastern Europe leads the world in alcohol use disorders. In contrast, the continent of Africa, the Middle East, Central America, and South Asia have relatively low rates of substance use disorders. The findings of a better outcome for schizophrenia in traditional societies are compelling enough to set a research agenda that may lead to important insights concerning the management and treatment of schizophrenia in both developing and developed countries.

To sum up, the most striking finding from long-term follow-up studies is the high proportion of patients who recover completely or with mild residual abnormalities after decades of severe illness. This contrasts with the ingrained image of schizophrenia as an intractable, deteriorating illness that many clinicians tend to adopt on the basis of a limited follow-up horizon and patient samples selected for their unfavorable course and treatment response. It is unlikely that the high percentage of recoveries in the long-term studies could be explained by affective illness or brief transient psychoses being misdiagnosed as schizophrenia. Similarly, unlikely is the attribution of all the good outcomes to antipsychotic treatment since comparable proportions of improvement of recovery had been reported for patients who never received antipsychotic medications. A tentative conclusion is that schizophrenia is not an invariably chronic deteriorating disorder. It should be noted, however, that some studies suggest a recent trend of worsening clinical and social outcomes in patients with schizophrenia in both developed countries and developing countries. People with schizophrenia are profoundly sensitive to their immediate social environment. Increasing social and economic stresses experienced by both rural and urban communities in many developing countries in the last two decades may be eroding the traditional support systems, thus resulting in worse outcomes.

THE BURDEN OF COMORBIDITY

Physical Disease

There is significant medical comorbidity in schizophrenia, comprising mainly common diseases that affect patients with schizophrenia more frequently than attributable to chance, as well as certain rare conditions or abnormalities that tend to co-occur with the disorder. Physical disease in patients with schizophrenia tends to be seriously undetected and underdiagnosed. Between 46% and 80% of inpatients with schizophrenia and between 20% and 43% of outpatients have been found in different surveys to have concurrent medical illnesses. Persons with schizophrenia, and especially those who are homeless or injection drug users, are at increased risk for potentially life-threatening communicable diseases, such as HIV/AIDS, hepatitis C, and tuberculosis.

However, it is the chronic noncommunicable diseases, that in the past decades, are taking an increasingly heavy toll on people with schizophrenia. Cardiovascular disease (CVD) and the associated premature mortality are significantly increased in schizophrenia. A recent comprehensive meta-analysis of mortality in schizophrenia reported a standardized mortality ratio (SMR) of 2.54. It has been estimated that the high prevalence of CVD impacts the life expectancy of people with schizophrenia by an average reduction of 18.7 years for men and 16.3 for women and that the gap between those affected with the disorder and the general population is increasing. Although suicide rates in schizophrenia are high and equal to those for major depression, recent studies suggest that CVD is at present the main cause of premature death, already evident in early or middle adulthood. The contribution of “atypical” antipsychotic medication to weight gain and hence to metabolic imbalance and type II diabetes is likely, but its effects are collinear with other modifiable lifestyle risk factors, including nutrition, smoking, and lack of physical activity. Obesity and the concomitant metabolic syndrome involving insulin resistance and type II diabetes are at present common problems in patients with schizophrenia.

Certain rare genetic or idiopathic disorders, such as metachromatic leucodystrophy, acute intermittent porphyria and coeliac disease, and dysmorphic features such as high-steepled palate, malformed ears, and other minor physical anomalies, also tend to co-occur with schizophrenia. The genetic syndrome that involves a chromosome 22q11.2 deletion (DiGeorge Syndrome) involves schizophrenia in about 25% of individuals. There are typical associated physical features with 22q11.2 deletion syndrome, including mild facial dysmorphia and congenital anomalies, as well as congenital heart defects, velopharyngeal insufficiency, and submucosal cleft palate in some cases.

Recently, studies have also found that schizophrenia patients infected with the novel coronavirus, COVID-19, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are twice as likely to be hospitalized for this infection (OR = 2.15) and remained significant even after controlling for sociodemographic and smoking, obesity, diabetes, hyperlipidemia, chronic obstructive pulmonary disease, and ischemic heart disease (OR = 1.88). In addition, compared to healthy controls, schizophrenia patients were about three times as likely to have infection mortality (OR = 3.27), and this remained significant after adjusting for sociodemographic (OR = 2.79) and clinical factors (OR = 2.77).

Substance Abuse

Substance abuse is, at present, by far the most common associated health problem among patients with schizophrenia and may involve any drug of abuse or a polydrug combination. The addictive use of cannabis, stimulants, and nicotine is disproportionately high among patients with schizophrenia and may be related to the underlying neurobiology of the disorder. Reasons for the comorbidity of substance use disorders and schizophrenia are unclear, and several theories have been proposed to explain this, including common genetic vulnerabilities and neural networks, cumulative risk and environmental stressors, or attempts at self-medication.

Alcohol and illicit drug use in schizophrenia patients are relevant clinically, as they worsen the overall course of the illness and increase morbidity and mortality. Affected individuals are more likely to get hospitalized or become homeless, violent, suicidal, or incarcerated. It is also linked with decreased treatment compliance and risk of relapse. In a systemic review and meta-analysis of 123 studies on substance use disorders in schizophrenia in community and clinical settings between 1990 and 2017, the overall prevalence was 0.417. The overall prevalence of alcohol use disorder was 0.243, with the prevalence in Asia being lower (0.077) than that in other countries. Illicit drug use disorder use in schizophrenia patients was 0.275 and higher in those with first-episode psychosis (0.362) compared to those with schizophrenia (0.197). The highest prevalence of illicit drug use disorder was found among patients in North America and Australia. The point estimate for prevalence for stimulant use disorder in schizophrenia patients was 0.073, while that for opioid use disorder was 0.051. Male patients also have a two to three times greater risk of substance use disorders than females.

About 3.9% of the global adult population uses cannabis, which is higher than all the other illicit substances combined. Cannabis use has particular relevance to schizophrenia as it has been directly linked to an increased risk of developing psychosis. Cannabis has been extensively reported to induce psychotic experiences (PEs), and active cannabis users report higher ratings of positive symptoms. Previous meta-analyses indicate that the risk of developing psychotic disorders is increased twofold in those with schizophrenia. A systematic meta-analysis of 10 selected studies involving 66,816 individuals showed that among the heaviest cannabis users, the risk of schizophrenia and other psychotic disorders was even higher (OR = 3.90). In addition to the degree of cannabis exposure, existing evidence indicates that the use of high-potency cannabis and early onset of use are stronger risk factors for developing psychotic disorders. A systematic review of published data on cannabis exposure and the onset of schizophrenia concluded that early use increased the risk of psychosis in a dose-related manner, especially in persons at high genetic risk of schizophrenia. The point estimate for prevalence for studies of subjects with current or lifetime CUD was 0.262. First-episode patients had higher prevalence rates (0.356) of CUD compared to chronic schizophrenia patients (0.208). The highest CUD prevalence in schizophrenia patients was found in the United States, Australia, and Western Europe. The lowest rates of CUD globally are seen in sub-Saharan Africa and North Africa/Middle East, and South Asia.

VARIATIONS IN MORTALITY

In both developing and developed countries, schizophrenia is associated with excess mortality, well documented by epidemiologic studies on large cohorts over extended periods. Those with schizophrenia have been known to have higher rates of somatic disorders, including cardiovascular and diabetes, compared to control populations, as well as increased tobacco, alcohol, and illicit substances use. In addition, increased sedentary behavior, poor dietary habits, and poor health-seeking behavior negatively influence health outcomes. Finally, schizophrenia patients have an increased suicide risk by as much as 22 times compared to the general population, particularly early after symptom onset. A meta-analysis of studies involving 247,603 patients across Africa (n = 1; Ethiopia), Asia (n = 1; Taiwan), Australia (n = 1), Europe (n = 7), and North America (n = 3) found that, on average, those with schizophrenia had 14.5 years of potential life lost. Men lost about two years of potential more than women with schizophrenia, and life expectancy was about 8 years lower for men than women patients, and the difference was primarily due to a male excess in suicides and accidents. Globally, the loss was the least in Asia and the greatest in Africa. Overall weighted life expectancy was 65.7 years in schizophrenia patients and was lowest in Asia and Africa.

Population-based data from Norway, 1926 to 1941 and 1950 to 1974 indicate that while the total mortality of psychiatric patients had been decreasing, the relative mortality of patients with schizophrenia remained unchanged at a level higher than twice that of the general population. Similar findings have been reported from other European countries and North America, with an SMR of 2.6 or higher for patients with schizophrenia, which corresponds to about a 20% reduction in life expectancy. In Taiwan, data collected over a 15-year period indicate that of all mental disorders followed up, schizophrenia was associated with the highest mortality, representing an 80% increase over the mortality of the general population. Data from successive patient cohorts in Denmark, the United Kingdom, and Australia suggest a trend of increasing mortality in the early stages of the disorder. In Denmark, the 5-year cumulated SMR increased from 5.30 (males) and 2.27 (females) for the period 1971 to 1973 to 7.79 (males) and 4.52 (females) for the period 1980 to 1982 (56). Particularly striking was the SMR of 16.4 for males in the first year after a diagnosis of schizophrenia.

Whereas in the past, the excess mortality among individuals with schizophrenia was due mainly to communicable diseases such as tuberculosis, the leading causes of premature death among patients with schizophrenia at present are CVD, suicide, accidents, and common physical diseases. Suicide accounts for 28% of the excess mortality in schizophrenia, with aggregated SMR of 9.6 for males and 6.8 for females. In developed countries, the suicide risk associated with schizophrenia is now nearly as high as that associated with major depression. In China, the *relative risk* (RR) of suicide in individuals with schizophrenia, compared to those without the disorder, has been estimated at 23.8. Suicide in schizophrenia may occur at any stage of the disorder, but the risk is particularly high in the first 6 months after the first psychotic episode, as well as following periods of frequent hospital admission and discharge. Several risk factors relatively specific to schizophrenia have been suggested: being young and male, experiencing multiple relapses and remissions, comorbid substance abuse, awareness of the deteriorating course of the condition, and a sense of hopelessness with loss of faith in treatment. In the Australian study, suicide risk was highest in the first 7 days after discharge from inpatient care. The trend seems to parallel the significant reduction in the number of psychiatric beds. Whether the increased suicide mortality is associated with the shift in the management of schizophrenia from hospital to community care remains to be established.

Unnatural causes of death apart, in both developed and developing countries, schizophrenia is associated with excess mortality from common diseases, caused or exacerbated by poor self-care, inadequate nutrition, heavy smoking, and medical neglect. At least part of this excess mortality is preventable.

CRIMINAL OFFENDING AND VICTIMIZATION

Much remains to be understood about the risk of offending by individuals with schizophrenia and related psychotic disorders. Inconsistent research findings have been attributed to the use of varied and sometimes flawed methodologies and biased samples, such as inpatient and prison populations. The risk of violent offending has been estimated to be much higher in people with severe mental illness than that in the general population, partly attributable to the presence of substance abuse. In a large-scale Western Australian study of a cohort of all people born (1955 to 1969), arrest history was documented for the period 1985 to 1996 and cross-linked with the psychiatric case register. The prevalence of mental illness among offenders was 11.1%; 6.5% of offenders had substance use disorders, but only 1.7% had schizophrenia. Notably, for the majority of offenders with schizophrenia, the first arrest preceded first contact with mental health services. Given a large proportion of schizophrenia offenders being arrested prior to their first contact with psychiatric services, there are important implications for mental health and criminal justice policy and practice. Very similar findings were reported from a meta-analysis of nine studies, indicating that a substantial proportion of first-episode psychosis individuals commit offenses before presenting for treatment. Another aspect of the problem is victimization—violent assaults on people with severe mental illness. Among adults with psychotic disorders, especially women, the annual prevalence of any victimization was 38.6%, and of violent victimization 16.4%—nearly five times higher than the general population rate.

CLINICAL HIGH-RISK FOR PSYCHOSIS AND SUBTHRESHOLD PSYCHOTIC EXPERIENCES

Over the past three decades, the study of prepsychotic states has become a rapidly expanding area of research, often coupled with intervention programs and dedicated services. Clinical constructs have evolved to capture potential prodromal syndromes, with terms such as the “clinical high-risk (CHR)” or “ultra-high-risk (UHR)” states. Recognizing these diagnostic syndromes can facilitate early intervention, which can improve long-term outcomes and potentially the progression of symptoms to full-blown psychotic disorders. The role of these syndromes as a diagnosis, however, is being debated, and a related condition has been included as a condition for future study in the DSM-5.

The CHR criteria for the psychosis-risk state are based on either attenuated psychotic symptoms (APS), brief intermittent psychotic episodes (BIPS), or having genetic risk with a marked deterioration in psychosocial functioning (GRD). CHR individuals also tend to have small-to-medium neurocognitive impairments, intermediate between healthy controls and those with schizophrenia. The most commonly used interview measures used to assess CHR features are the Comprehensive Assessment of At-Risk Mental State (CAARMS), widely used in Australia, Asia, and Europe, and the Structured Assessment of Psychosis-Risk Syndromes (SIPS), most commonly used in North America. Basic symptom (BS) criteria have been proposed as symptoms that may identify an earlier prodromal state than the CHR criteria and involve subjectively experienced disturbances in mental processes, including thinking, speech, attention, perception, drive and stress tolerance. While psychosis-risk states are less severe than psychotic disorders, they often are associated with distress and poor functioning. In one study, 89% of adolescents meeting APS were reported to be distressed by the experienced symptoms.

Putative prodromal symptoms are pleiotropic and often do not convert to psychotic disorders like schizophrenia, although the risk of developing psychosis in those in the CHR state is substantially greater than in the general population. Most individuals meeting CHR criteria do not convert to a psychotic disorder over time. In a recent meta-analysis of 9,222 meeting CHR criteria, the cumulative conversion risk for psychosis was 0.09 at 6 months, 0.15 at 12 months, 0.20 at 18 months, 0.19 at 24 months, 0.25 at 30 months, 0.25 at 36 months, 0.27 at 48 months, and 0.28 at more than 48 months. The nonconverting group typically shows improvement in symptoms and functioning, though on average, they remain at a lower level of functioning than the population controls. Compared to those in the general population who meet CHR criteria, help-seeking CHR individuals have a higher conversion risk and are also more likely to exhibit distress and impairment.

Population-based studies have found that a history of subthreshold PEs is not uncommon in the general population, and these can potentially impact everyday functioning. The estimated prevalence of these PEs has varied widely across communities and is influenced by the method of data collection or screening questionnaire used. Screening instruments that probe a larger number of PEs can result in higher prevalence estimates than instruments with fewer questions. In addition, dichotomous (i.e., yes/no) psychosis questionnaires often underestimate prevalence rates compared to dimensional questionnaires, which probe symptom frequency. Past PE estimates in community samples have ranged from 4% to over 70%. A replication study of the National Comorbidity Study found that 9.1% of the U.S. population reported at least one PE. The NEMESIS study found a 17.5% rate of at least one PE in a general population in the Netherlands. PE rates are generally found to be higher among women than in men, in stark contrast to sex differences typically found in psychotic disorders.

The WHO’s World Health Survey, composed of 256,445 subjects across 52 countries, found that 12.5% of the overall population responded positively to having had at least one PE over the last year. However, weighted and sex-age standardized PE prevalence estimates varied highly across the countries surveyed, from 0.66 in Vietnam and 1.5 in Bosnia and Herzegovina to 45.8 in Nepal and 32.0 in Brazil. Average standardized prevalence estimates for PEs were 8.83 for high and medium-high-income countries and 12.90 for low- and low-medium-income countries. A subsequent cross-national survey examined data from 31,261 subjects across 18 countries as part of the WHO World Mental Health Survey and reported mean lifetime PE rates of 5.8%. Lifetime PE rates in this study were significantly higher among those from high-income (6.8%) and upper–middle-income (7.2%) countries than among those from low- and low–middle-income countries (3.2%). Overall health status inversely correlates with the number of PEs. PE severity correlates with increased anxiety, depression, and perceived stress, and may reflect a vulnerability to psychotic, anxiety, mood, and substance use disorders.

SOCIAL AND ECONOMIC COSTS

The health, social and economic costs of schizophrenia are high. In Western countries, the estimated direct costs of schizophrenia, according to the WHO, are between 1.6% and 2.6% of the total health care expenditures. This accounts for 7% to 12% of the gross national product. A systematic review of 56 economic burden studies in schizophrenia up till 2014 found that annual costs ranged from $95 million in Puerto Rico to $102 billion in the United States. The economic burden of schizophrenia as a function of the country’s gross domestic product varied between 0.02% in the United Kingdom to 1.65% in Sweden. Most of the cost of schizophrenia is due to productivity losses, indirect costs contributing to 45% to 85% of the overall costs associated with schizophrenia across most national studies. A study in the Netherlands, however, reported indirect costs of only 8%, while two studies in Nigeria reported indirect costs of 12% and 27% respectively. The accuracy of indirect costs, however, depends on the cost estimation method used. Indirect costs refer to productivity losses due to morbidity and premature mortality and include absenteeism or sick leave, decreased work productivity, unemployment, permanent disability, or early retirement for individuals with schizophrenia or caregivers. Direct medical costs refer to costs of hospital and physician visits, long-term institutional care, and other health professionals’ care, as well as medication, diagnostic tests, or medical supplies. Direct nonmedical costs include costs of transportation, food and lodging associated with health visits, and social services.

Although the generic cost-driving factors associated with schizophrenia are likely to be similar around the world (management of the chronic or relapsing symptoms and impairments, provision of inpatient and outpatient care and medication, mortality, lost productivity and unemployability, impact on the family, and impact on the community), their relative weights and, hence, the structure of the direct and indirect costs of the illness is likely to vary considerably. Hospital or other residential care, which accounts for more than 75% of the direct costs of schizophrenia in high-income countries, is likely to account for a smaller fraction in developing countries because of lower staffing and equipment costs. On the contrary, the proportion of direct costs attributable to the dispensing of antipsychotic medication, which is less than 5% of the total direct costs in developed countries, is likely to be higher in developing countries.

It is difficult to estimate the economic impact of schizophrenia on families in developing countries. A study conducted in Nigeria revealed that caregiver *opportunity costs* (e.g., family members' lost productivity or income) were not of a different order from those in the developed world. Yet, the aggregate family costs in developing countries may, in fact, be substantially higher since a larger proportion of schizophrenic patients live with their families, as compared with patients in Western societies, and the family is likely to be the first line of treatment and management of psychotic episodes. The cost of purchasing prescribed maintenance medication is usually borne by the family. Additional direct costs may be incurred by the necessity to travel, often long distances, to the nearest hospital or clinic, as well as by payment for the services of traditional healers. In many traditional communities, the stigma associated with mental illness may affect the family as a whole and restrict, for instance, marital opportunities for younger family members. While lost educational opportunities are likely to be a problem, lost paid employment is more difficult to quantify, and therefore, less likely to appear prominently in estimates of the indirect illness burden in developing countries. On the contrary, reintegration of a family member who has suffered a psychotic episode into the domestic economy may be much easier to achieve than formal employment, and this may be one factor in the better and longer remissions of patients with schizophrenia observed in developing countries.

Notwithstanding the variations in the course and outcome discussed previously, schizophrenia is associated with a greater burden of chronic disability than any other mental disorder. Both the positive and negative symptoms of the disease interfere seriously with a person's capacity to cope with the demands of daily living. Patients with schizophrenia experience particular difficulty in dealing with complex demands and environments, especially those that involve social interaction and decoding of social communication. Moreover, the onset is usually at a developmental stage of incomplete social maturation, educational attainment, and acquisition of occupational skills. The intrusion of psychosis at this stage results in a severely truncated repertoire of social skills and lifelong socioeconomic disadvantage. These factors are exacerbated by the societal reaction to individuals manifesting behaviors associated with “insanity,” which generally involves stigma and social exclusion.

These adverse factors interact to cause the *social breakdown syndrome,* a term introduced by Ernest Gruenberg (1915 to 1991) to designate the cluster of secondary impairments in schizophrenia resulting in a loss of social support networks and a greatly diminished quality of life for a substantial proportion of patients. Many patients with schizophrenia end up on the streets or in the criminal justice system and are exposed to abuse, even in psychiatric hospitals. Such outcomes are not uncommon in either developed or developing countries, although, in the latter settings, traditional family and community structures are still capable of providing a protective environment, and probably fewer patients are marginalized by society.

An important aspect of the economics of schizophrenia is the so-called *funding imbalance effect:* studies indicate that 97% of the total lifetime costs of schizophrenia are incurred by fewer than 50% of the patients diagnosed with the disorder. While this finding points to a hardcore subset of cases with severe chronic illness, multiple disabilities, and excessive dependence on services and other support, it also suggests that in more than 50% of those with schizophrenia, the disorder is less disabling or treatment more effective.

Another aspect of the social and economic costs of schizophrenia is the commonly perceived association between the disease and criminal behavior. Offenses committed by persons with schizophrenia tend to receive wide media coverage and reinforce popular ideas about the dangerousness associated with mental illness. It is important to dispel such prejudicial attitudes. Carefully designed studies in Europe and elsewhere indicate that a small proportion of patients with schizophrenia tend to be overrepresented among the perpetrators of violent offenses, including homicide. However, the *population-attributable fraction* of such offenses committed by persons with schizophrenia is negligible compared with the total number of offenses in the community. Moreover, the rate of arrest and incarceration is likely to be high among patients with schizophrenia because of their conspicuous behavior and appearance rather than the seriousness of the offense. Issues of protecting patients’ rights must, therefore, be an important part of antistigma campaigns in both developing and developed countries.

FUTURE DIRECTIONS

Important insights into the nature and causes of schizophrenia have been gained from population-based studies, although essential questions still remain unanswered. The clinical syndrome of schizophrenia appears to be robust and identifiable reliably in diverse populations and cultures, suggesting that a common pathophysiology, and possibly, a common genetic predisposition may underlie its manifestations. At the level of large population aggregates, no major differences in incidence and morbid risk have to date been detected, though small geographical area variations exist and appear to be related to a mix of locally operating risk factors whose effects may be attenuated in large, heterogeneous populations.

The study of “atypical” populations, such as genetic isolates or minority groups, may be capable of detecting unusual variations in the incidence of schizophrenia that could provide novel clues to the etiology and pathogenesis of the disorder. Notwithstanding the difficulties in the genetic dissection of complex disorders, emerging powerful methods of genomic analysis will eventually identify polymorphisms and haplotypes associated with schizophrenia risk. The majority are likely to be in genes of small effect, although one cannot rule out the possibility that rare mutations of moderate or major effect will also be found, especially in isolated populations or at the level of neurocognitive and neurophysiologic abnormalities associated with the disorder. Establishing their population frequency and associations with a variety of phenotypes, including premorbid personality traits, will be a major task for comparative epidemiology.

Current epidemiologic research is increasingly making use of existing large databases, such as cumulative case registers or birth cohorts, to test hypotheses about risk factors in case-control designs. Methods of genetic epidemiology are being integrated into population-based studies. These trends predict an important future role for epidemiology in the unraveling of gene-environment interactions that are likely to be the key to the understanding of the etiology of schizophrenia. In this context, research into psychotic disorders in non-Western populations can provide valuable information on genetic heterogeneity, the impact of the environment, and the course and outcome of schizophrenia. Both traditional communities and societies undergoing a transition in their social and economic organization can contribute to a better understanding of the relationships between culture and mental disorder and the variety of human experiences in coping with mental illness.

The vast majority of people whose lives are profoundly affected by schizophrenia reside in developing countries. As indicated previously, the incidence and prevalence of schizophrenia in developing countries are not different from what is being found in the developed world or may even be lower. Although a greater proportion of people with schizophrenia in developing countries experience longer periods of symptomatic recovery than patients in the West, the burden of illness is severe and affects the productivity and quality of life of many families. The predicament of patients and families is likely to worsen as many of these countries are experiencing the pressures of economic restructuring, increasing income inequality, unemployment, and cuts in public spending on health care. The traditional economic and psychosocial resources of the family, which in the past have been capable of absorbing much of the impact of severe mental illness, may quickly become eroded.

Considering the challenges in managing the impact of established psychotic disorders, fostering a preventive approach is becoming increasingly popular. Long-term outcomes in those with schizophrenia are known to improve with earlier treatment and shorter durations of untreated psychosis (DUP). The median DUP in the United States is estimated at 1 to 3 years, much higher than is recommended by the WHO of 12 weeks or lower. Long DUP is associated with worse symptomatic and social outcomes, decreased chance of remission, underemployment, and quality of life. DUP in schizophrenia is determined by delays in accessing mental health services, and earlier recognition shows the most promise in lessening the global burden of the disease. Universal preventive approaches, such as clinical trials of dietary phosphatidylcholine supplementation in healthy pregnant women beginning in the second trimester, may reduce the likelihood of abnormal fetal brain development. Other candidate universal interventions for primary prevention of schizophrenia may include school-based interventions, e.g., targeting victimization and substance abuse. In particular, minimizing cannabis use early in developing and reducing use frequency may improve mental health literacy in those who are at high risk for schizophrenia and can help facilitate earlier intervention and engagement with services. The establishment of first-episode psychosis or CHR services has burgeoned around the world, but primarily in North America, Australia, and Europe. These coordinated specialty care (CSC) services generally involve psychotherapy, psychopharmacology, family education and support, case management, and vocational and educational support tailored to early psychosis youth. Individuals who later develop psychosis after engaging with CHR-P services demonstrate a significant decrease in DUP compared to those who do not get these services. CSC services for first-episode psychosis have led to improved quality of life, reduced psychotic symptoms, and greater involvement in work and school compared to community care after 2 years compared to usual care.

A common denominator in the great variety of situations and systems of care for people afflicted with psychotic illnesses in the developing world is that the modest successes achieved in some of these countries are vulnerable to the local repercussions of economic globalization. In all societies, the stigma surrounding mental illness is likely to mean that people with severe mental disorders are among the first to be disadvantaged by shrinking government health expenditures. It is, therefore, imperative to initiate proactive measures on an international scale to forestall such developments.

FURTHER READINGS

Bennett D, Rosenheck R. Socioeconomic status and the effectiveness of treatment for first-episode psychosis. *Health Serv Res*. 2021;56(3):409–417.

Bhugra D, Bhui K, eds. *Textbook of Cultural Psychiatry*. Cambridge University Press; 2007.

Bitan DT, Krieger I, Kridin K, et al. COVID-19 prevalence and mortality among schizophrenia patients: a large-scale retrospective cohort study. *Schizophr Bull.* 2021;47(5):1211–1217.

Bourque F, van der Ven E, Malla A. A meta-analysis of the risk for psychotic disorders among first- and second-generation immigrants. *Psychol Med*. 2011;41(5):897–910.

Chong HY, Teoh SL, Bin-ChiaWu D, Kotirum S, Chiou C-F, Chaiyakunapruk N. Global economic burden of schizophrenia: a systematic review. *Neuropsychiatr Dis Treat*. 2016;12:357–373.

de Pablo GS, Radua J, Pereira J, et al. Probability of transition to psychosis in individuals at clinical high risk: an updated meta-analysis. *JAMA Psychiatry*. 2021;78(9):970–978.

Fearon P, Kirkbride JB, Morgan C, et al. Incidence of schizophrenia and other psychoses in ethnic minority groups: results from the MRC AESOP Study. *Psychol Med.* 2006;36(11):1541–1550.

Fusar-Poli P, McGorry PD, Kane JM. Improving outcomes of first-episode psychosis: an overview. *World Psychiatry*. 2017;16(3):251–265.

GBD 2016 Alcohol and Drug Use Collaborators. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Psychiatry*. 2018;5(12):987–1012.

He H, Liu Q, Li N, et al. Trends in the incidence and DALYs of schizophrenia at the global, regional and national levels: results from the Global Burden of Disease Study 2017. *Epidemiol Psychiatr Sci*. 2020;29:e91.

Hjorthøj C, Stürup AE, McGrath JJ, Nordentoft M. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. *Lancet Psychiatry*. 2017;4(4):295–301.

*Hopper K, Harrison G, Janca A, Sartorius N, eds. *Recovery From Schizophrenia. An International Perspective. A Report from the WHO Collaborative Project, The International Study of Schizophrenia*. Oxford University Press; 2007.

Hunt GE, Large MM, Cleary M, Lai HMX, Saunders JB. Prevalence of comorbid substance use in schizophrenia spectrum disorders in community and clinical settings, 1990–2017: systematic review and meta-analysis. *Drug Alcohol Depend*. 2018;191:234–258.

Jongsma HE, Turner C, Kirkbride JB, Jones PB. International incidence of psychotic disorders, 2002–17: a systematic review and meta-analysis. *Lancet Public Health*. 2019;4(5):e229–44.

Kane JM, Robinson DG, Schooler NR, et al. Comprehensive versus usual community care for first-episode psychosis: 2-year outcomes from the NIMH RAISE early treatment program. *Am J Psychiatry*. 2016:173(4):362–372.

Mamah D, Mutiso VN, Ndetei DM. Psychotic-like experiences among 9,564 Kenyan adolescents and young adults. *Psychiatry Res.* 2021;302:113994.

Marconi A, Di Forti M, Lewis CM, Murray RM, Vassos E. Meta-analysis of the association between the level of cannabis use and risk of psychosis. *Schizophr Bull*. 2016;42(5):1262–1269.

McGrath JJ, Saha S, Al-Hamzawi A, et al. Psychotic experiences in the general population: a cross-national analysis based on 31,261 respondents from 18 countries. *JAMA Psychiatry*. 2015;72(7):697–705.

Moreno-Küstner B, Martin C, Pastor L. Prevalence of psychotic disorders and its association with methodological issues. A systematic review and meta-analyses. *PLoS One*. 2018;13(4):e0195687.

*Murray RM, Jones PB, Susser E, van Os J, Cannon M, eds. *The Epidemiology of Schizophrenia*. Cambridge University Press; 2003.

Nuevo R, Chatterji S, Verdes E, Naidoo N, Arango C, Ayuso-Mateos JL. The continuum of psychotic symptoms in the general population: a cross-national study. *Schizophr Bull*. 2012;38(3):475–485.

Olbert CM, Nagendra A, Buck B. Meta-analysis of black vs. white racial disparity in schizophrenia diagnosis in the United States: do structured assessments attenuate racial disparities? *J Abnorm Psychol*. 2018;127(1):104–115.

*Perälä J, Suvisaari J, Saarni SI, et al. Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Arch Gen Psychiatry*. 2007;64(1):19–28.

Phillips MR, Yang G, Li S, Li Y. Suicide and the unique prevalence pattern of schizophrenia in mainland China: a retrospective observational study. *Lancet*. 2004;364(9439):1062–1068.

Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS Med*. 2005;2(5):e411.

Schwartz RC, Blankenship DM. Racial disparities in psychotic disorder diagnosis: a review of empirical literature. *World J Psychiatry*. 2014;4(4):133–140.

Scully PJ, Owens JM, Kinsella A, Waddington JL. Schizophrenia, schizoaffective and bipolar disorder within an epidemiologically complete, homogeneous population in rural Ireland: small area variation in rate. *Schizophr Res*. 2004;67(2–3):143–155.

Semple DM, McIntosh AM, Lawrie SM. Cannabis as a risk factor for psychosis: systematic review. *J Psychopharmacol*. 2005;19(2):187–194.

Wu L-T, Zhu H, Swartz MS. Trends in cannabis use disorders among racial/ethnic population groups in the United States. *Drug Alcohol Depend*. 2016;165:181–190.

12.4 Genetics of Schizophrenia

Kimberley M. Kendall, B.Sc. (Hons.), M.B.B.Ch., Ph.D., M.R.C.Psych., Elliott Rees, B.Sc. (Hons.), Ph.D., and Michael J. Owen, B.Sc. (Hons.), M.B., Ch.B., Ph.D., F.R.C.P.Sych.

It has long been recognized that schizophrenia tends to run in families. Systematic attempts to examine the genetic basis of schizophrenia began in the early 1900s but started to gain traction in the latter half of the century.

FAMILY, TWIN, AND ADOPTION STUDIES

Attempts to establish that a disorder has a genetic susceptibility often begin with the demonstration that relatives of probands have an increased risk of developing the disorder (Fig. 12.4–1). Family studies demonstrated that schizophrenia has a high degree of familial loading. For example, siblings have around an 8- to 10-fold increased risk of developing the disorder compared to the general population. For many years, evidence from family studies was interpreted as showing that psychiatric disorders largely "breed true." That is to say that relatives of an index case are at increased risk only of the index disorder. However, in recent years, as data from better-powered family studies have accumulated, it has become clear that there is evidence for shared as well as independent genetic risk between schizophrenia and other adult psychiatric disorders, particularly bipolar disorder, and between schizophrenia and childhood neurodevelopmental disorders (NDDs).

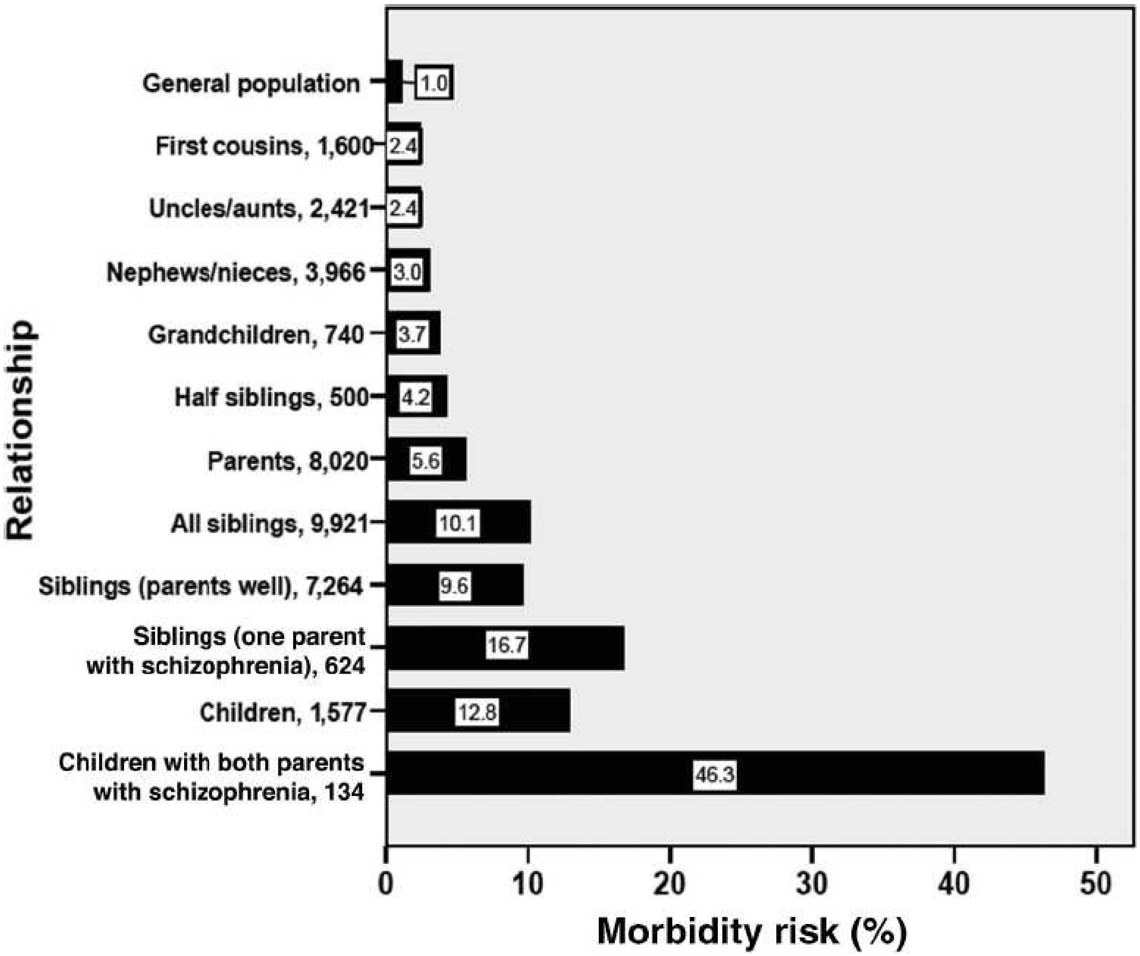

FIGURE 12.4–1. Morbidity risk for schizophrenia in the relatives of probands with schizophrenia. (The numbers are based on the study data from Gottesman II, Shields J. *Schizophrenia: The Epigenetic Puzzle*. Cambridge University Press; 1982.)

Family studies are unable to distinguish between risks conferred by genetic factors and environmental factors. In order to understand the basis of increased familial risk for schizophrenia, researchers turned to twin studies. Twins brought up in the same household share many aspects of their environment but different proportions of their DNA sequence depending on the nature of their twin relationship. Monozygotic (MZ) twins share all of their variable DNA sequence, whereas dizygotic (DZ) twins share, on average, 50%. Twin studies examine the proportion of twin pairs where both twins are affected or the proportion of cotwins who are also affected by a trait. This is expressed as the concordance rate. If MZ and DZ twins share their environments approximately equally (the shared environment assumption), a higher concordance rate in MZ twins implies that the disorder is, at least partly, genetic. If the concordance rate in MZ twins is less than 100%, this suggests that environmental factors also play a role. Meta-analyses of twin studies have estimated that the heritability of schizophrenia, the proportion of the phenotypic variance attributable to genetic factors, is around 81%.

The equal environment assumption has been called into question by the suggestion that MZ twins tend to share their environments to a greater extent than DZ twins. The separation of contributory factors in families is difficult because children share both genes and environment with their parents and siblings. Adoption studies provide an opportunity to investigate the genetic basis of a disorder while controlling for environmental factors. In these studies, adopted children, their biologic parents, and adoptive parents are studied. If there is a genetic component to a disorder, there should be a greater similarity between adopted children and their biologic parents than between adopted children and their adoptive parents. Adoption studies in schizophrenia established that children of biologic parents with schizophrenia have at least a 10% risk of developing the disorder or a related disorder. This risk is very similar to that reported in family studies, suggesting that most of the familial transmission of the disorder is genetic.

ADVANCES IN THE MOLECULAR GENETICS OF SCHIZOPHRENIA

Family, twin, and adoption studies successfully demonstrated that genetic factors play an important role in conferring the risk of schizophrenia. However, identification of risk at the level of DNA variation initially proved challenging, slow to advance, and difficult to replicate. The advent of more advanced techniques, including genome-wide association (GWA) studies, copy number variant (CNV) analysis, and next-generation sequencing, has resulted in greater success and optimism for the future.

Genome-Wide Association Studies

GWA studies examine common alleles, defined as variants observed in >1% of the general population, throughout the genome for association with a particular trait. The HapMap Project annotation of millions of single nucleotide polymorphisms (SNPs) demonstrated that many SNPs are highly correlated with neighboring SNPs. Therefore, it is not necessary to genotype all of them in order to gain nearly complete coverage of common variations in the genome. Furthermore, array-based technology has made it possible to genotype millions of SNPs on a single array, making the examination of common genetic variation across the whole genome in large samples a much less daunting prospect.

The earliest GWA studies were not particularly fruitful, but they did provide some useful lessons, in particular, the need to study extremely large samples due to the requirement to correct for the enormous multiple testing involved and the small odds ratios for risk conferred by common variants. It was estimated that GWA studies require a P value of $<5 \times 10^{-8}$ to achieve significance at a genome-wide level. This requires sample sizes in the order of tens of thousands, which could only be reached via collaboration between multiple research groups. The largest and most successful schizophrenia GWA studies to date were conducted by the Psychiatric Genomics Consortium (PGC), a collaboration bringing together data from research groups around the world. Studies by the PGC aggregate the most available schizophrenia samples with GWA genotypes into a single analysis and thus provide the most powerful studies of common alleles in schizophrenia.

The most recent PGC study analyzed SNP genotype data from 90 ancestry-matched, nonoverlapping case-control samples (67,390 cases and 94,015 controls; ~80% of the samples were of European ancestry, and ~20% were of East Asian ancestry). After a cohort association analysis, 7,858,078 common SNPs were meta-analyzed for association with schizophrenia. The results from this discovery analysis are depicted in Figure 12.4–2.

The PGC then meta-analyzed SNPs achieving a P value of $<1 \times 10^{-5}$ in this discovery sample with independent data from the deCODE genetics sample (1,979 cases and 142,627). This meta-analysis, which consisted of 69,369 cases and 236,642 controls, discovered 342 linkage-disequilibrium independent common allele associations that exceeded genome-wide significance ($P <5 \times 10^{-8}$). These significant associations mapped to 287 distinct genomic loci.

Schizophrenia common risk alleles are overrepresented among genes that are highly expressed in the brain, with many different brain regions showing evidence for association. However, when considering specific brain cell types and tissues, schizophrenia common risk alleles are concentrated among genes expressed in excitatory and inhibitory neurons, with the strongest enrichment observed for genes expressed in glutamatergic neurons in deep layers of the cortex, amygdala, and hippocampus. Genes related to particular biologic processes are also enriched for schizophrenia common alleles, such as neuronal excitability, development, and structure.

Moreover, schizophrenia common risk alleles have provided strong evidence for synaptic dysfunction in schizophrenia pathogenesis, with associations reported in genes with pre- and postsynaptic locations.

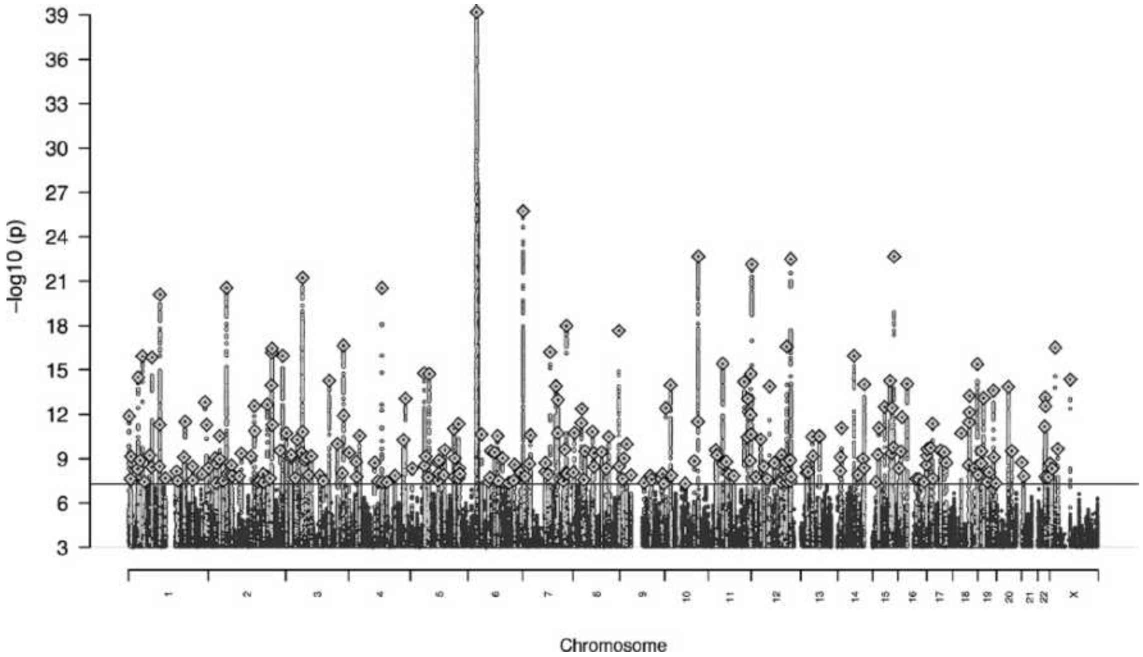

FIGURE 12.4–2. Manhattan plot showing the discovery results from the most recent GWAS of schizophrenia. x-axis–chromosomal position; y-axis–significance of the association (−log10[P]); horizontal line–genome-wide level of significance (5 × 10^{-8}).

Some genomic loci harboring schizophrenia common allele associations contain genes thought to be related to the etiology and treatment of schizophrenia, including the dopamine receptor D2 (*DRD2*), glutamatergic neurotransmission, and synaptic plasticity (*GRM3*, *GRIN2A*, and *SRR*), and calcium signaling (*CACNA1C*, *CACNB2*, and *CACNA1I*). However, due to the complex linkage disequilibrium structure surrounding associated SNPs, the causal variant or gene is not usually known. The PGC study used a method called fine mapping, which, together with functional genomic data, helped to prioritize 120 genes within associated loci as being the most likely causal gene. These prioritized genes were indeed enriched for characteristics known to be pathogenic for neuropsychiatric disorders, such as brain expression and intolerance to highly damaging classes of mutation, such as loss-of-function variants. Thirteen of the prioritized genes had either pre- or postsynaptic locations and functions related to voltage-gated calcium and chloride channels (*CACNA1C*, *CLCN3*), metabotropic receptors (glutamate and GABA), the ligand-gated NMDA receptor subunit (*GRIN2A*), synaptic organization and differentiation (*ADGRB3*, *LRRC4B*, *GPM6A*, *IL1RAPL1*), the trans-synaptic signaling complex (*IL1RAPL1*, *SNAP91),* and modulation of chemical transmission (*MAPK3*, *DCC*, *CLCN3*).

It seems likely that the number of genome-wide significant common allele associations identified in schizophrenia will continue to increase as larger samples are analyzed. This has become apparent with the development of methods, such as the polygenic risk score, which quantify the *en masse* effects of common risk alleles, including those risk alleles that are not individually supported at genome-wide levels of significance. These methods have estimated that relatively common, small-effect alleles (odds ratios <1.3) account for at least 25% of the total liability to schizophrenia (about 33% of genetic liability). However, current polygenic risk scores for schizophrenia explain only 7.7% of the variance in liability and thus are unable to accurately predict schizophrenia diagnosis in the general population (Area Under the Receiver Operating Characteristic Curve of 0.71). Nevertheless, polygenic risk scores show great promise in research settings for stratifying patients and may most rapidly find clinical application in groups at elevated risk of schizophrenia, for example, in helping predict which individuals in psychosis high-risk groups will later develop the disorder.

GWA studies have also established that other psychiatric disorders, such as bipolar disorder, attention-deficit hyperactivity disorder, and major depression, have a highly polygenic genetic basis. By comparing the results from GWA studies across psychiatric disorders, it has become increasingly apparent that some common risk alleles for schizophrenia also confer risk for other psychiatric disorders. In particular, common risk alleles for schizophrenia are highly correlated with those for bipolar disorder. Schizophrenia common risk alleles are also correlated, but to a lesser extent, with childhood-onset neuropsychiatric disorders, such as attention-deficit hyperactivity disorder, autism spectrum disorder (ASD), and developmental disorders. Moreover, studies that combine GWA data across schizophrenia and bipolar disorder have identified both shared and divergent common allele associations across these disorders. These findings suggest that, in part, the pathophysiologic mechanisms underlying schizophrenia are shared with other psychiatric disorders.

CNVs and Schizophrenia

CNVs are the deletion or duplication of segments of DNA greater than 1 kb in size. The earliest report of an association between a rare CNV and the risk of schizophrenia was for a deletion of around 2.5 Mb in size on the long arm of chromosome 22. The 22q11.2 deletion had long been known to cause a severe congenital disorder variously called DiGeorge syndrome, velocardiofacial syndrome, or Sphrintzen syndrome. Carriers of this CNV were shown to have up to 30% risk of developing schizophrenia or another psychotic disorder.

Subsequently, the development of high-resolution and high throughput microarray technologies allowed researchers to interrogate the whole genome and to find other CNVs also associated with an increased risk of schizophrenia. The first new CNV locus to be identified was a single gene–*NRXN1* (neurexin 1) on chromosome 2. Deletions in this region are of variable size and can affect different sections of the gene. In 2009, Rujescu et al. showed that it is deletion CNVs that intersect the exons of *NRXN1* that increase the risk of schizophrenia. *NRXN1* encodes a cell adhesion molecule, which is specific to synapses in the brain. It mediates interactions between pre- and postsynaptic structures and plays an important role in the formation, maintenance, and release of neurotransmitters.

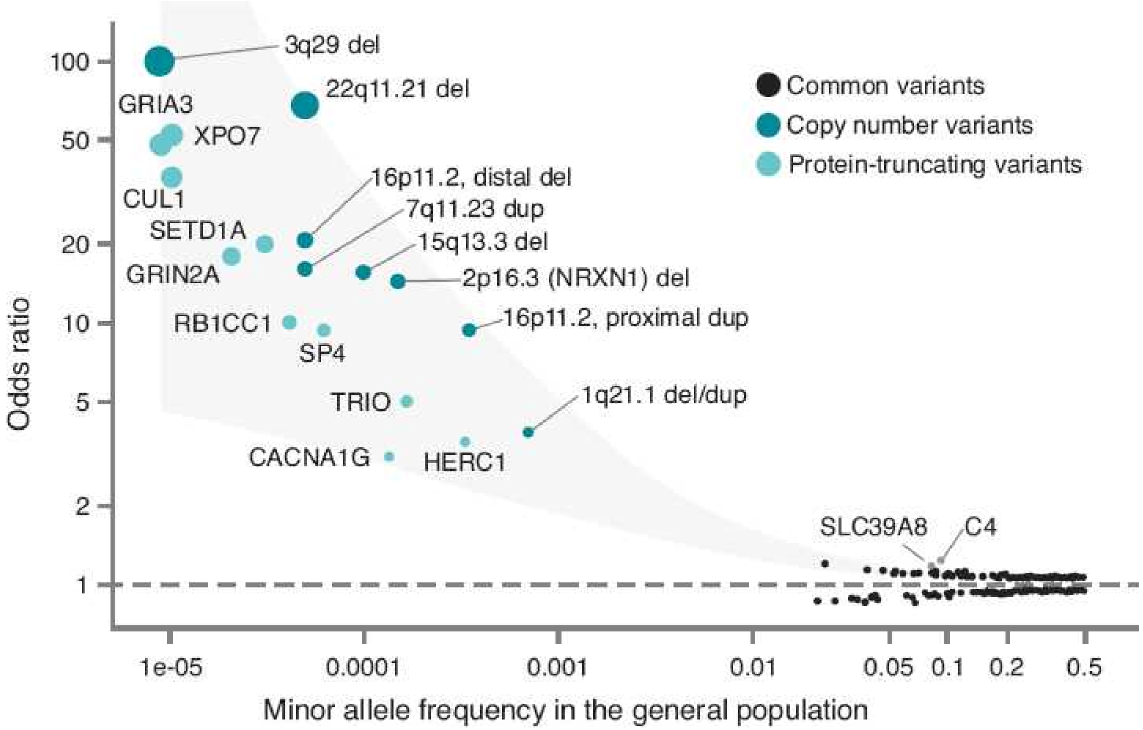

FIGURE 12.4–3. The genetic architecture of schizophrenia. This diagram shows significant genetic associations for schizophrenia from the most recent GWAS, CNV, and sequencing studies. y-axis–odds ratio; x-axis–minor allele frequency in the general population. (See eBook for color images.)

Additional CNV loci have since been identified as conferring an increased risk of schizophrenia (Fig. 12.4–3). To date, a total of 12 CNVs at 11 individual loci have been identified as associated with the disorder at genome-wide levels of statistical significance ($P < 4.1 \times 10^{-4}$)–deletions at 1q21.1, *NRXN1* (exonic), 3q29, 15q11.2, 15q13.3, 16p12.1, and 22q11.2, and duplications at 1q21.1, the Williams–Beuren syndrome locus at 7q11.23, the Angelman/Prader Willi syndrome locus at 15q11-q13, 16p13.11, and 16p11.2. Many schizophrenia-associated CNVs share certain features. They are rare and therefore require very large samples to confirm their role. Many disrupt multiple genes, which likely explains their high levels of pathogenicity. Many are flanked by large stretches of homologous DNA called low-copy repeats that can misalign during meiosis or mitosis and cause the DNA between them to be deleted or duplicated. Therefore, they are usually of the same size in different individuals, and their mutation rates are high compared to other CNV loci. Finally, all CNVs associated with schizophrenia have also been implicated in autism and intellectual disability (ID).

Large studies of rare CNVs in schizophrenia have confirmed the role of previously implicated loci in schizophrenia risk. However, when these were excluded, there remained a residual burden of rare CNVs in schizophrenia, suggesting there are further risk loci yet to be implicated. CNVs at these additional loci may be very rare, requiring even larger samples to confirm their role. The largest schizophrenia CNV study to date examined the effects, on schizophrenia risk, of 51 CNVs previously implicated in ID (excluding 11 known schizophrenia loci). ID CNVs were enriched in schizophrenia cases, and this study implicated the novel locus 16p12.1 deletion in the risk of schizophrenia.

Penetrance of CNVs. All of the individual CNVs described confer high odds ratios for schizophrenia, in the range of 3 to 58 (Fig. 12.4–3). However, each of these CNVs has also been implicated in the risk of other developmental disorders such as developmental delay, ASD, or congenital malformations. When considering the penetrance of CNVs, for example, for the purposes of genetic counseling, one should take into account the risk of developing any serious disorder, not just schizophrenia. A study of the penetrance of rare CNVs found estimates for developmental delay/ASD/congenital malformations to be higher than that for schizophrenia. Taken together, the penetrance for any serious developmental disorder is substantial, averaging just over 40%, and is accounted for mostly by nonpsychotic disorders.

Sequencing Studies

When examining association with disease, GWA studies are used to detect common variations, particularly SNPs, and the genotyping platforms used can also detect CNVs. However, these platforms will not detect rare single nucleotide variants (SNVs) and small insertions and deletions (indels). Exome-sequencing, which involves sequencing the protein-coding regions in the genome, is now widely used to identify rare SNVs and indels (collectively termed rare coding variants [RCVs]) and test them for association with disease.

Early exome sequencing studies of schizophrenia were underpowered to identify individual genes enriched for RCVs with genome-wide levels of significance, but they were able to demonstrate that people with schizophrenia have an elevated burden of RCVs within specific groups of genes. For example, RCVs in people with schizophrenia were found to be enriched among genes related to glutamatergic postsynaptic proteins, as well as more broadly among ~3,500 genes for which damaging mutations are known to be subjected to strong levels of purifying selection in the general population.

In 2016, an analysis of exome-sequencing data from 4,264 schizophrenia cases, 9,343 controls, and 1,077 schizophrenia trios identified RCVs in *SETD1A* to confer substantial risk for schizophrenia odds ratio = 37.6 (95% CI = 8 – 353). This finding represented the first gene associated with RCVs in schizophrenia with genome-wide levels of significance. Rare mutations in *SETD1A* were also shown to increase the risk for severe developmental disorders, and several schizophrenia carriers of *SETD1A* mutations also had a comorbid ID. *SETD1A* encodes a methyltransferase that catalyzes the methylation of lysine residues in histone H3 and is known to influence the transcriptional regulation of genes important for aspects of cell cycle regulation. Association between *SETD1A* and schizophrenia, therefore, suggests disrupted epigenetic modification as a potential mechanism underlying schizophrenia.

To advance the discovery of individual genes associated with RCVs in schizophrenia, the Schizophrenia Exome Sequencing Meta-analysis (SCHEMA) consortium was established in 2017 to amalgamate new and existing sequencing data sets collected from diverse global populations. After quality control, the SCHEMA study evaluated RCVs in 24,248 cases and 97,322 controls, and de novo variants from 3,402 trios. This analysis identified 32 genes enriched for RCVs at a false-discovery rate <5%, of which 10 genes surpassed genome-wide levels of significance (i.e., Bonferroni correction for 22,833 tests). RCVs within the 10 genome-wide significant genes conferred substantial risks of schizophrenia, with estimated odds ratios ranging between 3 and 50. The functions of the proteins encoded by these 10 genes are diverse and include ion transport (*CACNA1G*, *GRIN2A*, *GRIA3*), neuronal migration and growth (*TRIO*), transcriptional regulation (*SP4*, *RB1CC1*, *SETD1A*), nuclear transport (*XPO7*), and ubiquitin ligation (*CUL1*, *HERC1*). The findings from the SCHEMA study provide strong support for specific mechanistic hypotheses for schizophrenia; for example, an association of *GRIN2A*, which encodes an NMDA receptor subunit, and *GRIA3*, which encodes an AMPA receptor subunit, support dysregulation of the glutamatergic system in schizophrenia pathophysiology.

Sequencing studies have provided further evidence that schizophrenia and early onset NDDs, such as ASD and ID, share specific risk genes. Groups of genes affected by de novo mutations in people with NDDs are enriched for similar types of mutation in schizophrenia, and there is emerging evidence that within shared risk genes, the same mutation can sometimes confer the risk of schizophrenia and NDDs. However, a small number of genes have been reported to be enriched for different types of mutation in schizophrenia and NDDs; for example, while *TRIO*, *GRIN2A,* and *CACNA1G* are enriched primarily for protein-truncating variants in schizophrenia, only missense variants in these genes are enriched in NDDs. This suggests that within some shared risk genes, different changes in the gene's function can confer risk to different disorders.

CONVERGENCE ONTO BIOLOGIC MECHANISMS

As indicated, despite the undoubted complexity of the genetic picture that is emerging, work using different approaches is beginning to reveal some convergence onto a coherent set of biologic processes. Results from GWA studies, CNV, and sequencing studies all point to a functionally related set of synaptic proteins involved in synaptic plasticity, learning, and memory. Among the gene sets implicated are the ARC and NMDAR complexes, targets of FMRP, and voltage-gated calcium channels. Genes more widely expressed in brain tissues, as well as genes expressed in both excitatory and inhibitory neurons, are associated with both common and rare variants in schizophrenia. Genes prioritized by the fine-mapping approaches used in the PGC analysis of common alleles are enriched for rare variants in the SCHEMA analysis, and two of the top 10 most significant SCHEMA genes (*GRIN2A* and *SP4*) have robust common allele associations in the PGC study. These findings demonstrate that a wide spectrum of risk alleles disrupt specific biologic processes in people with schizophrenia and that insights into disease pathology can be gained by integrating findings from sequencing and SNP genotype data. Although it seems highly likely that there are other systems and mechanisms involved, including the dopamine system, and that these can be expected to emerge with further and larger GWA studies and sequencing studies, the current findings offer numerous entry points for neuroscientists to probe the biologic basis of schizophrenia.

THE EMERGING GENETIC ARCHITECTURE OF SCHIZOPHRENIA

While much genetic risk is still unexplained at the level of individual risk variants, the findings to date support a general framework for the genetic architecture of schizophrenia. As expected from genetic epidemiology and population genetics, there is now empirical evidence for a spectrum of allelic risk. There is a marked inverse relationship between effect size and frequency, as predicted from the effects of natural selection in a disorder associated with marked reductions in fecundity. There are contributions from alleles that are common in the population. The effect sizes of these are small, enabling them to escape the effects of natural selection, but as a group, they probably account for about a third of genetic liability. There is also a contribution from rare alleles, some of which have large effects on individual risk of disease pending their removal from the population by selection. The relationship between effect size and allele frequency is illustrated in Figure 12.4–3. The relative contribution of alleles of different effect sizes and frequencies is unclear, but this should be clarified as the larger whole exome and whole genome sequencing studies are conducted. However, it seems likely that alleles across the risk spectrum will be found with their effect sizes constrained by natural selection.

It has been shown that people with schizophrenia who carry rare risk CNVs still have an elevated burden of common risk alleles compared with controls, which suggests that cases who carry rare high penetrant mutations and those who do not have overlapping etiologies. However, among people with schizophrenia, those who carry a rare, highly penetrant variant have a lower burden of common risk alleles compared to cases without a rare mutation. This is consistent with a liability threshold model, where both common and rare mutations act additively to confer the risk of schizophrenia.

NEURODEVELOPMENTAL SPECTRUM

An important implication of recent studies is that genetic risk does not recognize our current definitions of disease. Extensive pleiotropy with respect to diagnosis has been demonstrated for common alleles at the level of individual SNPs, genes, and the *en masse* effects of multiple common risk alleles and for rare variants, including both CNVs and SNVs. These findings, combined with the recognition of the phenotypic similarities between some disorders, led to the development of the neurodevelopmental spectrum. This model proposes the conceptualization of schizophrenia and disorders considered to have a neurodevelopmental etiology, such as ID and ASD, as lying on an etiologic and phenotypic spectrum. The greatest level of neurodevelopmental impairment is observed in ID, and this corresponds to a greater loading of deleterious genetic variants. As one moves across the spectrum, the level of neurodevelopmental impairment and the rate of deleterious genetic variants decreases. Taken together, the existing evidence brings into question whether current diagnostic categories are the most appropriate for future genetic studies, for research aiming to understand relationships between genetic risk and clinical outcome, and to best understand pathophysiology.

CLINICAL IMPLICATIONS

As the body of evidence for the genetic basis of schizophrenia and related disorders increases, so do the clinical implications for individuals with these disorders and their relatives. Interpreting and communicating these findings is not straightforward but is of vital importance.

In the case of schizophrenia, one may inherit an increased susceptibility to developing the disorder. This does not make it inevitable that the disorder will develop, as even MZ twins are concordant for the disorder in only around half of cases. Genetic counseling may be appropriate to enable individuals to develop a greater understanding of the nature of their disorder and its familial basis. However, the complex genetic architecture of schizophrenia means that genetic testing is unlikely to be appropriate for all individuals with the disorder. The exception is for carriers of specific CNVs, for which reasonably accurate estimates of penetrance are available. Importantly, while these CNVs increase the risk of disorders across the neurodevelopmental spectrum, they also increase the risk of cognitive impairment and physical health phenotypes with implications for morbidity and mortality. Establishing CNV carrier status may inform clinical care by providing information on the potential physical conditions for which monitoring should be instigated. A key piece of work will be developing criteria for which CNV, and other genetic testing, should be offered. In the future, it may also be possible to use polygenic risk scoring to stratify individuals and tailor medication regimes in a biologically informed manner.

FURTHER READINGS

Bassett AS, Marshall CR, Lionel AC, Chow EWC, Scherer SW. Copy number variations and risk for schizophrenia in 22q11.2 deletion syndrome. *Hum Mol Genet*. 2008;17(24):4045–4053.

Crawford K, Bracher-Smith M, Kendall KM, et al. Medical consequences of pathogenic CNVs in adults: analysis of the UK biobank. *J Med Genet*. 2019;56:131–138.

Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature*. 2014;506:179–184.

Gottesman II, Shields J. *Schizophrenia, the Epigenetic Puzzle*. Cambridge University Press; 1982.

Hall J, Trent S, Thomas KL, O’Donovan MC, Owen MJ. Genetic risk for schizophrenia: convergence on synaptic pathways involved in plasticity. *Biol Psychiatry*. 2015;77(1):52–58.

Ingraham LJ, Kety SS. Adoption studies of schizophrenia. *Am J Med Genet*. 2000;97(1):18–22.

International HapMap Consortium, Frazer KA, Ballinger DG, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007;449(7164):851–861.

Kendall KM, Rees E, Escott-Price V, et al. Cognitive performance among carriers of pathogenic copy number variants: analysis of 152,000 UK biobank subjects. *Biol Psychiatry*. 2017;82(2):103–110.

Kirov G, Dilihan G, Chen W. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet*. 2008;17(3):458–465.

Kirov G, Rees E, Walters JT, et al. The penetrance of copy number variations for schizophrenia and developmental delay. *Biol Psychiatry*. 2014;75(5):378–385.

Marshall CR, Howrigan DP, Merico D, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet*. 2017;49(1):27–35.

Murray GK, Lin T, Austin J, McGrath JJ, Hickie IB, Wray NR. Could polygenic risk scores be useful in psychiatry?: a review. *JAMA Psychiatry*. 2021;78(2):210–219.

O’Donovan MC, Owen MJ. The implications of the shared genetics of psychiatric disorders. *Nat Med*. 2016;22:1214–1219.

Owen MJ, O'Donovan MC. Schizophrenia and the neurodevelopmental continuum: evidence from genomics. *World Psychiatry*. 2017;16(3):227–235.

Rees E, Han J, Morgan J, et al. De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. *Nat Neurosci*. 2020;23(2):179–184.

Rees E, Kendall K, Pardiñas AF, et al. Analysis of intellectual disability copy number variants for association with schizophrenia. *JAMA Psychiatry*. 2016;73(9):963–969.

Rees E, Owen MJ. Translating insights from neuropsychiatric genetics and genomics for precision psychiatry. *Genome Med*. 2020;12(1):43.

Rees E, Walters JTR, Georgieva L, et al. Analysis of copy number variations at 15 schizophrenia-associated loci. *Br J Psychiatry*. 2014;204(2):108–114.

Ruderfer DM, Ripke S, McQuillin A, et al. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell*. 2018;173(7):1705–1715.e16.

Rujescu D, Ingason A, Cichon S, et al. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet*. 2009;18(5):988–996.

Trubetskoy V, Pardiñas A, Qi T et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604:502–508.

Singh T, Kurki MI, Curtis D, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci*. 2016;19(4):571–577.

Singh T, Poterba T, Curtis D et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*. 2022;604:509–516.

Singh T, Walters JTR, Johnstone M, et al. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat Genet*. 2017;49(8):1167–1173.

Sullivan PF, Agrawal A, Bulik CM, et al. Psychiatric genomics: an update and an agenda. *Am J Psychiatry*. 2018;175(1):15–27.

Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003;60(12):1187–1192.

Tansey KE, Rees E, Linden DE, et al. Common alleles contribute to schizophrenia in CNV carriers. *Mol Psychiatry*. 2016;21(8):1085–1089.

Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–678.

12.5 The Clinical Epidemiology of Psychosis Spectrum Disorder

JIM VAN OS, M.D., PH.D., AND ULRICH REININGHAUS, PH.D.

WHAT IS CLINICAL EPIDEMIOLOGY?

Clinical epidemiology is the basic science of preventive medicine. It examines the variation in susceptibility, onset, course, and outcome of a disease and the reasons behind this variation within defined populations. It is the science of making predictions about the emergence (onset) and persistence (prognosis) of disease through the systematic identification and evaluation of inherited and acquired characteristics, attributes, behaviors, and exposures, known as risk factors (determinants). These risk factors increase an individual's likelihood of developing (1) a precursor of the clinical disorder (a condition or state preceding the recognized onset of a disease), (2) a clinical disease, or (3) a particular course and outcome trajectory for the disease process. Determinants rarely act in isolation to increase risk; rather, their effects are cumulative. Such complementary actions can be additive (the linear sum of the individual effects) or synergistic (interaction of effects) when individual risk factor effects are modified by particular environmental conditions, biologic stochastic events, or other known risk factors. Further, there is no reason to assume that the risk factors for the emergence of a clinical disorder and those influencing persistence are congruent. Evidence that the same risk factor is implicated in both onset and persistence of a disease (discrete major effect) could aid our understanding of etiologic mechanisms. However, the lack of discrete effects suggests the course of a disease is not predetermined and is potentially modifiable after onset.

Clinical epidemiology attempts to harness the heterogeneity observed in risk factor profiles, course, and outcomes and uses this variation to generate hypotheses and develop empirically based conceptual frameworks (models) of mental health as well as public mental health services, including mental health promotion, primary, secondary, and tertiary prevention strategies. Identifying the risk factors operating prior to the onset of the disease is fundamental to effective evidence-based primary (universal) prevention, while the risk factors predicting course and outcome help identify patients at higher risk of poor outcome, so guiding the development of secondary (selective) and tertiary (indicated) preventative strategies and interventions tailored for specific individuals in clinical settings.

Extensive heterogeneity is the hallmark of psychosis spectrum disorder, with manifest variation in (sensitivity for) risk factors for onset, symptomatology, course, and outcome. Predictive power is typically low and attributable fractions are small. Recent years have seen a transformation in our understanding of psychosis spectrum disorder. Based on evidence on shared aetiologic factors as well as a transdiagnostic symptom dimension that cuts across diagnostic boundaries of nonaffective and affective psychotic disorders in the bipolar-psychosis network on intermediate phenotypes (B-SNIP) and European Network of National Networks studying Gene-Environment Interactions in Psychosis (EU-GEI) consortia, it has been proposed to reconceptualize these disorders as “psychosis spectrum disorder” or, in the Hierarchical Taxonomy of Psychopathology (HiTOP), as “psychosis super spectrum.” The median age of onset of disorders in the psychosis spectrum is 25, with an IQR of 20 to 34. The total lifetime prevalence of disorders in the psychosis spectrum is 3.5%, of which less than a third meets criteria for “schizophrenia.” It has been proposed that for the purpose of clinical epidemiology, the diagnosis of the “psychosis spectrum” is best approached from the perspective of clinical characterization. Clinical characterization is focused on specifically increasing patient value by combining or replacing categorical ICD/DSM diagnosis with transdiagnostic, personalized, and contextual information on symptoms, clinical severity, clinical staging, and antecedent and concomitant variables. These efforts, taken together, are ultimately geared to produce new evidence on the interface between transdiagnostic risk factors, mechanisms, and phenotypic constructs as a means of connecting process and structure in understanding more fully what underpins psychotic disorders. A central tenet of modern risk factor epidemiology is that virtually all diseases are a consequence of the interplay between genetics and environmental characteristics. Therefore, recent studies investigate how differences in genetic endowment moderate the effect of environmental risk factors (gene–environment interactions) across the psychosis spectrum.

RISK FACTORS FOR THE ONSET OF PSYCHOSIS SPECTRUM DISORDER

Genetic Risk: Indirect Evidence from Epidemiologic Studies

Family Studies: Family studies compare the morbidity risk (the proportion of affected to unaffected relatives, usually adjusted for variation in nonaffected participants time at risk) for probands with the diagnosis of schizophrenia or psychosis spectrum disorder, to the morbidity risk in the general population or demographically matched control groups. Results from a systematic review of the studies published during the period 1980 to 2003 found the first-degree relatives of probands to have a higher morbidity risk for psychosis spectrum disorder as compared to the relatives of controls.

Estimates ranged from 2% to 9% (in studies that adjusted for time at risk using Weinberg, Strömgren, or Kaplan–Meier methods) compared to a morbidity risk in the relatives of controls at about 0.5%. The effect varied according to the diagnostic classification employed in the study, with broader definitions of the diagnosis of schizophrenia, such as Research Diagnostic Criteria, tending to show larger estimated effects than the narrower definitions employed in post-third editions of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM). Similarly, studies using control groups unscreened for mental disorders showed more conservative estimates. There was also evidence of clinical heterogeneity, with the relatives of early-onset cases having a higher morbidity risk compared to the relatives of late-onset cases. Further studies have examined familiar aggregation for categorical subtypes, symptom dimensions derived from factor analysis, and individual symptoms. The findings are inconsistent and show significant methodologic heterogeneity. Most family studies have investigated morbidity risk in first-degree relatives; the few studies of second- and third-degree relatives suggest an increase in risk with increasing relationship closeness, the structural form of which may not be linear.

Adoption Studies: Most children inherit their genetic characteristics as well as their early rearing environment from their biologic parents. Adoption provides an opportunity to disentangle these genetic and rearing effects. Evidence from adoption studies have primarily come from a small group of studies. An adoptee's family study, using the Danish national linked data set, compared rates of the diagnosis of schizophrenia in the adoptive and biologic families of adoptees with a diagnosis of schizophrenia. Results reported over the past 30 years (using different diagnostic criteria) consistently show the biologic relatives to be at a higher risk of a diagnosis of schizophrenia than the adoptive relatives, who have a similar estimated risk to the rest of the general population. The second strategy compares the rates of a diagnosis of schizophrenia in adopted children for parents with and without a diagnosis of schizophrenia. Early studies using this methodologic approach consistently demonstrated higher rates of schizophrenia in the adopted offspring of affected mothers, compared to the rates in the offspring of the unaffected controls. A recent Finnish study has improved on these earlier methodologies by identifying larger samples, using standardized interviews with home assessments and DSM operationalized diagnostic criteria. Its results support the early findings, showing a morbidity risk in the adopted offspring of mothers with a diagnosis of schizophrenia of 5.1%, as compared with 1.6% in the adopted offspring of control mothers, a finding that has been replicated in a recent, large Swedish registry study.

Twin Studies: There are some studies that compare the concordance rates for monozygotic (identical) twins (assumed to inherit all their genes in common) with those of dizygotic (nonidentical) twins (who on average share around 50% of their genes). Such studies further assume that twin pairs are equally exposed to environmental risk factors for psychosis spectrum disorder. Statistical combination of the published studies that used DSM-III-R criteria to define caseness estimated pooled concordance rates of 50% in monozygotic twins and 4.1% in the dizygotic pairs; the pooled estimates of concordance for studies using different diagnostic classification systems are actually very similar. The concordance rate for monozygotic twins was higher in twins with an early, rather than a late, onset of psychosis spectrum disorder. Twin studies investigating classical schizophrenia subtypes show higher monozygotic concordance rates for the hebephrenic subtype compared to the paranoid type but this may not be a conclusive finding. Combination twin and family studies find the risk for a diagnosis of schizophrenia in the offspring of the identical and unaffected twin to be the same as that seen for the offspring of the affected twin.

SPECIFICITY OF HERITABILITY

Relationship to Attenuated Psychosis Phenotypes

Clinicians have long recognized that some close relatives of patients with psychosis spectrum disorder, although never frankly psychotic, display symptoms and signs similar to the pathologic experiences seen in the clinical disorder. Some are attenuated forms of the positive/disorganized symptoms of psychotic disorder (odd behavior, odd speech, muddled thinking, loose associations, ideas of reference, magical thinking, suspiciousness, and illusions), while others resemble the negative symptoms (social withdrawal, aloofness, guardedness, poor rapport). The DSM classification system has combined these quasipsychotic (schizotypic) features under the diagnostic category of schizotypal personality disorder. As schizotypal, schizoid, and paranoid personality disorders are found to co-occur, they have been grouped together under the rubric Cluster A personality disorder within the manual. Family studies measuring lifetime prevalence for DSM Cluster A personality disorders (some studies have also included avoidant personality disorder) in first-degree relatives of patients with schizophrenia (and other psychotic disorders including psychotic bipolar disorder) find rates in the relatives of the probands (range: lifetime prevalence 2.8% to 8.2%) that are two to four times the rates found in the control families. The significant heterogeneity in estimated lifetime prevalence is primarily due to methodologic differences. Adoption and twin studies confirm this familiar aggregation. The familiar coaggregation of pathologic and attenuated schizotypal symptoms suggests that there may be a single continuum of liability across the schizotypic features and pathologic psychotic symptoms.

A recent systematic review of genome-wide association studies (GWAS) of attenuated psychotic symptoms and negative symptom (PENS) showed that in the most recent and well-powered studies, polygenic risk score (PRS) for schizophrenia was associated with all types of psychotic experience and negative symptom traits and that negative symptom traits also share genetic influences with major depression.

Relationship of Schizophrenia to Other Psychotic Disorders and Bipolar Disorder

Family and adoption studies of schizophrenia have generally shown an excess of risk in biologic relatives, not only for schizophrenia, but also for a broad range of psychotic disorders, such as schizophreniform, schizoaffective, delusional, and atypical delusional disorders. Biologic relatives also have higher rates of nonpsychotic disorders. A large, recent registry study of Swedish families reported evidence that schizophrenia and bipolar disorder partly share common genetic determinants. An extended adoption and cross-generational study form the same country showed that cross-generational transmission between bipolar disorder and broad schizophrenia appears to be entirely genetic with a moderate genetic correlation.

These studies do not help us elucidate the nature of this shared liability. A possible explanation would be an underlying continuous common liability for all psychotic disorders including bipolar disorder. Findings from GWAS such as those conducted by the Cross-Disorder Group of the Psychiatric Genomics Consortium indicate highly significant polygenic overlap between schizophrenia and not only bipolar but also major depressive disorder. These findings are complemented by evidence on shared intermediate and clinical phenotypes of schizophrenia, schizoaffective, and bipolar disorder from the EU-GEI and B-SNIP consortia. The Brainstorm Consortium quantified the genetic sharing of 25 brain disorders from GWAS of 265,218 patients and 784,643 control participants and assessed their relationship to 17 phenotypes from 1,191,588 individuals. The results showed that psychiatric disorders (psychotic and nonpsychotic) share common variant risk, whereas neurologic disorders appear more distinct from one another and from the psychiatric disorders. Overall, around two-thirds of genetic associations are common to schizophrenia, bipolar disorder, and major depressive disorder and overlaps also exist with genetic variants contributing to autism, attention-deficit/hyperactivity disorder, and intellectual disabilities. These findings suggest that PRS for mental disorders to a large extent represents transdiagnostic risk for mental suffering. PRS for schizophrenia in particular is associated with a variety of disorders, quality-of-life, and subclinical multidimensional phenotypes. Indeed, investigation of electronic health records from the United States reveals that PRS for schizophrenia is associated with not only a diagnosis of schizophrenia, but also diagnoses of other related psychiatric and medical conditions. Overall, the current evidence points toward a model of liability, which incorporates a common (shared) genetic susceptibility for psychosis spectrum disorder and many other mental disorders and traits.

GENETIC MODEL OF LIABILITY FOR PSYCHOSIS SPECTRUM DISORDER

The family, adoption, and twin studies provide irrevocable evidence for the involvement of a major genetic component in the liability for psychosis spectrum disorder. However, the familiar distributions are inconsistent with simple Mendelian modes of inheritance (single major locus) and are more compatible with a polygenetic mode of inheritance. One possible model is the *polygenetic threshold model*, which assumes most of the risk of psychosis spectrum disorder is due to additive effects of multiple small genes at different loci, and only a small amount of risk is associated with rare variants and copy number variants. However, the polygenetic threshold model does not adequately account for the exponential rise in the risk of psychosis spectrum disorder observed with increasing familiar closeness, which is more consistent with (at least some) epistatic gene–gene interaction. Nonetheless, the monozygotic/dizygotic twin correlations are significantly lower than would be expected with a full epistatic model. Recent years have seen rapid advances through large-scale collaboration in GWAS, which has generated replicated findings on a number of common risk alleles. However, the common variants identified to date explain not all of the genetic risks of psychosis spectrum disorder as established in family and twin studies. Molecular genetic analysis allows the estimation of a model that predicts trait values from genotype data expressed as a PRS. The amount of phenotypic variance explained by PRS is typically much lower than the amount of additive genetic variance estimated across twin studies. For arguably the most investigated mental disorder, schizophrenia, with an estimated twin heritability of 80%, tens of thousands of markers explain only 7% of the variance on the liability scale and around 24% of the variance on the observed 0 to 1 scale derived from the logistic regression model. Thus, a considerable “heritability gap” remains; the origin of which may represent environmental effects, indirect genetic effects within the family, rare genetic variants, gene-environment interplay, assortative mating, or other factors. It is likely that both genetic and environmental factors contribute to the development of psychosis spectrum disorder. The mechanisms by which these factors interact to influence risk remain unclear but will probably involve the complex interplay of both additive and interacting genes, environmental factors, gene-environment interactions (genetic moderation of environmental factors and environmental moderation of gene expression), as well as neural, cognitive, and affective mechanisms underlying the interplay of genetic and environmental factors. These are likely to be continuously distributed within the population. This model of risk may vary within and across individuals, families, or clinically defined subgroups of the disease, depending on the specific risk factors active within the person at different stages of development. If this probabilistic model of risk transmission is accepted, then environmental risk factors appear to be necessary complementary components in the development of psychosis spectrum disorder in many, if not all, cases.

Table 12.5–1.
Causal Criteria for the Onset of Psychosis

1. **Strength of the association**
   The argument states that strong associations are more likely to be causal because if they could be explained by another factor, the effect of the confounder would have to be even stronger than the observed association. While weaker associations are more likely due to noise and bias, it is important to remember that measures of association (e.g., the relative risk for odds ratios) are characteristic of the sample under study and depend on the relative prevalence of other causal factors.
2. **Dose-response relationship (biologic gradient)**
   Defines a unidirectional trend in disease frequency with increasing levels of exposure; however, this is neither necessary nor sufficient for causality as the presence of an unadjusted confounder will also show a gradient. Further, some causal relationships show single jumps (thresholds).
3. **Temporal association**
   This is a *sine qua non* for causality. Requires that for a factor to be considered causal, it must precede the disease. This is irrefutable; however, it does not provide evidence for or against causation when the cause is present prior to the onset of the disorder.
4. **Consistency of findings**
   Is the repeated observation of an association in different populations under different circumstances? However, lack of consistency does not rule out a causal association because some risk factors only produce effects under certain situations where a complement of causes is present (or has been present). Consistency can only really be assessed after the relevant details of the causal mechanism are understood.
5. **Specificity of effect**
   This is still emphasized in psychiatric epidemiology, perhaps due to the continual nosologic debates regarding the classification of disorders. However, there are no good grounds to refute a causal interpretation of an exposure because it is related to multiple effects or disease outcomes.
6. **Biologic plausibility**
   Plausibility is important; however, biologically plausible hypotheses can be constructed that are later found to be erroneous, and vice versa, causal associations can be dismissed due to current knowledge or belief systems.
7. **Coherence**
   Related to biologic plausibility and, therefore, similar problems, the

association should not conflict with what is known about the natural history and biology of the disease.

8. **Experimental evidence**
   Commentators are unclear about what is meant by experimental evidence, whether it is evidence coming from laboratory work or human experiments such as the removal of an exposure. However, this is a test for causality rather than a criterion.
9. **Analogy**
   Analogy may provide a source for more elaborate hypotheses about the associations but does not falsify a hypothesis.

Critique of the Causal Criterion

With the exception of temporality, none of the above situations can be viewed as necessary or sufficient criteria for determining whether an observed association is causal. Causal inference cannot be deduced with certainty.

Environmental Risk Factors for the Onset of Psychosis Spectrum Disorder

A number of environmental factors are statistically associated with an increased risk for onset of psychosis spectrum disorder. Evidence of association, however, does not equate to causality. Decisions regarding causation require deductive judgments. In order to aid this logic, causal criteria have been developed (Table 12.5–1). This section uses the most frequently applied criteria to structure the review of the current evidence for the risk factors for the onset of psychosis spectrum disorder. However, it is important to stress that these criteria should not be used in a mechanistic fashion, as all have exceptions and are useful at best as a framework for thought.

Table 12.5–2.
Commonalities in Patterns of Environmental Risk in Psychosis

The developmental window of exposure	The best-established risk factors impact the postnatal period of childhood–adolescence
Pleiotropy	The same risk factor impacts nonspecifically both psychotic and nonpsychotic disorder
Aetiologic heterogeneity	The same disorder is nonspecifically associated with a range of different risk factors
Operates across the severity spectrum	Environmental risks are associated with both attenuated psychotic symptoms and the clinical psychotic syndrome
Effects are additive	Effects of different environmental risk factors are independent and add to each other
More exposure is more symptom admixture and more severe psychopathology	The more severe the psychopathology (i.e., the more symptom domains involved), the greater the level of environmental exposure
Evidence of differential genetic sensitivity	There is emerging evidence that genetic factors make individuals differentially sensitive to the risk-increasing effect of environmental factors
Evidence of dose-response	Many risk factors have established evidence of a dose-response relationship in association with a psychotic disorder
Evidence of a degree of genetic confounding	There is some evidence that risk factors are, in part, driven by genetic factors. However, even so, they likely remain causally associated with a psychotic disorder.

Although there is a wide range of different risk factors, they have a number of commonalities that are highlighted below (Table 12.5–2).

Risk and Protective Factors Operating During Early Development

Early signals in relation to the later development of psychotic disorder have been studied extensively, and although there is some evidence for association, interpretation is difficult due to the long period between exposure and outcome and the many possible confounders, particularly genetic confounding.

A recent systematic review covering 152 studies and 98 risk or protective factors estimated that the following early exposures may be associated with psychotic disorder: maternal age younger than 20 years (OR 1.17) and 30 to 34 years (OR 1.05); paternal age younger than 20 years (OR 1.31) and older than 35 years (OR 1.28); any maternal (OR 4.60) or paternal (OR 2.73) psychopathology; maternal psychosis (OR 7.61) and affective disorder (OR 2.26); three or more pregnancies (OR 1.30); herpes simplex 2 (OR 1.35); maternal infections not otherwise specified (NOS; OR 1.27); suboptimal number of antenatal visits (OR 1.83); winter (OR 1.05) and winter to spring (OR 1.05) season of birth in the northern hemisphere; maternal stress (NOS; OR 2.40); famine (OR 1.61); any famine or nutritional deficits in pregnancy (OR 1.40); maternal hypertension (OR 1.40); hypoxia (OR 1.63); ruptured (OR 1.86) and premature rupture (OR 2.29) of membranes; polyhydramnios (OR 3.05); definite obstetric complications (NOS; OR 1.83); birth weights of less than 2,000 g (OR 1.84), less than 2,500 g (OR 1.53), or 2,500 to 2,999 g (OR 1.23); birth length less than 49 cm (OR 1.17); small for gestational age (OR 1.40); premature birth (OR 1.35), and congenital malformations (OR 2.35). Significant protective factors were maternal ages 20 to 24 years (OR 0.93) and 25 to 29 years (OR 0.92), nulliparity (OR 0.91), and birth weights 3,500 to 3,999 g (OR 0.90) or more than 4,000 g (OR 0.86).

Although the results were corrected for publication biases and sensitivity and metaregression analyses confirmed the robustness of these findings for most factors, interpretation is difficult. It is likely that at least some of the factors anticipate unmeasured environmental risks associated with other, later risk factors such as childhood adversity, tobacco smoking, minority group, urban environment, and drug use. The factors may also generally reflect current—and later—severe social adversity and inequality, contributing to high levels of allostatic load and increased risk for psychotic disorder. It is also likely that some of the reported risks are mediated by genetic risk for psychotic disorder. For example, maternal mental ill health may reflect psychotic disorder genetic risk and have no direct causal impact on the risk of psychotic disorder in the offspring. Other factors may also apply. For example, a large Danish study suggested that the association with higher paternal age may be due simply to selection into late fatherhood accompanying a predisposition to psychosis spectrum disorder.

Winter birth may be a proxy indicator for some seasonally fluctuating environmental factors. The most popular hypotheses relate to seasonal variation in exposure to intrauterine viral infections around the time of birth or variation in light, temperature/weather, or external toxins. For a dose-response relationship, the effect size increases with latitude (although there are fewer studies from the lower latitudes of the northern hemisphere in the pooled analyses) until about 50 degrees latitude, where it then starts to fall, suggesting a nonlinear inverted U-shaped association, which may in part be explained by methodologic heterogeneity across studies. For temporal association, exposure clearly predates the diagnosis of psychosis spectrum disorder. Studies from northern hemisphere countries consistently show excess winter births in people with psychosis spectrum disorder; however, the effect seems to be greater for those born in urban as compared to rural settings. A meta-analysis of studies carried out in the southern hemisphere did not find an effect for the season of birth (pooled odds ratio 1.04; 95% CI, 0.99 to 1.08). Studies examining season of birth association with other psychiatric conditions are small, and their findings are inconsistent. The largest and most robust study carried out in Denmark, did not find an association between the season of birth and the development of a bipolar disorder, suggesting some specificity of effect for schizophrenia (spectrum) disorder as compared with bipolar disorder. Few studies have looked at the age of onset in relation to the seasonality of birth in people with a psychotic spectrum disorder. The largest study to date found higher rates of winter births in patients who developed psychosis spectrum disorder later in life; however, most studies are underpowered to investigate this issue.

The listed pregnancy and birth complications cover a broad range of pre- and perinatal events, but even a large systematic review of prospective population-based studies (period 1995 to 2001) was unable to identify one individual complication or unifying mechanism. Taken together, findings may implicate an increased risk of hypoxia.

Studies to date have mostly been underpowered to uncover variation in effect size (odds ratio) across groups ordered by degree of exposure to pregnancy and birth complications. One population-based cohort study from Helsinki used continuous measures and found an effective gradient of increasing risk with decreasing birth weight, length at birth, and placental weight (controlled for family history of psychosis); and the National Collaborative Perinatal Project found risk increased when multiple hypoxia-associated complications were present. However, it is difficult to interpret these gradients while the underlying mechanism remains elusive. Exposure clearly predates diagnosis of psychosis spectrum disorder, although fetal growth retardation may be an early indicator of aberrant neurodevelopment. The associations for individual obstetric complications with the later development of psychosis spectrum disorder are mostly negative and inconsistent; however, statistically pooled estimations suggest a modest effect with the broad category of pregnancy and birth complications.

Pregnancy, birth, and neonatal complications do not act independently of one another; however, current studies are underpowered to detect interaction effects. There have been only a few small studies examining the differential effects of pregnancy and birth complications and symptom dimensions or subtypes of psychotic disorder. A recent systematic review of pregnancy and birth complications and bipolar disorder estimated the pooled odds ratio for exposure to pregnancy and birth complications and later development of bipolar disorder to be 1.01 (95% CI, 0.76 to 1.35) compared with controls and 0.61 (95% CI, 0.39 to 0.95) compared with people who later developed schizophrenia. Furthermore, a Danish National Register study did not find fetal growth indicators to be associated with the development of bipolar disorder. This suggests some specificity for schizophrenia (spectrum) disorder among psychiatric conditions. Indirect evidence of gene–environment interactions comes from a population-based cohort study in Helsinki that found that a history of fetal hypoxia was associated with increased structural brain abnormalities (reduced gray matter and increased ventricular size) among patients with psychotic disorder and their nonaffected siblings but not among controls at lower genetic risk for the disorder. Moreover, the effect size was greater for these patients compared to the unaffected siblings, suggesting that while they show some sensitivity to the psychosis-promoting effects of hypoxia, there is a quantitative difference in this sensitivity. Evidence of gene–environment correlation from a meta-analytic study of pregnancy and birth complications in children born to parents with psychotic disorder suggests individuals with psychotic disorder have an increased risk of obstetric complications, low birth weight, and poor neonatal condition. The risk is greater for mothers with psychotic disorder, but is not confined to mothers with onset prior to delivery or to the births of the children who develop psychosis spectrum disorder themselves later in life. These findings possibly reflect a gene-environment correlation whereby the genetic liability for psychosis in the parent increases the likelihood of social adversity that is associated with obstetric complications.

RISK FACTORS OPERATING DURING CHILDHOOD AND ADOLESCENCE

Urban Birth and Upbringing

Numerous studies to date have investigated variations in the incidence of psychotic disorders in terms of place. The latest and largest international incidence study from the EU-GEI consortium thoroughly documented marked geographic variation in the incidence of psychosis spectrum disorder across 17 catchment areas and six countries. Over 20 studies have demonstrated an association between urban living and rates of psychosis spectrum disorder in countries from the Global North. Generally, studies show a twofold increase in the risk of psychosis spectrum disorder in urban as compared to rural settings, as evidenced by a recent meta-analysis. Further, the weight of evidence supports a risk gradient, with unidirectional increasing rates of psychosis spectrum disorder with increasing levels of urbanicity. There is compelling evidence suggesting that elevated rates of psychosis spectrum disorder in urban areas are not due to social drift during the prodrome (reverse causation). There is some evidence; however, that part of the association between urban environment and psychosis spectrum disorder may be due to genetic confounding (genetic variation predisposing to psychosis spectrum disorder also influencing urban residence). It is urban exposure prior to the onset of the disorder that seems to be more important, rather than the level of exposure at the time of illness onset. A case register study from the Netherlands demonstrated an increased risk of psychosis spectrum disorder for people born in urban areas, irrespective of where they were living at the time of presentation, while those living in urban areas were only at increased risk if they had also been born within an urban area. Furthermore, a Danish study measured the effect of cumulative urban exposure on the risk of psychosis spectrum disorder and found the first 15 years of life were the most critical, as well as evidence of a dose-response relationship. Analyses from the same cohort further suggest that people conceived (including the period of gestational development) in urban areas but reared in less urban areas may be at higher risk of psychosis spectrum disorder. This phenomenon of increased risk seen when there is a mismatch between in-utero developmental environments versus the environment in adult life is recognized in other chronic complex diseases, such as cardiovascular disease. The findings are consistent with similar effect sizes seen across studies from different geographical areas. The effect is not confounded by other known risks, such as pregnancy and birth complications, and is present after controlling for sociodemographic variables. Most analyses have examined the association of urban living within the broad psychosis category rather than specific diagnoses. Studies that have compared across diagnostic categories have found the risk to be relatively specific for schizophrenia (spectrum) disorders compared to bipolar disorder. Studies investigating the family history and urban exposure support a possible interaction with genetic liability. For example, in the Danish cohort, those *without* a family history of psychosis spectrum disorder had a much smaller risk of psychosis spectrum disorder (0.054% increased risk per unit change of urbanicity) compared to those with a family history (0.22% increased risk, per unit change of urbanicity). New evidence shows that urbanicity effects may be context-specific, as indicated by absent or reversed effects for psychosis spectrum disorder in southern European countries. Similarly, more recent evidence on the prevalence of attenuated psychotic symptoms from the World Health Survey (WHS) conducted in countries from the Global South does not show a consistent pattern of higher prevalence estimates in urban areas. Therefore, urban-rural differences may vary in pattern across the globe and it would be ill-advised to assume city living has the same meaning worldwide. The contrasting findings highlight the strong need for adopting a global perspective in investigating geographic variation in the incidence and prevalence of psychosis spectrum disorder, such as in the INTREPID (“India, Nigeria, Trinidad: Researching Psychosis in Diverse Settings”) research program.

Could group-level social processes account for the urban upbringing effect? Growing up in an urban environment in a country of the Global North is a proxy environmental risk factor. Recent research has started to examine the social and economic structure of neighborhoods as a possible explanation for this effect. The social environment varies widely and systematically across dimensions of material deprivation, neighborhood organization (social fragmentation), and ethnic composition. Studies have consistently found a relationship (not necessarily linear) between material deprivation and risk of psychosis spectrum disorder, although study design has generally been cross-sectional, precluding assessment of the direction of the relationship, and could be accounted for by people drifting into more deprived areas before presenting to clinical services. Ecologic studies have consistently shown area-level social fragmentation to influence strongly the rates of psychosis spectrum disorder. A small number of recent studies have examined levels of social cohesion and found neighborhoods richly endowed with stocks of social capital (interpersonal trust, norms of reciprocity, mutual aid) to have reduced rates of psychosis independent of material deprivation. Ethnic density is an area-level characteristic, defined as the percentage of composition (in a particular geographical area) of people with a given ethnicity, and it has been found to be protective. That is, the higher the area’s ethnic density, the lower the rates of psychosis spectrum disorder for the ethnic group under study, independent of area measures of deprivation, a finding that has been consistently reported in a number of studies. Taken together, these reports suggest a possible neighborhood-level social process influencing the risk of psychosis spectrum disorder and emphasize the important role of social context in pathways to psychosis, which will require further elaboration and replication.

Minority Position and Migration

Many studies have reported increased rates of affective and nonaffective psychosis in migrants, especially among the second generation born in the new homeland. Increased risk is found among black Caribbean and Black African immigrants to the United Kingdom; Surinamese, Dutch Antillean, and Moroccan immigrants to the Netherlands; Australian, African, and Greenland immigrants to Denmark; and Greek immigrants to Belgium. Meta-analyses of incidence studies found a higher risk of psychosis spectrum disorder in almost all immigrant groups compared with the majority population. Although the pooled effect sizes suggest a threefold increase in risk, there was evidence of systematic heterogeneity across studies that could not be accounted for by variation in study design. Analyzing first- and second-generation migrants separately results in a pooled relative risk of around 2.5 for the first generation and a higher pooled relative risk for the second generation. However, there is no conclusive evidence that the risk in second-generation migrants is higher than in first-generation migrants. There is also a higher risk of psychosis spectrum disorder in migrants from developing countries; however, the most conspicuous finding was the increased risk associated with being Black, with a pooled relative risk of around 5. Migrant status is a binary exposure; one is either a migrant or one is not. Therefore, a dose-response association will only be testable after the underlying mechanism operating to increase the risk is established.

Meta-analysis of suitable studies has suggested that migration prior to age 18 is associated with a greater risk of psychotic disorder compared to migration in early adulthood.

The selective migration of people at increased risk of psychosis is possible (reverse causation). However, this does not explain the increased risk apparent in groups who migrated *en masse*, such as the Surinamese mass migration to the Netherlands. Further, it cannot explain the increased risk found in second-generation immigrants. Increased risk of psychosis spectrum disorder in migrant groups is a replicated and consistent finding. The validity of this finding has been intensely debated, however. Increased rates do not appear to be explained by psychiatrists diagnosing psychosis spectrum disorder differently in migrant groups, differential service use, or under enumeration, which may have hampered some of the earlier studies. Poor obstetric care in migrant mothers has also been suggested as an explanation, but there is no evidence to support this view. While there has been only a limited amount of research on candidate explanatory factors of the high rates of psychosis spectrum disorder in migrant and ethnic minority groups, the evidence reported to date by investigations such as the etiology and ethnicity in psychosis and other psychoses (AESOP) study; a large epidemiologic incidence and case-control study of first episode psychosis conducted in the United Kingdom suggests that it is a broad range of social adverse experiences in both childhood and adulthood that may combine and interact in complex ways to increase risk in some groups. Finally, studies mostly from the United Kingdom have found the risk of psychosis spectrum disorder to vary across different ethnic groups. The Black Caribbean and Black African groups in the United Kingdom are at higher risk than the Asian group, as are the Surinamese and Moroccan groups compared to the Turkish group in the Netherlands. The association of effect is not specific to schizophrenia, with rates of nonaffective and bipolar disorder also higher in migrants. For instance, the AESOP study found a sevenfold increase in mania among the Black Caribbean and Black African groups compared to the majority group and a twofold increase in mania among other migrant groups. There is no difference in the age of onset in migrant groups as compared to the majority population.

Use of Cannabis, Tobacco, and Other Drugs

There is little doubt that cannabis intoxication is associated with transient psychotic symptoms in some individuals. Furthermore, case-control studies have consistently demonstrated higher rates of cannabis use among patients with psychosis spectrum disorder as compared with controls. Prospective studies additionally have found higher rates of psychosis spectrum disorder in users of cannabis. The pooled analysis of the association between cannabis and psychosis indicates approximately a fourfold increase in risk for the heaviest users and a twofold increase for the average cannabis user in comparison to nonusers.

Studies examining cannabis use and the subsequent development of other psychiatric conditions are small and their findings are inconsistent, but they may indicate an impact on anxiety, depression, and mania. The association between prior cannabis use and subsequent psychosis spectrum disorder is a consistent finding across studies. Adjusting for other drug abuse, known risk factors of psychosis spectrum disorder, and sociodemographic variables attenuates the effect, but it remains statistically significant. A family study of adolescents with acute psychosis found a tenfold increase in the morbidity risk for psychosis in relatives of probands who tested positive for cannabis use compared to the relatives of probands who tested negative for cannabis, suggesting gene-environment interaction.

Observational studies on the association between cannabis use and psychosis outcomes commonly suffer from so-called omitted variable bias. This refers to the possibility that the observed link is due to any of a range of plausible confounders beyond simple demographics and direct confounders that can never be measured precisely in combination such as genetic vulnerabilities, cognitive abilities, fetal exposures, early development, and parental rearing practices. One way of comprehensively addressing omitted variable bias is to study within-person associations of cannabis and psychosis outcomes using panel data. Panel data (the same subjects measured at two or more points in time) consist of repeated measures that are nested within individuals. The repeated measures are referred to as the within-person level or level 1 units, whereas the persons are referred to as the between-person level or level 2 units. Studies on the association between cannabis use and psychosis over time are usually analyzed at the between-person level or the mixed within-person and between-person level. However, the relationship between cannabis use and psychosis may well differ across levels (i.e., between-person versus within-person levels). A classic example involves the relationship between physical exercise and heart rate: the within-person analysis will reveal a positive association in that more strenuous exercise leads to a higher heart rate. However, in contrast, this association is likely to be reversed at the between-person level, as people who exercise a lot, on average, tend to have a lower heart rate due to their overall better fitness. Panel data of individuals measured repeatedly over time allow control for fixed characteristics, including full nongenetic and genetic risk, whether they are measured/measurable or not. Three studies applied this within-person paradigm to cannabis and clinical and attenuated psychosis outcomes. All three found strong evidence for an association between cannabis and psychosis, and one additionally found no evidence for the self-medication hypothesis.

Another plausible epidemiologic paradigm to examine the causality of the association between cannabis and psychosis is to demonstrate an association between the level of use in the general population and the rate of psychotic disorder in that population. Using this paradigm, researchers in the EU-GEI consortium showed differences in the frequency of daily cannabis use and in the use of high-potency cannabis contributed to striking variation in the incidence of psychotic disorder across 11 studied sites in six countries.

Cannabis use is reliably associated with an earlier age of onset. Research suggests that the association between cannabis use and earlier onset of psychosis spectrum disorder is robust and is not the result either of tobacco smoking by cannabis-using patients or the other potentially confounding factors. This supports the hypothesis that, in some patients, cannabis use plays a causal role in the development of psychotic disorder.

Use of Tobacco and Other Drugs: Accumulating evidence suggests that tobacco use may be associated with schizophrenia, with findings from Mendelian randomization studies suggesting a possible causal link. A major challenge of studying tobacco use is confounding other smoking-based drugs such as cannabis. Tobacco smoking, unlike cannabis use, is not associated with differences in age at onset.

It is likely that stimulants as a class contribute to the overall burden of psychosis. The available evidence indicates a consistent relationship between methamphetamine use and psychosis, as well as depression and an increased risk of mortality due to suicide and overdose. Both lifetime and current cocaine users show a high prevalence of cocaine-induced psychosis, the prevalence of cocaine-induced psychotic symptoms being higher than the prevalence of the cocaine-induced psychotic disorder.

STRESSFUL LIFE EVENTS AND EARLY CHILDHOOD ADVERSITY

Many studies report an excess of stressful life events in the few weeks prior to the onset of psychotic and affective disorders. A recent systematic review and meta-analysis of studies published during the period 1968 to 2012 suggest a more than threefold increased risk of psychosis spectrum disorder in those exposed to adult life events. However, the authors noted several methodologic limitations of included studies, such as small sample sizes and the use of checklist measures. Although these life events may precipitate the disorder, there is some evidence to suggest they reflect a concatenation of stressful life events over a more sustained period before the onset of psychosis spectrum disorder. However, it is difficult to establish whether increased stress is a cause or a consequence of the development of the psychotic disorder. This is difficult to disentangle with traditional risk factor methodologies; gene-environmental interaction designs may be a more fruitful way of exploring the effect of stress.

Early childhood adversity studies describe a range of clustered adverse experiences, including sexual, physical, and emotional abuse and neglect. There are now a modest number of population-based studies that were recently combined into a meta-analysis (covering the period from 1980 to 2012), suggesting that the risk of attenuated psychotic symptoms and psychotic disorder is increased almost threefold in those exposed to early childhood adversity. There is evidence that childhood adversity-related psychosis is associated with more severe psychopathology as indicated by higher levels of dimensional admixture of different symptom domains. There is also some evidence that childhood adversity is associated with cognitive impairment and, therefore, with some of the childhood neurodevelopmental factors associated with an adult psychotic disorder. A meta-analysis of possible mediating psychological mechanisms found mediating roles of dissociation, emotional dysregulation, and post-traumatic stress disorder symptoms (avoidance, numbing, and hyperarousal) between developmental trauma and hallucinations.

Gene-Environment Interaction

Traditional epidemiologic risk factor methodologies measure the linear association of environmental exposures to risk for the onset and persistence of psychosis. Such studies demonstrate that people respond differently to the same environmental exposure. Some individuals seem to have a low risk of developing the disorder as a result of environmental exposure, while others are more vulnerable; in some circumstances, persons with increased liability for a disorder seem to have a higher exposure frequency to noxious environmental agents, as compared with the general population. Variation in genotype may, in part, account for this heterogeneity. There are two most often studied gene-environment relationships. *Gene-environment* *correlation* occurs when a genotype could increase the likelihood of a person becoming exposed to an environmental risk factor. For example, biologic parents not only determine a child's genotype but also provide their early rearing environment; or a child's temperament (at least partially genetically determined) influences how they select and/or shapes their environment. Some studies have suggested, for example, that the association between urban environment and psychosis may be due, at least in part, to gene-environment correlation. However, gene-environment correlation does not preclude the environmental factor from still having a causal impact. For example, in the Environmental Risk (E-Risk) Longitudinal Twin Study, a nationally-representative cohort of 2,232 British twins born in 1994 to 1995 and followed to age 18 (93% retention), socioenvironmental risks (urbanicity, air pollution, neighborhood deprivation, neighborhood crime, neighborhood disorder, social cohesion, residential mobility, family poverty) were associated with molecular genetic PRS for schizophrenia and depression, but the associations between socioenvironmental risk and adolescent attenuated psychotic symptoms existed above and beyond this gene-environment correlation. *Gene-environment interaction (synergism)* (G × E) posits that instead of inheriting a disease state per se, a person actually inherits a set of genetic susceptibility factors that act to moderate the effects of (causal) environmental exposures. If this assumption is correct, integrated epidemiologic approaches which examine the joint influence of genes and environment on the risk for psychosis spectrum disorder, will improve the accuracy and precision of the measured risk estimates and the predictive power of environmental risk factors.

G × E is notoriously difficult to study. However, with the advent of direct molecular measures of genetic risk, some progress has been made. Initially, studies modeled interactions between (functional) variants in a candidate gene and a measure of the environment. This resulted in some early results on, for example, the interaction between cannabis use and genetic variation in AKT1. However, these results likely represent false positives. A more robust approach involves the use of summary measures of polygenic risk and environmental risk. Thus, the utilization of PRS as a single metric of molecular genetic risk has considerably increased the power to detect associations with phenotypes as well as gene-environment interactions. Currently, PRS for schizophrenia of an individual can be estimated by summing the log odds ratios of individual single nucleotide polymorphisms multiplied by the number of risk alleles present at the corresponding loci. Similarly, recent work on the exposome paradigm has shown that it is possible to investigate the sum of nongenetic “risk” and apply predictive modeling approaches to construct an exposome score (ES; an aggregated score of nongenetic exposures) for schizophrenia. Evidence was found in the EU-GEI consortium and replicated in another study for additive interaction of molecular genetic risk state for schizophrenia with ES and, separately, the presence of lifetime regular cannabis use and exposure to early-life adversities. The results suggest that the etiopathogenesis of schizophrenia involves genetic underpinnings that act by making individuals more sensitive to the effects of some environmental exposures.

PREMORBID VULNERABILITY INDICATORS/MARKERS

Premorbid Indicators: Early Developmental Abnormalities

As a group, individuals who develop psychosis spectrum disorder as adults are more likely to manifest subtle motor/language and cognitive abnormalities during childhood and adolescence compared with individuals who do not develop psychosis spectrum disorder. There does not appear to be any association between developmental abnormalities and prenatal insults, suggesting they may be indicators of genetic liability. A series of large birth cohort studies have consistently found evidence for delays in the attainment of developmental milestones for children who later develop psychosis spectrum disorder. A meta-analysis focusing on motor milestones found that walking unsupported, standing unsupported, and sitting unsupported were significantly associated with adult schizophrenia risk. Abnormalities in motor milestones are most conspicuous before the age of 2, while receptive language abnormalities are most prominent between the ages of 5 and 15. The relationship between the age of acquiring motor/language milestones and the risk of psychosis spectrum disorder appears to be linear, making a possible developmental subgroup unlikely. Furthermore, severe motor/language abnormalities tend to predict an earlier age of onset for psychosis spectrum disorder. A study from Philadelphia found increased rates of motor and language abnormalities, not only in the children who later developed psychosis spectrum disorder but also in their siblings who remained unaffected, suggesting manifest motor abnormalities may be markers of underlying genetic liability. The early emerging motor and language deficits appear to be relatively specific for psychosis spectrum disorder. As a group, children who later develop psychosis spectrum disorder score lower on standardized measures of intelligence compared to their unaffected counterparts throughout childhood. However, the difference widens around the ages of about 4 to 7. Studies that compared affected and unaffected siblings found that the children who later developed psychosis spectrum disorder had a lower intelligence quotient compared to their unaffected siblings. This deviance in cognitive impairment seems to be relatively specific to schizophrenia (spectrum) disorder and is not present in bipolar disorder. Twin and family studies have shown that the alterations in adult cognitive performance seen in patients and their first-degree relatives are markers for schizophrenia genetic liability. It is likely that the developmental cognitive alterations similarly reflect the influence of genetic factors associated with schizophrenia.

As a group, the children who later develop psychosis spectrum disorder are more likely to show signs of subtle social maladjustment, for example, poor relationships with other children, social isolation, lone play, and social anxiety as compared to their peers. However, these subtle social markers are common among children who later develop a broad range of different psychiatric disorders. Although developmental abnormalities are common in children who later develop psychosis spectrum disorder, many of them have normal developmental histories, which results in low sensitivity (the ability to predict children who will later develop psychosis spectrum disorder correctly). Furthermore, they are relatively prevalent in the general population (estimated at around 15%), which consequently reduces their specificity (the ability to correctly predict children who will not develop psychosis spectrum disorder) and, in turn, makes them poor predictors of future psychosis spectrum disorder (positive predictive power estimates range from 3% to 5%). Developmental abnormalities may be most useful as a possible measurable indicator of genetic liability for psychosis spectrum disorder.

Premorbid Indicators: Psychopathology

Childhood psychiatric disorders: Longitudinal research from Israel and Denmark shows that a diagnosis of psychotic disorder is commonly preceded, sometimes by decades, by almost any other mental disorder suggesting the pathway from early psychopathology to schizophrenia is heterotypic. This was confirmed by the Dunedin Multidisciplinary Health and Development study, a longitudinal follow-up of a representative birth cohort of 1,037 children born in Dunedin, New Zealand, during the year 1972 to 1973. Its strength lies in the fact that it has sequentially assessed not only childhood developmental indices but also serially collected self-reported psychopathology (ages 11, 13, 15, 18, and 21 years), with further psychiatric interviews carried out at ages 11 and 26 years. The study found that generally, there was child-to-adulthood syndrome homotypic continuity found for anxiety, depression, and antisocial disorders. However, no such syndrome continuity was evident for adult schizophreniform disorder, which was heterotypically preceded by a range of different childhood disorders, including anxiety, depression, conduct/oppositional disorder, and attention-deficit hyperactivity disorder (ADHD). This group of juvenile diagnoses did not specifically predict adult psychotic disorder but were, in fact, the harbinger of many different adulthood psychiatric illnesses. Such findings clearly demonstrate significant heterotypic psychopathology prior to adult-onset psychosis spectrum disorder. However, the juvenile categorical diagnoses are demonstrably poor discriminators for what adult psychiatric disorder a child is likely to develop in the future.

Attenuated psychotic symptoms: Many studies have reported on the homotypic association between childhood attenuated psychotic symptoms and later psychotic disorder. For example, in the Dunedin Multidisciplinary Health and Development study, attenuated psychotic symptoms at age 11 were associated with increased rates of a psychometrically defined psychotic disorder outcome by age 38. More severe childhood psychotic symptoms carried higher risk than less severe expression of psychosis. Replication of this transition over time from attenuated psychotic symptoms to full-blown psychotic disorder, with evidence for dose-response, comes from longitudinal studies in the Netherlands (NEMESIS and NEMESIS-2), Germany (EDSP), Turkey (TURK-SCH), and the United Kingdom (ALSPAC). A limitation of these studies is the use of a nonclinical, variably, and rather arbitrary defined psychometric definition of psychotic disorder. The exception was a longitudinal study in Israel, which relied on the outcome of a registry-based clinical diagnosis of psychotic disorder requiring hospital admission. A stratified full probability sample of 4,914 persons aged 25 to 34 years who were screened for psychopathology in the 1980s. Subsequent psychiatric hospitalization was ascertained using the psychiatric hospitalization registry, with a mean follow-up of 24 years. Early attenuated psychotic symptoms predicted risk of later hospitalization for nonaffective psychotic disorders, mostly during the 5 years after baseline (adjusted odds ratio = 4.31; 95% CI, 2.21 to 8.41; positive-predictive value = 1.27%; population attributable risk fraction = 33%). Also, attenuated psychotic symptoms increased the risk of later hospitalization for other psychiatric disorders (adjusted odds ratio = 2.21; 95% CI, 1.02 to 4.82). Early negative symptoms (five interviewer-rated items assessing speech deficits, flat affect, and poor hygiene) were also associated with an increased risk of later schizophrenia only in the presence of attenuated psychotic symptoms (OR = 13.0, 9% CI, 2.1 to 79.4).

HIGH-RISK MODELS OF PSYCHOSIS ONSET

Strong heterotypic associations between prior psychopathology and later psychotic disorder outcomes may indicate a possibility for prevention. As a result, there is interest in advancing research on the predictive modeling of psychosis, similar to other areas of medicine such as cardiovascular medicine, oncology, and dementia care. Finding ways to understand and predict important health transitions and the factors causally impacting these represent a growing part of health expenditure at the interface of clinical medicine and public health.

Table 12.5–3.
High-Risk Strategies to Study the Onset of Psychosis and Develop Early Intervention

	Clinical high-risk studies	Population-based cohort studies of attenuated psychotic symptoms	Follow-back samples of patients with a psychotic disorder
Sampling	Opportunistic samples through selected searches and advertising, sometimes part of special “prodromal service”	Representative population-based samples	Representative samples of patients
Sample risk enrichment strategy	Yes	No	No
Risk predictors used	Almost entirely based on the construct of attenuated psychotic symptoms, based on clinical interview	Attenuated psychotic symptoms and other measures of psychopathology based on self-report or clinical interview	Measures of psychopathology elicited through follow-back clinical interview
Role prediction nonpsychotic psychopathology	Exclusion variable (attenuated psychotic symptoms are not “better explained” by another DSM disorder)	Included as an essential mechanistic factor of psychosis onset	Included as an essential mechanistic factor of psychosis onset

Prediction paradigm	Homotopic (psychosis → psychosis)	Homotypic and heterotypic (nonpsychotic → psychosis)	Homotopic and heterotypic
Baseline mental health status	The great majority are already ill with anxiety disorder, depression, substance use disorder	Illness rate representative of the general population	All diagnosed with a psychotic disorder
Psychotic disorder outcome	Arbitrary psychometric, typically based on an increase in a continuous positive psychotic symptom scale	Mostly arbitrary psychometric, typically based on composite measures of psychosis and functioning, need for care or care use	Natural clinical psychotic disorder
Transition rates	Less than 10% per year	Less than 1% per year	100% (per definition)
The main factor driving a transition	Sample risk enrichment strategy	Risk over time	Risk over time
Length of follow-up	Mostly 1–3 years	Mostly 5–20 years	NA (follow-back)

Psychosis research has been slow to develop strong links with public health approaches. One reason is that the incidence of psychosis spectrum disorder is low and the latency long. Much of the risk contributing to adult mental ill health is developmental. To conduct decades-long cohort studies to study the relationship between developmental risk and the onset of adult psychopathology is impractical and expensive. Although some impressive work has been carried out in, for example, the U.K. MRC birth cohort studies, the New Zealand Dunedin cohort study, the Israeli draft board studies, and the Scandinavian birth cohort studies, there is an urgent need for data on the clinical ontogenesis and prediction in the psychosis spectrum.

To this end, three types of approaches have been developed in recent years. These include the selected *clinical high-risk approach*, the representative *population-based attenuated symptom approach,* and the *clinical sample follow-back approach*, depicted in the table (Table 12.5–3).

The clinical high-risk approach can bypass the problem of low incidence and long latency inherent to population-based approaches to psychosis. It has been extensively piloted, resulting in a burgeoning literature on prognostic prediction in so-called “ultra–high-risk” or “clinical high-risk” samples. The results of these, however, cannot be translated to psychosis incidence in the general population—and clinical practice—as they are based on nonepidemiologic opportunity sampling of selected help-seeking individuals who majorly already have a nonpsychotic disorder (anxiety, depression, substance use), yielding sample-specific results that are unlikely to be either representative or generalizable. In addition, the outcome of “transition” in clinical high-risk studies is not clinical but psychometric, typically based on an arbitrary increase in positive psychotic symptoms on a 5- or 6-point scale. Also, the relatively high rate of transitions in these individuals (10% to 15% over a period of 2 years) largely arises as a function of risk enrichment strategies embedded in specific sampling procedures, further limiting their use as a model of the onset of psychosis in the general population. Finally, risk and transition are modeled as a homotypic process (attenuated psychosis gives rise to psychotic disorder), whereas population-based epidemiologic evidence discussed above indicates that more clinical psychosis is attributable heterotypically (i.e., nonpsychotic symptoms give rise to psychotic disorder) to prior affective and other nonpsychotic dysregulation than to prior subthreshold psychosis.

Another way to study the development of psychotic disorder, developed in parallel to the development of the “clinical high-risk” or “ultra–high-risk approach,” is the prospective population-based cohort approach, focusing on psychopathology risk factors for mental health outcomes, including psychotic disorder. The exposure used in these studies typically involves attenuated psychotic symptoms (also referred to as “psychotic experiences”) in combination with other nonpsychotic psychopathologic predictors. For example, the German EDSP and the Dutch NEMESIS-2 studies managed to follow representative cohorts of around respectively 3,000 and 6,600 individuals over a period of approximately 10 years, conducting four extensive structured interviews over that period. While these cohort studies yield interesting findings about the ontogenesis of psychosis (typically at much lower yearly transition rates than clinical high-risk/ultra–high-risk studies), early-risk states typically are described at the level of attenuated psychotic symptoms, which can be difficult to measure in general population settings. Some studies, however, like the Dutch NEMESIS-2 and German EDSP studies, used clinical interviews to assess attenuated psychotic symptoms. Nevertheless, although rare transitions to full psychotic disorder in these cohort studies can be used to study real population-based transitions, analyses are often hampered by low power and difficulties accurately pinpointing transitions, given the low frequency of interviews over the follow-up period.

Despite limitations inherent to the population-based approaches of psychosis prediction, a consistent finding is the importance of nonpsychotic psychopathology. The actual emergence of psychosis is not an isolated "onset" of a specific phenotype with homotypic progression, as in the "clinical high-risk" or "ultra–high-risk" paradigm, but rather is an indicator of increasing severity/complexity of an underlying heterotypic progression of mixed psychopathology, particularly affective dysregulation, giving rise to psychosis (Fig. 12.5–1). These findings echo the increasing awareness that genetic and environmental risk and the clinical manifestation of early psychopathology, including psychosis, to a large degree is "transdiagnostic." The broader, transpsychopathological, and heterotypic approaches have conceptual and statistical advantages over a more narrow focus on psychosis in clinical high-risk and population-based approaches.

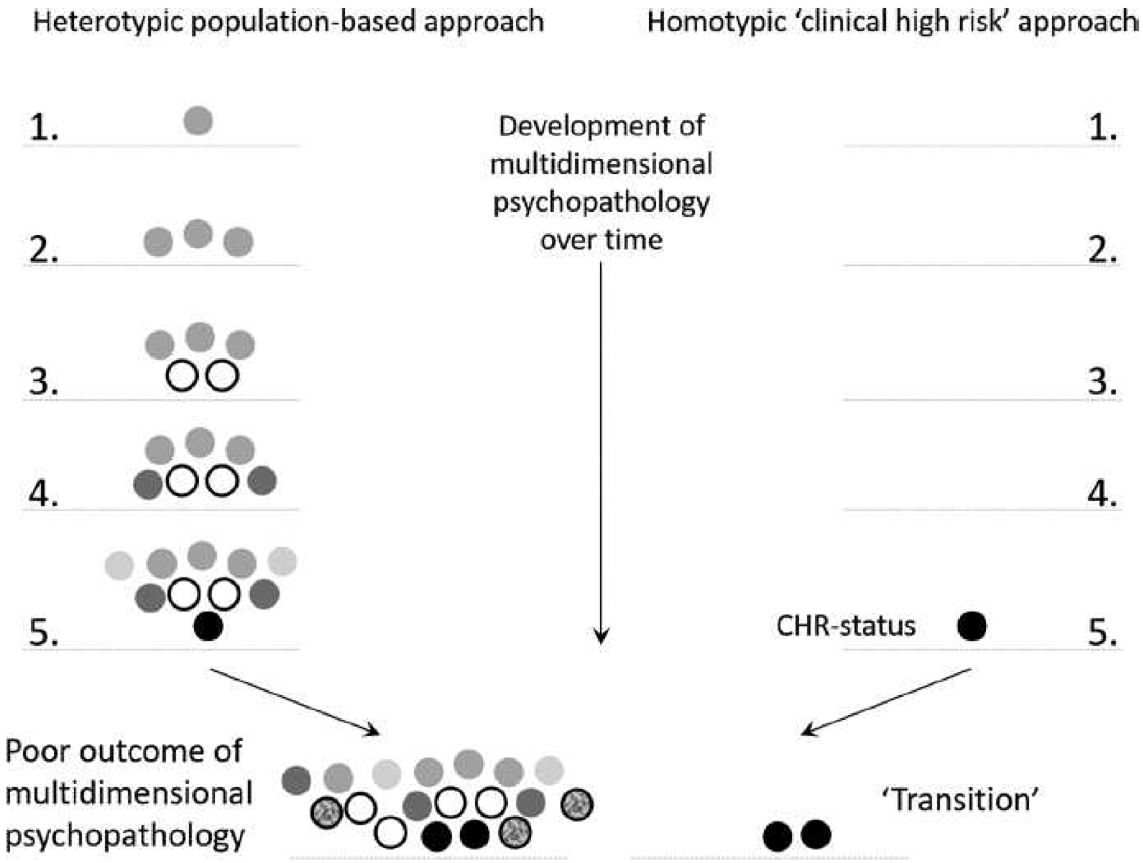
Heterotypic population-based approach
Homotypic 'clinical high risk' approach
1.
2.
3.
4.
5.
Development of multidimensional psychopathology over time
1.
2.
3.
4.
CHR-status
5.
Poor outcome of multidimensional psychopathology
'Transition'

FIGURE 12.5–1. Conceptual and practical contrast between the heterotypic and homotypic prediction of psychosis onset. The figure shows the relative "blindness" of the homotypic clinical high-risk (CHR)/transition paradigm for the heterotypic development of psychotic disorder. On the **left** is the natural development of multidimensional psychopathology over time (five phases) as established in recent prospective population-based cohort studies and retrospective follow-back studies. *Black circles* indicate (attenuated) positive psychotic symptoms. Other gray-scale circles indicate other psychopathology (e.g., anxiety, depression, obsessions, somatic concern). As the CHR paradigm ignores multidimensional nonpsychotic psychopathology, it remains "blind" and only "sees" homotypic psychotic phenomena as precursors of schizophrenia-"transition" (i.e., more severe psychosis; below on the **right**), while these phenomena are, in fact, a marker of relative poor outcome of heterotypic multidimensional psychopathology, diagnosable as a psychotic disorder in DSM-5 (below on the **left**). The restricted focus on positive symptoms in the CHR paradigm means that considerable potential for prevention in phases 1 to 4 is missed.

Recently, researchers have resurrected a time-honored method addressing the issue of studying psychosis onset and early intervention; the epidemiologic follow-back study. In a recent study in Canada, an intensive follow-back methodology was used to study the onset of psychosis in an epidemiologically representative incidence sample (n = 430), similar to the seminal work of Heinz Häfner in Germany, but modernizing it to fit the research framework of the current early intervention era. The authors found that individuals experiencing a first episode psychotic disorder could be traced back to the group with the most complex nonpsychotic syndromes, indexed as the number of preonset nonpsychotic symptoms identified. They established that (i) initial attenuated psychotic symptoms were preceded in time by nonpsychotic symptoms, (ii) nonpsychotic symptoms were more frequent and complex in those with attenuated psychotic symptoms, and (iii) higher levels of nonpsychotic symptoms were predictive of more subthreshold psychotic symptoms. The authors thus suggest that what is commonly considered the onset of a “psychotic disorder” may, in fact, represent an outcome of a prior nonpsychotic state.

These findings were confirmed in another prospective population-based cohort study, which showed that the risk in “psychosis high-risk” states was mediated not by attenuated psychosis per se but specifically by the combination of attenuated psychosis and nonpsychotic disorder. Psychosis high-risk research, therefore, may benefit from a reconceptualization from a focus on attenuated psychotic symptoms, with the exclusion of nonpsychotic DSM disorders providing a “better explanation” of the psychotic state, toward a focus on poor-outcome nonpsychotic disorders, characterized by a degree of psychosis admixture, on the pathway to psychotic disorder outcomes.

THE COURSE AND OUTCOME OF PSYCHOSIS SPECTRUM DISORDER

Prognostic Pessimism and Psychosis: Does the Evidence Stack Up?

Origins of prognostic pessimism: The pervading clinical and societal view of the diagnosis of schizophrenia is one of a uniformly “debilitating” and “deteriorating” disorder of poor outcome, the basis of which probably stems from clinically biased observations and a degree of conceptual confusion. For example, a carefully conducted Finnish study showed that the prevalence of the group of highly related (affective and nonaffective) psychotic disorders, that is, the psychosis spectrum, is as high as 3.5%, of which less than a third had a diagnosis of schizophrenia. As the different psychotic disorders share aetiologic factors, symptoms, and treatments, with no “point of rarity” separating them in these domains, schizophrenia may be considered the (relatively) poor outcome subgroup of a general broader psychosis spectrum syndrome. As schizophrenia is much more studied than other psychotic syndromes, the relatively poor outcome of schizophrenia has affected expectations of psychosis prognosis across the range of psychosis expression. The idea that a diagnosis of schizophrenia refers to a deteriorating disorder originates directly from the work of Kraepelin, who systematically investigated the course of psychoses in patients admitted to a large custodial institution in Germany during the late 19th century. He combined a number of clinical syndromes into a single disease entity (dementia praecox) because they appeared to share a poor longitudinal course (estimated range of permanent cure: 2.6% to 16.0%). However, the asylum provided long-term residential care for people with the most severe, complex, and debilitating conditions (i.e., a biased sample, skewed toward the worst outcomes—a mechanism known as “hospital bias” or “Berkson bias”), indicating that Kraepelin seminal description is unlikely to fully reflect the complete range of outcomes for the diagnosis of schizophrenia.

Table 12.5–4.
Methodologic Heterogeneity: Limiting Comparability between Studies of Course and Outcome of Psychosis

1. ***Variation in the populations from which patients are selected.*** Many studies have relied on in-patient samples only. However, approximately 20–40% (depending on the country and region) of patients with psychosis are never admitted to a hospital, and they may differ from those admitted in important variables associated with outcome.
2. ***Variation in diagnostic criteria.*** Lack of reliability between different classification systems may result in different samples being identified. Also, if broad concepts of psychosis are employed, the results tend to be more favorable than studies using narrow definitions.
3. ***Variation in the length of illness before entry into the follow-up study.*** If this varies, then subjects with an admixture of follow-up epochs and levels of chronicity are being examined.
4. ***Variation in length of follow-up*.**
5. ***Variability of attrition rates.*** In any follow-up study, a proportion of participants will be lost to follow-up due to death, migration, or refusal or are untraceable for other reasons. Many published studies have high rates of attrition (greater than 20%); further, if the analyses were done at a time when the researcher was unable to take into account this attrition (using imputation methods or other latent variable methods), loss to follow-up is likely to be related to particular patterns of course and outcome, which will bias the outcome.
6. ***Variation in methods used to assess course and outcome*.** There is no single measure of course and outcome for complex disorders, and qualitative summary terms (e.g., recovery, improvement, or deterioration) are likely to obfuscate intraindividual variability over time, and the terms are used inconsistently in the literature. The outcome is multidimensional (at least four domains should be considered, namely, symptom severity, functional impairment, disability in social and occupational roles, and existential recovery). There is very little work on the domain of existential recovery (recovery of perspective and the experience of a meaningful life despite problems) over time. The four domains are likely to be related, and few studies have used data reduction techniques to deal with the problems of multiple analyses. Further, few studies examine serial follow-up measures, and a single outcome measure will be less representative of a longitudinal course the longer the follow-up period is.
7. ***Variation in the characteristics of the general population*.** For example, social outcomes (e.g., unemployment) will be influenced by

the general unemployment rates in the area. Low and Middle Income Countries (LAMIC) countries may differ in the structure of mental health care and social care, as well as social cohesion and family structure, possibly impacting outcomes in many ways.

8. ***Variation in statistical techniques and adjustment of confounding.*** Statistical methods, which deal with multiple comparisons made at different time points, were not available or not utilized at the time of publication of the study. Further studies have often not had the resources to collect adequate data on potential confounding variables when assessing potential predictors (or prognostic indicators).
9. ***Variation in long-term management of psychosis spectrum disorder.*** Treatment may vary across time and place, which may differentially modify the natural history of the disorder. LAMIC countries may differ in structure of mental health care and social care, possibly impacting outcome in many ways.

Although most clinicians no longer work in long-term custodial settings, clinical caseloads remain Berkson-biased toward patients who require sustained contact with mental health services and have a continued need for care, rather than the representative sample of all persons ever diagnosed with psychosis spectrum disorder; therefore perpetuating the clinical illusion of the diagnosis of schizophrenia referring to an *invariably* chronic and deteriorating disorder. During the past century, many studies have attempted to examine the outcome of psychosis spectrum disorder yet, despite this effort, the data on longitudinal course remain weak, primarily due to the inherent and complex methodologic difficulties associated with carrying out follow-up studies, which are especially sensitive to confounding, bias and diagnostic selection (i.e., narrow schizophrenia vs. broad psychosis spectrum disorder) at baseline (Table 12.5–4). There is actually little information regarding the natural history of psychosis spectrum disorder in the literature, as, throughout history, psychosis has tended to result in custodial care or other medical, religious, or cultural interventions, which could potentially modify the longitudinal trajectory of the disorder. Since the 1950s, the vast majority of individuals diagnosed with psychosis have received antipsychotic treatment, so it is unlikely that contemporary studies actually reflect natural history. A study from India was able to examine outcomes in individuals who had never received treatment for psychotic disorder (a closer approximation to natural history, notwithstanding the potential influence of cultural context).

The course for the untreated group was heterogeneous and rather similar to the course of those who had received treatment.

Most descriptions of the longitudinal (symptom) pattern of psychosis have to date, been derived from systematic clinical observation (rather than statistical modeling techniques). Clinical researchers working during the early part of the 20th century followed patients with psychotic disorders over the long term and noted that the significant variability in the course of treatment could be usefully described by a number of course typologies (European researchers recognizing eight categories, while American observers described 12). These longitudinal patterns (slightly modified) have been incorporated into the current diagnostic classification systems as additional descriptors. The World Health Organization has described operationalized (symptom) descriptors for the course (Table 12.5–5) that are useful and reliable in clinical and research settings. The outcome is a multidimensional construct that, at minimum, requires a description of domains for (1) clinical (symptoms and treatment), (2) social (e.g., independent living, maintenance of social relationships, and employment), and (3) existential (meaningful life despite mental health problems). Such domains are likely, in part, to overlap. However, studies examining risk factors for the persistence of psychosis and disease-environment interactions should measure symptom course and social and existential domains independently, as the different domains probably do not covary directly but are more likely to be the result of different influences, at least in part (Tables 12.5–6 and 12.5–7).

Table 12.5–5.
World Health Organization Life Chart Description of Clinical Symptom Course

1. Single-psychotic episode followed by complete remission
2. Single-psychotic episode followed by incomplete remission
3. Two or more psychotic episodes with complete remission between episodes
4. Two or more episodes, with incomplete remission between episodes
5. Continuous (unremitting) chronic psychosis

Based on World Health Organization. *WHO Coordinated Multi-Centre Study on the Course and Outcome of Schizophrenia*. World Health Organization; 1992.

There is a small body of literature comparing the patterns and frequencies of outcomes using different perspectives, but few studies have examined multiple perspectives in the same sample. One such study carried out in South Verona's community care service, found a distinction between objective indicators and patient-reported outcomes. Patients were especially troubled by persistent unmet social, existential (help in the recovery of perspective), and functional needs. This may have important implications for the clinical management of people with psychosis, particularly since mental health services currently emphasize pharmacologic and psychological treatments while de-emphasizing social, existential, and functional interventions, yet social and existential unmet needs seem to influence disability over the longitudinal course.

Table 12.5–6.
Variation in Proportion of Good Outcome due to Variation in Outcome

Study	*N*	Period	Event Rate	Upper Limit	Lower Limit	Definition of Good Symptom Outcome
Muller, 1951 (Germany)	200	30	0.33	0.26	0.40	Recovered or substantially improved at the time of discharge (residual symptoms)
Bland, 1978 (Alberta)	43	14	0.21	0.09	0.33	Symptomatic recovery with no social or intellectual deficit through follow-up
Stephens, 1978 (Baltimore)	349	>5	0.24	0.20	0.29	"Recovered" included patients with symptoms but no social impairment
Ciompi, 1980 (Switzerland)	289	37	0.27	0.22	0.32	Remitted completely or had only mild residual symptoms
Salokangas, 1983 (Finland)	161	8	0.55	0.47	0.62	Complete recovery or occasional mild psychotic symptoms
Rabiner, 1986 (New York)	36	1	0.56	0.39	0.72	Symptomatic remission for at least 3 months prior to assessment
Sartorius, 1986 (Multinational)	1,352	2	0.39	0.36	0.42	One episode with no or minimal symptoms at follow-up
Shepherd, 1989 (United Kingdom)	107	5	0.22	0.15	0.30	Had no relapse during follow-up period
McCreadie, 1989 (Scotland)	49	2	0.37	0.23	0.50	Asymptomatic and functioning adequately at the time of follow-up
Marneros, 1992 (Germany)	355	25	0.07	0.04	0.10	Full remission after first episode
Thara, 1994 (India)	76	10	0.17	0.09	0.26	Complete recovery without relapse during follow-up period
an der Heiden, 1995 (Germany)	56	14	0.25	0.14	0.36	Full recovery over follow-up period
Mason, 1996 (United Kingdom)	58	13	0.52	0.39	0.65	Full remission in 2 years prior to outcome assessment
Wieselgren, 1996 (Sweden)	101	5	0.30	0.21	0.39	Dichotomized variable "good outcome" versus "poor outcome"
Ganev, 1998 (Bulgaria)	55	16	0.38	0.25	0.51	Complete remission in 2 years prior to follow-up assessment
Wiersma, 1998 (Netherlands)	82	15	0.27	0.17	0.36	Complete remission never readmitted
Vazquez–Barquero, 1999 (Spain)	76	10	0.32	0.21	0.42	Symptomatic recovery may have residual symptoms
Stirling, 2003 (United Kingdom)	49	10	0.47	0.33	0.61	Categorical measure Global Assessment of Functioning (GAF) score = 4 at the time of follow-up

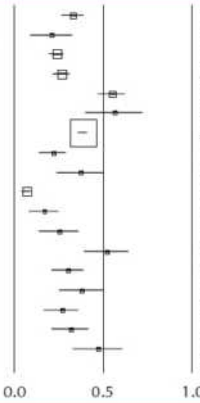

Broad Pattern of Course and Outcome

The first 5 to 10 years of psychosis spectrum disorder are considered "critical," referring to the, to date, nonconfirmed hypothesis that early energetic treatment may causally impact the later course of the syndrome. After the first episode, around 90% of patients will experience a rapid remission of symptoms. After 5 years, however, 80% will have experienced one or more relapses. With each episode, a small proportion of patients will develop a continuous illness course, displaying a mix of persistent positive symptoms, severe negative symptoms, cognitive difficulties, catatonia, and a state of demoralization and internalized stigma. After the first 10 years after onset, the illness course tends to plateau. However, cross-sectional outcome measures of psychopathology do not differ substantially according to study duration, suggesting that there is no clear pattern of deterioration or "progression," although this may occur in a subgroup of patients. A systematic review of prospectively designed studies examining the outcome of first episode psychosis (and other psychosis) published during the period 1966 to 2003 found good clinical outcomes in 42%, intermediate outcomes in 35%, and poor outcomes in 27% of the samples with a diagnosis of psychosis. In a recent 10-year follow-up of the AESOP first-episode psychosis cohort, 23% of cases did not experience a remission, 13% remitted and remained symptom-free, with an intermediate group of 44% having had at least one remission of 6 months or longer, and at least one episode of 6 months or longer. Moreover, using growth mixture modeling to identify latent trajectory classes that accounted for heterogeneity in the patterns of change in psychotic symptoms over time, four trajectories of psychotic symptoms were empirically identified in the AESOP study: (1) remitting-improving (58.5%); (2) late decline (5.6%); (3) late improvement (5.4%); and (4) persistent (30.6%). In addition, although recently, life expectancy in the general population has risen, there was further evidence from this study that the excess mortality, which has been repeatedly reported in people with psychosis, remains high. The absence of a uniform definition for outcome across studies makes direct comparison difficult. This frequency distribution is comparable with the findings from an earlier meta-analysis of long-term follow-up studies (published between the turn of the 20th century and 1998 with a follow-up period greater than 1 year). The evidence from recent systematic reviews is summarized in Table 12.5–8.

Table 12.5–7.
Variation in Proportion of Poor Outcome due to Variation in Outcome

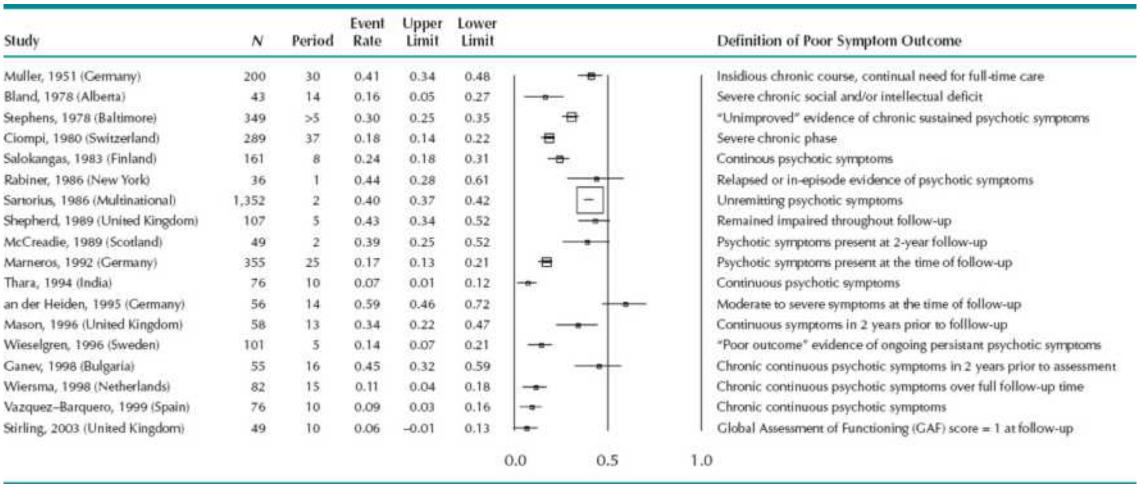

Study	*N*	Period	Event Rate	Upper Limit	Lower Limit	Definition of Poor Symptom Outcome
Muller, 1951 (Germany)	200	30	0.41	0.34	0.48	Insidious chronic course, continual need for full-time care
Bland, 1978 (Alberta)	43	14	0.16	0.05	0.27	Severe chronic social and/or intellectual deficit
Stephens, 1978 (Baltimore)	349	>5	0.30	0.25	0.35	"Unimproved" evidence of chronic sustained psychotic symptoms
Ciompi, 1980 (Switzerland)	289	37	0.18	0.14	0.22	Severe chronic phase
Salokangas, 1983 (Finland)	161	8	0.24	0.18	0.31	Continous psychotic symptoms
Rabiner, 1986 (New York)	36	1	0.44	0.28	0.61	Relapsed or in-episode evidence of psychotic symptoms
Sartorius, 1986 (Multinational)	1,352	2	0.40	0.37	0.42	Unremitting psychotic symptoms
Shepherd, 1989 (United Kingdom)	107	5	0.43	0.34	0.52	Remained impaired throughout follow-up
McCreadie, 1989 (Scotland)	49	2	0.39	0.25	0.52	Psychotic symptoms present at 2-year follow-up
Marneros, 1992 (Germany)	355	25	0.17	0.13	0.21	Psychotic symptoms present at the time of follow-up
Thara, 1994 (India)	76	10	0.07	0.01	0.12	Continuous psychotic symptoms
an der Heiden, 1995 (Germany)	56	14	0.59	0.46	0.72	Moderate to severe symptoms at the time of follow-up
Mason, 1996 (United Kingdom)	58	13	0.34	0.22	0.47	Continuous symptoms in 2 years prior to folllow-up
Wieselgren, 1996 (Sweden)	101	5	0.14	0.07	0.21	"Poor outcome" evidence of ongoing persistant psychotic symptoms
Ganev, 1998 (Bulgaria)	55	16	0.45	0.32	0.59	Chronic continuous psychotic symptoms in 2 years prior to assessment
Wiersma, 1998 (Netherlands)	82	15	0.11	0.04	0.18	Chronic continuous psychotic symptoms over full follow-up time
Vazquez–Barquero, 1999 (Spain)	76	10	0.09	0.03	0.16	Chronic continuous psychotic symptoms
Stirling, 2003 (United Kingdom)	49	10	0.06	−0.01	0.13	Global Assessment of Functioning (GAF) score = 1 at follow-up

Table 12.5–8.

Summary of Findings from Follow-Up Studies on Course and Outcome of Psychosis Spectrum Disorder

1. The course of psychosis is highly variable both *within* and *between* patients, even after taking account of methodologic heterogeneity across studies. There is a broad range of possible course patterns ranging from complete recovery to continuous unremitting psychopathology, cognitive performance, and social functioning. Between such extremes, a substantial number of patients present with multiple episodes of psychosis interspersed with partial remission.
2. Less than half of patients diagnosed with psychosis show substantial clinical improvement after follow-up time averaging 6 years. Although these estimates are more optimistic than the earlier reports based on observations in institutional settings, they importantly emphasize the fact that psychosis is more often associated with persistent chronic disability that profoundly affects a person's development and quality of life.
3. Examining the trends in course and outcome over the past century revealed substantial gains in favorable outcome from the 1920s until the 1970s, after which time the gradient seemed to reverse (slightly). This biphasic trend may reflect developments in treatment, changes in diagnostic criteria across the decades, as well as possible sampling bias. Moreover, a secular trend (benign metamorphosis) toward better outcome and the virtual disappearance of the most lethal forms of catatonia was evident prior to the introduction of antipsychotics. The reasons for this trend are unclear, but commentators have speculated that they reflect changes in care and expectations within health service settings, or the development of psychological, social, and rehabilitation approaches to treatment. Unfortunately, the current cannon of outcome studies (as published) does not lend itself to easily disentangling direct disease effects from environmental factors that may influence the persistence of the disorder.
4. Course descriptors vary as a function of the length of follow-up, with the greatest variability being in the first 5 years, after which time the course tends to plateau. However, cross-sectional outcome measures of psychopathology do not differ greatly according to study duration, suggesting there is no clear pattern of deterioration.
5. On average, patients with a diagnosis of schizophrenia have the poorest outcome, with schizoaffective patients occupying an intermediate position between schizophrenia and affective psychosis; however, diagnostic categories of psychosis poorly predict outcome. Further, course and outcome estimates vary depending on the

diagnostic classification used. Patients diagnosed using broadly defined schizophrenia have generally better outcomes compared to the narrowly defined (post-DSM-III) schizophrenia.

6. There is no reliable set of predictors yet identified for course and outcome. Furthermore, the predictive value of putative prognostic indicators is generally poor and likely to depend on the duration of illness (stage) and the projection time. Significant methodologic differences in the definition of predictor variables, adjustments for other known risk factors, differential attrition rates (loss to follow-up), and variation in outcome measurement reduce the comparability of studies on the prediction of prognosis and make exploratory subgroup analyses difficult, despite a large number of published studies. Best estimates to date suggest that all known prognostic factors or psychopathology explain only a small amount of the total variance in outcome (estimates at around 30%, depending on the outcome domain under study), which translates into low predictive power.
7. Very little is known about outcomes of existential recovery (recovery of perspective and experience of a meaningful life despite difficulties), despite growing evidence that existential recovery can be attained by many patients.

Risk Factor for the Persistence of Psychosis (5 Years' Follow-Up)

Although there have been considerable advances and concerted research efforts into the determinants of the onset of psychosis, there has been a paucity of studies examining the determinants of persistence over the past decade. The body of research evidence reflects this neglect. Many early studies of family risk and outcome of psychosis found an association between the good outcome in psychosis and a family history of affective disorders. However, the findings were inconsistent and, at times, contradictory. There was significant methodologic heterogeneity that could potentially bias the results, such as retrospective design, insufficient adjustment for potential confounders (family size, age, and sex of the relatives), the use of global unidimensional outcome measures, and mixed samples (recent onset and chronic disorders studied together).

There have been few recent studies examining the effect of family risk on the course of psychosis. The largest study to date examined outcome and family history in a sample of 253 DSM-III patients with psychosis and 723 first-degree relatives. There was a higher morbidity risk of psychosis (spectrum) disorder in the relatives of probands with poor outcome disorders, although this did not reach statistical significance. However, there was a (statistically significant) higher morbidity risk of affective disorder in the relatives of probands with good outcome disorder. A study from Camberwell examined the associations between 4-year outcomes and familiar loading for a range of psychotic disorders in a sample of 150 consecutively admitted patients with recent onset disorder. Higher familiar loading for psychosis predicted higher rates of persistent negative symptoms (estimated odds ratio 1.5), more time in the hospital, and poorer social functioning at follow-up. The effect of familiar loading was greatest in the group of patients with a diagnosis of schizophrenia compared to other psychotic disorders. These findings suggest poor outcome disorders, with higher levels of negative symptoms that may “breed true” within families. However, recent studies using molecular measures of PRS for schizophrenia did not find clinically meaningful associations with measures of course and outcome. There are relatively few studies examining the relationship between pregnancy and birth complications, and psychosis, and results are mixed.

There is evidence that childhood cognitive ability is associated with the outcome; lower intelligence has been shown to predict unfavorable clinical and functional outcomes at follow-up. There is interest in the question of to what degree any association between cognitive ability and outcome may be mediated by metacognition (referring to the cognitive processes involved in thinking about thinking) and by factors that may confound cognitive ability measurement in patients with a psychotic disorder such as motivation, effort, defeatist performance beliefs, stress, anxiety, and physical inactivity. Studies also have found that subtle social premorbid adjustment is associated with a less remitting course and poorer functional outcomes. There are now a number of relatively small- to medium-term (2 to 5 years) follow-up studies comparing the course and outcome of patients with psychosis exposed to cannabis versus nonexposed patients. They are methodologically heterogeneous, but some consistency across studies is evident. The cannabis-exposed patients have more positive symptoms (but not negative symptoms), and the course tends to be relatively unremitting. A poorer outcome in the domains of work and education has been demonstrated, which are most likely reflecting factors that independently increase the likelihood of early cannabis use and later education and labor difficulties. A review of reviews concluded that cannabis use was associated with increased relapse rates, more hospitalizations, and more pronounced positive symptoms in patients with a psychotic disorder. By contrast, psychosis that follows stressful life events may have a better outcome. The risk factors for the persistence of psychosis have been less intensively and comprehensively studied than the risk factors for the onset of psychosis. Further research is required to draw any firm conclusions that risk factors for onset may also be determinants influencing outcome (discrete effect models) but is worthy of further investigation. Discrete effects can be broadly grouped into two groups: (1) early developmental abnormalities and premorbid cognitive alterations, which may be more strongly associated with a clinical symptom profile characterized by cognitive alterations, negative symptoms, insidious onset, unremitting course, and functional deterioration, which are linked to the neurodevelopment model of psychosis; (2) adverse life events and migrant status, which may have a less specific association with the schizophrenia syndrome and whose clinical correlates suggest less prominent negative symptoms, acute onset, remitting course, and better functional outcomes, which may be linked to a mechanism of social stress sensitivity.

TOWARD AN EMPIRICALLY BASED MODEL OF PSYCHOSIS SPECTRUM DISORDER: A WORK IN PROGRESS

Understanding cause is fundamental to the development of scientific knowledge and the development of empirically based models of disease. Modern medicine has developed a deductive system for evaluating causation based on Koch postulates and Hill criteria (and their elaborations). The integrated interpretation of epidemiologic, clinical, and experimental research aided by such causal criteria has resulted in the successful identification of important risk factors for some leading causes of morbidity and mortality (e.g., cardiovascular disease), some forms of cancer, and infectious diseases. Why has this type of research, which has been reasonably successful for other diseases, been less successful for psychosis spectrum disorder? At least three (interrelated) factors may be impeding progress: (1) the lack of a sufficiently validated operational definition for psychosis outcomes (caseness definition). (2) The disorder has a spectrum-like distribution with a complex time course and a broad range of ages of onset. The most common course pattern is episodic (single or multiple), with delusions and hallucinations being prominent symptoms; however, mounting evidence from population studies suggests that these symptoms are also relatively prevalent in the general population and other psychiatric disorders. A more transdiagnostic approach toward research may be required. Furthermore, risk factors for psychosis probably operate throughout the life course, perhaps even prior to birth. (3) Classical epidemiologic strategies assume a static unidirectional cause–effect linear association, which may be of limited usefulness if psychosis is better conceptualized as a dynamic transitional process of small iterations of phenotypic change mediated by the complex interplay and coevolution of environmental, mental, and biologic risk factors that are continuously distributed in the general population. While the most widely accepted contemporary model of psychosis is based on a neurodevelopmental process, in recent years, there has been a substantial move toward an integrated model of psychosis based on consistent and strong evidence that implicates socioenvironmental and attenuated psychosis risk factors in the development of psychosis.

Neurodevelopmental Model of Psychosis Spectrum Disorder

The broad neurodevelopmental hypothesis posits that psychosis is the behavioral outcome of subtle deviances in early brain development, the full effect of which is not manifest until adolescence or early childhood. There is some evidence for the broad neurodevelopmental model, in agreement with the developmental abnormalities and possible prenatal risk factors reviewed in this section. Further evidence is the absence of inflammatory reactions in the brains of people with a diagnosis of psychosis. Ultrastructural refinement at the synapse is shaped by patterns of environmental input that occur across the lifespan and seems to have functional significance. Such plasticity in the connectivity of neurons is thought to underlie new learning and the development of new schemata and mental skills. There is generally a long latency period between exposure (before adulthood) to known risk factors and the onset of psychosis. Such a delayed effect may be consistent with a causal model where the environmental risk factor increases a person’s liability, but the full manifestation of the disorder, beyond attenuated symptoms, requires the addition or interaction of other complementary (gene or environmental) causal components.

Epidemiologic Psychosis Proneness—Persistence—Impairment Model

Emerging evidence from general population studies shows that attenuated psychotic symptoms are relatively common (prevalence of around 8%, based on a recent meta-analysis). However, most are brief, self-limiting developmental phenomena. The reason why approximately 15% of children have attenuated psychotic symptoms during development is not fully established, but it may, in part, reflect the shared genetic liability for psychosis continuously distributed in the general population. In two large independent general population samples, NEMESIS and the German Early Developmental Stages of Psychopathology Study (EDSD), investigators examined the hypothesis that these relatively common, developmental attenuated psychotic symptoms could become abnormally persistent when synergistically combined with known risk factors for the onset of psychotic disorders (the study used cannabis, urban upbringing, and developmental trauma). Over a follow-up period of 3 years, the studies found the rates of attenuated psychotic symptoms that *persisted* in being relatively low, 26% (NEMESIS) and 31% (EDSP). However, rates of persistence increased progressively with higher baseline doses of exposure to known risk factors. A systematic review established that childhood trauma was associated with a greater degree of persistence of attenuated psychotic symptoms over time. In addition, over a 10-year follow-up in the EDSP study, it was established that the greater the risk of persistence of attenuated psychotic symptoms over time, the greater the risk of eventual transition to clinical psychotic disorder. This suggests that environmental determinants of psychotic disorders may, in part, operate by driving up the rates of affective dysregulation and, if severe enough, attenuated psychotic symptoms, which become persistent in individuals with a genetic liability for psychosis. A dynamic model of psychosis evolution or proneness-persistence-impairment model may best conceptualize this trajectory from a developmental expression of affective dysregulation and psychosis liability (transitory subclinical psychotic symptoms) to more intrusive persistent subclinical psychotic symptoms to significant impairment and dysfunction and the eventual onset of psychotic disorder (Fig. 12.5–2). Such a model is supported by recent meta-analytic evidence and provides a framework for the further investigation of the ontogenesis of psychotic disorders. The psychosis–proneness—persistence–impairment model of psychosis offers a possible epidemiologic insight into the underlying mechanism for the development of psychosis.

Integrated Model of Psychosis Spectrum Disorder

The strong and consistent evidence that implicates socioenvironmental risk factors in the etiology of psychosis and the evidence suggesting a progression from psychosis proneness to persistence and, finally, full-blown psychosis, reviewed in this section, has recently led investigators to posit an integrated model of psychosis. In this model, the neurodevelopmental model is extended to and combined with other hypotheses, including sociodevelopmental, experiential, cognitive, neurobiologic, and G × E/epigenetic. Given the variation in incidence and expression of psychosis across different social contexts, this model introduces a level of explanation focusing on the impact of the environment, in interaction with genetic risk, on cognitive schemata as well as on brain neurobiologic mechanisms across the life course. So, for example, exposure to trauma during childhood may lead to the development of subtle developmental delays, cognitive alterations, and affective dysregulation and, thereby, predispose individuals to a paranoid way of thinking, which, when coupled with a sensitized dopamine system, may manifest as psychosis proneness which, in combination with further environmental impact, may result in persistence over time and increase liability to psychotic disorder. Given that every human being has thousands of genetic risk variants for psychosis, suggesting implication in natural processes like the ability to distill signals and meaning from environmental noise, psychosis susceptibility genes may interact with environmental factors to affect signal-to-noise ratio across the life course, an effect which may be mediated by an (epigenetically) sensitized dopamine system.

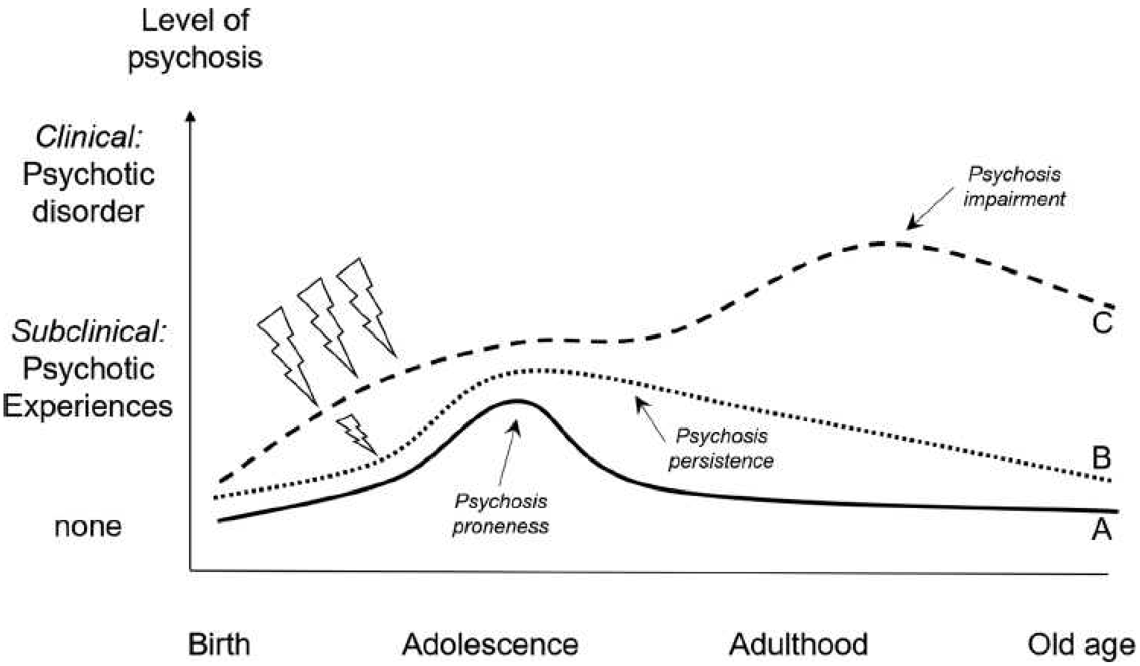
Level of psychosis
Clinical: Psychotic disorder
Subclinical: Psychotic Experiences
none
Psychosis impairment
Psychosis persistence
Psychosis proneness
C
B
A
Birth
Adolescence
Adulthood
Old age

FIGURE 12.5–2. Epidemiologic psychosis—proneness–persistence–impairment model of psychosis onset. Person A has "normal" developmental expression of subclinical psychotic experiences (*psychosis proneness*) that are transient. Person B has similar expression but longer *persistence* due to additional but mild environmental exposure. Person C has longer persistence due to severe, repeated environmental exposure and transition to clinical psychotic disorder with significant *impairment*.

FURTHER READINGS

*Anttila V, Bulik-Sullivan B, Finucane HK, et al; Brainstorm Consortium. Analysis of shared heritability in common disorders of the brain. *Science*. 2018;360(6395):eaap8757.

Cupo L, McIlwaine SV, Daneault JG, et al. Timing, distribution, and relationship between nonpsychotic and subthreshold psychotic symptoms prior to emergence of a first episode of psychosis. *Schizophr Bull*. 2021;47:604–614.

DeVylder JE, Kelleher I, Lalane M, et al. Association of urbanicity with psychosis in low- and middle-income countries. *JAMA Psychiatry*. 2018;75(7):678–686.

*Di Forti M, Quattrone D, Freeman TP, et al. The contribution of cannabis use to variation in the incidence of psychotic disorder across Europe (EU-GEI): a multicentre case-control study. *Lancet Psychiatry.* 2019;6(5):427–436.

*Guloksuz S, Pries L-K, Delespaul P, et al. Examining the independent and joint effects of molecular genetic liability and environmental exposures in schizophrenia: results from the EUGEI study. *World Psychiatry.* 2019;18:173–182.

Guloksuz S, Pries L-K, Ten Have M, et al. Association of preceding psychosis risk states and non-psychotic mental disorders with incidence of clinical psychosis in the general population: a prospective study in the NEMESIS-2 cohort. *World Psychiatry.* 2020;19(2):199–205.

Hasmi L, Pries L-K, Ten Have M, et al. What makes the psychosis 'clinical high risk' state risky: psychosis itself or the co-presence of a non-psychotic disorder? *Epidemiol Psychiatr Sci*. 2021;30:e53.

*Hegarty JD, Baldessarini RJ, Tohen M, Waternaux C, Oepen G. One hundred years of psychosis: a meta-analysis of the outcome literature. *Am J Psychiatry*. 1994;151(10):1409–1416.

Howes OD, Murray RM. Psychosis: an integrated sociodevelopmental–cognitive model. *Lancet.* 2014;383:1677–1687.

*Jongsma HE, Gayer-Anderson C, Lasalvia A, et al. Treated incidence of psychotic disorders in the multinational EU-GEI study. *JAMA Psychiatry.* 2018;75:36–46.

*Kirkbride JB, Keyes KM, Susser E. City living and psychotic disorders-implications of global heterogeneity for theory development. *JAMA Psychiatry*. 2018;75(12):1211–1212.

*Linscott RJ, van Os J. An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychol Med*. 2013;43:1133–1149.

*Maj M, van Os J, De Hert M, et al. The clinical characterization of the patient with primary psychosis aimed at personalization of management. *World Psychiatry*. 2021;20:4–33.

*Marconi A, Di Forti M, Lewis CM, Murray RM, Vassos E. Meta-analysis of the association between the level of cannabis use and risk of psychosis. *Schizophr Bull*. 2016;42:1262–1269.

Menezes NM, Arenovich T, Zipursky RB. A systematic review of longitudinal outcome studies of first-episode psychosis. *Psychol Med*. 2006;36(10):1349–1362.

*Morgan C, Dazzan P, Lappin J, et al. Rethinking the course of psychotic disorders: modelling long-term symptom trajectories. *Psychol Med.* 2021;52(13):2641–2650.

Morgan C, Hibben M, Esan O, et al. Searching for psychosis: INTREPID (1): systems for detecting untreated and first-episode cases of psychosis in diverse settings. *Soc Psychiatry Psychiatr Epidemiol*. 2015;50:879–893.

Moritz S, Silverstein SM, Beblo T, et al. Much of the neurocognitive impairment in schizophrenia is due to factors other than schizophrenia itself: implications for research and treatment. *Schizophrenia Bulletin Open*. 2021;2(1).

Newbury JB, Arseneault L, Caspi A, et al. Association between genetic and socioenvironmental risk for schizophrenia during upbringing in a UK longitudinal cohort. *Psychol Med*. 2020;52:1527–1537.

Plana-Ripoll O, Pedersen CB, Holtz Y, et al. Exploring comorbidity within mental disorders among a Danish national population. *JAMA Psychiatry*. 2019;76:259–270.

*Pries L-K, Lage-Castellanos A, Delespaul P, et al. Estimating exposome score for schizophrenia using predictive modeling approach in two independent samples: the results from the EUGEI study. *Schizophr Bull*. 2019;45:960–965.

*Reininghaus U, Bohnke JR, Chavez-Baldini U, et al. Transdiagnostic dimensions of psychosis in the bipolar-psychosis network on intermediate phenotypes (B-SNIP). *World Psychiatry*. 2019;18(1):67–76.

*Rothman KJ, Greenland S. Causation and causal inference. In: Rothman KJ, Greenland S, eds. *Modern Epidemiology*. 2nd ed. Lippincott-Raven Publishers; 1998:7–29.

Selten JP, van der Ven E, Termorshuizen F. Migration and psychosis: a meta-analysis of incidence studies. *Psychol Med*. 2020;50:303–313.

Trotta A, Murray RM, Fisher HL. The impact of childhood adversity on the persistence of psychotic symptoms: a systematic review and meta-analysis. *Psychol Med*. 2015;45:2481–2498.
*van Os J, Kenis G, Rutten BP. The environment and schizophrenia. *Nature*. 2010;468:203–212.
van Os J, Reininghaus U. Psychosis as a transdiagnostic and extended phenotype in the general population. *World Psychiatry*. 2016;15(2):118–124.
*Varese F, Smeets F, Drukker M, et al. Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective- and cross-sectional cohort studies. *Schizophr Bull*. 2012;38(4):661–671.
Weiser M, Reichenberg A, Rabinowitz J, et al. Association between nonpsychotic psychiatric diagnoses in adolescent males and subsequent onset of schizophrenia. *JAMA Psychiatry*. 2001;58(10):959–964.
Werbeloff N, Drukker M, Dohrenwend BP, et al. Self-reported attenuated psychotic symptoms as forerunners of severe mental disorders later in life. *Arch Gen Psychiatry*. 2012;69:467–475.

12.6 Cellular and Molecular Neuropathology of Schizophrenia

ALI S. IMAMI, M.B.B.S., M.S., NICHOLAS D. HENKEL, PH.D., JUSTIN F. CREEDEN, PH.D., AND ROBERT E. MCCULLUMSMITH, M.D., PH.D.

Schizophrenia is a chronic and debilitating brain disease. Despite concerted efforts, fully understanding the pathophysiology of the illness has been a challenge to the field. A factor in this has been the scarcity of reliable direct data, as direct observation of anatomical and molecular characterization is only possible with postmortem human brains. It was not until the 1980s, with the emergence of more sophisticated tools and techniques, that there was a deeper interest in the characterization of the pathophysiology of schizophrenia.

While there is no obvious, identifiable neuropathologic lesion that causes schizophrenia, there is a slew of subtle pathologies that underlie the illness. Over the past 40 years, many postmortem studies have been published, examining the various angles of the pathology at the tissue and molecular level. However, due to incomplete guiding hypotheses, the postmortem literature on schizophrenia is very broad and, at times, fragmented. Nonetheless, a consensus has emerged that suggests localization to specific brain regions such as the prefrontal cortical regions, thalamus, and limbic structures in the brain.

MORPHOMETRIC STUDIES AND CYTOARCHITECTURAL CHANGES

Gross anatomical and histologic changes have been a consistent feature of schizophrenia. There have been several studies that imply changes in the architecture of frontal lobe regions and their gross structure.

Whole Brain

The most consistent observation in schizophrenia is the normal macroscopic appearance of the postmortem brain. However, over the course of the past three decades, postmortem brain tissue studies have begun to provide clues into anatomical and histologic correlates of schizophrenia pathology. It should be noted that their findings are disparate, sometimes discrepant, or even conflicting. This may be due to disease heterogeneity, methodologic issues such as quality of postmortem tissue, sample size, or confounders of postmortem research such as tissue quality. Nonetheless, a coherent picture supported by converging data is beginning to emerge.

A meta-analysis looking at postmortem brain weight found a small but significant reduction in schizophrenia that inversely correlates with the age of onset. Consistent with this finding, structural imaging studies demonstrate that brain volumes are smaller in the prodrome, at the time of first break psychosis, and in the chronically ill. A recent longitudinal study reported reduced prefrontal and temporal lobe volumes in individuals with a high genetic load for developing schizophrenia. Those individuals who did go on to develop schizophrenia showed even greater reductions in prefrontal volumes. Furthermore, there is evidence that gray matter volume reductions may be progressive in the initial stages of the illness. Regional postmortem studies have identified evidence of subtle changes in cellular morphology, with regions most often implicated being the prefrontal cortex (PFC), thalamus, and medial temporal lobe (Table 12.6–1). An initial study in 1982 found that a large proportion of patients with schizophrenia showed evidence of gliosis. While several studies since then have not supported this finding, more recent studies suggest that the different types of glia may play roles in schizophrenia pathology. The absence of other classical attributes of neurodegenerative diseases, such as obvious neuron loss or neurofibrillary tangles, has led to the belief that schizophrenia is not a degenerative disease. Rather, the reported glia-related changes raise the possibility of an immunologic or inflammatory component to disease pathophysiology. A landmark study in 1976 reported ventricular dilatation in subjects with chronic schizophrenia. Since then, numerous studies have replicated this observation, primarily enlargement of the third and lateral ventricles in subjects with schizophrenia. This observation, however, lacks diagnostic specificity, as it can occur in numerous other brain disorders, such as Alzheimer disease and hydrocephalus. In addition, high-risk and family studies show that enlarged ventricular volume is also found in unaffected relatives who share genetic material (and, by extension, genetic risk for schizophrenia).

Table 12.6–1.
Summary of Golgi Studies Examining Dendritic Spine Density in Schizophrenia

Region	Reported Finding	Observations
Prefrontal cortex	Lower	In layer 3
Temporal cortex	Lower	In layer 3
Visual cortex	Unchanged	In layer 3
Hippocampus		
Subiculum	Lower	Apical dendrites
CA3	Higher	Apical dendrites

A more recent observation is related to the laterality of the brain's hemispheres. There has been evidence that there is loss of laterality in the schizophrenia brain in patients with earlier onset. This observation may interact with other variables like sex, disease severity, and duration of illness. In addition, abnormal gyrification has been frequently observed in schizophrenia.

Prefrontal Cortex

The PFC has been hypothesized as a major site of dysfunction in schizophrenia. There is reduced task-related prefrontal activation in patients with the illness and prominent deficits in certain PFC-dependent cognitive functions. Structural imaging studies have provided evidence that frontal lobe volume is reduced in schizophrenia. Stereologic analysis of postmortem brain samples has refined this observation by ascribing this volume deficit primarily to a 3% to 12% reduction in the gray matter of schizophrenia subjects compared to controls. Gray matter is composed of neurons, glial cells, and neuropil that include neuronal processes such as dendritic trees and axonal arbors. Initial studies using a two-dimensional stereologic approach detected neuronal loss in the PFC with no evidence of change in glial numbers.

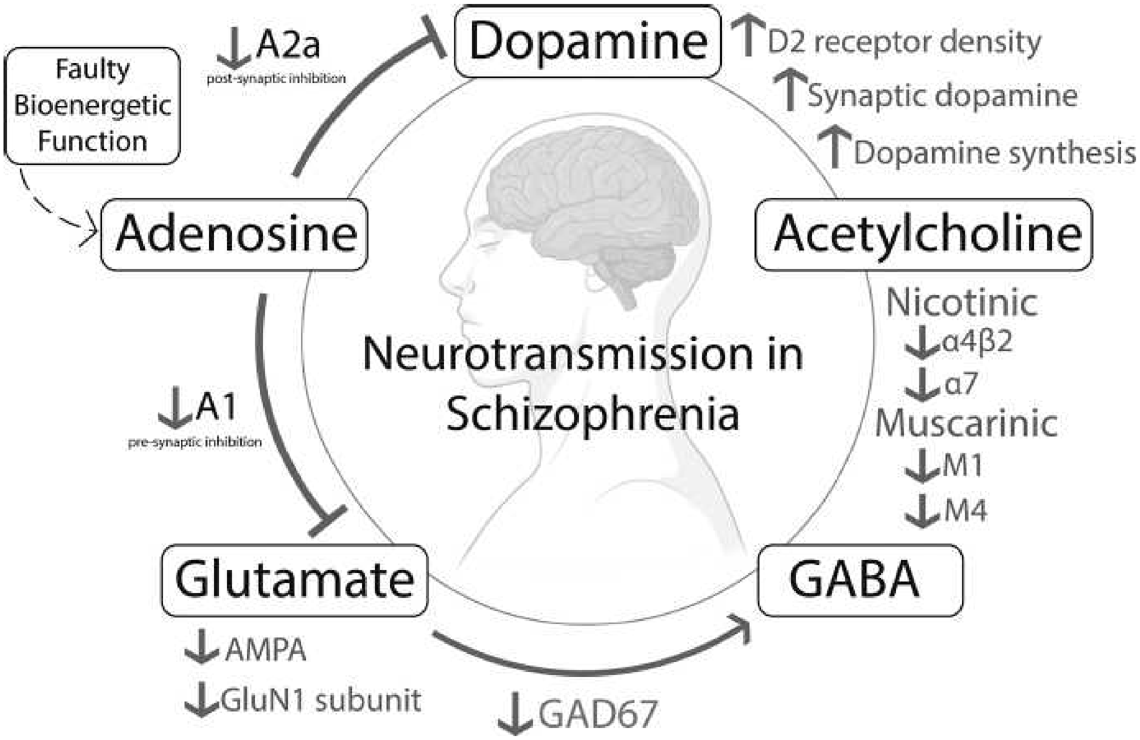
Faulty Bioenergetic Function
A2a
post-synaptic inhibition
Dopamine
D2 receptor density
Synaptic dopamine
Dopamine synthesis
Adenosine
Acetylcholine
Nicotinic
α4β2
α7
Muscarinic
M1
M4
Neurotransmission in Schizophrenia
A1
pre-synaptic inhibition
Glutamate
GABA
AMPA
GluN1 subunit
GAD67

FIGURE 12.6–1. Perturbed neurotransmitter systems in schizophrenia. From top, moving counterclockwise: the biochemical evidence for increased dopaminergic neurotransmission has been limited to the D2 receptor. Hyperdopinamergia, on the other hand, has been well supported in live-human, imaging studies. Adenosine signaling inhibits the release of dopamine and glutamate. The canonical adenosine receptors, A1 and A2, transcript expression is decreased in schizophrenia. These receptors may modulate glutamate and dopamine release and signaling at the pre- and postsynapse, respectively. Though further work is required, adenosine receptor expression changes may be attributed to faulty bioenergetic function in the disease. Further, while changes in adenosine modulate glutamatergic neurotransmission, glutamate receptor changes also have been reported in the AMPA receptor and in the GluN1 subunit of the NMDA receptor. GABA levels may be altered as the enzyme GAD67, which converts glutamate to GABA, is decreased in the prefrontal cortex. Of all the neurotransmission systems, the muscarinic and nicotinic acetylcholine receptor changes are the consistently reported in postmortem brain and in preclinical models of schizophrenia.

More recent studies have employed a three-dimensional stereologic approach that considers cell size and number per unit volume. These studies have not found changes in neuron numbers, suggesting that the volume decrease is not attributed to neuronal loss but rather reductions in neuropil. Indeed, quantitative analyses of neuron number and density in the PFC (Brodmann areas 9 and 46) demonstrated increases in neuronal density, smaller pyramidal somal cell volumes, and reduced neuropil in schizophrenia in layer 3 (Fig. 12.6–1) but not in layer 5 or 6. The reductions in neuropil suggest that schizophrenia is associated with a quantitative deficit in neuronal connections. Several studies have examined dendritic spine density in schizophrenia (Table 12.6–1) and consistently find reductions in spine density in layer 3 of the PFC. The layer 3 pyramidal neurons serve as a major site for the integration of information from other intrinsic cortical neurons and input from the thalamus, implicating these neural pathways in schizophrenia pathophysiology.

The issue of the influence of long-term antipsychotic treatment and its possible effect on these cellular measures has been examined. Exposure of nonhuman primates to chronic antipsychotic medication does not appear to alter the cortical neuronal density, suggesting that the elevated neuronal density found in schizophrenia is not a medication effect. Interestingly, in these chronically treated animals, the numbers of glia increased as much as 33% in certain cortical layers. The implication of this is not yet clear but may suggest a role for glia in terms of antipsychotic response.

Hippocampal Formation

Magnetic resonance imaging (MRI) studies find consistent reductions in hippocampal or amygdala–hippocampal volumes in schizophrenia that may be lateralized to the left side. Meta-analyses report hippocampal reductions of the magnitude of 4% in schizophrenia. Smaller hippocampal volumes are present at the time of disease onset and in those with high genetic loading for the illness. Human postmortem studies suggest an anteroposterior gradient of the neuropathologic findings, with the anterior part putatively more affected than the posterior part. If the volume decrease holds true, what causes it is still a matter of debate. The current consensus is that there are no significant differences in neuronal density in the hippocampus in schizophrenia, although it remains a possibility that there are cell-type–specific differences in subregions of the hippocampus. Studies of neuronal size have demonstrated somewhat more consistent results. As in the PFC, smaller pyramidal neurons have been reported in the hippocampus, but with normal-sized neurons in the visual and motor cortex in schizophrenia. A question that remains to be answered pertains to the timeline of changes. Although reduced hippocampal volumes can be identified at the beginning of the illness, imaging data attest to a progressive course (i.e., hippocampal volumes continue to decrease with disease advancement), which may suggest a potential toxic effect of the ongoing disease or drug treatment, particularly first-generation antipsychotic medications. Two Golgi impregnation studies have examined spine density in hippocampal pathways. One reported a reduction in spine density, while another recent study reported a specific increase in spine density on CA3 pyramidal cell apical dendrites postulated to reflect an increase in CA3 basal activity, thereby raising the possibility of subregional changes within the hippocampus. In the entorhinal cortex, early studies showed neurons in aberrant and poorly formed clusters. Since the spatial arrangement of cells is largely determined during brain development in utero, a disruption in this arrangement could reflect a developmental abnormality. Well-designed follow-up studies considering the rostrocaudal heterogeneity of the entorhinal cortex have not replicated these findings, and the results remain controversial. Similarly, initial reports of reduced number and smaller size of neurons in the entorhinal cortex have not held up to replication.

Thalamus

The thalamus, because of its pivotal position in relationship to afferent sensory information and as a station in the cortical–subcortical circuits, is a brain region of interest. The volume, cell number, and cell densities have been investigated in several thalamic nuclei, including the mediodorsal (MDL) and anterior thalamic nuclei, where the main efferent connections to the PFC and limbic regions arise. In the MDL nuclei, total neuronal counts are decreased, although not in all studies. These reductions in neuron number are reported to be in the parvocellular and densocellular parts of the MDL, which project primarily to the dorsolateral PFC (DLPFC). More recent studies with larger sample sizes have not, however, replicated these findings of reduced neuronal number.

Table 12.6–2. Summary of Main Morphologic Findings in Human Postmortem Studies in Schizophrenia Compared to Control Cases

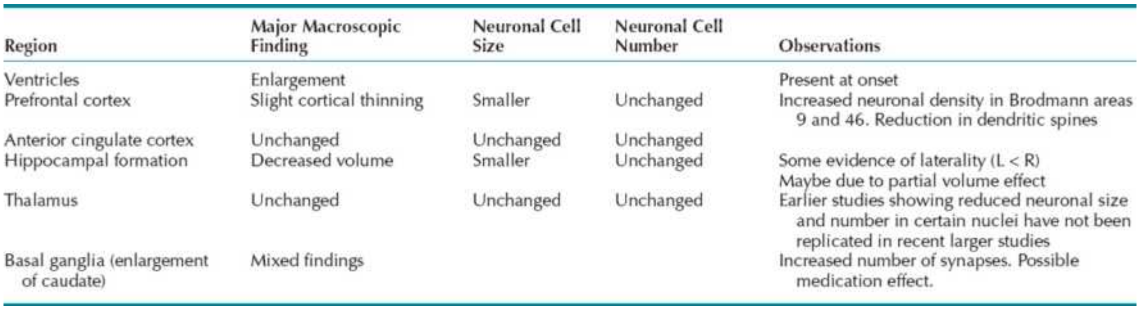

Region	Major Macroscopic Finding	Neuronal Cell Size	Neuronal Cell Number	Observations
Ventricles	Enlargement			Present at onset
Prefrontal cortex	Slight cortical thinning	Smaller	Unchanged	Increased neuronal density in Brodmann areas 9 and 46. Reduction in dendritic spines
Anterior cingulate cortex	Unchanged	Unchanged	Unchanged	
Hippocampal formation	Decreased volume	Smaller	Unchanged	Some evidence of laterality (L < R) Maybe due to partial volume effect
Thalamus	Unchanged	Unchanged	Unchanged	Earlier studies showing reduced neuronal size and number in certain nuclei have not been replicated in recent larger studies
Basal ganglia (enlargement of caudate)	Mixed findings			Increased number of synapses. Possible medication effect.

Basal Ganglia

Postmortem and in vivo imaging studies have reported volumetric increases in the caudate–putamen complex in patients with schizophrenia. Although animal and human studies initially suggested that the volumetric increase could be explained by antipsychotic medications, electron microscopic studies suggest an increase in the number of synapses in the caudate and putamen, suggesting increased neuronal activity in the basal ganglia. Whether this is disease related or a medication effect is not yet conclusive.

Table 12.6–2 summarizes the morphologic findings seen in schizophrenia.

NEUROTRANSMITTER SYSTEMS

Neurochemical studies began in schizophrenia at a time when scientists were anticipating a single neurobiologic defect. Although no single defect has been found, investigators interpret studies in a broader systems context. Nonetheless, findings are often still organized by neurotransmitter system.

Dopamine

In the 1960s, it was proposed that hyperdopaminergic activity may be responsible for the positive psychotic symptoms in schizophrenia. This was based on the pharmacologic observation that blockade of dopamine receptor subtype 2 (D_2) reduced positive psychotic symptoms. A series of studies in the nonhuman primate showed that a reduction in dopamine transmission in the PFC was associated with cognitive deficits such as those found in schizophrenia. This led to a modification of the dopamine hypothesis to the presence of an imbalance of dopamine transmission in schizophrenia, with dopaminergic hyperactivity in subcortical regions but hypoactivity in prefrontal cortical regions. These findings served as the compelling impetus to study biochemical measures of dopamine neurotransmission in the human postmortem brain in schizophrenia.

Although the dopamine hypothesis has been the prevailing hypothesis for decades, evidence of dopamine dysfunction in human postmortem studies has not been forthcoming.

Studies examining D_1 expression in the PFC and striatum have been largely negative. Increases in D_2 receptors are a consistent finding in several postmortem studies, but since chronic antipsychotic treatment is known to upregulate D_2 receptors, these findings have been largely attributed to a medication effect. Consistent with this notion, the density of D_2 receptors is increased in schizophrenia, while the density of D_1 receptors is not changed. D_3 receptor expression was increased in the ventral striatum of medication-free cases of schizophrenia but normal in subjects treated with antipsychotics. D_3 messenger RNA levels are, however, reported to be normal in the accumbens. There was considerable interest in the D_4 receptor, fueled in part by the high affinity of clozapine for D_4 receptors, but neither genetic nor postmortem studies have been able to provide conclusive associations between this receptor and schizophrenia. In addition, ligand-binding studies specifically targeting the D_4 have not been efficacious in mitigating psychosis. Finally, dopamine active transporter (DAT) expression in the striatum has been examined by several research groups, the majority finding that levels are unaltered in schizophrenia.

On the other hand, studies with living schizophrenia subjects have yielded support for hyperdopaminergia. In vivo imaging studies have demonstrated pronounced differences in dopamine neurotransmission in schizophrenia subjects. Schizophrenia subjects exhibit increased nascent dopamine synthesis, have higher subcortical dopamine content, and, when exposed to amphetamines, release more dopamine. Increased dopamine synthesis and synaptic content are not only highly correlated with the presence of positive symptoms but are also associated with the resolution of positive symptoms when antipsychotic treatment is initiated. However, the measurement of dopamine and metabolites in lumbar cerebrospinal fluid (CSF) as a reflection of neurotransmitter activity in the brain has not contributed to the understanding of schizophrenia pathophysiology.

Glutamate

The glutamate hypothesis of schizophrenia is centered on the clinical observation that N-methyl-D-aspartate (NMDA) receptor antagonists, such as phencyclidine (PCP) and ketamine produce a syndrome that is indistinguishable from schizophrenia.

Glutamate is the major excitatory amino acid neurotransmitter in the brain and is known to activate both ionotropic and metabotropic glutamate receptors (mGluRs).

The ionotropic receptors, α-amino-3-hydroxyl-5-methyl-4-isoxazolepropionic acid (AMPA), kainite (KA), and NMDA receptors, have been extensively studied at the mRNA, protein, and ligand-binding level in the cortex, striatum, thalamus, and limbic regions in schizophrenia.

In schizophrenia, hippocampal AMPA receptor subunit transcript and protein expression are consistently decreased. Likewise, the NMDA GluN1 subunit is downregulated in the temporal cortex, hippocampus, and thalamus. However, the NMDA GluN2B subunit is upregulated in the temporal cortex and the hippocampus. Among the mGluRs, mGluR2 and 3 have been implicated in both patients with and animal models of schizophrenia.

Studies showing a hypoactive NMDA Receptors are supported by further studies showing reduced expression of the protein complexes that mediate their downstream signaling, such as PSD-95, neurofilament light, and SAP102 in the PFC. There is emerging evidence that hypofunction of NMDARs may also be attributed to epigenetic changes that occur early in development. Such changes may impact other neurotransmitter systems.

γ-Aminobutyric Acid

γ-Aminobutyric acid (GABA) is synthesized from glutamate by glutamic acid decarboxylase (GAD). There are two isoforms of this enzyme, GAD_{65} and GAD_{67}, which are encoded by the genes *Gad1* and *Gad2,* respectively. A consistent finding in schizophrenia is the reduction in GAD_{67} mRNA levels in the PFC in schizophrenia. Since neuronal numbers are unchanged in the PFC, the decrease in GAD67 is likely due to decreased expression per neuron. Interestingly, the reduction in GAD67 found in the PFC is associated with reductions of GABA transporters in parvalbumin (PV)-positive interneurons. This effect, reportedly, is not due to antipsychotic treatment. With respect to interneuron subtypes, somatostatin interneurons, as well as the expression of somatostatin, are decreased. Broadly, GABA is modulated by several genes implicated in schizophrenia, such as *BDNF*, *DISC1*, and *NRG1*, which impact the migration of PV+ interneurons to the cerebral cortex, GAD67 expression, and development of GABA circuitry.

These changes suggest reduced synthesis and reuptake of GABA in these neurons and, thereby, altered GABAergic transmission. Dysfunction of GABAergic neurotransmission correlates with cognitive dysfunction in schizophrenia.

Acetylcholine

Cholinergic neurotransmission is integral to normal cognition and memory processing, functions disrupted in schizophrenia. Postmortem evidence for abnormalities in this system has shown decreased levels of the nicotinic (both the α7- and the α4β2-nACHR subtypes) and muscarinic receptors (namely the M_1 and M_4 receptors) in the hippocampus, frontal cortex, thalamus, and striatum in schizophrenia subjects. Moreover, genetic polymorphisms in the M_5 (*Chrm5*) and the α7-nACHR (*Chrna7*) genes are linked to a higher risk of developing schizophrenia.

While there is a substantial body of preclinical evidence suggesting the procognitive effects of modulating α7-nACHRs in rodent models of schizophrenia, human trials have often been unsuccessful. Recent evidence suggests that positive allosteric modulation of M1 receptors is impaired in schizophrenia, which may account for some of these failures. Reconciling the differences will require close examination of the study variables, including frequency, duration, and route of administration, as well as patient populations.

Further research is needed to confirm how these biochemical and genomic changes contribute to schizophrenia.

Adenosine

Adenosine is a neuromodulator of both the glutamate and dopamine neurotransmitter systems. Adenosine acts on purinergic receptors (e.g., A1 and A2A receptors) in the brain. Presynaptic activation of adenosine A1 receptors inhibits the release of glutamate, while postsynaptic activation reduces NMDA-receptor functioning. In addition, adenosine signaling through postsynaptic A2A receptors inhibits D2 receptor signaling through a heteromeric complex. For this reason, dysfunction of the adenosine system may play a role in the hyperdopaminergic and hypoglutamatergic states in schizophrenia.

Changes in the expression of the transcript and protein levels of enzymes that regulate extracellular adenosine levels as well as adenosine (i.e., purinergic) receptors have been consistently reported in postmortem schizophrenia studies.

Ectonucleoside triphosphate diphosphohydrolase (ENTPD)-1, the enzyme responsible for hydrolyzing adenosine triphosphate (ATP) to adenosine monophosphate (AMP) and adenosine diphosphate (ADP), transcript expression was decreased in the striatum. Transcript expression of both ENTPD-1 and -2 was decreased in astrocytes, but not neurons, isolated from the DLPFC. Interestingly, adenosine deaminase (ADA), which breaks adenosine down to inosine, is increased specifically in DLPFC neurons.

While no changes were found in A1 receptor protein expression in the striatum of schizophrenia subjects, A1 receptor transcript was reduced in an enriched population of pyramidal neurons from the DLPFC. In the striatum, a reduction in the transcript and the protein levels of A2A receptor was attributed to hypermethylation of the A2A coding region. However, the density of these A2A receptors was 70% higher in schizophrenia patients, possibly to compensate for low adenosine levels.

Conclusion

In summary, molecular abnormalities are found in several anatomic regions and in several neurotransmitter systems in schizophrenia. Identification of primary pathology from epiphenomenon is essential. For example, neurotransmitter systems are dynamic, and disruption of one system would lead to compensatory mechanisms in other relevant pathways. A sound strategy to adopt once an abnormality is discovered is to first confirm the finding by replication in a second cohort and then determine whether the molecular abnormality is part of the primary pathology. Converging data from in vivo human studies, human postmortem studies, and animal model studies would provide clues to the primary pathology.

BIOENERGETIC DYSFUNCTION

A substantial and diverse body of literature suggests that the pathophysiology of schizophrenia is related to dysregulation of bioenergetic function. Perturbations of various metabolic pathways across different brain regions have been reported in relation to the symptomology of schizophrenia. While a whole body of literature exists regarding the changes in several bioenergetic subprocesses, changes in glucose metabolism and mitochondrial oxidative phosphorylation are the best documented in the postmortem brain, especially in the DLPFC.

Glycolysis

Glucose enters neurons and astrocytes via glucose transporters (GLUTs). Decreases in the transcripts of GLUT1 and GLUT3 in pyramidal DLPFC neurons of schizophrenia subjects have been reported. In these cells, glucose is phosphorylated by hexokinase. In schizophrenia, the cytosolic to mitochondrial hexokinase ratio is increased in the DLPFC, suggesting an uncoupling of bioenergetic processes that produce ATP. Consistent with these findings, transcript expression of other regulatory glycolytic enzymes, namely hexokinase, phosphofructokinase, and glucophosphoisomerase, is decreased in schizophrenia. Hexokinase activity was also decreased, which suggests a decrease in glycolytic capacity. Evidence suggests that this finding was not secondary to antipsychotic treatment.

Beyond these biochemical findings, genes for key regulatory glycolytic enzymes, such as the rate-limiting enzyme complex, phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFKFB), hexokinase, and pyruvate kinase, fall within schizophrenia “risk” genomic regions in individuals of European decent.

Tricarboxylic Acid Cycle

The tricarboxylic acid (TCA) cycle is the final shared pathway for the aerobic oxidation of molecules to produce energy in cells. These enzymes have been extensively studied in the schizophrenia brain.

Transcriptomic studies have consistently reported significant decreases in the expression of TCA cycle genes in the DLPFC in schizophrenia. A report comparing the DLPFC in 13 pairs of schizophrenia and healthy controls showed alterations of several critical enzymes in the TCA cycles. The activities of enzymes in the first half of the TCA cycle (namely, aconitase, ketoglutarate dehydrogenase complex, and succinate thiokinase) were reduced in schizophrenia subjects. In contrast, enzymes in the second half of the cycle (succinate dehydrogenase and malate dehydrogenase [MDH]) were increased. Interestingly, pyruvate dehydrogenase complex (PDHC), citrate synthase (CS), isocitrate dehydrogenase (ICDH), and fumarase activities remained the same. Aside from the DLPFC, protein expression of the β-subunit of the PDHC was decreased in the schizophrenia striatum.

Oxidative Phosphorylation

Numerous studies suggest that oxidative phosphorylation is dysregulated in schizophrenia. Decreases in transcript levels of electron transport chain (ETC) complexes and ATP-synthase have been found in the PFC of subjects with schizophrenia. Concordant with these findings, decreases in transcripts for mitochondrial pathways were also reported in DLPFC layers III and V in schizophrenia isolated by laser capture microdissection. Proteomic analyses by Liquid Chromatography with Tandem Mass Spectrometry (LC-MS/MS) found a decrease in complexes I and III of the ETC. However, the activity of complex IV, but not complex I, is changed in the schizophrenia DLPFC. However, perturbation of ETC complexes has been reported in other brain regions. Complex I and III activities are reduced in the temporal cortices and in the basal ganglia. Complex IV activity was also reduced in the temporal cortices. Finally, in the dentate gyrus of the hippocampus, transcript expression of ATP synthase, cytochrome c, succinate dehydrogenase, and reduced nicotinamide adenine dinucleotide (NADH) dehydrogenase are all decreased.

Mitochondria

Abnormalities in the mitochondria and mitochondrial function are commonly reported in schizophrenia research across various substrates: postmortem brain, cell lines, and animal models. While there is a plethora of evidence suggesting dysfunction of both the TCA and oxidative phosphorylation, which take place in the mitochondria, there is variability in reporting morphologic and structural mitochondrial data. Due to the breadth and the diversity of the methodologies, brain regions studied, cell lines used, and subject diversity, reconciling the findings to draw a definitive conclusion is a challenge to the field.

In general, there appears to be a decrease in the number and the size of mitochondria across brain regions and cell types (both neurons and glia). However, this evidence is challenged by studies that have reported equal mitochondria, size, and density in the schizophrenia brain compared to controls.

Conclusion

The evidence for bioenergetic dysfunction in schizophrenia is accumulating. Further work is required to delineate the cellular specificity (i.e., interneuron, pyramidal neuron, glial, or global deficits) or the regional changes (i.e., frontal cortex, striatum, temporal lobes, etc.) of these perturbations and characterize the causative metabolic pathways (i.e., glycolysis, oxidative phosphorylation, or other metabolic pathways, etc.). How these deficits manifest in the behavioral and symptomology of schizophrenia has yet to be fully detailed.

SCHIZOPHRENIA AS A NEURODEVELOPMENTAL DISORDER

The development of the brain is a complex process that involves an intricate sequence of events marked by sensitive or critical time periods when neural circuits are maximally plastic. This plasticity allows essential developmental changes to occur but also confers vulnerability. Disruptions to essential neurodevelopmental processes can cause permanent alterations in brain circuitry and behavior. Schizophrenia demonstrates approximately 79% heritability, while proband concordance rates are reported as 33% for monozygotic twins and 7% for dizygotic twins. Although the etiology of schizophrenia remains unknown, these data suggest that most schizophrenia cases are the result of neurodevelopmental interactions between genetic predispositions and environmental factors.

Genetic Predisposition and Environmental Factors

Schizophrenia is a highly heritable and polygenic illness. Genome-wide association studies (GWAS) and copy number variation (CNV) studies have implicated many common alleles of small effect size as well as rare CNVs of large effect size in the development of schizophrenia. Polygenic risk scores (PRS) aggregate the risk scores of gene variants to estimate overall genetic predisposition. Expression profiles for some of these putative risk genes, including neuregulin 1 (*NRG1*), dystrobrevin-binding protein 1 (*DTNBP1*), disrupted in schizophrenia 1 scaffold protein (*DISC1*), and regulator of G-protein signaling 4 (*RGS4*) have been characterized in human postmortem tissue. The human postmortem brain remains an invaluable resource for schizophrenia research. In addition to gene expression profiles and variations in copy number, human postmortem tissue studies have identified single nucleotide polymorphisms (SNPs) risk variants for schizophrenia, such as rs1344706 (A/C) in the zinc finger protein 804A gene (*ZNF804A).* In addition to risk estimates, genetic studies are also beginning to inform schizophrenia diagnosis and treatment. By leveraging genetic findings with biomarker identification, neurobiologic strata are beginning to emerge. Understanding neurodevelopmental abnormalities within the context of neurobiologic subtypes (biotypes) provides insight into schizophrenia as a neurodevelopmental disorder driven, in part, by patient-specific genetic variations. Environmental contributions are also well described. Socioeconomic factors, fetal and perinatal infections, nutritional deficiencies, birth complications, and cannabis exposure are epidemiologically associated with neurodevelopmental abnormalities and schizophrenia. The development of the brain is a protracted process. As such, environmental factors span childhood, adolescence, and adulthood.

Functionally, schizophrenia symptoms derive from neurotransmitter imbalances within the brain. Genetic studies suggest that inherited or de novo mutations associated with synaptic dysfunction play an etiologic role in the disease mechanisms of schizophrenia. Many recent efforts to identify genetic risk factors for schizophrenia converge upon the salience of excitatory phenotypes.

Epigenetic Factors

Epigenetic processes dynamically regulate gene expression independently of DNA sequence variation and can regulate key neurobiologic processes. Modification of the chromatin structure of DNA (chromatin remodeling) is associated with reversible modifications of DNA rather than with inherited variability within the DNA sequences and is, therefore, called epigenetic modification. This important process is tightly regulated by histones that control how tightly DNA is bound and whether the copy machinery will have access to individual genes. Epigenetic regulation of chromatin can occur via several mechanisms such as DNA methylation and posttranslational modifications of histones. There appear to be periods during development when epigenomic changes take place in the human brain. For example, DNA methylation patterns show a unique temporal pattern across the lifespan, with the greatest changes occurring during the prenatal period, slowing after birth, and gradual slowing with aging. DNA methyltransferases (*DNMTs*), which mediate DNA methylation, are highly expressed in the brain and are overexpressed in GABAergic neurons in the PFC of schizophrenia subjects. A differential DNA methylation profile has been reported in leukocytes and postmortem brains in schizophrenia. One genome-wide methylation study using human postmortem schizophrenia tissue identified 2,929 unique genes that were differentially methylated, many of which were also identified in prior studies using leukocytes. Environmental factors can also influence epigenetic changes. Maternal malnutrition, certain intrauterine conditions, and viral infections can lead to hypermethylation of DNA. The epigenome might serve as a substrate that links environmental exposures, genetic variants, and psychopathology.

Brain Regions

Since the first study over a decade ago, several transcriptome studies have been conducted using postmortem tissue. Gene expression changes are found in several brain regions, including the PFC, hippocampus, and temporal cortices. Correlation networks are increasingly being used in bioinformatics applications. Weighted gene co-expression network analysis (WGCNA) is a method used to describe the correlation patterns among genes across microarray samples. This approach, successfully used in other biologic contexts such as cancer, can be used to find clusters (modules) of highly correlated genes and to relate modules to one another. Pathways identified in the schizophrenia PFC include oxidative phosphorylation, myelination, synaptic transmission, and immune function, while data from the hippocampus show deficits in immune response and inflammation, transcriptional regulation, and epigenetic control in PV-containing interneurons.

Neuronal Migration and Neurogenesis

During the second trimester of fetal development, neurons migrate upward from the ventricular wall to their target cortical layer. Initial postmortem studies in cortical tissue from individuals with schizophrenia found ectopic neurons and abnormal cytoarchitecture in the PFC and entorhinal cortex. These data were consistent with an impairment of neuronal migration of these cells into the cortex during their critical developmental period (second trimester) in schizophrenia and have been interpreted as evidence of schizophrenia as a developmental disorder. While intriguing, precise and carefully designed follow-up studies have not been able to replicate or confirm these initial findings. With recent advances in the field of adult neurogenesis, modern tools have become available that support the study of the proliferation and differentiation of stem cells in the brain. The use of antibodies previously used in cancer research has recently been applied in histologic studies of the postmortem human hippocampus, including Ki-67 and proliferating cell nuclear antigen. There are initial observations that suggest a reduction in cell division in the hippocampus in schizophrenia. Although provocative and interesting, replication studies are ongoing.

Spinogenesis and Pruning

Adult schizophrenia subjects have fewer synaptic connections in specific brain regions, and there is some evidence, although divergent and inconsistent, that synaptic proteins such as synaptophysin (*SYP*), growth-associated protein 43 (*GAP43*), complexin 1/2 (*CPLX1/2*), and synaptosome-associated protein 25 (*SNAP25*) are reduced in these individuals. Prior electron microscopy studies show that cortical synaptic density reaches maximum levels at 2 to 4 years of age—at which time, density is approximately double normal adult levels—with a reduction in the number of synapses (i.e., synaptic pruning) occurring mainly during adolescence. This prompts the suggestion of schizophrenia as a defect of excessive pruning during adolescence. However, data collected from subjects with ages ranging from newborn to 91 years old suggest that synaptic elimination may be a more extended process, continuing throughout the third decade of life before synaptic density stabilizes at adult levels. While the hypothesis of excessive pruning may still hold true, these new studies suggest pruning may occur over a broader time course. Excessive synaptic pruning caused by hyperactivity of the complement system has also been proposed. Another possible explanation for reduced spine density in certain brain regions in schizophrenia could be related to altered development and stabilization of dendritic spines.

Myelination

Myelination is critical for the proper axonal conductance and communication between brain regions. The first 2 years of life see the rapid myelination of axonal fibers that continue into adulthood following a region-specific course. In the PFC, for example, myelination occurs well into the second decade of life. Abnormalities in myelin-related genes are observed in schizophrenia. The protracted development course provides opportunities for genetic and environmental exposures to interfere with normal myelin maturation in the brain.

Neurologic Soft Signs

Neurologic soft signs (NSS) are clinically observable deficits in sensory integration, motor coordination, developmental reflexes, and other central nervous system functions. While not specific to schizophrenia, NSS suggest developmental abnormalities in schizophrenia patients. Subtle physical deviations indicative of abnormal prenatal development also exist. Patients with high CNVs of schizophrenia-associated genes can also present minor physical anomalies.

THE INFLAMMATION HYPOTHESIS

There is gradually accumulating evidence that hints toward the role of inflammation in the pathogenesis of schizophrenia. While the inflammation hypothesis is not new for schizophrenia and psychiatric disorders in general, the evidence now hints at a strong relationship.

The CNS Inflammatory machinery includes cellular components like microglia, astrocytes, and immune cells from the peripheral immune system and active components like C-reactive proteins (CRP) and cytokines. The exact role of inflammatory markers like CRP and cytokines is still unclear. A

2019 epidemiologic study found evidence that CRP has a protective effect against schizophrenia. There is also evidence of an increase in proinflammatory cytokines in patients with psychosis. This is supplemented by evidence of increased functioning and decreased positive symptoms in animal models. Finally, some of the SNPs associated with the highest risk for schizophrenia are found in or near genes that are immune system constituents.

The role of inflammation in schizophrenia is rapidly evolving and has significant therapeutic implications. Accounting for inflammation may allow identification of better treatments for schizophrenia.

INTEGRATIVE PATHOPHYSIOLOGY OF SCHIZOPHRENIA

There has been progress over the past decades that has led to landmark discoveries in the pathophysiology and treatment of the disease. There is neither a single brain region nor a single neurochemical alteration, but several associated with schizophrenia.

The PFC, hippocampus, and thalamus are most often implicated, perhaps in part because these are the most studied. At the cellular level, reduced gray matter volumes, reduced size of neurons but without cell loss, and reduced dendritic arborization and spines are the main observations seen in schizophrenia. White matter changes also contribute to connectivity deficits between brain regions implicated in schizophrenia. The true cause remains unknown, but the implication of these findings is one of reduced connectivity in the brain in schizophrenia.

Neurochemical deficits in neurotransmitter systems indicate abnormal synaptic communication between neurons. There may be specific neurotransmitter systems involved in certain cortical regions, with deficits in the GABAergic system being the most consistent. A plausible formulation is that schizophrenia is a disease of abnormal connectivity, one that may occur at the cellular level, synaptic level, or circuit level.

The nature, location, and timing of pathology seem to be key factors in the development of schizophrenia and warrant further exploration. The advent of next-generation sequencing (NGS) allows for the rapid sequencing of entire genomes, transcriptomes, and epigenomes from specific brain regions, specific cell types, and specific time points in development. The collaborative use of modern translational research methodologies and tools will be critical to the successful pursuit of understanding the pathophysiology of schizophrenia.

FURTHER READINGS

Akbarian S. Epigenetic mechanisms in schizophrenia. *Dialogues Clin Neurosci*. 2014;16(3):405–417.

Bakhshi K, Chance SA. The neuropathology of schizophrenia: a selective review of past studies and emerging themes in brain structure and cytoarchitecture. *Neuroscience*. 2015;303:82–102.

Bernstein HG, Steiner J., Guest PC, Dobrowolny H, Bogerts B. Glial cells as key players in schizophrenia pathology: recent insights and concepts of therapy. *Schizophr Res*. 2015;161(1):4–18.

Buckley PF. Neuroinflammation and schizophrenia. *Curr Psychiatry Rep*. 2019;21(8):72.

Catts VS, Fung SJ, Long LE, et al. Rethinking schizophrenia in the context of normal neurodevelopment. *Front Cell Neurosci*. 2013;7:60.

Catts VS, Lai YL, Weickert CS, Weickert TW, Catts SV. A quantitative review of the postmortem evidence for decreased cortical N-methyl-d-aspartate receptor expression levels in schizophrenia: how can we link molecular abnormalities to mismatch negativity deficits? *Biol Psychol*. 2016;116:57–67.

Eastwood SL. The synaptic pathology of schizophrenia: is aberrant neurodevelopment and plasticity to blame? In: Smythies *J, ed. International Review of Neurobiology.* Vol 59. Disorders of Synaptic Plasticity and Schizophrenia. Academic Press; 2004:47–72.

Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature*. 2014;506(7487):179–184.

Hagen CM, Gonçalves VF, Hedley PL, et al. Schizophrenia-associated mt-DNA SNPs exhibit highly variable haplogroup affiliation and nuclear ancestry: bi-genomic dependence raises major concerns for link to disease. *PLoS One*. 2018;13(12):e0208828.

Hahn CG, Banerjee A, MacDonald ML, et al. The Post-synaptic density of human postmortem brain tissues: an experimental study paradigm for neuropsychiatric illnesses. *PLoS One*. 2009;4(4):e5251.

Harrison PJ. The hippocampus in schizophrenia: a review of the neuropathological evidence and its pathophysiological implications. *Psychopharmacology (Berl)*. 2004;174(1):151–162.

Hopper S, Pavey GM, Gogos A, Dean B. Widespread changes in positive allosteric modulation of the muscarinic M1 receptor in some participants with schizophrenia. *Int J Neuropsychopharmacol*. 2019;22(10):640–650.

Jaaro-Peled H, Sawa A. Neurodevelopmental factors in schizophrenia. *Psychiatr Clin North Am*. 2020;43(2):263–274.

Katsel P, Davis KL, Haroutunian V. Variations in myelin and oligodendrocyte-related gene expression across multiple brain regions in schizophrenia: a gene ontology study. *Schizophr Res*. 2005;79(2):157–173.

Kim Y, Vadodaria KC, Lenkei Z, et al. Mitochondria, metabolism, and redox mechanisms in psychiatric disorders. *Antioxid Redox Signal*. 2019;31(4):275–317.

Kleinman JE, Law AJ, Lipska BK, et al. Genetic neuropathology of schizophrenia: new approaches to an old question and new uses for postmortem human brains. *Biol Psychiatry*. 2011;69(2):140–145.

Konopaske GT, Lange N, Coyle JT, Benes FM. Prefrontal cortical dendritic spine pathology in schizophrenia and bipolar disorder. *JAMA Psychiatry*. 2014;71(12):1323–1331.

Lin BD, Alkema A, Peters T, et al. Assessing causal links between metabolic traits, inflammation and schizophrenia: a univariable and multivariable, bidirectional Mendelian-randomization study. *Int J Epidemiol*. 2019;48(5):1505–1514.

Meador-Woodruff JH, Healy DJ. Glutamate receptor expression in schizophrenic brain. *Brain Res Brain Res Rev*. 2000;31(2):288–294.

Müller N. Inflammation in schizophrenia: pathogenetic aspects and therapeutic considerations. *Schizophr Bull*. 2018;44(5):973–982.

Perez JM, Berto S, Gleason K, et al. Hippocampal subfield transcriptome analysis in schizophrenia psychosis. *Mol Psychiatry*. 2021;26(6):2577–2589.

Pidsley R, Mill J. Epigenetic studies of psychosis: current findings, methodological approaches, and implications for postmortem research. *Biol Psychiatry*. 2011;69(2):146–156.

Schmidt MJ, Mirnics K. Neurodevelopment, GABA system dysfunction, and schizophrenia. *Neuropsychopharmacology*. 2015;40(1):190–206.

Selemon LD, Goldman-Rakic PS. The reduced neuropil hypothesis: a circuit based model of schizophrenia. *Biol Psychiatry*. 1999;45(1):17–25.

Sequeira PA, Martin MV, Vawter MP. The first decade and beyond of transcriptional profiling in schizophrenia. *Neurobiol Dis*. 2012;45(1):23–36.

Serati M, Delvecchio G, Orsenigo G, et al. The role of the subplate in schizophrenia and autism: a systematic review. *Neuroscience*. 2019;408:58–67.

Tarazi FI, Zhang K, Baldessarini RJ. Dopamine D4 receptors: beyond schizophrenia. *J Recept Signal Transduct Res*. 2004;24(3):131–147.

Wockner LF, Noble EP, Lawford BR, et al. Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. *Transl Psychiatry*. 2014;4(1):e339.

12.7 Structural Brain Imaging in Schizophrenia

Amanda E. Lyall, Ph.D., Marek Kubicki, M.D., Ph.D., and Martha E. Shenton, Ph.D.

THE BRAIN AND SCHIZOPHRENIA: HISTORICAL PERSPECTIVE

After more than 100 years of research, the neuropathology of schizophrenia still remains unknown. This is despite the fact that both Eugen Bleuler and Emil Kraepelin, who first described “dementia praecox” and “the schizophrenias,” postulated that brain abnormalities would ultimately be linked to the etiology of schizophrenia. The conviction that serious psychiatric illness was likely the result of an unhealthy brain was further fueled by critically important inroads into Huntington chorea, some forms of epilepsy, Pick’s disease, tertiary syphilis, and Alzheimer’s disease that were being made during this same time period. Alois Alzheimer, in fact, was one of the first to investigate the neuropathology of schizophrenia, although he later went on to investigate other illnesses that he believed were more tractable brain disorders.

Researchers at the end of the 19th and beginning of the 20th centuries were nevertheless hopeful that they would succeed in identifying neuropathologic features of schizophrenia. Results from postmortem studies, however, were ultimately disappointing, often leading to conflicting findings due in part to the crude measurement tools available and, in part, due to the high expectations, researchers had that they would observe more remarkable abnormalities. This state of affairs led to a waning interest in the neuropathology of schizophrenia, and although a small number of researchers continued to investigate the brain and its role in schizophrenia, the general consensus, as clearly articulated by the eminent neurologist Fred Plum, was that “schizophrenia is a graveyard of neuropathologists.” Consequently, research investigating brain abnormalities using histologic methods in schizophrenia lost favor.

Early Applications of Neuroimaging to the Study of Schizophrenia

It was not until 1976 that interest in investigating brain abnormalities in schizophrenia was renewed with the first computer-assisted tomography (CT) study of schizophrenia by Johnstone and colleagues. These investigators observed enlarged lateral ventricles in individuals with schizophrenia compared to healthy control (HC) subjects. This single study led to a proliferation of CT studies as well as a reinvigoration of postmortem studies investigating schizophrenia, with the overwhelming majority confirming the original finding of enlarged lateral ventricles. This early study was the impetus for later investigations of structural brain anomalies in schizophrenia where the goal was to identify markers of pathologic processes in the brain. This interest has continued to the present day.

Following the introduction of CT, nuclear magnetic resonance (NMR) was developed in the 1970s, based on the seminal work of Peter Mansfield and Paul Lauterbur. These two scientists shared the Nobel Prize in Medicine in 2003 for their independent contributions to the theory that led to the development of magnetic resonance imaging (MRI). MRI was then introduced and soon became the most powerful tool for visualizing soft tissue contrast in the brain (e.g., gray and white matter). Moreover, and in contrast to CT that uses ionizing radiation x-rays, MRI uses magnetic field and radio waves, which have no known adverse effects, and only a small number of contraindications. MRI can thus be used as a research tool in at-risk populations, such as children and pregnant women, and it can be repeated as needed, thereby opening the door to longitudinal studies to investigate brain changes over time, or to monitor more precisely treatment efficacy.

MR scanners were first built and clinically used in the 1980s, and today they are used routinely to visualize the anatomy and morphology of the body and the brain. The use of MRI in clinical research made it possible to evaluate specific brain regions *in vivo.* The promise of what MRI offered the field led to a proliferation of studies in schizophrenia. MRI studies conducted between 1988 and 2000 led to compelling evidence that brain abnormalities were present in individuals with schizophrenia. These studies also observed, not surprisingly, that most abnormalities were small and subtle, rather than large and dramatic, suggesting still further that more advanced and accurate measurements were needed to detect and understand such brain abnormalities. As will be detailed below, a number of novel imaging technologies have since become available that make possible even more precise identification and characterization of subtle structural abnormalities in schizophrenia.

The time period between 1988 and 2000 also witnessed a burgeoning of neuroimaging studies in schizophrenia “that have led to more definitive findings of brain abnormalities than in any other time period in the history of schizophrenia research” (see review by Shenton and colleagues in 2001). Combined with the wider application of MRI methodologies, there were also unparalleled advances in MR image acquisition protocols. For example, image acquisition paradigms evolved from evaluating a single 1-cm slice in the brain on MR scanners with low field strength (0.5 Tesla [T]) to acquiring 1-mm or thinner slices throughout the entire brain on MR scanners with higher field strengths (specifically 1.5 T and 3 T). This change in technology made it possible to evaluate the entire brain using thin, contiguous slices, thereby resulting in greater anatomical detail. Developments of novel sequences, such as those utilized for diffusion magnetic resonance imaging (dMRI), a technique most known for its ability to indirectly interrogate white matter microstructure through the detection and measurement of water diffusion in the brain, were significant accomplishments that furthered the imaging field. Likewise, the subsequent expansion of cutting-edge image analysis tools for both MRI and dMRI acquisitions led to an unprecedented number of novel findings in schizophrenia. For example, the introduction of semiautomated segmentation algorithms to classify different brain tissues into gray matter, white matter, and cerebrospinal fluid (CSF) and the development of newer automated atlas-based parcellation methods have made it possible to conduct larger studies with increased accuracy, greater speed, and higher reliability when compared to previous studies that employed manual tracing methods (i.e., manual region of interest-ROI methods). Figure 12.7–1 shows a coronal slice through the brain where the temporal lobe, amygdala, parahippocampal gyrus, and superior temporal gyrus have been manually drawn. Figure 12.7–2 shows a lateral view of the brain where the surface shows the different lobes of the brain. Such parcellations are now commonly done using automated procedures, with manual editing only being conducted when necessary.

Recent advances in imaging technology during this period also made it possible to associate gray matter and white matter abnormalities with genetic, environmental, behavioral, and neuropsychological measures in those suffering from schizophrenia. Significant efforts are also being made to link multimodal imaging techniques (e.g., structural MRI with functional MRI) within the same cohort of subjects so as to explore more fully the entire brain and its role in schizophrenia. It has also become increasingly clear that inroads to understanding the biologic underpinnings and developmental progression of this disorder are contingent upon more methodologically rigorous studies that will introduce measures with greater sensitivity for detecting changes in both brain structure and function. In summary, the 21st century has been and will continue to be an exciting time for brain imaging studies in schizophrenia, as advances in imaging technology over the past 30 years make new insights into the brain and its processes possible.

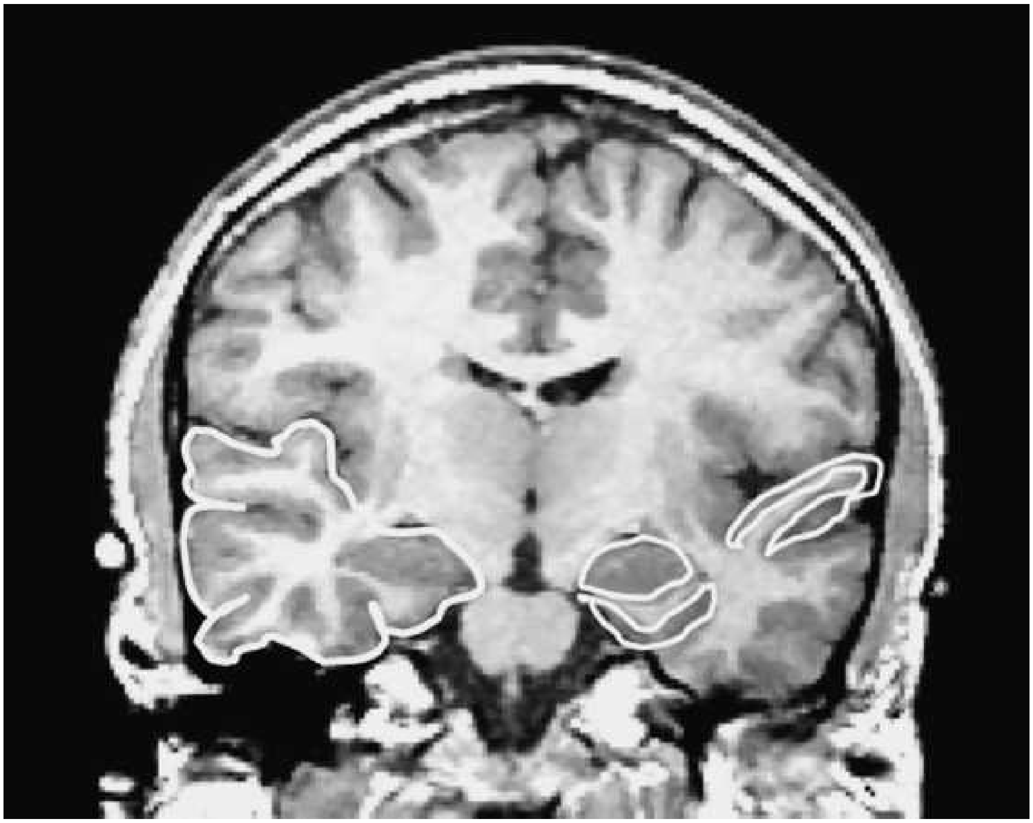

FIGURE 12.7–1. Coronal 1.5-mm slice that shows the medial temporal lobe region of the brain, with the temporal lobe delineated in *white* on the **left** side of the image (subject's **right**). Regions delineated in *white* on the right side of the brain (subject's **left**) are the superior temporal gyrus, which borders the Sylvian fissure, the amygdala (the *almond-shaped object*), and the parahippocampal gyrus, inferior to the amygdala. (From Shenton ME, Kikinis R, Jolesz FA, Pollak SD, LeMay M: Abnormalities of the Left Temporal Lobe and Thought Disorder in Schizophrenia—A Quantitative Magnetic Resonance Imaging Study. *N Engl J Med.* 1992;327:604–612. Copyright © 1992 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)

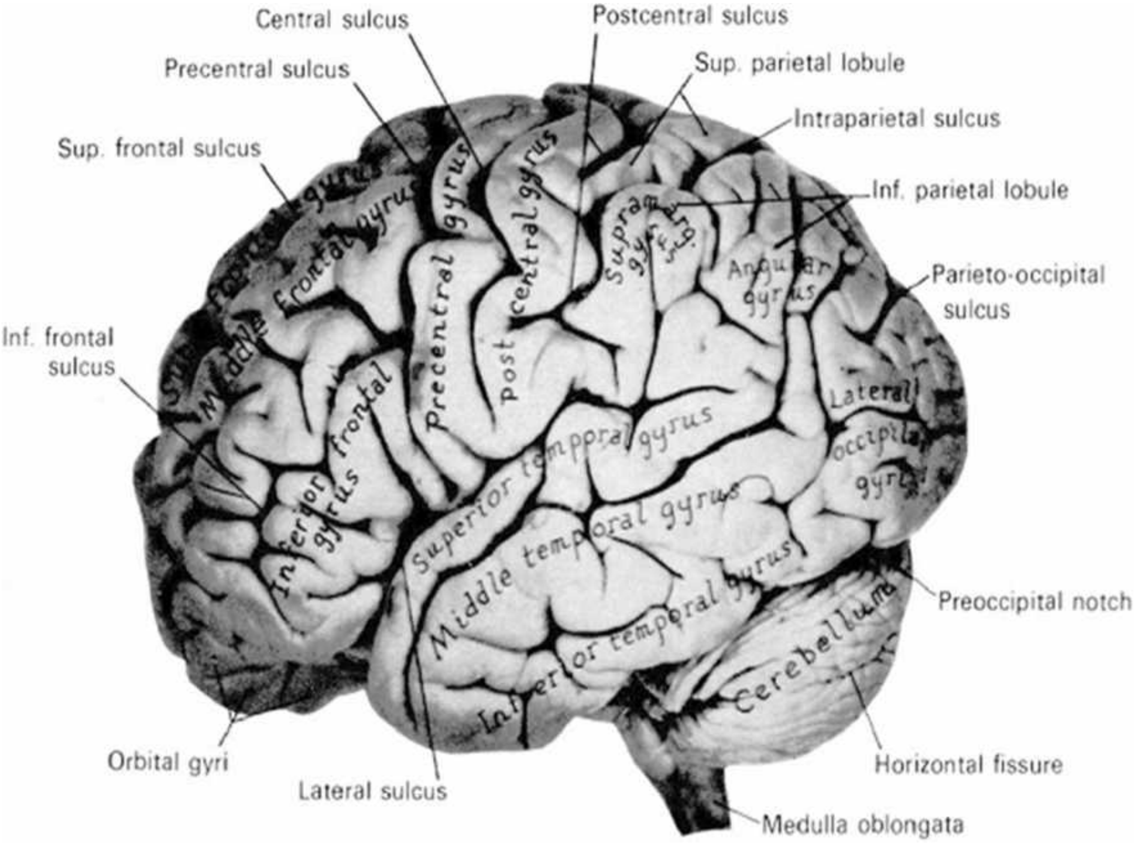
Central sulcus
Postcentral sulcus
Precentral sulcus
Sup. parietal lobule
Intraparietal sulcus
Sup. frontal sulcus
Inf. parietal lobule
Parieto-occipital sulcus
Inf. frontal sulcus
Superior frontal gyrus
Middle frontal gyrus
Precentral gyrus
Postcentral gyrus
Supramarginal gyrus
Angular gyrus
Lateral occipital gyri
Inferior frontal gyrus
Superior temporal gyrus
Middle temporal gyrus
Inferior temporal gyrus
Cerebellum
Preoccipital notch
Orbital gyri
Lateral sulcus
Horizontal fissure
Medulla oblongata

FIGURE 12.7–2. Lateral view of the human brain. (Reprinted with permission from Carpenter MB. *Human Neuroanatomy*. 8th ed, Williams & Wilkins; 1983.)

Advances Inspired by Seminal Studies and the Path Forward

As described above, advances in imaging were critical for confirming the presence of structural brain abnormalities in schizophrenia. These early seminal works inspired new studies that include (1) more homogeneous patient groups to explore relationships between structural abnormalities and clinical symptoms; (2) at-risk populations, based on increased clinical or genetic high risk; (3) applications of newer, more sensitive acquisition and analysis methods, such as dMRI to investigate brain microstructure and connectivity; and, (4) multimodal imaging to investigate brain structure and function within the same patient (e.g., fMRI, dMRI, or magnetic resonance spectroscopy [MRS], to name just a few). It is also clear that future investigations are needed to elucidate the natural history of when brain abnormalities first appear and how they may change over time. This includes a shift in the types of clinical populations investigated, with the field showing an increase in the number of studies of first-episode individuals and of those individuals with an increased clinical or genetic high risk for schizophrenia and other psychiatric illnesses, as opposed to studying individuals with more advanced illness.

The first studies of schizophrenia, as noted earlier, used relatively crude neuroimaging techniques by today's standards. Nonetheless, and as noted above, seminal early studies confirmed that schizophrenia is, in fact, "a brain disease." In 2001, Shenton and colleagues reviewed 193 peer-reviewed sMRI studies of schizophrenia conducted between 1988 and 2000. This review confirmed a number of findings described earlier in CT and postmortem studies. For example, 80% of the studies reviewed reported enlarged lateral ventricles in patients with schizophrenia compared to controls. This was a common finding in early CT studies and has proven to be the most replicated finding in schizophrenia. The second most replicated finding in this review was deficits in the medial temporal lobe, where 74% of studies reported volumetric reductions in medial temporal lobe structures, most notably in the amygdala, hippocampus, parahippocampal gyrus, and proximal neocortical areas, such as the superior temporal gyrus. Other abnormalities included cavum septi pellucidi (92%), third ventricle enlargement (73% of studies), parietal lobe volume reduction (67%), frontal lobe volume reduction (59%), occipital lobe volume reduction (44%), as well as thalamus (42%) and cerebellar volume reduction (31%). Basal ganglia showed abnormalities (68%, however, some increased volume, some decreased volume), as did the corpus callosum (63% showed abnormalities). Work following this time period has led to a further understanding of brain structural abnormalities in schizophrenia, through the incorporation of newer tools, more accurate measurements, and information gleaned from other imaging modalities. As will be made clear in our selective review below, a further appreciation of the importance of how brain circuitry abnormalities and alterations to brain structure occur, and how these aberrations relate to both the development of specific cognitive deficits and the manifestation of the clinical symptoms of schizophrenia, are topics of interest in current research.

The vast majority of studies today are no longer investigating whether schizophrenia is "a brain disease" but are asking instead: "What sort of disease is it?" Researchers are also more interested in posing questions such as: "Are there disease-specific structural abnormalities in schizophrenia compared to other disorders with psychotic features? At what point in brain development do brain abnormalities first appear? Do these brain abnormalities change over time? If brain abnormalities change over time, do they have a neurodevelopmental origin, are they the result of neurodegenerative processes, or both? Will the early treatment be effective at mitigating or preventing the progression of the disorder? What impact does medication/polypharmacy have on the brain in the short and/or long term? Could early treatment help attenuate cognitive decline? Is schizophrenia truly one illness or a spectrum?" A focus on answering these questions will bring us to a closer understanding of the nature and time course of schizophrenia. One such recent initiative is the Accelerated Medicine Partnership in Schizophrenia (AMP SCZ), "a public-private partnership between the National Institutes of Health (NIH), the U.S. Food and Drug Administration (FDA), the European Medicines Agency, and multiple public and private organizations. The overall aim of the AMP Schizophrenia (SCZ) initiative is to generate tools that will considerably improve success in developing early-stage interventions for patients who are at risk of developing schizophrenia" (https://www.nimh.nih.gov/research/research-funded-by-nimh/research-initiatives/accelerating-medicines-partnership-schizophrenia-amp-scz).

OVERVIEW OF CHAPTER

The goal of this chapter is to review progress in structural imaging findings from studies using T1-based structural MRI (sMRI) and dMRI techniques that further our understanding of schizophrenia. In this chapter, we provide a selective review of what we believe are among the most impactful publications over the last 10 years. We exclude case studies, studies with no control group, studies with small sample sizes ($n < 20$), and studies investigating schizotypal personality disorder or childhood-onset schizophrenia. We include important review articles that appropriately and accurately summarize the state of the field for a given area. In addition, we highlight studies that, in our opinion, will provide critical insights for future neuroimaging investigations into the underlying neuropathophysiology of schizophrenia.

In this chapter, the first section presents findings from studies that focus only on the use of sMRI methodologies, such as cortical gray matter volume and cortical thickness, followed by a section where we discuss studies that have utilized dMRI approaches, such as diffusion tensor imaging (DTI) and tractography. For both imaging modalities, we first present studies where individuals with schizophrenia who are at different stages of illness are collapsed into one larger analysis in an effort to increase sample size and to identify consistent abnormalities that have been found in patients with schizophrenia compared to healthy controls (HCs) regardless of illness stage. After this, we present studies which focus on discrete illness stages, first describing studies which center on individuals with chronic schizophrenia, then with studies on early course schizophrenia, and then a summary of studies carried out in individuals at risk for developing schizophrenia. To conclude each section, we also present studies that represent cutting-edge advances for each modality. At the end of this chapter, we examine the findings and then provide recommendations for future directions and considerations for brain imaging research in schizophrenia that will bring us closer to understanding, treating, and preventing this serious mental illness.

Structural Magnetic Resonance Imaging Findings in Schizophrenia

As discussed previously, the number of structural neuroimaging studies of schizophrenia has continued to increase at an exponential rate since the early 90s. While initial sMRI studies tended to measure values like cortical volume in small, well-defined populations, advances in computational power and image analysis techniques have provided a wider range of faster, more automated, and more precise measurement tools. This has led to large, multisite prospective and retrospective studies investigating morphometric brain changes in not tens, but thousands of individuals with schizophrenia. These “big data” approaches, which will be discussed at length later in the chapter, are represented by studies from the Enhanced NeuroImaging Genetic through Meta-Analysis (ENIGMA) Consortium. Of further note, several recent meta-analyses from the ENIGMA Consortium confirm widespread gray and white matter volumetric reductions in schizophrenia that were summarized in the earlier review in 2001 by Shenton and colleagues.

In one study, van Erp and colleagues utilized data from 2,028 individuals with schizophrenia and 2,540 HCs to investigate volumetric differences in subcortical structures. In this meta-analysis, those with schizophrenia exhibited reductions in the volume of the hippocampus, amygdala, thalamus, and nucleus accumbens. This study also confirmed prior reports that individuals with schizophrenia have reductions in overall intracranial volume. These individuals also showed increased volumes in the globus pallidus and the lateral ventricles. The majority of these differences exhibited effect sizes ranging from small to medium, with the largest effect being observed in the hippocampus (Cohen’s d = −0.46). In a separate 2018 meta-analysis, van Erp and colleagues applied the same approach to cortical thickness and surface area data from 4,474 subjects with schizophrenia and 5,089 HCs. In this study, the authors observed global reductions in cortical thickness and surface area, though to a lesser degree, in both hemispheres in schizophrenia. After taking into account global cortical thickness and surface area losses, it was found that those individuals showed additional regional changes in cortical thickness (but not the cortical area) of the fusiform gyrus and several temporal regions, in addition to thickness reductions in the ventrolateral prefrontal cortex, insula, and anterior cingulate cortex. Interestingly, those with schizophrenia also showed areas of thicker cortex in the motor, somatosensory, parietal, and rostral anterior cingulate cortex compared to healthy individuals.

Taken together, these findings confirm the presence of subtle but widespread case-control differences in gray matter in the brain in individuals diagnosed with schizophrenia. Smaller intracranial volumes and larger lateral ventricle volumes thus continue to be core features of the illness. Small but replicable reductions in the volume of subcortical structures, such as the hippocampus and thalamus, also continue to be observed. Cortical tissue reduction, compared with HCs, also appears to be more dramatic in frontal and temporal cortices as well as in several limbic areas, specifically the insula and the anterior cingulate cortex. Aberrant cortical thickening appears to be found in several areas, including the primary cortex, such as motor and sensory areas, as well as in the parietal cortex.

The studies reviewed above, however, do not provide a dynamic window into the brain changes along the schizophrenia time course. In the coming section, we introduce several recent studies which have provided insight into stage-specific volumetric changes in schizophrenia.

Chronic Schizophrenia. Chronic schizophrenia represents the most severe end of the clinical spectrum. Most of these individuals have received medications over a long period of time, and have been living with a chronic illness, both factors that may confound structural imaging findings. More recent studies have attempted to disentangle the effects of chronicity, medication, and aging on findings in chronic schizophrenia populations. In addition, in the last decade, there has been support for the idea of “accelerated aging” in schizophrenia. This hypothesis supports the idea that tissue reduction observed in chronic schizophrenia may be the result of exaggerated age-related decline. One example of this approach is a study by Cropley and colleagues, who used sMRI data of 326 subjects with schizophrenia (average duration of illness 15.9 years) and 197 age-matched HCs. They found that those with schizophrenia exhibit large differences in total gray matter volume when compared to HCs, which begins early in adulthood and extends into middle age. More importantly, the rate of gray matter volume loss was significantly greater compared to matched controls until middle age, with patients between the ages of 20 and 45 showing a 30% to 60% faster rate of gray matter volume loss. Interestingly, after the age of 50, the rate of loss begins to slow, at which time those with schizophrenia appear to age similarly to HCs. When investigating region specific, as opposed to global differences, the authors reported the fastest rate of volume loss in the frontal, temporal, occipital, and cingulate cortices in those individuals who were between 30 and 45 years of age. The authors also found that gray matter volume loss tended to predate pathologic changes in white matter (to be discussed later). Taken together, the results of this study provide evidence for the presence of an accelerated rate of gray matter loss in schizophrenia that the authors argue would normally be associated with healthy aging but at an earlier age.

Another key area of recent interest in chronic schizophrenia research is the investigation of the long-term impact of antipsychotic medication on brain structure. While this is still an area of active debate, recent work by Liu and colleagues provides some key insights. In this study, the authors utilized sMRI data from 35 never-treated individuals with chronic schizophrenia (average duration of illness 20 years), 20 chronic risperidone monotherapy individuals, 20 chronic clozapine monotherapy individuals (both matched to the never treated group for duration of illness), and 55 HCs. In their study, the authors found that compared to controls, both the never-treated and treated individuals exhibited significant gray matter volume loss in the bilateral rostral middle frontal and medial frontal cortex, as well as reductions in the left fusiform gyrus, left pericalcarine, and left banks of the superior temporal sulcus. Of further note, increased gray matter volume was found in the cuneus in both never-treated and treated individuals, but the never-treated individuals also showed an increase in both left and right frontal poles. When comparing the never-treated group to the treated groups, it was revealed that the treated group showed volume reductions in bilateral frontal, prefrontal, and temporal areas, as well as in the left inferior parietal lobule. When comparing individuals receiving risperidone monotherapy to those receiving clozapine monotherapy, it was observed that clozapine individuals had more severe reductions in gray matter volume, predominantly in the prefrontal cortex. In addition, gray matter reductions were shown to be negatively correlated with medication dosage as well as with the duration of illness and symptom severity in treated individuals, while changes in gray matter were not found to be related to clinical symptoms or duration of illness in the never-treated individuals.

In summary, while all individuals diagnosed with schizophrenia, whether treated or not, exhibit significant gray matter loss compared to HCs, this study showed that those who were treated showed further gray matter loss compared to their never-treated counterparts. As such, the work of Liu and colleagues provides preliminary evidence that cumulative medication exposure may be a potential contributing factor that influences the progression of gray matter deficits in chronic patients, which is consistent with previous animal work, and could partially explain cortical changes observed in the “accelerated aging” experiments described earlier in this section.

Moving forward, research in chronic schizophrenia may benefit from a greater focus on underrepresented groups, such as individuals who are treatment resistant or with a late onset of illness, as this may help to further our knowledge regarding factors that contribute to pathologic features observed in the brain in those in the chronic stages of schizophrenia.

First-Episode and Early-Course Schizophrenia Patients. Since the early 2000s, neuroimaging studies in schizophrenia have shifted from the study of chronic, clinically stable, and medicated individuals to investigating those who are in the early stages of illness, particularly after the onset of the first psychotic episode. Such studies are critical for differentiating between primary and secondary markers of illness. In these studies, confounds commonly observed in chronic populations, such as medication, are minimal, thereby making it possible for researchers to understand better the location or nature of structural abnormalities that are either contributing or present at the transition to illness. These studies tend to focus on those who experienced their first psychotic episode within the last 5 years. Recent work that focuses on the recruitment and longitudinal monitoring of antipsychotic naive subjects has also been critical to understanding the impact of medication on the brain in the early stages of the illness.

One of the seminal publications in clinical neuroimaging is a study carried out by Andreasen and colleagues in 2011, the Iowa Longitudinal Study. This groundbreaking effort was a prospective study of 543 first-episode psychosis subjects who were followed for up to 18 years. The authors included a total of 952 scans from 202 individuals with schizophrenia and

125 HCs to determine whether or not progressive brain changes occur after the onset of the illness. Each participant in this study had a minimum of two scans and a maximum of five scans over a 15-year period. The authors observed significant gray and white matter tissue loss in these individuals compared to controls. While the amount of tissue loss was relatively small, these losses occurred predominantly during the first 2 years after illness onset for nearly all gray and white matter measures. Overall, 34% of those diagnosed with the first episode of schizophrenia exhibited reductions in brain volume at a rate of 0.5% a year, with 14% exhibiting reductions in brain volume at a rate of 1% a year. Frontal gray matter and white matter volumes were most affected, though the white matter was more severely impacted. The authors also reported a significant increase in CSF volume in the first-episode subjects that was concurrent with the early gray and white matter tissue loss. When investigating remission status in the first-episode group, throughout the course of this longitudinal study, the authors reported that those first-episode individuals with the smallest increases in CSF were more likely to have longer periods of remission. This study aligns well with a study by Cropley and colleagues, described earlier, which also showed, albeit in a different type of analysis, progressive reductions in gray matter volume in the early stages of illness. As such, there is converging evidence that the greatest amount of tissue loss occurs in the earlier stages of illness, in the first few years after the emergence of frank psychosis. The authors suggest that this may be a consequence of impaired synaptic plasticity, particularly in the late-developing frontal lobe.

While tissue loss is likely to occur in the early stages of illness, it is still unclear what leads to the sudden loss of gray and white matter at this stage of illness. To address this question, in the last 5 to 10 years, there have been several lines of evidence that implicate immunologic processes in the pathophysiology of schizophrenia. These data range from studies demonstrating high levels of proinflammatory cytokines in the blood and CSF of early-course psychosis to postmortem studies demonstrating evidence for the upregulation of genes related to immune system functioning. While it has yet to be determined if immunologic activation is a cause or an effect of the transition to psychosis in the early course of the illness, accumulating evidence suggests its involvement.

Work by Núñez and coauthors provides evidence in support of the hypothesis linking immune activation to gray matter pathologies in the early course of psychosis. In this study, the authors used structural imaging data from 137 first-episode individuals and 81 HCs, as well as blood samples that were collected at a baseline visit or at a 2-month follow-up. The number of neutrophils, one of the subtypes of the white blood cells, was significantly higher in first-episode individuals compared to HCs, whereas the number of erythrocytes (i.e., red blood cells) was significantly lower. In the early-psychosis group, a high neutrophil count was also associated with smaller gray matter volumes and larger CSF volumes. Interestingly, none of the blood cell types were related to white matter volume. These relationships were also not observed in the HCs. The authors then performed a voxel-based morphometry analysis to determine whether or not there were specific gray matter areas that were more susceptible to higher neutrophil counts in early-course psychosis individuals. In fact, reductions in gray matter volume within the right anterior insula, right temporal pole, right entorhinal cortex, right middle temporal gyrus, and right inferior temporal gyrus were significantly correlated with the neutrophil count in first-episode individuals. Interestingly, neutrophils were also directly associated with the total score and the positive score from the Positive and Negative Syndrome Scale (PANSS), a clinical inventory of symptoms for psychotic disorders.

This study thus provided preliminary evidence for immune dysregulation as a potential link to tissue loss in the early stages of illness. Given the role of neutrophils in neuroimmune modulation, this study may provide a foundation for the development of novel treatment strategies focusing on the immune system. It is important to note, however, that the immune dysregulation hypothesis is only one of many theories purported to explain the pathologic tissue loss in schizophrenia. Others include oxidative stress and redox dysregulation, glutamatergic dysregulation, dopaminergic dysregulation, excessive synaptic pruning, aberrant neurodevelopment, and maturation, accelerated aging (discussed earlier), and the compromised functional integration of information due to faulty neurocircuitry, also known as the dysconnectivity hypothesis. Unfortunately, a comprehensive discussion of these models is outside the scope of the present chapter.

In summary, reductions in gray matter are already present at the time of the first psychotic episode. After illness onset, the rate of decline in cortical volumes appears to be faster in those with schizophrenia when compared to healthy individuals, at least for the first few years after disease onset. There is also evidence that gray matter reductions may be linked, at least in a subset of subjects, to immune activation that takes place at the onset of illness. However, this should be interpreted with caution, and future studies are needed to determine the functional consequences of immune activation to determine its role in early-course schizophrenia.

High-Risk Populations. An important question to answer in schizophrenia, as noted previously, is whether brain abnormalities are present prior to the onset of illness (i.e., before psychosis begins). To some extent, this question can be addressed by longitudinal studies that begin prior to illness onset, in what is called the *prodrome period*, or "*attenuated psychosis syndrome,*" that latter term proposed by the DSM-5. A further question to be addressed is whether brain abnormalities observed are also observed in family members without psychosis, which could indicate a potential genetic contribution to observed brain abnormalities independent of the presentation of psychotic symptoms. In this section, we first discuss studies that focus on sMRI findings in individuals exhibiting subthreshold symptoms and are deemed clinically at risk for developing schizophrenia. In the second section, we present a brief summary of sMRI studies focusing on unaffected family members who are deemed genetically at risk for developing schizophrenia.

CLINICAL HIGH RISK. There are several large research consortia around the world that focus on those who are at risk for developing psychosis. Due to the relatively low conversion rate to schizophrenia from those at high risk (less than one-third of those at clinical high risk [CHR] will convert), the number of subjects needed to investigate this population is understandably large. Nonetheless, to date, several studies have successfully investigated high-risk populations, including but not limited to the Edinburgh High-Risk Study (EHRS), the Melbourne Personal Assessment and Crisis Evaluation (PACE) study, the North American Prodrome Longitudinal Study (NAPLS) and the Shanghai At-Risk for Psychosis (SHARP) study. One of the main goals of high-risk studies is to identify consistent differences between individuals who convert compared to those who do not. Similarly, when comparing at-risk individuals to HCs allows for the identification of structural changes related to a more generalized risk of psychopathology, as opposed to the specific risk of developing a psychotic disorder. The long-term goal of identifying these structural alterations is that they may be useful early-stage biomarkers to stratify those most in need of treatment or intervention. Early intervention has, in fact, been shown to have significant efficacy in mitigating the severity of the illness and, in the future, could potentially even prevent onset.

In a publication from the NAPLS consortium in 2015, Cannon and coauthors investigated a cohort of 274 clinical high-risk cases, where 35 converted to psychosis. Follow-up scans were performed 12 months after baseline or after conversion to psychosis. No significant differences were found at baseline for cortical thickness and subcortical or ventricular volumes. However, CHR individuals who converted to psychosis exhibited a significantly faster rate of gray matter loss in right superior frontal, right middle frontal, and right medial frontal gyri, as well as a significantly larger expansion in third ventricular volume compared to nonconverters and HCs. In addition, the authors found that converters with no exposure to antipsychotic medications during the interscan interval actually showed a greater reduction in right prefrontal cortical thickness and greater expansion of third ventricle volume compared to nonconverters (with and without antipsychotic exposure) and HCs. This finding suggests that the differential rate of gray matter loss between converters and nonconverters was not explained by exposure to antipsychotic medications, further suggesting that cortical thinning may be the result of the pathophysiology of the onset of psychosis.

In 2019, Ding and coauthors conducted a meta-analysis of 14 studies that used voxel-based morphometry to study subjects at risk for psychosis (of which the prior study discussed was included), comprised of data from 743 at-risk individuals and 588 HCs. In this meta-analysis, reductions in gray matter volume were found in several frontal regions, including the gyrus rectus, right superior frontal gyrus, and left superior frontal gyrus. Increased gray matter volume was observed in the bilateral cingulate cortex, right fusiform gyrus, left superior temporal gyrus, and right thalamus. The authors also conducted a systematic review of studies that focused on cortical thickness as the measure of interest. Nine studies, comprising a population of 700 at-risk subjects and 469 HCs, were identified. Four studies reported cortical thinning in frontal, temporal, and occipital areas, while three studies found no differences. The remaining two studies reported increased cortical thickness, with one study observing increases in parietal and occipital cortices and the other describing an increase in the right frontal pole (though this finding was observed prior to being corrected for multiple comparisons). Among these nine studies, five included longitudinal scanning paradigms, yet only three employed cortical thickness. When looking at all five studies, individuals who eventually converted to psychosis were reported to have increased gray matter in the right superior, inferior, frontal, and Heschl's gyrus when compared to HCs. Yet, when confined to studies utilizing cortical thickness, two of the three studies showed reduced cortical thickness in the right superior, middle, and medial orbital frontal gyrus and left angular gyrus, precuneus, and temporoparietooccipital regions in those who converted. In the discussion, the authors highlight several key findings from their work, most notably that many of the regions which were identified as being abnormal in high-risk individuals are also commonly reported in first-episode populations as well, specifically the widespread reductions in frontal gray matter. Despite these findings, the authors further commented on the limited number of longitudinal studies that fail to find differences between converters and nonconverters. This lack of consistency in the findings within the high-risk field can be attributed to methodologic differences (e.g., inclusion/exclusion criteria, different pre- and postprocessing methods across samples), although one of the most compelling reasons for such inconsistencies is the heterogeneity in those who are at risk for psychosis, suggesting the importance of stratifying subjects into more homogeneous groupings.

GENETIC AND FAMILIAL HIGH RISK. Studying unaffected family members of individuals with psychosis is also important. First, these individuals have not been hospitalized for a psychiatric disorder and, therefore, are not suffering from the possible putative or neurotoxic effects of psychosis. Second, unaffected family members are generally not medicated with neuroleptic medications, thus removing the potential confound neuroleptics likely have on structural brain findings. Information acquired by investigating unaffected family members thus likely reflects an increased genetic vulnerability to developing schizophrenia, as opposed to illness-specific features. Importantly, however, this notion needs to be tempered with the knowledge that genetic factors may interact with environmental factors, thus making it difficult to determine with certainty that brain abnormalities observed in family members of patients with schizophrenia are necessarily driven only by genetic factors.

A recent review by Bois and colleagues in 2014 summarizes sMRI findings from multiple familial and genetic high-risk studies and meta-analyses. Cross-sectional analyses found decreased gray matter volume in genetic high-risk subjects in the dorsolateral prefrontal cortex, Broca area, Wernicke area, bilateral anterior cingulate cortex, right superior temporal gyrus, left precuneus, left medial frontal gyrus, right middle frontal gyrus, and bilateral hippocampal/parahippocampal regions. Similar results have been reported in studies employing cortical surface measures, with siblings and unaffected relatives showing decreased cortical thickness and surface area in the cingulate gyrus, superior temporal gyrus, and prefrontal cortex when compared to HCs. Investigations into gyrification patterns in familial high-risk individuals have proven to be generally inconclusive, with some studies reporting significant differences in cortical gyrification in unaffected relatives and siblings, whereas others have not. Abnormal gyrification patterns are posited to be indicative of aberrant pre- or perinatal neurodevelopment. Equivocal findings in gyrification in familial risk studies may indicate that the unaffected relatives of individuals with schizophrenia might not have experienced a potential early-stage insult, thereby contributing to their resilience against a transition to psychosis despite their genetic predisposition.

Longitudinal familial risk studies, which follow family members over time to determine if familial high-risk individuals convert to psychosis, also provide a powerful neuroimaging paradigm to assess potential neurodevelopmental progression. Despite the generally low conversion rate seen in familial high-risk studies, comparisons by Bois and colleagues in 2014 between high-risk converters and nonconverters have found that subjects that convert to psychosis exhibit different developmental trajectories from nonconverters and matched unaffected controls. Subjects who convert to psychosis tend to show reductions in cortical thickness over time in frontal, temporal, and parietal areas, as well as localized reductions in cortical surface area, while subjects who do not convert and controls exhibit cortical thickening or surface area expansion over time. These studies demonstrate that structural differences in schizophrenia may precede symptom onset and therefore provide researchers and clinicians with an avenue for the identification of structural biomarkers of risk that may be used for targeted therapeutic interventions, which might mitigate or prevent the onset of psychosis.

Looking Forward: Advances in Structural MRI for the Future. The majority of neuroimaging investigations continue to employ case-control comparisons to determine the location and extent of structural changes in the brains of individuals with schizophrenia or those who are at risk of developing the disorder. While this approach continues to be the bedrock of psychiatric neuroimaging studies, this paradigm does not consider the potential importance of the heterogeneity of the clinical presentations for individuals with schizophrenia or the potential heterogeneity of brain changes that could occur. New applications of a technique called normative modeling, an analysis that investigates the individual deviations from a normative range (usually defined by a population of individuals without severe mental illness), allow for the more refined investigation of the extent to which changes in an individual align with what is observed in a population of healthy individuals. In a study by Lv and coauthors, the investigators applied normative modeling to cortical thickness values derived from structural imaging data from 322 patients with schizophrenia and 195 HCs. When using a case-control approach, significant reductions in cortical thickness were found in 60 of 68 cortical regions examined. However, when applying the normative modeling approach, the majority of individuals with schizophrenia fell within the normative range. In fact, no more than 18% of patients were found to deviate from the normative range. Infranormal (lower than normal) deviations were commonly observed in the temporal and ventromedial prefrontal cortices, but interestingly only 10% to 15% of patients exhibited significant deviations in these areas. Supranormal (higher than normal) deviations were found in the paracentral lobule but only in 3% of the patients. These data provide strong preliminary evidence that there is significant heterogeneity in the location of anatomical differences found in individuals with schizophrenia and that it is likely that case-control comparisons do not appropriately reflect pathologies that would be found in individual patients. Given the results of this study, it is clear that new approaches which aim to refocus research investigations to the study of the individual as opposed to the “average patient” are needed. This has led to the emergence of “precision medicine” approaches, which we discuss in greater detail in the concluding sections.

As we have shown, reductions in gray matter are one of the most consistent findings in sMRI studies of schizophrenia. However, neuroimaging tools remain limited in their ability to probe the underlying cellular and molecular mechanisms that may lead to schizophrenia. Large-scale, genome-wide association studies (GWAS), which can consist of genetic samples of more than 20,000 people, have identified over 100 genes associated with schizophrenia. A greater understanding of the link between schizophrenia risk genes and brain abnormalities would provide much-needed validation of potential neuroimaging biomarkers of psychosis. However, at present, there is no structural imaging sample large enough to conduct a genome-wide neuroimaging association study. In a study by Ji and colleagues, the authors chose to utilize a transcription-neuroimaging association analysis, an approach that does not require extremely large samples. These authors used this approach to determine which schizophrenia-related genes, previously identified by GWAS, were associated with gray matter volume reductions. sMRI data were compiled from five independent samples that were harmonized to generate a population of 276 patients with schizophrenia and 284 HCs. Gene expression data were derived from six postmortem brains from the Allen Human Brain Atlas (http://human.brain-map.org) and were then correlated with the harmonized imaging data. The authors reported that 98 genes showed significant correlations with changes in gray matter volume between patients with schizophrenia and HCs. The 98 identified genes were implicated in critical functions like synaptic transmission, central nervous system development, and cell projection. For example, the authors found a strong positive correlation between gray matter volume reductions in patients and *GRIA1*, a gene linked to the α-amino-3-hydroxy-5-methyl-4-isoxazoleproprionic acid (AMPA) receptor. In patients, higher levels of *GRIA1* were associated with a greater degree of volume loss, a finding which is consistent with other studies that have observed aberrations in synaptic plasticity and glutamatergic transmission in schizophrenia. While this approach is indirect, as the genetic and imaging data were not obtained from the same individuals, approaches such as this provide significant insights into the molecular mechanisms that may underlie structural pathologies commonly observed in schizophrenia populations and will ultimately help to identify candidate markers for future studies. Interdisciplinary approaches exemplified by this study are increasingly needed and will be discussed in depth later in this chapter.

Summary. In summary, structural MRI has proven to be an extremely powerful tool in identifying structural abnormalities in gray matter in patients with schizophrenia. Differences in cortical and subcortical structures are reported at all stages of the disorder, from onset to chronicity, as well as in individuals who are at increased risk for developing the disorder. In addition, large cross-sectional, as well as longitudinal, studies have begun to chart the complex dynamics of gray matter pathology, with its accelerated trajectory around the psychotic onset and possibly later during early aging. At the same time, the newest studies highlight the heterogeneity of schizophrenia pathology, indicating that there might be more than one path to developing the disorder. This includes the already discussed potential influence of genes and environment (in both risk and resilience to develop schizophrenia), aberrant cellular and/or neuroinflammatory pathways, as well as other factors not discussed here, such as hormonal imbalances or cellular senescence. As the field progresses, structural neuroimaging will continue to move toward larger, longitudinal studies where imaging and nonimaging biomarkers will be combined with sophisticated analytic.

Diffusion Magnetic Resonance Imaging Findings in Schizophrenia

Since the early 2000s, there has been a growing interest in understanding the role that white matter may play in the pathophysiologic of schizophrenia. This has been jointly motivated by the development of imaging modalities specifically targeting white matter, such as dMRI, a method capable of visualizing and quantifying major axonal connections of the brain, and by the observation that the pattern and number of gray matter abnormalities observed in schizophrenia are consistent with a disturbance of connectivity within and between brain regions. In addition, transition to psychosis tends to occur in early adulthood, a period that overlaps with the rapid maturation of the majority of white matter tracts in the brain, while gray matter matures earlier. Moreover, neuropathologic and genetic studies suggest the specific involvement of oligodendrocytes, the neural cells that form myelin and provide protection to axons, as well as improve communication between brain areas in the pathophysiology of schizophrenia. All this evidence warrants further research into evaluating the degree to which the white matter abnormalities may contribute to schizophrenia pathophysiology. Since dMRI is the leading method in evaluating white matter structural pathology in the living human brain, in this section, we will focus on schizophrenia findings using dMRI methodology. To parallel the structure of the preceding section, we will first introduce studies that provide an overview of dMRI findings in schizophrenia. This will be followed by a presentation of stage-specific studies (i.e., chronic, first-episode, and clinical high-risk populations), again following the information provided previously for sMRI findings in schizophrenia.

DMRI methodology was introduced in the early 80s, but the first fully quantitative method used to estimate white matter microstructural integrity, called DTI, was introduced in 1994. The main measure derived from this method, called Fractional Anisotropy (FA), was sensitive to multiple microstructural changes within the white matter tissue, including myelin thickness, axonal number, axonal density, fiber coherence, and extracellular water content. The first DTI studies in schizophrenia evaluated the whole brain in relatively small samples with 20 or fewer subjects. However, even with these small sample sizes, reductions in FA were reported in several brain regions.

Over the last 20 years, there has been a dramatic increase in the number of studies focusing on the use of dMRI to investigate individuals with schizophrenia. More importantly, many of the datasets that were acquired are being made publicly available as part of a large-scale effort to promote the development of collaborative “big data” approaches. The availability of these datasets has motivated investigators to develop analytic methods that could combine and analyze multisite–multicenter studies that could leverage much greater statistical power than earlier single-site studies.

One of the first such studies was published in 2018. In this study, Kelly and her colleagues within the ENIGMA Consortium published the results from a meta-analysis on diffusion MR data compiled from 29 independent international sites. Their sample consisted of 1,963 patients with schizophrenia and 2,359 HCs. Of the 25 anatomical regions that were included in their meta-analysis, 20 exhibited significantly lower FA in individuals with schizophrenia compared to controls. The largest effect sizes were found for average whole-brain FA (Cohen’s d = –0.42), followed by the anterior corona radiata (Cohen’s d = –0.40) and the corpus callosum (Cohen’s d = –0.40). These findings align well with several previous research reports in smaller samples which have also shown widespread reductions in FA, as well as more specific reductions in key frontal, temporal, and commissural tracts. Larger effect sizes were also observed for FA when compared to the other DTI indices, mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD), though significantly higher MD and RD were reported in the schizophrenia cohort. This study also found no associations between current medication dosage and age of onset, two key clinical variables of interest. In sum, the results of the meta-analysis by Kelly and colleagues provided further evidence for the presence of microstructural white matter pathology in schizophrenia.

Following this study, Cetin-Karayumak and coauthors took a different approach by using diffusion MR harmonization methods, as opposed to a meta-analytic approach, to combine and analyze together extant dMRI schizophrenia datasets. Briefly, the purpose of harmonization methods is to remove the site-, scanner-, and sequence-specific differences that are generated during data acquisition. Since no two scanners are alike, even if diffusion MR data were to come from two sites that use similar hardware (i.e., same scanner), software (e.g., scanner software upgrades, pre- and postprocessing steps), or sequences, there will still be inherent differences in the data that cannot be controlled for statistically (i.e., including “site” as a covariate). Harmonization methods remove these variabilities from the raw diffusion image at the signal level, thereby allowing for imaging data from independent sites to be compiled and analyzed as if they were acquired from the same scanner with the same sequence. This groundbreaking approach holds significant promise for the future of large-scale data analytics in imaging.

When applying this harmonization method, Cetin-Karayumak and coauthors were able to determine the pattern of global and tract-specific age-related FA differences across the illness course in a sample of 600 patients with schizophrenia and 492 HCs. In whole-brain white matter, FA was significantly reduced in patients across the entire age range studied (14 to 65 years). Reductions in whole-brain FA were greatest in more chronic patients (up to 7% reduction in FA), with smaller differences being observed earlier in the course of illness (minimum reduction: 1.5%). Moreover, individuals diagnosed with schizophrenia were found to reach peak FA (a putative indicator of white matter maturity) 5 years earlier than their healthy counterparts (patient peak FA: 27 years; healthy control peak FA: 33 years). When investigating regional white matter trajectories, the authors observed three distinct patterns of age-related white matter pathologies in schizophrenia. First, neurodevelopmental abnormalities were observed in limbic white matter tracts (e.g., cingulum bundle), manifested as small to medium differences in FA that were stable across all ages. Second, abnormal maturation (i.e., earlier peak) and moderate evidence of accelerated aging were observed in the long-range association fibers (e.g., superior longitudinal fasciculus). Finally, significant deterioration from the onset of illness was observed in commissural tracts (e.g., corpus callosum). One of the strengths of this study is that in addition to showing whole-brain reductions in white matter FA across the entire illness spectrum, this study also demonstrated that there may be significant age-related differences that are tract specific. The use of the diffusion MR harmonization approach was thus critical in being able to model these differences over the illness course because many imaging studies, as we will show and discuss in the subsequent section, tend to focus on the recruitment samples at discrete stages of illness.

In summary, similar to studies using structural MR to study gray matter pathology, the use of diffusion MR has also revealed widespread abnormalities in white matter microstructure in patients with schizophrenia, further suggesting the developmental as well as progressive nature of such abnormalities. Below, we further discuss recently published, and in our opinion, important to the field, dMRI studies investigating discrete schizophrenia populations (chronic, first episode, and CHR) in the context of psychopathology and confounding factors, as well as their potential biologic correlates.

Chronic Patients. Similar to investigations employing sMRI, many of the initial dMRI studies of schizophrenia chose to focus on individuals in the chronic stage of illness. In these studies, the most common index evaluated was FA. A comprehensive review of dMRI findings was compiled in 2014 by Wheeler and colleagues. They report that patients with chronic schizophrenia exhibit decreased FA within the prefrontal and temporal lobes. Specifically, tracts connecting frontal and temporal areas, including the uncinate fasciculus, cingulum bundle, corpus callosum, and internal capsule, were the most consistently reported in the field.

As discussed in several sections of this chapter, one of the prevailing theories regarding structural pathologies in schizophrenia is the idea that those with schizophrenia are experiencing a period of "accelerated aging." In the earlier section focusing on sMRI techniques, we introduced the 2017 study by Cropley and colleagues, which showed evidence of higher rates of gray matter loss in subjects with schizophrenia compared to HCs. The authors also included diffusion MR images in their analysis to investigate white matter FA. Overall, those with schizophrenia exhibited a significantly faster rate of FA decline than controls. In fact, the slope of the line describing the relationship between age and FA was 60% steeper in those with schizophrenia compared to HCs. Significant reductions in FA were observed starting around 35 years of age, after which the reductions continued to increase until the early 60s (maximum difference observed: 3%). Early localized reductions, which also did not emerge until 35 years of age, were primarily found in the genu and body of the corpus callosum and the superior longitudinal fasciculus. With increasing age, localized reductions continued to progress, encompassing more areas until almost all areas of the cortex and the cerebellum exhibited FA reductions. One of the most significant findings of this paper was that in addition to demonstrating that both gray and white matter exhibit evidence of "accelerated aging," gray matter pathologies were present earlier than white matter pathologies. This is significant because it may suggest that white matter pathologies arise as a result of gray matter loss. However, the alternative hypothesis must also be considered—that aberrant processes relating to the development, maturation, and aging of both gray and white matter tissue is just as likely, particularly given that gray matter develops at a younger age than white matter which continues to develop up until the third decade of life. In any case, this study demonstrated that white matter exhibits a deterioration with increasing chronicity, supporting the findings also presented by Cetin-Karayumak, which were reviewed in an earlier section of this chapter.

Aligning with the notion of chronicity-related pathologies, the impact of antipsychotic medication on white matter is also an area of significant interest in the field. While some studies argue that antipsychotic use is detrimental to white matter, others have found evidence that antipsychotic use can ultimately lead to white matter improvements. In contrast to gray matter, which, as noted previously, matures earlier in development, white matter does not mature until the third decade of life. This means that white matter is still actively developing when many individuals are first placed on antipsychotic medications. Neuroimaging studies that have access to chronic, but never-treated patients, while rare, are critical for parsing out the impact of disease processes from those occurring as a result of treatment. In 2018, Xiao and others presented the results of a diffusion dMRI study that sought to determine the extent and location of differences in FA between 31 never-treated patients (average duration of illness: 21.3 years), 46 duration-matched individuals with schizophrenia who were treated with antipsychotics, and 58 HCs. Overall reductions in FA were found in those with schizophrenia compared to HCs. However, greater reductions were observed in the schizophrenia cohort who were never treated compared to those who were treated, specifically in the left anterior thalamic radiation, left cingulum–hippocampal pathway, genu, and splenium of the corpus callosum left superior longitudinal fasciculus and right uncinate fasciculus. In addition, never-treated individuals with schizophrenia exhibited a faster age-related decline compared with the treated the schizophrenia and HC groups. This faster age-related deterioration in FA was primarily observed in the corpus callosum. This report lends further evidence to earlier studies that argue for the beneficial impact of medication on white matter health. Interestingly, this is in opposition to the study presented early in the section on sMRI, which suggested that antipsychotic medication, particularly clozapine, was related to more pronounced gray matter loss. Thus, given the likely interrelatedness of gray and white matter pathologies in schizophrenia, it will be important to fully elucidate the timing and exact biologic nature of medication-induced brain changes in future studies in this area.

In summary, studies using diffusion MR approaches in chronic schizophrenia show reductions in FA in a large proportion of brain regions. Many of the reductions are the result of an exaggerated age-related decline in FA that begins in the commissural tracts and ultimately expands to all parts of the brain. Long-term medication may potentially lead to improvements in white matter in the later stages of illness. However, this is still being investigated. Future studies in chronic schizophrenia need to include multiple additional factors that likely influence the integrity of white matter and may contribute to the pathology observed in more chronic stages of the disease. These include chronic substance abuse, social isolation, and/or comorbid medical disorders (including metabolic syndrome, which is prevalent in schizophrenia).

First-Episode and Early-Course Patients. Similar to the pattern of gray matter changes, white matter also appears to exhibit widespread but less pronounced changes in the early stages of the illness compared to chronic stages. Studies conducted in the early 2000s, many of which had small sample sizes and relatively low statistical power, have been now replaced by larger multisite mega-analyses, novel analytic approaches which pool raw data from multiple studies conducted under comparable conditions (in contrast to meta-analyses which solely rely on summary statistics). In one such review by Wheeler, authors reported the involvement of white matter tracts (frontotemporal and interhemispheric connections) in the pathology of early course schizophrenia which was also observed in chronic schizophrenia. These investigators, however, noted that the findings in the early stages of the disease are less consistent across studies than those reported in chronic schizophrenia. Based on data from the available studies, the authors also concluded that medication was not likely the direct cause of white matter pathology, as studies with medication-naive subjects also report reductions in FA.

While dMRI is used to characterize the location and the extent of white matter pathology in first-episode and early-course schizophrenia, there are also other metrics used to characterize white matter alterations that contribute to our understanding of schizophrenia pathophysiology. Over the last several years, the focus has also shifted from the leading hypothesis that white matter anomalies in schizophrenia involve myelin pathology to the alternative hypothesis discussed earlier that implicates an immunologic brain response, especially in the early phases of the disorder. Immune activation is purported to negatively affect white matter integrity and lead to neurodegenerative and/or progressive changes in the later stages of the illness. In order to test such hypotheses, new models based on diffusion MR have emerged. One such model is called free water (FW) imaging. In this model, the diffusion signal is broken down into two components. The first component measures the amount of unrestricted or “free” diffusion that is taking place in the extracellular space (FW), while the other measures the amount and orientation of restricted water diffusion taking place in close proximity to cellular processes (FA of the tissue—FAt). Both the traditional DTI model and the FW model were applied to diffusion MR data from 63 first-episode patients and 70 matched HCs in a study by Lyall and colleagues. When applying the traditional DTI model, FA reductions in individuals with schizophrenia compared to controls were seen in most white matter areas (consistent with Wheeler and other studies). However, when applying the FW model, widespread and significant increases in FW showed a high degree of spatial overlap with the observed FA reductions. In contrast to earlier studies suggesting potential widespread white matter pathology in first-episode populations, the authors only observed small, localized reductions in the tissue-related measure FAt, which were mainly restricted to the body and genu of the corpus callosum, the left and posterior corona radiata, and the left superior longitudinal fasciculus. The results of this study are important because they suggest that there may be two co-occurring pathologies that are present in the brains of individuals who recently converted to psychosis. The first pathology is a widespread increase in extracellular water that may potentially be related to immune activation, while the second may reflect white matter-specific tissue aberrations (possibly related to myelin degeneration). Future studies are needed to understand better the timing and clinical correlates of such changes, but the initial findings discussed here have helped to pave the way for the development of more targeted medical treatment that will be based on biologic, rather than clinical biomarkers of early psychosis.

High-Risk Populations. Over the past few decades, there has been a gradual shift from characterizing actively ill chronic and early-course individuals with schizophrenia to dMRI investigations focusing on individuals at high risk for psychosis. The aim of this research is to understand better potential factors that contribute to the transition to psychosis from those who exhibit evidence of subthreshold clinical symptoms (CHR individuals) or those who have a first-degree relative with a confirmed diagnosis of schizophrenia (genetic and/or familial high risk). By studying the brains of these individuals, researchers hope to identify features that could be helpful for diagnosis, patient stratification, outcome prediction, symptom mitigation, treatment options, and, ultimately, prevention. While this field has grown in size, the application of dMRI has become a useful tool to understand the degree to which structural connectivity changes can confer and/or contribute to increased risk. Here we present a selection of this emerging field.

CLINICAL HIGH RISK. As highlighted in the preceding section, newer dMRI techniques, such as FW imaging, are being applied in clinical settings. This is particularly true for the investigation of individuals who are at clinically high risk for developing schizophrenia. This is due, in part, to the inconsistent findings in the dMRI literature for CHR populations which, in turn, is likely due to this cohort being highly heterogeneous. As a result, more sophisticated methods are necessary to identify reliable and objective markers for critically important clinical outcomes, such as the likelihood of transitioning to psychosis, functional/cognitive decline, and other adverse outcomes. One example of the application of newer models in CHR populations is a study by Tang and colleagues, where the authors utilized dMRI data from 50 CHR individuals, 40 of whom were not medicated, and 50 matched controls. The group acquired multishell dMRI data, a more advanced acquisition protocol that allows for more complex biophysical models of water diffusion to be applied. The investigators used FW imaging to investigate whole-brain white matter microstructure and observed a reduction in the tissue-related measure FAt but not in the extracellular measure of FW in the CHR group when compared to controls. Reductions in FAt in the CHR group were also associated with lower Global Assessment of Function (GAF). Taken in the context of previous studies which have utilized FW imaging, specifically those studies which focused on the investigation of first-episode psychosis populations, the lack of FW elevations was surprising but led to the conclusion that FW elevations may only be present at the onset of psychosis, but not in the at-risk state. The reductions in FAt (which, as suggested earlier, might reflect myelin pathology) were interpreted as evidence of a potential neurodevelopmental white matter aberration, predisposing individuals to exhibiting subthreshold clinical symptoms.

Recent results from the NAPLS-3 consortium are also consistent with the Tang et al. study cited above. Here, Di Biase and coauthors leveraged a longitudinal multishell dMRI data collection paradigm (dMRI data collected one to five times over the course of 1 year for each participant) from 286 subjects (aged 12 to 32 years). Within this sample, there were 25 CHR individuals who transitioned to psychosis during the follow-up period (CHR-P), 205 CHR individuals who did not transition (CHR-U), and 56 HCs. At baseline, there were no group differences observed between the CHR-P, CHR-U, and HC for either FAt or FW. However, when investigating the relationships of these measures across the age range, the authors observed that both CHR-P and CHR-U populations exhibited initially higher FAt when compared to controls. After adolescence, FAt gradually decreased for both CHR-P and CHR-U individuals, ultimately falling below their HC counterparts as they entered adulthood (~30 years of age). Interestingly, the rate of FAt decrease was greater in the CHR-P population than in the CHR-U population, suggesting greater developmental deterioration in CHR-P individuals nearing transition. When taken together, these findings provide evidence that the CHR state may be characterized by subtle neurodevelopmental differences in white matter microstructure and, more importantly, that transition to psychosis may not be a period during which dramatic WM structural changes occur. Thus, in contrast to findings of gray matter that were previously discussed, where significant gray matter loss is observed after the transition to illness, this study showed that subtle, developmental differences in white matter maturation may be central to the manifestation of psychosis and further supports the need for well-designed and well-powered longitudinal, multimodal studies.

GENETIC AND FAMILIAL HIGH RISK. Genetic and familial high-risk populations are also critical for the identification or characterization of features more likely to be associated with genetic, trait-related factors as opposed to the presentation of subthreshold symptoms, which would be considered state related. In these studies, dMRI data from unaffected family members are usually compared to control subjects and to their affected family members in order to determine the extent to which white matter displays potential genetically mediated phenotypes.

One recent representative example of this approach is a study by de Leeuw and colleagues. Instead of performing a whole-brain approach, the authors chose to specifically investigate frontostriatal white matter tracts in adolescents (aged 10 to 18 years) who had a parent diagnosed with schizophrenia. The decision to focus on these specific tracts was motivated by prior work performed by the research team, which demonstrated compromised structural connectivity in the tract, which connected the left nucleus accumbens and left dorsolateral prefrontal cortex in individuals with schizophrenia. Imaging data from 27 adolescent offspring of schizophrenia individuals and 32 matched healthy adolescents was investigated using DTI, and findings showed increased FA in the frontostriatal white matter tract in high-risk adolescents relative to controls. In addition, while FA showed age-related increases in the control subjects in this tract, the same increase was not observed in the genetic high-risk subjects. The greater FA, as well as the abnormal relationship between FA and age in the genetically high-risk individuals, was suggested to represent an acceleration of the maturation of white matter, which may take place in individuals who are at an elevated risk for psychosis. These findings align well with previous studies reviewed in this chapter that implicate abnormal white matter development at the time of the emergence of psychotic symptoms. They further suggest that there is likely a genetically mediated component to these neurodevelopmental aberrations.

Looking Forward: Advances in Diffusion MRI for the Future. While we tended to focus on studies that applied FW imaging in several previous sections, there are several other advanced models for dMRI widely used, which should be recognized. In the interest of brevity, we will only focus on one of these methods: neurite orientation dispersion and density imaging (NODDI). NODDI approaches require the collection of multishell diffusion data. As a result, the NODDI approach cannot be applied to older, legacy datasets. Similar to FW imaging, NODDI models several water compartments that are proposed to represent different biologic processes of the underlying tissue. One example of the NODDI approach in the study of schizophrenia is a study by Rae and colleagues. In this publication, the authors apply both the traditional DTI approach in addition to NODDI in 35 individuals with first-episode psychosis and 19 matched controls. As previously reported, individuals with first-episode psychosis exhibited lower FA in the corticospinal tracts as well as several commissural and association tracts. However, when applied NODDI, the authors found that FA reductions were co-localized with a reduction in neurite density index (NDI), as opposed to fiber arrangement (compartment termed orientation distribution index—ODI). The authors interpreted these differences to be indicative of aberrant processes that would most likley affected axonal fibers number, density and/or myelination as opposed to processes that would affect the spatial organization of fiber tracts. In addition to providing further evidence for a neurodevelopmental origin for schizophrenia white matter pathology, this also demonstrates that the utilization of novel tools can provide an avenue for the generation of new biologic hypotheses for schizophrenia.

Summary. In summary, dMRI has proven to be an extremely powerful tool in identifying white matter abnormalities in patients with schizophrenia.

Differences in white matter axonal connections are reported at all stages of the disorder, from onset to chronicity, as well as in individuals who are at increased risk for developing the disorder. Recent advances in dMRI processing and analysis including novel methods for data aggregation, have allowed for analysis of large, multisite datasets and a greater understanding of the timing of white matter pathology. In addition, the development of novel, biologically sensitive diffusion measures have shed light on the possible biologic nature of those abnormalities. White matter tissue pathology (most likely myelin abnormalities) appears to predate schizophrenia onset and might be a good biomarker for conversion to psychosis in clinical high-risk individuals. White matter tissue pathology also seems to follow that of gray matter, and progress over time, but unlike gray matter, chronic medication does not seem to negatively affect white matter health. In addition, it is suggested from some findings in the literature that the onset of psychosis may be characterized by increased extracellular water, which, in turn, might be linked to possible neuroinflammatory pathways. Finally, findings from familial studies seem to suggest that there might be a factor of white matter resilience in healthy offspring or siblings of schizophrenia patients.

FUTURE DIRECTIONS FOR STRUCTURAL BRAIN IMAGING IN SCHIZOPHRENIA

As noted in the information presented in this chapter, there have been incredible advances gleaned from studies of structural brain imaging in schizophrenia. Since its inception, neuroimaging has proven to be a powerful way to establish schizophrenia more clearly as a “brain disorder,” as well as to interrogate, *in vivo*, both the biologic nature and the trajectory of the illness. However, despite a large number of structural imaging studies, the field has yet to achieve the goal of producing clinically relevant imaging-based “biomarkers of schizophrenia.” This goal is motivated by the fact that within practically all other fields of medicine, a series of diagnostic evaluations can be carried out upon entry into a hospital. Unfortunately, psychiatry still relies on observation or report of symptoms and patient illness history, which can vary significantly between and within individuals. Similarly, many structural imaging research studies continue to rely upon case-control comparisons, many of which can be underpowered and therefore lack generalizability to the entire population. Despite the enormous contributions toward a greater understanding of schizophrenia, the field of structural imaging has yet to provide a feature or set of features that will yield sufficient diagnostic information that is clinically relevant to a given individual. Thus, while structural imaging tools continue to be valuable in research studies, reliable biomarkers remain elusive, and the field needs to change both conceptually and methodologically with respect to how we approach the study of schizophrenia. Below we present some recommendations for future approaches that may bring the field closer to this overarching goal.

Biologic Classifications for Schizophrenia

It is becoming more widely accepted that the clinically defined disorder known as “schizophrenia” may potentially be an assemblage of subtle biologic changes that present along an illness spectrum. As a result, the acknowledgment of a model in which psychiatric illnesses are characterized as a set of potentially overlapping biologic pathologies instead of discrete entities is shaping pathways forward in the field. Initiatives such as the National Institute of Mental Health’s Research Domain Criteria (RDoC) framework identified the need for more biologically based approaches to the study of psychiatric illnesses. Imaging techniques are uniquely suited to address this need as they are presently the only avenue to investigate human brain structure and function *in vivo.* Along these same lines, there has already been a considerable effort to define, based on imaging and electrophysiologic profiles, the so-called “biotypes” along the psychosis spectrum (e.g., see the work of Clementz et al., 2016). Such efforts should benefit further from including morphometric as well as more biologically sensitive neuroimaging measures discussed in the previous sections of this chapter.

Big Data Meets Precision Medicine

One of the biggest limitations of the clinical neuroimaging field is the recruitment of psychotic populations for research studies. Small samples and the resultant limited power in statistical analyses are recurrent issues in neuroimaging that likely contribute to replicability problems within the field. To overcome such limitations, and as showcased in this chapter, many groups have chosen to employ either meta-analytic or harmonization methods to increase sample size in order to ascertain a greater understanding of the general trends within a specific area of research. These approaches are aiding “big data” analytical approaches, which have garnered significant popularity within the clinical imaging field in the last decade. In fact, many large-scale databases have already been created, several of which are freely available to the public (e.g., fBIRN, NAPLS, PRONIA, Australian Schizophrenia Research Bank, BSNIP, Human Connectome Project). The compilation of these resources, along with the advancement of image analysis methodologies, has also made it possible for research teams around the world to employ another important emergent group of analysis tools: machine learning.

Due to the current shift in focus toward the identification of biologic constructs underlying psychopathologic features of psychotic illnesses, many researchers have started to embrace a bottom-up, data-driven approach to the stratification and/or classification of patients with psychiatric disorders. In these paradigms, subgroups of individuals can be created based on objective and measurable phenotypes derived from an assortment of modalities, whether it be neuroimaging metrics, physiologic assays, or genetic information. Two classes of approaches exist: (1) supervised methods (e.g., support vector machines) and (2) unsupervised methods (e.g., principal component analysis). Supervised methods are best used for predictions, whereas unsupervised methods are commonly used for clustering, where groups are defined without any *a priori* knowledge (as in the context of the “biotypes” example discussed earlier). With the increased availability of large data repositories, research teams can begin to apply these approaches to search for viable, replicable biomarkers of disease risk, onset, and outcome.

The complementary approach to big data is precision medicine. Here, individual patient information such as illness history, family history, medication history, trauma exposure, genetics, and nutrition, are incorporated into a framework that allows for more detailed and patient-specific questions. The main difference of this approach, compared to a big data approach, is that the individual nature of the illness is recognized, as opposed to an individual serving as a single data point in a larger analysis. Precision medicine also encourages the development of small, well-designed, hypothesis-driven, and, most importantly, patient-centered studies. When successful, many of these studies can have a significantly higher impact on the life of a patient and are much more easily transferred to a clinical setting. The major limitation of precision medicine approaches is the limited generalizability to larger populations of individuals with similar diagnoses.

Interestingly, despite appearing to represent two ends of a methodologic spectrum, “big data” and precision medicine approaches have been shown to work synergistically. In fact, for many machine learning approaches, a model can be trained on a set of data and then subsequently tested or challenged with the introduction of a new individual's data. These types of approaches are best used in classification paradigms for individuals who may be appropriate for specific types of treatment or for patient stratification. With the imaging community having greater access to larger repositories of high-dimensional data (both imaging and nonimaging), it is becoming possible to develop new hypotheses from the smaller groups of patients, which can then be tested in a larger, more representative samples for that given clinical population, and vice versa.

Studying Schizophrenia as a Lifespan Disorder

Schizophrenia is a chronic and debilitating life-long illness with likely neurodevelopmental origins. As such, it is important to recognize that it is a disorder with pathologies that evolve over time. In order to appropriately use the power of neuroimaging to study the structural changes that are linked to schizophrenia pathophysiology, there need to be concerted and continuous efforts to design effective, well-powered, longitudinal imaging studies. Multimodal acquisitions, such as the collection of both functional and structural imaging data, along with other important nonimaging markers like electrophysiology, blood, genetics, and behavioral inventories, will be critical. Ensuring that these collections occur at multiple points in a patient's illness course (e.g., before and after the onset of illness, before and after initiation or alteration of medication, during periods of high symptom burden vs. low symptom burden) will provide key steps forward in understanding the dynamic interplay that likely takes place between several biologic systems in the brain and body and how they contribute to the structural alterations observed in schizophrenia.

FURTHER READINGS

Andreasen NC, Nopoulos P, Magnotta V, Pierson R, Ziebell S, Ho BC. Progressive brain change in schizophrenia: a prospective longitudinal study of first-episode schizophrenia. *Biol Psychiatry.* 2011;70(7):672–679.

Bois C, Whalley HC, McIntosh AM, Lawrie SM. Structural magnetic resonance imaging markers of susceptibility and transition to schizophrenia: a review of familial and clinical high risk population studies. *J Psychopharmacol*. 2015;29(2):144–154.

Cannon TD, Chung Y, He G, et al. Progressive reduction in cortical thickness as psychosis develops: a multisite longitudinal neuroimaging study of youth at elevated clinical risk. *Biol Psychiatry*. 2015;77(2):147–157.

Cetin-Karayumak S, Di Biase MA, Chunga N, et al. White matter abnormalities across the lifespan of schizophrenia: a harmonized multisite diffusion MRI study. *Mol Psychiatry*. 2020;25(12):3208–3219.

Clementz BA, Sweeney JA, Hamm JP, et al. Identification of distinct psychosis biotypes using brain-based biomarkers. *Am J Psychiatry*. 2016;173(4):373–384.

Cropley VL, Klauser P, Lenroot RK, et al. Accelerated gray and white matter deterioration with age in schizophrenia. *Am J Psychiatry*. 2017;174(3):286–295.

de Leeuw M, Bohlken MM, Mandl RC, Hillegers MH, Kahn RS, Vink M. Changes in white matter organization in adolescent offspring of schizophrenia patients. *Neuropsychopharmacology*. 2017;42(2):495–501.

Di Biase MA, Cetin-Karayumak S, Lyall AE, et al. White matter changes in psychosis risk relate to development and are not impacted by the transition to psychosis. *Mol Psychiatry*. 2021;26:6833–6844.

Ding Y, Ou Y, Pan P, et al. brain structural abnormalities as potential markers for detecting individuals with ultra-high risk for psychosis: a systematic review and meta-analysis. *Schizophr Res.* 2019;209:22–31.

Ji Y, Zhang X, Wang Z, et al. Genes associated with gray matter volume alterations in schizophrenia. *Neuroimage*. 2021;225:117526.

Kelly S, Jahanshad N, Zalesky A, et al. Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group. *Mol Psychiatry*. 2018;23(5):1261–1269.

Liu N, Xiao Y, Zhang W, et al. Characteristics of gray matter alterations in never-treated and treated chronic schizophrenia patients. *Transl Psychiatry.* 2020;10(1):136.

Lv J, Di Biase M, Cash RFH, et al. Individual deviations from normative models of brain structure in a large cross-sectional schizophrenia cohort. *Mol Psychiatry*. 2020;26: 3512–3523.

Lyall AE, Pasternak O, Robinson DG, et al. Greater extracellular free-water in first-episode psychosis predicts better neurocognitive functioning. *Mol Psychiatry*. 2018;23(3):701–707.

Núñez C, Stephan-Otto C, Usall J, et al; PEPs group. Neutrophil count is associated with reduced gray matter and enlarged ventricles in first-episode psychosis. *Schizophr Bull*. 2019;45(4):846–858.

Rae CL, Davies G, Garfinkel SN, et al. Deficits in neurite density underlie white matter structure abnormalities in first-episode psychosis. *Biol Psychiatry*. 2017;82(10):716–725.

Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. *Schizophr Res*. 2001;49(1–2):1–52.

Tang Y, Pasternak O, Kubicki M, et al. Altered cellular white matter but not extracellular free water on diffusion MRI in individuals at clinical high risk for psychosis. *Am J Psychiatry*. 2019;176(10):820–828.

van Erp TGM, Hibar DP, Rasmussen JM, et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA Consortium. *Mol Psychiatry.* 2016;21(4):547–553.

van Erp TGM, Walton E, Hibar DP, et al. Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the enhancing neuro imaging genetics through meta analysis (ENIGMA) Consortium. *Biol Psychiatry.* 2018;84:(9):644–654.

Wheeler AL, Voineskos AN. A review of structural neuroimaging in schizophrenia: from connectivity to connectomics. *Front Hum Neurosci*. 2014;8:653.

Xiao Y, Sun H, Shi S, et al. White matter abnormalities in never-treated patients with long-term schizophrenia. *Am J Psychiatry*. 2018;175(11):1129–1136.

12.8 Functional Brain Imaging in Schizophrenia

RAQUEL E. GUR, M.D., PH.D., AND RUBEN C. GUR, PH.D.

Functional neuroimaging has become a major translational tool for advancing the understanding of the neurobiology of schizophrenia. Neuroimaging affords dissecting complex mental processes disrupted in schizophrenia to probe the nature and extent of aberrant activity in the underlying brain circuitry. The functional imaging methodology can thus help incorporate cognitive and affective neuroscience paradigms in efforts to understand brain regulation of behavior. The challenge is to harness the power of functional neuroimaging to clarify the neuropsychiatric manifestations of schizophrenia, integrating this knowledge with advances in genomics. Functional neuroimaging, applied in combination with pharmacologic probes, can augment the bridge between basic and clinical neuroscience. Such advances are helping elucidate mechanisms underlying schizophrenia with implications for treatment.

In the years since the previous edition of this textbook, the literature on functional neuroimaging in schizophrenia has continued to grow, especially in the application of functional magnetic resonance imaging (fMRI) methods applied for activation and for examining resting-state connectivity. This chapter will highlight fMRI applications in schizophrenia and clinical risk populations, summarizing main areas of progress, common issues, and potential future directions.

MAGNETIC RESONANCE FUNCTIONAL IMAGING METHODS

The introduction of fMRI was an exciting development in functional imaging research. Functional MRI methods offer multiple advantages over positron emission tomography (PET) for imaging brain function, including higher spatial resolution, higher temporal resolution, noninvasiveness, lack of ionizing radiation, direct correlation with anatomical imaging, greater repeatability, and economy. Disadvantages of fMRI techniques include the presence of loud background noise generated by the gradients, difficulties in presenting stimuli and performing tasks in the magnet bore, claustrophobia, low signal-to-noise ratios for most methods, and lack of quantitation in physiologic units for most methods. Many of these disadvantages can be overcome through the use of specialized equipment compatible with the MRI environment (Fig. 12.8–1). Considering the costs versus benefits and availability of fMRI, most PET studies in schizophrenia now examine receptor function, whereas brain-behavior research is conducted with fMRI. The most commonly applied methods are blood oxygenation level-dependent (BOLD) and arterial spin–labeled (ASL) imaging of cerebral blood flow rates in physiologic units (mL/100 g/min).

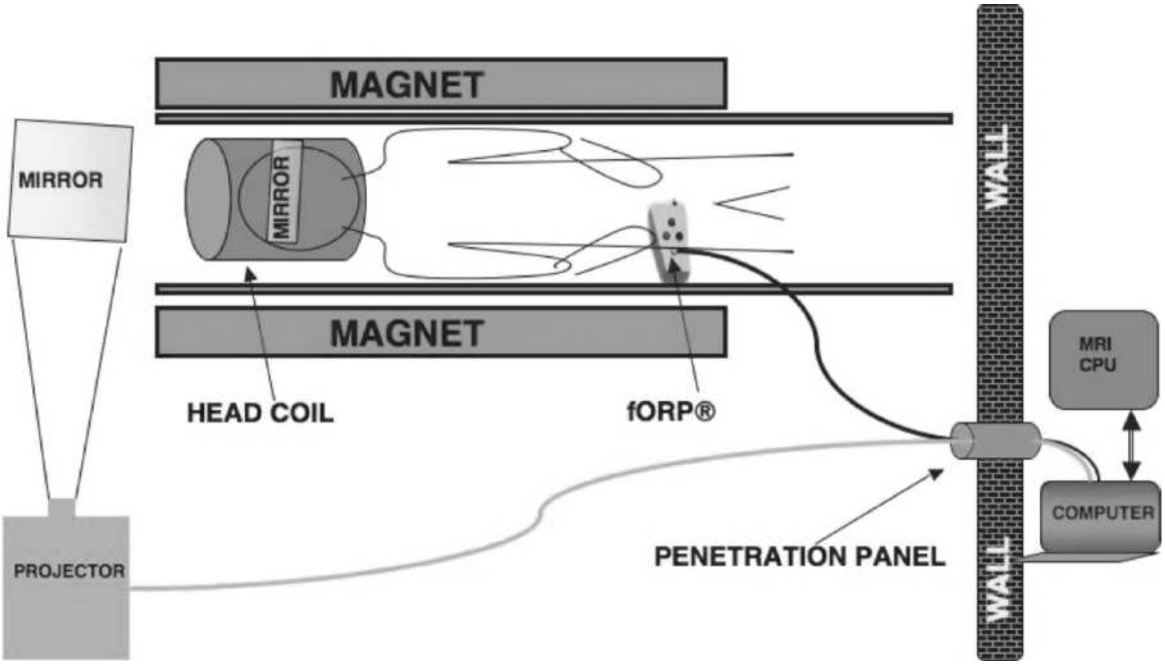

FIGURE 12.8–1. A schematic functional magnetic resonance image setup. fORP, fiber-optic response pad.

Blood Oxygenation Level–Dependent Imaging

The technique relies on the magnetic susceptibility effects of deoxyhemoglobin, which cause regional signal decreases in imaging sequences that are sensitive to susceptibility (e.g., echoplanar or routine gradient echo sequences). With regional brain activation studies, a net increase in signal intensity is observed in regions activated by the task. The increase in image intensity corresponds to a local decrease in deoxyhemoglobin. This effect is attributed to a greater increase in regional blood flow than in regional oxygen consumption, as indicated by PET measurements of blood flow and oxygen consumption associated with brain activation. A variety of pulse sequences can be used to obtain BOLD measures. Many simple activation paradigms have been tested, and activation has been observed with both fast and slow imaging. A typical response is a 1% to 25% change in regional image intensity that develops over 3 to 8 seconds following task initiation. The susceptibility effects of deoxyhemoglobin are field dependent. Thus, a scanner with 1.5 T field strength would typically record signal changes with functional activation of about 0.25% to 5%, while at 3 T and 7 T, scanning changes even higher than 25% have been observed. The main advantage of novel ultrafast imaging, especially using multiband parallel acquisition, is that the time course of signal change can be observed, and multislice imaging can be carried out in a reasonable time period.

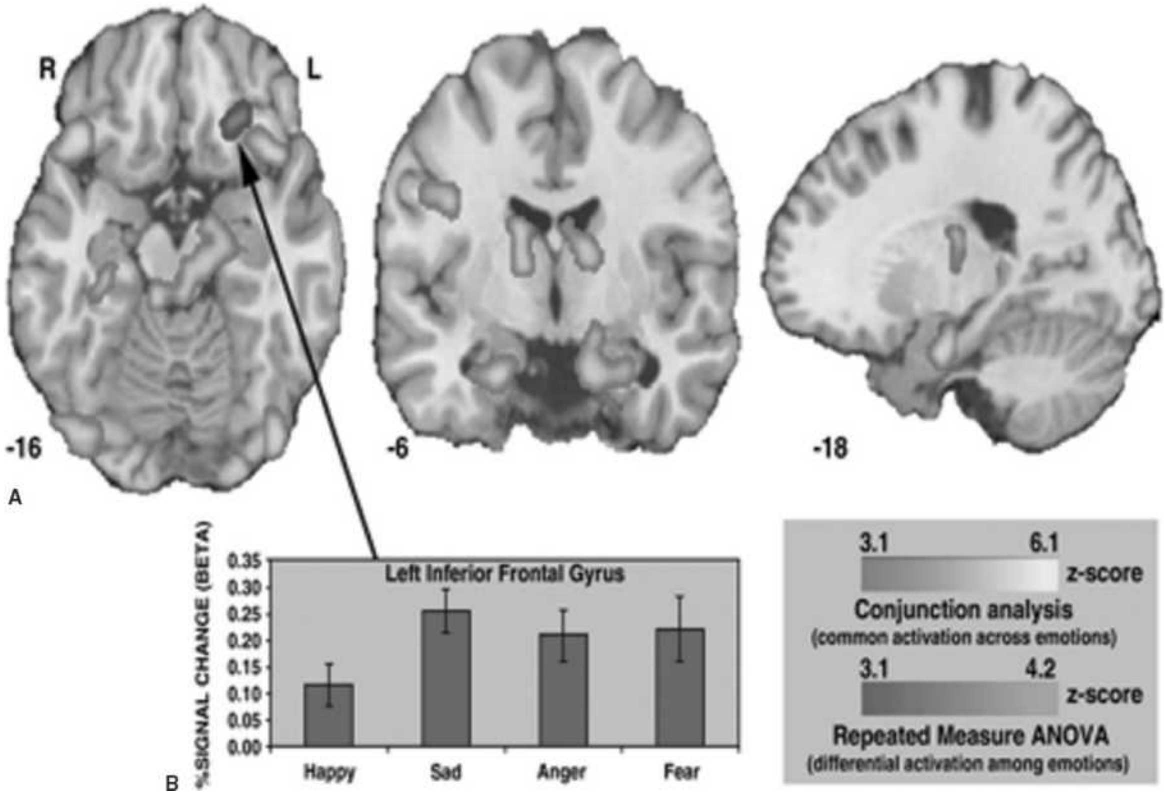
R
L
-16
-6
-18
A
B
Left Inferior Frontal Gyrus
% SIGNAL CHANGE (BETA)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Happy
Sad
Anger
Fear
3.1
6.1
z-score
Conjunction analysis
(common activation across emotions)
3.1
4.2
z-score
Repeated Measure ANOVA
(differential activation among emotions)

FIGURE 12.8–2. Block analysis results. **A:** Group analysis (*n* = 17) showing average brain activation for all target emotions and target emotion-sensitive activation in the inferior frontal gyrus. **B:** Mean percentage signal change for inferior frontal gyrus by target emotion. Images are displayed over a Talairach-normalized template in radiological convention (left hemisphere to viewer's right). (Reprinted from Loughead J, Gur RC, Elliott M, Gur RE. Neural circuitry for accurate identification of facial emotions. *Brain Res.* 2008;1194:37–44. Copyright © 2007, with permission from Elsevier. doi: 10.1016/j.brainres.2007.10.105.)

BOLD studies have initially relied primarily on blocked conditions where signal change was integrated over periods ranging from 20 seconds to minutes. In such designs, a sequence of stimuli is presented, and either the type of stimulus changes in blocks or the task may change. The signal is contrasted between different blocks. For example, in a study investigating neural substrates for identifying facial expressions, a sequence of faces is presented, separated by a control stimulus. The face stimuli repeat throughout the study, but the task changes, creating four blocks or conditions: (1) Push button for happy; (2) Push button for sad; (3) Push button for anger; (4) Push button for fear. Comparing signals integrated across face trials to that of the control stimulus will identify brain regions recruited for facial affect identification. Contrasting each condition with the baseline and with each other would highlight activation specific for each block of a specific target emotion (Fig. 12.8–2).

Event-Related fMRI

This method is a variant of the BOLD technology where, rather than aggregating the tasks into blocks, an estimate of the hemodynamic response is obtained by interspersing the appearance of stimuli and contrasting the signal following stimulus presentation to that following a control stimulus. These contrasts provide information on brain regions whose activation is time-locked to the appearance of the specific stimulus class. For example, if the task is to push the button for a specific emotion such as fear, the question may be asked: Which brain regions activate every time a fearful face appears, compared to when the face is neutral? This is a different question from that posed by the blocked analysis. The condition is the same, searching for fearful expressions that are intermixed with other emotions and with neutral faces. However, the event-related analysis will identify those brain regions within the facial affect–processing network that is specifically engaged when a fearful face appears rather than a neutral face. Furthermore, an event-related analysis can identify brain regions whose activation is associated with correct responses and separate them from brain regions associated with errors. This feature permits a closer link between brain activation with performance. An fMRI study can be done with blocked and event-related paradigms, or, by carefully spacing stimulus classes, hybrid designs can be applied where blocked analysis can identify brain regions engaged in specific tasks, and event-related analysis can be used to establish time-locked components of the hemodynamic response related to specific stimulus classes. For example, in the facial affect identification task, the block design revealed a circuitry engaged in this task that included sensory, limbic, and evaluative components. However, the spacing of the presentation of stimuli within blocks enabled contrasting hemodynamic changes associated with the appearance of fearful faces to those associated with neutral faces, and this contrast showed greater limbic activation associated with the correct identification of fearful and angry faces in healthy people (Fig. 12.8–3).

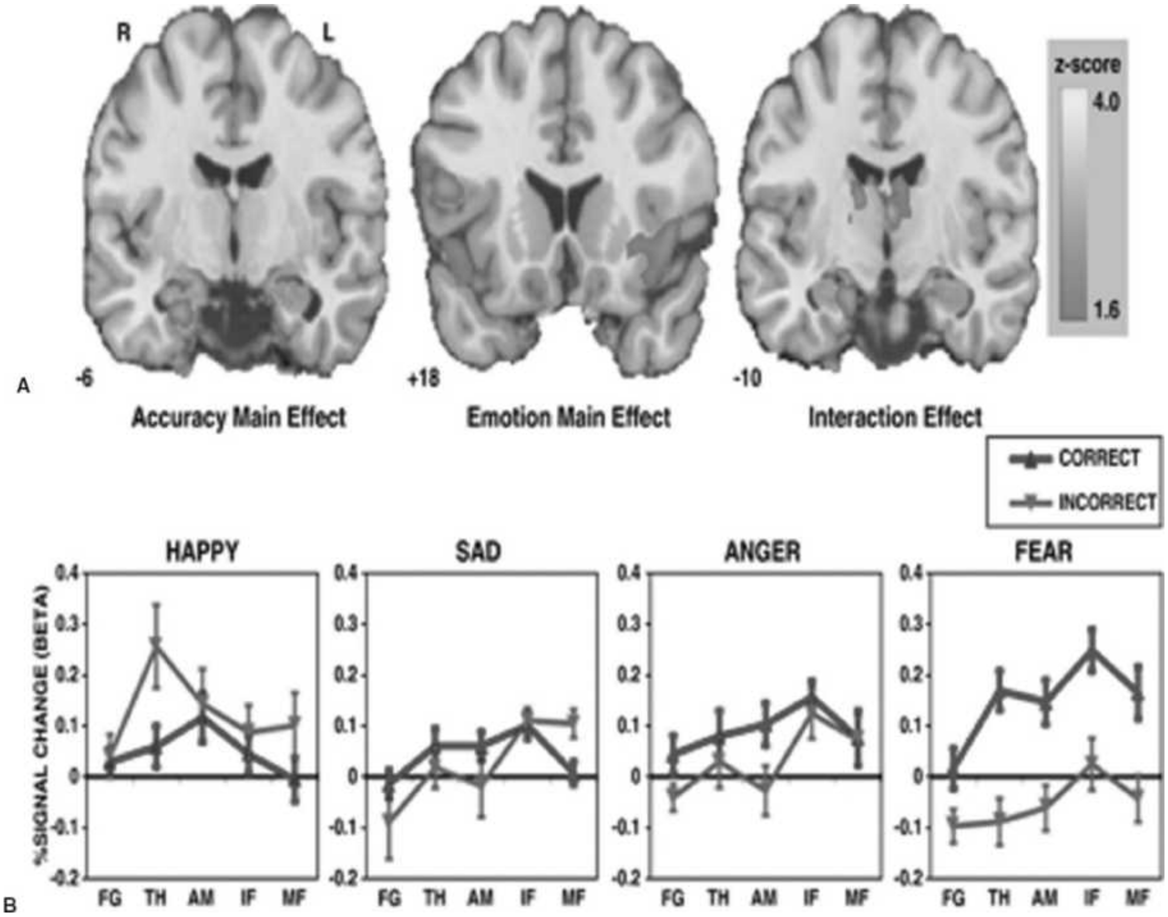
R
L
z-score
4.0
1.6
A
-6
+18
-10
Accuracy Main Effect
Emotion Main Effect
Interaction Effect
CORRECT
INCORRECT
HAPPY
SAD
ANGER
FEAR
% SIGNAL CHANGE (BETA)
0.4
0.3
0.2
0.1
0
-0.1
-0.2
FG
TH
AM
IF
MF
B

FIGURE 12.8–3. Event-related results. **A:** Emotion (happy, sad, anger, fear), accuracy (correct vs. incorrect), and interaction effects for whole brain repeated measure ANOVA highlighting mesial temporal lobe regions. **B:** Mean percentage signal change for regions showing main or interaction effects (error bars represent S.E.M. [standard error of the mean]). Images are displayed over a Talairach-normalized template in radiologic convention (left hemisphere to viewer's right), masked by block activation and thresholded *P* = .05. (Reprinted from Brain Res 1194. Loughead J, Gur RC, Elliott M, Gur RE. Neural circuitry for accurate identification of facial emotions, 37–44. Copyright © 2007, with permission from Elsevier. doi: 10.1016/j.brainres.2007.10.105.)

Arterial Spin Labeling

This technique provides a quantitative perfusion imaging method that uses magnetization tagging of endogenous arterial water to determine perfusion of brain tissue by comparing images obtained with and without labeling of the arterial supply. The method is analogous to steady-state techniques used in PET since the regional signal intensity is dependent on the arterial blood flow, which delivers labeled spins, and the T1 relaxation rate, which causes the labeling to decay. This technique has the important advantage of providing quantitative cerebral blood flow parameters in physiologic units (milliliter per 100 g of tissue per minute). Furthermore, perfusion is measured in brain parenchyma directly and is thus better localized than measurements obtained using an intravascular tracer, which is most sensitive to venous outflow effects. There may also be less motion sensitivity than BOLD (Fig. 12.8–4).

Analytic Approaches

The application of analytic methods in fMRI has been increasingly standardized and includes motion correction, high-pass filtering, spatial smoothing, and scaling. The resulting translational motion parameters should be examined to ensure that there is no excessive motion. Non-brain areas should be removed and the functional images coregistered to the structural volume. This image is then usually transformed into a standard anatomical space, and transformation parameters are later applied to statistical images after analysis is carried out in the subject's space. A more advanced spatial normalization procedure for fMRI data involves the application of the high-dimensional deformations determined from the anatomic images to the coregistered fMRI images. This provides higher registration accuracy across subjects and, therefore, potentially improves sensitivity in activation detection. This approach also allows correlating brain volumes with fMRI activity voxel-wise, as structural and functional images follow the same transformation to the standard space.

Subject-Level Time Series. This statistical analysis is carried out where condition events are modeled with a canonical hemodynamic response function and, when appropriate, its temporal derivative. Whole brain statistical analysis is completed for each individual in subject space, and resulting contrast maps, obtained through linear contrasts of event types, are spatially normalized.

Group-Level Random Effects. This analysis is performed, including within- and between-group comparisons. Significance thresholds are based on spatial extent (k) and peak height (u). Often, a height threshold corresponding to an uncorrected P value of .005 is used, requiring a minimum cluster $P \leq .05$. This results in a corrected probability of $P < .05$, based on the theory of Gaussian fields.

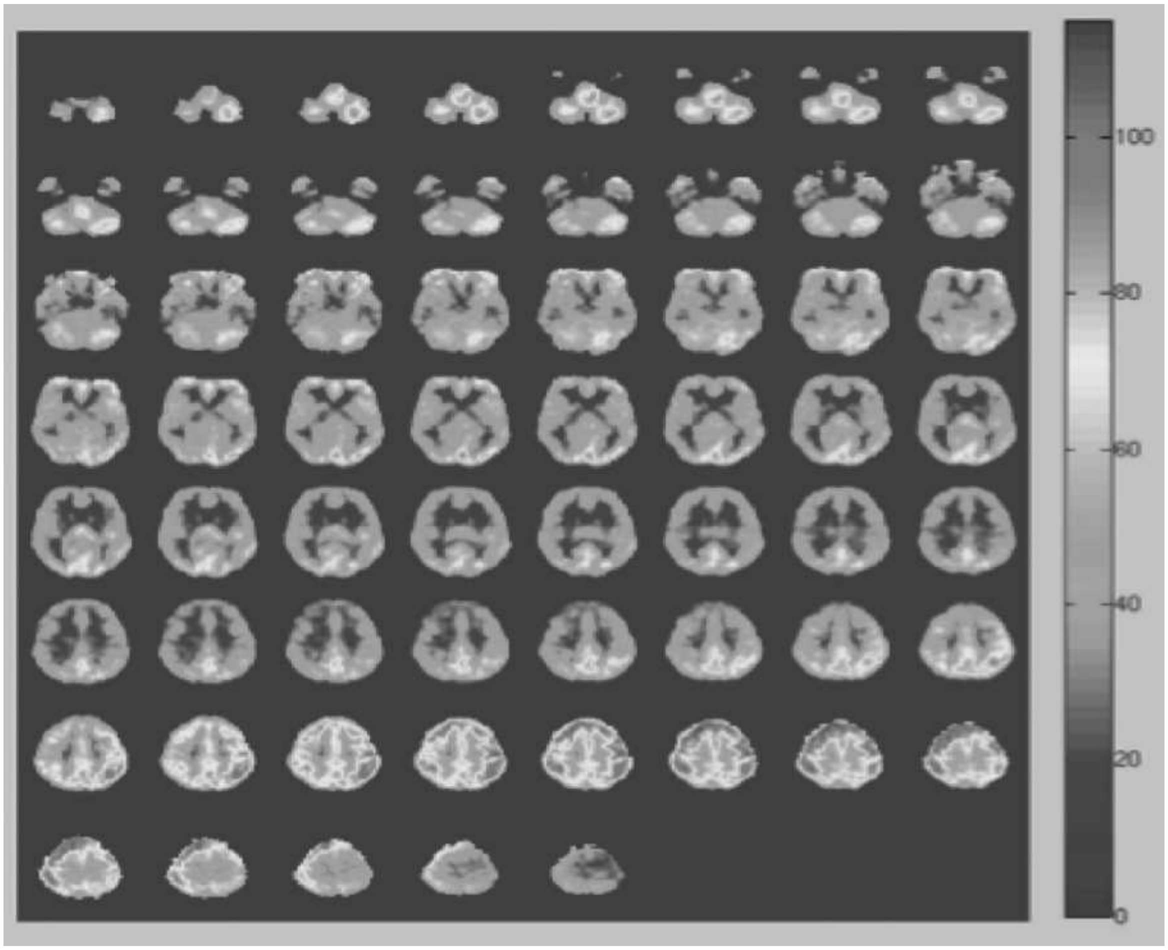

FIGURE 12.8–4. Map of quantitative blood flow obtained in a healthy individual with arterial spin labeling magnetic resonance imaging. (See eBook for color image.)

In addition to the whole brain analysis, a region-of-interest analysis can be performed to obtain the percentage signal change values for specific nodes in a hypothesized network. Predictor variables are a set of lagged indicators of stimulus type associated with each scan. Convolution with a priori hemodynamic response functions is not required. Resulting regression coefficients (β parameters) represent unbiased estimators of incremental hemodynamic responses at each time point, beginning at stimulus presentation and continuing for 16 seconds poststimulus. These data can be extracted for offline analysis.

Pattern Analysis and Classification. Voxel-by-voxel or region-of-interest-by-region-of-interest analysis of data is not able to capture correlations among structural and functional measurements at different brain locations. Therefore, spatially distributed and complex structural and functional patterns are better captured with high-dimensional pattern analysis and classification methods.

Connectivity Analysis. Although most fMRI studies have focused on examining brain regions activated by specific tasks, increased attention is being drawn to examining intercorrelations among brain regions. Such "connectivity analyses" can help identify the strength of association among nodes in the network and to establish in clinical populations whether deficits relate to reduced connectivity. Defined as the temporal correlations in neural activity among brain regions, functional connectivity provides a conceptual and methodologic framework for identifying spatially distributed patterns of brain activity and for quantifying inter-regional interactions. Several specific approaches have been implemented, and their relative merits are being investigated. The time series correlation method examines inter-regional correlations within individual subjects over the time course of an experiment and has been effectively applied to measure functional connectivity at a resting "default mode" state as well as across a wide range of cognitive and physiologic states.

More recently, graph theory has been applied to the analysis of functional connectivity. This approach considers the brain as consisting of nodes ("vertices") that communicate with each other, with an index of the strength of connectivity represented by edges. The values determining the significance of edges could be any index of relatedness, such as coordinated fluctuations in activity measured in changes in the BOLD signal time series. Graph theory permits the identification of networks and quantitation of the predominance and strength of within- and between-networks communications. For example, each node may show fluctuations that are synchronous not only with adjacent nodes but also with nodes that are in a homotopic region or nodes that belong to a different region entirely. Graph theory can calculate parameters of "community structure," such as modularity and cohesiveness. Graph theory has helped identify networks such as the default mode, salience, and cognitive control.

The methodologic developments summarized previously for measuring regional brain activation have been embraced by investigators because of the diverse parameters of brain structure and function they offer. Furthermore, MR scanning facilities are available in academic settings around the world. Consequently, numerous studies have applied neurobehavioral paradigms in fMRI studies aimed at dissecting complex behavior in schizophrenia. This has resulted in a large body of literature, with converging results that are contributing potential biomarkers that can advance the search for dimensional underpinnings of schizophrenia along the continuum of psychosis.

FUNCTIONAL MAGNETIC RESONANCE IMAGING IN SCHIZOPHRENIA

The literature on fMRI studies in schizophrenia has continued to grow with efforts to examine samples early in the course of psychosis and across psychotic disorders. Such efforts have capitalized on functional brain abnormalities identified in schizophrenia, including abnormal activity in response to motor tasks, working memory, attention, word fluency, emotion processing, and medication effects. These functional neuroimaging studies have suggested links between deficits in performance to abnormalities in activation patterns. Before summarizing such findings, it is important to outline some of the special challenges confronting the application of functional neuroimaging in schizophrenia. Although these challenges are not unique to functional imaging studies, they should be addressed systematically as the field matures.

Special Considerations in fMRI Applications in Schizophrenia

In addition to methodologic considerations that are common to all fMRI studies, several factors merit further consideration in the study of schizophrenia. These are likely to affect data interpretation and generalizability and relate broadly to several categories.

Clinical Variables

SAMPLE CHARACTERISTICS. The heterogeneity of schizophrenia can influence fMRI results. Samples vary in age, gender, and clinical status. Regarding sample size, most fMRI studies include relatively small samples that are not sufficiently powered to address pertinent clinical questions. Data on healthy people in fMRI studies are often obtained from university undergraduates, and small samples have sufficient power to detect task effects. However, in the study of schizophrenia, such samples could be marginally adequate to demonstrate differences between patients and healthy comparison participants but not informative for correlation with important clinical variables or for comparison with other clinical samples. The heterogeneity of schizophrenia may preclude adequate representation of some diagnostic subcategories, which can be further confounded by demographic characteristics. For example, patients are likely to be older than college undergraduates unless they are studied early in the course of the illness or have an early age of onset. Age and gender are important moderating variables in examining the study of brain and behavior, and they are even more so in schizophrenia. The age of onset is earlier in men, and premorbid functioning and outcome are worse. These considerations entail the recruitment of a comparison sample that is sociodemographically balanced to that of patients, and there must be a sufficiently large sample of patients to examine the possible effects of age and chronicity of illness on the measures obtained.

MEDICATIONS. Another relevant potential confound that needs consideration when applying functional neuroimaging in schizophrenia is the medication status of the participants. Unless studied at first presentation, when neuroleptic-naïve, patients are likely to be treated with psychoactive medications that affect brain function, and it is unclear how these effects will be reflected in BOLD signal changes. Short of controlled clinical trials, there are no simple and ethically acceptable ways to study or control potential medication effects. It is important to document medication status and examine, to the extent possible, whether they can account for the differences observed between patients and controls.

SYMPTOM SEVERITY. The clinical status of patients with schizophrenia is a factor that, in addition to influencing the feasibility of participation, may also compound the complexity of data interpretation. This factor is especially relevant when attempting to link imaging data to clinical measures of symptomatology. For example, men are likely to have more prominent negative symptoms than women, and, therefore, samples need to be adequately powered to clarify this potential confound. Obviously, the exclusion of acutely ill patients may limit the generalizability of data to patients who are clinically stable or who have more negative symptoms.

PERFORMANCE. A critical yet often overlooked issue in understanding the results of functional neuroimaging studies that involve neurobehavioral probes is how to deal with performance. The typical goal of tasks in functional imaging studies with healthy people is to activate a specific neural circuitry, and performance is usually required to be perfect or is used as a "nuisance" variable. When the aim is to identify a set of regions that get uniquely activated for a specific behavioral process or domain, making the task difficult will introduce individual differences that can diminish experimental effects and may engage extraneous regions related to frustration, anxiety, or motivation. This methodologic requirement for tasks that do not produce variable performance becomes problematic in studies in which the goal is to examine neural substrates for behavioral deficits. Thus, when comparing patients with schizophrenia to healthy controls on tasks where patients are putatively impaired, should it still be insisted that performance is perfect for patients? Such methodologic considerations have compelled some investigators to seek ways to equate difficulty levels when contrasting activation between patients and controls by making an easier version of the task for patients through titrating performance. This practice, however, may result in comparing activation maps between groups who received different tasks. Other approaches include changing the level of prescanning training to maximize patients' performance. However, such practices could result in group differences due to habituation.

Arguably, different paradigms are appropriate at different stages of the research. To establish whether the neural system under investigation engages the same regions in patients with schizophrenia as in controls, it would make sense, even when investigating a domain in which impairment is expected in patients, to make the task so easy that patients and controls will perform near perfection. Failure of patients to activate a specific region under these conditions will indicate that they do not recruit the requisite circuitry for that domain, which may explain why they fail when the going gets tough. However, in the next phase of the research, it would be important to make the task harder so that individual differences in activation could be correlated with individual differences in performance both in patients and in control groups.

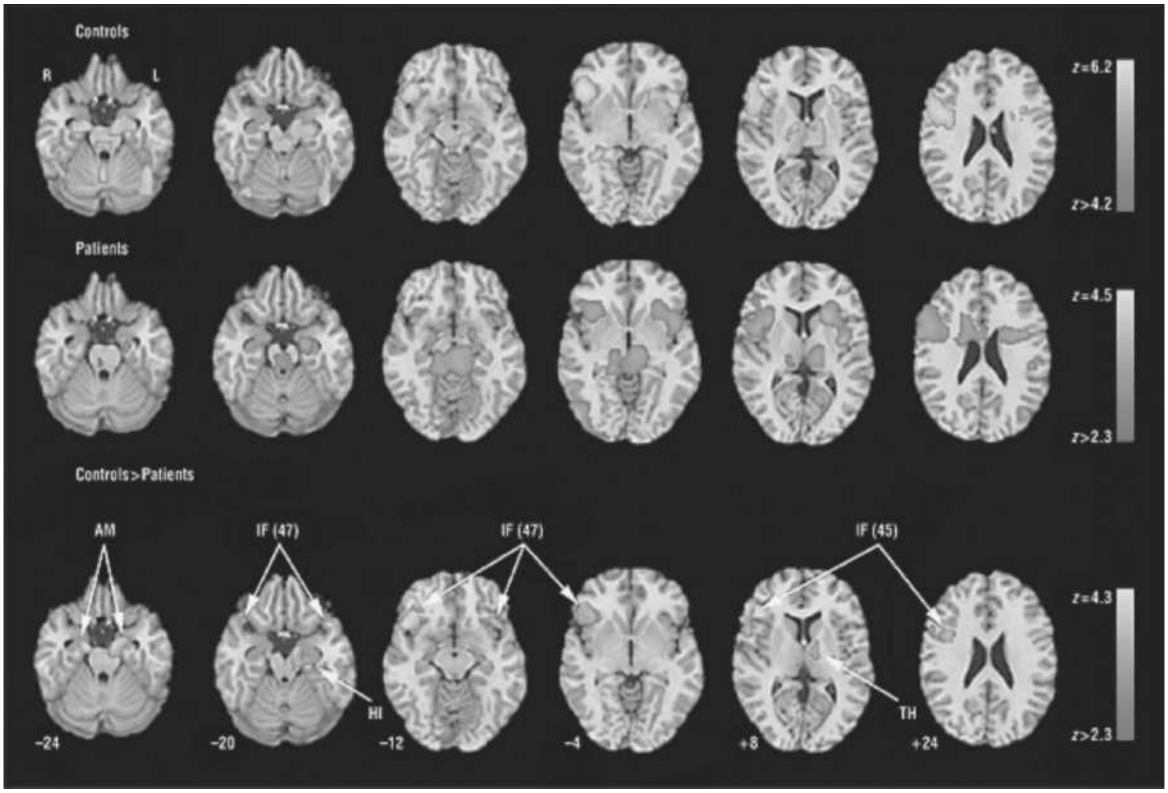

FIGURE 12.8–5. Regions activated for emotion identification task relative to baseline (block analysis) in controls (*upper row*), patients (*middle row*), and the controls–patients contrast (*bottom row*). No patients–controls contrast survived correction. Significance thresholds are based on spatial extent using a height of $z \geq 3.1$ and a cluster probability of $P \leq .05$. Images are displayed over a Talairach-normalized template in radiological convention (left hemisphere to viewer's right). The *z*-level coordinates are provided. AM, amygdala; IF (47), inferior frontal (Brodmann area 47); HI, hippocampus; IF (45), inferior frontal (Brodmann area 45); TH, thalamus. (Reproduced with permission from Gur RE, Loughead J, Kohler CG, et al. Limbic activation associated with misidentification of fearful faces and flat affect in schizophrenia. *Arch Gen Psychiatry*. 2007;64:1356–1366. Copyright © 2007 American Medical Association. All rights reserved. doi:10.1001/archpsyc.64.12.1356.)

Event-related paradigms can help address this issue to some extent because they enable the separation of correct from incorrect responses, thereby "leveling the playing field" with regard to performance. For example, on the emotion identification task described previously, a blocked analysis showed that patients with schizophrenia activated the same regions as controls but to a lesser degree (Fig. 12.8–5). An event-related analysis, however, showed a more dramatic difference when comparing activation time linked to correct and incorrect responses.

Activation was similar in patients and controls for correct responses. Yet, for threat-related expressions (angry and fearful faces), unlike controls in which greater limbic activation portended correct identification, in patients, greater activation in these regions was associated with incorrect responses (Fig. 12.8–6).

Major Findings in fMRI Studies in Schizophrenia

Cognition. Impaired cognition in schizophrenia has stimulated efforts to examine brain processes related to information processing cascades. Diffuse deficits have been documented, with relatively greater impairment in executive functions and in learning and memory. These deficits have been related to frontotemporal systems. Such impairments are core features of schizophrenia, important for elucidating underlying mechanisms, as buttressed by their prominence at initial illness presentation, limited amelioration with symptomatic relief, link to functional outcome, and utility as endophenotypes or biomarkers (quantitative intermediate phenotypic measures).

Neurobehavioral probes with functional imaging provide a powerful method for exploring the neural circuitry underlying such observed deficits. Abnormal activations in the ventromedial and superior temporal lobe, prefrontal cortices, and limbic structures were documented with memory and executive tasks. These complex processes could be compromised as downstream effects of sensory integration deficits.

Verbal learning deficits are well-established in schizophrenia. They are prominent at early clinical presentation, persist despite therapeutic intervention and amelioration of symptoms, and relate to functional outcomes. The relevance of learning and memory deficits in schizophrenia provides an impetus to integrate well-developed paradigms in cognitive neuroscience to advance the understanding of the neurobiology of these deficits.

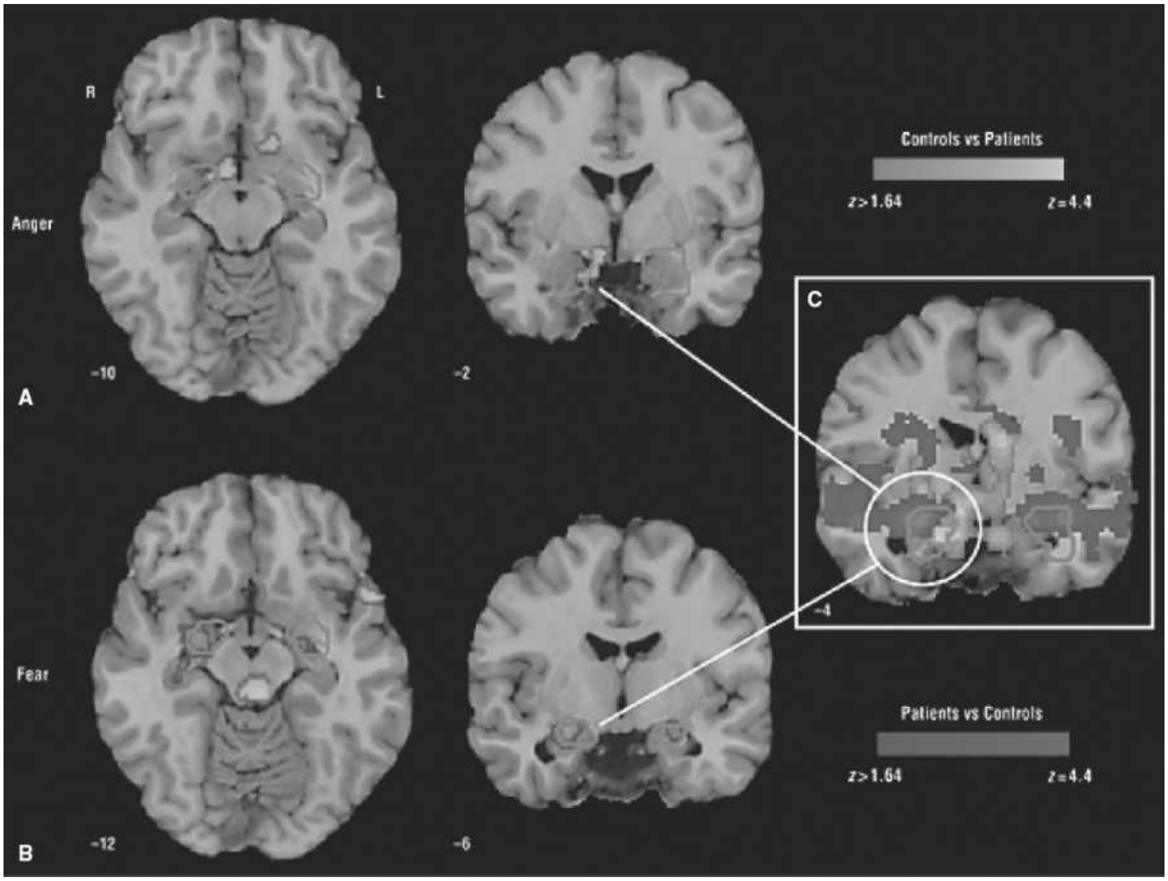

FIGURE 12.8–6. Activation maps showing peak amygdala response for anger (**A**) and fear (**B**) conditions for the event-related analysis. Images are displayed over a Talairach-normalized template in radiologic convention and thresholded at $z \geq 3.1$, uncorrected (>25 continuous voxels). Outline shows extent of atlas-derived amygdala regions of interest (Wake Forest University pickatlas) used for percentage of signal change extraction. Inset (**C**) highlights "patients versus controls" incorrect responses superimposed on "controls versus patients" correct responses at $z \geq 1.64$, uncorrected. (Reproduced with permission from Gur RE, Loughead J, Kohler CG, et al. Limbic activation associated with misidentification of fearful faces and flat affect in schizophrenia. *Arch Gen Psychiatry*. 2007;64:1356–1366. Copyright © 2007 American Medical Association. All rights reserved. doi:10.1001/archpsyc.64.12.1356.)

The application of fMRI enables researchers to examine distinct memory processes such as encoding and retrieval. Studies in patients, compared to healthy participants, have consistently demonstrated that during the learning phase, there are frontotemporal abnormalities. Most studies on schizophrenia have reported decreased activation of the frontal cortex, especially the inferior prefrontal region. The data are less consistent with regard to the temporal lobe. Most studies have observed decreased activation in patients' hippocampus and parahippocampal gyrus, but other studies have noted increased activation. As emphasized earlier, performance may be a contributing factor as the studies differ in task difficulty and their approach to performance variance.

Examining frontal lobe function, event-related fMRI studies have shown deficits in working memory and cognitive control in schizophrenia. Studies that focus on the temporal lobe in schizophrenia have progressed from language-processing tasks to the examination of the hippocampus and parahippocampal gyrus and probes evaluating long-term memory. Such reports noted that hippocampal dysfunction was often concomitant with aberrations in prefrontal function, suggesting that frontotemporal connectivity is disrupted in schizophrenia. As summarized previously, advances in image analysis methods enable researchers to measure the connectivity of distributed brain circuitry and test hypotheses implicating specific network models. Verbal episodic memory is differentially affected in schizophrenia, in part due to the use of inefficient organizational strategies. Indeed, providing a semantic organizational strategy during word encoding normalizes subsequent recognition accuracy in patients. This finding suggests that deficits in verbal memory may result in part from aberrant frontotemporal interactions, a hypothesis that fits well with the prevailing view of schizophrenia as a disorder of disturbed neural and psychological integration.

Supporting this hypothesis, functional neuroimaging studies in schizophrenia have revealed abnormalities in both frontal and temporal activities. Earlier PET studies showed that healthy people, when performing verbal tasks, activated prefrontal regions and deactivated the superior temporal gyrus. In contrast, patients with schizophrenia showed less superior temporal gyrus deactivation and a positive correlation between activity in the prefrontal and superior temporal gyrus regions. Further evidence of reciprocal changes in left frontal and left temporal regions was provided in fMRI studies of word fluency and in studies of word encoding and recognition.

Defined as the temporal correlations in neural activity among brain regions, functional connectivity provides a conceptual and methodologic framework for identifying spatially distributed patterns of brain activity and for quantifying inter-regional interactions. Several specific approaches have been implemented as outlined previously, and their relative merits have not been established. The time series correlation method examines inter-regional correlations within individual subjects over the time course of an experiment and has been effectively applied to measure functional connectivity across a wide range of cognitive and physiologic states.

In a study examining within-subject correlations of frontal and temporal time series during verbal encoding, aberrant frontotemporal connectivity was noted in schizophrenia, confirming previous reports. The study also identified distinct alterations within the dorsal and ventral prefrontal cortex. Relative to healthy controls, patients with schizophrenia had reduced connectivity between the dorsolateral prefrontal cortex and temporal lobe areas, including the parahippocampus and superior temporal gyrus. Patients, however, showed *increased* connectivity between a region of the ventrolateral prefrontal cortex and these same temporal lobe regions. Higher temporal–dorsolateral prefrontal cortex connectivity during encoding was associated with better subsequent recognition accuracy in healthy participants but not patients. Temporal–ventrolateral prefrontal cortex connectivity was uncorrelated with recognition accuracy in either group. The results suggest that reduced temporal–dorsolateral prefrontal cortex connectivity in schizophrenia could underlie encoding deficits, and increased temporal–ventrolateral prefrontal cortex connectivity may represent an ineffective compensatory effort (Fig. 12.8–7).

Notwithstanding extensive literature applying fMRI in schizophrenia, data on early information processing are more limited. A visual stimulation study demonstrated normal activation of visual, motor, somatosensory, and supplementary motor regions to a flashing checkerboard paradigm with a motor response. However, abnormal activation was reported for more complex sensory conditions, especially when requiring integration. A block design study presented simultaneous visual and auditory stimuli to first-episode neuroleptic-naive patients. Reduced activation in the parietal lobes, right thalamus, and prefrontal cortex, implicated in the dorsal visual processing pathway, was observed in patients. Diminished activation in patients in prefrontal regions implicated in regulating inhibition was reported in a study of the inhibitory P300 for a no-go condition. A three-stimulus auditory oddball task showed diffuse cortical and subcortical hypofunction during target detection and novelty processing in schizophrenia. Individuals with prodromal symptoms demonstrated smaller differential activation in frontal regions between relevant and irrelevant stimuli. In a visual oddball study, patients and healthy participants showed activation patterns specific to targets and novels, and activation of both neural systems was associated with faster performance. Specifically, reduced activation in regions involved in target and novelty processing in patients was accompanied by increased activation in circuits related to elaborated stimulus processing. For targets, abnormal activation was noted in regions related to ideational and visual association, and for novels, patients overactivated sensory and frontal areas related to visual–spatial processing and working memory (Fig. 12.8–8). Thus, the attenuated electrophysiologic response to targets seen in event-related potential studies may relate to insufficient top-down activation of target circuitry, while the attenuated-evoked response to novel distractors reflects overprocessing of bottom-up events. Notably, abnormal activation in patients was associated with more severe symptoms.

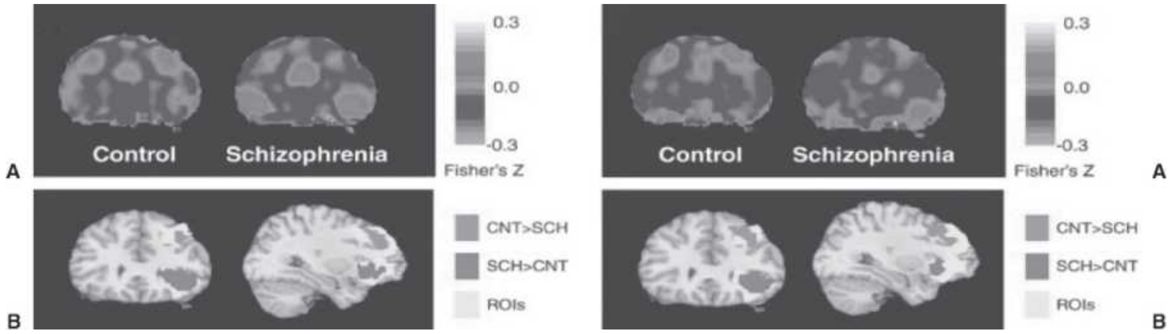

FIGURE 12.8–7. Group differences in left STG–DLPFC (superior temporal gyrus–dorsolateral prefrontal cortex) and STG–VLPFC (superior temporal gyrus–ventrolateral prefrontal cortex) connectivity (*left column*) and in left PHIP–DLPFC (parahippocampal gyrus–dorsolateral prefrontal cortex) and PHIP–VLPFC connectivity (*right column*). **A:** Left column, group-averaged STG time series correlation maps for control group (*left*) and schizophrenia group (*right*), from a representative coronal plane ($y = 30$). A region of DLPFC shows greater connectivity (measured as average Fisher's *Z*-transformed correlation coefficients) with STG in controls, while a region in VLPFC shows greater connectivity with STG in patients. **B:** Results of voxel-by-voxel *t* tests are overlaid on the two prefrontal regions of interests (ROIs) and a standard reference brain (Colin, MNI [Montréal Neurological Institute]), in the same coronal plane. The group differences seen within DLPFC and VLPFC are statistically significant. **A:** Right column, group-averaged PHIP time series correlation maps reveal a region within DLPFC, showing greater connectivity with PHIP in controls, while a region within VLPFC shows greater connectivity with PHIP in patients. **B:** Group differences seen within DLPFC and VLPFC are statistically significant. (Reprinted from Wolf DH, Gur RC, Valdez JN, Loughead J, et al. Alterations of fronto-temporal connectivity during word encoding in schizophrenia. *Psychiatry Res.* 2007;154(3):221–232. Copyright © 2007, with permission from Elsevier. doi: 10.1016/j.pscychresns.2006.11.008.)

Emotion. Social cognition, the cognitive processes involved in the perception, interpretation, and processing of social information, is increasingly recognized as a core domain of deficit in schizophrenia. Neurobiologic models of social cognition have confirmed that the processing of social information requires complex and synergistic interactions between several neural regions, which encompass the basic perception of social stimuli and the later evaluation of these stimuli.

The increased interest in affective neuroscience has likewise generated several tasks that probe emotion processing and social cognition. Such tasks may have relevance to schizophrenia, perhaps especially for understanding the intransigence of negative symptoms and poor social functioning. For example, functional imaging studies have demonstrated the role of limbic response in the identification of facial emotions in healthy people and diminished response in schizophrenia. As can be seen in the middle row of Figure 12.8–2, top-down (task-related) limbic activation is diminished in schizophrenia. However, patients showed abnormally increased limbic activation time-locked to the appearance of threat-related facial emotions of anger and fear. Furthermore, increased amygdala activation for fear was associated in patients both with failure to identify the emotion and with more severe flat affect (Fig. 12.8–9).

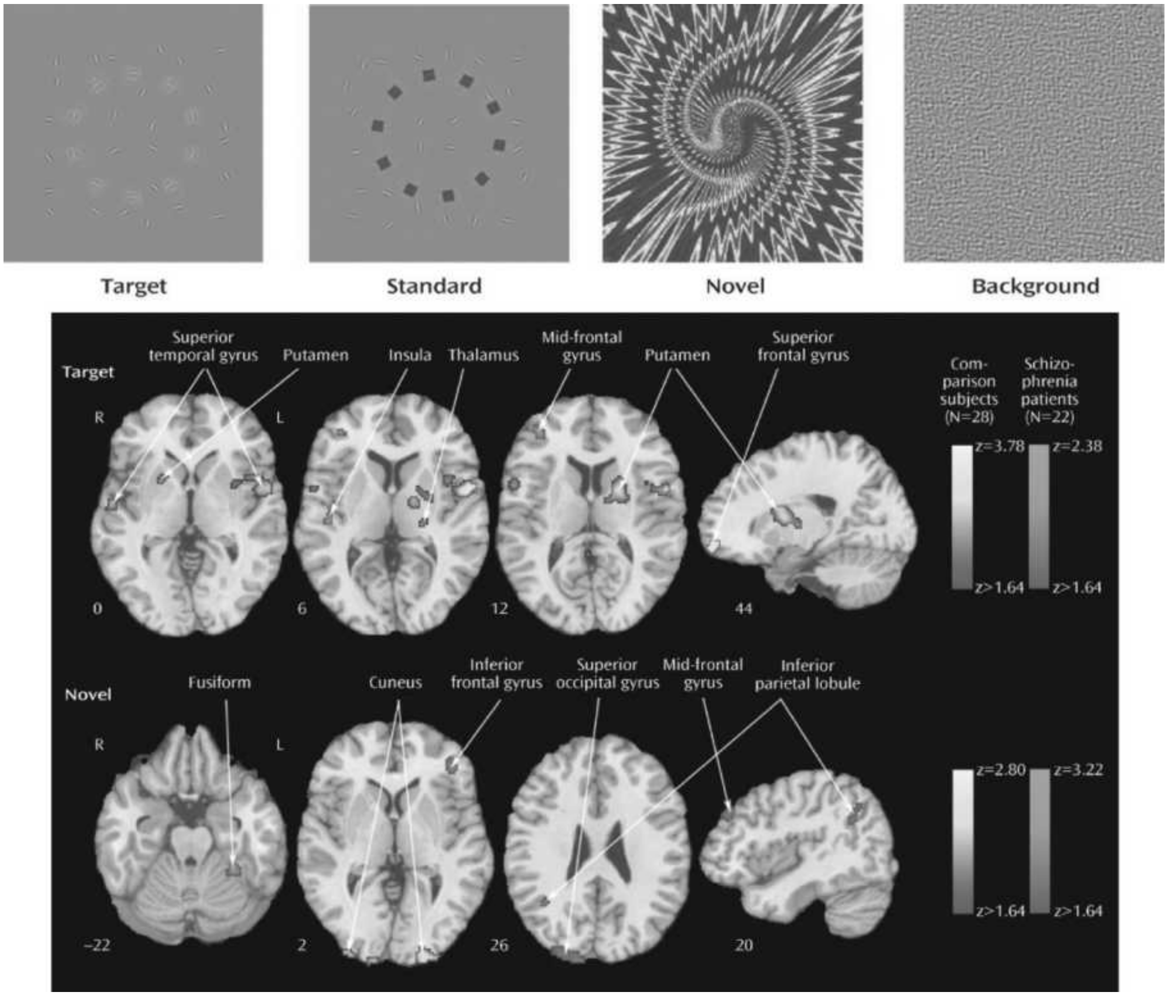
Target
Standard
Novel
Background
Target
Superior temporal gyrus
Putamen
Insula
Thalamus
Mid-frontal gyrus
Putamen
Superior frontal gyrus
Com-parison subjects (N=28)
Schizo-phrenia patients (N=22)
R
L
z=3.78
z=2.38
z>1.64
z>1.64
0
6
12
44
Novel
Fusiform
Cuneus
Inferior frontal gyrus
Superior occipital gyrus
Mid-frontal gyrus
Inferior parietal lobule
R
L
z=2.80
z=3.22
z>1.64
z>1.64
–22
2
26
20

FIGURE 12.8–8. Examples of stimuli used in “oddball” studies of attentional processing (*top*) and contrast images of patients with schizophrenia and comparison subjects for target and novel stimuli. Greater activation in patients is depicted by the scale on the right, whereas greater activation in comparison subjects is shown by the scale on the left. Images are in radiological convention (left hemisphere to viewer’s right). (See eBook for color image.) (Reprinted with permission from Gur RE, Turetsky BI, Loughead J, et al. Visual attention circuitry in schizophrenia investigated with oddball event-related functional magnetic resonance imaging. *Am J Psychiatry.* 2007;164(3):442–449. (Copyright © 2007). American Psychiatric Association. All Rights Reserved. doi: 10.1176/ajp.2007.164.3.442.)

Impaired emotional functioning is a prominent feature of schizophrenia. Although positive symptoms have traditionally attracted more attention and targeted treatment, negative symptoms, including flat affect, are increasingly recognized as more debilitating and resistant to intervention. Early studies reported impaired identification of facial affect in schizophrenia, but there was limited information on its relationship to clinical features and cognition.

Although there is growing literature in healthy people elucidating the role of the amygdala in affect processing, fewer functional neuroimaging studies have examined this circuitry in schizophrenia. Consistently in fMRI and PET studies with emotionally valenced pictures, faces, and odors, patients showed diminished activation in limbic and paralimbic regions, especially the amygdala. Abnormal activation of related cortical regions has also been observed. Functional MRI studies with block design report a lack of amygdala activation for sad mood induction, also evident in unaffected siblings. Decreased activation and accuracy in identifying expressions of fear, anger, and disgust contrasted with responses to mild happiness, especially in nonparanoids. Presentation of fearful and neutral faces showed a disconnection between brain activity and arousal in paranoid patients, where increased arousal was associated with decreased amygdala/medial prefrontal activity. In a study using a block design alternating between emotion and age identification, patients showed decreased activation in the left amygdala and bilateral hippocampus for the emotion condition. In a study examining intensity judgment, patients had an *exaggerated* amygdala response: Positive faces induced right amygdala activation in both groups, while negative faces activated only the right amygdala in controls and bilaterally in patients. Picture categorization (pleasant, unpleasant, neutral) produced less activity in amygdaloid–hippocampal and cortical–basal ganglia-thalamic circuitry in patients.

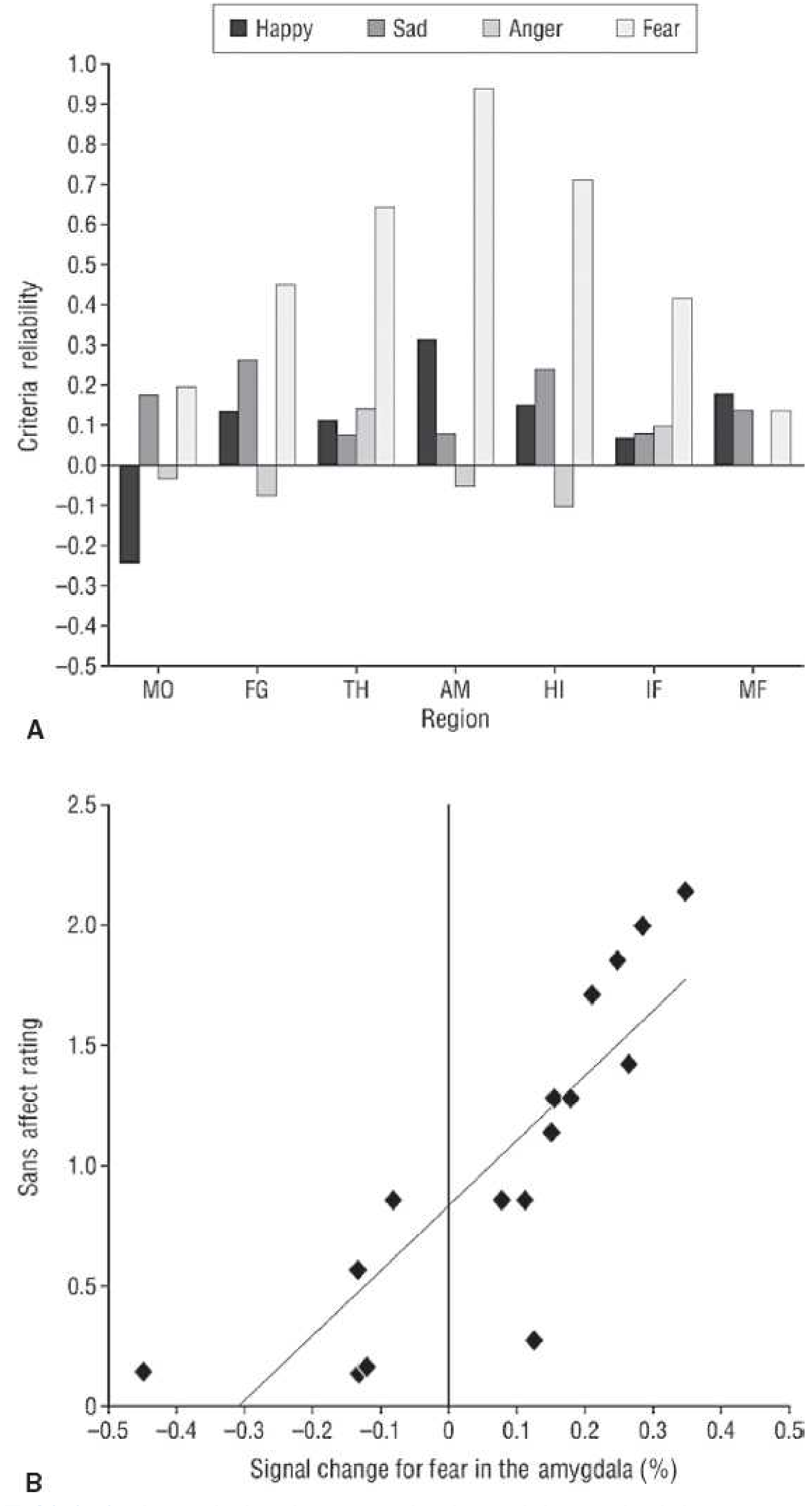

FIGURE 12.8–9. Association between brain activity and clinical measures. **A:** Correlations between event-related activation for the four emotional expressions in activated regions and severity of clinical ratings for flat affect.

B: Scatterplot of the association between percentage of signal change for the appearance of fear expressions and severity of flat affect. MO, midoccipital; FG, fusiform gyrus; TH, thalamus; AM, amygdala; HI, hippocampus; IF, inferior frontal; MF, midfrontal. (Reproduced with permission from Gur RE, Loughead J, Kohler CG, et al. Limbic activation associated with misidentification of fearful faces and flat affect in schizophrenia. *Arch Gen Psychiatry.* 2007;64:1356–1366. Copyright © 2007 American Medical Association. All rights reserved. doi:10.1001/archpsyc.64.12.1356.)

The growing fMRI literature complements PET studies that also showed abnormal cerebral blood flow in patients engaged in emotion-processing tasks. Pictures with aversive and nonaversive content activated the right insula in patients and controls, but patients showed greater activation of the medial prefrontal cortex for aversive stimuli. Furthermore, controls activated bilateral amygdaloid and orbitofrontal regions for nonaversive pictures, but patients activated only the left orbitofrontal cortex. Activity in the left amygdala correlated with positive symptoms. In a subsequent study, emotionally salient pictures produced abnormal activation in patients' ventral striatum and amygdala and increased tonic activation of the amygdala. Notably, a greater abnormality was associated with less severe positive and negative symptoms. The amygdala did not activate in patients for unpleasant pictures during valence categorization. Areas of decreased cerebral blood flow included the prefrontal cortex and subcortical regions such as the thalamus and cerebellum.

These studies of emotion processing in schizophrenia vary in methodology and design. Nonetheless, there seems to be considerable convergence of evidence that patients show abnormal activation in the amygdala and associated regions. It also appears that the abnormalities are more pronounced for negatively valenced stimuli, although no further differentiation for specific emotions has been established. In most studies, patients' performance was carefully examined, and tasks were often constructed so as to minimize performance differences to avoid confounding physiologic measures. In all cases in which performance was evaluated, it did not explain the difference in activation patterns. However, reduced task-related (top-down) activation could reflect bottom-up interference from abnormally increased amygdala activation related to stimulus valence. Thus, if the task is to identify emotions, the sensitivity of the amygdala to specific stimuli may disrupt the cortical processes required for categorization and response. Increased activation associated with incorrect responding was also seen for faces expressing anger. Thus, it appears that threat-related faces induce greater activation in patients in a way that interferes with the top-down process, thereby producing errors. Whenever this response differed between patients and controls, it indicated more error-associated activation for fear or anger in patients.

Bottom-up interference in cognitive functioning likely operates by disrupting the coordinated activation of regions required for its support. Although there is considerable work on conditions associated with task activation, recent efforts have been aimed at studying inter-regional connectivity. These studies have used structural equation modeling that permits the assessment of whether specific causal models linking activation of one region to another are not rejected by the data. To examine connectivity, it is useful to consider the pattern of correlations among activation parameters obtained from regions recruited for a specific task. These correlations would reflect the extent to which activation in one region is proportional to that in another. Controls have high-specific connectivity for activation to the top-down task and for bottom-up activation associated with correct responding. Patients, by contrast, show abnormally high connectivity associated with activation to threat-related stimuli, leading to incorrect responses. Thus, disrupted operation of a top-down process in schizophrenia could be linked to excessive connectivity for bottom-up amygdala activation related to stimulus valence. Such disruptive activation could not only reduce connectivity among regions participating in the top-down process but also trigger the recruitment of other regions, including cortical ones, that may further disrupt task-related processing. Therefore, abnormal amygdala activation will lead to dysfunctional activation in other regions involved in information processing and disrupt downstream processes involving complex behavior. This will manifest in cognitive deficits.

Ultimately, the importance of abnormal activation depends on whether it relates to disease variables. This question has received less attention, and results diverge. Studies show abnormally increased amygdala activation associated with more severe positive symptoms, both positive and negative symptoms, or less severe symptoms. It may be that the direction of such correlations depends on whether the abnormality is that of over- or underactivation and whether the activation potentiates symptoms or reflects a compensatory process. Although correlations between activation parameters and symptom severity are small or moderate, abnormally increased bottom-up amygdala activation for fearful faces was correlated highly with increased severity of affective flattening. Determining such correlations could help guide efforts to ameliorate symptoms with knowledge of regions where activation change can be palliative.

A likely specific consequence of abnormal bottom-up activation of the amygdala is the disruption of memory processes. There is considerable evidence that the amygdala interacts closely with the hippocampus in the formation of episodic memory, not only when affective valence is prominent. This area needs further investigation.

The convergence of findings pointing to the amygdala and its connectivity as a likely source for the devastating effects of schizophrenia propels a search for factors that could modulate amygdala response. Fortunately, there is an expansion of rigorous studies with fMRI in healthy people that have attempted to identify such modulators in efforts to construct mechanistic models of affect regulation. Emerging findings from meta-analyses support the notion of a fear-specific response of the amygdala and partially support the laterality hypothesis with increased left amygdala activation across tasks. Potential contributions of modulating processes such as attention, habituation, arousal and anxiety, learning, and extinction have been examined. Overall, while the amygdala response to fearful expressions can be influenced by such factors, the response is quite robust and specific for fear and could be neither explained nor completely extinguished by such manipulations. There is evidence that fear-related stimuli are especially potent activators of the amygdala, although cognitive processing modulates stimulus effects. The amygdala's extensive connections with subcortical regions are likewise essential for its role in affect processing. It has been proposed that rapid visual input from a subcortical thalamic pathway provides an early alarm signal for threat detection, whereas slower input through a geniculostriate pathway enables a more detailed evaluation of environmental input. This slower pathway could explain why amygdala activation seems to habituate rapidly in passive viewing but shows no signs of habituation when the task involves top-down processing where affective valence is relevant.

Studies attempting to elucidate mechanisms underlying fear response have often used faces to link with animal evolutionary approaches. Affective information on the face was systematically manipulated by methods including spatial frequency, binocular suppression, chimeric faces, masked fearful eye whites, and gaze direction. These methods have suggested that amygdala response can be potentiated or suppressed by such factors. For example, the amygdala response is stronger when information is presented in low rather than high-spatial frequency, and the effect is especially pronounced for fear in the upper part of the face, presumably the eye region. Notably, bilateral amygdala lesions specifically diminish a patient’s ability to utilize fear-related information in the eye region and direct the patient to note that the eyes could temporarily reverse this deficit. This finding is encouraging in that rehabilitation could possibly ameliorate deficits.

Gaze direction is another modulator of amygdala response. In healthy people, directed fear and averted anger expressions produced greater amygdala responses than averted fear and directed anger. This effect was interpreted to suggest that the amygdala is especially sensitive to threat-related ambiguity. Since poor eye contact is a major feature of impaired social interactions in schizophrenia, gaze direction, and evaluation of the patient’s ability to use information in the eye region seem the most appropriate target for the next phase of the investigation. Although scan path studies reported a restricted range of visual scanning in schizophrenia for happy, sad, and neutral faces, these studies have not indicated a difference between the eyes and the mouth regions. However, fear and anger stimuli have not been used. Because amygdala damage leads to avoidance of the eyes region in fearful faces, and the eyes region is important for distinguishing anger from fear, the difficulty of patients in identifying fear could be related to impairment in recognizing changes in the eyes region, thereby diminishing amygdala response to fear in the eyes relative to the mouth region. As with other bottom-up activation abnormalities, it is expected that the severity of these abnormalities is associated with poorer eye contact and affective flattening.

Pharmacology. The literature on the possible effects of antipsychotic agents on BOLD signal change is limited. Few studies have examined neuroleptic-naive patients, let alone in a repeated measures design. Furthermore, small samples and methodologically limited designs have precluded systematic examination of the possible effects of therapeutics. Nonetheless, the recognition that impaired cognition is a hallmark of schizophrenia and that fMRI enables researchers to examine the brain systems underlying the aberrant functioning has extended studies of cognition to fMRI.

Examples of studies include a working memory evaluation with an n-back paradigm after patients switched from first-generation antipsychotics to second-generation agents. Increased dorsolateral prefrontal cortex and parietal cortex activity were reported in the patients with the new treatment. Similarly, normalization of brain activity was reported in patients treated with long-acting risperidone (Risperdal) compared to conventional depot medication while performing an n-back task. Normalization of prepulse inhibition, which is impaired in schizophrenia, was also noted in patients treated with olanzapine (Zyprexa) and risperidone compared to those on first-generation antipsychotics. Studies that examined motor control in relation to medications have also noted improved brain activity in patients treated with second-generation agents compared to first-generation antipsychotics.

The well-established literature on the heritability of endophenotypic brain behavior measures in unaffected family members of individuals with schizophrenia can provide a path to investigate the potential effect of psychoactive medications on brain function in unmedicated family members. For example, an abnormally strong alprazolam-induced reduction in amygdala and hippocampus activation during emotion identification was observed in first-degree family members relative to healthy controls, compared to an equal reduction in both groups for the emotion memory task. An aberrant response may be related to unmasking underlying amygdalar and GABAergic abnormalities. Such pharmacologic fMRI paradigms could prove useful for developing drugs targeting specific neural circuits in schizophrenia.

Given the limited number and scope of the available literature, there is no conclusive evidence yet, as double-blind studies are lacking, and most reports consist of cross-sectional design. Thus, it is difficult to conclude that the reported “normalizing” effects of the medications are indeed attributable to specific medication effects and are not part of clinical stabilization or practice effects. Furthermore, as no absolute measures of brain activity are provided with BOLD, the relative change needs to be interpreted with caution. Pharmacologic fMRI is a potentially powerful methodology when integrated with well-designed and informative pharmacologic designs.

Genetics. The genomic architecture of schizophrenia has been investigated in large samples that have relied on the clinical phenotype with more recent efforts, including endophenotypes. As detailed in other chapters, it has become increasingly clear that this heritable, heterogeneous disorder does not have shared genetic contributions of relatively few genes. On the contrary, as more research emerges there seems an ever-growing list of foci, each with a small effect, that is associated with the disorder.

The application of fMRI in genetic paradigms is more limited relative to structural MRI. Initial studies have examined genetic effects indirectly by comparing activation patterns in probands with schizophrenia, unaffected siblings, and healthy comparison subjects with no family history of schizophrenia. Such studies have demonstrated, for example, that there were abnormalities in siblings that were less severe than those seen in affected individuals. Such findings support the application of fMRI as a quantitative phenotypic marker of schizophrenia.

The availability of sequencing of the human genome offers unprecedented opportunities for direct evaluation of the effects of genetic variability on brain activity. Early work exploiting this potential has demonstrated such effects in healthy people by comparing activation patterns between genotypically characterized groups. Studies applying genetic strategies used functional polymorphisms to group individuals for comparisons. For example, a common Val108/158Met substitution in the gene for catechol-O-methyltransferase (COMT) leads to decreased activity of this enzyme in dopamine catabolism and has been linked to decreased prefrontal cortical activity. Studies, therefore, examined val/met polymorphism and prefrontal cortex activation. Individuals homozygous for the met allele had diminished prefrontal and hippocampal engagement while performing episodic memory encoding and retrieval compared to val/val subjects. Notably, however, a recent meta-analysis of fMRI working memory studies in people with schizophrenia and healthy controls did not find evidence for an impact of COMT Val158Met genotype on brain activation. Diverse analytic approaches and small sample sizes are likely to contribute to inconsistent findings.

Using a similar approach, a risk haplotype was examined in *GRM3,* a gene encoding a metabotropic glutamate receptor. The findings were of reduced neuronal function in the prefrontal cortex and impaired activation in the hippocampus during the performance of a verbal memory task. The risk allele in the promoter region was associated with decreased activation of the prefrontal and temporal lobe cortex.

The incorporation of functional neuroimaging into genomic studies has the potential for constructing mechanistic models for the pathophysiology of schizophrenia but poses additional challenges. Since the expected effects of any single polymorphism are small, it is essential to pay close attention to experimental design in neuroimaging genomics. It is important to use well-characterized neurobehavioral probes to evaluate and control confounding variables such as age, gender, basal abilities, performance, and clinical status. In addition, neuroimaging genomic studies require large samples that can optimally be obtained by multisite studies. Such efforts introduce additional variability related to scanner characteristics and image acquisition. Nonetheless, the field is positioned to make new discoveries that will illuminate neural pathways to specific features of schizophrenia in ways that will lead to novel interventions.

Functional Neuroimaging in Clinical At-Risk Samples

The investigation of schizophrenia as a brain disorder began with studies that evaluated chronic patients, establishing neurocognitive deficits associated with abnormalities in parameters of brain structure and function. With the increased availability of noninvasive neuroimaging methodology, studies of younger patients became feasible, and this work indicated that brain abnormalities and neurocognitive deficits are already evident in first-episode individuals. This work has produced two shifts of emphasis in the study of psychosis that can advance the understanding of the underlying neurobiology. The first is a shift to study even earlier phases of psychosis–the clinical risk stage–the second is a shift from dichotomous symptom categorization to quantitative dimensional endophenotypes, which can be examined across psychosis. This development will benefit genomic investigations because such vulnerability indices are heritable brain and neurobehavioral parameters that are trait markers related to illness.

Functional MRI has been applied to individuals at risk for psychosis, commonly in small samples of people at genetic risk because of family history or help-seeking individuals with attenuated psychosis spectrum symptoms. Neurobehavioral probes applied were those that have shown differences between patients with schizophrenia and controls such as working memory using the n-back paradigm. Overall, psychosis risk groups show decreased activation in the BOLD response in the dorsolateral and medial prefrontal regions. The pattern of activity is similar to that seen early in the course of schizophrenia, but less pronounced abnormalities are observed. Similar to data obtained from patients with schizophrenia, individuals at risk show *increased* amygdala activation for the identification of threat-related expressions (Fig. 12.8–10).

To evaluate activation changes with disease progression, longitudinal designs are necessary. Such designs have been applied in several fMRI studies. This small literature suggests that individuals who transition to psychosis differ from those who do not, with the latter group showing normalization of brain activity. Thus, the application of fMRI holds promise as a tool that may facilitate elucidating the brain circuitry dysfunction underlying the psychotic process.

Furthermore, prior fMRI studies of psychosis risk have typically examined only categorical effects, lacking adequate power to identify correlates of illness dimensions. Therefore, it remains unclear to what extent the reported abnormalities relate to positive psychotic symptoms that are used to select the samples versus negative symptoms or cognitive deficits that are likely also present.

Dysconnectivity of resting BOLD has been reported in several brain networks in adults with schizophrenia, including the default mode and the cingulo-opercular network. More recently, youth at risk for psychosis showed multifocal dysconnectivity, implicating the bilateral anterior cingulate, frontal pole, medial temporal lobe, opercular cortex, and right orbitofrontal cortex. These results were driven by hyperconnectivity among default mode regions and diminished connectivity among cingulo-opercular regions, as well as diminished coupling between frontal and default mode regions. Thus, functional dysconnectivity in psychosis spectrum youths corresponds to abnormalities reported in adults with established psychotic disorders.

Challenges for Functional Neuroimaging Research in Schizophrenia

Considerable advances with functional neuroimaging technologies have been made in efforts to understand the neurobiology of schizophrenia. fMRI has become the dominant method to examine brain systems because it provides noninvasive measures with good anatomic resolution and increasingly reliable quantitation. Abnormalities in schizophrenia have been documented at many levels of neurobehavioral and neurophysiologic processing and in multiple neural systems. However, these diffuse abnormalities, which implicate neurodevelopmental aberrations, are not fully expressed as a disorder until after brain development and maturation are largely completed.

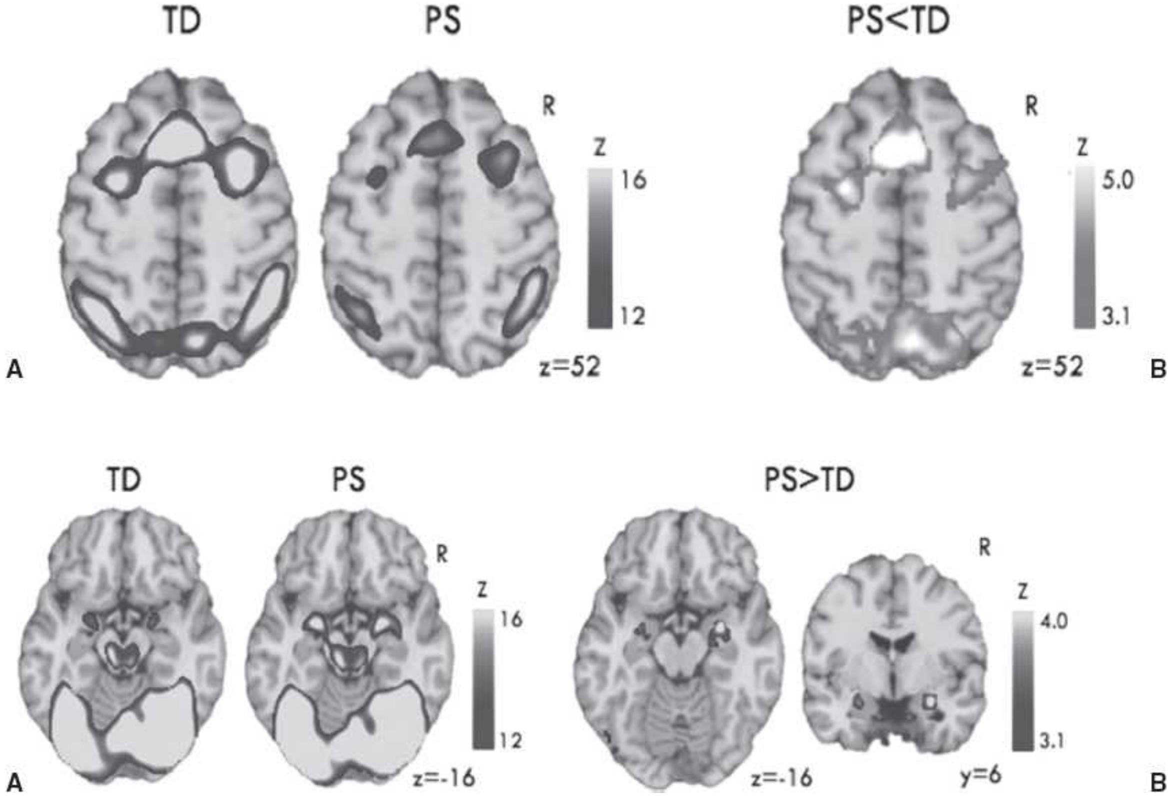

FIGURE 12.8–10. Aberrant activation in youth with psychosis spectrum symptoms. Top panel shows the working memory n-back task effect and bottom panel shows effects of emotion identification task. For top panel: **A:** The contrast evaluating the effect of working memory load (2-back greater than 0-back) robustly recruited executive function regions, including dorsal anterior cingulate and bilateral dorsolateral prefrontal cortex, with a qualitatively similar regional pattern in both the psychosis spectrum (PS) and typically developing (TD) groups. **B:** The PS group showed significantly reduced activation in these regions (significant clusters shown in light gray). R indicates right. For bottom panel: **A:** Facial emotion-processing activated amygdala and other regions implicated in emotion and face processing, with a qualitatively similar regional pattern in both the PS and TD groups. **B:** The PS group showed significantly greater activation in bilateral amygdala (significant clusters shown in black). R indicates right. (Reproduced with permission from Wolf DH, Satterthwaite TD, Calkins ME, et al. Functional neuroimaging abnormalities in youth with psychosis spectrum symptoms. *JAMA Psychiatry.* 2015;72(5):456–465. Copyright © 2015 American Medical Association. All rights reserved. doi:10.1001/jamapsychiatry.2014.3169.)

The recognition that the underlying pathologic processes of schizophrenia are present throughout the central nervous system (CNS) suggests that there may be some core defects in neurotransmission whereby the ability of neurons to respond to and process stimuli is curtailed. However, within the context of global abnormalities, some neural systems are differentially affected. Brain regions with greater vulnerability, such as the hippocampus, prefrontal cortex, and olfactory system, are dynamic and maintain a high degree of plasticity during brain maturation and thereafter. Thus, neuronal plasticity, evident in molecular, chemical, morphologic, and physiologic modifiability with neural activity, is abnormal in schizophrenia. This may ultimately be manifested in the array of complex features that characterize schizophrenia. However, functions mediated through other neural systems, where greater plasticity is not required, show subtle abnormalities, and motor abilities, sensation, and autonomic functions are relatively spared.

These challenges are further compounded by the absence of suitable animal models for the disorder as phenotypically defined, buttressing the need to identify endophenotypes that can be studied in both humans and animals. For example, sensorimotor gating and encoding can be characterized in animals and in humans, as can fear conditioning and affiliative behavior. The identification of such quantitative phenotypes in schizophrenia that have their counterparts in rodents enables researchers to investigate the genetic, molecular, biochemical, anatomic, and physiologic aspects of these behaviors in ways not possible in humans. Findings from studies in rodents will provide new information about these behaviors and elucidate the effects of genetic, pharmacologic, and behavioral manipulations, which can then be pursued in humans. This top-down–bottom-up interchange holds great promise for advancing the understanding of the pathophysiology of schizophrenia, where functional neuroimaging provides a bridge between the molecular and the system levels.

FURTHER READINGS

Anticevic A, Repovs G, Barch DM. Working memory encoding and maintenance deficits in schizophrenia: neural evidence for activation and deactivation abnormalities. *Schizophr Bull*. 2013;39:168–178.

Carter CS, Barch DM. CNTRICS Executive Committee: imaging biomarkers for treatment development for impaired cognition: report of the sixth CNTRICS meeting: biomarkers recommended for further development. *Schizophr Bull*. 2012;38:26–33.

*Dickinson D, Straub RE, Trampush JW, et al. Differential effects of common variants in SCN2A on general cognitive ability, brain physiology, and messenger RNA expression in schizophrenia cases and control individuals. *JAMA Psychiatry*. 2014;71:647–656.

Ford JM, Palzes VA, Roach BJ, et al.; Functional Imaging Biomedical Informatics Research Network. Visual hallucinations are associated with hyperconnectivity between the amygdala and visual cortex in people with a diagnosis of schizophrenia. *Schizophr Bull*. 2015;41:223–232.

*Fornito A, Bullmore ET. Connectomics: a new paradigm for understanding brain disease. *Eur Neuropsychopharmacol*. 2015;25:733–748.

*Fusar-Poli P, Borgwardt S, Bechdolf A, et al. The psychosis high-risk state: a comprehensive state-of-the-art review. *JAMA Psychiatry*. 2013;70:107–120.

Grimm O, Heinz A, Walter H, et al. Striatal response to reward anticipation: evidence for a systems-level intermediate phenotype for schizophrenia. *JAMA Psychiatry*. 2014;71:531–539.

Gur RE, Gur RC. Functional magnetic resonance imaging in schizophrenia. *Dialogues Clin Neurosci*. 2010;12:333–343.

Hager BM, Keshavan MS. Neuroimaging biomarkers for psychosis. *Curr Behav Neurosci Rep*. 2015;2015:1–10.

*Libby LA, Ragland JD. FMRI as a measure of cognition related brain circuitry in schizophrenia. *Curr Top Behav Neurosci*. 2012;11:253–267.

Lui S, Yao L, Xiao Y, et al. Resting-state brain function in schizophrenia and psychotic bipolar probands and their first-degree relatives. *Psychol Med*. 2015;45:97–108.

*Nickl-Jockschat T, Janouschek H, Eickhoff SB, Eickhoff CR. Lack of meta-analytic evidence for an impact of COMT Val158Met genotype on brain activation during working memory tasks. *Biol Psychiatry*. 2015;78(11):e43–e46.

*Röder CH, Dieleman S, van der Veen FM, Linden D. Systematic review of the influence of antipsychotics on the blood oxygenation level-dependent signal of functional magnetic resonance imaging. *Curr Med Chem*. 2013;20:448–461.

Satterthwaite TD, Vandekar SN, Wolf DH, et al. Connectome-wide network analysis of youth with psychosis-spectrum symptoms. *Mol Psychiatry*. 2015;20(12):1508–1515.

*Taylor SF, Kang J, Brege IS, Tso IF, Hosanagar A, Johnson TD. Meta-analysis of functional neuroimaging studies of emotion perception and experience in schizophrenia. *Biol Psychiatry*. 2012;71:136–145.

Wolf DH, Satterthwaite TD, Calkins ME, et al. Functional neuroimaging abnormalities in youth with psychosis spectrum symptoms. *JAMA Psychiatry*. 2015;72:456–465.

Wolf DH, Satterthwaite TD, Loughead J, et al. Amygdala abnormalities in first-degree relatives of individuals with schizophrenia unmasked by benzodiazepine challenge. *Psychopharmacology (Berl)*. 2011;218:503–512.

*Yan C, Yang T, Yu QJ, et al. Rostral medial prefrontal dysfunctions and consummatory pleasure in schizophrenia: a meta-analysis of functional imaging studies. *Psychiatry Res*. 2015;231:187–196.

▲ 12.9 Molecular Brain Imaging in Schizophrenia

Dean F. Wong, M.D., Ph.D., Ganesh B. Chand, Ph.D., Gerhard Gründer, M.D., Jimmy E. Jakobsson, Ph.D., and Albert Gjedde, M.D., D.Sc.

Molecular brain imaging is the visualization of in vivo brain chemistry in humans and animals and has undergone remarkable growth since the first imaging of neuroreceptors in humans four decades ago. Molecular brain imaging uses radiologic techniques that present unique opportunities to study disease mechanisms and therapeutic options in people with schizophrenia (see Table 12.9–1 for a glossary of terms). Additionally, molecular brain imaging is a powerful tool for the documentation of animal models that simulate the symptoms and signs of schizophrenia as part of parallel studies of animals and humans. Furthermore, molecular brain imaging provides the means to determine the effects of therapeutic interventions on variables such as receptor occupancy in animal models, healthy adult human volunteers, and people with schizophrenia.

MOLECULAR BRAIN IMAGING AND THEORIES OF SCHIZOPHRENIA

Arvid Carlsson's *dopamine hypothesis of schizophrenia* focused on the hyperactivity of the subcortical dopamine system as the basis of the pathogenesis and pathophysiology of schizophrenia. The hypothesis is closely associated with the original recognition of the role of dopamine independent of its additional role as a precursor of norepinephrine some 50 years ago. The dopamine hypothesis originated from the finding that traditional neuroleptic drugs such as chlorpromazine (Thorazine) antagonize dopamine activity by blocking postsynaptic D_2-like dopamine receptors. Subsequent research into schizophrenia provided the basis of additional hypotheses, including the serotonin hypothesis of schizophrenia, substantiated by studies of the mechanisms of action of atypical second-generation antipsychotics and, more recently, the glutamate hypothesis of schizophrenia. Molecular imaging provided the means to test these hypotheses directly in living humans with schizophrenia. Consequently, the original dopamine hypothesis that schizophrenia results from excessive dopaminergic activity required expansion to reflect contributions from dysfunction of serotonergic, glutamatergic, and several other neurotransmitter systems.

The study of molecular imaging in schizophrenia has a historical root in the very basis of neuroreceptor imaging with positron emission tomography (PET) and single-photon emission computed tomography (SPECT). Indeed the first imaging study in 1983 of a living human brain used the antipsychotic drug spiperone (Spiroperidol), first labeled with ^{11}C from [^{11}C]iodomethane and later with fluorine-18 (^{18}F). This advance was followed by the isotopic drug [^{11}C]raclopride (no difference of chemical structure or pharmacology), also an antipsychotic as well as a D_2 and D_3 antagonist that had been studied in clinical trials of humans. Although neither antipsychotic remained in use clinically in the United States, the radiolabeled PET analogs both led the way to the study of D_2-like dopamine receptors and the entire field of investigation of neuroreceptors as an essential new subspecialty of neuroimaging with tremendous application in neuropsychiatry. Both radiotracers quickly were shown to be replaceable by labeled versions of common antipsychotics such as haloperidol (Haldol), which ultimately led to occupancy studies and therapeutic drug dose levels now being visualized in the human brain.

The initial dopamine studies raised the interest in the serotonin system which is now held to be important to some aspects of schizophrenia symptomatology and potentially to treatment. Indeed, starting with clozapine (Clozaril) and subsequent later-generation antipsychotics such as risperidone (Risperdal) and olanzapine (Zyprexa), many atypical antipsychotics have both serotonergic and dopaminergic receptor-binding properties. Hence, the development of molecular imaging of the serotonin sites was a natural progression.

More recently, a “glutamate hypothesis” of schizophrenia emerged, as an extension of the dopamine hypothesis. The prevailing theory involves a hypoglutamatergic state of frontal cortical regions in subjects with schizophrenia, leading to a hyperdopaminergic state of the striatum. The glutamate hypothesis helps explain the evidence of psychostimulant (e.g., amphetamine) induced dopamine release (DAR), although substantial evidence for or against it remains lacking. Similarly, the actions of *N*-methyl-D-aspartate (NMDA) antagonists such as ketamine (Ketalar) that affect a phencyclidine (PCP) site (an ion channel) open the possibility of measuring DAR changes in schizophrenia and are now actively pursued.

Thus, neuroreceptor imaging as a technical achievement of neuropsychopharmacology and radioisotope methodology both paralleled and benefitted from research into the pathophysiology and treatment of schizophrenia. Once it is possible to study the glutamate, dopamine, serotonin, and other important systems in the same individuals and similar cohorts, the combined practices of neuroreceptor imaging will continue to help to continue to push forward the frontiers of schizophrenia research.

Synergy has developed among, on one side, the molecular imaging by PET or SPECT of radioisotopes of drugs, as well as by magnetic resonance spectroscopy

(MRS), structural (sMRI), and functional magnetic resonance imaging (fMRI) and, on the other side, our current knowledge of schizophrenia, including its pathophysiology and treatment. The strength of MRS is the ability to quantify amino acids (such as γ-aminobutyric acid [GABA] and glutamate). These magnetic resonance approaches complement the information obtained by the PET and SPECT molecular imaging records of signals from neurotransmitter metabolism and transporter and receptor activity that is evident in the brain, albeit the latter at lower levels of resolution. It is important to consider the contribution of changes in brain structure (sMRI) in the interpretation of neurochemical and functional imaging data methods (fMRI). Thus, the PET, SPECT, and MR methods are unique sources of information applicable to the resolution of specific research questions relevant to the pathophysiology and potential treatment of psychiatric diseases and disorders.

Table 12.9–1.
Glossary of Terms, Abbreviations, and Definitions for Molecular Brain Imaging in Schizophrenia

Term	Abbreviation	Definition
α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid	AMPA	A glutamate agonist
α-methyl-para-tyrosine	AMPT	A pharmacologic agent to deplete presynaptic dopamine by inhibition of tyrosine hydroxylase
β-carbomethoxy-3-β-(4-iodophenyl)-tropane	β-CIT	An agent to bind to dopamine transporters (DATs) and serotonin transporters (SERTs)
Binding affinity	$1/K_D$	Tightness of the attachment between a neurotransmitter and its neuroreceptor
Binding potential	BP	An estimate of the density of the indicated receptors in specific brain regions estimated by the ratio of B_{max} to K_D
$[^{11}C]$-3-amino-4-[2-(dimethylaminomethylphenylthio)]benzonitrile	$[^{11}C]$DASB	A radiotracer to bind to serotonin transporters (SERTs)
Dopa-decarboxylase	DDC	The rate-limiting enzyme in the

		presynaptic metabolism of dopamine
Dopamine receptor	D_2/D_3 R	D_2/D_3 R represents the receptors labeled by benzamide PET radiotracers, e.g., [^{11}C]raclopride.
Dopamine release	DAR	The increment in intrasynaptic dopamine following physiologic stress or pharmacologic stimulation by a dopaminergic agent, e.g., amphetamine and methylphenidate
Dopamine transporter	DAT	The reuptake site for dopamine from the synapse into the presynaptic neuron to be packaged again for rerelease into the synapse
(+)-enantiomer of 1,2,3,4a,5,6-hexahydro-9-hydroxy-4-n-propyl- 4H-naphth[1,2-b][1,4]-oxazine	(+)-PHNO	A potent dopamine D_2/D_3 agonist
Equilibrium dissociation constant	K_D	The inverse of the affinity of radioligand binding

Functional magnetic resonance imaging	fMRI	A brain imaging method detecting functional changes using magnetic waves
γ-aminobutyric acid	GABA	An inhibitory neurotransmitter
[^{123}I]iodobenzamide	[^{123}I]IBZM	An iodinated benzamide
Lysergic acid diethylamide	LSD	A hallucinogen
Molecular imaging	MI	The use of molecular-based probes to visualize physiologic process
Magnetic resonance spectroscopy	MRS	The visualization of in vivo brain chemistry using magnetic waves
Maximal binding potential	B_{max}	The theoretical density of the indicated neuroreceptor in portions of the brain
N-methyl-D-aspartic acid	NMDA	An agonist of the NMDA glutamate receptor
3-*N*-methylspiperone	NMSP	The methylated derivative of spiperone, a potent agent to bind to dopamine D_2/D_3/D_4 and

		serotonin (5-HT_{2A}) receptors
Positron emission tomography	PET	A technique for the visualization of positrons in an organ after the administration of a radioactive compound
3-quinuclidinyl benzilate	QNB	A potent antagonist of muscarinic receptors
Serotonin	5-HT_{2A}	An excitatory neurotransmitter
Serotonin receptor	5-HT_{2A} R	Molecules on the postsynaptic site to activate the neuron after being stimulated by serotonin
Serotonin transporter	SERT	The reuptake site for serotonin from the synapse into the presynaptic neuron to be packaged again for rerelease into the synapse
Single photon emission computed tomography	SPECT	A technique for visualization of single photons in an organ after the administration of a radioactive compound
2-deoxy-2-[^{18}F]fluoro-D-glucose	[^{18}F]FDG	Radioactive sugar

$H_2[^{15}O]$	$H_2[^{15}O]$	Radioactive water

Molecular Imaging of Cerebral Blood Flow and Cerebral Metabolism

Prior to the first neuroimaging receptor studies in schizophrenia in the mid-1980s, decades of imaging from the late 1940s involved cerebral blood flow measurements with 15% nitrous oxide, krypton ($[^{85}Kr]Kr$), xenon ($[^{133}Xe]Xe$), or radioactive water ($[^{15}O]H_2O$), and from the late 1970s measurements of regional cerebral glucose metabolism using 2-deoxy-2-$[^{18}F]$fluoro-D-glucose ($[^{18}F]$FDG) PET. The PET studies with FDG suggested the presence of hypofrontality in patients with schizophrenia, and PET with ($[^{15}O]H_2O$), demonstrated reduced blood flow to the frontal cortex, anterior cingulate, cerebellum, and thalamus in the left hemisphere, with possible compensatory blood flow increases to the right hemisphere. Nonetheless, the findings were difficult to relate to brain neurochemistry, a correlation that is now possible with molecular imaging studies.

The imaging of receptors began with tracers derived from drugs primarily used as antipsychotics and now extends to other neurotransmitter systems as well as to pre- and intrasynaptic binding sites. The methods of molecular imaging (see Section 1.17 on radiotracer imaging with PET and SPECT) revolutionized the field of central nervous system (CNS) neuroimaging by adopting a common methodology applied to neuropsychiatric disorders in general and, more recently, to drug development. While the initial imaging research primarily focused on schizophrenia, it also led the way to the study of common neuroreceptor systems in other disorders and organs (Table 12.9–2).

Molecular Imaging of D_2-Like Receptors

Historical advances in both molecular imaging and schizophrenia research extend beyond investigations of postmortem autoradiography and tissue samples to include the in vivo imaging of radiolabeled drugs such as $[^{11}C]$chlorpromazine. The first successful in vivo molecular imaging, a translation of in vitro neurochemistry, began with the imaging of $[^{11}C]$3-*N*-methylspiperone (NMSP), a methylated derivative of spiperone, one of the highest affinity antipsychotics tested in clinical trials in the 1970s. The methylation of spiperone provided the technical tool applicable to the imaging of D_2-like dopamine receptors with the object of quantifying the density of dopamine receptors in the basal ganglia, as well as serotonin (5-HT_{2A}) receptors in the cortex. Now, more specific D_2 and D_3 ligands exist, such as $[^{11}C]$raclopride and $[^{18}F]$fallypride.

Table 12.9–2.

Abbreviated and Chemical Names of Available Nuclear Neuroimaging Radiotracers and the Indicated Neurotransmitter Systems

Target	Abbreviated Name	Chemical Name
Acetylcholinesterase (AChE) activity	[^{11}C]MP4A	*N*-[^{11}C]methylpiperidin-4-yl acetate
Cannabinoid CB_1 receptors	[^{3}H]MePPEP	[^{3}H](3*R*,5*R*)-5-(3-methoxy-phenyl)-3-((*R*)-1-phenyl-ethylamino)-1-(4-trifluoromethyl-phenyl)-pyrrolidin-2-one
Cannabinoid CB_1 receptors	[^{11}C]OMAR	1-(2,4-dichlorophenyl)-4-cyano-[^{11}C]-5-(4methoxyphenyl)-*N*-(piperidin-1-yl)-1*H*-pyrazole-3-carboxamide
Cannabinoid CB_1 receptors	[^{18}F]MK-9470	*N*-[2-(3-cyano-phenyl)-3-(4-(2-[^{18}F]-fluoroethoxy)phenyl)-1-methylpropyl]-2-(5-methyl-2-pyridyloxy)-2-methylproponamide
Dopamine synthesis and storage	6-[^{18}F]DOPA	6-[^{18}F]dihydroxyphenylalanine
Dopamine D_1 receptors	(+)-[^{11}C]NNC112	(+)-8-chloro-5-(7-benzofuranyl)-7-hydroxy-3-[^{11}C]methy1-2,3,4,5-tetrahydro-1H-3-benzazepine
Dopamine D_1 receptors	[^{11}C]SCH23390	(*R*)-(+)-8-chloro-2,3,4,5-tetrahydro-3-[^{11}C]methyl-5-phenyl-1*H*-3-benzazepin-7-ol
Dopamine D_2 receptors	[^{11}C]RAC	[^{11}C]raclopride
Dopamine D_2 receptors	[^{18}F]NMB	[^{18}F](*N*-methyl)benperidol
Dopamine D_2 receptors	[^{123}I]IBZM	(*S*)-*N*-[(1-ethyl 2-pyrrolidinyl)] methyl-2-hydroxy-3-[^{123}I]iodo-6-methoxybenzamide
Extrastriatal dopamine D_2 receptors	[^{11}C]FLB 457	(*S*)-*N*-((1-ethyl-2-pyrrolidinyl)methyl)-5-bromo-[^{11}C]2,3-dimethoxybenzamide
Dopamine D_2/D_3 receptors	[^{18}F]FAL	[^{18}F]fallypride
D_3 receptors	[^{11}C]-(+)-PHNO	[^{11}C]-(+)-4-propyl-3,4,4a,5,6,10b-hexahydro-2H-naphtho[1,2-b][1,4]oxazin-9-ol
Dopamine transporter	[^{123}I]β-CIT	β-CIT2-betacarbomethoxy-3 beta-(4-

		[^{123}I]iodophenyl)tropane
γ-aminobutyric acid (GABA)	[^{18}F]FMZ	[^{18}F]fluoroflumazenil
Glycine transporter type 1 (GlyT$_1$)	[^{11}C]GSK931145	N-((1-([^{11}CH$_3$]dimethylamino)cyclopentyl)(phenyl)methyl)-2,6-dimethylbenzamide
Glycine transporter type 1 (GlyT$_1$)	[^{11}C]R05013853	[5-[^{11}C]methanesulfonyl-2-((*S*)-2,2,2-trifluoro-1-methyl-ethoxy)-phenyl]-[5-(tetrahydro-pyran-4-yl)-1,3-dihydro-isoindol-2-yl]-methanone
Muscarinic acetylcholine receptor (mAChR)	[^{11}C]NMPB	[^{11}C]*N*-methyl-4-piperidyl benzylate
Muscarinic acetylcholine receptor subtype 1 (mAChR$_1$)	[^{11}C]GSK1034702	[^{11}C]4-fluoro-6-methyl-1-(1-(tetrahydro-2H-pyran-4-yl)piperidin-4-yl)-1,3-dihydro-2H-benzo[d]imidazol-2-one
Nicotinic acetylcholine receptor (nAChR) subtype α4β2	[^{18}F]AZAN	(−)-2-(6-[^{18}F]fluoro-2,3'-bipyridin-5'-yl)-7-methyl-7-aza-bicyclo[2.2.1]heptane
Nicotinic acetylcholine receptor (nAChR) subtype α4β2	2-[^{18}F]FA	(*S*)-3-(azetidin-2-ylmethoxy)-2-[^{18}F]fluoropyridine
Nicotinic acetylcholine receptor (nAChR) subtype α4β2	6-[^{18}F]FA	6-[^{18}F]fluoro-3-(2(*S*)-azetidinylmethoxy)pyridine
Nicotinic acetylcholine receptor (nAChR) subtype α4β2	5-[^{123}I]IA	(*S*)-5-[^{123}I]iodo-3-(2-azetidinylmethoxy)pyridine
Nicotinic acetylcholine receptor (nAChR) subtype α7	[^{11}C]CHIBA-1001	4-[^{11}C]methylphenyl 2,5-diazabicyclo[3.2.2]nonane-2-carboxylate
Nicotinic acetylcholine receptor (nAChR) subtype α7	[^{18}F]ASEM	[^{18}F]3-(1,4-diazabicyclo[3.2.2]nonan-4-yl)-6-fluorodibenzo[b,d]thiophene 5,5-dioxide
Opioid receptors	[^{18}F]MK-0911	1-(2-[^{18}F]fluoroethyl)-3-(3-(hydroxymethyl)-1-(spiro[2.5]octan-4-

		ylmethyl)piperidin-4-yl)-1,3-dihydro-2H-benzo[d]imidazol-2-one
Phosphodiesterase 10A (PDE$_{10A}$) receptors	[^{18}F]JNJ42259152	2-[4-[1-(2-[^{18}F]fluoroethyl)-4-(4-pyridinyl)-1*H*-pyrazol-3-yl]phenoxy]methyl]-3,5-dimethyl-pyridine
Phosphodiesterase 10A (PDE$_{10A}$) receptors	[^{18}F]MNI-654	2-(2-(3-(1-(2-[^{18}F]fluoroethyl)-1*H*-indazol-6-yl)-7-methyl-4-oxo-3,4-dihydroquinazolin-2-yl)ethyl)-4-isopropoxyisoindoline-1,3-dione
Phosphodiesterase 10A (PDE$_{10A}$) receptors	[^{18}F]MNI-659	2-(2-(3-(4-(2-[^{18}F]fluoroethoxy)phenyl)-7-methyl-4-oxo-3,4-dihydroquinazolin-2-yl)ethyl)-4-isopropoxyisoindoline-1,3-dione
Serotonin 5-HT2a receptors	[^{11}C]MDL100907	(*R*)-(+)-[3-O-methyl-[^{11}C]-4-(1-hydroxy-1-(2,3-dimethoxyphenyl)methyl)-*N*-2-(4-fluoro-phenylethyl)piperidine
Serotonin 5-HT2a receptor	[^{18}F]altanserin	{3-(2-[4-(4-[^{18}F]fluorobenzoyl)piperidin-1-yl]-ethyl)-1,2-dihydro-2-thioxo-quinazolinone, *1*}
Serotonin (5-HT) transporter	[^{11}C]DASB	[^{11}C]-3-amino-4-(2-dimethylaminomethylphenylsulfanyl)-benzonitrile
Translocator protein (TSPO)	[^{11}C]PK11195	*N*-butan-2-yl-1-(2-[^{11}C]chlorophenyl)-*N*-methylisoquinoline-3-carboxamide
Translocator protein (TSPO)	[^{18}F]FEPPA	N-(2-(2-[^{18}F]fluoroethoxy)benzyl)-N-(4-phenoxypyridin-3-yl)acetamide
Vesicular acetylcholine transporter	[^{123}I]IBVM	[^{123}I]iodobenzovesamicol
Vesicular monoamine transporter type 2 (VMAT-2)	[^{11}C]DTBZ	[^{11}C]dihydrotetrabenazine

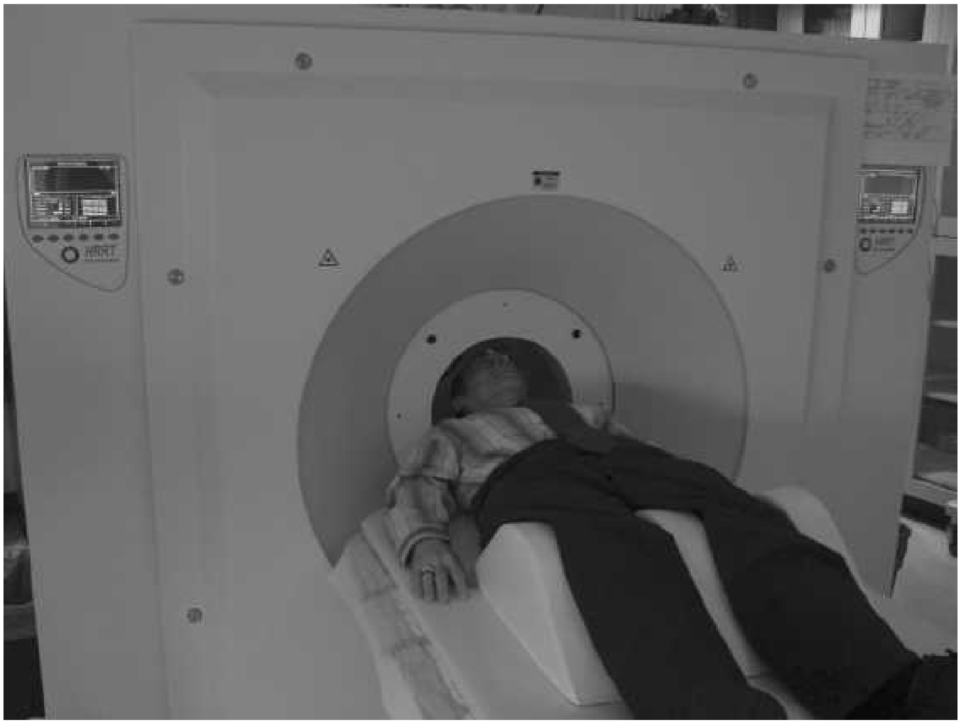

FIGURE 12.9–1. High-resolution positron emission tomography (PET) scanner.

TYPES OF MOLECULAR IMAGING

Molecular imaging techniques include PET, SPECT, MRS, and fMRI. They are functional brain imaging techniques that visualize the metabolism and other physiologic processes in the living brain.

Figure 12.9–1 illustrates a state-of-the-art PET device used in these studies. Functional brain imaging seeks to identify and characterize the dynamic chemical and biologic changes of brain work under varying conditions. Functional brain imaging contrasts with structural brain imaging that involves procedures to precisely demarcate anatomical regions of the brain. Thus, molecular brain imaging is a subclass of functional brain imaging.

Indirect Molecular Imaging Estimates

Ideally, concentrations of neurotransmitters would be measured in the living human brain. Unfortunately, procedures to directly measure these molecules are not yet available. Therefore, molecular brain imaging requires indirect techniques to estimate the density and distributions of neurotransmitters in the living brain. For example, there are no noninvasive procedures that directly measure glucose concentration in the living human brain. Instead, PET with [^{18}F]FDG is a technique that yields the density and distribution of [^{18}F]FDG molecules that trace the fate of glucose and its metabolism in the brain. Similarly, the concentrations of dopamine and serotonin cannot be directly measured in the brain. Instead, radiotracers exist that trace the density and distribution of endogenous substrates of the presynaptic dopamine (DAT) and serotonin (SERT) transporters, the postsynaptic dopamine (D_1–D_5) and serotonin (5-HT) receptors, the latter of which there are at least 14 different subtypes and opiate receptors of the mu subtype. For example, molecular imaging demonstrates the alterations in the density of dopamine D_2 receptors with the menstrual cycle, during which the binding of the dopamine D_2 receptor is reduced in the follicular phase and increased in the periovulatory and luteal phases of the estrous cycle in some early PET studies. Molecular imaging also was used to confirm the dopamine hypothesis that people with schizophrenia have an excess of intrasynaptic dopamine in the brain. The dopamine hypothesis represents a cornerstone of schizophrenia pathophysiology, and drugs that block the binding of dopamine to dopamine receptors remain a fundamental tool of the pharmacology of schizophrenia.

Pharmacologic Challenges

Pharmacologic challenges are experimental tools used to test the effects of the administration of drugs hypothesized to alter the levels of neurotransmitters in regions of the brain. Examples of pharmacologic challenges of the dopaminergic system include the administration of pharmacologic agents to increase, for example, amphetamine (Dexedrine), or decrease, for example, α-methyl-para-tyrosine (AMPT), metyrosine (Demser), intrasynaptic dopamine given before PET scans and to estimate the occupancy of dopamine receptors by endogenous dopamine in the brain. These techniques demonstrate that intrasynaptic dopamine is increased following the administration of amphetamine. After AMPT administration, patients with schizophrenia still have more dopamine available than normal, although both are decreased compared to nontreated AMPT subjects. This is also consistent with elevations of baseline dopamine D_2 receptors in schizophrenia.

Grace (1991) hypothesized that, in contrast to healthy normal controls, people with schizophrenia demonstrate abnormally elevated intrasynaptic levels of dopamine in both (1) the resting, basal, unstimulated, tonic state and (2) the phasic state of maximal arousal like the one induced pharmacologically by amphetamine or another agent as methylphenidate that produces maximal intrasynaptic levels of dopamine. The value of pharmacologic challenges in the molecular imaging of schizophrenia has been demonstrated by studies that have confirmed the hypothesis of Grace that extrasynaptic dopamine is abnormally reduced in the tonic state and intrasynaptic dopamine is abnormally elevated in the phasic states in people with schizophrenia. Pharmacologic challenges also represent tools to evaluate the mechanisms of action of novel agents. For example, agents hypothesized to alter the intrasynaptic concentration of dopamine can be administered before scans compared with the placebo states to determine the increase or decrease of intrasynaptic dopamine (Table 12.9–3).

MOLECULAR IMAGING IN PATHOPHYSIOLOGY

Molecular Imaging of Dopamine in Schizophrenia

The development of molecular imaging studies resulted from the successful radiolabeling of dopamine receptor antagonists, “typical” antipsychotic drugs (now called “first-generation”), in the mid-1980s, consistent with the investigation of postsynaptic dopamine D_2-like receptors.

Estimation of the Binding Dopamine D_2-Like Receptors. The study of the binding of D_2-like receptors is typically made in a single PET or SPECT imaging session at high specific activity of the tracer (i.e., with low mass of the nonpharmacologic moiety injected) designed to yield the maximum value of the binding potential (BP), equal to the ratio of available receptor density to binding affinity. The BP has been determined with two main classes of radiotracers of dopamine receptors. The first class of radiotracers that visualize D_2 or D_3 receptors in the brain includes butyrophenones such as [^{11}C]NMSP and [^{18}F]spiperone derivatives. The second class of radiotracers that visualize D_2-like receptors in the brain includes radiolabeled benzamides such as [^{11}C]raclopride (Fig. 12.9–2), [^{18}F]fallypride ([^{18}F]FP) (see Fig. 12.9–2), or [^{123}I]iodobenzamide ([^{123}I]IBZM).

Dopaminergic Neuroreceptor Mapping In Schizophrenia. Dopamine belongs to the group of monoamines that arise from the decarboxylation of large neutral and aromatic amino acids in brain tissue. Depending on the specific precursor, they fall into two categories, catecholamines that derive from phenylalanine but in the brain only from tyrosine, and the indoleamines that derive from tryptophan. Neurotransmission is the term applied to the exchange among brain cells of specific information encoded in chemicals known as transmitters. Neurotransmission includes the steps of synthesis, vesicular storage, release, reuptake, and removal or breakdown of neurotransmitters. Extra- or intracellular proteins function as enzymes, receptors, or transporters that implement the individual steps.

Table 12.9–3.
Imaging Studies of Dopamine Neurotransmission in Schizophrenia: Observations and Possible Mechanisms

Site	Observation	Salient and Related Findings	Possible Mechanism
D$_2$ (striatum)	Elevations with [^{11}C]NMSP (butyrophenone)	Varied postmortem changes, internalization, monomer elevation	Decreased tonic DA
D$_2$ (striatum)	No change (benzamides) except following AMPT using [^{123}I]IBZM or [^{11}C]raclopride	Due to competition with intrasynaptic DA or dimer configuration, unmasking of D$_2$ due to intrasynaptic DA removal with AMPT	Increased phasic DA Decreased tonic DA
DDC (striatum)	Elevation with [^{18}F]fluorodopa	Majority of imaging studies	Decreased tonic DA
DA (striatum) release after AMP challenge	Elevation with [^{123}I]IBZM or [^{11}C]raclopride	Majority of imaging studies	Increased phasic DA
DA (striatum) release after ketamine challenge	Elevation with [^{11}C]raclopride Elevation with [^{11}C]FLB457	Glutamate antagonist stimulates hypoglutamatergic state	Increased phasic DA
Prefrontal D$_1$	Decrease with [^{11}C]SCH23390		Unknown
ACCX D$_2$	Decrease with [^{11}C]FLB457	Postmortem tyrosine hydroxylase immunoreactive changes	Unknown

[^{11}C]NMSP, (3-*N*-[^{11}C]methyl)spiperone; DDC, dopa decarboxylase; DA, dopamine; AMPT, α-methylparatyrosine; [$^{123\text{I}}$]IBZM, (*S*)-*N*-[(1-ethyl 2-pyrrolidinyl)]methyl-2-hydroxy-3-[$^{123\text{I}}$]iodo-6-methoxybenzamide; AMP, amphetamine; [^{11}C]FLB457, (*S*)-*N*-((1-ethyl-2-pyrrolidinyl)methyl)-5-bromo-2[^{11}C], 3-dimethoxybenzamide; [^{11}C]SCH23390, (*R*)-(+)-8-chloro-2,3,4,5-tetrahydro-3-[^{11}C]methyl-5-phenyl-1*H*-3-benzazepin-7-ol; ACCX, anterior cingulate cortex.

Reproduced with permission from Wong DF. In vivo imaging of D2 dopamine receptors in schizophrenia: the ups and downs of neuroimaging research [Erratum]. *Arch Gen Psychiatry.*

2002;59(5):440. doi:10.1001/archpsyc.59.5.440. Copyright © 2022 American Medical Association. All rights reserved.

Transmitter molecules are ligands because they engage in binding (“ligation”) as the first step of a chain of interactions that differs according to the proteins that do the binding. *Receptors* return the ligand to the original pool of unbound molecules after a delay that depends on the affinity of the receptor toward the ligand. *Transporters* relocate the ligand to a different pool of unbound ligands. *Enzymes* release the ligand to the same pool but in a new chemical form. Exogenous or artificial ligands may differ so much from the native ligand that transporters or enzymes fail to carry out their primary function and instead act as receptors by simply returning the ligand to its original pool. Depending on the degree of interference, exogenous or artificial ligands may be antagonists or partial agonists of the action of the native transmitter. We list the main extra- and intracellular sites of binding, reuptake, and metabolism of dopamine in Table 12.9–1 and with the radiotracers in Table 12.9–2.

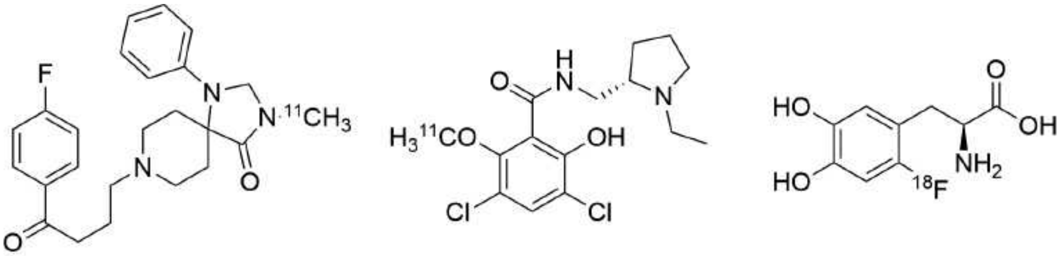

N-[^{11}C]methylspiperone
[^{11}C]NMSP

[^{11}C]Raclopride

[^{18}F]Fluorodopamine
[^{18}F]FDOPA

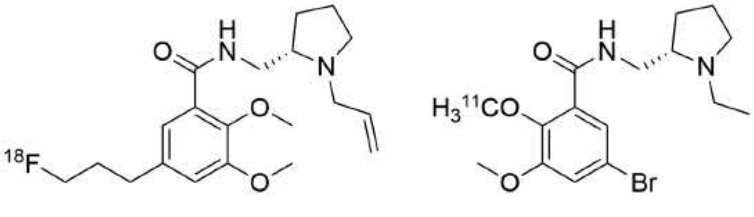

[^{18}F]Fallypride

[^{11}C]FLB 457

FIGURE 12.9–2. Structures of typical PET dopamine radiotracers.

The expression of receptors, transporters, and enzymes is highly variable and depends on the intra- and extracellular environments, making the results of PET and SPECT studies contingent upon specific circumstances at the time of the study, as is also true of transporters. When baboons and humans received the first tropane-based PET probes of the dopamine transporters, researchers generally held that the number of transporters remained constant. We now know that the number of dopamine transporters is subject to regulation by intrinsic mechanisms as well as pharmacologic manipulation.

Methods of Dopaminergic Neuroreceptor Mapping. PET studies yield multiple brain images of dopaminergic neurotransmission acquired as dynamic series of tomographic frames, typically 30 to 50 frames obtained as two- or three-dimensional images in sessions of 60 to 120 minutes. Because of the short half-lives, this duration is the limit of most studies with ^{11}C or ^{18}F. In contrast, SPECT studies with ^{123}I may continue for 6 or more hours. In two-dimensional imaging, the tomographic record consists of 15 to 64 or more sections of the brain, depending on the tomograph. Arterial sampling and brain imaging together establish the concentrations of the tracer in the blood and brain as functions of time. The records then form the basis for subsequent kinetic analyses that yield the specifics of dopaminergic neurotransmission (Fig. 12.9–3). The choice of the right tracers is the key to the successful imaging of animals or humans. With PET, the half-life of tracers limits the duration of imaging to the first hour of recording in the case of most of the ^{11}C-labeled (20-minute half-life) tracers. The pharmacology and binding properties of the tracers, in turn, dictate the usefulness of this information in relation to the relevant proteins and not solely to the properties of the ligand.

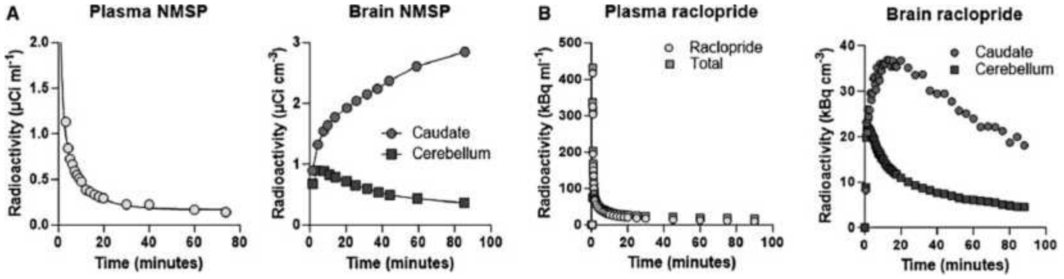

FIGURE 12.9–3. Typical time activity curves for kinetic modeling of dopamine receptors. Part **A** is arterial input plasma curve (**left**) and caudate and cerebellum time activity curve of PET imaging of [^{11}]CNMSP over 90 minutes (**right**) and is an illustration of irreversible binding to dopamine D$_2$/D$_3$ receptors; Part **B** is arterial plasma input curve (**left**) where total is the [^{11}C]raclopride including metabolites and the raclopride is the lower curve where only the parent [^{11}C]raclopride in plasma is shown and caudate and cerebellum time activity curve of PET imaging of [^{11}C]raclopride over 90 minutes (**right**) and is an illustration of reversible binding to dopamine D$_2$/D$_3$ receptors. (Data from Wong DF, Wagner Jr. HN, Tune LE, Dannals RF, et al. Positron emission tomography reveals elevated D$_2$ dopamine receptors in drug-naive schizophrenics. *Science*. 1986;234(4783):1558–1563. https://doi.org/10.1126/science.2878495; Phan JA, Wong DF, Chang NHS, Kumakura Y, Bauer WR, Gjedde A. Transient equilibrium determination of dopamine D$_2$/D$_3$ receptor densities and affinities in brain. *Front Nucl Med*. 2022;2:1030387. doi: 10.3389/fnume.2022.1030387.)

An important challenge to PET imaging is the necessary anesthesia of animals used in research. It has been possible in very few places to train nonhuman primates to remain awake while seated comfortably on a horizontally suspended tomograph. In other places, it is necessary to use a within-subject design of drug studies to reduce the variability of the PET measures. In some preclinical or clinical studies of animals, investigators did not routinely sort out all factors, such as the size of the brain, genotype, age, latent disease states, personality, and adaptation to stress and drugs. Another important factor is the sample size. Typical numbers of humans in PET studies range from 10 to 20, and populations of 30 subjects in single studies are rare. Adequate numbers are necessary to counteract the confounding influence of the tomography's finite power of resolution of recorded radioactivity, known as the "partial volume" effect. The effect is most prominent in small structures that typically are below a diameter of 2 to 3 mm or even 1 cm, depending on the resolution of the PET scanner. Current state-of-the-art PET scanners are now close to 2.5 to 4 mm in the ability to resolve structures without loss of quantification or to resolve two radioactive sites rather than smeared as one.

Tracers of Dopaminergic Neurotransmission. Table 12.9–2 lists a large number of dopaminergic tracers tested with variable success since 1983. Unfortunately, of the hundreds or thousands of compounds developed by the pharmaceutical industry, universities, or governmental institutions, only a tiny fraction reached the stage of successful labeling. Radioligands are useful only when they meet specific requirements of selectivity, affinity, and kinetics of binding, transport, or metabolism. Only when they meet these requirements can reliable estimates of receptor density or neurotransmitter synthesis, release, or metabolism, be obtained noninvasively.

Three tracers illustrate the historical advance of dopaminergic tracer development and the problems associated with the use of the tracers. In 1983, the experimental neuroleptic spiroperidol (spiperone, 8-[4-(4-fluorophenyl)-4-oxo-butyl]-1-phenyl-1,3,8-triazaspiro [4.5]decan-4-one) served as the precursor by methylation of the radioligand [^{11}C]3-*N*-methylspiperone (NMSP), for which the dopamine D_2-like (D_2, D_3, and D_4) receptors have high affinity. The radioligand yielded the very first D_2-like receptor images in the human brain with PET. However, the selectivity was not exclusive for this receptor subtype, as the serotonin 5-HT_{2A} receptors had a significant affinity. In contrast, useful ligands frequently have binding characteristics that include high selectivity but only moderate affinity, commensurate with BPs between unity and as much as approximately 6.

Currently, the most common dopamine receptor ligand used with PET is [^{11}C]raclopride (3,5-Dichloro-*N*-[(2*S*)-1-ethyl-2-pyrrolidinyl]methyl]-2-hydroxy-[^{11}C]6-methoxybenzamide), derived from another class of chemical, the benzamides. Benzamides are derivatives of benzoic acid, and substituted benzamides often serve as neuroleptics or antipsychotics in psychiatric practice. Two active substances from the group of benzamides are in clinical use, sulpiride and amisulpride (see reference to LB-102 methyl-amisulpride below). Raclopride binds reversibly and selectively to the dopamine D_2 and D_3 receptors, with BPs close to 3 (Gjedde 2003), making it an ideal tracer of dopaminergic binding sites in high-density locations such as the striatum. In the cerebral cortex, the lower density renders the BP too low to be of routine usefulness. Several alternatives with higher affinity are available for this purpose. Tracer [^{18}F]fallypride ([5-(3-[^{18}F]fluoropropyl)-2,3-dimethyl-*N*-[(2*S*)-1-(2-propenyl)-2-pyrrolidinyl]-methyl]-benzamide) is a radioligand with BPs of 0.5 to 1.5 in the cerebral cortex. The higher affinity elevates the retention in the striatum, where the BP is so high that binding is delivery-dependent.

In the case of metabolic substrates, the product must be retained in the brain as an index of the activity of the enzyme, provided the transport across the blood-brain barrier is not rate-limiting for the uptake. The first study with PET of the metabolism of the dopamine precursor tracer [^{18}F]fluorodopa ([^{18}F]FDOPA) was published in 1983 with an uptake that appeared to be irreversible, although the retention may reflect only a fraction of the tracer. This meant that subsequent researchers have attempted a host of different kinetic analyses to solve the issue of the specificity of the signal. It is now clear that the retention primarily reflects the transport of the labeled product fluorodopamine (FDA) into vesicles, and the retention depends on the irreversibility of this transport. Two processes interfere with the retention: FDA is subject to oxidation by monoamine oxidases, and release of FDA from the vesicles exposes the released FDA to renewed metabolism after reuptake into the terminals. Therefore, low retention can reflect both reduced vesicular incorporation due to loss of terminals or loss of dopa decarboxylase (DDC) activity, as well as inhibition of vesicular transporters, or increased release of FDA in response to increased dopaminergic neurotransmission.

As markers of dopaminergic activity in presynaptic terminals, alternatives to FDOPA are tracers of the dopamine reuptake sites, based on the tropane structure known as nortropane (8-azabicyclo[3.2.1]octane). Tropane is a nitrogenous bicyclic organic compound, mainly famous for the large group of alkaloids that derive from it, including, among others, atropine and cocaine. Several potentially irreversible ligands for the cocaine receptor site on the dopamine transporter derive from cocaine. The cocaine analog CFT (2-carboxymethoxy-3-(4-fluorophenyl) tropane or WIN-35,428) first arose as a radiolabeled probe for PET and SPECT by Madras and coworkers and later developed into the PET tracer [^{11}C]WIN-35,428 by Kuhar and coworkers. The tracer binds irreversibly to the cocaine site on the dopamine transporter (DAT) but also to the serotonin (5HTT or SERT) and norepinephrine (NET) transporters. The vesicular monoamine transporter (VMAT2) is located in the membranes of terminal vesicles and can be imaged with the radioligand [^{11}C]D-dihydrotetrabenazine ([^{11}C]DTBZ). However, the imaging of the VMAT2 transporter may not settle the "transporter regulation issue," as VMAT2 also appears to be regulated (see Table 12.9–2).

Estimation of Absolute Dopamine D_2-Like Receptor Density. In contrast to the typical approach to measures the BP of D_2-like receptors, the use of at least two PET scans provides the basis of calculation of the unique ability to measure absolute receptor density (B_{max}). The measurement of absolute B_{max} can be carried out by performing at least two brain imaging sessions with the administration of [^{11}C]raclopride or [^{11}C]NMSP, with at least one after a known potent dose of a cold, nonradioactive D_2-like receptor antagonist such as unlabeled raclopride itself or haloperidol, respectively, to reduce some of the binding. Following some additional mathematical modeling, these results then provide an estimate of the B_{max} and binding affinity (K_D). These results correspond to the results from autoradiography studies or in vitro homogenate studies.

Subsequent simplified studies involved single high-specific activity injections of radiotracer mass of less than 20 μg, carried out in order to estimate the BP, that is, the ratio of the maximum number of receptors B_{max} to the half-saturation concentration K_D. Nonetheless, in order to separate these variables of B_{max} and K_D, it is necessary to carry out receptor saturation studies with varying unlabeled mass injections. In general, the absolute B_{max} studies have tended to show elevations ranging from small to large as much as a threefold increase with [^{11}C]NMSP, a butyrophenone. Although B_{max} elevations have been consistently shown with [^{11}C]NMSP, no changes have been demonstrated with B_{max} using benzamides such as [^{11}C]raclopride or [^{123}I]IBZM. Interestingly, one of the characteristics of these benzamides is that they are also quite susceptible to endogenous dopamine and recently have gained popularity for measuring the effect of pharmacologic challenges or psychological stress that alter endogenous dopamine levels. Some have suggested that the differences in B_{max} between raclopride and NMSP could be attributed to the preferential affinity for a monomer versus a dimer form of the dopamine receptor in the case of the [^{11}C]NMSP and possibly receptor internalization and susceptibility to endogenous neurotransmitter for [^{11}C]raclopride. Regardless of the reasons for the differences, meta-analysis has shown that there is a small 12% significant elevation of striatal D_2 receptors with an effect size of 0.51, and there is an association between increased D_2 receptor density and positive symptoms.

Dopamine Precursor Estimates. Dopamine precursor measures using [^{18}F]FDOPA or [^{11}C]DOPA were among the first markers of dopamine neurotransmission to be studied. When properly quantified, these measurements reflect DDC rates of activity or represent some measure of dopamine presynaptic function. A number of these studies have demonstrated that the dopa uptake was higher in people with schizophrenia and other psychoses in contrast to healthy matched controls. Elevated [^{18}F]fluorodopamine ([^{18}F]FDOPA) turnover has been shown to coexist with decreased steady-state storage of FDOPA and its decarboxylated metabolites in the caudate nucleus and amygdala of untreated people with schizophrenia. Interestingly, the decreased steady-state storage of FDOPA in the amygdala correlated inversely with positive symptoms of schizophrenia, suggesting a role for this brain structure in the pathophysiology of the disease. Hence, the striatal findings are among the most substantial evidence of a "biomarker" in schizophrenia and are consistent with the concept of decreased tonic dopamine and elevated phasic dopamine proposed by Grace. More recently, in studies by Jauhar with [^{18}F]FDOPA, Ki^{cer} has been shown to be elevated in responders compared to nonresponders and also positively correlated with PANSS positive, negative, and total scores and negatively with Global Assessment of Function (GAF).

Regarding possible effects of antipsychotic medication, Jauhar studied 17 patients with schizophrenia on medication at therapeutic doses. There was no significant effect of treatment on dopamine synthesis capacity in the whole striatum ($P = .47$), thalamus, or midbrain, nor was there any significant relationship between change in dopamine synthesis capacity and change in positive ($r = 0.35$, $P = .13$), negative, or total psychotic symptoms. They concluded that dopamine synthesis capacity is unaltered by antipsychotic treatment and therapeutic effects are not mediated by these dopamine antagonist medications. However, this is contrary to what others found, such as Gründer with a decrease in dopamine synthesis capacity with treatment.

Other Dopamine Subtypes. There is considerable interest in the D_3 receptor as a potential therapeutic target for schizophrenia. Unfortunately, no studies are explicitly able to image the D_3 receptor in vivo without also measuring the much more densely distributed D_2 receptors, for example (^{11}C PHNO). A high-affinity selective D_3 radioligand is needed. Most measures of D_3 are quantified in brain regions of a relatively high density of D_3 as the ventral striatum, substantia nigra, and globus pallidus.

Dopamine D_1 receptors may also play a role in the development and maintenance of the symptoms and signs of schizophrenia. Several studies have shown no significant difference in D_1 receptors in untreated patients with schizophrenia. However, a twin study showed reduced striatal D_1 receptors in the affected twins who were on medications compared to the unaffected cotwins. Two studies have reported associations between D_1 receptor density in the dorsolateral prefrontal cortex and impairments in working memory performance and negative symptoms. The limitations of these studies of the D_1 receptors are that the radiotracers are not selective for D_1 receptors but also bind to 5-HT_{2A} receptors. Thus, further radiotracer development is needed for the D_1 receptor site. Several years ago, D_4 was considered a serious candidate for antidepressant action, but initial PET ligands and small therapeutic trials for schizophrenia failed. It is still of interest, perhaps in other areas of neuropsychiatry, and is a binding site for [^{11}C]NMSP.

Measuring Intrasynaptic Dopamine. In the past decade, an important property probably unique to radiolabeled benzamides has been found. Specifically, the D_2-like receptor binding of [^{123}I]IBZM and [^{11}C]raclopride (see Fig. 12.9–2) and probably [^{18}F]fallypride (see Fig. 12.9–2) can decline in the presence of a pharmacologic challenge with a pharmacologic agent that induces increased or decreased levels of intrasynaptic dopamine. In the case of increased intrasynaptic dopamine, challenges with dopaminergic agents, including methylphenidate (Ritalin) as well as amphetamines, have been employed. Investigators have shown that these benzamides can have a reduced BP in the presence of an intravenous amphetamine challenge (0.3 mg/kg body weight). Researchers showed significantly greater decreases in BPs in patients with schizophrenia in contrast to healthy normal control volunteers. There is still some controversy about the meaning of these decreases in BP. What is probably most defendable is that there is an increase in occupancy of the D_2-like sites such that the radiolabeled benzamides cannot bind to them. This reduction of occupancy could be due to the increased intrasynaptic dopamine released into the synapse through the use of psychostimulants. However, a change in the affinity state, which might increase with dopamine, could be occurring. These results, however, are not without controversy. Instead of the radiotracer binding returning to baseline as one might expect after a bolus amphetamine challenge, it stays depressed for many hours, sometimes exceeding 6 to 8 hours, which could be consistent with other factors than simply intrasynaptic DAR. Furthermore, at least one study has not demonstrated significant elevations in DAR, which in itself is an interesting finding suggestive of a potential heterogeneity in dopamine response to challenges with stimulants. The heterogeneity may be related to heterogeneity in clinical symptoms, as at least one study has shown that DAR is greater in patients who have more positive symptoms at the time of scanning.

It is interesting to note that one other neuropsychiatric disorder, Tourette syndrome, has shown highly significant elevations of DAR after amphetamine challenge and bolus intravenous [^{11}C]raclopride PET studies. These DAR measures seem to be greater with Tourette syndrome patients with comorbid obsessive-compulsive disorder. Hence, the elevated DAR appears not to be specific to schizophrenia.

Molecular Imaging after Dopamine Depletion. Another procedure used for changing intrasynaptic dopamine is dopamine depletion. This has been typically tried either with reserpine (Serpalan), as is typical with animals, or with AMPT, a drug that is approved in humans for pheochromocytoma. Using such AMPT challenges has allowed the demonstration of elevation in BPs, presumably reflecting the decrease of competition with intrasynaptic dopamine. Such measures have allowed an estimate of baseline or basal dopamine in the synapse. One such study demonstrated that the elevation or change in BP post-AMPT dopamine depletion is greater in people with schizophrenia than in healthy controls. Patients with greater elevations in D_2 receptor binding demonstrated greater improvement in positive symptoms with antipsychotic treatment. This is consistent not only with an increase in basal intrasynaptic dopamine but also with elevated basal densities of elevated D_2 receptors. Thus these findings support the results of a meta-analysis that shows a small but definite elevation of D_2 receptors in schizophrenia.

In schizophrenia, PET reveals increased capacity of dopamine synthesis from DOPA, first shown by increased tonic and phasic DAR, elevated density of dopamine D_2-like receptors, and increased extracellular dopamine. These findings explain how antagonists of dopaminergic neurotransmission can mitigate the worst effects of the disease, at least for some subtypes with positive symptoms. Still, complete reversal did not happen, probably because the primary disorder is cortical. It is likely that the increased synthesis capacity is due to an increase in enzyme activity, with the actual increase of dopamine production depending on the concentration of the precursor DOPA.

Future Molecular Imaging of Agonists of Dopamine and Glutamate. In the future dopamine agonists, such as the ^{11}C-labeled (+)-enantiomer of 1,3,4,6,7,11b-Hexahydro-9,10-dimethoxy-3-(2-methylpropyl)-2*H*-benzo[*a*]quinolizin-2-ol ([^{11}C](+)PHNO), or glutamate agonists, such as [^{11}C]α-amino-3-hydroxyl-5-methyl-4-isoxazole-propionate ([^{11}C]AMPA), may help determine whether there is truly intrasynaptic DAR. Dopamine agonists are more reflective of the competition with the intrasynaptic dopamine itself. On the other hand, a recent study with [^{11}C](+)PHNO did not demonstrate a significant change in DAR, probably related to experimental differences.

Extrastriatal Dopamine D_2-Like Receptors. Striatal D_2 radioligands are not ideal for extrastriatal imaging even though they have been examined longitudinally and have shown some limited imaging, such as the thalamus and some cortical regions. With the advent of higher affinity radiotracers of the D_2-like receptors, such as [^{18}F]fallypride and [^{11}C]FLB 457, it has been possible to image extrastriatal dopamine D_2-like receptors in the cortex in schizophrenia, with both

PET and SPECT. Several PET-SPECT studies have been carried out to date, all showing decreases or no change. Some studies have inconsistently demonstrated decreased binding in the thalamus, temporal cortex, amygdala, and cingulate cortex. Recent data showed DA deficits extending to most cortical regions and even to other extrastriatal subcortical regions not previously considered to be “hypodopaminergic” in schizophrenia. These findings yield a description of extrastriatal and striatal dopaminergic dysregulation in schizophrenia.

As the vast majority of studies focused on the dopamine system in patients with schizophrenia, and this is the use for which most PET and SPECT radioligands were developed and validated, it is helpful to present a summary in the shape of Table 12.9–2. As described above, there are increasing lines of evidence that the glutamatergic system is dysfunctional in schizophrenia. This, in turn, may lead to especially positive symptoms in schizophrenia, where antipsychotic medications appear to have benefits. As described by studies of Grace and others, one unifying theory may be described by two states of dopamine: A tonic state usually associated with extrasynaptic dopamine (homeostatic or basal level) and a phasic state, which is more episodic and perhaps more associated with symptomatology such as hallucinations usually associated with intrasynaptic dopamine. As shown in Table 12.9–3, the pre-, post-, and intrasynaptic measurements of dopamine would be quite compatible with this phasic–tonic dopamine theory. In this theory, patients with schizophrenia have lower tonic dopamine and hence less inhibition of their phasic dopamine (thereby alluding to the hyperdopaminergic state). Indeed, as shown in Table 12.9–3, lower tonic dopamine would be consistent with upregulation of the presynaptic enzymes of DDC as measured by [^{18}F]FDOPA (multiple studies showing elevations in schizophrenia), elevated amphetamine-induced DAR (two centers, three studies), and elevated D_2-like receptor postsynaptic in response to the lower tonic dopamine (multiple studies, primarily one site). In the future, it may be possible to integrate other abnormalities, such as serotonin. Once glutamate studies have been fully verified with multiple radioligands, this should also contribute to this model.

Molecular Imaging of Serotonin in Schizophrenia

The second major neurotransmitter system studied in schizophrenia is serotonin. This has most recently been driven by theories of the mechanisms of action of second-generation antipsychotics. However, the role of serotonin in schizophrenia can also be traced back to the psychotic symptomatology from the hallucinogenic effects of lysergic acid diethylamide (LSD), which is chemically similar to serotonin.

Estimation of the Density of Serotonin 2A Receptors and Serotonin 1A Receptors. Initial brain imaging studies of serotonin examined the 5-HT$_{2A}$

receptors. Indeed 5-HT$_{2A}$ receptors were originally imaged in the cortex with nonselective radiotracers such as [^{11}C]NMSP. A number of postmortem studies have suggested decreased 5-HT$_{2A}$ binding in the prefrontal cortex, but PET studies in drug-free or drug-naive patients with schizophrenia have found normal 5-HT$_{2A}$ receptor binding. One study demonstrated significant decreases in the prefrontal cortex in drug-naive schizophrenic patients, and another study demonstrated a decrease in cortical binding in drug-free patients. Estimation of the density and distribution of serotonin receptors in the living human brain is an area that merits further investigation as more 5-HT$_{2A}$ radioligands are now available such as [^{18}F]altanserin.

More recently, the ability to measure 5-HT$_{1A}$ receptors has led to investigations showing increases in binding in multiple brain regions, including the medial temporal cortex, using PET imaging with [^{11}C]WAY100635, a potent 5-HT$_{1A}$ antagonist. Subsequent studies have shown decreased binding using the same radiotracer in the amygdala, while a third showed no differences. Only some of the imaging studies have confirmed the results of postmortem studies, which suggest increases in 5-HT$_{1A}$ in cortical areas, but more novel radiotracers show promise for use in schizophrenia.

Estimation of the Density of the Serotonin Transporters (SERT). The second major serotonin imaging area focuses on the serotonin transporters (SERT). Initial studies carried out with nonselective SPECT radiotracers such as [^{123}I]β-carbomethoxy-3-β-(4-iodophenyl)-tropane ([^{123}I]-β-CIT) and amino-4-(2-[^{11}C]dimethylaminomethyl-phenylsulfanyl)-benzonitrile ([^{11}C]DASB), which labeled SERT in the midbrain and where no differences were found. This is despite the fact that some postmortem studies have shown decreases in SERT in the frontal cortex and the cingulate cortex.

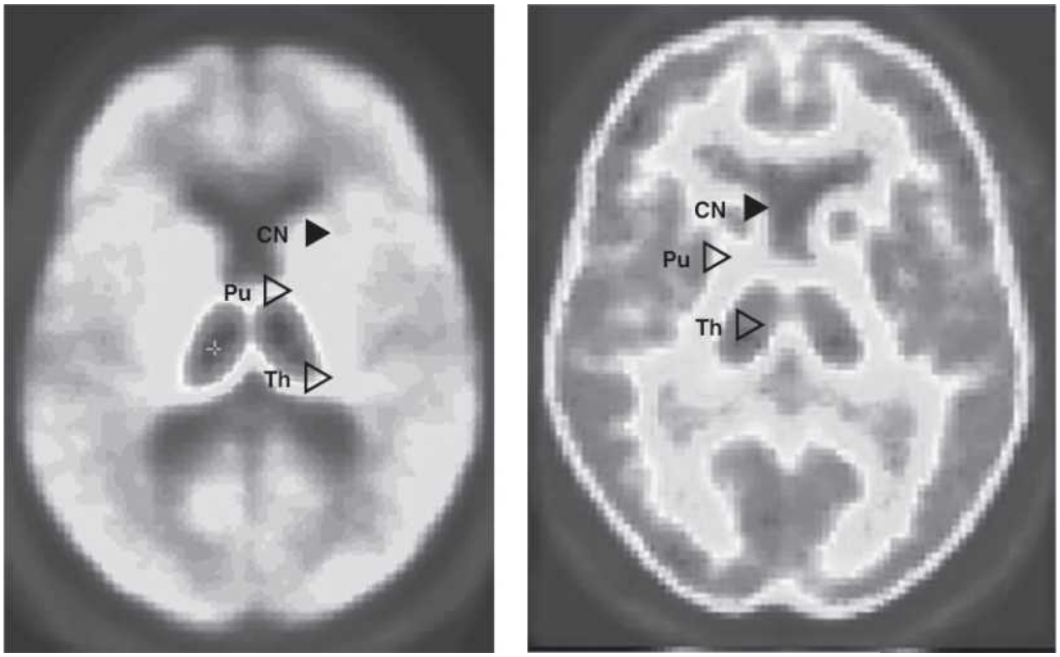

FIGURE 12.9–4. Alpha 4 beta 2 and alpha 7 nAChR human PET scans. (**left**) alpha 4 beta 2 nAChR axial slice of healthy volunteer showing high thalamic uptake and (**right**) alpha 7 nAChR axial slice of healthy volunteer showing more uniform cortical uptake.

Molecular Imaging of Serotonin Release. Despite a long attempt to measure intrasynaptic serotonin release in analogy to DAR, this may have been accomplished only recently. Erritzoe used the 5-HT_{2A} receptor agonist radioligand, [^{11}C]CIMBI-36, and 0.5 mg/kg oral D-amphetamine challenge to report synaptic 5-HT changes of up to 14%. This opens opportunities for investigating such studies in schizophrenia.

Molecular Imaging of Cholinergic Receptors in Schizophrenia

Acetylcholine represents another key neurotransmitter system likely to play a role in the pathogenesis and pathophysiology of schizophrenia. The neuronal receptors for acetylcholine in the brain are subdivided into two primary classes: Muscarinic and nicotinic. Both muscarinic and nicotinic acetylcholine receptors participate in the development and maintenance of schizophrenia.

Estimation of the Density of Muscarinic Acetylcholine Receptors. One study of the muscarinic acetylcholine receptor (mAChR) system using SPECT and 3-quinuclidinyl-4-[^{123}I]iodoenzilate([^{123}I]IQNB), a potent antagonist of muscarinic receptors, has shown a very marked decrease of mAChR in schizophrenia, consistent with some postmortem studies. However, the functional role of this finding is not fully understood. Using the same radiotracer, the investigators demonstrated greater mAChR occupancy by clozapine than olanzapine, consistent with the observation of lesser extrapyramidal side effects (EPS) associated with the two drugs.

Estimation of the Density of Nicotinic Acetylcholine Receptors. Some preliminary studies are proceeding to determine if the decreases in nicotinic acetylcholine receptors (nAChRs) observed in postmortem studies are also observed with PET imaging. This is quite relevant to schizophrenia, given the high prevalence of smoking among patients with schizophrenia. More recently, it has been possible to image alpha-7 nAChR, the subtype most associated with cognition in schizophrenia, which Wong et al. have shown using [^{18}F]ASEM, as shown in Figure 12.9–4. The first successful human PET tracers show evidence of some reduction in V_T in schizophrenia, as predicted by the pioneering postmortem work of Freedman. The evidence that not all subjects had reductions suggests this could be related to symptoms and severity of illness and will await larger samples. Also, the evidence that in vivo alpha-7 nAChR occupancy could be measured with unlabeled DXMB-A (GTS-21), an IND-approved first-generation drug, offers opportunities for target engagement of next-generation alpha-7 treatments for cognition.

Molecular Imaging of Histamine H_1 Receptors in Schizophrenia

The histamine H_1 system has also been implicated as having a role in the pathophysiology of schizophrenia. Molecular imaging with [^{11}C]doxepin has shown reductions of BP values for H_1 receptors in the frontal and prefrontal cortices and the cingulate gyrus among medicated patients with schizophrenia.

Molecular Imaging of Benzodiazepine Receptors in Schizophrenia

Deficient inhibitory neurotransmission of GABA, an inhibitory neurotransmitter, has also been implicated in the pathophysiology of schizophrenia based on postmortem studies. However, PET studies of the benzodiazepine receptors have been inconclusive.

Molecular Imaging Studies of the Glutamate System in Schizophrenia

Glutamate, the excitatory neurotransmitter, likely plays multiple roles in the pathogenesis and pathology of schizophrenia. One approach is to directly image one of the sites of the NMDA receptor channel, the PCP site. This was done with preliminary studies with *N*-(1-naphthyl)-*N*′-(3-[^{123}I]iodophenyl)-*N*-methylguanidine([^{123}I]CNS1261), which binds to the PCP site in the thalamus and other brain regions. This radiotracer binding is reduced by ketamine, an antagonist of the NMDA receptor, which also binds to this site in part. Although initial radiotracer binding studies did not show differences in patients with schizophrenia compared to healthy controls, treatment with clozapine did reduce the binding of [^{123}I]CNS1261. This suggests that clozapine exerts some of its action through the PCP site of the glutamate system. Since negative symptoms of schizophrenia were measured to be increased, one possibility is that the glutamate hypothesis of schizophrenia may include reductions of glutamate that lead to negative symptoms. Furthermore, blockade of the NMDA receptor by glutamate antagonists may precipitate positive symptoms of schizophrenia, as shown by the administration of ketamine in humans. At the same time as the ketamine challenges, negative symptoms of schizophrenia are increased as measured by the Brief Psychiatric Rating Scale (BPRS). Thus, the glutamate hypothesis of schizophrenia proposes that reductions of glutamate in the brain lead to negative symptoms of schizophrenia.

Even in healthy controls, ketamine induces hallucinations; hence the ketamine challenge is often considered a human model of mimicking schizophrenia symptoms. Studies involving ketamine infusion have been shown to significantly reduce the binding of [^{11}C]FLB 457 in the posterior cingulate/retrosplenial cortices, which suggests intrasynaptic DAR due to competition of endogenous dopamine with the radiotracer binding. This has further been supported by microdialysis studies in rodents. At the same time, ketamine-induced psychotic symptoms were associated with the changes in the radiotracer binding.

These results suggested that cortical dopamine mechanisms may relate to ketamine-induced psychotic-like symptoms in humans. The second study showed that psychotomimetic doses of (*S*)-ketamine decreased the in vivo binding of [^{11}C]raclopride to striatal D_2-like receptors along with eliciting psychotic-like symptoms with alterations in mood and cognition, and hallucinations. This decreased the raclopride BP by approximately 17% in the ventral striatum, 14% in the caudate, and 14% in the putamen, suggesting an increase in intersynaptic dopamine (i.e., DAR). The change in raclopride BP in the ventral striatum also correlated with heightened mood, including euphoria and grandiosity. This showed additional evidence that the glutamate NMDA receptor may contribute to psychotic symptoms.

Even in some but not all normal volunteers, ketamine-induced changes in [^{11}C]raclopride occurred, that were greater than placebo. In the studies where the changes exceeded placebo, they were not greater than those caused by the intravenous amphetamine, but the ketamine-induced changes were correlated with symptoms seen in schizophrenia. In summary, there is early evidence of this relationship between indirect NMDA antagonists and enhanced DAR in the striatum and elsewhere and extrastriatal areas. Such challenge studies with NMDA antagonists may simulate a hypoglutamatergic state, which is increasingly thought to be consistent with schizophrenia and the downstream effects on dopamine and may also underlie positive symptoms.

Interestingly, studies with acute ketamine did not alter serotonin binding with [^{18}F]altanserin in healthy volunteers. The authors considered this to be more of a technical finding. [^{18}F]altanserin may be more stable in acute ketamine challenges but might be useful in studying chronic plastic effects of ketamine on 5-HT$_{2A}$ receptors. This does not imply that NMDA antagonists have no effect on serotonin, which remains an unanswered question.

Glutamate and dopamine likely interact in people with schizophrenia. Blockade of the NMDA receptors by glutamate antagonists may precipitate positive symptoms of schizophrenia and produce DAR in the lateral prefrontal and anterior cingulate cortex.

A recent review of molecular imaging of the glutamate system has suggested that, except for ligands of the GluN2B subunit of NMDA receptors, there has been little success in developing radiotracers for imaging ionotropic glutamate receptors. In contrast, mGluR receptor imaging, especially mGluR5, has been used in schizophrenia, fragile X, autism spectrum disorders, substance abuse, alcohol use disorders, and aging and has been more successful.

Cannabinoid PET Imaging in Schizophrenia

An important neurotransmitter system in schizophrenia that may play a role in the developing brain and the risk for psychosis is the cannabinoid system. Three PET ligands ([^{11}C]OMAR, [^{18}F]MPEP, and [^{18}F]MK9470) have been developed and applied to cannabinoid receptor type 1 (CB1) imaging in schizophrenia. One of the first studies of [^{11}C]OMAR showed a fall with age, and a subset of patients with schizophrenia showed elevations beyond the 95% prediction limits of the distribution volume V_T versus age. Another study with a larger sample found decreases in binding with 25 male-only SCZ subjects (SCZs) (18 antipsychotic treated and 7 antipsychotic free amygdala, caudate, posterior cingulate cortex, hippocampus, hypothalamus, and insula. Also, a study using a different CB1 tracer ([^{18}F]MK-9470) showed elevations in the ventral striatum related to negative symptoms. Despite the discrepancies, a subsequent schizophrenia postmortem study in collaboration with the Lewis and Wong labs with [^{3}H]OMAR in the same tissue that showed decreases in mRNA CB1 showed elevations consistent with the original Wong et al. 2010 findings. Also relevant is the finding of elevated male CB1 compared to females using [^{18}F]FMPEP-d$_2$, where CB1R availability was, on average, 41% higher in males with a regionally specific effect larger in the posterior cingulate and retrosplenial cortices. [^{11}C]OMAR healthy control females had significantly higher V_T estimates than males in all brain regions examined. This may help explain more recent findings using the 3rd generation PET CB1 ligand [^{18}F]FMPEP-d$_2$ and [^{11}C]MePPEP, where availability of CB1R was lower in antipsychotic-treated and untreated cohorts relative to matched controls. Exploratory analyses indicated that greater reductions in CB1R levels were associated with greater symptom severity and poorer cognitive functioning in male patients. These discrepancies and sex differences need to be addressed with larger samples and perhaps more optimal CB1 PET radioligands.

Opiate PET scanning in Schizophrenia

Using [^{11}C]carfentanil PET scans in 19 schizophrenia patients and 20 age-matched healthy controls, Ashok et al. showed a significantly lower MOR availability in patients with schizophrenia in the striatum (Cohen's d = 0.7) and the hedonic network. They also reported a marked global increase in inter-regional covariance of MOR availability in schizophrenia, largely due to increased cortical-subcortical covariance. This may relate in part to the opiate-dopamine interaction.

Molecular Imaging of Other Neurochemical Systems

Although PDE10A, a second messenger site, has been considered a therapeutic site, Marques et al. found that patients with schizophrenia have normal availability of PDE10A in brain regions thought to be involved in the pathophysiology of this disorder and concluded that it does not support an altered PDE10A availability in schizophrenia. Unfortunately, PDE10A-targeted medications have not shown efficacy in schizophrenia.

Neuroinflammation in Schizophrenia. Neuroinflammation has been associated with schizophrenia throughout the literature. This has included elevated cytokines that may be due to activated microglia. The neuroinflammatory observation seems to be most associated with early illness stages and patients with acute psychosis symptom exacerbations. The current standard in TSPO PET studies for investigating neuroinflammation in schizophrenia shows discrepancies in outcomes. Moreover, this method requires the additional step and cost of a priori genotyping. Studies with TSPO PET with first and second generation radioligands have not convincingly shown elevations in chronic schizophrenia. TSPO suffers from variable signal-to-noise and, more seriously, the need to genotype subjects to separate high and possibly midaffinity binders which can be studied but omitting those with low-affinity binding. Recent reviews have cast doubt on TSPO as a schizophrenia imaging target, and meta-analyses have shown only small elevations and others none. Recent reviews have even suggested decreases in schizophrenia and acute psychosis. Despite inconsistency with TSPO PET, studies have shown elevations of plasma inflammatory markers, including IL-6 in plasma and CSF. Cox-2 is elevated in schizophrenia, but PET ligands are not yet well developed nor for cannabinoid (CB2). Recently, [^{11}C]S1P1 has been developed for human use and may be a promising alternative. Findings of Chand and colleagues have found a direct agreement of mRNA, protein S1P1 (the latter which is measured by [^{3}H]S1P1 autoradiography as a surrogate for PET) in one of two subgroups of patients who died with schizophrenia where both mRNA and autoradiography shows elevations compared to another subgroup where there was no elevations by comparison nor different from controls. It is provocative but conceivable that this elevation of S1P1 in one subgroup of patients with schizophrenia corresponds to one of the two distinct subtypes of patients identified by sMRI by Chand in two large machine learning studies (see below section) with clinical differences as well. Thus there is great potential for studying neuroinflammation (or related mechanisms) with other neuroinflammatory tracers like S1P1.

MOLECULAR IMAGING IN DRUG DEVELOPMENT FOR TREATMENTS IN SCHIZOPHRENIA

One of the most challenging aspects of psychiatric pharmacotherapy is the administration of clinically effective drug dosages that cause no or minimal side effects. A recent review outlines many of the challenges of antipsychotic action. At a given dose, plasma concentrations of psychotropic drugs vary extremely because of large inter-individual differences in absorption, metabolism, and excretion. PET and, more recently, SPECT have been used extensively to characterize the relationships between the occupancy of target molecules in the brain (neurotransmitter receptors and transporters) and plasma concentrations of the respective drug. This has been attempted to correlate with clinical efficacy and side effects. PET occupancy measures have now reached the status of almost surrogate markers that have been proven to be critical in decision-making in the development of new psychotropic drugs, especially for choosing a dose for expanded clinical trials. Determination of the appropriate occupancy level for various classes of drugs is vital to streamlining drug development. These studies have also helped elucidate the mechanism of action of some psychotropic drugs, including some of the newer antipsychotics.

D_2 Receptor Occupancy and Beneficial and Adverse Effects of Antipsychotics

With regard to the relationship between drug kinetics in plasma and the brain, as well as clinical effects and side effects, antipsychotics belong to the best-explored drugs. Lars Farde and collaborators demonstrated that clinically effective doses of antipsychotics occupy D_2-like dopamine receptors in the range between 65% and 90%. The suggestion of a “therapeutic window” between 60% and 80% striatal D_2 receptor occupancy for sufficient treatment response and a “ceiling” of around 80% occupancy for EPS was later confirmed by the same and a number of other groups (Fig. 12.9–5). Although the class of antipsychotics is characterized by marked pharmacologic heterogeneity, the general rule that the incidence of EPS increases in a dose-dependent manner does also apply to most of the second-generation antipsychotics, including olanzapine and risperidone. They lose their "atypical" properties when striatal D_2 receptor occupancy is raised above a threshold of approximately 80%. The 5-HT_2 antagonism, which characterizes most of the second-generation antipsychotics, seems to protect from EPS only at moderate doses; this preservation is lost at striatal D_2 receptor occupancies above the 80% threshold. Thus, when the doses of these drugs are raised above a certain threshold (usually approximately 6 mg for risperidone, 30 mg for olanzapine, and 160 mg for ziprasidone [Geodon]), striatal D_2 occupancy increases to levels that are associated with a higher incidence of EPS.

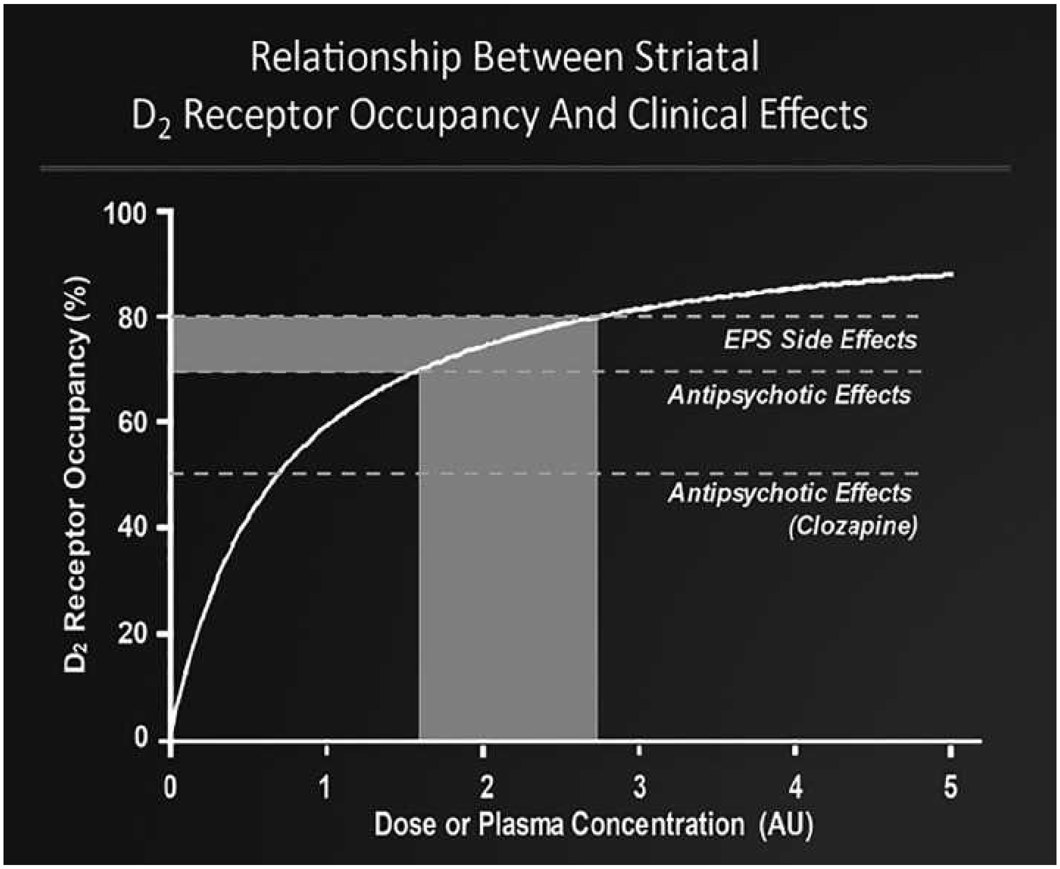

FIGURE 12.9–5. Previously unpublished composite of two studies reported by Farde et al. (1988, 1992), showing degree of haloperidol saturation of dopamine D_2/D_3 receptors measured by PET of [11C]raclopride binding under different degrees of competition. Abscissa unit is arbitrary relative haloperidol concentration. (Data from Farde L, Wiesel FA, Halldin C, Sedvall G. Central D2-dopamine receptor occupancy in schizophrenic patients treated with antipsychotic drugs. *Arch Gen Psychiatry.* 1988;45(1):71–76. doi: 10.1001/archpsyc.1988.01800250087012; Farde L, Nordström AL, Wiesel FA, Pauli S, Halldin C, Sedvall G. Positron emission tomographic analysis of central D1 and D2 dopamine receptor occupancy in patients treated with classical neuroleptics and clozapine. Relation to extrapyramidal side effects. *Arch Gen Psychiatry.* 1992;49(7):538–544. doi:10.1001/archpsyc.1992.01820070032005.)

However, there are a few exceptions to this general rule. Some second-generation antipsychotics do not induce EPS, even when their doses (or plasma concentrations) are raised to unusually high values. In those cases, the upper threshold of the "therapeutic window" is defined by other side effects rather than EPS. This pharmacologically heterogeneous group of antipsychotics can be differentiated by the distinct characteristics of these drugs in PET studies. Clozapine and quetiapine (Seroquel) occupy a maximum of 60% to 70% of striatal D_2-like dopamine receptors, even at extremely high doses or plasma levels. Twenty-four hours after the last administration of quetiapine, a significant occupancy of striatal D_2 receptors is not detectable anymore. These particular characteristics of clozapine and quetiapine are most likely due to their low affinity rather than their rapid dissociation (k_{off}) from the D_2 receptor. Thus, the tolerability and safety of these drugs are limited by anticholinergic, antihistaminic, and α-adrenergic rather than EPS side effects. On the other hand, the dopamine partial agonist antipsychotics (aripiprazole, [Abilify]; brexpiprazole, [Rexulti]; cariprazine, [Vraylar]) lead to complete saturation of striatal D_2 receptors at clinically used doses, although the incidence of EPS is very low. Mechanisms of action of aripiprazole have been reviewed for PET and in vitro/in vivo characterization for possible future medications. Again, the tolerability of these three drugs is not restricted by the occurrence of EPS but by other side effects such as psychomotor activation (Fig. 12.9–6). The characterization of aripiprazole, brexpiprazole, and cariprazine as partial dopamine agonists explains this unique feature. Thus, the "therapeutic window" for treatment with most antipsychotic drugs is defined downward by a minimum D_2 receptor occupancy, which must be achieved, and upward by the EPS threshold. This rule applies both to first- and second-generation antipsychotics, with the exceptions of dopamine partial agonists and some low-affinity antagonists, including clozapine and quetiapine.

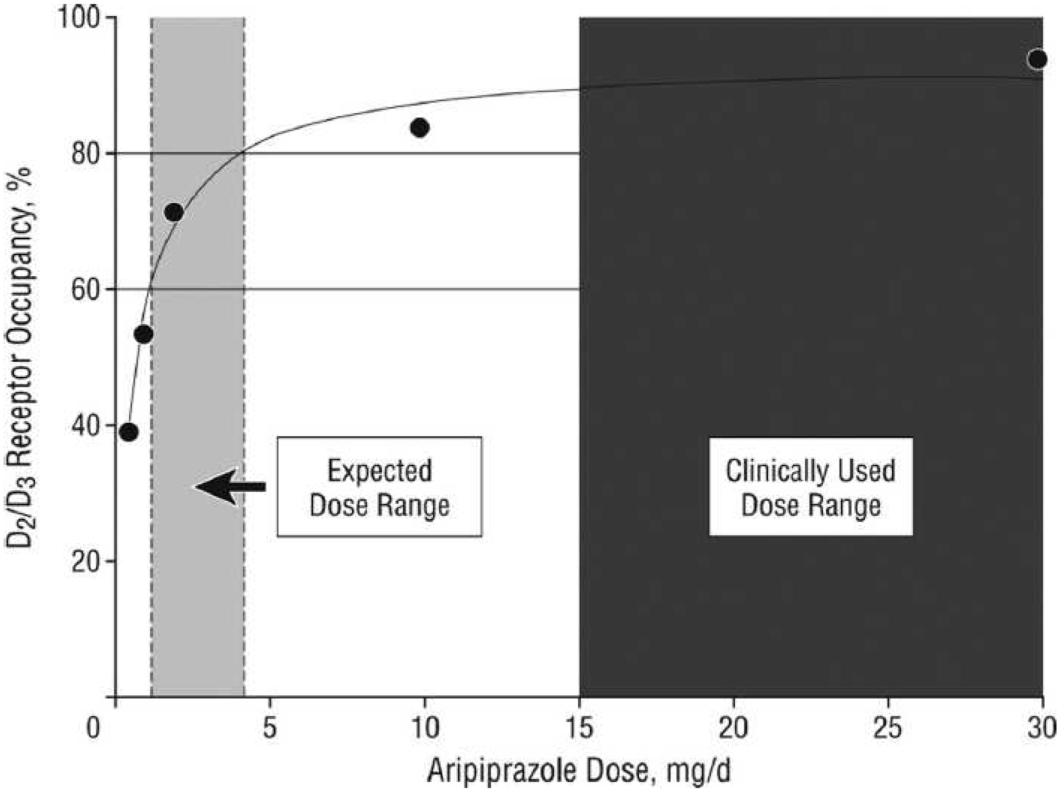

FIGURE 12.9–6. Relationship between aripiprazole (Abilify) daily dose, mean D_2 and D_3 receptor occupancy in the putamen, and clinically expected and used aripiprazole doses. *Horizontal lines* indicate the hypothesized therapeutic window between 60% and 80% striatal D_2 receptor occupancy. The *light-shaded area* designates the dose range that should lead to receptor occupancy within this therapeutic window, whereas the *dark-shaded area* indicates the dose range that has proved to have antipsychotic efficacy. (Reproduced with permission from Gründer G, Carlsson A, Wong DF. Mechanism of new antipsychotic medications: occupancy is not just antagonism. *Arch Gen Psych.* 2003;60(10):974–977. doi:10.1001/archpsyc.60.10.974. Copyright © 2003 American Medical Association. All rights reserved.)

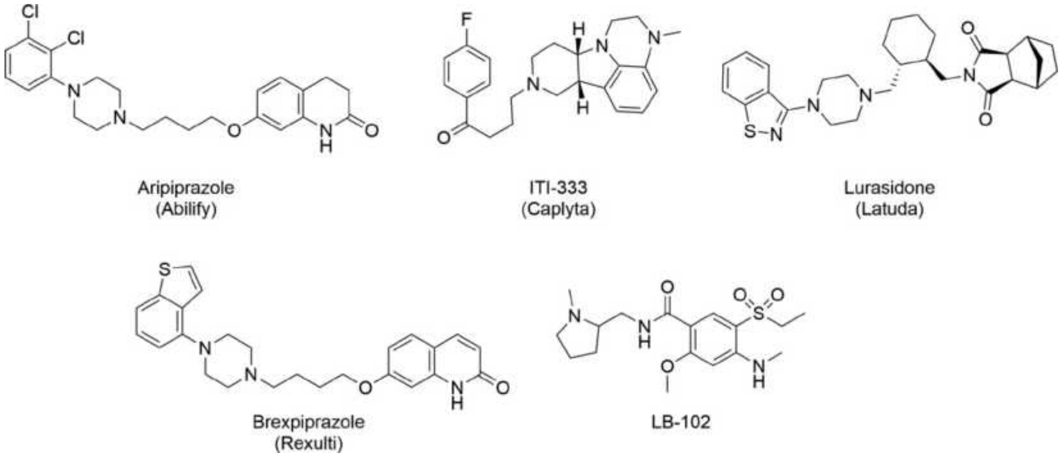

FIGURE 12.9–7. Chemical Structures of FDA approved (four) or under review (one-) antispsychotics. Aripiprazole (Abilify), Lumateperidone (Caplyta), Lurasidone (Latuda), Brexpiprazole (Rexulti) all FDA-approved and n-methylamisulpride (LB-102) in phase 2 pending all have used PET for target engagement of the dose in phase 1 leading to dosing in schizophrenia.

Receptor occupancy of the D_3 receptor has been suggested as a possible treatment for schizophrenia. Blonanserin has a high affinity for D_3 and D_2 receptors in vitro. [^{11}C](+)PHNO has been used to directly measure both D_3 and D_2 in healthy humans after a single dose of 12 mg blonanserin for the D_2-rich region was 0.39 ng/mL, and EC_{50} for the D_3-rich region was 0.40 ng/mL and target engagement in schizophrenia, with an average occupancy by blonanserin (12.8 ± 5.6 mg/day) of EC50 was 0.30 ng/mL for D_2 receptor for caudate and putamen.

New uses of first-generation, previously efficacious antipsychotics include nonracemic amisulpride (SEP-4199) and methyl amisulpride (LB-102) (Fig. 12.9–7).

Extrastriatal Occupancy/Target Engagement. However, all the observations mentioned above were made with reference to receptor occupancy in striatal structures only. With the advent of high-affinity radiotracers belonging to the class of substituted benzamides, such as [^{18}F]fallypride, it became possible to visualize and quantify extrastriatal dopamine receptors. Earlier PET studies with the moderate affinity D_2 or D_3 antagonist [^{11}C]raclopride demonstrated that the second-generation antipsychotics, clozapine, and quetiapine occupy striatal D_2 or D_3 receptors to a significantly lesser extent than other antipsychotics. It was later consistently shown in studies with high-affinity ligands that these two compounds nonetheless occupy a significantly higher proportion of temporal-limbic than striatal D_2 or D_3 receptors. Investigators have demonstrated that the clinical antipsychotic efficacy of clozapine is likely more related to its binding in the temporal cortex than to its striatal binding. This “preferential” extrastriatal binding of second-generation antipsychotics has since been shown for several other second-generation antipsychotics, at least to a modest extent (Figs. 12.9–7 to 12.9–9).

These findings have important implications not only for drug development but also for clinical practice. Once a plasma concentration range that leads into the “therapeutic occupancy window” has been established using PET, it can be tested to assess if this range is the most useful in large patient populations. This has indeed been demonstrated for various antipsychotic drugs. A well-documented and clinically important example is risperidone. PET studies show that at 6 mg of risperidone daily, the highest dose administered under usual clinical conditions, a significant proportion of patients present with striatal D_2-like occupancies sufficient to cause EPS. In contrast, a daily oral dose of risperidone of 3 to 4 mg produces a striatal D_2-like occupancy in the range of 70% to 80% in most patients. The available data suggest that the 80% threshold associated with a higher probability of EPS is reached with a plasma concentration of approximately 40 ng/mL. A plasma concentration of approximately 15 ng/mL is associated with 60% striatal D_2 occupancy. The same therapeutic window seems to apply to long-acting injectable risperidone. Clinical studies with long-acting injectable risperidone at various doses have demonstrated that a dose of 75 mg (used in clinical testing but is not approved) is associated with a relatively high EPS incidence and correspondingly high plasma concentrations above 40 ng/mL. Most patients treated within the approved dose range of 25 to 50 mg every 2 weeks present with plasma concentrations in the range of 15 to 40 ng/mL. Indeed, the 25 mg dose might be too low in some patients. This example illustrates that the knowledge of an individual’s plasma level can provide meaningful information on psychotropic drug concentrations at brain target molecules.

Dopamine Receptor Occupancy and Relapse. Using PET occupancy studies as the rationale, a recent study has suggested that when antipsychotics are reduced, it should be done gradually (over months or years) and in a hyperbolic manner (to reduce D_2 blockade “evenly”). For example, reducing by one quarter (or one half) of the most recent dose of antipsychotic, equivalent approximately to a reduction of 5% (or 10%) age points of its D_2 blockade, sequentially (so that reductions become smaller and smaller in size as total dose decreases), at intervals of 3 to 6 months, titrated to individual tolerance.

Serotonin Occupancy by PET. For almost all of the newer, second-generation, and partial agonist antipsychotics, high occupancy of serotonin 5-HT_{2A} receptors by these drugs has been documented. While 5-HT_{2A} occupancy is substantially higher than D_2 occupancy under treatment with the D_2 antagonists, including olanzapine, quetiapine, and risperidone, 5-HT_{2A} occupancy under treatment with one of the three available partial agonists (aripiprazole, brexpiprazole, cariprazine) is lower than D_2 occupancy, in the range between 20% and 40%. The contribution of antagonism by this receptor to the clinical effects of these compounds remains controversial; it may reduce the risk for EPS and possibly improve negative symptoms and cognitive dysfunction. With [^{11}C]NMSP, Gründer et al. first showed the dissociation of brain persistent 5-HT_{2A} occupancy of a single dose of 10 and 20 mg of unlabeled MDL100,907, a selective serotonin 5-HT_{2A}-selective antagonist, versus the relatively rapidly declining plasma concentration. Occupancy studies of quetiapine and [^{18}F]altanserin as the radiotracer have shown a significant positive correlation between neocortical binding prior to treatment and subsequent increases in BMI. At follow-up, the 5-HT_{2A} receptor occupancy was positively correlated with BMI increase.

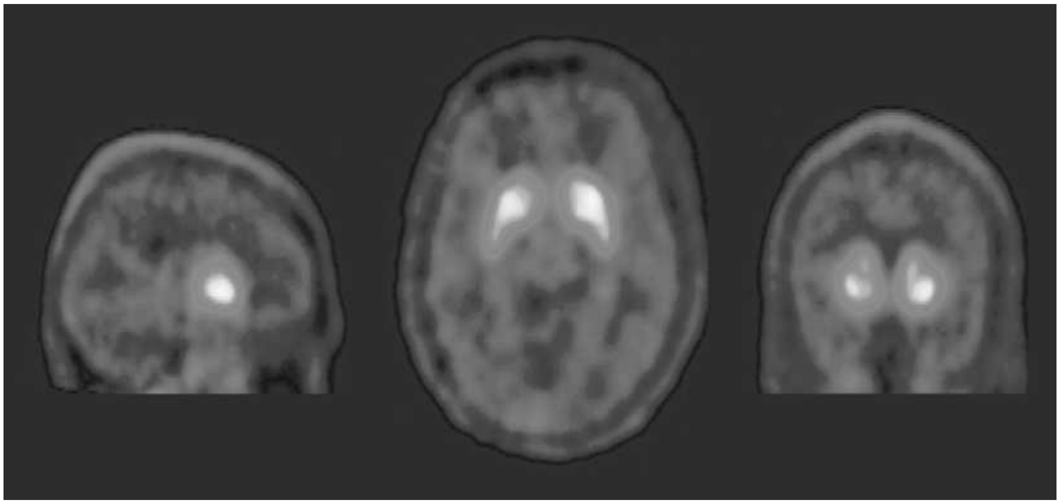

FIGURE 12.9–8. Extrastriatal dopamine binding with [^{18}F]Fallypride. Representative sagittal (**left**), transversal (**middle**), and coronal (**right**) images from positron emission tomography (PET) with [^{18}F]fallypride ([^{18}F]FP), a radiotracer for extrastriatal D_2/D_3 receptors in the brain, of a 24-year-old man with his first episode of schizophrenia treated with aripiprazole until 78 hours before the tomography. The average striatal D_2/D_3 receptor occupancy was 82%. At the moment of tomography the aripiprazole serum concentration was 106 ng/mL. (Reprinted with permission from Gründer G, Fellows C, Janouschek H, et al. Brain and plasma pharmacokinetics of aripiprazole in patients with schizophrenia: an [18F]Fallypride PET Study. *Am J Psychiatry.* 2008;165(8):994 (Copyright © 2008). American Psychiatric Association. All Rights Reserved. doi: 10.1176/appi.ajp.2008.07101574.)

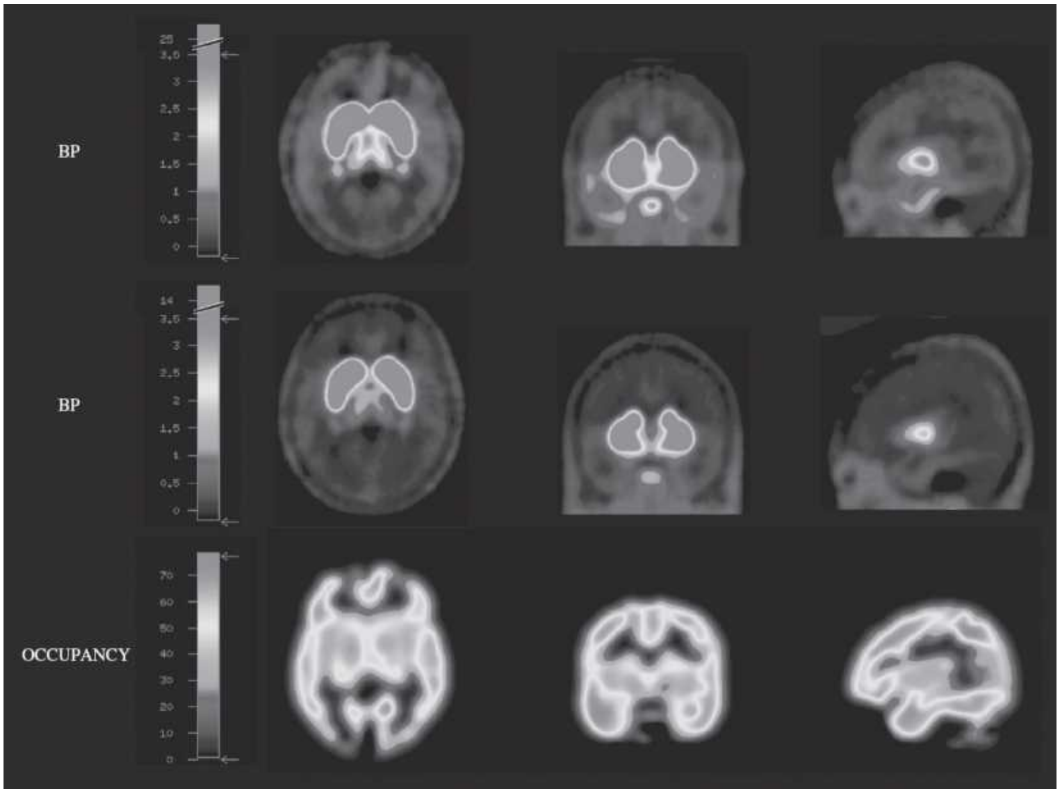

FIGURE 12.9–9. $[^{18}F]$Fallypride target engagement of cortical and subcortical dopamine D_2/D_3 receptors. $[^{18}F]$fallypride positron emission tomography (PET) parametric mean binding potential (BP) images in normal volunteers (*n* = 7, **top row**) and in clozapine (Clozaril)-treated patients with schizophrenia (*n* = 15, **middle row**). The **bottom row** illustrates the D_2/D_3 receptor occupancy in each single voxel in the patient group, clearly demonstrating the significantly higher binding in cortical regions compared to the striatum. (Reprinted from Gründer G, Landvogt C, Vernaleken I, et al. The striatal and extrastriatal D2/D3 receptor-binding profile of clozapine in patients with schizophrenia. *Neuropsychopharmacology.* 2006;31:1031, with permission from Springer Nature. doi:10.1038/sj.npp.1300931.)

Occupancy Studies of Other Molecular Targets with Potential Relevance in Schizophrenia

Because of its increasing relevance for the pathophysiology of schizophrenia, molecular targets related to glutamatergic neurotransmission have become a topic of major interest in molecular imaging research for more than a decade (see above). Target engagement of some new drugs binding to these targets has been evaluated in PET studies that provided new insights into the clinical relevance of these compounds. One of the targets that attracted significant attention was the glycine transporter type 1 (GlyT1). In preclinical models, GlyT1 inhibitors markedly increase the glycine concentration in the glutamate synapse, putatively at least partly reversing the deficit in glutamate neurotransmission that is believed to underlie aspects of the psychopathology of schizophrenia. Several radiotracers for the GlyT1 have been developed and subsequently used in occupancy studies with unlabeled inhibitors Wong and colleagues showed that the GlyT1 inhibitor bitopertin occupies the GlyT1 transporter in a dose-dependent fashion, with the highest administered dose, 175 mg daily, occupying around 85% of the target. However, in clinical studies, the dose with the best, albeit not sufficient, efficacy against negative symptoms was 10 mg daily. At that dose, a maximum of only about 40% of the GlyT1 is occupied. Studies in monkeys support the view that GlyT1 inhibitors exert beneficial effects on prefrontal function in cognitively disturbed animals at GlyT1 occupancy values of 10% to 30% and that these effects are lost at higher occupancies. It is presently unclear how these findings have to be interpreted and whether the clinical development of this class of drugs should be further pursued. PET tracers for other glutamatergic targets, such as metabotropic glutamate receptors, have been developed; to our knowledge, none of them have advanced to application in human PET studies of target engagement of investigational compounds, in part due to the absence of robust PET ligands such as for $mGluR_{23}$ a previous important therapeutic target in schizophrenia.

ISSUES OF QUANTIFICATION OF NEURORECEPTORS, TRANSPORTERS, AND INTRASYNAPTIC NEUROTRANSMITTERS

Controversy over Dopamine Molecular Imaging Markers

The original findings in postmortem studies of Dopamine D_2/D_3 receptors suggested elevations consistent with the dopamine hypothesis of schizophrenia. These findings in the late 1970s were potentially confounded by medication effects. The advent of molecular imaging in vivo in schizophrenia allowed studies without the confound of medication effects. Initial studies with $[^{11}C]$NMSP were first reported in 1986 by Wong et al. and replicated with his colleagues, and $[^{11}C]$raclopride by Farde et al. in the same time period did not show elevations and hence the inconsistent results; however, recent meta-analyses have shown average elevations of dopamine D_2/D_3 receptors across a number of D_2/D_3 PET radiotracers.

More consistent findings have revealed elevated FDOPA metabolism measures in schizophrenia with and without treatment. Another but less consistently reproduced finding is elevated DA release following intravenous amphetamine challenge. Also, at least two studies showed increased release of DA concentration following an intravenous amphetamine challenge. Here we describe the details and potential pitfalls of quantification of pre versus postmeasures of intrasynaptic DA in schizophrenia. Finally, the details of occupancy (i.e., target engagement) of drugs in relation to therapy are nontrivial in the absence of a proper region of reference with no specific binding, as is the case with multiple neuroreceptor systems. Here, we describe the use of so-called inhibition plots of drug occupancy in cases of absent reference regions of no binding:

Reports from three studies claimed increased DA release after amphetamine administration, judged from the reduction of the BP of tracer benzamide ligands ([^{11}C]raclopride or [^{123}I]IBZM), calculated relative to the cerebellum as a reference region with no binding. However, the reference region approach can be shown to potentially underestimate the change of BP change to such an extent that no significant change is evident, although the lack of a measurable change is no proof of absent elevated DAR. The only definitive method of accurate determination of the decline of BP depends on measures of tracer concentrations in the arterial circulation that supplies the regions in question, such as TREMBLE (Transient Receptor Equilibrium Method of Bolus Ligand Estimation), see below.

It has been shown that inhibition of dopamine synthesis through AMPT results in lower estimates of BP in patients with schizophrenia versus healthy control subjects. Although there is no difference between BP estimates in patients with schizophrenia and healthy control subjects with [^{11}C]raclopride imaging of dopamine receptors in the absence of AMPT, the evidence suggests higher baseline levels of intrasynaptic DA in the patients with schizophrenia. The same finding was evident from the increased apparent affinity of [^{11}C]NMSP in patients with schizophrenia (Gjedde and Wong, 2001). This can also be shown from the increased baseline DA levels in patients with schizophrenia when the measure of BP is consistent with elevated values of receptor density ($B_{\max}$) in the patients. When baseline BP estimates of patients with schizophrenia are greater than baseline BP estimates from healthy control subjects in the absence of elevated intrasynaptic DA and identical binding affinities (K_D), it follows that the maximum number of receptors in patients with schizophrenia ($B_{\max}{}^{\mathrm{SCZ}}$), equal to the BP relative to the affinity constant (BP^{SCZ}/K_D), is greater than the number of receptors in control subjects ($B_{\max}{}^{\mathrm{ctr}}$) that equals the BP relative to affinity (BP^{ctr}/K_D) in this group of subjects, as further evidence of the presence of elevated D_2/D_3 receptor numbers as measured by PET in the patients with schizophrenia.

Pharmacokinetics of Dopaminergic Neurotransmission. In Figure 12.9–10, the 50% inhibitory concentration or IC_{50} of the tracer depends both on the receptor's affinity toward the tracer and the concentrations of competitors.

The most commonly reported variable obtained with imaging of dopaminergic neurotransmission is the ratio between the density of receptors and the affinity constant, known as the BP. Both terms are part of Michaelis' and Menten's steady-state solution (1913) to the binding equations as follows,

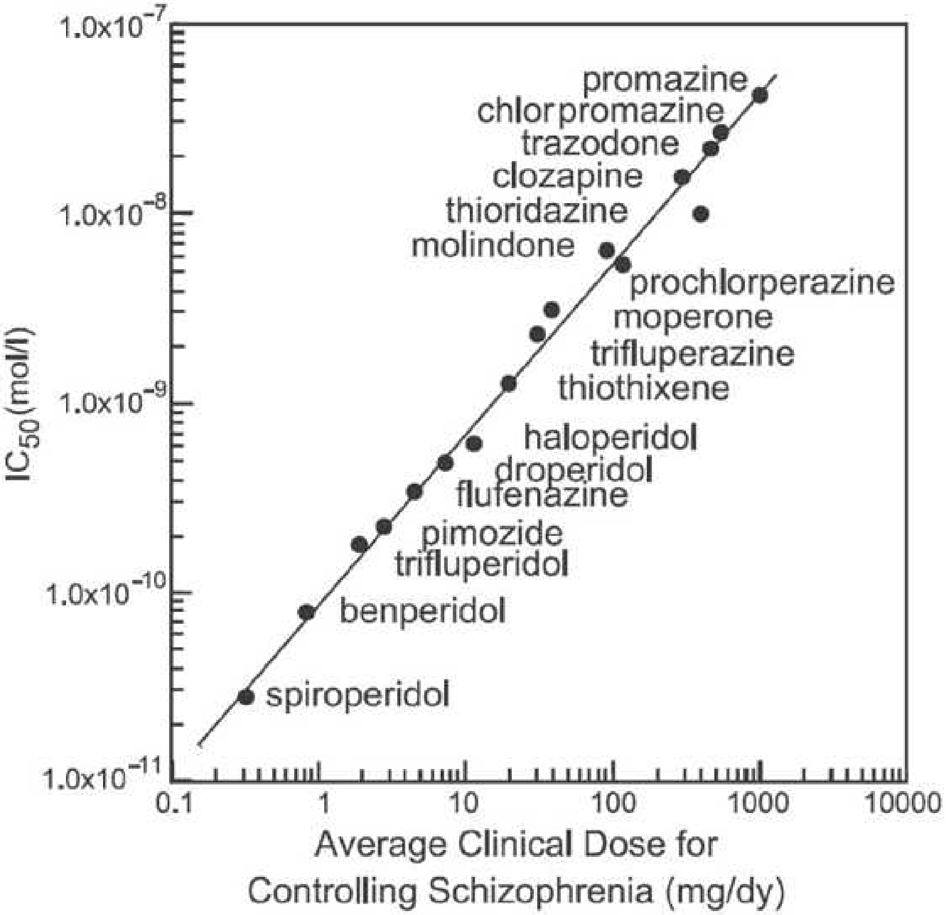

FIGURE 12.9–10. In vitro versus in vivo action of first-generation antipsychotics. Relationship between IC50 values at dopamine receptors and therapeutic dose of neuroleptics in treatment of schizophrenia. The IC50 values are the concentrations of the antipsychotic drugs that reduce the stereospecific component of [3H]haloperidol binding by 50%. The abscissa indicates the average values of doses used in schizophrenic patients. (Redrawn from Seeman P, Lee T, Chau-Wong M, Wong K. Antipsychotic drug doses and neuroleptic/dopamine receptors. *Nature.* 1976;261:717–719, with permission from Springer Nature. doi: 10.1038/261717a0.)

$$B = \frac{B_{\max}\ C}{K_D + C} \tag{1}$$

where B is the bound quantity of the ligand, $B_{\max}$, the maximum number of binding sites, C is the unbound ligand's concentration, and K_D is the Michaelis–Menten dissociation constant or half-saturation binding concentration of the ligand. The BP is the ratio between the bound and unbound ("free") quantities of a ligand, including an endogenous ligand, relative to the receptor's affinity for this ligand, that is, multiplied by the Michaelis–Menten half-saturation constant of the receptor toward the ligand in question. The two definitions of the BP express the potential in terms of available receptors (i.e., receptors not occupied by ligands) or in terms of the concentrations of ligands, as follows,

$$BP_{\mathrm{ND}} = \frac{B^*}{C^*} = \frac{B_{\max}}{C^* + K_{\mathrm{D}}^* \quad 1 + \frac{C_{\mathrm{i}}}{K_{\mathrm{I}}}} = \frac{K_{\mathrm{I}} B_{\max}}{K_{\mathrm{D}}^* \left(K_{\mathrm{I}} + C_{\mathrm{I}} \right)} = \quad \frac{K_{\mathrm{I}}}{K_{\mathrm{D}}^*} \quad \frac{B_{\mathrm{i}}}{C_{\mathrm{i}}} \tag{2}$$

where the asterisks refer to the radiotracer. The symbol BP_{ND} represents the BP, and the remaining symbols represent the masses or concentrations of specifically bound ligand (B^*), unbound tracer (C^*), the maximum number of binding sites ($B_{\max}$), affinity toward the ligand (K_{D}), and the affinity toward any competing endogenous ligand (K_{I}) with a presence as mass or concentration in the vicinity of the receptors (C_{i}) sufficient to occupy a significant number of receptor sites (B_{i}) as mass or concentration (Fig. 12.9–11). Note that the BP is a unitless ratio that requires that the variables of the ratio have identical units.

The specific definitions depend on the process and the conditions, but generally, the magnitude of k_3 is related to the quantity of protein and the concentration of the tracer, while k_4 is the decay constant of the pool of retained substance. Depending on the steady state, the definitions are,

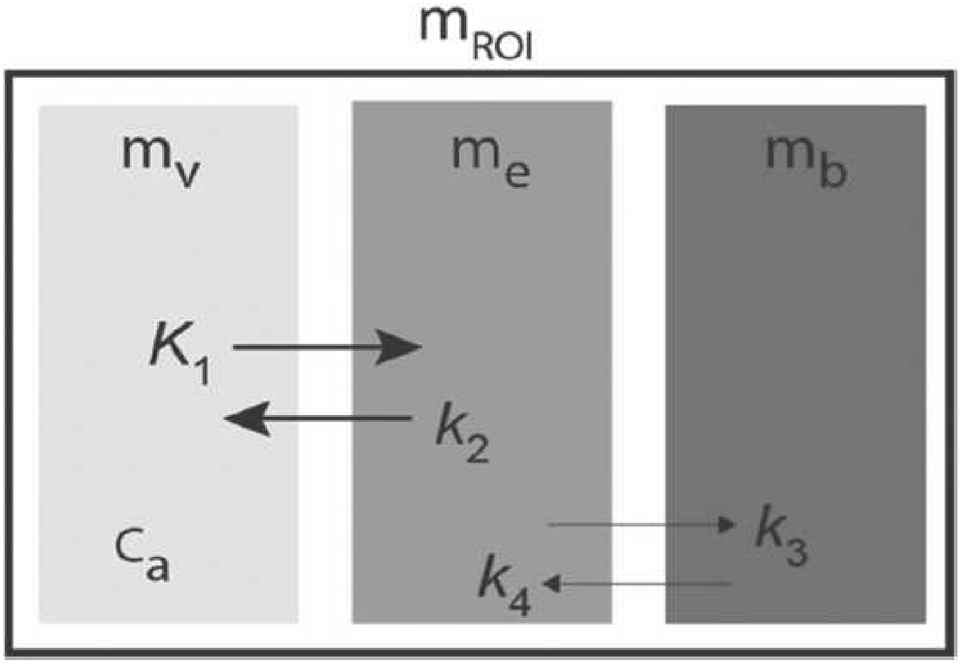

FIGURE 12.9–11. Three-compartment model. Three-compartment model of receptor binding of ligand. The transfer coefficients are clearances (units of flow) or rate constants (units of reciprocal time or clearance relative to volume of distribution), the former shown by upper-case symbols, the latter by lower-case symbols. The clearance shown as K1 is the blood–brain clearance, and the rate constant shown as K_2 is the fractional clearance from the precursor pool in brain. For definition of remaining rate constants, see text. (Reprinted from Phan JA, Wong DF, Chang NHS, et al. Transient equilibrium determination of dopamine D2/D3 receptor densities and affinities in brain. *Front Nucl Med.* 2022;2:1030387. © 2022 Phan, Wong, Chang, Kumakura, Bauer and Gjedde. doi:10.3389/fnume.2022.1030387.)

$$k_3 = k_{\mathrm{on}}\left(B_{\max} - B\right) \qquad (3)$$

and

$$k_4 = k_{\mathrm{off}} \qquad (4)$$

or

$$k_3 = k_{\mathrm{on}}\ B_{\max} \qquad (5)$$

and

$$k_4 = k_{\mathrm{off}}\ \ 1 + \frac{C}{K_{\mathrm{D}}} \qquad (6)$$

where k_3 and k_4 define the respective association and dissociation constants, of which only k_4 is a symbol that represents rate constants. The ratio k_3/k_4 is a function of the amount of protein (such as the density of receptors) and the affinity for the tracer. The ratio is, therefore, an index of the tissue's potential for retention. In the case of binding, the definitions show that this ratio is also the BP defined earlier. The exact solutions are those given by Gjedde et al. (1986) with the calculation of the receptor density from multiple estimates of the BP when the quantity of bound ligand is known for each estimate. The bound quantities, in principle, include both endogenous and exogenous ligands, such as a drug given to reduce the BP of the tracer by competition for binding.

Estimates of the receptor density are not routinely measured in human PET studies anymore because they require significant occupation of the receptors by a ligand with a specific pharmacodynamic effect. The procedure consists of giving a pharmacologically active dose of a ligand together with the tracer (e.g., 7.5 mg oral haloperidol in the case of [^{11}C]NMSP by Wong et al. (1986, 1997 or with unlabeled raclopride together with tracer [^{11}C]raclopride in the case of Farde et al. 1986)). As many as five different estimates of the BP have been made to complete Scatchard or Eadie–Hofstee plots of the binding, as shown in Figure 12.9–12, where the specific activity was reduced by increasing amounts of carrier-added unlabeled raclopride.

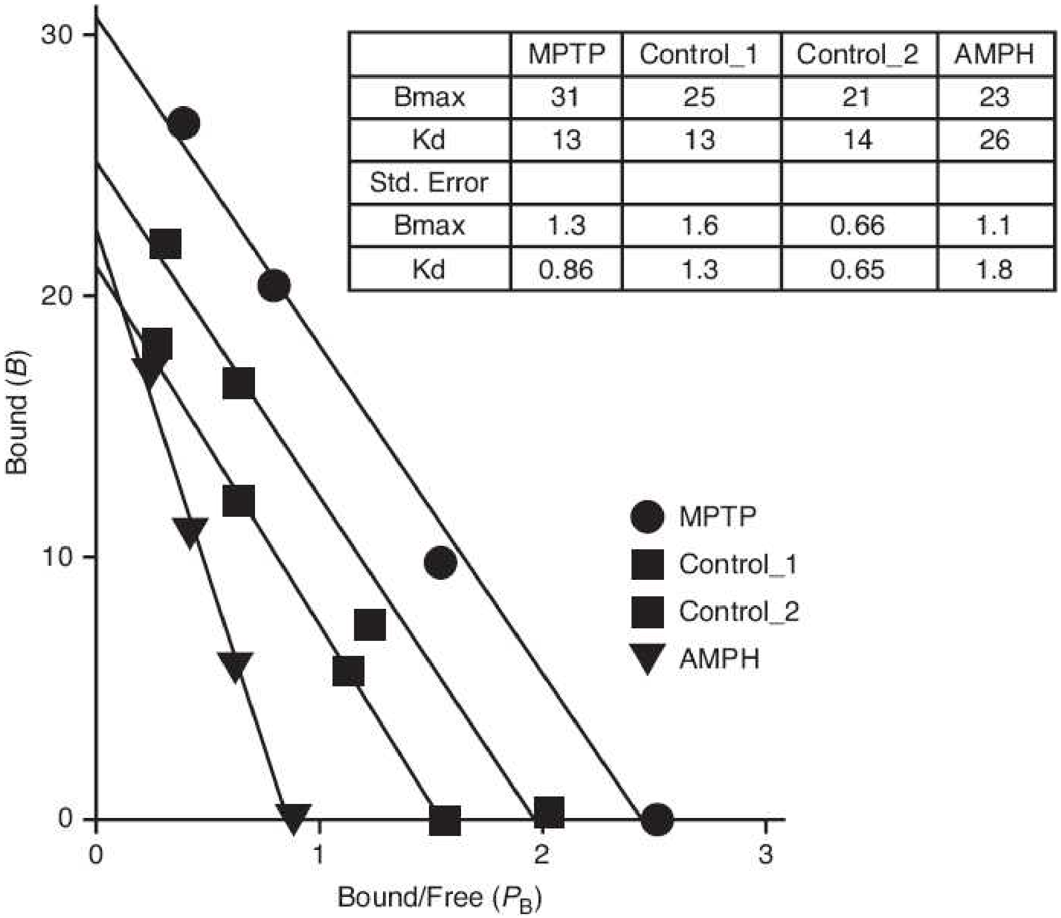

	MPTP	Control_1	Control_2	AMPH
Bmax	31	25	21	23
Kd	13	13	14	26
Std. Error				
Bmax	1.3	1.6	0.66	1.1
Kd	0.86	1.3	0.65	1.8

FIGURE 12.9–12. B_{max} and K_d estimates from multiple specific activities with [^{11}C]raclopride in nonhuman primates studies. Eadie–Hofstee plots of dopamine receptor density and radioligand affinity, determined with [^{11}C]raclopride in monkeys rendered parkinsonian by means of MPTP administration, in comparison with control monkeys. Eadie–Hofstee and Scatchard plots are linear renditions of the equation that defines the binding potential. This is also the equation of the Scatchard plot, in which the abscissa is the total quantity of bound ligand (**B**), usually determined from the specific activity and binding potential of the radioligand. The ordinate is the binding potential (BP_{ND}), determined as the ratio of bound to unbound ligand from binding regions-of-interest and a reference region of no binding, or from regression analysis estimates of the rate constants k_3 and k_4 or their ratio k_3/k_4, which equals the binding potential at steady-state. (The Eadie–Hofstee plot is drawn from data reported by Holden JE, Jivan S, Ruth TJ, Doudet DJ. In vivo receptor assay with multiple ligand concentrations: an equilibrium approach. *J Cereb Blood Flow Metab*. 2002;22(9):1132–1141.)

Calculation of Receptor Occupancy

For antipsychotics as well as a number of antidepressants, occupancies of 60% to 80% of dopamine receptors yield the best therapeutic effect. The receptor occupancy plotted against the plasma level of the drug essentially is a Michaelis–Menten graph of the therapeutic "window of opportunity" for a neuroleptic. This also appears to be true for partial agonists as well as antagonists of the dopamine receptors. This calculation is straightforward when there is a reference region, such as when measuring D_2/D_3 receptor occupancy.

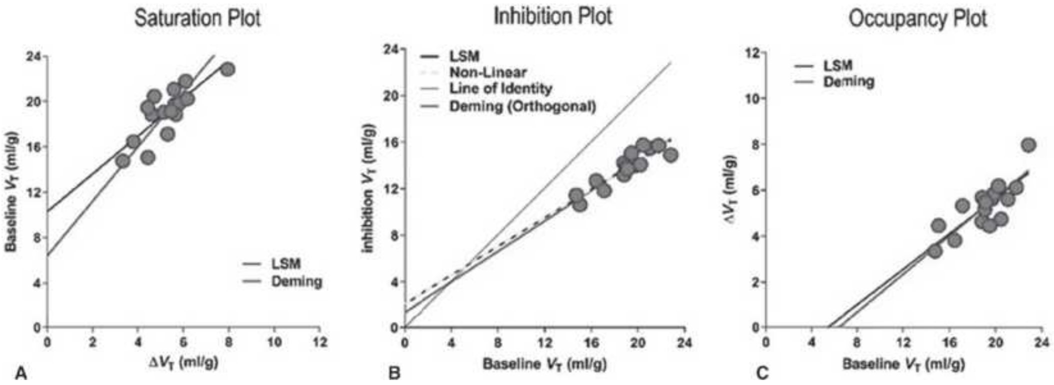

FIGURE 12.9–13. Three inhibition plots for target engagement in absence of reference region. Three linearization plots, Saturation, Inhibition, and Occupancy plots, plotted from data acquired by Horti et al. (2014) for an inhibition dose of 0.5 mg, as discussed in relation to Equations (7)–(13) in the text and by Khodaii et al. (2022). (Data from Horti AG, Gao Y, Kuwabara H, et al. 18F-ASEM, a radiolabeled antagonist for imaging the α7-nicotinic acetylcholine receptor with PET. *J Nucl Med.* 2014;55(4):672–677. doi:10.2967/jnumed.113.132068.) (See eBook for color images.)

Neuroreceptor studies of the brain in vivo using PET require comparisons of the magnitudes of BPs (BP_{ND}) of radiopharmaceutical receptor ligands at more or less inhibited receptor states from obtaining estimates of receptor density and affinity. The quantitative determination of BP_{ND} uses a fundamental equation of receptor availability to obtain separate estimates of radioligand BPs of a specific radioligand (Equation 2).

Application of any one of the three linearizations shown in Figure 12.9–13 is the first step toward the determination of the BP (BP_{ND}, also known as receptor availability) that is the foundation of the receptor binding analysis. The need for these equations arises when radioligands are applied for which a proper reference region with no specific binding of the ligand either does not exist or is not known to exist. In these cases, the three linearized versions of the basic receptor availability equation have been derived for the purpose of estimating the magnitude of the reference distribution volume of the radioligand in a region of nondisplaceable (i.e., nonspecific) binding (V_{ND}) by simple linear regression. The symbol V_{ND} refers to the distribution volume of the nondisplaceable tracer in the nonbinding compartment, also known as the partition volume or partition coefficient of the ligand.

Three different graphical plots emerged when the authors of publications by Lassen et al. (1995), Gjedde and Wong (2000), and Cunningham et al. (2010) linearized the basic equation of receptor availability to obtain the plots that we here refer to as the Saturation, Inhibition, and Occupancy plots (Equations 11 to 13 below). The plot names avoid the uncertainty of the authors associated with the Occupancy Plot solution (Equation 12), referred to by some authors as the Lassen Plot, although "Lassen" is the name more appropriately applied to the plot (Equation 11) that Lassen et al. actually derived and reported.

In the present examination of the plots of competition, we linearized the formulation of the fractional receptor availability (Equation 7) into the three equations underlying the different regressions that we refer to as the Saturation, Inhibition, and Occupancy plots. The purpose of all three linearizations is to obtain an estimate of the reference volume of distribution V_{ND} required to calculate the BP_{ND} of a radioligand. The quantitative determination of the BP (BP_{ND}) uses a fundamental equation of receptor availability to obtain separate estimates of radioligand volumes of distribution of a specific radioligand that yield a variable now generally known as "availability" (α),

$$\alpha = 1 - s = \frac{V_{T(i)} - V_{ND}}{V_{T(b)} - V_{ND}} \qquad (7)$$

that is, the formulation of relative or fractional receptor availability (1 – s) in terms of the relevant volumes of distribution when the term s represents the occupancy of the ligand. The volume of distribution terms represents the baseline and inhibition states of the radioligand, including the values of the total volumes of distribution (V_T). Of the terms, $V_{T(i)}$ represents the degree of inhibition by binding, that is, the apparent total volume of distribution of the sum of the specifically bound and nonspecifically dissolved ligand molecules occupying the receptor. The term $V_{T(b)}$ represents the baseline state of the apparent total volume of distribution, that is, when the receptor is not occupied by a specific inhibitor. Application of any one of the three linearizations presented here is the first step toward the determination of the magnitudes of BP_{ND} (or receptor availabilities) that is the foundation of the receptor binding analysis. The term BP_{ND} enters into the particular Eadie–Hofstee version of the linearized Michaelis–Menten equation that yields both the maximum binding (B_{max}) and the affinity coefficient (Michaelis half-saturation content or mass), K_D', of the receptors,

$$B = B_{max} - K_D \, BP_{ND} \qquad (8)$$

where B is the quantity of bound ligand. As above, the symbol BP_{ND} represents the ratio of the volumes of distribution of specifically bound ("displaceable") and nonspecifically bound ("nondisplaceable") ligand quantities. To determine the BP_{ND} of a radioligand, the volumes of distribution are entered into the relationship that defines the BP_{ND} (Equation 2),

$$BP_{\mathrm{ND}} = \frac{V_{\mathrm{T}} - V_{\mathrm{ND}}}{V_{\mathrm{ND}}} \tag{9}$$

which is applicable both to the receptor binding baseline and to multiple degrees of receptor blockade, provided the V_{ND} estimate is unaffected by the blockade. To calculate the value of BP_{ND}, it is necessary to know the distribution of unbound ligands in a region of no binding. However, the linearizations presented below serve to determine the reference volume of distribution of radioligands when the value of the volume (i.e., as applicable to a region of absent specific binding) is not known. From the volumes of distribution of the radioligand in the absence of displaceable binding, three different linearizations yield the BP_{ND} of radioligands, as shown in Figure 12.9–13.

Saturation Plot. As a novel steady-state approach to the determination of $\mathrm{BP}_{\mathrm{ND}}$ of tracers with an unknown reference volume of distribution, Lassen et al. in 1995 proposed to compare two levels of receptor occupancy, one essentially at zero for the labeled tracer itself and the other in the midrange of occupancy by addition of unlabeled ligand. To obtain the volume of nonspecific binding, Lassen et al. (1995) linearized Equation (7) to obtain the form of the plot we here call the Saturation plot. The plot yields the estimate of V_{ND} by plotting the baseline volume of distribution ($\mathrm{V}_{\mathrm{T(b)}}$) as a function of the difference between the baseline and inhibition volumes of distribution,

$$V_{\mathrm{T}} = V_{\mathrm{T(b)}} - V_{\mathrm{T(i)}} \tag{10}$$

that yields the basic Equation 11 of the Saturation plot,

$$V_{\mathrm{T(b)}} = (1/s) \quad V_{\mathrm{T}} + V_{\mathrm{ND}} \tag{11}$$

where the estimate of V_{ND} is the ordinate intercept of the linear regression, and the estimate of the reciprocal of the saturation (1/*s*) is the slope of the regression.

Inhibition Plot. Realizing that the uncertain choice of a reference volume of distribution of the ligand can lead to erroneous estimations of the occupancy, Gjedde and Wong (2000) proposed to linearize Equation (7) to obtain the form of the Inhibition Plot. The plot yields the estimate of V_{ND} by relating the inhibited volume $V_{\mathrm{T(i)}}$ to the baseline volume $V_{\mathrm{T(b)}}$ by linear regression, as shown in Equation 12 (also see Table 12.9–2),

$$V_{\mathrm{T(i)}} = (1 - s) V_{\mathrm{T(b)}} + s V_{\mathrm{ND}} \tag{12}$$

where the estimate of V_{ND} is obtained as the intercept of the linear regression line with the line of identity.

Occupancy Plot. In 2010, Cunningham et al. inverted the axes of the Saturation plot to obtain the plot of Equation 13 below. The authors presented the axes of the inverted Saturation plot as the plot we here call the Occupancy Plot to avoid the lack of specificity of the term "Lassen Plot." The linearization of the Occupancy Plot treats the differences between the volumes of distribution at the baseline and challenge conditions, represented by V_T as for the Saturation plot, shown here as a function of the baseline volume of distribution,

$$V_T = sV_{\mathrm{T(b)}} - sV_{\mathrm{ND}} \qquad (13)$$

where V_{ND} is the abscissa intercept. It is evident from the derivations that the Saturation and Occupancy plots have mutually inverted axes.

Based on goodness of fit parameters and with two regression methods (Least Squares and Deming II), we have shown the Inhibition and Occupancy plots to have higher degrees of convergence, most likely because of the absent commingling of the dependent and independent variables characteristic of the Saturation or original "Lassen" plot. Concerning the effect of the regression method (Least squares and Deming II), the average differences of results of the Inhibition and Occupancy plot linearizations on average are less than 1% and, as such negligible. In contrast, there is more than a 35% difference between the results of the Saturation plots by Least Squares and Deming II regressions because of the violation of the negligible variability rule for independent variables. Noise analysis shows that the three plots result in the same parameter estimates in the absence of noise, but in the presence of noise, the Inhibition and Occupancy plots yield higher and closer degrees of convergence.

Also, the inhibition plot has been extended to include additional effects of plasma and brain protein binding (Phan et al., 2017). Accurate quantification of target engagement (occupancy) of receptor-mediated antipsychotic treatments is vital, especially in the absence of a reference region for the receptor of interest. This is more common for newer receptor targets (e.g., CB_1, or α7 nAChR) regardless of the PET ligand, so the issue is fundamental to quantification in the assessments of new treatments offered to patients with schizophrenia.

MOLECULAR IMAGING AND GENETICS

Schizophrenia has a heritability of about 80%, but the detailed molecular genetic basis of the disease remains to be fully determined. Molecular brain imaging, as applied to the study of schizophrenia, can be a useful means to quantify mediators of this genetic vulnerability. Two approaches have been applied. Molecular imaging studies have been performed in individuals at genetic risk for schizophrenia. Increased cortical D_1 and caudate D_2 densities have been observed in monozygotic compared to dizygotic siblings of patients with schizophrenia. First-degree relatives of patients with schizophrenia have displayed increased uptake of [^{18}F]6-fluorodopa ([^{18}F]FDOPA) in caudate-putamen. These observations in at-risk samples are consistent with neuroreceptor binding studies in patients with schizophrenia who have demonstrated increased D_1, D_2, and FDOPA uptake. These studies highlight the potential role of molecular brain imaging in identifying genetic markers for schizophrenia as well as individuals at risk for the disease who might benefit from pharmacologic intervention. Furthermore, the potential role of molecular brain imaging in identifying genetic markers for schizophrenia is confirmed by the findings of a decrease in the prefrontal and cortical binding of 5-HT_{2A} receptors using PET and [^{18}F]altanserin in subjects at enhanced risk of schizophrenia. Those subjects at risk who developed first-episode psychosis showed an additional decrease of 5-HT_{2A} receptor BP in the caudate in comparison to the “nonconverters,” suggesting that the progressive reduction of 5-HT_{2A} receptor binding in the caudate could provide an indicator of illness activity and help to predict imminent conversion to schizophrenia. This study’s findings, although interesting, may be diminished by the very low concentration of 5-HT_{2A} receptors in the caudate, which makes it problematic to measure their binding. The second approach is to characterize genetic polymorphisms of the same transporters or receptors that are being imaged.

The use of molecular imaging also seems helpful in clarifying the contribution that different genotypes associated with dopamine receptors play in defining the phenotypic expression of the receptors themselves in schizophrenia. Impaired DAT function has been thought to be involved in antipsychotic-induced EPS. Nonetheless, studies so far conducted have been negative in this regard. The frequency of the more common alleles of the variable number of tandem repeats (VNTR) polymorphism of the DAT gene (SLC6A3) was not preferentially associated with patients who had EPS. Furthermore, no significant differences in DAT BP have been found among the three major VNTR genotype groups. Additionally, no significant association has been found between VNTR polymorphism and amphetamine-induced DAR and DAT BP density in patients with schizophrenia. These studies seem to suggest a lesser role for DAT as compared to D_2 or D_1 receptors in the pathophysiology of schizophrenia. Although not in patients with schizophrenia but with significant implications for patients with this disease is a study to determine whether common gene variants of the brain dopamine pathway explain the variability in nicotine-induced DAR. For subjects who smoked during imaging, those with at least one 9 allele of the dopamine transporter variable nucleotide tandem repeat, fewer than seven repeats of the D_4 variable nucleotide tandem repeat, and the Val/Val catechol-O-methyltransferase (COMT) genotype had greater decreases in BP (an indirect measure of DAR) with smoking than those with the alternate genotypes. An overall decrease in ventral caudate/nucleus accumbens BP in those who smoked compared with those who did not smoke was also found. These findings suggest that dopamine system genotype variabilities explain a significant proportion of the interindividual variability in smoking-induced DAR and indicate that smoking-induced DAR has a genetic predisposition. COMT has been shown to be critical for prefrontal dopamine flux and prefrontal cortex-dependent cognition and activation. Several potentially functional variants in the gene have been identified, but considerable controversy exists regarding the contribution of individual alleles and haplotypes to the risk for schizophrenia.

A pattern of inefficient prefrontal cortex engagement during working memory studied with fMRI has been found in normal volunteers in association with the Val158Met polymorphism of the COMT gene. A polymorphism of the metabotropic glutamate receptor 3 or GRM3, a G protein-linked receptor modulating synaptic glutamate, has also been associated, even in normal subjects, with a pattern of inefficient prefrontal cortex engagement during working memory studied with fMRI analogous to that found with COMT. A recent study has observed an apparent epistatic interaction of these two genes on the engagement of the prefrontal cortex during working memory. Specifically, the *GRM3* genotype putatively associated with suboptimal glutamatergic signaling was significantly associated with inefficient prefrontal engagement and altered prefrontal-parietal coupling on the background of the COMT Val-homozygous genotype. Conversely, COMT Met-homozygous background mediated against the effect of *GRM3* genotype. These findings extend putative brain dopaminergic and glutamatergic relationships indexed by COMT and *GRM3* to a systems-level interaction in human cortical circuits implicated in working memory dysfunction, such as in schizophrenia.

ROLE OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN SCHIZOPHRENIA SUBTYPING AND HETEROGENEITY ANALYSIS

Brain imaging methods, especially PET and structural MRI (sMRI), followed by advanced analytical approaches, have emerged as one of the encouraging research pillars in schizophrenia and many other mental diseases. Machine learning (ML) is a subfield that lies in a broadly named artificial intelligence (AI) field and is the current state-of-the-art standard that has emerged substantially in recent years. However, our success is heavily dependent on the accurate quantification of these imaging measures, as the effects of interest in schizophrenia are subtle, spatially distributed, and complicated in nature. The majority of current image analytics commonly overlook the heterogeneity of imaging phenotypes of schizophrenia. This means if certain neuroanatomical and molecular patterns are observed in some but not in all individuals with schizophrenia, these analytics are prone to confound the observed brain patterns.

There is a large body of literature suggesting that individuals with schizophrenia exhibit highly heterogeneous genetic profiles, clinical symptoms, illness course, and treatment response. Despite extensive efforts, understanding schizophrenia mechanisms and developing targeted, personalized clinical treatment are not encouraging, partly because there is still a substantive need for highly accurate and reproducible analytical methods.

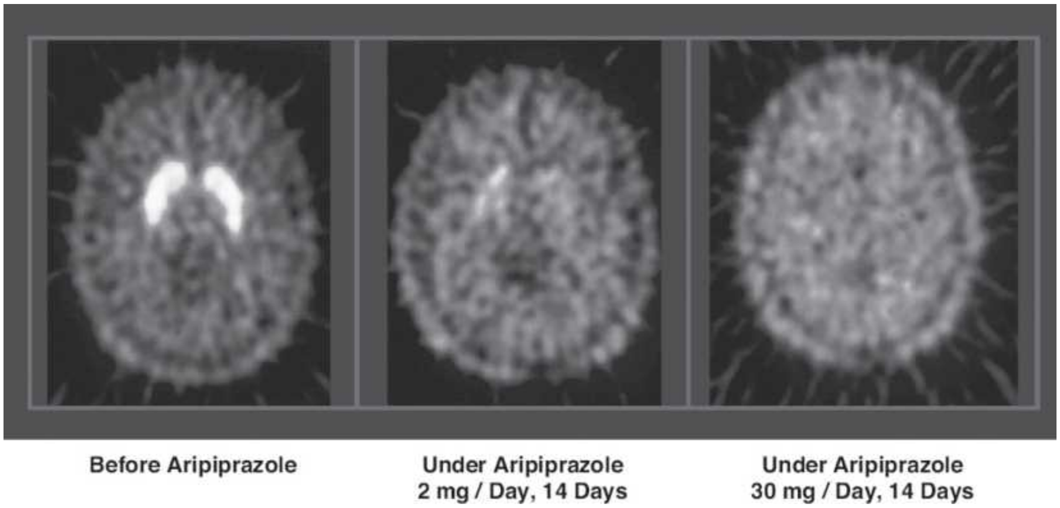
Before Aripiprazole
Under Aripiprazole 2 mg / Day, 14 Days
Under Aripiprazole 30 mg / Day, 14 Days

FIGURE 12.9–14. Dopamine D_2/D_3 PET with occupancy. Representative transversal images from positron emission tomography (PET) with [^{11}C]raclopride, a radiotracer for striatal D_2/D_3 receptors in the brain, of a healthy adult volunteer without medication (**left**) and after treatment for 14 days with either 2 mg (**middle**) or 30-mg (**right**) aripiprazole (Abilify). The baseline study (**left**) demonstrates marked uptake of the radiotracer in the striatum (caudate nucleus and putamen) indicating a high density of striatal D_2/D_3 receptors. After treatment with 2 mg aripiprazole for 14 days, approximately 70% of the D_2/D_3 receptors in the striatum are occupied by aripiprazole so the radiotracer uptake is markedly reduced (**middle**). After treatment with 30-mg aripiprazole for 14 days, almost all of the D_2/D_3 receptors in the striatum are occupied by aripiprazole so the radiotracer uptake is almost completely blocked (**right**). These results confirm that aripiprazole occupies the D_2/D_3 receptors in the striata of healthy adults in a dose-dependent manner, and that clinical doses of the drug lead to almost complete saturation of striatal D_2/D_3 receptors. (Modified from Yokoi F, Gründer G, Biziere K, et al. Dopamine D2 and D3 receptor occupancy in normal humans treated with the antipsychotic drug aripiprazole (OPC 14597): A study using positron emission tomography and [11C]raclopride. *Neuropsychopharmacology.* 2002;27:248, with permission from Springer Nature. doi:10.1016/S0893-133X(02)00304-4.)

A recent study by Chand and colleagues speaks directly to this type of weakness using newly developed data-driven ML methods in large international consortium-based schizophrenia and control samples. This study was different from the previous subtyping, where schizophrenia patients are typically subtyped using only the schizophrenia patients, which risks that the subtypes are confounded by the staging of disease patterns, age, sex, and imaging protocols based on the similarity metrics. However, in Chand and colleagues, schizophrenia patients are subtyped based on their relative deviations in diverse directions from controls, where the potential confounds such as age, sex, race, and imaging protocols are nullified by balancing patient and control groups.

In the report by Chand and colleagues, the first subtype was characterized by widespread reductions of white matter volume and gray matter volume correlating with illness duration and poorer premorbid functioning. In contrast, the second subtype was associated with overall normal brain structure except for enlarged striatal and internal capsule volumes. The prevalence of the first subtype of schizophrenia was ∼63%, and the second subtype of schizophrenia was ∼37%. Thus if those patients are grouped together (without subtyping) and compared with controls, the imaging patterns will be similar to that of the higher prevalence group (∼63%, the first subtype), and the patterns of the lower prevalence group are confounded in the current standard case-control comparisons. These findings have been recently replicated using a different newly developed multi-scale ML methodology. These neuroanatomical subtype signatures of schizophrenia have also been replicated in an independent dataset, and their expressions are quantified in psychosis-spectrum/typically developing youths and adults. Schizophrenia subtype signature expressions are associated differentially with psychosis symptoms, cognition, and genetic risk, potentially reflecting underlying neurobiologic vulnerability. Data-driven ML approaches of these studies can be applied to PET imaging with different radioligands and/or by integrating *complementary* information of sMRI and PET to better parse out the underlying disease heterogeneity in schizophrenia. Imaging measures are highly dimensional, but relating them to clinical, cognitive, molecular, or other risk factors is often challenging and requires an informed data summarization-based framework. The ML approach by Chand and colleagues places each subject in a 2D space formed by the two schizophrenia types and thus helps simplify the relationships among these different measures, which are interpretable and clinically valuable. Future PET studies in schizophrenia will benefit by incorporating such newly developed data-driven ML/AI methodologies. One of the potential challenges is that current PET studies in schizophrenia have smaller sample sizes. The molecular imaging communities can thus encourage data sharing to achieve larger samples needed to apply and validate ML models. The optimistic goal is to extract critical information on the heterogeneity of schizophrenia and contribute to precision medicine of diagnosis and treatment of schizophrenia.

MOLECULAR IMAGING IN CLINICAL PRACTICE

Molecular Imaging in the Treatment of Schizophrenia

Molecular imaging may become a tool to evaluate treatment effectiveness for people with schizophrenia. For example, traditional antipsychotic agents typically produce maximal therapeutic effects with minimal adverse effects when 75% to 80% of the dopamine D_2 receptors are occupied. Similarly, receptor occupancy can be demonstrated by means of PET for dopamine receptors with atypical antipsychotic agents (Fig. 12.9–14). Thus, administration of a molecular imaging procedure to an unimproved person with schizophrenia before and after typical therapeutic doses of a dopamine D_2 receptor blocking drug may demonstrate markedly low occupancy of dopamine D_2 receptors, thus explaining why the person does not benefit from a typical dose of a drug. Alternatively, if the occupancy is not low, the imaging study can help to identify groups of patients who might benefit from medications other than dopamine antagonists. Similarly, molecular imaging of dopamine and serotonin transporters and receptors before and after the administration of novel therapeutic agents may provide evidence for the typical effects of innovative therapeutic agents for schizophrenia. Thus, molecular imaging of neurotransmitter systems in people with schizophrenia may become tools to monitor putative therapeutic interventions. Additionally, molecular imaging of neurotransmitter systems in people with schizophrenia may provide evidence for or against future therapeutic interventions for a particular individual with schizophrenia.

FUTURE DIRECTIONS

The future of molecular imaging in schizophrenia is closely linked to the progress in the discovery and understanding of the molecular mechanisms related to the etiology and pathophysiology of the disease. The role of the glutamatergic system has become increasingly clear over the past decade. Both NMDA and non-NMDA glutamate receptors have been implicated in the pathophysiology of positive, negative, and cognitive symptoms of schizophrenia. A hypofunction of cortical glutamate has been postulated to interact with subcortical dopamine to cause increased release of presynaptic dopamine and elevated numbers of postsynaptic dopamine receptors. There is a need to develop ligands for glutamatergic receptors like metabotropic glutamatergic receptors type 2/3 (mglu2/3), whose agonist has recently been utilized for the treatment of schizophrenia. Likewise, there exists a tremendous future role for ligands for (1) the glycine and serine modulatory sites of the NMDA receptor and (2) the enzyme racemase that regulates the production of d-serine in the brain. Furthermore, the study of nicotinic receptors in vivo in patients with schizophrenia has also been limited by the availability of probes (especially the α7 subtype) for these receptors that likely play a significant role in the mechanisms of schizophrenia. New probes are also needed to image those neuronal structures (i.e., cytoskeleton proteins) involved in neuronal migration and synaptic formation, and that form the basis of the neurodevelopment hypothesis of schizophrenia.

FURTHER READINGS

Abi-Dargham A, Rodenhiser J, Printz D, et al. Increased baseline occupancy of D2 receptors by dopamine in schizophrenia. *Proc Natl Acad Sci U S A*. 2000;97(14):8104–8109.

Ashok AH, Myers J, Reis Marques T, Rabiner EA, Howes OD. Reduced mu opioid receptor availability in schizophrenia revealed with [11C]-carfentanil positron emission tomographic Imaging. *Nat Commun*. 2019;10.

Borgan F, Laurikainen H, Veronese M, et al.; METSY Group. In vivo availability of cannabinoid 1 receptor levels in patients with first-episode psychosis. *JAMA Psychiatry*. 2019;76(10):1074–1084.

Chand GB, Dwyer DB, Erus G, et al. Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. *Brain*. 2020;143(3):1027–1038.

Chand GB, Singhal P, Dwyer DB, et al. Schizophrenia imaging signatures and their associations with cognition, psychopathology, and genetics in the general population. *Am J Psychiatry*. 2022;179(9):650–660.

Erritzoe D, Ashok AH, Searle GE, et al. Serotonin release measured in the human brain: a PET study with [(11)C]CIMBI-36 and d-amphetamine challenge. *Neuropsychopharmacology*. 2020;45(5):804–810.

Farde L, Hall H, Ehrin E, Sedvall G. Quantitative analysis of D2 dopamine receptor binding in the living human brain by PET. *Science*. 1986;231(4735):258–261.

Farde L, Wiesel FA, Halldin C, Sedvall G. Central D2-dopamine receptor occupancy in schizophrenic patients treated with antipsychotic drugs. *Arch Gen Psychiatry*. 1988;45(1):71–76.

Gjedde A, Wong DF. Quantification of neuroreceptors in living human brain v: endogenous neurotransmitter inhibition of haloperidol binding in psychosis. *J Cereb Blood Flow Metab*. 2001;21(8):982–994.

Gjedde A, Wong DF, Rosa-Neto P, Cumming P. Mapping neuroreceptors at work: on the definition and interpretation of binding potentials after 20 years of progress. *Int Rev Neurobiol*. 2005;63:1–20.

Grace AA. Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience*. 1991;41(1):1–24.

Grunder G, Carlsson A, Wong DF. Mechanism of new antipsychotic medications: occupancy is not just antagonism. *Arch Gen Psychiatry*. 2003;60(10):974–977.

Grunder G, Landvogt C, Vernaleken I, et al. The striatal and extrastriatal D2/D3 receptor-binding profile of clozapine in patients with schizophrenia. *Neuropsychopharmacology*. 2006;31(5):1027–1035.

Hirvonen J, van Erp TG, Huttunen J, et al. Increased caudate dopamine D2 receptor availability as a genetic marker for schizophrenia. *Arch Gen Psychiatry*. 2005;62(4):371–378.

Jauhar S, Veronese M, Nour MM, et al. Determinants of treatment response in first-episode psychosis: an (18)F-DOPA PET study. *Mol Psychiatry*. 2019;24(10):1502–1512.

Kapur S, Zipursky RB, Remington G. Clinical and theoretical implications of 5-HT2 and D2 receptor occupancy of clozapine, risperidone, and olanzapine in schizophrenia. *Am J Psychiatry*. 1999;156(2):286–293.

Kenk M, Selvanathan T, Rao N, et al. Imaging neuroinflammation in gray and white matter in schizophrenia: an in-vivo PET study with [18F]-FEPPA. *Schizophr Bull*. 2015;41(1):85–93.

Khodaii J, Araj-Khodaei M, Vafaee MS, Wong DF, Gjedde A. Relative strengths of three linearizations of receptor availability: saturation, inhibition, and occupancy plots. *J Nucl Med*. 2022;63(2):294–301.

Kim JH, Marton J, Ametamey SM, Cumming P. A review of molecular imaging of glutamate receptors. *Molecules*. 2020;25(20):4749.

Lafuente A, Bernardo M, Mas S, et al. Dopamine transporter (DAT) genotype (VNTR) and phenotype in extrapyramidal symptoms induced by antipsychotics. *Schizophr Res*. 2007;90(1–3):115–122.

Laruelle M, Abi-Dargham A, van Dyck CH, et al. Single photon emission computerized tomography imaging of amphetamine-induced dopamine release in drug-free schizophrenic subjects. *Proc Natl Acad Sci U S A*. 1996;93(17):9235–9240.

Marques TR, Ashok AH, Pillinger T, et al. Neuroinflammation in schizophrenia: meta-analysis of in vivo microglial imaging studies. *Psychol Med*. 2019;49(13):2186–2196.

Phan JA, Landau AM, Jakobsen S, Wong DF, Gjedde A. Radioligand binding analysis of alpha 2 adrenoceptors with [(11)C]yohimbine in brain in vivo: extended inhibition plot correction for plasma protein binding. *Sci Rep*. 2017;7(1):15979.

Phan JA, Wong DF, Chang NHS, Kumakura Y, Bauer WR, Gjedde A. Transient equilibrium determination of dopamine D2/D3 receptor densities and affinities in brain. *Front Nucl Med*. 2022;2:1030387.

Ranganathan M, Cortes-Briones J, Radhakrishnan R, et al. Reduced brain cannabinoid receptor availability in schizophrenia. *Biol Psychiatry*. 2016;79(12):997–1005.

Reith J, Benkelfat C, Sherwin A, et al. Elevated dopa decarboxylase activity in living brain of patients with psychosis. *Proc Natl Acad Sci U S A*. 1994;91(24):11651–11654.

Sakayori T, Tateno A, Arakawa R, Kim WC, Okubo Y. Evaluation of dopamine D3 receptor occupancy by blonanserin using [(11)C]-(+)-PHNO in schizophrenia patients. *Psychopharmacology (Berl)*. 2021;238(5):1343–1350.

Sigvard AK, Nielsen MØ, Gjedde A, et al. Dopaminergic activity in antipsychotic-naïve patients assessed with positron emission tomography before and after partial dopamine D2 receptor agonist treatment: association with psychotic symptoms and treatment response. *Biol Psychiatry*. 2022;91(2):236–245.

Volk DW, Eggan SM, Horti AG, Wong DF, Lewis DA. Reciprocal alterations in cortical cannabinoid receptor 1 binding relative to protein immunoreactivity and transcript levels in schizophrenia. *Schizophr Res*. 2014;159(1):124–129.

Wagner HN Jr, Burns HD, Dannals RF, et al. Imaging dopamine receptors in the human brain by positron tomography. *Science*. 1983;221(4617):1264–1266.

Wong DF, Brasic JR, Singer HS, et al. Mechanisms of dopaminergic and serotonergic neurotransmission in Tourette syndrome: clues from an in vivo neurochemistry study with PET. *Neuropsychopharmacology*. 2008;33(6):1239–1251.

Wong DF, Kuwabara H, Horti AG, et al. Brain PET imaging of alpha7-nAChR with [^{18}F]ASEM: reproducibility, occupancy, receptor density, and changes in schizophrenia. *Int J Neuropsychopharmacol*. 2018;21(7):656–667.

Wong DF, Kuwabara H, Horti AG, et al. Quantification of cerebral cannabinoid receptors subtype 1 (CB1) in healthy subjects and schizophrenia by the novel PET radioligand [11C]OMAR. *Neuroimage*. 2010;52(4):1505–1513.

Wong DF, Pearlson GD, Tune LE, et al. Quantification of neuroreceptors in the living human brain: IV. Effect of aging and elevations of D2-like receptors in schizophrenia and bipolar illness. *J Cereb Blood Flow Metab*. 1997;17(3):331–342.

Wong DF, Wagner HN Jr, Tune LE, et al. Positron emission tomography reveals elevated D2 dopamine receptors in drug-naive schizophrenics. *Science*. 1986;234(4783):1558–1563.

12.10 Neurocognition in Schizophrenia

Philip D. Harvey, Ph.D.

NEUROCOGNITIVE IMPAIRMENT IN SCHIZOPHRENIA

Neurocognition as a Core Feature of Schizophrenia

Neurocognitive impairment in schizophrenia is clinically relevant and profound. Patients with schizophrenia perform one to two standard deviations below healthy controls on various neurocognitive tests. The severity of this impairment is greatest in the domains of memory, attention, working memory, problem-solving, processing speed, and social cognition. These deficits are present prior to the initiation of antipsychotic treatment and are not caused by psychotic symptoms in those patients who are able to complete testing, which includes the overwhelming majority. Many of the various neurocognitive deficits in schizophrenia have been shown to be associated with functional outcomes such as difficulty with community functioning, difficulty with instrumental and problem-solving skills, reduced success in psychosocial rehabilitation programs, and the inability to maintain successful employment. Neurocognitive deficits are better able to explain important functional outcomes such as work performance and independent living than positive or negative symptoms.

The emphasis on neurocognitive deficits in schizophrenia goes beyond their severity and relation to functional outcomes. Neurocognitive deficits are present to an extent in some patients with schizophrenia prior to the onset of psychosis, are prominent during the prodrome, and are correlated with measurable brain dysfunction more than any other aspect of the illness. Perhaps most importantly, neurocognition is increasingly considered a primary target for treatment.

Despite the importance of neurocognitive impairment in the biology, function, and treatment of schizophrenia, and despite the emphasis placed on neurocognitive impairment by Emil Kraepelin and Eugen Bleuler over 100 years ago, it still is not included as a criterion for diagnosis in the Fifth Edition of the *Diagnostic and Statistical Manual For Mental Disorders* or the eleventh edition of *International Statistical Classification of Diseases and Related Health Problems* (ICD-11) criteria. It is noteworthy, however, that this newest edition of the DSM has extensive discussions of cognition and its relation to disability. These discussions include the recommendation that assessment of cognition and everyday functioning should be a routine part of the assessment of schizophrenia, and the assessment scales section of the DSM-5 (Section 3) includes an assessment of the severity and types of cognitive deficits. Thus, recognition of the importance of cognition in schizophrenia is now truly “official” and viewed as centrally important to the assessment of the condition.

Profile and Magnitude of Neurocognitive Impairment Associated with Schizophrenia

The level of severity of impairment in performance on neurocognitive tests is strong evidence of the importance of neurocognitive deficits in schizophrenia. Neurocognitive deficits appear to be present in patients with schizophrenia early in the course of the illness. A number of studies have demonstrated that, on various neurocognitive test domains, unmedicated patients with either first-episode or chronic schizophrenia perform close to two standard deviations below healthy controls (Fig. 12.10–1). Further, those cases which are apparently in a prodromal state and covert to psychosis have more severe cognitive deficits at the time of the detection of their state than those cases that do not convert. The most striking aspect of the profile of neurocognitive deficits in patients with schizophrenia is that so few neurocognitive functions remain similar in level of performance compared to healthy controls. In fact, an estimate of the severity of neurocognitive impairment in patients with schizophrenia, based on an examination of the most methodologically sound studies completed, suggests that many important neurocognitive functions are in the impaired range (one to two standard deviations below the normal mean). Meta-analysis of over 300 studies shows a consistent and stable difference between patients with schizophrenia and healthy controls in a wide range of domains of neurocognitive functioning (Fig. 12.10–2).

Neurocognitive tests often assess more than one domain of functioning, and many tests do not fit neatly into a single domain. Thus, descriptions of the profile of neurocognitive deficits in schizophrenia have varied across literature reviews and can be confusing to the nonpsychologist. The domains that are most consistently cited as being severely impaired in schizophrenia are verbal memory, executive functions, attention or vigilance, verbal fluency, and motor speed. Deficits in social cognition also appear to be severe, but there is less normative data available on healthy people to provide information about the severity of impairment. The opinion of a group of experts who served on the Neurocognition Subcommittee for the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Project is that the most important domains of neurocognitive deficit in schizophrenia are working memory, attention/vigilance, verbal learning and memory, visual learning and memory, reasoning and problem-solving, speed of processing, and social cognition. These domains were chosen because they were viewed as substantially impaired, separable from one another, and to have demonstrated relationships to functional outcomes such as independent community living.

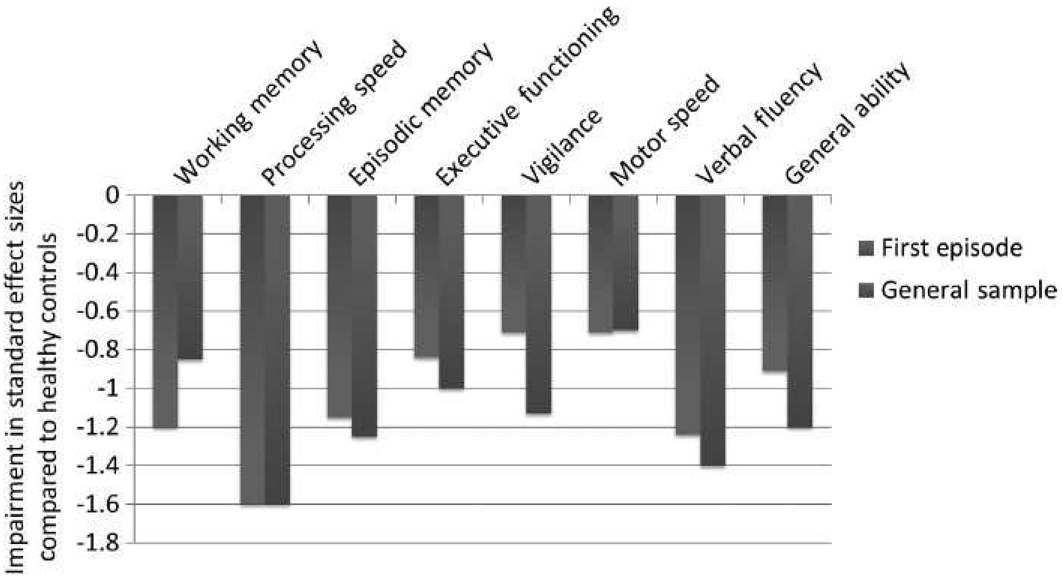

FIGURE 12.10–1. Impairment in major cognitive domains in first-episode patients compared to patients with schizophrenia in general. Based on two systematic meta-analyses: Mesolam-Gately et al. (2009) for first-episode patients and Dickinson et al. (2007) for patients in general. (Data from Mesholam-Gately RI, Guiliano AJ, Goff KP. Neurocognition in first episode schizophrenia: a meta-analytic review. *Neuropsychology*. 2009;23:315–336; Dickinson D, Ramsey ME, Gold JM. Overlooking the obvious: a meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. *Arch Gen Psychiatry.* 2007;64:532–542.)

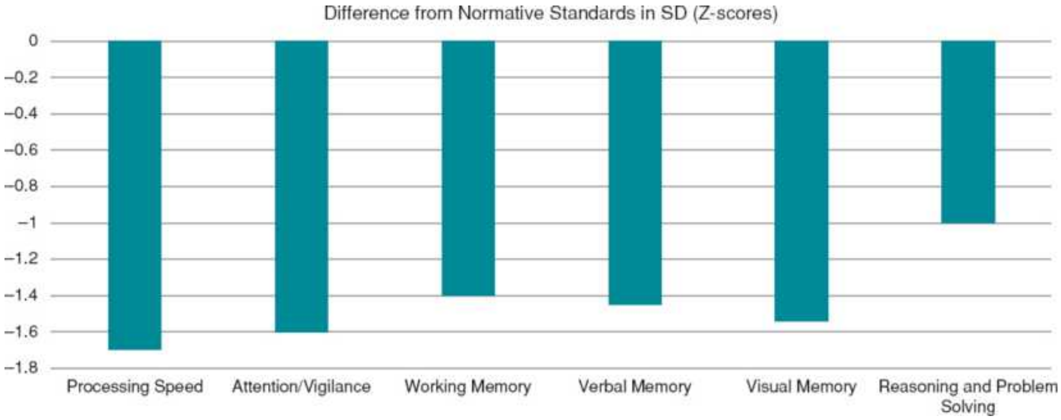

FIGURE 12.10–2. Effect sizes for six different domains on the Measurement and Treatment Research to improve cognition in schizophrenia (MATRICS) batter. Data were collected at screening assessments for participants who were enrolling into cognitive enhancement pharmacologic trials (*n* = 2,616). The processing speed domain has three tests, symbol coding, Trail Making Part A, and Animal naming. Working memory has two tests: the Maryland Letter Number Span and WAIS Spatial Span. All other domains are defined by a single test.

Attention/Vigilance. *Vigilance* refers to the ability to maintain attention over time. A standard vigilance test used in many studies is the continuous performance test (CPT). The “Identical Pairs” version of the CPT requires that patients attend to a series of two- to four-digit numbers presented sequentially on a computer screen at a rate of one per second. Patients respond with a button press on the computer mouse or keyboard each time a number is identical to the previous number. This type of CPT reveals moderately severe vigilance impairments in patients with schizophrenia. Many neurocognitive tests require intact vigilance for good performance, even if the test itself is not a measure of “pure” vigilance.

Impairments in vigilance can result in difficulty following social conversations and an inability to follow important instructions regarding treatment, therapy, or work functions. Simple activities like reading or watching television can become labored or impossible. Reviews of the literature have suggested that vigilance deficits in patients with schizophrenia are related to various aspects of the outcome, including social deficits, community functioning, and skills acquisition.

Verbal Learning and Memory. *Verbal memory* functioning includes, but is not limited to, abilities associated with learning new information, retaining newly learned information over time, and recognizing previously presented material. In general, patients show larger deficits in learning than in retention. The findings for recognition are more equivocal, with most studies suggesting relatively mild deficits, although large deficits have also been reported.

The tests used to measure learning typically involve the ability to learn lists of words or written passages. Verbal list learning tasks usually require the patient to listen to 12 to 16 words, then immediately recall as many of the words as possible. Using the California Verbal Learning Test, healthy controls can recall about 8 of 16 words after the first trial; patients with schizophrenia can recall only about 5. After five consecutive trials of the same word list, most controls can recall at least 13 of the words, while patients with schizophrenia, on average, can recall only nine. Thus, patients with schizophrenia are not only impaired in their ability to immediately recall verbal material but they are also impaired compared to controls in their ability to learn over time. Patients are also impaired in recalling more engaging verbal material, such as stories. Much empirical evidence points to the connection of verbal memory impairment with impairments in everyday activities and occupational deficits in patients with schizophrenia. Success in many modern endeavors requires the ability to learn information that is presented in a verbal context. Interestingly, there is little evidence that people with schizophrenia forget more verbal material than they have learned than healthy controls, with rapid forgetting being much more common in cortical dementias such as Alzheimer disease.

Visual Learning and Memory. Since visual information is not as easily expressed as verbal information, fewer tests sensitive to the deficits of schizophrenia have been developed. Most tests expose subjects to one or more verbal figures and require the subjects to draw them from memory or to indicate which among an array of figures was previously presented.

Studies of the relationship between poor visual memory and functional outcome have yielded mixed findings. Visual memory has been found to correlate modestly with employment status, job tenure, psychosocial rehabilitation success, social functioning, quality of life ratings, and strongly with functional capacity. Other studies have reported no significant correlations. It may be that tests of visual memory are simply less sensitive to an impairment or have other measurement problems.

Reasoning and Problem Solving. Although there are numerous tests of reasoning and problem-solving, also described in some contexts as “executive functioning,” the most well-known and frequently utilized in schizophrenia research is the Wisconsin Card Sorting Test (WCST). In this test, patients are given a deck of cards with various numbers of colored shapes on them and are asked to match their cards to four “key” cards that also have shapes on them that differ by color, form (i.e., shape), and number. The first principle to which the subject needs to learn to sort the cards is *color*. After a patient demonstrates that they have learned that *color* is the correct sorting principle, the principle changes to *form* without warning. Repeated sorting attempts by the previously correct principle are referred to as *perseverations*, and patients with schizophrenia, like patients with frontal lobe damage, often make these errors. In fact, the very poor performance of patients with schizophrenia on the WCST and the reduced activity of the dorsolateral prefrontal cortex during the performance of this test led to a widespread pursuit of the hypothesis of frontal hypoactivation in schizophrenia. It is important to note, however, that the WCST measures a variety of neurocognitive functions, including vigilance as well as visual memory, and is not a “pure” measure of executive functions.

The rules of society and the workplace change regularly, and success in these arenas is often measured by one’s ability to adapt to changes. Patients with schizophrenia who are impaired in measures of reasoning and problem-solving often have difficulty adapting to the rapidly changing world around them.

Speed of Processing. Many neurocognitive tests require subjects to process information rapidly and can be compromised by impairments in processing speed. A standard example of this type of task is the Wechsler Adult Intelligence Scale Digit Symbol Test. Each numeral (1 through 9) is associated with a different simple symbol. Subjects are required to copy as many of the symbols associated with the numerals as possible in 120 seconds. This nonspecific neurocognitive impairment has been found to correlate with a variety of clinically important features of schizophrenia, such as daily life activities, job tenure, and independent living status. It is also sensitive to medication side effects such as somnolence and extrapyramidal symptoms. It has been argued that processing speed tasks such as digit symbol and symbol coding may be the strongest neurocognitive discriminator between schizophrenia patients and healthy controls. Neurocognitive data from 1,331 schizophrenia patients assessed prior to treatment randomization for the National Institute of Mental Health (NIMH) Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) schizophrenia trial suggested that a digit symbol task accounted for more variance in the overall composite score than any other measure. Thus, while seemingly a small component of the profile of neurocognitive deficits in schizophrenia, brief measures of processing speed have a very important role in characterizing relevant aspects of neurocognitive impairment.

The relevance of these deficits is clearly demonstrated by the effect of reduced processing speed on the ability of some schizophrenia patients to keep in step with the task-oriented jobs that are frequently held by patients with schizophrenia. Psychomotor slowing has the potential to affect many different aspects of everyday life as well as hamper performance on many other cognitive testing measures.

Verbal Fluency. Most neurocognitive assessments in treatment studies of schizophrenia have included measures of verbal fluency as a separate domain of functioning. Most tests of verbal fluency measure either phonologic fluency (also referred to as letter fluency) or semantic fluency. *Phonologic fluency* refers to a patient's ability to produce as many words as possible beginning with a particular letter (e.g., "F") within, for instance, 60 seconds. *Semantic fluency* refers to the ability to produce words within a particular meaning-based category, such as "animals." Not only do schizophrenia patients produce fewer words than normal controls, but they are more likely than healthy people to produce words that are not outside the required category. Impaired verbal fluency can damage functioning in social and vocational settings by making communication difficult and awkward.

Immediate/Working Memory. *Immediate memory* refers to the ability to maintain a limited amount of information "online" for a brief period of time (usually a few seconds). Repeating a string of digits (digits forward) is an example of immediate memory. The definition of working memory, on the other hand, is more complex and varies across studies. Some investigators consider working memory to be synonymous with immediate memory, whereas others describe that it should require some manipulation of the information being held online. For example, repeating a series of digits in the reverse order than they were presented (digits backward) requires maintenance of the initial string and an active manipulation because the information needs to be both held online and then subsequently reordered. Patients typically show deficits on both tasks. Visual working memory tasks involve keeping a visuospatial stimulus, such as a dot on a computer screen or a series of visual objects, in mind while working on a related or unrelated task.

Working memory has been described by various authors as a core component of the neurocognitive impairment in schizophrenia and is related to functional outcomes such as employment status and job tenure. Much of the clinical relevance of working memory deficits in schizophrenia comes from strong correlations that working memory measures have with a variety of other neurocognitive domains impaired in schizophrenia, such as attention, planning, memory, and intelligence. In addition, the advanced understanding of the neuroanatomy of working memory functions in human and nonhuman primates adds to our understanding of impairments in working memory in schizophrenia. This neuroanatomical work has suggested that neural circuitry that includes prefrontal cortical regions mediates aspects of working memory functions and that this circuitry may be impaired in schizophrenia.

Social Cognition. Theory-of-mind skills and social and emotional perception have been the general focus of the literature on social cognition in schizophrenia. *Theory of mind* is the ability to infer another's intentions and/or to understand the mental states of others. Individuals with schizophrenia perform poorly on measures of theory-of-mind abilities. The evidence regarding whether impairments in theory-of-mind skills are independent of a general neurocognitive deficit is mixed, although the most recent data suggest minimal overlap. Facial affect recognition and social cue perception are the two general areas into which studies of social perception in schizophrenia can be broken down. Reviews of the literature on facial affect recognition suggest that individuals with schizophrenia have stable deficits on tests of facial affect perception compared with healthy controls and psychiatric controls and that perception of negative emotions and fear may be particularly impaired. Tests of social cue perception use more dynamic stimuli that require multiple sensory modalities, such as watching videotapes of persons interacting. Patients with schizophrenia show consistent impairments in these tasks. In particular, they have more difficulty discerning other individuals' goals and intentions than what they are wearing or saying.

Although a variety of reports have described the strong relationship between general neurocognitive impairments and social deficits in schizophrenia, there is growing evidence that social cognition is more strongly related to social impairments in schizophrenia, even after controlling for performance on neurocognitive tasks. Path analysis has been used to suggest that the relationship between neurocognitive impairment and social deficits is almost entirely mediated by impairments in social cognition.

Prevalence of Neurocognitive Deficits in Schizophrenia

About 15% of patients with schizophrenia (and 85% of the general population) are rated as "unimpaired" by clinical neuropsychological assessment. However, these patients tend to have the highest levels of premorbid functioning and demonstrate neurocognitive functioning that is considerably below what would be expected of them based on their premorbid levels and the educational level of their parents. Data from 150 patients with schizophrenia suggest that up to 98% of patients perform more poorly on neurocognitive tests than would be predicted by their parent's level of education. In addition, comparisons of monozygotic twins discordant for a formal diagnosis of schizophrenia suggest that almost all affected twins perform worse than their unaffected twin on neurocognitive tests, although the unaffected twins are performing more poorly than expectations based on previous educational attainment. Therefore, it is likely that almost all patients with schizophrenia are functioning below the level that would be expected in the absence of the illness.

Experts in neurocognition and schizophrenia have come to a clear consensus that neurocognitive impairment is a core feature of the illness. The profile of deficits is broad, severe, and likely present in most, if not all, patients. Neurocognitive impairment has clear clinical relevance, as neurocognitive impairment interferes with the everyday lives of patients in many important ways, from limiting social relationships to reducing the likelihood of employment. Further, cognitive impairments in schizophrenia are similar across different countries, educational systems, and racial and ethnic groups, suggesting the minimal influence of environmental and social variables.

NATURAL HISTORY OF NEUROCOGNITIVE IMPAIRMENT

Although the development of neurocognitive deficits in schizophrenia patients follows a predictable pattern over time, it, of course, can vary greatly in any individual. In most patients, detectable deficits may be present in childhood, followed by a decline in neurocognitive function that occurs sometime prior to the first episode. However, many individuals with schizophrenia demonstrate no obvious neurocognitive deficits prior to the onset of psychosis. In most patients, the severity of neurocognitive impairments becomes greater around the time that psychosis develops. The long-term stability of neurocognitive impairment over time in the majority of patients is clear. It is also clear that progress of deficits over the illness is generally absent, even in very poor outcome cases, until later life and then is limited to a subset of cases (Fig. 12.10–3).

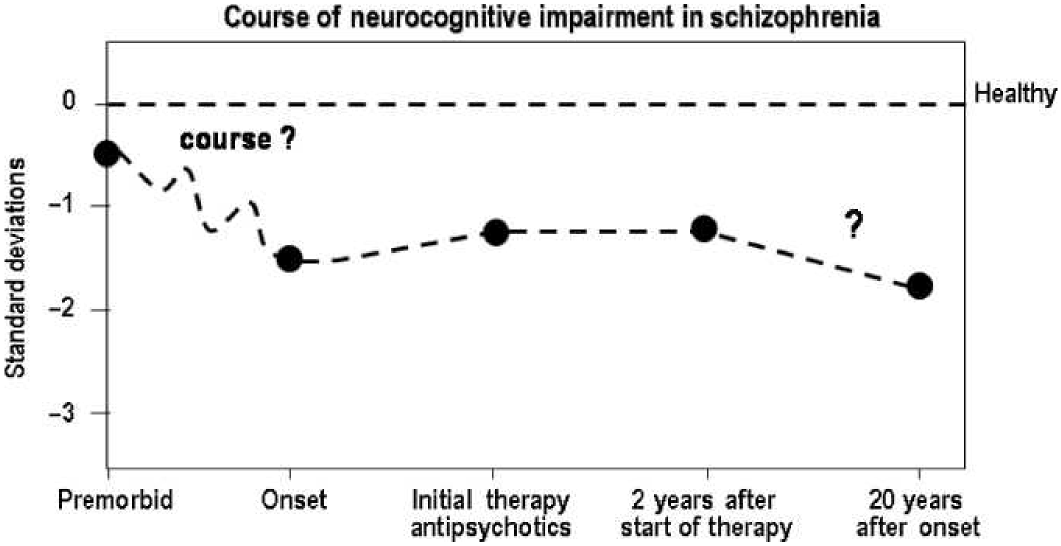

FIGURE 12.10–3. Course of neurocognitive impairment in schizophrenia. The Y axis is the number of standard deviations below the mean of healthy controls throughout the course of the illness. The course of cognitive decline between the premorbid phase and onset of illness is not well known. It is controversial whether patients with chronic schizophrenia have cognitive decline over time.

The Genetics of Neurocognition in Schizophrenia

Neurocognitive impairment in patients with schizophrenia has a clear heritable component. First-degree relatives of individuals with schizophrenia are impaired on a variety of neurocognitive measures, with effect sizes ranging from small to medium. The relationship between the heritability of schizophrenia and related neurocognitive deficits has been estimated, with overall intelligence quotient (IQ) and working memory demonstrating phenotypic correlations with the presence of schizophrenia that appear slightly higher than other aspects of neurocognition. The heritability of neurocognitive impairments in schizophrenia families appears to be moderate t ($h^2 = 0.25$ to 0.40) and larger than their environmental correlations. A number of aspects of neurocognition have been tentatively linked to specific genes or single nucleotide polymorphisms (SNPs). Some of these findings have been replicated, and most genes associated with cognitive deficits to date implicate the functioning of the glutamatergic system. Searches for the genomics of cognition in schizophrenia may lead to clues about the origin of disability, the etiology of schizophrenia, and similarities and differences between schizophrenia and other severe mental illnesses such as bipolar disorder.

Deficits in Children at Risk for Schizophrenia and Follow-Back Studies

Individuals who are genetically vulnerable to schizophrenia have notable neurocognitive impairments, while individuals who are examined with neurocognitive assessments before they develop schizophrenia are found to have impairments in a variety of areas. High-risk studies of children with one or two biologic parents with schizophrenia have suggested that attention deficits can predict which children will develop schizophrenia in the future.

Several studies have utilized the “follow-back” method of identifying adult patients with schizophrenia and then linking these patients with their records of neurocognitive assessments performed while they were children or adolescents. In the UK National Survey of Health and Development Study, data from 5,362 people indicated that children who went on to develop schizophrenia as adults differed significantly from the general population in a variety of neurocognitive and behavioral domains. Low educational test scores in verbal, nonverbal, and mathematics at all ages assessed were significant risk factors. Similar findings were generated from a population-based study that investigated the risk of schizophrenia in a sample of 50,000 18-year-old males conscripted into the Swedish army between 1969 and 1970. In the United States, scores from grades 4, 8, and 11 on the Iowa Tests for 70 children who later developed schizophrenia show no significant differences from controls in grades 4 and 8. However, for children who developed schizophrenia, test scores dropped significantly between grades 8 and 11, corresponding to the onset of puberty. Yet limitations of this study, including a small sample size and a nonrandom sample, restrict the generalizability of these results.

Several studies have utilized a link between the Israeli Draft Board Registry and the National Psychiatric Hospitalization Case Registry. Israeli law requires that all adolescents between the ages of 16 and 17 undergo a preinduction assessment to determine their intellectual, medical, and psychiatric eligibility for military service. This assessment is compulsory and is administered to the entire unselected population of Israeli adolescents. It includes individuals who will be eligible for military service, as well as those who will be excused from service based on medical, psychiatric, or social reasons.

The results suggest that neurocognitive functions are significantly impaired in adolescents who are later hospitalized for schizophrenia. These deficits precede the onset of psychosis in young people destined to develop schizophrenia. Along with social isolation and organizational ability, neurocognitive deficits are significant predictors of which young people will eventually develop a psychotic disorder. One study from this series examined the neurocognitive performance of 44 patients with a first episode of schizophrenia who had previously undergone neurocognitive assessment as part of their registration with the Israeli Draft Board. The stability of the deficits in these patients suggests that most of the neurocognitive impairment seen occurs prior to the first psychotic episode. These follow-back studies consistently demonstrate that neurocognitive impairment precedes the onset of illness in schizophrenia and schizoaffective illness. However, the impairments are small on average and, to date, have not proven useful for identifying cases at higher risk for developing schizophrenia.

Prodrome

Neurocognitive deficits are also found in individuals who are identified as being at “ultrahigh” risk for schizophrenia by virtue of their family history of schizophrenia in a first-degree relative or the manifestation of mild signs and symptoms consistent with the prodromal symptoms of schizophrenia. Preliminary data also suggests that olfactory identification and verbal memory deficits may help to predict which individuals at ultrahigh risk will develop schizophrenia. These findings are particularly interesting given that neurogenesis continues later into adulthood in the brain regions that contribute to these functions. However, at present, the ability of these neurocognitive deficits to predict which young individuals will eventually develop schizophrenia is weak. Since schizophrenia is a rare illness (about 1% of the population) and the percentage of individuals at ultrahigh risk for schizophrenia who eventually develop the illness ranges from 10% to 40%, a predictive factor must be highly specific to be useful in clinical practice; otherwise, an unacceptably large number of young people will be falsely identified as vulnerable to schizophrenia. The results from large-scale studies have shown that in cases with identifiable prodromal symptoms, cognitive impairments are present at the time of the detection of prodrome, and those cases with the greatest impairments are at the highest risk of developing psychosis.

First Episode of Psychosis Studies

Once psychosis develops, neurocognitive deficits are severe. Patients with a first episode of schizophrenia who have never taken antipsychotic medication already exhibit neurocognitive impairment of a magnitude similar to cases with an established illness. Motor functions and language functions may be more mildly impaired prior to treatment with antipsychotic medication. Initial treatment with antipsychotic medications does little to benefit these neurocognitive deficits.

Neurocognitive Changes in Patients Recovering from Acute Exacerbation

Several studies have investigated the longitudinal course of neurocognitive deficits in schizophrenia by repeatedly assessing neurocognition in samples that combine first-episode patients and those who have had an acute exacerbation of their symptoms. Patients recovering from an acute exacerbation of illness do not appear to demonstrate improvements in the severity of their neurocognitive impairment, despite clear improvements in symptoms with treatment.

Longitudinal Change in Neurocognitive Function

There are many age-related changes in neurocognitive functioning in the healthy population, especially in motor speed and memory functions. The magnitude of change in long-term memory and the ability to access and utilize previously learned material is more modest. Reading and vocabulary are tests of “old learning” or “crystallized intelligence.” Baseline functioning can thus be estimated from performance on these tests that show minimal aging-related changes. The prospect of detecting neurocognitive decline with aging is diminished due to floor effects on numerous tests. The types of neurocognitive functions that change with normal aging, such as new learning and recall, attention, and processing speed, are related to the deficits most commonly observed in schizophrenia patients early in life. Thus, aging-related changes may affect those functions that are already severely impaired in adults with schizophrenia.

The notion of neurodegeneration in schizophrenia has been controversial. Evidence has been presented that suggests that schizophrenia is a neurodegenerative process, and some have concluded that schizophrenia is progressive. However, with a few notable exceptions, cross-sectional studies generally have not found evidence of increased neurocognitive impairment in association with duration of illness, and older patients under the age of 65 do not manifest greater impairment than younger patients. Definitive answers to questions regarding the possible progression of neurocognitive impairment in schizophrenia must come from longitudinal studies.

Recent studies have shown a complex relationship between relapse, re-treatment with antipsychotic medications, changes in cortical structure, and changes in cognition. Several studies have shown that there is a loss of white and gray matter and ventricular enlargement seen following relapses due to medication discontinuation after initially successful treatment. One study has shown that these progressive changes in brain volume are correlated with the progression of impairments in executive functioning. The mechanism through which brain volume is reduced following relapse is unclear and why only some studies show a cognitive decline in early-course patients is unknown, but attention to this topic may help to understand both the course of cognitive functioning in relapsing patients and the initial origin of these cognitive deficits.

Neurocognitive Functioning in Older Adult Patients

Neurocognitive impairment in schizophrenia may worsen over time in at least a subgroup of older adult patients with schizophrenia. In older schizophrenic patients with a lifetime of poor functional outcomes, prominent neurocognitive impairments resembling dementia have been reported.

Based on cross-sectional studies, older adult patients with schizophrenia appear to show some decline in neurocognitive function toward the end of life. However, this decline may be restricted to those patients who had an early onset of illness, a long period of psychosis prior to antipsychotic treatment, or a lifetime of poor functioning. Some of the inconsistency of these results may derive from the subject selection processes in these studies. The patients who are available for these studies tend to be residents at tertiary care facilities that house only the patients with the greatest deterioration in functional ability. Finally, the neurocognitive decline in some older adult patients with schizophrenia has been found to be associated with tardive dyskinesia and neurologic dysfunction and could potentially be an outcome of decades of antipsychotic treatments and unenriched environments.

Summary

Several studies have demonstrated that neurocognitive impairment is present in a milder form prior to the onset of psychosis in young people destined to develop schizophrenia. Neurocognitive impairment is severe in patients who have experienced their first psychotic episode, even before antipsychotic treatment is initiated. Although some early-phase patients may demonstrate a slight improvement in neurocognitive impairment with treatment, many patients' neurocognitive abilities do not improve at all. Following this early phase, neurocognitive impairment appears to be remarkably consistent compared to functioning early in life, even in the presence of positive and negative symptom changes, although few longitudinal studies have had follow-up periods of longer than 5 years. The course of neurocognitive impairment in older adult patients is uncertain, but some studies suggest that patients with the most chronic courses of illness may manifest a further, albeit gradual, neurocognitive decline in the latest years of life.

RELATIONSHIP OF NEUROCOGNITIVE IMPAIRMENT TO SCHIZOPHRENIA SYMPTOMS

This section addresses the relationship between neurocognitive impairment and symptoms. If neurocognitive deficits were the result of the symptoms of the illness, then the deficits would disappear simultaneously with symptom reduction. However, this does not occur in patients with schizophrenia. In contrast to some patients with psychotic bipolar illness, whose performance on neurocognitive tests may improve somewhat when their psychosis remits, the performance of the patients with schizophrenia does not change when psychotic symptoms are minimized.

Cross-Sectional Studies of the Relation to Symptoms of Schizophrenia

Neurocognitive deficits are largely separate from both the positive and negative symptoms of the disorder. Cross-sectional relationships between neurocognition and positive symptoms, particularly for hallucinations and delusions, are usually quite weak. Generally, the variance shared by these variables is considerably less than 10%. Relationships tend to be slightly stronger between neurocognition and negative or “deficit” symptoms.

Positive Symptoms

It has been repeatedly reported that neurocognitive ability is not strongly correlated with the severity of psychotic symptoms in patients with schizophrenia. During periods of acute exacerbation, the severity of positive symptoms has been found to be significantly correlated with *better* performance in select domains of neurocognitive function. Although some exceptions exist, such as isolated reports of significant correlations of positive symptoms with working memory, source monitoring, and auditory distractibility, the overall trend is for general neurocognitive impairment not to be correlated with positive symptoms. The correlations between positive symptoms and neurocognitive function in five different domains and an overall composite score collected from 1,331 schizophrenia patients during the baseline period of the NIMH CATIE trial were near zero, with *P* values greater than 0.05 despite the tremendous statistical power in this study. Given that over 91% of patients were able to complete neurocognitive testing in this all-comer trial, it is clear that the absence of a relationship between neurocognition and positive symptoms is generalizable. This low correlation is also found in various patient samples, including first-episode, chronic, and older adult samples, supporting the idea that positive symptoms are not the sole cause of the neurocognitive impairment found in any schizophrenia patient population.

Various factors may mediate the absence of a relationship between neurocognitive deficits and positive symptoms. The low validity and temporal stability of positive symptom assessment may be a crucial factor. The subjective report of a psychotic patient may not reflect the true level of the patient's psychosis, and these symptoms change considerably over time compared to neurocognitive deficits, which are temporally stable. Another factor is that the exclusion of patients who are too psychotic to be tested may weaken any potential correlation between the severity of psychosis and neurocognitive impairment. However, the 91% success rate in CATIE suggests that this limitation may be overstated. Finally, those patients who have more intact neurocognitive abilities may be better able to recall and express their internal state, including detailed delusions and hallucinations. These patients would thus receive higher scores on positive symptom rating scales, resulting in a reduction in any potential correlation between positive symptoms and neurocognitive impairment. There are clear limitations to the assessment of positive symptoms, and the absence of cross-sectional correlations between positive symptoms severity and neurocognitive impairment may be more complex than can be determined in typical research designs.

Negative Symptoms

Neurocognitive dysfunction is correlated with various types of negative symptoms. The greater variance shared between neurocognition and negative symptoms compared with positive symptoms may result from measurement overlap. For instance, tests of verbal fluency and the negative symptom of “poverty of speech” both measure the level at which a patient generates speech. A patient who generates speech at a slow rate will do so during a test of verbal fluency as well as in an interview during which they are being rated for poverty of speech.

Empirical studies suggest that motor functions are strongly correlated with negative symptoms, job success, and outcomes. Deficient motor skills are represented in both the negative symptom and the neurocognitive dysfunction domain since symptoms such as blunted affect and motor retardation are observational measures of motor functioning. Thus, impaired motor skills lie at the core of negative symptoms of schizophrenia in many ways.

The association between reduced motivation and poor performance of patients with schizophrenia on neurocognitive tests is complex. Although monetary reinforcement has been shown to improve performance on effortful neurocognitive tests such as the WCST in some studies, others have not. Increases in pupil size are associated with increased neurocognitive processing demands. Therefore, pupillary response can measure engagement in a task. If the neurocognitive deficits of schizophrenic patients were due to a lack of interest or motivation, their pupillary response would be low throughout the period of neurocognitive assessment. Only during high processing conditions do patients have abnormal pupillary responses, suggesting they put forth a normal amount of effort when being given neurocognitive tests of modest difficulty, yet their decreased processing capacity leads them to be unable to engage in difficult tasks. Other studies in this have similarly found that patients do not increase their effort as task difficulty increases, even in the face of financial rewards that differentially reward more effort. These findings do not suggest that performance on neurocognitive tests is affected in general by poor motivation but that motivation does need to be considered as a factor.

On the contrary, neurocognitive deficits may cause reduced motivation or the appearance of reduced motivation. First, patients who are cognitively capable of engaging in a task or a social interaction may *appear* not to be motivated to succeed in that endeavor. Further, individuals with neurocognitive deficits are often less likely to be motivated to have goals and pursue them. Patients with neurocognitive impairment are likely to be met with failure if they attempt to pursue employment, social, and even recreational avenues that require a suitable neurocognitive skill. Repeated failures are likely to cause discouragement and reduced motivation in anyone, including people with schizophrenia.

In sum, a cross-sectional correlation between negative symptoms and neurocognitive impairment has been consistently reported, but the magnitude of this relationship is modest. The directionality of the relationship between these aspects of schizophrenia is complex and not well understood.

Formal Thought Disorder

Deficits in semantic memory may lie at the heart of the neurocognition–thought disorder relation. This argument is supported by empirical data suggesting that the difference between semantic fluency (the number of names of animals produced in 60 seconds) and phonologic fluency (the number of words beginning with “F” produced in 60 seconds), an indication of the severity of the impairment of the “semantic network” in schizophrenic patients, predicted the severity of their formal thought disorder. Thus, a patient’s inability to have verbal information available (referred to as *semantic priming*) may be the most important neurocognitive factor in formal thought disorder. Other hypotheses regarding the cognitive origins of formal thought disorder include deficits in working memory (i.e., being confused about what has already been said in a conversation) and attentional focus.

Affective Symptoms

Since many patients with schizophrenia also report depressed mood, if not a full depressive disorder, and since depression is associated with some neurocognitive impairment, the role of depressed mood in neurocognitive impairment is important. Very few studies have examined this relationship directly. Furthermore, the distinction between depressive symptoms and negative symptoms is sometimes difficult for raters and clinicians to make, although they appear to be separate dimensions of schizophrenia based on factor analyses. Higher depression scores are significantly correlated with poorer verbal memory, which remains even when psychomotor retardation and processing speed performance are controlled statistically. Thus, depression may influence the association between negative symptoms and neurocognitive impairments, and it may also have a direct deleterious effect on some domains of neurocognitive impairment.

The cross-sectional correlations between neurocognitive impairment and symptoms are weaker than might be expected and vary depending on the symptom domain. The consistency of this finding strongly supports the idea that neurocognitive impairment is not caused by psychosis. The episodic nature of symptoms in schizophrenia in the context of the high temporal stability of neurocognitive impairment contributes to their disassociation.

Longitudinal Course of Neurocognition and Symptoms

The relative independence of neurocognitive impairment and symptoms in cross-sectional studies appears to be supported by longitudinal studies. When patients with schizophrenia are successfully treated with antipsychotic medications, the severity of their symptoms is substantially reduced. However, if neurocognitive function improves with first-generation antipsychotic treatment, the effects are small.

The stability of neurocognitive impairment occurs in the context of frequent variability of the symptoms of schizophrenia, especially the positive symptoms. When patients are tested during a psychotic exacerbation and then again when symptoms have improved considerably, 1 to 2 years later, the neurocognitive function is very stable over time. Over a 1-year interval, neurocognitive deficits and negative symptoms are significantly correlated at each of the assessments, but there is little predictive relationship between these variables over time. If negative symptoms were somehow causing poor performance on neurocognitive tests (or vice-versa), then a longitudinal relationship between these variables would be expected. Second, when the differential relationship between neurocognitive and negative symptoms and adaptive functioning deficits has been examined, neurocognitive impairments have been found to be more strongly correlated with functional deficits, particularly in everyday activities and vocational outcomes, than negative symptoms. Third, in those patients whose functional status worsens over time, changes in neurocognitive impairment predict the level of change, while changes in negative symptoms do not. These data provide yet another line of support for the relative independence of these domains and the importance of neurocognitive impairment in the mediation of functional difficulties.

Another way to investigate the longitudinal relationship between neurocognitive impairment and symptoms is to calculate the correlation between change in neurocognitive impairment and symptom change. This approach addresses the issue of *pseudospecificity*, that is, changes in neurocognition caused by other clinical changes. Several studies have performed such analyses, and though results are mixed, the findings have generally suggested that symptom change does not contribute substantially to neurocognitive change.

The literature consistently demonstrates that neurocognitive deficits are not caused by the symptoms of schizophrenia. Most impressive are the negligible cross-sectional correlations between various domains of neurocognitive function and the positive symptoms of the illness. Although other aspects of schizophrenia, such as negative symptoms, appear to correlate with neurocognitive impairment, there is as much or more evidence supporting the idea that neurocognitive impairment causes negative symptoms, just as there is evidence that negative symptoms cause neurocognitive impairment. Although there is limited evidence of a relationship between changes in neurocognitive impairment and changes in symptoms over time, there is certainly no clear evidence that changes in neurocognitive impairment are caused by symptom changes.

CLINICAL IMPORTANCE OF NEUROCOGNITIVE IMPAIRMENT

Cross-Sectional Correlations with Function

Although the literature on neurocognitive impairment suggests that it is not strongly correlated with symptoms, it has been consistently proven to be related to a variety of other important aspects of the illness, such as social functioning, functional impairments, unemployment, quality of life, relapse prevention, medical status, and economic cost. Neurocognitive impairment also has considerable power to predict functional status years later.

Relationship of Neurocognitive Impairment to Functioning

Functional outcome in schizophrenia is difficult to define and measure. The three types of functional outcomes that most studies of neurocognitive deficits have examined are community (social and occupational) outcome, the ability to solve simulations of interpersonal interactions, and success in psychosocial rehabilitation programs. Strong support for associations between key areas of neurocognition and functional outcome are offered by several reviews of the literature.

The connection between actual community functioning and neurocognition has been weaker than the strong link between laboratory measures of community functioning or skills acquisition and neurocognition. In a 1-year follow-up study of patients with first-episode psychosis, the size of the correlations between the number and quality of actual social relations and various neurocognitive measures were found to be in the 0.25 to 0.35 range.

All of the key neurocognitive domains reviewed above have demonstrated significant relationships to functional outcome, with effect sizes in the medium range. Neurocognitive impairments are also correlated with deficits in the performance of specific skills critical for independent living.

Neurocognitive Impairment and Unemployment

Several studies of neurocognitive performance and vocational functioning have been conducted in rehabilitation settings. Lower ratings of work behavior and performance are related to poorer baseline scores on verbal memory tests and the WCST. Additionally, improvement in patient work performance in a 6-month work rehabilitation program was predicted by better baseline performance on various neurocognitive tests.

Studies examining correlates of competitive employment have consistently emphasized the role of neurocognition. Patients enrolled in school full-time or holding competitive employment show superior performance across measures of working memory, sustained attention, problem-solving, and episodic memory when compared to unemployed patients, with the scores of part-time workers falling between the other two groups. In addition, vocational functioning is significantly associated with performance on the speed of processing tasks such as Trailmaking A and B.

The determinants of employment in patients with schizophrenia are complex and involve a variety of socioeconomic variables that are outside a clinician's control, such as the availability of disability payments. However, beyond these societal factors, neurocognitive performance plays a more important role than clinical symptoms in the ability of patients with schizophrenia to work and in the ability to benefit from work support programs.

Neurocognitive Impairment and Quality of Life

Quality of life is often defined by the quality of a person's social, occupational, and interpersonal aspects of life, all of which are related to neurocognitive functioning. Reductions in quality of life are more strongly associated with neurocognitive deficits than other symptomatic features of the illness. Specifically, the relationship between subjective experience and social functioning has been shown to be mediated by executive functioning. Patients have less realistic impressions of their social functioning if they also have more severe executive deficits. One study of patients with schizophrenia that examined the relationship between coping abilities and neurocognitive dysfunction suggested that more severe executive and memory deficits are related to decreased use of coping mechanisms. Neurocognitive deficits have also been found to predict the long-term quality of life in a few studies. However, persistent symptoms also have an impact on quality of life, and clinicians need to balance the impact of treatment on these two aspects of the disorder separately.

Neurocognitive Impairment and Relapse Prevention

Neurocognitive functions have been shown to be the strongest predictors of a patient's ability to manage medications. Neurocognitive deficits contribute to patterns of medication mismanagement that are associated with poor adherence and risk of relapse. In one study, memory impairment was the strongest predictor of partial compliance. Patients performing poorly in medication management tests have also demonstrated poor global scores on a cognitive assessment battery. Decreased medication compliance in schizophrenic patients has been shown to be related to poor performance on tests of attention and visual memory. Since antipsychotic medication has little impact on neurocognitive deficits, there can be a fair amount of confidence that the direction of causation is running from neurocognitive impairment to decreased medication compliance.

Neurocognitive Impairment and Medical Comorbidity

Neurocognitive impairment is also related to medical comorbidities in schizophrenia. Neurocognitive deficits related to organization (such as reasoning and problem-solving) directly affect patients' ability to seek treatment for medical problems. In older adult patients with schizophrenia, neurocognitive and functional impairments predict the later incidence of new-onset medical problems, while medical problems do not predict the subsequent worsening of neurocognitive and self-care deficits. The inability of patients with schizophrenia to reduce damaging habits such as smoking has been correlated with deficits in memory and attention. Thus, neurocognitive impairments directly affect new-onset medical problems in older patients.

The Costs of Neurocognitive Impairment

Neurocognitive impairment is also a major factor in the direct and indirect costs associated with schizophrenia. Factors leading to the increased cost include loss of ability for self-care, level of inpatient and outpatient care needed, and loss of productivity (for both patient and caretaker). Although neurocognitive impairment by itself is rarely a reason for acute hospitalization of a patient with schizophrenia, it may contribute to the overall length of hospitalization. It may also lead to early admission to nursing homes or long-term care facilities in the case of older adult patients and other patients who are consistently unable to live in the community.

Neurocognitive impairment is the single greatest contributor to the functional disability of schizophrenia. Patients whose symptoms have remitted continue to experience profound neurocognitive deficits that interrupt their ability to function independently, benefit from rehabilitation programs, find and maintain employment, and experience healthy levels of quality of life.

NEUROCOGNITION AND FUNCTIONAL CAPACITY

In the last years, research has focused on the relationships between neurocognitive performance and the ability to perform everyday functional skills. These skills, including financial skills, shopping, travel and transit, and self-maintenance, have been shown to relate to real-world deficits in these areas. Functional capacity assessment instruments, such as the University of California–San Diego Performance-based Skills Assessment (UPSA), require patients to complete a series of challenging and relevant everyday tasks such as making a doctor’s appointment, paying a bill, or planning for a leisure activity. One possible mechanism of the impact of neurocognitive deficits on everyday disability is by inducing impairments in the ability to perform the skilled acts that are required for everyday functioning. Research studies have shown very high and consistent correlations between neurocognitive performance and functional capacity measures, and these same studies have shown that functional capacity measures may be more closely related to real-world functional outcomes than neurocognitive deficits are.

TREATMENT OF NEUROCOGNITIVE IMPAIRMENT

As described in the preceding sections, neurocognitive impairment in patients with schizophrenia has a devastating impact on various aspects of their illness and their lives. The search for treatments that can improve neurocognition has accelerated in recent years. Although antipsychotic medications have a pronounced beneficial impact on psychotic symptoms, there is currently little pharmacologic relief that can be provided for the neurocognitive impairment found in patients with schizophrenia.

Impact of Antipsychotic Treatment on Neurocognition

For decades, patients with schizophrenia were treated with excessively large doses of conventional, first-generation, “typical” antipsychotics, which may have had direct adverse neurocognitive effects. Larger doses of conventional antipsychotics cause lethargy, somnolence, and extrapyramidal symptoms, all of which impair neurocognition. In addition, the anticholinergic medications that are used to control side effects cause additional neurocognitive impairment. Further, lowering dosages of conventional antipsychotic medications has some modest neurocognitive benefits, supporting the idea that the negative impact of first-generation antipsychotics may be dose-dependent. As lower doses did not make neurocognitive impairment less severe than that seen in never-medicated patients, the impact of these older antipsychotic drugs at lower doses is less clear.

Although many studies and meta-analyses have suggested that second-generation, “atypical” antipsychotic treatments provide greater neurocognitive benefit to schizophrenia patients than first-generation antipsychotics, many of these studies have had substantial weaknesses, such as small sample sizes, short duration of treatment, no comparator or a comparator of relatively high doses of first-generation antipsychotic treatment, and inattention to important clinical factors such as the relationship of neurocognitive improvement with symptom change, anticholinergic treatment, and change in extrapyramidal symptoms. The impact of these various design weaknesses on study results is controversial.

The NIMH CATIE schizophrenia trial provided the opportunity to examine the neurocognitive effects of antipsychotics without these methodologic weaknesses. This study assessed the neurocognitive effects of four second-generation medications, olanzapine (Zyprexa), quetiapine (Seroquel), risperidone (Risperdal), and ziprasidone (Geodon), and a representative first-generation medication, perphenazine (Trilafon), using randomized, controlled, double-blind methods in 817 patients. The neurocognitive measures in this trial improved in each treatment group after 2 months of treatment, and none of the groups were statistically distinguishable from the others, including patients receiving the older medication, perphenazine. However, the amount of improvement in these patients was interpreted to be consistent with practice effects and expectation biases, suggesting that the effect of antipsychotic medications on these patients was negligible.

Less is known about the effects of antipsychotic treatments on patients with first-episode psychosis. Two industry-sponsored studies have suggested that olanzapine and risperidone improve neurocognition more than even low doses of haloperidol (Haldol) in first-episode patients, but the lack of improvement in haloperidol patients appears to be associated with other clinical changes and side effects in these patients. The amount of neurocognitive improvement in these studies is also small and may be consistent with practice and expectation effects.

The Search for Pharmacologic Treatments for Neurocognition

Although neurocognition is increasingly viewed as a central feature of the disorder, the inability of antipsychotic treatment to improve neurocognition has led to broad efforts to identify pharmacologic treatments that may provide neurocognitive benefit to patients with schizophrenia. Currently, many large- and medium-sized pharmaceutical companies have compounds in their development pipelines that are being considered as possible treatments for the neurocognitive deficits of schizophrenia. There have been some failures and false starts in this area, and at present, there are no U.S. Food and Drug Administration (FDA) approved treatments for neurocognition in schizophrenia.

Cognitive Remediation

A variety of cognitive remediation programs have been developed to address neurocognitive deficits in schizophrenia, with recent refinements of these programs going beyond older drill and practice exercises and moving toward teaching strategies to improve neurocognitive functioning, compensatory strategies to reduce the effects of persistent neurocognitive impairments, and group exercises aimed at the deployment of the trained skills. In contrast to earlier strategies that converted neurocognitive rehabilitation programs designed for individuals with brain injury following healthy premorbid histories and were found to be too rote and tedious for patients with schizophrenia, these newer strategies are engaging and perceived as entertaining by patients. The potential for the success of these programs to improve neurocognition and functional outcomes in patients with schizophrenia has been supported by a number of recent studies.

A meta-analysis of 26 randomized clinical trials involving a total of 1,151 patients concluded that neurocognitive remediation produces moderate improvements in neurocognitive performance and, when combined with psychiatric rehabilitation, also improves functional outcomes. Unfortunately, the largest of the trials had only 67 patients, and each study was conducted in a single or a small number of settings with highly trained research personnel. Since that time, there have been several larger studies that supported the benefits of cognitive remediation, particularly in concert with psychosocial interventions, for improving both cognitive test performance and improving everyday outcomes. Further, as in the pharmaceutical industry, the tendency of the developers of neurocognitive remediation programs to evaluate the efficacy of their own programs without bias has been questioned. Thus the generalizability of these findings is limited, and the real-world applicability of neurocognitive rehabilitation is unknown.

One relevant controversy in this research area is that pharmacologic interventions to improve neurocognition may only be minimally effective if patients with schizophrenia are not exposed to rich neurocognitive environments or receive targeted psychosocial interventions. A new compound that has the potential to have a beneficial neurocognitive effect by altering the neurochemistry of neurocognition in the human brain may be prevented from demonstrating benefit if patients do not have sufficient daily neurocognitive stimulation, which is frequently the case in many patients with schizophrenia. It is likely that one of the mechanisms of neurocognitive improvement in schizophrenia will be enhanced neuroplasticity, which is often experience-dependent. Thus, neurocognitive enhancement trials in schizophrenia may require neurocognitive remediation in all subjects to test the effectiveness of new compounds adequately.

Table 12.10–1.
Advantages and Disadvantages of Selective Neurocognitive Assessment Strategies

Performance-Based: Neuropsychological Tests		
	Advantages	**Disadvantages**
MCCB	Developed by an expert panel relatively full coverage	Some domains defined by a single test
BACS	Takes 30 minutes Available in a tablet version	Domain-level analysis not possible
rBANS	Abbreviated assessment Co-normed with other measures	Does not assess some critical domains
Performance-Based: Functional Capacity		
UPSA	Highly correlated with cognitive test performance Abbreviated highly valid versions available	Some items are now outdated Only a single Form
VRFCAT	Engaging VR assessment with multiple forms All instructions and responses recorded	The evidence for sensitivity to treatment responsive is not extensive
Interview-Based Functional Capacity		
SCoRS/CAI	Extensive data on convergence with performance-based tests Very low burden	Informants required for valid assessment data

MCCB, MATRICS Consensus Cognitive Battery; BACS, Brief Assessment of Cognition in Schizophrenia; r-BANS, Repeatable Battery for the Assessment of Neuropsychological Status; UPSA, UCSD Performance-Based Skills Assessment; VRFCAT, Virtual reality Functional Capacity Assessment Task; SCoRS, Schizophrenia Cognition Rating Scale; CAI, Cognitive Assessment Interview.

Summary

There are currently no clearly effective treatments for neurocognitive impairment in schizophrenia. Larger doses of first-generation antipsychotics, perhaps especially the higher potency medications like haloperidol, may have a deleterious effect. Despite initial hopes to the contrary, it appears as though even the second-generation antipsychotic medications do not provide more than minimal benefit for neurocognitive impairment. A broad effort is currently underway to find new medications that may serve as “cotreatments” to antipsychotics to benefit neurocognition specifically. A few limited studies have suggested that neurocognitive remediation programs may have modest beneficial effects on neurocognition in patients with schizophrenia, and large-scale efforts are needed.

METHODS OF ASSESSMENT

A large number of tests have been developed to assess neurocognition, and various batteries comprised of these tasks have been employed in clinical practice and research with schizophrenia patients (Table 12.10–1). The most useful method of assessment will depend on the clinical or research questions being addressed, the study settings, and the available resources. A battery comprised of many different tests that may take several hours to complete has the advantage of enhanced reliability of neurocognitive estimates and will allow the assessment of various domains of neurocognition. However, these long test batteries have the disadvantage of a patient, tester, and resource burden. In addition, missing data and attrition rates may increase, which may be higher in most impaired subjects, thus skewing results. If the research or clinical question involves the efficacy of treatment intervention on neurocognition, a composite score from a small battery may be sufficient and allow for a larger number of patients to be assessed while requiring less staff resources. If, however, the question involves relative strengths and weaknesses of various neurocognitive domains, a thorough battery composed of multiple tasks in each domain may be required. If the question is primarily focused on one neurocognitive domain, a brief general battery in combination with multiple tests of the domain of primary interest is likely to be appropriate.

A recent development in neurocognitive assessment for clinical trials is the availability of computerized test batteries that allow direct data transfer to study databases. These methods offer the hope of minimizing rater error and reducing the costs required for human quality assurance. However, many of these methods have not been fully validated, and, therefore, results obtained using them must be evaluated carefully and scientifically. Many of these tests have been made commercially available without nearly as much information on their validity as would be hoped for.

Complex experimental procedures derived from cognitive neuroscience have also been used to characterize detailed aspects of the neurocognitive impairments in schizophrenia. These procedures have the potential to be useful since the underlying neurobiologic circuitry that mediates these functions has been uncovered in animal and human studies. However, care must be taken when inferring an etiologic basis or treatment strategy for schizophrenia based on the performance of patients on these tasks. The outstanding question remains whether interventions that improve the performance of patients on these tasks would have any effect on functionally relevant cognitive ability or on functional outcomes, as correlations with everyday outcomes have sometimes been found to be low. These cognitive neuroscience tasks have been particularly appealing for use in genetic studies as tools to quickly assess an endophenotype that might reflect a specific genotypic vulnerability to schizophrenia.

Finally, neurocognitive deficits in schizophrenia and the impairments in everyday functioning that they mediate have begun to be measured with methods that may more closely track changes in functional outcomes. Interview-based assessments of neurocognition, such as the Scale for Cognition in Schizophrenia (SCoRS) and the Cognitive Assessment Interview (CAI), have been developed for use in clinical practice and clinical trials. These measures involve addressing a series of questions about the severity of neurocognitive impairment and how these impairments reduce functioning in daily life. Their strength is that they can be completed in a short time period (15–30 minutes) and yield a reliable estimate. However, information obtained directly from patients with schizophrenia tends to have reduced reliability and validity, and a caregiver or other informant is usually needed to provide information from an objective perspective. Performance on these tests is highly correlated with levels of functioning in everyday life, and it is believed that improvements in these tasks herald the potential for real-world change.

FURTHER READINGS

Bilder RM, Goldman RS, Robinson D, et al. Neuropsychology of first-episode schizophrenia: initial characterization and clinical correlates. *Am J Psychiatry*. 2000;157(4):549–559.

Bowie CR, McGurk SM, Mausbach BT, Patterson TL, Harvey PD. Combined cognitive remediation and functional skills training for schizophrenia: effects on cognition, functional competence, and real-world behavior. *Am J Psychiatry.* 2012;169:710–718.

Bowie CR, Reichenberg A, Patterson TL, Heaton RK, Harvey PD. Determinants of real-world functional performance in schizophrenia subjects: correlations with cognition, functional capacity, and symptoms. *Am J Psychiatry*. 2006;163(3):418–425.

Buchanan RW, Keefe RSE, Umbricht D, Green MF, Laughren T, Marder SR. The FDA-NIMH-MATRICS guidelines for clinical trial design of cognitive-enhancing drugs: what do we know 5 years later? *Schizophr Bull*. 2011;37:1209–1217.

Greenwood TA, Lazzeroni LC, Murray SS, et al. Analysis of 94 candidate genes and twelve endophenotypes for schizophrenia from the Consortium on the Genetics of Schizophrenia. *Am J Psychiatry.* 2011;168:930–946.

Harvey PD. Pharmacological cognitive enhancement in schizophrenia. *Neuropsychol Rev*. 2009;19:324–335.

Harvey PD, Keefe RSE. Studies of cognitive change in patients with schizophrenia following novel antipsychotic treatment. *Am J Psychiatry*. 2001;158(2):176–184.

Heinrichs RW, Zakzanis KK. Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology*. 1998;12:426–445.

Keefe RSE, Bilder RM, Davis SM, et al. Neurocognitive effects of antipsychotic medications in patients with chronic schizophrenia in the CATIE trial. *Arch Gen Psychiatry*. 2007;64(6):633–647.

Keefe RSE, Buchanan RW, Marder SR, et al. Clinical trials of potential cognitive-enhancing drugs in schizophrenia: what have we learned so far? *Schizophr Bull.* 2013;39:417–435.

Keshavan MS, Vinogradov S, Rumsey J, Sherrill J, Wagner A. Cognitive training in mental disorders: update and future directions. *Am J Psychiatry.* 2014;171:510–522.

Seidman L, Giuliano AJ, Meyer EC, et al. Neuropsychology of the prodrome to psychosis in the NAPLS consortium: relationship to family history and conversion to psychosis. *Arch Gen Psychiatry.* 2010;67:578–588.

Vinogradov S, Fisher M, Holland C, Shelly W, Wolkowitz O, Mellon SH. Is serum brain-derived neurotrophic factor a biomarker for cognitive enhancement in schizophrenia? *Biol Psychiatry.* 2009;66:549–553.

Vinogradov S, Fisher M, Warm H, Holland C, Kirshner MA, Pollock BG. The cognitive cost of anticholinergic burden: decreased response to cognitive training in schizophrenia. *Am J Psychiatry.* 2009;166:1055–1062.

Wykes T, Huddy V, Cellard C, McGurk SR, Czobor P. A meta-analysis of cognitive remediation for schizophrenia: methodology and effect sizes. *Am J Psychiatry*. 2011;168:472–485.

12.11 Phenotypes of Psychosis

BRETT CLEMENTZ, PH.D.

Every case of psychosis defies explanation. Consider the anatomy professor at an esteemed university. He is a small man with small features, Einstein-esque hair, a crackling intellect, and long history of academic success. His eyes are unremarkable, windows to the soul of the common man or genius. Without warning or expectation, the professor develops a florid psychosis, with a disturbance pervasive and profound. A meager improvement over many years becomes life in a state hospital's locked ward, with vanishing hope of recovery or even semblance of an independent existence. Reminiscent of John Nash, the professor still expresses complex thoughts, despite vivid and tormenting visual hallucinations that treatments leave unaltered.

To see the professor, it is necessary to schedule an appointment. Because like scientists at Los Alamos during the race to thermonuclear deliverance, the professor works at the forefront of psychiatry. There is little time. The mission is too important. His laboratory contains a wooden, dark-stained, heavy, and worn library table in the middle of a secure facility. The table has held many stories and sustained many dreams. The professor sits on the long side, facing the door, sometimes rising to pace. The pressure to complete his mission is palpable.

Piles of papers litter the table, scattered in no apparent order. Each page has a freehand drawing of the human eye. Combining exceptional artistic and intellectual talents and deep anatomical knowledge, the professor compulsively draws the retina. Beautiful. Anatomically precise. Each retina contains prisms, refracting machines for morphing the visual world. Simple prisms, elaborate designs, small prisms, large ones, mixed small and large, all within each retina. “The core of psychosis is in the eye,” the professor believes. He carefully considers each design, asking whether the prismatic machine modifies visual input to illuminate the dark side of the moon. To test the theory, he draws refracting lines on each design, from prism to prism, seeking the origin of his hallucinatory world. But each design is discarded in disgust with the realization of its failure. His eyes, these retinae, refuse to reveal their secrets. (A special thank you to Dr. Carol Tamminga for discussing the details of this case. Carol is a colleague, friend, intellectual powerhouse, and matchless role model for young women scientist-practitioners.)

A knowledgeable clinician may say this case is clear. It is there for all to see. The essence of schizophrenia, the professor’s diagnosis, the most devastating of serious psychiatric conditions, is apparent. The description supports the diagnostic conclusion. But does it really? The hallmark of science, including clinical science, is not pursuing support for your most cherished ideas but seeking their refutation. Superficially, the anatomy professor checks enough phenomenologic boxes for a schizophrenia diagnosis: chronic unremitting hallucinations, delusional elaboration of his warped sensory experience, a calamitous plunge from a previously high level of functioning, lack of interest in activity other than search for the cause of his hallucinatory torment, all against the backdrop of a clear sensorium. But there are inconsistencies: poor cognition is often believed to characterize schizophrenia—this was not true of our professor. Our psychiatric forebearers, historical giants like Kraepelin and Bleuler, believed disorders of thought characterize this most devastating of psychiatric conditions—this was not a feature of our professor. Schizophrenia psychosis typically onsets in the late teens to early 20s—but the professor became ill later in life after considerable occupational success.

This case, like every case of psychosis, defies explanation. How did this happen? Does focusing on clinical phenomenology alone, sensational bits of a clinical picture, bend our perceptions? Is there a different path to illuminating the specter of psychosis, what Irv Gottesman called the epigenetic puzzle?

DO CLINICAL EXPERIENTIAL PSYCHOSIS DEFINITIONS CAPTURE BIOLOGIC REALITY?

In 2013, a new Diagnostic and Statistical Manual of Mental Disorders (DSM-5) was published. This manual uses clinical phenomenologic and experiential features alone, not laboratory tests, to subtype psychosis syndromes, including schizophrenia, schizoaffective disorder, and bipolar disorder with psychosis. Operational diagnostic criteria, the backbone of these manuals, facilitate diagnostic *reliability* for bipolar disorder with psychosis and schizophrenia, although they are less successful for schizoaffective variants.

The *validity* of these diagnostic distinctions is debatable. Clinical features do not uniquely differentiate psychosis syndromes. This has been known for decades. A National psychiatric treasure, Dr. Will Carpenter of the Maryland Psychiatric Research Center and Chief Editor of *Schizophrenia Bulletin,* highlights this fact at every opportunity. Poor outcome patients (traditionally considered a hallmark of schizophrenia psychosis) often have prominent affective symptoms and bipolar disorder family histories. Considerable overlap exists across psychosis diagnoses on neurocognitive features, although schizophrenia cases tend to have more severe cognitive dysfunction and negative symptoms. Family, twin, and adoption studies indicate the importance of shared constitutional and genetic factors to risk for both bipolar disorder with psychosis and schizophrenia. DSM psychosis syndromes do not "breed true," a fact problematic for the Kraepelinian dichotomy (which persists in the psychiatric consciousness despite Kraepelin's 1920 deep concerns with its veracity). All antipsychotic drugs have a dopaminergic mechanism of action, are only partially effective, and treat psychosis generally, not specific clinical psychosis diagnoses. Apparently, to quote Dr. Steve Hyman, "DSM was a poor mirror of nature."

One could reasonably wonder whether DSM-like psychosis diagnoses have justified their pertinence. As theoretical constructs (hypotheses to be tested rather than entities to be reified), they are only modestly successful. There are certainly statistical tendencies for schizophrenia and bipolar disorder with psychosis to differ in predictable directions on a range of measures; but striking heterogeneity within and overlap between these diagnoses on a range of molecular, genetic, brain imaging, perceptual, neurocognitive, symptomatic, treatment response, and outcome variables challenge the neurobiologic reality of DSM psychosis definitions. Based on experience within the Psychiatric Genetics Consortium, Patrick Sullivan et al. stated the range of genetic findings "strongly suggest that our [current] diagnostic categories do not define pathophysiologic entities." As a result, it seems contrary to contemporary understanding to organize a presentation of laboratory phenotypes around the 2013 DSM psychosis diagnoses.

Experiences from other branches of medicine (oncology, cardiology, neurology) clarify that clinical features alone are largely insufficient for capturing biologically based disease entities. Would circumstances change for our professor if laboratory tests allowed for patient stratification and treatment selection? The National Academies of Sciences proposed revising the classification of disease across all medicine based on molecular and biomarker data rather than relying solely on clinical phenomenology, a charge not obviously embraced by framers of current psychiatric diagnostic manuals. In fairness, the DSM framers had few (if any) sturdy biomarker, molecular, or neurobiologic hooks on which to hang their diagnostic hats.

After decades of research on the neurobiology of psychosis, why is this the case? A standard and somewhat nihilistic answer is that psychiatry is hopelessly complex, its syndromes heterogeneous by their nature and beyond the reach of simple biologic explanations; they are qualitatively different from maladies handled by other branches of medicine. Perhaps, but in over 40 years since the revolutionary publication of DSM-III, the strategy for psychosis diagnoses has not changed, and its validity has not been seriously tested by DSM framers in relation to any competing alternatives. A different explanation is that continued reliance on DSM-type experiential diagnoses, which are allowed to supersede neurobiologic observations, impedes the scientific development of psychiatry enjoyed by other medical disciplines.

NEUROBIOLOGY OF DSM PSYCHOSIS DIAGNOSES

A challenge for modern psychiatry is to identify distinctive biologic substructures of its experientially defined syndromes. Clarifying even a few unequivocal etiologic and neurobiologic mechanisms for psychosis would herald a new revolution for psychiatry. Knowledge of mechanisms would help carve diseases from our current clinical syndromes (which are not diseases but constellations of clinical features untethered from neurobiology) and promote the development of targeted and successful treatments. This was an organizing goal of the Bipolar-Schizophrenia Network for Intermediate Phenotypes (B-SNIP) project from its inception in 2008. B-SNIP is a consortium of scientists, of which I am a member, with a common goal—to fortify the neurobiologic substance of our experientially defined psychosis syndromes. If that is not possible, develop a different means for organizing neurobiology within psychosis that promotes targeted treatment developments, then test that approach against the available alternatives. Other consortia, including the Consortium on the Genetics of Schizophrenia (COGS) and the North American Prodrome Longitudinal Study (NAPLS) group, have complementary goals.

B-SNIP borrowed elements from both dimensional and categorical approaches (which are complementary rather than competitive) for probing the neurobiologic underpinnings of psychosis. Simplistically, a dimensional model is a continuum of severity, with this continuum shading from disease into subsyndromal and normal variants on the construct(s) of interest (e.g., frontal lobe gray matter volume, cortical thickness, neuropsychological measures of cognitive control, oddities of perceptual experience). Categorical models capture qualitative, not just continuous quantitative (dimensional), differences between health and disease and between diseases. In their purest form, such diseases are qualitatively distinct from normalcy. It is unlikely any psychiatric brain disease is a categorical entity like Huntington disease (an autosomal dominant mutation of the Huntingtin gene). Even the characters of Mendel peas (e.g., yellow versus green color) were quantitative in nature (although “yellow” and “green” can be considered categorical descriptors). The nuances of approaches to nosology in psychiatry have been discussed with considerable sophistication in other publications.

B-SNIP was influenced by the proposals of Craddock, Owen, and colleagues on the reformulation of the psychoses. Initially, B-SNIP had two main organizing hypotheses for understanding psychosis neurobiology and its relationship to clinical manifestations. First, there is evidence for continua of severity on both neurobiologic and symptom features of relevance to psychosis, with schizoaffective variants bordered by psychotic bipolar disorder and schizophrenia on the extremes. For instance, one such continuum could be called “neuro-cognitive capacity,” with humans, not just persons with psychosis, distributed from deficit to surplus. It is evident our professor is closer to the surplus pole despite debilitating psychosis symptoms. A schizophrenia diagnosis alone does not capture this fact. Second, there are etiologic factors independent of a neurocognitive continuum. These factors, superimposed on that continuum, are responsible for cases tending toward a particular clinical manifestation. The cognitive capacity background upon which these etiologic “hits” occur could cause the sometimes confusing and seemingly arbitrary nature of clinical manifestations observed across psychosis cases. Which hit (or hits) did the professor possess that caused a debilitating psychosis despite substantial cognitive capacity? Identification of such hits must matter for treatment development and planning, but this information does not presently exist. Craddock and Owen (2010) formulated a comprehensive model that captures the essence of these propositions (with levels of analysis ranging from genes to intermediate neurobiologic and behavioral features to clinical phenotypic psychosis manifestations).

There are three keys to the success of B-SNIP aims. First, it is crucial to collect psychosis samples of sufficient size (across DSM diagnosis, what is called transdiagnostic assessment) to capture representative clinical and neurobiologic heterogeneity. Small samples account for inconsistent results in the psychosis literature and almost certainly fail to capture sufficient hard-to-classify cases. This was an avoidable obstacle. Larger samples also support more sophisticated analyses.

Second, psychosis-relevant features must be quantified at a level removed from clinical symptomatology. These measures need to encompass a broad range of the neurocognitive and physiologic correlates of psychosis. B-SNIP laboratory measures were selected given their promise for identifying neurobiologic deviations and understanding genetic risk for psychosis (so-called traditional endophenotypes, a term applied by Irv Gottesman to biomarkers that capture inherited predisposition to psychosis). Starting with these measures had a number of useful features: (i) in the late 2000s, they were the most researched endophenotypes for psychosis; (ii) findings would be easily replicable and portable across diverse clinical and laboratory settings; and (iii) starting at an intermediate level of neurobiologic deviation (between the more basic molecular and structural/functional anatomical and more superficial psychosocial functioning and clinical features) promoted shortening the causal distance between genes and clinical phenes. Associations can be tested in both downward and upward directions in this extensive and complex causal chain, strengthening the theoretical network.

Third, multiple statistically integrated biomarkers, rather than biomarkers considered individually, are required to capture the essence of psychosis disease factors. Epigenetics, etiologic heterogeneity, pleiotropy, and variable expressivity are facts of nature—they influence phenotypic manifestations. Any single biomarker almost certainly fails to precisely define a psychiatric disease. Just as one would never use a single item from the WAIS Information subscale to estimate full-scale IQ, it is overly optimistic to assume a single biomarker adequately assesses any brain function (e.g., context updating in working memory, behavioral inhibition, behavioral response speed). Single laboratory measures are often used in experimental psychopathology and psychiatry, but this weakness in psychosis biomarkers research is easily avoided. B-SNIP statistically integrated multiple individual measures to assess brain functions of interest, enhancing signal-to-noise (e.g., antisaccade response latency and prosaccade latency both index speed of visual orienting; the auditory ERP N100 response during the paired-stimuli task assessing a highly similar neural response as the N100 response during the oddball task).

HYPOTHESIS TESTING AND REPLICATION IN PSYCHOLOGY AND PSYCHIATRY

In psychology and psychiatry, the typical approach is to do statistical tests of differences between groups (it is considered a success to reject the null hypothesis that groups are the same on some measure of interest). This is a logically flawed method for building a knowledge base. Platt (1964), in his “Strong Inference” paper, addressed this point, which was echoed in psychosis research by Dr. Carpenter and colleagues in their 1993 “Strong Inference: Theory Testing and the Neuroanatomy of Schizophrenia.” Typically, we have no other approach for probing the veridicality of our syndromes (our theories are just not as strong as in the “hard sciences,” although we obviously hope to get there). In B-SNIP, we constructed a scenario whereby rejecting the null hypothesis falsified our theory (our theory being that a neurobiologic approach to psychosis subtyping, what we call Biotypes, captures reproducible structure within idiopathic psychosis—structures that may translate to etiologic knowledge and neurobiologically-informed treatments). This scenario also involved extensive replication and cross-validation, both of which are critical to scientific advancement, both of which have been challenges for psychology and psychiatry. The next two sections cover this ground for DSM psychosis diagnoses and biotypes through a summary of Clementz et al. “Psychosis Biotypes: Replication and Validation from the B-SNIP Consortium.”

B-SNIP RESULTS USING DSM PSYCHOSIS DIAGNOSES AS THE GOLD STANDARD

The focus of B-SNIP is idiopathic psychosis, a clinical phenotype with reliable and quantifiable symptom manifestations. Its presence also suggests the possibility of a severe psychiatric condition in need of treatment. Although fuzzy at its borders, a clinical psychosis diagnosis can be made with acceptable inter-rater reliability. There was no doubt our professor crossed this boundary. The initial B-SNIP goal was to identify neurobiologic features via our dense biomarker panel that were (i) shared between and (ii) distinctive to the major DSM psychosis diagnosis categories (schizophrenia, schizoaffective disorder, bipolar disorder with psychosis). B-SNIP1 (the first iteration of our project) recruited 711 clinically stable persons with psychosis and a community comparison sample of n = 274. Our replication sample (part of the second iteration of our project, called B-SNIP2) included similar numbers of psychosis cases (n = 717) and healthy persons (n = 198). We performed dense phenotyping across multiple levels of analysis (symptoms, social functioning, cognitive, psychophysiologic, brain imaging, and molecular measures) for both samples.

A traditional approach questions whether symptoms distinguish the three psychosis groups. Linear combinations of multiple psychosis-related clinical features were created that best differentiated the three DSM psychosis diagnoses. A complete description of these findings is in our “Testing Psychosis Phenotypes from B-SNIP for Clinical Application,” published in 2020. One practical outcome of this probe is to help clinicians know what clinical features are most closely associated with a particular psychosis diagnosis.

The main outcome is illustrated in Figure 12.11–1. Two dimensions differentiated the DSM groups. One dimension we label ‘CONCORD’ (CONation-COgnition-Reality Distortions) because its primary symptom correlates are hallucinatory behavior, delusions, unusual thought content, lack of judgment and insight, and blunted affect. Schizophrenia cases scored highest and bipolar disorder with psychosis cases scored lowest on this dimension. The other dimension we label ‘affective instability’ because its symptom correlates are depression, sadness, concentration difficulties, anxiety, suicidal thoughts, and lassitude. The affective instability dimension differentiated schizoaffective disorder from schizophrenia and bipolar disorder with psychosis.

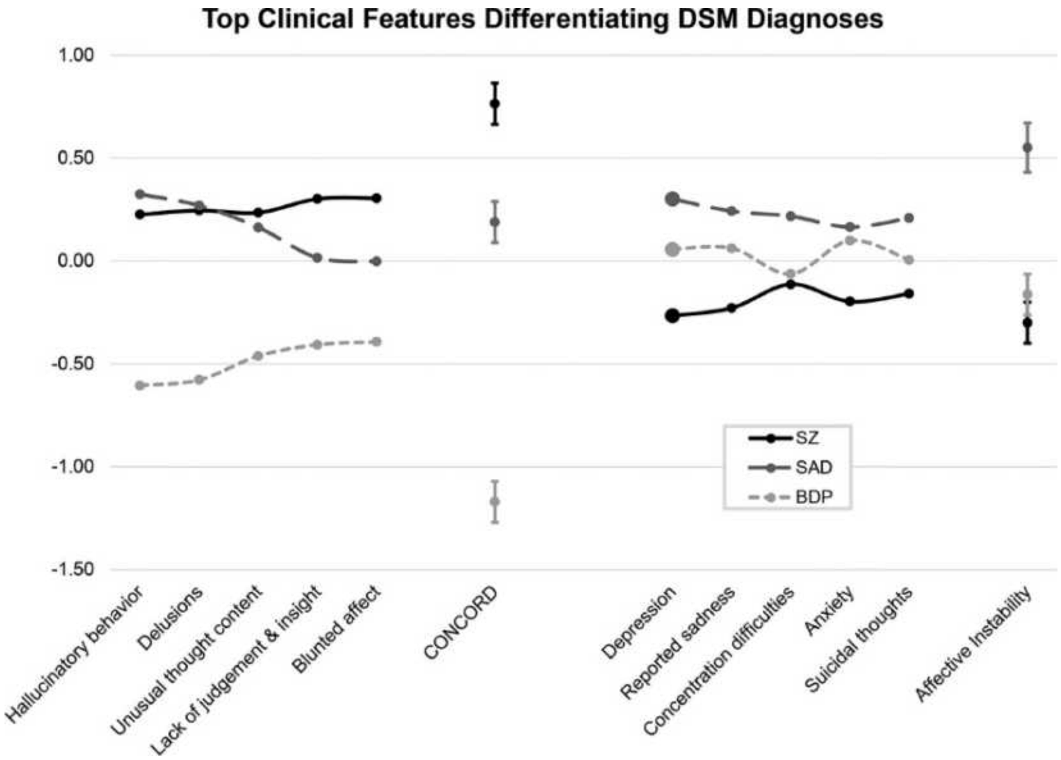

FIGURE 12.11–1. Clinical features differentiating DSM syndromes. Standardized scores (and standard errors) are on the y-axis across the two main clinical dimensions of CONCORD (conation—cognition—reality distortion) and affective instability for DSM syndromes. The top 5 clinical features for each dimension are also provided along with their standard scores. SZ, schizophrenia; SAD, schizoaffective disorder; BDP, bipolar disorder with psychosis. (Modified from Clementz BA, Trotti RL, Pearlson GD, Keshavan MS, et al. Testing psychosis phenotypes from Bipolar–Schizophrenia Network for Intermediate Phenotypes for clinical application: Biotype characteristics and targets. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2020;5(8):808–818. doi:10.1016/j.bpsc.2020.03.011.)

The symptoms differentiating DSM psychosis diagnoses are interesting for three main reasons. First, schizophrenia is not characterized by any special set of clinical features but rather by extreme scores on psychosis symptoms, which describes our professor—this dimension captures a severity continuum. Second, schizoaffective disorder, not bipolar disorder with psychosis, is characterized by affective instability, indicating that this often-maligned diagnosis has a distinguishing presentation. Third, bipolar disorder with psychosis is associated with low scores on CONCORD. Affective instability played no role in bipolar disorder's differentiation from the other DSM groups.

The symptomatic differentiations of the DSM psychosis groups are revealing. The next step investigated the neurobiologic uniqueness of the DSM psychosis diagnoses using our biomarker battery. An overview of the main biomarkers follows.

Brief Assessment of Cognition in Schizophrenia (BACS)

BACS covers multiple cognitive domains (verbal memory, processing speed, reasoning and problem solving, working memory), although a global neuropsychological functioning composite score integrating these domains yields the best measure of psychosis-related cognitive deviation. Dysfunction of cognition is a frequent feature of psychosis.

Pro- and Antisaccade Tasks

Prosaccade tasks require subjects to look as quickly and accurately as possible at a peripheral visual cue—they assess the speed of visual orienting (psychosis cases show slowed or speeded responses). Antisaccade tasks require subjects to look as quickly and accurately as possible at the mirror image location (same amplitude, opposite direction) of a peripheral visual cue, and they assess inhibitory control under perceptual conflict because the visual stimulus and required response location are incompatible (psychosis cases have delayed reaction times and increased error rates). Saccade latency (time from peripheral cue onset to saccade onset) and percentage of error responses (glances to the incorrect spatial location) are recorded for each condition.

Stop Signal Task (SST)

The SST measures the efficiency and adequacy of cognitive control when response activation and generation to a single stimulus location are placed in conflict (there are delayed responses and increased errors in psychosis). In a typical stop signal paradigm, trials begin with a central fixation cross, after which subjects are shown a ‘Go’ cue to the left or right. On a minority of trials, a “Stop” signal is presented at central fixation (with variable delays of, e.g., 50 ms to 250 ms after Go cue onset). Participants are instructed to respond to the Go cue as quickly as possible unless they encounter the Stop signal. Reaction time and proportion of Stop Signal errors are quantified.

Auditory Paired-Stimuli and Oddball-Evoked Brain Responses

Evoked brain responses to repetitive auditory stimuli (paired-stimuli task, used to measure what is sometimes called “P50 gating”) and predetermined auditory targets randomly interspersed with nontarget (or standard) auditory events (oddball task) are deviant in psychosis. These paradigms assess the neural dynamics of preparation for and recovery from auditory sensory activations, neural responses to stimulus salience, and neural differentiation of relevant from irrelevant auditory stimuli. For the typical paired stimuli task, subjects passively listen to multiple (120 to 150) binaural auditory click pairs with a short (500 ms) interval between the two clicks and a long (8 to 10 seconds) interval between the click pairs. This task ostensibly probes the ability to detect stimulus relevance (or reject stimulus irrelevance), a function deficient in some psychosis cases. For a typical auditory oddball task, subjects listen to hundreds (500 to 600) of auditory stimuli occurring every 1 to 1.5 seconds. Most of the stimuli are the same (1000-Hz tones) and are called “standards,” with 15% to 25% of the stimuli randomly interspersed with the standards, being different (1500-Hz tones)—these are called “targets.” Targets elicit a different brain response from the standards, with the most prominent being the so-called P300 (a positive voltage waveform in the event-related brain potential (ERP) most prominent over central parietal lobe sensors occurring 300 to 400 ms after stimulus onset). P300 measures updating of information in working (or available) memory, a function that is deficient in multiple psychiatric conditions.

Electroencephalography data can be scored in multiple ways. B-SNIP maximizes the use of available information, so auditory-evoked brain responses are scored in both the temporal and frequency domains. Rather than using a single or few sensors (the most typical approach in the psychosis literature), we integrate brain activity over 64 sensors distributed around the whole head to estimate patterns of neural response associated with auditory stimulus processing. Scoring data in the temporal domain yields information on the strength of brain signals in voltage units. These signals are typically used to measure ERPs at specific time points (50, 100, 200, or 300 ms after stimulus onset). Scoring in the frequency domain yields information on brain signals in particular frequency ranges (delta, theta, alpha, beta, gamma); this information also can be quantified as a function of time (e.g., gamma activity in the first 100 ms after stimulus onset). Extracting both voltage and frequency provides more information on complex brain responses than can be obtained using either quantification approach alone.

After data integration within the paradigm set (BACS, stop signal, saccade tasks, electroencephalography paradigms), the outcome is nine biomarker composite variables (what we call bio-factors) used to efficiently compare psychosis and healthy groups. Those bio-factors by DSM diagnosis are shown in Figure 12.11–2A. Bio-factors have familiar names (e.g., P300 ERP), but it is important to highlight that every one of them is a composite of multiple variables (e.g., the P300 ERP has primarily six temporal and frequency domain measures determining the score).

There are three noteworthy features of Figure 12.11–2A. First, as with most measures in the published literature, psychosis cases have deficient values (values below the mean of healthy persons) on most biofactors. Second, biofactor patterns between B-SNIP1 and replication samples for DSM diagnoses are remarkably similar; biofactor structure is reproducible using DSM diagnoses. Third, only the BACS significantly differentiates all four groups [schizophrenia < schizoaffective < bipolar < healthy], capturing a severity continuum. N100 ERP, P300 ERP, antisaccade, and paired-stimuli S2 ERP show similar, but less clear, severity continua. The P200 ERP is the only biofactor with modestly greater deviation for bipolar disorder with psychosis, and the SST biofactor identifies a DSM psychosis diagnoses with modest effectiveness.

BACS is the only individual biofactor to differentiate all three DSM psychosis subgroups. But using one measure, even a composite measure, is incompatible with the B-SNIP thesis that differentiating psychosis subgroups requires multiple measures. We tried to improve DSM group separations by using all biofactors simultaneously. The outcome of this multivariate analysis is a single severity dimension. The BACS, paired stimuli S2 ERP, antisaccade, and P300 ERP biofactors, in combination, increase the group separation from schizophrenia to bipolar disorder (the most extreme groups) over the BACS alone from 0.74 to 0.89 standard deviation units. Figure 12.11–3A illustrates that the distinction between DSM and healthy groups on this multivariate dimension is minimal. It is hard to imagine even this amount of biomarker data improving treatment selection and outcome predictions for our professor when relying on DSM diagnosis as the gold standard.

A prominent outcome of these B-SNIP biofactor analyses is, like the symptom data, a continuum of psychosis severity. Overlap on neurobiologic features is ubiquitous in the psychosis literature, in this case on measures ranging from basic sensory processing through sensorimotor performance, inhibitory control, target detection, memory functions, and general executive control abilities. DSM diagnosis groups show the same level of overlap on other structural and brain function measures. B-SNIP provides the largest and most comprehensive evaluation of biomarker-endophenotypes by DSM diagnostic status to date. The main outcomes do not support the thesis that DSM psychosis diagnoses capture neurobiologically distinct brain diseases. Differences in degree, rather than differences in kind, of deviation, described the overwhelming majority of DSM diagnosis variance. Perhaps B-SNIP used the wrong laboratory measures to address this issue (although measures were selected for their demonstrated success)—this possibility can be tested (see below).

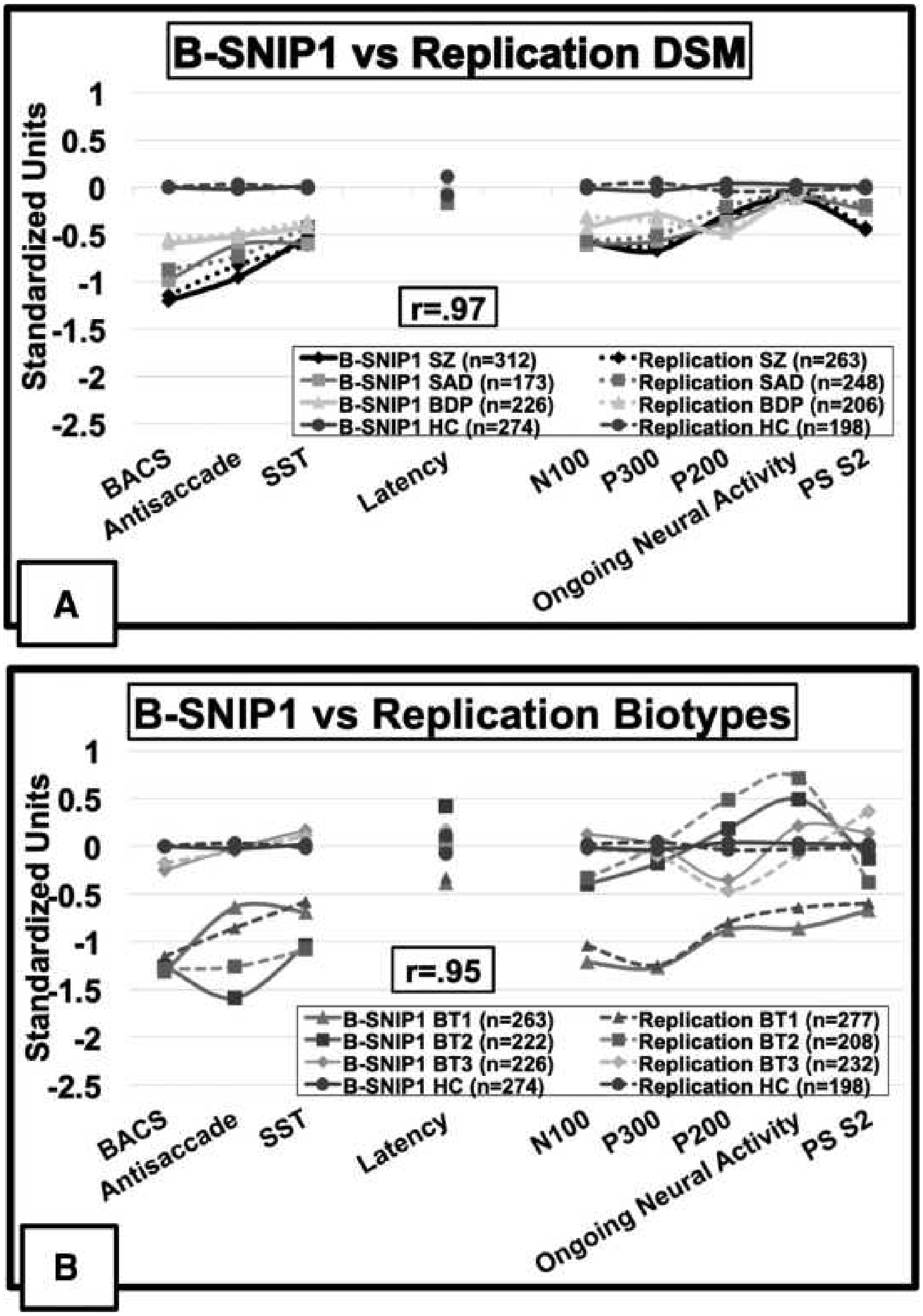

FIGURE 12.11–2. Biofactor patterns by DSM diagnosis (**A**) and Biotype (**B**). Biofactor means standardized scores (y-axis) are displayed by group and project. Solid lines indicate B-SNIP1 and dotted lines indicate replication sample. Boxed r-values are the correlations between the B-SNIP1 and replication samples. Lines link conceptually related variables (cognition variables on the left and EEG variables on the right). Legends show the number of observations by group. Lines with circle symbols show the healthy data, with their y-axis values adjusted so the average value of B-SNIP1 and the replication sample is zero; the healthy values are the same in both plots. The psychosis groups are displayed as a function of their difference from the healthy means. In relation to the healthy subjects, values below zero indicate deficient values and those above zero indicate exuberant values. (Reprinted from Clementz BA, Parker

DA, Trotti RL, McDowell JE, et al. Psychosis biotypes: replication and validation from the B-SNIP consortium. *Schizophrenia Bulletin.* 2022;48(1):56–68, by permission of Oxford University Press. doi: 10.1093/schbul/sbab090.)

PSYCHOSIS BIOTYPES

Many considerations, including the above analyses, led B-SNIP scientists to question the neurobiologic uniqueness of DSM psychosis diagnoses. B-SNIP's alternative hypothesis is that disease entities within idiopathic psychosis lie orthogonal to DSM definitions, a risky proposition in 2015 when we floated the idea. Work in other medical disciplines indicates that clinical diagnosis alone fails to capture etiologically meaningful diseases. Differentiations using biomarker data reshuffle cases with remarkably similar clinical presentations into distinct disease entities. There is no reason to believe idiopathic psychosis should be any different. As Fischer and Carpenter proposed in 2009: "Psychotic experience is to the diagnosis of mental illness as fever is to the diagnosis of infection—important but nondecisive in the differential diagnosis." If true, to probe the neurobiology of psychosis, we should sidestep DSM clinical diagnoses (except for the presence of psychosis). One alternative is to develop a classification of psychoses informed by clinical neuroscience.

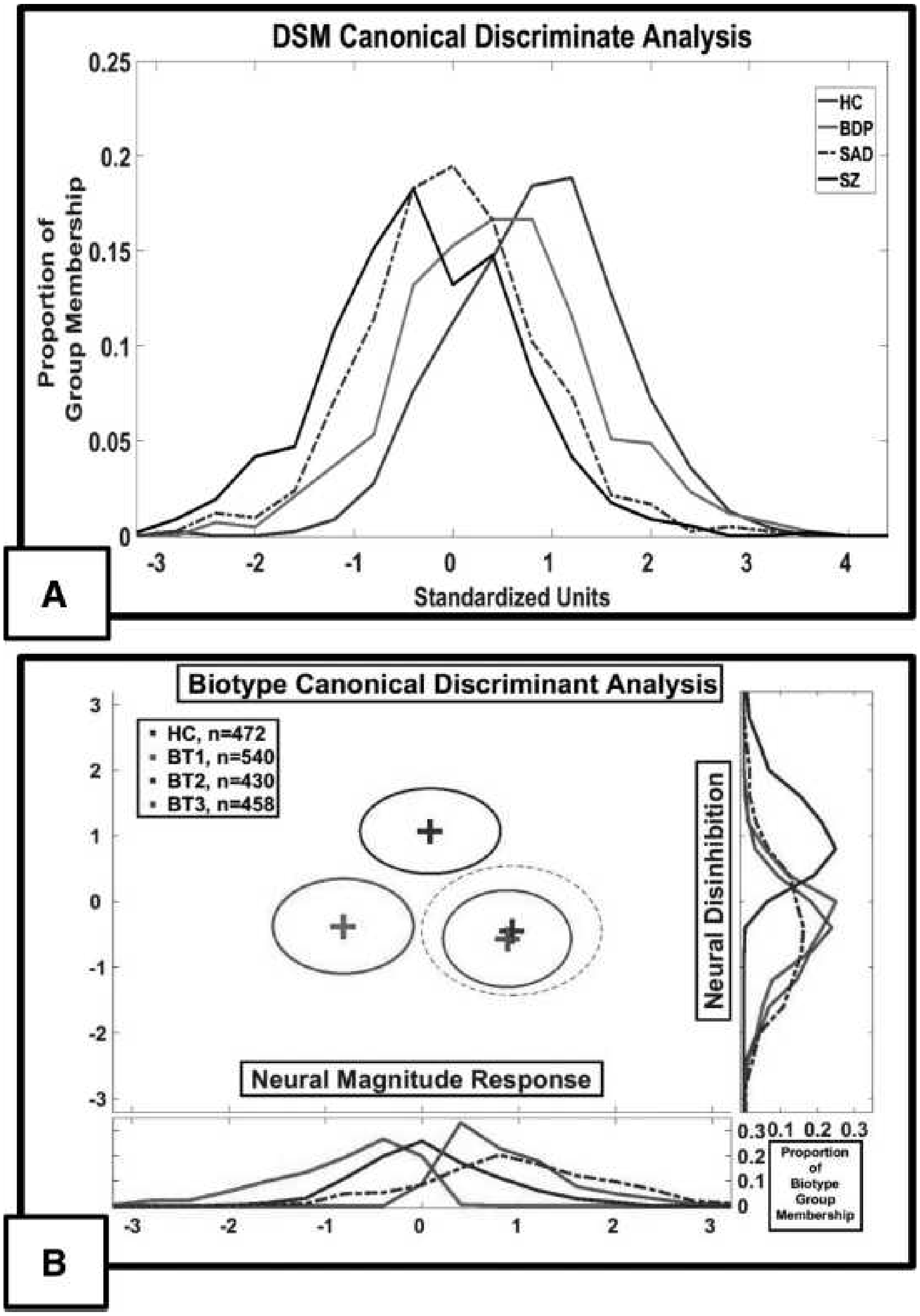

FIGURE 12.11–3. Group Separations by DSM diagnoses and Biotypes. B-SNIP1 and replication samples were combined for these analyses given their high degree of similarity. The outcome of multivariate discriminant analyses, using all biofactors to create functions that optimally separate groups are shown for DSM **(A)** and Biotypes **(B).** There was one significant function that differentiated the DSM diagnosis psychosis groups. Plots show proportion of cases within each group (y-axis) as a function of their standardized discriminant function scores (x-axis). There were two discriminant functions that differentiated the psychosis Biotype groups. The first function on the x-axis captured "Neural Response Magnitude," and the second function on the y-axis captured "Neural Disinhibition." Frequency polygons show the proportion of cases by group at the bottom (Neural Response Magnitude) and right (Neural Disinhibition) of the

central plot that shows the centroids and standard deviation ellipses by group. (See eBook for color images.) (Reprinted from Clementz BA, Parker DA, Trotti RL, McDowell JE, et al. Psychosis biotypes: replication and validation from the B-SNIP consortium. *Schizophrenia Bulletin.* 2022;48(1):56–68, by permission of Oxford University Press. doi: 10.1093/schbul/sbab090.)

B-SNIP leveraged variance in the biomarker panel to capture neurobiologically distinctive psychosis subgroups. The question is how to meaningfully parse psychosis-relevant biomarker variance. We chose a numerical taxonomy method called *k*-means clustering of the nine biofactors. This method forms groups of cases that are defined by their centroids (or center of mass in the nine biofactor space), with all cases in the group being as close to their respective centroid as possible. Only psychosis cases are used in these analyses because the biofactors differentiate psychosis and healthy persons, so the problem is meaningfully parsing variance within psychosis alone.

We quantitatively established that three subgroups best captured neurobiologic variance among the biofactors. Based on *k*-means analyses conducted separately for B-SNIP1 and replication samples, subjects are assigned to one of the three subgroups given their biofactor scores. We call the subgroups psychosis *Biotypes* (biologically distinctive types). Bio-factors by Biotype are shown in Figure 12.11–2B, which is analogous to Figure 12.11–2A, using the same subjects, but now grouped by Biotype.

There are three noteworthy features of Figure 12.11–2B. First, all biofactors differentiate the Biotypes, perhaps not surprising because numerical taxonomy formed maximally homogeneous and distinct subgroups. But the outcome does show that every one of the biofactors is useful, contrary to the DSM analyses. This is a curious outcome because our measures were selected based on results in the literature using DSM diagnoses. This is a lesson in the importance of using samples large and representative versus drawing conclusions from those small and unrepresentative. Second, biofactor patterns between B-SNIP1 and replication samples for psychosis Biotypes are highly similar, indicating the differentiating patterns are reproducible and robust. Third, specific patterns characterize psychosis Biotypes. Biotype-1 and Biotype-2 have a general cognitive deficit as measured by BACS. Biotype-1's most notable defining features are deficient neural activity revealed by low N100 and P300 ERP biofactors indicating difficulty detecting stimulus salience, low ongoing EEG neural activity, low amplitude responses to repeated auditory stimuli as indexed by the PS S2 ERP, combined with slowed response latencies to saccade stimuli. These cases capture the essence of cognitive and physiologic dysfunctions thought to characterize chronic unremitting psychosis. Biotype-2 defining features are greater deviation on cognitive tasks that require inhibitory control in sensorimotor performance as indexed by antisaccade and SST and excessive ongoing EEG neural activity not clearly locked to stimulus registration and the related accentuation of P200 ERP. Excess neural activity, including problems inhibiting behavioral outputs, describes these cases. While Biotype-3 has modest deviations in general cognition and P200 ERP and modestly larger responses to repeated auditory stimuli (PS S2 ERP), they are like healthy subjects on biofactors. Not surprisingly, these cases have the best social functioning and the mildest symptom presentations. Other analyses from our group show that these results cross-validate, have promising construct validity, and predict outcomes on other laboratory measures not used in Biotypes' creation. This is a promising start.

Just like for DSM diagnoses, we quantified the extent of Biotype group separations using all biofactors simultaneously. The outcome of this multivariate analysis is two dimensions, shown in Figure 12.11–3B. The first dimension, shown on the x-axis, we call "neural magnitude response" because of its substantial contributions from P300 ERP, N100 ERP, ongoing neural activity, and BACS biofactors. This function captures the deficient neural activation of Biotype-1. The second dimension, shown on the y-axis, we call "neural disinhibition" because of its substantial contributions from ongoing neural activity, antisaccade, P200 ERP, and SST biofactors. This function captures the neural overactivity and motor inhibitory failures typifying Biotype-2. These two functions provide respectable standard deviation separations between psychosis groups in Cartesian space (BT1 to BT2 = 2.07; BT1 to BT3 = 2.43; BT2 to BT3 = 2.64; BT3 and healthy did not significantly differ).

Biotype physiologies are remarkably different. It seems improbable the same treatments will address the core features of each subgroup. We can be certain that Biotype-1 physiology does not characterize our professor. Minimal laboratory measurement (less than an hour) would clarify his location in biofactor space. If we studied and diagnosed idiopathic psychosis in this fashion, could we improve our professor's treatment selection and outcome prediction?

As we did for DSM diagnoses, we created linear combinations of the psychosis-related clinical symptoms that best-differentiated Biotypes. The main outcome of this analysis is illustrated in Figure 12.11–4. Two dimensions differentiate the Biotypes. One dimension we call "cognitive deficit psychosis" because its primary clinical correlates are occupational and social deficit scales from the Birchwood, conceptual disorganization, and blunted affect. Biotype-1 cases, with the most profound neurocognitive and physiologic deficits, are the highest (worst scores) on this dimension, and Biotype-3 cases, most normal across biofactors, are the lowest. The other dimension we call 'neuro-affective dysregulation' because its primary clinical correlates are stereotyped thinking, mannerisms and posturing, anxiety, uncooperativeness, and disorientation. This dimension differentiates Biotype-2 from the other two Biotypes.

The symptom correlates of the Biotypes are interesting by virtue of their differences from DSM categories. This is not a surprise. We know DSM diagnoses and Biotypes are not redundant—Biotype-1 is mostly schizophrenia (40.7%) and schizoaffective disorder (42.3%), Biotype-2 is more schizophrenia (36.9%) and less bipolar disorder (21.8%), and Biotype-3 is disproportionately bipolar disorder (48.6%), but all DSM diagnoses are represented in all Biotypes. Unlike DSM diagnoses, so-called positive psychosis symptoms like hallucinations and delusions are not the main features differentiating Biotypes. This fact is consistent with the position that such clinical features are unhelpful for differentiating neurobiologically distinctive psychosis subtypes. Hallucinations and delusions may be akin to the symptoms of fever, headache, or fatigue and are not useful for differential diagnosis. Perhaps we do overvalue sensational bits of the clinical picture and are distracted from the features that best capture etiologic differentiation within idiopathic psychosis. Worth some thought anyway, especially because the clinical characteristics differentiating the Biotypes provide an intriguing relationship to neurobiology (i.e., Biotype-2, the cases with accentuated neural activity and motor inhibition problems also have differentiating symptom features concordant with neural dysregulation). It may be possible to identify symptomatic features characteristic of neurobiologic types working from the *bottom up* rather than from the *top down* (the approach taken by framers of the DSM). A related approach shows promise in diseases relevant to psychiatry, for instance, 22q11 deletion syndrome.

PSYCHOSIS SYNDROMES: DIFFERENCES IN KIND?

The answer to this question for DSM psychosis diagnoses has been known for some time. Using clinical symptoms alone fails to capture biologic reality. What if psychosis symptoms are like fever, an indicator of pathology, but not the disease in need of treatment? We must provide palliative care for troubling positive symptoms of psychosis, just like we provide antipyretics for pyrexia. The professor deserves such treatment. But perhaps in psychiatry, we confuse a symptom indicating a disease process for the disease process itself. Prominent psychiatric scientists, among them Drs. Carpenter, Casey, Hyman, Insel, and Jablensky question psychiatry's reliance on the clinical descriptive approach. They argue that such persistence impedes the discovery of etiologic mechanisms undergirding discrete disease entities and the development of successful treatments such discoveries may afford.

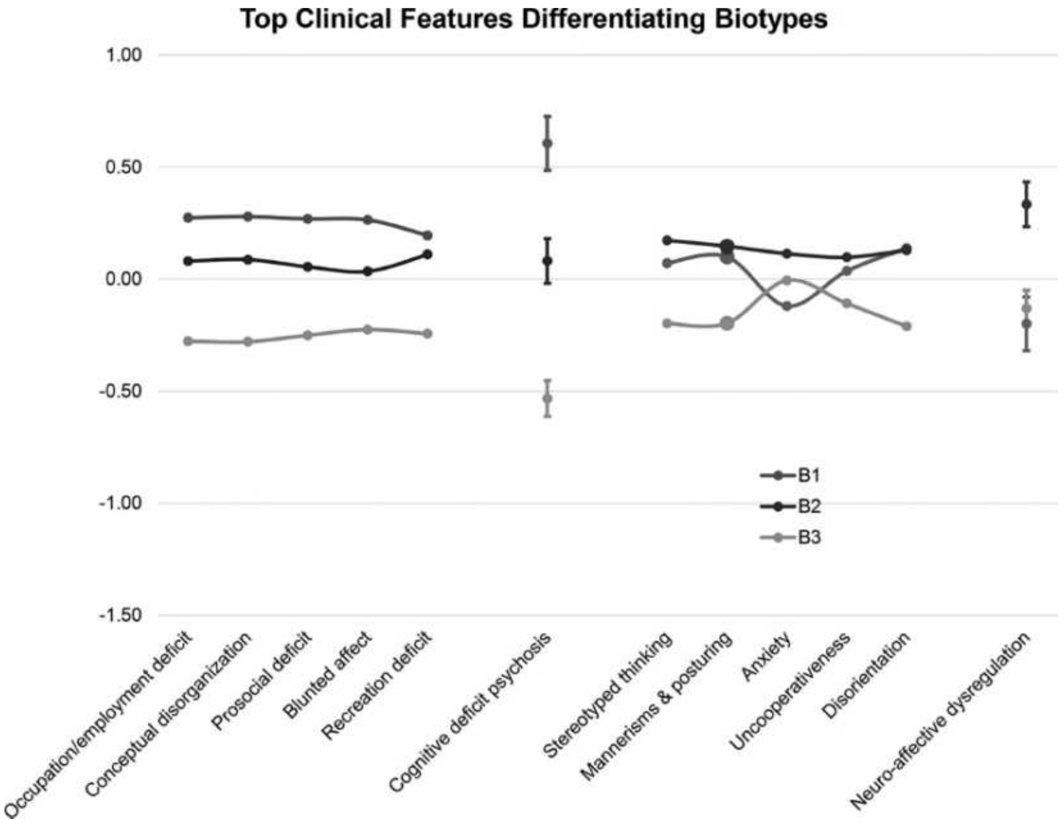

FIGURE 12.11–4. Clinical features differentiating psychosis biotypes. Standardized scores (and standard errors) are on the y-axis across the two main clinical dimensions of cognitive deficit psychosis and neuroaffective dysregulation for psychosis biotypes. The top five clinical features for each dimension are also provided along with their standard scores. B1, Biotype-1; B2, Biotype-2; B3, Biotype-3. (See eBook for color images.) (Modified from Clementz BA, Trotti RL, Pearlson GD, Keshavan MS, et al. Testing psychosis phenotypes from Bipolar–Schizophrenia Network for Intermediate Phenotypes for clinical application: Biotype characteristics and targets. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2020;5(8):808–818. doi:10.1016/j.bpsc.2020.03.011.)

An impediment to the development of psychiatric classifications with stronger biologic foundations is the lack of biologic data supporting such distinctions. This is a legitimate concern, but one with a certain level of circularity. Historically, DSM-type psychosis definitions are gold standards for determining the usefulness of neurobiologic measures. To the extent such definitions are orthogonal to neurobiologic reality, the ability to identify the etiologic distinctiveness of psychosis subtypes is severely limited. The B-SNIP project offers proof of concept that eschewing DSM-type clinical psychosis diagnoses makes possible the identification of neurobiologically distinctive psychosis subtypes.

The B-SNIP approach has limitations. The identification of subtypes is determined by the biomarkers used to dissect psychosis, the longitudinal characteristics of the Biotypes are uncertain, and there are probably more than three physiologically distinctive psychosis classes. But the verisimilitude of

Biotypes is supported by multiple external validating measures. Expanding these validations to pharmacologic profiles and molecular foundations is critical. B-SNIP is engaging in clinical trials, using specific treatments to target specific neurobiologic deviations, but so far, the clinical utility of this, or any other, neuroscience approach in psychiatry is unknown. Every patient, including our professor, deserves to know the outcomes that a rigorous approach to experimental medicine in psychiatry will provide. And even if psychiatry becomes a laboratory discipline, every patient deserves to have their troubles heard and understood by treatment staff. This is the way good medicine and good doctors operate. Becoming a laboratory discipline does not mean losing our humanity. It means improving the standard of care. At the very least, we should know if the entrenched approach, rooted in the DSM, is superior to competing alternatives. The B-SNIP consortium is excited for the future and the possibility that our work, and the work of similar research teams, will support a transformative reconceptualization of the psychoses. Maybe then, every case of psychosis, or at least some of them, will be crystal clear.

FURTHER READINGS

Bzdok D, Varoquaux G, Steyerberg EW. Prediction, not association, paves the road to precision medicine. *JAMA Psychiatry*. 2021;78(2):127–128.

Carpenter WT Jr, Buchanan RW, Kirkpatrick B, Tamminga C, Wood F. Strong inference, theory testing, and the neuroanatomy of schizophrenia. *Arch Gen Psychiatry*. 1993;50(10):825–831.

Casey BJ, Craddock N, Cuthbert BN, Hyman SE, Lee FS, Ressler KJ. DSM-5 and RDoC: progress in psychiatry research? *Nat Rev Neurosci*. 2013;14(11):810–814.

Clementz BA. Time for change in psychosis research. In: Tamminga CA, Ivleva E, Reininghaus U, van Os J, eds. *Psychotic Disorders: Comprehensive Conceptualization and Treatments*. Oxford University Press; 2021.

Clementz BA, Parker DA, Trotti RL, et al. Psychosis biotypes: replication and validation from the B-SNIP Consortium. *Schizophr Bull*. 2022;48(1):56–68.

Clementz BA, Trotti RL, Pearlson GD, et al. Testing psychosis phenotypes from bipolar-schizophrenia network for intermediate phenotypes for clinical application: biotype characteristics and targets. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020;5(8):808–818.

Craddock N, Owen MJ. The Kraepelinian dichotomy—going, going … but still not gone. *Br J Psychiatry*. 2010;196(2):92–95.

Fischer BA, Carpenter WT Jr. Will the Kraepelinian dichotomy survive DSM-V? *Neuropsychopharmacology*. 2009;34(9):2081–2087.

Gotra MY, Hill SK, Gershon ES, et al. Distinguishing patterns of impairment on inhibitory control and general cognitive ability among bipolar with and without psychosis, schizophrenia, and schizoaffective disorder. *Schizophr Res*. 2020;223:148–157.

Gottesman II, McGue M. Endophenotype. Cautin RL, Lilienfeld SO, eds. *The Encyclopedia of Clinical Psychology*, 1st ed. John Wiley & Sons, Inc; 2015.

Guimond S, Gu F, Shannon H, et al. A diagnosis and biotype comparison across the psychosis spectrum: investigating volume and shape amygdala-hippocampal differences from the B-SNIP study. *Schizophr Bull*. 2021;47(6):1706–1717.

Huang L-Y, Jackson BS, Rodrigue AL, et al. Antisaccade error rates and gap effects in psychosis syndromes from bipolar-schizophrenia network for intermediate phenotypes 2 (B-SNIP2).

Psychol Med. 2022;52(13):2692–2701.

Hyman SE. The diagnosis of mental disorders: the problem of reification. *Annu Rev Clin Psychol*. 2010;6:155–179.

Insel TR, Cuthbert BN. Medicine. Brain disorders? Precisely. *Science.* 2015;348(6234):499–500.

Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.

Jablensky A. The diagnostic concept of schizophrenia: its history, evolution, and future prospects. *Dialogues Clin Neurosci*. 2010;12(3):271–287.

Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry*. 2012;17(12):1174–1179.

Kendler KS, Parnas J, eds. *Philosophical Issues in Psychiatry ii: Nosology*. Oxford University Press; 2012.

Lichtenstein P, Yip BH, Björk C, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*. 2009;373(9659):234–239.

Mitchell KJ. What is complex about complex disorders? *Genome Biol*. 2012;13(1):237.

National Research Council. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. The National Academies Press; 2011.

Parker DA, Trotti RL, McDowell JE, et al. Auditory oddball responses across the schizophrenia-bipolar spectrum and their relationship to cognitive and clinical features. *Am J Psychiatry*. 2021;178(10):952–964.

Parker DA, Trotti RL, McDowell JE, et al. Auditory paired-stimuli responses across the psychosis and bipolar spectrum and their relationship to clinical features. *Biomark Neuropsychiatry*. 2020;3:100014.

Platt JR. Strong inference: certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*. 1964;146(3642):347–353.

Rzhetsky A, Wajngurt D, Park N, Zheng T: Probing genetic overlap among complex human phenotypes. *Proc Natl Acad Sci USA*. 2007;104(28):11694–11699.

Schneider M, Debbané M, Bassett AS, et al. Psychiatric disorders from childhood to adulthood in 22q11.2 deletion syndrome: results from the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. *Am J Psychiatry.* 2014;171(6):627–639.

Smoller JW. Disorders and borders: psychiatric genetics and nosology. *Am J Med Genet B Neuropsychiatr Genet*. 2013;162B(7):559–578.

Sullivan PF, Agrawal A, Bulik CM, et al. Psychiatric genomics: an update and an agenda. *Am J Psychiatry*. 2018;175(1):15–27.

Tamminga CA, Ivleva EI, Keshavan MS, et al. Clinical phenotypes of psychosis in the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *Am J Psychiatry.* 2013;170(11):1263–1274.

Turetsky BI, Dress EM, Braff DL, et al. The utility of P300 as a schizophrenia endophenotype and predictive biomarker: clinical and socio-demographic modulators in COGS-2. *Schizophr Res*. 2015;163(1–3):53–62.

12.12 Schizophrenia: Pharmacologic Treatment

John M. Kane, M.D., and Christoph U. Correll, M.D.

Nearly every patient with schizophrenia will benefit from pharmacologic treatment but to varying degrees. Drugs to treat psychosis (aka “antipsychotics”), currently all still reducing postsynaptic dopamine receptor activity, are the mainstay of pharmacologic treatment, being effective for reducing the impact of psychotic symptoms such as hallucinations, delusions, and disorganized thoughts and behavior, as well as related symptoms of agitation and aggression. In many patients, these symptoms can be completely eliminated or reduced to only a mild level. Medications can also decrease the likelihood that symptoms will recur. Nevertheless, negative symptoms and cognitive dysfunction are inadequately treated with currently available medications.

Controlling positive psychotic symptoms is often insufficient for improving the community functioning and the quality of life of patients with schizophrenia. There is also an increased awareness of the serious health problems of patients with schizophrenia and the possible contributing role of some medications. These observations have led to principles of drug treatment that emphasize the necessary, but not always sufficient, role of drugs for psychosis in helping patients achieves their personal goals.

HISTORY

The history of somatic therapies in schizophrenia can be divided into two eras, with the discovery of chlorpromazine (Thorazine)—the first clearly effective drug for psychosis—as the dividing line. During the late 19th and early 20th centuries, schizophrenia was viewed as associated with an inevitable deterioration into dementia. Patients were frequently hospitalized for long periods of time. Somatic treatments were used to help control the most severe symptoms of the disorder. Sedating agents such as bromides and barbiturates were used to control agitation, and treatments such as hydrotherapy and wet sheet packs were also used for calming effects. In the early 1920s, sleep treatment with barbiturates was introduced, based on the observation of improvement following an overdose of barbiturates. This involved maintaining patients in a highly sedated state for days.

Insulin coma treatment, introduced during the 1930s, involved gradually increasing doses of insulin causing a coma. After an hour of monitoring, glucose was administered, terminating the coma. Up to 20 insulin comas were widely used in the treatment of psychosis, suggesting some apparent effectiveness. It was abandoned when dopamine D2 antagonists were introduced.

Prefrontal lobotomy was proposed as a treatment for serious mental illnesses by Moniz in 1935. Frontal lobe extirpations in monkeys resulted in the appearance of less frustration. Although reports suggest that lobotomy may have been effective in reducing severe psychotic symptoms, worsening occurred in other areas, for example, personality deterioration with impulsive, psychopathic behaviors, as well as impairments in concept formation and planning. Psychosurgery was abandoned after the introduction of effective drugs.

Convulsive therapies were developed because some patients improved after a seizure. Drugs such as camphor and pentylenetetrazol (metrazol) were used initially to induce seizures but were abandoned after Cerletti and Bini introduced electrically induced convulsions. Initially, electroconvulsive therapy (ECT) was administered without anesthetics or muscle relaxants, inspiring fear in many patients, and the lack of muscle relaxants led to injuries from forceful muscle contractions. Modern ECT continues to have a role in certain patients with schizophrenia and is discussed in a later section.

The first effective drugs for psychosis were derived from the rauwolfia plant. Publications from the 1930s and 1940s suggested that these agents were effective for both hypertension and psychosis. Reserpine, the most potent of the rauwolfia alkaloids, which acts as an irreversible vesicular monoamine receptor (VMAT) inhibitor, depleting dopamine and other biogenic amines stored in presynaptic vesicles, was introduced in the early 1950s and was widely prescribed for schizophrenia and other psychotic illnesses. Studies comparing reserpine to dopamine receptor antagonists suggested similar efficacy. However, reserpine’s side effects, particularly depression, led to preference for dopamine receptor antagonists.

The discovery of chlorpromazine in the early 1950s may be the most important single contribution to the treatment of a psychiatric illness. Laborit, a surgeon in Paris, noticed that administering chlorpromazine to patients prior to surgery resulted in an unusual state in which they seemed less anxious regarding the procedure. In 1952 he convinced Delay and Deniker and other psychiatrists to administer chlorpromazine to psychotic and excited patients. The effects were extraordinary. Chlorpromazine, the prototypical first-generation drug for psychosis (aka FGA) acting as dopamine D2 antagonist, was effective at reducing hallucinations, delusions, and excitement, but caused side effects, including rigidity, tremor, and bradykinesia, that appeared similar to parkinsonism. Because chlorpromazine was effective and relatively easy to use, it and other drugs for psychosis were partially responsible for a substantial reduction in hospitalizations.

Thioridazine (Mellaril) and fluphenazine (Prolixin), as well as newer FGA classes, such as the butyrophenones (e.g., haloperidol [Haldol]) and the thioxanthenes (e.g., thiothixene [Navane]), were then developed. Although these newer agents differed in potency and side effects, all were similar in their effectiveness. Clozapine (Clozaril), the first effective drug for psychosis with negligible extrapyramidal side effects (EPS) and the first “second-generation antipsychotic” (SGA), was discovered in 1958 and first studied during the 1960s. However, in 1976 clozapine was found to have a substantial risk of agranulocytosis resulting in delayed introduction. In 1990, clozapine became available in the United States, for patients responding poorly to other agents. Subsequent SGAs (dopamine D2-serotonin 2A antagonists or partial dopamine D2 agonists) provided new alternatives-risperidone (Risperdal) introduced in 1994, olanzapine (Zyprexa) in 1996, quetiapine (Seroquel) in 1997, ziprasidone (Geodon) in 2001, aripiprazole (Abilify) in 2002, paliperidone (Invega) in 2007, asenapine (Saphris) in 2009, iloperidone (Fanapt) in 2009), lurasidone (Latuda) in 2010, brexpiprazole (Rexulti) in 2015, cariprazine (Vraylar) in 2015, lumateperone (Caplyta) in 2019 and olanzapine-samidorphan (Libalvi) in 2021. Table 12.12–1 summarizes pharmacologic, formulation, and dosing information.

PHASES OF TREATMENT IN SCHIZOPHRENIA

The acute stage is usually characterized by positive psychotic symptoms and related agitation/aggression that require immediate clinical attention. These symptoms may represent a first psychotic episode or, more commonly, a relapse in an individual who has experienced prior episodes. Treatment during this phase focuses on alleviating the most severe psychotic symptoms. Following the acute phase, usually 4 to 8 weeks, patients will generally enter a stabilization phase in which acute symptoms have been controlled, but risk for relapse is high with treatment interruption. During this (up to 6 months) phase, treatment focuses on consolidating therapeutic gains. Next is the maintenance phase with symptoms in relative remission or stable. The goals are to prevent psychotic relapse/exacerbation and improve functioning. Although controlling positive symptoms is a major goal minimization of negative symptoms and cognitive dysfunction is important since negative symptoms and cognitive disturbances can contribute to functional impairment more than positive symptoms.

FIRST EPISODE SCHIZOPHRENIA

The treatment of patients with early-phase illness differs from that of patients with chronic schizophrenia. First-episode patients are generally more treatment responsive, require lower drug doses (approximately 50%), and are more sensitive to side effects. They often have a difficult time accepting their illness and are at high risk for nonadherence, relapse, psychosocial deterioration, and suicidality. Although medication for psychosis is the cornerstone of management, multidisciplinary interventions, such as in coordinated specialty care, focusing on sustained engagement, relapse prevention, physical health, and functional recovery, are paramount.

EFFECTIVENESS OF DRUGS FOR PSYCHOSIS

A large body of evidence supports the effectiveness of dopamine receptor-blocking drugs. Many studies were carried out in the 1960s, addressing skepticism that these agents truly improved psychosis. The 1995 Schizophrenia Patient Outcomes Research Team (PORT) found that about 70% of patients treated with a drug for psychosis achieved relative remission or substantial improvement in contrast to 25% of patients treated with placebo. Recent meta-analyses of high-quality randomized controlled, parallel-design studies confirmed that dopamine receptor-blocking drugs are significantly superior to placebo for total symptomatology and positive symptoms.

TREATMENT OF ACUTE EPISODES

Indications for Pharmacologic Treatment

Nearly all schizophrenia patients with acute psychotic symptoms will benefit from a drug for psychosis, albeit to different degrees. Aside from relieving the patient from symptoms, there is some evidence indicating that lengthy delays in initiating dopamine receptor-blocking treatment may worsen the long-term course of schizophrenia. This evidence is summarized in a recent review by Howes et al. who found that delays in treatment were associated with significantly poorer results in 9 of 14 meta-analyzable outcomes, including worse symptomatic, social, functional, cognitive, and vocational outcomes. Although the reviewed studies have important limitations, it is probably prudent for clinicians to consider the possibility that untreated psychosis can result in poorer treatment response and prognosis, more frequent relapses, longer time spent in relapse, and, possibly, cortical volume reductions.

However, these data do not mean that all patients need to be treated immediately. A brief delay may permit clinicians to conduct a more thorough diagnostic evaluation and rule out causes of abnormal behavior, such as substance abuse, extreme stress, medical illnesses, and other psychiatric disorders.

Assessment

Whenever possible, patients should receive a physical, neurologic, and mental status examination and a laboratory evaluation before medications are started. Blood tests for complete blood cell count (CBC), electrolytes, fasting glucose and lipid profile, hemoglobin A1C, liver, renal, and thyroid function should be ordered. Other evaluations that should be considered are pregnancy tests in women, as well as human immunodeficiency virus (HIV) and syphilis tests when relevant. Individuals with schizophrenia are at a higher risk for cardiovascular disease than the population at large. As a result, an electrocardiogram (ECG) should probably be done at the onset of treatment for many patients with cardiac risk factors. In case of sudden onset or neurologic signs, brain magnetic resonance imaging (MRI) and the testing for infectious or autoantibody-related encephalitis should be considered. The presence of movement disorders, particularly pre-existing tardive dyskinesia (TD), should be assessed as this may influence the choice of medication.

Since dopamine receptor-blocking drugs are relatively safe, treatment can usually begin before the results of laboratory tests are known. An exception is clozapine treatment, which should only begin after the patient is confirmed as having a normal CBC. Under emergent conditions, for example, in potentially dangerous situations when patients refuse to cooperate with an evaluation, drugs for psychosis can be administered prior to a medical evaluation.

Selection of a Drug for Psychosis

Although the usefulness of the distinction has been questioned, currently available drugs for psychosis are all postsynaptic dopamine receptor antagonists that can be categorized into two main groups: FGAs, or dopamine D2 receptor antagonists, and SGAs, or dopamine D2-serotonin 2A antagonists and partial D2 agonist-serotonin 2A antagonists (see Table 12.12–1). The FGAs are further categorized as being low, mid, or high potency, with the higher potency drugs having more specificity and greater affinity for D_2 receptors, and a greater tendency to cause EPS. Lower potency drugs are less likely to cause EPS but tend to have more effects on other neurotransmitter receptors and are thus more likely to cause postural hypotension, sedation, weight gain, and anticholinergic effects.

Table 12.12–1. Pharmacologic, Formulation and Dosing Information for Second-Generation Antipsychotics and Selected First-Generation Antipsychotics

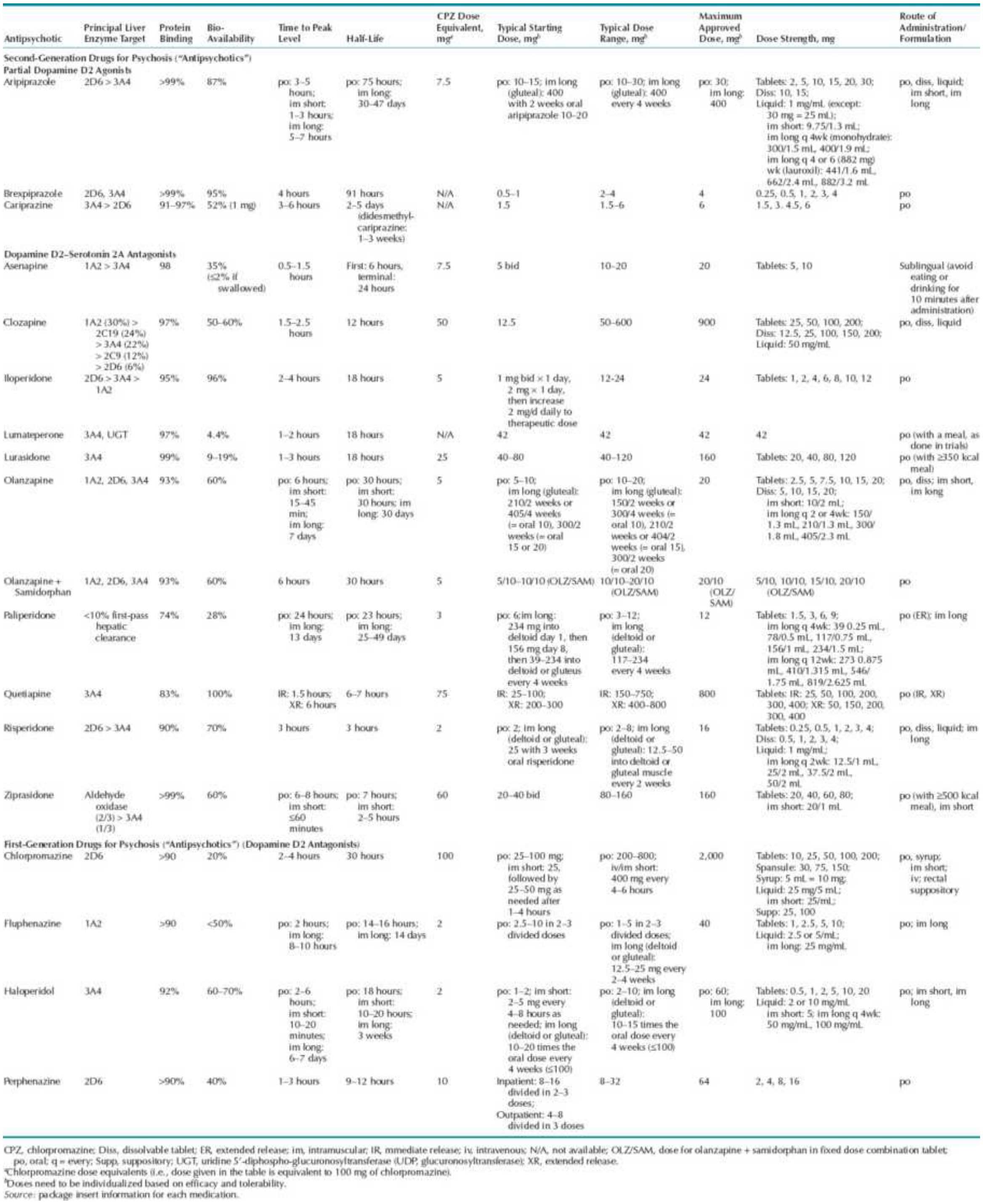

Antipsychotic	Principal Liver Enzyme Target	Protein Binding	Bio-Availability	Time to Peak Level	Half-Life	CPZ Dose Equivalent, mg[a]	Typical Starting Dose, mg[b]	Typical Dose Range, mg[b]	Maximum Approved Dose, mg[b]	Dose Strength, mg	Route of Administration/ Formulation
Second-Generation Drugs for Psychosis ("Antipsychotics")											
Partial Dopamine D2 Agonists											
Aripiprazole	2D6 > 3A4	>99%	87%	po: 3–5 hours; im short: 1–3 hours; im long: 5–7 hours	po: 75 hours; im long: 30-47 days	7.5	po: 10–15; im long (gluteal): 400 with 2 weeks oral aripiprazole 10-20	po: 10-30; im long (gluteal): 400 every 4 weeks	po: 30; im long: 400	Tablets: 2, 5, 10, 15, 20, 30; Diss: 10, 15; Liquid: 1 mg/mL (except: 30 mg = 25 mL); im short: 9.75/1.3 mL; im long q 4wk (monohydrate): 300/1.5 mL, 400/1.9 mL; im long q 4 or 6 (882 mg) wk (lauroxil): 441/1.6 mL, 662/2.4 mL, 882/3.2 mL	po, diss, liquid; im short, im long
Brexpiprazole	2D6, 3A4	>99%	95%	4 hours	91 hours	N/A	0.5–1	2-4	4	0.25, 0.5, 1, 2, 3, 4	po
Cariprazine	3A4 > 2D6	91–97%	52% (1 mg)	3–6 hours	2–5 days (didesmethyl-cariprazine: 1–3 weeks)	N/A	1.5	1.5–6	6	1.5, 3. 4.5, 6	po
Dopamine D2–Serotonin 2A Antagonists											
Asenapine	1A2 > 3A4	98	35% (≤2% if swallowed)	0.5–1.5 hours	First: 6 hours, terminal: 24 hours	7.5	5 bid	10–20	20	Tablets: 5, 10	Sublingual (avoid eating or drinking for 10 minutes after administration)
Clozapine	1A2 (30%) > 2C19 (24%) > 3A4 (22%) > 2C9 (12%) > 2D6 (6%)	97%	50-60%	1.5–2.5 hours	12 hours	50	12.5	50-600	900	Tablets: 25, 50, 100, 200; Diss: 12.5, 25, 100, 150, 200; Liquid: 50 mg/mL	po, diss, liquid
Iloperidone	2D6 > 3A4 > 1A2	95%	96%	2–4 hours	18 hours	5	1 mg bid × 1 day, 2 mg × 1 day, then increase 2 mg/d daily to therapeutic dose	12-24	24	Tablets: 1, 2, 4, 6, 8, 10, 12	po
Lumateperone	3A4, UGT	97%	4.4%	1–2 hours	18 hours	N/A	42	42	42	42	po (with a meal, as done in trials)
Lurasidone	3A4	99%	9–19%	1–3 hours	18 hours	25	40–80	40–120	160	Tablets: 20, 40, 80, 120	po (with ≥350 kcal meal)
Olanzapine	1A2, 2D6, 3A4	93%	60%	po: 6 hours; im short: 15–45 min; im long: 7 days	po: 30 hours; im short: 30 hours; im long: 30 days	5	po: 5–10; im long (gluteal): 210/2 weeks or 405/4 weeks (= oral 10), 300/2 weeks (= oral 15 or 20)	po: 10–20; im long (gluteal): 150/2 weeks or 300/4 weeks (= oral 10), 210/2 weeks or 404/2 weeks (= oral 15), 300/2 weeks (= oral 20)	20	Tablets: 2.5, 5, 7.5, 10, 15, 20; Diss: 5, 10, 15, 20; im short: 10/2 mL; im long q 2 or 4wk: 150/1.3 mL, 210/1.3 mL, 300/1.8 mL, 405/2.3 mL	po, diss; im short, im long
Olanzapine + Samidorphan	1A2, 2D6, 3A4	93%	60%	6 hours	30 hours	5	5/10–10/10 (OLZ/SAM)	10/10–20/10 (OLZ/SAM)	20/10 (OLZ/SAM)	5/10, 10/10, 15/10, 20/10 (OLZ/SAM)	po
Paliperidone	<10% first-pass hepatic clearance	74%	28%	po: 24 hours; im long: 13 days	po: 23 hours; im long: 25–49 days	3	po: 6;im long: 234 mg into deltoid day 1, then 156 mg day 8, then 39–234 into deltoid or gluteus every 4 weeks	po: 3–12; im long (deltoid or gluteal): 117–234 every 4 weeks	12	Tablets: 1.5, 3, 6, 9; im long q 4wk: 39 0.25 mL, 78/0.5 mL, 117/0.75 mL, 156/1 mL, 234/1.5 mL; im long q 12wk: 273 0.875 mL, 410/1.315 mL, 546/1.75 mL, 819/2.625 mL	po (ER); im long
Quetiapine	3A4	83%	100%	IR: 1.5 hours; XR: 6 hours	6–7 hours	75	IR: 25–100; XR: 200–300	IR: 150–750; XR: 400–800	800	Tablets: IR: 25, 50, 100, 200, 300, 400; XR: 50, 150, 200, 300, 400	po (IR, XR)
Risperidone	2D6 > 3A4	90%	70%	3 hours	3 hours	2	po: 2; im long (deltoid or gluteal): 25 with 3 weeks oral risperidone	po: 2–8; im long (deltoid or gluteal): 12.5–50 into deltoid or gluteal muscle every 2 weeks	16	Tablets: 0.25, 0.5, 1, 2, 3, 4; Diss: 0.5, 1, 2, 3, 4; Liquid: 1 mg/mL; im long q 2wk: 12.5/1 mL, 25/2 mL, 37.5/2 mL, 50/2 mL	po, diss, liquid; im long
Ziprasidone	Aldehyde oxidase (2/3) > 3A4 (1/3)	>99%	60%	po: 6–8 hours; im short: ≤60 minutes	po: 7 hours; im short: 2–5 hours	60	20–40 bid	80–160	160	Tablets: 20, 40, 60, 80; im short: 20/1 mL	po (with ≥500 kcal meal), im short
First-Generation Drugs for Psychosis ("Antipsychotics") (Dopamine D2 Antagonists)											
Chlorpromazine	2D6	>90	20%	2–4 hours	30 hours	100	po: 25–100 mg; im short: 25, followed by 25–50 mg as needed after 1–4 hours	po: 200–800; iv/im short: 400 mg every 4–6 hours	2,000	Tablets: 10, 25, 50, 100, 200; Spansule: 30, 75, 150; Syrup: 5 mL = 10 mg; Liquid: 25 mg/5 mL; im short: 25/mL; Supp: 25, 100	po, syrup; im short; iv; rectal suppository
Fluphenazine	1A2	>90	<50%	po: 2 hours; im long: 8–10 hours	po: 14–16 hours; im long: 14 days	2	po: 2.5–10 in 2–3 divided doses	po: 1–5 in 2–3 divided doses; im long (deltoid or gluteal): 12.5–25 mg every 2–4 weeks	40	Tablets: 1, 2.5, 5, 10; Liquid: 2.5 or 5/mL; im long: 25 mg/mL	po; im long
Haloperidol	3A4	92%	60–70%	po: 2–6 hours; im short: 10–20 minutes; im long: 6–7 days	po: 18 hours; im short: 10–20 hours; im long: 3 weeks	2	po: 1–2; im short: 2–5 mg every 4–8 hours as needed; im long (deltoid or gluteal): 10–20 times the oral dose every 4 weeks (≤100)	po: 2–10; im long (deltoid or gluteal): 10–15 times the oral dose every 4 weeks (≤100)	po: 60; im long: 100	Tablets: 0.5, 1, 2, 5, 10, 20 Liquid: 2 or 10 mg/mL im short: 5; im long q 4wk: 50 mg/mL, 100 mg/mL	po; im short, im long
Perphenazine	2D6	>90%	40%	1–3 hours	9–12 hours	10	Inpatient: 8–16 divided in 2–3 doses; Outpatient: 4–8 divided in 3 doses	8–32	64	2, 4, 8, 16	po

CPZ, chlorpromazine; Diss, dissolvable tablet; ER, extended release; im, intramuscular; IR, mmediate release; iv, intravenous; N/A, not available; OLZ/SAM, dose for olanzapine + samidorphan in fixed dose combination tablet; po, oral; q = every; Supp, suppository; UGT, uridine 5′-diphospho-glucuronosyltransferase (UDP, glucuronosyltransferase); XR, extended release.

[a] Chlorpromazine dose equivalents (i.e., dose given in the table is equivalent to 100 mg of chlorpromazine).

[b] Doses need to be individualized based on efficacy and tolerability.

Source: package insert information for each medication.

Comparisons of Drugs for Psychosis

Numerous studies have examined the comparative effectiveness and tolerability of various drugs for psychosis, and they generally appear to be similarly effective but with substantial differences in adverse effects. However, amisulpride (not available in the United States), olanzapine, and risperidone have been shown in some meta-analyses to produce somewhat greater symptom responses. Clozapine is more effective than other medications for individuals who have psychotic symptoms that persist despite adequate treatment.

Because of the perception of bias in industry-sponsored studies, government-funded agencies in the United Kingdom and the U.S.-sponsored studies intended to produce independent research findings regarding the longer-term effects of FGAs and SGAs. These studies, the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) schizophrenia trial in the United States and the Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS) in the United Kingdom confirmed that clozapine seemed advantageous for individuals with refractory symptoms, but did not identify clearly important advantages in effectiveness for any other drug or drug class for most patients. Both studies found that people tended to continue taking olanzapine longer than other medications, but neither study found greater reductions in symptoms or improvements in quality of life for olanzapine. Neither CATIE nor CUtLASS found substantial advantages in overall tolerability or acceptability for SGAs over FGAs. The CATIE study found that some drugs, in particular clozapine, olanzapine, and quetiapine, were associated with weight gain and serum lipid abnormalities that are risk factors for cardiovascular disease. These findings have been substantiated by subsequent studies and meta-analyses.

Route of Administration

Usually, patients should initially be treated with an oral medication. Short-acting intramuscular drugs are useful when the patient refuses oral dosing and when a very rapid onset is desired. Intramuscular administration of most drugs for psychosis results in peak plasma levels in about 30 minutes with clinical effects emerging within 15 to 30 minutes (see Table 12.12–1). Most orally administered drugs for psychosis result in a peak plasma level in 1 to 6 hours following administration (see Table 12.12–1).

Long-Acting Injectables

Since nonadherence to drugs for psychosis is common and one of the most relevant and preventable risk factors for insufficient response and relapse, long-acting injectable formulations (LAIs) of several drugs have been developed (see Table 12.12–1). These drugs differ from short-acting compounds in that they are usually started after an initial trial of the oral formulation and are released slowly over a period of several weeks, which is why it can take days or weeks to achieve therapeutic serum concentration and weeks to months to achieve a steady state. As a result, LAIs are helpful for continuation and maintenance treatment. In the United States, the two FGAs, fluphenazine, and haloperidol, have been formulated as oil-based esters formed between the alcohol group of the drug and a long-chain fatty acid. Currently, four SGAs, risperidone, olanzapine, paliperidone, and aripiprazole, are available as LAIs in the United States. Formulations of all four agents are water-based that require biweekly (risperidone, certain doses of olanzapine), monthly (aripiprazole, paliperidone, certain doses of olanzapine), 6-weekly or bimonthly (certain doses/formulations of aripiprazole), or 3-monthly or 6-monthly (certain doses of paliperidone) injections. Oral supplementation, LAI loading doses, or an LAI booster injection is necessary for some medications to initially reach therapeutic blood levels. LAIs of other SGAs are currently in development. Given the consequences of preventable relapse, LAIs should be considered for patients requiring maintenance treatment, including those in the early course of illness. However, relapse can occur on LAI treatment, in approximately 15% of patients in full symptomatic remission and up to 30% in patients not in remission, with substance use and TD increasing relapse risk on LAIs. While the general perception has been that most patients will refuse LAIs, recent data suggest that the majority will agree when LAIs are presented in a team-based shared-decision approach.

Starting Drugs for Psychosis

Prior to starting a drug for psychosis, in a shared decision model clinicians should describe the medication being suggested, its target symptoms, and its possible side effects, particularly those that are common and unpleasant (e.g., muscle stiffness, sedation, akathisia). Patients who are severely disturbed may be unable to participate meaningfully in this discussion. However, most will benefit from information about the benefits and risks associated with drugs for psychosis. Since patients with schizophrenia may be suspicious, it is particularly important to emphasize participation as collaborators in interpreting medication effects. It is helpful to also involve family members in decision-making.

In some settings, it is necessary for patients to give written or verbal consent prior to receiving a drug for psychosis. This can be a dilemma for patients who are conceptually disorganized and find it difficult to understand the risks and benefits of drug treatment. Under these circumstances, clinicians should adjust the complexity of the discussion to the patient’s mental state. Thus, it may be appropriate to provide a limited amount of information that focuses on the most common acute side effects of the medication when the patient is most seriously impaired. As the patient improves, clinicians may then elaborate on the costs and benefits of medication. For example, detailed discussions about TD or weight gain and metabolic side effects associated with chronic treatment may be deferred until the patient has improved and long-term maintenance is considered.

Thus, it is important for psychiatrists to evaluate whether acutely disturbed patients are able to participate meaningfully in decisions about their medication. Clinicians should become familiar with local and state laws that affect a patient’s right to refuse or accept drug treatment. The most difficult situation is when a patient who desperately needs medication refuses it. Under some conditions, family members who have been educated about schizophrenia may be helpful in convincing patients to accept medication. Every locality has provisions for treating patients against their will under emergency conditions. Some permit involuntary treatment when certain conditions are met. As patients improve, the majority will eventually accept their own need for medication. Many localities now have laws allowing for compulsory outpatient treatment, particularly for patients who have a history of violent or aggressive behavior.

Dosage Selection

Dosage selection is difficult because the physician is unable to titrate dose against clinical effects due to a variable time delay between initiation and clinical response. Recent studies and meta-analyses report that in 4 to 6 weeks of treatment, the greatest reduction in symptoms occurs first 1 to 2 weeks and little or no response during the first week or two (assuming a valid diagnosis, appropriate adherence, and therapeutic dose level) is a powerful predictor of subsequent lack of response.

The currently recommended dosage ranges for FGAs and SGAs are in the range of 300 to 600 mg daily of chlorpromazine equivalence. A number of dosage comparison studies and recent meta-analyses have failed to support the routine use of doses higher than the upper level of the approved dose range both in acute treatment and relapse prevention. However, the reverse has also shown to be true, in that doses below the lower level of the approved dose range are associated with a greater risk of nonresponse and relapse in group-level analyses. In the context of nonresponse to adequate doses of drugs for psychosis, it might be appropriate to assess drug blood levels or polymorphisms of the cytochrome-P450 enzymes that are responsible for the liver metabolism of the specific drug (see Table 12.12–1) in order to identify patients who are "pseudo"-resistant because of poor adherence or ultra-rapid metabolism.

Limited data from controlled trials of clozapine support the practice of using doses in the range of 300 to 500 mg daily. However, side effects, particularly sedation and orthostatic hypotension, are often limiting factors. The risk of seizures increases substantially above 600 mg. Plasma levels of 350 to 450 ng/mL, have been recommended, and measuring plasma levels is recommended for poor responders. Therapeutic dose levels of other drugs for psychosis are under-researched.

Large multicenter trials indicate that risperidone is most effective at 4 to 8 mg daily (see Table 12.12–1). Higher doses may lead to EPS and discontinuation without an advantage in increased effectiveness. In the United States, the average dose of risperidone prescribed for schizophrenia is slightly more than 4 mg daily. This suggests that a reasonable practice would be to manage patients with schizophrenia with 4 mg of risperidone and increase the dose if they fail to respond after 4 to 6 weeks. Olanzapine is usually effective in the range of 10 to 20 mg daily (see Table 12.12–1), although a number of case reports describe individuals who demonstrated optimal responses at doses of 25 mg and higher. Case reports also identify patients who had demonstrated substantial improvements when doses of quetiapine were raised well above 800 mg. At this time, there is no evidence that prescribing doses of ziprasidone greater than 160 mg daily increases its effectiveness (ziprasidone should be taken with at least a 500 kcal meal) (see

Table 12.12–1). Similarly, there is no evidence that doses of aripiprazole greater than 30 mg daily are more effective than the recommended range of 10 to 30 mg daily (see Table 12.12–1). Paliperidone's optimal dose range appears to be 6 to 12 mg daily, with higher doses associated with more adverse effects but not better effectiveness (see Table 12.12–1). Asenapine's optimal dose seems to be between 5 mg bid and 10 mg bid, taken sublingually (avoid eating or drinking for 10 minutes after administration) (see Table 12.12–1). Iloperidone's effective dose range is between 6 and 12 mg bid, but titration is needed to avoid orthostatic hypotension (i.e., 1 mg bid on day 1, 2 mg bid on day 2, then increase by 2 mg/day until a therapeutic dose is reached) (see Table 12.12–1). Lurasidone is effective between 40 and 160 mg (taken with at least a 350 kcal meal) (see Table 12.12–1). In the acute studies, brexpiprazole was initiated at 1 mg daily for 4 days, increased to 2 mg daily on day 5, and, if needed, to 4 mg daily on day 8. Cariprazine was initiated at 1.5 mg daily and increased every 1 to 2 days by 1.5 mg daily up to the upper approved dose of 6 mg daily. Finally, lumateperone is currently only approved as a single 42 mg daily dose, which means that no titration is needed (see Table 12.12–1). Some of the more rapid dose titration schemes are clinically adapted to being slower in outpatient and non-acute settings resulting in less initial adverse effect burden.

A number of recent findings suggest a reasonable strategy for treating acute schizophrenia. The dose of a postsynaptic dopamine receptor blocker that is likely to be effective is the dose that occupies an appropriate number of D_2 receptors. For dopamine D2 antagonists, this occupancy is approximately 60% to 70% of D2 receptors measured in the striatum, while for dopamine D2 partial agonists, this occupancy needs to be more than 90%, and for clozapine, quetiapine, and lumateperone approximately 40% postsynaptic dopamine D2 occupancy in the striatum appears to be sufficient for antipsychotic efficacy. The therapeutic response might depend to some extent on processes that occur after these receptors have been occupied for a period of time. This observation is supported by findings from both positron emission tomography (PET) scanning and the measurement of plasma homovanillic acid, which suggest that clinical improvement is not associated with the immediate effects of the drug on dopamine receptors, but on processes that occur later.

Initial dosage should follow labeling guidelines. Time-limited cotreatment with a benzodiazepine or another sedating medication can be helpful while awaiting response to the drug for psychosis. Since most of the improvement on the latter occurs in 2 to 4 weeks, patients should be observed for this interval before changing a drug; however, if there is little or no symptom improvement after 2 weeks, despite adherence to a therapeutic dose level, later response is unlikely to occur, and a switch or dose increase within the therapeutic range, should be considered.

Managing Agitation in Acute Psychosis

Agitation in acute schizophrenia can result from disturbing psychotic symptoms, such as frightening delusions or suspiciousness, or from other causes, including stimulant abuse or akathisia. Patients with akathisia can appear agitated when they experience a subjective feeling of motor restlessness. Differentiating akathisia from psychotic agitation can be difficult when patients are incapable of describing their internal experience. If patients are receiving an agent associated with EPS, usually an FGA, a trial with an anticholinergic medication (helpful for Parkinsonism) or propranolol (Inderal, helpful for akathisia) may be useful in making the distinction.

Clinicians have a number of options for managing psychotic agitation. Dopamine receptor blockers and benzodiazepines can result in relatively rapid calming. An advantage of a drug for psychosis is that a single intramuscular injection of haloperidol, fluphenazine, olanzapine, aripiprazole, or ziprasidone will often result in calming without an excess of sedation. The same is true for inhalable loxapine. Low-potency FGAs are often associated with sedation and postural hypotension, particularly when they are administered intramuscularly. Intramuscular ziprasidone, aripiprazole, and olanzapine are similar to their oral counterparts in not causing substantial EPS during acute treatment. This can be an important advantage over haloperidol or fluphenazine, which can cause frightening dystonia or akathisia in some patients, although coadministration of the anticholinergic benztropine (Cogentin) or the antihistamine promethazine (Phenergan) can reduce the risk of dystonia substantially. Rapidly dissolving oral formulation of olanzapine, risperidone, or aripiprazole may also be helpful as an alternative to an intramuscular injection.

Benzodiazepines are effective for psychotic agitation. Lorazepam (Ativan) has the advantage of reliable absorption when administered either orally or intramuscularly. The combination of lorazepam with a postsynaptic dopamine receptor blocker has been found to be safer and more effective than large doses of dopamine receptor blockers in controlling excitement and motor agitation. Due to the sedating properties of short-acting injectable olanzapine that may depress respiratory drive, it should not be combined with benzodiazepines within a 2-hour window prior to or after its administration. Finally, the alpha2 agonist dexmedetomidine, which has been available for decades as an I.V. anesthetic has recently been approved in an oral formulation to treat acute agitation in patients with schizophrenia or bipolar disorder.

MANAGING SIDE EFFECTS

Patients will frequently experience side effects of a drug for psychosis before they experience clinical improvement. Whereas a substantial clinical response may be delayed for days or weeks, side effects will often begin quickly. For low-potency drugs, these side effects are likely to include sedation, postural hypotension, and anticholinergic effects, whereas high-potency drugs are likely to cause EPS.

A patient’s interpretation of a drug’s effectiveness is often influenced by adverse effects and a challenge with psychotic individuals is maintaining the trust of those who may misinterpret experiences and become suspicious. Warning patients of potential side effects can lead to prompt management, improved trust, and better adherence.

Extrapyramidal Side Effects (EPS)

A widely accepted benefit of SGAs is reduced propensity for EPS. Even relatively low doses of high-potency FGAs (e.g., haloperidol 4 mg per day) are associated with more EPS than recommended doses of SGAs. Because EPS can occur with any currently available drug for psychosis, including SGAs, awareness of EPS remains critical.

A common form of EPS is akathisia, a side effect consisting of a subjective feeling of restlessness with or without restless movements, usually in the legs or feet. Patients who experience severe akathisia will often pace continuously or move their feet restlessly while they are sitting. Some complain that they are unable to feel comfortable, regardless of what they do. Severe akathisia can cause patients to feel anxious or irritable, and some reports suggest that severe akathisia can result in aggressive or suicidal acts. Researchers have estimated that 25% to 75% of patients treated with a high-potency FGAs will experience akathisia. Akathisia is also thought to be a correlate of poor drug response. The dopamine D2 partial agonists aripiprazole, brexpiprazole, and cariprazine can lead to early restlessness that may not be true akathisia, but rather due to early stimulatory effects that are minimized or controlled by switching slowly from a full dopamine D2 antagonist or a sedating or anticholinergic drug to the partial D2 agonist, using a lower starting dose or cotreating temporarily with a benzodiazepine or antihistamine. Often, when the initial restlessness subsides, dose escalation can be tolerated without development of true akathisia. Of note, akathisia rates were significantly reduced with lurasidone despite use of higher doses when the administration was moved from the morning to the evening, suggesting that movement-related side effects that may be peak concentration-related can be minimized when administering the drug in the evening.

Since patients may experience akathisia as irritability or agitation, asking patients whether they are restless or if they have difficulty sitting still can be helpful in early stages of treatment. At this point, a dosage adjustment, a centrally acting β-blocker such as propranolol (at doses between 30 and 90 mg daily or a benzodiazepine, which would also treat psychotic agitation, may provide relief.

Dystonias are probably the most frightening type of EPS. They are intermittent or sustained muscular spasms and abnormal postures affecting mainly the musculature of the head and neck but sometimes the trunk and lower extremities. Common forms of dystonia include abnormal positioning of the neck, impaired swallowing (dysphagia), hypertonic or enlarged tongue, and deviations of the eyes (oculogyric crisis). The rare form of laryngospasm is an acute emergency, requiring I.M or I.V treatment with an anticholinergic, antihistamine, or benzodiazepine. These reactions usually appear within the first few days of therapy. Dystonias are more likely to occur in younger patients, particularly young males.

Drug-induced parkinsonism consists of tremor, muscular rigidity, and a decrease in spontaneous movements resembling the movement disorder in idiopathic parkinsonism and affecting about 30% of patients treated with FGAs. The first evidence of drug-induced parkinsonism may be a diminished arm swing or decreased facial expressiveness. Akinesia can be subtle and can be mistaken for or exacerbate negative symptoms. Risk factors for drug-induced parkinsonism include increasing age and dose, FGA use, a history of parkinsonism, and underlying basal ganglia damage.

Clinicians have a number of alternatives: reducing the dose, most commonly of an FGA, adding an antiparkinson medication, or changing the patient to an SGA. The most effective medications are the anticholinergic antiparkinson drugs (which are described in another section). Although these medications are frequently effective, they also cause their own side effects, including dry mouth, constipation, blurred vision, and cognitive disturbance. Also, these medications are frequently only partially effective. Anticholinergic use can often be discontinued in the maintenance treatment phase in patients without enduring EPS.

If FGAs are prescribed, clinicians may consider prescribing prophylactic antiparkinson medications for patients who are likely to experience disturbing EPS. These include patients who have a history of EPS sensitivity or those who are being treated with high-potency FGAs. Prophylactic antiparkinson medications may also be indicated when high-potency drugs are prescribed for young men who tend to have an increased vulnerability for developing dystonias. These patients may be good candidates for SGAs. However, the need for chronic anticholinergic treatment should be avoided as indicated above.

Some individuals are highly sensitive to EPS at dosages necessary to control their psychosis. For many of these patients, medication side effects may seem worse than the illness itself. These patients should routinely be treated with drugs associated with lower incidence of EPS, which means that in general SGAs would be preferred to FGAs, although even SGAs can lead to EPS in a dose-dependent fashion. Clozapine, quetiapine, brexpiprazole and lumateperone seem to cause the least EPS, while risperidone, olanzapine, paliperidone, ziprasidone, iloperidone, asenapine, cariprazine, and aripiprazole may cause dose-related EPS.

Tardive Dyskinesia and Other Tardive Syndromes

As with EPS, TD is less common with the SGAs than with conventional drugs, although there is probably a range of risks within both groups of drugs. Prospective studies lasting >6 months demonstrate a significantly lower risk of TD with SGAs (especially aripiprazole and olanzapine) than haloperidol. These results were somewhat expected in that early occurring EPS are significant risk factors for TD. However, the risk of TD is not absent with SGAs, and clinicians must recognize and manage TD.

TD commonly consists of abnormal, involuntary movements of the mouth, face and tongue, trunk, and extremities. The oral-facial movements occur in about three-fourths of TD patients and can include lip smacking, sucking, puckering, and facial grimacing or irregular movements of the limbs, particularly choreoathetoid-like movements of the fingers and toes, and slow, writhing movements of the trunk. Younger patients with TD tend to develop slower athetoid movements of the trunk, extremities, and neck, but coexistence with tardive dystonic movements is not uncommon.

The abnormal movements of TD are usually reduced by voluntary movements of the affected areas and are increased by voluntary movements of unaffected areas increased with emotional arousal and absent during sleep. According to the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), the abnormal movements should be present for at least 4 weeks, and patients should have been exposed to a postsynaptic dopamine receptor blocker for at least 3 months (at least one month in the older adults), and the onset of the abnormal movements should occur either while the patient is receiving the drug or within 4 weeks of discontinuing an oral or 8 weeks after the withdrawal of an LAI.

Prevalence surveys indicate that 25% to 30% of patients chronically treated with an FGA and 7% to 20% of patients chronically treated with an SGA will exhibit symptoms of TD. Three to five percent of young patients receiving an FGA develop TD each year. The risk in older adult patients is much higher. Although seriously disabling dyskinesia is uncommon, in a small proportion of cases walking, breathing, eating, and talking are affected.

Recommendations for preventing and managing TD include: (1) establishing objective evidence that the drug for psychosis is indicated and effective for the individual; (2) utilizing the lowest effective dose of the drug; (3) prescribing cautiously with children, older adult patients, those with EPS and patients with mood disorders; (4) examining patients on a regular basis for evidence of TD; (5) when TD is diagnosed, consider alternatives to the drug for psychosis being used, obtain informed consent, and consider dosage reduction; (6) if the TD does not improve or worsens, consider discontinuing the offending medication if possible, switching from an FGA to an SGA or an SGA to clozapine, or using one of the two recently approved VMAT-2 inhibitors, deutetrabenazine, or valbenazine.

Routine monitoring of TD after reaching the maintenance dose should include individuals <50 years old examination every 6 months if on an FGA and every 12 months if on an SGA. Monitoring for individuals aged ≥50 and high-risk groups should be every 3 months, which includes patients who are sensitive to EPS or treated with an anticholinergic, individuals with affective or organic brain disorders, and those with diabetes, negative or cognitive symptoms, and comorbid substance use disorders.

Other Side Effects

Sedation and postural hypotension can be important side effects for patients who are being treated with low-potency FGAs, such as chlorpromazine or thioridazine, and with clozapine or iloperidone. These effects are often most severe during the initial dosing, and patients receiving these medications, particularly clozapine, may require weeks until they reach a therapeutic dose. Although most patients develop tolerance to sedation and postural hypotension, sedation may continue to be a problem that can interfere with functional attainment. Additionally, some agents have been related to insomnia. Neuroleptic malignant syndrome (NMS) is a rare (approximately 2 cases per 10,000 patient-years) but severe adverse effect of dopamine receptor blockers. Although LAIs cannot be stopped abruptly, mortality and adverse sequelae are not increased when NMS develops during LAI use vs. oral medication. Moreover, >90% of patients were successfully rechallenged with a drug for psychosis, although titration and dosing should likely be slower and lower.

All FGAs, as well as risperidone and paliperidone, elevate prolactin levels, which can result in suppression of the gonadotropin-releasing hormone leading to clinically important suppression in gonadal hormones, affecting the reproductive system, and sexual functioning, including amenorrhea or oligomenorrhea, galactorrhea, decreased libido, erectile dysfunction, or anorgasmia. In addition, elevated prolactin may cause decreases in bone density and lead to osteoporosis. Finally, prolactin-raising medications may increase the risk of breast cancer to a small but measurable degree.

Prior to the introduction of SGAs, elevated prolactin was a consequence of treatment with all drugs for psychosis. Clozapine, quetiapine, and lumateperone, as well as the three partial D2 agonists, aripiprazole, brexpiprazole, and cariprazine, do not appear to elevate prolactin above normal levels, with partial D2 agonist being able to decrease prolactin levels. As a result, when patients demonstrate side effects related to prolactin, changing patients to a prolactin-sparing agent may be effective. When the prolactin-raising agent cannot be changed, adding 5 to 15 mg of aripiprazole can normalize prolactin levels in a dose-dependent fashion.

Side Effects of Clozapine

As discussed elsewhere in this textbook, clozapine has a number of side effects. The most serious is agranulocytosis. This potentially fatal condition occurs in approximately 0.3% of patients treated with clozapine during the first year. Subsequently, the risk is substantially lower. Those receiving clozapine in the United States must have weekly blood monitoring for the first 6 months, biweekly for the next 6 months, and monthly thereafter.

Clozapine is associated with a higher risk of seizures than other drugs for psychosis (5% at >600 mg.) Seizures can usually be managed by reducing the dose and/or adding an anticonvulsant, usually valproate (Depacon). Myocarditis has been reported to occur in approximately 5 patients per 100,000 patient years, but emerging data suggest that less than half of rechallenged patients develop myocarditis again. Other side effects of clozapine include hypersalivation, sedation, tachycardia, weight gain, fever, and postural hypotension.

Metabolic Side Effects

Patients with schizophrenia are more likely to smoke and suffer from a number of illnesses, including coronary heart disease, diabetes, and hypertension. The high prevalence of these illnesses may explain why individuals with schizophrenia have a 15 to 20 years shorter life expectancy than the general population. This is partly explained by underlying genetic risk as well as inadequate medical care and unhealthy lifestyles-smoking, poor dietary habits, and lack of exercise.

Adverse effects can contribute to these problems. The risk for weight gain and abnormalities in glucose and lipid metabolism is greatest with clozapine, olanzapine, and chlorpromazine. The lowest risk is observed with aripiprazole, brexpiprazole, cariprazine, lumateperone, lurasidone, and ziprasidone, although problems may even occur with these agents, especially in first-episode drug naïve patients. All other agents have an intermediate risk for weight gain and glucose and lipid metabolism, with the exception of a higher risk for lipid abnormalities with quetiapine.

Health Monitoring in Patients Receiving Drugs for Psychosis

Routine cardiometabolic monitoring is recommended, at a minimum, for the assessment of body weight, blood pressure, and fasting glucose and lipids, ideally with fasting or random hemoglobin A1C at baseline, 3 months, and annually after drug initiation. More frequent assessments are recommended if significant weight gain occurs. When abnormalities are observed, healthy lifestyle counseling or interventions or switching to a lower-risk agent should be considered. If these fail or are not feasible, medications that target significant weight gain (metformin or topiramate) or that treat abnormal blood pressure, glucose, or lipid metabolism may be warranted, and consultation with medical specialists may be needed.

NEGATIVE, MOOD, AND COGNITIVE SYMPTOMS

Negative symptoms and cognitive impairments are major contributors to a social and vocational disability, resulting in placing a greater emphasis on treatment strategies for decreasing these impairments. Carpenter has made an important contribution to this area by classifying negative symptoms into primary and secondary categories. Secondary negative symptoms are those symptoms that may result from other conditions, such as depression, and positive symptoms, such as suspiciousness, sedation, or EPS. EPS is a common cause of secondary negative symptoms, particularly when patients are experiencing akinesia, a side effect that can manifest in decreased speech, motivation, and spontaneous gestures.

The management of secondary negative symptoms begins with the management of the condition that caused these symptoms. For depression, this may include the addition of an antidepressant medication (which can reduce negative symptoms even in the absence of overt depression); for EPS this may involve the addition of an antiparkinson medication, a dose reduction, or a change to another drug for psychosis, usually an SGA, that is associated with less EPS.

If the causes of secondary negative symptoms have been ruled out, the patient is likely to be demonstrating enduring primary negative symptoms. There is no strong evidence that any drug for psychosis is better for primary negative symptoms. Although some studies suggested that SGAs are more effective in reducing negative symptoms than FGAs, the advantage may be only in reducing secondary negative symptoms due to EPS. Meta-analyses have shown no significant advantages of SGAs over haloperidol in the reduction of negative symptoms. However, until this issue is decided by adequate controlled studies, it is probably reasonable for clinicians to consider changing to an SGA for patients with substantial negative symptoms in an attempt to minimize any that are secondary to EPS. Based on one 6-month study, cariprazine 4.5 mg daily was superior to risperidone 4 mg daily in reducing negative symptoms

Patients with schizophrenia frequently suffer from impairments in memory, attention, and information processing, interfering with social and vocational rehabilitation. As with negative symptoms, cognitive impairments can be secondary to other causes, including depression, substance abuse, or side effects, such as sedation or EPS. Additionally, the anticholinergic effects of either a drug for psychosis or an antiparkinson medication such as biperiden (Akineton) or benztropine can cause cognitive impairments. Decreasing the use of anticholinergic medication by changing to drugs that do not require antiparkinson medications, particularly SGAs, may be helpful. However, cognition was carefully monitored in the CATIE study and did not differ meaningfully among the FGA, perphenazine, and the tested SGAs.

SUICIDAL BEHAVIOR

Patients with schizophrenia and schizoaffective disorder are at risk for suicidal behavior. Approximately 20% to 40% of patients make suicide attempts, and 5% to 10% succeed. Suicidal behavior appears to be an independent domain from psychosis. However, depression and comorbid substance abuse increase the risk. Based on a large-scale study of clozapine (vs. olanzapine) in patients at risk for suicide, clozapine received an indication in the United States for the prevention of suicidal behavior. Clozapine should be considered for unremitting suicidal thoughts.

STRATEGIES FOR POOR RESPONDERS

Approximately 40% to 50% of acute psychoses will experience a remission of positive symptoms or experience only mild symptoms with treatment. The remaining 50% to 60% of patients will improve but will still demonstrate positive symptoms. Rather than categorizing patients into responders/nonresponders, consider the degree to which the illness is improved by medication. Some patients require chronic institutionalization. Others will respond to a medication with substantial improvement but still demonstrate persistent positive symptoms, such as hallucinations or delusions.

Before deciding that a particular medication is not working ensure that the diagnosis is correct, rule out substance-related or medical reasons, ensure that the patient has been adherent to the medication, and that they received an adequate trial. A 6-week duration on an adequate dose of a drug for psychosis represents a reasonable trial. If patients demonstrate even a mild amount of improvement during this period, it may be reasonable to wait before changing a medication, since some may improve at a steady rate for 3 to 6 months. However, early lack of at least minimal improvement (i.e., within the first 2 weeks) seems to identify a subgroup of patients who achieve a lower level of response even after 3 months of subsequent treatment. It may also be helpful to confirm that the patient is receiving an adequate dose by monitoring plasma concentration (if the patient is receiving a drug with an established therapeutic level) or to establish nonadherence. This information is available for a number of drugs, including clozapine, fluphenazine, haloperidol, perphenazine, risperidone, and trifluoperazine (Stelazine). A very low plasma concentration may indicate that the patient has been partially or noncompliant. It may also suggest that the patient is an ultrarapid metabolizer or the drug is not being adequately absorbed (e.g., if not taken with food in case of ziprasidone and lurasidone). Polymorphisms of the cytochrome-P450 enzyme system can be assessed to identify atypical metabolizer profiles. Under these conditions, raising the dose may be helpful. If the level is relatively high, clinicians should consider whether side effects may be interfering with therapeutic response.

With a poor response, many clinicians will consider raising the dose above the usual therapeutic level; however, nearly all studies found that doses above the labeled threshold were not associated with greater improvement. This suggests that changing to another drug is more likely to be helpful than changing to a high dose.

If a patient has responded poorly to an FGA, it is unlikely that this individual will do well on another FGA. Studies suggest that a poor response to one FGA is likely to be followed by a poor response to another. This has led many clinicians to change to an SGA in the hope that a somewhat different profile of receptor interactions may lead to a better response. However, this is often not the case.

Clozapine can be effective for patients who respond poorly to other drugs for psychosis. Important evidence of clozapine's advantages in this population comes from a multicenter trial reported by Kane et al. in 1988 in which clozapine was compared with chlorpromazine. Patients in this study were a severely psychotic group of individuals who had failed trials with at least three drugs for psychosis. Clozapine was significantly more effective than chlorpromazine in nearly every dimension of psychopathology. Thirty percent of clozapine patients met improvement criteria at 6 weeks compared to 4% of patients on chlorpromazine. As many as 60% of patients are likely to meet these same improvement criteria when maintained on therapeutic doses of clozapine for 6 months.

There is also evidence suggesting that risperidone and olanzapine may be helpful when an FGA is only partially effective. At least one controlled trial with each of these drugs has shown superiority to FGAs in poor or partial responders or "noninferiority" to clozapine. However, other studies suggest clozapine to be superior. The CATIE and CUtLASS trials also confirmed clozapine's superiority when FGAs or SGAs have failed. A number of meta-analyses have been conducted supporting clozapine's superiority in patients with refractory symptoms. Given clozapine's side effects profile, a case can be made for first-trying patients on risperidone or olanzapine when they have responded poorly to an FGA. If patients fail to respond adequately, a trial of clozapine is warranted. Since partial nonadherence is a common cause of treatment "resistance," a trial with an LAI should be considered before switching to clozapine.

There has been an increase in the use of polypharmacy. Although there are a variety of factors contributing to such practice, a poor initial response to monotherapy is often cited as a reason, but patients also end up on two or more drugs for psychosis if they start improving during a switch. There are very little controlled data supporting combining drugs for psychosis for treatment-resistant patients, except possibly for augmenting a dopamine antagonist with a partial D2 agonist for negative symptoms. However, in a nationwide database study from Finland, eight of the 10 most effective strategies for preventing hospitalization involved the use of two drugs for psychosis. These results could be biased by keeping patients on two drugs preferentially who improve during a switch, and most superior combinations were with clozapine or an LAI.

Other "augmentation" strategies for positive symptoms with benzodiazepines, lithium, antiepileptics, antidepressants, etc., have been studied to some extent, but controlled trials and meta-analyses are generally negative or biased toward small and poor-quality studies. These findings do not rule out the possibility of individual benefit for some patients; however, the data supporting the potential value of clozapine are far more consistent. The use of adjunctive medications to treat symptoms of affective disturbance, anxiety, agitation, etc., may be more appropriate.

ECT is also a treatment of last resort for refractory individuals, even those not adequately responding to clozapine.

MAINTENANCE THERAPY

During the maintenance phase, patients are often in a relative state of remission with only minimal positive psychotic symptoms or at a stable plateau with moderate symptoms. The goals are to prevent psychotic relapse or exacerbation and to improve functioning. Medications are important in preventing/delaying relapse and helping to manage functional impairments. The art of maintenance treatment is complicated by medication side effects, lack of patient insight, cognitive or organizational abilities, and varied motivation to continue long-term treatment. Titrating dosage in maintenance treatment is difficult because the relapse/exacerbation that can result from too low a dose is not likely to occur for weeks to months. Recent meta-analyses suggest that maintenance doses should neither be higher nor lower than the dose range for acute treatment.

Drug and Route of Administration for Maintenance Therapy

Those maintained on a drug for psychosis have a much lower relapse rate than patients who have their medications discontinued. Pooling large amounts of data suggests that 15% to 25% in a year will experience a relapse while receiving medications and 50% to 75% will relapse without medications. Clinicians are often tempted to discontinue medications in patients who have been well and stable for prolonged periods of time.

Unfortunately, these patients also have high relapse rates when their medications are discontinued. Other evidence indicates that patients who experience relapses while they are receiving a drug for psychosis have milder episodes than patients who relapse off medication. It has been suggested that patients who had their medications discontinued are more likely to show dangerous behavior and/or be admitted involuntarily.

TD has been a major concern in establishing the benefit-to-risk ratio of long-term treatment. Even with FGAs, experts agreed that the benefits of maintenance treatment outweighed the risks. As SGAs have been introduced with a substantively reduced risk of TD, one of the major concerns about long-term treatment has been diminished.

The evaluation of benefit-to-risk has been particularly challenging in those patients who have experienced only one psychotic episode and have responded well to treatment. It is important to recognize that although many such patients achieve remission, recovery (which includes sustained, relatively normal social and vocational adjustment) is far less common.

It has been well established that even patients who have had only one episode have at least a four in five chance of relapsing at least once over the next 5 years, and that stopping medication is the most significant risk factor. At the same time, it is often difficult for late adolescents or young adults to accept the nature of their illness and the need for ongoing pharmacotherapy. This represents a psychotherapeutic and psychoeducational challenge to clinicians and underscores the importance of combining psychosocial treatments, family therapy, vocational rehabilitation, and pharmacologic management.

Although published guidelines do not make definitive recommendations about the duration of maintenance treatment following the first episode, recent data suggest that 1 or 2 years might not be adequate. This is a particular concern when patients have achieved good employment status or are involved in educational programs, as they have an enormous amount to lose if they experience another psychotic episode. Furthermore, recent data indicate that all-cause mortality and mortality due to suicide and cardiovascular illness are reduced with continued versus discontinued medication treatment, particularly LAIs. The seeming paradox that medications that can increase cardiometabolic burden reduce cardiometabolic mortality has recently been elucidated, showing that in within-subject analyses, patients on drugs for psychosis were also more adherent to medications used to treat cardiovascular morbidity and prevent mortality.

It is generally recommended that multiepisode patients receive maintenance treatment for at least 5 years. However, this recommendation is based on controlled data until this time, and the implication that discontinuing medication at that point is not without substantial risk is a mistake. Therefore, many experts would recommend pharmacotherapy on an “indefinite” or “for the foreseeable future” basis.

The first 3 to 6 months following an acute episode or relapse is a period of particular vulnerability. With short lengths of hospital stay, adequate linkages with ambulatory programs are critical to ensure continuity of care. After stabilization for 6 months, gradual dosage reduction may be attempted, but its need is unclear if there are no limiting adverse effects that are dose-related, given data that doses below the minimally effective dose in an acute exacerbation significantly increase the risk of relapse. Given reduced concern regarding TD and the lack of a strong dose-response curve for weight gain, one could argue that there is less incentive to search for the lowest effective dose for maintenance treatment with SGAs.

There are some data, including a large meta-analysis, suggesting that the SGAs are more effective in preventing relapse than FGAs. Our impression at present is that this advantage is not due to enhanced compliance. Although there are data suggesting some improvement in adherence with SGAs, these differences are modest at best.

Nonadherence rates with long-term drug treatment are very high. Average estimates suggest that more than 40% to 50% of patients become at least partially noncompliant within 1 or 2 years. Given the high rates of relapse following discontinuation of medication and the potentially severe consequences (loss of job, interference with school, family burden, suicidality, homelessness, aggressive or violent behavior), attempts to enhance adherence are critical. Although psychosocial treatments focusing on adherence can be helpful, the use of LAIs should be considered a preventive measure and not just reserved for patients who have repeatedly experienced nonadherence and subsequent relapse.

Given these data, LAIs are a valuable treatment option that appears to be underutilized. Recently, the superiority of LAIs over oral drugs for psychosis with regards to hospitalization or relapse has been demonstrated across three research designs, including randomized controlled trials, large naturalistic cohort studies, and mirror image studies where patients serve as their own control.

There are a number of potential advantages of long-acting injectable medication. First, data show that nonadherence is less likely with LAIs. Second, clinicians immediately know when nonadherence occurs and have time to initiate appropriate interventions before the blood levels become nontherapeutic. Third, there is less day-to-day variability in blood levels, and peak concentrations are lower, potentially making it easier to establish minimum effective doses. Finally, many patients who have had experience with such treatment often prefer it, and most may agree to try an LAI when offered appropriately, for example, in a motivational interviewing shared decision-making framework.

INTEGRATING PHARMACOTHERAPY AND PSYCHOSOCIAL TREATMENT

Most patients with schizophrenia will benefit from a combination of pharmacotherapy and psychosocial treatments. This effect is particularly powerful in the case of coordinated specialty care. Recent improvements in both pharmacotherapy and psychosocial domains suggest that the overall outcome of this disorder can be improved if patients receive the optimal forms of both treatments at the appropriate stage of their illness. Psychosocial treatments are probably the most effective when patients have recovered from severe psychotic episodes. During the acute psychotic phase, clinical management should emphasize maintaining cooperativeness and trust. This focus is particularly important when there is overt suspiciousness or a tendency to misinterpret the intentions of the treatment team. A successful strategy is likely to include psychoeducation, including clear explanations of the rationale for treatment and possible drug side effects. Since family members may be important allies in ensuring cooperativeness, family psychoeducation programs are helpful during this phase.

It is difficult to generalize about the interactions of drugs for psychosis and psychosocial treatments for stable patients since psychosocial treatments can be quite different in their content and their goals. Nevertheless, it seems that effective pharmacotherapy benefits the efficacy of psychosocial treatments, while psychosocial interventions can benefit pharmacotherapy with improved adherence as one outcome.

Other studies have indicated that psychosocial treatments, particularly family treatment, may decrease the amount of stress that the patient experiences within the family and that this, in turn, decreases the amount of medication required by the patient and the risk of relapse. A recent meta-analysis of 20 psychological interventions targeting relapse prevention, given mainly adjunctively with pharmacotherapy for psychosis, found the following interventions to be superior to usual care at 1 year: family interventions, relapse prevention programs, cognitive behavioral therapy, family psychoeducation, coordinated specialty care, and patient psychoeducation.

DRUG DEVELOPMENT FOR SCHIZOPHRENIA

Studies, such as CATIE and CUtLASS, and meta-analyses indicate that both FGAs and SGAs have limits to their effectiveness. Although these agents are often effective in reducing the severity of positive symptoms, their effects on cognitive and negative symptoms are often inadequate. These domains are important since the long-term functional outcome—or the ability of patients to function in work, school, and social relationships—is related to the severity of these symptoms. Moreover, a significant subgroup of patients is partially or fully treatment-resistant regarding positive symptoms. Furthermore, the current availability of one single proven mechanism of action for the treatment of psychosis, postsynaptic dopamine receptor blockade, does not allow rational polypharmacy. These shortcomings have led to a search for medications that have a different mechanism of action and/or specific effects on residual positive, negative, and cognitive symptoms.

Ongoing activity has focused on identifying drugs for psychosis with mechanisms other than postsynaptic dopamine receptor blockade, hoping to improve the efficacy and tolerability of currently available agents but also to possibly facilitate treating patients who insufficiently benefit from available drugs. Recently, several novel mechanisms of action have been tested that show promise, including muscarinic autoreceptor activation, trace amine-associated receptor agonism, serotonin 2A antagonism/inverse agonism, and glycine transport reuptake inhibition. Whether any of these mechanisms with positive phase 2 data will lead to final approval remains to be seen.

Considerable attention has focused on cognitive impairments as a target for drug development in schizophrenia. This has included the development of, and agreement on, measurement batteries for cognitive dysfunction. The FDA and other agencies have clarified that to approve drugs for cognitive impairments associated with schizophrenia, the drug is required to separate from placebo both on a cognitive and a functional coprimary outcome measure. Currently, a few drugs addressing varied molecular targets are in different stages of development.

Another activity has focused on drugs that may improve negative symptoms of schizophrenia. Consensus on targets, designs, and measurement tools has prepared the ground for large development programs. However, like in the area of cognitive dysfunction, despite many attempts and promise in the area of proglutamatergic mechanisms, to date, no drugs have been approved for negative symptoms, though a number of agents are currently in clinical trials.

If studies of specific drugs for negative symptoms and cognitive impairments are successful, future treatment of schizophrenia may involve a drug for positive symptoms and comedications for other symptom domains, leading to improved functional outcomes. Whether any of these or other drug mechanisms will also enhance efficacy for positive symptoms remains to be seen, but additional mechanisms of action for the treatment of all domains of schizophrenia are highly desirable.

FURTHER READINGS

Bighelli I, Rodolico A, García-Mieres H, et al. Psychosocial and psychological interventions for relapse prevention in schizophrenia: a systematic review and network meta-analysis. *Lancet Psychiatry*. 2021;8(11):969–980.

Buchanan RW, Kreyenbuhl J, Kelly DL, et al. The 2009 schizophrenia PORT psychopharmacological treatment recommendations and summary statements. *Schizophr Bull*. 2010;36(1):71–93.

Carbon M, Kane JM, Leucht S, Correll CU. Tardive dyskinesia risk with first-and second-generation antipsychotics in comparative randomized controlled trials: a meta-analysis. *World Psychiatry*. 2018;17(3):330–340.

Ceraso A, Lin JJ, Schneider-Thoma J, et al. Maintenance treatment with antipsychotic drugs for schizophrenia. *Cochrane Database Syst Rev*. 2020;8(8):CD008016.

Correll CU, Galling B, Pawar A, et al. Comparison of early intervention services vs treatment as usual for early-phase psychosis: a systematic review, meta-analysis, and meta-regression. *JAMA Psychiatry.* 2018;75(6):555–565.

Correll CU, Rubio JM, Inczedy-Farkas G, et al. Efficacy of 42 pharmacologic cotreatment strategies added to antipsychotic monotherapy in schizophrenia: systematic overview and quality appraisal of the meta-analytic evidence. *JAMA Psychiatry*. 2017;74(7):675–684.

Correll CU, Rubio JM, Kane JM. What is the risk-benefit ratio of long-term antipsychotic treatment in people with schizophrenia? *World Psychiatry*. 2018;17(2):149–160.

Firth J, Siddiqi N, Koyanagi A, et al. The Lancet Psychiatry Commission: a blueprint for protecting physical health in people with mental illness. *Lancet Psychiatry*. 2019;6(8):675–712.

Galling B, Roldán A, Hagi K, et al. Antipsychotic augmentation vs. monotherapy in schizophrenia: systematic review, meta-analysis and meta-regression analysis. *World Psychiatry*. 2017;16(1):77–89.

Guinart D, Taipale H, Rubio JM, et al. Risk factors, incidence, and outcomes of neuroleptic malignant syndrome on long-acting injectable vs oral antipsychotics in a nationwide schizophrenia cohort. *Schizophr Bull.* 2021;47(6):1621–1630.

Højlund M, Kemp AF, Haddad PM, Neill JC, Correll CU. Standard versus reduced dose of antipsychotics for relapse prevention in multiepisode schizophrenia: a systematic review and meta-analysis of randomised controlled trials. *Lancet Psychiatry*. 2021;8(6):471–486.

Howes OD, McCutcheon R, Agid O, et al. Treatment-resistant schizophrenia: treatment response and resistance in psychosis (TRRIP) working group consensus guidelines on diagnosis and terminology. *Am J Psychiatry*. 2017;174(3):216–229.

Howes OD, Whitehurst T, Shatalina E, et al. The clinical significance of duration of untreated psychosis: an umbrella review and random-effects meta-analysis. *World Psychiatry*. 2021;20(1):75–95.

Huhn M, Nikolakopoulou A, Schneider-Thoma J, et al. Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multiepisode schizophrenia: a systematic review and network meta-analysis. *Lancet*. 2019;394(10202):939–951.

Jones PB, Barnes TRE, Davies L, et al. Randomized controlled trial of the effect on quality of life of second- vs first-generation antipsychotic drugs in schizophrenia: cost utility of the latest antipsychotic drugs in schizophrenia study (CUtLASS 1). *Arch Gen Psychiatry*. 2006;63(10):1079–1087.

Kahn RS, Fleischhacker WW, Boter H, et al. Effectiveness of antipsychotic drugs in first-episode schizophrenia and schizophreniform disorder: an open randomized clinical trial. *Lancet*. 2008;371(9618):1085–1097.

Kane J, Kishimoto K, Rubio J, Correll CU. Evaluation and management of treatment-resistant schizophrenia. In: UpToDate, Post TW, ed., *UpToDate* uptodate.com. Accessed on September 28, 2023.

Kane JM, Correll CU. Past and present progress in the pharmacologic treatment of schizophrenia. *J Clin Psychiatry*. 2010;79(9):1115–1124.

Keepers GA, Fochtmann LJ, Anzia JM, et al. The American Psychiatric Association Practice Guideline for the treatment of patients with schizophrenia. *Am J Psychiatry.* 2020;177(9):868–872.

Kishimoto T, Hagi K, Kurokawa S, Kane JM, Correll CU. Long-acting injectable versus oral antipsychotics for the maintenance treatment of schizophrenia: a systematic review and comparative meta-analysis of randomised, cohort, and pre-post studies. *Lancet Psychiatry*. 2021;8(5):387–404.

Krause M, Zhu Y, Huhn M, et al. Antipsychotic drugs for patients with schizophrenia and predominant or prominent negative symptoms: a systematic review and meta-analysis. *Eur Arch Psychiatry Clin Neurosci*. 2018;268(7):625–639.

Leucht S, Bauer S, Siafis S, et al. Examination of dosing of antipsychotic drugs for relapse prevention in patients with stable schizophrenia: a meta-analysis. *JAMA Psychiatry*. 2021;78(11):1238–1248.

Lewis SW, Barnes TRE, Davies L, et al. Randomized controlled trial of effect of prescription of clozapine versus other second-generation antipsychotic drugs in resistant schizophrenia. *Schizophr Bull*. 2006;32(4):715–723.

Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med*. 2005;353(12):1209–1223.

McEvoy JP, Lieberman JA, Stroup TS, et al. Effectiveness of clozapine versus olanzapine, quetiapine, and risperidone in patients with chronic schizophrenia who did not respond to prior atypical antipsychotic treatment. *Am J Psychiatry*. 2006;163(4):600–610.

Millan MJ, Agid Y, Brüne M, et al. Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nat Rev Drug Discov*. 2012;11(2):141–168.

Millan MJ, Fone K, Steckler T, Horan WP. Negative symptoms of schizophrenia: clinical characteristics, pathophysiological substrates, experimental models and prospects for improved treatment. *Eur Neuropsychopharmacol*. 2014;24(5):645–692.

Nielsen J, Correll CU, Manu P, Kane JM. Termination of clozapine treatment due to medical reasons: when is it warranted and how can it be avoided? *J Clin Psychiatry*. 2013;74(6):603–613.

Samara MT, Dold M, Gianatsi M, et al. Efficacy, acceptability, and tolerability of antipsychotics in treatment-resistant schizophrenia: a network meta-analysis. *JAMA Psychiatry*. 2016;73(3):199–210.

Samara MT, Leucht C, Leeflang MM, et al. Early improvement as a predictor of later response to antipsychotics in schizophrenia: a diagnostic test review. *Am J Psychiatry*. 2015;172(7):617–629.

Schoretsanitis G, Kane JM, Correll CU, et al. Blood levels to optimize antipsychotic treatment in clinical practice: a joint consensus statement of the American Society of Clinical Psychopharmacology and the Therapeutic Drug Monitoring Task Force of the Arbeitsgemeinschaft Für Neuropsychopharmakologie Und Pharmakopsychiatrie. *J Clin Psychiatry*. 2020;81(3):19cs13169.

Solmi M, Tiihonen J, Lähteenvuo M, et al. Antipsychotics use is associated with greater adherence to cardiometabolic medications in patients with schizophrenia: results from a nationwide, within-subject design study. *Schizophr Bull*. 2021;48(1):166–175.

Taipale H, Tanskanen A, Mehtälä J, et al. 20-year follow-up study of physical morbidity and mortality in relationship to antipsychotic treatment in a nationwide cohort of 62,250 patients with schizophrenia (FIN20). *World Psychiatry*. 2020;19(1):61–68.

Tiihonen J, Taipale H, Mehtälä J, et al. Association of antipsychotic polypharmacy vs monotherapy with psychiatric rehospitalization among adults with schizophrenia. *JAMA Psychiatry*. 2019;76(5):499–507.

Vancampfort D, Firth J, Correll CU, et al. The impact of pharmacological and non-pharmacological interventions to improve physical health outcomes in people with schizophrenia: a meta-review of meta-analyses of randomized controlled trials. *Focus (Am Psychiatr Publ).* 2021;19(1):116–128.

12.13 Psychiatric Rehabilitation

ROBERT E. DRAKE, M.D., PH.D., GARY R. BOND, PH.D., AND DROR BEN-ZEEV, PH.D.

Approximately 5% of adults in the United States experience disability due to a serious psychiatric illness such as schizophrenia. Psychiatric rehabilitation denotes a wide range of interventions to help people with disabilities due to mental illness improve their functioning, quality of life, and recovery by enhancing the opportunities, skills, and supports they need to succeed in regular adult roles and in the environments of their choice. Regular adult roles include living independently, attending school, being employed in a competitive job, having positive relationships with family, having friends, pursuing leisure activities, and having intimate relationships. Psychiatric rehabilitation emphasizes relative independence, self-management of illness, self-efficacy, and community integration, including avoidance of dependence on professionals and segregated settings.

Psychiatric rehabilitation has become a fundamental component of the treatment for schizophrenia, consistent with current knowledge regarding the etiology, pathophysiology, manifestations, treatment, and course of the disorder. Current evidence, reviewed in other chapters, indicates that schizophrenia is almost certainly a heterogeneous set of disorders causally related to genetic factors, prenatal and perinatal insults, and postnatal stress. Although the definitive neurobiologic substrate of the disorder remains uncertain, many patients manifest mild neurodevelopmental abnormalities related to attention, cognition, social function, and motor function in childhood, long before the onset of overt psychotic illness. Research on schizophrenia shows that many but not all patients exhibit abnormalities of neural structure, circuitry, and function. Clinically, the disorder is characterized by positive symptoms of psychosis, such as hallucinations and delusions; negative symptoms, such as lack of motivation and poverty of speech; cognitive deficits, such as problems with memory and executive functioning; and psychosocial difficulties, such as poor social role performance, chronic unemployment, high rates of substance abuse, and increased risk for homelessness, victimization, and other problems. Long-term follow-up studies have shown over many years that about 10% to 15% of schizophrenia patients recover completely, and most others improve significantly (Bleuler termed the usual course “healing with scarring”), but some continue to experience symptoms, cognitive difficulties, and psychosocial problems for decades.

Antipsychotic medications often reduce positive symptoms of psychosis, but they have limited impact on negative symptoms, cognitive impairment, and psychosocial functioning. For the great majority of patients, medications do not restore premorbid levels of functioning and do not produce normative role performance. Some 20% or more of schizophrenia patients have psychotic symptoms that do not respond to antipsychotic medications, and many experience only a partial symptomatic response. Further, medications do not reverse the consequences of stigma, impaired learning, failure to master adult developmental tasks, and social withdrawal. Therefore, maximizing functional recovery through psychiatric rehabilitation is critical for most patients.

Current interventions for schizophrenia are multifaceted. Historically, rehabilitation specialists distinguished rehabilitation from treatment by its focus on improving functional performance rather than controlling symptoms of illness. Rehabilitation specialists also argued that they focused on the patient’s strengths rather than deficits. In practice, however, these distinctions are not valid. Many cognitive behavioral, skills training, technology, and supportive interventions aim to improve the management of symptoms as well as functional status, and most rehabilitation specialists take strengths and deficits into account. In this book, several interventions for schizophrenia could be classified as treatment or rehabilitation, and the authors have divided up a selection of current interventions for the sake of illustration. Readers should consult both chapters.

This chapter illustrates several approaches to psychiatric rehabilitation with specific models and examples. Presenting the entire field is beyond our scope, but interested readers should consult textbooks listed at the end of the chapter.

HISTORY

Current approaches to psychiatric rehabilitation represent the confluence of several historical traditions. In the 19th century, the reform movement called moral treatment emphasized putting people with mental illness in benign environments and normal roles, though usually in institutions. Moral treatment recognized that participation in education, work, social activities, and other normal roles could have a healthful effect.

In the United States, government programs for the employment of people with disabilities began after World War I for people with physical disabilities and were subsequently extended to those with psychiatric disabilities in 1943. These early programs transferred ideas from physical rehabilitation to psychiatric rehabilitation and emphasized employment.

Deinstitutionalization and the development of community mental health centers in the 1960s and 1970s not only transferred the locus of care for people with schizophrenia to the community but also promoted interest in making services accessible, comprehensive, and focused on helping people to function in regular environments and roles. With leadership from the National Institute of Mental Health and the Center for Mental Health Services of the Substance Abuse and Mental Health Services Administration, the community mental health ideology evolved into the community support system model, which included approaches such as case management, assertive outreach, coordination, integration, continuity, advocacy, self-help, supported employment, supported education, supported housing, self-management, and an emphasis on personal goals. The community support system model enlarged the role of community mental health and rehabilitation beyond reversing the disabling effects of institutionalization to include the prevention of disability among young adults with severe psychiatric illness who had minimal experience in hospitals.

In the early years of deinstitutionalization, patients and nonprofessionals came together to form psychosocial clubs, such as Fountain House in New York and Thresholds in Chicago, to promote mutual aid and self-help. The early self-help clubs transformed into a range of psychosocial rehabilitation centers around the country. Despite great diversity, these centers consistently emphasized mutual support, strategies for coping, health promotion rather than a focus on illness, normative roles such as worker, and belief in the potential for productivity.

Current ideas regarding psychiatric rehabilitation emerged directly from these activities.

Developments in the fields of skills training and environmental modification have also stimulated modern approaches to rehabilitation. Skills training methods derived from social learning principles have been applied to numerous areas of everyday functioning: for example, daily living tasks, social activities, employment, education, housing, relating to the health care system, and managing one's illnesses. Approaches to making environments more accommodating are considerably more diverse. They range from collaborative work with families, landlords, and employers, to a wide range of strategies used in the community to remove barriers faced by people with mental illness related to societal stigma, legal constraints, economic disincentives, and rigidities in the health care system.

The consumer-initiated movement called "recovery" became an influential ideology for advocates and mental health care systems during the 1990s. Recovery has numerous definitions, but the ideology includes hope, self-management, pursuing personal goals, finding meaning in life, and participating fully in the community. Since rehabilitation has always been patient-centered, recovery goals for most patients are isomorphic with rehabilitation goals.

Most recently, health information technology has been expanding rapidly to enhance self-management and rehabilitation goals. Web-based and mobile technology tools give people with mental disorders and their families immediate access to the latest information, local resources, research, and rehabilitation tools. Increasingly, these tools are shifting control of decision-making and management from professionals to patients and families.

OVERVIEW OF TECHNIQUES

Psychiatric rehabilitation uses three basic approaches: creating opportunities, providing supports, and increasing skills. In practice, these three approaches are combined in nearly all models of psychiatric rehabilitation.

People with mental disorders have been discriminated against, stigmatized, and treated unfairly in nearly every society, even compared to people with other disabilities. In the past, long-term institutionalization was the primary mechanism; today segregation and lack of opportunity persist in the community. Sensationalism in the media constantly reinforces stigma. Psychiatrists and other professionals often convey stigmatizing attitudes and beliefs. Psychiatric rehabilitation aims to increase opportunities for normative housing, education, employment, socialization, leisure, and other activities that represent full citizenship under the rubric of community integration. To increase opportunities, rehabilitation specialists use legal mechanisms, regulations, community action, social networks, family supports, and personal relationships.

The second basic approach to psychiatric rehabilitation is to provide intensive supports. People with mental illness often need extra supports to succeed initially in functional roles. Once achieving success in a specific setting, such as school or work, supports are typically withdrawn substantially or completely, as the experience itself increases self-efficacy, functional skills, and natural supports.

Increasing personal skills for illness management and social functioning is the third basic technique of psychiatric rehabilitation. Due to cognitive, experiential, and neurologic difficulties, people with schizophrenia often need to enhance their abilities to manage their internal experiences, social environments, and roles they are pursuing in the community. Skills training has been a core feature of rehabilitation for decades. As we illustrate below, more of the learning is now shifting from professionals to technology tools.

We illustrate how opportunity, support, and skills combine by describing three current approaches: Housing First, Individual Placement and Support (IPS), and FOCUS. These specific examples cover only a small part of psychiatric rehabilitation but do demonstrate the enormous potential to improve life for people disabled by schizophrenia.

SUPPORTIVE HOUSING

Homelessness and Housing

Mental health researchers have often described homelessness as a psychiatric problem, suggesting that people cycle through homelessness and hospitalizations because of severe mental illness, poor treatment, or a combination of both. By contrast, homelessness researchers from other backgrounds, such as anthropology, epidemiology, and housing, consider homelessness to be primarily due to a lack of affordable and appropriate housing. Research shows that people with the most severe mental disorders and social disadvantage, for example, schizophrenia complicated by poverty and drug addiction, need safe housing, supports, and treatment.

The more difficult questions relate to the type of housing and connection to services: Do some people need to be in supervised, congregate housing, or is that a prejudice of mental health professionals from the deinstitutionalization era? If many people want independent housing, do they need to be stable and engaged in treatment first? Alternatively, can they succeed in independent housing before they are engaged and committed to treatment? Policymakers, researchers, clinicians, patients, and advocates have debated these issues for more than four decades. Meanwhile, homelessness persists and has worsened in many areas of the United States.

Housing Models

Traditional approaches to housing people with schizophrenia and serious mental illness used segregation, congregate living, and staff within the housing or linkage with required clinical services. Segregated housing included nursing homes, group homes, family care homes, boarding homes, and housing units within the same building and closely overseen by mental health workers. All of these models assumed that people with serious mental disorders were not competent to live on their own and needed constant supervision, which was most easily provided by grouping them together. As an alternative, people were placed in independent apartments, often owned by the mental health system, because they agreed to participate in mental health treatment.

In the early 1980s, research clinicians developed assertive community treatment (ACT) for the express purpose of helping people to leave the hospital and maintain tenure in the community. ACT initially involved moving a multidisciplinary team out of the hospital with the patients to provide comprehensive services 24 hours a day, 7 days a week. The team helped people to find housing and develop independent living skills, usually by providing intensive supports in the community. For example, care managers in the community would help people learn to shop in their neighborhoods, use public transportation, prepare meals in their apartments, take their medications, and find activities and friends. Learning these skills in natural environments rather than in treatment centers was more effective and obviated the need to transfer skills from the clinic to the community.

Encouraged by the success of ACT and as a reaction against the paternalistic psychiatric practices common in that era, advocates for community integration developed a set of housing principles that they labeled “supported housing.” Core principles included normalized permanent housing (regular tenancy agreements, assurance of privacy, location in integrated communities), client choice in type and location of residence, assistance in finding affordable housing, individualized and flexible support, separation of housing, and mental health services. Research showed that most people preferred living in their own apartments or homes rather than in supervised congregate housing and were able to live in scattered-site apartments.

The next question was readiness. Did people with mental illness need to be stable and participating in mental health care before finding an independent apartment? Studies of a supported housing model called Housing First answered this question. A systematic approach to reversing homelessness, Housing First combines two primary interventions: offering homeless people affordable housing of their choice as rapidly as possible and providing supportive services to help them succeed in housing. To ensure that residents can afford housing, some Housing First programs provide housing vouchers or subsidies. The approach does not require that people participate in treatment or any mental health program as a precondition for obtaining housing. Instead, Housing First assumes that people need housing as a foundation from which they can pursue other goals.

Typically, Housing First integrates housing assistance with ACT, intensive case management, or some other team-based approach to comprehensive services. The same team assists people in finding affordable housing and helps them learn to live independently and access other services when they have additional needs. Services are not mandatory, coercive, or chosen by professionals: tenants learn about services and make informed decisions, using a shared decision-making model.

James was a 50-year-old man who had been in and out of hospitals, homeless shelters, jails, and mental health centers for over 20 years. He often lived on the streets. Delusions, auditory hallucinations, and paranoia, all responding minimally to medications, accompanied him constantly. Regular marijuana use exacerbated his symptoms and got him entangled with drug dealers and the criminal justice system. James dressed eccentrically, wore a patch over one eye, and preached his religious beliefs in public settings, which occasioned rejections, confrontations with police, and trips to the emergency room. He seemed not to care about housing, failing to stay in several nursing homes, group homes, apartments, and shelters for more than a few weeks.

After a lengthy forensic hospitalization, James came to a Housing First team. Surprised that he was offered an apartment without preconditions, he readily agreed to move into his own place. The team helped him with furnishings, his landlord, shopping, and learning the new neighborhood.

James experienced many challenges, including disagreements with his landlord, restaurant managers, other tenants, drug dealers, and the police. But the team was available to intercede 24 hours a day, 7 days a week, and helped him to resolve many difficult situations. Listening respectfully to delusions and helping with practical matters was the central strategy. For example, James needed assistance to keep other drug users from moving into his apartment. He understood that to keep the apartment, which he liked, he needed to allow the team and his landlord to set limits with these acquaintances. Over time, James found delusional rationales for giving up marijuana, restricting his preaching, having a girlfriend, attending church, and taking a smaller dose of medications. Several team members admired his courage and integrity, and he often called on them for help. He did not return to homelessness, jail, or the hospital over several years.

Research on Supportive Housing

The absence of operationally defined program models limited most of the early research on supported housing and other residential models. The inconsistent and interchangeable use of the labels supported housing, Housing First, and the current term, supportive housing, confounded the problem. Researchers have not always agreed on the core components of supportive housing, and some principles are value statements not easily translated into objective criteria.

Housing First defined more clearly the core model components, and consequently research has yielded a coherent set of replicated findings. Large-scale pragmatic trials in the United States and Canada have demonstrated that most people with serious mental disorders can maintain independent housing over the long term when they have safe, affordable housing, and adequate psychosocial supports. Housing First studies have also shown that offering housing to people with substance use disorders, again without preconditions, does not lead to the dire consequences that paternalistic approaches assumed. Several studies have found that clients receiving permanent housing assistance are more satisfied with their housing and have reduced use of psychiatric hospitals and emergency rooms. Yet research has not demonstrated that improvement in housing outcomes generalizes to other areas of functioning or to clinical outcomes, such as substance use and psychiatric symptoms.

INDIVIDUAL PLACEMENT AND SUPPORT

History

Impairment of vocational role performance frequently complicates adjustment for patients with schizophrenia and other serious mental disorders, contributing to poverty, loss of self-efficacy, and other sequelae. Across the United States, less than 15% of patients are employed, even though 50% to 75% state that competitive employment is a primary goal. Due to patient interests and historical factors discussed earlier, vocational rehabilitation has always been a centerpiece of psychiatric rehabilitation.

Clinicians, rehabilitation specialists, researchers, and patients themselves have developed many approaches to help people with serious mental illness achieve employment over the last half-century. Most have involved train-and-place models, with extensive pre-employment assessments and preparatory experiences, such as counseling, day centers, sheltered workshops, transitional employment, and skills training. These approaches aimed to evaluate and improve the patient’s attitudes, skills, appearance, readiness, and other abilities presumed necessary to obtain and sustain competitive employment. Evaluations of these stepwise approaches to vocational rehabilitation found that they did not improve the rate of competitive employment; most patients became discouraged, disengaged, or stalled during the preparatory experiences.

Supported Employment

In the late 1980s, the emphasis of vocational rehabilitation began to shift to place-and-train models, termed supported employment, which began in the field of developmental disabilities. According to the federal definition, supported employment involves “competitive work in integrated work settings…consistent with the strengths, resources, priorities, concerns, abilities, capabilities, interests, and informed choice of the individuals, for individuals with the most significant disabilities for whom competitive employment has not traditionally occurred; or for whom competitive employment has been interrupted or intermittent as a result of a significant disability”. In contrast to other vocational approaches, supported employment programs (a) did not screen people for work readiness, but helped anyone who wanted to work; (b) did not provide intermediate work experiences, such as prevocational work units, transitional employment, or sheltered workshops; (c) actively facilitated job acquisition, often accompanying clients on interviews; and (d) provided ongoing support once the person was employed.

In the 1990s, a model of supported employment called IPS was developed specifically for people with severe mental illness appeared. IPS, also known as evidence-based supported employment, emphasizes the patient’s personal preferences for work, a rapid job search (without extensive pre-employment assessment or training), and individualized supports as needed. Employment specialists provide practical assistance at every stage of the employment process. Jobs are typically part-time, beginning at just a few hours per week and expanding over time. IPS emphasizes regular jobs in the competitive labor market, owned by the employee rather than the rehabilitation program, paid at or above minimum wage, and supervised by the employer.

Several features of IPS supported employment correlate with better employment outcomes and thus are considered evidence-based principles.

1. Zero Exclusion. Supported employment services are available for anyone who wants to work, regardless of symptoms, cognitive

impairments, work history, substance use, or other personal characteristics. Research shows that professionals cannot reliably identify who will succeed or fail in employment.

2. Competitive Employment. The goal of supported employment is competitive employment in work settings integrated into a community's economy. The great majority of clients who seek work want competitive employment, particularly if they are encouraged to believe that they can work. Few clients want sheltered or volunteer jobs.
3. Rapid Job Search. Clients seek employment directly rather than following lengthy pre-employment assessments, training, or practice. Rapid job search means applying for a job within 30 days and typically results in finding the first job within 4 months.
4. Job Development. The employment specialist conducts targeted job development, as opposed to locating any job, to help clients find jobs that match their skills and preferences. Job development often involves the entire team identifying or pursuing relationships with employers.
5. Integration of Rehabilitation and Mental Health Treatment. Supported employment should be an integral component of mental health treatment, not a separate service. Employment specialists are integrated into case management teams, ACT teams, or other multidisciplinary teams. Employment specialists meet regularly with their teams and collaborate on helping their clients to succeed in jobs.
6. Client Preferences. Services are based on clients' preferences and choices. Clients report higher satisfaction and longer job tenure when the jobs are consistent with their preferences.
7. Ongoing Support. Follow-along supports align with the individual client's needs and continue until the employee is independent. The multidisciplinary team provides a range of supports over time, with greater intensity as needed during times of crisis. For example, the psychiatrist may alter medications to help with the initial anxiety of working; the case manager may meet with the family to help them be supportive; and the employment specialist may provide benefits counseling, off-site supports, and skills training.
8. Individualized Benefits Counseling. Many clients who receive disability, welfare, or other benefits fear they could lose these benefits if they work. Individualized, accurate information about benefits often corrects misconceptions and helps clients make good decisions.

Employment specialists typically are college graduates supervised by master's-level vocational specialists. Employment specialists provide vocational assessment, placement, and support, but the multidisciplinary team collaborates throughout. The case manager and medical providers need to encourage clients to try employment. The entire team may help in finding a job and providing supports. For example, the case manager may help solve problems related to interpersonal difficulties on the job, the social worker may meet with the family to understand their concerns about employment, and the psychiatrist may provide extra medications to help with initial anxiety on the job or decrease medications to reduce side effects that interfere with working.

The key to success is finding the right match—a job that aligns with the client's interests, abilities, ways of coping with illness, needs regarding benefits, family situation, and so on. Trying different jobs often leads to an optimal job match.

Antonio was a 45-year-old man who had been a client of a mental health agency for over 10 years. He attended a rehabilitative day treatment program until it was converted to a supported employment program. His case manager encouraged him to think about the possibility of working part-time. Antonio told his case manager that he couldn't work because of schizophrenia and because he was helping to raise his two kids and needed to be home at 3 pm when they arrived home from school every day. The case manager explained to Antonio that going to work doesn't necessarily mean a 40-hour-a-week job and that lots of people in the agency's supported employment program work in part-time jobs, even jobs that are only a few hours a week.

Antonio agreed to meet one of the employment specialists to discuss the possibility of work. Over the next couple of weeks, the employment specialist met with Antonio several times, read his clinical record, and talked with his case manager and psychiatrist. The employment specialist learned that Antonio loved to drive his car, that he had attendance problems in past jobs, and that he was a kind, friendly, and likable person.

Antonio told the employment specialist that he was willing to do any job. He didn't have one specific job in mind. The employment specialist suggested a job at Meals on Wheels as a driver for lunch delivery. Antonio was hired and loved it right from the start. Absenteeism was never a problem because he knew that people were counting on him to deliver their meals. The hours were perfect (10 AM–2 PM), so he could be at home when his kids returned from school. He became good friends with the other people who worked there. He told his case manager that it was wonderful to be bringing home a paycheck again. And best of all, he said, was that his kids saw him going to work just like their friends' dads.

Research on IPS Supported Employment

Research has produced several consistent findings. Interventions that do not target job placement directly, such as housing, counseling, and medications, have very little impact on employment outcomes. In order to help patients achieve competitive employment, the intervention needs to be specifically directed at this goal.

Interventions that focus on pre-employment counseling, skills training, and practice work do not produce better vocational outcomes of any kind, including type, quality, amount, length, or satisfaction related to employment. Supported employment interventions that focus on direct and individualized placements with follow-along supports produce much better outcomes.

The evidence strongly favors IPS supported employment over other approaches to vocational rehabilitation. Nearly 40 randomized controlled trials show that IPS supported employment produces better employment outcomes, including higher rates of employment, greater earnings, more hours of work, and a longer tenure in jobs. The magnitude of these effects is substantial. For example, rates of employment are two or three times greater in IPS-supported employment programs than in the comparison conditions in the United States. Most of these trials are pragmatic trials conducted with typical patients, providers, and programs. Figure 12.13–1 summarizes competitive employment rates for IPS and control participants in 28 randomized trials.

All types of patients do better in IPS supported employment than in traditional vocational programs: for example, patients with poor vocational histories, those from minority backgrounds, specifically African Americans and Latinx people, those with co-occurring disorders, and inner-city residents who have recently been homeless. Research also shows that patients themselves express high satisfaction with IPS supported employment jobs and that they typically stay in first jobs for 5 to 6 months, just like the general population starting their first jobs. Several long-term follow-up studies indicate that patients maintain initial high employment rates over 5 to 10 years.

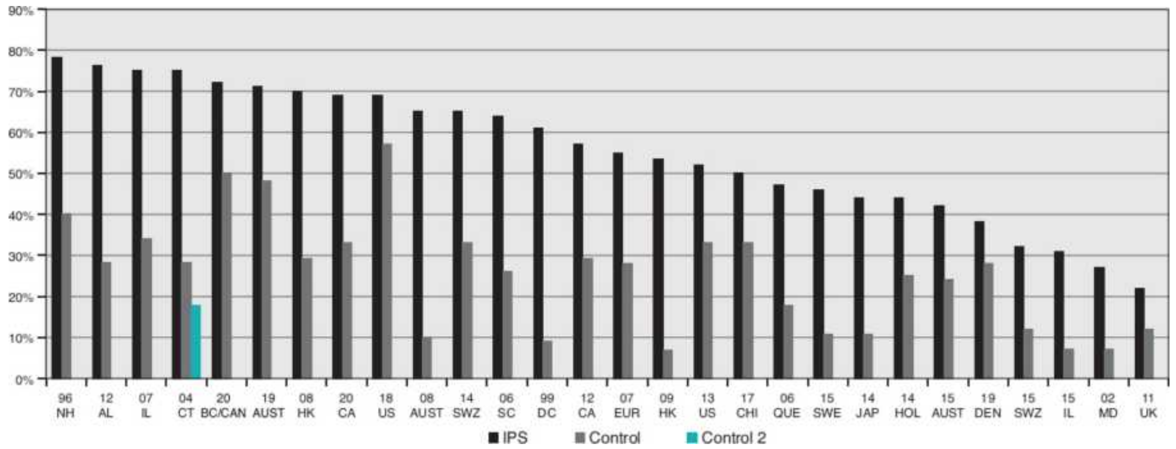

FIGURE 12.13–1. Competitive employment rates in 28 randomized controlled trials of individual placement and support.

Although some clinicians still fear that high expectations will destabilize patients with schizophrenia or other serious mental disorders, IPS supported employment services do not produce clinical instability, higher rates of relapse, or other untoward outcomes. On the contrary, patients who are currently in competitive employment appear to benefit in numerous ways and report higher satisfaction with services, better self-esteem, and greater quality of life.

TECHNOLOGY TOOLS

Technology Use Among People with Schizophrenia

People with psychiatric illnesses, including those with schizophrenia, use contemporary technologies in a manner similar to the general population. Survey studies suggest that 80% to 94% of people with schizophrenia access the internet regularly, use email, and use of social networking websites, online forums, chat rooms, and blogs. Moreover, people with serious mental illnesses such as schizophrenia are already using a host of digital resources for mental health reasons: obtaining information about mental disorders and treatments, searching for mental health services, connecting with peers with similar conditions, and engaging with online communities and supports. For example, people with psychiatric conditions are posting video testimonials in which they openly disclose their illness and provide first-person accounts of their symptoms and coping strategies on YouTube, a popular video-sharing website. Their posts receive numerous comments from viewers who describe their own experiences with illness, provide positive feedback, share symptom management suggestions, and thank the individual for sharing their video. In this manner, a virtual community can support people who may not have access to local support and help erode the stigma that often prevents people from disclosing their mental health struggles.

Mobile-cellular technology is pervasive—one of the fastest-adopted technologies in human history. The number of mobile phone subscriptions has gone from 1 billion in 2001 to over 7 billion in 2017, with over three-quarters of the world's mobile devices now being used in low- and middle-income countries. To contextualize the potential health care implications of such a ubiquitous resource, more people in the world now have access to mobile phones than to working toilets. When mobile phones first became commercially available at the end of the 20th century, they were inaccessible to people with significant disabilities and limited resources. But rapid advancements in the development of mobile hardware coupled with vast expansion in global mobile–cellular infrastructure have contributed to significant drops in the costs of devices, as well as access to data plans. Consequently, the majority of the population can make calls and send text messages using their personal mobile device with relative ease. In the United States, several federal and state programs now subsidize basic mobile phones and monthly call/text minutes to make them accessible to individuals below a certain income level. Thus, many more people with schizophrenia have access to a wealth of information and resources. Survey studies in the United States have shown that 86% to 97% of people with serious mental health conditions own and use mobile technologies, including cellular phones, text messaging devices for the hearing impaired, and smartphones (i.e., mobile phones with significant computational and data storage capacities). Similar access occurs in several European countries and Australia.

Both community-based outpatients and acutely ill inpatients with schizophrenia can use mobile devices successfully. Over two decades of clinical research using mobile devices, including beepers, personal digital assistants (PDAs), and smartphones to record behaviors, symptoms, and functioning, suggest that people with schizophrenia are quite open to using mobile self-tracking tools and that they experience little to no iatrogenic effects from doing so. Even highly symptomatic patients with paranoid schizophrenia have successfully engaged in Ecologic Momentary Assessment (EMA) studies that require them to self-report distressing thoughts and even illegal substance abuse using mobile devices. People with schizophrenia consistently report their interest in using technologic tools to support illness management and functional recovery. Research groups in different countries have developed innovative technologic approaches that aim to accomplish this goal.

Technologic Approaches for Assessment and Treatment

Web-Based Interventions. People with schizophrenia, their family members, and their supporters can access a range of resources on the internet, for example, illness management programs, chat forums, support groups, and informational webpages. Acceptability, usability, and helpfulness of web resources depend upon whether they were developed with close consideration of the needs of people with severe psychiatric illnesses (i.e., possible cognitive impairments, low literacy, and dexterity problems).

Web resources with appropriately designed user interfaces can be highly engaging and effective. For example, *Coping with Voices*, a self-directed web-based Cognitive Behavioral Intervention addresses the functional impact and distress caused by auditory hallucinations. Users progress through an interactive 10-lesson program at their own pace and revisit content from previous lessons. Video tutorials educate people about psychosis and dysfunctional cognitive styles using interactive games, quizzes, and practice opportunities. Participants who completed *Coping with Voices* experienced reductions in the severity and intensity of auditory hallucinations as well as other psychotic symptoms and general psychopathology.

Schizophrenia Online Access to Resources (SOAR) is a multifamily intervention for people with schizophrenia and their supports. The key elements of the *SOAR* intervention include the empathic engagement of users, education about illness and treatments, and the use of coping strategies. *SOAR* offers three therapy forums: one for people with schizophrenia, one for support persons, and one for both groups. Users can post questions and receive answers from trained clinicians within 24 to 48 hours. *SOAR* also contains a record of responses to previously posted questions, a library of relevant reading materials, and a list of community events, activities, and resources that may be of interest. People with schizophrenia who engage in *SOAR* experience a significant reduction in positive symptoms of psychosis. Support persons also showed a significant and sustained increase in knowledge about the course and prognosis of schizophrenia.

The *HORYZONS* system is an online platform developed for the long-term recovery of people experiencing a first episode of psychosis. The system delivers several evidence-based and interactive psychosocial interventions and facilitates a moderated online peer-to-peer social networking environment. Users respond to a set of standardized questions that guide the tailored modules that target risk factors for psychotic relapse and other areas that are salient to people in early recovery, including psychoeducation, early warning signs, depression, social anxiety, and stress management. First-episode patients who used the *HORYZONS* system reduced depression and found the approach to be engaging, empowering, and safe.

Mobile Health. The immense capabilities and functionality of mobile technology have led to the development of the interdisciplinary field of Mobile Health (mHealth). Mobile devices can support the delivery of health care in a variety of formats, enabling patients and providers to send and receive calls and text messages, access websites, provide clinical decision support, and capture and transmit data collected in real-time and in users' own environment. Mobile devices with computational capacity can also house health-focused applications (apps) designed to facilitate patient illness self-management.

FOCUS, the first smartphone intervention specifically designed for people with schizophrenia, was developed iteratively and in close consultation with patients and clinicians. The *FOCUS* system can be activated to prompt users multiple times a day with questions about their symptoms, social functioning, mood, medication adherence, and sleep, which appear on the device's touchscreen. Depending on users' responses, the system launches brief interactive interventions in the form of written text, images, and cartoon screens, designed to both give the individual skills they can use in the moment (e.g., relaxation strategies, medication reminders, social skills tips), as well as train the user in techniques that can be generalizable to a range of scenarios (e.g., cognitive restructuring, hypothesis testing, behavioral activation, distress tolerance). In addition to prompted interactions, all *FOCUS* interventions are available on-demand as part of a toolkit that can be accessed at any time. All user responses are recorded on the device, and the system can send the data to a remote server whenever the phone has connectivity. Authorized individuals (e.g., clinicians and researchers) can access a web-based dashboard summarizing all user responses shortly after they engage the system. People with schizophrenia find the *FOCUS* intervention highly acceptable and useable. Patients who used *FOCUS* experienced a significant reduction in symptoms of psychosis, general psychopathology, and depression. Findings from randomized controlled trials suggest that patients are much more likely to commence treatment and stay engaged in care when assigned to mHealth than a more standard clinic-based group intervention. Moreover, when patients use the *FOCUS* intervention, their use of other time and resource-intensive services drops by approximately 10%.

Lakesha is a 38-year-old woman with schizophrenia who receives community-based mental health services. Lakesha suffers from persecutory ideation and auditory hallucinations that warn her about making eye contact with strangers who are "out to humiliate her." Despite having access to a host of treatment options, she makes only sporadic use of the available resources because she is apprehensive about using public transportation to get to the clinic, and is concerned about "sitting in the waiting area with a bunch of crazies." Consequently, she is becoming increasingly isolative and is experiencing greater distress and impairment linked with her symptoms. Lakesha has an elderly mother with whom she is quite close but whom she has been visiting less and less because she is convinced that the local bus drivers intend to kidnap her. Lakesha's psychiatrist suggested she enroll in a treatment study in which participants are given the *FOCUS* smartphone system for 3 months. Initially, Lakesha thought she would not be able to use the intervention because she was unfamiliar with smartphones. However, after a 45-minute training session with the research staff, she used the device successfully and learned how to operate the touchscreen to respond to *FOCUS* questions and suggestions. Lakesha started using FOCUS that evening and continued to respond to clinical prompts daily. She found the guided breathing techniques to be particularly helpful as she could use them flexibly whenever she wanted, including on the bus ride to see her mother.

After 8 weeks of smartphone intervention use, Lakesha reported leaving her house daily to go to the local store, community center or to visit a neighbor. She reported feeling less anxious about others' intentions. While she still experienced auditory hallucinations regularly, she described being less concerned about them. She successfully integrated several *FOCUS* strategies into her illness management repertoire and is now able to engage in relaxation exercises and basic thought challenging of dysfunctional beliefs (e.g., "what is the evidence that I'm in danger when I leave the house?") independently.

Short Message Service (SMS) or text messaging approaches have also been used to support psychiatric rehabilitation for patients with schizophrenia. For example, *Mobile Assessment and Treatment for*

Schizophrenia (MATS) is a mobile phone intervention employed to administer questions and algorithm-driven follow-up suggestions to patients in the areas of medication adherence, social functioning, and management of auditory hallucinations. *MATS* initiates a sequence of brief automated back-and-forth exchanges, including illness-management suggestions drawn from cognitive behavioral treatments for psychosis.

In addition to enabling completely novel treatments, mobile technologies can supplement traditional models of care. Community-based research has shown that patients with psychosis and co-occurring substance use find texting to be a highly engaging method for connecting with their case manager. When a *Mobile Interventionist* used daily texting to check in and provide daily support, patients responded to 87% of the messages sent and rated the therapeutic alliance with the Mobile Interventionist highly. A randomized controlled trial showed that when ACT care is augmented with a texting mobile interventionist, patients with serious mental illness experience improved psychiatric symptom management and self-rated recovery than when receiving ACT alone.

Virtual Reality. Virtual Reality (VR) refers to multimedia techniques in which individuals engage in computer-generated scenarios on a computer screen or head-mounted stereoscopic display that creates a highly immersive 3D environment. VR has been used primarily to facilitate exposure therapy for people with anxiety disorders, but several studies have demonstrated its potential utility in supporting rehabilitation for people with psychosis. These include help identifying contextual triggers for persecutory ideation, learning skills, and practicing job interviews. Preliminary evidence suggests VR may be an effective tool, including benefits from continuous feedback from an on-screen virtual coach.

Technology-Assisted Psychiatric Rehabilitation During a Global Pandemic. The global coronavirus disease 2019 (COVID-19) has proven to have devastating effects on the lives of millions worldwide. Due to the highly infectious nature of COVID-19, comprehensive social distancing policies have been implemented broadly, blocking the ability of psychiatric rehabilitation professionals to provide in-person services. Disruption of clinic-based services coupled with a rise in the needs of people experiencing social isolation, economic upheaval, anxiety, and psychological trauma have propelled health care systems to embrace technology as a primary method to deliver services in a manner that is unprecedented. Major examples include the widespread implementation of videoconferencing technology as an alternative tool to conduct live telemental health sessions, a sharp increase in the use of clinical messaging services and companies that facilitate texting interactions between patients and remote providers, adoption of mental health apps and loosening of the regulatory restrictions on their use, and distribution of mental health resources through social media, platforms, and online streaming services. While it is not fully clear how the psychiatric rehabilitation landscape will look after the COVID-19 pandemic subsides, it is unlikely that we will see a full return to traditional in-person service models. Some pandemic-related changes in regulatory guidelines and policies that govern the use of technology in mental health care will likely remain. Thus, the coronavirus crisis and the collective response to emergent patient needs will likely prove to be a key catalyst for the adoption of digital mental health technologies.

SUMMARY

Psychiatric rehabilitation comprises a range of techniques and interventions that help people with disabilities related to psychiatric illness to manage their illness, succeed in regular social roles in the community, pursue their personal goals, and “recover” in the sense of moving beyond feeling dominated by illness to create a meaningful life. Rehabilitation requires creating normative opportunities, such as for education, employment, and housing; providing support as needed to ensure success; and helping people acquire the skills and natural supports they want to pursue life goals.

FURTHER READINGS

Alvarez-Jimenez M, Alcazar-Corcoles MA, González-Blanch C, Bendall S, Mcgorry PD, Gleeson JF. Online, social media and mobile technologies for psychosis treatment: A systematic review on novel user-led interventions. *Schizophr Res*. 2014;156(1):96–106.

Becker DR, Drake RE. *A working life for people with severe mental illness*. Oxford University Press; 2003.

Ben-Zeev D. The digital mental health genie is out of the bottle. *Psychiatr Ser*. 2020;71(12):1212–1213.

Ben-Zeev D, Kaiser SM, Krzos I. Remote “hovering” with individuals with psychotic disorders and substance abuse: Feasibility, patient engagement, and therapeutic alliance with a text-messaging mobile interventionist. *J Dual Diagn*. 2014;10(4):197–203.

Bond GR, Drake RE, Becker DR. An update on IPS Supported Employment. *World Psychiatry*. 2020;19:190–191.

Burns T, Catty J, Becker T, et al. The effectiveness of supported employment for people with severe mental illness: A randomised controlled trial. *Lancet*. 2007;370:1146–1152.

Corrigan PW, Mueser KT, Bond GR, Drake RE, Solomon P. *The Principles and Practice of Psychiatric Rehabilitation*. Guilford Press; 2008.

Drake RE, Bond GR, Becker DR. *Individual Placement and Support: An evidence-based approach to supported employment*. Oxford University Press; 2012.

Goering P, Veldhuizen S, Watson A, et al. National At Home/Chez Soi Final Report. Calgary, AB: Mental Health Commission of Canada. http://www.mentalhealthcommission.ca. 2014.

Gottlieb JD, Romeo KH, Penn DL, Mueser KT, Chiko BP. Web-based cognitive–behavioral therapy for auditory hallucinations in persons with psychosis: A pilot study. *Schizophr Res*. 2013;145(1):82–87.

Hoffmann H, Jäckel D, Glauser S, Mueser KT, Kupper Z. Long-term effectiveness of supported employment: five-year follow-up of a randomized controlled trial. *Am J Psychiatr*. 2014;171(11):1183–1190.

Liberman RP. *Recovery from Disability: Manual of Psychiatric Rehabilitation*. American Psychiatric Press; 2009.

Marshall T, Braude L, Goldberg RW, et al. Supported employment: assessing the evidence. *Psychiatr Ser*. 2014;65:16–23.

Pogue JA, Bond GR, Drake RE, Logsdon SM, Becker DR. Growth of IPS supported employment programs in the U.S.: An update. *Psychiatr Ser*. 2022;73(5):533–538.

Rog DJ, Marshall T, Dougherty RH, et al. Permanent supportive housing: assessing the evidence. *Psychiatr Ser*. 2014;65:287–294.

Rotondi AJ, Anderson CM, Haas GL, et al. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. *Psychiatr Ser*. 2010;61(11):1099–1105.

Swanson SJ, Becker DR. *IPS Supported Employment: A Practical Guide*. Dartmouth Psychiatric Research Center; 2013.

Tabol C, Drebing C, Rosenheck R. Studies of “supported” and “supportive” housing: A comprehensive review of model descriptions and measurement. *Eval Program Plann*. 2010;33:446–456.

Thornicroft G, Szmukler G, Mueser KT, Drake RE, eds. *Oxford Textbook of Community Mental Health*. Oxford Press; 2011.

Torous J, Myrick KJ, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. *JMIR Ment Health*. 2020;7(3):e18848.

Tsemberis S, Gulcur L, Nakae M. Housing First, consumer choice, and harm reduction for homeless individuals with a dual diagnosis. *Am J Public Health*. 2004;94:651–656.

12.14 Medical Health in Schizophrenia

ROBERT J. BOLAND, M.D.

We have known for centuries that people with schizophrenia have shorter lives than the rest of us. In the 1830s, the Council of the Statistical Society of London commissioned one of its fellows, William Farr, to investigate patient mortality in the county asylums. His result was the first study explicitly examining death rates in psychiatric patients, titled *Report on the Mortality of Lunatics*. Drawing on his observations of several English asylums, he detailed the high mortality and the discrepancies between male and female patients and economic groups.

Table 12.14–1.
The "Take Home"

Patients with Schizophrenia are more likely than the general population to

- Die prematurely
- Have a comorbid medical illness
- Be neglected by the medical establishment
- Have an unfavorable clinical outcome

In 1936, Ornulv Odegard, reporting on the mortality rates in Norwegian mental hospitals, observed that they most often died not from their mental illness but tuberculosis and other infectious diseases. Odegard believed that this was partly a function of their environment, with its crowding and poor nutrition. However, he also felt their mental illness contributed in that the apathy it caused left them unwilling to complain or seek medical treatment for their illnesses.

In 1959, when tobacco use was at its highest level, Albert Abraham published *Cardiovascular Disease in Psychotics*, demonstrating a high rate of cardiovascular disease (CVD) in hospitalized patients with schizophrenia. He rejected the popular view of the time that "the protective life of people with schizophrenia seems to guard them from early onset of degenerative diseases of the heart." By the late 1970s, the accumulated data made it clear that psychiatric patients had higher mortality rates than the general population. However, little was done to understand why this was so.

Today there is ever-growing literature investigating the importance of comorbidities in persons with psychotic disorders. Since 2000, almost 100 systematic reviews and meta-analyses have focused on the physical health comorbidities associated with various mental illnesses. For schizophrenia, there were 15 reviews done between 2000 and 2019. These include two reviews for CVD, five for cancer, two for diabetes, three for metabolic syndrome, and four for other disorders (such as osteoporosis, arthritis, and mixed illnesses).

Taken together, these studies show that physical comorbidities are the rule, not the exception, for patients with schizophrenia. Furthermore, the sociodemographic factors that affect persons with psychosis influence this frequency, and comorbid disorders are becoming increasingly common in older patients with severe mental illness.

Table 12.14–1 summarizes the major points that this chapter will discuss.

ALL-CAUSE MORTALITY IN SCHIZOPHRENIA

Individuals with schizophrenia and other chronic psychotic disorders die about 10 to 15 years prematurely, with men dying about 15 years prematurely and women dying, on average, 12 years. In addition, schizophrenia patients are at higher risk of developing various medical illnesses, are less likely to receive and benefit from appropriate medical treatment, and are more likely to experience unfavorable clinical outcomes. Many of these patients with comorbidity, perhaps as many as two-thirds, have more than one physical disease—"multimorbidity."

Some earlier studies use standardized mortality ratios (SMRs) to quantify the additional mortality risk associated with schizophrenia relative to the general population. SMRs are the ratio of observed to expected deaths within a specified sample or given age range (sometimes 100). The *expected* mortality rate is the rate seen in the general population, whereas the *observed* mortality rate is that seen in the population in question (e.g., those with schizophrenia). It is, therefore, a measure of relative and not absolute risk. More recent studies, such as the Global Burden of Disease (GBD) study, use such estimates as years of life lost of premature mortality (YLLs), calculated as the number of cause-specific deaths multiplied by a loss function specifying the years lost for deaths as a function of the age at which death occurs. This approach provides a somewhat more accurate estimate of the mortality cost. The most recent analysis from that study includes data up to 2019. However, much of the comorbidity analyses are from the 2010 analysis. That study found that the primary cause of premature death in patients with schizophrenia was due to chronic physical diseases, with CVD accounting for more than 33% of premature deaths. In contrast, suicide, homicide, and accidents accounted for less than 15% of the excess deaths.

The GBDs conclusion is consistent with earlier registry-based studies, such as by Brown and colleagues in 370 schizophrenia patients over 25 years, who found that 81% of patients with schizophrenia died from medical diseases. In their study, the most common diseases were cardiovascular and circulatory diseases, respiratory diseases including pneumonia, endocrine diseases including diabetes, nervous system diseases, and cancer.

In addition, population-based cohort studies consistently identify a significant increase in overall all-cause mortality ranging from two to four times higher in patients with schizophrenia compared to the general population, with somewhat smaller increases sometimes observed in exclusively inpatient samples. In a population-based sample of 7,784 inpatients with schizophrenia analyzed by Osby and colleagues, male and female subjects had an SMR of 2.8 and 2.4, respectively, for death due to any cause. All types of natural causes of death, except cancer in men and nervous system diseases in women, were significantly elevated. Overall, patients were more than twice as likely to die from cardiovascular, cerebrovascular, or respiratory disease. A study by Olfson and colleagues reported on SMRs for over 1 million U.S. Medicaid patients followed for almost 5 million person-years of follow-up, with over 74,000 deaths. Patients with schizophrenia were 3.7 times more likely to die in the follow-up period than the general population. CVD had an SMR of 3.6 and the highest mortality rate (403.2 per 100,000 person-years) of any individual cause. Cancer overall had the next highest mortality rate, with lung cancer contributing the most with an SMR of 2.4 and a mortality rate of 74.8 per 100,000 person-years. Elevated SMRs were also observed for chronic obstructive pulmonary disease (COPD) (SMR = 9.9) and influenza and pneumonia (SMR = 7.0), suggesting the effects of high smoking rates in this population, as well as for diabetes mellitus (SMR = 4.2, mortality rate = 61.8 per 100,000 person-years). Nonsuicide-related substance-induced death, primarily from alcohol or other drugs, was also a significant cause of death (95.2 per 100,000 person-years). Finally, accidental deaths (119.7 per 100,000 person-years) accounted for more than twice the number of deaths attributable to suicide (52.0 per 100,000 person-years). While much attention understandably goes to the effects of suicide in this population, this large and definitive study from the public sector underscores the critical effects of medical morbidity and mortality in this population.

MEDICAL CAUSES OF EXCESS MORTALITY IN SCHIZOPHRENIA

Despite finding that comorbid medical conditions, rather than unnatural causes of death such as suicide, confer most of the excess mortality seen in persons with schizophrenia, medical comorbidities often remain underdiagnosed and undertreated in these patients. Table 12.14–2 lists the most common medical causes of mortality in schizophrenia.

Table 12.14–2.
Common Medical Comorbidities in Schizophrenia

- Cardiovascular disease
  - Congestive heart failure
  - Cerebrovascular disease
  - Coronary heart disease
- Type 2 diabetes mellitus
- Infectious diseases
  - Pneumonia: COVID-19, other viruses, tuberculosis
  - HIV
  - Hepatitis B, C
- Cancer
  - Lung
  - Breast
- Pulmonary disease
  - COPD
  - Asthma

Cardiovascular Disease

Heart disease is the leading cause of death, accounting for about 1 in every 4 deaths in the United States, and it costs the United States about 363 billion dollars each year. It remains the leading cause regardless of sex, race, or ethnicity. However, non-Hispanic black persons are more than twice as likely to die of heart disease than other races or ethnicities. From 1999 through 2017, the death rates for heart disease have decreased in all groups. However, this decrease has slowed in recent years.

Although heart disease seems to be slowly declining in the general population, this improvement has not included patients with schizophrenia. As mentioned above, CVD is the leading cause of mortality in individuals with schizophrenia and is more common in this group than in the general population. For example, Fan and colleagues reviewed 13 studies with 422,698 patients with schizophrenia who had CVD outcome data. They calculated a pooled relative risk of 1.53 for CVD, 1.2 for coronary heart disease, 1.7 for stroke, and 1.81 for congestive heart failure (CHF). Similarly, a large meta-analysis by Correll and colleagues of more than 3 million patients found that the pooled CVD prevalence for all patients with serious mental illness (SMI) was 9.9%, significantly higher than the normal controls (odds ratio [OR] = 51.53, 95% CI: 1.27–1.83). In this study, schizophrenia was significantly associated with CVS, coronary heart disease, and cerebrovascular disease. The risk of CVD-related death was also higher than the controls, with a hazard ratio (HR) of 2.45 (95% CI: 1.64–3.65). Other studies have had similar findings with relative risks ranging from two to fivefold.

A systematic review focusing on acute coronary syndrome (ACS, a principal cause of cardiac mortality, is a sudden rupture of atherosclerotic plaques, causing thrombus formation and subsequent infarcts, CHF, and arrhythmias) found that death rates following ACS were higher in patients with schizophrenia than mentally healthy patients. The investigators also found that patients with ACS were at a greater risk for major adverse cardiac events (MACE) following hospitalization. They hypothesized that the causes were multifactorial, including patient factors (CVS risk, poor self-care, socioeconomic factors) and health care–related (treatment bias, prejudice, and suboptimal care).

The Framingham Heart Study (a long-term cohort study funded by the National Heart, Lung, and Blood Institute [NHLBI]) and other large population-based samples have identified modifiable risk factors for CVD and how these risk factors interact. The knowledge gained from these studies has important implications for decreasing cardiovascular-related morbidity and mortality, including in higher-risk populations such as people with schizophrenia. It is important to note that persons with schizophrenia have an increased prevalence of all key modifiable cardiovascular and metabolic risk factors, including obesity, smoking, hypertension, dyslipidemia, and hyperglycemia.

There is some disagreement in the literature as to the influence of sex on cardiac morbidity. For example, a study by Polcwiartek and colleagues looking at CHF data from 2002 to 2017 in more than 20,000 patients found that having an SMI increased the risk of death from CHF in men, but not women. In studying more than 8,000 patients with schizophrenia in the United Kingdom, McCreadie and colleagues found the risk for CHD and stroke higher in men with schizophrenia, but not women. Other studies find the risk to be higher for both sexes but highest in men. The greater risk for men is consistent with the general population.

Most concerning is the disparity in treatment for patients with schizophrenia who develop CVD. For example, Solmi and colleagues conducted a meta-analysis of 47 studies with a total of more than 1 million patients with comorbid CVD and mental illness. They found a significant disparity in screening between patients with schizophrenia and any CVD versus patients without a mental disorder (OR = 0.611, 95% CI: 0.554–0.673), a greater disparity than for any other mental disorder investigated.

Diabetes Mellitus

Diabetes-related morbidity and mortality significantly affect the general population, increasing overweight and obesity rates in developed countries. Diabetes is currently the seventh leading cause of death in the United States, with 34.2 million people having the disease. One in five people are unaware that they have the disease, and another 88 million people in the United States have prediabetes. After two decades of increases, the incidence of diabetes has slowly decreased from 2008 to 2017, and in 2019, 1.5 million new cases were diagnosed in U.S. adults.

The prevalence of type 2 diabetes mellitus is substantially higher in individuals with schizophrenia than in the general population, with prevalence estimates ranging from two to four times higher. Stubbs and colleagues reviewed 25 studies and found that the prevalence of type 2 diabetes mellitus in people with schizophrenia was 9.5%. Vancamport and colleagues of 22 studies estimated a pooled prevalence of 11.5% for patients with schizophrenia, with a risk of nearly double that in the general population. Similarly, Ward and Druss estimated a median prevalence of 13%. However, they reported a wide variation across different studies, owing to differences in study design, disease detection, and sample sizes. The comorbidity significantly increases the mortality risk for patients with schizophrenia; one population-based study in Hong Kong, for example, estimated an elevated mortality risk regardless of the level of disease or number of complications.

Underrecognition of schizophrenia as a marker for increased risk for diabetes may contribute to the impact of diabetes-related risk on the overall elevated mortality in the absence of extra efforts in primary or secondary prevention. The Canadian Diabetes Association and the American Association of Clinical Endocrinologists have added schizophrenia to the list of risk factors for diabetes in their screening guidelines. However, ongoing evidence generally documents low screening rates for abnormalities in plasma glucose or lipids in this population.

Infectious Disease

In developing countries, infectious disease remains a substantial cause of mortality for persons with and without schizophrenia. This fact has become remarkably more evident during the COVID-19 pandemic. Patients with schizophrenia are at higher risk than the general population for becoming infected with the SARS-CoV-2 and dying from it. A review of seven studies from various countries found increased rates in the range of an OR of 1.3 to 1.53. Similarly, all the studies with relevant data showed increased mortality rates, with ORs ranging from 2.25 to 4.36. Another meta-analysis estimated a mortality risk of about twice that for the general population (OR 2.28, using more than 100 studies). These studies provide compelling evidence to support the notion that the COVID-19 pandemic has disproportionally affected people with schizophrenia. The reasons are multiple and similar to those for other chronic diseases; however, some studies highlight the lower vaccination rates in patients with schizophrenia. In response, the centers for disease control and prevention (CDC) added schizophrenia spectrum disorders to its list of underlying medical conditions associated with a higher risk for severe COVID-19, including it in the category of disorders with solid evidence (along with mood disorders).

Even before the pandemic, infectious diseases were a significant problem among patients with schizophrenia. For example, the human immunodeficiency virus (HIV) infection rate has been estimated at 3% in patients with a severe mental illness, an eightfold increase over the general population rate. Data have shown that knowledge of exposure risks is deficient among these patients. Rates of hepatitis B (23.4%) and C (19.6%) infections are also significantly elevated among those with a severe mental illness, approximately 5 and 11 times the overall estimated population rates for these infections. Studies also consistently show a higher incidence of tuberculosis among patients with schizophrenia than the general population.

Cancer

According to the CDC, there were 1.7 million new cancer cases in the United States in 2018 (the most recent year with available data), and about 600 thousand people died of cancer that year. However, new cancers are decreasing somewhat, from an age-adjusted rate of 481.7 per 100,000 people in 1999 to 435.8 per 100,000 in 2018. Similarly, mortality has declined, with an age-adjusted cancer death rate of 200.7 per 100,000 in 1999 to 149.2 per 100,000 in 2018.

The association between schizophrenia and cancer risk is an area of some debate. Various studies disagree on whether patients with schizophrenia have a higher risk of cancer than the general population. The majority of studies suggest that the risk for cancer in patients with SMI is similar to the general population. Some, such as a meta-analysis by Li and colleagues, find a slightly decreased risk of cancer in patients with schizophrenia. An important variable is smoking, and studies that try to control for that tend to find much lower risks for cancer in patients with schizophrenia. However, attempts to account for smoking as a variable are inherently problematic given its prevalence in this population.

In addition, the mortality risk from cancer is higher for patients with SMIs. This likely relates to poorer self-care, including a lower likelihood of preventive care or following through with screening. Some psychotic patients are poor historians who may not accurately report symptoms that would usually trigger additional tests. Patients with schizophrenia are more likely to be homeless, with little protection from UV light and greater exposure to various environmental stresses. The metabolic side effects of antipsychotics may contribute to the cancer risk, as do worse dietary habits. In addition, there may be treatment biases, including a tendency to attribute physical symptoms to psychological causes or a nihilism around using potentially complex treatments. Most important is the prevalence of smoking—as discussed later, this remains highly prevalent in patients with schizophrenia.

The risk for cancer in patients with SMI may be higher in patients living in middle- or lower-income countries. One review by Park and colleagues, using 80 articles, found that cancer incidence is rising in these countries, both overall and particularly for patients with SMI. This increase relates to various factors. For example, although cervical cancer accounts for most cancer deaths in these countries, vaccinations (the human papillomavirus vaccine has been available since 2006) remain a low priority for these countries. In addition, cancer detection and screening are inadequate as well. Although these health care inadequacies are problems for the populations, it is worse for patients with schizophrenia.

Pulmonary Disease

Chronic lower respiratory disease, including COPD, is the fourth leading cause of death in the United States, after heart disease, cancer, and accidental deaths. CDC Data from 2019 reports 156,979 deaths from chronic lower respiratory diseases, including asthma, a death rate of 47.8 per 100,000. About 4.6% of Americans have been diagnosed with COPD, emphysema, or chronic bronchitis. However, that number is decreasing some. From 2000 through 2014, the age-adjusted rate for COPD decreased 22.5% for men and 3.8% for women 25 or older. The age-adjusted death rate also decreased for most groups, with the notable exception of black women, increasing 4.2%. In most cases, the cause of death was COPD itself.

Like other chronic diseases, people with schizophrenia have significantly higher COPD, asthma, and pneumonia rates. In a systematic review of 21 studies consisting of over 600 thousand patients with schizophrenia, Suetani and colleagues reported a prevalence of 7.7% for COPD, 7.5% for asthma, 10.3% for pneumonia, and 0.3% for tuberculosis. After adjusting for publication bias, the prevalence of COPD increased to 19.9%. In each case, this is higher than the general population. For example, the OR was 1.82 for COPD, 1.70 for asthma, and 2.62 for pneumonia. These rates are consistent with earlier studies. For example, Zareifopoulos and colleagues, using more restrictive inclusion criteria, reported that people with schizophrenia had an OR of 1.57 for COPD.

This increased risk undoubtedly relates to the several risk factors shared between schizophrenia and respiratory diseases. Most important, again, is smoking. Other factors include childhood exposure to air pollution, increased in patients with schizophrenia, and many socioeconomic factors. In addition, people with schizophrenia are less physically active than the general population. Also, antipsychotics may worsen respiratory diseases, mainly through weight can.

Oral Health in Schizophrenia

People with schizophrenia have poor dental health, among the poorest reported in the literature. Reports suggest that patients with schizophrenia brush their teeth irregularly, leading to functional difficulties such as difficulty with chewing and eating, in addition to increased risk for decay, infection, and inflammation.

MECHANISMS OF DISEASE

The previous sections have already shed some light on many of the reasons for the increased incidence of most major chronic physical disorders. Smoking, socioeconomic factors, psychological stress, and side effects of antipsychotics, particularly weight gain and metabolic changes, undoubtedly play essential roles. Patients with schizophrenia tend to have less insight into their illness and are less likely to seek medical care for physical complaints. When they do, health care professionals may dismiss their complaints as part of their mental disease. Even when diagnosed, patients with schizophrenia are less likely to receive optimal care, either due to their avoidance behaviors or owing to physician bias.

Apart from behavioral, social, and economic factors, schizophrenia may have direct physiologic effects that put patients more at risk for physical illnesses. For example, inflammation may have a role in schizophrenia and metabolic and other physical disorders. Some studies have found a correlation between markers of systemic inflammation (such as C-reactive protein and myeloperoxidase) and dyslipidemia in patients with schizophrenia.

Limited data suggest that increased activation of the hypothalamic–pituitary–adrenal (HPA) axis and the sympathetic nervous system can adversely affect the body. However, genetic factors may play a role. For example, one report found that 18% to 19% of individuals with schizophrenia had a family history of type 2 diabetes mellitus. In addition, some pharmacogenomic studies suggest that specific genetic variants that decrease the action of certain enzymes (such as methylenetetrahydrofolate reductase) are associated with an increased cardiovascular risk.

The physiologic effects are likely reciprocal as well. For example, some research suggests that metabolic dysfunction may directly affect brain structure and increase vulnerability to psychotic or other SMI.

Although research into possible direct physiologic effects is fascinating, it is preliminary. Even if more definitive studies follow, it remains likely that behavioral, lifestyle, and environmental factors offer the best explanation for why so many chronic physical illnesses are so common in patients with schizophrenia. This reality is perhaps good news, as we already have proven interventions to modify many of these risk factors. This chapter will discuss those after considering some of the potentially modifiable risk factors.

INCREASED PREVALENCE OF MODIFIABLE RISK FACTORS IN PATIENTS WITH SCHIZOPHRENIA

Key risk factors for the most significant causes of morbidity and mortality are all more prevalent in schizophrenia. These include public health priorities such as obesity, smoking, hypertension, dyslipidemia, and hyperglycemia. Thus, successfully addressing the excess burden observed in persons with schizophrenia depends on primary and secondary prevention efforts targeting modifiable risk factors. Table 12.14–3 discusses some of the most common modifiable risk factors in schizophrenia.

Table 12.14–3.
Modifiable Risk Factors in Patients with Schizophrenia

Smoking
Substance abuse
Metabolic syndrome
  - Overweight and obesity
  - Hypertension
  - Dyslipidemia
  - Hyperglycemia
  - Insulin resistance

Smoking and Substance Abuse

As we discuss throughout this chapter, cigarette smoking is a well-known risk factor, and it is associated with most of the diseases we have discussed. More than 16 million Americans are living with a smoking-related disease. Smokers die about 10 years earlier than nonsmokers, and smoking is responsible for more than 480,000 deaths every year in the United States, including more than 41,000 from secondhand smoke. This represents more than one in five deaths annually, of 1,300 deaths per day. Worldwide, tobacco use causes more than 7 million deaths per year. Although the prevalence of smoking has declined significantly, primarily due to public health campaigns and related cultural changes, smoking remains the leading cause of preventable disease, disability, and death.

While the overall rate of current smoking in the United States declined during this time from 42% to 18%, it remains very prevalent among people with mental illness, and perhaps as many as 70% to 85% of people with schizophrenia smoke. Persons with mental illness who smoke also tend to smoke more cigarettes than other smokers. Smokeless tobacco is also becoming more popular. Tobacco use is most common in people with mental illness younger than 45, who have low education levels and live in poverty. Co-occurring substance use is another influential risk factor. Some experts speculate that the high smoking rates are because nicotine’s stimulant effect can transiently improve some symptoms. However, quitting smoking also improves mood and quality of life in individuals with severe mental illness.

Besides nicotine, other substances are common for patients with schizophrenia, and perhaps half of patients with the disorder also suffer from a substance use disorder. The most common substances include alcohol, cannabis, and cocaine. The prevalence of each has an extensive range depending on the population studied and the methods used. However, it is about three times that of the general population for each of the drugs. In addition to the many health problems caused by these drugs, substance use can cause exacerbations of the user’s substance disorder, poor adherence to treatment, and an overall functional decline.

The exact link between substance use and schizophrenia is unclear, and experts have advanced various theories, many of which have commonsense appeal despite little empirical support. For example, the diathesis–stress model posits an interaction between neurobiologic vulnerabilities and environmental stress leading to schizophrenia and substance use. The self-medication hypothesis suggests that patients with schizophrenia use substances to address their symptoms. Alternatively, the two disorders may share a common pathophysiology with overlapping brain pathways. For example, schizophrenia may cause a deficit in the reward pathway, for which patients attempt to use substances to compensate. Admittedly, the theories are not mutually exclusive. However, there is little evidence for self-medication as a primary explanation, whereas there is growing evidence that schizophrenia and substance use disorders share common disordered circuits.

Metabolic Syndrome

The metabolic syndrome is not a separate entity from the other risk factors discussed here but instead brings together a constellation of several abnormal clinical and metabolic findings. These include adiposity, insulin resistance, hypertension, elevated triglycerides, and low high-density lipoprotein (HDL) levels. The most widely accepted definition of metabolic syndrome is, from the third adult treatment panel (ATP III) of the National Cholesterol Education Program (NCEP) and the adapted adult treatment panel (ATP III-A) of the NCEP Program proposed by the American Heart Association. Table 12.14–4 lists the guidelines.

Table 12.14–4. Definition of Metabolic Syndrome

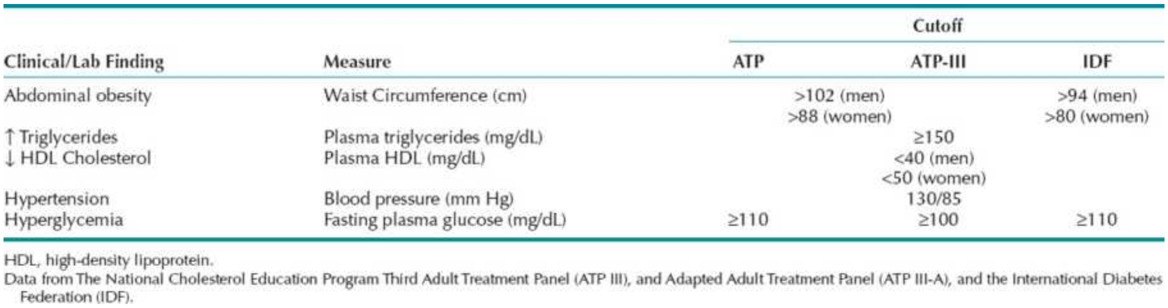

		Cutoff		
Clinical/Lab Finding	**Measure**	**ATP**	**ATP-III**	**IDF**
Abdominal obesity	Waist Circumference (cm)	>102 (men) >88 (women)		>94 (men) >80 (women)
↑ Triglycerides	Plasma triglycerides (mg/dL)		≥150	
↓ HDL Cholesterol	Plasma HDL (mg/dL)		<40 (men) <50 (women)	
Hypertension	Blood pressure (mm Hg)		130/85	
Hyperglycemia	Fasting plasma glucose (mg/dL)	≥110	≥100	≥110

HDL, high-density lipoprotein.
Data from The National Cholesterol Education Program Third Adult Treatment Panel (ATP III), and Adapted Adult Treatment Panel (ATP III-A), and the International Diabetes Federation (IDF).

ATP III-A modified this slightly by decreasing the cutoff for fasting glucose, and the International Diabetes Federation (IDF) stressed the importance of waist circumference by modifying the parameters to a lower cutoff. The general population prevalence of metabolic syndrome is approximately 34% in U.S. adults, similar to other developed countries. Like other risk factors, it is on the rise, increasing from 25% in 1988. The groups with the highest risk were non-Hispanic black females, whereas non-Hispanic black males are less likely to have metabolic syndromes than are non-Hispanic white males. Lower education and older age are also associated with the syndrome.

The metabolic syndrome and the composite syndrome components are major risk factors for type 2 diabetes mellitus and CVD. In addition, as discussed, obesity, dyslipidemia, hypertension, and type 2 diabetes mellitus are all common in people with schizophrenia. Most relevant studies indicate that the prevalence of metabolic syndrome in patients with schizophrenia ranges from about 40% to over 50%.

Antipsychotic medications are associated with the risk of metabolic syndrome to differing degrees, again generally in proportion to their risk for weight gain. For example, pooled data from two 26-week randomized double-blinded studies, one comparing aripiprazole with placebo and the other comparing aripiprazole with olanzapine, showed incidences of metabolic syndrome of 19.2 ± 4.0% for olanzapine, 12.8 ± 4.5% for placebo, and 7.6 ± 2.3% for aripiprazole.

Overweight and Obesity. Increased adiposity is associated with increased morbidity and mortality from hypertension, dyslipidemia, type 2 diabetes mellitus, coronary artery disease, stroke, and certain cancers, including colon, breast, endometrial, and gallbladder cancer. We usually use the body mass index (BMI) to define the conditions, in which we calculate weight in kilograms divided by the square of height in meters (the latter being proportional to body surface area). Persons with a BMI greater than 25 kg/m^2 are considered overweight, and a BMI greater than 30 kg/m^2 generally defines obesity. As of 2019 (the most recent available CDC data), approximately 42% of adults in the United States are obese. Obesity is increasing—in 2000, the rate was 30%. During the same span, severe obesity increased from about 5% to 9%. This upward trend appears to continue, both nationally and worldwide. Obesity disproportionally affects certain groups, as non-Hispanic Black adults have the highest age-adjusted obesity (50%), followed by Hispanic adults (45%), non-Hispanic White adults (42%), and non-Hispanic Asian adults (17%). It peaks in adults aged 40 to 59 years, and the rate has a slight decrease after age 60. The role of socioeconomic factors is complex. However, on average, people with college degrees have a lower obesity prevalence than those with less education. For women, obesity is lowest in the highest-income group, whereas it is lowest in men's low- and high-income groups. The effect of income is modified by sex, race, and ethnicity, as income seems to have little effect for non-Hispanic Black women, and for non-Hispanic Black men, obesity is higher in the highest-income group than the lowest-income group. The estimated annual medical cost of obesity in the United States was 147 billion in 2008, and medical costs for people with obesity are more than $1,900 higher each year than for people with a healthy weight.

Obesity is about twice as prevalent in people with schizophrenia in comparison to the general population. Evidence from numerous studies characterizes the problem. Substantial evidence suggests that treatment with psychotropic medications, particularly certain antipsychotic medications, increases body weight, and adiposity. Most antipsychotics lead to weight gain, although they differ in degree. Clozapine and olanzapine have the highest propensity, and haloperidol, lurasidone, and ziprasidone have the lowest. The mechanism of weight gain is varied, with histamine H1 receptors, D2 dopamine receptors, blockade of 5-hydroxytryptamine receptor 2C, and dysregulation of glucagon-like peptide-1 receptors all playing a role. Mood stabilizers, particularly sodium valproate, and, to a lesser degree, lithium can also increase weight. Carbamazepine has a lower propensity, and lamotrigine appears not to affect weight gain.

Hypertension. Schizophrenia patients are about twice as likely to have hypertension than members of the general population. Estimates of the rate of hypertension in schizophrenia patients range from 19% to roughly 47%. By comparison, approximately 15% of the general population has hypertension.

The mechanisms causing hypertension in patients with schizophrenia are multiple and include weight gain, insulin resistance, and increased sympathetic nervous system activity with resulting sodium retention. Again, many psychotropic medications put patients at risk for hypertension. Although the α_1-adrenoceptor antagonism associated with some medications can attenuate hypertension, the substantial weight gain overwhelms any protective effect. Cigarette smoking may also increase blood pressure both acutely due to the stimulant effects of nicotine and secondarily due to chronic smoking.

Dyslipidemia. Dyslipidemia is a term to describe abnormal lipid levels, such as increases in total or low-density lipoprotein (LDL) cholesterols and decreases in HDL cholesterols. It is responsible, at least in part, for as much as a third of ischemic heart disease and a fifth of cerebrovascular disease cases. In the United States, the prevalence of total cholesterol above 200 mg/dL is 38% higher in women (40%) than men (35%). In addition, about 29% of U.S. citizens have high levels of LDLs (at or above 130 mg/dL), with more men (30%) than women (27%).

Patients with schizophrenia are at even higher risk of dyslipidemia, and perhaps half of persons with schizophrenia have some measure of the disorder. Despite this, the disorder is often untreated in this population.

Some antipsychotic medications are associated with clinically significant increases in serum lipid concentrations, generally, but not exclusively, in association with their propensity to produce weight gain. The medications causing the highest risk include such atypical antipsychotics as clozapine and olanzapine and low-potency conventional antipsychotics. Risperidone and quetiapine have a moderate risk, and ziprasidone, aripiprazole, and high-potency conventional antipsychotics have the lowest risk.

Insulin Resistance and Hyperglycemia. The prevalence of insulin resistance, variously defined, in patients with schizophrenia is about 1.5 to 2 times the prevalence in the general population. Hyperglycemia in schizophrenia (glucose >100 mg/dL) is common, affecting up to 20% of patients in pooled samples. If uncontrolled, it can progress to type 2 diabetes mellitus. Most evidence implicates antipsychotics as increasing the risk for insulin resistance and hyperglycemia in patients with schizophrenia, as opposed to the disease state. Lifestyle can undoubtedly affect outcomes, with poor diet, inactivity, smoking, and lack of exercise commonly associated with overweight and obesity and insulin resistance in untreated nonpsychiatric samples and treated patients.

Different antipsychotic medications are associated with different levels of risk for insulin resistance and hyperglycemia, generally proportional to their association with drug-induced weight gain, although additional adiposity-independent effects may contribute to risk with some medications. For example, antipsychotic agents such as clozapine and olanzapine achieve relatively high intracellular concentrations where they can interact with intracellular glucose transporter proteins. This effect reduces glucose transport from the extracellular to intracellular space, potentially contributing to insulin resistance. The result is an increased risk for hyperglycemia, independent of any weight or adiposity effects.

PRIMARY AND SECONDARY PREVENTION

Cardiovascular Events

In the general U.S. population during the past decade, most of the observed reductions in CVD mortality are attributable to improvements in the treatment of acute cardiovascular events and long-term secondary prevention. For example, the decrease in the fatality rate for hospitalized

MI is likely due to the acute utilization of aspirin, thrombolytics, β-blockers, and angiotensin-converting enzyme (ACE) inhibitors. Longer-term post-MI use of aspirin, β-blockers, ACE inhibitors, and statins, as well as therapeutic lifestyle changes, have made additional critical contributions to reductions in MI-related mortality. Unfortunately, patients with schizophrenia who experience an acute MI are significantly less likely than the general population to receive drug therapies of proven benefit, including thrombolytics, aspirin, β-blockers, and ACE inhibitors. Patients with schizophrenia are also significantly less likely than members of the general population to undergo cardiac catheterizations and receive emergency angioplasties or coronary artery bypass grafts.

Smoking

As we discussed, the general decline in smoking rates is not apparent for patients with schizophrenia. Many patients with schizophrenia rely on state programs such as Medicaid, however many states have only limited coverage for smoking cessation treatments. Furthermore, there are often barriers to receiving these smoking cessation treatments, including requiring prior authorization and copayments and limiting the number of courses of treatment, all of which create additional hurdles for disabled patients with fewer resources and fixed incomes.

Most research on smoking cessation treatment does not include patients with severe mental illness. However, the limited evidence suggests that the same interventions that can help the general population can also help patients with schizophrenia. For example, some studies show that bupropion can increase quitting rates without exacerbating psychiatric symptoms. Other studies show similar positive results for varenicline, and in one meta-analysis, the quitting rates were four times that of the placebo group. There is less evidence for other medications, such as nicotine replacement. There is also promising but limited evidence for behavioral options, with the best evidence being for face-to-face (versus online) interventions.

Metabolic Risk

Screening for metabolic risk occurs at very low rates in patients with schizophrenia. This phenomenon occurs even in the context of treatment with antipsychotic medications associated with risk for disturbances in glucose and lipid metabolism, despite recommendations by the American

Psychiatric Association, the American Diabetes Association, other groups, and FDA warnings. The frequency of baseline glucose and lipid monitoring varies from 8% to less than 30%, and follow-up assessments occur in only 8.8% of patients. The NCEP guidelines recommend that physicians treat patients with diabetes mellitus as aggressively as patients with prior MI or stroke. However, patients with diabetes and schizophrenia are less likely than nonmentally ill diabetes patients to receive standard-of-care treatments.

The American Diabetes Association, the American Psychiatric Association, the American Association of Clinical Endocrinologists, and the North American Association for the Study of Obesity published recommendations for monitoring patients with diabetes in 2004. Table 12.14–5 summarizes these recommendations. The recommendations are for routine assessment, and specific clinical situations may warrant closer monitoring or adding other measures such as glycosylated hemoglobin (HbA_{1C}). For patients with hypertension, for example, the NHLBI recommends monthly monitoring of blood pressure following the initiation of antihypertensive therapy, with follow-up every 3 to 6 months once stabilizing the blood pressure.

Table 12.14–5.
Recommended Monitoring for Patients Taking Atypical Antipsychotics

Baseline, every 4 weeks for 12 weeks, then quarterly
  Weight and BMI
Baseline, 12 weeks, then annually
  Blood pressure
  Fasting blood glucose
Baseline, 12 weeks, every 5 years
  Fasting lipid profile
Baseline then Annually
  History (Personal, family)
  Waist circumference

Data from American Diabetes Association; American Psychiatric Association; American Association of Clinical Endocrinologists; North American Association for the Study of Obesity. Consensus development conference on antipsychotic drugs and obesity and diabetes. *Obes Res*. 2004;12(2):362–368.

The most effective strategy for treating hyperglycemia and other metabolic factors is to switch the antipsychotic. In addition, behavioral interventions, particularly those discussed below to address weight gain and physical fitness, can significantly impact at least specific indices, such as fasting blood glucose levels.

However, if switching antipsychotics and behavioral interventions fail, we should appropriately treat the risk factors using appropriate guidelines. For dyslipidemia, the American College of Cardiology/American Heart Association has developed guidelines for the general population, and these should also be the first step for patients with schizophrenia. Such recommendations include statin therapy and possibly other nonstatin agents such as ezetimibe or a bile acid sequestrant. However, some evidence suggests that patients on antipsychotics do not respond as well as other patients to standard treatments, such as statins, and may require other strategies. Physicians may need to consider off-label strategies such as metformin more aggressively than they might for other patients.

For hyperglycemia, behavioral interventions and metformin have both shown efficacy in patients with schizophrenia, although the data are mixed. As with weight, lifestyle interventions focusing on diet and physical activity may have the most significant benefit.

Weight and Physical Fitness

The most effective treatment for weight gain and related cardiovascular risks appears to be an antipsychotic switch similar to treating metabolic risks. The available evidence supports the strategy of switching from an agent with a high weight gain risk to one with lower risk would help reduce weight.

In addition, behavioral therapy can be an effective method for overweight or obese patients with schizophrenia to manage weight gain. The most successful interventions combine both diet and exercise, showing positive improvements in BMI and waist circumference.

Until recently, medications for weight loss had at best a modest effect. However the introduction of the glucagon-like peptide-1 receptor agonists offers a promising alternative which is being explored. There are currently 7 randomized controlled trials, and a pooled analysis from gives preliminary support for the safety and efficacy of these options, although additional studies are needed.

Increasing evidence points to excessively low physical fitness in this population, representing another modifiable CVD risk factor and an intervention opportunity. Numerous studies support the use of physical exercise, and it has the benefit of improving cardiovascular fitness and improving some of the core symptoms of schizophrenia and overall quality of life. However, the studies vary in the magnitude and significance of their results. This variation likely relates to the length of the studies and the intensity of the intervention. In addition, research is ongoing to determine whether any particular approach (e.g., strength vs. aerobic training) is best. However, it seems likely that any program that the patient will accept can be helpful.

The Role of the Psychiatrist in Aiding Prevention

Although psychiatrists usually prefer to defer to a primary care physician for many of the treatments described here, the psychiatrist should remain involved to encourage the rest of the treatment team to pursue aggressive treatment and avoid the therapeutic nihilism that may accompany the care of mentally ill patients. Thus, psychiatrists should continue to play a role in the care, not to help avoid undertreatment but to guide the medical team regarding the unique challenges inherent in patients with a severe mental illness.

Improvements in primary and secondary prevention offer the most significant potential impact on medical mortality in schizophrenia patients. However, clinicians will need to learn more about the potential risks to achieve such improvements better. We will need careful attention and more research on the appropriate role of cardiometabolic risk assessments during psychiatric treatment. We should also focus on the role of adjunctive pharmacotherapies and behavioral and physical fitness interventions that might target excess weight, dyslipidemia, hypertension, hyperglycemia, and smoking in this population. Without a greater degree of collaboration in the future between primary health care providers, endocrinologists, cardiologists, and psychiatrists, the significant burden of avoidable premature mortality from CVD in patients with severe mental illness is likely to continue and increase in severity.

SUMMARY AND RECOMMENDATIONS

Schizophrenia patients have elevated rates of medical comorbidities such as heart disease, cancer, and diabetes, in addition to other prevalent concomitant complications such as substance use. These commodities account for much, if not most, of the morbidity and mortality affecting these patients. Therefore, addressing prevention, screening, referral, and treatment of medical comorbidities is essential to improving health outcomes among patients in this population. In addition, there is a need for strategies that can provide clear guidance to administrators and medical directors on incorporating changes into already existing medical and psychiatric systems.

Although the information on the interactions of schizophrenia and other medical disorders grow exponentially, we can do a better job at standardizing methods and approaches so that pooled data will be more helpful. Even better would be more extensive prospective studies from different countries targeting different racial and ethnic groups.

As care providers, we have a particular responsibility to the many components of medical risk that we cause through our treatments. The Institute of Medicine has suggested that all psychiatric and medical systems involved in the care of patients with schizophrenia, including mental health, substance abuse treatment, general health care, and other services, need to effectively collaborate to coordinate the care of their patients and eliminate gaps and redundancies in necessary services. This collaboration should include establishing structured and routine comorbidity risk assessments, in addition to scheduled monitoring of antipsychotic treatment for side effects and adherence issues. Most importantly, the medical comorbidities of schizophrenia should receive the same established interventions provided to the general population. Implementation of such screening before the initiation of antipsychotic treatment and during treatment will likely improve outcomes among individuals with schizophrenia.

Although there is likely no perfect model for health care delivery to the seriously mentally ill, one crucial model is the integrated clinic in which medical care providers are on-site, colocated with psychiatrists and other team members. Data from collaborative care programs suggest that such an approach can improve service delivery and medical outcomes without increasing total costs. Regardless of the exact model, the psychiatrist will need to play a central role in a schizophrenia patient’s overall care. In many cases, the psychiatrist or other mental health team member has the best therapeutic relationship with the patient. The team should use that relationship to help assure adherence to all aspects of care, not just psychiatric. The mental health team member's role may involve coordination of care with a patient's primary care physician. Case management can also help immensely and appears to be a crucial ingredient for comprehensive treatment in patients with SMI and medical comorbidities.

There has been some optimism about the potential role of the medical home. The Agency for Healthcare Research and Quality (AHRQ) defines the medical home as a model of care with five functions and attributes: Comprehensive care, a patient-centered approach, coordination of care, accessible services, and a commitment to quality improvement and safety. Although such an approach could provide the type of coordination needed to help patients with schizophrenia, in practice, many of the existing examples do not include treatment for SMI. Some of the barriers are practical—primary care providers are often already overwhelmed and do not have the needed time to provide the counseling and monitoring needed for chronically mentally ill patients. In addition, there remain financial obstacles, and current attempts at mental health parity continue to fall short of the mark. The AHRQ has recommended several central policy and programmatic shifts to incorporate mental health care into the medical home model. These include such essentials as normalizing mental health into the mainstream of medical practice, integrating reimbursement mechanisms so that mental is not financially "carved out" from the rest of the health care system, creating a coordinated road map and plan to address complex mental health problems and then disseminate these to primary care practitioners.

In the short term, the existing mental health care system will have to put forth a significant effort, reallocating existing resources to coordinate screening, interventions, including needed referrals and follow-up monitoring. In addition, mental health care providers and systems will need to improve working relationships with primary care and specialty collaborators, including proactive efforts to facilitate patient evaluation and follow-up. Finally, future collaborations between primary health care providers, medical specialists, and psychiatrists will be required to reduce the enormous burden of avoidable premature mortality in patients with schizophrenia and reduce health care disparities between schizophrenia patients and the general population.

Recommendations

- Improve research on risk factors and interventions
  - Standard protocols
  - Standard measures
  - More prospective studies
  - Larger samples
- Improve and coordinate care between primary care and mental health specialists
- Screen for common medical comorbidities
- When discovered, treat comorbidities aggressively
  - Medication changes
  - Education
  - Pharmacologic interventions
  - Behavioral interventions
- Advocate for more resources for integrated care for the mentally ill.

ACKNOWLEDGMENTS

This chapter is a revision of the previous edition's chapter, authored by John W. Newcomer, M.D., Martin T. Strassnig, M.D., Peter A. Fahnestock, MD, and Michele L. Pergadia, Ph.D. The current chapter preserves much of their content and prose, and the author is indebted to the previous authors for their excellent scholarship and literary skill. Although much has happened since they wrote their chapter, and many facts and statistics are updated, the conclusions and recommendations of the prior authors remain essentially unchanged.

FURTHER READINGS

Agarwal SM, Vernon AC, Venkatasubramanian G, Hahn MK. Editorial: Cardiovascular and physical health in severe mental illness. *Front Psychiatry*. 2021;12:760250.

Arnold MJ, O'Malley PG, Downs JR. Key recommendations on managing dyslipidemia for cardiovascular risk reduction: stopping where the evidence does. *Am Fam Physician*. 2021;103(8):455–458.

Chan JKN, Wong CSM, Or PCF, Chen EYH, Chang WC. Diabetes complication burden and patterns and risk of mortality in people with schizophrenia and diabetes: A population-based cohort study with 16-year follow-up. *Eur Neuropsychopharmacol*. 2021;53:79–88.

*Charlson FJ, Baxter AJ, Dua T, et al. Excess mortality from mental, neurological, and substance use disorders in the Global Burden of Disease Study 2010. In: Patel V, Chisholm D, Dua T, Laxminarayan R, Medina-Mora ME, eds. *Mental, Neurological, and Substance Use Disorders: Disease Control Priorities*. 3rd ed (Volume 4). The International Bank for Reconstruction and Development/The World Bank; 2016. Chapter 3.

Correll CU, Solmi M, Veronese N, et al. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. *World Psychiatry*. 2017;16(2):163–180.

Fan Z, Wu Y, Shen J, Ji T, Zhan R. Schizophrenia and the risk of cardiovascular diseases: a meta-analysis of thirteen cohort studies. *J Psychiatr Res* 2013;47:1549–1556.

*Ferrari A. Global, regional, and national burden of 12 mental disorders in 204 2 countries and territories, 1990–2019: a systematic analysis from the 3 Global Burden of Disease Study 2019. *The Lancet Psychiatry (prepress)*. Accessed October 28, 2021. http://www.healthdata.org/sites/default/files/files/research_articles/2021/GBD_Mental-Disorders_Research-Article.pdf

*Firth J, Siddiqi N, Koyanagi A, et al. The Lancet Psychiatry Commission: a blueprint for protecting physical health in people with mental illness. *Lancet Psychiatry*. 2019;6(8):675–712.

García-Garcés L, Lacamara Cano S, Cebolla Meliá Y, et al. Comparison of three different exercise training modalities (aerobic, strength and mixed) in patients with schizophrenia: study protocol for a multicentre randomised wait-list controlled trial. *BMJ Open*. 2021;11(9):e046216.

Hannoodee H, Mahmoud AK, Nyein WYT, et al. The outcomes of acute coronary syndrome in patients suffering from schizophrenia: a systematic review. *Cureus*. 2021;13(8):e16998.

Kanagasundaram P, Lee J, Prasad F, et al. Pharmacological interventions to treat antipsychotic-induced dyslipidemia in schizophrenia patients: a systematic review and meta analysis. *Front Psychiatry*. 2021;12:642403.

Karaoulanis SE, Christodoulou NG. Do patients with schizophrenia have higher infection and mortality rates due to COVID-19? A systematic review. *Psychiatriki*. 2021;32(3):219–223.

Khaity A, Mostafa Al-Dardery N, Albakri K, et al. Glucagon-like peptide-1 receptor-agonists treatment for cardio-metabolic parameters in schizophrenia patients: a systematic review and meta-analysis. *Front Psychiatry*. 2023;14:1153648.

Khokhar JY, Dwiel LL, Henricks AM, Doucette WT, Green AI. The link between schizophrenia and substance use disorder: A unifying hypothesis. *Schizophr Res*. 2018;194:78–85.

Li H, Li J, Yu X, et al. The incidence rate of cancer in patients with schizophrenia: a meta-analysis of cohort studies. *Schizophr Res.* 2018;195:519–528.

Liu L, Ni SY, Yan W, et al. Mental and neurological disorders and risk of COVID-19 susceptibility, illness severity and mortality: A systematic review, meta-analysis and call for action. *E Clinical Medicine*. 2021;40:101111.

Minhas S, Patel JR, Malik M, Hana D, Hassan F, Khouzam RN, et al. Mind-body connection: cardiovascular sequelae of psychiatric illness. *Curr Probl Cardiol*.

2021;00:100959.

Mitchell AJ, Vankampfort D, Sweers K, Winkel RV, Yu W, Hert MD. Prevalence of metabolic syndrome and metabolic abnormalities in schizophrenia and related disorders—A systematic review and meta-analysis. *Schizopr Bull*. 2013;39(2):306–318.

Moore JX, Chaudhary N, Akinyemiju T. Trends in metabolic syndrome prevalence by Race/Ethnicity and sex in the US: National health and nutrition examination survey, 1988–2012. *Prev Chronic Dis.* 2017;14:E24.

Newcomer JW, Hennekens CH. Severe mental illness and risk of cardiovascular disease. *JAMA*. 2007;298:1794–1796.

Newcomer JW, Weiden PJ, Buchanan RW. Switching antipsychotic medications to reduce adverse event burden in schizophrenia: Establishing evidence-based practice. *J Clin Psychiatry*. 2013;74(11):1108–1120.

Palmer VJ, Lewis M, Stylianopolous V, Furler J. Primary care prevention of the cardiovascular health crisis for people with severe mental illnesses: The elephant in the room. *Aust J Gen Pract*. 2018;47(12):846–850.

Park SJ, Wai A, Pavithran K, Kunheri B, Valsraj K. Cancer and severe mental illness in low- and middle-income countries: The challenges and outlook for the future. *Psychooncology*. 2021;30(12):2002–2011.

*Patel V, Chisholm D, Parikh R, et al. Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities, 3rd edition. *Lancet*. 2016;387(10028):1672–1685.

Polcwiartek C, Loewenstein D, Friedman DJ, et al. Clinical heart failure among patients with and without severe mental illness and the association with long-term outcomes. *Circulation. Heart Failure.* 2021;14(10):e008364.

Solmi M, Fiedorowicz J, Poddighe L, et al. Disparities in screening and treatment of cardiovascular diseases in patients with mental disorders across the world: Systematic review and meta-analysis of 47 observational studies. *American J Psychiatry.* 2021:178(9):793–803.

Spanakis P, Peckham E, Young B, Heron P, Bailey D, Gilbody S. A systematic review of behavioural smoking cessation interventions for people with severe mental ill health—what works? *Addiction*. 2021;117(6):1526–1542.

Stubbs B, Koyanagi A, Veronese N, et al. Physical multimorbidity and psychosis: comprehensive cross sectional analysis including 242,952 people across 48 low- and middle-income countries. *BMC Medicine*. 2016;14:189.

Stubbs B, Vancampfort D, De Hert M, Mitchell AJ. The prevalence and predictors of type two diabetes mellitus in people with schizophrenia: a systematic review and comparative meta-analysis. *Acta Psychiatr Scand*. 2015;132:144–157.

Teh WL, Cetty L, Jeyagurunathan A, et al. Comorbid physical illnesses in adult outpatients with psychotic disorders: risk factors, psychological functioning, and quality of life outcomes. *Social psychiatry and psychiatric epidemiology*. 2021;56:9:1633–1643.

Tsoi DT, Porwal M, Webster AC. Interventions for smoking cessation and reduction in individuals with schizophrenia. *Cochrane Database Syst Rev*. 2013;2:CD007253.

Vancampfort D, Correll C, Galling B, et al. Diabetes mellitus in people with schizophrenia, bipolar disorder and major depressive disorder: A systematic review and large scale meta-analysis. *World Psychiatry*. 2016;15(2):166–174.

Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. *JAMA Psychiatry*. 2015;72(4):334–341.

Ward M, Druss B. The epidemiology of diabetes in psychotic disorders. *The Lancet Psychiatry*. 2015;2(5):431–451.

12.15 Recovery in Schizophrenia

DINA HOOSHYAR, M.D., M.P.H., CHACKU MATHAI, A.A.S., DREW MUSA, C.P.S., AND MARK RAGINS, M.D.

DEFINING RECOVERY AND ITS EVOLUTION

By the 1980s and 1990s, severe problems were emerging in the public community mental health system, and the dream of deinstitutionalization was fading. Failures were often attributed to a lack of funding and political will, but suspicions also began to arise that the “medical model” itself was inadequate to meet the needs of community-based mental health services. Fueling the dissatisfaction were outcome studies that showed that, contrary to common knowledge and expectations, people with schizophrenia recovered to live reasonably full lives, free from seriously disabling symptoms about two-thirds of the time. Why were people settling for stabilization and isolation if recovery is not just possible but probable? Maybe the diagnosis of schizophrenia should not impose a destiny of lifelong treatment, medication, disability, and dependence.

Alternative programs and approaches began to be developed to pursue better outcomes and lives:

- Ex-patient–controlled support groups, crisis respite, homeless outreach, and rights advocacy organizations, such as the Insane Liberation Front in Portland, Oregon, founded in 1970, the Mental Patients Liberation Project in New York City, and the Mental Patients Liberation Front in Boston (both founded in 1971), and the Network Against Psychiatric Assault in San Francisco in 1972 and other groups, demanded and obtained inclusion of their self-help initiatives in the National Institute of Mental Health’s (NIMH) Community Support Program division conferences in the late ‘70s.
- Fountain House in New York, clubhouses, and psychosocial rehabilitation programs addressed people’s loss of usefulness and meaningful, productive roles.

- Psychiatric rehabilitation championed at University of California at Los Angeles by R.P. Liberman and at Boston University by Bill Anthony used behavioral and medical rehabilitation techniques to build a variety of skills, including educational, employment, housing, and social skills.
- “Housing first” programs and “supported” programs with a variety of quality of life goals—for example, housing, employment, education, parenting, physical health care—emerged that focused on people learning by doing instead of being given treatments to prepare them to do things.
- Integrated dual diagnosis approaches for people with both mental illnesses and substance abuse ranged from Fred Osher’s early efforts at the NIMH to Ken Minkoff’s integrated systems of care to “harm reduction” and Miller and Rolnick’s “Motivational Interviewing.”
- Trauma-informed approaches were developed to reduce retraumatizing people while serving them and have been used to radically decrease seclusion and restraints in inpatient settings.
- Client choice, empowerment, and shared decision-making strategies were developed and used to enhance engagement with and participation in services.

Taken individually, each of these techniques adds to the “tool kit” of effective treatment approaches. More strikingly, however, was how they organically coalesced the Recovery Movement by the mid-1990s, no longer isolated “outsiders” but now part of a strong nationwide movement.

Previously within psychiatry, the concept of recovery had been used primarily within the context of substance-abuse recovery. Now “recovery” was being repurposed for serious mental illnesses.

The word “recovery” itself brings along with it a variety of traditions, histories, and values ranging from 12-step recovery to recovery from physical illness, to trauma recovery, to spiritual recoveries. This broad spectrum of grassroots, “bottom-up” influences gives the recovery movement impressive diversity and practicality and widespread appeal. It also makes it unwieldy. Even defining “recovery” has been difficult.

Two “consensus” approaches to defining recovery have been put forward by the Substance Abuse and Mental Health Services Administration (SAMHSA). The first, released in 2006, emphasized the “individual journey” aspect of recovery: “Mental Health Recovery is a journey of healing and transformation enabling a person with a mental health problem to live a meaningful life in a community of his or her choice while striving to achieve his or her full potential.” This approach emphasizes the individual’s struggles and triumphs in overcoming mental illnesses and conceptualizes service providers as guides, coaches, and fellow travelers promoting recovery journeys.

The second approach is to define recovery as achieving a quality of life despite the burdens of mental illness. SAMHSA’s 2011 “working definition” of recovery was: “A process of change through which individuals improve their health and wellness, live a self-directed life, and strive to reach their full potential.” It goes on to define four domains to support a life in recovery: *Health:* overcoming or managing one’s disease(s) as well as living in a physically and emotionally healthy way; *Home:* a stable and safe place to live; *Purpose:* meaningful daily activities, such as a job, school, volunteerism, family caretaking, or creative endeavors, and the independence, income and resources to participate in society; and *Community:* relationships and social networks that provide support, friendship, love, and hope.

JOURNEY OF RECOVERY

From an illness-centered point of view, the treatment of schizophrenia is based upon the course of the illness: It may begin with a prodromal phase before a “first break.” There is often a period of untreated psychosis before multiple attempts to acutely stabilize the illness are undermined by relapses. Over time the illness may stabilize and go into a residual phase or even remission, though acute relapses can always recur.

From a person-centered point of view, the service needs to depend less upon the diagnosis or stage of the illness and more upon how the illness is impacting the person and how they are relating to it—for example, out of control and at high risk, ambivalently engaging with helpers, taking self-responsibility and building coping skills, incorporating the illness into a rebuilt self-identity, or adapting to reconnect to meaningful community roles.

There are perceptive autobiographies written by people who have worked hard to recover from schizophrenia that everyone can all learn from, ranging from Joanne Greenberg’s fictionalized classic “I Never Promised You a Rose Garden,” describing her relationship with her psychiatrist Frieda Fromm-Reichmann at Chestnut Lodge and her return to the community in the premedication, psychoanalytic era, to the more contemporary “The Center Cannot Hold” describing Elyn Saks’ ongoing struggles with psychosis even as she has become a law school professor and associate dean at University of Southern California.

Patricia Deegan built her exemplary career promoting recovery practice upon her own experiences and her taped interviews of people describing their own recoveries. Based on these “lived experiences” of recovery, she created her “conspiracy of hope” promoting self-help “Coping with Voices” techniques and her “Common Ground” program promoting “personal medicine” and shared decision-making with psychiatrists. The approach of building strategies on personal stories and experiences has expanded to families in NAMI (National Alliance on Mental Illness) with their “In Our Own Voice” programs. Sharing personal experiences has proved a powerful tool for bridging differences and reducing antagonism to work effectively together.

The best way to learn about how someone’s brain functions is to ask them and then listen deeply. Their narratives tell us about important details of the recovery process and how to help people in their journeys while appreciating that everyone’s journey is unique. Listen deeply to the self-reflection and advice from one of our coauthors with lived experience of psychosis:

My name is Drew Musa. I'm a 27-year-old native New Yorker, and I've been diagnosed with schizophrenia. Due to my diagnosis, every day looks different. I can never expect what my mind is going to do every day. I experience auditory hallucinations where I hear things that others cannot. What I hear depends on where I am at that exact moment. I may hear voices through the TV, the radio, or just ambient sounds. My training as a peer specialist gave me the confidence to still have hope for the future regarding living with a diagnosis. It was through my Howie the Harp training that I learned to believe in myself again. Now thanks to the training, I know I can still live a normal life and not be ashamed of my diagnosis. When I first heard my diagnosis, I had no idea what it meant for me or if I would have a life worth living, but now I am no longer ashamed of my diagnosis. I see it as something I just live with.

The most challenging part of my recovery journey was the beginning. I was frustrated and scared that I was hearing things other people could not, which led to my first break. Fortunately, I've learned a lot since then. I've come a long way since my first hospitalization. The medication I take now is working for me. Now I always try to keep myself busy, whether it's playing games on my phone, cooking, or watching movies. I'm thankful I have such a great family who got me help as soon as possible. Thanks to my family, we found the OnTrack program (New York State's Office of Mental Health, Coordinated Specialty Care model for First Episode Psychosis). OnTrack is unlike any clinic I've been to. The program is based on peer-centered recovery, meaning there are no decisions made about me without me. The client is in the driver's seat, and the Ontrack staff is there to offer support and give their clients the tools to succeed. In addition, they helped me look for a job and offered me a music group that I go to regularly. I also speak with a therapist and a psychiatrist regularly as well. The coolest part of OnTrack is the way they treat me as a person. The biggest lesson I learned from OnTrack is to ask for help when I need it. I have great supports between OnTrack and my family, so I should utilize them as much as possible. I learned to be honest about what I'm experiencing and never be afraid to ask for help which is important because dealing with a mental health diagnosis on your own is difficult.

> If I could go back in time and give myself advice, I’d say don’t be afraid to ask for help, and things may seem bleak now, but things will get better. That’s the type of support I wish someone would’ve given me when I was newly diagnosed. As a peer, I now get to support people the way I wish I was supported. I try to tell newly diagnosed clients to be proactive in their recovery. You shouldn’t let people make decisions about your recovery without your consent, and don’t be afraid to ask questions when you don’t understand something like “what does the medication do exactly?” “What does that medical term mean?” and “Can we try a different medication?”

Elizabeth Kubler Ross made a groundbreaking shift by identifying five stages that a person goes through while dying (as opposed to stages the cancer illness goes through while it’s spreading and killing someone). One of the authors, Mark Ragins, has analogously described four common stages people go through while recovering from serious illnesses or destruction of any kind in his short book *A Road to Recovery.*

- Hope: Recovery begins with hope. To be motivating, hope has to be more than just an ideal. Hope has to take form as an actual image of how things could be if they were to improve. It must be personally meaningful for them to pursue it as a goal, persisting despite inevitable setbacks.
- Empowerment: To move forward, people need to have a sense of their own capability, their own power. Their hope needs to be focused on things they can do rather than new cures or fixes someone else will discover or give to them. It is often needed for someone else to believe in them before they’re strong enough to believe in themselves and to start focusing on their strengths instead of their losses. It also often takes some actual experience of success to really believe one can be successful. Waiting until someone is ready to move on can often be stagnating and disempowering because “readiness” often occurs only in retrospect after something has been done successfully.
- Self-Responsibility: At some point, most people who recover realize that no one else can do it for them, that they have to take charge of their Recoveries. People can be supported, but they can’t be caretaken or protected into recovery. Taking one’s own risks, setting one’s own goals and path, and learning one’s own lessons are essential parts of recovery.

- Meaningful Roles in Life: Ultimately, to recover, one must achieve some meaningful role apart from schizophrenia. Connectedness to other people, belonging, and feeling accepted, which may have for a while only been possible with others who have mental illnesses, within families, or with compassionate helpers, becomes possible in a variety of contexts. The isolation is increasingly broken, and life re-entered.

The journey of recovery should take place primarily in the community rather than in isolated islands of tolerance and asylum. The overarching social culture, as well as many local community cultures, often make a recovery difficult. Illness-centered, psychoeducational, antistigma efforts are likely less effective than relationship-based storytelling efforts. The overall goals are to increase community acceptance, inclusion, compassion, relationships, and belonging for people with schizophrenia. Since the middle ages, when a spiritual healing center for people with mental illnesses was established there, the small town of Gheel, Belgium, has maintained a pervasive counter-culture integrating people with severe mental illnesses within their families and community, proving it can be done.

Illness-Centered versus Person-Centered Recovery

The medical model conceptualizes the process of recovery as:

- The fundamental problem is the symptoms of their mental illness.
- Diagnosis, assessments, treatment, monitoring, and achieving success are focused on symptom reduction.
- Once the illness is treated, other services may be needed to help people deal with residual symptoms, keep symptoms away, and to rebuild lost functions or quality of life.
- Some fortunate people respond to treatment and other services to achieve recovery.

In contrast, the recovery model conceptualizes the process of recovery as:

- Mental illnesses are a highly individualized, complex personal reaction to a variety of biologic, emotional, developmental, traumatic, and social influences.
- Each person’s outcome is more dependent upon how they personally react, adapt, and cope than it is dependent upon the type or severity of

the symptoms themselves.

- Promoting positive personal reactions (e.g., maintaining hope, self-belief, managing risks, engaging and collaborating with helpers, taking self-responsibility, being motivated to build skills and supports, finding personal and social meanings, and avoiding isolation) throughout the course of each person's illness is crucial.
- Some fortunate people find enough inner strengths and resilience to make personal changes to achieve recovery.

These differences in conceptualization translate into three major transformations in service delivery: (1) from illness-centered to person-centered: *Recovery is something a person does, not what an illness does*. (2) From professionally driven to client-driven: *Recovery isn't something a professional can do to a person. It is a process they must engage in.* And (3) from deficit-based to strength-based: *Recovery doesn't happen when the person never has any more symptoms, even serious ones, but when they have the strength and resilience to handle whatever symptoms occur.*

Perhaps counterintuitively, this difference means that the recovery model is most important for people whose illnesses are most difficult to treat medically: traditionally, though perhaps not accurately, schizophrenia. The strength of the person is more relevant than the strength of their illness.

Within the medical model, persistent psychosis is viewed entirely as a reflection of brain abnormalities that lead to hallucinations, delusions, thought disorders, and often "negative symptoms" that must be addressed with antipsychotic medications. The recovery model is more likely to include the entire person in its conceptualization, leading to individualized, multidimensional personal formulations of the development and recovery from psychosis.

One of the author's current models is "The Psychosis Triangle" (Fig. 12.15–1).

The three interacting dimensions that lead people to have psychosis are as follows:

1. **Experiencing Reality**: There are changes in how they are experiencing "reality"—for example, hallucinations, delusions, paranoia, or spiritual experiences as well as changes in how they process reality, like making idiosyncratic connections or relying on irrational "primary process" thinking.

2. **Self-Identity**: There are changes in their internal processing and self-identity—for example, how they hold themselves together, balance rational and irrational portions, self-observation, or create their own stories. It can also include “anhedonia,” the inability to feel joy or pleasure in life.
3. **Relationships**: There are changes in how they relate to other people—for example, disconnected, missing “normal” social signals, hard to relate to, socially anxious, avoidant, or even paranoid.

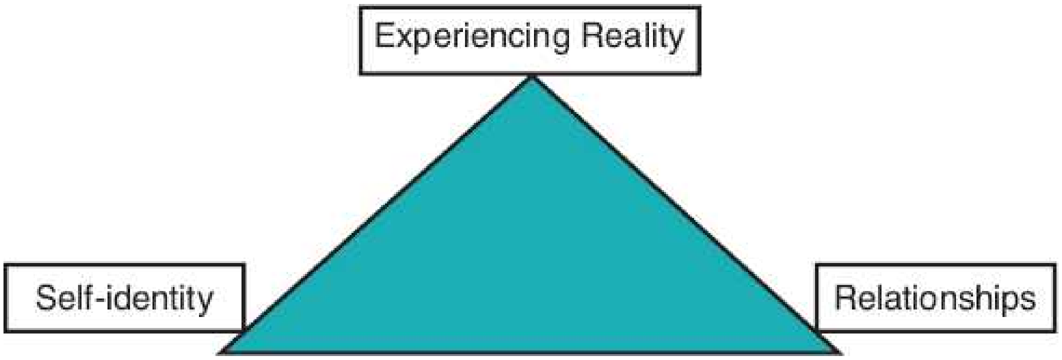

FIGURE 12.15–1. The Psychosis Triangle: Interacting dimensions of psychosis.

None of these dimensions falling apart on its own creates psychosis. All three of them must be seriously impaired for someone to become psychotic. Any of the three can act as the initial weak point, putting its neighboring dimensions at risk. Once two dimensions are compromised, the third dimension in the triangle is under severe strain, and if it also falls, psychosis is the result—or it could be a stronghold to rebuild the other two dimensions, leading to recovery. The “prodromal” period is usually a time when self-identity and relationships are deteriorating, making the person vulnerable to distortions in reality.

In contrast with the medical model that only focuses on the Experiencing Reality dimension, often sacrificing the other two dimensions along the way, the recovery model is open to all three dimensions.

Using all three dimensions allows us to integrate and appreciate not just biologic factors but also personal meaning, such as hope creation, role modeling, identity formation, and social factors, like structural racism, mentoring, and productive social integration.

One of our authors, Chacku Mathai, is a thoughtful peer advocate who effectively integrates all three dimensions (adapting to the unusual way he experiences reality including hearing voices, some of which are more trustworthy than others, intersecting personal identities, and navigating relationships within a racial, immigrant context). In this section, he models synthesizing his unique personal journey and experiences with the overarching social and professional contexts that envelop all of us.

Understanding the experiences of people diagnosed with schizophrenia and/or other psychotic disorders, especially experiences of recovery, is an incredible challenge due to the wide heterogeneity of experience, even among those with the same diagnosis. To add to the challenge, the history of psychiatry, psychiatric diagnosis, and the evolving characterization and naming of mental illness, psychosis, and schizophrenia are inextricably linked to the social and political forces, such as racism, that marginalized and oppressed certain groups throughout our history.

Accordingly, the authors in this section intend to offer a more personal, narrative approach to engaging these concerns in order to support and encourage greater reflection on structural practices and competencies that might matter most in supporting recovery. This is in contrast to a less effective narrative-sharing approach to address “stigma,” which in this author’s view is an inadequate and even euphemistic term that obscures the path to understanding the true depth and impact of bias, discrimination, and oppression in people’s lives.

Our hope is that the reader will experience the start of a uniquely crafted path toward an ongoing reflection and analysis of recovery experiences, including the journey of meaning-making in psychosis, dynamic qualities of a recovery-oriented approach, and the role of psychiatry in addressing social and political determinants of health to eliminate the racial disparities and inequities that continue to persist in psychiatry.

A Journey of Sense-Making With Psychosis

I struggled with an inner world that no one would understand, eventually resulting in a suicide attempt and drug overdoses when I was 15. Even after several years of abstinence from any of the substances I used to quell the stormy experiences within, I was struggling with hearing troubling voices and other extreme states. I saw shadowy figures jumping around my room, one of them always trying to enter my body. The only way to stop her was to scream. I also struggled with a belief that my parents were conspiring with others to poison my food and get rid of me. I caught my mother watching me eat. It was then that a convincing voice told me what she was doing. Although they denied it, I knew secrets about the family and broader society, including how teachers and schools were controlling people and our minds. Despite losing nearly a full year of high school to an intensive, long-term, outpatient day treatment program, I still managed to return and graduate high school with the advocacy and support of my family, although by a thin margin. Despite my suspicions and poor grades, I started some community college courses, only to then drop out in the middle of the semester.

The voices had finally convinced me that my father, and now the community college soccer coach, were conspiring against me. I struggled to find and keep work. Despite my limited income, I moved out of my parent's house to live with an older and supportive friend in his apartment. After volunteering a bit at a local animal shelter, I eventually found part-time work at a YMCA and took odd jobs swinging mops and delivering pizzas. I eventually found some full-time work in a local warehouse, loading trucks and shipping damaged or defective goods back to their manufacturers. My title was the Damaged and Defective (D and D) Clerk, and the irony of the narrative was not lost on me. As South Indian immigrants, my family and I did not plan for any of this. I felt like such a failure in our great experiment with life in America. I was supposed to be the first in my family to immigrate to the United States and study at a prominent university.

Through the mid to late 80s, I accepted my father's urging for me to check out a local teen recovery center he learned about from an acquaintance at his church. It was an old, renovated horse barn owned by a church across the parking lot from a high school. It was the first of its kind anywhere and was started by a group of families as a place for their teenage children to support each other as they returned to school and home from mental health and substance use treatment centers. I even surprised myself and my family as I discovered a peer community I could trust and a sense of purpose to pursue in this place we called “the Barn.”

During that time, I managed to start some of the first-ever peer support meetings and advocacy groups in that region at treatment centers, group homes, and even the local psychiatric center. I was asked to deliver public presentations about mental health and substance use conditions at schools, treatment programs, and police departments, calling for changes in the way we were perceived and responded to. I organized and spoke at local hearings with Congressional representatives and state legislators. I was even given an award by a local television station and the Junior League of Women Voters for my volunteer civic activities.

In 1990, as a 20-year-old local advocate with a personal psychiatric and substance use history, I attended a mental health fundraiser in my hometown. The guest speaker was Lionel Aldridge (February 14, 1941 to February 12, 1998). He was a Super Bowl Champion Defensive End for the Green Bay Packers, one of the pioneering Black men in the National Football League (NFL) in the 1960s, and a sports broadcaster in the 70s. He was also the first professional football player to share his story as a person diagnosed with paranoid schizophrenia as a mental health advocate.

Mr. Aldridge was speaking during the culmination of an era that strongly associated mental illness, substance use, and those of us who were Black, Indigenous, People of Color (BIPOC) with violence, crime, and poor moral character. I rushed to the front of the room after his presentation. Although his story was definitely inspiring on many levels, my hope was that he would answer some big questions that I was struggling with that whole year. Despite all the advocacy efforts, I still felt quite alone in my experience and perspective. I had no one left to ask. While I was aware of the family-run local and national mental health advocacy groups he was involved with, at the time, called the National Alliance for the Mentally Ill (NAMI), I was still just learning about other ex-patient–controlled groups that were organized around the country and would eventually be referred to as consumer/survivor/ex-patient advocacy groups. Until Mr. Aldridge, I had also yet to meet another person of color who spoke publicly about mental health and wondered why.

As the line of people waiting to speak with Mr. Aldridge grew longer behind me, I realized that I would only have his attention for one question. It was my turn next, and as I frantically tracked through my now racing thoughts, like a person trying to find a friend across a crowded room, one of my most trusted voices softly suggested, “Ask him about the cycles.”

She was referring to something my clinicians would say in our sessions and the various psychoeducation groups that my family and I attended. They explained what I was going through with illness-oriented language, such as decompensation, relapse, and remission. One of the main features of this explanatory model and approach to relapse prevention was that the course of mental illness was cyclical and deteriorating in nature and that if I could become aware of the early warning signs, I might have a better chance of getting the help I needed to manage it before things got worse for me. While I certainly appreciated the idea of a proactive plan for when I might be struggling, I was terrified by the idea that I might have to go through repeating cycles and waves of my worst experiences, only to eventually drown and require more treatment as I struggled beneath them. I kept hearing this theme, especially as a way to explain why one of my friends was struggling, had to go to the emergency room, or get hospitalized against their will.

This also just didn't jive with my own observation of my experiences; however, I couldn't yet articulate my take on this, and there didn't seem to be anyone else who might understand. So, after introducing myself as a psychiatric survivor/ex-patient who was diagnosed with schizoaffective disorder and cooccurring substance use disorders, I asked Mr. Aldridge, "What do you think about this idea that we are in a remission cycle of illness that gets worse over time, and that will eventually repeat itself as a relapse? Is this true?"

Mr. Aldridge looked at me with deep kindness and empathy that I will never forget. I remember noticing during his speech that he seemed to be free in a way that I only hoped could be true for me. In this moment with me, it seemed as though he was reaching back to reflect on a time when he was battling when he wasn't so free. He had no hesitation and responded with full authority. He also seemed to be speaking to all of us who were in line, "No man, that's not how it goes. The people that say this are looking at us from the outside. It only looks like the same old thing to them. We know that we are growing. We grow up, we grow out, we grow sideways, we grow down." As he was talking, I was imagining a sphere of my experiences, each experience as a swirling revolution of growth in every direction. That made more sense to me. They were different experiences…every time. Of course, they would look like cycles to the outside observer, or even me, especially if all we noticed were my repeated and seemingly failing ways of dealing with them. I would eventually learn that if I could understand the cycles as revolutions of growth and catalysts for change, I could also discover the unique gifts within them and new ways of going through them with the outside world.

This was one of several life-changing milestones in my own journey of sense-making about my experiences and led me to discover even more healing strategies, communities, and resources to turn to over the years. Even my family, who struggled to understand what I was going through as well, were willing to support my own strategies for recovery and reject the commonly held notions of the psychiatric system that I would be better off hospitalized, treated with psychiatric medications, and supported in the community with regular monitoring and supervision. Despite years of continuing conflict and my lack of trust in their motives, my parents still expressed their belief in me and supported my self-determination and choices.

What I Really Wanted

During my early treatment experiences, one of my therapists asked me if I could recall a time when I was happy. I remember thinking that this was one of the better questions she had asked. Although I still didn't answer her, my mind took flight toward a vivid memory that has never left me. Like a bird arriving at her favorite tree, I stood quietly as I watched myself, a 6-year-old boy, reach up to touch and gently hold the ear of the young elephant kneeling before me.

I was born in Kuwait. We had just started our life in the United States and were now on a visit to my parents' respective home villages in Kerala, India. At the time, elephants were used to clear the trees for farming and, noticing my delight at seeing them, my father asked one of the local handlers to let me sit on the elephant.

I pulled myself onto the elephant's back without hesitation. There was no question in my mind that this was where I belonged. As I ran my hands over the elephant's skin, I could feel him welcoming my touch. We were connected in a way that I didn't expect. As I looked up, I noticed my parents, uncles, aunts, cousins, local villagers, and the handler, all looking at me with pride and gratitude for my presence. They were seeing me and looking at me with great expectations. It was as if I was already doing something so immense that they couldn't wait to see what I would do next. What I could do next. There was a belief in me and what I was already offering to the world expressed in their eyes, by their gaze, that I will never forget; because this is what I really wanted but knew that I couldn't have in my life back in the United States.

A Structural Formulation of My Own: An Antiracist Approach to Recovery

Whenever I was asked to offer "my story" to various institutions as an advocate, I noticed that there were fundamental differences in how people, even Mr. Lionel Aldridge, were telling their stories publicly and how I wanted to tell my story. Most people offered statements like, “When I first became sick,” “After I was diagnosed,” or “Following my psychiatric hospitalization.” They also seemed to erase any contextual, marginalizing experiences from their story. BIPOC survivors seemed to avoid any references to race or racism. My story, however, would begin with, “When we moved to the United States.” I would then describe the racial tensions of the time, including how my family and I were treated with hostility and violence as people of color or “foreigners,” blaming us for taking White people’s jobs and moving into their neighborhoods. This became my personal strategy for opening the doors to important conversations about recovery, rights, and antiracism as an integrated framework for support.

Opportunities to train and influence communities across the country led me to realize that most of the behavioral health professionals, administrators, policy makers, police, teachers, and the general public still held the false belief that schizophrenia is a deteriorating illness of the brain. This was despite the international, long-term follow-up studies discussed earlier in this chapter that found that a high proportion of people diagnosed with schizophrenia completely or significantly recovered. It also seemed that most people were also not aware of the diverse and international presence of consumer/survivor/ex-patient–led organizations and public disclosure by those of us diagnosed with schizophrenia and other psychotic disorders who successfully live, learn, work, and play meaningful roles in society.

Many also still believed the debunked myth that people diagnosed with psychotic disorders, especially schizophrenia, are violent. This also seemed embedded as a racial bias and attitude that was maintained as a racist attitude of mental health systems, rationalizing policies, and practices that perpetuate significant racial disparities and inequities for BIPOC communities. For example, studies demonstrated that doctors over-diagnosed schizophrenia in African American men four times as often as in white patients, despite what we already know about schizophrenia as a disorder that affects all ethnic groups at the same rate.

An antiracist approach to recovery calls for:

- Commitment to recovery equity in communities of BIPOC
- Transform how we evaluate and understand racial disparities and inequities in behavioral health

- Defining racism as the powerful collection of policies and practices that perpetuate racial inequities and antiracism as the policies and practices that reverse them
- Utilize practice approaches such as structural competence and formulation that recognize that “place matters” and address social determinants of health, such as racism, as factors in how our health is influenced by where we are born, live, learn, work, and play.

Recovery Model Service Delivery

The challenges of both the medical model and recovery model services are the same. We can both work with people at any stage of illness or recovery, but truly integrating person-centered, client-driven, and strengths-based values while focusing on all three dimensions of psychosis leads to very different program designs and emphasis within each stage (Table 12.15–1).

Table 12.15–1. Many of the programs in the below table will be discussed later in the chapter. References are given for further exploration.

The most common “positive” outcomes from illness-centered treatment are someone with repressed psychotic symptoms but limited self-identity, roles, community integration, resilience, and protective factors, leaving them dependent on medications, vulnerable to relapses under stress, and with poor long-term health outcomes. The most common “positive” outcomes from person-centered approaches are someone who may have to continue to deal with psychotic symptoms while rebuilding a self-identity, purpose, relationships, and roles, either with ongoing medications (like Elyn Saks and Fred Frese) or without ongoing medications (like Dan Fisher and Mark Vonnegut).

Table 12.15–1.
Characteristics of the Medical Compared to the Recovery Model

	Medical Model (focus on experiencing reality only)	**Recovery Model** (includes focus on self-identity and relationships)
Public Health	Mental Health First Aid: Teaches people to recognize early signs of mental illness, destigmatize, and promote professional treatment	Emotional CPR: Teaches empathic listening and empowering relationships to help people build their own growth responses to emotional crisis
Individual Prevention	Preemptive antipsychotic medications for people with a family history of schizophrenia, “ultra-high risk”	Headspace: enhances social skills, relationships, and personal emotional skills for youth
Early Intervention “First break”	RAISE: Rapid medication stabilization, using hospitalizations if needed, psychoeducation into the chronic nature of illness and patient role, early rehabilitation to rebuild functioning	Open Dialogue: Intensive work with the natural social network to preserve relationships and roles while the person deals with unusual experiences
Intensive Ongoing Support	ACT: “hospital without walls,” multidisciplinary, community-based team insuring medication treatment and life supports	Full-Service Partnerships: Multidisciplinary, community-based team including specialists that support a wide variety of people’s quality of life goals through collaborative, growth-oriented partnerships building increasing personal self-responsibility
Crisis	Police, urgent care,	Soteria model crisis residential,

Response	Psychiatric Emergency Rooms, and involuntary hospitalizations designed to avoid the risk of danger and control people, usually with medications	peer respite programs designed to give space for people to regain self-control and learn what changes to make to stop suffering
Recovery	Illness Management Recovery: Focusing on chronic illness management	Clubhouse, Voice Hearers: Focusing on building positive personal roles and relationships

RECOVERY OUTCOMES

“Schizophrenia is a chronic disabling illness affecting approximately 1% of the U.S. population” is a familiar introduction to schizophrenia in the medical literature. The pessimistic tone of this introduction does not take into account the heterogeneity of outcomes associated with recovery in this population. Existing challenges to objectively measuring the longitudinal progress of people with schizophrenia include: (1) evolving, nonuniform definitions of schizophrenia across the world; (2) lacking universally accepted criteria for measuring recovery and rather measuring the nonexistence of symptoms associated with schizophrenia; and (3) having self-reinforcing conclusion (i.e., the bias of researcher/clinician that worst outcomes are typical based on the historical, mainstream medical society’s perception).

Although Emil Kraepelin’s intent was to advance the understanding of psychiatric conditions through the categorization of symptoms into manic depression and dementia praecox, a result was dementia praecox (i.e., schizophrenia) became associated with poor outcomes while manic depression with good outcomes. However, 17% and 8% to 13% of dementia praecox cases demonstrated long-term clinical improvement and long-term recovery, respectively. Eugen Bleuler, who coined the word “schizophrenia,” also propagated this sentiment when he revised his initial more optimistic assessment to one never associated with a “full restitutio ad integrum.”

In the 1970s, published studies contested the dogma that schizophrenia equals poor outcomes. The 1987 findings of the Vermont Longitudinal Research Project, which had a more rigorous study design as compared to some of the 1970s studies, supported the findings of these earlier studies. One-half to two-thirds of the Vermont study participants attained considerable improvement or recovery 20 to 30 years post completion of a comprehensive rehabilitation program and planned deinstitutionalization (Table 12.15–1). Vermont Longitudinal Research Project's cohort consisted of clients residing in Vermont State Hospital (*n* = 269) between January 1955 to December 1960 who were considered the most chronically mentally ill among the Vermont State Hospital's client population. They had not responded to chlorpromazine (Thorazine) treatment and had been referred to a comprehensive rehabilitation and community placement program sponsored by the Vermont State Hospital and Vocational Rehabilitation Division of the Vermont State Department of Education.

Despite the evidence that recovery occurs in schizophrenia, the concept that schizophrenia is a progressive disease and associated psychosis may be deleterious to the brain assisted in program development for minimizing the duration of untreated psychosis among young people experiencing a first episode of psychosis. PACE (Personal Assessment and Crisis Evaluation) Clinic was established in 1994 as the first clinical and research clinic in the world designed to prevent or delay the onset of psychosis among persons aged 14 to 25 years old with prodromal symptoms, brief self-resolving psychotic episodes, and a significant decrease in functioning in the past year. A meta-analysis of long-term outcome studies among first-episode psychosis in adults found remission over a mean follow-up of 5.5 years occurred in 58% of participants, and recovery over a mean follow-up of 7.2 years occurred in 38% of participants. This meta-analysis defined remission when symptoms were mild or absent and/or functional improvement occurred for at least greater than 6 months, while its definition of recovery required improvements to be sustained for at least 2 years. Predictors of treatment nonresponse among patients in their early stage of psychosis include lower premorbid functioning, lower level of education, negative symptoms from first psychotic episode, comorbid substance use, younger age at onset, lack of early response, non-adherence to treatment, and longer duration of untreated psychosis. Currently, a clear association between the duration of untreated psychosis and direct neurotoxicity does not exist.

Not all people with schizophrenia require continuous antipsychotic treatment for the remainder of their lives. Their characteristics include better premorbid developmental conditions, less vulnerability to anxiety and psychosis, and better neurocognitive skills. Another predictor of outcome is country of residence. A World Health Organization study found participants living in East Asia experienced the highest clinical remission rate at 84.4%; however, participants living in North America experienced the highest functional remission rate at 35.0%. Environmental factors associated with recovery include repeated opportunities for rehabilitation, positive expectations, and relationships, family psychoeducation, and access to good-quality mental health care (i.e., therapeutic alliance with clinician, satisfaction with antipsychotics, general medical condition care, assistance in managing deficits in resources, and responsiveness to needs and services).

The principle that schizophrenia is not uniformly synonymous with a chronic disabling illness was not a prevalent view during the *Diagnostic and Statistical Manual of Mental Disorders* (DSM)-III era. DSM-III described schizophrenia as “a complete return to premorbid level of functioning in individuals is so rare as to cast doubt upon the accuracy of the diagnosis.” However, DSM-5 reads: “course and outcome may not be reliably predicted. The course appears to be favorable in about 20% of those with schizophrenia, and a small number of individuals are reported to recover completely. However, most individuals with schizophrenia still require formal or informal daily living supports, and many remain chronically ill, with exacerbations and remissions of active symptoms, while others have a course of progressive deterioration.” As an appreciation of recovery being possible grows, a question naturally arises regarding what are the meaningful outcomes that should be evaluated to establish a practice as being a *recovery evidence-based practice*, especially given the majority of existing peer-reviewed literature analyzes symptoms rather than functional, everyday life task improvement.

EVIDENCE-BASED PRACTICES

The presented practices are a sampling of recovery evidence-based practices. More exhaustive lists can be found at such sites as the U.S.

Department of Health and Human Service’s SAMHSA’s Evidence-Based Practices Resource Center.

Pharmacotherapy

Pharmacotherapy is the mainstay of evidence-based treatments for schizophrenia. Yet, findings from a cohort of nearly 30,000 patients living with schizophrenia who were prescribed antipsychotics found that 43.7% were rehospitalized, and 71.7% experienced treatment failure over the observation period of approximately 7 years. These negative outcomes were the least among those prescribed clozapine and long-acting injections of antipsychotics. Neither first- nor second-generation antipsychotics effectively treat primary negative symptoms or cognitive impairments; second-generation antipsychotics are associated with minimal cognitive gains that are not associated with functional improvements. Antipsychotics improve functional impairment as a consequence of reducing positive symptoms; however, the main causes of functional impairment are negative symptoms and cognitive impairment. Clinical Antipsychotic Trials of Intervention Effectiveness found that before 18 months of antipsychotic treatment, 74% of the study’s population discontinued their antipsychotic due to such factors as inefficacy or intolerable side effects. This finding stresses the importance of the inclusion of nonpharmacologic therapies in tandem with pharmacotherapy in the recovery process for people with schizophrenia.

Assertive Community Treatment (ACT)

ACT is a comprehensive, individualized approach to delivering care for people whose functional impairments traditionally from chronic psychosis prevent them from effectively navigating their lives and health care systems in less intensive care models. ACT’s fundamental principles include integration of services, team approach (i.e., shared caseloads and daily team meetings), low client–staff ratios (i.e., 10:1 ratio, which varies depending on the severity of clients’ functional impairment), point-of-contact care delivery in the community, medication management, focus on everyday concerns, rapid access, assertive outreach (i.e., being persistent in offering services to clients who are not engaging in their care), and individualized services. Two original components of ACT have been revised: (1) time-unlimited services have been replaced with policies and practices that support encouraging graduation; (2) the multidisciplinary staff concept has been replaced with the concept that team members should continuously learn new competencies. ACT is one of the most researched treatment models and is associated with numerous beneficial outcomes (Table 12.15–2). However, critics view ACT as not being a recovery-orientated practice due to its “assertive outreach” model of repeatedly contacting clients who have repeatedly declined this service, its medication management component, some of its fidelity standards are clinician rather than client preference based, and utilization of behavioral contracts, outpatient commitments, and representative payeeship.

Table 12.15–2.
Illness-Centered Versus Person-Centered Prescribing

Illness-Centered Prescribing	Person-Centered Prescribing
1. Take a history and collect objective signs and symptoms.	1. Engage the person in the process, building a trusting relationship.
2. Make a diagnosis.	2. Find out their goals to motivate them and build hope for positive change.
3. Obtain informed consent (may include family psychoeducation).	3. Create a shared story of their lives, a “formulation” that stands up to self-reflection and makes them feel understood.
4. Prescribe medications.	4. Work together, using shared experience and decision-making, to learn how they can use medications and other tools to achieve their goals, change, and recover.
5. Assess the impact on target signs and symptoms of diagnosis and creation of undesirable side effects.	5. Support adaptations to their changed perceptions of reality and develop a recovery narrative.
6. Adjust medications until effect maximized.	6. Help them internalize the meaning of their changes, so they can be self-sustaining.
7. Continue effective medications to avoid relapse and deterioration.	
8. If medications have been effective and the patient is motivated, they may add rehabilitation to overcome disabilities and recovery services to enhance the quality of life.	

Supported Employment (SE)

SE's core principles include obtaining competitive employment rather than a sheltered workshop or transitional work participation, being available to all clients living with psychosis who are interested in competitive employment, designing job searches based on client employment preferences, engaging in rapid job search, integrating mental health and employment services, and providing individualized job supports and benefits counseling. Study findings favor SE with a large effect size in job acquisition, total weeks worked, and job tenure as compared to non-SE vocational programs. This employment has been associated with improved quality of life and well-being with mixed effects on clinical functioning.

Cognitive-Behavioral Therapy (CBT)

Findings from randomized controlled trials with CBT developed specifically for people with psychosis include no difference in psychotic symptom reduction or social function improvement as compared to an antipsychotic medication treatment-only control group; however, two people in the CBT group versus none in the antipsychotic medication group were hospitalized during the study period. Core components of CBT used in these trials included normalizing interpretation of events, examining the advantages/disadvantages of interpretations and responses, exploring causes of these events or interpretations, using behavioral experiments to test interpretations, considering alternative explanations, and developing coping strategies.

Housing First (HF)

HF's foundation is that permanent housing is a basic human right to be provided in tandem with a psychiatric rehabilitation approach that values client choice. "Continuum of Care," the traditional model, requires clients to complete a progressive housing program model starting from outreach to transitional housing to end with permanent housing. Given its progressive nature, if clients are unable to meet the "housing readiness" expectations for each step, such as maintaining sobriety or adhering to prescribed medications, then clients do not advance to permanent housing and/or potentially return to homelessness. Studies have shown that clients of HF, as compared to Continuum of Care, are more likely to experience fewer emergency department visits and hospitalizations, spend less time hospitalized, spend more days housed, and obtain housing earlier. No difference in substance use has been demonstrated.

Wellness Recovery Action Planning (WRAP)

WRAP is a peer-led, self-management intervention program for chronic mental health conditions. Unlike traditional medical models, WRAP does not depend on clients obtaining formal mental health care. WRAP also fosters an environment that empowers clients to build a life worth living, rather than simply managing symptoms, through emphasizing holistic health, wellness, strengths, and social support. WRAP intervention has been associated with a reduction in psychiatric symptoms and self-reported need for and use of formal mental health services and improvement in physical health, quality of life measures, and recovery-orientated perspectives such as hopefulness.

Motivational Interviewing (MI)

MI’s aim is to resolve ambivalence as the means to stimulate behavioral change. Its elements include rapport building, reflective listening, assessing the client’s motivation for change through the use of open-ended questions, highlighting the client’s change-related statements and efforts, demonstrating the gap between baseline and desired behaviors, soliciting permission prior to providing assessment or recommendations, utilizing resistance as prompt to alter approach, bolstering client’s self-efficacy for change, and creating an action plan that client verbalizes willingness to follow. Utilizing MI when providing treatment to people living with schizophrenia is associated with reducing symptom severity and rehospitalization and improving medication adherence, functioning, and insight into illness and/or treatment over a 6-month follow-up period.

Electroconvulsive Therapy (ECT)

ECT is indicated for the treatment of schizophrenia when antipsychotic therapy has not been associated with a beneficial response.

Low-Frequency Repetitive Transcranial Magnetic Stimulation (rTMS)

rTMS on the left temporoparietal cortex appears to be a valid treatment for auditory hallucinations in refractory schizophrenia.

EXEMPLARY PRACTICES

Among expert clinicians, the trend is to define practices as being the "gold standards" if they are "evidence-based." This paradigm does not address the challenges of creating evidence-based practices. Specifically, are the populations being enrolled in studies, which are the bases for evidence-based practices, truly representative of the general population living with mental health conditions who do not participate in these studies? Do these studies' questions, hypotheses, study design, and outcome measures take into consideration the perspective of people living with mental health conditions? This lack of representation limits the scope of evidence-based practices. Actions that might bridge this gap include active recruitment of people with mental health conditions into mental health fields given this level of training is required to appreciate the nuances of evidence-based practice creation and evaluation, self-disclosure by current mental health practitioners of their mental health conditions as a means to demonstrate that mental health policy and research decisions are not being created in isolation from clients, and awareness by mental health practitioners of practices that do not have the needed body of literature to determine if they are evidence-based practices but are associated with preliminary data to suggest beneficial outcomes. The presented practices are a sampling of recovery practices that do not have the needed body of literature to determine if they are "evidenced-based," but preliminary data suggest beneficial outcomes.

Crisis Intervention Team (CIT) Model

The number of people with serious mental illness in prisons and jails surpasses the number in state psychiatric hospitals by 10-fold. CIT's goal is to divert people with serious mental illness from the criminal justice system to services, if appropriate, through a collaborative approach between law enforcement and mental health systems. The main components of CIT include (1) 40 hours of mental health-related training for CIT officers; (2) training for call dispatchers to recognize calls potentially involving persons with serious mental illness and assigning them to CIT officers; (3) central designated psychiatric emergency drop-off site that accepts all people CIT officers bring for assessment. Initial findings suggest that CIT programs are associated with improved officer-level outcomes (officer satisfaction and self-perception of reduced use of force) and increased prebooking diversion from jails to psychiatric facilities. Peer-reviewed literature contains little evidence of CIT's impact on objective measures of arrests, officer injury, citizen injury, or use of force. Rogers Mental health advocates also question the effectiveness of CIT approaches, especially for BIPOC communities. CIT training does not include any content on racism or implicit bias for police to address.

Mental Health Courts (MHC)

MHC was created to address the over-representation and disproportionate jail stays of inmates with mental health illnesses; the U.S. jail system incarcerates approximately 2 million adults with serious mental illness annually and retains them 2.5 to 8 times longer than adults without mental health illnesses. MHC's goal is to divert people with mental illness from the criminal justice system to community treatment while maintaining public safety. Participation in MHCs is voluntary, given this decision is usually accompanied by a mandatory guilty plea; however, some courts expunge guilty pleas upon successful completion. Generally, clients are required to sign a written contract pledging their commitment to MHC's mandates, such as clients taking their prescribed medications. Penalties for nonadherence to this contract include increased supervision and reincarceration. Evidence suggests that important components of effective mental health courts include care managers or access to vocational and housing services. Long participation in MHCs has been associated with reduced recidivism rates without consistent findings that access to behavioral health services or psychiatric functions improves. This discrepancy then questions the criminalization hypothesis that mental illness causes criminal behavior. There are concerns about the impact of such long periods of involvement with MHCs on participants, as well as the disparities between the extensive length of involvement in MHC and traditional criminal court, leading to concerns that MHCs actually extend state control over individuals with serious mental illness rather than divert them.

Hearing Voices Network (HVN)

Collaborating with people who hear voices, social psychiatrist Marius Romme and researcher Sandra Escher promoted an approach that developed into HVN. The pillars of this approach emphasize accepting and making sense of these voices, being recovery orientated, and understanding the role of psychosocial adversity in voice onset and maintenance. A key part of this international network is supporting self-help groups called hearing voices groups (HVG). The aim of HVG includes providing psychologically safe environments to share the experience of hearing voices, promoting coping strategies, and cultivating a positive identity as people who hear voices. HVG consider the persons who hear voices as the experts in their experience, and as such, all perspectives about the existence of these voices are considered valid. Given that HVN does not consider hearing voices as pathologic, success is measured not by voice cessation but by understanding and accepting the experience of hearing voices and forming productive relationships with them. HVG members report an increased sense of social support, acceptance, and empowerment, along with improved coping strategies and quality of life. Critics of HVN also inaccurately perceive HVN as devaluing the benefits associated with antipsychotic treatment. HVN is not antipsychiatry, and members of HVN welcome and support each other to consider all pathways for addressing and making sense of their experiences, including the use of medications and psychiatry.

Open Dialogue

In 1984, the inpatient admission practices performed in a Finnish psychiatric hospital were changed to include patients and their family members in treatment-planning conversations. Open Dialogue has its origins in this transparency and includes seven principles: immediate help; social network perspective; flexibility and mobility; responsibility; psychological continuity; tolerance of uncertainty; and dialogue. Using this framework, Dialogic Practice promotes creating equal space for different points of view and thus reduces the gap between the identified “sick” and “well.” This equal footing dynamic increases the potential for people to be heard. In addition to relying on the personal ways therapists respond to people, Dialogic Practice has twelve key fidelity elements: having two (or more) therapists in the team meeting; having family and network participation; using open-ended questions; responding to clients’ utterances; emphasizing the present moment; eliciting multiple viewpoints; using a relational focus in the dialogue; responding to problem discourse or behavior in a matter-of-fact style and being attentive to meanings; emphasizing the clients’ own words and stories, not symptoms; having a conversation amongst professionals (reflections) in the treatment meetings; being transparent, and tolerating uncertainty. Olson Studies suggest that Open Dialogue decreases chronicity and disability associated with schizophrenia by half and is highly cost-effective as compared with geographic and historic controls. Critics of the available research on Open Dialogue grade it as very low quality.

Soteria

From 1969 to 1971, the original Soteria-California research project was conducted. Meaning “deliverance” in Greek, Soteria was the name of a 1915-vintage, 12-room house located in the Bay Area used as the location for the research’s experimental arm. The core principles of the Soteria paradigm include: creating a small, community-based therapeutic milieu; having a significant proportion of staff be lay staff; preserving personal power, social networks, and communal responsibilities; understanding and hence giving meaning to a person’s subjective psychotic experience by “being with” and “doing with,” having empathetic relationships and engaging in meaningful shared activities with clients, respectively; utilizing no or low-dose antipsychotic medications where clients take medications by choice and without coercion. This antipsychotic medication viewpoint stems from antipsychotic medications not being a cure-all for psychosis and from the desire to avoid unnecessary antipsychotic medication-associated side effects. The limited existing data suggests Soteria treatment produces comparable psychotic symptoms reduction with significantly lower doses of medications as compared to conventional hospital treatment, and clients strongly prefer it over conventional hospital treatment. The most controversial aspect of Soteria treatment is the concept that the therapeutic milieu in itself could be enough to treat acute psychosis. This hypothesis is based on the “affect-logic” theory that contends marked increase in affective tension in vulnerable people triggers psychosis. Thus, the sustained emotional relaxation cultivated in the Soteria therapeutic milieu causes symptom remission.

Trauma-Informed Approach

Causes of trauma include violence, abuse, neglect, loss, disaster, war, and other emotionally harmful events. Exposure to traumatic events is high. For example, among an online survey of U.S. adults, 89.7% reported exposure to a traumatic event. Addressing trauma is increasingly considered a vital health care component because traumatic exposure is almost ubiquitous and can have lasting adverse effects on people's functioning and mental, physical, social, emotional, and spiritual well-being. A trauma-informed approach includes all people at all levels of the organization or system realizing how trauma impacts groups and individuals, recognizing the signs of trauma, responding with a trauma-informed approach in all their interactions, and resisting retraumatization of both staff and clients. The six key principles of a trauma-informed approach include safety (staff and clients feel physically and psychologically safe throughout the organization); trustworthiness and transparency (the organization is transparent in its decisions with clients, family, and staff); peer support (people with lived experiences of trauma promote recovery through sharing their life experiences); collaboration and mutuality (power differences between staff and clients and among staff are leveled as the organization values everyone's contributions in providing a trauma-informed approach); empowerment, voice, and choice (the organization supports clients' shared decision-making, choice, and goal setting and provides needed supports for staff to do their work); cultural, historical, and gender issues (the organization addresses biases and offers responsive services). Inconsistent definitions of trauma-informed approaches and limited available information about measurement development processes limit the rigorous evaluation of determining the effectiveness of the trauma-informed approach.

ROLE OF PSYCHIATRIST

Psychiatrists are important staff members in a person's recovery despite the profound shortages of psychiatrists and limited time available because, all too often, they are the only staff. This situation is due to the crucial importance of medications and hospitalizations within the current system, both usually under direct psychiatrist control. There are important tools psychiatrists can incorporate into their practice.

The recovery model has frequently been presented to psychiatrists as either an antipsychiatry, antimedications movement that interferes with treatment and encourages people to get off their medications or as a useful supplement to psychiatric treatment, including a range of rehabilitative, supportive, self-help, and clubhouse services, especially useful once illnesses are successfully controlled. Instead, it can be a revolutionary paradigm, emphasizing person-centered, client-driven, strengths-based perspectives that transform our entire mental health system, including our daily prescribing practices, as was recommended by the President's New Freedom Commission on Mental Health back in 2003. Consider Table 12.15–2, which contrasts our usual prescribing practices with a transformed practice:

Once we free ourselves from the constraints of the medical model, we can find lots of collaborative uses for medications. Here are some examples of medication usage within various formulations from our journeys:

- The pills strengthen you mentally, so you can handle the stress you're under without fragmenting as much. They help "glue" you together. They also keep your sleep/wake cycle and basic functions under better control.
- The pills help your mind filter the amazing range of stimuli, including physical stimuli from your body, and internal thoughts that bombard it, so you can think more clearly and more linearly and have an easier time maintaining your personal balance and homeostasis, but this might make you feel less creative.
- The pills slow down your ability to jump rapidly between various thoughts and mental associations, so your thinking will be more grounded and normal and less ethereal and idiosyncratic. This may help you feel less lost, confused, and alone and help you be more productive.
- The pills help strengthen your rational, conscious mind so that your unconscious mind is more under control in both your sleep and waking hours.
- The pills help you stay in reality, so you're better able to deal with it instead of escaping into your imagination.
- The pills help strengthen your personal boundaries, so you won't pick up on other people's feelings as much as you walk by, and they won't impact you as much. This may make you less intuitive and lonelier but will make you more emotionally grounded and independent. This also makes it harder for other people to influence your mind outside of your personal control, regardless of the way they're trying to do it.
- The pills help make the "spiritual waters" you're swimming in shallower, so you won't drown as easily, but you also may not have as much contact with the immensity of the spiritual world, at least until you learn more about how to "swim" in these "deep waters."

- The pills help strengthen you, so the devil can't fool you or bully you as easily.
- The pills decrease the ability of voices to impact your self-consciousness and take over more of your mind.
- The pills partially counteract the disorganizing influence of drugs, alcohol, seizures, and other medical and neurologic conditions.

These aren't clever ways of manipulating people into taking medications who don't have insight into their psychiatric illness. They are authentic hypotheses for mechanisms of action for how these pills actually work. Remember that the classes of medications we now use have been named to correspond to the illness-centered model that corresponds to their FDA approval and their marketing campaign. For example, "antipsychotics" were originally "major tranquilizers" when the goal was to sedate institutionalized patients (or sometimes housewives), and now they've been rebranded as "mood stabilizers" so they can be marketed to people with major depression and bipolar disorder without psychosis. These class names aren't really any more legitimate than these psychological descriptions. These pills are all so complex neurochemically that there isn't a clear correlation between their biochemical actions and their psychological impacts. These descriptions haven't been studied, not because they're outlandish but because the medical model is narrow-minded and controls all the pharmaceutical research funding.

Keep in mind that medications aren't the only tool available. For example, some people might improve with better sleep, coping skills, cognitive reframing, using fewer drugs, talking with friends, relaxation, less isolation, etc. A package can be put together through shared trial-and-error for each person.

Taking Strengths Seriously and Building Resilience

- Building Social Determinants of Health: Often, the outcome of treatment depends more upon the social characteristics of the person than upon their illness itself. Integrating poverty services with mental health services without saying "that's not my job" will likely improve outcomes. Political and community activism supporting social policies that promote improved social determinants of health will also likely improve outcomes.

- Building protective factors: Protective factors have been shown to prevent risk factors from becoming predictive. For people experiencing repeated mental health crises, it may be more effective to build protective factors to prevent further mental health crises than to treat each crisis after it occurs.
- Building self-efficacy: The more someone believes and has experienced that what they do makes a difference in their life, the more likely they are to take self-responsible action to recover. Examples of self-efficacy include: Increasing the ability to affect inner states directly, increasing one's role in their own life instead of relying on others, developing specific goal-driven skills to achieve personal growth, making developmental progress through life's stages, building interpersonal and community-based efficacy, and increasing spirituality.
- Building resilience by finding strengths in struggles: Sometimes, the process of struggling helps people discover and develop strengths inside of themselves they didn't know they had, receive gifts from their wounds, and figure out what's really important in life. Examples of these strengths include internal qualities, external resources, meaning-making, and enhanced connectedness and belonging.

This example illustrates how the psychiatrist–patient relationship keeps evolving to meet the needs of each stage of the recovery journey. Notice the variety of interventions used by the team and the psychiatrist. Notice also how the three dimensions of the Psychosis Triangle (Experiencing Reality, Self-Identity, and Relationships) evolve in intertwined ways along the way, impacting each other positively and negatively.

Martha had suffered a great deal by the time I met her. She grew up in a small town in Missouri, the fourth of eight children. She had only a ninth-grade education and could barely read and write. When she was 15 years old, Martha's father left for St Louis to find work and her mother struggled to take care of the family even with his financial help. Martha started spending more time reading her Bible and began believing that God was telling her what to do. Unfortunately, the instructions included spending her day trying to hail down cars on the highway to teach people about the Bible and taking off all of her clothes in church to prove she was pure. After trying to deal with her for several years, Martha's family took her to a psychiatric hospital when she was standing over her mother with a knife one night. She was repeatedly put into seclusion and sedated for disrobing in the hospital. One day, she reported being raped by another patient, but the staff didn't believe Martha. After 2 weeks, she was discharged back home with a prescription she threw away because it drugged her. After 5 more years of psychosis, fighting with her family, never going to the clinic in the next town, police being called repeatedly on her, and several hospitalizations and jailings, she ran away to Los Angeles, where she arrived without any funds or friends. She was huddled in the corner, mumbling to herself, disheveled, holding tight to her Bible.

Help Them Trust Us

In a recovery-based treatment, clients do not trust their psychiatrist just because of their ability to treat them. They also trust their psychiatrist to personally care enough to create a relationship where they can find strengths in themselves to use their treatments and rebuild their lives.

Martha was wary of psychiatrists because she was afraid of being locked up, so staff arranged for me to meet her a few times informally, eating lunch at our client-run Deli. We talked about my memories of St Louis from medical school and about how she had helped raise her younger siblings. We shared family pictures from our wallets. I told Martha about a couple of other young women we'd met who came to Los Angeles all alone and how, even though it is hard, I thought she had the strength to make it. The third time, I helped Martha with forms to get welfare payments and a discount bus pass and offered to help her get social security income (SSI) if she came in for a full assessment. We let Martha use our showers and got her some clean clothes from our donations.

Help Them Regain Control of Their Lives

In a recovery-based treatment, the focus isn't on the psychiatrist taking care of them and helping them feel better. It is on them to take control of their own life so they can take care of themselves and function even if they're not always feeling well.

Martha enrolled in our program. We asked her what her goals were, and she said she wanted a place to stay of her own, someone to talk to and feel normal with, and to go to church. We introduced Martha to another member who was a little older and further on in her recovery, and they became friends even though Martha is hard to understand sometimes and still talked to voices. We loaned Martha money against her pending back check from SSI to get a cheap weekly hotel room. In return, she had to come to the Village and do some light office work for us for 1 hour a day. Martha's favorite job was helping to decorate the office for Christmas, but it made her miss her family. We helped Martha call them and tell them she was safe, and she sent off a box of presents for them. I showed her the award I'd gotten from the local Baptist church for community service and told her that although it isn't my job to help find business for churches, I could give her a list of local churches to try. She got dressed up to try the church her friend went to that caters to homeless people. As she was feeling stronger, she asked to meet our teacher to learn to read better. She became attached to him as a surrogate father figure, but mostly she was shocked and pleased with herself that she could learn to read when no one was yelling at her and hitting her when she made a mistake.

Help Them Rebuild Their Life

In a recovery-based treatment, the goal is not just to treat a mental illness. The goal is to help someone with a mental illness have a better life.

My report helped Martha get SSI, and we gave her the money for a deposit for her own apartment. Several staff and clients went to Martha’s housewarming party. I brought diet soda and chocolate. She said she was having trouble sleeping and wondered if I had any medications that might help. She noticed that other clients weren’t being drugged up here at the Village and had grown to trust me. I asked Martha about what was keeping her awake, and we ended up talking about how she was tormented by demons which she hated, as well as hearing God, which was a blessing. We talked about how the demons took advantage of her when she felt weak and vulnerable, and maybe medications could strengthen her to keep them away. We decided together that she would try taking low-dosage antipsychotic medications at bedtime, and I encouraged more nightly prayers to prepare her to sleep safely. She started coming to regular appointments with me in my office. As she felt stronger, she wanted to earn some money to donate to the church, and we helped Martha get a part-time training job at our Deli. She enjoyed preparing my daily salad the way I like it, without any carrots.

Help Them Heal

In a recovery-based treatment, healing isn’t just from the illness. Healing is also from the destruction and the rejection.

She was proud to share her success with her mother and younger brother, but she also began to realize how much she’d hurt them over the last few years. We began talking about her regrets more in our sessions than her medications, and she began to accept what she’d done enough to write a letter asking her mother for forgiveness. She never sent it, but she felt a weight lifted nonetheless. When our teacher left to take another job, she was devastated. She started isolating herself and stopped taking her medications, and the demons returned with a vengeance.

Take the Long View

In a recovery-based treatment, psychiatrists don't just assess and treat illnesses as detached professionals. Psychiatrist and patient walk together, sharing the joys and sufferings of recovery.

We resisted the impulse to hospitalize Martha and get her back on her medications, even though it was heartbreaking to see Martha deteriorate again because we didn't want to push her away. She lost her job at the Deli. She wouldn't respond to repeated phone calls, so I went with Martha's caseworker to Martha's apartment to reach out to her. She opened the door and broke into tears. As we talked, we realized together that the loss of our school teacher had reactivated the feelings of abandonment and betrayal she'd felt when Martha's father had left them. We talked about how she was older and stronger now and could handle this departure. She agreed to take the bottle of pills I'd brought in my pocket.

Help Them Move On

Even though many serious mental disorders persist indefinitely, a recovery-based treatment can prepare them to deal with whatever happens and keep going with their lives without needing to be in treatment forever.

After Martha rebuilt her life again, she enrolled in a WRAP group to develop better tools besides giving up and isolation to deal with life's inevitable losses in the future. And she'd have to do it without me coming to her home to help rescue her. I started spending less and less time helping her, and she spent more time mentoring new clients. The next Christmas, she saved money to take a bus back to Missouri, and this time did ask her mother for forgiveness which her mother gave her. After she returned to Los Angeles, she realized that Missouri really was her home, and we made arrangements for her to return home, including a phone call and letter from me so she could get medication from the clinic there. At our farewell party, she gave me a poster she'd made with her friend, as she was leaving, for my office wall that said, "The Village opens a door for us to have a better life. It's up to us to walk through it."

PEER SUPPORT SPECIALIST

Peer support is emotional support offered by a person with a mental health condition(s) to another who has a similar mental health condition(s) and works toward minimizing the inherent power imbalances between clients and providers. Peer Support Specialists (PSS) are people who are active in their recovery and potentially have obtained a PSS certificate. Over 30,000 PSS are working in the United States alone. Although fluctuating in its prominence in the history of psychiatry, the role of PSS can trace its roots to at least the “moral treatment era,” an 18th-century concept whose aim was to treat people with mental health disorders humanely. In 1935, Alcoholics Anonymous formed and utilized the peer support concept of a “sponsor.” During the 1970s, Mental Health Recovery Movement advocated for more inclusion of people with mental health disorders in their mental health care, and as such, the concept of incorporating people with mental health disorders into mental health treatment teams further emerged. In the early 1990s, the current version of PSS developed; however, their roles and responsibilities were limited as compared to today’s PSS.

Early roles for PSS generally included fewer clinical activities or activities that non-PSS clinicians were too busy to perform, such as driving clients to their appointments, filing, and answering phone calls. PSS initially did not participate in treatment planning given their marginalized, unclear clinical role. With time, PSS’ roles have become more clinical in nature to include assisting in identifying goals, teaching new skills, advocating, job coaching, providing counseling, and leading groups along with documenting these clinical interactions in the client’s medical records. One of the main roles of PSS is to disclose their mental health conditions to clients as a means to serve as a role model. This disclosure also potentially makes them more approachable to clients due to this acknowledged shared similar mental health-related experiences. These shared experiences can lay the foundation for engagement with clients whom non-PSS clinicians have been unable to reach. Another important role of PSS is to educate non-PSS clinicians about what methods might assist clients based on the PSS’ past experience with similar experiences and model engagement based on what matters most to clients.

SAMHSA considers the role of PSS as evidence-based practice in the context that “there is a need to identify specific attitudes, knowledge, and skills that are essential for peer support and effective operation of these services and programs.” Studies have shown that PSS is effective in engaging people in care, reducing emergency visits and hospital admissions, and improving social functioning. If PSS provides positive self-disclosure and role modeling to clients, these actions increase the clients’ self-care ability, decrease clients’ depression and psychosis, and increase clients’ sense of hope, control, ability to effect changes, community belonging, and life satisfaction. However, conclusions from limited meta-analytic reviews are PSS has a limited impact on measured outcomes.

Challenges faced by PSS include hiring officials at times being hesitant about hiring PSS for multiple reasons. One reason is concern that job-related stressors will disproportionately negatively impact PSS compared to non-PSS clinicians. As a result, “stability” eligibility criteria are sometimes utilized in hiring practices, such as a time requirement since last mental health hospitalization or length of sobriety. These eligibility restrictions are not based on any evidence-based predictive findings and potentially constitute discrimination based on disability rights legislation. Another concern is PSS might break clients’ confidentiality; however, like other clinicians, PSS has national practice guidelines that emphasize confidentiality and privacy and should be provided with on-the-job training and supervision, which also emphasize these topics. Yet another concern is that PSS will make the supervisor’s work harder. Initially, PSS might benefit from more supervision time to acclimate to working in their new environment than non-PSS clinicians; however, long-term benefits of improved client engagement and associated outcomes generally outweigh this time investment.

PSS do encounter unique on-the-job challenges that non-PSS clinicians do not encounter. The identification of therapeutic boundaries might be more challenging, given their past experiences and their role of being the more relatable people in the treatment team. For example, in the past, PSS could have been hospitalized with a client and, as such, has a preexisting relationship with this client as a client. Another unique challenge is clients often view PSS as reaching the pinnacle of success and, as such, are viewed as role models whom they overtly idolize. Simultaneously, some clients may be concerned about a possible relapse of the PSS, impugning the reputation of recovering persons as being able to perform in a competent manner. These challenges can usually be effectively dealt with in supervision with PSS' clinical supervisor either in individual or group meetings with other PSS and through structured training. In addition to challenges, the role of PSS often brings personal growth to both PSS and the other members of the treatment team. The PSS benefits from the experience of having fulfilling employment. The non-PSS clinicians are able to interact with people with openly disclosed mental health conditions who are contributing to their team's mission, challenging misconceptions about the abilities of people living with mental health conditions.

FURTHER READINGS

Ali SA, Mathur N, Malhotra AK, Braga RJ. Electroconvulsive therapy and schizophrenia: a systematic review. *Mol Neuropsychiatry*. 2019;5:75–83.

Barry CL, McGinty EE, Pescosolido BA, Goldman HH. Stigma, discrimination, treatment effectiveness, and policy: public views about drug addiction and mental illness. *Psychiatr Serv*. 2014;65(10):1269–1272.

Baxter AJ, Tweed EJ, Katikireddi SV, Thomson H. Effects of housing first approaches on health and well-being of adults who are homeless or at risk of homelessness: systematic review and meta-analysis of randomized controlled trials. *J Epidemiol Community Health*. 2019;73:379–387.

Blow FC, Zeber JE, McCarthy JF, Valenstein M, Gillon L, Bhingam CR. Ethnicity and diagnostic patterns in veterans with psychoses. *Soc Psychiatry Psychiatr Epidemiol*. 2004;39(10):841–851.

Bond GR, Drake RE. The critical ingredients of assertive community treatment. *World Psychiatry*. 2015;12(2):240–242.

Bozzatello P, Bellino S, Rocca P. Predictive factors of treatment resistance in first episode of psychosis: a systematic review. *Frontiers in Psychiatry*. 2019;10:67.

Calton T, Ferriter M, Huband N, et al. A systematic review of the Soteria paradigm for the treatment of people diagnosed with schizophrenia. *Schizophrenia Bulletin*. 2008;34(1):181–192.

Chamberlin J. *On Our Own: Patient-Controlled Alternatives to the Mental Health System*. McGraw-Hill; 1978.

Champine RB, Lang JM, Nelson AM, Hanson RF, Tebes JK. Systems measures of a trauma-informed approach: a systematic review. *Am J Community Psychol*. 2019;64(3–4):418–437.

Chien WT, Mui JHC, Cheung EFC, Gray R. Effects of motivational interviewing-based adherence therapy for schizophrenia spectrum disorders: a randomized controlled trial. *Trials*. 2015;16:270–284.

Cook JA, Goldrick V, Steigman PJ, et al. Impact of wellness recovery action planning on service utilization and need in a randomized controlled trial. *Psychiatric Rehabilitation*

Journal. 2013;36(4):250–257.

Cooper RE, Laxhman N, Crellin N, Moncrieff J, Priebe S. Psychosocial interventions for people with schizophrenia or psychosis on minimal or no antipsychotic medication: a systematic review. *Schizophrenia Research*. 2020;225:15–30.

Dixon LB, Goldman HH, Srihari VH, Kane JM. Transforming the treatment of schizophrenia in the United States: the RAISE Initiative. *Annual Review of Clinical Psychology*. 2018;14:237–258.

Drake RJ, Husain N, Marshall M, et al. Effect of delaying treatment of first-episode psychosis on symptoms and social outcomes: a longitudinal analysis and modelling study. *Lancet Psychiatry*. 2020;7:602–610.

*Frese FJ, Stanley J, Kress K, Vogel-Scibilia S. Integrating evidence-based practices and the recovery model. *Psychiatr Serv*. 2001;52:1462–1468.

Harding CM, Brooks GW, Ashikage T, Strauss JS, Breier A. The Vermont longitudinal study of persons with severe mental illness, I: methodology, study sample, and overall status 32 years later. *Am J Psychiat*. 1987;144:718–726.

Harrow M, Jobe TH, Faull RN. Does treatment of schizophrenia with antipsychotic medications eliminate or reduce psychosis? A 20-year multi-follow-up study. *Psychological Medicine*. 2014;44:3007–3016.

Hatton R. Open dialogue: the evidence and further research. *Psychiatric Services*. 2019;70(6):529–530.

Jackson V. In our own voice: African-American stories of oppression, survival and recovery in mental health systems. 2002;2:11–31.

Klee A, Chinman M, Kearney L. Peer specialist services: new frontiers and new roles. *Psychological Services*. 2019;16(3):353–359.

Lally J, Ajnakina O, Stubbs B, et al. Remission and recovery from first-episode psychosis in adults: systematic review and meta-analysis of long-term outcome studies. *Br. J. Psychiatry.* 2017;211:350–358.

Loong D, Bonato S, Barnsley J, Dewa CS. The effectiveness of mental health courts in reducing recidivism and police contact: a systematic review. *Community Mental Health Journal*. 2019;55:1073–1098.

McKay C, Nugent KL, Johnsen M, Eaton WW, Lidz CW. A systematic review of evidence for the clubhouse model of psychosocial rehabilitation. *Adm Policy Ment Health.* 2018;45(1):28–47.

Metzl JM. *The Protest Psychosis: How Schizophrenia Became a Black Disease*. Beacon Press; 2009 https://www.amazon.com/Protest-Psychosis-Schizophrenia-Became-Disease/dp/0807001279?asin=0807085928&revisionId=&format=4&depth=1

Mueser KT, Drake RE, Bond GR. Recent advances in supported employment for people with serious mental illness. *Curr Opin Psychiatry*. 2016;29(3):196–201.

Nischk D, Rusch J. What makes soteria work? On the effect of a therapeutic milieu on self-disturbances in the schizophrenia syndrome. *Psychopathology*. 2019;52:213–220.

Otani VHO, Shiozawa P, Cordeiro Q, Uchida RR. A systematic review and meta-analysis of the use of repetitive transcranial magnetic stimulation for auditory hallucinations treatment in refractory schizophrenic patients. *Int J Psychiatry Clin Pract*. 2015;19:228–233.

Pagdon S, Jones N, eds. 2020. Psychosis Outside the Box. V. 1. www.rethinkpsychosis.weebly.com

Rogers MS, McNiel DE, Binder RL. Effectiveness of police crisis intervention training programs. *J Am Acad Psychiatry Law*. 2019;47:414–421.

Styron T, Utter L, Davidson L. The hearing voices network: initial lessons and future directions for mental health professionals and systems of care. *Psychiatr Q*. 2017;88:769–785.

Tiihonen J, Mittendorfer-Rutz E, Majak M. et al. Real-world effectiveness of antipsychotic treatments in a nationwide cohort of 29 823 patients with schizophrenia. *JAMA Psychiatry*. 2017;4(7):686–693.

Yung A, Nelson B. The ultra-high risk concept—a review. *Canadian J Psych*. 2013;58(1):5–12.

12.16 Psychosis as a Defining Dimension in Schizophrenia

ELENA I. IVLEVA, M.D., PH.D., AND CAROL A. TAMMINGA, M.D.

DSM-5: DEFINITION OF PSYCHOSIS

In the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) and its text revision (DSM-5-TR), the psychosis syndrome is incorporated into a broad range of diagnoses, including schizophrenia, schizoaffective disorder, other psychotic disorders (e.g., delusional disorder, psychosis due to substance use or general medical condition), as well as attenuated psychosis syndrome and schizotypal personality disorder. The key symptom domains comprising “psychosis” include delusions, hallucinations, disorganized thought process (formal thought disorder), and disorganized or abnormal motor behavior (including catatonia). Despite the absence of unique pathognomonic psychosis features for each diagnosis, that is, any type of psychotic symptoms can occur in any diagnosis, DSM-5-TR emphasizes that some aspects of psychosis manifestations are characteristic for various psychosis diagnoses. In schizophrenia spectrum disorders (e.g., schizophrenia, schizoaffective disorder, schizophreniform disorder), psychotic symptoms are typically comprised of prominent hallucinations and/or delusions as well as formal thought disorder leading to rather severe behavioral manifestations (e.g., disorganized, bizarre, or catatonic behavior). In psychotic disorders due to general medical conditions, as well as in substance-induced psychotic disorders, psychosis refers to delusions and/or hallucinations temporally overlapping with the presumably causative event (medical illness, substance use), which do not occur exclusively during the course of a delirium. In such rare psychotic diagnoses as delusional disorder and shared psychotic disorder, psychosis is essentially equivalent to isolated persistent delusions. Importantly, only disorders grouped under the “Schizophrenia spectrum and other psychotic disorders” section of DSM-5 are conventionally characterized by psychotic symptoms as the key defining feature comprising the clinical “core” of the diagnosis. Nevertheless, multiple other psychiatric disorders, such as affective disorders, neurodevelopmental disorders, dementias, and some personality disorders (besides schizotypal type), may also manifest with psychotic symptoms as associated (“secondary”) diagnostic features.

In schizophrenia, psychosis presents in its most severe form, frequently with bizarre delusions and hallucinations leading to prominent behavioral and functional changes. Although any of five senses may be affected by hallucinatory experiences, the most common hallucinations in schizophrenia are auditory, with multiple voices conversing among themselves, commenting on the patient’s life or behavior, or talking directly to the patient. Visual hallucinations are rather common, whereas tactile, olfactory, and gustatory hallucinations are less typical and should alert the clinician to look for underlying neurologic or medical conditions that mimic psychotic disorder. Delusions in schizophrenia vary broadly in content with most classic examples including paranoid/persecutory delusions, ideas of reference, or spiritual/religious delusions. Although not entirely specific for schizophrenia, such delusions as being influenced or controlled by an outside entity or force (delusions of passivity or control), phenomena of thought insertion, withdrawal or broadcasting, and delusional preoccupation with esoteric, symbolic, abstract, or philosophical ideas are more typical for patients with schizophrenia than for other psychotic disorders. Importantly, Schneiderian psychotic features (i.e., presence of bizarre delusions or hallucinations consisting of a voice that keeps a running commentary on the person’s behavior or thoughts or multiple voices conversing with one another) have been de-emphasized in DSM-5 leaving criterion A to a broader range of symptoms including (any) delusions, (any) hallucinations, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms.

Although delusions and hallucinations are commonly perceived as core symptoms of schizophrenic psychosis, they are not an obligate component of the formal diagnosis of schizophrenia. According to DSM-5, since at least two Criterion A symptoms are required to fulfill the diagnosis, patients without either hallucinations or delusions could receive a diagnosis of schizophrenia; however, in these cases, other psychotic symptoms are required such as thought disorder (e.g., looseness of associations, thought blocking, derailment, tangentiality, circumstantiality, neologisms, verbigeration, echolalia, global incoherence), grossly disorganized behavior or catatonic symptoms. Although the phenomena of a gross impairment in reality testing and a loss of ego boundaries are not included in formal psychosis definitions, they are commonly seen in patients with schizophrenia, especially during acute psychotic states. These clinical entities describe the lack of a clear sense of where the patient’s own mind or body ends and where those of other people or inanimate objects begin. In the most severe manifestations, the patients may have experiences that they mentally or physically disintegrated or fused with an outside object or with the entire universe. These components of psychosis may be poorly verbalized by schizophrenia patients due to the complexity of the experiences; however, they can manifest in peculiar, impulsive, or dangerous behavior, and therefore, these symptoms should be carefully evaluated.

Schizophrenic psychosis is usually characterized by a chronic course with either continuous psychotic symptoms or intermittent periods of disease exacerbations and remissions. Course types with a single (or few) episodes followed by a long-term stable remission or full functional recovery are uncommon. However, in other psychotic disorders, the lifetime course of psychosis can be intermittent or fleeting, with relatively short episodes of hallucinations and delusions manifesting at the peak of severe depression or mania, decompensated medical conditions, or induced by substance exposure or medication. In the course of schizophrenia, psychotic symptoms tend to become less severe with time, whereas a decline in cognitive functioning and negative symptoms become more apparent. A further deterioration in the patient’s baseline functioning can follow each relapse of schizophrenic psychosis. This failure to return to baseline functioning after each relapse is commonly taken as the major distinction between schizophrenia and psychotic mood disorders, although recent studies report that about 20% to 25% of patients with affective disorders show considerable long-term impairment in psychosocial and basic daily functioning. Without treatment, psychosis in schizophrenia and related disorders is frequently incapacitating and contributes, along with negative symptoms and cognitive decline, to an overall poor prognosis of the illness.

CATEGORICAL AND DIMENSIONAL DIAGNOSTIC APPROACHES IN SCHIZOPHRENIA

The symptoms comprising DSM-5 criteria for schizophrenia do not define a diagnosis-specific syndrome and reflect substantial clinical heterogeneity of the illness. No single clinical sign is pathognomonic for schizophrenia; every symptom observed in schizophrenia can occur in different psychiatric disorders. The current categorical concept of schizophrenia (and other psychiatric diagnoses) is based on having a number of different combinations of symptoms, such that many variations of the defining and associated symptoms and symptom domains are possible. The phenomenology-driven diagnostic systems, while serving the purpose of clinical utility, have critical limitations resulting in classifications that are heterogeneous, lack clear boundaries, and, most importantly, neither map on emerging experimental biomarker constructs nor reflect causal biologic disease mechanisms. Despite considerable progress in understanding neurobiologic bases of psychiatric disorders, at this point, it is not possible to propose a biologically based diagnostic system due to a lack of defining knowledge about the neurophysiologic and molecular mechanisms of psychiatric disorders. One of the alternative diagnostic approaches espoused to reduce the clinical heterogeneity of psychiatric diagnoses is the identification of psychopathologic illness dimensions (such as psychosis) with a research and clinical focus on underlying disease neurobiology. Such symptom dimensions are comprised of groups of symptoms that resemble one another, occur together, and are, hypothetically, based on interrelated pathophysiologic and molecular mechanisms.

Three major symptom dimensions have been identified in schizophrenia: Psychosis, disorganization, and negative symptoms. The psychosis dimension includes distortions or exaggerations of inferential thinking (delusions) and perceptual disturbances (hallucinations). The disorganization domain manifests as illogical disorganized thought process and/or grossly disorganized behavior. Although schizophrenic disorganization is commonly perceived as one of the accompanying features of acute psychotic state which tend to improve as severity of psychosis diminishes, disorganized features are phenomenologically unique, and are specifically associated with poor prognosis, especially in cases of chronic presentation. Therefore, disorganization has been recognized as a distinct clinical domain of schizophrenia, not just a subtype of psychotic symptoms. Negative symptoms dimension includes restrictions in the range and intensity of emotional expression (affective flattening), in the fluency and productivity of thoughts and speech (alogia), and in the initiation of goal-directed behavior (avolition). Recently, a broader dimensional model of schizophrenia has been proposed, including cognitive dysfunction (i.e., deficits in general intelligence level, various aspects of memory, executive function, global processing speed, and attention) as well as affect dysregulation and mood symptoms (Fig. 12.16–1). Although neither is yet a part of the current DSM-5 diagnostic definition, cognitive dysfunction is known to be one of the defining features of the illness which frequently manifests long before first-break psychosis and is a strong predictor of functional outcome and overall prognosis.

Comparative studies contrasting dimensional composites with traditional diagnostic categories have shown that the dimensional approach may be superior in predicting course, demand for psychiatric care, treatment outcome, social adaptation, and global prognosis of illness. Originally, the dimensional models were developed to aid diagnostic classification. However, more recently these dimensional constructs have been targeted in various research paradigms, including genetic and disease biomarker studies, as well as in identifying clinical targets for novel drug development. Dimensional components within the schizophrenia diagnosis may be characterized by unique underlying neurobiologic mechanisms and appear to be, at least partially, pharmacologically independent. To date, effective treatments are available for only a few of these symptom dimensions. Psychotic symptoms may be satisfactorily controlled with the arsenal of first- and second-generation antipsychotic medications, although treatment-refractory cases are not uncommon. Affective symptoms are typically alleviated by antidepressants, mood stabilizers, and some antipsychotic agents. However, no effective treatments exist either for cognitive deficits or negative symptoms, although treatment development is evolving in these areas. Clinical experience, as well as well-controlled treatment trials, suggests that psychosis across many diagnostic categories, including such diagnoses as schizophrenia, bipolar disorder, psychotic depression, substance-induced psychoses, and various organic psychoses, show a similar pharmacologic treatment response. One of the current foci in drug development for schizophrenia is treatment for cognitive dysfunction, with hypothetical pharmacologic targets focused on glutamatergic, cholinergic, and serotoninergic neurotransmitter systems, as well as nonpharmacologic interventions (i.e., neuromodulation).

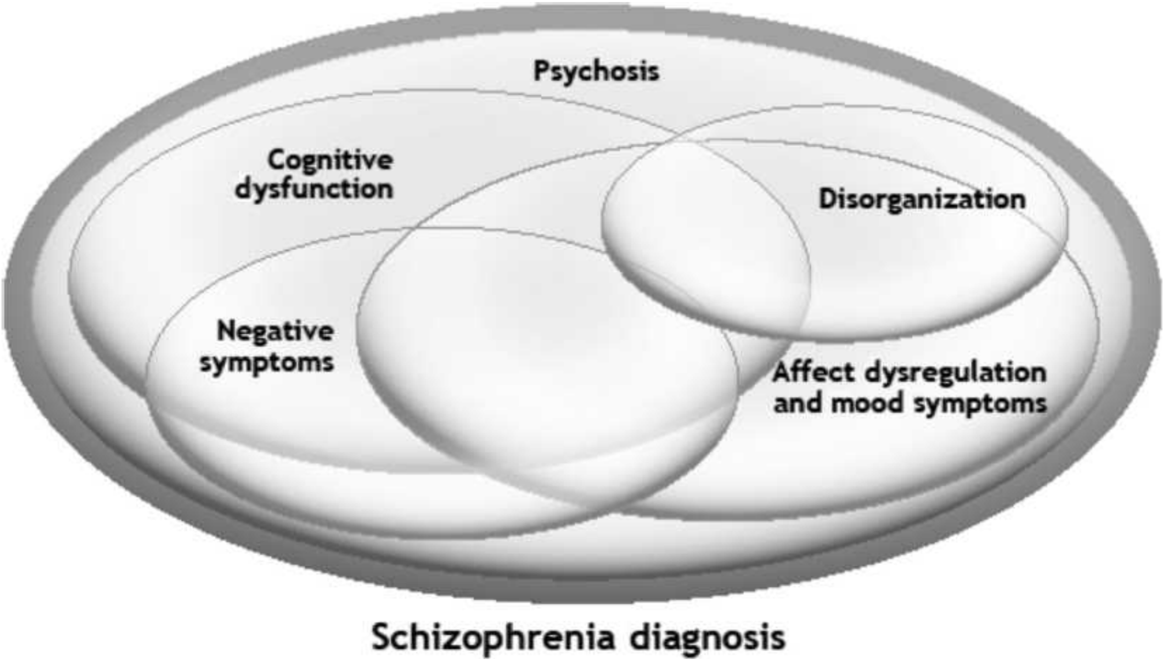

FIGURE 12.16–1. Symptom dimensions of schizophrenia.

PSYCHOSIS SPECTRUM DISORDERS: DOES PSYCHOSIS CUT ACROSS THE FORMAL DIAGNOSTIC CATEGORIES?

Among the clinical manifestations of psychiatric illnesses, psychosis represents perhaps the most dramatic and clinically captivating phenomenon. Historically, in early German and French schools of psychiatry in the 18th and early 19th centuries, psychosis was described as a single clinical entity and often was the hallmark of mental illness. Since 1896, when Emil Kraepelin demarcated “dementia praecox” from “manic depressive insanity,” the nosologic (categorical) mode of thinking and classification has been predominant in psychiatry worldwide. This is despite the fact that Kraepelin’s concept underwent a complex evolution that led him to a considerable personal uncertainty about making clear distinctions among psychosis spectrum disorders, such as dementia praecox, manic depressive psychosis, intermediate conditions, and paranoid disorders. In the current psychiatric classification, psychosis is acknowledged as a clinical phenotype within numerous psychiatric disorders, including schizophrenia, mood disorders, substance use disorders, cognitive and neurodevelopmental disorders, and disorders related to general medical conditions, although it is not an obligate component of their formal diagnoses (Fig. 12.16–2).

Large-scale research studies that utilized multivariate factor analytic approaches applied to samples spanning various categorical diagnoses suggest that psychosis represents a distinct clinical domain overlapping a broad spectrum of psychiatric disorders. The strategy of breaking down phenomena into distinct dimensions has been routinely used in many fields of medicine. But, at the present time, there is not a full consensus on how to subclassify various psychosis spectrum disorders into meaningful, biologically based dimensions. In recent studies of patients with psychotic disorders (schizophrenia, schizoaffective disorder, psychotic mood disorders, and other functional psychoses) and their biologic relatives, five distinct symptom domains spanning these diagnoses were described, including reality distortion (psychosis), disorganization, mania, depression, and negative symptoms commonly referred to as the “pentagonal model” of psychotic disorders. Different diagnostic categories showed distinct dimensional profiles, with schizophrenic persons scoring highest in the reality distortion, negative, and disorganized symptoms, and affective diagnoses scoring highest in manic and depressive dimensions and lower in the psychosis and negative symptoms. Interestingly, the mania factor is considered to be the best discriminator between schizophrenia and mood disorder diagnoses, while psychosis turned out to be a domain of overlap between these diagnostic categories. In all dimensional constructs, schizophrenia patients typically show higher variability compared to patients with mood disorders, suggesting that bipolar disorder diagnosis may signify a better-defined phenomenologic construct than schizophrenia, which is known to be highly heterogeneous. However, multiple questions remain, including whether or not the phenomenologic structure of psychosis is the same across diagnoses, what is the extent of similarities and uniqueness of underlying genetic and pathophysiologic mechanisms of psychosis is, and whether various psychosis spectrum disorders show the same molecular basis, treatment response, and functional outcomes.

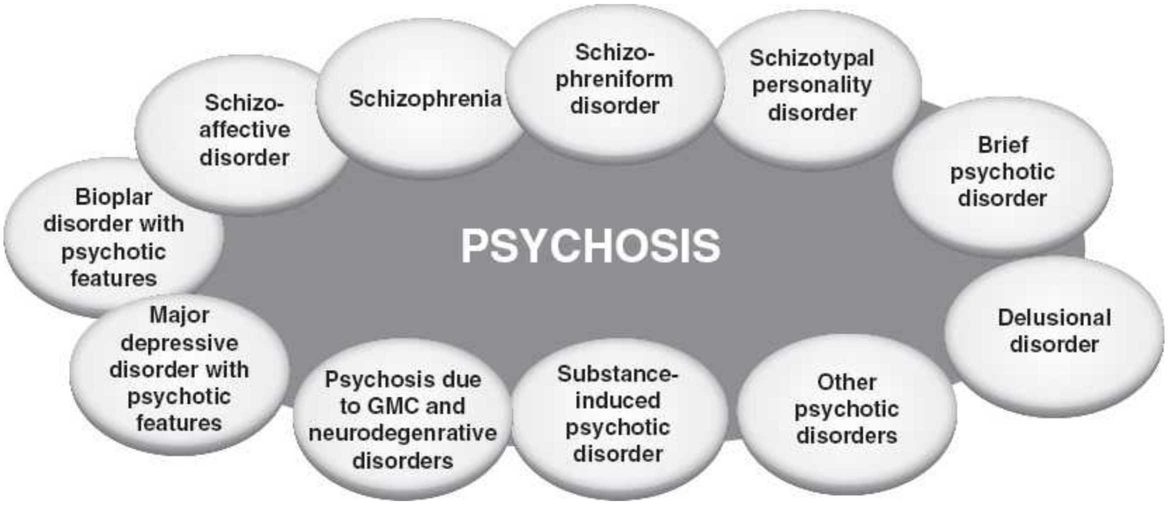
PSYCHOSIS
Bioplar disorder with psychotic features
Schizo-affective disorder
Schizophrenia
Schizo-phreniform disorder
Schizotypal personality disorder
Brief psychotic disorder
Delusional disorder
Other psychotic disorders
Substance-induced psychotic disorder
Psychosis due to GMC and neurodegenrative disorders
Major depressive disorder with psychotic features

FIGURE 12.16–2. DSM-5 diagnoses overlapping the psychosis dimension. GMC, general medical conditions.

BIOLOGIC FACTORS IN PSYCHOSIS: EVIDENCE FROM TRANSLATIONAL NEUROSCIENCE INVESTIGATIONS

Genetic Studies

Psychotic illnesses are broadly heterogeneous disorders, not caused by a single etiologic factor, but rather due to a complex interplay of various endogenous and environmental influences. There is consistent evidence that a genetic component contributes to the etiology of psychosis; however, little is known for certain about particular genomic regions or individual risk genes. Recent findings from large-scale whole genome studies provide evidence for contribution of numerous genes, each of a very small effect, to ethiopathogenesis of schizophrenia. Moreover, a substantial overlap in genetic susceptibility across a broad spectrum of psychiatric disorders, for example, schizophrenia, bipolar disorder, and autism, has been observed. Some of the identified candidate genes show strong associations with dimensional constructs, such as psychosis or mood symptoms, that are not projected directly onto categorical diagnoses. Hypothetically (not yet well-established) genes or sets of genes, interacting with epigenetic and environmental factors, may predetermine vulnerability to psychosis and other clinical disease manifestations. Depending on genetic influence and other endogenous (e.g., neurodevelopmental) and/or environmental interactions, psychosis may coexist with different clinical phenotypes (e.g., mood symptoms or cognitive dysfunction), composing epiphenomenal categorical diagnoses, such as schizophrenia. This conceptualization of psychosis draws preliminary support from epidemiologic and molecular genetic studies.

Family Studies. Family studies of schizophrenia and affective psychoses have established that psychosis strongly aggregates in families. Based on current knowledge, having a relative with schizophrenia or bipolar disorder is the single most powerful risk factor for developing psychosis. The lifetime risk for developing schizophrenia increases approximately 8- to 12-folds in first-degree biologic relatives of schizophrenic probands. Although the results of genetic studies in affective psychoses are less consistent, the familial aggregation of bipolar disorder and major depressive disorder is well-established. First-degree relatives of individuals with bipolar I disorder have elevated rates of bipolar I disorder (4% to 24%), bipolar II disorder (1% to 5%), and major depressive disorder (4% to 24%). While it was originally thought that schizophrenia and affective psychoses are inevitably segregated, recent reports challenge this. Family studies show that schizophrenia and affective psychoses frequently co-occur in the same families, suggesting shared familial risk. For example, a large epidemiologic study, using a genetically homogeneous population isolate in north-eastern Finland, reported that 16% of siblings of schizophrenic probands carried a lifetime diagnosis of psychotic disorder, including schizophrenia (0.5%), schizoaffective disorder (3.3%), bipolar I disorder (1.1%), major depressive disorder with psychotic features (2.2%), alcohol-induced (2.2%), and other substance-induced (1.1%) psychoses, delusional disorder (1.6%), and psychotic disorder, not otherwise specified (4.4%). Over half of the siblings (54%) in this study had a lifetime diagnosis of any mental illness, with major depressive disorder (15.3%) and anxiety disorders (12%) leading among nonpsychotic diagnoses. Numerous reports confirmed that the increased risk for psychotic illness in relatives of individuals with schizophrenia does not appear to be confined to schizophrenia alone. Likewise, bipolar illness has been associated with an increased risk of schizophrenia in relatives. In one of the recent family studies, it was reported that relatives of women with early onset of bipolar disorder had the highest morbid risks for both bipolar illness and schizophrenia. The presence of more than one patient with bipolar disorder in a family increased the risk for schizophrenia nearly fourfold. In a different study, it was shown that affective disorders were more frequently inherited from the same parental side of the family as schizophrenia psychosis, suggesting that the same genetic factors could contribute to susceptibility to both schizophrenic and affective psychoses. Several studies showed that, in fact, parents of schizophrenic probands suffer from affective disorders (especially, major depressive disorder) more often than from schizophrenia. Schizoaffective disorder occurs at similarly increased rates both in families of probands with schizophrenia and bipolar disorder. Likewise, schizophrenia and bipolar disorder have been shown to occur at increased rates in relatives of probands with schizoaffective disorder.

Further, there is growing evidence that certain clinical phenotypes and symptom dimensions may predict familial risk of psychosis across categorical diagnoses. For example, in a large family study from the Netherlands, a syndrome, characterized by bizarre behavior, inappropriate affect, catatonia, and poor rapport, predicted psychosis in biologic relatives of schizophrenic probands. The syndrome characterized by affective blunting and insidious early onset of illness predicted psychosis in the first-degree relatives of probands with various types of functional psychoses. A lifetime history of mania specifically predicted affective psychosis in the relatives of probands with various psychotic disorders. Similarly, in a recent study, conducted in a sample of psychotic inpatients and their 2,987 first-degree relatives in Spain, severe primary negative symptoms in probands were related to familial liability for schizophrenia, whereas familial morbid risk of mood disorders was predicted by mania, depression, and catatonia syndromes in probands. Importantly, this familial association pattern was completely independent of the probands' diagnosis of schizophrenia or major mood disorder. These studies provide evidence that familial liability for psychosis clearly overlaps DSM categories of schizophrenia and mood disorder. From a dimensional perspective, psychosis may represent a shared phenotype with unique genetic etiologies, running through family generations.

Twin Studies. Twin studies show that the concordance rate for schizophrenia is higher in monozygotic twins (47% to 56%) than in dizygotic twins (12% to 16%), suggesting a strong heritability component for the illness. It has been reported that the more severe the schizophrenia manifestations, the more likely it is for the twins to be concordant for the disorder. Some European studies have reported a concordance rate for monozygotic twins of over 80% in cases of severe schizophrenia with typical core symptoms. These observations suggest that the schizophrenia diagnosis includes clinically and etiologically heterogeneous subgroups with potentially distinct underlying genetic backgrounds. Further, twin studies suggest that a schizophrenia diagnosis in one twin equally increases the risk for schizophrenia and affective disorders in the co-twin. A substantial overlap in genetic risk for schizophrenia, schizoaffective, and manic phenotypes was reported in a sample of 77 monozygotic and 89 same-sex dizygotic twin pairs, ascertained from the Maudsley Twin Register in London. In this study, if one member of a monozygotic twin pair had schizophrenia, there was about an 8% chance of schizoaffective disorder diagnosis in the co-twin and an 8% risk of mania. Another report, based on the Maudsley twin series, showed that the maximum monozygotic/dizygotic concordance ratio was produced by a combination of schizophrenia, affective disorder with mood-incongruent psychotic features, schizotypal personality disorder, and atypical psychosis, suggesting that these psychosis spectrum disorders share a common genetic background.

Genetic Linkage Studies. Genetic linkage studies have identified several chromosomal regions showing linkage to both schizophrenia and affective psychosis, including 1q32, 10p11-15, 13q32, 18p11.2, and 22q11-13. In one study, significant linkage to psychotic bipolar disorder was obtained on chromosomes 9q31 and 8p21. Several other chromosomal sites were supportive of linkage with psychosis in bipolar pedigrees, including 5q, 6q, 8p, 13q, 15q, 17p, and 18q, which are also implicated in schizophrenia. A recent genome-wide linkage scan in schizoaffective disorder, using British and Irish pedigrees with at least one member diagnosed with schizoaffective disorder, bipolar type, confirmed the existence of loci that influence susceptibility across the functional psychosis spectrum. This study demonstrated genome-wide significant linkage at chromosome 1q42 and suggestive linkages at 22q11 and 19p13. The evidence for these linkages was equally contributed by schizophrenia pedigrees (where schizophrenia was predominant among relatives in addition to schizoaffective disorder) and bipolar disorder pedigrees (with higher proportion of affective psychosis). Noteworthy, two candidate genes, *DISC1* (disrupted in schizophrenia 1) and *COMT* (catechol-O-methyltransferase), which have been consistently implicated in schizophrenia and, more recently, in bipolar disorder, map to 1q42 and 22q11, respectively.

Studies of Individual Genes and Genome-Wide Association Studies. Studies of individual genes also support the evidence of overlap between schizophrenia and different psychotic disorders. In most studies, the candidate genes were first tested in schizophrenia, but a growing number of reports suggest that some of these genes may also be involved in etiologies of affective psychoses and even broader psychopathology, for example, autism spectrum disorders and epilepsy. In particular, recent genetic studies have shown increasing evidence for an overlap in genetic susceptibility across the traditional classification categories, including association findings at *DISC1*, *COMT*, *DAOA* (G72)/G30), (D-amino acid oxidase activator [G72]/G30), dysbindin (dystrobrevin-binding protein 1), *BDNF* (brain-derived neurotrophic factor), *NRG1* (neuregulin 1), and *RGS4* (regulator of G protein signaling 4). For instance, *NRG1* has been consistently implicated in schizophrenia; more recently, its relationship to bipolar disorder as a candidate susceptibility gene has also been reported. A recent study found the association of *NRG1* with a clinical phenotype of bipolar disorder with mood-incongruent psychotic symptoms, as well as with schizophrenia with lifetime manic episodes, suggesting that *NRG1* may confer susceptibility to a specific clinical phenotype with combined features of psychosis and mania. A number of independent genetic linkage and association studies in diverse populations support that variation in *DISC1* gene influences susceptibility to disorders of the psychosis spectrum, including schizophrenia, schizoaffective disorder, bipolar disorder, and major depressive disorder. Although this gene has been named “disrupted in schizophrenia,” in fact, the family in which the *DISC1* translocation was originally observed included cases of both psychosis and mood disorder. Recently several independent research groups have implicated the association of *DISC1* with bipolar disorder, and in one study, the strongest association was observed with schizoaffective disorder, suggesting that *DISC1* may influence susceptibility to a mixed psychosis–mood phenotype that cuts across the traditional schizophrenia/bipolar divide. *Dysbindin* is another example of a candidate gene that was initially implicated in schizophrenia. Nevertheless, recent reports have also demonstrated an association between a set of polymorphisms in this gene and a clinical subtype of bipolar disorder with recurrent psychotic episodes, specifically psychotic mania. Overall, these findings support variations in *NRG1*, *DISC1*, and dysbindin as potential genetic effects contributing (along with many other genetic and environmental factors) to susceptibility for psychosis or a mixed phenotype with features of both psychosis and mood symptoms rather than to the DSM "schizophrenia" construct, although these genetic markers were indeed originally tested as schizophrenia candidate risk genes.

Polymorphisms in *G72 (DAOA)/G30* have been broadly linked to both schizophrenia and bipolar disorder. Furthermore, a recent study from the United Kingdom, conducted with a large sample of schizophrenia and bipolar disorder cases, implicated several genetic variations in *G72 (DAOA)/G30* as specific risk markers for major mood episodes across the bipolar and schizophrenia diagnoses. Thus, although this locus was originally described as a schizophrenia risk gene, this large-scale report suggests that G72 (DAOA)/G30 appears to be more strongly associated with the mood symptoms domain, not psychosis, within the schizophrenia–bipolar diagnostic continuum.

Epidemiologic and molecular genetic studies support the hypothesis that psychosis is a clinical phenotype with multifactorial etiologies, including convincing heritability and genetic contributions. Family studies of schizophrenia and affective psychoses have established that psychosis aggregates in families. Twin studies have suggested high heritability estimates of psychosis with a complex mode of transmission. Recent whole genome linkage studies have identified a number of chromosomal loci that appear to influence susceptibility to psychosis, independently of diagnostic categories. Furthermore, studies identified several putative candidate genes that appear to be involved in genetic etiologies of schizophrenia and affective psychoses. Since schizophrenia is a relatively rare illness with presumably multiple small and/or very small genetic effects involved in its etiology, the majority of available to date genetic studies are still statistically underpowered. Geneticists suggest that tens of thousands of cases and controls may be needed to define firm associations. For example, the Schizophrenia Psychiatric Genome-Wide Association Consortium recently conducted a multistage genome-wide association study (GWAS) including 36,989 schizophrenia cases and 113,075 controls and identified 108 loci associated with schizophrenia at a genome-wide significance level. Examination of specific candidate genes at these loci suggested the involvement of dopaminergic regulation (*DRD2* gene), multiple genes involved in glutamatergic neurotransmission and synaptic plasticity (e.g., *GRM3, GRIN2A, SRR, GRIA1*), as well as neuronal calcium signaling (e.g., *CACNA1C, CACNB2*, *CACNA11*). Independent of genes expressed in the brain, associations were enriched among genes expressed in tissue that play an important role in immunity (in particular, relevant to B-lymphocyte lineages involved in acquired immunity), providing support for the hypothesized link between the immunologic effects and schizophrenia. In addition to polygenic, complex inheritance observed in schizophrenia and other psychotic disorders, there is also a strong indication that in some patients a deletion or insertion of a larger chromosomal region, so-called copy number variation (CNV), plays a crucial role in the disease pathogenesis. This could be specifically important in sporadic cases of schizophrenia since a higher frequency of *de novo* mutations has been associated with CNVs. Noteworthy, CNV markers linked to schizophrenia show substantial overlap with neurodevelopmental disorders, such as intellectual disability and autism spectrum disorders, providing support for partially overlapping etiologies underlying these clinical phenotypes.

Current clinical targets available for genetic studies, for example, “schizophrenia” or “bipolar disorder,” are highly heterogeneous and do not clearly map onto molecular or neurophysiologic disease models. Understanding the biologic effect of genetic risk markers is undoubtedly complex. Although several such candidate markers have been implicated, the associating variations are likely different in different disease subtypes and symptom clusters, and it is, therefore, difficult to determine the specific biologic effect of each risk gene characteristic for specific disease components. Various epigenetic factors, although insufficiently investigated so far, add to the complexity of the picture. In addition, environmental factors, perhaps interacting with risk genes, are likely to contribute and further complicate endogenous and environmental factors interplay. Although originally, the candidate risk genes were implicated in schizophrenia, recent findings provide evidence that they show the strongest associations with symptom dimensions, such as psychosis (e.g., *NRG1*, *DISC1*, and *dysbindin*) or mood symptoms (e.g., *G72/G30, BDNF*), across the schizophrenia–mood disorder continuum. Furthermore, recent GWAS studies support much broader genetic overlaps, from psychotic disorders to autism spectrum disorders, to epilepsy and neurodegenerative disorders, and to medical conditions (e.g., autoimmune disorders). Although intriguing, these observations are still preliminary and should be treated with caution, given the difficulties in establishing and defining evidence for genetic etiologies in complex, polygenetic disorders. Future genetic studies derived from large samples, targeting more homogeneous disease components (e.g., symptom dimensions) and/or biomarker-derived constructs underlying clinical phenomenology, are needed.

Emerging Brain-Based Biomarkers for Psychosis

Despite considerable progress in epidemiologic and molecular genetics of psychiatric syndromes made in the past decades, clinical phenotypes, such as schizophrenia or bipolar disorder, when directly applied in genetic studies, generally fail to generate consistent findings. Although multiple genome-wide “hotspots” and associated candidate genes have been implicated, the causative “psychosis genes” have not been identified. Even less is known about specific pathways and brain circuits through which these genetic effects may contribute to symptom manifestations of psychotic disorders. The inconsistent genetic findings may be, in part, explained by the presumed complexity of the pathophysiologic mechanisms of psychiatric disorders, resulting from a multicomponent interplay of genetic/epigenetic, endogenous, and environmental factors. In addition, the phenomenologic heterogeneity of psychotic disorders, as well as the lack of biologically based definitions and clear boundaries in the existing diagnostic categories, contribute to the limited success in identifying causative genetic etiologies. Based on these kinds of considerations, the focus of genetic and translational studies has recently shifted toward disease “intermediate phenotypes.” Intermediate phenotypes (often interchangeably called disease biomarkers) are specific, measurable alterations in brain function or structure or abnormalities in peripheral systems (e.g., inflammatory or endocrine effects) capturing components of disease pathogenesis which are thought to be “intermediate” between the aberrant genes and the clinical manifestations of the disease. These alterations are frequently referred to as “disease biomarkers” and can be assessed by objective neurophysiologic, neuroanatomical, or behavioral measures. An ideal intermediate phenotype for an illness would be a measure that is associated with the clinical syndrome, heritable and state-independent. In addition, it cosegregates in families while being expressed in unaffected family members and may serve as an informative predictor or risk marker observed long before full disease manifestations. Intermediate phenotypes are a promising area of study in the search for genetic factors in psychosis which may provide more direct clues to the genetic underpinnings than the clinical syndrome itself. On the other end, dimensions of illness, such as psychosis, may be particularly useful in identifying the underlying intermediate phenotypes, as dimensions provide less heterogeneous clinical constructs.

Biomarker-focused studies in psychotic disorders have led to the recognition of an array of putative disease biomarkers. These include oculomotor abnormalities (smooth pursuit and saccadic eye movement alterations); deficits in sensorimotor gating, measured with electroencephalography-based event-related potential (ERP) and prepulse inhibition (PPI) of startle response paradigms; abnormal neurocognitive performance; as well as structural and functional brain abnormalities captured with various neuroimaging methodologies. In contrast to the considerable amount of data on schizophrenia, much less is known about intermediate phenotypes of other psychotic disorders, although selected candidate biomarkers have recently begun to be explored in disorders of neurodevelopmental and affective spectrum. Similar to overlapping clinical symptom manifestations, the three most common psychotic disorders, schizophrenia, schizoaffective disorder, and bipolar disorder, appear to share common etiologic determinants and biomarker signatures (Fig. 12.16–3). Within the array of neurophysiologic, anatomic, and cognitive measures, there are specific elements that can be targeted as candidate psychosis biomarkers.

Oculomotor Alterations in Psychotic Disorders. Similar abnormalities in smooth pursuit eye movement (i.e., inability to closely follow a moving target) and saccadic eye movement disinhibition have been observed in schizophrenic and psychotic bipolar disorder patients, as well as in their first-degree relatives and persons with schizotypal personality disorder. These oculomotor phenotypes are thought to be mediated by the frontal cortex and associated circuits and have been linked to working visual–spatial memory function as well as deficits of inhibitory control. These oculomotor deficits are stable, trait-like characteristics observed in both acute (psychosis, acute mania) and remission disease phases and appear to be minimally influenced by medications. Recent genetic studies of oculomotor biomarkers suggested a linkage of smooth pursuit eye movement alterations to chromosome 6p21 in schizophrenia families. Other studies using fMRI simultaneously with eye tracking during the imaging sessions have reported that schizophrenic patients tend to activate different brain regions in order to maintain smooth visual tracking compared to healthy subjects. Patients showed reduced activation in frontal and, specifically, supplemental eye-field regions, medial superior temporal, and anterior cingulate cortices. Interestingly, activation in the medial occipitotemporal cortex was increased in patients compared to healthy individuals, perhaps, reflecting the recruitment of different brain regions in order to compensate for primary prefrontal deficits underlying smooth pursuit eye movement system dysfunction in schizophrenia.

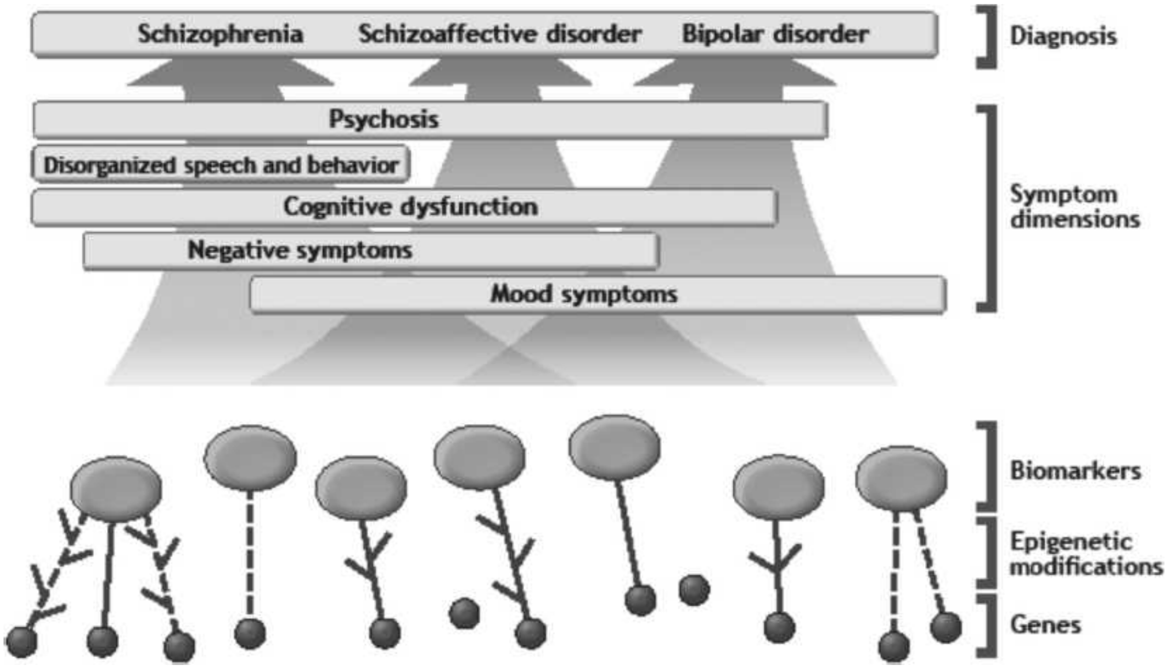
Schizophrenia
Schizoaffective disorder
Bipolar disorder
Diagnosis
Psychosis
Disorganized speech and behavior
Cognitive dysfunction
Negative symptoms
Mood symptoms
Symptom dimensions
Biomarkers
Epigenetic modifications
Genes

FIGURE 12.16–3. Clinical, biomarker, and genetic components in the schizophrenia–bipolar disorder continuum. Schizophrenia, schizoaffective disorder, and bipolar disorder are the three main psychiatric diagnoses overlapping the psychosis dimension. In addition, other symptom dimensions, such as cognitive dysfunction, disorganization, negative and affective symptoms, are variably presented in these diagnoses. These symptom manifestations are thought to result from a complex interplay of environmental and biologic factors, expressed in disease biomarkers such as altered brain anatomy, cognitive or neurophysiologic abnormalities, captured with electroencephalography, brain imaging, and other objective measurements. Some, though not all, of these biomarkers may be heritable and subserved by genetic and epigenetic effects.

Sensory and Sensory-Motor Gating Deficits in Psychotic Disorders. Sensory/sensory-motor gating is a neurophysiologic function of a normal brain that allows one to “filter out” irrelevant and/or excessive environmental stimuli. It has long been observed that patients with evolving psychosis and mania display hypersensitivity to environmental stimuli as well as aberrant stimuli attributions, thought to be mediated by altered sensory and sensory-motor gating mechanisms. These phenomena are typically measured by various electroencephalography (EEG) ERP paradigms as well as electromyography (EMG), for example, PPI paradigms.

ELECTROENCEPHALOGRAPHY-BASED BIOMARKERS IN PSYCHOTIC DISORDERS

Auditory Paired Stimuli ERPs. An overall pattern of gating deficits characteristic for psychosis is a reduced sensory, and sensory-motor gating function thought to capture neural “hypersensitivity” to environmental stimuli. In the standard auditory “paired stimuli” paradigm, the so-called “P50” response amplitude (i.e., the largest EEG wave within 50 ms after the stimulus presentation) is measured in response to each of two auditory “clicks.” In healthy individuals, the P50 wave following the second stimulus is lower than that of the first stimulus, reflecting normal sensory gating function. In contrast, both schizophrenia and bipolar patients (specifically, patients with acute mania) show diminished suppression of P50 ERP to the second stimulus, that is, insufficient sensory gating. These sensory gating deficits are thought to be stable “trait” measures largely independent of the current severity of psychotic or mood symptoms or treatment effects. Some studies report more prominent P50 abnormalities in patients with disorganized schizophrenia. In addition, associations between disrupted P50 and negative symptoms have been reported. An intriguing observation suggesting that sensory gating deficits may represent a unique psychosis biomarker has been recently reported in patients with bipolar disorder. Persons with bipolar I disorder and a lifetime history of psychosis and those with schizoaffective disorder showed P50 deficits similar to those observed in schizophrenia, whereas bipolar patients with no history of psychosis exhibited normal P50 suppression. “Schizophrenia-like” P50 suppression deficits are also observed in people with schizotypal personality disorder as well as in individuals at high risk for developing psychosis, such as subjects originating from high genetic density families (i.e., those with several psychotic family members) and individuals with schizophrenia prodrome. Evidence for a linkage between disrupted P50 gating and chromosome 15q14 has been reported in several studies, specifically to the promoter region of the α-7 nicotinic cholinergic receptor subunit gene, in both individuals with schizophrenia and their biologic relatives.

Odd Ball Paradigm and P300 ERP. P300 is typically measured in auditory or visual “oddball” paradigms where frequently presented stimuli (“targets”) are interleaved with infrequent and unexpected atypical (“oddball”) stimuli and refers to a positive EEG wave occurring within 300 ms following an “oddball” stimulus. Several large meta-analyses reported reduced P300 ERP amplitude and delayed P300 latency in probands with schizophrenia and their biologic relatives, as well as in “mild psychosis phenotypes” such as schizophrenia prodrome and schizotypy. Recently studies of P300 ERP have been extended to bipolar disorder and suggested similar “schizophrenia-like” alterations in P300 amplitude and latency in individuals with psychotic bipolar disorder and their relatives, albeit less consistently across different studies. In several studies, a frontal subcomponent of P300 ERP was reported to correlate with the severity of hallucinations in patients with schizophrenia, perhaps due to the shifting of attentional resources from external to internal (psychotic) stimuli. Similarly, patients with psychotic bipolar disorder showed reductions in frontal lobe P300 voltage, suggesting a component of overlapping neural substrate for dysfunction in both schizophrenia and affective psychosis. Diminished P300 amplitude and increased latency have also been linked to the severity of negative symptoms, illness duration, and early age at psychosis onset, as well as cognitive dysfunctions, that is, deficits in attention. Several studies conducted on healthy twin pairs suggested moderate heritability estimates for the P300 ERP biomarker. Interestingly, the P300 heritability in females appeared to be lower (45% compared to 56% in males), suggesting that endogenous or environmental factors may have a stronger modifying influence on this ERP biomarker in females. Genetic association studies linked P300 alterations to *COMT* and *NRG1* polymorphisms in patients with schizophrenia.

Sensorimotor Gating Deficits in Psychosis: Prepulse Inhibition. PPI paradigms measure inhibition of the startle reflex. In healthy individuals, a weaker “prepulse” stimulus leads to an inhibition of a response magnitude to a startle (“pulse”) stimulus, used as an index of sensorimotor gating. In individuals with psychotic disorders, insufficient PPI reflecting an exaggerated startle response is commonly observed. Diminished PPI has been reported in patients with schizophrenia, schizotypal personality disorder, and both affected and unaffected biologic relatives of schizophrenia probands. Similar findings have been noted in persons with bipolar disorder, especially during the acute manic phase, although the studies on bipolar disorder are few and less consistent. Several studies reported associations between PPI deficits and florid psychotic symptoms, such as disorganized thought processes and behavior, greater distractibility, and bizarre psychotic behavior. However, at least a few studies in schizophrenia observed disrupted PPI even in remission phases, supporting this candidate biomarker as a stable “trait” measure. Likewise, individuals with acute psychotic mania show significantly reduced PPI and less startle habituation, along with a significant reverse correlation between PPI sensorimotor gating and the severity of psychotic symptoms. In contrast, euthymic bipolar patients failed to demonstrate abnormalities in PPI. These observations of similar PPI deficits in schizophrenia and psychotic mania suggest that abnormal sensorimotor gating may represent one of the putative biomarkers subserving psychosis as a unique clinical dimension.

Cognitive Dysfunction as a Candidate Psychosis Biomarker. Cognitive dysfunction has recently drawn interest as a distinctive dimension across psychotic disorders, as well as a potential treatment target. The presence of specific cognitive deficits is not yet part of the diagnostic DSM-5 criteria for schizophrenia; nonetheless, it is recognized as a common feature of the disorder. Extensive cognition studies have shown stable, trait-like deficits in the domains of working memory, verbal episodic memory, attention, processing speed, problem-solving, and executive function in individuals with schizophrenia. Patients with bipolar disorder manifest deficits in overlapping cognitive domains, with attention, working memory, and verbal memory difficulties being among the most prominent, albeit generally less severe, than those found in schizophrenia. A considerable clinical heterogeneity of the bipolar phenotype correlates with variable presentations of cognitive deficits, with some patients manifesting more severe “schizophrenia-like” abnormalities and others showing little or no alterations in cognitive function. Studies demonstrate that a lifetime history of psychosis, a higher number of past manic episodes, a longer overall duration of illness, and a higher number of past hospitalizations are associated with more severe cognitive deficits in bipolar disorder. Moreover, bipolar patients with a lifetime history of psychosis appear to have a different profile of neurocognitive deficits than those without a history of psychosis, manifesting more severe impairment on measures of executive function and spatial working memory. More severe bipolar patients, especially those with psychosis, begin to resemble schizophrenia patients with regard to the level and profile of neurocognitive deficits. One of the commonly reported cognitive features that may help to disambiguate schizophrenia and bipolar disorder is a general intelligence level, especially premorbid intellectual functioning, with lower scores typically observed in patients with schizophrenia but normal/near normal general intelligence levels in bipolar disorder. Extensive research work has been done in an attempt to discover genetic associations for cognitive dysfunction in schizophrenia and related psychotic disorders, with most consistent findings supporting links for *COMT* and executive function, *DISC1* and verbal declarative and working memory deficits in schizophrenia, and less consistently, bipolar probands, and their biologic relatives.

Anatomical Brain Alterations in Psychotic Disorders. Anatomical brain changes have been examined as candidate psychosis biomarkers across schizophrenia and related psychotic disorders. Studies in schizophrenia have consistently shown a modest but measurable reduction in whole brain volume, particularly in gray matter (averaging 2% to 8% for various brain structures, based on meta-analyses), as well as an increase in ventricular size. Multiple areas of the brain have been reported to demonstrate volumetric changes, typically with the largest effect sizes in the frontotemporal regions but also in the parietal cortex and several subcortical structures (e.g., thalamus, basal ganglia). Several meta-analyses of volume change in schizophrenia reported whole brain and, particularly, frontotemporal gray matter volume reductions in individuals at the time of their first psychotic break, preceding any treatments.

The structural findings in psychotic mood disorders are less consistent, ranging from decreased, to normal, to increased gray and white matter volumes in various brain regions. Recent large-scale psychosis-focused studies report regionally overlapping but more localized and lesser in magnitude gray matter volume reductions in psychotic bipolar disorder compared to schizophrenia and schizoaffective disorder. The most reproducible regional volume reductions in psychotic bipolar disorder are typically localized to anterior limbic regions, that is, orbitofrontal, anterior cingulate, insular cortex, as well as amygdala and anterior temporal cortex. However, volume increases in the amygdala have also been reported. Some studies comparing individuals across the psychosis dimension find no abnormalities in gray matter volume, while volume deficits of white matter were evident throughout the neocortex and brainstem, with similar regions of white matter loss found in individuals with schizophrenia and bipolar disorder.

The question of whether volume loss is progressive during the lifetime course of psychosis has been explored. Measurable gray matter volume and cortical thickness changes are already observed in individuals near psychosis onset, progress with illness duration (more aggressively during the initial years of illness in some cases), and ultimately result in characteristic gray matter alterations found in chronic psychosis samples. Data from relatives with psychosis spectrum disorders (e.g., schizotypal personality disorder) demonstrate milder but similar gray matter volume reductions suggesting a heritable link between psychosis manifestations and brain structure alterations. Moreover, recent large-scale reports demonstrate overlapping gray matter volume reductions in the psychosis probands (including schizophrenia, schizoaffective disorder, and psychotic bipolar disorder) and their first-degree relatives with psychosis spectrum disorders, but normal gray matter characteristics in relatives unaffected by lifetime psychosis, even in those relatives with clear nonpsychotic diagnoses (e.g., mood disorders, anxiety disorders). These gray matter alterations may capture lifetime "psychosis burden," from extensive reductions in probands where psychosis is fully manifested to similarly distributed but milder alterations in relatives with mild psychosis phenotypes (e.g., schizophrenia spectrum personality disorders), to normal gray matter in relatives without lifetime psychosis. Moreover, associations between lifetime duration of psychosis and gray matter volume, with more loss observed as the illness progresses, have been reported in schizophrenia and schizoaffective probands in the frontal, temporal cortex, basal ganglia, and thalamus, the regions long implicated in psychosis in both early-course and chronic samples.

The exact mechanisms underlying these brain structure changes are unknown but can be interpreted in the context of available postmortem findings of cellular neocortical pathology in psychotic disorders. Histologic examination of schizophrenia cases indicates absence of neuronal loss, yet altered neuronal cell packing, presumably due to reduced interstitial neuropil, resulting in higher neuronal density and overall decreased cortical thickness. In contrast, analyses from bipolar cases show decreased neuronal and glial density but normal overall cortical thickness. Dendritic spine loss, as well as reduced dendritic length in dorsolateral prefrontal cortex, has been observed in both schizophrenia and bipolar postmortem tissue samples. These postmortem findings suggest both overlapping and unique anatomical underpinnings for schizophrenia and bipolar disorder, providing plausible cellular correlates for the structural brain changes captured with imaging techniques.

Furthermore, longitudinal studies provide support for the measurable effect of various psychotropic medications on brain structure. Long-term treatment with first-generation antipsychotics has been linked to increased gray matter volume in basal ganglia. Both first- and second-generation antipsychotics are associated with decreased gray and white matter volume in the frontotemporal regions. Lithium has been linked to higher gray matter volume/density in diffuse neocortical regions as well as a larger volume of specific subcortical structures, that is, the amygdala, in psychosis and mood disorder samples. Given that the majority of psychotic and mood disorder patients undergo years of treatment with various psychotropic agents, it is possible that the gray matter deficits consistently reported in patients with schizophrenia, and schizoaffective disorder are, at least in part, accounted for by the effect of a lifetime antipsychotics use, whereas a relative gray matter volume preservation in bipolar patients could be secondary to a medication effect (i.e., chronic treatment with lithium) and that, even if a “primary” disease-associated loss of neocortical volume exists in these patients, it could be obscured by a volume-enhancing effect of chronic lithium use. Disentangling “primary” disease and medication effects on brain structure presents considerable difficulty and requires carefully designed longitudinal research protocols.

Some, though not all studies, have reported associations between brain anatomy changes and clinical manifestations in schizophrenia and related disorders. In particular, reductions in the frontotemporal volume have been linked to severity of positive symptoms in schizophrenia and schizoaffective disorder; reductions in various structures within the medial temporal lobe have been shown to correlate with the severity of delusions. Additionally, volume reductions in the medial temporal lobe have been linked to declarative memory dysfunction in first-degree relatives of schizophrenic probands. Subsequent data suggest that the various subgroups within bipolar disorder may have different structural brain characteristics, depending on whether they have psychosis. The majority of studies report similar brain structure changes in persons with schizophrenia and psychotic bipolar disorder, perhaps as one of the overlapping intermediate phenotypes indicative of psychosis liability.

Biomarker-Based Approaches to Psychosis Dimension: Challenges and Future Directions. Disease biomarkers (sometimes referred to as “intermediate phenotypes”) represent smaller, homogeneous, measurable units of illness that are thought to be more proximal to the pathogenesis of illness and/or genetic etiologies than the global diagnosis constructs. At first, the intermediate phenotype/biomarker-focused approaches were used in schizophrenia research and resulted in the recognition of several reproducible cognitive, neurophysiologic, and neuroanatomical markers. More recently, these same candidate biomarkers have been used to explore different psychotic disorders in the hope of developing valid, objectively defined psychosis biomarkers. As reviewed above, among the variety of biomarker measures, there are specific elements that appear to be promising as indicators of the “psychosis dimension” cutting across DSM-based psychosis diagnoses. Further progress in psychosis research can move the field past early “proof of concept” biomarker studies. Parsing complex brain disorders, such as psychosis, based on biomarker approaches within and across psychiatric syndromes can provide more precise diagnostic algorithms and paths for the development of targeted treatments.

However, many challenges will need to be addressed in order to move the field toward successful translational and clinical implementation of biomarkers. To date, the majority of psychosis biomarker studies have focused on a single biomarker, for example, cognition or EEG. While it is unlikely that such single biomarker measures can adequately capture the complexity of brain alterations underlying psychosis, comprehensive biomarker batteries are lacking. It remains unknown which combination(s) of specific measures would represent the most parsimonious “psychosis biomarker battery.” There is a critical need for large-scale psychosis studies that incorporate multimodal biomarkers and link biomarker batteries to clinical manifestations. A few biomarker-focused psychosis consortia (e.g., the Bipolar-Schizophrenia Network for Intermediate Phenotypes [B-SNIP]) have made progress, but more studies—including biomarker replication in large independent samples—are needed.

Furthermore, little is known about specific functional, anatomical, and molecular features underlying current “psychosis biomarker” constructs. Unlike biomarkers in medicine, for example, glucose levels or Hemoglobin A1C for diabetes, currently available candidate biomarkers in psychiatry fail to capture specific mechanistic elements of complex brain disorders. Further basic neuroscience and translational work are urgently needed in order to link human *in vivo* brain-based biomarkers to potential molecular targets, for example, via animal models or live cell studies. In parallel, efforts are also needed to parse the biologic heterogeneity of psychiatric syndromes to guide successful clinical translation. Additionally, longitudinal studies are needed to investigate the prognostic value of psychosis biomarkers. Furthermore, pharmacologic target studies and clinical trials incorporating biomarker measures will be critical for capturing specific aspects of disease mechanisms and generating “actionable” biomarkers in psychiatry. Cumulatively, these strategies may move the field closer to elucidating disease mechanisms and exploiting mechanistically informed therapeutic targets—the bases for successful clinical translation.

At present, the clinical utility of the “psychosis biomarker” constructs is unclear. In contrast to many other fields of medicine where objective tools are the cornerstone of diagnostic algorithms, biomarker measures are not available to support the vast majority of psychiatric diagnoses. Experimental research biomarker batteries are complex and costly and require substantial technical and analytic expertise as well as further development in automatic quantification tools. It remains unclear which specific biomarkers (and/or biomarker batteries) are useful in informing clinical diagnostic algorithms and predicting treatment response and, thus, represent suitable candidates for implementation into clinical practice. Substantial work is needed to bridge the gap between research and clinical practice in order to demonstrate the translational applicability of biomarker findings in psychiatry.

Ultimately, to successfully address the challenging task of establishing clinical utility, the field will have to bring brain-based biomarkers to an individual patient level. This calls for sophisticated computational approaches, which would allow biomarker characterization and “biotyping” in individual patients, in real-time, with subsequent application of these biomarkers as a means to guide clinical decision-making. Changes in psychiatric training and in insurance policies, so they will support objective brain measures as part of routine clinical workups, are examples of challenging tasks lying ahead.

To conclude, significant challenges remain for direct clinical translation of biomarkers in psychiatry, though none are insurmountable in light of fast progress in other fields of medicine towards similar translational goals. The developing field of neurobiologic psychiatry offers unique paths to tackle these challenges via multidisciplinary efforts that bridge research and clinical practice. If successful, these efforts may generate stronger diagnostic algorithms and offer novel therapeutic targets for psychiatric disorders.

PSYCHOSIS AS A CLINICAL CONTINUUM: FROM ATTENUATED PSYCHOSIS PHENOTYPES TO FRANK PSYCHOTIC DISORDERS

Psychosis manifestations include a broad array of symptoms of various severities, which appear to lay on a continuum from mild subthreshold psychotic phenomena to frank and florid psychosis. Mild sporadic psychotic symptoms, such as illusions, mild hallucinations (e.g., undifferentiated noises or calling a person’s name), ideas of reference, magical thinking, unusual beliefs, and perceptual experiences, and the like are fairly common in otherwise healthy individuals who don’t carry any diagnosis of psychiatric illness. Although these symptoms are psychotic by nature, they are considered to be below the threshold for a formal psychosis diagnosis. It appears that there may be no discrete demarcations between frank psychosis and minor subclinical psychotic symptoms other than severity, breadth, and persistence; rather, psychosis represents a clinical continuum with multiple subtypes and grades of symptom manifestations.

Large longitudinal studies evaluated the frequency of mild psychotic symptoms in the general population and found that subclinical psychotic experiences are not uncommon. For example, the large community sample study using the British National Psychiatric Morbidity Survey data showed that 4.4% of the general population reported incident psychotic symptoms over an 18-month period. The vast majority of individuals who reported such psychotic symptoms did not meet the full diagnostic criteria for any psychiatric disorder. These observations may indicate the presence of latent continuum-like psychopathology in the general population. Moreover, they are consistent with a multifactorial etiologic model of schizophrenia where many different genes of small effect interact with one another and with environmental factors to result in the illness. Depending on a complex interplay of genetic and environmental influences, the manifestations of psychosis can vary from mild episodic experiences (e.g., déjà vu, infrequent thought blocking, or visual illusions) to moderate psychotic-like symptoms (e.g., magical thinking, ideas of reference, bodily illusion, or out-of-body experiences), or to severe and persistent frank psychosis, fulfilling DSM criteria (e.g., delusions or hallucinations). From the psychosis continuum perspective, subthreshold psychotic symptoms are considered indicators of proneness to psychosis, with clinical psychosis emerging perhaps when a critical genetic threshold is reached, accompanied by modifying or triggering epigenetic and environmental factors (Fig. 12.16–4).

Clinically, the continuum nature of psychosis is well illustrated by premorbid and prodromal symptoms in individuals with schizophrenia and affective psychoses, psychopathology in biologic relatives of psychotic probands, as well as by certain types of personality disorders, displaying a wide range of psychotic symptoms, varying from very mild psychotic-like phenomena to frank psychosis during decompensation phases.

Premorbid and Prodromal Signs and Symptoms in Schizophrenia and Affective Psychoses

The course of schizophrenia is generally conceptualized as starting with premorbid (preprodromal) signs and symptoms and then a prodromal phase of illness, leading to typical clinical manifestations as the illness evolves. However, it is not always possible to clearly demarcate these periods due to the presence of schizophrenia spectrum personality disorders and traits characteristic for a proportion of individuals who subsequently develop schizophrenia, as well as minor psychotic, affective, and behavioral symptoms that sometimes appear long before the debut of schizophrenia. In the typical, although not invariable, premorbid history of schizophrenia, patients have had schizoid or schizotypal personality features and are characterized as quiet, passive, introverted, emotionally aloof, and reclusive. Other characteristics, more typical for schizoid personality, include a low need for emotional bonds: In childhood, these individuals have few friends. By the adolescent period, they may show no interest in dating and avoid school sports and competitions. Instead, they prefer solitary activities and seem content by themselves. When schizotypal features are predominant, thinking, perception, and behavior appear to be odd or peculiar. Various psychotic-like symptoms, such as magical thinking, peculiar beliefs, ideas of reference, perceptual illusions, odd fantasies, or derealization–depersonalization symptoms, are commonly present. Especially during the adolescent period, peers may exclude them from group activities or tease them, as schizotypal individuals have poor communication skills and distorted social judgment. This often leads to increased anxiety, feelings of “not fitting in,” and further withdrawal from the social group. Although different types of personalities (including paranoid, avoidant, obsessive-compulsive, and, less commonly, borderline) are also seen in the premorbid period in individuals with schizophrenia, schizoid and schizotypal personality traits are believed to be the most common. Still, the presence of these premorbid and prodromal symptoms does not inevitably result in conversion to schizophrenia or other psychotic disorder.

Although the first hospitalization is often taken to be the beginning of the illness, in many cases, premorbid and prodromal signs and symptoms may have been present for months or years prior. The majority of schizophrenia cases are characterized by an insidious, gradual onset of psychosis with no clear date of its first manifestation. The prodromal symptoms may start with somatic complaints, such as headaches, digestive problems, autonomic symptoms, fatigue, and generalized weakness. Family and friends commonly notice that the person has changed, become withdrawn and isolated, and is no longer functioning well in occupational, personal, and social settings. Deterioration in personal hygiene and grooming decreased appetite, and peculiar eating habits are also common. During this stage, a patient may begin to develop new, unusual interests in abstract ideas, philosophy, occult, or religious topics. Gradually, although acute onset may also occur, subthreshold psychotic symptoms, such as unusual speech, strange perceptual experiences and bodily illusions, bizarre ideas, and markedly peculiar behavior, build up and subsequently progress to the level of frank psychosis in the form of hallucinations, delusions, typical thought and speech disorders, and/or grossly disorganized behavior.

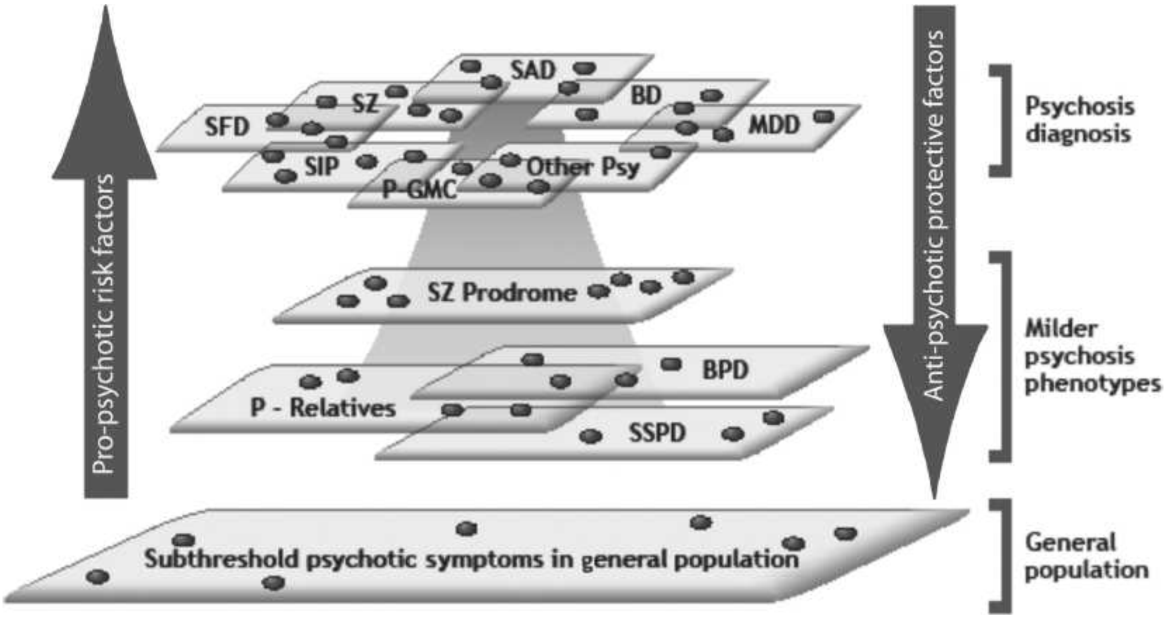
Pro-psychotic risk factors
SAD
SZ
BD
SFD
MDD
SIP
P-GMC
Other Psy
SZ Prodrome
BPD
P - Relatives
SSPD
Subthreshold psychotic symptoms in general population
Anti-psychotic protective factors
Psychosis diagnosis
Milder psychosis phenotypes
General population

FIGURE 12.16–4. Threshold model of psychosis. In the general population, considerable evidence indicates the presence of fleeting psychotic symptoms without any functional impact. When more severe or persistent, psychotic symptoms appear at subthreshold of formal psychosis diagnosis and manifest in entities like schizophrenia prodrome and schizophrenia spectrum personality disorders, as well as are frequently seen in biologic relatives of probands with psychotic illnesses. In a proportion of these individuals, psychosis may reach the level of florid clinical manifestations severe enough to fulfill criteria for DSM diagnosis. Genetic, epigenetic, other biologic, and environmental factors can protect against or advance psychosis. Examples of propsychotic risk factors include obstetrical complications, drug abuse, specifically, adolescent cannabis use, and psychosocial stressors. Less is known about antipsychotic resilience factors, though high premorbid functioning, structured environment, and cognitive remediation administered early in the course of evolving psychosis appear to have protective effects. •, candidate psychosis genes; P-Relatives, psychotic relatives of probands with psychotic disorder; SSPD, schizophrenia spectrum personality disorders (cluster A: Schizoid, paranoid, schizotypal); BPD, borderline personality disorders; SZ, schizophrenia; SAD, schizoaffective disorder; BD, bipolar disorder with psychotic features; MDD, major depressive disorder with psychotic features; SFD, schizophreniform disorder; SIP, substance-induced psychosis; P-GMC, psychosis due to general medical conditions; other Psy, other psychotic disorders.

The validity of the prodromal syndrome is uncertain, since symptoms are frequently recognized only after the diagnosis of schizophrenia has been made. During the prodromal phase, patients usually lack insight regarding their developing symptoms, although some verbalize feeling as if they have changed. Once schizophrenia is diagnosed, the retrospective remembrance and the patient’s subjective report of early signs and symptoms may be affected, therefore obtaining the information about the early (prodromal) period of illness from the patient’s family may be of significant help in building a diagnosis.

The presence of prodromal symptoms in affective psychoses is unclear. Although clinical observations suggest that such symptoms as mood lability, impulsivity, destructibility, and physical hyperactivity are often present long before the development of affective psychosis, the phenomenology of bipolar prodrome is not well characterized. One of the recent studies conducted with early onset bipolar I disordered individuals showed that there may be unique prodromal characteristics in patients with and without subsequent bipolar phenotype that manifests as psychotic mania. In this study, attenuated psychotic symptoms emerging late in the prodrome and increased energy and goal-directed activity were more common in individuals with later psychotic mania. Studies comparing prodromal symptoms in schizophrenia and affective psychosis revealed considerable overlap; for example, such symptoms as suspiciousness, subthreshold hallucinatory experiences, anxiety, insomnia, and impaired school or work functioning were frequently present during the prodromal period in both psychoses. On the other hand, depressed mood, suicidality, mood lability, difficulty thinking/communicating clearly, tiredness/lack of energy, obsessions/compulsions, and physical agitation were more specific for bipolar than schizophrenia prodrome.

Recently, substantial research and clinical efforts have focused on psychosis prodrome and other “clinically high risk” states, with overarching goals of identifying (i) clinical subgroups within highly heterogeneous prodromal states, (ii) predictive factors (including biomarkers) for conversion to fully manifested psychosis, and (iii) developing interventions that could prevent or, at least, delay conversion to psychosis. Recent studies have demonstrated that premorbid and prodromal symptoms, clinically well accepted as a phase in the course of schizophrenia, are not necessarily pathognomonic of schizophrenia psychosis. For example, in a study by the Melbourne group, high-risk individuals with either a family history of psychotic disorder, schizotypal personality disorder, subthreshold psychotic symptoms, or brief transient psychotic symptoms were followed for a 12-month period. In that group, 40.8% of the high-risk individuals subsequently developed psychosis; however, diagnoses ranged between schizophrenia, schizoaffective disorder, bipolar disorder, and major depression with psychosis. Other large-scale studies in this field suggest that approximately one-third of “at risk” individuals eventually convert to psychosis, one-third will continue to experience stable chronic mild psychotic phenomena without a full conversion to DSM-level psychotic illness, and one-third will improve over time, some to the level of full recovery. These results indicate that although early signs and symptoms, including prodromal psychotic-like phenomena, can be predictive of conversion to a spectrum of psychotic disorder, they do not indicate the exact type of psychosis, and the long-term outcomes are highly variable, ranging from fully manifested psychotic disorders to full recovery. Given these research outcomes, in DSM-5, a new diagnostic category, “Attenuated psychosis syndrome,” has been introduced. It refers to a broadly defined syndrome characterized by psychotic-like phenomena that are below a threshold for full psychosis, for example, the symptoms that are less severe and more transient and frequently accompanied by a relatively preserved insight. This diagnostic category is appropriate for some (but not all) individuals with prodromal or subthreshold symptoms and calls for a clinician’s judgment on where to draw the line for mild psychotic-like presentations that do not deserve DSM diagnosis and where it is appropriate to introduce this diagnosis.

Psychosis Manifestations as a Component of Personality Disorders

The occurrence of brief psychotic symptoms is common in individuals with personality disorders. Typically, psychotic experiences in these people are triggered by psychological (e.g., change in employment, loss of a relationship) or physiologic (e.g., medical illness) stressors, or substance use, when compensatory mechanisms weaken. Certain types of personality disorders are especially predisposed to psychosis, including cluster A and borderline personality disorders. In DSM-5, cluster A personality disorders include paranoid, schizoid, and schizotypal types. These personality disorders are also called *schizophrenia spectrum personality disorders* because their core features resemble symptoms of schizophrenia, albeit in a milder manifestation. In addition, evidence demonstrates the familial aggregation of these personality disorders in biologic relatives of probands with schizophrenia and other psychotic disorders.

Schizotypal Personality Disorder. Schizotypal personality disorder is described under the DSM-5-TR category of “Schizophrenia Spectrum and Other Psychotic Disorders,” emphasizing the long-acknowledged biologic link between this particular cluster A personality disorder and schizophrenia. Schizotypal personality disorder is characterized by cognitive and perceptual distortions, eccentric behavior, as well as a persistent pattern of social and interpersonal deficits. Mild psychotic symptoms, such as magical thinking, perceptual alterations, bizarre rituals, ideas of reference, paranoid ideation, derealization/depersonalization, idiosyncratic speech, and odd or eccentric behavior, are core features of schizotypal personality disorder. In response to stress, schizotypal individuals may decompensate and display frank psychotic symptoms (e.g., hallucinations, delusions, disorganization of thought process, and behavior); however, these episodes are generally brief and fragmentary. Usually, these psychotic symptoms are insufficient in duration and/or severity to draw an additional diagnosis of a psychotic disorder. Nevertheless, in some cases, clinically significant psychotic symptoms may develop that ultimately meet the criteria for brief psychotic disorder, schizophreniform disorder, delusional disorder, or evolve into schizophrenia. In these cases, a schizotypal personality disorder may be indicated in the diagnosis as a premorbid personality disorder, in parallel with a diagnosis of psychotic illness.

Evidence from research demonstrates that among cluster A personality disorders, schizotypy is most intimately related to schizophrenia and other psychoses. There is a higher density of schizotypal personality disorder cases in the biologic relatives of schizophrenia probands than in controls. Although schizotypal personality disorder occurs in 0.3% to 3% of the general population, studies have reported rates of 0.8% to 4.7% in the biologic relatives of psychotic patients. Recent genome-wide studies in schizophrenia and schizotypal personality disorder, conducted with a large sample of high-density schizophrenia families, have suggested that at least a subset of candidate schizophrenia susceptibility genes is also involved in the etiology of schizotypy in nonpsychotic relatives of schizophrenic probands.

In the large epidemiologic study called the Roscommon Family Study, which collected 534 probands with psychotic and affective disorders and 2,043 of their biologic relatives, schizotypal personality disorder showed a strong familial relationship not only with schizophrenia but also with the other major psychotic disorders. Further factor analysis from this research sample showed that schizotypy is a complex, multidimensional clinical construct whose various dimensions differ widely in the specificity with which they reflect the familial liability to psychosis. In this study, seven schizotypal dimensions were described, among which subpsychotic thought disorder, negative schizotypal signs, such as poor rapport and odd behavior, deficient occupational functioning, and social isolation/avoidance best disambiguated relatives of schizophrenic probands from relatives of matched controls. Different studies reported that the disorganization dimension of schizotypy was common to relatives of both schizophrenia and psychotic bipolar disorders but not in the relatives of nonpsychotic bipolar probands. Interestingly, schizotypal disorganization symptoms were suggested as a unique clinical phenotype associated with an increased familial risk for both schizophrenia and affective psychoses. From this perspective, schizotypy represents a subthreshold psychosis phenotype that cuts across the categorical diagnosis of schizophrenia and bipolar disorder.

Paranoid Personality Disorder. The essential feature of paranoid personality disorder is a pattern of long-standing suspiciousness and mistrust of people in general, such that their motives are interpreted as malevolent. Subthreshold psychotic symptoms are not uncommon in persons with paranoid personality disorder, of which ideas of reference and associated verbal and visual illusions are seen most often. Under stressful conditions, specifically in a situation naturally provoking paranoid feelings (e.g., when these patients are being criticized or disagreed with, relocating, or facing a new professional environment), they may decompensate and experience psychotic episodes. Among frank psychotic symptoms, paranoid and persecutory delusions are the most common, although brief visual and auditory hallucinations may also occur.

There is evidence of a genetic association between paranoid personality disorder and psychosis spectrum conditions. For example, in some instances, paranoid personality disorder may appear as the premorbid antecedent of schizophrenia or delusional disorder. An increased incidence of major depressive disorder with psychotic features has also been reported in patients with paranoid personality disorder. There have been few controlled studies of the familial risk of paranoid personality disorder based on direct interviews of relatives blind to proband’s diagnosis. However, there is substantial evidence for a familial relationship between paranoid personality disorder and psychotic disorders, with reports of a higher prevalence of paranoid personality disorder among first-degree relatives of schizophrenic, schizoaffective, and psychotic bipolar probands ranging from 1.7% to 8.6% in different studies (compared to 0.5% to 2.5% in the general population). In contrast to schizotypal personality disorder, paranoid personality disorder is less common in relatives of probands with affective psychoses, although one study reported an increased familial risk of paranoid personality disorder for unipolar depression (2.9%), which was higher than for relatives of schizophrenia patients (1.7%).

Schizoid Personality Disorder. Individuals with schizoid personality disorder are characterized by a pervasive pattern of detachment from social relationships and a restricted range of emotional expressions. This behavioral pattern begins in childhood and is usually clinically obvious by early adulthood, appearing in a variety of contexts. Although subclinical psychotic symptoms are not uncommon in persons with different cluster A personality disorders, individuals with a schizoid personality disorder do not show many unusual perceptual experiences, distorted thinking processes, or bizarre behavioral patterns. Although they appear self-absorbed, detached, and emotionally aloof, they have a normal capacity to recognize reality. On rare occasions, particularly in response to social or interpersonal stressful circumstances, individuals with this disorder may experience very brief psychotic episodes (typically lasting minutes to hours). The content of psychotic symptoms may vary but is usually limited to delusions of reference, paranoid or persecutory delusions, and brief visual and auditory hallucinations relevant to the social or interpersonal stressors.

Although a genetic association between schizoid personality disorder and functional psychoses has been suggested, the data to support this are contradictory. Some studies (e.g., the Roscommon Family Study, B-SNIP) have reported that schizoid personality disorder has significantly increased prevalence in the relatives of individuals with schizophrenia (1.2% to 3.4%), although this association is modest compared to schizotypal personality disorder. In contrast, several studies showed that the rate of schizoid personality disorder among biologic relatives of schizophrenic probands is not any higher than in the general population (reports vary between 0.3% to 0.7%). Similarly, reports of schizoid personality disorder cosegregated in families with affective psychoses are uncommon. One of the difficulties is that people with isolated schizoid personality disorder are not commonly seen in clinical settings, and the prevalence of this personality disorder in the general population is not clearly established. Schizoid personality disorder may appear as the premorbid precursor of schizophrenia; however, the proportion of patients who eventually convert to schizophrenia is unknown. Schizoid personality disorder most often co-occurs with schizotypal and paranoid personality disorders, as well as with avoidant personality disorders, which have also been reported to be associated with schizophrenia.

Borderline Personality Disorder. The crucial feature of borderline personality disorder is a pervasive pattern of instability of self-image, interpersonal relationships, and affective reactions, as well as marked impulsivity and self-damaging behavior. Borderline personality disorder is clinically perceived as a condition “on the border” of psychotic functioning; therefore, historically, in the psychoanalytic literature, it has been called “ambulatory schizophrenia” (a term introduced by Helene Deutsch) and “psychotic character disorder” (described by John Frosch). Individuals with borderline personality disorder are prone to transient psychotic episodes that occur most frequently during periods of stress, specifically in response to a real or imagined abandonment. In fact, psychotic-like symptoms are included as diagnostic criteria for borderline personality disorder (criterion 9: Transient, stress-related paranoid ideation or severe dissociative symptoms). Other broadly defined psychotic symptoms are also often reported, including ideas of reference, hypnagogic phenomena, and body-image distortions. More severe psychotic symptoms, including frank hallucinations and delusions, occur less frequently. Typically, the symptoms tend to be transient, lasting minutes to hours. However, one study by a Cornell group reported that 27% of the patients with borderline personality disorder had psychotic episodes lasting for weeks which were not attributed to comorbid affective or substance use disorders.

It is noteworthy that in some cases, psychotic symptoms in persons with borderline personality disorder may be of a factitious nature (e.g., Canadian studies report up to 13% of factitious psychotic symptoms in inpatients with borderline personality disorder). Thus a thorough psychiatric history, observation of the patient’s behavior, and objective information obtained from the family members are crucial. In some cases, the psychotic symptoms of borderline personality disorder respond well to external structuring and situational changes (e.g., the real or perceived return of subjectively valuable person or caregiver may result in the remission of the transient psychosis). However, in more severe cases, antipsychotic medications are warranted.

Borderline personality disorder shows a strong familial pattern. For example, the risk of acquiring borderline personality disorder for first-degree biologic relatives of those with the disorder is about five times greater than in the general population. There is also an increased familial association of borderline personality disorder with mood disorders, antisocial personality disorder, substance-related disorders, and eating disorders. Although considerable co-occurrence of borderline personality disorder with schizotypal personality disorder has been reported, as well as the occurrence of premorbid borderline personality features in persons with schizophrenia, the majority of studies do not support a familial link between borderline personality disorder and schizophrenia. The large-scale epidemiologic studies of schizophrenia-related personality disorders in relatives (e.g., the Roscommon Family Study) reported that borderline personality disorder was rare among first-degree relatives of schizophrenic probands but showed a modest clustering of cases in relatives of patients with psychotic and nonpsychotic affective disorders. Recent findings from a large multisite consortium in biomarkers of psychosis, B-SNIP, demonstrated that 1.3% of first-degree relatives of schizophrenia patients met rigorous DSM-IV-TR criteria for borderline personality disorder, whereas 4.1% and 3.2% of relatives of patients with schizoaffective disorder and psychotic bipolar I disorder, respectively, met such criteria. Family studies of borderline personality disorder have also indicated increased rates of mood disorders in the relatives of borderline probands compared with healthy controls. Recent studies have suggested that borderline personality disorder and bipolar disorder may exist on a spectrum and, perhaps, share partially overlapping pathogenesis and may be rooted in the same genetic etiology, similar to cluster A personality disorders and schizophrenia. However, the question of whether borderline personality disorder is related specifically to psychotic bipolar disorder and to psychosis, in general, remains open. Pilot studies suggest that persons who have borderline personality disorder with a history of psychotic-like symptoms show distinct eye-tracking abnormalities similar to those in patients with schizophrenia. This may suggest that borderline personality disorder with psychotic symptoms may fall into the psychosis spectrum disorders continuum and share some endophenotypic and genetic markers underlying psychosis. Psychotic-like symptoms and narrowly defined psychosis appears to be a promising target dimension for future research in borderline personality disorder.

In summary, clinical and research evidence suggests that psychosis is best conceived of as a clinical continuum with multiple subtypes and grades of symptoms, similar to hypertension or congestive heart failure. In fact, there appears to be no clear demarcation between frank, clinically significant psychosis and mild psychotic-like symptoms other than severity and duration. Large epidemiologic studies suggest that fleeting psychotic symptoms are common in the general population. The majority of individuals who have these psychotic experiences show no functional impairment and are considered to be free of any formal psychiatric diagnosis. These observations suggest that psychosis can manifest in mild forms long before its manifestation to a clinically significant level. Moreover, many of the individuals with such subthreshold symptoms will remain stable for years without further progression into a full psychosis phenotype. Only a modest proportion of them will convert to a psychotic illness in the future. These “at risk” individuals, while rarely evaluated in psychiatrists’ offices, represent a valuable population for studying proneness to psychosis as well as for exploring propsychotic risk factors and individual resilience against psychosis.

In more severe and persistent presentations, psychotic symptoms appear at the border of formal psychosis diagnosis and manifest in entities like schizophrenia prodrome and cluster A personality disorders, frequently seen in biologic relatives of probands with psychotic illnesses. Although cluster A personality disorders (in particular, schizotypy) are generally thought to be biologically related to schizophrenia, recent family and whole genome linkage studies have shown that schizotypal personality disorder may present a unique phenotype associated with an increased familial risk for psychosis in general, independent of the categorical diagnosis. Similarly, prodromal symptoms, which are classically conceptualized as a phase in the course of schizophrenia, appear to be not unique to schizophrenia psychosis. Instead, they signify a risk for a spectrum of functional psychoses but not the exact diagnosis that will emerge.

From the disease continuum perspective, the final diagnosis of a psychotic disorder is the end point of a longitudinal pathway from asymptomatic, to subthreshold psychotic experiences, to frank but infrequent psychotic symptoms, to the development of persistent psychosis. It is possible that this pathway is based on a dynamic matrix of genetic, gene expression, and epigenetic influences, which are in a complex interplay with environmental factors, with frank psychosis emerging when a critical genetic effects threshold is reached, accompanied by modifying or triggering epigenetic, individual biologic and/or environmental factors. Although subthreshold manifestations of psychosis are considered indicators of proneness to schizophrenia and other functional psychoses, they do not signify an inevitable conversion to psychosis. Whether or not an individual at risk will traverse this pathway may depend on the interaction between propsychotic factors (e.g., individual biologic vulnerability, obstetrical complications, maternal/in-utero immune stimulation, adolescent cannabis use, childhood trauma) and protective resilience factors (e.g., high premorbid functioning, stable social environment).

ACKNOWLEDGMENTS

We would like to thank Four Seasons Decorations, Dallas, TX, for their help with graphic design and illustrations for this chapter.

Further Readings

*Allardyce J, Gaebel W, Zielasek J, van Os J. Deconstructing psychosis conference February 2006: the validity of schizophrenia and alternative approaches to the classification of psychosis. *Schizophr Bull*. 2007;33:863–867.

Balaratnasingam S, Janca A. Normal personality, personality disorder and psychosis: current views and future perspectives. *Curr Opin Psychiatry*. 2015;28(1):30–34.

*Dikeos DG, Wickham H, McDonald C, et al. Distribution of symptom dimensions across Kraepelinian divisions. *Br J Psychiatry*. 2006;189:346–353.

Fanous AH, Neale MC, Gardner CO, et al. Significant correlation in linkage signals from genome-wide scans of schizophrenia and schizotypy. *Mol Psychiatry*. 2007;12:958–965.

Gershon ES, Pearlson GD, Keshavan MS, et al. Genetic analysis of deep phenotyping projects in common disorders. *Schizophr Res*. 2018;195:51–57.

Ivleva EI, Bidesi AS, Keshavan MS, et al. Using brain gray matter volume as an intermediate phenotype of psychosis: bipolar and schizophrenia network on intermediate phenotypes (B-SNIP). *Am J Psychiatry*. 2013;170(11):1285–1296.

Ivleva EI, Moates AF, Hamm JP, et al. Smooth pursuit eye tracking, prepulse inhibition and auditory paired stimuli processing endophenotypes across the schizophrenia – bipolar disorder psychosis dimension. *Schizophr Bull*. 2014;40(3):642–652.

Ivleva EI, Clementz BA, Dutcher AM, et al. Brain structure biomarkers in the psychosis biotypes: findings from the bipolar-schizophrenia network for intermediate phenotypes. *Biol Psychiatry*. 2017;82(1):26–39.

Ivleva EI, Turkozer BH, Sweeney JA. imaging-based subtyping for psychiatric syndromes. *Neuroimaging Clin N Am*. 2020;30(1):35–44.

Kato T. Whole genome/exome sequencing in mood and psychotic disorders. *Psychiatry Clin Neurosci*. 2015;69(2):65–76.

*Kendler KS, McGuire M, Gruenberg AM, O’Hare A, Spellman M, Walsh D. The Roscommon Family Study: III. Schizophrenia-related personality disorders in relatives. *Arch Gen Psychiatry*. 1993;50:781–788.

Kwapil TR, Barrantes-Vidal N. Schizotypy: looking back and moving forward. *Schizophr Bull*. 2015;41(Suppl 2):S366–S373.

Links PS, Steiner M, Mitton J. Characteristics of psychosis in borderline personality disorder. *Psychopathology*. 1989;22:188–193.

Massa N, Owens AV, Harmon W, et al. Relationship of prolonged acoustic startle latency to diagnosis and biotype in the bipolar-schizophrenia network on intermediate phenotypes (B-SNIP) cohort. *Schizophr Res*. 2020;216:357–366.

Narayanan B, Ethridge LE, O’Neil K, et al. Genetic sources of subcomponents of event-related potential in the dimension of psychosis analyzed from the B-SNIP study. *Am J Psychiatry*. 2015;172(5):466–478.

Owen MJ, Craddock N, Jablensky A. The genetic deconstruction of psychosis. *Schizophr Bull*. 2007;33:905–911.

Park N, Juo S, Cheng R, et al. Linkage analysis of psychosis in bipolar pedigrees suggests novel putative loci for bipolar disorder and shared susceptibility with schizophrenia. *Mol Psychiatry*. 2004;9:1091–1099.

Parker DA, Hamm JP, McDowell JE, et al. Auditory steady-state EEG response across the schizo-bipolar spectrum. *Schizophr Res*. 2019;209:218–226.

Pearlson GD, Folley BS. Endophenotypes, dimensions, risks: is psychosis analogous to common inherited medical illnesses? *Clin EEG Neurosci*. 2008;39(2):73–77.

Peralta V, Cuesta MJ. The relationship between syndromes of the psychotic illness and familial liability to schizophrenia and major mood disorders. *Schizophr Res*. 2007;91:200–209.

Reininghaus U, Böhnke JR, Chavez-Baldini UY, et al. Transdiagnostic dimensions of psychosis in the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *World Psychiatry*. 2019;18(1):67–76.

*Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421–427.

Tamminga CA, Ivleva EI, Keshavan MS, et al. Clinical phenotypes of psychosis in the Bipolar and Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *Am J Psychiatry*. 2013;170(11):1263–1274.

*Tamminga CA, Ivleva EI. Toward a dimensional understanding of psychosis and its treatment. Treatment of psychiatric disorders. In: Gabbard GO, ed. *Gabbard’s*

Treatments of Psychiatric Disorders. 5th ed. American Psychiatric Publishing; 2014.

Tamminga CA, Clementz BA, Pearlson G, et al. Biotyping in psychosis: using multiple computational approaches with one data set. *Neuropsychopharmacology*. 2021;46(1):143–155.

Thaker GK. Neurophysiological endophenotypes across bipolar and schizophrenia psychosis. *Schizophr Bull*. 2008;34(4):760–773.

Valles V, Van Os J, Guillamat R, et al. Increased morbid risk for schizophrenia in families of inpatients with bipolar illness. *Schizophr Res*. 2000;42:83–90.

van Os J, Linscott RJ, Myin-Germeys I, Delespaul P, Krabbendam L. A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychol Med*. 2009;39(2):179–195.

Webb CT, Levinson DF. Schizotypal and paranoid personality disorder in the relatives of patients with schizophrenia and affective disorders: a review. *Schizophr Res*. 1993;11:81–92.

Wiles NJ, Zammit S, Bebbington P, Singleton N, Meltzer H, Lewis G. Self-reported psychotic symptoms in the general population: results from the longitudinal study of the British National Psychiatric Morbidity Survey. *Br J Psychiatry*. 2006;188:519–526.

12.17 Other Psychotic Disorders

Ramin Mojtabai, M.D., Ph.D., M.P.H., Laura J. Fochtmann, M.D., M.B.I., and Evelyn J. Bromet, Ph.D.

This section describes disorders included in the section on other psychotic disorders in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) and its text revision (DSM-5-TR). It also describes the 11th edition of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-11) category of acute and transient psychotic disorder. The disorders discussed in this section occur less frequently and are less well understood than schizophrenia or mood disorders with psychotic features. Nevertheless, they can have profound consequences for affected individuals and are important to identify and treat.

ACUTE AND TRANSIENT PSYCHOTIC DISORDER AND BRIEF PSYCHOTIC DISORDER

History

Descriptions of psychotic disorders with an acute onset and transient course have a long history in psychiatry. Along with the classic syndromes of schizophrenia and affective psychoses, Emil Kraepelin (1856–1926) mentioned amentia, a psychotic disorder with a remitting course and favorable outcome, originally described by Theodor Meynert (1833–1898). Similar psychotic syndromes were described by other 19th- and 20th-century European psychiatrists under various terms such as “cycloid psychosis,” “bouffée délirante,” “reactive psychosis,” and others, as briefly described in Table 12.17–1. At the same time, investigators reported a high prevalence of acute psychotic episodes with a benign course in Africa, the Caribbean islands, and India.

The descriptions of these psychotic syndromes are rather similar, a remarkable observation since these descriptions grew out of different psychiatric traditions separated by cultural and linguistic boundaries. Although all portray a psychotic condition with acute onset and remitting course, the multiplicity of descriptions and labels has been a major impediment to better understanding and classifying these psychoses. In recent years, attempts have been made to surpass this barrier by defining a syndrome based on the common features of the traditional descriptions. In this context, an important step was the introduction of the acute and transient psychotic disorder in ICD-10 and its successor, the ICD-11. Many of these cases would be classified under the category of brief psychotic disorder in DSM-5-TR. Some would be categorized under schizophreniform and other schizophrenia spectrum disorders, as well as other psychotic disorders in the DSM-5-TR.

The historical roots of the DSM-5-TR diagnosis of brief psychotic disorder and the ICD-11 diagnosis of acute and transient psychotic disorder can be found in European psychiatry. The Scandinavian concept of reactive or psychogenic psychosis was especially influential in the formulation of the DSM-III brief reactive psychosis diagnosis, which was the forerunner of the DSM-5-TR brief psychotic disorder. Brief reactive psychosis was defined as a psychotic condition lasting less than 2 weeks that followed a significant psychosocial stressor, involved emotional turmoil, and was manifested by loosening of associations, delusions, hallucinations, or disorganized or catatonic behavior. This diagnosis was maintained in the DSM-III-R, but the allowable duration of symptoms was extended to 1 month. In DSM-IV and DSM-IV-TR, the diagnosis of brief reactive psychosis was replaced by brief psychotic disorder, eliminating an identifiable psychosocial stressor as a diagnostic criterion. The change was partly motivated by the observation that many cases of brief psychosis are not precipitated by a marked stressor and hence cannot be labeled as “reactive.” Furthermore, the DSM-III and DSM-III-R criterion of emotional turmoil was removed from the DSM-IV and DSM-IV-TR criteria for brief psychotic disorder. The aim of these revisions in later editions of the DSM was to broaden the scope of the diagnosis of brief psychotic disorder. Nonetheless, DSM-IV and DSM-IV-TR distinguished between brief psychotic disorder with and without marked stressors and indicated that the former group is equivalent to the brief reactive psychosis. There were few changes in the definition and characterization of brief psychotic disorder between the DSM-IV-TR and the DSM-5. The DSM-5 requires that one of the criteria be the presence of delusions, hallucinations, or disorganized speech. The DSM-5 also includes a “with catatonia” specifier and ratings of the current severity of symptoms on a scale ranging from not present to present and severe.

Comparative Nosology

Acute and transient psychotic disorder is defined by ICD-11 as a psychotic disorder characterized by acute onset of psychotic symptoms without a prodrome. The symptoms reach their maximum severity within 2 weeks and commonly last from a few days to 1 month but not more than 3 months. In addition to delusions, hallucinations, and disorganized thoughts, perplexity or confusion may also be present, as well as disturbances of mood. Furthermore, in some cases, catatonia-like psychomotor disturbances may be present. The symptoms and their severity typically change rapidly from day to day or even within a day. Brief psychotic disorder is defined by DSM-5-TR as a psychotic condition that involves the onset of psychotic symptoms that last 1 day or more but less than 1 month. Remission is full, and patients return to their premorbid level of functioning. Thus, most individuals diagnosed with brief psychotic disorder under DSM-5-TR would be classified as having acute and transient psychotic disorder under ICD-11.

There is no specific section for acute and transient psychotic disorder in the DSM-5-TR. As a result, patients who meet these ICD-11 criteria are classified under a number of different diagnostic categories in the DSM-5-TR. The large majority would meet the

DSM-5-TR criteria for brief psychotic disorder, schizophreniform disorder or other specified schizophrenia spectrum, and other psychotic disorder, although the actual proportion receiving each diagnosis varies considerably across studies. In a study from Germany, 62% of patients with ICD-10 acute and transient psychotic disorder also met the DSM-IV criteria for brief psychotic disorder, 31% for schizophreniform disorder, and 5% for psychotic disorder not otherwise specified (NOS).

By including a specifier for “with peripartum onset,” DSM-5-TR allows many cases of postpartum psychosis to be categorized under the brief psychotic disorder category. ICD-11, on the other hand, categorizes cases that start within about 6 weeks after delivery under the category of “mental or behavioral disorders associated with pregnancy, childbirth or the puerperium, with psychotic symptoms.” ICD-11 also requires that a diagnosis of a specific mental disorder be additionally assigned if the criteria for that disorder are met.

Table 12.17–1.

Partial Listing of the Classic Descriptions of Acute and Transient Psychotic Disorders

Diagnosis	Publications	Description
Amentia	Meynert T. Amentia, die Verwirrtheit. *Jahrb Psychiatr Neurol*. 1889;9:1.	A psychotic illness with acute onset characterized by confusion and perplexity, agitation, rapidly changing vivid hallucinations and delusions, misidentification phenomena, anxiety and apprehension, first introduced by Theodore Meynert (1833–1898). An association with physical illness and exhaustion was noted in some patients. Full recovery occurs in a few weeks or months. Although somewhat influential at its time, amentia never produced much interest in later years.
Cycloid psychosis	Kleist K. über zykloide, paranoide und epileptoide Psychosen und über die Frage der Degenerations-	A psychotic disorder with acute onset and good prognosis but frequent recurrences, characterized by

psychosen. *Schweizer Arch Neurol Psychiat*. 1928;23:1.

Leonhard K. Cycloid psychoses-endogenous psychoses which are neither schizophrenic nor manic depressive. *J Ment Sci*. 1961;107:632.

Perris C, Brockington IF. Cycloid psychoses and their relation to the major psychoses. In: Perris C, Struwe G, Jansson B, eds. *Biological Psychiatry*. Elsevier; 1981:447.

confusion, mood incongruent delusions, hallucinations, overwhelming anxiety, deep feelings of happiness or ecstasy, motility disturbances of akinetic or hyperkinetic type, a particular concern with death, mood swings, rapid change in symptoms within an episode. Two variants of the disorder were first described by Karl Kleist (1879–1960): confusional insanity characterized by contrasting phases of confused excitement and stupor and motility psychosis characterized by contrasting phases of hyperkinesis and akineses. A third variant, anxiety-elation psychosis was introduced by

		Karl Leonhard (1904–1988). The diagnosis of cycloid psychosis is still used by German, Scandinavian, and other European psychiatrists and was influential for the formulation of acute and transient psychotic disorders in ICD-10. Carlo Perris and Ian Brockington introduced a set of explicit criteria for the diagnosis of cycloid psychosis in 1981.
Bouffée délirante	Magnan V, Legrain M. *Les Dégénérés* (*Etat Mental et Syndromes Episodiques*). Rueff et Cie; 1895. Pull CB, Pull MC, Pichot P. Nosological position of schizoaffective psychoses in France. *Psychiatr Clin*. 1983;16:141.	A psychotic disorder with acute onset in a person without previous psychiatric history. The episode remits completely with no residual symptoms. The episodes are characterized by delusions, hallucinations, depersonalization and derealization, confusion, mood

change, and changing symptoms during the course of episode. Episodes are not due to organic or substance use. The syndrome was introduced by Valentin Magnan (1835–1916) and Paul-Maurice Legrain (1860–1939) in 1895 and was said to be caused by degeneration, a 19th-century view that attributed mental illness to the spread of modern civilization and urban life. Although this etiologic formulation is no longer accepted, the diagnosis is still used by French-speaking clinicians in Europe, West Africa, and the Caribbean. The concept of bouffée délirante was influential in formulating the ICD-

		10 acute and transient psychotic disorders. Due to its common occurrence in Africa and the Caribbean, it is also categorized as a culture-bound syndrome in the DSM-IV-TR. Charles Pull and colleagues proposed a set of explicit diagnostic criteria for bouffée délirante in 1983.
Psychogenic or reactive psychosis	Wimmer A. *Psykogene Sindssygdomsformer*. St. Hans Hospitals Jubilæumsskrift; 1916.	A psychotic disorder with acute onset following external stress. Compared to schizophrenia, the onset is more likely to be acute and to be later in life. Premorbid adjustment tends to be better than in schizophrenia. There are more affective and confusional symptoms and less bizarre symptoms, and there is less family history of

		schizophrenia. Prognosis is better than schizophrenia. Diagnoses of psychogenic or reactive psychoses were popular among Scandinavian psychiatrists in the early part of the 20th century and continued to be used throughout the past century but found limited use outside of these countries.
Schizophreniform psychosis or disorder	Langfeldt G. *The Schizophreniform States.* Ejnar Munksgaard; 1939. American Psychiatric Association. *Diagnostic and Statistical Manual*. 3rd ed. American Psychiatric Press; 1980.	The Norwegian psychiatrist Gabriel Langfeldt (1937–1966) introduced the concept of schizophreniform psychosis as a condition with a good outcome that had a sudden onset following an identifiable precipitating factor in an individual with well-adjusted premorbid personality. Patients

		often present with disturbance of mood and clouding of consciousness. The term, but not the concept, was adopted by the DSM-III as a nonaffective psychotic syndrome with schizophrenic symptoms, which is distinguished from schizophrenia by a duration of less than 6 months. The concept of schizophreniform disorder in the latter sense was perpetuated in the later editions of the DSM.
Oneirophrenia	Meduna LJ. *Oneirophrenia: The Confusional State*. University of Illinois Press; 1950.	Ladislas von Meduna (1896–1964) described oneirophrenia in 1939 as a syndrome characterized by acute onset of confusion, nightmare, or dream-like quality of all perceptions (hence the term

		“oneirophrenia”), extreme fear and anxiety, delusions, and visual hallucinations. The prognosis is generally good with full recovery. Meduna proposed an endocrinologic explanation for the syndrome.
Hysterical psychosis	Hollander MH, Hirsch SJ. Hysterical psychosis. *Am J Psychiatry*. 1964;120:1066.	Although the term hysterical psychoses had been in use since the early 20th century, Marc Hollander and Steven Hirsch presented the first formal description of the syndrome in 1964. These authors described a psychotic episode with sudden and dramatic onset related to a profoundly upsetting event in the context of a “hysterical” personality. Symptoms include hallucinations,

delusions, depersonalization, and disorganized behavior. The episode seldom lasts longer than 1–3 weeks.

Epidemiology

Acute and transient psychotic disorder is rare, especially in industrialized countries. As a result, data on its incidence or prevalence are limited. Much of the current knowledge of this disorder has come from clinical samples, but the brief duration of this condition means that many individuals with acute and transient psychotic disorder will never come to the attention of mental health services. Additional information about the epidemiology of this condition derives from international studies in developing and industrialized countries in the 1980s and 1990s. For example, a first-episode psychosis study from Nottingham, England, estimated the incidence rate of cases that remained diagnostically stable for 3 years to be 1.36 per 100,000 (0.74 per 100,000 for men and 1.99 for women). Another epidemiologic study reported a substantially higher incidence in developing countries compared with industrialized countries. These psychoses comprised 35% of all first-contact nonaffective psychoses in developing countries compared to only 6% in industrialized countries. The age of onset was also younger (the mid-20s) in developing than industrialized (mid-20s to 30s) countries.

Brief psychotic disorder is perhaps even less common than acute and transient psychotic disorder, although, as noted, the two disorders overlap considerably. In a sample of over 1,000 patients with nonorganic psychotic or major affective disorder hospitalized in Germany, individuals with brief psychotic disorder comprised only 2.5% of the cases. Similarly, in the Suffolk County, New York, first-admission cohort, a primary diagnosis of brief psychotic disorder was found in only 2% of the cohort sample. The incidence may be higher in first-onset samples compared to first-admission samples because some of these cases may remit before clinical attention is sought. A study of U.S. Air Force recruits during basic training reported a 6-week incidence rate of 1.43 cases of DSM-III-R brief reactive psychosis per 100,000 recruits. However, this number may be an overestimate of general population rates because basic training is a particularly stressful period.

Similar to acute and transient psychotic disorders, brief psychotic disorder is more common among women than men. Also similar to acute and transient psychotic disorders, the age of onset of brief psychotic disorder in industrialized countries appears to be older than in developing countries.

Etiology

Little is known about the etiology of acute and transient psychotic disorders, and even less is known about brief psychotic disorder. The available evidence for acute and transient psychotic disorders points to both biologic and sociocultural factors, including stress, as underlying risk factors. In one study from India, for example, a history of antecedent fever was reported in 47% of individuals with an acute and transient psychotic disorder. In almost all cases, the fever was of high grade, requiring medical attention or hospitalization. In all patients, the fever had resolved before the onset of psychotic symptoms, making it unlikely that the psychosis was the direct consequence of an acute infection. Similar reports from other low-income country settings and classic European psychiatric literature also note high rates of febrile illness prior to the development of acute and transient psychotic disorders, which may, in part, explain the higher incidence of these disorders in low-income countries where infectious diseases are more common than in high-income countries. A more recent study from India identified elevated levels of proinflammatory markers in a group of patients with acute and transient psychotic disorder compared with healthy controls.

Information on the biochemical, physiologic, and anatomical correlates of these disorders is limited, and the available data do not reveal any distinctive abnormalities that would suggest a specific etiology. A recent MRI study from Hungary identified reductions in the volume of the caudal/rostral middle frontal cortex and superior frontal cortex in patients with brief psychotic disorder compared to controls. These abnormalities were more pronounced in patients with multiple lifetime episodes. Researchers from Japan have noted hypothalamic–pituitary axis abnormalities in subsets of female patients with recurrent atypical psychoses who are more likely to experience episodes at the time of menstruation and parturition; however, the relevance of these findings to other patients with acute and transient psychotic disorders is unclear. Most available evidence on biologic correlates of these conditions is based on work involving patients with cycloid psychosis, which overlaps with ICD-11 acute transient psychotic disorder and DSM-5-TR brief psychotic disorder. Auditory evoked potential studies in patients with cycloid psychosis recorded P300 peaks over the left hemisphere, similar to normal controls, but patients with cycloid psychosis had a higher P300 amplitude than normal controls, suggesting a higher level of arousal. This finding is consistent with a regional cerebral blood flow study of cycloid psychosis that recorded increased hemispheric blood flow during the psychotic episode that was directly related to symptom severity and that returned to a normal level after remission.

No genetic studies and few family studies of ICD-11 acute and transient psychotic disorders or DSM-5-TR brief psychotic disorder have been conducted. A study from India reported a higher risk of acute transient psychotic disorder and a lower risk of schizophrenia and mood disorders in the first-degree relatives of probands compared with schizophrenia probands. Similar results were obtained in family studies of cycloid psychosis and reactive psychosis in other settings.

The high incidence rate of acute and transient psychotic disorder in low-income countries has also led to speculations about the role of sociocultural factors in the etiology of these syndromes. It has been suggested that rapid cultural change and modernization in low-income countries expose individuals to loss of status and stress arising from role confusion, making individuals vulnerable to psychotic reactions.

Acute stress (within about 2 weeks of onset of psychosis) was noted by ICD-10 as a feature of illness in some individuals with acute and transient psychotic disorder. However, the actual proportion of patients who experience acute stress in this time frame varies considerably across studies, ranging from 10% to 69%. Assessment of acute stress is also complicated by lack of a clear definition of such stress and the possible impact of recall bias. Furthermore, the stressors likely vary according to social and cultural context. For example, in one study from India, departure from or return to their parental village was a major stressor among women, whereas job-related problems were the main stressor among men. Association with social stressors is not limited to low-income countries and tentative links between the stress-associated social and economic consequences of COVID-19 epidemic and incident cases of brief psychotic disorder have been noted in a case series from Italy.

The specifiers “with” or “without marked stressor(s)” are noted in the DSM-5-TR criteria of brief psychotic disorder. DSM-5-TR also notes that these events “would be markedly stressful to almost anyone in similar circumstances in the individual’s culture.” However, there are few reliable data on the proportion of cases of brief psychotic disorder that are preceded by such stressors.

Clinical lore suggests that pre-existing personality disorders and psychoticism traits are associated with acute and transient psychotic disorder as well as brief psychotic disorder. However, reliable assessment of premorbid personality disorders in patients with psychotic conditions is often difficult and may be affected by recall bias or residual psychotic symptoms. This point was highlighted in a Danish study of acute and transient psychotic disorder in which the majority of patients were diagnosed with a pre-existing personality disorder shortly after recovery from psychosis, whereas a year later, only a minority received such a diagnosis. Similarly, a study from Germany did not find significant differences with regard to premorbid adjustment or the “big five” personality dimensions (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness) measured by self-report between patients who had recovered from an acute transient psychotic episode and normal controls.

A psychodynamic formulation of brief reactive psychosis links this condition to deficient ego strength. According to this formulation, psychosis occurs in response to an intense emotionally stressful event that, in the absence of support in the environment, overwhelms the individual's ego defense mechanisms.

Diagnosis and Clinical Features

Acute and Transient Psychotic Disorder. Recognition of acute and transient psychotic disorder early in its course is difficult. In the first few weeks of the illness, there are often few specific indicators that would distinguish acute and transient psychotic disorder from other nonaffective psychotic disorders with acute onset. Furthermore, even in retrospect, it is often difficult to assess the duration of a psychotic episode or distinguish spontaneous remission from remission due to antipsychotic medication treatment. Therefore, a definitive diagnosis of acute and transient psychotic disorder is often only possible retrospectively and when extensive information on the patient's history is available.

Some of the features of acute and transient psychotic disorder in ICD-11 are helpful in differential diagnosis: (a) acute onset of psychotic symptoms that emerge without a prodrome and reach their maximal severity within 2 weeks; (b) presence of perplexity or confusion; and (c) rapid change in the symptoms and their intensity, from day to day, or even within a day (sometimes referred to as "polymorphic symptomatology").

A diagnosis of acute and transient psychotic disorder requires ruling out another medical condition (e.g., a brain tumor) and the effect of a substance or medication on the central nervous system (CNS) (e.g., corticosteroids) or withdrawal from substances (e.g., alcohol withdrawal).

DSM-5-TR Brief Psychotic Disorder. The diagnosis of brief psychotic disorder in DSM-5-TR is based on the presence of one or more of the following psychotic symptoms for a period of at least 1 day but less than 1 month: (1) delusions, (2) hallucinations, (3) thought disorder as evidenced by disorganized speech, and (4) disorganized or catatonic behavior. DSM-5-TR specifies that one of the first three symptoms must be present for the diagnosis to be made. DSM-5-TR criteria for brief psychotic disorder allow for specifiers indicating the presence or absence of a marked stressor or stressors. These are defined as events that would be “markedly stressful to almost anyone in similar circumstances in the individual’s culture.” The manual does not specify stressors that would qualify as marked stressors, but examples reported in the literature include serious surgery or medical illness, immigration, war and other mass violence, torture, and even intensive training programs such as military training.

Although in some cases the temporal order of stressor and psychotic episode is clear, it is not always easy to determine whether a specific stressor was a precipitant or consequence of the illness. Other sources of information, such as information from spouse or friends about the level of functioning before the onset of episode, as well as history of response to similar stressors in the past, would be helpful in determining the order of events. The DSM-5-TR criteria for brief psychotic disorder also allow for a “with peripartum onset” specifier for episodes starting during pregnancy or within 4 weeks of giving birth. Comorbid catatonia can be identified by a “with catatonia” specifier and the additional diagnostic codes 293.89 (F06.1). Similar to other psychotic disorders in the DSM-5-TR, the current severity of psychotic symptoms can be optionally specified using a 5-point rating scale.

“Emotional turmoil” and “confusion” are not among the diagnostic criteria for brief psychotic disorder in DSM-5-TR. However, they are noted as typical experiences in the text of the manual. The text also notes rapid shifts from one intense affect to another in some individuals. A study of brief reactive psychosis from Swaziland, South Africa, also notes that the clinical presentation in that setting is characterized more by volatile affect and behavior, altered states of consciousness, and dissociative experiences than by clear-cut psychotic symptoms.

Differential Diagnosis

A definitive diagnosis of acute and transient psychotic disorder or brief psychotic disorder early in the first episode of illness is difficult if not impossible. Unless the psychosis has fully remitted, the duration of an episode of illness cannot be determined. Therefore, during the first few days or weeks of an episode, the diagnosis is usually provisional.

Many aspects of the differential diagnosis of ICD-11 acute and transient psychotic disorders relate to symptom duration. A provisional diagnosis of acute psychotic disorder would be changed to ICD-11 schizophrenia or other psychotic disorders if the illness lasts more than 3 months.

Similar considerations apply to DSM-5-TR brief psychotic disorder, in which the duration of an episode is defined as less than 1 month. If the duration of illness (from prodrome through recovery) exceeds 1 month, the diagnosis would need to change depending on other symptoms and the duration of the illness. A remission of psychotic symptoms within 1 month as a result of successful treatment also makes it difficult to distinguish brief psychotic disorder from other longer-lasting disorders. In such instances, the presence of continued residual symptoms or impairment may be helpful in ruling out a diagnosis of brief psychotic disorder. In many cases, however, only the long-term course of illness can resolve diagnostic uncertainties.

Perhaps the most critical differential diagnostic issue for both acute and transient psychotic disorder and brief psychotic disorder is to rule out delirium, psychotic disorder due to another medical condition, substance/medication-induced psychotic disorder, and drug or alcohol intoxication or withdrawal. Careful physical and laboratory examination is often indicated. Toxicologic screens and imaging studies may also be necessary. A positive drug or alcohol screen does not necessarily rule out a diagnosis of acute and transient psychotic disorder or brief psychotic disorder. Some patients with acute and transient psychotic disorder or brief psychotic disorder may use substances in an attempt to ameliorate their symptoms. The temporal relationship between onset of psychosis and the pattern of substance abuse may be helpful in determining if a psychotic disorder is substance or medication induced. Symptoms that persist for many days after all traces of the substance are eliminated from blood and urine would support a diagnosis of acute and transient psychotic disorder or brief psychotic disorder.

Both ICD-11 acute and transient psychotic disorder and DSM-5-TR brief psychotic disorder may be associated with confusion and perplexity, making it difficult to distinguish these disorders from delirium or medical causes of psychosis. Diurnal variation in consciousness, which is a characteristic of delirium, may also be confused with variable “polymorphic” psychotic symptoms seen in acute and transient psychotic disorder. However, compared with delirium, changes in consciousness in acute and transient psychotic disorder are generally less severe and less persistent. Furthermore, a medical cause of delirium can often be identified through history, physical examination, laboratory tests, or imaging studies.

Acute psychotic episodes have been reported in a number of medical conditions, including head trauma, cerebral anoxia, epilepsy, and endocrine disorders such as hyper- or hypothyroidism. Many of these episodes are associated with disturbances of consciousness. Differential diagnosis is often based on a thorough medical history and physical examination as well as laboratory investigations.

Good prognosis of brief psychotic disorder and of acute and transient psychotic disorder, variability of symptoms, and the presence of mood symptoms sometimes make it difficult to differentiate these diagnoses from mood disorders. The presence of a full-mood syndrome that meets the criteria for a manic or a depressive episode excludes a diagnosis of brief psychotic disorder.

Patients with personality disorders, especially borderline personality disorder, sometimes experience psychotic symptoms under stressful life circumstances. These symptoms are generally transient and do not require a separate diagnosis. If the symptoms last more than 1 day, an additional diagnosis of brief psychotic disorder may be warranted.

Finally, a number of "culture-bound" syndromes with acute onset of psychotic and dissociative symptoms have been described in various cultural settings. Many of these syndromes are transient and resolve spontaneously without antipsychotic medication treatment. The nosologic relationship of these conditions to acute and transient psychotic disorders or brief psychotic disorder remains unresolved.

Course and Prognosis

Acute and transient psychotic disorder has a favorable early course by definition. Full remission within 3 months is a diagnostic criterion for this disorder in the ICD-11. Although there are few first-admission studies that allow for reliable assessment of course and outcome, the available evidence indicates that most of the recovered patients remain well in ensuing years, and even among those who experience further episodes, these episodes are transient, and the outcome is generally favorable. ICD-11 includes specifiers for "first episode" and "multiple episodes." Recurrence rates of 39% to 47% have been reported in first-admission samples followed for 3 to 4 years, but full recovery between episodes is common. Only about 6% to 18% of patients have any residual symptoms at a later point in time, although some studies have shown that the duration of nonaffective psychoses with acute onset and remitting course may extend up to 6 months. A 12-year follow-up study from India also reported favorable long-term outcomes. There is a paucity of data on the long-term course and outcome of acute and transient psychotic disorders in industrialized countries. In a prospective study from Latvia, 40% of a group of first-admission patients with a diagnosis of acute and transient psychotic disorder were readmitted in a 26-month follow-up.

A number of studies from various countries have assessed the diagnostic stability of the ICD-10 acute and transient psychotic disorders. In studies from India, Denmark, Spain, and Latvia, between 27% and 48% of patients had a different diagnosis 1 to 3 years after their index episodes. Some studies have found that diagnostic stability was associated with the patients' clinical features. Polymorphic symptoms are associated with greater stability of diagnosis; whereas, presence of schizophrenia symptoms was associated with change of diagnosis at follow-up. The diagnostic stability of an acute and transient psychotic disorder may also vary according to the patient's gender. In a British study, the diagnoses of acute and transient psychotic disorder remained unchanged in 73% of women but only 14% of men at 3-year follow-up. Most changes were in the schizophrenia or mood disorder categories.

Brief psychotic disorder also has a benign early course by definition. Full remission within 1 month is a DSM-5-TR criterion for this diagnosis. Some individuals, however, experience relapses. In a study of consecutive admissions for psychosis in Germany, 83% of the patients initially diagnosed with brief psychotic disorder maintained this diagnosis when re-evaluated 2 years later, and 70% experienced relapses during the 2-year follow-up. Because changes in diagnosis are more likely to occur in the early course of psychotic conditions and the German study used cross-sectional diagnosis at the time of follow-up, the rates of diagnostic change and of relapse would likely be higher than in a first-admission or first-episode sample. Nevertheless, the social functioning of these individuals was much better than individuals with schizophrenia; over 90% were living independently, and over half were employed.

Several other studies have also examined the stability of the brief psychotic disorder diagnosis. In a prospective epidemiologic study of a first-admission cohort from Suffolk County, New York, only 27% of those receiving a consensus research diagnosis of brief psychotic disorder at 6-month follow-up maintained this diagnosis when re-evaluated at the 24-month follow-up, with 9% rediagnosed with schizophreniform disorder and 18% with psychotic disorder NOS, a residual diagnostic category in DSM-IV. The remaining individuals were rediagnosed with mood disorders (27%), schizophrenia (9%), or other conditions (9%). A Brazilian study of first-episode patients admitted to an emergency psychiatric unit also found that only about

23% of patients with a clinical discharge diagnosis of brief psychotic disorder maintained this diagnosis 1 year later. These findings suggest that assessment of the prognosis of brief psychotic disorder is confounded by the diagnostic instability of the disorder.

SCHIZOPHRENIFORM DISORDER

History

The concept of schizophreniform disorder was introduced in 1939 by the Norwegian psychiatrist Gabriel Langfeldt. He divided a group of patients who had initial symptom profiles consistent with schizophrenia into two groups based on their subsequent course. One group had a poor course and was referred to as "typical schizophrenia," whereas the "schizophreniform" group had a sudden onset, associated mood symptoms, and clouding of consciousness with a more benign course.

Forty years later, schizophreniform disorder was officially included in the DSM-III in the section on psychotic disorders not elsewhere classified. However, the term was used to describe patients with a different set of characteristics from those identified by Langfeldt. In fact, when the charts of 55 of Langfeldt's original patients diagnosed with schizophreniform psychosis were re-evaluated, most patients met the DSM-III-R criteria for a mood disorder. In contrast, DSM-III described a condition that was identical to schizophrenia except with regard to its duration, with the prodromal, active, and residual phases of schizophreniform disorder lasting more than 2 weeks but less than 6 months.

DSM-III-R added the requirement that a patient given a diagnosis of schizophreniform disorder should not meet the criteria for brief reactive psychosis (i.e., the duration of symptoms should not be less than a month). In addition, the presence or absence of a series of good prognostic features was added as a specifier in order to more closely approximate Langfeldt's original conception of the disorder. DSM-III-R also dropped the criterion B of schizophrenia (impairment in functioning) from the diagnostic criteria of schizophreniform disorder and noted that the diagnosis should be made provisionally if recovery had not yet occurred. Additional changes in the diagnostic criteria for schizophreniform disorder reflect changes to the underlying diagnostic criteria for schizophrenia between DSM-III and DSM-III-R (i.e., differences in the specific psychotic symptoms needed to establish a diagnosis, removal of the requirement for an age of onset of fewer than 45 years). In DSM-IV and DSM-IV-TR, the diagnostic criteria for schizophreniform disorder were maintained with little change. DSM-5 introduced a few changes consistent with changes in the A criterion for schizophrenia and specifiers added to virtually all psychotic disorder diagnoses. In addition, in keeping with the other psychotic disorder diagnoses, the DSM-5 diagnosis for schizophreniform disorder allows for a "with catatonia" specifier and ratings of current severity of symptoms.

The A criterion for schizophrenia in DSM-5 removed the higher weight given in prior editions of DSM to bizarre delusions, auditory hallucinations keeping up a running commentary on the person's behavior or thoughts, and auditory hallucinations of two or more voices conversing with each other. In DSM-III-R, DSM-IV, and DSM-IV-TR, presence of one of these symptoms was enough to meet the A criterion requirement; whereas, at least two symptoms were required for diagnosis if none of these highly weighted symptoms were present. In contrast, the DSM-5 gives equal weight to symptoms included under the A criterion. However, it requires that one of the symptoms be a delusion, hallucination, or disorganized speech.

Comparative Nosology

The nosology of schizophreniform disorder and, in particular, its relationship to the schizophrenia diagnosis has changed little in successive editions of the DSM. Similar to previous editions of the DSM, in DSM-5-TR, schizophreniform disorder is conceptualized as a variant of schizophrenia (Table 12.17–2). Thus, the first criterion specifically requires that a patient meet the A criterion for schizophrenia. The criteria further specify the minimum and maximum duration at least 1 month but less than 6 months, that the patient does not meet criteria for schizoaffective or mood disorder with psychotic features, and that the disorder is not due to the physiologic effects of a substance or another medical condition. With a similar constellation of symptoms but a duration of less than 1 month, a diagnosis of brief psychotic disorder would be made. If the duration of illness extended beyond 6 months and there is associated functional impairment, a diagnosis of schizophrenia would be warranted.

A comparable category of schizophreniform disorder does not exist in ICD-11. However, the duration of schizophrenia in ICD-11 is 1 month or longer. As such, many cases of DSM-5 schizophreniform disorder will be classified as schizophrenia in ICD-11.

Table 12.17–2.
DSM-5 Diagnostic Criteria for Schizophreniform Disorder

Diagnostic Criteria

295.40 (F20.81)

A. Two (or more) of the following, each present for a significant portion of time during a 1-month period (or less if successfully treated). At least one of these must be (1), (2), or (3):
   1. Delusions.
   2. Hallucinations.
   3. Disorganized speech (e.g., frequent derailment or incoherence).
   4. Grossly disorganized or catatonic behavior.
   5. Negative symptoms (i.e., diminished emotional expression or avolition).

B. An episode of the disorder lasts at least 1 month but less than 6 months. When the diagnosis must be made without waiting for recovery, it should be qualified as "provisional."

C. Schizoaffective disorder and depressive or bipolar disorder with psychotic features have been ruled out because either (1) no major depressive or manic episodes have occurred concurrently with the active-phase symptoms, or (2) if mood episodes have occurred during active-phase symptoms, they have been present for a minority of the total duration of the active and residual periods of the illness.

D. The disturbance is not attributable to the physiological effects of a substance (e.g., a drug of abuse, and a medication) or another medical condition.

Specify if:

With good prognostic features: This specifier requires the presence of at least two of the following features: onset of prominent psychotic symptoms within 4 weeks of the first noticeable change in usual behavior or functioning, confusion or perplexity, good premorbid social and occupational functioning, and absence of blunted or flat affect.

Without good prognostic features: This specifier is applied if two or more of the above features have not been present.

Specify if:

With catatonia (refer to the criteria for catatonia associated with another mental disorder, pp. 119–120, for definition).

Coding note: Use additional code 293.89 (F06.1) catatonia associated with schizophreniform disorder to indicate the presence of the comorbid catatonia.

Specify current severity:

Severity is rated by a quantitative assessment of the primary symptoms of psychosis, including delusions, hallucinations, disorganized speech, abnormal psychomotor behavior, and negative symptoms. Each of these symptoms may be rated for its current severity (most severe in the last 7 days) on a 5-point scale ranging from 0 (not present) to 4 (present and severe). (See Clinician-Rated Dimensions of Psychosis Symptom Severity in the chapter “Assessment Measures.”)

Note: Diagnosis of schizophreniform disorder can be made without using this severity specifier.

Reprinted with permission American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:111, 112.

Epidemiology

Data on the prevalence of schizophreniform disorder are limited. One study of DSM-IV schizophreniform disorder in the Finnish general population found a lifetime prevalence of 0.07% overall with a fivefold greater rate in men than in women. Somewhat higher estimates were noted in a systematic review of the literature, which found a 1-year prevalence of 0.09% and a lifetime prevalence of 0.11%, although a great deal of heterogeneity in the data was observed. In one study from Suffolk County, New York, approximately 5% of patients with a first-admission psychotic disorder met the diagnostic criteria of schizophreniform disorder on admission.

Etiology

The etiology of schizophreniform disorder remains unclear. Studies of risk factors for psychotic illness have typically combined results from individuals with schizophreniform disorder with those from individuals with schizophrenia, making it difficult to delineate risk factors for the development of schizophreniform disorder separate from schizophrenia. The heterogeneity in the course and outcome of schizophreniform disorder further complicates etiologic research. As currently conceptualized, this diagnostic category includes both individuals with schizophrenia in the early course of illness who are given the diagnosis of schizophreniform disorder provisionally and whose diagnosis would switch to schizophrenia in follow-up, as well as individuals with a remitting psychotic illness who would not meet the criteria for schizophrenia in future.

Nevertheless, available research supports the role of genetic factors in the etiology of schizophreniform disorder. One meta-analysis found that rates of schizophrenia and of mood disorder in relatives of schizophreniform probands did not differ from those rates in the relatives of probands with schizophrenia or mood disorders, respectively. A subsequent study showed an excess risk of schizophrenia spectrum illness in relatives of schizophreniform probands but no increase in the risk of mood disorder in family members. However, a third study of patients with schizophreniform disorders with “good prognostic features” showed an increased likelihood of mood disorder in first-degree relatives of probands. Despite this variability in the findings of family studies, there does seem to be an increase in the rates of other psychiatric illnesses among family members of those with schizophreniform disorder, suggesting that genetic factors may be playing a role in some patients.

Studies using brain imaging to look for abnormalities in individuals with schizophreniform disorder have had mixed results. One large study that examined hippocampal and amygdala volumes found no difference between patients with schizophreniform disorder and controls. Studies of ventricular enlargement in patients with schizophreniform disorder have produced mixed results. In one study, left ventricular enlargement was correlated with a progression of illness to schizophrenia or schizoaffective disorder. Although two studies showed no progressive increases in ventricular size over 2 to 7 years of follow-up for the schizophreniform patient group as a whole, outcomes were poorer in patients with initially enlarged ventricles than in those without ventricular enlargement. A more recent study found that hippocampal volume deficits were related to the prognosis of illness. These deficits were apparent at baseline only among patients whose illness persisted and whose diagnosis was changed to schizophrenia at 2-year follow-up, but not among those whose illness remitted and maintained a diagnosis of schizophreniform disorder.

Diagnostic and Clinical Features

The mode of onset of schizophreniform disorder is similar to that of schizophrenia. Although many patients with schizophreniform disorder may experience functional impairment during an episode, they are unlikely to report a progressive decline in social and occupational functioning. The initial symptom profile is the same as that of schizophrenia in that two or more psychotic symptoms (hallucinations, delusions, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms) must be present. There is an increased likelihood of emotional turmoil and confusion or perplexity, the presence of which may indicate a good prognosis. Although negative symptoms may be present, they are relatively uncommon in schizophreniform disorder and are considered poor prognostic features. By definition, patients with schizophreniform disorder return to their baseline state within 6 months. Otherwise, the provisional diagnosis of schizophreniform disorder needs to be changed to schizophrenia. In some individuals, schizophreniform disorder is episodic with more than one episode occurring after long periods of full remission.

Differential Diagnosis

It is important to first differentiate schizophreniform disorder from psychoses that can arise from medical conditions or use of substances. This is accomplished by taking a detailed history and physical examination, and where indicated, performing laboratory tests or imaging studies. Although it is not always possible to distinguish substance-induced psychosis from other psychotic disorders cross-sectionally, a rapid onset of psychotic symptoms in a patient with a significant substance history should raise the suspicion of a substance/medication-induced psychotic disorder. Because individuals with schizophreniform disorder also may have an acute onset of psychotic symptoms, longitudinal assessment is often crucial in distinguishing a substance/medication-induced psychotic disorder from schizophreniform disorder with incidental substance use. A detailed history of medication use, including over-the-counter medications and herbal products, is essential because such therapies may produce an acute psychosis. A substance-use history and toxicologic screen are also important for identifying substance-related psychosis and developing an appropriate treatment plan in an individual with a new onset of psychosis.

The duration of psychotic symptoms is one of the features that distinguishes schizophreniform disorder from other syndromes. Schizophrenia is diagnosed if the duration of illness lasts more than 6 months, whereas symptoms that occur for less than 1 month indicate a brief psychotic disorder. One of the difficulties in making a diagnosis of schizophreniform disorder involves dating the onset of a psychotic episode that evolves insidiously. Without a distinct symptom onset, it becomes difficult to decide whether the illness has actually lasted less than 6 months. In the event that the diagnosis is made before 6 months have elapsed, the diagnosis of schizophreniform disorder is considered provisional. For some patients, prolonged observation will be needed before the duration of symptomatology can be established with certainty, particularly if the patient has received antipsychotic treatment.

It is sometimes difficult to distinguish mood disorders with psychotic features from schizophreniform disorder. The acute onset of symptoms may resemble the onset of mood disorder episodes. In addition, mood symptoms such as loss of interest and pleasure may be difficult to distinguish from negative symptoms, avolition, and anhedonia. A thorough longitudinal history is important in elucidating the diagnosis since the presence of psychotic symptoms exclusively during periods of mood disturbance is an indication of a primary mood disorder.

Course and Prognosis

The course and prognosis of schizophreniform disorder are variable and depend largely on whether the diagnosis remains schizophreniform disorder or whether the provisional diagnosis of schizophreniform disorder changes over the course of longitudinal follow-up. There are conflicting data on the diagnostic stability of schizophreniform disorder. One small study found that all patients in a first-episode cohort who were initially diagnosed with schizophreniform disorder were subsequently diagnosed with schizophrenia or schizoaffective disorder at 2-year follow-up. However, most studies have found that 45% to 75% of patients who receive a diagnosis of schizophreniform disorder will have a subsequent change in diagnosis to schizophrenia or schizoaffective disorder with longitudinal assessment. Evolution of schizophreniform disorder to a mood disorder was less common, occurring in about 5% to 15% of patients in most studies with a maximal rate of 35% in a single study. Based on data from one study, greater rates of progression to a mood disorder may be seen in individuals with schizophreniform disorder who have good prognostic features. Only about 20% to 40% of patients maintained the diagnosis of schizophreniform disorder over follow-up periods of several years. With longer periods of follow-up, the proportion of individuals who retain the diagnosis of schizophreniform disorder is even smaller. In the Suffolk County, New York, first-admission sample, less than 5% of those who initially received a diagnosis of schizophreniform disorder retained that diagnosis at the time of the 10-year follow-up assessment. These findings underscore the importance of regarding the diagnosis of schizophreniform disorder as provisional when it is made cross-sectionally.

For individuals whose diagnosis remains schizophreniform disorder after longitudinal follow-up, the illness course is usually benign. In general, these patients will have met criteria for “good prognostic features” upon first presentation.

Individuals classified as “without good prognostic features” often have a course that resembles that of patients initially diagnosed with schizophrenia. For many of these patients, their diagnosis indeed changes to schizophrenia over time.

DELUSIONAL DISORDER

History

Delusional disorders, which were once referred to as paranoid disorders or paranoia, have been recognized since the time of the ancient Greeks. In the early 19th century, Johann Heinroth and later Jean Esquirol described a disorder that influenced the intellect in which irrational ideas and actions appeared to develop through plausible and logical reasoning. In 1863, Karl Kahlbaum, who referred to the syndrome as “diastrephia,” made the important observation that personality is preserved during the course of the illness. Late in the 19th century, Valentin Magnan described a chronic systematized delusional disorder in which delusional ideas were well circumscribed and highly structured.

Kraepelin described paranoia as a third form of psychosis, distinct from both schizophrenia and mood disorder and consisting of chronic, nonbizarre, well-systematized delusions without the typical deterioration of dementia praecox. He noted that “One invariably observes the insidious development of a permanent unshakable delusional system from inner causes in which clarity and order of thinking, willing and action are completely preserved.” In the sixth edition of his textbook, Kraepelin also introduced the concept of paranoia in the older adults, which would later be termed paraphrenia.

The distinction between paranoid disorder and paranoid schizophrenia was maintained by Eugen Bleuler. However, Kurt

Kolle reassessed Kraepelin’s patients at a later time and showed that their pattern of deterioration was similar to that of patients with dementia praecox. This view was expressed in DSM-II in which paranoid states were regarded as possible variants of schizophrenia.

Beginning with the third edition, the DSM adopted the Kraepelinian concept of paranoia as consisting of nonbizarre delusions without prominent hallucinations. DSM-III-R changed the name of the disorder to delusional disorder for clarity and to encompass other types of delusion in addition to persecutory delusions. The DSM-IV and DSM-IV-TR also acknowledged that delusions may show varying levels of conviction and that delusional beliefs may be partially responsive to evidence or reason. The major change between DSM-5 and prior editions of the DSM in the definition of delusional disorder is the removal of the requirement that the delusions be nonbizarre. Furthermore, in DSM-5, a specifier for delusions “with bizarre content” was added if the delusions are clearly implausible, not understandable, and not derived from ordinary life experiences. The example of bizarre delusion provided in the manual is the delusional belief that the person’s internal organs have been replaced with someone else’s organs without leaving any wounds or scars. The DSM-5 further added course specifiers and current severity specifiers to the criteria (Table 12.17–3).

Comparative Nosology

In the DSM-5-TR criteria for delusional disorder, the emphasis is on the presence of one or more delusions lasting for more than a month, in the absence of other criterion A symptoms of schizophrenia. Hallucinations, if present, are not prominent and are related to the content of delusions.

The ICD-11 description of delusional disorder is similar to the DSM-5-TR description except that the minimum duration is 3 months.

Notwithstanding these differences, delusional disorder is typically chronic in nature, and both systems of classification will most likely include similar patients. Similar to DSM-5-TR, the ICD-11 allows for the presence of transient and nonprominent perceptual disturbances, including hallucinations, illusions, and misidentifications of persons that are thematically related to the delusional content.

Epidemiology

Studies of the epidemiology of delusional disorders have been difficult to conduct because the condition occurs infrequently. However, in a general population study from Finland, the lifetime prevalence of DSM-IV delusional disorder was 0.18%. In clinical studies, individuals with delusional disorder represent from 1% to 4% of psychiatric admissions. The persecutory type is the most common type of delusional disorder. Other than the jealous type that appears to be more common in men, delusional disorders appear to be equally common in the two genders.

Etiology

The etiology of delusional disorder is not known and its infrequent occurrence limits the ability to examine genetic or other risk factors. However, small family studies have failed to identify strong genetic links between delusional disorder and schizophrenia. Abnormalities in grey matter volume and cortical surface have been noted in a number of imaging studies. Interestingly, the specific patterns of these deficits appear to vary according to the content of delusions. For example, lower gray matter volume in thalamic, striatal (putamen), insular, and medial prefrontal brain regions were found in delusional parasitosis (a form of somatic delusion); whereas, patients with nonsomatic delusional disorder showed lower gray matter volume in temporal cortical regions.

Table 12.17–3.
DSM-5 Diagnostic Criteria for Delusional Disorder

Diagnostic Criteria

297.1 (F22)

1. The presence of one (or more) delusions with a duration of 1 month or longer.
2. Criterion A for schizophrenia has never been met.

   Note: Hallucinations, if present, are not prominent and are related to the delusional theme (e.g., the sensation of being infested with insects associated with delusions of infestation).
3. Apart from the impact of the delusion(s) or its ramifications, functioning is not markedly impaired, and behavior is not obviously bizarre or odd.
4. If manic or major depressive episodes have occurred, these have been brief relative to the duration of the delusional periods.
5. The disturbance is not attributable to the physiological effects of a substance or another medical condition and is not better explained by another mental disorder, such as body dysmorphic disorder or obsessive–compulsive disorder.

Specify whether:

Erotomanic type: This subtype applies when the central theme of the delusion is that another person is in love with the individual.

Grandiose type: This subtype applies when the central theme of the delusion is the conviction of having some great (but unrecognized) talent or insight or having made some important discovery.

Jealous type: This subtype applies when the central theme of the individual’s delusion is that his or her spouse or lover is unfaithful.

Persecutory type: This subtype applies when the central theme of the delusion involves the individual’s belief that he or she is being conspired against, cheated, spied on, followed, poisoned or drugged, maliciously maligned, harassed, or obstructed in the pursuit of long-term goals.

Somatic type: This subtype applies when the central theme of the delusion involves bodily functions or sensations.

Mixed type: This subtype applies when no one delusional theme predominates.

Unspecified type: This subtype applies when the dominant delusional belief cannot be clearly determined or is not described in the specific types (e.g., referential delusions without a prominent persecutory or grandiose component).

Specify if:

With bizarre content: Delusions are deemed bizarre if they are clearly implausible, not understandable, and not derived from ordinary life experiences (e.g., an individual's belief that a stranger has removed his or her internal organs and replaced them with someone else's organs without leaving any wounds or scars).

Specify if:

The following course specifiers are only to be used after a 1-year duration of the disorder:

First episode, currently in acute episode: First manifestation of the disorder meeting the defining diagnostic symptom and time criteria. An *acute episode* is a time period in which the symptom criteria are fulfilled.

First episode, currently in partial remission: *Partial remission* is a time period during which an improvement after a previous episode is maintained and in which the defining criteria of the disorder are only partially fulfilled.

First episode, currently in full remission: *Full remission* is a period of time after a previous episode during which no disorder-specific symptoms are present.

Multiple episodes, currently in acute episode

Multiple episodes, currently in partial remission

Multiple episodes, currently in full remission

Continuous: Symptoms fulfilling the diagnostic symptom criteria of the disorder are remaining for the majority of the illness course, with subthreshold symptom periods being very brief relative to the overall course.

Unspecified

Specify current severity:

Severity is rated by a quantitative assessment of the primary symptoms of psychosis, including delusions, hallucinations, disorganized speech, abnormal psychomotor behavior, and negative symptoms. Each of these symptoms may be rated for its current severity (most severe in the last 7 days) on a 5-point scale ranging from 0 (not present) to 4 (present and severe). (See Clinician-Rated Dimensions of Psychosis Symptom Severity in the chapter "Assessment Measures.")

Note: Diagnosis of delusional disorder can be made without using this severity specifier.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:104–106.

Psychodynamic theories hypothesize that persecutory delusions, such as those found in delusional disorder, are a protective psychological response to stresses or conflicts that represent a profound threat to the self (e.g., fear of the unknown and adjustment to immigration). In responding to the threat, there is an affective withdrawal from intimate ties and an intense effort to retain the appearance of normalcy through defensive mechanisms such as projection. At its extreme, the paranoid defense fails to relieve anxiety and escalates to a distinct maladaptive state, which is postulated to result in the development of a delusional disorder.

Another theory of the etiology of delusional disorder is based on research by cognitive and experimental psychologists and suggests that persons with persecutory delusions selectively attend to threatening information, jump to conclusions on the basis of insufficient information, attribute negative events to external personal causes, and have difficulty in envisaging others' intention and motivations. Preferential recall of threatening episodes has also been postulated as a mechanism for reinforcing the delusional belief.

Diagnostic and Clinical Features

The DSM-5-TR lists five specific subtypes of delusional disorder based on their content: persecutory, jealous, somatic, grandiose, and erotomanic. It also allows for mixed and unspecified types.

Persecutory Type. The persecutory type is the best-recognized type of delusional disorder in which the individual is convinced that they (or someone to whom the person is close) are being malevolently treated or harmed in some way. These persecutory beliefs are often associated with querulousness, irritability, anger, and at times aggressive or even homicidal behavior. The individual may also become preoccupied with formal litigation. In contrast to the persecutory delusions of schizophrenia, the delusions in delusional disorder are systematized and coherent. With the exception of impairments related to the delusion itself, social functioning is not compromised. In some of these individuals, the demarcation between normality, pathologic behavior, overvalued ideas, and delusion may not be very clear.

Jealous Type. Feelings of jealousy can fall on a continuum from normal jealousy to pathologic jealousy and to jealousy that is delusional in nature, with the thresholds for abnormal jealousy being, in part, socially determined.

The most common delusion of jealousy is the belief that one's spouse or partner is unfaithful. In some instances, some degree of infidelity may actually be occurring, yet the magnitude of the jealous response and the "evidence" accumulated to support accusations of unfaithfulness may take on a delusional quality. Jealousy is dangerous, for it not only evokes anger but also arms jealous individuals with a sense of righteous indignation to justify their acts of aggression. Jealousy focuses primarily on the partner and can sometimes lead the partner or others to be harmed or even murdered.

Erotomanic Type. In the erotomanic type of delusional disorder, which has also been referred to as De Clerambault syndrome or *psychose passionelle*, the patient has the delusional conviction that another person, usually of higher status, is in love with them. Patients will generally describe the love object as having fallen in love with them first and as having made the first advances. Patients also describe elaborate rationalizations when the love object ignores or rejects their contacts and communications. Patients who develop erotomania often tend to be solitary, withdrawn, dependent, and sexually inhibited and have low levels of social or occupational functioning. Symptoms typically have an acute onset and become chronic.

Erotomanic symptoms may also occur in the context of other psychiatric disorders. In fact, the majority of patients with erotomanic symptoms will have a diagnosis other than delusional disorder.

Stalking may be observed in individuals with the erotomanic type of delusional disorder, although most stalkers do not meet the criteria for erotomania. Stalkers most often persecute their targets through unwanted communications that consist of frequent telephone calls, letters, emails, text or social media messages, or mail packages. Typically, stalkers will also follow their target and make unwelcome or intrusive and unexpected appearances in their target’s public domain. Along with this pattern of intrusive behavior is an implicit or explicit threat, which leads the stalking target to experience considerable fear. In some circumstances, the violence may escalate until the stalker actually physically attacks the target or individuals associated with the target.

Somatic Type. The somatic type of delusional disorder is the only one of the five specific types that was not included in Kraepelin’s original description of paranoia. It is diagnosed when the central themes of the delusional beliefs are of a hypochondriacal or somatic nature. As a result, in the older literature, this type of delusional disorder is sometimes referred to as monosymptomatic hypochondriasis. As with other subtypes of delusional disorder, the patient must have a clear sensorium and the symptoms may not result from an underlying general medical condition or another psychiatric disorder.

In the somatic type of delusional disorder, the onset of symptoms may be gradual or sudden. In most patients, the illness is unremitting, although the severity of the delusion may fluctuate. Anxiety and hyperalertness also characterize patients with this subtype. Some themes recur, such as concerns about infestation in delusional parasitosis, preoccupation with body features with dysmorphic delusions, and delusional concerns about body odor, which are sometimes referred to as “bromosis.” In delusional parasitosis, tactile sensory phenomena are often linked to the delusional beliefs.

Patients with the somatic type of delusional disorder rarely present for psychiatric evaluation, and when they do, it is usually in the context of a psychiatric consultation. Instead, these patients generally present to a specific medical specialist for evaluation. Thus, these individuals are more often encountered by dermatologists, plastic surgeons, urologists, and other medical specialists, including dentists.

Grandiose Type. The grandiose type of delusional disorder was included in Kraepelin’s original description of paranoia. However, it is relatively rare to encounter a patient who has grandiose delusions alone without associated symptoms of mania that would suggest a diagnosis of bipolar disorder with psychotic features, or other psychotic symptoms that would indicate a diagnosis of schizophrenia. The content of delusions in some other delusional disorder types may include elements of grandiosity. For example, an individual with a persecutory delusion may attribute the persecution to their own special qualities. However, these grandiose elements are rarely elaborated into a full-blown delusion.

Mixed Type. The mixed type of delusional disorder includes individuals who exhibit more than one of the preceding types simultaneously. These delusions may be bizarre in nature. Of note, under DSM-IV-TR, the presence of a bizarre delusion would have satisfied criterion A for schizophrenia or schizophreniform disorder and ruled out a diagnosis of delusional disorder. However, under DSM-5-TR, at least two of the criterion A symptoms need to be met for a diagnosis of schizophrenia or schizophreniform disorder, and delusional disorder can be diagnosed even if the delusions are bizarre.

Unspecified Type. Although the preceding themes are the most common themes observed among individuals with delusional disorder, other delusions may occur as well. One example is the Cotard syndrome, in which individuals report that they have lost all of their possessions and status as well as their entire being and even their internal organs. Other examples include the misidentification syndromes such as the Fregoli syndrome, in which strangers are misidentified as a single persecutor wearing different disguises, and the Capgras syndrome, in which the individual believes that their family members or other familiar people have been replaced by identical-looking imposters. Individuals experiencing misidentification delusions may act violently toward the presumed imposter, believing that the imposter is doing harm to the individual or a loved one.

Differential Diagnosis

The first step in the clinical assessment is establishing whether a delusion exists. Some statements that initially seem to be delusional may, in fact, be true. In contrast, reports of circumstances that initially seem believable may only be clearly identified as delusions as the symptoms worsen, the delusion becomes less encapsulated, or more information comes to light. At times, having multiple sources of information is helpful in deciding the delusional nature of a belief.

Certain behavioral features such as hypersensitivity, guardedness, evasiveness, secretiveness, litigiousness, overly detailed descriptions, hostility, and humorlessness are characteristic of the delusional patient yet are not pathognomonic. The patient may also be secretive, vague, and not voluntarily provide the clinician with necessary details. At other times, the patient may see no need to convey information, believing that the interviewer is part of the "conspiracy" and already knowledgeable about its specifics. For example, the patient may view the interview as a staged game where both parties know the hidden meanings behind the overt text. As a result, certain phrases are commonly heard from individuals with persecutory ideas in the context of delusional disorder (e.g., "Don't play games with me," "You know what 'they' have been doing").

Characteristic observations or reports about the patient's behavior (e.g., checking for microphones in persecutory delusions and checking the spouse's clothes in delusions of jealousy) may also lend support to the diagnosis. Even with such observations, it is sometimes difficult to distinguish between a true situation, an overvalued idea, or a delusion. Confabulation may, at times, be difficult to distinguish from delusions, although delusions are usually more systematized. Often, the inappropriateness of the patient's behavior or emotional response, rather than the simple truth or falsity of the belief, will indicate its delusional nature. Consequently, when evaluating individuals with possible delusional disorder, the most sensible approach is to gather as much data as possible from many sources and then utilize clinical judgment in determining whether a threshold indicating psychopathologic disturbance has been passed.

In making a diagnosis of delusional disorder, the first step is to eliminate general medical conditions as potential causes of delusions. Many general medical conditions may be associated with the development of delusions, and delusion may, at times, accompany a delirious state.

Metabolic conditions and disorders affecting the limbic system and basal ganglia are most often associated with the emergence of delusional beliefs. Complex delusions occur more frequently in patients with subcortical pathology. In Huntington disease and in individuals with idiopathic basal ganglia calcifications, for example, more than 50% of patients demonstrate delusions at some point in their illness. Anosognosia (i.e., nonrecognition of disability) and reduplicative paramnesia (i.e., individuals believing they are in different places at the same time) that accompany right-side cerebral infarctions may seem like delusional beliefs. Capgras syndrome has been observed in a number of medical disorders, including CNS lesions, vitamin B12 deficiency, hepatic encephalopathy, diabetes, and hypothyroidism. Focal lesions associated with such misidentification delusions have more often involved the right than the left hemisphere. Delusions of infestation, lycanthropy (i.e., the false belief that the patient is an animal, often a wolf or werewolf), and erotomania have been reported in small numbers of patients with epilepsy, CNS lesions, or toxic–metabolic disorders. In such cases, a diagnosis of psychotic disorder due to another medical condition is appropriate.

In dementia syndromes of all types, 15% to 50% of individuals experience delusions. Most of these delusions are simple, although a minority of patients may have complex, well-systematized delusions. Persecutory delusions are particularly common in early dementia and may be the first manifestation of a dementing illness in some individuals.

Once general medical conditions, delirium, and dementia have been eliminated as possible diagnoses, substance use and medications must be explored as possible causes. Among the substances of abuse, amphetamines and cocaine are most likely to result in delusional beliefs, which are typically persecutory in nature. Prescribed substances, over-the-counter medications, and herbal products such as ephedra may also be associated with the development of delusions. Delusional beliefs clearly linked to a substance of abuse or medication are categorized as substance/medication-induced psychotic disorder.

In differentiating delusional disorder from other psychotic disorders, a number of factors need to be taken into consideration. Individuals with schizophrenia will often have hallucinations or disorganized thoughts or behaviors that are not present in delusional disorder. In addition, prominent negative symptoms or deterioration in function from previous premorbid levels would suggest a diagnosis of schizophrenia. Functional impairment in delusional disorder is a consequence of the delusions themselves (e.g., being fired from work after harassing the boss in the context of an erotomanic delusion).

Mood disorders are also important in the differential diagnosis of delusional disorder. Mood symptoms are highly prevalent in delusional disorders. Symptoms of depression may seem to be an understandable response to the perceived delusional experiences (e.g., feeling overwhelmed and demoralized over “harassment” by neighbors or coworkers). In addition, given the prevalence of mood disorders, many individuals who present with a delusional disorder will be found to have had a history of depression. Somatic, grandiose, or persecutory delusions that are similar to those observed in patients with delusional disorder may also be seen in individuals with psychotic mood disorders. In distinguishing mood disorders with psychotic features from delusional disorder, it is therefore important to determine whether a full mood syndrome (manic, depressed, or mixed episode) is present or whether milder mood symptoms that do not meet the criteria for a mood disorder are present. It is also important to consider the temporal pattern of symptom onset and progression, recognizing that the boundaries between mood disorder and delusional disorder will generally become clear over time. If mood disorders have occurred concurrently with the delusions, their duration should have been brief relative to the duration of the delusions.

Perhaps the most difficult in terms of differential diagnosis is differentiating between an overvalued idea, an obsession with a lack of insight, and a delusion. For example, the somatic type of delusional disorder may be easily confused with illness anxiety disorder (hypochondriasis). In illness anxiety disorder, there is a preoccupation with fears or ideas of having a serious disease, and these concerns are based on the person’s misinterpretation of bodily symptoms. In addition, the somatic preoccupation persists despite appropriate medical evaluation and reassurance. Although the patient may be briefly relieved and begin to doubt their ideas of having an illness for a short while after each evaluation, the fear generally emerges again. Although many individuals with illness anxiety disorder have some realization of the excessive nature of their fears, some individuals may lack insight, which makes differential diagnosis difficult. In most individuals with illness anxiety disorder, however, there is a long pattern of being preoccupied with physical symptoms, and the preoccupations usually involve more than one symptom or organ system. Furthermore, somatic delusions are generally more bizarre than the health concerns in illness anxiety disorder.

In body dysmorphic disorder, there is a preoccupation with a slight or imagined defect in appearance. These individuals may recognize that their view of their appearance is distorted, yet a significant proportion of individuals with body dysmorphic disorder will hold their beliefs with delusional intensity. In these cases, a diagnosis of body dysmorphic disorder, with “absent insight/delusional beliefs” specifier is appropriate.

Similarly, reality testing may be poor in some individuals with obsessive–compulsive disorder (OCD). The DSM-5 diagnosis of OCD requires that the individual must recognize that the obsessions or compulsions are excessive or unreasonable at some time. DSM-5-TR includes a specifier for the degree of insight. In practice, patients show varying degrees of insight into the validity of their beliefs and some individuals do not recognize that their symptoms are unreasonable. If the obsession reaches “delusional proportions,” DSM-5-TR suggests that a diagnosis of OCD with the “absent insight/delusional beliefs” specifier is the appropriate diagnosis.

When persecutory ideas are present and patients are guarded about sharing full details of their delusional beliefs, it may be particularly difficult to differentiate delusional disorder from a paranoid personality disorder. In these cases, the longitudinal history will be important in establishing a diagnosis of a personality disorder and also in determining whether delusions are superimposed on the individual’s premorbid personality. If a patient met the criteria for a paranoid personality disorder prior to the onset of psychosis and continues to meet these criteria after the psychosis has subsided, such individuals would be given a diagnosis of delusional disorder as well as a diagnosis of paranoid personality disorder.

Course and Prognosis

The course of delusional disorder has been difficult to study for a number of reasons, including the relatively low prevalence, the unstable nature of the diagnosis, and the tendency for patients with delusional disorder to avoid treatment by medical professionals in general and by psychiatrists in particular. The age of onset of delusional disorder is generally in the mid- to late-30s, with the vast majority of individuals having an insidious onset of symptoms. In addition, delusional disorder is generally associated with less deterioration in functioning over time than schizophrenia or schizoaffective disorder. Although the specific delusions may persist and remain fixed, they are generally stable. Functioning tends to be preserved and negative outcomes from the disorder typically occur when individuals are unable to refrain from acting on their delusional beliefs. The impact of age of onset on the course is unclear, with some studies showing a more favorable prognosis with earlier age of onset and others showing no effect of age of onset on course. In some studies, the presence of precipitating factors has been associated with better outcomes, along with female sex and being married. The presence of co-occurring diagnoses, particularly depression, may be an additional factor that negatively influences the course of the disorder.

SCHIZOAFFECTIVE DISORDER

History

Kraepelin divided the psychoses into dementia praecox characterized by progressive deterioration and manic-depressive illness characterized by a remitting course; however, his rich clinical examples included many patients with features of both disorders. In the early 20th century, Lange, Bowman, and Dunton also described intermittent forms of dementia praecox that presented with periods of excitement or depression along with gradual deterioration in functioning. Bleuler believed that the existence of mood and schizophrenic symptoms during the same episode did not alter the diagnosis of schizophrenia and pointed out the need for prolonged observation to correctly differentiate schizophrenia from manic-depressive psychosis.

Jacob Kasanin, in a 1933 paper, was the first to coin the term schizoaffective psychosis to identify patients with characteristics of both schizophrenia and mood disorders. He based his observation on a distinct subgroup of nine young patients with good premorbid functioning, a sudden onset of both schizophrenic and mood symptoms, and good outcome. However, a review of his detailed case histories shows that all of these individuals could have been diagnosed as having a mood disorder with psychotic features under the DSM classification.

DSM-I described a schizoaffective type of schizophrenia reaction diagnosed in individuals “showing significant admixtures of schizophrenic and affective reactions.” The manual also noted that “The mental content may be predominantly schizophrenic, with pronounced elation or depression. Cases may show predominantly affective changes with schizophrenic-like thinking or bizarre behavior.” However, the long-term course of these individuals was said to be “basically schizophrenic in nature.” DSM-II maintained the diagnosis of schizophrenia, schizoaffective type, and further divided the category into excited and depressed subtypes.

Under DSM-III, disorders involving full affective syndromes with schizophrenia-like symptoms were classified as mood disorders with mood-incongruent psychotic symptoms. As long as the schizophrenic symptoms occurred at the same time as the mood disturbance, the patient was considered to have a mood disorder. The DSM-III relegated schizoaffective disorder to the psychotic disorders not elsewhere classified, reserved for cases that did not clearly meet the diagnostic criteria for schizophrenia or mood disorders. In the DSM-III-R, schizoaffective disorder became more clearly defined with the specification that “there have been delusions or hallucinations for at least 2 weeks, but no prominent mood symptoms” and that “the duration of all episodes of a mood syndrome has not been brief relative to the total duration of the psychotic disturbance." DSM-IV and DSM-IV-TR incorporated the stricter time frame of 1-month duration of schizophrenia symptoms (as part of criterion A). DSM-5 and DSM-5-TR further elaborated on the duration of mood symptoms by requiring that the symptoms for a major mood episode should be present for the majority of the total duration of the active and residual portions of the illness.

Comparative Nosology

Both the DSM-5-TR and ICD-11 criteria for schizoaffective disorder require that there be a period of concurrent psychosis and disturbance of mood (Table 12.17–4).

DSM-5-TR specifically requires "An uninterrupted period of illness during which there is a major mood episode (major depressive or manic) concurrent with Criterion A of Schizophrenia."

ICD-11 similarly requires that "diagnostic requirements of schizophrenia and a manic, mixed, or moderate or severe depressive episode are met within the same episode of illness, either simultaneously or within a few days of each other." However, DSM-5-TR requires a period of at least 2 weeks during the lifetime duration of illness in which delusions or hallucinations occur in the absence of a major mood episode. In addition, DSM-5-TR requires that "Symptoms that meet criteria for a mood episode are present for the majority of the total duration of the active and residual portions of the illness." In contrast, ICD-11 requires that mood and psychotic symptoms both be a prominent part of the clinical picture.

DSM-5-TR schizoaffective disorder is further divided into two subtypes based on whether or not manic episodes are present with or without major depressive episodes (bipolar type) or the mood episodes are limited to the major depressive episodes (depressive type).

Epidemiology

Few studies have examined the prevalence of schizoaffective disorder in general population samples. A general population study in Finland using DSM-IV criteria found a lifetime rate of schizoaffective disorder of 0.32%.

Table 12.17–4.
DSM-5 Diagnostic Criteria for Schizoaffective Disorder

Diagnostic Criteria

1. An uninterrupted period of illness during which there is a major mood episode (major depressive or manic) concurrent with Criterion A of schizophrenia.

   Note: The major depressive episode must include Criterion A1: Depressed mood.
2. Delusions or hallucinations for 2 or more weeks in the absence of a major mood episode (depressive or manic) during the lifetime duration of the illness.
3. Symptoms that meet criteria for a major mood episode are present for the majority of the total duration of the active and residual portions of the illness.
4. The disturbance is not attributable to the effects of a substance (e.g., a drug of abuse and a medication) or another medical condition.

Specify whether:

295.70 (F25.0) Bipolar type: This subtype applies if a manic episode is part of the presentation. Major depressive episodes may also occur.

295.70 (F25.1) Depressive type: This subtype applies if only major depressive episodes are part of the presentation.

Specify if:

With catatonia (refer to the criteria for catatonia associated with another mental disorder, pp. 119–120, for definition).

Coding note: Use additional code 293.89 (F06.1) catatonia associated with schizoaffective disorder to indicate the presence of the comorbid catatonia.

Specify if:

The following course specifiers are only to be used after a 1-year duration of the disorder and if they are not in contradiction to the diagnostic course criteria.

First episode, currently in acute episode: First manifestation of the disorder meeting the defining diagnostic symptom and time criteria. An *acute episode* is a time period in which the symptom criteria are fulfilled.

First episode, currently in partial remission: *Partial remission* is a time period during which an improvement after a previous episode is maintained and in which the defining criteria of the disorder are only partially fulfilled.

First episode, currently in full remission: *Full remission* is a period of time after a previous episode during which no disorder-specific symptoms are present.

Multiple episodes, currently in acute episode: Multiple episodes may be determined after a minimum of two episodes (i.e., after a first episode, a remission and a minimum of one relapse).

Multiple episodes, currently in partial remission

Multiple episodes, currently in full remission

Continuous: Symptoms fulfilling the diagnostic symptom criteria of the disorder are remaining for the majority of the illness course, with subthreshold symptom periods being very brief relative to the overall course.

Unspecified

Specify current severity:

Severity is rated by a quantitative assessment of the primary symptoms of psychosis, including delusions, hallucinations, disorganized speech, abnormal psychomotor behavior, and negative symptoms. Each of these symptoms may be rated for its current severity (most severe in the last 7 days) on a 5-point scale ranging from 0 (not present) to 4 (present and severe). (See Clinician-Rated Dimensions of Psychosis Symptom Severity in the chapter "Assessment Measures.")

Note: Diagnosis of schizoaffective disorder can be made without using this severity specifier.

Note: For additional information on Development and Course (age-related factors). Risk and Prognostic Factors (environmental risk factors), Culture-Related Diagnostic Issues, and Gender-Related Diagnostic Issues, see the corresponding sections in schizophrenia, bipolar I and II disorders, and major depressive disorder in their respective chapters.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:121, 122.

Gender differences in the rates of schizoaffective disorder in clinical samples generally parallel gender differences seen in mood disorders, with approximately equal numbers of men and women who have the bipolar subtype and a more than twofold female-to-male predominance among individuals with the depressed subtype of schizoaffective disorder. Individuals with schizoaffective disorder are more likely to be married or employed than individuals with schizophrenia but less likely than those with bipolar disorder. Other sociodemographic risk factors have not been well studied.

Etiology

In etiologic research, analyses have typically combined individuals with schizoaffective disorder and those with schizophrenia into a single category. As a result, little is known about the distinctive etiologic factors associated with schizoaffective disorder. However, genetic factors may play a role because relatives of probands with schizoaffective disorder have higher rates of mood disorder than relatives of probands with schizophrenia and higher rates of schizophrenia than relatives of probands with a mood disorder. Studies of the *disrupted-in-schizophrenia 1* (*DISC1*) gene, located on chromosome 1q42, suggest its possible involvement in schizoaffective disorder as well as schizophrenia and bipolar disorder.

Diagnostic and Clinical Features

According to the DSM-5-TR, a diagnosis of schizoaffective disorder is based on the presence of both criteria A for schizophrenia (i.e., delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms) and presence of concurrent symptoms meeting the criteria for a depressive or a manic episode. To distinguish depressive symptoms from negative symptoms of schizophrenia, the DSM-5-TR criteria for schizoaffective disorder, depressed type, emphasize that the A1 criterion of a major depressive episode must be fulfilled (i.e., the presence of “depressed mood” and not simply anhedonia and/or loss of interest).

Determining whether an individual fulfills the DSM-5-TR criteria for schizoaffective disorder also requires that the clinician take a detailed temporal history of the onset and cessation of mood and psychotic symptoms. To accomplish this, the patient must be able to clearly distinguish when mood and psychotic symptoms started and stopped. Because retrospective reporting is inherently unreliable and both types of symptoms can begin or end insidiously, this is not always a simple task for either the clinician or the patient. Without sufficient information about the length of psychosis and mood symptoms, it is difficult to know whether the group A criteria (psychotic symptoms of at least 1 month’s duration), the group B criteria (psychotic symptoms in the absence of the mood syndrome), and the group C criteria (mood symptoms occurring for the majority of the total duration of the active and residual portions of the illness) are fully met. It is particularly difficult to make retrospective estimates of the duration of these criteria. In a series of approximately 700 first-admission patients with psychotic features, 6% received a diagnosis of schizoaffective disorder at hospital discharge, but only half of these (3% of the total) met DSM-III-R criteria for schizoaffective disorder upon review by the research psychiatrists. Two years later, only one-third of those initially diagnosed with schizoaffective disorder were again diagnosed with this disorder, one-third were diagnosed with schizophrenia, and another third with an affective disorder. Recent phenomenologic research has failed to support the validity of the 2-week cutoff for duration of psychotic symptoms in the absence of mood episode for distinguishing schizoaffective disorder from other disorders with concurrent mood and psychotic symptoms.

In clinical practice, the diagnosis of schizoaffective disorder is more commonly made than justified by strict application of DSM-5-

TR criteria. Individuals with a chronic longitudinal course of illness and elements of psychosis and mood disturbance are sometimes erroneously classified as cases of schizoaffective disorder, as are some individuals who have psychosis in the absence of mood symptoms and prominent mood episodes but who do not fulfill criteria for mood episodes during a significant portion of the illness. Schizoaffective disorder has also been mistakenly viewed as a compromise between the diagnoses of schizophrenia and mood disorder when the exact diagnosis is unclear.

Differential Diagnosis

A critical step in differential diagnosis is determining whether a substance use disorder or a general medical condition could account for the patient's symptoms. A substance/medication-induced mood or psychotic disorder is diagnosed when there is evidence from the history, physical examination, or laboratory tests indicating that the symptoms are the direct physiologic consequence of a specific substance or medication. A number of prescribed and over-the-counter medications and substances of abuse may produce mood change and psychosis either with chronic use, acute intoxication, or withdrawal. For example, acute or chronic use of alcohol can produce hallucinosis and can also predispose the patient to a depressive state. The chronic use of cannabis may precipitate psychosis and can also lead to an anergic state, which shares many features with depression (amotivational syndrome). Chronic cocaine use may similarly be associated with psychosis as well as changes in mood. Thus, careful screening for general medical conditions and substance misuse is an essential aspect in the differential diagnosis of schizoaffective disorder. Similarly, mood or psychotic disorder due to another medical condition is diagnosed when there is evidence from the history, physical examination, or laboratory tests indicating that the symptoms are the direct consequence of another medical condition. A wide range of medical etiologies including neurologic disorders, infectious diseases, metabolic abnormalities, and endocrinopathies may result in a combination of psychosis and mood change that may resemble schizoaffective disorder.

In distinguishing between schizoaffective disorder and mood disorder with psychotic features, several elements should be taken into consideration. Many clinicians fail to assess the relative persistence of psychotic and mood symptoms between acute episodes, and others expand the definition of schizoaffective disorder to include chronic psychosis with a few prominent mood symptoms or mood disorders with an acute and severe psychotic presentation. Patients often view depression as more understandable or socially acceptable than psychosis and may retrospectively emphasize the duration of depressive symptoms in their illness. It is also difficult to disentangle negative symptoms from depressive symptoms by history alone. Without a clear history of a euphoric mood, retrospectively assessed manic episodes may be viewed as episodes of acute psychosis associated with agitation or irritability. Thus, it is important to take a thorough history using information from as many informants as possible to delineate whether criteria were met for both a mood episode and for schizophrenia. Obtaining a thorough temporal history can also be useful in distinguishing a mood disorder, typically depression that is superimposed on schizophrenia. Even with a detailed history, the diagnosis of schizoaffective disorder may only become clear over time.

Course and Prognosis

Although the course and prognosis of schizoaffective disorder may vary, the outcome is generally poorer than in mood disorders but somewhat better than schizophrenia. In the Suffolk County, New York, first-admission series, 75% of patients with this disorder did not remit fully over the first 2 years of follow-up, compared to 87% of patients with schizophrenia and 22% and 40% of those with bipolar disorder with psychotic features or major depressive disorder with psychotic features, respectively. Individuals with schizoaffective disorder may also have levels of neurocognitive impairment and deficits in social and occupational functioning that are greater than those of individuals with bipolar disorder. In one European study, the death rate of schizoaffective disorder due to unnatural causes was comparable to the death rate of bipolar disorder and greater than schizophrenia.

SUBSTANCE/MEDICATION-INDUCED PSYCHOTIC DISORDER

History

Psychotic symptoms in alcoholics and in opioid or hashish intoxication have been known for many years. In the DSM-I, these disorders were introduced under the rubrics of acute or chronic brain syndromes associated with alcohol intoxication, with drug or poison intoxication. However, the DSM-I did not distinguish between psychotic symptoms that occur in the context of delirium and those that occur in clear sensorium. These syndromes continued to be classified under organic brain syndromes in DSM-II. However, the DSM-II also introduced the categories of other alcoholic hallucinosis characterized by accusatory or threatening auditory hallucinations in a state of relatively clear consciousness and alcohol-paranoid state (alcoholic paranoia) that develops in chronic alcoholics and is characterized by excessive jealousy and delusions of infidelity by the spouse. The DSM-IV and DSM-IV-TR combined the organic hallucinosis and delusional syndromes into a single category of substance-induced psychotic disorders in the section on schizophrenia and other psychotic disorders and introduced specifiers for onset during intoxication and onset during withdrawal. This approach to classification of these disorders was maintained in the DSM-5 and DSM-5-TR, although the name of the disorder was changed to substance/medication-induced psychotic disorder.

Diagnostic and Clinical Features

According to the DSM-5-TR, substance/medication-induced psychotic disorder is characterized by the presence of delusions and/or hallucinations that have developed during intoxication or withdrawal from a medication or substance or soon after exposure to it. The substance or medication involved is one that is known to produce psychotic symptoms. Furthermore, the symptoms are not better explained by an independent psychotic disorder, such as schizophrenia. DSM-5-TR presents some features of the disorder that could be helpful in the differential diagnosis of an independent psychotic disorder, including psychotic symptoms that preceded substance/medication use, recurrent episodes of the disorder, or symptoms that last for a long period of time (1 month or longer) after withdrawal or intoxication symptoms are resolved, all suggesting an independent psychotic disorder.

Symptoms that occur exclusively during the course of delirium should not be diagnosed as a substance/medication-induced psychotic disorder. Also, in the context of intoxication or withdrawal, a diagnosis of substance/medication-induced psychotic disorder can only be made when the psychotic symptoms predominate the clinical picture and are sufficiently severe to need clinical attention. Specifiers “with onset during intoxication” and “with onset during withdrawal” can be used to signal association with intoxication or withdrawal, respectively, and to indicate that the diagnostic criteria for intoxication or withdrawal are met.

Comparative Nosology

Corresponding diagnoses are included in the ICD-11 under disorders due to each substance separately, including alcohol; cannabis; synthetic cannabinoids; opioids; sedatives, hypnotics, or anxiolytics; cocaine, stimulants including amphetamines, methamphetamine, or methcathinone; synthetic cathinones; hallucinogens; volatile inhalants, 3,4-methylenedioxymethamphetamine (MDMA), and 3,4-methylenedioxyamphetamine (MDA); and dissociative drugs including ketamine and phencyclidine (PCP). In addition, categories are included for psychotic disorder induced by other specified psychoactive substance, psychotic disorder induced by multiple specified psychoactive substances, and psychotic disorder induced by unknown or unspecified psychoactive substance. Alcohol- and stimulant-induced psychotic disorders include the subcategories with hallucinations, delusions, and mixed psychotic symptoms.

Epidemiology

The prevalence of substance/medication-induced psychotic disorder likely varies according to the distribution of substance and medication use in the population. In the Lundby study, the prevalence of substance-induced psychotic disorders was about six times more common among males than females with a lifetime prevalence of 0.59% in a sample of over 3,500 participants in rural communities in Sweden who were followed over a period of 50 years. In a study of 544 young people presenting with a first-episode psychosis in Australia, 10.3% (N = 56) were diagnosed with a substance-induced psychotic disorder. More than a third (35.6%) of those initially diagnosed with a substance-induced psychosis had a change of diagnosis to a primary schizophrenia spectrum or bipolar disorder after a median of 84 weeks. The prevalence of substance/medication-induced psychotic disorder is expectedly higher among individuals with concurrent psychosis and substance use. In a study of 400 New York City emergency department patients with psychosis in conjunction with substance use, 44% were diagnosed as cases of substance-induced psychotic disorder. When these patients were re-evaluated a year later, 25% of cases initially diagnosed as substance-induced were rediagnosed as cases of psychotic disorders not related to substances.

The substances involved are likely different in different age groups. Substances of abuse may be a more common cause of substance/medication-induced psychotic disorder in the younger age group; whereas medications in the context of treatment for multiple health conditions are a more common cause in the older age group—a distribution that follows the pattern of substance and medication use in these age groups.

Etiology

A broad range of substances have been implicated in cases of substance/medication-induced psychotic disorder. Withdrawal from alcohol, sedative–hypnotics, and anxiolytic drugs has also been associated with psychotic symptoms.

A number of common medication groups are associated with psychotic symptoms, including nonsteroidal anti-inflammatory medications, other over-the-counter medications (e.g., phenylephrine and pseudoephedrine), corticosteroids, gastrointestinal medications, muscle relaxants antihistamines, antihypertensive and cardiovascular medications, antimicrobial medications, and chemotherapeutic agents (e.g., cyclosporine and procarbazine). The list includes medications commonly used in psychiatry and neurology, anticonvulsants, antiparkinsonian medications, and disulfiram. In addition, some toxins commonly used in agriculture or industry have been reported to induce psychotic symptoms, including organophosphate insecticides, carbon monoxide, carbon dioxide, and volatile substances such as fuel or paint.

Diagnostic and Clinical Features

Prominent delusions and/or hallucinations that are judged to be due to the physiologic effects of a substance/medication are the core features of the diagnosis of substance/medication-induced psychotic disorder. The DSM-5-TR states that if reality testing remains intact, meaning that a person experiences a distorted perception but recognizes them as false and does not act on them, the disorder would be diagnosed as substance intoxication or substance withdrawal with the accompanying specifier “with perceptual disturbances.” The manual also notes that the psychotic symptoms are required to have developed during or soon after substance intoxication or withdrawal, as identified by specifiers “with onset during intoxication” and “with onset during withdrawal.” The diagnosis is recorded along with the name of the substance and, in the case of substances of abuse, the presence of a substance disorder if appropriate (e.g., cocaine-induced psychotic disorder with onset during intoxication, with severe cocaine use disorders).

Differential Diagnosis

A major differential diagnosis is between substance/medication-induced psychosis and an independent psychotic disorder that is not substance induced. The DSM-5-TR provides a few guidelines for differential diagnosis of nonsubstance/medication-induced psychotic disorders, including temporal order of symptom onset, course of psychotic symptoms, and past history. If the psychotic symptoms preceded the onset of the substance/medication use or the symptoms persist for a substantial period of time (e.g., about 1 month) after the acute withdrawal or severe intoxication, then the psychotic symptoms are more likely to be related to a primary psychotic disorder. A history of a recurrent primary psychotic disorder also argues against substance/medication-induced psychosis. Nevertheless, information about the temporal order of symptoms and substance use and past history are not always clear. Furthermore, a past history of psychotic disorder unrelated to substances does not rule out a current diagnosis of substance/medication-induced psychotic disorder.

Specific features of the disorder, such as atypical age of onset (e.g., the first onset of delusions in a person older than 35 years without a known history of a primary psychotic disorder) or the nature of symptoms, may provide some guidance in differentiating substance/medication-induced psychotic disorder from primary psychotic disorders. Certain types of psychotic symptoms and accompanying symptoms are more common with certain substances. For example, the tactile hallucination of bugs crawling under the skin (formication) is more common in cocaine-induced psychosis, whereas cannabis-induced psychotic disorder developing shortly after high-dose cannabis use usually involves persecutory delusions, anxiety, emotional lability, and depersonalization experiences. Alcohol-induced psychotic disorder with hallucinations often presents with auditory hallucinations after prolonged and heavy use of alcohol. Evidence of substance use from the history and physical examination or laboratory tests for substances that can be detected using such tests may be helpful in solidifying a diagnosis.

Research comparing the characteristics of patients with substance-induced psychosis and patients with primary psychosis and concurrent substance abuse has produced few consistent results. In a study from Australia among patients with first-episode psychosis, those with substance-induced psychosis were more likely to be male, unemployed, and have a comorbid substance use disorder. Patients with substance-induced psychosis were not different from those with primary psychotic disorders with regard to duration of untreated psychosis, severity of psychotic symptoms, time to remission, or rates of relapse. A review of six studies comparing substance-induced psychotic disorders with primary psychotic disorders found that compared to those with primary psychotic disorders, patients with substance-induced psychosis had less family history of psychotic disorders, a greater degree of insight, fewer positive and negative symptoms, more depression (only in cannabis-induced psychotic disorders), and more anxiety.

Assessment of cognitive and mood symptoms is vital for differential diagnosis from delirium and mood disorders. Hallucinations are relatively common in the context of delirium associated with substance intoxication or withdrawal and do not justify a separate diagnosis of substance/medication-induced psychotic disorder. Similarly, delusions in the context of dementia or mild neurocognitive disorder should be coded as major or mild neurocognitive disorder, with behavioral disturbance.

The DSM-5-TR clearly states that if the individual recognizes that the hallucinations are the result of substance or medication use and does not act on them, then these experiences should not count toward a diagnosis of substance/medication-induced psychosis. Some individuals who are intoxicated with stimulants, hallucinogens, or PCP, or withdrawing from alcohol or sedatives may experience altered perceptions that they recognize as such. In these cases, the appropriate diagnoses are substance intoxication or substance withdrawal with perceptual disturbances.

Prolonged use of hallucinogens is sometimes associated with recurrent hallucinations long after stopping the use of these drugs. Reality testing remains intact, and the person is aware that these experiences are related to drug effects. Such “flashbacks” should be diagnosed as hallucinogen-persisting perception disorder.

A difficult differential diagnosis is to separate the effects of medications from the effects of underlying medical or mental disorder for which the medication is prescribed. The clinical history, course of illness, and the nature and course of treatments may provide clues to this differential diagnosis. At times medication substitution or discontinuation may be the only way to definitively establish a causal link between medication use and psychosis. If the disturbance is deemed to be related to both a medical condition and substance/medication use, both diagnoses should be made (e.g., psychotic disorder due to systemic lupus erythematosus and steroid-induced psychotic disorder).

Course and Prognosis

The course and prognosis of substance/medication-induced psychotic disorder is generally favorable, and psychosis typically remits after the removal of the offending agent. However, there have been reports of continuing psychotic symptoms for weeks or months after stopping stimulants, cocaine, or PCP. Treatment with antipsychotic medications may be needed in cases of severe substance/medication-induced psychotic disorder associated with agitation, suicidal/homicidal ideation, or other life-threatening behavior. Hospitalization may be needed for the safety of the patient or others in some of these cases.

PSYCHOTIC DISORDER DUE TO ANOTHER MEDICAL CONDITION

History

An association between mental and physical illness has been noted since antiquity. The concept of “sympathetic insanity” that became popular in the 19th century hypothesized an association in normal and abnormal functioning between all body organs that leads to severe mental illness among individuals with sicknesses in other organs. In DSM-I, these disorders were introduced under the rubric of “acute brain disorders,” which included acute brain disorders associated with infections, trauma, circulatory disturbances, metabolic disturbances, convulsive disorder, and neoplasms. However, the DSM-I did not distinguish between psychotic symptoms that occur in the context of delirium and those that occur in conjunction with clear sensorium. The DSM-III introduced two categories of “organic delusional disorder” and “organic hallucinosis” that presented with psychotic symptoms in the absence of clouding of consciousness, as in delirium. The DSM-IV and DSM-IV-TR combined the two categories into one category of “psychotic disorder due to…[the general medical condition] and introduced specifiers: “with delusions” and “with hallucinations” to identify the predominant symptoms.” This approach to the classification of these disorders was maintained in the DSM-5 and DSM-5-TR.

Comparative Nosology

Corresponding diagnoses are included in the ICD-11 secondary psychotic syndrome under the overall category of “Secondary mental or behavioral syndromes associated with disorders or diseases classified elsewhere.” Similar to DSM-5-TR, the secondary psychotic syndrome in ICD-11 includes specifiers “with hallucinations,” “with delusions,” as well as “with hallucinations and delusions.” The clinician is instructed to code the causing medical condition.

Epidemiology

There are few studies on the population prevalence of “psychotic disorder due to another medical condition.” The prevalence of this disorder likely varies across different communities and, over time, corresponds to the prevalence of the common underlying disorders.

In the Lundby study, the prevalence of psychotic disorders related to a general medical condition did not vary across genders and was 0.76% over the lifetime in a sample of over 3,500 participants in rural communities in Sweden who were followed over a period of 50 years. The prevalence is likely higher in the old age group, reflecting the prevalence of the underlying disorders. The nature of medical disorders giving rise to psychotic symptoms may also change across the lifespan, with epilepsy, head trauma, and autoimmune diseases being more common causes earlier in life and stroke, anoxic events, and cardiovascular diseases more common later in life. There are growing reports of psychotic episodes linked to Covid-19 infection. However, these reports are based on case reports and case series. Possible causal links need to be investigated in future epidemiologic studies.

Etiology

The list of medical conditions associated with psychotic symptoms is quite long, including neurologic conditions (e.g., head trauma, neoplasms, cerebrovascular disease, Huntington disease, multiple sclerosis, epilepsy, visual nerve injury, deafness, migraine, and CNS infections), endocrine disorders (e.g., hyper- and hypothyroidism, hyper- and hypoparathyroidism, hyper- and hypoadrenocorticism), metabolic conditions (e.g., hypoxia, hypercarbia, hypoglycemia), fluid or electrolyte imbalances, hepatic insufficiency, kidney failure, and autoimmune disorders with CNS involvement (e.g., systemic lupus erythematosus, anti-NMDA receptor antibody encephalitis). Among these conditions, endocrine, metabolic, and autoimmune disorders are the most common contributors to psychotic disorder due to another medical condition. Psychotic symptoms are also observed in 2% to 7% of epilepsy patients and are specifically associated with the postictal phase of temporal lobe epilepsy. Table 12.17–5 provides a partial listing of medical conditions associated with psychotic symptoms.

Diagnostic and Clinical Features

The essential feature of psychotic disorder due to another medical condition is the existence of prominent delusions or hallucinations that are judged to be directly caused by another medical condition as evidenced by clinical history, physical examination findings, or laboratory data including neuropsychological testing and imaging studies. The psychotic symptoms are not better explained by another mental disorder, such as a mood disorder with psychotic features or schizophrenia and do not occur only in the course of a delirium. The DSM-5-TR criteria also require that the psychotic symptoms be associated with significant distress or impairment in functioning. The DSM-5-TR allows the clinician to identify the predominant psychotic symptoms by using specifiers “with delusions” or “with hallucinations” and to specify the severity of psychotic symptoms.

Although hallucinations in psychotic disorder due to another medical condition can occur in any sensory modality, certain types of hallucinations are suggestive of specific etiologies. For example, olfactory hallucinations are suggestive of temporal lobe epilepsy. Furthermore, visual hallucinations and hallucinations in more than one sensory modality may be more common in psychotic disorder due to another medical condition than in schizophrenia. Psychotic disorder due to another medical condition is generally not diagnosed if reality testing is preserved and the individual recognizes that the hallucinations result from the medical condition. When delusions are due to another medical condition, they are most commonly persecutory. The association between the content of delusions and particular medical conditions appears to be less specific than is the case for hallucinations.

Table 12.17–5.
A Partial List of Medical Disorders that May Cause a Psychotic Disorder due to Another Medical Condition

Neurodegenerative disorders	Huntington disease, basal ganglia calcification, multiple sclerosis, metachromatic leukodystrophy
Other CNS disorders	Brain tumors, especially temporal lobe and deep hemispheric tumors; epilepsy, especially complex partial seizure disorder; head trauma, anoxic brain injury, migraine
Vascular disease	Atherosclerotic vascular disease, especially when associated with diffuse, temporoparietal, or subcortical lesions; hypertensive encephalopathy; subarachnoid hemorrhage; subdural hematoma; fat embolism
Infectious disease	HIV/AIDS, encephalitis lethargica, Creutzfeldt–Jakob disease, syphilis, malaria, acute viral encephalitis
Metabolic disorder	Hypercalcemia, hyponatremia, hypoglycemia, uremia, hepatic encephalopathy, pancreatic encephalopathy, porphyria
Endocrinopathies	Addison disease, Cushing disease, hyper- or hypothyroidism, hyper or hypoparathyroidism, panhypopituitarism
Vitamin deficiencies	Vitamin B_{12} deficiency, folate deficiency, thiamine deficiency, niacin deficiency
Autoimmune diseases	Systemic lupus erythematosus, temporal arteritis, anti-NMDA receptor antibody encephalitis

AIDS, acquired immunodeficiency syndrome; CNS, central nervous system; HIV, human immunodeficiency virus; NMDA, *N*-methyl-D-aspartate.

Psychotic disorder due to another medical condition is coded in DSM-5-TR by recording the name of the medical condition (e.g., psychotic disorder due to malignant lung neoplasm, with delusions).

Differential Diagnosis

Psychotic disorder due to another medical condition must be distinguished from delirium, which is common in many acute medical conditions and may be associated with hallucinations and delusions. A separate diagnosis of psychotic disorder due to another medical condition is not justified if the psychotic symptoms only occur in the context of delirium. Delusions in the context of a neurocognitive disorder (e.g., dementia) would be diagnosed as a neurocognitive disorder with behavioral disturbance.

Severe medical conditions can be extremely stressful, especially if they lead to serious disability or prolonged hospitalization. Distinguishing between a psychologically mediated brief psychotic disorder with marked stressor and a psychotic disorder due to another medical condition is especially difficult in the absence of other information regarding past history, including information about the course of medical illness and psychosis.

Many patients with medical conditions also use medications. If there is evidence of recent (within 4 weeks) use of medications or substances of abuse, a substance/medication-induced psychotic disorder should be considered as part of the differential diagnosis. If the disturbance is judged to be due to both a medical condition and substance use, both diagnoses (i.e., psychotic disorder due to another medical condition and substance/medication-induced psychotic disorder) can be given.

Psychotic disorder due to another medical condition must be distinguished from other psychotic disorders including schizophrenia, delusional disorder, schizoaffective disorder, and/or mood disorders with psychotic features. The presence of features such as later age at onset or presence of visual or olfactory hallucinations may suggest psychotic disorder due to another medical condition or a substance/medication-induced psychotic disorder.

Course and Outcome

The course and outcome of psychotic disorder due to another medical condition depend on the course and outcome of the underlying medical illness. With appropriate management of the underlying medical disorder, the psychotic symptoms generally remit. Most cases are transient and very rarely do the psychotic symptoms continue after remission of the underlying medical disease.

CATATONIA

History

Catatonia was originally described by the German psychiatrist Karl Kahlbaum in the 1870s as a distinct mental disorder with psychiatric, motor, and behavioral symptoms. Although originally, Kahlbaum considered catatonia to have an overall benign course, reports of chronic catatonia led later psychiatrists, including Kraepelin, to reconsider and view catatonia as a chronic disorder similar to hebephrenia. This view of catatonia contributed to Kraepelin’s proposal that catatonia be viewed as a subtype of dementia praecox. While Kraepelin recognized that catatonic symptoms could occur in mood and organic disorders, he distinguished between catatonic dementia praecox and catatonic symptoms occurring in other disorders.

This view of catatonia as a distinct subtype of schizophrenia dominated modern classification systems, including the DSM and ICD. Multiple editions of DSM through DSM-IV-TR and ICD-10 discussed catatonia as a subtype of schizophrenia. DSM-III-R introduced the catatonic features specifier for mood disorders when the clinical picture of a mood disorder episode is dominated by two or more catatonic symptoms. The category of catatonic disorder due to a general medical condition was introduced in DSM-IV. Furthermore, the ICD-10 recognized a category of organic catatonic disorder in addition to catatonic schizophrenia.

In a clear break from the Kraepelinian conception of catatonia as a subtype of schizophrenia, DSM-5 and ICD-11 introduced a separate rubric for catatonia. This change was motivated by evidence regarding the occurrence of catatonia in a number of other psychiatric and medical disorders besides schizophrenia and mood disorders and evidence of the underrecognition of catatonia in current psychiatric practice. DSM-5 and DSM-5-TR recognize three types of catatonia: (1) catatonia associated with another mental disorder, (2) catatonic disorder due to another medical condition, and (3) unspecified catatonia. Similarly, ICD-11 identifies catatonia associated with another mental disorder, catatonia induced by substances or medications, and an unspecified category of catatonia.

Epidemiology

Epidemiologic research on the prevalence, incidence, and correlates of catatonia has been hampered by differences in the definition and method of assessment of catatonia. A review of 20th-century studies found that catatonia was reported in 7% to 17% of patients hospitalized with acute psychotic disorders and 13% to 31% of patients with mood disorders. A more recent review of 74 prevalence studies in clinical samples conducted between 1935 and 2017 reported that catatonia was present in 9% of patients. There is some evidence that the prevalence of the catatonic type of schizophrenia may have declined over the 20th century and that catatonia may be more common in developing countries than in industrialized countries. However, overall, the prevalence of catatonia syndrome does not appear to have changed markedly over time. Catatonia may be seen in individuals of all ages, including children and adolescents. Gender and age distributions of patients with catatonia likely reflect the epidemiology of the underlying disorder.

Etiology

The etiology of catatonia is unknown. The long and growing list of medical conditions associated with catatonia underscores the potential heterogeneity in the etiology of catatonia. The list includes major neurologic disorders such as epilepsy, encephalitis, and trauma; endocrine disorders including thyroid disease; metabolic disturbances such as uremia and hypercalcemia; paraneoplastic disorders; and autoimmune diseases. Catatonia can also occur in the context of a number of mental disorders, including psychotic and mood disorders, and neurodevelopmental disorders, including autism spectrum disorder.

Imaging studies among patients with hypokinetic catatonia in the context of psychotic disorders indicate increased activity in premotor areas which may be due to corticocortical inhibition or increased activity of inhibitory corticobasal projections. The favorable response of the disorder to GABAergic treatments (e.g., benzodiazepines) points to a role of the GABA system in catatonia. There is also a growing interest in the role of the NMDA receptors in catatonia based on numerous case reports associating anti-NMDA receptor antibody encephalitis with catatonic symptoms.

Diagnostic and Clinical Features

The essential feature of catatonia is a psychomotor disturbance that may involve decreased engagement with the interviewer during clinical interview or physical examination, decreased motor activity, or excessive and peculiar motor activity. The psychomotor disturbance may range from marked unresponsiveness to environmental stimuli (stupor) to marked agitation. In some cases, the individual's symptoms may quickly switch between decreased motor activity and excessive motor activity, making the diagnosis difficult. Some of the peculiar motor and verbal symptoms such as catalepsy (holding a posture induced by the examiner), posturing (active adoption of often awkward postures by the patient), waxy flexibility (slight resistance when the examiner tries to position the patient's body parts), mannerism (odd or caricature-like mimicking of normal actions), grimacing, echolalia, and echopraxia (mimicking another person's speech or behavior) are more pathognomonic and helpful for the diagnosis of a catatonic syndrome. Self-inflicted injuries and harm to others can also occur, particularly during periods of increased activity.

The DSM-5-TR diagnosis of catatonia requires that the clinical picture be dominated by at least 3 of the 12 characteristic symptoms, including stupor, catalepsy, waxy flexibility, mutism, negativism, posturing, mannerism, stereotypy, agitation, grimacing, echolalia, and echopraxia. The diagnosis of catatonia associated with another medical condition also requires evidence from clinical history, physical examination, or laboratory examinations findings (including imaging studies) that the catatonic symptoms are caused by another medical condition, that they are not better explained by another mental disorder and do not occur exclusively during the course of delirium, and that the symptoms cause significant distress or impairment in social or occupational functioning.

In coding catatonia associated with another mental disorder or medical disorder, the name of the associated mental or medical disorder is recorded first, followed by the specifiers “catatonia associated with…” in the case of mental disorders (e.g., “major depressive disorder; catatonia associated with major depressive disorder”) or “catatonic disorder due to…” in the case of medical conditions (e.g., “hypothyroidism; catatonic disorder due to hypothyroidism”).

The residual category of unspecified catatonia is reserved for situations in which the catatonic symptoms are present and cause clinically significant distress or impairment. However, either the number of symptoms does not reach the minimum required of three symptoms, the underlying medical or mental disorder is not clear, or there is insufficient information regarding the underlying cause (e.g., in the emergency department settings).

Differential Diagnosis

Movement disturbances are common during the course of delirium or neuroleptic malignant syndrome, and a separate diagnosis of catatonic disorder due to another medical condition is not justified in these cases. Catatonia-like symptoms are not uncommon among patients treated with antipsychotic medication (e.g., neuroleptic-induced acute dystonia). History of recent exposure to medications, especially first-generation antipsychotic medications, may help in distinguishing these diagnostic possibilities. Generalized rigidity is a cardinal feature of neuroleptic malignant syndrome, in which significant increases in creatine kinase levels are common. Autonomic instability and increases in temperature can be seen in individuals with catatonia, but these findings are typically more severe in neuroleptic malignant syndrome and can be potentially life-threatening.

Course and Prognosis

Catatonic symptoms sometimes remit with the management of the underlying psychiatric or medical condition. However, many individuals will also require specific treatment of the catatonic syndrome and early identification of catatonia is essential. Although a rare type of remitting and relapsing catatonia, known as “periodic catatonia” has been described in classic European literature, the majority of cases of catatonia do not have a recurring course.

Catatonic symptoms are especially responsive to treatment with high doses of benzodiazepines and electroconvulsive therapy (ECT), which may be indicated as the main or adjunctive treatment in the acute management of psychiatric patients presenting with severe catatonic symptoms. On the other hand, outcomes may be worsened by use of antipsychotic medications, particularly high-potency first-generation antipsychotic medications. During severe stages of immobility or increased activity, the individual may need careful supervision and intervention to prevent dehydration, malnutrition, exhaustion, hyperpyrexia, aspiration, embolic events, decubiti, or contractures of extremities.

OTHER SPECIFIED SCHIZOPHRENIA SPECTRUM AND OTHER PSYCHOTIC DISORDER AND

UNSPECIFIED SCHIZOPHRENIA SPECTRUM AND OTHER PSYCHOTIC DISORDER

History

The concept of “unspecified psychosis” was introduced in DSM-II but was specifically noted to be “for librarians and statisticians to use in coding incomplete diagnoses” rather than for use by clinicians. However, in DSM-III, the diagnosis of “atypical psychosis” was introduced as “a residual category for cases in which there are psychotic symptoms … that do not meet the criteria for any specific mental disorder.” DSM-III-R changed the name of the diagnosis to psychotic disorder NOS (atypical psychosis) and eliminated mention of the diagnosis as a “residual category.” Instead, it emphasized the use of the diagnosis for “Disorders in which there are psychotic symptoms (delusions, hallucinations, incoherence, marked loosening of associations, catatonic excitement or stupor, or grossly disorganized behavior) that do not meet the criteria for any other nonorganic psychotic disorder.” It also noted that the diagnosis of psychotic disorder NOS was preferable to deferring a diagnosis in individuals for whom there was inadequate information to make a specific diagnosis.

The concept of psychotic disorder NOS persisted in DSM-IV, although the references to atypical psychosis and nonorganic psychotic disorders were dropped. As in prior editions of DSM, there were no specific diagnostic criteria for this category. However, it was now defined as including psychotic symptoms “about which there is inadequate information to make a specific diagnosis or about which there is contradictory information, or disorders with psychotic symptoms that do not meet the criteria for any specific Psychotic Disorder.”

The DSM-5 divided the DSM-IV category of psychotic disorder NOS into two distinct categories of “other specified schizophrenia spectrum and other psychotic disorder” and “unspecified schizophrenia spectrum and other psychotic disorder” to underscore the reason why the disorder could not be classified as one of the specific schizophrenia spectrum disorders. DSM-5-TR maintained this approach.

While in the other specified schizophrenia spectrum and other psychotic disorders the characteristics of the condition clearly do not meet the criteria for a specific disorder, in the unspecified schizophrenia spectrum and other psychotic disorder there is inadequate information to make a more specific diagnosis. The latter diagnosis may thus be appropriate for acute care and emergency department settings where there is insufficient information available. Similar categories of “other specified primary psychotic disorder” and “schizophrenia or other primary psychotic disorder, unspecified” were included in ICD-11.

DSM-5-TR provided four examples of other specified schizophrenia spectrum and other psychotic disorder: (1) persistent auditory hallucinations, (2) delusions with significant overlapping mood episodes, (3) attenuated psychosis syndrome, and (4) delusional symptoms in a partner of an individual with delusional disorder. The last example, variously called “shared delusional disorder,” “shared psychotic disorder,” “induced psychosis,” and “folie à deux,” was discussed as a specific psychotic disorder in the previous edition of DSM and in ICD-10. The new category of attenuated psychosis syndrome was a new category in the DSM-5, and its introduction was associated with some controversy. These two categories are discussed in some detail later in the chapter.

Epidemiology

There is minimal information about the prevalence of other specified schizophrenia spectrum and other psychotic disorder and unspecified schizophrenia spectrum and other psychotic disorder in the community. In clinical practice, the psychotic disorder NOS diagnosis has been often used in first-admission or emergency psychiatric patients when there is inadequate information to make a diagnosis.

Etiology

Given the heterogeneity of the other specified and unspecified schizophrenia spectrum and other psychotic disorder diagnoses, it is not surprising that there has been minimal investigation into the etiologies of these disorders. However, one family study has examined the occurrence of other psychotic illnesses in relatives of probands with psychotic disorder NOS according to prior editions of the DSM and found a substantially elevated risk of schizophrenia and schizophrenia spectrum disorders in these family members but not mood disorders.

Diagnostic and Clinical Features

The diagnostic category of other specified schizophrenia spectrum and other psychotic disorder includes cases with psychotic symptoms and clinically significant distress or impairment that do not conform to diagnostic criteria for specific schizophrenia spectrum and other psychotic disorders. The specific presentation can be recorded as well. For example, for a patient who presents to the hospital with persistent auditory hallucinations that are not accompanied by mood disturbance and that are not consistent with symptom criteria for schizophrenia or any other DSM-5-TR category, a diagnosis of “other specified schizophrenia spectrum and other psychotic disorder, persistent auditory hallucinations” may be recorded. Other examples include psychotic symptoms that are below the threshold for a full psychotic episode (attenuated psychosis syndrome), delusional beliefs in partners of individuals with a primary delusional disorder, and persistent delusions with periods of overlapping mood symptoms that have been present for a substantial portion of the delusional disturbance.

A diagnosis of unspecified schizophrenia spectrum and other psychotic disorder may be appropriate for a patient presenting with psychotic symptoms causing clinically significant distress or impairment but where information regarding symptoms and history is not sufficient to make a diagnosis of a specific schizophrenia spectrum and other psychotic disorder. For example, the clinician may not be able to determine whether the psychosis is primary, due to a general medical condition, or the result of a substance or medication use. Such distinctions can be difficult to make, especially when adequate clinical history is not available (e.g., in the emergency department settings).

Differential Diagnosis

Because other specified and unspecified schizophrenia spectrum and other psychotic disorder are diagnoses of exclusion, the differential diagnosis includes other possible diagnoses with psychotic symptoms. Co-occurring substance misuse is one of the most frequent confounding factors in establishing a specific diagnosis, because substance use can be found in more than half of hospitalized patients with psychotic symptoms. Furthermore, the possibility that a general medical condition may be causing the psychotic symptoms must always be kept in mind.

If mood symptoms are present, the differential diagnosis will include major depressive disorder with psychotic features or bipolar disorder with psychotic features, depending on the type of mood change that is present. Schizoaffective disorder is also possible if the temporal requirements for the psychotic symptoms and mood changes are met.

If no mood symptoms are present and substance, medication, and medical disease can be ruled out as causes of psychotic symptoms, the differential diagnosis will include schizophrenia, schizophreniform disorder, brief psychotic disorder, and delusional disorder, depending on the duration and type of psychotic symptoms.

Course and Prognosis

The course and prognosis are highly variable as other specified and unspecified schizophrenia spectrum and other psychotic disorder are heterogeneous. Furthermore, the diagnosis of unspecified schizophrenia spectrum and other psychotic disorder is often a provisional diagnosis by design that would be changed upon gathering more information. In the first-admission study in Suffolk County, New York, 12% of individuals were diagnosed by the research team with psychotic disorder NOS at the time of their first admission, but only 4.8% received this diagnosis by the 6-month follow-up, and by 24-month follow-up, the rate was 3.9%. Of 25 individuals with a diagnosis of psychotic disorder NOS at 6-month follow-up, only 11 maintained this diagnosis by 24-month follow-up, a stability rate of 44%. Eight of the 25 subjects with psychotic disorder NOS had met the criteria for schizophrenia or schizoaffective disorder.

ATTENUATED PSYCHOSIS SYNDROME

The introduction of the new category of attenuated psychosis syndrome in the DSM-5 was subject to some controversy. While prodromal symptoms of schizophrenia have long been recognized by psychiatric investigators, earlier classifications did not include a specific category for these symptoms. Introduction of these symptoms in the diagnostic system was motivated by efforts over the past two decades to improve early detection of schizophrenia and to implement early treatment with the aim of reducing chronicity and long-term disability often associated with schizophrenia. In recognition of this aim, the category was initially labeled the “psychosis risk syndrome.” The name change to attenuated psychosis syndrome was intended to highlight the importance of current symptoms and the associated distress and disability as a focus of clinical attention rather than the future risk of a full-blown psychotic disorder.

The DSM-5-TR provides few guidelines and no discrete criteria for the diagnosis of attenuated psychosis syndrome. The manual mentions symptoms that are of lower severity are more transient or are associated with relatively intact insight as examples of attenuated psychosis syndrome. Although the definition specifies “psychotic-like symptoms that are below a threshold for full psychosis,” no cutoffs for the severity or duration of symptoms are given. Like other conditions included under other specified schizophrenia spectrum and other psychotic disorders, attenuated psychosis syndrome is associated with “clinically significant distress or impairment in social, occupational, or other important areas of functioning."

In a study from Switzerland, prevalence of attenuated psychosis syndrome in adolescents and young adults aged 15 to 40 years was estimated at 0.3%. Attenuated psychotic syndrome is associated with comorbid depressive disorder, suicidality, and functional impairment. The syndrome may be more common in males than females.

Etiology of attenuated psychosis syndrome remains unknown. Genetic risk, such as history of psychotic disorders in first-degree relatives, is a criterion in some alternative definitions of attenuated psychosis risk. Structural imaging studies of attenuated psychosis syndrome cases have produced mixed results. Some functional imaging studies have suggested abnormalities in dopamine and glutamate systems.

The future risk of a full-blown psychotic disorder among individuals with attenuated psychosis syndrome has been a major focus of research. In one meta-analysis of short-term studies, less than 40% developed a full-blown psychotic illness in a 3-year period. However, the risk of transition remained elevated beyond 3 years and was as high as 65% to 79% in a long-term study. In a more recent meta-analysis of 23 cohort studies, the risk of transition was 11% at 6 months, 15% at 12 months, 20% at 24 months, and 23% at 36 months.

There are few guidelines regarding the management of attenuated psychosis. Given that the majority of individuals with attenuated psychosis syndrome would not experience a full-blown psychotic illness in near future, treatment with antipsychotic medication may not be indicated for self-limited symptoms. In many cases, careful assessment and frequent monitoring may be appropriate.

DELUSIONAL SYMPTOMS IN PARTNER OF INDIVIDUAL WITH DELUSIONAL DISORDER

The category of “delusional symptoms in partner of individuals with delusional disorder” was first described in 1860 by Jules Baillarger, who called the syndrome “folie à communiquée,” and again described in 1877 by Charles Lasègue and Jean Pierre Falret who gave the syndrome the commonly used name, “folie à deux.” Over the years, many other names including shared psychotic disorder, shared delusional disorder, and induced psychotic disorder have been used to describe this syndrome in which two individuals with a close and generally long-term relationship share the same delusional belief. Shared paranoid disorder was introduced in DSM-III in the section on paranoid disorders and the persecutory nature of the shared delusions was emphasized. In DSM-III-R and later, DSM-IV and DSM-IV-TR, the content of the shared delusion was broadened to include all types of delusional beliefs. DSM-5-TR included this condition in the category of other specified schizophrenia spectrum and other psychotic disorder. Due to the rarity of cases that present to clinical settings, there is no information about the prevalence of delusional symptoms in the partners of individuals with delusional disorder.

Etiologic theories regarding delusional symptoms in the partner of an individual with delusional disorder are primarily based on family and interpersonal dynamics. In general, individuals who develop a shared delusion will tend to have had a submissive role in the relationship with the partner who was first to develop a delusional disorder. The person who develops the shared delusional symptoms may be particularly impressionable due to personality factors or intellectual limitations. Social isolation, poverty, language difficulties, and geographical isolation may also contribute to the development of the condition. In some cases, the shared delusion may involve more than two individuals and even the entire family unit. Persecutory delusional beliefs are most commonly seen in shared psychotic disorder, comprising about 70% of patients in one study; however, religious, grandiose, and somatic delusions may also be observed. Because of the relative rarity of shared delusional disorder, its clinical course and prognosis are unclear, although some reports of successful treatment are present in the literature.

FURTHER READINGS

Allebone J, Kanaan R, Wilson SJ. Systematic review of structural and functional brain alterations in psychosis of epilepsy. *J Neurol Neurosurg Psychiatry*. 2018;89:611.

Bogren M, Mattisson C, Isberg PE, Nettelbladt P. How common are psychotic and bipolar disorders? A 50-year follow-up of the Lundby population. *Nord J Psychiatry*. 2009;63:336.

Bromet EJ, Kotov R, Fochtmann LJ, et al. Diagnostic shifts during the decade following first admission for psychosis. *Am J Psychiatry*. 2011;168:1186.

Castagnini AC, Fusar-Poli P. Diagnostic validity of ICD-10 acute and transient psychotic disorders and DSM-5 brief psychotic disorder. *Eur Psychiatry*. 2017;45:104.

Caton CL, Drake RE, Hasin DS, et al. Differences between early-phase primary psychotic disorders with concurrent substance use and substance-induced psychoses. *Arch Gen Psychiatry*. 2005;62:137.

*Cheniaux E, Landeira-Fernandez J, Lessa Telles L, et al. Does schizoaffective disorder really exist? A systematic review of the studies that compared schizoaffective disorder with schizophrenia or mood disorders. *J Affect Disord*. 2008;106:209.

de Pablo GS, Catalan A, Fusar-Poli P. Clinical validity of DSM-5 attenuated psychosis syndrome: advances in diagnosis, prognosis, and treatment. *JAMA Psychiatry*. 2020;77:311–320.

Fennig S, Craig TJ, Bromet EJ. The consistency of DSM-III-R delusional disorder in a first-admission sample. *Psychopathology*. 1996;29:315.

Fiorentini A, Volonteri LS, Dragogna F, et al. Substance-induced psychoses: a critical review of the literature. *Curr Drug Abuse Rev*. 2011;4:228.

Fusar-Poli P, Cappucciati M, Bonoldi I, et al. Prognosis of brief psychotic episodes: a meta-analysis. *JAMA Psychiatry*. 2016;73:211–220.

Hensel JM, Chartier MJ, Ekuma O, et al. Risk and associated factors for a future schizophrenia diagnosis after an index diagnosis of unspecified psychotic disorder: a population-based study. *J Psychiatr Res*. 2019;114:105.

Huber M, Wolf RC, Lepping P, et al. Regional gray matter volume and structural network strength in somatic vs. non-somatic delusional disorders. *Prog Neuropsychopharmacol Biol Psychiatry*. 2018;82:115.

*Kendler KS. The clinical features of paranoia in the 20th century and their representation in diagnostic criteria from DSM-III through DSM-5. *Schizophr Bull*. 2017;43:332–343.

Kotov R, Leong SH, Mojtabai R, et al. Boundaries of schizoaffective disorder: revisiting Kraepelin. *JAMA Psychiatry*. 2013;70:1276.

Mahadevan J, Sundaresh A, Rajkumar RP, et al. An exploratory study of immune markers in acute and transient psychosis. *Asian J Psychiatr*. 2017;25:219.

*Marneros A, Pillman F. *Acute and Transient Psychoses*. Cambridge University Press; 2004.

McHugo M, Armstrong K, Roeske MJ, Woodward ND, Blackford JU, Heckers S. Hippocampal volume in early psychosis: a 2-year longitudinal study. *Transl Psychiatry*. 2020;10:306.

Naz B, Bromet EJ, Mojtabai R. Distinguishing between first-admission schizophreniform disorder and schizophrenia. *Schizophr Res*. 2003;62:51.

Nykiel SA, Baldessarini RJ, Bower MC, Goodwin J, Salvatore P. Psychosis NOS: search for diagnostic clarity. *Harv Rev Psychiatry*. 2008;16:55.

O'Connell J, Sunwoo M, McGorry P, O'Donoghue B. Characteristics and outcomes of young people with substance induced psychotic disorder. *Schizophr Res*. 2019;206:257.

Pollak TA, Lennox BR, Müller S, et al. Autoimmune psychosis: an international consensus on an approach to the diagnosis and management of psychosis of suspected autoimmune origin. *Lancet Psychiatry*. 2020;7:93.

Rutigliano G, Merlino S, Minichino A, et al. Long term outcomes of acute and transient psychotic disorders: the missed opportunity of preventive interventions. *Eur Psychiatry*. 2018;52:126.

Santelmann H, Franklin J, Bußhoff J, Baethge C. Interrater reliability of schizoaffective disorder compared with schizophrenia, bipolar disorder, and unipolar depression—a systematic review and meta-analysis. *Schizophr Res*. 2016;176:357–363.

*Schwartz JE, Fennig S, Tanenberg-Karant M, et al. Congruence of diagnoses 2 years after a first-admission diagnosis of psychosis. *Arch Gen Psychiatry*. 2000;57:593.

Solmi M, Pigato GG, Roiter B, et al. Prevalence of catatonia and its moderators in clinical samples: results from a meta-analysis and meta-regression analysis. *Schizophr Bull*. 2018;44:1133.

Susser E, Wanderling J. Epidemiology of nonaffective acute remitting psychosis vs schizophrenia: sex and sociocultural setting. *Arch Gen Psychiatry*. 1994;51:294.

Tondo L, Vázquez GH, Baethge C, et al. Comparison of psychotic bipolar disorder, schizoaffective disorder, and schizophrenia: an international, multisite study. *Acta Psychiatr Scand*. 2016;133:34–43.

Walther S, Stegmayer K, Wilson JE, Heckers S. Structure and neural mechanisms of catatonia. *Lancet Psychiatry*. 2019;6:610.

Wilson L, Szigeti A, Kearney A, Clarke M. Clinical characteristics of primary psychotic disorders with concurrent substance abuse and substance-induced psychotic disorders: a systematic review. *Schizophr Res*. 2018;197:78–86.

Wolf RC, Hildebrandt V, Schmitgen MM, et al. Aberrant gray matter volume and cortical surface in paranoid-type delusional disorder. *Neuropsychobiology*. 2020;79:335.

13

Mood Disorders

13.1 Mood Disorders: Historical Introduction and Conceptual Overview

Robert J. Boland, M.D.

CLINICAL AND PUBLIC HEALTH SCOPE OF MOOD DISORDERS

Epidemiology

Research on the epidemiology of depression, both in the United States and worldwide, has considerably broadened our understanding of the disorder. We have discovered that depression is much more common than once thought. It is also one of the more debilitating disorders known to humanity, as it often affects individuals during what should be their most productive years. Hence, the depressive disorders are one of the most costly disorders in society.

Measuring the true incidence of major depressive disorder (MDD) is complicated and highly dependent on the methods used to collect the sample, definitions of the disorder, and the particular instruments used to measure depression. For example, studies that use self-assessment tend to report higher incidences than studies using clinician-rated instruments. The timing of a study likely affects the results as well. As depressive illness become less stigmatized, persons may be more willing to endorse symptoms of the disorder to the well-meaning strangers who knock at their door, call their phone, or email them and identify themselves as researchers.

In a reasonably recent meta-analysis, which included 90 studies from 30 countries done from 1994 to 2014, with combined data on more than 1 million participants, the point prevalence for depression was 12.9%, the 1-year prevalence was 7.2%, and the lifetime prevalence was 10.8%. There was significant heterogeneity among the different studies.

In the United States, the most recent epidemiologic data is from the National Survey on Drug Use and Health (NSDUH), sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA). Using a definition of a major depressive episode based on *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), they found the overall 1-year prevalence of a major depressive episode to be 7.1%; hence, very similar to the international data.

Concepts of Mood Disorders

The broader rubric of affective disorder (which subsumes mood and anxiety disorders) has been conceptualized along two influential schools in the European tradition. Aubrey Lewis and his followers from the Maudsley school have promoted a continuum model—from anxiety disorders to mild neurotic depressions to severe endogenous and psychotic depressions. The Newcastle school, led by Martin Roth, has sharply demarcated those conditions from one another. Although vestiges of both approaches are still influential in clinical and basic research, European studies in German-speaking countries that subdivide mood disorders based on polarity now overshadow that approach. That sharp demarcation, in part supported by studies in the United States, has served as the basis for much recent research into the biology, treatment, and classification of mood disorders and is reflected in the fifth edition of the DSM-5 and the tenth revision and eleventh revisions of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10 and 11). Despite such official sanctions, many authorities continue to see considerable continuity between recurrent depressive and bipolar disorders. This inconsistency has led to widespread discussion and debate about the bipolar spectrum (BPS), which incorporates classic bipolar disorder, bipolar II, and recurrent depressions.

Childhood bipolarity, too, is receiving increasing clinical attention, thanks to the seminal work of Elizabeth Weller and colleagues, conducted initially at Ohio State University. In addition, clinical observations at the University of Tennessee on the juvenile offspring of adult patients with bipolar disorder have led to a greater appreciation of the bipolar nature of complex clinical presentations of affective illness in the juvenile offspring and kin of adult bipolar probands. More recent work by Biederman's group at Harvard suggests intriguing links between pediatric bipolar disorder and attention-deficit/hyperactivity disorder (ADHD).

Current concepts of mood disorders in the United States embrace a broad spectrum, including many conditions previously diagnosed as schizophrenia, personality disorder, or neurosis. The diagnostic shift occurred in part as a result of the U.S.–U.K. Diagnostic Project, which demonstrated that we often diagnosed schizophrenia at the expense of mood disorders. The availability of new and effective treatments and the unacceptable risk for tardive dyskinesia and suicide in persons with misdiagnosed mood disorders further broadened the conceptual boundaries. More generally, current research interest in mood disorders in the United States emanated from a landmark 1969 NIMH conference on the psychobiology of affective illnesses. The NIMH Collaborative Depression Study—a long-term prospective project deriving directly from recommendations made at the conference—has legitimatized the broader perspective.

Morbidity and Mortality

Unfortunately, findings published by Martin Keller and colleagues in the 1980s documenting gross undertreatment of mood disorders continue to describe the current treatment landscape worldwide. Whatever changes have occurred in diagnostic practice do not appear to have significantly affected the morbidity and mortality of mood disorders. This depressing news is all the more concerning given that the 1990s saw new classes of user-friendly antidepressant and mood-stabilizing agents, as well as depression-specific psychotherapies. Thus, the state of affairs results, in part, from the fact that clinical exposure to mood disorders in specialized (psychiatric) and primary care training is suboptimal. As highly prevalent and lethal disorders, mood disorders must command a more significant share in the clinical curriculum of both psychiatrists and general medical practitioners. Because most mood disorders are chronic and relapsing, long-term exposure to patients with these disorders in mood or bipolar clinics should be part of a young doctor's core knowledge. Unfortunately, few academic centers have appropriate clinics for this training, and those with the resources often devote them primarily to research. The primary goal of these clinics is the execution of research protocols rather than the acquisition of clinical experience in caring for such patients.

Nearly half of all cases of depression, just like those with adult-onset diabetes, remain undetected for years or inadequately controlled—both of which seem to lag behind hypertension, in which early detection and treatment have significantly reduced complications such as stroke. Efforts by patient-advocacy organizations—often in concert with national psychiatric organizations and governmental mental health agencies—appear to be increasing public and government awareness of mood disorders. Ultimately, however, the challenge is to provide all primary care physicians with the requisite hands-on experience in this prevalent group of disorders. User-friendly tools to detect suicidal patients in the general medical sector would further enhance preventive efforts. These would require significant changes in health care structure, including, but not limited to, the greater participation of nurses and social workers as liaisons in primary mental health and continuity of care for depressed or suicidal patients.

Because mood disorders underlie 50% to 70% of all suicides, effective treatment of these disorders on a national level should, in principle, drastically reduce this significant complication of mood disorders. That depressed older adult patients, often with medical comorbidity, constitute the highest-risk group for suicide yet escape clinical detection and treatment is particularly problematic for public health. In addition, clinical findings in recurrent mood disorders have shown the value of lithium prophylaxis for preventing suicide and overall mortality. Emerging data suggest that such benefits may accrue from all efficacious treatments for mood disorders. Suicide prevention at the national level, attributed to the widespread use of selective serotonin reuptake inhibitors (SSRIs) and primary care physician education has been achieved in many but not all European countries. Such prevention does not have a tradition in the United States, and, as mentioned earlier, suicide rates have increased in the United States.

DEFINITIONS

Mood disorders encompass many psychiatric disorders in which pathologic moods and related vegetative and psychomotor disturbances dominate the clinical picture. Known in previous editions of DSM as *affective disorders*, we prefer *mood disorders* because it refers to sustained emotional states, not merely to the present emotional state's external (affective) expression. Mood disorders are more syndromes than discrete diseases, with a cluster of signs and symptoms sustained over weeks to months. The symptoms represent a marked departure from a person's habitual functioning and tend to recur, often periodically or cyclically.

Major Depressive Disorder and Bipolar Disorder

MDD (unipolar depression) is the most common mood disorder. It may manifest as a single episode or as recurrent episodes. The course may be somewhat protracted—up to 2 years or longer—in those with the single-episode form. Whereas the prognosis for recovery from an acute episode is good for most patients with MDD, three of four patients experience recurrences throughout life and have varying degrees of residual symptoms between episodes. *Bipolar disorders* (previously called *manic-depressive psychosis*) consist of at least one hypomanic, manic, or mixed episode. *Mixed episodes* represent a simultaneous mixture of depressive and manic or hypomanic manifestations. Although a minority of patients experience only manic episodes, most bipolar disorder patients experience episodes of both polarities. Manias predominate in men, depression, and mixed states in women. Bipolar II disorder, which is not always easily discernible from recurrent MDD, illustrates the need for more research to elucidate the relationship between bipolar disorder and MDD.

The border with schizophrenia continues to be fuzzy. Modern research challenges the Kraepelinian dichotomy between dementia praecox and manic-depressive illness in favor of a spectrum concept that goes beyond mood-incongruent mood episodes and might spill over into "good-prognosis" schizophrenia.

Dysthymia and Cyclothymia

Clinically, major depressive episodes often arise from a low-grade, intermittent, and protracted depressive substrate known as *dysthymic disorder* (called persistent depressive disorder in DSM-5, but the terms are interchangeable). Likewise, many instances of bipolar disorders, especially ambulatory forms, represent episodes of mood disorder superimposed on a cyclothymic background, which is a biphasic alternating pattern of numerous brief periods of hypomania and numerous brief periods of depression. Dysthymic and cyclothymic disorders represent the two prevalent subthreshold mood conditions roughly corresponding to the basic temperamental dysregulations described by Emil Kraepelin and Ernst Kretschmer as predisposing to affective illness.

It is not always easy to demarcate full-blown syndromal episodes of depression and mania from their subthreshold counterparts commonly observed during the interepisodic periods. The subthreshold conditions appear to be fertile terrain for interpersonal conflicts and postaffective pathologic character developments that may ravage the lives of patients and their families. In North America and some Western European countries, clinicians often label such patients as a *borderline personality disorder*, which, unfortunately, often tends to obscure the affective origin of the presenting psychopathology.

Cyclothymic and dysthymic conditions also exist in the community without progression to full-blown mood episodes. As such, they are best considered, respectively, as trait bipolar and trait depressive conditions. Understanding the factors that mediate the transition from trait to clinical state is essential for preventing manic and major depressive episodes.

A fascinating development related to these temperaments pertains to the “positive” attributes of the depressive and related temperaments in subserving sensitivity to human suffering, thereby playing a role in altruistic behavior. The more tempestuous cyclothymic, based on recent data from different centers, appears to be the temperament involved in poetry, the arts, architecture, and other forms of creative human endeavor. In other words, the “dilute” forms of affective illness may have evolved as traits of fundamental importance for human culture.

Other Subthreshold Mood States

Epidemiologic studies in Europe and North America have also revealed other subsyndromal conditions with depressive and hypomanic manifestations with few symptoms (oligosymptomatic mood states) and short duration (brief episodes). Variously referred to as *minor*, *subsyndromal*, *brief*, or *intermittent*, these descriptions do not merely represent an arbitrary lowering of diagnostic thresholds. However, they also herald increasing realization of their importance in the early detection of at-risk individuals, as has happened in other medical fields (e.g., diabetes mellitus and essential hypertension). If disabling mood disorders afflict 5% to 8% of the general population (Epidemiologic catchment area [ECA] study), milder but still clinically significant mood disorders would raise lifetime rates to 17% (National Comorbidity Study [NCS]). If we add subclinical mood states, the figure doubles to involve one-third of the general population (as reported, e.g., by Kenneth Kendler and colleagues). New evidence from both Europe and the United States has shown that BPS conditions (bipolar I, bipolar II, and unspecified bipolar disorder in diagnostic manuals such as the DSM-5 and the ICD-10) may account for at least 50% of all mood disorders in the community and psychiatric practice.

Comorbidity in mood disorders involves considerable overlap with anxiety disorders. As summarized in an NIMH monograph, anxiety disorders can occur during an episode of depression, perhaps as a precursor to the depressive episode, and, less commonly, during the future course of a mood disorder. Those findings suggest that at least some depressive disorders share a common diathesis with certain anxiety disorders. More recent clinical experience suggests intriguing comorbidity patterns between bipolar II disorder on one hand and panic, obsessive-compulsive, and social phobic states on the other.

Furthermore, bipolar I and II disorders are particularly likely to be complicated by alcohol, stimulants, or both. In some cases, the alcohol or substance abuse represents attempts at “self-treatment” of the depression and associated anxiety or insomnia (or both) and attempts to maintain or enhance the positive moods in the case of mania and hypomania energy. Finally, physical illness—both systemic and cerebral—occurs in association with depressive disorders with a greater frequency than expected by chance alone. Unless properly treated, such depression negatively affects the prognosis of the physical disorder. More provocatively, there is current reawakening of the contribution of cerebral and cardiovascular factors to the origin of late-onset psychotic depressions (previously classified as *involutional melancholia*).

An integrated framework of pathogenesis is necessary for understanding psychopharmacologic, somatic, and psychotherapeutic approaches in the clinical management of patients with mood disorders. A historical perspective on current developments is also a valuable lesson in the study of mood disorders.

ANCIENT GREEK AND ROMAN DESCRIPTIONS

The Greeks and Romans described much of our current concepts of mood disorders, and we can trace much of our modern thinking to their descriptions. They coined the terms *melancholia* and *mania* and noted their relation. The ancients also hypothesized a temperamental origin for those disorders.

Melancholia

Hippocrates (460 to 357 BCE) described melancholia ("black bile") as a state of "aversion to food, despondency, sleeplessness, irritability, and restlessness." Thus, in choosing the condition's name, Greek physicians (who likely borrowed the concept from ancient Egyptians) postulated the earliest biochemical formulation of any mental disorder. They believed that the illness often arose from the substrate of the somber, melancholic temperament, which, under the influence of the planet Saturn, made the spleen secrete black bile, ultimately leading to mood darkening through its influence on the brain.

One Hippocratic aphorism recognized the close link between anxiety and depressive states: "Patients with fear or depression of long-standing are subject to melancholia." Hippocrates, who described the first historical case of melancholia, may have also been the first to describe a depressive mixed state, an activated form of depression:

> *A woman of Thasos became morose because of justifiable grief, and although she did not take to her bed, she had insomnia, loss of appetite . . . she complained of fears and talked much; she showed despondency and . . . talked at random and used foul language . . . many intense and continuous pains . . . she leapt up and could not be restrained.* [Hippocrates.]

According to Galen (131 to 210 CE), melancholia manifested in "fear and depression, discontent with life, and hatred of all persons." A few hundred years later, another Roman, Aurelianus, citing the now-lost works of Soranus of Ephesus, amplified the role of aggression in melancholia (and its link to suicide) and described how the illness assumed delusional coloring:

> *Animosity toward members of the household, sometimes a desire to live and at other times a longing for death, suspicion on the part of the patient that a plot is being hatched against him.*

In addition to natural melancholia, which, presumably, arose from an innate predisposition to overproduce the dark humor and led to a more severe form of the malady, ancient medicine recognized such environmental contributions to melancholia as immoderate consumption of wine, perturbations of the soul due to the passions (e.g., love), and disturbed sleep cycles. Autumn was considered the season most disposing to melancholy.

Mania

The ancient Greeks noted a state of raving madness with exalted mood, although it referred to a somewhat broader group of excited psychoses than that in modern nosology. Its relation to melancholia was probably noted as early as the first century BCE, but, according to Aurelianus, Soranus discounted it. Nonetheless, Soranus had observed the coexistence of manic and melancholic features during the same episode, consisting of continual wakefulness and fluctuating states of anger and merriment and, sometimes, of sadness and futility. Thus, Soranus seemed to have described what today are called *mixed episodes.* Natural melancholy was generally considered a chronic disorder, but Soranus noted the tendency for attacks to alternate with periods of remission.

Although others before him hinted at it, many credit Aretaeus of Cappadocia (ca. 150) for making the connection between the two major mood states: “It appears to me that melancholy is the commencement and a part of mania.” He described the cardinal manifestations of mania as it is known today:

> *There are infinite forms of mania, but the disease is one of them. If mania is associated with joy, the patient may laugh, play, dance night and day, and go to the market crowned as if a victor in some contest of skill. The ideas the patients have are infinite. They believe they are experts in astronomy, philosophy, or poetry.*

Aretaeus described the extreme psychotic excitement that could complicate the clinical picture of mania:

> *The patient may become excitable, suspicious, and irritable; hearing may become sharp . . . [and they might] get noises and buzzing in the ears; or may have visual hallucinations; bad dreams and his sexual desires may get uncontrollable; aroused to*

> *anger, he may become wholly mad and run unrestrainedly, roar aloud; kill his keepers, and lay violent hands upon himself.*

Noting the fluctuating nature of symptoms in the affectively ill, Aretaeus commented:

> *They are prone to change their mind readily; to become base, mean-spirited, illiberal, and in a little time extravagant, munificent, not from any virtue of the soul, but from the changeableness of the disease.*

Aretaeus was keenly aware of the characterologic distortions so commonly manifested during the different phases of cyclical mood disorders.

Finally, consolidating the knowledge of several centuries, Aretaeus described mania as a disease of adolescent and young men given intermittently to “active habits, drunkenness, lechery” and an immoderate lifestyle (what today might be called *cyclothymic disorder*). Exacerbations were most likely to occur in the spring.

Affective Temperaments

The concept of health and disease in ancient medicine was based on harmony and balance of the four humors, of which the ancients deemed sanguine humor the healthiest. Even desirable humor, however, such as blood, which made persons habitually active, amiable, and prone to jest, could, in excess, lead to the pathologic state of mania. The melancholic temperament, dominated by black bile and predisposed to pathologic melancholia, was described as lethargic, sullen, and given to brooding or contemplation; its modern counterpart might be persistent depressive disorder as defined in DSM-5. A long tradition dating back to Aristotle (384 to 322 BCE) attributed creative qualities to the otherwise tortured melancholic temperament in philosophy, the arts, poetry, and politics. The remaining two temperaments—choleric and phlegmatic—were less desirable, in that yellow bile made persons choleric (irritable, hostile, and given to rage), and phlegm made them phlegmatic (indolent, irresolute, and timid). The choleric and phlegmatic temperaments would probably be recognized today as borderline personality disorder and avoidant or schizoid personality disorder, respectively.

Many of the original Greek texts on melancholia were transmitted to posterity through medieval Arabic texts written by such scholars as Ishaq Ibn Imran and Abu Ali al-Hussain ibn Abdallah ibn Sina (known in the

West as Avicenna), and their Latin renditions by Constantinus Africanus. In describing different affective states, Avicenna developed the theory of the temperaments to its fullest. He speculated that a particular form of melancholia supervened “if black bile is mixed with phlegm,” when the illness was “coupled with inertia, lack of movement, and quiet.” Furthermore, mania was not necessarily linked to the sanguine (what today is termed *hyperthymic*) temperament, in that many forms of excited madness were believed to represent a mixture of black and yellow bile.

Avicenna further observed that the mixture of anger and restlessness in melancholia indicated that the disease was manic in nature and that the appearance of such signs and symptoms with violence heralded the transition from melancholia to mania. Those elaborations on Galen’s temperamental types might be considered the forerunners of current personality dimensions, deriving mood states from various mixtures of neuroticism and introversion–extroversion. (What both the ICD-10 and the DSM-5 describe as *cyclothymic disorder* represents the intense mood lability of high neuroticism coupled with cyclic alternation between extroversion and introversion.) Speculation on how we can understand diverse depressive phenomena as a mix of humors anticipated modern multiple-transmitter hypotheses of depression. Ishaq Ibn Imran summarized the existing knowledge of melancholia by considering the interaction of genetic factors (“injured prenatally as the result of the father’s sperm having been damaged”) with an exceptional temperament given to “mental overexertion”—although not necessarily physical overactivity—that, in turn, was associated with “disruption of the correct rhythms of sleeping and waking.” Those views, too, have a very modern ring to them.

MODERN ERA

The first English text (Fig. 13.1–1) entirely devoted to affective illness was Robert Burton’s *Anatomy of Melancholy*, published in 1621. A scholarly review of medical and philosophical wisdom accumulated in past centuries also anticipated many modern developments. The concept of affective disorder endorsed by Burton was rather broad (as it always has been in the United Kingdom), embracing mood disorders and many disorders that are today considered somatoform disorders, including hypochondriasis. Although he described “causeless” melancholia, Burton also categorized the various forms of love, melancholy, and grief. Particularly impressive was his catalog of causes, culminating in a grand conceptualization:

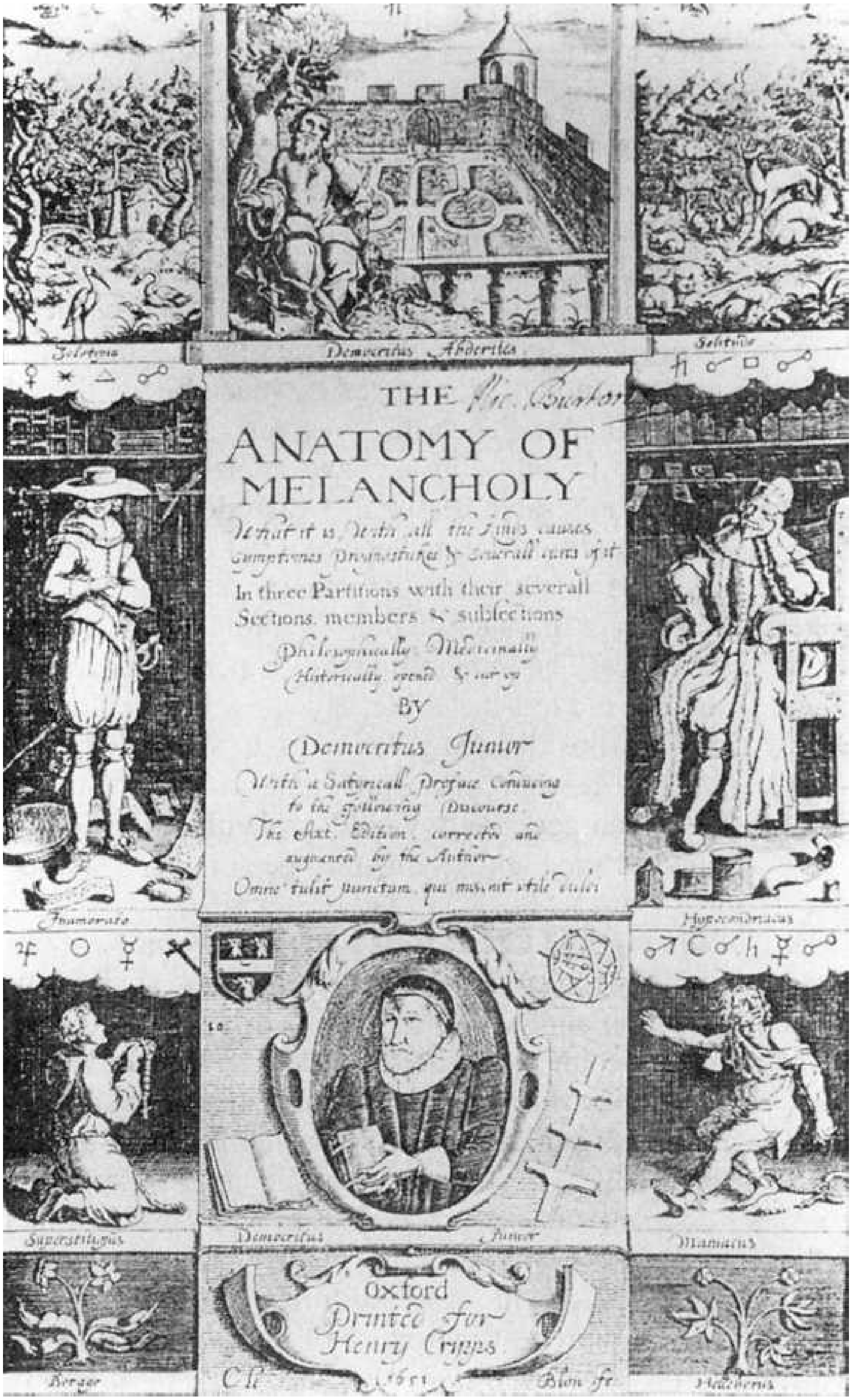

THE ANATOMY OF MELANCHOLY

What it is. With all the kindes causes symptomes, prognostickes & severall cures of it.

In three Partitions with their severall Sections, members & subsections

Philosophically, Medicinally, Historically opened & cut up

By

Democritus Junior

With a Satyricall Preface conducing to the following Discourse.

The Sixt Edition, corrected and augmented by the Author

Omne tulit punctum, qui miscuit utile dulci

Oxford
Printed for
Henry Cripps

1651

Zelotypia
Democritus Abderites
Solitudo
Inamorato
Hypocondriacus
Superstitiosus
Democritus Junior
Maniacus
Borrage
CL
Blon fe
Helleborus

FIGURE 13.1–1. Frontispiece of Robert Burton's *Anatomy of Melancholy* (1621).

> *Such as have the moon, Saturn, Mercury misaffected in their genitures … such as are born of melancholy parents; as offend in those six non-natural things … are of a high sanguine complexion … are solitary by nature, great students, given to much contemplation, lead a life out of action, are most subject to melancholy. Of sexes both, but men more often … Of seasons of the year, the autumn is most melancholy … Jubertus excepts neither young nor old.*

Burton's six non-natural things referred to such environmental factors as diet, alcohol, biologic rhythms, and perturbations induced by passions such as intense love. Burton did not definitively indicate age prevalences. Like nearly all of his predecessors, he favored male (rather than the currently reported female) preponderance.

Finally, Burton considered both the melancholic (contemplative) and the sanguine (hot-blooded) temperaments to be substrates of melancholia. Burton's work thus linked certain forms of depression with the softer expressions of the manic disposition, or bipolar II disorder, from which he appears to have suffered.

The 18th and 19th centuries introduced humane hospital care of the mentally ill, thereby permitting systematic clinical observation of the psychopathology and outcome of mood disorders.

Concept of Affective Disorder

Although Celsus (ca. 30CE) had described "forms of madness that go no further than sadness," the French alienist Jean-Philippe Esquirol (1840) may have been the first psychiatrist in modern times to suggest that a primary disturbance of mood might underlie many forms of depression and related paranoid psychoses. Until Esquirol's work, physicians considered melancholia a form of insanity (i.e., ascribed to deranged reasoning or thought disturbance). Esquirol's observations on melancholic patients led him to postulate that their insanity was partial (dominated by one delusion, a monomania) and that "the symptoms were the expression of the disorder of the affections. The source of the evil is in the passions." He coined the term *lypemania* (from the Greek, "sorrowful insanity") to give nosologic status to a subgroup of melancholic disorders that were affectively based. Esquirol cited Benjamin Rush (1745 to 1813), the father of American psychiatry, who had earlier described *tristimania*, a form of melancholia in which sadness predominated.

Esquirol's influence led other European psychiatrists to propose milder states of melancholia without delusions, which were eventually categorized as simple melancholia and, ultimately, as primary depressions. Such descriptions culminated in the Anglo-Saxon psychiatric term *affective disorder*, coined by Henry Maudsley (1835 to 1918), the renowned British psychiatrist commemorated by the eponymous London hospital.

Manic-Depressive Illness and the Question of Psychogenic Depressions

Since it was described two millennia ago, the connection between mania and depression has sporadically been rediscovered. However, two of Esquirol's disciples did the clinical work establishing the illness as a discrete nosologic entity in the 1850s. Jean-Pierre Falret called it *circular insanity*, and Jules Baillarger used the term *folie à double forme.* That accomplishment built on Philippe Pinel's reforms, which championed humane treatment of the mentally ill in Paris around the turn of the 18th century and emphasized systematic clinical observations of patients and detailed case records. French alienists made longitudinal observations on the same patient from one psychotic attack into another.

Furthermore, Esquirol had introduced the practice of chronicling events in statistical tables. Thus, French alienists applied the Hippocratic approach to defining a particular case by its onset, circumstances, course, and outcome to the affectedly ill. The humanitarian reforms introduced in the 19th century ensured that general health and nutrition standards would improve the outlook for the mentally ill—especially those with potentially reversible disorders such as affective disorders. By segregating the nondeteriorating mood disorders from other types of insanity, the French school then paved the way for the Kraepelinian system.

Kraepelin's (1856 to 1926) unique contribution was not so much his grouping together of all the forms of melancholia and mania, but his methodology and painstaking longitudinal observations, which established manic-depressive illness as a nosologic entity and (he hoped) a disease entity. He based his rationale on the following: (1) the various forms had common heredity measured as a function of familial aggregation of manic and depressive cases, (2) frequent transitions from one to the other occurred during longitudinal follow-up of patients, (3) a recurrent course with illness-free intervals characterized most cases, (4) the superimposed episodes were commonly opposite to the patient's habitual temperament —that is, mania on a depressive temperament and depression on a hypomanic temperament—and (5) both depressive and manic features could occur during the same episode (mixed states). Kraepelin developed his synthesis as early as the sixth (1899) edition of his *Lehrbuch der Psychiatrie* and most explicitly stated in the opening passages of the section on manic-depressive psychosis in the eighth edition (published in four volumes, 1909 to 1915):

> *Manic-depressive insanity … includes on the one hand the whole domain of so-called periodic and circular insanity, on the other hand simple mania, the greater part of the morbid states termed melancholia and also a not inconsiderable number of cases of confusional insanity. Lastly, we include here certain slight and slightest colorings of mood, some of them periodic, some of them continuously morbid, which on the one hand are to be regarded as the rudiment of more severe disorders, on the other hand, pass over without boundary into the domain of personal predisposition.* (Prof. Emil Kraepelin, Manic-depressive Insanity and Paranoia; translated by R. Mary Barclay, M.A., M.B.; edited by George M. Robertson, M.D., F.R.C.P. Edin. Edinburgh: E & S Livingstone; 1921.)

For Kraepelin, the core pathology of clinical depression consisted of lowered mood and slowed (retarded) physical and mental processes. In mania, by contrast, the mood was elated and physical and mental activity was accelerated. His earlier observations on what he termed *involutional melancholia* (referring to 40- to 65-year-old patients with extreme anxiety, irritability, agitation, and delusions) had led him to separate that entity from the broader manic-depressive rubric. In the eighth edition of *Lehrbuch der Psychiatrie,* however, he united melancholia with the manic-depressive group, with the justification that it was a unique form of mixed state and that follow-up conducted by his pupil Dreyfus had demonstrated unmistakable excited phases.

The classification of mood disorders is still evolving. Karl Leonhard in 1957, Jules Angst in 1966, Carlo Perris in 1966, and George Winokur, Paula Clayton, and Theodore Reich in 1969, working independently in four different countries, proposed that depressive disorders without manic or hypomanic episodes (unipolar depressive disorder) that appear in middle age and later are distinct from depressive episodes that begin at earlier ages and alternate with manic or hypomanic episodes (bipolar disorder). The main difference between the two affective subtypes is the more significant familial loading, especially for bipolar disorder among bipolar disorder probands.

Kraepelin had conceded the occurrence of psychogenic states of depression occasioned by situational misfortune. However, he believed that manic-depressive illness was hereditary, although he could not document postmortem anatomopathologic findings in the brains of manic-depressive patients. Therefore, manic-depression had to be considered a functional mental disorder in which they presumed that the brain disturbances lie in altered physiologic functions. Such biologic factors were deemed absent in the psychogenic depressions. Thus, Kraepelin's classification of mood disorders is both dualistic and unitary. It is dualistic to the extent that he designated them as either psychologically occasioned or somatically caused. It is unitary concerning disorders in the latter group, termed *endogenous affective disorders* (i.e., internal biologic causes). In other words, Kraepelin restricted the concept of clinical depression to what the DSM-5 terms *major depressive disorder with melancholic features*. Moreover, he postulated a continuum between that condition and what the DSM-5 and the ICD-10 and -11 now term *bipolar disorders*.

Cartesian thinking in 17th-century France conceptually separated mind from body, thereby providing physicians autonomy over the somatic sphere, free from interference by the Church. The dichotomous paradigm ensured that the complexities of mind–body interactions would not confound the study of the two aspects of the human organism. That is one reason Kraepelin's descriptive observations have proved valuable to subsequent generations of clinicians. Furthermore, his approach exemplifies the best tradition of scientific humanism in medicine: description and diagnostic categorization of an individual patient are necessary for the physician to apply the knowledge gained from past observation of similarly described and diagnosed patients. One limitation of the Kraepelinian approach is that it is not sufficiently articulate to account for mind–body interactions in the genesis of mental disorders because of its biologic reductionism.

Depressions as Psychobiologic Affective Reaction Types

It was the ambition of Swiss-born Adolf Meyer (1866 to 1950) to bridge the divide between psyche and soma. Meyer, who dominated psychiatry from his chair at Johns Hopkins University during the first half of the 20th century, coined the term *psychobiology* to emphasize that both psychological and biologic factors could enter into the causation of depressive and other mental disorders. Because of the nascent state of brain science during Meyer's time, he was more adept at biography than biology and paid greater attention to psychosocial causation. Meyer preferred the term *depression* (pressed down) to *melancholia* because of its lack of biologic connotation. He conceived of depressive states in terms of unspecified constitutional or biologic factors interacting with a series of life situations beginning at birth or even at conception. From that viewpoint arose the unique importance accorded personal history in depressive reactions to life events.

Meyer's terminologic revision left a somewhat confusing legacy in that we now use the term *depression* for a broad range of affective phenomena, ranging from sadness and adjustment disorders to clinical depression and bipolar disorders. Nosologic nuances to which Meyerians paid little attention, such as the difference between melancholic depression and more mundane depressions, are not just a matter of semantics. To the extent that those two forms of depression occur in different clinical settings, hypotheses based on one population may not apply to the other. For instance, uncontrollable traumatic events may have taught study participants to feel helpless or view the world negatively, but that does not equate with clinical depression. Nor does the process appear to be specific to depression. The failure to make such nosologic distinctions complicated trials that attempt to compare psychotherapy and pharmacotherapy for depressive disorders.

On the other hand, the Meyerian emphasis on biographical factors for the patient represented a more practical approach to depth psychology. Meyer's work also influenced recent sociologic interpretations of depression. In the final analysis, however, the Meyerian concern for the uniqueness of the individual has proved heuristically sterile. It de-emphasizes what is diagnostically common to different individuals, thereby obscuring the relevance of accrued clinical wisdom for the index patient. For that reason, the Meyerian approach, after enjoying clinical popularity for several decades in North America, has given way to neo-Kraepelinian rigor. However, the psychobiologic vision of bridging biology and psychology, one of the central preoccupations of psychiatric thought and research today, owes much to Meyer's legacy.

CONTEMPORARY MODELS OF DEPRESSION

Aggression-Turned-Inward Model

Sigmund Freud was initially interested in a psychoneural project for all mental phenomena. Limitations of the brain sciences of the day led him to adopt instead a model that relied on a concept of mental function borrowed from physics. Freud's disciple, Karl Abraham, initially suggested that depressed affect might derive from one's aggressive impulses directed against an ambivalently loved internalized object, and Freud later elaborated this idea. Abraham and Freud hypothesized that turned-inward anger was intended as punishment for the love object that had thwarted the depressed patient's need for dependency and love. Because the patient internalizes the object, they become the target of their own thanatotic impulses. A central element in those psychic operations was the depressed patient's ambivalence toward the object, which the patient perceived as a frustrating parent. Aggression directed at a loved object (parent) would then cause considerable guilt. In the extreme, such ambivalence, guilt, and retroflexed anger could lead to suicidal behavior.

Thus, this model implies that depression is an epiphenomenon of the transduction of thanatotic energy. Freud's earlier writings had similarly portrayed anxiety as being derived from the transformation of dammed-up sexual libido. Freud had hoped that these psychoanalytic constructs would eventually be localized neuroanatomically. However, we understand that the concepts are metaphors.

The concept of emotional behavior as an arena of incompatible forces confined to a psyche that is relatively impervious to current outside influences is the primary liability of the aggression-turned-inward model and, perhaps, of orthodox psychoanalysis itself. Although modern psychoanalysts have discarded the sexual-energy-transduction hypothesis of anxiety, they continue to use the aggression-turned-inward model clinically in modified versions. The lingering popularity of the model may be due, in part, to its compatibility with the clinical observation that many depressed patients suffer from a lack of assertion and outwardly directed aggressiveness. However, we encounter a substantial number of hostile, depressed patients in clinical practice (indeed, "depression with anger attacks" has been recently described), and clinical improvement in most patients typically leads to decreased, not increased, hostility. Such observations shed doubt on the aggression-turned-inward mechanism as a universal explanation for depressive behavior. Finally, little evidence supports the contention that outward expression of anger has therapeutic value in clinical depression.

Outwardly directed hostility in depression is not a new clinical observation; in fact, ancient Greek and Roman physicians had noted it. Hostility is best considered a manifestation rather than a cause of a depressive disorder, especially when mixed bipolar features attend the disorder. We can also understand the hostility of the depressed patient as an exaggerated reaction to frustrating love objects, as secondary to self-referential attribution, or simply as nonspecific irritability of an ego in affective turmoil. The latter could, in part, be a function of a concurrent personality disorder. Such commonsense explanations that do not invoke unobservable phenomena have greater appeal from heuristic and clinical perspectives.

Object Loss and Depression

Object loss refers to traumatic separation from significant objects of attachment. Ego psychology reformulations of the Abraham–Freud model focus on how these losses affect the ego, de-emphasizing other more abstract aspects. The depressant effect of separation events often resides in their symbolic meaning for a person and is difficult to measure objectively. However, love loss, bereavement, and other exits from the social scene, as defined by the London psychiatrist Eugene Paykel, are presently the concepts most commonly used in practice and research.

Although poets previously described love and melancholy, Freud first considered these systematically in his 1917 paper on mourning and melancholia. It appears that the transition from grief to pathologic depression occurs in no more than 10% of adults and 20% of children. These figures suggest that this transition occurs mainly in persons predisposed to mood disorders.

John Bowlby of the Tavistock Clinic, London, did a comprehensive clinical investigation of the child's attachment with the mother or mother substitutes during development, a bond considered the prototype for all subsequent bonds with other objects. Like many psychoanalytic explanations of adult symptom formation, the object loss model proposes a two-step hypothesis. The first step consists of early breaks in affectional bonds, which provide the behavioral predisposition to depression. The second involves adult losses, which revive the traumatic childhood loss and precipitate depressive episodes. However, the role of proximate separations in provoking depressive reactions rests on more solid clinical evidence than the hypothesized sensitization resulting from developmental object loss. Realizing this, Bowlby regarded childhood sensitization from early deprivation as a generic characterologic vulnerability, leading to many possible disorders.

Compared with aggression turned inward, object loss is more directly relevant to clinical depression, yet it is still pertinent to question whether it is an etiologic factor. Studies at the Wisconsin Primate Center indicated that individuals achieve optimal homeostasis with the environment most readily when securely attached to significant others. The dissolution of such ties appears to be relevant to the emergence of a broad range of psychopathologic disturbances than depression per se. A related methodologic question is whether object loss operates independently of other etiologic factors. For instance, a history of early breaks in attachment may reflect that one or both of the patient's parents had a mood disorder, resultant separation, divorce, or suicide.

On balance, the ego-psychological object loss model is conceptually superior to its id-psychological counterpart. In postulating an open system of exchange between a person and the environment, the model permits consideration of etiologic factors other than separation, such as heredity, character structure, and adequacy of social support—all of which might modulate the depressant effect of adult separation events. Conceptualizing the origin of depression along those lines is mainstream of current adaptation, homeostasis, and disease ideas. A vital treatment implication is the value of social support in preventing relapse and mitigating the chronicity of depression. That is, indeed, an ingredient in the interpersonal psychotherapy of depression, which can be conceptualized as a form of brief, focused, and practical psychodynamic therapy.

Loss of Self-Esteem and Depression

Reformulating the dynamics of depression as the ego suffering a collapse of self-esteem represents an additional conceptual break with the original Abraham–Freud formulation. Depression is said to originate from the ego's inability to give up unattainable goals and ideals. The model further posits internalized values of the ego (rather than thanatotic energy turned inward) impose the narcissistic injury that crushes the depressed patient's self-esteem. Because the ego construct is rooted in social and cultural reality, loss of self-esteem may result from symbolic losses involving power, status, roles, identity, values, and purpose for existence. This approach provides the clinician with a more flexible and pragmatic tool for understanding depressed persons than the classical model. This newer approach represents one of the first attempts to formulate depression in terms that subsequent psychological theory and research could operationalize in a more testable form.

Self-esteem is part of the habitual core of the individual and, hence, is integral to the personality structure. Indeed, low self-esteem conceived as a trait is a primary defining attribute of the depressive (melancholic) personality. Although understandably, environmental adversity can quickly sink such individuals into melancholia, it is not apparent why persons with higher self-esteem also succumb. To explain such cases, one must invoke an underlying instability in the system of self-esteem that renders it vulnerable to depression. The opposite is also known to occur; that is, manic episodes may develop from a baseline of low self-esteem, as in bipolar disorder patients with antecedent traits of shyness, insecurity, and dysthymia.

These considerations suggest that the fluctuations of self-esteem most central to the model of depression as loss of self-esteem represent a more fundamental mood dysregulation. Classic psychoanalysis considers this dysregulation to be of constitutional origin. Except for the idea that mania could be a way to deny painful affect, most psychoanalytic explanations for bipolar oscillations have not progressed beyond metapsychological jargon.

Cognitive Model

The cognitive model, developed by Aaron Beck at the University of Pennsylvania, hypothesizes that thinking along negative lines (e.g., thinking that one is helpless, unworthy, or useless) is the hallmark of clinical depression. Beck redefines depression as a cognitive triad. Patients think of themselves as powerless, interpret most events unfavorably vis-à-vis the self, and believe the future to be hopeless. More recent academic psychology formulations characterize these cognitions as a negative attributional style that is global, internal, and stable and exists in the form of latent mental schemata that generate biased interpretations of life events.

Because the cognitive model derives from retrospective observations of already depressed persons, proving causality is virtually impossible. Whether the negative mental schemata cause depression or are the result is difficult to know. The theoretical importance of the cognitive model lies in the conceptual bridge it provides between ego-psychological and behavioral models of depression. Most importantly, it has also lead to a popular psychotherapy method that addresses these negative attributions.

Therefore, the cognitive model has the cardinal virtue of focusing on critical reversible clinical dimensions of depressive illness, such as helplessness, hopelessness, and suicidal ideation, while providing a testable and practical psychotherapeutic approach. That approach, however, is less likely to succeed in patients with the full-blown melancholic manifestations of a depressive disorder. It is doubtful that negative cognitions alone could account for the profound disturbances in sleep, appetite, and autonomic and psychomotor functions encountered in melancholic depressions. Furthermore, conceptualizing a multifaceted disorder such as depression primarily or solely as a function of distorted cognitive processes is reminiscent of pre-Esquirolian notions that emphasized impaired reasoning in the development of depression. Finally, the extension of modified cognitive and cognitive-behavioral therapy to treat virtually all emotional disorders—even schizophrenia—reminds us of earlier grandiose claims made by psychoanalysts.

Learned Helplessness Model

The learned helplessness model is, in some ways, an experimental analog of the cognitive model. The model proposes that one learns the depressive posture from past situations in which one cannot avoid undesirable contingencies. The basis for this model is the research performed on dogs, which the researchers prevented from taking adaptive action to avoid unpleasant electrical shock. These dogs subsequently showed no motivation to escape such aversive stimuli, even when escape avenues were readily available. Armed with evidence from many such experiments, University of Pennsylvania psychologist Martin Seligman postulated a trait of learned helplessness formed from the accumulation of past episodes of uncontrollable helplessness. He defined learned helplessness as a belief that it is futile to initiate personal action to reverse aversive circumstances.

The learned helplessness paradigm is a general one and refers to a broader mental disposition than depression. Thus, it is potentially helpful in understanding such diverse conditions as social powerlessness, defeat in sporting events, and post-traumatic stress disorder (PTSD). In addition, past events might shape a characterologic cluster—consisting of passivity, lack of hostility, and self-blame—relevant to certain depressive phenomena. This paradigm may explain the low hostility observed in some patients during clinical depression. Learned helplessness could thereby provide plausible links between aspects of personal biography and clinical phenomenology in depressive disorders. Cognitive therapies capitalize on the concept and focus on developing new cognitive strategies geared to modifying expectations of uncontrollability. We can thus use experimental insights to inform our treatments.

Nonetheless, the clinician should be wary of unwarranted clinical extrapolations. For example, some therapists have argued that the depressed patient's passivity is "manipulative," serving to obtain interpersonal rewards. This dynamic may be significant in the development of the depressive character. That interpretation appears more relevant to selected aspects of depression than to the totality of the disorder. Depressive behavior and verbalizations have a powerful interpersonal impact, but casting depression as merely a masochistic lifestyle developed to secure interpersonal advantages represents a mechanistic circular argument that is disrespectful of the clinical agony of patients with mood disorders. Finally, although most formulations focusing on helplessness have emphasized acquisition through learning, recent experimental research in animals tends to implicate genetic factors in the vulnerability of learning to behave helplessly. The value of the helplessness paradigm may reside in its utility in predicting a variety of subthreshold affective disturbances generic to civilian reactions to adversity and trauma.

Depression and Reinforcement

Other behavioral investigators, such as the psychologist Peter Lewinsohn, developed clinical formulations of depression hinging on specific deficits in reinforcement mechanisms. According to the reinforcement model, depressive behavior may result from a lack of appropriate rewards, such as noncontingent rewards. The model identifies several contributory mechanisms. Some environments may consistently deprive persons of rewarding opportunities, thereby placing them in a chronic state of boredom, pleasurelessness, and, ultimately, despair. That reasoning, however, may offer more insight into social misery than clinical depression. A more plausible postulated mechanism is the provision of rewards that are not in response to the recipient's actions; in other words, the gratis provision of what a person considers undeserved rewards may lower self-esteem. Inadequate social skills may predispose to depression, as they decrease a person's chances of responding to potentially rewarding contingencies in the environment. Indeed, recent research on the relationship between personality and depression suggests that such deficits might underlie certain depressive states. Therefore, psychotherapeutic approaches designed to enlarge a patient's repertoire of social skills may prove valuable in preventing some types of depression.

The concepts of depression that have been derived from behavioral methodology and developed in the last several decades are scientifically articulate and, therefore, testable approaches to clinical depression. However, the critical distinction between depression on self-report inventories and clinical depression tends to be overlooked in investigations testing the reinforcement paradigm. Furthermore, the behavioral model does not address the distinct possibility that reinforcement deficits may, in part, represent the psychomotor deficits of depressive illness. Nevertheless, the behavioral model provides a conceptual bridge between purely psychological and emerging biologic conceptualizations of depression by focusing on reward mechanisms.

Biogenic Amine Imbalance

The formulation of sophisticated biologic explanations of mood disorders had to await the development of neurobiologic techniques that could probe the parts of the brain involved in emotions. Although we cannot directly observe the complex physiology of the limbic-diencephalic centers of emotional behavior, we have learned much from animal work. The limbic cortex links with the neocortex, the midbrain, and lower brain centers. The midbrain subserves higher symbolic functions, and the midbrain and lower brain centers control autonomic functions, hormonal production, sleep, and wakefulness. Thus, the limbic cortex influences all aspects of the body and mind. Norepinephrine-containing neurons influence many functions that are profoundly disturbed in melancholia, including mood, arousal, appetite, reward, and drives. Other biogenic amine neurotransmitters mediating such processes are the catecholamine dopamine and indoleamine serotonin. Dopamine is crucial for drive, pleasure, sex, and psychomotor activity, and serotonin regulates affects, aggression, sleep, and appetite. Cholinergic neurons, secreting acetylcholine at their dendritic terminals, are usually antagonistic to catecholaminergic neurons.

Although the opioid system might, on experimental and theoretical grounds, also serve as one of the neurochemical substrates for mood regulation, no cogent model of mood disorders involving that system has appeared. Likewise, biochemical formulations of mood disorders have paid relatively little attention to the major excitatory brain neurotransmitter glutamate and the inhibitory neurotransmitter γ-aminobutyric acid (GABA).

Biogenic Amine Hypotheses. Joseph Schildkraut at Harvard University and William Bunney and John Davis at NIMH published the first formal hypothesis connecting depletion or imbalance of biogenic amines (specifically norepinephrine) and clinical depression. Alec Coppen in England and I. P. Lapin and G. F. Oxenkrug in Russia emphasized the serotonin counterpart of the model. Both catecholamine and indoleamine hypotheses are essentially derived from two sets of pharmacologic observations. First, reserpine, which decreases blood pressure by depleting biogenic amine stores, precipitates clinical depression in some patients. Second, antidepressant medications, which alleviate clinical depression, raise the functional capacity of the biogenic amines in the brain. This thinking style is known as the *pharmacologic bridge*, extrapolating from evidence on the mechanism of drug action to the neurotransmitter pathologies presumed to underlie a given psychiatric disorder. Such pharmacologic strategies have been of heuristic value in developing research methods to investigate mood disorders and schizophrenia. Indeed, the research methodology developed by the relatively few investigators working in this area during the last half-century is among the most elegant in the history of psychiatry.

Variations of the biogenic amine model assign somewhat different relative weights to the biogenic amines norepinephrine and serotonin in the development of pathologic mood states. Arthur Prange and colleagues at the University of North Carolina formulated a permissive biogenic amine hypothesis in which serotonin deficits permit the expression of catecholamine-mediated depressive or manic states. Subsequent animal research showing that an intact serotonin system is necessary for the optimal functioning of noradrenergic neurons supported this hypothesis. Omission of tryptophan from the diet of antidepressant-responsive depressed patients may annul the efficacy of the antidepressant; among healthy volunteers, that special diet also induces sleep electroencephalographic characteristics of clinical depression. Although such findings are provocative, the precursor-loading strategy to increase the brain stores of serotonin (e.g., with l-tryptophan) has not been unequivocally successful in reversing clinical depression. Dietary loading with catecholamine precursors has fared even worse than serotonin-precursor loading in the treatment of depression.

The cholinergic–noradrenergic imbalance hypothesis proposed by David Janowsky and colleagues is another attempt to understand the roles of biogenic amines. This hypothesis, along with the related cholinergic supersensitivity hypothesis developed by J. Christian Gillin, has been tested extensively at the University of California at San Diego. Subsequent formulations by Larry Siever and Kenneth Davis at the Mount Sinai Hospital in New York have refocused on noradrenergic dysregulation. The model assumes oscillation from one output mode to the other at different phases of depressive illness. In a provocative extrapolation from that model, bipolar depression would have low noradrenergic output. However, many instances of MDD, as with some anxiety disorders, could be biochemically conceptualized as high-output conditions.

Despite more than four decades of extensive research and indirect evidence, researchers have not demonstrated deficiencies or excesses of biogenic amines in specific brain structures that are necessary or sufficient for the occurrence of mood disorders. It has not been possible to confirm the putative role of central norepinephrine in depression or discard it altogether. The role of dopamine as formulated, among others, by the Italian pharmacologist Gian Luigi Cessa, although studied less extensively than norepinephrine, deserves greater recognition in that it might have relevance to atypical and bipolar depression or to mania.

Preliminary data from a small brain imaging study has shown blunted serotonin responsivity in prefrontal and temporoparietal areas in unmedicated patients with MDD. Such data, considered in the context of the overall serotonin literature in depression, are provocative but not conclusive and illustrate that the case for serotonergic disturbance in depression continues to be indirect. Moreover, the putative permissive role of serotonin is more evident in aggressive suicide attempts. Serotonergic dysfunction might subserve other conditions characterized by lack of inhibitory control, such as obsessive-compulsive disorder (OCD), panic disorders, bulimia nervosa, certain forms of insomnia, alcoholism (alcohol use disorder), and impulse-ridden personality disorders. Such considerations have led Dutch psychiatrist Herman van Praag and colleagues to postulate a dimensional neurochemical disturbance generic to a large group of disorders within the traditional nosology. We could interpret this hypothesis as a challenge to psychiatric nosology or a statement of the need to supplement clinical classification with biochemical parameters. Both interpretations are in line with clinical observations during the last two decades, testifying to the high prevalence of comorbidity in depressive, other emotional disorders, and certain impulse-control disorders.

By implication, the preceding postulated biochemical faults are genetically determined. Although biogenic amine models of mood disorders were developed retrospectively from the pharmacologic action of antidepressant and thymoleptic agents, they have stimulated the development of new classes of antidepressants with more selective action on specific neurotransmitter receptors. Their introduction has virtually revolutionized the treatment of depression. Yet, the fundamental biochemistry of mood disorders is still far from being understood. Curiously, although selective in action, the new compounds working on the serotonin system have broad effectiveness for various mood-related conditions, such as dysthymic disorder, PTSD, OCD, panic disorder, social phobia, bulimia nervosa, and borderline personality disorder. Such data indirectly favor the hypothesis of an underlying biologic commonality to several of these disorders. In turn, the preceding considerations led to a provocative formulation of an increasingly prevalent “social syndrome” in populations experiencing social disruption, immigration, and abuse. This syndrome includes symptoms of anxiety, depression, violence-proneness, impulsivity, and suicidality. This formulation suggests a perturbation of the serotonin system, the most basic brain structure involved in human socialization and territoriality, coping with stress, danger, and survival.

New antidepressants with dual action on both serotonergic and noradrenergic receptors and emerging data on their possible greater efficacy in melancholic depressions suggest that the biochemistry of mood disorders is more complex than a single-neurotransmitter hypothesis can imply. The work of George Henninger and colleagues at Yale University indicates further that monoamines better explain how antidepressants facilitate recovery from depression than their being the fundamental causes of depression. Moreover, emerging biochemical paradigms are moving away from distal biochemical lesions to focus on molecular perturbations closest to the putative genetic underpinnings of mood disorders. The initial work focused on mood stabilizers’ mechanisms in bipolar disorder. However, researchers are now exploring second messenger systems, phosphorylation G proteins, signal transduction, DNA transcription, and messenger RNA translation. Again, such a search for molecular mechanisms represents “backward logic” from the putative mechanism of action of selected thymoleptic agents.

The same is true for Frederick Petty’s GABAergic and Shih-Jen Tsai’s brain-derived neurotrophic factor (BDNF) hypothesis of the origin of bipolar disorder. BDNF is responsible for the ongoing maintenance of neurons in the brain and disruption of it causes a reduction in neuronal number and size. It is challenging to measure BDNF in living subjects, but postmortem studies have suggested that deficits in BDNF may correlate with psychopathology. For example, some studies have found that in postmortem examination of persons who died from suicide, the average BDNF in the prefrontal cortex and hippocampus was lower than for control subjects. Indirect studies in live patients, looking at BDNF in plasma rather than the brain (the correlation between the two is not clear), found increases in serum BDNF in depressed patients who responded to antidepressants.

Neuroendocrine Links. Functionally inadequate mobilization of neurotransmitters in the face of continued or repeated stress, indirectly reflected in the pathologic modification of noradrenergic and serotonergic receptor function, could represent neurochemical final common pathways of homeostatic failure. Such mechanisms could also provide links with psychoendocrine dysfunction; the hypothesized neurotransmitter deficits may underlie the disinhibition of the hypothalamic–pituitary–adrenal axis, characterized by steroidal overproduction, the most widely studied endocrine disturbance in depressive illness. When challenged with dexamethasone, the altered axis resists suppression, which offered Bernard Carroll’s team (then at the University of Michigan) the possibility of developing the dexamethasone-suppression “test” (DST) for melancholia. This procedure is of uncertain specificity for depressive illness and is unsuitable to serve as a diagnostic test. However, that line of research has been helpful in pathogenetic understanding. For instance, it led to the demonstration by Charles Nemeroff (then at Emory University) of increased concentrations of corticotropin-releasing factor (CRF) in the cerebrospinal fluid (CSF) of patients with MDD. CRF also appears relevant to the pathophysiology of anxiety disorders, such as panic disorder and PTSD. The research of Florian Holsboer's group at Munich's Max Planck Institute has shown impaired glucocorticoid and mineralocorticoid receptor function in these disorders, with relevant pathophysiologic and therapeutic implications.

Another neuroendocrine index of noradrenergic dysregulation, blunted growth hormone response to the α_2-adrenergic receptor agonist clonidine, likewise points to limbic–diencephalic disturbance. However, studies performed in the United States suggest that it is positive in endogenous depression and severe anxiety disorder (panic disorder). Thyroid-stimulating hormone (TSH) blunting on thyrotropin stimulation, another common neuroendocrine disturbance in depression, also shows limited specificity.

However, what is remarkable is that the DST, clonidine, and thyrotropin challenge data, in aggregate, identify most persons with clinical depression. Such evidence of midbrain disturbance argues for considering clinical depression to be a legitimate disease. The disease concept of depression is further buttressed by computed tomography (CT) scans showing enlarged pituitary and adrenal glands, a state marker of depressive illness.

Stress and Depression. The concept of a pharmacologic bridge implies two-way traffic. The hypothesized chemical aberrations may be primary or genetically based. However, we should provide for the likelihood that psychological events that precipitate clinical depression might initiate or exacerbate neurochemical imbalance in vulnerable subjects. Support for this suggestion includes studies in animals in which early separation from peers and inescapable frustration profoundly alters the turnover of biogenic amines and postsynaptic receptor sensitivity. Thus, in genetically predisposed persons, environmental stressors might more easily lead to perturbations of limbic–diencephalic neurotransmitter balance. Finally, in vulnerable individuals, especially during the formative years of childhood, psychological mechanisms might more easily perturb midbrain neurochemistry. Traumatic experiences appear particularly potent in this regard. The hippocampus has been the subject of intense recent research as the possible neuroanatomical substrate linking such loss and trauma to adult depression. Ongoing ingenious experimental paradigms in primates and rodents continue to explore the role of early experience and stress in subsequent depressive-like behaviors in these animals. In humans, a new provocative finding indicates that a polymorphism of the serotonin transporter gene would identify who among traumatized children would develop adult depression. Likewise, a polymorphism of the monoamine oxidase (MAO) gene plays a significant role in determining who among battered children grows into an adult sociopath. Animal models of mania are sparse and problematic.

Neurophysiologic Approaches

Neuronal Hyperexcitability. Lithium is known to replace intracellular sodium and hyperpolarize the neuronal membrane, thereby decreasing neuronal excitability. Abnormalities in neuronal electrolyte balance (an excess of residual sodium, defined by radioisotope techniques) and hypothesized secondary neurophysiologic disturbances were the focus of British investigations by Alec Coppen and colleagues in the early 1960s. The data appear compatible with the hypothesized movement of excess sodium into the neuron during an episode of mood disorder and redistribution toward the preillness electrolyte balance across the neuronal membrane during recovery. Intraneuronal sodium leakage is postulated in both depressive and manic disorders but is deemed more extreme in the latter. Because the harmonious activity of the neuronal cell and, by implication, that of a group of neurons, depends on the electrical gradient maintained across its membrane by differential distribution of sodium, abnormalities in sodium concentrations and transport are hypothetically relevant to the production of an unstable state of neurophysiologic hyperexcitability. In formulating their thesis of neurophysiologic arousal in melancholic states, Joseph Mendels and Peter Whybrow (both of whom have worked at the University of Pennsylvania) have capitalized on the previous electrolyte disturbances. The view that mania represents a more extreme electrophysiologic dysfunction in the same direction as depression violates the commonsense notion of symptomatologic “opposition” between the two kinds of disorder. Yet, in part, it may account for the existence of mixed states in which symptoms of depression and mania coexist. The NIMH team led by Frederick Goodwin first showed that a substantial minority of depressed patients with a bipolar substrate respond to lithium salts, which further supports the concept of a common neurophysiologic denominator to mania and depression. Perturbations of calcium metabolism also appear relevant to bipolar patients. Therapeutic implications of this observation (e.g., the use of calcium channel inhibitors in bipolar I disorder) have not yielded consistent results. Finally, rubidium, another alkali metal, has been explored in the depressive phase of bipolar disorders, again with inconclusive results.

Immunologic Disturbance. Depressive disorders are associated with several immunologic abnormalities, including decreased lymphocyte proliferation in response to mitogens and other forms of impaired cellular immunity. These lymphocytes produce neuromodulators, such as CRF and cytokines, peptides known as interleukins. There appears to be an association with clinical severity, hypercortisolism, and immune dysfunction, and the cytokine interleukin-1 may induce gene activity for glucocorticoid synthesis.

Circadian Rhythms. It has long been understood that depression is associated with a premature loss of deep (slow-wave) sleep and an increase in nocturnal arousal. Four types of disturbance reflect the latter: (1) an increase in nocturnal awakenings, (2) a reduction in total sleep time, (3) increased phasic rapid eye movement (REM) sleep, and (4) increased core body temperature. The combination of increased REM drive and decreased slow-wave sleep results in a significant reduction in the first period of non-REM (NREM) sleep, a phenomenon referred to as reduced REM latency. Blunted secretion of growth hormone may cause this reduction in latency, which usually increases with sleep onset. It is very persistent, often lasting beyond recovery from depression. Some patients manifesting a characteristically abnormal sleep profile are less responsive to psychotherapy and have a higher risk of relapse or recurrence, and may benefit preferentially from pharmacotherapy.

Shortened REM latency, which David Kupfer and colleagues have extensively studied at the University of Pittsburgh, has been proposed as another laboratory “test” for depressive disorder. Shortened REM latency may serve as a trait marker for depression because it occurs in dysthymia and so-called borderline personality and among the clinically “well” offspring of adults with major depression.

Formulations of circadian rhythms by Thomas Wehr and Norman Rosenthal, working at NIMH, have focused on abnormalities in brain regulation of temperature, activity, and sleep cycles. Others have investigated the role of the pineal hormone melatonin in mood disorders without achieving consistent results. Applying circadian rhythm research concepts to women with mood disorders has also led to innovative methods, but, again, without definitive characterization of the neurophysiologic faults.

At a basic level, instrumental REM sleep deprivation in neonatal rats leads to adult “depression-like” behaviors. In human studies, sleep deprivation and exposure to bright white light correct phase disturbances and thereby terminate depressive episodes, especially in subjects with periodic and seasonal depressions. The average modern person is light-deprived, and phototherapy can benefit even those without clear-cut seasonal patterns. Unfortunately, except for their use in mild seasonal depressions and suppression of sleep deprivation to prevent hypomania in bipolar disorder, the preceding circadian studies have not had a tangible impact on practice. Their application in the large segment of mood-disordered patients remains cumbersome, if not elusive. Their effect can be better assessed at the level of theory. Although the specificity and efficacy of these neurophysiologic indices and manipulations for clinical depression and bipolar disorder require additional investigation, cumulatively, they point to midbrain dysregulation as the likely common neurophysiologic substrate of affective disorders. The preceding considerations further suggest that the ancient Greeks, who ascribed melancholia to malignant geophysical influences, did not indulge in mere poetic metaphor. The ancients had observed the disturbed circadian patterns and advocated their readjustment to restore euthymia.

Affective Dysregulation. A significant challenge for research in mood disorders is to characterize the fundamental molecular mechanisms that underlie the neurophysiologic rhythmopathies, which, in turn, might account for the recurrent nature of the affective pathology as envisioned by Kraepelin. For the most typical recurrent forms of the disorders, the constitutional foundations (manifested as cyclothymic and dysthymic traits or a broad range of emotional disequilibrium covered by the rubric of “neuroticism”) are so unstable that the illness may run its entire course more or less autonomously, with the environment mainly serving to turn on and off the more florid phases (episodes). Parisian psychiatrist Jean Delay, a pioneer in psychopharmacology in the 1950s, also emphasized affective dysregulation as the fundamental pathology in the spectrum of mood disorders. Robert Post, at NIMH, hypothesized that the electrophysiologic substrates could be so kindled that an infrequent and episodic disorder initially triggered by environmental stressors could assume an autonomous and recurrent course. He hypothesized that this phenomenon might occur because neuronal perturbations brought about by stressors in the early course of mood disorders get incorporated into the DNA. However, this fascinating kindling hypothesis does not seem to pertain to garden-variety mood disorders but those with extreme cyclicity. The monograph on manic-depressive illness by Fred Goodwin and Kay Jamison presents in-depth arguments for this cyclical paradigm of thymopathy.

Through its connection with the serotonin transporter, the amygdala appears to play a crucial role in the dysregulation of the emotional life of affectively ill patients. Data indicate that the dysregulation is both in function and anatomic structure.

THEORETICAL SYNTHESIS

Pathophysiologic Understanding

Modern psychobiology attempts to link experience and behavior to the central nervous system (CNS). Building conceptual bridges between the psychological and biologic approaches to mood disorders require sophisticated strategies beyond the Cartesian notion of limited mind–body interactions or the pineal gland and the generalizations of the Meyerian school.

William McKinney and Hagop Akiskal, in 1973, developed a conceptual framework that considers the affective syndromes as the final common pathway of various psychological and biologic processes. The overarching hypothesis is that psychological and biologic–etiologic factors converge in reversible deficits in the diencephalic substrates of pleasure and reward. Those areas of the brain subserve the functions that are disturbed in melancholia and mania. The integrative model links the central chemistry and physiology of reward mechanisms with the object loss and behavioral models of depression, both of which give singular importance to the depressant role of loss of rewarding interpersonal bonds. A vital element of the model is the circadian disturbances observed since ancient times in depressive and manic syndromes. Both syndromes are conceptualized as clinical manifestations of a disordered limbic system with its subcortical and prefrontal extensions. Brain imaging studies in melancholic patients by Wayne Drevets, originally at Washington University, have visualized limbic disturbances extending into subcortical structures and occurring primarily in patients with a familial diathesis for depression; the amygdala appears to be the focal limbic structure in the latter studies. Clinical experience and research data suggest that multiple factors described in the following subsections combine to produce or exacerbate dysregulation in these brain regions, leading to the final common pathway of clinical depression. The data on mania is more tentative and is mentioned when relevant.

Pathologic Findings from Brain Imaging Studies. The most consistent structural abnormality observed in depressive disorders is the increased frequency of abnormal hyperintensities in subcortical regions, such as periventricular regions, the basal ganglia, and the thalamus. More common among older adults, these hyperintensities appear to reflect the deleterious neurodegenerative effects of recurrent affective episodes. Ventricular enlargement, cortical atrophy, and sulcal widening also have been reported in some studies. Some depressed patients also may have reduced hippocampal or caudate nucleus volumes, or both, suggesting more focal defects in relevant neurobehavioral systems. Diffuse and focal areas of atrophy are associated with increased illness severity, bipolarity, and increased cortisol levels.

Also, specific structures appear reduced in volume in patients with depression, for example, the hippocampus. This reduction may be due to neuronal loss due to the neurotoxic effects of the increased cortisol found in depressed patients. Although many brain areas are susceptible to this toxic effect, the hippocampus is particularly vulnerable as it is rich in glutaminergic neurons. Other areas with reported brain loss include the prefrontal cortex, cingulate gyrus, and cerebellum.

The most widely replicated positron emission tomography (PET) finding in depression is decreased anterior brain metabolism, which is generally more pronounced on the left side. From a different vantage point, depression may cause a relative increase in nondominant hemispheric activity. Furthermore, a reversal of hypofrontality occurs after shifts from depression into hypomania, such that higher left hemisphere reductions occur in depression compared with more significant right hemisphere reductions in mania. Other studies have observed more specific reductions of reduced cerebral blood flow or metabolism, or both, in the dopaminergically innervated tracts of the mesocortical and mesolimbic systems in depression. Evidence suggests that antidepressants at least partially normalize these changes.

In addition to a global reduction of anterior cerebral metabolism, increased glucose metabolism occurs in several limbic regions, particularly among patients with relatively severe recurrent depression and a family history of mood disorder. During episodes of depression, increased glucose metabolism correlates with intrusive ruminations.

One meta-analysis found three consistent differences when comparing fMRI studies of depressed patients with healthy controls. They found activation of the pulvinar nucleus of the thalamus, a more significant response to negative stimuli in the amygdala, insula, and anterior cingulate, and a lesser response in the dorsal striatum and dorsal lateral prefrontal cortex.

Heredity. Numerous family, adoption, and twin studies have long documented the heritability of mood disorders. Recently, however, the primary focus of genetic studies has been to identify specific susceptibility genes using molecular genetic methods.

Depression aggregates within families, and meta-analytic studies suggest that first-degree relatives of persons with MDD have an increased odds ratio of 2.84. Only limited adoption studies have been done, of varying quality and mixed results and most of what we know about the genetic contribution to the disorder comes from twin studies, which estimate the heritability to be 37%. It appears that this heritability is higher in women. The twin studies also concluded that shared environmental factors did not play a significant role in the familial aggregation of the disorder.

Genome-wide association studies have not found significant associations. Although there are various explanations for the lack of replicated findings, it likely relates to the fact that GWAS is powered to detect common variants and that the genes relevant to depression risk are rare.

Candidate gene studies, which focus on specific genes of interest, have resulted in many reported findings. However, few of these are replicated. The most considerable focus has been on genes that might have a role in the proposed mechanism of depression, such as the gene involved in the encoding of serotonin (HTR1A), the serotonin transporter (5HTTP/SLC6A4), or the dopamine receptor (DRD4) and transporter (SLC6A3). Although some have replicated the most significant associations for these genes (most often for the 5HTTP/SLC6A4 gene), the effects sizes are small, and the sample sizes are smaller than those using GWAS, leading some experts to question whether these are false-negative findings. Suppose some or all of the candidate gene studies are significant. In that case, the most likely interpretation of available findings is that much, or most of, the genetic heritability for depression results from many genes, each contributing a small effect, which only becomes meaningful when combined across thousands of genes. It is also possible, however, that there is some rare, still undetected genetic variant.

Developmental Predisposition. Parents with mood disorders are often in conflict, leading to separation, divorce, and suicide. Heredity often determines the type of environment into which the child predisposed to mood disorder is born. Although not involved explicitly in causing mood disorder, developmental object loss might modify the expression of the illness, possibly by leading to earlier onset, more severe episodes, and an increased likelihood of personality disorder and suicide attempts. The serotonin transporter polymorphism mentioned earlier appears to mediate the relationship between early trauma and depression. Likewise, this polymorphism appears relevant to neuroticism and suicide attempts.

Temperament. Since ancient times, doctors have described persons prone to mania and melancholia as possessing specific temperamental attributes, representing variations on a theme now subsumed under cyclothymic, dysthymic, and anxious-inhibited temperaments, as well as the traits of high neuroticism describing emotionality. Many monozygotic twins discordant for full-blown mood disorders, studied by Aksel Bertelsen's Danish research team, exhibited affective instability with temperamental moodiness, which strongly suggests a genetic basis for these attributes. Research conducted by Kendler's team at the Medical College of Virginia further suggests that several temperamental attributes are transmitted as part of the genetic liability to mood disorders. Research has also identified such temperaments in the prepubertal offspring of parents with bipolar I disorders, suggesting that they precede by years to decades the overt onset of major mood disorder episodes. The atmosphere of high expressed emotion and the negative critical remarks by relatives and affectively unstable patients documented in the recent psychological literature on mood disorders often reflect the interpersonal clashes between patients and their temperamentally intense relatives. Thus, temperaments appear to be intimately involved in generating much interpersonal friction, emotional arousal, and sleep loss (just to cite common perturbations), thereby eliciting many of the life stressors that precipitate affective episodes. The use of stimulant drugs to self-treat lethargy or enhance hypomanic traits could further contribute to episode precipitation. As for the depressive disposition, the work of Maria Kovacs at the University of Pittsburgh has shown that dysthymia in children evolves into major depressive episodes postpubertally, of which a proportion switch to bipolar states. These data cohere with the work conducted at the University of Tennessee, Memphis, showing shortened REM latency in early-onset dysthymic subjects. The familial bipolar diathesis revealed in the Tennessee work and its tendency to switch to hypomania suggests commonalities between depressive and bipolar II disorders.

Life Events. Most individuals do not develop clinical depression when exposed to environmental adversity. Such adversity seems to play a pathogenic role primarily in those with an affective diathesis. The work of Kendler at the Medical College of Virginia indicates that genetic factors might underlie depressive disorder patients' susceptibility to life events. Furthermore, current data suggest that social stressors in the onset of depression relate more to the first few episodes of the illness. The evidence of this is less convincing for mania. At any rate, stressful events often appear to be triggered by the temperamental instability that precedes clinical episodes. Interpersonal losses are common events in the lives of individuals with intense temperaments. The arousal and sleep loss associated with such events can precipitate both depressive and manic states.

A recent study by Peter McGuffin's team at the Institute of Psychiatry, London, raised the possibility that one mechanism by which heredity produces depression is by creating environmental adversities in the lives of individuals predisposed to this illness. Independent groups of investigators have replicated this work. Whatever the origin of environmental adversity, it is a common clinical experience that loss represents an important, perhaps even central, theme in clinical depression. Variables that seem to modulate the effect of adult losses include concurrent life events, resultant changes in lifestyle, lack of interpersonal support, deficient social skills, and the symbolic meaning of the putative loss. The research program of George Brown and his followers in London capitalizes on the preceding considerations, particularly the importance of early and proximate losses in socioeconomically disadvantaged women who lack supportive relationships. However, that conceptualization downplays the degree to which the social context of the depression reflects the dysthymic temperamental liabilities of those depressed women. Recent research indicates that even social support is determined to a considerable degree by the genetic mechanisms that underlie mood disorders. Indeed, the short alleles of the serotonin transporter gene may mediate between adverse events and clinical depression.

Biologic Stressors. Many physical diseases and pharmacologic agents are known to precede the onset of both depressive and manic episodes. Like psychosocial stressors, however, other diseases and drugs do not generally seem to cause de novo episodes, but mobilize them in persons with a personal and family history of mood disorders. Thyroid disturbances have a role in practice because they are associated with rapid cycling in bipolar patients, especially women; lithium is often contributory to such disturbances occurring in the depressions in which bipolar women are not uncommonly “stuck.”

Neuroplasticity Hypotheses. Neuroplasticity hypotheses propose that the atrophy of already-developed neurons causes depression. In this version, rather than a deficit of nerve growth, atrophy occurs in already mature neurons. As noted, chronic stress can increase glucocorticoid levels, which can cause atrophy. It can also decrease BDNF expression, which is essential for the survival, growth, and differentiation of neurons in the brain; the result is atrophy. As this preferentially happens in the hippocampus, this can account for the finding of decreased hippocampal volume in depressed patients.

Sex. Clinical and epidemiologic studies agree that women are at higher risk for mood disorders, with the highest risk for depression. Despite many changes in the overall prevalence of depression, the female:male ratio of global disability from the disorder remains unchanged at 1.7:1. This ratio is consistent for developing and developed countries. The global consistency suggests an essential role for biologic sex differences instead of sociocultural and economic factors—the difference in rate starts at puberty and declines after age 65.

This observation appears to be, in part, a function of anxious-depressive traits represented by neuroticism. These traits have strong genetic determinants. Women have higher concentrations of MOA (the enzyme that breaks down monoamine transmitters) in the brain and more precarious thyroid status. In addition, low-estrogen and high-progesterone concentrations are possible mediating factors in postpartum depressions, premenstrual accentuation of affective instability, and women's vulnerability to the depressant effect of steroidal contraceptives. Finally, data point to the role of estradiol in depressions occurring during the transition to menopause.

Personality factors might also be relevant to the sex differences in depression. In collaborative work with University of Pisa psychiatrist Giulio Perugi, Akiskal proposed that female sex might favor the greater expression of dysthymic attributes, whereas hyperthymic traits appear to favor male sex. Those considerations tend to parallel the *ruminative* and *active* cognitive response styles reported by Susan Nolen-Hoeksema, originally at Stanford University, to distinguish the sexes. What specific sex-related biographical factors might interact with sex-related biologic factors to produce such trait differences is mainly unknown. An intriguing possibility is that women, because of temperamental inclination to depressive cognitions, might react more intensely to childhood adversities and be more specifically vulnerable to adult stressors related to bonding with men and child-rearing. Research by Mark George and colleagues has raised the provocative possibility that women over-respond to sad circumstances over a lifetime, thereby permanently altering anterior limbic and prefrontal brain function in a "depressive" direction. The high prevalence of anxious-depressive conditions in women—most pronounced at the clinical level—might be linked, hypothetically, to an evolutionary adaptive advantage conferred by traits of fear, inhibition, and avoidance to women who bear the responsibility of pregnancy and child-rearing. The higher prevalence of minor depressions in women, rediscovered in a recent Danish study, is in line with the previous hypothesis.

Integrative Approaches

It remains difficult, if not impossible, to incorporate all the various observations and research findings on depression into one grand unified theory. However, the various attempts by investigators and theoreticians to incorporate these findings have certain common features.

It seems clear that there exists some genetic vulnerability that puts persons at risk for depression. Such persons who then encounter stress, either external or internal, have epigenetic changes that differ from normal responses to stress; for example, they may be less able to transcribe certain growth factors in response to stress. This deficit results in subtle neuronal loss, and this loss is particularly apparent in some vulnerable areas (e.g., the hippocampus). The result is a disruption of the usual regulatory systems in the brain, causing changes in the expressions of transmitters, hormones, and other regulatory systems. This combination of effects causes the physical changes that we associate with depression and our behavioral response to this is the subjective feeling of depression.

Therapeutic Perspectives

The previous integrative model mandates the joint use of somatic-pharmacologic and psychosocial interventions. Although the milder forms of mood disorders are manageable with psychotherapy, somatic treatments are usually required to reverse the biologic disturbances in melancholia before the patient can respond to interpersonal feedback. Depressive disorders with psychotic features often necessitate more definitive somatic interventions, such as electroconvulsive therapy (ECT). Continued psychopharmacologic treatment is also effective in decreasing rates of relapse and future recurrence in most. Bipolar disorder and most forms of depression are considered lifelong illnesses that need indefinite maintenance pharmacotherapy.

Psychosocial therapy by skilled clinicians can provide support, combat demoralization, change maladaptive self-attributions, and improve conjugal and vocational functioning. Recent fascinating data support the notion that psychosocial interventions such as cognitive-behavioral therapy modulate cortical-limbic function described earlier as the final common pathway of depressive illness. Whether such therapy can also modify personality traits to fortify the patient against new episodes is a future research challenge. It might prove more profitable to attempt to help patients to explore professional and object choices that match their temperamental proclivities and assets, which, in turn, might provide them greater harmony and adaptation in life. Although we need to learn much about the indications for medication and psychotherapy in different subtypes of mood disorders, research not only does not support a negative interaction between the two forms of treatment but, on selected parameters, suggests additive and even synergistic interaction. There is a great need for patients, their families, and clinicians to understand how we can approach a biologically driven illness like depression from a pragmatic psychotherapeutic perspective.

The challenge for psychiatric research in the decade ahead is to elucidate the basic mechanisms by which the predisposing, precipitating, and mediating variables reviewed here and others yet to be identified interact to produce the final common path of decompensation in melancholia. Because of the heterogeneity of depressive conditions presenting as a psychobiologic final common clinical syndrome and because antidepressant agents, irrespective of specificity to one or another biogenic amine, are approximately equally effective in two-thirds of those with depressive disorders, the antidepressant agents may be acting not on the primary lesions of these disorders but a neurochemical substrate distal to the underlying biologic faults. Choice of antidepressants is still highly determined by the side effect profile least objectionable to a given patient's physical status, temperament, and lifestyle. That so many different classes of antidepressants—with different mechanisms of action—have been marketed since the 1990s represent indirect evidence for heterogeneity of putative biochemical lesions. The investigation of central neurotransmitter receptor function continues to occupy much current effort to delineate the mechanism of antidepressant action and side effects of classic agents, as well as the new compounds that have made the treatment of depression "clinician and patient friendly." Whether the study of specific receptors will unravel the molecular mystery of depression remains to be seen. Since the 1990s, studies have begun on antidepressant and mood-stabilizing effects on molecular mechanisms believed to be closer to the "genetic underpinnings" of mood disorders. Herein is the promise of the future—a new generation of psychiatrists conversant with both clinical phenomenology and molecular biology. Data suggest that the biologic specificity of genetic factors in mood disorders might translate into distinct temperamental dysregulations, which, in turn, might predispose to different affective subtypes.

Returning to the therapeutic arena, mounting clinical evidence indicates that, in a particular subgroup of depressed patients with bipolar disorder, antidepressants might provoke mixed episodes, hypomanic episodes, or both and possibly increase later cycling. The kindling-sensitization model suggests the utility of anticonvulsant medication on episode escalation and might represent yet another example of pathophysiologic intervention. Whatever the merit of this model, the 1990s witnessed intense clinical and research interest and U.S. Food and Drug Administration (FDA) approval of the clinical introduction of sodium valproate and lamotrigine for bipolar disorder, and many other promising anticonvulsants are being developed for that disorder. Anticonvulsant mood stabilizers appear to possess a broad spectrum of activity on bipolar disorders, including mixed dysphoric and rapid-cycling forms. Lithium, by contrast, seems more specific to euphoric, or “classic,” mania.

On the other hand, the introduction of atypical antipsychotics—such as olanzapine, risperidone, quetiapine, and aripiprazole—for mania raises intriguing questions about a common neural substrate for schizophrenic and manic disorders. Psychoeducational interventions geared to disturbed rhythms of the disorder represent another example of rational therapeutics. Mood clinics should help patients and their significant others dampen stimulation to keep it at an optimal level for depressed patients with cyclothymic traits. We should also aim to gradually eliminate all offending drugs (e.g., cocaine, caffeine, and sedative-hypnotic agents) and minimize circadian disruptions and sleep loss. The more significant challenge is to learn how to curb the ill-advised actions of patients with cyclical depressions. Unfortunately, current data indicate neurocognitive deficits, particularly in executive function, that underlie or complicate these disorders. Psychoeducation and psychotherapy have the task of ameliorating the resulting social problems, and despite much progress in this realm, these deficits remain stable over time. Compliance with mood-stabilizer regimens that, for many, would attenuate episodes and prevent such sequelae is challenging to achieve. We need further research on treatment or medication adherence techniques for promoting the more efficient use of mood stabilizers.

It is tempting to suggest that biogenic amines, the “humors” of modern psychobiology, play the same heuristic role as the ancient humors did for many centuries. The black humor, appropriately evoked in the construct of melancholia in DSM-5, may not have the same claim for etiologic relevance to depressive disorders as norepinephrine and serotonin, but at least it has a classic heritage. Dopamine, by contrast, may represent the sanguine humor that drives hypomanic temperaments and manic behavior. When we discover the genetic factors contributing to clinical depression and mania, in all likelihood, they will link to temperamental dispositions rather than full-blown affective disease phenotypes. The clinician will still need to interpret the myriad influences that affect such inclinations to produce disease in an individual patient—that is, fundamental scientific advances in mood disorders, rather than diminishing the role of practitioners, will increase it. Regrettably, despite destigmatization efforts and a rich armamentarium of therapeutic developments, the clinical care of affectively ill patients at the severe end of the spectrum continues to be grossly inadequate. More research will not improve this dismal situation unless we address severe mood disorders’ human and social dimensions with the requisite clinical and public health policies. Caring for patients with mental illness is a dimension distinct from evidence-based treatments along psychopharmacologic and psychotherapeutic lines: it requires allocating human resources and integrated mental health structures geared to the total patient.

In any discipline, scientific truth is a function of its technology, but understanding the phenomena under consideration is a philosophical temperament that seeks integration and the hope for a unified vision. Research into the causes and treatment of mood disorders has generated considerable recent data suitable for integrating theory and practice. Conceptualizing the origin and treatment of mood disorders can no longer be justified on the grounds of ideologic preference alone.

ACKNOWLEDGMENT

This chapter is a revision of a previous chapter written by Hagop S. Akiskal, M.D. Sadly, Dr. Asiskal passed away before the writing of this book.

FURTHER READINGS

Adams F, ed. *Aretaeus of Cappadocia: The Extant Works of Aretaeus, the Cappadocian*. Sydenham Society; 1856.

Akiskal HS. Dysthymia and cyclothymia in psychiatric practice a century after Kraepelin. *J Affect Disord*. 2001;62:17–31.

Akiskal HS, Akiskal KK. In search of Aristotle: temperament, human nature, melancholia, creativity and eminence. *J Affect Disord*. 2007;100:1–6.

*Akiskal HS, McKinney WT Jr. Depressive disorders: toward a unified hypothesis. *Science*. 1973;182:20–29.

Albert PR. Why is depression more prevalent in women? *J Psychiatry Neurosci*. 2015;40(4):219–221.

Badcock PB, Davey CG, Whittle S, Allen NB, Friston KJ. The depressed brain: an evolutionary systems theory. *Trends Cogn Sci*. 2017;21(3):182–194.

Berretta S, Pantazopoulos H, Lange N. Neuron numbers and volume of the amygdala in subjects diagnosed with bipolar disorder or schizophrenia. *Biol Psychiatry*. 2007;62:884–893.

Boku S, Nakagawa S, Toda H, Hishimoto A. Neural basis of major depressive disorder: beyond monoamine hypothesis. *Psychiatry Clin Neurosci*. 2018;72(1):3–12.

Brandon AR, Shivakumar G, Freeman MP. Depression: covering mood and vasomotor symptoms. *Curr Psychiatry*. 2008;7:39.

*Caspi A, Sugden K, Moffit TE, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*. 2003;301:386–389.

Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet*. 2018;391(10128):1357–1366.

Delay J. *Les Dèréglements de l'Humeur*. Presses Universitairesde France; 1946.

Depression Guideline Panel. *Depression in Primary Care (Clinical Practice Guideline, Number 5)*. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; 1993.

Flint J, Kendler KS. The genetics of major depression. *Neuron*. 2014;81(3):484–503.

Freud S. Mourning and melancholia. In: *Standard Edition of the Complete Psychological Works of Sigmund Freud*. Vol. 4. Hogarth Press; 1975.

Gershon ES, Hamovit J, Guroff JJ, et al. A family study of schizoaffective, bipolar I, bipolar II, unipolar, and normal control probands. *Arch Gen Psychiatry*. 1982;39:1157–1167.

*Goodwin FK, Jamison KR. *Manic-Depressive Illness*. 2nd ed. Oxford University Press; 2007.

Hamilton JP, Etkin A, Furman DJ, Lemus MG, Johnson RF, Gotlib IH. Functional neuroimaging of major depressive disorder: a meta-analysis and new integration of base line activation and neural response data. *Am J Psychiatry*. 2012;169(7):693–703.

Hariri AR, Mattay VS, Tessitore A, Kolachana B, Fera F. Serotonin transporter genetic variation and the response of the human amygdale. *Science*. 2002;19:400–403.

Kendler KS. Social support: a genetic-epidemiological analysis. *Am J Psychiatry*. 1997;154:1398–13404.

Klerman GL, Lavori PW, Rice J, et al. Birth-cohort trends in rates of major depressive disorder among relatives of patients with affective disorder. *Arch Gen Psychiatry*. 1985;42:689–693.

*Kraepelin E. *Manic-Depressive Insanity and Paranoia*. Livingstone; 1921.

Kupfer DJ. REM latency: a psychobiologic marker for primary depressive disease. *Biol Psychiatry*. 1976;11:159–174.

Kuswanto C, Chin R, Sum MY, et al. Shared and divergent neurocognitive impairments in adult patients with schizophrenia and bipolar disorder: whither the evidence? *Neurosci Biobehav Rev*. 2016;61:66–89.

Kuswanto C, Chin R, Sum MY, et al. The psychoses dissected: the kraepelinian dichotomy, a continuum or one disorder? Program No. S4. Annual Meeting of the American College of Neuropsychopharmacology, Boca Raton, FL; 2007.

Lewinsohn PM, Youngren MA, Grosscup SJ. Reinforcement and depression. In: Depue RA, ed. *The Psychobiology of Depressive Disorders; Implications for the Effects of Stress*. Academic Press; 1979.

Lewis A. States of depression: their clinical and aetiological differentiation. *Br Med*. 1938;J2:875.

Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of depression in the community from 30 countries between 1994 and 2014. *Sci Rep*. 2018;8(1):2861.

Malhi GS, Byrow Y, Fritz K, et al. Mood disorders: neurocognitive models. *Bipolar Disord*. 2015;17(Suppl 2):3–20.

Merikangas KR, Akiskal HS, Angst J, et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2007;64:543–552.

Murray CJL, Lopez AD, eds. *The Global Burden of Disease*. World Health Organization; 1996.

Nelson JC, Baumann P, Delucchi K, Joffe R, Katona C. A systematic review and meta-analysis of lithium augmentation of tricyclic and second generation antidepressants in major depression. *J Affect Disord*. 2014;168:269–275.

Perugi G, Musetti L, Simonini E, Piagentini F, Cassano GB. Gender mediated clinical features of depressive illness: the importance of temperamental differences. *Br J Psychiatry*. 1990;157:835–841.

Prange AJ Jr, Wilson IC, Lynn CW, Alltop LB, Stikeleather RA. L-tryptophan in mania: contribution to a permissive hypothesis of affective disorders. *Arch Gen Psychiatry*. 1974;30:56–62.

Results from the 2017 National Survey on Drug Use and Health: detailed tables, SAMHSA, CBHSQ. (n.d.). Retrieved April 1, 2019. https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHDetailedTabs2017/NSDUHDetailedTabs2017.htm#tab8-56A

Schildkraut JJ. The catecholamine hypothesis of affective disorders: a review of supporting evidence. *Am J Psychiatry*. 1965;122(5):509–522.

Seligman MEP. *Helplessness: On Depression, Development and Death*. Freeman; 1975.

Torres F, Barbosa M, Maciel P. Recurrent copy number variations as risk factors for neurodevelopmental disorders: critical overview and analysis of clinical implications. *J Med Genet*. 2016;53(2):73–90.

Zaninotto L, Solmi M, Toffanin T, Veronese N, Cloninger CR, Correll CU. A meta-analysis of temperament and character dimensions in patients with mood disorders: comparison to healthy controls and unaffected siblings. *J Affect Disord*. 2016;194:84–97.

13.2 Mood Disorders: Epidemiology

KATHLEEN RIES MERIKANGAS, PH.D., AND ZOLTÁN RIHMER, M.D., PH.D., D.SC.

Application of the tools of epidemiology to psychiatry over the past several decades has led to numerous advances in our understanding of the magnitude, correlates, patterns of comorbidity, course, impact, and service patterns of mood disorders. Recent contributions of epidemiology continue to be both methodologic developments and substantive knowledge derived from shifting the focus of research from clinical to population-based samples. Methodologic aspects still focus on defining clinical phenomenology including the overlap between mood disorder subtypes, the core features such activation as a core feature of mania, and the spectrum of manifestations of mood disorders. Substantive developments have emerged from the expansion of epidemiologic research from North America, Europe, and developed pan-Pacific countries to less economically advantaged countries across the world that provide global estimates of the tremendous burden of mental disorders and the glaring gap in access to treatment. There also have been advances in documenting medical mental comorbidity that has been facilitated by the increased use and availability of data from electronic health records. Integration of genetic and environmental risk factors is being accomplished with more sophisticated methods that may inform our ability to intervene by preventing these conditions. Increased awareness of diversity of gender, age, ethnicity, and class is leading to increased efforts to address differences in risk profiles in population strata.

This section will provide (1) an update prevalence rates of mood disorder subtypes from international studies, (2) a summary of the demographic and social correlates of mood disorder subtypes, (3) changes in service patterns for mood disorders in general community samples, and (4) update evidence on the individual and global impact of mood disorders.

DESCRIPTIVE EPIDEMIOLOGY

Bipolar Disorder

The largest series of international studies that provide prevalence rates of mood disorders is the World Health Organization (WHO) World Mental Health (WMH) surveys. Data on BPD was collected in a subset of 11 population-based surveys carried out in the Americas (São Paulo metropolitan area, Brazil; Colombia; Mexico; and the United States), Europe (Bulgaria and Romania), Asia (Shenzhen, People's Republic of China; Pondicherry region, India; and nine metropolitan areas in Japan), Lebanon, and New Zealand. Apart from Japan (unclustered 2-stage probability sample), all surveys were based on stratified multistage clustered area probability samples. In the past few decades, the lifetime prevalence of bipolar I disorder (BPI) has generally been estimated at about 1.0%. The most recent estimates from the largest population-based survey, the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) have reported substantially greater lifetime prevalence estimates of BPI of 3.3, bipolar II (BPII) of 1.1 with corresponding 12-month rates of 2.0 and 0.8, respectively. Recent studies have begun to include estimates for BPII and bipolar spectrum (BPS) disorders as well as BPI. As would be expected, prevalence tends to increase with successively more inclusive disorder definitions. For example, in the WMH Surveys, the cross-national lifetime prevalence of BPS disorders was 2.4%; of which 0.6% met criteria for BPI, 0.4% for BPII, and 1.4% for subthreshold bipolar disorder (sub-BPD) In many (but not all) studies assessing BPII, its prevalence is actually lower than that of BPI; this may indicate that the majority of people who experience hypomanic episodes also experience manic episodes, or may be explained by the major depressive episode (MDE) requirement for diagnosis of BPII.

An increasing number of studies have estimated the occurrence of BPD disorder in youth. Lifetime prevalence rates of BPD in youth range from <1% to 2.9% for BPI/II. A few studies have also estimated 12-month prevalence rates of bipolar disorder, ranging from 1.3% to 2.5%., suggesting that bipolar disorder tends to be enduring. Prevalence rates of mania range from 0.4% (12 months) to 2.0% (12 months; and rates of hypomania range from 0.1% (Lifetime) to 0.9% (6-month). The results of longitudinal studies converge in estimating the prevalence of bipolar disorder at between 1.4% and 2.1%, which approximates cross-sectional prevalence rates in adult samples. Meta-analysis of epidemiologic studies of BPD in children and adolescents yielded a mean prevalence of 1.8%.

Major Depression

The WMH surveys also provided the largest aggregate data on the prevalence of major depressive disorder (MDD). As shown in Figure 13.2–1, lifetime and 12-month prevalence rates were estimated in 18 countries, divided according to high- and low-middle income. The lifetime prevalence estimates average 11.1 (range 8.0 to 18.4) in low and 14.6 (range 6.6 to 21.0) in high-income countries, whereas the 12-month prevalence rates average 5.5 in high (range 2.2 to 8.3) and 5.9 (range 3.8 to 10.4) in low-income countries. More recent prevalence estimates from the NESARC study were 13.2 for lifetime and 5.3 for 12-month MDD. Although there is a wide range of estimates, the average rates of both lifetime and 12-month depression are consistent across studies that employ comparable methodology.

Several cross-sectional and prospective studies have also reported rates of MDD in youth. A recent meta of 41 studies conducted in 27 countries from every world region yielded a 12-month estimate of any depressive disorder of 2.6% (CI 95% 1.7–3.9). The prevalence of MDD increases significantly across adolescence, more marked among females than males. Most cases of MDD are associated with psychiatric comorbidity and severe role impairment, and a substantial minority reported suicidality. Treatment in any form was received by most adolescents with 12-month DSM-IV MDD (60.4%), but only a minority received treatment that was disorder-specific or from the mental health sector.

Numerous studies have begun to expand the diagnostic criteria for mental disorders to collect information on the spectra of expression of conditions. With respect to mood disorders, the spectrum concept of bipolar disorder long advocated by clinicians has now been validated in community samples of adults and children. Application of the spectrum concept of BPD has yielded rates of bipolarity of up to 5% and depression of up to 20%. Likewise, many individuals in the community may show some (a few or more) depressive symptoms that do not reach the severity or duration threshold for specific mood disorders in the DSM-5 system but nevertheless, also have substantial morbidity and dysfunction. Although these subthreshold disorders may be less severe forms of major depressive or bipolar disorders, they may also cause substantial suffering as well as disability. Subtypes of subthreshold depression identified in the Zurich Cohort Study and replicated in numerous subsequent studies include recurrent brief depressive disorder and minor depressive disorder (as well as recurrent brief hypomania).

CORRELATES OF MOOD DISORDERS

Sex/Gender

The most consistent finding across all the studies on the prevalence and incidence of unipolar major depression is that it is approximately twofold more common among women than men. This gender difference begins in early adulthood, is most pronounced in people between the ages of 30 and 45, and persists in the older adults. Because there are no data (aside from biologic–hormonal differences) that show that female gender per se means increased vulnerability for depression, increased stress sensitivity, maladaptive coping strategies, and multiple social roles (all of which are frequently seen in women), and substance use disorders that can mask depressive symptoms (more frequently seen in men) have been suggested for the explanation of the gender difference. In addition to these psychosocial theories, recent studies showed that because prior anxiety disorders are more common in women, preceding anxiety disorder may also be a significant factor contributing to the higher depressive morbidity in women. Thus, gender difference in unipolar major depression seems to result from the complex interaction of the mentioned biologic and psychosocial variables. Minor depressive disorder and recurrent brief depressive disorder are also more common among women, but the difference is not so marked as that in major depression.

In contrast to MDD, the gender ratio in BPD (all subtypes combined) is approximately 1:1. However, among bipolar II patients and in special subpopulations (mixed/dysphoric mania, mixed depressive episode, winter depression, bipolar depression with atypical clinical features, rapid cycling BPD), women are overrepresented. Looking at the depression–mania continuum as a whole spectrum, there is a clear trend: the higher the depressive component, the higher the proportion of women. Consequently, in the rare cases of unipolar mania (manic episode without any major or minor depression), men are markedly overrepresented. The gender differences in lifetime prevalence rates are more marked than in 1-year and current prevalence rates, which may be attributable—at least in part—to the stronger male tendency to forget previous episodes and deny emotionally negative events. Recent population-based epidemiologic surveys showed that the lifetime prevalence and 1-year prevalence of MDD, dysthymia, and bipolar (I + II) disorders were much higher among people with same-sex sexual behavior, (including bisexuality), particularly in the case of men.

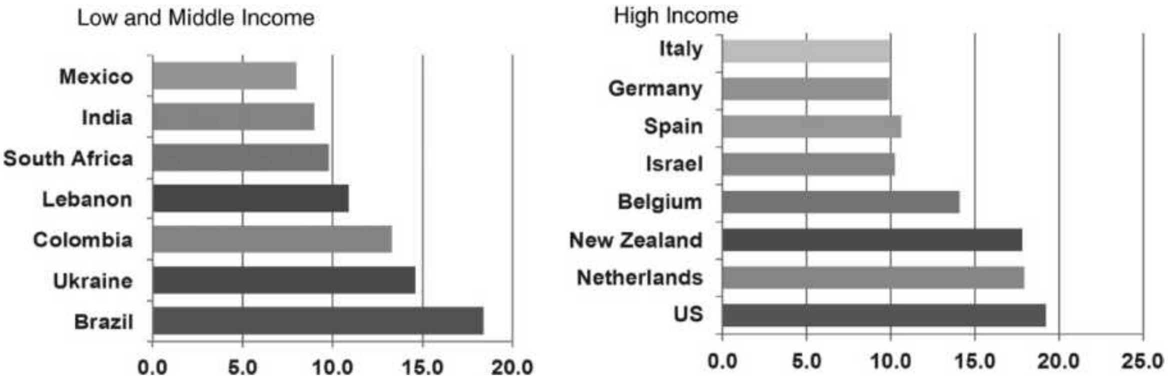

FIGURE 13.2–1. Lifetime prevalence of major depression in the world mental health surveys. (Data from Merikangas KR, Jin R, He JP, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch Gen Psychiatry*. 2011;68(3):241–251.)

Sexual orientation (lesbians, gay people, bisexual, and transgender [LGBT]) as a potential risk factor for mood disorders has also been increasingly recognized. People from the LGBT community are frequently not just victims of stigmatization, aggression, and physical/sexual abuse, but they are vulnerable to many major mental disorders, including BPD and MDD, anxiety as well as substance-use disorders. LGBT people have also been thought to be at elevated risk of mood disorders as well as suicide attempts and to lesser extent completed suicide. The elevated risk of suicidal behavior of transgender persons does not further increase after sex reassignment (hormonal and/or surgical) therapy. In addition to major suicide risk factors (including elevated rate of mood and anxiety disorders), commonly present in these subgroups, stigmatization, resulting in psychological discomfort is a central contributory factor.

Age

The higher lifetime prevalence of depression is much greater among people younger than 45 years, but the age of onset differs between unipolar and bipolar and familial and nonfamilial cases. The average age of onset of recurrent MDE falls between the ages of 30 and 35 years, whereas single-episode major depression that often lacks family history of mood disorders usually begins some years later. The genetic predisposition for depression decreases with age. However, social stressors appear to place younger individuals at a greater risk for depression than older adults. On the other hand, isolation, loss of interpersonal contacts, medical disorders, and disability play a more important role in the development of depression in later life. Early-onset depression is associated with a higher female-to-male ratio than late-onset depression. Because MDD is a frequent and highly recurrent illness, the probability of recurrence does not decrease with age. The incidence of MDD in old age is lower in both sexes, but first incidence and prevalence of minor depressive disorder show the opposite trend. The incidence of unipolar MDD in the postpartum period is just slightly increased, but this is not the case in the postmenopausal years. Pure (primary) dysthymic disorder typically starts in late adolescence or early adulthood, but, in the absence of appropriate treatment, almost all cases develop later into major depression. The first incidence of dysthymia is also common in old age, but this is typically the consequence of the adverse psychosocial and biologic conditions frequently seen among the older adults.

The age of onset of BPD, most commonly around 20 years of age, is substantially (about 10 years) lower than that of unipolar depression. Men with BPD appear to have a 4 to 5 years earlier onset than bipolar women. In more than half of the cases, onset is before the age of 20, frequently in late adolescence. In contrast to this, first-onset mania is very rare among older adults. The ages of onset of BPI and BPII disorders are similar, but there is a slight tendency for higher age of onset in BPII patients. The incidence of the depressive phase of BPD after childbirth is relatively high, and most patients with “postpartum depression” have bipolar disorder. Like MDD, people with BPD with positive family history of mood disorders are significantly younger at the beginning of the illness and need less stressors to precipitate the illness than those lacking such history.

Race and Ethnicity

It is well-known that cultural differences may influence the clinical presentation of depression and mania, leading to unreliable figures in prevalence studies performed by standard methodology. However, there is now consistent evidence that rates of mood disorders are lower in blacks and Hispanic than in whites. Because of their increased exposure to psychosocial stress and other risk factors for mood disorders, the lower rate in these ethnic subgroups is paradoxical.

Differences in mood disorders in immigrants have been of growing interest because of their implications for both treatment and prevention. European research suggests that immigrants are at greater risk of mental illness, and suicide than are nonimmigrants. Whereas immigrants into the United States are significantly *less* likely than U.S.-born individuals to meet criteria for a range of mental disorders, their offspring often have elevated rates of these conditions. With the increases in worldwide immigration, studies of differential risk to native-born will be a powerful way to identify cultural and environmental exposures related to mood disorders.

SOCIAL CORRELATES

Marital Status

The relationship between marital status and mood disorders is quite complex. For example, being single, divorced, or separated can be either a risk factor for depression or the result of the adverse life events generated by depressive or manic psychopathology, or both. MDD and BPD are most frequent among divorced, separated, or widowed individuals. Single women have lower rates of depression than married women do, but the opposite is true for men. However, being single because of having never married, because of the dissolution of a difficult marriage, and because of widowhood represent three very different conditions. The risk of a MDE is very high among recently widowed individuals of all ages but is particularly high in the older adults. Patients with mood disorder are over-represented among the divorced, and the rate of family breakdown (separation, divorce) is elevated slightly in dysthymic patients, substantially in major depressive patients, and markedly in bipolar I and bipolar II patients. The presence of mania, hypomania, or major depression is a strong predictor for future separation or divorce, which can cause serious distress for the patients and for their spouses and may generate negative life events for their children. These early negative life events (e.g., parental loss before adolescence) are well-known predisposing factors for adult mood disorders, particularly in the case of family loading (case of subjects with positive family history of mood disorders).

Socioeconomic Factors

Although the relationship between depressive symptoms and low social class is well-documented, most studies found only a weak (but consistent) correlation between MDD and BPI with lower socioeconomic status. Individuals with lower socioeconomic status have a lower level of educational, lower income, and poorer living conditions, as well as a higher rate of unemployment and, ultimately, homelessness. The proportion of MDE was approximately three times higher among individuals without a workplace than among those with one. However, as in the case of marital status and mood disorder, cause and effect may be reversed here, too. Mood disorder can easily lead to unemployment, divorce, or low income, resulting in regression on the social hierarchy scale. Because hypomania is not as disruptive as mania, in terms of academic and social carrier, the educational level of BPII patients is above the average and, in contrast to those with MDD, people with BPI and BPII tend to belong to higher social classes and are relatively overrepresented among socially active, creative people.

Residence

As urban communities are more stressful than rural ones, it is not surprising that most studies performed in Western societies concluded that major depression was more frequent in urban residents than in their rural counterparts. The urban–rural distinction tells us little about the real living and social conditions in general, but it can be a good marker for the density of the population and for other important sociologic variables.

Seasonal Factors

Even though more than two-thirds of patients with recurrent major mood disorders show irregular seasonal patterns individually, statistically, spring and fall are the peak times for depression, just as summer is for mania. Because seasonal affective disorders (fall–winter depression and spring–

summer depression) occur in approximately 20% to 25% of the patients with recurrent major mood disorders, it is possible that the seasonal pattern observed in MDD and BPD, in general, is the consequence—at least in part—of the characteristic annual rhythms of these specific seasonal subtypes. The seasonal profiles of committed and attempted suicides, the prescription of antidepressants and ECT, and the availability of ʟ-tryptophan (the main precursor of serotonin) are very similar to the seasonal onset of MDD. However, acute, and long-term pharmacotherapy of mood disorders (more precisely, its cessation) can change the seasonal pattern of depression and mania, which should be considered either in planning studies or in the interpretation of the results. On the other hand, the decreasing seasonal variation in suicides that has been observed in many countries in the last two decades may reflect, among others, the reduced rate of depression-related suicides in the given population.

There has been growing interest in circadian rhythms that underlie the temporal dynamics of suicide in terms of seasonality as well as shorter-term fluctuations. There has been an increasing focus on motor activity, now a core criterion for mania, as well as it’s inter-relationships with sleep and mood regulation, particularly with the increasing availability of mobile tools to track these systems in real-time.

Geographic Trends

There is a general, but weak, trend for lower prevalence of depression and higher rate of mania in regions located closer to the equator. Consistently, at least in the Northern Hemisphere, winter depression (which affects between 1% and 6% of the community) seems to be more frequent in countries situated farther from it. A significant positive correlation was found between prevalence of winter depression and latitude in North America (where its prevalence is twice as high as in Europe), whereas a similar, but only slight, tendency was observed in Europe. On the other hand, the distribution of summer depression across the latitudes shows the opposite tendency. However, the relationship between these two forms of seasonal affective disorder and latitude (e.g., the length of daily photoperiod) is weak. Other climatic and genetic influences, as well as social and cultural factors—which can also be interrelated with the daily photoperiod—may play a role, too. Regarding the West–East dimension, the lower prevalence rates of MDD and BPD reported from Far Eastern countries may be related primarily to cultural differences and methodologic shortcomings, but there is some recent evidence indicating that the prevalence of major mood disorders in these countries may, in fact, be lower.

PSYCHOSOCIAL FACTORS

Social Stressors

Social stressors, in general, have been well-recognized as risk factors for mood disorders. However, different kinds of social stressors (i.e., childhood vs. adulthood events, acute vs. chronic stressors, positive vs. negative life events) can play different roles in the predisposition and precipitation of depressive or manic disorders. In the case of major depression and bipolar disorder, the association of acute stressors and the onset of illness become progressively weaker with the increasing number of previous episodes, and patients at high-genetic risk for mood disorders commonly experience depressive or manic episodes without any negative life event. In the development of major mood disorders, chronic stressors (e.g., unemployment, difficult marriage) play a more important role than specific, acute stressors. However, accumulation of stressful negative life events is the strongest predisposing factor. The higher level and, probably, the different nature of social stressors in individuals living in urban communities may be among the main sources accounting for their higher psychiatric morbidity. The fact that subjective perception of life events is more important than the event itself makes the estimation of the causative role of life events more difficult. On the other hand, it has also been demonstrated that acute, positive life events (that are quite rare in Western communities) can also precipitate either major depression or mania in vulnerable individuals. Negative life events should not be considered as only predisposing or precipitating factors, as they are frequently the result of the behavior of patients with MDD and, particularly, patients with bipolar disorder.

Social Support

Social support can improve coping and modify (e.g., reduce or eliminate) the occurrence of psychosocial stressors or the adverse consequences of them. The consistent finding in the literature is that living alone, having low socioeconomic status, and being unemployed are significant risk factors for mood disorders. In other words, it means that weak or lacking social support (including social network, social interaction, and instrumental support) can also be considered a major risk factor. However, regarding social interactions, the frequency of the interactions is more important than the amount. Poor social support is related to onset, relapse, and recurrence of depression, but there is no evidence for the excess of depression among women is caused by reduced social support.

COMORBIDITY

Individuals with major mood disorders are at an increased risk of having one or more additional comorbid Axis I disorders. The most frequent disorders are alcohol abuse or dependence, panic disorder, obsessive-compulsive disorder (OCD), and social anxiety disorder. Conversely, individuals with substance use disorders and anxiety disorders also have an elevated risk of lifetime or current comorbid mood disorder. In both unipolar and bipolar disorder, men more frequently present with substance use disorders, whereas women more frequently present with comorbid anxiety and eating disorders. In general, bipolar patients more frequently show comorbidity of substance use and anxiety disorders than do patients with unipolar major depression. Both clinical and epidemiologic studies show that, if BPII disorder is also considered, comorbidity of substance use disorders and anxiety disorders is the highest in the BPII subgroup. Because comorbid substance use disorders and anxiety disorders worsen the prognosis of the illness and markedly increase the risk of suicide among those with MDD and BPD, the data presented above are consistent with previous findings showing that those with BPD (particularly BPII) are at the highest risk of suicide. These findings also indicate that Axis I comorbidity might be among the most important contributing factor to that risk. However, there is a substantial overlap between mood and substance use disorders. Data from the NESARC study revealed that 41% of those with alcohol use disorders met criteria for primary depression, 17% for concurrent depression, and 42% for secondary depression. This finding supports and extends the results of previous studies conducted among alcoholics in treatment and has strong diagnostic and treatment implications.

Untreated major mood disorders (and particularly BPD) carry extremely high risk of both completed and attempted suicide. This is partly because they use more violent/more lethal methods. If patients with major mood disorders attempt or commit suicide, they do it mostly during their severe MDE and less frequently in mixed affective episodes or dysphoric mania but very rarely during euphoric mania and euthymia suggesting that suicidal behavior in patients with major mood disorder is a “state-dependent” and severity-dependent. Consequently, the early recognition and appropriate acute and long-term treatment of mood disorders is crucial for suicide prevention. The careful and systematic exploration of clinically detectable suicide risk factors helps clinicians to identify suicidal patients but unfortunately, at the stage of our present knowledge we cannot predict and prevent all suicides.

There is growing interest in studies of comorbidity of mood disorders with medical conditions. Systematic collection of data on chronic conditions in the WMH Surveys provided systematic evaluation of patterns of comorbidity of mental disorder subtypes with the full range of common medical disorders, their impact on severity and course. Prospective studies have continued BPD. There is growing effort to identify biomarkers, particularly inflammatory/immunologic factors-related to the stress system, as potential explanations for comorbidity of mood disorders with cardiovascular disease, diabetes, and cancer. However, it has not been possible to distinguish whether there may underlying common pathophysiologic factors, or the comorbidity was attributable common risk factors, such as smoking, sedentary behavior, alcohol and drug use, and other unhealthy behaviors that are elevated among people with mood disorders.

SERVICE PATTERNS

Adults

Despite the great progress in the diagnosis and treatment of depressions in the last two decades, they remain underdiagnosed and undertreated. North American and European surveys show that approximately half of those who develop mood disorders seek treatment for them, but only a small fraction (about one-third) of recognized depressives receive appropriate treatment. The results of the NESARC study revealed that only 36.8% of those with a current mood disorder had sought disorder-specific treatment for a mood disorder. Interestingly, people with manic episodes were less likely than those with MDE to seek disorder-specific treatment. An increasing rate of health service utilization is related to increasing severity of depression and to comorbid psychiatric and medical disorders.

Even though many patients with depressive disorders seek help in primary care, general practitioners still have difficulties recognizing and treating depression. The current prevalence of MDD in primary care is approximately 10% to 15% and that of dysthymic disorder is approximately 6% to 8%. Paradoxically, because severe major depression is much more common among those with comorbid chronic medical disorders, depressions with significant somatic comorbidity remain unrecognized in primary care. Population-based studies also show that in individuals with two or more chronic physical illnesses, the 1-year prevalence of severe major depression is fourfold higher than that of individuals without such conditions. Women more often seek treatment for their depression and are more compliant with the treatment than are men, but, unfortunately, “male depression” is less frequently recognized. Because there is no gender difference in the response to antidepressant pharmacotherapy, these facts can help to explain the so-called suicide paradox—that is, even though major depression is much more common in women, it is men who are markedly over-represented among suicide victims. Besides general practitioners, other professionals (i.e., internists, cardiologists, neurologists) also frequently see depressed patients. The point prevalence of MDD in acute (medical–surgical hospital) care is also higher than 10%, a figure well more than the point prevalence rate for the general population. Concomitant depression increases the morbidity and mortality from concurrent medical illness, and patients with simultaneous medical disorder and depression are less compliant with treatment and take longer to recover than nondepressed medical patients.

People with MDD, particularly those who go unrecognized, frequently seek general health care, but depression, if correctly identified, can be effectively treated even in primary care. On the other hand, even though more than half of suicide victims contact different levels of health care 4 weeks before committing suicide, the rate of adequate (antidepressive) pharmacotherapy among depressed suicide victims is disturbingly low. However, epidemiologic and clinical studies, as well as the steadily increasing use of antidepressants (the latter is accompanied by a constant decline in suicide rates in North America, and in Europe, particularly in those countries with traditionally high-suicide rates), indicate that, in the last 20 to 25 years, the referral, recognition, and successful treatment of depression increased, suggesting that the prognosis for depressive illness, in general, is improving.

Youth

Despite the magnitude and serious consequences of depression in youth, only about one-quarter to less than half of those with mental disorders receive mental health services. Factors associated with service utilization include ethnicity, high-global impairment, comorbidity, prior history of depression, suicide attempt, and impact of the child's problem on the family. School services are the most common point of entry for children seeking services, although those who enter through the education sector are least likely to transition to specialty mental health services. The actual diagnostic process and services provided differ dramatically by the context of entry to service.

There has been accumulating controlled research on the effects of antidepressants in youth, but few studies of either the comparative efficacy of various agents that are commonly used in clinical practice or of individual and family therapies either in conjunction with or independent from drug treatments. There is now growing evidence for a stepped-care approach with increasingly intensive interventions moving from psychosocial to pharmacologic interventions, as well as their combinations.

IMPACT OF MOOD DISORDERS

The importance of role disability has become increasingly recognized as a major source of indirect costs of illness because of its high economic impact on ill workers, their employers, and society. The introduction of the concept of Disability Adjusted Life Years, which estimate the disease-specific reduction in life expectancy attributable to disability and increased mortality has highlighted the dramatic public health impact of mental disorders. By the year 2020, it is estimated that psychiatric and neurologic disorders will account for 15% of the total burden of all diseases. MDD is the leading cause of disability among those age 5 and over, and the second leading source of disease burden surpassing cardiovascular diseases, dementia, lung cancer, and diabetes.

In the most recent estimates of the impact of chronic diseases by the WHO, 2.5% of the total proportion of DALYs was attributable to MDD, and 0.5% to BPD. In terms of DALYs for Mental, Neurologic, and Substance Use Disorders, 24.5% was attributable to MDD and 5% to BPD. Comparative studies of role disability reveal that the effects of mood disorders outweigh those of most other mental and neurologic disorders with a few exceptions. The dramatic impact of mood disorders on lifetime disability highlights the importance of epidemiology in surveillance, understanding, and control of the major mental disorders.

The global burden of mental disorders in children and adolescents up to 24 years of age has also been examined. Like studies within the United States., there is a broad range of estimates of the rates of mental disorders. A recent review reported estimates that ranged from 8% (in the Netherlands) to 57% (for young people receiving services in five sectors of care in San Diego, California, USA). Results from The Australian National Survey of Mental Health and Well-being showed that at least 14% of adolescents younger than 18 years had a diagnosable mental or substance use disorder within the previous 12 months, and this figure rose to 27% in the 18- to 24-year age group. Meta-analysis of the childhood and adolescent data has shown that at least one out of every 4 to 5 young people in the general population will suffer from at least one mental disorder in any given year. However, there is much less information on the burden of mental disorders in developing countries and substantial cross-cultural variations are likely.

Mortality and Mood Disorders

There are now several prospective studies have investigated mortality among those with a range of mental disorders. Data from large population registries in Sweden and Denmark have shown that BPD is associated with substantial elevation in risk of both mortality in general, particularly death by suicide (see Chapter Pompili and Rihmer: Mood Disorders: Suicidal Behavior). The mortality risk associated with mood disorders tends to decrease with age; that is, the high-risk of death by suicide associated with affective disorders in the Danish Registry was much greater during the first 12 months following the initial admission compared to later years. No differences in excess mortality have been shown for different subgroups of mood disorders in either of these registries. A more recent prospective population-based study with direct examinations of both psychiatric and medical conditions, the CoLaus|PsyCoLaus study found that current Major Depression was associated with a nearly threefold increase in mortality after adjustment for socio-demographic and lifestyle characteristics, comorbid anxiety disorders, antidepressant use, and cardiovascular risk factors and diseases.

HISTORICAL TRENDS

Whereas early community surveys suggested a strong cohort effect for mood disorders, both retrospective and prospective studies show that there has been no direct evidence for a marked increase in the incidence and prevalence of depression over the past decades. The observed changes are mainly due to artifacts of memory and, to a lesser degree, to higher subjective awareness, better definitions, and methods of case identification, redistribution by age and sex (with higher rates among younger women being of recent origin), and increase in childhood-onset mood disorders. Similar findings have emerged for depression in youth with comprehensive meta-analyses showing that there has been no increase in rates of depression across the last three decades.

FURTHER READINGS

Akinhanmi MO, Biernacka JM, Strakowski SM, et al. Racial disparities in bipolar disorder treatment and research: a call to action. *Bipolar Disord*. 2018;20(6):506–514.

Angst J, Ajdacic-Gross V, Rossler W. Bipolar disorders in ICD-11: current status and strengths. *Int J Bipolar Disord*. 2020;8(1):3.

Angst J, Merikangas KR, Cui L, Van Meter A, Ajdacic-Gross V, Rossler W. Bipolar spectrum in major depressive disorders. *Eur Arch Psychiatry Clin Neurosci*. 2018;268(8):741–748.

Angst J, Rossler W, Ajdacic-Gross V, et al. Differences between unipolar mania and bipolar-I disorder: evidence from nine epidemiological studies. *Bipolar Disord*. 2019;21(5):437–448.

Bromet E, Andrade LH, Hwang I, et al. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med*. 2011;9:90.

Carpenter JS, Crouse JJ, Scott EM, et al. Circadian depression: a mood disorder phenotype. *Neurosci Biobehav Rev*. 2021;126:79–101.

Cuijpers P, Karyotaki E, de Wit L, Ebert DD. The effects of fifteen evidence-supported therapies for adult depression: a meta-analytic review. *Psychother Res*. 2020;30(3):279–293.

Faedda GL, Marangoni C, Serra G, et al. Precursors of bipolar disorders: a systematic literature review of prospective studies. *J Clin Psychiatry*. 2015;76(5):614–624.

Ferrari AJ, Charlson FJ, Norman RE, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med*. 2013;10(11):e1001547.

Ferrari AJ, Somerville AJ, Baxter AJ, et al. Global variation in the prevalence and incidence of major depressive disorder: a systematic review of the epidemiological literature. *Psychol Med*. 2013;43(3):471–481.

Hasin DS, Sarvet AL, Meyers JL, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry*. 2018;75(4):336–346.

Herrman H, Kieling C, McGorry P, Horton R, Sargent J, Patel V. Reducing the global burden of depression: a Lancet-World Psychiatric Association Commission. *Lancet*. 2019;393(10189):e42–e43.

Kessler RC, Bromet EJ. The epidemiology of depression across cultures. *Annu Rev Public Health*. 2013;34:119–138.

Konig H, Konig HH, Konnopka A. The excess costs of depression: a systematic review and meta-analysis. *Epidemiol Psychiatr Sci*. 2019;29:e30.

Mekonen T, Chan GCK, Connor JP, Hides L, Leung J. Estimating the global treatment rates for depression: a systematic review and meta-analysis. *J Affect Disord*. 2021;295:1234–1242.

Merikangas KR, Jin R, He JP, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch Gen Psychiatry*. 2011;68(3):241–251.

Merikangas KR, Lamers F. The ‘true’ prevalence of bipolar II disorder. *Curr Opin Psychiatry*. 2012;25(1):19–23.

Merikangas KR, Swendsen J, Hickie IB, et al. Real-time mobile monitoring of the dynamic associations among motor activity, energy, mood, and sleep in adults with bipolar disorder. *JAMA Psychiatry*. 2019;76(2):190–198.

Plans L, Barrot C, Nieto E, et al. Association between completed suicide and bipolar disorder: a systematic review of the literature. *J Affect Disord*. 2019;242:111–122.

Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *J Child Psychol Psychiatry*. 2015;56(3):345–365.

Ramsey CM, Spira AP, Mojtabai R, Eaton WW, Roth K, Lee HB. Lifetime manic spectrum episodes and all-cause mortality: 26-year follow-up of the NIMH Epidemiologic Catchment Area Study. *J Affect Disord*. 2013;151(1):337–342.

Remes O, Mendes JF, Templeton P. Biological, psychological, and social determinants of depression: a review of recent literature. *Brain Sci*. 2021;11(12):1633.

Rihmer Z, Gonda X, Dome P. Is mania the hypertension of the mood? Discussion of a hypothesis. *Curr Neuropharmacol*. 2017;15(3):424–433.

Rihmer Z, Gonda X, Dome P, Serafini G, Pompili M. Suicid risk in mood disorders—can we better prevent suicide than predict it? *Psychiatr Hung*. 2018;33(4):309–315.

Scott J, Murray G, Henry C, et al. Activation in bipolar disorders: a systematic review. *JAMA Psychiatry*. 2017;74(2):189–196.

Thom R, Silbersweig DA, Boland RJ. Major depressive disorder in medical illness: a review of assessment, prevalence, and treatment options. *Psychosom Med*. 2019;81(3):246–255.

Wardenaar KJ, van Loo HM, Cai T, et al. The effects of co-morbidity in defining major depression subtypes associated with long-term course and severity. *Psychol Med*. 2014;44(15):3289–3302.

Weinberger AH, Gbedemah M, Martinez AM, Nash D, Galea S, Goodwin RD. Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. *Psychol Med*. 2018;48(8):1308–1315.

Wirz-Justice A, Ajdacic V, Rossler W, Steinhausen HC, Angst J. Prevalence of seasonal depression in a prospective cohort study. *Eur Arch Psychiatry Clin Neurosci*. 2019;269(7):833–839.

Woody CA, Ferrari AJ, Siskind DJ, Whiteford HA, Harris MG. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. *J Affect Disord*. 2017;219:86–92.

13.3 Mood Disorders: Clinical Features

ROBERT J. BOLAND, M.D.

TERMINOLOGY

Mood disorders are a pervasive dysregulation of mood and psychomotor activity and related biorhythmic and cognitive disturbances. The rubric of *affective disorder,* which, in some European classifications, also subsumes morbid anxiety states, is increasingly being replaced by the nosologically more delimited concept of *mood disorder.* Thus, *mood disorder* is now the preferred term in the WHO's 10th edition of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) and the American Psychiatric Association's (APA's) fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5). Official mood disorder categories in current use include bipolar disorders (with manic or hypomanic, depressive, or mixed episodes) and major depressive disorders and their respective attenuated variants known as *cyclothymic* and *dysthymic disorders* (called persistent depressive disorder in DSM-5). The term *unipolar* is avoided in DSM-5 because major depressive disorder, especially when early in age at onset, is an unstable diagnosis and can switch to bipolar. Earlier editions had various types of depression, such as *endogenous depression, involutional melancholia,* and *psychotic depressive reaction;* however, starting with DSM-III, they were all incorporated in major depressive disorder. This change meant the end of the neurotic–endogenous distinction, but we now use the term *melancholic features* to qualify major depressive disorders in which biologic concomitants predominate. The United States and international classifications recognize the common occurrence of mixed anxiety–depressions, which underscores the value of the broader affective disorder rubric. It is uncertain whether classic neurasthenic conditions, which have recently reemerged and overlap to some extent with the so-called chronic fatigue syndrome, should be classified under this broad rubric.

RECENT DIAGNOSTIC TRENDS

The reshuffling and reclassification of various affective conditions into the mood disorders chapter of the APA classification in DSM-5 have, on balance, considerably broadened their boundaries. This change reflects, in part, developments in pharmacotherapy that have resulted in considerable alleviation of suffering for persons whose illnesses fall short of and sometimes beyond the boundaries of classic mood disorders. As a result, many individuals with recurrent mood disorders who would have been disabled can now expect to live relatively free of significant episodes and personal or occupational disruptions. Such gratifying results have, in turn, helped to destigmatize this group of disorders. Published self-revelations of famous persons with depressive and bipolar disorders have also helped to destigmatize the disorder. Unfortunately, access to competent mental health care is not available to large segments of the population.

Moreover, destigmatization, which has led to global efforts to detect and treat mood disorders—particularly depression—in public and private sectors, may have had unintended sequelae. Despite the apparent public health advantages of such destigmatization, some authorities nonetheless now contend that a global *depression-centric* perspective has emerged in psychiatric and general medical sectors at the expense of bipolar disorders and may have led to overtreatment with antidepressants and unmasking of latent bipolar disorders. Others believe that the opposite to be true—that is, that we overdiagnose bipolar disorders and overprescribe mood stabilizers. The data, however, indicate that, on balance, it is the continuing underdiagnosis of bipolar disorder that presents the greatest threat to public health. Indeed, recent epidemiologic data suggest considerable disability in untreated or inadequately treated bipolar disorder.

SPECTRUM OF MOOD DISORDERS

The preceding considerations notwithstanding, the increasing clinical diagnosis of major depressive and bipolar disorders should not be dismissed as a mere therapeutic fad. External validating strategies, such as familial-genetic studies and prospective follow-up, can now be used to buttress the broadened concept of mood disorders, including that of bipolar disorder. New research comparing monozygotic and dizygotic twins has demonstrated that the genetic propensity to mood disorders embraces entities that extend beyond endogenous depression (melancholia) to subsume a wider variety of depressions, including some encountered in persons in the community who have never received psychiatric treatment. Although such data might seem counterintuitive to those who would restrict depression to a core primary “biologic” disease, they suggest that the constitutional predisposition for affective dysregulation occurs in as many as one of every three persons. That ratio is similar to the proportion of those who progress to a complete depressive syndrome after bereavement, rhesus monkeys developing depressive-like behavior after a separation paradigm, and dogs that develop learned helplessness after inescapable shock. The fact that these rates are considerably higher than those observed in clinical populations suggests that many people have protective factors against major depressive episodes; alternatively, the data suggest that other factors determine which person with emotional distress becomes a clinical case. Therefore, the study of self-limiting affective conditions on the border of mood disorders can reveal a great deal about the nature of pathologic affective processes.

Evidence from long-term prospective studies indicates that mood disorders consist of various gradations of affective oscillations from the subsyndromal level to the syndromic level. Major episodes represent an operational convention to define a clinical threshold. Subchronic course, dominated by subthreshold symptoms, occurs in 50% to 60%. This is equally true for major depressive disorder and bipolar disorders. Detecting the illness in its subthreshold stage and treating interepisodic subthreshold phases are important clinical considerations.

The suffering and dysfunction resulting from mood disorders are among the most common reasons for consulting psychiatrists and other physicians. In fully developed cases of depression, all activity stops—including creative powers—and life is grim and in total disarray. Although a few talented individuals with milder forms of hypomania may have periods of great creativity, this is not the norm, and most individuals with bipolar disorder have poor judgment and insight, refuse treatment, and suffer a deteriorative course.

All great physicians of the past, beginning with Hippocrates, have devoted considerable space in their general medical texts to the clinical characterization of melancholic and manic states and their alternations in the same patient. The relationship between anxious and melancholic states was also known. A broad spectrum of affective disturbances—ranging from the relatively mild temperamental variants (represented in the official U.S. nosology by dysthymic and cyclothymic disorders) to their severest forms (including what today is considered mood disorder with mood-congruent and mood-incongruent psychotic features)—has been described in classical medical and psychiatric treatises. Finally, classical authors noted that melancholia and certain physical diseases shared seasonal exacerbation, and they described the common occurrence of alcohol indulgence, especially in individuals prone to mania. These boundary problems continue to pose challenges today, with the addition of substance use disorders.

AFFECTS, MOODS, TEMPERAMENTS, AND THE MORBID MOOD STATES

Ethologic Considerations

Affects and *moods* refer to different aspects of emotion. We communicate affect through facial expression, vocal inflection, gestures, and posture, and (according to current ethologic research) affect is intended to move human beings and other primates to appraise whether an individual is satisfied, distressed, disgusted, or in danger. Thus, joy, sadness, anger, and fear are basic affects that serve a communicative function in primates and many other mammalian species.

Affects tend to be short-lived expressions reflecting momentary emotional contingencies. Moods convey sustained emotions; they are more sustained than mere fleeting emotion, and we experience them inwardly as our current feeling state. We manifest our moods in subtle ways, and assessment requires an empathic interviewer. The words that persons use to describe their inner emotions may coincide with the technical terms used by researchers or clinicians and often vary from one culture to another. Furthermore, the inward emotion and the prevailing affective tone may be discordant. This conflict could be due to deliberate simulation (i.e., the person does not wish to reveal their inner emotion), or it could result from a pathologic lesion or process that has altered the emotions and their neural substrates. Thus, evaluating moods and affective expression requires considerable clinical experience. Such evaluation is obviously of great importance in the differential diagnosis. Finally, it is of great usefulness in the therapeutic process; the admonition to “go after the affect” in psychotherapy helps find where it hurts, thereby making treatment possible.

Sadness and Joy

The usual emotions of sadness and joy are part of everyday life, and we must differentiate them from major depressive disorder and mania. Sadness or routine depressive feelings are a universal human response to defeat, disappointment, or other adversities. The response may be adaptive, in an evolutionary sense, by permitting withdrawal to conserve inner resources, or it might signal the need for support from significant others. Transient depressive periods also occur as reactions to particular holidays or anniversaries. They can also occur during the premenstrual and postpartum periods. Called various names—*holiday blues, anniversary reactions, premenstrual tension*, and *maternity blues*—this emotional state is not psychopathologic per se, but individuals predisposed to mood disorder may develop clinical depression during such times.

Grief

Normal bereavement or grief, considered the prototype of reactive depression, occurs in response to significant separations and losses. Such losses might include death, divorce, romantic disappointment, leaving familiar environments, forced emigration, or civilian catastrophes. DSM-5 tends to limit the concept of normal grief to loss due to the death of a loved one. However, the work of Elie Karam and colleagues showed that bereavement and other losses associated with the civil war in Lebanon served as potent forces in depression formation. The same was true for the losses due to the earthquake in Armenia. The boundary between normal reactions and clinical depression is blurry in the face of such complex losses. In addition to depressed affect appropriate to the loss, we characterize bereavement reactions by the prominence of sympathetic arousal and restlessness. From an evolutionary perspective, these symptoms may represent physiologic and behavioral mechanisms to facilitate the search for the lost object. Like other adversities, bereavement and loss do not generally seem to cause depressive disorder, except in individuals predisposed to mood disorder. Reactions to significant catastrophes might represent a partial exception.

Elation

Colloquially, we associate the positive emotion of elation with success and achievement. In many cases, it is as simple as that. However, paradoxical depressions may also follow such positive events, possibly because of the increased responsibilities often faced alone. Elation is conceptualized psychodynamically as a defense against depression or as a denial of the pain of loss, as exemplified by *maniacal grief,* a rare form of bereavement reaction in which elated hyperactivity may replace the expected grief. The character of the “merry widow” in opera seems to be of similar derivation.

Other pseudomanic states include the brief energetic and unusually lucid periods encountered in dying patients or those who need to take superhuman action in the face of unusual duress, both of which some experts think of as a *flight into health.* In predisposed persons, such reactions might be the prelude to a genuine manic episode. Sleep deprivation, which commonly accompanies major stressors, might represent one of the intermediary mechanisms between a stressor and adverse clinical outcomes.

Affective Temperaments

Another mediating factor between normal and pathologic moods is temperament. Most persons have a characteristic pattern of basal affective oscillations that defines their temperament. For instance, some are easily moved to tears by sad or happy circumstances, whereas others tend to remain placid. Usually, oscillations in affective tone are relatively minor, resonate with day-to-day events, and do not interfere with functioning.

Some persons exhibit more significant variability of emotional responses whereby, with no apparent provocation, the person alternates between normal mood and sadness or elation. Temperament traits tend to cluster into basic types. Depressive episodes often complicate a worrying temperament associated with generalized anxiety disorder. The episode overlaps considerably with the depressive temperament; such inclination to melancholy makes the person more easily sink into weary and sad moods and occurs in 3% to 6% of the general population.

The hyperthymic temperament, in which the person is ordinarily cheerful, occurs in perhaps 2% to 5% of young adults. Similarly, the cyclothymic temperament, swinging between cheerful and sad moods, characterizes 4% to 6% of young adults. All four types have an early insidious onset and tend to persist throughout adult life. Of interest, these temperaments may be the prelude to episodes of depressive or hypomanic and manic polarity, which underscores the inherent instability of temperamental inclinations. Marked irritable–explosive traits occur in 2% to 3% of young persons and tend to attenuate by middle age. Current data suggest that such traits often coexist with the mood–labile cyclothymic type, representing this temperament’s dark “borderline” side.

Examining the traits associated with these temperaments can help support the hypothesis that these traits subserve social and evolutionary functions (Table 13.3–1). Thus, the person with a depressive temperament is hard working, dependable, sensitive to the suffering and needs of others, and suitable for jobs that require sustained devotion to meticulous details; we think of such persons as shouldering the burdens of existence without experiencing its pleasures. A person with the hyperthymic temperament, endowed with high levels of energy, extroversion, and humor, assumes leadership positions in society or excels as a performer in media or entertainment; such persons are also successful from a Darwinian perspective in that they are adept in amorous advances and engender a large number of offspring. In talented persons, the cyclothymic temperament, which alternates between sadness and elation, could inspire love and the emotional intensity needed for composing music, writing poetry, and painting. A person with an irritable temperament might be best suited for a military career or even revolutionary action. The danger with extreme temperaments is that they could swing too far in one direction or both directions (i.e., major depressive, manic, or mixed episodes). The use of such substances as alcohol, caffeine, and other stimulants might destabilize affective regulation in persons with those attributes. Some adolescent girls with irritable cyclothymia might develop the extreme emotional disequilibrium that, in contemporary psychiatry, is considered a borderline personality disorder. Temperament concepts can enrich understanding of the boundary between normal moods and emotional disorders and supplement the DSM-5 descriptors of personality disorders with valuable information about individual vulnerability and assets. A new instrument—the Temperament Evaluation of Memphis, Pisa, Paris, and San Diego—has been developed in an auto questionnaire version (TEMPS-A) and validated in at least 12 languages.

Table 13.3–1.
Attributes, Assets, and Liabilities of Depressive and Hyperthymic Types Derived from Classic Concepts of Temperaments

Depressive	Hyperthymic
Gloomy, incapable of fun, complaining	Cheerful and exuberant
Humorless	Articulate and jocular
Pessimistic and given to brooding	Overoptimistic and carefree
Guilt-prone, low self-esteem, and preoccupied with inadequacy or failure	Overconfident, self-assured, boastful, and megalomaniac
Introverted, with restricted social life	Extroverted and people seeking
Sluggish, living a life out of action	High energy level, full of plans
Few but constant interests	Versatile and broad interests
Passive and "sensitive"	Overinvolved and meddlesome
Reliable, dependable, and devoted	Uninhibited and stimulus seeking

Morbid Mood States

Mood disorders represent abnormal or extreme variations in mood and associated manifestations and have the following features: pathologic mood change, endoreactive moods, recurrence, and impairment.

Pathologic Mood Change. Pathologic moods differ from their normal counterparts by being out of proportion to any concurrent stressor or situation; being unresponsive to reassurance; being sustained for weeks, months, and, sometimes, years; and having a pervasive effect on the person that the mood seriously compromises judgment.

Endoreactive Moods. We diagnose depression and mania, respectively, when sadness or elation is overly intense and continues beyond the expected impact of a stressful life event. Indeed, the morbid mood might arise without apparent or significant life stress. We define the pathologic process in mood disorders, in part by the ease with which an intensely emotional state is released and, especially, by its tendency to persist autonomously even when the offending stressor is no longer operative. Rather than being endogenous (i.e., occurring in the absence of precipitants), mood disorders are best conceptualized as endoreactive (i.e., once released, they tend to persist autonomously). The homeostatic dyscontrol of mood, which is part of more pervasive mood dysregulation, resists the habitual baseline affective tone.

Recurrence. In a more descriptive vein, what sets mood disorders apart from their normal emotional counterparts is the clustering of signs and symptoms into discrete syndromes that typically recur on an episodic basis or pursue an intermittent, subthreshold course over many years, if not a lifetime. Cyclic course and, in some cases, regular periodic recurrence are other signs of mood dysregulation that are particularly relevant to bipolar disorder.

Impairment. Normative reactions to adversity and stress, including biologic stress, typically consist of transient admixtures of anxiety and dysphoria, which DSM-5 includes under the rubric of adjustment disorder with mixed anxiety and depressed mood. That is, self-limiting reactions are best qualified broadly as normal affective states that often produce minor or transient impairments in our core functions.

Although anxiety, irritability, and anger occur in various mood disorders, pathologically sustained mood states of depression and elation characterize those disorders. Morbid mood states (mood disorders) then consist of protracted emotional reactions that deepen or escalate, respectively, into clinical depression or mania, with a tendency to recur or to evolve into unremitting chronicity in 15% to 20% of cases. The contribution of temperamental peculiarities to such outcomes should be apparent. We base the impaired functioning characteristic of mood disorders on a combination of factors, including severity, autonomy, recurrence, and chronicity of the clinical features.

To recapitulate, dysregulation in mood disorders can take different forms. It may occur as a single severe episode that persists autonomously for many months and sometimes years, or it might recur with episodes of varying severity, years apart or in rapid succession, with or without interepisodic remission. In general, the earlier the age at onset, the more likely are recurrences, especially those of bipolar nature. Thus, depending on the course of the illness, impairment could be state dependent and could occur during an episode, or it could extend into the interepisodic period. According to the National Institute of Mental Health (NIMH) estimates, on average, a woman with bipolar disorder spends 12 years in florid episodes (often hospitalized), loses 14 years from a productive career and motherhood, and has her life curtailed by 9 years. The NIMH Collaborative Depression Study, which used weekly prospective observations for up to two decades, showed that patients with bipolar I disorder are symptomatic 47% of the time, and bipolar II disorder patients are symptomatic 53% of the time, much of it spent in subthreshold depression. In this study, the unipolar major depressive disorder was even more pervasive, with subacute chronicity 59% of the time.

Recent observations have also revealed another pattern of impairment. In dysthymic and cyclothymic disorders, which represent an intensification of temperamental instability, impairment is not due to the severity of the mood disturbance per se but to the cumulative impact of the dysregulation beginning in the juvenile or early adult years and continuing unabated or intermittently over long periods; hence, the frequent confusion with character pathology. Here, the impairment is more subtle but is pervasive. Persons with the cyclothymic disorder tend to be dilettantes, whereas those with the dysthymic disorder often lead morose and colorless lives. We should seek the fundamental causes of mood disorders in the preclinical expressions found in the offspring of adults with these disorders.

PSYCHOPATHOLOGY

Depressive Syndrome

Like other illnesses, depressive disorder clusters into signs and symptoms. These criteria attempt to set an operational threshold for the depressive disorder based on a specified number of items and their temporal patterns. A checklist cannot accomplish the diagnosis of clinical depression, but the DSM-5 diagnostic criteria for major depressive disorder provide a general guide. Only after an in-depth phenomenologic approach can a clinician ascertain the diagnosis of a depressive disorder. Disturbances in all four spheres (mood, psychomotor activity, cognitive, and vegetative) should be ordinarily present for a definitive diagnosis of major depressive disorder.

Mood Disturbances. Mood change, usually considered the sine qua non of morbid depression, manifests in a variety of disturbances, including (1) painful arousal, (2) hypersensitivity to unpleasant events, (3) insensitivity to pleasant events, (4) insensitivity to unpleasant events, (5) reduced anticipatory pleasure, (6) anhedonia or reduced consummatory pleasure, (7) affective blunting, and (8) apathy. The phenomenology and psychometric properties of this broad range of mood disturbances are under investigation at the Salpêtrière Hospital in Paris. The focus in the description that follows is primarily on painfully aroused mood (depression) and diminished capacity for pleasure (anhedonia), two mood disturbances given selective weight in DSM-5 and ICD-10 and 11.

DEPRESSED MOOD. The term *depressed mood* refers to negative affective arousal, variously described as *depressed, anguished, mournful, irritable,* or *anxious.* These descriptions tend to trivialize a morbidly painful emotion, typically experienced as worse than the severest physical pain. Thus, a depressed mood has a somatic quality that, in the extreme, is indescribably painful. Even when not so severe, depressive suffering is qualitatively distinct from its “neurotic” counterparts, taking the form of groundless apprehensions with severe inner turmoil and torment. This description is particularly apt for middle-aged persons and older adults, historically diagnosed as having *involutional melancholia.* The sustained nature of the mood permits no respite, although it tends to lift somewhat in the evening. Suicide may represent an attempt to find deliverance from such unrelenting psychic torment; a depressed person may consider death as comforting.

Patients with a milder form of the malady typically seen in primary care settings might deny experiencing mournful moods and instead complain of physical agony, such as headache, epigastric pain, or precordial distress, in the absence of any evidence of diagnosable physical illness. Past psychiatrists described such conditions as *depressio sine depressione* or *masked depression.* In such cases, often in older patients, the physician should examine the patient’s facial expression, vocal inflection, and overall appearance to help decide whether there is a mood disturbance behind the complaints.

ANHEDONIA AND LOSS OF INTEREST. Paradoxically, patients with a heightened perception of pain may also feel unable to experience normal emotions. Patients exhibiting the disturbance may lose the capacity to cry, a deficit that reverses as the depression lifts.

In evaluating anhedonia, inquiring whether the patient has lost the sense of pleasure is not enough; the clinician must document that the patient has given up previously enjoyed pastimes. When mild, patients with anhedonia have a decreased interest in life. Later, patients complain that they have lost all interest in things. These phenomena are reminiscent of Hamlet’s lament: “How weary, stale, flat, and unprofitable seem to me all the uses of the world” (William Shakespeare, *Hamlet,* Act I, Scene II). In the extreme, patients lose their feelings for their children or spouses, who once were a source of joy. Thus, the hedonic deficit in clinical depression might represent a particular instance of a more pervasive inability to experience emotions.

Patients with severe depression may complain of being emotionally cut off from others and experience depersonalization and a world that seems strange to them (derealization). The impact of the loss of emotional resonance can be so pervasive that patients may denounce values and beliefs that had previously given meaning to their lives. For instance, members of the clergy might present with the complaint that they no longer believe in the Church and have lost faith in God. The inability of the person with a depressive disorder to experience normal emotions (commonly observed among young depressed patients) differs from the flat affect of patients with schizophrenia in that the loss of emotions is itself experienced as painful; that is, the patient suffers immensely from the inability to experience emotions.

Psychomotor Disturbances. In depression, psychomotor changes consist of abnormalities in the motor expression of mental and emotional activity. In severe cases, these changes manifest in specific facial features (Fig. 13.3–1).

PSYCHOMOTOR AGITATION. Although agitation (pressured speech, restlessness, hand wringing, and hairpulling) is the more readily observed abnormality, it appears less specific to the illness than retardation (slowing of psychomotor activity). Psychophysiological studies have documented that such slowing often coexists with agitation.

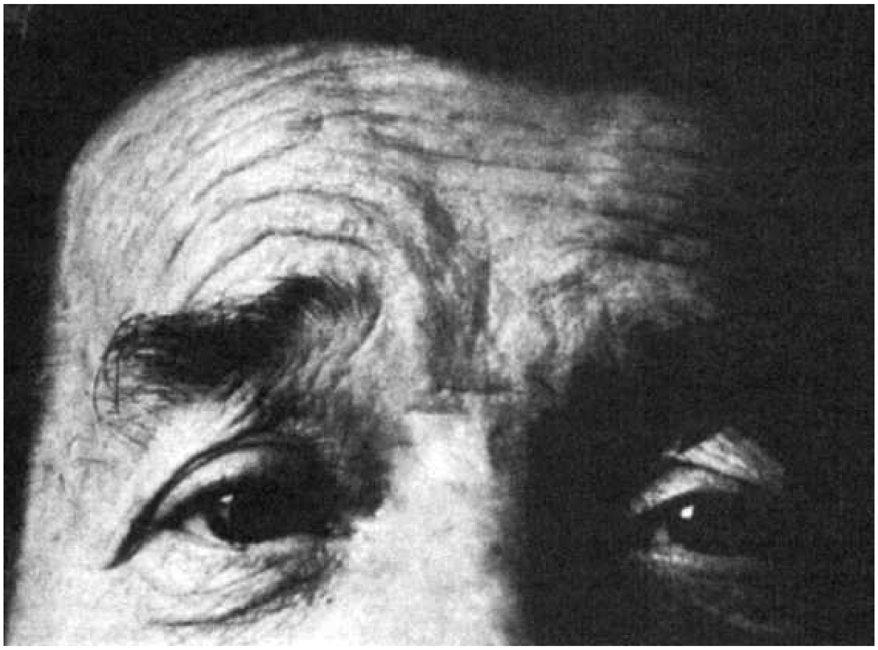

FIGURE 13.3–1. The Swiss neuropsychiatrist Otto Veraguth described a peculiar triangle-shaped fold in the nasal corner of the upper eyelid. The fold, often associated with depression, is referred to as Veraguth's fold. The photograph illustrates this physiognomic feature in a 50-year-old man during a major depressive episode. Veraguth's fold may also occur in persons who are not clinically depressed, usually while harboring a mild depressive affect. Distinct changes in the tone of the corrugator and zygomatic facial muscles accompany depression, as shown in electromyograms. (Courtesy of Heinz E. Lehmann, M.D.)

PSYCHOMOTOR RETARDATION. Underlying many of the deficits seen in clinical depression, some authorities believe psychomotor retardation to be the core, or primary, pathology in mood disorders. Morbid depression—what patients describe as being “down”—can be understood in terms of moderate-to-extreme psychomotor slowing. The patient experiences inertia, being unable to act physically and mentally. Brain imaging research revealing subcortical (extrapyramidal system) disturbances in mood disorders tend to support the centrality of psychomotor dysfunction in these disorders.

Long neglected in psychopathologic research, psychomotor retardation is one of the few symptoms of depression that is reasonably simple to measure objectively. The Salpêtrière Retardation Scale developed by Daniel Widlocher and colleagues places particular emphasis on the following disturbances: (1) paucity of spontaneous movements; (2) slumped posture with downcast gaze (Fig. 13.3–2); (3) overwhelming fatigue (patients complain that everything is an effort); (4) reduced flow and amplitude of speech and increased latency of responses, often giving rise to monosyllabic speech; (5) a subjective feeling that time is passing slowly or has stopped; (6) poor concentration and forgetfulness; (7) painful rumination or thinking that dwells on a few (usually unpleasant) topics; and (8) indecisiveness or an inability to make simple decisions.

DSM-5 places greater emphasis on the more easily observable objective or physical aspects of retardation. For the patient, however, the subjective sense of slowing is as pervasive and disabling. It is easier to measure this more psychological dimension of retardation in depressed persons with good verbal skills.

Ms. A, a 34-year-old literature professor, presented to a mood clinic with the following complaint: “I am in a daze, confused, disoriented, staring. My thoughts do not flow, my mind is arrested . . . I seem to lack any sense of direction, purpose. . . I have such an inertia, I cannot assert myself. I cannot fight; I have no will.”

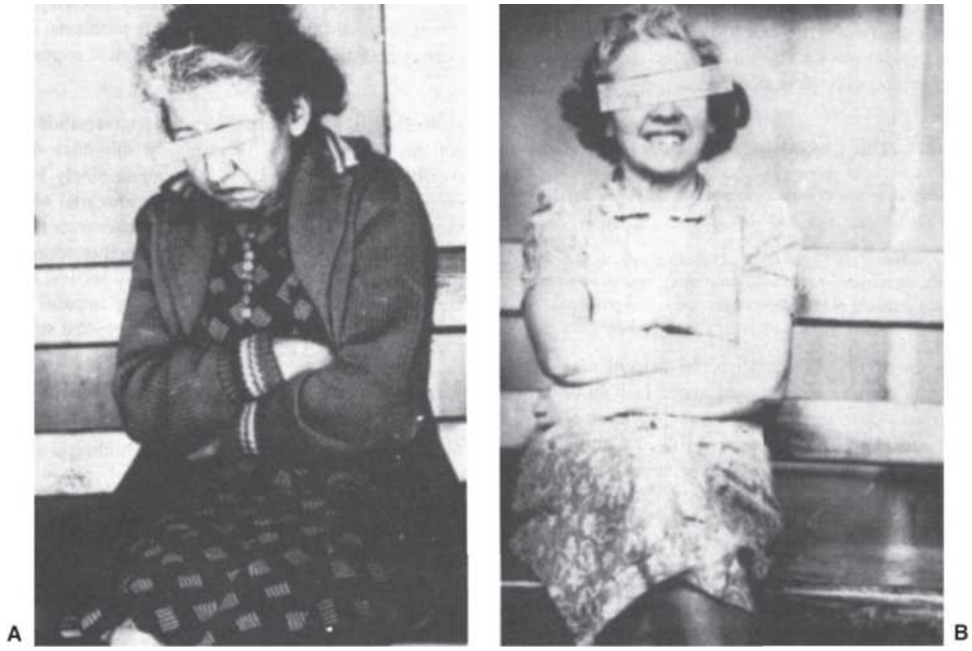

FIGURE 13.3–2. A 38-year-old woman during a state of deep retarded depression **(A)** and 2 months later, after recovery **(B)**. Note the turned-down corners of her mouth, stooped posture, drab clothing, and hairdo during the depressed episode. (Courtesy of Heinz E. Lehmann, M.D.)

Less linguistically sophisticated patients would simply complain of an inability to perform household chores or difficulty in concentrating on their studies. Such psychomotor deficits, in turn, underlie depressed patients' diminished efficiency or their inability to work.

STUPOR. Psychomotor slowing in young persons is sometimes so extreme that patients might slide into a stupor, unable to participate even in essential biologic functions, such as feeding themselves. Such an episode is often the precursor of bipolar disorder, which later declares itself in a manic episode. Today, we usually diagnose depressive disorders earlier and are more likely to encounter more subtle stupor.

A 20-year-old male college student seen in the emergency room spoke of "being stuck—as if I have fallen into a black hole and cannot get out." Further evaluation revealed that the patient was metaphorically describing his total loss of initiative and drive and felt engulfed by the disease process. A clinician without the requisite phenomenologic training might consider such a patient bizarre and perhaps even psychotic. Nevertheless, the patient responded dramatically to fluoxetine and, in 2 weeks, was back in school.

Cognitive Disturbances. The cognitive view of depression considers negative evaluations of the self, the world, and the future (the negative triad) central to understanding depressed mood and behavior. However, it is equally likely that the depressed mood colors perceptions of the self and others or that disturbed psychomotor activity leads to negative self-evaluations. Therefore, instead of being considered causal, the cognitive triad in depression is best approached empirically as a psychopathologic manifestation of depression. Depressed persons express faulty thinking patterns as (1) ideas of deprivation and loss; (2) low self-esteem and self-confidence; (3) self-reproach and pathologic guilt; (4) helplessness, hopelessness, and pessimism; and (5) recurrent thoughts of death and suicide.

The essential characteristic of depressive thinking is that the patient views everything in a highly negative light. The self-accusations are typically unjustified or are blown out of proportion. An example might be the case of a middle-aged woman who was tormented by guilt because, as a child, she had not repaid a nickel that she had borrowed from a classmate. Some of the thoughts may verge on the delusional. For instance, an internationally renowned scientist complained that he was “nothing.” Self-evaluations that indicate a shallow image of one’s self might, nonetheless, reflect an accurate perception of one’s impairment from psychomotor retardation.

In addition to content, the cognitive symptoms include various cognitive difficulties, the most pronounced being mental slowing. Patients may also have general difficulties with attention, concentration, and memory. This deterioration of cognitive functions can be so pronounced in older adult that they experience disorientation and confusion. At times the term “pseudodementia” has been used to describe this. However, that may be ill-advised as it implies that the cognitive symptoms are not genuine, which is not the case. Of most concern is the data suggesting that such cognitive symptoms persist beyond antidepressant response and impact the likelihood of recurrence. Because of these findings, there is a growing interest in recognizing and treating the cognitive symptoms of depression.

MOOD-CONGRUENT PSYCHOTIC FEATURES. In depressive disorders with psychotic features, negative thinking acquires grossly delusional proportions and is maintained with such conviction that the thoughts are not amenable to change by evidence to the contrary. According to Kurt Schneider, delusional thinking in depression derives from humankind’s four basic insecurities—those regarding health, financial status, moral worth, and relationship to others. Thus, severely depressed patients may have delusions of worthlessness and sinfulness, reference, and persecution: They believe that others single them out for their past mistakes and that everyone is aware of their incompetence. Persecutory ideation in depression is often *prosecutory* in that it derives from the belief that one deserves punishment for such transgressions. A severely depressed man may feel so incompetent in all areas of functioning, including the sexual sphere, that he may suspect his wife of having an affair (delusion of infidelity).

Other depressed persons believe that they have mismanaged their finances in such a way that their children will starve (delusions of poverty); that they are harboring an occult illness, such as cancer or the acquired immune deficiency syndrome (AIDS) (delusions of ill health); or that parts of their bodies are missing (nihilistic delusions). In more severe illness, the patient might feel that the world has changed and that calamity and destruction await everyone. In rare tragic instances, a parent with such delusions might kill their young children to save them from moral or physical decay and commit suicide. In women with psychotic depression, infanticide is most likely to occur in the postpartum period, often leading to all types of inappropriate interpretations in the media. Finally, a minority of depressed persons have fleeting auditory or visual hallucinations with extremely unpleasant content along the lines of their delusions (e.g., hearing accusatory voices or seeing themselves in coffins or graveyards). All of these psychotic experiences are genuine affective delusions or hallucinations. They are mood congruent in the sense that they are phenomenologically understandable in light of the prevailing pathologic mood.

MOOD-INCONGRUENT PSYCHOTIC FEATURES. Sometimes so-called first-rank or Schneiderian-type symptoms can arise in the setting of a major depressive episode.

A 42-year-old civil servant said that her depression was so paralyzing that she felt no personal initiative or volition; she believed that some malignant force had taken over her actions and commented on every action she was undertaking. The patient recovered fully with thymoleptic medication. There is no reason to believe that, in this patient, the feelings of somatic passivity and running commentary indicated a schizophrenic process.

Thus, with proper phenomenologic probing, certain classes of apparently mood-incongruent psychotic experiences can be understood as arising from the pathologic mood and the profound changes in psychomotor activity that accompany them. (In other instances, the clinician must seek a history of alcohol or substance use disorder or withdrawal as a putative explanation for mood incongruence in psychotic depression.) In brief, incidental Schneiderian first-rank symptoms should not distract from the diagnosis of an affective psychosis if otherwise typical signs and symptoms of mood disorder are present.

HOPELESSNESS AND SUICIDE. Given that most, if not all, clinically depressed patients find themselves locked in the private hell of their negative thoughts, it is not surprising that many untreated or inadequately treated patients give up hope of ever recovering and kill themselves. The suicide attempt is not, however, undertaken in the depth of melancholia. For example, when asked if she had any suicide plans, a severely depressed patient replied, “Doctor, I don’t exist—I am already dead.”

Thus, the risk of suicide is less pronounced during acute severe depression. Emil Kraepelin observed that when psychomotor activity is improving, yet mood and thinking are still dark, the patient is most likely to muster the requisite energy to commit the suicidal act. Aaron Beck’s work has shown that hopelessness on mental status evaluation in a patient recovering from depression should alert the clinician to the possibility of such an outcome. The discovery of a therapeutic modality that could improve psychomotor, mood, and cognitive components in tandem will constitute a significant clinical advance.

Despite a common belief otherwise, inquiring about suicide does not provoke it. On the contrary, patients are often relieved that the physician appreciates the magnitude of their suffering. Patients commonly express suicidal ideation indirectly (e.g., wish not to wake up or die from a malignant disease). Some depressed persons are tormented with suicidal obsessions and constantly resist unwanted urges or impulses to destroy themselves. Others might yield to such urges passively (e.g., careless driving or walking into high-speed traffic). A third group harbors elaborate plans, carefully preparing a will and taking out insurance. Deliberate planning indicates a high suicidal risk. The preceding examples are not exhaustive; they are simply reminders to clinicians with depressed patients to be alert always to the possibility of suicide.

Vegetative Disturbances. The Greeks considered depression a somatic illness and ascribed it to black bile; hence, *melancholia.* The mood change in depressive disorder includes measurable alterations of biorhythms that implicate midbrain dysfunction. Once the changes occur, they tend to be independent of the environment throughout much of the episode, and, as a consequence, they do not respond to interpersonal feedback of a pleasant or upbeat nature. The biologic concomitants of melancholia include profound reductions in appetite, sleep, and sexual functioning and alterations in other circadian rhythms, especially matinal worsening of mood and psychomotor performance. These disturbances are central to the concept of melancholia, a form of depression in which such biologic concomitants predominate. An equally important subgroup of depressed persons exhibits a reversal of the vegetative and circadian functions, increasing appetite and sleep—and sometimes in sexual functioning—and an evening worsening of mood; in this atypical pattern, patients characteristically exhibit mood reactivity and sensitivity to rejection. Marked retardation might herald a manic switch, whereas atypicality should raise the suspicion of bipolar II disorder. Table 13.3–2 summarizes the vegetative symptoms of depression.

ANOREXIA AND WEIGHT LOSS. The most reliable somatic indicators of the depressive disorder include anorexia and weight loss. In addition to being a hypothalamic-based disturbance in depression, anorexia might be secondary to blunted olfactory or taste sensations or a decreased enjoyment of food, or (rarely) it might result from a delusional belief that someone has poisoned one's food.

Inanition, especially in older adults, can lead to malnutrition and electrolyte disturbances that represent medical emergencies in their own right. If weight loss is severe, especially after 40 years of age, the psychiatrist should first use appropriate medical consultation to rule out the likelihood of an occult malignancy.

WEIGHT GAIN. Overeating, decreased activity, or both may result in weight gain. In middle-aged patients, it may aggravate pre-existing diabetes mellitus, hypertension, or coronary artery disease. In younger patients, especially women, weight problems may conform to a bulimic pattern that often expresses the depressive phase of a bipolar disorder with infrequent hypomanic periods (bipolar II disorder).

INSOMNIA. Sleep disturbance, a cardinal sign of depression, often is characterized by multiple awakenings, especially in the early hours of the morning, rather than by difficulty falling asleep. The light sleep of a depressed person, in part a reflection of the painful arousal of the disorder, tends to prolong the depressive agony over 24 hours. Thus, deep stages of sleep (III and IV) are decreased or deficient. The attempt to overcome the problem by drinking alcohol may initially succeed but ultimately aggravates the sleep patterns and insomnia. This paradox is also true for sedative–hypnotic agents, which a busy general practitioner might prescribe, having not recognized the underlying depressive condition. Although sedatives (including alcohol) effectively reduce the number of awakenings in the short term, they are not effective in the long run because they further diminish stage III and stage IV sleep. They are not antidepressants, and they tend to prolong the depression.

Table 13.3–2.
Neurovegetative Symptoms of Depression

Common:
- Fatigue, low energy
- Inattention
- Insomnia, early morning awakening
- Poor appetite, associated weight loss

Sometimes included:
- Decreased libido and sexual performance
- Menstrual irregularities
- Worse depression in the AM

HYPERSOMNIA. Young depressed patients, especially those with bipolar tendencies, often exhibit excessive sleep and have difficulty getting up in the morning.

Kevin, a 15-year-old boy, was referred to a sleep center to rule out narcolepsy. His main complaints were fatigue, boredom, and a need to sleep all the time. Although he had always started the day somewhat slowly, he now could not get out of bed to go to school. That alarmed his mother, prompting sleep consultation. Formerly a B student, he failed most of his courses in the 6 months before referral. Psychological counseling, predicated on the premise that his family's recent move from another city had led to Kevin's isolation, had not been beneficial. Extensive neurologic and general medical workup had also proven unremarkable. He slept 12 to 15 hours per day but denied cataplexy, sleep paralysis, and hypnagogic hallucinations. During the psychiatric interview, he denied being depressed but admitted that he had lost interest in everything except his dog. He had no drive, participated in no activities, and had gained 30 pounds in 6 months. He believed that he was "brain damaged" and wondered whether it was worth living like that. The question of suicide disturbed him, as it was contrary to his religious beliefs. These findings led to the prescription of an antidepressant. Not only did the antidepressant reverse the presenting complaints, but it also pushed him to the brink of a manic episode.

The affective nature of the disorder in such patients is often unrecognized, and observers attribute their behavior to "laziness." The vignette also illustrates the emergence of manic behavior during antidepressant treatment. Such shifts in polarity are typical in major depressive disorder and necessitate revising the diagnosis to a bipolar disorder. Hypersomnia associated with lethargy should raise a strong suspicion of bipolar disorder in young depressive patients. In older adults, organic or brain pathology should be suspected.

CIRCADIAN DYSREGULATION. Many circadian functions, such as temperature regulation and cortisol rhythms, are disrupted in major depressive disorder. Disturbances of sleep rhythms, however, have received the most significant research focus. These include deficits in stage IV or delta sleep and more intense REM activity in the first one-third of the night. More specific to depressive disorders—whether suffering from insomnia or hypersomnia—nearly two-thirds of patients exhibit a marked shortening of REM latency, the period from the onset of sleep to the first REM period. This abnormality is observed throughout the depressive episode and may also occur during relatively euthymic periods in persons with recurrent depression. The occurrence of short REM latency in the younger, clinically healthy relatives of the affectively ill may be an early presymptomatic warning sign of the disease. However, on closer scrutiny, these "well" relatives are often found to meet the criteria for subthreshold mood conditions.

Few data exist on the consistency of sleep electroencephalographic (EEG) abnormalities in patients from episode to episode. However, clinical experience suggests that a patient observed over time (even during the same episode) may exhibit insomnia and morning worsening of mood and activity during one period of the disorder and hypersomnia extending to late morning hours during another period. In either case, persons with a depressive disorder are characteristically tired in the morning, which means that even prolonged sleep is not refreshing. Most often, we see this propensity to exhibit such divergent patterns of sleep disturbance in bipolar disorders. Patients with major depressive disorder tend to exhibit insomnia more stereotypically episode after episode; despite extreme fatigue, they rarely oversleep. Such fatigue coexisting with negative affective arousal is even more exhausting.

SEASONALITY. Another classic biorhythmic disturbance in mood disorders is seasonal (especially autumn–winter) accentuation or precipitation of depression. Most of those patients experience increased energy and activation, if not frank hypomania, in the spring. In the fall and winter, they complain of fatigue, tend to crave sugars, and overeat and oversleep. The hypersomnia in some of these patients is associated with delayed (rather than short) REM latencies. These data suggest dysregulation of circadian rhythms in depressive disorders rather than mere phase advance. Given the biphasic nature of the clinical phenomenology, some autumn–winter depressions belong to the bipolar spectrum (type II). Although autumn–winter depression has received the most significant attention, there also exist summer depressions; the former appears to be related to a reduction in daylight (photoperiods), and the latter appears to be related to increased temperature.

SEXUAL DYSFUNCTION. Decreased sexual desire occurs in depressed men and women. In addition, some women experience a temporary interruption of their menses. Depressed women are often unresponsive to lovemaking or are disinclined to participate in it, a situation that could lead to marital conflict. Psychotherapists might mistakenly ascribe the depression to the marital conflict and devote unnecessarily zealous psychotherapeutic attention to those issues. Decreased or lost libido in men often results in erectile failure, which may prompt endocrinologic or urologic consultation. Again, clinicians may mistake depression for sexual dysfunction rather than the reverse, and the focus on the sexual complaint may delay definitive treatment. In the past, some men with depressive disorder received permanent penile implants before any depression treatment. This mistreatment is less likely to occur in the sildenafil era, but even treatment with such agents would not necessarily resolve erectile disorder in clinically depressed patients without competent treatment of the mood disorder. SSRIs typically aggravate sexual dysfunction, whereas bupropion may improve it.

A small subgroup of patients with depressive disorders may exhibit increased sexual drive or activity of a compulsive nature. These patients tend to have other atypical features as well; hence, the increased sexual drive can be considered the *fifth reverse vegetative sign* (after evening worsening of mood, initial insomnia, hypersomnia, and weight gain). Several patients have derived temporary reversal of their depression after intense sexual encounters. In these depressed persons, increased sexual drive, intensity, or both may indicate a mixed episode of bipolar disorder (type II). Further scrutiny in such cases often reveals a premorbid cyclothymic or hyperthymic temperament. Current data suggest that, depending on the breadth of the criteria used, 30% to 70% of depressions with atypical features belong to the bipolar spectrum.

Manic Syndrome

As with clinical depression, we can organize the psychopathology of mania in terms of mood, psychomotor, circadian, and cognitive disturbances. The clinical features of mania are generally the opposite of those of depression. Thus, instead of lowered mood, thinking, activity, and self-esteem, the patient has an elevated mood, rush of ideas, psychomotor acceleration, and grandiosity. Despite those contrasts, the two disorders share such symptoms as irritability, anger, insomnia, and agitation. An excess of the last symptoms of escalating intensity suggests a mixed phase or mixed episode of mania and depression co-occurring. Manic and mixed episodes represent the hallmark of what psychiatrists used to call *manic-depressive psychosis;* today, we call it *bipolar I disorder.*

Although milder or hypomanic features can contribute to success in business, leadership roles, and the arts, even mild manic symptomatology recurrences are typically disruptive. The elated mood tends to produce overoptimism concerning one’s abilities, which, coupled with the impulsivity characteristic of mania, often leads to disaster. Thus, accurate and early diagnosis is paramount.

Classic mania, as formulated in the DSM-5, is relatively easy to recognize. Misdiagnosis was once rampant in North American practice as clinicians confused severe mania with schizophrenia and confused its milder variants with normality or narcissistic and sociopathic personality disorders. As with misdiagnosed depression, these errors of clinical judgment represent a lack of familiarity with the phenomenology of the classic illness. Again, DSM-5 criteria provide only a guideline. The actual diagnosis requires a careful history and phenomenologic understanding by an empathic observer. The manic patient lifts the observer’s mood and makes the examiner smile and even laugh, but it can also be irritating. The patient’s speech is fast and may even appear to the novice psychiatry student as “loose,” but it can often be witty. Finally, the behavior is typically dramatic, expansive, and jesting. Current research indicates that social disinhibition and pathologic overfamiliarity with strangers represent cardinal features of mania which we can contrast with the withdrawal of most schizophrenic patients. For the experienced clinician, the overall gestalt experienced in the presence of manic patients is emotionally and qualitatively distinct from that of persons with schizophrenia or frontal lobe diseases; the latter conditions tend to leave the examiner cold. These considerations become apparent when the clinical observer systematically examines the psychopathology of mania in the areas of mood, behavior, and thinking.

Mood Disturbance. Mood disturbance in mania represents a contrast to that observed in depression, but not entirely.

MOOD ELEVATION. The mood in mania is classically one of elation, euphoria, and jubilation, typically associated with laughing, punning, and gesturing.

LABILITY AND IRRITABILITY. The prevailing positive mood in mania is not stable, and momentary crying or bursting into tears is common. In addition, the high is so excessive that many patients experience it as intense nervousness. When crossed, patients can become highly irritable and hostile. Thus, lability and irritable hostility are as much features of the manic mood as is elation. In mixed manic states, they dominate the clinical picture, giving rise to what is now termed *dysphoric mania* (and what Kraepelin characterized as *anxious-depressive mania*).

Psychomotor Acceleration. Accelerated psychomotor activity, the hallmark of mania, is characterized by overabundant energy and activity and rapid, pressured speech. Subjectively, the patient experiences an unusual sense of physical well-being (eutonia).

FLIGHT OF IDEAS. Thinking processes are accelerated, subjectively experienced as a flight of ideas, and thinking and perception are unusually sharp. The patient may speak with such pressure that associations are difficult to follow; such "clang" associations are often based on rhyming or chance perceptions and can be lightning fast. The pressure to speak may continue despite the development of hoarseness.

IMPULSIVE BEHAVIOR. Manic patients are typically impulsive, disinhibited, and meddlesome. Pathologic familiarity with total strangers is also a feature not explicitly listed in the DSM-5 schema for mania, yet it is one of its cardinal signs. They are intrusive in their increased involvement with others, leading to friction with family members, friends, and colleagues. They are distractible and move quickly, not only from one thought to another but also from one person to another, showing heightened interest in every new activity that strikes their fancy. They are indefatigable and engage in various activities in which they usually display poor social judgment. Examples include preaching or dancing in the street; abusing the use of long-distance calling; buying new cars, hundreds of records, expensive jewelry, or other unnecessary items; paying the bills of total strangers in bars; giving away furniture; marrying impulsively; engaging in risky business ventures; gambling; and taking sudden trips. Such pursuits can lead to personal and financial ruin.

DELIRIOUS MANIA. Although not listed in DSM-5, a highly severe expression of mania (once known as *Bell mania*), called delirious mania, involves frenzied physical activity that continues unabated and leads to delirium and disorientation—a life-threatening medical emergency. This complication, the manic counterpart of stupor, is rare today.

WITH CATATONIC FEATURES. As a symptom, catatonia can be present in several mental disorders. The presence of catatonic features in patients with mood disorders may have prognostic and treatment significance. The hallmark of symptoms of catatonia—stuporousness, blunted affect, extreme withdrawal, negativism, and marked psychomotor retardation—can be seen in both catatonic and noncatatonic schizophrenia, major depressive disorder (often with psychotic features), and medical and neurologic disorders. Clinicians often do not associate catatonic symptoms with bipolar I disorder because of the marked contrast between these symptoms of stuporous catatonia and the classic symptoms of mania.

VEGETATIVE DISTURBANCES. Vegetative disturbances are more challenging to evaluate in mania than in depression.

Hyposomnia. Hyposomnia is the cardinal sign of a decreased need for sleep—the patient sleeps only a few hours but feels energetic on awakening. Some patients may go sleepless for several days. This sleep deprivation could lead to a dangerous escalation of manic activity, which might continue despite signs of physical exhaustion.

Inattention to Nutrition. There does not seem to be a clinically significant level of appetite disturbance, but weight loss may occur because of increased activity and neglect of nutritional needs.

Sexual Excesses. Hypersexuality is a cardinal sign of mania. The sexual appetite is typically increased and may lead to sexual indiscretion. Married women with previously stable sexual lives may become promiscuous and involve themselves with strange and even potentially dangerous men. Men typically overindulge in alcohol, frequent bars, and may squander their savings on prostitutes. The sexual misadventures of manic patients result in marital disasters—hence the multiple separations or divorces that are almost pathognomonic of the disorder. The sexual excesses of bipolar patients can also put them at risk of sexually transmitted diseases, including AIDS.

Cognitive Distortions. Manic thinking is overly positive, optimistic, and expansive.

GRANDIOSITY, LACK OF INSIGHT, AND POOR JUDGMENT. The patient exhibits inflated self-esteem and a grandiose sense of confidence and achievements. Behind that facade, however, may be a vague and painful recognition that the positive self-concepts do not represent reality. However, such insight (if present at all) is transient, and manic patients are notoriously refractory to self-examination and insight. Denial and lack of insight are cardinal psychological derangements of mania. This lack of insight—coupled with poor judgment—leads manic patients to engage in activities that harm themselves and their loved ones. It also explains, in part, their nonadherence to medication regimens during the manic phase.

DELUSION FORMATION. Manic patients often harbor delusional beliefs. They may include delusions of exceptional mental and physical fitness and talent, wealth, aristocratic ancestry or another grandiose identity, or assistance (i.e., from powerful or supernatural people). They may also have delusions of reference or persecution, based on the belief that enemies observe or follow them out of jealousy at their unique abilities. At the height of the mania, patients may even see visions or hear voices congruent with their euphoric mood and grandiose self-image (e.g., they might see images of heaven or hear cherubs chanting songs to praise them). The denial characteristic of mania—and the frequently psychotic nature of episodes—means that clinicians must routinely obtain diagnostic information about past episodes from significant others. Lack of insight also, unfortunately, means that physicians may have to resort to involuntary hospitalization.

MOOD-INCONGRUENT PSYCHOSIS. Psychosis in the setting of mania and mixed manic episodes is typically mood congruent. The sense of physical well-being and mental alacrity is so extraordinary that it is understandable why manic patients believe that they possess superior powers or perhaps are great scientists or famous reformers. Moreover, their senses are so vivid that reality appears more luxurious or exotic. They may interpret their misperceptions as being visions.

Likewise, their thoughts are so rapid and vibrant that they feel that they can hear them. Thus, specific first-rank Schneiderian-type symptoms that have been traditionally considered mood incongruent can be understood phenomenologically to arise from the powerful mental experiences of mania.

A 37-year-old engineer had experienced three manic episodes for which he was hospitalized; before each episode, he had several weeks of moderate psychomotor retardation. Although he had responded to lithium each time, he had been reluctant to take it once outside the hospital and eventually refused to do so. Now that he was euthymic, after his third and most disruptive episode during which he had severely beaten his wife, he could more accurately explain how he felt when manic. He experienced mania as "God implanted in him," so he could serve as "testimony to man's communication with God." He elaborated as follows: "Ordinary mortals will never, never understand the supreme manic state which I am privileged to experience every few years. It is so vivid, so intense, so compelling. When I feel that way, there can be no other explanation: To be manic is, ultimately, to be God. God himself must be supermanic: I can feel it when mania enters through my left brain like laser beams, transforming my sluggish thoughts, recharging them, galvanizing them. My thoughts acquire such momentum. They rush out of my head to disseminate knowledge about the true nature of mania to psychiatrists and all other ordinary mortals. That's why I will never accept lithium again—to do so is to obstruct the divinity in me." Although he was on the brink of divorce, he would not yield to his wife's plea to go back on lithium.

The vignette illustrates the possibility that even some of the most psychotic manifestations of mania represent explanatory delusions, the patient's attempt to make sense of the experience of mania. Many manic patients abuse alcohol and stimulants to enhance their mental state; we can sometimes explain mood incongruence on that basis.

HYPOMANIA VERSUS MANIA. Nonpsychotic and nondisruptive variants of mania are much more common, and DSM-5 recognizes them as hypomanic episodes. They are the historical clinical marker for bipolar II. Diagnostically, significant others can provide a history of hypomania, as the subject may be unaware of it or may tend to deny it. Others experience irritable or dysphoric hypomanic episodes, which may be difficult to recognize clinically because of a stereotype that emphasizes the positive aspects of hypomania. Some patients feel that they benefit from the energy and confidence of hypomania. This potential benefit is true if hypomania is mild and “sunny.” But, many patients, especially those with “darker” irritable activation, experience impairment in occupational and interpersonal functioning over time. DSM-5 stipulates a minimum duration of 4 days for hypomania; however, current evidence indicates that bipolar II disorders with long (4 or more days) and short (2 to 3 days) hypomania are indistinguishable based on bipolar family history. It is, therefore, recommended that the threshold for detecting the duration of hypomania be set at 2 days. Ultimately, the duration of hypomanic experiences might be less important than the fact that they recur. In other words, the occurrence of brief recurrent hypomanias, even if their duration is 1 day, interspaced with major depressive episodes, can be taken as presumptive evidence for bipolar II. Finally, although DSM-5 states that treatment-emergent hypomania in a depressed patient does not count toward a diagnosis of bipolarity, prospective observations show that nearly all such episodes are followed eventually by spontaneous hypomania (or mania); moreover, family history for bipolar disorder is comparable in patients with spontaneous and antidepressant-associated hypomania. In brief, antidepressant-associated hypomanias are likely a genetically less penetrant variant of bipolarity (which, in the literature, are often referred to as *bipolar III disorders*).

DIAGNOSTIC CLASSIFICATION

Tables 13.3–3 through 13.3–7 list the most significant mood disorder diagnoses and compare the DSM-5 and ICD-11 approaches to diagnosing these disorders.

The classification of mood disorders in DSM-5 subsumes many patients seen in private, public, ambulatory, and inpatient settings. The main demarcation in that large clinical terrain is between bipolar and depressive disorders. Thus, bipolar disorders range from the classic manic and depressive episodes, often of psychotic intensity (bipolar I disorder), to recurrent major depressive episodes alternating with hypomanic episodes (bipolar II disorder) and cyclothymic mood swings. Likewise, depressive disorders include those with psychotic severity, melancholia, atypical features, and dysthymic variants.

We can distinguish the major and specific attenuated subtypes by their severity and duration. In dysthymic and cyclothymic disorders, a partial mood syndrome consisting of subthreshold features as subdepressive and hypomanic periods is maintained, intermittently or continuously, for at least 2 years. Subdepressive periods dominate in dysthymia; in cyclothymia, they alternate with brief hypomania. The onset is typically in adolescence or childhood, and most persons with these diagnoses first seen clinically in young adulthood have had low-grade mood symptoms for 5 to 10 years. Major mood disorders, which generally begin much later in life, require the presence of a full manic episode or a full depressive episode—sustained for at least 1 and 2 weeks, respectively—and an episodic course, typically permitting recovery or remission from episodes. As many as one-third of persons with major depressive disorders fail to achieve full symptomatic recovery and should thus be qualified as chronic or partial remission.

Dichotomy or Continuum?

Although we can discriminate between bipolar and depressive disorders clinically and therapeutically (Table 13.3–8), clinical observations testify to a vast overlap between those extremes. Thus, the distinctions among the various affective subtypes are not as hard and fast as DSM-5 attempts to portray. For instance, full-blown bipolar disorder can superimpose on a cyclothymic disorder that tends to persist after the resolution of manic or major depressive episodes. Even more common is major depressive disorder complicating cyclothymic disorder, which should be a significant course variant of bipolar II disorder. Likewise, recent evidence indicates that dysthymic disorder may precede major depressive disorder in as many as one-third of cases. DSM-5 recognizes this overlap by including "mixed" subtypes for both major depressive disorder and bipolar disorder. However, it does not entirely resolve the question of where the boundaries lie.

Moreover, as much as 50% of persons with a major depressive disorder during long-term prospective follow-up develop hypomanic or manic episodes and have bipolar disorder. In some, if not many instances, apparent switching of polarity might be due to earlier misclassification of bipolar disorder as major depressive disorder. Finally, there may be an unexpected crossing from dysthymic disorder to hypomanic or manic episodes, suggesting that some forms of dysthymic disorder are subaffective precursors of bipolar disorder. The concept of *subbipolar dysthymia* (not a formal category in DSM-5) can bring clinical attention to this subgroup of patients. Such observations are in line with Kraepelin's historic attempt to bring all mood disorders under one rubric. Epidemiologic studies in the community have also shown much fluidity among various subthreshold and major mood disorders.

Heterogeneity undoubtedly exists among mood disorders. However, the classic unipolar–bipolar distinction might not be the best way to capture it. The preceding observations suggest that much of the recurrent depressive terrain might be *pseudounipolar* (i.e., soft bipolar). The clinical significance of these considerations lies in the fact that many DSM-5 subtypes of mood disorders are not pure entities, and considerable overlap and switches in polarity take place. They also provide some rationale, for instance, for why lithium or other mood stabilizer augmentation may be effective in some unipolar depressions; such patients do not necessarily experience brief spontaneous hypomanic episodes but often exhibit high baseline level hyperthymic traits. Finally, several studies have shown that bipolar patients with cyclothymic premorbid adjustment and interepisodic adjustment are at considerable risk of *antidepressant-induced rapid cycling,* defined as a rapid succession of major episodes with few or no intervals of freedom.

As Kraepelin illustrated in his monograph, we can best capture the course graphically (Fig. 13.3–3). After diagramming 18 illustrative patterns for the entire spectrum of manic-depressive illness, Kraepelin declared that the illness pursued an indefinite number of courses. Although not represented in the official U.S. and international classifications, some of these course patterns are of considerable interest today. For instance, a biphasic course (a sequence in which episodes of opposite polarity succeed each other, followed by a free interval) is relevant to treatment response. Thus, depression followed by mania or hypomania—compared to mania or hypomania followed by depression—appears less responsive to lithium based on replicated studies.

DEPRESSIVE DISORDERS

The broad category of depressive disorders includes major depressive disorder, dysthymic disorder, and depressive disorder not otherwise specified.

Major Depressive Disorder

Episodes usually begin over a prodromal period of weeks to months. The DSM-5 and text revision, or DSM-5-TR diagnosis of major depressive disorder requires one of the following: (1) dysphoric mood or (2) decreased interest in usual activities. Table 13.3–3 lists the full range of signs and symptoms. Such symptoms continue for at least 2 weeks, and there is no alternative process that can otherwise explain the depressive symptoms (such as normal bereavement or other physical or mental disorders known to cause depressive symptoms). It can be a single episode or, commonly, recurrent, or both.

Comorbid Physical Disease. The preceding considerations raise the question of whether major depressive disorder should be limited to depressions of unknown etiology (i.e., those without documented physical causes). The DSM-5 approach has been that, when the cause is known, we should diagnose the condition as a mood disorder due to a general medical condition (that must be specified) or substance-induced mood disorder. The problem with this approach is that many common medical factors historically associated with depression do not seem to be causative in the etiologic sense but rather are triggering agents in otherwise predisposed persons. This problem is analogous to the role of life events, which we no longer use to distinguish between reactive and endogenous subtypes of depression. A more troubling implication is that major depressive disorders without demonstrable physical disease are not medical or otherwise biologic. More importantly, there appears to be no reliable or valid way for a clinician to decide that a depressive condition is due to a specified medical condition. In brief, the designation *due to a general medical condition* is cumbersome and redundant. We would be better off thinking of major depressive disorder as a final common pathway of multifactorial interacting factors—physical and psychological—a syndrome that we should diagnose irrespective of the presumed cause.

Table 13.3–3.
Major Depressive Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic Name	Major Depressive Disorder	Major Depressive Episode	Major Depressive Disorder
Duration	2-week period	N/A	
Symptoms	Dysphoria/Depressed mood Anhedonia Change in weight or appetite Change in sleep pattern Psychomotor disturbance: agitation or retardation Fatigue/decreased energy Depressive ruminations: worthlessness, guilt Poor concentration Recurrent thoughts of death, suicidal ideation, or a plan	Lowered mood Reduced energy Decreased activity Reduced capacity for enjoyment Reduced interest Reduced concentration Fatigue after even minimal effort Disturbed sleep/early morning awakening Disturbed appetite/weight loss Poor self-esteem Poor self-confidence Guilt or worthlessness Mood unreactive to circumstances Anhedonia Worse symptoms in the AM Psychomotor disturbance: agitation or retardation Loss of libido	**Affective cluster** ▶ Depressed mood ▶ Anhedonia **Cognitive-behavioral cluster** ▶ Poor concentration ▶ Low self-worth or inappropriate guilt ▶ Hopelessness ▶ Recurrent thoughts of death, suicidal ideation, or a plan **Neurovegetative cluster** ▶ Disrupted sleep ▶ Change in appetite ▶ Psychomotor agitation/retardation ▶ Fatigue/decreased energy
Required number of symptoms	5 (1 of these has to be symptom 1 or 2 listed above)	N/A	5 (≥1 from the affective cluster)
Psychosocial	Distress or impaired	Depends on	Significant impairment

consequences	functioning (social, occupational, or other significant areas)	severity	in ≥1 domain
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness History of: ▸ manic episode ▸ hypomanic episode	Adjustment disorder Conduct disorder Recurrent depressive disorder (which is considered a separate diagnosis)	Bereavement Mixed episode bipolar disorder Other medical consideration Effects of substance or medication
Symptoms specifiers	**With anxious distress** ▸ 2+ symptoms of anxiety **With mixed features** ▸ 3+ manic/hypomanic symptoms occurring during the depressive episode (if occur independently, diagnose bipolar disorder) **With melancholic features** ▸ Loss of pleasure or reactivity to pleasure ▸ 3+ of the following: ▸ Severe depression/despair ▸ Mood worse in AM ▸ Early morning awakening ▸ Psychomotor disturbance ▸ Anorexia/weight loss ▸ Guilt **With atypical features** ▸ Mood reactivity ▸ ≥2: ▸ Increased appetite/weight ▸ Hyposomnia ▸ Leaden paralysis ▸ Rejection sensitivity ▸ With mood-congruent psychotic features	Depressive reaction Psychogenic depression Reactive depression	Mild Moderate Severe Unspecified With/without psychotic symptoms With prominent anxiety symptoms With panic attacks Current depressive episode persistent Current depressive episode with melancholia With seasonal pattern Associated with pregnancy, childbirth, or puerperium

	► With mood-incongruent psychotic features **With psychotic features:** symptoms occur only during a depressive episode. Also specify if delusions are mood congruent or incongruent. **With catatonia** ► Must be present during most depressive episodes **With peripartum onset** **With seasonal pattern** ► Usually occurs during a specific season		
Course specifiers	**In partial remission:** full criteria no longer met **In full remission**: no symptoms of an episode for the past 2 months	**Recurrent depressive disorder** (coded as a separate disorder). Repeated episodes of the above symptoms. No mania.	**In partial remission** **In full remission**
Severity specifiers	**Mild:** Minimum # symptoms **Moderate:** between mild and severe **Severe:** # of symptoms, the severity of symptoms, or dysfunction is much more than required for a diagnosis	**Mild** ► 2–3 symptoms. ► Functions normally despite the distress **Moderate** ► 4+ symptoms ► Difficulty with functioning **Severe** ► Several symptoms marked and distressing	

- Loss of self-esteem/feels worthless and guilty
- Suicidal ideation/acts
- Somatic symptoms of depression

Severe with psychotic symptoms

- As above but with psychosis

Other

- Atypical depression
- Single episodes of “masked depression”

Unspecified

Table 13.3–4.
Persistent Depressive Disorder/Dythymia/Dysthymic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic Name	Persistent Depressive Disorder	Dysthymia	Dysthymic Disorder
Duration	Symptoms continuously present for ≥2 years No time during the affected period of no symptoms for >2 months		≥2 years
Symptoms	Depressed mood most of the time on a majority of days over a period of ≥2 years (NOTE: in children, irritability may be present, and duration must be for a period of ≥1 year) During the depressed period: ▸ Abnormal appetite (↑ or ↓) ▸ Abnormal sleep (↑ or ↓) ▸ ↓ energy ▸ ↓ self-esteem ▸ ↓ concentration or ↑ indecisiveness ▸ Hopelessness	Chronically depressed mood that does not meet criteria for a depressive episode, though criteria may have been met in the past.	Persistent depressed mood Additional symptoms of depression. May include: ▸ Anhedonia or decreased interest ▸ Poor concentration ▸ Low self-worth or inappropriate guilt ▸ Hopelessness ▸ Disturbed sleep ▸ Change in appetite ▸ Fatigue or low energy
Required number of symptoms	≥2 bulleted symptoms		
Psychosocial consequences	Marked distress and/or functional impairment		Marked distress and/or functional impairment
Exclusions	This cannot occur if a history of bipolar disorder Other medical condition Substance or medication, including withdrawal Other mental illness	Result of: ▸ Anxiety ▸ Depression ▸ Bereavement ▸ schizophrenia	A depressive episode in the first 2 years of the disorder History of manic, mixed, or

		hypomanic episodes Other medical condition Substance or medication, including withdrawal
Symptoms specifiers	**With pure dysthymic syndrome:** criteria for the depressive episode not met in the last 2 years **With persistent major depressive episode:** has met diagnostic criteria for a depressive episode for an entire 2-year period **With intermittent major depressive episode, with current episode:** though may currently meet criteria for major depressive disorder, has had a period of ≥8 weeks during the prior 2-year period during which symptoms were below the diagnostic threshold for a depressive episode **With intermittent major depressive episode, without current episode:** Does not currently meet criteria for a depressive episode, though has experienced ≥1 depressive episode over the course of the prior 2 years. **With anxious distress:** including ≥2 symptoms among feeling tense, restlessness, difficulty with concentration due to worrying, excessive fear without an identifiable cause, fear of loss of control **With mixed features—**full criteria met for either depressive or manic/hypomanic episode, with	

	additional symptoms of depressive or manic/hypomanic period without also meeting full criteria
	With melancholic features—involving loss of pleasure, reduced reactivity with ≥3 of the following: ▶ Depressed mood/despair ▶ Depressive symptoms worse in the morning ▶ Early morning awakening ▶ Psychomotor changes ▶ Loss of appetite/weight loss ▶ Excessive guilt
	With atypical features: during most days, an individual is experiencing mood reactivity with ≥2 symptoms: ▶ Weight gain or increased appetite ▶ Excessive sleeping ▶ Leaden paralysis ▶ A pattern of heightened interpersonal rejection sensitivity
	With mood-congruent psychotic features—mood-congruent delusions or hallucinations
	With mood-incongruent psychotic features—mood-incongruent delusions or hallucinations
Course specifiers	**With peripartum onset—**full criteria for a mood episode are met during pregnancy or within 4 weeks of delivery
Severity specifiers	**Mild—**few symptoms in excess of minimum diagnostic criteria, with no to minimal impairment and/or distress
	Moderate—moderate symptomatology and impairment

Severe—symptoms in marked excess of minimum diagnostic requirement, with marked distress and impairment in functioning

Table 13.3–5.
Bipolar 1 Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic Name	Bipolar 1 Disorder	Bipolar Affective Disorder ***NOTE*:** ICD-10 does not distinguish between Bipolar 1 and 2, requiring only a history of discrete episodes of mania, hypomania, and/or depression, with episodes demarcated by switches in mood/affect.	Bipolar 1 Disorder
Duration	Manic episode: 1 week+ Hypomanic episode: 4 days+ Major depressive episode: 2 weeks+	Manic episode: 1 week+ Hypomanic episode: 4 days+ Major depressive episode: 2 weeks+	1 week+
Symptoms	**Manic or hypomanic episodes** ▸ Abnormally ↑ or irritable mood (required) ▸ Grandiose thoughts ▸ ↓ need for sleep ▸ Pressured speech ▸ Racing and expansive thoughts ▸ Distractibility ▸ Hyperactivity ▸ Impulsivity/high-risk activities **Depressive episodes** Similar to that for major depressive disorder	History of episodes of mania, hypomania, and/or depression Switches in mood/affect **Mania** ▸ Abnormally ↑ or irritable mood (required) ▸ Abnormally elevated or irritable mood ▸ ↑ activity ▸ ↑ talkativeness ▸ Flight of ideas/racing thoughts ▸ Social disinhibition ▸ ↓ need for sleep ▸ Grandiose thoughts ▸ Distractibility ▸ Impulsivity/recklessness ▸ Hypersexuality **Hypomania** ▸ Abnormally ↑ mood (required) ▸ Psychomotor agitation ▸ ↑ talkativeness ▸ Poor concentration/distractibility ▸ ↓ need for sleep ▸ Hypersexuality	**Manic episode** ▸ Extreme mood state (euphoria, irritability of expansiveness) ▸ Increased activity ▸ Talkativeness ▸ Flight of ideas ▸ Increased self-esteem ▸ Decreased need for sleep ▸ Distractibility ▸ Impulsivity ▸ Increased libido, sociability, or goal-directed activity **Mixed episode** Several prominent manic and several prominent depressive symptoms

		► Impulsivity/ or ↑ spending ► Over-familiarity **Depressive episode** ► Depressed mood ► Loss of interest or pleasure ► Decreased energy *Additional symptoms:* ► Low self-esteem ► Excessive guilt or shame ► Recurrent thoughts of death or suicide ► Poor concentration ► Psychomotor changes ► Sleep disturbance ► Change in appetite and/or weight	
Required number of symptoms	≥1 manic episode Abnormally ↑ or irritable mood required 3 of the other symptoms (4 if irritable mood)	**Bipolar affective disorder, current episode hypomanic** Hypomania (elevated mood and ≥3 symptoms History of a prior affective episode (manic, hypomanic, depressed, mixed) **Bipolar affective disorder, current episode manic** Mania (elevated or irritable mood and ≥3 symptoms (4 if irritable mood) History of prior affective episodes (manic, hypomanic, depressed, mixed) **Bipolar disorder, current episode mild/moderate/severe depression** depressive episode (2/3 of the first 2 symptoms and 2 more of the additional symptoms) History of prior affective episodes (manic,	≥1 manic or mixed episode consisting of first 2 symptoms listed above, plus several of the other symptoms listed.

		hypomanic, depressed, mixed) **Bipolar disorder, current episode mixed** A mixture or rapid alternation of hypomanic, manic, and/or depressive symptoms (defined above) History of prior affective episodes (manic, hypomanic, depressed, mixed)	
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness	Psychoactive substance use Another mental disorder	Other medical condition Substance or medication, including withdrawal
Psychosocial Consequences	***Manic Episode:*** impaired functioning or needing hospitalization Hypomanic episode: No impairment or need for hospitalization Depressive episode: marked distress and/or psychosocial impairment		Impairment in ≥1 functional domain, or requiring intensive treatment (hospitalization), or includes hallucinations or delusion.
Symptom specifiers	**Mild—**2 symptoms **Moderate—**3 symptoms **Moderate-severe—4–5** symptoms **Severe—**4–5 symptoms + motor agitation **With mixed features**	**Current episode hypomanic** **Current episode manic without psychotic symptoms** **Current episode manic with psychotic symptoms** **Mood congruent** **Mood incongruent** **Current episode of depression** **Current episode mixed**	**Current/most recent episode:** Manic Hypomanic Depressive Mixed **With/without psychotic symptoms** **With prominent anxiety**

	Either a depressive or manic/hypomanic episode Additional symptoms of depressive or manic/hypomanic period (not full criteria) **With rapid cycling** ≥4 mood episodes in 1 year ≥2-month period of partial/full remission between episodes **With melancholic features** Similar to that for major depressive disorder **With atypical features** Similar to that for major depressive disorder **With anxious distress:** ≥2 symptoms among the following: Feeling tense Restlessness Difficulty with concentration due to worrying XS fear without cause Fear of loss of control	**Currently in remission**–no symptoms. History of previous episodes. ***With somatic syndrome*** *≥4 of the following:* ▸ *Loss of interest or pleasure* ▸ *Lack of emotional reactivity* ▸ *Early morning awakening* ▸ *Depression gets worse in the mornings* ▸ *Psychomotor changes* ▸ *Loss of appetite* ▸ *Weight loss* ▸ *Loss of sexual drive*	**symptoms** **With panic attacks** **With Seasonal pattern** **With rapid cycling** **Associated with pregnancy, childbirth or the puerperium**
Course specifiers	**With peripartum onset:** episode occurs during pregnancy or within 4 weeks after delivery		

	With seasonal pattern: pattern present for ≥2 years		
Severity specifiers		**For the current episode of depression** **Mild:** 4 symptoms **Moderate:** 6 symptoms, including ≥2 among loss of pleasure, depressed mood, and low energy **Severe depression** ≥8 symptoms, including all loss of pleasure, depressed mood, and low energy **without psychotic symptoms** **with psychotic symptoms** **mood congruent** **mood incongruent**	**Mild** **Moderate** **Severe** **Unspecified severity**

Diagnostic Threshold. Another question concerning the definition of major depressive disorder relates to the threshold at which a constellation of depressive features becomes a condition distinct from the ordinary blues. Within the current definition, if a person responds to a setback with lowered spirits and self-doubt, difficulty in sleeping and concentration, and decreased sexual interest for 14 days, they would qualify for the diagnosis of a major depressive disorder of mild intensity. Many clinicians would consider such a condition a relatively minor departure from normality, probably no more than an adjustment disorder. Criteria other than signs, symptoms, and duration are necessary to differentiate a major depressive disorder from adjustment reactions to life situations. The presence of the following characteristics might assist in such a differentiation:

Table 13.3–6.
Bipolar 2 Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic Name	Bipolar 2 Disorder	Bipolar Affective Disorder **_*NOTE_**: ICD-10 does not distinguish between Bipolar 1 and 2, requiring only a history of discrete episodes of mania, hypomania, and/or depression, with episodes demarcated by switches in mood/affect.	Bipolar Type II Disorder
Duration	**Hypomanic episode:** lasting ≥4 days, with symptoms, present most of the day for most days **Major depressive episode**: symptoms lasting ≥2 weeks	Manic episode: lasting ≥1 week Hypomanic Episode: lasting ≥4 days, Major depressive episode: symptoms lasting ≥2 weeks	Several days
Symptoms	Hypomanic episodes involve abnormally elevated or irritable mood with: ▶ Grandiosity ▶ Decreased need for sleep ▶ Pressured speech ▶ Racing thoughts with expansiveness of thought ▶ Distractibility ▶ Goal-directed or non-goal-directed hyperactivity ▶ Engagement in high-risk activity or increased impulsivity Depressive episode involving depressed mood and/or decreased	History of discrete episodes of mania, hypomania, and/or depression, with episodes demarcated by switches in mood/affect. Episodes are defined as follows: **Mania:** abnormally elevated or irritable mood with ≥3 of the following (≥4 if mood irritable) ▶ Increased activity ▶ Increased talkativeness ▶ Flight of ideas/racing thoughts ▶ Social disinhibition ▶ Decreased need for sleep ▶ Grandiosity ▶ Distractibility ▶ Increased recklessness and impulsivity ▶ Hypersexuality **Hypomania:** abnormally elevated mood with ≥3 of the following: ▶ Psychomotor agitation ▶ Increased talkativeness	**Hypomanic episode** Elevation of mood or irritability outside of the usual mood range Increased activity Talkativeness Flight of ideas Increased self-esteem Decreased need for sleep Distractibility Impulsivity Increased libido, sociability, or goal-directed activity **Depressive episode**

	interest/pleasure with: ▶ Unintentional weight loss or change in appetite ▶ Disruption of sleep ▶ Psychomotor changes ▶ Loss of energy ▶ Difficulty with concentration and/or attention ▶ Excessive feelings of worthlessness and/or guilt ▶ Recurrent thoughts of death, including passive and/or active suicidal ideation	▶ Poor concentration or increased distractibility ▶ Decreased need for sleep ▶ Hypersexuality ▶ Impulsivity or increased spending ▶ Over-familiarity Depressive episode: 4 total symptoms from among the following ≥2: ▶ Depressed mood ▶ Loss of interest or pleasure ▶ Decreased energy Additional symptoms: ▶ Low self-esteem ▶ Excessive guilt or shame ▶ Recurrent thoughts of death or suicide ▶ Poor concentration ▶ Psychomotor changes ▶ Sleep disturbance ▶ Change in appetite and/or weight	*See depressive disorders*
Required number of symptoms	Lifetime history of a hypomanic episode involving elevated/irritable mood with ≥3 of the above symptoms (4 if mood irritable) Lifetime history of a depressive episode with ≥5 of the above symptoms, which must include ≥1 depressed mood or loss of pleasure	**Bipolar affective disorder, current episode hypomanic—**meets the above criteria for hypomania (elevated mood with ≥3 listed symptoms) and has a history of prior affective episodes (manic, hypomanic, depressed, mixed) **Bipolar affective disorder, current episode manic—**meets above criteria for mania (elevated or irritable mood with ≥3–4 listed symptoms) and has a history of prior affective episodes (manic, hypomanic, depressed, mixed)	≥1 hypomanic episode ≥1 depressive episode

		Bipolar disorder, current episode mixed—mixture or rapid alteration in hypomanic, manic, and/or depressive symptoms, with symptoms, present most of the time for ≥2 weeks, in addition to having a history of prior affective episodes (manic, hypomanic, depressed, mixed) **Bipolar disorder, current episode mild/moderate/severe depression—**current symptoms meet criteria for a depressive episode, in addition to having a history of prior affective episodes (manic, hypomanic, depressed, mixed)	
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness	Psychoactive substance use Another mental disorder	H/o manic episodes Other medical condition Effect of substance or medication
Psychosocial consequences	Hypomanic episode: NO marked impairment in functioning and NO hospitalization necessitated Depressive episode: marked distress and/or impairment in psychosocial functioning		Does not cause marked functional impairment
Symptom specifiers	**Current episode:** ***Depressed*** ***Hypomanic*** **Additional specifiers to be added to the above:**	**Current episode hypomanic** **Current episode manic without psychotic symptoms** **Current episode manic with psychotic symptoms**	**Current/most recent episode** Hypomanic Depressive

With anxious distress: including ≥2 symptoms among feeling tense, restlessness, difficulty with concentration due to worrying, excessive fear without an identifiable cause, fear of loss of control

With mixed features—full criteria met for either depressive or manic/hypomanic episode, with additional symptoms of depressive or manic/hypomanic period without also meeting full criteria

With rapid cycling—the presence of ≥4 mood episodes over the course of a 12-month period, with ≥2-month period of partial/full remission between mood episodes

With melancholic features—involving loss of pleasure, reduced reactivity with ≥3 of the following:

- Depressed mood/despair
- Depressive symptoms worse in the morning
- Early morning awakening
- Psychomotor changes

- *With mood-congruent psychotic symptoms*
- *With mood-incongruent psychotic symptoms*

Current episode moderate or mild depression— ≥4 (mild) or 6 (moderate) symptoms present, including ≥2 among loss of pleasure, depressed mood, and low energy

With somatic syndrome—*≥4 of the following:*

- *Loss of interest or pleasure*
- *Lack of emotional reactivity*
- *Early morning awakening*
- *Depression gets worse in the mornings*
- *Psychomotor changes*
- *Loss of appetite*
- *Weight loss*
- *Loss of sexual drive*

Without somatic syndrome

Current episode severe depression without psychotic symptoms—≥8 symptoms present, including all 3 of loss of pleasure, depressed mood, and low energy

Current episode severe depression with psychotic symptoms—≥8 symptoms present, including all 3 of loss of pleasure, depressed mood, and low energy, occurring either

- *With mood-congruent psychotic symptoms*
- *With mood-incongruent psychotic symptoms*

Current episode mixed

Currently in remission—does not currently meet

With/without psychotic symptoms

With prominent anxiety symptoms

With panic attacks

With seasonal pattern

With rapid cycling

Associated with pregnancy, childbirth, or the puerperium

	▸ Loss of appetite/weight loss ▸ Excessive guilt	criteria for an affective episode despite a prior history of a manic or hypomanic episode
	With atypical features: during most days, an individual is experiencing mood reactivity with ≥2 symptoms of: ▸ Weight gain or increased appetite ▸ Excessive sleeping ▸ Leaden paralysis ▸ A pattern of heightened interpersonal rejection sensitivity	
	With mood-congruent psychotic features—mood-congruent delusions or hallucinations	
	With mood-incongruent psychotic features—mood-incongruent delusions or hallucinations	
	With catatonia—catatonic features present during most of the mood episode	
Course specifiers	**With peripartum onset—**full criteria for a mood episode are met during pregnancy or within 4 weeks of delivery	
	With seasonal pattern (for ≥1 type of episode)—pattern present for ≥2 years	

	In partial remission **In full remission**		
Severity specifiers	**Mild—**few symptoms in excess of minimum diagnostic criteria, with no to minimal impairment and/or distress **Moderate—**moderate symptomatology and impairment **Severe—**symptoms in marked excess of minimum diagnostic requirement, with marked distress and impairment in functioning	See the above criteria for mild, moderate, and severe depression	Mild Moderate Severe
Comments	**Note**: A diagnosis of Bipolar 2 cannot be made if there is a history of a manic episode, in which case a diagnosis of Bipolar 1 disorder would be made.		

Table 13.3–7.
Cyclothymic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic Name	Cyclothymic Disorder	Cyclothymia	Cyclothymic Disorder
Duration	Symptoms lasting ≥2 years (≥1 year in children), with depressive and hypomanic symptoms present for ≥half the time.	2-year period involving periods of depression and elevated mood, without ever meeting criteria for a depressive episode or manic episode.	≥2 years
Symptoms	Hypomanic episodes involve abnormally elevated or irritable mood with: ▸ Grandiosity ▸ Decreased need for sleep ▸ Pressured speech ▸ Racing thoughts with expansiveness of thought ▸ Distractibility ▸ Goal-directed or non-goal-directed hyperactivity ▸ Engagement in high-risk activity or increased impulsivity Depressive Episode involving depressed mood and/or decreased interest/pleasure with: ▸ Unintentional weight loss or change in appetite ▸ Disruption of sleep ▸ Psychomotor changes ▸ Loss of energy ▸ Difficulty with concentration and/or attention ▸ Excessive feelings of worthlessness and/or guilt	Periods of depressed mood involving: ▸ Low energy ▸ Insomnia ▸ Low self-esteem ▸ Poor concentration ▸ Social withdrawal ▸ Loss of pleasure/enjoyment ▸ Reduced talkativeness ▸ Excessive pessimism Period of elevated mood involving: ▸ Increased energy ▸ Decreased need for sleep ▸ Inflated self-esteem ▸ Increased creativity ▸ More talkative ▸ More socially outgoing ▸ Hypersexuality and increased participation in pleasurable activities ▸ Grandiosity	Mood instability, with numerous hypomanic (see Bipolar type II disorder) and depressive periods (see depressive disorders). Children instead may show irritability. Hypomanic periods may or may not have been severe or long enough to meet criteria for hypomanic episode.

	► Recurrent thoughts of death, including passive and/or active suicidal ideation		
Required number of symptoms	Symptoms of hypomania and depression present without ever meeting the full criteria for a depressive episode or manic episode	≥3 symptoms during periods of mood elevation without meeting the full criteria for a. manic episode ≥3 symptoms during periods of low mood without meeting the full criteria for a depressive episode	–
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illnesses (Bipolar I or II)	Bipolar affective disorder (though a history of bipolar affective disorder may be present)	Other medical condition Substance or medication, including withdrawal Other mental illness
Psychosocial consequences	Marked distress or impairment in areas of functioning		Marked distress or impairment in areas of functioning
Symptom specifiers	**With anxious distress:** including ≥2 symptoms among feeling tense, restlessness, difficulty with concentration due to worrying, excessive fear without an identifiable cause, fear of loss of control		
Course specifiers		**Early onset:** before the ’20s **Late onset:** onset between ages 30 and 50	
Severity specifiers	**Mild—**2 symptoms **Moderate—**3 symptoms		

Moderate-Severe—4 or 5 symptoms

Severe—4–5 symptoms with motor agitation

Table 13.3–8.
Differentiating Characteristics of Bipolar and Unipolar Depressions

	Bipolar	**Unipolar**
History of maina or hypomania (definitional)	Yes	No
Temperament and personality	Cyclothymic and extroverted	Dysthymic and introverted
Sex ratio	Equal	More women than men
Age of onset	Teens, 20s, and 30s	30s, 40s, and 50s
Postpartum episodes	More common	Less common
Onset of episode	Often abrupt	More insidious
Number of episodes	Numerous	Fewer
Duration of episode	3–6 months	3–12 months
Psychomotor activity	Retardation > agitation	Agitation > retardation
Sleep	Hypersomnia > insomnia	Insomnia > hypersomnia
Family history		
Bipolar disorder	Yes	±
Unipolar disorder	Yes	Yes
Alcoholism	Yes	Yes
Pharmacologic response		
Most antidepressants	Induce hypomania–mania	±
Lithium carbonate	Prophylaxis	±

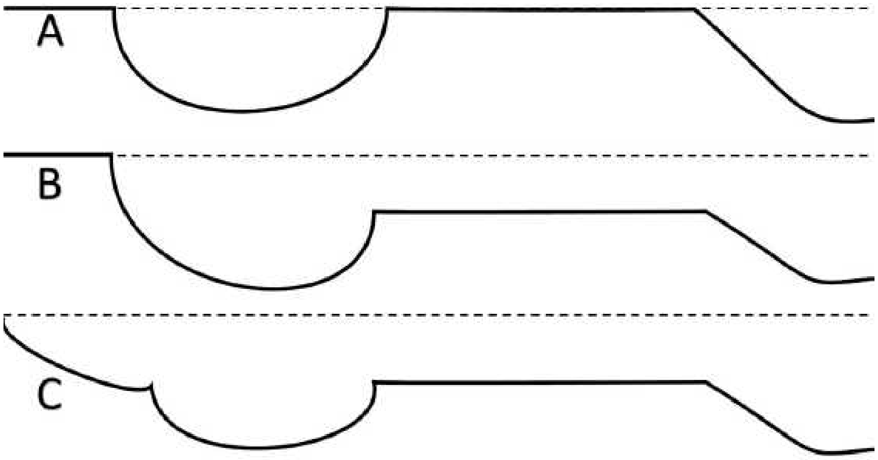

FIGURE 13.3–3. Typical courses of depression represented graphically. The *dotted line* indicates euthymia. **A:** Major depressive disorder, recurrent, with complete remission between episodes. **B:** Major depressive disorder, recurrent, with partial remission between episodes. **C:** Major depressive disorder, recurrent, with an overlapping dysthymic disorder, hence antecedent dysthymia and a return to that dysthymia between episodes, often referred to as double depression. (Data from Keller MB, Boland RJ, Leon A, Solomon D, Endicott J, Li C. Clinical course and outcome of unipolar major depression. In: Keller MB, Coryell WH, Maser JD, eds. *Clinical Guide to Depression and Bipolar Disorder: Findings from the Collaborative Depression Study*. American Psychiatric Press; 2013.)

- By definition, a major depressive disorder should be incapacitating. Previously, we paid much attention to the interpersonal consequences of depression. Recent evidence indicates that measurable deficits in work performance are often early manifestations. Afflicted persons also do not benefit from taking leisure time, and, hence, prescribing vacations is futile.
- We perceive major depressive disorder as a break from a person's usual or premorbid self, which can be so striking that patients may feel they are losing their minds. The critical point is that the patient and significant others can usually relate the onset of the illness to a given month or quarter of a year, which is not valid, for instance, for dysthymic disorder.
- The patient often experiences major depressive disorder as qualitatively distinct from grief or other understandable reactions to loss or adversity. William James described it as follows:

> *There is a pitch of unhappiness so great that the goods of nature may be entirely forgotten, and all sentiment of their existence vanish from the mental field. For this extremity of passion to be reached, something more is needed than adversity; the individual must in his own person become the prey of pathological melancholy. Such sensitiveness and susceptibility of mental pain is a rare occurrence where the nervous constitution is entirely normal: one seldom finds it in a healthy subject even where he is the victim of the most atrocious cruelties of outward fortune; it is an active anguish, a sort of psychical neuralgia wholly unknown to healthy life.*

Two additional features, when present, would further validate the diagnosis of major depressive disorder.

- History of past episodes.
- Consecutive-generation family history of mood disorder—especially when many family members are afflicted with depression or mood disorder—is characteristic of clinical

depression. For instance, one study that prospectively followed persons with minor or neurotic depression found that such pedigrees predicted the development of future major episodes. In clinical practice, these factors would strongly weigh whether to take the depression seriously.

Single-Episode and Recurrent Subtypes. A significant minority—perhaps one-third—of all major depressive episodes do not recur. Such patients tend to be older and less likely to have a positive family history of mood disorders and have a more protracted (1 to 2 years) course of the disorder. We should distinguish patients with a single-episode major depressive disorder from those experiencing their first episodes of recurrent major depressive disorder. The latter group tends to be younger, and the disorder is more likely to have been preceded by a depressive temperament or dysthymic disorder. Those who switch to bipolar disorder are more likely to have experienced recurrent depressions (as many as five or more episodes).

Research has established that recurrent major depressive disorders are more familial than their single-episode counterparts. The average length of episodes is 6 months, whereas the mean interval between episodes tends to vary (typically years). According to retrospective and prospective studies, the mean number of major episodes over a lifetime is five to six, in contrast to an average of eight to nine major episodes in bipolar disorder. These figures are likely an underestimate as they rely on such data as clinical referrals or hospitalization, or both.

Melancholic Features. *Melancholic features* is a qualifying phrase for major depressive disorders in which anhedonia, guilt, and psychomotor–vegetative disturbances dominate the clinical picture. In addition, severe suicidal ideation may be present.

Atypical Features. Reverse vegetative signs with rejection sensitivity, often contrasted to melancholia on phenomenologic and pharmacologic grounds, represent a major depressive disorder qualifier occurring in as many as one-third of all major depressive disorders. Atypical features are so common in bipolar disorder, especially bipolar II disorder, that some consider them to be clinical markers for soft bipolar disorders. Although there is no consensus on this question, before diagnosing major depressive disorder in such cases, it is clinically wise to exclude bipolar II disorder, especially where hostile–labile features predominate.

Mixed Features. DSM-5 added a category of “mixed features” to its major depressive subtypes. This subtype describes a category of patients with major depressive disorder or who display subsyndromal symptoms of mania or hypomania but do not meet the criteria for either syndrome. This subtype reflects the reality that many patients diagnosed with major depressive disorder have occasional mania-type symptoms. However, the new subtype blurs an already blurry boundary between depressive disorders and bipolar disorders. Some experts have questioned whether antidepressants are the preferred treatment for patients in this category or whether mood stabilizers are preferable. Existing studies have focused on atypical antipsychotics. We currently lack sufficient evidence to make a definitive statement about the implications of this category and await further longitudinal studies.

Psychotic Features. The separation of depressed patients with psychotic symptoms—*psychotic depression*—as a distinct subtype of depression was spurred by observations that patients whose depressive disorder included psychotic features had a differential treatment response than other depressive patients. Early subtyping treated psychotic depression as a severe type of major depression. However, newer data suggest that the subtype includes a range of severity. However, in most cases, patients with psychotic depression have more severe depression, including melancholic features. Perhaps 10% to 20% of depressed adults have psychotic symptoms, although clinicians often miss these symptoms in their depressed patients, so this may be an underestimate.

Typically, the symptoms are hallucinations or delusions. The delusions are classically considered to be mood congruent, often self-deprecatory or imagined poverty or illness. There are cases of mood-incongruent delusions, which seem to have a worse outcome, although this is not a consistent finding. The trait of psychosis is likely transdiagnostic, and relatives of patients with psychotic depression may have relatives with other psychotic illnesses. Patients with this subtype are more likely than other depressed patients to progress to a bipolar illness.

Chronic Depression. The symptom profile in chronic depressions usually displays low-grade intensity rather than severe syndromic chronicity. Severe depressive disorder, in its psychotic forms, is so agonizing that the patient is at risk of committing suicide before the disorder has a chance to become chronic. The psychotic symptoms often respond to medication or ECT, but residual depressive symptoms may linger for a long time. In other persons with chronic depressions, the chronicity arises from more mundane (nonpsychotic) major depressive episodes, representing depressive residua following one or several clinical episodes that fail to remit fully. Instead of the customary remission within 1 year, the patients are ill for years. The level of depression varies, fluctuating between syndromic illness and milder symptoms. A landmark analysis from the NIMH Collaborative Depression Study by Lewis Judd and colleagues showed that as many as 60% of patients with major depressive disorder have a subthreshold fluctuating chronic course.

Rather than exhibiting a frankly depressive mood, many persons with chronic depression experience deficits in their ability to enjoy leisure and display an attitude of irritable moroseness. They also show a sense of resignation, generalized fear of an inability to cope, adherence to rigid routines, and inhibited communication. Such deficits, along with irritable humor, tend to poison their conjugal lives: Their marriages are typically in a state of chronic deadlock, leading to neither divorce nor reconciliation. Somatic features dominate the residual phase in other patients, such as sleep and other vegetative or autonomic irregularities. The persistent sleep EEG (especially REM and delta phase) abnormalities that are indistinguishable from their acute depressed counterparts demonstrate that the interpersonal, conjugal, and autonomic manifestations represent unresolved depression. Unfortunately, self-treatment with alcohol or iatrogenic benzodiazepine dependence rather than definitive treatment of the ongoing low-grade depression is common.

Failure to recover from major depressive disorder is associated with increased familial loading for depression, disabled spouses, deaths of immediate family members, concurrent disabling medical disease, relying on depressant pharmacologic agents, or excessive alcohol and sedative–hypnotic agents. Persons with residual depression through the death or illness of significant others may then lack social support. Therefore, a thorough medical evaluation and socially supportive interventions should be essential ingredients of the overall approach to these patients.

Interpersonal disturbances in such patients are usually secondary to the distortions produced by longstanding depression. Therefore, observed pathologic characterologic changes—clinging or hostile dependence, demanding tendencies, touchiness, pessimism, and low self-esteem—are best considered as *postdepressive personality* changes. Dangerous stereotypical thinking holds that the disorder must have a characterologic substrate because a patient has not responded adequately to standard treatments (the illness has become chronic). The long duration of the disorder often leads the patient to identify with the failing functions of depression, producing the self-image of being a depressed person. This self-image represents a malignant cognitive manifestation of the depressive disorder and dictates vigorous treatment targeted at the mood disorder.

Dysthymic Disorder (Persistent Depressive Disorder in DSM-5)

Dysthymia, called persistent depressive disorder in DSM-5 (see Table 13.3–4), is distinguished from chronic depressive disorder because it is not a sequel to well-defined major depressive episodes. Instead, in the most typical cases, patients complain that they have always been depressed. Thus, most cases are of early onset, beginning in childhood or adolescence and certainly by the time patients reach their 20s. A late-onset subtype, much less prevalent and not well characterized clinically, has been identified among middle-aged and geriatric populations, mainly through epidemiologic studies in the community.

Although the dysthymic disorder category in DSM-5 can occur as a secondary complication of other psychiatric disorders, the core concept of dysthymic disorder refers to a subaffective disorder with (1) low-grade chronicity for at least 2 years, (2) insidious onset with origin often in childhood or adolescence, and (3) persistent or intermittent course. Although not part of the formal definition of dysthymic disorder, the family history is typically replete with depressive and bipolar disorders, which is one of the more robust findings supporting its link to primary mood disorder.

Social Adjustment. Dysthymic disorder is typically an ambulatory disorder compatible with relatively stable social functioning. However, the stability is precarious; recent data document that many patients invest whatever energy they have in work, leaving none for leisure and family or social activities, which results in marital friction. These empirical findings on persons with dysthymic disorder echo earlier formulations in German and Japanese literature. For instance, Kraepelin described such persons as follows: “Life with its activity is a burden which they habitually bear with dutiful self-denial without being compensated by the pleasures of existence.”

The dedication of persons with a dysthymic disorder to work may be an overcompensation and a defense against their battle with depressive disorganization and inertia. Nevertheless, Ernst Kretschmer suggested that such persons are the “backbone of society,” dedicating their lives to jobs that require dependability and great attention to detail. Epidemiologic studies have demonstrated that some persons with protracted dysthymic complaints, extending over many years, have never experienced clear-cut depressive episodes. Some may seek outpatient counseling and psychotherapy for what some clinicians might consider *existential depression,* with feelings of being empty and lacking any joy in life outside of their work. Some have described them as leading “monocategorical existences.” Others present clinically because their low-grade dysphoria has intensified into a major depressive disorder.

Course. An insidious onset of depression dating back to late childhood or the teens, preceding any superimposed major depressive episodes by years or even decades, represents the most typical developmental background of dysthymic disorder. A return to the low-grade depressive pattern is the rule after recovery from superimposed major depressive episodes, if any; hence, the designation *double depression* as a prominent course pattern. This pattern, commonly seen in clinical practice, consists of the baseline dysthymic disorder fluctuating in and out of depressive episodes. The more prototypical patients with dysthymic disorder often complain of having been depressed since birth or of feeling depressed all the time. In the apt words of Kurt Schneider, they seem to view themselves as belonging to an “aristocracy of suffering.” Such descriptions of chronic gloominess in the absence of more objective signs of depression earn these patients the label of *characterologic depression.* Further reinforcing the description is the fluctuating depressive picture that merges imperceptibly with the patient’s habitual personality makeup. This conceptual uncertainty notwithstanding, given the confluence of data on the efficacy of many classes of antidepressants, patients with dysthymia should not be denied the potential benefit of antidepressants.

Clinical Picture. The profile of dysthymic disorder overlaps with that of major depressive disorder but differs from it in that symptoms tend to outnumber signs (more subjective than objective depression). This means that marked disturbances in appetite and libido are uncharacteristic, and we do not see any psychomotor agitation or retardation. The result is depression with attenuated symptomatology. However, it is not uncommon to see subtle endogenous features: inertia and anhedonia that are characteristically worse in the morning.

Although dysthymic disorder represents a more restricted concept than its parent, neurotic depression, it is still quite heterogeneous. Anxiety is not a necessary part of its clinical picture, yet we often diagnose dysthymic disorder in patients with anxiety and neurotic disorders. That clinical situation is perhaps a secondary or *anxious dysthymia* or, in the framework of Peter Tyrer, as part of a *general neurotic syndrome.*

It is best to restrict dysthymic disorder to a primary disorder for greater operational clarity, which a nonmood disorder cannot explain. The essential features of such primary dysthymic disorder include habitual gloom, brooding, lack of joy in life, and preoccupation with inadequacy. Dysthymic disorder is a longstanding, fluctuating, low-grade depression experienced as part of the habitual self and representing an accentuation of traits observed in the depressive temperament. Sleep EEG data indicate that many persons with dysthymic disorder at baseline exhibit the sleep patterns of those with acute major depressive disorder, providing support for the constitutional nature of the disorder. Further evidence for that position comes from studies demonstrating high rates of familial affective disorder in dysthymic disorder or depressive temperament, or both.

The clinical picture of dysthymic disorder that emerges from the preceding description is quite varied, with some patients proceeding to major depression, whereas others manifest the pathology primarily at the personality level. The previous considerations suggest that a clinically satisfactory operationalism of dysthymia must include symptomatic, cognitive, and trait characteristics.

A 27-year-old male grade-school teacher presented with the chief complaint that life was a painful duty that had always lacked luster for him. He said that he felt "enveloped by a sense of gloom" that was nearly always with him. Although his peers respected him, he felt "like a grotesque failure, a self-concept I have had since childhood." He stated that he merely performed his responsibilities as a teacher and never derived any pleasure from anything he had done in life. He said he had never had any romantic feelings; sexual activity, in which he had engaged with two different women, had involved pleasureless orgasm. He said that he felt empty, going through life without any sense of direction, ambition, or passion, a realization that itself was tormenting. He had bought a pistol to end his "useless existence" but did not carry out suicide, believing that it would hurt his students and the small community where he lived.

Dysthymic Variants. Dysthymia is not uncommon in patients with chronically disabling physical disorders, particularly among older adults. Dysthymia-like clinically significant subthreshold depression lasting 6 or more months occurs in some neurologic conditions, including stroke. According to a WHO conference dedicated to the subject, depression worsens the prognosis for the underlying neurologic disease and deserves pharmacotherapy. Ongoing studies should provide more explicit clinical recommendations on this topic.

Prospective studies on children have revealed an *episodic course* of dysthymia with remissions, exacerbations, and eventual complications by major depressive episodes, 15% to 20% of which might even progress to hypomanic, manic, or mixed episodes postpuberty. Persons with dysthymic disorder presenting clinically as adults tend to pursue a chronic unipolar course complicated by major depression. They rarely develop spontaneous hypomania or mania. However, when treated with antidepressants, some may develop brief hypomanic switches that typically disappear when decreasing the antidepressant dose. In this particular subgroup of persons with dysthymic disorder, the family histories are often positive for bipolar disorder. These patients represent a clinical bridge between major depressive disorder and bipolar II disorder.

Premenstrual Dysphoric Disorder. Given the higher prevalence of depressive disorders in women, premenstrual affective changes—dysphoria, tension, irritability, hostility, and labile mood—have received clinical and, more recently, research attention. DSM-5 includes the diagnosis of premenstrual dysphoric disorder, focusing on these symptoms. As with most psychiatric disorders, it should be sufficient to affect social or work functioning or other mental and physical functions, such as concentration, energy, appetite, and sleep. It may also include gynecologic symptoms such as breast tenderness or swelling or bloating, or weight-gain sensations.

Although women with severe premenstrual complaints appear to have higher rates of lifetime major mood disorders, some data find that genetic and environmental factors contributing to premenstrual depression and major depressive disorders are primarily distinct. Furthermore, events such as migraine, epileptic attacks, and panic states may, in some instances, associate with the premenstrual phase. The preceding considerations suggest the hypothesis that premenstrual psychobiologic changes exacerbate different neuropsychiatric disorders to which women are otherwise predisposed. In DSM-5-TR Premenstrual Dysphoric Disorder is classified as a separate mood disorder, in which symptoms should be present in the week leading up to menses and then improving with the onset of menses and resolving the week afterward. This disorder is discussed in more detail in Chapter 31.1 Psychiatry and Reproductive Medicine.

Other Specified Depressive Disorder and Unspecified Depressive Disorder

Minor Depressive Disorder. In so-called minor depression, observed in primary care settings, the depression is subthreshold, milder than major depressive disorder, yet not protracted enough to be considered dysthymic. These varied manifestations of depression argue for a continuum model (Fig. 13.3–4) as initially envisaged by Kraepelin. Lewis Judd and collaborators at the University of California at San Diego have suggested that subthreshold depressive symptoms—without necessarily meeting the criterion for mood change—might represent the most common expressions of a depressive diathesis. From such a subsyndromal symptomatic depressive base, individuals predisposed to depressive illness fluctuate in and out of the various DSM-5 and subthreshold subtypes of depressive disorders. This viewpoint is presently most cogent for subsyndromal symptomatic depression that follows major depressive disorder, a strong predictor of subsequent frequent relapse or chronic course. There is a vital message for the clinician: Treat subsyndromal symptomatic depression residual to major depressive disorder.

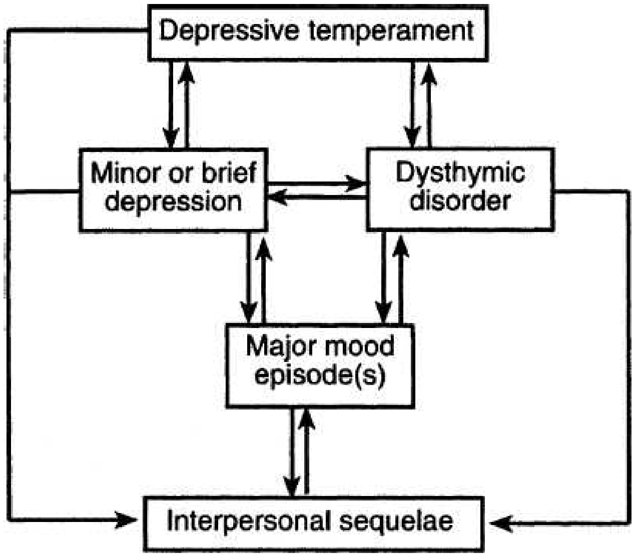

FIGURE 13.3–4. Relation of various depressive conditions supporting a spectrum concept. (From Akiskal HS. Dysthymia: clinical and external validity. *Acta Psychiatr Scand*. 1994;89(S383):19–23, with permission. doi:10.1111/j.1600-0447.1994.tb05879.x.)

Recurrent Brief Depressive Disorder. Recurrent brief depressive disorder derives from British work on young adults with frequent suicide attempts and epidemiologic studies conducted in a young adult cohort in Zurich. It is short-lived depressions that usually recur monthly but are *not* menstrually related. They could coexist with major depressive disorder or dysthymic disorder and with hypomania. Again, the existence of hypomanic-prone recurrent brief depressive disorder argues for a bridge between depressive and bipolar II disorders. Such patients are more prevalent in primary care than in psychiatric settings. Those seen in psychiatric settings are likely to be given personality disorder diagnoses, such as borderline personality disorder. The current nosologic status of recurrent brief depressive disorder is uncertain, but it testifies to Kraepelin's observation that many transitional forms link the depressive temperament to affective episodes:

> *A permanent gloomy stress in all the experiences of life usually perceptible already in youth, and may persist without essential change throughout the whole of life (or) there is actually an uninterrupted series of transitions to periodic melancholia in which the course is quite indefinite with irregular fluctuations and remissions.*

Reactive Depression. Classically, *reactive depression* results from a specific life event. In an ideal case, the depression would not have occurred without the event (e.g., love loss) to which it is a reaction. It continues as long as the event is present, and it terminates with the reversal of the event (e.g., return of the lover). We rarely see depressions exhibiting all of those features in clinical practice. With interpersonal support, most people can face life's reverses, which explains why reactive depression tends to be self-limiting. Hence, adjustment disorder is the more appropriate diagnosis for many cases of reactive depression.

Chronic Demoralization. Conceptually, however, one can envision chronically unsatisfactory life situations that might lead to chronic helplessness. However, such a condition, which could warrant chronic reactive depression, is a contradiction in terms. The question often raised is why a person would continue to stay in the situation. Sometimes psychodynamic authors invoke the concept of masochism to explain why certain persons cannot rid themselves of painful life situations, implying that they somehow contribute to their maintenance. Current thinking is that some presumed self-defeating traits are more situation specific than previously believed and might resolve with eliminating the situation. So-called self-defeating features are best considered psychodynamic mechanisms rather than indicators of a specific personality. This concept is not well understood and does not currently deserve to be considered a nosologic entity (it is not in DSM-5). Chronic adjustment disorder might describe the chronic demoralization observed among some individuals stuck in chronically unsatisfactory life situations. Others might fulfill the criteria for dysthymia. Finally, if the chronically unsatisfactory life conditions involve conjugal violence and trauma, PTSD may be the appropriate diagnosis.

Neurasthenia. A century-old term developed by the American neuropsychiatrist George Beard, *neurasthenia* refers to a more chronic stage of anxious–depressive symptomatology. The anxiety generated by overstimulation is so excessive that a chronic disposition to irritability, fatigue (especially mental fatigue), lethargy, and exhaustion replaces it. It is as if the patient's mind refuses to take on new stresses. The clinical picture described by Beard suggests that anxious manifestations were pre-eminent in his time. They included headache, scalp tenderness, backache, heavy limbs, vague neuralgias, yawning, dyspepsia, palpitations, sweating hands and feet, chills, flushing, sensitivity to weather changes, insomnia, nightmares, pantophobia, asthenopia, and tinnitus.

Although the diagnosis of neurasthenia is now used more in China than in the rest of the world, the recent worldwide popularity of the concept of chronic fatigue syndrome attests to the clinical acumen of classic physicians. Despite much energy invested in finding a viral or immunologic cause, current descriptions tend to suggest an anxiety or mood disorder basis for many, but certainly not all, those with the syndrome. Some patients with neurasthenia would meet the criteria for major depressive disorder with atypical features and a history of panic and phobic disorders. What circumstances would lead anxiety or depression to manifest primarily in fatigue is as elusive as it was 100 years ago. Like many other patients presenting to primary care settings with somatic complaints, those with chronic fatigue tend to denounce psychiatric diagnoses as inadequate explanations for their ills.

Postpsychotic Depressive Disorder of Schizophrenia. Postpsychotic depression can occur during the residual phase of schizophrenia. It includes the persistence of negative or positive symptoms in an attenuated form (e.g., odd beliefs, unusual perceptual experiences). The depressive symptoms include loss of interest or pleasure and a sad or depressed mood. The major depressive episode typically follows the active phase of the psychotic episode. Sometimes it may follow after a short or extended interval during which there are no psychotic symptoms. Mood symptoms due to the direct physiologic effects of a drug of abuse, a medication, or a general medical condition would suggest a different cause for the depressive symptoms.

In all postpsychotic depressions, one must first exclude a missed bipolar diagnosis. Negative symptoms due to classic antipsychotics—especially depot phenothiazines and those due to the residuum of schizophrenia once positive symptoms are under control—should be distinguished from the depressive episodes that complicate the course of schizophrenia in young, intelligent patients. This phenomenon is so common (at least 30% of patients with schizophrenia) that we consider it a part of the natural course of schizophrenia rather than a separate nosologic entity.

BIPOLAR DISORDERS

There are several types of bipolar disorder categories included in DSM-5. We categorize some on a descriptive level, such as bipolar I disorder, bipolar II disorder, cyclothymic disorder, and unspecified bipolar disorder. We classify others according to etiology, such as substance- or medication-induced bipolar disorder and bipolar disorder due to another medical condition. The clinical picture for each is similar regardless of etiology.

Tables 13.3–5 and 13.3–6 compare the DSM-5 and ICD-11 approaches to diagnosing bipolar disorders.

Bipolar I Disorder

Typically beginning in the teenage years, the 20s, or the 30s, the first episode of bipolar I disorder could be manic, depressive, or mixed. One typical onset is mild retarded depression or hypersomnia for a few weeks or months, switching into a manic episode. Others begin with a severely psychotic manic episode with schizophreniform features; only when a more classic manic episode occurs is the affective nature of the disorder clarified. In a third group, several depressive episodes occur before the first manic episode. A careful history from significant others often reveals hyperthymic or cyclothymic traits that antedated the frank onset of manic episodes by several years, if not longer.

A single manic episode describes patients having a first episode of mania. Most such patients eventually develop depressive episodes. We use the remaining subcategorization to specify the nature of the current or most recent episode in patients who have had recurrent mood episodes. For clinicians and researchers alike, it is more meaningful to chart a patient's course in color over time—for example, using red for manic, blue for depressive, and violet for mixed episodes, with hypomanic, dysthymic, and cyclothymic periods drawn in the appropriate colors on a smaller scale between the major episodes. Arrows can indicate life events, biologic stressors, and treatment on the time axis. Initially championed by Kraepelin, this approach is routinely used in mood or bipolar clinics to chart retrospectively the past course of illness in mood disorders, especially bipolar disorders. Robert Post at the NIMH has developed this approach into systematic clinical science. An international conference in Barcelona recommended adopting this approach for clinical use. Data from the NIMH Collaborative Depression Study—using a prospective method of life charting—of bipolar I disorder clinical cohorts followed in five major U.S. medical centers show more depression, especially its subthreshold variants, during an average of 12 years of follow-up. These patients were symptomatic nearly 50% of the weeks through the entire follow-up, suggesting a more pervasive course of illness than previously appreciated. Although such data indicate some degree of tertiary care bias in favor of patients willing to be retained in a long-term study, they underscore a clinical reality faced by many clinicians in their daily practice (see Fig. 13.3–5).

On average, manic episodes predominate in youth, and depressive episodes predominate in later years. Although the overall sex ratio is approximately 1:1, men, on average, undergo more manic episodes, and women experience more mixed and depressive episodes. Bipolar I disorder in children is not as rare as previously thought; however, most reported cases are in boys, and mixed manic (dysphoric–explosive) and rapid-cycling presentations are typical. Childhood-onset depression must also be considered a significant risk for ultimate bipolar transformation. This is based on the following characteristics: (1) early age at onset; (2) even sex ratio; (3) prominence of irritability, labile moods, and explosive anger, suggesting mixed episodes; (4) questionable response to antidepressants, hypomanic switches, or both; (5) high recurrence rate after depression; and (6) familial affective loading. Mania can also first appear after 65 years of age, although a diligent search often reveals a past mild, forgotten, or untreated depressive or excited episode in earlier years.

Acute Mania. Mania typically escalates over 1 to 2 weeks; more sudden onsets can also occur. The DSM-5 criteria stipulate (1) a distinct period that represents a break from premorbid functioning, (2) at least 1 week in length, (3) an elevated or irritable mood, (4) at least three or more classic manic signs and symptoms, and (5) the absence of any physical factors that could account for the clinical picture. The irritable mood in mania can deteriorate to cantankerous behavior, especially when rebuffing the person. Such patients are among the most aggressive seen in the emergency room. Florid grandiose psychosis with paranoid features, a typical presentation of mania, further contributes to the aggression. Alcohol use, observed in at least 50% of bipolar I patients (often during the manic phase), further disinhibits the patient and might lead to a dangerous frenzy. Such patients may attack loved ones and hurt them physically. So-called crimes of passion have been committed by patients harboring delusions of infidelity on the part of spouses or lovers, usually when the patient is under the influence of alcohol.

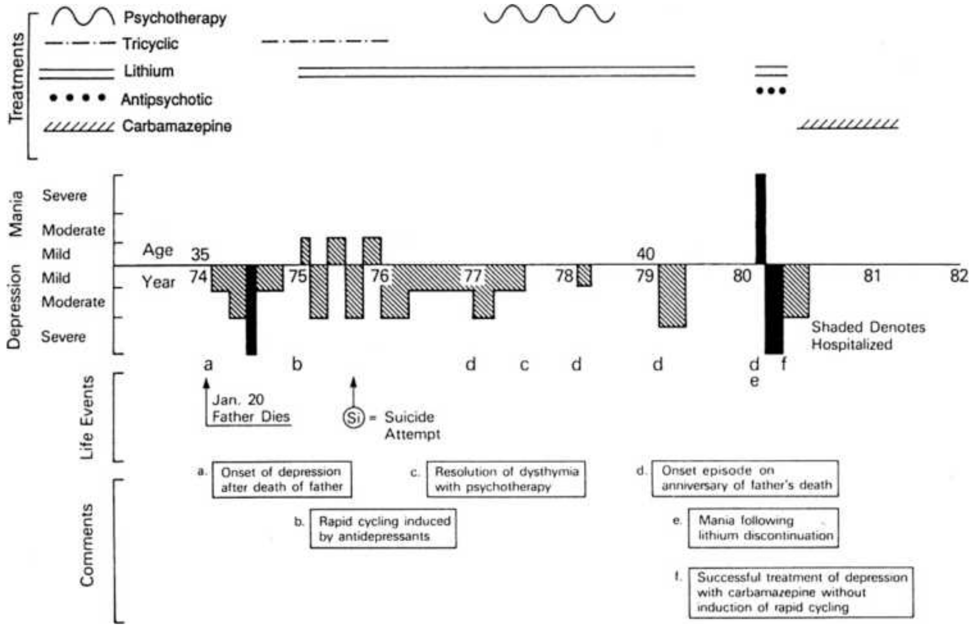

FIGURE 13.3–5. Graphing the course of a mood disorder. Prototype of a life chart. (Courtesy of Robert M. Post, M.D.)

Delusional, hallucinatory, even first-rank, psychotic experiences in mania can also occur. Recent research has documented that most formal thought disorders are common to schizophrenic and mood psychoses; only poverty of speech content (vagueness) emerges as significantly more common in schizophrenia. In addition, posturing and negativism may occur during manic periods. Finally, even cognitive deficits resembling dementia can occur. Mania is a phase of bipolar I disorder, which has strong genetic determinants. The evidence does not permit separating recurrent mania without depressive episodes from that type as a distinct nosologic entity.

Secondary Mania. Although there is some suggestion that postpartum mania without depression is distinct from familial bipolar I disorder (in which depressive, manic, and especially mixed manic episodes occur in the postpartum period), the evidence for a distinct puerperal mania is not compelling. Mania without prior bipolarity can arise in the setting of such somatic illnesses as thyrotoxicosis, systemic lupus erythematosus or its treatment with steroids, rheumatic chorea, multiple sclerosis (MS), Huntington disease, cerebrovascular disorder, diencephalic and third ventricular tumors, head trauma, complex partial seizures, syphilis, and (most recently) AIDS. The family history is reportedly low in such cases, suggesting a relatively low genetic predisposition and, thus, a lower risk of recurrence. These patients do not easily fit into the DSM-5 category of mood disorder due to a general medical condition because most of the conditions appear to be cerebral. Such factors may be more typical in manias of late life.

Less well-defined forms of mania are the so-called reactive manias. Personal loss and bereavement may be triggering factors, and the reaction is conceptualized psychodynamically as a denial of loss. Although such explanations may be plausible in individual cases, no systematic data suggest that these patients differ in family history from persons with other manias. The same is generally valid for depressed patients who switch to hypomania or mania after abusing stimulant drugs, taking antidepressants, or sleep deprivation. A bipolar diathesis usually manifests in a family history of mania or spontaneous excited episodes during prospective observation. First-onset manic episodes can also occur in persons who abruptly abstain from alcohol after one or more decades of chronic use and then develop classic bipolar I disorder.

Chronic Mania. About 5% of bipolar I disorder patients have a chronic manic course. These cases commonly represent a deteriorating course dominated by recurrent manic episodes grafted onto a hyperthymic baseline. Noncompliance with pharmacologic treatment is the rule. Recurrent excitement is personally reinforcing, subjective distress is minimal, and insight is seriously impaired. Thus, the patient sees no reason to adhere to treatment. Episodic or chronic alcohol abuse, prevalent in such patients, may contribute to the chronicity. Some authorities further consider comorbid cerebral pathology responsible for nonrecovery (and increased mortality) from manic excitements occurring in late life.

Grandiose delusions (e.g., delusions of inventive genius or aristocratic birth) are not uncommon in chronic mania and may lead to the mistaken diagnosis of paranoid schizophrenia. Because of their social deterioration, Kraepelin subsumed such patients under the category *manic dementia.* Organic factors, such as head trauma and chronic alcohol abuse, may contribute to the deterioration. Nonschizoid premorbid adjustment, a family history of bipolar I disorder, and the absence of flagrant formal thought disorder may help establish the affective basis of these poor-prognosis manic states.

Bipolar Mixed Phase. Momentary tearfulness and even depressed mood are commonly observed at the height of mania or during the transition from mania to retarded depression. These transient labile periods, which occur in most bipolar I disorder patients, must be contrasted with mixed episodes.

The latter, variously referred to as *mixed mania* or *dysphoric mania,* are characterized by dysphorically excited moods. These moods may include symptoms of irritability, anger, panic attacks, pressured speech, agitation, suicidal ideation, severe insomnia, grandiosity, hypersexuality, or persecutory delusions and confusion. It is challenging to distinguish severely psychotic mixed states that involve hallucinations and Schneiderian symptoms from schizoaffective disorder. A correct diagnosis helps avoid conventional antipsychotic drugs known to exacerbate the depressive component; in such patients, failure to use mood stabilizers can prolong the patient’s misery.

Research data from mood centers worldwide on mixed mania suggest that dysphoric mania—mania and full-blown depression co-occurring—is relatively uncommon. Two to four depressive symptoms from the list of depressed mood, helplessness, hopelessness, fatigue, anhedonia, guilt, and suicidal ideation, impulses, or both in the setting of a manic syndrome appear to suffice for the diagnosis of mixed manic states, which occur in 50%

of patients with bipolar disorder sometime during their lives. Mixed states occur predominantly in women with mania superimposed on a depressive temperament or a dysthymic baseline. The emerging conceptualization of mixed mania is a manic state intruding on long-term depressive traits. Other mixed manic states, especially in men, often arise from the interaction of substance or alcohol use in bipolar I disorder.

Depressive Phase. Psychomotor retardation, with or without hypersomnia, marks the uncomplicated depressive phase of bipolar I disorder. Onset and offset are often abrupt, although onset can also occur gradually over several weeks. Patients may recover into a free interval or may switch directly into mania. Switching into an excited phase is particularly likely when using antidepressants. However, not all patients develop mania after antidepressant treatment of bipolar depression. Some develop a mixed agitated depression; indeed, patients may linger for many months in a severe depressive phase with some manic admixtures, such as racing thoughts and sexual arousal. Delusional and hallucinatory experiences are less common in the depressive phase of bipolar I disorder than in the manic and mixed manic phases. Stupor is the more common psychotic presentation of bipolar depression, particularly in adolescents and young adults. Pseudodemented organic presentations appear to be the counterpart of stupor in older adults.

Cyclothymic Disorder

An attenuated bipolar disorder that typically begins insidiously before 21 years of age, cyclothymic disorder is characterized by frequent short cycles of subsyndromal depression and hypomania. Table 13.3–7 compares the diagnostic approaches to cyclothymic disorder in DSM-5 and ICD-11. The course of cyclothymia is continuous or intermittent, with infrequent periods of euthymia. Shifts in mood often lack adequate precipitants (e.g., sudden profound dejection with social withdrawal for a few days switching into cheerful, gregarious behavior). Circadian factors may account for some of the extremes of emotional lability, such as the person’s going to sleep in good spirits and waking up early with suicidal urges. The mood changes of cyclothymia are best described as *endoreactive* in the sense that endogenous overreactivity seems to determine the sudden shifts in mood and behavior (e.g., falling in love with a person one has just met and as quickly falling out of love).

Mood swings in these ambulatory patients are overshadowed by the chaos the swings wreak in their personal lives. Repeated romantic breakups or marital failures are common because of interpersonal friction and episodic promiscuous behavior. Uneven performance at school and work are also typical. Persons with cyclothymic disorder are dilettantes; they show great promise in many areas but rarely bring their efforts to fruition. As a result, their lives are often a string of improvident activities. Geographical instability is a characteristic feature; easily attracted to a new locale, job, or love partner, they soon lose interest and leave in dissatisfaction. Polysubstance abuse, which occurs in as many as 50% of such persons, is often an attempt at self-treatment. Suicidality has emerged as a public health hazard in adolescents. Clinicians may mistake this for borderline personality disorder, especially in young people, when they may suffer from an undiagnosed temperamental expression of bipolar disorder.

Bipolar II Disorder

Research conducted during the last three decades has shown that, between the extremes of manic-depressive illness and major depressive disorder, there exists an overlapping group of intermediary forms characterized by recurrent major depressive episodes and hypomania. Current data worldwide indicate that bipolar II disorder is more prevalent than bipolar I disorder. This prevalence certainly appears to be true in the outpatient setting, in which an average of 50% (and in one French national study up to 65%) of persons with major depressive disorder conform to the bipolar II disorder pattern. The depressive phase of bipolar disorder, particularly bipolar II, is a significant public health problem.

The following self-description provided by a 34-year-old poet illustrates the pattern:

> *I have known melancholy periods, lasting months at a time when I would be literally paralyzed: All mental activity comes to a screeching halt, and I cannot even utter one word. I become so dysfunctional that I was once hospitalized. Although the paralysis creeps into me insidiously—often lasting months—it typically reverses within hours. I am suddenly alive and vibrant. I cannot turn off my brain neither during the day nor at night; I usually go on celebrating like this for many days, needing no more than few hours of slumber each day.*

This vignette is nearly identical to the autobiographical description provided by the British poet William Cowper three centuries earlier:

> *I have known many a lifeless and unhallowed hour . . . long intervals of darkness interrupted by short returns of peace and joy . . . for many succeeding weeks to rejoice day and night was all my employment. Too happy to sleep much, I thought it was lost time that was spent on slumber.*

Although some hypomanias last for weeks, the hypomania at the end of depressive episodes in most bipolar II disorder patients does not last long; usually, a few days. As discussed earlier, hypomanias with short duration (less than 4 days) are familially similar to those with longer duration (more than 4 days). Another common form of bipolar II disorder is major depressive disorder superimposed on cyclothymic disorder, in which hypomania precedes and follows major depressive disorder, with the entire interepisodic period characterized by cyclothymic mood instability. Such patients tend to have an unstable course, stormy interpersonal relationships, and more irritable and hostile hypomanic episodes—hence the confusion with borderline personality disorder. Current data indicate that cyclothymic and noncyclothymic bipolar II disorder have similar rates of familial bipolarity, which suggests that the borderline characterologic features represent a complication of the unstable course that cyclothymia imparts to bipolar II disorder. Such individuals with unstable course have a high risk of suicide—indeed, based on research in Budapest; it appears that the high burden of suicide in mood disorders appears to be due to bipolar II disorder. Descriptively, *hypomania in bipolar II disorder* includes mini manic episodes occurring spontaneously. Bipolar II disorder—especially with superimposed major depressions and cyclothymia—is best characterized as cyclical or *cyclothymic depression.*

The depressive episodes of patients with bipolar disorder often have admixtures (e.g., flight of ideas and increased drives and impulsivity in sexual and other domains). These are likely depressive mixed states. DSM-5 is the first to describe such states. Depressive mixed states are not as severe as dysphoric mixed states but are refractory to antidepressants nonetheless and, according to current data, are overrepresented in suicide. Research conducted by Franco Benazzi and Hagop Akiskal indicates that such depressive mixed states occur in as much as 60% of cases of bipolar II disorder and as much as 30% of cases of “unipolar” major depressive disorder. Both groups have similar rates of familial bipolarity, suggesting that one of three patients with “unipolar” depression belongs to the bipolar II spectrum.

Hypomania. The common denominator of the soft spectrum of bipolar disorders is hypomania. *Hypomania* refers to a distinct period of at least a few days of mild elevation of mood, sharpened and positive thinking, and increased energy and activity levels, typically without the impairment characteristic of manic episodes. It is not merely a milder form of mania. Hypomania occurring as part of bipolar II disorder rarely progresses to manic psychosis, and insight is relatively preserved. Hypomania is distinguished from mere happiness in that it tends to recur (happiness does not!), and antidepressants can sometimes mobilize hypomania. In cyclothymic disorder, it alternates with mini depressions; in hyperthymic temperament, it constitutes the person’s habitual baseline. These definitions, developed by empirical research, recognize three patterns of hypomania: *episodes* heralding the termination of a retarded depressive state (bipolar II disorder), cyclic alternation with mini depressions (cyclothymic disorder), and an elevated baseline of high mood, activity, and cognition (hyperthymic or chronic hypomanic *traits*).

Because people experience hypomania as a rebound relief from depression or as a pleasant, relatively short-lived, ego-syntonic mood state, persons with bipolar II disorder rarely spontaneously report it or even see it as a problem. Skillful questioning is required to diagnose soft bipolar conditions; as in mania, collateral information from family members is crucial. During a patient interview, we typically ask questions like, “Have you had a distinct sustained high period when your mood was so intense that you felt nervous?” Alternatively, we might ask, “Have you had such energy that others could not keep up with you, or your thinking and perceptions were unusually vivid or rapid?” Many bipolar patients react to inquiries about “high moods” or “high periods” with a negative: “I am not a manic-depressive!” Benazzi and Akiskal suggested that it is better to inquire about signs and symptoms referring to behavioral activation initially. Once remembering energized periods, the patient is more likely to recall the mood change associated with these periods.

Clinical and epidemiologic studies in the United States and Europe have revealed a rich range of hypomanic manifestations, including increased cheerfulness and jocularity; gregariousness and people seeking; more interest in sex; talkativeness, self-confidence, and optimism; and decreased inhibitions and sleep need. The clinician must ascertain that those experiences are not due to stimulant use or alcohol withdrawal. Depressive and hypomanic periods are often not easily discerned because chronic caffeine use, stimulant abuse, or both may complicate the depression. In such instances, we should defer a diagnosis for 1 month after detoxification. Cocaine and stimulant use is also not uncommon in hypomania, especially in patients with hyperthymic traits, and constitutes an attempt to augment the baseline energy and mood rather than self-treatment of low moods. When in doubt, direct clinical observation of hypomania—sometimes elicited by antidepressant pharmacotherapy—provides definitive evidence for the bipolar nature of the disorder.

Seasonal Patterns. Seasonality occurs in many cyclic depressions, often with autumn or winter anergic depression and energetic periods in the spring. This natural propensity explains why phototherapy may provoke mild hypomanic switches. Seasonal depressions conform, in considerable measure, to the bipolar II disorder or bipolar III disorder pattern. Furthermore, preliminary evidence suggests that treatment with classic antidepressants disrupts the baseline seasonality, with the depressive phase appearing in the spring and summer. The changes that antidepressants induce in seasonal depressions probably represent a particular variant of the rapid-cycling phenomenon.

Temperament and Polarity of Episodes. Systematic clinical observations suggest that bipolar II disorder (characterized predominantly by depressive attacks) arises more often from a hyperthymic or cyclothymic baseline. Bipolar I disorder (defined by manic attacks) not uncommonly arises from the substrate of a depressive temperament. In line with these observations, a prospective 11-year NIMH study of major depressive disorder patients who switched to bipolar II disorder showed that *mood–labile* (cyclothymic) and *energetic–active* (hyperthymic) temperament traits were highly specific (86%) and reasonably sensitive (42%) predictors of such an outcome.

We usually think of *bipolarity* as alternating manic (or hypomanic) and depressive episodes. The previous data on temperaments suggest that a more fundamental characteristic of bipolarity is the reversal of temperament into its “opposite” episode (in the bipolar II disorder spectrum, from cyclothymia and hyperthymia to major depression). Such findings suggest that the intrusion of cyclothymic and hyperthymic traits into a depressive episode may underlie the instability of the bipolar II disorder subtype and could partly explain why bipolar II depression often has mixed features. These considerations may have important implications for preventing recurrence. For instance, a prospective study of the onset of bipolar disorder in the offspring or siblings of adults with the disorder found that children with depressive onsets as their first episode (typically treated with antidepressants) had significantly higher rates of recurrence than those with manic or mixed onsets (treated with lithium) during a 3-year prospective observation. It appears that temperamental instability in the depressive group might have predisposed them to the cycling effect of antidepressants.

Alcohol and Substance Abuse. New evidence supports the high prevalence of alcohol and substance abuse in mood disorder subtypes, especially those with interepisodic cyclothymia and hyperthymia. The relation appears to be particularly strong in the teenage and early adult years when taking such substances may represent self-medication for mood instability. It is not just self-treatment of selected symptoms associated with the down or up phases (e.g., alcohol to alleviate insomnia and nervousness characteristic of both phases); it also augments certain desired ends (e.g., stimulants to enhance performance or sexual behavior associated with hypomania). Alcohol may be sought both for its disinhibiting effects and sedation to calm the nervousness of the high periods. How many patients display alcohol and substance abuse secondary to an underlying bipolar diathesis remains to be determined. However, given findings suggesting a link between polysubstance abuse and suicide in adolescents with bipolar familial backgrounds, we should strongly consider using mood stabilizers in these adolescents. Although alcohol and stimulant use continues into adult years in many bipolar disorder patients, such use is often unrelated to familial alcoholism and frequently tends to dwindle during long-term follow-up. This finding supports the self-medication hypothesis.

Rapid-Cycling Bipolar Disorder. *Rapid cycling* is the occurrence of at least four episodes of depression and hypomania (or mania) per year. Thus, rapid cyclers are rarely free of affective symptoms and experience severe vocational and interpersonal incapacitation. Lithium is only modestly helpful to those patients, as are traditional antipsychotic agents; most antidepressants readily induce excited episodes and aggravate the rapid-cycling pattern. Balancing mood stabilizers, atypical antipsychotic drugs, and antidepressants is challenging. We should avoid antidepressants because depression follows every provoked hypomania, and the roller coaster is further prolonged. Many such patients require frequent hospitalization because they develop explosive excitement and precipitous descent into severe psychomotor inhibition. The disorder is a roller coaster nightmare for the patient, significant others, and the treating physician. Treating these patients is an art. Lamotrigine, a mood stabilizer that stabilizes “from below,” has been reported to be efficacious in this group of patients.

As expected, rapid cycling commonly arises from a cyclothymic substrate, which means that most rapid cyclers have bipolar II disorder. Factors favoring its occurrence include (1) female gender, (2) borderline hypothyroidism, (3) menopause, (4) temporal lobe dysrhythmias, (5) alcohol, minor tranquilizer, stimulant, or caffeine abuse, and (6) long-term, aggressive use of antidepressant medications. Most clinically identified patients are bipolar II women in middle age or upper socioeconomic status. Rapid cycling is uncommon from a bipolar I disorder base.

Leadership and Creativity. Persons with hyperthymic temperament and soft bipolar conditions, in general, possess assets that permit them to assume leadership roles in business, the professions, civic life, and politics. Increased energy, sharp thinking, self-confidence, and eloquence represent the virtues of an otherwise stormy life.

Creative achievement is relatively uncommon among those with the manic forms of the disorder, which are too severe and disorganizing to permit the necessary concentration and application. We do see some notable artistic achievements among those with soft bipolar disorders, especially cyclothymic disorders. Psychosis, including severe bipolar swings, is generally incompatible with creativity. That conclusion tends to refute the romantic tendency to idolize insanity as central to the creative process. Because talent is the necessary ingredient of creativity, how might soft bipolarity contribute? The simplest hypothesis is that depression might provide insights into the human condition, and the activation associated with hypomania helps produce the artistic work. A more profound interpretation suggests that the repeated self-doubt accompanying recurrent depression might be a significant ingredient of creativity, because an original artistic or scientific expression is often initially rejected. The self-confidence that accompanies repeated bouts of hypomania can help rehearse such ideas or expressions. Finally, the tempestuous object relations associated with bipolarity in the parent’s or the patient’s life often create unique biographical landmarks that the artist might then immortalize in their art.

Ultrarapid Cycling. In these patients, hypomanic or manic and depressive episodes alternate rapidly over a few days. In many, the alternation occurs regularly. The episodes in such cycling patients have greater amplitudes than in cyclothymia, leading to marked impairment. We might consider them as an intermediary between cyclothymia and full-blown rapid cycling. Of considerable theoretical and methodologic interest, they are challenging to manage clinically. Some individuals with this condition learn to change their work days to coincide with the hypomanic periods.

Recurrent Brief Hypomania. This hypomanic counterpart of recurrent brief depressive disorder is a variant of soft bipolarity, as described previously. It tends to coincide with cyclothymic depression, in which hypomanias are short (less than 4 days).

Hysteroid Dysphoria. Hysteroid dysphoria combines reverse vegetative signs with the following characteristics: (1) giddy responses to romantic opportunities and an avalanche of dysphoria (angry-depressive, even suicidal, responses) to romantic disappointment; (2) impaired anticipatory pleasure, yet the capability to respond with pleasure when others provide it (i.e., preservation of consummatory reward); and (3) craving for chocolate and sweets, which contain phenylethylamine compounds and sugars believed to facilitate cellular and neuronal intake of the amino acid L-tryptophan, hypothetically leading to the synthesis of endogenous antidepressants in the brain. Historically, clinicians used the term *hysteroid* to convey that the apparent character pathology was secondary to a biologic disturbance in the substrates governing affect, drives, and reward. The intense, unstable life of the patient with hysteroid dysphoria suggests links to cyclothymic disorder or bipolar II disorder. Researchers such as those at Columbia tend to subsume those patients under atypical depressions (some of which, as indicated, have bipolar affinities). Like patients with bipolar depression, they may respond preferentially to monoamine oxidase inhibitors (MAOIs). In brief, hysteroid dysphoria appears to be a variant of bipolar II disorder with cyclothymic-irritable traits. Clinicians often relegate such patients to the realm of borderline personality. The Pisa–San Diego collaboration (for which Dr. Akiskal was a member) suggested that cyclothymic traits represent the underlying diathesis of atypical depression, borderline personality, and bipolar II disorders.

UNSPECIFIED MOOD DISORDERS

After obtaining all diagnostic information, a clinician may find that some patients with affective symptoms do not meet the specific criteria for the mood conditions described thus far, nor do they fit the examples listed in the not otherwise specified categories for depressive and bipolar disorders. One example is individuals with acute suicidal ideation of an ego-dystonic nature, not accompanied by other affective symptoms. Family or close follow-up, or both, often suggests a more discrete mental condition. In general, we should consider such cases as undiagnosed mood disorders. However, there are two hybrid conditions that we have not addressed elsewhere in this chapter and can conveniently discuss here.

Mixed Anxiety–Depressive Disorder

Both anxiety and depression can occur together when confronted with a significant aversive life situation. The admixture implies that the psychopathology progresses from anxiety to depression, that the patient's mental state is still in flux, and that the ongoing dynamics partly explain the disorder's subacute or chronic nature. Anxious depression points to the common presence of anxiety in depressive states, especially its greater visibility when the depression is less prominent. Patients with the latter presentation are reportedly most prevalent in general medical settings. This prevalence should not come as a surprise, because depressive symptoms that motivate medical consultation tend to complicate generalized anxiety states with a subthreshold level of symptomatology. Some authorities argue that neurotic depressions arise as maladaptive responses to anxiety and, on that basis, suggest retaining the *neurotic depression* rubric. Some genetic data indirectly support the contention that certain (unipolar) depressive and (generalized) anxiety states are related. As currently defined, anxious depressions are heterogeneous. In patients refractory to anxiolytic or antidepressant treatment, practitioners must entertain the diagnosis of a complex bipolar II disorder with mixed features. Indeed, recent familial-genetic investigations suggest that bipolar II disorder with panic attacks might represent a particular form of bipolar disorder.

Atypical Depression

Initially developed in England and under investigation at Columbia University in New York City, *atypical depression* refers to fatigue superimposed on a history of somatic anxiety and phobias, together with reverse vegetative signs (mood worse in the evening, insomnia, tendency to oversleep and to overeat). Experience suggests that increased sexual interest or desire is another reverse vegetative sign, although it remains undescribed in this literature. Sleep is disturbed in the first half of the night in many persons with atypical depressive disorder, and so irritability, hypersomnolence, and daytime fatigue would be expected. Sensitive traits characterize the temperaments of these patients. The MAOIs and serotonergic antidepressants seem to show some specificity for such patients, which is why we take atypical depression seriously.

Other research reveals several classifications of reverse vegetative signs: (1) the anxious type just described or (2) a subtle bipolar subtype with protracted hyperphagic–hypersomnic–retarded dysthymic disorder with occasional brief, extroverted, hypomanic-type behavior, often elicited by antidepressants. Increasing evidence indicates considerable affinity between atypical depression and bipolar II and III disorders. Furthermore, many patients with dysthymic disorder exhibit atypical features at various times. Some research suggests that many patients with atypical depression meet the criteria for brief hypomania or cyclothymic disorder. Such patients also resemble those with hysteroid dysphoria, the precursor of the construct of atypical depression.

DIFFERENTIAL DIAGNOSIS

A missed mood disorder diagnosis means that the disorder does not receive specific treatment, which has serious consequences. Many such persons drop out of school or college, lose their jobs, or get divorced, and some commit suicide. Those with unexplained somatic symptoms are frequent users of the general health system. Others are unwell, despite interminable psychotherapy. Some, treated with conventional neuroleptics, develop tardive dyskinesia unnecessarily. As with other medical disorders for which specific treatments are available, accurate diagnosis and early treatment are within the purview of all physicians and mental health professionals. Psychiatrists, in particular, should develop the competence to detect the entire spectrum of mood disorders. Despite massive educational efforts, underdiagnosis and undertreatment of mood disorders remain serious problems worldwide. The most pernicious of these is downplaying the mood component because a clinician gives a personality disorder undue prominence.

Although much enthusiasm was generated three decades ago about the potential use of specific biologic markers (e.g., REM latency, DST, and thyrotropin-releasing hormone [TRH] test) to corroborate the differentiation of mood disorder from adjacent disorders, no definitive progress justifies their routine use in clinical practice. The same is true, at least for now, for brain imaging, neuropsychologic testing, and genetic screening. Given that the presentations can be unusual or confusing, we are best off applying a systematic clinical approach. Such an approach helps (1) to detail all clinical features of the current episode, (2) to elicit a history of more typical major mood episodes in the past, (3) to assess whether the presenting complaints recur periodically or cyclically, (4) to substantiate adequate social functioning between periods of illness, (5) to obtain a positive family history for a classic mood disorder and to construct a family pedigree, and (6) to document a history of unequivocal therapeutic response to thymoleptic medication or ECT in the patient or the family.

Normal Bereavement

Bereaved persons exhibit many depressive symptoms during the first 1 to 2 years after their loss, so how can we identify the 5% of bereaved persons who have progressed to a depressive disorder? Table 13.3–9 outlines some features that suggest that bereavement has become a “full-blown” depressive disorder.

In another form of bereavement depression, the patient simply pines away, unable to live without the departed person, usually a spouse. Although not necessarily pathologic by the preceding criteria, such persons do have a severe medical condition. Their immune function is often depressed, and their cardiovascular status is precarious. Death can ensue within a few months of that of a spouse, especially among older men. Such considerations (highlighted in the work of Sidney Zisook and his colleagues at the University of California at San Diego) suggest that it would be clinically unwise to withhold antidepressants from many persons experiencing an intensely mournful form of grief. A case vignette illustrates this point:

Table 13.3–9.
Signs that a Bereaved Person Has Progressed to a Depressive Disorder

Beliefs. Grieving persons and their relatives perceive bereavement as a normal reaction whereas those with depressive disorder often view themselves as sick and may actually believe they are losing their minds.

Emotional reactivity. Unlike the melancholic person, the grieving person reacts to the environment and tends to show a range of positive effects.

Psychomotor activity. Marked psychomotor retardation is not observed in normal grief.

Guilt. Although bereaved persons often feel guilty about not having done certain things that they believe might have saved the life of the deceased loved one (guilt of omission), they typically do not experience guilt of commission.

Delusions. Delusions of worthlessness or sin and psychotic experiences in general point toward mood disorder.

Suicidal ideation. Active suicidal ideation is rare in grief but common in major depressive disorder.

Behaviors. “Mummification” (i.e., keeping the belongings of the deceased person exactly as they were before their death) indicates serious psychopathology.

Anniversaries. Severe anniversary reactions should alert the clinician to the possibility of psychopathology.

From Akiskal. Mood disorders. In: Sadock BJ, Sadock VA, Ruiz P, eds. *Kaplan & Sadock’s Comprehensive Textbook of Psychiatry*. 10th ed. Wolters Kluwer; 2017.

A 75-year-old widow was brought to the treatment by her daughter because of severe insomnia and total loss of interest in daily routines after her husband’s death 1 year before. She had been agitated for the first 2 to 3 months and, after that, “sank into total inactivity—not wanting to get out of bed, not wanting to do anything, not wanting to go out.” According to her daughter, she was married at 21 years of age, had four children, and had been a homemaker until her husband’s death from a heart attack. Past psychiatric history was unrevealing; compulsive traits had characterized premorbid adjustment. She was dressed in black during the interview, appeared moderately slowed, and sobbed intermittently, saying, “I search everywhere for him … I don’t find him.” When asked about life, she said, “Everything I see is black.” Although she expressed no interest in food, she did not seem to have lost an appreciable amount of weight. The patient declined psychiatric care, stating that she “preferred to join her husband rather than get well.” She was too religious to commit suicide, but by refusing treatment, she felt that she would “pine away … find relief in death and reunion.”

Anxiety Disorders

Anxiety symptoms, including panic attacks, morbid fears, and obsessions, are common during depressive disorders, and depression is a common complication of anxiety states. Systematic British studies have shown that early-morning awakening, psychomotor retardation, self-reproach, hopelessness, and suicidal ideation are the most robust clinical markers of depression in that differential diagnosis. On follow-up of depressed patients, the manifestations tend to remit, whereas those with anxiety states continue to exhibit marked tension, phobias, panic attacks, vasomotor instability, feelings of unreality, and perceptual distortions, as well as hypochondriacal ideas. A predominance of such anxiety features antedating the present illness suggests the diagnosis of an anxiety disorder. Because anxiety disorders rarely first appear after 40 years of age, a late appearance of marked anxiety features strongly favors the diagnosis of melancholia. The clinical picture is often one of morbid groundless anxiety with somatization, hypochondriasis, and agitation. Superior response to ECT further supports the depressive nature of the illness. Periodic monosymptomatic phobic and obsessional states exist which we can regard as affective equivalents based on a family history of mood disorders and their response to thymoleptic agents. Recent data from a large clinical series suggest that 15% of cases have hypomanic symptoms; these patients likely have bipolar II disorder. Social phobias often usher in adolescent depression, even a bipolar II disorder; the latter is particularly likely when pairing the phobia with alcohol abuse.

We have not resolved the psychopathologic differentiation of anxiety and depressive states. Cognitive factors may differentiate them best (Table 13.3–10). Although recurrent (particularly retarded) major depressive disorder is a distinct disorder from anxiety states, some forms of depression may share a common diathesis with anxiety disorders, particularly generalized anxiety disorders. Before assigning patients to such putative mixed anxiety–depressive group (not yet an official nosologic entity), the clinician must note that it is best to consider anxiety arising primarily during depressive episodes as epiphenomenal to depressive disorder. The same is generally true for anxiety symptoms in a patient with a depressive disorder who uses alcohol, sedative–hypnotics, or stimulant drugs. Finally, anxiety symptoms could be prominent features of mixed bipolar states and complex partial seizures.

Table 13.3–10.
Unique Cross-Sectional Profiles of Clinical Anxiety and Depression

Anxiety	Depression
Hypervigilance	Psychomotor retardation
Severe tension and panic	Severe sadness
Perceived danger	Perceived loss
Phobic avoidance	Loss of interest—anhedonia
Doubt and uncertainty	Hopelessness—suicidal
Insecurity	Self-deprecation
Performance anxiety	Loss of libido
	Early-morning awakening
	Weight loss

Reprinted with permission from Akiskal HS. Toward a clinical understanding of the relationship of anxiety and depressive disorders. In: Maser JP, Cloninger CR, eds. *Comorbidity of Mood and Anxiety Disorders*. American Psychiatric Press; 1990.

Major Depressive Disorder versus Bipolar Disorder

This “internal” boundary question of whether a patient has major depressive disorder versus bipolar disorder has been discussed extensively throughout this work. With the increasing availability of mood-stabilizing agents, it has emerged as a significant challenge in clinical practice. Numerous studies have shown that bipolar disorder is confused not only with personality, substance use, and schizophrenic disorders but also with depressive and anxiety disorders. Table 13.3–11 summarizes the features predictive of bipolar disorder based on results of prospective and other systematic studies.

Table 13.3–11.
Features of a Depressive Episode that Are More Predictive of Bipolar Disorder

Early age at onset
Psychotic depression before 25 years of age
Postpartum depression, especially one with psychotic features
Rapid onset and offset of depressive episodes of short duration (less than 3 months)
Recurrent depression (more than five episodes)
Depression with marked psychomotor retardation
Atypical features (reverse vegetative signs)
Seasonality
Bipolar family history
High-density three-generation pedigrees
Trait mood lability (cyclothymia)
Hyperthymic temperament
Hypomania associated with antidepressants
Repeated (at least three times) loss of efficacy of antidepressants after initial response
Depressive mixed state (with psychomotor excitement, irritable hostility, racing thoughts, and sexual arousal during major depression)

From Akiskal. Mood disorders. In: Sadock BJ, Sadock VA, Ruiz P, eds. *Kaplan & Sadock’s Comprehensive Textbook of Psychiatry*. 10th ed. Wolters Kluwer; 2017.

More broad indicators of bipolarity include the following conditions, none of which, by itself, clinches a bipolar diagnosis but should raise clinical suspicion in that direction: agitated depression, cyclical depression, episodic sleep dysregulation, or a combination of these; refractory depression (failed antidepressants from three different classes); depression in someone with an extroverted profession; periodic impulsivity, such as gambling, sexual misconduct, and wanderlust, or periodic irritability, suicidal crises, or both; and depression with erratic personality disorders.

Personality Disorders

Most personality measures are state dependent. In general, clinicians should refrain from using personality disorder labels to describe patients with active affective illness and focus instead on the competent treatment of the mood disorder. Unfortunately, many clinicians ignore this advice. Even in patients with chronic or subthreshold mood disorders, we should consider personality maladjustment as postaffective, arising from the distortions and conflicts that affective disturbances produce in the patient's life. The most problematic of the personality labels used in patients with mood disorders is borderline personality disorder, usually applied to teenage and young adult women. Table 13.3–12 shows that the overlap between borderline personality and mood disorders is extensive so that giving a *borderline* diagnosis to a person with a mood disorder may be redundant. The use of personality disorder diagnoses may lead to a neglect of the mood disorder or, perhaps, half-hearted treatment of the mood disorder; we then blame the patient's nonresponse on their "self-defeating character" or "resistance to getting well," thus exculpating ourselves.

Although we need more systematic research on the complex interface of personality and mood disorders, clinically, they are often inseparable. As with alcohol and substance use disorders, it is generally preferable to diagnose mood disorders at the expense of personality disorders, which should not be difficult to justify in most cases that satisfy the validating strategies outlined previously. When personality and mood disorders coexist, it is good practice to defer personality disorder diagnoses and treat the mood disorder. Although not all personality disturbances recede with the competent treatment of mood disorders, so many experienced clinicians have seen such disturbances melt away with the successful resolution of the mood disorder that erring in favor of mood disorders is justified.

Table 13.3–12.
Overlap of Borderline Personality Disorder and Mood Disorders

Familial
High rates of mood disorder
Phenomenology
Dythymic disorder
Cyclothymic disorder
Bipolar II disorder
Mixed state
Pharmacologic response
Worsening on most antidepressants
Stabilization on anticonvulsants
Prospective course
Major mood episodes
Suicide

Data from Akiskal HS, Chen SE, Davis GC, Puzantian VR, Kashgarian M. Borderline: an adjective in search of a noun. *Clin Psychiatry.* 1985;46(2):41.

A 19-year-old single woman presented with the chief complaint that “all men are bastards.” Since her early teens, with the onset of her menses, she had complained of extreme variability in her moods on a nearly daily basis; irritability with hostile outbursts was her primary affect, although more-protracted hypersomnic depressions with multiple overdoses and wrist slashings had led to at least three hospitalizations. She also had migrainous headaches that, according to her mother, had motivated at least one of those overdoses. Despite her tempestuous and suicidal moods that led to these hospitalizations, she complained of “inner emptiness and a bottomless void.” She had used heroin, alcohol, and stimulants to overcome this troubling symptom. She also gave a history of ice cream craving and frequent purging. She was talented in English and wrote much-acclaimed papers on the American confessional poet Anne Sexton. She said she was mentally disturbed by a series of stepfathers who had forced “oral rape” on her between 11 and 15. She subsequently gave herself sexually to any man she met in bars, no longer knowing whether she was a “prostitute” or a “nice little girl.” On two occasions, she had inflicted cigarette burns inside her vagina “to feel something.” She had also engaged in a “brief lesbian relationship” that ultimately left her “emptier” and guilt ridden; nonetheless, she now believed that she should burn in hell because she could not get rid of “obsessing” about the excitement of mutual cunnilingus with her much older female partner. The patient’s mother, who owned an art gallery, had been married five times and gave a history of unmistakable hypomanic episodes; a maternal uncle had died from alcohol-induced cirrhosis. The patient’s father, a prominent lawyer, known for his “temper and wit,” had committed suicide. Her psychiatrist gave the patient phenelzine, at which point, the mother described her as “the sweet daughter she was before age 13.” At her next premenstrual phase, the patient developed insomnia, ran away from home at night, started “dancing like a go-go girl, met an incredibly handsome man” of 45 years of age (a pornography shop owner), and had a clandestine marriage to him. After many dose adjustments, she currently takes a combination of lithium and valproic acid. The patient now attends college and has completed four semesters in art history. In addition to control of her irritable and suicidal moods, bulimic and migraine attacks have abated considerably. She was able to annul her marriage, contending that she was not mentally competent at the time of the wedding. She is no longer promiscuous and now expresses fear of intimacy with men to whom she is attracted. She is receiving individual psychotherapy for this problem.

This case illustrates the intimate relationship among atypical depression, borderline personality, and bipolar II disorder. These three conditions may share an underlying psychobiologic or genetic diathesis. Clinicians often complain that even when diagnosing a mood disorder in a “borderline” patient, the antidepressant response is disappointing. The problem is that affective disorders in these patients usually conform to bipolar II disorder—often complicated by ultrarapid cycling—and many clinicians trained in an earlier era, including some with biologic orientation, may lack sufficient experience in the art of pharmacologically managing patients who markedly deviate from classic bipolar I disorder.

Alcohol and Substance Use Disorders

We cannot explain the high comorbidity of alcohol and substance use disorders with mood disorders as merely the chance occurrence of two prevalent disorders. Psychiatrists and other clinicians dealing with addiction do not sufficiently appreciate self-medication for mood disorders. Given the clinical dangers of missing an otherwise treatable disorder, we should seriously consider a mood disorder as the primary diagnosis if marked affective manifestations persist or escalate after detoxification (e.g., 1 month). This consideration also pertains to cyclothymic disorder and dysthymic disorder, which appear particularly likely to invite self-medication. The clinical validating strategies listed previously can further buttress a mood disorder diagnosis.

The DSM category of substance-induced mood disorder is challenging to validate clinically because, in the absence of an affective diathesis, detoxification should, in principle, rapidly repair affective disturbances in persons whose primary problem is that of substance abuse. A dual diagnosis of a mood disorder and a substance use disorder is a more realistic clinical approach to this group of patients. Bipolarity, particularly bipolar II disorder, should be sought in the interface of mood and substance use disorders.

A 27-year-old married businessman employed in an international family venture owned by his father presented with a court-ordered request for psychiatric treatment. He had been found “bringing” cocaine across the United States–Mexican border and was briefly jailed. He had used stimulants since his late teens to enhance his already high level of energy. His family was wealthy, and he had no difficulty affording cocaine. During the previous year, he had needed more cocaine because of more remarkable moodiness and fleeting suicidal ideation, which he linked to increasing tensions between him and his father: “My father was never satisfied with me and demanded greater and greater performance from me.” His arrest by police was a major embarrassment for him and his family and motivated his compliance with psychiatric hospitalization to detoxify him. He had not had cocaine for 10 days, exhibited marked lability of mood, and gradually sank into a severe hypersomnic-retarded depression of stuporous proportions. He was hospitalized on an inpatient psychiatric unit and treated with tranylcypromine 20 mg twice a day, and, within 10 days, he switched into hypomania, his mind “exploding with creativity and confidence,” marked jocularity and witticisms that entertained other patients, and marked seductiveness toward the nurses. His wife recalled that the patient previously had several such periods naturally (i.e., “off cocaine”), which had strained their marriage due to “brief sexual liaisons.” Reducing the tranylcypromine dose by 50% did not eliminate the hypomanic behavior, and lithium, 900 mg per day, was added. He has since been maintained on a combination of tranylcypromine and lithium for 4 years; he has not relapsed into cocaine use, and, following a few psychoeducational sessions involving father and spouse, relationships with family and spouse have been less tempestuous. (Since then, consultation was sought by the patient’s 60-year-old mother, an attractive, sophisticated woman who confessed that for years she had been engaged in “love relationships” with young artists, with, apparently, her husband’s “tacit consent;” since at least her mid-20s, she, by history, would meet the criteria for bipolar II disorder, only was treated “on the couch,” and her sister and brother had received treatment of “alcohol excesses.”) The patient stated that pharmacotherapy, which did require adjustment now and then, helped in balancing the “rough edges” of his “high-nervous temperament” and his “periodic lapse into paralyzing fatigue states that occurred at stressful times.”

If his doctors had assumed his mood states were merely due to cocaine withdrawal, they would never have treated his bipolar II disorder. There is emerging interest in treating dually diagnosed patients with anticonvulsant mood stabilizers and their judicious combination. Among the antidepressants, bupropion is a suitable agent to “replace” stimulant craving; serotonin reuptake inhibitors help concurrent social phobia and anxious components. The intention is to treat any substance-related withdrawal while also treating any underlying or emerging bipolar diathesis.

Physical Disease

Somatic complaints are common in depressive disorders. Some, such as vegetative disturbances, represent the hypothalamic pathology believed to underlie a depressive disorder. Autonomic arousal, commonly associated with depression, could explain such symptoms as palpitations, sweating, and headache. In some instances, the physical symptoms might reflect delusional experiences. The clinician must be vigilant about the likelihood that somatic complaints in depression can also reflect an underlying physical illness. Table 13.3–13 lists the most common medical conditions associated with depression. When depressive symptoms occur in the setting of physical illness, it is not always easy to determine whether they constitute a genuine depressive disorder. Before diagnosing depression, psychiatrists must ensure that they are not dealing with *pseudodepression*. By “pseudodepression,” we mean (1) functional loss due to physical illness; (2) vegetative signs, such as anorexia and weight loss, as manifestations of such an illness; (3) stress and demoralization secondary to the hospitalization; (4) pain and discomfort associated with the physical illness; and (5) medication adverse effects. However, nonpsychiatric physicians who manage such patients must consider the diagnosis of depression in the presence of persistent anhedonia; observed depressed mood with frequent crying; observed psychomotor retardation or agitation; indecisiveness; convictions of failure, worthlessness, or guilt; and suicidal ideation. The physician should also suspect clinical depression in all patients who refuse to participate in medical care.

Table 13.3–13.
Pharmacologic Factors and Physical Diseases Associated with Onset of Depression

- Pharmacologic
  - Steroidal contraceptives
  - Reserpine (Serpalan), á-methyldopa
  - Anticholinesterase insecticides
  - Amphetamine or cocaine withdrawal
  - Alcohol or sedative–hypnotic withdrawal
  - Cimetidine (Tagamet), indomethacin (Indocin)
  - Phenothiazine antipsychotic drugs
  - Thallium, mercury
  - Cycloserine (Seromycin)
  - Vincristine (Oncovin), vinblastine (Velban)
  - Interferon
- Endocrine–metabolic[a]
  - Hypothyroidism and hyperthyroidism
  - Hyperparathyroidism
  - Hypopituitarism
  - Addison disease
  - Cushing syndrome
  - Diabetes mellitus
- Infectious
  - General paresis (tertiary syphilis)
  - Toxoplasmosis
  - Influenza, viral pneumonia
  - Viral hepatitis
  - Infectious mononucleosis
  - Acquired immune deficiency syndrome
- Collagen
  - Rheumatoid arthritis
  - Lupus erythematosus
- Nutritional
  - Pellagra
  - Pernicious anemia
- Neurologic
  - Multiple sclerosis
  - Parkinson disease
  - Head trauma
  - Complex partial seizures
  - Sleep apnea
  - Cerebral tumors
  - Cerebrovascular infarction (and disease)
- Neoplastic
  - Abdominal malignancies
  - Disseminated carcinomatosis

[a] Cholesterol is not mentioned because low levels as a factor in depression have been inconsistently reported.

Diagnosing depression in medically ill older adult patients can be particularly difficult. We should diligently undertake this task, given the data showing that depression can accelerate mortality in the medically ill (especially in those with cardiovascular disease). Depressed older adults often deny being depressed but complain of anxiety, fatigue, and worsening memory. Hypochondriacal symptoms and pain are common. Patients may exhibit extreme negativism and querulousness when invited to participate in medical procedures; others develop inadequate fluid and food intake out of proportion to their physical conditions.

Another critical diagnostic problem at the interface of mood disorder and physical disease is the rare development of malignancy in patients with an established mood disorder. Patients who had responded well to a given antidepressant during previous episodes now have an unsatisfactory response to the same medication. Even a small dose may cause such alarming symptoms as agitation, dizziness, depersonalization, and illusions, which might indicate an occult malignancy, perhaps in the abdomen or the brain. The psychiatrist should always be vigilant for possible life-threatening physical diseases developing in patients with pre-established depressive disorder.

A 55-year-old woman had four previous episodes of severe depression that had responded to ECT, amitriptyline, or a combination in her native city in the Middle East. She immigrated to the United States at 43 years of age and encountered several significant stresses, including an unfamiliarity with English, her daughter dating a man of Chinese extraction, and a complicated series of operations for uterine prolapse. Then, her husband confessed that he had had an affair with a much younger woman. For months, the patient had complained of intermittent fatigue and expressed hostility toward her husband. He had slapped her on her face on two occasions. Her ensuing fifth depressive episode (with pain localized to her face) appeared entirely "understandable," but her physician's prescription of 25 mg of amitriptyline resulted in dizzy spells that culminated in syncope. Citalopram, 10 mg, did not fare any better. An extensive medical workup revealed a retroperitoneal lymphoma. She died 6 months later.

Stupor. Although less common today, stupor still raises a diagnostic problem differentiating between a mood disorder and somatic disease, and other psychiatric disorders. Depressive stupor is relatively easy to distinguish from so-called hysterical mutism; in the latter, the patient directs the behavior to significant others in the patient's environment. It is best to reserve the rubric of catatonic stupor for a phase of schizophrenia; in such patients, the schizophrenic origin of the catatonia might be apparent from the patient's history. Otherwise, most acute-onset stupors are probably of affective origin. The primary differential diagnosis here is from organic stupor (due to drugs or acute intracranial events); the physical and neurologic examination is not always decisive in such cases. Diagnosis depends on a high index of suspicion of possible somatic factors.

Dementia. We can usually distinguish the cognitive symptoms caused by depression (sometimes called "pseudodementia," although this can be a misleading term) by their acute onset without prior cognitive disturbance. Also, a personal or family history of past affective episodes, psychomotor retardation with reduced social interaction, self-reproach, and diurnal cognitive dysfunction (worse in the morning), can help with the differential. Often described are the subjective complaints of memory dysfunction, which appear much worse than what we find on objective testing. In some cases, we can reverse circumscribed memory deficits with proper encouragement or coaching. Sometimes the deficits seem to improve with sleep deprivation. The deficits do not improve with depression treatment for some patients, which may signal the early stages of a dementing disease. Such patients may be labile, aggressive, and disinhibited and may become easily confused. A family history of bipolar disorder or a history of mild manic-like expressions helps distinguish these patients from those whose cognitive deficits are due to depression. The distinction is essential because antidepressants and anti-Alzheimer agents often worsen patients with an underlying bipolar disorder, who would instead benefit from a mood stabilizer. Prospective follow-up reveals the definitive diagnosis. The main clinical message here is not to miss treatable or potentially reversible depressive and bipolar states in older adult patients with dementia-like illnesses or early- and even mid-stage dementia.

Chronic Fatigue Syndrome. Chronic fatigue syndrome is a complex differential diagnostic problem, given the subtle immunologic disturbances presumably associated with it. The following self-report by such a patient illustrates many uncertainties marking the present knowledge of the interface between the syndrome and mood disorders.

I am a 39-year-old, never-married woman, trained as a social worker but currently on disability. I have experienced extreme lethargy and fatigue for many years. I have always felt foggy headed and had trouble thinking and concentrating. My complaint is of fatigue, not of depression. My body feels like lead and aches all over. My brain feels achy and sore. I feel much worse in the morning, and I cannot get out of bed; I feel better at night. I feel bad every day. I ache all over as though someone had beaten me up. Exercise has been prescribed to me, but it makes me worse. Also, I am very sensitive to hot and cold. My sexual drive is low. I have a general feeling of anhedonia. As far back as I remember—in junior high school—I was always exhausted. I always complained about fatigue, not depression, because that has been the overwhelming problem. I feel the depression is secondary to the fatigue. In high school, I was a compulsive overeater, and I was bulimic for a few years, but it was never severe, and I was only about 10 pounds overweight. In those days, I would sleep 10 or 12 hours a night on the weekend and still feel exhausted; I could not get up for school on Monday. As an adolescent, I felt inferior. I could not make decisions. I did not want to go to camp or leave home for long periods of time—I felt so insecure. Recently, I had a sleep study done, which showed a short latency to stage REM sleep (49 minutes). I was diagnosed as having dysthymic disorder and began taking antidepressants. When I took tranylcypromine, it was the first time in my life that I felt like a normal person. I could play sports, had a sex drive, had energy, and was able to think clearly, but the benefits lasted for barely 2 months. My response was equally short lived to phenelzine, imipramine, selegiline, and bupropion. I have not responded to SSRIs at all. I also wish to point out that I had never experienced high periods before I took antidepressants. My main problem has always been one of exhaustion. When I responded to medications, they worked very quickly (within a few days), and I felt great, but they all stopped working after a short time. The dose would be raised, and, again, I would feel better. Eventually, when I got to high doses, I either could not tolerate the high dose, or the drug would no longer help. I have taken different combinations of drugs for 10 years, and I have not been able to feel well for more than 6 weeks at a time. Recently, I went to an immunologist. He said I have an abnormality in regulating antibody production and recommended gamma-globulin shots. They did not help. When I first started working, I always felt tired and foggy headed, so it was difficult to be sharp while at work. At times, I would close the door to my office and put my head down. Working has become increasingly difficult for me. I had two great jobs, which I blew. As of last year, I had to go on disability. I am desperate for relief, as my condition has drastically affected my life. Disability has been hard for me. I am single and have no other financial resources. I am very despondent, as I feel that my life is passing by without the hope of my ever really improving.

This clinical picture is compatible with a pseudounipolar or bipolar III disorder as described previously. If so, mood stabilizer augmentation of the tranylcypromine would have been a therapeutic choice to pursue. Some biologists and immunologists, and some psychiatrists, believe that abnormal substances circulate in the bloodstream, supplying the patients' brains. Some have suggested industrial toxins as a factor. Although awaiting more definitive research on the etiology of chronic fatigue syndrome, the psychiatrist can cautiously consider certain patients for thymoleptic trials. The following supports such a trial: The patient wakes up with fatigue and dread of facing the day; fatigue is part of a more generalized psychomotor inertia or lack of initiative; fatigue is associated with anhedonia, including sexual anhedonia; and fatigue coexists with anxious and pessimistic ruminations. Although none of the previous considerations alone is pathognomonic for depression, they point in that direction in aggregate. The occurrence of hypomanic-like periods (as in the previous vignette) further supports the link between some cases of chronic fatigue and mood disorder. Several recent neuroendocrine challenge studies suggest that some chronic fatigue patients might have a strong anxiety substrate. These observations do not suggest that chronic fatigue is mostly a matter of missed affective diagnoses with bipolar II and panic disorder features; yet, it would be a pity to miss potentially treatable diagnoses. A family or past personal history of classic affective illness or episodes should strongly weigh in this direction. Definitive data are lacking on the essential nature of chronic fatigue, and practitioners should be guided by their clinical experience with therapeutic trials of nonsedating thymoleptic agents while they await new research developments.

Schizophrenia

Despite some genetic overlap between schizophrenic and bipolar disorders, clinically, the two groups of disorders are in the main separate. Young patients with bipolar disorder might seem psychotic and disorganized and, thus, schizophrenic. Their thought processes are so rapid that they may seem loose, but, unlike patients with schizophrenia, they display an expansive and elated affect, which is often contagious. By contrast, the severely retarded bipolar depressive person, whose affect may superficially seem flat, rarely exhibits significant fragmentation of thought. The clinician, therefore, should place greater emphasis on the pattern of symptoms than on individual symptoms in the differential diagnosis of mood and schizophrenic psychoses. No pathognomonic differentiating signs and symptoms exist. Differential diagnosis should consider the overall clinical picture, phenomenology, family history, course, and associated features. Because the two groups of disorders entail different combinations of pharmacologic treatments on a long-term basis (such as mood stabilizers for bipolar disorder), the differential diagnosis is of paramount clinical importance. In the past, many bipolar patients, especially those with prominent manic features at onset, were labeled as having *acute schizophrenia* or *schizoaffective schizophrenia.* In the past, such misdiagnoses (which typically led to long-term treatment with conventional neuroleptic agents) were costly in terms of tardive dyskinesia, vocational and social decline, and even suicide. For instance, some patients with postpsychotic depressive disorder of schizophrenia have postmanic depressions but are treated with classic neuroleptic monotherapy without the benefit of more definitive thymoleptic agents.

Modern antipsychotics tend to keep many persons with schizophrenia out of the hospital. However, they do not always seem to prevent an overall downhill course. By contrast, the intermorbid periods in bipolar illness are relatively normal or even supernormal, and yet, over time, some social impairment may result from the accumulation of divorces, financial catastrophes, and ruined careers. (Although rapid-cycling disorders, which have sharply risen during the past two decades, cause considerable social impairment, mood symptoms are so prominent that differentiation from schizophrenia is generally not complicated; in addition, such patients usually display more classic bipolar phases before the rapid cycling.)

Postpsychotic depressions in persons with established schizophrenia are sometimes due to inadequate control of schizophrenic symptomology. In other patients, particularly more intelligent young patients with schizophrenia, they reflect the experience of losing one’s ego and sanity. It would be more meaningful to give such patients a diagnosis of schizophrenia and a superimposed major depressive diagnosis and treat the patient accordingly.

Schizoaffective Disorder

As the previously mentioned considerations suggest, depression in the setting of a schizophrenic disorder does not necessarily constitute a distinct nosologic entity. We should restrict the concept of schizoaffective (or cycloid) psychosis to recurrent psychoses, with full affective and schizophrenic symptoms occurring nearly simultaneously during each episode. We should not consider this diagnosis for mood psychosis in which mood-incongruent psychotic features (e.g., Schneiderian and Bleulerian symptoms) are better explained by one of the following: (1) affective psychosis superimposed on mental retardation, giving rise to extraordinarily hyperactive and bizarre manic behavior; (2) affective psychosis complicated by concurrent brain disease, substance abuse, or substance withdrawal, known to give rise to numerous Schneiderian symptoms; or (3) mixed episodes of bipolar disorder (which are notorious for signs and symptoms of psychotic disorganization). Official diagnostic systems, such as DSM-5, use the category of schizoaffective disorder broadly. Thus, patients with clear-cut manic episodes receive a schizoaffective diagnosis if delusions or hallucinations occur in the interepisodic period in the absence of prominent affective symptoms. Many psychotic symptoms in mood disorders are often explanatory (albeit delusional), whereby the patient tries to make sense of the core experiences of the affective illness. In patients with recurrent episodes, delusional thinking can carry over into the interepisodic period. Such patients are thus delusional in the absence of prominent mood symptoms and technically might be considered schizoaffective.

However, affective illness is typically a lifelong process, and limiting its features to discrete episodes is artificial. In the past, neuroleptic agents were prescribed on an as-needed basis to reduce the affective charge of those interepisodic delusions, but they did not effectively eliminate the affect-laden experiences. Atypical antipsychotic agents introduced during the 1990s are more likely to be beneficial in this regard. Augmenting with continued thymoleptic treatment (resorting to ECT, if necessary) and an empathic psychotherapeutic approach may prove more rewarding in the long run.

A 29-year-old female college graduate, a mother of two children, and wife of a bank president had experienced several manic and retarded depressive episodes that had responded to lithium carbonate. She was referred to me because she had developed the delusion that she was involved in an international plot. Careful probing revealed that the delusion represented further elaboration, in a rather fantastic fashion, of a grandiose delusion that she had experienced during her last postpartum manic episode. She believed that she had played an essential role in uncovering the plot, thereby becoming a national hero. Nobody knew about it, she contended, as the circumstances of the plot were top secret. She further believed that she had saved her country from the international scheme and suspected that she was singled out for persecution by the perpetrators of the plot. At one point, she had even entertained the idea that the plotters sent unusual radio communications to intercept and interrupt her thoughts. As is typical in such cases, she was on a heavy dose of a lithium–antipsychotic combination. The consultation was requested because the primary mood symptoms were under control, and yet she had not given up her grandiose delusion. She flippantly remarked that “I must be crazy to believe in my involvement in an international plot,” but she could not help but believe in it. Over several months, seen typically in 60-minute sessions weekly, the patient had developed sufficient trust that her psychiatrist could gently challenge her beliefs.

She was, in effect, told that her self-professed role in the international scheme was highly implausible and that someone with her superior education and high social standing could not entertain a belief, to use her own words, “as crazy as that.” She eventually broke into tears. She said that everyone in her family was so accomplished that she had to be involved in something grand to keep up with them. In effect, the international scheme was her only claim to fame. She said, “Nobody ever gives me credit for raising two kids and throwing parties for my husband’s business colleagues: My mother is a dean, my older brother holds high political office; my sister is a medical researcher with five discoveries to her credit [all true] and who am I? Nothing. Now, do you understand why I need to be a national hero?” As she alternated, over subsequent months, between such momentary flashes of insight and delusional denial, she was able to discontinue her antipsychotic medication gradually. Maintained on lithium, she now only makes passing reference to the grand scheme. She was encouraged to pursue her career goal toward a master’s degree in library science.

The vignette illustrates how we can fruitfully combine phenomenologic understanding, rational pharmacotherapy, and practical psychotherapeutic or vocational guidance in our approach to patients with psychotic mood disorders. It also indicates that problems with insight are not necessarily state dependent in mood disorders. Cognitive disturbances have recently been described in all mood disorders, even when they are nonpsychotic. At a more fundamental level, the vignette suggests that clinical diagnoses cannot be based entirely on operational criteria; the treatment response does not always change one’s opinion of a patient’s illness.

ACKNOWLEDGMENT

This chapter is a revision of a previous chapter written by Hagop S. Akiskal, M.D. Dr. Akiskal, a giant in this field, passed away in January of 2021.

FURTHER READINGS

Akiskal HS, Akiskal KK. TEMPS: temperament evaluation of Memphis, Pisa, Paris and San Diego. *J Affect Disord*. 2005;85(1–2):1–2.

Akiskal HS, Azorin JM, Hantouche EG. Proposed multidimensional structure of mania: beyond the euphoric-dysphoric dichotomy. *J Affect Disord*. 2003;73:7–18.

Akiskal HS. Bipolar depression: focus on phenomenology. *J Affect Disord*. 2005;84(2–3):105–106.

Akiskal HS, Benazzi F. Family history validation of the bipolar nature of depressive mixed states. *J Affect Disord*. 2003;73:113–122.

Akiskal HS, Benazzi F, Perugi G, Rihmer Z. Agitated "unipolar" depression re-conceptualized as a depressive mixed state: implications for the antidepressant-suicide controversy. *J Affect Disord*. 2005;85:245–258.

Akiskal HS, Bourgeois ML, Angst J, Post R, Möller HJ, Hirschfeld R. Re-evaluating the prevalence of and diagnostic composition within the broad clinical spectrum of bipolar disorders. *J Affect Disord*. 2000;59(Suppl 1):S5–S30.

Akiskal HS, Cassano GB. *Dysthymia and the Spectrum of Chronic Depressions*. Guilford Press; 1997.

Akiskal HS, Maser JD, Zeller PJ, et al. Switching from 'unipolar' to bipolar II: an 11-year prospective study of clinical and temperamental predictors in 559 patients. *Arch Gen Psychiatry*. 1995;52:114–123.

Alhajji L, Nemeroff CB. Personalized medicine and mood disorders. *Psychiatr Clin North Am*. 2015;38(3):395–403.

Angst J, Ajdacic-Gross V, Rössler W. Classification of mood disorders. *Psychiatr Pol*. 2015;49(4):663–671.

Badcock PB, Davey CG, Whittle S, Allen NB, Friston KJ. The depressed brain: an evolutionary systems theory. *Trends Cogn Sci*. 2017;21(3):182–194.

Balázs J, Benazzi F, Rihmer Z, Rihmer A, Akiskal KK, Akiskal HS. The close link between suicide attempts and mixed (bipolar) depression: implications for suicide prevention. *J Affect Disord*. 2006;91:133–138.

Coplan JD, Aaronson CJ, Panthangi V, Kim Y. Treating comorbid anxiety and depression: psychosocial and pharmacological approaches. *World J Psychiatry*. 2015;5(4):366–378.

Davidson JR, Miller RD, Turnbull CD, Sullivan JL. Atypical depression. *Arch Gen Psychiatry*. 1982;39:527–534.

Judd LL, Akiskal HS, Maser JD, et al. A prospective 12-year study of subsyndromal and syndromic depressive symptoms in unipolar major depressive disorders. *Arch Gen Psychiatry*. 1998;55:694–700.

Krause JS, Reed KS, McArdle JJ. Factor structure and predictive validity of somatic and nonsomatic symptoms from the patient health questionnaire-9: a longitudinal study after spinal cord injury. *Arch Phys Med Rehabil*. 2010;91(8):1218–1224.

Kuswanto C, Chin R, Sum MY, et al. Shared and divergent neurocognitive impairments in adult patients with schizophrenia and bipolar disorder: whither the evidence? *Neurosci Biobehav Rev*. 2016;61:66–89.

Liu X, Jiang K. Should major depressive disorder with mixed features be classified as a bipolar disorder? *Shanghai Arch Psychiatry*. 2014;26(5):294–296.

Malhi GS, Byrow Y, Fritz K, et al. Mood disorders: neurocognitive models. *Bipolar Disord*. 2015;17(Suppl 2):3–20.

Mordeno IG, Carpio JGE, Mendoza NB, Hall BJ. The latent structure of major depressive symptoms and its relationship with somatic disorder symptoms among Filipino female domestic workers in China. *Psychiatry Res.* 2018;70:587–594.

Perini G, Ramusino MC, Sinforiani E, Bernini S, Petrachi R, Costa A. Cognitive impairment in depression: recent advances and novel treatments. *Neuropsychiatr Dis Treat*. 2019;15:1249–1258.

Post RM, Altshuler LL, Kupka R, et al. More illness in offspring of bipolar patients from the U.S. compared to Europe. *J Affect Disord*. 2016;191:180–186.

Rumble ME, White KH, Benca RM. Sleep disturbances in mood disorders. *Psychiatr Clin North Am*. 2015;38(4):743–759.

Shim IH, Myong Bahk WM, Woo YS, Yoon BH. Pharmacological treatment of major depressive episodes with mixed features: a systematic review. *Clin Psychopharmacol Neurosci.* 2018;16(4):376–382.

Singh MK, Garrett AS, Chang KD. Using neuroimaging to evaluate and guide pharmacological and psychotherapeutic treatments for mood disorders in children. *CNS Spectr*. 2015;20(4):359–368.

Sotsky SM, Glass DR, Shea MT, et al. Patient predictors of response to psychotherapy and pharmacotherapy: findings in the NIMH Treatment of Depression Collaborative Research Program. *Am J Psychiatry*. 1991;148 (8):997–1008.

Torres F, Barbosa M, Maciel P. Recurrent copy number variations as risk factors for neurodevelopmental disorders: critical overview and analysis of clinical implications. *J Med Genet*. 2016;53(2):73–90.

Winokur G, Turvey C, Akiskal H, et al. Alcoholism and drug abuse in three groups—bipolar I, unipolars and their acquaintances. *J Affect Disord*. 1998;50:81–89.

Yarlagadda A, Acharya G, Kasaraneni J, Hampe CS, Clayton AH. Placental barrier and autism spectrum disorders: the role of prolactin and dopamine on the developing fetal brain. *Innov Clin Neurosci*. 2015;12(9–10):14–17.

Young AS, Fristad MA. Family-based interventions for childhood mood disorders. *Child Adolesc Psychiatr Clin N Am*. 2015;24(3):517–534.

Zaninotto L, Solmi M, Toffanin T, Veronese N, Cloninger CR, Correll CU. A meta-analysis of temperament and character dimensions in patients with mood disorders: comparison to healthy controls and unaffected siblings. *J Affect Disord*. 2016;194:84–97.

13.4 Mood Disorders: Intrapsychic and Interpersonal Aspects

John C. Markowitz, M.D., and Barbara L. Milrod, M.D.

The symptom criteria listed in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) reliably define mood disorders, but they do not completely describe these syndromes or explain how patients experience them. Depressed patients suffer emotionally, cognitively, and physically. Three key theoretical psychotherapeutic approaches have been applied to mood disorders: psychoanalytic/psychodynamic, cognitive, and interpersonal. They vary in their frequency of use in clinical practice and in the amount of research that they have received.

Theory has uses and limitations. Theory can help to organize clinical and research thinking about depression and can pinpoint aspects of the syndrome. A theory can provide a narrative thread or focus on anchoring a clinical formulation and giving coherence to a therapy. Indeed, some theoretical approach is necessary for clinicians to impose order on the overwhelming amount of clinical data with which patients present for treatment. Theories may also allow us to make predictions about treatment mechanisms and outcomes. We submit that an organizing theoretical framework is as essential to structure psychotherapy as grammar is to speech. Hence understanding the theoretical backgrounds of psychotherapies is crucial for the psychotherapeutic clinician.

Theories should be considered approximations of truth, needful of testing, and subject to disconfirmation. Thomas Huxley, in 1870 famously described “The great tragedy of Science—the slaying of a beautiful hypothesis by an ugly fact.” Dogma and ideology have done psychiatrists and other mental health professionals more harm than good, for example, in separating psychotherapy and pharmacotherapy into opposing camps—a rift from the 1950s that is only now healing. Inasmuch as the etiology of mood disorders is complex and no single theory comprehensively explains it, rigid adherence to any theory is unwarranted and potentially problematic. Clinicians should understand and use multiple theories, flexibly choosing the optimal theoretical perspective to apply to a given patient's illness and then persevering in using that model for an appropriate course of treatment. They should then proceed following that single, coherent theoretical approach rather than devolving into eclecticism. (The authors of this chapter recommend formal training in any treatment modality the clinician plans to undertake. Theoretical understanding does not suffice.) On the other hand, to develop clinical expertise in any therapy, it is essential to be able to understand patients' problems through the lens of its underlying theory.

Psychodynamic theory has the longest historical tradition. Both cognitive theory and psychodynamic theory largely focus on intrapsychic phenomena, whereas interpersonal theory, the newest and least formally developed, focuses more on interpersonal, extrapsychic reality. It is widely accepted that aspects of both cognitive and interpersonal theories derive from psychoanalysis. Each approach provides a potentially useful explanation of the plight of the depressed patient. Although the utility of therapies derived from cognitive theory (cognitive behavioral therapy [CBT]) and interpersonal theory (interpersonal psychotherapy [IPT]) has been better tested for depression than psychodynamic therapy to date (albeit psychodynamic therapies have been found efficacious for depression), no theories of depression have been formally tested in all of their assumptions. They are best considered working clinical models.

PSYCHODYNAMIC ASPECTS OF MOOD DISORDERS

A contemporary psychoanalytic understanding of mood disorders includes a comprehensive focus on biologic underpinnings, cognitive function, and interpersonal situation and style. What is unique to psychoanalysis is its attention to intrapsychic, unconscious pressures in its consideration of psychological symptoms, including mood disorders. In order to describe intrapsychic aspects of mood disorders, several basic psychoanalytic ideas about mental life first must be defined:

1. From a psychodynamic view, mental life exists on two levels: Both within the realm of consciousness and also within a less accessible realm described as "*the unconscious*." Psychic or emotional symptoms arise from aspects of mental life that are at least in part unconscious. This is true of mood disorders.

2. Psychoanalysts have found it useful to conceptualize the mind as comprising three basic theoretical psychic structures, *the id*, *the ego*, and *the superego*. In brief, *the id* is considered the aspect of the mind that subsumes the drive derivatives and desires. *The ego* serves as an intermediary between the id and the external world. It contains many intrapsychic functions, including motor action, perception, self-esteem, relationship to reality, and the ability to modulate the drives, and to modulate anxiety. Defense mechanisms, which are unconscious intrapsychic mechanisms employed to modulate both the drives and anxiety, are ego functions. *The superego* is conceptualized to comprise the person's value system. The superego can both punish and reward the person, depending upon whether or not their actions are consonant with their moral values.
3. Moods are pervasive ego states that color the entire ego with the same affect state. Unlike simple affective responses to events, they are not focused but general, either because the affect is too strong or because the ego is too weak to contain a focused response. Because moods tend to be generalized states, they typically involve some degree of denial of the opposite feelings. From a psychoanalytic vantage, moods carry an unconscious significance, notwithstanding their presumed biologic and neurotransmitter underpinnings.
4. Since Sigmund Freud's discovery of the importance of the unconscious in everyday mental life and of individuals' capacity to shut off unwanted, painful, emotionally laden experiences using defense mechanisms, our understanding of moods and symptoms has been enhanced by the recognition that these psychic phenomena represent a breakthrough of *unconscious fantasy* into consciousness. Persistent unconscious fantasies often underlie people's psychological symptoms, dreams, personalities, and important life choices. They can underlie depression as well.
5. Freud initially described another central principle in the organization of mental life: That people unconsciously avoid "unpleasure" and that ideas that produce unpleasure are screened from consciousness by "repression," or processes that we now call *defenses*. Clinically, the degree to which unpleasure is avoided varies from person to person, and Freud later modified his original theory about people avoiding unpleasure. In *Beyond the Pleasure Principle*, he attempted to describe intrapsychic mechanisms that could account for phenomena in which the pursuit of pleasure is given up for the more fundamental need to discharge intense emotional tensions to protect the individual.

> Thus, major depression is a condition in which the person is living beyond the pleasure principle.

PSYCHOANALYTIC DESCRIPTIONS OF MAJOR DEPRESSION

Psychoanalytic formulations highlight the phenomenology of aspects of major depression, helping to make the emotional backdrop of the illness more understandable. Although these theories provide a clinically rich set of ideas that can be useful in treating such patients, it should be emphasized that no theoretical positions in any analytic school are supported by systematic research. The field of psychoanalysis has been fairly resistant to research. Nor are various psychoanalytic formulations, despite their differences, mutually exclusive. The following sections outline some of these formulations and provide illustrative clinical examples.

Response to Loss/Anger Turned Inward

The classic psychoanalytic understandings of depression were stated by Karl Abraham, Freud, and Sandor Rado and emphasized depressed patients' reactions to object loss, in reality or in fantasy. In these formulations, the profound response to loss is believed to occur in part because the current loss invokes an earlier childhood loss, also either of a fantasy or a reality nature. These authors noted depressed patients' ambivalent or hostile object relations, along with object attachments characterized by excessive dependency, laced with an emphasis on need for gratification in emotionally charged relationships. Major depression only occurs after the tie to the object is shattered. In *Mourning and Melancholia*, Freud highlighted the way in which depressed patients irrationally attack themselves. In his formulation, this occurs because aspects of an ambivalent object become internalized or incorporated into the patient's sense of self, and hostility directed toward the object is instead directed at the self. This state of affairs serves to preserve the relationship with the other person (the object) in reality.

Stuart Asch described a variation on these dynamics in which some patients become depressed not because of object loss but due to maintaining a submissive tie to the object. This tie is maintained because the patient perceives separation to be an aggressive act with destructive consequences. Rage at having to play this part is directed toward the self, as the patient feels devalued because of the submission.

Joseph Sandler and Walter Joffe, also focusing on the phenomenon of loss leading to depression, called the Hampstead Index, a comprehensive clinical registry of childhood responses to abandonment and loss, for cases of childhood depression. They inferred that depression is a basic affective response to loss. They emphasized that more than just the other person is lost and that the child feels that they have also lost a sense of self in relationship to the lost object or a previous set of feelings about the self. Thus, as Freud and Abraham also noted, Sandler and Joffe emphasize a symbiotic or narcissistic tie to the object. They saw individuals predisposed to depression as struggling against feelings of helplessness and injured self-esteem in childhood.

Ms. A, a 26-year-old woman, presented with severe major depression in the context of her boyfriend B’s expressing increasing anger at her and his saying that he wanted to move out of their home. An anxious person at baseline, Ms. A took to her bed for several weeks, closing the blinds so that no light could enter her room. She stopped eating, lost weight, and slept most of the time. She dropped out of school and took a leave of absence from work. She thought almost constantly of suicide, although she had not picked a method and her deep religious commitment made this an unacceptable choice. In her first session, she spoke about how worthless she felt and how she blamed herself for her boyfriend wanting to break up with her. She said she thought that she was too “pointless,” “anxious,” and “boring” for him and that her presence stifled him. She expressed no anger toward B at the time of her presentation, saying that she “understood” why he was like this. She also said she hoped that this was just a “bad phase” that he would get over, although he told her repeatedly that he was dating another woman. In psychotherapy, it gradually emerged, however, that over the past month, B had been breaking dishes and furniture in their home. He yelled at Ms. A almost constantly. This patient’s depression remitted with psychodynamic psychotherapy only as she became aware of her own murderous rage toward B, at which time she began to view herself in a less negative and denigrated manner.

Guilt

Melanie Klein postulated that depressed patients fear that they cannot protect an idealized or good, internalized “other” from destructive, rageful impulses. Although emphasizing a different facet of major depression, this view coincides with Freud’s focus on the destruction of the object tie to major depression. As a result, the depressed patient’s characteristic guilt, inhibitions, and punitive superego develop. However, not all depressions are characterized by excessive guilt, and Klein’s description applies only to this subset of patients. Klein also highlighted the danger that depressed patients foresee in “triumphing” over parents or siblings via any life success: success is experienced as aggressively humiliating to loved ones or as damaging to others. Klein theorized that idealization and devaluation are “manic defenses” against the guilt and sense of loss experienced in the resulting depression.

Ms. C, a 23-year-old, became acutely depressed when she was accepted to a prestigious graduate school. Ms. C had been working diligently toward this acceptance for the past 4 years. She reported being “briefly happy, for about 20 minutes” when she learned the good news but rapidly slipped into a hopeless state in which she recurrently pondered the pointlessness of her aspirations, cried constantly, and had to physically stop herself from taking a lethal overdose of her roommate’s insulin. In treatment, she focused on her older brother, who had regularly insulted her throughout the course of her life, and how “he’s not doing well.” She found herself very worried about him. She mentioned that she was not used to being the “successful” one of the two of them. In connection with her depression, it emerged that Ms. C’s brother had had a severe, life-threatening, and disfiguring pediatric illness that had required much family time and attention throughout their childhood. Ms. C had become “used to” his insulting manner toward her. In fact, it seemed that she required her brother’s abuse of her in order not to feel overwhelmed by survivor guilt about being the “healthy, normal” child. “He might insult me, but I look up to him. I adore him. Any attention he pays to me is like a drug,” she said. Ms. C’s acceptance to graduate school had challenged her defensive and essential compensatory image of herself as being less successful or damaged, in comparison with her brother, thereby overwhelming her with guilt. Her depression remitted in psychodynamic psychotherapy as she better understood her identification with and fantasy submission to her brother.

Impairment in Self-Esteem Regulation

A general trait of patients with major depression is the loss of self-regard. Yet, the loss of self-regard can occur in the absence of depression. Edward Bibring disagreed with Klein’s formulation that emphasized the importance of a punitive superego and argued that conflicts about aggression and object loss were secondary determinants of depression. He viewed depression instead as resulting from a sense of helplessness, impaired self-esteem, and self-directed anger triggered by failures to live up to the narcissistic aspirations of any developmental phase.

Charles Brenner also deemphasized the classic psychoanalytic focus on object loss, seeing a depressive propensity as equally likely to connect with organizing fantasies of narcissistic injury, and particularly of castration. Brenner states that these fantasies are accompanied by reactive aggression against those blamed for the painful affects, with consequent guilt.

Mr. D, a 39-year-old successful and aggressive lawyer, had "never been depressed a day in my life" before developing a kidney stone that required a 5-day hospital stay full of medical interventions. He was well on discharge but developed major depression, with lethargy, decreased energy, hopelessness, and weight loss. He cried all the time. "I feel like a big, fat failure," he announced. Handsome but insecure, Mr. D prided himself on his physical prowess and appearance and on his ability to attract the amorous attentions of "beautiful women." "I'm not supposed to be sick," he said, "Especially down there. It makes me feel old, ugly, and unattractive. Disgusting, in fact." In describing his experiences in the hospital, he reported that the humiliation, or narcissistic blow to his self-esteem resulting from the illness and hospitalization, had so overwhelmed him ("I was hooked up to these disgusting tubes—even in my penis!" he wailed) that he had uncharacteristically not even complained about being placed into a four-bed hospital room with three homeless people who hadn't showered. "I figured they were in control, so I didn't even think about it," he said. His symptoms remitted in psychoanalysis with a therapeutic exploration of the way in which the hospitalization had in fantasy felt like a castration.

Many contemporary psychoanalysts amplify upon these models in their understanding of depression while acknowledging the prominence of impaired self-esteem regulation. Edith Jacobson emphasized the development of self and object representations in depressed patients. She noted depressed patients' disappointment with parental figures, resulting in devaluation and degradation of their images and self-representation, especially when a mature separation had not been achieved.

Inadequacy of Early Caregivers

Psychoanalysts have given a personal, intrapsychic face to the much later articulated and researched, well-known epidemiologic observation of the connection between parental (particularly maternal) depression and subsequent depression in children. Hans Kohut described depression as connected to experiences of profound emptiness in patients whose parents were unable to empathize with their early affective experiences. Such is the case, as many parents of depressed patients are themselves depressed. These patients crave compensatory relationships (“self-object” relationships, mirroring experiences, and idealizing relationships), leaving them vulnerable to disappointment, as real relationships cannot live up to these compensatory fantasies.

Leo Stone highlighted depressed patients’ refusal to accept the separateness and autonomy of the object. Again, this view emphasizes the symbiotic tie to the object to compensate for an incomplete sense of self. Stone suggested that depressed patients unconsciously coerce objects; they are disappointed in them and prone to envy and rage because of an early history of “oral frustration.” (The “oral frustration” aspect of this formulation is somewhat outdated in its drive-derivative focus and might be better understood from a relational perspective, of parental relationships being experienced as being disengaged or depriving.) Aggressive fantasies about disappointing and hurting loved ones give rise to the severe guilt with which these patients struggle.

Ms. E, a 21-year-old college student, presented with major depression and panic disorder since early adolescence. She reported hating herself, crying constantly, and feeling profoundly hopeless in part because of the chronicity of her illness. Even at the time of presentation, she noted her sensitivity to her mother's moods. "My mother's just always depressed, and it makes me so miserable, I just don't know what to do," she said. "I always want something from her, I don't even know what, but I never get it. She always says the wrong thing, talks about how disturbed I am, stuff like that, makes me feel bad about myself." In one session, E poignantly described her childhood: "I spent a lot of time with my mother, but she was always too tired; she never wanted to do anything or play with me. I remember building a house with blankets over the coffee table and peeking out, spying on her. She was always depressed and negative, like a negative sink in the room, making it empty and sad. I could never get her to do anything." This patient experienced extreme guilt in her psychotherapy when she began to talk about her mother's depression. "I feel so bad," she sobbed. "It's like I'm saying bad things about her. And I love her so much, and I know she loves me. I feel it's so disloyal of me." Her depression remitted in psychodynamic psychotherapy as she became more aware of and better able to tolerate her feelings of rage and disappointment with her mother.

Sidney Blatt, a psychoanalytic researcher, contrasted *anaclitic* depressed patients, anxiously attached individuals who struggle with excessive dependence on others, with *introjective* depressed patients, who are compulsively self-reliant. Blatt contended that anaclitic patients suffer more from feelings of loneliness, helplessness, and weakness, introjective patients from a sense of worthlessness, self-criticism, and guilt.

PSYCHOANALYTIC FORMULATIONS OF DYSTHYMIA

Psychoanalysts have largely written about major depression, but some have also addressed dysthymia, or what DSM-5 now calls persistent depression. Asch noted underlying masochistic pathology in dysthymic patients, a view that has been deemphasized by contemporary dysthymia researchers, who have underlined that chronic depression can appear phenomenologically like masochistic character structure simply because of its chronicity. (In other words, masochism is a redundant concept in chronic depression.) Milton Horowitz emphasizes the “pleasure of revenge” in the patient’s defeating all around them through failure, hopelessness, and negativity, and he regards this as more important than the experience of personal suffering for many such patients. David Milrod describes both the rewarding and the punitive aspects of superego function in response to narcissistic injury in patients with chronic dysthymia and self-pity.

Ms. F, a 20-year-old college student, came to psychotherapy complaining of chronic depression. She described herself as “dull, but difficult too.” While she enjoyed school, she had trouble finding things outside of school that she enjoyed doing. She was bored, lethargic, and unhappy most of the time. Although she had friends and a boyfriend—albeit a man much less desirable than she might otherwise have been expected to find—she resented their demands on her time, often because she found herself upset when she was with them. Her relationships with her boyfriend, some of her closer friends, and her mother were characterized by intermittent “crises,” events when Ms. F felt ignored, criticized, and abused by them, and she wound up locking herself in her room to cry, sometimes for days. As psychotherapy unfolded, it became clear that her boyfriend was very demanding of her time yet rarely actually spoke to her. For example, he preferred to have her “around” while he worked on his computer for hours on end, but when he went out “to hang out,” he preferred to spend time with his friends without her. Ms. F handled this situation with a combination of resignation and resentment: She stayed with him while he worked, then often ignored him at parties that they both attended in order to “get back at him.” In psychodynamic psychotherapy, her dysthymia and passive masochism resolved simultaneously, enabling her to stop feeling worthless in social situations and with her boyfriend and to become more assertive in all of her relationships.

SYNTHESIS: DYNAMICS OF DEPRESSION

Features common to many psychoanalytic theories of depression include feelings of exquisite narcissistic vulnerability stemming from a variety of sources, including early loss or experiences with parents perceived as traumatically unempathic, frustrating, or rejecting. A sense of helplessness or inadequacy in relation to these experiences, with accompanying fantasies of damage or castration, can contribute to this vulnerability. The resulting impairment in self-esteem regulation is common to all depressed patients, who are prone to a self-image of being unlovable, damaged, or inadequate.

Depressed patients perceive that they have either failed to live up to their ambitions or to their moral values in the ego ideal, the intrapsychic mechanism that triggers guilt in depression. Many psychoanalysts hypothesize that the resulting aggression toward a frustrating parent or toward the self as damaged contributes decisively to the propensity toward depression. In depressed patients, aggression is largely self-directed. Guilt (conscious or unconscious) or shame theoretically results from the patient's perceived sense of failure, with a diminished sense of self. Difficulties in self-esteem regulation contribute to a self-representation of being "bad" or shamefully out of control, aggravating the original problem in a vicious cycle.

The sine qua non of depression in this conceptualization is aggression directed toward the self-representation, aggression that proves uncontainable and spreads to a mood state. This may arise from a withdrawal of positive, or "libidinal," supplies to the self-representation and replacement with aggression in the following ways:

1. By not living up to personal aspirations (giving rise to shame, rather than guilt);
2. By not living up to the ego ideal (precipitating guilt);
3. In an interpersonal depression, as described by Freud, in which a symbiotic bond to an ambivalent object is shattered. From a psychoanalytic perspective, this type of depression is associated with more primitive pathology, and these patients are more likely to be in a borderline or psychotic spectrum because of the nature of the primitive, symbiotic interpersonal bond.

These different pathways trigger a common self-assessment of being worthless, bad, inadequate, or unlovable. The resulting hostility directed toward the self is a core feature of depression.

Retroflected hostility has been pinpointed in the historical psychoanalytic literature as also contributing to borderline and obsessional pathology. Indeed, many patients with these underlying character structures also experience major depression. From a psychoanalytic perspective, the essential feature of all depressions is self-directed aggression, which becomes a mood state. Difficulty with self-esteem regulation, which is not an essential component of other syndromes, is always prominent in depression, coupled with aggression turned inward that exacerbates guilt and shame.

PSYCHOANALYTIC FORMULATIONS OF MANIA

From a psychoanalytic perspective, the clinical presentation of mania arises as a result of a global, massive regression that affects all three psychic structures: The id, the ego, and the superego. The regression leads to a primitive mental state in which the pleasure principle is reinstated. In *Group Psychology and Analysis of the Ego*, Freud described mania as a fusion of the ego with its superego. Less cryptically, psychoanalysts have highlighted a common organizing fantasy in manic patients of a fantasy incorporation, or mystical union, of the patient with someone of great power, often an aristocrat, or God, as in the story of St. Theresa’s mystical union with God. Such organizing fantasies, couched in both sexual and “oral” terms, magically impart a sense of omnipotence and specialness to the patient, highlighting one aspect of the common phenomenology of mania. Bertram Lewin noted that in mania, early relationships with both parents that had become desexualized during the process of superego formation during latency become resexualized during the manic episode and that in fantasy, the manic patient is identified with both partners (male and female) during the sexual act. The states of mania and hypomania involve massive utilization of the defense mechanism of denial, in which sad and negative aspects of reality are entirely ignored. One author hypothesized that the manic episode was precipitated by the need to control the memory and experience of intolerable pain. In this way, mania could be considered a defensive reaction, different from depression, in response to pain and distress. True mania is not amenable to psychotherapy treatment alone, requiring pharmacologic interventions.

Ms. G, a 42-year-old housewife and mother of a 4-year-old boy, developed symptoms of hypomania and later of frank mania without psychosis when her only son was diagnosed with acute lymphocytic leukemia. A profoundly religious woman who had experienced 10 years of difficulty with conception, Ms. G was a devoted mother. She reported that she was usually rather “down.” Prior to her son’s illness, she used to joke that she had become pregnant with him by divine intervention. When her son was diagnosed and subsequently hospitalized, he required painful medical tests and emergency chemotherapy, which made him very ill. The doctors regularly barraged Ms. G with bad news about his prognosis during the first few weeks of his illness.

Ms. G was ever-present with her son at the hospital, never sleeping, always caring for him, yet the pediatricians noted that as the child became more debilitated and the prognosis more grim, she seemed to bubble over with renewed cheerfulness, good humor, and high spirits. She could not seem to stop herself from cracking jokes to the hospital staff during her son’s painful procedures, and as the jokes became louder and more inappropriate, the staff grew more concerned. During her subsequent psychiatric consultation (requested by the pediatric staff), Ms. G reported that her current “happiness and optimism” were justified by her sense of “oneness” with Mary, the Mother of God. “We are together now, she and I, and she has become a part of me. We have a special relationship,” she winked. Despite these statements, Ms. G was not psychotic and said that she was “speaking metaphorically, of course, only as a good Catholic would.” Her mania resolved when her son achieved remission and was discharged from the hospital.

COGNITIVE THEORY OF DEPRESSION

Learning theory has long been a branch of behavioral psychology. Aaron Beck, finding that psychoanalytic theory did not sufficiently explain dreams of depressed patients, developed a theory of depression based on educating the patient about their negative thinking or cognitions (“automatic thoughts”), then conducting behavioral tests to disconfirm depressively distorted cognitions. Beck and colleagues then successfully tested CBT, a treatment built on this theory, in clinical trials. The cognitive model is based on the recognition that people are not objective; rather, an individual's idiosyncratic perception of events affects their emotions and behaviors. Depressed individuals perceive reality in subjectively depressed ways. Elaborate discussions of cognitive theory exist, and cognitive explanations have been extended from their original depressive origins to a range of psychopathology. Brad Alford and Beck argue that cognitive theory provides a comprehensive and coherent paradigm for psychopathology.

Beck's initial observations about major depression have salience and simplicity worth repeating. He noted that depressed patients tend to have characteristically skewed and negative thoughts about (1) themselves, (2) their environment, and (3) the future, a cluster he termed the *cognitive triad*. Indeed, depressed individuals frequently report negative thinking about themselves: “I'm a loser,” “Everything I do goes wrong,” “I'm weak and defective.” The environment appears hostile and overwhelming: “Even if I felt capable—which I don't—there's no way I could cope with what I have to do”; “My friends will react badly if I try to speak up”; “She will reject me.” Finally, not only do things look grim in the present but there is no prospect of relief in the future: “It's never going to get any better.” These three aspects of negative perspective converge to provide a convincingly, despairingly bleak, and mood-congruent view of the world. This outlook helps to explain why depressed patients see no way out of their misery and contemplate suicide.

Cognitive theory has explored the form as well as the content of thinking characteristic of depressed patients. Not only are cognitions skewed to the negative and pessimistic, but particular types of distortions occur. Depressed individuals tend to engage in “all or nothing” dichotomous thinking: If things aren't entirely one way, then they must be the opposite. Depressed individuals make arbitrary (negative) inferences about events, selectively abstract negative details out of context, overgeneralize (concluding negative rules from single instances), magnify (the negative) and minimize (the positive), and take personally events that may not be directly about them. Depressed patients “catastrophize,” leaping from one imagined worst-case scenario to the next in disastrous cascade:

> He will take what I said as an insult, and will never speak to me again, and will tell all our friends. Then no one will like me, I'll be all alone, and my social life will be ruined

forever.

Depressed individuals selectively focus on negative evidence, failures, and setbacks that confirm their theories of defectiveness while ignoring or discounting the successes they have achieved as “flukes.”

J, a 19-year-old college sophomore, made two mistakes on a test. Although his overall grade was high and the mistakes themselves appeared minor to the therapist, J took them as proof that he was defective and deficient. His perfect score on the next test, by contrast, he discounted as “luck…the teacher was easy on me,” not as any evidence of his personal competence. If, while depressed, you consider yourself a loser, then you pay most attention to outcomes that fit that schema.

Beck made the important point that the depressed outlook is not objective but irrational. “Automatic thoughts” of characteristically self-deprecatory and hopeless nature pop involuntarily into the patient’s head. Because these thoughts are mood-congruent, depressed individuals find them believable and tend not to question them. The negative thinking characteristic of depressed patients is damaging in two respects. First, it is painful and depressing: Many patients are barraged with negative ideas about themselves and their situation, in effect a stream of insults. Second, these thoughts tend to inhibit action: If you know you are incompetent and going to fail, feelings of hopelessness and thoughts like, “Why bother?” follow. Depressive thinking patterns not only hurt, but they also paralyze individuals into inaction, which then leaves them more time to ruminate on their inactivity (“I’m not doing anything with my life”), and to suffer self-criticism. The deterioration of productive behavior due to negative thinking will then lead to more things going wrong in the patient’s life, providing further examples of the patient’s incompetence and reinforcing the negative thinking.

The case examples above provide good examples of self-critical thinking, which in cognitive theory would be understood as the irrational, automatic thoughts of a depressed patient: Ms. A felt “pointless” and “boring”; Mr. D “like a big, fat failure,” “old, ugly, and unattractive…disgusting”; Ms. F as “dull, but difficult too.”

Immediate and specific negative cognitions (“He doesn’t like me”) fit into larger, more basic, and stable patterns of self-conception called “schemas” or core beliefs. For depressed patients, these tend to be global, negative, all-or-nothing assumptions based on early childhood experience: “If I do something imperfectly, it’s (and I’m) worthless”; “If I’m not in total control, I’m completely helpless.” Early cognitive theory presented such negative thinking as a cause rather than an effect of depression. Although cognitions are clearly an important aspect of depression, it seems naive at this point to expect that this is the single etiology of a complex and multidetermined syndrome that has both genetic and environmental components. Hence over time, cognitive theory has tended to back away from that etiologic stance.

A general cognitive explanation of a major depressive episode is that a vulnerable individual, perhaps predisposed both by biology and by negative schemas based on early childhood, experiences a current life situation that evokes “automatic” thoughts. For example, a relatively minor setback at work might activate underlying schemas by eliciting the thoughts, “I’ve done a bad job,” and thereby “I’m a terrible worker,” “I’ve let the company and my coworkers down,” “My boss hates me,” and therefore “I’ll never succeed,” “I’m a failure.” Or a disagreement in a relationship might rouse ideas such as, “K doesn’t like me,” evoking and seemingly confirming the more general core beliefs: “I’m a terrible person,” “I’m a lousy parent,” and furthermore “No one could possibly love me.” The onset of these thoughts and accompanying depressive symptoms further compromises functioning, reinforcing the negative outlook. As the mood darkens and the negative thoughts become ever more credible and pervasive, the patient gives up activities (“Why bother?”) and feels still more helpless and hopeless.

Cognitive therapy, the treatment that follows from this approach, includes a Socratic discussion and evaluation of the patient’s thoughts, weighing the evidence supporting and contradicting such thoughts. The patient actively tests hypotheses based on automatic thoughts (“I’ll fail at whatever I do”) by attempting various selected behaviors as homework. As the patient learns to recognize the irrational nature of depressive thinking, they can challenge rather than simply believe it and can begin to extinguish such thinking, replacing automatic irrational thoughts with rational responses. Outcome research repeatedly has shown that this approach is efficacious in treating mood disorders and other psychiatric syndromes.

The decades since Beck developed this original approach to depression have yielded various CBT adaptations. More recent “third wave” cognitive approaches have emphasized mindfulness: learning to recognize that you have negative thoughts and feelings but letting them pass nonjudgmentally, learning to ignore them. Thus although CBT has diffused in various directions and adapted to the treatment of various disorders, the basic principles of Beck’s 1960s theory persists.

Dysthymic and Bipolar Disorders

Cognitive theory has not been particularly altered for dysthymic disorder. Rather, the chronicity of depression would simply seem to ingrain negative core beliefs more deeply. Cognitive behavioral analysis system of psychotherapy (CBASP), an eclectic treatment that James McCullough developed specifically to treat chronic depression, postulates that chronically depressed patients fail to behave appropriately in interpersonal encounters. CBASP combines elements of cognitive, behavioral, psychodynamic, and interpersonal therapies in focusing on the patient’s behaviors and thinking in interpersonal encounters, including those with the therapist.

Cognitive-behavioral treatments have been developed not only for unipolar depression but also as an adjunct to medication for bipolar disorder. Mania may be viewed as the mirror of depression, wherein overvalued positive ideas (“I’m special” or “I’m the greatest”) automatically arise, again distorting reality, but this time in the opposing direction. Thinking itself accelerates in mania. By and large, cognitive theory has had less to say about manic than depressed states, and adaptations of cognitive therapies for mania are more practical than theoretical. They focus on psychoeducation about the illness, behavioral stabilization, such as maintaining a regular sleep schedule and emphasize the importance of adherence to pharmacotherapy. Following these principles, Francisco Colom and Eduard Vieta have developed an empirically tested psychoeducational treatment for bipolar patients.

Aspects of cognitive theory have been tested for unipolar depression, with sometimes contradictory results. CBT, which directly addresses depressive, negative cognitions, and helps patients to weigh their evidence rather than simply believe them, has repeatedly demonstrated efficacy as a treatment for major depression and does improve negative cognitions. Yet other antidepressant treatments relieve depression and improve cognition without specifically addressing cognitions, and behavioral activation—a behavioral approach without the cognitive component—has been shown to be at least as efficacious as CBT. Efficacious treatments for depression may spread from a particular theorized target to have generalized benefits.

Comparison to Psychodynamic Theory

It will be seen that cognitive theory does not contradict and, to some degree, overlaps with psychoanalytic theory. The treatments emerging from these theories differ greatly, however, in large part due to their differing definitions of the role of the therapist, handling of the therapeutic relationship, and approach to the meaning of that relationship: In psychoanalytic terms, the transference. Both theories focus on intrapsychic processes and on self-critical thinking, but cognitive theory emphasizes the distinction between rational and irrational thinking and their connection to mood rather than to unconscious fantasy. In practice, psychodynamic and cognitive therapies also differ dramatically in their structure, application of homework, and therapist stance.

INTERPERSONAL THEORY OF DEPRESSION

Interpersonal theory dates back to the era after World War II when it arose as a heretical response to the more intrapsychic emphasis of psychoanalysis. Psychoanalytic theory emphasized the importance of early life experience, and many therapists at that time saw the patient's psychic structure as essentially formed by the end of adolescence. Psychiatrists such as Adolf Meyer, Harry Stack Sullivan, Erich Fromm, and Frieda Fromm-Reichmann challenged then-current theory by emphasizing the influence of the real impact of current life events on their patients' psychopathology, focusing on environmental and interpersonal encounters rather than underlying intrapsychic drives and structures.

Sullivan coined the term "interpersonal" as a rubric for considering current life experience. He scrutinized communications in the social field, a more external outlook than traditional psychoanalysis. The interpersonal theorists worked mainly with inpatients diagnosed with schizophrenia in a prepharmacologic era. Although their work stirred great controversy at the time and a Sullivanian school distinct from the psychoanalysis of drives and ego psychology still exists, Sullivan was trained in psychoanalysis and did not entirely disagree with his forebears. Over time, the rift between Sullivanians and other psychoanalysts has narrowed.

The consideration of current interpersonal factors gained currency over succeeding decades, and it is now mainstream clinical thinking that current life events and interpersonal functioning affect and are affected by psychopathology. A school of interpersonal psychoanalysis—not particularly focused on mood disorders—arose. Psychoanalytically trained therapists like Silvano Arieti and Jules Bemporad emphasized interpersonal factors in the treatment of depressed patients. Jack Anchin and Donald Kiesler have reviewed other interpersonally based psychotherapies. None of these theories has received empirical testing.

Researchers did develop a host of related data about interpersonal issues associated with depression. For example, research showed that interpersonal support protects an individual against depression: Having a confidant to talk to reduces the risk of developing a depressive episode. Major life stressors, including the death of a significant other, struggles in important relationships, and upheavals such as a change in marital status, housing, job status, or physical health, have been shown to increase the risk of depressive episodes in vulnerable individuals. Moreover, the onset of depressive episodes leads to deterioration in relationships and social functioning and generates negative life events.

John Bowlby postulated that people have an evolutionarily determined, instinctual drive to form emotional attachments. Animal evidence now supports this theory. This basic component of human nature ensures infant survival: Children need to have parents nearby or available for feeding and protection. As children develop, they begin to explore their environment, gradually moving out from the “secure base” of their attachment figure. Disruptions in this early caregiving connection may lead to vulnerability of attachment style. For example, the loss of one’s mother in the first decade of life has been shown to be a risk factor for subsequent depression. Children with insecure childhood attachments may not learn to ask for help from others. Individuals with insecure attachment styles may have difficulty developing comfortable relationships on which they can rely for support in times of need. When such vulnerable individuals face stressors or feel an absence or inadequacy of interpersonal support during times of stress, they may be helpless to respond effectively and prone to developing symptoms.

In the 1970s, when Gerald L. Klerman, Myrna M. Weissman, and their colleagues were designing a randomized controlled pharmacotherapy trial of outpatients with major depression, they recognized that many such patients received psychotherapy in community treatment. They sought accordingly to add psychotherapy to their trial but realized that it was unclear then (as now) what such community psychotherapy consisted of. They then developed IPT as a manualized, time-limited treatment for outpatients with major depressive disorder based on interpersonal and attachment theories as well as empirical evidence of the psychosocial nature of depression. IPT is thus one specific therapeutic application of more general interpersonal theory. IPT demonstrated efficacy for major depression in this and in numerous subsequent outcome trials.

In simplest terms, interpersonal theory as applied to IPT can be understood as a link between mood and events. For all people, upsetting external events evoke a sad or demoralized mood: In lay terms, they are “depressing.” For biologically or environmentally predisposed individuals, however, a sufficiently disturbing life event can trigger an episode of major depression. Examples of such life events are the death of a significant other (complicated bereavement), a problematic relationship (role dispute), or other major life changes (role transition). Once a depressive episode starts, its symptoms compromise functioning, producing more negative life events that further worsen mood in a vicious downward cycle. This formulation seems straightforward, even commonsensical, but depressed patients have a peculiar amnesia for external events and tend to blame themselves for how they feel or to see the depressed state as who they are. IPT helpfully employs a medical model to remind patients that they are ill, not defective, and that outside events may have contributed to their distress.

IPT therapists do not propose this as an etiologic theory of depression but as a pragmatic one: The depressive mood episode can be linked either to a precipitating life event or to consequent life events that become the focus for treatment. The IPT therapist defines major depression as a medical illness—a treatable medical problem that is not the patient’s fault—and links it to an interpersonal focus such as a role dispute. The therapeutic contract for the patient is to solve the interpersonal focus within a time-limited period. Solving the interpersonal problem is at once a realistic relief (e.g., improving a marriage), relieves depressive symptoms, conveys a sense of mastery over the environment, and builds interpersonal skills that may hopefully protect against future interpersonal triggers and depressive episodes. Typical areas of interpersonal skill-building are issues such as self-assertion, confrontation, effective expression of anger, and the taking of social risks.

Clinicians armed with differing theories approach the same material in different ways. IPT seeks a life event or interpersonal situation as a plausible and pragmatic framework on which to focus a time-limited treatment in which the patient can work on interpersonal skills. For example, the case of Ms. A might be conceptualized as major depression arising in the context of a role dispute with her boyfriend, B. Had she been treated in IPT, the focus might have been on recognition and appropriate expression of her own anger as part of learning to renegotiate that relationship to ensure her own needs were met. Alternatively, or additionally, clinical judgment might have defined the breakup of that relationship as a role transition that Ms. A needed to mourn and accept in order to move on to better relationships or activities. Ms. C's difficulty in tolerating her acceptance to graduate school might similarly be considered a role transition, as might Mr. D's bout with renal stones. Medical illnesses, even if transient, frequently provoke role transitions by shifting a patient's perception of his life trajectory; Mr. D may have interpreted his hospitalization as evidence of his fading potency, aging, and mortality. Ms. E's situation might conceivably be defined either as a role transition—adjusting to college life away from her mother—or as a role dispute with her mother.

Preliminary evidence supports the theory underlying IPT as a treatment. Patients in two IPT treatment trials were asked to report the degree to which they had resolved the interpersonal crisis (e.g., role transition) on which their treatment had focused. Subjects' report of making changes in this interpersonal focal area correlated with symptomatic improvement during the trial. We recently found that time-limited IPT improves separation anxiety in patients with PTSD and might work in part by repairing insecure attachment.

Dysthymic Disorder

An adaptation of IPT has been developed for dysthymic disorder, but differences in the treatment are largely those of technique. Patients with chronic depression tend to be socially isolated (Ms. F above is slightly unusual in having a boyfriend, although her choice of boyfriend was in keeping with the disorder) and invariably have interpersonal difficulties. If the syndrome began early in life, as is often the case, patients might never have developed important communication skills. Even more than people with acute depression, they tend to be passive and unassertive, to experience self-assertion as “selfish” rather than a healthy and sometimes necessary negotiation of one’s own goals in life, and to view anger as a “bad” feeling rather than a useful interpersonal signal that someone is bothering them. Even if they have had such capabilities in the past, their interpersonal functioning will have eroded under the pounding discouragement of chronic depression. Chronicity breeds resignation and entrenches maladaptive interpersonal patterns. The adaptation of IPT for chronic depression focuses on the patient’s inevitable confusion of chronic mood with who they truly are beneath it and seeks to disentangle the two by helping the patient to trust their feelings by handling events in a more assertive fashion.

Bipolar Disorder

In all of the theoretical approaches under discussion, mania has received less attention than unipolar depression. There is no real adaptation of interpersonal theory for mania. Clinically, it is apparent that manic patients lose social judgment and become overinvolved and labile in relationships with others. Ellen Frank and colleagues have adapted IPT for bipolar patients as an adjunctive treatment to pharmacotherapy by combining it with a behavioral strategy. Interpersonal social rhythms therapy (IPSRT) is a hybrid of IPT and social zeitgeber theory, which stresses the importance of maintaining diurnal structure through regular patterning of circadian events or social rhythms. In particular, this involves ensuring a regular bedtime to prevent the loss of sleep that often triggers a manic episode. As with CBT for bipolar disorder, this represents more of a clinically reasonable, pragmatic extension of an already established treatment than a theoretical evolution. This approach, described in a manual, has shown benefits as an adjunct to pharmacotherapy in randomized controlled trials for bipolar patients.

Future Directions

Psychoanalytic, cognitive, and interpersonal theories all provide insights into mood disorders. None is received truth, none can be considered fully tested in empirical studies, and all are surely incomplete. Nor has this section covered all theories of mood disorder: Biologic, behavioral, feminist, and many other theories have been proposed for depression. Psychoanalytic, cognitive, and interpersonal theories each capture important aspects of complex syndromes. In the hands of skilled clinicians, therapists, and patients may find each approach useful in conceptualizing the patient's status and organizing their treatment.

Future research may elaborate more comprehensive applications of these theories to bipolar disorder. And theory alone does not suffice to train a clinician: clinical experience and supervision, as well as personal therapy, help to expand clinical understanding and ability.

FURTHER READINGS

Abraham K. Notes on the psycho-analytical investigation and treatment of manic-depressive insanity and allied conditions. In: *Selected Papers on Psychoanalysis*. Hogarth Press; 1927:137.

Anchin JC, Kiesler DJ, eds. *Handbook of Interpersonal Psychotherapy*. Pergamon Press; 1982.

Arieti S, Bemporad J. *Severe and Mild Depressions*. Basic Books; 1978.

Basco M, Rush AJ. *Cognitive-Behavioral Therapy for Bipolar Disorder*. Guilford Press; 1996.

Beach SRH, Kogan SM, Brody GH, Chen YF, Lei MK, Murry VM. Change in caregiver depression as a function of the Strong African American Families Program. *J Fam Psychology*. 2008;22(2):241–252.

*Beck AT, Rush AJ, Shaw BF, Emery G. *Cognitive Therapy of Depression*. Guilford Press; 1979.

Bowlby J. *Attachment and Loss*. Basic Books; 1973.

Colom F, Vieta E. *Psychoeducational Manual for Bipolar Disorder*. Cambridge University Press; 2006.

Copeland LA, Zeber JE, Salloum IM, Pincus HA, Fine MJ, Kilbourne AM. Treatment adherence and illness insight in veterans with bipolar disorder. *J Nerv Ment Dis*. 2008;196:16–21.

DeRubeis RJ, Hollon SD, Amsterdam JD, et al. Cognitive therapy vs. medications in the treatment of moderate to severe depression. *Arch Gen Psychiatry*. 2005;62–416.

Eells TD, ed. *Handbook of Psychotherapy Case Formulation*. 2nd ed. Guilford Press; 2007.

Frank E. *Treating Bipolar Disorder: A Clinician's Guide to Interpersonal and Social Rhythm Therapy*. Guilford Press; 2005.

Freud S. Beyond the pleasure principle. In: *Standard Edition of the Complete Psychological Works of Sigmund Freud*. Vol. 18. WW Norton & Co; 1920:3–67.

*Freud S: Mourning and melancholia. In: *Standard Edition of the Complete Psychological Works of Sigmund Freud*. Vol. 14.WW Norton & Co; 1915;237–258.

Freud S: The ego and the Id. In: *Standard Edition of the Complete Psychological Works of Sigmund Freud*. Vol. 19. WW Norton & Co; 1923;3–66.

*Jacobson E. *Depression: Comparative Studies of Normal, Neurotic, and Psychotic Conditions*. International Universities Press; 1972.

Klein K, Juliet M. Mourning and its relation to manic-depressive states. *Int J Psychoanal*. 1940;21:125–153.

*Klerman GL, Weissman MM, Rounsaville BJ, Chevron ES. *Interpersonal Psychotherapy of Depression*. Basic Books; 1984.

Lewin BD. Comments on hypomanic and related states. In: Arlow JA, ed. *Selected Works of Bertram D. Lewin*. William Byrd Press; 1973:78.

*Lipsitz JD, Markowitz JC. Mechanisms of change in interpersonal psychotherapy. *Clin Psychol Rev*. 2013;33:1134–1147.

Lorenzo-Luaces L, German RE, DeRubeis RJ. It’s complicated: the relation between cognitive change procedures, cognitive change, and symptom change in cognitive therapy for depression. *Clin Psychol Rev*. 2015;41:3–15.

Markowitz JC. *Interpersonal Psychotherapy for Dysthymic Disorder*. American Psychiatric Press; 1998.

Markowitz JC, Milrod BL: Affectfocused and exposure-focused psychotherapies. *Am J Psychotherapy* (in press, 2023).

McCullough JP Jr. *Treatment for Chronic Depression: Cognitive Behavioral Analysis System of Psychotherapy*. Guilford Press; 2000.

*Milrod D. Self-pity, self-comforting, and the superego. *Psychoanal Study Child*. 1972;27:505–528.

Milrod B, Keefe JR, Choo T-H, et al. Separation anxiety in PTSD: a pilot prevalence and treatment study. *Depress Anxiety*. 2020;37:386–395.

Rado S. The problem of melancholia. *Int J Psychoanal*. 1928;9:420–438.

Rush AJ, Beck AT, Kovacs M, Weissenburger J, Hollon SD. Comparison of the effects of cognitive therapy and pharmacotherapy on hopelessness and self-concept. *Am J Psychiatry*. 1982;139:862–866.

Sandler J, Joffe WG. Notes on childhood depression. *Int J Psychoanal*. 1965;46:88–96.

Sullivan HS. *The Interpersonal Theory of Psychiatry*. WW Norton; 1953.

Weissman MM, Markowitz JC, Klerman GL. *Clinicians’ Quick Guide to Interpersonal Psychotherapy*. Oxford University Press; 2007.

Zeber JE, Copeland LA, Good CB, Fine MJ, Bauer MS, Kilbourne AM. Therapeutic alliance perceptions and medication adherence in patients with bipolar disorder. *J Affective Disord*. 2008;107(1–3):53–62.

▲ 13.5 Mood Disorders: Suicidal Behavior

MAURIZIO POMPILI, M.D., PH.D., AND ZOLTÁN RIHMER, M.D., PH.D., D.SC.

Suicidal behavior is a major public health concern worldwide, with approximately 800,000 people dying annually from completed suicide. Suicide attempts are at least 10 times higher than the number of suicides. Furthermore, suicidal ideation and planning, although difficult to ascertain, are much more frequent phenomena than suicidal behavior. Mood disorders have been unanimously associated with suicidal behavior due to the nature of the patient's symptoms, and mood-disorder patients die from suicide, on average, 10 years earlier than from other causes. Depression, both unipolar or bipolar, seems insidious in precipitating suicide risk. However, suicide is a multifactorial phenomenon, with an array of variables influencing such risk. Mood disorders are, therefore, major contributory but not exclusively factors for suicidal behavior. For instance, feelings of hopelessness (e.g., not having future expectations, etc.) are found to be more predictive of suicide risk than a diagnosis of depression *per se*. Therefore, it seems reasonable to acknowledge the interplay of mood disorders with other critical factors, such as familial–genetic disorders, personality, and psychosocial and demographic risk factors, which also play a significant role.

The exploration of these factors' role helps to better understand what makes the minority of patients with mood disorders suicidal. Furthermore, nowadays, clinicians are considering suicide risk from the neurodevelopmental perspective, comparing distal risk factors (family history of suicide, epigenetic factors, early life adversity, personality traits, and early substance abuse) with proximal risk factors (psychopathology, suicidal ideation, hopelessness, recent life events). History of untreated major mood disorders (particularly in previous suicide attempts) constitutes an important risk factor for many illness-related health and economic problems, such as loss of productivity, secondary substance-use disorders, family breakdown, and increased somatic morbidity and mortality. Given the high lifetime prevalence of unipolar major depressive, and bipolar disorders, they are among the most frequent and potentially life-threatening psychiatric illnesses. Despite the great clinical and public health significance of major mood disorders, people with these disorders are still under-referred, underdiagnosed, and undertreated. This is particularly true for those who die by suicide and for those who attempt suicide and receive medical attention. As killing oneself is considered a major cause of years of life lost, its prevention is receiving more and more attention.

FROM SUICIDE RISK ASSESSMENT TO SUICIDE RISK FORMULATION

Suicide risk assessment should be part of health professionals' skills, although it is often taken for granted that only mental health professionals should deal with such tasks. Recently, experts on suicide realized that collecting risk factors and assessing the risk yielded an incomplete process, lacking proper judgment on the risk level and proper interventions for that specific risk.

Experts on suicide introduced the suicide risk formulation (SRF) as a data-dependent process in the suicide risk assessment. SRF assigns a suicide risk level intended to inform decisions about triage, treatment, management, and preventive interventions. Considering that the traditional model points to a low, medium, and high risk of suicide, the risk formulations summarize the data into four different judgments to directly inform intervention plans (Fig. 13.5–1): (1) risk status (the patient's risk compared to a specified subpopulation), (2) status risk (the patient's risk concerning baseline or other specified time points), (3) available resources that the patient can draw on in the crisis, and (4) foreseeable changes that may exacerbate the risk. This approach aims to consider the risk factors coming from the literature and consider specific factors that surround a specific individual who may be threatened by specific and unitary sources of distress for which the risk of suicide is increased.

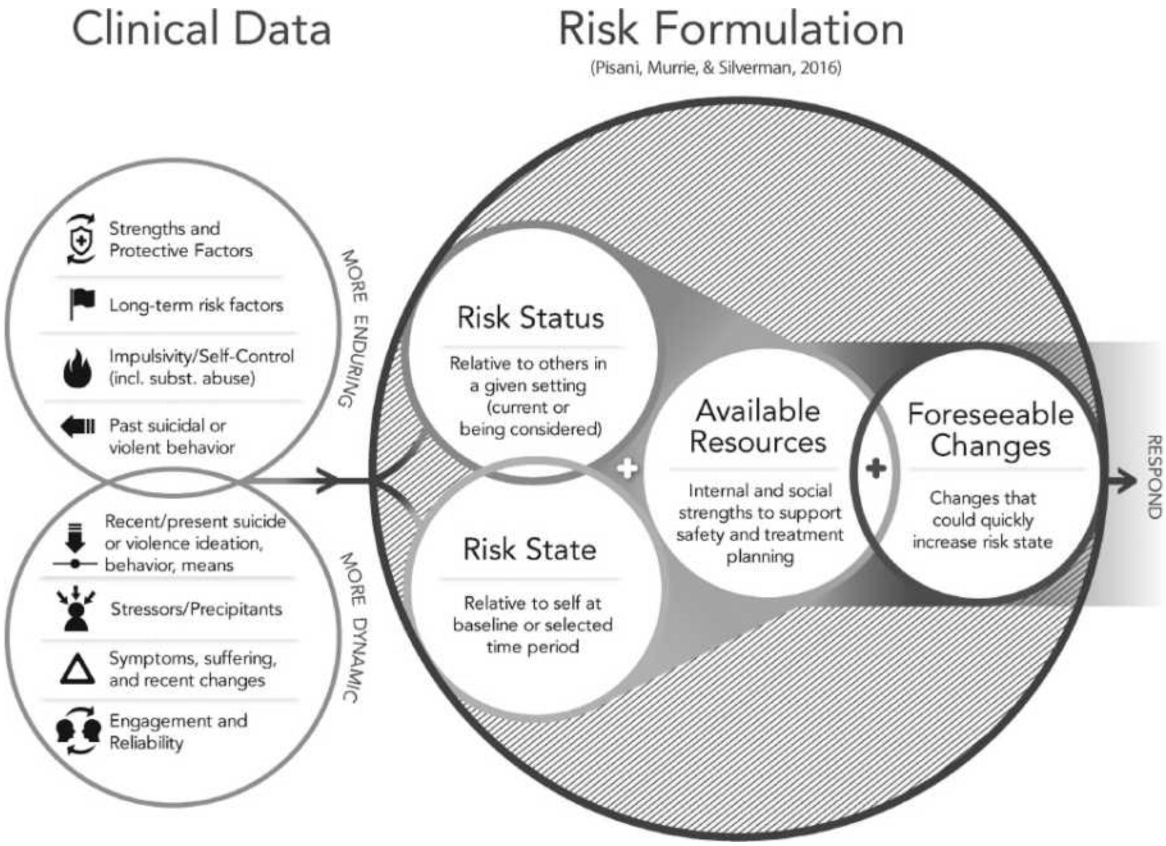
Clinical Data
Risk Formulation
(Pisani, Murrie, & Silverman, 2016)
Strengths and Protective Factors
Long-term risk factors
Impulsivity/Self-Control (incl. subst. abuse)
Past suicidal or violent behavior
MORE ENDURING
Recent/present suicide or violence ideation, behavior, means
Stressors/Precipitants
Symptoms, suffering, and recent changes
Engagement and Reliability
MORE DYNAMIC
Risk Status
Relative to others in a given setting (current or being considered)
Risk State
Relative to self at baseline or selected time period
Available Resources
Internal and social strengths to support safety and treatment planning
Foreseeable Changes
Changes that could quickly increase risk state
RESPOND

FIGURE 13.5–1. Prevention-oriented risk formulation. The model flows from left to right. The left side of the model shows eight domains of "clinical data" that a clinician gathers and synthesizes in collaboration with the patient and other individuals central to the patient's life and care. (Adapted from Pisani AR, Murrie DC, Silverman MM. Reformulating suicide risk formulation: from prediction to prevention. *Acad Psychiatry.* 2016;40:632–629. doi:10.1007/s40596-015-0434-6. License: http://creativecommons.org/licenses/by/4.0/)

Recent analyses of the study of suicide provided broader perspectives with new explanatory models. A paradigm change seems to be emerging, and the analyses of personality and a mentalistic process have substituted the classic medical model for explaining suicide risk. As suicide risk has been traditionally viewed as a symptom of psychiatric disorders, a paradigm shift is now necessary to address a much broader entity, pointing to the peculiarity of the suicidal dimension, partly overlapping with the dimension that includes psychiatric disorders. A more phenomenologic approach, uncovering the features of the suicidal mind, seems imperative. Integration between psychiatry and suicidology seems the new emerging paradigm, providing an understanding of unique individuals' suffering. Both data are emerging from empirical research and neuroimaging studies focused on the role of mental pain in shedding light on the extreme suffering of the mind. Such experience seems to share neuroanatomical circuits of somatic pain and appears a strategic point to investigate suicide risk.

According to this model, the suicidal individual experiences unbearable mental pain (also known as psychache) or suffering, and suicide might be, at least in part, an attempt to escape from this suffering. Among the constructs used to describe the wish to die, a simple but extraordinary model has proved, at least for its straightforwardness, to be useful in describing the suicidal mind. According to this model, suicide is an escape from intolerable suffering, emphasizing that suicide is not a movement toward death but rather an escape from intolerable emotion and unendurable or unacceptable anguish. Experiencing negative emotions, with an internal dialogue making the flow of consciousness painful and leading the individual to the ultimate conclusion, may be related to the fact that, if tormented individuals could somehow stop consciousness and still live, they would opt for that solution. Suicide occurs when that individual deems the psychache to be unbearable. Therefore, understanding the suicidal mind allows the clinician to familiarize with dynamic risk factors and address them through a personalized, empathic approach and critical appraisal of the suicide risk in mood disorders.

Clinicians who consider suicidal ideation to merely be a symptom of depression may miss the rare opportunity to get to know a very private aspect of patients and reduce their suffering. They may even uncover that they are treating not necessarily major depression with suicidal ideation but rather severe human sadness emerging with suicidal wishes. Mental health professionals should provide differential diagnoses for the various conditions for which they are requested to intervene and, regardless of what is assessed, deal with the need for treatment. In line with this, the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) states, “Diagnosis of a mental disorder should have clinical utility,” but “the diagnosis of a mental disorder is not equivalent to a need for treatment. Need for treatment is a complex clinical decision that takes into consideration symptom severity, symptom salience (e.g., the presence of suicidal ideation), the patient’s distress (mental pain).” Furthermore, “Clinicians may thus encounter individuals whose symptoms do not meet full criteria for a mental disorder but who demonstrate a clear need for treatment or care. The fact that some individuals do not show all symptoms indicative of a diagnosis should not be used to justify limiting their access to appropriate care” (18, p. 20).

EXPLORATION OF SUICIDE RISK IN MOOD DISORDERS: DATA FROM PSYCHOLOGICAL AUTOPSIES

Most of the data shedding light on suicide come from psychological autopsies. These procedures collect information from crucial persons—spouse, lover, parent, grown child, friend, colleague, physician, supervisor, or coworker—who knew the decedent to reconstruct their mental status before death. The literature has reported some of the limits of such procedures, pointing to a lack of standardized instruments for assessing psychopathology, life events, and the procedure for interviewing informants. Studies investigated concordance among different informants (caregivers, spouses, clinicians, etc.) on diagnoses of living cases and showed that agreement ranges from poor to moderately good. The authors concluded that the informant’s profile might significantly influence the information obtained on cases, with an overlap in data from different sources that are far from perfect.

Despite the limitations of psychological autopsy studies, we learn from such studies in several countries that up to 90% of those who die of suicide have a mental disorder, possibly a mood disorder. The retrospective reconstruction considers how much it can belong, for example, to a depressive disorder and how much it is a humanly understandable sadness reaction instead. Although the ratio of completed suicide attempts in the general population is approximately 15 to 40: 1, it is much lower (5 to 10: 1) among patients with major mood disorders, suggesting that these patients use more lethal (or more violent) methods of suicide. Patients with mood disorders, particularly bipolar patients, often have a low ratio of suicide attempts to completed suicides, indicating the relatively high mortality of suicide attempts in bipolar disorder. These results indicate that bipolar patients have an unusually high level of intention to die and high lethality of the methods used. Suicide and suicide attempts are two different but highly overlapping phenomena: about half of those who complete suicide have attempted suicide at least once previously. The first attempt (even if the method used is nonviolent or non-lethal) significantly increases the risk of suicide in the future. Such a fact is in part because those who repeatedly attempt suicide often change their method from nonviolent to violent or from nonlethal to lethal. Suicidal ideation, suicide attempt, and suicide are three different but highly overlapping phenomena. Previous suicide attempts and current major depression are the two strongest predictors of future suicides. The vast majority of suicide attempts or those who have completed suicide come from a population of people with current suicidal ideation, often in association with an untreated major depressive episode.

It has been reported that the standardized death rate from suicide in patients (unipolar and bipolar) with mood disorders greater than that of the general population is 10 to 30 times. While suicide is about the 10th most common cause of mortality in the general population in well-developed countries, it is among the most common leading causes of death among unipolar major depression and bipolar disorder. The life expectancy of these patients is decreased by about 10 to 12 years. Consolidated studies in the literature estimate that 15% to 19% of severe (mostly hospitalized) patients with unipolar and bipolar major depressive episodes would have died from suicide. However, the most recent studies, perhaps reflecting the most popular and effective treatments, have found that this figure is less than 10%, especially diversified for subpopulations, such as patients hospitalized for suicide risk, patients hospitalized for a mood disorder, and outpatient patients.

Other long-term follow-up studies also show that the suicide rate is higher in bipolar disorder (types I and II combined) than in unipolar depression. Similarly, population-based epidemiologic studies have also found a substantially higher rate of suicide attempts in bipolar I and II disorders than in patients with unipolar major depressive disorder: More than half of bipolar patients and about a third of unipolar depression patients report a life history of one or more suicide attempts, and suicidal ideation, a major precursor of suicidal behavior in patients with mood disorders, is even more common. Most studies show that of the three different clinical manifestations of major mood disorders (unipolar depression, bipolar I disorder, and bipolar II disorder), bipolar patients in general and bipolar II people, in particular, carry the highest risk of both attempted and committed suicide. Although the literature has indicated that patients with bipolar II disorder (BP-II) are at increased risk of suicide, such evidence does not appear to emerge from recent meta-analysis studies. However, BP-II accounts for 30% to 58% of all major depressions in psychiatric practice, and mixed states are very common in the BP-II spectrum. Furthermore, the cyclothymic temperament is the most prevalent in BP-II and has been shown to be a predictor of suicide, and rapid mood shifts are the hallmark of BP-II. Apparently, BP-II suicides are the most likely to be lethal, as they use the most aggressive methods.

The risk of complete and attempted suicide in bipolar disorder is most pronounced at a younger age, especially in the first few years after onset. However, bipolar patients with previous suicide attempts have an earlier age of onset than those with no prior suicidal tendency. Because the correct diagnosis of bipolar disorders usually occurs on average 8 to 10 years after the onset of the disease and delays an accurate diagnosis, as well as delays in the onset of mood stabilizers at the onset of the disease, increase the risk of suicidal behavior, these findings strongly suggest that early recognition of the bipolar nature of depressive disorders and adequate short- and long-term treatment are crucial to preventing suicide.

The first suicide attempt significantly increases the risk of future suicide attempts or completed suicide, and about 20% of those who attempt suicide will eventually die at their own hands. However, up to half of the suicide victims have made at least one previous suicide attempt, indicating that about 50% of suicide victims die from their first suicide act. Therefore, detecting the risk of suicide in patients with mood disorders and planning prevention strategies as early as possible, even before the first suicide act, are crucial in suicide prevention. This chapter summarizes the risk of suicide and the most clinically detectable protective factors in mood disorders and briefly highlights the most effective preventive strategies. It will focus on complete suicide (an act of deliberately taking one’s own life) and attempted suicide (an act of self-harm with the intent to end one’s life), both commonly referred to as “suicidal behavior,” but it will not cover deliberate self-harm (defined as an intentional injury to one’s body with no apparent suicidal intent).

CLINICALLY DETECTABLE SUICIDE RISK AND PROTECTIVE FACTORS IN PATIENTS WITH MOOD DISORDERS

Risk factors often yield too many false positives because they are static elements referred to other populations and do not necessarily convey suicide risk in a specific individual. Therefore, clinicians are requested to investigate classic risk factors and other, more dynamic conditions belonging to a given individual.

Because suicide is a rare event in the community, its precise prediction in individual cases is challenging but not impossible. However, suicidal behavior (completed suicide and suicide attempt) is quite frequent among patients with mood disorders who contact different health care levels some weeks or months before their death or attempt, which emphasizes the potentially key role of health care workers in suicide prevention. The scientific literature identifies several clinical, biologic, psychosocial, and demographic risk and protective factors that are statistically significant with suicidal behavior. A meta-analysis found that a family history of suicide was also reported, especially in cases of a family history of psychiatric disorder. Few clinical characteristics of depression were indicators of risk, just more severe psychopathology and hopelessness, with only a trend for suicidal ideation. Suicide risk factors are classified as short-term versus long-term, proximal versus distal, changeable versus unchangeable, psychiatric–medical versus sociodemographic, and clinical versus biologic.

In addition, considering the competence, the field of responsibility, and the changeable/unchangeable nature of these suicide risk factors, they can also be classified hierarchically. Although the statistical relationship between the different *tertiary (*demographic) risk factors like male gender, old age, etc., and *secondary* (psychosocial) risk factors like unemployment, living alone, etc., and suicidal behavior is well-demonstrated, their predictive utility is very limited in individual cases. *Primary* (psychiatric-medical) suicide risk factors, for example, untreated depression, suicide attempt, etc., are the most powerful and clinically useful predictors of suicidal behavior, especially if secondary and tertiary risk factors are also present.

As for biologic risk factors, a decreased (dysregulated) central serotonergic system in suicide victims and suicide attempters (particularly in those with the violent method), regardless of the psychiatric diagnosis, is a very strong correlate of suicidal behavior and is one of the most consistent neurobiologic findings in the field of psychiatry. Some studies, however, suggest that alterations in other neurotransmitter systems (norepinephrine, dopamine, GABA, and *N*-methyl-D-aspartate [NMDA]) and cholesterol, as well as omega-6 and omega-3 metabolism, might also be important in this respect. Studies consistently show that hyperactivity of the hypothalamic–pituitary–adrenal (HPA) axis (as reflected in the abnormal dexamethasone suppression test) is a robust predictor of future completed suicide, primarily in patients with major depressive episodes. As decreased central serotonin metabolism is more strongly related to impulsive–aggressive personality characteristics than to any specific psychiatric disorder, these trait-related features increase the likelihood of suicidal behavior, particularly in the simultaneous presence of an acute episode of major psychiatric disorder and adverse psychosocial situations, including lack of social support. Even though current studies on the biology of suicidal behavior are promising in everyday clinical practice (particularly in urgent need of action), they have limited practical value. Therefore, as discussed below, clinically explorable risk and protective factors are the most important aids in this complicated and highly responsible task.

Specific Clinical Characteristics of the Current Mood Episode

The interplay among factors associated with the history of the individual, proximal adverse life events, and psychopathology is often traceable in patients who attempt or complete suicide. In the majority of cases, mood-disorder patients are at higher risk of suicide during major depressive or mixed episodes and very rarely during euphoric mania or euthymia (clinical recovery). Patients suffering from mixed states, experiencing depression which escalates into mania with volatile and erratic moods associated with dysphoria and agitation, are more vulnerable to the precipitation of suicide risk.

Severe major depressive episodes and the simultaneous presence of major depressive and manic syndrome (DSM-IV mixed affective episode and DSM-5 major depressive episode or manic episode “with mixed features”) are the most important time-related risk periods of suicidal behavior in patients with mood disorders. Both the DSM-IV mixed affective episode and DSM-5 mania or major depressive episode “with mixed features” are characterized not only by increased suicide risk but also by an earlier age of onset, higher frequency of manic and depressive episodes, higher rate of rapid cycling course, as well as substance use and anxiety disorder comorbidity, all of which increase the risk of suicidal behavior.

According to both clinical and research observations, depression of suicide victims differs qualitatively from living depressed controls. Suicidal depressives often experience more severe forms of depression, with features highly related to suicide risks, such as insomnia, hopelessness, anxiety, agitation, weight or appetite loss, feelings of worthlessness or inappropriate guilt, and thoughts of death or suicidal ideation. Comorbidity with substance (including alcohol) use disorder(s) is also often observed in depressed suicide victims. Patients with mood disorders who exhibit volatile and erratic moods associated with dysphoria and agitation or present the classic mixed states have a notably higher risk of suicide. Predisposing factors interact with precipitating stressors, but deaths, separations, and other major losses, scandals, or imprisonment rarely precipitate suicide in the absence of a psychiatric disorder. Current mood states are critical determinants of suicidal risk in patients with mood disorders, particularly depressive and dysphoric-irritable mixed states (DSM-5 major depressive or manic episode with mixed features) that together are present in at least 75% of suicides in bipolar patients. In contrast, suicide is extremely rare in pure mania, hypomania, and after clinical recovery.

Subthreshold lifetime or current (hypo)manic symptoms during DSM-IV unipolar major depressive disorder (subthreshold bipolar depression and mixed depression that shows great overlap with the clinical category of agitated depression) or the DSM-5 category of the major depressive or manic episode with mixed features also increases the risk of suicidal behavior. Studies also show that the psychotic nature of major mood episodes also increases the risk of suicide, particularly in patients with bipolar depression. In contrast to a major depressive episode, minor depression and pure dysthymic disorder (dysthymia without “comorbid” major depression) do not markedly increase the risk of suicide. Still, many patients with these milder forms of mood disorders develop major depressive episodes and, consequently, an increased risk of suicide during the long-term course.

Comorbid Psychiatric, Personality, and Medical Disorders

Patients with mood disorders with comorbid anxiety disorders, substance use disorders, cigarette smoking, personality disorders (mainly borderline-type), and serious, disabling, painful, and fatal medical illness are also at increased risk of attempted and committed suicide. Young persons with bipolar disorder and ADHD have an increased likelihood of suicidal behavior than young patients who have only bipolar disorder. Besides the well-known role of chronic alcoholism, acute alcohol intake is also a significant risk factor for attempted and completed suicide, even in nonalcoholic depressives. It could trigger the actual suicidal behavior and increases the lethality of suicidal acts. Strong evidence suggests that in addition to alcohol and substance use disorders, cigarette smoking is also positively related to suicidal behavior in the general population and—according to most studies—also in samples of patients with major mood disorders. Major depression, bipolar disorder, and schizophrenia are associated with higher cigarette smoking rates than that of the general population, but smoking remains a significant suicide risk factor even after controlling for psychiatric morbidity. Smokers, in general, are more impulsive, and current smokers with major depressive disorder show a greater number of depressive symptoms and higher levels of suicidality (suicide attempts, suicidal ideations) during their depressive episode than those with major depressive disorder who have never smoked. Patients suffering from disabling, painful, and/or life-threatening medical disorders, including cancer, are also at elevated risk of suicide, mainly when depression is also present. The actual risk of suicide in patients with cancer varies with the anatomical location, stage of cancer, and the result of cancer therapy. Despite recent cancer treatments, this diagnosis is still a major stressor for many patients, especially those with other psychiatric or psychosocial suicide risk factors.

Prior Course of the Mood Disorder

Suicide risk appears to be particularly high, primarily early in the course of bipolar illness. This underlines the importance of precise diagnosis and effective treatment of specific mood disorders as early as possible. Clinically detectable suicide risk factors related to the prior course of the illness in patients with mood disorders are early onset of the unipolar or bipolar disorder, relatively early stage of the illness, first episode depression, predominantly depressed polarity during the prior course of the illness (which is typical for bipolar II disorder), rapid cycling bipolar disorder, and, of course, suicidal ideation or suicide attempt in the past; all of these have been repeatedly demonstrated to increase the risk of attempted or completed suicide in patients with mood disorders.

Affective Temperaments and Specific Personality Features

Affective temperaments have been studied in connection with suicide risk. The literature has indicated that hyperthymic temperament is protective against suicide risk, whereas cyclothymic, irritable, depressive, and anxious temperaments were associated with suicide risk. Some temperaments may present features that are more related to suicide. For instance, the role of cyclothymic temperament in suicidal behavior is supported by the finding that anger and hostility are most common in remitted bipolar and major depressive patients with a cyclothymic affective temperament. The role of unstable affective temperament in suicidal behavior is also demonstrated by findings showing that cyclothymic temperament significantly predicts future bipolar transformation and suicide attempts in adult and juvenile depressives, and bipolar I, bipolar II, and unipolar major depressive disorder patients with cyclothymic temperament have a significantly higher rate of prior suicide attempts and lifetime or current suicidal ideation than noncyclothymic depressives. The persistent, frequent, and marked instability of mood, thinking, and behavior is the most characteristic feature of cyclothymia, the attenuated manifestation of major bipolar mood disorders. However, marked affective temperament, which can be detected in about 15% to 20% of the general population, become suicide risk factor only during major depressive episodes. People with marked affective temperaments have a much higher chance of developing major mood disorders. Patients with cyclothymic and irritable affective temperaments who attempt suicide significantly more frequently report childhood physical and/or sexual abuse, suggesting that besides impulsivity, cyclothymic or irritable temperaments are further mediating variables between these early negative life events and adult suicidal behavior.

Impulsivity, aggressivity, pessimism, and a few reasons for living, like personality traits, markedly increase the risk of all forms of suicidal behavior when combined with depression. There is a strong association between impulsiveness and bipolar disorder, and impulsivity becomes more pronounced during depressive or (hypo)manic episodes. Even modest manic symptoms during bipolar depressive episodes are associated with greater impulsivity and higher suicide attempts. Pessimistic personality features, cognitive rigidity, ruminative thinking, maladaptive coping strategies, and disturbed neurocognitive functions like impaired decision-making, as could also be detected by specific neurocognitive tests, are also characteristics of most suicidal depressives.

The interaction between personality features and illness characteristics in the suicidal process is best formulated in the “stress-diathesis model,” which proposes that suicidal behavior is determined not only by the stressor (acute major psychiatric illness or acute major personal crisis) but also by a diathesis or predisposition (impulsive, aggressive, pessimistic personality traits).

Previous Suicide Attempt

The literature consistently reports that previous suicide attempts, particularly in violent or more lethal methods, are the most powerful single predictor of future attempts and fatal suicide in patients with major mood disorders or other psychiatric diagnoses. A recent study, analyzing 101 articles in the international literature, found that one-third of unipolar and about 30% to 50% of bipolar depressives attempt suicide at least once during their lifetime, and suicidal behavior shows high consistency across major depressive episodes. Suicide attempters are, on average, 35% more frequent among women than men. And despite some consolidated observations, recent studies found no difference in suicide attempter risk between types I and II bipolar disorder.

Bipolar patients use more violent and more lethal suicide methods than unipolar depressives, which are more common in males. Nonviolent or low-lethality suicide attempts do not mean low suicide risk, as those who repeatedly attempt suicide commonly change their method from nonviolent to violent, but the opposite pattern is rare. Bipolar depressives with a history of suicide attempts have more severe symptomatology in general. They report more hopelessness, self-blame, guilt, and current suicidal ideation, and they more frequently show aggressive–impulsive personality traits than nonsuicidal bipolar depressives do. Therefore, higher rates of suicidal behavior (mainly completed suicide) in patients with bipolar disorder than in unipolar major depression may be partly due to a specific effect of bipolar disorder, resulting in more dangerous or more lethal suicidal behaviors.

Family History of Suicidal Behavior

Suicide risk is notably associated with the heritability of endophenotypes, and no real genetic pattern seems satisfactory for explaining the role of a family history of suicide. It appears that cultural and coping styles are transferred across generations more than specific genetic transmission.

However, family, twin, and adoption studies show that a family history of suicidal behavior and major mood disorders in first- and second-degree relatives are among the strongest risk factors for both attempted and completed suicide in psychiatric patients in general and in patients with a major mood disorder in particular. The heritability index of completed suicide is approximately 40%, and it is even higher among depression-related suicides. However, the familial component of suicidal behavior seems to be partly independent of specific psychiatric disorders, as relatives of suicidal persons are 10 times more likely than relatives of comparison subjects to attempt or complete suicide after controlling for psychopathology. Similarly, it emerged from the literature that the children of parents with mood disorders who attempted suicide were five times more likely to attempt suicide, even after controlling other offspring risk factors for suicidal behavior, such as previous suicide attempt and history of mood disorder (both assessed at baseline) or mood disorder at the time point before the suicide attempt. Bipolar and unipolar major depressive patients who have a family history of suicidal behavior and who were exposed to childhood physical and/or sexual abuse are at greater risk of suicidal behavior. Impulsivity and irritable temperament seem to link childhood abuse and suicidal behavior, particularly in pure or mixed major depressive episodes. Research data, including twin and adoption studies, show that the familial aggregation of suicidal behavior has several components: genetic background of major psychiatric disorders, genetic–familial nature of impulsive–aggressive personality traits irrespective of a psychiatric disorder, and, in some cases, copycat mechanisms. However, it should be noted that the “suggestive” effect of a suicidal family member becomes active mainly when a major depressive or other severe mental disorder develops.

Negative Life Events and Adverse Life Situations

As a multifactorial phenomenon encompassing both distal and proximal risk factors, recent negative life events may trigger suicidal behavior in vulnerable persons, particularly in high-risk groups, such as those with major depression. Adverse, unwanted, or stressful life events play an important role in the suicidal process as predisposing (childhood events, including physical and sexual abuse) and precipitating (adulthood events) factors. Although negative life events do not inevitably lead to suicidal behavior in healthy persons or even in high-risk groups such as psychiatric (particularly mood disorder) patients, they may actually trigger the suicidal process, particularly if other risk factors or adverse situations are also present. One-third of patients with unipolar major depression or bipolar disorder report a history of childhood physical and sexual abuse or neglect, and depressed patients with such a history have an earlier age of onset of bipolar illness, greater psychiatric comorbidity, and increased rates of suicide attempts. These adverse childhood events may lead to developmental and neuroendocrinologic scars that render the subject vulnerable to suicidal behavior in a crisis situation during adulthood, including severe major depression or bipolar disorder.

Affect mastery, or the capacity to regulate affect, is gradually acquired after birth–it is not present early on. Only with time and reasonable parenting do children attain the capacity to regulate themselves and keep themselves in reasonable emotional composure. The various developmental stages bring different emotional stresses to children. They must learn to deal with separations, frustrations, or other emotionally challenging circumstances. Thus, the importance of good early attachment experiences is unambiguous. The parent who is frightened, insecure, or vindictive prevents the child from making a “secure attachment.” Insecure attachment with affective dysregulation, which is now identified in various central nervous system activities and HPA-axis activity, is considered a suicide risk.

About half of all completed suicides in bipolar and unipolar mood disorders are associated with recent severe, acute negative life events or adverse life situations, such as the death of a relative or close friend, isolation, living alone, separation, or divorce, and being unmarried. However, the stressors are commonly dependent on the victims’ own behavior, particularly in bipolar I or bipolar II disorder. Hypomanic episodes, but mainly manic episodes, can easily lead to financial extravagance, aggressive–impulsive behavior, or episodic promiscuity, thus generating several interpersonal conflicts, marital breakdown, and new negative life events, all of which have a negative impact on the further course of the illness and on the family atmosphere of the patients as well, which may ultimately trigger a new depressive episode and suicidal behavior. Other relevant stressful events identified as suicide risk factors include interpersonal or occupational difficulties; personal or economic losses; retirement; bereavement; social isolation; feeling stigmatized, as also commonly seen in patients with mood disorders; and limited access to support or health care or social services. Permanent adverse life situations (e.g., unemployment and social isolation), as well as acute psychosocial stressors (e.g., loss events and financial breakdowns) in adult patients with unipolar or bipolar disorders, are useful indicators of suicidality in clinical practice, primarily if other risk factors are also present. It should be noted that unemployment is frequently the consequence of a major mood disorder, particularly in comorbid substance-use disorder cases. The risk of suicide is very high in the first few days of hospitalization and a few days and weeks after discharge from an inpatient psychiatric department, particularly in the case of unplanned discharge, in patients with a short hospital stay, in patients with a high number of previous hospitalizations, and, of course, in the case of lack of follow-up care. Doctors should be alert when a patient discharged from an inpatient or outpatient psychiatric department seeks help for any psychological or even medical problem.

Critical Appraisal of Suicide Risk in Mood Disorders and Clinical Utility of Suicide Risk Factors

A mental state examination is an integral part of the assessment. The suicide risk factors in mood disorders are shown in Table 13.5–1. This table organizes the risk factors from the clinical examination point of view when clinicians collect information about the present clinical manifestations and patients' history with mood disorders. However, these risk factors have different clinical relevance and diverse predictive utility; illness-related factors are more alarming and more powerful than personality-related risk factors. Suicide risk factors have an additive effect; the more risk factors present, the higher the suicide risk.

Understanding the suicidal mind requires knowledge of the individual's perturbed state in crisis since this provides the motivation for the individual to contemplate suicide. Therefore, asking where the suffering comes from and how it has changed and becoming more acute is a method of intervention that, although intuitive and straightforward, is often forgotten by those responsible for managing the person in crisis. Clinicians should interpret the internal debate of the individual in crisis. Patients experience ambivalence about dying, and if clinicians can empathize with the suffering of patients, they can invigorate the wish to live. Patients often have an internal dialogue with ruminations that need to be interrupted with pharmacologic agents, reducing anguish and despair, and other nonpharmacologic treatments. In a phenomenologic view of the suicidal mind, perturbation supplies the motivation for suicide, whereas lethality is the fatal trigger. Suicidologists believe that suicidal individuals feel driven to suicide as an escape from intolerable mental pain and feel that suicide is the only option left. To explain some features of the suicidal mind, the concept of “constriction” is often defined as a type of tunnel vision with which the suicidal individual cannot consider a valid option to reduce the source of suffering. In such a condition, suicidal individuals experience a particular way of thinking, often described as dichotomous. The options considered by the suicidal subject as valid to reduce his suffering are either a magical immediate resolution of his state or a concrete option, which is suicide. It seems that, although there may be effective supports from family and friends, the individual is unable to benefit from them. The pleasant memories and their history concerning others are not helpful, and the individual focuses on intolerable emotions and how to escape from them.

Table 13.5–1.
Clinically Explorable Suicide Risk Factors in Mood Disorders

I. Illness-related risk-factors
   1. Risk factors related to current mood episodes
      a. Severe major depressive episode
         Current suicide attempt, plan, ideation
         Hopelessness, guilt, pessimism, few reasons for living, unusual behavior (i.e., making a will = testament, etc.)
         Major depressive episode with mixed features (mixed depression), agitated depression
         Past mania or hypomania (bipolar I or II diagnosis)
         Insomnia, anxiety
         Appetite and/or weight loss
         Comorbid anxiety/anxiety disorder disorders, substance use disorder, acute alcohol intake, cigarette smoking
         Comorbid personality disorder(s)
         Comorbid disable medical disorder(s)
         Lacking medical care and lacking family/social support
         Noncompliance, treatment-resistance
         First few days/weeks of the treatment (particularly if appropriate care and comedication are lacking) and first few weeks after hospital discharge
      b. Major depressive or manic episode with mixed features
   2. Risk factors related to the prior course of the mood disorder
      Previous suicide attempt/ideation (particularly in the case of violent/highly lethal methods)
      Early-onset/early stage of the illness
      Predominantly depressive course/high proportion spent in major depressive or mixed mood episode, first episode depression
      Rapid cycling course of bipolar I or II disorder

II. Personality-related and historical risk factors
   1. Risk factors related to personality features
      Aggressive/impulsive/pessimistic personality traits
      Cyclothymic, irritable, depressive, anxious affective temperament
   2. Risk factors related to personal history and/or family history
      Early childhood adverse events (parental loss, separation, emotional, physical, and sexual abuse)

Permanent adverse life situations (unemployment, financial problems, isolation, living alone, chronic/disabling medical disorders)

Acute psychosocial stressors (loss events, acute financial catastrophe)

Family history of depressive or bipolar disorders (first- and second-degree relatives)

Family history of suicide and/or suicide attempt (first- and second-degree relatives)

In this perspective, the authors referred to the cessation concept in which the suicidal individual concludes that death will relieve his suffering. During this internal debate, helpers can deliver prevention action as long as they pay attention to both clinical signs and warning signs. It is essential to monitor suicide risk at all times by taking into consideration warning signs for suicide, such as any change in habits, especially if insomnia is presented, and any reference to the wish to die. People may feel trapped and may engage in maladaptive behavior, such as drinking alcohol and using psychoactive substances. Suicidal individuals also often put their affairs in order and give away symbolic items, as if they wish another person would take care of a prized possession, regardless of its economic value.

Table 13.5–2.
Suicidal Risks

- I Ideation—threatened or communicated
- S Substance abuse—excessive or increased
- P Purposeless—no reasons for living, anhedonia
- A Anxiety, agitation, and insomnia
- T Trapped—feeling no way out, perceived burdensomeness
- H Hopelessness
- W Withdrawal—from friends, family, society
- A Anger (uncontrolled)—rage, seeking revenge
- R Recklessness—risky acts, unthinking
- M Mood changes (dramatic)

Source: American Association of Suicidology.

The American Association of Suicidology proposed the acronym "Is Path Warm" (Table 13.5–2) to remind clinicians of potential risks factors.

Clinicians should be aware of communications about suicide; very often, patients talk about suicide, death, and/or having no reason to live. Most suicidal individuals give definite warnings of their suicidal intentions. Still, significant others are either unaware of the significance of these warnings or do not know how to respond to them. However, suicidal intent is sometimes denied or minimized. Clinicians should never forget to ask about suicide; questioning patients about suicidal thoughts and plans are always indicated and crucial for proper prevention. Talking about suicide does not increase the risk. During a suicidal crisis, patients may withdraw from friends and/or social activities, and they may also lose interest in hobbies, work, school, and so on. Suicidal individuals may also prepare for death by writing out a will (unexpectedly) and making final arrangements, giving away prized possessions, clearing up business matters about survivors, accumulating drugs or poisons for taking their life, and so on. Because these warning signs, which are highly indicative of current suicidal danger, are reported mostly by patients' relatives, keeping contact with family members or with significant others is also important. Several suicidal intent scales have been developed that could be useful in special circumstances, but they are routinely used mostly in clinical research. Table 13.5–3 lists some tips for managing a suicide crisis.

Suicide Protective Factors in Patients with Mood Disorders

In contrast to the numerous suicide risk factors, only a few circumstances are known to have a protective effect against suicidality, and suicidal behavior is always the result of the complex interplay between risk and protective factors. Good family and social support, pregnancy and the postpartum period, having a large number of children, holding strong religious beliefs (regardless of the nature of religion), regular physical activity, restricting lethal suicide methods whenever possible (e.g., to reduce domestic and car exhaust gas toxicity, barriers at train stations and bridges), and introducing stricter laws on drug and gun control seem to have some protective effect. Recent studies have found that optimistic personality features and hyperthymic temperament may also be a protective factor and that there is a strong relationship between psychological resilience and hyperthymic temperament in both depressive and healthy individuals. However, one of the most extensively studied and changeable suicide protective factors in major mood disorders is acute and long-term treatment, both pharmacologic and nonpharmacologic.

Table 13.5–3.
Tips for Managing the Suicide Crisis

How to Communicate

- Listen carefully, calmly
- Understand the feelings of the other with empathy
- Emit nonverbal signals of acceptance and respect
- Express respect for the opinions and values of the person in crisis
- Speak honestly and simply
- Express concern, care, and solidarity
- Focus on the feelings of the person in crisis

How not to Communicate

- Interrupt too often
- Express discomfort
- Give the impression of being busy and hasty
- Give orders
- Make intrusive or unclear statements
- Ask too many questions

Useful Questions

- Are you feeling sad?
- Do you feel that no one is taking care of you?
- Do you think it is not worth living?
- Do you think you would like to die by suicide?

Suicide Planning Investigation

- Have you ever made plans to end your life?
- Do you have an idea how to do this?

Investigation of Possible Methods of Suicide

- Do you have drugs, firearms, or other means to commit suicide?
- Are they easily accessible and available?

Investigation Over a Specific Time Frame

- Have you decided when you want to end your life?
- When are you going to do this?

Suicidal behavior, as a final act, is always the result of a dynamic interplay between risk and protective factors. However, in individual cases, suicide risk factors have more practical relevance, and the presence of one or more suicide risk factors is more important than the lack of protective ones. In most cases, more than one suicide risk factor is present simultaneously, and the effect is cumulative: The higher the number of risk factors, the higher the suicide risk. They are usually closely related to each other. For example, older adult major depressives frequently have multiple medical disorders, and they commonly experience loss events and isolation; major depression easily leads to secondary alcoholism, loss of job, family breakdown, financial problems, and so on. When a clinician considers the clinically relevant suicide risk factors in patients with mood disorders, suicidal behavior is predictable but, of course, not guaranteed in the majority of cases. Severe major depression with agitation, anxiety, insomnia, or hopelessness, particularly in combination with a prior suicide attempt and, of course, with current suicidal ideation, are the most important signs for identifying acute or hyperacute suicide risk that requires prompt action.

SUICIDAL RISK EMERGING DURING ANTIDEPRESSANT TREATMENT: THE ROLE OF UNDERLYING BIPOLARITY

Given that suicidality is the most visible outcome of *untreated or inadequately treated* depressive illness, the U.S. Food and Drug Administration (FDA) warning of emergent suicidality in children and adolescents based on the antidepressant arm of placebo-controlled randomized trials has created understandable concern in clinical practice. The issues involved are of broader public health importance for all patients, not just for youngster patients. In 2004, the FDA advised that antidepressants could sometimes be related to suicidal behavior, which mainly pertains to children and adolescents in their database. Newer meta-analyses have confirmed the FDA's position for a slightly increased risk of suicidality (suicidal ideation and attempts but not completed suicide) in juvenile patients in randomized controlled antidepressant trials on unipolar major depression. As in other branches of medicine, psychiatrists must always be vigilant of iatrogenesis's rare risk when prescribing potent agents like antidepressants to patients with depressive disorders where the risk of suicidality is inherent and quite common. They could increase the risk only in a small, vulnerable subpopulation (mostly in young depressives who are predominantly prebipolar); therefore, in everyday clinical practice, the suicide preventive effect of antidepressants highly overshadows this unwanted iatrogenic.

There is no doubt that suicide mortality is much higher among untreated depressives than among antidepressant-treated depressives. The rarely occurring suicidal behavior in patients taking antidepressants is mostly the consequence of the lack of antidepressant effect (antidepressant-associated suicidality) and is rarely the result of the suicide-inducing potential of antidepressants (antidepressant-induced suicidality), as antidepressants can sometimes worsen depression as they can mobilize intradepressive hypomanic symptoms resulting in mixed/agitated clinical picture, frequently called also as “activation syndrome.” It would appear that in highly suicidal individuals, the administration may pose risks, especially when such individuals experience agitation, akathisia, anxiety, irritable dysphoria, insomnia, anger, aggression, or threats of aggression. Patients may also be at higher risk of suicide during antidepressant therapy when experiencing worsening depression, hopelessness, anguish, helplessness, social isolation, or current stressors (loss, separation, scandal, financial crisis). This is probably because these agents do not calm down the anguish and despair, and the effects mentioned above may add consistent sufferance to already fragile subjects. It is also important to emphasize that removing the antidepressant or adding a centrally active agent with mood-stabilizing, sedative, anxiolytic, antimanic, antipsychotic, or antiaggressive activity led to a favorable outcome in most cases. Furthermore, because antidepressants may not be appropriate in some mood disorders, the precise and early diagnosis of bipolar disorder, even in first-episode depressives, and the formal recognition of bipolar depressive mixed states are crucial.

PREVENTION OF SUICIDE IN PATIENTS WITH MOOD DISORDERS

Prevention of suicide in patients with depressive disorders is not only a great challenge for everyone who cares for or has contact with them but could also be among the most reliable indicators of the efficacy as well. As much of the putative psychosocial and demographic suicide risk factors are not modifiable in the frame of individual health care, suicide preventive efforts of psychiatrists and other medical professionals should focus on the treatable and amendable contributory psychiatric disorders involved in such behavior. The health care system's role in suicide prevention varies as a function of the setting (e.g., psychiatric settings vs. primary care practice), with specialized suicide prevention centers (where they exist) playing the most important role. Because suicidal behavior is a complex phenomenon, even in patients with mood disorders, its prevention should also be complex. The majority of suicide attempts and completed suicides in patients with mood disorders are preventable. As suicidal behavior in patients with mood disorders is very rare in the absence of a major depressive or mixed affective episode, reducing the time spent in these high-risk states is essential for their suicide prevention. While successful acute treatment can only prevent suicide connected with the given mood episode, given the highly recurrent nature of major mood disorders, adequate long-term therapy can provide effective protection for a longer period.

In acute suicide danger, the patient needs close observation and urgent hospitalization, even against their own wish. Crisis intervention, whenever needed, is also helpful. After an open discussion with the patient and relatives, involuntary admission is rarely needed. It should be explained that hospitalization will be as short as possible and that it is best for the patient. In cases when acute hospitalization is not indicated, close observation by family members and removing possible means of suicide (i.e., guns, drugs, pesticides, and car keys) is advised, as is consultation with a local outpatient psychiatrist in the GP and other nonpsychiatric practice. As anxiety, psychomotor agitation, and insomnia markedly increase acute suicide risk, prompt pharmacologic treatment of these symptoms is always needed, and these treatments provide immediate results. In the most severe suicidal (sometimes stuporous or catatonic) depressives, electroconvulsive therapy (ECT) is the most useful treatment strategy.

Pharmacologic and Nonpharmacologic Approaches

Large-scale, retrospective, and prospective open-naturalistic and controlled long-term clinical studies, including severely ill, comorbid, frequently suicidal depressives (most often inpatients), have shown that appropriate acute and long-term pharmacotherapy (antidepressants in unipolar depression and mood stabilizers in bipolar disorder) markedly reduces suicide morbidity and mortality—and probable other forms of aggressive, violent behaviors—even in this high-risk population. Bipolar patients on mood stabilizers with or without antidepressants or antipsychotics, and unipolar depressives on antidepressants, show a 70% to 80% (range 56% to 93%) risk reduction of completed and attempted suicide compared to those without pharmacologic treatment. Although long-term treatment with lithium and antiepileptic mood stabilizers results in a similar reduction in suicide mortality, lithium may have some superiority in preventing suicide. Register-based observational cohort studies also show that former inpatients with unipolar major depression who continued treatment with antidepressants, and former inpatients with bipolar disorder who continuously took mood stabilizers (lithium or antiepileptic mood stabilizers), had a markedly decreased rate of completed suicide compared with those who stopped taking antidepressants or mood stabilizers.

This marked antisuicidal potential of lithium seems to be more than the simple result of its **episode-prophylactic effect**, as it has been shown that during the long-term lithium prophylaxis of recurrent bipolar or unipolar affective disorder patients with at least one prior suicide attempt, a significant reduction in the number of suicide attempts was found not only in the excellent responders but also in moderate responders and nonresponders. This finding’s clinical importance is that in the case of lithium nonresponse, when the patient has one or more suicide risk factors, instead of switching from lithium to another mood stabilizer, the clinician should retain lithium (even on a lower dose) and combine it with another mood stabilizer. The well-documented strong antisuicidal effect of lithium in patients with bipolar and unipolar major mood disorder has also been supported by results derived from the general population: Investigating the lithium levels in tap water and drinking water in Texas, Japan, Austria, and Greece, lithium levels were significantly and negatively associated with suicide rates. The evidence that suicide rates among depressed patients have progressively and significantly lowered through the first half of the “pretreatment era” (1900 to 1939), “ECT era” (1940 to 1959), and “antidepressant era” (1960 to 1992) also supports the suicide-preventive effect of antidepressants in depressed patients. As comorbid anxiety, agitation, and insomnia increase the short-term suicide risk for depressed patients, anxiolytics, atypical antipsychotics, and sleep-promoting drugs—mostly for short-term use—are also part of the therapeutic arsenal.

Recent pharmacotherapy innovations pointed to the role of glutamate, the most abundant excitatory neurotransmitter in the brain. Esketamine, a glutamatergic NMDAR antagonist, proved to reduce depressive symptoms faster than typical antidepressants. Esketamine blocks NMDA receptor-mediated activation of specific GABA interneurons. GABA interneurons are nerve cells that are inhibitory in function; they dampen the release of glutamate. Esketamine-mediated inhibition of these GABA interneurons could enhance the release of glutamate and, hence, greater glutamate action on AMPA receptors.

Such enantiomer of ketamine was then approved in combination with oral antidepressants to treat the major depressive disorder with suicidal ideation or behavior or in some geographical regions to treat major depressive disorder with psychiatric emergencies. Studies demonstrated efficacy for improvement of depression after 24 hours from first administration but not for suicide risk *per se*, probably due to extra clinical care also delivered to the placebo plus oral antidepressant arm.

Pharmacotherapy, however, is a necessary but insufficient method of reducing suicidal behavior in depressed patients, and psychosocial interventions are always needed. Psychoeducation and supportive psychotherapy are also always needed. There is some evidence that concurrent depression-focused psychotherapies, combined with pharmacotherapy, improve patients' adherence, increase pharmacotherapy effectiveness, and contribute to suicide prevention for severe unipolar or bipolar disorders. The interaction between pharmacotherapy and psychosocial interventions is quite complex. It has been reported, for example, that successful episode-preventive medication with mood stabilizers in bipolar patients counteracted dysfunctional cognitions (including lowered self-esteem), and adjunctive cognitive therapy could help to optimize the long-term course of bipolar illness.

However, psychotherapy for depression *alone* is not enough for preventing depression-related suicide. A systematic review and meta-analysis show that there is little evidence for the assumption that suicidality in depressed patients can be reduced with psychotherapy for depression despite decreasing depressive symptomatology, indicating that in suicidal depressives, psychotherapy and psychosocial interventions should always be combined with appropriate acute (and, if necessary, long-term) pharmacotherapy. Concurrent depression-focused psychotherapies, combined with pharmacotherapy, improve patients' adherence, increase pharmacotherapy effectiveness, and contribute to suicide prevention for severe recurrent unipolar or bipolar disorders. Both hopelessness–pessimism and aggressiveness–impulsiveness may be amenable to cognitive-behavioral therapy and pharmacotherapy (e.g., selective serotonin reuptake inhibitor [SSRI] antidepressants, lithium, and other mood stabilizers).

Suicide Prevention in Primary Care

Although more than two-thirds of suicide victims contact different levels of health care (mostly in General Practice (GP) and psychiatrists) during the last few weeks or months before their death, the rate of recognition of depression and adequate pharmacotherapy among depressed suicide victims is around 20% to 25%, which is disturbingly low. While the current prevalence of major mood disorders in the primary care practice is around 8% to 12%, many depressed patients are not recognized by their GPs. However, more recent studies reported much higher recognition and treatment of depression in primary care practice (62% to 85%), and 33% to 50% of them were treated with antidepressants. Physician factors related to poor recognition of depression include lack of experience, insufficient or suboptimal knowledge about symptoms, treatment and good prognosis in treated depression, prejudices about mental illness, lack of postgraduate psychiatric training, insufficient interview skills, lack of cooperation with psychiatrists, and a low level of empathy.

Although suicidal behavior is a relatively rare event in primary care, given that depression is quite frequent in GP and is very common among completed suicides, major depression, particularly in combination with past or current suicidal behavior, should be taken very seriously in primary care. Compared to nonsuicidal patients, suicide victims visit their GPs much more frequently in the last four weeks of their life. Many physicians and even mental health care professionals avoid discussing suicide directly and frankly with patients for fear of provoking suicidal behavior or, more likely, because of personal discomfort. However, asking questions about suicidal ideation and past suicide attempts does not trigger suicide. This is particularly true if such a discussion is accompanied by some sentences explaining that depressive disorders can be successfully treated and that suicidal intent will vanish after (or even before) the recovery from depression. Asking simple questions (e.g., “What do you think about the future”? “Do you feel that life is not worth living”?) can easily facilitate further, deeper, and honest discussion on the topic of suicide. This is beneficial, as many patients think they are alone or unique in their suicidal ideas. Leaflets, posters, and fliers left in the waiting room and providing appropriate websites indicating the symptoms, dangers, and treatable nature of depression and preventable nature of suicide are also beneficial.

Postgraduate training on better recognition and management of depressive disorders among psychiatrists and primary care appears to prevent suicide. The first example of the significant role of GPs in suicide prevention comes from the Swedish Gotland Study. The study demonstrated that after a short intensive postgraduate training for GPs on Gotland’s Island on the diagnosis and treatment of depression, suicide rates and the number of depression-related hospital admissions declined significantly, and antidepressant prescription increased markedly a few years after the training. The rate of depressive suicides among all suicides decreased substantially after the training, indicating that the decline in suicide mortality after the education resulted directly from a robust decrease in depressive suicides and suggesting that this result might not be caused by random fluctuation of other unknown factors. However, the decline in depressive suicides after the GP training mainly resulted from a reduction in female depressive suicides. Few suicidal males were known to the local medical services, although many were known to the police and social welfare services.

However, primary care education in isolation does not have any significant long-term effect on suicide mortality, and primarily complex, permanent, and multilevel educational and organizational interventions are useful. The multilevel community program should be based on the involvement and the strong collaboration with psychiatrists, GPs, gatekeepers, and media. The effect of these multilevel community suicide prevention programs has been demonstrated recently by the German Nuremberg Alliance Against Depression Study and the International European Alliance Against Depression Project, a multifaceted community-based action program against depression and suicide. Better management of depression requires improved recognition and treatment skills from the doctors and good compliance. Nonadherence to antidepressant therapy is one of the most common causes of treatment failure. GPs should work in close and permanent collaboration with the local mental health services. Outpatient psychiatric consultation is also helpful in differential diagnostic problems, treatment resistance, and comorbid substance-use disorder, regardless of whether the patient is suicidal. If long-term prophylactic pharmacotherapy is needed (bipolar disorder, recurrent unipolar major depression), the GP may direct the patient to a psychiatrist to optimize the therapy. Although depressive disorders represent an important target for suicide prevention in old age, there should be increased attention also for psychosocial and somatic–medical risk factors in this age range.

Regular aftercare with fixed appointments is highly recommended, particularly for patients with previous suicide attempts. Psychological support should be available for each suicidally depressed patient. This is important because the actual clinical picture immediately after a suicide attempt is often misleading due to the cathartic effect of self-aggression, resulting in a short-lived but sometimes marked improvement in depression. This can also serve as one of the explanations for why some health care workers misinterpret suicide attempts as manipulative acts.

Health care professionals, of course, are unable to prevent all suicides, including those occurring in the context of patients with mood disorders. Nevertheless, the current theoretical knowledge and the available treatment and preventive strategies are sufficient to prevent many, probably most, of them. The most important methods of suicide prevention in patients with mood disorders are shown in Table 13.5–4. However, as health and social care workers can only help those who contact them, public education media campaigns on symptoms, dangers, treatability, and referral pathways are vitally important.

Table 13.5–4.
Suicide Prevention Strategies in Patients with Mood Disorders

I. Elimination of acute suicide danger (physical inhibition, emergency hospitalization, sedation, anxiolysis, crisis-intervention)

II. Improving the early diagnosis and treatment of mood disorders with particular regard to the soft (subthreshold) clinical manifestations of bipolar disorder

1. Education of health care workers, patients, relatives, and gatekeepers (clergymen, policemen, teachers, social workers, peer-helpers, etc.)
2. Adequate acute and long-term treatment/aftercare (pharmacotherapy, nonpharmacologic interventions such as psychoeducation, psychotherapy, cognitive therapy, family counseling/family therapy, regular long-term care/contact, etc.)

III. Improving the patients' compliance (psychoeducation, psychotherapy, cognitive therapy, etc.)

IV. Education of the public via printed and electronic media, including the internet

1. Education on the symptoms, dangers, and treatable nature of mood disorders and the preventable nature of suicidal behavior
2. Reducing the stigma against mood disorders and suicide
3. Providing information on how and where to get help in the case of mood disorder and suicide crisis

FURTHER READINGS

Akiskal HS. Targeting suicide prevention to modifiable risk factors: has bipolar II been overlooked? *Acta Psychiatr Scand*. 2007;116:395–402.

Akiskal HS, Benazzi F. Psychopathologic correlates of suicidal ideation in major depressive outpatients: is it all due to unrecognized (bipolar) depressive mixed states? *Psychopathology*. 2005;38:273–280.

Angst J, Cui L, Swendsen J, et al. Major depressive disorder with subthreshold bipolarity in the National Comorbidity Survey Replication. *Am J Psychiatry*. 2010;167:1194–1201.

Bostwick JM, Pankratz VS. Affective disorders and suicide risk: a reexamination. *Am J Psychiatry*. 2000;157:1925–1932.

Brent DA, Melhem NM, Oquendo M, et al. Familial pathways to early-onset suicide attempt: a 5.6-year prospective study. *JAMA Psychiatry*. 2015;72:160–168.

Cuijpers P, de Beurs DP, van Spijker BA, Berking M, Andersson G, Kerkhof AJFM. The effects of psychotherapy for adult depression on suicidality and hopelessness: a systematic review and meta-analysis. *J Affect Disord*. 2013;144:183–190.

Fawcett J. Depressive disorders. In: Simon RI, Hales RE, eds. *The American Psychiatric Publishing Textbook of Suicide Assessment and Management*. 2nd ed. American Psychiatric Publishing; 2012;109–123.

Goodwin FK, Jamison KR. *Manic Depressive Illness. Bipolar Disorders and Recurrent Depression*. Oxford University Press; 2007.

Haukka J, Suominen K, Partonen T, Lönqvist J. Determinants and outcomes of serious attempted suicide: a nationwide study in Finland; 1996–2003. *Am J Epidemiol*. 2008;167(10):1155–1163.

Hawton K, Casañas I Comabella C, Haw C, Saunders K. Risk factors for suicide in individuals with depression: a systematic review. *J Affect Disord*. 2013;147(1–3):17–28.

Isacsson G, Ahlner J. Antidepressants and the risk of suicide in young persons—prescription trends and toxicological analyses. *Acta Psychiatr Scand*. 2014;129(4):296–302.

Isometsa E. Suicidal behaviour in mood disorders—who, when, and why? *Can J Psychiatry*. 2014;59(3):120–130.

Nordentoft M, Mortensen PB, Pedersen CB. Absolute risk of suicide after first hospital contact in mental disorder. *Arch Gen Psychiatry*. 2011;68(10):1058–1064.

O'Connor RC, Nock MN. The psychology of suicidal behaviour. *Lancet Psychiatry*. 2014;1:73–83.

Pisani AR, Murrie DC, Silverman MM. Reformulating suicide risk formulation: from prediction to prevention. *Acad Psychiatry*. 2016;40(4):623–629.

Pompili M. Critical appraisal of major depression with suicidal ideation. *Ann Gen Psychiatry*. 2019;18:7.

Pompili M, Erbuto D, Innamorati M, et al. The relationship between mental pain, suicide risk, and childhood traumatic experiences: results from a multicenter study. *J Clin Psychiatry*. 2022;83(4):21m14176.

Pompili M, Gonda X, Serafini G, et al. Epidemiology of suicide in bipolar disorders: a systematic review of the literature. *Bipolar Disord*. 2013;15(5):457–490.

Pompili M, Innamorati M, Erbuto D, et al. High depression symptomatology and mental pain characterize suicidal psychiatric patients. *Eur Psychiatry*. 2022;65(1):e54.

Pompili M, Innamorati M, Rihmer Z, et al. Cyclothymic-depressive-anxious temperament pattern is related to suicide risk in 346 patients with major mood disorders. *J Affect Disord*. 2012;136(3):405–411.

Pompili M, Rihmer Z, Innamorati M, Lester D, Girardi P, Tatarelli R. Assessment and treatment of suicide risk in bipolar disorders. *Expert Rev Neurother*. 2009;9:109–136.

Pouliot L, De Leo D. Critical issues in psychological autopsy studies. *Suicide Life Threat Behav*. 2006;36(5):491–510.

Rihmer Z. Depression and suicide and attempted suicide. In: O'Connor R, Platt S, Gordon J, eds. *International Handbook of Suicide Prevention: Research, Policy and Practice*. Wiley Blackwell; 2011.

Rihmer Z, Gonda X. Pharmacological prevention of suicide in patients with major mood disorders. *Neurosci Biobehav Rev*. 2013;37(10 Pt 1):2398–2403.

Rihmer Z, Rutz W. Treatment of attempted suicide and suicidal patients in primary care. In: Wasserman D, Wasserman C, eds. *Oxford Textbook of Suicidology and Suicide Prevention*. Oxford University Press; 2009.

Schaffer A, Isometsa ET, Tondo L, et al. International society for bipolar disorders task force on suicide: meta-analyses and meta-regression of correlates of suicide attempts and suicide deaths in bipolar disorder. *Bipolar Disord*. 2015;17:1–16.

Shneidman ES. Suicide as psychache. *J Nerv Ment Dis*. 1993;181:145–147.

Tondo L, Pompili M, Forte A, Baldessarini RJ. Suicide attempts in bipolar disorders: comprehensive review of 101 reports. *Acta Psychiatr Scand*. 2016;133(3):174–186.

Turecki G, Brent DA, Gunnell D, et al. Suicide and suicide risk. *Nat Rev Dis Primers*. 2019;5(1):74.

13.6 Mood Disorders: Pharmacologic Treatment of Depression and Bipolar Disorders

Lakshmi N. Yatham, M.B.B.S., F.R.C.P.C.S., M.R.C.Psych(UK), M.B.A. (Exec), Sidney H. Kennedy, M.D., F.R.C.P.C., F.R.C.Psych., Ayal Schaffer, M.D., F.R.C.P.C., and Wegdan Rashad, M.D.

Mood disorders comprise two separate chapters in the DSM-5: “Depressive Disorders” and “Bipolar and Related Disorders.” Among the major changes, mixed episode as a diagnostic entity was eliminated and a course specifier “with mixed features” has been added for both major depressive episodes (MDEs) and manic/hypomanic episodes. This allows an individual in an MDE with three or more (hypo)manic symptoms and no prior manic/hypomanic episodes to remain within the MDD domain, rather than be classified as a forme fruste of bipolar disorder.

Other changes within the Depressive disorders chapter are the inclusion of disruptive mood dysregulation disorder (DMDD) and premenstrual dysphoric disorder as new diagnoses, and the replacement of chronic depression and dysthymia with a single diagnostic entity—persistent depressive disorder (PDD). The essential feature of PDD is depressed mood and at least two depression symptoms present for 2 years or more (1 year for children or adolescents). There is still the possibility that MDEs may occur intermittently with PDD (i.e., PDD with intermittent MDE) or persistent MDEs may occur with PDD (i.e., PDD with persistent MDE). Abnormally and persistently increased goal-directed activity or energy has been added to the stem criterion for a manic/hypomanic episode. Further, manic/hypomanic episodes emerging during antidepressant treatment are indicative of bipolar disorder provided the symptoms persist beyond the physiologic effects of treatment. The not otherwise specified (NOS) category has been replaced with “other specified depressive disorder,” “other specified bipolar and related disorder,” and “unspecified bipolar and related disorder.” Unfortunately, there is a limited evidence base from clinical studies to inform treatment recommendations for these additional diagnostic categories. Therefore, the focus of this chapter will be on the treatment of MDD and bipolar I and II disorder.

TREATMENT OF MAJOR DEPRESSIVE DISORDER—FOCUS ON PSYCHOPHARMACOLOGY

MDD is typically a disorder of recurrent depressive episodes rather than a single episode. Several treatment options are available for managing patients with MDD and the list of treatment options continues to expand, although progress in identifying novel strategies has been slow. This section will outline the foundations of management for MDD from a psychopharmacologic perspective.

Since MDD is not a homogeneous entity, even in the absence of co-occurring psychiatric disorders, there is major symptom heterogeneity among persons presenting with an MDE: hence, diverse treatment modalities have a place in the management of MDE. The reader is referred to other chapters in this book for the management of MDD at different phases across the lifespan: pregnancy and postpartum, and the management of MDD in children and adolescents as well as older individuals.

Foundations of Management

Core Concepts. A thorough biopsychosocial assessment, including attention to safety, is the nucleus to quality care. In the psychiatric assessment, it is important to consider other diagnostic possibilities such as bipolar disorder, and complete a full characterization of clinical aspects that may include anxiety disorders, substance use, suicidality, psychotic features, and other psychiatric comorbidities. Reports on psychiatric comorbidity indicate that anxiety disorders (60%), substance use disorders (25%), and impulse control disorders (30%) are highly prevalent among patients with lifetime MDD diagnosis. Information on previous psychiatric treatments and sensitivity to psychopharmacology is also required.

An assessment of physical health, functionality, sociodemographic, and family history as well as concurrent prescription medications can guide the treatment plan. Collecting collateral information is valuable at this stage.

A medical workup should focus on patient symptoms and evidence of change. There is also a bidirectional relationship between MDD and physical health. Individuals with prior depressive episodes have high risk of medical disorders including heart disease and diabetes, while patients with medical disorders, including cardiovascular disease and stroke, chronic pain, obesity, diabetes, epilepsy, and other central nervous system (CNS) disorders, and cancer have two- to fourfold increases in risk of MDD.

The next step includes formulating a diagnosis and differential diagnosis followed by a dialogue on treatment options and recommendations. It is important to establish a therapeutic alliance and to engage in a dialogue on perceptions about psychiatric treatments and antidepressants, in particular. The psychiatrist often plays the role of an educator and "myth buster." Many patients will also have searched the Internet for information and will welcome a chance to discuss the materials they have collected online: it is important to explore these perceptions and often misperceptions. These discussions are likely to recur during ongoing supportive care and it is important to establish good rapport at the initial visit.

Effective management of MDD requires a focus on both acute and maintenance aspects of the disorder from the outset (see Table 13.6–1). It is imperative to recognize the patient as a partner in determining goals of treatment. The primary goal in the acute phase is to achieve remission—a state that comes close to being symptom-free (typically remission represents a score of ≤7 on the Hamilton Rating Scale for Depression [HAM-D] or ≤10 on the Montgomery-Asberg Depression Rating Scale [MADRS]). Since failure to achieve remission is associated with a higher risk of recurrence, this has replaced the previously accepted treatment aim of response, defined as a reduction of ≥50% on one of the above scales as a goal of treatment.

Although symptomatic response or remission are the traditional goals of treatment, it is increasingly recognized that improvement in function (e.g., return to work or other meaningful activities) and overall quality of life are the priority goals of most patients: it is, therefore, important to listen to their apprehensions and expectations and provide education about depression and treatment options.

Evaluate Adherence. The first step in exploring minimal or no response is evaluating adherence, which is a pivotal component of any therapy. Unfortunately, many patients neglect to take their medication on a frequent basis, missing doses for days or weeks, or discontinuing altogether. This increases the risk of relapse or recurrence. “Forgetting” to take medication as prescribed is more likely to occur as patients begin to feel better and about 50% admit to decreased adherence by 3 months. On the positive side, there is evidence that adherence improves in the weeks leading up to and just after a follow-up appointment. Although not currently a routine practice, plasma monitoring of antidepressant blood levels provides an additional method to assess adherence. There is also evidence that a brief telephone contact to monitor symptoms and side effects early in the course of treatment improves adherence and ultimately outcome. Strategies for improving adherence include fostering a stronger therapeutic alliance, patient education, teaching self-management, and harnessing digital tools.

Table 13.6–1.
Phases of Antidepressant Treatment

Treatment Phase	Duration	Goals	Activities
Acute and continuation	8–12 weeks	▶ Achieve symptomatic remission ▶ Monitor side effects ▶ Restore function	▶ Establish therapeutic alliance ▶ Provide psychoeducation ▶ Select optimal antidepressant treatment(s) ▶ Apply supportive and measurement-based care ▶ Monitor progress
Maintenance	6–24 months or longer	▶ Return to full function and quality of life ▶ Prevention of recurrence	▶ Continue psychoeducation ▶ Rehabilitate ▶ Manage comorbidities ▶ Monitor for recurrence

Modified from Lam RW, McIntosh D, Wang J, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: section 1. Disease burden and principles of care. *Can J Psychiatry*. 2016;61(9):510–523.

Self-management techniques may help patients to understand and manage their symptoms and psychosocial sequelae. Behavioral activation techniques, communication skills, coping with difficult emotion, engaging in healthy lifestyle habits (integrating nutrient-dense foods and exercise), relapse prevention planning, skill development, and self-monitoring. Through these skills, individuals are less reliant on health care providers and in turn are more empowered and self-efficacious. Patients may be helpful in attending peer support services.

Monitor Outcomes with Measurement-Based Care. Implementing measurement-based care requires routine use of standard rating scales to identify and track changes in symptoms and function during an MDE. The advantages of administering a rating scale include longitudinal monitoring and promoting a therapeutic alliance to achieve better outcomes.

Several clinical instruments are readily available and completion times range from 5 to 20 minutes depending on the scale length. Abbreviated versions such as the HAMD-7 item scale are also available. The Patient Health Questionnaire (PHQ-9), a self-report scale containing the nine items listed in DSM-5, rated according to their frequency of occurrence and the Quick Inventory for Depression Symptomatology (QIDS), available in both observer-rated and self-report forms are among the most frequently used scales. Early reduction of at least 20%–in severity on a rating scale predicts greater likelihood of response at 6 to 8 weeks and failure to achieve a 20% reduction should trigger dose increase or other options (switching or augmentation). The GAD-7 compliments these depression scales in assessing anxiety. Two easy-to-administer scales that evaluate functional outcomes are the Sheehan Disability Scale (SDS) and the Lam Employment Absence and Productivity Scale (LEAPS). In this era of remote care, these scales can be completed on electronic personal devices and may be combined with objective data such as sleep–wake patterns, body temperature, or physical activity.

Digital Tools. There is burgeoning evidence that internet and digital applications on computers and mobile devices can improve access to care, care quality, and clinical outcomes. E-mental health applications are widely available for screening, assessment, and monitoring as well as platforms for self-management and psychotherapy. However, clinicians should be aware of the potential harms of using and recommending e-mental health applications and become familiar with attention to validity, privacy, and practical aspects. Websites that rate and rank digital applications are increasingly available including appraisal frameworks such as the American Psychiatric Association App evaluation model.

After addressing goals of treatment, a comprehensive management plan should be outlined including safety, social supports, adherence, and clinical follow-up. Regular follow-up appointments are important, particularly in the first 2 to 3 months. The first follow-up visit should generally be 2 weeks after initiating treatment, although high-risk patients with suicidal intent or psychotic features should be seen sooner and more frequently. At each visit, the psychiatrist should review adherence, symptom severity and change, side effects, and function, as well as evaluate psychosocial stressors and supports.

Most patients who achieve remission will do so by 6 to 8 weeks, although a minority may achieve early response (within the first 4 weeks). Typical rates of remission are between 30% to 40% with most antidepressants and response rates are generally in the 50% to 65% range, for patients who are not considered treatment resistant.

Sequencing and Staging of Treatment. Approximately half of individuals with MDD do not respond to their first prescribed antidepressant treatment, and 20% to 30% have not responded after at least two treatment trials. This introduces the possibility of treatment-resistant depression (TRD) defined as failure to respond to at least two antidepressant trials, adequate in dose and duration. To better define and manage TRD, several staging models have been proposed.

Antidepressant Treatment Strategies

Recognize First-, Second-, and Third-Line Approaches. Selecting an antidepressant requires a balance between physician expertise and patient preference, considering symptom profile, past history of treatment response and tolerability, potential for drug–drug interactions, and in many instances, cost. First-line antidepressants include selective serotonin reuptake inhibitors (SSRIs), SNRIs, other second-generation, and novel antidepressants (see Table 13.6–2). These agents have comparable efficacy but superior tolerability to the first-generation MAOI and tricyclic antidepressant (TCA) agents, which are considered third-line treatments for MDD. Second-line options include novel antidepressants and adjunctive treatments. The sequenced treatment alternatives to relieve depression (STAR*D) remains the largest effectiveness trial to directly evaluate remission rates among sequential monotherapies and adjunctive antidepressant treatment strategies in MDD. These findings are complemented by two network analyses by Cipriani and colleagues in 2009 and 2018. Combining knowledge from these trials with course specifiers for individual patient episodes of depression can help to refine treatment selection.

Table 13.6–2.
Recommendation Level and Drug Class with Dose Range for Treatment of Major Depressive Disorder

Recommendation Level	Drug Class	Recommended Dose Range (mg) per day
1	SSRI	Varies
1	SNRI	Varies
1	Agomelatine[a]	25–50
1	Bupropion	150–450
1	Mianserin[a]	60–120
1	Mirtazapine	15–60
1	Vortioxetine	10–20
2	Tricyclic Antidepressants	Varies
2	Moclobemide[b]	300–600
2	Quetiapine	150–400
2	Selegiline transdermal	6–12
2	Trazodone	150–300
2	Vilazodone	20–40
3	Phenelzine	45–90
3	Reboxetine[a]	8–10
3	Tranylcypromine	20–60

[a] Not available in North America.

[b] Not available in the United States, available in Canada.

Adapted from Vaccarino SR, Kennedy SH. Treatment-resistant depression. In: Vazquez GH, Zarate CA, Brietzke E, eds. *Ketamine for Treatment-Resistant Depression: Neurobiology and Applications*. 1st ed. Academic Press; 2021, 33–84.

COMPARATIVE ANTIDEPRESSANT STUDIES. In STAR*D, medications from each of the above drug classes were administered over four stages, starting with approximately 3,000 participants who received citalopram in

Phase I. Based on the QIDS Self Report scale, 33% of those in Phase I achieved remission, and 47% achieved response. Those who did not achieve remission in Phase I were allocated to Phase II, a comparison of switching and adjunctive treatments: to another SSRI (sertraline); an SNRI (venlafaxine); bupropion, or CBT in which remission rates ranged from 18% (sertraline switch) to 30% (buspirone augmentation). Despite the large initial sample size, the number of subjects evaluated in the third and fourth phases was substantially smaller, and it is difficult to draw conclusions about the relative effectiveness of TCA, MAOI, and other adjunctive strategies.

Overall, STAR*D highlighted the relatively low rate of remission (approximately 30%) following an adequate trial of citalopram. In addition, the study did not support conventional thinking that switching to a different drug class (e.g., from SSRI to SNRI) following failure with an SSRI was superior to switching within the same class. There was also no difference in outcomes between switching and adjunctive treatment strategies. While the time to achieve remission was longer in those who received CBT, the overall rates of remission did not differ significantly from pharmacotherapy strategies. STAR*D also highlighted the marked reduction in remission rates after two consecutive trials and the high rates of dropout after each phase.

The first network meta-analysis examined 12 second-generation antidepressants from 117 published trials that collectively included almost 26,000 participants. To incorporate all trials, the authors accepted different severity rating measures (HAM-D, MADRS, and Clinical Global Impression) with “response” as the primary efficacy outcome. The percentage of patients dropping out was a proxy for intolerability. Escitalopram, mirtazapine, sertraline, and venlafaxine were ranked as the top four antidepressants for efficacy, although escitalopram and sertraline were preferred when tolerability and efficacy rankings were combined. While sample size was a major strength in this study, methodologic differences across trials may have influenced conclusions. Other studies suggest comparable efficacy between bupropion and SSRIs, as well as superiority for escitalopram compared to other SSRIs. However, many have argued that differences among first-line antidepressants are minimal. A limitation of this trial was the absence of a placebo comparison.

The subsequent meta-analysis by the same investigators involved nearly 116,500 participants. They compared the efficacy and acceptability of 21 antidepressants with both placebo-controlled and head-to-head studies. All antidepressants were more effective than placebo. Regarding acceptability, only agomelatine and fluoxetine were associated with fewer dropouts than placebo. In head-to-head studies, agomelatine, amitriptyline, escitalopram, mirtazapine, paroxetine, venlafaxine, and vortioxetine were more effective than other antidepressants, whereas fluoxetine, fluvoxamine, reboxetine, and trazodone were the least efficacious drugs.

COURSE SPECIFIERS FOR DEPRESSIVE EPISODES. Course specifiers for depressive episodes also provide some guidance when selecting a medication. In the case of anxious distress (tense, restless, fearful, etc.), it would be preferable to select an antidepressant with established efficacy in the treatment of generalized anxiety disorder (GAD). There is also evidence that patients who display melancholic features (specifically anhedonia and lack of reactivity) may have a different response to current first-line antidepressants (e.g., SNRI and SSRI) compared to those with nonmelancholic features. Historically, patients with atypical features (preservation of mood reactivity, hypersomnia, hyperphagia, weight gain, and a heightened sensitivity to interpersonal rejection) were thought to respond better to MAOIs, but it appears that current first-line antidepressants are equally effective in patients with atypical features. Two specifiers that are most likely to influence treatment selection are "with seasonal pattern," where light therapy is recommended, at least as an adjunctive treatment, and with "mood congruent or mood incongruent delusions," where either an antidepressant/antipsychotic combination therapy or ECT is recommended.

An additional specifier in DSM-5 is MDD with mixed features. Currently, it is unclear whether patients within this subtype will continue to meet criteria for MDD or switch to bipolar disorder. Lurasidone is an appropriate first-line therapy currently and has produced improvement in both depressive and hypomanic symptoms in patients with MDD with mixed features.

SELECTING A FIRST-LINE ANTIDEPRESSANT. Generally, clinicians should select a first-line agent from the SSRI or SNRI classes. While overall efficacy is likely similar among these agents, differences in their receptor pharmacology and the adverse event profile may influence treatment selection. For instance, the SSRI antidepressants share blockade of the serotonin (5HT) transporter as their common mechanism of action, but they differ in their actions on additional receptors. Differences among SNRI antidepressants are even more distinct. Venlafaxine, duloxetine, and desvenlafaxine preferentially block the 5HT transporter, while levomilnacipran predominantly blocks the norepinephrine (NE) transporter. Among other first-line antidepressants, differences in transporter and receptor affinities are also important, for example, activating effect of bupropion and sedating effect of mirtazapine. These differences allow clinicians to tailor the selection of antidepressants for individual patients and increase acceptability.

Some second-generation antidepressants with unique mechanisms of action also have noradrenergic effects. Bupropion is both an antidepressant and a medication for smoking cessation with predominantly noradrenergic and dopaminergic effects. It is an activating antidepressant; hence, it is frequently prescribed in combination with a more sedating 5HT-enhancing antidepressant. Mirtazapine promotes release of both 5HT and NE through blockade of α2-noradrenergic receptors. Its sedating profile is often beneficial for patients with depression and severe insomnia. Mirtazapine is also associated with weight gain. Trazodone is another antidepressant with dual effects on 5HT and NE receptors. Because of its highly sedating properties, it is more often prescribed at subtherapeutic antidepressant doses (50 to 100 mg) as a hypnotic. Nefazodone has a similar mode of action to trazodone with additional NE transporter-blocking effects. Despite a favorable efficacy profile, reports of life-threatening hepatotoxicity have resulted in the virtual discontinuation of its use. Moclobemide is a reversible inhibitor of MAO-A that may be prescribed to treat MDD without dietary precautions, although it is not available in the United States. For a list of additional side effects of antidepressants discussed above, see Table 13.6–3.

Another major consideration when choosing an antidepressant is the side effect profile of the drug and patient sensitivity to side effects. Gastrointestinal, CNS, and sexual side effects are particularly common with SSRIs, while patients with pre-existing hypertension may be more prone to further elevation in blood pressure with an SNRI such as venlafaxine. Prolongation of QTc interval has also been noted at higher doses with various antidepressants, including citalopram. The relative frequencies of side effects associated with SSRIs and SNRIs are listed in

Table 13.6–3. Clinicians need to be vigilant about the emergence of hypo(manic) symptoms during treatment with antidepressants, particularly in patients with a family history of bipolar disorder and in those who display reverse vegetative symptoms as a part of an MDE. It is important to highlight these risk factors in such patients. The risk of antidepressant-induced mood elevations in patients who meet criteria for an MDE with mixed features in the context of MDD is likely higher, and hence such patients also need to be educated and monitored closely.

While emergent suicidal ideation is of greatest concern in adolescents and young adults during antidepressant clinical trials resulting in “black box warnings” about prescribing SSRIs and other antidepressants to this age group, suicidal ideation should also be a consideration in ongoing assessment of MDD in adults. However, the relationship between increased suicidal ideation and completed suicide remains controversial. Other rare but serious adverse effects of antidepressants include the serotonin syndrome, increased upper gastrointestinal bleeding (particularly when antidepressants and nonsteroidal anti-inflammatory agents are coprescribed); hyponatremia has also been of concern, especially in older adult patients with depression, and most antidepressants, particularly bupropion reduce the seizure threshold.

NOVEL ANTIDEPRESSANTS. Several newer antidepressants exhibit serotonin multimodal actions. Vilazodone is an SSRI with additional 5HT-1A partial agonist effects, providing the effect of combining a 5HT-1A partial agonist, such as buspirone, with an SSRI. Favorable aspects of the drug include low rates of sexual side effects, lack of cardiac arrhythmias, and minimal weight gain. Vortioxetine is also a 5HT transporter blocker, with additional receptor effects including 5HT3 and 5HT7 antagonism, as well as 5HT1A and 5HT1B agonism. During clinical trial development, measures of cognition including executive function, verbal learning, memory, and processing speed were evaluated and found to be significantly improved.

Table 13.6–3.

Prevalence of Side Effects among Selective Serotonin-Reuptake Inhibitors and Serotonin- and Norepinephrine-Reuptake Inhibitors

Drug Class	Recommended Dose Range (mg)	Side Effect Frequency	
		10–30%	>30%
SSRI			
Citalopram	20–40	Nausea, dry mouth, sweating	None
Escitalopram	10–20	Male sexual dysfunction and nausea	None
Fluoxetine	20–60	Nausea, dry mouth, somnolence, nervousness, anxiety, insomnia, tremor, anorexia	None
Fluvoxamine	100–300	Dry mouth, headaches, somnolence, agitation, insomnia, sweating, tremor, anorexia, dizziness, constipation	Nausea
Paroxetine	20–60	Nausea, diarrhea, dry mouth, headaches, somnolence, insomnia, sweating, asthenia, male sexual dysfunction, dizziness	None
Sertraline	50–200	Nausea, diarrhea, dry mouth, headaches, somnolence, insomnia, fatigue, tremor, male sexual dysfunction, dizziness	None
SNRI			
Venlafaxine	75–375	Headaches, somnolence, dry mouth, dizziness, nervousness, insomnia, sweating,	Nausea

		male sexual dysfunction	
Desvenlafaxine	50–100	Dry mouth, dizziness, nausea, sweating	None
Duloxetine	30–120	Nausea, dry mouth, constipation, insomnia, male sexual dysfunction	None
Levomilnacipran	20–80	Nausea, dry mouth, headaches, male sexual dysfunction	None
Other Second-Generation and Novel Antidepressants			
Agomelatine[a]	25–50	Dizziness	None
Bupropion	150–450	Insomnia, dry mouth, nausea	Headaches
Mirtazapine	15–60	Dry mouth, constipation, increased appetite, weight gain	Somnolence
Moclobemide[a]	300–600	None	None
Vilazodone	10–40	Diarrhea, nausea, headaches	None
Vortioxetine	10–20	Nausea	None

[a] Not available in the United States.

Data from Kennedy SH, Lam RW, McIntyre RS, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 Clinical Guidelines for the management of adults with major depressive disorder: section 3. pharmacological treatments. *Can J Psychiatry*. 2016;61(9):540–560; Kennedy SH, Rizvi SJ. Chapter 246: SSRIs and related compounds. In: Stolerman I, ed. *Encyclopedia of Pharmacology*. Springer; 2010.

Agomelatine is a melatonin 1 and 2 receptor agonist. Additional 5-HT2C antagonist effects facilitate release of dopamine in the prefrontal cortex and may account for its favorable effect on anhedonia and minimal sexual side effects. This drug was also associated with elevation of hepatic enzymes and monitoring of liver function is recommended.

SECOND-LINE APPROACHES TO TREATMENT (OPTIMIZATION, SWITCHING, AND ADJUNCTIVE THERAPIES). When a patient has received at least 4 to 6 weeks of treatment with a first-line antidepressant and has shown no improvement despite a dose increase, the following options should be considered:

- further **dose increase**, particularly when the medication appears to be well tolerated;
- **switching** to a different antidepressant within the same or a different class; and
- **adjunctive therapy** with an established agent such as an atypical antipsychotic, lithium, triiodothyronine, or a second antidepressant with a different neurotransmitter action (e.g., bupropion added to an SSRI).

Invariably, optimization is the first step, particularly when the drug is well tolerated, is at a relatively low therapeutic dose or if past treatment history suggests tolerance for high doses of other medications. Extending the initial trial beyond 4 to 6 weeks is indicated if there is an incremental but slow improvement during the first 4 weeks. In a minority of instances, a dose reduction may optimize effectiveness and tolerability.

Potential advantages of switching from one monotherapy to another are cost, single drug side effects, and simpler dosing regimen. These factors are likely to be associated with improved adherence, compared to adjunctive treatment regimens. Two clinician-monitored methods for switching monotherapies are typically used. The first requires a complete tapering of the initial drug before starting the second drug. This is known as the “V-method.” The second “X-method” involves a cross-gradual tapering of the initial drug while adding the second drug. This method minimizes discontinuation-emergent symptoms and is preferred when there are similar pharmacodynamic drug profiles (e.g., switching from one SSRI to another or from an SSRI to an SNRI); however, it should not be used when the potential for a hazardous drug–drug interaction exists (e.g., switching from an MAOI drug to a TCA or SSRI).

APPLY ADJUNCTIVE STRATEGIES. In clinical practice, most adjunctive treatments are administered sequentially, adding the second agent only after the first has been ineffective or partially effective. This second-line approach was used at several stages in the STAR*D trial. In a few studies, concurrent combined treatments have been initiated from the outset with inconclusive results. Until there are specific indications for this approach, sequential adjunctive treatments are recommended.

There is a large body of evidence to support the adjunctive use of second-generation antipsychotics (SGAs) also known as dopamine and serotonin modulators in the treatment of MDD, and TRD. Although they share a common dopamine D2 antagonist action, these agents differ in their affinities for D2, D3, and other receptors. This justifies switching from one SGA to another in the absence of a favorable response or if adverse effects of the first agent are intolerable. Based on a series of studies, the combination of olanzapine and fluoxetine was superior to either agent alone for TRD. Although the only SSRI involved in these trials was fluoxetine, it is likely that olanzapine (5 to 10 mg) would also be effective when combined with other antidepressants. Quetiapine (150 to 300 mg) is the only atypical agent to demonstrate antidepressant efficacy as monotherapy in the treatment of MDD and is also effective when combined with a range of antidepressants in patients with previous TRD. Augmentation with risperidone was efficacious in two double-blind, placebo-controlled trials. Aripiprazole (5 to 15 mg) is an effective adjunctive therapy in MDD. Interestingly, patients who either worsened or failed to improve with a first antidepressant treatment still showed significant benefits with adjunctive aripiprazole. Like aripiprazole, brexpiprazole acts as a dopamine and serotonin receptor partial agonist and has shown efficacy as an adjunctive agent in MDD. Despite similar efficacy, brexpiprazole shows improved tolerability (e.g., potentially less akathisia and extrapyramidal symptoms) compared to aripiprazole. Adjunctive treatment with ziprasidone in a TRD trial was not significantly better than continuation of the SSRI (sertraline), supporting the view that SGA adjunct efficacy is not a class effect but depends on the individual receptor profiles of these agents. Cariprazine (see bipolar section) acts as a partial agonist at the dopamine D3, D2, and 5-HT1A receptors, it is an antagonist at the 5-HT2A receptors. Cariprazine is currently approved for both manic and depressive episodes of bipolar disorder and is under investigation as an adjunctive agent for MDD treatment. Although SGAs provide a valuable additional treatment option for TRD, their potential to cause serious side effects, including weight gain, dyslipidemia, hyperglycemia, akathisia, extrapyramidal symptoms, and QTc interval prolongation, is of concern. This is most relevant when long-term treatment is anticipated. An unanswered question is how long to continue any adjunctive agent particularly when the additional therapy carries a hazardous side effect burden (e.g., weight gain or hyperlipidemia).

Several agents including lithium, triiodothyronine (T3), and buspirone have been displaced as first-line adjunctive agents in favor of SGAs. Most studies of lithium involved augmentation of TCAs and the daily dose varied widely (250 to 1,200 mg); nevertheless, two recent meta-analyses support the efficacy of adjunctive lithium therapy. There is also limited support for the use of T3 at a dose of 25 to 50 μg/day as an augmentation strategy. At the third level of treatment in STAR*D, lithium, and T3 both produced modest increases in remission. Buspirone, a partial postsynaptic 5-HT1A agonist, was also evaluated in the STAR*D trial and was comparable in effectiveness to bupropion in some, but not all, outcome measures in the augmentation of citalopram.

When patients fail to respond to the above strategies, psychotherapy, and various forms of neurostimulation should be considered. Among the psychotherapies, most evidence supports CBT or modified versions including internet-based CBT (refer to section on psychotherapy).

Electroconvulsive therapy (ECT) (refer to neurostimulation section) is very effective in antidepressant treatment-refractory patients with response rates as high as 70% and is specifically indicated for certain groups of patients with depression, including those with acute suicidal ideation or with mood congruent or incongruent delusions. Among subsequent forms of neurostimulation, repetitive transcranial magnetic stimulation (rTMS) has the most evidence to support its role as an adjunctive treatment.

CONSIDER THIRD-LINE ANTIDEPRESSANTS. The use of tricyclic and MAOI antidepressants is limited by their potential for side effects and toxicity. Nevertheless, they should be considered for particularly difficult-to-treat cases of depression. Among the TCAs, amitriptyline is the most evaluated. It has mixed reuptake inhibition effects on both 5HT and NE transporters with additional cholinergic, histaminergic, serotonergic, and noradrenergic receptor actions. These multiple effects are responsible for the wide range of side effects including sedation, weight gain, urinary retention, and cardiac conduction defects. Nortriptyline is a metabolite of amitriptyline that was previously recommended to treat late-life depression due to a less severe side effect profile. Imipramine, desipramine, and clomipramine are also TCAs that are typically reserved for third-line use, although clomipramine is indicated for the treatment of obsessive-compulsive disorder (OCD) and not MDD in the United States. In all cases, plasma monitoring may be helpful in determining appropriate doses.

Tranylcypromine, phenelzine, and isocarboxazid are irreversible inhibitors of both MAO-A and MAO-B. While this mode of action provides a broad-spectrum inhibition of monoamine reuptake and a favorable antidepressant effect, it also increases the risk of orthostatic hypotension, and in the presence of tyramine-rich foods may produce a life-threatening hypertensive crisis. This necessitates a low tyramine diet during MAOI treatment. Hence, these agents are generally reserved for otherwise treatment-resistant patients. Selegiline is an irreversible inhibitor of MAO-B that has been used as an adjunct therapy for Parkinson disease and is available in both oral and transdermal patch delivery systems to treat MDD. Its use is generally restricted to TRD patients. Moclobemide is a selective and reversible MAO-A inhibitor with a generally safer tolerability profile, although it is not approved for use as an antidepressant in the United States.

Future Directions. Tailoring management to a selective subgroup of patients shows promising results as demonstrated in a precision medicine–based, “fast-fail” approach there is evidence that engaging κ-opioid-receptor antagonism improves reward-circuit dysfunction and anhedonia in MDD. This approach aims to eliminate ineffective therapies at an early phase of clinical trials, thereby avoiding costly large trials that fail to show efficacy.

Ketamine and its enantiomer esketamine, both N-methyl-D-aspartate (NMDA) receptor antagonists display antidepressant efficacy as adjunctive agents in TRD, delivered respectively as intravenous and intranasal formulations. Current practice involves four to six ketamine infusions over 2 to 3 weeks as an initial course of treatment, but there are no clear guidelines on maintenance therapy. In the case of esketamine, continuation treatment once weekly or once every 2 weeks proved superior to antidepressant plus placebo in delaying relapse. There is still a lack of consensus on postketamine/esketamine maintenance treatment. Adverse behavioral (e.g., dissociative symptoms) and physiologic (e.g., hypertension) events require careful monitoring during treatment.

Ketamine has also been administered in oral, intranasal, sublingual, and intramuscular (IM) formulations particularly in combination with psychotherapy, but variability of absorption through these routes makes it difficult to identify appropriate doses for these routes of administration.

Although there is considerable interest in the adjunctive role of anti-inflammatory agents to treat MDD, no single compound has emerged with clear antidepressant benefits. Evidence that proinflammatory cytokines such as interleukin 1 (IL-1) and IL-6 are frequently elevated in depression has stimulated interest in the potential role of anti-inflammatory agents as antidepressants. Celecoxib, minocycline, and infliximab are examples of anti-inflammatory agents with potential antidepressant properties. Pramipexole, a D3 dopamine agonist has been shown to relieve depression in Parkinson disease and at higher doses in patients with TRD, including bipolar disorder.

The past decade has seen a revival in human psychedelic drug research. One of the earlier studies (2016) demonstrated that administering a single, 25-mg psilocybin (magic mushroom) dose with psychotherapy for resistant depression markedly reduced depressive symptoms from 1 week to 3 months after treatment. Since then, trials have continued to expand for wider populations including patients with terminal illness, PTSD, and substance use disorders. In a recent double-blind, randomized controlled trial psilocybin was comparable to escitalopram in reducing depression. Other psychedelic agents including LSD and MDMA are also being investigated.

Maintaining Remission

Address Risk Factors for Relapse and Recurrence. The primary aim of maintenance treatment is to prevent the recurrence of a new episode. It should be considered for patients with MDD who have experienced recurrent episodes or PDD. Maintenance medication is more effective than placebo in virtually all studies to date. Given the high rate of relapse for individuals who have had three or more episodes, maintenance therapy is recommended.

In general, the management of PDD is not significantly different from that of MDD. While there may be differences in the tailoring of treatment plans based on symptom severity, and chronicity, the general principles of pharmacotherapy and psychotherapy remain the same. The cognitive-behavioral analysis system of psychotherapy (CBASP) is typically recommended for the treatment of PDD, but it has yet to become a standard of care and other evidence-based psychotherapies including cognitive behavioral therapy and interpersonal therapy should also be considered.

Patients who consistently relapse following antidepressant discontinuation, those with recurrent psychotic episodes, persistent residual symptoms, or a strong family history of mood disorders are candidates for prolonged maintenance therapy. In any of these circumstances, clinicians and patients should decide collaboratively whether to initiate maintenance treatment or to maintain contact and rapidly reinstate treatment if there are emergent symptoms. If a new episode develops when the patient is free of treatment, then early intervention shortens the length of the new episode. There are virtually no data to suggest when to discontinue maintenance medication, although a study examining the impact of discontinuing venlafaxine after either 1 year or 2 years of maintenance treatment supported the benefits of longer-term treatment. When the decision to discontinue treatment is made, careful monitoring is required during gradual discontinuation, particularly for the first 6 months as this is a high-risk period for recurrence.

TREATMENT OF BIPOLAR DISORDERS

The discovery of the utility of atypical antipsychotics and some anticonvulsants such as valproate and lamotrigine for bipolar disorder in the 1990s heralded a new era in the pharmacologic treatment of bipolar disorders. The availability of several newer treatments for the management of bipolar disorder has presented clinicians with further options to tailor the treatment to suit individual patient needs based on efficacy, safety, tolerability, and patient preferences. The past two decades have also witnessed significant progress in understanding the adjunctive role of psychological and behavioral treatments in managing bipolar disorder.

This section will address the foundations of the management of bipolar disorder and specific treatment options for the management of mania, bipolar depression, and prophylaxis of bipolar disorder. Clinical recommendations are provided for the management of each phase of bipolar disorder with guidance on first-, second-, and third-line choices, but it is important to acknowledge that these recommendations are mainly applicable to bipolar I disorder. This section also briefly covers the management of bipolar II disorder as well as the management of comorbidity in bipolar patients. The reader is advised to refer to other sections in the textbook for managing bipolar disorder during pregnancy and the postpartum period as well as in children and adolescents, and older adults.

Foundations of Management

Diagnose and Intervene Early. Although mania is the defining feature of bipolar I disorder, depressive episodes and subsyndromal depressive symptoms are more predominant in the overall course of bipolar I disorder. Indeed, prospective cohort studies suggest that patients with bipolar I disorder spend more time (a ratio of 3:1) experiencing depressive episodes/symptoms than manic episodes/symptoms. Depressive symptoms are even more prevalent in bipolar II patients, with the average patient spending 37 times more days experiencing depressive symptoms than hypomanic symptoms.

Patients with bipolar disorder are also more likely to seek help when they are depressed and often are not aware or do not easily identify information about previous manic/hypomanic episodes. Hence, clinicians must be vigilant to the possibility of bipolar disorder in every patient presenting with depressive symptoms. Studies suggest that about two-thirds of patients with bipolar disorder are misdiagnosed initially, and major depressive disorder is the most common misdiagnosis. While screening instruments such as the Mood Disorders Questionnaire are useful, they do not replace a thorough clinical assessment that systematically probes for a previous history of manic/hypomanic symptoms. The probabilistic approach, which takes into account the presence or absence of certain clinical or disease variables, can also assist in the diagnostic evaluation process. For instance, the probability of depression being part of a bipolar disorder is higher in patients who have a family history of bipolar disorder and if their depressive episode includes psychotic features, reverse vegetative symptoms, or mood lability as a part of their clinical presentation. Management decisions can then be informed based on the level of probability of having bipolar disorder. In the context of diagnostic uncertainty, patients with a very high likelihood may benefit from a focus on treatments used for bipolar disorder, whereas those with a low likelihood of bipolar disorder may be reasonably treated for nonbipolar depression, but with ongoing monitoring for the emergence of any manic/hypomanic symptoms.

Early intervention has the potential to improve outcomes in bipolar disorder. Indeed, studies in first-episode mania patients suggest that those that remain episode-free after their first manic episode show improvements (reversal of cognitive deficits) in cognitive functioning and minimal decline in age-related gray matter volumes.

Conduct Systems Review, Baseline Investigations, and Implement Safety Monitoring for Psychotropic Medication. In addition to obtaining a comprehensive psychiatric history to diagnose bipolar disorder, clinicians are advised to conduct a systems review of medical history with particular attention to personal and family history of cardiovascular, endocrine, and metabolic disorders as the incidence of medical comorbidities is higher in bipolar patients compared with the general population. Many psychotropic medications have a tendency to cause weight gain, and some are associated with an increased likelihood of metabolic syndrome. Therefore, it is important to conduct some baseline laboratory investigations and obtain weight and waist circumference as early as possible in the treatment course, ideally baseline prior to initiating pharmacotherapy (see Fig. 13.6–1 for details) for bipolar disorder. Patients that require treatment with lithium, valproate, carbamazepine, or atypical antipsychotics need additional specific lab investigations prior to initiating treatment. During follow-up visits, clinicians are advised to routinely monitor for medication-induced adverse events at each scheduled clinical follow-up and institute appropriate strategies to address them. Patients should have weight, appropriate laboratory investigations, and serum levels of lithium, valproate, or carbamazepine measured at regular intervals as outlined in Figure 13.6–1.

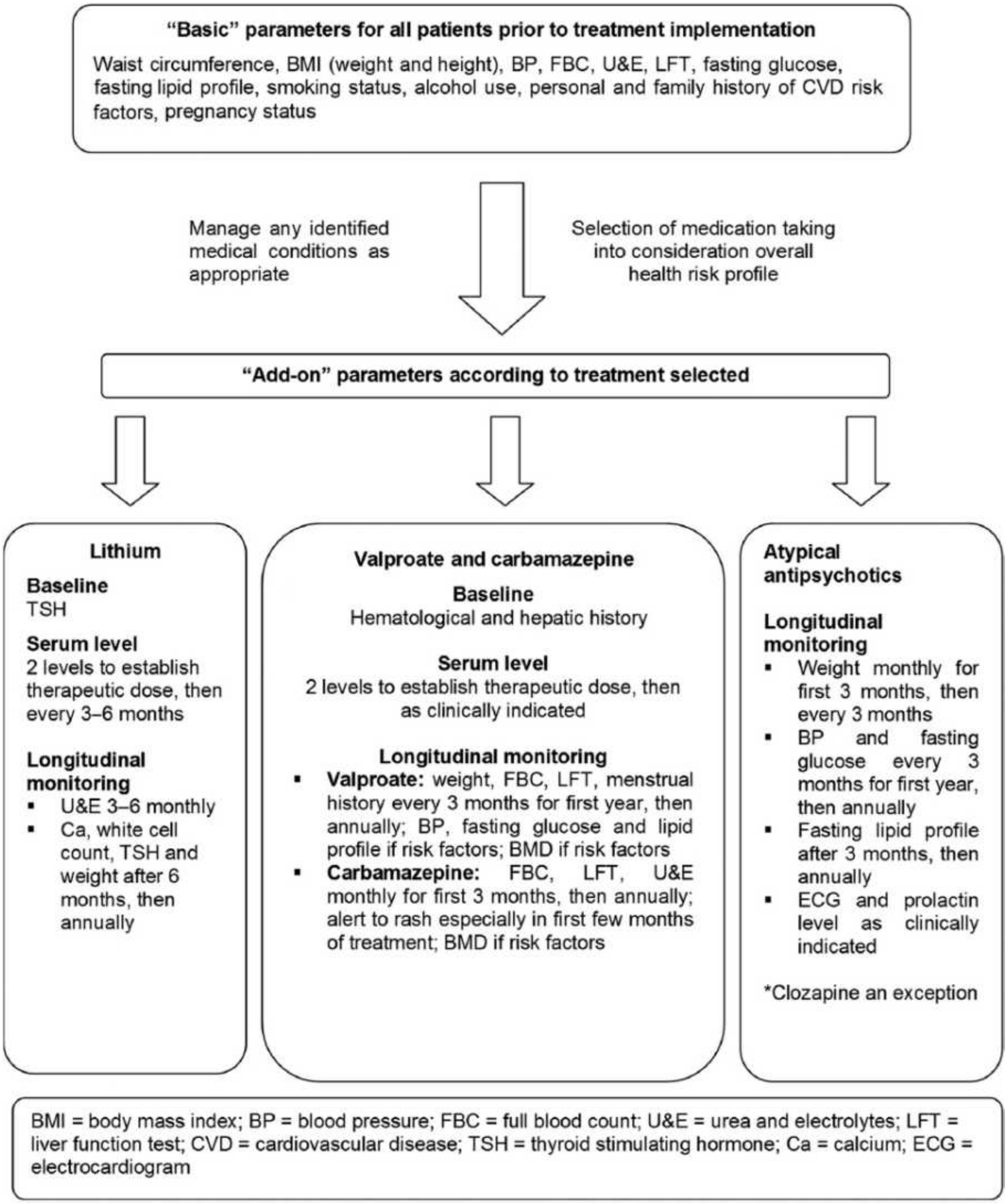

"Basic" parameters for all patients prior to treatment implementation

Waist circumference, BMI (weight and height), BP, FBC, U&E, LFT, fasting glucose, fasting lipid profile, smoking status, alcohol use, personal and family history of CVD risk factors, pregnancy status

Manage any identified medical conditions as appropriate

Selection of medication taking into consideration overall health risk profile

"Add-on" parameters according to treatment selected

Lithium

Baseline
TSH

Serum level
2 levels to establish therapeutic dose, then every 3–6 months

Longitudinal monitoring

- U&E 3–6 monthly
- Ca, white cell count, TSH and weight after 6 months, then annually

Valproate and carbamazepine

Baseline
Hematological and hepatic history

Serum level
2 levels to establish therapeutic dose, then as clinically indicated

Longitudinal monitoring

- **Valproate:** weight, FBC, LFT, menstrual history every 3 months for first year, then annually; BP, fasting glucose and lipid profile if risk factors; BMD if risk factors
- **Carbamazepine:** FBC, LFT, U&E monthly for first 3 months, then annually; alert to rash especially in first few months of treatment; BMD if risk factors

Atypical antipsychotics

Longitudinal monitoring

- Weight monthly for first 3 months, then every 3 months
- BP and fasting glucose every 3 months for first year, then annually
- Fasting lipid profile after 3 months, then annually
- ECG and prolactin level as clinically indicated

*Clozapine an exception

BMI = body mass index; BP = blood pressure; FBC = full blood count; U&E = urea and electrolytes; LFT = liver function test; CVD = cardiovascular disease; TSH = thyroid stimulating hormone; Ca = calcium; ECG = electrocardiogram

FIGURE 13.6–1. Recommendations for treatment monitoring in bipolar disorder. (Adapted with permission from Ng F, Mammen OK, Wilting I, et al. The International Society for Bipolar Disorders (ISBD) consensus guidelines for the safety monitoring of bipolar disorder treatments. *Bipolar Disorders.* 2009;11(6):559–595. © 2009 The Authors. doi: 10.1111/j.1399-5618.2009.00737.x.)

Treat the Disorder and Not Just Acute Episodes. Pharmacotherapy is the cornerstone of the management of bipolar disorder. When initiating treatment of acute episodes, while stabilization of the acute episode is the immediate goal, it is essential for clinicians to bear in mind that bipolar disorder is a recurrent illness, and hence prevention of further mood episodes is an equally important consideration in treatment selection. Therefore, when choosing a medication to treat an acute episode, those options that have proven efficacy in treating not only the index episode but also in preventing the recurrence of both manic and depressive episodes should be considered first before selecting medications that have efficacy for only acute episodes. There is growing consensus that mood episode recurrences are “toxic” to the brain and are associated with cumulative reductions in brain volumes, lower brain-derived neurotrophic factor (BDNF) levels, cognitive impairment, and poorer response to pharmacotherapy, and thus prevention of recurrence of mood episodes is paramount in the management of bipolar disorder. Consensus is growing that prophylaxis should be offered for all patients with bipolar disorder, including for those after the first manic episode, as prospective naturalistic cohort studies suggest that 60% to 75% of first-episode manic patients experience a recurrence within 4 years. Adjunctive psychological treatments such as psychoeducation promote treatment adherence and reduce the risk of relapse of mood episodes in bipolar patients, and hence should be initiated as soon as the patient is amenable to such strategies.

Treat Acute Episodes to Full Remission. Acute manic, hypomanic, and depressive episodes must be treated to full and sustained remission. This is because several studies now suggest that residual subsyndromal mood symptoms not only significantly affect the quality of life and functioning but also strongly predict relapse of subsequent full-blown mood episodes. Patients who recovered from an index mood episode with residual mood symptoms experience a new episode 3.4 times sooner than those that recovered from an index episode with no residual symptoms. Median time to recurrence of a mood episode is five times shorter in those with residual symptoms compared with those with full recovery from an index episode.

Assess for and Treat Comorbidities. Psychiatric and medical comorbidities are common in bipolar disorder. While the estimates of the prevalence vary, at least 75% of patients with bipolar disorder have at least one-lifetime comorbid psychiatric disorder. Alcohol or substance use disorders are present in at least 40% of patients, and some type of anxiety disorder is present in 50% to 75% of patients. There is evidence that patients with alcohol and substance use disorders or anxiety disorder comorbidities take longer to recover from an index episode and are more likely to relapse compared with those without these comorbidities. Hence, the optimal management of bipolar disorder should include not only treatment of mood symptoms/episodes associated with this condition but also assessing for comorbidities and instituting appropriate pharmacologic/psychological management strategies to address comorbidities in order to improve outcomes. During the maintenance treatment, clinicians should routinely monitor for the onset of new psychiatric, substance, or alcohol use comorbidities and manage them as clinically indicated.

Address Cognitive Impairment. Patients with bipolar disorder often complain of cognitive difficulties. A thorough clinical assessment should be conducted in such patients to address contributory factors such as comorbid substance use (e.g., marijuana), psychotropic medication-induced confusion or sedation, lithium-induced hypothyroidism, valproate-induced hyperammonemia, etc. for cognitive difficulties. Evidence has accumulated over the past 20 years to suggest that cognitive dysfunction can be an inherent part of bipolar disorder for nearly two-thirds of patients. Impairments are seen in several domains, including processing speed, verbal learning, working memory, and executive function, and such impairments are present not only during acute mood episodes but even during euthymic periods. The effect sizes for the magnitude of impairment during euthymia are in the range of 0.6 to 0.8. Cognitive impairments, particularly in the domains of verbal learning and executive function, correlate with functional impairments. There is increasing interest in examining the efficacy of pharmacologic treatments and cognitive remediation to address cognitive dysfunction. While evidence for efficacy is limited, there is early preliminary data supporting the use of lurasidone or nonpharmacologic interventions such as cognitive remediation.

Monitor for Treatment Adherence. While the rates of treatment nonadherence vary depending upon how it is defined, there is evidence that up to 60% of patients with bipolar disorder are nonadherent or poorly adherent to pharmacotherapy. The consequences of poor adherence to pharmacotherapy include illness destabilization with increased relapse/recurrence of mood episodes, higher suicide risk, increased hospitalization, disability, and poor functioning. Several factors such as poor insight/denial of illness, concerns about or experience of side effects of medications, inadequate efficacy of medications, younger age, low level of education, or cognitive impairment all contribute to poor treatment adherence. It is important for clinicians to be cognizant of demographic variables and explore illness-related factors, attitudes toward illness, and treatment-related factors that predict poor adherence and address them routinely at clinical follow-up visits. There is evidence that psychoeducation improves treatment adherence. More recently, customized treatment adherence programs have been developed that include modules on psychoeducation, substance abuse, medication routines, and communication with providers, and these programs can be tailored to address the specific reasons for the treatment of nonadherence in individual patients.

Monitor for the Emergence of Mood Symptoms and Suicide Risk. There is evidence that bipolar patients with subsyndromal mood symptoms are more likely to relapse into a full-blown mood episode compared with those without such symptoms. Therefore, it is prudent for clinicians to routinely screen for emergent subsyndromal symptoms at each clinical visit. Clinicians may find the Young Mania Rating Scale (YMRS) and MADRS particularly helpful in systematically assessing and quantifying subsyndromal symptoms. The reasons for the emergence of such symptoms may include poor treatment adherence, comorbidity, or partial efficacy of ongoing treatments, which should be identified when possible and addressed with psychoeducation or changes to pharmacotherapy as appropriate in order to rapidly resolve such symptoms. Subsyndromal depressive symptoms can also be treated with psychological strategies such as interpersonal and social rhythm therapy (IPSRT) or CBT.

Clinicians should also routinely monitor for the emergence of suicidal ideation not only when treating acute episodes but also during the maintenance treatment. This is because suicidal ideation is common in patients with bipolar disorder and that between 25% to 50% of patients with bipolar disorder make a suicide attempt at least once during their lifetime. The rates of lifetime risk of suicide deaths in bipolar patients range from 6% to 19% depending on the sample, with more recent studies in less severely ill populations suggesting lower rates. Regardless, the standardized mortality ratios for suicide in bipolar disorder patients range from 10 to 30, which means that bipolar patients are 10 to 30 times more likely to die of suicide compared with the general population. A recent International Society for Bipolar Disorders (ISBD) Task Force identified several risk factors for suicide attempts and suicide deaths (see Table 13.6–4) in bipolar disorder.

Clinicians should be more vigilant for suicidal ideation and other warning signs of suicide in those with an increasing number of risk factors. The risk of suicide attempt and suicide death is higher early in the course of the illness, during acute mood episodes (especially in the presence of mixed features) or hospitalization, and within 4 to 12 weeks of discharge from the hospital. When suicidal ideation is endorsed, clinicians must take appropriate steps immediately to ensure the safety of the patient, and these may include hospitalization, preventing/reducing access to methods of suicide (e.g., medication overdose, access to firearms), establishing a therapeutic alliance, and involving family members/friends when possible in care, and addressing the illness-related and psychosocial factors that led to the emergence of suicidal ideation.

Table 13.6–4. Factors Associated with Suicide Attempts and Suicide Deaths in Bipolar Disorder

Variables	Increased Likelihood of Suicide Attempts in Bipolar Disorder	Increased Likelihood of Suicide in Bipolar Disorder
Sociodemographics		
Sex	Female	Male Females have higher SMR
Age	Younger Older age—higher lethality attempts	No clear effect Older age—higher ratio of deaths/attempts
Race	Minorities—specific to youth only	
Marital status	Single, divorced, single parents	
Religious affiliation	Discordant results	
Characteristics of Bipolar Disorder		
Age of illness onset	Younger age	
Duration of illness	Discordant results	
Polarity of 1st episode	Depression Mixed symptoms Mania—more violent attempts	
Predominant polarity	Depressive	
Polarity of current episode	Depressive episodes Mixed states	Depressive episodes Mixed states Manic states with psychotic features
Other mood episode	Mixed features	Hopelessness

characteristics	Greater number and severity of episodes/rapid cycling Anxiety Atypical features Suicidal ideation	Psychomotor agitation
Psychosis at any time	Discordant results	No effect
Bipolar disorder subtype	No clear effect	No clear effect
Comorbidities		
Psychiatric comorbidity	Substance use disorder Cigarette smoking Coffee intake Anxiety disorder Obesity or high BMI Eating disorder	Substance use disorder—no clear effect Anxiety disorder
Personality disorders and characteristics	Personality disorders—including borderline or cluster bipolar disorder in particular Aggression or irritability Impulsivity (discordant results)	
Other Clinical Variables		
First-degree family history	Mood disorders Bipolar disorder Suicide	Mood disorders Bipolar disorder Suicide
Prior suicide attempts	Presence	Presence
Early life trauma	Childhood abuse Higher early-life stress	
Psychosocial precipitants	Interpersonal problems Occupational problems Bereavement	Present within a week prior to death
Sexual	Presence	

dysfunction

SMR, standardized mortality ratio.

From Schaffer A, Isometsä ET, Azorin J, et al. A review of factors associated with greater likelihood of suicide attempts and suicide deaths in bipolar disorder: part II of a report of the International Society For Bipolar Disorders Task Force on suicide in bipolar disorder. *Aust N Z J Psychiatry*. 2015;49(11):1006–1020.

Acute Mania: Treatment Options

Lithium. There is robust evidence for the efficacy of lithium in treating acute mania. When adequately dosed, the therapeutic effect of lithium becomes apparent within 7 days as the separation from placebo was seen as early as day 7 in clinical trials. While the response rates to lithium varied between the trials, on average, about 50% of patients met the criteria for response to lithium in these 3-week trials.

Lithium is typically started at 900 mg/day, although lower starting doses may be required for some patients who are more sensitive to side effects. Serum levels should be measured after 5 days of consecutive therapy, 12 hours after the last dose. Serum levels between 0.8 and 1.2 mEq/L are likely required to treat acute mania effectively, although lower levels (0.4 to 0.8 mEq/L) may be appropriate in older adults. Lithium levels of ≥1.5 mEq/L can lead to symptoms of lithium toxicity, and this should be treated as a medical emergency. Patients should be educated and clinically monitored for signs of lithium toxicity that may include gastrointestinal (vomiting and diarrhea), neurologic (coarse tremor, agitation, dysarthria, drowsiness, lethargy, ataxia), renal (polyuria and polydipsia), and cardiovascular (syncope, dizziness, arrhythmias) symptoms.

Anticonvulsants

CARBAMAZEPINE. Carbamazepine is as effective as lithium and typical antipsychotics in treating acute mania. Carbamazepine extended-release (ER) is also effective, with response rates between 40% and 60% that are similar to the aggregate response rates of 52% observed with carbamazepine in older studies.

Carbamazepine is typically started at 100 to 200 mg bid, and the dose is gradually titrated upward to achieve a therapeutic effect. Most patients will need between 400 and 800 mg/day. The therapeutic effect is observed typically within 1 week to 10 days. There is no evidence to suggest a relationship between serum carbamazepine levels and acute therapeutic efficacy in mania.

OXCARBAZEPINE. While oxcarbazepine may be better tolerated than carbamazepine, the data for its efficacy in treating acute mania is conflicting. While four small controlled studies in adult manic patients support the efficacy, a recent larger clinical trial in children and adolescents with mania that compared oxcarbazepine with placebo found no differences in improvement in manic symptoms between the two groups. Given the lack of larger placebo-controlled trials in adults with modern methodology and a negative trial in children and adolescents with acute mania, further studies are clearly needed to determine the efficacy of oxcarbazepine for acute mania. This statement is consistent with a recent Cochrane review which concluded that there is currently insufficient evidence for the efficacy of oxcarbazepine in treating acute mood episodes in bipolar disorder.

DIVALPROEX/VALPROATE. Divalproex consists of sodium valproate and valproic acid in a 1:1 molar ratio. The landmark trial by Bowden and colleagues (1994) that compared the efficacy of divalproex with placebo and lithium led to the approval of divalproex for the treatment of acute mania. Divalproex is as effective as lithium in adult manic patients. Two out of three studies that compared divalproex with olanzapine showed similar efficacy at 3 weeks. Divalproex has similar efficacy to quetiapine in treating acute mania in adolescents.

The therapeutic effect of divalproex is apparent by day 7, although studies with divalproex oral loading (20 mg/kg/day or 30 mg/kg/day on days 1 and 2 followed by 20 mg/kg/day) suggest a more rapid onset of action with effect apparent by day 5 or sooner. Posthoc analyses of data from these trials suggest that divalproex has a broader spectrum of efficacy with effect in manic patients with and without concurrent depressive symptoms, psychotic symptoms, rapid cycling, and regardless of the number of prior mood episodes.

There is a linear relationship between serum valproate levels and therapeutic efficacy for acute mania, with higher serum levels yielding greater efficacy. The evidence suggests that optimal efficacy may not be obtained when serum levels are below 94 µg/mL. Hence, clinicians should titrate the dose of valproate to achieve serum valproate levels to this target range. If tolerability is not a concern, the doses could be further increased if necessary as some patients may require higher levels for optimal efficacy.

OTHER ANTICONVULSANTS. Lamotrigine, gabapentin, topiramate, and licarbazepine are not effective in treating acute mania as none are separated from placebo in controlled clinical trials in improving manic symptoms. Lamotrigine or topiramate adjunctive therapies are also ineffective in treating acute mania.

Other anticonvulsants such as tiagabine, pregabalin, levetiracetam, felbamate, and zonisamide have not been assessed in randomized, double-blind controlled trials for acute mania.

Conventional/First-Generation Antipsychotics. Haloperidol is the most widely investigated first-generation antipsychotic for acute mania. Several modern placebo-controlled trials of atypical antipsychotics such as risperidone, olanzapine, quetiapine, ziprasidone, and aripiprazole used haloperidol as an active comparator. In all these studies, haloperidol was more effective than placebo, confirming the efficacy of haloperidol in treating acute mania.

The efficacy of haloperidol was apparent at early as day 2 in some trials. The doses of haloperidol used in these trials ranged from 2 to 15 mg per day, although in the ziprasidone trial, up to 30 mg/day of haloperidol was allowed. These studies demonstrated that haloperidol was as effective as risperidone, olanzapine, and aripiprazole and more effective than quetiapine and ziprasidone. Extrapyramidal symptoms are common with haloperidol, and patients with bipolar disorder are more prone to movement disorders such as tardive dyskinesia. Furthermore, more patients treated with haloperidol experienced depression in some of these trials compared with placebo or atypical antipsychotics.

It is believed that the efficacy of haloperidol for acute mania is related to its ability to reduce dopamine transmission through the blockade of dopamine D2 receptors. Since all conventional antipsychotics reduce dopamine transmission through their effects on D2 receptors, they are likely effective in treating acute mania. Chlorpromazine has been compared with lithium and found to have similar efficacy overall, although it appears to be more effective in “highly active” manic patients, which is likely related to its sedating effects, while lithium has better efficacy in treating core manic symptoms. The loxapine-inhaled formulation is effective y in treating agitation in acutely manic patients and found to be effective, but its utility in treating core manic symptoms has not been assessed in placebo-controlled trials.

Haloperidol in combination with lithium or valproate is more effective than lithium or valproate monotherapy. Similarly, valproate add-on to haloperidol/perphenazine is more effective than monotherapy with haloperidol/perphenazine.

Atypical Antipsychotics

MONOTHERAPY. Several atypical antipsychotics are currently in use for the treatment of psychosis. Of these, risperidone, olanzapine, quetiapine, ziprasidone, aripiprazole, paliperidone, asenapine, and cariprazine have proven efficacy and hence approved for the treatment of acute mania in select countries. Cariprazine, a new atypical antipsychotic with D2 and D3 partial agonist properties, is effective and was recently approved by the FDA for the treatment of acute mania. Table 13.6–5 provides a summary of the efficacy of atypical antipsychotics and the dose ranges commonly used in clinical practice. Clozapine and lurasidone are likely effective, but no placebo-controlled trials have assessed their efficacy for acute mania to date. Surprisingly, brexpiprazole was not found to be effective for acute mania in RCTs and hence is not recommended.

The only study that compared two atypical antipsychotics head-to-head found no difference in efficacy between risperidone and olanzapine for acute mania. While other atypical antipsychotics have not been compared directly in double-blind, controlled trials, there is no reason to believe that there are any clinically meaningful significant differences in efficacy between various atypical antipsychotics, given the broad similarity in effect sizes and placebo-corrected response rates.

Table 13.6–5.

Atypical Antipsychotics for Bipolar Disorder: Efficacy Summary and Dose Ranges

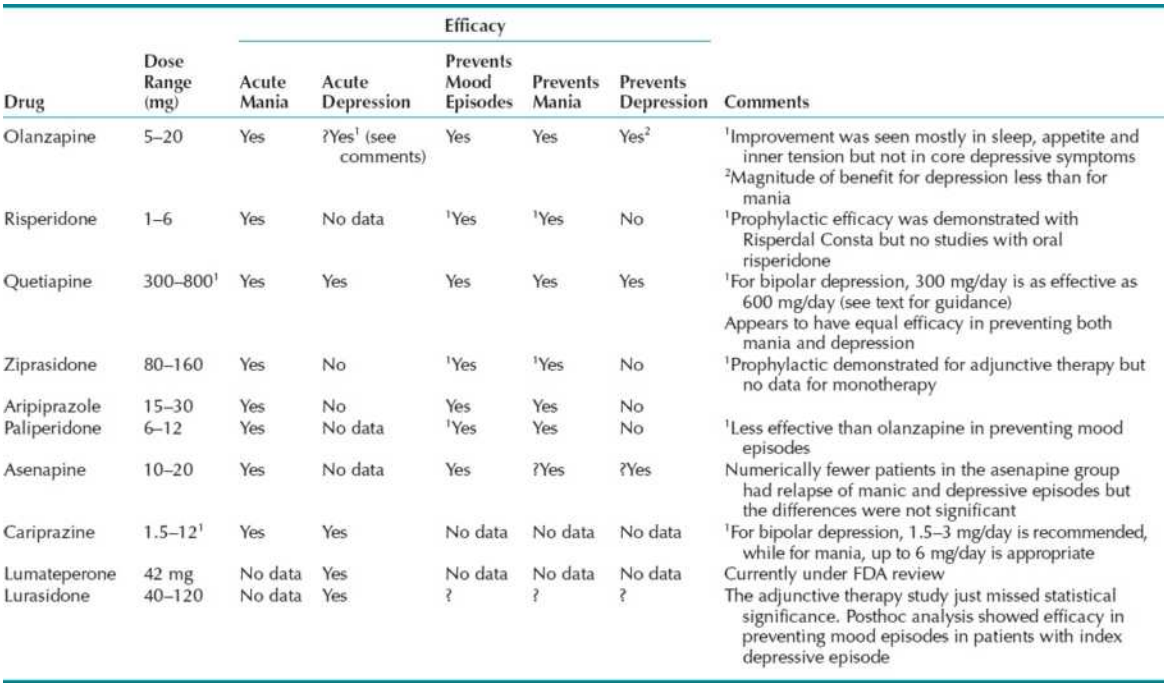

Drug	Dose Range (mg)	Efficacy					Comments
		Acute Mania	Acute Depression	Prevents Mood Episodes	Prevents Mania	Prevents Depression	
Olanzapine	5–20	Yes	?Yes[1] (see comments)	Yes	Yes	Yes[2]	[1] Improvement was seen mostly in sleep, appetite and inner tension but not in core depressive symptoms [2] Magnitude of benefit for depression less than for mania
Risperidone	1–6	Yes	No data	[1]Yes	[1]Yes	No	[1] Prophylactic efficacy was demonstrated with Risperdal Consta but no studies with oral risperidone
Quetiapine	300–800[1]	Yes	Yes	Yes	Yes	Yes	[1] For bipolar depression, 300 mg/day is as effective as 600 mg/day (see text for guidance) Appears to have equal efficacy in preventing both mania and depression
Ziprasidone	80–160	Yes	No	[1]Yes	[1]Yes	No	[1] Prophylactic demonstrated for adjunctive therapy but no data for monotherapy
Aripiprazole	15–30	Yes	No	Yes	Yes	No	
Paliperidone	6–12	Yes	No data	[1]Yes	Yes	No	[1] Less effective than olanzapine in preventing mood episodes
Asenapine	10–20	Yes	No data	Yes	?Yes	?Yes	Numerically fewer patients in the asenapine group had relapse of manic and depressive episodes but the differences were not significant
Cariprazine	1.5–12[1]	Yes	Yes	No data	No data	No data	[1] For bipolar depression, 1.5–3 mg/day is recommended, while for mania, up to 6 mg/day is appropriate
Lumateperone	42 mg	No data	Yes	No data	No data	No data	Currently under FDA review
Lurasidone	40–120	No data	Yes	?	?	?	The adjunctive therapy study just missed statistical significance. Posthoc analysis showed efficacy in preventing mood episodes in patients with index depressive episode

If a medication is effective rapidly, it can shorten the duration of the manic episode and the hospital length of stay. When atypical antipsychotics are dosed adequately, some improvement in manic symptoms is apparent typically within the first week with all atypical antipsychotics. While there is a perception that sedating atypical antipsychotic drugs may have a more rapid onset of action, this is not borne out in clinical trials. For instance, although aripiprazole and ziprasidone are less sedating, their therapeutic effects on mania were apparent at first postbaseline measurements of day 4 and day 2, respectively, in clinical trials.

DSM-5 has eliminated the mixed episodes category and introduced mixed features as a course specifier for manic as well as depressive episodes. This is to reflect the fact that the symptoms of opposite polarity are present in nearly two-thirds of patients with acute mania or acute depression. Therefore, it is not surprising that several acute mania trials with atypical antipsychotics have included some patients with various definitions of mixed episodes or symptoms. Posthoc analyses from these trials suggest that asenapine, cariprazine, ziprasidone, aripiprazole, or olanzapine are effective in treating manic symptoms in patients with acute mania with mixed features, while asenapine, cariprazine, and ziprasidone also have some data supporting treatment of mixed depressive features during acute mania. Further, a meta-analysis suggests that atypical antipsychotics as a group are effective in treating acute mixed episodes, and the effect size for their efficacy in treating mixed episodes is in the medium range which is similar to their magnitude of effect in manic patients. Based on this, it is reasonable to conclude that atypical antipsychotics, unlike conventional first-generation antipsychotics, at the very least do not worsen or induce depressive symptoms in manic patients.

COMBINATION/ADD-ON THERAPY. There is evidence that as many as 80% of manic patients in clinical settings are treated with a combination of a mood stabilizer and an antipsychotic. In some cases, both medications are started at the same time (combination therapy), while in other cases, one medication is started, and the other is added at a later point (add-on therapy).

In clinical trials of add-on therapy, although patients met entry criteria for mania severity, refractoriness to existing therapy was not established prospectively in many studies. Hence, it is unknown if the add-on studies truly assessed the efficacy in refractory manic patients.

Regardless, there is evidence that response rates are greater when manic patients are treated with a combination/add-on of risperidone, olanzapine, quetiapine, aripiprazole, or asenapine with lithium or valproate compared with lithium or valproate monotherapy. On average, 20% more manic patients will respond with combination/add-on therapy compared with lithium or valproate monotherapy. While different combinations of treatments are not directly compared, based on differences in response rates observed between monotherapy and combination therapies in various studies, there does not appear to be any clinically meaningful differences between different combinations of atypical antipsychotic and lithium or valproate. The only exception is ziprasidone adjunctive therapy which was not superior to lithium or valproate monotherapy in the only clinical trial that assessed this strategy.

There is preliminary evidence that adding an atypical antipsychotic to carbamazepine does not confer additional efficacy. For instance, risperidone or olanzapine combined with carbamazepine is no more effective than carbamazepine monotherapy. This is likely because carbamazepine, through induction of hepatic microsomal enzymes, reduces serum levels of antipsychotics and thus affects their efficacy. In theory, this could be overcome with higher doses of the atypical antipsychotic, but that has not been tested in a controlled fashion.

Some clinicians believe that antipsychotic monotherapy is as effective as the combination strategy. There is preliminary evidence to suggest that this is clearly not the case. For instance, a large double-blind trial suggested that lithium given in combination with quetiapine XR was more effective than quetiapine XR monotherapy. Similarly, valproate in combination with haloperidol/perphenazine was more effective than monotherapy with an antipsychotic.

It is, however, important to note that greater efficacy of combination therapy comes with a trade-off in terms of adverse events as combination therapy is often associated with more adverse events than monotherapy.

Miscellaneous Agents. Tamoxifen, a drug used in the treatment of breast cancer, is a protein kinase c (PKC) inhibitor. There is evidence that treatment with lithium or valproate alters the PKC signaling pathway and, more specifically, reduces specific isoforms of this enzyme. There is evidence from five small double-blind RCTs (two monotherapy and three adjunctive therapies) that tamoxifen is effective in treating acute mania. However, clinicians must be aware that tamoxifen use in patients with breast cancer is associated with a two- to sevenfold increase in endometrial carcinoma. Further, there is no longer-term safety data for tamoxifen in bipolar patients. Similarly, two RCTs demonstrated that endoxifen, an active metabolite of tamoxifen, is as effective as valproate in treating acute mania at doses of 8 mg/day.

Verapamil was reported to be effective in two small earlier controlled trials and some open trials, but a larger double-blind, placebo-controlled trial found no evidence for the efficacy of verapamil in treating acute mania. Two very small trials suggest that nimodipine may be useful in treating manic symptoms, but larger trials are needed before any conclusions about its efficacy can be drawn. There is some evidence that clonazepam is useful in treating acute manic symptoms. While calcitonin was reported to be effective in treating acute manic symptoms in a trial conducted in the 80s, a more recent double-blind trial that compared calcitonin nasal spray versus saline nasal spray in refractory manic patients found no difference between the two groups. Celecoxib adjunctive therapy has shown efficacy in two small RCTs.

Clinical Recommendations for the Management of Acute Mania

Manic episodes can be a part of bipolar I disorder or secondary to an underlying medical condition, neurologic condition, head injury, substance use disorder (e.g., cocaine or amphetamines), or induced by medications such as steroids. When possible, verbal de-escalation techniques should be tried to calm the manic patient and obtain a thorough psychiatric and medical history. Manic patients are often uncooperative and lack insight; collateral information is vital in such cases. Patients should be assessed for the risk of aggressive behavior, the severity of mania, the presence of psychotic features, and insight, all of which inform the clinician about the treatment setting for further management. Patients with severe mania and aggressive behavior with little insight will require hospitalization, while those with mild-to-moderate mania may be treated in an outpatient setting if adequate family support is available and if the patient has some insight.

A thorough review of medication history should be conducted. If a patient had been taking antidepressant medication, this should be discontinued. Similarly, other medications that might be potentially contributing to manic symptoms, such as stimulants or steroids, should be discontinued. A physical examination and appropriate baseline medical investigations should be conducted before definitive pharmacologic treatment is instituted, although this may not be possible in some cases. If the patient has a history of substance use, this should be simultaneously addressed while mania is being treated. Psychoeducation should commence as soon as the patient is behaviorally stabilized and is amenable.

Management in an Emergency Setting. Patients presenting with a manic episode in an emergency setting are often agitated, irritable, uncooperative, and lack insight. In some cases, they may be verbally or physically aggressive and may escalate to becoming assaultive and violent. Hence, ensuring the safety of the acutely manic patient and those around is a high priority. It is important to move the patient to a low stimulus and comfortable environment as soon as possible.

The severity of agitation and aggressive behavior and the patient's willingness to accept treatment dictate whether medications can be given orally or in an injectable form. Oral medications are less likely to disrupt the therapeutic alliance and hence are preferable but have the disadvantage of slower onset of action. Further, although oral administration of benzodiazepines such as lorazepam, first-generation antipsychotics such as haloperidol and loxapine, and SGAs (risperidone, olanzapine, quetiapine, aripiprazole, ziprasidone, asenapine) is a common clinical practice in the emergency setting to control agitation and aggressive behavior, they are not widely studied in controlled clinical trials for efficacy in treating agitation and aggressive behavior in mania. Nevertheless, clinicians often employ this strategy in many parts of the world. While haloperidol and several SGAs have proven efficacy in treating acute mania, there is little evidence to suggest that lorazepam is effective in treating core manic symptoms, although this may be useful in controlling agitation and calming the patient. Hence lorazepam monotherapy for controlling agitation in mania is not recommended, but this can be given in conjunction with an antipsychotic and, in particular, first-generation antipsychotics such as haloperidol to reduce the risk of extrapyramidal symptoms and to enhance the calming effect. Clonazepam has shown some efficacy in treating manic symptoms, and this can be given in conjunction with a mood stabilizer or an antipsychotic medication with proven antimanic effects.

While lithium and carbamazepine are effective in treating acute mania, they have not been studied for their efficacy in treating acute agitation and aggressive behavior in acute mania, and hence are not recommended as monotherapy for treating agitation in an emergency setting. Valproate oral loading dose has a similar faster onset of action as the antipsychotics in treating acute mania but has not been studied for controlling agitation. For patients who have a tendency to “cheek” the medication, orally disintegrating tablets (ODT) or liquid form of antipsychotics is preferable. However, there is no convincing evidence to support that these forms of administration have a faster onset of action in comparison to conventional tablet forms.

Two novel forms of oral administration of antipsychotics have shown rapid efficacy in controlling agitation. In a recent clinical trial, sublingual asenapine was shown to rapidly control agitation within 2 hours in psychiatric patients that included manic patients with a number needed to treat (NNT) of 3. Sublingual asenapine bypasses the first-pass metabolism provided that the patient does not drink or eat for 10 minutes after the administration, and this may explain the rapid effect. Similarly, loxapine aerosol delivered through inhalation with a staccato system is also rapidly effective in controlling agitation in manic patients with an NNT of 3 achieved within 30 minutes. Loxapine inhalation (ADASUVE, 10 mg) is approved by the FDA in the United States as well as by the European Medicines Agency (EMA) for the treatment of agitation in mania. Since ADASUVE can cause bronchospasm that can potentially lead to respiratory distress and arrest, it is approved under the risk evaluation and mitigation strategy (REMS) program to mitigate the risk of bronchospasm. This basically means that ADASUVE can be administered only in an enrolled healthcare facility. Given the risk of bronchospasm, it is contraindicated in patients with a history of respiratory conditions such as asthma, chronic obstructive pulmonary disease (COPD), or other lung diseases associated with bronchospasm.

When rapid control of agitation is required to ensure safety and if a patient is unwilling to take the medication orally, IM administration of an antipsychotic should be considered. Haloperidol and second-generation antipsychotic medications such as olanzapine, aripiprazole, and ziprasidone IM formulations are rapidly effective within 2 hours in controlling agitation with an NNT of 3. Clinicians, however, must recognize that IM administration of medication against patients' wishes can be a very negative experience for the patient and may interfere with the establishment of a therapeutic alliance. Hence, if IM formulations are used to manage agitation and aggressive behavior, patients should be switched to oral medications as soon as the patient is willing to take the medication.

Treatment Selection for Acute Mania. If a manic patient was given an atypical antipsychotic with proven efficacy for mania to control agitation in the emergency setting and if this was well tolerated and improved agitation, this should be continued and the doses adjusted as clinically appropriate to treat acute mania. However, if first-generation antipsychotics such as haloperidol, loxapine, or chlorpromazine were used to control agitation, patients should be switched to a first-line antimanic agent as soon as possible. Although first-generation antipsychotics are likely as effective in treating acute mania, they are not recommended as first-line antimanic agents by most experts because first-generation antipsychotics such as haloperidol have a tendency to induce depressive symptoms and increase the risk of depressive episodes in bipolar patients.

FIRST-LINE TREATMENTS. Lithium, valproate, and several atypical antipsychotics are all considered first-line agents for treating acute mania (see Table 13.6–6) as they all are effective and the response rates are broadly similar. Although multiple treatment meta-analyses suggest that risperidone may be more effective, this may have been accounted for by variability in study populations included in different trials. Thus it is unlikely that there are any clinically meaningful significant differences in efficacy between these agents, and hence, monotherapy with any of these agents would be considered an appropriate first-line option for acute mania.

Some guidelines, such as the Canadian Network for Mood and Anxiety Treatments (CANMAT) that were published in collaboration with the ISBD, recommend a combination of an atypical antipsychotic with lithium or valproate also as a first-line option. This is because combination therapy in general results in approximately 15% to 20% more patients responding compared with monotherapy. As with monotherapy, there does not appear to be any clinically meaningful significant differences in efficacy between various proven combinations for acute mania. It is, however, important for clinicians to remember that combination therapy is likely associated with more adverse events.

Table 13.6–6.
Recommendations for Pharmacologic Treatment of Acute Mania

First-line	**Monotherapy:** Lithium, divalproex, divalproex ER, risperidone, quetiapine, quetiapine XR, aripiprazole, asenapine, paliperidone ER, cariprazine **Adjunctive therapy with lithium or divalproex:** Risperidone, quetiapine, aripiprazole, asenapine
Second-line	**Monotherapy:** Olanzapine, ziprasidone, carbamazepine, carbamazepine ER, ECT, haloperidol **Combination therapy:** Lithium + divalproex, olanzapine plus lithium or divalproex
Third-line	**Monotherapy:** Chlorpromazine, clozapine, tamoxifen **Combination therapy:** Lithium or divalproex + haloperidol, lithium + carbamazepine, adjunctive tamoxifen
Not recommended	**Monotherapy:** Gabapentin, topiramate, lamotrigine, licarbazepine, zonisamide, valnoctamide, tiagabine **Combination therapy:** Risperidone + carbamazepine, olanzapine + carbamazepine

ECT, electroconvulsive therapy; XR or ER, extended-release.

Modified with permission from Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord.* 2018;20:97–170. doi:10.1111/bdi.12609. © 2018 The Authors. Published by John Wiley & Sons Ltd.

Which First-Line Treatment?. Given that various monotherapy and combination therapy first-line options are available for treating acute mania, clinicians have to make a choice as to which monotherapy or combination therapy should be used to treat a given patient with acute mania.

Since there are no consistent efficacy differences between various first-line monotherapies and between different first-line combination therapies, physicians should consider clinical presentation, previous history of response/nonresponse, adverse events, and patient preference in deciding which monotherapy or combination to use to treat mania in a given patient. Patients are more likely to respond to combination therapy compared with monotherapy, but this comes with additional adverse event liability. Therefore, clinicians must carefully weigh the risks and benefits of monotherapy versus combination therapy for each patient before making a decision on the most appropriate strategy.

If a decision is made to use monotherapy, patients presenting with acute agitation and aggressive behavior are commonly started on an atypical antipsychotic given their effectiveness in behavioral control and more rapid onset of action compared with mood stabilizers such as lithium. Lithium is the preferred agent in patients with “classical euphoric grandiose” type of mania, whereas those with dysphoric/irritable mania, mixed features, comorbid alcohol or substance abuse, or brain trauma may respond better to valproate. Similarly, patients with multiple previous mood episodes tend to respond poorly to lithium, while this does not predict poor response to valproate. Atypical antipsychotics, and in particular, olanzapine, asenapine, aripiprazole, and ziprasidone, appear to be equally effective for patients with classical mania as well as those with mixed episodes/mixed features. While clinicians tend to believe that presence of psychotic features may require treatment with an antipsychotic, the evidence from clinical trials suggests lithium, valproate, and atypical antipsychotics are equally effective in treating mania in patients with and without psychosis. However, clinical trial data notwithstanding, if a manic patient who is on lithium or valproate continues to exhibit psychotic symptoms, clinicians should consider adding an antipsychotic. Given that there are clear differences in adverse event profiles between various antimanic agents, this is also a very important consideration in treatment selection. For instance, olanzapine is less appropriate for manic patients who are overweight or obese or have a personal or family history of diabetes. Similarly, for patients with kidney disease with compromised kidney function, lithium is not appropriate.

Since there are no efficacy differences between combinations of different atypical antipsychotics (except ziprasidone) with lithium or valproate, the decision with regard to which combination to use for a given patient is also determined by taking into consideration previous history of response, adverse event profile of medications, clinical presentation, and preference of the patient.

What If a Patient Does Not Respond to a First-Line Treatment? Since several first-line monotherapy and combination treatment options are available, if one first-line treatment is ineffective, it is appropriate to switch the patient to another first-line treatment. Most manic patients who respond to therapy will show some improvement in manic symptoms within the first 2 weeks. If no improvement is observed during this time period, a switch to a different first-line treatment should be considered. For instance, if a patient is not responding to an atypical antipsychotic monotherapy, a switch to lithium or valproate or vice versa might be appropriate. A switch to a different monotherapy or combination could also be considered for tolerability reasons. In real-world clinical practice, it is unusual to encounter a patient who does not respond to one of the several first-line treatment options that are currently available to treat acute mania.

SECOND-LINE TREATMENTS. These should be considered mainly for those patients who have not responded to several first-line therapies.

Haloperidol is an effective treatment of acute mania. However, haloperidol has a tendency to induce depressive symptoms/switch patients into depression. Therefore, despite excellent efficacy for acute mania, haloperidol is recommended only as a second-line option.

Carbamazepine is also as effective in treating acute mania as other first-line antimanic agents, but tolerability is a significant concern with carbamazepine. Further, carbamazepine induces hepatic microsomal enzymes, which results in the induction of its own metabolism as well as the metabolism of other drugs, and this complicates the management of bipolar disorder. Weight gain and metabolic risk, however, are lower with carbamazepine.

The efficacy of lithium and valproate combination for acute mania is not examined in large double-blind, placebo-controlled trials. Nonetheless, given that both lithium and valproate are effective in monotherapy, it is likely that combination may provide some benefit for patients who are refractory to monotherapy with either agent or a combination of either agent with an atypical antipsychotic.

THIRD-LINE TREATMENTS. If a patient is a nonresponder to all first-line and second-line treatments, third-line options should be considered. These include first-generation antipsychotics such as chlorpromazine, the atypical antipsychotic clozapine, tamoxifen, or neuromodulation treatments such as rTMS.

TREATMENTS NOT RECOMMENDED. Given the efficacy of valproate and carbamazepine for acute mania, clinicians tend to believe that other anticonvulsants are also effective in treating acute mania. However, there is evidence that gabapentin, topiramate, lamotrigine, Licarbazepine, zonisamide, and valnoctamide are ineffective for acute mania and hence should not be used. Brexpiprazole was not effective in RCTs of mania and hence is not recommended for treating acute mania. Similarly, it is not advisable to combine atypical antipsychotics such as risperidone or olanzapine with carbamazepine as no additional benefit in efficacy was identified with this strategy.

ECT in Acute Mania. Although there is a paucity of well-designed studies assessing the efficacy of ECT for acute mania, it is believed that ECT is highly effective for acute mania based on the data from retrospective and some prospective level 3 evidence studies. ECT may be used as a first-line option in the first trimester of pregnancy or in patients with delirious mania where rapid response is required. In other situations, given the cognitive adverse event profile, ECT should be used primarily in only those manic patients who have failed several first and possibly second-line treatments.

Acute Bipolar Depression: Treatment Options

There is a significant unmet need for the treatments of acute bipolar depression as few drugs have proven efficacy.

Lithium. The evidence for the efficacy of lithium monotherapy in treating acute bipolar depression is sparse. Earlier small trials that used cross-over designs with short duration of lithium and placebo periods reported that response rates were significantly higher during lithium periods compared with placebo periods. The mean response rate to lithium in these studies was 76%. About half of the patients experienced a relapse of depression when lithium was substituted with a placebo.

However, the efficacy of lithium was not confirmed in a more recent large double-blind, placebo-controlled parallel design study as lithium was not different from placebo in improving depressive symptoms in bipolar I and bipolar II patients. The mean serum lithium levels in this study were 0.61 mEq/L, and therefore, it is possible that higher levels might have led to greater efficacy. One study that compared lithium versus lithium plus paroxetine showed that the latter was more effective than lithium alone in patients with low serum lithium levels, but in those that had serum lithium levels ≥0.8 mEq/L, there was no difference in efficacy, lending some support to the notion that higher serum lithium levels are required for lithium to be effective for treating acute bipolar depression.

Studies that compared antidepressant/other-agent adjunctive therapy to mood stabilizer with mood stabilizer monotherapy (that included patients on lithium) in general found no differences in efficacy between the two groups. Based on this, it has been suggested that mood stabilizer (lithium) monotherapy is effective. Lithium has been reported to have an antisuicidal effect beyond its mood-stabilizing properties, although prospective randomized controlled data is still required to confirm this. Furthermore, lithium is effective in preventing relapse of depressive episodes, and based on the efficacy of lithium in various phases of bipolar disorder, CANMAT/ISBD guidelines recommend lithium as one of the first-line treatments for bipolar depression.

Anticonvulsants

DIVALPROEX/VALPROATE. Similar to lithium, the data for the efficacy of divalproex in acute bipolar depression is also sparse. Divalproex was superior to placebo in three out of four small placebo-controlled trials. A meta-analysis of these trials suggests that divalproex is effective and that approximately 40% of patients will respond and meet the criteria for remission. It is currently unknown if there is a relationship between serum valproate levels and therapeutic efficacy for improving depressive symptoms.

LAMOTRIGINE. Five double-blind, placebo-controlled studies with reasonable sample sizes of 7 to 10 weeks duration failed on the primary measures to demonstrate the efficacy of lamotrigine for treating acute bipolar depression. Since the lamotrigine dose needs to be titrated up gradually due to concerns about the risk of Stevens–Johnson syndrome with rapid up-titration, it is possible that 7- to 10-week study durations were not sufficient to demonstrate the efficacy given that the target dose of 200 mg was not reached until the beginning of week 5 or 6 in these clinical trials.

A pooled analysis of response rates from these trials suggests that lamotrigine is effective in treating acute bipolar depression, but the magnitude of the effect is small. However, the therapeutic effect was larger in those that had more severe depression.

There is evidence that lamotrigine addition provides a therapeutic benefit in bipolar-depressed patients who fail to respond to lithium monotherapy.

OTHER ANTICONVULSANTS. There is only limited data for carbamazepine as its efficacy was assessed in only three small placebo-controlled trials that had a total of 40 bipolar-depressed patients. The mean response rates to carbamazepine were 68%; relapse occurred in 50% of these patients with placebo substitution.

Gabapentin and levetiracetam are found to be ineffective in small controlled trials. No placebo-controlled data exist for other anticonvulsants.

Atypical Antipsychotics. While all atypical antipsychotics are effective in treating mania, only a few of these have proven efficacy in treating bipolar depression. There is substantial evidence for the efficacy of quetiapine for bipolar I and II depression. However, clinical trials assessed only two fixed doses (300 mg/day and 600 mg/day), and there was no significant difference in efficacy between these two doses. Thus, 300 mg/day of quetiapine is generally considered the target dose for the improvement of depressive symptoms. Although the efficacy of lower doses of quetiapine has not been studied in acute bipolar depression in clinical trials, experience from clinical practice as well as trials in MDD suggests that some bipolar depressed patients are likely to respond to lower doses. Lurasidone monotherapy and adjunctive therapy to lithium or valproate have also been shown to be efficacious in treating acute bipolar depression and have received FDA approval for this indication. While efficacy has been established across the dose range of 20 to 120 mg, the lack of a clear signal for greater efficacy at the higher dose range supports the clinical impression that the 20 to 60 mg dose range may be sufficient for most patients.

Olanzapine was more effective than placebo in clinical trials of bipolar depression, but the magnitude of benefit was small. Olanzapine plus fluoxetine combination is clearly effective and is approved by the FDA for the treatment of acute bipolar depression. Although not tested, it is likely that other SSRIs given in conjunction with olanzapine are also effective.

Cariprazine, a D2 and D3 receptor partial agonist, at 1.5 mg/day was superior to placebo in three double-blind trials and at 3 mg/day in one RCT in improving depression in acute bipolar I depression. Based on these data, cariprazine was recently approved by the FDA for the treatment of acute bipolar I depression. Lumateperone, a novel antipsychotic, has been shown to be effective at 42 mg/day as a monotherapy and in combination with a mood stabilizer in improving depressive symptoms in patients with acute f bipolar I or II depression and the FDA review is currently underway.

Two placebo-controlled RCTs of aripiprazole did not find efficacy for treating acute bipolar depression, although the studies allowed flexible doses between 5 to 30 mg/day, and the mean doses in these studies were around 16 mg/day, which may have been too high. More patients in the aripiprazole group dropped out from these clinical trials, and evidence of aripiprazole efficacy at lower doses for add-on treatment in MDD provides relevant information. Thus, it is conceivable that tolerability related to higher doses may have impacted the efficacy. Ziprasidone monotherapy and adjunctive therapy are not effective in treating acute bipolar depression.

Antidepressants. There is a continuing controversy about the role of antidepressants in treating acute bipolar depression. There is consensus that monotherapy with antidepressant medications is not appropriate for bipolar I depression because of the concerns of manic/hypomanic switch and possible destabilization of the course of bipolar disorder. This advice is consistent with the recent data from a Swedish Registry study which demonstrated that manic switch is about 2.8 times greater in bipolar patients treated with antidepressant monotherapy compared with those taking antidepressants in conjunction with mood stabilizers. There is consensus that the risk of manic/hypomanic switch is higher with TCAs, older MAOIs, and possibly with SNRI antidepressants such as venlafaxine.

While adjunctive antidepressant therapy is widely used by clinicians to treat acute bipolar depression, the evidence for such a strategy from clinical trials is conflicting. Fluoxetine adjunctive therapy to olanzapine is effective, but paroxetine, citalopram, bupropion, and agomelatine adjunctive therapy failed to provide benefit.

However, two out of three meta-analyses support the efficacy of antidepressants. Further, another recent meta-analysis that calculated the NNT for the various treatments of bipolar depression reported an NNT of 5.75 for antidepressants which is in the same range as the NNT of 5.62 for quetiapine and 5.24 for lurasidone. However, these meta-analyses included studies of both antidepressant monotherapy and combination therapy. This limits the applicability of findings from these meta-analyses to real-world practice as most experts do not recommend antidepressant monotherapy for treating acute bipolar I depression. A more recent meta-analysis that specifically examined the efficacy of modern non-SNRI antidepressant adjunctive therapy supports efficacy, but the magnitude of benefit is small. There is uncertainty in the literature regarding the precise switch rates with acute antidepressant use, with results ranging from 0% to 19% in more recent studies and as high as 50% with the TCA desipramine. The ISBD Task Force on Antidepressant Use in Bipolar Disorders provides a balanced consensus view that discourages antidepressant use, especially among patients at higher risk of switch or cycling, but also acknowledges that for some patients, they can play a helpful role in treating acute depressive symptoms. The CANMAT/ISBD Bipolar Guidelines provide a second-line recommendation for add-on antidepressant use in bipolar I depression.

Adjunctive Therapy with Dopamine Agonists/Psychostimulants. There is conflicting evidence for the efficacy of modafinil/armodafinil adjunctive therapy. While modafinil and a phase 2 study of armodafinil were positive, only one out of three phase 3 studies of armodafinil adjunctive therapy supported its efficacy for bipolar depression. Armodafinil was well tolerated in these trials with no increase in manic switches. However, clinical experience suggests that the risk of a switch with modafinil/armodafinil is an important clinical consideration.

Pramipexole adjunctive therapy has been shown to be useful in two small trials of patients with refractory acute bipolar depression, although only one of these included patients with bipolar I depression.

Omega-3 Fatty Acid Adjunctive Therapy. Although there is evidence for the efficacy of adjunctive omega-3 fatty acid therapy in major depression, few studies assessed its efficacy in acute bipolar depression. The results of these studies are conflicting, but a meta-analysis of the data from these trials suggests efficacy. The evidence suggests that potential therapeutic benefits are likely to be derived from eicosapentaenoic acid (EPA) preparation. The omega-3 fatty acid preparations are in general well tolerated; a few patients report gastrointestinal symptoms such as diarrhea and bloating.

Nutritional Supplements as Adjuncts. N-acetylcysteine (NAC) is a glutathione precursor and hence is expected to have protective effects against oxidative stress. There was preliminary evidence that NAC, given at 2 g/day, may provide some benefit in treating depressive symptoms as one out of two studies that initially assessed the efficacy of adjunctive NAC versus placebo reported improvement in depressive symptoms in patients with bipolar disorder that included patients with bipolar I, bipolar II, and bipolar NOS, however recent data have been less promising. There are no significant tolerability issues with NAC.

Glutamatergic Modulators. A growing number of small studies, some of which included treatment-resistant bipolar depressed patients, support the efficacy of the noncompetitive NMDA antagonist ketamine. When infused in subanesthetic doses of 0.5 mg/kg over 40 minutes as adjunctive therapy, ketamine has been reported to rapidly improve depressive symptoms in patients with TRD. A recent systematic review concluded that a single ketamine infusion was associated with higher rates of response and remission on days 1, 3, and 7. The efficacy appears to be greater in those with unipolar depression than in those with bipolar depression. Ketamine may have antisuicidal effects. There is preliminary evidence that the benefits of ketamine could be maintained with further infusions or IM or oral or sublingual or intranasal administration of ketamine. However, the administration of riluzole, another glutamatergic modulator, was not more effective than placebo in maintaining the antidepressant effects of ketamine. Ketamine infusion is associated with transient psychotomimetic effects as well as sedation, dizziness, lightheadedness, incoordination, blurry vision, and feeling “strange or unreal.” The intranasal enantiomer esketamine is now available. However, extensive data on bipolar depression is lacking.

Limited open-label data suggest that riluzole at doses of 50 to 200 mg may be useful in treating bipolar depressive symptoms.

Somatic Treatments

ELECTROCONVULSIVE THERAPY. There is consensus among the experts that ECT is one of the most effective treatments for acute bipolar depression. Response rates to ECT are in the range of 50% to 70% in refractory major depressive disorder and bipolar depression. A recent meta-analysis concluded that there is no difference in efficacy of ECT between major depressive disorder and bipolar depression. There is some evidence that patients with bipolar depression respond faster and require fewer treatments to improve compared with those with unipolar depression. It is not surprising that a recent study found that ECT is more effective than algorithm-based pharmacologic treatments in patients with bipolar depression who have failed at least two previous pharmacologic treatments. Most patients with bipolar depression require six to eight treatments to respond.

Maintenance pharmacotherapy is necessary for the vast majority of patients to prevent relapse of depressive and manic episodes. This should be commenced, when possible, prior to completion of the ECT course. While ECT is very effective, cognitive side effects can often be problematic. Therefore, clinicians need to pay particular attention to various ECT stimulation parameters such as electrode placement, pulse width, and electrical dose in order to minimize cognitive side effects.

MAGNETIC SEIZURE THERAPY. Magnetic seizure therapy (MST) is a relatively new technique that involves inducing seizures through magnetic stimulation of the brain. There is preliminary uncontrolled evidence that MST is effective in treating depression. However, few patients with bipolar depression were included in these reports. Further, while MST appears to be effective, head-to-head comparison with ECT suggests that response and remission rates are better with ECT compared with MST. It is conceivable that further refinements in stimulation parameters of MST might lead to better efficacy. As expected, MST is associated with a much safer cognitive side effect profile as memory, language, and practice seem to be relatively unaffected by MST.

TRANSCRANIAL MAGNETIC STIMULATION. The data with regard to the efficacy of rTMS in acute bipolar depression is limited, although there is extensive literature supporting its efficacy in treating major depressive disorder. A meta-analysis that included data on 181 patients with different subtypes of bipolar disorder and different stimulation parameters concluded that overall, there is support for the efficacy of rTMS in treating bipolar depression with an NNT of 6. However, two small recent RCTs that assessed the efficacy of theta burst stimulation in improving depressive symptoms in bipolar depression failed to support the efficacy, thus raising questions about its utility in treating acute bipolar depression. However, given the small number of studies and differences in methodology, clearly, further controlled trials are needed to fully understand its role in treating bipolar depression.

TRANSCRANIAL DIRECT CURRENT STIMULATION (TCDS). Transcranial direct current stimulation (tCDS) involves the application of low-amplitude direct current to stimulate brain areas through surface scalp electrodes. There is emerging evidence from a systematic review that included seven reports and a recent sham-controlled randomized trial supporting the efficacy of tCDS in treating depressive symptoms in bipolar disorder. The sham-controlled trial showed that response rates were 68% in tCDS group compared with the 30% response rate in the sham treatment group, with an NNT of 2.69. This strategy may be considered for patients with bipolar depression who do not respond to standard first and second-line treatments or those that have difficulties in tolerating the side effects of medications.

Clinical Recommendations for the Management of Acute Bipolar Depression

Treatment goals in a patient with bipolar depression include treating depression to full remission without destabilizing the course of bipolar disorder while ensuring the safety of the patient. The risk of suicide is higher during acute depressive episodes and especially so in depressed patients with mixed features. Therefore, clinicians must carefully assess for suicide risk, and appropriate steps must be taken to ensure the safety of the patient. The severity of depression, presence of mixed features or psychotic features, substance abuse or other psychiatric comorbidities, and the severity of suicidal ideation should guide the clinician with regard to the most appropriate treatment setting for managing depression in each patient.

Adjunctive psychological treatments such as CBT, IPSRT, and family-focused therapy (FFT) are beneficial in reducing time to recovery, improving treatment adherence and recovery rates, and reducing subsyndromal symptoms in bipolar depressed patients and, therefore, form an important part of the optimal management of bipolar depression.

Treatment Selection for Acute Bipolar Depression. Many patients with bipolar disorder experience breakthrough depressive episodes. This may be because patients had been taking inadequate doses of maintenance medication, were partially or fully nonadherent to medications, or the maintenance medications were ineffective in preventing the depressive recurrence or relapse. A thorough clinical assessment that gathers information about the course of bipolar disorder (e.g., number of previous manic and depressive episodes, their duration, and severity), the treatment history (current and previous medications, what worked and what did not work, tolerability issues), and adherence to treatment should help the clinician in determining the most appropriate treatment choices for each given patient.

First-Line Treatments. The evidence for the efficacy of many commonly used medications for treating depression in bipolar disorder is less robust and often conflicting. Therefore, it is not surprising that there is no general consensus among experts about the most appropriate first-line treatment options for bipolar depression. The CANMAT/ISBD group that published the most recent treatment guidelines took into account several factors such as efficacy from controlled trials, meta-analyses, clinical experience, the safety of the medication, and long-term efficacy in arriving at the first-line treatment recommendations for acute bipolar depression. A modified list of CANMAT recommendations for bipolar depression is provided in Table 13.6–7.

WHICH FIRST-LINE TREATMENT? If a patient had been taking one of the first-line treatments, optimizing the current treatment is the most appropriate first step before other interventions are considered. Similarly, first-line treatment should be commenced in patients who are drug-free or not taking first-line medications. While there is no clear research evidence to guide clinicians between various first-line monotherapy options, previous treatment history, illness course, symptom profile of depression, side effect profile of medications, and patient preferences are all important considerations in determining treatment choices among the first-line therapies. The 2018 CANMAT/ISBD guidelines also introduced a hierarchical approach that emphasized the importance of long-term treatment planning even for the management of acute episodes. This hierarchical approach encouraged incorporating all aspects of the levels of evidence across all phases of illness, safety and tolerability concerns during the acute and maintenance phases, and risk of a switch when arriving at the sequence of interventions. Given that quetiapine has shown efficacy in acute mania, acute bipolar depression, and prevention of both mania and depression, it is recommended at the top of the hierarchy of first-line treatments. However, patient-specific factors such as certain clinical features, previous history of response or nonresponse, preference for a medication based on the tolerability profile of an agent may override such hierarchy. For instance, lithium monotherapy may be more appropriate for patients with a history of classical manic episodes, an illness course that has mania, depression, and euthymia patterns, and a family history of bipolar disorder. Bipolar I patients with frequent depressions and infrequent mild manias are suitable for lamotrigine monotherapy, while quetiapine is a better option for those with frequent depressions as well as manic episodes. Those with reverse vegetative symptoms such as hypersomnia and hyperphagia may be reluctant to take quetiapine because of its sedating and appetite-stimulating properties, although there is no reason to believe that it is less effective for this population. Lamotrigine dose needs to be escalated gradually due to concerns of skin rash with rapid dose escalation, and hence although it is effective in severely depressed patients, it may not be the most appropriate option given that it takes a few weeks to reach an effective therapeutic dose. Lurasidone is relatively weight-neutral and is preferred by patients who are reluctant to take medications that can cause weight gain. While the pharmacologic profile of this agent suggests that it may prevent mood episodes, given that the only maintenance study failed to demonstrate the efficacy, it may be prudent to consider combining lurasidone with lithium or valproate. Cariprazine has shown efficacy in monotherapy at 1.5 mg daily, which can be the starting and therapeutic dose for this agent, and it appears to have efficacy in treating a broader group of bipolar-depressed patients. As well, it is also relatively weight neutral and has minimal impact on blood sugar or lipid levels. Thus, it is also an excellent option for bipolar-depressed patients. While the pharmacologic profile of this agent suggests that it might prevent mood episodes, the RCT data are not available as yet; hence, it may be prudent to monitor patients on cariprazine monotherapy closely for the emergence of any mood symptoms and add mood stabilizer therapy as appropriate.

Table 13.6–7.
Recommendations for Pharmacologic Treatment of Acute Bipolar I Depression

First-line	**Monotherapy:** Quetiapine/quetiapine XR, lithium, lamotrigine, lurasidone, cariprazine **Combination therapy:** Lithium or divalproex + lurasidone, adjunctive lamotrigine
Second-line	**Monotherapy:** Divalproex, ECT[a], lumateperone **Combination therapy:** Adjunctive SSRI/bupropion to a mood stabilizer or atypical antipsychotic, adjunctive lumateperone
Third-line	**Monotherapy: Carbamazepine, Olanzapine** **Combination therapy:** Adjunctive aripiprazole, adjunctive armodafinil/modafinil, adjunctive asenapine, adjunctive eicosapentaenoic acid (EPS), adjunctive ketamine (IV), adjunctive light therapy ± total sleep deprivation, adjunctive levothyroxine, adjunctive pramipexole, adjunctive tCDS, adjunctive SNRI/MAOI antidepressant
Not recommended	**Monotherapy:** Antidepressant, Gabapentin, aripiprazole, ziprasidone **Combination therapy:** Adjunctive ziprasidone, adjunctive levetiracetam, adjunctive mifepristone

[a] Can be a first-line treatment of severe depression with suicidal ideation or in the first trimester of pregnancy.

SSRI, selective serotonin reuptake inhibitor; MAOI, monoamine oxidase inhibitor; ECT, electroconvulsive therapy;

Modified with permission from Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord.* 2018;20:97–170. doi: 10.1111/bdi.12609. © 2018 The Authors. Published by John Wiley & Sons Ltd.

WHAT IF A PATIENT DOES NOT RESPOND TO A FIRST-LINE TREATMENT? If optimization of monotherapy with a first-line agent is ineffective, a switch to another first-line monotherapy therapy (e.g., a switch from lithium to quetiapine or vice versa or to cariprazine or lurasidone) or combination therapy (e.g., lamotrigine plus lithium or lurasidone or cariprazine plus a mood stabilizer) is appropriate.

Second-Line Treatments. Lumaterperone has excellent safety and tolerability profile and hence is a good option for patients that have trouble tolerating first-line agents for bipolar depression. The addition of an SSRI or bupropion could be considered for bipolar depressed patients who were unresponsive to optimization of lithium or valproate monotherapy, especially if there is a previous history of response to adjunctive antidepressant therapy. However, antidepressants are not appropriate for patients with a history of rapid cycling or those with mixed features. If antidepressants are used for treating acute bipolar depression, an attempt should be made to taper and discontinue them within 8 weeks of remission of depression.

ECT should be considered for patients who have failed several first-line treatments or for whom the rapid onset of response is paramount (e.g., in the presence of catatonia). ECT may be considered earlier for bipolar depressed patients with psychotic features, those at high risk of suicide, the older adults, and those with rapidly deteriorating physical status due to poor oral intake because of depression.

Valproate monotherapy is appropriate for bipolar depressed patients who had tolerability issues with other first-line monotherapy options. While no data exists, it may also be useful for bipolar depression with mixed features. Olanzapine-fluoxetine combination would also be a reasonable option for patients who remain unresponsive to treatment. Antidepressant adjunctive therapy to other atypical antipsychotics such as aripiprazole or quetiapine may be helpful for some patients.

Third-Line Treatments. These options are only appropriate for patients who failed many first- and second-line options. If these strategies are employed, clinicians need to educate patients and their family members about the limited evidence base to support their use.

Treatments Not Recommended. Treatment with antidepressant monotherapy is not recommended. As noted earlier, aripiprazole monotherapy also has negative data for acute bipolar I depression treatment. However, it can be effective in treating and preventing mania, and hence, adding an SSRI or bupropion to aripiprazole could be an acceptable strategy in patients who were refractory to all the other first and second-line treatments.

Maintenance Therapy/Prophylaxis: Treatment Options

Lithium. Lithium is still considered the “gold standard” by many experts for the maintenance treatment of bipolar disorder. Lithium is effective in preventing mood episodes not only in bipolar I patients who had responded to lithium during acute episodes (enriched for lithium response) but also in those who had been treated with a different psychotropic medication for acute mood episodes (not enriched for lithium response). Lithium is effective in preventing both manic and depressive episodes, although the magnitude of benefit may be slightly lower for depression than it is for preventing mania. While earlier studies suggested that around 70% of patients with bipolar disorder remained episode-free with lithium, more recent studies suggest that the response rates are probably below 50%. This may be due to several reasons, including the cohort effect and the fact that these studies recruited patients who were responders to other psychotropic medication during acute episodes.

There is evidence that lithium plus valproate combination is more effective than valproate alone in preventing relapse of mood episodes. In bipolar patients who are refractory to therapy with lithium or carbamazepine, the combination appears to be helpful in reducing the risk of relapse of mood episodes.

It is recommended that lithium is given as a single daily dose as there is some evidence that multiple daily doses of lithium are more commonly associated with a greater incidence of renal side effects and pathologic changes in kidneys. Further, treatment adherence is likely better with a single daily dose.

There is evidence that serum lithium levels at or below 0.6 mEq/L are less effective. Therefore, for prophylaxis, serum lithium levels should be maintained between 0.8 and 1.2 mEq/L. For older adults with bipolar disorder, there is expert consensus that lower levels are preferred. Serum levels of 0.4 to 0.8 mmol/L are recommended for those ages 60 to 79 years, and levels of 0.4 to 0.7 mmol/L for those ages 80 or above. Lithium levels should be measured after the target dose has been achieved, in the steady-state (approximately 5 days after the last dose change), about 12 hours after the last dose. Serum levels should be repeated until two consecutive levels within the therapeutic range are obtained for the same dose.

Patients receiving maintenance lithium therapy should be monitored as per modified ISBD safety monitoring guidelines (see Fig. 13.6–1) or more frequently if clinically indicated.

Patients on maintenance treatment of lithium should be educated about symptoms of lithium toxicity and the preventive measures. Risk factors for lithium toxicity include renal diseases, dehydration, drug interactions, older adults, and those with organic brain syndromes. Lithium toxicity is a medical emergency, and if not treated, can result in irreversible organ damage and death.

If lithium is to be discontinued for any clinical reason in bipolar patients whose serum lithium levels had been in a therapeutic range, it should be tapered and discontinued over a period of several weeks. This is because there is evidence that the time to 50% risk of recurrence of mood episodes is four times shorter with abrupt discontinuation (1 to 14 days) compared with gradual (15 to 30 days) discontinuation of lithium.

Anticonvulsants

CARBAMAZEPINE. There is limited data for the efficacy of carbamazepine from placebo-controlled trials. However, most, although not all, controlled studies that compared carbamazepine with lithium showed similar efficacy. While no study showed that carbamazepine was more effective than lithium, there is some evidence that carbamazepine may be more effective than lithium in patients with “nonclassical bipolar disorder” (i.e., patients with bipolar II disorder/bipolar NOS or with mood-incongruent delusions or with anxiety disorders or substance use comorbidity or mixed states).

The clinical trials have not specifically examined the relationship between serum carbamazepine levels and therapeutic efficacy in maintenance treatment. Hence, it is currently unknown what levels might offer optimal prophylactic efficacy. In clinical practice, serum carbamazepine levels are typically maintained between 4 and 12 μg/L, the same range used for treating patients with epilepsy.

Carbamazepine induces hepatic microsomal enzymes that metabolize carbamazepine. As a result, within the first few weeks of commencing carbamazepine, the clearance of carbamazepine increases by approximately twofold; hence, dosage adjustments are necessary in order to keep serum carbamazepine levels within the clinically accepted range. Carbamazepine causes leukopenia in 10% to 20% of patients, and it occurs commonly within the first 3 months of commencing treatment and resolves with discontinuation of carbamazepine. Agranulocytosis and aplastic anemia are rare but have a more rapid onset and occur in an unpredictable pattern. Therefore, periodic clinical monitoring is not likely to be of much help. Clinicians need to be vigilant to this possibility when hematologic symptoms emerge clinically.

There are no controlled data to support the efficacy of oxcarbazepine in the maintenance treatment of bipolar disorder.

VALPROATE. In contrast to lithium which has robust data for its efficacy in maintenance treatment, only a single large double-blind, controlled trial assessed the efficacy of valproate in the maintenance treatment of bipolar disorder. In this trial, valproate was not superior to placebo on the primary measure of time to any mood episode. This study was considered a failed trial for a variety of reasons, including the fact that lithium was also not superior to placebo in preventing relapse of mood episodes. Valproate, interestingly, was superior to placebo on some secondary measures, including time to discontinuation for a recurrent mood episode or depressive episode. In the subgroup of patients who were treated with valproate during the open-label phase, valproate was superior to placebo in time to any mood episode. Further, patients randomized to valproate were significantly less likely to drop out for a mood episode and less likely to be prematurely terminated for any reason. Thus, in aggregate, these data suggest that valproate may be beneficial for preventing mood episodes in patients who respond to valproate during acute episodes.

A randomized controlled study found no difference in efficacy between valproate and lithium in patients with rapid cycling bipolar disorder who were stabilized on a combination of valproate plus lithium. Consistent with this, the results of a Cochrane review indicated that valproate was more effective than placebo in preventing study withdrawal due to any mood episode and that there was no difference in efficacy between lithium and valproate. There is evidence that fewer patients drop out of treatment with valproate compared with placebo or lithium. This may explain why valproate is widely used for maintenance treatment.

Although there is a linear relationship between serum levels of valproate and improvement in acute manic symptoms, there is no data to guide the clinicians with regard to serum valproate levels for the maintenance treatment of bipolar disorder. Therefore, it is a common practice to maintain the same valproate dose that was effective in treating mania for maintenance treatment, provided serum levels are within the range used for epilepsy (45 to 125 µg/mL).

Most patients prefer to take valproate as a single bedtime dose. Valproate can cause nausea and vomiting, particularly in the first few days of commencement of therapy. Some patients experience weight gain and hair loss. Women taking valproate should be monitored for polycystic ovary syndrome with questions about menstrual irregularities, fertility issues, hirsutism, and galactorrhea.

LAMOTRIGINE. Lamotrigine is approved for prophylaxis of mood episodes in bipolar I disorder by the FDA. Lamotrigine has robust efficacy in preventing depressive relapses/recurrences, but it only has marginal efficacy in preventing manic episodes. There is no data to suggest a relationship between serum levels of lamotrigine and prophylactic efficacy.

A large randomized open study that compared lamotrigine with lithium found no statistically significant differences in efficacy, but as expected, the hazard ratios for lamotrigine relative to lithium were higher (1.91) for mania and lower (0.69) for depression. In rapid cycling bipolar disorder, lamotrigine was no more effective than placebo on the primary measure of time to additional pharmacotherapy, but lamotrigine patients stayed significantly longer in the study compared with those that received placebo.

Lamotrigine is titrated upward slowly due to concerns about skin rash with rapid titration. Other risk factors for skin rash include younger age, a prior history of rash with another drug, and taking lamotrigine in conjunction with valproate. Rash more commonly occurs within the first few weeks of therapy. The incidence of benign skin rash is about 10% in prospective studies. The incidence of serious skin rash (i.e., Stevens–Johnson syndrome or toxic epidermal necrolysis) is rare (about 1 to 10 per 10,000) with slow dose titration. Nearly 50% of serious skin rash are seen in patients who receive lamotrigine in conjunction with valproate.

If a patient on lamotrigine therapy develops a skin rash, this needs to be evaluated carefully to determine if it is a benign or serious rash. When in doubt, if immediate dermatologic consultation is not available, lamotrigine should be discontinued, and the patient should be carefully monitored. Those with serious skin rash require urgent medical evaluation and treatment.

Apart from the cutaneous reactions described above, lamotrigine does not cause weight gain and is in general well tolerated.

OTHER ANTICONVULSANTS. Gabapentin is not effective for treating acute mania or bipolar depression. In a small controlled trial, surprisingly, gabapentin adjunctive therapy was beneficial in maintaining mood stability. There are no controlled maintenance data for topiramate, levetiracetam, felbamate, or other anticonvulsants.

Antipsychotics

TYPICAL (CONVENTIONAL) ANTIPSYCHOTICS. Conventional antipsychotics are likely effective in preventing manic episodes. However, there is evidence that conventional antipsychotic therapy is associated with an increased risk of depressive symptoms and depressive episodes. Further, patients with bipolar disorder are at higher risk of developing tardive dyskinesia with conventional antipsychotics. Therefore, therapy with conventional antipsychotics, including depot preparations, should be mainly reserved for those patients who have failed treatment with traditional mood stabilizers and atypical antipsychotics.

ATYPICAL ANTIPSYCHOTICS. Atypical antipsychotics are widely used for treating acute mania or depression, and most available agents in North America have proven efficacy from double-blind, controlled trials. If a manic patient responds to an atypical antipsychotic agent, whether continuation of that atypical antipsychotic is effective in preventing relapse of mood episodes is an important clinical question.

From the longer-term maintenance perspective, both oral aripiprazole monotherapy (10 to 30 mg/day) and aripiprazole once-monthly injection (400 mg) are effective in preventing mood episodes and manic episodes but not depressive episodes. Similarly, aripiprazole adjunctive therapy to lithium or valproate is also beneficial in preventing manic episodes.

Continuation of olanzapine in manic/mixed episode patients who respond to olanzapine is effective in preventing manic, mixed, and depressive episodes, but the magnitude of benefit appears to be greater for preventing mania and mixed episodes relative to depression. Olanzapine appears to be as effective as lithium and valproate in preventing mood episodes. Olanzapine adjunctive therapy to lithium or divalproex appears to be useful in preventing symptomatic relapse of mood episodes.

While oral risperidone has not been assessed for its efficacy, manic patients who responded to oral risperidone and switched to risperidone long-acting injectable (Risperdal Consta) are less likely to experience a relapse of mood episodes. Like aripiprazole, Risperdal Consta is effective in preventing manic episodes but not depressive episodes. Most patients require about 25 mg of Risperdal Consta every 2 weeks for preventing manic episodes. Risperdal Consta adjunctive therapy is also effective in preventing relapse of mood episodes and manic episodes but not depressive episodes in frequently relapsing bipolar patients.

Paliperidone ER monotherapy (mean dose of 6 mg/day) continuation in manic patients who responded to paliperidone is effective in prolonging time to and reducing the risk of relapse of mood episodes and manic episodes but not depressive episodes. In schizoaffective patients, paliperidone palmitate once-monthly (between 78 and 234 mg monthly) given as either monotherapy or adjunctive therapy is also beneficial in reducing the risk of relapse of manic, depressive, and psychotic symptoms.

Quetiapine is effective in preventing mood episodes, manic/mixed episodes, and depressive episodes in bipolar patients who responded to quetiapine during acute mania or bipolar depression. Quetiapine appears to be equally effective in preventing both mania and depression. The mean doses of quetiapine in the maintenance studies ranged from 527 to 562 mg/day. Quetiapine adjunctive therapy to lithium or divalproex is also effective in preventing both manic and depressive episodes, and the magnitude of benefit of adjunctive therapy for prevention of mania and depression is similar with hazard ratios around 0.3. This means that there is approximately a 70% reduction in risk of manic and depressive episodes with quetiapine adjunctive therapy compared with lithium or valproate monotherapy.

A double-blind placebo-controlled trial showed that asenapine monotherapy was associated with a longer time to recurrence of any mood episode in recently stabilized manic/mixed episode patients. Asenapine was significantly more effective than placebo in preventing relapse of both manic and depressive mood episodes; however, similar to olanzapine, the magnitude of prophylactic benefit with asenapine was greater for preventing mania than in preventing depression.

Lurasidone adjunctive therapy reduced the risk of recurrence of a mood episode by 29% relative to placebo adjunctive therapy, but this difference just failed to reach statistical significance. Lurasidone adjunctive therapy, however, was superior to placebo adjunctive therapy in those who did not have rapid cycling BD and in patients with an index episode of depression in reducing the risk of recurrence of any mood episode.

While ziprasidone monotherapy is effective in treating acute mania, there is no data to guide clinicians with regard to whether the continuation of ziprasidone in such patients is useful in preventing mood episodes. However, ziprasidone adjunctive therapy to lithium or valproate is effective in preventing mood episodes and manic episodes but not depressive episodes.

There are no long-term data as yet for cariprazine, but a large RCT is currently underway, and the results are awaited in 2022.

Clozapine is likely effective at least in preventing mania, but no controlled studies have been conducted to date. In refractory bipolar/schizoaffective patients, clozapine add-on has been reported to be beneficial in reducing manic and psychotic but not depressive symptom scores on rating scales.

ANTIDEPRESSANTS. There is evidence that TCAs increase the risk of manic switch and are not recommended for treating acute bipolar depression. The maintenance studies that compared imipramine with lithium or lithium plus imipramine confirmed an increased risk of a manic switch with both imipramine mono and adjunctive therapy and hence TCAs are clearly not appropriate for maintenance therapy of bipolar disorder.

If modern antidepressants are used as adjuncts to treat acute bipolar depression, is the continuation of antidepressant therapy beneficial? This issue was addressed in a Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) trial that randomized patients who responded to adjunctive antidepressant therapy to continuation for 1 to 3 years versus discontinuation within 2 weeks. The results showed that the continuation of antidepressant therapy was associated with a trend toward less depressive symptom burden without increasing manic/hypomanic symptoms. Further, the time to a depressive episode was significantly delayed in the antidepressant continuation group. There was no evidence of an increase in manic episode risk. Interestingly, a rapid-cycling course predicted a threefold increase in depressive episodes in patients who continued antidepressants versus those who discontinued. Therefore, clinicians are advised to carefully assess the risks and benefits of the need for long-term antidepressant adjunctive therapy for each patient and tailor the treatment accordingly.

Clinical Recommendations for Maintenance Treatment/Prophylaxis of Bipolar I Disorder

Most experts agree with the need for long-term prophylaxis in all patients with bipolar I disorder, and that prophylaxis should begin after the first manic episode. This is because bipolar disorder is a highly recurrent illness, frequently characterized by residual symptoms and chronicity. Over 90% of patients have a recurrence of a mood episode within 5 years. A recent meta-analysis suggests that recurrence rates are lower in first manic episode patients, with about 60% experiencing a mood episode within 4 years. Bipolar I patients experience syndromal or subsyndromal symptoms an average of 50% of the time, with depressive symptom periods predominating by a ratio of 3:1 to manic/hypomanic symptoms. There is evidence that subsyndromal symptoms increase the risk of full-blown mood episodes. Multiple mood episodes are more likely to be associated with cognitive impairment, structural brain changes, and functional impairment. Frequent episodes alter the response to treatment such that those that have had multiple mood episodes are less likely to respond to treatment during acute mood episodes as well as to prophylactic pharmacotherapy and adjunct psychotherapy.

Therefore, the goals of the long-term treatment include prevention of mood episodes, prevention of/improvement in subsyndromal symptoms, prevention of suicidal behavior, and improvement in treatment adherence, quality of life, cognitive, social, and work function.

While pharmacotherapy is the mainstay of the long-term prophylactic treatment of bipolar disorder, there is evidence that adjunct psychological treatments are helpful in reducing the risk of relapse in bipolar patients. Group or individual psychoeducation is effective in reducing relapse rates by 25% to 30%. Patients should be educated about the illness, adverse effects of medication, early warning symptoms of mood episodes, triggers for mood episodes such as substance use and psychosocial stressors, as well as coping strategies. IPSRT and FFT are also effective. Bipolar patients who recovered from mood episodes tend to benefit from adjunct individual CBT that reduces relapse rates and prolongs time to mood episodes. The STEP-BD study reported that recovery rates at 1 year were significantly greater in acute bipolar depressed patients who received CBT, IPSRT, or FFT compared with those that received collaborative care. Thus, adjunct psychological treatments are essential components of the long-term treatment of bipolar disorder to improve treatment adherence and mood symptoms and reduce relapse rates.

Treatment Selection for Maintenance Treatment. Several factors are important in aiding the treatment selection for the maintenance treatment. These include information about the course of the illness in the patient, response (lack of response) and tolerability to various treatments during acute episodes and maintenance treatment, predictors of response to various medications, the predominant polarity of the illness, comorbidity, and whether the most recent episode was manic or depressive and what treatment/s were effective in remitting this episode. Family history of bipolar disorder, if present, and treatment response in family members are also helpful in some cases.

FIRST-LINE TREATMENTS. Lithium, valproate, lamotrigine, and some atypical antipsychotics are effective and are recommended as first-line agents for the maintenance treatment of bipolar disorder (see Table 13.6–8). However, it is important to note that the benefits of these agents in reducing relapse rates have been demonstrated in bipolar patients who were responders to acute therapy with the same agents. The only exception to this is lithium that has been shown to be effective in the maintenance treatment regardless of whether patients were treated for an acute episode with lithium or not.

Which First-Line Treatment? If a patient was recently treated for an acute manic episode with monotherapy with one of the first-line agents listed in Table 13.6–8, it is advisable to continue the same medication for the maintenance treatment. Although risperidone and ziprasidone are first-line agents for acute mania, they are only listed as third-line treatments for maintenance therapy as there are no efficacy data for maintenance therapy of oral formulations with these agents. However, given that the other atypical agents that are effective in mania are also effective in maintenance treatment, clinicians may choose to continue these agents for the maintenance treatment after discussing the potential benefits and risks with such a strategy. Although paliperidone ER (>6 mg) is effective in maintenance treatment, it was less effective than olanzapine, and hence, it is listed as a second-line treatment for maintenance, but in manic patients who were responders to paliperidone ER, clinicians may also choose to continue this medication for prophylaxis.

Table 13.6–8.
Recommendations for Maintenance Pharmacotherapy of Bipolar Disorder

First-line	**Monotherapy:** Lithium, quetiapine, divalproex, lamotrigine (limited efficacy in preventing mania), asenapine, aripiprazole oral or LAI (mainly for preventing mania) **Adjunctive therapy with lithium or divalproex:** Quetiapine, aripiprazole (mainly for preventing mania)
Second-line	**Monotherapy:** Olanzapine[a], risperidone LAI (mainly for preventing mania), Carbamazepine, paliperidone ER (>6 mg), paliperidone LAI **Combination therapy:** Adjunctive risperidone LAI, lurasidone added to lithium or divalproex, ziprasidone added to lithium or divalproex (mainly for preventing mania), Lithium + divalproex, lithium + carbamazepine, lithium or divalproex + atypical antipsychotic
Third-line	**Monotherapy: Ziprasidone[b], risperidone[b]** **Adjunctive therapy:** Aripiprazole and lamotrigine, adjunctive clozapine, olanzapine plus fluoxetine, ECT
Not recommended	**Monotherapy:** Gabapentin, topiramate, antidepressants, perphenazine

[a] Given the metabolic side effects, use should be carefully monitored.

[b] Manic patients who responded to these, continuation is appropriate as a first-line.

LAI, long-acting injection; ECT, electroconvulsive therapy; ER, extended-release.

Modified with permission from Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord.* 2018;20:97–170. doi:10.1111/bdi.12609. © 2018 The Authors. Published by John Wiley & Sons Ltd.

If a patient's illness course is characterized by predominant depressive polarity, clinicians should consider adding lamotrigine or lithium to atypical antipsychotic agents (such as aripiprazole, risperidone, or ziprasidone). Alternatively, they could replace these agents with either lithium, quetiapine, asenapine, or lamotrigine for prophylaxis.

For manic patients who were treated with a combination of lithium or valproate with an atypical antipsychotic, there is evidence that continuation of combination therapy is more effective than lithium or valproate monotherapy for preventing relapse of mood episodes. A recent study that compared different durations of atypical antipsychotic adjunctive therapy with risperidone or olanzapine for the maintenance treatment reported that adjunctive therapy is beneficial for the first 6 months after remission of mania, but continuation beyond this period was associated with significant weight gain liability with no apparent additional benefit in efficacy. Hence, clinicians are advised to carefully assess the need for the continuation of atypical antipsychotic adjunctive therapy beyond 6 months after mania remission. For instance, a longer duration of adjunctive therapy may be appropriate in bipolar patients with a previous history of insufficient prophylaxis with lithium or valproate monotherapy.

Patients who respond to treatment with quetiapine or lithium for acute bipolar depression are advised to continue these agents for maintenance treatment of bipolar disorder. In a bipolar depressed patient treated with lurasidone, clinicians may choose to continue this medication for prophylaxis despite lack of maintenance data given its D2 blockade (likely confers acute and prophylactic antimanic efficacy) properties, especially given the fact that lurasidone adjunctive therapy was effective in preventing relapse of mood episodes in those with an index depressive episode. Lamotrigine monotherapy is effective mainly in preventing depression, and hence in bipolar patients with previous severe manic episodes, clinicians are well-advised to add a prophylactic antimanic agent such as lithium or an atypical antipsychotic for optimal prophylaxis. Bipolar depressed patients who respond to cariprazine are also advised to continue this agent for maintenance treatment given its efficacy in treating acute mania and acute bipolar depression despite lack of maintenance data, while they are being monitored closely for emergence of mood episodes.

If a bipolar depressed patient responded to adjunctive modern antidepressant therapy, is the continuation of adjunctive therapy with antidepressant beneficial? There is preliminary evidence that continuation in nonrapid cycling bipolar patients is beneficial in reducing the risk of depressive relapses without destabilizing mood. The CANMAT/ISBD guidelines recommend tapering and discontinuing antidepressants about 8 weeks after remission of depression, based on the assumption that antidepressants are primarily needed to cover the natural duration of the depressive episode, and beyond this, there is a risk of potential destabilization. The data from a recently completed Canadian trial are awaited for more definitive guidance on this issue.

In making a decision with regard to the maintenance medication for bipolar patients who are currently not on any treatment, clinicians need to consider several factors outlined in the treatment selection section above. If lithium has not been tried previously, given the robust evidence for efficacy in preventing both mania and depression and possibly suicidal behavior, it is the most appropriate first-line agent for such patients. Quetiapine monotherapy also has a wealth of evidence and clinical experience as a maintenance treatment. Patients with a history of alcohol or substance abuse or mixed features may be more appropriate for valproate therapy. For those with predominant manic polarity, any first-line agent with the exception of lamotrigine would be appropriate.

What If a First-Line Agent Is Ineffective? If a patient experiences breakthrough subsyndromal symptoms or relapse while on monotherapy with one of the first-line agents, a switch to an alternative first-line agent or a combination therapy is appropriate. For instance, if a patient on lithium experiences depressive symptoms, consider adding lamotrigine, lurasidone, or quetiapine. If manic symptoms emerge on lithium, consider adding an atypical antipsychotic or valproate or if there are tolerability issues, consider a switch to an alternative first-line agent. It is recommended that monotherapy or combination therapy with several first-line agents be tried before embarking on second-line strategies for maintenance therapy.

SECOND-LINE TREATMENTS. Olanzapine has strong data supporting maintenance use for the prevention of both mania and depression. However, given greater liability with regard to weight gain and metabolic effects, it should be considered after the series of first-line agents. Given that the evidence for prophylactic efficacy of carbamazepine and paliperidone is less robust, these options are mainly considered for bipolar patients who failed or had tolerability issues to monotherapy with first-line agents. Carbamazepine does not generally cause weight gain and is effective in preventing relapses, particularly in patients with atypical forms of bipolar disorder.

However, many patients who failed several first-line agents are unlikely to respond to monotherapy. Hence, combination therapy strategies outlined in Table 13.6–8 are most appropriate for such patients.

THIRD-LINE TREATMENTS. Given the efficacy of cariprazine for acute mania and depression, the expert opinion is that it is likely effective in maintenance treatment and should be considered if other first and second-line options have failed to prevent mood episodes. Clinical experience and some controlled data suggest that clozapine adjunctive therapy is very effective in refractory bipolar disorder. However, in addition to the risk of agranulocytosis, clozapine is associated with significant weight gain and metabolic side effects, and hence patients need to be monitored closely for these adverse events and treated appropriately. For some refractory patients, maintenance ECT is an option. Aripiprazole and lamotrigine combination has some evidence for maintenance treatment.

Treatment of Bipolar II Disorder

Although bipolar II disorder is common and associated with significant morbidity and mortality, few clinical trials investigated the efficacy of treatments for various phases of bipolar II disorder. The limited data that is currently available is derived from studies of bipolar I disorder or major depressive disorder that had included a few patients with bipolar II disorder. Therefore, in the absence of controlled clinical trial data, by necessity, extrapolation is made from the treatment studies in bipolar I disorder to aid clinical decisions in managing bipolar II disorder.

Treatment of Hypomania. It is generally believed that medications that are effective in treating acute mania are also effective in treating hypomania. Indeed, small placebo-controlled trials or open-label studies demonstrated that valproate, quetiapine, and risperidone are effective in treating hypomanic symptoms. These findings are consistent with the clinical experience of experts in the field.

Thus, bipolar II patients with persistent hypomanic symptoms could be treated with an antimanic agent to rapidly alleviate hypomanic symptoms. Prior history of treatment response, adverse effects of medications, and patient preferences are important considerations in treatment selection. If a patient had been taking a prophylactic antimanic agent, an increase in the dose of that medication might be sufficient in some cases.

Medications that could potentially be contributing to hypomanic symptoms, such as antidepressants, should be discontinued.

Treatment of Bipolar II Depression. The data for the efficacy of various treatments is even more limited for treating acute bipolar II depression relative to bipolar I depression. Quetiapine is the only medication with some evidence for efficacy from placebo-controlled trials with adequate sample sizes in treating acute bipolar II depression. In three out of five studies, quetiapine/(XR) was more effective than placebo in improving depression, with improvement apparent as early as week 1. Hence, quetiapine is an appropriate first-line option for bipolar II depressed patients who are drug naïve or have never tried quetiapine.

Most experts recommend lithium, lamotrigine, valproate, and lurasidone as appropriate monotherapy agents for some patients with bipolar II depression despite limited and conflicting data. Recent lumateperone monotherapy trial suggests that efficacy was robust in bipolar II depressed patients; hence, this should also be considered an appropriate option for bipolar II depressed patients.

Modern antidepressants are widely used to treat bipolar II depression, although the evidence is limited and conflicting. Given the limited effective options for treating bipolar II depression, modern antidepressant adjunctive therapy is appropriate for treating some patients with bipolar II depression. A meta-analysis suggests that the risk of antidepressant mood elevation is lower in bipolar II disorder compared with bipolar I disorder (7% in acute and 14% in maintenance trials vs. 14% and 23%, respectively for bipolar I). Therefore, there is less disagreement among experts about the use of modern antidepressant adjunctive therapy for bipolar II depression. This advice is consistent with the recommendations of a recent ISBD Task Force that endorses its use in bipolar II depression except in patients with rapid cycling bipolar disorder and in those with mixed features. Although a series of small studies suggest that antidepressant monotherapy is effective and is not associated with increased risk of a switch, such a strategy is not recommended for most bipolar II depressed patients.

ECT is an effective treatment for bipolar II depression. Although the evidence is very limited, adjunctive therapy with pramipexole, modafinil, omega-3 fatty acids, or N-acetyl cysteine could be considered for treating bipolar II depressed patients who are refractory to the above strategies.

Maintenance Treatment of Bipolar II Disorder. Syndromal or subsyndromal depressive symptoms comprise the vast majority of symptomatic periods in patients with bipolar II disorder. Therefore, the major challenge in the management of bipolar II disorder is not only to treat depressive episodes to full remission but also to prevent relapse/recurrence of depressive episodes without increasing the risk of a hypomanic switch.

Clinicians typically continue the medication that worked for the acute episode for the maintenance treatment. While such strategy has not been widely studied in bipolar II disorder, there is some evidence that bipolar II depressed patients who respond to quetiapine or lamotrigine benefit from the continuation of the same medication for the maintenance treatment. Lithium is beneficial for maintenance regardless of whether it was used to treat acute episodes or not. If a patient has responded to adjunctive modern antidepressant therapy, the continuation of the combination may be beneficial for some patients. Some studies suggest that patients who responded to fluoxetine monotherapy may benefit from a continuation of this medication for prophylaxis. Although these studies did not report increased hypomanic switch rates, clinicians are well-advised to warn patients about the potential risk of hypomanic/manic switch if such a strategy was chosen. If the above options are ineffective, open-label data and clinical experience support the use of combinations of mood stabilizers or a mood stabilizer and an atypical antipsychotic with or without a modern antidepressant or pramipexole or other agents such as NAC or omega-3 fatty acids.

Treatment of Bipolar Disorder with Comorbidity

Both psychiatric and medical comorbidity is highly prevalent in patients with bipolar disorder and impacts the clinical outcome. In general, bipolar patients with comorbidities tend to have poorer outcomes as they are less likely to respond or take a longer time to respond to treatments and are more likely to relapse. Therefore, the optimal management of such patients includes treating not only symptoms associated with bipolar disorder but also addressing comorbidities without destabilizing bipolar disorder. However, clinical trials conducted for regulatory approval of treatments for bipolar disorder have systematically excluded patients with active comorbidities. Therefore, in the absence of robust evidence-based strategies for the management of patients with bipolar disorder with psychiatric comorbidity, treatment decisions for the management of comorbidity have to be made by necessity, based on the secondary analysis of data on comorbid symptoms in trials of bipolar disorder and the treatment efficacy data from clinical trials of the other conditions not in the context of bipolar disorder. In choosing treatments for managing comorbidity, clinicians must carefully weigh the risks of adjunctive treatments, such as their propensity to destabilize bipolar disorder and adverse events versus benefits of treating comorbidity and the consequential positive impact on the outcome of bipolar disorder.

Anxiety Disorders. Since several psychological treatments such as CBT have been shown to be effective in treating primary anxiety disorders, their use is preferable in treating the comorbid anxiety disorders in bipolar patients as they pose little risk of illness destabilization.

GENERALIZED ANXIETY DISORDER. GAD/anxiety symptoms are common in bipolar disorder, impact clinical outcomes, and hence must be treated. In addition to psychological treatments, if a patient has been taking an anticonvulsant such as lamotrigine or valproate or an atypical antipsychotic such as quetiapine or olanzapine plus fluoxetine combination for bipolar disorder, these should be optimized as there is some evidence from secondary analysis of clinical trials or open-label data to support their efficacy. If optimization of ongoing pharmacotherapy is ineffective in resolving symptoms, medications that have been shown to be effective in treating primary anxiety disorder should be considered. Among these, pregabalin and quetiapine are preferable as they are not associated with the risk of destabilization of bipolar disorder. Gabapentin is also useful as anecdotal evidence supports its efficacy in treating comorbid anxiety symptoms in bipolar patients. Some SSRI antidepressants (escitalopram, paroxetine, sertraline) are effective and widely used in clinical practice. Similarly, some newer antidepressants such as agomelatine and vortioxetine are also effective in treating GAD; the risk of illness destabilization appears to be lower with agomelatine, while no data are available as yet for vortioxetine. However, it must be remembered that since anxiety disorders/symptoms pose a chronic symptom burden in bipolar disorder, the long-term use of SSRIs or other antidepressants in bipolar patients may pose a considerable risk of illness destabilization. Hence, other strategies outlined above should be tried first before considering the use of SSRIs. If SSRIs are used on a longer-term basis, patients need to be on an adequate cover with antimanic mood-stabilizing medications (i.e., a combination of two mood stabilizers or a mood stabilizer and an atypical antipsychotic). Although SNRIs such as duloxetine and venlafaxine are effective in primary anxiety disorder, their use in bipolar patients is not recommended due to their potential higher risk of mood destabilization.

SOCIAL ANXIETY DISORDER. While the estimates of a social anxiety disorder (SAD) in bipolar disorder vary, the National Comorbidity Survey reported that 39% of patients with any type of bipolar disorder have SAD. There are virtually no controlled efficacy data in this comorbid population. CBT is very effective in treating primary SAD and likely useful in comorbid patients as well. Among the agents shown to be effective in treating primary SAD, pregabalin and gabapentin are preferred options as they have no propensity to destabilize bipolar disorder. In nonresponders, SSRIs could be considered provided patients with bipolar disorder are on adequate mood stabilization. Beta-blockers such as propranolol and atenolol are not recommended as they do not appear to be beneficial in primary SAD.

PANIC DISORDER. Between 14% to 27% of patients with bipolar disorder have panic disorder, and up to 35% of patients with bipolar disorder report experiencing panic attacks. There is limited open-label data to support the use of valproate in bipolar disorder with comorbid panic disorder. There is posthoc evidence that gabapentin is useful in severe PD. If optimization of anticonvulsant therapy is ineffective, the addition of an atypical antipsychotic such as quetiapine or risperidone may be appropriate. As with other anxiety disorders, SSRIs are effective, but these should be given preferably in conjunction with an atypical antipsychotic and possibly a mood stabilizer in order to minimize the risk of manic/hypomanic switch. Clonazepam, alprazolam, and lorazepam are also effective, and they may be appropriate for short-term use to control panic attacks.

POST-TRAUMATIC STRESS DISORDER. Lifetime post-traumatic stress disorder (PTSD) rates range from 16% to 39% in patients with bipolar disorder. Childhood sexual abuse, adult sexual assault, and adult survival of the suicide, homicide, or accidental death of a close friend or relative are commonly associated with comorbid PTSD in bipolar disorder.

There are virtually no controlled data on bipolar disorder with PTSD comorbidity. Anticonvulsants are likely beneficial and should be preferred particularly in light of their efficacy in reducing cravings for alcohol, as comorbid alcohol abuse is common in patients with PTSD. Atypical antipsychotic mono or adjunctive therapy has been reported to be beneficial for primary PTSD, particularly in treating symptoms of intrusion, and hence are appropriate options in anticonvulsant nonresponders. As with other anxiety disorders, SSRIs are effective but should be used preferably in conjunction with an atypical antipsychotic. Benzodiazepines are not effective, and there is anecdotal evidence that they may worsen symptoms.

OBSESSIVE-COMPULSIVE DISORDER AND RELATED DISORDERS. About 10% to 25% of patients with bipolar disorder have a lifetime history of OCD. The symptoms of OCD are particularly exaggerated during depressive episodes. Sexual and religious obsessions tend to be more common in bipolar disorder, and their severity tends to wax and wane over a lifetime. Psychological treatments such as CBT should be offered first. There are no controlled data for pharmacotherapy in the comorbid patient population, but there is some evidence that adjunctive treatment with lamotrigine, topiramate, or atypical antipsychotics such as risperidone, aripiprazole, olanzapine, or quetiapine is effective in treating primary refractory OCD. Since these medications, with the exception of topiramate, are effective in treating bipolar disorder, they should be optimized first. In those that continue to manifest OCD symptoms, therapy with topiramate or a combination of an atypical antipsychotic with an SSRI (i.e., escitalopram, fluoxetine, fluvoxamine, paroxetine, and sertraline) may be considered. The need for ongoing SSRI treatment should be evaluated periodically, and the risk versus benefit analysis should determine the duration of SSRI treatment. While clomipramine is effective in primary OCD, it is not recommended in bipolar patients due to the higher risk of a manic switch.

Alcohol and Substance Use Disorders. About 50% of patients with bipolar disorder have alcohol or substance use comorbidity. There is evidence that valproate is effective in reducing various indices of alcohol use in bipolar patients with comorbid alcohol use, and hence is a preferred option in this population. Lamotrigine is also beneficial for some patients with alcohol comorbidity. Other anticonvulsants such as gabapentin and topiramate have shown efficacy in primary alcoholism and could be used as adjuncts to primary treatments for bipolar disorder in some patients to address alcohol comorbidity. While naltrexone and disulfiram are effective in primary alcohol use disorder, disulfiram is not recommended in bipolar patients given its inhibitory effects on dopamine β-hydroxylase and the potential for worsening of mania. There is some evidence for naltrexone add-on therapy in bipolar patients with alcohol comorbidity. While quetiapine is effective in primary bipolar disorder, it does not appear to have any utility in reducing alcohol-related behaviors.

Valproate, lamotrigine, quetiapine, and risperidone mono or add-on therapy appear to provide some benefit in patients with cocaine abuse. Add-on NAC or modafinil may be useful in some patients.

Attention-Deficit Hyperactivity Disorder. The estimates for the prevalence of ADHD in patients with bipolar disorder range from 9% to 30%, and therefore, all patients with bipolar disorder should be routinely screened for comorbid ADHD. Bipolar patients with ADHD tend to have an earlier age at onset of mood symptoms, more frequent mood episodes, more suicide attempts, and other psychiatric comorbidities. ADHD symptoms persist during euthymia and contribute to impairment in function. The diagnosis of comorbid ADHD in bipolar patients can be challenging, given the overlapping symptoms. Therefore, clinicians need to be aware of the possibility of both under-diagnosis and over-diagnosis of ADHD in bipolar disorder. Persistent impulsivity and the impaired ability for sustained attention despite treatment with mood-stabilizing medications and the absence of other symptoms of mood episodes should raise the index of suspicion for ADHD.

If ADHD symptoms and functional impairment persist despite psychoeducation and behavioral interventions in bipolar patients optimized on mood-stabilizing medications, the institution of specific pharmacotherapy targeting ADHD symptoms is appropriate. However, clinicians must be aware that ADHD treatments can destabilize the course of bipolar disorder. Adjunctive bupropion is often the first choice, particularly in bipolar patients with comorbid ADHD and frequent depressive episodes/depressive symptoms, as it is less likely to induce manic switch compared with stimulants. In nonresponders, adjunctive modafinil or mixed amphetamine salts, or methylphenidate could be considered. If a patient responds, clinicians are advised to periodically evaluate the need for ongoing treatment with stimulants and offer psychoeducation to patients on monitoring for the emergence of manic symptoms.

FUTURE DIRECTIONS IN THE TREATMENT OF MOOD DISORDERS

An ultimate goal is a stratified approach to the personalized treatment of major depressive disorder, bipolar disorders, and other mood disorders. However, this is limited by the failure of clinical phenotypes to inform the selection of psychotropic medication. While the search for validated molecular, neuroimaging, and other biomarkers to subgroups within the major depressive disorder and bipolar disorders is underway, this has not yet influenced treatment selection or patient outcomes.

Newer therapeutic approaches are also emphasizing the rapidity of response, which matches the acute and severe nature of acute mood episodes. This approach may eventually lead to more discrete care pathways that utilize specific treatment goals at various stages of management, even within acute episodes.

Newer studies are also emphasizing the importance of therapeutic targets beyond only mood symptoms. For instance, clinical trials that utilize cognitive functioning as a primary outcome offer the hope for additional interventions that help patients across the broad spectrum of illness characteristics.

Further Readings

Aboraya A, Nasrallah HA, Elswick DE, et al. Measurement-based care in psychiatry: past, present, and future. *Innov Clin Neurosci*. 2018;15(11–12):13–26.

Arnow BA, Blasey C, Williams LM, et al. Depression subtypes in predicting antidepressant response: a report from the iSPOT-D trial. *Am J Psychiatry*. 2015;172(8):743–750.

Carhart-Harris R, Giribaldi B, Watts R, et al. Trial of psilocybin versus escitalopram for depression. *N Engl J Med*. 2021;384(15):1402–1411.

Cipriani A, Barbui C, Salanti G, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *Lancet*. 2011;378(9799):1306–1315.

Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet*. 2009;373(9665):746–758.

Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet*. 2018;391(10128):1357–1366.

Clayton AH, Alkis AR, Parikh NB, Votta JG. Sexual dysfunction due to psychotropic medications. *Psychiatr Clin North Am*. 2016;39(3):427–463.

Daly EJ, Trivedi MH, Janik A, et al. Efficacy of esketamine nasal spray plus oral antidepressant treatment for relapse prevention in patients with treatment-resistant depression: a randomized clinical trial. *JAMA Psychiatry*. 2019;76(9):893–903.

Dore J, Turnipseed B, Dwyer S, et al. Ketamine assisted psychotherapy (KAP): patient demographics, clinical data and outcomes in three large practices administering ketamine with psychotherapy. *J Psychoactive Drugs*. 2019;51(2):189–198.

Fava M, Durgam S, Earley W, et al. Efficacy of adjunctive low-dose cariprazine in major depressive disorder: a randomized, double-blind, placebo-controlled trial. *Int Clin Psychopharmacol*. 2018;33(6):312–321.

Frankel JS, Schwartz TL. Brexpiprazole and cariprazine: distinguishing two new atypical antipsychotics from the original dopamine stabilizer aripiprazole. *Ther Adv Psychopharmacol*. 2017;7(1):29–41.

Furukawa TA, Efthimiou O, Weitz ES, et al. Cognitive-behavioral analysis system of psychotherapy, drug, or their combination for persistent depressive disorder: personalizing the treatment choice using individual participant data network metaregression. *Psychother Psychosom*. 2018;87(3):140–153.

Ghaemi SN, Ostacher MM, El-Mallakh RS, et al. Antidepressant discontinuation in bipolar depression: a systematic treatment enhancement program for bipolar disorder (STEP-BD) randomized clinical trial of long-term effectiveness and safety. *J Clin Psychiatry*. 2010;71(4):372–378.

Gignac A, McGirr A, Lam RW, Yatham LN. Recovery and recurrence following a first-episode of mania: a systematic review and meta-analysis of prospectively characterized cohorts. *J Clin Psychiatry*. 2015;76(9):1241–1248.

Ginsberg LD. Impact of drug tolerability on the selection of antidepressant treatment in patients with major depressive disorder. *CNS Spectr*. 2009;14(12):8–14.

Kennedy SH, Lam RW, McIntyre RS, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: section 3. Pharmacological treatments. *Can J Psychiatry*. 2016;61(9):540–560.

Krystal AD, Pizzagalli DA, Mathew SJ, et al. The first implementation of the NIMH FAST-FAIL approach to psychiatric drug development. *Nat Rev Drug Discov*. 2019;18(1):82–84.

Lam RW, Parikh SV, Michalak EE, Dewa CS, Kennedy SH. Canadian Network for Mood and Anxiety Treatments (CANMAT) consensus recommendations for functional outcomes in major depressive disorder. *Ann Clin Psychiatry*. 2015;27(2):142–149.

Ng F, Mammen OK, Wilting I, et al. The International Society for Bipolar Disorders (ISBD) consensus guidelines for the safety monitoring of bipolar disorder treatments. *Bipolar*

Disord. 2009;11(6):559–595.

Nutt D, Carhart-Harris R. The currentstatus of psychedelics in psychiatry. *JAMA Psychiatry*. 2021;78(2):121–122.

Pacchiarotti I, Bond DJ, Baldessarini RJ, et al. The International Society for Bipolar Disorders (ISBD) Task Force report on antidepressant use in bipolar disorders. *Am J Psychiatry*. 2013;170(11):1249–1262.

Parikh SV, Huniewicz P. E-health: an overview of the uses of the Internet, social media, apps, and websites for mood disorders. *Curr Opin Psychiatry*. 2015;28(1):13–17.

Raison CL, Rutherford RE, Woolwine BJ, et al. A randomized controlled trial of the tumor necrosis factor antagonist infliximab for treatment-resistant depression: the role of baseline inflammatory biomarkers. *JAMA Psychiatry*. 2013;70(1):31–41.

Schaffer A, Isometsä ET, Azorin J, et al. A review of factors associated with greater likelihood of suicide attempts and suicide deaths in bipolar disorder: part II of a Report of the International Society for Bipolar Disorders Task Force on Suicide in Bipolar Disorder. *Aust N Z J Psychiatry*. 2015;49(11):1006–1020.

Sim K, Lau WK, Sim J, Sum MY, Baldessarini RJ. Prevention of relapse and recurrence in adults with major depressive disorder: systematic review and meta-analyses of controlled trials. *Int J Neuropsychopharmacol*. 2015;19(2):pyv076.

Strawbridge R, Carter B, Marwood L, et al. Augmentation therapies for treatment-resistant depression: systematic review and meta-analysis. *Br J Psychiatry*. 2019;214(1):42–51.

Swartz MA, Thase ME. Pharmacotherapy for the treatment of acute bipolar II depression: current evidence. *J Clin Psychiatry*. 2011;72(3):356–366.

Thase ME, Gelenberg A, Kornstein SG, et al. Comparing venlafaxine extended release and fluoxetine for preventing the recurrence of major depression: results from the PREVENT study. *J Psychiatr Res*. 2011;45(3):412–420.

Yatham LN, Beaulieu S, Schaffer A, et al. Optimal duration of risperidone or olanzapine adjunctive therapy to mood stabilizer following remission of a manic episode: a CANMAT Randomized double blind placebo controlled trial. *Mol Psychiatry*. 2016;21(8):1050–1056.

Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord*. 2018;20(2):97–170.

13.7 Mood Disorders: Psychotherapy

John R. Mcquaid, Ph.D., and Michael E. Thase, M.D.

Psychotherapy, in various forms, has been one of the cornerstones of therapeutics for mood disorders for decades. Even before controlled studies were available to document the utility of these approaches, there was substantial clinical support or face validity for treating depression psychotherapeutically; depression frequently occurs during times of significant psychosocial stress and often is associated with relevant intra- and interpersonal vulnerabilities. As the field has become more evidence based and awareness of issues such as cost-effectiveness and disseminability has grown, greater emphasis has been placed on models of therapy that are procedurally specified and empirically tested. These trends were first evident in the introduction of shorter-term or time-limited therapies, particularly those that emphasize more pragmatic behavioral, cognitive behavioral, or interpersonal formulations and interventions. More recently, time-limited and manual-based models of psychodynamic psychotherapy have been introduced for the treatment of mood disorders. All of these therapies aim to improve upon the considerable beneficial effects of supportive clinical management, with the potential for theoretically driven, mode-specific benefits in addition to the relief of depressive symptoms. This chapter will review the current status of these therapies.

HISTORY OF PSYCHOTHERAPY FOR DEPRESSION

Psychotherapy for depression arose from psychoanalytic and subsequent psychodynamic models of intervention. Early theories conceptualized depression as an intrapsychic phenomenon stemming from internalized anger that arose from unconscious conflicts. Therefore, treatment was long term and focused on developing awareness of these internal conflicts through identification and interpretation of historical experiences that reflected such conflicts. Intervention emphasized developing insight into these conflicts through the use of free association and the therapist's interpretations. This treatment also used the patient's projections onto the therapist (the transference neurosis) as a means of studying early object relationships and conflicts.

In contrast to more traditional approaches, behavioral theories of depression treatment emphasized observable factors, such as decreased participation in potentially rewarding or pleasurable activities, and suggested that the negative emotional symptoms of depression could be understood as a consequence of withdrawal of reinforcement. Treatment interventions initially were guided by learning theories, but by the 1970s had broadened to include cognitive components (e.g., to increase self-efficacy or decrease negative thoughts or beliefs about self, world, and future). Some models emphasized social skills or problem-solving strategies, whereas others focused more on increasing positive activities or remediating cognitive symptoms and vulnerabilities. Whatever the differences in specific models of CBT, all emphasize the role of the therapist as a teacher of specific skills that are intended to help the depressed person reduce depressive symptoms and/or improve coping. Moreover, virtually all models of CBT emphasize the importance of homework assignments to enhance mastery and generalization of these skills in "real-world" environments.

Arguably, the most impactful of the models of CBT to emerge from the 1970s was the approach developed by Beck and colleagues. More commonly known as cognitive therapy (CT), this form of therapy focuses more on identifying the negatively valenced automatic thoughts associated with depressed moods and using strategies to both test the accuracy of the negative thoughts and consider more rational (and less depressogenic) alternatives. As many people in the midst of a depressive state evince various kinds of cognitive distortions (e.g., thinking in a more overly general and absolute way or selective recall of negative mood–relevant memories or associations), a skilled CT therapist can help most nonpsychotically depressed people see examples of "cause and effect" relationships between cognition, mood, and behavior within the first hour of treatment.

Whereas early models of behavior therapy either minimized the value of the therapeutic relationship or were silent about its importance, Beck and colleagues emphasized a collaborative approach between the therapist and the patient, referred to as collaborative empiricism, and specifically structured therapy sessions in discrete 10- to 20-minute segments to ensure that the patient had the opportunity to give feedback (e.g., whether the particular topic was personally relevant or made sense) and express reservations. A collaborative process using Socratic questioning was similarly incorporated to guide and facilitate mastery of techniques for testing and remediating negative thoughts and beliefs, and the patient practices these techniques.

Beyond CT's initial focus on automatic negative thoughts ("surface" cognitions), Beck's model of therapy also emphasized that vulnerability to depression was linked to "deeper" forms of cognition, including dysfunctional attitudes and schema. Schemas are underlying cognitive structures that help the brain organize information in memory, guide awareness, and predispose action. The complex thoughts and actions involved in tying a bow are an example of a simple, nonaffectively relevant schema. Schemas that are associated with depression are often reflected by automatic negative thoughts or abiding beliefs about competence ("I'm just not good enough") or intimacy ("No one will ever really love me"), which may be activated by thematically relevant life events. Thus, another distinguishing feature of CT is a specific stress–diathesis model of vulnerability. As such, schemas are "silent" (i.e., they must be inferred from the predominant themes or beliefs revealed in automatic negative thoughts), and Beck's model of CT did deviate significantly from the zeitgeist of the behavior therapy movement. Nevertheless, the fully developed model of therapy developed by Beck and colleagues incorporated more conventional behavioral strategies such as activity scheduling and graded task assignments.

Interpersonal psychotherapy (IPT), developed by Klerman, Weissman, and colleagues, emerged in the 1980s as the leading alternative to CBT. Drawing upon psychodynamically influenced interpersonal models (Meyer and Sullivan) and attachment theory (Bowlby), IPT incorporated many of the more pragmatic methods used by social workers in case management to develop a time-limited therapy that addresses the common problematic patterns in relationships that plague the lives of people with depression, including unresolved grief, role disputes, role transitions, and interpersonal deficits. Beyond eliciting an interpersonal inventory and identifying the area or areas of interpersonal difficulty of greatest relevance to a particular patient, the interventions used in IPT are relatively eclectic and can include psychoeducation, nondirective questioning and empathic support, and role-playing and social problem solving. IPT thus may be particularly well suited for more traditionally trained psychotherapists or counselors who find it difficult to implement the structure, activity level, and the more directive posture that characterize most models of CBT. Significantly, IPT was the first psychotherapy outside of CBT to be evaluated in comparison to antidepressant medications using RCT methods.

Several adaptations of IPT have a particular utility for patients with more complex depressive disorders. Frank and colleagues incorporated the importance of social rhythm disruption and the adaptive value of social rhythm stability (i.e., the consistency of regular activities such as bedtime and meal times) in their adaptation of IPT, which was primarily targeted for recurrent forms of affective illness, including bipolar disorder. Known as interpersonal and social rhythm therapy (IPSRT), this approach adds an assessment of social rhythm stability and activity scheduling tailored to strengthen the role of social *zeitgebers* (time givers) in helping patients prevent recurrent affective episodes and promote recovery. Shear and colleagues adapted IPT to focus on complicated bereavement, including clinical presentations in which the thoughts, feelings, and behaviors pertaining to bereavement overlap with the signs/symptoms of PTSD.

More recent innovations in psychotherapy have incorporated techniques from a long-standing meditation tradition and done so using an evidence-based approach. Mindfulness meditation, developed from Buddhist techniques, emphasizes two skills: observing one’s perceived sensations and accepting and experiencing those sensations nonjudgmentally. In the application of this technique to depression, sensations include cognition and emotional states, as well as input through the five senses. Segal, Teasdale, and colleagues propose that these techniques are relevant for three reasons. First, the training in observation facilitates gathering evidence, which is a central component of CBT. Second, learning to view thoughts—even the most negative or damning ones—as discrete, limited experiences helps patients dampen affective arousal in response to them, which in turn improves their ability to consider alternate perspectives. Third, the emphasis in meditation on nonjudgmental observing moves the therapy toward a focus on acceptance rather than change per se. For more chronic depressive disorders, a focus on acceptance—rather than change—can offer a novel therapeutic component, particularly when the ability to tolerate distress has interfered with coping behaviors.

Several models of treatment now incorporate mindfulness techniques and are often referred to as “third-wave” techniques, reflecting the path from behavioral to cognitive to now mindfulness and acceptance-based interventions within the CBT tradition. Mindfulness-based cognitive therapy (MBCT) is a treatment intervention specifically developed to reduce the risk of relapse in individuals who have a history of recurrent depression. Acceptance and commitment therapy (ACT), developed by Steven Hayes and colleagues, emphasizes identifying core values and committing to acting on those values rather than focusing on avoiding suffering. This model was particularly designed for addressing chronic depression. Another closely related therapy, dialectical behavior therapy (DBT), has recently been adapted for working with patients with more treatment-resistant depression. Initially developed by Linehan and colleagues, DBT is a behaviorally oriented system of therapy that also places a strong emphasis on mindfulness and acceptance and has well-replicated effects on the reduction of suicidal ideations and behaviors.

Family support is an important asset in life and is generally a favorable prognostic indicator across mood disorders. Conversely, there is an increased risk of relapse/recurrence for patients whose families show high levels of criticism and emotional expression. Drawing upon these observations, Miklowitz and colleagues developed family-focused therapy (FFT), which has primarily been adapted for the treatment of adolescents and young adults with bipolar disorder. This treatment focuses on educating both patients and family members on the disorder and helping them gain a shared perspective of the disorder. From this shared perspective, the therapist guides the family in developing a relapse prevention plan, including ways that the family can identify symptom exacerbations, an agreement about how they will share this observation with the patient, and an understanding of how the patient will respond to this information. Options for managing symptoms are generated (e.g., visiting the doctor for medication adjustment or increasing structure in the home). Family members are also trained to communicate effectively in ways with the patient and reduce the level of hostility and criticism in the environment.

REVIEW OF RESEARCH FINDINGS

Unipolar Disorders

Individual Psychotherapy. A number of forms of psychotherapy (see Table 13.7–1) have been shown to be either efficacious or probably efficacious for outpatients with major depressive disorder. These treatments have been shown to be effective in outpatients treated in a variety of settings (e.g., psychiatric clinics, psychology clinics, college counseling centers, community mental health centers, and primary care clinics), and meta-analyses suggest that across modes of therapy, differences tend to be modest to nonexistent. By contrast, no form of psychotherapy is indicated as the primary treatment for bipolar disorder, although a growing literature supports the value of several different forms of psychotherapy as adjuncts to ongoing therapy with mood stabilizers.

CBT has been evaluated in numerous RCTs, including placebo-controlled studies that included first- and second-generation antidepressants as active comparators. Although the findings of one very influential early multicenter study, the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program (TDCRP), suggested that CT may be less effective than pharmacotherapy for the treatment of patients with more severe depressive symptoms, the findings of a number of other RCTs do not support such a relationship. Rather, patients with higher symptom severity scores tend to respond slower or less favorably to many different interventions. In general, depressed patients treated with CBT are as likely to respond/remit across 8 to 16 weeks of treatment as are patients treated with antidepressant medications.

Table 13.7–1.
Efficacious Treatments for Major Depression

Treatment	Conceptualization of Disorder Etiology	Sample Interventions
Behavioral therapy	Deficit of reinforcers, including pleasant activities and positive interpersonal contacts	Increase activity level Structured goal setting Interpersonal skills training
Cognitive behavioral therapy	Interaction of beliefs with matching stressor	Identify and challenge automatic thoughts Engage in activities that provide evidence disproving dysfunctional beliefs Modify core beliefs by reviewing evidence
Interpersonal psychotherapy	Interpersonal vulnerabilities arising from early attachment and learned relationship patterns	Develop awareness of patterns in primary relationships and the therapeutic relationship Interpersonal skills training Communication analysis
Behavioral marital therapy	Marital distress increases stress while impairing support resources	Assertive communication training Active listening exercises Problem-solving skills Increasing reinforcing behaviors toward spouse

A number of studies have likewise demonstrated that IPT also is an effective treatment for depressed outpatients, and there is particularly strong literature supporting the use of IPT for the treatment of depression during and after pregnancy. In aggregate, IPT and CBT appear to be comparably effective across 2 to 4 months of treatment.

The results of several studies suggest that simpler behavioral interventions may be as effective for depressed outpatients as more elaborate models of therapy. The performance of a “pure” behavioral activation (BA) intervention was particularly strong for patients with more severe levels of depression in one study. Whether BA and CT or IPT might be sequenced for patients at increased risk of relapse or recurrence is an interesting topic for future research, particularly because there is growing evidence that CBT and IPT curricula can be provided online, with a relatively modest amount of therapist support.

Although it was once true that there were too few studies of time-limited forms of psychodynamic psychotherapy, there is now a sufficient body of research to conclude that these are efficacious for the treatment of depressed outpatients (as compared to waiting list control conditions) and have outcomes across 8 to 16 weeks that are comparable to other therapies, such as CBT or IPT, or antidepressant medication. To date, no reliable indicators of preferential responsiveness to psychodynamic psychotherapy versus other forms of psychotherapy have emerged.

Across studies and types of psychotherapies, there is evidence that a response to psychotherapy may be more durable or persistent for weeks or months after therapy is stopped than is the case following the withdrawal of antidepressant medications. As a result, differences in the cost of delivering care that might favor pharmacotherapy in the short run may dissipate or actually shift in favor of psychotherapy when the need for continuation and maintenance phase pharmacotherapy is taken into account. Longer-term models of psychotherapy have been developed for patients who do not obtain full remission during a typically adequate course of individual therapy. In one study of incompletely remitted patients judged to be at high risk for relapse, the provision of 8 months of continuation phase CBT had a relapse prevention effect that was comparable to pharmacotherapy. It is less certain that an effective course of time-limited psychotherapy actually reduces the longer-term risks of recurrent depression. In the classic study of maintenance phase therapies conducted by Frank and colleagues, patients with highly recurrent forms of major depressive disorder who recovered with the combination of IPT and imipramine were significantly more like to remain well when their maintenance treatment included the active imipramine than when they received monthly sessions of “maintenance” IPT. Nevertheless, other studies have indicated that the addition of CBT or MBCT during an ongoing course of continuation or maintenance phase therapy may improve residual symptoms and reduce the risk of recurrent illness after discontinuation of antidepressant medication.

Group and Couples Therapy. A number of studies indicate that cognitive and behavioral therapies are well suited for delivery in groups. Group delivery of CBT may not only be a cost-effective alternative to individual therapy but may have particular value for groups with subsyndromal symptoms, for whom individual treatment might be judged not to be necessary. Depressed individuals with marital discord similarly have been shown to benefit from behavioral marital therapy (BMT). In controlled studies, BMT appears to have an impact on depressive symptoms that is comparable to individual CBT, with the added value of greater improvements in marital functioning than CBT.

Combined Psychotherapy and Pharmacotherapy. Many psychiatrists favor combining psychotherapy and pharmacotherapy, particularly for the treatment of outpatients with more chronic, severe, or treatment-resistant forms of depression. Although the effect sizes typically favoring combined treatment over competently administered monotherapies have tended to be modest, a pooled analysis of participants from a series of studies conducted at the University of Pittsburgh found that the advantage of combined treatment was particularly large for patients with more severe, recurrent depressive disorders. Results of several larger-scale, multicenter studies likewise showed advantages for patients with more severe or chronic depression. Similarly, several studies that have adapted IPT and CBT to inpatient settings have shown that the addition of focused therapy improves outcomes in comparison to treatment as usual conditions that include antidepressant medication and milieu therapies. As noted earlier, sequential studies of pharmacotherapy followed by focused psychotherapy have likewise shown advantages for promoting recovery or facilitating discontinuation of pharmacotherapy.

Bipolar Disorders

In addition to the models of psychotherapy described earlier, group psychoeducation has emerged as a cost-effective form of adjunctive intervention for people with bipolar disorder. As developed by Colom and colleagues in Barcelona, this practical treatment can be conducted in groups as large as 30 to 50 participants and their families and focuses on risk factors, warning signs, and coping strategies. Although the acute efficacy of psychoeducation has never been studied for bipolar depression, it has been shown in several studies to have a significantly positive effect on treatment adherence and relapse prevention.

With respect to the treatment of bipolar disorder, no form of psychotherapy—provided as a monotherapy—has been shown to be an effective treatment of mania or mixed states. Like group psychoeducation, most formal therapies have been studied as adjuncts to pharmacotherapy with mood stabilizers with the goal of preventing relapses or recurrences. Looking across interventions, the evidence is fairly convincing that adjunctive therapies, including psychoeducation, IPSRT, FFT, and CBT, significantly increase well time and reduce the risk of relapses and recurrence. In fact, the magnitude of benefit for adjunctive psychotherapy, as compared to no adjunctive treatment or treatment as usual alone, is comparable to the prophylactic effects of mood stabilizers such as divalproex or lamotrigine in placebo-controlled trials (see Table 13.7–2).

The evidence for the utility of focused psychotherapies for the treatment of acute bipolar depressive episodes carries the same caveat that pertains to antidepressant pharmacotherapy. In one component study of the large-scale U.S. study known by the acronym STEP-BD, adjunctive psychotherapy with FFT, IPSRT, or CBT had significant effects on both depressive symptoms and social functioning. The nature of this study did not permit valid comparisons among the psychotherapies. Given the hazards of treatment-emergent affective switches and induction of rapid cycling associated with antidepressants, the potential for focused psychotherapies for adjunctive treatment of bipolar I depressive episodes warrants further exploration. Likewise, the use of focused psychotherapy alone for selected patients with bipolar II depressive episodes deserves systematic study.

APPLYING THE RESEARCH TO CLINICAL PRACTICE

There have been important advances in understanding the roles of psychotherapy in the treatment of mood disorders, and several good psychotherapy options for both unipolar and bipolar disorders now exist. Nonetheless, clinicians are often faced with providing treatment to patients who may not match the typical participant in research trials and need to extrapolate from the available data to develop a treatment plan for their patients. Of course, several important critical questions remain, including which treatment is best for which patient, what are optimal doses (e.g., number and length of sessions), and how best to address comorbid disorders. It is also not clear how critical-specific techniques associated with these interventions are for the successful treatment of mood disorders. A study of the therapeutic process of IPT (i.e., the actual in-session behaviors of the therapist and patient) found that the process was more consistent with ideal CBT than ideal IPT. A meta-analysis examining comparisons between specific psychotherapies and placebo conditions found that the difference between treatments was related to the adequacy of the placebo design. Studies with placebo comparisons deemed adequate (e.g., using nonspecific therapy techniques for the placebo condition and matching for the level of contact with therapists) showed smaller differences between active interventions and placebo than those studies with placebo conditions that were inadequate (e.g., less therapist contact or constraint of topics to issues not related to the patient's disorder). These findings suggest that differences between psychotherapy techniques may be less critical than some of the shared components (e.g., positive therapeutic relationship, shared goals, and facilitated behavior change). There are also no definitive answers regarding the underlying models of depression. Although BA, CBT, and IPT are based on empirical models of psychopathology that all have supportive evidence, the development of these treatments and the conceptualization of depression are extrapolations beyond the available data.

Table 13.7–2.
Efficacious Treatments for Bipolar Disorder

Treatment	Conceptualization of Disorder Etiology	Sample Interventions
Cognitive behavioral therapy	Biologic vulnerability interacting with stress Skills deficits limit ability to manage symptoms	Identify and challenge automatic thoughts that interfere with treatment adherence Engage in rewarding activities that provide increased routine and stability Practice communication skills with providers
Interpersonal and social rhythm therapy	Interpersonal vulnerabilities arising from early attachment and learned relationship patterns, plus disruption of social rhythms	Develop awareness of patterns in primary relationships and the therapeutic relationship Track and stabilize social rhythms Interpersonal skills training Communication analysis
Family-focused therapy	Biologic vulnerability exacerbated by negative expressed emotion in family environment.	Education regarding the disorder, including precipitants, risk factors, and effective treatment Establish relapse prevention plan, agreed upon by all involved family members Communication skills training

Clinicians then are faced with how to use this knowledge in treating patients. One key approach, regardless of the model, is the development of a case formulation to guide treatment.

TREATMENT OF MAJOR DEPRESSIVE DISORDER WITH CBT

CBT is based on the assumption that dysfunctional beliefs, learned in response to life experiences, lead to dysfunctional behavioral and emotional responses. Treatment, therefore, modifies these beliefs to reduce problematic responses and increase functional responses. Patients learn to modify beliefs by first learning to observe their thinking, affect, and behavior and then evaluating how thoughts influence mood and actions. As the patient becomes aware of these processes, the therapist teaches skills in questioning and testing the validity of beliefs and developing alternative, more accurate, functional beliefs. The long-term goal of treatment is for patients to learn the skills to such an extent that they are able to be their own therapist.

Assessment and measurement by the therapist and the patient are central to effective CBT. The therapist teaches the patient to quantify the issues of concern. Patients learn to observe and define their own assumptions that lead to depression and then engage in a series of experiments to improve their mood. Patients learn to measure the effects of these experiments so that they test the previous assumptions and can develop new, more constructive beliefs and actions that reduce depression and other problematic moods and behaviors.

A 23-year-old, never married Black woman presented to the university mental health clinic. She was in the midst of her first semester at law school, and her primary care physician referred her to the clinic after the patient reported frequent crying and feeling hopeless. She also reported that her mood was depressed nearly every day, and she was experiencing anhedonia, early, middle, and late insomnia, decreased appetite and weight loss, difficulty concentrating, and almost daily thoughts of death and life not being worth living.

CBT, like most other psychotherapies, uses a case conceptualization approach to determine the treatment of each patient. In CBT, the conceptualization is developed based on an understanding of how different aspects of cognition are related. Aaron Beck and later, Judith Beck proposed that cognition could be divided into three levels: automatic thoughts, intermediate beliefs, and cognitive schemata. The most accessible, labeled, automatic thoughts represented the conscious response to stimuli. For example, the previously mentioned patient, when studying, often thought, “I’ll never understand this.”

The second level, intermediate beliefs, is assumptions about the self, the world, and the future that lead to the automatic thought occurring in response to a particular stimulus. Using the case of the law student presented here, the patient believed that she was not competent and that she was not good enough; she perceived the world (in this case, professors and classmates) as attacking and critical. She anticipated that the future would be filled with failure. At times when things went well, she was not unduly distressed. However, in the face of failure or criticism, she quickly experienced self-critical thoughts that were consistent with these beliefs.

The content (e.g., the beliefs) and the organization of that content define the third level, cognitive schema. An individual’s schema determines which stimuli are most likely noticed and encoded in memory, which stimuli are ignored or discounted, how encoded information is linked or associated in memory, and which memories are most easily recalled. For the previously mentioned patient, she may suffer from a schema that could be conceptualized as, “I’m not good enough.” She tends to notice information consistent with this belief and tends to ignore or discount other information.

As the therapist completes an initial assessment and establishes a treatment plan, a central goal is to elaborate a case conceptualization that will facilitate treatment. By understanding the unique meaning of experiences to the patient, the therapist can be more effective in identifying interventions that truly test the beliefs and produce meaningful improvement.

Several authors have described models for developing case formulations in CBT, and they generally are based on gathering evidence to propose an underlying schema. Once the automatic thoughts, beliefs, and schema are identified, interventions are chosen based on these hypothesized structures. The therapist continues to gather evidence throughout treatment, refining the formulation over time.

Initial Assessment

Specificity and accurate data gathering are key focus of CBT. In particular, assessment of diagnosis and symptoms is necessary. Primary assessment preferably includes standardized measures, such as the Structured Clinical Interview for DSM-5 (SCID), the Hamilton Rating Scale for Depression, and the Beck Depression Index (BDI). The assessment also needs to include the dimensions of functioning: interpersonal relationships, work and achievement, health, and recreation. As treatment progresses, a formulation is developed regarding the patient’s case conceptualization, as well as particular cognitive and behavioral coping strategies that the patient uses.

The previously presented patient experienced impairments in several areas of functioning. The patient had not contacted family or friends for several weeks, which was unusual for her. She had no current close friends at her school and had not had a dating relationship for several years. She reported that she was skipping classes and that she believed she was failing her program. The patient made statements such as, “I’m just a total loser” and “I don’t belong.” Despite excellent grades in college, she believed she was only accepted to law school due to her minority status. She primarily coped with stress by ruminating and had stopped many rewarding activities (e.g., exercise).

She had several relevant experiences from her early childhood. The patient had been adopted shortly after her birth by a White couple and was their only child. She grew up in an almost exclusively White neighborhood. She reported that, although she had some friends, she had no African American friends and that she often felt discriminated against by Whites and the few African Americans with whom she had contact, who criticized her for “trying to be White.” The patient described that, despite having devoted parents, she had a sense of not being good enough and being out of place since approximately 6 years of age, when she first started school.

Course of Therapy

In the initial session, the therapist has several goals, including (1) orienting the patient to the treatment approach, (2) establishing rapport, (3) identifying problems and treatment targets, (4) assessing current symptom and problem severity, (5) providing some initial symptom relief, and (6) introducing an initial homework assignment. In reality, these components tend to overlap. For example, it is often helpful to illustrate the CBT model using the life experience of the patient, and in doing so, to provide a chance at some symptom relief. An example could be processing a situation, such as being late to a first therapy appointment.

Therapist: You mentioned that you were upset because you were late. Can you tell me what you were thinking when you realized you were late?

Patient: I figured you'd be angry.

Therapist: Okay, and when you thought I'd be angry, how did you feel?

Patient: Pretty nervous.

Therapist: So that is what we will focus on in here, how your thinking in different situations influences how you feel. In this case, you were late, thought "he'll be angry," and felt nervous. Our goal is to test out the beliefs, like "he'll be angry," and change them when they aren't healthy or accurate. The great thing is that you sort of implicitly tested out the belief that I'll be angry by showing up. Was I angry?

Patient: You didn't seem angry.

Therapist: How anxious did you feel once we started talking?

Patient: Well, I'm feeling more comfortable.

Therapist: What thoughts do you have now about me being angry?

Patient: I don't think you're angry.

Therapist: So, this is an example of what we're going to be doing in therapy—I'll be helping you to identify thoughts that make you feel bad. We're then going to work together to come up with ways to check them out and change them if they're not true or accurate. You checked out whether I was angry by observing me, and you changed your thought, and now, I get the impression that you feel better.

In this example, the model of CBT is described for the patient and illustrated using her own thoughts, behaviors, and feelings. In particular, the effect of behavioral testing of beliefs (i.e., coming into session late and finding out if the therapist actually was angry) is pointed out. To increase her sense of efficacy, the therapist points out that her current actions can lead to mood change (feeling more relaxed after coming in late) and that she is already demonstrating some of the necessary skills to benefit from CBT.

To facilitate skills, an agenda is explicitly set forth and agreed on by the therapist and the patient every session. In the early sessions, the patient is oriented and trained in thought tracking and activity tracking. The decision of which thoughts or behaviors are targets for intervention is based on the specific needs of the patient, which the therapist assesses using the case formulation. In the case described, the patient reported feeling “miserable” when she arrived at class. The therapist used the situation to train the patient in using a thought record, as shown in Figure 13.7–1. The specific situation is identified, and the emotional response and the thought or thoughts are recorded as well as the patient’s belief in the thought.

Patients learn techniques for modifying unhelpful thinking, including identifying factual inaccuracies (e.g., “I don’t understand anything” is all-or-nothing thinking) and biased information processing (ignoring evidence that she does know some things) and generating more accurate thoughts by the use of logic and testing beliefs.

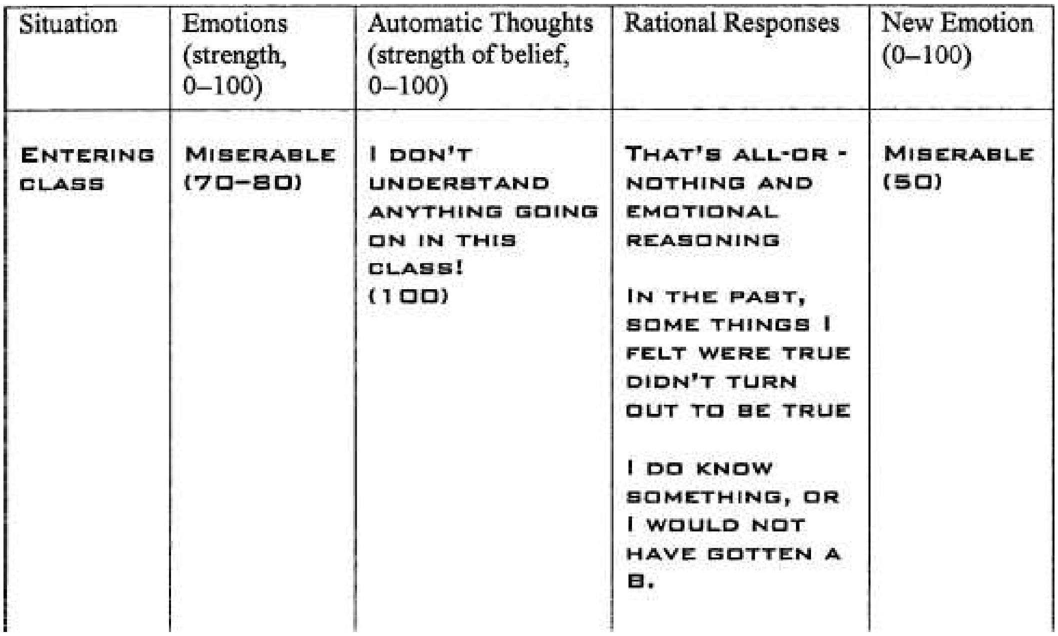

Situation	Emotions (strength, 0–100)	Automatic Thoughts (strength of belief, 0–100)	Rational Responses	New Emotion (0–100)
ENTERING CLASS	MISERABLE (70–80)	I DON'T UNDERSTAND ANYTHING GOING ON IN THIS CLASS! (100)	THAT'S ALL-OR - NOTHING AND EMOTIONAL REASONING IN THE PAST, SOME THINGS I FELT WERE TRUE DIDN'T TURN OUT TO BE TRUE I DO KNOW SOMETHING, OR I WOULD NOT HAVE GOTTEN A B.	MISERABLE (50)

FIGURE 13.7–1. Thought record example.

Therapist: Okay, in this situation, you are feeling miserable. Now let's check this thought. What problems do you see with the thought, "I don't understand anything?"
Patient: Well, it's pretty all-or-nothing, but it feels like it's true.
Therapist: Can you think of any times when you felt something was true, and it wasn't?
Patient: Sure, I felt sure I wouldn't get into law school or college, for that matter.
Therapist: What does that tell you about this thought?
Patient: Well, it might not be true either.
Therapist: Okay, what would be other evidence that this thought isn't completely true?
Patient: Well, I do know some things.
Therapist: What is your evidence? Have you taken any tests or completed any assignments?
Patient: Yeah—and I got a B on the first test.
Therapist: So, what can we write in the rational responses column?
Patient: It isn't all-or-nothing; I do know some stuff.
Therapist: How miserable do you feel if you tell yourself that?
Patient: Well, a little less.

In this case, the patient experiences several types of distortion, including all-or-nothing thinking and emotional reasoning (i.e., "I feel it is true; therefore, it must be true"). The patient is asked to generate evidence that challenges the negative belief. The evidence reduces but does not eliminate the negative feeling. This is often the case, particularly when working with retrospective examples and patients experiencing significant depression. One approach to increase the efficacy of thought challenging and evidence gathering in CBT is by developing behavioral experiments to test the belief. In this case, the patient later conducted an experiment to test the "I don't know anything" belief. She wrote down her own answers to questions asked in the class, as well as the answers that other students made. She found that her knowledge was on par with her classmates, and she was able to answer most of the questions. This experiment also addressed her "belonging" because she got evidence of her equivalence to her peers.

As treatment progresses, the therapist points out recurring themes and helps the patient to identify and challenge core beliefs. In this example, the patient repeatedly experienced thoughts about not being good enough. She said that she felt this way all her life and attributed this to her sense of being different from her parents and her European American friends and her experience of persistent racism.

One approach to changing schema is to generate alternative core beliefs and to help the patient to identify behaviors and cognitions that are consistent with the new beliefs. In this case, a belief of “I belong just like everyone belongs; others may not believe it or act that way, but I do” was consistent with the patient’s values and overriding experience. The therapist would therefore challenge thoughts by asking, “is that the ‘old self’ or ‘new self’ talking?” This provided a shorthand for identifying dysfunctional thoughts and generating new ones consistent with the alternative schema.

In the final sessions, the tasks shift to reviewing progress, identifying potential future challenges, and learning to administer interventions independently. Returning to the example, the patient showed significant improvement in her mood. Interventions that had been beneficial were attending the university gym on a regular schedule (activity scheduling), joining a study group (activity scheduling, assertiveness training, and social interaction), and gathering evidence regarding beliefs about being good enough and belonging (thought challenging and experiments). In addition, another student had invited the patient on several dates. At the end of therapy, the patient had a depression score on the BDI in the nondepressed range, her school attendance and grades had improved, and she had daily contact with friends and family.

In preparation for termination, it is important to identify the improvements that the patient has made, the skills that have been learned, and potential challenges that may arise in the future. This client identified her risk factors for increased depression as a rejection if she started dating or failure at school. She also noted that, although she now did not fear attending class and she was doing well, she still was not talking in class unless she was called on.

Therapist: Is not answering questions in class a problem?
Patient: Yeah.
Therapist: Okay, you've been doing this a while; what do you think would be a good homework for you?
Patient: Well, I guess I could just plan to answer one or two questions in class and see what happens.
Therapist: When you say "see what happens," what could you track?
Patient: Maybe how many I get right and my thoughts and feelings after each question.
Therapist: Sounds good. What are some things you can do to make that easy in the class setting?

Because CBT is a skills-based model, it is critical that the patient demonstrates an ability to use skills independently of the therapist. In this case, she is able to take on the role of therapist and assign herself a task that can help her to modify her beliefs and behavior.

ADJUNCTIVE PSYCHOTHERAPY FOR BIPOLAR DISORDER

Psychotherapy for bipolar disorder is an adjunct to pharmacotherapy for the management of symptoms. As such, interventions including CBT, FFT, and IPSRT emphasize education about the disorder, learning to track symptoms and warning signs of relapse, and improving adherence to pharmacotherapy. In CBT, additional interventions include identifying thoughts and beliefs that can either (a) exacerbate symptoms or (b) increase the risk of relapse. In CBT and IPSRT, there is an emphasis on establishing stable routines to reduce triggers for mania. Finally, FFT emphasizes the role of educating the entire family system about the disorder and enlisting family members to aid patients with the management of their disorder. The current vignette describes a case from a cognitive behavioral conceptualization that incorporates techniques from other approaches.

A 47-year-old divorced Caucasian man was referred for psychotherapy by his treating psychiatrist. The patient was currently unemployed, although he volunteered 5 hours per week at his church. Although he had a long history of symptoms, this was the first time that he had been referred for psychotherapy. At the time of referral, the patient reported mixed symptoms of depression and mania. Depressive symptoms included depressed and sad mood, feelings of guilt, and suicidal ideation. Currently, the patient reported manic symptoms of irritability, racing thoughts, decreased need for sleep, and impulsive spending. In addition, the patient was preoccupied with religious ideation.

In considering integrative interventions, it is key to have an organizing structure with which to work. In this current case, the psychotherapist starts from a cognitive behavioral framework but then incorporates techniques associated with other approaches based on the formulation.

Initial Assessment

As with a unipolar disorder, an accurate assessment of diagnosis and symptoms is necessary for effective treatment. In addition, it is particularly important in the treatment of bipolar disorder to assess the level of activity, stability of routine, and supportive social networks. This is also an important point in identifying patient beliefs regarding bipolar disorder. These could include beliefs about the benefits of mania, fears that treatment would eliminate these benefits, concerns about medication side effects, or questioning the diagnosis. These issues will be key targets of treatment in any intervention.

The current patient presented for treatment secondary to several significant life events. He reported that he had several arguments with other volunteers at his church. This led to a warning from his pastor that he would not be able to continue volunteering if these conflicts continued. The patient was also under severe financial stress. Although he received disability income, he had spent the entirety of his most recent check on religious memorabilia and charitable donations and was at risk of eviction from his subsidized housing.

The patient had a long history of his disorder. He reported that he was raised in an intact family by his mother and father and had two younger sisters. He reported that his childhood was uneventful, and his parents were loving and supportive. He had done well in high school and had joined the army. He served 4 years in the military as a communications specialist. While in the military, he married a German woman whom he met while stationed in Germany.

His first symptoms occurred while in the military. He reported a manic episode leading to hospitalization, characterized by religious delusions, elated mood, goal-directed behavior (attempting to convert other soldiers to his beliefs), excessive spending, alcohol and drug use, and sexual acting out. This first episode led to his medical discharge from the military and divorce from his wife. After discharge from the hospital, the patient reported ongoing depressive symptoms extending several years, punctuated by periods of mixed depression and mania. He reported three additional hospitalizations, one following an additional manic episode and two following suicide attempts secondary to severe depression.

Course of Therapy

Psychotherapy for bipolar disorder is an adjunctive treatment to pharmacotherapy, and initial sessions reflect this. As with unipolar disorders, treatment included education about the treatment model, establishing rapport, identifying problems and goals, and assessing symptoms. In addition, the therapist assesses knowledge about bipolar disorder and provides education to both expand on that understanding and correct any misconceptions.

Therapist: What do you know about your diagnosis of bipolar one?

Patient: Well, it means I have ups and downs, and they get pretty bad.

Therapist: What else do you know about it?

Patient: Well, my doctor says I have to take my meds or else it will get worse.

Therapist: How do you feel about taking your medication?

Patient: I really don't like to.

Therapist: How come?

Patient: I'm just not me when I'm taking them. I feel like I am slowed down, and I just can't think. I just don't know if they help. I'm taking them now, and I don't feel much better. And they don't fix the real problems I've got.

Therapist: Which are?

Patient: I have no friends, I have no girlfriend, I have no job, I have no money, no one helps me, my family's disowned me. And when I'm taking my medication, I just become a lump, and none of that gets better.

Therapist: That sounds pretty depressing. I am going to write down those things you listed as items for your problem list, including job, money, dating, family, friends, and medications.

Patient: Okay.

Therapist: And one thing that stands out to me is that you aren't happy with taking medications. That is an important thing for us to deal with as well.

In this example, the question about the patient's understanding of his disorder was to actually get information for the problem list. In addition, it elicits information about concerns regarding medication. The therapist here now has the opportunity to both demonstrate the development of the problem list and, over time, educate the patient about medication use.

One issue that complicates treatment when patients are receiving combined pharmacotherapy and psychotherapy is whether the psychotherapist is also the prescribing provider. In the case where there is one provider, coordination of pharmacotherapy and psychotherapy is obviously easier, but frequently the prescriber does not have time and/or expertise to provide psychotherapy. Therefore, frequently separate providers do psychotherapy and pharmacotherapy, and therefore it is critical for these providers to be in frequent contact and coordinate.

From a CBT perspective, the key aspects of psychotherapy for bipolar disorder will include tracking symptoms and functioning, identifying and modifying cognitions that may exacerbate symptoms or impair functioning, and behavioral interventions designed to increase rewarding activities without producing manic activation. To do this, interventions can include those from other models, including the education and involvement of family or other support resources and stabilization of social rhythms through the establishment of a consistent routine as in IPSRT.

In this situation, identifying warning signs can help the patient apply skills to prevent full symptom exacerbation. However, patients often don’t notice as symptoms are worsening, and therefore it is useful to enlist family members or other support to provide feedback. For this to be effective, it is helpful to (a) meet with the support person and educate them about the disorder and the goal of tracking warning signs and (b) to work with both the patient and the support person to agree on how the support person will inform the patient about concerns and how the patient will respond. In the current case, the patient’s pastor eventually attended a session with the patient and developed a plan for tracking symptoms and responding when symptoms were exacerbated. This both provided pragmatic support and also challenged the patient’s perception that no one cared. In this case, the fact that the support person was the pastor was also beneficial, given that the patient had symptoms of hyperreligiosity when becoming manic. By educating the pastor and patient regarding these symptoms, they could quickly notice when there was a change and intervene through thought challenging, working with structure, attending to medication adherence, and working with the prescribing provider to adjust medications.

This example reviews identifying strategies to reduce the risk of relapse following recent symptom exacerbation.

Therapist: What do you think led to you feeling worse?

Patient: I think that I felt like I had to do all the work for the church to prepare for the holidays. I started doing more and more, and I got really stressed.

Therapist: All right, so taking on too much is a problem. One thing we're going to do today is work on a list of warning signs that your symptoms are getting worse. Then we have to figure out how to help you notice when they come up and do something about them. Do you have anyone around whom you trust who could help you notice when this is happening?

Patient: No. You know that.

Therapist: I was thinking about co-workers, like one of the other volunteers or your pastor. You've mentioned that he has been really supportive, and he knows about your symptoms.

Patient: Yeah, I know he gets worried about me.

Therapist: One thing we can do is practice asking him to give you feedback about your actions and to let you know when he notices warning signs that your symptoms are getting worse.

Patient: I just worry that I'm asking too much.

Therapist: Well, let's practice in here how you would ask. Also, let's generate some other ways to notice warning signs.

During the course of treatment of bipolar disorder, the role of the therapist is to facilitate the patient identifying warning signs, implementing skills practice, and providing support and encouragement (Table 13.7–3).

Table 13.7–3.
Examples of Warning Signs

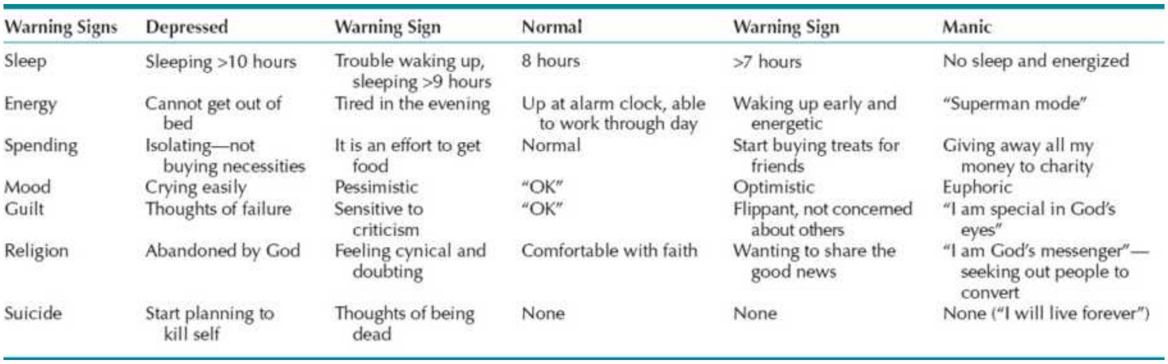

Warning Signs	Depressed	Warning Sign	Normal	Warning Sign	Manic
Sleep	Sleeping >10 hours	Trouble waking up, sleeping >9 hours	8 hours	>7 hours	No sleep and energized
Energy	Cannot get out of bed	Tired in the evening	Up at alarm clock, able to work through day	Waking up early and energetic	“Superman mode”
Spending	Isolating—not buying necessities	It is an effort to get food	Normal	Start buying treats for friends	Giving away all my money to charity
Mood	Crying easily	Pessimistic	“OK”	Optimistic	Euphoric
Guilt	Thoughts of failure	Sensitive to criticism	“OK”	Flippant, not concerned about others	“I am special in God’s eyes”
Religion	Abandoned by God	Feeling cynical and doubting	Comfortable with faith	Wanting to share the good news	“I am God’s messenger”—seeking out people to convert
Suicide	Start planning to kill self	Thoughts of being dead	None	None	None (“I will live forever”)

Later sessions emphasize not only the continued practice of learned skills but challenging core beliefs that may contribute to the risk of symptom relapse. For the current patient, the belief “I’m damaged and will never be accepted by anyone” was a key theme that came up. He perceived his bipolar disorder as a flaw that meant that he could never have a “normal” life and that others would always reject him. Assumptions that arose from this belief were that he could never have a romantic relationship, that he could never reconnect with his family, and that he could never have paid employment. He also was ambivalent about his treatment (particularly medication), doubting that he could make any improvement. Therefore, the later sessions focused on both challenging unhelpful beliefs and developing alternatives to unhelpful core beliefs.

At the end of treatment, the goals are for the patient to be able to identify symptom changes, have a range of skills with which to respond to those changes, and have improved communication with their pharmacotherapist so that medication adherence is consistent and appropriate changes in pharmacotherapy can be made as needed. In addition, if family members are available, the goal is that the patient and family have developed an improved understanding of the disorder and communication skills so as to facilitate a stable mood.

FUTURE DIRECTIONS

Some basic guidelines can be derived from treatment. In general, it is incumbent on the clinician to apply interventions for which there is empirical support. In the case of mood disorders, there are several psychotherapeutic options that are now available. There are some patient characteristics that could recommend treatment, including the presence of significant marital distress, amenability to a particular validated treatment modality, and chronic depression.

An important component of good psychotherapy, regardless of the model, is that the therapist has a treatment rationale that guides decisions regarding interventions. This is a shared theme of all of the current validated treatments for mood disorders. However, such a model is not likely sufficient for effective treatment. Psychotherapies generally are built around mediational models of change, with modification of the target symptom (e.g., depressed mood) being achieved through intervention on a more accessible construct (e.g., an irrational belief, a rewarding activity, increase in medication adherence, or a transference reaction in therapy). What distinguishes excellent clinicians, regardless of orientation, is their ability to hold a conceptual framework in mind and to apply that framework to the unique challenges of the patient.

FURTHER READINGS

*Beck JS. *Cognitive Therapy: Basics and Beyond*. Guilford Press; 1995.

Beck AT, Rush AJ, Shaw BF, Emery G. *Cognitive Therapy of Depression*. Guilford Press; 1979.

Chiang K-J, Tsai J-C, Liu D, Lin C-H, Chiu H-L, Chou K-R. Efficacy of cognitive-behavioral therapy in patients with bipolar disorder: a meta-analysis of randomized controlled trials. *PloS One*. 2017;12(5):e0176849.

*Cuijpers P, Karyotaki E, de Wit L, Ebert DD. The effects of fifteen evidence-supported therapies for adult depression: a meta-analytic review. *Psychother Res*. 2020;30(3):279–293.

Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. *JAMA Psychiatry*. 2019;76(7):700–707.

*Cuijpers P, Noma H, Karyotaki E, Vinkers CH, Cipriani A, Furukawa TA. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry*. 2020;19(1):92–107.

Dimidjian S, Hollon SD, Dobson KS, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol*. 2006;74:658–670.

Driessen E, Abbass AA, Barber JP, et al. Which patients benefit specifically from short-term psychodynamic psychotherapy (STPP) for depression? Study protocol of a systematic review and meta-analysis of individual participant data. *BMJ Open*. 2018;8(2):e018900.

Ekers D, Webster L, Van Straten A, Cuijpers P, Richards D, Gilbody S. Behavioural activation for depression; an update of meta-analysis of effectiveness and sub group analysis. *PLoS One*. 2014;9(6):e100100.

Elkin I, Shea MT, Watkins JT, et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Arch Gen Psychiatry*. 1989;46(11):971–982.

Frank E. *Treating Bipolar Disorder: A Clinician's Guide to Interpersonal and Social Rhythm Therapy*. Guilford Press; 2005.

Frank E, Kupfer DJ, Perel JM, et al. Three year outcomes for maintenance therapies in recurrent depression. *Arch Gen Psychiatry*. 1990;47(12):1093–1099.

Frank E, Kupfer DJ, Thase ME, et al. Two-year outcomes for interpersonal and social rhythm therapy in individuals with bipolar I disorder. *Arch Gen Psychiatry*. 2005;62(9):996–1004.

*Furukawa TA, Shinohara K, Sahker E, et al. Initial treatment choices to achieve sustained response in major depression: a systematic review and network meta-analysis. *World Psychiatry*. 2021;20(3):387–396.

*Furukawa TA, Weitz ES, Tanaka S, et al. Initial severity of depression and efficacy of cognitive-behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. *Br J Psychiatry*. 2017;210(3):190–196.

Hollon SD, DeRubeis RJ, Fawcett J, et al. Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder: a randomized clinical trial. *JAMA Psychiatry*. 2014;71(10):1157.

Jarrett RB, Minhajuddin A, Gershenfeld H, Friedman ES, Thase ME. Preventing depressive relapse and recurrence in higher-risk cognitive therapy responders: a randomized trial of continuation phase cognitive therapy, fluoxetine, or matched pill placebo. *JAMA Psychiatry*. 2013;70(11):1152.

*Karyotaki E, Efthimiou O, Miguel C, et al; Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry*. 2021;78(4):361–371.

Keller MB, McCullough JP, Klein DN, et al. A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *N Engl J Med*. 2000;342(20):1462–1470.

*Klerman GL, Weissman MM, Rounsaville BJ, Chevron ES. *Interpersonal Psychotherapy of Depression*. Basic Books; 1984.

Linehan MM, Korslund KE, Harned MS, et al. Dialectical behavior therapy for high suicide risk in individuals with borderline personality disorder: a randomized clinical trial and component analysis. *JAMA Psychiatry*. 2015;72(5):475–482.

Miklowitz DJ, Efthimiou O, Furukawa TA, et al. Adjunctive psychotherapy for bipolar disorder: a systematic review and component network meta-analysis. *JAMA Psychiatry*. 2021;78(2):141–150.

*Miklowitz DJ, Goldstein MJ. *Bipolar Disorder: A Family Focused Treatment Approach*. Guilford Press; 1997.

Miklowitz DJ, Otto MW, Frank E, et al. Psychosocial treatments for bipolar depression: a 1-year randomized trial from the Systematic Treatment Enhancement Program. *Arch Gen Psychiatry*. 2007;64(4):419–426.

*Segal ZV, Williams JMG, Teasdale JD. *Mindfulness-Based Cognitive Therapy for Depression: A New Approach to Preventing Relapse*. Guilford Press; 2002.

Thase ME, Greenhouse JB, Frank E, et al. Treatment of major depression with psychotherapy or psychotherapy-pharmacotherapy combinations. *Arch Gen Psychiatry*. 1997;54(11):1009.

Thase ME, Wright JH, Eells TD, et al. Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT. *Am J Psychiatry*. 2018;175(3):242–250.

*Weitz ES, Hollon SD, Twisk J, et al. Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: an individual patient data meta-analysis. *JAMA Psychiatry*. 2015;72(11):1102–1109.

Wright JH, Brown GK, Thase ME, Basco MR. *Learning Cognitive-Behavior Therapy: An Illustrated Guide*. American Psychiatric Publishing; 2017.

Wright JH, Owen JJ, Richards D, et al. Computer-assisted cognitive-behavior therapy for depression: a systematic review and meta-analysis. *J Clin Psychiatry*. 2019;80(2):18r12188.

▲ 13.8 Mood Disorders: Neurobiology

MICHAEL E. THASE, M.D., VLADIMIR MALETIC, M.D., M.S., AND CHARLES L. RAISON, M.D.

It has been thought for centuries that at least some forms of depression are either caused by or maintained by disturbances of brain function and since the 1960s, it has been possible to study selected neurobiologic processes that are linked to the etiology and pathogenesis of mood disorders. Several key sets of observations have informed research in this area. First, the heritability of mood disorders has suggested that the neurobiologic underpinnings of depression may be related to specific genes, of which more than 100 have been identified. Second, a more detailed understanding of the neurobiology of stress response has informed interactive stress-diathesis models of vulnerability. Third, the discovery of the first generation of “somatic” treatments (i.e., ECT and the TCA and monoamine oxidase inhibitor [MAOI] antidepressants) in the 1940s and 1950s pointed to potentially reversible neurobiologic targets for intervention. As the methodology for studying the neurobiology of mood disorders has grown more sophisticated, investigations utilizing indirect indicators of brain function such as the levels of monoamine metabolites or cortisol, urine, plasma, or cerebrospinal fluid (CSF), have largely been supplanted by translationally guided studies of gene transcripts and proteomics. Likewise, coarse measures of regional brain function, such as recordings of evoked potentials or electroencephalographic (EEG) activity patterns during waking and sleeping, largely have given way to neuroimaging strategies that permit the activity of regions or specific neural circuits to be examined at rest and during provocative challenges. This chapter will survey the neurobiology of mood disorders, emphasizing clinically relevant associations with phenomenology, course of illness, and treatment outcomes.

CLINICAL PHENOMENOLOGY

The signs, symptoms, and subjective experiences associated with depression have long suggested dysfunction of basic central nervous system (CNS) processes. With respect to cortical function, depression involves multiple disturbances of information processing. Most depressed people automatically interpret experiences from a negative perspective, and their access to memory is negatively biased. In more severe depressive states, cognition and problem-solving skills are further compromised by poor concentration and decreased ability to use abstract thought. A virtual monologue of negative thoughts and images seems to run on autopilot, and unlike normal states of sadness, ventilation to a confidante has little beneficial effect. In more extreme cases, delusions or hallucinations, or both, grossly distort reality testing. These neurocognitive changes point to dysfunction involving the hippocampus, prefrontal cortex (PFC), the amygdala, and other limbic structures.

Another biologically based characteristic of depression involves decreased interest and a loss of mood reactivity. Spontaneous, goal-directed activities decrease, and events that *should* improve mood have little or no effect. One correlate of loss of interest is decreased reinforcer salience. Even basic appetitive functions, such as appetite and libido, are diminished in severe depression. Anhedonia and decreased appetitive behavior point to dysfunction of the neural circuits involved in the anticipation and consummation of rewards, which involve the thalamus, hypothalamus, nucleus accumbens, anterior cingulate, and PFC.

Alterations in the initiation, expression, and spontaneity of movement are common in severe depression. A background state of decreased animation, referred to as psychomotor retardation, is sometimes accompanied by a superimposed state of agitation. In addition to the pacing and frequent postural shifts, stereotypical signs or behaviors, such as a furrowed brow, hair pulling, biting at the lips or nail beds, and compulsive scratching, are observed in severe depressive states. Anhedonia and psychomotor disturbances are more pronounced among older depressed patients and in recurrent depressive episodes, with agitation typically limited to the most severe depressive syndromes. These observable signs point to the dysfunction of subcortical circuits connecting the thalamus, basal ganglia, and striatum.

Almost all depressed people report low energy or fatigue and some difficulty involving sleep. Although insomnia is much more common (prevalence rates of two-thirds to four-fifths in most clinical samples), hypersomnolence is not uncommon, particularly earlier in life. Clinicians have long noted an association of hypersomnia with other “reversed” neurovegetative signs and symptoms, such as increased appetite and weight gain. The classical sleep disturbances of depression, such as difficulty maintaining sleep and early morning awakening, also tend to be more pronounced and severe in older individuals with a history of recurrent illness. Such terminal insomnia is in turn linked to diurnal mood variation, in which morning is the worst time of the day. Disturbances of these normally well-entrained circadian rhythms are indicative of dysregulation of nuclei in the thalamus, hypothalamus, and brain stem.

For decades, clinical researchers have viewed the tendency for these signs and symptoms to coaggregate or cluster together to reflect the neurobiologic basis for the classical subtypes of depression. For example, anhedonia, loss of mood reactivity, and psychomotor disturbance are more commonly associated with weight loss, terminal insomnia, diurnal mood disturbance, or a combination of these. This constellation forms the basis of the syndrome of *melancholia* (also known over the past five decades as *autonomous*, *biologic*, or *endogenous* forms of depression). As these depressions are often recurrent, and discrete episodes can begin without the provocation of a significant stressor, they were proposed to have an *internal* (i.e., neurobiologic) etiology. The fact that such depressions were slow to remit spontaneously, yet were quite responsive to ECT further reinforced perceptions of underlying biologic dysfunction. The stable nature of this syndrome over the centuries has been recognized by the contemporary use of the ancient Greek term *melancholia*, even though black bile is no longer implicated in the pathophysiology of severe depressive states.

In contrast to *melancholia*, the depressions more commonly seen in young adult life often seemed inextricably tied to personality problems, interpersonal problems, and maladaptive responses to psychosocial difficulties. The terms *nonendogenous* and *reactive* were used to represent the observation that depression in young adults is almost invariably associated with stressful life events. As suggested by the names, these depressions were presumed to be mediated by nonbiologic factors related to personality and resilience to stress. The association of an early age of onset with co-occurring anxiety (both symptoms and discrete disorders)—coupled with a low incidence of psychotic symptoms (i.e., delusions or hallucinations)—further solidified impressions that younger depressed patients had a *neurotic* disorder. Another hallmark of a putatively neurotic disorder, persistence or chronicity, also characterized by many early-onset depressions. The presumed primacy of psychosocial factors in neurotic or nonendogenous depression was further suggested by the too frequently poor responses to ECT and TCAs. That the onset of depression in early life may, of course, distort personality development, limit one's repertoire of coping skills and diminish resilience, change brain structure and function, and alter drug responsivity (i.e., pathoplasty) was not yet appreciated. It was also not yet known that early-onset depressions are associated with a relatively high familial loading in genetic studies (i.e., a hallmark of a biologically mediated disorder). Likewise, neither the heritability of neuroticism nor the neural basis of the characteristically increased reactivity to emotional stimuli had yet been established.

One historically important subgroup of patients with nonendogenous or neurotic depression was defined in part by *reverse* neurovegetative symptoms (i.e., increased appetite, weight gain, and hypersomnolence). As the patients treated by the hospital-based psychopathologists of the 1940s and 1950s seldom had such symptoms, this presentation was described as *atypical depression*. The notion of atypicality was reinforced by a relatively poor response to ECT and TCAs, and by contrast, a relatively preferential response to MAOIs. Importantly, as clinical research shifted to ambulatory settings, and greater numbers of younger and less severely symptomatic people began to seek treatment for mood disorders, it was recognized that about 15% to 25% of depressed people have reverse vegetative features, which means that—despite the name—this presentation was not so atypical. Furthermore, contemporary studies used the correlation between polygenic risk scores, functional imaging studies, and phenotypical and metabolic markers in order to “dissect” depression subtypes based on their neurobiologic underpinning. Atypical depression was noted to share risk genes with increased appetite/BMI/obesity, upregulated inflammatory signaling (higher C-reactive protein [CRP]), insulin resistance, elevated leptin and leptin resistance, and increased insula response to appetitive stimuli.

RISK FACTORS

Genetic Influences

Mood disorders and suicide clearly run in families, yet the contributions of nature and nurture have provoked intense debates for decades. It is now well established that bipolar affective disorder is more heritable than other mood disorders; heritability increases in proportion to the amount of shared genetic material, and early age of onset is associated with greater heritability. Thus, identical twins have a greater heritable risk than fraternal twins, who are at approximately the same risk as other first-degree relatives (i.e., siblings and parents). First-degree relatives, in turn, have a greater shared risk than half-siblings, grandparents, or cousins. Research comparing same-sex fraternal and identical twins and studies using the adopted-away paradigm further demonstrated that heritable risk transcends environmental influences. Research using modern molecular genetic methods will eventually lead to the identification of the gene products that transmit this risk. Already, it is clear that a number of alleles are associated with a small but replicable increase in the risk of depression, yet others may be associated with relative resilience. Other genes, while showing no relation to the risk of depression, are relevant to response to specific antidepressants, primarily by increasing or decreasing drug metabolism via Cytochrome P450 enzymes.

Although the significance of genetics cannot be disputed, considerable variability exists within and across groups with comparable heritable risks. For example, identical twins do not have 100% concordance of bipolar affective disorder. Early-onset depressions are more heritable, and conversely, late-onset mood disorders (i.e., after 60 years of age) have the lowest rates of heritability, presumably because of the greater impact of acquired risk factors, including cerebrovascular disease and selected dementing illnesses. In some pedigrees, men are at greater risk of alcoholism and sociopathy, and women are at greater risk of depression. There is some evidence that different environmental experiences may influence whether a woman develops depression or a generalized anxiety disorder. Conversely, for men, alcohol or drug use disorders may constitute an alternate illness trajectory to early-onset depressive disorders. Genetics thus constitutes just one pathway of vulnerability.

Predispositions to various emotional states (i.e., mood set points or temperament) also appear to be partly heritable. Of particular importance is a temperament characterized by excessive emotional reactivity and behavioral inhibition. Modern theories of temperament and personality converge around several related dimensions, including high levels of harm avoidance (neuroticism) and low levels of novelty seeking (introversion).

These temperaments, which can be recognized in infancy, are associated with increased vulnerability to depression and anxiety, presumably via increased limbic reactivity and heightened sensitivity to the impact of stressful life events. Childhood-onset of dysthymia similarly presages extremely high rates of depression and bipolar disorder in adulthood.

One reason for inconsistency in genetic research is the process known as epigenetics. This term refers to functionally relevant modifications of genes—typically induced by DNA methylation or histone modification—in response to environmental stimulation that do not actually change the DNA sequence of the gene. Epigenetic changes in gene expression, both activating and inhibiting, are typically irreversible (i.e., will last for the life of the cell) and may persist across multiple generations. Intensive ongoing research indicates that some of the changes in brain function attributable to the impact of chronic stress may be mediated by epigenetic changes in key regions of the brain.

Recent synthetic approaches attempting to elucidate common etiologies within the otherwise biologically heterogeneous condition of MDD combined large-scale genome-wide association studies (GWAS) with postmortem cytology and functional and structural imaging studies. Consistent associations were noted between principal neuroimaging findings in depressed patients and downregulated genetic markers for cortical somatostatin (SST)-expressing gamma aminobutyric acid (GABA) interneurons and astrocytes. Polygenic SST interneuron markers were most expressed in the subgenual anterior cingulate (sgACC), medial PFC (mPFC), anterior insula, and temporal lobes coinciding with brain areas where imaging studies confirmed cortical thinning and aberrant connectivity in depressed patients. By linking specific genetic changes to functional/structural consequences in the brain, these types of combined methodologies may help both identify depressogenic mechanisms and more biologically homogeneous subtypes of depression.

ETIOLOGY OF EMOTIONS

Some consider depression to be an extreme expression of sadness, a normal mood, with elation to be the healthier counterpart of the euphoria of mania. Such continuity between normal and pathologic mood states is illustrated by bereavement. On the one hand, grief is a universal experience, yet, on the other hand, it can segue into a severe and disabling depressive state. Although there are less obvious parallels to mania, the intensity of new romantic love or other peak emotional experiences is associated with changes in perception, cognition, behavior, and judgment that border on manic excitement. The euphorigenic effects of cocaine and other psychostimulants amply document that the hard wiring for manic states exists in many if exposed to the right neurochemical milieu.

There is now considerable agreement that several basic emotions are observed across human cultures. Sadness and crying, for example, may be universally recognized interpersonal distress signals. Moreover, the capacity to recognize the emotional states of others develops at such an early age that innate processes are undoubtedly implicated.

At the risk of anthropomorphizing excessively, basic emotional states are also observable across mammalian species. Behavioral predispositions toward aggressivity (or passivity) and social dominance (or subordination) appear to extend across vertebrate species. However, by virtue of a large neocortex, humans are distinguished by a greater capacity to modify the expression of basic emotional states via abstraction, integration, and synthesis of complex symbolic representations, communication of experiences to others in direct or elaborated forms, and development of self-concepts that guide behavior, both in relation to others and an anticipated future. The domains of competence are similarly broader and diverse. Whereas a highly competent primate may quickly gain a position of dominance if placed in a new troupe, only the human can intentionally misrepresent their competence, conceal motivations, or be thwarted by archival documentation of past behaviors.

When compared to other mammals, the relatively greater importance and intensity of attachment bonds facilitate the protracted task of human child-rearing and enhance the advantages of kinship. Indeed, for hundreds of thousands of years, the ancestors of man lived in a world in which survival into adulthood was uncommon. Indeed, it is only in the past few centuries that average life expectancy has exceeded 40 years. Young humans are arguably the most vulnerable of mammals and certainly remain dependent on their caregivers for the longest period of time. Certain necessary divisions of labor no doubt increased the chances of viable offspring and may have slowly shaped sex differences in affectivity and affiliative behavior that in turn may help to underpin gender differences in the risk of depressive disorders. Ethologic considerations also may help to explain why rates of depression appear to be increasing. The profound changes in social systems and lifestyles that have taken place in the past several centuries have forced adaptations that are too rapid for natural selection. In relative terms, the breathtaking sociocultural and technologic changes of the past century have taken place in less than 0.00001th of hominid experience! The stressors faced in modern life thus present complexities that despite such a large neocortex and complex social networks, sometimes overwhelm basic adaptive capacities.

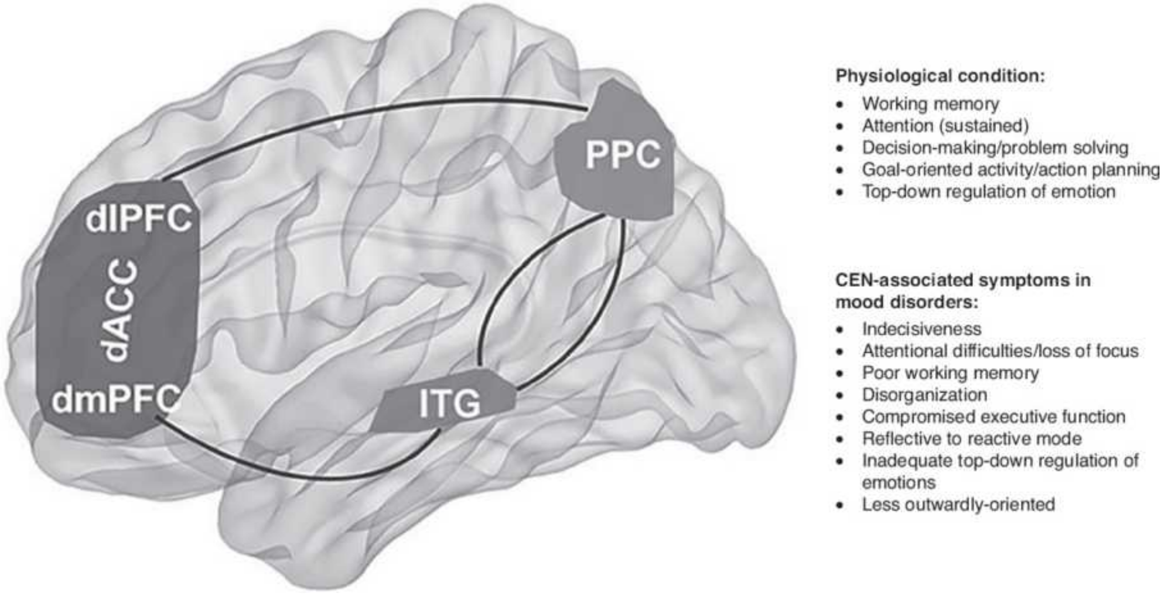

Physiological condition:

- Working memory
- Attention (sustained)
- Decision-making/problem solving
- Goal-oriented activity/action planning
- Top-down regulation of emotion

CEN-associated symptoms in mood disorders:

- Indecisiveness
- Attentional difficulties/loss of focus
- Poor working memory
- Disorganization
- Compromised executive function
- Reflective to reactive mode
- Inadequate top-down regulation of emotions
- Less outwardly-oriented

FIGURE 13.8–1. The central executive network (CEN). CEN is centered on the dorsolateral prefrontal cortex (dlPFC) and posterior parietal cortex (PPC), and also includes the dorsomedial prefrontal cortex (dmPFC), dorsal ACC (dACC), and inferior temporal gyrus (ITG). (Figure reprinted from Bi B, Che D, Bai Y. Neural network of bipolar disorder: Toward integration of neuroimaging and neurocircuit-based treatment strategies. *Transl Psychiatry.* 2022;12:143. doi: 10.1038/s41398-022-01917-x. License: https://creativecommons.org/licenses/by/4.0/.)

EMOTIONAL PROCESSING AND THE BRAIN

Modern affective neuroscience focuses on the importance of four brain regions in the regulation of normal emotions: the PFC, the anterior cingulate, the hippocampus, and the amygdala (Fig. 13.8–1). The PFC is viewed as the structure that holds representations of goals and appropriate responses to obtain these goals. Such activities are particularly important when multiple, conflicting behavioral responses are possible or when it is necessary to override affective arousal. The dorsolateral region of the PFC (dlPFC) and posterior parietal cortex (PPC) are particularly active in the suppression of negative affect and are key constituents of the cognitive control network (CNN), topographically coinciding with the central executive network (CEN), and for the most part with the frontoparietal network (FPN) (see Fig. 13.8–1). In addition to "top-down" regulation of emotions, the CNN/CEN also supports working memory and allocation of attention. The mPFC along with the posterior cingulate cortex (PCC), inferior parietal lobe, and temporal cortical structures, including the hippocampus (see below), are important hubs in the default mode network (DMN)(Fig. 13.8–2), an interconnected circuit of regions that are active during resting wakefulness, self-referential thinking (i.e., autobiographical memory), processing social information, planning for the future or daydreaming. Contemporary studies have implicated DMN structures in the planning stages of higher cognitive activity and creative processes. The DMN appears to show both hyperconnectivity *and* hypoconnectivity in depressive disorders, which predisposes to negative recall and a failure to inhibit rumination. More recent meta-analyses have investigated the individual roles of DMN subsystems that contribute to MDD. The “Core DMN,” composed of anterior mPFC and PCC, plays a key role in self-referential processing. The “dmPFC complex,” together with the temporoparietal junction (TPJ) and lateral temporal cortex, participates in mentalizing and metacognitive representation of self and others. Finally, the medial-temporal cortex (MTL) subsystem, which also includes the dmPFC, extended hippocampus, and posterior inferior parietal cortex, underpins both reminiscing and projecting one’s future.

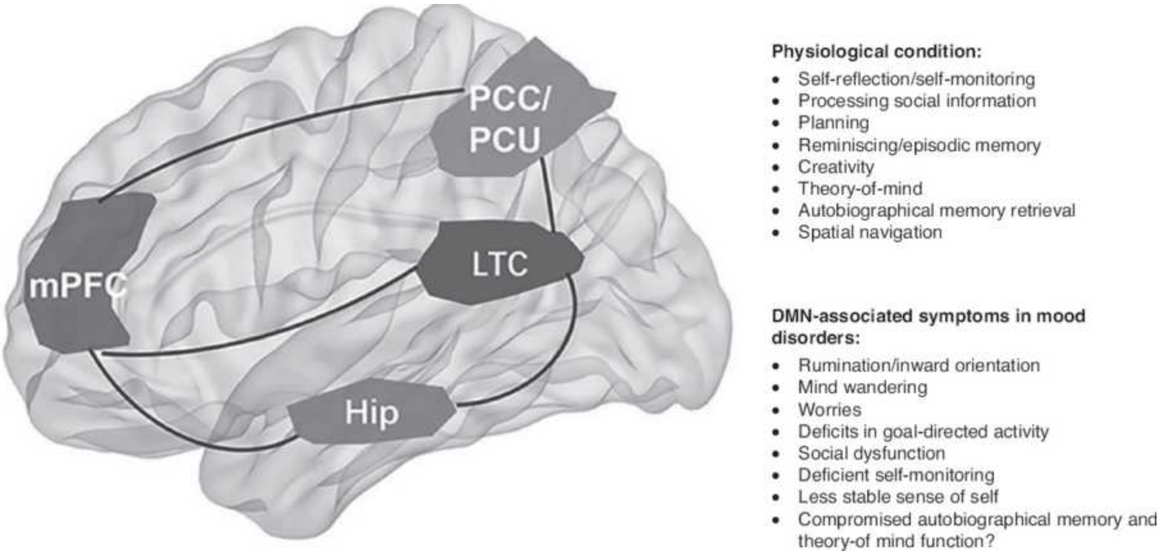
PCC/
PCU
mPFC
LTC
Hip
Physiological condition:
• Self-reflection/self-monitoring
• Processing social information
• Planning
• Reminiscing/episodic memory
• Creativity
• Theory-of-mind
• Autobiographical memory retrieval
• Spatial navigation
DMN-associated symptoms in mood disorders:
• Rumination/inward orientation
• Mind wandering
• Worries
• Deficits in goal-directed activity
• Social dysfunction
• Deficient self-monitoring
• Less stable sense of self
• Compromised autobiographical memory and theory-of mind function?

FIGURE 13.8–2. The default-mode network (DMN). DMN consists of two core regions: the medial prefrontal cortex (mPFC) and the posterior cingulate cortex/precuneus (PCC/PCU), and the inferior parietal lobule (IPL). The lateral temporal cortex (LTC) and the hippocampal formation (Hip) are often found as being strongly related to the DMN, but likely constitute a subsystem within the DMN. (Figure reprinted from Bi B, Che D, Bai Y. Neural network of bipolar disorder: Toward integration of neuroimaging and neurocircuit-based treatment strategies. *Transl Psychiatry.* 2022;12:143. doi: 10.1038/s41398-022-01917-x. License: https://creativecommons.org/licenses/by/4.0/.)

The largest meta-analyses examining the relationship between MDD and DMN (sample size of several hundred patients) have confirmed that MDD is highly heterogeneous while simultaneously establishing overall hypoconnectivity in the DMN-core as a moderate risk marker of MDD. Global DMN hypoconnectivity has also been linked with severe negative bias and anhedonia. Furthermore, DMN core and dmPFC subsystem dysfunction appear to be responsible for ruminations in a subpopulation of MDD patients. Extending these findings, recent studies have reported an association between greater "internal" DMN connectivity, as well as connectivity between DMN and SMN and FPN with response to antidepressants.

There is also evidence of some hemispherical specialization in PFC function. For example, left-sided activation of regions of the PFC is more involved in goal-directed or appetitive behaviors, whereas regions of the right PFC are implicated in avoidance behaviors and inhibition of appetitive pursuits.

The anterior cingulate cortex (ACC) is thought to serve as the point of integration of attentional and emotional inputs. Two subdivisions have been identified: an affective subdivision in the rostral and ventral regions of the ACC and a cognitive subdivision involving the dorsal ACC. The former subdivision shares extensive connections with other limbic regions, and the latter interacts more with the PFC and other cortical regions. It is proposed that activation of the ACC as part of a CCN facilitates effortful control of emotional arousal, particularly when goal attainment has been thwarted or when novel problems have been encountered. There is evidence that connectivity among the components of the CCN is diminished in depressive states, and if persistent, such decreased connectivity may contribute to relapse/recurrence risk.

The salience network (SAN) attributes emotional meaning to sensory experiences including anxious and depressed feelings. Abbreviation SAN will be used in place of the more commonly utilized SN for the purpose of disambiguation from the substantia nigra (SN). Furthermore, SAN structures are key participants in motivational and reward signaling. Moreover, aberrant connectivity between components of SAN (anterior ACC) and DMN (PCC) has been associated with the intensity of depressive ruminations and the duration of depressive episodes. It is as if the negative affect generated by SAN to fully engage CCN/CEN during cognitive tasks without adequately suppressing DMN activity may hinder cognitive performance due to interference stemming from ruminations and/or daydreaming. This pattern of maladaptive functional activity is more common in individuals with child/adolescent onset of depression. Diminished SAN connectivity has been linked with several clinical correlates of the depressed state, such as inattention, negative emotional bias, anxious avoidance, and anhedonia.

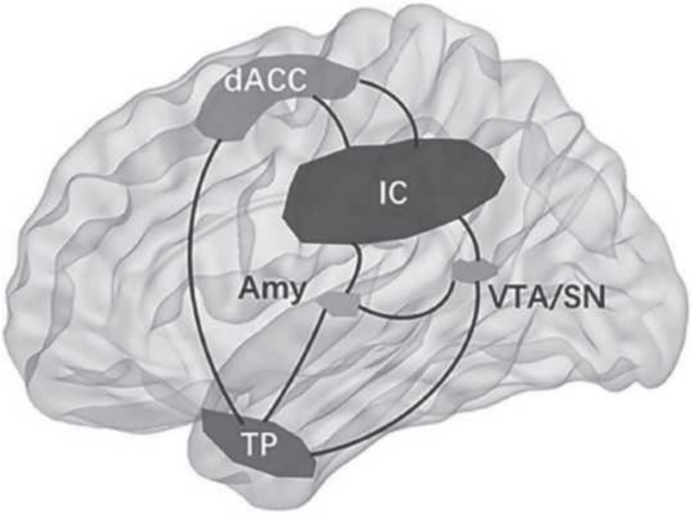

Physiological condition:

- Salience attribution
- Detection and autonomic reaction to salient sensory stimuli
- Interoceptive awareness
- Reward signaling
- Involved in switching between DMN and CEN

SN-associated symptoms in mood disorders:

- Low motivation/enthusiasm
- Anhedonia
- Negative emotions (anxiety and depression), anxious avoidance
- Inadequate emotional control
- Inattentiveness
- Symptom severity
- Elevated pre-treatment SN activity is associated with antidepressant and rTMS response

FIGURE 13.8–3. The salience network (SAN). The salience network (SAN) consists of the insular cortex (IC), dorsal anterior cingulate cortex (dACC), amygdala (Amy), ventral tegmental area (VTA), substantia nigra (SN). (Figure reprinted from Bi B, Che D, Bai Y. Neural network of bipolar disorder: Toward integration of neuroimaging and neurocircuit-based treatment strategies. *Transl Psychiatry.* 2022;12:143. doi:10.1038/s41398-022-01917-x. License: https://creativecommons.org/licenses/by/4.0/.)

The hippocampus is most clearly involved in various forms of learning and memory, including fear conditioning, as well as inhibitory regulation of HPA-axis activity. Importantly, the hippocampus is also one of the few sites of the brain capable of neurogenesis (i.e., the process of differentiation of stem cells into potentially functional neurons). Emotional or contextual learning appears to involve a direct connection between the hippocampus and the amygdala.

The amygdala appears to be a crucial way station for processing novel stimuli of emotional significance and coordinating or organizing cortical responses. Located just above the hippocampi bilaterally, the amygdala has long been viewed as the functional center of the limbic system. Although most research has focused on the role of the amygdala in responding to fearful or painful stimuli, novelty and ambiguity can also bring the amygdala "on line."

Finally, one recent review of intrinsic functional networks includes a sensory-motor or somatomotor network (SMN) (Fig. 13.8–4). The sensory-motor network is composed of the middle cingulate cortex (MCC), dorsal striatum, ventral thalamic nuclei, sensory (postcentral) gyrus, motor (precentral) gyrus, and pre- and supplemental motor areas. The SMN supports sensory processing, as well as planning and execution of complex motor behaviors. It has been posited that mania is associated with elevated activity of SAN and SMN and diminished DMN-based self-monitoring. In contrast, depressive states may be marked by reduced SMN activity, coinciding with psychomotor retardation and increased DMN function, reflected by internally focused thinking and rumination. The ratio between DMN and SMN activity was found to strongly correlate with the intensity of depression. Psychomotor abnormalities due to aberrant functional network activity may represent a trans-diagnostic phenomenon. Furthermore, "supremacy" of SMN relative to DMN activity is most likely clinically associated with compromised self-monitoring, sensory hyperacuity, emotional reactivity, and psychomotor agitation in the context of mania but also agitated unipolar and bipolar depression. Based on a recent large meta-analysis of the brain's functional connectome involving over 600 MDD patients, internal hypoconnectivity within SMN seems to be the most stable imaging marker of MDD.

Any cogent discussion of functional neural circuits involved in the etiology of mood disorders would be incomplete without mention of the reward circuit and cortico–striatal–thalamo–cortical (CSTC) circuit (see Fig. 13.8–5). Malfunction of the reward circuit, comprised of the dopaminergic ventral tegmental area (VTA), ventral striatum, rostral ACC, and ventromedial (vmPFC)/orbital PFC, likely underpins anhedonia, loss of interest, low motivational states, and even some cognitive manifestations of depressive episodes. The CSTC circuit is a highly integrated "updating" system comprised of three parallel feedback loops. In some ways, CSTC functions as a "backbone" of other major functional networks (SAN, DMN, CEN, SMN) as it shares anatomical locations and pathways with the hubs of these networks and continuously updates them with new multimodal information. The CSTC comprises three parallel loops: (1) the associative/cognitive division, which includes dlPFC, dorsal caudate, pallidum, SN, pars reticulata (SNr), and ventral anterior (VA), and mid-dorsal (MD) thalamic nuclei; (2) the affective/limbic division, which includes vmPFC/ACC, ventral striatum, pallidum, SNr and VA, and MD thalamus; and (3) the sensory-motor division, which encompasses sensory and motor cortical areas, dorsal putamen, pallidum, SNr and ventrolateral (VL) thalamus (see Fig. 13.8–5).

Elevated dopaminergic transmission in the associative striatum and increased activity of the sensory-motor division putatively underpin psychotic symptoms, agitation, and emotional reactivity in bipolar mania, while diminished activity in the sensory-motor division may coincide with psychomotor retardation in the context of a depressive episode. Inadequate functioning of the associative/cognitive division of the CSTC loop may also be responsible for cognitive and executive function deficits in mood disorders. Moreover, reduced activity in the ventral affective/limbic division of the CSTC loop has been implicated in anhedonia and cognitive/emotional blunting during a depressive episode. Novel approaches to elucidating the role of major functional networks in generating the symptoms of mood disorders have shifted the focus away from stationary connectedness to a more dynamic perspective that highlights disturbance in their adaptive coupling and the relative duration of specific network activity (e.g., level of depression correlated with duration of DMN activity).

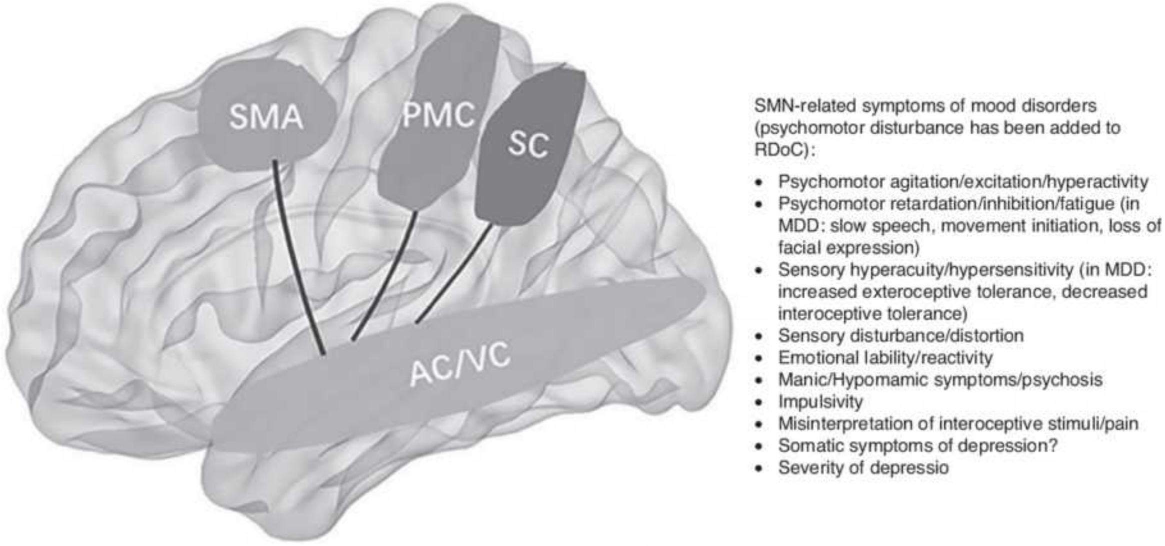

SMN-related symptoms of mood disorders (psychomotor disturbance has been added to RDoC):

- Psychomotor agitation/excitation/hyperactivity
- Psychomotor retardation/inhibition/fatigue (in MDD: slow speech, movement initiation, loss of facial expression)
- Sensory hyperacuity/hypersensitivity (in MDD: increased exteroceptive tolerance, decreased interoceptive tolerance)
- Sensory disturbance/distortion
- Emotional lability/reactivity
- Manic/Hypomamic symptoms/psychosis
- Impulsivity
- Misinterpretation of interoceptive stimuli/pain
- Somatic symptoms of depression?
- Severity of depressio

FIGURE 13.8–4. The somatomotor/sensorimotor network (SMN). SMN—which comprises the middle cingulate cortex (MCC), dorsal striatum, ventral nuclei of thalamus and postcentral gyrus (somatosensory cortex [SC]), premotor (PMC), and supplemental motor areas (SMAs)—is involved in sensory processing and motor functions. Auditory and visual cortices (AC/VC) are sometimes included. (Figure reprinted from Bi B, Che D, Bai Y. Neural network of bipolar disorder: Toward integration of neuroimaging and neurocircuit-based treatment strategies. *Transl Psychiatry.* 2022;12:143. doi:10.1038/s41398-022-01917-x. License: https://creativecommons.org/licenses/by/4.0/.)

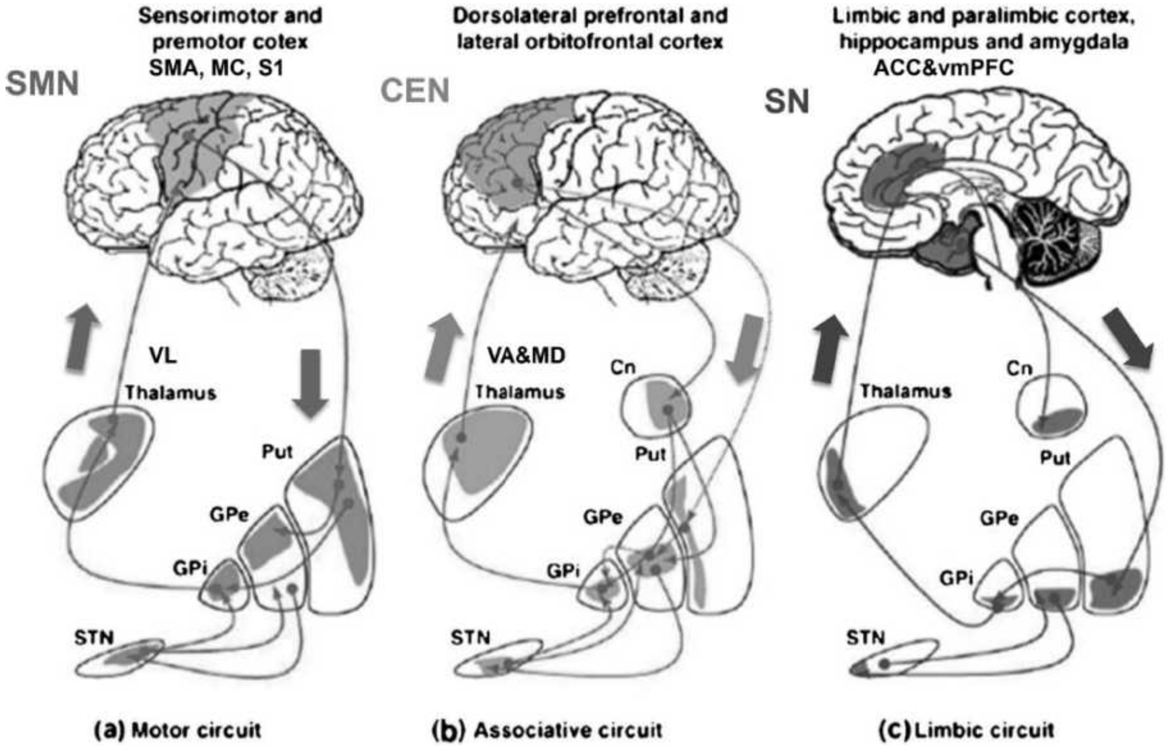
Sensorimotor and premotor cotex SMA, MC, S1
SMN
VL
Thalamus
Put
GPe
GPi
STN
(a) Motor circuit
Dorsolateral prefrontal and lateral orbitofrontal cortex
CEN
VA&MD
Thalamus
Cn
Put
GPe
GPi
STN
(b) Associative circuit
Limbic and paralimbic cortex, hippocampus and amygdala ACC&vmPFC
SN
Thalamus
Cn
Put
GPe
GPi
STN
(c) Limbic circuit

FIGURE 13.8–5. Sensory-motor, associative-cognitive, and limbic divisions of cortico–striatal–thalamo–cortical (CSTC) loop. Putamen (Put), globus pallidus pars interna (GPi), globus pallidus pars externa (Gpe), the subthalamic nucleus (STN), caudate nucleus (Cn); VA, ventral anterior thalamic nucleus; MD, mediodorsal thalamic nucleus; VL, ventral lateral thalamic nucleus; SMA, supplemental motor area; MC, motor cortex; S1, sensory postcentral gyrus; vmPFC, ventromedial prefrontal cortex. (Figure adapted from Lapidus KAB, Stern ER, Berlin HA, et al. Neuromodulation for obsessive–compulsive disorder. *Neurotherapeutics*. 2014;11:485–495. https://doi.org/10.1007/s13311-014-0287-9.)

STRESS AND ANIMAL MODELS OF DEPRESSION

Studies of smaller mammals and nonhuman primates have demonstrated that acute stress responses involve activation of central and peripheral components of two interactive psychoneuroendocrine systems: the cortical HPA axis and the sympathomedullary (SM) axis. Acute responses to stress are phenomenologically and neurobiologically more akin to fear and anxiety than depression. However, whereas stress initially signals threat, it is the anticipation of loss and the meaning of the loss—as well as the loss of hope—that can provoke the transition from anxiety to sadness and despair.

Acute stress activates cell bodies of norepinephrine (NE) neurons in the locus coeruleus (LC), whose ascending axons signal noradrenergically-mediated increases in cortical arousal and whose descending axons signal increased adrenergic (predominantly epinephrine) output by the sympathetic nervous system and adrenal medulla. Behaviorally, the perception of acute stress elicits perceptual vigilance, preparedness for response, and inhibition of appetitive activities, such as foraging and pursuit of sexual activity.

Stress simultaneously triggers and releases corticotropin-releasing hormone (CRH) from neurons in the hypothalamus and cerebral cortex. Hypothalamic CRH activates the synthesis and release of adrenocorticotropic hormone (ACTH) from the anterior pituitary, which in turn, triggers the release of cortisol and other glucocorticoids (from the adrenal cortex). CRH also synergistically increases LC activity, and directly or indirectly, increases the synthesis of other stress-reactive gene products and anti-inflammatory responses, and decreases the synthesis of key neuropeptides such as brain-derived neurotrophic factor (BDNF). The excitatory amino acid glutamate may synergistically heighten CNS arousal in response to stress. The acute response to stress is counterbalanced by homeostatic or adaptive mechanisms. These include feedback inhibition by glucocorticoid receptors in the hippocampus and pituitary, downregulation of postsynaptic noradrenergic receptors, and inhibitory auto- and heteroreceptors on presynaptic NE neurons. As is discussed subsequently, parallel input from serotoninergic, GABA, and glycinergic neurons exert dampening or inhibitory effects.

Exposure to prolonged, inescapable stress is associated with numerous adaptations in neurobehavioral responses. Although CRH levels in the brain and corticosteroid levels in the periphery may remain elevated, levels of NE, 5-HT, dopamine (DA), and GABA in the brain stem and forebrain eventually decrease. Animals trapped in such a state eventually cease trying to resist or to escape and show decreased grooming and appetitive behavior. The behavior persists outside of the experimental situation. There are significant individual differences in the development of such learned helplessness, as well as differences across pedigrees, strains, and species. Like other forms of chronic stress, the negative behavioral effects of learned helplessness may be sustained by epigenetic changes in key regions of the brain. Importantly, antidepressant drugs have been shown to prevent, attenuate, or reverse learned helplessness across species.

Studies of the experiences of primates in the wild, of course, are more relevant to the stresses faced by people than a rat's response to an inescapable, painful electric shock. These naturalistic studies demonstrate not only that central 5-HT activity is partly under genetic control, but also that loss of a dominant role within a primate social hierarchy results in decreased 5-HT neurotransmission and increased HPA. In the wild, animals with low levels of the serotoninergic metabolite 5-hydroxyindoleacetic acid (5-HIAA) are likely to be more aggressive, more socially subordinate, and less sexually active. When social status is manipulated by creating a new troupe from a cohort of subordinate primates, a new dominance hierarchy emerges, and the more dominant animals experience a corresponding increase in 5-HT function. Conversely, during times of adversity, such as drought or famine, it is the socially dominant primates that experience the largest increases in HPA activity.

MALTREATMENT AND EARLY ADVERSITY

Physical, verbal, and sexual abuse and parental neglect have an indelible effect on one's life trajectory. Although it has been known for decades that such maltreatment plays a key role in the pathogenesis of PTSD and borderline personality disorder, it is now clear that such a history results in a two- to threefold increase in the risk of depression and is associated with important phenomenologic and biologic differences in depressive states.

Studies of animal models confirm that lasting alterations in neuroendocrine and behavioral responses can result from severe early stress. Animal studies further indicate that even transient periods of maternal deprivation can alter subsequent responses to stress. This vulnerability has been shown to be the result of enduring changes in gene expression. For example, protracted stress thus can induce changes in the functional status of neurons via inducing epigenetic changes and may suppress neurogenesis in the hippocampus. Recent studies in depressed humans indicate that a history of early trauma is associated with long-lasting changes in responsivity of the HPA axis, including accompanied by structural changes in the hippocampus and perhaps other focal regions within the cerebral cortex.

MONOAMINE SYSTEMS: FUNCTION AND DYSFUNCTION

An area of scientific inquiry is chosen in part because of the knowledge base and experimental paradigms available to investigators. In the early 1960s, nobel prize-winning research had just elucidated the importance of monoamines in neurotransmission. Moreover, it was possible to measure the metabolites of the catecholamine NE in various body fluids and 5-HIAA, the principal metabolite of 5-HT, in CSF. Visualization of the functional brain, by contrast, was essentially limited by the low sensitivity of recording neuronal activity from the surface electrodes used for EEGs. Specifically, whereas the paroxysmal spikes in electrical activity caused by epileptogenic foci or the diffuse slowing associated with delirium could be documented by EEGs, the more subtle changes in brain activity or regional cerebral metabolism associated with affective states could not.

In addition, there were multiple lines of evidence from pharmacologic studies implicating perturbations of monoamine systems in the therapeutic and iatrogenic effects of various medications. The relevant pharmacotherapies of the time, the TCAs and MAOIs, were known to increase the amount of neurotransmitter available at monoamine synapses. Conversely, the monoamine-depleting compound reserpine was known to cause depression as a side effect during the treatment of hypertension. Intoxication with potent noradrenergic agonists, such as amphetamine and cocaine, also was known to induce states of sleepless euphoria and activated psychosis, followed by a crashing, depression-like withdrawal.

The early monoamine hypotheses—formulated by many of the pioneers of the first generation of psychopharmacologists, such as Joseph J. Schildkraut, Alexander H. Glassman, Arthur J. Prange Jr., John Davis, and Herman van Praag, are now obsolete as it is known that these ancient neuromodulatory systems comprise only a small fraction of the brain’s neurons. Nevertheless, NE and 5-HT neurotransmission is still thought to play a role in the pathophysiology of mood disorders, and most antidepressant medications initiate their actions through these neurons. A third monoamine, the catecholamine DA, was thought to play a greater role in the pathophysiology and treatment of schizophrenia and thus received less emphasis. Nevertheless, evidence of DA dysfunction in at least some forms of depression has emerged. Knowledge about monoamine receptor families, second messengers, neurokinins, excitatory amino acids, and gene transcription factors has added layers of complexity to the original monoamine hypotheses. As a result, there has been a progressive shift away from focusing on disturbances of single neurotransmitter systems in favor of studying neurobehavioral systems, neural circuits, and more intricate neuroregulatory mechanisms.

Supporting this shift away from a single neurotransmitter focus, recent clinical studies and meta-analyses correlating CSF monoamine metabolites with clinical diagnosis, as well as GWAS and studies using pharmacologic probes to evaluate monoamine transmitter/receptor function in MDD have cast further doubt on the primacy of monoamine dysfunction in MDD. A meta-analysis involving 11 individual studies comparing CSF levels of the NE metabolite 3-methoxy-4-hydroxyphenylglycol (MHPG) and the 5-HT metabolite 5-HIAA found no significant difference between controls and MDD patients. Another study evaluating altered monoamine signaling in MDD by using various pharmacologic “probes” found that approximately 40% of MDD patients had no measurable aberrations of NE and 5-HT signaling. Similarly, a GWAS study found no evidence of involvement of monoamine-related genes in the etiology of MDD.

When considered together, recent studies indicate two more indirect relationships between monoamine and the pathogenesis and treatment of MDD. In regard to pathogenesis, it is increasingly clear that monoamines play an important role in “fine-tuning” the activities and coupling of functional brain networks. Such an explanatory model recognizes the “two-pronged” influence that monoamine projections may exert on functional networks involved in the regulation of affect, cognition, and motoric activity. Monoamine projections emanating from the brainstem directly innervate the hubs of major functional networks (SN, DMN, CEN, and SMN), primarily regulating their basal activity and connectivity. In addition, monoamines moderate a more dynamic influx of “updating signals” originating from the CSTC loop. Dopaminergic fibers originating from the SN exert an influence on the associative/cognitive and motor striatum, while the VTA provides input to the limbic/affective section of the striatum. DA fibers also project to the dorsomedial thalamus, thereby regulating ascending glutamatergic projections (final branch, closing the CSTC loop) from the basal ganglia and thalamus to the cortical components of major functional networks. In a similar fashion, noradrenergic fibers from the LC and serotoninergic fibers from the nuclei raphe primarily provide input into thalamic sections (limbic/affective, associative/cognitive, and sensory/motor) of the CSTC loop. These specialized thalamic areas are the origin of glutamatergic fibers projecting to the cortical SAN, CEN, DMN, and SMN hubs. In conclusion, even in circumstances where primary monoamine dysfunction is not a key etiologic factor in mood disorders, their balance, and influence on local glutamate and GABA circuits still have an important role in managing baseline activity of major functional networks and the dynamic of their adaptive coupling and uncoupling, and can therefore be a relevant target for pharmacotherapeutic interventions in mood disorders.

These "fine-tuning" modulatory activities of monoamine neurotransmitters are consistent with multiple findings suggesting that altered monoamine signaling reflects a potential compensatory response rather than a primary pathophysiologic basis of mood disorders. For example, several studies have reported elevated levels of 5-HT and NE metabolites in untreated MDD patients, which were "normalized" after successful antidepressant treatment. This notion is further supported by the fact that monoamine-enhancing antidepressant interventions produce a delayed clinical response (measured in weeks), despite the rapid elevation of brain monoamines (measured in hours), directly registered by in vivo microdialysis in animal models. In summary, the monoaminergic systems are thus now viewed primarily as broader neuromodulatory systems, and disturbances are as likely to be secondary or epiphenomenal effects as they are directly or causally related to etiology and pathogenesis.

NORADRENERGIC SYSTEMS

The cell bodies of almost all noradrenergic neurons in the brain are located in the LC of the brain stem and project rostrally to the thalamus, hypothalamus, ventral striatum, limbic system, and cerebral cortex (see Fig. 13.8–6). There are significant gender-related differences, as females relative to males have approximately 20% more NE fibers emanating from the LC on each side of the brain, as well as much richer axonal arborization at the termination points for these projections. Noradrenergic projections from the LC have a modular organization, subserving HPA/endocrine regulation, emotional modulation, cognition, and sensory and motor functions. This diffuse distribution belies NE's role in initiating and maintaining limbic and cortical arousal, as well as modulation of other neuronal systems. Noradrenergic projections to the amygdala and hippocampus are implicated in emotional memory, and behavioral sensitization to stress and prolonged activation of the LC contributes to the state of learned helplessness. The medial forebrain bundle (MFB) is the key ascending NE pathway to anterior cortical structures. Stimulation of the MFB elicits increased levels of goal-directed and reward-seeking behavior. Sustained stress eventually results in decreased MFB neurotransmission, which may account for anergia, anhedonia, and diminished libido in depression. The LC also is the origin of neurons that project rostrally to interact with the cell bodies of the sympathetic nervous system, and hence, the adrenal medulla. The LC is phasically active (see Fig. 13.8–7). Increased activity is provoked by the perception of novel stimuli; decreased activity accompanies eating and sleeping. As noted previously, there is an additive interaction (i.e., positive feedback loop) between noradrenergic neurons in the LC and CRH from the hypothalamus, as well as cortical regions. Other negative feedback loops must exert inhibitory control over this alerting synergism to prevent sustained pathologic activation. For example, increased noradrenergic output stimulates α_2 heteroreceptors on colocalized inhibitory 5-HT neurons, and a sustained increase in CRH drives results in a decreased hypothalamic release of ACTH which blunts HPA activity. Changes in cognitive processing of sensory input can further dampen corticolimbic and sympathoadrenal responses to sustained threatening or painful stimuli.

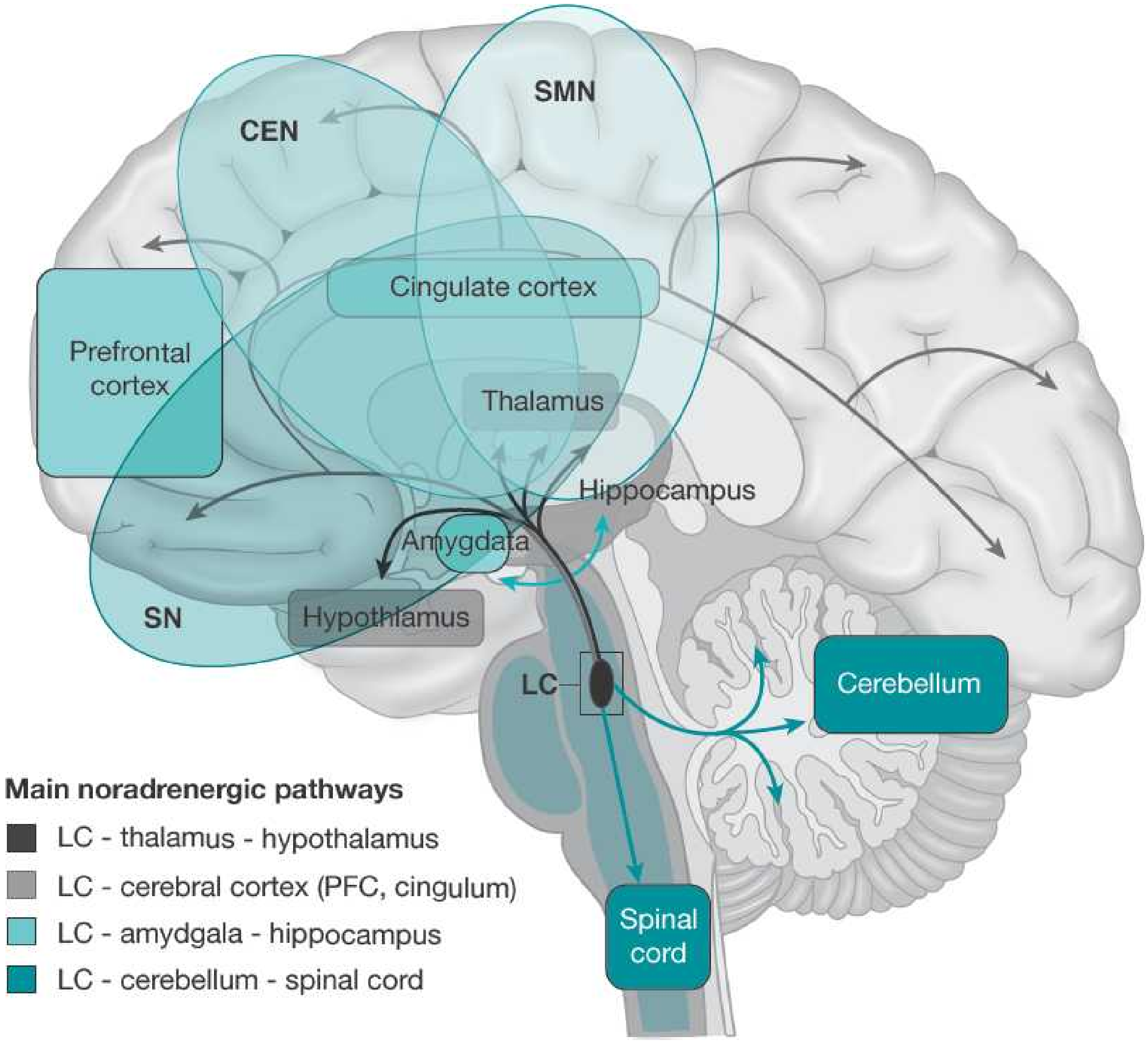
SMN
CEN
Cingulate cortex
Prefrontal cortex
Thalamus
Hippocampus
Amygdala
SN
Hypothalamus
LC
Cerebellum
Spinal cord
Main noradrenergic pathways
LC - thalamus - hypothalamus
LC - cerebral cortex (PFC, cingulum)
LC - amydgala - hippocampus
LC - cerebellum - spinal cord

FIGURE 13.8–6. Noradrenergic projections. Noradrenergic cells are mainly concentrated in the pontine nucleus *locus coeruleus* (LC) that projects diffusely to the entire cortex (mostly prefrontal cortex and cingulum) and the limbic system (mainly hippocampus and amygdala), thalamus/hypothalamus, cerebellum, and spinal cord. (From Carandini T, Cercignani M, Galimberti D, Scarpini E, Bozzali M. The distinct roles of monoamines in multiple sclerosis: a bridge between the immune and nervous systems? *Brain Behav Immun*. 2021;94:381–391.)

Recent functional imaging studies have reported a high degree of functional connectedness between noradrenergic-LC and hubs of the CEN, sensory-motor network (SMN), and SAN. In contrast to its presumed regulatory role in CEN and SMN function and connectedness, LC/NE seems to have a rather limited role in influencing DMN function. Given the primary internal focus of the DMN, this finding supports conclusions from the literature proposing a primary role of NE in “externally directed processes.” Given extant literature implicating disruption in adaptive shifting from internal self-preoccupation to externally focused adaptive coping in MDD, NE/glutamate dysfunction is a likely candidate.

Although past literature has predominantly qualified MDD as a “low norepinephrine” condition, accumulating evidence points in the opposite direction. A study that measured CSF concentration of MHPG in patients suffering from treatment-refractory mood disorders discovered a significant correlation between the “accumulated mood burden,” a composite measure based on the number and severity of mood episodes, and increased NE turnover, reflected by MHPG CSF concentration. Another study found an association between saliva MHPG levels and intensity of depression measured by a standardized scale. Moreover, several studies report “normalization” of elevated CSF and plasma NE levels in patients suffering from severe melancholic depression following successful ECT and antidepressant treatment. A hyperaroused subgroup of MDD patients identified by an automated EEG analysis, exhibited perhaps surprisingly, prominent concentration difficulties and fatigue. Other studies have found increased LC, ascending reticular activating system (ARAS), amygdala, and ACC activity in MDD patients suffering from insomnia (ACC and amygdala are hubs of SAN).

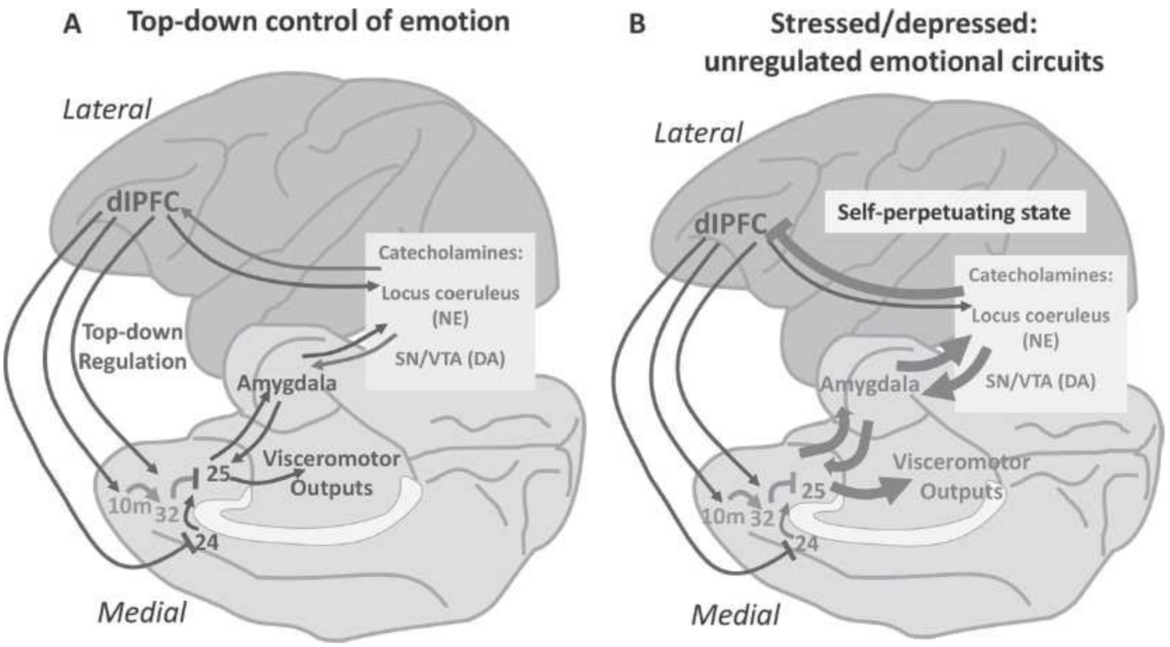
A
Top-down control of emotion
Lateral
dlPFC
Top-down Regulation
Catecholamines:
Locus coeruleus (NE)
SN/VTA (DA)
Amygdala
Visceromotor Outputs
10m
32
25
24
Medial
B
Stressed/depressed: unregulated emotional circuits
Lateral
dlPFC
Self-perpetuating state
Catecholamines:
Locus coeruleus (NE)
SN/VTA (DA)
Amygdala
Visceromotor Outputs
10m
32
25
24
Medial

FIGURE 13.8–7. Reflective to reactive: Catecholamine regulation of prefrontal cortex, limbic and paralimbic areas. Anterior cingulate cortex BA24; BA25 (subgenual cingulate); BA10, frontopolar or dorsomedial PFC; dlPFC, dorsolateral PFC; BA32, dorsal anterior cingulate; SNc, substantia nigra pars compacta, supplies striatum with dopamine. (Figure reprinted from Yang S, Seo H, Wang M, Arnsten AFT. NMDAR neurotransmission needed for persistent neuronal firing: potential roles in mental disorders. *Front Psychiatry.* 2021;12:654322. Copyright © 2021 Yang, Seo, Wang and Arnsten. doi: 10.3389/fpsyt.2021.654322. License: https://creativecommons.org/licenses/by/4.0/)

How does one reconcile these apparently contradictory findings? Persistently elevated LC activity/hyperarousal has been shown to disrupt the function of PFC areas associated with CEN and motor cortical areas associated with SMN. Therefore, disrupted regulation of the LC/arousal system in MDD, which would generate excessive NE signaling, is likely to produce disturbance of cognitive/executive function, including “top-down” regulation of emotions and sensory/psychomotor modulation. This imbalance could underpin impaired concentration, executive function, working memory, and psychomotor retardation. Moreover, a high frequency of LC firing promotes the release of a “cotransmitter” galanin, which acts as a functional antagonist of NE and DA transmission. Consistent with common manifestations of MDD, increased NE signaling in the amygdala and ventral portions of the ACC (SAN) has been linked with negative affect, anxiety, and diminished motivation and reward response. Therefore, it is plausible that in terms of arousal/LC/NE signaling, MDD patients may be a heterogeneous group. While some experience hyperarousal, it is also possible that some may be hypoaroused, while others have no disturbance in NE signaling. In addition, NE/LC transmission acts as a “master switch,” providing a smooth transition between the patterns of connectedness of major functional networks in order to successfully adapt to changes in the environment. It still remains to be ascertained if malfunction of the LC/arousal system may be responsible for the failure of MDD patients to pivot from their inwardly focused ruminative mode to an externally focused actively coping mode.

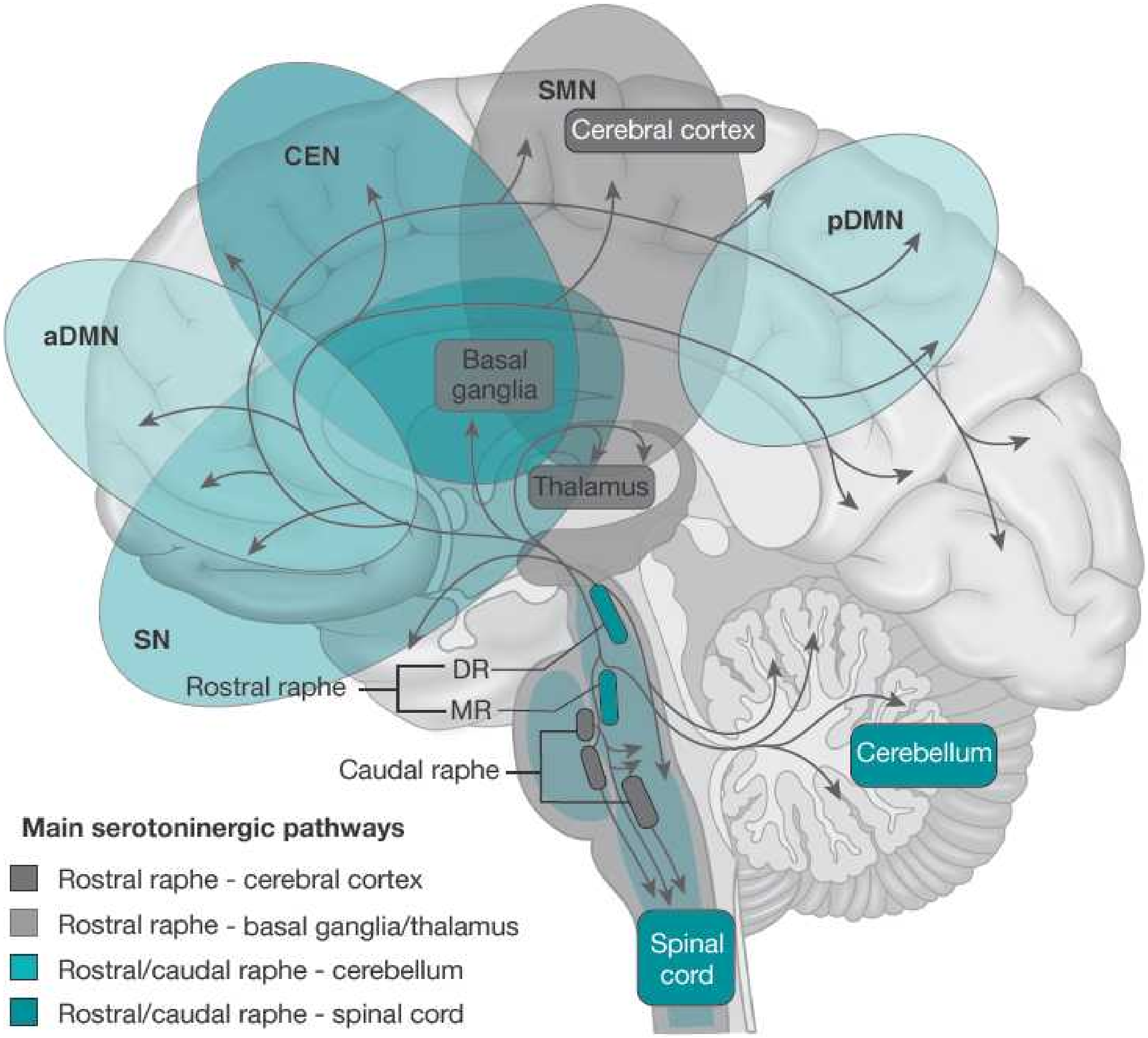
SMN
Cerebral cortex
CEN
pDMN
aDMN
Basal ganglia
Thalamus
SN
Rostral raphe
DR
MR
Caudal raphe
Cerebellum
Spinal cord
Main serotoninergic pathways
Rostral raphe - cerebral cortex
Rostral raphe - basal ganglia/thalamus
Rostral/caudal raphe - cerebellum
Rostral/caudal raphe - spinal cord

FIGURE 13.8–8. Main serotoninergic projections. Serotoninergic neurons are concentrated in the brainstem in two functional groups: (1) the caudal raphe nuclei (raphe magnus, raphe obscurus, and raphe pallidus) project to the lower brainstem and spinal cord; (2) the rostral raphe nuclei (median raphe [MR] and dorsal raphe [DR]) project to the entire neocortex, as well as to the thalamus and basal ganglia. The rostral raphe nuclei send a large contingent of fibers to the cerebellum and spinal cord. (From Carandini T, Cercignani M, Galimberti D, Scarpini E, Bozzali M. The distinct roles of monoamines in multiple sclerosis: a bridge between the immune and nervous systems? *Brain Behav Immun*. 2021;94:381–391.)

Aberrations in NE transmission have also been noted in bipolar disorder. Elevation of CSF-NE levels correlates with the manic state as well as associated mixed symptoms, such as anxiety and anger, typifying the state of “dysphoric mania.” Moreover, research has established a correlation between plasma levels of MHPG and the intensity of manic symptoms as well as their response to treatment.

SEROTONINERGIC SYSTEMS

Serotoninergic neurons project from the brain stem dorsal raphe nuclei to the cerebral cortex, hypothalamus, thalamus, basal ganglia, septum, and hippocampus (see Fig. 13.8–8). The 5-HT pathways have inhibitory (mediated by 5-HT1 and 5-HT5 receptors) and facilitatory functions (mediated by 5-HT3,4,6 and 7 receptors) in the brain. Both in cortical and subcortical areas, 5-HT neurons synapse with GABA and glutamate neurons, thereby regulating the balance between excitation and inhibition. Furthermore, the 5-HT system interfaces with acetylcholine (ACh), cortisol, prolactin, oxytocin, substance P, and vasopressin. Emphasizing its ubiquitous role, some authors point out that virtually all neurons and glial cells in the CNS express some type of 5-HT receptors. For example, much evidence suggests that 5-HT is an important regulator of sleep, appetite, body temperature, metabolism, energy homeostasis, stress response, anxiety, pain regulation, cognition, memory, reward processing, and libido. Moreover, 5-HT inhibits aggressive behavior across mammalian and reptilian species. Serotoninergic neurons projecting to the suprachiasmatic nucleus (SCN) of the hypothalamus help regulate circadian rhythms (e.g., sleep–wake cycles, body temperature, and HPA-axis function). The 5-HT also permits or facilitates goal-directed motor and appetitive behaviors in conjunction with NE and DA.

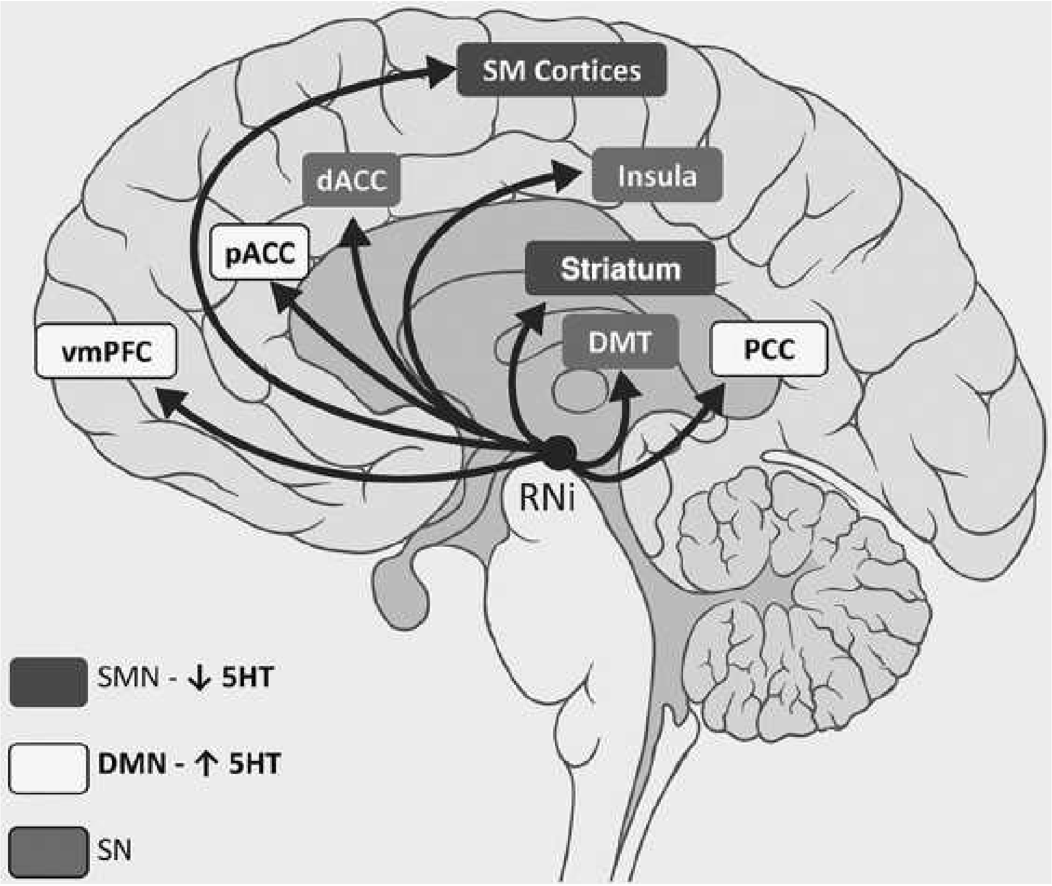

FIGURE 13.8–9. The role of serotonin in regulation of functional networks. DMT, dorsomedial thalamus; pACC, perigenual anterior cingulate cortex; PCC, posterior cingulate cortex; vmPFC, ventromedial prefrontal cortex; SM, sensorimotor; RNi, raphe nuclei; dACC, dorsal anterior cingulate cortex; SMN, sensorimotor network; DMN, default-mode network; SN, salience network. (Modified from Conio B, Martino M, Magioncalda P, et al. Opposite effects of dopamine and serotonin on resting-state networks: Review and implications for psychiatric disorders. *Mol Psychiatry.* 2020;25:82–93. doi:10.1038/s41380-019-0406-4.)

As noted previously, there is some evidence that 5-HT neurotransmission is partly under genetic control. The 5-HT “tone” also shows a seasonal fluctuation and shows a characteristic temporal response to stress. Specifically, acute stress increases 5-HT release transiently, whereas chronic stress eventually causes decreased 5-HT activity and depletion of 5-HT stores. Chronic stress may also increase the synthesis of 5-HT_{1A} autoreceptors in the dorsal raphé nucleus, which further decreases 5-HT transmission. Elevated glucocorticoid levels tend to enhance 5-HT functioning and thus may have significant compensatory effects in chronic stress. In addition, 5-HT has an important role during neurodevelopment when it participates in synapse formation, regulation of adhesion molecules, and neural network construction and connectivity.

Generally speaking, 5-HT and DA appear to have opposing roles in the regulation of major functional networks in healthy subjects. Serotonin tends to attenuate activity in the SMN network while catecholamines elevate it (see Fig. 13.8–9). This is in line with reports that 5-HT-enhancing psychotropics cause a sense of fatigue and blunted sensory experience, while catecholamine-boosting agents (e.g., stimulants) may produce a state of agitation and increased sensory acuity. Diminished DA transmission from the SN, along with lower 5-HT output from nuclei raphe, may underpin psychomotor retardation in mood disorders, while excessive DA output from SN and inadequate 5-HT transmission may be responsible for psychomotor agitation, observed both in depressive and manic states (see Fig. 13.8–10). In terms of 5-HT modulatory influence on the DMN network, the situation appears much more complex.

In healthy individuals, suppression of 5-HT transmission induced by acute tryptophan depletion (ATD) paradigm compromised the function of several DMN hubs. This change may be associated with the onset of depressed mood, anger/hostility, and diminished vigor. It is plausible that disturbance of 5-HT transmission in MDD may be related to a reduction of functional connectivity and perfusion within DMN, as well as aberrant DMN connectedness with other limbic and paralimbic formations. In line with this hypothesis, intravenous citalopram (an SSRI) “reintegrated” the precuneus with other DMN components in MDD patients while diminishing its connectivity with the amygdala (an SN component). Other studies have noted that aberrant connections between DMN and sgACC in MDD may be associated with ruminative tendencies and disease duration. Failure of DMN areas to “decouple” from CEN areas during cognitive tasks may be the basis for impaired working memory in MDD, which often manifests clinically as self-focused thoughts that interfere with externally oriented problem-solving. In conclusion, 5-HT elevation due to SSRI treatment may attenuate connectedness between core DMN areas in MDD (supporting more accurate self-assessment and adaptive future planning).

An increase in 5-HT activity may reduce aberrant connections between DMN and SAN and CEN brain structures, which might help to explain antidepressant effects on depressive symptomatology. Surprisingly, 5-HT influence on DMN may not be unidirectional, and recent dynamic imaging studies suggest that core DMN areas exert top-down regulation of 5-HT and DA release. In apparent contradistinction to SSRI effects, psychedelic drugs with putative antidepressant effects, such as potent 5-HT2A agonists LSD and psilocybin, have been noted to produce acute disruptions in DMN connectivity. Disruption of typical DMN connectedness has been associated with "self-dissolution" and altered social perception, commonly associated with psychedelic use. It is difficult to ascertain how much of the psychedelic effect is directly attributable to 5-HT2A agonism in DMN because these agents also disrupt the activity of SAN.

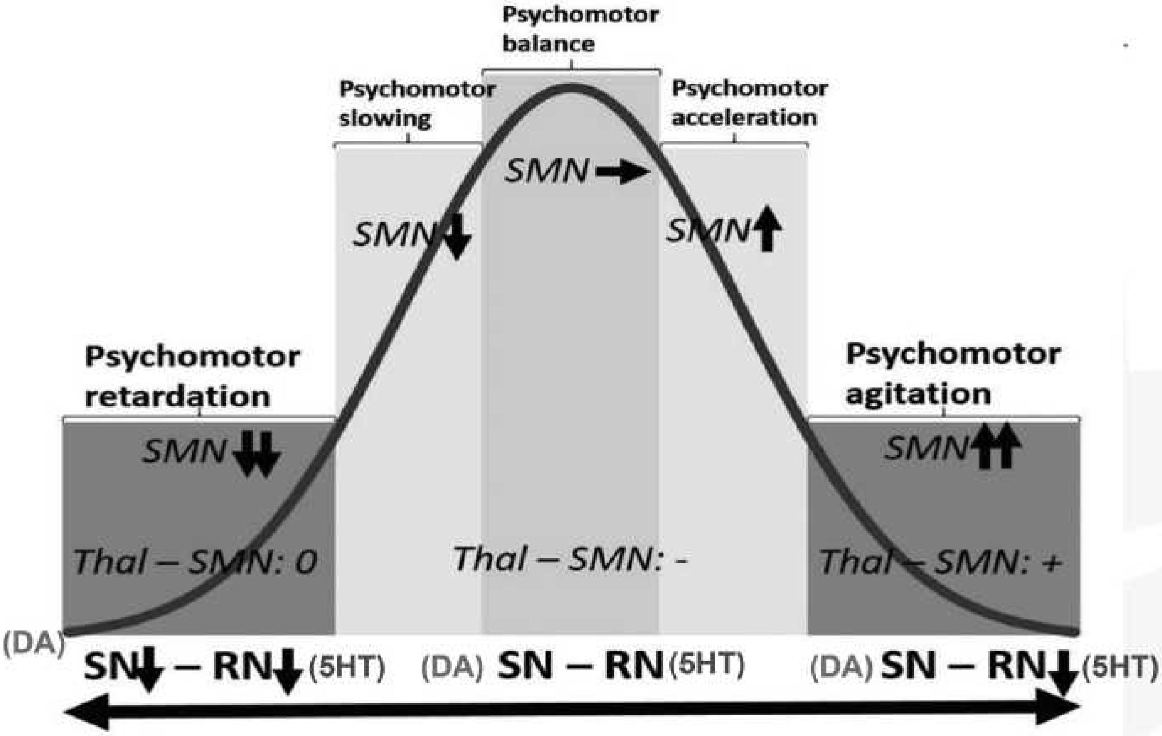
Psychomotor balance
Psychomotor slowing
Psychomotor acceleration
SMN →
SMN ↓
SMN ↑
Psychomotor retardation
Psychomotor agitation
SMN ↓↓
SMN ↑↑
Thal – SMN: 0
Thal – SMN: -
Thal – SMN: +
(DA) SN↓ – RN↓ (5HT)
(DA) SN – RN (5HT)
(DA) SN – RN↓ (5HT)

FIGURE 13.8–10. Psychomotor balance, functional networks and monoamine neurotransmitters. SN, substantia nigra (DA); RN, raphe nucleus (5-HT); BG, basal ganglia, Thal thalamus; SMN, sensorimotor network; DMN, default-mode network; GA, global activity. (From Northoff G, Hirjak D, Wolf RC, et al. All roads lead to the motor cortex: Psychomotor mechanisms and their biochemical modulation in psychiatric disorders. *Mol Psychiatry.* 2021;26(1):92–102. doi:10.1038/s41380-020-0814-5.)

Serotonin influence on the SMN is more straightforward. Increased 5-HT transmission from the nuclei raphe has been consistently associated with suppressed activity in the SMN, consistent with the high expression of inhibitory 5-HT1A receptors in the sensory and motor cortical areas. Serotoninergic regulation of the SMN is of utmost clinical importance since compromised SMN resting and dynamic connectedness appears to be one of the central features of MDD, most likely underpinning psychomotor retardation. SSRI treatment of MDD may be a mixed blessing because it is often associated with complaints of fatigue and blunted sensory experience, phenomena consistent with down-regulated SMN function. Moreover, meta-analyses have revealed that SNRIs and a DNRI agent (bupropion) provide greater relief of psychomotor symptoms of MDD compared to SSRIs, which are either minimally effective or a hindrance.

Functional imaging studies have consistently implicated aberrant activation of SAN components, such as dACC, insula, dorsomedial thalamus, nucleus accumbens, and amygdala (some authors also include VTA and ventrolateral PFC), in the pathophysiology of depression. Functional connectivity between serotoninergic nuclei raphe and SAN hubs, insula, mid-ACC, and amygdala, has been reported to be altered in MDD patients compared to controls. Successful SSRI treatment tends to “normalize” amygdala activity. As a reminder, the degree of amygdala activity and its connectedness with the sgACC has been associated with the duration and intensity of negative affect in MDD. Greater amygdala, insula, and dorsal ACC response to emotional stimuli in MDD patients are noted to be a positive predictor of SSRI response. Administration of an SSRI was associated with greater suppression of nucleus accumbens response to an erotic stimulus in healthy individuals relative to placebo and bupropion. This finding may explain some aspects of the SSRI-induced sexual dysfunction. Moreover, SSRI administration in preclinical studies has been associated with the suppression of VTA dopaminergic neurons, hypothetically mediated by 5-HT2C receptors. Extending these findings, the use of the antidepressant agomelatine, a melatonin receptor agonist and 5-HT2C antagonist, resulted in the rapid improvement of anhedonia in MDD patients, measured by a standardized scale.

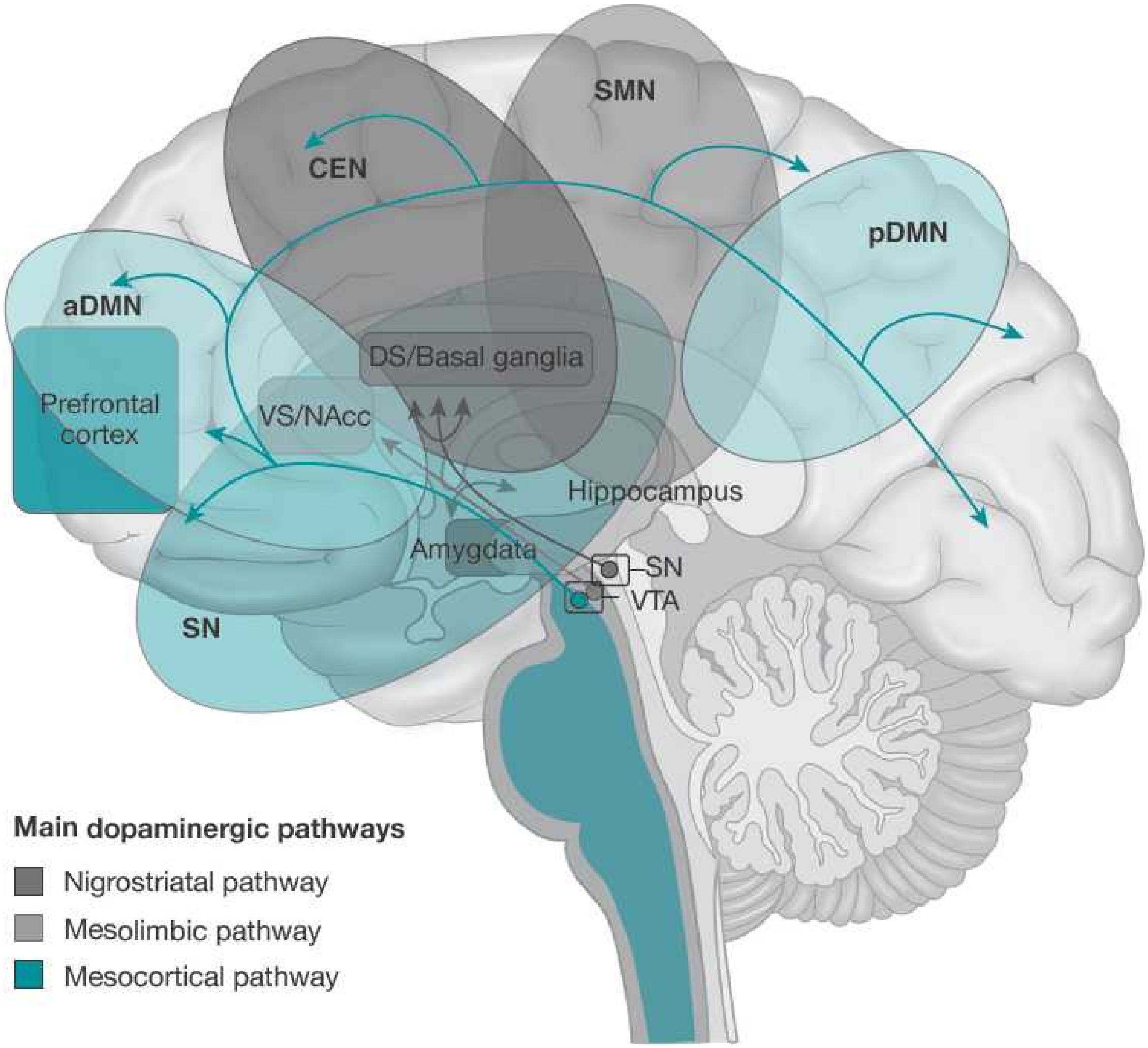
SMN
CEN
pDMN
aDMN
DS/Basal ganglia
Prefrontal cortex
VS/NAcc
Hippocampus
Amygdata
SN
VTA
SN
Main dopaminergic pathways
Nigrostriatal pathway
Mesolimbic pathway
Mesocortical pathway

FIGURE 13.8–11. Main dopaminergic projections. (1) the *substantia nigra* (SN) that projects to the dorsal striatum (DS) via the nigrostriatal pathway, which is part of the basal ganglia loop; (2) the ventral tegmental area (VTA) that is the origin of the mesocorticolimbic pathway. The mesolimbic portion projects from the VTA to the ventral striatum (VS)/nucleus accumbens (Nacc) and limbic system (mainly hippocampus and amygdala). The mesocortical part connects the VTA with the prefrontal cortex and the entire neocortex. (From Carandini T, Cercignani M, Galimberti D, Scarpini E, Bozzali M. The distinct roles of monoamines in multiple sclerosis: a bridge between the immune and nervous systems? *Brain Behav Immun*. 2021;94:381–391.)

Finally, diminished connectivity between the serotoninergic nuclei raphe and basal ganglia/thalamus (BG/THAL), with a resulting decrease in DMN activity, has been associated with bipolar mania. Compromised DMN activity also may underpin decreased self-monitoring in mania.

DOPAMINERGIC SYSTEMS

There are four relatively discrete DA pathways in the brain (see Fig. 13.8–11). The tuberoinfundibular system projects from cell bodies in the hypothalamus to the pituitary stalk, exerting inhibitory control over prolactin secretion. The nigrostriatal system originates from cell bodies in the SN and projects to the basal ganglia, regulating involuntary motor activity. Recent studies have also reported sparse projections from SN to lateral aspects of PFC. The cell bodies of the mesolimbic pathway are located in the ventral tegmentum and project diffusely to the nucleus accumbens, amygdala, hippocampus, medial dorsal nucleus of the thalamus, and cingulate gyrus. This pathway modulates emotional expression, learning and reinforcement, and hedonic capacity. The fourth DA pathway, also originating in the ventral tegmentum, is the mesocortical pathway, which projects to the orbitofrontal and the prefrontal cortical regions and helps to regulate motivation, concentration, and initiation of goal-directed and complex executive cognitive tasks. Decreased mesocortical and mesolimbic DA activity has obvious implications for the cognitive, motor, and hedonic disturbances that are associated with depression and are a target for action for some investigational compounds. Across pathways, increased DA activity is potentiated by nicotinic inputs and glucocorticoids, and DA levels are inversely correlated with indices of brain 5-HT activity. The nigrostriatal pathway innervates important striatal relay structures, which are components of the corticostriatal–thalamo–cortical (CSTC) loop. While the VTA provides dopaminergic input for the limbic/affective division of the striatum, the cognitive/associative and sensory-motor striatum are innervated by SN. Excessive DA release in the associative striatum has been linked with psychotic symptoms in the context of bipolar mania.

Dopaminergic signaling and its balance with other monoamines is a key determinant of the activity of major functional networks, and by extension, regulation of affect, motor behavior, and cognition. One of the indicators of CNS DA signaling is the CSF concentration of its principal metabolite, homovanillic acid (HVA). In a recent study, severely depressed patients had significantly reduced CSF concentration of HVA compared to controls. Furthermore, depression scores were inversely correlated with CSF-HVA levels in depressed patients, suggesting that severe MDD may in part be a hypodopaminergic state.

In contrast, increased dopaminergic input into BG/THAL CSTC relay structures is associated with increased activation of the (SMN), which most likely underpins psychomotor agitation in the context of bipolar mania. It is plausible that increased activity of the SMN, combined with inadequate DMN-based self-monitoring, maybe the neurobiologic basis for transdiagnostic manifestations of psychomotor agitation. In support of this theory, functional imaging studies have reported increased connectivity between the dopaminergic VTA and the motor cortex in MDD patients. On the other hand, attenuated DA and 5-HT signaling in the BG/THAL may translate into decreased SMN activation and ensuing psychomotor retardation, a prominent feature of depression. In aggregate, data support what we observe clinically; a major depressive episode may have two separate biotypes with matching phenotypes: agitated depression and depression with psychomotor retardation.

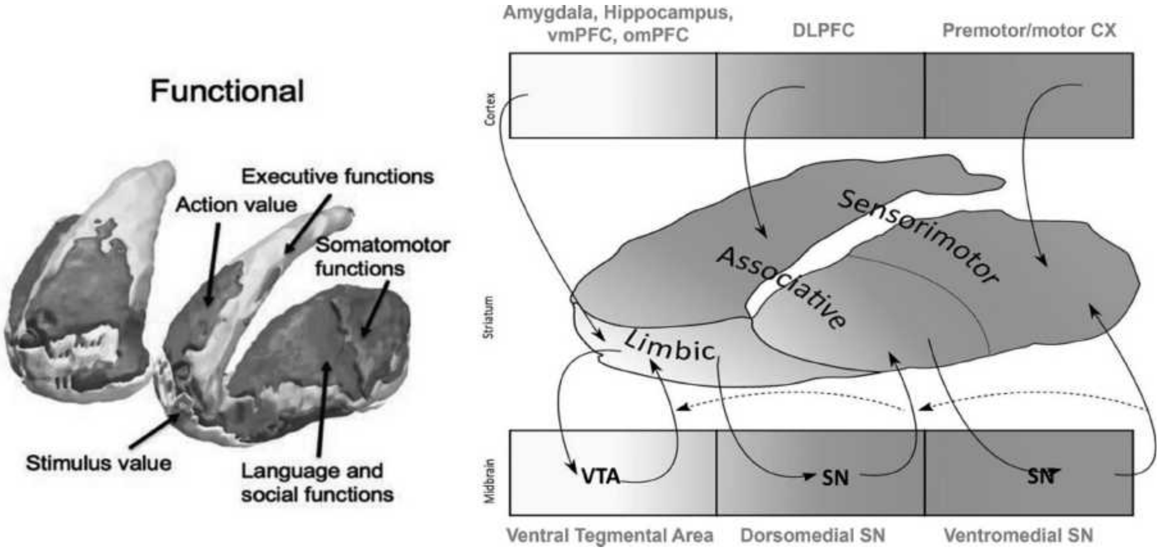
Functional
Executive functions
Action value
Somatomotor functions
Stimulus value
Language and social functions
Amygdala, Hippocampus, vmPFC, omPFC
DLPFC
Premotor/motor CX
Cortex
Striatum
Midbrain
Sensorimotor
Associative
Limbic
VTA
SN
SN
Ventral Tegmental Area
Dorsomedial SN
Ventromedial SN

FIGURE 13.8–12. Dopamine transmission and functional organization of the striatum. (See eBook for color images.) (Reprinted from McCutcheon RA, Abi-Dargham A, Howes OD. Schizophrenia, dopamine and striatum: From biology to symptoms. *Trends Neurosci.* 2019;42(3):205–220; Basile GA, Bertino S, Bramanti A, et al. Striatal topographical organization: bridging the gap between molecules, connectivity and behavior. *Eur J Histochem.* 2021;5(s1):3284. doi:10.4081/ejh.2021.3284. License: https://creativecommons.org/licenses/by/4.0/.)

As a general rule, DA transmission, predominantly mediated by D1 receptors in the hubs of the CEN, has an inverted “U” shaped relationship with alertness, sustained attention, working memory, and executive function. Both inadequate and excessive D1 signaling may result in distractibility, inattentiveness, and impaired working memory, as well as impairment of the executive function. Excessive stress and MDD may act synergistically via dysregulated catecholamine input into the DLPFC to compromise CEN-related top-down regulation of emotion (see Fig. 13.8–12). Indeed, increased DA signaling in the amygdala (a key SAN area) has been associated with an increased response to negative emotional stimuli and excessive preoccupation with internal bodily processes in MDD. It is possible that increased stress contributes to the elevation of negative emotionality in MDD, both by increasing catecholamine activity in limbic and paralimbic SAN regions and by reducing the DLPFC-mediated top-down regulation. Indeed, functional imaging studies have observed increased connectedness between the VTA and insula (key SAN hub) in MDD patients. Such increased connectedness between the insula, a focal target of all interoceptive inputs, and the VTA may be responsible for the over-attribution of emotional salience to internal stimuli in MDD patients. Indeed, imaging findings support the over-activation of brain areas related to interoception and relatively attenuated brain responses to exteroceptive stimulation. On the other hand, lower functional connectivity between the dopaminergic VTA and the dACC (a SAN hub) is reflected in more pronounced somatic symptoms of MDD, such as appetite disturbance, fatigue, and loss of sexual interest. Furthermore, multiple studies have reported decreased response of the nucleus accumbens (another major SAN node) to positive and rewarding stimuli in depressed individuals relative to controls. As a matter of fact, compromised NAcc response to positive stimuli has been deemed a predictor of response to bupropion (a DA/NE modulating antidepressant). Moreover, contemporary imaging studies have provided new insight into the relationship between DLPFC and the regulation of mesolimbic DA transmission stemming from VTA. Not only does VTA via mesocortical pathway provide innervation for DLPFC, a CEN hub, but in a retrograde fashion, the DLPFC via polysynaptic connections also modulates dopaminergic influx into the nucleus accumbens, possibly explaining how our interest can drive motivated behaviors. In addition, decreased DA signaling in the ventral striatum/ventral pallidum via D2/D3 receptors has been noted to correlate with higher anxiety in MDD patients.

Finally, the elevation of striatal DA signaling has been consistently associated with suppression of the DMN and an increase in SAN connectedness in healthy individuals (see Fig. 13.8–13). Dysregulated DA signaling in BG/THAL structures of the CSTC loop with resulting increase in glutamatergic flow from the thalamus to the CEN and SMN regions, along with reduced serotoninergic input in BG/THAL and diminished activation of the DMN network, has been associated with mania. The above-described pattern of major brain network activity/connectivity may underpin clinical manifestations of mania, such as inattentiveness, impaired executive function, problematic decision-making and working memory (all CEN-related), and affective reactivity due to compromised top-down emotional regulation. Pathologically enhanced SMN activation may be associated with hyperactive behavior and increased sensory acuity, while compromised DMN function may be clinically reflected in inadequate self-monitoring, planning, and maladaptive processing of social cues, all common clinical features of the manic state.

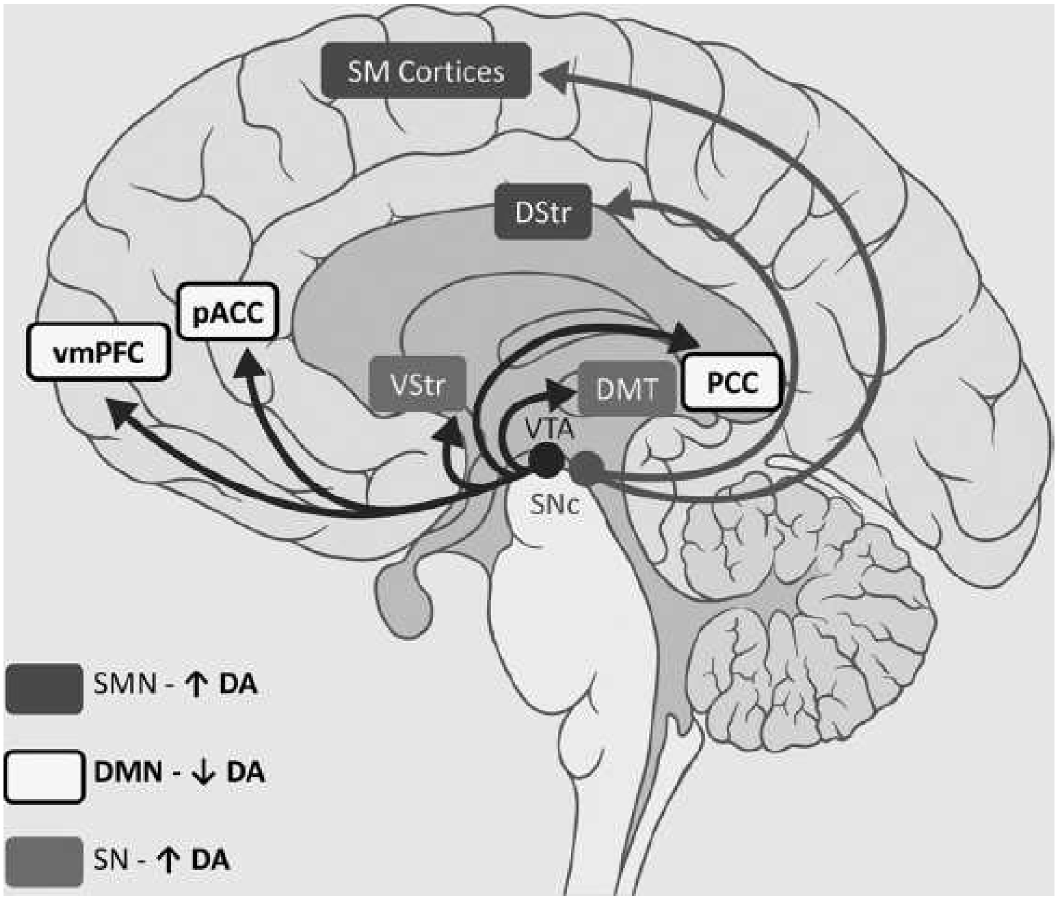

FIGURE 13.8–13. The role of dopamine in regulation of functional networks. SNc, substantia nigra pars compacta; VTA, ventral tegmental area; DStr, dorsal striatum; VStr, ventral striatum; DMT, dorsomedial thalamus; pACC, perigenual anterior cingulate cortex; PCC, posterior cingulate cortex; vmPFC, ventromedial prefrontal cortex; SM, sensorimotor; RNi, raphe nuclei; dACC, dorsal anterior cingulate cortex; RSNs, resting state networks; SMN, sensorimotor network; DMN, default-mode network; SAN, salience network. (Modified from Conio B, Martino M, Magioncalda P, et al. Opposite effects of dopamine and serotonin on resting-state networks: Review and implications for psychiatric disorders. *Mol Psychiatry.* 2020;25:82–93. doi:10.1038/s41380-019-0406-4.)

SUMMARY OF RESEARCH ON MONOAMINES IN DEPRESSION

After nearly 50 years of research, it can be said that some depressed people manifest one or more abnormalities of monoamine neurotransmission. Decreased central NE activity can be inferred, in part, from decreased urinary excretion of the metabolite MHPG, although a partly overlapping subgroup of patients manifests elevated circulating levels of NE and its metabolites. This suggests a dissociation of NE activity in the MFB and SM systems. As reviewed in earlier editions of this chapter, 5-HT dysfunction has been documented in overlapping subgroups of patients using a variety of methods, ranging from low CSF levels of 5-HIAA to decreased regional cerebral metabolism in response to 5-HT agonists, such as D-fenfluramine, and as discussed subsequently, decreased 5-HT neurotransmission also can be inferred from the findings of HPA and EEG sleep studies. Low levels of the DA metabolite HVA have been implicated in psychomotor retardation. Some evidence suggests that increased DA activity may contribute to the development of delusions and hallucinations.

Burgeoning evidence from biochemical, neuroimmunologic, endocrine, and functional imaging studies implicates either primary or secondary malfunction of monoamine signaling in the pathophysiology of mood disorders. Moreover, monoamine projections have an established role in regulating the connectedness and function of major functional networks involved in the regulation of mood, cognition, motoric activity, and sensory processing (SAN, SMN, DMN, and CEN). Furthermore, dysfunction of major functional networks closely correlates with symptoms of mood disorders, offering these theories good face validity. In addition, monoamine signaling has a prominent role in regulating the “updating” function of the CSTC loop. One can conceive of harmonious and adaptive monoamine signaling as the basis for maintenance of dynamic homeostasis (allostasis) involving major brain functional networks and their dysfunction as a principal contributor to a disequilibrium underlying the “allostatic overload” in mood disorders. Finally, decades of research have established an important role of agents modulating monoamine metabolism, reuptake, and receptor activity in the treatment of mood disorders.

OTHER NEUROTRANSMITTER DISTURBANCES

ACh is found in neurons that are distributed diffusely throughout the cerebral cortex. Cholinergic neurons have reciprocal or interactive relationships with all three monoamine systems. Abnormal levels of choline, which is a precursor to ACh, have been found at autopsy in the brains of some depressed patients, perhaps reflecting abnormalities in cell phospholipid composition. In an animal model of depression, strains of mice that are super- or subsensitive to cholinergic agonists have been found to be prone to or more resistant to developing learned helplessness. Cholinergic agonists can induce changes in HPA activity and sleep that mimic those associated with severe depression. Indeed, some remitted patients with mood disorders, as well as their never-ill first-degree relatives, have a trait-like increase in sensitivity to cholinergic agonists. Decades ago, a theory was put forth whereby a high cholinergic/adrenergic ratio represents a risk for depression, whereas a low ratio may be linked with the manic state. Cholinergic agonist and antagonist drugs have differential clinical effects on depression and mania. Agonists can produce lethargy, anergia, and psychomotor retardation in normal subjects, can exacerbate symptoms of depression, and can reduce symptoms of mania. These effects generally are not sufficiently robust to have clinical applications, although some recent research has suggested that parenteral scopolamine can have rapid antidepressant effects. Presumably, scopolamine blocks M1 muscarinic receptors on GABA interneurons, thereby decreasing their signaling and diminishing the ensuing inhibition of glutamatergic neurons. The resulting increase in the glutamatergic transmission is putatively associated with the observed rapid antidepressant response. More conventional antidepressants that target serotonergic or adrenergic neurons also indirectly decrease cholinergic tone.

GABA interneurons have an inhibitory effect on ascending monoamine pathways, particularly the mesocortical and mesolimbic systems. Furthermore, mPFC GABA interneurons modulate the signal-to-noise ratio of the incoming transmission and the output volume of cortical pyramidal glutamatergic neurons. While the signal-to-noise ratio of incoming signals is primarily regulated by somatostatin-expressing (SST) GABA interneurons, the glutamatergic output is modulated by parvalbumin-expressing (PV) GABA interneurons. Long glutamatergic projections from the mPFC to brainstem nuclei have a cardinal role in the modulation of ascending monoamine fibers. Reductions in GABA have been observed in plasma, CSF, and brain GABA levels in depression. Animal studies have also found that chronic stress can reduce and eventually deplete GABA levels. By contrast, GABA receptors are upregulated by antidepressants, and some GABAergic medications have weak antidepressant effects. The most compelling evidence of GABA involvement in the pathophysiology of

MDD comes from ENIGMA consortium studies which integrated mega-GWAS findings with functional and structural imaging studies, postmortem cytologic findings, and clinical data from tens of thousands of MDD patients. Two cell types that were most closely associated with MDD etiopathogenesis were astrocytes and cortical SST GABA interneurons. Downstream effects of PFC SST GABA interneuron dysfunction include diminished transmission from pyramidal neurons to the amygdala, thalamus, hypothalamus, BNST, and brainstem monoaminergic nuclei. Therefore, deficits in descending glutamatergic transmission, emotional regulation, endocrine control, arousal modulation, autonomic function, and monoamine signaling in mood disorders can all be potentially traced back to altered SST GABA interneuron density and function. Recent functional imaging studies have discovered that rapid GABA and glutamate circuits have a central role in “switching,” establishing, and shifting connections between major functional networks, such as SAN, CEN, and DMN. It is plausible that altered GABA signaling may have repercussions on the previously described disruption of the major functional network dynamics in mood disorders. Clinical studies have established a link between reduced size, density, and neurotransmitter synthesis in dlPFC GABA interneurons and MDD. In addition, the synthesis of allopregnanolone, a naturally occurring neurosteroid ligand of GABA-A receptors has been reported to be lower in DLPFC of depressed patients relative to controls. Moreover, other studies reported a correlation between ACC GABA levels and anhedonia scores in young patients suffering from MDD. In addition, lower GABA ACC levels are also associated with reduced hippocampal volumes in treatment-resistant MDD patients but not ones who have responded to treatment. Furthermore, inadequate GABA neuronal activity in the ACC combined with structural changes in the hippocampus may interfere with the HPA stress response and contribute to the endocrine manifestations of MDD. Moreover, altered GABA synthesis has also been reported in the amygdala of MDD patients compared to controls. This finding may have implications for increased amygdala response to negatively valenced stimuli, frequently observed in individuals suffering from MDD. Extending these findings, synthetic neurosteroids are being developed for the treatment of MDD. Unlike benzodiazepines which predominantly modulate the gamma-subunit containing synaptic GABA-A receptors involved in phasic signaling, neurosteroids also bind to extrasynaptic GABA-A receptors, containing the delta-subunit involved in tonic regulation of GABA signaling (see Fig. 13.8–14). Brexanolone, a synthetic neurosteroid that functions as a positive allosteric modulator of GABA-A receptors, is already approved for the treatment of major depressive episodes with peripartum onset. Another similar agent, zuranolone, is in the late stage of development for the treatment of MDD.

In the CNS, two amino acids—glutamate and glycine—are respectively the major excitatory and inhibitory neurotransmitters. Glutamate and glycine bind to sites associated with the *N*-methyl-D-aspartate (NMDA) receptor, and an excess of glutamatergic stimulation can cause neurotoxic effects. It appears that glutamate binding to extra-synaptic NMDA receptors, containing the N2B subunit, with glycine as a coagonist produces excitotoxicity and suppression of BDNF synthesis. In contrast, activation of synaptic NMDA receptors containing the N2A subunit, with d-serine as a coagonist, promotes BDNF synthesis and is neuroprotective. In addition to NMDA receptors, glutamate binds to two other ionotropic receptors: α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPA) and kainate receptors.

While ionotropic receptors rapidly mediate rapid glutamate transmission, eight subtypes of metabotropic glutamate receptors (mGlu) are involved in slower (i.e., seconds to minutes) signaling. Importantly, the hippocampus has a high concentration of NMDA receptors. Glutamate may thus work in tandem with hypercortisolemia to mediate the deleterious neurocognitive effects of severe recurrent depression. Aberrant glutamate transmission may be proximally implicated in atrophic changes documented in some regions of the cerebral cortex, perhaps by first reducing the connectivity of neurons and volume of supporting glia. Indeed, imaging studies have confirmed glutamate markers in hippocampi of chronically depressed patients compared to healthy controls; such changes are evident even among people suffering their first-lifetime depressive episode. Furthermore, another study has noted a reduction of glutamate in the vmPFC, a key area in emotional regulation of patients suffering from recurrent or chronic forms of depression but not in controls or first-episode depressed patients. Interestingly, there is also a significant correlation between the duration of depressive illness and the reduction of vmPFC glutamate concentrations. Glutamate levels in ACC also are lower in both unipolar and bipolar depression compared to controls. Altered glutamate transmission was also noted in the basal ganglia of MDD patients relative to controls. Elevated glutamate levels in this key component of the CSTC loop were associated with both anhedonia and psychomotor retardation in MDD patients. The potential significance of aberrant glutamate signaling via NMDA receptors in the pathophysiology of depressive states is suggested by growing evidence that subanesthetic doses of ketamine and esketamine, which noncompetitively antagonizes NMDA receptors, exert large and rapid antidepressant effects. Studies of continued therapy (e.g., weekly or every other week) with intranasal esketamine indicate that treatment benefits persist for months or even years. Both racemic ketamine and esketamine show a preference for N2B containing extrasynaptic NMDA receptors. Theoretically, antagonism at NMDA receptors located on GABA interneurons reduces their input, thereby promoting glutamatergic bursts from the postsynaptic neurons.

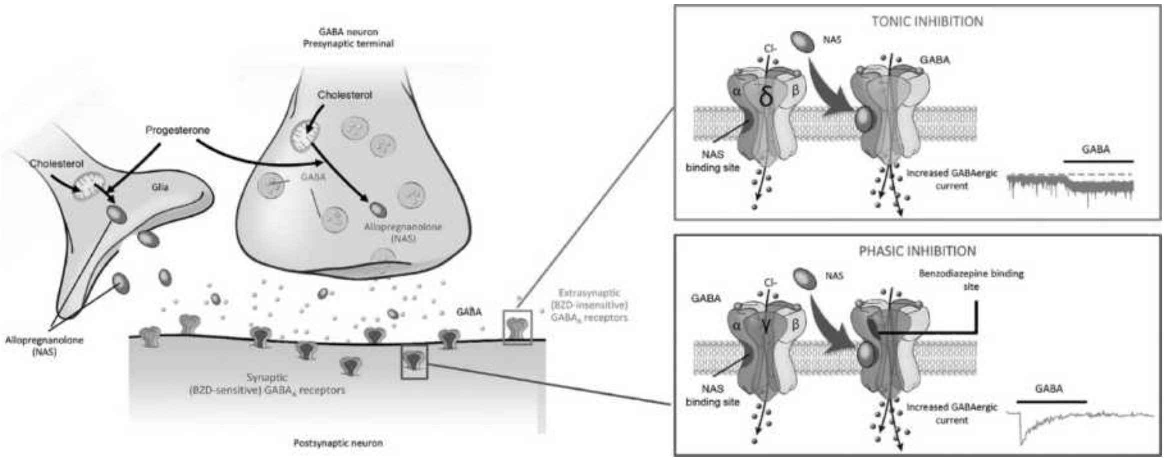
GABA neuron
Presynaptic terminal
Cholesterol
Progesterone
Cholesterol
Glia
GABA
Allopregnanolone
(NAS)
Allopregnanolone
(NAS)
GABA
Extrasynaptic
(BZD-insensitive)
GABA$_A$ receptors
Synaptic
(BZD-sensitive) GABA$_A$ receptors
Postsynaptic neuron
TONIC INHIBITION
Cl-
NAS
GABA
α
δ
β
NAS
binding site
Increased GABAergic
current
GABA
PHASIC INHIBITION
Cl-
NAS
Benzodiazepine binding
site
GABA
α
γ
β
NAS
binding site
Increased GABAergic
current
GABA

FIGURE 13.8–14. Relationship between GABA and neuroactive steroids. NAS binding to GABA$_A$ receptors resulting in potentiation of both synaptic and extrasynaptic GABA$_A$ receptor populations and leads to increased GABA$_A$ receptor surface expression. Extrasynaptic GABA$_A$ receptors containing delta subunits are insensitive to benzodiazepines.[1-8] HSD, hydroxysteroid dehydrogenase; StAR, steroidogenic acute regulatory; MBR, mitochondrial benzodiazepine receptor. (See eBook for color images.) (Data from Jacob TC, Moss SJ, Jurd R. GABA(A) receptor trafficking and its role in the dynamic modulation of neuronal inhibition. *Nature Rev Neurosci.* 2008;9:331–343; Reddy DS, Estes WA. Clinical potential of neurosteroids for CNS disorders. *Trends Pharmacol Sci.* 2016;37:543–561.)

Increased glutamatergic signaling via AMPA receptors has been associated with inhibition of eukaryotic elongation factor 2 (eEF2) and consequently enhanced synthesis of BDNF. Moreover, elevated BDNF signaling activates an intracellular cascade supporting mTOR (mammalian target of rapamycin) synthesis and a downstream increase in synaptogenesis. In contrast to intrasynaptic NMDA receptor activation, extrasynaptic NMDA activation is associated with mitochondrial dysfunction, suppression of BDNF synthesis, and ultimately, cell death. Therefore, extrasynaptic NMDA inhibitors such as esketamine and ketamine are likely to support BDNF synthesis via the reduction of extrasynaptic NMDA-mediated glutamate transmission (see Fig. 13.8–15).

Subanesthetic doses of ketamine and esketamine influence the activity of major functional networks both in healthy volunteers and depressed patients. An MRI study compared ketamine infusion with placebo in a group of healthy volunteers. Ketamine infusions produce a significant increase in dACC activity, suggesting positive modulation of a key SAN hub. However, a study comparing depressed patients and healthy controls found that ketamine infusions increased activity in the CEN and reduced activity in the DMN and SAN relative to the placebo group. The impact of ketamine on burst firing in the lateral habenula (LHb) has been a recent focus of interest. Burst firing of LHb has been associated with suppression of monoamine release and intense negative affect in MDD patients.

Convergent lines of evidence implicate changes in glia function and density in the pathohistology of mood disorders. Large-scale studies of mood disorders incorporating genetic data, postmortem cytology, and functional and structural imaging studies have found that transcriptional signatures related to astrocytes, microglia, pyramidal neurons, and SST-expressing GABA interneurons have the greatest correlation with a decreased thickness of cortical areas in mood disorders. Furthermore, postmortem cytology in MDD demonstrated a reduction of glial density and numbers in several brain areas involved in mood regulation and the stress response. A substantial decline in both numbers and density of astroglia and oligodendroglia has been reported in the sgACC, DLPFC, orbitofrontal cortex, and amygdala of unmedicated bipolar and MDD patients relative to controls. In addition, a recent transcriptional study of MDD patients discovered the aberrant expression of 17 genes related to oligodendroglia function, suggesting that reduced oligodendroglia numbers may be a consequence of genetic vulnerability. A 29% reduction of oligodendroglial density was reported in the DLPFC of bipolar patients compared to controls. Given the oligodendroglia’s role in myelination of long white matter tracts, these cytologic findings may underpin imaging reports of disrupted functional and structural connectivity between key prefrontal and limbic areas involved in emotional regulation.

Postmortem studies using immunohistochemistry evaluated microglia density in the DLPFC, ACC, mediodorsal thalamus, and hippocampus of mood-disordered patients. Significant microgliosis (i.e., increased number of microglia) was noted in affective disorder patients who committed suicide but not in depressed individuals who died due to other causes. Microgliosis in this study was interpreted as a marker of stress prior to suicide. Microglial pathology may interact with dysregulated peripheral inflammatory signaling in MDD and bipolar disorder. In response to excessive peripheral inflammatory signaling in mood disorders, a pathogen-associated molecular pattern (PAMP) occurs in the CNS. Specifically, activated microglia respond by releasing copious amounts of IL-1, IL-6, TNF-alpha, PGs, NO, and H_2O_2, which in turn prompts astrocytes to release more of the similar inflammatory molecules. Due to aberrant astroglia-microglia communication, oligodendroglia are likely to suffer oxidative damage related to overexposure to reactive oxygen and nitrogen species (ROS and RNS) and maladaptive activation of its NMDA receptors. Furthermore, TNF-alpha released by microglia and astrocytes, combined with reduced BDNF support, has a toxic effect on oligodendroglia. The sum of these aberrant activities likely contributes to demyelination and compromised communication within and between mood-modulatory networks. Moreover, aberrant microglia function in mood disorders may be related to inflammatory, neurotrophic, and neurotransmitter disturbance. Proinflammatory cytokines and prostaglandins jointly induce indoleamine 2,3-dioxygenase (IDO), an enzyme that converts tryptophan to kynurenine and quinolinic acid (QUIN), thereby reducing its availability for 5-HT and melatonin synthesis (see Fig. 13.8–16). In addition, proinflammatory cytokines accelerate 5-HT and DA reuptake from the synaptic cleft, further disrupting their role in neurotransmission. Moreover, QUIN is a robust NMDA agonist and stimulator of glutamate release, most likely contributing to excitotoxicity by binding to extrasynaptic NMDA receptors. In vitro studies have provided evidence that activation of microglial NMDA receptors drives the release of inflammatory cytokines. It is conceivable that there is a vicious circle present in mood disorders, whereby microglial activation induces increased extrasynaptic NMDA signaling, which in turn perpetuates microglial proinflammatory activity. Although the use of translocator protein (TSPO) as a marker of microglial activation is not without controversy, imaging studies (PET) have reported increased TSPO in the PCC, ACC, insula, and hippocampus of MDD patients compared to controls. Furthermore, the duration of untreated depression correlates with TSPO elevation in the PFC, ACC, and insula, while the duration of antidepressant treatment is counter-correlated with TSPO concentration in the same brain areas. Extending these findings, imaging studies have reported a correlation between PFC TSPO binding and cognitive dysfunction in MDD. Interestingly, ketamine may reduce microglial inflammatory signaling.

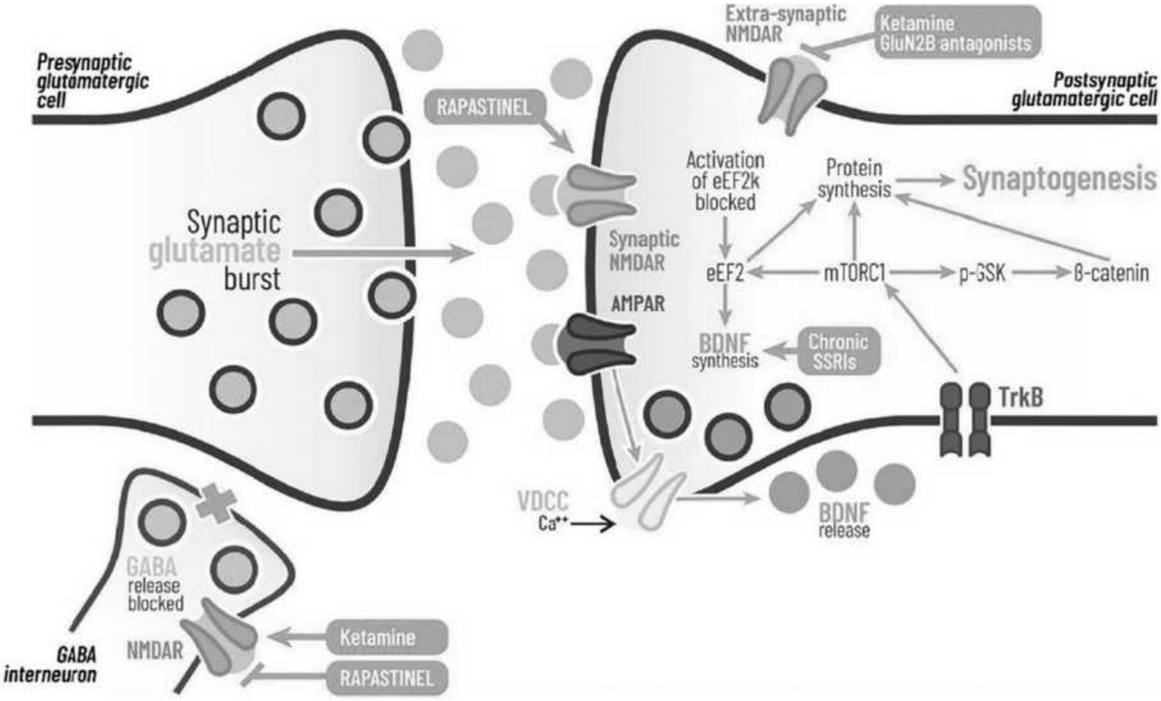
Presynaptic glutamatergic cell
RAPASTINEL
Extra-synaptic NMDAR
Ketamine GluN2B antagonists
Postsynaptic glutamatergic cell
Activation of eEF2k blocked
Protein synthesis
Synaptogenesis
Synaptic glutamate burst
Synaptic NMDAR
AMPAR
eEF2
mTORC1
p-GSK
β-catenin
BDNF synthesis
Chronic SSRIs
TrkB
VDCC
Ca++
BDNF release
GABA release blocked
NMDAR
GABA interneuron
Ketamine
RAPASTINEL

FIGURE 13.8–15. Hypothetical mechanism of action of NMDA modulating medications. (Reprinted from Delfino RS, Surjan J, Bandeira ID, et al. NMDA antagonists and their role in the management of bipolar disorder: A review. *Curr Behav Neurosci Rep.* 2020;7:76–85. doi:10.1007/s40473-020-00201-w.)

Only indirect evidence of altered astrocyte function in mood disorders is available. Elevated inflammatory signaling in mood disorders precipitates activation of IDO in astrocytes, which propagates conversion of tryptophan into kynurenic acid (KA), a known antagonist of dopamine and NMDA receptors. Furthermore, astrocyte dysfunction may alter GABA/glutamate balance in another important way. Recent research has identified astrocytes as an important source of GABA in the CNS. In addition to its role in transmission, GABA may act as an anti-inflammatory molecule by attenuating microglial and astrocytic release of inflammatory cytokines. In this manner, astrocytes, via autocrine signaling mediated by GABA-A receptors, can reduce their own inflammatory signaling. Accumulation of inflammatory and oxidative compounds in MDD may cause astrocytic apoptosis (established by postmortem studies) and precipitate a shift in the microglia/astroglia balance, favoring microglial activity, thus producing excessive glutamatergic transmission relative to GABA output further perpetuating a proinflammatory milieu. In addition, since microglia and astrocytes are principal sources of BDNF, their compromised function and numbers may contribute to alterations in neurotrophic signaling in MDD. Due to histologically established astrocytic apoptosis in MDD, their processes no longer effectively seal off synaptic space, allowing glutamate spillover where it binds to extrasynaptic NMDA receptors. In contradistinction with activation of synaptic NMDA receptors, which leads to enhancement of BDNF synthesis and release, glutamate binding to extrasynaptic NMDA receptors robustly suppresses BDNF synthesis.

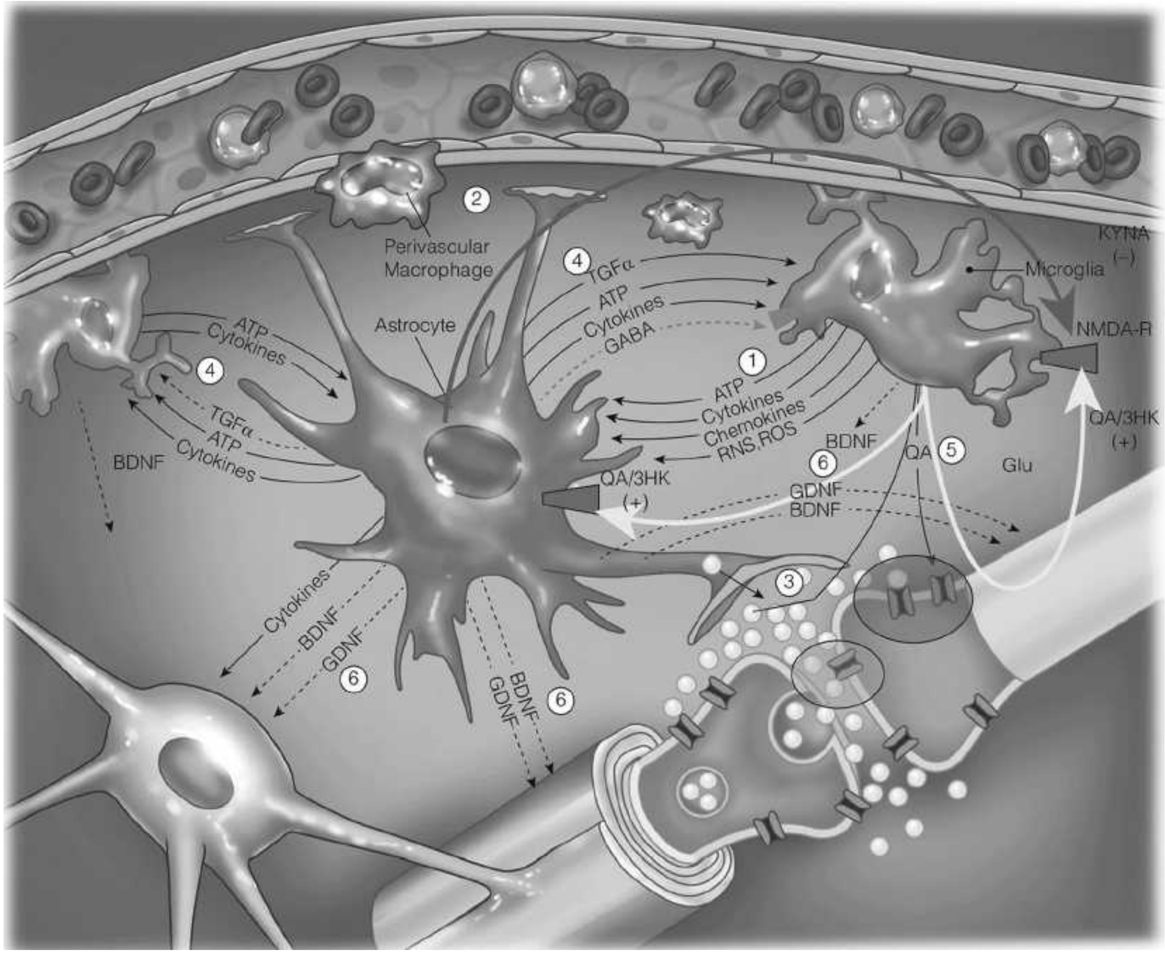
②
Perivascular Macrophage
Astrocyte
④
TGFα
ATP
Cytokines
GABA
KYNA (−)
Microglia
NMDA-R
ATP
Cytokines
④
TGFα
ATP
Cytokines
BDNF
①
ATP
Cytokines
Chemokines
RNS,ROS
BDNF
QA
⑤
Glu
QA/3HK (+)
⑥
GDNF
BDNF
QA/3HK (+)
③
Cytokines
BDNF
GDNF
⑥
BDNF
GDNF
⑥

FIGURE 13.8–16. Interactions at the glia–synaptic junction in mood disorders NMDA-R, *N*-methyl-D-aspartate receptor (glutamate); QA, quinolinic acid; KA, kynurenic acid; 3HK, 3-hydroxy-kynurenine; BDNF, brain-derived neurotrophic factor; GDNF, glial-derived neurotrophic factor; TGF, transforming growth factor; ATP, adenosine triphosphate; Glu, glutamate; GABA, gamma aminobutyric acid. (Reprinted from Maletic V, DeMuri-Maletic B. Bipolar disorders and their clinical management, Part I: Epidemiology, etiology, genetics, and neurobiology. In: Black DW, ed. *Scientific American Psychiatry. Decker Intellectual Properties*; 2019. doi: 10.2310/7800.13099.)

Unlike widespread and prominent glial abnormalities, changes in neuronal morphology and density in MDD appear to be more subtle and regionally limited. A number of postmortem studies noted a greater reduction in pyramidal cell size in the hippocampus, ACC, DLPFC (BA 9), and OPFC (BA 47) of MDD patients relative to controls. The distribution of neuronal pathology remarkably mirrors glial reductions and findings from structural imaging studies. Furthermore, changes in pyramidal neuron morphology and density were also reported in bipolar disorder.

Morphologic changes are predominantly related to neuronal apoptosis and thinning of interneuronal neuropil. Nevertheless, a postmortem study reported a 16% to 22% reduction of pyramidal neuron density in the DLPFC of bipolar patients relative to controls. This finding is particularly intriguing since affected pyramidal neurons are targets of ascending thalamic glutamatergic fibers and represent a cortical destination of the CSTC loop. In addition, DLPFC pyramidal neurons provide top-down feedback to the ACC and amygdala. Therefore, one can plausibly speculate that morphologic and ensuing functional changes in these pyramidal neurons may be related to cognitive difficulties and compromised top-down emotional regulation in bipolar disorder. Moreover, a significant decrease in ACC neuronal size and hippocampal neuronal density separated bipolar patients from controls. It is likely that these postmortem findings in bipolar patients have repercussions regarding affective modulation, declarative and spatial memory, and possibly HPA regulation.

NEUROPLASTICITY AND NEUROTROPHIC CHANGES IN MOOD DISORDERS

Changes in several neurotrophic factors have been detected in patients suffering from mood disorders. Probably, the best studied is BDNF. BDNF has an important role in neuronal maturation, differentiation and survival, synaptic plasticity, long-term memory consolidation, HPA activity regulation, slow-wave sleep modulation, as well as regulation of serotonin, glutamate, and GABA release. Several convergent factors present in MDD converge to impede BDNF synthesis and signaling, such as unfavorable genetic polymorphism, epigenetic methylation of BDNF promoter gene, loss of diurnal rhythm of cortisol secretion, elevation of inflammatory signals, and impaired astroglia function. Burgeoning research has detected a reduction in CSF levels of BDNF in MDD patients relative to controls and an association between the CSF BDNF levels and suicidal ideation in depressed patients. Although the majority of studies reported lower peripheral BDNF levels in MDD patients, particularly in suicidal individuals and ones with the persistent disease form, compared to healthy controls, the correlation between the severity of depression and disease recurrence has been inconsistent. Different antidepressant treatments, including tricyclic and heterocyclic antidepressants, MAOIs, SSRIs, SNRIs, selective norepinephrine uptake inhibitors (NRIs), and ketamine, as well as nonpharmacologic interventions such as ECT, repetitive transcranial magnetic stimulation (rTMS), exercise, and psychotherapy. Peripheral BDNF levels may be an indicator of antidepressant response. A recent study reported that an increase in serum BDNF levels to 126% from baseline, following a week of antidepressant treatment, combined with a 50% reduction in their HAMD-17 over the same time period, predicted the ultimate treatment response with 100% specificity! However, despite successful antidepressant treatment, peripheral BDNF levels in remitted depressed patients most often lag behind BDNF concentrations in healthy controls. A number of clinical studies have noted diminished peripheral BDNF levels in depressed and manic bipolar patients. Furthermore, low BDNF levels have been associated with the clinical severity of depression and mania. One study reported normalization of BDNF levels in conjunction with a successful treatment outcome. A reciprocal relationship between BDNF and inflammatory mediators is an important marker of the progression of bipolar disease.

Glia-derived neurotrophic factor (GDNF) plays an important role in the maintenance of neuronal health, neuroplasticity, microglia function, GABA, and monoamine transmission and cognition. Research has discovered reduced serum levels of GDNF in depressed patients compared to healthy controls. Furthermore, 8 weeks of antidepressant treatment was associated with a significant increase in serum GDNF in depressed patients relative to controls. Both elevated and reduced levels of GDNF have been reported in mania. Unlike the scenario in MDD, bipolar depression does not appear to have a meaningful impact on peripheral GDNF levels. Pharmacologic studies have detected a “normalizing” effect of antipsychotic and mood-stabilizer treatment in manic patients.

SECOND MESSENGERS AND INTRACELLULAR CASCADES

The binding of a neurotransmitter and a postsynaptic receptor triggers a cascade of membrane-bound and intracellular processes mediated by second messenger systems. Receptors on cell membranes interact with the intracellular environment via guanine nucleotide-binding proteins (G proteins). The G proteins, in turn, connect to various intracellular enzymes (e.g., adenylate cyclase, phospholipase C, and phosphodiesterase) that regulate the utilization of energy and formation of second messengers, such as cyclic nucleotide (e.g., cAMP and cyclic guanosine monophosphate [cGMP]), as well as phosphatidylinositols (e.g., inositol trisphosphate and diacylglycerol) and calcium–calmodulin.

Second messengers regulate the function of neuronal membrane ion channels, neurotransmitter synthesis and release, and protein kinase activity. Protein kinase, for example, catalyzes phosphorylation, an energy-liberating process involved in the synthesis and degradation of receptors, ion channels, G proteins, gene transcription, and translation factors. Some studies have reported abnormalities in platelet adenylate cyclase activity, phosphoinositide hydrolysis, intracellular calcium metabolism, and G-protein function in depressive disorders. Moreover, the synaptic effects of antidepressants initiate a series of intracellular reactions that ultimately alter gene activity, resulting in the downregulation of selected monoamine receptors and increased synthesis of BDNF. There is increasing evidence that mood-stabilizing drugs similarly act on G proteins or other second messengers.

ALTERATIONS OF HORMONAL REGULATION

HPA Activity

As noted previously, elevated HPA activity is a hallmark of mammalian stress responses and one of the clearest links between depression and the biology of stress. Elevated HPA activity in depression has been documented via excretion of urinary free cortisol (UFC), 24-hour (or shorter time segments) IV collections of plasma cortisol levels, salivary cortisol levels, and tests of the integrity of feedback inhibition. A disturbance of feedback inhibition is tested by administration of dexamethasone (0.5 to 2.0 mg), a potent synthetic glucocorticoid, which normally suppresses HPA-axis activity for 24 hours. Nonsuppression of cortisol secretion at 8:00 AM the following morning or subsequent escape from suppression at 4:00 PM, or 11:00 PM is indicative of impaired feedback inhibition. Hypersecretion of cortisol and dexamethasone nonsuppression are imperfectly correlated (approximately 60% concordance), suggesting that they may measure overlapping but different aspects of HPA dysfunction. Hypercortisolemia in depression suggests one or more of the following central disturbances: decreased inhibitory 5-HT tone; increased drive from NE, ACh, or CRH; or decreased feedback inhibition from the hippocampus. About 20% to 40% of depressed outpatients and 40% to 60% of depressed inpatients have some evidence of increased HPA activity. Older patients, particularly those with highly recurrent or psychotic depressive disorders, are the most likely to manifest increased HPA activity. Although hypercortisolism is one of the best-replicated biologic correlates of melancholia or endogenous depression, it is hardly a specific abnormality. For example, a short period of starvation or several weeks of partial sleep deprivation can induce hypercortisolism in otherwise healthy people. Neither the sensitivity nor the specificity of these tests of feedback inhibition is sufficient for use as a diagnostic test, and adrenocortical hyperactivity (albeit usually less prevalent) is observed in mania, schizophrenia, dementia, and other psychiatric disorders.

Elevated HPA activity in depression is typically not associated with the physical stigmata of Cushing syndrome but is sufficient to be implicated in the genesis of neurocognitive and neuroimmunologic disturbances. Patients with increased HPA activity are typically less responsive to attention placebo interventions and psychotherapy. Hypercortisolemic depressed patients thus may have a relatively greater need for active pharmacotherapy or ECT. Hypercortisolism usually resolves with effective treatment. When persistent despite effective treatment, increased HPA activity is associated with a high risk of relapse. This is presumably a consequence of the incomplete resolution of the pathophysiology of the depressive episode. One implication of these relationships is that interventions that suppress HPA activity, including dexamethasone and the cortisol synthesis inhibitor ketoconazole, are sometimes used to treat patients with more refractory depressive disorders. Attempts to develop antidepressant medications that directly antagonize central CRH receptors have to date been unsuccessful.

Thyroid Axis Activity

Approximately 5% to 10% of people evaluated for depression have previously undetected hypothyroidism, as reflected by low levels of circulating thyroid hormone, an elevated basal TSH level, or an increased TSH response to a 500-mg infusion of the hypothalamic neuropeptide thyroid-releasing hormone (TRH). Such abnormalities are often associated with elevated antithyroid antibody levels, and unless corrected with hormone replacement therapy, may compromise response to antidepressant medication. A larger subgroup of depressed patients (20% to 30%) shows a blunted TSH response to the TRH challenge. This type of response would usually suggest hyperthyroidism, yet few depressed patients have clinically significant elevations of thyroid hormones. A blunted TSH response in a euthyroid person thus may result from pituitary downregulation consequent to increased TRH "drive." As neurons containing TRH have been identified in a variety of cortical regions, this abnormality may have a suprahypothalamic origin. Increased central TRH secretion, in turn, could result from a homeostatic response to decreased noradrenergic neurotransmission. The therapeutic benefits of adjunctive therapy with 1-tri-iodothyronine (T_3) or other thyroid hormones may be mediated by a dampening of this failed homeostatic response. This abnormality may be most common in individuals who have reduced ability to convert thyroxine to T_3. The major therapeutic implication of a blunted TSH response is evidence of an increased risk of relapse despite preventive antidepressant therapy. Of note, unlike the dexamethasone suppression test (DST), blunted TSH response to TRH often does not normalize with effective treatment.

GROWTH HORMONE (GH)

Growth hormone (GH) is secreted from the anterior pituitary after stimulation by NE and DA. Secretion is inhibited by SST, a hypothalamic neuropeptide, and CRH. Pulsatile GH secretion follows a 24-hour circadian rhythm with a characteristic secretory surge during the first few hours of sleep. The most consistent finding in depression is a blunted GH response to clonidine, an α_2-receptor agonist. Secretory responses after the onset of sleep or the administration of nonselective adrenergic agonists such as desipramine are also blunted in depression. Although the hypothalamus has the highest concentrations of SST, significant concentrations are also found in the amygdala, hippocampus, nucleus accumbens, PFC, and LC. In addition to inhibition of GH release, SST inhibits release or antagonizes the effects of CRH, GABA, and TSH. Decreased CSF SST levels have been reported in depression, and increased levels have been observed in mania.

Prolactin

Prolactin is released from the pituitary by 5-HT stimulation and inhibited by DA. Most studies have not found significant abnormalities of basal or circadian prolactin secretion in depression, although a blunted prolactin response to various 5-HT agonists has been described. This response is uncommon among premenopausal women, suggesting that estrogen has a moderating effect.

ALTERATIONS OF SLEEP NEUROPHYSIOLOGY

Sleep is regulated by complex, temporally, and environmentally cued interactions between the monoamine systems, neuroendocrine secretion, and neuropeptides. At the broadest levels, there is a seasonal tendency for longer and deeper sleep in winter and a circadian propensity for the onset of sleep (i.e., after dark and after midday). The onset of sleep is promoted by a surge in the secretion of the pineal hormone melatonin (after the onset of darkness). Sleep onset is followed by the nocturnal surge in GH secretion, which occurs within the first 90 minutes of sleep onset. A reduction in core body temperature and diurnally low levels of cortisol secretion further promote the maintenance of sleep. Orchestrated across each 24-hour period is an oscillating, 90-minute, infradian cycle defined by rapid eye movement (REM) sleep. Within each cycle, there is a characteristic progression from light to deeper levels of sleep (defined by different types of EEG activity), culminating in the paradoxical central activation of REM sleep, during which most of the night's dreaming occurs. An 8-hour night of sleep thus typically includes four or five cycles consisting of nonrapid eye movement (NREM) sleep and REM sleep. The propensity for deep sleep is greatest within the first 3 hours after sleep onset. REM sleep periods, by contrast, tend to become longer and more intense as the night of sleep progresses.

Prefrontal cortical metabolism is normally decreased during NREM sleep, a time of physical and metabolic rest. The frontal cortex is essentially offline during deep sleep, and the characteristic rhythm of brain activity consists of slow (delta) desynchronized waves of thalamocortical origin. REM sleep, by contrast, is characterized by fast, low-amplitude electrical activity and increased glucose metabolism in the limbic system. REMs are under the direct control of cholinergic neurons in the pons, which is tonically inhibited by the reticular activating system (predominately by histaminergic and noradrenergic neurotransmission) during wakefulness. During sleep, inhibitory 5-HT projections from the dorsal raphe nuclei phasically suppress REM. Pharmacologic manipulations that increase central cholinergic activity lighten sleep and increase phasic REM activity. Dietary depletion of 5-HT and exogenous administration of glucocorticoids similarly can increase phasic REM indices. Injections of CRH and ingestion of potent noradrenergic agonists decrease total sleep time, reduce slow-wave sleep, and suppress REM sleep.

Depression is associated with decreased deep (slow-wave) sleep and an increase in nocturnal arousal. The latter is reflected by four types of disturbance: (1) an increase in nocturnal awakenings, (2) a reduction in total sleep time, (3) increased phasic REM sleep, and (4) increased core body temperature. The combination of increased REM “drive” and decreased slow-wave sleep results in a significant reduction in the first period of NREM sleep, a phenomenon referred to as reduced REM latency. Results of family and twin studies suggest that these related abnormalities are partly heritable. Consistent with the expected behavior of a heritable trait, reduced REM latency and deficits of slow-wave sleep typically persist after recovery from a depressive episode. Blunted secretion of GH after sleep onset is associated with decreased slow-wave sleep and shows similar state-independent or trait-like behavior. Difficulties maintaining sleep and increased phasic REM sleep are associated with hypercortisolism and increased limbic blood flow and glucose metabolism; although some forms of pharmacotherapy can distort sleep profiles, recovery from depression usually is accompanied by at least a partial reversal of these abnormalities. The combination of reduced REM latency, increased REM density and decreased sleep maintenance identifies approximately 40% of depressed outpatients and 80% of depressed inpatients. False-negative studies are commonly seen in younger, hypersomnolent patients, who may actually experience an increase in slow-wave sleep during episodes of depression. Approximately, 10% of otherwise healthy individuals have abnormal sleep profiles, and like dexamethasone nonsuppression, profiles associated with depression are seen in patients with other psychiatric disorders.

Despite clear limitations as a diagnostic test, EEG sleep recordings continue to be an important research tool. For example, patients manifesting a characteristically abnormal sleep profile have been found to be less responsive to psychotherapy and to have a greater risk of relapse or recurrence, and may benefit preferentially from pharmacotherapy. Most antidepressants suppress REM sleep, probably by directly or indirectly activating postsynaptic 5-HT$_{1A}$ receptors. The efficacy of several antidepressants that do not suppress REM sleep, including bupropion, nefazodone, and agomelatine, in patients with pathologically increased REM sleep, warrants further study. The effects of antidepressants on sleep maintenance and slow-wave sleep are more variable and are largely mediated by antihistaminic effects and antagonism of 5-HT$_2$ and possibly melatonin receptors.

STRUCTURAL AND FUNCTIONAL BRAIN IMAGING

Modern magnetic resonance imaging (MRI) scans permit visualization of the living brain, with high resolution of cortical structures such as the hippocampus and anterior cingulate, as well as white matter lesions. One abnormality commonly observed in depressive disorders is an increased frequency of hyperintensities in subcortical regions such as periventricular regions, the basal ganglia, and the thalamus. More common in bipolar I disorder and among older adults, these hyperintensities appear to reflect atherosclerotic change, and as such, illustrate the deleterious effects of the interaction of aging and recurrent affective episodes. Ventricular enlargement, cortical atrophy, and sulcal widening have been observed in some studies of patients with more severe recurrent affective illness. More recent studies utilizing higher-resolution MRI scans have documented reduced volume in a number of relevant regions, including the hippocampus, medial orbital cortex, and anterior cingulate. Diffuse and focal areas of atrophy have been associated with increased illness severity and increased cortisol levels. In several studies, the hippocampal volume reduction has been associated with measures of illness chronicity, such as lifetime days of untreated depression. Volume reduction is thought to partly result from a reduction in the glial matter that supports neurons; suppression of BDNF-mediated connectivity, and ultimately, neurogenesis also may be implicated. Nevertheless, the reduced hippocampal volume has been documented in younger individuals experiencing a first depressive episode and has been shown to be partly under heritable control in primates.

Positron emission tomography (PET) and functional MRI (fMRI) scanning are now widely used for research aimed at visualizing brain activity during rest and various states of activation. Normal sadness is associated with an increase in blood flow and neuronal activity in the thalamus and mPFC, which may be thought of as a nonspecific change associated with diverse emotional responses. More specific activation is seen in the left amygdala, hippocampal formation, and parahippocampal gyrus. Sadness generated by one's own thoughts (as opposed to a video scenario) and anticipatory anxiety are associated with a relative increase in blood flow to the anterior insular cortex. The most widely replicated fMRI findings associated with depression pertain to resting-state connectivity, which is abnormally increased in the DMN and decreased connectivity within the CNN.

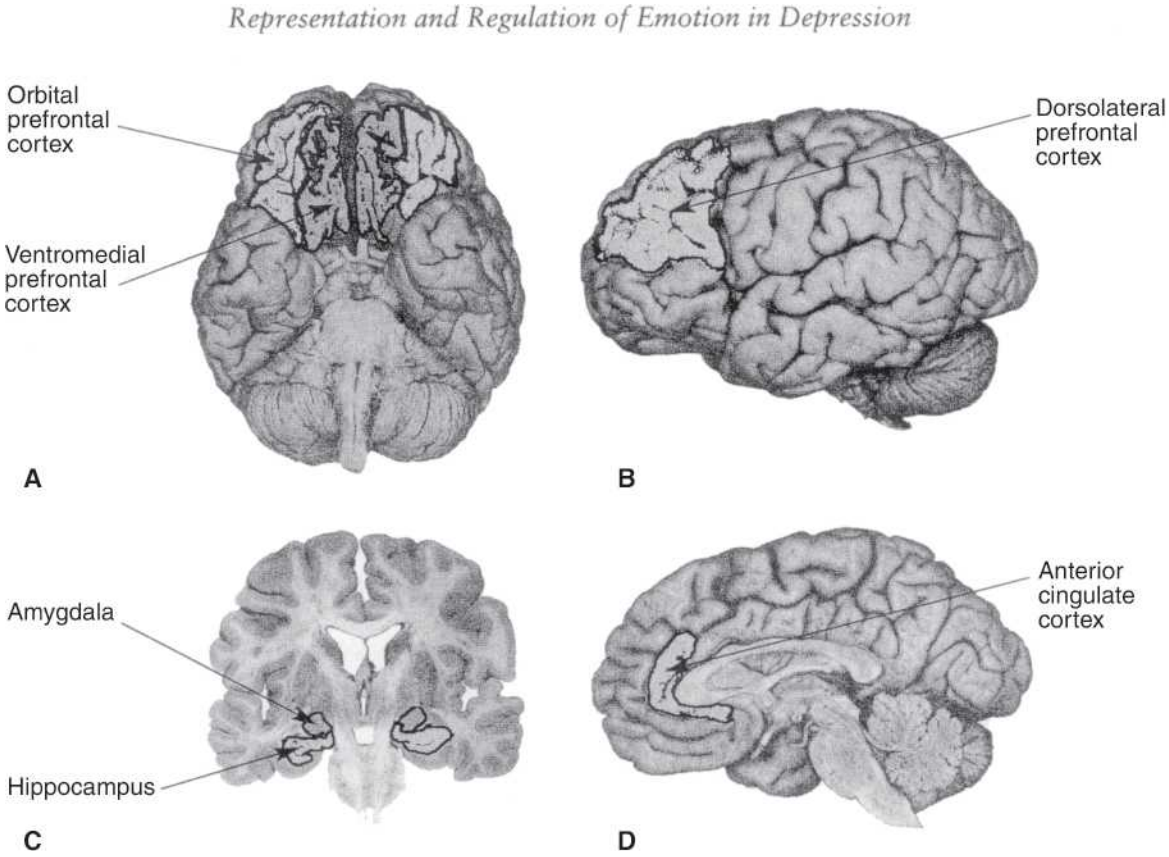
Representation and Regulation of Emotion in Depression
Orbital prefrontal cortex
Ventromedial prefrontal cortex
Dorsolateral prefrontal cortex
A
B
Amygdala
Hippocampus
Anterior cingulate cortex
C
D

FIGURE 13.8–17. Key brain regions involved in affect and mood disorders. **A:** Orbital prefrontal cortex and the ventromedial prefrontal cortex. **B:** Dorsolateral prefrontal cortex. **C:** Hippocampus and amygdala. **D:** Anterior cingulated cortex.

The most widely replicated PET finding in depression is decreased anterior brain metabolism, which is generally more pronounced on the left side. Thus, depression may be characterized by a relative increase in nondominant hemispheric activity. Decreased cerebral blood and reduced glucose metabolism in regions such as the dorsolateral PFC appear to be severity-dependent and are correlated with difficulties inhibiting negative affective expression, reduced abstraction, and impaired executive problem-solving skills. Abnormalities of regional cerebral metabolism have been observed in unipolar and bipolar depressions and appear to be somewhat state-dependent. Furthermore, hypofrontality tends to reverse in concert with shifts from depression into hypomania, such that there are greater left-hemisphere reductions in depression compared to greater right-hemisphere reductions in mania. Other studies have observed more specific reductions in reduced cerebral blood flow or metabolism, or both, in the dopaminergically innervated tracts of the mesocortical and mesolimbic systems in depression. Again, there is evidence that antidepressants at least partially normalize these changes.

In addition to a global reduction in anterior cerebral metabolism, increased glucose metabolism has been observed in several limbic regions, particularly among patients with relatively severe recurrent depression and a family history of mood disorder (see Fig. 13.8–17). During episodes of depression, increased glucose metabolism is correlated with intrusive ruminations. Of note, increased metabolic activity in the amygdalae and decreased activity in the dorsolateral PFC appear to be independent dysfunctions. Increased paralimbic metabolism appears to be reversible with effective pharmacotherapy, but, in one study, the abnormality re-emerged when recently remitted patients were restudied off medication. State-dependent or trait-like paralimbic hypermetabolism has been proposed to be an “emotional amplifier” that helps to distort the signal of various stressors in those at high risk of depression.

IMMUNOLOGIC DISTURBANCE

Overview

Although long conceptualized as separate systems with different biologic goals, it is now widely understood that the CNS and immune system are better conceptualized as an interconnected network tasked with protecting organismal integrity from threats, both internal (i.e., infectious, cancer) and external (i.e., predators, threatening conspecifics). In addition to sharing the capacity to learn from experience, the CNS and immune system are biologically interdigitated in ways that would have astounded both neurologists and immunologists practicing their crafts in the mid-20th century. Initial findings linked depression with immunosuppression, as measured by reduced adaptive immune responses to immune stimuli. Subsequent studies highlighted associations between activation of the innate immune system (i.e., inflammation) and the development of depression. In general, these studies have carried the day in terms of scientific focus, but more recent research is again highlighting important and complex associations between adaptive immunity (especially T-cell function) and stress-related conditions such as MDD. In this section, we focus on the more robust literature linking inflammatory processes to the pathogenesis and treatment of depression.

Evidence that Inflammation Is Associated with MDD: Cross-Sectional and Inflammatory Stimulation Studies

Given our general failure to identify unequivocal biologic alterations associated with MDD, it is quite striking that the top contender for such an alteration may be increased peripheral inflammation. Hundreds of studies and multiple meta-analyses confirm that, as a group, individuals with MDD demonstrate increased circulating levels of a number of inflammatory mediators, with the most reliable increases seen for the proinflammatory cytokines interleukin (IL)-6 and tumor necrosis factor (TNF) and the acute phase reactant CRP. Gene expression studies suggest these inflammatory molecules likely derive from activated peripheral blood immune cells, given findings that MDD is associated with activation of canonical inflammatory signaling pathways, such as nuclear factor-kappa-beta and the NLRP3 inflammasome which activates the potent proinflammatory cytokine IL-1-beta. MDD has also been repeatedly associated with increased inflammatory activity in the CNS, as evidenced by increased levels of inflammatory mediators in CSF and activation of microglial cells, which are the resident immune cells in the CNS. A robust literature also reports increased inflammatory mediators in the brain parenchyma of individuals with depression in postmortem studies, especially in depressed individuals who died by suicide.

These types of cross-sectional and postmortem studies can report association but are only suggestive of causation. However, two other lines of research provide strong support for the ability of inflammatory processes to produce depression and related emotional states, such as anxiety, fatigue, and irritability. The first line of evidence comes from many longitudinal studies showing that individuals with increased inflammatory biomarkers are more likely than others to develop depression in the future. Related literature from large population-based datasets finds that the occurrence of serious infectious or autoimmune illness (both of which are associated with inflammatory activation) predicts the later development of depression, with infections serious enough to require hospitalization being associated with a 42% increase risk of subsequent suicide.

The second line of evidence supporting the depressogenic potential of inflammation comes from very consistent literature showing that inflammatory stimuli induce symptoms commonly seen in depression, as well as closely related symptoms such as anxiety and irritability. Of these inflammatory stimuli, the most extensively studied is treatment with the inflammatory cytokine interferon (IFN)-alpha, which was widely used for many years as a treatment for certain cancers (e.g., malignant melanoma) and infections (e.g., hepatitis C). Because patients would voluntarily expose themselves to IFN-alpha on a long-term basis, this treatment modality provided a natural experiment in the behavioral and neurologic effects of chronic inflammation. Depending on the dose, chronic exposure to IFN-alpha produces the full syndrome of MDD in 30% to 50% of individuals not depressed prior to commencing the treatment and induces far higher rates of subsyndromic mood disturbance. That inflammatory processes are responsible for these behavioral changes comes from many lines of evidence, including the observation that increased induction of the inflammatory second messenger p38 mitogen-activated protein kinase in response to the first injection of IFN-alpha predicts the development of depression a month later with a high degree of accuracy. Evidence that the behavioral state induced by IFN-alpha is a “true” depression comes from studies showing a high degree of symptom overlap between IFN-alpha-induced depression and depression in the general population, as well as from studies showing that the depressive state induced by IFN-alpha can be prevented and/or treated with standard antidepressant medications. In addition to IFN-alpha, other shorter-term immune stimuli, such as typhoid vaccine and lipopolysaccharide, have been shown to induce time-limited depressive symptoms. Intriguingly, females appear to be more vulnerable to the depressogenic effects of these brief inflammatory stimuli, which may contribute to the higher rate of depression seen in females worldwide.

It is important to note that the degree of inflammatory activation seen in populations of medically healthy individuals with depression is typically mild, with mean values frequently in the upper range of normal for biomarkers such as CRP, which have standardized values. Nonetheless, because individuals living in industrial societies tend to demonstrate fairly stable levels of inflammation across time (in contradistinction to individuals in hunter-gatherer societies who show far more intraindividual variation), even these mild increases in inflammatory biomarkers are known to significantly increase the risk for a range of medical conditions (e.g., diabetes, cardiovascular disease, dementia). Given this, it shouldn’t be surprising that the modest increases observed in depression are also able to produce profound alterations in mood, cognition, and behavior.

Inflammation Causes Depression, But Depression Is Not an Inflammatory Condition

The finding that groups of depressed people have elevated inflammation compared to nondepressed comparison groups could imply two very different possibilities. It might be that almost every person with MDD has mildly increased inflammation, which would be consistent with being a “final common pathway for the development of MDD.” Under this scenario, MDD would be reconceptualized as an inflammatory condition. On the other hand, it might be that the average increase in inflammation seen in groups of depressed individuals reflects the effect of a subgroup of depressed individuals with more markedly increased inflammation who are pulling up the mean value for the depressed group as a whole. In fact, it is now established that this latter scenario is the case: only a subgroup of depressed individuals demonstrate increased inflammation (as measured by a CRP >3 mg/L), with this group typically comprising 30% to 45% of the larger depressed population, depending on a variety of other factors independently associated with inflammation, including obesity, childhood adversity, socioeconomic disadvantage, and treatment resistance. This means that the majority of depressed individuals do not show signs of elevated peripheral inflammation. While it is possible that inflammation may contribute to depressive pathogenesis in some individuals with normal inflammatory biomarkers but enhanced downstream sensitivity to their effects, most current data argue against this. As we discuss below, depressed patients with increased inflammation show differences in brain functional connectivity, metabolism, and response to anti-inflammatory medications and standard antidepressants. All of this points to the fact that inflammation likely only contributes substantially to depression in individuals with increased inflammation, whereas other brain–body pathways, either separate or downstream from inflammation, are more relevant to pathogenesis in individuals with MDD but lower levels of peripheral inflammation.

Inflammation Promotes Specific Symptoms with Transdiagnostic Implications. Just as MDD is not always caused by inflammation, inflammation does not appear to be an “equal opportunity” promoter of depressive symptoms. Indeed, convergent data indicate that inflammation is especially likely to induce the interrelated symptoms of anhedonia and psychomotor slowing, as well as suicidal behavior. Deficits in anhedonia and psychomotor activity have been repeatedly linked to the tropism of inflammatory mediators for the basal ganglia, where cytokines have been shown to impair dopamine signaling directly and indirectly while promoting glutamatergic excitotoxicity, leading to reduced basal ganglia functionality and impaired connectivity, both internally within the striatum and between the striatum and other areas implicated in depression such as vmPFC. Anhedonia is a complex phenomenon, and inflammatory processes do an admirable job of deconstructing it into its more basic and homogeneous elements. Studies in animals and humans indicate that inflammation impairs effort-based motivation, reinforcement learning, and reward anticipation without impacting the capacity to experience reward while it is occurring. Anyone who has experienced the difficulty of making an effort or looking forward to things while depressed will understand this constellation of effects.

It is increasingly recognized that associations between inflammation and depression in the modern world might be a legacy of adaptive advantages in pathogen-host defense conferred by depressive symptoms and biology across hominid evolution. However, the fact that inflammation increases the risk of suicidal ideation and behavior is a challenge that all immune-based and other types of adaptive explanations for depression will have to address. It is hard to imagine many scenarios in which suicide, especially early in life, might advance survival and reproduction.

Findings in support of an association between inflammation and suicide come from several sources. Postmortem studies have observed increased levels of inflammatory mediators in the brains of depressed individuals who died by suicide versus depressed and nondepressed individuals who perished from other causes. Postmortem studies also indicate that patients with either depression or schizophrenia who die by suicide have increased microgliosis, a marker of increased inflammatory microglial activation. Markers of increased inflammation in postmortem brain samples have been found in both adults and adolescents who died by suicide. Suicidal thoughts and behavior in living individuals have been repeatedly associated with increased CNS and peripheral inflammatory cytokines when compared to patients with similar disorders without suicidal issues. Finally, suicidal ideation was among the most reduced depressive symptoms in medically healthy patients with MDD and elevated inflammation (CRP >5 mg/L) treated with the TNF antagonist infliximab.

Of all the contributions that inflammation has made to our understanding of depression, perhaps none is more intriguing than the critique it offers to current psychiatric diagnostic nosology. On the one hand, inflammatory processes have more causative power over certain depressive symptoms than others, thereby providing what may be the field’s best evidence for a subtype of depression more biologically homogeneous than the disorder as a whole. On the other hand, inflammation promotes symptoms that are strikingly transdiagnostic, suggesting that immune-driven depression has deeper ties with elements of other psychiatric conditions than would be apparent from their current nosologic status as separate disorders. This is clearly the case with anhedonia, which is a prominent component of the negative symptom complex of schizophrenia and is common in individuals struggling with post-traumatic stress disorder. That inflammation simultaneously promotes specific symptoms that are not specific to any particular psychiatric condition may help explain the well-established observation that most psychiatric conditions (e.g., schizophrenia, bipolar disorder, PTSD) show as strong—or stronger—associations with increased inflammation as does MDD.

Mechanisms by which Inflammation Promotes Depressive Symptoms

Overview. A good rule of thumb for understanding the role of peripheral inflammation in the pathogenesis of MDD is that inflammation can cause any and all biologic effects that have been observed to differ between depressed and nondepressed individuals. Indeed, the breadth of depression-relevant CNS changes induced by inflammation contributed to earlier conceptualizations of MDD as an inflammatory disorder. As we discuss below, we now know this is not true, and that in fact, inflammation is probably only directly relevant to a subgroup of depressed individuals. Nonetheless, the range of depressive signatures induced by inflammation is impressive and includes; (1) glucocorticoid resistance, (2) flattening of the diurnal cortisol rhythm, (3) changes in sleep architecture, (4) alterations in monoamine and glutamatergic neurotransmission, and (5) loss of glial cell numbers and function. Recent research suggests that changes in neurotransmission and impairment of glial cell function, with resultant effects on functional brain circuits, may be especially relevant for depressive symptoms, such as anhedonia, and are therefore the focus of what follows.

Pathways by which Inflammatory Mediators Access the CNS. Given the centrality of the brain to cognition, emotion, and behavior, one would de facto assume that peripheral inflammatory mediators must have pathways by which they impact CNS activity, and indeed this is the case. It has long been known that inflammatory cytokines can access the brain via fenestrated capillaries in circumventricular organs, via saturable transporters, and via stimulation of peripheral sensory afferent nerves, which project to the nucleus of the solitary track, with subsequent activation of catecholaminergic projections to higher brain areas. Recent years have identified other mechanisms that speak to an even more intimate and profound connection between peripheral immunity and the brain. We now know that the brain can respond to an infectious and noninfectious threat by the release of chemokine (C–C motif) ligand 2 (CCL2), which results in monocytic and T-cells being released from bone marrow that subsequently traffic to the CNS. Depending on the initial stimulus, these cells can either cross into the brain parenchyma or congregate in the meninges, from which they secrete cytokines that influence aspects of CNS function highly relevant to depression. It appears that the behavioral effects of this immune cell trafficking depend, at least to some measure, on whether the cells involved demonstrate an inflammatory (e.g., “M1 type monocytes, CD8 effector T-cells) or an anti-inflammatory (e.g., “M2 type” monocytes, regulatory T-cells [Treg]) phenotype. Studies in rodents demonstrate that stress resilience (as well as learning from stressors) requires the trafficking of immune cells into the subarachnoid space, from whence they secrete the cytokine IL-4, which crosses into the CNS to induce BDNF in the hippocampus. Importantly, a wide range of antidepressant modalities likewise increase BDNF in the hippocampus. A final mechanism by which peripheral inflammation can signal the CNS is via disruption of the blood–brain barrier (BBB), allowing cytokines direct access to the brain parenchyma. Recent animal work indicates that rather than promoting wholesale BBB disruption that would indicate random pathology, psychological stressors that activate peripheral inflammation produce targeted BBB disruption in areas in which cytokines have specific behavioral effects that—while contributing to depression—likely served adaptive purposes across evolutionary time. In particular, stress-induced inflammation has been shown in animal models to reduce the expression of claudin 5 with resultant BBB disruption in the area of the nucleus accumbens, a brain region known to play a key role in the ability of inflammation to induce anhedonia. Similar decreases in claudin 5 have been found in the nucleus accumbens of postmortem brains of individuals with depression.

Effects of Inflammation on CNS. This chapter provides various perspectives on the relevance of monoamine neurotransmitters to the pathogenesis of depression. An immunologic point of view comports with the idea that monoaminergic signaling may not be the *sine qua non* of biologic abnormalities that produce depression but still may play an important role in subgroups of depressed individuals, in this case, individuals with increased inflammation. Of the monoamines impacted by inflammation, dopamine stands out as the most effected and best studied for its role in contributing to anhedonia and psychomotor slowing, both symptoms common to depression and other mental illnesses (e.g., negative symptoms of schizophrenia). Multiple lines of evidence in animals and humans demonstrate that inflammatory mediators target all aspects of dopamine signaling. Inflammation attenuates the production of dopamine via the oxidative stress-induced destruction of tetrahydrobiopterin (BH4), a cofactor essential for the conversion of precursors such as phenylalanine to monoamines, including dopamine. In humans, IFN-alpha administration increased the phenylalanine/tyrosine ratio (a marker for decreased BH4 activity), which was associated with increased fatigue and lower concentrations of CSF dopamine and the dopamine metabolite HVA. Inflammation antagonizes dopamine release from presynaptic nerve terminals via reductions in VMAT2, which packages neurotransmitters for release across the synaptic cleft. Finally, inflammation has been shown to impair postsynaptic D2 receptor function.

While recent attention has turned to dopamine, earlier studies tended to focus more exclusively on the effects of inflammation on serotonin metabolism. Inflammatory cytokines such as TNF and IFN-gamma induce the enzyme indoleamine 2,3-dioxygenase (IDO), which plays a key role in constraining immune activation by shunting tryptophan from the production of serotonin to the production of kynurenine. A number of studies found that increased peripheral IDO activity predicted depression in response to IFN-alpha treatment, raising the possibility that inflammation vitiated serotonin availability in the CNS. However, studies of the CFS showed that tryptophan levels were normal during IFN-alpha treatment and showed no association with depressive symptoms, whereas kynurenine levels were markedly increased and were associated with the development of depressive symptoms. Kynurenine is actively transported into the CNS, where it is metabolized into several neuroactive compounds, especially the neurotoxic metabolite QUIN, which has been associated with depressive symptoms in patients receiving IFN-alpha and in medically healthy patients with MDD and elevated levels of inflammation. Increased CNS QUIN in medically healthy patients with MDD has also been associated with volume loss in the striatum and treatment resistance.

In the well-functioning brain, glutamatergic neurotransmission is tightly controlled and generally restricted to synaptic signaling, where it plays an essential role in long-term potentiation and other processes necessary for learning, memory, and other cognitive functions. This control is required given the excitotoxic/neurotoxic potential of glutamate when present in excessive amounts or in inappropriate locations within the CNS. Glial cells, which surround neural synapses, play a definitive role in this regulatory control. In particular, astrocytes are responsible for removing 90% of synaptic glutamate via excitatory amino acid transporters (EAATs). Even before the advent of ketamine as a treatment for MDD, which operates by antagonizing the glutamatergic NMDA receptor, significant evidence implicated alterations in glutamate, and its metabolites in the pathogenesis of depression and reductions in astrocytes and oligodendrocytes as well as EAATs are among the most consistent findings seen in postmortem brains of individuals with MDD. Inflammatory mediators have been found to induce all these changes. TNF and IL-1 beta have been shown to decrease the expression of EAATs on glial cells, reverse the polarity of the transporters such that glutamate is released instead of taken up and via these processes promote astroglia and oligodendrocyte death, further reducing the capacity of these cells to regulate glutamate levels. By enhancing the function of cystine/glutamate (xCT) transporters on glial cells and by inducing the binding of adenosine triphosphate (ATP) to purinergic P2X7 receptors, inflammatory mediators (and related processes such as oxidative stress) further extrude glutamate into extracellular spaces. This process allows glutamate to engage with extrasynaptic NMDA receptors located in postsynaptic cell processes. Unlike synaptic NMDA receptors, which stimulate growth factor release, stimulation of extrasynaptic NMDA receptors attenuates growth factors such as BDNF, with subsequent loss of neurotrophic support for the relevant neural elements. QUIN has also been shown to stimulate extrasynaptic NMDA receptors, highlighting the relevance of this neurotoxic pathway to the depressogenic effects of inflammation. Neuroimaging studies also point to a role for glutamatergic signaling in the capacity of inflammation to induce depression. Studies using magnetic resonance spectroscopy have observed increased glutamate in the left striatum, and dorsal anterior cingulate gyrus in patients undergoing treatment with IFN-alpha, and medically healthy individuals with MDD and increased inflammation also demonstrate increased left striatal glutamate. In both cases, increased striatal glutamate was associated with features of anhedonia and psychomotor performance.

The effects of inflammation on neurochemical functioning in CNS can be seen writ large in neuroimaging studies showing associations between immune activation and changes in patterns of functional brain activity. Studies of individuals receiving either chronic (IFN-alpha) or acute (LPS, typhoid vaccine) immune stimulation are concordant with studies in patients with MDD and increased inflammation in showing reduced activity in the basal ganglia and increased activity in areas of the cingulate cortex important for modulation of mood, assessment of risk, and error detection.

Multiple imaging modalities have demonstrated altered/impaired basal ganglia activity in patients receiving IFN-alpha, and these changes reliably track with anhedonia (ventral striatum) and psychomotor slowing (motor areas of basal ganglia). fMRI studies in healthy volunteers receiving an endotoxin challenge show a similar blunting of ventral striatal activity in anticipation of reward, with this blunted response associated with the emergence of depressive symptoms. Typhoid vaccination was found to decrease functional connectivity within mesolimbic circuits, including the nucleus accumbens and sgACC cortex, with these changes associated with increased IL-6 and depressive symptoms. In unmedicated medically healthy patients with MDD, increased concentrations of plasma CRP are associated with decreased functional connectivity between the ventral striatum and vmPFC, which correlated with increased anhedonia. In the same population, decreased functional connectivity between the dorsal striatum, vmPFC, and the presupplementary motor area was associated with psychomotor slowing. Inflammatory stimuli (e.g., IFN-alpha, endotoxin) have also been shown to activate brain areas comprising threat circuitry, including the posterior-anterior cingulate, amygdala, insula, and hippocampus. For example, treatment with IFN-alpha in patients with chronic hepatitis C virus infection increased right amygdala reactivity to sad faces. Interestingly, pharmacologic blockade of TNF signaling in patients with rheumatoid arthritis produced the opposite effect. Taken as a whole, an important implication of the imaging studies is that inflammation produces regional and circuit-level alterations in brain function that are associated with symptoms that are also most prominently induced by inflammatory stimuli (i.e., anhedonia, psychomotor slowing, anxiety) but that likely cut across currently recognized psychiatric conditions.

Treatment Implications

If inflammation causes depression, the obvious conclusion is that reducing inflammation should reduce depression in patients with elevated inflammation. Multiple lines of evidence suggest that this is true. However, because MDD is not an inflammatory condition per se, it is not true that all patients with depression will benefit from pharmacologic strategies for reducing inflammation. Moreover, because inflammation promotes some depressive symptoms more than others, anti-inflammatory medications would be predicted to show a larger benefit for these symptoms than for symptoms less robustly induced by inflammation. Again, evidence suggests this is the case.

It is a paradox that the most extensively studied anti-inflammatory agents are the least informative in terms of evaluating the value of reducing inflammation as a depression treatment strategy. Taken together, many studies have examined the antidepressant potential of nonsteroidal anti-inflammatory drugs (NSAIDs), cyclooxygenase 2 (COX 2) inhibitors, omega-3 fatty acids, minocycline, and other agents that–while having anti-inflammatory properties–also have a range of off-target effects. These off-target effects have been shown in various studies to have both anti- and prodepressant properties, thereby confounding any observed associations between these nonspecific anti-inflammatories and effects on depression. Not surprisingly, studies of these agents as antidepressants paint a mixed picture. Several meta-analyses suggested benefits, but these included studies with effect sizes likely to be unrealistic. More recent large randomized, placebo-controlled trials have generally not shown that these agents have consistent antidepressant effects in general populations of medically healthy individuals (i.e., in depressed populations not enriched for increased inflammation).

Fortunately, a number of studies have examined the antidepressant effects of cytokine antagonists. These medications block specific cytokines with minimal or no off-target effects, making the establishment of causality between blocking inflammation and any subsequent behavioral effects tractable. Moreover, these agents are large molecules that do not readily cross the BBB. Because of this, any behavioral effects can likely be ascribed to the modulation of peripheral inflammation, thereby providing a new line of evidence that bodily inflammation can drive brain function in ways productive of mental illness.

The antidepressant properties of cytokine antagonists have been examined in patients with primary autoimmune conditions and in medically healthy patients with MDD, and the results provide consistent evidence for an antidepressant effect in patients with elevated inflammation and no antidepressant signal in those with normal levels of inflammatory biomarkers. A mega-analysis of studies in autoimmune conditions (10, 743) that examined antagonists against IL-6, Il-12/23, and TNF found an overall modest antidepressant effect (*d* = 0.29), with the strongest benefits for IL-6 and IL-12/23 antagonists (trend level for TNF). These effects were only observed in the substratum of patients with elevated depressive symptoms, and results were attenuated when adjusted for improvements in the underlying medical condition.

While suggestive, results in patients with autoimmune conditions are confounded by low rates of depression prior to treatment and improvements in the underlying medical disease, which can secondarily improve depression. The obvious resolution to these limitations is to examine cytokine antagonists in medically healthy individuals with depression. Two studies with the TNF antagonist infliximab and one study with the IL-6 antagonist sirukumab have done this and provided remarkably consistent findings. These agents do not show an antidepressant effect in general populations of medically healthy individuals with depression (either unipolar or bipolar) but do provide a modest benefit in patients with elevated inflammation prior to treatment and perhaps in individuals with early life adversity, which has been repeatedly observed to be associated with increased inflammation (at rest and in response to stress) in adulthood. However, the level of inflammation required to expose the antidepressant effect is quite high (CRP >5 mg/L for infliximab and >8 mg/L for sirukumab), suggesting that the majority of depressed patients will not benefit from anti-inflammatory treatments. In fact, results from the two studies of infliximab indicate that TNF blockade may be counter-productive in patients with depression and lower levels of inflammation. In 60 patients with treatment-resistant depression, those with pretreatment CRP less than 5 mg/L showed a significantly larger antidepressant effect following treatment with saltwater placebo than treatment with the TNF antagonist. Numerically similar results were found in a study of infliximab for bipolar depression. Interestingly, the largest randomized trial of omega-3 fatty acids as monotherapy for MDD found a similar pattern: patients with lower levels of pretreatment inflammation responded better to placebo than the fatty acid. Similarly, a large randomized, placebo-controlled trial of omega-3 fatty acids as a preventive measure in adults 50 years and older found an increased prevalence of depression over time in those who received the fatty acid.

The literature is mixed on whether standard antidepressants (and antidepressant augmenting agents such as atypical antipsychotics) have anti-inflammatory effects and whether any such effects might contribute to their therapeutic efficacy. More consistent have been studies showing that increased pretreatment inflammatory biomarkers are associated with a poorer response to both antidepressants and psychotherapy, but even here, there is a wrinkle. More recent studies suggest a picture more in line with results from studies of anti-inflammatory agents, which is that patients with higher and lower levels of inflammation likewise respond differently to standard agents. Although based on post hoc findings as of this writing, several studies indicate that patients with elevated peripheral inflammation (often measured as CRP >1 mg/L) respond less well to serotonergic antidepressants and preferentially to agents with catecholaminergic–and especially dopaminergic–activity. This pattern has also been observed for the dopaminergic modulating agent lurasidone in bipolar depression, where response showed a linear relation with CRP concentrations, and only patients with elevated CRP showed a benefit of lurasidone over placebo. Some evidence suggests that increased inflammation may also predict response to ketamine. Taken as a whole, these results are conceptually gratifying, given convergent evidence that inflammatory processes preferentially disrupt dopaminergic and glutamatergic signaling while being less relevant for serotonin activity in the CNS.

Future Directions. Despite remarkable progress over the last 20 years, many key questions remain regarding the role of peripheral and CNS immunity in the pathogenesis and treatment of MDD. Most studies to date have examined innate immune processes, in particular inflammation, with far less work being done to clarify the extent to which adaptive immune functions may also contribute to depression or behavioral resilience. The relevance of this gap in our knowledge is highlighted by recent animal studies showing that T-cells directly interact with CNS processes to promote behavioral processes of direct relevance to stress resilience, learning, and memory—all of which are important for MDD. When it comes to the role of innate immune inflammation in the pathogenesis of depression, data increasingly suggest that inflammation might be best considered as akin to climate and its effects on downstream mediators as akin to weather. Inflammatory molecules are among the most ancient of life’s inventions, and their effects on mood are likely the result of producing a background climate conducive to depression driven more immediately by downstream mediators, such as monoamine neurotransmitters, glutamate, and kynurenine pathway metabolites. The degree to which these downstream mediators are specific for inflammation or relevant to depression more broadly is also a key question, as this will influence the search for optimal predictive biomarkers, or collections of biomarkers, to identify subgroups of depressed patients most likely to benefit from anti-inflammatory treatment strategies. A better understanding of the pathophysiology of an immune-based subtype of depression will also likely help identify optimal anti-inflammatory treatment strategies and whether some of these strategies are most effective broadly or whether even more fine-grained subtypes of inflammatory depression will be identified and found to respond to the targeting of specific inflammatory mediators. Finally, the recurrent observation that anti-inflammatory agents may be harmful compared to placebo in patients with MDD and low levels of peripheral inflammation deserves more attention and study, both for the obvious pragmatic reason of patient safety and because it might suggest that immune processes are relevant for these patients also, but in ways that are different and would require different treatment approaches than in those with high inflammation.

SUMMARY

More severe and persistent depressive states, including those classified as major depressive disorder, are associated with a wide range of neurobiologic disturbances, which in turn are linked to at least some of the differences observed in clinical presentations and response to specific treatments. Some of the disturbances are better understood as *traits*, which may be either inherited or acquired, whereas others are clearly state-dependent and are reversible by treatment or spontaneous remission. Some of the state-dependent abnormalities associated with major depressive disorder, which occur more commonly in older patients with more severe symptomatology, include increased phasic REM sleep, poor sleep maintenance, hypercortisolism, impaired cellular immunity, reductions in anterior cerebral blood flow and glucose metabolism, and increased glucose metabolism in the amygdala. Together, these changes appear to reflect the progressive effects of otherwise adaptive short-term responses to sustained stress. Once manifest in this form, such severe or melancholic depressive episodes tend to be longer, more disabling, more prone to relapse, and more likely to benefit from pharmacotherapy or ECT (vis-à-vis nonspecific or psychotherapeutic interventions).

More trait-like abnormalities may include decreased slow-wave sleep, reduced REM latency, blunted nocturnal GH secretion, and increased connectivity within the DMN. These abnormalities are associated with an early age of onset and increased vulnerability to recurrent illness. The heritability of these abnormalities, which at least partly reflects decreased 5-HT inhibition of responses to stress or threat, is both inferred from family studies and other at-risk paradigms and has been linked to inheritance of the *s* allele of the 5-HT transporter.

Examples of more persistent but acquired abnormalities may include global and focal changes, such as volume reduction in the hippocampus and anterior cingulate, hypertrophy of the adrenal cortex, periventricular hyperintensities, and alterations in CRH synthesis. Blunted TSH response (to a TRH infusion) and dexamethasone nonsuppression may represent *hybrids* in that these abnormalities can be slow to normalize, and when persistent after remission, convey high risk of relapse. Aging and the accumulating consequences of recurrent depressive episodes are inextricably connected. However, the diseases of aging that ravage brain function definitely increase the risk of depression and decrease responsivity to conventional forms of treatment. The late-onset form of depression associated with cerebral atherosclerosis illustrates the subtle interplay between brain function and mood disorder.

There is no doubt that vulnerability to mood disorders is partly heritable. This type of heritability is most likely polygenetic, and in all likelihood, will be best understood through models that include gene-environment interactions. Nevertheless, increased heritability is associated with an earlier age of onset, greater comorbidity, increased risk of recurrent illness, and an increased likelihood of hypomanic or manic episodes. Is it ironic that two of the more heritable forms of mood disorder, early-onset chronic depressions, and bipolar depression, are relatively less likely to be associated with state-dependent neurobiologic disturbances? Different genetic vulnerabilities, different clinical presentations, and age-dependent differences in the effects of depression on brain function depression are likely to explain this apparent paradox.

Ultimately, depression remains a fundamentally human experience, partly because people can all relate to sadness, grief, and the heartbreak of lost love and partly because of the contrived nature of experiments that subject animals to prolonged, inescapable stress. Strong and sustaining affective bonds and an enduring sense of self-worth and competence are simply too important, and assaults on these fundamental aspects of well-being are too frequent for some humans to withstand without becoming depressed. That some are more vulnerable to depression than others is without question, as is the association of stress, depression, and numerous reproducible changes in brain function. Understanding the mechanisms of successful and unsuccessful adaptation and elucidating the alterations in brain function that predispose to and maintain depressive disorders represent the best hope to prevent and relieve the misery and suffering of tens of millions of people.

FURTHER READINGS

Amstadter AB, Maes HH, Sheerin CM, Myers JM, Kendler KS. The relationship between genetic and environmental influences on resilience and on common internalizing and externalizing psychiatric disorders. *Soc Psychiatry Psychiatr Epidemiol*. 2016;51(5):669–678.

Anderson KM, Collins MA, Kong R, et al. Convergent molecular, cellular, and cortical neuroimaging signatures of major depressive disorder. *Proc Natl Acad Sci U S A*. 2020;117(40):25138–25149.

Badini I, Coleman JRI, Hagenaars SP, et al. Depression with atypical neurovegetative symptoms shares genetic predisposition with immuno-metabolic traits and alcohol consumption. *Psychol Med*. 2022;52(4):726–736.

Bogdan R, Pagliaccio D, Baranger DA, Hariri AR. Genetic moderation of stress effects on corticolimbic circuitry. *Neuropsychopharmacology*. 2016;41(1):275–296.

Conio B, Martino M, Magioncalda P, et al. Opposite effects of dopamine and serotonin on resting-state networks: review and implications for psychiatric disorders. *Mol Psychiatry*. 2020;25(1):82–93.

Drevets WC, Wittenberg GM, Bullmore ET, Manji HK. Immune targets for therapeutic development in depression: towards precision medicine. *Nat Rev Drug Discov*. 2022;21(3):224–244.

Dunlop K, Talishinsky A, Liston C. Intrinsic brain network biomarkers of antidepressant response: a review. *Curr Psychiatry Rep*. 2019;21(9):87.

Dutta A, McKie S, Downey D, et al. Regional default mode network connectivity in major depressive disorder: modulation by acute intravenous citalopram. *Transl Psychiatry*. 2019;9(1):116.

Gudayol-Ferré E, Peró-Cebollero M, González-Garrido AA, Guàrdia-Olmos J. Changes in brain connectivity related to the treatment of depression measured through fMRI: a systematic review. *Front Hum Neurosci*. 2015;9:582.

Homan P, Neumeister A, Nugent AC, Charney DS, Drevets WC, Hasler G. Serotonin versus catecholamine deficiency: behavioral and neural effects of experimental depletion in remitted depression. *Transl Psychiatry*. 2015;5(3):e532.

Kautzky A, Baldinger P, Souery D, et al. The combined effect of genetic polymorphisms and clinical parameters on treatment outcome in treatment-resistant depression. *Eur Neuropsychopharmacol*. 2015;25(4):441–453.

Liu G-L, Cui Y-F, Lu C, Zhao P. Ketamine a dissociative anesthetic: neurobiology and biomolecular exploration in depression. *Chem Biol Interact*. 2020;319:109006.

Lucido MJ, Bekhbat M, Goldsmith DR, et al. Aiding and abetting anhedonia: impact of inflammation on the brain and pharmacological implications. *Pharmacol Rev*. 2021;73(3):1084–1117.

Maletic V, Raison C. Integrated neurobiology of bipolar disorder. *Front Psychiatry*. 2014;5:98.

Miller AH, Raison CL. The role of inflammation in depression: from evolutionary imperative to modern treatment target. *Nat Rev Immunol*. 2016;16(1):22–34.

Nestler EJ. Role of the brain's reward circuitry in depression: transcriptional mechanisms. *Int Rev Neurobiol*. 2015;124:151–170.

Northoff G, Hirjak D, Wolf RC, Magioncalda P, Martino M. All roads lead to the motor cortex: psychomotor mechanisms and their biochemical modulation in psychiatric disorders. *Mol Psychiatry*. 2021;26(1):92–102.

Phillips ML, Chase HW, Sheline YI, et al. Identifying predictors, moderators, and mediators of antidepressant response in major depressive disorder: neuroimaging approaches. *Am J Psychiatry*. 2015;172(2):124–138.

Piguet C, Karahanoğlu FI, Saccaro LF, Van De Ville D, Vuilleumier P. Mood disorders disrupt the functional dynamics, not spatial organization of brain resting state networks. *Neuroimage Clin*. 2021;32:102833.

Plante DT. The evolving nexus of sleep and depression. *Am J Psychiatry*. 2021;178(10): 896–902.

Posner J, Cha J, Wang Z, et al. Increased default mode network connectivity in individuals at high familial risk for depression. *Neuropsychopharmacology*. 2016;41(7):1759–1767.

Pourhamzeh M, Moravej FG, Arabi M, et al. The roles of serotonin in neuropsychiatric disorders. *Cell Mol Neurobiol*. 2022;42(6):1671–1692.

Schmaal L, Veltman DJ, van Erp TGM, et al. Subcortical brain alterations in major depressive disorder: findings from the ENIGMA major depressive disorder working group. *Mol Psychiatry*. 2016;21(6):806–812.

Siegle GJ, Thompson WK, Collier A, et al. Toward clinically useful neuroimaging in depression treatment: prognostic utility of subgenual cingulate activity for determining depression outcome in cognitive therapy across studies, scanners, and patient characteristics. *Arch Gen Psychiatry*. 2012;69(9):913–924.

Simmons WK, Burrows K, Avery JA, et al. Appetite changes reveal depression subgroups with distinct endocrine, metabolic, and immune states. *Mol Psychiatry*. 2020;25(7):1457–1468.

Smart C, Strathdee G, Watson S, Murgatroyd C, McAllister-Williams RH. Early life trauma, depression and the glucocorticoid receptor gene–an epigenetic perspective. *Psychol Med*. 2015;45(16):3393–3410.

Tafet GE, Nemeroff CB. The links between stress and depression: psychoneuroendocrinological, genetic, and environmental interactions. *J Neuropsychiatry Clin Neurosci*. 2016;28(2):77–88.

Thase ME. Depression and sleep: pathophysiology and treatment. *Dialogues Clin Neurosci*. 2006;8(2):217–226.

Thase ME. Using biomarkers to predict treatment response in major depressive disorder: evidence from past and present studies. *Dialogues Clin Neurosci*. 2014;16(4):539–544.

Zhou H-X, Chen X, Shen Y-Q, et al. Rumination and the default mode network: meta-analysis of brain imaging studies and implications for depression. *Neuroimage*. 2020;206:116287.

14

Anxiety Disorders

14.1 Anxiety Disorders: Introduction and Overview

KATHLEEN RIES MERIKANGAS, PH.D.

The anxiety disorders are the most prevalent psychiatric syndromes in the U.S. population. Nearly one-fifth (17%) of adults report a lifetime history of one of the major anxiety disorders, and 1 in 10 suffer from a current anxiety disorder. Anxiety disorders are associated with intense subjective distress and social impairment, and their childhood and adolescent onset can interfere with the accomplishment of social, educational, and occupational functioning. The most recent estimates of the Global Burden of Anxiety reveal that anxiety disorders explain 10% of the disability-adjusted life years for all mental, neurologic, and substance use disorders, second only to major depression. Despite their high burden, anxiety disorders have the lowest proportion of professional treatment of all of the classes of mental disorders. The sections of this chapter document the phenomenology and classification of anxiety disorders in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), epidemiology, risk factors and impact, biologic correlates including biochemical, neuroimaging, and genetics factors, and somatic and behavioral treatments.

BACKGROUND ON FEAR AND ANXIETY

Fear and anxiety can be conceptualized as two key core negative emotions. As such, definitions of “fear” and “anxiety” must be based on broader definitions of emotions in general. For the current set of sections, the term “emotion” refers to the brain state associated with the perception of a motivationally salient stimulus, a stimulus that creates a need for the organism to act. Motivational stimuli can, in turn, be divided into “rewards” and “punishments,” where a “reward” refers to a stimulus for which an organism will expend effort to approach, and a “punishment” refers to a stimulus for which an organism will expend effort to avoid.

“Fear” refers to the specific set of emotions or brain states that are elicited in an organism when it confronts danger. Many current approaches refer to “fears” in the plural because “fears” are viewed as a collection of closely related but distinguishable brain states, as opposed to one single state. This view reflects the fact that different forms of danger elicit different neural responses and associated differences in information processing and behavior: The response to a conspecific or some other form of innately dangerous stimulus can be distinguished, and both of these can, in turn, be distinguished from the response to objects that are viewed as dangerous through the effects of learning from experience. The term “danger,” in turn, refers to any stimulus or situation that is capable of producing harm to the organisms. The act of encountering a specifically dangerous object, such as a predator, can also be conceptualized as a “threat.” Moreover, because organisms tend to exert effort to avoid virtually all threats or dangerous scenarios, “threats” and “dangerous scenarios” can also be conceptualized as “punishments.”

The term “fear” can be distinguished from the term “anxiety” in a few respects. “Fear” is typically taken to refer to brain states associated with presentations of overtly dangerous stimuli. For example, fear is the emotion aroused when an organism immediately confronts a predator or a dangerous conspecific appearing in close proximity directly in front of the organism. As such, “fear” represents an acute, immediate reaction manifest in abrupt encounters with highly salient threats. In research with rodents, this definition has, at times, been used to model “normal” responses in humans to dangers. For example, this type of “fear” response in the rodent is often modeled with the classic conditioning paradigm in which a rodent is placed in a cage and exposed to a cue, such as a tone or a light, that predicts the immediate occurrence of an unconditioned aversive stimulus (UCS), such as a shock. In this fear-conditioning paradigm, the cue is conceptualized as a conditioned stimulus (CS+), and the organism's response to the cue can be viewed as an analog of "fear."

Unlike "fear," from this perspective, "anxiety" refers to brain states elicited by signals that predict impending but not immediately present danger. Thus, unlike "fear," "anxiety" involves a more sustained change in the brain, manifesting when a threat is still relatively removed from the organism in a spatial or temporal context. For example, "anxiety" might be conceptualized as the emotion aroused when an organism enters a deserted, isolated context where an encounter with a predator appears likely, despite the isolated nature of the context and the absence of any overt cues indicating that the predator is lurking. In the fear-conditioning experiment, "anxiety" is considered an analog of pathologic reactions to danger in humans, as occurs in anxiety disorders. This view reflects the relatively pervasive, chronic state of negative affect in clinical syndromes. In rodent experimental studies using fear conditioning, "anxiety" can be contrasted with "fear," which manifests on the immediate confrontation of the CS+; "anxiety" typically is considered as manifesting in the general context of the fear-conditioning experiment. Thus, anxiety might be elicited when the organism is placed in the cage where the fear-conditioning experiment was previously conducted.

Thus, "fear" and "anxiety" have been distinguished on both temporal and spatial grounds. From the temporal perspective, fear refers to the acute, immediate response to suddenly appearing, imminent danger, whereas anxiety refers to the sustained, insidious response to danger, as might manifest when the degree to which a threat is present remains ambiguous. From the spatial perspective, fear refers to the acute response to proximal threats, whereas anxiety refers to the response to distal threats. As such, "fear" in rodents can be considered an analog of "normal" human emotions elicited by thoughts about danger. "Anxiety" can be considered an analog of "abnormal" human emotions in similar scenarios. Parenthetically, when an acute, proximal threat is particularly dangerous, however, the emotional state elicited in the organism might better be characterized as "panic" as opposed to "fear."

Conceptualizations of "normal" and "abnormal" fears and anxieties emerging from research directly among humans differ somewhat from conceptualizations in rodent-based or other basic research. Definitions in humans rest heavily on the presence of impairment, a disruption in normal functioning, or the presence of "clinically significant" distress. Basing definitions of abnormal anxiety on impairment or clinical judgments about distress does create some problems for attempts to conduct cross-species and neuroscience-focused research. The degree to which impairment or distress might manifest can be heavily influenced by the environmental circumstances to which the individual is exposed. Thus, two individuals manifesting identical neural responses to a dangerous stimulus might be classified differently in terms of the degree to which their associated anxiety can be considered "normal." If one individual faces hurdles not encountered by the other due to the nature of their environment and associated resources, these two individuals, who function identically from a neuroscience perspective, will be classified differently from a clinical perspective.

Thus, as delineated in the chapters pertaining to etiology, particularly those on neurophysiology, neurochemistry, and brain circuitry, the field is entering a period in which ongoing research in basic brain function and its cross-talk with the environmental context and stimuli will enhance our understanding of the pathophysiology of disturbances in fear and anxiety. This knowledge has major future implications in its future potential for treatment and prevention.

DSM-5 MODIFICATIONS IN ANXIETY DISORDERS

The major subtypes of anxiety disorders in the DSM-5 include panic disorder (with or without agoraphobia), agoraphobia (without a history of panic disorder), specific phobia, social phobia, and generalized anxiety disorder (GAD). Revisions to the classification of anxiety disorders in the DSM-5 involve removing obsessive-compulsive disorder, and post-traumatic disorder have been subsumed under newly created "obsessive-compulsive and related disorders" and "trauma- and stressor-related disorders" categories, respectively. Therefore, both obsessive-compulsive disorder and post-traumatic stress disorder are not considered in this chapter. Other modifications to the proposed DSM-5 anxiety disorders category include the addition of separation anxiety disorder (contained under disorders usually first diagnosed in infancy, childhood, or adolescence in the DSM-IV), the identification of agoraphobia as a distinct and codable disorder (diagnosed only with reference to panic disorder in the DSM-IV), minor revisions to criterion language to enhance clarity, objectivity, and consistency across the anxiety disorders, and the relabeling of social phobia as social anxiety disorder (SAD). As such, the term “social phobia” will now be replaced by “social anxiety disorder.”

ASSESSMENT OF ANXIETY

Anxiety may be assessed at the symptom or disorder level. Among adults, the most commonly used symptom measures include self-report checklists of both state and trait anxiety, fears, phobias, and avoidance, as well as clinician-administered symptom checklists. Anxiety disorders among adults are assessed by structured and semistructured diagnostic interviews designed for administration by either lay interviewers (e.g., the Composite International Diagnostic Interview [CIDI], the Diagnostic Interview Schedule for DSM-IV [DIS-IV]) or experienced clinicians (e.g., the Structured Clinical Interview for DSM Disorders [SCID], the Anxiety Disorder Interview Schedule-Revised [ADIS-R]). The ADIS-R is the most comprehensive diagnostic interview for the assessment of the diagnostic criteria of anxiety and affective disorders.

Assessment of symptoms and disorders of anxiety among children can involve obtaining information from the child alone or, more routinely, from individuals who observe the child in multiple contexts (i.e., caregivers and/or teachers), in addition to the child themself. There are also several structured and semistructured diagnostic interviews of child anxiety disorders, including the Diagnostic Interview Schedule for Children (NIMH-DISC-IV), the CIDI modified for adolescents, the Schedule for Affective Disorders and Schizophrenia for School-Age Children (KSADS-PL); and the Anxiety Disorders Interview Schedule for Children (ADIS-IV-C and P).

Several psychophysiological indicators of anxiety have also been used to examine both human and animal anxiety. Experimental models which induce stress and measure autonomic output to test the human “fight or flight” response to threat have been used to study the range of triggers, correlates, and responses to fear-provoking situations. Behavioral tasks, such as giving a speech or receiving the threat of shock, have been used to experimentally induce anxiety states in unaffected individuals as well as those with ongoing anxiety disorders. Measures of changes in pulse, skin conductance, heart rate, and temperature regulation, as well as self-reported observations of facial expression, blushing, and other overt signs of anxiety, are presumed to provide a more accurate depiction of anxiety than self-reports or interviews about typical response patterns to stress.

OVERVIEW OF ANXIETY DISORDERS SECTION

The Anxiety Disorders section of this textbook covers the topics of phenomenology, epidemiology, genetics, neuroimaging, neurochemical aspects, and treatment of anxiety disorders.

In Chapter 14.2, Ryan J. Kimmel and Peter Roy-Byrne delineate the key **Clinical Features of the Anxiety Disorders**. The authors keep in line with the changes in the DSM-5. The DSM-5 removed post-traumatic stress disorder and obsessive-compulsive disorder but continues to link panic, generalized anxiety, and phobic anxiety together. This section summarizes the key symptoms and discusses differential diagnosis as it pertains to delineating presentations that appear to occur in multiple disorders. The authors provide the historical background and possible implications of the changes (or the lack thereof) to each disorder in the DSM-5.

Chapter 14.3 on the **Epidemiology of Anxiety Disorders** by Kathleen R. Merikangas and Michelle Theodory provides an update on the magnitude of anxiety disorders in adults and children in community surveys across the world. They show anxiety disorders are the most prevalent of the classes of mental disorders across the world and are associated with substantial individual and societal impact. They also document the large gap between the high magnitude of impairing anxiety disorders and low mental health treatment rates remains a serious challenge in public health, despite the efficacy of behavioral and pharmacologic treatments.

Chapter 14.4 provides an extensive summary by Costi, Feder, Iacoviello, Murrough, and Charney of the **Neurobiology and Neuroscience of Anxiety Disorders.** They summarize evidence from preclinical and clinical research that has implicated a range of neural systems through interactions with cortical and subcortical neural circuits involved in fear responses in the etiology of fear and anxiety responses. They review the role of neurochemical systems involved in anxiety disorders and their treatment targets, including the norepinephrine and serotonin systems, but also neurosteroids, opioid peptides, neuropeptide Y (NPY), and other neurochemicals involved in the modulation of anxiety. They further describe the growing number of studies that investigate target-selective agents that may enhance the benefits of exposure-based or other forms of psychotherapy. They highlight the importance of continued close collaboration between basic and clinical researchers required to advance the current understanding of the neurobiology of anxiety disorders.

In Chapter 14.5, Drs Kitt, Cohodes, and Gee present a detailed review of **Neuroimaging and the Neuroanatomical Circuits in Anxiety Disorders**. They provide an excellent summary of the prevailing models of anxiety disorders and their relevance to each of the anxiety disorders. They show that the application of neuroimaging technology to investigations of the neurobiologic basis of anxiety disorders is yielding seminal information about the functional anatomical and neurochemical correlates of both normal and pathologic anxiety and fear states. Neuroimaging data from studies of panic disorder, specific phobias, SAD, and GAD support disease models that depend on altered interactions among the amygdala, hippocampus, and medial and orbital prefrontal cortex (PFC). They further highlight the growing effort to delineate the neurodevelopmental processes associated with anxiety disorders in childhood and adolescence and characterize the heterogeneity of anxiety states through approaches that integrate imaging data with genetic and neurochemical information.

A review of the current state of **Cognitive-Behavioral Therapy of Anxiety Disorders** is presented by Jonathan D. Huppert and Edna B. Foa (CBT) in Chapter **14.6**. There is evidence for the efficacy of CBT programs in the treatment of all anxiety disorders across children, adolescents, adults, and geriatric populations. Exposure in vivo to feared but otherwise safe stimuli and imaginal exposure to unwanted and upsetting thoughts are essential components of most CBT programs for anxiety. The cessation of safety behaviors or other subtle avoidance behaviors is essential for the anxiety disorders in order to disconfirm the patients' erroneous perceptions. The addition of interoceptive exposure appears to make a significant contribution to the treatment of panic disorder. In addition, exposure is typically conducted in the presence of cognitive challenging or discussions, helping the patient to attend to the corrective information. Anxiety management training and cognitive therapy have also been shown to be effective in the treatment of GAD. Cognitive therapy that includes behavioral experiments, which inherently involves in vivo exposure, is also an effective treatment for panic disorder and SAD. Overall, it seems that the integration of exposure therapy with discussions that increase the attention of patients to the disconfirmation of their expectations of negative outcomes to exposures to the feared stimuli is the best method for optimizing treatment outcomes.

In Chapter **14.7**, Drs Ravindran and Stein review the current evidence for the application and evidence for specific **Somatic Therapies for Anxiety Disorders.** Aside from reviewing evidence for specific pharmacologic agents in treating each of the subtypes of anxiety disorders, they also highlight the importance of the diagnostic process and comprehensive evaluation of comorbidity, laboratory studies, and contextual life factors in treatment decision making. Physical examination and laboratory tests contribute to a comprehensive evaluation. They further describe methods to maintain treatment adherence, minimize side effects, and manage discontinuation of these medications when indicated. They also highlight the growing role of the specialty psychiatrist in consulting with primary care about the treatment of anxiety disorders.

In summary, these sections provide an update on the current state of knowledge on the magnitude, clinical presentations, impact, etiologic factors, and treatments of anxiety disorders. There has been substantial progress in each of these domains since the last edition of this textbook. However, the challenges of etiologic and phenotypic heterogeneity, lack of validity of diagnostic thresholds that distinguish the disorder from basic human emotions of fear and anxiety, indistinct boundaries within specific anxiety subtypes and between anxiety disorders, and mood and other emotional and behavior disorders continue to complicate our understanding of their causal mechanisms.

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. 2013.

Dunsmoor JE, Paz R. Fear generalization and anxiety: behavioral and neural mechanisms. *Biol Psychiatry*. 2015;78(5):336–343.

Ganella DE, Kim JH. Developmental rodent models of fear and anxiety: from neurobiology to pharmacology. *Br J Pharmacol*. 2014;171(20):4556–4574.

Gilmartin MR, Balderston NL, Helmstetter FJ. Prefrontal cortical regulation of fear learning. *Trends Neurosci*. 2014;37(8):455–464.

Morrison FG, Ressler KJ. From the neurobiology of extinction to improved clinical treatments. *Depress Anxiety*. 2014;31(4):279–290.

Patel V, Chisholm D, Parikh R, et al. Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities, 3rd edition. *Lancet*. 2016;387(10028):1672–1685.

Tallot L, Doyère V, Sullivan RM. Developmental emergence of fear/threat learning: neurobiology, associations and timing. *Genes Brain Behav*. 2016;15(1):144–154.

14.2 Clinical Features of the Anxiety Disorders

RYAN J. KIMMEL, M.D., JOSEPH M. CERIMELE, M.D., M.P.H., AND PETER ROY-BYRNE, M.D.

With the shift from anxiety neurosis in early conceptualizations of anxiety, three different spheres of anxiety have been delineated by clinicians: phobias (context-dependent fears), generalized anxiety (context-independent fears), and panic. Research that began in the 1960s suggested that obsessive-compulsive, phobic, nonspecific anxiety, and panic symptom subsets could be grouped together in part due to their common, positive response to various antidepressants. In the 1970s, interest in postcombat stress reactions rose as veterans returned from the Vietnam War. The cumulative research from the preceding decades led the authors of the DSM-III to organize an “Anxiety Disorders” section that included phobic disorders, generalized anxiety disorder (GAD), panic disorder, obsessive-compulsive disorder (OCD), and post-traumatic stress disorder (PTSD).

Though the DSM-IV refined the existing criteria and added a few new diagnoses, its overarching organization of the “Anxiety Disorders” section did not depart from the DSM-III nomenclature. The DSM-5, on the other hand, divided the DSM-IV’s anxiety spectrum into three pieces by separating out PTSD and OCD. Moreover, the DSM-5 asserted a close relationship between the classical anxiety disorders and a number of diagnoses that had been in other, nonanxiety classifications. Thus, the DSM-5’s “Anxiety Disorders” section contains separation anxiety disorder, selective mutism, specific phobia, social anxiety disorder, panic disorder, agoraphobia, GAD, substance/medication-induced anxiety disorder, anxiety disorder due to another medical condition, other specified anxiety disorder, and unspecified anxiety disorder.

The DSM-5 subsumed PTSD, reactive attachment disorder, disinhibited social engagement disorder, acute stress disorder, and adjustment disorder into a section entitled “Trauma- and Stressor-Related Disorders.” The final grouping is codified within the DSM-5 as the “Obsessive-Compulsive and Related Disorders” and includes OCD, body dysmorphic disorder, hoarding disorder, trichotillomania, and excoriation disorder.

With PTSD and OCD removed from the DSM-5 anxiety disorders section, modern nomenclature has made an interesting circuit. Emil Kraeplin, Anna Freud, John Bowlby, and other anxiety theorists from the prior century might recognize the logic behind the DSM-5’s decision to recircumscribe its “anxiety disorders,” gathering together panic, generalized anxiety, phobias, and separation anxiety. Every generation, however, has researched, argued, and reshuffled the anxiety spectrum to see if drawing new parallels and new divisions would yield improved insight and enhanced treatment efficacy.

Keeping in line with the current nosology, the following material describes the contents of DSM-5’s anxiety disorders section, whereas the trauma-related and obsessive-compulsive spectra are detailed in other chapters. Etiology and treatment are briefly referenced but are more extensively covered in other sections of this book. Differential diagnosis is discussed, especially as it pertains to delineating between presentations that appear to straddle multiple disorders in the DSM-5’s anxiety disorder grouping. However, given that anxiety-like symptoms appear in most of the diagnostic clusters in section II of the DSM-5, this section cannot provide an exhaustive differential diagnosis. The references included at the end of this section are a good source of cataloged confounders, as is the DSM-5 itself. Moreover, the reader should refer to the DSM-5 available in order to cross-reference the specific diagnostic criteria and their descriptions that could not be repeated verbatim in our text below.

PANIC DISORDER AND THE PANIC ATTACK SPECIFIER

Definition, Diagnosis, and Clinical Features

The DSM-5 defines a panic attack as a sudden, time-limited (minutes to an hour) burst of extreme anxiety accompanied by at least four of a cluster of cardiorespiratory, gastrointestinal, otoneurologic, autonomic, or cognitive symptoms (the specific symptoms are listed in Criterion A of Table 14.2–1). Culture-specific symptoms may be characteristic of panic attacks experienced by patients in other regions or communities, but typical panic symptoms are usually present, as well. Thus, the culture-specific symptoms do not supplant the need for four symptoms from the DSM-5's list of panic attack criteria.

A single panic attack is not necessarily indicative of a psychiatric disorder. More importantly, panic attacks can be a feature of a wide variety of psychiatric and medical diagnoses. Thus, DSM-5's Panic Attack Specifier may be added to other diagnoses to describe the experience and possible impairment in a given patient.

Panic disorder involves recurrent panic episodes accompanied by either a persistent fear of future panic attacks or a change in behavior designed to avoid panic triggers or situations. The behavior focuses on situations where the impact of a panic attack would be especially disconcerting or efforts to gain reassurance that nothing is wrong medically (by making ER visits or having repetitive medical testing) (see Table 14.2–1). At least some panic attacks have to "come out of the blue" without an obvious trigger, but many will be the result of a specific trigger or associated with situations where a specific trigger is anticipated.

Table 14.2–1.
Criteria for Panic Disorder

A. Recurrent unexpected panic attacks. A panic attack is an abrupt surge of intense fear or intense discomfort that reaches a peak within minutes, and during which time four (or more) of the following symptoms occur:

   Note: The abrupt surge can occur from a calm state or an anxious state.

   1. Palpitations, pounding heart, or accelerated heart rate.
   2. Sweating.
   3. Trembling or shaking.
   4. Sensations of shortness of breath or smothering.
   5. Feelings of choking.
   6. Chest pain or discomfort.
   7. Nausea or abdominal distress.
   8. Feeling dizzy, unsteady, light-headed, or faint.
   9. Chills or heat sensations.
   10. Paresthesias (numbness or tingling sensations).
   11. Derealization (feelings of unreality) or depersonalization (being detached from oneself).
   12. Fear of losing control or “going crazy.”
   13. Fear of dying.

   Note: Culture-specific symptoms (e.g., tinnitus, neck soreness, headache, uncontrollable screaming or crying) may be seen. Such symptoms should not count as one of the four required symptoms.

B. At least one of the attacks has been followed by 1 month (or more) of one or both of the following:
   1. Persistent concern or worry about additional panic attacks or their consequences (e.g., losing control, having a heart attack, “going crazy”).
   2. A significant maladaptive change in behavior related to the attacks (e.g., behaviors designed to avoid having panic attacks, such as avoidance of exercise or unfamiliar situations).

C. The disturbance is not attributable to the physiological effects of a substance (e.g., a drug of abuse, a medication) or another medical condition (e.g., hyperthyroidism, cardiopulmonary disorders).

D. The disturbance is not better explained by another mental disorder (e.g., the panic attacks do not occur only in response to feared social situations, as in social anxiety disorder; in response to circumscribed

> phobic objects or situations, as in specific phobia; in response to obsessions, as in obsessive-compulsive disorder; in response to reminders of traumatic events, as in posttraumatic stress disorder; or in response to separation from attachment figures, as in separation anxiety disorder).

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:235, 236.

The requirement that at least some panic attacks seem to be unexpected helps separate the definition of panic disorder from, for example, the definition of a specific phobia of heights with a panic specifier. While an exposure to heights may predictably trigger a panic attack, unless there are context-independent panic attacks, the patient would not meet criteria for panic disorder.

History and Comparative Nosology

Though the term “panic attack” is a 20th-century construct, descriptions of what is now referred to as “Panic Disorder” date back to prior centuries. The disorder was identifiable enough that several decades before DSM-III, psychopharmacology researchers were able to clearly identify patients with recurrent panic attacks who experienced therapeutic benefit in early studies of the efficacy of antidepressants.

The biggest debate over the last 20 years regarding the panic disorder diagnosis has been in its relation to agoraphobia. The DSM-IV contained diagnoses of panic disorder with agoraphobia and panic disorder without agoraphobia. This framed agoraphobia as a fear of public situations that developed as the result of panic attacks. However, the high prevalence of agoraphobia alone, without panic attacks, in community surveys worldwide led the DSM-5 committee to include agoraphobia distinctly from panic disorder. This newly independent agoraphobia diagnosis will be discussed later in this chapter.

Epidemiology

Successive iterations of the DSM have consistently distinguished single panic attacks from panic disorder because only about 1 in 25 people will meet criteria for panic disorder in their lifetime, whereas at least 1 in 7 (with some measurements closer to 1 in 3) will have a panic attack. Panic disorder has a high rate of comorbidity with affective disorders, substance use disorders, and other anxiety and trauma-related disorders. The age of onset of panic disorder peaks in the late teens and early 20s. Twice as many women as men are affected.

Differential Diagnosis, Etiology, Course, and Treatment

A variety of medical conditions may produce panic attack–like symptoms. These include brain tumors and epilepsy involving temporal lobe and adjacent subcortical structures; certain endocrinopathies such as hyperthyroidism and pheochromocytoma; and cardiopulmonary conditions causing chest pain, palpitations, or respiratory distress (e.g., myocardial infarction, cardiac arrhythmias, asthma, and pulmonary emboli). A number of medications can induce panic symptoms, including corticosteroids, stimulants, certain hormones, and some asthma medications. Illicit stimulants and hallucinogens can induce panic symptoms, as can caffeine and marijuana.

While family and twin studies clearly suggest that panic disorder has a heritable component (estimated at 35% to 40%), there is insufficient information to direct diagnosis or treatment. There is an increased frequency and intensity of stressful life events just prior to the onset of panic disorder, suggesting they serve as triggers of an underlying genetic vulnerability. Certain “anxious temperaments,” as well as childhood trauma, have been linked to higher rates of panic disorder in adulthood.

Panic disorder tends to be chronic, though the course is often fluctuating. Even after treatment to remission, the rate of relapse is high. For example, naturalistic studies often demonstrate a >50% relapse rate within 12 months of discontinuing an effective antidepressant. Given the impact of panic disorder on psychosocial functioning, as well as the increased rate of suicide attempts in these patients, ongoing monitoring is warranted if the patient chooses to disengage from active treatment.

Serotonergic antidepressants are the mainstay of pharmacologic treatment. Several benzodiazepines have FDA approval for panic disorder, and their efficacy and associated risks are discussed elsewhere in this book. There is also significant data to support treatment with cognitive behavioral therapy (CBT) and emerging, though still limited, evidence that other psychotherapies, such as psychodynamic and acceptance and commitment therapies, may be effective.

AGORAPHOBIA

Definition, Diagnosis, and Clinical Features

In DSM-5, agoraphobia is characterized by fear (imminent), anxiety (anticipated), and avoidance of multiple public situations where the individual feels that help is either not available or where escape would be difficult if they were to experience panic, anxiety, or other distressing or embarrassing symptoms. The fear or anxiety is “out of proportion” to the real threat, resulting in avoidance of the situation. The emotions (fear or anxiety) or behavior (avoidance) must persist for at least 6 months.

Patients with agoraphobia may develop “panic-like” symptoms, per DSM-5, without meeting full criteria for the panic attack specifier. Patients also may have other psychosomatic symptoms, often gastrointestinal or autonomic, associated with the fear or anxiety. Older individuals with multiple medical comorbidities may develop agoraphobia related to distressing symptoms of their medical condition (e.g., Parkinson disease). This combination of fear and psychosomatic symptoms leads to major life dysfunction. In some cases, patients may avoid all employment and in-person social interactions.

History and Comparative Nosology

In DSM-IV, agoraphobia was subsumed under panic disorder and thought only to develop in someone suffering from panic attacks. Clinicians thus had the option of coding for panic disorder with or without agoraphobia. However, more recent evidence has shown that many people with agoraphobia, perhaps over half, do not suffer from discernable panic attacks. Moreover, agoraphobia is not uncommon in children and adolescents who have never reported a panic attack. This difference has been shown to emerge from differences in clinical and nonclinical samples because many individuals with agoraphobia alone fail to seek treatment because of their extreme avoidance.

Epidemiology

Similar to panic disorder, more women than men have agoraphobia. The age of onset peaks in the late teens to early 20s. The measured prevalence of agoraphobia in specific clinical settings may evolve as the DSM-5’s recognition of agoraphobia without panic disorder spurs clinicians to screen and consider the disorder in patients who do not present with panic attacks. Other anxiety disorders are seen alongside agoraphobia in comorbidity rates that often exceed 50%. Comorbid depressive disorders are seen in 33% to 52% of cases, with some suggestion that the presence of comorbid panic attacks increases the risk of comorbid depressive episodes.

Differential Diagnosis, Etiology, Course, and Treatment

Although many of the other anxiety disorders are marked by severe anxiety and even panic, consideration of the specific context triggering panic–anxiety can help with differential diagnosis. Specific phobia is limited to one stimulus, whereas agoraphobia's diagnostic criteria require at least two situations. Determining the reason for avoidance can provide further clues. For example, a patient who is loath to walk across a bridge may be afraid of heights (specific phobia) or may perceive the bridge to be an exposed, open area with limited exits should some frightening event occur (agoraphobia). Similarly, patients with separation anxiety disorder may refuse to leave home because of fear of permanent separation from an attachment figure, patients with PTSD may avoid specific stimuli that trigger traumatic memories in certain community situations, patients with social anxiety disorder may prefer home to situations where they might be judged and embarrassed, and patients with OCD may be effectively homebound due to specific obsessions about contamination when outside. Thus, these disorders can usually be distinguished from agoraphobia by the breadth of the avoidance and the motivation for the behavior.

Twin studies have suggested that agoraphobia may be slightly more heritable than panic disorder. Similar to panic disorder, there are certain personality factors, such as anxiety sensitivity, that serve as risk factors. Agoraphobia tends to persist if not treated. It is presumed, based on the improvement in agoraphobia symptoms in DSM-IV era studies of panic disorder with agoraphobia, that the same pharmacotherapies that are efficacious for panic disorder would also be efficacious for agoraphobia. DSM-5's designation of agoraphobia as an independent diagnosis will help spawn prospective trials and retrospective, naturalistic studies that will test this assumption. In contrast to the pharmacotherapy literature, a number of studies already document the efficacy of behavioral treatment for agoraphobia independent of panic disorder.

SOCIAL ANXIETY DISORDER

Definition, Diagnosis, and Clinical Features

Social anxiety disorder is a fear of one or more social situations where for a period lasting longer than 6 months, an individual is concerned about being harshly judged and evaluated for their performance or interpersonal interactions. There is an out-of-proportion fear of being shamed or negatively evaluated. In many cases, the social situation(s) is avoided altogether or tolerated only with significant anxiety. Many people describe increased anxiety in social situations, but a key to the social anxiety diagnosis, per DSM-5, is a "significant impairment" in an important facet of life functioning.

Disabling performance anxiety when speaking in front of a crowd is a common example of social anxiety disorder, as is severe anxiety at having to meet and engage with new people. However, patients can also fear having to do simple, everyday tasks in any setting that could be loosely described as "social." That is, unlike speaking in front of a crowd, the feared situations in social anxiety disorder need not be ones where it would be common and reasonable to expect some level of judgment. An example would be a fear of writing one's signature in front of others out of concern that the manner of writing, no matter how ordinary, will be a source of derision. This fear may be compounded with anxiety about appearing anxious.

History and Comparative Nosology

The formulation of social anxiety disorder evolved from DSM-III to DSM-5. Prior classifications included distinguishing between "specific" and "generalized" social anxiety. Because most individuals with social anxiety disorder fear and avoid multiple social situations, the "specific" specifier was rarely used clinically and subsequently omitted from classification. The DSM-5 conceives the generalized pattern to be the norm, and the only specifier is "performance only," which is by far the most common type of circumstance-specific social anxiety.

The other nosology debate that has impacted this diagnosis over the last several decades is whether the old, "generalized" subtype of social anxiety disorder could be distinguished from avoidant personality disorder. At times, avoidant personality disorder has been considered a severe form of social anxiety disorder. While there is an unusually high rate of comorbidity between the two disorders as currently defined, it is notable that less than half of those with avoidant personality disorder have concurrent social anxiety disorder. Moreover, some theories align avoidant personality disorder closer to schizophrenia than social anxiety. The DSM-5 thus characterizes avoidant personality as phenomenologically distinct from the anxiety spectrum, though debates remain.

Epidemiology

Estimates of social anxiety disorder in the United States include a 12-month prevalence of 8% and a lifetime prevalence of 13%. Peak incidence occurs in the early teenage years (often in association with puberty and the transition from junior high school to high school or another increase in social context complexity). The prevalence is higher in women compared to men. Depressive episodes and substance use are common in those with social anxiety disorder.

Differential Diagnosis, Etiology, Course, and Treatment

Avoiding social interactions is central to social anxiety disorder but can also occur in other conditions, suggesting a need to consider a broad differential diagnosis. Alternate explanations for avoiding social interactions include other anxiety disorders such as panic disorder (differentiated by the presence of panic attacks and associated avoidance) or GAD (where the worry is broad), avoidant personality disorder (differentiated by avoidance of more than social interactions), development disorders such as autism spectrum disorders (where communication skills are more limited), and psychotic disorders (where insight is generally low). As mentioned above, avoidant personality disorder has been the subject of debate over whether it is distinct from social anxiety disorder, with the DSM-5 separating the two.

Individuals with social anxiety disorder have an elevated risk of comorbid mood and anxiety disorders. Commonly, individuals with social anxiety use nonprescribed anxiolytic and sedative medications and alcohol and cannabis for self-medication. The comorbidity between social anxiety disorder and selective mutism is discussed elsewhere in this chapter. Social anxiety disorder is frequently associated with autism spectrum disorder. Social anxiety disorder also occurs at rates above the population baseline in patients with schizophrenia.

Risk factors for social anxiety disorder include female gender, family history, and childhood signs of behavioral inhibition. There is insufficient data on specific genetic factors mediating the increased familial risk, but parenting styles may also contribute to this familiality. The Mini-Social Phobia Inventory (Mini-SPIN) is an appropriate screening tool for adults.

Agoraphobia symptoms can resemble social anxiety disorder or PTSD, though the context-dependent fears in the same situation would be different. For example, a patient with social anxiety disorder might dislike a crowded party due to feeling that many people are passing judgment. A patient with agoraphobia might avoid the same crowded party out of concern that, were they to develop anxiety, it would be difficult to sprint for the exit. A patient with PTSD might find their desire for hypervigilance is overwhelmed by the multiple stimuli in a crowded room.

The risk of suicide attempts is elevated in patients with social anxiety disorder. This, along with the significant functional impairment associated with this disorder, should motivate clinicians to initiate and monitor treatment for these patients. CBT and IPT have demonstrated efficacy for social anxiety disorder, with greater effect from CBT. The first-line pharmacologic treatments are serotonergic agents, though PRN medications, especially for the performance subtype (beta-blockers), have been used effectively. Social anxiety disorder is a chronic condition with a high rate of symptom re-emergence after symptom remission is achieved via SSRIs, for example. However, durable remission, even after cessation of CBT and/or antidepressants, has been seen in a small proportion of patients.

SPECIFIC PHOBIA

Definition, Diagnosis, and Clinical Features

Specific phobias are fears of specific objects or situations, such as spiders or blood draws, that go beyond the true threat of the stimulus and cause avoidance and functional impairment (Table 14.2–2). To meet criteria, the individual must experience fear during nearly every confrontation with the phobic stimulus. The phobia, by rule, lasts longer than 6 months. Many individuals have multiple phobic stimuli, in which case each would be coded separately. ICD-10 has specific codes for phobias of animals, natural environment (heights, storms, water), fear of blood, fear of injections, fear of transfusions, fear of other medical care, fear of injury, fear of a specific situation (elevators, airplanes, enclosed spaces, etc.), and fear of other (anything that does not fit into one the above).

History and Comparative Nosology

Circumscribed phobias have long been recognized. They are amenable to nonpharmacologic interventions and, as such, have been central to the delineation of developmental theories and psychotherapy strategies for over a century. A significant development over the course of the DSM iterations has been the separation and diagnostic refinement of social anxiety disorder (formerly social phobia), agoraphobia (formerly subsumed under panic disorder), and specific phobia (formerly simple phobia). The number of subcategories of specific phobia, as well as the value of trying to group the various phobic stimuli, continues to evolve in successive DSM iterations. Determining how much distress and impairment is necessary to separate the range of normal from the range of pathologic remains a nosologic challenge.

Table 14.2–2.

Criteria for Specific Phobia

A. Marked fear or anxiety about a specific object or situation (e.g., flying, heights, animals, receiving an injection, seeing blood).
   Note: In children, the fear or anxiety may be expressed by crying, tantrums, freezing, or clinging.
B. The phobic object or situation almost always provokes immediate fear or anxiety.
C. The phobic object or situation is actively avoided or endured with intense fear or anxiety.
D. The fear or anxiety is out of proportion to the actual danger posed by the specific object or situation and to the sociocultural context.
E. The fear, anxiety, or avoidance is persistent, typically lasting for 6 months or more.
F. The fear, anxiety, or avoidance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
G. The disturbance is not better explained by the symptoms of another mental disorder, including fear, anxiety, and avoidance of situations associated with panic-like symptoms or other incapacitating symptoms (as in agoraphobia); objects or situations related to obsessions (as in obsessive-compulsive disorder); reminders of traumatic events (as in posttraumatic stress disorder); separation from home or attachment figures (as in separation anxiety disorder); or social situations (as in social anxiety disorder).

***Specify* if:**

Code based on the phobic stimulus:

300.29 (F40.218) Animal (e.g., spiders, insects, dogs).

300.29 (F40.228) Natural environment (e.g., heights, storms, water).

300.29 (F40.23x) Blood-injection-injury (e.g., needles, invasive medical procedures

Coding note: Select specific ICD-10-CM code as follows: **F40.230** fear of blood; **F40.231** fear of injections and transfusions; **F40.232** fear of other medical care; or **F40.233** fear of injury.

300.29 (F40.248) Situational (e.g., airplanes, elevators, enclosed places).

300.29 (F40.298) Other (e.g., situations that may lead to choking or vomiting; in children, e.g., loud sounds or costumed characters.

Coding note: When more than one phobic stimulus is present, code all ICD-10-CM codes that apply (e.g., for fear of snakes and flying, F40.218 specific phobia, animal, and F 40.248 specific phobia, situational).

Reprinted with permission from the American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:224, 225.

Epidemiology

Lifetime prevalence estimates in the United States are from data that is over 20 years old. At that time, the lifetime prevalence of specific phobia was estimated at approximately 9%. The highest prevalence occurs in the teenage years. Female gender and younger age were associated with increased risk of specific phobia. Most individuals experienced multiple fears.

Differential Diagnosis, Etiology, Course, and Treatment

Delineating between the context-dependent phobias can be challenging in clinical practice, specifically when eliciting a history from a patient. Though some patients develop a specific phobia only after a traumatic event involving a threatening phobic stimulus, these patients, by definition, do not develop the breadth of symptoms that meet criteria for PTSD. Similarly, patients can develop a specific phobia to an event following a panic attack, but panic disorder requires that at least some panic attacks that are unexpected rather than just a consequence of a known phobic stimulus. As articulated previously, agoraphobia and specific phobia may be difficult to distinguish if the phobic stimulus is in a specific, public location. However, patients with agoraphobia ought to fear multiple public situations. Moreover, patients with agoraphobia specifically fear the panic or shame that can develop in public situations.

Some specific phobias, such as animal phobias, have a unique age of onset, suggesting a nontraumatic etiology perhaps more based in evolutionary preparedness. Other specific phobias, such as blood–injury–illness phobia, have a unique physiology (i.e., bradycardia and hypotension as opposed to the usual tachycardia and hypertensive reaction) and a greater rate of familiality that suggest still another distinct etiology.

Twin and family studies have demonstrated that there is an elevated familial/genetic risk for phobic disorders in general but not for the type of specific phobia. For many phobias, environmental factors, such as trauma, may be more important. As with other anxiety disorders, cognitive factors and conditioning may also be involved in their development.

Many individuals experience a chronic course of specific phobia. However, effective treatments exist. Patients often recognize that their anxiety exceeds the true threat of the stimulus, and this insight can often motivate treatment seeking. The treatment of choice for specific phobia is a behavioral therapy series that includes exposure-based, systematic desensitization. New evidence suggests efficacy for exposure therapy delivery through virtual reality in some circumstances. There are small, placebo-controlled studies involving serotonergic antidepressants, but the data is very limited. A recent meta-analysis of trials augmenting exposure-based psychotherapy for anxiety disorders with D-cycloserine included n = 124 individuals with specific phobia and showed a small short-term augmentation effect independent of antidepressant treatment status across all trials, though it did not report subgroup results for specific phobia. The data requires ongoing interpretation and study replication in more individuals with specific phobia to avoid harm via administration of this medication in the wrong context. There is limited long-term data on the durability of remission in pharmacologic studies.

GENERALIZED ANXIETY DISORDER

Definition, Diagnosis, and Clinical Features

GAD is characterized by an uncontrollable worry, more days than not, lasting greater than 6 months, and causing significant impairment. Accurate classification of GAD includes the worry being associated with at least three of the following six symptoms: “restlessness or feeling keyed up or on edge; being easily fatigued; difficulty concentrating or mind going blank; irritability; muscle tension; or sleep disturbance” (Table 14.2–3). All three symptoms need not be present simultaneously every day. The worry usually involves several to many domains of everyday life, such as usual daily activities, timeliness, finances, or health. These are issues that may concern the average person on a daily or weekly basis. However, patients with GAD tend to exaggerate the likelihood and severity of the stressor to the point where negative outcomes or catastrophe seems possible, likely, and imminent. Another key to the GAD diagnosis is that individuals have difficulty prioritizing more pressing matters that pop up on a day-to-day basis over other worries and concerns. This inappropriate prioritization of certain anxieties contributes to the pathologic effect that GAD has on psychosocial functioning.

Table 14.2–3.
Criteria for Generalized Anxiety Disorder

A. Excessive anxiety and worry (apprehensive expectation), occurring more days than not for at least 6 months, about a number of events or activities (such as work or school performance).

B. The individual finds it difficult to control the worry.

C. The anxiety and worry are associated with three (or more) of the following six symptoms (with at least some symptoms having been present for more days than not for the past 6 months):

   Note: Only one item is required in children.

   1. Restlessness or feeling keyed up or on edge.
   2. Being easily fatigued.
   3. Difficulty concentrating or mind going blank.
   4. Irritability.
   5. Muscle tension.
   6. Sleep disturbance (difficulty falling or staying asleep, or restless, unsatisfying sleep).

D. The anxiety, worry, or physical symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

E. The disturbance is not attributable to the physiological effects of a substance (e.g., a drug of abuse, a medication) or another medical condition (e.g., hyperthyroidism).

F. The disturbance is not better explained by another mental disorder (e.g., anxiety or worry about having panic attacks in panic disorder, negative evaluation in social anxiety disorder [social phobia], contamination or other obsessions in obsessive-compulsive disorder, separation from attachment figures in separation anxiety disorder, reminders of traumatic events in post-traumatic stress disorder, gaining weight in anorexia nervosa, physical complaints in somatic symptom disorder, perceived appearance flaws in body dysmorphic disorder, having a serious illness in illness anxiety disorder, or the content of delusional beliefs in schizophrenia or delusional disorder).

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:250, 251.

History and Comparative Nosology

The definition and diagnostic criteria of GAD have been iteratively changed over time since its inclusion in DSM-III. Alterations have ranged from broadly theoretical, such as whether GAD is a diagnosis of exclusion for patients whose anxiety does not seem to meet criteria for another anxiety disorder, to empirically detailed, such as how many concurrent symptoms are necessary over what duration of time. The high rate of comorbidity of GAD with other anxiety disorders and major depressive disorder makes it difficult to isolate GAD for the purposes of evaluating the fidelity of the criteria, limits tracking the durability of the diagnosis prospectively, and complicates evaluation of treatment effects. In the last 15 years, the pace of the nosologic evolution of GAD has slowed dramatically, as evidenced by the diagnostic agreement between DSM-IV and DSM-5. This may reflect a lack of consensus on proposed changes to the diagnostic criteria and/or indicate an awareness of the value of maintaining nosologic stability for the purposes of treatment efficacy research.

Epidemiology

The estimated lifetime prevalence of GAD is approximately 5%. Incidence is elevated in early adulthood and older adulthood. GAD is more common in women than in men. The high lifetime psychiatric comorbidity has led some to view GAD as a prodromal or residual phase of a major depression, though there is insufficient empirical support for this view. Other common comorbidities include a broad range of anxiety spectrum disorders and substance use disorders. Clinical epidemiology commonly involves presentation in primary care settings, often with physical symptoms.

Differential Diagnosis, Etiology, Course, and Treatment

For some individuals, GAD distinguishes itself from other anxiety diagnoses with its relatively high severity and persistent symptoms rather than the episodic course seen in the more context-dependent anxiety disorders. However, GAD may still fluctuate in severity over time. The biggest challenge, given the comorbidity between GAD and major depressive disorder, can be distinguishing whether an individual’s high anxiety during a major depressive episode represents a symptom of that individual’s depression or whether they have an independent, comorbid disorder. The easiest way to determine this would be to evaluate the anxiety during periods of euthymia, though this often requires a long period of observation.

Studies have identified genetic associations with GAD, though these are not specific enough to guide diagnosis or treatment. As with many other anxiety disorders, childhood maltreatment and abuse are associated with an increased incidence of GAD. The wide variety of neurologic, medical, medication, and chronic substance etiologies listed in the differential for panic disorder can also present with GAD-like symptoms. Beyond maintaining a high index of suspicion, presentations that include a marked change in intensity of GAD symptoms, a marked decline in functioning, an atypical age of symptom onset, or lack of response to standard treatments may prompt a broad medical assessment. The evolving criteria associated with successive DSM iterations complicates medication efficacy comparisons. For example, buspirone garnered FDA approval under a pre–DSM-IV set of GAD diagnostic criteria. Placebo-controlled efficacy trials of benzodiazepines, similarly, were studied in the DSM-III and pre–DSM-III eras. It would be difficult to replicate the depth and breadth of these trials with every new DSM. One of the original motivations to delineate generalized (context-independent) fears and panic was early data suggesting a difference in the efficacy of antidepressants between the two. However, over the last 30 years, a variety of antidepressant medications including tricyclics, SSRIs, and SNRIs, have demonstrated efficacy for both panic disorder and GAD.

Multiyear, naturalistic studies of GAD show that relapse is common. Factors associated with persistence of GAD include higher symptom severity, longer duration of avoidance behavior, and high behavioral inhibition. Thus, studies that follow patients for only a few months after cessation of an effective treatment modality may give clinicians and patients the false expectation of sustained remission. Clinical population-based strategies, such as those involving routine identification of individuals with GAD and subsequent symptom monitoring, stepped care strategies starting with monitoring and lower intensity treatments, treatment intensification for individuals not responding to initial steps in treatment, and case review by psychiatrists, are associated with significantly improved outcomes. The GAD-7 scale is a brief instrument validated for screening and treatment monitoring of individuals with GAD. Patients with GAD are at an increased risk for suicide attempts, and thus, screening for and responding to suicidal ideation is recommended.

Long-term studies of GAD symptom relapse rates following discontinuation of antidepressants after treatment response have led to recommendations that pharmacologic treatment continue for longer than 6 months, though longer follow-up studies are needed. Given that context-independent anxiety disorders have long been formulated as separate from panic and context-dependent phobias, there is a significant body of literature on psychotherapy for the GAD spectrum, including CBT, relaxation, imagery exposure, and various meditation modalities. Studies with follow-up after treatment completion longer than 12 months show that treatment with CBT is associated with better outcomes compared to control conditions.

SELECTIVE MUTISM

Definition, Diagnosis, and Clinical Features

Selective mutism is a disorder wherein children are unable to speak in certain social situations. These children are capable of speaking and do so when at home. Though the lack of speech at school is what often leads to an evaluation with a mental health provider, the disorder can be evident during other interactions between the child and less-familiar adults. The mutism must impact social or educational functioning for greater than a month in order to meet DSM-5's diagnostic criteria. The social situation wherein the symptoms manifest must not be one where a relative mutism might be expected, such as just after starting school or while integrating into a setting with a new language.

History and Comparative Nosology

In order to reframe the discussion of this disorder toward themes of anxiety and phobia, DSM-III's "Elective Mutism" term was switched to "Selective Mutism" in the DSM-IV, thus removing the implication that the symptoms were born of defiance. DSM-5's placement of selective mutism in the anxiety disorders section stems from evidence that anxiety (most often an element of social phobia) is often present. However, selective mutism's age of onset, as well as its heterogeneity in terms of comorbid social anxiety symptoms, developmental disorders, communication disorders, and many other symptom subsets, argue that it is a distinct diagnosis from social phobia.

Epidemiology

The rarity of selective mutism makes it a challenge to quantify, but most studies of children find a prevalence in the range of 1%. While some studies suggest a higher prevalence in girls, others suggest a more even distribution between boys and girls. Social anxiety disorder is likely the most common comorbidity, though social anxiety disorder has an average age of onset in the early teens, whereas selective mutism usually presents by age 5.

Differential Diagnosis, Etiology, Course, and Treatment

Communication disorders, autism spectrum, and developmental delay would be important etiologies to rule out when making the diagnosis of selective mutism. If there is a broader autism spectrum diagnosis with prominent communication symptoms that better explains the symptoms, the selective mutism diagnosis is not applied. Selective mutism should prompt further consideration and screening for comorbid anxiety disorders. A wide variety of genetic, temperament, environmental, and neurodevelopmental contributors have been suggested to contribute to the etiology of selective mutism.

CBT and behavioral therapies have shown efficacy in various studies. Programs that include parental and school-based interventions may have the most promise. These CBT studies also demonstrate that the course of this disorder need not be chronic. Following CBT interventions, long-term studies often report that >50% of children demonstrate full remission. However, long-term follow-up studies have also suggested that children and young adults who had selective mutism go on to demonstrate a higher-than-population-baseline rate of phobic disorders. Some SSRIs have demonstrated partial efficacy, though the studies are small, and there is a low rate of full remission. Overall, the field continues to make strides with regard to selective mutism, with the formulation of this diagnosis as an anxiety disorder appearing to accelerate the development of effective interventions.

SEPARATION ANXIETY DISORDER

Patients with separation anxiety disorder experience excessive distress when anticipating separation or when physically apart from figures of meaningful attachment. These patients ruminate on the possibility of any harm that might befall them or their loved ones. This worry then leads to the avoidance of activities outside of the presence of the attached figure. Patients can also fear sleep, manifesting a kind of vigilance to reduce the possibility of a separation event. Even when they do manage to sleep, patients with separation anxiety disorder often have nightmares regarding separation. Finally, the excessive worry can be associated with somatic symptoms of anxiety. Symptoms of separation anxiety disorder in children need only persist for 4 weeks to meet criteria. Symptoms must persist longer than 6 months to meet criteria for the adult-onset variant.

History and Comparative Nosology

The epidemiologic evidence documenting a high rate of onset occurring after 18 years old led to a removal of DSM-IV's age-of-onset restriction. The inclusion of the adult-onset formulation moved separation anxiety disorder from DSM-IV's "Disorders Usually First Diagnosed in Infancy, Childhood or Adolescence" to DSM-5's "Anxiety Disorders." The prominent symptom of worry ties separation anxiety disorder to the other classical diagnoses in the anxiety disorders section.

Epidemiology

In a recent worldwide study, the lifetime prevalence was 4.8%, with 43% of these patients experiencing onset after age 18. The prevalence of childhood-onset separation anxiety disorder is more common in girls than boys, but the female:male ratio in adult-onset separation anxiety disorder is closer to even. More than half of patients show a remission of their disorder within the first 10 years after onset. Over the course of the lifetime, separation anxiety disorder is associated with an increased risk of a broad range of comorbid mood and anxiety disorders, including major depressive disorder, bipolar disorder, panic disorder, OCD, social anxiety disorder, and specific phobia. The data on the relative rates of the various comorbidities will likely continue to evolve as DSM-5's recognition of the adult-onset variant dramatically expands the population cohort available for research participation. Prevalence data and improved screening tools for the adult-onset group are important because the presence of comorbid separation anxiety disorder predicts reduced efficacy for CBT in studies of panic disorder, GAD, and social phobia. CBT interventions may need to be specifically tailored to include addressing separation anxiety symptoms when this diagnosis is comorbid with other anxiety disorders.

Differential Diagnosis, Etiology, Course, and Treatment

Patients with separation anxiety disorder have emotional experiences and behaviors when out alone in public that overlap with agoraphobia, though the fear in separation anxiety disorder is that something bad will happen to the attachment figure during the separation. The fear in agoraphobia, on the other hand, involves one's own safety and is secondary to the unavailability of timely help or escape should a frightening or embarrassing event occur. Similarly, patients with separation anxiety disorder might experience panic-like symptoms, or even symptoms meeting full criteria for the panic attack specifier. However, panic disorder requires that some of the attacks be context independent and "out of the blue," whereas panic symptoms in separation anxiety disorder would have their origins in a very specific theme of separation. While fear about the well-being of a loved one could be a theme in GAD, the worries in GAD involve a broader spectrum of subject matters and situations, rather than the more circumscribed worry about separation in separation anxiety disorder. There are adult-onset courses of separation anxiety disorder that seem, at least initially, to be features of a complicated bereavement process, suggesting an overlap with either PTSD or complicated grief. Whether this yields treatment modalities that target trauma- or grief-related symptoms in patients with adult-onset separation anxiety disorder remains to be seen.

Separation anxiety disorder aggregates in families, though what portion of this is genetic and what portion may be due to environmental modeling is not clear. Various biologic correlates are being pursued, including studies of oxytocin receptor polymorphisms, serotonin transporter variants, platelet 18 kDa translocator protein density, and dopamine D4 receptor polymorphisms, to name a few. Interestingly, both separation anxiety disorder and panic disorder are associated with carbon dioxide sensitivity in adults.

The lifetime prevalence of traumatic experience is significantly associated with the lifetime prevalence of separation anxiety disorder. There are no large-scale, controlled intervention studies to guide treatment in adult-onset separation anxiety. Given the efficacy of CBT in childhood-onset separation anxiety disorder, this modality is usually the first intervention in the adult-onset variant. However, it is not known whether CBT interventions, or which facets of these interventions, are efficacious for the adult-onset variant. The data on pharmacologic interventions in adult-onset separation anxiety are still preliminary.

In childhood-onset separation anxiety disorder, studies show benefit for CBT, as well as for programs that include a component of education for parents and a focus on social skills. CBT strategies often have to be modified based on the age of the child, and there are few studies that include patients under the age of 7, for example. The data on pharmacotherapy for children with separation anxiety is mixed, with SSRIs sometimes employed for CBT nonresponders. While there are children who appear to be refractory to current treatment programs, the format and targets of CBT interventions for separation anxiety disorder continue to evolve and demonstrate improved efficacy.

SUBSTANCE/MEDICATION-INDUCED ANXIETY DISORDER AND ANXIETY DISORDER DUE TO ANOTHER MEDICAL CONDITION

Substance/medication-induced anxiety disorder can manifest as either generalized anxiety or panic. In order to meet criteria, there must be a temporal association between the symptoms and either intoxication or withdrawal from the substance or medication. In addition, the substance or medication in question must be known to be associated with the production of these symptoms. The DSM-5 contains documentation of specific ICD-10 codes for individual inciting agents, including alcohol, caffeine, cannabis, phencyclidine, other hallucinogen, inhalant, opioid, sedative/hypnotic/anxiolytic, amphetamine (or other stimulant), cocaine, and other (or unknown) substance. The “other” substances known to produce anxiety symptoms include a wide range of endocrine and neuromodulatory agents. The ICD-10 mandates a comment about whether there is a comorbid substance use disorder, as well as a comment on the severity of the substance use disorder. There are also coding specifiers regarding whether the symptoms were the result of intoxication, withdrawal, or medication initiation.

Ideally, the anxiety symptoms resolve with the removal of the offending agent or the resolution of the withdrawal. If the anxiety disorder persists for many months after the inciting incident, one may need to consider whether the patient would merit the diagnosis of a primary anxiety disorder. This point is interesting and part of a longer-lasting debate about whether, for example, there is a prolonged withdrawal syndrome from benzodiazepines that can last months to years. In general, there is not good evidence for toxic or withdrawal effects of any substance lasting for durations longer than the physiologic duration of the withdrawal and, perhaps, the duration required to cognitively and psychologically “readjust.”

Anxiety disorder due to another medical condition can similarly present with anxiety or panic, though the symptoms are temporally related to the onset or exacerbation of a diagnosed medical condition whose pathophysiology is known to yield anxiety symptoms. As was noted in the panic disorder section, there are numerous endocrine, metabolic, oncologic, cardiopulmonary, infectious, and neurologic etiologies that can manifest anxiety or panic symptoms.

FUTURE DIRECTIONS

The development of the nosology for anxiety disorders began with organized, direct observation. Differential responses to various psychotherapy and psychopharmacology interventions have contributed in part to our current delineation of the disorders, though there is substantial overlap between psychopharmacologic efficacy among the disorders, and even the specific CBT packages for distinct disorders may not be as necessary, with transdiagnostic (unified) CBT protocols now shown to have efficacy across disorders. Epidemiologic studies that help identify course, comorbidity, cultural differences, and biases have also helped shape the modern nomenclature. Biomarkers, genetic studies, gene–gene interactions, epigenetics, pharmacogenetics, neuroimaging, and the other developing tools of neuroscience have yet to be incorporated into successive iterations of the DSM but will hopefully contribute in the future to the increasingly detailed nosology of the anxiety disorders. It is interesting that, with the culling of the trauma-associated and obsessive-compulsive spectra, DSM-5’s anxiety disorders diagnostic grouping appears more neatly bound. However, the extraordinarily high rate of comorbidity between the component diagnoses of the anxiety disorders, trauma-associated disorders, and OCDs argues that anxiety disorder etiology and treatment should not be viewed in isolation. Moreover, anxiety symptoms, or at least end-symptom presentations that look strikingly similar to facets of the anxiety disorders, appear in nearly every diagnostic cluster of Section II of the DSM-5. Thus, while there is diagnostic and treatment research value in the demarcation of the anxiety disorders, there is also value in seeing the forest through the trees.

FURTHER READINGS

American Psychiatric Association. *Desk Reference to the Diagnostic Criteria From DSM-5*. American Psychiatric Publishing; 2013.

Batelaan NM, de Graaf R, Penninx BWJH, van Balkom AJLM, Vollebergh WAM, Beekman ATF. The 2-year prognosis of panic episodes in the general population. *Psychol Med*. 2010;40(1):147.

Bergman RL, Piacentini J, McCracken JT. Prevalence and description of selective mutism in a school-based sample. *J Am Acad Child Adolesc Psychiatry*. 2002;41(8):938–946.

Bögels SM, Alden L, Beidel DC, et al. Social anxiety disorder: questions and answers for the DSM-V. *Depress Anxiety*. 2010;27(2):168–189.

Bögels SM, Knappe S, Clark LA. Adult separation anxiety disorder in DSM-5. *Clin Psychol Rev*. 2013;33(5):663–674.

Bruce SE, Yonkers KA, Otto MW, et al. Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: a 12-year prospective study. *Am J Psychiatry*. 2005;162(6):1179–1187.

Donker T, Cornelisz I, van Klaveren C, et al. Effectiveness of self-guided app-based virtual reality cognitive behavior therapy for acrophobia: a randomized clinical trial. *JAMA Psychiatry*. 2019;76(7):682–690.

Ehrenreich JT, Santucci LC, Weiner CL. Separation anxiety disorder in youth: phenomenology, assessment, and treatment. *Psicol Conductual*. 2008;16(3):389–412.

Goodwin RD, Roy-Byrne P. Panic and suicidal ideation and suicide attempts: results from the National Comorbidity Survey. *Depress Anxiety*. 2006;23(3):124–132.

Hasin DS, Grant BF. The National Epidemiologic Survey on Alcohol and Related conditions (NESARC) Waves 1 and 2: review and summary of findings. *Soc Psychiatry Posychiatr Epidemiol*. 2015;50(11):1609–1640.

Hovenkamp-Hermelink JHM, Jeronimus BF, Myroniuk S, Riese H, Schoevers RA. Predictors of persistence of anxiety disorders across the lifespan: a systematic review. *Lancet Psychiatry*. 2021;8(5):428–443.

Hudson JL, Rapee RM, Lyneham HJ, McLellan LF, Wuthrich VM, Schniering CA. Comparing outcomes for children with different anxiety disorders following cognitive behavioural therapy. *Behav Res Ther*. 2015;72:30–37.

Kalin NH. Novel insights into pathological anxiety and anxiety-related disorders. *Am J Psychiatry*. 2020;177(3):187–189.

Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):593–602.

Kessler RC, Chiu WT, Jin R, Ruscio AM, Shear K, Walters EE. The epidemiology of panic attacks, panic disorder, and agoraphobia in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2006;63(4):415–424.

Kessler RC, McGonagle KA, Zhao S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry*. 1994;51(1):8–19.

Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H-U. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *Int J Methods Psychiatr Res*. 2012;21(3):169–184.

Klauke B, Deckert J, Reif A, Pauli P, Domschke K. Life events in panic disorder-an update on “candidate stressors”. *Depress Anxiety*. 2010;27(8):716.

Leichsenring F, Leweke F. Social anxiety disorder. *N Engl J Med*. 2017;376(23):2255–2264.

Mataix-Cols D, de la Cruz LF, Monzani B, et al; the DCS Anxiety Consortium. D-Cycloserine augmentation of exposure-based cognitive behavior therapy for anxiety, obsessive-compulsive, and posttraumatic stress disorders: a systematic review and meta-analysis of individual participant data. *JAMA Psychiatry*. 2017;74(5):501–510.

Mohatt J, Bennett SM, Walkup JT. Treatment of separation, generalized, and social anxiety disorders in youths. *Am J Psychiatry*. 2014;171(7):741–748.

Pennix BWJH, Pine DS, Holmes EA, Reif A. Anxiety disorders. *Lancet*. 2021;397(10277):914–927.

Rochester J, Baldwin DS. Adult separation anxiety disorder: accepted but little understood. *Hum Psychopharmacol*. 2015;30(1):1–3.

Rollman BL, Belnap BH, Mazumdar S, et al. Telephone-delivered stepped collaborative care for treating anxiety in primary care: a randomized controlled trial. *J Gen Intern Med*. 2017;32(3):245–255.

Silove D, Alonso J, Bromet E, et al. Pediatric-onset and adult-onset separation anxiety disorder across countries in the World Mental Health Survey. *Am J Psychiatry*. 2015;172(7):647–656.

Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006;166(10):1092–1097.

Stangier U, Schramm E, Heidenreich T, Berger M, Clark DM. Cognitive therapy vs interpersonal psychotherapy in social anxiety disorder: a randomized controlled trial. *Arch Gen Psychiatry*. 2011;68(7):692–700.

van Dis EAM, van Veen SC, Hagenaars MA, et al. Long-term outcomes of cognitive behavioral therapy for anxiety-related disorders. *JAMA Psychiatry*. 2020;77(3):265–273.

Wittchen HU, Gloster AT, Beesdo-Baum K, Fava GA, Craske MG. Agoraphobia: a review of the diagnostic classificatory position and criteria. *Depress Anxiety*. 2010;27(2):113–133.

Wittchen HU, Nocon A, Beesdo K, et al. Agoraphobia and panic. Prospective-longitudinal relations suggest a rethinking of diagnostic concepts. *Psychother Psychosom*. 2008;77(3):147.

14.3 Epidemiology of Anxiety Disorders

KATHLEEN RIES MERIKANGAS, PH.D., AND MICHELLE M. THEODORY, B.A.

Population-based studies across the world have shown that anxiety disorders are the most common class of mental disorders. Epidemiology has made a valuable contribution to our understanding of anxiety disorders by documenting the prevalence, correlates, and natural history of anxiety disorders and highlighting the gap between untreated anxiety disorders in the general population compared to cases seen in clinical settings. During the past few years, more valid estimates of the global burden of anxiety disorders have also become available.

The aims of this chapter are: (1) to describe the general methods of epidemiology that have contributed to our understanding of anxiety disorders; (2) to summarize the prevalence and correlates of DSM-IV anxiety disorders in adults and children from community samples; and (3) to describe the individual and societal impact of anxiety disorders.

BACKGROUND: PSYCHIATRIC EPIDEMIOLOGY

Epidemiology is the study of the distribution and determinants of diseases in human populations. Epidemiologic studies are concerned with the extent and types of illnesses in groups of people and with the factors that influence their distribution. Epidemiologists investigate the interactions that may occur among the host, agent, and environment to produce a disease state. Although the tools of epidemiology were developed to identify causes of infectious diseases, their application to chronic diseases such as cancer, heart disease, and mental disorders has also enhanced our understanding of their multifactorial etiology.

The ultimate goal of epidemiologic studies is to identify the *etiology* of a disease in order to prevent or intervene in the progression of the disorder. To achieve this goal, epidemiologic studies generally proceed from studies that specify the prevalence and distribution of a disease within a population by person, place, and time (*descriptive* epidemiology) to more focused studies of the determinants of disease in specific groups (*analytic* epidemiology). Descriptive epidemiologic studies are important in specifying the rates and distribution of disorders in the general population. Identification of differential distribution by sex, age, ethnicity, geographic site, or exposure to particular risk factors provides clues that may be tested systematically. Case-control studies generally proceed from retrospective designs and define cases and controls by the presence or absence of a disease in order to identify the disease's potential associations with a particular risk factor or set of risk factors. Cohort studies identify exposure to a putative risk factor in a group of people at baseline and then follow them retrospectively or prospectively to examine the differential incidence of the disease over time.

Community study data can be examined to identify biases that may exist in treated populations and to construct case registries from which persons may serve as probands for analytic epidemiologic studies. Such attention to sampling issues is a major contribution of the epidemiologic approach, as individuals identified in clinical settings often constitute the tip of the iceberg of the disease and may not be representative of the general population of similarly affected individuals with respect to demographic, social, or clinical characteristics.

Application of Epidemiology to Psychiatry

The traditional contributions of the application of the tools of epidemiology to psychiatry include (1) methodological developments, including the introduction of structured and semistructured diagnostic interviews; (2) estimation of the prevalence and correlates of mental disorders across the world; and (3) identification of risk factors for the development of mental disorders.

Community studies of mental disorders have also highlighted the need to expand the categorical diagnostic criteria for mental disorders to collect information on the spectra of expression of their underlying domains. For example, findings from the National Comorbidity Survey-Replication (NCS-R) demonstrated the validity of the spectrum concept of bipolar disorder that had been proposed by clinicians for more than a decade. Prospective follow-up of community samples of youth has also demonstrated the heterogeneity of different patterns of developmental manifestations of anxiety subtypes and their underlying dimensions of fear and anxiety. Therefore, the imposition of arbitrary thresholds in the contemporary categorical diagnostic system may lead to an inaccurate classification of these conditions. Likewise, the tendency for individuals to meet criteria for more than one condition both at the same time and across their lifetimes, or comorbidity, has been consistently demonstrated in community samples, highlighting the lack of distinct boundaries between classes of mental disorders, particularly anxiety and mood disorders.

One of the most important contributions of epidemiology to our understanding of mental disorders is the identification of the large number of individuals who suffer from one or more of these conditions yet have not received a clinical intervention. There is a growing gap between cases seen in specialty mental health settings and people with anxiety in the general community. As expected, clinical samples have a disproportionate representation of severe cases with multiple comorbid conditions and substantially greater impairment than those who do not receive mental health treatment. Below, we provide updated estimates of the prevalence of anxiety disorders.

MAGNITUDE OF ANXIETY DISORDERS

Adults

Anxiety disorders have emerged as the most prevalent mental disorders in both adult and child populations. The accumulating literature suggests that the prevalence of anxiety disorders in youth is similar, if not higher, to that reported in adults, thereby indicating the importance of a life-course approach to the study of anxiety. During the past decade, the results of several large-scale studies of cross-national and cross-cultural samples have emerged. The largest global collaboration to date is the aforementioned WHO World Mental Health 2000 Initiative (WMH2000), which encompasses a collection of nationally or regionally representative surveys on the prevalence, risk factors, correlates, and service patterns for mental disorders from over 25 countries, representing all regions of the world, with a total sample size exceeding 130,000 individuals. There has also been progress in descriptive epidemiology in countries outside of the United States and Europe that have been traditionally the sole source of population-based data on mental disorders. Below, we provide an update on the prevalence of both the aggregate rates of anxiety as well as those of specific anxiety disorders.

The *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)* criteria removed post-traumatic stress disorder and obsessive-compulsive disorder from the anxiety disorder section, so we limit our review to the major subgroups of anxiety disorders, including DSM-5 anxiety disorders. Separation anxiety disorder and selective mutism are now included as anxiety disorders, and the former's diagnostic criteria no longer specify the age of onset. Other major changes include the separate diagnoses of panic disorder and agoraphobia, the removal of the requirement that individuals over age 18 must recognize that their anxiety is excessive or unreasonable, and the extension of the 6-month duration criterion. However, because the DSM-5 was introduced during the past year, not all studies have updated prevalence rates based on the new criteria. Several studies suggest that modifications in the anxiety disorders section of the DSM-5 will not have a substantial impact on the prevalence rates of specific subtypes of anxiety.

Table 14.3–1 presents an overview of recent national and international community studies of *DSM-IV* and *DSM-V* anxiety disorders in adults.

Panic Disorder. The prevalence of DSM-5 panic disorder has been examined across the globe, and the rates remain fairly consistent across studies. Most studies yield 12-month prevalence rates between 0.2% and 1.1%. The highest rate was reported in the NESARC-III Study in the United States (3.1%), whereas the lowest 12-month prevalence was reported by the WMH in Nigeria (0.1%). The lifetime prevalence rates of panic disorder range from 0.2% in Nigeria to a high of 4.7% in the United States, with most studies falling between 1% and 2%.

Generalized Anxiety Disorder. Generalized anxiety disorder (GAD) has also been extensively examined in samples from several continents. Whereas the 12-month prevalence rate of GAD falls between 4.0% and 5.3% in both large-scale studies in the United States (WMH and NESARC-III), the rates in other areas of the world ranged from 0% in Nigeria to 4.3% in Spain (Murcia). The lifetime prevalence rates of GAD were even more variable, with the lowest rate reported in Nigeria (0.1%) and the highest in Australia (8.0%).

Agoraphobia. Apart from an elevated 12-month prevalence of agoraphobia reported in South Africa (4.8%), the rates of agoraphobia were remarkably consistent across regions. Nine separate international studies reported estimates ranging from 0% (China) to 0.8% (United States). The lifetime prevalence estimates were similarly consistent, with the lifetime prevalence of DSM-IV agoraphobia ranging anywhere from 0.2% (China) to 2.9% of the population (Brazil) and a median of 0.9%.

Table 14.3–1. Lifetime and 12-Month Prevalence Rates of DSM-IV and DSM-V Anxiety Disorders in International Community Studies of Adults

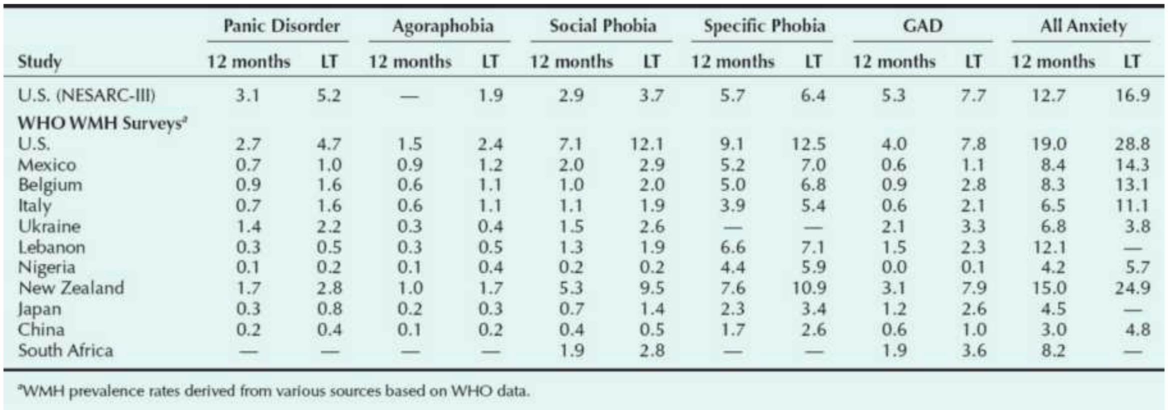

	Panic Disorder		Agoraphobia		Social Phobia		Specific Phobia		GAD		All Anxiety	
Study	12 months	LT	12 months	LT	12 months	LT	12 months	LT	12 months	LT	12 months	LT
U.S. (NESARC-III)	3.1	5.2	—	1.9	2.9	3.7	5.7	6.4	5.3	7.7	12.7	16.9
WHO WMH Surveys[a]												
U.S.	2.7	4.7	1.5	2.4	7.1	12.1	9.1	12.5	4.0	7.8	19.0	28.8
Mexico	0.7	1.0	0.9	1.2	2.0	2.9	5.2	7.0	0.6	1.1	8.4	14.3
Belgium	0.9	1.6	0.6	1.1	1.0	2.0	5.0	6.8	0.9	2.8	8.3	13.1
Italy	0.7	1.6	0.6	1.1	1.1	1.9	3.9	5.4	0.6	2.1	6.5	11.1
Ukraine	1.4	2.2	0.3	0.4	1.5	2.6	—	—	2.1	3.3	6.8	3.8
Lebanon	0.3	0.5	0.3	0.5	1.3	1.9	6.6	7.1	1.5	2.3	12.1	—
Nigeria	0.1	0.2	0.1	0.4	0.2	0.2	4.4	5.9	0.0	0.1	4.2	5.7
New Zealand	1.7	2.8	1.0	1.7	5.3	9.5	7.6	10.9	3.1	7.9	15.0	24.9
Japan	0.3	0.8	0.2	0.3	0.7	1.4	2.3	3.4	1.2	2.6	4.5	—
China	0.2	0.4	0.1	0.2	0.4	0.5	1.7	2.6	0.6	1.0	3.0	4.8
South Africa	—	—	—	—	1.9	2.8	—	—	1.9	3.6	8.2	—

[a]WMH prevalence rates derived from various sources based on WHO data.

Social Anxiety Disorder. The 12-month prevalence of social phobia varied across regions, with the highest rates reported in the United States in the WMH survey (7.1%) and the lowest appearing in Nigeria (0.2). However, it cannot be determined whether differences in these rates are the result of methodological or translation inconsistencies or true cross-cultural differences in defining social fears. Even within the United States, the 12-month prevalence dramatically differed, with the WMH study reporting 7.1% in comparison to the 2.9% prevalence reported from the NESARC-III study that was based on DSM-5. The estimates of the lifetime prevalence of *DSM-IV* social phobia were also inconsistent across studies, with a range spanning from 1.2% in East Mediterranean countries to a high of 9.5% in New Zealand and 12.1% in the United States.

Specific Phobia. Similar to the estimates reported for social anxiety disorder, the 12-month and lifetime prevalence rates of DSM-IV-specific phobia varied across studies. The 12-month prevalence estimates, according to the NCS-R, ranged from 1.9% (China) to 9.1% (United States). Estimates for lifetime-specific phobia were lowest in Italy (1.5%) and highest in New Zealand (10.8%), with a median of 6%. As with other disorders, the reasons for these inconsistencies need to be examined further in order to distinguish between true differences in prevalence and those attributable to methodological differences between studies.

Separation Anxiety Disorder. The *DSM-5* criteria no longer include an age-at-onset restriction because a substantial number of adults report onsets after age 18. The 12-month prevalence rate averaged 1.0% across 20 countries in the WMH survey. The lowest was 0.0% in Nigeria and Italy, and the highest was 2.7% in Colombia. The lifetime prevalence rate averaged 3.1%, and the rates ranged from 0.1% (Nigeria) to 7.9% (Colombia).

Children and Adolescents

Estimates of prevalence rates of anxiety disorders in children have been much more variable than those of adults because of substantial differences in study methods and developmental manifestations of anxiety and fear that may influence prevalence estimates. The chief differences include the age and sex composition of the samples, methods for assessing anxiety disorders (i.e., self-report vs. diagnostic interview); informant source (i.e., parent, child, teacher), number and type of anxiety disorders included in the study, and definition of diagnostic criteria for each disorder. A sophisticated meta-analysis of 41 studies conducted in 27 countries by Polanczyk yielded a pooled worldwide 12-month prevalence of any anxiety disorder of 6.5% (CI 95% 4.7 to 9.1).

The only study that included the full range of anxiety disorders in US adolescents was the National Comorbidity Adolescent Supplement (NCS-A), which was an extension of the NCS-R, using the same diagnostic methodology (Table 14.3–2). These studies yielded high aggregate lifetime prevalence rates of anxiety in adults (28.8%) and youth (31.9%). However, lifetime prevalence rates of disorders with severe impairment were substantially lower than lifetime DSM-IV disorders: GAD-0.9%; social phobia-1.3%, and 0.6% of those with separation anxiety and specific phobias. The finding that the prevalence rates in youth approximate those of adults suggests that the onset of most anxiety disorders occurs in childhood and adolescence. In fact, the rates of aggregate anxiety were actually higher in adolescents than in adults in the NCS-R for both 12-month (24.9% vs. 18.1%) and lifetime (31.9% vs. 28.8%) prevalence (see Table 14.3–2). By contrast, the rates of GAD

(5.7%), panic disorder (4.7%), and social anxiety disorder (12.1%) in adults were greater than those in adolescents, with lifetime rates of 2.2%, 2.3%, and 9.1%, respectively.

Data from several prospective community studies have followed community samples of children and young adults for more than a decade. These studies provide information on the cumulative incidence and stability of anxiety disorders in early and middle adulthood. The Zurich Cohort Study of a community sample of 19- to 20-year-old adults from Zurich, Switzerland, provides unique information on the prevalence and continuity of anxiety disorders across 30 years as they progressed from early to middle adulthood. The cumulative prevalence rate of anxiety disorders across the lifetime was 30%. Similar to cross-sectional studies, specific phobia was the most common disorder (26.9%), followed by GAD (20.8%), social anxiety (12.6%), agoraphobia (6.8%), and panic disorder (6.1%). Classification of the spectrum of disorders in the Zurich Cohort Study has further shown that there are a large proportion of people from the community whose symptoms fall below the diagnostic thresholds for mood and anxiety disorders but nevertheless exhibit impairment and/or distress. Consideration of subthreshold manifestations of anxiety disorders leads to further elevation of the rates of these conditions in the general population.

Table 14.3–2.
12-Month and Lifetime Rates of DSM-IV Anxiety Disorders in the National Comorbidity Survey Replication (Adults) and Supplement (Adolescents)

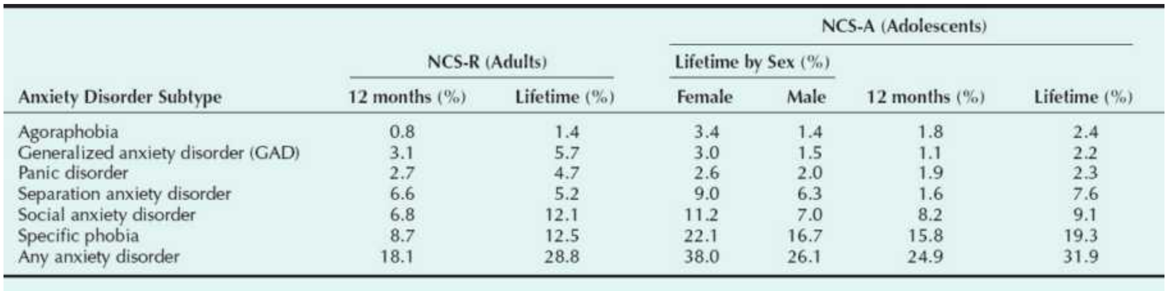

	NCS-R (Adults)		NCS-A (Adolescents)			
			Lifetime by Sex (%)			
Anxiety Disorder Subtype	12 months (%)	Lifetime (%)	Female	Male	12 months (%)	Lifetime (%)
Agoraphobia	0.8	1.4	3.4	1.4	1.8	2.4
Generalized anxiety disorder (GAD)	3.1	5.7	3.0	1.5	1.1	2.2
Panic disorder	2.7	4.7	2.6	2.0	1.9	2.3
Separation anxiety disorder	6.6	5.2	9.0	6.3	1.6	7.6
Social anxiety disorder	6.8	12.1	11.2	7.0	8.2	9.1
Specific phobia	8.7	12.5	22.1	16.7	15.8	19.3
Any anxiety disorder	18.1	28.8	38.0	26.1	24.9	31.9

Similar prevalence rates emerged from The Early Developmental Stages of Psychopathology Study (EDSP), a 10-year prospective longitudinal study of a community sample of youth ages 14 to 24 to age 33. Cumulative incidence rates were: total anxiety disorders—31.3%; specific phobia—7.0%; social phobia—6.0%; GAD—4.3%; and panic—

3.3%. Projected rates to age 33 were almost double those at the initial interview, demonstrating that the peak period for the onset of anxiety disorders is in late adolescence and early adulthood. Another prospective study of youth that assessed 1,420 participants in the southeastern United States up to 11 times between ages 9 and 26 was the Great Smoky Mountain Study (GSMS). The cumulative incidence rates of anxiety across the study were 22.7% for all anxiety, 9.7% for GAD, 6.1% for agoraphobia, 5.0% for separation anxiety disorder, 4.8% for panic disorder, 4.2% for social phobia, and 2.2% for specific phobia.

A recent follow-up of about 14,000 participants in the Avon Longitudinal Study of Parents and Children (ALSPAC) birth cohort in the United Kingdom identified sources of heterogeneity in the long-term trajectories of different anxiety disorders over time. They further identified that parental mood and anxiety disorders and sleeping difficulties in childhood were associated with the persistence of anxiety disorders from childhood to adulthood. Some of these differential correlates are described in the section below.

CORRELATES AND RISK FACTORS

Sex Differences

The results of community studies in adults reveal that women have greater rates of almost all anxiety disorders. Despite differences in the magnitude of the rates of specific anxiety disorders across studies, the sex ratio is strikingly similar. Women have an approximately twofold elevation in lifetime rates of panic disorder, GAD, agoraphobia, and simple phobia than males in nearly all of the studies. Studies of youth also report that girls tend to have more of all subtypes of anxiety disorders, irrespective of the age composition of the sample.

Epidemiologic surveys of adults reveal that the female preponderance of anxiety disorders is present across all stages of life but is most pronounced throughout early and mid-adulthood. The rates of anxiety disorders in males are also rather constant throughout adult life, whereas the rates in females peak in the fourth and fifth decades of life and decrease thereafter. The increased rates in females are present across all ages and do not diminish as the rates of anxiety decrease in late life.

Age of Onset

Retrospective reports of adults and adolescents with anxiety disorders suggest that the onset generally occurs in childhood or adolescence. For example, results of the NCS-R reveal that anxiety disorders have the earliest age of onset of all of the major classes of mental and behavioral disorders, with a median onset by the age of 12. The NCS-A found an even earlier median onset at age 6. The onset is far earlier than the onset of mood disorders or substance use disorders and comparable to that of impulse control disorders. The prospective GSMS found a U-shaped age prevalence curve for any anxiety disorder in which a sharp decrease in mid-childhood is followed by increasing levels from early adolescence to young adulthood.

Although there is substantial variation across studies, the results of prospective community-based research reveal differential peak periods of onset of specific subtypes of anxiety: separation anxiety and specific phobias in middle childhood; social anxiety disorder in middle adolescence; agoraphobia, and panic disorder from late adolescence to young adulthood; GAD in young adulthood.

Several follow-up studies of children and adolescents have shown that there is stability in both social anxiety disorder and simple phobia, whereas GAD and panic symptoms and disorders are more variable and tend to overlap with episodes of depression.

Social Class and Ethnicity

Community studies have consistently found that rates of anxiety disorders, in general, are greater among those at lower levels of socioeconomic status and education levels, but the associations might be more complex due to the significant interactions among the sociodemographic characteristics. For example, anxiety disorders were negatively associated with income and education levels in the NCS. There was almost a twofold difference between lifetime rates of anxiety disorders in individuals in the highest income bracket over the lowest and those who completed over 16 years of school versus less than 11. However, the association with education was not found in the NCS-R, and the association with income was not found in the NCS-A.

Certain anxiety disorders also seem to be elevated in specific occupations. According to the NEMESIS study, anxiety disorders were higher in homemakers and those unemployed or with a disability.

Similarly, the NCS found higher levels of GAD and phobias in homemakers, students, and other unemployed persons.

The research on how race/ethnicity and social class impact the prevalence of anxiety disorders is inconclusive. While several community studies have yielded greater rates of anxiety disorders, particularly phobic disorders, among African Americans, others have shown the opposite. For example, a recent study on 5,424 Hispanics, non-Hispanic Blacks, and non-Hispanic Whites in the NCS-R showed that both minority groups had a lower risk for GAD and social anxiety disorder, and non-Hispanic Blacks specifically had a lower risk for panic disorder. There was also a lower risk among Hispanics, relative to non-Hispanic Whites, but only among the younger cohort. The interaction between race/ethnicity and education may also be important as they identified that this lower risk among minorities was more pronounced at lower levels of education. In contrast, the NCS-A showed that non-Hispanic black adolescents had higher rates of anxiety disorders than non-Hispanic White adolescents, and no two-way interaction for anxiety disorders was found. The reasons for ethnic and social class differences have not yet been evaluated systematically; however, both methodological factors and differences in exposure to stressors have been advanced as possible explanations.

Patterns of Comorbidity

Subtypes of anxiety disorders tend to co-occur both cross-sectionally and across the lifetime. Comorbidity between anxiety disorders and other mental disorders is already apparent in childhood and adolescence. Anxiety disorders are associated with all of the other major classes of disorders, including mood disorders, disruptive behaviors, eating disorders, and substance use disorders. The co-occurrence of anxiety disorders and mood disorders is so common that there is emerging evidence that anxiety disorders may be part of the developmental sequence in which anxiety is expressed early in life, followed by depression in adulthood.

Both family and twin studies have been used to examine sources of overlap within the anxiety disorders and between the anxiety disorders and other syndromes, including depression, eating disorders, and substance abuse. Family studies have demonstrated the independence of familial aggregation of panic and phobias. With respect to comorbidity, panic disorder, GAD and depression have been shown to share common familial and genetic liability, whereas there is substantial evidence for the independent etiology of anxiety disorders and substance use disorders. Similar results have emerged from studies of symptoms of anxiety and depression in both youth and adults in which anxiety and depression were found to result from a common genetic diathesis.

Familial and Genetic Factors

The familial aggregation of all of the major subtypes of anxiety disorders has been well-established, and the substantial role of genetic factors underlying familial transmission has been well-documented. However, the relatively moderate magnitude of heritability also strongly implicates environmental etiologic factors. Components of anxiety that may be inherited include somatic manifestations of anxiety such as physiologic responses, which include pulse, respiration rate, and galvanic skin response. There is also substantial evidence for shared genetic risk factors for mood disorders and anxiety disorders that may manifest differentially across developmental stages. Increased rates of anxiety symptoms and disorders have also been found among offspring of parents with anxiety disorders which suggests that there may be underlying psychological or biological vulnerability factors for anxiety disorders in general which may already manifest in children prior to puberty. Children at risk for anxiety disorders throughout life are characterized by behavioral inhibition, autonomic reactivity, somatic symptoms, social fears, enhanced startle reflex, and respiratory sensitivity. These findings suggest that anxiety and fear states are highly heterogeneous and that future studies need to investigate the extent to which the components of anxiety result from common versus unique genetic factors and the role of environmental factors, either biological or social, in either potentiating or suppressing their expression.

Temperament/Personality

One of the earliest indicators of vulnerability to the development of anxiety is behavioral inhibition, characterized by increased physiological reactivity or behavioral withdrawal in the face of novel stimuli or challenging situations. Behavioral inhibition may be a manifestation of a biological predisposition characterized by both overt behavioral and physiologic measures.

Likewise, anxiety sensitivity is another potential sensitive and specific trait marker for the development of anxiety disorders. Anxiety sensitivity is characterized by beliefs that anxiety sensations are indicative of harmful physiological, psychological, or social consequences (e.g., fainting or an impending heart attack). Anxiety sensitivity has been shown to precede the development of anxiety symptoms and disorders and to have specificity in predicting the development of anxiety as opposed to depression.

Medical Symptoms/Disorders

There has been a striking increase in studies of anxiety disorders among a range of medical conditions, including diabetes, cardiovascular disease, respiratory disorders, and neurologic disorders (i.e., epilepsy, migraine, multiple sclerosis, Parkinson's disease). Medical comorbidity is far more strongly associated with GAD and panic disorder than phobic disorders. There is heterogeneity in the explanations for comorbidity that may reflect common underlying diatheses, anxiety disorders as complications of disability from the medical condition, or anxiety as a side effect of medications used to treat the disorder. Community studies have also provided an important basis for identifying the extent to which clinical samples are biased by increased treatment-seeking among those with comorbid medical or mental disorders. These findings highlight the importance of a comprehensive evaluation of concurrent medical conditions in the treatment of people with anxiety disorders.

Exposure

The role of life experiences in the etiology of anxiety states, particularly phobias and panic disorders, has been widely studied. Life events have often been designated a causal role in the onset of phobias, which are linked inherently to particular events or objects. More broadly, life experiences that are likely to trigger specific future problems with safety and security are perceived to predict the onset of anxiety disorders. While it is likely that exposure may trigger the fear response, there is compelling evidence for the evolutionary tendency toward fear and avoidance of potential predators. Therefore, phobic states may first become apparent without exposure. The ability to extinguish such fears may be protective against the development of clinically significant phobias. Recent research suggests such life experiences that also trigger immediate distress (e.g., parental divorce or unemployment) might also more broadly precipitate the onset of comorbid depression and anxiety. In evaluating the evidence on the causal role of life experiences, it is critical to consider the subtypes of anxiety disorders separately. Moreover, stressful life events may interact with other risk factors, such as a family history of depression precipitating episodes of panic.

COURSE AND PERSISTENCE

There has been limited information on the persistence of anxiety disorders across the lifespan. Several prospective studies have also shown that the specific subtypes of anxiety fluctuate over time, particularly generalized anxiety and panic attacks. By contrast, phobic states, particularly social anxiety, tend to be more stable across the lifespan. A recent review of predictors of persistent anxiety disorders that included nearly four dozen studies and 30,000 patients with anxiety disorders found that poor treatment response, greater severity and duration of anxiety disorders, and a profile of specific clinical and psychological characteristics were associated with greater persistence of anxiety. Finally, there is also evidence that anxiety disorders may be associated with increased mortality, likely attributable to respiratory and cardiovascular correlates of these conditions.

IMPACT OF ANXIETY DISORDERS

Although the individual impact of anxiety disorders has been well-recognized, early estimates of the global burden of disease tended to neglect anxiety disorders as a significant source of lifetime disability. Recent studies of the global burden of disease have employed the concept of Disability Adjusted Life Years (DALYs) as an index of the impact of particular diseases on mortality and disability. Earlier estimates were based on a limited number of anxiety disorders, particularly obsessive-compulsive disorder and post-traumatic stress disorder (PTSD), that are no longer included in the class of anxiety disorders. Estimates of both person-years lived with disability (YLDs) and years of lost life (YLL) are used as a measure of the societal impact of anxiety disorders. The most recent global estimate of YLDs is 370/100,000 population for all anxiety disorders, and the total DALY for all anxiety disorders is about 459 for females and 282 for males. The peak disability occurs in the adolescent and young adult age group from 10 to 24 years. Anxiety is the eighth leading cause of YLDs after low back pain, depressive disorders, headaches, diabetes, hearing loss, other musculoskeletal disorders, gynecological diseases, and neck pain and falls, with similar patterns across country-specific income levels. These findings highlight the significant public health impact of anxiety disorders, particularly when coupled with other comorbid conditions.

Evidence for the impact of anxiety on the life course of individuals who suffer from one or more of these conditions has emerged from prospective follow-up studies of children with anxiety into adulthood. The most important consequences are comorbid mental and physical conditions, substance use disorders, and in a small proportion, suicide. The 15-year follow-up of the Smoky Mountains study revealed that youth with anxiety disorders during childhood had substantial impairment in one or more of the domains assessed, including health, financial, and interpersonal functioning in adulthood. Results also showed that youth with GAD tended to have poor functioning across all domains, whereas those with social phobia only displayed impairment in the interpersonal domain, and those with separation anxiety had more health problems than other youth. These studies illustrate that the burden of anxiety disorders is attributable to both their childhood and adolescent-onset and to their continuity across the lifespan. In consideration of their association with the subsequent development of substance use disorders, behavior and mood disorders, these findings indicate the critical need for intervention at the early stages of these conditions.

FUTURE DIRECTIONS

This chapter provides an update on the magnitude of anxiety disorders in adults and children in community surveys across the world. Epidemiology is expanding beyond the North American/European boundaries of earlier work to achieve greater representation of the global prevalence and impact of mental disorders. The most consistent finding across studies is the high prevalence of anxiety disorders and the consistency in correlates of anxiety disorders across geographic settings and methods of assessment. Estimates of global disability associated with anxiety disorders document the substantial individual and societal impact of anxiety disorders. The impact of anxiety disorders on lower educational and occupational attainment, missed days in one’s role, and comorbid associations with other mental and physical disorders highlight their public health relevance. The huge gaps between the high magnitude of impairing anxiety disorders and low mental health treatment rates remain a serious challenge in public health. This is of even greater significance because of the well-established efficacious behavioral and pharmacologic treatments for these conditions. The growing integration of pediatrics and medicine with psychiatry may facilitate efforts to provide treatment to a larger segment of the population and enhance recognition of comorbidity as an important source of disability from anxiety disorders.

The population surveys described herein have been supplanted to some extent by the increasing focus on biobanks, treatment registries, and pharmacologic and insurance databases that are becoming increasingly available. This work will require new tools to derive estimates of the generalizability of these data and may lead to increased efforts to move beyond diagnostic codes in determining population prevalence. In addition, as neuroscience and genetics inform knowledge regarding neural processes underlying anxiety disorders and the role of genetic and environmental factors in their evolution, studies of treatment and prevention strategies will assume increasing importance in reducing the magnitude and burden of this major source of mental disorders.

FURTHER READINGS

Alonso J, Liu Z, Evans-Lacko S, et al. Treatment gap for anxiety disorders is global: results of the World Mental Health Surveys in 21 countries. *Depress Anxiety*. 2018;35(3):195–208.

Angst J, Paksarian D, Cui L, et al. The epidemiology of common mental disorders from age 20 to 50: results from the prospective Zurich Cohort Study. *Epidemiol Psychiatr Sci*. 2016;25(1):24–32.

Comer JS, Blanco C, Hasin DS, et al. Health-related quality of life across the anxiety disorders: results from the national epidemiologic survey on alcohol and related conditions (NESARC). *J Clin Psychiatry*. 2011;72(1):43–50.

Costello EJ, Egger H, Angold A. 10-year research update review: the epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. *J Am Acad Child Adolesc Psychiatry*. 2005;44(10):972–986.

de Jonge P, Roest AM, Lim CCW, et al. Cross-national epidemiology of panic disorder and panic attacks in the world mental health surveys. *Depress Anxiety*. 2016;33(12):1155–1177.

de Jonge P, Wardenaar KJ, Lim CCW, et al. The cross-national structure of mental disorders: results from the World Mental Health Surveys. *Psychol Med*. 2018;48(12):2073–2084.

Eaton WW, Merikangas KR. Psychiatric epidemiology: progress and prospects in the year 2000. *Epidemiol Rev*. 2000;22(1):29–34.

Harford TC, Chen CM, Kerridge BT, Grant BF. Self-and other-directed forms of violence and their relationship with lifetime DSM-5 psychiatric disorders: results from the National Epidemiologic Survey on Alcohol Related Conditions−III (NESARC−III). *Psychiatry Res*. 2018;262:384–392.

Hovenkamp-Hermelink JHM, Jeronimus BF, Myroniuk S, Riese H, Schoevers RA. Predictors of persistence of anxiety disorders across the lifespan: a systematic review. *Lancet Psychiatry*. 2021;8(5):428–443.

Iza M, Olfson M, Vermes D, Hoffer M, Wang S, Blanco C. Probability and predictors of first treatment contact for anxiety disorders in the United States: analysis of data from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *J Clin Psychiatry*. 2013;74(11):1093–1100.

Lateef T, He J-P, Nelson K, et al. Physical-mental comorbidity of pediatric migraine in the Philadelphia Neurodevelopmental Cohort. *J Pediatr*. 2019;205:210–217.

McGrath JJ, Lim CCW, Plana-Ripoll O, et al. Comorbidity within mental disorders: a comprehensive analysis based on 145 990 survey respondents from 27 countries. *Epidemiol Psychiatr Sci*. 2020;29:e153.

Meier SM, Mattheisen M, Mors O, Mortensen PB, Laursen TM, Penninx BW. Increased mortality among people with anxiety disorders: total population study. *Br J Psychiatry*. 2016;209(3):216–221.

Merikangas KR, He JP, Burstein M, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *J Am Acad Child Adolesc Psychiatry*. 2010;49(10):980–989.

Merikangas KR, He JP, Burstein M, et al. Service utilization for lifetime mental disorders in U.S. adolescents: results of the National Comorbidity Survey-Adolescent Supplement (NCS-A). *J Am Acad Child Adolesc Psychiatry*. 2011;50(1):32–45.

Merikangas KR, He JP, Rapoport J, Vitiello B, Olfson M. Medication use in US youth with mental disorders. *JAMA Pediatr*. 2013;167(2):141–148.

Olfson M, Blanco C, Wall MM, Liu S-M, Grant BF. Treatment of common mental disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions-III. *J Clin Psychiatry*. 2019;80(3):18m12532.

Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *J Child Psychol Psychiatry*. 2015;56(3):345–365.

Roest AM, de Jonge P, Lim CCW, et al. Agoraphobia. In: Scott KM, de Jonge P, Stein DJ, Kessler RC, eds. *Mental Disorders Around the World: Facts and figures from the WHO World Mental Health Surveys*. 5th ed. Cambridge University Press; 2018:106–119.

Ruscio AM, Hallion LS, Lim CCW, et al. Cross-sectional comparison of the epidemiology of DSM-5 generalized anxiety disorder across the globe. *JAMA Psychiatry*. 2017;74(5):465–475.

Stein DJ, Lim CC, Roest AM, et al. The cross-national epidemiology of social anxiety disorder: data from the World Mental Health Survey Initiative. *BMC Med*.

2017;15(1):143.

Wardenaar KJ, Lim CCW, Al-Hamzawi AO, et al. The cross-national epidemiology of specific phobia in the World Mental Health Surveys. *Psychol Med*. 2017;47(10):1744–1760.

14.4 Anxiety Disorders: Neurobiology and Neuroscience

Sara Costi, M.D., Adriana Feder, M.D., Brian Iacoviello, Ph.D., James W. Murrough, M.D., Ph.D., and Dennis S. Charney, M.D.

Anxiety disorders encompass a heterogeneous group of conditions, and our understanding of their underlying neurobiologic basis continues to evolve. Evidence from preclinical and clinical research has implicated a range of neural systems in the etiology of fear and anxiety responses. Dysfunction in these systems also plays a key role in the development of anxiety disorders through interactions with cortical and subcortical neural circuits involved in fear responses. In animals, threat detection and defense responses rely on nonconscious mechanisms, which in turn lead to responses in the brain and body that eventually result in conscious fear. Although threat responses have been extensively studied in animals and continue to inform the understanding of anxiety disorders in humans, they cannot inform on the subjective experience of fear and anxiety that are uniquely humans. Thus, continued close collaboration between basic and clinical researchers is required to advance the current understanding of the neurobiology of anxiety disorders. Advances in functional neuroimaging have also allowed for greater understanding of key brain regions in humans with and without anxiety disorders and during emotion, learning, or anxiety-inducing tasks. In addition, progress in genetics allows for increased understanding of vulnerability to anxiety disorders or a poorer treatment response.

NEUROCHEMICAL BASIS OF FEAR AND ANXIETY

The human body responds to perceived threats and fear-inducing stimuli by activating the release of specific neurotransmitters and neuropeptides located in key brain regions (Fig. 14.4–1). The “fight-or-flight” reaction can be adaptive in that increased vigilance, recall of potentially dangerous elements, and rapid physiologic response (evidenced by increased pupillary dilation, increased heart rate, and rapid breathing) prepare the organism for impending danger before quickly subsiding once the fear-inducing trigger is no longer present. On the other hand, overgeneralization of fear-learning to harmless stimuli is thought to characterize anxiety disorders.

Neurochemicals released during stress or periods of prolonged anxiety can result in chronic alterations to a range of neurobiologic systems. Severe stress can also have maladaptive consequences when it causes biologic systems to be chronically activated. Based on preclinical and clinical models of anxiety, there is evidence that exposure to early-life stressors or trauma may result in predisposition to the development of anxiety disorders later in life, by alterations in neuroendocrine systems or perhaps by direct neurotoxic effects. Identification of mechanisms underlying individual differences in susceptibility versus resilience to stressors is an area of active investigation. This chapter reviews current knowledge regarding the neurobiologic systems involved in anxiety and fear response, including neurotransmitters, neuropeptides, and hormones, and summarizes the main findings pertinent to anxiety disorders from animal models of anxiety and from human studies. A figure showing the brain’s key regions involved in response to fear and anxiety is included below (see Fig. 14.4–1).

NORADRENERGIC SYSTEM

Norepinephrine (NE) is a catecholamine released in the central nervous system (CNS) primarily from the locus coeruleus (LC), which is located in the pons. NE is a product of hydroxylation of dopamine, which is a metabolite of tyrosine. NE has several functions, including regulation of CNS cells and circuits. Acute threats result in the release of NE, a key neurotransmitter involved in the “fight-or-flight” response, characterized by coordinated physiologic and behavioral responses. Both “fight” and “flight” are characterized by increased arousal and a higher level of vigilance, critical for assessing threat and reacting quickly in a dangerous situation.

While fight or flight is adaptive responses to a potential threat, overactivity of the NE system is associated with an aberrant response to threat and a prolonged anxious state.

At the brain level, the NE system receives input from and sends projections to a variety of regions, modulating neurons that release a range of other neurotransmitters, including dopamine, serotonin, glutamate and γ-amino butyric acid (GABA). Neurons in the LC and projections from the LC to cortical and limbic regions are involved in the response to fearful or threat stimuli and adaptation to acute and chronic stress. NE is also released by sympathetic nervous system ganglia and via synaptic vesicles in the adrenal medulla. It is responsible for regulating the body's behavioral and physiologic responses to stress through activity of the adrenergic receptors or adrenoreceptors. These receptors are G-protein–coupled and are responsible for controlling NE release, in part via a negative feedback mechanism. α-1 adrenergic receptors are postsynaptic, whereas α-2 adrenergic receptors are presynaptic. Animal studies of these receptors have led to a greater understanding of the role of NE in anxiety disorders. Agonists of the α-2 adrenergic receptors, such as clonidine and guanfacine, are antiadrenergic, whereas α-2 antagonists, such as idazoxan and yohimbine, result in increased NE release. In addition, certain environmental stressors in animal studies, such as acute cold restraint and inescapable stress, cause a reduction in α-2 adrenergic receptors density in the hippocampus and amygdala. Mice lacking the α-2 receptor exhibit heightened autonomic activation and concomitant anxiety behaviors after administration of D-amphetamine.

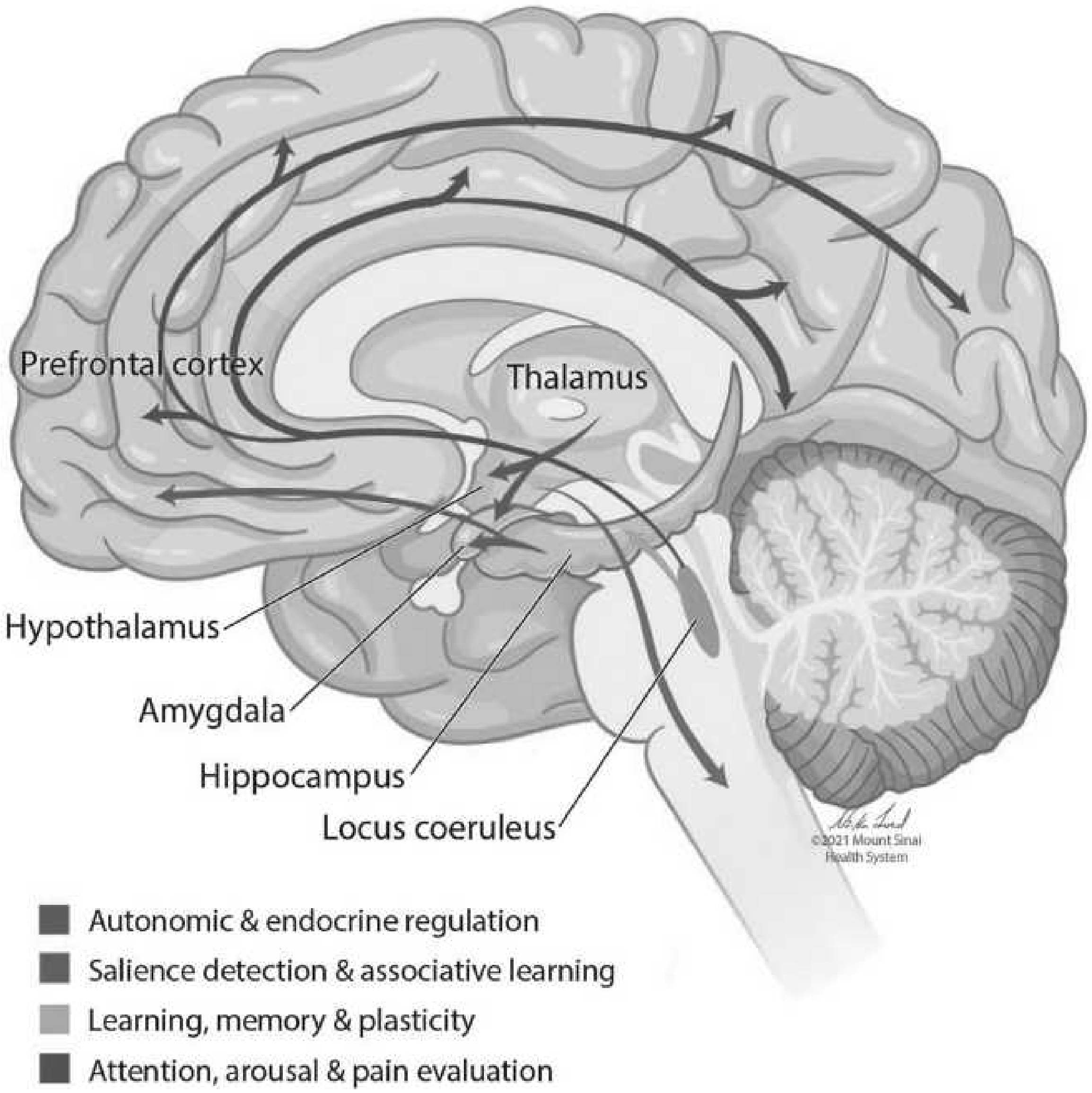
Prefrontal cortex
Thalamus
Hypothalamus
Amygdala
Hippocampus
Locus coeruleus
©2021 Mount Sinai Health System
■ Autonomic & endocrine regulation
■ Salience detection & associative learning
■ Learning, memory & plasticity
■ Attention, arousal & pain evaluation

FIGURE 14.4–1. Schematic representation of the major brain systems involved in response to fear and anxiety. Projections from the thalamus regulate the autonomic and endocrine responses; projections from the hippocampus and amygdala modulate salience detection and associative learning; projections from the locus coeruleus innervate the cortex for regulation of attention, arousal, and pain evaluation. (See eBook for color images.) (Printed with permission from © Mount Sinai Health System.)

Proxy measures of NE system activation can be measured during a state of anxiety or threat in humans. Activation of the sympathoadrenal system can be measured with either plasma or urine sample since circulating levels of NE have been shown to correlate with its concentration in the urine. While 24-hour urinary epinephrine excretion has been found to positively correlate with anxiety symptom severity, 24-hour urinary NE excretion did not correlate with anxiety or depression in subjects with untreated high blood pressure. Similarly, pupil dilation and 3-methoxy-4-hydroxyphenylglycol (MHPG)—a major metabolite of NE—can be measured in plasma and urine, in both humans and animals, in response to a threat. Specifically, dysfunction of the NE system has been demonstrated in panic disorder (PD) and phobias. There has been longstanding clinical evidence that propranolol, a nonselective β-adrenergic receptor-blocking agent, is useful in the management of social anxiety disorder (SAD), particularly performance anxiety, although it has no direct impact on anticipatory anxiety. Beta-blockers reduce peripheral autonomic symptoms such as hand tremors and tachycardia, often associated with social anxiety. Further, proadrenergic agents such as yohimbine have been reported to be panicogenic in animal and human models of anxiety. Yohimbine is known to precipitate panic attacks in patients with PD but not in healthy subjects. In a study of children with various anxiety disorders, yohimbine caused greater anxiety responses than in the healthy control group. A more selective α-2 receptor antagonist than yohimbine, idazoxan, has been shown to provoke panic attacks in healthy subjects. Patients with PD also have an abnormal response to the α-2 receptor agonist clonidine, with greater decreases in blood pressure and plasma MHPG compared to controls. However, CSF concentration of MHPG has not been found to significantly differ in PD patients compared to non-PD depressed subjects.

Studies of phobic individuals have demonstrated higher baseline rates of NE and other catecholamines and increased release of NE in response to anxiety-inducing triggers. This is not surprising, given the role that NE plays in fear conditioning and how repeated exposure to a stressful stimulus causes a potentiated release of NE, reflecting a process of sensitization. Patients with generalized anxiety disorder (GAD) also show higher plasma levels of NE and MHPG than healthy controls. Their response to yohimbine, however, is generally similar to that of healthy controls. Guanfacine, aα-2 receptor agonist approved in extended-release form for attention-deficit/hyperactivity disorder (ADHD), has been primarily tested for post-traumatic stress disorder (PTSD) but showed no difference compared to placebo in clinical trials. In a trial of guanfacine in children and adolescents aged 6 to 17 years with anxiety disorders including GAD, SAD, and separation anxiety disorder, no differences were observed between treatment groups on anxiety, although a higher proportion of subjects receiving guanfacine showed greater clinical improvement as judged at the investigator compared to placebo.

To summarize, patients with different types of anxiety disorders exhibit symptoms consistent with an exaggerated noradrenergic system output, including excessive worry, avoidance, heightened arousal, and startle, along with increased autonomic and sympathetic activation. Drugs that act to attenuate this effect by decreasing the firing of NE neurons, including substances such as alcohol, opioids, and benzodiazepines, are commonly used and abused by patients with anxiety disorders as a form of self-medication. Drugs that increase NE signaling, such as SNRIs and tricyclic antidepressants, are thought to treat anxiety via eventual adaptive changes in NE (and serotonin) receptors occurring weeks later.

HYPOTHALAMIC–PITUITARY–ADRENAL AXIS

Cortisol is a glucocorticoid synthesized in the adrenal glands. It contributes to increased arousal, vigilance, focused attention and memory formation, inhibition of the growth and reproductive systems, and containment of the immune response. It is also known to regulate, marshal, and replenish energy stores. There is ample support for the notion that many types of psychological stress increase the synthesis and release of cortisol. Cortisol is a principal contributor in regulating hippocampus, amygdala, and prefrontal cortex (PFC) activity. It has been shown to enhance amygdala activity, increase corticotropin-releasing hormone (CRH), messenger ribonucleic acid (mRNA) concentrations in the central nucleus of the amygdala (CEA), increase the effects of CRH on conditioned fear, and facilitate the encoding of emotion-related memory. Of interest, this glucocorticoid-mediated facilitation of emotional memory encoding has been found to be contingent on endogenous arousal-induced activation of NE in the amygdala. In addition, glucocorticoids such as cortisol have biphasic effects on cognitive function, hippocampal excitability, and memory. Such effects may account, at least in part, for the adaptive role of cortisol in modulating behavior in response to acute stress.

The stress-induced increase in cortisol release is regulated through a complex negative feedback system mediated via moderate-to-low-affinity glucocorticoid receptors (GRs) and high-affinity mineralocorticoid receptors (MRs). The hippocampus has a very high concentration of GR receptors and possibly also MR receptors and plays an important role in the negative feedback of the HPA axis. If stress-induced cortisol secretion is not contained, persistently elevated cortisol levels can cause diverse harmful effects both peripherally and in the CNS. Peripheral adverse effects include hypertension, osteoporosis, immunosuppression, insulin resistance, dyslipidemia, coagulation abnormalities, and ultimately atherosclerosis and cardiovascular disease. The hippocampus is particularly sensitive to many types of insults. High glucocorticoid levels impair cell survival, alter cell metabolism, and induce changes in cell morphology. The stress-related impairment in hippocampal-dependent cognitive and memory function may be a consequence of this cellular damage. Furthermore, there is accumulating evidence that the amygdala may play an important role in mediating the negative effects of stress on the hippocampus.

A large number of studies have investigated the role of the HPA axis in anxiety disorders, examining plasma, urine, or saliva samples either in basal conditions or after pharmacologic or behavioral stressor tests. Tests used to probe HPA axis function in patients with anxiety disorders have included examination of cortisol suppression after administration of low-dose dexamethasone (DEX), the response of adrenocorticotropic hormone (ACTH) after CRH administration, and the DEX/CRH challenge test, a sensitive and specific technique for assessing HPA axis function. Results of these studies are summarized below. In patients with PD, analyses of plasma cortisol levels have yielded mixed findings, with elevated daytime or nighttime levels observed in some studies but not others. Similarly, urinary-free cortisol has been reported to be within range or elevated. Plasma ACTH has been found to be elevated in patients compared with controls. Blunted ACTH responses to CRH have been reported in PD patients in some but not all studies. While circulating cortisol levels marginally increase during naturally occurring panic attacks as measured with indwelling catheters in patients resting in bed, a significant increase in salivary cortisol concentration has been reported in PD patients during spontaneous panic attacks compared to levels measured on panic-free days on the same subjects, at a similar time of the day.

To further understand the relationship between panic attacks and cortisol, several trials have employed panic provocation tests. Although in the majority of studies, patients who developed a panic attack following lactate infusion generally did not show elevations in ACTH or cortisol, one study found that cortisol levels dropped during lactate-induced panic in both patients and controls. By contrast, patients who developed panic attacks following yohimbine administration showed a significant increase in plasma cortisol levels compared to healthy volunteers. Oral administration of the partial serotonin agonist meta-chlorophenylpiperazine (mCPP) or oral caffeine was associated with an increase in plasma cortisol in both patients and controls. However, recent evidence from a placebo-controlled study suggests that increases in plasma cortisol, ACTH, and dehydroepiandrosterone sulfate (DHEAS) observed after a 400-mg caffeine challenge administered as instant coffee in PD patients are not associated with the occurrence of panic attacks. Cognitive intervention in PD has been shown to significantly reduce the elevation of cortisol and ACTH by pentagastrin challenge, although it had no effect on panic symptoms per se. Thus, HPA axis hypersensitivity in PD may be activated by environmental contextual cues such as novelty, whereas baseline activity might remain normal. Finally, while most studies found a normal response to the DEX suppression test in PD subjects, results from the CRH stimulation test have been discordant, with some studies reporting a reduced ACTH response and ACTH/cortisol ratio, and others showing inconclusive findings.

Despite some discrepancies in the literature, most evidence suggests that GAD is also associated with HPA axis abnormalities. Specifically, it is still unclear if GAD is associated with elevated cortisol levels. While some studies report similar levels of baseline cortisol and cortisol responses to challenge tests, others report abnormal (increased or decreased) HPA axis activity in these patients. The DEX suppression test showed a reduction in negative feedback sensitivity of the HPA axis in GAD patients. Interestingly, cortisol levels, as well as severity of anxiety symptoms, were demonstrated to decrease after treatment, independently from the kind of therapy provided, benzodiazepine, SSRI, or psychotherapy. Studies analyzing cortisol concentrations in hair (a measure of long-term cortisol levels) showed 50% to 60 lower hair cortisol concentrations in GAD patients compared to healthy volunteers, supporting the hypothesis that chronic anxiety is associated with a downregulation of the HPA axis.

As the HPA axis has been found to play a pivotal role in social interactions in animal studies, the role of cortisol in the regulation of behavioral responses to social stressors has been investigated in SAD. Several studies reported no abnormalities in basal HPA axis function in adult subjects suffering from SAD by measuring urinary free cortisol levels, free cortisol/creatinine ratio, and the 24-hour excretion of urinary free cortisol. While a study found lower basal plasma cortisol levels during morning hours in patients with SAD compared to healthy controls, a recent study failed to confirm the association between SAD and baseline saliva cortisol during morning hours in a large epidemiologic cohort of 1,050 subjects aged 14 to 21 years old. In the same study, baseline hair cortisol was inversely correlated to SAD in males but not females. Further, longitudinal analyses available in a subgroup of the sample (n = 605) showed a positive association between hair cortisol change between baseline and 1 year follow-up. Overall, this data suggests that both gender and age may mediate the relationship between cortisol and SAD.

Another study found no differences in the level of cortisol following DEX test in SAD subjects. Data from studies in children and adolescents with SAD confronted with a psychosocial stressor show conflicting results, with some reporting similar salivary cortisol levels compared to controls and others increased salivary cortisol levels in children and adolescents with SAD. A recent study conducted on 35 adults with SAD and 35 healthy controls showed a blunted plasma cortisol in SAD subjects compared to matching healthy volunteers, while no difference emerged for ACTH levels between groups.

Finally, HPA axis abnormalities have also been reported in patients with specific phobias (SPs), wherein exposure to the phobic object resulted in an increase in cortisol levels. Interestingly, one study reported a decrease in salivary cortisol levels collected in the afternoon hours following an intensive course of psychotherapy in individuals with SP from the first assessment (following a 2-minute exposure) to the final assessment (following a 15-minute exposure).

There has been ongoing research on the use of glucocorticoids in the treatment of anxiety disorders. In preclinical and clinical studies, glucocorticoids have been shown to block the recall of traumatic memories. Cortisol was also found to reduce the activation of fear responses in patients with SAD and spider phobia. In one of these studies, the oral administration of cortisone shortly before a social stressor significantly reduced social fear prior to, during, and in the aftermath of a stressor in individuals with SAD; cortisol levels correlated with fear ratings in the placebo group, suggesting a role for glucocorticoids in reducing the intensity of phobic fear. In a double-blind placebo-controlled study, oral cortisol administration was shown to enhance therapeutic response to in vivo exposure-based group therapy in patients with SP of spiders. In preclinical studies, RU-486 (mifepristone), a GR antagonist, impaired the recall of stressful memories in rats. In a pilot study conducted in older adults with anxiety disorder (GAD, PD, or anxiety disorder not otherwise specified) and cooccurring cognitive dysfunction, administration of mifepristone was associated with improvements in memory, executive functions, and severity of worry in patients with elevated cortisol levels at baseline. PH94B, an intranasally administered neurosteroidal aerosol, has been investigated for the treatment of SAD in a randomized, double-blind, placebo-controlled study, wherein adult females with SAD were randomized to placebo or PH94B nasal spray 15 minutes prior to a laboratory-simulated public speaking and social interaction challenge. PH94B was associated with significant improvement in anxiety symptoms compared to placebo. Another study assessed the use of PH94B for SAD in a randomized, double-blind, placebo-controlled trial where males and females with SAD were randomized to either placebo or PH94, and self-administered the nasal spray prior to a distressing social interaction or performance. PH94B was superior to placebo in decreasing symptoms of anxiety. Further studies of glucocorticoids in humans are underway.

CORTICOTROPIN-RELEASING HORMONE

CRH, often referred to as corticotropin-releasing factor (CRF), is also a principal component of the response to stress, coordinating the adaptive behavioral and physiologic responses that occur after stress exposure. Driven by stress, hypothalamic levels of CRH are increased, and the HPA axis becomes active or perhaps even hyperactive, resulting in increased levels of cortisol and DHEA. CRH also exerts extrahypothalamic effects, which may be no less important than those mediated by the HPA axis. CRH-containing neurons are broadly distributed within the brain and can be found in the prefrontal and cingulate cortices, the CEA, the bed nucleus of the stria terminalis (BNST), the nucleus accumbens (NAc), the periaqueductal gray (PAG) matter, and brainstem nuclei such as the LC and dorsal and median raphe. The activation of amygdala CRH neurons triggers fear-related behaviors, and the engagement of cortical CRH neurons may lessen reward expectation. CRH is also involved in the regulation of diverse neurovegetative functions, such as food intake, sexual activity, and growth and reproduction. In most cases, the administration of CRH inhibits these behaviors, whereas the administration of CRH antagonists increases them.

It has been shown that early-life stress may bring about long-term alterations in brain CRH concentration. Furthermore, both CRH secretion and the individual response to heightened CRH levels apparently depend on numerous environmental factors such as current and/or past social environment, current and/or past trauma history, and behavioral dominance. Continuous high levels of both hypothalamic and extrahypothalamic CRH strongly contribute to psychobiologic allostatic load. Preclinical studies have demonstrated that basal cerebrospinal fluid (CSF) CRH is higher in primates that have experienced early-life stress. The association between CRH and behavior occurs in a context-dependent manner. Exposure to excessive levels of CRH in early life is associated with hippocampal damage later in life. Of interest, intracerebroventricular injection of CRH to the hippocampus produces an increase in MR levels, suggesting that CRH is an important regulator of HPA axis function. A study found that an SNP in the CRH gene was associated with behavioral inhibition, a childhood trait that predicts adult anxiety. Still, the exact functional impact of the SNP on the manifestation of the CRH gene is yet to be determined.

It has been suggested that CRH-1 and CRH-2 receptors have a major role in the regulation of anxiety. Both types of receptors are found in the pituitary and throughout the neocortex (especially in prefrontal, cingulate, striatal, and insular cortices), amygdala, and hippocampal formation in the primate brain. However, other regions of the brain display predominantly one type of CRH receptor: CRH-1 receptors can be found within the LC, nucleus of the solitary tract, thalamus, and striatum, whereas CRH-2 receptors are chiefly located in the choroid plexus, certain hypothalamic nuclei, the nucleus prepositus, and the BNST. The varied distribution of these receptors suggests that each has a distinct role within the primate brain.

In rodents, CRH-1–deficient mice exhibit less anxiety-like behavior and a dampened response to stress. Conversely, CRH-2–deficient mice show more anxiety-like behavior and an exaggerated response to stress. Thus, the activation of CRH-1 receptors may contribute to increased anxiety-like responses, whereas the stimulation of CRH-2 may be pivotal in mediating anxiolytic behavior. Maintaining the appropriate functional equilibrium in the contribution of each of the two CRH receptor subtypes to brain CRH pathways might be essential to maintain the protective and adaptive psychological and physiologic response to stressors. Although the study of CRH receptor function in humans is limited, genetic investigations have reported an association between nucleotide polymorphisms of the CRH-1 receptor gene and PD. There are also ongoing efforts to develop CRH receptor ligands for single-photon emission tomography (SPECT) and PET imaging. In light of the anxiogenic effects of CRH-1–receptor stimulation described previously, CRH-1 antagonists have been developed. To date, however, no CRH-1 receptor antagonist has demonstrated safety and efficacy in phase III trials for a stress-related psychiatric disorder. A large multicenter clinical trial failed to demonstrate superiority of the selective CRH-1 receptor antagonist pexacerfont over placebo in patients with GAD. Clinical trials of CRH-1 receptor antagonists verucerfont and emicerfont for SAD have been completed, but results have not been published yet.

DOPAMINERGIC SYSTEM

Dopamine, a catecholamine, is involved in goal-directed and reward-seeking behaviors. Dopamine, however, is also released in response to acute stress in a number of specific brain areas important in the regulation of affective behavior, including the basolateral nucleus of the amygdala, the NAc, and the medial prefrontal cortex (mPFC). The dopamine innervation of the mPFC appears to be particularly vulnerable to stress. Thus, low-intensity stress (such as that associated with conditioned fear) or brief exposure to stress increases dopamine release and metabolism in this region, even without any notable changes in other mesotelencephalic dopamine regions. This sensitivity of dopamine receptors in the mPFC is supported by animal studies showing that low-intensity electric foot shock increases in vivo tyrosine hydroxylase and dopamine turnover in the mPFC but not in the NAc or striatum. In other areas receiving dopamine innervation, stress can enhance dopamine release and metabolism, provided that greater-intensity or longer-duration stress is applied. Thus, sensitivity and response of different brain regions to dopamine activation differ, with mPFC dopamine innervation activated prior to mesolimbic and nigrostriatal systems, and mesolimbic dopamine innervation more sensitive to stress than striatal dopamine innervation.

There is preclinical evidence that the extent of activation of the mesocortical dopamine system to stress may be, in part, genetically determined. It has been further suggested that extreme mesocortical dopamine release by stressful events may represent a vulnerability to depression and may lead to helpless reactions through an inhibition of subcortical dopamine transmission. Lesions of the amygdala before and after training in a conditioned stress model block stress-induced mPFC dopamine metabolic activation, suggesting amygdala control of stress-induced dopamine activation and a role in integrating the behavioral and neuroendocrine components of the stress response. On the other hand, lesions of mPFC dopamine neurons delay extinction (no effect on acquisition) of the conditioned fear response, indicating that prefrontal dopamine neurons are involved in facilitating the extinction of the fear response. One way to reconcile these two seemingly contradictory sets of data is to suggest that there is an optimal range for stress-induced increases in mPFC dopamine release to facilitate adaptive behavior responses. Too much mPFC dopamine release produces cognitive impairment, and an inhibition in NAc dopamine activity results in abnormalities in motivation and reward mechanisms. Insufficient prefrontal cortical dopamine release delays the extinction of conditioned fear. In humans, preliminary studies have reported a reduced density of the dopamine transporter and D_2 receptors in patients with SAD. Notably, increased anxiety is among the symptoms usually associated with dopamine agonist withdrawal in patients treated for conditions like Parkinson disease or restless legs syndrome.

One research group found a higher concentration of the dopamine metabolite HVA in plasma in PD patients with high levels of anxiety and frequent panic attacks. PD patients were also shown to have a larger growth hormone response to the DAT agonist apomorphine than patients with depression. Although no direct alteration in CSF HVA level has been demonstrated in PD patients, and no correlations between anxiety severity and panic attacks have been found, another study conducted on patients suffering from PD with concomitant SAD found lower CSF HVA levels.

SEROTONERGIC SYSTEM

Serotonin (5-HT), the release of which is regulated by the dorsal raphe nuclei, is secreted in response to stress and has been studied extensively in the context of anxiety disorders. Serotonin has been shown to have a dual role in anxiety. The serotonin hypothesis of anxiety suggests that 5-HT may be anxiogenic through its action on the PFC and amygdala, causing a heightened awareness of threats, and anxiolytic by its action on the dorsal periaqueductal gray (dlPAG), thereby inhibiting fight-or-flight behaviors. This model helps to explain the role of serotonin in producing panic and anticipatory anxiety, as well as excessive worry associated with generalized anxiety.

Various serotonin receptor subtypes have been shown to be involved in anxiety, in particular, 5-HT$_{1A}$ and 5-HT$_{2A}$. The 5-HT$_{1A}$ receptor density is high in the cortex, hippocampus, amygdala, dlPAG, and raphe nuclei (presynaptic). Animal models initially suggested a role of 5-HT$_{1A}$ receptors in the development of chronic anxiety and helped to generate animal models of anxiety-related disorders. However, it is not clear whether these models constitute relevant models for PD in humans. The development of highly selective 5-HT$_{1A}$ receptor radioligands such as [^{18}F] *trans*-4-fluoro-*N*-2-[4-(2-methoxyphenyl) piperazin-1-yl]ethyl]-*N*-(2-pyridyl)cyclohexanecarboxamide (FCWAY) made it possible to assess central 5-HT$_{1A}$ receptor binding in PD using PET. In PD patients, including those who also had depression, receptors were reduced by an average of nearly one-third in the anterior cingulate in the front middle part of the brain, the posterior cingulate, the rear middle part of the brain, and the raphe, as well as in the midbrain.

Preclinical studies of the genetic expression of the 5-HT$_{1A}$ receptor have shown that rodents lacking the receptor exhibit higher levels of anxiety. The opposite has also been shown in mice that have greater expression of 5-HT$_{1A}$. Of interest, animal studies using models of early-life stress, such as maternal separation, have shown that alterations to the 5-HT$_{1A}$ receptor that occur early in life, such as the prenatal period, can cause persisting changes to the anxious phenotype, even after the restoration of the receptors at a later time. There is no such long-term effect if the receptors are reduced and then reinstated later in adult animals. These findings may be due to the effects of increased CRH and cortisol caused by the early stress, which thereby downregulate the number of 5-HT receptors. There may also be a genetic component.

Studies in humans have used 5-HT$_{1A}$–binding ligands with neuroimaging to determine receptor density; they have reported reduced receptor binding in patients with PD and SAD. Interestingly, antidepressants may exert their anxiolytic effects through different receptor subtypes depending on the anxiety disorder involved. Chronic treatment with SSRIs in PD patients is thought to suppress panic attacks by enhancing the responsiveness of 5-HT$_{1A}$ and 5-HT$_{2A}$ receptors in the dorsal PAG matter, while in patients with generalized anxiety, SSRIs might act through the desensitization of 5-HT$_{2C}$ receptors and, possibly, through stimulation of 5-HT$_{1A}$ receptors in the forebrain.

Analysis of peripheral levels of 5-HT has led to mixed results. Two studies have found lower plasma levels of 5-HT in individuals with PD compared to healthy volunteers. The majority of studies of 5-HT reuptake binding sites in platelets of patients with PD reported no differences compared to controls, while it appears to be decreased in GAD. In individuals with SAD, the density of platelet 5-HT$_2$ receptors has been found to be similar compared to controls, and it was associated with anxiety symptom severity. Levels of 5-HT and its metabolite 5-hydroxyindoleacetic acid (5-HIAA) were also measured in CSF in PD patients and healthy volunteers and did not appear to differ between the two groups.

One method of ascertaining the role of 5-HT in anxiety disorders has been to use pharmacologic challenges such as tryptophan depletion (TD), in which an amino acid mixture deficient in l-tryptophan, a precursor of serotonin, is administered, thereby resulting in an acute decrease in tryptophan and thus a decrease in brain serotonin levels. TD has also been shown to trigger depressive symptoms even in individuals with a family history of depression. While more consistent in individuals with a history of depression, the results of TD challenges in individuals with anxiety disorders have been mixed. In studies of PD patients, TD reduced tryptophan levels and increased anxiety and was associated with a heightened panicogenic effect of CO2 challenge compared to controls. There are, however, studies of patients with PD in which TD did not produce anxiety, including a study using a cholecystokinin-4 (CCK-4) panic challenge. Another pharmacologic challenge of the serotonin system involves the 5-HT agonist mCPP, used as a probe of the 5-HT$_2$ receptor. Some but not all studies of oral and intravenous mCPP have shown increased anxiety in PD after the administration of the drug. Patients with SAD on SSRIs were also reported to have greater sensitivity to TD challenges, whereas another study found that mCPP did not induce a greater rate of panic attacks than in healthy controls.

Serotonergic agents, such as SSRIs, serotonin–norepinephrine reuptake inhibitors (SNRIs; venlafaxine and duloxetine), mirtazapine, and buspirone remain effective treatments for PD, social phobia, and GAD. Recently, novel compounds acting on 5-HT receptors with more favorable tolerability profiles have been tested. For instance, some trials explored the efficacy of vilazodone, an SSRI with partial agonistic properties on 5-HT1A, for GAD. However, evidence from a recent meta-analysis appears to not support the use of vilazodone in GAD. Conversely, one randomized, double-blind, placebo-controlled trial of vilazodone in SAD showed potential efficacy. Gepirone, a selective 5-HT1A receptor partial agonist, has been shown to have anxiolytic properties, and other 5-HT1A agonists have been investigated for the treatment of anxiety disorders. A clinical trial of PRX-00023 was conducted in GAD patients, but it failed to differentiate from placebo. Studies of other selective 5-HT1A partial agonists are underway.

γ-Amino Butyric Acid and the Benzodiazepine System

GABA is the primary inhibitory neurotransmitter in the brain and a metabolic product of glutamate, the excitatory neurotransmitter. There are two subtypes of GABA receptors: GABA-A and GABA-B. GABA-B receptors are slow-acting channels and are agonized by agents such as baclofen (a drug used as a muscle relaxant). GABA-A receptors are fast-acting, ligand-gated channels made up of five protein subunits and consisting of multiple families of receptors. A wide range of drugs act as agonists of the GABA-A receptor, including barbiturates, benzodiazepines, alcohol, anesthetics, neuroactive steroids, and several anticonvulsant medications. Not surprisingly, all of these agents have been shown to have anxiolytic effects. In particular, benzodiazepines have long been used in the treatment of anxiety disorders, in particular PD, phobias, and GAD.

Central benzodiazepine receptors have also been found in the human brain. They form a complex with GABA-A receptors and have different binding sites but positive allosteric effects on one another. The presence of these benzodiazepine receptors led to reports of endogenous benzodiazepine receptor ligands, also called endozepines, in the brain and speculations that they might function in the modulation of fear and learning. The exact role of these compounds in anxiety, however, is yet to be fully understood.

The role of GABA and benzodiazepine receptors in anxiety is well-documented in animal studies. It has been reported that benzodiazepine partial agonists reduce anxiety-like behaviors in rats, whereas GABA inverse agonists produce anxiety in certain learning and behavioral tasks. There is also ample supporting evidence in humans. Studies using benzodiazepine antagonists have shown these agents to be anxiogenic. Flumazenil, an antagonist used in emergency settings to treat benzodiazepine overdose, has been shown to be a panicogenic agent in patients with PD but not in healthy controls. Clinical studies of GABA levels in the blood of patients with anxiety disorders have been limited. Studies of patients with PD compared to healthy controls have failed to show a change in plasma GABA at baseline and after a CCK-4 challenge.

Neuroimaging studies have enabled the visualization of GABA and benzodiazepine receptors in the human brain. Several studies of PD patients have shown reduced GABA-A and benzodiazepine binding, using PET and SPECT imaging, in areas such as the cortex, hippocampus, and insula when compared to controls. A study of PD patients taking medications (antidepressants and/or benzodiazepines) using proton magnetic resonance spectroscopy showed reduced GABA levels in the anterior cingulate and basal ganglia. Although benzodiazepines continue to be among the most widely prescribed medications worldwide, they are no longer considered first-line treatment for PD, SAD, or GAD, having been replaced with SSRIs, cognitive-behavioral therapy, and other types of therapy. Part of this change in prescribing practices is due to concerns about the risk of tolerance, the potential for abuse or dependence, sedative properties, and potential long-term effects on cognition and memory. Despite these concerns, there has been consistent evidence that benzodiazepines are safe and effective in the treatment of anxiety when prescribed appropriately.

Despite the fact that several GABA-A receptor subtype agonists failed to reach the market due to lack of efficacy, evidence from preclinical studies has yielded several agents that are currently in development, including SAGE-217, a GABA-A positive allosteric modulator. While SAGE-217 failed to distinguish itself from placebo for treatment-resistant depression, it is currently being studied for the treatment of GAD. Further, newer agents that enhance GABA function have been explored for the treatment of anxiety disorders. These drugs include several anticonvulsants that work in part as GABA agonists, including tiagabine (Gabitril), a selective GABA reuptake inhibitor (SGRI), and vigabatrin, an inhibitor of GABA transaminase (the latter is not approved for use in the United States). Although controlled studies of tiagabine reported equivocal findings in GAD patients and did not show a difference from placebo in PD, tiagabine appeared effective in improving symptoms of anxiety in a small sample of SAD patients (eight adults aged 21 to 39 years old). Results from clinical trials of other anticonvulsants, such as topiramate, in PD have been mixed. Further studies using larger sample sizes are indicated before the efficacy of these agents as primary treatments for anxiety disorders can be fully ascertained. Gabapentin and pregabalin, although structurally related to GABA, do not act on GABA receptors but instead, bind to a subunit of voltage-sensitive calcium channels.

Neurosteroids, such as allopregnanolone and pregnanolone (see separate section below), have been demonstrated to act as modulators of GABAergic neurotransmission mainly through a stereoselective interaction with GABA-A receptors. In animal models, studies of a synthetic analog of allopregnanolone reported an anxiolytic effect similar to benzodiazepines in the absence of tolerance phenomena. Targeting GABA-A receptors, and in particular the development of compounds selective for GABA-A receptor subtypes, appears to be an intriguing pharmacologic approach for the treatment of anxiety disorders. This approach, also taking advantage of the rapid turnover of GABA-A receptors, could yield a potential drug with high selectivity and reduced adverse effects. To date, results from animal trials appear promising, but further investigations in patients with anxiety disorders are required.

GLUTAMATE SYSTEM

Glutamate is the primary excitatory neurotransmitter of the CNS, is the precursor compound of GABA, and modulates pathways that mediate panic and defensive responses. Results from human and animal studies suggest that panic might arise from an imbalance between tonic

GABAergic inhibition and glutamate-mediated excitation, with predominance of the latter. For example, reduced GABAergic tone has been reported in patients with PD. Further, glutamate is known to regulate other neurotransmitters involved in PD, including serotonin and NE.

Glutamate binds to two families of receptors: metabotropic and ionotropic receptors. Metabotropic receptors (mGluRs), which are G-protein–coupled membrane receptors, are made up of eight subtypes and are believed to serve as allosteric modulators. Ionotropic receptors, which are ligand-gated channels, include *N*-methyl-ᴅ-aspartate (NMDA), α-amino-3-hydroxy-5-methylisoxasole-4-propionic acid (AMPA), and kainate receptors. These receptors have become a focus of investigation in the area of anxiety disorders, primarily PD.

Preclinical studies have shown anxiolytic properties of group-2 and group-3 mGlu-receptor agonists, and, more recently, of selective group-2 receptor type 2 (mGluR2) agonists. Systemic infusion of these glutamate receptor agonists prevented panic-like responses after infusion of sodium lactate in panic-vulnerable rats, showing efficacy against panic similar to that of comparison drug alprazolam in this animal model of panic. Despite several studies showing the anxiolytic effect of mGlu modulators in animal models, trials on human samples have not been as promising. With regard to human studies, a small randomized, double-blind clinical trial of an mGluR2/3 agonist, LY354740, in patients with PD, failed to show treatment effects compared to placebo. Similarly, JNJ-40411813 (ADX-71149), an mGluR2 positive allosteric modulator, was studied in an RCT for anxious depression but did not show efficacy. Another randomized, double-blind, placebo-controlled trial of the mGluR2/3 agonist, LY354740, in patients with GAD showed greater improvement in anxiety scores compared to placebo, but the study was discontinued due to findings of convulsions in animal studies of the compound. This is an understandable risk, given that glutamate-enhancing drugs have neuroexcitatory properties and that glutamate inhibitors are commonly used as anticonvulsants. Similarly, there are studies of mGlu1, mGlu5, and mGluR7 antagonists and AMPA/kainate–receptor blockers showing antianxiety properties in animals. Further studies in humans are needed before the safety and tolerability of these compounds can be better understood.

NMDA receptors are prominently involved in the conditioning of fear learning via entry of calcium through the receptors and voltage-gated calcium channels. Blockade of the NMDA receptor with antagonists such as (2R)-amino-5-phosphonovaleric acid (APV) can block fear acquisition, and some studies show that it may even block expression. In panic-prone rats, injection of NMDA antagonist MK-801 into the dorsomedial/perifornical hypothalamus blocked panic-like behavior induced by sodium lactate in a dose-dependent fashion. The NMDA receptor antagonist ketamine, which at subanesthetic doses has demonstrated efficacy in patients with treatment-resistant depression, has shown efficacy for PTSD in recent studies and mixed results in patients with obsessive-compulsive disorder, both conditions formerly classified as anxiety disorders. A recent randomized clinical trial investigating the effect of repeated ketamine infusions compared to the active control midazolam for chronic PTSD showed a significantly greater improvement in PTSD symptoms severity in the ketamine group from baseline to week 2 (following the six infusions), with a median time to loss of response of 27.5 days following the last infusion. Similarly, few studies have explored the effect of ketamine in patients with conditions currently classified as anxiety disorders. Two small preliminary studies (one open-label and the other exploratory double-blind, compared to midazolam) evaluated the efficacy of ketamine in patients with treatment-refractory SAD and/or GAD without current depression, using an ascending dose design (0.25, 0.5, 1 mg/kg subcutaneously), administered weekly; in both studies, ketamine was associated with a quick reduction in anxiety symptoms, lasting up to 7 days. A double-blind, proof-of-concept RCT with a crossover design tested the effect of intravenously administered ketamine (0.5 mg/kg) compared to saline (administered 28 days apart) in patients with SAD. In this study, ketamine was associated with significantly greater improvement in anxiety than saline in the first 2 weeks following infusion, measured with the Liebowitz Social Anxiety Scale.

NMDA receptors are also critical for fear extinction, with preclinical studies showing that blockade of this pathway inhibits extinction. These findings were supported by animal research showing that a partial agonist at the glycine modulatory site on the NMDA receptor, D-cycloserine (DCS), a drug previously used for the treatment of tuberculosis, could facilitate extinction learning. D-cycloserine is among the most widely studied glutamatergic agents in anxiety. Healthy volunteers receiving DCS during a stressful learning task had less stress and better performance than those receiving placebo. In addition, a small study using functional magnetic resonance imaging showed that healthy volunteers receiving DCS had a more blunted amygdala response to fearful faces compared to those receiving placebo. Several studies have shown the efficacy of DCS as an adjunct to exposure therapy or cognitive behavioral therapy (CBT) in patients with acrophobia (fear of heights), SAD, and PD. Negative studies have also been reported, including a study of DCS in spider phobia. A recent review reported no difference between DCS and placebo in augmentation of psychotherapy for anxiety disorders in children and adolescents. Overall, results from DCS augmentation have been inconsistent, with initial trials showing promising results that have not been replicated in larger studies.

Another potential strategy consists of inhibiting the glycine transporter, thereby increasing the availability of glycine at the NMDA receptor complex glycine modulatory site. Glycine transporter inhibitor Org 25935 was evaluated in a multicenter clinical trial as augmentation to CBT in patients with PD; panic symptoms decreased in every treatment group, but there was no statistically significant difference between Org 25935 and placebo.

Other glutamate-modulating agents, such as pregabalin, tiagabine, and lamotrigine (Lamictal), have been studied for GAD and SAD. Although pregabalin demonstrated efficacy in several randomized controlled trials in patients with GAD, it has not received formal approval by the U.S. Food and Drug Administration (FDA) for the treatment of this condition. Pregabalin, however, is recommended as a first line of treatment for GAD comorbid with epilepsy. Riluzole (Rilutek), a drug approved for amyotrophic lateral sclerosis, was studied in an 8-week, open-label trial in patients with GAD and found to be effective in lowering anxiety scores. Functional neuroimaging studies of open-label riluzole in patients with GAD showed changes in hippocampal volumes and N-acetylaspartate (NAA) concentrations that correlated with improvement in anxiety symptoms in patients compared with healthy volunteers. More recently, an analog of riluzole, troriluzole (BHV-4157), failed to differentiate from placebo in a large Phase III 8-week trial in GAD. The study of glutamate in larger-scale human studies should continue to provide novel therapeutic agents for anxiety disorders.

NEUROSTEROIDS

Baulieu introduced the term “neurosteroid” in the 1980s to describe endogenous steroids synthesized in the CNS from cholesterol or sterols. Neurosteroids act on neuronal function and regulate neuronal excitability. They are highly selective and positive allosteric modulators of the GABA-A receptor, thereby making them potential anxiolytic agents, a finding that has been supported in animal models. Neurosteroids are thought to act on the amygdala by affecting intracellular receptors in a process involving NMDA receptors. They additionally regulate CRH gene expression, thereby reversing the anxiogenic effects of CRH. The progesterone metabolites believed to be involved in fear and anxiety include 3α,5α-tetrahydroprogesterone (THP) and 3α,5β-THP (pregnanolone).

Preclinical studies have shown that progesterone is anxiolytic, likely due to its effect of increasing allopregnanolone. Progesterone is increased during pregnancy, and it has been shown in animal studies that this is associated with a reduction in anxiety behaviors and that this anxiolytic effect is reversed by finasteride, an inhibitor of the 5α-reductase enzyme that is involved in steroid biosynthesis. One study in humans reported that pregnant women had higher levels of progesterone, allopregnanolone, and tetrahydrodeoxycorticosterone (THDOC), which were believed to be protective against maternal anxiety. A study of pregnant women with preexisting PD found a reduction in anxiety symptoms during the pregnancy and rebound anxiety during the postpartum period, which is consistent with a rapid drop in hormone levels. The effect of hormone levels on anxiety has been studied during the menstrual cycle as well. One study reported that women with PD had elevated peripheral concentrations of allopregnanolone both during the follicular and premenstrual phase and suggested that neurosteroids may be a protective factor against spontaneous panic attacks. Of note, the FDA-approved brexanolone for postpartum depression, a condition characterized by a rapid decrease in the concentration of allopregnanolone and progesterone after childbirth, especially in women with postpartum “blues.” A single infusion of brexanolone showed rapid antidepressant effects, as documented in an open-label and two double-blind, randomized clinical trials (RCTs).

Patients with PD have been found to have increased allopregnanolone levels and decreased isopregnanolone levels. In a study of panic attacks induced by CCK-4 and lactate, patients with PD showed a decrease in allopregnanolone and pregnanolone, suggesting decreased GABA tone not found in healthy controls. It is possible that alterations in neurosteroid levels might be related to the pathophysiology of panic attacks in PD and not simply reflects state anxiety. No data are available concerning alteration of neurosteroid levels during spontaneous panic attacks. Studies investigating the impact of pharmacotherapy on allopregnanolone levels in PD patients revealed conflicting results, with one trial showing no effect of paroxetine on allopregnanolone levels or concentrations of other neuroactive steroids, but another one showing a significant increase in allopregnanolone levels after 2 months of treatment with paroxetine.

Studies of neurosteroid levels in patients with other anxiety disorders have yielded inconsistent findings. Two small-scale investigations reported low pregnanolone sulfate levels in both SAD and GAD in both cases when compared to healthy controls. One group, however, demonstrated no difference in levels of DHEA, DHEA sulfate (DHEAS), or ALLO in patients with SAD compared to healthy volunteers. Two compounds with neurosteroidal activity, mifepristone (RU486) and PH94B, have been investigated for the treatment of anxiety symptoms (see “Hypothalamic–Pituitary–Adrenal Axis” section for more details). Studies of synthetic analogs of neurosteroids are underway, with no published studies in anxiety disorders to date.

ARGININE VASOPRESSIN

Arginine vasopressin (AVP) is released by the parvocellular neurons found in the paraventricular nucleus (PVN). AVP is best known for its effects on the control of body fluid levels, blood pressure, and blood sugar. Vasopressin is used as a treatment for nocturnal enuresis as well. It has also been linked to pain sensitivity, circadian rhythms, regulation of sleep, and learning and cognition. AVP exerts its effect through two receptors, V1a and V1b, located throughout the brain, including the septum, cortex, and hippocampus.

Although CRH is believed to be the critical factor in controlling the release of ACTH, AVP serves as an important mediator. The release of ACTH during periods of acute stress, as in rat studies using the forced swim test, can be suppressed by simultaneous administration of a CRH receptor 1 antagonist and an AVP V1b antagonist. Mice with the V1a receptor knock-out showed reduced anxiety on similar stress tasks. Similarly, mice with overexpression of the V1a receptor exhibited heightened anxiety-like behaviors. The role of AVP in anxiety is supported by a human genetic study, which found an association between an SNP within the AVP V1b receptor gene and PD, compared with healthy controls. With regard to treatment, preclinical studies have yielded mixed results of AVP V1b antagonists on anxiety-like behaviors. Clinical studies investigating a V1b antagonist in patients with GAD failed to demonstrate clinical efficacy. A few studies testing the effects of SRX246, an AVP V1a receptor antagonist, have yielded initial encouraging results. In one of these, a 7-day treatment with SRX246 was associated with a reduction of amygdala response to angry faces following an acute intravenous administration of vasopressin in a sample of 29 healthy males.

In addition, there are limited reports of AVP administration in anxiety disorders. A trial in healthy male and female volunteers found that the administration of intranasal AVP reduced skin conductance and heart rate responses during an angry-faces paradigm and that there was a difference between men and women in the social recognition paradigm, supporting the preclinical findings of sex differences in AVP.

OXYTOCIN

Oxytocin, a 9–amino acid peptide hormone synthesized in the paraventricular and supraoptic nuclei of the hypothalamus, and released from neurons within the hypothalamus and also into the bloodstream via the posterior pituitary, has been shown to interact with central reward processing regions, modulate social-emotional functions, and attenuate stress responses via interaction with the HPA axis. The anxiolytic, stress-attenuating, and prosocial effects of oxytocin were first discovered in rodents. Anxiogenic and stressful stimuli activate the synthesis and release of oxytocin, which modulates the HPA axis and anxiety behaviors.

In human studies, circulating plasma oxytocin level has been associated with anxiety, while the CSF level of oxytocin appears to be negatively associated with anxiety in children and adult subjects. Additional support for the role of oxytocin in anxiety is also suggested by a report of an association between polymorphisms in the oxytocin receptor gene and increased risk for anxiety, stress, and depression in individuals with early-life stress exposure. Further, intranasal administration of oxytocin has been associated with increased prosocial behavior and improvement in emotion recognition and social memory. Its anxiolytic and stress-attenuating effects have stimulated significant interest in its potential therapeutic effect for anxiety disorders. A study using intranasal oxytocin in healthy male volunteers reported reduced amygdala activation to fearful faces, lower anxiety scores, and improved social cognition compared to intranasal placebo. In another study of men with early parental separation, a risk factor for anxiety and depressive disorders in adults, intranasal oxytocin administration attenuated decreases in salivary cortisol compared to placebo, thus dampening the HPA axis alteration associated with early separation.

A series of studies in healthy and anxious volunteers have shown decreased anxiety symptoms during variations of the Trier Social Stress Test (TSST) after intranasal oxytocin administration. In one study of patients with GAD, daily oxytocin administration over 3 weeks was beneficial, especially in male patients. Oxytocin facilitates extinction of social fear. In patients with SAD, intranasal oxytocin alone or as augmentation of exposure therapy was associated with more positive self-evaluation after public performance. In a recent neuroimaging study in patients with SAD, oxytocin administered intranasally was associated with greater enhancement of amygdala-frontal resting state connectivity in patients with higher social anxiety. Oxytocin also enhanced functional connectivity between the amygdala and other brain regions (insula, cingulate gyrus) in patients with SAD but not in healthy controls in response to fearful faces. Oxytocin has also been shown to attenuate the neural response to a negative social interaction task toward other same-sex humans in men, within the amygdala and anterior insula, as measured with fMRI, suggesting a potential effect of oxytocin in reducing the stress of negative social interactions among healthy men. Interestingly, this effect was not detected in female participants. A study in patients with SAD and healthy controls found that decreased methylation of the oxytocin receptor gene was associated with SAD, increased cortisol response to the TSST, and increased amygdala reactivity during processing of words relevant to social anxiety. A trial is currently underway to explore the levels of oxytocin prestressor and after-stress exposure using the TSST in children and adolescents with anxiety disorders. Further studies of oxytocin administration in patients with SAD are necessary to develop this intervention for potential use in clinical practice.

NEUROPEPTIDE Y

Neuropeptide Y (NPY), a 36–amino acid peptide, is one of the most abundant and highly evolutionarily conserved polypeptides found in the brain. Its highest concentrations are in the LC, hypothalamus, septum, and PAG, with moderate levels in the hippocampus, amygdala, and brainstem. NPY is linked with various roles in humans, such as feeding, sleep, and anxiety regulation, although via distinct mechanisms. There is ample preclinical evidence that NPY has anxiolytic effects. Of the five NPY receptor subtypes found in mammals (Y1 to Y5), the NPY-Y1 receptor has been most closely studied in anxiety. Studies of Y-1–receptor knockout mice and administration of Y-1 antagonists increased anxiety-like behaviors in the animals. In addition, NPY has been shown to be involved in fear consolidation, with preclinical studies showing that the administration of NPY impairs the retention of traumatic memories and reduces anxiety during stressful tasks. NPY also mediates the response to chronic stress by increasing expression of amygdala NPY mRNA. A study of rats exposed to maternal separation reported that treatment with acupuncture increased NPY expression in the basolateral amygdala and reduced anxious behaviors. Recent evidence from preclinical studies also suggests a role for NPY in mood regulation. Mice with decreased NPY gene expression in the NAc exhibited behavioral despair. Anxiety-like behavior appeared significantly increased following toxin-mediated ablation of NAc NPY neurons, while chemogenetic activation of NAc NPY neurons reduced anxiety behavior in mice. These data suggest that NPY may play a role in the psychobiology of stress responses and resilience.

Few studies have investigated the effects of NPY administration in clinical samples. The susceptibility of NPY to gastrointestinal degradation and the limited transportation through the blood–brain barrier had implications on the chosen route of peptide administration. Thus, an intranasal direct nose-to-brain route may represent the most clinically relevant procedure to explore the effect of NPY on the clinical population. The only study to date on NPY administration for anxiety was conducted in individuals with chronic PTSD. In this study, 26 subjects were randomized to receive either intranasal NPY or placebo using a cross-over, ascending dose design. The study sample was divided into five cohorts reflecting the single ascending dose study: 1.4 mg (n = 3), 2.8 mg (n = 6), 4.6 mg (n = 5), 6.8 mg (n = 6), and 9.6 mg (n = 6). Each subject received intranasal NPY or intranasal placebo saline on separate treatment days 1 week apart. Assessments were conducted pretreatment and following a trauma script symptom provocation procedure after receiving the study drug or placebo. Safety and tolerability of NPY represented the primary outcome of the study, while NPY efficacy on anxiety and PTSD symptoms severity as a function of dose were explored as secondary outcomes. Twenty-four subjects completed both treatments, and NPY was well-tolerated up to the highest dose. Higher doses of NPY were associated with greater improvement in anxiety symptoms compared to placebo, as measured with the Beck Anxiety Inventory, immediately following symptom provocation. Other measures of anxiety (State-Trait Anxiety Inventory–State Form) and PTSD (Impact of Event Scale-Revised) symptoms did show a statistically significant difference between NPY and placebo, although the pattern of the dose-response relationship was similar. Recently, a small trial showed potential benefit of intranasal NPY administration for individuals with depression. In this study, 30 subjects with major depressive disorder (MDD) on a stable dose of a conventional antidepressant received 6.8 mg NPY ($n = 12$) or placebo ($n = 18$) in a double-blind, randomized fashion. Effects on depressive symptoms severity were collected at several time points (baseline, 1, 5, 24 and 48 hours following administration) using clinician-administered rating scales (Montgomery-Åsberg Depression Rating Scale). NPY was superior to placebo at the 24-hour time point and showed similar a trend 1- and 5 hours following administration, albeit not reaching significance. However, the improvement in depression severity at 48 hours was no longer significant. While there are currently no active trials of NPY administration for GAD, PD, or SAD, the encouraging initial findings in PTSD and MDD suggest the need for further studies in individuals with mood and anxiety disorders.

GALANIN

Galanin, a 30–amino acid peptide, has a wide variety of functions in humans, including modulation of pain, learning, feeding and gastrointestinal motility, and cardiovascular control. Galanin, which is coexpressed with NE neurons in the LC, and thus projects to structures involved in the fear pathway (including the amygdala, hippocampus, and PFC), has been linked to regulation of mood and fear response. It also acts as an autoreceptor to reduce firing of the LC. Galanin acts on three subtypes of receptors: Gal-R_1, Gal-R_2, and Gal-R_3. Although preclinical data are inconsistent, the three receptors have been shown to have differing effects on anxiety-like behaviors. Mice lacking the Gal-R_1 receptor exhibited anxiety-like behaviors, whereas the injection of receptor agonists produced a reduction of such behaviors. Similarly, studies of Gal-R_3 antagonists in rodents showed them to have antidepressant and anxiolytic properties in several behavioral tasks. Overall, increases in central galanin, either by overexpression or via intracerebral administration, have been demonstrated to regulate anxiety behaviors in rats. More recent studies suggest that galanin might have anxiolytic effects under high stress/high noradrenergic activation. A trial in mice using galnon, a nonselective galanin-receptor agonist, showed that the drug reduced anxiety during certain stressful tasks, such as the forced swim test and the elevated zero maze, and that this effect was reversed by administration of a galanin-receptor antagonist. This finding has also been reported in another animal study using forced swim, a model of depression. The reports suggest that galanin's effects on mood and anxiety may be through its inhibition of hippocampal serotonin receptors.

There are very few human studies of galanin in patients with anxiety disorders. A genetic study of men and women with PD found an association between two of six haplotypes of a galanin gene polymorphism and severity of illness, but only in women; this finding was later replicated in a larger sample. Despite the therapeutic potential of galanin, there are no known published trials of galanin modulators as potential therapeutic agents in anxiety disorders.

CHOLECYSTOKININ

CCK is a neuropeptide derived from a 112–amino acid precursor compound and is classified by its various fragment peptides, of which CCK octapeptide (CCK-8) is the most abundant. CCK helps regulate the digestion of proteins and lipids in the gastrointestinal system. However, CCK is also found in the brain, in key regions of the fear network (amygdala, hypothalamus, periaqueductal grey), and interacts with several neurotransmitters and neuropeptides such as serotonin, noradrenaline, GABA, endocannabinoids, and NPY. CCK acts via two types of receptors, CCK-1 and CCK-2 (also referred to as CCK_A and CCK_B, respectively). Specifically, CCK-2 has been linked with anxiety-like behaviors in mice and with panic attacks in humans. Further, while CCK-8 acts on both receptor subtypes, CCK tetrapeptide (CCK-4) is a primary agonist of CCK-2 receptors. CCK has its highest levels in the cerebral cortex, hippocampus, amygdala, caudate, and putamen, with intermediate levels in the thalamus and hypothalamus. CCK receptors are found in greatest density in the cortex, hypothalamus, substantia nigra, and PAG.

It has been shown in various preclinical studies that CCK-4 is anxiogenic and that this effect is blocked by CCK-2 antagonists. CCK-8 has shown a similar effect on inducing anxiety in rats, but only after intracranial administration, suggesting that the drug does not cross the blood–brain barrier as effectively as CCK-4. Studies in humans have also shown similar results. Studies of both healthy subjects and patients with PD have shown the panicogenic effects of CCK-4 and of CCK-5, or pentagastrin. These effects could be blocked by the administration of CCK antagonists or by treatment with GABA agonists such as vigabatrin or tiagabine. In addition, patients with PD have been shown to have lower CSF levels of CCK versus healthy controls, although it is unclear whether this is due to higher CCK turnover or greater receptor sensitivity. In addition, a recent study of healthy volunteers undergoing a CCK-4 challenge reported that cardiovascular changes may not be correlated with subjective panic, which suggests that this may not be an ideal model for studying the effect of anxiolytic drugs on panic unless strict criteria are used.

The aforementioned studies supported the role of CCK in anxiety and prompted the investigation of CCK-2–receptor antagonists as potential antianxiety drugs. Although several CCK-2 receptor antagonists have been developed, none have been demonstrated to be effective in improving anxiety symptoms in humans. In a randomized, double-blind, placebo-controlled study, the effect of CI-988, a CCK-2 antagonist, on anxiety scores in GAD patients did not differ from placebo. Another study in GAD patients found that CI-988 failed to abate the anxiogenic effects of mCPP, and a similar study showed that it failed to treat the anxiety symptoms induced by CCK-4 in patients with PD. Trials of CCK-2 antagonists in PD showed that they were not significantly better than placebo. This has led to the suggestion that the preclinical data were not as clear-cut in defining the significance of CCK-2 in anxiety modulation. The role of CCK as a possible target of novel therapeutics remains an area of investigation.

ENDOCANNABINOID SYSTEM

Preclinical studies have stimulated a growing interest in the endocannabinoid system, involved in modulating sleep, appetite, pain, emotional memory, and fear responses, as a potential therapeutic target in patients with anxiety disorders. This system comprises polyunsaturated fatty acid endogenous ligands, anandamide (AEA), 2-arachidonoylglycerol (2-AG), and two G-coupled receptors, cannabinoid receptors 1 (CB1) and 2 (CB2). CB1 and CB2 receptors are widely distributed in the frontal cortex and subcortical structures, including the amygdala, basal ganglia, hippocampus, and other structures. In particular, CB1 agonists show a biphasic effect characterized by anxiolytic properties at low doses and anxiogenic properties at high doses.

Findings from preclinical studies suggest that basal endocannabinoid levels are low and increase during exposure to threatening stimuli, counteracting fear responses under conditions of threat. In humans, new studies are focusing on targeting this system in patients with PTSD, which is no longer classified as an anxiety disorder. Cannabidiol (CBD), a cannabis sativa constituent, binds to the CB1 receptor and several others, including the serotonin 5-HT$_{1A}$ receptor. CBD has been studied in animal models of general anxiety and stress-induced anxiety. In human studies, CBD significantly reduced anxiety during simulated public speaking in healthy volunteers and patients with SAD. Conversely, rimonabant, a CB1 antagonist/inverse agonist, increased anxiety in anticipation and during simulated public speaking in healthy volunteers. Further, in a recent study of adolescents with SAD, CBD was associated with an improvement in anxiety symptoms compared to placebo. Currently, a clinical trial exploring the effect of flexible dosages of CBD in patients with PD, GAD, SAD, or agoraphobia is in progress.

While direct CB receptor modulators might potentially worsen anxiety or depression, inhibitors of fatty acid amide hydrolase (FAAH), the principal catabolic enzyme for AEA, show promise. Reduced FAAH expression has been associated with enhanced extinction learning in animal and human neuroimaging studies. Knock-in mice with a polymorphism associated with reduced expression and activity of FAAH exhibited enhanced fronto-amygala connectivity, increased fear extinction learning, and decreased anxiety-like behaviors. However, the role of cannabinoid receptors in mediating memory consolidation and reconsolidation processes appears to be complex, considering that both endocannabinoid agonists and antagonists have been involved in fear memory reconsolidation. Further, administration of FAAH inhibitors to rodents was associated with decreased anxiety-like behaviors during experimental tasks designed to increase anxiety, such as the elevated-plus maze and the light-dark box. Clinical trials of these novel compounds are needed in social anxiety and other anxiety disorders.

The implication of endocannabinoids in anxiety is in accordance with dense expression of CB1 receptors in brain regions related to anxiety, fear, and aversion, including the mPFC, hippocampus, amygdala, and PAG. As one example supporting the role of the endocannabinoid system in anxiety, CB1 receptor availability in the amygdala was found to mediate attention bias to threat cues in a sample of PTSD patients. A partial agonist of CB1, delta-9-tetrahydrocannabinol (THC), appears to have anxiolytic effect in low doses but can induce anxiety and panic attacks in high doses. Similarly, nabilone is a synthetic cannabinoid with high affinity to CB1 and CB2 receptors. Studies in humans assessing the effect of nabilone on pain and anxiety symptoms showed discordant results, with some trials reporting an improvement in pain symptoms and a decrease in anxiety severity and others that did not show an effect on anxiety symptoms. Two trials studied the effect of nabilone in patients with anxiety symptoms. One enrolled eight anxious volunteers who received placebo and three dose levels of nabilone at 1-week intervals using a single-blind design and showed anxiolytic effects only in a small subgroup, while in another small open-label study, patients with anxiety were treated with nabilone for 28 days and had significant improvement in anxiety symptoms. Finally, preclinical studies showed that anxiogenic-like effects of CB1 antagonists tend to be more evident when animals are subjected to high levels of aversion. Anandamide-hydrolysis inhibitors are more efficacious as anxiolytic drugs when tested in a highly aversive environment. The basal levels of endocannabinoid synthesis and release tend to be low; however, the activity of this system is enhanced in response to neural activation when experimental animals are exposed to threatening stimuli when endocannabinoids would work to counteract.

INFLAMMATORY AND IMMUNOLOGIC FACTORS

Recent evidence suggests a role of inflammatory processes in the pathophysiology of anxiety, with few studies reporting elevated levels of circulating inflammatory markers in anxiety disorders. Stress leads to activation of the HPA axis and the sympathetic nervous system, with the resulting release of catecholamines, which in turn modulate the release of inflammatory mediators (such as cytokines) through α- and β-adrenoceptors displayed on immune cells. Although some studies have observed higher circulating levels of cytokines in individuals with PD, others have failed to replicate this finding and instead showed reduced plasma levels of some cytokines (such as Interferon-γ), similar to findings in animal models. While a few studies have observed higher levels of C-reactive protein (CRP) in GAD patients, CRP levels appear reduced in SAD patients, especially in females. However, the limited number of studies conducted in individuals with anxiety disorders and the presence of multiple confounding factors (age, gender, BMI) that may interfere with the concentration of circulating cytokines preclude any final conclusions on the role of inflammation in the pathophysiology of these disorders. Further studies in this area are warranted.

FUTURE DIRECTIONS

Knowledge about anxiety disorders has arisen from various animal models of stress and fear, using primarily Pavlovian fear conditioning as a basis of understanding. Although such studies have provided invaluable information about the mechanisms of the fear response, translating this work to humans remains challenging. Anxiety disorders are highly unique in their etiology, susceptibility, presentation, treatment, and prognosis. Preclinical research will continue to refine the current understanding of the role and integration of neurochemical systems both in producing fear responses and in modulating anxiety responses in animal models. Additional progress will come from optimizing current animal models of anxiety and incorporating emerging understandings of the intricate and fine-grained neural circuits involved in the experience of anxiety. Further, animal studies are employing novel techniques including in vivo chemogenetics and optogenetics, coupled with retrograde tracing to probe neuro-biologic systems involved in anxiety. New insights from basic research can then be applied to human studies in the search for novel biomarkers of anxiety disorders and new treatment approaches. A compelling example that has recently emerged involves the identification of a neural circuit in the amygdala, and the cellular mechanisms within this circuit, that give rise to anxiety. Through a series of experiments in mice, researchers have identified this circuit within the central amygdala as critical to the emergence of anxiety and have begun to elucidate the role of dynorphin in reducing inhibition of somatostatin-expressing (SOM+) neurons in the central amygdala. This work has begun to reveal the potential role of dynorphin in the aberrant functioning of these neurons within the central amygdala circuit as well as the role of these SOM+ neurons' aberrant activation in the experience of anxiety. An increased understanding of these circuits is likely to inform the treatment potential of dynorphin inhibition for anxiety in humans.

Future research in humans will benefit from continued fine-tuning of intermediate phenotypes for anxiety disorders, yielding a more refined clinical categorization of disorders that will facilitate translational research. During the last two decades, a growing number of studies have investigated target-selective agents to enhance the benefits of exposure-based or other forms of psychotherapy. Mechanistically novel compounds targeting glutamate, neuropeptide, and endocannabinoid systems are already under development and will continue to be tested in clinical trials. In addition to the neurochemical systems reviewed in this chapter, potentially promising avenues for future research include the neurotrophin, renin–angiotensin, and acetylcholine systems, as well as the immune system, among others. Finally, a new wave of studies in humans that integrates neural circuitry data with burgeoning work in psychiatric genomics promises to expand the current understanding of anxiety disorders. This work is likely to yield novel and more specific biomarkers of anxiety disorders that can lead to more tailored treatment interventions for these highly common and debilitating mental disorders.

FURTHER READINGS

Bandelow B, Baldwin D, Abelli M, et al. Biological markers for anxiety disorders, OCD and PTSD: a consensus statement. Part II: neurochemistry, neurophysiology and neurocognition. *World J Biol Psychiatry*. 2017;18(3):162–214.

Bukalo O, Pinard CR, Holmes A. Mechanisms to medicines: elucidating neural and molecular substrates of fear extinction to identify novel treatments for anxiety disorders. *Br J Pharmacol*. 2014;171:4690–4718.

Chen X, Hackett PD, DeMarco AC, et al. Effects of oxytocin and vasopressin on the neural response to unreciprocated cooperation within brain regions involved in stress and anxiety in men and women. *Brain Imaging Behav*. 2016;10(2):581–593.

Dunsmoor JE, Paz R. Fear generalization and anxiety: behavioral and neural mechanisms. *Biol Psychiatry*. 2015;78:336–343.

Farb DH, Ratner MH. Targeting the modulation of neural circuitry for the treatment of anxiety disorders. *Pharmacol Rev*. 2014;66:1002–1032.

Garakani A, Murrough J, Freire R, et al. Pharmacotherapy of anxiety disorders: current and emerging treatment options. *Front Psychiatry*. 2020;11:595584.

Graeff FG, Zangrossi H Jr. The dual role of serotonin in defense and the mode of action of antidepressants on generalized anxiety and panic disorders. *Cent Nerv Syst Agents Med Chem*. 2010;10:207–217.

Griebel G, Holmes A. 50 years of hurdles and hope in anxiolytic drug discovery. *Nature Rev*. 2013;12:667–687.

Guastella AJ, Richardson R, Lovibond PF, et al. A randomized controlled trial of D-cycloserine enhancement of exposure therapy for social anxiety disorder. *Biol Psychiatry*. 2008;63:544–549.

Harvey BH, Shahid M. Metabotropic and ionotropic glutamate receptors as neurobiological targets in anxiety and stress-related disorders: focus on pharmacology and preclinical translational models. *Pharmacol Biochem Behav*. 2012;100:775–800.

Kalk NJ, Nutt DJ, Lingford-Hughes AR. The role of central noradrenergic dysregulation in anxiety disorders: evidence from clinical studies. *J Psychopharmacol*. 2011;25:3–16.

Kent JM, Daly E, Kezic I, et al. Efficacy and safety of an adjunctive mGlu2 receptor positive allosteric modulator to a SSRI/SNRI in anxious depression. *Prog Neuropsychopharmacol Biol Psychiatry*. 2016;67:66–73.

Kische H, Ollmann TM, Voss C, et al. Associations of saliva cortisol and hair cortisol with generalized anxiety, social anxiety, and major depressive disorder: an epidemiological cohort study in adolescents and young adults. *Psychoneuroendocrinology*. 2021;126:105167.

Kormos V, Gaszner B. Role of neuropeptides in anxiety, stress, and depression: from animals to humans. *Neuropeptides*. 2013;47:401–419.

Linares IM, Chags MHN, Machado-de-Sousa JP, Crippa JAS, Hallak JEC. Neuroimaging correlates of pharmacological and psychological treatments for specific phobia. *CNS Neurol Disord Drug Targets*. 2014;13:1021–1025.

Long Z, Medlock C, Dzemidzic M, Shin YW, Goddard AW, Dydak U. Decreased GABA levels in anterior cingulate cortex/medial prefrontal cortex in panic disorder. *Prog Neuropsychopharmacol Biol Psychiatry*. 2013;44:131–135.

Morrison FG, Ressler KJ. From the neurobiology of extinction to improved clinical treatments. *Depress Anxiety*. 2014;31:279–290.

Neumann ID, Slattery D. Oxytocin in general anxiety and social fear: a translational approach. *Biol Psychiatry*. 2016;79(3):213–221.

Papagianni EP, Stevenson CW. Cannabinoid regulation of fear and anxiety: an update. *Review Curr Psychiatry Rep*. 2019;21(6):38.

Perna G, Schruers K, Alciati A, Caldirola D. Novel investigational therapeutics for panic disorder. *Expert Opin Investig Drugs*. 2015;24:491–505.

Petrowski K, Schmalbach I, Strunk A, Hoyer J, Kirschbaum C, Joraschky P. Cortisol reactivity in social anxiety disorder: a highly standardized and controlled study. *Psychoneuroendocrinology*. 2021;123:104913.

Preter M, Klein DF. Panic, suffocation false alarms, separation anxiety and endogenous opioids. *Prog Neuropsychopharmacol Biol Psychiatry*. 2008;32:603–612.

Prevot TD, Li G, Vidojevic A, et al. Novel benzodiazepine-like ligands with various anxiolytic, antidepressant, or pro-cognitive profiles. *Mol Neuropsychiatry*. 2019;5(2):84–97.

Santos M, D'Amico D, Dierssen M. From neural to genetic substrates of panic disorder: Insights from human and mouse studies. *Eur J Pharmacol*. 2015;759:127–141.

Schüle C, Nothdurfter C, Rupprecht R. The role of allopregnanolone in depression and anxiety. *Prog Neurobiol*. 2014;113:79–87.

Singewald N, Schmuckermair C, Whittle N, Holmes A, Ressler KJ. Pharmacology of cognitive enhancers for exposure-based therapy of fear, anxiety and trauma-related disorders. *Pharmacol Ther*. 2015;149:150–190.

Slee A, Nazareth I, Bondaronek P, Liu Y, Cheng Z, Freemantle N. Pharmacological treatments for generalised anxiety disorder: a systematic review and network meta-analysis. *Lancet*. 2019;393:768–777.

Turna J, Patterson B, Van Ameringen M. Is cannabis treatment for anxiety, mood, and related disorders ready for prime time? *Depress Anxiety*. 2017;34:1006–1017.

Van't Veer A, Carlezon WA Jr. Role of kappa-opioid receptors in stress and anxiety-related behavior. *Psychopharmacology (Berl)*. 2013;229:435–452.

Zorumski CF, Paul SM, Covey DF, Mennerick S. Neurosteroids as novel antidepressants and anxiolytics: GABA-A receptors and beyond. *Neurobiol Stress*. 2019;11:100196.

14.5 Neuroimaging and the Neuroanatomical Circuits Implicated in Anxiety Disorders

ELIZABETH R. KITT, M.S., M.PHIL., EMILY M. COHODES, M.S., M.PHIL., AND DYLAN G. GEE, PH.D.

Convergent advances in techniques for examining the biological correlates of emotional behavior in animals (i.e., neurophysiological, lesion, and molecular techniques) and fear and anxiety in humans (i.e., neuropharmacological and neuroimaging approaches) have exponentially improved understanding of the neuroanatomical, cellular, and molecular correlates of fear and anxiety. By capitalizing on the unique advantages of these two streams of knowledge, prior research has elucidated the neural circuitry involved in adaptive fear and anxiety. Building off this foundation, a growing body of research has begun to explicate how dysfunction in these circuits may contribute to maladaptive anxiety in the form of anxiety disorders. Embedded in this research is a growing understanding of the neurodevelopmental processes that may contribute to risk for pediatric anxiety disorders.

Originally extrapolated from fear-conditioning research in animals, prevailing models of neural circuitry underlying anxiety disorders in humans may now be directly examined through advances in techniques of in vivo human imaging. These models emphasize interactions within limbic circuitry (especially the prefrontal and sensory cortical regions), the amygdala, hippocampus, hypothalamus, bed nucleus of stria terminalis (BNST), and periaqueductal gray (PAG). This section reviews prevailing models of neuroanatomical circuits implicated in fear and stress-related disorders. Through an emphasis on human neuroimaging studies, this section further demonstrates how modern human neuroimaging approaches provide essential insight into the development and dysfunction of neural circuitry underlying these disorders.

BRIDGING FROM PHENOMENOLOGY TO NEUROCIRCUITRY

Characterizing the neural circuitry underlying anxiety disorders involves a series of stepwise progressions in understanding that ultimately bridge knowledge about phenomenology with growing insight into abnormalities in neurocircuitry. Specifically, previous work has identified neural circuitry underlying these disorders by: (1) examining the phenomenology of the disorder to identify functional domains that appear to be affected; (2) establishing the mediating anatomy of these functional domains in unaffected individuals; (3) investigating whether abnormalities exist at nodes within those circuits in individuals with the disorder; and, finally, (4) demonstrating associations between such abnormalities and the clinical features of the disorder. In the context of anxiety disorders, several key functional domains have emerged, including persistent or enduring learning about threat-related stimuli, exaggerated fear responses, impaired extinction learning or retention, difficulty suppressing attention and emotional and/or behavioral responses to disorder-relevant stimuli, overgeneralization of possible threat, perceptual hypersensitivity to threat-related stimuli, and heightened anxiety sensitivity. These functional domains and their associated neural circuitry will be discussed in detail in the following sections of this chapter.

In general, two parallel lines of research have proved particularly valuable for this process of investigating the neural circuitry involved in anxiety disorders. First, classic fear conditioning paradigms used in both animal studies and neuroimaging studies of humans without psychiatric disorders have elucidated the neuroanatomical circuits mediating anxiety, fear, and stress. Second, neuroimaging studies of individuals with anxiety disorders have begun to provide convergent evidence suggesting how these mediating neuroanatomical circuits are altered in the context of these disorders.

PRECLINICAL STUDIES OF FEAR LEARNING

Capitalizing on the large degree of cross-species conservation in the neural circuitry underlying fear learning, preclinical work in animals has laid the foundation for the design and interpretation of functional imaging studies of fear and anxiety in humans. Preclinical work in this area has commonly examined the anatomical correlates of fear learning in the context of Pavlovian fear conditioning, fear-potentiated startle (FPS), and other theoretically anxiety-related paradigms, such as the open field paradigm, most typically in rodents.

Classic fear conditioning paradigms present a highly influential and pertinent example through which to discuss the impact of preclinical work on our understanding of neurocircuitry implicated in anxiety, fear, and stress-related disorders. In these paradigms, an animal is exposed to a sequence of intrinsically neutral sensory stimuli (e.g., auditory tones). These stimuli are conditioned over the course of numerous trials, with one of these stimuli paired with an inherently aversive stimulus called the unconditioned stimulus (UCS) (e.g., a shock), and others presented without the UCS. The conditioned stimulus (CS) that is associated with the UCS is referred to as the CS+, whereas the CS that is presented without the shock is referred to as the CS-. Following conditioning via repeated exposure to these stimuli, exposure to the CS+ alone (i.e., without the UCS) provokes defensive behaviors or actions that otherwise suggest preparation for an expected aversive event. These behaviors, such as increased startle or freezing in rodents, resemble expected behaviors associated with fear and anxiety.

Some research paradigms follow this fear conditioning with extinction learning. In extinction paradigms, a sequence of CS+ stimuli is presented without an accompanying UCS, thus suggesting that the CS+ no longer predicts the UCS. After a temporal delay (e.g., the following day), the animal’s response to the CS+ is investigated as an indicator of the retention of this extinction learning. Previously extinguished fears may return due to a variety of reasons. Importantly, extinction does not erase the memory of the original pairing between the CS+ and the UCS; rather, extinction leads to the creation and bolstering of a competing safety memory. Thus, previously extinguished fears can return due to the competition between the new learning with the original fear memory. In addition, the fear response can return following the presentation of unsignaled UCS (i.e., *reinstatement*) or following the passage of time or presentation of the CS+ after extinction (i.e., *spontaneous recovery*).

Paradigms that investigate contextual fear learning offer particular relevance for the phenomenology of situation-specific anxiety. For instance, the initial fear conditioning might occur in a specific context (Cx+; e.g., a chamber with particular features) and the extinction learning in a different context (Cx−; e.g., a different chamber). In this circumstance, testing the day after extinction training reveals differential extinction retention depending on the context of the testing trials. Specifically, extinction learning is typically preferentially exhibited in the Cx− (the extinction context), whereas the conditioned fear response is typically larger in the Cx+ (the conditioning context).

Expanding upon prior work in classic extinction, recent work has examined the role of memory reconsolidation in the reduction of fear. When memories are retrieved, they return to a labile state before they restabilize. Memory reconsolidation rests on this dynamic state of memories; recalling a memory and thereby making it modifiable allows for updating this memory with new information. In the process of memory reconsolidation, extinction thus occurs within a specific temporal window of reconsolidation starting approximately 10 minutes after retrieval and continuing for at least 1 hour. In contrast to classic extinction, in which the new extinction memory competes with the original fear memory, memory reconsolidation involves updating the original fear memory such that the old fear response is no longer expressed.

Findings from preclinical studies of fear learning have important implications for the nature and treatment of anxiety disorders in humans. Individuals with anxiety disorders may show perturbations in functional domains linked with fear learning and extinctions, such as heightened learned-fear responses, limited extinction learning or retention of extinction memories, or limited recognition of safe contexts. Furthermore, this work has important implications for treatment. Exposure-based behavioral therapies for anxiety disorders explicitly build on the principles of fear extinction. Moreover, the effectiveness of behavioral therapies and interventions may rely on extinction learning, retention, and generalization across contexts.

ANATOMICAL CORRELATES OF FEAR AND ANXIETY

Preclinical studies have afforded the mapping of neural circuits that support core functions implicated in the emotional states of fear and anxiety (Fig. 14.5–1). Emotional processing can be divided into four components: evaluation, expression, experience, and modulation. *Emotional evaluation* involves assessing a stimulus’s valence (e.g., positive or negative) and its relation to prior experiences and environmental contexts in which the stimulus occurs. *Emotional expression* comprises the array of behavioral, endocrine, and autonomic manifestations of the emotional response. *Emotional experience* conveys the subjective feeling accompanying the response. Importantly, this subjective experience has primarily been addressed in humans due to their capacity to verbally describe their emotional experiences. Each of these components of emotional processing is *modulated* by complex neurobiological systems that prevent emotional responses from becoming persistent, excessive, or maladaptive.

Fear Circuitry

In general, the anatomical systems underlying fear are organized to permit both rapid responses to basic perceptual elements of potentially threatening stimuli and longer latency responses to more complex sensory stimuli and environmental contexts. The amygdala plays a critical role in fear learning, including the formation of learned-fear associations and the organization of fear responses. The amygdala responds to the inherently emotionally salient UCS and, in a learning-dependent manner, to the CS that predicts the UCS.

The amygdala is comprised of numerous subnuclei, collectively referred to as the amygdaloid complex, each of which serves a distinct function. The lateral nucleus of the amygdala (LA) acts as the primary sensory interface of the amygdala, receiving sensory information from the thalamus and cortex. Single neurons within the LA respond to auditory, visual, and somatic stimuli, with information about the CS and UCS within these domains converging in the LA, whereas neurons in the cortical and medial nuclei of the amygdala respond to olfactory stimuli via direct projections from the olfactory bulb. By bypassing cortical processing, the projections from the sensory thalamus to the LA are posited to support rapid conditioning to simple auditory or visual stimuli, thus permitting unconscious fear responses. In line with this theory, lesions in the auditory cortex do not prevent conditioning to single auditory tones. Conversely, projections from the primary sensory and sensory association cortices to the LA may be critical for aspects of conditioned responding to increasingly complex sensory stimuli. These relations are domain-specific (e.g., disruption of the projections from the auditory thalamus and cortex to LA specifically impedes fear conditioning to auditory stimuli and learned-fear responses to previously conditioned auditory stimuli).

After sensory input enters the LA, the neural representation of the stimulus is distributed in parallel to other amygdaloid nuclei, where it is modulated by numerous diverse functional systems. The most extensive amygdalar projections of the LA include reciprocal projections to the basal (BA) and accessory basal nuclei, as well as the central nucleus of the amygdala (CE). The direct projection from LA to CE may be sufficient to mediate conditioning to simple sensory stimuli, as lesions to either the LA or CE (but not other amygdalar nuclei) inhibit fear conditioning to a tone CS. Projections from LA to BA are also involved in forming long-term memory traces for fear conditioning. Functional inactivation of the LA and BA prior to Pavlovian fear conditioning interferes with learning, whereas inactivation immediately following conditioning does not affect memory consolidation. The BA nuclei have widespread intranuclear connections and also project to other amygdalar nuclei, including the LA and CE. The BA nuclei may also modulate neuronal responses of both the prefrontal cortex (PFC) and the LA and CE via their extensive reciprocal projections with the orbital and medial PFC.

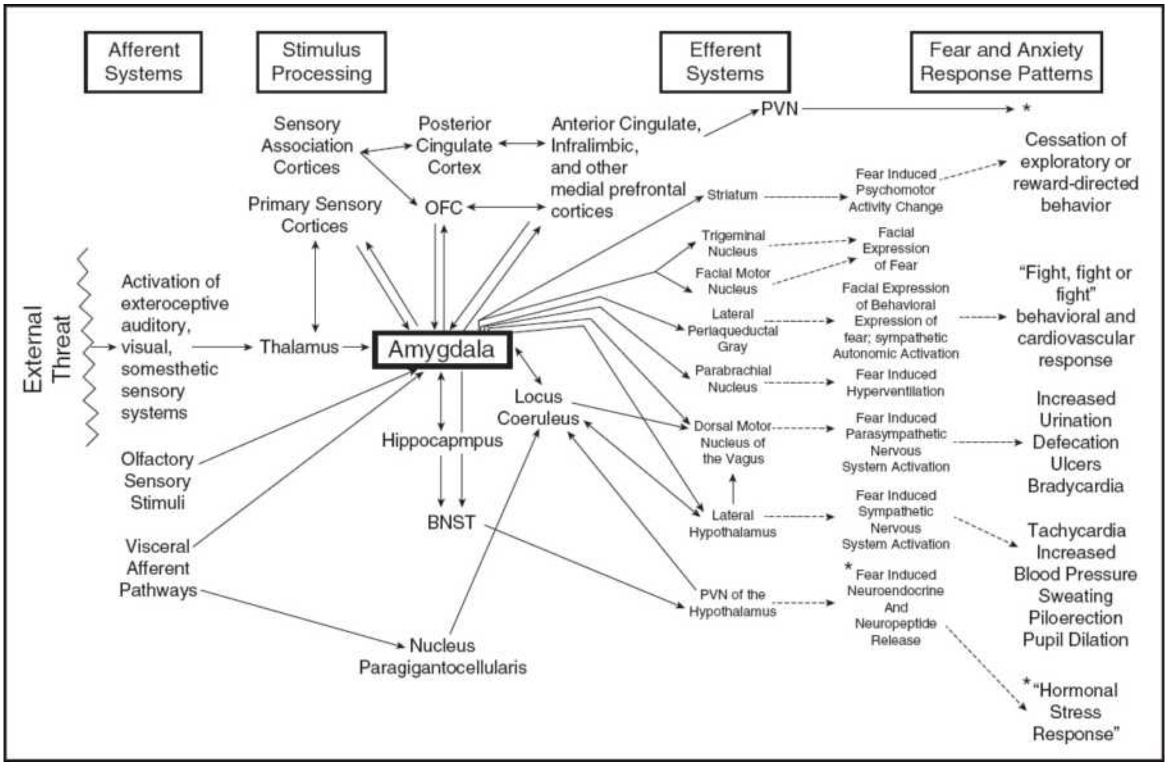
Afferent Systems
Stimulus Processing
Efferent Systems
Fear and Anxiety Response Patterns
PVN
*
Sensory Association Cortices
Posterior Cingulate Cortex
Anterior Cingulate, Infralimbic, and other medial prefrontal cortices
Primary Sensory Cortices
OFC
Striatum
Fear Induced Psychomotor Activity Change
Cessation of exploratory or reward-directed behavior
Trigeminal Nucleus
Facial Expression of Fear
Facial Motor Nucleus
External Threat
Activation of exteroceptive auditory, visual, somesthetic sensory systems
Thalamus
Amygdala
Lateral Periaqueductal Gray
Facial Expression of Behavioral Expression of fear; sympathetic Autonomic Activation
"Fight, fight or fight" behavioral and cardiovascular response
Parabrachial Nucleus
Fear Induced Hyperventilation
Locus Coeruleus
Dorsal Motor Nucleus of the Vagus
Fear Induced Parasympathetic Nervous System Activation
Increased Urination Defecation Ulcers Bradycardia
Olfactory Sensory Stimuli
Hippocapmpus
BNST
Lateral Hypothalamus
Fear Induced Sympathetic Nervous System Activation
Tachycardia Increased Blood Pressure Sweating Piloerection Pupil Dilation
Visceral Afferent Pathways
PVN of the Hypothalamus
* Fear Induced Neuroendocrine And Neuropeptide Release
Nucleus Paragigantocellularis
*"Hormonal Stress Response"

FIGURE 14.5–1. Neuroanatomical circuitry underlying fear and anxiety, organized by systems that evaluate the salience of potentially threatening stimuli (to the left) and systems that organize and mediate the expression of fear and anxiety behaviors (to the right). BNST, bed nucleus of the stria terminalis; OFC, orbitofrontal cortex; PVN, paraventricular nucleus. (Reproduced with permission from Charney DS, Drevets WC. The neurobiological basis of anxiety disorders. In: Davis K, Charney DS, Coyle J, Nemeroff CB, eds. *Neuropsychopharmacology. The Fifth Generation of Progress*. Lippincott Williams & Wilkins; 2002:901.)

Both the amygdala and cortical areas are implicated in plasticity-related fear learning. Rapid conditioning of fear responses to potentially dangerous stimuli depends on amygdala plasticity, which involves long-term potentiation. Conversely, learning involving higher cognitive (i.e., attentional or mnemonic) processing of fear experiences may rely on cortical neuronal plasticity, which is influenced by the amygdala. Cortical regions are believed to establish explicit or declarative memories about the fear-related event through interactions with the medial temporal lobe memory system. The amygdala appears to influence cortex-based memories via late plastic components of the auditory cortex neuronal responses to a CS. Single-unit recording during fear conditioning has indicated that some auditory cortex neurons develop late-conditioned responses (i.e., the conditioned response does not develop until 500 to

1,500 ms after CS onset) that are resistant to extinction. These late-conditioned auditory cortical neurons take more trials to show a conditioned response and respond more slowly than LA neurons within trials. Furthermore, amygdala lesions prevent these late-developed conditioned responses.

By contrast, other auditory cortex neurons show early (i.e., less than 50 ms after CS onset) plasticity during fear conditioning. This short-latency plasticity involves the conditioning-based enhancement of pre-existing electrophysiological responses of auditory cortex neurons to the CS and appears to rely on input from the auditory thalamus. Furthermore, this early response is not affected by amygdala lesions; however, the amygdala appears to be involved in preventing the extinction of these early responses, as they are extinguished more quickly in animals with amygdala lesions.

Neural Circuitry Underlying Fear Expression

The amygdaloid output nuclei, particularly the CE, receive convergent input from multiple amygdala regions and organize behavioral responses in line with the sum of the neuronal activity produced by multiple amygdala nuclei. Through CE projections to nuclei in the medulla, midbrain, and hypothalamus, the CE serves as the interface between the amygdala and the motor, behavioral, autonomic, and neuroendocrine system responses involved in fear. For example, the amygdala facilitates the release of stress-related corticotropin-releasing hormone (CRH) through intrinsic CRH-containing neurons within the amygdala and bisynaptic (double GABAergic [GABA, γ-aminobutyric acid]) projections to the paraventricular nucleus of the hypothalamus (PVN; the location of the majority of CRH-synthesizing neurons in the hypothalamus). Stimulation of the CE results in responses similar to those elicited by fear-conditioned stimuli, and lesions of the CE prevent the expression of several types of fear responses. In contrast, lesioning of specific structures efferent to the CE (e.g., the lateral hypothalamus or PAG) results in selective deficits in behavioral fear responses (e.g., cardiovascular or somatobehavioral, respectively). The amygdala also projects to regions involved in the organization of motor responses to threatening stimuli, such as the thalamus, nucleus accumbens, ventromedial caudate, and ventral putamen. Moreover, the amygdala may further impact motor behavior via projections to the hypothalamus and PAG.

Both the CE and BNST can directly activate structures in the hypothalamus and brainstem that mediate the expression of emotional behavior. In particular, the BNST has been implicated in sustained states of anxiety, and anxiety-like responses elicited by exposure to a threatening environment for several minutes or by intraventricular administration of CRH are specifically mediated via the BNST rather than the CE. Given these data, this system is hypothesized to be involved in mediating anxiety during exposure to less explicit or less-well-defined sensory cues and/or contexts that occur over a longer duration.

Fear Regulation Circuitry

Multiple areas of the medial and orbital PFC are involved in modulating anxiety and other emotional responses. These areas are posited to play a role in the interpretation of the higher-order significance of experiential stimuli, the modification of behavioral responses based on reinforcement contingencies, and the prediction of social outcomes of behavioral responses to emotional events. The amygdala modulates PFC neuronal activity, and the PFC modulates amygdala-mediated responses to emotionally salient stimuli through extensive reciprocal projections. Through input to the BA and the intercalated cells (a group of GABAergic neurons that regulate input from the BA to the CE), the mPFC further regulates amygdala output.

Areas of the mPFC implicated in anxiety and fear-related behavior include the infralimbic cortex (IL), prelimbic cortex, anterior cingulate cortex (ACC), and an anterior region extending from the rostral ACC (BA 24, 32) toward the frontal pole. Current theory suggests that reciprocal projections between the amygdala and mPFC play a critical role in reducing fear responses and extinguishing behavioral fear-conditioned responses that are no longer reinforced. In rats, lesions of the ACC result in enhanced freezing behavior in response to a fear-conditioned tone, thus suggesting that this region is involved in fear reduction. Additionally, neurons in the rat prelimbic cortex show reduced spontaneous firing in the presence of a fear-conditioned, aversive tone. This reduction is inversely proportional to the magnitude of the rat’s fear and negatively correlated with increases in amygdala neuronal activity. Projections from the prelimbic cortex (posited to correspond to the dorsal ACC in humans) to the amygdala have also been shown to increase the magnitude of fear responses to fear-conditioned stimuli (Fig. 14.5–2). Furthermore, lesions of the IL (believed to correspond to the ventromedial PFC [vmPFC] in humans) specifically interfere with extinction recall following long delays between extinction learning and re-exposure to the original CS.

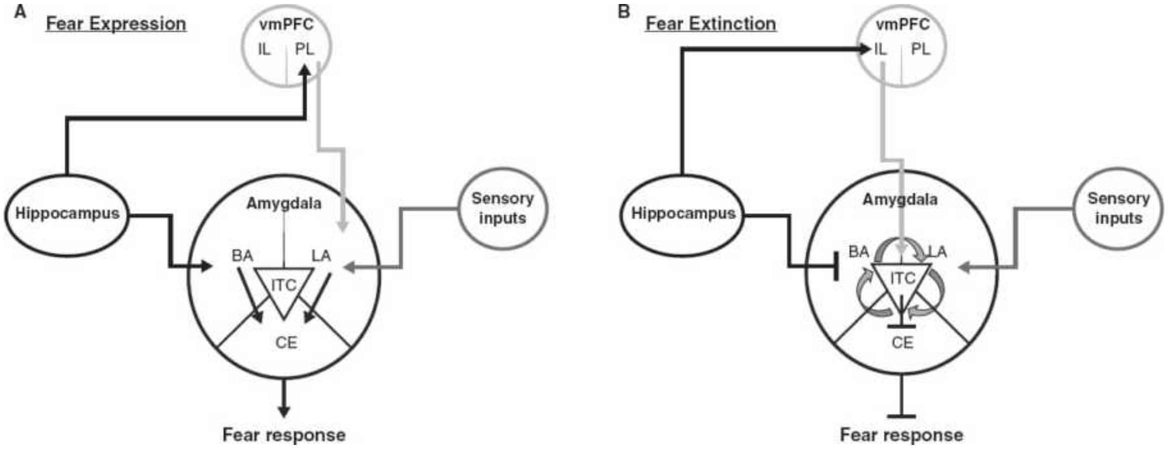

FIGURE 14.5–2. Neural circuitry of fear expression and extinction. **(A)** During acquisition and expression of conditioned fear, projections from PL and thalamic nuclei (mediating converging sensory information) excited LA neurons, while hippocampal projections (mediating contextual inputs) lead to extinction of BA neurons directly or indirectly via connections with PL. LA and BA neurons activate CE output neurons, which project to downstream brainstem and hypothalamic nuclei responsible for mediating physiologic responses, resulting in fear expression. **(B)** During extinction of conditioned fear, hippocampal projections (mediating contextual inputs) lead to divergent excitation of IL neurons and inhibition of BA neurons. IL projections directly activate GABAergic ITC cells within the amygdala. Integration of ITC, BA, and LA inputs during extinction results in a suppression of CE output neurons, resulting in a lack of physiologic response and suppression of fear expression. Arrowheads delineate pathway excitation; straight ends delineate pathway inhibition. *For simplicity, connection arrows are delineated as being unidirectional, although bidirectional projections exist. (Reproduced from Pattwell SS, Lee FS, Casey BJ: Fear learning and memory across adolescent development: Hormones and Behavior Special Issue: Puberty and Adolescence. *Horm Behav.* 2013;64(2):380–389. Copyright © 2013, with permission from Elsevier. doi:10.1016/j.yhbeh.2013.01.016.)

In humans, the areas of the ACC located anterior to the genu of the corpus callosum (i.e., the pregenual ACC [pgACC]) show elevated hemodynamic activity during elicited anxiety states, both in healthy participants and in participants with anxiety disorders. Electrical stimulation of this area is associated with vocalization in animals and a sense of fear, panic, or foreboding in humans. The subgenual ACC (sgACC) has been implicated in adaptive sadness, major depressive disorder (MDD), bipolar disorder, PTSD, and GAD. Specifically, patients with GAD show reduced gray matter volume compared to healthy controls.

Both the sgACC and pgACC have reciprocal projections with areas involved in emotional behavior, including the amygdala, hypothalamus, nucleus accumbens, posterior orbitofrontal cortex (OFC), PAG, VTA, raphe, locus coeruleus (LC), and nucleus tractus solitarius. In humans, lesions of the mPFC that include the pgACC and sgACC are associated with atypical automatic responses to emotionally provocative stimuli, inability to experience emotion related to concepts, and failure to use probabilistic information regarding possible aversive or rewarding social consequences to direct social behavior. In rats, lesions of the vmPFC (including IL, prelimbic, and ACC cortices) affect corticosterone secretion, sympathetic autonomic responses, and gastric stress pathology during restraint stress or exposure to fear-conditioned stimuli. Furthermore, stimulation of GC receptors in the rat ACC inhibits the release of stress-induced corticosterone.

In addition to these regions, more anterior mPFC areas are also implicated in emotional responses and evaluations in healthy humans. When anticipating an electrical shock, healthy humans show cerebral blood flow (CBF) increases in the rostral mPFC (near anterior BA 24, BA 32, and rostral BA 9). The magnitude of the change in CBF is inversely correlated with changes in anxiety ratings and heart rate. In addition, rats show exaggerated heart rate responses to fear-conditioned stimuli following lesions of the rostral mPFC, and stimulation of this region in rats reduces defensive behavior and cardiovascular responses induced by amygdala stimulation. In primates, the amygdala shares extensive reciprocal connections with BA 24 and 32; however, it shares only sparse connections with BA 9 cortex. Nevertheless, all three areas send extensive efferent projections to the PAG and hypothalamus, and these projections may modulate cardiovascular responses implicated in emotional behavior.

Role of the Hippocampus and Other Structures in Contextual Fear

The development of emotional responses to the environment involves interactions of the amygdala with the hippocampus, perirhinal cortex, ventrolateral PFC (vlPFC), and anterior insula. Projections from the hippocampal formation to the amygdala via the fornix are involved in spatial contextual conditioning. For instance, lesions affecting these projections specifically inhibit fear conditioning to the context (i.e., particular chamber or location within a maze) in which fear conditioning previously occurred. By supplying contextual information about the degree of threat or safety in the environment, the hippocampus modulates frontoamygdala function and influences whether the original fear memory or extinction memory is expressed. This modulation of frontoamygdala function relies on projections from the ventral hippocampus to the BA, which can further modulate subsequent fear responses via projections to the CE. Moreover, connections between the ventral hippocampus and the vmPFC (or IL in rodents) modulate extinction learning using contextual cues in the surrounding environment. For example, inactivation of the hippocampus prior to extinction training results in impaired retrieval of cued extinction memories. In these ways, hippocampally mediated control of frontoamygdala connections enhances an organism's flexibility to respond to potentially dangerous stimuli in a manner appropriate to the current environment. Building off this foundation, recent work has suggested an increasingly central role of the hippocampus in the consolidation of contextual fear memories. Recent evidence suggests that contextual fear memories may be consolidated and maintained by both the amygdala and hippocampus, in contrast to the previous model of single contextual representations from the hippocampus unidirectionally influencing the associative contextual memory in the amygdala. While this model requires future investigation, it speaks to the important role of the hippocampus in contextual fear learning.

The perirhinal cortex shares reciprocal projections with the amygdala and is theoretically implicated in conveying information about complex visual stimuli to the amygdala during the presentation of fear-conditioned visual stimuli. Lesions of the anterior perirhinal cortex, basolateral nucleus of the amygdala, or CE can completely eradicate the behavioral fear response (e.g., FPS) during exposure to some conditioned visual stimuli. Conversely, removal of the visual cortex, insular cortex, mPFC, and posterior perirhinal cortex does not affect the magnitude of FPS, and lesions of the frontal lobe only partially diminish FPS. Lesions of the lateral geniculate nucleus, which sends input regarding conditioned visual stimuli to the perirhinal cortex, can also inhibit FPS. Finally, the anterior perirhinal cortex receives afferent projections from structures implicated in modulating behavioral responses to fear-conditioned stimuli, such as the visual cortices and the ACC, infralimbic cortex, and parietal cortex. Lesions placed in the anterior perirhinal cortex following fear conditioning impede the expression of conditioned fear responses elicited by visual and auditory stimuli in contexts that differ from the original conditioning contexts. In contrast, lesions of the vlPFC and anterior insula do not affect conditioned learning or response extinction but do reduce fear reactivity to contextual stimuli.

The temporopolar cortex is theorized to play a role in modulating autonomic aspects of emotional responses and in processing emotionally provocative visual stimuli. Electrical stimulation of sites within the temporopolar cortex can affect a range of autonomic nervous system functions. In humans, exposure to emotionally provocative stimuli (whether sad, disgusting, or happy) is associated with increased neurophysiological activity in the anterior temporopolar cortex compared to exposure to neutral visual stimuli or to elicitation of emotional states via recall of autobiographical information. In this manner, areas within the temporopolar cortex may function as sensory association areas involved in evaluating the emotional salience of actual or anticipated stimuli and in modulating autonomic responses to such stimuli.

Emotion Processing and Other Related Circuitry

In both humans and nonhuman primates, other areas implicated in fear and anxiety include the posterior and lateral OFC, anterior insula, and vlPFC. Experimentally induced anxiety is associated with increased physiological activity in these regions in healthy human participants and participants with specific phobia or panic disorder (PD). The relation of physiological activity in the OFC and vlPFC with anxiety symptoms is complex. Specifically, CBF increases in the posterior OFC are observed in participants with animal phobias during exposure to phobic stimuli and in healthy participants with induced sadness; however, this change in CBF is inversely correlated with changes in anxiety.

Consistent with these findings, electrophysiological and lesion studies have demonstrated the role of the OFC in modulating fearful, defensive, and reward-directed behavioral and visceral responses in line with changing reinforcement contingencies. During the delay between stimulus and response, nearly one-half of OFC pyramidal neurons alter their firing rates in relation to the presence or absence of reinforcement. These cells are posited to be involved in extinguishing unreinforced responses to appetitive or aversive stimuli. Furthermore, the posterior and lateral OFC and amygdala modulate each other’s neural transmission through projections with each other and with overlapping areas within the striatum, hypothalamus, and PAG. For example, concurrent stimulation of OFC sites and the amygdala reduce or eliminate the behavioral and cardiovascular responses associated with stimulation of the amygdala; however, stimulation of OFC sites on their own does not produce these effects.

In humans, OFC lesions are associated with impaired performance on tasks that require the use of reward- or punishment-related information, continued use of unreinforced behavioral strategies, and difficulty switching intellectual strategies in response to changing task demands. Similarly, in monkeys, surgical lateral OFC/vlPFC lesions are associated with difficulty learning to inhibit prepotent responses to nonreinforced stimuli in response to changing reinforcement contingencies, a phenomenon known as “perseverative interference.” These lesion data suggest that the OFC activation seen during normative anxiety or obsessional states may reflect endogenous attempts to attenuate emotional expression or to interrupt unreinforced aversive thoughts or emotions. In contrast, dysfunction of the OFC may impair one’s ability to inhibit nonreinforced or maladaptive emotional, cognitive, and behavioral responses to social interactions and sensory or visceral stimuli and thus contribute to pathological anxiety and obsessional states.

Areas within the posterior cingulate cortex, including the retrosplenial cortex, show increased CBF in response to various types of aversive stimuli. The retrosplenial cortex, in particular, is believed to play a role in both cue-specific and contextual learning and memory. The posterior cingulate cortex is believed to function as a sensory association cortex that processes the affective salience of experiential stimuli. The posterior cingulate cortex then relays information to limbic circuitry through a major projection to the ACC.

Medial cerebellar activity shows increased activation in both healthy participants undergoing experimental anxiety or sadness inductions and in participants with anxiety disorders. A meta-analysis investigating the anatomical correlates of multiple types of anxiety across positron emission tomography (PET) studies identified the medial cerebellum and midbrain (near the PAG) as regions in which CBF increased across all normal and pathologic anxiety states studied. The medial cerebellum contains nuclei that functionally complement the hypothalamus in the modulation of autonomic responses. In addition, most or all regions within the cerebral cortex form closed, multisynaptic circuits through the cerebellum, although the functional significance of these anatomical loops remains unclear.

Anxiety- and fear-related changes in regional physiological activity in other structures have been less reliably replicated. Several studies have reported anxiety-related CBF changes in the ventral striatum or the lateral temporal and inferior parietal cortical areas. These areas serve as sensory association cortices, and anxiety-related physiological activity in these regions may reflect sensory processing of anxiety-provoking stimuli or feedback to these regions from the amygdala or other subcortical regions.

Role of the Amygdala in Biologically Relevant Learning and Emotional Memory

In addition to its role in fear conditioning, evidence suggests the amygdala more broadly facilitates biologically relevant learning, including responding to aversive and appetitive CS, uncertainty, associability, violations of expectations, innate fear, processing affective elements of social interactions, and nonspecific attention and arousal that facilitates learning. In the context of social dominance hierarchies, monkeys with amygdala lesions tend to become more submissive to dominant conspecifics and to show attenuated aggression and fear. These amygdala-lesioned monkeys also appear not to recognize socially significant signals in the behavior or facial expressions of other monkeys and, ostensibly as a consequence, are generally killed by other monkeys in natural setting primate colonies. In humans, the amygdala has been found to play a role in evaluating social behavior (specifically, the emotional information conveyed by human facial expression or voice intonation). Humans with amygdala lesions show impaired ability to recognize fear in facial expressions and both fear and anger in voice intonation. In humans without amygdala lesions, viewing faces showing emotional expressions of varying types is associated with increased amygdala activity in PET and fMRI studies. This response is automatic, with the amygdala responding even when face stimuli are presented for so little time that the participants are not explicitly aware of seeing them.

The amygdala is also involved in the recall of emotional or arousing memories. In humans, the recollection of specific emotional events has been associated with bursts of electroencephalographic (EEG) activity in the amygdala. Furthermore, electrical stimulation of the human amygdala can evoke experiences of fear or anxiety and the recollection of emotionally arousing life events from distant memory. Amygdala activation also enhances the strength of long-term memory for emotional stimuli or contexts represented in other cortical memory circuits. These circuits are posited to involve the medial temporal lobe memory system, which shares extensive projections with the amygdala that support interactions between storage and explicit recall of emotionally salient memories. Moreover, acquisition of fear-conditioned responses requires an intact central noradrenergic system, as memory consolidation is critically dependent on noradrenergic stimulation of β- and α-1-adrenoreceptors in the basolateral nucleus of the amygdala for at least some form of emotional learning. Notably, glucocorticoid (GC) hormone effects on noradrenergic neurons facilitate the activation of norepinephrine release during emotional learning.

Structures Mediating Endocrine and Autonomic Responses of Fear or Stress

The generation of adaptive responses to threat or stress is facilitated by the peripheral hormonal and autonomic responses to threat mediated by the hypothalamic–pituitary–adrenal (HPA) axis and the sympathetic and parasympathetic autonomic nervous systems. Cardiovascular and endocrine responses to threat and stress are modulated by projections from the ACC, anterior insula, and posterior OFC to the hypothalamus. Stimulation of the lateral nucleus of the hypothalamus via afferent projections from the CE, BNST, or ventral striatum activates the sympathetic nervous system, thereby producing increases in blood pressure and heart rate, sweating, piloerection, and pupillary dilation. During stress, the secretion of CRH from the PVN of the hypothalamus results in increased peripheral adrenocorticotropic hormone (ACTH) levels, which in turn stimulates the adrenal glands to secrete cortisol, the so-called “stress hormone.”

Regional differences in the regulation of CRH function by GC-receptor stimulation and stress appear to play important roles in the mediation of fear and anxiety. Feedback inhibition of CRH release by GC-receptor stimulation (i.e., suppressing HPA-axis activity) occurs at the level of the PVN and anterior pituitary. Specifically, systematically administered GC hormones reduce CRH release in the PVN and CRH-receptor expression in the anterior pituitary. Conversely, administered GC hormones can increase CRH mRNA expression within the CE and BNST. Ultimately, this positive feedback of GC hormones on extrahypothalamic CRH function in the amygdala and the BNST may amplify stress responses.

The innervation of the parasympathetic nervous system through afferent projections from limbic structures is thought to mediate visceral symptoms associated with anxiety, such as alterations in cardiovascular, gastrointestinal, and genitourinary function. The major efferent projections of the parasympathetic nervous system to the viscera include the vagus and splanchnic nerves. The vagal nuclei receive afferent projections from the lateral hypothalamus, PVN, LC, amygdala, and infralimbic cortex, whereas the splanchnic nerves receive afferent connections from the LC.

Regions within the orbital and medial PFC are also implicated in the modulation of peripheral responses to stress, including heart rate, blood pressure, and secretion of GC hormones. In turn, neuronal activity within these areas of the PFC is modulated by an array of neurotransmitter systems that are activated in response to threats and stressors. For instance, by modulating neuronal activity in the PFC, the noradrenergic, dopaminergic, and serotonergic systems are involved in enhancing vigilance, regulating goal-directed behavior, and facilitating decision-making involving probabilities of rewarding versus aversive outcomes.

OVERVIEW OF NEUROIMAGING TECHNOLOGY

Structural Neuroimaging Techniques

Using structural imaging techniques, experimenters have examined regional size, shape, and tissue composition of brain regions and have investigated how these features relate to clinical symptoms in individuals with anxiety and stress-related disorders. Volumetric magnetic resonance imaging (vMRI) enables investigators to obtain indices of size or features of shape from high-resolution structural MRI data using automated or semiautomated segmentation of brain regions of interest. This approach capitalizes on techniques for parcellating cortical sub-regions, automated procedures for measuring cortical thickness, and voxel-based morphometry (VBM), which is an automated method for identifying voxels where structural differences can be inferred based on signal intensity differences in structural MRI scans. Furthermore, diffusion tensor imaging

(DTI) enables experimenters to index regularity and myelination of white matter tracts based on the degree of anisotropic diffusion (fractional anisotropy; FA).

Functional Neuroimaging Techniques and Paradigms

A variety of functional neuroimaging techniques and experimental paradigms have been developed to delineate the neurophysiological correlates of normal brain functions and the pathophysiological correlates of psychiatric disorders. One such technique is PET scans, which can be conducted with tracers to measure regional cerebral blood flow (rCBF) or glucose metabolism. Another popular method is fMRI, which is used to measure blood oxygenation level–dependent (BOLD) signal changes. Furthermore, EEG and magnetoencephalography (MEG) measure changes in electrical potentials and the external magnetic field, respectively, generated by brain activity. Importantly, these profiles of brain activity are sensitive to the cognitive-behavioral state of the individual at the time of tracer distribution or image acquisition.

Experimental paradigms capitalize on this sensitivity through a variety of state manipulations, and they may rely on an array of possible principal statistical comparisons to probe specific questions. In *neutral state* paradigms, participants are scanned while performing a nonspecific continuous task or during a nominal “resting” state. These paradigms thus support testing of hypotheses regarding between-group differences that do not particularly consider momentary state variables.

In *symptom provocation* paradigms, participants are scanned during both a symptomatic state (i.e., following the induction or exacerbation of symptoms) and during a control condition. Symptoms may be provoked using either behavioral or pharmacological challenges, or experiments may be designed to capture spontaneously occurring symptomatic states without additional provocation. These paradigms support testing of within-group hypotheses regarding the mediating anatomy of the symptomatic state, as well as group-by-condition interactions probing differential responses in patient versus control groups.

In *cognitive/behavioral activation* paradigms, participants are scanned while performing specially designed cognitive/behavioral tasks that activate particular brain systems of interest, thus increasing sensitivity and specificity. As in symptom provocation paradigms, both within-group analyses and group-by-condition interactions may be used to establish the mediating anatomy of functional domains of interest and to compare the functional responsivity or integrity of target brain systems in patient versus control groups, respectively.

Each of these types of paradigms may be used in longitudinal studies. In *developmental* studies, longitudinal designs are used to examine brain measures across childhood and adolescence in individuals with or at risk for a disorder of interest. In *treatment* studies, participants are scanned before and after receiving treatment. Within-group analyses are then used to examine changes in brain activity profiles that are associated with symptom improvement. Such analyses may inform the mechanisms underlying effective treatment. Additional analyses of pre-treatment data may further identify pretreatment brain activity characteristics that predict response to treatment.

HUMAN NEUROIMAGING STUDIES OF FEAR CONDITIONING AND EXTINCTION

In line with findings from experimental animals, a growing body of neuroimaging studies in healthy humans has investigated the functional anatomy associated with conditioned fear learning, extinction, and extinction retention. In addition to neuroimaging techniques, these studies often use peripheral physiological endpoints (e.g., skin conductance) as indices of conditioned fear in the pursuit of identifying patterns of functional activation and connectivity associated with fear conditioning.

Many studies have demonstrated increased amygdala activation to the CS+ compared to the CS− during fear conditioning and have shown that amygdala activation during acquisition correlates with the conditioned fear response. Studies have also found activation of the vmPFC (particularly the pgACC) and vmPFC-amygdala connectivity, especially during extinction. For example, during testing of extinction following a 24-hour delay, vmPFC activation predicted extinction learning retention and correlated with amygdala activation. This finding is consistent with preclinical work indicating that the IL inhibits amygdala response during extinction recall. In a separate study, cortical thickness at a specific prefrontal location within medial OFC directly correlated with the magnitude of extinction recall in humans. In contrast, human neuroimaging studies have suggested that memory reconsolidation is primarily mediated by the amygdala and may not rely on prefrontal engagement.

Human neuroimaging studies and lesion studies have supported the role of the hippocampus in contextual learning. Neuroimaging techniques have demonstrated activation of the hippocampus during contextual conditioning and in context-dependent extinction retention. Furthermore, consistent with work in experimental animals, humans with hippocampal lesions show impairment in contextually mediated reinstatement of extinguished fear.

In summary, these data implicate the amygdala, vmPFC, and hippocampus in human fear conditioning and extinction. By demonstrating the neural substrates of these functions in healthy humans, this research provides important tools for examining the functional anatomy underlying anxiety- and stress-related disorders.

COGNITIVE AND EMOTIONAL ACTIVATION PROBES

Human neuroimaging paradigms have employed cognitive and emotional components to study emotion processing, anxiety, and the neural circuitry hypothesized to underly anxiety and stress-related disorders. Most commonly, these paradigms employ pictures of emotionally valenced visual stimuli, such as emotionally expressive human faces. Given humans’ vast experience with emotional facial expressions and the reinforcement provided by the information obtained about the environment through emotionally expressive faces, facial expressions of emotions can be conceptualized as conditioned stimuli.

Passive viewing of human face stimuli has been found to reliably evoke amygdala activation. The amygdala responds to fearful, happy, neutral, and angry faces; however, it appears to have a greater response to fearful faces. Some studies have suggested that the amygdala may “preferentially” respond to fearful faces in part because fearful faces predict uncertainty (i.e., increased probability of threat of unknown nature or location, as opposed to the direct threat entailed by angry faces). In line with its reactivity to uncertainty, the amygdala also shows robust activation to surprised faces, which are inherently ambiguous (i.e., they could predict either a positive or negative event). Of note, both fearful and surprised faces predict a salient but unknown event, and they share common features (e.g., widened eyes).

In addition to the amygdala, numerous other brain areas (e.g., visuotemporal cortices, PFC) are implicated in the processing of face stimuli. Thus, studying the neural circuitry underlying the processing of emotional faces allows for the assessment of the extended functional anatomical networks that are recruited in the evaluation of emotional stimuli. In particular, presenting face stimuli using backward-masking techniques (i.e., the subject is not explicitly aware of seeing the masked face) provides an opportunity to dissociate automatic amygdala responses to stimuli from top-down influences related to cognitive processes. Moreover, repeated presentation of face stimuli can be used to examine profiles of amygdala response habituation. These principles have been found to extend to viewing emotionally valenced visual stimuli other than emotional faces (such as those included in the International Affective Picture System), as studies have reported robust and reliable amygdala activation to emotionally valenced versus “neutral” visual stimuli.

Given that anxiety disorders are characterized by emotional interference and negative attentional biases (i.e., difficulty suppressing attention to and redirecting attention away from threat-related stimuli), functional neuroimaging paradigms that measure these processes play a large role in the literature on anxiety disorders. The ACC is critically involved in signaling the presence of conflict. Specifically, the dorsal ACC tends to signal cognitive interference (e.g., as probed by the Counting Stroop Task), whereas the ventral ACC tends to signal emotional interference (e.g., as probed by the Emotional Counting Stroop Task). The lateral PFC, on the other hand, is more directly involved in augmenting cognitive control and mediating attentional biases.

Another domain studied using functional neuroimaging is the enhancement of long-term memory for an emotional event associated with brain activation during the event. In healthy humans, the magnitude of amygdala activation when reading stories of differing emotional valence correlated with both the negative emotional intensity and the subsequent ability to recall the content of the story. Furthermore, activation in the amygdala and hippocampus during memory encoding correlated with enhanced episodic memory for both pleasant and aversive visual stimuli. These data suggest that the role of the amygdala in modulating emotional memory may more generally depend on the degree of arousal or the behavioral salience associated with target information and not exclusively on the negative emotional intensity, as was previously believed.

Finally, given that interoceptive sensitivity can be heightened in anxiety disorders (e.g., in PD), interoceptive functioning has been explored in functional neuroimaging studies. Using tasks such as perceiving one's own heartbeat, viewing faces expressing disgust, viewing disgusting pictures, and generating internal images of disgust, neuroimaging work has implicated the insular cortex in interoceptive function.

NEUROIMAGING STUDIES OF ANXIETY DISORDERS

Extrapolating from the previously described data in experimental animal models of fear and stress, theorists have posed hypotheses regarding dysfunction in homologous structures in humans with anxiety disorders. Specifically, individuals with anxiety disorders are hypothesized to show atypical functional connectivity of the amygdala with the mPFC and/or the hippocampus that contributes to intrinsically exaggerated amygdala hyperresponsivity and/or deficient top-down modulation of the amygdala response. Fear conditioning and extinction studies have found both behavioral and neuroimaging data supporting these hypotheses for some anxiety disorders. The following sections discuss the current theories and associated evidence base within specific anxiety disorders.

Panic Disorder

Models of PD must explain both the behavioral sequelae of panic attacks (e.g., agoraphobic avoidance) and the occurrence of spontaneous panic episodes. The behavioral sequelae can putatively be explained as a product of fear learning and operant conditioning processes. Several hypotheses have emerged regarding the occurrence of spontaneous panic episodes. One hypothesis suggests that typical physiological anxiety responses mediated by anxiety/fear circuitry may be spontaneously recruited as an aberrant event due to homeostatic deficits. This hypothesis is exemplified by the "false suffocation alarm" model and theories regarding monoaminergic dysregulation. Another hypothesis holds that failures in the systems responsible for modulating normal anxiety responses lead to the escalation of what would be minor anxiety responses into excessive, maladaptive responses. Such a process might involve deficits in the hippocampus or other regions within the temporal cortex. Finally, a third hypothesis proposes that panic attacks that are described as spontaneous may, in fact, reflect fear responses to stimuli that are implicitly processed without conscious awareness. According to this model, panic attacks might involve amygdala hyperresponsivity to subtle environmental cues, such that implicit cues can trigger full-scale fear responses. Moreover, dysfunction in hippocampal or parahippocampal regions may underlie the absence of conscious awareness, as dysfunction in these regions may allow for abnormal escalation in response to contextual stimuli that are not explicitly recognized as inherently threatening.

In general, a broad range of disparate components of neurocircuitry has been implicated in PD. Notable elements include dysregulated ascending noradrenergic and/or serotonergic systems, aberrant responsivity to panicogenic substances (e.g., CO_2) at the level of the brainstem, global cerebral abnormalities in lactate metabolism, and abnormal interactions among hippocampus, parahippocampal cortex, and amygdala.

Functional Imaging Findings in Panic Disorder. In PD, the baseline state is characterized by mild to moderate levels of chronic anxiety (i.e., “anticipatory anxiety”). The neural correlates of this clinical observation have been examined using resting-state neuroimaging, which measures brain activity in the absence of a task (i.e., while the participant is at rest). Patients with PD show altered resting-state rCBF and functional connectivity in both PET and fMRI studies. Specifically, patients show altered activation in the parahippocampal gyrus, hippocampus, and superior temporal gyrus, as well as increased ACC-precuneus functional connectivity. These findings point to distinct patterns of alterations in brain function in PD, even in the absence of symptom provocation. However, the extent to which these alterations create versus reflect patients’ chronic anxious states remains unclear.

Symptom provocation studies of PD have relied on both pharmacologic challenges and cognitive/emotional challenges to induce symptoms. SPECT and PET studies using pharmacologic challenges have had mixed results. Following pharmacologic challenge, patients who experienced induced panic attacks have been found to have global reductions in CBF, reduced rCBF in the frontal cortex, and increased rCBF in both the putamen and ACC, compared to both healthy controls and patients who did not experience induced panic attacks. Findings regarding the insula activity during panic attacks have been inconsistent, with some studies reporting reduced rCBF and others reporting increased rCBF. In addition to pharmacologic challenges, some studies have examined patterns of rCBF in the midst of spontaneous panic attacks that have occurred during imaging. These results have been based on a small number of participants and have been mixed, with recorded responses including decreased rCBF in the right OFC, subgenual PFC, ACC, and anterior temporal cortex in one case; increased amygdala, insula, and PFC activation in another; and increased PFC activation in a third case.

Cognitive/emotional challenges (such as script-driven panic-related imagery, agoraphobic-specific pictures, and panic-related visual stimuli) have also been used to probe the neural correlates of anxiety in PD. Patients have been found to show exaggerated activation in the inferior frontal cortex, cingulate cortex, OFC, hippocampus, insula, and striatum in response to such panic-related stimuli. Other studies have used emotional stimuli that are not disorder-specific, such as emotional faces, to probe neural changes associated with PD. In general, patients have been found to show increased activation in the amygdala and insula and reduced activation in the vmPFC and ACC in response to emotional faces; however, findings have been mixed, particularly for the amygdala and ACC. Consistent with the general pattern of intolerance of uncertainty that is characteristic of anxiety disorders, patterns of increased insula activation have been found in response to unpredictable, but not predictable, aversive stimuli.

Treatment studies have begun to identify potential mechanisms and predictors of treatment response in PD. Compared to nonmedicated patients, patients undergoing pharmacologic treatment have been found to demonstrate distinct patterns of brain function and structure following symptom remission, such as increased FA in the uncinate fasciculus. Furthermore, symptom remission following cognitive behavioral therapy (CBT) has been associated with reduced IFG activation and greater functional connectivity of the IFG with the amygdala, hippocampus, mPFC, and lateral PFC. In addition to these efforts to identify mechanisms underlying effective treatment, other studies have aimed to identify predictors of treatment response. Stronger negative functional connectivity between the ACC and the amygdala at baseline has been found to predict future response to both pharmacotherapy and CBT. Taken together, these studies suggest that alterations in frontolimbic connections may underlie effective treatments for PD.

Structural Imaging Findings in Panic Disorder. Compared to healthy controls, patients with PD have been found to exhibit smaller OFC, putamen, and overall temporal lobe volumes; increased insula and brainstem volumes; and lower gray matter density in parahippocampal cortex. Furthermore, some evidence suggests that smaller amygdala volume may be negatively associated with anxiety severity in PD; however, other studies have not detected abnormal amygdala volumes in this population.

Summary. In summary, the neuroimaging data implicate abnormalities in several brain regions in PD. Data suggest that PD is associated with abnormal hippocampal/parahippocampal activity at rest, as well as reduced activity in widespread cortical regions, such as the PFC. Morphometric MRI studies have also detected abnormalities in the gray matter volume of the parahippocampal/temporal cortex. Finally, regional abnormalities within the temporal lobes provide support for theories of dysfunctional top-down influence over the amygdala via the temporal cortex and/or the hippocampus. Such abnormalities may play a role in the liability for developing PD following an initial panic attack.

Specific Phobias

Prominent models suggest that specific phobias represent the product of dysregulation in systems related to assessing potentially threatening, anxiety-provoking, or disgusting stimuli or situations. To date, neuroimaging data has begun to provide an empirical model for such neurocircuitry-based models of specific phobias. These theories initially focused on the amygdala due to its crucial role in mediating both innate and conditioned fear. However, human neuroimaging data has more directly implicated anterior paralimbic regions and sensory cortical regions known to interact with the amygdala. In light of these data, these regions have taken an increasingly central role in theories of neurocircuitry underlying specific phobias.

Functional Imaging Findings in Specific Phobias. Meta-analyses have detected consistent findings of increased insula and amygdala activation to phobic stimuli in patients with specific phobias. Furthermore, meta-analyses have reported preliminary evidence for increased activation in the ACC, OFC, thalamus, and cerebellum. The majority of such studies of neural activation in specific phobias have relied on fMRI and PET symptom-provocation studies using phobia-related stimuli. For example, one study exposed patients with spider phobia to pictures of spiders or neutral pictures under predictable (phasic) or unpredictable (sustained) conditions. Phasic fear of phobic stimuli was associated with elevated amygdala activation, whereas sustained fear of phobic stimuli was associated with elevated BNST and ACC activation. Patients with spider phobia also showed stronger BNST-amygdala functional connectivity compared to healthy controls. These findings demonstrate the dissociation between the role of the amygdala in acute fear and the role of the BNST in sustained anxiety.

Studies using nonphobic stimuli (e.g., emotional faces) have not detected heightened amygdala activity. Thus, amygdala hyperreactivity in participants with specific phobias appears to be unique to phobic stimuli. Notably, disruptions in cortical control regions or their connectivity with limbic and paralimbic regions may underlie this hypersensitivity to fear-related cues. For instance, a study of cognitive reappraisal in spider phobia instructed participants to use specific, effortful regulation strategies (i.e., up-regulate or down-regulate) or to automatically regulate (i.e., simply look) while presented with pictures of spiders and non–spider-related aversive pictures. Compared to controls, patients with spider phobia showed reduced vmPFC activation during automatic regulation of spider versus non–spider-related aversive pictures, as well as reduced dACC/dmPFC activation during effortful regulation (down-regulate versus simply look) to spider versus non–spider-related aversive pictures.

Several studies have attempted to identify neural substrates that may be common to all specific phobias or distinct to subtypes (e.g., spider vs. blood-injection-injury). For example, patients with spider phobia and those with blood-injection-injury phobia both show faster time-to-peak in amygdala reactivity compared to healthy controls. However, spider phobia may be more related to alterations in the dACC and insula, whereas blood-injection-injury may be more related to alterations in the PFC, thalamus, and occipito–temporo–parietal regions. Furthermore, evidence suggests that increased activation in sensory association cortices corresponds to the sensory modality associated with the symptomatic state in subtypes of specific phobias. In this context, it is noteworthy to recall that the amygdala sends anatomical projections to sensory cortices, and stimulation of these projections “tunes” (i.e., facilitates) responsivity in those areas during amygdala activity (e.g., in response to threatening stimuli).

In addition to elucidating the neurocircuitry underlying specific phobia, neuroimaging has also been used to examine potential mechanisms involved in effective treatment. A meta-analysis reported that exposure therapy for specific phobias was associated with reduced activation in the frontal cortex, limbic regions, and cerebellum. Moreover, findings from a longitudinal study of patients with spider phobia who engaged in a single exposure therapy session (3 hour or less) indicate that successful exposure therapy leads to lasting reorganization of brain responses to previously feared stimuli, such that limbic activation is normalized without the need for active prefrontal down-regulation. Prior to treatment, patients showed increased amygdala, insula, and ACC activation to phobic versus neutral stimuli. Shortly following exposure therapy, effective treatment was associated with reduced amygdala activation and increased prefrontal engagement. Six months later, participants continued to show reduced amygdala activation but in the absence of increased prefrontal engagement. Moreover, lower activation in the visual cortex shortly following the exposure therapy session predicted better treatment outcomes 6 months later.

Structural Imaging Findings in Specific Phobias. Structural MRI studies have identified several volumetric differences associated with specific phobias in general and with particular subtypes. Participants with animal phobias have been found to show increased cortical thickness in the pgACC, insular cortex, and posterior cingulate cortex. Participants with dental phobia (a subtype of blood-injury-injection phobia) have been found to show increased gray matter volume in the sgACC, insula, OFC, and PFC. Finally, participants with spider phobia have been found to show cortical thinning in the ACC, as well as reduced amygdala volume that correlates negatively with symptom severity. The specificity exhibited by these subtype-specific findings may relate to the complex neurocircuitry underlying the representation of these multifaceted categories of stimuli.

Summary. Current evidence implicates activation of anterior paralimbic regions (i.e., anterior insula) and sensory association cortices corresponding to the sensory modality associated with the symptomatic state in specific phobia. These results are consistent with a model of hypersensitivity in the system responsible for evaluation of and response to specific threat-related cues—a system that crucially involves the amygdala. The amygdala hyperresponsivity detected in specific phobia appears to selectively occur in response to phobic-related stimuli and may relate to disrupted interactions with cortical control regions that are typically involved in emotion regulation.

Social Anxiety Disorder

Similar to specific phobia, models of social anxiety disorder (SAD) center around increased sensitivity to the potential threat of a specific type of stimuli—in this case, socially relevant stimuli. Some possibilities for this increased sensitivity involve enhanced sensitivity to social threat or scrutiny. Other hypotheses suggest that SAD may involve heightened responses to novel social stimuli compounded by a decreased reward value of positive social stimuli. The majority of research paradigms in this area have examined primary responses to social stimuli (e.g., faces) of varying valence, arousal, and novelty. In addition, in line with hypothesized models of SAD, research paradigms have probed self- versus other-focused attention.

Functional Imaging Findings in Social Anxiety Disorder. A variety of symptom provocation paradigms, including public speaking challenges and exposure to contemptuous faces, have been used to examine neural mechanisms of SAD. Such symptom provocation studies using both PET and fMRI consistently demonstrate heightened amygdala activity in patients with SAD compared to healthy controls. This elevated amygdala activity has been found to persist across conditions, such as presentation of facial expressions of both low and high intensity. Such findings suggest that the amygdala may be activated at a relatively low threshold of social stimulus in SAD. Furthermore, several studies have reported correlations between patients' amygdala activation and the severity of their social anxiety. In addition to these findings of altered amygdala function, heightened insula activation to emotional stimuli, and lower striatal activation during implicit learning have also been implicated in SAD. Preliminary evidence further suggests altered function of the OFC, ACC, mPFC, and parahippocampal gyrus. Functional connectivity analyses suggest that weaker frontolimbic connections may underlie this heightened emotional reactivity (e.g., in the amygdala and insula).

Several studies have demonstrated that patients with SAD selectively show altered neural responses to social stimuli (and not to nonsocial stimuli). For instance, hyperresponsivity of the amygdala in SAD was detected in response to exposure to human face stimuli and not to aversive odors. Moreover, patients have been found to exhibit exaggerated amygdala and mPFC activation to self-referential critical comments but not to critical comments of others or to neutral or positive comments about the self or others. This suggests that processing of self-focused criticism or judgment may be particularly altered in SAD.

Treatment studies have identified possible neural mechanisms underlying effective treatments for SAD. As compared to patients in a wait-list control condition, patients who received CBT showed a greater reduction in negative emotion and corresponding greater dlPFC and dmPFC activation and stronger negative dmPFC-amygdala connectivity during cognitive reappraisal. Patients who received CBT also exhibited earlier onset of dmPFC activation. Several studies have also examined neural predictors of treatment response in SAD. CBT response has been shown to be predicted by pretreatment activation of the PFC and amygdala, as well as amygdala-ACC functional connectivity. Specifically, patients with greater regulatory capacity over emotional reactivity to faces at baseline have been found to be more likely to benefit from treatment. Thus, highlighting the potential contributions of neuroimaging for personalized medicine, studies have found that neural measures enhance the prediction of treatment response in SAD over and above clinical measures alone.

Structural Imaging Findings in Social Anxiety Disorder. Consistent with fMRI findings of weaker amygdala-OFC functional connectivity, structural neuroimaging studies have detected lower FA in the uncinate fasciculus in SAD. Furthermore, patients with SAD have been found to show greater cortical thickness in regions including the ACC, insula, dlPFC, and parietal cortex. Other studies have found lower gray matter volume in areas such as the lateral OFC, parahippocampus, and cerebellum.

Summary. Functional neuroimaging data implicates exaggerated responsivity to social stimuli in medial temporal lobe structures, particularly the amygdala, in SAD. Furthermore, patients with SAD may show weaker frontolimbic connectivity, and changes in frontolimbic connectivity may underlie effective treatment. These data are consistent with the theoretical model of SAD stemming from a face evaluation system that is hypersensitive to social threat or salience, perhaps in part due to weakened cortical control.

Generalized Anxiety Disorder

Models of neural mechanisms underlying generalized anxiety disorder (GAD) have emerged more recently and received relatively less attention than those of other anxiety disorders. Similar to other anxiety disorders, models of GAD include a focus on fear learning. However, satisfactory models of GAD must also account for the overgeneralization of fear that characterizes this disorder. Thus, neuroimaging studies have focused on core features of emotion dysregulation and excessive worry, as well as the overgeneralization of conditioned fear. Of note, efforts to examine neurocircuitry underlying GAD have been complicated by high rates of comorbidity with MDD and other anxiety disorders. To address this challenge, the studies presented in this section have attempted to recruit samples of “pure” GAD or to control for comorbidities by comparing patients with only GAD to those with comorbidities.

Functional Imaging Findings in Generalized Anxiety Disorder. Inconsistent findings on amygdala function in GAD suggest a more complex picture than the classic pattern of amygdala hyperreactivity seen across most other anxiety and stress-related disorders. Whereas some fMRI studies have found no differences or even reductions in amygdala reactivity to fearful faces in GAD, evidence suggests that patients with GAD may exhibit nonspecific amygdala hyperreactivity. For instance, patients with GAD have been found to show heightened amygdala activation to both negative and neutral stimuli, as well as to both incongruent and congruent trials during emotional interference.

Patients with GAD may also display altered BNST function, given the role of BNST in sustained anxiety. Compared to controls, patients with GAD show lower amygdala activation but higher BNST activation and higher sustained arousal to a gambling task designed to induce high uncertainty and anxiety about monetary loss. Furthermore, a recent study provided evidence for this temporal dissociation of BNST and amygdala function during the anticipation of threat, reporting that patients with GAD showed elevated phasic amygdala activity to threat anticipation, followed at a delay by heightened, sustained BNST activity (Fig. 14.5–3). Together, these findings may reflect earlier disengagement of the amygdala response to acute stress or anticipatory threat coupled with increased BNST activation that maintains a more sustained state of anxiety.

Consistent with clinical observations of sustained worry and anxiety, patients with GAD appear to show impaired prefrontal function. During a worry induction procedure, both patients with GAD and healthy controls activated the dACC and dmPFC, part of the “aversive amplification circuit”; however, while this activation normalized in healthy controls in resting-state scans following the worry induction, these regions showed persistent activation in participants with GAD. Moreover, activation in the dACC/dmPFC correlated with patients’ self-reported worry symptoms. Another line of work has investigated prefrontal alterations that may relate to the overgeneralization of conditioned fear in GAD. Building off classical fear conditioning paradigms, fear generalization paradigms examine participants’ responses to nonthreatening target stimuli that are increasingly dissimilar from the trained CS+ (e.g., the same shape as the CS+ but of increasingly smaller size). Patients with GAD show more overgeneralization (i.e., greater fear of the generalized stimuli), as well as less recruitment of the vmPFC. This reduced vmPFC recruitment is associated with poorer discrimination between a CS+ and a generalized stimulus. These findings point to potential neural substrates of fear overgeneralization and the failure to discriminate between threat and safety in GAD.

Connectivity analyses have provided critical information regarding the neural bases of emotion dysregulation in GAD. Several independent studies have provided evidence of altered corticolimbic connectivity in GAD. For example, an implicit emotion regulation task tested adaptation to emotional conflict through regulatory processes by asking participants to identify a target facial expression (fearful or happy). These facial expressions were overlayed with emotional words (“fear” or “happy”) that either matched or conflicted with the expressions. Healthy participants showed strong negative pgACC-amygdala coupling during this task, consistent with the downregulation of emotional reactivity; however, patients did not activate the pgACC or show this negative pgACC-amygdala connectivity. Moreover, patients with GAD have also been found to exhibit disrupted functional connectivity of amygdala subregions at rest, thus providing additional support for altered amygdala connectivity.

Together, these findings suggest that GAD may be characterized by impaired interactions between cortical regions that are important for exerting control (e.g., the ACC) and subcortical limbic structures that may underlie deficits in emotion regulation (e.g., the amygdala).

Finally, several studies have begun to examine possible predictors or mechanisms of effective treatment for GAD. Better response to venlafaxine has been found to be associated with higher anticipatory ACC activation and with higher rostral ACC and lower amygdala activation to fearful faces pretreatment. Furthermore, treatment with citalopram resulted in lower activation in prefrontal and limbic regions to worry-related stimuli post-treatment. Finally, patients who received CBT showed less amygdala and sgACC activation to fearful and angry faces, as well as increased activation to happy faces.

Structural Imaging Findings in Generalized Anxiety Disorder. Consistent with findings of altered corticolimbic functional connectivity, reduced FA has been observed in the bilateral uncinate fasciculus in patients with GAD. Moreover, lower FA in the uncinate fasciculus was associated with weaker negative pregenual cingulate-amygdala functional connectivity during anticipation of aversive stimuli. Both structural and functional connectivity changes have been associated with poorer discrimination between threat and safety in GAD. Taken together, these combined structural and functional results suggest potential neural bases for the emotion dysregulation and fear generalization that are key features of GAD.

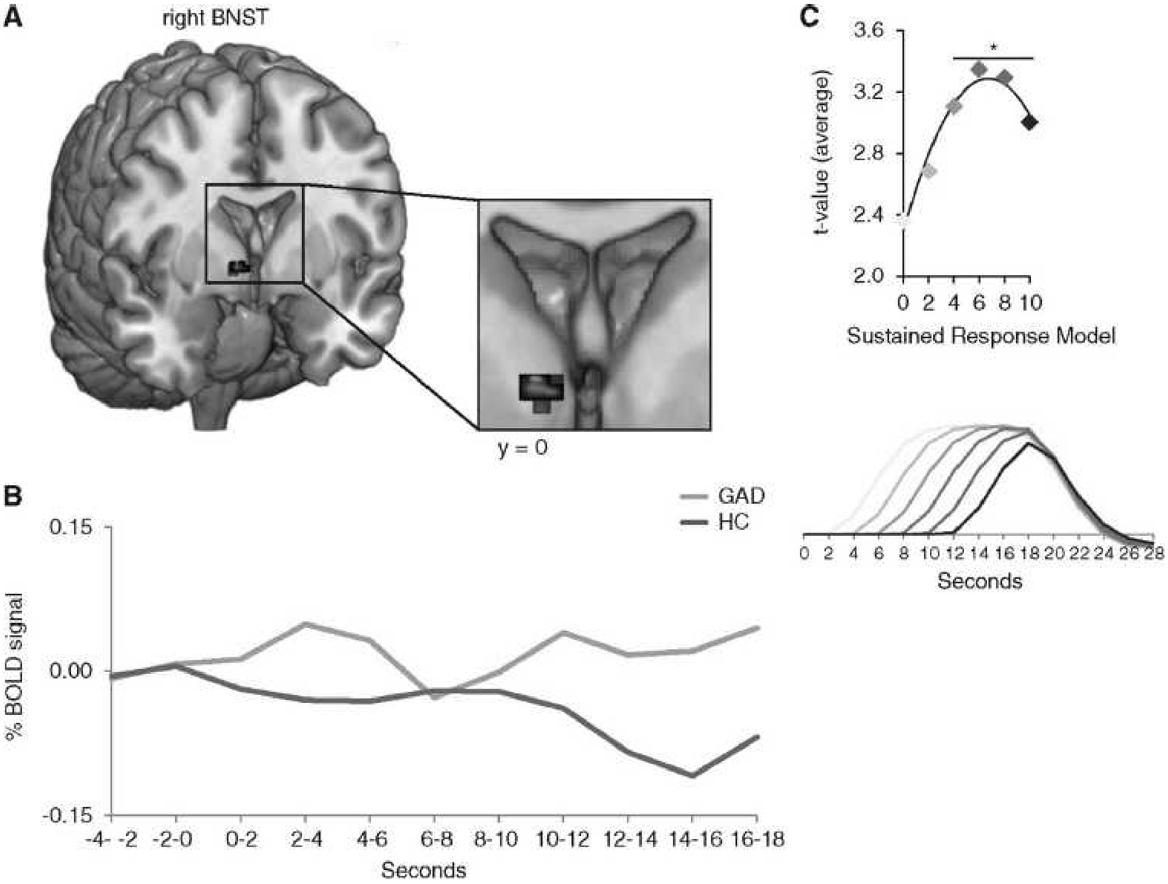

FIGURE 14.5–3. Elevated bed nucleus of the stria terminalis activity to threat minus neutral anticipation in patients relative to healthy controls. **(A)** Generalized anxiety disorder (GAD) patients relative to healthy controls (HC) responded with elevated sustained, but delayed bed nucleus of the stria terminalis (BNST) activity to threat minus neutral anticipation for all Sustained Response General Linear Models (GLM) with a delayed onset of at least 4 s in a highly overlapping cluster. **(B)** Time courses of the blood oxygen level-dependent (BOLD) signal change, extracted from the overlapping BNST cluster (of the 4, 6, 8, and 10 s Sustained Response GLMs) and averaged across time points and participants per group, are displayed for the anticipatory epoch (threat minus neutral anticipation) starting 4 s before cue onset. **(C)** Quadratic regression analyses on the average *t*-values for the contrast threat minus neutral anticipation in GAD patients as compared to HC extracted from the overlapping cluster of all significant Sustained Response GLMs, revealed a significant quadratic relationship. The vertex of the function was at 6.69 s, suggesting a delayed onset of about 6 to 7 s of the differential sustained BNST effect. The lower graph displays all Sustained Response Models. $^{*}P < .05$. (See eBook for color images.) (Reproduced with permission from Buff C, Brinkmann L, Bruchmann M, et al. Activity alterations in the bed nucleus of the stria terminalis and amygdala during threat anticipation in generalized anxiety disorder. *Soc Cogn Affect Neurosci.* 2017;12(11):1766–1774, by permission of Oxford University Press. doi:10.1093/scan/nsx103.)

Summary. Unlike the clear pattern of amygdala hyperreactivity seen in other anxiety disorders, the BNST appears to play a critical role in the sustained state of anxiety characteristic of GAD. Furthermore, disruptions in functional (ACC-amygdala) and structural (uncinate fasciculus) connectivity appear to underlie key clinical features of GAD, including emotion dysregulation, which appears to relate to weaker frontoamygdala connectivity, and fear generalization, which may be associated with alterations in vmPFC function and connectivity. Finally, initial evidence suggests that both pharmacologic interventions and CBT may reduce symptoms by altering frontoamygdala circuitry.

Commonalities and Specificity Across Anxiety Disorders

The current neuroimaging literature suggests several commonalities among disorders. The amygdala has received the most attention and has been implicated across these disorders. Patients with PD exhibit amygdala hyperreactivity to both disorder-specific and general threat-related stimuli. Nonspecific amygdala hyperreactivity may also be seen in GAD, although more consistent results implicate the BNST. Conversely, amygdala hyperreactivity in specific phobias and SAD appears to be selective for disorder-specific stimuli. Notably, heightened activity in the anterior insula has also been observed across all of these disorders.

Data implicate the abnormalities in the hippocampus in PD. Temporal lobe volumetric abnormalities are reported in extrahippocampal regions, and most resting-state data suggest a rightward shift in laterality of function. Less compelling evidence implicates the hippocampus in specific phobias, SAD, or GAD.

With respect to the PFC, PD is associated with deficient recruitment of the pgACC. In specific phobias, pgACC cortical thickness appears to be increased, although evidence of PFC dysfunction during emotional processing in specific phobias is lacking. Findings are mixed for SAD, with some studies reporting diminished neurophysiological responses in the pgACC but others failing to detect this finding. The vmPFC has also been implicated in overgeneralization of fear in GAD.

These profiles of similarities and differences among anxiety and stress-related disorders are consistent with the nosologic categorization of these disorders as distinct from one another but with sufficient similarities to merit inclusion as anxiety disorders. There has been considerably less neuroimaging work on other anxiety disorders (e.g., separation anxiety, selective mutism, adjustment disorder); thus, the extent to which these disorders have shared versus unique neural substrates will require further investigation.

The Developmental Perspective and Markers of Risk, Resilience, and Treatment

Anxiety disorders are the most common group of psychiatric disorders in childhood and adolescence, with estimated prevalence rates in adolescence as high as 32%. Moreover, anxiety disorders across the lifespan peak during adolescence. Understanding neurodevelopmental changes and the role these changes play in both illness onset and treatment efficacy is crucial to reducing the burden of anxiety and stress-related disorders. Developmental neuroimaging studies can thus have a major impact on understanding risk, resilience, and treatment for pediatric anxiety disorders.

Converging evidence from humans and experimental animals highlights dynamic changes in the neural circuitry supporting fear learning and extinction across development. The amygdala undergoes rapid structural and functional development early in life, followed by continuous change across development. Likewise, the hippocampus appears to develop earlier in life; however, the PFC shows protracted development and does not reach a mature functional architecture until early adulthood. Although the basic structure of the amygdala is present at birth, the fine-tuning of its connections with PFC and hippocampus continues throughout childhood and adolescence.

This earlier maturation of subcortical regions supporting fear reactivity, in conjunction with the protracted maturation of PFC, may result in altered fear regulation during adolescence. These developmental changes in neural function and connectivity are consistent with the changes observed in the processes of fear learning that are mediated by these structures. Evidence suggests that simple discrimination between a CS+ and a CS- improves linearly with age; however, the capacity to express and extinguish fear appears to undergo nonlinear changes across normative development. Translational evidence points to a period of diminished extinction of cued fear during adolescence, relative to childhood and adulthood, which may relate to protracted maturational changes in the PFC and its connections with the amygdala.

In addition to studies examining patterns of fear conditioning across normative development, a growing line of work has investigated fear conditioning in youth with anxiety disorders. A recent meta-analysis found that, although youth with and without anxiety disorders exhibited similar patterns of fear acquisition and extinction, fear responses to individual stimuli were of greater magnitude in youth with anxiety disorders. Several other studies have examined extinction recall across development in the context of anxiety disorders. In a recent study, 200 participants with and without anxiety disorders across a broad range of ages (8- to 50-years-old) viewed morphed stimuli of varying similarity to previously extinguished threat cues while in the scanner. Age was found to moderate anxiety-related differences in amygdala-vmPFC connectivity while processing threat-related cues. Specifically, adults with anxiety disorders showed lower amygdala-vmPFC connectivity to morphed stimuli that were increasingly more dissimilar to the originally conditioned CS+ (i.e., contained increasingly more safety-related information); however, this anxiety-related effect was reversed in youths (Fig. 14.5–4). Although based on the relatively small number of initial studies conducted to date, these findings highlight the importance of examining fear conditioning in youth versus adults with anxiety disorders.

Developmental changes in behavior and neurocircuitry have important implications for the effect of stress and anxiety on the brain. During developmental periods of heightened neuroplasticity (i.e., *sensitive periods*), environmental influences may have a particularly potent effect on neurodevelopment and may lead to a series of cascades that ultimately influence behavior (in a positive or negative manner). In other words, the brain may be more capable of remodeling itself during sensitive periods compared to windows of reduced plasticity; however, this remodeling is not inherently positive or negative. Sensitive periods may render the developing brain more vulnerable to the effects of later stress, but they could also serve as windows of opportunity, during which there is increased potential for positive adaptation or effective intervention. Dynamic neural and behavioral changes likely influence the effect of stress on the brain and the manner in which youth cope with stressors at unique developmental time periods. Ultimately, the risk for anxiety disorders may be increased by stressors that exacerbate transient imbalances in brain circuitry. Neuroimaging studies across childhood and adolescence can inform the timing and nature of sensitive periods with direct relevance for the risk and treatment of anxiety disorders.

Particularly given the potential early age of onset and chronic course of anxiety disorders (when not effectively treated), developmental neuroimaging studies offer great promise for informing risk factors for anxiety disorders across the lifespan. Even in adulthood, individual differences often stem from early-life environments/influences (e.g., parental modeling, stressful events), genetic factors, or temperament. Elucidating risk conferred by such early individual differences is critical for informing prevention and early intervention. However, distinguishing markers of risk from markers of disease remains challenging. Both the promise and challenges of this work are highlighted by the example of behavioral inhibition (BI), a behavioral phenotype characterized by reactivity to novel stimuli and withdrawal/avoidance. This temperament is evident in early childhood but confers increased risk for pediatric anxiety disorders, which in turn confers increased risk for adult anxiety disorders. BI has been associated with exaggerated amygdala responses to novel versus neutral faces, a finding that has also been observed in adult anxiety. This underscores the potential danger of interpreting between-group differences in cross-sectional studies of adults with versus without a given disorder with respect to interpreting between-group differences; although investigators may be inclined to attribute such differences to the pathophysiology of disease, such experiments cannot rule out the possibility that observed differences instead represent markers of risk. To fully address these confounding influences, investigators must study comparison groups of participants who have the risk factor but not the full disease. Moreover, brain-based indicators of risk must account for age, sex, medication effects, and other pertinent temporal variables (e.g., menstrual phase, diurnal variation).

Overall, developmental neuroimaging studies have provided critical insight into the neural substrates of fear and anxiety in children and adolescents, with initial evidence of how developmental processes may relate to risk for anxiety disorders. However, developmental neuroimaging efforts must also contend with unique challenges, such as the need to dissociate disorder effects from developmental processes. Furthermore, methods must be adapted to be developmentally appropriate. Fear conditioning paradigms used with children and adolescents have traditionally employed UCSs that have been adapted to be appropriate for children and adolescents, such as appropriately aversive noises, air puffs, or images. Moreover, neuroimaging research in children and adolescents has primarily relied on fMRI paradigms to avoid exposing children to the limited radiation inherent to PET studies.

Due to high rates of comorbidities among anxiety disorders in youth, a large proportion of studies have examined pediatric anxiety disorders as a unit rather than distinguishing among specific disorders. For example, across a sample of youth with GAD, SAD, and/or separation anxiety, increased activation in the inferior and superior frontal gyri (including the dlPFC and vlPFC) was associated with improved responses to either CBT or pharmacologic treatment. The bulk of research on specific types of anxiety disorders has primarily focused on SAD and GAD, whereas PD, specific phobias, selective mutism, and separation anxiety disorder remain less explored. Future work will be necessary to elucidate the neural substrates of these disorders during development and to understand predictors of treatment response in children and adolescents with specific anxiety disorders.

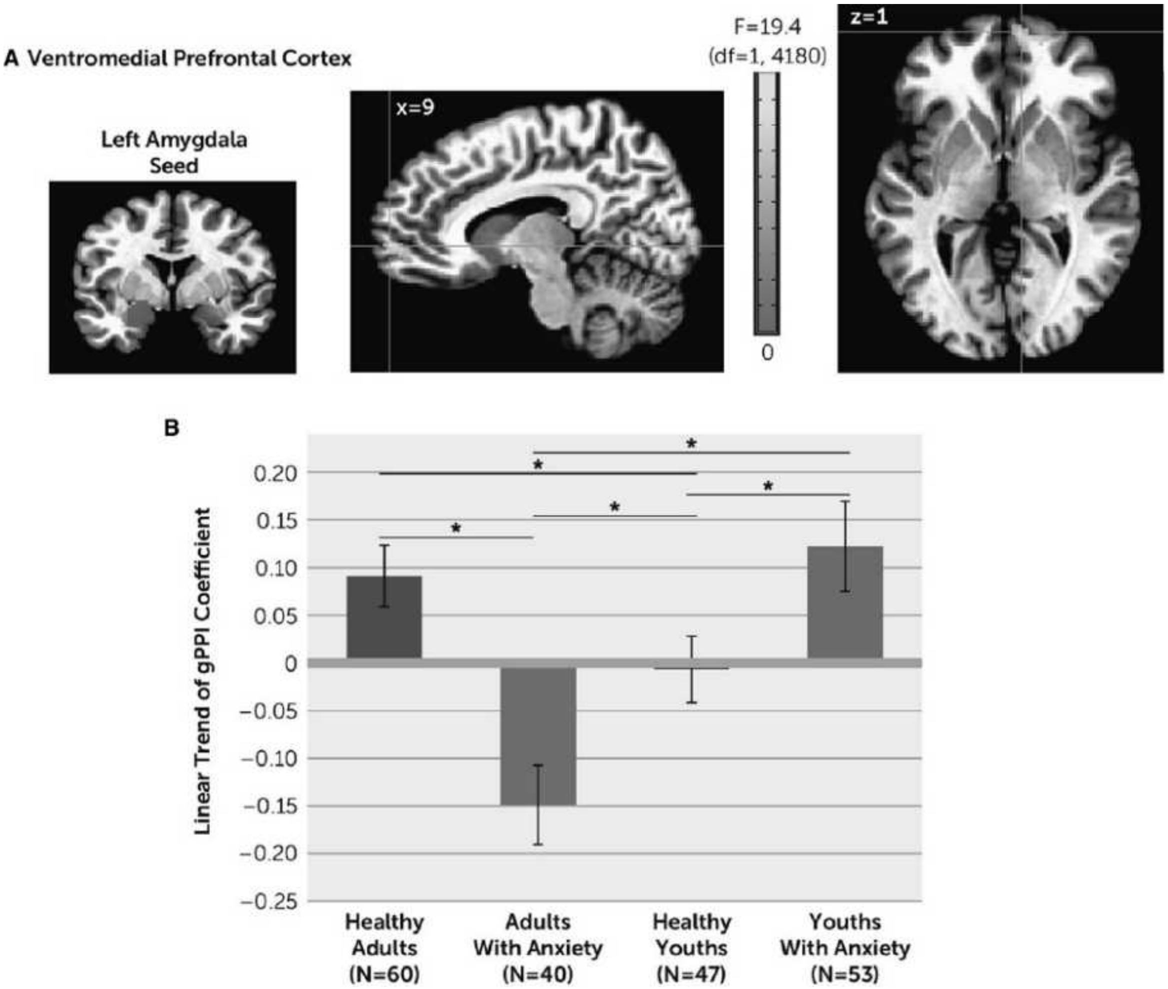

FIGURE 14.5–4. Left amygdala connectivity with the ventromedial prefrontal cortex (vmPFC) among youths and adults with anxiety disorders and healthy subjects. Whole-brain generalized psychophysiologic interaction (gPPI) analysis revealed a significant interaction of anxiety diagnosis, age, and linear trend in the task-related functional connectivity between the left amygdala seed and the vmPFC (panel **A**). Images are shown in neurologic convention (i.e., left is left) and thresholded at F > 10.76, df = 1, 4180, *P* < .001, cluster size > 57 voxels (890.625 mm^3). To decompose the complex interaction effects, mean extracted values for the vmPFC cluster are plotted separately by group, based on anxiety diagnosis (healthy, anxiety) and age (median split: adults, youths). In the graph (panel **B**), the y-axis shows the linear trend of the gPPI coefficient averaged across participants in each group. Error bars indicate standard deviation. (See eBook for color images.) (Reprinted with permission from the Gold AL, Abend R, Britton JC, et al. Age differences in the neural correlates of anxiety disorders: an fMRI study of response to learned threat. *Am J Psychiatry*. 2020;177(5):460.)

Social Anxiety Disorder in Children and Adolescents. Social stimuli become increasingly salient in normative development in adolescence.

Thus, it is perhaps unsurprising that the incidence of SAD peaks during this developmental period. Using increasingly naturalistic paradigms, findings suggest important changes in brain responses to social stimuli in adolescents with SAD. For example, during a paradigm mimicking an online chatroom, adolescents with SAD showed increased amygdala activation when anticipating evaluation from peers they rated as undesirable for social interaction (“rejected peers”). These adolescents also displayed increased positive vmPFC-amygdala connectivity, which correlated with symptom severity. Other work with this task revealed that socially anxious adolescents show reduced activation in the nucleus accumbens when anticipating feedback from peers they rated as desirable (“selected peers”). Moreover, anxiety interacted with age. Among socially anxious participants, age and rACC activation were positively associated with rejected peers but negatively associated with selected peers; however, the opposite pattern was observed for controls. Interpreting these findings, the authors proposed that increased regulatory resources may be devoted to potentially adverse social outcomes as individuals with SAD mature. An age-by-anxiety interaction was also observed in striatal activation and connectivity in an fMRI paradigm in which participants received feedback from peers. Adolescents with SAD showed increased striatal activation and negative frontostriatal functional connectivity during unexpected positive feedback from highly valued peers, but this pattern was not detected for adults with SAD or for nonanxious adults or adolescents. Finally, a longitudinal study using both EEG and MRI reported that N400 amplitudes (the event-related potential that occurs at 400 ms after stimulus presentation believed to relate to processing of information in human faces) at ages 8 to 9 in healthy participants predicted social anxiety symptoms at age 14 to 15. Although preliminary, this suggests that limbic reactivity may predate or index risk for SAD. While much remains unknown about the interaction between anxiety and developmental processes, these findings emphasize the importance of investigating potential differences in the mechanisms underlying SAD across development.

Generalized Anxiety Disorder in Children and Adolescents. In line with findings in adults, studies of youth with GAD have reported inconsistent findings regarding amygdala reactivity. In one study, youth with GAD showed elevated amygdala, vPFC, and ACC activation to fearful faces than happy faces when attending to their own fear, whereas this pattern of elevated activation was not seen in control youth. In contrast, another study did not detect differences in amygdala reactivity to angry faces in youth with GAD versus controls. However, this may be a function of the design of this second study, as subsequent studies have shown heightened amygdala activity to masked angry faces. Youth with GAD's amygdala reactivity has been found to positively correlate with the severity of their anxiety, and they also exhibit weaker vlPFC-amygdala functional connectivity. Furthermore, youth with GAD show elevated vlPFC activation, and vlPFC activation is inversely correlated with the severity of anxiety. Together, these findings suggest that the vlPFC may play a compensatory role in pediatric GAD. Though less is known about resting-state functional connectivity in pediatric GAD, evidence suggests that, as in adults, adolescents with GAD also display altered amygdala-mPFC functional connectivity at rest. In addition to these findings of functional alterations, several preliminary studies point to potential alterations in brain structure underlying pediatric GAD. To date, existing data suggest that youth with GAD have larger amygdala volumes and increased cortical thickness in regions including the vmPFC, inferior and middle temporal cortex, and lateral occipital cortex.

Initial evidence points to possible neural predictors and mechanisms of effective treatment for pediatric GAD. Higher amygdala reactivity pretreatment has been associated with greater symptom improvement in youth with GAD following CBT or psychopharmacologic treatment. Furthermore, these treatments appear to be associated with higher post-treatment vlPFC activation. Across these characterization and treatment findings, additional studies will be necessary to evaluate the extent to which these developmental findings are similar to or distinct from adults with GAD.

Summary of Neuroimaging Findings in Pediatric Samples. Children and adolescents are highly vulnerable to anxiety disorders. The growing literature on pediatric anxiety has begun to identify common neural mechanisms that are known to underlie anxiety disorders in adults, such as alterations in the amygdala and PFC. However, further research is necessary to elucidate the dynamic longitudinal developmental changes that occur across childhood and adolescence in youth with anxiety disorders. Samples that span across developmental periods to include children, adolescents, and adults will contribute to our understanding of anxiety- and stress-related changes to neurodevelopmental trajectories. Finally, identifying sensitive periods and risk/resilience factors will ultimately be essential for early identification and intervention efforts and for optimizing treatments for youth.

FUTURE DIRECTIONS

Using neuroimaging techniques to investigate the neurobiological bases of anxiety disorders continues to yield seminal information regarding the functional anatomical and neurochemical correlates of both normal and pathologic anxiety and fear states. Converging evidence from the systems neuroscience data obtained from human neuroimaging studies and the physiological and anatomical data obtained from preclinical studies with experimental animals has facilitated the development of general neurocircuitry models of anxiety disorders. Moreover, studies of specific disorders (PD, specific phobias, SAD, and GAD) have supported disease models that involve altered function of and interactions among the amygdala, hippocampus, and medial and orbital PFC.

Investigating the neurodevelopmental processes associated with pediatric anxiety disorders will be crucial to support the early identification of risk and optimization of treatment for youth with anxiety disorders. Innovative study designs, including developmental studies, are also needed to disentangle core substrates of pathophysiology from confounding neural substrates of risk, compensatory processes, and epiphenomena. Moreover, the neurobiological heterogeneity encompassed within criteria-based psychiatric diagnoses presents an additional challenge, which may be reflected in the mixed findings observed across some studies. Research that integrates neuroimaging with genetic and neurochemical data may help to characterize this heterogeneity. Ultimately, such integrative work is optimally poised to contribute to the refinement of phenotypes of anxiety disorders, the elucidation of their pathogenesis, and the development of personalized medicine approaches that tailor treatments based on individual differences and developmental stages.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIH) Director's Early Independence Award (DP5OD021370) to D.G.G., Brain and Behavior Research Foundation (National Alliance for Research on Schizophrenia and Depression; NARSAD) Young Investigator Award to D.G.G., Jacobs Foundation Early Career Research Fellowship to D.G.G., The Society for Clinical Child and Adolescent Psychology (Division 53 of the American Psychological Association) Richard "Dick" Abidin Early Career Award and Grant to D.G.G., National Institute of Mental Health (R61MH115113) to D.G.G., National Science Foundation Graduate Research Fellowship Program Awards to E.M.C. and E.R.K. (NSF DGE-1752134), The Society for Clinical Child and Adolescent Psychology (Division 53 of the American Psychological Association) Donald Routh Dissertation Grant to E.M.C., the American Psychological Foundation Elizabeth Munsterberg Koppitz Child Psychology Graduate Fellowship to E.M.C., a Dissertation Funding Award from the Society for Research in Child Development to E.M.C., and a Dissertation Research Award from the American Psychological Association to E.M.C.

FURTHER READINGS

Blair K, Geraci M, Devido J, et al. Neural responses to self- and other referential praise and criticism in generalized social phobia. *Arch Gen Psychiatry*. 2008;65(10):1176–1184.

Buff C, Brinkmann L, Bruchmann M, et al. Activity alterations in the bed nucleus of the stria terminalis and amygdala during threat anticipation in generalized anxiety disorder. *Soc Cogn Affect Neurosci*. 2017;12(11):1766–1774.

Charney DS, Drevets WC. The neurobiological basis of anxiety disorders. In: Davis K, Charney DS, Coyle J, Nemeroff CB, eds. *Neuropsychopharmacology. The Fifth Generation of Progress*. Lippincott Williams & Wilkins; 2002:901.

Chen Y, Cui Q, Fan YS, et al. Progressive brain structural alterations assessed via causal analysis in patients with generalized anxiety disorder. *Neuropsychopharmacology*. 2020;45(10):1689–1697.

Drysdale AT, Hartley CA, Pattwell SS, et al. Fear and anxiety from principle to practice: implications for when to treat youth with anxiety disorders. *Biol Psychiatry*. 2014;75(11):e19–e20.

Dvir M, Horovitz O, Aderka IM, Shechner T. Fear conditioning and extinction in anxious and non-anxious youth: a meta-analysis. *Behav Res Ther*. 2019;120:103431.

Etkin A, Prater KE, Hoeft F, Menon V, Schatzberg AF. Failure of anterior cingulate activation and connectivity with the amygdala during implicit regulation of emotional processing in generalized anxiety disorder. *Am J Psychiatry*. 2010;167(5):545–554.

Etkin A, Wager TD. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am J Psychiatry*.

2007;164(10):1476–1488.

Gold AL, Abend R, Britton JC, et al. Age differences in the neural correlates of anxiety disorders: an fMRI study of response to learned threat. *Am J Psychiatry*. 2020;177(5):454–463.

Goldin PR, Ziv M, Jazaieri H, Hahn K, Heimberg R, Gross JJ. Impact of cognitive behavioral therapy for social anxiety disorder on the neural dynamics of cognitive reappraisal of negative self-beliefs: randomized clinical trial. *JAMA Psychiatry*. 2013;70(10):1048–1056.

Guyer AE, Lau JYF, McClure-Tone EB, et al. Amygdala and ventrolateral prefrontal cortex function during anticipated peer evaluation in pediatric social anxiety. *Arch Gen Psychiatry*. 2008;65(11):1303–1312.

Hauner KK, Mineka S, Voss JL, Paller KA. Exposure therapy triggers lasting reorganization of neural fear processing. *Proc Natl Acad Sci USA.* 2012;109(23):9203–9208.

Klumpp H, Keutmann MK, Fitzgerald DA, Shankman SA, Phan KL. Resting-state amygdala-prefrontal connectivity predicts symptom change after cognitive behavioral therapy in generalized social anxiety disorder. *Biol Mood Anxiety Disord*. 2014;4(1):14.

Kujawa A, Swain JE, Hanna GL, et al. Prefrontal reactivity to social signals of threat as a predictor of treatment response in anxious youth. *Neuropsychopharmacology*. 2016;41(8):1983–1990.

LaBar KS, Gatenby C, Gore JC, LeDoux JE, Phelps EA. Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron*. 1998;20:937–945.

LeDoux JE. *The Emotional Brain*. Simon and Schuster; 1996.

Lissek S. Toward an account of clinical anxiety predicated on basic, neurally mapped mechanisms of Pavlovian fear-learning: the case for conditioned overgeneralization. *Depress Anxiety*. 2012;29(4):257–263.

McClure EB, Monk CS, Nelson EE, et al. Abnormal attention modulation of fear circuit function in pediatric generalized anxiety disorder. *Arch Gen Psychiatry*. 2007;64(1):97–106.

Monk CS, Telzer EH, Mogg K, et al. Amygdala and ventrolateral prefrontal cortex activation to masked angry faces in children and adolescents with generalized anxiety disorder. *Arch Gen Psychiatry*. 2008;65(5):568–576.

Munsterkotter AL, Notzon S, Redlich R, et al. Spider or no spider? Neural correlates of sustained and phasic fear in spider phobia. *Depress Anxiety*. 2015;32(9):656–663.

Myers KM, Davis M. Behavioral and neural analysis of extinction. *Neuron*. 2002;36(4):567–584.

Nash JR, Sargent PA, Rabiner EA, et al. Serotonin 5-HT 1A receptor binding in people with panic disorder: positron emission tomography study. *Br J Psychiatry*. 2008;193(3):229–234.

Pattwell SS, Lee FS, Casey BJ. Fear learning and memory across adolescent development: hormones and behavior special issue: puberty and adolescence. *Horm Behav*. 2013;64(2):380–389.

Phelps EA, Delgado MR, Nearing KI, LeDoux JE. Extinction learning in humans: role of the amygdala and vmPFC. *Neuron*. 2004;43(6):897–905.

Pine DS. Research review: a neuroscience framework for pediatric anxiety disorders. *J Child Psychol Psychiatry*. 2007;48:631–648.

Rauch SL, Drevets WC. Neuroimaging and the neuroanatomy of stress-induced and fear circuitry disorders: the agenda for future research. In: Andrews G, Charney DS, Sirovatka PJ, Regier DA, eds. *Stress-Induced and Fear Circuitry Disorders: Refining the Research Agenda for DSM-V*. American Psychiatric Association; 2008;235.

Shin LM, Liberzon I. The neurocircuitry of fear, stress, and anxiety disorders. *Neuropsychopharmacology*. 2010;35:169–191.

Tromp DPM, Grupe DW, Oathes DJ, et al. Reduced structural connectivity of a major frontolimbic pathway in generalized anxiety disorder. *Arch Gen Psychiatry*. 2012;69(9):925–934.

Whalen PJ, Phelps EA. *The Human Amygdala*. The Guilford Press; 2009.

14.6 Anxiety Disorders: Cognitive-Behavioral Therapy

JONATHAN D. HUPPERT, PH.D., AND EDNA B. FOA, PH.D.

This section focuses on cognitive-behavioral therapy (CBT) for *pathological anxiety* that may lead to significant impairment in functioning or substantial distress. The fifth edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) specifies five primary anxiety disorders: phobias, panic disorder, agoraphobia, social anxiety disorder, and generalized anxiety disorder (GAD). Common to the anxiety disorders are intense anxiety related to exaggerated and/or unrealistic cognitions about future threats of harm (either physical or psychological), physiologic responses when confronted with anxiety-relevant stimuli, and behavioral tendencies to escape from or avoid anxiety triggers and to prevent anticipated harm. Whereas the DSM-5 includes obsessive-compulsive disorder (OCD), post-traumatic distress disorder (PTSD), and illness anxiety disorder (hypochondria) in separate categories from the anxiety disorders, most of the principles that will be discussed here also apply to both OCD and PTSD. With the exception of panic disorder, in all anxiety disorders, the focus is on the harm related to the specific source of fear rather than on the panic itself.

ASPECTS OF THE ANXIETY DISORDERS PARTICULARLY RELEVANT TO CBT

Each anxiety disorder is distinguished by specific symptoms and specific negative cognitions. Panic attacks in panic disorder, for example, are discrete episodes of intense anxiety associated with cognitive symptoms (e.g., thoughts that one is going to die, go crazy, or lose control), along with a fear that the attacks will recur. The central feature of panic disorder, according to CBT theories, therefore, is the fear of one's own anxiety reactions because they are perceived as harmful (also known as "fear of fear"). Conceptualizations similar to the fear of fear for panic disorder can be made for social anxiety disorder, in which the focus is on the fear of embarrassment and rejection; OCD, in which the fear is of an idiosyncratic concern (contamination, fire, being an evil or harmful person, etc., often related to one's values and being responsible for negative outcomes); generalized anxiety, in which the worry is about the uncertainty of future negative outcomes; PTSD, in which the fear is about the consequences of thinking about the trauma or approaching situations that are similar to the trauma with the expectation that of another traumatic event occurring in these situations, or inability to tolerate distress while thinking about the trauma; and specific phobia, in which the fear is circumscribed to a specific situation or object.

Fortunately, effective psychological and pharmacologic treatments have been developed for the anxiety disorders. In fact, CBT has been endorsed as a first-line treatment for anxiety disorders by the United Kingdom's National Institute of Health and Clinical Excellence, the American Psychiatric Association Treatment Guidelines, and many other healthcare and government agencies around the world. These guidelines have been influenced by research suggesting that CBT for anxiety disorders is more cost-effective than medication or other treatments in the long term, and patient preferences tend to be 3:1 for psychosocial interventions over medical ones. We now discuss the theory and practice of CBT for anxiety disorders and review evidence of its efficacy.

EMOTIONAL PROCESSING THEORY: A PSYCHOLOGICAL MODEL OF ANXIETY DISORDERS AND THEIR TREATMENT

Before describing the components of CBT and its efficacy in treating various anxiety disorders, a general psychological theory of anxiety disorders and the hypothesized conditions necessary for successful treatment are presented. Emotional processing theory, developed by Edna Foa and Michael Kozak, is based on the idea that a cognitive *fear structure* that includes representations of stimulus, response, and meaning serves as a blueprint for avoiding or escaping from danger. The theory integrates information processing concepts and recent learning theories of acquisition and extinction. It posits that environmental stimuli activate matching information in the fear structure, resulting in the spreading of activation throughout the rest of the structure via associative connections, thereby activating the fear structure. A fear structure is adaptive if the stimuli represented in it are objectively harmful and the responses represented in it lead to effectively avoiding, escaping, or coping with the threat. A fear structure is pathologic, as is the case in anxiety disorders when the associations among stimulus, response, and meaning representations do not accurately reflect reality, such that harmless stimuli or responses are erroneously interpreted as being dangerous. For example, the fear structure for an individual with a dog phobia would include associations among representations of dogs; meaning representations of dogs as being dangerous and uncontrollable; and responses such as physiologic arousal and various defensive behaviors (e.g., calling for help, running away). The fear structure in panic disorder is characterized by fear of physical sensations associated with panic symptoms (e.g., rapid breathing, heart palpitations) and interpretation of these sensations as indications of danger, such as a heart attack or going crazy. As a result of the erroneous meaning associated with these sensations, people with panic disorder avoid situations that they anticipated to give rise to panic or cause similar sensations, such as intense physical activity. The fear structure of social anxiety disorder represents an erroneous association between social interactions and negative consequences, such as being perceived as boring or stupid, which will lead to rejection. The fear structure in GAD is much more diffuse, as there is often no central or “core fear.”

Emotional Processing

Emotional processing refers to the changes in pathologic elements of fear (emotional) structure (e.g., most dogs are not dangerous) that result in long-term fear reduction and recovery of the anxiety disorder. According to emotional processing theory, two conditions are necessary for emotional processing to take place. First, the fear structure must be activated either through a match with stimuli in the environment (e.g., confronting the feared stimulus) or through symbolic means (e.g., imagining or thinking about the feared stimuli). Second, new information that is *incompatible* with the pathologic aspects of the fear structure must be available and incorporated into the structure, thereby altering the structure and creating a realistic or “nonfear” structure via the new, competing associations. These two conditions have been supported via research in animal and human studies, the latter of which includes both basic research and treatment process research. Indeed, this theory is consistent with current models of neuroscience, suggesting that the brain is constantly scanning, processing the environment, and making predictions. When surprising information counter to predictions occurs, this leads to a prediction error (a signal that something needs to be processed), and oftentimes error correction, or updating one’s predictive model.

Emotional processing theory explicitly aimed to explain the phenomenology and mechanisms of anxiety disorders and the efficacy of exposure therapy in treating pathologic anxiety. Exposure therapy involves helping patients systematically confront feared but otherwise safe stimuli in a manner that promotes eventual fear reduction. The efficacy of exposure therapy in reducing anxiety has been widely demonstrated, and exposure procedures (i.e., confrontation with feared stimuli and responses followed by changes in their meaning) are a basic component of nearly every CBT program that has been shown to be effective in the treatment of the various anxiety disorders. The relevance of emotional processing theory to other treatments for anxiety, specifically anxiety management training and cognitive therapy, is addressed later.

Recent elaborations of emotional processing theory specify two indicators of emotional processing that can be measured separately from treatment outcome. Following from the proposal that activation of the fear structure is necessary for change in the structure to occur, one indicator of emotional processing is sufficient emotional activation of the fear structure (emotional engagement), as reflected in the assessment of fear and distress such as facial reactions, self-reports of fear and anxiety, or physiologic reactions such as increased skin conductance or heart rate when the person confronts or thinks about the feared stimulus or response. The second indicator of emotional processing is a reduction in the levels of fear or distress across repeated sessions, or *between-session habituation (extinction or learning)*. Fear reduction obtained in one session that does not at least partially transfer to a subsequent session (i.e., the lack of between-session learning) indicates that the fear structure was not altered and suggests that any previously observed fear reduction may have been brought about by some kind of compulsion, safety behavior, or other form of defensive tactic. By contrast, fear reduction that does transfer across sessions strongly suggests that the exposure experiences have led to some enduring change in the fear structure (typically of the meaning of the stimuli and or responses). Other indicators of between-session emotional processing may include reductions in fear-related cognitions or changes in biased cognitive processes, such as reductions in the investment of attentional resources to threat stimuli or decreased negative and/or increased positive interpretations of ambiguous situations.

Earlier formulations of emotional processing theory included a third indicator of emotional processing: Gradual reduction in these indices of emotional reactivity to the feared situation over the course of an exposure exercise. In the context of exposure therapy, this is called *within-session habituation*. However, acute fear reduction can occur for reasons other than habituation, such as distraction or engaging in compulsions or other safety behaviors. Such measures limit emotional engagement and prevent the encoding of disconfirming information, thereby preventing therapeutic changes from taking place in the fear structure. For example, if a person with panic disorder fears dying from a panic attack experiences an attack, thus activating the fear structure, but then immediately sits down and drinks water to prevent dying, they will not learn that the panic attack is not dangerous and attribute the lack of anticipated harm to drinking water. Given the potential unreliability of this indicator and many null findings, revisions of emotional processing theory suggest that within-session habituation, although a possible indicator, is neither necessary nor sufficient.

Several studies conducted across a range of disorders, including specific phobias, social anxiety disorder, panic disorder, OCD, and PTSD, investigated the relationship between indicators of emotional processing and treatment outcomes of exposure therapy. Many of these studies showed a relationship between initial activation and long-term outcome and between changes in either cognition or fear between sessions and outcome. The relationship between within-session changes in fear or cognitions and outcome was less consistent, though some recent studies using more fine-grained analyses have found positive results.

Sources of Corrective Information and Conditions That Interfere With Emotional Processing

It is relatively easy to see how exposure therapy promotes emotional engagement as patients are encouraged to think and talk about feared thoughts, images, and memories; confront feared objects and situations; and engage in feared activities. It may not be as easy to see what sources of corrective or fear-disconfirming information are contained in an exposure exercise.

One important source of corrective information is within-session fear reduction, which provides new response and meaning information embedded in the experience of the *reduction* of physiologic responding in the presence of feared stimuli that is incompatible with prior response information and the incorporation of nonthreat meaning related to this response. Take, for example, a patient with panic disorder; the experience of habituation disconfirms their belief that anxiety will persist unless the person escapes the situation, and it corrects misperceptions that physiologic responses associated with fear and anxiety lead to heart attack or to going crazy. In-session activation also frequently teaches the patient that the magnitude of the anxiety is not as great as initially expected. Experiencing the anxiety also frequently teaches the patient that they can tolerate the anxiety and that it does not last indefinitely.

Repeated and prolonged exposure may also provide corrective information regarding the actual likelihood or significance of feared consequences. Anxious people frequently overestimate the probability of feared events and the significance of those consequences, should they happen. For example, a person with social anxiety may make an unrealistically high estimate of the probability of being rejected in social interactions. Repeated exposure allows the person to test whether the feared consequence happens. Moreover, even if it does happen—for example, the person says something embarrassing—the person could learn that these consequences, although unpleasant, are not catastrophic.

Exposure to intrusive and unwanted thoughts can help a person distinguish between unpleasant thoughts that occur in their mind and reality or separate thoughts from overt actions. Exposure therapy allows these individuals to learn that they can think of upsetting thoughts without acting on them and that thinking about upsetting consequences does not make them happen.

Explicit provision of corrective information can also be extremely important and facilitate new associations within a fear structure. For example, psychoeducation about the causes and consequences of panic attacks can substantially reduce anxiety and panic attacks in approximately 30% of patients with panic disorder. This education includes corrective information regarding the likelihood of fainting because of panic attacks (which is close to zero in individuals other than those with blood-injection phobias) and distinguishing heart attacks and strokes from panic attacks, and a description of the biopsychosocial model of panic. On the other hand, many individuals require more experiential exercises to incorporate such corrective information. For example, individuals with social anxiety disorder can benefit significantly from the provision of video and confederate feedback after an exposure to make explicit some of the corrective information that would otherwise not be observed. At times, within-session habituation does not occur in these circumstances, but between-session habituation does, making the latter a more reliable indicator of emotional processing than the former.

Failures of Emotional Processing

If repeated exposure to feared stimuli contains corrective or fear-disconfirming information, then why do most anxiety disorders have a chronic course? An important feature of all anxiety disorders, in addition to the experience of intense distress on exposure to the presence of the feared stimuli, is a strong tendency to escape from feared situations or avoid them altogether. To the extent that such escape and avoidance strategies are effective at reducing anxiety, they are strengthened through the principle of negative reinforcement, making it even more likely that the person will engage in escape or avoidance in the future. Moreover, successful escape or avoidance limits the person’s exposure to fear-disconfirming information. For example, a person who is afraid of dogs and successfully avoids even friendly dogs will not have the opportunity to learn that dogs can be not only safe but also fun to play with, and thus they will maintain the belief that avoiding dogs has prevented harm from occurring to them.

One intervention studied in the context of exposure therapy that has been found to promote acute reductions in distress while in the presence of feared stimuli but that interferes with between-session habituation is the use of benzodiazepine medications. When taken shortly before exposure to a feared stimulus, benzodiazepines can significantly reduce anxiety. As with distraction, however, the fear reduction that occurs with the medication does not transfer to subsequent tests in a nonmedicated condition and, indeed, appears to interfere with the normal process of cognitive changes and between-session habituation. Researchers demonstrated this among a group of patients undergoing treatment for fear of flying that involved taking two flights scheduled one week apart. Ninety minutes before taking the first flight, half of the patients took the high-potency benzodiazepine alprazolam, and the remaining patients took a placebo. Patients taking the active medication reported significantly less anxiety during the first trip than patients who took the placebo. Before the second flight, all patients were given a placebo pill. Patients who had taken the placebo before the first flight showed a reduction in their anxiety on the second flight, whereas those who had taken the active medication showed an increase in anxiety on the second flight, showing lack of extinction without fear activation. Similar findings were found in animal studies of extinction.

Another intervention that has led to more mixed findings is distraction during exposure. Studies have differed substantially in how distraction was manipulated, the population that was studied, and the duration of the exposure sessions, which may account for some of the variability in results across studies. Overall, a recent meta-analysis showed that, under some circumstances, distraction can produce acute reductions in anxiety that are equivalent to focused exposures. Results suggest that these effects are found more when multiple sessions are studied, and distraction is more interpersonal/interactive in nature. It is possible that engagement of others during exposure recruits interpersonal emotion regulatory systems or that implicit learning via modeling or subtle verbal or nonverbal cues facilitates processing. Alternatively, such interactions may help facilitate engagement of approach/reward systems and deactivation of threat/fear systems. A more recent theory suggests that distraction is an effective emotion regulation process when emotions are too intense and resources are viewed as too limited to cope. Distraction in such cases may moderate anxiety to a point that it is manageable to process. At times, uninstructed exposure seems to be superior to either focused or distracted exposure, suggesting that distraction may not be effective in some circumstances. It should be noted that the degree of distraction (e.g., complete engagement of cognitive resources on another task versus divided attention, splitting one's resources between the exposure and another task) may have different outcomes and that most of the studies conducted to date used analog participants or patients with specific phobia. Therefore, more research is needed for more complex anxiety and related disorders.

Other Approaches to Understanding the Mechanisms of Exposure and CBT for Anxiety Disorders

It is beyond the scope of this chapter to describe in detail all the theories attempting to explain the mechanisms of exposure and CBT for anxiety disorders. However, two additional approaches should be noted, given their prominence in the literature. First, the belief disconfirmation approach was developed first by Aaron T. Beck and colleagues. The main aspect of this model is that thoughts and beliefs lead to emotions and behaviors. This approach was initially called cognitive therapy to contrast it with behavior therapy and exposure approaches. The belief disconfirmation model suggests that explicit discussions, contrived "experiments," and experiential exercises are used to dispel maladaptive beliefs the patient holds that are not realistic. Typically, such models do not attempt to explain implicit learning or concepts such as conditioning, habituation, or extinction. Ultimately, cognitive theories have become integrated with behavioral theories, including emotional processing theory, and their theoretical and practical distinctions are at most subtle. Indeed, in practice, most researchers and clinicians today call their treatments cognitive behavioral therapy rather than pure behavioral or pure cognitive therapy.

A second approach that has been used recently by Michelle Craske and colleagues to contrast with emotional processing theory is the inhibitory learning approach. The basis of inhibitory learning is that exposure leads to fear reduction due to inhibition of fear associations. This model is based predominantly on modern, neo-classical learning theory. Accordingly, fear is thought to occur due to pairing of neutral stimuli with an aversive stimulus (classical conditioning). Fear is then "unlearned" through both habituation and extinction, leading to an inhibition of the association. The primary issue suggested here is that associations are not replaced but only inhibited. Thus, they can be reactivated when the inhibitory processes weaken or fail, leading to relapse. There are competing models with the idea of inhibitory processes, such as retrieval competition (suggesting that new and old memories both exist and compete to be expressed) and memory reconsolidation (suggesting that memory is constantly being reconstructed and not static). Some suggested implications of the inhibitory model are that it is essential to design exposures to maximize expectancy violation (or prediction errors; common in many CBT models today, including cognitive therapy and emotional processing), remove safety signals, vary exposures, and change contexts of exposures. Most of these implications are not unique to inhibitory theory and are included in most current CBT models, including those of belief disconfirmation and emotional processing, as well as models that are not readily integrated into those models. Thus, many of the models proposed to explain the mechanisms of exposure and CBT for anxiety use different language to ultimately describe very similar processes. It is likely that the best explanatory model integrates all of these models, thereby incorporating their large overlap and adopting the subtle unique differences.

COGNITIVE-BEHAVIORAL THERAPY PROCEDURES FOR ANXIETY DISORDERS

Most CBT programs for anxiety disorders begin with an assessment, patient education, and specific treatment planning. The actual treatment procedures involve at least one of four components: (1) exposure to thoughts, objects, situations, and physiologic sensations that are not dangerous but are nonetheless feared, avoided, or endured with great distress; (2) training in general anxiety or stress management techniques (including acceptance or mindfulness); (3) use of cognitive therapy techniques; and (4) training in specific skills, such as dating skills, assertiveness, and so forth.

Assessment, Psychoeducation, and Treatment Planning

CBT begins with a thorough assessment of the patient’s presenting complaints using empirically validated assessment procedures. Self-report and interviewer rating measures have been developed for the diagnosis and assessment of severity of each anxiety disorder. The assessment may also include measures of related psychopathology, such as depression and general anxiety, and specific theoretically relevant variables associated with the disorder of interest, such as worry-related metacognitions in GAD and anxiety sensitivity in panic disorder. Typically, assessments are repeated over the course and at the end of treatment for the purpose of establishing a diagnosis, selecting treatment procedures, monitoring the process of treatment, and evaluating treatment outcomes.

The next step of CBT is psychoeducation. Patients are typically provided with a cognitive-behavioral formulation of their specific diagnosis and a related treatment rationale that serves as a guiding framework for future treatment sessions. In addition, alternative treatments shown to have similar efficacy are discussed to allow the patient to make an informed choice about treatment.

As part of the initial assessment and patient education, the therapist and patient work together to identify specific targets for treatment and to work out the details of the treatment plan. An example of this is the development of a hierarchy for conducting in vivo exposure exercises or determining the patient's core beliefs and core fears to determine the proper exposures and their course. In addition, expectations are set that the patient will engage in homework between sessions to increase mastery of the therapy skills and self-monitoring procedures for symptoms.

Exposure Therapy

Exposure therapy involves intentionally confronting feared but otherwise not dangerous objects, situations, thoughts, memories, and physical sensations for the purpose of reducing fear reactions and avoidance associated with the same or similar stimuli. Systematic desensitization, where gradual imaginal and in vivo exposure was combined with relaxation, was the first exposure therapy technique to undergo scientific investigation. Although an effective treatment for some anxiety disorders, it has generally fallen out of use among researchers and cognitive-behavioral therapists due to the finding that concurrent relaxation was deemed unnecessary and possibly contraindicated. In addition, it does not seem necessary to progress systematically up the hierarchy (though it can be helpful at times). Finally, systematic desensitization did not appear helpful for OCD or agoraphobia, while more intense in vivo exposures were helpful for these as well as specific phobias. The contemporary use of exposure therapy may be usefully divided into three classes of procedures: In vivo exposure, imaginal exposure, and interoceptive exposure. These are typically conducted while encouraging the patient to prevent safety behaviors, or strategies used to prevent negative outcomes but prevent learning.

In Vivo Exposure. In vivo exposure involves helping patients directly confront feared objects, activities, and situations. It is usually conducted in a graduated fashion according to a mutually agreed-on (between patient and therapist) hierarchy. For example, a hierarchy for a specific animal phobia, such as snakes or spiders, may begin with looking at pictures and other representations of the feared animal, followed by looking at the actual animal kept in a cage, first at a distance and then gradually moving closer. Depending on the animal, subsequent steps may involve touching and handling the animal, perhaps first while wearing a glove and then without the glove. These steps may be repeated across several different examples of the animal, differing in such dimensions as size and activity level.

In the case of social anxiety disorder, individuals are asked to engage in conversations and other interpersonal situations. Recent evidence suggests that such exposures maximize the incorporation of corrective information by ensuring that patients with social anxiety make explicit, specific predictions prior to their exposures, focus their attention outward and drop safety behaviors during the exposure, and receive feedback from those they interacted with and by carefully planned postexposure review of videotapes of the exposure. In the case of panic disorder, individuals expose themselves to situations in which they fear having panic attacks, such as riding in elevators, waiting in line at a store, and so on. GAD does not have standard in vivo exposure as part of its treatment because there is not usually a specific source of fear.

Imaginal Exposure. Imaginal exposure typically involves having the patient close their eyes and imagine feared stimuli as vividly as possible. One common form of imaginal exposure is imaginal flooding, in which the patient imagines the worst possible scenario repeatedly in order to help them imagine the full stream of occurrences that would need to happen in order for the feared outcomes to occur. Such imaginal exposure helps the patient process the fear, which is often avoided both behaviorally (by avoiding contact with feared objects) and cognitively and emotionally (i.e., by avoiding thinking thoroughly about the feared consequences or avoiding how they would feel in the feared situations). Imaginal exposure has two general uses. The primary use is to help patients confront feared thoughts, images, and memories in order to process them. For example, individuals with social anxiety may experience extreme distress over thoughts of being rejected. Imaginal exposure to feared thoughts and consequences is used to promote reduction of emotional reactivity to the scenarios, to decrease the belief in the likelihood that the event will occur, to learn that the consequences are less severe than imagined, and to help the patient distinguish between thoughts and reality. In the case of phobias where there is a clear traumatic cause (e.g., fear of driving after a car accident), imaginal exposure is used to help the patient confront their memory of the traumatic event. Here again, the procedure promotes habituation to the memory and helps the patient to distinguish between the actual traumatic event, which was dangerous, and the current memory of the event, which, although distressing, is not harmful. With patients who have PTSD, imaginal exposure to traumatic memories is also used to help the patient get a new perspective of the trauma and how they behaved during the trauma. For example, a patient who was raped and blamed herself for not fighting during the rape might realize during imaginal exposure that she fought vigorously during the rape; a soldier who blamed himself for not saving his friend during the battle might realize during imaginal exposure that he was too far away from his friend to be able to save him. In these cases, imaginal exposure also creates or organizes a coherent narrative, which may help the patient to cope with their fears.

A second use of imaginal exposure is in lieu of in vivo exposure when arranging direct contact with the feared situation is not safe or feasible or as a preparatory exercise to facilitate subsequent in vivo exposures. In these cases, the patient imagines confronting the feared object or engaging in the feared task. When the patient’s level of anticipatory anxiety is reduced, in vivo exposure may be conducted with the actual feared object or task, assuming that it is both safe and available.

Recently, another imaginal procedure that builds from imaginal exposure has started to enter the repertoire of CBT: imagery rescripting. In imagery rescripting, after eliciting a past memory, one retells the memory from the perspective of the current self or another individual, helping the patient in the past cope better, get their needs fulfilled, and ultimately distinguish the past event from their current life. They then retell the experience of the original memory, now incorporating the new information. This intervention has been shown effective for PTSD, and preliminary studies have suggested potential use in GAD, OCD, and more.

Interoceptive Exposure. Interoceptive exposure is the most recent form of exposure therapy to be introduced. This procedure is designed to induce feared physiologic sensations under controlled circumstances. Interoceptive exposure exercises are most commonly used in the treatment of panic disorder and certain specific phobias, for example, the fear of vomiting, in which internal cues such as those associated with physiologic arousal (e.g., rapid heart rate, dizziness, tingling in the fingertips) or gastric upset (e.g., after eating a spicy meal) elicit fear, anxiety, and further arousal.

Several specific exercises have been developed by David Barlow and colleagues to induce specific panic-like sensations. For example, hyperventilating produces shortness of breath, dizziness, and tingling in the fingers and around the mouth and can produce a sense of unreality. Breathing through a thin straw produces the sensation of not getting enough air. Spinning in a chair or spinning in place produces dizziness and, in some people, mild nausea. Other possible activities include aerobic exercise to induce rapid heart rate and shortness of breath, consuming mild stimulants (e.g., eating chocolate-covered espresso beans), or eating spicy foods. The goal is to find an activity that produces similar sensations to the ones that patients find unduly distressing and attempt to avoid (e.g., the core panic symptoms in panic disorder). As with other forms of exposure, the purpose of inducing these sensations is to learn that these sensations are not dangerous. By intentionally inducing these feared sensations, the patients can learn to dissociate the sensations from danger per se by learning that these sensations do not cause harm and also gain a sense of control and mastery over their anxiety by accepting it. These exercises also promote habituation of fear in response to them (i.e., the fear of fear or anxiety sensitivity) and distinguish these sensations, which are normal responses to the various exercises, from physiologic sensations that may actually signal a significant health problem. More recently, there have been suggestions and research supporting the use of interoceptive exposure as an exercise that assists patients in tolerating negative activating emotions (arousal), even outside the context of panic disorder. This is consistent with the notion that exposures (all types) in the context of acceptance of negative emotions (as opposed to attempts to suppress them) facilitate improvement during exposure.

Cognitive Therapy

The basic assumption of cognitive therapy is that people’s beliefs and appraisals of events and situations are at least as important in determining their emotional reactions as the actual events and situations. Problematic emotions such as anxiety are frequently the result of unrealistic and unhelpful beliefs about the world, the self, and others. The goal of cognitive therapy is to help patients to identify unhelpful cognitions and cognitive distortions (e.g., probability overestimation, catastrophic thinking, all-or-nothing thinking, overgeneralization, and only considering evidence that is consistent with existing beliefs) and to modify them. Three of the most common traditional cognitive therapy procedures are (1) the use of Socratic dialogue, in which the therapist uses a series of questions to help the patient identify and challenge unhelpful beliefs or to uncover evidence disconfirming fear-related beliefs; (2) the downward-arrow technique, in which the therapist helps to uncover deeper beliefs and meanings by repeatedly asking for greater clarification; and (3) the use of thought records, in which patients record automatic beliefs, list evidence for and against those beliefs, review potential cognitive errors that may be reflected in those beliefs, and then generate more realistic and helpful beliefs. All these procedures can be conducted formally as traditionally intended or in a more integrated fashion with exposures, as is more typically done in contemporary CBT. Indeed, behavioral experiments epitomize this integration. The goal of a behavioral experiment is to test beliefs about the nature and probability of feared consequences under conditions that will optimize exposure to, and the encoding of, disconfirming information. Accordingly, before conducting a behavioral experiment, the patient makes detailed predictions about what they believe will happen under a particular set of circumstances and then conducts the relevant experiment to see what happens. For example, consider a person with panic disorder who fears they will go crazy if they feel dizzy. A behavioral experiment to test predictions about the likelihood of going crazy might be to spin in a chair slowly at first, then work up to spinning quickly, and see what the consequences of feeling intense dizziness are. Each time patient feels dizzy without going crazy (operationalized idiosyncratically for the given patient), the patient is confronted with evidence that disconfirms their beliefs about the probability of going crazy. The conceptual emphasis in cognitive therapy on changing beliefs through exposure to disconfirming evidence is consistent with the tenets of emotional processing theory. The practical and empirical issues involve determining the best ways to accomplish this goal.

Anxiety Management and Stress Inoculation Training

Anxiety management training is a general approach to treating anxiety, of which Meichenbaum’s stress inoculation training is one specific example. Emotions in general, and stress and anxiety, reflect the activity of three loosely coupled and interacting “channels,” or modes, of responding: Cognitive-phenomenologic, physiologic, and behavioral. Stress inoculation training involves training patients in applying specific skills and techniques designed to address each of these response modes in a flexible manner tailored to the individual’s needs. The assumption underlying anxiety management training is that some people lack the skill of managing their anxiety and need to learn these skills. For example, spiraling negative cognitions that often occur during times of stress may be interrupted by the use of thought stopping, in which the patient thinks the word “stop” in their mind while vividly imagining the word “stop” and then replaces it with positive or coping self-statements. The physiologic manifestations of anxiety and stress (e.g., hyperventilation and its consequences, muscle tension, and gastric symptoms) may be addressed through training in diaphragmatic breathing and progressive muscle relaxation (e.g., Jacobson “tense and relax” method). Changes in overt behavior are promoted through role-playing and covert rehearsal, in which patients practice new behaviors while executing other coping skills either with another person or in their mind. Exposure to feared situations either in vivo or in the imagination is frequently a part of stress inoculation training but is usually for the purpose of practicing new skills and is accompanied by the use of other coping skills. From the perspective of emotional processing theory, stress inoculation training brings about successful coping with fear and anxiety, which provides new meaningful information related to personal competence that counters typical fear-related perception of oneself as incompetent.

Acceptance and Mindfulness-Based Training

Over the past two decades, there has been an increase in interest in combining exposure and other forms of procedures that aim at coping with anxiety via mindfulness, meditation, and a related concept frequently associated with mindfulness called “acceptance.” The interest is due to several CBT treatments that have their roots in Buddhism and have been integrated with humanism and CBT. Three of the most popular treatments are acceptance and commitment therapy (ACT), mindfulness-based stress reduction (MBSR), and mindfulness-based cognitive therapy. All of these and other versions have in common the idea that fighting one’s anxiety (in the case of anxiety disorders) leads to a vicious cycle of attempting to suppress or otherwise force the downregulation of anxiety, which typically backfires. Therefore, acceptance-based techniques emphasize the need to first acknowledge that the anxiety exists and to try to objectively and nonjudgmentally observe it. Importantly, much of this concept was part of earlier versions of exposure therapies, but the aspect of radical acceptance was not emphasized. While some argue that acceptance-based techniques are in direct contrast with habituation models, anxiety reduction models, and cognitive therapy, we believe that all these techniques likely facilitate one another and that such techniques are easily integrated into exposure. Indeed, acceptance techniques and their integration with the other procedures may facilitate emotional processing.

Interpersonal Skills Training

Research has suggested that some individuals with social anxiety disorder, as well as those with other anxiety disorders, sometimes have interpersonal deficits that contribute to the maintenance of their anxiety and potentially negatively affect treatment. Thus, some people with anxiety lack important skills needed to interact effectively with others. Consequently, they experience fewer interpersonal rewards and more punishments, leading them to avoid social interactions when possible, thereby further limiting their ability to acquire effective social skills and decreasing their social support. Accordingly, comprehensive treatment for anxiety provides explicit training to help patients acquire and use appropriate social skills and thereby reverse the negative cycle when necessary. These skills include assertiveness training and training in initiating, maintaining, and ending conversations. At times, it may be that training in such skills helps individuals have more confidence in their own skills and stop engaging in avoidance or safety behaviors. In addition, some research groups have suggested adding more interpersonal components to CBT to further help patients learn that the goals of interpersonal interactions are not to pass a test but rather to seek intimacy or simply to be in a relationship.

Efficacy of Cognitive-Behavior Therapy

Phobias and Agoraphobia. A large body of research, beginning in the mid-1960s, has clearly demonstrated the efficacy of several exposure therapy protocols in reducing simple phobias and fear of public speaking. This research establishes exposure therapy as the treatment of choice for these disorders. The earliest studies focused on systematic desensitization. In the first controlled study, systematic desensitization was compared with a waitlist control condition in the treatment of students with a fear of snakes. In the first phase of the study, participants in the desensitization group separately practiced relaxation and created a hierarchy. In the second phase, participants in the desensitization group paired the feared images from the hierarchy with relaxation. Neither group showed any changes during the first phase, but the desensitization group showed significant improvement across the second phase, whereas the control group showed no change.

In a follow-up study, desensitization was found to be superior to a pseudotherapy that, like desensitization, included training in relaxation and hierarchy construction. After the initial preparation, participants in the desensitization group paired the feared images with relaxation, whereas the pseudotherapy group engaged in positive imagery during relaxation. Another study compared the efficacy of five sessions of systematic desensitization for public speaking fears and other forms of social anxiety with waitlist and two alternative treatment conditions (insight-oriented psychotherapy and an attention-placebo condition in which subjects were given a placebo pill but told that it was a drug that reduced anxiety in stressful situations and then conducted a task that was said to be stressful but actually induced drowsiness). Results revealed that all three treatments were superior to the waitlist and that systematic desensitization was superior to the alternative treatments, which did not differ from one another. This pattern of outcome was maintained for up to 2 years after the completion of treatment.

Research in the 1970s began to investigate the “flooding” approach to imaginal and in vivo procedures, which involved repeated and prolonged exposure to fear cues of high intensity without relaxation. Results from the earliest of these studies provided evidence for the efficacy of all three approaches (systematic desensitization, imaginal exposure, and in vivo exposure), but the results were mixed regarding the relative efficacy of the treatments. One group found that both systematic desensitization and imaginal exposure were more effective than a nonspecific treatment control condition among a mixed sample of patients with specific phobias and patients with agoraphobia. However, they did not replicate the finding of superiority of imaginal exposure over desensitization for agoraphobia.

A more consistent pattern of outcome in favor of a particular form of in vivo exposure, called *participant modeling* or *guided mastery*, over other forms of exposure therapy in the treatment of phobias and agoraphobia has been reported in a series of studies by Albert Bandura and colleagues. In participant modeling, the therapist actively models handling the feared object for the patient and may use various “response induction aids” to help the patient successfully to handle the feared object. For example, in treating a patient with a spider phobia, the therapist may first demonstrate how to touch, handle, and control a tarantula. The therapist would then coach the patient in performing the same behaviors, first while wearing protective gloves and then without gloves. This active approach to in vivo exposure may be contrasted with more-passive approaches in which the patient may be instructed to sit and observe a snake in a cage and attempt contact with the object when their fear at the earlier step has decreased.

Systematic desensitization, prolonged imaginal exposure, and prolonged in vivo exposure have all been found to be effective in the treatment of phobias, but in vivo exposure was generally more effective than imaginal exposure, both with phobias and agoraphobia. Prolonged in vivo exposure is at least as effective as, and often more effective than, desensitization and prolonged imaginal exposure. The participant modeling variation of in vivo exposure is particularly effective in treating a range of phobias, including small-animal phobias, acrophobia, and agoraphobia. Some recent studies have suggested that in vivo exposure plus a mildly distracting social interaction is more effective and tolerable than full focus on the feared stimulus. Studies have also found virtual reality-based exposure therapy to be effective.

Panic Disorder. According to the cognitive-behavioral model of panic disorder, a consequence of experiencing unexpected panic attacks for some individuals is the development of a vicious cycle. The cycle begins with hypervigilance for somatic cues indicative of a panic attack that, if detected, are interpreted negatively (e.g., as an indicator of having a heart attack or going crazy). The negative interpretation of somatic sensations results in increased anxiety and arousal, spiraling into a panic attack. Agoraphobic avoidance may also develop as an attempt to avoid experiencing panic attacks in situations in which help may not be available. Cognitive-behavioral treatments based on this model contain several components: (1) psychoeducation regarding the cognitive-behavioral model of panic disorder, (2) cognitive restructuring to challenge catastrophic cognitions about the likelihood and significance of panic attacks, (3) interoceptive exposure to feared somatic sensations to reduce the “fear of fear,” and (4) in vivo exposure to avoided cues, particularly in cases in which agoraphobic avoidance has developed. Oftentimes, later phases of treatment include combinations of phases 3 and 4 to help the patient learn to deal with physical sensations in the context of formerly avoided situations. Relaxation and breathing retraining were early components of CBT protocols, but research has suggested that they are at best unnecessary and possibly lead to greater dropout or interference with treatment gains.

The efficacy of CBT programs based on this model has been demonstrated in a series of well-controlled studies and meta-analyses. Between 41% and 87% of patients receiving CBT are panic free after treatment, compared to 36% receiving relaxation only, 50% receiving alprazolam, 20% receiving imipramine (Tofranil), 13% to 36% receiving placebo, and 30% to 33% undergoing waitlist. In the one study that combined imipramine with CBT (Barlow et al., 2000), the combination treatment was found to be slightly more effective than CBT alone but not CBT combined with placebo. In this study, 6 months after treatment discontinuation, CBT alone and CBT plus placebo were superior to CBT plus imipramine, suggesting that medication may interfere with the maintenance of CBT gains after medication discontinuation. Several other research groups investigated variations of CBT, with similar results, replicating the superiority of CBT over relaxation, supportive counseling, and waitlist. Another large, multicenter study examined CBT as a first-line treatment for panic with agoraphobia. Individuals who did not respond to 10 to 12 sessions of CBT for panic and agoraphobia were randomized to either more sessions of CBT or paroxetine. Individuals who received paroxetine responded more quickly to treatment than continued CBT, though individuals who received CBT had caught up by 9-month follow-up. Individuals who responded to treatment were randomized to maintenance CBT or no maintenance treatment. Maintenance sessions decreased relapse significantly (5%) in comparison to those not randomized to maintenance treatment (18%; White et al., 2010).

Several studies attempted to identify which components add to treatment outcomes. Two studies found that adding cognitive restructuring to in vivo exposure substantially improved outcomes, although other researchers have not been able to replicate this. Another study compared CBT consisting of education, cognitive restructuring, and interoceptive exposure with the same program plus relaxation and found that adding relaxation did not improve outcomes. A 1997 study compared two treatments that incorporated cognitive restructuring and in vivo exposure. In addition, one included interoceptive exposure, and the other group substituted breathing retraining. Both treatments were associated with improvement, but the group receiving interoceptive exposure had superior outcome on panic frequency at post-treatment and follow-up assessments. In a study using a crossover design, patients were assigned to receive four sessions of cognitive therapy followed by four sessions of interoceptive exposure or the reverse. Both treatments were associated with improvement, and more improvement was observed for the first intervention administered, but there was no difference in the efficacy of the two components. An additional study compared cognitive therapy without interoceptive exposure to interoceptive exposure without cognitive therapy and found similar positive outcomes in both groups that were equivalent to other studies examining combined packages. Another study compared education, cognitive restructuring, and in vivo exposure with or without the inclusion of breathing retraining and found no reduction in treatment efficacy. Three variations of self-conducted exposure therapy—in vivo exposure, interoceptive exposure, and the combination of in vivo plus interoceptive exposure—were compared with a waitlist control in another study. All three treatments were superior to the waitlist condition, but there were no significant differences among treatment conditions.

Treatment for panic disorder has largely consisted of multicomponent packages consisting of education about the cognitive-behavioral model of panic along with one or more of the following: Diaphragmatic breathing, relaxation, cognitive therapy, in vivo exposure, and interoceptive exposure. Dismantling studies have been partially successful in identifying the active ingredients, which seem to be cognitive restructuring combined with in vivo or interoceptive exposure. Indeed, the addition of interoceptive exposure may be particularly helpful. In a meta-analysis of 43 controlled studies of treatment for panic disorder, the largest mean effect size was obtained for the combination of cognitive therapy plus interoceptive exposure (.88), followed by CBT without interoceptive exposure (.68), CBT plus medication (.56), and medication alone (.47). Moreover, dropout rates from CBT were lower than dropout rates from medication conditions (alone or in combination with therapy). Another review by the Cochrane Collaboration found 21 trials comparing CBT to medications and their combination. Results suggested that combination treatment is superior to medication alone. Combination treatment also had superior symptom reduction to CBT alone in the acute phase but also had more dropouts. After treatment discontinuation, no difference between CBT and combined treatment was found. Some trials showed that CBT for panic can also be helpful for treating medication-resistant patients, as well as for helping patients discontinue medications, including benzodiazepines and selective serotonin reuptake inhibitors (SSRIs). Recent trials also showed the ability to disseminate treatment for panic in community mental health clinics in addition to primary care. Other studies showed that CBT is effective for nocturnal panic and that focused CBT on panic also significantly reduces comorbid conditions. In addition, several trials suggested the efficacy of group CBT treatment for panic disorder. More recently, numerous studies have demonstrated the efficacy of online or internet-based, therapist-guided interventions for panic disorder. Finally, some studies showed promising results for early intervention with CBT in individuals at risk for developing panic disorder.

Social Anxiety Disorder

Individuals with social anxiety disorder experience marked and persistent fear and avoidance of one or more social or performance situations. The person may fear doing or saying something embarrassing or manifesting signs of anxiety (e.g., blushing, sweating, trembling) in such situations. Thus, social anxiety disorder can be conceptualized as a fear of embarrassment or social rejection. Most individuals with social anxiety disorders fear and avoid multiple situations, falling within the generalized subtype of social anxiety disorder. Recent research has suggested that individuals who suffer from social anxiety disorder have a fear of positive evaluation as well as a fear of negative evaluation. They also tend to have diminished positive affect, particularly due to a lack of pride (more than joy), which has been theorized to reflect some underlying low self-esteem.

Treatments for social phobia have generally consisted of exposure therapy, cognitive therapy, and the combination of exposure plus cognitive therapy (sometimes also including social skills training). A network meta-analysis summarized the outcome of 101 studies of 41 different treatments or control conditions for social anxiety. Most active treatments were superior to waitlist. However, individual CBT had the largest effect size (greater than group CBT and other treatments) and was significantly more effective than psychodynamic therapy, supportive therapy, or mindfulness. The results of individual dismantling studies yield an ambiguous picture. In the first such study, exposure alone was compared with exposure plus cognitive therapy. Although both groups improved, the outcome was somewhat better for the combined treatment, particularly at the 3-month follow-up assessment. In a second study by the same group, the efficacy of each component alone, as well as the combination of the two, was compared with a waitlist control group. Compared to the waitlist control group, all three treatments were found to be effective in reducing social anxiety. Again, however, outcome was best among subjects who received the combined treatment. By contrast, another study did not find evidence for superiority of combined treatment over exposure alone. On some measures, patients who received exposure only had numerically better outcome than those who received the combined program.

One treatment program that has received considerable research attention is Richard Heimberg’s cognitive-behavioral group therapy (CBGT), consisting of in-session exposure to feared social situations, cognitive restructuring, and corresponding homework. The group format not only serves to make treatment more affordable, but it is also believed that group treatment provides a great resource for conducting exposures, obtaining feedback, and learning vicariously. One study found CBGT to be superior to waitlist, and three studies found CBGT to be superior to a group attention control condition, in which patients received education about social anxiety and nondirective group therapy. Heimberg and colleagues also compared CBGT with phenelzine (Nardil) and placebo. Although there were no differences between CBGT and phenelzine immediately after treatment, there were more instances of relapse for patients receiving phenelzine during the 6-month follow-up. Another study also compared CBGT with the same treatment components administered individually and found the two methods to be comparable. However, more recent studies have suggested that individual treatment is superior to group treatment. One of the largest randomized trials found that group CBT including social skills training yielded similar outcomes to fluoxetine and the combination of fluoxetine plus CBT, and all these treatments were superior to placebo.

The most promising results have come from a version of CBT developed by David M. Clark and colleagues. Several randomized studies have been published, showing large effects compared to fluoxetine, exposure alone, psychodynamic therapy, and group CBT (although not CBGT). Recent data support the efficacy of this treatment in a fee-for-service clinic. Some differences between the newer program and other CBT for social anxiety include a stronger emphasis on identifying and eliminating safety behaviors and shifting attention outward from the self. In addition, exposures are integrated carefully with cognitive challenging in the context of making specific predictions and evaluating these predictions post hoc with feedback from multiple sources, including videotape and those with whom the patient interacted during the exposure. Much of this is consistent with modern theories of CBT discussed above. One of the latest advances in this treatment is an online internet administration of much of the program, using videos, chats, and other interactive activities. Results seem to suggest excellent outcomes with less time spent by the therapist (most of which is online), though considerable patient investment in time on their own. Indeed, multiple studies of internet-based CBT have been conducted and reviewed in meta-analyses, suggesting good efficacy.

Generalized Anxiety Disorder (GAD)

The key feature of GAD is the presence of excessive worry accompanied by anxiety and physiologic symptoms of tension, irritability, and sleep disturbance. Pathologic worry, as occurs in GAD, is distinguished from normal worry in that the frequency, intensity, and duration of the worry, in addition to the attendant anxiety, greatly exceed the actual probability or impact of the feared event and that the worry is experienced as difficult to control. GAD is the least focused of anxiety disorders and has the greatest overlap with depressive disorders. GAD is one of the only psychiatric disorders with a specific cognitive criterion A variable in the DSM-5: excessive, uncontrollable worry. Individuals with GAD tend to worry about minor matters (a sometimes-distinguishing feature of GAD) as well as major ones (health, children's future, etc.). In addition to worry, individuals with GAD tend to have physiologic responses that are less autonomic and more related to stress (e.g., sleep disturbance, muscle tension, irritability).

CBT packages for GAD have typically included training in relaxation and cognitive therapy. In addition, some packages have included some form of exposure therapy to the worry content and problem-solving. Relaxation alone, cognitive therapy alone, and their combination (with and without the further addition of exposure) have all been found to be effective in reducing worry and anxiety compared to waitlist. In addition, CBT has been shown to be more effective than supportive control conditions. One meta-analysis found that CBT and relaxation did not yield different results for GAD. Several meta-analyses and reviews have found CBT superior to waitlist and placebo conditions. One meta-analysis found that CBT effects were somewhat smaller than those for medications; another more recent review found CBT to be superior. The general results of these trials have been positive, but there is a consensus that improvements in treatment efficacy are necessary, perhaps more than in CBT for other anxiety disorders. Several attempts have been made to improve treatment outcomes for GAD. One study added more treatment sessions for individuals predicted to have poorer outcome, with nonsignificant improvement in outcome compared to standard treatment. Others attempted to modify or add to existing models of CBT. Refinement of the CBT model includes a greater focus on meta-worry (worry about worry) or a greater focus on the intolerance of uncertainty and the separation of worry into worry about current problems (addressed by problem-solving) and worry about future unknown events (addressed by increasing skills in tolerating uncertainty, including imaginal exposure). Additional areas and techniques to integrate with current CBT models include interpersonal treatment strategies, emotion regulation strategies, mindfulness, and experiential avoidance strategies, additional focus on achieving positive well-being, and schema-focused treatments. Most data to date have not shown that these techniques provide improved outcomes compared to standard CBT packages, though one study focusing on emotion regulation found larger effect sizes. As with the other disorders, internet-based treatments for GAD are also showing efficacy and great promise.

Future Directions

The current literature review demonstrates the efficacy of CBT programs in the treatment of anxiety disorders. These treatments appear to be effective in children, adolescents, adults, and geriatric populations and are typically more effective than treatment as usual in the short and long term. Exposure in vivo to feared but otherwise safe stimuli and imaginal exposure to unwanted and upsetting thoughts are essential components of most CBT programs for anxiety. The cessation of safety behaviors or other subtle avoidance behaviors is essential for the anxiety disorders in order to disconfirm the patients' erroneous perceptions. The addition of interoceptive exposure appears to make a significant contribution to the treatment of panic disorder. In addition, exposure is typically conducted in the presence of cognitive challenges or discussions, helping the patient to attend to the corrective information. Anxiety management training and cognitive therapy have also been shown to be effective in the treatment of GAD. Cognitive therapy that includes behavioral experiments, which inherently involves in vivo exposure, is also an effective treatment for panic disorder and social anxiety disorder.

Several researchers have demonstrated the efficacy of multicomponent CBT packages that include components of anxiety management, cognitive therapy, and exposure. However, few of the relevant dismantling studies showed significantly improved outcome for these combined treatments as compared to the individual components. This may, to some extent, reflect low statistical power to detect the additive effects of treatments that individually are effective. Consistent with this hypothesis, meta-analyses have sometimes detected evidence for the superiority of combined treatments (typically of cognitive therapy plus exposure) that were not as evident in individual studies. Overall, it seems that the integration of exposure therapy with discussions that increase patients' attention to the disconfirmation of their expectations of negative outcomes to exposures to the feared stimuli is the best method for optimizing treatment outcome. Perhaps the addition of cognitive therapy to exposure enhances outcome because it serves to enhance attention to disconfirming information. It is important to note that, as was hypothesized by emotional processing theory, exposure should match the fear of the patient. It follows that exposures that maximize opportunities for correction of erroneous associations by matching the patient's fears carefully and providing corrective information via these experiences are likely to lead to the best outcomes. The need for relaxation or similar coping or management strategies seems questionable.

Regarding the execution of these specific techniques, it is important to note that some studies suggested that outcomes can differ across therapists using the same techniques, and some studies showed that expert sites are more effective in delivering the same techniques than nonexpert sites (although this literature also includes data to suggest that some nonexpert sites can obtain similar effects for some treatments). When therapist or site differences occur, it is due to two aspects of treatment delivery that likely interact. First, execution of these techniques typically requires experience and competence in formulating and articulating the treatment rationale (i.e., creating a flexible hierarchy that targets core fears, determining the proper focus of imaginal exposures, etc.), addressing the many overt and covert avoidance behaviors effectively, and knowing how to push exposures to their optimal level. In addition, although formation of the therapeutic alliance, warmth, empathy, and other nonspecific therapeutic skills are not the central focus of CBT, they are still important components that likely interact with the focused treatment techniques, facilitating compliance and thus better outcome. However, having nonspecific characteristics without the expertise in technique is insufficient for provision of optimal treatment. Therefore, when experts are not available, it is essential to monitor patient progress carefully (also true when experts are available), and if there is no progress after 10 sessions of treatment, referral to an expert out of one's locale may be warranted. Alternatively, there is increasing data to suggest that therapist-supported internet-based programs can be effective treatments, which yield promising results that often are similar to face-to-face treatments.

Despite the success of CBT in the treatment of anxiety disorders, it is important to acknowledge two limitations. First, not every patient is responsive to treatment, and many people judged to have responded to treatment continue to experience significant residual symptoms. Thus, one avenue for future research is to identify ways to enhance existing treatments or to develop new treatments that are more effective. As noted, however, attempts to augment one psychological component with another treatment have not generally been very successful in significantly enhancing treatment outcome. Some of the recent improvements in outcome appear to be due to techniques that are developed based on specific maintenance mechanisms identified in studies on experimental psychopathology. Hopefully, continued study and improvement in understanding the psychological mechanisms involved in each of the anxiety disorders will help to improve outcome.

Another approach to potentially improving outcome is to combine psychotherapy with standard types of pharmacotherapy. This section did not provide coverage of the many studies that have investigated the efficacy of combining medication and CBT. The data tend to show that combining medication for anxiety disorders, such as SSRIs, with CBT either provides a small advantage or does not enhance outcome beyond those achieved by CBT alone, whereas adding CBT to medication enhances medication outcome. Most studies of combined treatments, whether combining two forms of CBT or combining CBT with medication, implemented both treatments simultaneously. This strategy may actually serve to minimize the ability to detect the effects of combined treatments when the individual treatments are generally effective. An alternative strategy would be to use an augmentation strategy, in which partial responders to one active treatment are then randomized to continue with the original treatment alone or to add a second treatment, as was done in the panic study described above.

The second limitation of CBT is its relative inaccessibility to people who do not live in large cities or near a university-based medical school or graduate program in clinical psychology. Greater research is needed to determine the factors that limit the use of CBT among community-based clinicians and the development of models for the dissemination of these treatments to make them more widely available. Some advancements in this area include using self-help and stepped-care approaches to treatment and using internet-based interventions.

FURTHER READINGS

Antony MM, Stein MB. *Oxford Handbook of Anxiety and Related Disorders*. Oxford University Press; 2008.

Bandelow B, Reitt M, Röver C, Michaelis S, Görlich Y, Wedekind D. Efficacy of treatments for anxiety disorders: a meta-analysis. *Int Clin Psychopharmacol*. 2015;30:183–192.

*Bandura A. *Principles of Behavior Modification*. Holt, Rinehart & Winston; 1969.

Barlow DH. *Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic*. 2nd ed. Guilford Press; 2002.

Barlow DH. *Clinical Handbook of Psychological Disorders*. 6th ed. Guilford Press; 2021.

*Beck AT. *Cognitive Therapy and the Emotional Disorders*. International University Press; 1976.

Carl E, Witcraft SM, Kauffman BY, et al. Psychological and pharmacological treatments for generalized anxiety disorder (GAD): a meta-analysis of randomized controlled trials. *Cog Behav Ther*. 2020;49(1):1–21.

Carpenter JK, Andrews LA, Witcraft SM, Powers MB, Smits JAJ, Hofmann SG. Cognitive behavioral therapy for anxiety and related disorders: a meta-analysis of randomized placebo-controlled trials. *Depress Anxiety*. 2018;35(6):502–514.

Chambless DL, Baker MJ, Baucom DH, Beutler LE, Calhoun KS. Update on empirically validated therapies II. *Clin Psychol*. 1998;51:3.

Clark DA, Beck AT. *Cognitive Therapy of Anxiety Disorders: Science and Practice*. Guilford Press; 2011.

Clark DM, Ehlers A, McManus F, et al. Cognitive therapy versus fluoxetine in generalized social phobia: a randomized placebo-controlled trial. *J Consult Clin Psychol*. 2003;71:1058–1067.

Craske MG, Treanor M, Conway CC, Zbozinek T, Vervliet B. Maximizing exposure therapy: an inhibitory learning approach. *Behav Res Ther*. 2014;58:10–23.

Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CF. Adding psychotherapy to antidepressant medication in depression and anxiety disorders: a meta-analysis. *World Psychiatry*. 2014;13(1):56–67.

Cuijpers P, Sijbrandij M, Koole S, Huibers M, Berking M, Andersson G. Psychological treatment of generalized anxiety disorder: a meta-analysis. *Clin Psychol Rev*. 2014;34(2):130–140.

Davidson JRT, Foa EB, Huppert JD, et al. Fluoxetine, comprehensive cognitive behavioral therapy, and placebo in generalized social phobia. *Arch Gen Psychiatry*. 2004;61:1005–1013.

Eilert N, Enrique A, Wogan R, Mooney O, Timulak L, Richards D. The effectiveness of Internet-delivered treatment for generalized anxiety disorder: an updated systematic review and meta-analysis. *Depress Anxiety*. 2021;38(2):196–219.

Foa EB, Huppert JD, Cahill SP. Update on emotional processing theory. In: Rothbaum BO, ed. *The Nature and Treatment of Pathological Anxiety*. Guilford Press; 2006:3.

*Foa EB, Kozak MJ. Emotional processing of fear: exposure to corrective information. *Psychol Bull*. 1986;99:20–35.

Foa EB, Kozak MJ. Treatment of anxiety disorders: implications for psychopathology. In: Tuma AH, Maser JD, eds. *Anxiety and the Anxiety Disorders.* Lawrence Erlbaum

Associates, Inc.; 1985:421–452.
Guo S, Deng W, Wang H, et al. The efficacy of internet-based cognitive behavioural therapy for social anxiety disorder: a systematic review and meta-analysis. *Clin Psychol Psychother*. 2020;28:656–668.
Heimberg RG, Becker RE. *Cognitive–Behavioral Group Therapy for Social Phobia: Basic Mechanisms and Clinical Strategies*. Guilford Press; 2002.
Heimberg RG, Turk CL, Mennin DS. *Generalized Anxiety Disorder: Advances in Research and Practice*. Guilford Press; 2004.
Lang PJ, Lazovik AD. Experimental desensitization of a phobia. *J Abnorm Soc Psychol*. 1963;66:519–525.
Mayo-Wilson E, Dias S, Mavranezouli I, et al. Psychological and pharmacological interventions for social anxiety disorder in adults: a systematic review and network meta-analysis. *Lancet Psychia*. 2014;1(5):368–376.
McGlynn FD, Mealiea WL, Landau DL. The current status of systematic desensitization. *Clin Psychol Rev*. 1981;1:149.
Meichenbaum D. *Cognitive Behavior Modification*. General Learning Press; 1974.
Mitte K. A meta-analysis of the efficacy of psycho- and pharmacotherapy in panic disorder with and without agoraphobia. *J Affect Disord*. 2005;88:27–45.
Podină IR, Koster EHW, Philippot P, Dethier V, David DO. Optimal attentional focus during exposure in specific phobia: a meta-analysis. *Clin Psychol Rev*. 2013;33(8):1172–1183.
Pompoli A, Furukawa TA, Efthimiou O, Imai H, Tajika A, Salanti G. Dismantling cognitive-behaviour therapy for panic disorder: a systematic review and component network meta-analysis. *Psychol Med*. 2018;48(12):1945–1953.
Sheppes G, Suri G, Gross JJ. Emotion regulation and psychopathology. *Ann Rev Clin Psychol*. 2015;11:379-405.
Stech EP, Lim J, Upton EL, Newby JM. Internet-delivered cognitive behavioral therapy for panic disorder with or without agoraphobia: a systematic review and meta-analysis. *Cogn Behav Ther*. 2020;49(4):270–293.
Van Dis EAM, van Veen SC, Hagenaars MA, et al. Long-term outcomes of cognitive behavioral therapy for anxiety-related disorders: a systematic review and meta-analysis. *JAMA Psychiatry*. 2020;77(3):265–273.
Watts SE, Turnell A, Kladnitski N, Newby JM, Andrews G. Treatment-as-usual (TAU) is anything but usual: a meta-analysis of CBT versus TAU for anxiety and depression. *J Affect Disord*. 2015;175:152–167.
Wolitzky-Taylor KB, Horowitz JD, Powers MB, Telch MJ. Psychological approaches in the treatment of specific phobias: a meta-analysis. *Clin Psychol Rev*. 2008;28:1021–1037.
*Wolpe J. *Psychotherapy by Reciprocal Inhibition*. Stanford University Press; 1958.

14.7 Somatic Therapies for Anxiety Disorders

Lakshmi N. Ravindran, M.D., F.R.C.P.C., and Murray B. Stein, M.D., M.P.H.

Anxiety disorders are associated with a high-burden of illness with significant personal and societal consequences. However, not all patients with anxiety symptoms require somatic treatments. In a significant proportion of anxious patients, symptoms are transient and subsyndromal. There have also been concerns that the “medicalization” of certain symptoms, such as normal worries and shyness, may lead to unnecessary pharmacologic interventions. However, studies confirm that “overdiagnosis” is relatively low and that anxiety disorders are frequently under-recognized and untreated, particularly in primary care settings. This is attributed to factors including lack of awareness of symptom significance and stigma associated with receiving psychiatric help. Unfortunately, a proportion of patients that do seek help do not always receive optimal care. This is due to a myriad of reasons, including: the tendency of anxious patients to present with physical rather than psychological symptoms, their reluctance to report emotional symptoms, and clinicians’ failure to inquire about psychological issues. Thus, unsurprisingly, detection rates of anxiety and depression in primary care are estimated to be less than half, and among them, only a quarter receive appropriate evidence-based care. However, detection rates increase substantially with increased symptom severity and repeated clinic attendance. The diagnosis of anxiety disorders, and provision of appropriate care, are enhanced through systematic screening using questionnaires such as the Generalized Anxiety Disorder Scale (GAD-7), which has been validated for use in primary care settings. Education of clinicians, in particular general practitioners, about subgroups vulnerable to anxiety disorders, including those with chronic physical illness and adverse social determinants, would enhance early recognition and intervention. If detected and correctly diagnosed, most patients with anxiety disorders can be effectively managed in primary care.

Table 14.7–1.
Pharmacotherapy of Anxiety Disorders: Key Principles

- Diagnostic assessment
  - Determine primary diagnosis
  - Assess medical or psychiatric comorbidity
  - Document contributions of psychosocial stressors
  - Determine baseline symptom severity/frequency
- Physical examination and laboratory investigation
- Psychoeducation
- Choice of pharmacologic agent
  - Based on evidence
  - Family history
  - Past response
  - Drug interactions
  - Cost and access
- Monitor treatment program with validated tools

In many patients, the lack of confidence in the efficacy of medication leads to poor treatment adherence, specifically in those with somatic symptoms of anxiety. Psychoeducation for patients on the side effects of psychotropic treatment is also particularly relevant, as patients suffering from anxiety disorders are more prone to experience adverse events.

PRINCIPLES OF MANAGEMENT

As with all psychological illnesses, the management of anxiety disorders should follow recognized principles (see Table 14.7–1). This includes determining the primary diagnosis using standard diagnostic criteria as well as the presence of comorbid psychiatric and medical conditions, including the use of medications. Assessing the contribution of life stressors and familial-genetic factors, a screening physical examination and laboratory tests contribute to a comprehensive evaluation.

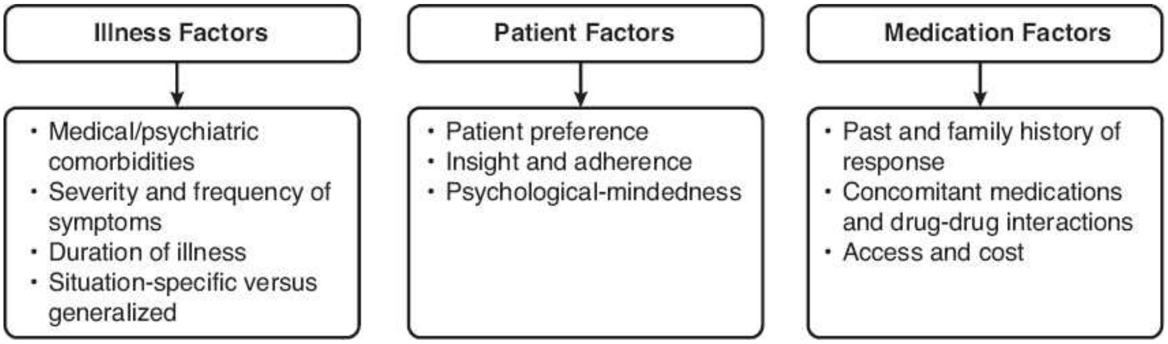
Illness Factors
Medical/psychiatric comorbidities
Severity and frequency of symptoms
Duration of illness
Situation-specific versus generalized
Patient Factors
Patient preference
Insight and adherence
Psychological-mindedness
Medication Factors
Past and family history of response
Concomitant medications and drug-drug interactions
Access and cost

FIGURE 14.7–1. Factors influencing choice of treatment.

After assessment, the recommended initial interventions for anxiety disorders include pharmacotherapy, empirically supported psychotherapy (such as cognitive-behavioral therapy [CBT]), or a combination of the two. These interventions are supplemented by psychoeducation and systematic monitoring of progress. The specific choice of interventions is informed by several factors (see Fig. 14.7–1) based on a holistic understanding of personal and family history, in conjunction with knowledge of the evidence based as well as factors such as access and cost. Medications and psychological therapies are reported to have similar benefits in the acute phase, but their comparative efficacy over the long term has not been well evaluated. While it would be intuitive to assume that the combination would be superior to either intervention, the evidence remains inconclusive due to lack of long-term data.

PHARMACOLOGIC OPTIONS FOR TREATING ANXIETY DISORDERS

Several classes of agents have been evaluated in the treatment of anxiety and related disorders. These include antidepressants, benzodiazepines (BZDs), anticonvulsants, and antipsychotics, among others. There are significant differences in efficacy and tolerability between classes as well as among agents within the same class. There is also significant variability in the degree of symptom alleviation and the time points of optimal benefit in the illness trajectory.

Antidepressants

Selective serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine reuptake inhibitors (SNRIs) are considered first-line agents for the treatment of anxiety disorders. Several novel agents, as well as tricyclic antidepressants (TCAs), monoamine oxidase inhibitors (MAOIs), and reversible inhibitors of monoamine oxidase type A (RIMAs), are used as second-line agents for specific subgroups of patients.

Selective Serotonin Reuptake Inhibitors (SSRIs) and Serotonin-Norepinephrine Reuptake Inhibitors (SNRIs). SSRIs are effective in treating a broad spectrum of anxiety disorders and are often considered the first medication treatment option. SSRIs include fluoxetine, sertraline, paroxetine, fluvoxamine, citalopram, and escitalopram. There are several SNRIs available, including venlafaxine, desvenlafaxine, duloxetine, and levomilnacipran. Most guidelines recommend SSRIs or venlafaxine as the initial antidepressant treatment for anxiety disorders.

Treatment is usually initiated at low doses to minimize an initial anxiogenic effect (which is almost always short-lived) and a slow up-titration to an optimal dose range. Although an initial response may be seen as early as 2 to 4 weeks, a trial of up to 8 to 12 weeks at a therapeutic dose is necessary to assess efficacy. In some patients, time to response and remission may be prolonged even up to 6 months after initial dosing. The aim of treatment should always be remission, as patients exhibiting a partial response and residual symptoms continue to have functional impairment and carry a greater likelihood of relapse. Even after achievement of remission, relapse or recurrence is not uncommon when medications are discontinued. As such, the use of SSRIs and SNRIs long-term for patients with anxiety disorders is not uncommon. The practice of gradual tapering before discontinuation and the adjunctive use of CBT may mitigate the likelihood of relapse.

The significant advantages of SSRIs and SNRIs over the older TCAs and MAOIs are their tolerability and low likelihood of abuse and overdose. Although there are individual differences between SSRI/SNRI agents, they are generally well tolerated. Potential troublesome adverse effects include initial nervousness, insomnia, gastrointestinal issues, agitation, tremors, headache, and dizziness, many of which improve with continued use. If these side effects persist for more than a few weeks, it may be necessary to switch to a different agent or reduce the dose and introduce augmenting agents. Although rare, it is also worth noting the more dangerous and potentially life-threatening side effects of these agents, including seizures and induction of mania. Treatment-emergent sexual dysfunction across various domains (desire, arousal, erectile, ejaculatory, and orgasmic dysfunction) are also commonly reported side effects of SSRIs, though their frequency varies to some extent within the SSRI class (slightly less with SNRIs) and to a significant extent between antidepressant classes (see Table 14.7–2). Since there is usually little or no likelihood of tolerance development for SSRI-induced sexual dysfunction, a switch to another class of agents, dose reduction, or the adjunctive use of sildenafil or similar agents may be needed. Agents with less propensity for sexual side effects include agomelatine (not available in the United States), bupropion, mirtazapine, or moclobemide (not available in the United States). Long-term use of SSRIs has been associated with several serious but uncommon adverse effects, including serotonin syndrome and neuroleptic malignant syndrome. Hyponatremia, agranulocytosis, osteoporosis, and fractures in older adults have been reported in a small subgroup of patients. Increased risk of gastrointestinal bleeding, specifically when prescribed with NSAIDs, is another well-documented serious adverse event.

Table 14.7–2.
Frequency of Sexual Dysfunction among Most Commonly Prescribed Agents for Anxiety Disorders

Class	Drug	Frequency of Sexual Dysfunction
SSRI	Desvenlafaxine	<10%
	Citalopram	10–30%
	Escitalopram	10–30%
	Fluoxetine	>30%
	Fluvoxamine	>30%
	Paroxetine	>30%
	Sertraline	>30%
SNRI	Reboxetine	<10%
	Duloxetine	10–30%
	Levomilnacipram	10–30%
	Venlafaxine	10–30%
RIMA	Moclobemide	<10%
Other Psychotropics	Agomelatine	<10%
	Buspirone	<10%
	Mirtazapine	<10%
Atypical Antipsychotics	Quetiapine	10–30%
Anticonvulsant	Pregabalin	10–30%

With a similar adverse effect profile to SSRIs, SNRIs are generally well tolerated; among them, venlafaxine extended-release is the most prescribed globally. Its common adverse effects include dry mouth, increased sweating, and constipation. Other notable side effects include headache, insomnia, diarrhea, and sexual dysfunction; the severity of most side effects increases with higher doses. At the upper end of the therapeutic range (e.g., venlafaxine 225 mg or higher), routine blood pressure monitoring is also warranted because some patients (fewer than

5%) experience dose-related hypertension. Venlafaxine at lower doses (i.e., <150 mg) is said to have mainly serotonergic effects. Thus, it should be prescribed at higher doses to obtain the dual reuptake inhibition. A newer SNRI, duloxetine, has a similar efficacy and tolerability profile to venlafaxine, including discontinuation syndrome, and is not recommended for individuals with liver dysfunction. More recently, desvenlafaxine, the major metabolite of venlafaxine, was released as a standalone compound. It has primarily been investigated as a treatment for major depression, although it is often used off-label for other conditions, including anxiety disorders (GAD, SAD, and PD in particular). It is thought to offer the advantage of dual uptake inhibition even at lower doses, unlike venlafaxine, where it occurs only in the upper dose range. To date, reports of tolerability relative to its parent compound are quite favorable.

Levomilnacipran is an antidepressant recently introduced into the North American market. Similar to other SNRIs, levomilnacipran is safe to use and generally well tolerated among patients who may report mild side effects such as nausea, sweating, and agitation following use. There is also a need to monitor an increase in blood pressure following initiation of use. The advantage of levomilnacipran is that the degree of norepinephrine reuptake inhibition is said to be proportionally the highest (compared to the serotonin effect) among all SNRIs. Published evidence of levomilnacipran's use in pure anxiety conditions is lacking, but the literature suggests individuals with anxiety conditions comorbid with depression are known to benefit.

Both with SSRIs and SNRIs, abrupt stopping can lead to discontinuation syndrome. Commonly reported symptoms of this condition, which tend to occur 2 to 4 days after medication cessation, include increased anxiety, irritability, tearfulness, dizziness, lightheadedness, malaise, sleep disturbance, and concentration difficulties. Although discontinuation symptoms can occur with all SSRIs and SNRIs, they are more common with agents with shorter half-lives (e.g., paroxetine and venlafaxine). Patients should discontinue SSRIs or SNRIs only under the supervision of their physician, with gradual taper; if discontinuation symptoms emerge while tapering, the dose should be increased to stop symptoms, followed by a restart of the discontinuation at a slower rate. An alternate option is the use of fluoxetine as a transition agent because of its long plasma half-life.

Ongoing controversy concerning antidepressant use and increased suicidality in the initial few months of treatment prompted the FDA to issue a black box warning on all antidepressant medication, cautioning against their use in children and young adults. This led to more systematic evaluations of the risk of such occurrence, which found no clear evidence to support an increase in risk of suicidality in individuals aged 24 and above over the short-term. Their use may, in fact, confer a degree of protection in the adult and older populations. With child and young adult samples, independent and more recent meta-analyses show a 1.5- to 2-fold risk of increasing suicidal thoughts and behaviors with newer antidepressants, but not in the number of completed suicides and with only a very small increase of absolute risk. Given the low-absolute risk and the significant impact of anxiety disorders in children and adolescents (the known age of onset), the use of pharmacotherapy in these populations is warranted, particularly in the more severely affected.

Tricyclic and Heterocyclic Antidepressants (TCAs and HCAs). TCAs were the earliest proven agents in the treatment of anxiety disorders until SSRIs were invented. TCAs and related HCAs used for treatment of anxiety disorders include desipramine, nortriptyline, clomipramine, amitriptyline, and doxepin. TCAs and HCAs work by inhibiting the reuptake of multiple monoamines, including norepinephrine and serotonin, in the synaptic cleft. They also block cholinergic, histaminic, and alpha-1 adrenergic receptors, which contributes to their significant side-effect profile. These include sedation, weight gain, hypotension, dry mouth, blurred vision, and constipation, among others. Subgroups of patients often report an initial “jitteriness” when treatment is initiated with TCAs, which can often be mitigated with low doses and a gradual up-titration to a therapeutic level. TCAs are suggested to be less efficacious than the SSRIs/SNRIs for anxiety disorders and also carry a greater risk of serious adverse effects, including a lowered seizure threshold and cardiotoxicity, specifically from overdose. In addition, drug–drug interactions and discontinuation syndrome are not insignificant issues with TCAs.

Monoamine Oxidase Inhibitors (MAOIs). MAOIs are effective antidepressants whose use has been superseded by equally effective but significantly safer SSRIs and SNRIs, relegating them to second-line use. The two most commonly used MAOIs in North America are phenelzine and tranylcypromine. The benefit of phenelzine is well documented in panic disorder and social anxiety disorder. MAOIs irreversibly inactivate both isoforms of the enzyme monoamine oxidase (MAO-A and MAO-B), leading to a relative increase in monoamines serotonin, norepinephrine, and dopamine in the synaptic cleft. The irreversibility of the inhibition requires patients to have a minimum 2-week "washout" period when switching from an MAOI to any other antidepressant class or even from one MAOI to another. The enzymatic inhibition is generic for monoamines and reduces the ability of the body to break down tyramine, leading to the risk of a hypertensive crisis after consumption of food rich in this monoamine. Though this reaction is rare, patients on MAOIs need to follow a low-tyramine diet during treatment and for an additional 2 weeks after discontinuation. The more common and more benign side effects of MAOIs include postural hypotension, insomnia, and weight gain. Similar to TCAs, patients taking MAOIs should be monitored closely for drug–drug interactions and the possibility of overdose.

Reversible Inhibitor of Monoamine Oxidase Type A (RIMAs). RIMAs are unavailable in the United States, although one member of this class, moclobemide, is available in Canada and several other countries globally. Randomized clinical trials (RCTs) have shown that moclobemide is effective in reducing symptoms of depression and several anxiety disorders (SAD in particular); however, anecdotal reports suggest that it may be less efficacious than MAOIs. Clinicians have also noted the gradual development of tolerance in a proportion of patients, leading to the need for an incremental increase in dose. Close monitoring is advised when used in patients with hypertension. The key advantage of RIMAs over MAOIs is that dietary restrictions can be less stringent but are still recommended when RIMAs are used in higher doses. More common and more benign adverse effects include insomnia, dizziness, agitation, and dry mouth; most are transient and self-limiting. RIMAs, when available, represent an alternative to SSRIs or SNRIs for patients with SAD who have experienced significant side effects with those agents.

Novel Antidepressants. Mirtazapine is a noradrenergic and specific serotonergic antidepressant (NaSSA). It has not been extensively studied in anxiety disorders but is claimed to be a useful agent for treating symptoms of panic disorder and GAD, as well as anxiety symptoms in older adult patients. Adverse effects are generally immediate and short-term and include sedation, dry mouth, and dizziness. Mirtazapine is relatively free of sexual side effects but has a propensity to increase appetite and cause weight gain. Additionally, its favorable effect on sleep often makes it a useful adjunctive agent for patients with insomnia.

Buproprion is a dual norepinephrine dopamine reuptake inhibitor that is not often used as a first-line agent for the treatment of anxiety disorders. This is in part due to a clinical impression that it may trigger agitation and worsen anxiety, at least in initial therapy. It is more often used as an adjunctive agent to address residual symptoms and improve sexual dysfunction.

Vortioxetine is a newer agent with a multimodal mechanism of action on the serotonergic system. It acts as an agonist and partial agonist at 5-HT_{1A} and 5-HT_{1B} receptors, respectively, and as an antagonist at 5-HT_3, 5-HT_7, and 5-HT_{1B} receptors. To date, the limited clinical trials of its use in anxiety have primarily focused on GAD, where results remain preliminary. However, its neurochemical profile and the reported reduction of anxiety symptoms in patients with MDD would suggest benefit.

Agomelatine is a novel agent, not available in the United States. Its action is said to occur through agonism of melatonin receptors MT1 and MT2 and antagonism of 5-HT_{2C} receptors, leading to increased monoamine neurotransmission. The evidence for its benefit in treating the symptoms of anxiety disorders is still preliminary, but there is promising evidence for its use in GAD and its prophylaxis. It has the advantage of a low likelihood of sexual dysfunction and weight gain and does not cause discontinuation symptoms. Adverse effects, mostly gastrointestinal symptoms and fatigue, are mild and self-limiting. Regular monitoring of liver function is advised in the early period of treatment.

Vilazodone, recently introduced in North America, works by combining selective serotonin reuptake inhibition with partial agonism of the 5-HT_{1A} receptor. Approved for use in depression, a recent trial found that a flexible dose (20 to 40 mg/day) of vilazodone incurred a statistically significant reduction of symptoms in patients with GAD when compared to placebo. However, more studies are needed to provide a more robust understanding of its use in anxiety disorders as well as its long-term effects. Finally, milnacipran is an SNRI available in several countries worldwide but not currently approved in North America. Two small open trials have suggested potential utility in GAD and PD, but controlled trials are needed to confirm benefit.

Benzodiazepines (BZDs)

BZDs have been used for several decades for the treatment of anxiety disorders. Their therapeutic effect is mediated through augmentation of GABAergic neurotransmission, specifically in “fear circuits” mediated by the amygdala. In addition to anxiolytic effects, they also possess sedative, anticonvulsant, and muscle relaxant properties. Most BZDs (e.g., clonazepam) are metabolized by the liver, primarily through CYP-mediated oxidation, but others (e.g., lorazepam, oxazepam, and temazepam) are metabolized via a conjugation process, and thus are the preferred choice for individuals with impaired liver function. Common adverse effects of BZDs include mild dizziness and sedation, both of which are generally attenuated by time or dose reduction. Patients should be cautioned about using heavy or dangerous machinery and driving, especially at initiation and after a dose increase. Consumption of alcohol should be avoided as it can intensify the effects of BZDs. Although potential for abuse and development of tolerance to therapeutic effects are of some concern, when used appropriately under medical supervision, BZDs are efficacious (even when used chronically), have a rapid onset of action, and are generally well tolerated. They are particularly useful for as-needed (PRN) use for symptomatic relief of those with occasional panic attacks or performance anxiety. However, PRN use of BZDs is generally discouraged in individuals with frequent panic attacks, where regular use of medication to prevent panic (rather than try to abort an attack once it has started) is recommended.

BZDs are classified based on their half-life as short-, intermediate-, and long-acting. Agents with a short or intermediate half-life (e.g., alprazolam or lorazepam) are the most commonly prescribed agents for as-needed use but are also prescribed for regular use, although long-acting agents (e.g., clonazepam) are often preferred for regular use. BZDs are often prescribed for a brief period (3 to 4 weeks) when antidepressant treatment is initiated to improve delayed or partial response or to help relieve adverse effects from the antidepressant, such as insomnia or jitteriness. At other times, BZD monotherapy can be justified when other treatments have failed or are not tolerated and when anxiety is persistent, severe, and disabling. BZDs are also often used as an adjunct to antidepressant therapy in instances where the antidepressant provides some, but not optimal, anxiety reduction.

Some patients have been noted to remain stable on BZDs for many years with continued therapeutic response and no evidence of tolerance or abuse. However, BZD use is best avoided in individuals with histories of alcohol or substance abuse unless there are compelling clinical reasons (i.e., nonresponse to other agents); under such circumstances, care should be under the supervision of a psychiatrist.

Anticonvulsants

Antiepileptic drugs have long been used in the treatment of mood disorders, but their use in anxiety disorders is still mostly anecdotal. Controlled studies support the benefit of pregabalin in these conditions, particularly in GAD, panic disorder, and social anxiety disorder. Pregabalin, a structural analog of GABA, has evidence from several good-quality RCTs supporting its use as a monotherapy as well as an adjunctive agent therapy in acute treatment and prophylaxis of anxiety disorders. It may be of specific benefit when there are comorbid pain conditions and when insomnia is prominent. Dizziness and drowsiness are common side effects, and weight gain occurs in a minority of patients. It can be used safely in patients with hepatic dysfunction but is best avoided in patients with renal impairment. Other anticonvulsant medications, often used as adjunctive therapies to antidepressants, can be considered on a case-by-case basis in the treatment of patients with refractory anxiety disorders. These medications include gabapentin, levetiracetam, topiramate, tiagabine, and lamotrigine; however, evidence of efficacy from RCTs for most of these agents is lacking or inconclusive.

Antipsychotics

The use of typical antipsychotics for the treatment of anxiety disorders is limited by their significant adverse effects, including risk of tardive dyskinesia (TD) and extrapyramidal symptoms (EPS). The “atypical” antipsychotics (AAPs) include olanzapine, risperidone, quetiapine, ziprasidone, aripiprazole, and lurasidone, among others. AAPs are distinguished from the older generation of antipsychotics in part by their therapeutic effect being mostly mediated by the serotonergic system and, thus, lower propensity for extrapyramidal adverse events. Several AAPs have been shown to be effective augmentation agents to antidepressants for treatment of depression. Evidence for their benefit in anxiety disorders is less robust, with the possible exception of quetiapine, which has been shown to be an effective agent for the treatment of GAD. Use of AAPs is limited by their propensity to cause weight gain and metabolic syndrome.

Other Antianxiety Medications

Beta-adrenergic antagonists, also called beta-blockers, are primarily useful as needed for performance anxiety. They block the peripheral effects of adrenaline and noradrenaline and thus reduce symptoms such as tachycardia and tremors, which patients with performance anxiety often tend to experience. Propranolol and atenolol are the commonly prescribed agents with some variability in the dosing. The dose range for propranolol is usually between 20 and 40 mg, and that for atenolol is between 25 and 50 mg. It has been suggested that targeting a heart rate reduction by 10 beats/minute and evaluating the patient's subjective feedback on performance is a useful strategy to determine the optimum dose. While published literature indicates a marked reduction of somatic anxiety symptoms, clinical impression suggests a less robust benefit. Beta-blockers used short-term as needed are generally well tolerated, but occasional side effects include dizziness, vertigo, or sleep disturbance. Their use in patients with certain cardiac or pulmonary diseases, angle-closure glaucoma, or diabetes is usually contraindicated.

Buspirone, an azapirone, is effective in the treatment of chronic anxiety states, primarily GAD. Buspirone is usually prescribed in two or three divided doses at a dose range of 20 to 60 mg. The time course of response to buspirone is similar to that of antidepressants, and it may take time to reach full therapeutic benefit. Hydroxyzine is an antihistamine with preliminary evidence of benefit in the treatment of GAD. While initial symptom relief can be experienced within the first 15 to 20 minutes of administration, it can take up to several weeks to achieve an optimal therapeutic effect. Hydroxyzine can be used for both short-term and long-term anxiety symptoms but may lead to mild side effects, including dry mouth, sedation, and tremors. It may be a good alternative to BZDs for some patients, although tolerance to the anxiolytic effects of hydroxyzine may occur rather quickly.

Treatment Resistant Anxiety Disorders

Partial or nonresponse to pharmacotherapy is not uncommon with anxiety disorders. In such cases, confirmation of the diagnosis (including ruling out potentially contributing medical conditions, such as thyroid disorders), evaluating the impact of potential comorbidities (e.g., substance use), and pharmacokinetic interactions, as well as validating adequate trials (i.e., adequate dose and duration) are prudent first steps. In many patients, treatment resistance occurs even when pharmacotherapy is used in combination with CBT. Literature on treatment-resistant anxiety remains sparse, with few well-conducted clinical trials. Augmentation agents commonly used include anticonvulsants, AAPs, and BDZ. A common strategy in practice for addressing treatment-resistant GAD is the augmentation of SSRIs with pregabalin or BZDs. Despite lack of controlled trials and mixed evidence, in clinical practice, gabapentin is often used as an adjunct for refractory anxiety disorders (GAD and PD), particularly in cases of comorbid alcohol use disorder. The literature on the benefit of AAPs is mixed. Several small case series/open trials of olanzapine and aripiprazole indicate their utility of use as augmentation while data with quetiapine augmentation is mixed. With all AAPs, the risk of adverse metabolic effects needs evaluation and long-term monitoring. For treatment-resistant PD, there is evidence for adding clonazepam to an SSRI if a patient is not amenable to or suitable for CBT.

Novel Therapeutics for Anxiety

Despite the relative wealth of pharmacologic options for the treatment of anxiety disorders, clinicians continue to struggle with a subset of patients who are refractory to conventional treatments. There is an increasing interest in several novel strategies, including the use of neurostimulation therapy as well as several unique pharmacologic compounds, complementary and alternative (CAM) therapies, and NMDA receptor agents for this population.

Repetitive transcranial magnetic stimulation (rTMS) has shown promise as a treatment for depression or depression comorbid with anxiety disorder; however, evidence for rTMS applied for pure anxiety disorders is sparse. Case reports have suggested utility for rTMS in SAD and PD, but a double-blind trial failed to show benefit over placebo for rTMS treatment for PD. More promisingly, a small trial of fMRI-guided rTMS for GAD showed positive results following acute treatment, with gains generally maintained at 6-month follow-up; however, more controlled studies are needed to confirm these findings.

Transcranial direct current stimulation (tDCS), which involves application of a low-intensity direct electrical current to specific brain areas to modulate cortical activity, has preliminarily demonstrated benefit in treatment-resistant depression. There are reports of its efficacy in GAD and PD, but it needs more systematic evaluation. Cranial electrotherapy stimulation, another novel, noninvasive technique using low-intensity alternating current, has shown early signals of benefit but needs further study.

There is emerging evidence regarding the use of neurosteroids for the treatment of anxiety disorders. Due to their ability to rapidly impact the central nervous system, there is some indication that their use can relieve anxiety symptoms. In particular, three good-quality trials have shown a nasally administered neurosteroid (compound PH94B) to be superior to placebo in treating symptoms of social anxiety disorder; this drug is still under investigation.

Ketamine and its analogs are seen as promising new interventions for treatment-resistant depression. Significant relief of anxiety symptoms in these patients suggests its benefit for depression comorbid with anxiety disorders and potentially even refractory anxiety conditions. While high-quality data for the latter is lacking, smaller studies have shown intravenous ketamine to produce a marked improvement in anxiety symptoms in patients with SAD and GAD. Results with intranasal esketamine are pending.

With the recent legalization of cannabis in many countries worldwide, as well as its increasing use for multiple medical conditions, use of cannabis for psychiatric conditions has become commonplace. However, to date, systematic trials of cannabinoids for anxiety disorders are few. The only meta-analysis to date on the topic suggests some utility for cannabinoids in SAD and GAD. However, methodologic issues, including heterogeneity in study design, preparation differences, duration of treatment, absence of studies of long-term efficacy, and publication bias, mitigate making any clear conclusions.

Consumers, particularly those with mood and anxiety disorders, are increasingly turning to CAM to either self-treat or augment their more conventional medications. These therapies include herbal medications, acupuncture, biofeedback, and yoga, among others. However, many herbal preparations have the potential to cause drug–drug interactions, so it is imperative that clinicians routinely check for the use of these substances. Two meta-analyses noted that kava, one of the most common herbs taken for anxiety, does demonstrate benefit for mild to moderate anxiety, but its utility in specific anxiety disorders is still unclear, and there is also possible risk of hepatotoxicity and documented interactions with medications metabolized through the cytochrome P450 enzyme system. Other herbal remedies with purported antianxiety properties but insufficient clinical data to support their use include valerian root, passiflora (passion flower), inositol, St. John wort, gotu kola (*Centella asiatica*), California poppy, hawthorn, Ginkgo biloba, lemon balm, and sour date seed. Aerobic exercise has some data for utility in anxiety conditions and, given its myriad health benefits, should be routinely recommended. Preliminary evidence suggests benefits of yoga for performance and some other forms of anxiety.

TREATMENT RESPONSE AND COURSE OF ILLNESS ACROSS ANXIETY DISORDERS

In general, all first-line agents are thought to be equally effective, but in practice, SSRIs are often chosen as first-line options since they are often better tolerated than SNRIs. Starting with lower doses and increasing them gradually, as well as adding short-term concomitant BZDs, may improve their efficacy and tolerability. Partial or nonresponse to the initial trial of a first-line antidepressant may necessitate a switch to another first- or second-line agent from a different class. Augmentation with BZDs or pregabalin also has evidence for benefit. Guidelines recommend combining pharmacotherapy with psychological treatment as a first-line option for patients showing partial or nonresponse to medications.

Although individually, either pharmacotherapy or psychotherapy (e.g., CBT) show efficacy in the treatment of anxiety disorders, their combination has not been extensively studied. Available data shows inconsistent benefits for the combination; this may be because one treatment modality alone is sufficiently good that showing an additive (or better) effect is difficult without large sample sizes or because superiority of the combination to individual therapies in acute treatment fails to be sustained long-term. Access to CBT is often limited or expensive in many settings. If available, it is sometimes recommended that pharmacotherapy and CBT be started simultaneously. However, for many patients with partial or nonresponse to pharmacotherapy, sequential provision of CBT is a pragmatic option.

Empirical data on the optimal duration of pharmacotherapy are not extensive for most anxiety disorders. For many patients, at least 8 to 12 weeks of medications at therapeutic doses may be needed to evaluate the efficacy of acute treatment. Systematic data on long-term maintenance is limited for anxiety disorders. However, several RCTs do show that 20% to 50% of patients with chronic anxiety disorders relapse within several months of antidepressant discontinuation. It is thus recommended that such patients be maintained on treatment for at least 1 to 2 years before considering dose reduction or discontinuation and that any discontinuation be done by gradual decrements of 10% to 25% every 1 to 2 months while observing for recurrence.

The degree of response to pharmacotherapy differs among the anxiety disorders. Panic disorder is among the most responsive, with efficacy seen in 60% to 70% of treated subjects. However, for patients with comorbid agoraphobia, the response rates may be lower, and some patients may require higher doses of medication or concomitant CBT. Response rates for effective pharmacotherapy in GAD and SAD are slightly lower, with 50% to 60% of patients having a good clinical outcome.

Table 14.7–3.
Evidence-Based Recommendations for Monotherapy of Anxiety Disorders

Classes and Agents	**Anxiety Disorders**		
	SAD	**GAD**	**PD**
SSRIs	First-line	First-line	First-line
SNRIs	First-line	First-line	First-line
TCAs	Not recommended	Second-line	Second-line
MAOIs	Second-line	Insufficient evidence	Second-line
RIMA	Second-line	Insufficient evidence	Insufficient evidence
Buspirone (serotonergic agent)	Second-line	Second-line	Second-line
Novel Antidepressants			
Agomelatine	Needs further evaluation	Preliminary evidence	Needs further evaluation
Buproprion	Needs further evaluation	Needs further evaluation	Needs further evaluation
Mirtazapine	Needs further evaluation	Preliminary evidence	Preliminary evidence
Vortioxetine	Needs further evaluation	Preliminary evidence	Needs further evaluation
Vilazodone	Needs further evaluation	Preliminary evidence	Needs further evaluation
Milnacipran	Needs further evaluation	Preliminary evidence	Preliminary evidence
Benzodiazepines	Second-line	Second-line	Second-line
Quetiapine (AAPs)	Not recommended	Second-line	Insufficient evidence
Pregabalin (anticonvulsant)	First-line	Second-line	Insufficient evidence

SSRI, selective serotonin reuptake inhibitor; SNRI, serotonin norepinephrine reuptake inhibitor; TCA, tricyclic antidepressants; MAOIs, monoamine oxidase inhibitors; RIMA, reversible inhibitors of monoamine oxidase A; AAPs, atypical antipsychotics.

Note: SSRIs may be useful as second line for specific phobia nonresponsive to psychotherapy, and beta-blockers are recommended only for performance anxiety.

MANAGEMENT OF INDIVIDUAL ANXIETY CONDITIONS

The following sections review the literature for the treatment of individual anxiety conditions and make medication recommendations based on the strength of evidence available for different classes of drugs. See Table 14.7–3 for a summary of recommendations and Table 14.7–4 for general pointers.

Generalized Anxiety Disorder

GAD, which is characterized by excessive and frequent worries in multiple domains accompanied by psychic and somatic symptoms of anxiety, often presents with physical symptoms (e.g., fatigue, muscle tension, sleep disturbance) and usually in primary care settings. Comorbidities include depression, other anxiety disorders, or substance use are common. GAD often follows a chronic, fluctuating course, but periods of remission may occur.

Large RCTs, as well as systematic reviews and meta-analyses, confirm the benefit of SSRIs (fluoxetine, citalopram, escitalopram, paroxetine, sertraline) and SNRIs (venlafaxine and duloxetine) as first-line choices for the treatment of GAD. Novel agents agomelatine, mirtazapine, pregabalin, buspirone, quetiapine, and vortioxetine are suggested as alternative options. Two double-blind trials and one prevention trial of agomelatine for GAD indicated that agomelatine treatment improved clinical response and remission of anxiety symptoms and reduced the rate of relapse. The TCA imipramine has also been shown to be beneficial, but it is often used as a second-line option due to adverse effects. SSRIs are usually the initial choice, followed by pregabalin. Another highly prescribed class of agents for GAD are BZDs, which include alprazolam, lorazepam, diazepam, and bromazepam. Evidence for their benefit is strong, but they are often used only short term due to risk of tolerance and dependence. While the AAP quetiapine has significant efficacy in GAD, guidelines designate it as a second-line intervention due to adverse metabolic effects and low acceptability and tolerability. Buspirone is often used similarly due to lack of comparative data with antidepressants. Other AAPs (olanzapine, risperidone, and aripiprazole) and bupropion, which have lower evidence of efficacy, may also be useful as adjunctive therapies. Among the newer antidepressants, vilazodone has shown promise for the acute treatment of GAD, but the two published meta-analyses reported contradictory findings, and there are no comparator data. Similarly, studies of another newer antidepressant, vortioxetine, yielded inconsistent results. Beta-blocker propranolol, anticonvulsant tiagabine, and AAP ziprasidone have not been shown to be effective in GAD.

Table 14.7–4.
Optimizing Pharmacotherapy

- Remission as goal of treatment
- Use first-line agent
- Adjunctive benzodiazepines short-term for selected subgroups
- Avoid quick dose escalation and high-dose use
- Patient education regarding adverse effects
- Trial of 8–12 weeks of pharmacotherapy at optimum dose
- Maintenance treatment recommended for a year

For relapse prevention, there is good evidence to support the use of SSRIs and SNRIs as maintenance therapy. A similar benefit has also been demonstrated with pregabalin and quetiapine. The benefit of continued maintenance medications may last at least 6 months, and cognitive behavior therapy (CBT) may be similarly effective.

Panic Disorder

Panic disorder is characterized by the presence of recurring, unexpected panic attacks. While periods between panic attacks are relatively symptom-free, persistent concern about future attacks and avoidance behavior is common. Meta-analyses and RCTs have confirmed the benefit of several classes of antidepressants for panic disorder. SSRIs (citalopram, fluoxetine, paroxetine, sertraline, fluoxetine, and escitalopram) and the SNRI venlafaxine are often recommended as initial therapy (though SSRIs tend to be the preferred choice). TCAs (in particular, clomipramine and imipramine), as well as MAOIs (phenelzine and tranylcypromine), have equal efficacy but are often used as second- or third-line treatment because of their adverse event profiles. Other second-line agents include mirtazapine, bupropion, and reboxetine (not currently available in North America). Several BZDs (alprazolam, lorazepam, diazepam, and clonazepam) have also been proven effective in reducing panic symptoms, particularly in short-term treatment. Although classified as second-line interventions due to the risk of dependence and withdrawal, they are useful as concomitant short-term treatment to reduce or prevent the panic attacks and reduce anticipatory anxiety. BZDs also have the added benefit of a rapid onset of action. In contrast, AAPs have only preliminary evidence of benefit and are restricted to adjunctive third-line options for cases of partial or nonresponse to first- or second-line agents. Studies with the beta-blocker propranolol and anticonvulsants tiagabine, gabapentin, and carbamazepine have reported negative findings, while there are only preliminary reports evaluating pregabalin.

Maintenance studies in panic disorder have reported that many treatment responders show further improvement over time with antidepressant therapy effective in preventing relapse, with data up to 3 years. In addition to most first-line agents, which are effective in maintenance therapy, there is evidence that BZDs (alprazolam and clonazepam) may provide effective treatment for maintenance without the development of tolerance. In general, maintenance treatment for at least a year is recommended for most patients. Once a decision is made to stop medication, gradual tapering and slow discontinuation are recommended, particularly with BZDs.

Agoraphobia

Agoraphobia is **characterized by anxiety or fear of being in a situation from which escape or obtaining aid would be difficult.** Most presentations of agoraphobia are accompanied by panic attacks or PD, and agoraphobic symptoms improve in parallel with panic symptoms upon appropriate pharmacotherapy for PD. In other cases, agoraphobic symptoms often overlap with those of social anxiety disorder, GAD, and avoidant personality disorder. The efficacy of pharmacotherapy in “pure” agoraphobia has not been well evaluated; early studies with BZDs (alprazolam and diazepam), the TCA imipramine, and the MAOI phenelzine had negative findings, but little research has been done since.

Published literature confirms that pharmacologic treatment of common comorbidities, including depression and other anxiety disorders, often results in reduction of agoraphobia. Psychotherapeutic interventions, in particular, CBT with a focus on exposure therapy, appear to be effective interventions. It has also been suggested that patients with “pure” agoraphobia are often housebound and avoid seeking help, leading to an underestimation of its true prevalence and limited research on its treatment.

Social Anxiety Disorder

With an estimated lifetime prevalence of 8% to 12%, social anxiety disorder is a common condition. Its key symptom is excessive discomfort in social situations for fear of embarrassment or negative evaluation, which often leads to avoidant behavior. Evidence from meta-analyses, systematic reviews, and large RCTS suggest that all SSRIs, SNRIs (venlafaxine and duloxetine), pregabalin, and clonazepam, have strong evidence for benefit and are recommended as first-line treatments for SAD. The MAOI phenelzine is also very efficacious for SAD, but because of a poor adverse effect profile, it is used as a second-line agent for refractory illness. Other second-line options include the novel antidepressant mirtazapine and RIMA moclobemide, as well as BZDs bromazepam and alprazolam, and gabapentin. Due to negative findings, TCAs, the novel agent buspirone, and the AAP quetiapine are not recommended. While beta-blockers propranolol and atenolol can be beneficial in performance anxiety, they do not appear to help in the generalized subtype of SAD. There is also very limited positive evidence regarding the use of ketamine as a treatment for SAD.

While routine use of high doses of antidepressants is not recommended for SAD, it may benefit a subgroup to attain and maintain remission. However, lack of remission, even with optimum treatment, is not uncommon. Unfortunately, there are few systematic studies of strategies to treat refractory SAD. Switching to an alternative first-line agent or phenelzine, augmentation with pregabalin or clonazepam, or combination with CBT are among the common strategies employed to treat refractory SAD.

Longer-term studies have noted that following acute treatment, a significant proportion of patients continue to show incremental benefit, increasing the overall rate of response. Relapse prevention RCTs confirms the benefit of several agents, including SSRIs, pregabalin, and clonazepam, over a 6-month period. As with acute treatment, psychological interventions, including CBT and mindfulness-based stress reduction, have been documented as efficacious in reducing SAD symptoms.

Specific Phobia

The key feature of specific phobia is an intense fear or anxiety of a specific object or situation, leading to its avoidance. Examples of common specific phobias include fear of airplanes, insects, and heights. Subsyndromal presentations are common, and lifetime prevalence is estimated at 10% to 13%, with many affected reporting multiple fears. Clinical presentations are rarely for the phobia itself but more often for comorbid depression and other anxiety disorders.

Psychological interventions, particularly exposure-based therapies, are considered to have the most evidence for efficacy in treating specific phobias. The benefit of pharmacotherapy has not been well studied, but very limited evidence suggests that SSRIs (escitalopram and paroxetine) may be useful as second-line treatments if patients do not respond to psychotherapy. So far, BZDs have not been found to be beneficial even as adjuncts to exposure therapy and may even reduce the efficacy of exposure-based psychological treatments. Thus, in general, there is little evidence to support the benefit of pharmacotherapy in the management of specific phobias. Of note, the presence of comorbid depression and other anxiety disorders, which are common, would lead to a recommendation of treatment with appropriate first-line pharmacologic agents.

Separation Anxiety Disorder

Separation anxiety disorder is characterized by an extreme fear or anxiety of being away from attachment figures, often accompanied by worry of harm befalling them. This often leads to avoidance of being away from home or loved ones. Historically recognized primarily as a disorder of childhood and adolescence, DSM-5 classified SpAD as an anxiety disorder. Its lifetime prevalence rates are estimated to be between 4% and 8%, and it is reported to be more common in those with a family history of agoraphobia and panic disorder. Comorbidity with a wide variety of psychiatric conditions is common.

Pharmacotherapy of SpAD has been primarily derived from research in children and adolescents, as SpAD has been historically under-recognized in adults. In studies looking at mixed populations of youth with GAD, SAD, and SpAD, SSRIs have shown good evidence of benefit. However, trials for TCAs (imipramine) and BDZ (clonazepam) reported negative findings. Most recently, a small RCT did not find vilazodone to be superior to placebo for Adult SpAD after 12 weeks of treatment.

Selective Mutism

Selective mutism is a disorder typically seen in childhood whereby individuals who are physically able to speak and understand speech fail to do so in selective situations (e.g., school). Children with selective mutism will readily speak in other more comfortable situations (e.g., home). Selective mutism is relatively rare, with prevalence rates reported to be between 0.02% and 1.9%; it is increasingly recognized as a forerunner of social anxiety disorder.

While psychosocial interventions are preferentially used in treatment of selective mutism, case reports and open trial studies suggest that SSRIs (e.g., fluoxetine and fluvoxamine) may be helpful, especially in more severe cases. The black box warnings issued by the FDA regarding increased risks of suicidality with SSRI use in children and adolescents means these medications are now generally reserved for young people who do not respond to psychological interventions. Nevertheless, the findings of meta-analyses that the increase in absolute risk of suicide is very small, and recognition of the serious functional morbidity associated with chronic anxiety and avoidance in children and adolescents, would suggest that the use of pharmacotherapy with appropriate monitoring should not be unduly withheld in these situations.

TREATMENT OF ANXIETY DISORDERS IN CHILDREN AND ADOLESCENTS

Anxiety disorders are the most prevalent mental health conditions in young people and, if left untreated, can result in long-term psychological impairment, poor treatment outcomes, and increased risk of psychiatric comorbidities later in life. Several RCTs have provided conclusive evidence that SSRIs and SNRIs are useful in the treatment of certain anxiety disorders in children and adolescents, particularly GAD, SAD, and separation anxiety disorder in youth. Preliminary data also suggests that the combination of medication and CBT is better than either intervention alone for separation anxiety in children. Similarly, while psychosocial interventions are preferentially used in treatment of selective mutism, early case reports and pilot studies suggest that SSRIs (particularly fluoxetine and fluvoxamine) may be helpful, especially in more severe cases. The black box warnings issued by the FDA regarding increased risks of suicidality with SSRI use in children and adolescents means these medications are now generally reserved for young people who do not respond to psychological interventions. However, meta-analyses suggest that the increase in absolute risk of suicide is very small. Together with observed recognition of the serious functional morbidity associated with chronic anxiety and avoidance in children and adolescents, it is suggested that use of pharmacotherapy should be considered whenever it might be beneficial and safe.

SUMMARY

Pharmacotherapy is a key treatment option for anxiety disorders. Among them, SSRIs and SNRIs remain the first-line option, with novel agents, pregabalin, and other psychotropic medications relegated to second-line use. When considering psychopharmacologic options for treatment of anxiety disorders, clinicians are encouraged to take a holistic approach that considers a patient’s personal and family history, published literature on the subject, as well as the accessibility and cost of the intervention. Mild to moderate side effects are not uncommon following initiation of medical intervention, and thus it is recommended that physicians provide adequate psychoeducation (about side effects and overall setting expectations about the anticipated time course of response, etc.), start patients on a low dose and up-titrate to the optimal therapeutic dose. Similarly, discontinuation of treatment must be done gradually with vigilant monitoring for signs of withdrawal or relapse. Given the increasing burden of disease due to anxiety disorders, it is of utmost importance that practitioners, especially those in primary care settings, inquire about psychological symptoms and that patients receive appropriate psychoeducation to increase help-seeking behavior and treatment adherence. Whereas physicians will often be in the role of medication prescribers for anxiety disorders, they can and should also play an important role in making patients aware of and referring them to evidence-based psychological therapies, such as CBT.

A large majority of pharmacologic clinical trials to date have focused on pure anxiety disorders, often excluding common comorbidities such as major depression. This has been helpful in ascertaining the effects of medications on individual anxiety disorders but does not reflect real-world anxiety populations. It would be helpful for future research to systematically include common comorbid populations (e.g., SAD and PD) and pharmacologic combinations in treatment-resistant anxiety disorders. As noted above, there is also a need to further develop alternative biologic therapies for treatment-refractory patients, particularly as we continue to further understand the underlying pathophysiology of these illnesses as well as to take into account patient preference for more complementary types of treatments.

FURTHER READINGS

Baldwin DS, Anderson IM, Nutt DJ, et al. Evidence-based pharmacological treatment of anxiety disorders, post-traumatic stress disorder and obsessive-compulsive disorder: a revision of the 2005 guidelines from the British Association for Psychopharmacology. *J Psychopharmacol*. 2014;28:403–439.

Cirillo P, Gold AK, Nardi AE, et al. Transcranial magnetic stimulation in anxiety and trauma-related disorders: a systematic review and meta-analysis. *Brain Behav.* 2019;9:e01284.

**Craske MG, Stein MB, Eley TC, et al. Anxiety disorders. *Nat Rev Dis Primers*. 2017;3:17024.

**Cuijpers P, Sijbrandij M, Koole SL, et al. Adding psychotherapy to antidepressant medication in depression and anxiety disorders: a meta-analysis. *World Psychiatry*. 2014;13:56–67.

**Garakani A, Murrough JW, Freire RC, et al. Pharmacotherapy of anxiety disorders: current and emerging treatment options. *Focus*. 2021;19:222–242.

Glue P, Neehoff SM, Medicott NJ, Gray A, Kibby G, McNaughton N. Safety and efficacy of maintenance ketamine treatment in patients with treatment-refractory generalized anxiety and social anxiety disorders. *J Psychopharmacol.* 2018;32:663–667.

Hua A, Major N. Selective mutism. *Curr Opin Pediatr*. 2016;28:114–120.

Jakubovski E, Johnson JA, Nasir M, Muller-vahl K, Bloch MH. Systematic review and meta-analysis: dose-response curve of SSRIs and SNRIs in anxiety disorders. *Depress Anxiety*. 2019;36:198–212.

**Katzman MA, Bleau P, Blier P, et al. Canadian clinical practice guidelines for the management of anxiety, posttraumatic stress and obsessive-compulsive disorders. *BMC Psychiatry*. 2014;14(Suppl 1):S1.

Meuret AE, Tunnell N, Roque A. Anxiety disorders and medical comorbidity: treatment implications. *Adv Exp Med Biol*. 2020;1191:237–261.

Montejo AL, Montejo L, Navarro-Cremades F. Sexual side-effects of antidepressant and antipsychotic drugs. *Curr Opin Psychiatry*. 2015;28:418–423.

Purohit MP, Zafonte RD, Sherman LM, et al. Neuropsychiatric symptoms and expenditure on complementary and alternative medicine. *J Clin Psychiatry*. 2015;76:e870–e876.

Rabheru K, Richter PM, Robichaud M, Walker JR. Canadian clinical practice guidelines for the management of anxiety, posttraumatic stress and obsessive-compulsive disorders. *BMC Psychiatry*. 2014;14(Suppl 1):S1.

Schneier FR, Moskow DM, Choo T-H, Galfalvy H, Campeas R, Sanchez-lackey A. A randomized controlled pilot trial of vilazodone for adult separation anxiety disorder. *Depress Anxiety*. 2017;34(12):1085–1095.

**Strawn JR, Welge JA, Wehry AM, Keeshin B, Rynn MA. Efficacy and tolerability of antidepressants in pediatric anxiety disorders: a systematic review and meta-analysis. *Depress Anxiety*. 2015;32:149–157.

**Van Leeuwen E, van Driel ML, Horowitz MA, et al. Approaches for discontinuation versus continuation of long-term antidepressant use for depressive and anxiety disorders in adults. *Cochrane Database Syst Rev*. 2021;4(4):CD013495.

15

Obsessive-Compulsive and Related Disorders

Dan J. Stein, M.D., Ph.D., and Christine Lochner, Ph.D.

Obsessive-compulsive and related disorders (OCRDs) constitute a new chapter in both the DSM-5 and the ICD-11. In DSM-5, and its text revision (DSM-5-TR) this grouping of disorders includes obsessive-compulsive disorder (OCD), body dysmorphic disorder (BDD), hoarding disorder (HD), trichotillomania (TTM) (hair-pulling disorder), and excoriation (skin-picking) disorder, while ICD-11 additionally includes olfactory reference disorder, hypochondriasis, and Tourette syndrome. Both nomenclatures include the OCRDs in a separate chapter, given the evidence that these share some phenomenologic and psychobiological features, and to improve the recognition and management of these underdiagnosed and undertreated conditions. Here, each of the DSM-5-TR OCRDs will be covered in turn.

OBSESSIVE-COMPULSIVE DISORDER

Introduction

OCD is the paradigmatic OCRD and the one on which most research has been done. Like several other OCRDs, it is characterized by repetitive, intrusive thoughts and repeated mental acts or behaviors in response to those thoughts. OCD was once central to psychodynamic thinking, but it is now an exemplar par excellence of a neuropsychiatric disorder with a specific neurocircuitry proposed. Furthermore, whereas OCD was once considered rare, it is now known to be quite prevalent—patients are often embarrassed by and ashamed of their symptoms, and careful screening is needed to ensure case-finding. Finally, whereas OCD was once considered very refractory to treatment, the introduction of specific pharmacotherapeutic and psychotherapeutic interventions has led to more favorable outcomes for many.

Definition

The presence of obsessions or compulsions characterizes OCD. Obsessions are intrusive and unwanted repetitive thoughts, urges, or impulses that often lead to a marked increase in anxiety or distress. Compulsions are repetitive behaviors or mental acts in response to obsessions or performed in a rigid rule-bound way. Some patients may have only obsessions or only compulsions, but both obsessions and compulsions are present in most cases.

History

Although there are earlier descriptions of OCD symptoms in the medical literature, the English term derives from Westphal's description of *zwangsvorstellung* in the late 19th century, translated as "obsession" or "compulsion." Janet subsequently made an important contribution to the literature, describing the role of "incompleteness" in OCD patients, as well as symptoms such as tics. Freud's theories on the role of the unconscious in obsessive-compulsive symptoms were particularly influential on clinical practice in the early part of the 20th century, as was his description of a spectrum of symptoms ranging from obsessive-compulsive personality traits through obsessive-compulsive neurosis and on to obsessive-compulsive psychosis.

In the middle of the 20th century, case reports indicated that the serotonergic tricyclic antidepressant, clomipramine, was useful in OCD. Randomized controlled trials confirmed the efficacy of clomipramine and gave impetus to a serotonin hypothesis of OCD. The introduction of behavior therapy and selective serotonin reuptake inhibitors (SSRIs) to treat OCD were additional significant steps forward. The finding that unique patterns of activation accompanied OCD on brain imaging and that both of these interventions helped normalize the altered functional neuroanatomy of OCD gave significant impetus to further research on this condition. Over time, other interventions, including dopaminergic and glutamatergic agents and neurosurgical interventions, were also developed for treatment-refractory OCD.

The finding that clomipramine was more effective than desipramine, a noradrenergic tricyclic, in OCD and a range of other conditions characterized by intrusive thoughts or repetitive behaviors helped consolidate the notion of an obsessive-compulsive spectrum of disorders. SSRIs proved to be helpful for several of these conditions, but not all. Overlapping, but somewhat different behavioral approaches, have also proven helpful for these conditions. At the start of the 21st century, cutting-edge research methods, including neurogenetics, have been used to help delineate the relationship between OCRDs, particularly OCD and tic disorders.

Comparative Nosology

DSM-III provided operational criteria for OCD, defining obsessions and compulsions along the lines of the definition above. DSM-IV introduced the notion that compulsions can be mental acts. DSM-5 made relatively subtle adjustments to the DSM-IV criteria to further clarify the symptoms, for example, replacing the term “impulse” with “urge” (to avoid confusion with impulse-control disorders) and noting that obsessions cause anxiety in most individuals (given that this is not the case in all individuals). However, in DSM-5, OCD was shifted into a new chapter on OCRDs, a more substantial change. Another significant change from DSM-IV was the extension of the insight specifier to include absent insight; this emphasizes to clinicians that some patients with OCD (around 4%) may have delusional thoughts, yet they should be diagnosed with OCD rather than with a psychotic disorder.

Table 15–1.
Obsessive-Compulsive and Related Disorders

DSM-5-TR		ICD-11	
Disorder	**Diagnostic Criteria**	**Disorder**	**Diagnostic Guidelines**
Obsessive-compulsive disorder	Intrusive and unwanted recurrent and persistent thoughts, urges, or impulses (obsessions) or repetitive behaviors or mental acts that the individual feels driven to perform in response to an obsession or according to rules that must be applied rigidly (compulsions).	Olfactory reference disorder	Persistent preoccupation with the belief that one is emitting a perceived foul or offensive body odor or breath. The individual engages in repetitive and excessive behaviors in response to the odor concerns.
Body dysmorphic disorder	Preoccupation with one or more perceived defects or flaws in physical appearance that are not observable or appear slight to others. The	Hypochondriasis (health anxiety disorder)	Persistent preoccupation or fear about having serious, progressive, or life-threatening illness. Preoccupation is accompanied

	individual has performed repetitive behaviors or mental acts in response to the appearance concerns.		by repetitive and excessive health-related behaviors or maladaptive avoidance behavior related to health.
Hoarding disorder	Persistent difficulty discarding or parting with possessions, regardless of their actual value, accumulates possessions that congest and clutter living areas compromising their intended use.	Tourette Syndrome	Presence of chronic motor tics and vocal tics, beginning during childhood. Tics are rapid, nonrhythmic, and recurrent movements or vocalizations.
Trichotillomania (hair-pulling disorder)	Recurrent pulling of hair, resulting in hair loss, with repeated unsuccessful attempts to reduce or stop pulling.		
Excoriation (skin-picking) disorder	Recurrent picking of skin, resulting in		

lesions, with repeated unsuccessful attempts to reduce or stop pulling.

The ICD-10 description of OCD does not include the notion that obsessions give rise to compulsions. There is certainly support in the literature for the idea that compulsions may precede obsessions, and indeed some have even suggested that the disorder be termed compulsive-obsessive disorder. At the same time, from a treatment planning perspective, it may be useful for clinicians to understand the cycle of obsessions and compulsions, and ICD-11 emphasizes that in many patients with OCD, obsessions lead to anxiety or distress, and compulsions are undertaken to reduce these negative emotions. ICD-11 also has a new chapter on OCRDs and emphasizes that some patients with OCD have absent insight.

The DSM-5-TR chapter on OCRDs includes OCD, BDD, HD, TTM (hair-pulling disorder), and excoriation (skin-picking) disorder. ICD-11 lists both TTM and excoriation disorder as body-focused repetitive behavioral disorders (BFRBDs) and includes olfactory reference disorder, hypochondriasis, and Tourette syndrome (Table 15–1). ICD-11 emphasizes that whereas patients with BDD are preoccupied with flawed appearance, those with olfactory reference disorder are preoccupied with smelling bad, with both groups displaying repetitive mental acts and behaviors in response to these concerns. DSM-5-TR includes illness anxiety disorder in the somatoform disorders, whereas ICD-11 emphasizes that this condition is characterized by intrusive, repetitive thoughts and repetitive behaviors in response and lists it under both OCRDs and anxiety or fear-related disorders. Both DSM and ICD-11 recognize the relationship between OCD and tics; DSM-5-TR classifies Tourette disorder as a neurodevelopmental disorder and includes a tic-related specifier for OCD, whereas ICD-11 lists Tourette syndrome in three places: neurodevelopmental disorders, OCRDs, and disorders of the nervous system.

Epidemiology

Individuals with OCD are characteristically aware that their compulsions are excessive. In the past, they may have been too embarrassed to seek professional help and unlikely to have received efficacious intervention—factors that may have led to early authors describing OCD as a rare condition. The Epidemiological Catchment Area study found that OCD was, in fact, one of the most prevalent of all psychiatric disorders, and follow-up cross-national surveys confirmed its presence across several different countries. The National Comorbidity Survey Replication (NCS-R) was able to delineate clinical from subclinical obsessive-compulsive symptoms in a rigorous way, and findings were that OCD had a lifetime prevalence of 2.3%, but that subclinical OCD was much more prevalent.

Several additional findings with public health relevance emerge from community and clinical studies. First, although the exact prevalence of OCD differs from study to study, OCD is a relatively common disorder in most countries surveyed, and subclinical OCD is even more prevalent. Second, OCD is accompanied by significant comorbidity, with particularly high rates of OCRDs, mood and anxiety disorders, as well as suicide ideation. Third, OCD is accompanied by significant morbidity, including impairment in social, academic, and occupational functioning and reduced quality of life comparable to that seen in other severe medical and mental disorders. Early Global Burden of Disease data indicated that OCD was one of the most disabling of all medical disorders, with more recent iterations emphasizing the crucial contributions of anxiety disorders (including OCD) as a whole. Fourth, there is significant underdiagnosis and undertreatment of OCD even in high-income countries. Fifth, partly because of this underdiagnosis and undertreatment, OCD is associated with high economic costs, both for the individual and society.

OCD has a median age of onset of 19 in the NCS-R. Approximately one-quarter of cases have onset by age 10, with early onset more common in males. In other cases, OCD may have a later onset; in females, onset is often during adolescence, with some developing OCD during pregnancy or the puerperium. Onset after 30 is unusual. OCD has surprisingly similar features in children and adults, but compulsions are typically more central to the presentation earlier in life, with some children either not experiencing obsessions or being unable to articulate them. Community studies indicate a higher prevalence of OCD in females, but there may be roughly equal numbers of male and female patients in clinical contexts. Males and females have similar symptoms, but males are more likely than females to have comorbid tics. Individuals with OCD have an increased family history of OCD and OCRDs, including tic disorders.

Etiology

OCD is increasingly viewed as a neuropsychiatric disorder mediated by specific neurocircuits. From a cognitive-affective neuroscience perspective, OCD appears to be characterized by deficits in cognitive control, with a shift from goal-directed to habit-driven behavior. Aspects of cognitive dyscontrol may include alterations in inhibitory control (e.g., inability to withhold task-irrelevant responses), task shifting (e.g., increased cognitive inflexibility), and error-monitoring (e.g., altered detection of response errors); such alterations may contribute to repetitive, habitual behavior. Early clinical work established that OCD could be precipitated by damage to striatal circuitry, as seen in the global influenza epidemic in the early 20th century. More recent advances in structural and functional brain imaging have delineated the role of these and related circuits in mediating OCD, and translational work in animal models has allowed further delineation of the neuroanatomy of cognitive dyscontrol and habit-driven behavior. In recent decades, a range of laboratory and clinical work has focused on the neurotransmitters and other molecules found in these circuits and potentially involved in the pathogenesis of OCD.

Cognitive and neuropsychological studies of OCD have found a range of alterations, including impaired executive function, with evidence for cognitive inflexibility and excessive habit formation. Such work is consistent with the hypothesis that OCD is characterized by cognitive dyscontrol, with a shift to habit-driven behavior. A second group of studies has emphasized alterations in affective and reward processing in OCD, with perhaps the most specific findings related to perceptions of disgust. Patients with OCD appear to show increased recognition of disgust compared to other emotions. This finding is consistent with the view that OCD differs fundamentally from anxiety disorders, which are characterized by alterations in fear processing. These cognitive-affective findings, as well as work on temperamental risk for OCD (such as increased behavioral inhibition), are consistent with a view that emphasizes that OCD symptoms, particularly in subclinical form, can be understood through the lens of evolutionary theory; they have an adaptive role in protecting individuals from harmful stimuli, whether by grooming to ensure the removal of harmful organisms or via the avoidance of disgust cues.

Several brain regions, including the anterior cingulate, orbitofrontal cortex, and striatum, have consistently been implicated in OCD, although there is also evidence that a wide range of other regions play a role. One hypothesis has been that cortico–striatal–thalamic–cortical (CSTC) circuitry plays a critical role in mediating the cognitive-affective impairments seen in OCD, with activation or inhibition of different components of this circuitry driving the compulsive and impulsive features of OCRDs. Laboratory research has indicated that these circuits control automated behaviors such as habits and grooming. Clinical research has indicated that there are gray matter volumetric changes in components of this circuitry and alterations in frontostriatal connectivity and a range of white matter tracts in OCD. A growing database of functional imaging studies has found increased activity in these circuits both at rest and exposed to feared stimuli. Early imaging work was seminal in showing that both pharmacotherapy and cognitive-behavioral therapy (CBT) could normalize functional alterations in CSTC circuitry in OCD, and the relevant evidence base has continued to grow.

Magnetic resonance spectroscopy and molecular imaging studies have begun to delineate the specific neurochemical alterations found in CSTC and other circuitry implicated in OCD. Such work has been complemented by work on blood and cerebrospinal fluid (CSF) metabolites and studies of pharmacologic “challenges” in OCD, pharmacotherapeutic dissection, and neurogenetics. This work has indicated that many neurotransmitter systems contribute to OCD, including serotonergic, dopaminergic, glutamatergic, and GABAergic systems. A serotonin hypothesis of OCD has been influential; a range of evidence points to the involvement of this system in OCD, and patients respond to treatment with serotonin reuptake inhibitors (SRIs) but not to treatment with noradrenergic reuptake inhibitors. At the same time, there is relatively little evidence that altered serotonergic function plays a causal role in OCD. Furthermore, a range of evidence points to the involvement of the dopaminergic system in OCD, and, notably, augmentation of serotonin reuptake blockers with dopamine receptor antagonists is considered a first-line intervention for treatment-refractory OCD. There has been recent interest in additional systems, including the glutamatergic system in OCD, with evidence that some glutamatergic agents, such as memantine, may be helpful in this condition.

Genetic findings may increasingly influence future work on the cognitive-affective neuroscience of OCD. Twin and family studies indicate that genetic susceptibility plays a role in OCD, and more so in childhood-onset OCD. To date, OCD has been associated with serotonin-related polymorphisms and variants involved in catecholamine modulation; taken together, meta-analyses of candidate gene studies suggest that OCD may be associated with multiple genes of relatively small effect. Growing data from genome-wide association studies support this hypothesis, and, notably, the most significant findings have included genes that contribute to glutamatergic function. It is also noteworthy that gene variants in OCD may be associated with some specific brain imaging findings. The future hope is that a better understanding of the relevant intermediate phenotypes involved in OCD and their underlying molecular neurobiology will help development of a personalized medicine approach.

Diagnosis and Clinical Features

OCD can readily be diagnosed based on a thorough psychiatric history and examination. While patients often attempt to neutralize obsessions with compulsions, these behaviors or mental acts are not realistic in terms of what they are designed to prevent or are excessive. The DSM-5-TR diagnostic criteria for OCD address the presence of obsessions and compulsions and emphasize that these symptoms are time-consuming or distressing, impairing, are not due to substances, medications, or other medical conditions, and are not due to other psychiatric disorders (Table 15–2). Repetitive thoughts and behaviors can be seen in a range of other mental disorders, as is discussed in more detail below. When assessing the negative impact of OCD symptoms, it is also essential to assess avoidance of triggers of obsessions and compulsions. In addition, key features to ascertain during the history and examination of an individual with OCD include the age of onset and course of the symptoms and any possible precipitants, the type and severity of symptoms, the extent of insight and impairment, and the presence of comorbid disorders and symptoms.

As noted above, OCD may have an early age of onset; while there is often no specific precipitant, OCD onset has been linked to various infections, certain substances, and traumatic or stressful life events. In factor analytic studies of OCD, symptoms fall on several limited dimensions, including contamination/cleansing, forbidden thoughts/checking, symmetry/ordering, and hoarding symptoms. While patients with OCD typically have good insight into the fact that their OCD beliefs are not true, some patients have poor insight, and a small percentage have absent insight—their thoughts are, in fact, delusional. As these patients may be misdiagnosed with a psychotic disorder and treated with inappropriately high doses of antipsychotic medication, DSM-5 and DSM-5-TR extend the insight specifier to include absent insight/delusional OCD.

Table 15–2.
DSM-5-TR Diagnostic Criteria for OCD

A. Presence of obsessions, compulsions, or both:

Obsessions are defined by (1) and (2):

1. Recurrent and persistent thoughts, urges, or images that are experienced, at some point during the disturbance, as intrusive and unwanted, and that in most individuals cause marked anxiety or distress.
2. The individual attempts to ignore or suppress such thoughts, urges, or images, or to neutralize them with some other thought or action (i.e., by performing a compulsion).

Compulsions as defined by (1) and (2):

1. Repetitive behaviors (e.g., handwashing, ordering, checking) or mental acts (e.g., praying, counting, repeating words silently) that the person feels driven to perform in response to an obsession or according to rules that must be applied rigidly.
2. The behaviors or mental acts are aimed at preventing or reducing distress, or preventing some dreaded event or situation; however, these behaviors or mental acts either are not connected in a realistic way with what they are designed to neutralize or prevent, or are clearly excessive.

Note: Young children may not be able to articulate the aims of these behaviors or mental acts.

B. The obsessions or compulsions are time-consuming (e.g., take more than 1 hour per day) or cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

C. The obsessive-compulsive symptoms are not attributable to the direct physiologic effects of a substance (e.g., a drug of abuse, a medication) or another medical condition.

D. The disturbance is not better explained by the symptoms of another mental disorder (e.g., excessive worries, as in generalized anxiety disorder; preoccupation with appearance, as in body dysmorphic disorder; difficulty discarding or parting with possessions, as in hoarding disorder; hair pulling, as in trichotillomania [hair-pulling disorder]; skin picking, as in excoriation [skin-picking] disorder; stereotypies, as in stereotypic movement disorder; ritualized eating behavior, as in an eating disorder; preoccupation with having an illness, as in illness anxiety disorder; sexual urges or fantasies, as in a paraphilic disorder; impulses, as in disruptive behavior disorders;

guilty ruminations, as in major depressive disorder; thought insertion or delusional preoccupations, as in a psychotic disorder; or repetitive patterns of behavior, as in autism spectrum disorder).

***Specify* if:**

With good or fair insight: The individual recognizes that obsessive-compulsive disorder beliefs are definitely or probably not true or that they may not be true.

With poor insight: The individual thinks obsessive-compulsive disorder beliefs are probably true.

With absent insight/delusional: The individual is completely convinced that obsessive-compulsive disorder beliefs are true.

***Specify* if:**

Tic-related: The individual has a current or past history of a tic disorder.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:265, 266.

DSM-5-TR also includes a tic-related specifier for OCD, as many patients with OCD have current or past tics. OCD with tics is more commonly seen in males, has an earlier age of onset, is more likely to be accompanied by specific OCD symptom dimensions (e.g., symmetry/ordering), and may be less likely to respond to SSRIs. Sensory phenomena or premonitory urges can precede both OCD symptoms and tics, and in the case of complex tics, it may be quite challenging to differentiate tics from compulsions. Given their comorbidity and the close familial relationship between OCD and tic disorders, there is a strong argument that Tourette disorder (TD, characterized by motor and vocal tics) is the OCRD par excellence. Nevertheless, DSM-5-TR retains TD as a neurodevelopmental disorder. In contrast, ICD-11 lists Tourette syndrome under neurodevelopmental disorders, OCRDs, and movement disorders.

Individuals with OCD are more likely to suffer from other OCRDs than those without OCD, and they are also more likely to provide a family history of such conditions. However, the most common comorbid disorder in OCD patients presenting to the clinic is depression. From a clinical perspective, it is crucially important to assess for suicidal symptoms, which are quite prevalent in OCD. Comorbid anxiety disorders are also common in OCD. Notably, substance use disorders are more common in community studies than in clinical studies of OCD. Individuals with OCD may also have several personality disorders, including schizotypal and obsessive-compulsive personality disorder (OCPD). There is an argument that OCPD—although not characterized by obsessions and compulsions, but rather by a pattern of excessive perfectionism and need for control—is best conceptualized as an OCD spectrum condition; OCPD may share neuropsychological impairments with OCD and may perhaps respond to treatment with SSRIs.

There is a growing literature on diagnostic interviews for screening for OCD and ensuring accurate differential diagnosis, and symptom measures for assessing the severity of OCD symptoms. DSM-5-TR includes a Self-Rated Level 1 Cross-Cutting Symptom Measure with two items for OCRDs, and this can be followed up by a Level 2 Repetitive Thoughts and Behaviors Scale. The Level 1 items focus on “unpleasant thoughts, urges, or images that repeatedly enter your mind” and about “feeling driven to perform certain behaviors or mental acts over and over again.” The Level 2 questions are adapted from the Florida Obsessive-Compulsive Inventory (FOCI) Severity Scale and assess the time occupied by symptoms, distress caused by the symptoms, difficulty in controlling the symptoms, avoidance caused by the symptoms, and interference associated with the symptoms. The Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) is a clinician-rated instrument that assesses similar items and is commonly used as a primary outcome measure in clinical trials. Adaptations of this scale include a version that provides a score for each major OCD symptom dimension. The Structured Interview for OCRDs provides questions that target DSM criteria for each OCRD and key issues critical in different diagnoses (e.g., are symptoms exacerbated by substance use), allowing a comprehensive assessment of each of these conditions.

Diagnosis and assessment should also cover several other key issues. First, it is crucial to ascertain the patient’s explanatory model of OCD. Many patients are quite comfortable with a biomedical explanation of OCD. However, others may have entirely different models, and it is then imperative to provide psychoeducation and negotiate a joint approach to management. In occasional cases, obsessions and compulsions may themselves interfere with management (e.g., patients may view medications as contaminating). Second, it is essential to ascertain how family members are impacted by or accommodate symptoms. This is particularly important when planning CBT, which demands increased exposure to feared stimuli and refraining from compulsions. Third, it is vital to take a meticulous medical history. There is a growing literature indicating that OCD may be precipitated by Streptococcal infection or a range of other insults, so a comprehensive medical history is warranted.

In children with OCD, the aim of compulsions may not be easily articulated. It is also relevant to differentiate normal developmental rituals from OCD in this age group. In children and adolescents with OCD, it is vital to assess the impact of OCD on developmental outcomes, ask about a family history of OCD, and assess how symptoms impact or are accommodated by the family. Males and females have similar presentations, although males are more likely to have earlier-onset, certain OCD symptoms, and comorbid tics, and females may have onset during pregnancy and the puerperium, and perhaps exacerbation of symptoms during the premenstrual period. OCD symptoms appear largely independent of culture and geography, although the local context may influence the specific content of symptoms.

Pathology and Laboratory Examination

At the current time, there are no biomarkers for OCD. Brain imaging may ultimately be clinically helpful in predicting response to treatment, but presently it remains a research tool. Similarly, while neurogenetic investigations have been valuable in the research context, they do not have sufficient clinical utility for more widespread use, except for obsessive-compulsive symptoms secondary to conditions such as autistic spectrum disorders, where such investigations may be helpful.

Differential Diagnosis

The psychiatric disorder hierarchy criterion in DSM-5-TR lists a large set of psychiatric conditions that should be differentiated from OCD: generalized anxiety disorder, BDD, HD, TTM, excoriation disorder, stereotypic movement disorder, eating disorders, substance-related and addictive disorders, illness anxiety disorder, paraphilic disorders, disruptive, impulsive-control and conduct disorders, major depressive disorder, schizophrenia spectrum and other psychotic disorders, and autism spectrum disorders. Worries in generalized anxiety disorder are about real-life concerns. They tend to be less irrational and egodystonic than OCD. Similarly, the ruminations of depression and the preoccupations of mania are typically mood-congruent and egosyntonic; in neither case are symptoms neutralized by compulsions. In BDD, illness anxiety disorder, TTM, excoriation disorder, stereotypic movement disorder, anxiety disorders, and eating disorders, the focus of preoccupations and repetitive behaviors is distinct from that typically seen in OCD.

Hoarding behaviors can be seen in OCD; for example, patients may hoard newspaper articles to prevent harm should they forget an important matter. However, in most cases of HD, there is no specific obsession. Patients with OCD may experience a “just right” feeling on completion of compulsions, but in paraphilic disorders, substance-related and addictive disorders such as gambling disorder and disruptive, impulse-control, and conduct disorders, repetitive behaviors are typically gratifying. Thus, although terms such as compulsive gambling or compulsive stealing are used, these conditions should not be conflated with OCD. While some patients with OCD have absent insight into their symptoms, the delusional beliefs are related to OCD, and they do not have other features of schizophrenia spectrum disorders. A subgroup of OCD patients suffers from comorbid schizophrenia spectrum disorders, including schizotypal personality disorder. Similarly, while patients with autistic disorder often have repetitive symptoms, they also have a broad array of other characteristic symptoms.

Course and Prognosis

OCD is generally chronic, and symptoms may persist for decades without intervention. Nevertheless, growing data emphasize that long-term outcomes can be positive. Thus, patients should be encouraged to remain hopeful, reminded of the availability of a range of different evidence-based interventions, and strongly encouraged to persist with treatment.

Treatment

Psychoeducation is a crucial component of the treatment of OCD. Exploring the patient’s explanatory model of OCD is a useful first step in key elements of treatment, such as understanding the patient’s suffering, building a collaborative approach to care, and instilling hope. Many patients and families are relieved to learn basic facts about OCD, including its high prevalence, the fact that many others with this condition are embarrassed by or ashamed of symptoms, that it is not due to commonly voiced factors (e.g., bad mothering), and that treatments are efficacious for many. Consumer advocacy organizations, such as the Obsessive-Compulsive Foundation, may play a crucial role in conveying such information.

Pharmacotherapy. Seminal early work reported that patients with OCD responded to clomipramine, a relatively serotonergic tricyclic antidepressant. Subsequent randomized controlled trials showed that clomipramine was efficacious in children, adolescents, and adults with OCD. Furthermore, this agent was more effective than the relatively noradrenergic tricyclic antidepressant, desipramine. Clomipramine became the first registered agent for the treatment of OCD, and a "serotonin hypothesis" of OCD influenced a good deal of subsequent research on both OCD and OCRDs.

The subsequent introduction of the SSRIs was significant in that these provided a set of efficacious and well-tolerated agents. In meta-analyses of SRIs in OCD, clomipramine has the largest effect size, but this may reflect that these were simply the earliest trials. There is equal efficacy in head-to-head comparisons of clomipramine and SSRIs, but SSRIs are better tolerated. Several of the SSRIs are Food and Drug Administration (FDA) or European Medicines Agency (EMA) approved for OCD, including childhood OCD, and these agents are therefore considered first-line for OCD pharmacotherapy.

Critical clinical questions include the optimal dose and duration of SRIs in the pharmacotherapy of OCD. Although there is not a very steep dose–response curve, meta-analysis indicates that higher doses of SRIs are associated with better outcomes in OCD. Therefore, clinical guidelines suggest increasing the dosage of SRIs over time in OCD to relatively high ranges (e.g., escitalopram 30 mg, fluoxetine 80 mg, sertraline 200 mg), with some patients responding to even higher doses. However, exceeding the recommended doses is not advisable for clomipramine or citalopram due to safety concerns. There is also some evidence that response to SRIs is slower in OCD than in depression; clinicians are therefore encouraged to prescribe an SRI for at least 12 weeks before deciding to switch agents or augment with another psychotropic. A general rule of thumb is that the dose required for a response should be continued for at least 1 to 2 years. There is a real risk of relapse when discontinuing medication, so this should be done very gradually, with small adjustments every few months.

Patients who fail to respond to one SRI may respond to another. In patients who fail to respond to an SRI, reasonable options include augmentation with additional medications. A range of medications has also been used to augment SRIs, with the most extensive evidence base comprising studies of low-dose dopamine receptor antagonists (e.g., risperidone) or partial agonists (e.g., aripiprazole) in adult OCD. Some studies have indicated that these are particularly useful in patients with tics, but not all data are consistent. From a clinical perspective, it is notable that when a response to these agents is seen, it typically occurs within a relatively short time (i.e., 4 weeks). Despite the relatively low doses used in this setting, it is important to monitor adverse events. There is some indication that agents with increased D2 and D3 binding (e.g., aripiprazole) are more effective in treatment-refractory OCD, but this hypothesis requires further investigation.

Outcomes have been negative, inconsistent, or modest for other potential SSRI augmentation agents. Perhaps most promising are small studies that have indicated that certain glutamatergic agents, such as memantine, riluzole, ketamine, lamotrigine, and topiramate, may be helpful in this condition. However, these remain proof-of-principle investigations. Sometimes low doses of clomipramine are used to augment SSRIs, but careful electrocardiographic monitoring is required. D-cycloserine is a partial glutamatergic agonist that has been used to augment CBT in OCD and anxiety disorders, but efficacy data in OCD are currently inconsistent. Choice of medication in refractory patients is guided by a range of clinical considerations, including the presence of comorbid disorders. Monotherapy with irreversible monoamine oxidase inhibitors remains a potentially useful option in refractory cases.

Psychotherapy. There is a long history of understanding obsessive-compulsive symptoms from a psychodynamic perspective. However, there is no rigorous evidence that a psychodynamic approach is useful in treating OCD. Instead, empirical research on the psychotherapy of OCD has emerged from behavioral and cognitive perspectives. Exposure and response prevention (ERP) has proven to be a particularly effective component of treatment, with more recent work also focusing on the potential role of cognitive and mindfulness approaches.

To some extent, most clinicians, including primary care physicians involved in diagnosing and treating OCD, can apply the principles of exposure and response prevention. The essential principle is minimizing avoidance of distressing cues and encouraging in vivo exposure to feared situations and imaginal exposure to feared consequences. After such exposures, it is important not to engage in the usual compulsive response. A hierarchy of gradual exposure to more distressing cues is then developed and followed. Self-help volumes and internet programs have been developed to assist with the process. There may, however, be significant resistance to undertaking exposure and responsive prevention, in which case a skilled cognitive-behavioral therapist is needed. As in the case of pharmacotherapy, it is essential to ensure that the dose and duration of exposure and response prevention treatment are sufficient.

More cognitive approaches to the psychotherapy of OCD have also been developed. These focus on thought distortions, for example, those related to inflated responsibility. The therapist and patient collaborate to test such distortions and replace them with more adaptive thinking patterns. Although the evidence base is not as well developed, mindfulness techniques may help supplement these approaches. Family sessions may also be a beneficial component of psychotherapy, particularly in younger patients; families are likely to benefit from a better understanding of OCD and exposure and response prevention principles.

Combined Treatment. Clinical guidelines indicate that either pharmacotherapy or psychotherapy is a reasonable first-line intervention for the management of OCD. In younger patients, it seems reasonable to begin CBT first. It seems reasonable to begin pharmacotherapy in patients with more severe symptoms or comorbid depression. There is some evidence from rigorous comparisons of CBT and pharmacotherapy and their combination that combined behavior therapy and SRIs are particularly useful, including in children and adolescents. Further, one careful study in adults indicated that ERP augmentation of SSRIs was more efficacious than risperidone augmentation. In very refractory cases, partial hospitalization for combined treatment may also be helpful. In everyday clinical practice, judicious sequencing of interventions may be used; for example, patients on pharmacotherapy may be encouraged to decrease avoidance and increase exposure while gradually tapering dosage.

Somatic Therapies. Targeted neurosurgical lesions to CSTC tracts have been shown efficacious in a proportion of highly treatment-refractory patients. Initial work focused on stereotactic neurosurgical lesions, with subsequent work using radiosurgical techniques without craniotomy. The more recent focus of neurosurgical work has been on deep brain stimulation (DBS) by implanting electrodes to stimulate specific brain regions. As a reversible, nonablative alternative to cingulotomy/capsulotomy, DBS has shown some promise for treatment-resistant OCD. A small number of centers worldwide have the skills to do such work, and the FDA has approved targeting of the ventral striatum/ventral capsule. Transcranial magnetic stimulation (TMS) has also been studied for OCD, and deep TMS provided in tandem with CBT is FDA approved for treatment-resistant OCD. Electroconvulsive therapy is not thought to be efficacious for OCD.

Figure 15–1 summarizes the epidemiology, pathology, diagnosis, treatment, and prognosis for OCD.

BODY DYSMORPHIC DISORDER

Introduction

Although long described in the medical literature, BDD was rarely recognized by clinicians or studied by researchers. It was only introduced into the official nomenclature in DSM-III-R and was only recognized for the first time as an independent disorder in the ICD system in ICD-11. Since its introduction to the nomenclature, there has been a substantial increase in the literature on this underrecognized condition. Recent decades have witnessed a growing awareness of its prevalence and morbidity—embarrassment and shame about symptoms mean that careful screening is needed to ensure case-finding. Fortunately, approaches to evaluation and treatment developed for OCD and appropriately modified for BDD have led to improved outcomes.

Definition

BDD is characterized by persistent preoccupations about one or more perceived defects or flaws in one's appearance. The defects or flaws appear slight or are not observable to others. In response to their appearance concerns, individuals with BDD demonstrate a range of mental acts or behaviors, including comparing themselves with others, checking in the mirror, or camouflaging their perceived flaws.

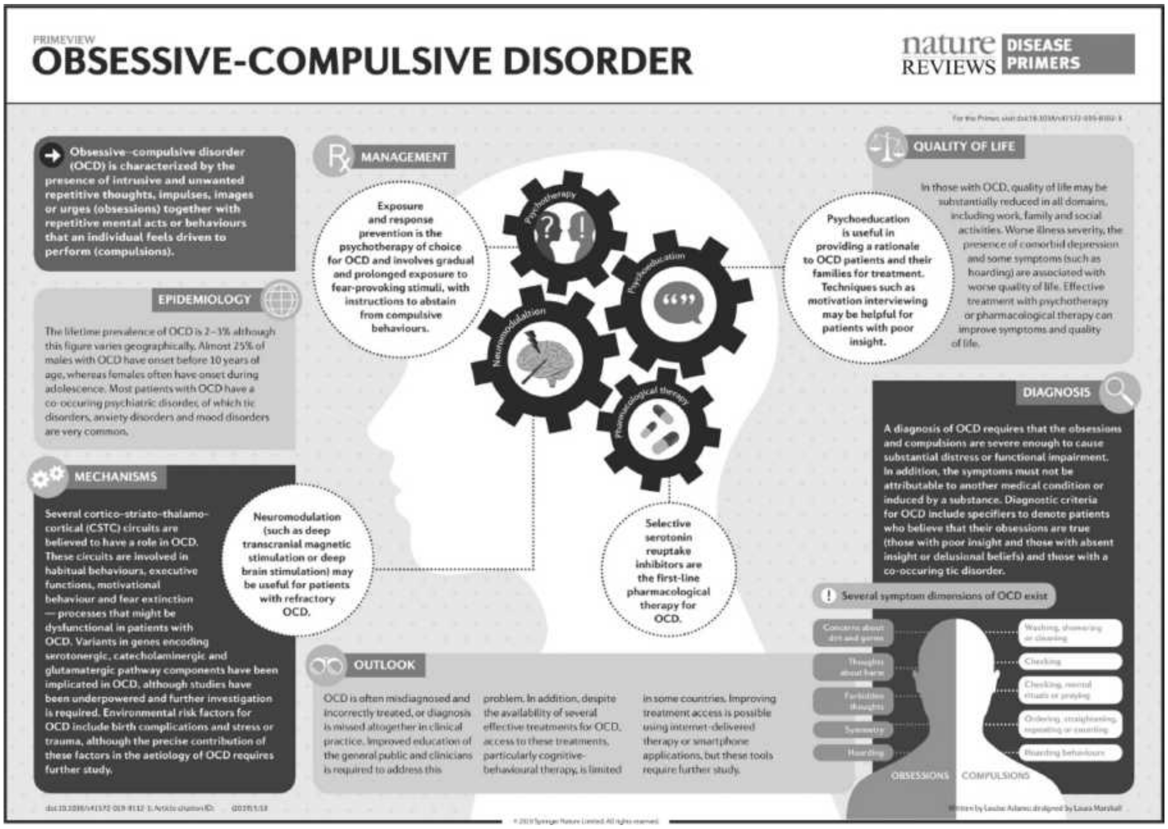

PRIMEVIEW

OBSESSIVE-COMPULSIVE DISORDER

nature REVIEWS DISEASE PRIMERS

Obsessive–compulsive disorder (OCD) is characterized by the presence of intrusive and unwanted repetitive thoughts, impulses, images or urges (obsessions) together with repetitive mental acts or behaviours that an individual feels driven to perform (compulsions).

EPIDEMIOLOGY

The lifetime prevalence of OCD is 2–3% although this figure varies geographically. Almost 25% of males with OCD have onset before 10 years of age, whereas females often have onset during adolescence. Most patients with OCD have a co-occuring psychiatric disorder, of which tic disorders, anxiety disorders and mood disorders are very common.

MECHANISMS

Several cortico-striato-thalamo-cortical (CSTC) circuits are believed to have a role in OCD. These circuits are involved in habitual behaviours, executive functions, motivational behaviour and fear extinction — processes that might be dysfunctional in patients with OCD. Variants in genes encoding serotonergic, catecholaminergic and glutamatergic pathway components have been implicated in OCD, although studies have been underpowered and further investigation is required. Environmental risk factors for OCD include birth complications and stress or trauma, although the precise contribution of these factors in the aetiology of OCD requires further study.

MANAGEMENT

Exposure and response prevention is the psychotherapy of choice for OCD and involves gradual and prolonged exposure to fear-provoking stimuli, with instructions to abstain from compulsive behaviours.

Psychoeducation is useful in providing a rationale to OCD patients and their families for treatment. Techniques such as motivation interviewing may be helpful for patients with poor insight.

Neuromodulation (such as deep transcranial magnetic stimulation or deep brain stimulation) may be useful for patients with refractory OCD.

Selective serotonin reuptake inhibitors are the first-line pharmacological therapy for OCD.

QUALITY OF LIFE

In those with OCD, quality of life may be substantially reduced in all domains, including work, family and social activities. Worse illness severity, the presence of comorbid depression and some symptoms (such as hoarding) are associated with worse quality of life. Effective treatment with psychotherapy or pharmacological therapy can improve symptoms and quality of life.

DIAGNOSIS

A diagnosis of OCD requires that the obsessions and compulsions are severe enough to cause substantial distress or functional impairment. In addition, the symptoms must not be attributable to another medical condition or induced by a substance. Diagnostic criteria for OCD include specifiers to denote patients who believe that their obsessions are true (those with poor insight and those with absent insight or delusional beliefs) and those with a co-occuring tic disorder.

OUTLOOK

OCD is often misdiagnosed and incorrectly treated, or diagnosis is missed altogether in clinical practice. Improved education of the general public and clinicians is required to address this problem. In addition, despite the availability of several effective treatments for OCD, access to these treatments, particularly cognitive-behavioural therapy, is limited in some countries. Improving treatment access is possible using internet-delivered therapy or smartphone applications, but these tools require further study.

Psychotherapy
Psychoeducation
Neuromodulation
Pharmacological therapy
Several symptom dimensions of OCD exist
Concerns about dirt and germs
Thoughts about harm
Forbidden thoughts
Symmetry
Hoarding
Washing, showering or cleaning
Checking
Checking, mental rituals or praying
Ordering, straightening, repeating or counting
Hoarding behaviours
OBSESSIONS
COMPULSIONS

Written by Louise Adams; designed by Laura Marshall

© 2019 Springer Nature Limited. All rights reserved.

FIGURE 15–1. Key points about OCD. (Reprinted with permisison from Springer Nature. Obsessive–compulsive disorder. *Nat Rev Dis Primers.* 2019;5:53. doi:10.1038/s41572-019-0112-1.)

History

The term dysmorphophobia was introduced more than a century ago by an Italian psychiatrist, Enrico Morselli (1852–1929). It is derived from “dysmorphia,” meaning misshapenness or ugliness. Subsequently, both Janet and Freud described cases with BDD symptoms, conceptualizing them as obsessive-compulsive conditions. The introduction of BDD into the official nomenclature in 1987 helped increase clinical awareness of this condition as an independent entity and of how similar BDD was to OCD—which had been found to respond to selective pharmacotherapies and psychotherapies—gave impetus to ongoing research.

Comparative Nosology

DSM-5 made several changes from DSM-IV. The wording was altered to be clearer and perhaps less pejorative to patients. An additional diagnostic criterion was added to address BDD's repetitive mental acts or behaviors. Significantly, an insight specifier was added, including absent insight, so that patients with delusional BDD are now classified as having an OCRD rather than a psychotic disorder. In ICD-10, BDD was collapsed together with hypochondriacal disorder. ICD-11 follows DSM-5 in recognizing BDD as a separate disorder in the chapter on OCRDs.

Epidemiology

BDD has rarely been included in major nationwide community surveys of the prevalence of mental disorders. However, point prevalence estimates in the United States and elsewhere have ranged from 1.7% to 2.4%, indicating that this is one of the more common OCRDs. Furthermore, clinical studies have found high rates of BDD in general adult psychiatric inpatients, cosmetic surgery clinics, and dermatology clinics.

Community and clinical studies confirm that BDD is associated with significant morbidity and comorbidity. Patients are often highly impaired, with very poor quality of life. Compared with OCD, there are high rates of depression and suicidality, further contributing to the burden of disease associated with this condition. The mean age of onset is in adolescence, and as in the case of OCD, community studies indicate a higher prevalence of females with BDD, while clinical reports indicate a roughly even male-to-female ratio.

Etiology

There is evidence that specific cognitive-affective impairments characterize BDD; for example, neuropsychological studies have suggested deficits in executive functioning, including cognitive inflexibility and visual processing. Brain imaging and neurogenetic studies have provided insights into the potential mechanisms that underlie such deficits and potential overlaps with other OCRDs. At the same time, the current evidence base remains relatively thin, and the etiology of BDD is still far from clear.

The few structural and functional brain imaging studies in BDD have tentatively suggested some involvement of the CSTC circuits that appear to play a role in OCD. At the same time, other neurocircuitry may also be relevant, for example, that involved in visual processing. Family and twin data provide preliminary evidence for genetic susceptibility and a relationship to OCD. However, there is a paucity of data on specific genetic variants underlying susceptibility to BDD. While childhood adversity, including a history of teasing, has been associated with BDD, it is not clear if this association is a specific one.

Diagnosis and Clinical Features

As in the case of OCD, the diagnosis of BDD is based on a rigorous psychiatric history and examination. The DSM-5-TR diagnostic criteria for BDD address the presence of appearance preoccupations, repetitive behaviors, or mental acts in response to these preoccupations, the consequent distress and impairment, and emphasize that such symptoms are not due to other medical conditions or other psychiatric disorders. It is crucial for clinicians to also assess avoidance due to being BDD symptoms; this can range from minor social avoidance to housebound. Again, key features to ascertain during the history and examination include the age of onset and course of the symptoms, the type and severity of symptoms, the extent of insight and impairment, and the presence of comorbid disorders and symptoms. Table 15–3 lists the DSM-5-TR criteria for BDD.

Patients incorrectly perceive themselves as ugly or grotesque when they look normal or attractive. Symptoms may focus on any body area, but they most commonly focus on the face and head, with concerns most often involving the skin, nose shape and size, and hair. Individuals may be concerned with one feature or many; throughout their illness, patients are preoccupied with about five to seven different body areas. More than a quarter of patients are concerned with the symmetry of their appearance.

Table 15–3.
Body Dysmorphic Disorder

Diagnostic Criteria

A. Preoccupation with one or more perceived defects or flaws in physical appearance that are not observable or appear slight to others.

B. At some point during the course of the disorder, the individual has performed repetitive behaviors (e.g., mirror checking, excessive grooming, skin picking, reassurance seeking) or mental acts (e.g., comparing his or her appearance with that of others) in response to the appearance concerns.

C. The preoccupation causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

D. The appearance preoccupation is not better explained by concerns with body fat or weight in an individual whose symptoms meet diagnostic criteria for an eating disorder.

Specify if:

- With muscle dysmorphia: The individual is preoccupied with the idea that his or her body build is too small or insufficiently muscular. This specifier is used even if the individual is preoccupied with other body areas, which is often the case.

Specify if:

- Indicate degree of insight regarding body dysmorphic disorder beliefs (e.g., “I look ugly” or “I look deformed”).
- With good or fair insight: The individual recognizes that the body dysmorphic disorder beliefs are definitely or probably not true or that they may or may not be true.
- With poor insight: The individual thinks that the body dysmorphic disorder beliefs are probably true.
- With absent insight/delusional beliefs: The individual is completely convinced that the body dysmorphic disorder beliefs are true.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:271, 272.

Most individuals with BDD have ideas of reference, falsely believing that people are staring at or talking about their deformities. BDD by proxy is rarely described in the literature, that is, preoccupation with the appearance of others, such as a family member.

Most patients attempt to camouflage their perceived flaws. Common repetitive behaviors in BDD include comparing body parts with the same areas in others, checking body parts in the mirror or other reflective surfaces, grooming (including applying make-up, hair-pulling, and skin-picking), seeking reassurance, touching body parts to check them, changing clothes, dieting, tanning, exercising, lifting weights, and undergoing cosmetic procedures. Such symptoms can involve serious harm to self (e.g., skin-picking can lead to significant medical complications) and others.

On one survey, 40% of cosmetic surgeons reported that a dissatisfied BDD patient had threatened them legally or physically, and there are occasional reports of homicide.

Individuals with BDD generally have less insight into the excessiveness and irrationality of their beliefs than those with OCD. The section on BDD in DSM-5-TR includes a specifier for level of insight. Only a quarter of BDD patients have fair or good insight, and around one-third have absent insight. Patients with delusional BDD have greater symptom severity and more significant morbidity. Individuals with BDD sometimes focus on whether their body build is too small or not sufficiently muscular; such so-called muscle dysmorphia is also available as a DSM-5-TR BDD specifier. Patients with muscle dysmorphia mostly look normal, though some may be unusually muscular due to body-building. They are more likely to be male and are more likely to have abnormal eating behavior (e.g., eating high-protein meals) and abuse substances (e.g., anabolic steroids) related to their appearance concerns. Compared to other patients with BDD, they may also be more suicidal.

Of the various OCRDs, it appears that BDD is the one most associated with comorbid depression with a lifetime rate of 75%, and suicidality with 80% having experienced suicidal ideation. About one-third of BDD patients have a lifetime history of OCD, and around 30% of BDD patients experience panic attacks triggered by appearance concerns. BDD is also associated with high levels of rejection sensitivity and low self-esteem. Many patients with BDD use substances to self-medicate social anxiety or emotional pain symptoms. Several measures are now available to assess BDD symptoms and their severity. The Y-BOCS has been adapted for use in BDD, and the Brown Assessment of Beliefs Scale can be used to assess insight. Practical screening questions for BDD include “Are you very worried about your appearance in any way?” and “Are you unhappy with how you look?”

A majority of BDD cases have onset before 18; in children and adolescents, it is essential to assess the impact of BDD symptoms on developmental trajectories and family functioning. Males and females have similar presentations, although males are more likely to be preoccupied with thinning hair, small penis size, and small body build, while females are more likely to be preoccupied with weight, breasts, hips, legs, and excessive body hair. Women are more likely to camouflage their bodies, check mirrors, and pick their skin. BDD symptoms appear universal, although local contexts may influence the precise range of appearance concerns. In the East, a variant of taijin kyofusho, termed *shubo-kyofu*, is characterized by a fear that one’s ugliness will offend others.

Pathology and Laboratory Features

At the current time, there are no biomarkers for BDD. Brain imaging may ultimately be clinically helpful in predicting response to treatment, but presently it remains a research tool. Similarly, while neurogenetic investigations have been employed in the research context, they do not have sufficient clinical utility for more widespread use.

Differential Diagnosis

The specific focus of BDD on appearance concerns serves to differentiate it from other conditions characterized by repetitive behaviors (e.g., OCD, TTM, excoriation disorder), anxiety or social avoidance (e.g., generalized anxiety disorder, panic disorder, social anxiety disorder, agoraphobia), depression (e.g., major depressive disorder), or delusional beliefs (e.g., psychotic disorder). For example, some patients with BDD have hair-pulling, and many have skin-picking, but in BDD, such behaviors are in response to concerns about body flaws, while in TTM and excoriation disorder this is not the case.

Differentiating BDD from eating disorders may require care; if the patient’s appearance concerns focus only on body fat or weight, and symptoms meet the criteria for an eating disorder, then an eating disorder should be diagnosed. That said, BDD and eating disorders are often comorbid. BDD should not be diagnosed when an individual is only concerned with their genitals or secondary sex characteristics and when symptoms meet the criteria for gender dysphoria.

Olfactory reference disorder is characterized by repetitive concerns about emitting a foul odor from the body or mouth (i.e., halitosis), which leads to a range of repetitive behaviors such as repetitive washing, changing clothing, or toothbrushing. Compared to OCD and BDD, those with olfactory reference disorder are even more likely to demonstrate poor or no insight. Nevertheless, it may be helpful to classify olfactory reference disorder as an OCRD, given that some patients with this presentation appear to respond to treatment with SSRIs.

Illness anxiety disorder (in DSM-5-TR) and hypochondriasis (in ICD-11) are repetitive thoughts about being ill rather than about body flaws. As in the case of BDD, however, there are compulsive responses to the somatic preoccupations (e.g., checking, visiting physicians), and treatment approaches overlap to some extent with those used for OCD.

Course and Prognosis

The course of BDD is generally chronic, and in the absence of intervention, symptoms may persist for decades. Earlier age of onset and more severe symptoms at intake may predict a worse course.

Treatment

Psychoeducation is a critical component of the treatment of BDD. Exploring the patient’s explanatory model is a practical first step to understanding the patient’s suffering, building a shared decision-making approach to care, and instilling hope. Basic facts, such as the frequent response of symptoms to treatment, may bring some relief. While patients cannot simply be talked out of their symptoms, key facts, such as that cosmetic procedures may exacerbate symptoms, may need to be presented. Engaging patients at the start of treatment is critical and challenging when there are low levels of insight. Given high rates of suicidality in BDD, evaluating and monitoring such symptoms is essential throughout treatment.

Pharmacotherapy. Clomipramine is more effective than desipramine in BDD, and limited but rigorous data from randomized controlled trials indicate that SSRIs are efficacious in treating BDD, including delusional BDD. Thus SSRIs are first-line agents for the pharmacotherapy of BDD, including delusional BDD. There is also some limited evidence for the efficacy of these agents in children and adolescents with BDD. Unfortunately, the pharmaceutical industry has not undertaken registration trials.

There is some evidence that, as in the case of OCD, doses higher than those ordinarily used in the treatment of depression are beneficial in BDD and that a longer treatment duration is needed to see the full effects of treatment. Again, the duration of treatment should be longer rather than shorter, with gradual tapering of dosage and close monitoring of emergent symptoms.

Patients who do not respond to one SRI may respond to another. There is anecdotal evidence that buspirone augmentation of SRIs is useful in partial responders. The very limited controlled data on the augmentation of SRIs with low-dose dopamine receptor antagonists are not favorable. However, in clinical practice, strategies used for treatment-resistant BDD include adding antipsychotics such as aripiprazole and monotherapy with irreversible monoamine oxidase inhibitors.

Psychotherapy. CBT for OCD has been modified for BDD, and studies have indicated efficacy, including for individuals with delusional BDD. Core elements of such treatments include exposure, response prevention, and cognitive restructuring. As in the case of OCD, the dose and duration of treatment should be sufficient; some have suggested weekly CBT sessions for at least 6 months, with daily homework.

Modifications of OCD manuals are necessary to address the substantial social avoidance often present in BDD (e.g., use of behavioral experiments to test the accuracy of beliefs about others) and to target visual processing abnormalities (e.g., use of perceptual retraining to develop a more holistic view of appearance). Further, given that BDD is accompanied by ideas of reference and often by poor or absent insight, motivational interviewing may be needed when ambivalence about treatment is expressed, and cognitive interventions may need to be more complex and more intensive than in OCD. Habit reversal may be needed for BDD-related skin-picking, hair-pulling, and body touching.

Combined Treatment. There are no data from rigorous trials comparing pharmacotherapy and psychotherapy in BDD. Nevertheless, it would seem sensible to make judicious use of both in the clinical setting. Thus, pharmacotherapy may be helpful at the start of treatment if patients are too depressed to undergo CBT for BDD. Psychotherapy may be helpful while gradually tapering medication dosage over the long term. A judicious combination or sequencing of modalities may be needed in clinical practice.

HOARDING DISORDER

Introduction

Hoarding symptoms have long been described in the medical and psychiatric literature, but HD is a recent entry into the official nomenclature, appearing first in DSM-5. This will hopefully give further impetus to research on this condition. Like so many other OCRDs, it is prevalent and impairing but underdiagnosed and undertreated.

Definition

HD involves accumulating possessions due to persistent and profound difficulty discarding or giving away possessions. This leads to significant congestion, clutter, and substantive distress or impairment, differentiating normal collecting from the clinical disorder.

History

Freud described hoarding symptoms as a component of OCPD. In the DSM-III era, clinicians interested in OCD typically viewed hoarding as a subtype of OCD. During work on DSM-5, a review of the growing literature on hoarding symptoms emphasized that many patients with clinical hoarding did not meet diagnostic criteria for either OCPD or OCD and noted important differences between OCD and hoarding in phenomenology and psychobiology, and so concluded that HD should be included in the nomenclature as a separate diagnostic entity. Hoarding symptoms are a predictor of a worse response to treatment for OCD, and one hope is that more specific diagnostic categories will help lead to better outcomes for patients.

Comparative Nosology

Hoarding is mentioned in DSM-IV in the section on OCPD but is not described in the equivalent ICD-10 category of anankastic personality disorder. DSM-IV also noted that when differentiating OCD from OCPD, a diagnosis of OCD should be considered when hoarding is extreme. DSM-5, and its text revision (DSM-5-TR) for the first time, provides diagnostic criteria for HD and includes this condition in the OCRD section. Similarly, ICD-11 includes HD as a separate entity. There are distinctions between DSM-5-TR and the ICD-11 HD; for example, ICD-11 emphasizes that excessive acquisition is seen in most patients with this condition, whereas DSM-5-TR includes excessive acquisition as a specifier.

Epidemiology

DSM-5 and DSM-5-TR diagnostic criteria for HD have not yet been widely employed in epidemiologic studies. However, preliminary data indicate that HD is remarkably prevalent in Western countries, with one study employing DSM-5 criteria yielding a point prevalence of 1.5%. Epidemiologic data indicate that HD is associated with significant comorbidity and morbidity. In a study using DSM-5 criteria, the male-to-female ratio was about 1. In contrast, females are more common in clinics than males, perhaps indicating greater insight. Unlike the picture seen in most OCRDs, the prevalence of hoarding increases with age, and in clinical settings, the average age of first treatment is around 50. As in the case of other OCRDs, most studies have been done in high-income countries, and much further work is needed in low- and middle-income countries.

Etiology

Patients with HD may demonstrate impairment in various neuropsychological domains, including spatial planning, working memory, response inhibition, and set-shifting. Structural and functional brain imaging studies have indicated that the relevant neurocircuitry overlaps partially with that involved in OCD. Research on patients with neurologic disorders has suggested that ventromedial prefrontal-anterior cingulate cortices and medial temporal regions are involved in hoarding.

Animal studies have pointed to a possible role for the dopaminergic system in hoarding. Twin studies indicate genetic susceptibility for developing pathologic hoarding. While there is a relative dearth of neurogenetic studies, GWAS data from OCD studies provide a preliminary indication that a range of genetic variants are involved in mediating hoarding symptoms.

Diagnosis and Clinical Features

The diagnosis of HD is made based on a rigorous psychiatric history and examination. DSM-5-TR diagnostic criteria for HD emphasize that individuals have difficulty discarding or parting with possessions, regardless of their actual value. This difficulty is due to a perceived need to save the items and the associated distress from discarding them. It results in congestion and clutter so that the use of living areas for their intended use is severely compromised. If living areas are uncluttered, it is likely that others assisted the individual. For example, friends or family members may agree to store items in their own homes. Hoarding is associated with clinically significant distress or impairment and is not attributable to another medical condition or mental disorder. Table 15–4 lists the DSM-5-TR diagnostic criteria for HD.

Key features to ascertain during the history and examination include the age of onset and course of the symptoms, the type and severity of symptoms, the extent of insight and impairment, and the presence of comorbid disorders and symptoms. Many objects may be hoarded, but they are usually everyday items (e.g., newspapers, books, clothing), and the resulting clutter is usually highly disorganized. There is literature on animal hoarding, which some conceptualize as a subtype of HD. Patients cite various reasons for hoarding, including utilitarian value, sentimental attachment, concerns about losing information, and not being wasteful. In some cases, exposure to traumatic or stressful events may precede the onset or exacerbation of hoarding.

Table 15–4.
Hoarding Disorder

Diagnostic Criteria

A. Persistent difficulty discarding or parting with possessions, regardless of their actual value.
B. This difficulty is due to a perceived need to save the items and to distress associated with discarding them.
C. The difficulty discarding possessions results in the accumulation of possessions that congest and clutter active living areas and substantially compromises their intended use. If living areas are uncluttered, it is only because of the interventions of third parties (e.g., family members, cleaners, authorities).
D. The hoarding causes clinically significant distress or impairment in social, occupational, or other important areas of functioning (including maintaining a safe environment for self and others).
E. The hoarding is not attributable to another medical condition (e.g., brain injury, cerebrovascular disease, Prader–Willi syndrome).
F. The hoarding is not better explained by the symptoms of another mental disorder (e.g., obsessions in obsessive-compulsive disorder, decreased energy in major depressive disorder, delusions in schizophrenia or another psychotic disorder, cognitive deficits in major neurocognitive disorder, restricted interests in autism spectrum disorder).

Specify if:

- With excessive acquisition: If difficulty discarding possessions is accompanied by excessive acquisition of items that are not needed or for which there is no available space.

Specify if:

- With good or fair insight: The individual recognizes that hoarding-related beliefs and behaviors (pertaining to difficulty discarding items, clutter, or excessive acquisition) are problematic.
- With poor insight: The individual is mostly convinced that hoarding-related beliefs and behaviors (pertaining to difficulty discarding items, clutter, or excessive acquisition) are not problematic despite evidence to the contrary.
- With absent insight/delusional beliefs: The individual is completely convinced that hoarding-related beliefs and behaviors (pertaining to difficulty discarding items, clutter, or excessive acquisition) are not problematic despite evidence to the contrary.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:277.

Hoarding may be associated with a broad range of sequelae, including health and safety problems (e.g., patients may not have sufficient space to cook or keep clean and may be at increased risk for a broad range of illnesses). “Clutter avalanches” may lead to falls and even mortality. In some individuals, conditions become entirely unsanitary; this is termed “domestic squalor.” Family relationships can be negatively impacted, and in some cases, dependents may need to be removed. Threats of eviction or forced clearing of possessions may lead to substantial distress and even suicide.

DSM-5-TR provides two specifiers for HD. The excessive acquisition is specified if difficulty discarding possessions is accompanied by excessive acquisition of items that are not needed or for which there is no available space. As in OCD and BDD, insight can be specified as good or fair, poor, or absent/delusional. However, the definition is framed somewhat differently from that in OCD and BDD; in absent/delusional insight, the individual is wholly convinced that hoarding-related beliefs and behaviors are not problematic despite evidence to the contrary. History from other informants may be critical for accurately evaluating symptoms when insight is absent.

Commonly associated features of HD include indecisiveness and procrastination, difficulty planning and organizing tasks, and distractibility and avoidance. Comorbid OCD is elevated in HD compared with the general population. Prevalent comorbid disorders in HD include generalized anxiety disorder and major depressive disorder. Clinical evidence suggests that other OCRDs may also be more commonly comorbid than expected by chance. In older adult patients with HD, there may be medical complications of HD. Several different measures, including the Saving Inventory-Revised and the Hoarding Rating Scale, can be used to assess the severity of hoarding symptoms. Home visits, or home photographs, are helpful for the accurate assessment of symptoms.

Pathology and Laboratory Examination

There are no biomarkers for HD.

Differential Diagnosis

Differentiating normal collecting from HD is not typically problematic. In normal collecting, accumulation is confined to a narrow range of objects, acquired in a structured way, and then well organized. Furthermore, collecting is a pleasurable activity that does not lead to impairment. In contrast, pathologic hoarding is characterized by the accumulation of a range of objects, unstructured or impulsive acquisition, haphazard clutter, as well as consequent distress and impairment.

One key differential diagnosis to consider is that of OCD. In OCD, hoarding is undertaken in response to an unwanted obsession, most commonly due to symmetry obsessions, feelings of incompleteness, or fears of harm or contamination (e.g., an object may be kept in case discarding that object leads to harm). In contrast, in HD, there is no associated obsession; instead, the individual wishes to retain saved items and is distressed by the thought of discarding them. In some cases, both OCD and comorbid HD may be diagnosed.

Although the DSM-5 proposal was for OCPD to remove the hoarding criterion, no changes were made to the DSM-IV diagnostic criteria for personality disorders. Hoarding symptoms may also be seen in a range of other psychiatric disorders, including autism spectrum disorders (where collections may reflect a particular interest) and psychotic disorders. HD should only be diagnosed if the hoarding symptoms are independent of these other conditions; that is when they have a different onset and course.

Several neurodevelopmental and neurodegenerative conditions are associated with hoarding, for example, Prader–Willi Syndrome (where individuals are characterized by hoarding of food, obesity, and skin-picking) and Alzheimer disease. Damage to particular brain regions, including the anterior ventromedial prefrontal and cingulate cortices, has been associated with excessive and indiscriminate hoarding activity. Saving “bizarre” items (e.g., feces, rotten food) or severe squalor may point to a diagnosis other than HD.

Course and Prognosis

HD has a somewhat later onset than most OCRDs. Without intervention, its course appears to be often chronic and progressive. Hoarding symptoms often begin in childhood and adolescence, with diagnostic criteria met only in the 30s and worsening symptoms in each subsequent decade.

Treatment

Pharmacotherapy. The literature on pharmacotherapy of OCD has indicated that hoarding symptoms predict less favorable outcomes to treatment with SSRIs. Nevertheless, these agents remain the first-line intervention for HD in clinical practice, and uncontrolled data indicate that these agents, as well as the serotonin–norepinephrine reuptake inhibitor venlafaxine, may be efficacious for this indication. The number of studies on the pharmacotherapy of HD is relatively small, but, interestingly, some case reports have noted a response to glutamate modulators.

Psychotherapy. CBT has been adapted for HD, and research to date indicates that this manualized treatment is superior to a wait-list. Components include psychoeducation and goal setting, organization skills training, practice with resisting acquisition, and cognitive restructuring. Many patients have limited insight, so motivational interview techniques may be needed to engage patients in treatment. Group therapy is considered a cost-effective alternative to individual CBT. Involving family members and social services may be critical, particularly when insight is absent.

Combined Treatment. There are no rigorous studies of combined treatment in HD, and clinical judgment is used to determine the optimal combination and sequencing of different modalities.

TRICHOTILLOMANIA (HAIR-PULLING DISORDER)

Introduction

The term TTM was coined by a French dermatologist, François Hallopeau (1842–1919), in the 19th century. In the early part of the 20th century, a large case series of patients was described by two surgeons, who operated on individuals with trichobezoars (due to swallowing of hair after hair-pulling). Attention was directed to TTM when researchers discovered that, like OCD, TTM responded selectively to clomipramine. Interest in the disorder increased when consumer advocacy organizations rose in response to the demand for appropriate recognition and efficacious interventions for this underrecognized and undertreated condition.

Definition

TTM is characterized by pulling out of one’s hair. This is not done for cosmetic reasons, but instead, individuals often describe an irresistible urge to pull out their hair.

History

Pulling out of one’s hair has long been described in the medical literature as a symptom of psychopathology, with descriptions dating at least as far back as Hippocrates. Hallopeau’s term derives from the Greek for hair (trich) and pull (till). However, TTM is a relatively late entrant to the nomenclature; it was included for the first time in DSM-III-R. In DSM-5, TTM was first characterized as an OCRD given its overlapping features with OCD. In ICD-11, TTM and excoriation (skin-picking) disorders are classified as BFRBDs, appropriately recognizing this overlooked set of conditions.

Comparative Nosology

In DSM-IV, TTM was classified as an impulse control disorder not otherwise classified. Accordingly, diagnostic criteria include a description of rising tension prior to hair-pulling, followed by gratification during hair-pulling. DSM-5 reviews noted that many patients with TTM did not meet diagnostic criteria for these two phenomena. Patients who did not meet these diagnostic criteria did not demonstrate clinically meaningful differences, so they suggested replacing this with a criterion referring to unsuccessful attempts to stop hair-pulling.

Epidemiology

Studies of select populations, such as college students, indicate that noncosmetic hair-pulling has a prevalence of 10% or higher, with 1% to 3% meeting DSM-5 diagnostic criteria for TTM. Age of onset is typically at menarche, and female to male ratio is about 9:1. Pediatric samples include more girls but with a less skewed gender ratio. There is also growing evidence of the comorbidity and morbidity associated with TTM; this includes comorbid psychiatric disorders, various medical sequelae (e.g., trichobezoars), and decreased quality of life.

Etiology

From a cognitive-affective neuroscience perspective, neuropsychological investigations of TTM have demonstrated some evidence of impairments, including deficits in working memory and visual–spatial learning. Structural and functional brain imaging studies have suggested some involvement of CSTC circuits relevant to habit learning, although the database of studies is sparse. There is also evidence of the involvement of other regions, such as those involved in reward processing and affect regulation, which may well be relevant to TTM.

Animal models of TTM include “barbering” and have investigated the role of various molecular systems in this behavior. Family and twin studies provide evidence for genetic susceptibility and a relationship to OCD and other BFRBDs. Several interesting candidate genes have preliminarily been associated with TTM, including SAPAP3 and SLITRK1, but unbiased genome-wide association studies have not yet described the genetic architecture of this disorder. There is evidence of increased childhood trauma in TTM, but this association is likely not specific to this particular OCRD.

Diagnosis and Clinical Features

The diagnosis of TTM is made based on a rigorous psychiatric history and examination. DSM-5-TR criteria refer to hair-pulling resulting in hair loss, repeated attempts to decrease or stop hair-pulling, clinically significant distress or impairment, with the hair-pulling or loss not attributable to another medical condition, nor better explained by the symptoms of another medical disorder. Key features to ascertain during the history and examination include the age of onset and course of the symptoms, the type of symptoms, the extent of impairment, and the presence of comorbid disorders and symptoms. Table 15–5 lists the diagnostic criteria for TTM.

Hair-pulling may be from any body area but is most commonly from the scalp, eyebrows, and eyelashes. It is worth inquiring about hair-pulling from the pubic area as this may be more embarrassing for patients to volunteer. Patients often pull hair daily, either more continuously or in fewer isolated episodes. Most patients pull with their fingers, but some use tweezers or other instruments. Patients may have a range of habits associated with hair-pulling, for example, stroking or playing with hair, choosing a particular kind of hair to pull, and mouthing the hair once it is pulled. It is also important to inquire about trichophagy (compulsive eating of hair), partly because of associated shame and partly because it may have serious medical sequelae. Some individuals pull hair from dolls or other objects, pets, or other people.

States of both hyperarousal (e.g., stress) and hypoarousal (e.g., boredom) may trigger hair-pulling. It has been suggested that there are two significant kinds of hair-pulling; focused and automatic. In focused hair-pulling, the person is aware of an urge to pull, tries to resist this, and then gives in. In automatic hair-pulling, the person has less awareness of the hair-pulling, and before they realize it, has pulled out several hairs. Most people with TTM report both kinds of hair-pulling, but individuals differ in the extent to which they employ focused or automatic hair-pulling. Hair-pulling may occasionally be seen in infants and toddlers; this is not well studied but may be self-soothing and self-limited.

TTM is associated with substantial morbidity and comorbidity. First, TTM is associated with significant emotional distress. Repeated failures to control hair-pulling may lead to shame, frustration, and decreased self-esteem. Time-consuming camouflage and avoidance of situations or activities where hair loss may be discovered may further contribute to impairment. Comorbid conditions in TTM include other BFRBDs, with excoriation (skin-picking) disorder the most common. OCD is also more prevalent in TTM than in the general population. More than half of treatment-seeking individuals with TTM have a comorbid psychiatric disorder, with mood and anxiety disorders the most common. A range of measures is now available to assess hair-pulling severity and type, including the Massachusetts General Hospital Hairpulling Scale and the Milwaukee Inventory for Subtypes of TTM. Photographs of pulled areas may also be helpful to assess progress during treatment.

Table 15–5.
Trichotillomania (Hair-Pulling Disorder)

Diagnostic Criteria

A. Recurrent pulling out of one’s hair, resulting in hair loss.
B. Repeated attempts to decrease or stop hair pulling.
C. The hair pulling causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The hair pulling or hair loss is not attributable to another medical condition (e.g., a dermatological condition).
E. The hair pulling is not better explained by the symptoms of another mental disorder (e.g., attempts to improve a perceived defect or flaw in appearance in body dysmorphic disorder).

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:281.

Hair-pulling symptoms are somewhat similar across the life course, although younger patients may be less likely to report urges or arousal prior to pulling or gratification and relief after pulling. Females and males also have similar symptoms, but there is some evidence that males have slight differences in hair-pulling patterns (e.g., pulling more commonly from face and chest) and are more likely to have comorbid tics. While there are limited data on TTM from different countries and cultures symptoms seem relatively similar across the world.

Pathology and Laboratory Examination

The majority of patients with TTM can describe their symptoms. In cases of alopecia, where hair-pulling is denied and the diagnosis is unclear, trichoscopy may be helpful.

Differential Diagnosis

TTM should be differentiated from hair-pulling for cosmetic or cultural reasons and from other repetitive hair behaviors that do not lead to pulling. In TTM, there is no associated intrusive obsession or appearance concern, differentiating it from hair-pulling due to OCD (with symmetry concerns) or BDD (with body appearance concerns). Similarly, on the rare occasion that hair-pulling is in response to a psychotic symptom or a substance, TTM should not be diagnosed. Alopecia may occur for a range of reasons other than TTM.

Course and Prognosis

There are relatively few data on the long-term course of TTM. In the absence of intervention, its course is likely often chronic. However, some data indicate that remission does occur in a proportion of cases.

Treatment

Psychoeducation is a critical component of the treatment of TTM. As in other OCRDs, exploring the patient's explanatory model is a useful first step in progressing key elements of treatment such as understanding the patient's suffering, building a collaborative approach to care, and instilling hope. Many patients and families are relieved to learn basic facts about TTM, including its high prevalence, the fact that many others with TTM are embarrassed by or ashamed of symptoms, that it is not due to commonly voiced factors (e.g., bad mothering), and that treatments are efficacious for many. Consumer advocacy organizations, such as the Trichotillomania Learning Centre (TLC) Foundation for Body-Focused Repetitive Behaviors, may aid in conveying such information.

Pharmacotherapy. Early data indicated that TTM, like OCD and BDD, responded more robustly to clomipramine than desipramine. Unfortunately, subsequent SSRI trials in TTM have not consistently demonstrated efficacy. The most significant positive clinical trial in adults with TTM to date is with N-acetylcysteine (NAC) (1,200 to 2,400 mg/day), a nutraceutical that acts on the glutamatergic system. However, a trial of NAC in pediatric TTM was negative. Small controlled trials indicate that dopamine receptor blockers may also be helpful in these conditions. Given its favorable side-effect profile, it may be reasonable to use NAC as a first-line agent in adults with TTM, use SSRIs when patients have comorbid conditions for which these agents are efficacious, and consider low-dose dopamine receptor antagonists in more treatment-refractory cases.

Psychotherapy. Habit reversal therapy (HRT) is a set of cognitive-behavioral techniques that have been shown efficacious in treating childhood and adult TTM. Components of HRT include awareness training, competing response training, and social support. Awareness training includes elements such as self-monitoring and collecting pulled hair. In competing response training, the patient substitutes hair-pulling with an incompatible action (e.g., fist-clenching). Social support involves using another individual to reinforce the competing response and may be particularly useful in pediatric TTM. In addition, stimulus control (SC) is used to change the patient's environment so that hair-pulling is more effortful and less reinforcing. There is also preliminary evidence that web-based administration of HRT/SC may be efficacious. Finally, controlled trials of augmentation of HRC/SC with acceptance and commitment therapy, dialectical behavior therapy, or cognitive therapy, which focus on strategies such as affect regulation, enhanced motivation, relapse prevention, or cognitive restructuring, have also suggested efficacy in TTM.

Combined Treatment. There is preliminary evidence that SSRIs in combination with psychotherapy are more beneficial than either treatment alone in some individuals. From a clinical perspective, judicious combination and sequencing of different treatment modalities may be helpful.

EXCORIATION (SKIN-PICKING) DISORDER

Introduction

Cases of "neurotic excoriation" or "psychogenic excoriation" have long been described in the medical literature. However, excoriation (skin-picking) disorder is new to DSM-5. This debut reflects a gradually accumulating research literature, and hopefully, the inclusion of skin-picking disorder (SPD) in the nomenclature will give impetus to even more investigation and bring this often overlooked condition to the attention of clinicians. In ICD-11, TTM and excoriation disorder are listed together under the umbrella of BFRBDs, given their many similarities.

Definition

SPD is characterized by recurrent skin-picking, resulting in skin lesions. Individuals with skin-picking attempt to decrease or stop skin-picking but cannot do so. While some picking of scabs is likely universal, there is associated clinically significant distress or impairment in the case of SPD.

History

This disorder was first described in detail by a French dermatologist, Louis-Anne-Jean Brocq (1856–1928), who noted the development of skin-picking in young women with acne (“acne excoriee”). The field of psychodermatology was influenced by psychodynamic constructs, as evidenced by the subsequent term “neurotic excoriation.” In 1978, Robert D. Griesemer reported that 2% of dermatology patients had SPD. Early work with the SSRIs suggested that these agents were helpful in SPD, and productive questions arose about the relationship between OCD, TTM, and SPD. Consumer advocacy associations have further raised awareness of TTM, SPD, and their frequent co-occurrence.

Comparative Nosology

In DSM-IV, SPD is mentioned briefly as an “impulse control disorder not otherwise specified” in the chapter on “impulse control disorders not elsewhere classified.” SPD is classified in the nomenclature for the first time in DSM-5. This inclusion recognized the increasing scientific evidence for the diagnostic validity and clinical utility of this entity, as well as the voice of patients and patient advocates. Diagnostic criteria in DSM-5 and DSM-5-TR closely reflect TTM, consistent with the many phenomenologic similarities between these conditions. Diagnostic guidelines in ICD-11 are also similar, but both TTM and SPD fall under the umbrella of BFRBDs. The most common of these conditions are TTM and SPD, but they include a range of other conditions with repetitive behaviors that damage the integument, including, for example, severe nail-biting, lip-chewing, and cheek chewing.

Epidemiology

Community-based prevalence studies of SPD have found a point prevalence of 1.4% to 5.4%. Age of onset is somewhat more variable than TTM, although the mean age is 12 years, typically coinciding with the onset of puberty. Although many more females than males have SPD, the ratio is more even than in the case of TTM. There is also growing evidence of the morbidity and comorbidity associated with SPD; this includes comorbid psychiatric disorders, various medical sequelae (e.g., infection), and decreased quality of life.

Etiology

From a cognitive-affective neuroscience perspective, neuropsychological investigations of SPD have demonstrated some overlap with OCD in that there is evidence of motor impulsivity. Brain imaging studies have similarly suggested some involvement of CSTC circuitry, although the database of investigations is sparse, and other regions also appear to be implicated.

Animal data, the finding that skin-picking is precipitated by dopaminergic stimulants, and respond to dopamine receptor antagonists suggests that the dopaminergic system plays a role in SPD. Family and twin studies provide evidence for genetic susceptibility and a relationship with OCD and other BFRBDs. Several interesting candidate genes have preliminarily been associated with SPD, including SAPAP3, but unbiased whole-genome association studies have not been undertaken.

Diagnosis and Clinical Features

The diagnosis of SPD is based on a rigorous psychiatric history and examination. DSM-5-TR diagnostic criteria for SPD refer to skin-picking resulting in skin lesions, repeated attempts to decrease or stop skin-picking, clinically significant distress or impairment, with the skin-picking not attributable to another medical condition, nor better explained by the symptoms of another medical disorder. Key features to ascertain during the history and examination include the age of onset and course of the symptoms, the type of symptoms, the extent of impairment, and the presence of comorbid disorders and symptoms. Table 15–6 lists the diagnostic criteria for Excoriation (Skin-Picking) Disorder.

Table 15–6.
Excoriation (Skin-Picking) Disorder

Diagnostic Criteria

A. Recurrent skin picking resulting in skin lesions.
B. Repeated attempts to decrease or stop skin picking.
C. The skin picking causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The skin picking is not attributable to the physiological effects of a substance (e.g., cocaine) or another medical condition (e.g., scabies).
E. The skin picking is not better explained by symptoms of another mental disorder (e.g., delusions or tactile hallucinations in a psychotic disorder, attempts to improve a perceived defect or flaw in appearance in body dysmorphic disorder, stereotypies in stereotypic movement disorder, or intention to harm oneself in nonsuicidal self-injury).

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:284.

Skin-picking may be from any body area but is most commonly from the face, followed by the hands, fingers, arms, and legs. Patients may pick from multiple sites; one study found that patients picked from an average of 4.5 sites. Most patients pick with their fingernails, but some use tweezers or other instruments. Patients often skin pick daily, either more continuously or in fewer isolated episodes. Individuals may have a range of habits associated with skin-picking, for example, stroking or playing with the skin, choosing a particular scab to pick, and mouthing or swallowing the scab once it is picked. Some individuals pick the skin from other people.

Patients may report that skin-picking began after a dermatologic condition such as acne but continued even with the resolution of that condition. States of both hyperarousal (e.g., stress) and hypoarousal (e.g., boredom) may trigger skin-picking. As with hair-pulling, it has been suggested that there are two major kinds of skin-picking: focused and automatic. In focused skin-picking, the person is aware of an urge to pick, attempts to resist, and then gives in. In automatic skin-picking, the person has less awareness of the skin-picking, and before they realize it, has picked again. Most people with SPD report both kinds of skin-picking, but individuals differ in the extent to which they employ focused or automatic skin-picking.

SPD is associated with substantial morbidity and comorbidity. First, SPD is associated with significant emotional distress. Repeated failures to control skin-picking may lead to shame, depression, and self-medication. Time-consuming camouflage and avoidance of situations or activities where skin lesions may be discovered may further contribute to impairment. Comorbid conditions in SPD include other BFRBDs, with TTM being the most common. OCD and BDD are also more prevalent in SPD than in the general population. Other common comorbid psychiatric disorders in SPD include mood and anxiety disorders. A range of measures are available to assess skin-picking severity and type, including the Skin Picking Impact Scale. A thorough physical examination may be helpful to assess the extent and severity of skin-picking, and photographs of skin lesions may help assess progress.

Skin-picking symptoms are somewhat similar across the life course, although further work is needed in this area. In one study, males had a later age of onset, more severe symptoms, and greater impairment than females with SPD. While there are limited data on SPD from different countries and cultures, symptoms seem relatively similar across the world.

Pathology and Laboratory Examination

Most patients with SPD can describe their symptoms, and dermatologic examination reveals classical features. Thus, for example, lesions often have linear or circular shapes, are often grouped at easily accessible sites, and may be at various stages (from small superficial erosions to deep ulcerations to scars).

Differential Diagnosis

Skin-picking is quite common in BDD, but it is done in response to concerns about body appearance in these individuals, and SPD should not be diagnosed. Similarly, skin-picking may also occur in response to obsessions about contamination, delusions, or tactile hallucinations related to infestation (parasitosis) or may be induced by dopaminergic psychostimulants (e.g., cocaine). Skin picking may also be due to a factitious disorder (dermatitis artefacta). Finally, skin-picking may be seen in response to a broad range of dermatologic conditions associated with itching (such as scabies), as well as in several medical conditions (such as Prader–Willi syndrome).

Course and Prognosis

Available data on the long-term course of SPD suggest that this is often chronic, albeit with fluctuations in severity over time in the absence of intervention. As is the case in other OCRDs, individuals often report that they have not sought medical attention because they are embarrassed by their symptoms, feel they should stop the symptoms themselves, or are unaware that the symptoms may be part of a known condition with efficacious treatments.

Treatment

Psychoeducation is a key component of the treatment of SPD. As in other OCRDs, exploring the patient's explanatory model is a useful first step in understanding the patient's suffering, building a collaborative approach to care, and instilling hope. Many patients and families are relieved to learn basic facts about SPD, including its high prevalence, the fact that many others with SPD are embarrassed by/ashamed of symptoms, that it is not due to commonly voiced factors (e.g., bad mothering), and that treatments are efficacious for many. Consumer advocacy organizations, such as the TLC Foundation for Body-Focused Repetitive Behaviors, may aid in conveying such information.

Pharmacotherapy. There is some evidence from randomized controlled trials that SPD responds to SSRIs, although there are few studies with small samples. The largest positive clinical trial in adults with SPD to date is once again with NAC, a nutraceutical that acts on the glutamatergic system. Preliminary data show that low-dose dopamine receptor antagonists may also be helpful in SPD, but the cost–benefit ratio should be carefully weighed. Data on the efficacy of lamotrigine in SPD are inconsistent.

Psychotherapy. There is a good deal of overlap in psychotherapy for TTM and SPD. HRT has been shown efficacious in treating childhood and adult SPD compared to wait-list controls. Components of HRT for SPD have included awareness training, competing response training, and social support. Awareness training includes elements such as self-monitoring. In competing response training, the patient substitutes skin-picking with an incompatible action (e.g., fist-clenching). Web-based HRT may be helpful. Social support involves using another individual to help reinforce the competing response. Manualized CBT treatment for SPD includes cognitive restructuring and relapse prevention. There is interest in acceptance-enhanced augmentation of HRT for SPD, but this has not been rigorously studied.

Combined Treatment. There are no rigorous studies comparing pharmacotherapy and psychotherapy in SPD. From a clinical perspective, judicious combination and sequencing of modalities may be helpful.

OCRD DUE TO ANOTHER MEDICAL DISORDER

This category is new to DSM-5-TR and ICD-11, reflecting that in a new chapter on OCRDs, it is relevant to exclude conditions due to another medical disorder. To determine whether OCRD symptoms are etiologically linked to another medical disorder, consideration should be given to the presence of a clear temporal relationship between such symptoms and the medical disorder. As noted earlier, obsessive-compulsive symptoms have long been described in the context of various neurologic lesions, most notable in the aftermath of the influenza epidemic in the early 20th century. At postmortem, patients had striatal lesions. More recent work has documented the occurrence of OCD due to several other striatal pathologies, including stroke and infection. For example, it is notable that a significant proportion of cases of Sydenham chorea have obsessive-compulsive symptoms.

Pediatric neuropsychiatric disorders associated with Streptococcus (PANDAS) is a condition thought similar to Sydenham chorea, in that a Streptococcal infection is hypothesized to trigger an autoimmune reaction, resulting in sudden onset of obsessive-compulsive and other symptoms. Rigorous research has demonstrated that cases of PANDAS respond to immunotherapeutic interventions. At the same time, this is a relatively new area of research, and some have argued that sudden-onset obsessive-compulsive symptoms can be seen after a broad range of insults.

SUBSTANCE/MEDICATION-INDUCED OCRD

There is evidence that particular substances or medications can trigger or exacerbate obsessive-compulsive symptoms. For example, skin-picking may be seen in patients with cocaine intoxication. Such work is theoretically interesting as it points to the potential role of the dopaminergic system in OCRDs. However, there is less evidence that chronic symptoms of OCRDs are due to substances or medications.

OTHER SPECIFIED AND UNSPECIFIED OCRDS

DSM-5-TR notes a range of other specified and unspecified OCRDs. Other specified OCRDs include obsessional jealousy and cultural variants such as *shubo-kyofu*, which involves repeated concerns about body appearance, and *koro*, which is a concern that the penis will recede into the body. Other specified OCRDs are also diagnosed when the presentation differs fundamentally from those outlined in the chapter; for example, BDD symptoms in patients who have visible body flaws or do not have repetitive behaviors.

CONCLUSION

The DSM-5-TR chapter on OCRDs aims to improve the assessment and management of these underrecognized and undertreated conditions, which are highly prevalent and associated with a great deal of comorbidity and morbidity. ICD-11 has a similar chapter but includes a broader spectrum of conditions. Both DSM-5-TR and ICD-11 list OCRDs adjacent to anxiety disorders to emphasize the overlap between these groups of conditions. Any particular approach to the meta-structure of the nomenclature arguably has pros and cons; on the one hand, the OCRD chapter highlights significant phenomenologic and psychobiologic similarities across these conditions, but there is the risk that naïve clinicians may overlook important differences required in their evaluation and treatment. Certainly, when assessing an individual with any one OCRD, it may well be helpful to ask about the presence of other OCRDs in the individual and family members. There are some similarities in pharmacotherapy and psychotherapy approaches across these conditions. At the same time, as in other DSM-5-TR and ICD-11 chapters, it is important to emphasize that there are also important distinctions that must be made between these conditions and consequent differences in treatment approaches.

FURTHER READINGS

Angelakis I, Gooding P, Tarrier N, Panagioti M. Suicidality in obsessive compulsive disorder (OCD): a systematic review and meta-analysis. *Clin Psychol Rev*. 2015;39:1–15.

Bienvenu OJ, Samuels JF, Wuyek LA, et al. Is obsessive-compulsive disorder an anxiety disorder, and what, if any, are spectrum conditions? A family study perspective. *Psychol Med*. 2012;42:1–13.

Chang K, Frankovich J, Cooperstock M, et al. Clinical evaluation of youth with pediatric acute-onset neuropsychiatric syndrome (PANS): recommendations from the 2013 PANS Consensus Conference. *J Child Adolesc Psychopharmacol*. 2015;25:3–13.

Davidson EJ, Dozier ME, Pittman JOE, et al. Recent advances in research on hoarding. *Curr Psychiatry Rep*. 2019;21:91.

Dougherty DD, Brennan BP, Stewart SE, Wilhelm S, Widge AS, Rauch SL. Neuroscientifically informed formulation and treatment planning for patients with obsessive-compulsive disorder: a review. *JAMA Psychiatry*. 2018;75:1081–1087.

Ducasse D, Boyer L, Michel P, et al. D2 and D3 dopamine receptor affinity predicts effectiveness of antipsychotic drugs in obsessive-compulsive disorders: a metaregression analysis. *Psychopharmacology (Berl)*. 2014;231:3765–3770.

Fineberg NA, Reghunandanan S, Kolli S, Atmaca M. Obsessive-compulsive (anankastic) personality disorder: toward the ICD-11 classification. *Braz J Psychiatry*. 2014;36(Suppl 1):40–50.

Grant JE, Odlaug BL, Chamberlain SR, Keuthen NJ, Lochner C, Stein DJ. Skin picking disorder. *Am J Psychiatry*. 2012;169:1143–1149.

Kariuki-Nyuthe C, Gomez-Mancilla B, Stein DJ. Obsessive compulsive disorder and the glutamatergic system. *Curr Opin Psychiatry*. 2014;27:32–37.

Lewis-Fernandez R, Hinton DE, Laria AJ, et al. Culture and the anxiety disorders: recommendations for DSM-V. *Depress Anxiety*. 2010;27:212–229.

Lochner C, Fineberg NA, Zohar J, et al. Comorbidity in obsessive-compulsive disorder (OCD): a report from the International College of Obsessive-Compulsive Spectrum Disorders (ICOCS). *Compr Psychiatry*. 2014;55:1513–1519.

Macy AS, Theo JN, Kaufmann SC, et al. Quality of life in obsessive compulsive disorder. *CNS spectrums*. 2013;18:21–33.

Mataix-Cols D, Frost RO, Pertusa A, et al. Hoarding disorder: a new diagnosis for DSM-V? *Depress Anxiety*. 2010;27:556–572.

Phillips KA, Stein DJ, Rauch SL, et al. Should an obsessive-compulsive spectrum grouping of disorders be included in DSM-V? *Depress Anxiety*. 2010;27:528–555.

Radua J, Grau M, van den Heuvel OA, et al. Multimodal voxel-based meta-analysis of white matter abnormalities in obsessive-compulsive disorder. *Neuropsychopharmacology*. 2014;39(7):1547–1557.

Robbins TW, Vaghi MM, Banca P. Obsessive-compulsive disorder: puzzles and prospects. *Neuron*. 2019;102:27–47.

Romanelli RJ, Wu FM, Gamba R, Mojtabai R, Segal JB. Behavioral therapy and serotonin reuptake inhibitor pharmacotherapy in the treatment of obsessive-compulsive disorder: a systematic review and meta-analysis of head-to-head randomized controlled trials. *Depress Anxiety*. 2014;31:641–652.

Ruscio AM, Stein DJ, Chiu WT, Kessler RC. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Mol Psychiatry*. 2010;15:53–63.

Russell EJ, Fawcett JM, Mazmanian D. Risk of obsessive-compulsive disorder in pregnant and postpartum women: a meta-analysis. *J Clin Psychiatry*. 2013;74:377–385.

Sharma E, Thennarasu K, Reddy YCJ. Long-term outcome of obsessive-compulsive disorder in adults: a meta-analysis. *J Clin Psychiatry*. 2014;75:1019–1027.

Stein DJ. Obsessive-compulsive disorder and global mental health. *Indian J Psychiatry*. 2019;61(Suppl 1):S4–S8.

Stein DJ, Costa D, Lochner C, et al. Obsessive-compulsive disorder. *Nat Rev Dis Primers*. 2019;5:52.

Stein DJ, Fontenelle LF, Reed GM. Obsessive-compulsive and related disorders in ICD-11. *Braz J Psychiatry*. 2014;36(Suppl 1):1–2.

Stein DJ, Grant JE, Franklin ME, et al. Trichotillomania (hair pulling disorder), skin picking disorder, and stereotypic movement disorder: toward DSM-V. *Depress Anxiety*. 2010;27:611–626.

Strauss C, Hale L, Stobie B. A meta-analytic review of the relationship between family accommodation and OCD symptom severity. *J Anxiety Disord*. 2015;33:95–102.

Taylor S. Molecular genetics of obsessive-compulsive disorder: a comprehensive meta-analysis of genetic association studies. *Mol Psychiatry*. 2013;18:799–805.

Vicheva P, Butler M, Shotbolt P. Deep brain stimulation for obsessive-compulsive disorder: A systematic review of randomised controlled trials. *Neurosci Biobehav Rev*. 2020;109:129-138.

16

Pathological and Problem Gambling: Gambling Disorder

HARVEY ROY GREENBERG, M.D.

Pathological gambling has always been a characteristic of humans, a scourge in the shadows. Because of an unprecedented worldwide increase in gambling opportunities over decades, the illness can no longer be condemned nor ignored. It is a universal public health problem, often as crippling as schizophrenia, major affective illness, and anxiety disorders.

Mental health professionals have become more adept at diagnosing pathologic gambling. Unfortunately, efforts at increasing the public awareness of its ravages lag significantly behind the attention paid to other psychiatric problems which do not bear the unwarranted attainder of sinfulness.

One cannot precisely gauge the take from legitimate gambling in the United States. A reasonable guesstimate is approximately 150 billion dollars. The annual profit from American licit and illicit gambling enterprises is at least several hundred billion dollars, as much as 1% of the U.S. gross national product. Worldwide gambling profit is estimated at one-quarter to one-half trillion dollars.

The so-called brick-and-mortar gambling locales, casinos, bingo and videogame halls, etc., have spread from Biloxi to Buffalo, London to Las Vegas, and Moscow to Montreal.

Macao, Singapore, and Hong Kong are the Xanadus of today. Their pleasure domes dwarf Las Vegas' gaudy palaces.

A "whale" is casino slang for a multizillionaire gambler. Casinos make much of their money from whales, many of whom come from abroad. (Most high-end casinos have a private mini-casino for a "whale.") At sumptuous casinos within casinos, multibillionaire whales bet a developing nation's wealth on the turn of a card. Workaday employees are more enthralled by the weekly football pool than their jobs. The weekly church bingo night seems infinitely more exciting to some than the humdrum of their daily lives. Impoverished individuals may wager their last few dollars on ubiquitous state lotteries.

Recreational gamblers are more eager than ever to toss away dollars or euros, pounds or yen. Casino resort vacationers routinely figure acceptable losses into their holidays. But as surely as the house percentage destroys an ordinary roulette player's stake, the lives of a predictable percentage of gamblers will be ground down; careers ruined and family lives devastated, as they sink ever deeper into the idiosyncratic world of the pathologic gambler.

HISTORICAL BACKGROUND

David Schwartz's *Roll the Bones: The History of Gambling (Casino Edition)* comprises an engrossing account of the archeology, anthropology, and history of gambling. Schwartz describes the ubiquity of gambling in virtually every place and era from West to East, primeval to postmodern. One envisions Cro-Magnons wagering flints on how far Og or Magog could chuck his spear, or whether Og would beat Magog in a prehistoric wrestling match.

Some forms of gambling originally had a religious basis, evolving from divination pitched at sussing out the future intentions of the gods. One of the oldest of these is sortilege, interpreting the patterns in a random throw of sticks, stones, or bones. It was only a matter of time before a punter of yore conceived it would be entertaining and profitable to bet on the immediate outcome of rolling the bones, rather than the outcome of a war next year.

Astragali, the knucklebones of goats and sheep, often figured in the divination of antiquity. Shaved into cubes, marked with numbers or letters, they became forerunners of dice. Early examples have been exhumed from the tombs of Mesopotamia Ur and the palaces of Babylon.

The ancient Greeks and Romans were devotees of divining as well as extravagant gambling. Rolling the bones, once used to predict the future, now determined who went home broke at the end of the evening. Romans of every class—emperors and plebes—were ardent dicers. (Lest torture and execution follow, tyrants like Nero and Caligula were allowed to win large night after night.) Rabid fans also handicapped gladiatorial combats.

In the ancient East, punters wagered on combat between men or beasts, birds, and insects. Cockfighting was popular in India several millennia ago. In the gambling houses of dynastic Chinese towns, enthusiasts bet on matches between quails, fish, and crickets.

Board and tile games were popular in antiquity, played for pleasure or serious coin. Chinese gambling houses featured “mahjong-like games.” Pharaonic Egyptians enjoyed early versions of checkers and backgammon. Modern chess derives partly from the ancient Indian game of chaturanga, in which moves of the pieces were dictated by the throw of a die.

Gambling flourished in Europe throughout the medieval period and Renaissance. Crusaders played at hazard to pass tedious hours while laying siege to Arab castles, or staked fortune, limb, and life in savage jousting tournaments. Dicing was the most common gambling activity for the gentry and commoner of the early renaissance.

“Contests between humans, animals, and both were denounced by authority, but nevertheless drew crowds.” Bear baiting, forbidden in Elizabethan London, flourished just outside the city proper. It pleasured people of the better or lesser sort as much as the dramas staged in nearby theaters like Shakespeare’s Globe. Audiences of bloody tragedies like “Hamlet” could smell genuine blood in the air from the nearby bear pits.

In the first half of the 15th century, a new gambling craze swept across Asia and Europe—card playing. With more variations possible, card games quickly became dicing's chief rival. Cards also figured prominently in divination—and still are. The Tarot deck used for fortune-telling derives from the deck employed several hundred years ago in Italian card play.

Gambling historians believe that playing cards originated in China and Korea, then were brought to Europe by travelers, traders, sailors, and mountebanks. Florence became a major center for the manufacture and export of cards. The deck of the common player was crude and cheap. The nobility commissioned exquisitely crafted decks at extravagant cost.

Venice figured significantly in the history of Renaissance wagering. The first legitimate European casinos were founded there: a stiff percentage of their earnings went to the Doge and his cronies, one of the first examples in early modernity of a government-gambling connection.

Lotteries were even more likely to be joined to the political fortunes of the Italian city-states that sponsored them. Venetians, Genoese, and Romans were notable devotees (the Papacy-sanctioned lotteries in 1732). The fiscal debacles that helped precipitate the French revolution were in part stoked by the diversion of state wealth into the pockets of lottery profiteers (Voltaire included).

Probability theory was arguably first applied to gaming by Gerolamo Cardano, a 15th-century physician and pathologic gambler. Before Cardano, wins or losses were chalked up to the whims of fortune. With a mathematical grasp of odds, shrewder bets could be laid. Claudius' elaborate wager in the duel between Laertes and Hamlet is based on the assumption that Laertes is the better swordsman. "Hence have we odds," says the wily king, even though he has already plotted a lethal payoff for his nephew.

Blaise Pascal, the 17th-century philosopher, mathematician, and theologian, also studied probability theory. Pascal's wager famously proved that belief in God was the safest bet in the Heavenly House. Pascal is also credited with crafting a precursor of the modern roulette wheel, supposedly while trying to build a perpetual motion machine.

Roman soldiers allegedly had already invented a crude version of roulette by upending a chariot and wagering on the spin of its wheel. Although the attribution may be apocryphal, soldiers have always been ardent gamblers, a function of the perennial boredom of hurry up and wait military life.

During the Age of Enlightenment, gamblers throughout Europe continued to wager heavily against all reason. English gentlemen of the Enlightenment and Regency periods bet hugely on cards and dice as well as horse races, dog/rat fights, and savage bare-knuckle boxing contests.

Many a Lord lost his estate at the tables, then looked to marrying an heiress to mend his fortunes. Groundlings mostly wagered at dice. They also subscribed farthings to lotteries sponsored by the Crown to establish hospitals and charitable institutions. These honorable and occasionally rigged enterprises were the antecedents of the numbers racket and legitimate lottery play of today.

At least since the Renaissance, gambling condemned by authority morphed into respectable and powerful businesses. Once more, Venice, hub of international trade, became a singularly auspicious locale for gambling on or against the arrival of rich cargo ships from distant climes.

The venerable Lloyds of London was a coffeehouse-cum-bourse before it became a respectable insurance firm. Like the first casinos, early stock markets stood on equivocal moral and legal footing enormously lucrative for owners and the state, yet at the same time harshly condemned from the pulpit, by the press, and the courts for promoting gambling, as it were, by others means.

During the 18th and 19th centuries, elegant gambling halls were erected in spas frequented by the idle, where the pursuit of health served as a convenient pretext for wooing Lady Luck. The watering holes of Baden-Baden, Monte Carlo, Nice, and Bath were notorious for high-stakes play. Roulette was the most popular table game across the social spectrum; faro and variants of baccarat and blackjack were also played in the public halls. Separate private rooms were set aside for high-stake card games of whist and piquet.

The fortunes of the great spas eventually waned, although their vestiges still can be found in Nice and Monte Carlo. European public gambling severely declined after World War II because of postwar social and economic devastation. However, renewed prosperity quickly led to the return and profitability of gambling. As always, the authorities took their cut from the proceeds, even as they intermittently denounced wagering. With the rise of powerful international gaming corporations, legitimate gambling has achieved undreamed-of profit in Europe as well as Asia (*vide infra*).

Gambling in America

Long before Europeans set foot on the American continent, indigenous tribes were eagerly betting on games and sports. *Roll the Bones* contains an excellent discussion of early Indian gambling. Despite the Puritan ethic of the early New World settlers, gambling quickly became an omnipresent and predictably ambivalent presence in precolonial America.

The pervasive double standard perennially applied to gambling is piquantly illustrated by the gaming behavior of the Founding Fathers. George Washington played cards almost nightly during the darkest years of the American Revolution. He also issued an edict strictly forbidding wagering among his officers and troops, proclaiming that:

> . . . at this time of public distress, men must find enough to do in service to their God and their country without abandoning themselves to vice and immorality . . .

Luckily for the nation, Washington was a better general than gambler. His meticulous expense accounts reveal he was consistently down a few pounds every year, predominantly at whist.

During the 18th century, American gentry imitated the gambling habits of their English counterparts, wagering on card games, horse racing, boxing, and wrestling matches. Everyday Americans diced as busily as their English brethren. Bearbaiting and similar blood sports were not as popular as abroad, except for cockfighting.

American variations of European rummy- and bridge-like games emerged during the first half of the 19th century. The game known today as poker derived from European models. Craps, which had once been the rage of well-heeled Parisians after the French revolution, became hugely popular among average Americans. By the Civil War, it was the favorite dice game among soldiers on both sides of the conflict.

As the borders of the nation expanded, the gambler followed close upon the heels of the prospector, the homesteader, and the entrepreneur. Gambling flourished everywhere—in the marginal milieus of mining camps and cattle towns; on Mississippi riverboats; in bustling New Orleans (where Bourbon Street was rechristened the Rue De Craps); in the slums and salons of major cities; and in posh summer resorts where political and financial makers and shakers congregated.

American gambling legends spread. During a draw poker session, a known cheater declares he has beaten Wild Bill Hickock with a flush. But Bill calls a full house, turns over a pair of sixes and one ace. He draws his gun, says "Here's my six …," then plunges his knife into the table—"and here's my spike! (gambling lingo of that time for an ace)." The cheater turns pale and folds.

The professional gambler's persona acquired an ineluctable shimmer compounded out of a yeasty, penny-dreadful blend of romance and danger. Gambling idioms entered the vernacular, for example, passing the buck, from the practice of denoting the dealer in a poker game by sticking a buckhorn knife in the table before him.

By mid-19th century, gambling in the United States figured prominently in coalitions between criminal elements and political machines. In the 20th century, widespread corruption of police, judicial, and other government officials accompanied the consolidation of powerful empires built around gambling, prostitution, loan-sharking, and bootlegging after the Volstead Act was passed. The mob/government symbiosis especially flourished in Chicago, New York, Atlantic City, New Orleans, Little Rock, Phoenix, and Las Vegas.

Vegas was the brainchild of the notorious Bugsy Siegal, gunned down in the 1950s at the order of Mafia capos unhappy about his managerial skills. As the city grew, mob muscle power controlled casinos as well as illicit gambling, loan-sharking, and related criminal enterprises throughout the 1950s and early 1960s.

Eventually, the mob's iron control yielded to legitimate business interests. With the bad old Ratpack days gone, casinos on the Las Vegas Strip and elsewhere reinvented themselves into family entertainment centers. Middle- and working-class patrons with relatively modest budgets became prime targets of opportunity. Conventioneers replaced the high-rolling "whales" of the past. Major money began to be reaped from slot machines rather than table games.

At this writing, family fun still exists in Las Vegas. But one also encounters a sanitized version of the mobbed-up 50s, catering to a younger clientele with money to burn. In garish, theme-based resort/gambling complexes like the Venetian and Caesar's Palace, tourists of every stripe enjoy carefully marketed wretched excess; ever more elaborate slot-gaming technology, celebrity/Cirque de Soleil megashows, and dancing clubs. Five-star restaurants rub elbows with cheap noodle bars and Burger Kings. The heady promise of unzipped sexuality pervades the scene.

The Las Vegas model has spread to casino/resort complexes elsewhere in America, Asia, and throughout the world. The multinational gaming corporations which run them have branched out into running film and record companies.

Since the 1990s, brick-and-mortar casinos have been challenged by the internet's virtual casino, with its ever-expanding cornucopia of gaming websites. Internet poker garnered the greatest attention in America until it was banned several years ago. It is now being slowly relegalized, probably because of its potential for generating tax income for states in straightened circumstances. Meanwhile, internet sport and poker gambling continue to thrive worldwide, available at the tap of a computer key.

NOSOLOGIC CONSIDERATIONS

Pathologic gambling was designated as a separate diagnostic entity in the third edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III). Parameters of the disorder were altered in the revision of the DSM-III, reformulated yet again for the fourth edition (DSM-IV), and were retained essentially unchanged in the revision of that edition (DSM-IV-TR). The illness continued to be subsumed in the category of impulse control disorders: not otherwise specified (which also included pyromania, kleptomania, trichotillomania, etc.).

The DSM-5 and DSM-5-TR

Substantive laboratory and clinical research over the past decade have definitively shown that pathologic gambling should be considered an addiction, rather than an impulse control disorder or obsessional variant. Accordingly, the DSM-5 removed it from the group of impulse control disorders, not otherwise specified, and situated it in a new category of non–substance-related disorders. This was retained in the DSM-5-TR.

The DSM-5 also replaced pathologic/problem gambling with gambling disorder. The renamed diagnosis has a tame, homogenized ring. It does not intimate at first glance the devastating nature of the illness, as it should for the unknowledgeable reader. Pathologic gambling disorder or problem/pathologic gambling disorder would have been more appropriate if there must be a disorder at all. Problem gambling is also not mentioned as such in the DSM-5-TR's gambling disorder.

Most of the criteria for gambling disorder essentially replicate earlier criteria for pathologic gambling with various emendations. Some of these seem rather notional, and there are a few significant omissions.

To meet the diagnosis of any subcategory of gambling disorder, four of the following nine criteria must exist for 12 months:

1. Wagering increasing amounts of money in order to achieve the peculiar excitement of the illness.
2. Restlessness and irritability accompanying efforts to decrease wagering, or abstain from it altogether.
3. Repeated unsuccessful efforts to control, diminish, or stop gambling.
4. An overwhelming preoccupation with past and future actual gambling activities and gambling-related endeavors such as handicapping a horse race.
5. Gambling in the context of dysphoric affect, anxiety, depression, etc. (One submits that the causes of dysphoria are not cited at sufficient length, nor is it sufficiently emphasized that dysphoria may be a product of pathologic gambling itself.)
6. Attempts to get even by repeatedly returning to gamble (chasing losses).
7. Lying to significant others in order to conceal the extent of gambling activity.
8. Gambling-related jeopardizing or loss of significant family and social relationships, jobs, and educational/career opportunities.
9. Escalation of the search for sources of money to pay off gambling debts. Loan-sharking is not specifically mentioned.

The DSM-5-TR indicates that a diagnosis of gambling disorder cannot be made if excessive wagering is better explained by a manic episode. (Other disorders and conditions which can precipitate pathologic gambling are not cited, e.g., delusional schizophrenia, or the administration of dopamine agonists in the treatment of Parkinson disease.)

The DSM-5-TR distinguishes between episodic and persistent forms of gambling disorder. (The factors which precipitate relapse are not cited.) The time necessary for achieving early or sustained remission is notionally set at 3 and 12 months, respectively. The distinction between mild, moderate, and severe types of gambling disorder is based on how many of the nine essential criteria are met at any given time in a patient’s course.

Reflecting one's position on appropriate diagnosis, pathologic gambling is used below instead of gambling disorder in discussing the illness' severest presentation. (The majority of clinical studies at this time still employ pathologic gambling [PG].) The abbreviation is frequently used in clinical practice and research. Problem gambling is also employed rather than the mild subset of the DSM-5-TR gambling disorder which is characterized by presenting fewer of the criteria.

EPIDEMIOLOGY

Evidence-based research across the world consistently indicates that approximately 2% to 3% of any population are problem gamblers who do not meet diagnostic criteria for full-fledged pathologic gambling; 0.5% to 1% of any population do meet those criteria. Some problem gamblers remain stable although compromised throughout their lives; others shift back and forth between problem and pathologic gambling or slide permanently into pathologic gambling.

A definite connection has been established between new casino construction and the rise of problem/pathologic gambling in the surrounding area, which did not exist before. Over 3 to 5 years' postconstruction, the prevalence rate then increases to approximately 3%.

The escalation of casino and internet gambling over the past several decades has been paralleled by an increase in nonpathologic and problem/pathologic gambling. In this context, it is inevitable that many individuals with no previous history of problem/pathologic gambling will have their latent disorder activated. (Quite possibly, these subjects were involved in some other form of addictive activity before gambling became available.)

The typical patient in earlier studies was a middle-aged White male from a comfortable economic background. However, reliable comprehensive surveys of treatment as well as nontreatment populations indicate that problem/pathologic gambling now cuts across every ethnic, class, age, and occupational divide. (According to anonymous sources within the casino world, physicians are among their most consistent heavy players and losers.)

The prevalence of pathologic gambling has especially increased in the following groups over the past two decades:

1. The poor: Particularly poor minorities dwelling in urban ghettos.
2. Adolescents and young adults: This group notably includes teenage internet poker players, who can gain access to the virtual tables without identification. In general, peer pressure to “get with it and join the game” plays a major role in adolescent gambling of any sort. Male teenage gamblers predominate significantly over females. Adolescent gamblers are particularly likely to have parents, grandparents, and other relatives with serious gambling pathology. Many preadolescent gamblers also bet on internet poker, Magic Cards, fantasy sports contests, and videogames without a cash payoff. These young gamers are thought to be at risk for developing gambling difficulties later on.
3. The elderly: Particularly retirees with an inadequate social network, who live alone, have limited recreational or educational interests, or are homebound because of health problems. Problem drinking or frank alcoholism frequently accompanies gambling problems in this cohort, often with no prior history of alcohol abuse or gambling difficulties. “Day Trip” gambling by seniors in assisted living and nursing home facilities is often encouraged by nearby casinos and slot parlors. Inevitably, some day trippers become pathologic gamblers.
4. Women: Several decades ago, male pathological gamblers were said to outnumber females by 5 to 1. Recent studies indicate a shift toward gender equality. Although evidence-based study is limited, it is postulated that more women gamble today because of their increased presence in the workplace, with more ready cash at hand.

   Women formerly favored bingo, lotto, and slot machine play, but now participate in every wagering activity. Most women begin intemperate wagering later than men but are quicker to become full-fledged pathological gamblers. Gambling-related suicidality is higher in this group.

5. Prisoners: Studies indicate that approximately one-third of American prison inmates have formidable gambling issues. Many bring a previous gambling problem to incarceration. But some gambling begins in prison as an endemic solution to the tedium of prison life. Offenders with no previous difficulties often become pathologically involved with card games like poker, hearts, backgammon, or chess. Cigarettes, drugs, privileges, and cash comprise the common currency of prison gambling. Loan-sharking is common. Failure to pay debts can exact brutal punishment, even death.

   Prisoners with gambling problems frequently suffer from cross-addiction to alcohol and other substances, as well as other serious psychiatric disorders. Unfortunately, prison systems rarely have treatment programs for pathological gamblers, compared with programs for substance and alcohol addiction.

FAMILY HISTORY

The family histories of pathologic gamblers reveal that one or more parents, grandparents, or significant relatives are heavy or frankly pathologic gamblers. Family members also have increased rates of affective illness—notably major depressive and bipolar disorder; substance abuse, particularly of alcohol, cocaine, and nicotine; and also experienced verbal and sexual abuse in childhood. Attention deficit hyperactivity disorder (ADHD) is supposedly found more often in family members with pathological gambling, but to date, there is no hard research on this score.

One conceives pathologic gambling as the dark underbelly of the American dream, for the patholoogical gambler’s family of origin typically prizes competition, is highly materialistic, and has little inclination to save or budget. Intense admiration for wealth and its signatures is common. Frequently a parent or grandparent has become well-to-do through hard work. Sadly, the same individual may have gambled away his business.

CLINICAL FEATURES

Lesieur's description of the pathologic gambler's signs, symptoms, and life trajectory in 1977 is as relevant today. His typical subject was a middle-class married male in his 30s or 40s. This type is still quite common, but the clinician will encounter many similarities, differences, and variations of the model as a function of age, gender, ethnicity, and social class.

The pathological gambler's intense preoccupation with betting usually begins in adolescence but can emerge even in the preteen years. In his youth, he acquires considerable gambling skills, often in a number of activities. A patholoogical gambler often experiences exhilarating big wins early in his career, which augments an ineffable sense of specialness.

Because of his desire to make it big, he commonly drops out of high school or college and moves away from home. Frequent gambling continues into adulthood but is reasonably well integrated into his life. Most pathological gamblers marry, raise children, work hard, often achieve considerable success, and live large. Contrary to a popular image of the pathological gambler as a callous, entitled narcissist, he is usually deeply committed to family life and takes great pride in being a good provider and parent.

At some point, the pathological gambler becomes embroiled in an ever-escalating, ever-tightening spiral of gambling involvements and options. He may start losing in the context of life stress; by wagering in new locales such as casinos and the internet; or for no discernible reason whatsoever. Instead of stopping and cutting losses, which he has often been able to do before, he begins to chase, calling on more and more options to recoup mounting losses.

The pathological gambler taps into sources of cash for gambling such as family checking and savings accounts; piggy bank and rainy-day stashes; his personal or workplace business; legal and illegal lending institutions; credit cards; and fellow gamblers.

As ever more options are used, they are also used up. The spiral tightens, until fewer and fewer options remain. The pathological gambler becomes trapped in a formidably powerful, self-enclosed system which ensures his inevitable downfall. He bets ever-larger sums and becomes embroiled in more types of gambling for longer periods of time. Greater losses occur because of the grinding down of natural and house odds, as well as uncharacteristic, impulsive, foolish play.

Most pathologic gamblers are highly knowledgeable. But their expertise steadily crumbles as loss accrues. Indebtedness is compounded by the exorbitant interest rates (vigorish or the vig) on loans floated by sharks and credit card advances. When every option is exhausted, and every source is dried up, the pathological gambler reaches the disastrous moment of closure.

Crimes such as embezzlement, passing bad checks, and other types of fraud may be undertaken to get even. Female pathological gamblers may become prostitutes. Such offenses were once believed to be hallmarks of end-stage pathologic gambling. Recent studies indicate that petty criminality often begins earlier and continues throughout the patient's career. However, pathologic gamblers without genuine antisocial or sociopathic traits rarely become involved in major criminal enterprises.

The revelation of the pathological gambler's transgressions is excruciatingly humiliating, and the consequences are often catastrophic. Exposure may precipitate imprisonment, professional disgrace, divorce, escape to another state or country, or the ultimate deliverance—suicide.

But it is more likely for exposure to be followed by a favorable outcome with restitution of debts undertaken by family or friends; remission of debts by a bookmaker, bank, or loan shark; an employer's forgiveness; a kind judge who orders probation rather than incarceration. Typically, the pardoner is deeply moved by one of the innumerable “sad tales” every pathological gambler always has at their beck and call.

The craving for action is central to the pathological gambler's modus vivendi. A pathological gambler frequently rationalizes this overwhelming desire by convincing himself that one simply must be wherever the action takes in order to catch up and stop chasing. But it is the pull of action itself that draws the pathologic gambler ever deeper into the gambling vortex. It matters little whether he is losing or catches up. The ineluctable lure of action is captured by Dostoyevsky’s mot “le jeu pour le jeur” (“the game for its own sake”).

Trippett observes that:

> Supposedly . . . action speaks to the process of betting. I suspect it speaks more deeply to what happens within the gambler. He places the bet, juices flow, he really feels . . . alive. Action: when the bet is on, his existence is confirmed . . .

The idiosyncratic action rush is inflected by the peculiar ambiance of each gambling locale; the racetrack’s steamy milieu, polluted by rumor and false report; the clockless eternal present of the casino, with its idiosyncratic symphony of sounds—the click of a roulette wheel, the mellow warbling of slot machines, the joyful uproar around a craps table when a winner on a roll makes his number again.

The action high can be obtained in so-called action systems: a pool hall; bowling alley; the backroom of a candy store; and a diner—where bookmakers, pros, and recreational and pathologic gamblers congregate. Here every kind of play is found or can be arranged.

Some PGs stay wedded to the game they have always played as they become enmeshed. However, a chasing pathological gambler usually finds that one kind of gambling is insufficient to satisfy his craving. So he branches out into other gambling activities. A lack of expertise in a new game often compounds losses, accelerating the chase.

As the spiral deepens, the trajectory of action inevitably extends beyond the time consumed in betting to embrace longer time before and after wagering—planning action; setting up a poker game; handicapping races at several tracks during a single day; getting to action as in leaving work prematurely to get to a racetrack in time for the daily double; dealing with the consequences of action—collecting winnings or, more often, dashing madly about to cover losses; making interminable excuses to a bookie for coming up short; to an employer for lateness and absences; to his wife for coming home at 4 AM only to leave a few hours later.

The pathologic gambler is a master at manipulating cash flow—moving money. His previous business skills are now directed at the business of enabling play. He shuffles loans, bank accounts, and cash, from one source to another in dazzling displays of financial legerdemain which would do a Ponzi proud. His hectic strategies in moving money actually become an integral ingredient of the ominous, raw excitement of the action-seeking lifestyle.

The pathological gambler in serious chase mode is like a juggler, desperately fielding an ever-increasing number of plates, oranges, and bowling clubs tossed at him from offstage. Like his psychological twin, the adolescent, he is supremely engaged in an eternal present. A central cause of mounting losses is his utter incapacity to envision a future more than a few days away, a time when he won’t be able to secure the cash to forestall closure and exposure.

The pathological gambler’s trajectory over years is consistently marked by repeated cycles of spiraling, closure, exposure, bailout, and relapse. Relapse is often precipitated by the prompting of fellow pathological gamblers who cannot abide the presence of an abstainer; by the intense boredom accompanying abstinence; or by the very fact that the problems precipitated by closure have been resolved. As the horror of indebtedness fades from memory—the memory for gambling disasters is notably short—the pursuit of action is all too easily resumed.

Pathologic gambling occurs in companionable social circumstances at a blackjack or craps or poker table, at a horse or dog track, or football game. Most pathological gamblers are outwardly gregarious, with an apparently wide network of relationships including other gamblers. These often have a curious ephemeral quality. For several hours of poker play, the man beside you is your best friend before he gets up and vanishes forever into the night. The decision to gamble with fellow pathological gamblers at the expense of the sufferer’s family and social life is an especially significant choice in the course of the illness.

A lone-wolf pathological gambler is rare. In aid of rationalizing schizoid tendencies, he will often assert that he shuns fellow gamblers to avoid bad advice or loaning hard-won cash. Most internet gamblers are not lone wolves. Computer play commonly alternates with social gambling. However, when internet waging creates the pathologic spiral in one so predisposed, outside wagering activities along with other social outlets fall away.

The induction of a gambling spiral has been ascribed *inter alia* to marital problems, death of a loved one, the birth of a child, and business reverses or successes. However, on closer examination, these may not really be the actual causes of chasing but *post hoc propter hoc* events. Marital problems may be the consequence, not the precipitant of pathologic gambling.

Whatever stressor precipitates the chase, the pathological gambler becomes an extraordinary source of stress for others, at home and at work. For many, the hallmark of the illness' increasing severity is not simply the time devoted to gambling, or the preoccupation with gambling, or mounting indebtedness—all wrenching enough. It is rather the devastating impact of pathologic gambling upon every relationship outside gambling.

As the chase escalates, the pathological gambler's quality of life progressively deteriorates. For years, spouse and family may have been totally unaware of the extent and severity of gambling, until indebtedness drives the patient to invade family funds. Exposure is especially calamitous for the pathological gambler's spouse, shaking confidence in the pathological gambler's successful image and evoking angry feelings of betrayal.

Exposure regularly is followed by impassioned promises to cease gambling, leading to tearful reconciliation. But as the pathological gambler continues repeated chasing cycles and deceitful tactics to avoid the spouse's complaints, an atmosphere of pervasive mistrust comes to pervade the marriage. It degenerates into squalid ploys and counterploys revolving around the pathological gambler's concealment strategies, and the spouse's maneuvers to protect themself and the children.

Despite suffering, the pathological gambler's spouse does not easily oppose the pathological gambler. Once forged, the dysfunctional marital bind may continue for years before the spouse issues a final ultimatum to cease or face divorce. Research does not support the common belief that spouses of pathological gamblers are inherently masochistic. Nor do they subtly encourage the pathological gambler's illness.

It is rather that a wife, husband, or lover becomes locked into a uniquely compelling interpersonal system that eventually justifies its own existence. The spouse genuinely cares for the pathological gambler and tends to deny or excuse gambling behavior, especially with the pathological gambler's convincing expressions of guilt. Even when a spouse winds up loathing the pathological gambler, they may be unable to break free of the system because they fear being unable to support themself, or realistically cannot do so (one notes the similarity to the battered spouse's plight).

The pathologic gambler's children are tragic victims of the illness. They bear the burden of neglectfulness, of parental disharmony, and of actual abuse. Predictably they have a high rate of psychiatric symptoms and behavioral problems. By that poignant process through which one identifies with a traumatizing parent, they also stand at considerable risk of becoming pathological gamblers themselves.

The crumbling of the pathological gambler's value system is as traumatic to themself as others. The pathological gambler has always treasured honor; loathes themself for abrogating it. Like addicts in the throes of their illness, the pathological gambler becomes demoralized and exhibits ego-alien and indeed abhorrent behavior, which would never exist were they not caught up in the chase. In end-stage gambling, the pathologic gambler's entire existence revolves around play. At this point, the Pathological gambler may paradoxically derive little real pleasure from gambling. Winning has long since ceased to be an end, rather the means of securing further action.

Some Pathological gamblers cease serious gambling in later life. Cessation does not appear related to treatment of any sort. Akin to the burn out of drug addicts, the craving for action seems to simply wither away for obscure reasons. Mutatis mutandis, a Pathological gambler may escalate chasing in old age, as gambling becomes life's raison d'être.

Personality Characteristics

Pathologic gamblers display a strikingly similar cluster of personality traits. Depending on age, gender, and social class, some may be exaggerated or absent.

In pure culture, pathologic gamblers are competitive, going back as far back as latency or early adolescence. They are intelligent, although usually not deeply intellectual.

They consider themselves rugged individualists; they are highly independent and overconfident to the point of abrasive. A keen interest in sports participation and viewing parallels their interest in gambling. They are perennial optimists and deniers, and they possess a distinctive gallows humor:

Joe: Which fighter are you betting on tonight?
Jim: Both.
Joe: But you'll only come out even.
Jim: I need the money!!!

Pathological gamblers are by turns touchingly loyal, or gaggingly insensitive, notably when chasing. Despite their apparent sociability, they are often beset by profound feelings of loneliness. They do not easily express feelings, indeed may be so unable to get in touch with their emotions as to be alexithymic.

Pathological gamblers as well as successful professional gamblers often project an air of grandiose elitism. They believe they always know the score at gambling and life in general, for which they hold gambling as a Hobbesian dog-eat-dog metaphor. The pride of pathologic gamblers in their idiosyncratic vocation, even at its disastrous end-stage, is poignant.

COMORBIDITY

Extensive research indicates that impressive comorbidity exists between gambling and affective disorders, notably major depression and bipolarity, and addictive disorders, notably to alcohol and cocaine, nicotine, and caffeine.

There is significant comorbidity between pathologic gambling and ADHD in adolescent as well as adult problem/pathologic gamblers. One has the sense that many pathological gamblers have a forme fruste of full-blown ADHD.

Various personality disorders have been described in PGs, especially narcissistic, borderline, and antisocial disorders. Although many pathologic gamblers have obsessive personality traits, full-blown obsessive-compulsive disorder (OCD) is rare.

Mutatis mutandis, a higher rate of pathologic gambling has been discovered in patients with a primary diagnosis of affective illness (particularly bipolar) and schizophrenia. One’s clinical impression is that patients with generalized anxiety disorder, panic disorder; narcissistic, borderline, and antisocial personality disorders are more likely to have gambling problems. However, the relationship has yet to be proven.

The pathologic gambler’s course is often worsened by serious comorbid conditions, particularly alcohol and drug addiction. The easy accessibility of these substances in or near gambling milieus further disinhibits the Pathological gambler, resulting in wilder play and greater losses.

Exacerbation of an affective disorder may be a primary precipitant of a gambling binge. Heavy indebtedness may precipitate a dysthymic state or frank major depression, which then triggers off more betting binges.

Pathologic gamblers paradoxically are often both hypochondriacal and dreadfully neglectful of their physical well-being. Pathological gamblers eat poorly and excessively; are chronically sleep deprived and underexercised. They are more vulnerable to hypercholesterolemia, arteriosclerotic cardiovascular disease, diabetes, peptic/gastric ulcer, and hepatic and pulmonary disorders (often a function of nicotine abuse).

DIFFERENTIAL DIAGNOSIS

Making a diagnosis of pathologic gambling is often problematic because pathological gamblers so often minimize, rationalize, or absolutely deny their difficulties. Should doubts arise, the clinical interview may be supplemented by reasonably reliable tests such as the South Oaks Gambling Screen (SOGS) and the National Opinion Research Center DSM Screen for Gambling Problems (NODS). SOGS is better known and frequently used in pathological gambler research. NODS is a population-based telephone interview that identifies gambling problems according to DSM criteria. Although initially useful, an extensive clinical interview obviously provides the most accurate information for the diagnosis to be made.

A problem gambler is most likely to be seen first by a clergyperson or general physician. Referral to a psychiatrist or other mental health practitioner usually occurs further down the line. Because of the patient's intense denial, both laypeople and practitioners need to recognize the symptoms and behavior that will differentiate problem/pathologic gambling from the recreational betting practiced by at least 85% of Americans.

Ordinary recreational wagering is based on a mix of conscious and unconscious factors unique to each unproblematic bettor, for example:

1. A desire for relaxation in a companionable social setting, away from the everyday grind. Recreational gamblers usually play on designated occasions such as a weekly poker game. They have a firm grip on acceptable losses.
2. The profit motive. One often hears players proclaim they don't care if they win or lose, it is the fun that counts. If pressed, however, most recreational gamblers will admit that they would rather win than lose.
3. The inherent pleasure of exercising a variety of skills and ego functions depending on the nature of the game or sport.
4. The satisfaction of nonpathologic competitive drives.
5. The attraction of risk taking in a betting situation where a negative outcome produces minimal if any psychic trauma. In a quixotic world where mortality is the only certainty, yet the time of

its coming ever in doubt, gambling allows one to indulge in the illusions of omnipotent control over destiny.

6. The impression of control is fostered by games in which the chance element can be reduced by studying the parameters of play and doping out the odds, for example, racetrack handicapping. When loss does occur, one is gently reacquainted with the limitation of one's powers.
7. The gratification of other obscure, possibly conflict-based drives and motives stemming from early stages of psychological development, etc.

For the professional gambler, wagering is a business, plain and simple. Contrary to the myth of the glamorous high roller, most pros avoid the limelight if only to avoid taxes. Their expertise is impressive. They are summarily prepared to take requisite risks; coolly acknowledge the losing streaks which come with the territory; and are able to weather losses on any given day without losing control. However, given the inherent nature of the work and personality factors, a pro is much more likely to slip into pathological gambler chase mode than the ordinary gambler.

Acute substance abuse accompanied by temporary disinhibition may precipitate a gambling bout without other symptoms of pathologic gambling. Alcohol and cocaine are particular offenders in this regard.

Gambling binges are occasionally seen in the setting of an acute or chronic schizophrenic delusional system, for example, a patient who believed God had ordered him to stake his life savings at the dog track so as to hasten the end of days foretold in Revelations.

Excessive wagering may occur during the manic stage of a bipolar disorder in patients who have been unproblematic gamblers before or have never gambled at all. The correlation between gambling episodes and mood swings in pathologic gamblers with significant comorbid affective illness is a vexed question—obvious in some pathological gamblers, either nonexistent or obscure in others.

Bipolar pathological gamblers with mania in remission nevertheless can appear as wired to the universe as a patient in a full-blown manic state; crackling with tension as they pursue their dizzy dance over the abyss. However, in the former case, the pathological gambler quickly returns to a more temperate mood once play has ceased. He may be reactively and temporarily elated after a win, or just as appropriately depressed after a loss. In any case, a mood-disordered problem/pathologic gambler will generally wager excessively whether clinically manic, depressed, or euthymic.

Individuals with antisocial personality disorder frequently gamble, but the nature of their behavior reflects their essential criminality and utter lack of empathy. They will do anything possible to win and are chronic cheaters. They are often poor players and blame others for their losses rather than recognize their own inadequate play. An antisocial individual may retaliate violently after a loss; indeed, may subsequently rob an individual to whom he lost, or hold up a game where he lost.

Unlike antisocial gamblers, the majority of pathologic gamblers have a better work record, a more stable family life, and a reasonably scrupulous moral set before their illness catches up with them. They are ashamed when driven to crime—usually petty—out of profound desperation, to cover losses and keep gambling.

Although gambling may seem to have an obsessive element, it is rare for an individual with obsessive-compulsive illness to become a pathologic gambler. Unlike the tormenting ego-alien symptoms of an OCD patient, the pathologic gambler’s overwhelming preoccupation with play is usually quite ego-syntonic, even persisting when the patient has reached a terminal, end stage of gambling.

Bouts of pathologic gambling are rarely due to compromise of cerebral function by injury, infection, or neoplasm. Intemperate wagering in patients with Parkinson disease treated with dopamine agonists is further discussed below.

ETIOLOGY

For centuries, pathologic gambling was chiefly viewed as a moral stain rather than a medical problem. Gamblers were said to suffer from an engrained weakness of character that rendered them susceptible to temptation from the dark side. The evils of gambling were associated with the evils of drink, sexual depravity, and low companions.

Pseudopsychological theories surfaced in the 19th century, ridden with sanctimonious moralizing. The advent of psychoanalysis spurred awareness that pathologic gambling might have a scientific, rather than a "weak-brain" or theologic basis. During the last two decades, exploration of the pathological gambler's psyche has been joined by impressive research into the illness' genetics and neuroscience.

Psychoanalytic Theories

The literature on pathologic gambling was sparse at the advent of psychoanalysis. Theories chiefly addressed the gambler's childhood trauma and psychic complexes.

Early analytic investigators drew analogies between pathologic gambling and sexual excitement. It was commonly supposed that compulsive wagering comprised some sort of neurotic substitute for mature erotic fulfillment.

However, players state, and researchers have affirmed, that the unique thrill of gambling action is utterly divorced from Eros. Nell Kimball, a madam of the frontier days whose establishment provided prostitutes and a gambling house declared that "if you give a man a choice between gambling and sex, he will take gambling any time." The majority of pathologic gamblers are fully capable of enjoying sex. All in all, the displacement of libido from bedroom to bookie must be rated a very vexed question.

The first psychoanalytic theories now appear simplistic or simply wrongheaded, but the findings of depth psychology over a century should not be dismissed out of hand. While analytically oriented therapy without other adjuvants will not benefit most pathological gamblers, it does yield helpful insights into the gambler's inner life and motivation, his vertiginous fluctuations between exhilaration and despair.

Simmel published the first account of a pathologic gambler's analysis in 1920. He thought that pregenital and Oedipal conflicts were implicated in the illness. Stekel described the gambler's frequent mood swings, preoccupation with superstition and rituals, and fantasies of omnipotence. He classified the pathological gambler as a compulsive neurotic with latent homosexual tendencies.

In his famous 1928 monograph, "Dostoevsky and Parricide," Freud likened pathologic gambling to other obsessive-compulsive behaviors. He argued that the great writer's gambling addiction stemmed from a displacement of childhood masturbatory urges. The frenetic drivenness, escalating tension, and use of the hands in roulette and other games were autoerotic equivalents.

Fenischel astutely compared the gambler's cycles of triumphant wins and desolating losses with manic-depressive cycling between giddy elation and devastating despair. He speculated that pathologic gamblers believe they are entitled to special protections and gratifications, based on grandiose infantile expectations of parental bounty. Gambling is the magical means by which fate is compelled to act on the patient's behalf. In a variation of the *Schicksalneurosis,* the destiny neurosis, the pathologic gambler surrenders autonomy and responsibility for his actions, perceiving himself either as destiny's favorite or victim, depending upon whether he is winning or losing.

Kris theorized that the pleasure of the pathological gambler's wagering articulates with guilt-ridden Oedipal sexual tensions so that the heady joy of action becomes ineluctably fused with dread. The patient's ego is unable to control what it has initiated. He becomes trapped in a vicious cycle of anxiety: the need for reassurance of his survival through winning; more play; mounting losses; and the ominous escalation of tension. Gambling evolves from an innocent pastime into a literal matter of life and death.

Edmund Bergler's theories about pathologic gambling were immensely influential for decades. His signal assumption was that the pathologic gambler is driven by a desire to lose. Bergler asserts that the pathological gambler bears profound grudges, first against parents, then later authority figures who are resented for imposing the reality principle on his infantile narcissism and omnipotent fantasies.

According to Bergler, aggressive impulses unconsciously directed against the misperceived deprivers of nursery days cause inordinate guilt and an intense need for punishment. Gambling recapitulates the cycle of childhood rage, aggression, and remorse. Pleasure in wagering reflects the exhilaration of grandiose rebellion; pain arises from the morbid certainty of punishment.

As the pathologic gambler’s anguish about losses becomes erotized, he develops psychic masochism: savoring his victimhood, relishing injustice collecting, reveling in self-pity, and spurious righteous indignation. Lady Luck at base is his Bad Mother. He hopes his suffering will compel her to restore the blissful infantile nurturing she snatched from him.

While Bergler’s ideas are now viewed as at least questionable, they nevertheless have provided grounds for more sophisticated psychoanalytic study.

Whatever the pathologic gambler’s other key conflicts, one speculates that the pathological gambler’s devotion to wagering constitutes an uncanny spilt religion. His plethora of superstitions and rituals seem pitched at winning Fate over to his side, in aid of dealing with a devastating fear of death which has largely gone unappreciated in the literature. One wonders if the pathological gambler may court the erratic favors of Lady Luck in aid of making the ultimate score—a triumph over death itself.

MATERIALIST RESEARCH AND HYPOTHESES

Imaging Studies

More questions have been raised than answered about the neurobiology of the pathologic gambler. Nevertheless, researchers are beginning to make progress in mapping cerebral centers and pathways that mediate judgment, mood regulation, decision making, and impulse control. These are compromised in pathologic gambling, as they so often are in substance addictions.

Patients with cerebral trauma, tumors, or infectious processes that damage the ventromedial prefrontal cortex and amygdala *inter alia* may exhibit denial or unawareness of situations that require common sense judgments. Actions with immediate rewards are favored, while long-term negative consequences are ignored. Experts speculate that pathological gamblers may have some sort of analogous cerebral compromise which facilitates immoderate betting activity.

Many imaging studies have focused on the ventral and ventrolateral striatal prefrontal cortex and connections with other brain loci mediating prudent judgment. Experts speculate that suppression of activity within these pathways may cause a blunted neurophysiologic response to rewards and losses, in turn eroding the ability to make appropriate decisions. Pathological gamblers, like substance abusers, regularly seek short-term gratification without regard for later consequence.

The role of visual cueing in pathologic gambling has been studied with various imaging techniques. Pathological gamblers viewing videos of gaming scenarios demonstrate heightened activity in pathways between *inter alia* the right dorsolateral prefrontal cortex, right parahippocampal gyrus, and the left occipital cortex. Corollary conscious cravings to gamble were experienced by subjects after exposure to the gambling videos.

The impact of risk taking upon the pathological gambler’s brain has been demonstrated in PET research, in which subjects playing for real rather than simulated money showed an increased metabolic rate in the primary visual cortex, cingulate gyrus, putamen, and various prefrontal areas.

These findings are intriguing, but only scratch the surface. Any episode of recreational as well as pathologic gambling involves many other activities than betting. A bout includes arousal of the urge to gamble; planning the time, place, and choice of gambling activity; experiencing the ebb and flow of action, for example, at the racetrack or card table; dealing with the outcome of winning or losing (e.g., moving money to cover losses). Investigators are still far from being able to monitor what precisely the pathological gambler’s brain is doing, which cerebral loci and pathways drive the doing, over the entire trajectory of a gambling episode.

Neurotransmitter Research

Many recent studies indicate that pathological gamblers suffer from complex neurotransmitter dysregulation similar to the abnormalities found in substance abusers and patients with other behavioral/impulsive problems. Imbalances across the entire neurotransmitter spectrum—serotonergic, noradrenergic, dopaminergic, GABA, etc.—have been variously postulated as significant factors in the dysfunction of cerebral sites and pathways mediating behavioral arousal, inhibition/disinhibition, reward/reinforcement, and psychophysiological stress.

The particular focus on the role of increased dopaminergic activity in pathologic gambling arose from the serendipitous discovery that approximately one-third of patients being treated for Parkinson disease with dopamine agonists develop the illness without any previous history. The gambling excesses of these patients usually disappear upon withdrawal of the offending drug.

Pathologic gambling in the Parkinsonian group treated with dopamine agonists has been ascribed *inter alia* to disproportionate stimulation of dopamine receptors localized in mesolimbic pathways, which are thought to be involved in reward seeking.

However, one notes that two-thirds of the Parkinsonian patients studied do not develop pathologic gambling, indicating that other factors must play a role in the afflicted third: pre-existing cerebral damage due to Parkinson disease itself, genetic, family, and sociocultural influences, etc.

Other Predisposing Materialist Factors

Adolescents, particularly male teenagers, have a higher incidence of impulsivity and stimulus seeking than other age groups, as well as a higher incidence of recreational gambling, arguably attributable to the cerebral and hormonal changes of that period. While it has not been established that average teenagers are more predisposed to pathologic gambling than adults, adolescents with ADHD are significantly comorbid for pathologic gambling and vice versa.

A small study of adult nongambling subjects exposed to high- and low-risk gambling experiments showed that those preferring stimulus-rich environments consistently sought long shots, wagering more to win more. One wonders if such subjects have some sort of inborn predisposition toward stimulus seeking, which given other factors, might lead them to become more attracted to social, and even problematic, gambling.

Twin studies suggest that pathological gamblers, beyond identifying with a gambling relative, may have an inherited genetic potential to develop the illness.

Variations in several alleles have been discovered in pathological gamblers, but more research is needed to establish whether a predisposition for problem/pathologic gambling exists at a chromosomal level.

Behavioral/cognitive research focuses on pathologic gambling as a learned, disastrous habit. Based on the available research as well as one’s own clinical work, it appears more likely that the pathological gambler’s habit is driven by neurobiologic factors, rather than simple introjection of a gambling relative’s addiction. Hence the discussion of behavioral/cognitive factors has been placed in the category of materialist hypotheses.

Many pathologic gamblers exhibit sensitivity to a gamut of reinforcers directly or indirectly related to wagering. An example of direct or primary reinforcement is the slot machines’ lure of high return for a repetitious low-response small investment.

An example of indirect or secondary reinforcement is the “comp” offered by casinos, depending on hours of play, player’s wealth, etc. The comp can include a special room rate or no charge at all, free food, drinks, etc.

Contemporary casino design, technology, and management practices demonstrate an elegantly sinister deployment of virtually every behavioral stimulus for a vulnerable problem or pathologic gambler to bet over his head, not with it.

Pathologic gamblers are exceptionally prone to betting-related cognitive biases and distortions. These include discounting the large sum of money owed as a result of making small bets toward achieving the big score; magnifying one's own skills while minimizing the real talents of other players; losing sight of the amount of gambling time needed to recover losses; selectively remembering memories of wins over losses; the abiding conviction that a chain of losses must herald an impending win, especially after a near miss (the so-called gambler's fallacy, but actually one of many).

A Darwinian Sidebar

At first glance, gambling would seem to have little if any evolutionary value. If the impulse to gamble is inherently unhelpful for survival today, why has it not simply mutated away?

There is obvious survival value in prudently choosing to stay inside the cave while outside the winds rage and the beasts howl. *Arguendo* the chance of group survival was improved when some Og or Magog proposed a better, if more dangerous, plan to kill a mammoth in aid of providing better sustenance.

But what if the daring caveman subsequently began to wonder if a life better than bone-chilling subsistence lay over the next mountain, or beyond? It is not unreasonable to suppose that he, his children, or grandchildren would eventually decide to defy the prevailing belief that monstrous creatures or molten pits of fire lay over that mountain and seek a land where the weather was mild and food plentiful.

One conjectures that certain individuals have always been more genetically/neurobiologically hardwired, as well as characterologically predisposed, to appropriate risk taking, whether in geographical exploration, warfare, scientific research, finance, athletics, etc. Not only would these risk takers enjoy the prospect of their perilous venture's success, but they might also savor the peril *sui generis* to an appropriate extent.

Now, entertain the possibility that viable risk-taking turns pathologic for a small subset of the daring ones. In such cases, the risk inherent in a dangerous enterprise may become the sole end, rather than merely one aspect of attaining whatever achievement is being sought. Throughout history, perilous enterprises have suffered or even failed because a pathologic risk taker courted greater danger than warranted for its own sweet sake. (One would expect to find a wide range of behavior between appropriate and absolutely pathologic risk taking.)

Whyte acutely speculates that the appropriate risk taker is able to see a useful, albeit perilous pattern of opportunity in the environment, and takes prudent advantage of it, but an inappropriate risk taker may play down the pattern's inherent danger, indeed misperceive that one risk or another should never be taken, no matter how auspicious the intended outcome.

From this perspective, one conceptualizes gambling of any sort, recreational to pathologic, as a spin-off of the Cro-Magnon's hard-wired risk-taking behavior. With no mammoth to slaughter, with no journeying into frightening terra incognita, wagering at best provides a pleasurable anticipation in the moment, joined to the hope of a successful outcome; at the very worst, the pathologic gambler savors pleasure even when he knows the odds are stacked against him. Research cited above indicates that the pathological gambler's pleasure might actually be heightened by the heady promise of ultimate success, after heavy initial losses.

TREATMENT

Psychotherapy

Virtually every modality has been used to treat the pathologic gambler. For decades, results were ungratifying and relapse rates high. Better understanding of the disorder's psychology and neurobiology have improved outcome somewhat—this in the context that only 10% to 20% of pathological gamblers actually seek or are compelled to seek professional help. Success hinges upon carefully individualizing treatment, case by case. Combined approaches are frequently required. It is not unusual for a patient to need psychotherapy; cognitive/behavioral work; medication; couple and family therapy, and membership in a Gamblers Anonymous (GA) group. Recent data provide the most support for cognitive behavioral therapy (CBT) and motivational interviewing.

Pathological gamblers make extremely difficult patients. They rarely want help, persistently deny flagrant problems, and can charm the pants off an unwary therapist. The average pathologic gambler is often forced into treatment only after disclosure precipitates an ultimatum to get help or suffer divorce, prosecution, etc.

The pathological gambler's enormous pride, formidable denial, alexithymic lack of introspection, chronic impatience, and incurable optimism, all militate against forming a viable working alliance, particularly in psychoanalytically oriented therapy. These patients usually present a host of dire symptoms and problems related to the chase—depersonalization, hypochondriasis, panic attacks, substance abuse (notably of alcohol, cocaine, and tobacco)—which do not readily yield to talking approaches. The patient is also likely to suffer from other physical illnesses that have gone untreated or poorly attended to because of his own neglect—obesity, peptic ulcer, etc.

Pathologic gamblers expect magical interventions and frequently quit treatment when miracles are not forthcoming. They often go through the motions of patienthood, while remaining inwardly convinced that a turn in luck and improved cash flow are all that is needed to resolve their problem. They are usually scornful of someone who doesn't gamble. However, the therapist who is a gambling novice can put his ignorance to work by having a client teach him the ropes, much as an adolescent therapist will ask a stony silent youngster to instruct him about rap music.

Psychoanalytically oriented therapy with a pathologic gambler is stormy indeed, akin to working with a resistant adolescent. One routinely expects missed sessions; relapses; and financial crises which cause disruption or cessation of treatment due to inability to pay.

Firm, but kind limit setting is crucial. In most cases, the patient is asked to pay for each session on the spot, regardless of the cost. Greater responsibility in handling fees is often a sign of improvement. One has found that carrying a pathological gambler seriously undermines therapy, as does lowering a fee because of gambling losses unless serious physical illness supervenes, or the patient loses a job for reasons not related to gambling.

The therapist should be vigilant vis-à-vis the veracity of the patient's explanation about problematic treatment issues, as well as serious family and work difficulties. A pathological gambler in chase mode will minimize troubles or justify them with a "sad tale." Reality can be treated like toffee stretched here or there according to whatever will work. As in the treatment of seriously disturbed children and adolescents, it may be necessary to contact significant others to get a clearer picture of what is truly going on. The patient should be informed of this possibility at the beginning of treatment, with the understanding that the therapist would rather maintain confidentiality if the patient will hold up his end of the contract.

Cognitive and behavioral strategies appeal to the pathologic gambler's inherent pragmatism. These approaches aim at decreasing or redirecting the need for action, as well as demonstrating specifically how one or another gambling milieu incites or facilitates excessive gambling. Patients are instructed in relaxation exercises to decrease tension; taught to identify gambling triggers; substitute meaningful rewards for the imagined gain of gambling, etc.

Workbooks enable gamblers to impartially assess the time and cash spent gambling, as well as the time consumed by thinking about gambling. Other helpful CBT methods include stopping techniques, structured motivational sessions which encourage the pathological gambler to think ahead and explore future goals, and strategies that help the pathological gambler recognize the irrationality of his particular gambling fallacies.

Evidence-based research suggests that a cohort of pathologic gamblers will be more likely to accept therapy of some sort, or stay in therapy longer, when a therapist does not insist on total abstinence, at least early in the game. Indeed, some gamblers become abstinent once they get a clear understanding of their dysfunctional behavior.

Pathological gamblers require a tolerant, noncritical attitude regardless of the treatment modality employed. With their testy pride and razor-keen ability to sense disapproval, they will quickly reject help if they suss out even the smallest tincture of moralizing or criticizing.

The attractiveness of group techniques for pathologic gamblers is not surprising, given their extraversion. Group, family, and couple therapy can all be helpful. The distressing impact upon significant others often registers more compellingly in family sessions than in individual therapy alone.

A number of other treatment modalities are undoing investigation, with some promise. These include mindfullness-based therapy, neuromodulation with rapid transcranial magnetic stimulation (rTMS) and transcranial direct current stimulation (tDCS), and cognitive remediation. More research is needed in these areas.

Psychopharmacology

Psychopharmacologic treatment, once largely unsuccessful, now frequently plays a significant role in management. Success has been reported with antidepressants, notably the SSRIs and bupropion; mood stabilizers including lithium and lamotrigine; atypical antipsychotics; opioid antagonists such as naltrexone; and N-acetylcysteine. (The use of benzodiazepine drugs should be avoided or carefully monitored because of the potential for triggering, or increasing, comorbid substance addiction.)

In many patients, it is difficult to determine whether an antidepressant or mood stabilizer is alleviating gambling cravings directly or through treating the comorbid condition—especially affective/bipolar disorders. Many short-term outcome studies have demonstrated improvement. Unfortunately, there are much fewer studies of long-term effectiveness. The same problem exists with other treatment strategies. Pathologic gambling is not a monolithic entity. Different pathological gamblers may respond robustly to a medication that somehow addresses their specific neurophysiology.

Self-Help Interventions

GA is an international self-help organization administered by abstinent pathologic gamblers, with branches in many large cities and small towns, here and abroad. GA groups are found throughout the first world, but can be hard to come by in developing countries.

GA is run along traditional 12-step lines, analogous to Alcoholics Anonymous, with similar groups for relatives. These can be very worthwhile when the pathologic gambler himself refuses help. Cross-addicted pathologic gamblers often attend 12-step groups related to other forms of substance abuse (principally alcohol and cocaine) and indebtedness.

In a GA meeting, clients are encouraged to confront their problems in the company of others thoroughly acquainted with every turn and twist of chasing. Sponsors can be provided. Many new members eventually become sponsors themselves. Organizing financial restitution is a crucial aspect of GA work. Unfortunately, GA dropout rate is high, especially for clients facing exposure for the first time, and when GA is the only source of help. Many pathological gamblers quit GA groups after a bailout: under the rubric of “no debts, no problem.”

Other Interventions

A surprising number of pathologic gamblers, especially adolescents, are unknowledgeable about their illness. Education about substance abuse is common in elementary and high school, but education about problem/pathologic gambling remains sparse. Several studies show that pathological gamblers willing to learn more about their disorder from an instruction manual, structured telephone interviews, and GA interventions have a better likelihood of ceasing or at least reducing their wagering.

It may be necessary to halt a debilitating gambling cycle by hospitalization, taking the pathological gambler out of his terribly stressful milieu. With remission of more serious symptoms, an effective outpatient treatment plan can be organized. At this time, there are more outpatient programs for pathologic gamblers, but still relatively few inpatient units devoted to working with pathologic gamblers.

A bingeing or end-stage pathological gambler will usually be placed in a standard medical hospital setting, or shoehorned into inpatient drug and alcohol programs. Intriguingly, a hospitalized pathologic gambler may suffer initially from physiologic as well as emotional distress analogous to a hospitalized drug addict undergoing detoxification.

Given the pathological gambler's poor health habits, many clients require a complete physical and neurologic examination, appropriate laboratory studies, and thorough evaluation of nutritional and exercise status beyond evaluating and treating their chief physical disorder(s).

GAMBLING AND THE LAW

Federal, state, and local legal systems have traditionally been oriented toward the punishment of pathologic gamblers, rather than their rehabilitation.

Approximately 60% of pathologic gamblers commit illegal acts over the course of their illness. As noted, the majority of pathological gambler crimes are nonviolent—bouncing checks, financial scams, and prostitution. Earlier in the course of his illness, a pathological gambler under arrest is unlikely to be incarcerated, but bailed out and quickly returned to the street where he is likely to relapse again. Incarceration itself can paradoxically become a disguised bailout, forestalling restitution or eliminating it altogether.

Experts in the field have repeatedly told legal authorities that rather than jailing most pathological gamblers—particularly the majority who do not demonstrate antisocial pathology—these offenders should be offered the option to discontinue gambling, make restitution, and undergo some form of treatment as a condition of strict probation. Several projects modeled after drug-related diversion programs have been set up across the United States, with promising results. Funding is a recurrent problem.

At this writing, most states have set up agencies and programs to deal with pathologic gambling. The sum set aside for counseling derived from lotteries, taxing race tracks, and casinos is still relatively small, compared to cash committed to drug and alcohol treatment programs.

Casinos that once refused to admit that pathologic wagering ever existed have begun to fund research into the prevention and treatment of the illness. Some casinos allow compulsive gamblers to set limits on their betting in advance, or even to bar entrance to a pathological gambler when so requested. Casino employees are also being taught to identify out-of-control gamblers and give information about sources of help. Unfortunately, some of these efforts turn out to be mere window dressing. It is exceedingly difficult to create programs to convince the clientele to bet with your head, not over it, while developing evermore creative methods of separating punters from their money.

Attempts to outlaw gambling by local, state, and federal authorities have proven no more successful than the prohibition against selling alcohol during the Roaring Twenties. Protests from antigambling groups often receive only token attention. Across America, straightened economic conditions and formidable lobbying by a flourishing gaming industry are chief causes of nullification of antigambling legislation as more casinos rise.

Internet poker was outlawed several years ago in the United States after cheating scandals and embezzlement schemes were uncovered. No such prohibitions have taken place elsewhere in the world. At this writing, online poker and sports wagering have been re-established in several states. Over the next several years, most states are expected to follow in re-establishing online casino and sports wagering.

THE FUTURE

The Future for the Gambling Industry

The business of gambling has always been fraught with risk *sui generis*. Sponsors of illegal gambling enterprises, from floating crap games to unlicensed casinos, traditionally face the possibility of prosecution by authority as well as criminal harassment for protection. Legal gambling business is exquisitely sensitive to the fickle winds of fashion and the fashionable; the pressure of antigambling forces; and the vicissitudes of socioeconomic circumstances. War is no friend to gambling, except among the warriors.

“Roll the Bones” tells a fascinating and often dispiriting tale of the rise and fall of gambling meccas over the past three centuries. During the past 70 years, Las Vegas casinos have undergone several boom-and-bust cycles.

The proliferation of brick-and-mortar casinos, notably tribal enterprises, poses a challenge to the gaming industry. Although there are more gamblers than ever, they have more places to gamble, diminishing the chances of small as well as large casinos to stay winners.

Today, the greatest gambling industry profit is scored in Asia, particularly in Hong Kong, Singapore, and Macau. The latter’s megacasinos have generated higher yearly returns than Las Vegas since 2008. The Asian boom is linked to rising middle-class prosperity across the Pacific Rim, as well as the preference of regional whales to wager colossally closer to home. Japan’s pachinko machine parlors continue to generate cascades of yen each year (the government, deriving much coin from the parlors, has gotten around its own gambling ban by declaring that pachinko is a painless pastime).

Punters will always seek out real gambling locales, but one predicts that the greatest future profit for the industry will be reaped from the internet virtual casino, especially with the return of the American poker and sports network to the scene.

The Future for Problem/Pathologic Gamblers

Ominously, millions of preteens and children visit internet sites which, intentionally or otherwise, teach them the rudiments of gambling in games where wagering wins prizes instead of cash. One wonders how many of these youngsters will become serious adult recreational gamblers, and for some tomorrow's problem/pathologic gamblers.

Even were gambling opportunities to escalate only modestly over the next decade (unlikely), and the world population increases modestly (also unlikely), one forecasts that the silent public health crisis caused by problem and pathologic gambling will grow even more dire, as more of the afflicted make their debut. Pathological gamblers have a higher mortality rate than ordinary citizens, but they won't perish fast enough to mitigate disaster.

The avalanche of future pathological gamblers certainly could be reduced by a combination of truly useful legal measures, for example, mandating more diversion and treatment programs for gamblers charged with minor criminal offenses; increased public and physician awareness; and the development of more effective treatments.

Educating the public and health professionals about pathologic gambling has increased substantively over the last 20 years. Film, broadcast, print, and internet media devote more time and space to the problem. One lauds the work of groups like the National Council on Problem Gambling (NCPG) (see next section). Nevertheless, it is disheartening to discover that many people, lay and professional, still denounce the sick gambler as a reprobate.

Research into the cause and treatment of gambling problems has increased dramatically and will continue. It is possible that neurobiologic study will yield more effective medications. Perhaps future science will engineer a better fit between the patient and medication.

One predicts there will be refinement of existing cognitive/behavioral strategies, as well as the development of new techniques, as discussed above. Marital, family, and group therapies including GA hopefully will flourish with heightened public awareness.

Brief talking therapies will continue, and perhaps new strategies will be developed. Prolonged psychoanalytically oriented individual treatment will diminish, if only because most insurance schemes do not support it in the United States and worldwide.

The chronic obstacle for treatment success, now and in the future, remains the afflicted gambler's formidable resistance to getting help. The number of pathological gamblers entering any form of therapy is still scant. A prospective patient frequently presents in the advanced stage of the disorder, which is extremely difficult to treat, a situation that one surmises won't change much in the near future. Early identification of fledgling pathological gamblers more likely to accept treatment would seem to be vital, but it is uncertain whether significant programs toward that end can be developed in this country and across the world, or whether the average pathological gambler-to-be would accept them.

THE NATIONAL COUNCIL ON PROBLEM GAMBLING

The NCPG is a nationwide advocacy organization for pathologic gamblers and their families. Membership includes former pathological gamblers, mental health practitioners, researchers, lawyers, and legislators. NCPG is an invaluable source of cutting-edge information about every aspect of gambling. It also maintains a hotline for pathological gamblers, relatives, and friends which offers immediate help and provides information about gambling treatment across the United States. Address: 730 11th Street NW, Suite 601, Washington, DC 20001. https://www.ncpgambling.org/. 1-800-GAMBLER.

CLINICAL VIGNETTES

INTERNET GAMBLING

Jim, a 19-year-old freshman, was introduced to internet poker by his roommate. Initially, he played free games. He quickly graduated to low-stake cash games. After winning several hundred dollars, he thought he was ready to take on big-league high-stakes play.

By midterm, Jim had lost several thousand dollars and was spending most days on his computer, playing several games simultaneously. He neglected studies, slept through examinations, and ate irregularly and poorly. He quickly exhausted the money his parents sent every month, maxed out several credit cards, and began borrowing heavily from friends. Eventually, he was placed on academic probation and contemplated suicide. Luckily, a friend succeeded in getting him to go to the student health center, and he was able to get straight through individual therapy and GA membership.

GAMBLING IN THE ELDERLY

Fred, a 79-year-old Air Force sergeant, moved to Las Vegas after retiring in his mid-40s. He and his wife were easily able to afford a small house 10 miles out of town. For the next 20 years, he worked as a casino security officer, then retired.

Fred lived comfortably for another decade until his wife died from breast cancer. He was inconsolable, hid his grief beneath a stoic façade, and became increasingly depressed. He had never gambled much and had only visited the Strip for the occasional dinner or show.

To alleviate grief and boredom, he began working again as a part-time guard at the casino. Its glitter and excitement made him feel better; he found the warbling of slot machines peculiarly soothing. He began playing a few coins before returning home. Within a month, he was spending every evening after work and during the weekend at the slots. He enjoyed the company of other devotees, swapping stories about hot machines and big cash-outs.

At first, Fred hardly noticed that his hobby was cutting into his finances. He told himself that he was only betting quarters and that his wife would want him to enjoy himself. But within 6 months, he had poured his entire savings into the machines. He could not meet bills and cut back on food and other essentials.

He realized how sick his gambling had become when he began thinking he might sell the home his wife loved so much to support his habit. He joined GA and subsequently exchanged his job for work at a supermarket near his house. “I never believed this could happen to me,” he says, “just as sure as I am it could happen tomorrow.”

A CLASSIC PATHOLOGIC GAMBLER

Gerry is a 35-year-old former auto dealership owner. Two of his uncles were compulsive gamblers, and his paternal grandfather had been hospitalized for bipolar illness. He began playing poker and going to the racetrack when he was 15. He dropped out of college after a few months and became a car salesman. Soon, he was promoted to showroom manager, then went out on his own. By 32, he was a multimillionaire owner of a dealership chain, happily married with two children.

Throughout his life, Gerry continued to gamble frequently. He was a successful weekend sports bettor, as well as a consistent winner in weekly gin rummy, poker games, and occasional jaunts to Las Vegas and Atlantic City.

Gerry had always wanted a third child. His wife developed eclampsia, and the baby was premature and died of respiratory failure. Gerry became depressed for the first time in his life. He visited casinos more often, and gradually increased his bets at blackjack and craps. His card games became boring—the action was zilch. He began frequenting a mob-backed local poker parlor which featured megastake action.

Over several years, Gerry slipped into a typical gambling spiral. He accumulated several million dollars in debts and invaded his personal bank accounts, including his children's college money. He maxed out all his credit cards and began borrowing from loan sharks at exorbitant rates. He thought of killing himself in a car crash so that insurance would take care of his family.

Gerry's illness was exposed when his Porsche was repossessed one Sunday morning. His wife threatened to divorce him when she learned the extent of his losses, but her wealthy father intervened and bailed him out. He swore never to gamble again, entered GA, and within 2 months resumed chasing. Over the next 10 years, Gerry endured four more episodes of recovery and relapse. His wife left him, he lost his dealerships, and he was forced into bankruptcy.

He finally signed himself into a medical center with a pilot dual-diagnosis recovery program, where he was diagnosed with atypical bipolar disorder. His treatment included GA meetings, individual and family counseling, and pharmacotherapy with bupropion and lamotrigine.

Gerry eventually reconciled with his wife and family. He started selling cars again, now lives modestly, still attends meetings regularly, but thinks of himself as a gambler who could relapse on any given day.

A PATHOLOGIC LOTTERY GAMBLER

Nelly, a 35-year-old housewife, had been playing the State Lotto for a dollar every Saturday for many years. She rotated numbers based on her children's birthdays, her anniversary, home address, etc. Occasionally, she won a few dollars, which she always reinvested in more tickets.

Then Nelly missed a 5-million-dollar jackpot because she inadvertently had reversed the last two numbers of her license plate. She felt inexplicably cheated. Thoughts about her near triumph increasingly filled her mind, and she became obsessed with "finding my way back to the right combination—it's gotta happen."

Nelly began playing Lotto every day and soon was buying 20 tickets. She also went to church bingo games several times a week, then over weekends. She gradually distanced herself from her family and friends and became profoundly depressed.

In order to finance her wagering, Nelly spent the money she had carefully saved over a decade, then surreptitiously sold jewelry inherited from her mother. Overwhelmed with guilt, she finally revealed her problem to her husband as well as a sympathetic priest. She reluctantly agreed to enter treatment.

She responded well to family support, medication, and individual and family therapy, but shunned GA groups because she "doesn't want to hang around with a bunch of bums every day slurping coffee. Everyone thinks the lottery is harmless, but it nearly destroyed my life."

AN INCARCERATED PATHOLOGIC GAMBLER

Lonnie, an 18-year-old young man with ADHD, received a 3-year prison sentence for low-level drug dealing. On the outside, he dealt with the chronic boredom characteristic of that disorder by snorting cocaine and playing hearts. He considered himself a master at the game.

He dealt with the tedium of prison life by playing hearts at every chance. He knew, yet denied, that he was up against career criminals who had honed their skills over years of play. He began borrowing heavily from the loan shark of a feared gang. When he fell behind, he was repeatedly beaten and raped by gang members. Traumatized both physically and psychologically, he hung himself.

Additional Information: Harvey Roy Greenberg, MD, died June 11, 2022.

FURTHER READINGS

Achab S, Khazaal Y. Psychopharmacological treatment in pathological gambling: a critical review. *Curr Pharm Des*. 2011;17(14):1389–1395.

Aragay N, Roca A, Garcia B, et al. Pathological gambling in a psychiatric sample. *Compr Psychiatry*. 2012;53(1):9–14.

Balodis IM, Kober H, Worhunsky PD, Stevens MC, Pearlson GD, Potenza MN. Diminished frontostriatal activity during processing of monetary rewards and losses in pathological gambling. *Biol Psychiatry*. 2012;71(8):749–757.

Bellegarde JD, Potenza MN, eds. Neurobiology of pathological gambling. In: Ross D, Kincaid H, Spurret D, Collins P. *What Is Addiction*? MIT Press; 2010:27–51.

Bergler E. *The Psychology of Gambling*. International University Press; 1957.

Billieux J, Van der Linden M, Khazaal Y, Zullino D, Clark L. Trait gambling cognitions predicting near-miss experiences persistence in laboratory slot-machine gambling. *Br J Psychol*. 2012;103(3):412–427.

Blaszczynski A. *Overcoming Compulsive Gambling: A Self-Help Guide Using Cognitive Behavioral Techniques*. Robinson; 1998.

Bolen DW, Boyd WH. Gambling and the gambler: a review and preliminary findings. *Arch Gen Psychiatry*. 1968;18(5):617–630.

Brasfield H, Febres J, Shorey R, et al. Male batterers' alcohol use and gambling behavior. *J Gambl Stud*. 2012;28(1):77–88.

Brewer JA, Potenza MN, Desair RA. Differential associations between problem and pathological gambling and psychiatric disorders in individuals with and without alcohol abuse or dependence. *CNS Spectr*. 2010;15(1):33–44.

Clark L. From impulsivity to addiction: gambling disorder and beyond. *Psychiatric Times*. 2015. https://www.psychiatrictimes.com/view/impulsivity-addiction-gambling-disorder-and-beyond

Crockford DN, Goodyear B, Edwards J, Quickfall J, el-Guebaly N. Cue-induced brain activity in pathological gamblers. *Biol Psychiatry*. 2005;58(10):787–795.

Cowlishaw S, Merkouris S, Dowling N, Anderson C, Jackson A, Thomas S. Psychological therapies for pathological and problem gambling. *Cochrane Database Syst Rev*. 2012;11:CD008937.

Dodd ML, Klos KJ, Bower JH, Geda YE, Josephs KA, Ahlskog JE. Pathological gambling caused by drugs used to treat Parkinson disease. *Arch Neurol*. 2005;62(9):1377–1381.

Fenichel O. *Psychoanalytic Theory of Neurosis*. WW Norton; 1945.

Freud S. Dostoyevsky and parricide. *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. 1961;21:173–196.

Grant JE, Donahue CB, Odlaug BL, Kim SW. A 6-month follow-up of imaginal desensitzation plus motivational interviewing in the treatment of pathological gambling. *Ann Clin Psychiatry*. 2011;23(1):3–10.

Gyollai A, Griffiths MD, Barta C, et al. The genetics of problem gambling: a systematic review. *Curr Pharm Des*. 2014;20(25):3993–3999.

Hodgins DC, Stea JN, Grant JE. Gambling disorders. Vol. 378. 2011. www.thelancet.com

Joutsa J, Johansson J, Niemelä S, et al. Mesolimbic dopamine release is linked to symptom severity in pathological gambling. *Neuroimage*. 2012;60(4):1992–1999.

Kennedy SH, Fulton K, Welsh BR, et al. Frequency and correlates of gambling problems in outpatients with major depressive disorder and bipolar disorder. *Can J Psychiatry*. 2010;55(9):568–576.

Kozlowski C. Demand for stimulation and probability preferences in gambling decision.

Pol Psychol Bull. 1977;8(2):67–73.

Lesieur HR. *The Chase: Career of the Compulsive Gambler*. Anchor Books; 1977.

Limbrick-Oldfield EH, van Holst RJ, Clark L. Fronto-striatal dysregulation in drug addiction and pathological gambling: consistent inconsistencies? *Neuroimage Clin*. 2013;2:385–393.

Lobo DSS, Kennedy JL. Genetic aspects of pathological gambling: a complex disorder with shared genetic vulnerabilities. *Addiction*. 2009;104(9):1454–1465.

Meng Y, Deng W, Wang H, Guo W, Li T. The prefrontal dysfunction in individuals with Internet gaming disorder: a meta-analysis of functional magnetic resonance imaging studies. *Addict Biol*. 2015;20(4):799–808.

Potenza MN. Neurobiology of gambling behaviors. *Curr Opin Neurobiol*. 2013;23(4):660–667.

Schreiber L, Odlaug BL, Kim SW, Grant JE. Characteristics of pathological gamblers with a problem gambling parent. *Am J Addict*. 2009;18(6):462–469.

Schwartz DG. *Roll the Bones: The History of Gambling: (Casino Edition)*. Winchester Books; 2013.

Stefanovics EA, Potenza MN. Update on gambling disorder. *Psychiatr Clin North Am*. 2022;45(3):483–502.

Stekel W. The Gambler. In: *Disorders of the Instincts and the Emotions; transl*. James S. Teslar, Vol. 2, Ch. 13, Williams and Norgate, 1929.

Studer B, Apergis-Schoute AM, Robbins TW, Clark L. What are the odds? The neural correlates of active choice during gambling. *Front Neurosci*. 2012;6:46.

Toneatto T. Cognitive psychopathology of problem gambling. *Subst Use Misuse*. 1999;34(11):1593–1604.

Weintraub D, Koester J, Potenza MN, et al. Impulse control disorders in Parkinson disease: a cross-sectional study of 3090 patients. *Arch Neurol*. 2010;67(5):589–595.

17

Post-Traumatic Stress Disorder

Arieh Y. Shalev, M.D., and Charles R. Marmar, M.D.

Post-traumatic stress disorder (PTSD) is classified in the Fifth Edition of the American Psychiatric Association (APA) Diagnostic and Statistical Manual (DSM-5) as a Trauma and Stressor-Related Disorder. Like other DSM-5 disorders, PTSD is defined by its clinical features and, specifically, by four groups of symptoms that develop following exposure to a traumatic event. PTSD diagnostic criteria allow clinicians to reliably identify and diagnose affected individuals and provide a framework for research. PTSD symptoms do not directly reflect the disorders' underlying neuropathologic processes. They are best understood *as varying behavioral expression of an underlying neurobehavioral systems' pathology, shaped by individual, environmental, and cultural factors*.

Individuals who suffer from PTSD relive distressing instances of the traumatic event with vivid emotional proximity and high, imperative intensity. They organize their lives trying to contain and mitigate the persistent effects of past traumatic experience. For those traumatized in a warzone, "*the war never ends.*" Victims of rape, assault, or torture describe difficulties engaging and trusting other humans. Repeatedly re-experiencing the trauma in the present, PTSD patients' lives are dominated by effortful attempts to avoid concrete and mental reminders including thoughts, images, and feelings associated with the traumatic event. They scan the environment for threat signals, which they fearfully expect and remain on guard, tense, restless, and exhausted.

PTSD has been diagnosed across continents and cultures. It is a common pathologic outcome of a wide variety of traumatic events, from wars and disasters to individual events such as road traffic or work accidents. The concrete content of PTSD symptoms (e.g., specific situations avoided, nightmare themes) reflects aspects of the triggering events and thus may differ from one individual to another. Core symptom patterns however (e.g., insomnia, exaggerated startle response, or emotional numbing), are similar across traumatic events, nationalities, races, social classes, and cultures, suggesting a common underlying neurobehavioral foundation.

Symptom persistence despite threat termination and inability to regain a sense of safety are core features of the disorder and determine its “post-traumatic” or postexposure signature. Another core feature is the involuntary, uncontrollable, and intense nature of the symptoms. PTSD patients live under the shadow of past trauma. They are unable to distance themselves emotionally from their traumatic experiences, which are ever-present, powerful, and dominant. These features differentiate PTSD from an immediate reaction to traumatic events and from remote memories of trauma exposure. The latter are not forcefully “relived” but rather remembered. They are not ever-present, but rather occasional, such as on anniversaries or memorial days. They do not dominate behavior. Those with PTSD report frequent, intrusive, and distressing “flashbulb” recollections of the traumatic events during daytime or in nightmares. They respond to reminders of the event by involuntary and uncontrollable alarm, and defensively avoid external and internal stimuli reminiscent of the trauma. They are tense, anxiously aroused, expecting harm, and at the same time tired, emotionally numb, and unable to enjoy previously pleasurable activities. Many express these symptoms for years and even decades.

Understanding the bio-psycho-social mechanisms leading from trauma to PTSD is central to predicting and preventing the disorder. Chronologically, these mechanisms combine pretrauma vulnerability factors, trauma exposure impact, and postexposure “symptom maintaining” factors. Vertically, each of these categories comprises molecular, brain circuit, neurobehavioral, cognitive, emotional, interpersonal, and social dimensions. The etiology and the pathogenesis of PTSD are, therefore, complex and multifactorial, triggered but not fully explained by the traumatic event.

DEFINITION

PTSD is a trauma and stress-related disorder, defined by the co-occurrence of re-experiencing, avoidance, negative beliefs, and hyperarousal symptoms in survivors of significant adversity. PTSD is a common outcome of all types of traumatic events, from the most horrifying and protracted (e.g., captivity and torture) to shorter events or incidents (e.g., accidents). The construct of PTSD has, in fact, replaced several event-related "syndromes" that pre-existed its definition (e.g., "*concentration camp syndrome*," "*war neurosis*," or "*rape victim syndrome*"). Unlike most other mental disorders, the diagnosis of PTSD relies on associating current symptoms with a triggering traumatic event (e.g., a triggering life event is not a required diagnostic feature of depression or psychosis). The association of PTSD symptoms with the traumatic event can be chronologic (symptoms starting shortly after the event) or content-related (intrusive recollections of aspects of the traumatic event, or avoidance of specific reminders).

The essential features of PTSD are the development of characteristic symptoms and associated functioning difficulties following direct (i.e., being a victim of) or indirect exposure (witnessing, being told about) to one or more traumatic events (see Table 17–1).

Like other complex disorders, the clinical presentation of PTSD may vary among individuals. Fear-based re-experiencing, emotional, and behavioral symptoms may predominate in some individuals whereas others may primarily show anhedonia, dysphoria, negative cognitions, emotional numbing, social isolation, irritability, or insomnia.

Table 17–1.
DSM-5 Diagnostic Criteria for Post-Traumatic Stress Disorder

Post-Traumatic Stress Disorder DSM-5 Diagnostic Criteria

A. Exposure to actual or threatened death, serious injury, or sexual violence in one (or more) of the following ways:
   1. Directly experiencing the traumatic event(s).
   2. Witnessing, in person, the event(s) as it occurred to others.
   3. Learning that the traumatic event(s) occurred to a close family member or close friend. In cases of actual or threatened death of a family member or friend, the event(s) must have been violent or accidental.
   4. Experiencing repeated or extreme exposure to aversive details of the traumatic event(s) (e.g., first responders collecting human remains: police officers repeatedly exposed to details of child abuse).

   Note: Criterion A4 does not apply to exposure through electronic media, television, movies, or pictures unless this exposure is work related.

B. Presence of one (or more) of the following intrusion symptoms associated with the traumatic event(s), beginning after the traumatic event(s) occurred:
   1. Recurrent, involuntary, and intrusive distressing memories of the traumatic event(s). **Note:** In children older than 6 years, repetitive play may occur in which themes or aspects of the traumatic event(s) are expressed.

2. Recurrent distressing dreams in which the content and/or effect of the dream are related to the traumatic event(s).
3. Dissociative reactions (e.g., flashbacks) in which the individual feels or acts as if the traumatic event(s) were recurring. (Such reactions may occur on a continuum, with the most extreme expression being a complete loss of awareness of present surroundings.)
4. Intense or prolonged psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event(s).
5. Marked physiologic reactions to internal or external cues that symbolize or resemble an aspect of the traumatic event(s).

C. Persistent avoidance of stimuli associated with the traumatic event(s), beginning after the traumatic event(s) occurred, as evidenced by one or both of the following:

1. Avoidance of or efforts to avoid distressing memories, thoughts, or feelings about or closely associated with the traumatic event(s).
2. Avoidance of or efforts to avoid external reminders (people, places, conversations, activities, objects, situations) that arouse distressing memories, thoughts, or feelings about or closely associated with the traumatic event(s).

D. Negative alterations in cognitions and mood associated with the traumatic event(s), beginning or worsening after the traumatic event(s) occurred, as evidenced by two (or more) of the following:

1. Inability to remember an important aspect of the traumatic event(s) (typically due to dissociative amnesia and not to other factors such as head injury, alcohol, or drugs).
2. Persistent and exaggerated negative beliefs or expectations about oneself, others, or the world (e.g., “I am bad,” “No one can be trusted,” ‘The world is completely dangerous,” “My whole nervous system is permanently ruined”).
3. Persistent, distorted cognitions about the cause or consequences of the traumatic event(s) that lead the individual to blame him-/herself or others.
4. Persistent negative emotional state (e.g., fear, horror, anger, guilt, or shame).
5. Markedly diminished interest or participation in significant activities.
6. Feelings of detachment or estrangement from others.
7. Persistent inability to experience positive emotions (e.g., inability to experience happiness, satisfaction, or loving feelings).

E. Marked alterations in arousal and reactivity associated with the traumatic event(s), beginning or worsening after the traumatic event(s) occurred, as evidenced by two (or more) of the following:

1. Irritable behavior and angry outbursts (with little or no provocation) typically expressed as verbal or physical aggression toward people or objects.
2. Reckless or self-destructive behavior.
3. Hypervigilance.
4. Exaggerated startle response.
5. Problems with concentration.

    6. Sleep disturbance (e.g., difficulty falling or staying asleep or restless sleep).

F. Duration of the disturbance (criteria B, C, D, and E) is more than 1 month.

G. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

H. The disturbance is not attributable to the physiologic effects of a substance (e.g., medication, alcohol) or another medical condition.

Specify whether:

With dissociative symptoms: The individual's symptoms meet the criteria for post-traumatic stress disorder, and in addition, in response to the stressor, the individual experiences persistent or recurrent symptoms of either of the following:

1. Depersonalization: Persistent or recurrent experiences of feeling detached from, and as if one were an outside observer of, one's mental processes or body (e.g., feeling as though one were in a dream; feeling a sense of unreality of self or body or of time moving slowly).
2. Derealization: Persistent or recurrent experiences of unreality of surroundings (e.g., the world around the individual is experienced as unreal, dreamlike, distant, or distorted).

   Note: To use this subtype, the dissociative symptoms must not be attributable to the physiologic effects of a substance (e.g., blackouts, behavior during alcohol intoxication) or another medical condition (e.g., complex partial seizures).

Specify if:

With delayed expression: If the full diagnostic criteria are not met until at least 6 months after the event (although the onset and expression of some symptoms may be immediate).

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:301–304.

DSM-5 parses PTSD symptoms into four subcategories or diagnostic criteria: re-experiencing, avoidance, negative cognitions, and hyperarousal. To qualify for a diagnosis of PTSD, a trauma-exposed individual must express a predefined minimum number of symptoms in each category. The numerous combinations of qualifying symptoms allowed by DSM-5 can accommodate significant heterogeneity of individual presentations. Although such flexibility has been criticized as being too permissive and allowing different kinds of patients under the same diagnosis, such variety is in line with the multiple ways in which dysfunctions of specific molecular pathways and brain circuits are expressed by individuals who have different neurocognitive, emotional, and interpersonal resources, different life history, environmental pressures, beliefs, and understanding of causality. PTSD symptoms are best understood, therefore, as varying behavioral indicators of underlying neurobiology.

The *directly experienced traumatic* events (PTSD criterion A) include, but are not limited to, exposure to war as a combatant or civilian, threatened or actual physical assault, threatened or actual sexual violence, being kidnapped, being taken hostage, terrorist attack, torture, incarceration as a prisoner of war, natural or human-made disasters, and severe motor vehicle accidents. Medical disorders that qualify as traumatic events involve sudden, catastrophic illnesses. *Witnessed events* include, but are not limited to, observing threatened or serious injury, unnatural death, physical or sexual abuse of another person due to violent assault, domestic violence, accident, war or disaster, or a medical catastrophe. The latter has received ample demonstration among agony- and death-exposed health-care personnel, patients, and the public during the recent COVID-19 epidemic. *Indirect exposure* through learning about an event is limited to experiences affecting close relatives or friends and experiences that are violent or accidental (e.g., violent personal assault, suicide, serious accident, and serious injury).

Characteristic symptoms resulting from trauma exposure include persistent intrusion symptoms associated with the traumatic event (criterion B), persistent avoidance of stimuli associated with the traumatic event (criterion C), negative alterations in cognitions or mood associated with the event (criterion D), and marked alterations in arousal and reactivity associated with the traumatic event (criterion E). The full symptom picture must be present for more than 1 month (criterion F), and cause clinically significant distress or impairment in social, occupational, or other important areas of functioning (criterion G) and should not be attributable to the physiologic effects of a substance or another medical condition.

Symptoms of intrusion (five symptoms, one needed to meet PTSD diagnostic criterion) reflect the unrelenting, unavoidable, and uncontrollable presence of the traumatic event in the person's life. Symptoms must be recurrent, intense, and marked. They evoke significant distress and, often, physiologic alarm reactions (heart pounding, cold sweats). Most importantly, they are experienced and lived "here and now" and not as retrospective, distant recall of past events. PTSD intrusion symptoms differ, therefore, from other autobiographic or episodic memories in that they are not perceived as having occurred "there" and "then" but rather as recurring and current. Their emotional "here and now" presence coexists with the patient's clear understanding that the traumatic event occurred in the past. At their extreme expression (criterion B3; dissociative reactions), however, they literally transport the person back to the traumatic situation, overshadowing the sense of current reality.

Symptoms of avoidance (two symptoms, one needed) similarly reflect a direct linkage with the traumatic event. Criterion C2 (avoidance of reminders) reflects threat avoidance acquired during the traumatic event and reinforced thereafter, in line with the fear conditioning theory of PTSD. Studies have shown that the C2 criterion (avoidance of places and situations) tend to expand with time, such that new objects and situations are avoided (e.g., avoidance of combat situations may extend to avoiding weapons, uniforms, crowds, mail envelops reminding those used to call people to reserve duties, etc.). Such "inappropriate generalization" can progress to a point of confining severely affected PTSD patients to restricted environments with "no surprises." Criterion C1 is equally intriguing in that it reflects an effort to avoid inner (mental) representations of the traumatic event, as if these representations are analogous to external reminders. In that sense, trauma survivors with PTSD carry with them internalized representations of the trauma that, when accessed, evoke significant distress and are intensely avoided. Just being removed from a traumatic situation is not enough to reduce PTSD patients' avoidance: they avoid thoughts and memories. This clinical feature resembles the "obsession" component of obsessive-compulsive disorder (OCD), where ruminations and thoughts become as distressing as actual encounters with feared objects. Patients expressing C1 have been described as constantly trying to push thoughts and images out of their consciousness and thereby paradoxically reinforcing their presence and never letting them fade away. Individuals who recover from traumatic experiences, in contrast, progressively forget and don’t have to deliberately suppress their recollections.

The *negative alterations in cognitions and mood associated with the traumatic event* criteria (seven symptoms, two needed) is a new DSM-5 cluster, which did not exist in DSM-IV. It expands the scope of PTSD symptoms beyond reactions to the traumatic event into general dispositions toward life, other humans, and society. The “diminished interest” (D5), “detachment and estrangement” (D6), and “inability to experience positive emotions” (D6) criteria are remnants of one of the theoretical sources of PTSD, namely the loss and grief analogy, in which, during acute grief, unrelated aspects of life lose their appeal. Freudian theory suggested that the bereaved are not free to re-engage emotionally until they mourn their loss. Those symptoms mark a profound dysphoria and anhedonia and overlap with depression symptoms. The “inability to recall” criterion (D1) follows a long historical thread starting with the construct of “repression” (of traumatic memories), but also supported by a more modern construct of “failure to encode” traumatic experiences in memory at the height of their occurrence. Inability to recall is one of the most difficult symptoms to ascertain clinically. Other negative alterations criteria reflect changes in cognitive schemata and belief (e.g., sustained belief that the world is dangerous, other people untrustworthy, and oneself worthless). Indeed, what characterizes PTSD negative cognition is its permanence: the “world is *always* dangerous” (not sometimes); “I am irreversibly worthless” or unable to cope with another trauma (a construct known as “low self-efficacy”). Symptoms under criterion D are remarkably heterogeneous and not event specific. They nonetheless reflect the extent to which PTSD is associated with widespread modification of personal dispositions, mood, and beliefs—far beyond the direct effect of the traumatic event.

Finally, the *alteration in arousal (E criterion, five symptoms, two needed)* combines an unrelenting scanning of the environment for threat signals (hypervigilance) with irritability, insomnia, loss of concentration, exaggerated startle, and reckless behavior. As with the “negative alterations” criterion, symptoms may begin with the traumatic event or worsen after it but do not have a direct link with the content of the event or refer to content-specific aspects of the event. Frequently observed in PTSD patients, the resulting irritability, outburst of anger, restlessness, and at times progression to violence are often directed at family members and close relationships and further enhance the patient's isolation and loneliness.

Clinicians using DSM-5 diagnostic criteria for PTSD should be aware of the syndrome's inherent heterogeneity, varying individual presentations, and co-occurrence with other mental disorders. They must rely on symptoms with a direct link to the traumatic event when making a diagnosis of PTSD. Relative to DSM-IV, DSM-5 raised the bar for identifying PTSD. Furthermore, it identifies slightly different subsets of trauma survivors as having PTSD, such that some individuals with DSM-IV PTSD may not meet DSM-5 strict diagnostic criteria and vice versa. A formal diagnosis of DSM-5 PTSD, therefore, clearly identifies subjects in need of help. However, it should not be used to exclude other, slightly different distressing responses to traumatic events from receiving assistance, help, and treatment.

HISTORY

Traumatic syndromes resembling PTSD have been described and sharply debated during the 18th, 19th, and 20th centuries. "Stragglers" in the American Civil War were described as "*trembling, staring into the middle distance and jumping at any loud noise.*" Survivors of ship explosions in Toulon in 1907 and 1911 exhibited *"recapitulation of the scene, terrifying dreams, diffuse anxiety, fatigue, and various minor phobias.*" World War I trench warfare survivors were labeled with "shell shock" and "war neurosis," the first suggesting an organic cause (traumatic brain injury) and the second a psychological origin. The term "combat fatigue," used by the U.S. army during World War II and "combat operational stress" used by the U.S. military in Iraq and Afghanistan, implies a normal reaction to extreme demands rather than a mental disorder. Resembling current PTSD symptoms, the manifestations of "operational fatigue" included, "*Irritability, fatigue, difficulties falling asleep, startle reaction, depression, tremor and evidence of sympathetic over-reactivity, difficulties in concentration and mental confusion, preoccupation with combat experiences, nightmares and battle dreams, phobias, personality changes, and increased alcoholism.*" Symptoms of "traumatic neurosis" similarly included, "*Fixation on the trauma, typical dream life, contraction* *of general level of functioning, irritability, and proclivity to explosive aggressive reactions.*"

Despite clinical observations during World War I and World War II, traumatic stress disorders were not included in the first two Diagnostic and Statistical Manuals of mental disorders, DSM-I (1952 to 1968) and DSM-II (1968 to 1980). DSM-I included a category of "Gross Stress Reaction" of a maximum 2-month duration, thereby implying that, unless there is other underlying psychopathology, reactions to major stressors are transient. Those who remain ill after a traumatic exposure were seen as having another disorder (e.g., neurosis, depressive illness). DSM-II eliminated the "Gross Stress Reaction" and did not have a diagnostic category for reactions to major stressors. Ironically, therefore, American psychiatrists did not have a diagnostic category to accommodate war-related disorders during one of the harshest U.S. military engagements, the Vietnam War.

The introduction of PTSD into DSM-III (1980) followed a delayed realization of psychopathologic sequelae of the Vietnam War. The APA Task Force on Nomenclature (the DSM-III Task Force) reviewed over 700 clinical observations of psychologically disabled veterans and rape victims. It coined the term PTSD and organized its symptoms into three diagnostic criteria (re-experiencing of the trauma, numbing of responsiveness, and other, mainly hyperarousal symptoms) following Kardiner's description of "traumatic neuroses" and UCSF researcher Mardi Horowitz's delineation of "stress response syndrome." Like other DSM-III disorders, PTSD diagnostic criteria were an optimized template, based on tradition and preliminary studies, and meant to be validated by empirical research: the APA *Committee on Reactive Disorders* included PTSD in DSM-III as a "*consensus*" rather than "*empirically validated*" disorder.

Empirical findings during the first decades after the publication of DSM-III refined PTSD diagnostic criteria, and these refinements were included in subsequent editions of the diagnostic and statistical manual: DSM-III-R (1987) and DSM-IV (1994). For example, survivor guilt, initially included in DSM-III, was found to be rare and excluded from subsequent DSM editions. Physiologic reactivity to reminders was moved from the arousal to the re-experiencing criterion. The most radical change, however, involved the definition of a "traumatic event" (PTSD criterion A). The disorder was originally seen as resulting from extreme events that were "*out of the realm of normal life experience*." Based on evidence that PTSD also occurred in survivors of less dramatic occurrences (e.g., road traffic accidents, natural disasters), the array of traumatic events leading to PTSD was extended in DSM-IV and the "universally devastating" qualification was replaced by the "individually shattering" one. Traumatic events in DSM-IV also included instances of witnessing trauma to others (such as witnessing an execution or body handling). DSM-IV added a requirement of experiencing *intense fear, helplessness, or horror at the time of exposure* to its traumatic event criterion. DSM-IV also introduced a related syndrome, acute stress disorder, occurring during the first days and weeks following trauma exposure and characterized by re-experiencing, avoidance, arousal, and by prominent symptoms of dissociation.

DSM-5 (2013; Table 17–1) eliminated the requirement for strong emotional reaction at the time of the trauma, introduced a new symptom cluster, namely the *negative alterations in cognitions and mood,* removed PTSD from the group of anxiety disorders, and created a new category of trauma and stressor-related disorders. Underlying DSM-5 changes were studies showing limited emotional reactions at the time of the trauma among some combat-exposed military personnel and first responders who subsequently developed the full PTSD syndrome, as well as theoretical considerations regarding the importance of alterations in worldview and mood in individuals with prolonged PTSD.

Reliability and Validity of the Syndrome

Originally introduced as a disorder awaiting empirical validation, the nature and persistence of PTSD have been repeatedly validated. Early studies evaluating the construct validity of PTSD found it to be very good and to generally match DSM-III original symptom clusters (re-experiencing, avoidance, and hyperarousal). More recent work suggested that avoidance symptoms comprise two distinct subcategories: specific avoidance of trauma memories and reminders, and a broader restriction or alteration of social engagement and mood. DSM-5 criterion D reflects these new findings.

Originally referred to as a *normal reaction to abnormal circumstances*, PTSD is currently recognized as a psychopathologic consequence of trauma exposure, occurring in some individuals and sparing others. The traumatic event is seen as triggering rather than causing PTSD. *The* *etiology of PTSD is multicausal, involving pre-exposure vulnerability (risk) factors, exposure characteristics, and postexposure recovery environment at biologic, neurobehavioral, and psychosocial levels*. Accordingly, PTSD observable diagnostic criteria are symptomatic expressions of underlying neuropathology, which current research is progressively evaluating.

PTSD symptoms and their communication by those with the disorder are influenced by self-awareness, social norms, belief systems, desirability of disclosure, stigma, and help-seeking behavior. Numerous survivors misinterpret their symptoms as reflecting a normal response, or delay communicating their symptoms or seeking help. In U.S. military personnel with PTSD, for example, the median time from the end of the last deployment to engagement in mental health care was over 2 years. Contrary to a prevalent late 20th-century criticism, PTSD is *not* a psychosocial construct. It is a major disturbance of the central nervous system's (CNS) threat sensitivity, threat detection, emotion regulation, and processing of reward, triggered or exacerbated by trauma exposure, usually present well before formal help-seeking.

Innovation in PTSD Classification and Subtyping

Recent years have seen a growing interest in parsing PTSD biologic heterogeneity and identifying clinically and biologically pertinent subtypes. Psychophysiological studies have identified a large subgroup of PTSD patients who did not respond to trauma reminders by hyperarousal and, instead, developed dissociative symptoms. A series of clinical, psychometric, and functional brain imaging studies subsequently laid the foundations for a dissociative subtype of PTSD, currently recognized by DSM-5, and characterized by symptoms of depersonalization and derealization, gaps in awareness and memory, and frequent reckless or self-destructive behavior, and characterized biologically by an overmodulation of the amygdala alarm signaling. Data from the World Mental Health Surveys, a large and demographically broad sample, estimated the prevalence of dissociative subtype at 14.4% of those with PTSD.

The 11th revision of the International Classification of Diseases (ICD-11) reduced the number of PTSD symptom criteria (intrusive and distressing re-experiencing of the traumatic event; avoidance of internal and external reminders of the traumatic event; hypervigilance and enhanced startle reaction to stimuli) and, at the same time, defined complex PTSD. The new disorder describes the more complex reactions that are typical of individuals exposed to chronic trauma that include, in addition to PTSD symptoms, disturbances of emotion regulation, self-identity, and relational capacities. Studies of U.S. military personnel and large international samples have shown that numerous individuals diagnosed with PTSD using DSM-5 criteria do not meet ICD-11 PTSD criteria and vice versa, calling for caution in using either classification to exclude traumatized patients who are clinically distressed and functionally impaired.

COMPARATIVE NOSOLOGY

Previously classified among the anxiety disorders (DSM-III through DSM-IV-TR), research has shown that many PTSD patients do not show prominent anxiety symptoms, but rather anhedonia, dysphoria, irritability, or dissociation. These findings have justified the inclusion of PTSD in a new trauma and stress-related disorders category. The latter is defined by a linkage between current symptoms and past exposure to external stressors. Other disorders in that category include acute stress disorder, adjustment disorder, reactive attachment disorder, and disinhibited social engagement disorder.

PTSD often co-occurs with other mood and anxiety disorders: about 50% of adults with PTSD concurrently meet diagnostic criteria for major depression, and the lifetime comorbidity of the two disorders is about 90%. Military veterans and survivors of interpersonal trauma demonstrate higher rates of comorbid PTSD and major depressive disorder (MDD). PTSD and depression may co-occur shortly after trauma exposure or develop years later. These high co-occurrence rates emphasize the need to routinely assess MDD among individuals with PTSD.

Most trauma survivors do not develop PTSD. Contrasting with PTSD patients, trauma survivors who do not develop PTSD often express transient initial symptoms and may be occasionally reminded of past trauma (e.g., at “anniversary” dates), but their lives are not dominated by the imperative, uncontrollable, and inescapable presence of the traumatic experiences. PTSD, therefore, is a distinct pathologic consequence of trauma exposure and postexposure disorder-maintaining factors such as lack of social support and repeated exposure to stress. Traumatic events are better construed as a *necessary but not sufficient cause of PTSD*. Along with PTSD, traumatic events can accelerate or exacerbate other

DSM-5 disorders, such as major depression or substance misuse. When trauma survivors meet both PTSD and depression diagnostic criteria, both disorders can be diagnosed as comorbid, and clinicians have to prioritize their responses to address the most prominent, or life-threatening, condition.

EPIDEMIOLOGY

Prevalence of PTSD in the Community

Trauma exposure is common, with estimates of lifetime trauma exposure in the United States ranging from 50% to 89%. The most prevalent types of traumatic events in the United States are physical or sexual assault (52%) and being involved in an accident or fire (50%). Lifetime exposure to multiple events is common: in a worldwide study of 69,000 adults, 70% of the responders reported lifetime exposure to a traumatic event, and a striking 30.5% were exposed to four traumatic events or more.

The National Comorbidity Survey Replication (NCS-R) surveyed a nationally representative sample of 9,282 Americans aged 18 years and older. PTSD was assessed among 5,692 participants, using DSM-IV criteria. The NCS-R estimated the lifetime prevalence of PTSD among adult Americans to be 6.8%. Current past year PTSD prevalence was estimated at 3.5%. The lifetime prevalence of PTSD among men was 3.6% and among women was 9.7%. The 12-month prevalence was 1.8% among men and 5.2% among women.

Despite the high prevalence of trauma exposure, only a minority of those exposed develop PTSD. The *conditional prevalence* of developing PTSD upon trauma exposure varies by trauma types and gender: from 65% for men and 46% for women following sexual assault, to 2% for men and 22% for women following physical assault. The conditional prevalence of PTSD following gender-neutral accidents is similar in males and females. Approximately 17% to 33% of people with acute PTSD go on to develop chronic PTSD, with survivors of interpersonal violence taking longer to remit. A minority of trauma-exposed individuals experience delayed onset of PTSD. Importantly, most individuals with PTSD (>50%) do not seek treatment. Furthermore, only 58% of individuals with PTSD in treatment are seen by a psychiatrist, clinical psychologist, or other mental health professional. Importantly, more than a third of individuals with PTSD have been found to have persistent PTSD (even for many years), regardless of treatment received.

Traumatic events among children trigger PTSD as well as a wide array of mental and developmental disorders. In a 2019 population-representative birth-cohort study of 2,232 children in England and Wales, 31.1% of the participants reported trauma exposure by age 18, and 7.8% had PTSD. Trauma-exposed children also reported major depressive episodes (29.2%), conduct disorders (22.9%), self-harm (25.0%), and suicide attempts (8.3%), and those with PTSD reported more frequent major depressive episodes (54.7%) and a staggering 20.1% experienced suicide attempts. Similarly, in a 2007 representative population sample study of 1,420 children in the United States, two-thirds of children had experienced at least one traumatic event by age 16, with 13.4% of these children developing some PTSD symptoms; fewer than 1% met DSM-IV PTSD diagnostic criteria. The 6-month prevalence of PTSD in a household probability sample of 4,023 *adolescents* between the ages of 12 and 17 was estimated to be 3.7% for boys and 6.3% for girls.

Prevalence of PTSD in Veterans

The National Vietnam Veterans Readjustment Study (NVVRS), conducted between 1986 and 1988, interviewed 3,016 American Vietnam Veterans and civilian controls selected to provide a representative sample of those who served in the armed forces during the Vietnam era. The estimated lifetime prevalence of PTSD among these veterans was 30.9% for men and 26.9% for women. Among Vietnam theater veterans, 15.2% of males and 8.1% of females were currently diagnosed with PTSD at the time the study was conducted. A 25-year follow-up of the NVVRS, the National Vietnam Veteran Longitudinal Study (NVVLS), was conducted between 2011 and 2013, assessing the course of PTSD 40 years after the war. The estimate for current DSM-5 warzone PTSD was 4.5%, 10.8% for current plus subthreshold warzone PTSD, and 17.0% for lifetime warzone PTSD.

Comorbid major depression occurred in a third of veterans with current and subthreshold warzone PTSD. Over the past 25 years, 16% of theater veterans reported a greater than 20-point increase, and 7.6% reported a greater than 20-point decrease in PTSD symptoms. The Health of Vietnam-Era Women's Study estimated that 15.9% of female veterans had current PTSD 40 years after the Vietnam War. In 2008, the RAND Corporation estimated the prevalence of PTSD to be 13.8% in a household telephone survey of a representative sample of Iraq and Afghanistan veterans. These findings underscore the need for mental health services over many decades for Vietnam veterans with PTSD symptoms and the urgent need to provide care for a new generation of warfighters.

Prevalence Following Mass Trauma and Disasters

The population prevalence of PTSD among communities exposed to mass-trauma and disaster has been estimated by studies of the World Trade Center attacks of September 11, 2001, the 1995 Oklahoma City bombing, and the 2005 “Katrina” hurricane disaster in New Orleans. Shortly after trauma exposure, studies found that 4% to 7% of all residents had probable PTSD. These numbers declined from 1% to 2% after 9 to 12 months, showing both a general trend toward community recovery and a significant proportion of persistently affected individuals. Higher prevalence of PTSD was observed among rescuers, directly exposed survivors, and low-income minorities.

Contrasting with the formers’ limited impact phase, recent COVID epidemics (SARS, MERS, and COVID-19) are prolonged, deadly, and lingering events. In line with DSM-5 inclusion among potentially traumatic events of “*repeated or extreme exposure to aversive details of the traumatic event(s),*” COVID virus trauma concerns both afflicted individuals, first responders, frontline healthcare workers, and threatened communities and combine both strictly postexposure, residual symptoms, and peritraumatic responses to ongoing threat.

Estimates of COVID-19 PTSD prevalence vary by type of exposure. Stressful exposure among health-care personnel, for example, include managing critical medical situations, witnessing agony and death, operating in crowded settings, and daily exposure to life threat. PTSD symptoms of concern were found in over 20% of health care workers during COVID-19 infection waves. Among survivors of critical medical conditions and ICU admission, 30.2% reported severe PTSD symptoms 30 to 90 days after discharge, and 17% had major depression. Delirium during treatment and residual symptoms contributed to PTSD severity.

ETIOLOGY

Risk Factors for PTSD

In considering risk factors for PTSD, one should separate those informing PTSD risk in the general population, those specifically affecting trauma-exposed individuals (e.g., combat veterans, disaster survivors), and those affecting exposed individuals who develop early PTSD symptoms. Population risk factors encompass exposure risk, propensity to develop PTSD upon exposure, and postexposure PTSD memory consolidation and persistence factors. Risk factors among trauma survivors are more focused, and often of greater clinical relevance.

In a meta-analysis of clinical, biographical, and epidemiologic risk factors for PTSD in the community, it was reported that female gender, age at trauma, race, lower education, childhood abuse, greater severity of trauma exposure, lack of social support, and additional life stress increased risk for PTSD. A second meta-analysis identified seven predictors: prior trauma, prior psychological adjustment, family history of psychopathology, greater perceived life threat during the trauma, lower post-trauma social support, greater emotional distress during exposure, and greater dissociation during exposure. A large epidemiologic study found that exposure to interpersonal violence and have experienced four or more traumatic events carry a higher PTSD risk.

Among trauma-exposed survivors, a meta-analysis of risk factors for combat-related PTSD identified lower education, nonofficer ranks, army service, combat specialization, higher numbers of deployments, longer cumulative length of deployments, more adverse life events, prior trauma exposure, and prior psychological problems as PTSD risk factors. Aspects of deployment that increase the risk of PTSD include higher levels of combat exposure, discharging a weapon, witnessing someone being wounded or killed, and other deployment-related stressors. Another study reported that killing enemy combatants, prisoners of war, and civilians in the warzone increased the risk of PTSD.

Risk factors should be differentiated from risk indicators; the former being mechanistically related to PTSD pathogenesis and potentially informing PTSD prevention, whereas the latter informing PTSD likelihood but unrelated to its etiology (e.g., early symptoms). A longitudinal study of 2,473 civilian trauma survivors identified early PTSD symptom severity as a reliable risk indicator of chronic PTSD, and female gender, lifetime exposure to personal violence, and lower education as potent pre-exposure modifiers of the risk inferred from early symptoms.

Because the etiology of PTSD is complex and multifactorial, involving interacting biologic, personal, and psycho-social contributing factors, the contribution of a single risk factor (e.g., gender) hinges on the presence of other interacting factors (e.g., trauma type). A recent longitudinal machine learning study, for example, found that cortisol levels shortly after trauma exposure interact with lifetime history of child abuse to increase PTSD likelihood.

A few of the above-mentioned risk factors can be modified and are thereby amenable to preventive interventions. Interpersonal conflict, including disorganization and severance of attachment ties after trauma exposures are such “malleable” risk factors and should be addressed during a disaster’s aftermath. Several biologic risk factors, such as low cortisol levels after trauma exposure or high adrenergic drive, have been the subject of early prevention studies with limited success.

Genetics and Risk for PTSD

Individuals’ genetic propensity to develop PTSD has been intensely researched during the last decades. The initial hopes to uncover a single, major genetic PTSD vulnerability trait have been replaced by a polygenic minor effects approach, and more recently by emerging attempts to evaluate the cascading effects, leading from genetic endowment, environmental pressure, gene expression, protein synthesis, brain cellular, and connectivity effects, and, ultimately, clusters of PTSD symptoms. Consequently, while a portion of PTSD risk is inherited, the specific mechanisms linking genetic endowment to expressing PTSD remain one of biologic psychiatry’s greatest challenges.

Based on early twin studies of Vietnam Veterans, it is estimated that 30% of the variance in the risk for PTSD is determined by genes alone. Researchers who studied 222 monozygotic and 184 dizygotic twin pairs reported PTSD symptoms to be moderately heritable with the remaining variability accounted for by unique environmental experiences. A number of common genetic variants have been associated with PTSD, including polymorphisms in FKBP5, PACAP1, COMT, DRD2, GABA alpha-2 receptor, G protein signaling 2 (RSG2), an SNP in an intergenic region of the fourth chromosome, and an estrogen response element on ADCYAP R1. In addition, the s/s genotype of the serotonin transporter gene interacts with childhood adversity to increase PTSD risk.

A recent genome-wide association study (GWAS) of **over 2,50,000 participants of the Million Veteran Program** identified a tight coherence of a PTSD symptom factor that shares genetic variance with internalizing (mood-anxiety-neuroticism) symptom factor often observed in PTSD. The study further identified an enrichment in tissue-specific genetically regulated gene expression in several cortical and subcortical regions related to PTSD pathophysiology and pathogenesis. These results validate the biologic coherence of the PTSD syndrome, inform its relationship to comorbid anxiety and depressive disorders, and provide new brain tissue considerations for treatment.

The genetic vulnerability for PTSD is shared with other mental disorders: a study of **Vietnam veteran twin pairs reported that monozygotic** twins of veterans with combat-related PTSD had more mood disorder symptoms than monozygotic cotwins of combat controls or dizygotic cotwins of veterans with PTSD. Recent studies by **the Psychiatric Genetic Consortium PTSD (PGC-PTSD)** currently evaluate very large samples of trauma-exposed and PTSD patients. A 2020 PGC GWAS of 32,000 PTSD cases and 100,000 trauma-exposed controls found a significant shared genetic risk of PTSD with other psychiatric disorders. It also found higher heritability in females compared with male individuals. Extending beyond GWAS, the recent PGC studies address PTSD epigenetics, gene expression, imaging, and integrative systems biology with the ultimate goal of elucidating the molecular architecture of this complex disorder.

In a multiethnic cohort of over 30,000 PTSD cases and 170,000 controls, Nievergelt and colleagues, conducted a GWAS of PTSD. Three significant loci were identified, two in European, one in African ancestry, and elsewhere, with heritability estimates varying by sex, ranging from 5% to 20%. Five genes were associated with PTSD in those of European ancestry (CDHHC14, a Parkinson disease gene PARK2, KAZN, TMEM51-AS1, and ZNF813) and five distinct genes in those of African ancestry (LINC02335I, MIR5007, TUC338, LINC02571, and HLA-B). Two additional genes, SH3RF3 and PODXL, were identified from 18,222 protein-coding genes, associated with European but not African ancestry. The two gene sets identified in those of European ancestry suggest a role for the immune system in PTSD. Genetic signals associated with re-experiencing symptoms in this cohort were strongly correlated with markers associated with re-experiencing symptoms in an independent cohort.

DIAGNOSIS AND CLINICAL FEATURES

As mentioned previously, the diagnosis of PTSD relies on identifying a traumatic event and confirming the presence of symptoms belonging to four diagnostic criteria: intrusion, avoidance, negative alteration in cognition, and hyperarousal.

To a degree, PTSD symptoms, despite their apparent heterogeneity, reflect an inner logic and theoretically meaningful functional interrelations. Traumatic events remain lively and ever-distressing, they are experienced as personal defeat, surrender to overwhelming stress, or failure to react properly, hence their endlessly distressing rumination and lack of emotional resolution. Because they remain extremely painful, utterly uncontrollable and persistent, individuals with PTSD avoid stimuli that can provoke this reliving of them, including inner states of mind, thoughts, emotions, and bodily experiences. Fending threat, either external or internal, becomes an imperative priority and leads to always remaining on guard, ever more vigilant, attending to and seeking signs of potential threat, and inability to allocate attentional resources needed to deal with real-life challenges. Ultimately, PTSD patients may feel defeated by the intensity, uncontrollability, and timelessness of their post-traumatic distress and flashbulb memories, and their Sisyphean attempts to escape their predicament or regulate their emotions, and thus are likely pessimistic, moody, and distant.

Importantly, post-traumatic psychopathology is not limited to PTSD, and milder and transient reactions are more frequent than lingering PTSD. Even within PTSD, the disorder can be expressed in a myriad of ways and with different configurations of its qualifying symptoms. PTSD, therefore, should be thought of as a prototypical reaction, a psychobiologic trap in which what, during healthy trauma recovery, helps in elaborating and overcoming the traumatic experiences (e.g., recall and reappraisal of traumatic experiences, self-protection, and adequate attention to existing stressors) becomes senseless, timeless, intolerable, and inescapable.

Some researchers have suggested that PTSD is best construed continuously rather than categorically. According to this view, PTSD merely represents the endpoint of a stress response continuum, whereas partial PTSD would represent a subsyndromic degree of severity somewhere between no or mild response and full-blown PTSD. Recent work, however, suggests that distress and impairment do not smoothly vary as a function of increasing symptoms; those victims who meet full criteria for PTSD are more impaired than those who merely qualify for partial PTSD, a finding more consistent with a categorical rather than a dimensional construal of the illness. A recent longitudinal study has shown that PTSD symptom severity fluctuates with time in the disorder's first years, during which time survivors eventually lose and then regain their formal PTSD diagnosis. Clinicians should be aware of such fluctuations and of the consequent need to attend to distressed or impaired help-seeking PTSD survivors as clinically required, including those with subthreshold symptoms.

Controversies Concerning PTSD's Validity and Social Origin

PTSD was born at an age of increased sensitivity to social justice, human rights, protective networks, and duties of care. From its inception, therefore, the disorder was associated with advocacy and a sense of mission among many of its proponents. PTSD has also been used extensively to identify and infer psychological damage and claim compensation. Surpassing its intended role as a descriptor of a mental disorder, PTSD has become a social, cultural, and economic phenomenon.

Additionally, because of its captivating simplicity, PTSD was quickly reified. Rather than a descriptor of an underlying condition, PTSD became the condition itself. Referring to PTSD as a real and distinct entity overlooks the difference between a set of symptoms and an underlying condition. This distinction is too frequently forgotten.

By another DSM rule, complex clinical presentations that follow trauma are subsumed as comorbidities that are, coexisting but distinct disorders. PTSD and major depression, for example, overlap in over 40% of the cases at any given time and co-occur in 95% of the cases in a lifetime perspective. PTSD and major depression are nonetheless seen as distinct entities, particularly when it comes to defining research groups and/or therapies. The amount of information lost by anchoring research and practices to categorical entities is hard to estimate.

Similarly, in a typical posthoc attribution error ("after this, therefore because of this"), the association between PTSD and the traumatic event is quickly perceived as causal. Part of the popularity of PTSD is due to its symbolizing such highly desirable, simplified, intuitive causal attribution.

A recent effort to supersede DSM-5 diagnostic categories with a cross-diagnostic grid of neurobehavioral dysfunctional dimensions–the **Research Domaine Criteria (RDoCs**) initiative–has redefined recent PTSD research strategies and goals and fostered studies of PTSD subcomponents. However, the RDoCs system is yet to deliver a compilation of illness dimensions translatable to clinical practice. To fully apprehend the tension between RDoCs dimensional approach and DSM-5 diagnostics, one should remember that PTSD diagnostic criteria are clinical indicators of an underlying and complex neuropathology which, heuristically (that is, toward better problem-solving and discovery), can be parsed into contributing dimensions. The RDoCs tentative approach is such an effort, and as such, it doesn't truly challenge the role of clinical diagnostic practices.

PATHOLOGY AND LABORATORY EXAMINATION

PTSD epitomizes a mental disorder occurring at the junction between external reality and a responsive brain: it is a maladaptive outcome of a host-environment interaction gone wrong. Several clinical features of PTSD inform the underlying CNS mechanisms: the disorder emerges after exposure to traumatic events, and therefore must centrally involve new learning and CNS plasticity. Its core symptoms of re-experiencing and avoidance directly involve reminders of the traumatic event, which implies acquisition, consolidation, and conservation of aversive memories. The presence of hypervigilance, hyperarousal, and cognitive and emotional alterations extend the range of symptoms beyond direct responses to reminders and suggest an overgeneralization of aversive learning and dysregulation of emotions. Acute PTSD symptoms often emerge shortly after trauma exposure and do not remit in survivors who develop PTSD, making PTSD a likely “disorder of recovery,” and evoking alterations of fear extinction and inability to process safety signals. PTSD patients remain exceedingly reactive to ongoing life stressors and thus show altered perception of threat signals and heightened threat sensitivity.

Not surprisingly, studies of the neurobiology of PTSD have been extremely productive. To make sense of that abundant literature, this chapter addresses three layers of findings: (a) how the CNS mediates known PTSD symptoms and related behavior, (b) alteration in neuroregulatory biologic systems (e.g., neuroendocrine, neuropeptide, inflammatory) that moderate CNS responses, and (c) basic neurobehavioral alterations, genetic endowment, and epigenetic regulation, often present before the traumatic event, and increasing CNS' vulnerability to develop PTSD.

In reading that literature, one should bear in mind that, while several biologic perturbations have been associated with PTSD, none has proven to be characteristic of the disorder in all affected individuals or account for the large proportion of the disorder's causation. Such a gap between group-based trends and findings and individual causation characterizes complex and multicausal disorders, in which there are many pathways to an outcome and individual-specific convergences of risk factors. To fully translate the complex biology of PTSD into practical prediction, diagnostics, or therapeutics, one must resort to advanced bioinformatics models, which are currently being developed. This section covers the existing evidence and the few newly discovered links between layers of biologic alterations and environmental pressures.

Psychophysiologic Studies

Bearing different names (e.g., Soldier Heart; Shell Shock), the historical equivalents of PTSD have been associated with "physical" or "physiologic" symptoms. Studies performed during World War I (1917 to 1918) recorded exaggerated heart rate responses to gunfire, sulfuric flames, epinephrine injection, and CO_2 inhalation among war veterans with combat stress reactions. Abraham Kardiner, whose description of War Neurosis inspired the DSM-III definition of PTSD, defined these conditions as "physio-neuroses," involving bodily alarm responses to stressful combat reminders. Shortly after the inclusion in DSM-III, PTSD was equated with a "conditioned emotional response," similar to those observed in animal experiments of fear conditioning. Studies performed established the script-driven imagery protocol (having PTSD patients listen to a verbal description of their traumatic events while measuring their heart rate, perspiration, and facial muscle tension) as a standardized and reliable procedure to elicit physiologic responses to individual trauma memories and showed that these responses were significantly higher in trauma survivors with PTSD, relative to those who did not develop PTSD. Another line of studies, initiated in Jerusalem in 1992, showed heightened auditory startle responses in PTSD.

Psychophysiologic measurements added an objective dimension to PTSD. As such, they have helped deflect some of the initial skepticism regarding the disorder’s validity. They have also opened ways to experimentally elicit PTSD symptoms, which was later used in functional brain-imaging “symptom provocation” experiments (refer to *Functional Brain Imaging*). Psychophysiologic measurements have also shown that PTSD patients fail to extinguish conditioned fear responses, which is in line with their inability to recover from the early responses to traumatic events. Studies have also shown that PTSD patients fail to retain extinction, once they have acquired it, and return to reacting to adversity despite having safely recovered from it. Additionally, studies indicate an association between elevated heart rate shortly after traumatic events (e.g., during emergency department admission after injury) and subsequent PTSD, an association that reflects the intensity of the unconditioned responses (UCRs) during trauma exposure. Finally, decreased heart rate variability (HRV) has been robustly associated with PTSD and linked to emotional cognition and emotion regulation.

CNS Mediation of PTSD Clinical Manifestations

Several brain circuits have been implied in the mediation of PTSD clinical manifestations Those include *fear learning, emotion regulation, executive control,* and *contextual learning* (see Fig. 17–1). Experimental methods used to explore these circuits encompass psychophysiology, neurocognitive assessments, and brain imaging.

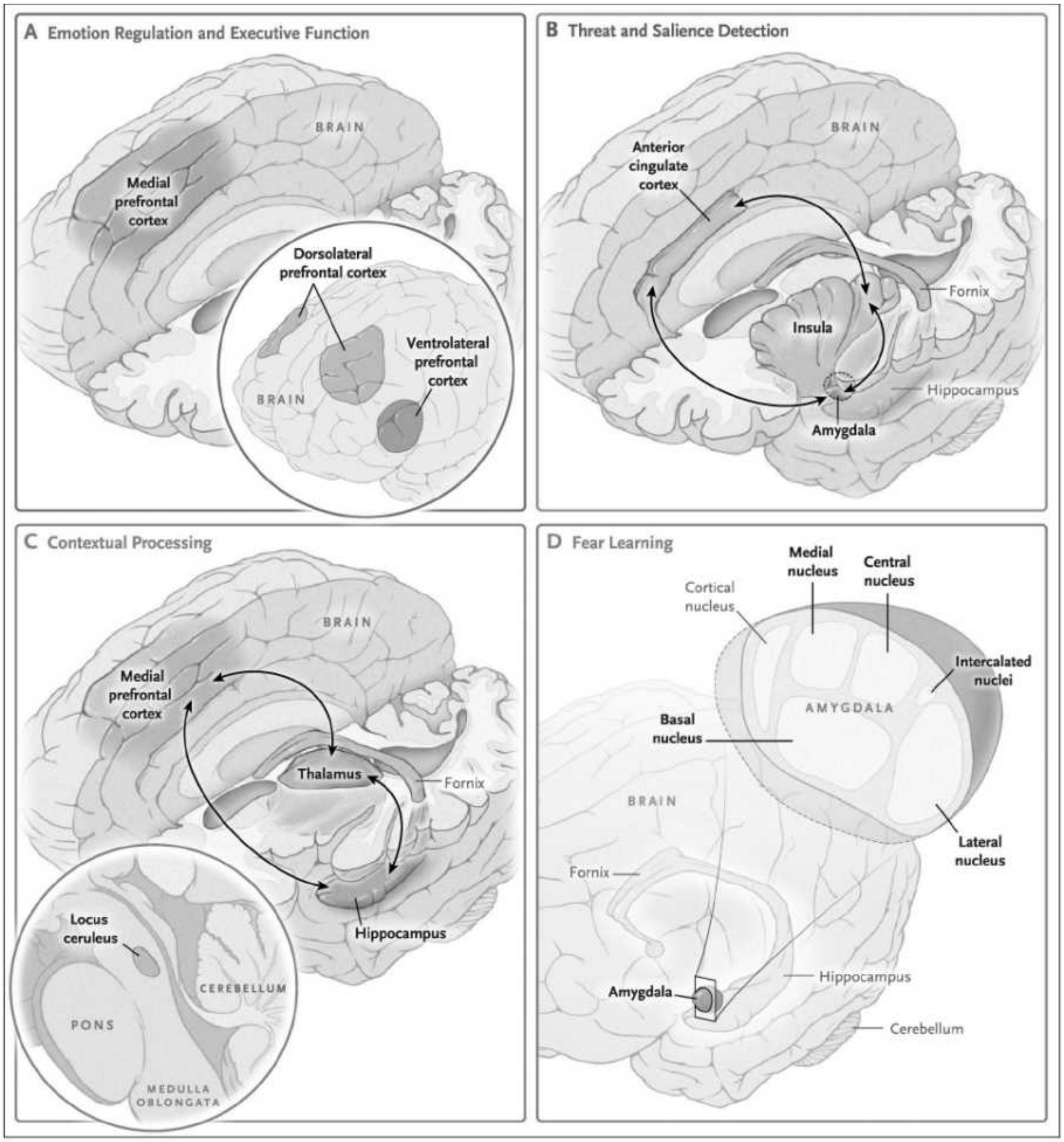
A Emotion Regulation and Executive Function
BRAIN
Medial prefrontal cortex
Dorsolateral prefrontal cortex
Ventrolateral prefrontal cortex
BRAIN
B Threat and Salience Detection
BRAIN
Anterior cingulate cortex
Fornix
Insula
Hippocampus
Amygdala
C Contextual Processing
BRAIN
Medial prefrontal cortex
Thalamus
Fornix
Hippocampus
Locus ceruleus
CEREBELLUM
PONS
MEDULLA OBLONGATA
D Fear Learning
Medial nucleus
Central nucleus
Cortical nucleus
Intercalated nuclei
AMYGDALA
Basal nucleus
BRAIN
Lateral nucleus
Fornix
Hippocampus
Amygdala
Cerebellum

FIGURE 17–1. Brain regions implicated in the pathophysiology of posttraumatic stress disorder (PTSD). Shown are the known connectivity paths within four dysfunctional circuits that play a part in the psychopathology of PTSD: emotion regulation and executive function, threat detection, contextual processing, and fear learning. (From Shalev A, Liberzon I, Marmar C. Posttraumatic stress disorder. *NEJM.* 2017;376(25):2459–2469. Copyright © 2017 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society. doi:10.1056/NEJMra1612499.)

Fear Conditioning Theory of PTSD. This model equates the acquisition and the maintenance of PTSD symptoms with the consolidation and extinction of conditioned fear responses. Within this model, the traumatic event is the conditioned stimulus (CS), the immediate responses to trauma are UCRs, and stimuli (sights, sounds) present during the traumatic event acquire the ability to elicit conditioned responses (CRs) resembling the UCR in the absence of real threat. A series of psychophysiologic studies has buttressed the fear-conditioning model of PTSD by consistently showing intense physiologic responses to trauma reminders in PTSD.

CNS Circuits Involved in Fear Conditioning. *Fear conditioning* and *extinction* are adaptive learning processes. The neural circuitry that mediates these processes is essential to survival and conserved across species. It commands the body's hard-wired, involuntary, instantaneous defensive responses to imminent threat (a "species-specific defense response"). This highly conserved stress response system includes sympathetic and parasympathetic system activation, cardiovascular, and respiratory reactions, activation of the hypothalamic–pituitary–adrenal (HPA) axis, defensive fight, flight, and freezing behaviors, and changes in information processing. It is orchestrated by the central nucleus of the amygdala, a brain structure that connects to critical brain areas responsible for response execution. The activity of the central nucleus is modulated by layers of midbrain nuclei capable of filtering and evaluating threat stimuli (such as the basolateral amygdala [BLA]) and further by cortical and subcortical structures, such as the hippocampus, insula, and prefrontal structures. Each of these modulatory layers is highly plastic and therefore subject to new learning throughout life. PTSD is thought to involve new learning in these modulatory systems as follows.

During exposure to extreme threat, previously neutral stimuli overlapping in time with threat stimuli pass through the thalamus to the basolateral nucleus of the amygdala, which activates the central nucleus of the amygdala to enable defensive responding. The association between threat or "unconditioned" stimuli (US) and the previously neutral but currently CS is made in the basolateral nucleus of the amygdala. Once formed this association enables an exposure to the CS to trigger a defensive response. The magnitude of the defense response is modulated by brainstem serotonergic, noradrenergic, and dopaminergic neurons in the ventral tegmental area, dorsal and median raphe nuclei, and locus coeruleus. These monoamines affect both the magnitude of UCRs and the consolidation of US–CS associations. The balance of neurotransmitters and neuromodulators activated during unconditioned response also influences the extent to which contextual stimuli, more loosely associated in time with the US, may contribute to later overgeneralization of threat stimuli.

Activity within the amygdala is under inhibitory control of the prefrontal cortex (PFC). Specifically, glutamatergic neurons from the PFC activate inhibitory interneurons within the BLA, which in turn suppress BLA outputs to the Central Nucleus of the Amygdala. The monoaminergic effects are further modulated by neuroactive steroids and neuropeptides released by the adrenal gland during stress that further magnify, mitigate, or otherwise shape fear acquisition, extinction, and recall.

Consolidation, Extinction, Extinction Recall, and Reconsolidation. Associations encoded in the amygdala during traumatic events tend to be robust and resistant to change. Their behavioral expression, however, is subject to extinction, which typically occurs when the CS is no longer associated with threat in a safe environment. Neuronal mechanisms underlying extinction include PFC inhibitory input to the BLA. Extinction involves active learning and as such is subject to recall and forgetting. Prolonged stress after the traumatic event can compromise the PFC ability to efficiently extinguish fear memories, whereas soothing interpersonal contacts and a safe environment may enhance it. In the aftermath of traumatic events, therefore, high levels of continuing stress and absence of soothing contacts may decrease the likelihood of efficient extinction and contribute to the risk of developing PTSD.

Threat Detection and Emotion Regulation

Other specific circuits implicated in PTSD development mediate *threat information processing,* mediated by the amygdala, dorsal Anterior Cingulate (dACC), and insula cortex, and *regulatory control mechanisms* that involve the hippocampus and medial and lateral PFC regions (dMPFC and dLPFC). These regions undergo plastic modifications upon exposure to potentially traumatic events, which in turn modify future processing of stressful encounters. The amygdala assigns rapid “threat-potential” tagging of environmental cues, and its high reactivity may be a risk factor for PTSD. The amygdala-associated PFC regions and insula have been implicated in detecting stimulus saliency and biasing attention to threat. *Emotional regulation* circuits select appropriate response to threat—including emotional response—and involve the medial prefrontal cortex (MPFC) and dorsolateral prefrontal cortex (DLPFC). Evidence from neuronal network studies demonstrates segregation among a *salience network* (SN) involving regions of the insula, ACC, and amygdala, an *attention network* (AN) involving the DLPFC and VLPFC, and a *default mode network* (DMN) including the MPFC and hippocampus.

Neuroimaging findings in PTSD patients found abnormalities in threat detection and emotion regulation regions with increased reactivity of the extended amygdala, insula, and dorsal ACC regions, and decreased regulatory activity in subregions of the prefrontal lobe. In both PTSD and MDD, fMRI studies have shown excessive reactivity to potential threat stimuli in the amygdala and insula and hypoactivation of the PFC. Prefrontal deficits involving ventromedial prefrontal cortex (vmPFC) have been implicated in emotion regulation during changing environmental demands in PTSD, MDD, and generalized anxiety disorder. Resting state connectivity studies in PTSD similarly show reduced network coupling within the DMN, greater coupling within the SN, and decreased segregation of the DMN and the SN, suggesting an overengagement of saliency detection in PTSD. Recent studies further suggest that trauma exposure alters DMN connectivity shortly after its occurrence.

Context Appraisal and Update

Among the core clinical features of PTSD is the persistence of fear-driven avoidant behavior, heightened reactivity to reminders, and hypervigilance despite reaching safety and living out of danger. Such symptoms have been interpreted as reflecting an impaired capacity to update one’s contextual appraisal from threat to safety. Inappropriately and persistently interpreting situations and signals as threatening and dangerous and difficulty in differentiating safety from threat may result from improper contextual update processing. Contextual processing is mediated by the vmPFC, the hippocampus, and the thalamus and requires low adrenergic drive during sleep. A diminished functional activity of the vmPFC in patients with PTSD has been linked to abnormal processing of contextual information, as well as to impaired extinction recall.

Cognitive Control, Memory, and Attention

Salient PTSD symptoms include inability to concentrate, memory disturbances, and difficulties planning and monitoring one’s activity. Activity in dACC and frontoparietal attentional networks have been implicated in cognitive control, including performance monitoring, response inhibition, task shifting, working memory, attention, and memory. This type of regulation also disengages medial prefrontal–medial parietal DMN regions involved in internally oriented, self-directed mentation. DMN regions are overactive or underdynamic in PTSD, resulting in poorer ability to disengage from threat responses and discriminate truly threatening stimuli from remote reminders of traumatic experiences.

Modulatory Processes Underlying PTSD Neuropathology

Neuroendocrine Modulation. Several “stress” hormones have been implied in the etiology, pathogenesis, and psychopathology of PTSD, but none are disorder-specific and together they exercise complex and sometimes opposing neuro-modulatory effects. The HPA axis and sympathetic nervous system are important stress-reactive systems and thus central to understanding the disorder. They contribute to brain information processing, reactivity, and behavior in response to threat. Briefly stated, stress pathways involving cortisol signaling via the HPA axis are altered with blunted cortisol responses to stressors as a result of enhanced negative feedback from increased glucocorticoid receptor sensitivity in PTSD, with consequent influences on the HPA axis, sympathetic nervous system, brainstem, amygdala, and potentially accounting for co-occurring diagnostic patterns (PTSD, depression, chronic pain, refractory substance dependence).

Neurosteroids. Cortisol is centrally involved in stress response, and its regulation in PTSD has been extensively studied. Schematically, a surge in cortisol secretion by the adrenals follows the initial adrenergic response to stress. This surge, then, downregulates the brain’s initial adrenergic response, via negative feedback mechanism, thereby attenuating the acute response to stress and the related US–CS pairing. Early studies revealed a pattern of low cortisol output and increased sensitivity to glucocorticoid negative feedback in PTSD, but the finding has not been consistent across samples of PTSD patients, and during the acute aftermath of traumatic events. Sources of the variability in these studies include different experimental designs, population characteristics, trauma characteristics (e.g., single occurrence vs. prolonged exposure [PE] to stressors) and variation in individual genetic endowment. Differences between findings may, therefore, reflect *differences between individuals' HPA axis sensitivity,* which accounts of whole groups are unable to discern. In other words, altered HPA axis response to stress may affect some individuals (e.g., those with history of child abuse, variation in genetic endowment, or different epigenetic regulation history) but not others. It is also important to evaluate the roles of cortisol in the pathophysiology of PTSD in the context of other relevant neurohormonal responses as follows:

Neuropeptide Y (NPY) has anxiolytic and stress resilience effects in numerous animal models and human studies. It stimulates neurogenesis in the hippocampus and as such may support recovery from stress. For example, higher plasma NPY levels during distressing mock captivity training were associated with stress resilience. NPY has also several bodily trophic effects including efficient restoration of homeostasis, improved sleep quality, lower heart rate response to adrenergic stimulation, and better energy balance, all of which may affect the recovery from early PTSD symptoms.

Allopregnanolone/Pregnanolone are neuroactive steroids that selectively enhance the effect of the brain principal inhibitory neurotransmitter GABA. They also have neuroprotective effects such as increased myelination, decreased apoptosis (programmed cell death), and increased neurogenesis. Studies have shown lower levels of these hormones in PTSD and depression and negative correlations with symptom severity.

Dehydroepiandrosterone (DHEA) is the immediate precursor for androgen. It is secreted from the adrenal gland along with cortisol in response to adrenocorticotropic hormone (ACTH). DHEA antagonizes $GABA_A$ receptors and facilitates *N*-methyl-D-aspartate (NMDA) receptor function. In humans, an increasing DHEA level under stress appears to confer psychological and neurocognitive resilience and may protect against negative health outcomes.

Catecholamines. Exaggerated noradrenergic signaling is related to deficits in frontal lobe executive function resulting in undermodulated amygdala reactivity as well as sleep fragmentation and nightmares, which are common in PTSD. An exaggerated adrenergic response during trauma exposure and in the hours following exposure may contribute to the formation of more durable emotional memories. Evidence concerning the presence of sympathetic system hyper-reactivity in PTSD has been generally consistent. Noradrenergic hyper-reactivity may be a pretrauma vulnerability trait in some individuals or may develop as a result of trauma exposure in others. Factors that may contribute to the increased release of norepinephrine (NE) in response to traumatic stress or conditioned reminders of the trauma include decreased expression, affinity, or function of alpha 2-adrenergic auto receptors, or genetic or stress-induced decrements in NPY. Factors that may enhance amygdala reactivity may include a downregulation of intra-amygdala synaptic GABA receptors or reductions in GABAergic neuroactive steroids such as allopregnanolone.

Illustrating the role of catecholamines in maintaining PTSD symptoms, the alpha 2-antagonist yohimbine induces extreme hyperarousal, flashbacks, vivid intrusive memories, panic, and increases in heart rate and systolic blood pressure in male combat veterans with PTSD. Yohimbine also suppresses frontal lobe metabolism to a greater degree among PTSD patients. Resulting from these observations, alpha 1-adrenergic postsynaptic receptor antagonists, such as prazosin, have been used for the treatment of nightmares in PTSD and evaluated as an early prevention of PTSD (see “Treatment” section).

Dopamine release in the amygdala promotes the expression of unconditioned and conditioned stress responding. Additionally, dopamine may inhibit PFC projections into the BLA and thus diminish the inhibitory control over learned fear responses and lead to increasing hypervigilance and sensitivity to environmental cues. Dopamine is also highly relevant to reward signaling in the nucleus accumbens, a process that appears to be downregulated in PTSD (see also PTSD criterion D symptoms).

Neuropharmacologic treatment and genetic epidemiologic studies suggest that the serotonin (5-HT) system impacts both PTSD risk and symptom severity. Administration of meta-chlorophenylpiperazine (mCPP), which interacts with multiple 5-HT receptor subtypes and with the 5-HT transporter to release 5-HT, provokes anxiety, panic attacks, and PTSD symptoms, including flashbacks, and cognitive changes that are reversible by mixed 5-HT1c/5-HT2 antagonists. Selective serotonin reuptake inhibitors, however, do not reduce the likelihood of developing PTSD when administered shortly after trauma exposure.

Genetic Contribution to PTSD Pathophysiology

The genetic contribution to PTSD is multifactorial, person-specific, and significantly affected by early life experiences. The inheritance of PTSD is polygenic, with no single gene variant explaining more than a small percentage of the variance. Nonetheless, a large twin study has shown significant (>30%) inheritance of PTSD and PTSD symptoms, but reanalyses of the same database reported substantial genetic overlap between MDD and PTSD. As in other areas of psychiatry, the translation of heritability to specific gene variants is a work in progress, and the distance between genetic endowment and clinical expression is currently insurmountable. Studies to date have, therefore, focused on gene variants that affect neuronal and neuro-modulatory processes leading to PTSD.

Gene variants in the *NPY and noradrenergic signaling* pathways have been found to impact physiologic processes of relevance to PTSD, such as heart rate, blood pressure, NE release, and negative emotional reactions in response to low-dose yohimbine. These sympathetic system stress reactions would be expected to increase risk for PTSD, but whether any of these latter gene variants influence risk for PTSD is yet to be examined. Similarly, to the extent that the BDNF gene affects synaptic plasticity, neurogenesis, and cognition, it might indirectly affect the likelihood of developing PTSD or remaining resilient in the aftermath of traumatic events. Numerous genes have polymorphisms that affect corticotropin-releasing factor (CRF), ACTH, or cortisol responses to stress (e.g., polymorphisms of genes for the mu-opioid receptor, catechol-O-methyl-transferase [COMT]) and angiotensin I–converting enzyme (ACE-I), the glucocorticoid receptor, the ACTH receptor, and CRF and the CRF receptor.

Gene–environment interactions that affect PTSD risk have been associated with the short allele of the serotonin transporter gene promoter region (SLC6A4). In a group of Rwandan genocide refugees, the SLC6A4 gene promoter region polymorphism was found to increase the risk for PTSD development in the context of relatively low-trauma load. In 582 European American and 670 African American individuals, possessing either one or two alleles of the SLC6A4 promoter polymorphism interacted with adult trauma and childhood adversity to increase the risk for PTSD, especially for those with high rates of both types of trauma exposure. Similarly, both childhood trauma *and* FKBP5 lower expression are required to increase PTSD risk.

It is currently clear that *PTSD is a disorder with a highly polygenic architecture.* Thus, to capture a substantial portion of the genetic contribution, effects from many gene variants need to be aggregated. Recent studies of polygenic risk score (PRS) offer a new way to use genetic endowment to infer PTSD risk. PRS is an estimate of an individuals' genetic liability to develop a disorder, given their genotype profile and available GWAS. PRSs also assess shared etiology between phenotypes and evaluate the clinical utility of genetic data for complex disease. The Psychiatric Genomics Consortium—Post-traumatic Stress Disorder Group has recently reported an analysis of 20,730 trauma survivors with or without PTSD in which they found a PTSD polygenic risk profile that overlapped with that for schizophrenia.

Epigenetic Changes Associated with PTSD

The epigenetic regulation of gene transcription may provide new opportunities for understanding PTSD. Epigenetic effects consist of coordinated modification of gene promoters and chromatin such that access to the gene by regulatory elements is either enhanced or limited, including lower methylation of the glucocorticoid receptor gene F1 promoter in peripheral blood cells in veterans. Of note, research has shown differences between groups of persons with and without PTSD in methylation patterns of DNA in white blood cells. A cluster of unmethylated immune system function genes also characterized PTSD, which otherwise showed a greater number of methylated genes in general. In addition, there were group differences in methylation of the dinucleotide methyl transferase (DNMT) type 3 beta and 3L genes, which are involved in *de novo* rather than maintenance methylation, suggesting that specific adaptive epigenetic capacities themselves may be altered in PTSD.

The extent to which epigenetic mechanisms, operating after trauma exposure, underlie some of the persistent effects of exposure is an exciting new perspective. Research showed that immobilization stress in rodents *reduced* expression of BDNF gene transcripts I and IV in the hippocampus in association with decreases in the level of histone acetylation of the P1 and P4 promoters. Conversely, research has shown that extinction of conditioned fear was accompanied by an *increase* in the levels of histone acetylation of the gene promoters for BDNF transcripts I and IV. Together this work suggests that activation of specific PFC circuits *in combination* with agents that facilitate epigenetic processes that enhance synaptic plasticity and stabilization of new neuronal circuits may contribute to the persistence of PTSD and recovery from PTSD. Recent studies have shown that several epigenetic features of PTSD are modified after successful treatment or following spontaneous recovery. A successful treatment of PTSD was accompanied by significant changes in DNA *methylation at 12 differentially methylated regions in one study.*

Structural and Functional Neuroanatomy

Structural Brain Imaging of PTSD. The first reported structural abnormality in PTSD was a reduced left hippocampal volume in Vietnam veterans with PTSD. This finding had been interpreted as reflecting the hippocampus' vulnerability to excessive stress hormone levels due to prolonged stress. Numerous studies subsequently reported left, right, or bilateral hippocampal volume reduction in PTSD. Other studies, particularly in younger veterans, have failed to replicate this finding, but meta-analyses generally support the existence of reduced hippocampal volume in PTSD. Findings of smaller hippocampus volume in PTSD, however, lack disorder specificity (similar findings have been shown in schizophrenia and affective disorders). There is also a controversy regarding their relationship with PTSD. Originally thought to be a consequence of developing PTSD due to a neurotoxic effect of protracted or "allostatic" stress, a longitudinal study of hippocampal volume in early PTSD has not shown a hippocampus shrinking effect and an elegant twin study of Vietnam veterans found smaller hippocampi in unexposed identical twins of war veterans with chronic PTSD, thereby positioning smaller hippocampi as a vulnerability factor, rather than a consequence of PTSD.

More recent studies reported white matter volume reductions in the corpus callosum and gray matter reduction in the amygdala and the insula. The evidence concerning structural alterations of the amygdala is quite weak. Two meta-analyses reached somewhat different conclusions. One found that there is overall evidence for smaller left amygdala volumes in adults with PTSD, but not in pediatric PTSD, whereas the other did not find significant volumetric alteration in either population (Fig. 17–1).

Several studies have pointed to abnormal structural characteristics in the *anterior cingulate cortex (ACC),* including reduced ACC gray matter volumes or white matter abnormalities of the cingulum. These reduced volumes may be particularly characteristic of the rostral and subcallosal portions of the ACC, rather than the dorsal ACC. Rostral ACC volume has been found to predict response to cognitive behavioral interventions such that larger volumes were associated with greater symptom reduction. A study of pregenual ACC volumes in monozygotic twins discordant for PTSD suggested that reduced volumes might reflect an acquired trait of PTSD rather than a vulnerability factor.

Functional Brain Imaging. Functional neuroimaging studies of PTSD include both resting and activation studies. The latter include symptom provocation and processing of both emotional and neutral stimuli. Within the limbic core, differences between PTSD and controls have been reported for the amygdala, and less so for the hippocampus. Outside the limbic core the vast majority of reported differences in activation between PTSD and controls occur in the posterior, dorsal, and anterior cingulate, occipital and MPFC, and dorsolateral PFC. The same overall pattern can be detected for neutral tasks: most abnormal activations appear to occur in the dorsal trend. There are, nonetheless, several differences: the rostral ACC and subcallosal ACC do not show unique patterns of activity in neutral conditions, and only the dorsal ACC appears to be recruited differently in PTSD compared with controls. There is also less consistency across studies, which may reflect the more varied tasks and smaller number of published studies.

Sleep

Up to 90% of people with PTSD report sleep disturbances such as nightmares and insomnia. In studies examining the physiologic basis for sleep disturbances in PTSD, both *macrolevel* sleep architecture (i.e., time spent in various sleep stages overnight) and *microlevel* physiologic measures of autonomic system activity during sleep (i.e., heart-rate variability, electrodermal activity) have been implicated. The frequent nightmares that are often the hallmark of PTSD are thought to occur in rapid eye movement (REM) sleep. Additionally, REM has also been shown to play a role in the maintenance of fear memories. Sleep abnormalities also characterize the early phases of PTSD: while it is unclear why some survivors develop PTSD and others do not, persistent sleep abnormalities, which are present in upwards of 66% of those who develop the disorder, have been suggested as a contributing factor. More specifically, recent studies have shown that sleep disturbances, measured prior to exposure to traumatic stress, predict the development of PTSD.

Summary of the Biology of PTSD

Arguably the most intriguing feature of PTSD is its persistence, that is, its nonextinction despite the absence of further stress. Longitudinal studies have confirmed a frequent pattern of immediate expression of PTSD symptoms followed by recovery in most of those exposed, and persistence of symptoms in those who “develop” chronic PTSD. The conditioned fear analogy similarly predicts extinction in the absence of reinforcement, which does not occur in PTSD. A neurobiologic account of PTSD, therefore, must also consider the disorder’s maintaining factors. Several theoretical models account for the persistence of PTSD, including a *kindling hypothesis,* according to which patterns of pathologic neuronal coactivation become “entrenched” by repeated use; an *allostatic stress hypothesis,* presuming a “wear and tear” of CNS emotional-control systems (e.g., the hippocampus) by chronic stress, and what might be referred to as a *subsystem clash* or *truncated response* view, according to which one brain subsystem blocks another from achieving the full processing of its inherent tasks.

DIFFERENTIAL DIAGNOSIS

PTSD cannot be diagnosed in the absence of exposure to a traumatic stressor. Indeed, it is the memory of trauma that generates and knits together the otherwise disparate symptoms of PTSD, many of which overlap with other mood and anxiety disorders. Hence, identifying the etiologically significant trauma is the crucial starting point for differential diagnosis. The symptoms must persist for at least 1 month following exposure to the stressor, thereby distinguishing PTSD from acute stress disorder, whose symptoms, by definition, cannot persist for longer than 1 month. PTSD arises in response to a traumatic stressor, whereas the diagnosis of adjustment disorder is suitable for people who experience anxiety, mood, or other symptoms following exposure to nontraumatic life events, such as loss, separation, or illness.

PTSD comprises two kinds of symptoms: event-related and nonevent-related responses. Intrusive thoughts, nightmares, flashbacks, and avoidance of reminders are event-related and thereby facilitate differential diagnosis by distinguishing PTSD from other syndromes that develop or are exacerbated after exposure to traumatic stressors. For example, OCD intrusive thoughts concern current and future threats, such as contamination. Likewise, generalized anxiety disorder involves envisioning and worrying about threats that lie in the future.

PTSD nonevent-related responses lack direct reference to the traumatic event. These include loss of interest in activities, a sense of foreshortened future, difficulty in experiencing loving feelings toward others, exaggerated startle response, difficulty sleeping, difficulty concentrating, and irritability. These symptoms present two challenges during differential diagnosis. First, the assessor must determine that these symptoms followed in the wake of trauma rather than predating exposure to trauma. Many people have sleep problems, are irritable, and so forth, and one cannot automatically assume that the presence of these problems signifies PTSD. Indeed, a person exposed to a trauma who exhibits irritability, difficulty concentrating, and insomnia must not be assumed to have "PTSD symptoms" unless these difficulties postdate the trauma or are exacerbated by the trauma and are accompanied by trauma-related symptoms. Automatically interpreting insomnia in a trauma victim as a "symptom of PTSD" makes no more sense than calling coughing a "symptom of bacterial pneumonia" in the absence of a confirming culture.

The other challenge arises when assessors must determine whether apparent PTSD symptoms are direct consequences of memory for trauma, or if they signify the presence of another disorder. For example, sleep disturbance, irritability, difficulty envisioning one's future, loss of loving feelings, and loss of interest in activities are common symptoms of MDD—a condition often comorbid with PTSD. Likewise, other symptoms, such as hypervigilance and startle, may occur in other anxiety disorders as well. As a rule, therefore, PTSD requires a fuller presence of typical symptoms and a substantiated link in both content and timing to trauma exposure.

PTSD seldom occurs by itself, and the presence of other DSM-5 mental disorders does not preclude PTSD. For example, researchers studying Vietnam veterans found that 84% of outpatients, as well as inpatients with PTSD, met criteria for comorbid conditions, often alcoholism and antisocial personality. In another study of help-seeking veterans, researchers diagnosed current major depression, panic disorder, bipolar disorder, and social phobia more often among those with PTSD than among those without PTSD. Comorbidity is also extremely common in epidemiologic samples. In the NVVRS, 98.8% of the men who had a lifetime diagnosis of PTSD also qualified for at least one other mental disorder, whereas only 40.6% of those without PTSD did so. The most common comorbid disorders were alcohol abuse, depression, and generalized anxiety disorder. Comorbidity is common in epidemiology studies of civilian PTSD as well. In one study conducted in the Detroit area, about 80% of subjects with PTSD had at least one other disorder at some point during their lives. In a representative survey of the U.S. population, the lifetime comorbidity rates were 88.3% in men and 79% in women with PTSD. Major depression and alcohol dependence (AD) were common comorbid conditions. Retrospectively reported ages of onset suggested that other anxiety disorders often predated PTSD, whereas alcohol and mood disorders usually developed after PTSD.

In cases of comorbidity, clinicians are often challenged by identifying the primary disorder, or, in better terms, the disorder, or the configuration of symptoms that requires the most urgent attention. Studies of PTSD suggest that treating “secondary” substance misuse is critical for addressing PTSD in treatment, ideally prior to initiating PTSD treatment. When not possible, simultaneously treating PTSD and substance use disorders (SUDs) is feasible, thus emphasizing harm reduction when abstinence cannot be achieved. Similarly, PTSD patients with prominent depression, suicidal ideation, severe impulse dyscontrol, or paranoid features should first be protected from the immediate consequences of these symptom configurations.

COURSE AND PROGNOSIS

Immediately following exposure to a traumatic event, many people experience shock, nightmares, intrusive thoughts about the trauma, and so forth. Indeed, it would be surprising if severely trauma-exposed individuals had no emotional reaction whatsoever to horrific events. Accordingly, a diagnosis of PTSD requires symptoms that persist for at least 1 month following the trauma.

For most people exposed to trauma, symptoms do wane. For example, in a study of help-seeking rape victims, researchers found that 95% met PTSD symptom criteria within 2 weeks of the rape. Yet the proportion of victims who still met symptom criteria at 1, 3, and 6 months after the rape declined to 63.3%, 45.9%, and 41.7%, respectively. Among victims of nonsexual assault, 64.7% met PTSD symptom criteria 1 week after the trauma, whereas the proportion still meeting criteria at 1, 3, 6, and 9 months after the assault declined to 36.7%, 14.6%, 11.5%, and 0%, respectively.

Epidemiologic data from the National Comorbidity Study (NCS) indicate that the median time for PTSD to remit is 36 months for individuals who sought help for any mental health problem (not necessarily for PTSD) and about 64 months for individuals who never sought help for a mental health problem. Approximately one-third of those who qualify for PTSD had a chronic course.

A recent review of studies concerning delayed-onset PTSD confirmed that *symptom* onset is almost never delayed. This is a striking contrast to the original claim in the text of DSM-III that symptoms often only surface months or years after the trauma. In some rare cases, a person may develop immediate symptoms but fall just short of meeting full criteria for PTSD. A worsening of symptoms several months or years later may result in the person finally qualifying for the diagnosis. At first glance, this would appear to constitute a delayed onset of the PTSD *syndrome,* but it is not as if the person has been wholly asymptomatic during the period following the trauma and prior to meeting full criteria. For example, in one study of Israeli combat veterans, most cases of apparent delayed-onset PTSD involved either exacerbation of chronic, subsyndromic symptoms or delayed help-seeking in those with chronic PTSD. For example, some veterans finally met full criteria for PTSD after receiving notice that they were about to be reactivated for military service. Studies of war and holocaust survivors have documented an emergence of PTSD symptoms upon aging or during illness. Stressful or traumatic events reminiscent of the original trauma may trigger a re-emergence of disabling PTSD symptoms in individuals previously expressing limited distress or dysfunction.

TREATMENT

Psychological Treatment of PTSD in Adults

A review of 70 studies of psychotherapies involving a total of 4,761 participants showed that many of the studies were rated as being at risk of bias, and sample sizes were generally small with limited follow-up data.

There was evidence that individual trauma-focused cognitive behavior therapy (TFCBT) including PE therapy, which focuses on re-experiencing the traumatic event through repeatedly engaging with the memories (imaginal exposure) and everyday reminders (in vivo exposure) rather than avoiding triggers, eye movement desensitization and reprocessing therapy (EMDR), which includes repeatedly recalling distressing images while receiving sensory inputs, and non-TFCBT including present centered therapy (PCT), which focuses on current relationship and work challenges rather than the trauma, are equally effective immediately post-treatment, with some evidence that TFCBT and EMDR are superior to non-TFCBT between 1 and 4 months following treatment. Individual TFCBT, EMDR, and non-TFCBT were found to be more effective than other therapies. There was evidence of greater dropout in active treatment groups. There is also evidence that cognitive processing therapy (CPT), which emphasizes correcting faulty attributions including post-traumatic overgeneralizations of the world as dangerous, uncontrollable, and unpredictable, is effective for adults with chronic PTSD. A recent study in adults with PTSD reported that IPT, focusing on role expectations in dyadic relationships, was comparable to PE for treatment completers, with fewer dropouts in the IPT group for participants with comorbid depression. This study raises the provocative question, “Is exposure necessary” ? for successfully treating PTSD.

Treatment delivery has become a focus of research following the COVID-19 prohibition of face-to-face therapies and other known logistics and cost challenges of delivering direct-contact therapies. A 2021 Cochrane review of internet-based cognitive or behavioral therapy (I-C/BT) for PTSD found some beneficial effects of I-C/BT for PTSD. The certainty of the evidence was very low due to the small number of included trials.

Psychotherapy for Military-Related PTSD

A review of 37 controlled trials found that military-related PTSD is complex and difficult to treat. Trauma-focused treatments and CPT are associated with improvement in symptoms in approximately 60% of veterans but are limited by relatively high dropout rates, and treated patients often remain symptomatic, with more than two-thirds retaining their diagnosis. Non–trauma-focused treatments, which are less demanding for patients and professionals, provide a reasonable option over trauma-focused therapy, particularly if this leads to continued willingness to remain in care.

Psychological Therapies for the Treatment of PTSD in Children and Adolescents

A review of 14 controlled studies, including 758 participants exposed to sexual abuse, civil violence, natural disaster, domestic violence, and motor vehicle accidents, found no study was rated at high risk for selection or detection bias but a minority were rated at a high risk for attrition, reporting, and other bias. Across all psychological therapies, improvement was significantly better within a month of completing psychological therapy compared to a control group. The psychological therapy for which there was the best evidence of effectiveness was CBT, with greater improvement lasting up to a year.

Pharmacotherapy for Adults with Chronic PTSD

A meta-analysis was conducted to compare the efficacy of different medications for treating PTSD. The largest body of evidence for short- and long-term efficacy of medication currently exists for SSRIs, and the strongest evidence is for sertraline and paroxetine which are approved by the FDA for adults with PTSD. There are promising findings for the noradrenergic and serotonergic reuptake inhibitor venlafaxine and the atypical antipsychotic risperidone. Evidence for the effectiveness of benzodiazepines is lacking, despite their continued use in clinical practice. Finally, the alpha 1-adrenergic antagonist prazosin and the atypical antipsychotics show some efficacy in treatment-resistant PTSD, and prazosin shows preliminary efficacy for treating nightmares. Insomnia is prevalent and disabling in chronic PTSD, with low-dose trazadone preferable to benzodiazepines. The recently developed orexin antagonist suvorexant is promising for managing trauma-related insomnia. Medications in the pipeline are targeting endocannabinoids, including fatty acid amide hydrolase (FAAH) inhibitors, nabilone, and cannabidiol, and other novel targets including glutamate, brain-derived neurotrophic factor (BDNF), and the oxytocin receptors.

In a recent systematic review of cannabis in the management of PTSD, cannabis was associated with a reduction in overall PTSD symptoms and improved quality of life. Dry mouth, headaches, and psychoactive effects such as agitation and euphoria were the commonly reported adverse effects. Cannabis was generally well-tolerated, but small proportions of patients experienced worsening PTSD symptoms. Overall, evidence to date stems from low-quality and high-risk-of-bias studies, such that the use of cannabis in PTSD cannot be unequivocally recommended.

Pharmacologic Treatment of Comorbid PTSD and Substance Use Disorder

Researchers have reviewed advances in pharmacologic treatment of comorbid PTSD and SUD. A randomized clinical trial of sertraline did not show overall efficacy for comorbid PTSD and alcohol use disorder (AUD), although it may have efficacy among light drinkers. Another clinical trial demonstrated the efficacy of both disulfiram and naltrexone for the treatment of AUD in individuals with PTSD. A recent clinical trial suggested that NE reuptake inhibitors may also have efficacy for the treatment of comorbid PTSD and AUD. Noradrenergic medications that are promising for comorbid PTSD and SUD include prazosin, guanfacine, and atomoxetine. Promising glutamate/GABA medications include topiramate, memantine, acamprosate, N-acetylcysteine (NAC), and ketamine. The safety and efficacy of these medications for the treatment of PTSD and SUDs need to be tested in controlled clinical trials.

Combined Pharmacotherapy and Psychological Treatments for PTSD

Researchers have assessed whether the combination of psychological therapy and pharmacotherapy is more effective for treating PTSD than either of these interventions alone. Patients of any age or gender, with chronic or recent onset PTSD arising from any type of event relevant to the diagnostic criteria were included. Four trials were eligible for inclusion, one of these trials (*n* = 24) was on children and adolescents. All used an SSRI and PE or a cognitive behavioral intervention. There was no strong evidence for superiority of combined treatment over psychological therapy or pharmacotherapy used as monotherapy. The authors conclude that there is not enough evidence to support or refute the effectiveness of combined pharmacotherapy and psychological therapy compared to either of these interventions alone and that large randomized controlled trials are urgently required.

A recent study evaluated the effect of 3, 4-methylenedioxymethamphetamine (midomafetamine or MDMA) as add-on therapy provided in conjunction with CBT and found the combination

(MDMA + CBT) more efficiently reduced PTSD symptoms than CBT alone. MDMA is widely used as recreational drug (alias Ectasy, E, or Molly) for its hedonic, altered sensation and increased empathy, and energy. Chemically MDMA is a substituted amphetamine primarily increasing serotonin, dopamine, and noradrenaline transmission. Importantly, MDMA alone has no known therapeutic efficacy in PTSD and has a significant abuse potential.

Nontraditional, Alternative, and Complementary Treatments

There is preliminary evidence to support the efficacy of acupuncture, breathing and muscle relaxation, mindfulness meditation, and yoga in reducing PTSD symptoms and associated distress. A systematic review assessed the evidence for 15 novel interventions for the treatment of PTSD. The majority of the 19 studies were of poor quality, hampered by methodologic limitations, such as small sample sizes and lack of control group. There were 4 interventions, however, stemming from a mind–body philosophy (acupuncture, emotional freedom technique, mantra-based meditation, and yoga) that had moderate quality evidence from mostly small- to moderate-sized randomized controlled trials. Importantly, judging complementary interventions against PTSD symptom reduction might miss an important effect on well-being, self-monitoring, and self-confidence, which are critically important for many chronic PTSD patients.

Novel and Emerging Treatment Modalities

Recent advances in understanding PTSD neuropathology have identified specific brain areas mechanistically related to PTSD pathophysiology and expression. The latter constitute potential targets for electrical, electromagnetic, or biofeedback neuromodulation.

A meta-analysis of 13 studies (549 participants) of repetitive transcranial magnetic stimulation (rTMS) has shown that rTMS was superior to sham comparisons in reducing PTSD and depression severity. The quality of evidence, however, was limited by small samples sizes, treatment heterogeneity, inconsistent results, and an imprecise pooled effect. Further research is required to advance the evidence of this treatment.

Neurofeedback (NF) studies that train PTSD participants to reduce the amygdala activities are under development with promising preliminary results. Functional MRI and MRI-derived brain-area specific EEG NF allow humans to voluntarily self-modulate neural patterns in specific brain areas, thereby opening the gate for brain-guided psychiatric interventions. This new framework, termed “process-based NF,” relies on targeting specific dysfunctional mental processes by modifying their underlying neural mechanisms via neuro-feedback interfaces. Studies targeting the amygdala as primary NF target in PTSD provided preliminary evidence for efficacy in reducing PTSD symptoms and for a sustained learning effect after treatment termination.

Early and Preventive Intervention

Unlike other mental disorders, PTSD often follows a distinct triggering event and has a clear onset point. Moreover, numerous trauma-exposed individuals are brought to the immediate attention of emergency care services and helpers. This creates a unique opportunity to identify individuals at risk and initiate preventive interventions.

Individuals’ reactions to traumatic events follow characteristic trajectories that include persistent lack of symptoms, initial distress followed by recovery, and nonremitting symptoms. A less frequent symptom trajectory consists of early resilience followed by delayed onset of symptoms. These trajectories define three primary goals for early interventions: (a) mitigate the development of early symptoms, (b) increase the likelihood of remission in survivors with initial symptoms, and (c) prevent the delayed emergence of PTSD. Studies addressing the first goal include “stress-management,” “preparation and training,” or “need-based assistance” (food, shelter, soothing human contact), as well as pharmacologic and psychological prevention provided immediately after trauma exposure. Studies addressing the second and third goals must identify survivors at risk and use specific intervention protocols.

Early Psychological Interventions

DEBRIEFING. Psychological debriefing, a widely used method in the 1980s to 1990s, aimed at preventing long-term post-traumatic symptoms by promoting emotional processing of traumatic events. Debriefing was offered unselectively to anyone exposed to a potentially traumatic event. The method typically involved a single session within several days of the trauma exposure, either in a group or individually, and included general education about the possible effects of trauma exposure, as well as chronologic retellings of the recent traumatic event. The method has face validity and is still well-known, and therefore may be expected by laypeople when confronted with traumatic events. However, well-conducted studies showed no evidence of beneficial effects and even suggested that debriefing may have a negative effect on recovery. After a negative Cochrane review first published in 1997, most treatment guidelines have been updated to recommend against providing psychological debriefing on a routine basis for adults after trauma.

EARLY COGNITIVE BEHAVIORAL THERAPY. Cognitive-behavioral therapy (CBT) can involve different strategies with distinct aims. *Exposure-based CBT,* exemplified by the PE protocol, aims to achieve and maintain fear extinction through repeated exposure to trauma-related stimuli in a safe context, thereby providing a sense of control over reactions and reducing avoidance. *Cognitive-based CBT* approaches aim to change the way patients continue to react to trauma-related reminders and to behavioral restrictions and rules derived from the traumatic experiences, and to negative appraisal of self and others by challenging the patient’s beliefs about the meaning and current implication of the trauma. CBT is offered individually or in a small group to people who report symptoms (else there are no “intervention targets”) and typically involves several weekly sessions, homework, and in vivo training exercises. The treatment may continue for over 3 months and require significant skills from the therapists.

Exposure-based CBT efficiently reduces PTSD symptoms in selected samples. A study was conducted using modified PE in rape, assault, and motor vehicle accident survivors around 12 hours after trauma and found lower PTSD symptoms in the intervention group at 4 and 12 weeks after trauma, mainly for sexual assault victims. Another study found 5 weeks of exposure-based CBT to be effective in reducing PTSD in participants who meet acute stress disorder diagnostic criteria. Research has also found a reduction of PTSD symptoms at 13 months, but not 3 months, after the traumatic events. In the largest study to date, researchers showed that both exposure and nonexposure-based CBT efficiently reduced the prevalence and the intensity of PTSD symptoms 5 months after trauma exposure, and that the outcome of delayed CBT (starting 5 months after the event), assessed 9 months after trauma exposure, did not differ from that of treatment administered within a month of the traumatic event.

However, a small study with 3 weeks of PE did not find significant symptom improvement in the PE group compared to supportive counseling. Cognitive-based CBT has also shown effectiveness in some but not all studies. Researchers compared CBT to waitlist control subjects with acute PTSD, and found that CBT accelerated recovery, but made no long-term difference.

Modifications of CBT-based intervention showed varied results. A telephone-based CBT study in patients with a cardioverter defibrillator installed reported significant improvements in PTSD symptoms in the CBT group. Researchers developed a self-guided internet-based intervention (trauma TIPS) based on CBT to prevent the onset of PTSD symptoms. The result did not support the efficacy of trauma TIPS.

Cognitive behavioral therapy is currently the mainstay of early prevention of PTSD. Several considerations, however, make the systematic implementation of CBT a major challenge. First, a meta-analysis of early interventions has indicated that CBT is only efficient in participants with diagnosable PTSD at treatment onset. Other findings have similarly demonstrated that survivors with subthreshold PTSD symptoms recover as well with or without CBT. Furthermore, research has shown the known efficiency of CBT is limited to survivors of sexual assault, showing no effect among accident victims. Moreover, attempts to deliver diluted versions of CBT (such as telephone or web-based communication) yielded negative results. Interestingly, studies do not show a reduced effect of CBT when administered later in the development of PTSD (e.g., 5 months after trauma exposure instead of 1 month). CBT, therefore, is best positioned by current research as the treatment for ascertained clinical cases, optimally provided at some distance from the traumatic event. On the positive side, the early outcome of CBT is maintained over time, and thus this intervention is virtually the unique means to stop the progression of early PTSD into chronicity. In its best performance, however, early CBT leaves a substantial number of survivors unaffected, and thus should be supplemented by “second-step” interventions or by interventions specifically affecting participants in the nonremission trajectory.

Pharmacologic Interventions for Preventing PTSD. Various pharmacologic agents have been examined in the prevention of post-traumatic symptoms. A recent review article concluded that, in general, there is moderate quality evidence for the efficacy of hydrocortisone, and no evidence for propranolol, escitalopram, temazepam, and gabapentin. This field is rapidly developing as the neurobiologic processes underlying symptoms start to be clarified by more studies.

HYDROCORTISONE. Hydrocortisone has been shown to be effective, especially in patients who have never been treated for psychiatric disorders. The underlying mechanism is not entirely understood. One hypothesis is that hydrocortisone can facilitate extinction learning through both nongenomic and genomic effects. Some also believe that high-dose exogenous hydrocortisone administered shortly after trauma may promote recovery through enhancing synaptic plasticity and connectivity. An animal model showed significantly increased dendritic growth and spine density, with increased levels of BDNF and obtunded postsynaptic density-95 (PSD-95) levels in steroid-treated stressed rats. Increasing cortisol levels also counters adrenergic activation, thereby reducing fear conditioning.

PROPANOLOL. Propranolol is a beta-adrenergic antagonist that crosses the blood–brain barrier and therefore is capable of reducing the CNS adrenergic drive associated with defensive threat responses. Experimental studies of propranolol in healthy subjects have shown that its administration prior to exposure to potentially traumatic narratives reduced the recollection of stressful elements of the narrative without affecting the general recall. It was proposed as a prime candidate to affect traumatic recall in PTSD. Early treatment with propranolol aims to preventing overconsolidation of traumatic memories by blocking the memory-enhancing influence of stress hormones. Theoretically, therefore, it should be initiated while memories of the trauma are being encoded and consolidated, preferably within hours of the traumatic event. However, controlled studies in the first hours after exposure have failed to show a preventive effect of propranolol on PTSD symptoms, although they showed a reduction in physiologic responses to reminders. The documented gap between propranolol’s effect on physiologic responses to trauma reminders and its lack of effect of PTSD symptoms could be interpreted as suggesting that the pathogenesis of PTSD is not limited to amygdala-mediated threat conditioning and involves other brain areas and memory systems.

BENZODIAZEPHINES. Benzodiazepines are a group of GABA agonists and thereby enhance the inhibitory transmission in wide areas of the brain. They are used as tranquilizers and sleep inducers but also interfere with long-term potentiation, learning, and memory. As such they were positioned as capable of reducing excessive "learning" occurring during or following trauma exposure. However, three human studies have shown that participants receiving benzodiazepines fared worse than those not treated at all. An animal "predator stress" study found that administering diazepam shortly after predator odor exposure enhanced the acquisition of long-term fear responses. While the exact mechanism of benzodiazepines' PTSD-enhancing effect is unknown, it is possible that these compounds interfere with postevent learning of extinction. Intriguingly, current data on benzodiazepines rely on small studies, while these compounds are widely used to mitigate distress following various stressors. Nonetheless, benzodiazepines are not recommended for use in the aftermath of traumatic events.

MORPHINE. Animal studies suggest morphine can produce retrograde amnesia for contextual conditioned fear, possibly through decreasing cAMP or activating NMDA receptors in the hippocampus. Observational studies of hospital patients also suggested a possible beneficial effect of morphine administration within 48 hours after trauma exposure to survivors who had pain, and a similar result was reported, retrospectively, in 696 military personnel with severe combat injury. Given the retrospective nature of most studies, more research is needed to separate a specific effect of morphine from a generic "analgesic" effect. Pain after trauma exposure is a potent predictor of PTSD. It is unclear, therefore, whether morphine has any preventive value in trauma survivors without physical pain.

Other and Investigational Approaches. Oxytocin is involved in emotion stress regulation, social engagement, and attachment, and preliminary studies suggest that it may buffer the development of PTSD when administered shortly after trauma exposure. NPY is another neuroendocrine candidate for early intervention. An animal study showed that delivery of NPY to rats' brains has a pronounced effect on reducing the development of PTSD-like symptoms, possibly through modifying stress-triggered dysregulation of HPA axis and central noradrenergic activity. Besides hormonal intervention, neurobehavioral retraining is also being tested for its capacity to reduce negative emotional processing and enhance executive control. Emerging evidence of impaired emotion regulation and executive functions in PTSD is likely to generate other early neurocognitive approaches.

Placing Prevention in Context. Most studies above summarize the “clinical” implementation of preventive interventions, that is, specialized care provided to survivors screened or formally diagnosed within a healthcare setting. They are therefore best construed as secondary and tertiary prevention. Studies to date, however, have not followed a primary prevention model. Challenging the restricted “clinical” model, Zatzick et al. evaluated a stepped collaborative care model, which introduced care managers to address each patient’s unique needs and coordinated different components of interventions (e.g., CBT, motivational interviewing, pharmacotherapy). The care team repeatedly measured patients’ symptoms and adjusted levels of care accordingly. Results showed the feasibility and effectiveness of the method, with reduced PTSD symptoms in the intervention group.

IMPLICATIONS AND FUTURE DIRECTIONS

Studies on PTSD prevention demonstrated significant variability in study methodology. Differences in methodology need to be carefully examined before one can infer from study results and determine the most effective intervention for a certain population. Current evidence from research also represents a great degree of heterogeneity in sampling. Different studies recruited people with different features (e.g., trauma types, injury severity, and peritrauma reactions). Among all traditional prevention methods, TFCBT has been mostly reported as an effective approach, but only in people with acute stress symptoms. Taking into consideration the heterogeneity of trauma survivors, people with different ages, genders, trauma types, and potentially different genetic features, childhood experiences, and recovery environments may need different optimal strategies. For example, brief TFCBT was reported to be more effective in victims of traffic accidents, and exposure therapy was shown to be more beneficial to sexual assault victims or people with high genetic risk. No intervention is effective for all trauma survivors.

Most importantly, PTSD is likely multicausal, and, as such, individuals with differing vulnerabilities and different exposure and postexposure circumstances may come to express the PTSD symptom complex through individual-specific pathways—and be responsive to individual-specific interventions. One way to advance the prevention of PTSD is to better map the variety of paths leading to this condition and map those paths into subsets of trauma-exposed individuals. Subtyping PTSD for treatment stratification is a promising strategy for advancing precision medicine. Once such knowledge becomes available, target-specific early intervention might replace generic treatment protocols applied to aggregates. Several steps should be taken in future studies. Psychiatrists need to step beyond diagnosis-based screenings and develop more complex and accurate methods to predict individual-level risks of chronic symptoms and impairment after traumatic events. By enhancing prediction models, intervention studies can take the important step to selecting the most relevant sample for a more rigorous study design and the best clinical interest. At the same time, research needs to continue exploring and confirming the underlying mechanisms of post-traumatic pathogenesis. Increasing knowledge to understand the mechanisms of post-traumatic psychopathology will open opportunities to explore more targeted interventions outside the range of traditional treatment-derived methods. These targeted intervention methods can allow clinicians to focus on both specific population and specific pathologic processes.

FURTHER READINGS

Amos T, Stein DJ, Ipser JC. Pharmacological interventions for preventing post-traumatic stress disorder (PTSD). *Cochrane Database Syst Rev*. 2014;(7):CD006239.

Benjet C, Bromet E, Karam EG, et al. The epidemiology of traumatic event exposure worldwide: results from the World Mental Health Survey Consortium. *Psychol Med*. 2016;46(2):327–343.

Bisson JI, Roberts NP, Andrew M, Cooper R, Lewis C. Psychological therapies for chronic post-traumatic stress disorder (PTSD) in adults. *Cochrane Database Sys Rev*. 2013;12:CD003388.

Brewin CR, Andrews B, Valentine JD. Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *J Consult Clin Psychol*. 2000;68(5):748–766.

Dohrenwend BP, Turner JB, Turse NA, Adams BG, Koenen KC, Marshal R. The psychological risks of Vietnam for U.S. veterans: a revisit with new data and methods. *Science*. 2006;313(5789):979–982.

Etkin A, Maron-Katz A, Wu W, et al. Using fMRI connectivity to define a treatment-resistant form of post-traumatic stress disorder. *Sci Transl Med*.

2019;11(486):eaal3236.

Forneris CA, Gartlehner G, Brownley KA, et al. Interventions to prevent post-traumatic stress disorder: a systematic review. *Am J Prev Med*. 2013;44(6):635–650.

Galatzer-Levy IR, Ankri Y, Freedman S, et al. Early PTSD symptom trajectories: persistence, recovery, and response to treatment: results from the Jerusalem Trauma Outreach and Prevention Study (J-TOPS). *PLoS One*. 2013;8(8):e70084.

Gilbertson MW, Shenton ME, Ciszewski A, et al. Smaller hippocampal volume predicts pathologic vulnerability to psychological trauma. *Nat Neurosci*. 2002;5:1242–1247.

Greenberg N, Weston D, Hall C, Caulfield T, Williamson V, Fong K. Mental health of staff working in intensive care during Covid-19. *Occup Med (Lond)*. 2021;71(2):62–67.

Hetrick SE, Purcell R, Garner B, Parslow R. Combined pharmacotherapy and psychological therapies for posttraumatic stress disorder (PTSD). *Cochrane Database Syst Rev*. 2010;7:CD007316.

Kearns MC, Ressler KJ, Zatzick D, Rothbaum BO. Early interventions for PTSD: a review. *Depress Anxiety*. 2012;29(10):833–842.

Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):617–627.

Lanius RA, Brand B, Vermetten E, Frewen PA, Spiegel D. The dissociative subtype of posttraumatic stress disorder: rationale, clinical and neurobiological evidence, and implications. *Depress Anxiety*. 2012;29(8):701–708.

Lewis SJ, Arseneault L, Caspi A, et al. The epidemiology of trauma and post-traumatic stress disorder in a representative cohort of young people in England and Wales. *Lancet Psychiatry*. 2019;6(3):247–256.

Li Y, Scherer N, Felix L, Kuper H. Prevalence of depression, anxiety and post-traumatic stress disorder in health care workers during the COVID-19 pandemic: a systematic review and meta-analysis. *PLoS One*. 2021;16(3):e0246454.

Lubianiker N, Goldway N, Fruchtman-Steinbok T, et al. Process-based framework for precise neuromodulation. *Nat Hum Behav*. 2019;3(5):436-445.

Marmar CR, Schlenger W, Henn-Haase C, et al. Course of posttraumatic stress disorder 40 years after the Vietnam war: findings from the National Vietnam Veterans Longitudinal Study. *JAMA Psychiatry*. 2015;72(9):875–881.

Nievergelt CM, Ashley-Koch AE, Dalvie S, et al. Genomic approaches to posttraumatic stress disorder: the Psychiatric Genomic Consortium Initiative. *Biol Psychiatry*. 2018;83(10):831–839.

Ozer EJ, Best SR, Lipsey TL, Weiss DS. Predictors of posttraumatic stress disorder and symptoms in adults: a meta-analysis. *Psychol Bull*. 2003;129(1):52–73.

Pitman RK, Rasmusson AM, Koenen KC, et al. Biological studies of post-traumatic stress disorder. *Nature Rev Neurosci*. 2012;13(11):769–787.

Rehman Y, Saini A, Huang S, Sood E, Gill R, Yanikomeroglu S. Cannabis in the management of PTSD: a systematic review. *AIMS Neurosci*. 2021;8(3):414-434.

Roberts NP, Kitchiner NJ, Kenardy J, Bisson JI. Early psychological interventions to treat acute traumatic stress symptoms. *Cochrane Database Syst Rev*. 2010;(3):CD007944.

Shalev AY, Gevonden M, Ratanatharathorn A, et al. Estimating the risk of PTSD in recent trauma survivors: results of the International Consortium to Predict PTSD (ICPP). *World Psychiatry*. 2019;18(1):77–87.

Simon N, Robertson L, Lewis C, et al. Internet-based cognitive and behavioural therapies for post-traumatic stress disorder (PTSD) in adults. *Cochrane Database Syst Rev*. 2021;5:CD011710.

Steenkamp MM, Litz BT, Hoge CW, Marmar CR. Psychotherapy for military-related PTSD: a review of randomized clinical trials. *JAMA*. 2015;314(5):489–500.

Stein MB, Levey DF, Cheng Z, et al. Genome-wide association analyses of post-traumatic stress disorder and its symptom subdomains in the Million Veteran Program. *Nat Genet*. 2021;53(2):174–184.

Vinkers CH, Geuze E, van Rooij SJH, et al. Successful treatment of post-traumatic stress disorder reverses DNA methylation marks. *Mol Psychiatry*. 2021;26(4):1264–1271.

Zatzick D, Roy-Byrne P, Russo J, et al. A randomized effectiveness trial of stepped collaborative care for acutely injured trauma survivors. *Arch Gen Psychiatry*. 2004;61(5):498–506.

Zhu K, Niu Z, Freudenheim JL, et al. COVID-19 related symptoms of anxiety, depression, and PTSD among US adults. *Psychiatry Res*. 2021;301:113959.

18

Somatic Symptom and Related Disorders

Javier I. Escobar, M.D., M.Sc., and Joel E. Dimsdale, M.D.

It is common to experience and communicate psychological distress in the form of physical symptoms and to seek medical attention for these symptoms. Often, these physical symptoms remain poorly explained and are associated with increased medical visits, unnecessary medical tests, and the performance of procedures that may result in iatrogenic complications. For many decades, the term “somatization” was used to describe these phenomena. This was gradually replaced by more descriptive terms such as “medically unexplained symptoms,” “unexplained symptoms,” and “functional somatic symptoms.” The word “idiopathic” (meaning “arising spontaneously or from an obscure or unknown cause”) is probably the most technically correct way to label the physical symptoms displayed by many patients with psychiatric disorders that remain poorly explained after detailed medical scrutiny and laboratory/radiologic assessments. The evidence for the “psychogenesis” of many of these somatic presentations was originally derived from theoretical postulates such as Wilhelm Stekel’s notion of somatization and has been nurtured by the frequent association of these symptoms with psychological distress and psychopathology. However, many of these somatic manifestations can arise spontaneously, and their causes can remain obscure or unknown. The most distinctive characteristic of many patients with somatic symptom disorders (SDDs) is not the symptoms they complain about but the way in which they present and interpret them. Despite the relevance that these idiopathic physical symptoms have to medicine and psychiatry, their definition is difficult, and their classification remains controversial. This has led to frequent revisions of nomenclatures that alter the criteria used to designate “cases” and complicate the clinical recognition and management of these syndromes, recently and notably in *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5).

Everyone experiences somatic symptoms with a certain frequency, and most are able to cope with them effectively. However, some people’s lives are overwhelmed by their somatic concerns. Sometimes, the somatic concerns stem from well-established major medical illnesses; sometimes, the origins of the concerns are never quite clear. What is common in both situations is that patients’ lives are shadowed by disproportionate and excessive thoughts, feelings, and behaviors that center on the perceived somatic sensations. DSM-5 incorporated this perspective because the older perspective (simply counting or cataloging symptoms labeled as medically unexplained) embodied in the previous somatoform disorders’ category in DSM-III and IV, was unreliable and such perspective often put the doctor and the patient at odds over the question of the legitimacy and “reality” of the patient’s symptoms and personal suffering. In fact, research has shown that regardless of medical explanations, multiple physical symptoms predict disability and psychopathology (Fig. 18–1).

DEFINITION

According to the fifth edition of the American Psychiatric Association’s DSM-5, the key characteristic of SSDs is the presence of one or more persistent somatic symptoms that are associated with excessive thoughts, feelings, and behaviors related to the symptoms. The stipulation that these symptoms be medically unexplained has been removed because of the recognition that such a distinction in itself is unreliable and because psychiatrists commonly treat patients with medically established diagnoses who are troubled by or preoccupied with their physical symptoms disproportionately. Furthermore, DSM-5 acknowledges that frequently, SSDs co-occur with other psychiatric disorders. The diagnosis of a somatic symptom disorder is only made when there is significant distress or impairment associated with these chronic somatic symptoms.

In DSM-IV and earlier editions of the DSM, disorders such as factitious disorder and psychological factors affecting other medical conditions (PFAOMC) were grouped elsewhere. This led to some confusion in the differential diagnosis. As a result, such disorders are now included within the rubric of the SSDs chapter. Due to its strong linkage and similarity to obsessive-compulsive disorders, body dysmorphic disorder has now been moved out of the SSDs and is now included in the obsessive-compulsive disorders group. Because pain symptoms are quite common and the requirement that symptoms be unexplained has been removed, pain disorder as a psychiatric diagnosis has been deleted in DSM-5, and these syndromes are now subsumed under somatic symptom disorder. Also, largely because of its pejorative connotation, the classical term "hypochondriasis" has been removed, and the disorder is renamed illness anxiety disorder in DSM-5 (Fig. 18–2).

Thus, the SSD group in DSM-5 includes the following disorders: "somatic symptom disorder," "illness anxiety disorder," "functional neurologic symptom disorder (conversion disorder)," "psychological factors affecting another medical condition," and "factitious disorder." Somatic syndromes that are below threshold are classified under two residual categories, "other specified" and "unspecified somatic symptom and related disorders."

In the DSM-5 field trials, somatic symptom disorder showed high levels of reliability, more robust even than those for more traditional diagnoses (Fig. 18–3).

HISTORY

Somatic presentations that have a presumed psychological origin have always mystified the medical establishment and have taken new forms and shapes as cultures evolve, and medical paradigms shift. Throughout the ages, multiple folk concepts and terms have been used to designate the phenomena, such as "hysteria," "hypochondria," "spleen," "English malady," "surmenage," "humoral disorders," "neurovegetative dysfunction," "neurasthenia," "neurovegetative dystonias," "psychosomatic disorders," and many others. The idea that psychological processes are involved in these somatic experiences appears to have ancient origins but waxed and waned periodically. In the 17th century, Thomas Sydenham attributed symptoms and signs such as "pain," "convulsions," "diarrhea," and "dropsy" to disturbances of the mind.

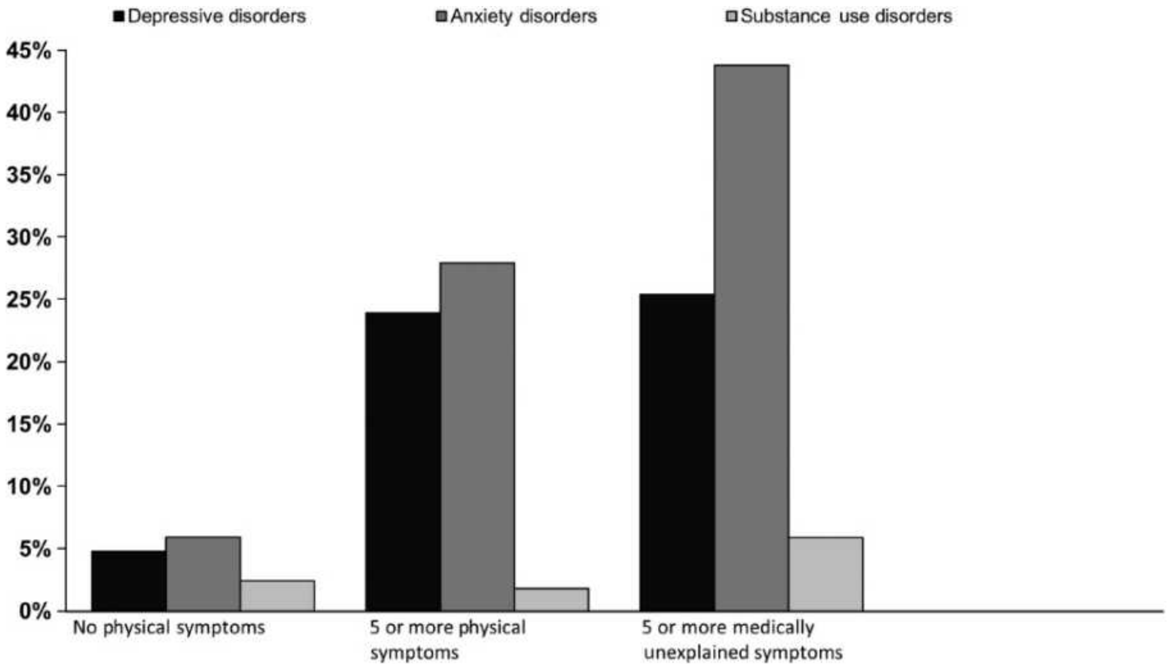

FIGURE 18–1. Whether medically explained or not, high levels of somatic symptoms augur psychopathology in general populations. (Data from Escobar JI, Cook B, Chen CN, et al. Whether medically explained or not, three or more concurrent somatic symptoms predict psychopathology and service use in community populations. *J Psychosom Res.* 2010;69(1):1–8.)

Hysteria and Hypochondria

The more dramatic somatic manifestations—“hysteria” and “hypochondria”—were key syndromes in traditional psychopathology and remained influential in medical discourse at the turn of the 20th century. The long survival of these labels (e.g., “hysteria,” “hypochondria” in the United States, and “neurasthenia” in China) despite efforts to eliminate them attests to their historical and practical value, but also to the fragility and inadequacy of current classifications. Michel Foucault included “hysteria” and “hypochondria” among the four classic syndromes in psychopathology together with “mania” and “melancholia.” He indicated that from ancient times, somatic presentations have not fitted well with the schemes of traditional psychopathology and have been viewed as outcasts in medicine and psychopathology.

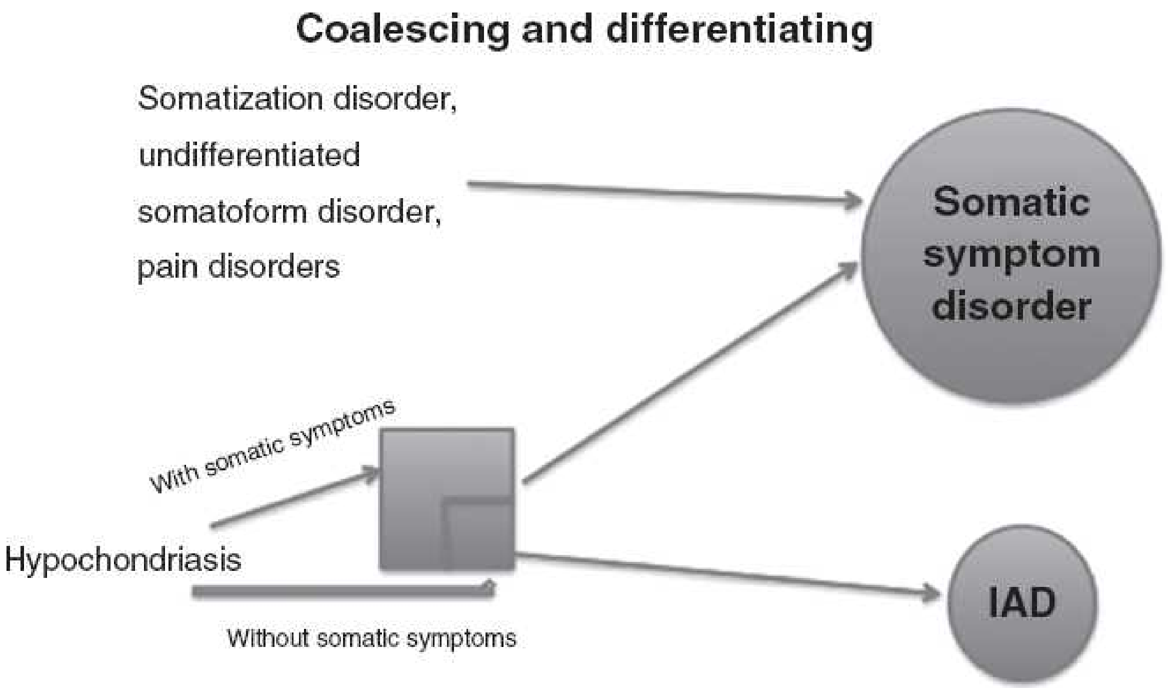
Coalescing and differentiating
Somatization disorder, undifferentiated somatoform disorder, pain disorders
Somatic symptom disorder
With somatic symptoms
Hypochondriasis
Without somatic symptoms
IAD

FIGURE 18–2. Main changes from DSM-IV to DSM-5. (Reprinted from Dimsdale JE, Creed F, Escobar JI, et al. Somatic symptom disorder: an important change in DSM. *J Psychosom Res.* 75(3):223–228. Copyright © 2013, with permission from Elsevier. doi: 10.1016/j.jpsychores.2013.06.033.)

Somatic Presentations in Classical Psychopathology

In his “History of Mental Symptoms,” the psychiatric historian German Berrios describes how somatic syndromes (hysteria and hypochondria) were discarded in 19th-century nosologies and that terms such as “dysmorphophobia” and *cenestopathie* were the precursors of the psychosomatic syndromes.

Kurt Schneider viewed somatic presentations not as a separate group but as key components of other psychopathological syndromes. For example, he considered “anguish” as a corporal sensation localized in the chest. However, in Schneider’s psychopathology, somatoform phenomena appear to have the best fit among his “psychopathic personalities,” specifically the group he labeled “asthenic psychopaths.” Schneider characterized them as “somatically labile” or “somatopaths”—individuals who completely focused their attention on their bodies, fatigued easily, and suffered from insomnia, headaches, and heart, bladder, and mental disturbances. Schneider also trivialized somatic presentations by asserting, for example, “occupation often dissipates all these problems.”

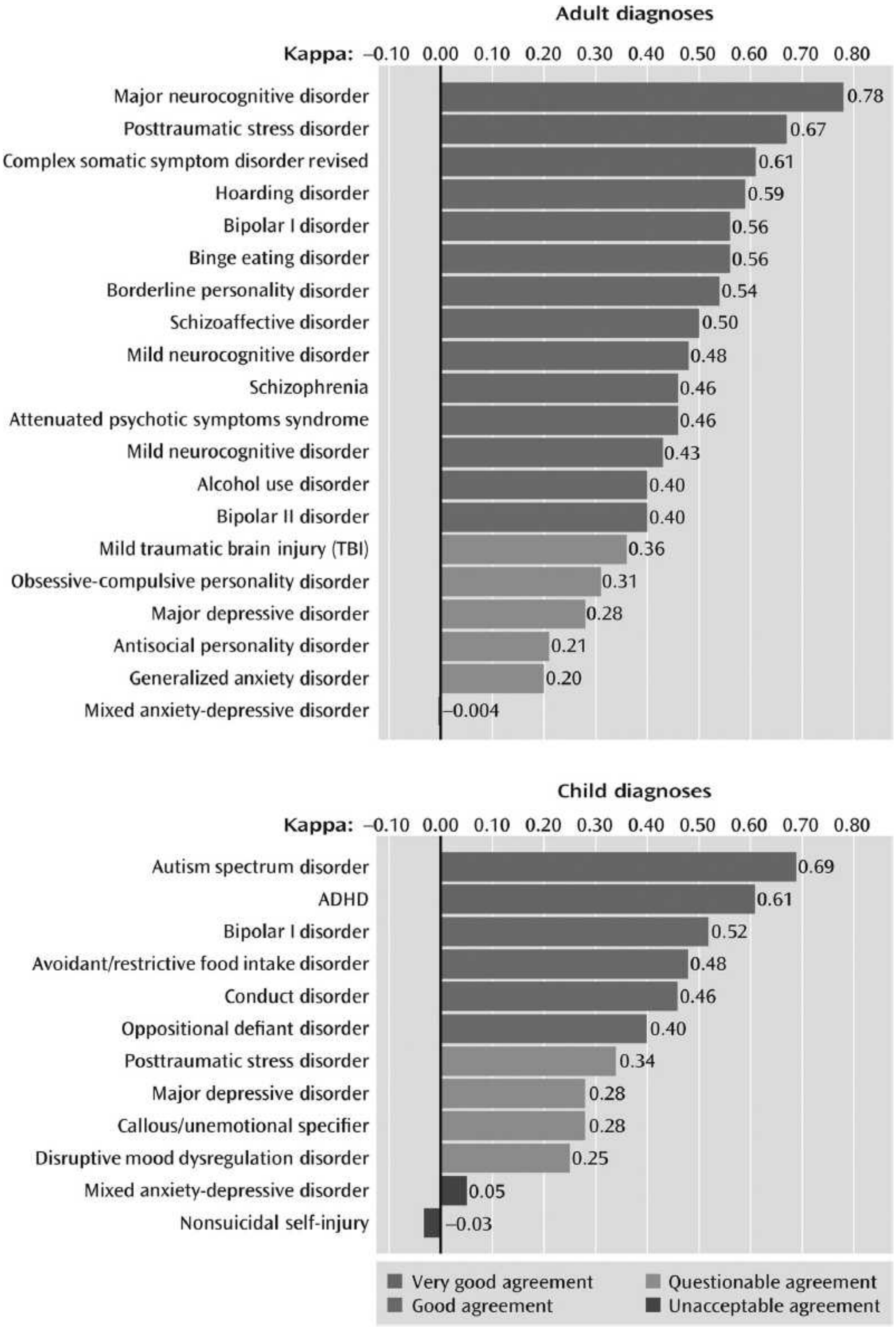

FIGURE 18–3. Reliability of somatic symptom disorders relative to other DSM-5 disorders. (Reprinted with permission from Freedman R, Lewis DA,

Michels R, et al. The initial field trials of DSM-5: new blooms and old thorns. *Am J Psychiatry*. 2013;170(1):2.)

Karl Jaspers mentioned "hypochondria," "hysteria," and the "somatopsychic," emphasizing the polarity that existed at that time regarding physical and psychological approaches to these problems. He affirmed that a healthy person lives in his body and does not think much of it and added that the great majority of physical suffering is due to psychological reflection and not to manifested physical disease. Thus, in Jaspers' opinion, many of the somatic presentations tend to become fixed and repetitive and can be understood as part of personal experience, situations, and conflicts.

The Swiss psychopathologist Eugen Bleuler dedicated very little attention to somatic phenomena, considering them to be outside the "endogenous" realm to which he devoted the most attention in his psychopathology. While Bleuler alluded to "hysterical syndromes" and "neurasthenic syndrome," he did not mention hypochondria or psychosomatic phenomena.

Henry Ey, an influential French psychiatrist and author of a well-known textbook on psychiatry in France, reflected the influence of psychoanalytic theories on European psychiatry during the first half of the 20th century. In his textbook, he devoted a full chapter to hysteria, which he defined as "the somatic hyper-expressivity of unconscious ideas, images, and affects." Ey did not mention hypochondriasis but included a long list of psychosomatic disorders that he labeled "organ neuroses," similar to the approaches of Otto Fenichel and other psychoanalysts who are well known in North America.

The Spanish psychiatrist Juan José Lopez-Ibor, strongly influenced by German psychiatry, was perhaps the first to highlight the somatic ("endogenous" or "biologic") component of some forms of anxiety and depression, proposing antidepressant treatments for them as early as the 1960s and thus anticipating by a number of years the development of the panic disorder concept in North American psychiatry. As to phenomena relevant to somatic presentations, Lopez Ibor defined anxiety as "vital anxiety or anguish," in his view, a corporeal sensation, coining the Spanish term *timopatia ansiosa* (anxious thymopathy), not referring to the thymus gland but to the Greek root *thymos,* meaning "soul, spirit, mind or temperament." Lopez Ibor also placed "hysteria," "hypochondria," and "psychosomatic disorders" among his "disorders of mood."

Briquet Syndrome

The description in 1859 of a syndrome with multiple motor and sensory symptoms by the French neurologist Pierre Briquet allowed an initial separation of somatization from conversion phenomena. This approach initiated a trend to emphasize the descriptive aspects of these syndromes, gradually detaching them from psychological causality and placing a particular focus on the process of eliciting individual physical symptoms. This would eventually lead to the “symptom counting process” that would characterize somatization disorder under the Feighner, research diagnostic criteria (RDC), and DSM (DSM-III and IV) criteria.

Somatization

The psychoanalyst Wilhelm Stekel used this term for the first time in 1911, defining it as the bodily expression of a deep somatic neurosis or the process of how neurotic conflicts are presented as physical symptoms. Although the term “somatization” is credited to Stekel, its formal entry into medical discourse is due to the work and influence of Zbigniew J. Lipowski, who provided a broader view of somatization, defining it as the tendency to experience, conceptualize, and/or communicate psychological states or contents as body sensations, functional changes, or somatic metaphors. By the time “somatization disorder” was enthroned as a full category in official nosology, an extensive body of research, principally conducted by investigators at Washington University/Barnes Hospital in St. Louis, had produced empirical evidence for the utility of its forbearer “Hysteria/Briquet syndrome” in tertiary care settings in North America.

The St. Louis Criteria

In 1951, James J. Purtell, Eli Robins, and Mandel Cohen, investigators from Washington University in St. Louis, compared the incidence of specific physical and mental symptoms among women with and without a diagnosis of hysteria and highlighted the polysymptomatic and chronic nature of this disorder. Next, Michael J. Perley and Samuel B. Guze, in 1962, formulated for the first time a numerical definition of the symptoms needed to diagnose hysteria, which included the presence of at least 15 individual symptoms distributed over at least 9 of 10 groups of possible symptoms. These criteria were a modification of those proposed by E.

Robins and P. O’Neal in a 1953 study of children that required at least 18 symptoms from at least 7 symptom groups. The St. Louis group began adding other physical symptoms, psychiatric symptoms, and personality characteristics on top of the “pure” neurologic symptoms that were included in the original Briquet syndrome. This grouping together of somatic and psychiatric symptoms was again used, with a few modifications, in the criteria proposed by John P. Feighner and collaborators for use in psychiatric research. In the Feighner criteria, the threshold for a hysteria diagnosis was elevated to at least 25 symptoms distributed in 9 of 10 possible groups. These criteria also required that the disorder should start prior to the age of 30 years.

COMPARATIVE NOSOLOGY

The Concept of “Somatoform”

Psychiatric nomenclatures (DSM-III and IV) relegated somatic phenomena to a secondary level and classified somatic syndromes of presumed psychological origin under the heading of “somatoform disorders,” a category that is hierarchically inferior to other categories included in these nosologies (e.g., mood disorders). In the 1970s, the American nosologist Robert Spitzer coined the term “somatoform” (meaning somatic-like) to group these disorders among his RDC, which made its formal entry in North American psychiatry with the publication of the DSM-III in the 1970s. Although the creation of the somatoform disorders category was lauded as a nosologic advance by some investigators, in the real world, patients presenting with multiple unexplained physical symptoms did not fit neatly into the taxonomies provided by the DSM and international classification of diseases (ICD) systems. Even when somatoform criteria are met, there is also significant overlap with several other psychiatric syndromes, a reminder that nosologies are still quite imperfect and in need of continuous distillation. The restrictive quality of categorical somatoform diagnoses left out many of the “real cases,” and this led investigators to propose dimensional classifications. These dimensions included broader, less restrictive categories as an alternative for research, such as “abridged somatization,” “multisomatoform disorder,” and others. The use of dimensional constructs rather than categories may have particular utility in primary care. In reviewing the literature on physical symptoms, it appears that there are two main dimensions that may have practical value. One is characterized by high levels of physical symptoms (the former somatization) and the other by amplification of somatic presentations (hypochondriasis or health anxiety). This dimensional approach has been taken into account in the current nosologic revision (DSM-5) and is included in Section III of the manual (Table 18–1).

Differences between DSM and ICD Classification of Somatic Symptom Disorders

There were several differences between DSM-IV and ICD-10 in the way they classified these disorders. DSM-IV listed all these disorders under the somatoform disorders category. Chapter F45 of the ICD-10 included “somatization disorder” (this had a different symptom threshold from that in DSM-IV), “undifferentiated somatoform disorder,” “hypochondriacal disorder,” other somatomorph disorders, and somatomorph disorder not otherwise specified (note the use of “somatomorph” instead of “somatoform”) and two disorders not listed in DSM-IV, “somatomorph autonomic dysfunction” and “persistent somatomorph pain disorder.” In ICD-10, conversion and dissociative disorders were placed in a separate category under the heading of “dissociative disorders (conversion)” while two disorders not in DSM-IV, “neurasthenia” and “Briquet disorder,” were listed in Chapter F48 under “other neurotic disorders.”

Table 18–1.
DSM-5 Diagnostic Criteria for Somatic Symptom Disorder

300.82 (F45.1) Somatic Symptom Disorder

A. One or more somatic symptoms that are distressing or result in significant disruption of daily life

B. Excessive thoughts, feelings, or behaviors related to the somatic symptoms or associated health concerns as manifested by at least one of the following:

   1. Disproportionate and persistent thoughts about the seriousness of one’s symptoms
   2. Persistently high level of anxiety about health or symptoms
   3. Excessive time and energy devoted to these symptoms or health concerns

C. Although anyone somatic symptom may not be continuously present, the state of being symptomatic is persistent (typically more than 6 months)

Specify if:

With predominant pain (previously pain disorder): This specifier is for individuals whose somatic symptoms predominantly involve pain.

Specify if:

Persistent: A persistent course is characterized by severe symptoms, marked impairment, and long duration (more than 6 months).

Specify current severity:

Mild: Only one of the symptoms specified in Criterion B is fulfilled.

Moderate: Two or more of the symptoms specified in Criterion B are fulfilled.

Severe: Two or more of the symptoms specified in Criterion B are fulfilled, plus there are multiple somatic complaints (or one very severe somatic symptom).

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

ICD-11. World health organization (WHO) released the ICD-11, with an implementation date of 2022.

ICD-11 classifies some of the disorders presenting primarily with somatic symptoms under the new category “disorders of bodily distress and bodily experience.” This label encompasses two disorders: “bodily distress disorder” and “body integrity dysphoria.” The replacement of ICD-10’s “somatoform disorder” with “bodily distress disorder” (BDD) was a major change in ICD-11.

ICD-11’s “bodily distress disorder” is characterized by the presence of bodily symptoms that are distressing to the individual and by an “excessive” level of attention directed toward the symptoms, which leads to repeated contact with health care providers. The disorder is conceptualized as existing on a continuum of severity and can be qualified accordingly (mild, moderate, or severe) depending on the impact on the level of functioning. It is important to emphasize that as in the case with DSM-5’s SSDs, BDD was defined primarily by the presence of essential psychological features, such as psychological distress and “excessive” thoughts and behaviors, rather than by the absence of medical explanations for the symptoms, as had been the case for somatoform disorders in ICD-10 and DSM-IV.

“Body integrity dysphoria” is another newly introduced diagnosis in ICD. It is a rare and severe disorder (there are only about 200 cases of BID recorded in the medical literature) in which patients wish to cut off their healthy limb (e.g., one’s arm or leg) because they feel there is a mismatch between their mental and physical image. BDD has been included in the 11th edition of the (ICD-11) as a “disorder of bodily distress or bodily experience” together with “bodily stress disorder.”

The 11th Edition of the International Classification of Diseases (ICD-11) allocates somatic symptom syndromes into the following categories:

- *Disorders of Bodily Distress and Bodily Experience* (“bodily distress disorder”; “body integrity dysphoria”).
- *Obsessive-Compulsive and Related Disorders* (“hypochondriasis”; “body dysmorphic disorder”).
- *Factitious Disorders* (“factitious disorder”)
- *Psychological and Behavioral Factors Affecting Disorders or Diseases Classified Elsewhere.*
- *Dissociative disorders* (“conversion disorder”; “dissociative neurologic symptom disorder”).

In contrast, all the above-listed ICD-11 diagnoses (with the exception of BID and body dysmorphic disorder) are subsumed under the single category of “somatic symptom disorders” in DSM-5.

Functional Somatic Syndromes

Table 18–2 lists a conglomerate of syndromes frequently diagnosed in medical practice. These consist of clusters of “idiopathic” physical symptoms. Despite the fact that these functional diagnoses are “legitimately” used in many specialties, their association with psychopathology and psychological distress suggests that many of them would fall within the realm of SSD. Excellent reviews of these syndromes have appeared in leading journals on both sides of the Atlantic (e.g., *Lancet, JAMA*), highlighting their common ancestry, as well as their epidemiologic, clinical, and psychopathological aspects. Besides the high levels of idiopathic symptoms, these syndromes also share common elements such as the following:

- The absence of a gold standard against which a specific diagnosis can be confirmed or ruled out.
- No consistent explanation emanates from physical and laboratory assessments.
- The absence of clearly articulated pathophysiology, despite a panoply of putative mechanisms (e.g., symptom amplification, muscle contraction, catecholamine release, persistent neurobiologic dysfunction, neurologic hyper-reactivity, elevated cortisol, etc.).
- The simultaneous presence of multiple unexplained physical symptoms originating from several different organ systems.
- High levels of psychiatric symptoms and “comorbidities.”
- Comparable responses to certain psychological (e.g., cognitive-behavioral therapy [CBT]) and pharmacologic (e.g., antidepressants) interventions.

ICD-11. The EURONET-SOMA workgroup (European Network to improve diagnosis, treatment, and health care for patients with persistent somatic symptoms), a panel of 29 researchers from Denmark, the Netherlands, Sweden, Norway, Latvia, Belgium, United Kingdom, Germany, and Russia was charged with developing a research agenda and work toward a common understanding of the terminology, conceptualization, and management of persistent somatic symptoms and for interdisciplinary agreement on a consistent diagnostic classification. This group recently proposed the new label, “Functional Somatic Disorders,” which they view as a more neutral construct, since in their view, “it is neither purely somatic nor purely mental.” This term would encompass a majority of the functional somatic syndromes listed in Table 18–2.

Table 18–2
Functional Somatic Syndromes across Medical Specialties

Specialty Area	Functional Syndrome(s)
Allergy	Food allergies
Cardiology	Atypical chest pain, noncardiac pain, mitral valve prolapse
Dentistry	Temporomandibular joint syndrome, atypical facial pain
Ear, nose, and throat	Tinnitus, dizziness, globus syndrome
Gastroenterology	Irritable bowel, nonulcer dyspepsia
Internal medicine	Chronic fatigue syndrome, chronic Lyme disease, hypoglycemia, chronic candidiasis
Military medicine	Gulf war syndrome, chronic multisymptom illness
Neurology	Tension headache, pseudoseizures
Obstetrics/Gynecology	Premenstrual syndrome, chronic pelvic pain
Occupational medicine	Multiple chemical sensitivity, sick-building syndrome, repetitive strain injury
Orthopedics	Carpal tunnel syndrome, chronic low back pain, chronic whiplash syndrome
Plastic surgery	Silicone-associated connective tissue disease
Pulmonary medicine	Dyspnea, habit cough, laryngeal dysfunction, hyperventilation
Rehabilitation medicine	Fibrositis, fibromyalgia
Urology	Interstitial cystitis

Data from Henningsen P, Zipfel S, Herzog W. Management of functional somatic syndromes. *Lancet*. 2007;369:946. doi:10.1016/S0140-6736(07)60159-7.

EPIDEMIOLOGY

When available, specific epidemiologic data will be provided for each individual disorder.

In general, somatic presentations are very common in all countries and cultures. A WHO study done in nine countries showed that only one-third of patients presenting to primary care had a documentable physical illness. At least one additional third had a mental disorder. Physical symptoms can be considered the most common presentation of mental disorders worldwide, and their frequency seems to be on the increase. Canadian historian Edward Shorter pointed out that the shift in somatic symptoms commonly presented in medical and psychiatric settings seems to be related to prevailing medical models. He noted that according to U.S. household surveys, there was a 150% increase in “episodes of illness” reported by the U.S. population between the 1930s and the 1980s, despite the enormous improvements in health care. According to the data, more than one-half of all patients presenting to primary care may present with “idiopathic physical symptoms.” Yet, despite their predominantly somatic presentations, such patients do not appear to have a diagnosable physical illness. In fact, the proportion of patients in general medical settings with “idiopathic” physical symptoms for which no organic cause can be found has been estimated to be in the 20% to 80% range. It is quite common for these “idiopathic” physical symptoms to co-occur with psychiatric disorders—specifically with anxiety and mood disorders. With regard to recent data from countries outside the United States, more than 20% of the general population in Germany reported at least one “idiopathic” physical symptom leading to severe impairment. In Switzerland, about 80% of young adults reported somatic problems such as sleep disorders, backache, headache, and stomach or bowel complaints. In addition, research in primary care has shown that regardless of the presence or absence of medical explanations, as the number of individual somatic symptoms increases, so do the odds that the person will also meet the criteria for common psychiatric disorders, in particular depression and anxiety disorders.

Gender and Somatic Presentations

Epidemiologic studies in the United States show a higher prevalence of somatic symptom-related disorders (SSRDs) among females compared to males.

Somatic Presentations in Children and Adolescents

While the reporting of physical symptoms appears to increase with age, somatic presentations are also very common in children. Several reports, primarily from consultation-liaison child psychiatrists, highlight the relevance and frequency of DSM-5 SSRDs in children and adolescents in pediatric settings. According to these reports, SSRDs account for 15% to 20% of psychiatric consults in the pediatric hospital setting and have been reported as the second most common reason for psychiatric consultations related to suicide attempts. As part of a growing effort to standardize the evaluation and management of pediatric SSRDs, the American Academy of Child and Adolescent Psychiatry (AACAP) has published a clinical consensus pathway for SSRD evaluation and management.

However, general pediatricians continue to utilize such labels as “idiopathic’ or “amplified pain syndrome” (APS) in the case of children presenting medically unexplained musculoskeletal pain, complex regional pain syndromes, headaches, or functional stomach pain. Some disagree with such terminology because they regard it as stigmatizing.

ETIOLOGY

Somatic presentations of mental disorders should be conceptualized using a biopsychosocial model that includes genetic and biologic vulnerability, environmental influences, as well as psychological and behavioral elements. According to this model, a number of contributing factors converge into the final common pathway of “somatic presentations.” This appears to be a complex process in which genetic and biologic vulnerability (increased sensitivity to pain, proprioceptive acuity), early traumatic experiences (violence, abuse, deprivations), learning factors (attention obtained from having an illness, lack of reinforcement of nonsomatic expressions of distress) and psychological elements, lead to a characterologic style and a behavioral repertoire that focus on the somatic and conveys a diminished ability to express emotionally laden feelings. These traits may be more likely displayed in certain cultural/social frames that devalue or stigmatize psychological suffering in contrast to “genuine” medical disorders. This is reinforced or stimulated by psychiatric comorbidities, and in some instances, secondary gain (such as financial compensation).

Environmental and Cultural Factors

“Idiopathic” physical symptoms have been reported in most cultural groups, although the types of symptoms and their prevalence appear to differ across cultures. The phenomenon of “somatization” has also been well documented in international studies. Differences in medical care across cultures obviously affect the presentation, recognition, and management of these somatic presentations. In general, “idiopathic” physical symptoms appear to be more common among individuals from lower socioeconomic strata, patients from developing countries, and, in the United States, those from certain ethnic groups.

In the case of Hispanic populations, a relatively high frequency of SSRDs has been reported in epidemiologic studies, particularly among Puerto Rican and Mexican American populations.

In the case of Asian populations, both Asians in general and also Asian Americans, somatic symptom presentations also appear to be relatively common and are frequent manifestations of distress.

In these and other cultural groups, clinical presentations dominated by idiopathic physical symptoms may represent “idioms of distress.” Lay explanations embedded in culture concerning the origin of behavioral symptoms (e.g., *ataque de nervios* among Puerto Ricans) may play a role. Variations in symptom presentation are likely the result of the interaction of multiple factors within cultural contexts that affect how individuals identify and classify bodily sensations, perceive illness, and seek medical attention for them. Somatic presentations frequently follow traumatic events. Thus, somatic presentations can be viewed as expressions of personal suffering inserted in a cultural and social context. L.J. Kirmayer outlined a useful process for framing somatic symptoms that incorporates disease/disorder, symbolic expression of intrapsychic conflict, psychopathology, metaphorical expression of distress, positioning in the world, and formulation of commentary or social protest. According to Kirmayer, “somatization” and “psychologization” are not mutually exclusive ways in which to express distress. He advocated the adoption by clinical investigators of a sociocultural perspective to unravel the meaning of these somatic symptoms and to design effective therapeutic strategies.

Biologic Factors

There is surprisingly limited information on the biology of somatoform disorders. Neurophysiologic studies had suggested that somatoform and functional somatic syndromes might be related to proprioceptive acuity, abnormality of autonomic and proprioceptive responses, problems with the pituitary–hypothalamic axis, and other nonspecific findings. More recent studies using brain imaging techniques point to interesting differences in patients with conversion disorder and functional somatic syndromes such as irritable bowel syndrome compared to controls or patients with demonstrated organic pathology. A study by Jon Stone and coworkers using functional magnetic resonance imaging (fMRI) showed a different type of brain activation in patients with conversion disorder, which suggested a more complex level of mental activity in conversion than in control patients. Another study by Bruce D. Naliboff and collaborators examined cerebral activation in patients with irritable bowel syndrome and control subjects during recto sigmoid stimulation. The results showed differences in the activity of the right hemisphere and hypoperfusion of the nondominant hemisphere in the patients with functional GI syndromes. A study on a large adolescent cohort (TRAIL study) suggested specific associations between cortisol responses and certain clusters of functional somatic symptoms. SSDs have generally not been studied in contemporary genetic research. However, epidemiologic studies have suggested that there is a familial predisposition in the case of functional syndromes such as fibromyalgia. Studies in the 1950s and 1960s had also suggested that conduct disorder during childhood predisposed toward hysteria (somatization disorder) during adulthood and that antisocial personality in males may be associated with hysteria in female relatives.

Using fMRI, Korean investigators assessed functional connectivity between different areas of the brain in patients with SSDs compared to controls. They report that in three of the four brain networks studied, functional connectivity was greater in the patients with SSD compared to controls. Scores on the somato-sensory amplification scale were correlated with the level of functional connectivity. Based on these findings, SSD may be related to changes in sensory processing of pain and other symptoms, a process that is influenced by emotional processing. According to the authors, “patients with SSD have a deficit in attention, leading to misperception of external stimuli and failure to regulate bodily functions aimed at interactions with external stimuli.” Such observations may lend new insights into the underlying neurobiology of SSD.

Psychological Factors

Psychological factors play a significant role in somatic symptom presentations. Such traits as suggestibility, dramatic demeanor, flair, and flamboyance were all related to the classic notion of hysteria. Some of these traits survived in more recent nomenclatures and were included in the personality disorders category (“hysterical personality” and, more recently, “histrionic personality”). The concept of “alexithymia,” used to describe an inability to express and normally process emotions, has also been associated with a tendency to emphasize somatic rather than psychological symptom presentations.

Attitudinal/Behavioral Factors

Learned behaviors and the attainment of “secondary” or “psychological” gain and “advantages” derived from the assumption of the sick role are all factors that may also play a role in the initiation and maintenance of somatoform symptoms.

DIAGNOSIS AND CLINICAL FEATURES

Two separate case reports will be used here to illustrate typical presentations and degrees of severity of SSD. These case reports will highlight diagnostic, prognostic, and management issues for these syndromes.

MILD SOMATIC SYMPTOM DISORDER

Mrs. M, a 40-year-old immigrant from Latin America, presented to the primary care clinic with health anxiety and a number of specific somatic complaints. She had no previous psychiatric or medical history. She consumed alcohol occasionally but denied any other substance abuse. A systematic assessment disclosed the presence of several physical symptoms in the last few years, mostly related to pain, that she was constantly worried about and that had resulted in either medical consultation or significant distress. These included low back pain, stomach pain, pain in arms, headaches, and “dizzy” spells. Upon careful probing, these symptoms did not appear to have a clear medical explanation. Some of these symptoms had first occurred early in life (before the age of 15 years) and had continued intermittently. The primary care evaluation of mental disorders (PRIME-MD) elicited diagnoses of minor depression and anxiety disorder not otherwise specified. Depressive symptoms included diminished interest or pleasure, decreased energy/fatigue, and feelings of worthlessness. Anxiety symptoms included a subjective sense of worry, agitation, muscle pain, concentration difficulties, and irritability. These symptoms did not occur nearly every day but only a few days per week. Although she reported occasional thoughts that life was not worth living, she denied suicidal thoughts. The patient was managed in the primary care setting with a modified cognitive intervention that included relaxation techniques (diaphragmatic breathing, a recommendation for brief daily naps, etc.) and resulted in a relatively fast improvement of somatic, mood, and anxiety symptoms. She did not require psychotropic medications. Treatment was ended after the 10th weekly session, and the improvement in physical symptoms was maintained during a 1-year follow-up period.

SEVERE, PERSISTENT, SOMATIC SYMPTOM DISORDER

Mr. K, a white man in his mid-50s, consulted a general medicine clinic complaining of gastrointestinal problems. Major presenting symptoms were a long list of physical symptoms and concerns mostly related to the gastrointestinal system. These included abdominal pain, left lower quadrant cramps, bloating, persistent sense of fullness in stomach hours after eating, intolerance to foods, constipation, decrease in physical stamina, heart palpitations, and feelings that “skin is getting yellow” and “not getting enough oxygen.” A review of systems disclosed disturbances from virtually every organ system, including tired eyes with blurred vision, sore throat and “lump” in the throat, heart palpitations, irregular heartbeat, dizziness, trouble breathing, and general weakness. The patient reported that symptoms started prior to the age of 30 years. For more than a decade, psychiatrists, general practitioners, and all kinds of medical specialists, including surgeons, had seen him. He used the Internet constantly and traveled extensively in search of expert evaluations, seeking new procedures and diagnostic assessments. He had undergone repeated colonoscopies, sigmoidoscopies, computed tomographic (CT) scans, MRI studies, and ultrasound examinations of the abdomen that had revealed Barrett esophagus but no other pathology. He was on disability and had been unable to work for more than two years due to his condition. About 3 years before his visit to the clinic, his abdominal complaints and his fixed belief that he had an intestinal obstruction led to an exploratory surgical intervention for the first time, apparently with negative findings. However, according to the patient, the surgery “got things even worse,” and since then, he had been operated on at least five other occasions. During these surgeries, he has undergone subtotal colectomies and ileostomies due to possible “adhesions” to rule out “mechanical” obstruction. However, available records from some of the surgeries do not disclose any specific pathology other than “intractable constipation.” Pathologic specimens were also inconclusive. A complete physical and neurologic examination was normal except for an examination of the abdomen, which revealed multiple abdominal scars. Right ileostomy was present, with soft stool in the bag and active bowel sounds. There was no point tenderness and no abdominal distension. During the examination, the patient kept pointing to an area of “hardness” in the left lower quadrant that he thought was a “tight muscle strangling his bowels.” However, the examination did not disclose any palpable mass. Skin and extremities were all within normal limits, and all joints had a full range of motion and no swelling. Musculature was well developed. Neurologic examination was within normal limits. The patient was scheduled for brief monthly visits by the primary care physician, during which the doctor performed brief physicals, reassured the patient, and allowed the patient to talk about “stressors.” The physician avoided invasive tests or diagnostic procedures, did not prescribe any medications, and avoided telling the patient that the symptoms were mental or “all in his head.” The primary care physician then referred the patient back to psychiatry. The psychiatrist confirmed a long list of physical symptoms that started before the age of 30 years, most of which remained medically unexplained. The psychiatric examination revealed some anxiety symptoms, including apprehension, tension, uneasiness, and somatic components such as blushing and palpitations that seemed particularly prominent in social situations. Possible symptoms of depression included mild dysphoria, low energy, and sleep disturbance, all of which the patient blamed on his “medical” problems. The mental status examination showed that Mr. K’s mood was rather somber and pessimistic, although he denied feeling sad or depressed. Affect was irritable. He was somatically focused and had little if any, psychological insight. The examination revealed the presence of a few life stressors (unemployment, financial problems, and family issues) that the patient quickly discounted as unimportant. The psychiatrist diagnosed DSM-5 Somatic Symptom Disorder, severe type. Although the patient continued to deny having any psychiatric problems or any need for psychiatric intervention or treatment, he agreed to a few regular visits to continue to assess his situation. He refused to have anyone from his family involved in this process. Efforts to engage the patient with formal therapy such as CBT or a medication trial were all futile, so he was seen only for “supportive psychotherapy” with the hope of developing rapport and preventing additional iatrogenic complications. During the follow-up period, the patient was operated on at least one more time and continued to complain of abdominal bloating and constipation and to rely on laxatives. The belief that there was a mechanical obstruction of the intestines continued to be firmly held by the patient and bordered on the delusional. However, he continued to refuse pharmacologic treatment. The only medication he accepted was a low-dose benzodiazepine for anxiety. He continued to monitor his intestinal function 24 hours per day and to seek evaluation by prominent specialists, traveling to high-profile specialty centers far from home in search of solutions.

Comorbidity of Somatic Symptom Disorders

Patients with SSDs present complicated social histories, including instances of abuse and neglect. The presenting symptoms can be quite broad, including features of mood, anxiety, personality disorders, and, to some extent, substance use disorders (at least the abuse or misuse of prescription medications). In primary care studies, greater than 90% of patients presenting with high levels of physical symptoms qualify for a depression or anxiety diagnosis. The comorbidity with other disorders is also quite elevated, indicating significant overlap across the various categories. Thus, greater than 60% of patients with a dissociative disorder also meet the criteria for somatization disorder.

DIFFERENTIAL DIAGNOSIS

Because SSDs are placed at the crossroad between physical and mental disorders, their differential diagnosis tends to be quite inclusive and elaborate. However, there are several features of these disorders that can help the differential diagnosis. Research in primary care has shown that as the number of physical symptoms increases, so does the likelihood that those individuals will also meet the criteria for a psychiatric disorder, not a medical disorder.

The following features can help in deciding whether idiopathic physical symptoms may have a psychiatric etiology:

- The symptoms coexist with major psychiatric disorders such as depression or panic.
- The symptoms closely follow traumatic events.
- The symptoms lead to psychological “gratification” or “secondary gain.”
- The symptoms represent predictable personality traits (coping mechanisms).

- The symptoms become persistent, join a conglomerate of other symptoms, and convey such attitudes as overuse of medical services and dissatisfaction with medical care.

The more of the foregoing features that are present, the more likely it is that symptoms may be classified as SSD. Differential diagnosis from other psychiatric disorders is also difficult because many of the somatic symptoms may be related to a psychiatric disorder such as depression (e.g., pain symptoms), anxiety (e.g., cardiorespiratory and GI symptoms), or even psychotic disorders (somatic delusions). However, the presence of a high level of idiopathic physical symptoms has to be taken into account even if they appear in the context of what is viewed as another primary disorder such as depression or anxiety because they are likely to affect symptom severity, treatment response, and level of disability.

COURSE AND PROGNOSIS

The course of most SSDs tends to be chronic, which is to be expected in the case of syndromes so closely associated with personality and cognitive styles. Epidemiologic studies have shown that the type and number of physical symptoms often change during follow-up periods, suggesting a certain inconsistency in symptom presentation. However, according to international studies, although idiopathic symptoms change, the number of symptoms present appears to remain high. The presence of high levels of idiopathic physical symptoms appears to be very disabling, leading to higher levels of disability than those reported for most medical and psychiatric disorders. Research in primary care has shown that psychiatric disorders such as depression or anxiety become more severe and disabling if they are associated with high levels of medically unexplained physical symptoms.

TREATMENT

Establishing an Alliance

Patients with severe and protracted somatic syndromes, as in the case presented earlier, approach medical encounters with a mixture of magical, unrealistic expectations, pessimism, and distrust of the profession. These attitudes are based on their previous experiences with physicians conveying a lack of interest or disbelief in the patient’s complaints and suffering. Building a trusting alliance in this context must begin with respect for the patient’s symptoms and an acknowledgment of their validity. Active, receptive listening, tolerance for repetition, and a “neutral” approach (avoiding being dismissive, confrontational, or overly reassuring) are essential skills. It takes time and persistence to develop a trusting, therapeutic partnership.

Taking the History

Many of these patients usually bring with them “thick” charts. These include descriptions of many clinical encounters and multiple tests and procedures that are often redundant and ordered without clear rationale. The prospect of reviewing these medical records is challenging and often leads to a negative attitude from the physician. The physician should keep an open mind, despite forewarnings in the medical records, and perform an independent assessment of the patient. An emphasis on psychological questioning and interpretations should be avoided at this stage. Premature reassurance, although seemingly appropriate from the physician’s perspective, may be perceived by the patient as disinterest or dismissiveness.

Reassurance

The humanistic principles of the medical profession are well summarized in the maxim attributed to Oliver Wendell Holmes that the role of the physician is “to comfort always.” Beyond the specificity of diagnoses and the effectiveness of treatments, the traditional practice of medicine gave high importance to bedside manners and interpersonal issues. However, 21st-century medicine, with its reliance on technology and the changes in practice patterns (e.g., managed care and complex electronic medical records), leaves little time for face-to-face interactions, which is gradually eroding the doctor-patient relationship. It would appear that the profession has drifted away from the role “to comfort,” meaning “to give strength and hope; to ease the grief or trouble; to cheer,” to “to reassure,” simply meaning “to assure anew; or restore to confidence.” In this new context, “reassurance” simply means letting patients know that their symptoms do not appear to be caused by physical disease. Unfortunately, this does not seem to work very well in the case of many patients presenting with “idiopathic” physical symptoms. The timing and degree of reassurance must be based on the adequacy of data and the trust and security of the relationship.

Physical versus Psychological Focus

Patients with SSDs appear to feel most comfortable providing details of symptoms in physical or somatic terms. However, they may also bring a repertoire of significant events (e.g., losses, trauma, disappointments, and deprivations) that may be related to the onset or persistence of symptoms and need to be explored at some point. However, many patients do not readily acknowledge or recognize “emotional” issues and will feel more comfortable dealing with questions related to their physical symptoms than questions related to psychological issues. Taking a history of multiple physical complaints can be carried over into subsequent appointments. A thorough physical examination should follow history taking at the initial visit, and briefer physical assessments should be performed in subsequent visits. An inclusive drug history, including prescription and nonprescription remedies, also should be completed. As history taking moves along, attitudes, beliefs, and attributions should become clearer and patterns of interaction and illness behavior discernible. Distorted beliefs, contradictory ideas, and fears can be addressed at several points during the process of gathering historical data.

General Therapeutic Approach

Because of the refractoriness of idiopathic physical symptoms to traditional interventions, physicians and therapists tend to be more successful when adopting a “caring” rather than a “curing” approach with these patients. Restraint in the use of medication is advised unless these are clearly indicated for the relief of comorbid affective and anxiety symptoms. Clear assignment of appointment times at regular but relatively infrequent intervals (about once per month) is preferable to random appointments based on whim or fluctuation in symptoms. A clearly spelled-out treatment plan is highly desirable. As the rapport and trust increase, the patient may be more willing to discuss their social world, family relationships, and the “stressors” that may be contributing to the symptoms. Although true insight may never fully develop, some awareness of emotional components could make the patient more receptive to referral for specialized behavioral intervention. Even if a referral is successful, the patient should continue to see both the primary care physician and the psychiatrist. An ongoing working collaboration between the psychiatrist and the primary care physician should lead to better outcomes.

SPECIFIC SOMATIC SYMPTOM AND RELATED DISORDERS

Somatic Symptom Disorder

In DSM-III, “somatization disorder” was the direct descendant of the hysteria/Briquet syndrome diagnosis described in the pioneer studies of the St. Louis group as one of the few valid diagnoses in psychiatry. This separated the concept of multiple idiopathic physical symptoms from the concepts of conversion and dissociative hysteria, which became separate diagnostic entities in the new nosologies. Once incorporated into the DSM process, “somatoform disorder” underwent various changes during the subsequent revisions of the nomenclature up to DSM-5. In clinical practice, somatization disorder appeared to be quite rare, and there was disillusionment about its relevance in primary care settings and, in general, pervasive pessimism about its treatment and prognosis. “Full” somatization disorder (in DSM-III and IV) represented the severity extreme in dimensional models of somatization, and there was ample debate in the field about its survival in future editions of the DSM. Indeed, DSM-5 eliminated this diagnosis, merging it into the broader category of somatic symptom disorder.

As discussed earlier, hysteria was the precursor of the concept of somatization and also of the DSM diagnosis of somatization disorder. In the golden age of descriptive psychopathology, the formal description of a syndrome with multiple neurologic symptoms/complaints without organic findings (still called hysteria) was made by Pierre Briquet in France in 1859, and this began the separation of this polysymptomatic disorder from conversion hysteria. Toward the middle of the 20th century, North American nosologists from Washington University and Barnes Hospital in St. Louis started using a systematic approach to validate psychiatric diagnoses. One of their first targets was the diagnosis of hysteria. These investigators proposed a “revised” Briquet syndrome that included many other physical symptoms besides the neurologic ones, as well as psychiatric and behavioral components. In the DSM classification, before

DSM-III established the syndrome as somatization disorder, it had been called “conversion reaction” in DSM-I (1951) and “hysterical neurosis, conversion type” in DSM-II (1968).

Comparative Nosology

EVOLUTION OF THE DIAGNOSIS OF SOMATIZATION DISORDER. The confusion that arose in differentiating hysteria from hysterical conversion reactions in DSM-III was at least temporarily resolved by retaining Briquet hysteria as somatization disorder and hysterical conversion as conversion disorder. Briquet virtual encyclopedia of purely neurologic symptoms was reduced in DSM-III to 14 symptoms for women and 12 for men out of a possible 37 symptoms attributable to several organ systems (cardiorespiratory, musculoskeletal, genitourinary, neurologic, and GI). Rather arbitrarily, DSM-III-R revised the criteria to require 13 symptoms for both males and females. Finally, in DSM-IV, the criteria were changed once again. This time the revisions were stimulated by a need to simplify the criteria and by the reliance on epidemiologic data sets from outside of the United States. The same data set of adopted Swedish community residents for whom detailed medical records were available and which had led to the distinction of subtypes of somatization in the 1980s (high-frequency vs. low-frequency somatizers) was used again for this purpose. A method based on discriminant analysis was derived to detect “somatizers,” and this led to four discrete sets, including bodily pains, GI symptoms, pseudo neurologic symptoms, and sexual or menstrual symptoms, that were the basis for the new diagnostic formulation. Thus, DSM-IV continued the deflationary trend and modified the criteria to have at least eight physical complaints referable to four pain sites or functions (e.g., back, chest, urination), two nonpain GI symptoms (e.g., nausea, bloating), one nonpain sexual or reproductive system symptom (e.g., menstrual irregularity, loss of libido), and one pseudo neurologic symptom (e.g., urinary retention, aphonic, blindness). Somatic symptoms should remain medically unexplained to quality as somatoform disorder symptoms, although “exaggerations” of ordinarily expected symptoms of coexisting physical disease can also be considered in the symptom count. Somatoform disorder symptoms generally occur over a period of several years and are accompanied by significant impairment of social and occupational function and high utilization of medical resources, usually resulting in either ineffective and/or unnecessary medical/surgical treatments as illustrated in the case presented earlier.

DSM-5'S SOMATIC SYMPTOM DISORDERS. DSM-5 took a new tack, completely redesigning this category, renaming it "somatic symptom and related disorders." There was concern about the very limited use of somatoform disorder diagnoses both in psychiatric practice and primary care. Somatization disorder in DSM-III and IV was a restrictive category, based on simple counts of medically unexplained symptoms, thus neglecting important attitudinal and behavioral elements. Moreover, the reliance on an absence of medical explanations to catalog symptoms was not precise or practical. Therefore, DSM-5 added key attitudinal and behavioral requirements (B criterion) and no longer require that symptoms be medically unexplained.

ICD-11'S BODILY DISTRESS DISORDER (BDD). As briefly described above, BDD is characterized by the presence of bodily symptoms that are distressing to the individual, but the diagnosis also requires that the individual places "excessive" personal attention on the symptoms. Thus, a key feature of both BDD and SSD is the specific focus on psychologic/emotional features of these disorders, such as excessive thoughts, feelings, or behaviors related to the symptoms and associated health concerns. Similar to what was the case with SSD in DSM-5, BDD frequently involves multiple bodily symptoms, but a single, severe symptom such as pain is acceptable for diagnostic purposes.

DELETING, RENAMING, AND MERGING OTHER DSM-IV SOMATIC CATEGORIES IN DSM-5. In DSM-IV, *hypochondriasis* overlapped with *somatization disorder,* and the label itself was stigmatizing. *Pain disorders* were quite nonspecific, and these diagnoses were predominantly used in settings devoted to pain management. The most common somatoform diagnoses used in practice were *undifferentiated somatization disorder* and *somatoform disorder not otherwise specified (NOS),* both quite imprecise and rather meaningless labels. Thus, DSM-5 eliminated a few categories, renamed most, and merged others (see Table 18–1 for criteria for DSM-5's somatic symptoms and related diagnoses).

Epidemiology. The prevalence of somatic symptom disorder is not yet known. Because "medical explanation" for the symptom is no longer an exclusion and the somatic symptom threshold is much lower than that of DSM-IV, it has been argued that the prevalence of the DSM-5 disorder could be much higher than that of “full” somatization disorder in DSM-IV. However, the requirement in DSM-5 for “disproportionate and excessive thoughts, feelings, and behaviors” in addition to the reported somatic symptoms makes the criteria more restrictive, and this may affect its overall prevalence in the opposite direction, that is, lowering the prevalence. This conundrum will only be resolved after numerous well-designed, systematic studies are performed in various communities and clinical settings. These studies may require the use of more sophisticated instruments such as the Schedules for Clinical Assessment in Neuropsychiatry-Present State Examination (SCAN-PSE) that can elicit both the symptoms and the psychological/behavioral components of this disorder.

Worldwide, a large majority of the comparative surveys that examined the old diagnosis of somatization disorder consisted of simple counts of physical symptoms with some effort made to rule out medical explanations. In many studies, the presence of high levels of physical symptoms was derived from physical symptom inventories such as those included in the symptom checklist 90 (SCL-90), the general health questionnaire (GHQ), and particularly the diagnostic interview schedule (DIS) and the Composite International Diagnostic Interview (CIDI). The somatic syndromes were articulated as either “full” or “abridged” somatization disorders. Problems with these studies include variations in key instruments, ways of survey administration, format of answers, and the reliance on lifetime rather than current symptoms. In general, somatoform disorder and abridged somatoform disorder have been reported to be more common among females, individuals from lower socioeconomic strata, and, in the United States at least, among people from certain ethnic groups, such as Latinos. In addition, the comorbidity of high levels of these symptoms with psychiatric syndromes such as depression and anxiety has been well documented for most cultures and ethnic groups, although there appears to be cross-cultural variation in frequency and severity of the associated physical symptoms.

GENERAL POPULATIONS. Studies in the United States, Puerto Rico, Germany, and Italy found lifetime prevalence rates of “full” somatization disorder ranging from 0.1% in the United States to 0.8% in Germany, whereas lifetime rates of “abridged” somatization disorder ranged from 5.6% in Germany to 19% in Puerto Rico. Much greater prevalence rates have been reported for more broadly defined somatic symptom clusters. For example, 22% of the general population in Germany reported at least one unexplained physical symptom leading to severe impairment. Unfortunately, some of the most recent large-scale epidemiologic studies in the United States and elsewhere, such as the national comorbidity survey (NCS), the national epidemiologic survey of alcohol and related conditions (NESARC), and the world mental health surveys, did not systematically assess unexplained physical symptoms, possibly due to the time and effort it takes to go through long and tedious physical symptom inventories.

Etiology

GENETIC AND BIOLOGIC FACTORS. Some of the early family studies of the St. Louis group had suggested an association between hysteria/Briquet syndrome in females and antisocial personality in their male first-degree relatives. In addition, studies of adopted-away children highlighted the presence of alcoholism and violence in biologic fathers of women with somatization disorder. However, no family studies have been performed since the new diagnostic criteria were developed.

Treatment

GENERAL CONSIDERATIONS. Historically, the treatment of patients with severe SSDs has been considered difficult and frustrating. Encounters of these patients with their physicians generally evoke frustration on both sides. When Freud began his practice as a clinical neurologist in Vienna, many of his patients (clearly many of those on whom the original theories of psychoanalysis were based) were “difficult” patients with dramatic physical presentations who were referred by other physicians who had failed to help them. Although the form of symptoms has been reshaped by cultural change and shifts in medical paradigms, this situation is not that different from the realities faced by many physicians today. Such a context represents a profound challenge to the most well-intentioned, dedicated physicians.

Besides the numerous physical symptoms, somatic symptom disorder embodies traits that complicate therapeutic encounters and may render patients refractory to treatment. These traits include denial of psychological components; chronicity; lack of response to traditional types of reassurance; unshakeable belief about the presence of physical disease; high/inappropriate utilization of medical resources; risk of “occult” medical comorbidities; sensitivity to rejection; and frequent dysfunctional patient-physician relationships. Therapeutic goals should be modest at first and be limited to small, attainable gains such as a decrease in medical visits, a commitment to a single primary care physician, and the avoidance of unnecessary tests and procedures. The physician must be prepared for a long-term commitment to the patient. Specialist consultations should be limited and restricted to cases with evidence of comorbid physical conditions. Some of the elements of CBT, such as diary keeping, should be incorporated into the primary care physician’s treatment strategy. Recommendations of such activities as exercise, yoga, relaxation, meditation, and massage may be useful and are generally better accepted by these patients than psychological recommendations. The major challenge, therefore, continues to be finding practical and effective ways to help primary care physicians to recognize, understand, and manage these patients in the context of a supportive doctor-patient relationship at the primary care site. The current emphasis on learned behaviors leads to the choice of cognitive-behavioral interventions, whereas other conceptualizations may encourage more relational, interpersonal, or psychodynamic approaches. Psychopharmacological approaches may be used independently or combined with other interventions. Until recently, interventions with somatoform disorder patients generally had a pessimistic outlook, and there was a paucity of well-controlled studies. Other issues that delayed effective interventions included the fact that somatoform disorder patients were more likely to be seen in primary care settings where therapists may be poorly prepared to work with demanding, frustrating patients, and comorbid physical disease and psychiatric disease are frequent confounders.

PHARMACOTHERAPY. Medications have a limited role in the management of the patient with somatoform disorder except in the presence of clearly delineated anxious, depressive, or psychotic symptoms. Despite the development of new pharmacologic agents such as dual (norepinephrine and serotonin) reuptake inhibitors and a preliminary report on the possible efficacy of gabapentin, there have been relatively few studies that employ a purely pharmacologic approach in treating somatoform or SSDs.

SMITH'S CONSULTATION LETTER. A groundbreaking controlled study by Richard Smith and colleagues at the University of Arkansas for Medical Sciences demonstrated the effectiveness of a rather simple intervention. This consisted of a brief "consultation letter" intended for primary care physicians providing them with "dos and don'ts" regarding their encounters with patients with multiple medically unexplained physical symptoms and briefly instructing them on the use of a few key management techniques. The letter urged the physicians to see these patients during regularly scheduled appointments, perform brief physical examinations focusing on the area of discomfort at each visit, avoid unnecessary diagnostic procedures, invasive treatments, and hospitalizations, and avoid using statements such as "symptoms are all in your head," and briefly allow/encourage patients to talk about "stressors." Although patients did not show a significant change in terms of somatic symptom relief per se, their functional capacity improved significantly, and there was a decrease in their utilization of health resources that generated significant cost savings.

REASSURANCE AND SUPPORTIVE THERAPY. It is generally accepted that patients with somatoform disorder respond poorly to the simple reassurance that there are no positive findings in the physical and laboratory workup. Although no adequately controlled studies of psychotherapeutic intervention for "full" somatoform disorder existed prior to the study by Smith and colleagues, a number of studies of undifferentiated forms of somatoform disorder had suggested therapeutic benefits from the presence of an accepting attitude in the therapist, a shift in the patient's attention from somatic to emotional features, and the use of group approaches focusing on explanation, support, relaxation, and cognitive-behavioral approaches emphasizing adaptation to chronic somatic distress.

PSYCHOTHERAPY. There are reports indicating that brief psychodynamic therapy of unexplained somatic symptoms may be effective, and at least one of these studies was a randomized, clinical study comparing brief psychodynamic therapy to regular medical treatment alone. However, most of the recent evidence in terms of randomized clinical trials involves CBT.

Cognitive-Behavioral Therapy. A body of research has reported on the efficacy of CBT for the treatment of full somatization disorder and related syndromes. CBT-type interventions appear to help patients by modifying thoughts and behaviors associated with somatization. In the typical CBT program, patients are systematically exposed to a number of behavioral techniques, which include relaxation training and graded increases in activities. From a cognitive perspective, CBT helps these patients to identify thoughts that contribute to increased stress, inactivity, and health concern. Often, patients with somatoform disorder syndromes may have a tendency to think catastrophically about their physical symptoms. Such thoughts lead them to conclude that they are sick and must limit physical activity, creating a cycle that perpetuates the somatic process. There are several large, well-designed studies that support the use of CBT-type interventions for patients suffering from somatization disorder and related syndromes. In primary care settings, studies between 1995 and 2000 strongly supported the efficacy of CBT for functional somatic syndromes.

In Germany, the group led by Professor Winfried Rief did a study in 140 primary care patients randomized to either *one* session of CBT or “standard medical care” and reported excellent improvement and very good acceptance of CBT. It must be noted that in Germany, there exist psychosomatic units in hospitals that specialize in managing these patients (and have the time and the needed resources to do so thoroughly). It is generally acknowledged that to effectively treat patients with somatic symptom disorder, the primary physicians must be recruited as participants. In the United States, Robert Smith and colleagues reported on a randomized clinical trial of patients with “high service utilization” due to high levels of medically unexplained symptoms treated with cognitive-behavioral and pharmacologic interventions administered by primary care clinicians (nurses). This intervention led to significant clinical improvement in mental and physical functioning.

Pivotal Study. In a 2006 study performed at a mental health care system in New Jersey by L. A. Allen and coworkers, 84 patients meeting full criteria for somatization disorder were randomly assigned to either standard medical care augmented by a psychiatric consultation letter or to a 10-session, manualized, individually administered CBT added to the psychiatric consultation letter. The primary outcome measure was the severity scale of the clinical global impression scale for somatization disorder (CGI-SD). Secondary outcome measures were responder status as determined by clinical ratings, self-report measures of physical functioning and somatic symptomatology, and health care utilization assessed via medical records. Results of this study showed that 15 months after baseline, somatization symptoms were significantly ($P < .001$) less severe in the group treated with CBT. Patients treated with CBT also were significantly more likely to be rated as either “very much improved” or “much improved” than patients treated with only augmented standard medical care (39.5% vs. 4.9%). CBT was associated with greater improvements in self-reported functioning and symptomatology and led to a greater decrease in healthcare costs. The authors concluded that for patients diagnosed with somatoform disorder, CBT might produce clinical benefits beyond those that result from current state-of-the-art treatments.

Illness Anxiety Disorder

Together with conversion hysteria, the old notion of hypochondria has had a long and eventful history in psychiatry. Given the stigma attached to the original term, the name changed in DSM-5. DSM-5 recognizes that most cases of hypochondriasis can be subsumed under the rubric of SSD. However, a minority of patients with hypochondriasis (~20%) are now given a diagnosis of illness anxiety disorder. The distinction lies in the severity of the somatic symptoms. Where somatic symptoms are relatively minor, and patients instead focus on concerns that they will get sick or have an undiagnosed illness, the diagnosis is illness anxiety disorder. When there is a preoccupation with health concerns superimposed by substantial somatic symptoms, the diagnosis is somatic symptom disorder. Phenomenologically, health anxiety can be broken down into several components, including idiopathic physical symptoms, such as amplification of bodily sensations; certain characterologic features, such as low self-esteem, guilt, passivity, and personal suffering; and learned behaviors, such as the assumption of the sick role. The core features include disease conviction, disease fear, and bodily preoccupation.

Definition. In DSM-5, key features of health anxiety are preoccupation with physical symptoms and the belief that they may be due to a serious physical illness that persists despite medical reassurance or objective evidence on the contrary. Some experts subscribe to a dimensional concept of health anxiety, according to which these symptoms exist on a continuum from the heightened awareness of bodily function to extreme delusional bodily preoccupation. Indeed, the fear of disease can become so intense that it has been given a phobic connotation, as is the case with historical terms such as "syphilophobia," "thanatophobia," "acquired immunodeficiency syndrome (AIDS)-phobia," and several others. These fears are particularly prevalent at times when certain diseases are thought to be incurable or terminal. The preoccupation with health issues can also become fixed and reach delusional (psychotic) proportions.

History. Hippocrates suggested that this disorder had its anatomical origin in organs located under the hypochondrium—literally below the cartilage or ribcage—and this notion was expanded by other classical Greeks such as Plato and Diocles. In the 16th and 17th centuries, Robert Burton, Thomas Willis, and Thomas Sydenham attributed the origin of hypochondria to "disturbances of the mind," and it was viewed by some as the male equivalent of hysteria. Since those early times, clinical descriptions of hypochondria have seemed very consistent with those formalized in modern diagnostic systems. In the mid-1600s, the French dramatist Molière in his satire "Le Malade imaginaire," depicted a hypochondriac who feared the ministrations of physicians. Ironically, Molière, who suffered from pulmonary tuberculosis, collapsed on stage while playing the lead role and died later that evening.

Comparative Nosology. In DSM-5, illness anxiety is defined as the preoccupation with fears of having, or the idea that one has, a serious disease based on the person's misinterpretation of bodily symptoms, with a duration of at least 6 months. The preoccupation persists despite appropriate medical evaluation and reassurance; the belief is not of delusional intensity and is not restricted to a circumscribed concern about appearance. The preoccupation causes clinically significant distress or impairment in social, occupational, or other important areas of functioning, and the preoccupation is not better accounted for by another disorder such as anxiety disorder or depressive disorder.

ICD-11 retained the diagnosis of “hypochondriasis” and defined it as “a preoccupation with or fear about the possibility of having a serious, progressive or life-threatening illness, accompanied by either repetitive and excessive health-related behaviors, such as repeatedly checking the body for evidence of illness or maladaptive avoidance behavior.” ICD-11 included “hypochondriasis” under the *obsessive-compulsive and related disorders* grouping based on shared phenomenologic features (e.g., repetitive thoughts about having an illness, repeated/excessive behaviors driven by the preoccupation) and similar treatment response. In ICD-11, the presence of somatic symptoms is not an essential feature of “hypochondriasis.”

Epidemiology. The prevalence of health anxiety can vary according to how it is defined, how the criteria are elicited, and the type of settings studied. It has been estimated that it may affect as many as 4% to 6% of all patients. Health fears seem quite common in the general population. A national study in Germany found that 10% of the population reported “health anxiety” or the conviction of having a serious disease in the absence of medical explanations. A study in Italy reported a 1-year prevalence of health anxiety of about 5%. In primary care, studies using various methodologies have shown prevalence rates of health anxiety in the 3% to 7% range. Psychiatric comorbidities are quite common among patients with health anxiety, including depression (40%), panic disorder (10% to 20%), OCD, and generalized anxiety disorder (5% to 10%). Longitudinal studies have also shown that the course of health anxiety can become chronic and incapacitating, with a majority (greater than 60%) of the cases still showing symptoms after several years of follow-up.

Treatment

REASSURANCE. The role of reassurance in health anxiety disorder has been a controversial one. Although the criteria include the inability to reassure the patient as a criterion for diagnosis, virtually all therapeutic interventions include reassurance as one of their main features. The ease with which patients with health anxiety can be successfully reassured may be related to patient factors such as chronicity, the severity of symptoms, and personality characteristics, and also to the attitudes and treatment style of the therapist. Recommendations for approaching patients with health anxiety are generally similar to those already given for approaching patients with somatic symptom disorder. However, in general, patients with health anxiety seem less demanding than those with somatic symptom disorder, at least the severe types, and some may even respond to properly conveyed reassurances. Once patients with health anxiety are systematically engaged in a therapeutic relationship, opportunities for insight development may arise.

PHARMACOLOGIC TREATMENT. Little information is available on the pharmacologic treatment of hypochondriasis. Most reports have been anecdotal, single-blind studies. Such agents as tricyclics, selective serotonin reuptake inhibitors (SSRIs), and nefazodone (Serzone) have been reported to be clinically useful, particularly in health anxiety associated with depressive symptoms. In secondary syndromes, symptoms of health anxiety have been reported to improve in parallel with anxiety or depressive symptoms. Antipsychotic agents such as risperidone have been recommended for delusional syndromes. No study has established the superiority of one drug over another in managing health anxiety. Brian Fallon and colleagues conducted a double-blind, placebo-controlled clinical trial in patients with health anxiety (hypochondriasis). Following a placebo run-in, the investigators randomly assigned patients to receive fluoxetine (Prozac) in dosages ranging between 20 and 80 mg per day or a placebo. Fifty-seven patients were enrolled. Twelve were discontinued during the placebo run-in, and 45 were randomized to either fluoxetine or placebo for 12 weeks. Responders to acute treatment (much/very much improved in CGI) entered a 12-week maintenance phase to week 24. Sustained responders at week 24 entered a double-masked discontinuation phase. More patients improved on fluoxetine compared to placebo starting at week 8 (50% vs. 19%, $P = .03$). The acute response was maintained until week 24, with more responders. The authors conclude that fluoxetine is a moderately effective and well-tolerated treatment for hypochondriasis.

PSYCHOTHERAPIES. Various group and individual therapies, including psychodynamic types, have been proposed over the years to manage patients with health anxiety. Unfortunately, there is a paucity of controlled studies. Robert Kellner, one of the pioneers in the treatment of somatic symptom syndromes, was one of the first to report positive results of behavioral interventions in health anxiety by focusing on patients' misinformation and cognitive distortions.

Cognitive-Behavioral Therapy. Several recent controlled studies have confirmed the earlier suggestions of the efficacy of CBT interventions. This substantive evidence places CBT as the prototype, first-line treatment for this condition. For example, a review of several controlled studies showed that CBT-treated patients improved significantly more than control subjects in a great majority of the studies, with a "trend" toward improvement observed in the remaining trials.

Pivotal Study. In a 2004 study by Barsky and Ahern, patients from primary care practices and volunteers responding to public announcements, all with high hypochondriasis scores, were assigned to one of two treatment groups, either a six-session, individual CBT intervention or a "medical care as usual" group. The CBT was accompanied by a consultation letter sent to the patient's primary care physician. Main outcome measures included hypochondriacal beliefs, health fears, somatic symptoms, role function, and impairment. One hundred and two individuals were assigned to CBT, and 85 were assigned to "medical care as usual." At the 12-month follow-up, CBT patients had significantly lower levels of hypochondriacal symptoms, beliefs, and attitudes ($P < .001$) and health-related anxiety ($P = .009$) than those in "medical care as usual." They also had significantly less impairment of social role functioning ($P = .05$) and intermediate activities of daily living ($P < .001$). Hypochondriacal somatic symptoms were not improved significantly by treatment. Barsky and Ahern concluded that this brief, individual CBT intervention, developed specifically to alter hypochondriacal thinking and restructure hypochondriacal beliefs, appears to have significant beneficial long-term effects on the symptoms of hypochondriasis.

Functional Neurologic Symptom Disorder (Conversion Disorder)

In DSM-IV, conversion disorder is the closest remnant of the old concept of hysteria. In its classic form (a motor or sensory deficit), it now appears to be a rare disorder, at least in Western societies. Although it was a key component of the classic syndrome of hysteria, conversion was separated from hysteria, starting with the Briquet and the St. Louis criteria, and passed as a separate category into the DSM system. In DSM-IV, the diagnosis of conversion disorder continued to invoke etiologic assumptions (psychological factors) against the otherwise “atheoretical” background of the system. DSM-5 made two crucial modifications in diagnosing this condition. First, it changed the principal name of the disorder (“functional neurologic symptom disorder”) in recognition that this term is more familiar to neurologists and arguably more acceptable to patients. Second, it clarifies that the diagnosis is made when the symptom is incompatible with known neurologic disorders. This “incompatibility” criterion is important because clinicians can be led astray by cognitive bias in evaluating such patients.

Definition. A conversion reaction is a rather acute and temporary loss or alteration in motor or sensory function that is incompatible with known neurologic disorders. Because the onset frequently coincides with psychological issues (conflict), early theorists speculated that such issues were “converted” to neurologic symptoms. The problem is that stress is omnipresent in life, and many patients present with conversion when no stressor has been demonstrated. The classic syndromes resemble neurologic syndromes. Conversion motor symptoms mimic syndromes such as paralysis, ataxia, dysphagia, or seizure disorder (“pseudoseizures”), and the sensory ones mimic neurologic deficits such as blindness, deafness, or anesthesia. There also can be disturbances of consciousness (amnesia, fainting spells). However, many clinicians continue to reserve the term conversion reaction for syndromes mimicking a neurologic disease. In ICD-10 but not DSM-5, dissociation and dissociative disorders are attached to the somatoform category, closely linked to conversion syndromes. Patients with conversion disorder usually present with symptoms suggestive of neurologic diseases such as muscle weakness, gait disturbance, blindness, aphonia, deafness, convulsions, or tremors. About one-third of patients diagnosed with conversion disorder presenting with motor symptoms also meet criteria for other Axis I psychiatric diagnoses and 50% meet criteria for Axis II diagnoses.

History. The concept of *hysteria,* derived from the Greek word for womb or uterus, implied an unwanted migration of the organ to higher sites. It was apparently known even to ancient Egyptians. In the Middle Ages, hysterical symptoms were attributed to demonic influences, and their placement at the moral level retarded the medical debate. During the

Renaissance, hysteria returned to medicine and was considered a somatic disorder by physicians, who implied a connection or pathway between the uterus and the brain. By the middle of the 19th century, Briquet provided a detailed clinical description of a somatic syndrome affecting young women to which he gave the name hysteria and whose origin he situated in the brain. Charcot described the classic form as “la grande crise hystérique,” which included phases such as the “prodromal,” “epileptoid,” “trance,” and “terminal or verbal” phases. He also proposed the term “functional lesion” in an effort to resolve the absence of physical findings. His followers, Babinsky, Janet, and Freud, continued to emphasize psychological factors. By the time Freud proposed his psychodynamic theory, Paracelsus, Thomas Willis, Sydenham, Mesmer, and Pinel had already suggested that hysteria had a psychogenic origin. Freud and Breuer jointly reported the first case of hysterical conversion, the case of Anna O. They theorized that symptoms of hysteria represented unwanted emotional distress or conflict that was suppressed and kept unconscious by the individual, only to appear in the form of medically unexplainable bodily symptoms. Freud named this process “somatic compliance” or “conversion,” and thus, with the case report of Anna O, both “conversion hysteria” and the “talking cure” were born. Since then, however, conversion hysteria has gradually lost its central role in psychopathology, being more loosely described as hysteria, hysterical conversion, or simply conversion, and the term is often used in situations without clear evidence of psychological determination.

Comparative Nosology. In DSM-I (1951), hysteria became conversion reaction; in DSM-II (1968), hysterical neurosis, conversion type; and finally, in DSM-III (1980), conversion disorder. In DSM-IV, conversion disorder is defined as the presence of one or more symptoms or deficits affecting voluntary motor or sensory function, suggesting a neurologic or other medical condition; judged to be associated with psychological conflict or other stressors, not intentionally produced or feigned; and not fully explained by other medical conditions, use of substances, or cultural nuances.

Although DSM-IV purported to be atheoretical, it continued to pay homage to psychodynamic tradition in this instance. DSM-5 emphasizes that the neurologic symptoms are incompatible with recognized neurologic or medical disorders and provides a specifier to denote whether or not a psychological stressor has been identified.

Epidemiology. The epidemiologic information on conversion disorder is limited. Estimates vary broadly: Less than 1% in the general population, 5% to 14% among general hospital medical/surgical referrals to psychiatry consultation services, and 5% to 25% in treated psychiatric outpatients. The disorder appears to be more frequent in females and can be seen in children as young as 7 or 8 years old. It is rare after the age of 35 years.

Etiology

PERSONALITY FACTORS. Some of the traditional features of hysteria included a detached, unemotional, calmed attitude in front of what appears to be a severe and turbulent illness, a trait known as “la belle indifference.” “Histrionic personality” is currently the term that summarizes the drama, flair, flamboyance, and exhibitionism attributed to these patients. However, none of these is given any relevance for the DSM-5 diagnosis of conversion disorder.

BIOLOGIC FACTORS. The neurophysiologic conceptualization speaks of an inherent defect in certain brain functions, especially those in the dominant hemisphere, that may interfere with verbal associations. In addition, functional MRI studies have shown differences in brain activation between patients with conversion disorder and controls.

PSYCHOLOGICAL FACTORS. The behavioral theory attributes conversion disorder to faulty childhood learning, with the nonadaptive behavioral responses used for secondary gain and control of interpersonal relationships. The psychoanalytic theory, on the other hand, describes symptoms as compromise formations with a primary gain of conflict resolution through the partial expression of the conflict without conscious awareness of its significance. Some have suggested a strong relationship between childhood traumatization by sexual or physical abuse and a later propensity for conversion disorder. Other studies, however, do not confirm such an association.

Treatment. Although a variety of explanations and treatments have been reported in the literature, systematic research on the treatment of conversion disorder has lagged behind the treatment of other disorders such as depression. Improved guidelines for treatment would be an important contribution, but these need to wait on methodologically controlled studies and long-term follow-up. Although psychological etiology is a requisite of DSM-IV diagnosis, it is often difficult to identify it early in treatment and may not reveal itself until extensive additional history is obtained. Screening methods for assessing trauma, dissociative experience, and post-traumatic stress disorder (PTSD) should be helpful in this process. Any assessment of conversion symptoms must take into account idiosyncratic cultural or religious beliefs that may require specific culture-bound interventions for the resolution of conflict and symptom amelioration. DSM-5 requires that careful examination document that the neurologic symptom is incompatible with known neurologic disorders.

APPROACH TO THE PATIENT. The usual caveats on history taking described for other disorders also apply here. There should be special attention given to a history of trauma, sexual and physical abuse, and a family history of conversion symptoms. The physical examination must pay particular attention to ruling out neurologic diseases such as multiple sclerosis and other peripheral and central nervous system disorders. Routine laboratory studies are indicated, as well as electroencephalograms (to distinguish between epilepsy and psychogenic nonepileptic seizures) and other special studies (e.g., MRI, X-rays, spinal tap, etc.) to rule out other possible organic etiologies.

REASSURANCE. Many conversion syndromes have an acute, benign course and may remit spontaneously with understanding and support. Early intervention can forestall potential chronicity and the progression into a well-entrenched somatization disorder.

HYPNOTHERAPY. This is perhaps the oldest treatment used for conversion disorder. However, there are very few randomized, controlled trials, and these seem to be inconclusive because hypnotizability was not predictive of treatment outcome.

SODIUM AMYTAL. The use of sodium amytal for diagnosing conversion disorder, uncovering traumatic events, and providing therapeutic relief was for some time a common approach to the patient with conversion hysteria. However, it is now rarely used because a number of observations questioned its clinical value.

PSYCHOTHERAPY. Once chronicity has developed, intensive treatment may use all treatment modalities, including hospitalization, individual or group therapy, insight-oriented therapies, behavioral techniques, hypnosis, sodium amytal interview, physical therapy, biofeedback, relaxation training, and medication (primarily for comorbid anxiety, depression, or other somatoform disorders). Psychological interpretations or explanations do not work well early in the process, but reassuring patients that critical tests are normal and that symptoms will eventually improve may be helpful. Any implication to the patient that they are malingering is counterproductive. Behavioral interventions should focus on improving self-esteem, the capacity for emotional expression and assertiveness, and the ability to communicate comfortably with others. It can be helpful to explore the patient's illness beliefs, predisposing vulnerabilities, precipitants, and perpetuating factors.

PHARMACOLOGICAL TREATMENT. Accompanying comorbid depression, anxiety, and behavior problems may respond to pharmacologic interventions. The use of hypnotic or narcoleptic techniques, if these are being considered, must be tentatively offered to patients whose fear of passivity or loss of control may induce overwhelming anxiety.

PHYSICAL THERAPY. With chronic conversion, muscle contractures can occur, and physical therapy is necessary. Even in the absence of such contractures, however, many conversion patients find that physical therapy can be helpful for muscular symptoms or balance problems. The process of slowly progressive exercises and activities can help restore functioning.

Definition. SSD patients presenting predominantly with pain symptoms:

These patients complain of severe, persistent pain in one or more physical areas that are of sufficient severity to warrant clinical attention and cause significant impairment in function. Like the other presentations of SSD, there is no assumption that the symptom is psychogenic or that there can be a clear-cut distinction between "physical" pain and "psychological" pain. The diagnosis of SSD with pain as a specifier is made when the symptoms of pain are accompanied by extensive cognitive, affective, and behavioral repertoires. Differential diagnosis of “psychologically” versus “physiologically” induced pain is difficult. This task is made more complicated by incidental findings such as arthritic and degenerative changes in joints, vertebral disc problems, muscle spasms, and so on. In the presence of other psychiatric disorders such as depression or anxiety, it is often impossible to ascertain whether the pain is primary or secondary (due to the other disorders).

Theories of psychogenic pain formation invoke neural pathways, but they can be best understood if they are framed within a more comprehensive biopsychosocial model. The type and quality of pain response have been attributed to genetic and physiologic vulnerabilities by some authors, whereas others emphasize psychosocial factors as the major determinants of the symptoms, including behavioral mechanisms. The notion of a “pain-prone personality” has waned in recent years, but it was part of a psychodynamic concept of chronic pain, with pain viewed as a psychological defense against guilt and anxiety possibly representing self-punishment. Patients with chronic pain often meet the criteria for DSM personality disorders, but no specific personality profile has emerged as more relevant than others. Pain in the context of other depressive symptoms and in the absence of dysphoria or sadness was often called “depressive equivalent” or “masked depression,” but this has never been officially endorsed. For a long time, psychogenic pain complaints were considered to be closely related to hysterical phenomena and described as part of classical hysterical syndromes. In previous classifications, a pain of unclear origin had been called “idiopathic,” a term that continues to have relevance in recent classifications of pain in specialties such as neurology. In the United States, pain was separated from conversion disorders, and the category of “pain-prone disorder” was first proposed in the RCD, possibly stimulated by a proliferation of pain treatment programs in the 1970s and 1980s. This led to the DSM category. DSM-III defined it as “psychogenic pain,” DSM-III-R as “somatoform pain disorder,” DSM-IV as simply “pain disorder,” and DSM-5 merging it into SSD.

Those who subscribe to the pathophysiologic relevance of brain neurotransmitters to pain syndromes are likely to focus more on the pharmacologic management of pain disorders. For decades, tricyclic antidepressants, particularly the more serotonergic ones (e.g., amitriptyline has been used in open trials, and the evidence suggests that they are effective for pain syndromes) at doses lower than the ones traditionally prescribed for depression (e.g., amitriptyline 75 to 100 mg per day).

There have been several randomized, controlled clinical studies on various pain syndromes. A placebo-controlled study of low-dose imipramine in patients with unexplained chest pain (“abnormal cardiac pain perception”) showed significantly better improvement in pain scores among patients receiving imipramine. In a placebo-controlled study of phenelzine, imipramine, and placebo in patients with mixed pain/depression syndromes, phenelzine was found to be more effective than imipramine and placebo. A recent meta-analysis of controlled studies on various functional pain syndromes showed a robust effect of antidepressant medications, particularly tricyclics, on pain and functional symptoms. It has long been noted that pain is often associated with depression. Yet, Smith found only 13 studies from 1975 to 1990 that explored this relationship. Of these studies, only four addressed the treatment of depression and coexisting pain. All were open-trial designs, with and without antidepressants (amitriptyline or doxepin), and comprehensive “pain programs.” All showed impressive rates of response to depression scores when pain was the focus of the treatment in these comorbid states. The development of dual antidepressants such as duloxetine has reawakened interest in examining somatic components of depression such as unexplained physical pain. Although this may have begun as a marketing strategy, it has served to reinforce the observation of clinicians of a significant association of depression with somatic syndromes such as pain.

Cognitive-Behavioral Therapy. CBT in noncardiac chest pain was studied in a randomized, controlled trial by a British group. Thirty-seven subjects agreed to participate in the study after being reassured by a cardiologist that they did not have heart disease. The study found superiority of CBT over the control group in improving functional limitation and decreased tendency to worry about physical symptoms. A meta-analysis of 25 trials of treatment with CBT in various functional syndromes, including pain, showed that CBT was effective for a variety of chronic pain conditions. Additional studies comparing different therapeutic modalities may help to sharpen their specificity. Although psychodynamic treatments have fallen somewhat into disfavor, the assessment of pain programs that have this orientation can help shed light on the quality of patient-therapist interactions that predict good outcomes. Finally, collaborative approaches between mental health professionals and primary care physicians should contribute to more integrated treatment approaches that may ease the burden of the therapists and maximize the therapeutic gains.

Psychological Factors Affecting Other Medical Conditions (PFAOMC)

According to DSM-5, the A criterion for this condition is the presence of a medical symptom or condition while the B criterion, the essential feature, is the presence of one or more clinically significant psychological or behavioral factors that adversely affect a medical condition by increasing the risk for suffering, death, or disability. Severity ranges from “mild” to “extreme.”

Psychological or behavioral factors include psychological distress, patterns of interpersonal interaction, coping styles, and maladaptive health behaviors such as denial of symptoms or poor adherence to medical recommendations.

Common clinical examples include denial and refusal of treatment for an acute condition (such as myocardial infarct or abdominal emergencies) by individuals with certain personality styles (e.g., domineering or controlling), the exacerbation of asthma or irritable bowel attacks by anxiety, and the manipulation of insulin by an individual with diabetes, or diuretics in the case of hypertensive patients, in efforts to lose weight.

The diagnosis should be only made when the effect of the psychological factor on the medical condition is unambiguous and leads to documentable effects on the course and outcome of the medical condition.

The reverse situation, the psychiatric or psychological consequences of having a medical condition, is more properly classified as adjustment disorder. The prevalence of PFAOMC is unknown. In mental health systems, the diagnosis is most frequently made in consultation-liaison psychiatry settings.

The differential diagnosis can be complex and should include other somatic symptom disorders (somatic symptom disorder, conversion disorder, factitious disorder), as well as personality disorders (e.g., borderline personality disorder).

Treatment. Treatment frequently involves communication with the patient's primary medical team as well as family. The psychoeducational intervention clarifies the role that emotional and behavioral factors play in aggravating the underlying medical condition. Medication to treat another underlying psychiatric disorder may be necessary.

Mr. A, a 55-year-old man, was hospitalized in the intensive care unit (ICU) following a cardiac arrest. He had been experiencing severe substernal chest pain but "ignored it." He became diaphoretic and reported reluctantly to a nearby emergency room, where he had a cardiac arrest while awaiting evaluation. He was successfully resuscitated and transferred to the ICU. He emphatically rejected having had a heart attack, took his EKG leads off, and was preparing to leave the unit against medical advice. The psychiatric consultant obtained a history that the patient's father and brother had died of coronary disease. At the time of the initial assessment, the patient was irritable and somewhat anxious. He agreed to stay in the ICU overnight after his wife insisted that he stay, and the psychiatrist explained the tests further. He was given a low dose of a benzodiazepine, and on the next day, Mr. A was agreeable to remain in the hospital for further treatment.

Factitious Disorder

This category was listed as a separate category in the previous DSM editions.

DSM-5 added it to the somatic symptom and related disorders category because the symptoms these patients present with are most frequently somatic, and a differential diagnosis with other SSDs such as conversion disorder (functional neurologic disorder) is essential.

These patients feign, misrepresent, simulate or cause signs or symptoms of physical illness on themselves. A subtype of factitious disorder is factitious disorder imposed on another (previously known as factitious disorder by proxy).

DSM-5 distinguishes factitious disorder from malingering. According to DSM-5, malingering is differentiated from factitious disorder by the intentional reporting of symptoms for personal gain such as financial compensation or getting time off work. Thus, a distinguishing feature of factitious disorder in DSM-5 is the falsification of signs or symptoms or the induction of disease such as skin lacerations, adding blood to urine, inducing GI bleeding, laxative-induced diarrhea, induced fever, and many others in the absence of obvious external rewards.

Conversion disorder is an important differential diagnosis in patients presenting with neurologic symptoms (fits, convulsions, paralysis, anesthesia, and so on), as unequivocal evidence of voluntary and deceptive fabrication of symptoms may be difficult to obtain.

Historically, simulated physical symptoms have been recorded since the time of the Greeks (Galen's report of induced physical symptoms such as vomiting or rectal bleeding). In recently released Vatican archives, self-induced stigmata, discovered as part of inquisition trials, is an interesting observation. In more recent times, factitious disorder was known as "Munchausen syndrome" following the 1951 publication by Asher in the *Lancet*.

The epidemiology of Factitious Disorder is not known. Etiologic factors are generally not clear, except in cases where religious beliefs or familiarity with the health field play a role in symptom formation. A common denominator nowadays is that these patients tend to be avid medical service seekers, curiously not a criterion included in defining the disorder.

Treatment is basically conservative, aimed at preventing further complications.

Other Specified Somatic Symptom and Related Disorders

Included here are below threshold somatic syndromes, including brief somatic symptom disorder, brief illness anxiety disorder (both with less than 6 months duration of symptoms), illness anxiety disorder without excessive health-related behaviors (criterion D not met), and pseudocyesis.

Unspecified Somatic Symptom and Related Disorders and Other Below-Threshold Somatic Symptom Syndromes

Several investigators responding to the expressed need of clinicians and researchers, particularly in primary care, have proposed below-threshold subtypes that may capture many of the patients often seen in primary care who do not meet the full criteria for somatic symptom disorder.

Definition. Syndromes that fall below the threshold for an SSD may include the old notions of "somatization syndrome," "abridged somatization disorder," and "polymorphous," "multisomatoform," or "polysomatoform" disorder.

History. Because of the difficulty in applying the unwieldy diagnostic criteria of full DSM-IV somatization somatoform disorder in primary care, based on a clinical study, Escobar proposed an abridged construct of somatization that used a threshold of four idiopathic physical symptoms for males and six for females. This construct was shown to have similar discriminating power as the full somatization disorder. The abridged construct was systematized in 1987, incorporating the physical symptoms elicited with the DIS. Using these criteria, Kirmayer and Robbins in 1991 outlined three nonoverlapping syndromes in patients seen in a primary care service. They reported that the majority of patients presenting with medically unexplained somatic symptoms fitted one of the three definitions, each implying a different pathogenic process. In 1993, TK Craig and collaborators in the South London Somatization Study also used the abridged criteria for somatization and proposed four syndromes in primary care: "psychologizers," "somatizers," and "mixed" and "physical" cases. The notions of "psychologizers" and "somatizers" were also used in Spain by A. Lobo and collaborators. Other authors, such as K. Kroenke in the United States and W. Rief in Germany, also proposed abridged constructs of somatization.

Epidemiology. Overall, it has been estimated that in up to two-thirds of patients presenting to primary care, no allopathic disease entity can be differentially diagnosed and unequivocally established as the principal determinant of the presenting somatic complaints. In the United States, the proportion of patients with full DSM-IV somatization disorder and abridged somatoform disorder in primary care settings has been estimated to be 3% and 20%, respectively. Of the 20% or so primary care patients with high levels of idiopathic physical symptoms, most of them (75% or so) meet the criteria for a major psychiatric disorder, lending credence to the statement that idiopathic physical symptoms may be the most typical way for common psychiatric disorders such as depression and anxiety to manifest themselves in primary care. An international study sponsored by the WHO that took place in 14 different countries examined in detail somatic symptoms, as well as various concepts of somatization in primary care, and how these related to depression and other syndromes. The study found that idiopathic physical symptoms defined either as a somatoform disorder or abridged somatoform disorder were extremely common (range 53% to 65%). No major differences were noted across countries when somatization was defined as either denial of psychological symptoms or reporting of medically unexplained somatic symptoms. However, when somatization was defined according to presenting symptoms, the study found greater variation among the centers. An important observation of this study was confirming that reports on lifetime somatic symptoms are unstable, particularly when somatic symptom thresholds are high (somatoform disorder). Lower symptom counts (abridged somatoform disorder) fared better, but even here, half of the symptoms reported at the initial interview were forgotten one year later. Because, according to this research, diagnoses based on lifetime recall appear to rest on an unreliable foundation, the use of current rather than lifetime somatic symptoms should be the rule. Idiopathic symptoms may be simply a reflection of the availability of particular health services in a given region, as suggested by the WHO collaborative study mentioned above. There is also some evidence that there are fewer "somatizers" in settings in which a more personalized type of care is common than in those settings in which it is rare.

Treatment

GENERAL OUTLINE FOR TREATMENT OF IDIOPATHIC PHYSICAL SYMPTOMS IN PRIMARY CARE. The primary care setting has been dubbed the "de facto mental health system," and because this is more readily available across cultures than the more formal mental health system, comparative international studies have often taken place in that setting. Studies in primary care are also of particular importance with regard to somatic presentations because physical symptoms presented at this level can be taken to have clinical importance to the person experiencing them. In the 1920s, Francis W. Peabody, Harvard Professor of Medicine at Boston City Hospital, wrote that teachers and students are at risk of "serious error in their attitude toward a large group of patients who do not show objective, organic pathologic conditions, and who are generally spoken of as having nothing the matter with them." Peabody stated that treatment was the responsibility of the internist and general practitioner and that its effectiveness resided in the quality of a caring patient–doctor relationship.

Although some of these “difficult” patients eventually end up in psychiatric clinics, many remain in primary care. Over the years, the care of these patients has been negatively affected by the lack of an integrated healthcare system. Given the dimensional characteristics of these disorders, having a broad range of symptoms, symptom severity, and functional impairment, the “one size fits all” approach is grossly unrealistic. Therefore, a personalized, multistage approach should be developed. In primary care, patients with idiopathic symptoms often have symptoms of common mental disorders such as depression and anxiety. The initial step should focus on providing symptom relief for the very complaints that prompted patients to seek treatment in the first place. In the case of these patients, these are somatic complaints that frequently express idioms of distress. One would then expect that a positive effect on physical symptoms would make nonpsychologically minded patients more amenable to psychiatric approaches and interventions. Thus, the first steps in this process seem crucial.

- *Step 1:* This starts with detailed scrutiny of medically unexplained physical symptoms that will identify patients in primary care who are likely to be in need of mental health treatment.
- *Step 2:* This entails the use of simple, brief modules, such as psychoeducation, biofeedback, and stress reduction instructional videos. These should target primarily the type and intensity of physical symptoms with which a given patient presents, under the supervision of primary care providers (e.g., nurse practitioners) who have been trained to assess patients and even provide some simple interventions.
- *Step 3:* If symptoms do not improve with the first two steps, the next one entails greater onsite involvement by mental health specialists. It may include expert assessment and referral to a time-limited, CBT-type intervention administered on-site in primary care that focuses on the physical symptoms and has been shown to be beneficial for both physical symptoms and, secondarily, symptoms of depression and anxiety.
- *Step 4:* This and subsequent steps include a menu of options for the continuing care of such patients, including psychotherapy and drug treatments either alone or in combination. These may require referral

to mental health clinics and treatment coordinated by a team of mental health clinicians.

Pivotal Study. A 2007 study by Escobar and coworkers tested the effectiveness of a CBT-type intervention delivered in primary care for patients presenting with medically unexplained physical symptoms. Patients were randomly assigned to receive either the new intervention in addition to a consultation letter or "usual clinical care" plus a consultation letter alone. All treatments and assessments took place at the same primary care clinic where patients sought care. The results showed that a significantly higher proportion of patients in the intervention group had their physical symptoms rated as "much–very much improved" compared to those in the other group (60% vs. 25.8%; odds ratio, 4.1; 95% confidence interval [CI], 1.9 to 8.8; $P < .001$). The intervention on unexplained physical symptoms was best noticed at treatment completion, led to significant relief of symptoms in more than half of the patients, and persisted for up to 6 months after the intervention, although its effectiveness did gradually diminish. The intervention also led to significant improvement in self-reported physical symptoms, self-ratings of physical symptoms, and clinician-rated depression scales at the post-treatment assessment. The authors of the study concluded that a time-limited CBT-type intervention significantly ameliorated presenting physical complaints of patients seen in primary care and offer an alternative for managing these common and problematic presentations in primary care settings.

GENERAL GUIDELINES FOR TREATMENT OF SOMATIC SYMPTOM DISORDERS

Reviews of controlled studies have shown a good efficacy of psychotherapy, particularly of CBT type, and some evidence for the efficacy of drug therapy in the treatment of a number of SSD and functional somatic syndromes. This should dispel the therapeutic nihilism that has surrounded the treatment of these patients for quite some time. Moreover, in a recent update (up to 2006), Kroenke found 34 randomized clinical trials on somatoform disorders, which included more than 1,000 patients. Of these, 13 were CBT trials, 5 were antidepressant trials, and 16 involved other modalities. Overall, CBT showed the highest efficacy, with virtually all reported studies showing a significant effect of CBT on symptoms compared to no treatment. Results were particularly favorable in the case of patients with various functional symptom syndromes and high levels of medically unexplained physical symptoms, as well as for patients with hypochondriasis and body dysmorphic disorder. Antidepressants seem to work in painful syndromes, and the data favor older antidepressants (tricyclics) over new ones (SSRIs). Other interventions, including CBT, the use of a psychiatric consultation letter, brief psychodynamic therapy, exercise, and biofeedback, have also shown beneficial effects on some of these syndromes. A word of caution should be added, however, to temper the enthusiasm on the efficacy of CBT. Even though the pivotal CBT trials described here were randomized clinical trials, in most instances, the comparison group was restricted to a “medical care as usual” group or to the simple addition of the consultation letter. Thus, it cannot be ruled out that the better outcomes in the CBT-treated patients may be at least partly related to the increased attention and the additional time and effort invested in the case of the experimental (CBT) treatment. Therefore, future trials should rigorously control for this factor.

THE FUTURE OF SOMATIC SYMPTOM DISORDERS—TOWARD A NEW DIAGNOSTIC CLASSIFICATION?

Psychological causation and modulation, as well as behavioral and attitudinal aspects, of the somatic symptom syndromes, have been neglected in past diagnostic systems, possibly because they are difficult to define operationally. In an influential book, McHugh and Slavney stated that a somatic presentation (“hysteria” in this instance) “is *not something the patient has; it is something the patient does; that is, it is a behavior.*” To some extent, this perspective was incorporated into the new criteria delineated in DSM-5. The tendency to always view certain somatic presentations as the physical expression of a primary mental disorder was evident in criteria such as those proposed by Bridges and Goldberg in England, which emphasized the need for a proper psychiatric diagnosis made by a competent clinician to confirm that the somatic presentations are or are not connected to a psychiatric disorder.

Other interesting models to dissect ambiguous somatic symptom syndromes have been proposed. For example, Richard S. Brown separates somatic presentations according to three explanatory themes, namely “dissociation, conversion, and somatization.” Harold Merskey and Francois Mai distinguish two types of somatic symptom presentations. The first comprises symptoms such as unexplained pain and “functional” cardiovascular or GI symptoms, which are involuntary, automatic, related to autonomic arousal, and appear in temporal relationship to the experience of “stress” or jointly with depression/anxiety syndromes. This type seems particularly common in primary care. The other and more ambiguous set depends on thoughts, embodied intentions, or ideas that are not recognized or acknowledged, expressed as unexplained neurologic symptoms. This subtype may represent the more severe, polymorphous, treatment-refractory syndromes that are more likely to be seen in specialty mental health environments but are also seen in medical and surgical tertiary care centers. These models make sense intuitively, but they have not been assessed in systematic studies. A European study showed that the expression of physical and psychological complaints is ubiquitous, starting at an early age and that they seem to follow separate paths regarding their expression, evolution, and recognition.

NIMH Research Domain Criteria (RDoC)

This alternative diagnostic approach, developed by Bruce Cuthbert and Thomas Insel at NIMH, relies on behavioral neuroscience, brain “circuits,” “cells,” and “self-report scales.” RDoC is a dimensional approach for use in research, and its main goal is to define clinical characteristics that may reflect the biology of the illness with greater precision, hopefully leading to more specific therapeutic targets compared to the broad overlapping categories of the DSM. While not yet proposed by the RDoC workgroup, it may be useful to consider dimensional approaches as a new research strategy to examine SSD.

Instruments to Assess Somatic Symptom Disorders

The present state examination (PSE), a phenomenologically rich semistructured instrument that requires clinical expertise and significant training and includes excellent symptomatic descriptions, maybe a useful tool for assessing DSM-5 categories of somatic symptom and related disorders in both clinical and epidemiologic studies.

There are numerous widely used scales for assessing somatic complaints, but the diagnosis of SSD requires more than just chronic distressing and impairing somatic complaints. Instead, it requires assessments of excessive dysfunctional cognitions (catastrophizing) and behaviors. A number of inventories such as the Whitley scale or the pain catastrophizing scale may be helpful, but the anchors and cut points need to be derived in careful research lest the diagnosis is overapplied.

FURTHER READINGS

Adams C, Anderson J, Madva EN, LaFrance WC Jr, Perez DL. You've made the diagnosis of functional neurological disorder: now what? *Pract Neurol*. 2018;18:323–330.

Allen LA, Woolfolk RL, Escobar JI, Gara M, Hamer RM. Cognitive behavior therapy for somatization disorder: a randomized controlled trial. *Arch Intern Med*. 2006;166:1512–1518.

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Bagayogo IP, Interian A, Escobar JI. Transcultural aspects of somatic symptoms in the context of depressive disorders. *Adv Psychosom Med*. 2013;33:64–74.

Barsky AJ, Ahern DK. Cognitive behavior therapy for hypochondriasis: a randomized controlled trial. *JAMA*. 2004;291:1464–1470.

Berrios G. *The History of Mental Symptoms*. Cambridge University Press; 1996.

Bleuler E. *Lehrbuch der Psychiatrie [Textbook of Psychiatry]*. Springer-Verlag; 1955.

Braus BR, Rummans TA, Lapid MI, et al. Clinicians and cognitive bias: a case of frontotemporal dementia misdiagnosed as conversion disorder. *Am J Psychiatry*. 2019;176:690–693

Briquet P. *Traité de l'hysterie*. Bailliere; 1859.

Burton C, Fink P, Henningsen P, Löwe B, Rief W; EURONET-SOMA Group. Functional somatic disorders: discussion paper for a new common classification for research and clinical use. *BMC Med*. 2020;18(1):34.

Creed FH, Davies I, Jackson J, et al. The epidemiology of multiple somatic symptoms. *J Psychosom Res*. 2012;72(4):311–317.

Dimsdale JE, Creed F, Escobar J, et al. Somatic symptom disorder: an important change in DSM. *J Psychosom Res*. 2013;75(3):223–228.

Dimsdale JE, Levenson J. What's next for somatic symptom disorder? *Am J Psychiatry*. 2013;170:1393–1395.

Escobar JI, Burnam MA, Karno M, Forsythe A, Golding JM. Somatization in the community. *Arch Gen Psychiatry*. 1987;44:713–718.

Escobar JI, Cook B, Chen CN, et al. Whether medically explained or not, three or more concurrent somatic symptoms predict psychopathology and service use in community populations. *J Psychosom Res*. 2010;69:1–8.

Escobar JI, Gara MA, Diaz-Martinez A, et al. Effectiveness of a time-limited, cognitive behavior therapy–type intervention among primary care patients with medically unexplained symptoms. *Ann Fam Med*. 2007;5:328–335.

Escobar JI, Gara M, Silver RC, Waitzkin H, Holman A, Compton W. Somatization disorder in primary care. *Br J Psychiatry*. 1998;173:262–266.

Ey H. *Manuel de Psychiatrie*. Masson; 1965.

Fallon BA, Petkova E, Skritskaya N, et al. A double-masked placebo-controlled study of fluoxetine for hypochondriasis. *J Clin Psychophamacol*. 2008;28:638–645.

Feinstein RE, deGruy FV. Difficult patients: personality disorders and somatoform complaints. In: Rakel RE, ed. *Textbook of Family Medicine*. 8th ed. Elsevier; 2011: Chapter 46.

First MB, Gaebel W, Maj M, et al. An organization- and category-level comparison of diagnostic requirements for mental disorders in ICD-11 and DSM-5. *World Psychiatry*. 2021;20:34–51.

Foucault M. Folie et deraison. In: *Histoire de la folie*. Librarie Plon; 1961.

Grover S, Ghosh A. Somatic symptom and related disorders in Asians and Asian Americans. *Asian J Psychiatr*. 2014;7(1):77–79.

Janssens KAM, Oldehinkel AJ, Verhulst FC, Hunfeld JAM, Ormel J, Rosmalen JGM. Symptom-specific associations between low cortisol responses and functional somatic symptoms, the TRAILS study. *Psychoneuroendocrinology*. 2012;37(3):332–340.

Jaspers K. *General Psychopathology*. Johns Hopkins University Press; 1997.

Katsamanis M, Lehrer PM, Escobar JI, Gara MA, Kotay A, Liu R. Psychophysiologic treatment for patients with medically unexplained symptoms: a randomized controlled trial. *Psychosomatics*. 2011;52(3):218–229.

Kim SM, Hong JS, Min KJ, Han DH. Brain functional connectivity in patients with somatic symptom disorder. *Psychosom Med*. 2019;81(3):313–318.

Kroenke K. Efficacy of treatment for somatoform disorders: a review of randomized controlled trials. *Psychosom Med*. 2007;69:881–888.

Laferton JAC, Stenzel NM, Rief W, Klaus K, Brähler E, Mewes R. Screening for DSM-5 somatic symptom disorder: diagnostic accuracy of self-report measures within a population sample. *Psychosom Med*. 2017;79:974–981.

Lipowski ZJ. Somatization: the concept and its clinical applications. *Am J Psychiatry*. 1988;145:1358–1368.

Lipsitt D, Escobar J. Psychotherapy of somatoform disorders. In: Gabbard GO, Beck JS, Holmes J, eds. *Oxford Textbook of Psychotherapy*. Oxford University Press; 2005:247.

Malas N, Donohue L, Cook RJ, Leber SM, Kullgren KA. Pediatric somatic symptom and related disorders: primary care provider perspectives. *Clin Pediatr (Phila)*. 2018;57(4):377–388.

McHugh PR, Slavney PR. *The Perspectives of Psychiatry*. Johns Hopkins University Press; 1998:223.

Perley MJ, Guze SB. Hysteria: the stability and usefulness of clinical criteria. A quantitative study based on a follow-up period of six to eight years in 39 patients. *N Engl J Med*. 1962;266(9):421–426.

Scarella TM, Boland RJ, Barsky AJ. Illness anxiety disorder: psychopathology, epidemiology, clinical characteristics, and treatment. *Psychosom Med*. 2019;81:398–407.

Schneider K. *Clinical Psychopathology*. Grune & Stratton; 1959.

Sharma N, Daley M, Vente T. Pediatric somatic symptom disorder: an overview. *Psychiatric Annals*. 2017;47(10):507–511.

Sharpe M. Somatic symptoms: beyond “medically unexplained.” *Br J Psychiatry*. 2013;203(5):320–321.

Shorter E. *From the Mind into the Body: The Cultural Origin of Psychosomatic Symptoms*. Free Press; 1994:90–117.

Smith GR, Monson RA, Ray DC. Psychiatric consultation in somatization disorder. A randomized controlled study. *N Engl J Med*. 1986;314:1407–1413.

Stekel W. *Die Sprache des Traumes*. Bergmann; 1911.

Weisman A, Quintner J, Masharawi Y. Amplified pain syndromes: an unsupportable assumption. *JAMA Pediatr*. 2021;175(6):557–558.

Zaroff CM, Davis JM, Chio PH, Madhavan D. Somatic presentations of distress in China. *Aust N Z J Psychiatry*. 2012;46(11):1053–1057.

19

Factitious Disorder

DAVIN K. QUINN, M.D., DORA WANG, M.D., M.A., SETH POWSNER, M.D., AND STUART J. EISENDRATH, M.D.

Patients with factitious disorder simulate, induce, or aggravate illness to receive medical attention, regardless of whether they are ill or not. These patients may inflict painful, deforming, or even life-threatening injury upon themselves, their children, or other dependents. The primary motivation is attention and care as a *patient*. Factitious disorder patients, in contrast to malingerers, lack an obvious external reward, such as avoidance of duties or financial gain.

Factitious means “artificial, false,” from the Latin *facticius,* “made by art.” The art and artifice of patients with factitious disorder often lead to high hospital drama. Clinicians may exclaim in frustration, “He’s not really sick! He’s doing it to himself!” and dismiss or refuse to treat patients with factitious disorder. Countertransference can be a major obstacle to the proper treatment of these patients who are at risk and suffering a psychiatric disturbance.

Factitious disorders can lead to morbidity or even mortality. Therefore, the medical and psychiatric needs of these patients must be taken seriously, even though presenting complaints are by definition falsified.

An operating room technician, the daughter of a physician, repetitively injected herself with *Pseudomonas,* which caused multiple bouts of sepsis, bilateral renal failure, and ultimately her death. Such deaths are not uncommon in severe factitious disorders.

In factitious disorder imposed on another, also known as factitious disorder by proxy, *Munchausen syndrome by proxy*, and “fabricated illness in a child,” the patient seeks medical attention for another individual, most commonly their own preverbal infant. Factitious disorder imposed on another is recognized in all 50 states as child abuse, elder abuse, or abuse of a vulnerable, dependent adult. Clinicians must prioritize the safety of the victim, which may involve notifying government agencies responsible for the protection of these groups.

Factitious disorders imposed on self and imposed on another should not be considered diagnoses of exclusion but should be actively pursued, given the potentially high stakes, especially when dependent abuse is suspected. Clinicians should be mindful of minimizing unnecessary invasive medical procedures and of trying to steer patients toward the psychiatric care they may need in face-saving, nonthreatening ways. Indeed, factitious illness behavior often represents an underlying psychiatric disturbance, such as a personality disorder, mood disorder, or substance use disorder. Ironically, even patients presenting with factitious psychological complaints, such as feigned bereavement, usually suffer from another psychiatric condition for which they are not seeking help.

Limited empirical knowledge is available about the etiology, epidemiology, course, prognosis, and effective treatment of factitious disorders. Most knowledge comes from case reports, and this information is suspect given that these patients, by definition, falsify their histories. Methodologic problems are inherent in studying deceptive patients. Systematic studies of reported factitious disorder cases have inherent limitations, and federally funded investigation is nonexistent.

It is likely that every clinician will cross paths with a patient with a factitious disorder at some point in their career. A survey of the literature on factitious disorders may engender conclusions of therapeutic nihilism, given the few descriptions of successful treatment and the many examples of patients with recurrent factitious behaviors over decades. However, *Munchausen syndrome*, the prototypical factitious disorder and the first to gain wide medical interest, is now understood to be a chronic severe variant that comprises only a small portion of all cases of factitious disorder. Clinicians may do well to remember that deception is a pervasive, normal behavior in human interactions and politics. In medical settings, patients may engage in a wide range of deceptive behaviors for a host of reasons, including as a coping strategy under conditions of stress, as exemplified by a moving first-person description of recovery from factitious disorder published by Bass et al. in 2013. Conceptualizing deception and excessive health care seeking on a spectrum of normal to abnormal behaviors will help clinicians understand that not all factitious illness behavior is as refractory or as chronic as that demonstrated by patients with Munchausen syndrome.

DEFINITION

The main clinical feature of factitious disorder is the falsification of physical or psychological signs or symptoms or the induction, or exaggeration of injury, or disease. It must be associated with identified deception and be evident in the absence of obvious external rewards, although at times, this may be difficult to determine. An individual may present oneself as ill, impaired, or injured, in which case the diagnosis is *factitious disorder imposed on self*. If an individual presents another individual as ill, the diagnosis is *factitious disorder imposed on another*. Table 19–1 summarizes the definitions in the text revision of the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) and the International Classification of Diseases 11th Revision (ICD-11). Factitious disorders are now considered part of the somatic symptom and related disorders category, given that both disorders involve manifestations of predominantly somatic symptoms and are most often encountered in the medical setting. There are also many areas of overlap between these disorders in comorbidities and treatment.

Munchausen syndrome, the eponymous term coined by Richard Asher in his landmark 1951 publication, was known as *chronic factitious disorder with predominantly physical signs and symptoms* in the lexicon of DSM-IV. Some practitioners use the term Munchausen syndrome for a chronic, severe, refractory form of factitious disorder, in which deceptive illness behavior becomes a lifestyle, precluding stable relationships or employment. Munchausen syndrome comprises approximately 10% of all cases of factitious disorder. Constantly seeking medical care and hospitalization, such patients often assume grandiose false identities, sometimes claiming to be royalty, relatives of celebrities, or figures in important historical events. They travel from hospital to hospital, and when they become well known in one city, take their deception on the road to begin the behavior anew elsewhere. Previous terms applied to these patients included *hospital hoboes, hospital addicts,* and *professional patients.* Two distinguishing features of Munchausen syndrome beyond the simulation of disease are *pseudologia fantastica*—the telling of vague, self-aggrandizing, heroic tales often containing a kernel of truth—and *peregrination*—the tendency to travel widely. There is evidence to support a second durable phenotype of factitious disorder imposed on self, called *common factitious disorder,* in which factitious behavior is confined to one locality and a relatively circumscribed set of complaints. The prototypical patient tends to be young, female, socially connected, employed, and working in health care.

Table 19–1.
Factitious Disorder Imposed on Self

	DSM-5	ICD-11
Name	Factitious disorder	Factitious disorder
Duration		
Symptoms	**Factitious disorder imposed on self:** ▸ Act of falsifying symptoms and clinical findings in one's self ▸ Presenting one's self to others as ill ▸ Deceptive behaviors present outside of identifiable external reward or motivation **Factitious disorder imposed on another:** ▸ Act of falsifying symptoms/injury/disease and clinical findings in another individual ▸ Presenting individual in whom symptoms have been falsified to others as being ill ▸ Deceptive behaviors present outside of identifiable external reward or motivation	**Factitious disorder imposed on self:** ▸ Feigning, falsifying, or inducing signs, symptoms, or injuries ▸ Deceptive behaviors present ▸ If pre-existing disorder present, aggravates signs or symptoms or injuries associated ▸ Seeks treatment and presents as ill/impaired **Factitious disorder imposed on another:** ▸ As above, but enacted on another person
# Symptoms needed		
Exclusion (not the result of)	Another mental disorder	Excoriation disorder Malingering
Psychosocial impact		
Symptom specifiers		
Severity specifiers		
Course specifiers	**Single episode** **Recurrent episodes (**2 or more discrete events)	
Comments	*Note: In factitious disorder imposed on another, the perpetrator receives the diagnosis, not the victim.*	

In factitious disorder imposed on another, a person intentionally simulates illness in another individual, often someone who is dependent on the perpetrator for care. In 95% of all reported cases, the perpetrator is a mother feigning or producing illness in her own child. Factitious disorder imposed on another may also be committed against an adult, such as an elder or spouse. Epidemics of hospital deaths have been attributed to medical personnel inducing or exacerbating illnesses in patients. Because factitious disorder imposed on another almost always constitutes child abuse or criminal abuse, the forensic terms “perpetrator” and “victim” are used even in the medical literature. Subsequent to the DSM-5, factitious disorder imposed on another replaces the terms factitious disorder by proxy and Munchausen syndrome by proxy in psychiatric nomenclature. There is an ongoing debate in the pediatric literature about the appropriate terminology for the condition. *Caregiver-fabricated illness in a child* is the term used by the American Academy of Pediatrics Committee on Child Abuse and Neglect, as it focuses attention on the harm caused to a child and less on the motivations of the perpetrator, consistent with the approach to diagnosing other forms of child maltreatment.

HISTORY

In the second century of the Common Era, the Greek-born physician Galen wrote of patients inducing or simulating symptoms such as vomiting or rectal bleeding. The Bible relates accounts of self-inflicted injury. In the Middle Ages, "hysterics" reportedly put leeches in their mouths to simulate hemoptysis and abraded their skin to reproduce skin conditions. However, judging historical accounts through contemporary mindsets is often problematic, and it is difficult to say whether these accounts actually represent factitious disorder.

In 1838, the Scottish military physician Hector Gavin published an essay, "On the Feigned and Factitious Diseases of Soldiers and Seamen, on the Means Used to Simulate or Produce Them, and on the Best Modes of Discovering Impostors." Although most of Gavin's subjects were malingering to escape duty in the high-casualty Napoleonic Wars, he also noted that the motive of some was simply "to excite compassion or interest" and that "some soldiers, indeed, without any ulterior object, seem to experience an unaccountable gratification in deceiving their officers, comrades, and surgeon."

Jean-Marie Charcot, around 1890, used the term *mania operativa activa* to describe a young girl who continually sought surgery for pain in a knee joint until her medical care seeking resulted in a surgeon amputating the leg. No pathology was ever found in the leg. In 1901, the Swiss physician Henri Secretan lent his name to a peculiar syndrome of nonhealing, traumatically induced edema of the dorsum of the hand. G. Reading, in 1980, confirmed that Secretan syndrome is factitiously produced. In 1934, Karl Menninger described "polysurgical addiction."

Interest in factitious disorder increased markedly when the term Munchausen syndrome was coined by Asher in 1951 to describe three patients with false abdominal complaints, all of whom used multiple identities and sought care at multiple hospitals. In prefacing the case reports, he wrote:

> Here is described a common syndrome which most doctors have seen, but about which little has been written. Like the famous Baron von Munchausen, the persons affected have always traveled widely; and their stories, like those attributed to him, are both dramatic and untruthful. Accordingly, the syndrome is respectfully dedicated to the Baron and named after him.

Hieronymus Carl Friedrich Freiherr von Münchhausen (1720–1797) was a nobleman who served in the Russian army in the war against Turkey (Fig. 19–1). After retirement, he entertained friends with embellished stories of his war adventures. He was turned into a literary figure by the author Rudolph Eric Raspe, a friend of the Baron, who fled Germany for England when caught embezzling from a museum. To pay off debts, Raspe published an account of the baron's tales in 1785. The spelling of the Baron's name has evolved over time to its current most commonly used but likely misspelled form, Munchausen.

In 1968, Spiro noted that of the 38 cases of Munchausen syndrome then published, none involved a detailed psychiatric workup, and less than half were evaluated by a psychiatrist. He advocated a greater understanding of these patients and presented the first detailed psychiatric case study, confirming information with collateral sources. He criticized the "Munchausen" label and recommended the less pejorative term *chronic factitious symptomatology.*

Factitious disorder with psychological symptoms was first described by Alan Gelenberg in 1977, who mused that whereas other factitious disorder patients avoided psychiatrists, his patient, a war veteran, gained admission to more than 30 psychiatric hospitals within a few years, usually feigning depression and suicidality under various pseudonyms.

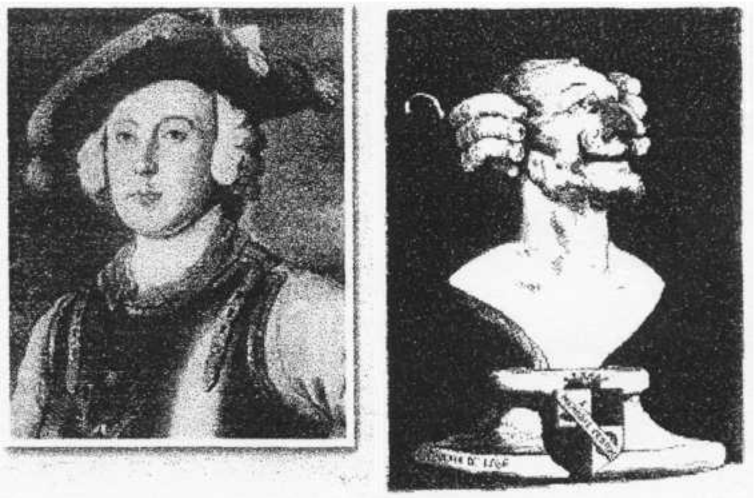

FIGURE 19–1. Hieronymus Carl Friedrich Freiherr von Münchhausen (1720–1797). **Left:** The baron wears military armor in this 1750 portrait by G. Bruckner. A nobleman who served the Russian army in the war against the Turks, the Baron entertained friends with embellished stories of his war adventures in his retirement. His tales gained fame when published by Rudolph E. Raspe. **Right:** The baron appears as a caricature in this drawing by 19th-century artist Gustave Doré. The Latin inscription running diagonally across the shield says “mendace veritas” (ambivalence between lie and truth). Like the Baron, patients with factitious disorders are real persons deserving of respect, even though they often present themselves as caricatures. (Portrait courtesy of Bernhard Wiebel, http://www.Muenchhausen.ch. The actual portrait was lost in World War II. Caricature from Gustave Doré. The Adventures of Baron Munchausen, One Hundred and Sixty Illustrations by Gustave Doré. New York: Pantheon Books; 1944.)

“Munchausen syndrome by proxy” was first described in 1977 by British pediatrician Roy Meadow. He described one mother who caused salt poisoning in her child and another mother whose repeated fabrications of urinary tract infections in her daughter resulted in 12 hospitalizations, more than 150 urine cultures, six examinations under anesthesia, five cystoscopies, and seven major x-ray procedures—all by the age of 6 years. When the girl was admitted for observation, the diagnosis of Munchausen syndrome by proxy was made based on urine samples that were bloody whenever collected by the mother but normal when collected by the nurse. Of note, the mother had also sought medical treatment for factitiously induced urinary tract infections in herself, an early indication of the intergenerational transmission of illness behavior that is often seen in factitious disorders.

COMPARATIVE NOSOLOGY

The DSM-III, in 1980, was the first edition of the DSM to recognize factitious disorder. It focused on Munchausen syndrome, calling it the “prototype” of all factitious disorders. Factitious disorder with psychological symptoms was also recognized, as well as a category called atypical factitious disorder that included what later was termed factitious disorder with predominantly physical signs and symptoms.

Subsequent editions of the DSM increasingly recognized the rarity of Munchausen syndrome and placed more emphasis on a greater spectrum of factitious disorders. The DSM-III-R recognized factitious disorder with physical symptoms and factitious disorder with psychological symptoms. The emphasis on Munchausen syndrome was lessened. Factitious disorder with combined physical and psychological symptoms made its appearance under the category of factitious disorder not otherwise specified (NOS).

In contrast, the DSM-IV defined a single category, factitious disorder, with three types: (1) with predominantly psychological signs and symptoms, (2) with predominantly physical signs and symptoms, and (3) with combined psychological and physical signs and symptoms. Factitious disorder NOS was exemplified by factitious disorder by proxy, a disorder cited as a category deserving of more research, with research criteria listed. The DSM-IV-TR criteria for factitious disorders were unchanged from the DSM-IV criteria.

DSM-5 in 2014 introduced several changes to the diagnostic category. First, it formally recognized factitious disorder by proxy by dividing the general category of factitious disorders into two groups: factitious disorder imposed on self and factitious disorder imposed on another. The diagnosis of factitious disorder NOS was eliminated. Second, there was no distinction made between physical or psychological presentations. Factitious disorders are specified as being a single episode or recurrent episodes. Third, the criteria no longer required that illness induction be conscious or intentional, recognizing the difficulty inherent in making such a determination. Instead, evidence of deception in the absence of clear material gain (also known as "secondary gain") is sufficient to establish factitiousness. If secondary gain is present, such as financial remuneration, relief from duties or legal responsibility, or receipt of controlled substances or narcotics, the deceptive behavior is considered malingering, which is not considered a psychiatric disorder.

The ICD-11 is closely aligned with DSM-5. It lists factitious disorder under the category of "mental, behavioral, or neurodevelopmental disorders" and has subtypes of factitious disorder imposed on self (6D50), factitious disorder imposed on another (6D51), and factitious disorder, unspecified (6D5Z). Falsified illness manifestations are also listed separately under specific medical disorders, such as factitious hypoglycemia, urticaria, cheilitis, purpura, panniculitis, and hyperthyroidism. ICD-11 no longer includes "compensation neurosis" as a related disorder, which in ICD-10 was defined as the exaggeration of physical symptoms that are due to a confirmed physical disorder, caused by the psychological state of the individual in which distress and anxiety about prolonged or progressive disability feature prominently.

ICD-11 aligns with DSM-5 in its distinction between factitious disorder from malingering (QC30) as depending on the absence of external incentives or rewards.

In reality, intent, motive, and awareness are difficult to discern for a clinician with only brief or intermittent interactions with a patient, and each of these occurs on a spectrum. Although the DSM-5 and the ICD-10 and 11 require exclusion of malingering to diagnose a factitious disorder, these two conditions easily and frequently coexist: For example, patients who habitually gratify themselves in the sick role may discover that they can also receive disability payments concurrently or patients who initially fabricated illness to obtain narcotics for intoxication may continue to deceive even when such medications are no longer forthcoming.

Similarly, factitious disorders can coexist, particularly over time, with unconsciously produced illness behaviors such as functional neurologic symptom disorder, somatic symptom disorder, and dissociative disorders. This longitudinal shifting in the production and motivation of somatic symptoms and the difficulty in determining the exact contributors to any one behavior led to the categorization of factitious disorders as a type of somatic symptom disorder in DSM-5. Seen from the perspective of somatic symptoms as an expression of emotional distress, factitious disorder patients and somatic symptom disorder patients are both "engagement seeking."

The evolution of factitious disorder nosology continues along several schools of thought. One perspective has argued for the greater incorporation of performance-validity and symptom-validity testing into diagnostic criteria. Another perspective put forth by Bass and colleagues argues that a person's psychological and social milieu is a greater contributor to most factitious presentations than any specific psychiatric diagnosis. Third, neuroimaging studies on deception indicate the activation of distinct networks in individuals instructed to feign symptoms during functional magnetic resonance imaging (fMRI), hinting at the possible future use of such techniques to refine the diagnostic process.

Despite these theoretical controversies, factitious disorder imposed on another, still known as Munchausen syndrome by proxy in the legal literature, is well recognized in the legal system. All 50 states require the reporting of factitious disorder imposed on a child to protective services as a form of child abuse. Adults who, by virtue of a medical or psychiatric condition, are mentally or physically disabled or reliant on another individual for assistance with activities of daily living are also at risk for becoming victims of abuse via factitious disorder imposed on another. With recognition in the lay press and medical literature of adult dependents as potential victims of an individual with factitious disorder imposed on another, clinicians should be aware that required reporting also extends to older adult patients and to the developmentally delayed.

Neither the DSM-5 nor the ICD-11 makes mention of variants of factitious disorder imposed on another, in which individuals impose their psychiatric disorders or emotional needs upon their children, dependents, spouses, or even entire communities. These phenomena can range from the mundane (*achievement by proxy*) to the mildly disruptive (*hypochondriasis by proxy* resulting in repetitive pediatric visits and unnecessary tests without fabrication or induction of illness) to the psychologically damaging (*gaslighting*) and the physically harmful (*anorexia nervosa by proxy* and *malingering by proxy*).

EPIDEMIOLOGY

Factitious Disorder Imposed on Self

The prevalence of factitious disorder is unknown, in part because of the role of deception in this population, but is estimated to be approximately 1% of the health care–seeking population. Estimates of prevalence are invariably skewed by methodologic limitations of the studies, such as the type of data studied (physician surveys, demographic databases), the study population and setting (hospital ward, outpatient subspecialist, disability office), small sample sizes, recall bias, and publication bias. Studies of specialists and disease-specific tests appear to overestimate prevalence, likely because factitious disorder patients are more likely to see subspecialists and undergo extensive testing, whereas large database studies and systematic reviews are likely to underestimate prevalence due to lack of clinician awareness. There is a large population of people who never present to medical care but engage in disease-simulating behavior, such as those who play the sick role with family, friends, coworkers, social service agencies, online communities, or veterinarians and thus would not be included in medical epidemiologic estimates. Online support groups, social media platforms, and the internet create ripe conditions for factitious disorder patients to garner attention, owing to the ease of creating false identities, lack of face-to-face interaction, and the ability to broadcast to large sympathetic audiences.

In Sutherland and Rodin's 1990 study of patients referred to a psychiatric consultation service in a general hospital, 10 out of 1,288 patients met the criteria for factitious disorder, resulting in a prevalence of 0.8%. A review by Kocalevent et al. in 2005 of 18 studies reporting a prevalence of factitious disorder found estimates varied from a minimum of 0.032% to a maximum of 9.26%, with a weighted mean of 0.9%. In a 2012 occupational medicine clinic sample of 400 patients tested for maximum grip strength, 32 were found to be deceptive in their performance, but only 3 were diagnosed with factitious disorder (0.75%). Fliege et al., in 2007, surveyed 109 senior physicians from internal medicine, surgery, neurology, and dermatology for the estimated prevalence of ICD-10 factitious disorders in their yearly practice. The mean estimated prevalence was 1.3%, the modal prevalence was 1%, and individual estimates ranged from 0.01% to 15%. In 2010, Swanson et al. surveyed 213 physicians regarding their estimates of rates of patients with medically unexplained symptoms. Out of all 4,159 patients seen in 1 day by these physicians, only 1.5% of patients were diagnosed with either a somatoform or factitious disorder diagnosis, with the majority likely somatoform disorders.

Psychiatric settings are somewhat less likely to engage factitious disorder patients: Bhugra et al. estimated psychiatric Munchausen syndrome at 0.5% of adult psychiatric admissions younger than the age of 65 years, based on four diagnoses of Munchausen syndrome out of 775 admissions. In a study of 4,500 visits to a psychiatric emergency room, only 7, or 0.15%, represented factitious disorder. Similarly, a 1994 study estimated the prevalence at 0.14% in a community sample. In a reflection of the overlap between somatic symptoms and factitiousness, Fliege prospectively followed 194 patients admitted to a psychosomatic unit and found 9.7% (19) manifested factitious illness behavior.

Database studies yield significantly lower rates of factitious disorder. In 2013, the DSM-5 working group on factitious and somatoform disorders examined data from the National Hospital Discharge Survey to determine the assignment rate of various psychiatric disorders. Out of more than 3.8 million discharges tracked, a factitious disorder diagnosis was assigned in approximately 9 cases per 100,000, far below the expected prevalence estimate. It was theorized that making a factitious disorder diagnosis was likely unpalatable to physicians and could negatively affect insurance reimbursements. However, all psychiatric disorders, including depression and generalized anxiety, were assigned at rates below their expected prevalence, indicating that a factitious disorder diagnosis is not necessarily less acceptable than any other psychiatric diagnosis. In a study of 28 million military health care visits from 2005 to 2011, Lande and Williams found that only a minuscule number (89 patients) were diagnosed with factitious disorder. However, it is worth noting that in the military, malingering and self-harm are prosecutable offenses.

Profiles of patients with factitious disorder have increased in the last 5 years. Krahn et al. retrospectively reviewed 93 cases of factitious disorder diagnosed over a span of 21 years in a general hospital. Seventy-two percent were female, with a mean age of 31. Two-thirds of the female factitious disorder patients possessed health care training or jobs. Follow-up and treatment engagement were sparse in this group, and only 17% ever acknowledged simulating illness. Reich et al. found a similar profile in their series of 41 cases over 10 years: predominantly female, unmarried, with an average age of 33 years, working in medical jobs, and manifesting emotional and sexual inhibition. Yates, in 2016, performed a comprehensive systematic review of 455 adult factitious disorder cases published between 1965 and 2015.

Of these, 33% involved males, the mean age was 34.2 years, and in 122 cases, the patient had a health care or laboratory job, most commonly nursing. Of the cases with a history of health care work, 114 of these were female, confirming this particular patient profile known as “common factitious disorder” or “nonperegrinating factitious disorder” by some groups. In contrast, Munchausen syndrome features peregrination or frequent wide-ranging travel to seek medical care and *pseudologia fantastica.* In the same series described above, the male patients were predominantly middle aged, unemployed, and unmarried, with antisocial traits or histories, and 61% engaged in peregrination.

Factitious Disorder Imposed on Another

Factitious disorder imposed on another is most commonly perpetrated by mothers against infants or young children. The estimated incidence is approximately 0.5 to 2.0 cases per 100,000 children under age 16. In the United States, it accounts for less than 0.04%, or 1,000 of 3 million cases of child abuse, although it may well be underrecognized. Of infants brought to an Australian clinic for apparent life-threatening episodes, an estimated 1.5% represented factitious disorder imposed on another. A clinical practice at Great Ormond Street Hospital in England reported that in 20 years, 43 children from 37 families were diagnosed as having induced illnesses.

Rosenberg’s review of 117 cases of factitious disorder imposed on another revealed valuable information about the typical perpetrators and victims. Male and female children were equally victimized. The mean age of a child at diagnosis was 3.3 years, with the onset of symptoms occurring at a mean of 1.24 years earlier. All perpetrators were mothers, with 98% being biologic mothers and 2% adoptive. Paternal collusion was suspected in only 1.5% of the cases. In half of the cases, illness was actively produced in the child, whereas in 25%, illness was simulated without direct infliction on the child. In 25% of the cases, illness was both simulated and produced.

In the same review, 10% of perpetrators were thought to have Munchausen syndrome themselves, and an additional 14% showed features of the syndrome. Data collected from the Great Ormand Street Hospital clinic over a 20-year period indicated that about one-third of perpetrating mothers also had a history of factitious disorder imposed on self. Nearly half of them described serious marital problems. About half of the perpetrating caregivers had a history of psychiatric symptoms, and one-third gave histories of emotional neglect or physical abuse. Fathers were generally absent or peripheral.

Bass and Jones examined 28 individuals referred for the treatment for factitious disorder imposed on another. Fifty-seven percent had evidence of a DSM-IV somatoform disorder, and 64% had a history of a factitious disorder imposed on self. Thirty-two percent manifested pseudoseizures and 61% engaged in pseudologia fantastica. Early abuse experiences were seen to lead to pathologic lying and finally to the fabrication of illness in the victim.

ETIOLOGY

Factitious Disorder Imposed on Self

The etiology of factitious disorder is unknown, and a variety of causes likely explain the wide spectrum of factitious illness behavior. The underlying motivations for the behaviors are believed to be unconscious, although this aspect of the disease is becoming increasingly irrelevant for diagnosis. Two factors underlie most cases of factitious disorder: (1) an affinity for the medical system and (2) poor, maladaptive coping skills. A majority of factitious disorder patients have medical training, such as in nursing. In a review of six case series comprising a total of 165 patients, Reich and Gottfried found that 60% worked in a medical profession. This approximate ratio of two-thirds having medical background has been confirmed in systematic reviews by Yates et al. and Caselli et al., as well as single-site case series by Ximenez et al. in 2020. The subjects seemed motivated toward their career choices by genuine lifelong preoccupations with health. Indeed, many with factitious disorder see health care providers as allies, not adversaries. In this case series, many patients had personal ties with their physicians, having worked in their offices or having babysat for them. Having a medical background also permits a patient to engage in more sophisticated fabrications of illness and to conceal their deceptions more skillfully.

Coping deficits are widely noted. Many factitious disorder patients seem to come from large families or to have been neglected as children. They may have grown up without consistent nurturing conducive to the development of mature coping. The poor coping skills of these patients are often symptomatic of an Axis II disorder, such as borderline, narcissistic, dependent, or antisocial personality disorder. Increasingly, factitious disorder is being placed within the broader concepts of “illness behavior” and the “sick role,” which include conditionally acceptable and socially advantageous behaviors that persons engage in as part of a therapeutic contract between themselves and the health care system (“the doctor tries to heal the patient; the patient tries to get better”). Physical diseases are more readily accepted as entries into the sick role compared to emotional or psychological diseases; thus, factitiousness is seen as a type of adaptive behavior to enter into and maintain a therapeutic contract that confers psychological sustenance, while secretly undermining it in order to prolong it.

Factitious disorder may be the presenting symptom of an Axis I disorder. Many case reports indicate a lessening or alleviation of factitious illness behavior when major depression is treated. Other case reports point to illness anxiety disorder, previously known as hypochondriasis, as an underlying factor. For example, a 27-year-old physician simulated insulinoma by injecting himself with insulin. When insulin and a syringe were found in a toilet tank, he confessed that he was preoccupied with the thought that he had pancreatic cancer and was trying to provoke further investigation. A 15-year-old boy who was convinced he had a lesion in his urogenital system simulated hematuria to encourage investigation. These vignettes highlight how factitious behavior may be deployed in the service of getting the attention of medical professionals to focus on a perceived concern. Substance abuse, and to a lesser degree psychotic disorders, and developmental disorders have also been implicated in factitious illness behavior.

No genetic or familial inheritance pattern has been shown. Intergenerational transmission of susceptibility to factitious disorder is possible, as cases and reviews highlight how families in which a child of a parent with factitious disorder goes on to develop factitious disorder as an adult.

While intact neurologic and cognitive functioning is observed in most factitious disorder patients, a neurobiologic basis of some factitious behaviors is suggested by reports of abnormal brain scans and deficits in neuropsychological testing in several studies. Pankratz and Lezak, in 1988, reviewed 25 cases of factitious disorder in which neuropsychological testing was obtained. Approximately one-third of those patients manifested significant neuropsychological deficits. Lawrie et al. in 1993 described a case of a 34-year-old woman who developed personality change and a full-blown Munchausen syndrome after a cholecystectomy. Electroencephalogram demonstrated focal slowing in the right temporal lobe, and testing revealed attentional and memory deficits with a significant visuospatial constructional apraxia. Single-photon emission computed tomography (SPECT) demonstrated reduced uptake in bilateral frontal and left temporal cortices. Mountz et al. in 1995 reported the results of a SPECT scan in a woman with severe Munchausen syndrome. Regional cerebral blood flow in the right hemithalamus was markedly increased, in line with a theory of right hemisphere abnormalities in factitious disorder put forth by Pankratz.

There is one case report of fMRI performed by Spence et al. in a woman convicted of child abuse as a result of factitious disorder imposed on another. She was scanned both while endorsing her version of events in which she is innocent and while she endorsed the account of her accusers. Longer response times and greater activation of ventrolateral prefrontal areas were observed on repeated scans when she endorsed her accusers’ version of events, indicating a change in brain state depending on “truthful” versus “false” endorsement. While this, of course, cannot be considered a biomarker for innocence or guilt, this finding confirms what has been found in studies using neuroimaging to study simulation.

Zheltyakova found regions such as the bilateral temporoparietal junction, right superior temporal sulcus, and left precuneus to have increased blood oxygen level–dependent signal during actions with deceptive intentions, Different and more complex patterns of activation were observed by Stone et al. in bilateral frontal and subcortical areas when patients with conversion weakness attempt to move the paretic limb versus when healthy controls are asked to simulate a weak limb. This line of research suggests that while clinical assessment and treatment of somatic symptom disorders and factitious disorders are converging, deception may involve different brain functions than somatic symptom disorders.

Psychodynamic theories have focused on the concepts of mastery, masochism, and mothering. Striving for mastery may especially apply to factitious disorder patients with predominantly psychological signs and symptoms. For example, patients with factitious psychosis often progress to develop a genuine psychotic disorder, which suggests that the feigning of psychosis may represent a defense or a way of feeling in control of initial psychotic symptoms. Because many patients with factitious physical symptoms seem to have suffered traumatic illnesses as children, adult factitious illness behavior may represent an attempt to master and to feel in control of situations in ways in which they never did as children. They demand or refuse procedures and leave the hospital against medical advice when they feel they are losing control.

Masochism may be a reason that patients repetitively endure painful or deforming surgeries and procedures, such as amputations of limbs and fingers or exploratory abdominal surgeries that result in extensive scarring. The patient relives childhood physical or emotional abuse at the hands of the medical staff in a repetition compulsion. The physician and the medical system at large become symbolic parents against whom the patient re-enacts dependency, idealization, and anger. The system responds with caring, but also with ostensible physical (e.g., surgery) and emotional (e.g., caregiver anger) abuse and, too often, ultimately with rejection and abandonment. Indeed, for those with chronic factitious disorder, the medical system becomes the main object relation.

Factitious illness behavior often occurs when a patient experiences a loss, such as the death of a relative or an occupational loss. Securing the attention of medical clinicians, family, and friends may be a way of obtaining emotional solace without directly confronting the loss. Dependency and narcissistic needs are fulfilled. Factitious illness behavior may also serve the function of providing some patients a role, identity, or purpose. In the role of a patient, they are masterful orchestrators of medical drama. In practicing *pseudologica fantastica,* they often construct desirable and interesting identities.

Behavioral theories postulate that early in life, these patients received positive reinforcement when sick and perhaps gained nurturing from the medical community that they did not receive at home. Perhaps they learned to see the medical system as a source of caring and emotional support. Alternatively, many of these patients lived in large, neglectful households and became the center of focus only when ill.

Factitious Disorder Imposed on Another

Psychodynamic theories predominately explain factitious disorder imposed on another as objectification of the child to serve the parent’s psychological needs. Schreier and Libow called factitious disorder imposed on another a *perversion of mothering* in which a child is dehumanized by the mother and instead serves as a fetishized object through which the mother’s dependency needs are met. Eminson and Postlethwaite argued that two axes are at play: the desire to consult and the inability of the parent to distinguish parental needs from the child’s needs.

Rosenberg noted a *disorder of empathy* among perpetrating mothers, along with pervasive themes of loneliness and isolation, often under circumstances of uninvolved or absent husbands. The mother seeks a relationship with the physician, who substitutes for the uninvolved husband. The relationship with the physician is often ambivalent, characterized by idealization and belittling. The more severe the child’s alleged illness, the more the mother fulfills her own need for caretaking, both vicariously through the child and more directly in her relationship with the physician.

A common view is that those who impose factitious disorder on another have suffered a loss, such as spousal abandonment or loss of contact with another child. Whether perpetrating mothers tend to have suffered abuse is controversial because many such reports of abuse have subsequently been disqualified by collateral sources. Four factors that suggest a caregiver at risk of fabricating illness in a child include (1) appearing to need or thrive on attention from physicians; (2) insisting the child cannot cope without the caregiver's ongoing attention; (3) being directly involved in a medical profession or having expertise or familiarity with medical knowledge and terminology; and (4) having a history of factitious or somatic symptom disorder. It should be noted that these characteristics overlap with those of caregivers who are advocates for children with genuine illnesses and that a caregiver with factitious disorder imposed on another may manifest none of these traits.

DIAGNOSIS AND CLINICAL FEATURES

Table 19–1 summarizes the definitions in the text revision of the DSM-5 and the ICD-11.

Factitious Disorder Imposed on Self

The diagnosis of factitious disorder should be actively pursued and not considered a diagnosis of exclusion. This is recommended for the following reasons: (1) satisfactory exclusion of all other possibilities may be difficult to accomplish or impossible; (2) early diagnosis can minimize the patient's self-harm, as well as harm through iatrogenic complications; (3) confirmation of deceptive behavior requires positive proof that is usually not forthcoming without proactive investigation, and (4) failure to consider the diagnosis is the most common reason for missing a diagnosis of factitious disorder.

Table 19–2 lists reported presentations of factitious disorders along with means used to simulate illness, features suggestive of fabrication, and possible means of detection. The list is a testament to the creativity and resourcefulness of these patients, as well as to their willingness to undergo pain, injury, and inconvenience. Diseases of every organ system have been simulated, limited only by the patient's creativity, knowledge, and access to technologic means. Often, the factitious patient's knowledge of disease exceeds that of the evaluating physician, given the tendency of the patient to be medically trained, as well as the tendency to present to emergency rooms or clinics during evenings or weekends when less experienced resident physicians and staff are on duty. It has been noted by Sinha et al. that factitious patients may emphasize signs and symptoms leading to protocol-driven or "fast-track" admission workups and may be attracted to specialties with more complex disorders and overlapping differential diagnoses.

Factitious symptoms can be (1) *fabricated,* for example, by giving a false history of cancer, acquired immune deficiency syndrome (AIDS), or another illness; (2) *feigned,* for example, by faking symptoms such as pain or seizures; (3) *induced,* by actively producing symptoms through self-infliction of injury or through injection or ingestion; or (4) *aggravated,* such as by manipulating a wound so that it will not heal.

Table 19–2.
Methods of Factitious Symptom Production, Suggestive Signs, and Confirmatory Tests by Systems

Symptom	Method of Factitious Symptom Manufacture	Signs Suggestive of Factitiousness	Test for Factitious Method
Infectious Disease			
Fever	Injecting infectious material into vein/intravenous line	Higher than 41°C	Monitoring while temperature taken
	Ingesting thyroid hormone	Not accompanied by other vital sign abnormalities	Using electronic thermometer reading skin surface
	Drinking hot fluids	Does not follow diurnal pattern	Recording oral and rectal temperatures simultaneous
	Manipulating thermometers with lightbulb or heating pad	No diaphoresis with rapid defervescence	Noting thermometer brand names/serial numbers
	Substituting another thermometer		Measuring fresh urine specim temperature
	Warm wax/wet cotton in ears		
Bacteremia	Injection of contaminated substance	Polymicrobial bacteremia	Culture, whole-genome sequencing to establish reinfection vs. relapse
		Absence of urologic/biliary/GI obstruction	
		Stool flora/pet flora noted	
HIV/AIDS	False history/reports	Normal CD4 count/undetectable viral load/Antibody negative	Repeat HIV ELISA, western b viral load
			Confirm test results with laboratory
Gastrointestinal			
Diarrhea	Laxative abuse (magnesium, castor oil, phenolphthalein)	Metabolic alkalosis	Detection of laxative in urine stool (usually need multiple tests)
		Acute hyperchloremic metabolic acidosis with normal anion gap	3-day stool collection
		Decreased serum bicarbonate with metabolic acidosis	Urine screen for phenolphthaleins, anthraquinones, bisacodyl
		High daily stool volume	Stool screen for magnesium mmol/L, phosphate
		Melanosis coli on sigmoidoscopy	
		Cathartic colon on barium enema	

		Low urine potassium concentration	
		High fecal fluid potassium concentration	
		Fecal fluid osmolality <290 mOsm/kg	
	Adding water to stool sample	High stool osmolar gap >125 mOsmol/kg	
Vomiting	Ipecac/cathartic abuse	Hypokalemia	Detection of ipecac in stool by chromatography
		High urine potassium concentration	Serum/urine emetine levels
		Low urine chloride concentration	EKG with abnormalities c/w ipecac toxicity
		Metabolic alkalosis with increased serum bicarbonate	
Pancreatitis	Spitting into urine sample (salivary amylase)	Hyperamylasuria with normal serum amylase	Obtain monitored sample
GI bleeding	Injection of blood from transfusions	Nasogastric tube shows blood despite normal endoscopy	Radiolabeling transfusions
	Ingestion of NSAIDs/salicylates	Single-stripe sign-on colonoscopy	
Obstruction	Ingestion of loperamide		HPLC of serum/stool for motil slowing agent
Renal			
Diuresis	Diuretic abuse	High urine potassium concentration	Detection of diuretics by chromatography
Bartter syndrome	Loop diuretic abuse	Hypokalemia, urine chloride low or variable	Renal biopsy for juxtaglomeru hyperplasia
	Self-induced vomiting		Screen urine for diuretics
Metabolic			
Hypervitaminosis A	Vitamin A abuse	Increased gamma-glutamyltransferase	Increased serum/tissue levels vitamin A, retinoic acid derivatives
		Increased bilirubin	
Hypokalemia	Laxative abuse	See above	
Hyperkalemia	Injection of urine into blood sample	Potassium level incompatible with life	Monitoring sample gathering yields normal result
Hypernatremia	Salt load	Fractional excretion of sodium high	
		Gastric aspirate salt concentration >200 mmol/L	

Urologic			
Hematuria	Adding blood from another wound/meat	Lack of red blood cell/hemoglobin casts	"3-tube" urinalysis (more bloo first tube if urethral trauma)
	Traumatizing urethra	Lack of distorted red blood cells	Physical examination
	Inserting foreign bodies into the bladder		Radiography
	Ingestion of anticoagulants		
	Addition of coloring to urine		
Proteinuria	Injection of egg protein in bladder	Large day-to-day variations in urine protein concentration	Electrophoresis of urine protein/pure albumin
		Serum albumin concentration remains in normal range	Large albumin band on UPE without transferrin increase
		Lack of other signs of nephrotic syndrome	Antibody confirmation of huma albumin
Bacteriuria	Injecting bacteria into bladder or urine specimens		
Urinary calculi	Adding pepper grains to urine		Infrared spectrophotometry
	Inserting stones into bladder		Chemical analysis
	Submitting quartz, feldspar as samples		X-ray diffraction
			X-ray crystallography
Hematologic			
Anemia	Self-bloodletting	Decreased serum iron, ferritin, iron-binding capacity	Iron 59 elimination studies
		Decreased bone marrow iron concentration	Urine/stool iron levels
		No evidence of bleeding	Blood typing
Sickle cell crisis		Serum protein electrophoresis normal	Genetic testing
			Hemoglobin electrophoresis
Pancytopenia	Ingestion of chemotherapy		
Hemorrhage/purpura	Ingestion of anticoagulant (rodenticide/coumadin)	Prolonged prothrombin time (PT)	Warfarin/brodifacoum/heparin assay
	Heparin injection	Prolonged partial thromboplastin time (PTT) with normal PT	PTT measured every 2 hours under observation

			Reversal with protamine sulfa
			Normal reptilase time
			Failure of PTT to correct in 1: mixture with normal plasma
			Correction of PTT with hepari removal measures
	Ingestion of quinidine	Purpura with thrombocytopenia	Detection of quinidine-depen antiplatelet antibodies
Endocrine			
Hypoglycemia	Injection of insulin	Serum insulin concentration >100 mU/L	Normal proinsulin levels (increased in insulinoma)
	Ingestion of oral hypoglycemic (sulfonylurea)		High insulin/low C-peptide lev
	Ingestion of tolbutamide		Insulin/C-peptide ratio >1.0
	Manipulation of testing strips		Extreme elevations in insulin levels
			Insulin antibodies (less reliabl with human recombinant insulin)
			Failure of glucagon to produc peptide
			Glyburide assay in serum/urin
			Tolbutamide assay in serum
Hyperglycemia	Insufficient administration of insulin		
Hyperthyroidism	Ingestion of thyroid hormone	Goiter/eye findings absent	Serum thyroglobulin low-norm or undetectable
		Increased thyroxine or liothyronine levels	Radioactive iodine uptake low
		Low TSH in thyroid storm	
Cushing syndrome	Injecting/ingesting glucocorticoid	Plasma/urinary cortisol increased	Detection of synthetic glucocorticoids in serum/uri
	Adding glucocorticoid to urine	Plasma ACTH low/undetectable	
		Normal corticosterone/high cortisol	Serum cortisol and corticoster levels
Pheochromocytoma	Injection of epinephrine, isoproterenol	Normal chromogranin A level	Epinephrine/norepinephrine levels in serum/urine
	Ingestion of stimulant	Lack of metabolites of epinephrine (metanephrines)	Metanephrine/normetanephri levels in serum/urine
	Injection of epinephrine into urine sample	Lack of norepinephrine increase after	Provocative (glucagon)/suppression

		glucagon	(clonidine) tests
		Normal response to clonidine suppression test	44-metaiodobenzylguanidine nuclear scan
Hyperaldosteronism	Ingestion of black licorice/glycyrrhizic acid	Treatment-resistant hypokalemia	Serum glycyrrhizic acid level
		Metabolic alkalosis	
		Hypernatremia	
Cardiovascular			
Arrhythmia	Rearranging of EKG leads	Unusual pattern on EKG	Supervise placement of EKG leads
	Ingestion of digitalis, beta-blockers, calcium channel blockers		Serum levels of drugs
			Electrochemiluminescence assays detect BBs in urine
Hypertension	Valsalva maneuver	Office BPs normal, home BPs high	Observe BP checks
	Ingestion of stimulants		Serum/urine assay for pseudoephedrine
Infarction	Beating self with towel to raise creatine kinase	EKGs don’t change	Repeat EKGs
		Enzymes do not rise appropriately	Repeat troponins
Dermatologic			
Cheilitis granulomatosa	Self-inoculation with polyvinylpyrrolidone (PVP)		Liver/lymph node biopsies sho PVP in histiocytes
Dermatitis artefacta	Self-mutilation	Distribution of lesions in reachable areas	Skin biopsy shows mechanica trauma, necrosis, blood
Erythematous lesion	Applying alcohol to lesion		
Nonhealing wounds	Injection of air/contaminants/foreign bodies	Heal when casted	Apply fluorescein to wound, examine hands/nails for fluorescence
		Resistant to treatment	
		Subcutaneous emphysema	
Obstetrical			
Ectopic pregnancy	Injection of hCG	Negative urine b-hCG	Negative ultrasound
		Widely varying b-hCG levels	
Vaginal discharge	Intravaginal insertion	Inconsistent pH	
		Vaginal wall shows abrasion/denies intercourse	

Neurologic			
Movement disorders	Ingestion of neuroleptics to induce parkinsonism	MRI/EEG/EMG inconsistent	Serum levels of neuroleptic
		Hoover sign/tremor entrainable/distractable	
		Prolactin elevated with antipsychotics	
Multiple sclerosis		Examination inconsistent	CSF protein electrophoresis
Pseudoseizures		History inconsistent	Video EEG
		Semiology inconsistent	
		Never observed seizing by others	
Pulmonary			
Asthma	Tape interfering with pulse oximeters	Intractable hypoxia	Lung biopsy
	Inhalation of irritant (talc)	Intractable status asthmaticus	Serum salicylate level
	Ingestion of allergens		
Cystic fibrosis	Manipulation of sweat test	Normal x-rays	Sweat potassium
		Normal examination/lack of clubbing	
Hemoptysis	Ingestion of anticoagulants	Normal bronchoscopy/lack of blood	
	Cough abrasion of lung		
	Self-induced trauma to respiratory tract		
Respiratory failure		Incongruous arterial blood gas	
		Marked hypocapnia	
Rheumatologic			
Arthritis	Insertion of metal fragments in joints		X-rays
			Synovial aspirate analysis
Lupus	Feigned history/borderline positive ANA	Repeat serum autoimmune tests negative	Repeat serum tests
		Normal complement in active flare	
Vasculitis	Injection of contaminants to create purpura		Tissue biopsy of lesions revea foreign material
Ophthalmologic			
Conjunctivitis	Instillation of foreign substance	Purulent discharge purposely left on skin and eyelashes	Corneal examination/conjunct biopsy

		Severity of discharge greater than severity of redness Severity of swelling less than severity of redness Mainly inferior conjunctival involvement Cornea uninvolved
Anisocoria	Instillation of atropine eye drops	

HPLC, high-performance liquid chromatography; UPEP, urine protein electrophoresis; ACTH, adrenocorticotropic hormone; ANA, antinuclear antibody.

Table 19–3 lists clues that should trigger suspicion of factitious disorder. Factitious disorder should be suspected whenever medical signs or symptoms defy conventional medical understanding or when they do not respond to usual medical treatment—for example, when a wound refuses to heal or when test results show a pattern inconsistent with usual disease presentation. Factitious disorder patients may also demonstrate an exceptional eagerness to undergo invasive or extensive testing. They may deny access to collateral sources of information, refuse to sign releases of information, and refuse to give contact information for family, friends, or former physicians. An extensive medical history, evidence of multiple surgeries, and reports of multiple drug allergies are also clues for the astute clinician. These patients often have jobs in the medical profession, have few visitors, and have been known to forecast the progression of their symptoms. They may supply falsified transcripts, test results, medical records, photos, personal documents, and physician reports. There are even cases of patients bringing in pictures and x-rays of pathologic conditions downloaded from the internet and purported to be their own.

Simply considering factitious disorder is the first important step toward diagnosis. Subsequently, information supporting the diagnosis should be collected. A review of past medical records is mandatory and may reveal inconsistencies. Collateral information from family, friends, or other health care providers may likewise show inconsistencies. Ward clerks may be the first to notice waxing and waning levels of distress depending on when the patient thinks clinicians are observing. The collection of laboratory specimens under close observation can minimize contamination or manipulation of samples. Laboratory values can provide diagnostic clues, as when microbiology reports reveal unusual pathogens or polymicrobial findings without apparent cause.

Table 19–3.
Clues That Should Trigger Suspicion of Factitious Disorder

1. The patient has sought treatment at various different hospitals or clinics
2. The patient is an inconsistent, selective, or misleading informant; they resist allowing the treatment team access to outside sources of information
3. The course of the illness is atypical and does not follow the natural history of the presumed disease
4. A remarkable number of tests, consultations, and medical and surgical treatments have been done to little or no avail
5. The magnitude of symptoms consistently exceeds objective pathology, or symptoms have proved to be exaggerated by the patient
6. Some findings are discovered to have been self-induced or at least worsened through self-manipulation
7. The patient might eagerly agree to or request invasive medical procedures or surgery
8. Physical evidence of a factitious cause might be discovered during the course of treatment
9. The patient predicts deteriorations, or there are exacerbations shortly before their scheduled discharge
10. A diagnosis of factitious disorder has been explicitly considered by at least one health care professional
11. The patient is noncompliant with diagnostic or treatment recommendations or is disruptive on the unit
12. Evidence from laboratory or other tests disputes information provided by the patient
13. The patient has a history of work in the health care field
14. The patient engages in gratuitous, self-aggrandizing lying
15. While seeking medical or surgical intervention, the patient opposes psychiatric assessment

Factitious psychological symptoms are more challenging to diagnose because of the lack of clear objective markers for psychiatric symptoms. Methods of confirmation applicable with factitious physical disorders, such as contradictory laboratory results or findings on room searches, are difficult to apply here. Nevertheless, certain features of the patient's presentation can alert the psychiatrist that the patient may be simulating illness. As with physical factitious disorders, the patient may present with unusual symptoms that fail to correspond to any recognizable diagnosis. For example, one patient reported no other psychotic symptoms except seeing the entire cast of a television show emerge from her closet. Other features include worsening of symptoms when the patient is aware of being observed, inconsistencies in the patient's story over time, and the patient's overeagerness to recount symptoms of the illness. Such patients are often suggestible and readily admit to additional symptoms on questioning. They may refuse to cooperate with obtaining collateral information, and untraceable prior health care providers are not unusual. On admission to the ward, patients may reveal familiarity with hospital routine while denying previous hospitalizations. They may exhibit dramatic and unusual reactions to medications. They may demonstrate attention-getting tactics by breaking ward rules. Visitors are usually few or absent. Symptoms may acutely worsen when discharge is imminent.

In contrast to patients with physical factitious disorders, who tend to avoid psychiatric care, patients with factitious psychological symptoms actively seek contact with the psychiatric system and readily acknowledge the presence of a psychiatric disorder, even if it may not be the one from which the patient actually suffers. Feigned bereavement and then psychosis appears to be the most common presenting symptoms. Ironically, one patient sought psychiatric admission for his Munchausen syndrome, claiming that he feigned physical illness when in fact, there was no evidence of this.

In a series of 20 patients who presented with factitious bereavement, 15 also exhibited a history of factitious physical symptoms. A majority of them met the criteria for other psychiatric disorders. They typically reported recent, dramatic, violent, and often multiple deaths of loved ones, whereas collateral information showed that, in fact, no deaths had occurred. Another series of 12 cases of factitious bereavement yielded similar findings of complaints of violent, dramatic deaths and referral from medical wards for supposed physical illness. Although the complaints of bereavement were factitious, these patients exhibited prominent symptoms of depression. A theory about factitious bereavement is that patients are expressing the underlying truth of their emotional state, if not the factual truth. In contrast to the stigma associated with psychiatric illness, sympathy and care are usually offered to those in mourning. This may be why factitious bereavement is the most commonly seen factitious psychological disorder. One patient stated that she fabricated the complaint of bereavement to rationalize her depression.

In a series of 219 consecutive cases of psychosis, 9 (0.04%) met the criteria for factitious disorder. All nine patients demonstrated severe personality disorder. These patients were doing poorly when followed up 4 to 7 years later, with multiple hospitalizations and poor quality of life. Their outcome was no better than for schizophrenics concurrently followed. In another study of six patients who presented with feigned psychosis, five of the six were diagnosed with schizophrenia on follow-up 3 months to 10 years later. Consistent with psychodynamic theories about factitious disorders representing an attempt to feel mastery, feigning psychosis may have been an attempt to feel control over early psychotic symptoms.

In patients who complain of neuropsychiatric or cognitive symptoms, neuropsychological testing can be useful in establishing response bias, unusual patterns of deficits, poor levels of effort, or overreporting of symptoms in a manner that suggests a motivation to appear ill. As in the assessment of malingering, these tests may come in the form of embedded tests or stand-alone symptom validity tests and have been extensively reviewed and approved for this use by the American Academy of Clinical Neuropsychologists. Examples of embedded tests include subscales and indices within the Personality Assessment Inventory (PAI); the Meyers index, which uses seven subscales of the Minnesota Multiphasic Personality Inventory (MMPI); the Fake Bad Scale, drawn from 43 questions on the MMPI that correlate with simulation; reliable digit span tests of memory. Stand-alone tests for simulating symptoms include the Structured Inventory of Malingered Symptomatology (SIMS), the Structured Interview of Reported Symptoms (SIRS), and the Miller Forensic Assessment of Symptoms Test (M-FAST), and the Test of Memory Malingering (TOMM). Important in the use of neuropsychological testing to supplement the diagnosis of factitious disorder is an understanding of its limitations: While a set of given tests can detect a pattern suggestive of simulation, it cannot ascertain motivation or etiology of the pattern. In other words, neuropsychological testing cannot make a diagnosis of factitious disorder—assessment by the clinician is required.

Conclusive confirmation of factitious disorder can be difficult. Covert surveillance, hidden video cameras, or searching the patient’s belongings for syringes or illness-inducing substances can be employed after considering the patient’s right to privacy and the bounds of the patient–doctor relationship. Internet searches and scans of social media may uncover pictures, text, news articles, or legal records that contradict the patient’s history or presentation. Legal counsel is advisable, and in some cases, court orders should be obtained. Consultation from a bioethics team can help weigh the benefits and risks of violations of privacy versus morbidity from factitious disorder. This phase of acquiring information to confirm a diagnosis is often a time of conflict for staff, who may have split opinions about the patient and about the means being employed to confirm the diagnosis. Regular interdisciplinary meetings are helpful.

However, the obtaining of this collateral and confirmatory information to aid in the diagnosis of a factitious disorder brings up several ethical dilemmas:

1. Privacy: Patients have a right to privacy concerning their personal health information and activities, and providers are bound by biomedical ethical principles such as nonmaleficence (doing no harm) and patient autonomy. On the other hand, providers must also consider ethical principles of beneficence (doing what is good for the patient) and justice (doing what is right for society), and patients are expected to participate in their health care in good faith. When the risk of harm from suspected self-induced behavior becomes great enough, decisions to override patient preferences and seek information from covert searches or videotaping may be made, typically in consultation with legal counsel.
2. Confidentiality: Patients with factitious disorder will frequently refuse to release medical records to their providers, hampering the ability to diagnose and manage factitious behaviors. The U.S. Health Insurance Portability and Accountability Act (HIPAA) allows health care providers to release medical records to one another for treatment purposes without an individual patient’s consent, as long as

both health care providers had a relationship with the patient and the records pertain to that relationship. Unauthorized disclosing of information to third parties that are not covered entities under the HIPAA law will need legal vetting to ensure compliance. Searching pharmacy databases and electronic medical records exchanges outside of a health care system represent a powerful modern tool to identify evidence of factitious disorder, and should also be informed by consultation with legal counsel.

3. Stigma: Careful workup for genuine organic causes, judicious use of indicated testing, and application of evidence-based treatment must be performed even in circumstances where a factitious disorder is heavily suspected. Devising tests for factitious illness and obtaining a resistant patient's consent for such tests requires a skillful approach that is nonconfrontational but truthful. There is no role for deceptiveness on the part of the provider in the diagnosis of factitious disorder, such as misrepresenting the reasons for tests or giving placebos disguised as active treatment. The intense negative countertransference on the part of health care staff deceived by patients with factitious disorder is formidable and can result in distortions of usual care, violations of protocols and regulations, and even malpractice. If a patient becomes pejoratively labeled as a "faker" or "somatizer," it is very possible that a condition that has a genuine organic basis may go undetected or untreated.

Munchausen Syndrome

Diagnosis of Munchausen syndrome is often simpler than in other cases of factitious disorder because of the tendency for dramatic, exaggerated presentations. Treatment and management, however, are usually far more challenging. These patients often are new to a particular area, being prone to peregrination. Munchausen syndrome patients may appear eerily comfortable in hospital settings, immediately talking to nurses, physicians, and medical staff as peers. Some show up for hospital admission wearing surgical scrubs. They demonstrate *pseudologia fantastica,* a specific syndrome of autobiographical lying with four features as defined by King and Ford: (1) the stories are not entirely improbable and are built upon a basis of truth; (2) the stories are enduring; (3) the stories are not told purely for personal gain, and have a self-aggrandizing quality; and (4) they are not delusions, in that the patient can admit to falsehoods when confronted with conflicting facts. Munchausen patients often manifest hostility when their needs are not met, or they are confronted by staff regarding suspicions of deception and will then leave the hospital abruptly. They repeatedly present under different names to different hospitals, employing a repertoire of symptoms and presentations to gain admission and treatment. The cost of this disorder can be tremendous, as illustrated by the British report of a "million-dollar man" who, over a 13-year period, spent 1,300 days in psychiatric units, 556 days in prison, and 354 days in medical care for 261 hospital admissions. All diagnostic considerations that apply to factitious disorder also apply here. Because the risk of morbidity and mortality may be higher in these patients, it may be prudent to involve bioethics consultations and hospital legal counsel early.

Factitious Disorder Imposed on Another

Previously called *Munchausen syndrome by proxy,* the essential feature of this disorder is the intentional feigning or production of physical or psychological symptoms in another individual who is under the perpetrator's care. The perpetrator's motive is to assume the sick role by proxy. Mothers of preverbal infants are the most common perpetrators, although fathers, other family members, babysitters, and even medical professionals also have been implicated. Victims can also be spouses, older adult parents, hospital patients, or anyone under the care of a perpetrator. In all cases, this behavior is considered criminal abuse and is mandated to be reported to government authorities.

In an unusual case of adult factitious disorder imposed on another, a 34-year-old man drugged his wife with sleeping pills in her coffee, then injected gasoline into her skin to cause lesions from which she eventually died. Subsequently, he hired a female babysitter to care for his children and repeated his actions with her. In each case, he assumed the role of the concerned caretaker in the center of medical drama. He was finally arrested for murder, and in prison, he sought to work in the infirmary.

Factitious disorder imposed on another has also been cited as the etiology for epidemics of death at hospitals and nursing homes. Perpetrators have largely been nurses and aides who produced illness through various means, such as injection of insulin, lidocaine, digoxin, or other substances. In Great

Britain in 1993, pediatric nurse Beverly Allitt was convicted of killing four children under her care and was given 13 life sentences. Florida registered nurse Bobbie Sue Terrell was accused of a rash of murders at a nursing home where she worked, and in 1988, she plead guilty to reduced charges of second-degree murder. Apparently, after learning of her infertility, she began demonstrating self-injurious behavior and factitious disorder. Kristen Gilbert, a nurse at the Veterans Affairs Medical Center in Northampton, Massachusetts, from 1989 to 1996, was found guilty of having induced cardiac arrest in at least four patients by injecting epinephrine into their intravenous fluids and was suspected of potentially having sickened or killed hundreds of patients. Known as the "Angel of Death" because of the anecdotal increase in deaths during the shifts she worked, Gilbert was reported to have a history of pathologic lying in childhood, including faking a suicide note. During the investigation into the deaths, Gilbert called in a bomb threat to the hospital.

Most reported cases involve mothers and their children. The challenges of recognizing this condition are illustrated by unfortunate examples. Nancy Reagan presented a Mother of the Year award in 1988 to foster mother Yvonne Eldridge for her valorous care of ill children. Eldridge was later recognized as a factitious disorder imposed on another perpetrator and was convicted of child abuse in 1996. Hillary Rodham Clinton publicly recognized Kathy Bush's dedication to her daughter. Bush was later recognized as a perpetrator of factitious disorder imposed on another and was sent to prison in 2002 for child abuse that had resulted in 200 hospitalizations, 40 surgeries, and 15 bouts of sepsis for her daughter.

Sudden infant death syndrome (SIDS) was initially thought to have a strong genetic component because it often occurred in siblings. Today, when siblings die of SIDS, infanticide is suspected. Indeed, a study of 81 children who died of SIDS or natural causes and who later were found to have died at the hands of their parents showed that half of the perpetrating parents had factitious disorder or another somatic symptom disorder. The comorbidity of factitious disorder and factitious disorder by proxy is estimated to be from 10% to 30%. When one child is victimized, there is a high likelihood a sibling will also be victimized, usually serially.

Table 19–4 lists clues that should trigger suspicion of factitious disorder imposed on another.

The variety of medical presentations of factitious disorder imposed on another is impressive. In the first comprehensive review of the disorder, published in 1987, Rosenberg described 68 induced or fabricated signs or symptoms. The most common presentations among 117 cases were bleeding (44%), seizures (42%), central nervous system depression (19%), apnea (15%), diarrhea (11%), vomiting (10%), fever (10%), and rash (9%). Many children had more than one presentation. Twenty-five percent of cases involved simulation of illness, 50% involved production of illness, and in the remaining 25%, simulation and illness production were concurrent.

Table 19–4.
Clues Triggering Suspicion for Factitious Disorder Imposed on Self

1. Diagnosis does not match the objective findings
2. Signs or symptoms are bizarre
3. Caregiver or suspected offender does not express relief or pleasure when told that dependent is improving or that dependent does not have a particular illness
4. Inconsistent histories of symptoms from different observers
5. Caregiver insists on invasive or painful procedures or hospitalizations
6. Caregiver's behavior does not match expressed distress or report of symptoms (e.g., unusually calm)
7. Signs and symptoms begin only in the presence of one caregiver
8. Sibling or another dependent has or had an unusual or unexplained illness or death
9. Sensitivity to multiple environmental substances or medicines
10. Failure of the dependent's illness to respond to its normal treatments or unusual intolerance to those treatments
11. Caregiver publicly solicits sympathy or donations or benefits because of the dependent's rare illness
12. Extensive unusual illness history in the caregiver or caregiver's family; caregiver's history of somatization disorders
13. Caregiver seeks other medical opinions when told the dependent does not have illness
14. Caregiver perseverates about borderline abnormal results of no clinical relevance despite repeated reassurance, or refutes the validity of normal results

Perpetrating caregivers usually appear to be concerned and interested in their children's care. They tend to be exemplary in their interactions with medical staff, enlisting support and sympathy, often crossing professional boundaries by eating meals with staff or helping nurses with duties. They may demonstrate unusual willingness or even excitement at the prospect of invasive procedures for their children. However, analyses of reported cases indicate that factitious behavior continued during hospitalization in 54.4% of cases and that the death of the victim was caused in 7.4% of cases, indicating the potentially fatal consequences if the perpetrator is not stopped in their abuse.

Factitious disorder imposed on another should not be a diagnosis of exclusion, but confirmatory evidence should be actively pursued with the goal of lessening the risk to the victim and assuring the victim's safety. The recommended steps for diagnosing a suspected factitious disorder imposed on another by the American Academy of Pediatrics are outlined in Table 19–5.

The gold standard for confirming factitious disorder imposed on another is covert video surveillance that may record a perpetrator causing harm. Examples of documented harms include caregivers suffocating children, causing fractures, and administering poisons, or injecting harmful substances into intravenous lines. Such videos have also shown mothers appearing concerned in the presence of staff but behaving indifferently toward their children when unaware of being watched.

Despite the benefits of objective documentation of abuse, the use of covert video surveillance is not without risk: (1) damage may be done to the child prior to health care intervention; (2) it may interrupt a potential therapeutic alliance; (3) some consider covert taping to violate the parents' right to privacy or represent entrapment; (4) video footage can be difficult to interpret and may result in false accusations; and (5) covert taping can be costly and require constant real-time monitoring to prevent harm. Covert video should only be undertaken after consultation with legal counsel because a court order may be legally required. A bioethics consultation can help weigh the potential benefits to the child versus compromises of privacy for the parent. An institutional protocol should be in place regarding covert taping, including plans for monitoring and interventions, if the child is found to be at risk.

Table 19–5.
Diagnostic Principles and Tools for Factitious Disorder Imposed on Another

1. Ask: Are the history, signs, and symptoms of disease credible?
2. Ask: Is the child receiving unnecessary and harmful or potentially harmful medical care?
3. Ask: If so, who is instigating the evaluations and treatment?
4. Multidisciplinary evaluation involving medical, psychosocial, child protective, and legal professionals
5. Consultation with a specialist in child abuse pediatrics/elder abuse
6. Complete review of the medical record
7. Construction of a chronologic summary of medical contacts with attention to:
    - use of multiple medical facilities
    - excessive or inappropriate pattern of utilization
    - pattern of missed appointments and discharge against medical advice
    - pattern of misrepresentation of opinions of physicians about the child's medical problems
8. Attribution of all medical information to its source (e.g., Did the nurse observe apnea or did the caretaker inform the nurse of it?)
9. Creation of a table with the following elements for each health contact: Name; date; location; reason for contact; reported symptoms as stated by caregiver; objective signs documented by physician; conclusions/diagnoses made; treatment provided; efficacy of treatment
10. Ask: Does the history provided by the caregiver match the history in the medical record and/or the diagnosis made by the physician?
11. Contact involved physicians and solicit concerns about possible fabrication of illness
12. Have all involved staff meet and develop a consensus management plan
13. Interview the child separately from the caregiver, if verbal, for their recollection of any symptoms, including when and where they occurred
14. Take a careful social and family history, including for any unusual or frequent illnesses in the extended family and siblings
15. Perform diagnostic testing for fabricated illness alongside tests for organic etiologies
16. Separate the child from the suspected caregiver and observe for improvement
17. Employ covert video surveillance while the child is hospitalized, with legal and ethical consultation

Other means of confirming factitious disorder imposed on another include searching a mother's belongings for illness-inducing agents, reviewing collateral information and past medical records for inconsistencies, gathering information on siblings, recording temporal associations between parental visits and a child's signs and symptoms, observing the child's well-being when they are removed from the parent's care, and analyzing specimens taken in the presence of the parent compared to those taken in the parent's absence.

The international literature on factitious disorder imposed on another indicates that signs and symptoms appear to be consistent across the world. Perpetrators are usually mothers, and serial abuse of children is common. This does not appear to be a phenomenon exclusive to medicalized societies.

Pathology and Laboratory Reports

No laboratory or pathology tests are diagnostic of factitious disorders, although they may help to confirm the diagnosis by demonstrating deception. In many cases, a laboratory finding may be the only objective sign that factitious behavior is taking place.

See Table 19–2 for laboratory tests that can be helpful in determining that a given sign or symptom has been artificially induced.

DIFFERENTIAL DIAGNOSIS

The lynchpin of diagnosis of factitious disorder imposed on self or another is the confirmation of deceptive behavior perpetrated by the patient in the service of simulating or exacerbating a disease.

An underlying physical or psychiatric disorder is the main consideration in the differential diagnosis of suspected factitious disorder. Unusual, inconsistent, or poorly characterized symptoms must still be explored for nonfactitious etiologies according to best practices, evidence-based guidelines, and clinical indications.

A 46-year-old woman with type 2 diabetes manifested multiple episodes of hypoglycemia despite several inpatient hospitalizations to optimize her diet and insulin regimen. She was suspected of factitious disorder until it was discovered she had developed antibodies to insulin.

Underlying or prodromal psychiatric/neurologic illnesses may initially be mistaken for psychological factitious disorders. Several case reports document the development of full-blown psychotic disorder following the emergence of factitious illness behavior. Cognitive disorders due to cerebral lesions, neurodegenerative disorders, and delirium may lead to self-induced illness and deceptive behavior. In cases of frontotemporal dementia, the deceptions are usually unsophisticated, and the attempts to obfuscate inept. Neuropsychological and functional testing may help to delineate the cognitive deficits in such cases.

Even when deception is present, genuine comorbid illness is more often the rule than the exception in factitious disorders. Factitious disorder patients frequently aggravate an actual physical illness. Many patients who feign pseudoseizures also have true seizures. Many who manipulate blood sugars to produce symptoms are also diabetic. Munchausen syndrome patients may become addicted to narcotics or learn to seek disability payments and thereby also demonstrate substance abuse and malingering. One area of diagnostic differentiation concerns psychiatric disorders, which can present similarly to factitious disorder with unusual or inconsistent physical or psychological symptoms and may even be comorbid when deceptive behavior is positively confirmed.

In *somatic symptom disorder,* there are one or more physical symptoms that cause disproportionate distress manifesting in thoughts, feelings, or behaviors. If the physical symptom does not have a clear etiology or does not conform to typical clinical phenomenology, suspicions may be raised for factitious behaviors, but positive evidence for deceptiveness will be lacking. In *illness anxiety disorder* (formerly hypochondriasis), normal physiologic sensations are perceived as harbingers of dread disease despite negative tests and reassurance to the contrary. It is important to remember that a presentation that does not conform to an identifiable medical or mental disorder, or the diagnosis of factitious disorder, does not necessarily rule out the presence of a true *medical condition not associated with intentional symptom falsification.*

If a symptom is neurologic and is not consistent with neurologic pathophysiology, then *functional neurologic symptom disorder* (FND) should be diagnosed.

A 35-year-old man presented to emergency care with rapidly progressive lower extremity flaccid paralysis following a mild head injury. Medical workup, including electromyography, could not discern any neurologic disorder. There was no evidence of deception or material gain in this patient, but when a random loud noise startled him, the patient leaped off the examination table quite nimbly and was diagnosed with FND.

The use of video electroencephalographic monitoring allows for the rapid capturing and diagnosis of nonresponsive episodes mimicking absence seizures, complex partial seizures, or syncope.

Dissociative disorders such as *dissociative amnesia* or *dissociative identity disorder* may involve episodes of inability to recall important personal information too extensive to be due to ordinary forgetfulness. These episodes can resemble psychological presentations of factitious disorders in which amnesia or cognitive impairment features prominently. However, they do not involve deceptive behaviors.

The second area of diagnostic differentiation concerns self-inflicted harm. A patient with *borderline personality disorder* may engage in self-destructive, physically injurious, or psychologically damaging behavior resulting in illness (such as overdosing on acetaminophen or cutting one’s arm) for the purposes of stress reduction, pain relief, or conveying distress. If the induction of injury is associated with deception, then factitious disorder may be diagnosed. However, patients who have attempted suicide and caused physical harm may lie about their suicidal ideation, intent, plan, or the aftermath of the attempt. This may be in the service of avoiding detection or hospitalization and carrying out another attempt successfully. This is not factitious disorder nor malingering *per se,* but the expectable behavior of a person bent on suicide and must be screened for with a high level of suspicion.

If a patient's habitual behaviors contribute to worsening health, such as overeating leading to type 2 diabetes, work-related stress exacerbating hypertension, or cigarette smoking contributing to emphysema, and there is no overt deception in the patient's presentation for care, then the appropriate diagnosis is *psychological factors affecting other medical conditions.* This is a new somatic symptom disorder in DSM-5 and is meant to capture psychological states which complicate the course of a physical disease, interfere with its treatment, and represent a well-established health risk. When assessing whether deception is involved, it is important to remember that patients commonly minimize reporting of behaviors that they expect health care providers to criticize or that are not socially, culturally, or legally sanctioned. This should not automatically trigger a diagnosis of factitious disorder without a full exploration of the reasons for minimization. Likewise, ostensibly deceptive behavior may be a result of a lack of cultural understanding on the part of the health care provider, a language barrier between provider and patient, or due to violence or the threat of violence in the patient's life. As Bass and others have recommended, diagnosing a patient's deceptive behavior as a disorder is less important than understanding the underpinnings of the behavior.

A 48-year-old woman with intractable hypertensive urgency endorsed taking her medications faithfully to her primary care physician. However, with intensive case management involvement, it became clear she did not take her medications as prescribed. The team discovered that she was being forced into prostitution by her husband under threat of death, something that she never volunteered to her providers.

Once confirmed, if deceptive behavior is performed to obtain a material gain such as to procure disability payments, narcotics, or an excuse from work, *malingering* is the appropriate label. DSM-5 categorizes malingering as a "condition that may be a focus of clinical attention," defining it as the production of grossly exaggerated or feigned symptoms motivated by an external incentive. DSM-5 lists four conditions under which malingering should be strongly suspected, including medicolegal context, a discrepancy between self-report and medical findings, poor patient cooperation, and antisocial personality disorder; however, these are considered neither sensitive nor specific. Ganser syndrome, also known as "prison psychosis," presents a related challenge. It was first described in 1897, consisting of twilight states, memory disturbances, and *vorbeireden,* a German term loosely translated as "talking at cross purposes." Ganser syndrome is also known as the "syndrome of approximate answers," in which answers to simple questions are erroneous in a manner that implies knowledge of the correct answer. An example would be answering "3" to the question of "How many legs does a horse have?" Ganser syndrome may be due to dissociation, malingering, or organic causes.

In actuality, whether a behavior arises from external or internal origins can be difficult to discern. A clinician may be unaware of a patient's upcoming court date and thus potential gain from being sick. For these reasons, Feldman and coworkers argued that these disorders fall along a continuum: "there is little to be gained, and much to be lost, by the use of the current practice of viewing the somatoform disorders, factitious disorder, and malingering as discrete and distinct clinical entities." Conscious versus unconscious production of symptoms can be especially difficult to discern for psychological signs and symptoms, as there are few objective tests of mental status. It is for this reason that the DSM-5 no longer requires clinicians to determine that patients are explicitly aware of their deceptive behavior, only that there is deception, and not for evident secondary gain.

Mr. J, a 60-year-old man, presented to his local emergency department complaining of nausea, vomiting, and dizziness. He was admitted for possible bleeding peptic ulcer and was also found to have a skin rash and depression. His rash was diagnosed as scabies, but Mr. J refused to shower or change out of his dirty garments. He then requested to be discharged. Psychiatric consultation was requested for evaluation of depression.

Mr. J stated that he was grieving the deaths of his younger daughter and wife in a car accident 2 and a half years ago and that his life had been disintegrating since then. He was staying at friends' houses after leaving his home with his older daughter, as he "couldn't stand being there" any longer because of the grief. He had not sought treatment or professional help, he said, because "I don't want to bother my family or doctors with anything." However, he did endorse having "snapped" and having two "breakdowns" during which he could not remember where he was or who he was, and was hospitalized at a psychiatric unit in another state, but could not recall any details of this incident. He refused to give any contact information for family or friends, citing concerns about bothering them too much and refused treatment for depression.

His social history included employment as a chemical engineer in his own laboratory for the nearby military base, a PhD in parapsychology and forensic psychology, training as a paramedic, a father who was a fighter pilot, and a mother who was a physician. He wistfully recalled his family being stationed overseas. He related that his daughters both held advanced degrees, and his younger daughter was on her way to start a position at a prestigious university hospital when she was killed in the accident.

Inconsistencies between his presentation and his history prompted a search for records in the wider hospital system. Records filed under the same name and date of birth at another facility detailed a significantly different history. When he was approached with these disparate facts, he angrily denied them. He was discharged the next day after it was determined he was not a danger to himself or others.

Two weeks later, Mr. J presented with worsening altered mental status, shortness of breath, abdominal pain, nausea, cough, dysphagia, and vertigo. He was found to have pneumonia, a posterior inferior cerebellar artery (PICA) stroke, a pulmonary embolus, bilateral lower extremity venous thromboses, and esophagitis causing gastrointestinal bleeding. He was treated with antibiotics, anticoagulation, inferior vena cava filter placement, and gastric tube placement. Magnetic resonance brain imaging revealed an old left temporal lobe infarct with encephalomalacia and gliosis, in addition to the PICA infarct.

Psychiatric evaluation after Mr. J's stroke revealed significantly impaired cognition (12/25 on the Montreal Cognitive Assessment) with deficits in short-term memory, attention, and executive function. He could not recall what he was being told by hospital staff about his condition and needed frequent repetition of information. He then revealed that he could not recall any remote memories about his early life, such as schooling, family possessions, or even where he was born. He continued to refuse to speak with family or friends. However, hospital staff had to contact his daughter so that someone could serve as a surrogate decision-maker.

Over several weeks, Mr. J became more attentive and his short-term memory improved. He revealed that, in fact, his wife and daughter had not died in a car accident 2 years ago, that this was a story lifted from the life of his friend "Bill." He could offer no explanation for why he had given this as his life history. He also recounted that he was interested in parapsychology but denied having an advanced degree or practicing in this field. He seemed surprised by the details of his life that were obtained from outside hospital records. Interestingly, he complained that the hospital staff was "keeping secrets" from him. He was transferred to the outside hospital, familiar with his history, for longer-term care.

Three months later, Mr. J presented to the local hospital with a supratherapeutic level of anticoagulation and pain in his lower extremities. It was discovered that when Mr. J had been transferred to the subacute facility 3 months prior, the facility received a telephone call from a hospital in another state, stating that Mr. J was in fact not Mr. J, because the "real" Mr. J was admitted at their facility—family positively confirmed that their relative was the person at the other hospital.

The patient who had suffered scabies, cerebellar stroke, and pulmonary emboli was not the person he said he was, nor was he the person the outside hospital thought he was—he was Mr. R.

When Mr. R was interviewed about his identity, he could not recall any episodic history of his life prior to his hospitalizations, nor could he recall his parents' names, wife's name, whether he had children, where he hailed from, or his employment history. He did not recall his previous deceptive stories. When presented with pictures of famous persons and allusions to famous world events, he denied knowing anyone named Hitler, Marilyn Monroe, or Ronald Reagan. He stated that he recognized Jesus from having read the Bible at the outside hospital and also knew of Martin Luther of the Reformation from reading. His short-term memory, attention, praxis, and executive function seemed very intact. He was able to recall numerous accurate details about his recent hospitalizations. He refused to give details of where he had been living and who had been helping him, always out of a concern to "not bother them."

Mr. R was hospitalized on a psychiatric ward out of concern that his amnesia, whether factitious or organic, was interfering with his ability to manage his medical issues, in particular the anticoagulation. He refused to engage in formal cognitive testing but accepted treatment for depression and aftercare planning and was discharged with appropriate appointments and services in place.

After discharge, Mr. R regularly attended an anticoagulation clinic, physical therapy, wound clinic, and psychiatric appointments. During his mental health appointments, he endorsed that parts of his remote memory were coming back, including how his family lived overseas, his father served in the military, and his mother was a physician. However, he could not call to mind any memory of his family's appearances or names. He later claimed that at his living situation (a group home, but what he termed "someone's garage"), he fixed computers that others brought to him, that others said he was "trained as a paramedic" and that he was "head of an Information Technology department."

DISCUSSION

Mr. R was finally diagnosed with *factitious disorder imposed on self* based on his deceptive behavior (giving a false name and history) in the service of simulating disease (feigned grief and amnesia). Mr. R admitted that he had acted deceptively, making this a conscious, willful act, although he never articulated his motivation. His life history contained fantastical elements, and he tended toward embellishment and confabulation (*pseudologia fantastica*). No clear material gain was ever established.

This case serves as a cautionary tale against assuming that deceptive behavior implies a lack of organic pathology. Mr. R had an ischemic stroke and evidence of prior strokes causing organic cognitive and memory impairment, but he also gave a false identity.

The available information makes it impossible to determine whether Mr. R's previous amnestic episodes during times of stress were real; one may rightly wonder what effect his old temporal lobe stroke had on his memory and whether it played any role in his deceptive behavior.

Factitious Disorder Imposed on Another

When considering the differential diagnosis of this disorder, in addition to the considerations for factitious disorder imposed on self, clinicians should keep in mind the wide range of behavior among normal parents who seek medical care for their children. Again, according to Meadow, "exaggeration and mild deception are part of everyday behavior." The possibility of factitious disorder and malingering by children themselves should be considered, especially in older children and teenagers. Approximately 14% of victims of factitious disorder imposed on another were found to be in collusion with the perpetrator. Other *by-proxy* possibilities should also be considered because children may manifest psychopathologies of their parents. For example, in hypochondriasis by proxy, a hypochondriac mother preoccupied with her child's health may repetitively seek pediatric care leading to unnecessary procedures and iatrogenic illness for the child. In anorexia nervosa by proxy, an anorexic mother may restrict her child's food due to fears of excessive weight in her child. A mother with malingering by proxy may put her child through multiple evaluations to maintain disability or welfare payments. A paranoid father with a history of psychosis feared that his son was being poisoned by breast milk and insisted that the emergency room check his son's hair for mercury. It is important to remember that putting a child through unnecessary tests or treatments which can cause harm is medical child abuse, regardless of the perpetrator's motives.

Other *by-proxy* syndromes described in the literature include *masquerade syndrome* (in which illness fabrication results in the child's increasing dependency on the mother), *mothering to death* (in which the child is confined to a sick role as if the child were ill while avoiding physicians and agencies), *extreme illness exaggeration* (in which a parent exaggerates their child's symptoms in an effort to increase a pediatrician's attention to the child), and *achievement by proxy* (as in youth sports). *Gaslighting* is a nonmedical term used to indicate when a person induces another to doubt their own sanity through psychological manipulation, even to the point of being hospitalized. It may be a form of malingering by proxy when external incentives are present.

COURSE AND PROGNOSIS

Factitious Disorder Imposed on Self

The course of factitious disorder is difficult to determine, as many patients drop out of treatment once a factitious disorder has been raised. The wide spectrum of this disorder should be kept in mind when considering course and prognosis. At one end of the spectrum, factitious illness behavior could be considered within the normal range, as when a patient magnifies his symptoms to seek reassurance from a physician or when a patient lies about completing her antibiotic course for cellulitis in order to maintain positive regard. Further along the spectrum, factitious illness behavior becomes a maladaptive coping mechanism but not a chronic pattern of deception. These less pernicious cases may respond to a clinician's sensitive inquiry into the precipitants to factitious behavior and coaching the patient to find more adaptive solutions. Concurrent mood, anxiety, or substance abuse disorders bode a better prognosis; comorbid personality disorders, especially antisocial personality disorder, bode a poorer prognosis. According to some experts, many factitious disorders remit around the age of 40 years, corresponding to the age of remission for many with borderline personality disorder. Unfortunately, Munchausen syndrome tends to have an unremitting, refractory course and is associated with substantial morbidity and mortality. After efforts to engage a patient in treatment fail, management should be directed toward harm reduction rather than cure.

In Sutherland and Rodin's series of ten patients diagnosed on a psychiatric consultation service with factitious disorder, only one out of nine patients confronted with the diagnoses admitted to factitious behavior, and only one engaged in psychiatric treatment. In Krahn's series, only one in six of confronted patients admitted to factitious behavior, and only 12% agreed to psychiatric treatment. These studies would seem to support the contention that confrontation is unlikely to helpfully engage a patient. However, in Reich and Gottfried's series of factitious disorder patients, of 33 confronted with the diagnosis, the majority improved, including four chronic cases (only 13 ever acknowledged their role). A "supportive confrontation" conceptualization may help to frame the interaction as an attempt to ally with the patient's and provider's common goals when the willingness of the patient to engage in rehabilitation is perceived to be favorable. In more resistant patients, nonconfrontational approaches may permit the patient to save face while maintaining the therapeutic relationship. At times, confrontation is unavoidable, as in cases when "therapeutic discharge" is required to prevent further harm and mischief from occurring when a patient is refusing to leave. It should be remembered that this population is at high risk of self-injurious behavior when attempting to confront factitious behavior, and plans for preventing self-harm should be in place prior to the confrontation.

The course of factitious disorder is usually intermittent: Isolated events and frequently recurring/unremitting episodes are less common. Onset is usually in early adulthood, often after hospitalization for a medical condition or a mental disorder. In individuals with recurrent episodes of falsification or signs and symptoms of illness and/or induction of injury, this pattern of successive deceptive interactions with medical personnel may become lifelong.

Once clinicians recognize a factitious physical disorder, they should be alert to the possibility of patient-initiated litigation to defend the “realness” of their self-induced illness. Litigation often fulfills the same psychodynamic motivations that led to the original factitious illness behavior. Actions brought against clinicians may provide an additional outlet for the patient’s anger, as well as provide an increased level of care or attention. Litigation may be initiated by the patient’s family, who may not understand that the patient’s illness is self-inflicted. Litigation may be a marker for increased risk of death. In a study of 20 factitious disorder patients with concurrent lawsuits, 4 cases ended in death.

In general, psychological presentations of factitious disorder imposed on self have a poor prognosis comparable to that of Munchausen syndrome. A 1982 study of patients with factitious psychosis showed that 4 to 7 years later, these patients functioned at a level comparable to that of schizophrenic patients in the same study. Other studies found that feigned psychosis is commonly a precursor to actual psychosis, leading Hays to comment, “Acting crazy may bode more ill than being crazy.”

Mr. K is a 43-year-old male who has presented repeatedly since age 26 to regional tertiary care hospitals across two states. His initial presentations in emergency departments consisted of repeated rattlesnake, cobra, and viper bites, which he said occurred because of working as a volunteer “snake wrangler.” His work reportedly entailed catching snakes, raising them, feeding them, and giving them medications, but he also reported bites from picking up snakes he found by the road, sometimes while inebriated.

His past psychiatric history was noteworthy for several hospitalizations as a child for suicidal ideation and behavior, which Mr. K attributed to “wanting to get out of the house while [his mother] was menopausal.” He had been given a diagnosis of conduct disorder. Even as a teenager, he was noted to frequently sustain snake bites and to not comply with recommended treatment.

FIRST STAGE

Since his early 30s, Mr. K presented in excess of 30 times for the treatment of snakebite envenomation. He received experimental antibody-fragment antivenin preparations without ill effects so frequently (>150 vials) that his treatment was reported as robust evidence of the safety of the preparation. Complications of his snake bites included anaphylaxis, often requiring intensive care, compartment syndromes in both his arms treated with fasciotomies, and finger and toe amputations. Mr. K denied any intent to be bitten, saying these bites came as a result of his normal snake handling. Yet, he expressed no remorse or desire to change his behaviors. He stated during one admission that snake handling was his only source of income, although he also described performing construction work at other times. Over the course of his hospitalizations for anaphylactic symptoms of snakebite, he began to decline antivenin, withhold history, and request escalating amounts of pain medication. During one admission, he assaulted a health care worker because he did not feel he had been given enough analgesic.

SECOND STAGE

At age 37, he presented to the emergency department with chandelier lightbulbs in his rectum, stating that they were inserted by assailants in the city park. These were removed fully intact, but he presented again 2 weeks later, saying glass shards were retained in his rectum, adamantly denying he had inserted them himself. Proctoscopy revealed glass shards that could not be removed, and colectomy with colostomy was performed.

At age 39, he presented with accidental ethylene glycol ingestion, which he said he mistook for his energy drink while working on an automobile engine with a friend. The next year, he presented having ingested 13 shards of glass and two razor blades, which he told an admitting physician were shoved down his throat by the same assailants who had inserted the lightbulbs in his rectum. He later told another physician that he had been working construction, holding screws in his mouth, and had accidentally ingested them. An x-ray confirmed the presence of several bolts and screws in his stomach.

At age 41, he presented having swallowed fishhooks and screws and inserted a lightbulb into his rectum. This was the first time he admitted that this was self-inflicted. He endorsed depression and anxiety as factors driving him to perform these acts.

THIRD STAGE

In his 42nd year, he had 15 admissions at a local tertiary hospital for intentionally ingesting nail clippers, bottle tops, nails, screws, scissors, tent stakes, and knives and inserting lightbulbs into his rectum. He presented with snake bites, but less frequently than earlier in his history. He admitted that he had thoughts of intentionally getting bitten in order to kill himself. He occasionally complained of auditory command hallucinations telling him to ingest objects, such as a letter opener, but this was infrequent. On most admissions, he was able to say that negative feelings such as depression, loneliness, stress, or anxiety triggered the impulse to harm himself.

During stages two and three, psychiatric consultations were obtained every admission. The consultants attempted to assess without confrontation, to address reinforcing factors such as pain and sedative medications, to identify and address psychiatric symptoms that might prompt Mr. K’s behaviors, and to build his coping skills. Of note, medication trials including fluoxetine, buspirone, lorazepam, aripiprazole, olanzapine, and trazodone were uniformly ineffective, or Mr. K did not take them. He continually asked for benzodiazepines for anxiety and opiates for chronic pain.

FOURTH STAGE

At age 43, from the fall into that winter, he had no presentations to the local tertiary care hospital. Then he presented having swallowed a tent stake, resulting in esophageal tear and pneumothorax treated with thoracotomy and multiple chest tubes. During that admission, he cited factors such as having an apartment, roommates, and friends caring for him, which kept him from ingesting objects for that span of time. It was evident that social support and positive relationships aided in Mr. K's sense of stability, whereas ruptures in these areas led to further self-destructive behaviors. Of note, an interview with one of these friends indicated she thought he was diagnosed with cancer and had been hospitalized with a collapsed lung.

At age 44, during the spring, Mr. K did not present to the tertiary center. In June, he presented after swallowing brush bristles. By this time, he had a caseworker, disability income, and a representative payee. The precipitant for his relapse was eviction for not following his landlord's rules.

Of note, a review of controlled substance prescriptions indicated that at age 43, Mr. K had visited 13 local providers between April and September, including eight dentists, obtained 36 prescriptions for oxycodone, hydrocodone, or alprazolam. None had written for longer than a week's worth of medication, and the prescriptions involved seven separate pharmacies.

DISCUSSION

Mr. K manifested factitious disorder imposed on self. At least initially, there was no identified gain except to play the role of a patient. Later, narcotic analgesics appeared to become a motivator for Mr. K. His eventual presentations with swallowed and inserted objects went far beyond what is typical for narcotic-seeking malingerers. He eventually admitted to fabricating stories, which confirmed the conscious and willful dimension of his behavior. He fit the typical profile of a patient with Munchausen syndrome: a young male with antisocial traits, who demonstrated a prolonged, chronic course with intermittent hostility, travels to multiple institutions and practitioners (peregrination), and deceptive stories that strained one's credulity (*pseudologia fantastica*).

This case exemplifies how a factitious behavior can be a prequel or sequel of several other disorders during a course of illness. What seemed to start as an occupational hazard (bites while handling snakes) became a source of narcotics (malingering). One behavior became a grab bag of self-destructive acts (ingesting and inserting objects), about which Mr. K was untruthful (factitious disorder). Finally, when Mr. K admitted to his self-destructive tendencies, these appeared to be stereotyped emotional responses accompanying poor distress tolerance (personality disorder). Mr. K's presentations became less frequent with interventions for personality-disordered patients: stable living and financial arrangements, a sympathetic support system, and case management with frequent contacts. In addition to the phenomenologic similarities between factitious disorder, malingering, personality disorders, and somatic symptom disorders, the fact that treatment of factitious disorder bears similarity to treatment for somatization and personality disorder may support the perspective that factitious behavior is simply a dimension of other disorders and not a separate entity.

Factitious Disorder Imposed on Another

Victims of factitious disorder imposed on another are at very high risk. Mortality rate estimates of victims range from 6% to 22%, with deaths commonly occurring through suffocation or poisoning. In Rosenberg's 1987 review, 10 of 117 children died, yielding a mortality rate of 9%, with the youngest children being the most vulnerable to death. A sobering statistic is that in 20% of the deaths, caretakers were confronted with the diagnosis of factitious disorder imposed on another, but the children were not removed from their custody. Of the survivors, 8% suffered long-term morbidity, including destructive joint changes, limp, cerebral palsy, cortical blindness, and psychiatric problems. Approximately 75% of the morbidity was precipitated by caregivers' behaviors occurring while the child was hospitalized. In 25% of the cases, morbidity was solely iatrogenic, caused by medical procedures and investigations.

Factitious disorder imposed on another is often recurrent. In a 1998 follow-up study by Davis et al., 17% of victims who returned home without restrictions were abused again. In cases where the child was poisoned or suffocated, repeat abuse rates were as high as 50%, in some cases leading to death. These statistics reinforce the importance of prompt diagnosis, proper intervention, and protection.

Siblings of victims are at great risk. Studies indicate that 9% to 29% of siblings die, underscoring the importance of checking and protecting all children of a given perpetrator. Siblings tend to be the focus serially rather than simultaneously, and so clinicians should always be alert for past and future victims. In 1990, Meadow studied 27 young children suffocated by their mothers. Over 13 years, nine died, and one suffered severe brain damage. These 27 children had 18 siblings who had died “suddenly and unexpectedly in early life.” In another study of 32 children presenting with factitious epilepsy, 21% of siblings (7 of 33) had died of SIDS. In a study of suffocation cases, 21% of siblings (3 of 14) had died unexpectedly. In another study of 56 victims of factitious disorder imposed on another, 39% of siblings were subjected to illness fabrication. Among Rosenberg’s review of 117 cases, 10 siblings had died under “unusual circumstances.”

TREATMENT AND MANAGEMENT

Guidelines for the treatment and management of factitious disorder are given in Table 19–6.

Factitious Disorder Imposed on Self

Three major goals should guide the treatment and management of factitious disorders: (1) to reduce the risk of morbidity and mortality, (2) to address the underlying emotional needs or psychiatric diagnosis that may be driving factitious illness behavior, and (3) to be mindful of legal and ethical issues.

1. Risk reduction: The Hippocratic doctrine of “first do no harm” should be foremost. Prompt recognition of factitious disorder can reduce morbidity or mortality. Active pursuit of the diagnosis and timely management are essential. For example, in the case of a woman who reported an extensive family history of breast cancer and sought prophylactic mastectomy, early suspicion of factitious disorder may have prevented the unnecessary bilateral mastectomies that were performed. Recognition of factitious asthma or arthritis prevents the administration of steroid medications, which pose significant side effects. Invasive, potentially harmful tests, procedures, and treatments can be minimized. Once symptoms are confirmed as factitious, clinicians should administer medical treatment according to their clinical judgment, considering objective evidence and keeping in mind that the patient’s complaints and requests can be deceptive.

Table 19–6.
Guidelines for Management and Treatment of Factitious Disorder

Keep in mind that active pursuit of a prompt diagnosis can minimize the risk of morbidity and mortality.
Minimize harm. Avoid unnecessary tests and procedures, especially if they are invasive. Treat according to clinical judgment, keeping in mind that subjective complaints may be deceptive.
Arrange regular interdisciplinary meetings to reduce conflict and splitting among staff. Manage staff countertransference.
Consider facilitating healing by using the double-bind technique or face-saving behavioral strategies such as self-hypnosis or biofeedback.
Steer the patient toward psychiatric treatment in an empathic, nonconfrontational, face-saving manner. Avoid aggressive direct confrontation.
Treat underlying psychiatric disturbances such as Axis I disorders and Axis II disorders. In psychotherapy, address coping strategies and emotional conflicts.
Appoint a primary care provider as a gatekeeper for all medical and psychiatric treatment.
Consider involving risk management and bioethicists from an early point.
Consider appointing a guardian for medical and psychiatric decisions.
As a behavioral disincentive, consider prosecution for fraud.

Management of countertransference is a priority to reduce risk because a clinician's negative feelings can interfere with appropriate patient care. If possible, psychiatric consultation should be promptly obtained. Good communication among all involved is essential because these patients are prone to cause confusion and splitting of caretakers. Regular interdisciplinary meetings are helpful.

Psychiatrists should be aware of their own countertransference reactions and help medical clinicians cope with their own reactions to these patients who falsify symptoms but who also have genuine needs. Negative countertransference can lead to therapeutic nihilism, breaches of confidentiality, inappropriate treatment, or denial of care. Countertransference feelings of frustration and anger can more constructively be understood as projective identification and, thus, a window toward understanding the patient's longstanding feelings. It is essential to preserve good staff communication and to formulate a clear management plan involving the whole team to minimize splitting. Allowing staff to express negative views and emotions at team meetings while providing empathic support and helpful formulations of the patient can reduce unintended retribution against the patient. The psychiatrist or psychotherapist may also be prone to negative countertransference arising from feelings of being "duped" or manipulated and should seek appropriate supervision.

On an outpatient basis, all medical care should be directed through a single primary care physician through whom all care is coordinated. This *gatekeeper* can minimize the unnecessary repetition of tests and treatments. However, it should be remembered that the patient may seek care from clinicians in another system, in another city, or even in another state or country. Appointments with the gatekeeper primary care physician should be regular and frequent and not dependent on medical crises. This fosters object constancy while minimizing the patient's need to induce illness to seek medical attention. As a means of last resort, legal disincentives might be pursued. In Arizona, a woman with factitious disorder was prosecuted for fraudulent procurement of medical, psychiatric, and dental services: she pled guilty. Laws in several states specifically address this type of action.

2. Addressing psychiatric issues: Allowing the patient to save face is essential toward establishing a therapeutic alliance and toward preventing the patient from simply taking the factitious illness behavior elsewhere. This is especially true because factitious disorder patients usually have immature personalities or personality disorders that make them especially sensitive to narcissistic injury. Direct or aggressive confrontation is generally not effective. The patient usually responds with anger and denial and may leave the hospital against medical advice, only to perpetuate their factitious illness behavior elsewhere. This kind of confrontation may amplify the patient's psychological need for mastery through deception. Clinicians should keep in mind that confession is not a necessary aspect of management or treatment.

Several behavioral strategies have been successfully used to facilitate the healing of factitiously produced symptoms in face-saving ways. Eisendrath advocated a *double-bind* technique whereby patients are told that if the symptoms are genuine, then they should improve with the treatment administered. If the symptoms do not improve, then they must be factitious. This technique can be used with nonhealing wounds, factitious paralysis, or psychological factitious symptoms that should improve with medications. For example, factitious patients who chronically induce or worsen wounds can be informed, "This treatment should heal your wound. If it doesn't, we have no choice but to conclude that this is due to a factitious disorder."

Self-hypnosis and biofeedback can allow a patient to relinquish factitious behavior in a face-saving manner. For example, a patient can be told that under self-hypnosis, blood flow to a wound can be increased and healing promoted. In this manner, the patient can take control of healing the wound rather than seek control by worsening symptoms. Positive feedback should be given to patients when their efforts result in healing. Biofeedback can be used in a similar way.

Once factitious disorder is confirmed, or evidence is deemed sufficient, the patient should be gently and artfully steered toward ongoing psychiatric treatment in a sympathetic manner acceptable to the patient. For example, the patient can be told that their factitious illness behavior is an expression of great emotional need or distress. The clinician can make empathic statements about the neglect, abuse, or trauma that the patient may have undergone. The patient should then be engaged in care in a psychiatric setting in which underlying emotional issues can be addressed and where the risk of morbidity and mortality is substantially less than in a medical setting. Because these patients may crave to be understood and to have their emotional needs recognized, this type of sympathetic confrontation that focuses on addressing the patient's genuine rather than factitious needs can be well accepted by patients.

Psychiatric treatment should first focus on underlying or comorbid psychiatric disorders or on the emotional distress that might have precipitated the factitious illness behavior.

Comorbid mental illnesses must be recognized and treated appropriately. For example, most cases of feigned bereavement are comorbid with major depression. One patient claimed to have feigned bereavement to justify depression. Another falsely claimed to have been recently raped and sought help for post-traumatic stress disorder. Although she was not recently raped, she had suffered longstanding childhood sexual abuse. Feigned mental illnesses may express an emotional truth for the patient, if not factual truth. Comorbid personality disorders are more common than affective or psychotic disorders. Borderline personality disorder is the most common comorbid diagnosis. Antisocial personality traits are also common, especially in patients exhibiting *pseudologica fantastica* and Munchausen syndrome. Pharmacologic and psychotherapeutic treatments should be employed according to the diagnosis. Substance abuse and dependence may require intensive outpatient or residential inpatient treatment. Other than targeting comorbid psychiatric disorders, there is no standard pharmacologic treatment for factitious disorder.

Once in psychiatric treatment, the greatest challenge is getting the patient to engage and commit to long-term therapy. Brief regularly scheduled contact on a time-contingent rather than a distress-contingent basis is the mainstay of therapy. Consistent long-term psychotherapy is aimed at enabling patients to express their feelings, gain insight and coping skills, and provide a reliable and supportive outlet for communication.

Relapses of factitious behavior should be expected, just as with substance abuse. Clinicians should not despair at relapses but instead, take them as opportunities to further understand the patient. For example, if a patient has a pattern of relapsing under conditions of new romantic involvements or arguments with authority figures, this provides insight into the patient's vulnerabilities and may be addressed in psychotherapy.

3. Managing legal and ethical issues: Early involvement of risk management experts and bioethicists is prudent in the care of factitious disorder patients. Patients' right to privacy and implicit injunctions against unwarranted searches and seizures are major issues. The once-common practice of alerting all emergency rooms about factitious disorder patients is no longer routinely utilized because of heightened sensitivity about confidentiality rights. Nevertheless, this kind of widespread alerting can be done with the consent of the patient. For example, once a patient acknowledges having factitious

disorder, the patient can be asked, “I’d like to minimize harm to you. Can I have your permission to alert emergency rooms and some doctors in the community?”

Similarly, the patient’s right to confidentiality should be respected when gathering collateral information in nonemergent situations. Verbal permission or, preferably, signed releases of information should be obtained prior to contacting collateral sources, except in emergency situations. Clinicians should be careful about revealing information to the patient’s employers, friends, or family. The diagnosis of factitious disorder must never be revealed, even to spouses or significant others, without a patient’s explicit permission, as this will likely lead to accusations of a confidentiality breach. Searching a patient’s hospital room or personal belongings for illness-inducing means may facilitate diagnosis and lessen morbidity, but this should only be done after consulting with risk management or other legal counsel because it goes beyond the contractual and fiduciary bounds of the doctor–patient relationship. Covert surveillance should only be undertaken after careful legal consultation. It should also be kept in mind that simply because an action is legal, it is not necessarily ethical. Therefore, the involvement of bioethicists can be helpful.

Factitious Disorder Imposed on Another

Protection of the victim is the first priority. Active pursuit of the diagnosis and then prompt intervention is essential for minimizing the risk of morbidity and mortality to the victim, who is usually a vulnerable child, as over 50% of reported cases described ongoing induction of illness in the hospital setting. Because young children are the most commonly victimized, Table 19–7 will prioritize the management of pediatric factitious disorder imposed on another.

Table 19–7.
Pediatric Factitious Disorder Imposed on Another—Basic Principles of Management

Make sure the child is safe.
Make sure the child’s future safety is also assured.
Allow treatment to occur in the least restrictive setting possible.
A pediatrician should serve as “gatekeeper” for medical care utilization.
All other physicians should coordinate care with the gatekeeper.
Child protective services should be informed whenever a child is harmed.
Family psychotherapy and/or individual psychotherapy should be instituted for the perpetrating parent and the child.
Health insurance companies, school officials, and other nonmedical sources should be asked to report possible medical abuse to the physician gatekeeper. Permission of a parent or of child protective services must first be obtained.
The possibility should be considered of admitting the child to an inpatient or partial hospital setting to facilitate diagnostic monitoring of symptoms and to institute a treatment plan.
The child may require placement in another family. The perpetrating parent may need to be removed from the child through criminal prosecution and incarceration.

When gathering clinical information and evidence, clinicians should keep in mind that if the mother feels that she is suspected or is losing control of the situation, she is likely to remove the child and seek care elsewhere. When evidence of factitious disorder imposed on another is felt to be sufficient, both parents should be informed together. In this manner, a nonperpetrating parent, usually the father, might be enlisted as an ally.

Child protective services should be informed, and a legal hold should be instituted if clinicians feel that parents might flee with the child. If harm has been done to the child, reporting of factitious disorder imposed on another to child abuse protection authorities is mandatory in all 50 states.

In Rosenberg’s review of 117 cases of factitious disorder imposed on another, 9% of the involved children died, several of whom were sent home with their parents after parents were confronted with the diagnosis. Given the high rate of mortality and morbidity, children should be removed from the parents until further assessment indicates that it is safe for the family to be reunited. Child protective services usually take responsibility for this action. When this is done, there is a high likelihood that the mother will need intervention for suicidal ideation. Investigation into possible abuse of siblings should also be initiated.

After separation, the child must be treated for ongoing medical problems, as well as for psychological problems. For many children, longstanding refractory medical illnesses will resolve once they are separated from perpetrators. Many such children suffer from post-traumatic stress disorder and should be treated. As is typical of those with post-traumatic stress disorder, many victims avoid medical care as adults. On the other hand, others develop factitious disorder. By adolescence, many victims of factitious disorder imposed on another collude with parents in perpetrating deception.

The perpetrator, usually the mother, should also undergo treatment to diagnose and address underlying psychiatric conditions. The perpetrator should be evaluated to determine whether it is likely that they will ever achieve “good enough” parent status and whether reunification with their child is a realistic goal. Those who actively induce illness in their children are less likely to ever be adequate parents. If reunification is unrealistic, then psychotherapy should focus on the parent’s underlying psychopathology, as well as the loss of a child. If reunification is the goal, then treatment should address emotional maturation, the ability to put the child’s needs before their own, relinquishing the defense of denial, and learning to seek help and to express themselves in more appropriate ways.

Sanders et al. in 2019 outlined a system of therapeutic goals for perpetrators to adhere to, the ACCEPTS treatment model:

ACKNOWLEDGE (AC): acknowledgment and taking responsibility for inappropriate behaviors that harmed or could have harmed the child by describing their maladaptive behaviors in detail and to accept and understand how these behaviors placed the child victim(s) at risk.

COPING (C): development of a wider range of effective coping strategies to manage personal emotional needs, recognizing that abusive behavior is no longer a coping option.

EMPATHY (E): development of the ability to empathize with the child and experience an appropriate emotional response to the harm and suffering that the child experienced as a result of the abuser’s past maladaptive behaviors.

PARENTING (P): development of appropriate parenting skills and the ability to prioritize the child’s needs over those of the abuser.

TAKING CHARGE (T): recognition of the perpetrator’s own power and how to utilize it appropriately.

SUPPORT (S): As abusive behaviors may persist despite intervention, a support and monitoring system by child protective services and/or health professionals is necessary. If reunification occurs, a primary pediatrician should act as a gatekeeper who monitors and approves all medical care.

Although the literature on factitious disorder imposed on another focuses on the most severe cases, mild cases do occur. Meadow advocated that in these mild cases, parents should be supported and prevented from putting their children through needless investigations and treatments. Psychotherapy should be undertaken. The prognosis in these cases may not be so grim. Factors associated with positive outcomes according to Bools et al.’s study included (1) continuous positive input from spouse or grandparents; (2) successful short-term foster care before returning to live with the perpetrating caregiver; (3) the perpetrator’s long-term relationship with a social worker; (4) successful remarriage for the perpetrator; (5) early adoption of the victim; and (6) long-term foster care placement.

There is little literature on the treatment of factitious disorder imposed on another between adults. In these cases, removal of the victim through legal means is usually not an option if the victim is a legally independent adult, even if they might be emotionally dependent on and enmeshed with the perpetrator. Unless the victim has a legally dependent status, adult victims must initiate separation. Psychiatric comorbidity in the victim, if severe enough, may render the individual incapacitated and requiring hospitalization or even guardianship. Adult protective services and/or agencies overseeing the protection of developmentally disabled individuals must be notified when adult factitious disorder imposed on another is suspected.

In cases in which health care providers are suspected of perpetrating illness among their patients, legal prosecution should be instituted. Simply removing the provider is not adequate because the perpetrator is likely simply to seek another job, where more patients will be at risk.

FUTURE DIRECTIONS

Further studies are needed to elucidate risk factors for developing factitious disorder, best practices for management, and outcomes of specific treatments. With an understanding that deception in the medical setting is commonplace, research into the underpinnings of this behavior may help to identify means of reducing episodes of deception both in patients with factitious disorders and in the general public.

FURTHER READINGS

Asher R. Munchausen's syndrome. *Lancet*. 1951;1:339.

*Bass C, Halligan P. Factitious disorders and malingering: challenges for clinical assessment and management. *Lancet*. 2014;383:1422–1432.

Bass C, Jones D. Psychopathology of perpetrators of fabricated or induced illness in children: case series. *Br J Psychiatry*. 2011;199:113–118.

Bass C, Taylor M. Recovery from chronic factitious disorder (Munchausen's syndrome): a personal account. *Personal Ment Health*. 2013;7:80–83.

Bools CN, Neale BA, Meadow SR. Follow up of victims of fabricated illness (Munchausen syndrome by proxy). *Arch Dis Child*. 1993;69:625–630.

Burton MC, Warren MB, Lapid MI, Bostwick JM. Munchausen syndrome by adult proxy: a review of the literature. *J Hosp Med*. 2015;10(1):32–35.

*Davis P, McClure RJ, Rolfe K, et al. Procedures, placement and risks of further abuse after Munchausen syndrome by proxy, non-accidental poisoning, and non-accidental suffocation. *Arch Dis Child*. 1998;78:217–221.

Eisendrath SJ. Factitious physical disorders: treatment without confrontation. *Psychosomatics*. 1989;30:383.

Eisendrath SJ, McNiel DE. Factitious disorders in civil litigation: twenty cases illustrating the spectrum of abnormal illness-affirming behavior. *J Am Acad Psychiatry Law*. 2002;30:391.

*Flaherty EG, MacMillan HL; Committee on Child Abuse and Neglect. Caregiver-fabricated illness in a child: a manifestation of child maltreatment. *Pediatrics*. 2013;132:590.

Fliege H, Lee JR, Grimm A, Fydrich T, Klapp BF. Axis I comorbidity and psychopathologic correlates of autodestructive syndromes. *Compr Psychiatry*. 2009;50:327–334.

Gelenberg AJ. Munchausen's syndrome with a psychiatric presentation. *Dis Nerv Syst*. 1977;38:378.

Gregory RJ, Jindal S. Factitious disorder on an inpatient psychiatry ward. *Am J Orthopsychiatry*. 2006;76(1):31–36.

*Kenedi CA, Shirey KG, Hoffa M, et al. Laboratory diagnosis of factitious disorder: a systematic review of tools useful in the diagnosis of Munchausen's syndrome. *N Z Med J*. 2011;124(1342):66–81.

Kinns H, Housley D, Freedman DB. Munchausen syndrome and factitious disorder: the role of the laboratory in its detection and diagnosis. *Ann Clin Biochem*. 2013;50:194.

Krahn LE, Li H, O'Connor MK. Patients who strive to be ill: factitious disorder with physical symptoms. *Am J Psychiatry*. 2003;160:1163–1168.

Lawrie SM, Goodwin G, Masterton G. Munchausen's syndrome and organic brain disorder. *Br J Psychiatry*. 1993;162:545–549.

McClure RJ, Davis PM, Meadow SR, Silbert JR. Epidemiology of Munchausen syndrome by proxy, non-accidental poisoning, and non-accidental suffocation. *Arch Dis Child*. 1996;75:57–61.

McCullumsmith CB, Ford CV. Simulated illness: the factitious disorders and malingering. *Psychiatr Clin North Am*. 2011;34:621–641.

Meadow R. Management of Munchausen syndrome by proxy. *Arch Dis Child*. 1985;60:385.

Meadow R. Unnatural sudden infant death syndrome. *Arch Dis Child*. 1999;80:7.

Olry R, Haines DE. Historical and literary roots of Münchhausen syndromes: as intriguing as the syndromes themselves. *Prog Brain Res*. 2013;206:123–141.

Pankratz L, Lezak MD. Cerebral dysfunction in the Munchausen syndrome. *Hillside J Clin Psychiatry*. 1987;9(2):195–206.

Poole CJM. Illness deception and work: incidence, manifestations and detection. *Occup Med (Lond)*. 2010;60:127–132.

Reich P, Gottfried LA. Factitious disorders in a teaching hospital. *Ann Intern Med*. 1983;99:240.

Rosenberg DA. Web of deceit: a literature review of Munchausen syndrome by proxy. *Child Abuse Neglect*. 1987;11:547.

*Spence SA, Kaylor-Hughes CJ. Looking for truth and finding lies: the prospects for a nascent neuroimaging of deception. *Neurocase*. 2008;14:68–81.

Sutherland AJ, Rodin GM. Factitious disorders in a general hospital setting: clinical features and a review of the literature. *Psychosomatics*. 1990;31:392.

Wallach J. Laboratory diagnosis of factitious disorders. *Arch Intern Med*. 1994;154:1690.

Yates GP, Bass C. The perpetrators of medical child abuse (Munchausen syndrome by proxy)—a systematic review of 796 cases. *Child Abuse Negl*. 2017;72:45–53.

Yates GP, Feldman MD. Factitious disorder: a systematic review of 455 cases in the professional literature. *Gen Hosp Psychiatry*. 2016;41:20–28.

20

Dissociative Disorders

Richard J. Loewenstein, M.D., Roberto Lewis-Fernández, M.D., M.T.S., and Paul Frewen, Ph.D., C.Psych.

The *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition, Text Revision (DSM-5-TR) conceptualizes dissociation as a disruption, interruption, and/or discontinuity of the normal, subjective integration of potentially any aspect of experience and cognition, including behavior, memory, identity, consciousness, emotion, perception, body representation, and motor control. In more acute pathologic dissociative presentations, dissociation is primarily related to traumatic and/or overwhelming experiences. In lifelong dissociative adaptations such as dissociative identity disorder, dissociative symptoms can occur both in traumatic contexts and for nontraumatic events in everyday life. Overall, in dissociation, aspects of psychobiological functioning that should be associated, coordinated, and/or linked are not. This does not occur randomly but is due to potentially explicable psychobiological processes. Dissociation is described as both a dimensional, adaptational process and as a set of categorically distinct psychopathologies. In general, contributions to the literature on dissociation do not make this crucial distinction. This has led to conceptual confusion and, at worst, almost a Tower of Babel effect between different schools of thought writing about dissociation (see the sections on Nosology and Etiology).

The DSM-5-TR dissociative disorders are dissociative identity disorder, dissociative amnesia; depersonalization/derealization disorder; other specified dissociative disorders; and unspecified dissociative disorder. The International Classification of Diseases, 11th Edition (ICD-11) aligns the definition of dissociation more closely with the DSM-5-TR definition, particularly as compared with ICD-10. In ICD-11, “Dissociative disorders are characterized by involuntary disruption or discontinuity in the normal integration of one or more of the following: identity, sensations, perceptions, affects, thoughts, memories, control over bodily movements, or behavior.” Unlike DSM-5-TR, ICD-11 continues to place conversion disorder as dissociative neurologic symptom disorder, whereas the DSM-5-TR persists in classifying these conditions under somatic symptom disorders as functional neurologic symptom disorder (conversion disorder). In addition to the DSM-5-TR dissociative disorders and functional neurologic symptom disorder, ICD-11 includes trance disorder, possession trance disorder, partial dissociative identity disorder, and secondary dissociative syndrome, where dissociative symptoms are “a direct pathophysiological consequence of a health condition not classified under mental and behavioral disorders,” for example, focal seizures with impaired awareness (focal unaware seizures in ICD-11 [previously complex-partial seizures]).

In the following sections, many issues common to all the dissociative disorders are discussed together, and then each of the dissociative disorders is reviewed in greater depth.

HISTORY

The history of dissociation and dissociative disorders traverses the modern history of psychiatry from its very beginning to the present time. Embedded in this history are some of the most complex, important, and often controversial issues in psychiatry and some of the most important figures in the early development of the field of psychiatry, like Pierre Janet and Sigmund Freud. From the beginning of the study of dissociative disorders, controversies have also related to larger social, political, and cultural issues. Indeed, since the early 19th century, these disorders have generated media attention and led to an awareness of the public about them and debates about them. Controversies about dissociation and dissociative disorders will be addressed in later sections of this chapter.

The modern study of dissociation begins at the end of the 18th century with the shift of interest in phenomena of trance and possession from the religious to the medical realm, at least for the individuals in the scientific, medical, and increasingly secularized parts of society. Within the Western world, to this day, dissociative trance and possession experiences have continued to be normative in some religious groups and cultural contexts. From the 19th century until the present time, various forms of the paranormal and the supernatural, including spiritism, mediumship, astrology, and occultism—often associated with dissociative types of experiences—have continued to evoke interest among many members of the public.

In 1791, Eberhardt Gmelin described a German woman who allternatively exchanged her peasant personality for that of an aristocratic French lady, each amnesic for the other's existence. By the early 19th century, such cases, with diagnoses of dual, double, or duplex consciousness, were being regularly reported on both sides of the Atlantic. In 1811, Benjamin Rush, considered the father of American psychiatry, included a classic example in his medical school lectures. Rush proposed that dual personality reflected a functional disconnection between the two cerebral hemispheres. In the same year, the case of Mary Reynolds, who became the American archetype of dissociative identity disorder for the remainder of the century, was first published.

In Europe, the disciples of Franz Anton Mesmer developed the first systematic descriptions of what is now called *hypnosis.* These thinkers developed an interest in *magnetic diseases* (from Mesmer's theories of animal magnetism): Amnesia, fugue, somnambulistic states, and alternating or multiple personality that could be treated with *artificial somnambulism,* that is, hypnosis. These *magnetizers* were mostly outside the mainstream of European academic medicine, and their works are not widely known. Among them, Antoine Despine, a French family doctor, wrote systematic case studies and reports about several series of multiple personality patients and their treatment, including child and adolescent cases. In his cases, Despine alluded to childhood histories of sexual abuse that could not be written about. Also, he described multiple personality cases with numerous self-states, beyond the dual and triplex cases described in most case histories of the time, as well as posttraumatic and conversion symptoms. Many other 19th-century patients also had an extensive conversion and somatoform symptoms as part of the clinical presentation.

Psychiatric historians have charted the shift in interest from somnambulism to multiple personality to hysteria and back over the course of the 19th century, with *hysteria* finally seen as a unifying concept for all these conditions. Many 19th-century debates about dissociation prefigured those that continue into the 21st century. Paul Briquet, who wrote a famous study of hysteria in the mid-19th century, disputed the prevailing notion that the disorder was caused by female sexual frustration or hypersexuality. He viewed hysteria as a disorder of the brain, not the uterus. Briquet documented that traumatic, overwhelming, and grief-engendering experiences were strongly associated with hysteria. Mid-19th century France also witnessed the rise of popular movements such as spiritism, which attempted to study trance states and other dissociative activities (e.g., automatic writing) as forms of spirit communication. The work of Allan Kardec (pseudonym for Hippolyte de Rivail) was disseminated widely, sparking healing movements targeting—and mediated by—spirit possession, particularly across Latin America.

The famous 19th-century French neurologist Jean-Martin Charcot synthesized the teachings of the magnetizers with those of the more accepted medical and psychiatric establishment, including Briquet's etiologic theories. Charcot influenced major figures in psychiatry and neurology, such as Sigmund Freud, Pierre Janet, Georges Gilles de la Tourette, and Joseph Babinski. As is well known, he gave public demonstrations of female hysterical patients and hypnosis at his Paris hospital La Salpetrière, where the neuropsychiatric wards primarily housed indigent women. Charcot's weekly demonstrations of hysterical symptoms were open to the public and were considered major social events in Parisian society of the time. Hysterical individuals were depicted in numerous highly popular French plays, novels, short stories, and poetry; some of Charcot's patients became celebrities, as did Charcot himself.

Through these demonstrations and frequent interaction with Charcot and his associates, hysterical patients were readily prompted in various ways to perform spectacularly for him since gaining favor with Charcot often led to better treatment by the staff. Despite the relatively poor conditions for patients at La Salpetrière, these were often a considerable improvement over the poverty and misery of their outside lives. Some were magnetized (i.e., hypnotized) on the wards or encouraged by sycophantic followers of Charcot to act as dramatically as possible for the great man.

After his death, two of Charcot's most important successors, Babinski and Janet, took divergent views of hysteria. Babinski took the radically nihilistic view that all hysteria was caused by *suggestion* and could be removed by *persuasion* or *countersuggestion*. He coined the term *pithiatism* as a substitute for hysteria. The Nancy school of hypnosis, led by Hippolyte Bernheim, believed that hypnosis was not necessarily a sign of pathology, as Charcot believed, but an effect of suggestion. Babinski and Bernheim argued that the symptoms of Charcot's patients were artifacts of suggestion and contagion, not of bona fide disorders. This view continues today, most prominently in the thinking of Paul McHugh, a vehement critic of the current trauma-based theories of dissociation.

On the other hand, Janet, who remained at La Salpetrière after Charcot's death, emphasized the traumatic origins of hysterical symptoms. Janet's research and clinical theory, in particular the emphasis on the role of traumatic antecedents of dissociation, are widely regarded as the foundation for modern views of dissociation, although, during his lifetime, Janet never developed a school of followers to promulgate his ideas.

During Charcot's time, stage hypnotism began—as a pseudo-medical performance with a "doctor" and his "hysterical" subject, including subjects drawn from the audience. Many of the symptoms of Charcot's patients are duplicated today in stage hypnosis demonstrations. Modern hypnosis research has been concerned with separating what is essential about the hypnotic state and the hypnotic capacity of the subject with "demand characteristics," sometimes called sociocognitive aspects of hypnosis. The latter are the expectations subjects bring about how they are to behave while "being hypnotized"—that are different from the core features of the hypnotic state itself. In large part, demand characteristics account for the behavior of individuals in stage hypnotism demonstrations (see sections on Iatrogenic and Sociocognitive Model of Dissociation and Culture, Dissociative Disorders, and Sociocognitive Phenomena).

Trauma and the Patients at La Salpetrière

Recent scholarship has uncovered the medical records of the female patients at La Salpetrière during the 19th and early 20th centuries, including those studied by Charcot, Babinski, Janet, et al. in their research on hysteria—and some of Charcot's most famous demonstration cases. Supporting Janet's views, the hospital records document histories of multiple forms of severe trauma, loss, and social dislocation among the women.

These women were young, generally between 15 and 30 years old, and came from backgrounds of extreme poverty, illiteracy, violence, and traumatic loss. Many had brutal, incestuously abusive alcoholic fathers or stepfathers. Often, the women were the *de facto* caretakers for their younger siblings and witnessed the traumatic deaths of many of these children from tuberculosis, other infections, fevers, convulsions, etc., as well as witnessing the deaths of other family members. Some were illegitimate and shuttled between violent homes and religious institutions characterized by austerity and harsh discipline.

In addition, the later life histories of the hysterical women reported in the hospital records detailed rape, beatings, sexual exploitation by employers, and severe physical injuries due to trauma, in addition to earlier life traumas. Also, many had experienced wartime trauma during the 1870 Franco-Prussian war. They came from working-class neighborhoods directly involved in the street warfare in 1871 when the Paris Commune was crushed, a major fire devastated Paris, and thousands of communards were summarily executed in the streets of Paris. In addition, many of the "treatments" of the time were likely more traumatic than therapeutic: ovarian compression (often for hours) with a specific "ovarian compressor," applied electric currents and cervical cauterization with a hot iron. Drug treatment with opium, ether, chloroform, and/or amyl nitrate was also very common. Finally, based on the records, a number of these women were impregnated by Charcot's students. Nonetheless, the hospital was a more benign environment than their lives outside La Salpetrière. They were poor, massively traumatized, often uneducated, and, outside the hospital, were only allotted the most despised jobs in Parisian society. The accompanying case describes the history of profound trauma suffered by Blanche Wittmann, *la reine des* hystériques (the queen of hysterics), as well as her later life. She is also depicted in the classic painting: A Clinical Lesson at the Salpêtrière ("*Une leçon clinique à la Salpêtrière*") by Pierre Aristide André Brouillet (1857–1914) (see Fig. 20–1).

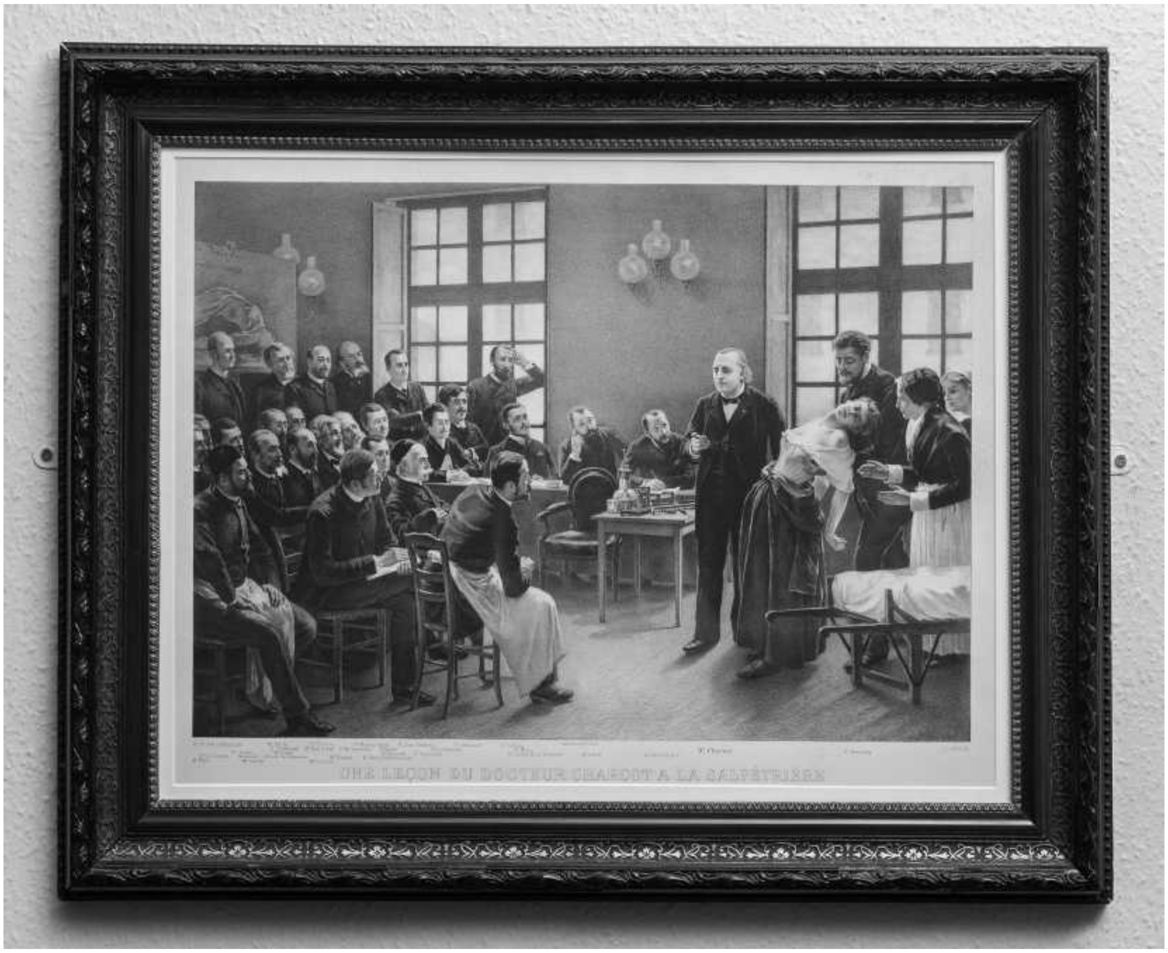
UNE LEÇON DU DOCTEUR CHARCOT A LA SALPÊTRIÈRE

FIGURE 20–1. A Clinical Lesson at the Salpêtrière ("*Une leçon clinique à la Salpêtrière*") by Pierre Aristide André Brouillet (1857–1914). Center: Charcot, Joseph Babinski, Blanche Wittmann. Others include Theodule Ribot (far left), Georges Gilles de la Tourette (front in the apron), and Paul Richer (just to the right of Charcot's arm). (© Freud Museum London.)

Marie (Blanche) Wittmann had a history of profound medical, emotional, physical, and sexual trauma in her early life. This included severe illnesses and infections before the age of 2; a violent father who was committed for life to a psychiatric hospital when she was 12; losses of five of her eight siblings; and a mother who was "nervous" and died when she was 15. She was apprenticed to a furrier who repeatedly physically and sexually assaulted her. A young man she loved died suddenly when she was 15. She had behavioral problems from early childhood and could not remain in school; thus, she was uneducated and only could be employed in the most menial and despised jobs. She suffered from frequent hysterical attacks and was hospitalized as well as working as a "ward girl," someone who did the most reviled tasks in the hospital. Admitted to La Salpetrière in 1877 at age 18, she assumed the name "Blanche." She became famous in Paris for her symptoms during Charcot's public demonstrations of hypnosis and hysteria. However, in 1888, treated by Jules Janet, he discovered a classic "dual personality" with a second self-state completely free of all the hysterical symptoms and reactions. This self-state was grounded, had a positive outlook, and reported co-consciousness (awareness of the thoughts, feelings, and actions) with the hysterical, depressed state. He attempted to increase the dominance of this apparently healthier state with some success. After Charcot's death in 1893, Wittmann shifted states again. She denied her whole prior history and went to work in the radiology laboratory of La Salpetrière. There she experienced severe radiation burns and stoically underwent repeated amputations of her fingers, hand, and arm bilaterally until she died of radiation-induced cancer in 1913.

History of Dissociation in the 19th and Early 20th Centuries

At the end of the 19th century, Janet in France, Morton Prince and William James in the United States, and others across Europe were engaged in a lively transatlantic discussion about possible psychological and neurologic mechanisms underlying cases of multiple personality, amnesia, and fugue. The scientific investigation of dissociation dates to the experiments of Prince in 1908, using a crude polygraph to measure galvanic skin resistance (GSR) across the personality states of a dissociative identity disorder patient—a physiologic measure that remains of interest today.

Prince reported seeing differential GSR reactivity to words that were emotionally laden for one personality state but not for another. Over the next century, several dozen psychophysiological and cognitive performance studies were added to the literature. Many of these reported findings suggesting differential responses across dissociative identity disorder self-states. Medical models of the time-invoked phenomena such as sleep and dreams, hypnosis and somnambulism, epilepsy, and disconnections between the cerebral hemispheres to explain multiplicity.

As the 20th century began, however, interest in dissociation waned, subsumed under the somatoform hysteria construct. Alternative theories of hysteria, like those of psychoanalysis, began their ascendance. However, clinical interest in dissociative phenomena has reoccurred in every war since the turn of the century with the observation of amnesia, fugues, and conversion symptoms in traumatized soldiers. Subsequently, these observations tended to be forgotten by the psychiatric community as the wars themselves faded from memory.

For example, Babinski's ideas about hysteria, renamed "pithiatism"—a disorder defined by symptoms created and removed by suggestion—had considerable influence on European psychiatrists until the onset of World War I. At that time, with the development of the concept of *shell shock,* psychiatrists systematically reported that virtually all the symptoms of dissociation, amnesia, automatism, and hysteria could be found among battle-traumatized soldiers, plainly contradicting Babinski's ideas and supporting those of Janet. During World War I and its aftermath, many military psychiatrists adopted psychodynamic and/or hypnotherapeutic treatment models for shell shock patients. Following Babinski, other military psychiatrists viewed hysterical symptoms in combat veterans as signs of mendacity and cowardice. They adopted "treatments" like applying powerful electric shocks to body areas manifesting hysterical symptoms. Similar conflicts have persisted from subsequent wars to today. Despite the increasing acceptance of combat-related PTSD, skepticism about its legitimacy and/or frequency continues within the military, the U.S. Veterans Health Administration, and among some psychiatrists and psychologists.

Janet and Freud

Freud and Joseph Breuer's revered *Studies in Hysteria* (1893) referenced Janet's work, used similar notions of "hypnoid states," fixed ideas, and traumatic etiology, and introduced the cathartic method of cure for hysteria. Anna O, the archetypal Freudian "hysteric," is specifically described as having a dual consciousness (personality) and a plethora of dissociative symptoms, such as complex amnesias for the current and historical experience; spontaneous age regressions and trances; depersonalization; fluctuations in handwriting, handedness, and language; and alternating states of consciousness (see the section on Diagnosis of Dissociative Identity Disorder).

Freud rejected his early dissociationist ideas, including the notion that hysteria and related conditions were caused by childhood sexual and/or physical violence (the "*seduction theory*"). Later psychoanalysts asserted that Janet dishonestly claimed priority for ideas related to hysteria, even though Janet's work clearly antedated Freud's, and Janet was referenced specifically in the *Studies in Hysteria.* Janet, for his part, was uncharacteristically publicly angry at this failure to acknowledge his work. Throughout his career, he remained critical of Freud and psychoanalytic ideas.

Despite the Janetian ideas found in early Freudian works on hysteria, until recently, psychoanalytic theory has had difficulty accounting for the phenomenon of dissociation. Freud's disciple Sandor Ferenczi wrote papers that described the sequelae of childhood sexual abuse and resulted in profound "splitting of the personality," analogous to modern descriptions of dissociative identity disorder. Ferenczi's insistence on the importance of actual traumas in psychopathology, as well as other issues, led to his expulsion from institutional psychoanalysis. British psychoanalysts from the Object Relations School, Harry Guntrip and Ronald Fairbairn, both had experience with sexually abused children. They emphasized the "splitting" of the ego as the result of these early traumas. Thus, the psychoanalytic term *splitting* frequently has been equated with dissociation, notwithstanding the criticism among some psychoanalysts of the reification of the splitting metaphor and the difficulty aligning the range of dissociative phenomena with the splitting concept.

David Rappaport, Merton Gill, Margaret Brenman, Charles Fischer, and Elisabeth Geleerd, among others, wrote important psychoanalytic papers and books on fugue, amnesia, hypnosis, and dissociation in the 1940s and 1950s, with Rappaport, following Janet, presciently enumerating the conceptual and phenomenologic differences between repression and dissociation. However, psychoanalytic thinkers have largely ignored this work, as it was hard to align with psychoanalytic theory of the time, including the notions that incest reports were “fantasies” and that hypnosis was anathema in analytic psychotherapy since Freud rejected this technique early in his work. Recently, many psychoanalysts have again recognized the reality of childhood and adult trauma. They have attempted to reintegrate dissociation into psychoanalytic thinking. However, this has been problematic because of the persistence of the conceptual legacy of splitting and the difficulty of fitting the full spectrum of dissociative psychopathology into this and other psychoanalytic constructs. Psychoanalysts of the relational school, such as Phillip Bromberg, have broadened the dissociation concept so widely that it is seen as characteristic of human psychological processes in general and of the experience of the psychoanalyst in all psychotherapy, not just therapy of individuals with dissociative disorders.

It is unfortunate that disputes over the primacy of ideas led to a schism between Janet and the Freudians. Modern conceptualization of dissociative disorders often involves a synthesis of Janetian ideas and Freudian ones. For example, current theories of treatment include the Janetian notions of systematic work with traumatic memories, using adjunctive hypnotic and imagery techniques, and challenges to posttraumatic cognitive distortions, using cognitive and behavioral psychotherapy, psychoeducation, and building life skills. The psychoanalytic concepts of transference, countertransference, object relations, projective identification, intrapsychic defense, and conflict, among others, are also essential to understanding modern therapy of dissociative patients.

Based on his work with the severely ill patients of La Salpetrière, Janet viewed hysterics as liable to psychic disintegration with stress. Freud, working with well-to-do, bourgeois patients in his private practice, viewed hysterics as psychologically strong, attempting to survive extreme conflicts leading to symptoms. Both these ideas are relevant to the modern conceptualization of dissociative disorders and their treatment. Research has shown that dissociative disorder patients are often highly symptomatic and impaired. However, studies have also shown that early-life dissociation may be a resilience factor, particularly in the most severe, early-life onset dissociative disorder, dissociative identity disorder. Early psychological “compartmentalization” of traumatic experiences appears to allow more normal development in and preservation of cognitive and creative abilities, along with a capacity for humor, hope, human attachment, and insight, despite reported histories of profound, repeated early life trauma (see the section on Personality Organization, Personality Disorders, and Dissociative Identity Disorder). Although requiring a more extended discussion than is possible here, clinicians and researchers typically do not identify that thinking about the mind typically involves physical metaphors like “splitting,” “compartmentalization,” etc. This may result in reification of the mind and may be a particular problem in conceptualization of dissociation and dissociative disorders, especially dissociative identity disorder.

Janet continued to be active in psychology and psychiatry until his death in 1947, writing on many different clinical, theoretical, and philosophical subjects. Janet never developed a formal school of psychology nor a group of disciples committed to his theories. He was deeply concerned with the protection of patient confidentiality, perhaps in part as a reaction to the excesses of the Charcot era. Accordingly, he had his library of detailed case histories, including those of the most famous classic hysterics, destroyed at his death. Because of this, and because of the primacy of the psychoanalytic theories in the latter part of the 20th century, his work was largely neglected until the recent resurgence of interest in dissociation.

Janet’s contributions to modern dissociation theory include (1) recognition of the causal role of trauma in dissociation in a stress-diathesis model; (2) the taxonometric approach—Janet believed that dissociation was not on a continuum from normal psychological experience; (3) recognition that the underlying mental mechanisms of somatoform and dissociative disorders were similar; (4) the role of *fixed ideas* in dissociative disorders and attempts to find the most fundamental problematic ideas that influence the patient’s behavior; and (5) treatment using a multimodal approach with hypnotic and trauma-focused methods combined with cognitive therapy, behavioral therapy, and life-skills building.

Dissociation During and After World War II

Clinicians treating battlefield casualties in World Wars I and II described dissociative symptoms, such as amnesia, fugue, automatisms, and somatoform symptoms, in traumatized soldiers as part of *traumatic war neurosis,* as it came to be known in World War II. The extent of amnesia described in psychiatric battlefield casualties in World War II ranged from relatively brief periods of time to complete amnesia for life history and identity (generalized dissociative amnesia), as well as fugue episodes.

Treatment of these World War II amnesia cases included hypnosis, narcosynthesis with a mixture of hypnosis and sodium amobarbital (Amytal), individual and group psychotherapy, or a combination of these. Attempts were made to treat these soldiers on the front lines and return them rapidly to combat. Detailed case descriptions of dissociative amnesia and fugue can be found in clinical studies from that time, as well as studies of large cohorts of patients. Researchers have reported that soldiers from the Korean conflict and the Vietnam war also reported amnesia as part of the posttraumatic syndromes related to combat experiences and were treated with similar modalities. Dissociative symptoms, dissociative disorders, and/or the dissociative subtype of PTSD continue to be described in U.S. combat veterans from the Gulf War and U.S. and NATO veterans of the wars in Iraq and Afghanistan. Cross-nationally, dissociation and/or dissociative disorders have been reported among combat veterans with PTSD. These include Israeli combat veterans from the 1967 Arab-Israeli War, with significantly higher levels of dissociation in former prisoners of war; Iranian veterans with PTSD from the Iran-Iraq War (DSM-IV dissociative amnesia, dissociative fugue, depersonalization disorder); and Turkish professional soldiers with PTSD.

Waning Awareness of Dissociation

Despite the extraordinary interest in multiple personality throughout the 19th century, the study of dissociation mostly disappeared after the beginning of the 20th century. Authorities suggest that this was related to the dominance of Freudian paradigms of hysteria; the disrepute into which hypnosis fell at this time; the rise of the Bleulerian concept of schizophrenia, which may have subsumed dissociative patients (see the section on Other Dissociative Disorders); and the loss of interest in the works of Janet, Prince, and others who had been crucial in the development of models of dissociation. From the 1930s to the late 20th century, some classical Freudian psychoanalysts, and later the Sullivanian psychoanalysts, began to treat “schizophrenics” with psychodynamic therapy. Simultaneously, partly in reaction to some of the barbaric “treatments” of the time—insulin and metrazol shock, intrathecal horse serum, colloidal gold infusions, and multiple daily ECT courses—there was a lack of concern about symptoms and diagnosis, viewing these as dehumanizing. This coincided with the expansion of the diagnosis of schizophrenia to encompass a wide swath of patients, following Bleulerian concepts of the relative unimportance of "surface" symptoms. Case studies of "successful" psychodynamic treatments using these models suggest that the responsive cases better fit the symptoms, trauma histories, and dynamics of dissociative disorder presentations, not those of individuals with psychotic disorders.

Nonetheless, throughout much of the 20th century, the psychiatric, psychological, and psychoanalytic literature described individual cases or small case series of individuals with dual or multiple personality and/or generalized dissociative amnesia, with and without dissociative fugue. From the 1890s to the present (but declining since the 1920s), newspaper accounts have documented numerous individuals with amnesia and fugue who were brought to hospitals without recall of their past history and/or personal identity. Conventional wisdom still held that these disorders were extraordinarily rare, especially multiple personality; thus, individual case studies, such as the famous *Three Faces of Eve,* were not only published in the professional literature but entered popular media as well. Nonetheless, systematic reviews of the literature periodically supported the validity of the diagnostic construct of dissociative identity disorder.

History of Depersonalization/Derealization Disorder. First described by Maurice Krishnaber in 1872, depersonalization was formally named in 1898 by Ludovic Dugas, who sought to convey "the feeling of loss of ego." Freud, Janet, Eugen Bleuler, and other 19th- and early 20th-century authorities reported patients with symptoms of depersonalization and derealization. Freud himself described his own symptoms of depersonalization and derealization in 1904 as part of his self-analysis, when he finally was able to visit the Acropolis. A monograph by Paul Schilder in 1914 is regarded as a turning point in psychiatric interest. Classic studies by Wilhelm Mayer-Gross in 1935, Harold Shorvonn in 1946, and Jerome Saperstein in 1949 furthered the delineation of a syndrome of chronic depersonalization. Brian Ackner enumerated the essential features of current diagnostic definitions in 1954. These include "(1) the feeling of unreality or strangeness apropos the self; (2) the retention of insight and lack of delusional elaboration; (3) the lack of affective response (‘numbness’) except for the discomfort regarding depersonalization; and (4) the unpleasant property that may vary in intensity inversely with the subject’s familiarity with the phenomenon.”

Later 20th Century History

The interest in trauma and dissociation began anew in the 1970s and 1980s. Social and cultural factors leading to this revival are thought to include the return from war of the Vietnam veterans and the systematic academic study of their psychiatric problems; the recognition of the prevalence of childhood physical and sexual abuse with its associated mandated reporting; systematic research on child abuse and family violence; the rise of feminism with its critique of psychological theories that ascribed sexual abuse reports to fantasy; increased academic rigor in theories about hypnosis, such as Ernest Hilgard’s *neodissociation* theory; greater acceptance of academic hypnosis research; popular interest in multiple personality owing to works such as *Sybil;* and the publication of DSM-III (1980), with its promulgation of diagnostic criteria for the dissociative disorders and posttraumatic stress disorder (PTSD).

Cornelia Wilbur. The contemporary era in the study of dissociative identity disorder began with the work of Arnold Ludwig and colleagues, including Cornelia Wilbur, at the University of Kentucky during the 1970s, primarily studying differential findings in the psychobiology of dissociative identity disorder self-states. Cornelia Wilbur, widely identified with the case of *Sybil,* was influential in describing the clinical features of the modern construct of dissociative identity disorder and identifying the role of childhood trauma as a major factor in the etiology of the disorder. In the mid-20th century, Wilbur, a psychoanalyst, described several cases of dissociative identity disorder, all with more complex self-state systems than the classical dual and duplex personality cases and all with histories of extreme early life maltreatment, including incest. However, her writings were rejected by psychoanalytic journals, primarily due to the prevailing *zeitgeist* that multiple personality and incest were one-in-a-million conditions, and that female reports of childhood sexual abuse were virtually all Oedipal fantasies. Finally, in frustration, Wilbur published the *Sybil* case in the popular press, with no expectation that it would be read widely, if at all.

DSM-III and Beyond

Since the publication of the DSM-III in 1980, substantial systematic research has developed about dissociative disorders, particularly dissociative identity disorder, dissociative amnesia, and depersonalization/derealization disorder. Many academic studies have also investigated dissociation as a quantifiable trait in clinical and nonclinical samples. Studies have also looked at the prevalence and phenomenology of amnesia, particularly as it relates to the question of delayed recall of traumatic experiences.

Beginning in the early 1980s, Frank W. Putnam, Jr. at the National Institute of Mental Health (NIMH) was one of the first to develop the modern scientific study of dissociative disorders and dissociation. He undertook the first contemporary systematic epidemiologic, phenomenologic, and psychophysiological studies of dissociative identity disorder, including attempts to repeat the classical studies of reportedly compartmentalized memory and differing psychophysiological responses of the self-states. Among his major findings were the *significantly greater variability* among dissociative identity self-states compared with simulating controls on a variety of psychophysiological variables such as event-related potentials (ERPs), 16-lead EEG, and autonomic functioning. He showed that the self-states did not exhibit the kind of test–retest similarity on these measures found in simulating control subjects. This was due to the inability of patients to maintain a specific state, shifting of states during research trials, and variably recurrent intrusion and interference among states during testing.

These findings led him to posit that dissociative identity disorder naturalistically existed more as "*the state of multiple states*," with state shifting leading to overlap, intrusion, and interference among personality states. The latter were described as constructed around specific emotions, a sense of self, and a group of state-dependent memories. This was different from the classical understanding of the disorder as characterized by several distinct, dramatically switching self-states that were highly defined, complex, and invariant when studied from one time to the next, maintaining "executive control."

Putnam and his colleague Eve B. Carlson developed the *Dissociative Experiences Scale* (DES), the first screening inventory for trait dissociation, and measured *dissociativity* in many diagnostic groups compared to normal controls, including individuals with PTSD, affective disorders, psychotic disorders, anxiety disorders, personality disorders, and dissociative disorders. The DES has been translated into more than 40 languages, and studies across cultural settings reveal strong similarities between Western and non-Western samples, attesting to the universality of dissociation/dissociative disorders. Putnam also contributed major works on the diagnosis and treatment of dissociative identity disorder and was among the first to systematically study dissociation in children and adolescents, developing measures comparable to the DES for younger individuals. This led to his promulgation of the discrete behavioral states (DBSs) model of dissociation, discussed in depth in the sections on Etiology and on Dissociative Identity Disorder.

Putnam also began a groundbreaking, longitudinal case-control study of girls with documented histories of incest that, with a greater than 90% retention rate, has continued for over 25 years, now following the children of the original participants. This study includes diverse measures of hypothalamic–pituitary-adrenal (HPA) axis functioning, dissociation, hypnotizability, development of clinical disorders, and social and school functioning, among others. These studies established the biologic mechanisms underlying accelerated puberty, increased morbid obesity, and advanced epigenetic aging in sexually abused women and girls.

David Spiegel of Stanford University studied cases of dissociative identity disorder and wrote influential papers on the relationship between hypnosis, trauma, PTSD, dissociation, and dissociative disorders. He and his colleagues also conducted a series of studies on acute dissociation in victims of natural disasters in California. This work led to the DSM-IV diagnostic criteria for acute stress disorder (ASD), now subsumed in the DSM-5 trauma- and stressor-related disorders category. Under Dr. Spiegel's direction, the DSM-IV changed the names of the dissociative disorders to emphasize their dissociative character and make them more clinically and theoretically relevant. He also took the lead in the changes in diagnostic criteria for dissociative disorders in DSM-5. He led major studies on dissociation and PTSD and directed the international studies that were a major factor in the inclusion of a dissociative subtype of PTSD in DSM-5. Also, Dr. Spiegel helped coordinate the publication of major papers in support of the trauma model (TM) of dissociation that challenges the sociocognitive and fantasy models (SCM/FM) of dissociation and of dissociative identity disorder (see sections on Etiological Theories).

International and Cross-Cultural Research on Dissociation and Dissociative Disorders. Beginning in the 1980s, there developed major international research efforts on dissociation and the dissociative disorders. Judith Armstrong, Bethany Brand, Suzette Boon, Chui-De Chiu, Phillip Coons, Martin Dorahy, Nel Draijer, Milissa Kaufman, Ruth Lanius, Lauren LeBois, Roberto Lewis-Fernández, Richard J. Loewenstein, Alfonso Martínez-Taboas, Warwick Middleton, Colin Ross, Vedat Sar, Daphne Simeon, Marlene Steinberg, Masahiro Umesue, Onno van der Hart, Eric Vermetten, and other researchers in the United States, Canada, Europe, Australia, New Zealand, Latin America, Turkey, Japan, China, and Taiwan began systematic research studies on the phenomenology, epidemiology, diagnosis, psychobiology, and treatment of dissociative disorders, especially dissociative identity disorder, in children, adolescents, and adults.

As part of this globalized investigation of dissociation, anthropologically informed research on pathologic possession and other trance states began to be conducted cross-culturally with psychometrically sound measures of dissociation. These studies revealed notable similarities between well-characterized cases of pathologic possession and some features of dissociative identity disorder, such as loss of personal identity followed by assumption of an alternate (usually spiritual) identity, high indices of trait and state dissociation, and elevated trauma exposure. These research findings paved the way for the inclusion of some pathologic possession presentations under the rubric of dissociative identity disorder in DSM-5. In ICD-11, by contrast, possession trance disorder, along with trance disorder, are included as diagnostic categories distinct from dissociative identity disorder under the rubric of trance disorder and possession trance disorder (the rationale for this classification is presented in the section on Comparative Nosology). Research by Roberto Lewis-Fernández and colleagues on the dissociative nature of some acute reactions to traumatic events, such as some instances of *ataque de nervios* (attack of nerves) in Latin America, also led to the inclusion of an acute dissociation category under the other specified dissociative disorders section in DSM-5.

Richard P. Kluft made many contributions to the literature on dissociation and dissociative identity disorder. He wrote extensively about the psychotherapy of dissociative identity disorder and the phenomenology of the disorder throughout the life cycle. He was the first to identify the high frequency in dissociative identity disorder of apparent *first-rank symptoms* (FRSs) of psychosis, albeit experienced quite differently in dissociative identity disorder and indicative of the overlap and interference of states described by Putnam. He is the only scholar to write extensively about the fusion and integration of self-states in the treatment of dissociative identity disorder.

The International Society for the Study of Trauma and Dissociation (ISSTD) began under a somewhat different name in 1984. This society has published three editions of Guidelines for the Treatment of Dissociative Identity Disorder in Adults (3rd revision in 2011) and Guidelines for the Evaluation and Treatment of Dissociative Symptoms in Children and Adolescents (2004). Both guidelines are currently under revision. Jennifer Freyd of the University of Oregon is the editor-in-chief of the ISSTD's journal, *The Journal of Trauma and Dissociation* (JTD). From 1988 to 1997, Richard Kluft and Catherine Fine edited the society's first journal, *Dissociation* (available at Scholarsbank.uoregon.edu/xmlui/handle/1794/1129). In 2006, Suzette Boon in the Netherlands and Bettina Overcamp in Germany chartered The European Society for Trauma and Dissociation (ESTD), now affiliated with the Association Francophone du Trauma et la Dissociation (AFTD). With Cyril Tarquinio as editor-in-chief, the ESTD publishes *The European Journal of Trauma & Dissociation / Revue Européenne du Trauma et de la Dissociation*,

Contemporary Study of Depersonalization/Derealization Disorder

Recent systematic research of two large cohorts of well-characterized participants with depersonalization/derealization disorder has yielded strikingly consistent findings. These were reported by Daphne Simeon of Mount Sinai Medical School in the United States and by Anthony David at the Depersonalization Research Unit (DRU) at the Maudsley Hospital, Kings College, London. Clinical features, course and prognosis, and cognitive and neurobiologic correlates have emerged that clearly distinguish this disorder from other psychiatric disorders with symptoms of depersonalization. Mauricio Sierra and German E. Berrios, also affiliated with the DRU, developed the most widely used instrument for screening and diagnosis of this disorder, the *Cambridge Depersonalization Scale* (CDS). Both the U.S. and the British groups have studied treatment and treatment outcomes of patients with depersonalization/derealization disorder, including psychopharmacology and psychotherapy (see the section on Depersonalization/Derealization Disorder Treatment).

DISSOCIATIVE DISORDERS, MODERN SOCIETY, AND THE MEDIA

The study of dissociative disorders appears to be inseparable from the social and cultural ambiance in which it occurs. Contemporary academic interest in trauma and its clinical outcomes has been paralleled by social and religious movements focusing on self-identification as a victim of childhood abuse and the seeking of recovery through a variety of therapeutic and self-help modalities. Furthermore, some feminist theorists argued that healing from childhood sexual abuse mandated confrontation with alleged perpetrators in personal meetings, in the courts, or both.

Legal theorists concerned that victims of childhood maltreatment did not have the standing to sue for damages or to bring criminal actions under usual statutes of limitations lobbied state legislatures to expand the length of time that victims could seek legal redress. This was based on the idea that many survivors had amnesia for abuse or did not recognize the harm that had been done to them until many years after the abuse had occurred.

Many states did modify the statute of limitations for childhood sexual abuse, allowing the statute to begin to toll after the person "recovered" memory of the abuse or, alternatively, after recognizing the harm that had been done, whether amnesia was present or not. Successful legal actions were brought under these definitions, with courts accepting expert testimony concerning the validity of "repressed memory" for childhood maltreatment. Other state courts have explicitly rejected testimony based on delayed recall of traumatic memory.

In the 1980s, clinicians working with dissociative disorders began to report patient accounts of multi-perpetrator, multi-victim, organized sadistic abuse in the context of supposed underground cult groups. These chilling accounts were amplified by highly publicized daycare abuse cases that described multiple perpetrators, in some cases with elements of ritualized, occultist abuse. Several state governments developed commissions to investigate allegations of "ritual abuse" and claimed to have found evidence of these activities. The media also developed an intense interest with the publication of popular autobiographical accounts of individuals with reported histories of occultist abuse, as well as frequent news stories and television specials on this and related topics.

Unfortunately, some clinicians working with these dissociative patients chose to respond to their patients' reports in a highly publicized and sensationalized manner rather than subjecting them to dispassionate and rigorous scrutiny. As discussed later, one of the classic countertransferences in trauma treatment, whether for combat veterans or dissociative disorder trauma survivors, is overidentification with the patient with loss of the treatment boundaries. The therapist takes on an advocacy role because of some "special" aspect of the patient's trauma(s). This situation invariably leads to negative outcome for the patient(s) and often for the therapist(s). Some of these clinicians painted themselves as crusaders against a vast criminal conspiracy that threatened the nation, presenting themselves and their patients on television and in the newspapers. Patients in some specialized dissociative disorder units shared detailed recollections of trauma in group therapy and in the hospital milieu, with relatively little concern for possible cross-contamination or confusion about the origin of accounts of trauma. Cautionary or critical voices among dissociative disorder researchers and clinicians were ignored or rebuked. Forensic and clinical research was disregarded that cast doubts on the veracity of some of these patients' accounts and suggested alternative mechanisms for many of these "ritual abuse" narratives. Simultaneously, police were involved in the arrest and prosecution of individuals accused of the sadistic sexual abuse of children in daycare settings.

Backlash

Ultimately, an extensive backlash occurred that included the dismissal of all the daycare cases, even those with solid, independent confirmation of the abuse, and increased skepticism about the reliability of children's reports of maltreatment, particularly for sexual abuse. Popular works that encouraged readers to work on recovery from abuse were castigated as suggesting abuse to millions when none had occurred. Allegations of delayed recall for childhood trauma (unfortunately often termed "*recovered memory"*) in clinical and forensic cases were disparaged as based on erroneous views of human memory and cognition, although no one had questioned previously the phenomenon of amnesia related to combat, rape, or other types of traumas. The dissociative disorders construct was dismissed completely as engendered by suggestion and sociocognitive factors (see the section on Etiology). Therapy based on working with traumatic memories was discounted as "risky." The reports of an extensive prevalence of childhood abuse in the general population and its actual impact on psychopathology were questioned. Some critics raised doubts about the validity of the diagnostic construct of PTSD.

False Memory Syndrome. From 1992 until its demise in 2019, The False Memory Syndrome Foundation (FMSF) had a significant impact on this debate. The founders of this organization sought out "accused" families, like themselves, some of whom had had successful criminal or civil judgments brought against them because of their children's allegations of abuse. This organization, accompanied by several attorneys, developed a legal theory that there was no such thing as a delayed recall for trauma or dissociative identity disorder. The organization's argument went as follows: poorly trained clinicians convinced credulous (usually female) patients that their difficulties were all based on a history of repressed childhood abuse and engaged in risky therapeutic techniques (*guided imagery, hypnosis*) that led patients to believe that these confabulations were accurate. The patients then were mistakenly urged to confront or to stop all contact with family members accused of sexual abuse, ritual abuse, and other forms of maltreatment.

A group of attorneys and academic expert witnesses sought out patients and ex-patients to bring civil cases throughout the country in which therapists were sued for malpractice or were subject to licensing board complaints, or both, based on the FMSF position. Recanting former victims and/or "accused" parents sued mental health providers alleging malpractice and/or alienation of affections, respectively, based on the notion that there is no such thing as dissociation, amnesia, "repression" of memory, and dissociative disorders. The FMSF had a sophisticated public relations arm, and the media now became enthralled by weeping "recanters" reuniting with their previously abjured families. Many of these cases were settled. Several resulted in jury findings for the plaintiffs with significant damage awards. Others concluded with juries finding for the defendant's doctors. A federal mail fraud case against several clinicians ended in a mistrial before the prosecution's case had concluded, with the prosecution declining to retry the case.

Only one published study has systematically assessed recanters of abuse allegations, amnesia, and dissociative identity disorder. It found that virtually all these individuals had long psychiatric histories, with PTSD and dissociative symptoms predating contact with defendant clinicians. These patient plaintiffs had significant personality disorders in which they readily took the victim role and looked to forces outside themselves for an explanation of their problems. Previously, some had been involved in successful civil and criminal actions against their parents. Factitious elements were prominent in many patients' presentations. With treatment, most had *improved* with respect to their dissociative psychopathology, but not in their characterologic difficulties. Later, usually after the termination of treatment, the patients shifted their perception of victimization from their parents to their prior therapists, now putting their attorneys in the role of rescuer, usually with the promise of large sums of money if they prevailed in their lawsuits. The experience of many experts suggests that some therapies that resulted in false-memory lawsuits could have resulted in litigation for treatment below the standard of care for dissociative disorders.

Despite its popularity in the media and psychology textbooks, *false-memory syndrome* as a clinical construct has never been operationalized, studied, or validated. As the false-memory fad began to fade, the polarized, extreme views about the delayed recall of trauma have been better resolved through research and increased communication between trauma and cognitive researchers, who reject both the extreme false memory view as well as the extreme credulous view that holds that all reported traumatic memories are accurate (see the section on Etiology). "False-memory" experts continue to appear for the defense in high-profile sexual assault or murder cases against prominent, wealthy men and women, in civil and criminal litigation against accused pedophile clergy, and even in asbestos litigation to question the reliability of plaintiffs' memories. In these recent cases, most juries have found criminal defendants guilty and, in civil cases, liable for damages.

Persistence of Controversy About Dissociation and Dissociative Disorders

Despite the extensive body of systematic, international research on dissociation and dissociative disorders since the 1980s, passionate disagreements remain surrounding the disorders that evolved from classic *hysteria.* These academic and social debates parallel many of those that polarized the field in the days of Babinski and Janet. No other group of disorders in psychiatry can evoke so much heated controversy, conflict, and fervent debate.

Skeptical pieces about dissociative amnesia, the nature of dissociation, and dissociative identity disorder continue to appear in the media and psychiatric journals. It is unclear why skepticism specifically focuses on disorders most associated with severe adversity during early life, primarily intrafamilial maltreatment. Some of the same authors have questioned data on the prevalence of combat trauma and PTSD in Vietnam veterans, as well as the validity of the PTSD diagnosis, suggesting that combat veterans with PTSD are mostly malingerers.

Why should there be such a passionate debate? Why should appeals to the results of scientific research not resolve it? More so than other psychiatric conditions, these disorders touch on many complex and contentious aspects of personal, political, philosophical, and even religious beliefs. These include the fundamental nature of the self, the mind, memory, volition, and consciousness, as well as the responsibility of individuals for their behavior. In psychiatry and psychology, dissociative disorders are the focus of controversies, including longstanding debates between mentalists and behaviorists; between *psychodynamically oriented* and *behaviorally oriented* clinicians; between various researchers in cognitive psychology; between cognitive researchers and clinical researchers and practitioners, and between different theoretical schools at odds over the nature of hypnosis. Some Western fundamentalist Christian clinicians view *actual* demonic possession as part of the differential diagnosis of dissociative disorders.

Issues here include the significance of early trauma for human psychopathology; the existence of *unconscious* mental life and *intrapsychic defenses,* such as *repression* and *dissociation;* the nature of memory; and the nature of hypnosis, and whether it involves altered states of consciousness or is only a “learned social role.” The existence of dissociative amnesias and delayed recall of reported traumatic childhood events raise difficult-to-answer questions about the reliability of traumatic memory.

When viewed from a larger sociopolitical perspective, dissociation theory intersects with many of the most controversial social issues of modern times. Recent systematic research consistently has found a robust relationship between dissociation and traumatic experiences, increasing with earlier, more severe, and prolonged trauma, as well as for subgroups of severely maltreated individuals under conditions of captivity. The role of trauma, particularly intergenerational violence and sexual abuse, crosses into historically taboo subjects, such as rape, incest, child abuse, domestic violence, trafficking of both children and adults in prostitution and/or pornography, and their actual prevalence across societies. This leads to larger legal, social, and cultural questions related to peace and war, the meaning of violence cross-culturally, including differing religious views about the relationship between men, women, and children, and the nature of the family.

Dissociation and the Media

The media presentation of dissociative disorders also plays a role in controversies. From the mid-19th century to the present, the popular media has nurtured a public fascination with various forms of dissociation, especially “multiple personality” but also fugue, amnesia, and crimes committed under the influence of hypnosis or in somnambulistic states. Accounts of 19th-century cases were carried in popular magazines such as *Harpers,* and famous cases, such as Morton Prince’s Miss Beauchamps (1905), became the subject of books and plays. Robert Louis Stevenson’s fictional *The Strange Case of Dr. Jekyll and Mr. Hyde* (1886) is the most well-known of these works. Searches of newspaper archives from the mid-19th century onward show many cases of amnesia and fugue described in the popular press (now also on television and the internet), often with the goal of identifying the amnestic individual or reporting the successful identification of the person and reunion with the family.

Starting in the late 20th century, popular first-person accounts of multiple personality, such as *The Three Faces of Eve, Sybil,* and others, continued in this vein. The widespread use of multiple personality as a fictional plot device—generally to give a bizarre twist to the story—has contributed to the public and professional confusion that surrounds this disorder. Similarly, amnesia for life circumstances and for trauma is also a recurrent story mechanism in contemporary novels, films, and television—almost never portrayed accurately. Media stereotyping occurs for many forms of mental illness, but the distortions typical in most fictional depictions of dissociative disorders are particularly misleading.

Unfortunately, it often seems that most mental health practitioners also have received their training in dissociative disorders from the media, thus leading to misdiagnosis and mistreatment of dissociative patients, often accompanied by blanket skepticism: “I don’t believe in that.” Experts in dissociative disorders do not “believe in” these media portrayals either. As delineated throughout this chapter, the clinical phenomenology of dissociative disorders, as well as the evidence-based treatment models, bears little resemblance to these media depictions.

Highly publicized criminal and civil cases, such as the Hillside Strangler, the Franklin murder case, the Ramona case, and many others, have repeatedly brought additional media attention to claims of multiple personality, amnesia, and recovered memory. These contentious cases, with their warring academic experts and sensationalized media depictions, have also fueled popular stereotyping of dissociative disorders. Media spin about these cases has often been inaccurate and misleading.

Popular television interview programs have featured therapists and dissociative identity disorder patients, often with interviewers and the audience demanding to see the patients “switch.” The ISSTD treatment guidelines specifically enjoin clinicians from involving themselves and their patients in the media. Thus, for the most part, only poorly trained and/or exploitative clinicians have been involved in these media debacles. Due to their extensive traumatization, dissociative identity disorder patients are often highly susceptible to later exploitation, including involvement in the media without informed consent about how damaging this can be, as well as the permanent record that will be left in the age of the Internet.

Groups of individuals with dissociative identity disorder have developed “self-help” groups on social media. In general, these have resulted in disastrous outcomes when exploitative individuals have taken advantage of group members, group members have denigrated treatment for the disorder, and/or flaunted and encouraged eating disorders, nonsuicidal self-injury, and/or suicidal behavior.

More recently, individuals with dissociative identity disorder—or feigned dissociative identity disorder—have created YouTube, Tik-Tok, and/or Facebook channels to display their purported self-states, including individual Facebook pages for these entities. This may lead to increased skepticism of the disorder by viewers, antitherapeutic reinforcement of self-states as separate “people,” and vulnerability to exploitation. Also, there are reports that social media promulgation of the disorder has led to cases of “imitative” dissociative identity disorder, where individuals who do not meet diagnostic criteria for the disorder—as well as their concerned others—come to believe that they suffer from this condition.

Institutional and Cultural Trauma

Across the world, the reported failure of many institutions to respond to reports of trauma has also led to intense, polarized, and often politicized controversies. For example, in addition to lingering controversy and stigma about combat-related PTSD, the presence of increasing numbers of women in the military and in the U.S. Veterans Health System has led to an awareness of military sexual harassment and sexual assault, affecting both men and women.

Clergy sexual abuse scandals have involved the Catholic Church and many other religious groups. Religious hierarchies have failed to respond appropriately to these abuse reports, sacrificing the well-being of children to that of the accused clergy. Internationally, a few clerical administrators have been removed from their positions, and some were prosecuted and convicted. The recent, highly publicized #Me-Too movement has led to an international, cross-cultural outcry from women about gender-based violence, including sexual harassment and assault, as endemic in virtually all societies across the world and tolerated, if not encouraged, by most institutions and industries. These include government, business, military, all levels of education, athletics, media, technology, religious, medical, mental health, and entertainment. Similarly, there is increased awareness of violence against gay, lesbian, and transgender people.

In Australia, Canada, and the United States, there are documented reports of institutional sexual, physical, emotional, spiritual, and cultural maltreatment of indigenous children who had been removed from their families and cultural groups. The goal was to instill Western cultural and religious beliefs and to extirpate indigenous societies. Children’s bodies who died under suspicious circumstances have been found buried on the grounds of the institutions. In parts of the United States and Europe, similar types of abuse and deaths have been described in secular agencies for juvenile lawbreakers and in religious institutions for unwed mothers and their babies.

At a broader societal level, these can be conceptualized in terms of the long-term, traumatizing effects of historical trauma as well as systemic racism, sexism, and heterosexism. There are clinical and research data that systemic racism itself, including microaggressions, engenders symptoms of PTSD, dissociation, and depression.

Posttraumatic and Dissociative Responses to Institutional and Cultural Betrayal Trauma

Jennifer Freyd and her colleagues have evaluated the adverse effects of betrayal that results from childhood maltreatment by caregivers and/or others on whom the developing child must depend. Compared with traumatized individuals who do not report betrayal, these effects include higher levels of dissociation (particularly dissociative amnesia), alexithymia, anxiety, depression, nonsuicidal self-injury, and physical health complaints (see sections of Etiological Theories of Dissociation and Dissociative Disorders and sections on Dissociative Identity Disorder and Dissociative Amnesia).

Institutional Betrayal Trauma. Freyd and her colleagues have studied the adverse psychological effects of *institutional betrayal trauma*. This is defined as harm by institutions to those dependent on them, including failure to prevent or respond supportively to wrongdoings within the institution itself when there is a reasonable expectation of protection. Institutional betrayal can involve pragmatic harms (e.g., multiple adversities due to systemic racism) and psychological forms of distress. The pragmatic harms of racism can include multigenerational disparities in economic conditions; inadequate access to health care, particularly mental health care; lack of access to good educational resources; government policies that negatively affect social cohesion and support, such as the destruction of healthy neighborhoods of communities of color, replacing them with highways or other projects, with the dispersal of community members; de facto housing segregation into neighborhoods that are nearer to toxic pollution sources, more affected by adverse climate change, and with less access to safe water and adequate nutrition; higher violent crime rates along with a dearth of trust in police agencies, and many others.

Institutional betrayal trauma in sexually traumatized female college students was uniquely associated with dissociative symptoms and health problems, even when controlling for a history of personal betrayal trauma. In female military service members and veterans with a history of military sexual trauma, institutional betrayal was significantly associated with more severe depression and PTSD symptoms, including dissociative flashbacks and avoidance, the latter involving dissociation. They were significantly less likely to seek treatment in the VA system because of institutional mistrust.

Cultural Betrayal Trauma. Jennifer M. Gómez has defined another form of betrayal trauma, *cultural* or *ethnocultural betrayal trauma.* Cultural betrayal negatively affects the sense of connection, attachment, and dependence that minoritized group members may rely on as a form of self-protection from the effects of systemic racism and related societal harms. Within-group violence in minoritized populations is an implicit violation of this in-group trust, a cultural betrayal. This is the first factor in cultural betrayal trauma as intra-cultural trust may be a protective factor against the societal institutional betrayals inherent in systemic racism. Also, intra-cultural pressure may lead to cultural betrayal when within-group dynamics in a minoritized group silence or punish victims of cultural betrayal trauma (e.g., because of fears that this victimization would reflect negatively on the entire minoritized group). An illustration is the public suggestion by the president of a historically Black university that female African American students not report rape by male African American students to protect the latter from the societal trauma of the mass incarceration of Black men and to keep the issue within the Black community. Studies using this model among some U.S. ethnoracial minoritized groups (e.g., African Americans, U.S. Latinx, Asian Americans) have found that, even when controlling for trauma (e.g., sexual assault), cultural betrayal and intra-group pressure can lead to significantly increased dissociation, posttraumatic stress symptoms, dissociative hallucinations, internalized prejudice, and trauma-related deformation of ethnic identity.

COMPARATIVE NOSOLOGY

Diagnostic Criteria: DSM III-DSM 5-TR

The modern conceptualization of dissociative disorders as a distinct group began with the introduction of the DSM-III in 1980. Although the DSM-I distinguished dissociative reactions from other reactions, the DSM-II subsumed dissociative conditions under the superordinate category of hysterical neurosis. The latter was conceptualized as conversion or dissociative subtypes.

In the DSM-III, conditions that developed from the 19th-century construct of hysteria were distributed among different diagnostic categories: dissociative disorders, somatoform disorders—now somatic symptom and related disorders in DSM-5-TR—PTSD, and histrionic and borderline personality disorders. The DSM-III, as did the 10th edition of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), took the stance that the label *hysteria* had become so imprecise and variously defined that it had become meaningless. In the DSM-III, these conditions were defined and organized separately so that systematic empirical research could better clarify their reliability, validity, and possible relationship to each other. The DSM-III dissociative disorders were psychogenic amnesia, psychogenic fugue, multiple personality disorder, and depersonalization disorder.

The evolution from the DSM-III conceptualization of dissociation and the dissociative disorders diagnostic criteria through DSM-III-R, DSM-IV, DSM-IV-TR, DSM-5, and DSM-5-TR has been based on an increasingly sophisticated body of international research on the phenomenology, course, etiology, treatment, and treatment outcome of the dissociative disorders (see subsequent sections). These data support the robust relationship between dissociative disorders with traumatic and/or overwhelming experiences (see sections on History, Epidemiology, and Etiology). This research has been buttressed by the development of a variety of symptom screening measures for dissociation and dissociative disorders, as well as clinician-administered interviews for dissociative symptoms and diagnosis of dissociative disorders. The most important of these measures are detailed in Table 20–1.

The DSM-III criteria for dissociative disorders were established by expert consensus. By the time DSM-III-R was published, systematic research on dissociative disorders had increased substantially. Little changed in the diagnostic criteria for depersonalization disorder, psychogenic amnesia, and psychogenic fugue between DSM-III and DSM-III-R. However, the criteria for multiple personality disorder changed considerably to make them more flexible, less reified, and more compatible with research findings on the phenomenology of self-states.

In the DSM-IV (and IV-TR), the names of most of the disorders were changed to specify their dissociative nature more clearly (e.g., psychogenic amnesia became dissociative amnesia and multiple personality disorder became dissociative identity disorder). The criteria for dissociative amnesia specified that the amnesia usually occurred for traumatic or stressful circumstances. Dissociative fugue now could be diagnosed in the absence of the assumption of another identity. Amnesia was added as a criterion symptom for dissociative identity disorder. The disorders were structured to reflect pathologies of memory, identity, and perception, respectively.

The DSM-IV (and TR) category of dissociative disorders not otherwise specified (NOS) contained several examples that became particular areas of attention in DSM-5: “Clinical presentations similar to dissociative identity disorder that fail to meet full criteria for this disorder. Examples include presentations in which a) there are not two or more distinct personality states or b) amnesia for important personal information does not occur;” “Derealization unaccompanied by depersonalization in adults;” and “Dissociative trance disorder…[and] Possession trance disorder…[where the]…dissociative or trance disorder is not a normal part of a broadly accepted collective cultural or religious practice.”

Table 20–1.
Most Commonly Used Reliable and Valid Screening Measures and Structured Diagnostic Interviews for Dissociation and Dissociative Disorders

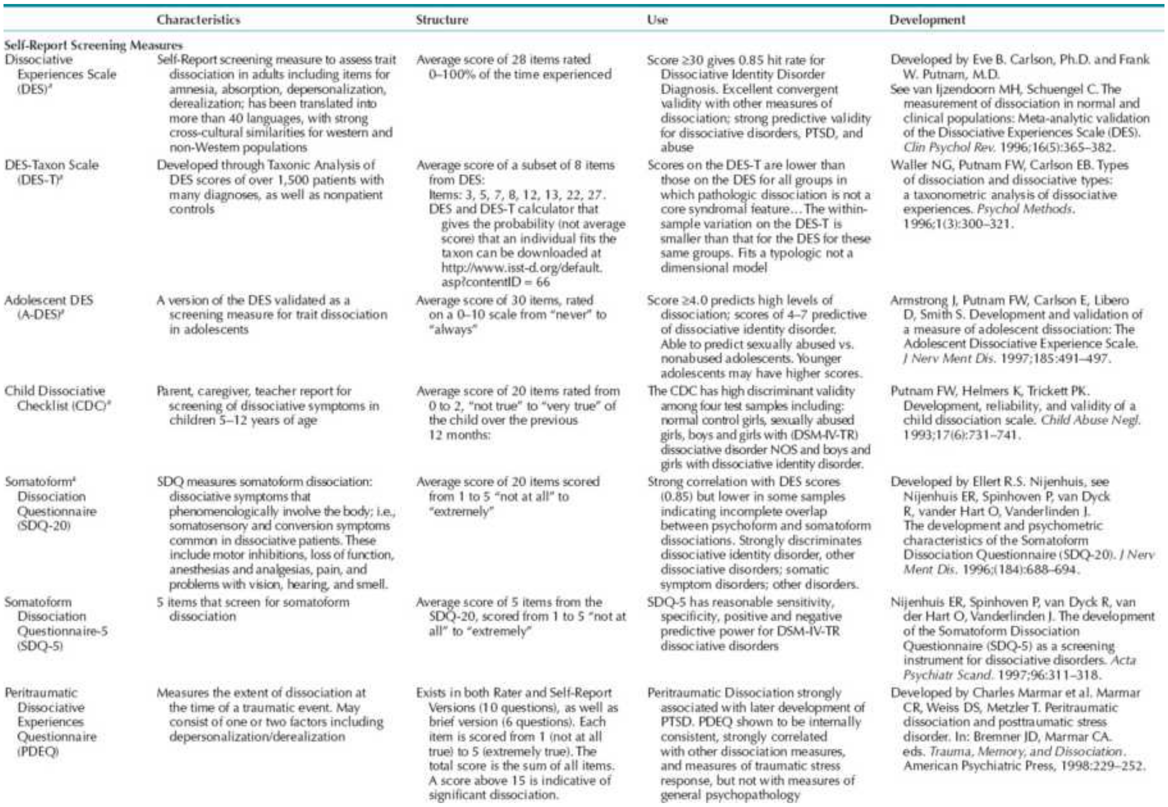

	Characteristics	Structure	Use	Development
Self-Report Screening Measures				
Dissociative Experiences Scale (DES)[a]	Self-Report screening measure to assess trait dissociation in adults including items for amnesia, absorption, depersonalization, derealization; has been translated into more than 40 languages, with strong cross-cultural similarities for western and non-Western populations	Average score of 28 items rated 0–100% of the time experienced	Score ≥30 gives 0.85 hit rate for Dissociative Identity Disorder Diagnosis. Excellent convergent validity with other measures of dissociation; strong predictive validity for dissociative disorders, PTSD, and abuse	Developed by Eve B. Carlson, Ph.D. and Frank W. Putnam, M.D. See van Ijzendoorn MH, Schuengel C. The measurement of dissociation in normal and clinical populations: Meta-analytic validation of the Dissociative Experiences Scale (DES). *Clin Psychol Rev.* 1996;16(5):365–382.
DES-Taxon Scale (DES-T)[a]	Developed through Taxonic Analysis of DES scores of over 1,500 patients with many diagnoses, as well as nonpatient controls	Average score of a subset of 8 items from DES: Items: 3, 5, 7, 8, 12, 13, 22, 27. DES and DES-T calculator that gives the probability (not average score) that an individual fits the taxon can be downloaded at http://www.isst-d.org/default.asp?contentID = 66	Scores on the DES-T are lower than those on the DES for all groups in which pathologic dissociation is not a core syndromal feature…The within-sample variation on the DES-T is smaller than that for the DES for these same groups. Fits a typologic not a dimensional model	Waller NG, Putnam FW, Carlson EB. Types of dissociation and dissociative types: a taxonometric analysis of dissociative experiences. *Psychol Methods.* 1996;1(3):300–321.
Adolescent DES (A-DES)[a]	A version of the DES validated as a screening measure for trait dissociation in adolescents	Average score of 30 items, rated on a 0–10 scale from "never" to "always"	Score ≥4.0 predicts high levels of dissociation; scores of 4–7 predictive of dissociative identity disorder. Able to predict sexually abused vs. nonabused adolescents. Younger adolescents may have higher scores.	Armstrong J, Putnam FW, Carlson E, Libero D, Smith S. Development and validation of a measure of adolescent dissociation: The Adolescent Dissociative Experience Scale. *J Nerv Ment Dis.* 1997;185:491–497.
Child Dissociative Checklist (CDC)[a]	Parent, caregiver, teacher report for screening of dissociative symptoms in children 5–12 years of age	Average score of 20 items rated from 0 to 2, "not true" to "very true" of the child over the previous 12 months:	The CDC has high discriminant validity among four test samples including: normal control girls, sexually abused girls, boys and girls with (DSM-IV-TR) dissociative disorder NOS and boys and girls with dissociative identity disorder.	Putnam FW, Helmers K, Trickett PK. Development, reliability, and validity of a child dissociation scale. *Child Abuse Negl.* 1993;17(6):731–741.
Somatoform[a] Dissociation Questionnaire (SDQ-20)	SDQ measures somatoform dissociation: dissociative symptoms that phenomenologically involve the body; i.e., somatosensory and conversion symptoms common in dissociative patients. These include motor inhibitions, loss of function, anesthesias and analgesias, pain, and problems with vision, hearing, and smell.	Average score of 20 items scored from 1 to 5 "not at all" to "extremely"	Strong correlation with DES scores (0.85) but lower in some samples indicating incomplete overlap between psychoform and somatoform dissociations. Strongly discriminates dissociative identity disorder, other dissociative disorders; somatic symptom disorders; other disorders.	Developed by Ellert R.S. Nijenhuis, see Nijenhuis ER, Spinhoven P, van Dyck R, vander Hart O, Vanderlinden J. The development and psychometric characteristics of the Somatoform Dissociation Questionnaire (SDQ-20). *J Nerv Ment Dis.* 1996;(184):688–694.
Somatoform Dissociation Questionnaire-5 (SDQ-5)	5 items that screen for somatoform dissociation	Average score of 5 items from the SDQ-20, scored from 1 to 5 "not at all" to "extremely"	SDQ-5 has reasonable sensitivity, specificity, positive and negative predictive power for DSM-IV-TR dissociative disorders	Nijenhuis ER, Spinhoven P, van Dyck R, van der Hart O, Vanderlinden J. The development of the Somatoform Dissociation Questionnaire (SDQ-5) as a screening instrument for dissociative disorders. *Acta Psychiatr Scand.* 1997;96:311–318.
Peritraumatic Dissociative Experiences Questionnaire (PDEQ)	Measures the extent of dissociation at the time of a traumatic event. May consist of one or two factors including depersonalization/derealization	Exists in both Rater and Self-Report Versions (10 questions), as well as brief version (6 questions). Each item is scored from 1 (not at all true) to 5 (extremely true). The total score is the sum of all items. A score above 15 is indicative of significant dissociation.	Peritraumatic Dissociation strongly associated with later development of PTSD. PDEQ shown to be internally consistent, strongly correlated with other dissociation measures, and measures of traumatic stress response, but not with measures of general psychopathology	Developed by Charles Marmar et al. Marmar CR, Weiss DS, Metzler T. Peritraumatic dissociation and posttraumatic stress disorder. In: Bremner JD, Marmar CA. eds. *Trauma, Memory, and Dissociation*. American Psychiatric Press, 1998:229–252.

Self-Report Screening Measures				
Cambridge Depersonalization Scale (CDS)	Assessment measure for depersonalization and derealization symptoms	29 questions rated for frequency and duration over past 6 months. Score >70 indicative of depersonalization/derealization disorder. Significantly higher than patients with anxiety disorders, mood disorders, and partial-complex seizures. Dissociative identity disorder patients often score >200.	Four factors account for 73% of the variance on the CDS: anomalous body experience; emotional numbing; anomalous subjective recall; and alienation from surroundings	Developed by Mauricio Sierra and German E. Berrios. Sierra M, Berrios GE. The Cambridge Depersonalization Scale: a new instrument for the measurement of depersonalization. *Psychiatr Res*. 2000;93(2):153–164.
Comprehensive Self-Report (Diagnostic Measures)				
Multidimensional Inventory of Dissociation (MID)[a] There is also an MID for adolescents.	Diagnostic self-report for dissociative disorders, PTSD, somatic symptom disorders, borderline personality disorder features	218 items self-report rating how often someone experiences symptoms from 0 ("never") to 10 (always); only dissociation measure that has validity scales; 30–90 minutes. Scoring program available (see below)	The MID demonstrated internal reliability, temporal stability, convergent validity, discriminant validity, and construct validity. The MID also exhibited incremental validity over the DES.	Developed by Paul Dell, Ph.D. Dell PF. The multidimensional inventory of dissociation (MID): A comprehensive measure of pathological dissociation. *J Trauma Dissociation*. 2006;7(2):77–106.
Multi-scale Dissociation Inventory (MDI). Available at http://www.johnbriere.com/multiscale.htm	Multiscale measure of dissociation with a 5-point Likert format that generates six dissociative symptom subscales and a total dissociation scale. Normed on 444 trauma exposed individuals in the community	30 items; approximately 15 minutes to complete. Scales include disengagement; depersonalization; derealization; emotional constriction (numbing); memory disturbance; identity dissociation; and total dissociation score	The scales show their expected relationships to trauma history and to other known dissociation scales, including the DES. The MDI has very-good-to-excellent sensitivity and specificity.	Developed by John Briere, Ph.D. Briere J, Weathers FM, Runtz M. Is dissociation a multidimensional construct? Data from the Multiscale Dissociation Inventory. *J Trauma Stress*. 2005;18:221–232.
Clinician-Administered Dissociative States Scale (CADSS)	Clinician-administered inventory that measures current state dissociation; has 28 items rated by the interviewer from 0 ("never") to 4 ("extreme")	Administered by a clinician usually in the context of an experimental study, to assess symptoms of amnesia, depersonalization, and derealization. The CADSS has proven useful in pharmacologic studies (e.g., of the dissociative agent ketamine in the treatment of depression), and in military stress studies	The CADSS has good convergent validity with other measures of dissociation and high interrater and test–retest reliability.	Developed by J. Douglas Bremner. Bremner JD, Krystal JH, Putnam FW, et al. Measurement of dissociative states with the clinician-administered dissociative states scale (CADSS). *J Trauma Stress*. 1998;11:125–136.
Structured Clinical Interview for DSM-IV-TR Dissociative Disorders, Revised (SCID-D-R)- being updated for DSM-5	Semi-structured, clinician-administered interview; 276 questions on a 4-point scale; 1–3 hours, if dissociative symptoms are present; much less time for nondissociative patients. Widely seen as "gold standard" for research studies requiring a diagnosis, and for differential diagnosis clinical cases.	Assesses the presence and severity of amnesias, identity confusion and alteration, depersonalization, and derealization, and renders a DSM-IV-TR diagnosis for all five dissociative disorders and for acute stress disorder.	The SCID-D-R has good-to-excellent interrater and test–retest reliability and well-established validity in numerous studies. It has been translated into at least a dozen languages.	Developed by Marlene Steinberg, M.D. Steinberg M. The Structured Clinical Interview for DSM-IV Dissociative Disorders-revised (SCID-D-R). American Psychiatric Press; 1994.
Dissociative Disorders Interview Schedule (DDIS). Has been updated for DSM-5. The DSM-5 version is available at http://www.rossinst.com/ddis.html	Clinician-administered interview 132 items. Strict yes–no format for all questions; may limit utility in situations where feigning is a concern. Requires approximately 30–60 minutes to conduct with dissociative identity disorder patients.	Inquires about, in addition to dissociative symptoms, child abuse history, substance abuse, paranormal experiences, and symptoms of major depressive, somatoform, and borderline personality disorders; yes–no format limits utility for diagnosis of borderline personality disorder	Except for depersonalization disorder, inter-rater reliability is acceptable, and convergent validity includes strong correlations with the DES, the SCID-D, and clinical diagnoses of dissociative disorders.	Developed by Colin Ross, M.D. Ross CA, Heber S, Norton GR, Anderson D, Barchet P. The dissociative disorders interview schedule: A structured interview. *Dissociation*. 1989;(2):169–178.

[a] DES, DES-T, ADES, CDC, SDQ, MID 6 and MID scoring program, and DES scoring program available from International Society for the Study of Trauma and Dissociation. Available at https://netforum.avectra.com/eweb/DynamicPage.aspx?Site=ISSTD&WebCode=SCREENING

DSM-5 (and TR) Dissociative Disorders and the DSM-5 (and TR) Trauma- and Stressor-Related Disorders

The DSM-5 made major changes to the classification of the posttraumatic and dissociative disorders that were retained in DSM-5-TR. PTSD, ASD, adjustment disorder, and several other conditions were placed in a separate category of trauma- and stressor-related disorders. Because of the robust relationship between dissociative disorders and antecedent trauma, dissociative disorders were considered for inclusion in the trauma- and stressor-related category. However, unlike the disorders in this new category, the diagnostic criteria for dissociative disorders do not require exposure to a stressor or a traumatic event as a diagnostic criterion.

Nonetheless, to show the strong relationship between dissociation and trauma, the DSM-5 (and TR) chapter on dissociative disorders is placed immediately after that of the trauma- and stressor-related disorders. This relationship was explicitly described in the introduction to the dissociative disorders section. The DSM-5 (and TR) diagnostic criteria for PTSD and ASD specifically describe flashbacks as dissociative. The PTSD criteria include a *dissociative subtype,* characterized by depersonalization and/or derealization in response to posttraumatic reminders, often, but not invariably, alternating with the typical intrusive symptoms that characterize nondissociative forms of PTSD. The ASD diagnostic criteria include dissociative symptoms such as depersonalization, derealization, altered time sense, and dissociative amnesia. Recent studies suggest that there may also be a dissociative subtype of ASD. The World Mental Health Survey data from 25,018 respondents from 16 high, middle-high, and low/lower-income countries found a 12-month prevalence of the dissociative subtype of PTSD in 14.4% of individuals with PTSD.

As will be discussed later in this chapter, additional research has begun to clarify the relationship between the dissociative subtype of PTSD and other dissociative disorders.

Changes to the Dissociative Disorders in DSM-5 (and TR)

Significant alterations to the dissociative disorders' criteria in DSM-5 (retained in DSM-5-TR) include changes to the definition of dissociation, the diagnostic criteria for dissociative identity disorder, dissociative amnesia, and depersonalization disorder, and the examples included under other specified dissociative disorders.

The Definition of Dissociation

The definition of dissociation in DSM-5 changed considerably from prior editions of the DSM. The definition encompassed phenomena that may be relatively acute—such as an episode of dissociative amnesia that is recognized shortly after a soldier is removed from combat—as well as childhood-onset complex dissociative disorders that are frequently chronic, like dissociative identity disorder. As part of the latter, individuals may report or display a plethora of dissociative symptoms, including subjective self-division, chronic complex dissociative amnesia, dissociative fugue, overlap and intrusion among self-states, and depersonalization and derealization, among others. Also, the definition of dissociation had to encompass culturally diverse experiences of pathologic dissociation, such as pathologic possession (see below).

As described in the introductory section, the DSM-5 (and TR) definition states that dissociation can affect the normal integration of every aspect of psychobiological functioning. Also, dissociative symptoms can be classified as "negative" and "positive," similarly to the definition of psychotic symptoms. The definition specifies that dissociation can be experienced not only as an absence, failure, or lapse, such as an awareness of dissociative amnesia (a negative symptom). But it can also be experienced as a disruption, interruption, and/or discontinuity suffered as recurrent, jarring, involuntary intrusions into executive functioning and sense of self. For instance, several minutes of rapid shifting of dissociative identity disorder self-states (a positive symptom). Negative and positive dissociative symptoms may occur sequentially. For example, after several minutes of shifting states, the individual with dissociative identity disorder may not recall what has just occurred.

DSM-5 (AND TR) DIAGNOSTIC CRITERIA FOR THE DISSOCIATIVE DISORDERS

Dissociative Identity Disorder

Studies of the phenomenology of over 1,000 dissociative identity disorder individuals studied mostly in the United States and Europe with the Structured Clinical Interview for DSM-IV-TR Dissociative Disorders (SCID-D-R), the Dissociative Disorders Interview Schedule (DDIS), or the Multiaxial Inventory of Dissociation (MID) found that over 90% of these individuals endorsed the following symptoms, in order of frequency: amnesia (100%); subjective self-alteration (98%); symptoms of overlap, interference, and intrusion among self-states, often characterized as variants of the FRSs of psychotic individuals (94%); awareness of the voices of self-states that are usually experienced within the mind (92%); and spontaneous trance behaviors (91%). Less commonly, dissociative identity disorder individuals described somatoform symptoms (83%), depersonalization (77%), and derealization (63%). In addition, these studies found that, unless the person was in crisis, only 15% of dissociative identity disorder individuals regularly showed clear-cut switching behaviors between distinctly differentiated self-states.

These and other studies found that over 40% of DSM-IV-TR dissociative disorder diagnoses were for dissociative disorder NOS, most commonly for Example 1 of the dissociative disorders NOS criteria. Clinicians familiar with dissociative identity disorder patients often used this NOS example *as a specific diagnosis*: a disorder with the clinical features, course, comorbidities, antecedent developmental trauma history, and treatment response of dissociative identity disorder but without “distinct” identities that take “control” of the person’s behavior through a process of “switching,” often a subtle process in and of itself. This diagnostic option has been codified in ICD-11 as *partial dissociative identity disorder*; a controversial category described later in this section.

DSM-5 diagnostic criteria for dissociative identity disorder moved completely away from the notion of relatively stable (“enduring”) personality states that switch (“take control”) of behavior to a more user-friendly enumeration of the distinct subjective differences seen in dissociative identity disorder presentations. These can include marked discontinuity in the subjective sense of self or the self as an agent of behavior. In addition, the person may experience and/or display discontinuities in mood and affect; manifest behavior; apparent state of consciousness; availability of autobiographical, procedural, and/or semantic memory; perception of the environment; cognitive schema; and/or sensorimotor functioning. The latter may involve conversion symptoms, especially in some non-Western cultural contexts (see the section on Diagnosis and Clinical Features). Moreover, DSM-5 criteria allowed the diagnosis to be made based on reports by the patient or others of the self-state discontinuities, not only observed by the clinician of these states shifts. The DSM-5 diagnostic criteria for dissociative identity disorder also modified the amnesia criterion (B) based on phenomenologic studies. Changes in the amnesia criterion emphasize the complexity of reported memory disturbances and that patients often report amnesia for ongoing experience (see the section on Diagnosis and Clinical Features).

It remains a matter of clinical judgment just how “distinct” personality or possession states have to be or how “marked” the discontinuities in the sense of self and agency to meet the A criterion for dissociative identity disorder. Despite the use of terms like “distinct” and “marked” in the criteria, all the data on dissociative identity disorder supports the subtlety of the symptoms of this disorder. The *distinctness* of self-states does not necessarily refer to the secondary elaboration of the states into dramatically different manifest presentations (see the section on Dissociative Identity Disorder). Rather, the term refers to the extent of compartmentalized information manifested by the identity states, accompanied by a subjective sense of self and a capacity for behavioral enactment. Similarly, the term *marked* refers to the disruptiveness of the symptoms to the person, not necessarily to the dramatic presentation of dissociative symptoms.

In addition, after an extensive review of the cultural literature, the DSM-5 (and TR) integrated the DSM-IV-TR dissociative disorder NOS example concerning an experience of pathologic possession into the diagnostic criteria for dissociative identity disorder. (Non-possession dissociative trance remains as one of the examples for other specified dissociative disorder.) This decision was based on the phenomenologic similarity between dissociative pathologic possession states and dissociative identity disorder; their shared dissociative origin; the coexistence in the same patient of identities attributed to either possessing entities or to endogenously generated self-states; the occurrence of amnesia for the experiences during the possession experiences in pathologic possession; and shared risk factors, such as high rates of lifetime traumatic experiences. While pathologic possession states are often described as more transient and episodic than those of non-possession dissociative identity disorder, further research is needed to assess whether these presentations of pathologic possession represent the overt appearance of phenomena that persist in more subtle form over time. As described above, most non–possession-form dissociative identity disorder individuals may have subtle presentations of core dissociative identity disorder symptoms and/or may have a relapsing/remitting course and appear to be relatively asymptomatic unless experiencing a clinical decompensation.

DSM-5 and DSM-5-TR carefully elucidate that not all presentations of possession meet the criteria for dissociative identity disorder. In addition to meeting criterion A (description of dissociative symptoms), individuals with a presumptive presentation of possession-form dissociative identity disorder must also meet criterion B (amnesia), criterion C (the possession state must be pathological, that is, lead to distress or impairment) and criterion D (not a normal part of a broadly accepted cultural or religious practice). Most instances of possession experience in the world are normal, embedded in spiritual activities, or do not entail dissociative experiences, functioning instead to communicate distress in a spiritual language (e.g., as a cause of any number of physical or mental symptoms).

The addition of pathologic possession to the DSM-5 (and TR) dissociative identity disorder criteria indicates that, around the world, dissociative identity disorder presents in two main forms: non–possession-form or possession-form. These are not mutually exclusive: individuals in one cultural setting may experience either form, or a single individual may experience both forms. Typically, in non–possession-form dissociative identity disorder, the personality states are experienced as *internal* aspects of the person. In possession-form dissociative identity disorder, by contrast, the alternative identities are typically experienced as *external* possessing agents, usually of supernatural or spiritual origin (e.g., demons, spirits). The main contrast between these two forms of the disorder appears to be shaped by the patient's sociocultural milieu. That is, possession-form dissociative identity disorder arises in contexts where possessing agents have ontologic validity (are considered "real"), such as in diverse traditional South Asian cultures or some fundamentalist religions in the United States. By contrast, non–possession-form dissociative identity disorder draws its alternative identities from individualistically focused experiences, such as stages of life (e.g., childhood) or symbolic, trauma-based intrapsychic dynamics (e.g., "protector," "sexualized") (see the section on Dissociative Identity Disorder).

Despite their phenomenologic similarities, several differences between dissociative pathologic possession states and non–possession-form dissociative identity disorder keep open the question of whether, as in ICD-11, pathologic possession states should be classified as a distinct nosologic category from a dissociative identity disorder. Most cases of "spirit possession" in two community studies in Africa did not report amnesia for the possession state. ICD-11 would include amnestic and nonamnestic cases under a single category of possession trance disorder, whereas they would be separated into distinct diagnoses in DSM-5-TR. Many individuals in these African studies also described alterations of consciousness consistent with ICD-11 trance disorder that predated the possession trance, linking the two disorders that were separated in DSM-5. Finally, very high rates of full therapeutic improvement have been reported for individuals with pathologic possession episodes participating in traditional healing, in contrast with the gradual improvement typical of dissociative identity disorder, possibly suggesting that the two conditions differ in severity. However, scant details on the course of these pathologic possession episodes are available to corroborate these claims of therapeutic success.

One pattern is so common as to be almost prototypical: presentations of dissociative possession states that are unwanted and cause distress only to evolve, typically over a period of months, into experiences that are willed, controlled, and nondistressing, usually as functioning traditional healers. It remains unclear whether these conditions meet the criteria for dissociative identity disorder or should be considered transient episodes of distress that meet exclusion criterion D because they fulfill cultural expectations for broadly accepted cultural or religious practices of spiritual healing. This judgment may be more readily made in retrospect; the level of distress during the initial possession episodes of "healers-in-training" may resemble that of individuals who do not improve so quickly and completely. Additional research is needed to resolve these nosologic questions. In some clinical settings, the inclusion of possession-form presentations in DSM-5 appears to have improved detection and treatment response for individuals who otherwise would meet dissociative identity disorder criteria but whose identification might be missed due to clinicians' lack of familiarity with this presentation.

The evolution of the diagnostic criteria for dissociative identity disorder in DSM-III, DSM-IIIR, DSM-IV-TR, and DSM-5 (identical in DSM-5-TR) is found in Table 20–2.

ICD-11 Diagnostic Requirements for Dissociative Identity Disorder and Partial Dissociative Identity Disorder. It is beyond the scope of this chapter to extensively discuss the ICD-11 diagnostic requirements for the dissociative disorders. Besides the nosologic categorization of dissociative presentations of pathologic possession states, the ICD-11 diagnostic requirements for dissociative identity disorder differ in important ways from the DSM-5 (and TR) criteria. ICD-11 also emphasizes specific types of identity states, notably ones that "function in everyday life" and that "take executive control" of the person's behavior and interactions with others. In addition, ICD-11 has created a new *partial dissociative identity disorder* diagnosis. This is characterized by "…[O]ne personality state [that] is dominant and normally functions in daily life (e.g., parenting, work), but is intruded upon by one or more non-dominant personality states (dissociative intrusions)." This process may lead to symptoms of intrusion, influence, overlap, and interference but these "non-dominant" self-states only take "executive control" for limited times under specific circumstances.

Table 20–2. Comparison of DSM-III, IIIR, ICD-10, DSM-IV-TR, DSM-5 Diagnostic Criteria for Dissociative Identity Disorder

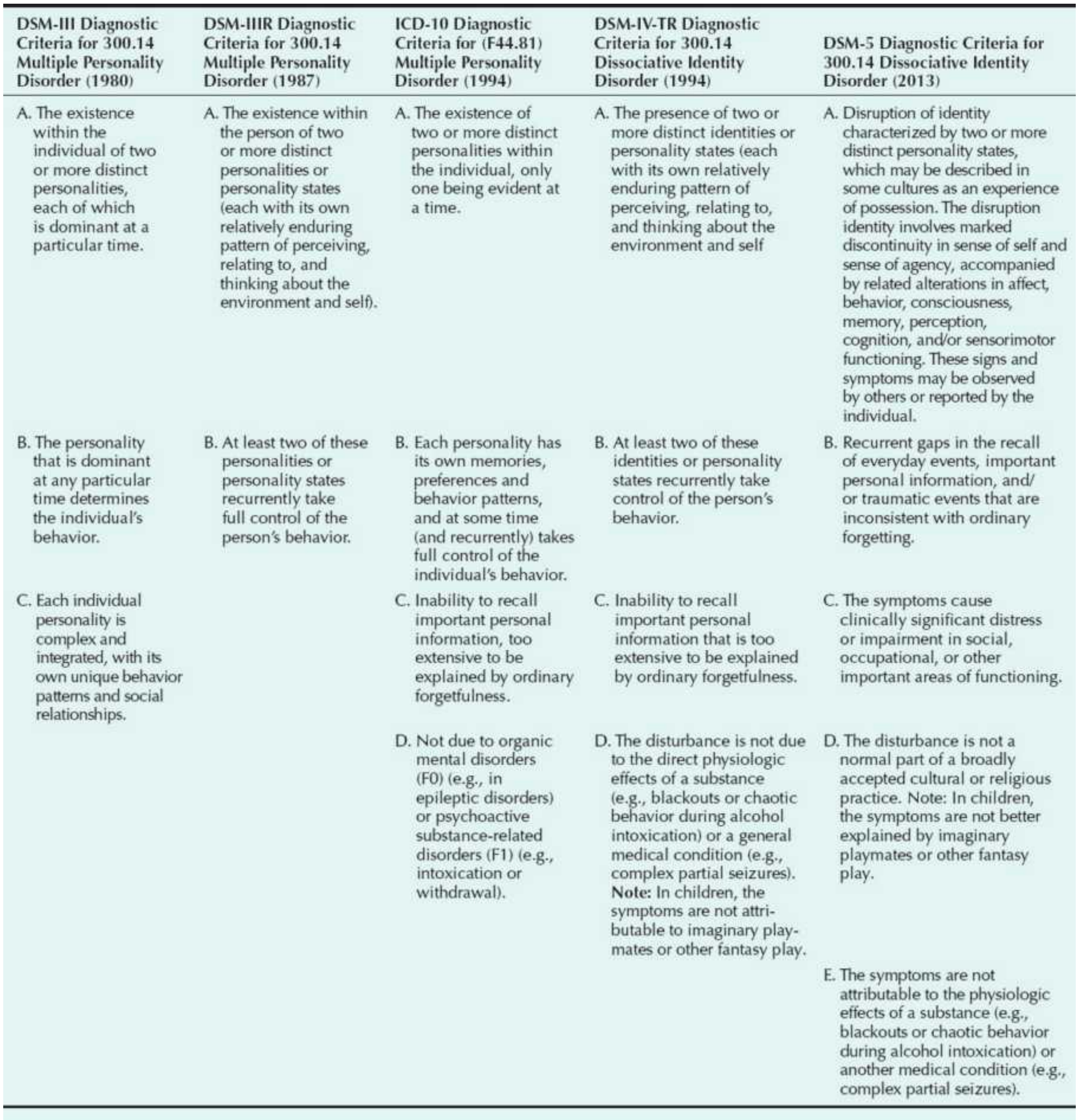

DSM-III Diagnostic Criteria for 300.14 Multiple Personality Disorder (1980)	DSM-IIIR Diagnostic Criteria for 300.14 Multiple Personality Disorder (1987)	ICD-10 Diagnostic Criteria for (F44.81) Multiple Personality Disorder (1994)	DSM-IV-TR Diagnostic Criteria for 300.14 Dissociative Identity Disorder (1994)	DSM-5 Diagnostic Criteria for 300.14 Dissociative Identity Disorder (2013)
A. The existence within the individual of two or more distinct personalities, each of which is dominant at a particular time.	A. The existence within the person of two or more distinct personalities or personality states (each with its own relatively enduring pattern of perceiving, relating to, and thinking about the environment and self).	A. The existence of two or more distinct personalities within the individual, only one being evident at a time.	A. The presence of two or more distinct identities or personality states (each with its own relatively enduring pattern of perceiving, relating to, and thinking about the environment and self	A. Disruption of identity characterized by two or more distinct personality states, which may be described in some cultures as an experience of possession. The disruption identity involves marked discontinuity in sense of self and sense of agency, accompanied by related alterations in affect, behavior, consciousness, memory, perception, cognition, and/or sensorimotor functioning. These signs and symptoms may be observed by others or reported by the individual.
B. The personality that is dominant at any particular time determines the individual's behavior.	B. At least two of these personalities or personality states recurrently take full control of the person's behavior.	B. Each personality has its own memories, preferences and behavior patterns, and at some time (and recurrently) takes full control of the individual's behavior.	B. At least two of these identities or personality states recurrently take control of the person's behavior.	B. Recurrent gaps in the recall of everyday events, important personal information, and/ or traumatic events that are inconsistent with ordinary forgetting.
C. Each individual personality is complex and integrated, with its own unique behavior patterns and social relationships.		C. Inability to recall important personal information, too extensive to be explained by ordinary forgetfulness.	C. Inability to recall important personal information that is too extensive to be explained by ordinary forgetfulness.	C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
		D. Not due to organic mental disorders (F0) (e.g., in epileptic disorders) or psychoactive substance-related disorders (F1) (e.g., intoxication or withdrawal).	D. The disturbance is not due to the direct physiologic effects of a substance (e.g., blackouts or chaotic behavior during alcohol intoxication) or a general medical condition (e.g., complex partial seizures). **Note:** In children, the symptoms are not attributable to imaginary playmates or other fantasy play.	D. The disturbance is not a normal part of a broadly accepted cultural or religious practice. Note: In children, the symptoms are not better explained by imaginary playmates or other fantasy play.
				E. The symptoms are not attributable to the physiologic effects of a substance (e.g., blackouts or chaotic behavior during alcohol intoxication) or another medical condition (e.g., complex partial seizures).

Building on the psychobiological research at NIMH in the 1980s, the extensive phenomenologic studies that led to the DSM-5 diagnostic criteria showed that the naturalistic condition of dissociative identity disorder is one of multiple, simultaneously overlapping states, more like the ICD-11 partial dissociative identity disorder diagnosis. A recent study of 169 experienced clinicians who assessed the impact of the change from DSM-IV-TR to DSM-5 diagnostic criteria for dissociative identity disorder found that diagnostic accuracy was improved in DSM-5 by removal of the requirement that self-states take control of behavior in any way.

Most presentations falling under the rubric of partial dissociative identity disorder in ICD-11 would thus qualify for a dissociative identity disorder diagnosis in DSM-5 (and TR). The new ICD categorization is based on the construct described by structural dissociation theory. Here it is postulated that the "personality" "splits": into secondary and tertiary "structural dissociation" because of traumatic experiences of various degrees of severity, justifying separate diagnoses of partial and full dissociative identity disorder (see sections on Etiological Theories of Dissociation and Dissociative Identity Disorder for a full discussion). Nosologically, the "partial" rubric provides a specified diagnosis for patients who meet DSM criteria for dissociative identity disorder except for the amnesia criterion, now conceptualized under the "other" specified dissociative disorder category in DSM-5 (and TR). It remains a question for research whether the construct of partial dissociative identity disorder will allow clinicians to better identify subtle presentations of dissociative identity disorder. Alternatively, the division of dissociative identity disorder into two categories ("partial" and "full") may prove confusing, as they suggest different severity. The DSM-5(and TR) diagnostic criteria were designed to decrease the high rate of diagnosis of DSM-IV-TR dissociative disorder NOS Example 1. The data on which this change was based did not find differences in symptoms, course, or treatment outcome between DSM-IV-TR dissociative identity disorder and most cases of Example 1 dissociative disorder NOS.

Structural dissociation theory is not accepted by many dissociation/dissociative disorder experts. Critique includes whether its language for mental processes is overly simplified, concrete, and reified, without sufficient acknowledgment that these are metaphorical constructs (e.g., "dissociated parts of the personality," "retraction of the field of consciousness"); that it only describes mental "structure" without sufficient attention to process (e.g., dimensional aspects of dissociation); that it reasons backward from adults with dissociative identity disorder in creating an overarching etiological theory of dissociative, trauma- and stressor-related disorders, somatic symptom disorders, and borderline personality disorder; and that it conflates secondary structuring of dissociative identity disorder self-states (e.g., culturally accepted roles in

Western society like “functioning in everyday life”) with the basic phenomenology of self-states (see section on Discrete Behavioral States Theory and Dissociative Identity Disorder).

Other differences between ICD-11 and DSM-5-TR include that, after DSM-III, the subsequent DSM diagnostic criteria abandoned the notion of “executive control” by specific self-states. Due to the overlap/intrusion symptoms, “control” typically involves a summation of self-states that coordinate and cooperate versus self-states in conflict with these and/or other self-states. It is the internal conflicts that typically lead to increased symptoms and life difficulties. Dissociative identity disorder self-states tend to be organized not as a collection of separate states but into complex, dynamic self-state systems.

Thus, it is difficult for clinicians to define “control” of behavior in dissociative identity disorder other than by detailed moment-to-moment and/or exhaustive retrospective assessment of the subjective states of mind of the individual in specific situations. Even this is limited by the willingness of the individual patient to be forthcoming—or even have adequate awareness of what has led to a particular situation. Also, this exegesis is affected by the experience of the clinician in tracking state shifts and understanding the trauma-based psychodynamics that underlie dissociative identity disorder. Finally, the ICD-11 terms “control,” “executive control,” and “dominant”/ “non-dominant” self-states are difficult to operationally define (e.g., what is “executive control” in bipolar disorder?).

Dissociative Amnesia

There are few systematic studies of dissociative amnesia in the literature. Studies of severe PTSD in traumatized groups have shown very high rates of the PTSD dissociative amnesia criterion. Also, many studies have documented delayed recall of trauma and/or the association of autobiographical memory deficits with trauma. Based on close reading of these studies, many of the subjects would meet the diagnostic criteria for dissociative amnesia. In DSM-5-TR, dissociative amnesia is defined in terms of an alteration of *autobiographical memory* that can be either recent or remote but that typically presents as retrospective gaps and/or subjectively fragmented memory for life history. In some very severe cases of generalized dissociative amnesia, individuals have been shown to lose semantic and procedural memory as well (e.g., what is a contact lens and how to use it). Also, the criteria now define the most common types of dissociative amnesia: localized, selective, or generalized dissociative amnesia (see Table 20–3 for types of dissociative amnesia). In ICD-11, following the diagnostic requirements in prior ICD editions, the autobiographical memory deficits are typically characterized by recent memory deficits associated with adverse life events, personal or interpersonal conflicts, or stress; more severe traumatic circumstances are associated with more persistent and refractory amnesia.

Dissociative Fugue as a Specifier of Dissociative Amnesia. There is little systematic literature on dissociative fugue, and the best studies remain detailed descriptive case series from the 1940s and 1950s. Both DSM-5-TR and ICD-11 agree that there are insufficient data to validate dissociative fugue as a disorder distinct from dissociative amnesia or dissociative identity disorder. Prior DSM versions of dissociative fugue included a criterion of “assumption of a new identity (partial or complete).” In DSM-5-TR and ICD-11, the development of a “new identity” during a fugue episode would meet diagnostic criteria for dissociative identity disorder. Other individuals with dissociative fugue do not differ clinically from those who meet diagnostic criteria for generalized dissociative amnesia other than for their peripatetic behavior.

Diagnostic Criteria for Depersonalization/Derealization Disorder

As opposed to DSM-IV-TR, the ICD-10, ICD-11, and DSM-5-TR agree that this disorder is typified by both depersonalization and derealization. Also, ICD-11 now includes depersonalization/derealization disorder among the dissociative disorders. This is based on research showing that depersonalization and derealization are prevalent in other dissociative disorders such as dissociative identity disorder. Depersonalization and derealization are quintessential responses to acute trauma (*peritraumatic dissociation*) and have been reported as highly prevalent in those subjected to motor vehicle accidents, emotional/verbal abuse, and imprisonment, among other exposures (see the section on Epidemiology). The dissociative subtype of PTSD is based on the symptoms of depersonalization and/or derealization in response to trauma cues.

Table 20–3.
Types of Dissociative Amnesia

Localized	Inability to recall events related to a circumscribed period of time
Selective	Ability to remember some, but not all, of the events occurring during a circumscribed period of time
Generalized	Failure to recall one's entire life
Continuous	Failure to recall successive events as they occur
Systematized	Amnesia for certain categories of memory, such as all memories relating to one's family or to a particular person

Several authorities have proposed that depersonalization is a "hard-wired" inhibitory response—part of the human psychobiological survival system—that allows for attenuation of anxiety and psychobiological hyperarousal to help preserve adaptive behavior and conservation of physical resources under conditions of threat and danger. Disorders related to depersonalization and derealization may represent the generalization and persistence of this response outside the context of immediate threat, just as PTSD fear reactivity can be understood, in part, as a generalization of the adaptive conditioned fear response beyond the immediate context where it was generated.

Studies of two large cohorts have persuasively documented a consistent nosology for DSM-5-TR depersonalization/derealization disorder. These studies described consistent age of onset, acute and remote antecedents, as well as distinct symptom clusters, cognitive organization, course, treatment response (primarily *lack* of treatment response), and gender ratio. Factor analysis from these studies suggests that symptoms of depersonalization/derealization disorder are characterized by four to five factors: numbing, unreality of self, unreality of other, temporal disintegration, and perceptual alterations, including micropsia/macropsia, out-of-body experiences, and feelings of unreality of objects and of one's own body.

Mr. J, a high-functioning professional without a psychiatric history, described that, while a college student in a geology class, he was involved in a field trip where he was excavating a small cave. He was working inside the cave when a landslide blocked the entrance. Although there was complete darkness in the cramped space, he noticed, with an unemotional sense of detachment, that he could observe himself as if from a distance. He reported having no sense of time. He did not experience fear or concern about his situation. He noted with interest the sounds of his classmates and professor shouting and digging at the front of the cave. He recalls that, as soon as he saw a beam of light come into the cave, he felt that he had re-entered his body, and suddenly felt anxiety and pain, and noticed bleeding where the rocks from the avalanche had struck his face. He still suffers from a partial facial nerve paralysis due to this event. He denies any subsequent depersonalization or derealization experiences.

Other Specified Dissociative Disorder

In DSM-5-TR, this category includes four examples: (a) chronic and/or recurrent mixed dissociative symptoms, such as identity disturbance associated with less-than-marked discontinuities in the sense of self and agency or alterations of identity or episodes of possession in an individual who reports no dissociative amnesia. This aligns somewhat with the ICD-11 partial dissociative identity disorder diagnosis; (b) identity disturbance in response to *coercive persuasion*; (c) *acute dissociative reactions to stressful events*, typically encompassing brief episodes of mixed dissociative symptoms that occur in the context of trauma or extreme stress. These are distinct from brief episodes of monosymptomatic dissociative disorders (e.g., dissociative amnesia) by combining a mix of dissociative symptoms but without the identity alteration of dissociative identity disorder. These episodes show substantial cultural variation. Some presentations of cultural concepts of distress included in DSM-5-TR (e.g., *ataque de nervios*) may meet the criteria for this condition; and (d) dissociative trance characterized by acute narrowing or complete loss of awareness of immediate surroundings manifesting as profound unresponsiveness or insensitivity to environmental stimuli. The trance state may be accompanied by minor stereotyped behaviors (e.g., finger movements) of which the individual is unaware and/or that (s)he cannot control, as well as transient paralysis or loss of consciousness. As in prior editions of the ICD, ICD-11 places conversion disorders in the dissociative disorders section as a dissociative disorder of movement, sensation, or cognition. These conditions are included in the DSM-5-TR somatic symptom disorders section as functional neurologic symptom disorder (conversion disorder).

VALIDITY OF THE DISSOCIATIVE DISORDERS

As noted in the section on History, despite substantial systematic research about dissociation, the dissociative disorders and their relationship to trauma, debate continues about the validity of the dissociation concept, dissociative amnesia, and dissociative identity disorder.

Validity of the Dissociative Identity Disorder Construct

The debate about the existence of dissociative identity disorder has waxed and waned for more than a century. Sufficient data on pathologic dissociation, in general, and on dissociative identity disorder itself have accrued to judge this condition by the same standards that are applied to the validity of other psychiatric diagnoses.

The validity of psychiatric diagnosis is not defined by a single, unitary construct. There are no absolute criteria, models, guidelines, or standards to determine whether a psychiatric diagnosis is “valid.” A summary of the literature on psychiatric diagnostic validity combines the following elements: the history of the disorder; a detailed clinical description of the disorder that is repeatedly and independently replicated (content validity); longitudinal studies of symptom stability over time; the existence of a syndrome (at least one criterion symptom has ≥50% probability of being accompanied by another); diagnostic criteria and validated diagnostic and screening measures (discriminant validity); laboratory studies and/or valid and reliable psychological tests consistent with the clinical picture (criterion validity); etiologic studies; family studies; treatment outcome studies; taxonomic analysis; and exclusion criteria (e.g., is not better explained by another disorder, does not cause distress, and/or is biased based on race/ethnicity, gender, or culture-related factors). Data supporting the validity of dissociative identity disorder will be discussed in-depth in later sections.

Dissociative identity disorder, described throughout the 19th and early 20th-century psychiatry literature, is one of the oldest diagnoses in the literature. In the early to mid-20th century, for a variety of reasons, dissociative identity disorder was subsumed under other constructs and thought to be, at best, vanishingly rare (see the section on History). Later 20th-century reviews of cases and case series supported the construct and historical validity of dissociative identity disorder. Subsequently, clinical phenomenologic studies of dissociative identity disorder—including from North and South America, Europe, the Middle East, and Asia—document the presence of a core dissociative psychopathology in dissociative identity disorder. Case series and prospective epidemiologic studies document both symptom stability over time as well as syndrome predictability. The core features of the disorder have been documented across the life cycle: in childhood, adolescence, adulthood, and geriatric cases.

Construct or discriminant validity requires that the disorder can be differentiated from other disorders. Structured interviews such as the SCID-D-R and comprehensive symptom-screening measures such as the MID show that dissociative identity disorder patients can be discriminated from normal individuals and patients with Cluster B personality disorders, PTSD, psychotic disorders, and mood disorders, among others. Likewise, psychological test batteries and experimental cognitive and psychophysiological studies have discriminated dissociative identity disorder patients from those with other disorders, including simulators.

Criterion-related validity requires that laboratory tests or reliable psychological tests be consistent with the defined clinical picture. This stipulation is met by the reliable and valid diagnostic interviews and scales increasingly used in dissociation research. Psychological assessment protocols show a unique psychological profile that discriminates dissociative identity disorder from schizophrenia, depression, borderline and other personality disorders, nondissociative forms of PTSD and ASD, feigning individuals, and normal controls, among others.

Studies using neurobiologic and psychophysiological measurements of dissociative identity disorder subjects compared to simulating controls have invariably shown significant differences between the real subjects and simulators. Also, actual dissociative identity disorder subjects generally show unique patterns of neurobiologic and/or psychophysiological variables uncharacteristic of other psychiatric disorders. When assessed in rigorous hypnotizability research protocols, across studies, individuals with dissociative identity disorder score significantly higher than patients with borderline personality disorder, PTSD, anxiety disorders, mood disorders, schizophrenia, and nonclinical controls.

Etiologically, as discussed in detail in subsequent sections, every study of dissociative identity disorder in children, adolescents, and adults documents the highest rates of childhood maltreatment and adversity compared to any other clinical group. Some studies substantiate trauma reports with records from hospitals, schools, child protective services, and other agencies, as well as confirmation by siblings or other relatives. Small clinical case series demonstrate multigenerational dissociative disorders, including dissociative identity disorder, as well as substantiated multigenerational childhood maltreatment.

As described later in this chapter, every retrospective and prospective treatment outcome study utilizing the triphasic trauma treatment model for dissociative identity disorder has shown significant patient improvement in many important symptom and functional domains. Randomized controlled clinical trials are currently in development. Taxon analysis has defined a categorically distinct “pathologic dissociation” taxon, encompassing severe dissociative symptoms characteristic of the most severe, complex dissociative disorders, including dissociative identity disorder. From the syndrome perspective, studies find that the conditional probability of one DES-T symptom predicting another is 100%. Based on the DES-T, regardless of base rates across samples, the conditional probabilities among amnesia, depersonalization, derealization, identity alteration/confusion, and hearing voices were 0.92, 0.92, 0.93, 0.76, and 0.94, respectively. Finally, these patients are highly distressed, with substantial rates of suicide attempts and nonsuicidal self-injury. If anything, there is stigma and bias *against* this diagnosis in the professional community.

In summary, these unique findings cannot be better explained by another disorder. Dissociative identity disorder satisfies all of the widely accepted standards for psychiatric validity and should be regarded as a legitimate disorder requiring an informed diagnostic and treatment approach.

Validity of Dissociative Amnesia

Studies that address the validity of dissociative amnesia have primarily centered on the existence of the phenomenon, not on the study of the clinical disorder. As noted previously, *all* studies that have attempted to investigate the presence of amnesia for traumatic events have reported that it occurs. However, recent studies using the MID found that individuals with dissociative amnesia can be distinguished from those with dissociative identity disorder. The former described retrospective autobiographical memory deficits. The latter demonstrated *both* retrospective autobiographical memory deficits and contemporaneous memory lacunae for current, ongoing activities, well-learned skills, and acquisition of possessions. International case series of individuals with dissociative amnesia document the phenomenology, psychological and neuropsychological profile, and neurobiology of dissociative amnesia, as well as a consistent history of severe childhood and/or adult interpersonal trauma (see sections on Neurobiology and on Dissociative Amnesia).

Alternative hypotheses about trauma and memory claim that trauma virtually always causes an indelible inscribing of memory in the brain and that dissociative amnesia cannot exist. This debate has also focused on the existence of hypothesized psychological mechanisms for the reported amnesia, such as repression or dissociation. Critics suggest that more mundane processes, such as failure to report, cognitive avoidance, or ordinary forgetting, might account for the apparently delayed recall, although it is difficult to understand how these alternative mechanisms could account for the repeatedly documented findings of acute or delayed amnesia resulting from trauma such as combat or rape. Also, dissociative amnesia remains a criterion symptom for both PTSD and ASD. Most critics of the dissociative amnesia concept do not appreciate that breakdown of dissociative amnesia generally results in intrusive PTSD symptoms: intense, highly disturbing perceptional, somatosensory, and emotional intrusions, and/or full flashbacks, with a fragmented narrative, not linear, quotidian autobiographical memory. The construct *of state-dependent learning and memory*, as shown in international phenomenologic, neuropsychological, and neurobiologic studies of dissociative amnesia, provides support for the view that trauma can be *both* registered in the brain in some form and can be relatively inaccessible to retrieval.

Research shows that individuals who report problems with the recall of trauma may describe several subjective processes, including difficulties recalling some or all traumatic life events and/or life periods in which the trauma occurred. Others report deliberate attempts to avoid thinking about the traumatic events due to subjective distress or shame. Others describe a lack of ability to appreciate the abusive nature of trauma until they were older and became aware of the devastating effects of trauma on their lives. In clinical practice, *all* these mechanisms may account for aspects of dissociative amnesia in an individual patient.

Studies have corroborated the traumatic events recalled by individuals with dissociative amnesia. Types of corroboration include contemporaneous records from medical, social service, and diocesan church sources; verbal or written admission by perpetrators; criminal convictions or civil judgments against perpetrators; reports by family members; and information from witnesses or other victims of the same perpetrator. Meta-analyses show no significant differences in the accuracy of delayed recall of trauma compared to accounts of individuals who report continuous memory for trauma. Contrary to stereotype, studies show that delayed recall of trauma most frequently occurs outside of psychotherapy. Corroboration studies of delayed trauma memories also show that recall may be generally accurate, partially accurate, or completely confabulated. There are no significant differences in accuracy between delayed recall and reportedly “continuous” memories of traumatic events. Finally, the concept of dissociative amnesia for combat or other trauma was widely accepted—even by clinicians and researchers who later eschewed belief in it—until the issue led to litigation around childhood sexual abuse (see the section on History).

EPIDEMIOLOGY OF THE DISSOCIATIVE DISORDERS

Epidemiologic studies in North America, Europe, Africa, Asia, Australia, Latin America, and Turkey find that pathologic dissociation and dissociative disorders are common in samples of the general population, as well as in psychiatric inpatients, outpatients, emergency department patients, patients in alcohol and substance abuse treatment, and in special populations with high rates of early childhood trauma, such as women in prostitution. Also, studies of both clinical and community populations with histories of trauma show a subgroup with substantial rates of dissociation and dissociative disorders.

GENERAL POPULATION STUDIES OF DISSOCIATIVE DISORDERS

Table 20–4 shows the findings for the prevalence of DSM-IV-defined dissociative disorders in three representative general population samples ascertained by structured diagnostic inventories: Winnipeg, Canada, New York State, USA and a female sample in Sivas, a city in central Turkey. As shown in the table, the overall prevalence of dissociative disorders ranged from 9.1% to 18.3%. Based on these studies, one researcher has conservatively estimated the lifetime prevalence of dissociative disorders in the community at about 10%.

A random sample of 1,055 adults in Winnipeg, Canada, were given the DES, and a subsample of 502 respondents that did not differ significantly demographically from the larger sample was subsequently reevaluated with the DDIS. Findings include a left-skewed distribution on the DES, with only a small subset of the population 3% to 5% endorsing many forms of dissociative experience, consistent with the construct of *pathologic dissociation* and a categorical taxon (see the section on Theories of Dissociation). This group was distinct from all other diagnostic categories, with symptoms characteristic of dissociative identity disorder and DSM-IV-TR dissociative disorder NOS. Overall, 12.2% of study participants met DSM-IV diagnostic criteria for a lifetime diagnosis of a dissociative disorder.

There was a 9.1% 1-year prevalence of dissociative disorders in a longitudinal, community-based study of 658 adults in New York State. DSM-IV Dissociative disorder NOS was the most common diagnosis, with a prevalence of 4.3%. Dissociative disorders were associated with clinically significant impairment and, as in other studies, with comorbid mood, anxiety, and personality disorders. Individuals with dissociative identity disorder were the most impaired of the dissociative disorders group.

Table 20–4.
Prevalence of Dissociative Disorders in General Population Samples Ascertained with Structured Diagnostic Instruments

Study	Ross (1997)	Johnson et al. (2006)	Sar et al. (2007)
Measures	DES and DDIS	Adapted from DES, SCID, SCID-II, and GAFS	DDIS, SCID-PTSD, and SCID-II
Number of subjects	502	658	628 (female)
Diagnosis	Subjects (%)	Subjects (%)	Subjects (%)
Dissociative amnesia	6.0	1.8	7.3
Dissociative fugue	0	0	0.2
Dissociative identity disorder	1.3	1.5	1.1
Depersonalization disorder	2.8	0.8	1.4
Dissociative disorder not otherwise specified (NOS)	0.2	4.3	8.3
Dissociative disorder NOS with multiple personality states			4.1
Dissociative disorder NOS with indirect cues for personality states			2.4
Derealization without depersonalization			1.1
Dissociative trance disorder			0.6
All dissociative disorders	12.2	9.1	18.3

DDIS, Dissociative Disorders Interview Schedule; DES, Dissociative Experiences Scale; GAFS, Global Assessment of Functioning Scale; SCID, Structured Clinical Interview for DSM-IV-TR Dissociative Disorders; SCID-PTSD.

Data from Ross CA. Dissociative Identity Disorder: Diagnosis, Clinical Features, and Treatment of Multiple Personality. John Wiley & Sons; 1997:109; Johnson JG, Cohen P, Kasen K, et al. Dissociative disorders among adults in the community, impaired functioning, and axis I and II comorbidity. *Psychiatry Res*. 2006;40(2):131; Sar V, Akyüz G, Doğan O. Prevalence of dissociative disorders among women in the general population. *Psychiatry Res*. 2007;149:169.

In a representative sample of 628 women in Sivas, Turkey, assessed with the DDIS, the PTSD module of the SCID-I, and the SCID-II, 18.3% met the criteria for a lifetime DSM-IV diagnosis of a dissociative disorder. Dissociative identity disorder represented about 1.1% of the sample; like in the New York State study, dissociative disorder NOS was the most prevalent dissociative disorder at 8.3%, with about half of these partially meeting criteria for dissociative identity disorder. Diagnosis of a dissociative disorder was significantly associated with other comorbid conditions such as PTSD, major depression, somatization disorder, borderline personality disorder, and a history of attempted suicide.

Pathologic Dissociation in the General Population

There are two Finnish general population studies of dissociation: a 3-year prospective study of 2,001 adults and a cross-sectional study of 1,424 adolescents ages 13 to 18. The adult studies used the DES, the DES-Taxon scale (DES-T), the Beck Depression Inventory (BDI), the Somatoform Dissociation Questionnaire (SDQ), the Toronto Alexithymia Scale (TAS), and an instrument based on measures used in the Adverse Childhood Experiences (ACE) study (discussed later in this chapter). A subset of the original adult cohort (N = 1,399) was followed up 3 years later with the DES, DES-T, and BDI. The adolescent participants completed the Adolescent DES (A-DES) and measures of academic performance, competence, social relationships, and alcohol and drug use.

In the adult study, just as in the Winnipeg study, the investigators found a skewed distribution of the DES, with about 3.5% of the general population fitting the taxonomic pathologic dissociation construct. Pathologic dissociation was strongly associated with depression, suicidality, alexithymia, and alcohol abuse. SDQ scores increased in a graded fashion with each ACE: childhood intrafamilial violence, including childhood emotional, sexual, and physical abuse; familial alcohol abuse; and poor relationship between the parents. The combination of high SDQ scores and high DES scores differentiated a specific subgroup with higher depression and suicidality, as well as many life stresses and problems with work, finances, and health.

At 3-year follow-up, membership in the subgroup with high-dissociation scores was more stable than membership in the pathologic dissociation taxon. About a third of the original high-dissociation group still showed high-dissociation scores, but the other two-thirds had been replaced by an equal number of different individuals with DES scores above the cutoff of 20. However, only four of the original individuals who fit the pathologic taxon remained in the taxon, although the total number of individuals in the taxon remained the same (*n* = 39; 3.9% at 3 years). Improvement in depression and suicidality was associated with lower-dissociation scores, indicating that worsening dissociation may be a response to stressful events and greater levels of symptoms in several diagnostic domains. Problems with work performance and younger age also predicted consistently higher-dissociation scores.

In the adolescent sample, higher-dissociation scores were correlated with daily smoking, frequent use of alcohol, abuse of legal drugs, cannabis use, social isolation, and poor school performance. There was no assessment of maltreatment in these adolescents. In longitudinal studies of abused children smoking, alcohol and drug use, social isolation, and poor school performance are all associated with childhood adversity and maltreatment (see discussion of the ACE study below).

In the adults, higher-dissociation scores were associated with more ACEs and higher depression, suicidality, and life adversities. These findings are consistent with studies finding that, in many individuals, overt dissociative disorder symptoms occur with a remitting/relapsing course, only fully emerging during times of life stress, trauma, and/or adversity. These may be accompanied by increased PTSD symptoms, depression, substance abuse, and self-destructive and suicidal behavior—all associated with trauma.

An island-wide representative community study in Puerto Rico of 891 youth ages 11 to 17 examined the point prevalence of pathologic dissociation, defined by a cutoff on an 8-item version of the A-DES keyed to the pathologic taxon. As in the Finnish research, this study also found a skewed population distribution of dissociation, with nearly 53% of adolescents reporting an A-DES score of 0. Pathologic dissociation was reported by 4.9% of adolescents, with no significant differences by age or gender. In multivariate analyses, pathologic dissociation was associated with exposure to violence and elevated levels of physical abuse. Individuals exposed to multiple forms of victimization were significantly more likely to endorse pathologic dissociation.

Two community-based studies in Brazil examined the prevalence of dissociative disorders using standardized instruments. However, variations in inclusion criteria and clustering of specific disorders in reporting complicate cross-national comparisons. One representative community study of 3 large metropolitan areas in Brazil (*N* = 6,476) using a clinician-rated symptom screener based on DSM-III found a lifetime prevalence of 2.8% to 8.1% for "somato-dissociative disorders" depending on the urban area; however, the mixing of somatoform and dissociative categories complicates the interpretation of this finding. A different representative community sample in a large section of São Paulo (*N* = 1,464) examined ICD-10–defined dissociative disorders using the lay-administered Composite International Diagnostic Interview (CIDI), revealing a prevalence of 2.2% (lifetime), 0.5% (12 months), and 0.3% (1 month). However, ICD-10 differs substantially from DSM in the criteria for dissociative disorders.

Dissociative Identity Disorder

Across community-based studies, dissociative identity disorder had a prevalence of about 1% to 1.5% in the general population. In the New York and Turkish studies, DSM-IV-TR dissociative disorder NOS was the most frequent diagnosis, 4.3% and 8.3%, respectively. In the Turkish study, about half of these NOS individuals (4.1% of the whole sample) were described as similar to dissociative disorder NOS Example 1 in DSM-IV-TR, "Dissociative disorder NOS with multiple personality states" (see Nosology). Thus, the improved diagnostic precision of DSM-5-TR diagnostic criteria may increase the prevalence of dissociative identity disorder by decreasing prevalence of Example 1.

Dissociative Identity Disorder and Gender in the General Population. Developmental studies indicate that the ratio of female to male dissociative identity disorder cases steadily increases from 1 to 1 in early childhood to approximately 8 to 1 by late adolescence in North American samples. Proposed explanations include gender-related differences in the types, age of onset, and duration of maltreatment experienced by men and women; differences in clinical presentations (i.e., male cases are more likely to be missed); and the possibility that more male dissociative identity disorder cases end up in the criminal justice or alcohol and drug treatment systems, or both, rather than the mental health system.

However, this finding has not been replicated in Turkish studies of adolescents with dissociative identity disorder or in the WHO study of the PTSD dissociative subtype, where the male gender was more prevalent. The Finnish longitudinal study found a 1:1 ratio of male: female patients in the pathologic dissociation taxon. No significant gender differences were found in the Puerto Rico study of community-based adolescents. A study of the entire male and female population of Taiwanese prisons (*N*≈60,000) found that the most common diagnoses in *both* genders were among the ICD-10 dissociative/conversion disorders.

Sociocultural effects are implicated in these discrepant findings. Consistently in cross-national and cross-ethnic studies, men are less likely than women to seek mental health treatment. Due to clinical stereotypes, childhood maltreatment, especially sexual abuse, may be less frequently ascertained in men. Males may be even more reluctant than women to discuss childhood or adult rape. Traditional gender roles for men tend to increase the expression of externalizing behaviors compared to women, possibly skewing to higher rates of incarceration for traumatized men. Western/secular depictions of dissociative identity disorder are primarily based on media stereotypes, including popular first-person autobiographical accounts, movies, and television shows. The stereotypical dissociative identity disorder patient is a bright, creative, histrionic, sexualized White female, with theatrical switching among clearly defined, dramatically different personality states. Due to this inaccurate diagnostic framework, most female dissociative identity disorder patients are misdiagnosed, often for long periods of time. Males with dissociative identity disorder are even less likely to be correctly diagnosed. Media stereotypes depict men with this diagnosis as criminals, typically sexually sadistic serial killers. In the form of racism-based institutional betrayal by the mental health system—and despite well-known high rates of inner-city violence, trauma, and social adversities—men and women of color may be even *less likely* to be evaluated for trauma- and stressor-related and/or dissociative disorders. Men of color may be particularly profiled as antisocial and/or psychotic, leading to a lack of appropriate diagnosis and treatment, especially for trauma/dissociative disorders.

Epidemiology of Pathologic Possession Form of Dissociative Identity Disorder

Cases of pathologic possession with dissociative symptoms have been reported in many countries, including China, India, Turkey, Iran, the Republic of Singapore, Brazil, Puerto Rico, Mozambique, and Uganda. Some case series of dissociative identity disorder in the United States and Canada have described patients who experienced self-states as possessing entities or were judged as possessed by co-religionists and subjected to exorcisms or similar practices to attempt to expel the personality states. In these cases, the phenomenology of dissociative identity disorder was influenced by these individuals' participation in religious groups that emphasized normative possession (e.g., Pentecostals, Charismatics).

Two epidemiologic studies in conflict regions of Africa shed light on the prevalence of pathologic possession trance. A randomized household study (*N* = 941) of individuals ages 13 to 60 residing in districts of rural Mozambique particularly affected by civil war revealed a point prevalence of 11% for "spirit" possession; distress was reported in 85% of cases, and loss of personal consciousness during possession in 95%. In 44% of cases, possession was attributed to more than one spirit, a characteristic associated with higher morbidity. However, amnesia was not assessed, preventing the application of full DSM-5-TR criteria. In multivariable analyses, individuals suffering from spirit possession were more likely than nonpossessed respondents to be women, have physical symptoms, and have suffered the loss of a child; there was also a trend for higher endorsement of traumatic experiences on the Harvard Trauma Questionnaire.

A community-based study (*N* = 1,113) of youth ages 12 to 25 living in randomly selected households in areas of Northern Uganda also torn by civil conflict examined the 1-month prevalence of pathologic spirit possession using DSM-IV research criteria for dissociative trance disorder. These pathologic possession experiences were attributed to *cen*, considered the spirits of individuals murdered in the civil war. Nine percent of respondents reported symptoms of identity alteration attributed to possession; the same proportion also endorsed the amnesia criterion.

Four or more criteria linked to DSM-IV dissociative trance disorder were endorsed by 8.2% of the sample, suggesting a high likelihood of meeting DSM-5-TR criteria for possession-type dissociative identity disorder; 45% of these also met DSM-IV PTSD criteria. Compared to individuals with low likelihood, those with a high likelihood of pathologic dissociation were significantly more likely to be female and to have suffered various forms of war-related trauma. In multivariable analyses, symptoms of spirit possession were significantly associated with functional impairment, suicide risk, perceived discrimination, and physical complaints.

An in-depth case series of 119 pathologic possession cases in Uganda confirmed the association between pathologic spirit possession and traumatic exposure, revealing significantly higher traumatic reports on standardized measures compared with 71 randomly selected mentally healthy inhabitants of the same villages. Scores on these trauma scales were strongly correlated with symptoms of pathologic possession and with higher levels of dissociative symptoms as measured by the DES and the SDQ.

Earlier epidemiologic studies also confirm the elevated global prevalence of pathologic possession experiences. In rural Indian communities, the period prevalence of pathologic possession has been estimated at 0.97% (over 6 months) to 3.5% (over 1 year), depending on the region, the sample, and the method of assessment. The period prevalence of culturally congruent, nonpathologic possession trance contributed an additional 0.2% to 1.8% over 6 to 12 months in the same studies. In a representative community sample of 4,129 residents of rural Baluchistan (southeast Iran), the point prevalence of a local pathologic possession state (“Djinnati”) among female residents ages 15 to 60 was 1%; no men reported the condition. Individuals who were Djinnati had significantly higher mean scores on the DES than a random subsample of nonaffected study participants.

In the Turkish study of 628 women in Sivas (previously discussed in this chapter), 20.2% of the women described paranormal experiences (e.g., precognition; encountering poltergeists or ghosts), and 2.1% reported dissociative possession experiences. Women with a dissociative disorder reported all types of paranormal and dissociative possession experiences significantly more frequently than those without. One subgroup composing 7.8% of the sample accounted for 91.6% of the reported possession experiences. They were characterized by the most severe trauma in both childhood and adulthood; the most severe dissociative disorders (e.g., dissociative identity disorder and DSM-IV-TR dissociative disorder NOS); and multiple comorbidities commonly associated with trauma, such as major depression, PTSD, somatic symptom disorders, suicidal and self-destructive behavior, and features of borderline personality disorder. About half of the possession experiences were attributed to a dead or living person and the other half to supernatural entities (e.g., demon, jinn) or forces.

In the non–possession-form of dissociative identity disorder in secular psychiatric settings, patients frequently report the experience of “introject” self-states that literally embody other people in the patient’s past or current life. These are sometimes experienced as a benevolent person, but mostly they represent prior or current reported perpetrators. This phenomenon may be experienced as so ego-alien that the person reports that they *feel* possessed. However, these individuals usually do not believe that they actually are possessed. Future studies will need to address the question of belief in possession as opposed to the dissociative individual *feeling* possessed.

Epidemiology of Dissociative Amnesia and the Dissociative Fugue Subtype of Dissociative Amnesia

Dissociative amnesia, as defined by the DSM-IV-TR, has been reported in approximately 1.8% (1-year prevalence) to 6% (lifetime prevalence) of general population samples, with 7.3% of the general population sample of Turkish women in Sivas meeting diagnostic criteria for dissociative amnesia in their lifetime. In these studies, there were low rates of the *dissociative fugue subtype:* 0% in Winnipeg and New York State and 0.2% in Sivas. The low prevalence of dissociative fugue in general population studies was one of the factors contributing to the decision to change this from a separate diagnosis to a subtype of dissociative amnesia.

Autobiographical Memory Disturbances for Childhood and Trauma

Childhood autobiographical memory disturbances (CAMD) are a common presentation of dissociative amnesia patients, with selective, localized, and, often, systematic amnesia for early life history (see Table 20–4). This finding has empirical support in data from the ACE study, a collaboration between the Centers for Disease Control (CDC) and Kaiser Permanente HMO in San Diego, CA, involving over 17,000 employed patients of a

California HMO. It is one of the largest population studies to assess the relationship between childhood adversities and later life medical, psychiatric, and behavioral problems. In the ACE study, during a comprehensive physical examination, patients filled out a questionnaire about 11 different childhood adversities and about their current health status. This included questions about recurrent childhood contact with sexual, physical, and/or emotional abuse; physical or emotional neglect; witnessing domestic violence; the presence of a substance-abusing member of the household; parental mental illness; parental incarceration; and parental absence or loss. The ACE study documented that increasing numbers of ACEs were, in a stepwise fashion, associated with a variety of adverse mental, physical, and behavioral health outcomes (see Tables 20–5 and 20–6 for a list of ACEs and their relationship to a variety of clinical disorders and behavioral problems). Studies show that ACEs are synergistic, with childhood sexual abuse being the most malignant ACE, pairing most frequently with other ACEs, followed by physical abuse, neglect, and intimate partner violence

A recent study from Hong Kong by Hong Wang Fong, Colin Ross and colleagues found that, in 202 outpatients from outpatient mental health clinics, just as in Western samples, increasing numbers of ACEs predicted both psychoform and somatoform pathological dissociation, as measured by the DES, DES-T, and the SDQ.

As part of the ACE study, CAMD was ascertained in a cross-sectional subsample of 9,460 relatively healthy adults in relationship to 8 different ACEs. Participants were asked: “Are there any large parts of your childhood after age 4 that you cannot remember?” This question is like the mental status questions for dissociative amnesia described in the section on clinical diagnosis. Also, the question is congruent with the DSM-5 Criterion A for dissociative amnesia: “An inability to recall important autobiographical information, usually of a traumatic or stressful nature, that is inconsistent with ordinary forgetting.”

If respondents reported affirmatively to this question, they were then asked to identify periods in 2-year increments, between the ages of 4 and 18, for which they had memory deficits. Overall, the age-standardized lifetime prevalence for CAMD was 18% (*N* = 1,502); the lifetime prevalence was 15% among men and 19% among women. Strong associations were observed between CAMD and each ACE. For both men and women, as the number of ACEs increased, the prevalence of CAMD increased in a graded fashion. Overall, adults with an ACE score ≥6 had nearly six times the odds of having autobiographical memory deficits compared to adults with an ACE score of zero. Seventy-seven percent of those reporting CAMD identified memory deficits for ages 4 to 6; 57% for ages 7 to 9; 32% for ages 10 to 12; 15% for ages 13 to 15; and 9% for ages 16 to 18.

Table 20–5.
Types of Adverse Childhood Experiences from the ACE Study

- Abuse
  - Physical abuse
  - Sexual abuse
  - Psychological abuse
- Neglect
  - Emotional neglect
  - Physical neglect
- Household dysfunction
  - Mother treated violently
  - Household substance abuse
  - Household mental Illness
  - Parental separation or divorce
  - Incarcerated household member

Data from https://www.cdc.gov/violenceprevention/aces/

Table 20–6.

Psychiatric, Behavioral, and Medical Outcomes of Childhood Traumas and Adversities: From the Adverse Childhood Experiences Study and Other Studies (Outcomes marked with an asterisk are from the ACE study)

Psychiatric
Depression*
Posttraumatic stress disorder (PTSD)
Pathologic dissociation
Alcohol and substance use disorders*
Panic and anxiety disorders*
Suicidal and self-destructive behavior*
Death by suicide*
Somatic symptom disorders
Eating disorders
Personality disorders
Multiple DSM diagnoses*
Hallucinations*
Autobiographical memory disturbances*

Behavioral
Multigenerational victimization and abuse
Victim of repeated sexual assaults
Violent behavior
Intimate partner violence—both as victim and perpetrator*
Smoking*
Impaired school/worker performance*
High-risk sexual behaviors in women*
Earlier sexual behaviors*
Early pregnancy as either father or mother*—teen pregnancy*
Unexpected pregnancy*
>30 Sexual partners in women*

Medical
Reduced health-related quality of life*
Greater healthcare utilization
HPA-axis dysfunction
Heart disease—ischemic heart disease*
Lung disease—COPD*
Liver disease*
Autoimmune diseases*
Kidney Disease

Chronic pain
Dementia (older adult Japanese cohort)
Morbid obesity*
Headaches*
Gastroesophageal reflux disease
Early menarche
Pregnancy complications and higher risk of children with autism spectrum disorders (risk factor: childhood sexual abuse)
Spontaneous abortions and fetal death*
Children with behavioral health problems

Data from Loewenstein R, Brand B, Gilbert LE, et al. Treating complex trauma survivors. *Psychiatric Times*. 2014;31:40–45.

Sexual and physical abuse in childhood were strongly related to autobiographical memory deficits. For women, the odds of reported amnesia in the sexually abused group were approximately double that of the nonabused group. Among abused men, the odds for CAMD were about 1.5 times higher than for the nonabused group. Sexual abuse frequency and severity were significantly related to autobiographical memory deficits, as was a close relationship between victim and perpetrator. Similarly, the presence of childhood physical abuse significantly increased the amnesia reports. Here again, women and men had, respectively, 2 and 1.5 times the odds of reported amnesia compared to individuals not reporting physical abuse. The combination of sexual and physical abuse was associated with a significantly greater extent of CAMD than a single type of abuse. The effects of comorbid depression could not fully explain the reported amnesia, as nondepressed individuals with higher rates of sexual and/or physical abuse showed significantly higher rates of CAMD than nondepressed, nontraumatized individuals. The authors concluded that ACEs, particularly sexual and physical abuse, have a specific, overarching effect on autobiographical memory, not just an effect on recall deficits for specific traumatic events. They hypothesized that, in the individuals reporting memory loss, cumulative traumatic experiences were encoded differently than ordinary autobiographical memory.

Table 20–7.
Factors Leading to Dissociative Amnesia after Traumatic Experiences

Trauma caused by human assault rather than natural disaster
Repeated traumatization as opposed to single traumatic events
Longer duration of trauma
Fear of death or significant harm during trauma
Trauma caused by multiple perpetrators
Close relationship between perpetrator and victim
Betrayal by a caretaker as part of maltreatment, especially sexual abuse
Threats of death or significant harm by perpetrator if the victim discloses their identity or information regarding the traumatic experience
Violence of trauma (i.e., physical injury caused by the trauma)
Earlier age at onset of trauma

Note: No factor has exclusively been associated with dissociative amnesia.

Adapted with permission from the *Treatment of Psychiatric Disorders*. 3rd ed. Vol 2. American Psychiatric Association; 2001.

Factors thought to predispose to trauma-related amnesia are listed in Table 20–7. However, none of these factors has been shown to exclusively distinguish individuals with amnesia for trauma from those without amnesia.

Delayed Recall for Early Life Trauma

Over 70 studies have confirmed the presence of delayed recall of sexual abuse, physical abuse, or both in clinical, community, and forensic populations, including retrospective, prospective, and longitudinal study designs. In various studies of clinical populations and individuals who identified themselves as survivors of childhood abuse, 16% to almost 100% reported amnesia for the abuse at some time in their lives, averaging approximately 30% across studies, though at a later point, they were able to recall the abuse. Lower prevalence was found in outpatients relative to inpatients. In studies of college undergraduates and women in the community, the prevalence of amnesia for abuse ranged from 13 to more than 50%.

A random sample of 505 men and women in the general population found that approximately 21% reported a history of childhood sexual abuse. Of these, 20% described full amnesia, and 22% reported partial amnesia for the abuse at some time in their lives. Somewhat lower rates of amnesia for physical and emotional abuse have been reported in clinical and nonclinical samples: about 20% of participants endorse this finding across studies.

Longitudinal and prospective studies, primarily following up large samples of individuals with childhood sexual trauma documented in court or medical reports, have found that almost 40% of respondents failed to recall a forensically or medically documented episode of childhood sexual abuse when interviewed 20 years or more later, even though, in some cases, other episodes of trauma or maltreatment were recalled. Some respondents with documented trauma denied any history of maltreatment at all.

Studies have compared both delayed recall and continuous memory for sexual abuse in a variety of studies. These have included studies of sexually abused individuals whose perpetrators admitted to the sexual abuse and many of whom had physical injuries related to the abuse. In these and other studies, researchers have concluded that, when corroboration is obtained, there is *no difference in accuracy between trauma memories that are continuous compared to those with delayed recall. Either type of memory may be essentially accurate, partially accurate, or wholly confabulated.*

Epidemiology of Dissociative Fugue Subtype of Dissociative Amnesia. As noted above, community studies have found almost no cases of dissociative fugue. Dissociative fugue is thought to be more common during natural disasters, wartime, or times of major social dislocation and violence, although no systematic data exist on this point. Most of the pre—DSM-III cases are difficult to assess because the diagnostic conventions are so different. Most cases in the literature describe men with dissociative fugue, primarily in military samples, although historical studies describe an “epidemic” of apparent dissociative fugue in men in late 19th-century France. Dissociative fugue is usually described in adults, although 20th-century case reports have described adolescents with fugue behaviors. Dissociative fugue as a disorder or as a symptom of other dissociative identity disorder may be harder to discern in children, as children may have a much more limited living space in which to wander in an altered state of consciousness.

Prevalence of Amnesia in Traumatized Populations

A variety of studies have looked at the prevalence of amnesia for specific types of traumatic experiences: combat, sexual abuse, physical abuse, and emotional abuse. Other studies have reported amnesia along with other symptoms of PTSD in various traumatized groups, including survivors of genocide.

Across studies of active-duty soldiers in World War II, the average point prevalence of amnesia in the immediate aftermath of the combat ranged from 5% to 35%. Only a small percentage of these soldiers were reported to have suffered significant head injuries as a direct cause of their amnesia. In one study of 1,000 soldiers just after the Battle of Dunkirk, 35% of those exposed to the most intense combat had amnesia for the battle. On average, 14.4% of the soldiers in this sample reported amnesia. Past or family history of dissociation or hysterical symptoms was associated with amnesia in soldiers with minimal battlefield exposure.

Semi-structured interview data using the SCID-D-R have shown significantly higher amnesia scores in Vietnam-era veterans with PTSD compared to combat veteran controls without PTSD. Studies of Nazi Holocaust survivors, Korean War veterans, Vietnam-era combat veterans with PTSD and adult survivors of childhood sexual abuse have repeatedly documented deficits in explicit memory on standardized memory tests as compared to matched controls.

Forty percent of a sample of 50 randomly selected survivors of the Cambodian Holocaust endorsed amnesia on PTSD diagnostic inventories. The average DES score in this group of severely traumatized individuals was 37.1. Amnesia was reported in 3% to 10% of Nazi Holocaust survivors assessed with a PTSD inventory. The highest prevalence was found in tattooed survivors of Auschwitz. Studies of natural disasters, such as the San Francisco earthquake in 1989, have found that 3% to 5% of survivors report amnesia for at least some aspects of these events.

Epidemiology of Depersonalization and Derealization Symptoms

Transient experiences of depersonalization and derealization are common in normal and clinical populations. They are the third most reported psychiatric symptoms after depression and anxiety. Transient symptoms of depersonalization/derealization in the general population are common, with a lifetime prevalence of 26% to 74% and 31% to 66% at the time of a traumatic event. A survey of a random community-based sample of 1,000 adults in the rural South found a 1-year prevalence of 19% for depersonalization and 14% for derealization. Several studies have identified age as a significant factor for transient experiences of depersonalization and derealization, with adolescents and young adults reporting the highest rates in normal population samples, although a recent German community sample did not find such a difference. In one study, 46% of college students reported at least one significant episode of depersonalization within the prior year.

Depersonalization/derealization can also occur with illicit drug use, especially marijuana, hallucinogens, ketamine, and 3,4-methylenedioxymethamphetamine (MDMA or ecstasy). Studies have found that about 15% of depersonalization/derealization disorder cases are triggered by drug use. Transient depersonalization has also been described after certain types of meditation, deep hypnosis, extended mirror or crystal gazing, and sensory deprivation experiences. It is common after mild-to-moderate head injury, where there is little or no loss of consciousness, but significantly less likely if unconsciousness lasts for more than 30 minutes. One study estimated that at least 20% of minor head injury patients experience significant depersonalization and derealization. Depersonalization/derealization symptoms also occur in patients with epilepsy, both as a manifestation of the aura as well as during the seizure itself, particularly in focal impaired awareness seizure (previously labeled complex-partial seizure) patients.

Depersonalization/Derealization during and after Life-Threatening Experiences. Acute depersonalization/derealization is also common after life-threatening or near-death experiences, with or without serious bodily injury. Several studies of accident victims have found that up to 66% of individuals with a life-threatening experience report at least transient depersonalization during the event or immediately thereafter.

As will be discussed further in the section on etiology, measures of depersonalization and derealization increased significantly in studies of general infantry and Special Forces soldiers after undergoing survival school training, including experiences of uncontrollable stress, semi-starvation, sleep deprivation, lack of control over hygiene, and external control of movement, communications, and social interactions.

Depersonalization/Derealization and the Dissociative Subtype of PTSD. Individuals with PTSD who experience depersonalization and derealization symptoms in response to trauma cues appear to be suffering from a distinct dissociative subtype of PTSD (see the section on PTSD). Studies have found that 15% to 30% of individuals with PTSD will meet the criteria for the dissociative PTSD subtype. Individuals with the dissociative subtype have specific fMRI and autonomic nervous system patterns when exposed to personal trauma scripts, including activation of medial prefrontal cortex and dorsal anterior cingulate; decreased activation of amygdala and insula; and no change, or even decrease, in heart rate. Individuals with this subtype have been found in a number of studies, including an international WHO study of 16 high, upper-middle, and low-middle income countries, respectively, with over 25,000 participants. In this study, the dissociative subtype of PTSD was associated with multiple types of earlier trauma, male sex, severe PTSD flashbacks, role impairment, disability, and suicidality, among other factors. Recent research has expanded our understanding of the neurobiology of the subtype, involving many other brain areas, for example, the periaqueductal gray.

Epidemiology of Depersonalization and Derealization Disorder

Community-based prevalence of DSM-IV-TR depersonalization/derealization disorder ranged from a 1-year prevalence of 0.8% in New York to 2.8% lifetime prevalence in Winnipeg. The Turkish study found a lifetime prevalence for DSM-IV-TR depersonalization/derealization disorder of 1.4%, but also included lifetime prevalence for DSM-IV-TR dissociative disorder NOS Example 2: “derealization without depersonalization” (prevalence 1.1%) (see Table 20–4). The DSM-5 depersonalization/derealization disorder diagnostic criteria specify that the presence of either or both symptoms is sufficient to make the diagnosis (see the section on Nosology). The actual general population prevalence of depersonalization/derealization disorder based on this sample could be around 2.5%, similar to that found in Winnipeg. Further evidence for this comes from a representative sample of the German population (*N* = 2,512), where the 6-month prevalence of depersonalization/derealization disorder was 3.4%. In this study, there were no age or gender differences in caseness for depersonalization/derealization disorder. The disorder was strongly associated with depression and anxiety. However, principal component analysis clearly showed that depersonalization/derealization disorder is a separate psychopathological entity, and not just a concomitant of anxiety and/or mood disorders.

STUDIES OF DISSOCIATIVE DISORDERS IN CLINICAL POPULATIONS

The frequency of dissociative disorders in clinical populations varies depending on the type of population studied (e.g., state hospital vs. private psychiatric hospital), assessment methodology, and geographic setting. Most studies use a two-tiered approach, where the DES is used to screen the target population, followed by a diagnostic interview (the DDIS and/or the SCID-D-R for individuals above a DES cutoff score, generally 20 to 30). These studies have been criticized for using relatively high cutoff scores that are more likely to capture individuals with the most severe dissociative psychopathology, such as dissociative identity disorder, not those with dissociative amnesia or depersonalization/derealization disorder. Very few cases of dissociative fugue have been found in any of these studies. In general, across all these studies, high rates of childhood and adult trauma are significantly associated with high-dissociation scores on the DES and dissociative disorder diagnoses. Types of traumas include sexual abuse, physical abuse, emotional abuse, neglect, medical trauma, adult rape, adult physical assault, intimate partner violence, and traumatic bereavement.

Many of these studies employed the DDIS as the diagnostic interview, which uses a simple “yes–no” format to inquire about the presence or absence of depression, somatic symptom disorders, and features of borderline personality disorder, as well as history of various forms of childhood maltreatment. All of these are common comorbidities in patients with dissociative disorders. Subsequent U.S. and Turkish DDIS studies of dissociative identity disorder found that high rates of borderline personality disorder symptoms indicate a decompensated patient with severe posttraumatic and dissociative symptoms, not an actual, reliable diagnosis of borderline personality disorder.

Inpatient Studies

In studies of over 2,300 adult and adolescent inpatients from Canada, China, Finland, Germany, Switzerland, the Netherlands, Turkey, and the United States, between 12% and 30% of patients scored above the DES screening target score. These studies report a wide range in point prevalence of dissociative disorders—between 1.7 (China) and 40.8% (United States) (mean = 15.8%; median = 12%)—although no study failed to find individuals with dissociative disorders. Across these inpatient studies, the frequency of dissociative identity disorder ranged from 0.4% of admissions to a Swiss hospital to 7.5% of admissions in a private psychiatric hospital in the United States (mean = 3%; median = 2%). In one Turkish inpatient study, almost half of patients diagnosed with DSM-III-R dissociative identity disorder (then, multiple personality disorder) believed themselves possessed by a jinn, a specific type of possessing entity from Islamic folklore.

The point prevalence of dissociative amnesia in psychiatric inpatients has been reported as 0.2% to 11%, depending on the sample, with the lowest rate in an inpatient sample in China. The point prevalence of depersonalization/derealization disorder in psychiatric inpatients ranges from less than 1% to 16%, depending on the sample and the method of assessment. Earlier studies reported higher rates of the symptom as these failed to differentiate carefully between depersonalization/derealization disorder or this as a symptom of other disorders, for example, schizophrenia or panic disorder.

A recent systematic study of a mixed group of Swiss outpatients and partial hospital patients, using standardized measures of disability, confirmed the finding of the New York State community sample: marked disability in the dissociative disorder group, with greatest disability in patients meeting structured interview, DSM-IV-TR diagnostic criteria for dissociative identity disorder, or dissociative disorder NOS, Example 1. This study concluded that the dissociative disorder patients fit the construct of *severe mental illness.*

Outpatient Studies

In studies of over 800 adult and adolescent outpatients in various settings from the United States, Europe, Turkey, and China, the point prevalence of dissociative disorders ranged from 12% to 29% (mean = 14.8; median = 13.8), with 2% to 6% meeting diagnostic criteria for dissociative identity disorder (mean = 5; median = 2.5). The point prevalence of DSM-IV-TR dissociative amnesia has been between 1.3% and 10% of adult outpatients. The point prevalence of DSM-IV-TR depersonalization/derealization disorder has been reported to range from <1% to 5% of adult outpatients.

In a study of 116 adolescents ages 11 to 17 admitted to a Turkish outpatient clinic who were assessed with the A-DES and the SCID-D for DSM-IV-TR, 45.2% had a dissociative disorder, 16.4 had dissociative identity disorder, and 28.8% had dissociative disorder NOS (primarily Example 1). There were high rates of traumatic experiences among dissociative adolescents, but not significantly higher than among adolescents who did not meet diagnostic criteria for a dissociative disorder.

Dissociative Disorders in Non-Western/Secular Settings. Shekhar Saxena and colleagues reviewed chart information on more than 4,000 Indian outpatients; 104 had a diagnosis of a dissociative disorder. Only 5% to 10% of these outpatients received a specified dissociative disorder diagnosis according to DSM-III or DSM-III-R. By contrast, 85.5% of dissociative disorder cases were covered by the specified dissociative diagnoses in ICD-10. Seventy-four percent of patients had dissociative convulsions, 9.5% trance and possession disorders, and 2% dissociative movement disorder. This supports the observation that, outside of Western/secular settings, dissociative disorders frequently come to clinical attention due to conversion and/or trance/possession symptoms.

Emergency Department. In a study of Turkish emergency department patients, about a third of 43 consecutive patients screened with the DES subsequently met DSM-IV-TR diagnostic criteria for a dissociative disorder when subsequently administered the DDIS. Of these, 13% met diagnostic criteria for dissociative identity disorder, 13% for dissociative disorder NOS, and 7% for dissociative amnesia. These patients also had high rates of childhood trauma and self-destructive and suicidal behavior; the latter frequently leading to treatment at more restrictive levels of care. Failure to recognize and treat these acute presentations of dissociative disorders may lead to extended periods of costly and ineffective treatment (see section on Treatment).

Dissociative Disorders in Substance Use Disorder Populations

Dissociative disorders and childhood and adult trauma are frequently found in studies of inpatient and outpatient substance abuse treatment programs. This is consistent with the high rate of current or past comorbid substance abuse in patients with dissociative disorders from other clinical settings. In a series of studies in the United States, Canada, Germany, and Turkey, 985 detoxified inpatients and outpatients in substance abuse treatment were screened for pathologic dissociation and DSM-III-R or DSM-IV-TR dissociative disorders with the DES and the DDIS and SCID-D, respectively. Males predominated in the samples. Rates of dissociative disorders ranged from 15% to 39% (mean = 24.3; median = 21.6); with 6% to 14% for dissociative identity disorder (mean = 8.25; median = 6.5); 0.5% to 27% for dissociative amnesia or dissociative fugue (mean = 8.6; median = 3.45); and 0.5% to 16% for depersonalization/derealization disorder (mean = 4.65; median = 1). The remainder received a diagnosis of dissociative disorder NOS. The wide range of findings in these studies likely reflects differing methodologies. These include varying DES cutoffs, diverse male/female ratios across samples, differences between DSM-III-R and DSM-IV diagnostic criteria, relative rate of alcohol abuse alone (associated with a lower prevalence of dissociative disorders in all studies), and younger age of participants (associated with higher rates of dissociative disorders).

Dissociative disorder diagnoses were strongly associated with a more refractory course—including more frequent treatment discontinuation, substance use, self-destructive and suicidal behavior, and history of childhood abuse and neglect. Women were more likely to receive a dissociative disorder diagnosis and had more severe and earlier forms of childhood trauma. Emotional abuse emerged as a particularly strong factor in the relationship between dissociative disorders and substance abuse. As in other clinical studies using the DDIS, the patients had high rates of comorbid depression, somatic symptoms disorders, and borderline personality features, the latter reflecting decompensation.

Dissociative Disorders in Samples of Women in Prostitution

In his treatise on hysteria, Briquet compared the prevalence of hysteria in nuns, house servants, and prostitutes. He found the highest rates in prostitutes, due, he speculated, to their traumatic and tumultuous lives, with intermediate rates in house servants and the lowest in the nuns. This comparison supported his view that hysteria was a brain disease, associated with stressful and traumatic life experiences, and not a disease of disordered sexuality.

In modern clinical studies of individuals with dissociative identity disorder, as many as 20% reported being in prostitution at some time in their lives. Contemporary studies of women in prostitution have found high rates of childhood abuse: 65% to 90% childhood physical abuse, 74% to 99% intrafamilial and/or extrafamilial childhood sexual abuse, and 98% emotional abuse. Similarly, women in prostitution suffer very high rates of adult rape, life-threatening physical assault, physical and neurologic injuries, homelessness, and emotional abuse/degradation by pimps, their associates, and clients.

Not surprisingly, very high rates of PTSD and dissociation have been found in samples of women in prostitution, with about 70% meeting diagnostic criteria for PTSD. Consistent with the pathologic dissociation construct, in some studies the DES scores of women in prostitution average >30, indicating severe dissociative psychopathology. Also, 67% of 40 women in prostitution (including exotic dancers) met diagnostic criteria for a dissociative disorder: 20% with dissociative identity disorder, 30% with dissociative amnesia, and 17.5% with depersonalization/derealization disorder.

GENETICS AND DISSOCIATION

Studies comparing child and adult cohorts of adoptive siblings, fraternal, and identical twins indicate that genetics account for around 45% to 50% of the interindividual variance in dissociative symptoms, with “non-shared,” stressful, and traumatic environmental experiences accounting for most of the additional variance. Genetic studies of dissociation suggest that there is a complex interplay between genetic factors and type, timing, and chronicity of trauma. Most studies indicating an epigenetic or candidate gene x maltreatment interaction related to dissociation (and for other stress-related disorders, e.g., depression, PTSD) are limited because of different study populations; differing ascertainment of types of maltreatment; retrospective, cross-sectional assessments of past events; and study of only one gene locus rather than interactions of multiple SNPs/epigenetic effects. Studies have linked dissociation to the interaction of early childhood trauma and chronic traumatic experiences with specific single-nucleotide polymorphisms (SNPs) in genes related to serotonergic (*5-HTTLPR*), dopaminergic (*COMT*), and BDNF systems, as well the HPA axis (*FKBP5*).

FKBP5 and Dissociation

Studies of GxE (e.g., maltreatment) interactions have focused on *FKBP5* as it is a glucocorticoid receptor (GR)-regulating co-chaperone of stress proteins involved in regulating the sensitivity and binding affinity of the GR. Induced by cortisol, *FKBP5* acts within a negative feedback loop to promote the transcription of ACTH and plasma cortisol. Specific *FKBP5* haplotypes and SNPs are associated with altered stress responsivity in healthy and psychiatric populations; non-suppression or hyposecretion of the HPA axis; and prolonged cortisol response after psychosocial stress in infants, adolescents, and adults. *FKBP5* genotypes are associated with alterations in brain function and structures (e.g., amygdala and hippocampus) in areas that affect emotional regulation, learning, memory, and inhibition. Although the direction of effects varies in different study populations—for example, by ethnicity, sex, outcome type—*FKBP5* gene variants are thought to act as a premorbid risk or protective factor for psychopathology in the context of environmental exposure to stress, specifically childhood adversity. Studies across ethnically diverse populations suggest that childhood adversity, but not adult adversity, interacts with *FKBP5* haplotypes and/or SNPs leading to adult PTSD and depression, both common outcomes of trauma.

In a study of 261 ethnically diverse, low-income women with a reported childhood maltreatment history, the *FKBP5 CATT* haplotype moderated indirect effects between maltreatment and symptoms of dissociation via "limbic irritability," with higher levels of dissociation in haplotype carriers exposed to maltreatment than noncarriers. "Limbic irritability" is a construct said to encompass non–seizure-based symptoms of focal unaware seizures (formerly partial-complex seizures): paroxysmal somatic disturbances, brief hallucinatory events, visual phenomena, automatism, and dissociative experiences—although all these symptoms are better conceptualized as dissociative.

In a cohort of hospitalized, severely injured children and adolescents after motor vehicle accidents, stabbings, gunshot wounds, or other traumatic events, two specific polymorphisms (*rs3800373, rs1360780*) in the *FKBP5* gene significantly mediated peritraumatic dissociation, an important risk factor for later PTSD.

In a large, prospective, community sample of carefully documented (e.g., social service substantiation) maltreated and nonmaltreated, low socioeconomic-status adolescents were grouped across four SNPs of

FKBP5. Dissociative symptoms were significantly higher among adolescents with no copies of the *CATT* haplotype who also experienced infancy-onset, chronic maltreatment over multiple developmental periods, compared to nonmaltreated adolescents or those with later-onset and/or less chronic maltreatment. There was a significant interactive effect of timing/chronicity of maltreatment and the *CATT* haplotype on dissociative symptoms. A secondary analysis suggested a protective effect from later dissociation of the *CATT* haplotype in adolescents with earlier and more chronic maltreatment.

Finally, a genome-wide association study of dissociative symptoms in individuals with combat-related PTSD identified 10 SNPs potentially related to a genetic basis for depersonalization and derealization symptoms, the most significant among which was r263232 located in the adenylyl cyclase 8 (*ADCY8*) gene on chromosome 8 which is known to be involved in long-term potentiation and synaptic plasticity. This finding may be relevant in clarifying specific learning and memory disturbance (e.g., dissociative amnesia) found in the dissociative disorders.

NEUROBIOLOGY OF DISSOCIATION AND DISSOCIATIVE DISORDERS

Neurobiologic study into the nature of dissociation does not fit tidily within the diagnostic system for dissociative disorders, but can be summarized more effectively along three major theoretical fronts. The first involves investigations into various trauma-related DBSs, including trauma-related altered states of consciousness (TRA; e.g., depersonalization and derealization) that are conceptualized as regulatory responses to fear or other extreme emotions in the context of perceived or actual threat (see section on Theories of Dissociation). These states are consistent with animal models of defensive responding and associated with objective behavioral markers including freezing and tonic immobility, cognitive nonresponse (e.g., blank stare or stupor, “trance-like behavior”), and increased pain threshold. However, TRASC are more adequately defined, following the research paradigm of *neurophenomenology,* by additionally appealing to their subjective experiential elements, which can include the 4-D model of trauma-related dissociation, an alteration in time sense (temporality seemingly “standing still”), flashbacks and “reliving” phenomena, voice hearing, depersonalization, and emotional numbing.

Acute, bottom-up, state-dependent stimulus-response processes most easily explain such states across sensorimotor, emotional, and somatosensory systems, including as peritraumatic dissociative responses and in the context of traumatic reminders. However, dissociative disorders can also be conceptualized as chronically evoked defensive states.

A second line of research investigates the biologic bases of dissociative amnesia, particularly for traumatic life events. Research is beginning to attribute to disturbances in specific neural systems the failed encoding, failed recall, and ultimate recovery of traumatic memories frequently observed in persons with dissociative disorders. The most implicated systems include hippocampal, amygdalar, and prefrontal modulation of amnestic barriers.

Finally, dissociation has also been described as a division or multiplicity of identity or sense of self. Such self/identity-state dissociation primarily implies a top-down organization of bottom-up processes via higher-order, superordinate "self" representations. Putative neurobiologic factors involved in TRASC, dissociative amnesia, and identity-state dissociation are elaborated below.

Neurobiology of Depersonalization and Trauma-Related Altered States of Consciousness

Several theorists have applied animal models as a basis for understanding dissociative responses in the context of peri- and posttraumatic experience. In general, their descriptions are consistent with the early concept of a "defense cascade" as evoked in sequenced fashion proportional to the physical closeness of a predator and the corresponding likelihood of its attack. After the animal orients toward the threat at a distance, a freezing response is usually elicited wherein the animal, while remaining alert of the predator's movements, keeps still to avoid detection. Flight or fight behaviors may ensue, however, when a predator signals impending attack, for example, by charging. However, if an organism determines it is unable to defend itself, it may assume a defensive posture of *submissive immobility,* which some authors term "freeze–fright" and "death feigning" responses. These defensive maneuvers have been regarded by some as the most likely behavioral counterparts to subjective experiences of dissociative TRASC such as including a distortion of temporal processing (altered time perception), depersonalization, emotional and physical numbing, and subsequent amnesia (i.e., limited encoding) for events taking place during the dissociative episode (e.g., mind “going blank”). Such lower-order defensive responses are thought to be adaptive when reserved for dire circumstances during which active defense (fight or flight) is not possible or would likely occasion an increased likelihood of harm. However, these responses become maladaptive particularly when chronically evoked involuntarily as the primary recourse to repeated trauma exposure, such as in circumstances of ongoing intimate partner violence and childhood abuse. Different stages of the defensive cascade are associated with different neurophysiologic responses: ranging from those exciting vigilance and motor reactivity to those promoting motor inactivity, reduced cognitive awareness, and anesthesia. As the neurophysiology of early orienting, flight, and fight responses has been described elsewhere including in the chapters focusing on Anxiety Disorders and PTSD, the authors focus on the immobile, submissive responses here.

Maggie Schauer and Thomas Elbert have suggested that traumatic events that provoke marked psychobiological arousal may lead to a prolonged immobile defensive posture in humans, like the animal defense of “playing dead.” They hypothesize that this leads to functional sensory deafferentation at the level of the thalamus, gating what are otherwise overwhelming incoming visual, auditory, somatosensory, and proprioceptive information. This survival-based psychobiological process may compromise sensory integration and modulation at limbic (amygdala, hippocampus) and higher cortical centers (e.g., frontal cortex, anterior cingulate, occipital cortex), that is, dissociation at the neurobiologic level.

Jonathan Krystal and others have suggested that if these defensive postures are prolonged, they may present phenomenologically as quasi-catatonic states. However, these states are more appropriately understood as indicating a profound level of helplessness marked by hypoarousal, physical analgesia, and a lowering of level of consciousness. Thus, the lack of sensory/perceptual/affective integration and the disturbance in consciousness leads to unintegrated, fragmented encoding of events. This could underlie the posttraumatic alteration in memory processes described in the ACE study. Phenomenologically, this is experienced as dissociative amnesia (see sections on Epidemiology and Dissociative Amnesia).

Similarly, Steven Porges's *polyvagal theory* may offer another aspect of the psychobiology of these processes. Porges expanded the classical notion of the relationship between the parasympathetic and sympathetic components of the nervous system during the stress response. He differentiates between two vagal systems in the brainstem medulla: that is, between the ventral vagus that regulates cardiac output as a means of facilitating prosocial engagement and the dorsal vagus that responds to threat with behavioral immobilization, hypoarousal, and analgesia. The latter may be one mediator of the metabolically conservative defensive stance of the immobile, dissociative state. These freezing behaviors are commonly reported by individuals experiencing extreme trauma where they cannot fight or flee, for example, in situations of rape and/or captivity. Individuals with severe early life trauma histories with dissociative disorders often report involuntary immobilization and freezing in response to earlier life danger that persists into their later lives. Consistent with this, studies of autonomic response in dissociative disorder individuals, when hearing personal trauma scripts, show that some of these subjects decrease, or do not increase, heart rate, blood pressure, and related autonomic functions.

Also, hypoarousal has been observed in persons with depersonalization/derealization disorder in the form of lower skin conductance levels (SCLs) and slower response-time when viewing unpleasant stimuli, interpreted as autonomic blunting. Persons with depersonalization/derealization disorder were also unable to increase SCLs with real-time biofeedback.

HPA Axis Function and Depersonalization. In contrast, studies of HPA axis functioning in persons with dissociative disorders have evidenced both hyper- and hypoactivation, although discrepant study findings may be attributable to measurement of urinary versus plasma cortisol, as well as measurement of basal HPA axis functioning compared with measurement in response to stimulus-evoked stress responses. For example, in an initial study of persons with depersonalization/derealization disorder symptoms, dissociation was unrelated to cortisol levels under basal condition but there were increases in cortisol levels when subjects experienced acutely stressful stimuli. Some researchers have observed increased plasma and urinary cortisol in depersonalization/derealization disorder compared to individuals with PTSD and other disorders, and greater resistance to, and faster escape from, dexamethasone suppression. However, an inverse relationship between severity of dissociation and cortisol stress reactivity has also been found, complicating interpretation of the data. The latter is consistent with Ruth Lanius and colleagues' conceptualization of dissociation as an "overmodulating" stress response that inhibits reactivity to trauma cues. Variable findings may also reflect the state-dependent nature of stress-response system reactivity in persons with dissociative disorders, including as a function of environmental context, as well as the presence versus absence of trauma triggers.

CNS Response and Depersonalization. In the 1950s, Penfield described "queer sensations of not being present and floating away" with stimulation of the superior and middle temporal gyrus in epilepsy patients. He insightfully theorized that these "illusions of unfamiliarity, strangeness and remoteness" involved an "alteration in the usual mechanism of comparison of immediate sensory perceptions with memory records." At the time, he claimed that such perceptual illusions could be produced by cortical stimulation "only in the temporal region, perhaps extending somewhat into the occipital cortex." Subsequently, in studies of persons without psychiatric illness undergoing neurosurgery, direct stimulation of the temporoparietal junction (TPJ) (particularly in the right hemisphere) also produced out-of-body experiences. Moreover, decreased response within the superior and middle temporal gyrus has been observed in depersonalization/derealization disorder in various studies, as well as decreased activity in orbitofrontal cortex, medial frontal cortex, left frontal-temporal cortex, and left caudate. In contrast, elevated activity has been found in parietal cortex and left occipital lobe, with certain effects contingent on specific symptoms such as negative affect.

In various social-cognitive neuroscience studies, the TPJ and extra-striate body area have been implicated in self-embodiment experiences. Accordingly, dysfunction in these areas figures prominently in neuroanatomical models for depersonalization experiences. Decreased volume of middle temporal cortex has also been observed as a function of suppressing emotional responses in depersonalization/derealization disorder, whereas prefrontal cortices may be increased in thickness, possibly associated with enhanced capacity for emotional suppression. Moreover, recent studies have shown that transcranial magnetic stimulation of the left dorsolateral prefrontal cortex (DLPFC), ventrolateral prefrontal cortex (VLPFC), and TPJ lead to decreased symptoms in depersonalization/derealization disorder patients. On the other hand, during resting state in healthy volunteers, deep transcranial magnetic stimulation over the medial prefrontal cortex induced depersonalization and a lowering of self-awareness. It was hypothesized that this result was caused by introduction of an error signal to cortical midline structures implicated by neuroimaging studies in self-referential processing. These experimental treatments indicate that medial prefrontal, lateral prefrontal, and temporoparietal cortical regions are involved in depersonalization experiences and in a diminished sense of self-embodiment.

Mauricio Sierra and German E. Berrios put forth the “corticolimbic disconnection hypothesis” of depersonalization/derealization, positing bilateral corticolimbic disconnection with prefrontal activation and limbic inhibition, resulting in hypoemotionality and attentional difficulties. Lanius and colleagues also applied this model to understanding the contrast between a state of hypoarousal and “overmodulation” of emotion in dissociative PTSD and the hyperarousal and emotional “undermodulation” typically seen in nondissociative presentations of PTSD. This model has been described as a putative neurobiological basis for the dissociative subtype of PTSD.

Two distinct neurobiologic components of the depersonalization/derealization experience have been proposed, subsumed by distinct neurocircuitry: “visual derealization,” associated with occipitotemporal dysfunction and “body alienation,” associated with parietal dysfunction. A review of recent neurologic literature provides convergent data for significant roles for the temporal and inferior parietal lobes and other transmodal sensory cortical areas in mediating depersonalization-like experiences. Depersonalization/derealization is common in seizure patients, especially in temporal lobe epilepsy with left-sided foci. Inferior parietal and angular gyrus tumors can also manifest with depersonalization symptoms. Structural lesions underlying “neglect” syndromes have been found to be concentrated in the right inferior parietal lobule. Patients with parietal lobe epilepsy frequently report somatosensory aura, disturbances of body image, vertiginous sensations, and visual illusions overlapping with the symptoms of depersonalization/derealization disorder. Differences in the amplitude of distinct EEG oscillatory frequencies have also been found to be gross markers of different arousal states in persons with dissociative disorders. For example, the slower waves of the theta frequency band (4 to 7 Hz) were associated with increasing dissociative states in a control population. However, contrasting results implicating decreasing dissociative symptoms with increasing amplitude of theta frequencies were observed in a psychiatric sample with mixed diagnoses. In summary, disturbances in bilateral inferior frontal, medial prefrontal, cingulate, insular, and temporoparietal cortex are all implicated in engendering experiences of depersonalization/derealization.

Neurochemical Systems and Depersonalization/Derealization. Dissociative experiences presenting in the form of TRASC have also been proposed to be dependent on specific neurochemical systems. Such hypotheses are informed by a considerable literature describing dissociative experiences in the context of illicit drug use, including depersonalization/derealization and a distorted sense of time. Drugs that have been reported to produce these dissociative symptoms include marijuana, hallucinogens, opiates, cocaine and its relatives, and dissociative anesthetics such as ketamine (Ketalar) and its relatives. Studies also implicate neuromodulators including serotonin and norepinephrine. The array of implicated pharmacologic pathways and their putative range of neurochemical actions would suggest that several neurotransmitter systems might be involved in producing dissociative reactions.

A series of studies using ketamine, an *N*-methyl-D-aspartate (NMDA) antagonist that increases glutamate release, has suggested that dysregulation of the NMDA glutamate receptor may play a central role in dissociative symptoms. Ketamine produces dose-dependent increases in dissociative symptoms including slower perception of time, tunnel vision, derealization, and depersonalization. Excess glutamatergic activation may contribute to structural changes in the hippocampus evidenced in MRI studies of patients with dissociative identity disorder. Pretreatment with a benzodiazepine or with lamotrigine, an anticonvulsant that decreases glutamate release, reduces by about half the dissociative effects of ketamine. NMDA receptors are widely distributed and have excitatory functions at hippocampal and cortical association areas. They also promote neuroplasticity and potentiate new learning. Krystal has proposed that NMDA-mediated integration of various brain areas may be necessary for cohesive conscious experience and therefore that dissociation may involve disruption of corticocortical, thalamocortical, amygdalocortical, and hippocampocortical connectivity.

Perceptual changes, derealization, depersonalization, and other acute dissociative symptoms have also been proposed to have a negative relationship with levels of neuropeptide Y (NPY) which is known to have anxiolytic effects. Charles A. Morgan and colleagues performed a series of naturalistic studies of soldiers undergoing high stress training in preparation for entry into the special forces, 95% of whom experienced acute dissociative symptoms as measured by the CADDS (see section on Theories of Dissociation). They found that NPY was elevated during these stresses and *higher* in Special Forces soldiers after 24 hours of stress than in general recruits. NPY levels were *inversely* correlated with self-reported dissociation severity. NPY was also correlated with cortisol and more adaptive behavioral responses. NPY is a neurotransmitter/modulator found in high concentration in cell bodies and terminals of the amygdala. In preclinical studies, NPY has shown anxiolytic and stress-lowering effects. Levels of NPY in Special Forces soldiers were inversely correlated with measures of dissociation, but positively correlated with both cortisol levels and behavioral performance under stress. Twenty-four–hour levels of NPY were at normal baseline in Special Forces soldiers but significantly lower than baseline in the other soldiers. The findings on this stress-hardy group underscore the importance of psychological adaptation to prior trauma on responses to current stressors. The NPY findings may suggest a neurobiological protective system that antedated stress and/or a neurobiological system demonstrating a more resilient adaptation to stress. NPY may have a role in regulation of anxiety and dissociation in high-stress situations. Higher levels of dehydroepiandrosterone sulfate (DHEA-S) in plasma were also associated with lower self-reported dissociation and improved behavioral adaptation to stress.

Norepinephrine is a neurotransmitter central to facilitating alertness, selective attention, and enhanced memory encoding under stress. Several studies support the notion of noradrenergic system involvement in depersonalization/derealization symptoms. Daphne Simeon and colleagues observed that decreased urinary norepinephrine is associated with increasing depersonalization/derealization symptom severity. In a small sample with chronic depersonalization/derealization, 24-hour urine norepinephrine was found to be strongly inversely correlated to dissociation severity. Similar findings have been reported in a sample with acute peritraumatic dissociation. In another study, administration of a norepinephrine reuptake inhibitor–induced depersonalization/derealization, while discontinuing the drug resulted in improvement.

Other studies suggest that depersonalization/derealization can be induced by serotonin agonists. Neurochemical challenge studies with the 5-HT_{2C} agonist m-chlorophenylpiperazine have demonstrated the induction of depersonalization/derealization in subjects with various diagnoses including social phobia, borderline personality disorder, and OCD, as well as the induction of flashbacks and dissociative symptoms in a subgroup of patients with PTSD. A PET imaging study with the hallucinogenic 5-$HT_{1A/2A}$ agonist psilocybin resulted in increased dopamine in striatum that correlated with depersonalization/derealization severity.

Other neurochemical systems putatively implicated in depersonalization and derealization include the cannabinoid system. Cannabinoids, such as marijuana, have been consistently shown to induce depersonalization/derealization, with a pronounced component of temporal disintegration. In addition to their action at the cannabinoid CB receptors, cannabinoids block NMDA receptors at sites distinct from other noncompetitive NMDA antagonists. A PET study using intravenous tetrahydrocannabinol (active compound in marijuana) found that cerebral blood flow increases in the right frontal and anterior cingulate cortex correlated with depersonalization/derealization severity, while there was a subcortical cerebral blood flow decrease in amygdala, hippocampus, basal ganglia, and thalamus.

The opioid system has also been implicated in depersonalization/derealization disorder states. The endogenous opioid system mediates stress-induced analgesia. In PTSD patients, analgesic response to combat stimuli can be blocked by pretreatment with the opioid antagonist naloxone. The κ opioid agonist enadoline induces a depersonalization-like syndrome in healthy subjects compared to placebo, with perceptual disturbances and a sense of detachment. Along these lines, opiate antagonists have been reported to reduce dissociation, such as high-dose naltrexone in borderline personality disorder and intravenous naloxone in chronic depersonalization/derealization. The opioid antagonist nalmefene has been reported to decrease emotional numbing in veterans with PTSD.

Dissociative Amnesia

As described in the sections on Epidemiology and Dissociative Amnesia, traumatic events, particularly cumulative traumatic events during development, are associated with a variety of autobiographical memory deficits. However, there are few systematic studies of the underlying neurobiology mediating dissociative amnesia per se. Neurobiologic findings in clinical populations have been primarily reported in single case studies or small case series without adequate diagnostic assessment to rule out dissociative identity disorder, history of trauma, or comorbid disorders such as PTSD.

Nevertheless, neuroimaging studies in healthy persons and lesion studies in neurologic patients strongly implicate crucial roles for the hippocampus, temporal lobe, temporoparietal junction, posterior cingulate, precuneus, and medial prefrontal cortex in the formation and recall of episodic memories. As malfunctioning of these structures has variably been implicated in neuroimaging studies of dissociative disorders across various experimental paradigms, such disturbances may partially account for amnesia for traumatic events and an overall impoverishment of memory for childhood in traumatized persons with dissociative disorders. Deficits in encoding of traumatic events and/or inhibition of retrieval of memory for the events are important models to help explain dissociative amnesia. Both processes may be relevant to amnesia symptoms in a specific patient. However, for clarity, data on encoding and retrieval models of dissociative amnesia will be described separately.

Peritraumatic Dissociative Amnesia: Encoding. Stimuli that are mildly to moderately emotionally provocative are known to be associated with a memorial advantage over neutral cues, partially mediated by response in neural networks involved in salience processing that favor their attention during encoding, for example, via the amygdala, insula, and anterior cingulate. However, highly threatening experiences, such as traumatic events, may cause an individual to orient *away* from the source of threat in fear (e.g., by averting gaze) or to defensively retreat “inward” (i.e., emotional suppression) or “outward” (i.e., out-of-body experience, depersonalization) as automatic peritraumatic emotional self-regulation strategies. Each of these emotion regulatory responses has been associated with a diminution of salience network processing, which in theory should predict reduced attention and encoding. By parsimony, these studies suggest that impoverished encoding is one neural pathway of amnesia for traumatic events.

Consistent with this interpretation, a recent study provided a striking experimental demonstration of the impairment of autobiographical memory by depersonalization, as evoked by a multisensory perceptual illusion during encoding. Study participants evidenced impoverished recall of scripted life events encoded in an “out-of-body” (depersonalized) condition. This was induced by altering subjects’ visual and somatosensory experience to give the false impression that their embodied frame of reference emanated from a place 2 m behind or beside their actual physical position. Participants rated events encoded from this depersonalized perspective as less vivid and recalled these events more poorly. Moreover, a neuroimaging study showed that impoverished recall was associated with reduced left posterior hippocampal activation in early rehearsal trials, with greater activation in later trials. This is a reversal of the observed and expected effect for encoding under natural conditions. The authors concluded that a dissociative out-of-body experience during an acutely stressful event could, by itself, impair the encoding mechanism and produce fragmented, spatiotemporally disorganized memories.

Neuroimaging has also been used to study laboratory paradigms of active memory suppression or repression. For example, an influential study conducted in healthy persons examined neural response during conditions in which participants were either instructed to think about or “not think” about (i.e., suppress memory of) the second member of a previously seen pair of words. Subsequently, cued recall of suppressed words was inferior to recall of baseline words and even generalized to novel test cues. Neural networks implicated in this suppression response involved bilateral prefrontal cortex, paralimbic, subcortical, and parietal integration areas and reduced activation bilaterally in hippocampal areas. Moreover, hippocampal activation patterns differed from trials involving ordinary forgetting. These were greater in response to suppression-induced forgetting compared to items where suppression failed: that is, items that were successfully recalled despite attempted suppression.

The authors concluded that activation of a broad neural network was involved in memory suppression, including hippocampus, DLPFC, anterior cingulate cortex, medial-temporal lobe, dorsal premotor cortex, presupplementary motor area, and intraparietal areas. The authors described a neurobiologic model for memory control in response to trauma, where, via encoding effects, these networks are involved in producing lasting amnesia in response to traumatic events.

Neurobiology of Dissociative Amnesia: Disturbed Retrieval. Although no neurobiologic studies of dissociative amnesia have employed large samples, case studies of patients with generalized dissociative amnesia or dissociative fugue subtype with loss of personal identity have shown unusual metabolic and activation patterns in PET scanning and SPECT studies. A PET study of 14 amnestic patients tested with PET in a resting state showed decreased metabolic rate within right inferolateral prefrontal cortex. These findings complemented those in a neuropsychological case series of persons with amnestic syndromes prompted by traumatic and/or stressful events demonstrating deficient response in right frontotemporal cortex. Other findings included, respectively, reversal of the usual patterns of right and left hemisphere activation during autobiographical and semantic recall. These were similar to findings in individuals with traumatic brain damage and related memory deficits, and those with disruptions or functional disconnections between major neural networks.

Replicated neurobiologic case studies of dissociative amnesia have shown deficits in activation of the hippocampus and medial temporal lobe, with possible inhibition by frontal systems. Reduced occipital involvement during episodic recall has also been documented. These findings are consistent with fMRI studies of posthypnotic amnesia in healthy subjects that showed reduced response in anterior temporal and occipital cortex, with increased response in left rostrolateral prefrontal cortex. These patterns normalized after discontinuation of the hypnotic suggestion for amnesia. Nonhypnotized participants exhibited the expected habituation effect to repeated presentation of previously encoded cues. Participants in the hypnotic amnesia condition had *decreased* neural response during early recognition trials, as compared to later trials following removal of the hypnotic suggestion for amnesia. These findings implicate deficits in retrieval processes, rather than the deficient encoding in depersonalization and dissociative amnesia.

Hans J. Markowitsch has reported similarities between organic and psychogenic (i.e., dissociative) retrograde amnesia, both thought to be caused by a desynchronization or blockade in neural processing ("mnestic block syndrome") related to altered brain metabolism possibly including changes in various neurotransmitter and hormonal systems, such as GABA-agonists, glucocorticoids, and acetylcholine. Current neurobiologic theories of dissociation suggest a possible role for glutamate in at least some dissociative symptoms, warranting further pharmacologic investigation.

Markowitsch and his colleague Angela Staniloiu have studied several large samples of individuals with dissociative amnesia. In addition to the above, they posit that psychological trauma in early life induces a stress-hormone-related-and mediated neurotoxic cascade resulting in desynchronization between frontal lobe systems important for autonoetic consciousness (awareness of our self in past, present, future and related to development of sense of self and identity) and the temporal lobe/amygdalar systems that underlies emotional processing of memory. They view this desynchronization as precipitated by traumatic and adverse life experiences, particularly in childhood, adolescence, and early adulthood. These may be modulated by genetic, epigenetic, critical period, familial, and sociocultural factors as well as medical and psychiatric comorbidities. They hypothesize that glucocorticoid stress hormones bind to receptors in the amygdala and hippocampus leading to changes in morphology and/or function of these systems and impairment in encoding and/or retrieval of episodic-autobiographical memory. In case studies, they found support for this hypothesis with FDG-PET showing overall reduction in cerebral metabolism and in memory-sensitive areas of the temporal and diencephalic regions, as well as bilateral hippocampus. They view "mnestic block syndrome" as demonstrating that memories are encoded in brain, but full conscious access is blocked.

Michael C. Anderson and colleagues reviewed a large group of studies that attempt to show that memory can be actively suppressed, either at the point of memory encoding or by retrieval inhibition. Many of these studies utilize the laboratory paradigm of "Think, Not Think" (TNT) instructions: to study the variety of neurobiologic processes that may be involved subjects are instructed to recall or forget word pairs or whole lists of words. In these studies, similar to the model of frontal systems suppressing amygdala and insula in the dissociative subtype of PTSD, DLPFC, and VLPFC appear to be involved directly in suppression of activity in hippocampus, the fusiform gyrus (perceptual awareness of objects), and occipital lobe (visual perceptual priming). Neural networks that account for the findings in the TNT studies are similar to those found in the fMRI study of posthypnotic amnesia. Also, as discussed below, in PET studies of dissociative identity disorder, similar findings of frontal activation and suppression of neural systems involved in autobiographical memory and self-referential processes occurred in self-states that did not experience personal trauma scripts as autobiographical memory.

Neurobiology of Self/Identity-State Dissociation

In contrast to the literature implicating dysregulation of specific neurochemical systems in the experience of depersonalization and derealization, it is unlikely that any single or simple combination of a few neurotransmitter systems could underlie the highly complex symptom picture of dissociative identity disorder. This picture includes symptoms found in each of the other dissociative disorders—depersonalization/derealization and dissociative amnesia—for past and current life experiences, as well as dissociative fugue experiences. In addition, individuals with dissociative identity disorder experience symptoms of identity confusion and identity alteration; symptoms of overlap, interference, and intrusion of identities; and spontaneous age regression, analgesia, multimodal hallucinations, and negative hallucinations (lack of perception of sensory stimuli) consistent with spontaneous dissociative, posttraumatic, and autohypnotic processes (also, see section on Diagnosis).

Historically, early medical models attempted to explain dissociative identity disorder by invoking phenomena such as sleep and dreams, hypnosis and somnambulism, epilepsy, and disconnections between the cerebral hemispheres to explain the multiplicity of identity and sense of self that characterizes this disorder. The scientific investigation of identity-state dissociation dates to the experiments of Prince in 1908, who, by way of a crude polygraph to measure GSR, demonstrated distinct responses to emotionally valenced words across alternative identity states in a dissociative identity disorder patient. This basic experimental paradigm, involving a comparison of biologic measures in different dissociative identity disorder self-states, continues to dominate the literature to this day. Indeed, over the next century dozens of psychophysiological and cognitive performance studies were published along similar lines. Often, these lacked methodologic rigor either due to small sample sizes, single case study design, or deficiencies in experimental control. In contrast, the work of Frank W. Putnam and Antje A.T.S. Reinders stands out as a notable exception. Their studies included rigorous control groups as well as study of dissociative identity disorder subjects under a variety of conditions, including hypnotic trance and states of relaxation, among others.

Putnam's research group has demonstrated unique autonomic system (ANS) reactivity between alternative identity states in dissociative identity disorder patients and dysregulation of the EEG during transitions between identity states. Nevertheless, an important observation in these studies was the instability of both ANS and CNS measures not only across but also *within* different identity-state presentations, the magnitude of which was found to be greater than that observed in nonpsychiatric control groups. Rather than implicating a model of sustained differences between alternative identity states in persons with dissociative identity disorder, these studies found general disorganization in both temporal and spatial integration of ANS and CNS. Consistent with this, beyond the gross measures of amplitude of various bandwidths of the EEG typically employed in early studies, more recent literature is also turning to the investigation of functional connectivity and signal complexity as possible EEG markers of dissociative disorders and the general effects of complex trauma histories.

For example, an EEG study of α-wave coherence using five simulator controls and five dissociative identity disorder subjects found significant differences in the latter group between identity states in six brain regions, whereas there were no differences in the control group. A study comparing ERPs elicited by words learned in the same or in different identity states for four dissociative identity disorder patients found little support for the existence of amnesia between personality states when compared to controls from another study who deliberately concealed knowledge of previously learned words. However, given that the ERP is highly relevant to the early detection of salience and basic stimulus processing, the current authors think it unlikely that this very early indicator of cognitive processing would strongly distinguish identity states in dissociative identity disorder. The latter are more likely to be distinguished during poststimulus phases typified by cognitive elaboration and higher-level cognition partially overlapping other EEG markers such as the phenomenon of pre- and poststimulus EEG alpha desynchronization (see subsection on Inter-Identity Amnesia in the section on Dissociative Identity Disorder). A Quantitative EEG (QEEG) study, found that EEG-β activity in frontal and temporal lobes varied between different identity states in dissociative identity disorder patients. Other studies also suggest putative roles for primary sensory and motor cortices, and frontal and prefrontal cortex and nucleus accumbens, during transitions between identity states.

General neural markers for dissociative identity disorder include Vedat Sar and colleagues' findings regarding decreased perfusion in the orbitofrontal region coupled with increased blood flow within medial and superior frontal regions, all associated with self-referential processing in controls; these researchers further identified increased perfusion in occipital cortex. Left temporal involvement in dissociative identity disorder also has been implicated in SPECT studies, which may partially underlie auditory hallucinations (voice hearing) frequently reported by such patients. Strikingly, hippocampal volume was also reported to be 19% smaller, and amygdalar volume 32% smaller, in patients with dissociative identity disorder compared to healthy controls. Reduced volumes in the parahippocampal gyrus were also reported and correlated with severity of both psychoform and somatoform dissociation symptoms.

Reinders' group has reported elevated heart rate, higher systolic blood pressure, and lower heart rate variability in persons with dissociative identity disorder in response to audio-recorded personal trauma narratives when individuals experienced the narrative as autobiographical in nature (in a Traumatic Identity State [TIS]). Their biologic indices differed when they experienced the event as non–self-referential in a different self-state (Neutral Identity State [NIS]); under those conditions, their autonomic responses and brain activation patterns were similar to those of individuals with the dissociative subtype of PTSD. These researchers have extended the alternative identity-state paradigm to fMRI neuroimaging. They found that response within regions of the brain previously associated with "self-referential processing" in controls (thinking about oneself and evaluating whether a stimulus such as a personality trait descriptor is self-referent) differed between TIS and NIS in persons with dissociative identity disorder. These differences were noted most notably in the medial prefrontal cortex and midline posterior cortex (posterior cingulate, precuneus). Moreover, these effects cannot be simulated by trained nonpsychiatric controls. Differences from controls have also been observed in regions associated with emotional processing (amygdala, insula) and body representation (TPJ). Resting state metabolism and functional connectivity also differed between identity states in persons with dissociative identity disorder, indicating a level of internal (in)coherence and (de)stabilization that endures beyond the momentary, stimulus-reactive experimental design employed by most previous studies.

Reinders and colleagues have continued to expand the fMRI findings in response to the critics of the TM of dissociation. In studies comparing the two previously described types of dissociative identity disorder identity states ("traumatic" and "neutral") in high and low "fantasy-prone" simulators, Reinders and colleagues replicated and extended the findings from the PET studies. Both types of patient states were significantly different from simulators on brain activation and autonomic variables, particularly from the high fantasy-prone group. TIS differed from simulators on autonomic activation and on activation of the insula and dorsal striatum. Replicating the findings from the prior studies, and similar to individuals with the dissociative PTSD subtype, NIS showed bilateral activation of the superior frontal gyrus, prefrontal cortex, cingulate, posterior association areas, and parahippocampal gyri. Brain deactivation patterns in individuals with dissociative identity disorder were independent of fantasy-proneness. Significantly differently from both control groups, in response to traumatic and neutral scripts, TIS showed activation of the striatum, particularly the caudate nucleus not the hippocampus. This was interpreted as processing of trauma memory and maintenance of state stability by the caudate, not the hippocampus, in dissociative identity disorder. In fact, dissociative identity disorder patients were more like *low* fantasy-prone than high fantasy-prone simulators, showing less reactivity in emotional brain areas. This is consistent with the psychological profile data that, in dissociative identity disorder, dissociation is an ideational, distancing defense recruited to attempt to downregulate emotional stimuli (see section on personality organization of Dissociative Identity Disorder).

Data on the neurobiologic basis of dissociative identity disorder are growing rapidly. However, a factor that has significantly impeded progress in this domain is the (at least implicit) concern of many studies to use neurobiologic measures to cross-validate the existence of the disorder, as opposed to utilizing research methods to better understand the unique neurobiologic and psychophysiological profiles of this group of patients. Studies have frequently been designed to reveal biologic differences between personality states in dissociative identity disorder, often compared with simulator controls. As with attempts to demonstrate out-of-body states in depersonalization/derealization disorder, the focus of much research has been to replicate the finding that such psychological states are indeed real. This methodology would scarcely ever be expected for inclusion in biologic studies of other disorders, for example, depression, schizophrenia, PTSD, or borderline personality disorder. Instead, the veracity of that self-reported symptomatology is taken for granted, with biological studies then designed to interrogate its etiology rather than serving to discount concerns that the self-reported symptoms might be factitious. Also, studies of dissociative identity disorder rarely include male subjects, although epidemiologic data show that the male-to-female ratio in general population studies is about equal.

More recent neurobiologic studies have begun to address these issues. Sima Chalavi and colleagues found that, although both individuals with PTSD and those with dissociative identity disorder comorbid with PTSD had smaller hippocampal volume than healthy controls, the comorbid-disorder group differentiated from the PTSD group in the abnormal shape and significantly smaller volume of their CA2-3, CA4-DG and (pre)subiculum. This correlated with more severe histories of childhood trauma and dissociative symptoms. In a multi-center, case-control study, Reinders’ group evaluated cortical morphology of women with dissociative identity disorder compared to healthy controls. The dissociative identity disorder group showed reduced cortical volume, thickness, and surface area relative to controls. There were significant, widespread, spatially dependent abnormalities in gray and white matter in numerous areas of the frontal cortex, anterior cingulate cortex, temporal, parietal, and occipital cortices and cerebellum. Women with dissociative identity disorder show extensive, significant volumetric reductions of regional gray matter in the insula, the cingulate cortex, the dorsolateral, superior, medial, and orbitofrontal prefrontal cortex, and the superior and inferior temporal lobe. Differences in cortical thickness and surface area were largely nonoverlapping and each contributed separately to cortical volume. Pattern-recognition models of brain biomarkers differentiated patients with dissociative identity disorder from controls with a sensitivity of 71.88% and a specificity of 73.81%, like what has been found for other psychiatric disorders.

Lauren LeBois, Milissa Kaufman, and colleagues developed an ingenious paradigm to test the capacity of dissociative identity disorder patients to recognize their own faces. Phenomenologically, individuals with dissociative identity disorder report discomfort, detachment, alienation, and fear about their bodies and faces. Often, they are phobic of being seen, as well as of seeing themselves in the mirror, frequently describing lack of self-recognition and/or seeing shifting “other people” instead of themselves in mirrors (e.g., “a blonde little girl,” “a scary witch,” “a big, burly man”). A backward-masking study in which healthy controls and individuals with dissociative identity disorder were asked to rate familiarity of their own faces, a stranger’s face, and that of a famous person was conducted. Dissociative identity disorder patients showed a significant lack of self-bias effect across face types. However, patients who had progressed in treatment had recovered this capacity, possibly due to improved internal cooperation, coordination, cooperation, and awareness among self/identity states.

These investigators performed a novel, large-scale connectivity analysis using machine learning on patients with dissociative identity disorder. The goal was to develop brain-based measures that were sufficiently sensitive and robust to assess pathologic dissociation apart from self-reports. The models showed a moderate ability to estimate dissociation within individuals after controlling for childhood trauma and PTSD severity. Connections that contributed most to the model involved the default mode and frontoparietal control networks, the latter replicating findings from multiple studies of dissociation using differing methodologies (see above). This work represented a first step in developing functional brain biomarkers for dissociation in individual subjects.

The studies of Chalavi et al. and LeBois et al. begin to move away from the neurobiological studies of dissociation and dissociative disorders where the goal is to respond to critics and “prove” that these conditions are “real.” In a recent review, Constance Dalenberg and colleagues described the use of multiple rhetorical, not scientific, arguments in most critiques of the TM of dissociation, and the reality of dissociative disorders. These include—among others—“moving the goalposts,” asking for a “super study,” and/or utilizing “motivated skepticism” to discount every study supporting the TM, always demanding some additional “proof.”

Rather, what the literature requires is a greater number of comparative studies, for example, between different dissociative disorders, and between dissociative and nondissociative trauma-related disorders. A dimensional perspective may also be useful in this regard. Frontal, limbic, temporal, and parietal cortices are nonspecifically implicated in both depersonalization/derealization disorder and dissociative identity disorder, and may relate to alterations in self-referential processing, emotional numbing, and felt disembodiment common to both disorders. However specific distinguishing clinical features (e.g., voice hearing in dissociative identity disorder) may require study of specific markers appropriate to differential diagnosis, for example, within auditory association cortex. Longitudinal and naturalistic studies utilizing biological measures will be needed, as well as neuroimaging studies of combined psychological and neurobiological interventions (e.g., neurofeedback, rTMS) to better appreciate the clinical significance of the biologic markers so far identified for the dissociative disorders, and to help develop more effective treatment for these patients.

ETIOLOGY OF DISSOCIATION AND DISSOCIATIVE DISORDERS

This section discusses etiologic theories about dissociation and dissociative disorders. A comprehensive etiologic model involves integrating components of several of these as well as cognitive and neurobiologic correlates (see Neurobiology).

Discrete Behavioral States (DBS) Model

The concept of *state of consciousness*—as a distinct unit of behavioral organization—has been part of the psychological vocabulary for several centuries. DBSs are defined as repeating patterns of mind, body, and brain variables/dimensions that for a finite period are correlated with and/or influence how a person thinks, feels, and acts. Each distinct DBS—combinations of states—uniquely (state-dependently) influences perception, cognition, memory, emotion, motivation, values, psychophysiology, and interpersonal interactions.

The DBS model, developed by Frank W. Putnam, Jr., is a transtheoretical, translational *theory of human consciousness*. DBS theory elucidates human experience and behavior across myriad normal and pathologic domains. Discrete states of being are so pervasive and ubiquitous in our moment-to-moment lives as to be essentially invisible unless they shift in dramatic, unexpected ways and, even then, we usually do not conceptualize these changes in terms of behavioral “states.” In the “atheoretical” DSM-5, the term “state” occurs over 100 times. DBST is an organizing framework for a vast array of normal and pathologic human behavior and experience, individually, in relationships, and in groups. The power of DBST is based on substantial support from numerous domains of research, including trans-species chronobiology, child development, neurobiology, psychopathology, and many others—from subcellular to social systems.

The notion of DBSs is inherent in all theories of dissociation and the dissociative disorders. As such, DBST subsumes and is a superordinate, organizing framework for all other theoretical models discussed in this chapter. Transdiagnostically, DBST postulates that dissociative disorders belong to a large group of psychiatric conditions characterized by rapid—often environmentally triggered—discrete shifts in states of consciousness, “state-change disorders.” Others include mood disorders (e.g., bipolar disorder), panic disorder, periodic catatonia, substance use disorders, trauma- and stressor-related disorders, conversion disorder, sleep disorders, borderline personality disorder, impulse control disorders, delirium, paraphilias, most feeding and eating disorders, brief psychotic disorder, and many others. In state-change disorders, there are gradual or rapid switches or shifts between two or more distinct states of consciousness, at least one of which is pathological. Switches among dissociative identity disorder self-states, panic attacks, and PTSD flashbacks are examples of dysfunctional behavioral states, often triggered by specific stimuli or interactions that rapidly reorganize an individual’s emotions, thinking, and behavior.

The DBS model postulates three general levels of psychopathology inherent in all *state-change* disorders. The first is intrinsic to the nature of the pathologic states per se, for example, flashbacks, numbing, depression, mania, dissociative identity self-states, depersonalization/ derealization. The second results from the individual’s inability to self-modulate the expression of these dysfunctional states or to integrate them into their life, such as by becoming unable to work. The third level results from maladaptive attempts to block or minimize effects of painful, distressing, pathologic states, such as attempts to self-medicate mood disorder symptoms; self-injury to modulate intense emotional states like shame, anger, and sadness; and/or social withdrawal to avoid triggering PTSD reactivity. Transdiagnostically, the DBS model predicts that effective treatment should facilitate better self-modulation of intense or dysphoric states; address the state dependency of dysfunctional perceptions, cognitions, and assumptions; and assist patients in finding recovery-oriented solutions to modulate symptomatic states.

Behavioral States as the Basic Unit of Human Development. In the 1960s, a group of child psychologists and psychiatrists formally operationalized the study of DBSs in normal human infants. Heinz Prechtl, Robert Emde, Robert Harmon, and Peter Wolff empirically validated a multidimensional, ethologic classification system used to study normal infants. By using four to five variables (e.g., heart and respiratory rate, extremity muscle tone, skin perfusion, vocalization, facial expression), researchers could delineate a basic set of DBSs shared by most full-term, newborn infants. Each state was time-limited, with a distinct switch between them. Caregiver interactions could evoke or suppress certain states to shape the infant's behavior. With maturation, additional behavioral states were added, and the pathways of connections among the distinct states become more complex, corresponding to the child's increased behavioral repertoire. Fundamental to healthy emotional regulation, the developing child increases the ability to self-modulate the expression of these DBSs and to integrate the many additional states related to normal development. More recent research characterizes the interplay of DBS between caregiver and child as essential to developing normal attachment, capacities for behavioral state regulation, brain executive functions, and other vital human capabilities.

Theories of Dissociation

In the literature, the term *dissociation* is conceptualized in over 25 different, and sometimes diametrically opposed, ways. Substantial imprecision can result in both clinical and scholarly discussions unless one is clear about which definition is being used (see Table 20–8 for a partial list of the often-contradictory ways dissociation has been defined in the literature). For example, dissociation has been hypothesized as a normal psychobiological capacity to filter perception to focus attention on salient information (e.g., being unaware of the sensation of one's body while sitting in a chair). Most theories of dissociation conceptualize it as a *dimensional, adaptational process* (e.g., as a response to trauma) and subsequently as an intrapsychic defense to prevent full awareness of past traumatic experiences and their impact on the person. In the DSM and ICD systems, dissociation describes *categorical distinctions among psychopathological entities:* the dissociative *disorders.*

Discrete Behavioral State Model, Psychopathology, and Dissociation

The DBS model conceptualizes dissociative phenomena as distinct states of consciousness differing in terms of the individual's senses of self, identity, and physical being; access to explicit, implicit, and /or autobiographical memory; and accompanied by specific patterns of state-dependent psychobiological activation (e.g., in the ANS). Disturbances of self-organization are the nucleus around which dissociative states organize—just as distinct moods serve to organize the DBSs of bipolar patients. This is the case in dissociative amnesia, as autobiographical memory continuity and coherence is fundamental to sense of self.

Table 20–8.
Different Conceptualizations of Dissociation

1. Dissociation is described as a *normal* part of the perceptual filtering system of the brain/mind that allows for normal functioning by helping the mind focus on relevant stimuli and filtering out other stimuli.
2. A fundamental aspect of hypnotic experience, related to the capacity for divided consciousness/attention, and/or separation of cognitive and behavioral schemas (e.g., losing track of driving for several minutes).
3. Theories of mind that describe divided consciousness in normal, hypnotic, and pathologic states (e.g., Ernst Hilgard's *Neo-Dissociation Theory*).
4. Continuum Model: *A continuum from normal to pathologic,* with normal capacities for absorption, focused attention, and imaginal involvement at one and end, and dissociative disorders at the other.
5. Taxon Model (mixed dimensional /categorical): Two *different dimensional continua of dissociative phenomena* that are categorically different: *normal dissociation* and *pathologic dissociation.*
6. A *psychobiological process* that is involved in human response to acute and/or chronic trauma.
7. An intrapsychic defense that is understood in psychodynamic terms, e.g., protective against full recall of traumatic experiences and impacts, e.g., trauma-based intrapsychic conflict, caregiver betrayal.
8. In the DSM and ICD systems, dissociation is used to describe *categorical distinctions among psychopathological entities:* the dissociative *disorders.*

The DBS model conceptualizes dissociation as basic to the psychobiology of the human response to life-threatening danger: a protective activation of altered states of consciousness that change perception, pain sensation, time sense, access to memory, and senses of self and identity. Trauma-related memories and affects are encoded during these altered states. Upon return to baseline, the person has less access to the dissociated information, with dissociative amnesia for at least some part of the traumatic events. However, the dissociated memories and affects can manifest themselves in nonverbal forms as posttraumatic nightmares, reenactments, intrusive imagery, and somatoform symptoms. Not only is there amnesia for the traumatic event but the person also frequently has dissociated basic assumptions about the self, relationships, other people, and the nature of the world that have been altered by the trauma.

In dissociative disorders related to childhood adversity, the DBS model postulates that early childhood trauma, disturbances in attachment, or both (see section on Type D attachment and betrayal trauma) disrupt the processes essential to self-modulation and the developmental integration of new behavioral states.

Taxon Model of Pathologic Dissociation

Recent interest in the *taxon model* resurrects a century-old debate about whether dissociation occurs along a continuum proceeding from normal to pathologic or whether pathologic dissociation, such as dissociative identity disorder, represents a distinct type (a taxon) of psychological organization. As data with dissociation measures accrued across clinical and community-based populations, it became apparent that there existed a distinct group of high-scoring individuals. These observations led to research on the possible existence of a dissociative type of individual differing significantly on dissociation measures from other individuals, irrespective of their psychiatric status. Using newly developed taxonometric approaches, researchers examined item-response data for manifestations of a latent class variable. They identified eight DES items (Table 20–9), composing the DES-T (Taxon) subscale, that robustly differentiate the most severely dissociative patients from other psychiatric patients and normal controls, who almost never endorse these items.

Table 20–9.
Dissociative Experiences Scale Items Comprising the Dissociative Experiences Scale Taxon

Subjects are asked to rate the percentage of time, from 0% to 100%, that they have the experience.

3. Some people have the experience of finding themselves in a place and having no idea how they got there.
5. Some people have the experience of finding new things among their belongings that they do not remember buying.
7. Some people sometimes have the experience of feeling as though they are standing next to themselves or watching themselves do something, and they actually see themselves as if they were looking at another person.
8. Some people are told that they sometimes do not recognize friends or family members.
12. Some people have the experience of feeling that other people, objects, and the world around them are not real.
13. Some people have the experience of feeling that their body does not seem to belong to them.
22. Some people sometimes find that, in one situation, they may act so differently compared to another situation that they feel almost as if they were two different people.
27. Some people sometimes find that they hear voices inside their head that tell them to do things or comment on things that they are doing.

Adapted with permission from Waller NG, Putnam FW, Carlson EB. Types of dissociation and dissociative types: A taxonometric analysis of dissociative experiences. *Psychol Methods*. 1996;1(3):300–321. doi:10.1037/1082-989X.1.3.300. Copyright © 1996, American Psychological Association.

Studies have subsequently affirmed that a taxonic approach is a clinically useful way to identify a subgroup of dissociative patients within any given diagnosis (e.g., the dissociative subtype of PTSD) who differ from the rest of the group on symptoms and features. In psychophysiological investigations of dissociation, simply dividing samples into high and low scorers has identified different physiologic and cognitive responses to traumatic reminders. The taxon model implies a different developmental scenario than the continuum model as well as a different approach to treatment. The typologic differences between high dissociators and virtually everyone else supports a genetic predisposition (see section on Genetics) and/or a fundamentally different early developmental trajectory that has long-lasting effects on brain connectivity. Research linking dissociation with type D attachment disturbances (discussed later in this chapter) offers another potential environmental factor that may mediate the relationship between genetic diathesis and expression of pathologic dissociation. A large, prospective, general population study from Finland found a point prevalence for the dissociative taxon of 3.4%, consistent with retrospective re-analyses of DES data.

Dissociation as a Transdiagnostic Process

In a recent meta-analysis of 219 studies of dissociation measured by the DES, involving over 15,000 patients in 19 diagnostic categories, Lisa Lyssenko and colleagues concluded that dissociation is an important *transdiagnostic process* that often indicates poorer outcome to standard treatments and should routinely be assessed. The ACE Studies, the National Comorbidity Studies, and other research found that at least subgroups of many common psychiatric disorders are associated with higher rates of childhood trauma. Thus, intuitively, dissociative symptoms likely would be present among many diagnostic groups. Mean DES scores were highest, consistent with the pathologic dissociation taxon model, for dissociative identity disorder (48.7), followed by all dissociative disorders combined (38.9), PTSD (28.6), borderline personality disorder (27.9), functional neurologic symptom disorder (conversion disorder) (25.6), anorexia nervosa (24.1), and bulimia nervosa (22.0). Mean DES scores for nonclinical populations tend to cluster around 8 (see Table 20–10 for the full range of scores).

Table 20–10.
DES Mean Scores Across Diagnostic Groups

- *Dissociative identity disorder: 48.7
- *Dissociative disorders: 38.9
- *PTSD: 28.6
- *Borderline personality disorder: 27.9
- *Conversion disorder: 25.6
- *Anorexia nervosa: 24.1
- *Bulimia nervosa 22.0
- *Substance-related and addictive disorders, including gambling disorder: 19.9
- *Alcohol use disorder: 19.7
- *Non-conversion somatic symptom disorders: 18.8
- Schizophrenia: 17.8
- *Other substance-related disorders 17.7
- Panic disorder 15.6
- Depressive disorders: 15.3
- OCD 15.3
- *Anxiety disorders: 15.2
- Bipolar disorder 14.8

Disorders marked with an asterisk have at least subgroups with high rates of childhood maltreatment and neglect, as well as adult traumas.

Adapted with permission from the Lyssenko L, Schmahl C, Bockhacker L, Vonderlin R, Bohus M, Kleindienst N. Dissociation in psychiatric disorders: a meta-analysis of studies using the dissociative experiences scale. *Am J Psychiatry*. 2018;175(1):39.

DES sub-group factor scores could only be calculated for the cohorts of individuals with dissociative disorders, borderline personality disorder, and schizophrenia. Dissociative disorder patients had significantly higher mean scores on all DES factors: amnesia (12), absorption (12), and depersonalization/derealization (17). The mean scores for the other diagnostic groups were amnesia (4) absorption (4) and depersonalization/derealization (4,5); individuals with schizophrenia scored higher than borderline personality disorder patients. These subgroup findings are consistent with studies using other measures (e.g., the SCID-D) that compared dissociative identity disorder and borderline personality disorder (see section on Differential Diagnosis of Dissociative Identity Disorder), showing overall lower scores and a different pattern of scores on standardized dissociation measures for individuals with borderline personality disorder. Apart from individuals with dissociative disorders, there are no data on the pathologic dissociation taxon in any diagnostic group other than PTSD. These findings suggest that elevated dissociation scores across diagnoses may, in part, represent different patterns than that of the dissociative taxon.

The Hierarchical Taxonomy of Psychopathology (HiTOP) model uses the observed covariation of symptoms to organize psychopathological syndromes dimensionally by grouping related symptoms and their components/subtypes. Comorbidity is assessed by combining co-occurring syndromes into spectra. In a prospective study of 486 community volunteers, Stephanie Ellickson-Larew and colleagues assessed dissociation based on the HiTOP model. Participants, approximately half of whom were African American, were evaluated at baseline with two 3-hour sessions for self-report measures and a diagnostic clinical interview, respectively. Three hundred participants returned an average of 9.5 months later to complete 3 hours of self-report measures. The authors derived three facets for dissociation: depersonalization/derealization, memory disturbances (amnesia), and fantasy, the latter related to absorption. Bivariate and multivariate analyses across both self-report and interview data showed that all the dissociation facets related broadly across psychopathology dimensions, including internalizing, externalizing, and psychoticism. When controlling for neuroticism, memory disturbances and depersonalization/derealization (the latter to a lesser degree) maintained these relationships. These patterns established the transdiagnostic nature of dissociation.

At follow-up, depersonalization/derealization and memory disturbances maintained strong predictive power across symptoms; fantasy was predictive of mania. After controlling for neuroticism, depersonalization/derealization maintained strong predictive power for internalizing symptoms and had substantial predictive power for increases in psychopathology, particularly suicidality. The memory disturbances facet was the strongest correlate of current psychopathology. Fantasy was weaker in predictive power, more for externalizing symptoms.

This study strongly supported dissociation as an important transdiagnostic factor that was related to all diagnoses tested in this model, including PTSD, fear- and distress-based disorders such as anxiety and mood disorders, substance abuse, OCD, mania, and psychotic disorders. These authors argued that dissociative symptoms may be a cross-cutting indicator of the effect of trauma on psychopathology in general as well as associated with poorer treatment outcome. This group strongly recommended that dissociation be assessed systematically in all diagnostic assessments.

Dissociation: A Major Risk Factor for Suicide Attempts and Nonsuicidal Self-Injury

All studies that have evaluated the question have shown that dissociation and dissociative disorder diagnosis are independent risk factors for nonsuicidal self-injury and suicide attempts, particularly multiple suicide attempts. Clinical studies of dissociative identity disorder ascertain that 90% to 100% of these patients report a past or current history of suicidal ideation; 60% to 78% endorse prior, often multiple suicide attempts; 78% endorse current or past nonsuicidal self-injury. Clinical case series report that patients with generalized dissociative amnesia are at high risk for suicide attempts if the amnesia is too quickly broken down and there are inadequate safeguards to protect the patient.

Logistic regression analysis of dissociative disorder patients in an inner-city clinic found that dissociation was the only remaining significant relationship with multiple suicide attempts, after PTSD, substance abuse, trauma history, and borderline personality disorder were entered into the analysis. Meta-analyses of suicidality and nonsuicidal self-injury find that higher DES scores and dissociative disorder diagnosis predict suicide attempts and nonsuicidal self-injury. Patients with a dissociative disorder diagnosis are more likely to report previous suicidal ideation and nonsuicidal self-injury compared to other patient groups. In studies using a DSM-based methodology to diagnose dissociative disorders, results remain highly significant for both outcomes. Accordingly, dissociation is an independent risk factor for dangerousness to self and should be assessed in all clinical evaluations for safety issues and in all research studies of suicide risk.

Dissociation as a Response to Trauma

As noted in the historical section, there is a longstanding debate about the relationship of dissociation to antecedent traumatic experiences. Since the 1980s, multiple lines of evidence have linked dissociative disorders with antecedent trauma, building a strong case for its role as a necessary antecedent to the development of pathologic dissociation. For each of the DSM-5-TR dissociative disorders, there now exist multiple independent case series, including non-Western samples, documenting unusually high rates of trauma in these patients, including case series where one or more traumatic events have been verified for most patients. Several hundred peer-reviewed studies have found significantly higher levels of dissociation, as measured by the DES and other well-validated instruments, in traumatized groups compared to nontraumatized clinical and general population samples.

Across multiple international, cross-cultural studies dissociative disorders are associated with more severe and chronic trauma, early-life trauma, and attachment disturbances with primary childhood caregivers. Early life adversities strongly associated with dissociative disorders include childhood sexual abuse and physical abuse, emotional abuse, and neglect and can include nonmaltreatment forms of childhood trauma such as multiple, painful medical procedures and wartime dislocation. Survivors of more severe and chronic adult trauma, including prisoners of war, torture survivors, women in prostitution and trafficking, and women subjected to intimate partner violence, show elevated dissociation scores. Recently, Ruben Vonderlin and colleagues performed a meta-analysis of 65 DES studies with 7,532 individuals reporting abuse or neglect. A childhood history of abuse or neglect predicted significantly higher DES scores compared to nonabused subjects. Childhood sexual and physical abuse were associated with higher DES scores, compared to emotional abuse and neglect, although in these groups dissociation scores were significantly higher than in subjects without a childhood trauma history. The combination of physical and sexual abuse resulted in even higher mean DES scores. Higher DES scores were related to earlier onset of abuse, longer duration of abuse, and parental abuse. These were highly intercorrelated, consistent with other studies of intrafamilial maltreatment (e.g., in the second National Comorbidity Survey), finding in these highly pathologic families such strong intercorrelations between all abuse types and family dysfunction that these cannot be readily parceled out.

Dissociation is instrumental in shaping outcomes related to traumatic experience and psychophysiological responses to future stress and trauma. Peritraumatic dissociation—including symptoms of depersonalization/derealization, disorientation, time distortion, trance- /automaton-like behavior, and amnesia measured right after traumatic events (or described retrospectively)—predicts subsequent PTSD. This can help identify individuals at risk for PTSD after acute trauma. Dissociation has been shown to mediate the relationship between antecedent trauma and subsequent psychopathology, including PTSD. This means that—once the level of dissociation is statistically controlled—the strength of the relationship between trauma and psychopathology significantly decreases or disappears. High levels of dissociation also alter the psychophysiological responses of traumatized individuals to traumatic reminders. As noted above, multiple lines of evidence support the existence of a dissociative subtype of PTSD (see chapter on PTSD). This subtype is characterized by earlier and more severe trauma, more severe PTSD symptoms, specific autonomic and neurobiologic circuitry patterns in response to trauma scripts, and possible SNPs associated with the subtype, among other differences.

Dissociation in Complex and Dissociative PTSD. In the ICD-11 field trials of complex PTSD, individuals with this ICD-11 diagnosis reported significantly more types of developmental and lifetime trauma—including neglect—as well as significantly higher overall dissociation scores, but not limited to depersonalization/derealization, when compared with traumatized controls and individuals with only ICD-11 PTSD. In a national U.S. study of over 3,000 treatment-seeking adolescents (ages 12 to 16; 60.5% female) a latent class analysis evaluated the relationship of PTSD to dissociation. Two models were selected; the first, the DSM-5 depersonalization/derealization model, had five classes. A second, expanded dissociation model identified an additional class characterized by dissociative amnesia and detached arousal. This second model significantly improved differentiation of the high-dissociation—high-PTSD group from the high-PTSD—low-dissociation group. The high-dissociation—high-PTSD group represented 14.7% of the sample, consistent with prior studies of the PTSD dissociative subtype. The high-dissociation PTSD and the PTSD groups did not differ significantly on number or types of maltreatment nor on most demographic variables.

Compared to the high-dissociation/high-PTSD group, the dissociative amnesia group reported decreased PTSD intrusion/reexperiencing symptoms, fewer trauma experiences, and more numbing, detachment, and amnesia. The former had high levels of intrusive PTSD symptoms as well as depersonalization/derealization. The authors hypothesized that the dissociative amnesia may lead to underreporting of trauma, as described in adult populations. These findings are consistent with clinical observations that breakdown in dissociative amnesia results in increased PTSD intrusive symptoms.

Studies in Diverse Cultural Settings. The relationship between traumatic exposure and dissociative disorders has been confirmed in diverse cultural settings. Several studies in Africa have examined the association of traumatic exposure and dissociative pathology, especially in regions affected by civil war. In Uganda, for example, separate studies by two groups of investigators found a significant association between episodes of pathologic "spirit" possession and traumatic exposure. Frank Neuner and colleagues interviewed 1,113 youth ages 12 to 25 with pathologic dissociation attributed to spirit possession, most of whom had been abducted as child soldiers. In multivariable analysis, pathologic possession was associated with higher trauma exposure, such as being forced to kill someone, after adjusting for poverty, education, sex, and other factors; spirit possession was also associated with a greater number of traumatic event types. Dissociative possession symptoms were significantly correlated with symptoms of PTSD (Spearman's rho = 0.36) and depression (rho = 0.37); in addition, 44.5% of individuals meeting a threshold criterion for dissociative possession met PTSD criteria, compared to 9.8% of those with low levels or no spirit possession.

In a second study in Uganda conducted in a relatively peaceful post-conflict zone, Marjolein van Duijl and colleagues compared 119 individuals reporting pathologic spirit possession with 71 residents of the same villages without pathologic possession experiences using the Harvard Trauma Questionnaire and the Traumatic Experiences Checklist. Mean scores on both trauma assessments were higher in individuals with pathologic possession and these scores were significantly correlated with the number of spirit possession symptoms and scores on the DES and the SDQ.

In a representative community-based sample of youths ages 11 to 17 in Puerto Rico, pathologic dissociative symptoms on the Adolescent DES were significantly higher among participants who reported severe physical abuse and/or were exposed to violence. Compared to individuals with conversion but without a dissociative disorder, individuals in Turkey who also met criteria for a dissociative disorder were more likely to endorse childhood trauma. Several studies conducted in clinical and nonclinical populations in Brazil found a correlation between DES scores and measures of traumatic exposure.

At the same time, international studies have noted that some forms of dissociation—pathologic possession, psychogenic nonepileptic seizures, conversion disorder with dissociative features, dissociative cultural idioms of distress such as *ataque de nervios* (attacks of nerves)—can be associated with significant but nontraumatic stressors. For example, these include school and family conflicts among youth, troubled romantic affairs, emotional neglect, unemployment, and insecure attachment. It is possible that some cultural characteristics—such as the avoidance of strong negative affectivity to maintain equanimity and/or the development of a more "porous" sense of self open to transpersonal spiritual experiences—may nurture dissociative capacity, which may develop into pathologic dissociation upon exposure to severe but nontraumatic stressors. However, only some of these studies systematically assessed for the presence of childhood trauma that could play a role in predisposing individuals to dissociative disorders once exposed to these nontraumatic stressors. This is similar to descriptions of precipitants to episodes of severe dissociation in the classic cases and case series on generalized dissociative amnesia and dissociative fugue. Here, extreme intrapsychic conflict and/or acute loses or stresses appeared to precipitate most of these events. In virtually every study, however, careful reading showed that these individuals had suffered from severe prior traumas such as wartime combat, experiences of violence or sexual assault, and/or childhood sexual abuse and/or physical abuse. Further research is necessary to clarify this issue.

Studies of Military Survival School Subjects. Remarkable research conducted by Morgan and colleagues with military Special Operations units under extremely stressful conditions are as close as one can ethically come to an experimental model of trauma-induced dissociation. Military Survival, Evasions, Resistance, and Escape (SERE) training is offered by all branches of the U.S. military and is based on American POW experiences during WWII, and the Korean and Vietnam wars. This involves days of inescapable stressors, including semi-starvation, exhaustion, sleep deprivation, lack of control over hygiene and bodily functions, abusive interrogations, hooding, and lack of control over one's movements, social contact, and communication. Working with both U.S. and European units undergoing SERE and similar training, as well as the grueling Combat Diver Qualification Course, Morgan examined the neurobiologic and psychological effects of intense stress uniformly applied to drug-free, healthy, nonclinical subjects. The neuroendocrine stress effects measured in these soldiers were equal to or exceeded those reported for individuals subjected to life-threatening experiences. These stressful experiences were positively associated with dissociation measures. Significant differences were found between pre- and posttest scores with the greatest effects for depersonalization items (e.g., looking at the world through a fog, time slowing, "spacing out"). Life threat from prior trauma was significantly correlated with pre- and poststress dissociative symptoms. In general, the prestress pattern of elevated dissociation predicted poorer performance during these high-stress situations, including problems with attention, working memory, and problem-solving. Dissociation scores also accounted for 41% of the variance in health complaints after the stress experience. The significant relationship between the pre- and poststress change in dissociation score and somatic symptoms is congruent with many clinical studies reporting linkages between somatization and dissociation.

Morgan's systematic studies are consistent with observational reports of the high frequency of dissociation, especially depersonalization symptoms, in normal individuals exposed to life-threatening stress. The ubiquity of acutely high levels of dissociation during extreme stress indicates that other factors, perhaps prior life-threatening trauma, are operative in the peritraumatic dissociation–PTSD linkage discussed previously. The highest-dissociation scores were associated with intense subjective appraisal of life threat from past trauma. However, one highly traumatized subgroup, Special Forces soldiers, had *low life-threat appraisals of their past trauma* and, consequently, the *lowest pre- and poststress change in dissociation scores.* As described in the neurobiology section, after interrogation stress during survival training, Special Forces soldiers showed significantly higher levels of Neuropeptide Y (NPY) compared with soldiers in the general forces, suggesting a role of NPY in regulation of anxiety and/or dissociation during high-stress situations. Clinically, such individuals report the capacity to remain calm and focused during extreme danger, even when suffering from PTSD that is disruptive in other contexts. They describe the ability to ramp down their stress response, achieving resolute, unemotional clarity evaluate and manage threat.

In summary, several independent lines of evidence point to a causal relationship between trauma and dissociation. This relationship is apparent for many forms of trauma and has been found across various cultural settings in which it has been investigated. Significant levels of dissociation in the immediate aftermath of traumatic events appear to predispose to developing PTSD and related psychopathology, although additional factors involving the psychological appraisal of prior trauma also appear to be involved. Dissociation appears to be a major mediating process between traumatic experiences and subsequent psychopathology and thus is an important target for prevention and early intervention efforts. Stress-resistant individuals can be characterized, in part, by their lack of dissociative responses to significant stressors and/or their ability to—consciously or unconsciously—modulate their stress response during threat.

Culture, Dissociative Disorders, and Sociocognitive Phenomena

Cultural and sociocognitive factors shape the presentation of all psychiatric disorders. Certain symptoms may be emphasized above others when experiencing or describing disorders (e.g., psychological vs. somatic aspects of depression) and specific experiences may be common in certain settings but rare in others due to societal-structural differences as well as diverse meaning traditions. For example, conversion-like dissociative symptoms are more common in India than in certain affluent settings in the United States. Attributions of pathologic experience also vary. Nineteenth-century individuals with paranoid schizophrenia did not have delusions of being controlled over the internet.

The inclusion in the DSM-5 of the pathologic possession form of dissociative identity disorder was based on cross-cultural research that described sociocognitive differences in the structuring of the disorder between secularized (often Western) contexts and nonsecularized (often non-Western) ones. In the literature on dissociative identity disorder, various developmental, social, cultural, intrapsychic, interpersonal, and cognitive factors, including the impact of psychotherapy, are hypothesized to account for the *secondary structuring* of the dissociative process into the individualized identities characteristic of a given person in a specific cultural context. Influences may include family factors (e.g., contradictory parental demands for the child to be a boy or a girl); religious and other cultural and social influences (e.g., characters from books, films, television, or video games); the accepted existence of nonmaterial entities (e.g., gods, demons, possessing spirits); and specific traumatic experiences (e.g., being sexually trafficked in childhood and having to create identities to fit the fantasies of “purchasers”; being a child soldier); among others.

In clinical case series, developmental factors affect the presentation of the dissociative identity disorder self-states in important ways. Usually, children with the disorder manifest less overt switching and self-states are described more as elaborated and autonomous imaginary companions. During adolescence, the person may develop more complex, secondarily structured self-states that relate to many aspects of psychological life. Intelligence and creativity may lead to greater elaboration of identity states in some individuals with dissociative identity disorder, as does posttraumatic withdrawal from outside reality due to trauma and neglect; these may lead to the creation of more elaborate fantasy worlds in which to “live.” Cross-cultural research could help describe childhood forms of pathologic possession in cultures that support these belief systems. Future research may trace in greater detail how culture interacts with biology in a developmental timeframe and in the context of an individual’s exposure to experiences that shape the expression of culturally diverse phenomenologies of dissociative psychopathology (see Developmental Models of Dissociation).

Type D Attachment

Attachment pathology in infancy has been linked with later dissociation in prospective studies. Disorganized (Type D) attachment in infancy as coded by the Strange Situation protocol (between 12 and 20 months) is manifest as conflicted, disoriented, and/or fearful behavior toward the parent. Disorganized attachment is most common in severely maltreated children but can occur with nonabusive traumas, losses, and unresolved parental trauma and loss. Frightened, frightening, or dissociative caregiver behavior is strongly implicated in the development of disorganized attachment; and in the case of the dissociative parent, may contribute to the multigenerational transmission of trauma and dissociation.

Two prospective, longitudinal research studies following families at high risk for maltreatment found that infants with disorganized attachment had significantly higher-dissociation scores in late adolescence and early adulthood compared with infants classified with other categories of attachment. In one study, Type D attachment accounted for almost half of the variance in dissociation scores 20 years later. Disorganized attachment in young adulthood partially mediates the relationship between the severity of childhood maltreatment and dissociation scores. A prospective, case-control study following sexually abused and nonabused comparison girls into adulthood found that increased maternal dissociation predicted harsh parenting of the type associated with the development of disorganized attachment in offspring. Consistent with these childhood studies, a study of 554 healthy adults found a large effect size (d = 1.43) for the combination of verbal/emotional abuse and witnessing domestic violence in childhood in predicting dissociation, comparable to the effect of familial sexual abuse.

Betrayal Trauma

Beyond the multiple associations among various types of traumatic experiences and dissociative sequelae, Jennifer Freyd has formulated a specific *Betrayal Trauma Theory* (BTT), derived from social psychological principles and developmental attachment theory (see section on History). BTT began as an attempt to explain childhood amnesia for trauma as related to the meaning of the negative event as a betrayal by a trusted, needed other, rather than simply related to the frequency and aversiveness of traumatic experiences. Betrayal influences how the event is processed and remembered. In this model, supported by data in clinical and nonclinical samples, a child is more likely to develop amnesia when subjected to abuse by a family member or another person whom the child trusts or on whom the child is dependent. The amnesia is protective of the child's developmentally mediated need for attachment, allowing for preservation of emotional and cognitive growth despite the abuse.

It is hypothesized that psychological schema about the abuse (i.e., the betrayal) are subjectively sequestered from mental mechanisms that control attachment and attachment behavior. BTT is consistent with the DBS model and the Type D attachment findings. It elucidates cognitive factors that may influence dissociation of state-related information. Traumas involving betrayal are processed differently at the cognitive level from other autobiographical experiences, leading to state-dependent memory deficits. BTT helps explain the lower prevalence of amnesia for wartime trauma, natural disasters, and the Nazi Holocaust, among others, as compared to intrafamilial childhood sexual or physical abuse. Difficulty in being able to "share" the traumatic experiences also may contribute to amnesia for trauma. In addition to the neurobiologic factors that inhibit verbal access to such events, they are less likely to be shared because of threats, shame, developmental inability to find words for the experience, reliance on adults to define childhood experience, and intense confusion and emotional arousal engendered by the abuse.

Compared with traumatized individuals not reporting betrayal, those reporting betrayal show higher levels of dissociation, particularly dissociative amnesia, alexithymia, anxiety, depression, nonsuicidal self-injury, and physical health complaints. In a community study, “double betrayal” was defined as sexual abuse by a caregiver followed by a negative response to disclosure. This was associated with more severe dissociative amnesia and higher rates of revictimization.

Betrayal trauma and lack of integration of betrayal information—called *betrayal blindness*—have also been studied across diverse types of experience. These include personal betrayals, such as marital infidelity, as well as institutional betrayal (e.g., the Catholic Church in the clergy abuse cases) and cultural betrayal (e.g., a member of a cultural minority group feeling betrayed by another member of the same group). Betrayal blindness may be a common human response to these situations, with intense emotional reactions when awareness of the betrayal becomes less dissociated from the person’s awareness.

A unified information processing and betrayal model predicts the relative likelihood of amnesia under varying circumstances, such as high threat/low betrayal, high threat/high betrayal, and so forth. However, as described above, various factors may lead to disturbances of autobiographical memory for traumatic experiences, although betrayal trauma may be particularly salient as a factor in the lack of integration of trauma-related information. BTT may also help explain liability to revictimization in survivors of childhood trauma. When the survivor encounters potentially exploitative individuals, the “betrayal detection system” becomes relatively unavailable to consciousness. This, as well as the over-learned freeze response to threat developed in childhood (see section on Neurobiology), may explain how dissociation is a risk factor for revictimization.

Developmental Model of Dissociative Disorders

Modern interest in a developmental model of dissociative disorders, especially dissociative identity disorder, can be traced to pioneers who examined children with the condition, such as Cornelia Wilbur and Richard Kluft. Long-forgotten case reports were later discovered in which 19th-century clinicians used phrases in Latin to hint at incest and unspeakable familial trauma. Small case series and reviews of childhood dissociative disorders appeared in the mid-to-late 1980s, with more systematic studies coming in the 2000s. As with the adult case series, histories of severe trauma were commonly noted (see section on Dissociative Identity Disorder).

The core phenomenology of dissociative disorders in children is largely like that in adults (see section on Culture, Dissociative Disorders, and Sociocognitive Phenomena). However, as with most childhood manifestations of lifelong psychiatric disorders, the day-to-day symptoms vary with the child's age. In general, older children and adolescents are more overtly symptomatic than younger children. In part, this may reflect the ability of older children to better report subjective distress as well as their greater opportunity to manifest psychopathology in a variety of settings.

In a pooled sample of 177 child and adolescent dissociative disorder cases, amnesias, identity disturbances, and auditory hallucinations increased with age, whereas trance-like and "spacey" behavior was ubiquitous across all age groups. Related psychopathology, such as suicidal ideation, self-mutilation, and somatization, increased with age in parallel with dissociative symptoms. Boys and girls did not differ on dissociative symptoms but girls were significantly more symptomatic for anxiety and phobic symptoms, PTSD, sleep disturbances, sexual acting out, and somatization. Boys showed more externalizing behaviors such as aggression and violence. These findings closely parallel the gender differences in adult clinical profiles. Here males with dissociative identity disorder report higher rates of violence, criminal involvement, and alcohol and substance use disorders. Children and adolescents diagnosed with dissociative identity disorder were the most symptomatic across all age groups, which is consonant with the clinical belief that dissociative identity disorder is the most severe of the dissociative disorders.

Validated child and adolescent dissociation measures have been administered to various clinical and nonclinical samples (see

Table 20–1). Scores were significantly higher for traumatized (typically child abuse) versus nontraumatized respondents across all age groups. As in the adult studies, levels of dissociation were also significantly associated with more general psychopathology. Among maltreated preschoolers, there were robust correlations with internalizing and externalizing behavior problems for girls and boys, respectively. A comparison of maltreated and nontraumatized preschoolers matched on family constellation and demographics found that the maltreated group had significantly increased levels of dissociation a year later, with the physically abused children accounting for the greatest increase in scores. The controls showed substantial decreases in their dissociation scores over the same period, in line with the often-reported decrease in dissociation scores with age in normal children.

Studies have explored the possible mediating role of dissociation in the development of psychopathology in sexually abused children and adolescents. Measuring psychopathology with a variety of standard measures, the main effect of sexual abuse on behavioral problems disappeared when levels of dissociation were controlled for, as measured by a self-report measure (A-DES) or by observer report (Child Dissociation Checklist [CDC]). The possibility that dissociation is a critical mediating variable for subsequent psychopathology has significant implications for early intervention with maltreated and traumatized children. Assessment of dissociation should be a standard part of the evaluation of traumatized children and adolescents and significant elevations should be addressed in the treatment plan. This would include diagnostic assessment for specific dissociative disorders, for example, dissociative identity disorder, and development of a treatment plan that specifically targets dissociative symptoms/disorders. The latter is critical as studies of adult trauma survivors show that, without specific attention to dissociation, standard trauma-informed treatments are likely to be partially effective at best.

Search for Developmental Precursors and Substrates of Dissociation

Imaginary Companionship. The focus on child dissociative identity disorder led clinical investigators to initially focus on childhood fantasy phenomena, such as imaginary companionship, as possible developmental precursors for dissociative disorders. Imaginary companions are reported in 20% to 60% of normal children depending on the age of the child and the definition used. Normal imaginary companionship is widely regarded as benign and is commonly considered to be a sign of creativity in young children but it becomes increasingly suspect in older children and adolescents and is thought to be always pathologic in adults. In dissociative children, the rates of imaginary companionship range from 42% to 84%, with the highest rates reported for children diagnosed with dissociative identity disorder.

Comparing the characteristics of imaginary companions in maltreated and non-maltreated boys showed that maltreated boys averaged 6.4 entities compared to 2.5 entities for the non-maltreated boys. The latter first appeared between 2 and 4 years of age and had all disappeared by 8 years of age. The imaginary companions reported by the boys in residential treatment, however, served other functions, including as (1) helpers and comforters, (2) powerful protectors, and (3) family members. In some cases, the boys reported feeling as if they had lost control and their imaginary companions took over their behavior. The maltreated boys reported that their imaginary companions were still present at a mean age of 10.6 years, well after they had disappeared in the nonmaltreated boys. The imaginary companions of the maltreated boys often had names such as *God, the devil,* and *guardian angel.* Authorities theorize that some of the imaginary companions found in maltreated children eventually represent the dissociative identity disorder self/identity states.

Developmental Mediation of Psychopathology by Dissociation

The negative effects of elevated dissociation on child development may operate by hindering the development of a sense of self, emotional regulation, impulse control, information processing, and coping with stressors. Children assemble and seek to integrate a complex, multidimensional sense of self over the course of development. Dissociative components, such as autobiographical amnesias, depersonalization, and passive influence experiences, interfere with the integration of self and the development of a unified sense of self-agency.

Dissociative amnesias also disrupt the child's understanding of cause-and-effect sequences, so that negative or risky behaviors may appear unconnected from their consequences, experienced as "coming out of the blue." This lack of connection can interfere with the ability of dissociative children to learn from experience. The impact of dissociation on the metacognitive function of self-monitoring behavior can disrupt integration of experiences across contexts, further complicating the child's ability to learn and to practice self-control, particularly in the context of stressors. Increased dissociation is well correlated with increased aggression, impulsivity, and poorer social skills in children and adolescents.

Prospective studies of sexually abused girls compared with carefully matched controls show that dissociation is negatively associated with competent learning and overall classroom performance and is strongly predictive of school avoidance. The combination of cognitive dysfunctions and negative school trajectories puts the dissociative child at an academic disadvantage that has profound implications for adult attainment.

Chronic depersonalization, closely associated with emotional numbing in PTSD patients, is thought to promote a sense of detachment from self that fosters risky and self-destructive behavior, such as nonsuicidal self-injury (see section on Dissociation and Dangerousness to Self). Dissociative patients often describe self-inflicting pain in attempts to break through profound states of depersonalization and numbness. At other times, they report feeling nothing as they cut or burn themselves. Alienation from self is also thought to play a role in the high rate of suicide attempts in dissociative patients. As noted above, dissociation has been implicated in predisposing an individual to revictimization. In aggregate, these experiences expose dissociative individuals to further traumatization, which takes a cumulative toll over the individual's life. Identification of protective environmental factors may help stimulate intervention models that prevent the development of significant dissociation in acutely traumatized children.

4-D Model of Trauma-Related Dissociation. The "4-D Model of the Trauma-Related Dissociation" has been articulated by Paul Frewen and Ruth A. Lanius. This framework, based on work from philosophy and consciousness studies, elaborates the concept of trauma-related altered states of consciousness (TRASC) introduced by the DBS model. The 4-D model proposes a four-dimensional neurophenomenology as a means of describing specific experiential states: (1) time and memory, (2) verbal/thought, (3) embodiment, and (4) emotion. The neurophenomenologic approach is based on subjects' personal descriptions of experiences that are correlated with neurobiologic measures. Imaging studies of individuals with PTSD and in community samples have shown differences in fMRI patterns between individuals with and without dissociative symptoms, irrespective of diagnosis. Those without dissociative symptoms have shown patterns of distress associated with the neurobiologic normal waking state of consciousness (NWC) as opposed to TRASC, characteristic of distress in individuals with the dissociative subtype of PTSD and with dissociative experiences related to trauma. The 4-D model, consistent with the developmental DBS model, provides a testable way to link neurocircuitry patterns to abnormalities of emotion regulation and self-referential appraisal. These, along with other factors such as attachment disturbances (see sections on Betrayal Trauma and Type-D Attachment), may be important in development of dissociative disorders such dissociative identity disorder and depersonalization/derealization disorder.

Hypnotic Model. Recent reviews have revived the hypnotic model of dissociation as particularly etiologic for dissociative identity disorder. This model hypothesizes that a traumatized child uses their innate hypnotic capacity to induce autohypnosis as a defense against overwhelming or repetitive traumatic experiences. In dissociative identity disorder, with continuous use, the autohypnotic state is personified into self-states. Support for this model includes that dissociative patients are the most highly hypnotizable group of individuals, especially those with dissociative identity disorder. Many clinical phenomena associated with pathologic dissociation—such as trance states, age regression, auditory hallucinations, and amnesias—can, in a hypnotic state, be transiently produced in highly hypnotizable, normal individuals. Finally, two studies suggested that childhood trauma might increase hypnotizability.

Among problems with this model, it reasons retrospectively from adult patients to hypothesized mechanisms in children. Hypnotic capacity is difficult to assess in children at least until early adolescence. More systematic research has dispelled the notion that childhood trauma generally increases hypnotizability. At least six studies found no differences between traumatized and nontraumatized adult and child samples. In clinical and general population samples, there are no discernible dose effects between measures of trauma severity and hypnotizability. However, a small group of individuals, termed *double dissociators,* score high on measures of dissociation and hypnotizability. These individuals generally have histories of earlier and more severe trauma. This observation is most compatible with the taxonic model of pathologic dissociation and may explain the greater hypnotizability of dissociative identity disorder patients. There may be different types of hypnotizability, with high dissociators making up a distinct type. However, other hypnotizability studies have found subgroups of highly hypnotizable people in nondissociative samples. In dissociative identity disorder, autohypnosis may be involved in secondary elaboration of self-states, in attempts to attenuate intrusive symptoms, and as a self-protective process during maltreatment. Autohypnosis is a core aspect of dissociative trance disorder (see section on Other Specified Dissociative Disorders).

Psychoanalytic Models. As described in the section on History of Dissociation, psychoanalytic theory has had difficulty with the concept of dissociation since the unfortunate schism of Freudian and Janetian ideas in the late 19th and early 20th centuries. In addition, in psychoanalytic writings, the term trauma was broadened to refer to almost any developmental stressor, a conceptual difficulty recognized by some psychoanalytic writers such as Anna Freud when she studied British children with wartime trauma caused by bombings of the United Kingdom by Germany during World War II. Also, psychoanalytic theory became focused on the intrapsychic appraisal and symbolization of events rather than on interplay between the severity of traumatic events and subjective appraisal.

Other than in the work of David Rappaport and his associates at the Menninger Clinic in the 1940s that attempted to integrate Freudian and Janetian concepts, as well as a few papers on “hypnoid states,” until recently most psychoanalytic writing has centered on the concept of “splitting” as equivalent to dissociation. On the other hand, psychoanalyst Steven Marmer, based on his experience treating dissociative identity disorder patients, posited intrapsychic transitional object phenomena as a specific developmental aspect of the creation of dissociative identity disorder self-states. Some theorists of the relational school of psychoanalysis now view dissociation as a general psychological phenomenon that underlies all neurotic, personality disorder, and posttraumatic symptoms. However, in general, psychoanalytic theory appears to have difficulty viewing dissociation as a specific response to trauma, both developmentally and in adults, and as a specific intrapsychic defense against subsequent awareness of trauma and its impact on the person.

Structural Dissociation Theory. Onno van der Hart and Ellert Nijenhuis have attempted to develop a comprehensive theory, *Structural Dissociation Theory* (SDT), postulating that a basic dissociative “split” in personality accounts for posttraumatic manifestations and disorders. SDT is especially popular in Europe as an etiologic theory and treatment model. This theory proposes that dissociative disorders, PTSD, somatoform disorders, and some cases of borderline personality disorder are fundamentally “dissociative” in that severe traumatic experiences led to a “split” between an *Apparently Normal Personality* (ANP) and an *Emotional Personality* (EP). The hypothesized ANP is constricted and avoidant, unaware of traumatic memories, externally focused on daily life, and phobic of internal experiences. The EP is filled with the pain and distress of unintegrated, unprocessed traumatic memory. SDT theorists have been the guiding force in the ICD-11 diagnostic criteria for dissociative identity disorder and partial dissociative identity disorder.

In this theory, PTSD results from a single EP and the ANP; more complex dissociative disorders involve “secondary” and “tertiary” structural dissociations with, respectively, multiple EPs and then multiple ANPs. “Tertiary structural dissociation” results in the development of dissociative identity disorder and secondary structural dissociation in “less complex” dissociative disorders. SDT follows DBST in conceptualizing dissociative identity disorder as resulting from an early developmental failure to consolidate a consistent sense of self across different emotional states, social contexts, and relationships. However, SDT postulates that the alternation of approach/avoidance behavior of infants and toddlers with Type D attachment patterns is an expression of an alternation of an “ANP” and an “EP,” an adult-based construct not consistent with accepted models of the psychobiological development of human children,

Problems with SDT include its overly reified conception of the mind as a structure, with a concretized “personality” that is “split” and made up of “parts.” Dissociative disorders are better conceptualized as disorders of self-organization, with development of complex dynamic self-state systems, not a lesion of the “personality.” SDT is based on reasoning backward from adult dissociative identity disorder rather than forward from the developmental perspective of the DBST. It minimizes the complexity of child development by positing hypothesized adult psychological “structures” in infants and toddlers. In addition, meta-analyses of PTSD and dissociation are consistent with the model that dissociation is either best understood as a subtype or a component of PTSD, not that dissociation underlies all forms of PTSD. Other problems include the unwieldy, reified, and difficult-to-operationalize conceptual language (e.g., “apparently,” “normal,” and “personality”) that minimizes the notion of dissociation as a process or set of processes in favor of metaphors for “structure.” The SDT model of dissociation posits universally rigid amnestic barriers between ANP and EP and sequential activation of differentiated states rather than the shifting, variable symptom presentation that characterizes most dissociative identity disorder individuals with overlap, interference, influence, and intrusion between states, as well as cooperation among self-states/systems.

Iatrogenic, Sociocognitive, Fantasy, and Sleep Models of Dissociation and the Dissociative Disorders. Despite research data on the relationship between trauma and dissociation, some authorities continue to question the existence of dissociation, the trauma–dissociation relationship, and/or the validity of the dissociative disorders, particularly dissociative identity disorder and dissociative amnesia (see section on History). This skepticism has taken for forms: *the iatrogenic model* (IM), the *sociocognitive model* (SCM), the *fantasy model* (FM*),* and, most recently, *the sleep-dissociation model* (SDM).

IATROGENIC AND SOCIOCOGNITIVE MODEL OF DISSOCIATION. The IM posits that dissociative identity disorder and dissociative amnesia are not authentic psychiatric disorders but instead the product of direct or indirect cuing and suggestion during psychotherapy by credulous therapists to highly “suggestible,” hypnotizable, and “fantasy-prone” patients. These are supposed to be characterized by dependent, borderline, and/or histrionic traits and to be desperate for acceptance and a sense of identity. The therapist, through an ideologic belief in dissociative identity disorder or “repressed memory,” engages in a “risky” “recovered memory therapy” (a nonexistent, pejorative “treatment” created as a forensic and public relations gimmick) that focuses only on retrieval of as many trauma memories as possible using "suggestive techniques," including hypnosis, and pharmacologically facilitated interviews. The therapist thereby implicitly and/or explicitly cues and encourages the patient to undertake the role of the trauma survivor with recovered repressed memories and/or dissociative identity disorder. The patient develops a real, albeit iatrogenic, factitious disorder, now truly believing in the roles that they enact. Advocates of these models posited that dissociative disorder cases were only found "in a few centers" run by "enthusiasts" who churned out an "epidemic" of dissociative identity disorder and recovered memory cases.

The SCM is similar to the IM, but adds a factor related to the broader cultural and social influence—often promulgated by the internet, mass and social media, self-help books, and/or victim support groups—in which people, primarily in North America, come to accept the possibility of *multiple identity enactments* or a history of forgotten childhood victimization, or both, as part of an established social role. In this view, therapy may not be a necessary factor in the development of a dissociative disorder; a susceptible person may develop the disorder simply through more general social influence. The SCM for dissociation is constructed as an analog to the Sociocognitive Model of Hypnosis, where hypnotizability is not related to an altered state of consciousness or discontinuities in experience but solely is a learned social role, that of "being hypnotized."

Treatment requires removing the patient from the therapy, ignoring any reference to self-states or trauma memories, educating the patient about the factitious nature of these phenomena, and beginning "real" therapy for other issues. If the patient has become estranged from their family of origin due to belief in the false abuse history, reunification with the family is deemed important as well.

Recent comprehensive reviews of the IM/SCM for dissociative disorders find that virtually no empirical, scientific studies have been performed in clinical populations to examine this model. The original database supporting these constructs consists of a few heterogeneous studies from the 1940s and 1950s, mostly consisting of anecdotal case reports. These demonstrated that some degree of role enactment of an alternative identity occurred with hypnosis, automatic writing, and strong repeated suggestion. Many subjects were the students of the researchers and knowledgeable of the goals of the experiments. The role enactments were limited to the experimental situation. One of the researchers, studying a traumatized soldier, concluded that hypnosis allowed access to authentic, previously dissociated self-states, not to artifacts of the procedure.

Research with thousands of subjects undergoing hypnosis over several decades has shown that a subgroup of highly hypnotizable individuals can exhibit limited, temporary apparent self-states after hypnotic suggestion. These states were phenomenologically dissimilar from clinical dissociative identity disorder and were limited to the experimental situation. The researchers noted that they could not determine whether these procedures uncovered previously existing dissociated aspects of the mind or created them de novo.

A few studies of college volunteers in various hypnotized and nonhypnotized conditions found that highly hypnotizable, “fantasy-prone” students would enact the role of a dissociative identity disorder murderer with amnesia when given a series of robustly explicit suggestions to do so. This research has been criticized on several grounds, including lack of diagnostic controls (e.g., enacting another psychiatric disorder, such as a murderer with schizophrenia); lack of equivalence of the research design with the clinical situation; conflation of dissociative identity disorder with role enactments, rather than the complex polysymptomatic picture described in the clinical and research literature; lack of attention to strong expectancies and demand characteristics of the research design; and failure to account for preexisting dissociated self-states and traumatic experiences.

Finally, the large, international, cross-cultural epidemiologic and phenomenologic studies of dissociative disorders cannot be accounted for by the IM/SCM postulates. Cases were ascertained by administering valid and reliable screening measures and structured interviews to a wide array of patients in myriad settings, hardly specialized “dissociative clinics,” including inpatient, outpatient, partial hospital, state hospital, private psychiatric hospital, substance abuse treatment, and emergency departments in high-, middle-, and low-income countries (see section on Epidemiology).

The Fantasy Model (FM). The next iteration of this skepticism is the FM. In this model, dissociation is not a phenomenon that follows traumatic or overwhelming experiences. It is a personality trait consisting of a mild form of cognitive/executive function impairment that predisposes an individual to being suggestible and generally prone to fantasy and to have a specific liability to confabulated traumatic experiences. FM proponents aver that the relationship between trauma and dissociation is weak or nonexistent. Being fantasy prone was originally described in nonclinical samples, seen in individuals with high hypnotizability, vivid fantasy lives, and a capacity to spontaneously create intense perceptual, sensory, and imagery experiences, often more vivid than actual reality. Fantasizers also described vivid recall of childhood and multiple religious and paranormal experiences.

Systematic Comparison of the Trauma and Fantasy Model of Dissociation

A general critique of the literature on fantasy proneness posits that the basic psychometric inventories for fantasy proneness do not measure a specific construct but rather diverse factors apparently related to potential clinical and normative phenomena, most of which are unrelated to being fantasy prone. Second, fantasy proneness and dissociation measures have been found to overlap. Third, some studies have found a relationship between fantasy proneness and childhood trauma, potentially explaining the correlation between dissociation and fantasy proneness measures.

In 2012, Constance Dalenberg and colleagues published a comprehensive, rigorous meta-analysis that examined eight contrasting predictions from the TM and the FM of dissociation. These findings are summarized in Table 20–11.

This review only included studies with sufficient data to calculate effect sizes, sample sizes greater than 50, and whose design was unlikely to restrict the range of the trauma or dissociation variables. A particularly strong point was a review of the relationship between various suggestibility paradigms and dissociation. Major reviews of the term *suggestibility* cite the multiple definitions and explanations that have been used both for the phenomenon of *suggestion* and the hypothesized trait of *suggestibility.* These have included studies of hypnotic phenomena, unconsciously produced motor behaviors, dissociative possession states, false confessions, susceptibility to advertising, responses to psychotherapy, behavioral conditioning, and memory distortions, among many others. Also, one type of suggestibility may not correlate with other types.

Table 20–11.
Eight Predictions of the Trauma Model versus the Fantasy Model of Dissociation

In contrast to predictions associated with the Fantasy Model, The Trauma Model for Dissociation was confirmed on the following eight predictions:

1. The relationship between trauma and dissociation was consistent and moderate in strength across studies.
2. The relationship between trauma and dissociation remained significant when proven, well-supported, and/or objective measures of trauma were used, not just self-report.
3. Dissociation was temporally related to trauma and trauma treatment; symptoms tended to decrease with time, and were responsive to trauma treatment, rather than inflate and increase with treatment.
4. Dissociation was predictive of trauma history when fantasy-proneness was controlled, but not vice versa.
5. Dissociation was not reliably associated with suggestibility on a series of suggestibility paradigms.
    a. Event suggestibility
        i. For nonautobiographical events (post-event misinformation)
        ii. For autobiographical events (“false memory” creation)
    b. Interrogatory suggestibility (IS) (liability to false confessions)
    c. Imagination inflation
    d. Source monitoring errors
6. There was no relationship between dissociation and inaccuracy of documented memory with delayed recall (“recovered memory”); i.e., *both* continuous and delayed recall memories may be accurate, partially accurate, or confabulated.
7. Dissociation negatively related to general measures of narrative cohesion, indicating that dissociation led to difficulties in recall, not inflation of false trauma accounts.
8. Neurobiologic research supported the trauma theory of dissociation as a regulatory response to fear or other extreme emotion with measurable biologic correlates.

Data from Dalenberg CJ, Brand BL, Gleaves DH, et al. Evaluation of the evidence for the trauma and fantasy models of dissociation. *Psychol Bull*. 2012;138(3):550–

588. doi:10.1037/a0027447.

In the review by Dalenberg and colleagues—across various suggestibility paradigms, including those hypothesized to etiologically generate false trauma memories—*dissociation was not reliably related to suggestibility.* Indeed, the average *weighted effect sizes between all forms of suggestibility and dissociation accounted for only 1% to 3% of the variance.* Consistent with previous research, the review suggested that a subgroup of dissociative individuals may be more prone to generation of confabulated memory, although studies have not distinguished between false memory, false belief, and acquiescence to an authority figure in giving responses. In addition, studies have not tested *false nonbelief,* that is, subjects who come to deny actual memories (i.e., parallel to false trauma recantation) when suggested to do so by an authority figure. In response to the Dalenberg review, advocates of the IM/SCM/FM models, in a major concession, acknowledged for the first time that some delayed recall of trauma memories may be veridical.

Based on the evidence for the eight predictions of the TM of Dissociation, Dalenberg and colleagues concluded that there is *strong empirical support for the hypothesis that trauma causes dissociation* and that dissociation remains related to trauma history after adjusting for fantasy proneness. They found little empirical support that the dissociation–trauma relationship is due to confabulated memories of trauma, various types of suggestibility, and/or fantasy proneness.

A recent meta-analysis of 104 international studies of 31,905 college students by Mary-Anne Kate and colleagues tested several assumptions of the TM versus the FM. Findings included that the prevalence of dissociative disorders in this sample (11%) fit the predictions of the TM model, not the FM (4%). They concluded that rates of dissociative disorders and elevated dissociation better fit the international prevalence of traumatic experiences (12%), not media exposure rates. Prevalence rates of dissociation/dissociative disorders have remained stable over time, not fluctuant as would be consistent with the ebb and flow of a fad. Prevalence rates of dissociative disorders were similar in college students to other populations, a finding contrary to FM postulates. Studies of dissociation by FM proponents found *higher-dissociation scores*, but did not use validated screening or diagnostic inventories, and may have primed subjects by prior administration of scales of fantasy proneness. Overall, this large meta-analysis supported the TM over the FM.

DISSOCIATION, DISSOCIATIVE IDENTITY DISORDER, AND INTERROGATORY SUGGESTIBILITY. The Gudjonsson Suggestibility Scale (GSS) is a robust measure of susceptibility to external pressure to produce confabulated narrative accounts, developed to assess *Interrogatory Suggestibility* (IS), the liability to false confessions. The GSS includes subscales assessing the respondent's liability to yield to leading questions and to shift (confabulate) responses under pressure from the examiner. Ninety-two adults recruited from the community were studied with measures of fantasy proneness, dissociation, trauma history, and the GSS. Participants' dissociation scores correlated significantly with trauma history and fantasy proneness, but not with GSS scores. Moreover, dissociation was negatively correlated in a sub-study with recall of the events of 9/11. In another study, hospitalized women with recovered childhood sexual abuse memories were significantly less suggestible on the GSS compared with general psychiatric controls from the same hospital.

These data are consistent with preliminary findings from studies of dissociative identity disorder that showed lower GSS scores compared with subjects with PTSD, trauma controls, and individuals with borderline personality disorder. Dissociative identity disorder subjects were *more* resistant than the other groups to interpersonal persuasion. A different study of SCID-D–diagnosed dissociative identity disorder patients failed to confirm the basic clinical profile posited by the SCM and replicated the finding that dissociative identity disorder patients did not show elevated interrogatory suggestibility. In particular, the dissociative identity disorder patients were highly introversive on the Eysenck personality inventory and showed low histrionic personality features on the Million Multiaxial Clinical Inventory. Most dissociative identity disorder respondents were low or moderately high on fantasy proneness, with a subgroup showing high scores. This may indicate that there is a subgroup of dissociative identity disorder individuals with high fantasy-prone scores who may have more difficulty with reality testing and possibly greater liability to forms of suggestibility and production of factitious symptoms, self-states, and personal history.

THE DISSOCIATION/PATHOLOGIC SLEEP MODEL. After the refutation of the FM, the critics of trauma-dissociation theory have moved on to conceptualize dissociation as caused by sleep problems, specifically a subclinical form of narcolepsy characterized by sleep paralysis and hypnogogic and hypnopompic hallucinations. This view denies that sleep disruption is symptomatic of posttraumatic dissociation itself, together with common comorbidities (e.g., PTSD, major depressive disorder). Rather, sleep problems are thought to be *causative* of dissociation, which, possibly accompanied by childhood trauma, leads to cognitive errors, problems with source monitoring, and promulgation of false memories of trauma. These studies downplay the idea that sleep disorders are the result of dissociation, trauma, and comorbid disorders (e.g., PTSD) and may secondarily contribute to the level or intensity of symptoms and disability. One study compared 12 female dissociative identity disorder patients, with 27 PTSD patients, and 55 normal controls on retrospective sleep assessments, the DES, measures of fantasy-proneness, and “cognitive efficiency.” Diagnoses of the dissociative identity disorder patients were ascertained with the SCID-D only, and the PTSD patients with the Clinician Administered PTSD Scale (CAPS) only. The authors concluded that intrusions of sleep phenomena into waking consciousness led to dissociative experiences. However, as 80% to 100% of dissociative identity disorder patients meet diagnostic criteria for PTSD, failure to assess dissociative identity disorder patients for PTSD makes the study uninterpretable. The authors did not explore the complex sleep pathology of dissociative identity disorder patients that includes severe PTSD intrusions, nocturnal activation of self-states, and awakening in flashbacks with fugue behavior. Also, due to dissociation, PTSD, and autohypnosis, many waking dissociative symptoms naturalistically have an oneiroid quality.

The transdiagnostic prevalence of dissociation suggests that sleep problems are related to underlying psychopathology, possibly exacerbated by dissociation, not causative of dissociation. Improvement in sleep may have general ameliorative effect on all types of psychopathologies, including dissociation. Amelioration of sleep and dissociative symptoms does not indicate a causal role for sleep problems in dissociation. Finally, this theory focuses on symptoms that no longer are part of the diagnostic criteria for narcolepsy. General population studies have found that hypnogogic/hypnopompic hallucinations and sleep paralysis are far more common in the general population than the prevalence of narcolepsy. Thus, there is minimal empirical support for the role of disordered sleep as etiologic in the development of dissociation and dissociative disorders.

CONCLUSION. In a recent review, Constance Dalenberg and colleagues described the use of eight nonscientific rhetorical arguments found in most critiques of the TM of dissociation. These include (a) extraordinary claims = anything that you believe that I do not; (b) *ipse dixit* or the bare assertion fallacy; (c) false consensus; (d) the “ever-shifting” goalpost (e.g., non-trauma dissociation theories); (e) ad hominem arguments and accusations of these against those with whom we disagree; (f) motivated skepticism; (g) demand for a super-study; and (h) arguing from authority (*ad verecundiam*). In addition, as in the sleep studies, there is a basic failure to understand the phenomenology of dissociation and dissociative identity disorder and how these affect all domains of functioning. Motivated skepticism is shown by, even in studies involving dissociative identity disorder individuals, failure to assess them for PTSD, to assess for trauma history, and/or attempt to specifically replicate methodology from trauma-model-based studies of dissociation/dissociative disorders.

DISSOCIATIVE IDENTITY DISORDER

Dissociative identity disorder has been researched most extensively of all the dissociative disorders. It is the paradigmatic dissociative psychopathology in that the symptoms of all the other dissociative disorders are commonly found in patients with dissociative identity disorder: Amnesias, fugues, depersonalization, derealization, and similar symptoms. The DSM-5-TR diagnostic criteria for dissociative identity disorder are discussed in the section on Nosology and are listed in Table 20–3.

Etiology of Dissociative Identity Disorder—Trauma

In all studies that have systematically examined this question, non–possession-form dissociative identity disorder is strongly linked to extreme, chronic, and early childhood maltreatment. Across a wide variety of studies, in both men and women, the rates of reported severe childhood trauma for children and adults with non-possession dissociative identity disorder range from 85% to 97% of cases. Sexual and physical maltreatment, usually in combination, are the most frequently reported sources of childhood trauma, although other kinds of trauma have been reported, such as multiple painful medical and surgical procedures during childhood, and wartime trauma. Emotional abuse, neglect, and experiencing and/or witnessing family violence have also been cited as traumatic antecedents of dissociative identity disorder, usually in combination with physical and sexual violence. Studies in the National Comorbidity Study Replication (NCS-R) have found that sexual, physical, and emotional abuse, neglect, intimate partner violence, and severe family dysfunction covary together so extensively in abusive families that no one of these can easily be partialled out from any other in attempting to understand the impact of early life maltreatment. Across studies (e.g., the NCS-R and a large population sample of children and adolescents from the National Childhood Traumatic Stress Network [NCTSN]), childhood sexual abuse is the most potent “synergistic” developmental adversity, combining with more ACEs than any other ACE.

Data on the role of trauma in possession-form dissociative identity disorder, although more limited, confirm the association between this presentation and traumatic exposure. Two epidemiologic community-based studies in Africa report the widespread attribution of pathologic dissociative possession to the restless spirits of individuals who perished in local civil wars, linking “spirit” possession and violent death. One study in Northern Uganda found that individuals with high likelihood of pathologic dissociation were significantly more likely to have suffered various forms of war-related trauma; 45% of individuals with this condition also met DSM-IV criteria for PTSD. Marjolein Van Duijl and colleagues conducted a systematic study of traumatic exposure among 119 possessed patients who visited traditional healers in Uganda. These individuals scored higher on the Harvard Trauma Questionnaire and the Traumatic Experiences Checklist than 71 randomly selected healthy inhabitants of the same villages, and scores on both trauma scales were strongly correlated with symptoms of dissociation ($r = 0.64$ to 0.69) and pathologic possession ($r = 0.51$ to 0.64).

Nontraumatic stressors, such as romantic disappointments and work or religious conflicts, have triggered other cases of pathologic dissociative possession. However, systematic assessment of childhood trauma is usually not conducted in these studies, precluding the exclusion of traumatic antecedents in the genesis of this condition. Other studies find a partial relationship with trauma; for example, one community-based epidemiologic study in Mozambique ($N = 941$) found only a trend-level association between pathologic dissociative possession states and the Harvard Trauma Questionnaire, although sufferers of this condition were more likely to have experienced the death of a child. Pathologic possession states may encompass presentations that do not meet criteria for dissociative identity disorder; those that do, however, appear more likely to be associated with traumatic etiologies.

The posttraumatic etiology of dissociative identity disorder is supported by data that many disorders, symptoms, life experiences, and maladaptive behaviors typical of traumatized individuals are commonly found in dissociative identity disorder. For example, the National Comorbidity Study-1 (NCS-1) found that the most common outcomes of trauma included PTSD, depression, and substance/alcohol abuse. Studies of dissociative identity disorder have found comorbid PTSD in 79% to 100% of cases, current or past depression in 83% to 96%, and a past or current history of substance abuse in 83% to 96% of cases. High rates of childhood abuse and maltreatment characterize subsamples of individuals with somatic symptom and related disorders, especially functional neurologic symptom (conversion) disorder and somatic symptom disorder. Across studies, 40% to 92% of individuals with dissociative identity disorder report somatoform symptoms and 35% to 61% have a comorbid somatic symptom disorder. Data from Turkey suggest that pseudoneurologic conversion symptoms are the most common presenting symptom of persons with dissociative identity disorder; some of these Turkish patients attributed their self-states to possession, often combined with non–possession-form self-states. In one study by Vedat Sar and colleagues, nearly half of 38 consecutive psychiatric outpatients diagnosed with conversion disorder met DSM-IV criteria for a dissociative disorder, including 17% with dissociative identity disorder and nearly three-quarters with dissociative disorder NOS. A subgroup of North American/European dissociative identity disorder patients come to clinical attention for a primary functional neurologic symptom disorder, especially nonepileptic seizures and sensory–motor conversion symptoms.

Individuals with dissociative identity disorder frequently meet diagnostic criteria for somatic symptom disorder, often with predominant pain—particularly headaches, abdominal pain, and pelvic pain, but also migratory pain syndromes. Childhood sexual abuse survivors with psychophysiological disorders are more likely than controls to have a lower threshold for experiencing physiologic phenomena as noxious or painful. Conversely, repeated childhood abuse can lead to *anesthesia* for physical pain, such as during dental procedures. This can create the paradoxical situation that dissociative identity disorder patients may be highly sensitized to chronic pain patterns associated with childhood sexual abuse (e.g., chronic pelvic pain, headaches, irritable bowel syndrome) and may have numerous, costly, unproductive workups for these symptoms, including exploratory surgery.

Dangerousness to Self and Others and Dissociative Identity Disorder

Suicidal, self-destructive, violent, and high-risk behaviors are also strongly associated with trauma and, particularly, with early life trauma and repeated traumatic experiences. Across studies, 66% to 80% of individuals with dissociative identity disorder reported a history of suicide attempts, and 92% to 100% reported current or past suicidal ideation. Individuals with dissociative identity disorder have higher rates of multiple suicide attempts than individuals with borderline personality disorder, PTSD, and substance abuse disorders who do not have a comorbid dissociative disorder. In a 30-month prospective, longitudinal treatment outcome study described in detail below, 2 patients—1.8% of the final sample of 119—died by suicide. This is likely an unrepresentatively low rate, based on participation in a treatment outcome study that demonstrated an overall decrease in suicidality among its participants. Case series also have described death by suicide in dissociative identity disorder patients, including in possession-form cases.

Studies with pre–DSM-5-TR criteria show that 78% of individuals with dissociative identity disorder have engaged in self-harm behaviors. These include severe forms of nonsuicidal self-injury such as major lacerations requiring sutures or surgical repair to limbs, genitals, and abdomen; severe burns including use of acids and caustic substances; ingestion of caustic compounds such as drain or oven cleaner; deliberate fracture of limbs; genital mutilation; and severe head-banging resulting in bruising, bleeding, and/or skull fractures, among many others.

Clinical case series document that a subgroup of individuals with dissociative identity disorder engage in violent behavior; some have been incarcerated for violent crimes including murder, severe assaults, and rape. A case series of male patients with dissociative identity disorder revealed that 20% admitted to perpetrating homicide and 13% to rape. In this study, 10% of females with the disorder also reported perpetrating homicide. In a clinical study of mothers with dissociative identity disorder, 16% engaged in significant maltreatment of their own children and 45% were compromised or impaired as mothers. Abusive and/or neglectful mothers contribute the multigenerational transmission of trauma.

Table 20–12 enumerates many of the safety issues that occur in patients with dissociative identity disorder.

Like other victims of early life trauma, individuals with dissociative identity disorder have high rates of adult traumatization, in particular rape. In one study, 60% of women and 17% of men with dissociative identity disorder reported being raped as adults. Nearly 60% of women with dissociative identity disorder reported suffering intimate partner violence or other forms of violence in adulthood; 36% reported experiencing violence in the past year. These rates of rape and intimate partner violence are notably higher than those for the general U.S. population.

As described in the Epidemiology section, a subgroup of individuals with dissociative identity disorder report being trafficked, often beginning in childhood, including being sold as children to be raped by male or female adults, most commonly family members, and sometimes recorded on video or transmitted via live-streaming. Many of these individuals also report that they continue to be trafficked into adulthood, sometimes by the same perpetrators, often members of their families of origin. Recent clinical case series document that a subgroup of individuals with dissociative identity disorder continues, even into middle age, to be incestuously raped and/or trafficked by family members, typically male family members, but generally known about, condoned, and even participated in by female family members. This is consistent with recent international reviews of public criminal cases describing adults who, often for years, have been entrapped into emotional, physical, and sexual violence by family members. This phenomenon urgently requires further study and antihuman trafficking legislation should take such dynamics into consideration.

Table 20–12.
Types of Dangerousness to Self and Others Typically Found in Dissociative Identity Disorder Patients

1. Dangerousness to self
    - a. Self-destructive behavior
    - b. Nonsuicidal self-injury
    - c. Suicide attempts
    - d. Death by suicide
2. Dangerousness to others
    - a. Abuse of minor children of the patient and/or neglect leading to abuse of children by others, intra-famillialy and/or extra-famillialy
    - b. Intimate partner violence
    - c. Assault, sexual assault, arson, murder, serial murder, etc.
3. High-risk behaviors, e.g., repeated reckless driving, risky sexual behaviors
4. Substance-related and addictive disorders
5. Eating disorders
6. Enmeshment in/victim of abusive/traumatizing relationships
    - a. Enmeshment with abusive family of origin
    - b. Intimate partner violence
    - c. Being trafficked into prostitution
7. Lack of food, clothing, and/or shelter
8. Lack of access to and/or avoidance of adequate medical care

Despite abundant evidence of the relationship between trauma and dissociative identity disorder, critics have raised questions about the validity of dissociative identity disorder patients' self-reports of childhood trauma. Recent studies, including large samples of maltreated children with dissociative disorders and intensively validated case studies, nevertheless have provided rigorous independent corroboration of the patients’ reports of maltreatment. These studies continue to strongly support a developmental linkage between childhood trauma and dissociative identity disorder. In general, studies involving adult self-reports of childhood trauma have not shown these reports to be any less valid than self-reports of other clinical phenomena, such as self-reports of a history of smoking.

Discrete Behavioral States and Dissociative Identity Disorder

The DBSs model, discussed in the prior section on Etiology, is a particularly helpful overarching construct for conceptualizing dissociative identity disorder. It is grounded in data on child development and on the alterations that maltreatment and/or severe early stress causes on the trajectory of development in the sense of self and identity, autobiographical memory, capacity for attachment, emotion regulation, cognitive schema, ability to maintain and/or modulate behavioral states, and beliefs about the perpetrators, the self, and the world.

According to the DBS model, *dissociative identity disorder is conceptualized as a childhood-onset posttraumatic developmental disorder* in which the traumatized child is unable to complete the normal developmental processes involved in consolidating a core and singular sense of self. Instead, repeated early trauma disrupts the usual unification of sense of self, resulting in the development of multiple independent or DBSs. In normal development, adults implicitly recognize state changes in children and attempt to assist the child in reducing distressing states (e.g., feeling hungry, wet, teething, tantrums, ill) and/or in returning to better states (e.g., feeling calm, interested, sleepy). Adult attunement to the child helps develop the child’s capacities to recognize and self-modulate their own states, and, ultimately, to develop a singular and internally consistent sense of self across states, relationships, and contexts, including distressing ones.

In comparison, traumatic experiences—particularly repeated, unpredictable maltreatment perpetrated by caregivers—not only produce extreme, overwhelming states, but also leave the child to attempt to recover from these states, typically by a global dissociative shut-down. Abusive parents usually provide little or highly inconsistent soothing experiences, especially after having just hurt the child. They may compound the difficulty by blaming, mocking, and humiliating the child for being distressed, and/or literally telling the child that the distress, as well as many other characteristics of the child, causes the parent’s abusiveness.

Due to disturbed caretaker–child attachment and parenting, repeated early trauma disrupts the development of normal metacognitive processes involved in the consolidation of a unified sense of self across different relational (e.g., with parents, peers, others); social (e.g., at school, at home, with peers); and emotional states (e.g., terrified, humiliated, enraged, mixed). The models of Type D attachment and BTT fit observations of children and adults with dissociative identity disorder. Typically, among the psychodynamic underpinning of self-states, there is at least one group that denies trauma and reports strong attachment to at least some childhood caregivers, citing their positive qualities. These states are usually in profound conflict with other self-states reporting revulsion and hatred of the same caregivers, and who describe profound maltreatment-based betrayals.

The child who develops dissociative identity disorder often experiences marked discontinuities between the reality of a chronically abusive, violent home life and the presentation of the family as “perfect” in the outside world (e.g., at church, school, public events). The dissociative child may also experience other areas of life as additional disconnected realities: school, athletics, artistic endeavors, and so on. Thus, the developing self-states may be experienced as encompassing alternate, often contradictory realities, leading some to refer to the disorder as a *multiple reality disorder.*

Over time, the child’s developing internal sense of self may come to reflect the disorganization and conflict evident across the various relationships and environments with which they are occupied. Different self-states can become subjectively personified and develop along independent developmental trajectories. In adolescence, “separation and individuation,” including issues of sexuality and autonomy from family, are major developmental tasks. This is difficult to negotiate for dissociative adolescents, who may experience elaboration and secondary subjective structuring of self-states. This process can involve creation of states with different roles (e.g., “everyday life,” “school,” “home”); intrapsychic and emotional functions (e.g., “protector,” “angry one”); relationships to trauma (e.g., one who “knows,” one who has no memory), and creative secondary elaborations (e.g., names, accents, wardrobe, hair style, behaviors).

Making Sense of Self-States

In dissociative identity disorder *self-states are defined as discrete behavioral states, each organized around a prevailing affect, a sense of self (often including a distinct body image), a set of—usually limited—state-dependent autobiographical memories, and a discrete behavioral repertoire, with a capacity to affect the person’s behavior.* In other words, *dissociative identity disorder self-states are multiple, relatively separate centers of information processing in the mind, capable of relatively independent thought, emotion, sense of self, and memory, as well as the capacity to—directly or indirectly—control the person’s behavior.* Despite its current name referencing “identities,” developmental deficiencies in integration of the *self* and *self-systems* are the basic psychobiological process in dissociative identity disorder.

Usually, over developmental time, dissociative identity disorder self-states organize into multiple interacting *self-state systems.* The subjective experiences and behavior of the dissociative identity disorder individual can be best understood in terms of the model of *complex dynamic systems.* On a moment-to-moment basis, these “outputs” depend on the cooperation, collaboration, and communication among some self-states/systems that may be variably in conflict with other self-states/systems, typically resulting in symptoms of overlap, interference, influence, and intrusion. It is beyond the scope of this chapter to provide in-depth discussion of the concepts of “self” (and self-systems), “identity,” and “personality.” In the literature, these are defined differently by different schools of psychology, philosophy, and anthropology. Sometimes the definitions are tautological, for example, “self” is defined as “identity.” These terms have “fuzzy boundaries” and are hard to define without reference to each other. These constructs also may be best conceptualized as interactions among complex, dynamic psychological systems. For simplicity, subjectively, self is “who” we are—and there may be many “whos” in different situations and contexts over the lifespan. Identity is “what” we are: a whole host of (usually) longer-term dispositions having to do with ethnicity, gender, age, sexual orientation, political affiliation, profession, etc. These can be *negatively* imposed from outside by bullying, racism, genocide, and, among others, the child abuse survivor being told abuse is their fault. “Personality” is defined differently by various schools of thought, including theoretical nonclinical domains, but is defined by long-term, relatively stable, psychobiological dispositions and patterns, within which self and identity states are shaped. Manifest personality symptoms can be state-dependent (e.g., a depressed individual appears to meet diagnostic criteria for a personality disorder). Often, both mood and apparent personality disorder symptoms remit with successful antidepressant treatment.

The DBS developmental model of dissociative identity disorder postulates that, contrary to typical conceptualizations, there is no “real” singular self, identity or “personality” from whom “alters split off,” even when a patient themselves endorses this view. The seemingly “real” self-state(s) are commonly conflated with those claiming the legal name of the human being. Rather, here, their mind is made up of a combination of all self-states, dissociative defenses, and nondissociative psychobiological processes. Developmentally, the individual with dissociative identity disorder is not a “shattered mirror” with a finite number of “fragments” or “parts.” The individual’s mind encompasses complex, dynamic, potentially imaginatively infinite systems of DBSs, more like a network of three-dimensional jigsaw puzzles. Different self-states often encompass painful psychological issues for the individual, frequently taking the form of polarized pairs representing antithetical positions, although self-states embodying more neutral and conflict-free processes also commonly occur. In some individuals with dissociative identity disorder virtually every aspect of mental life is personified.

Individuals with the disorder show considerable variability in terms of the complexity and therapeutic tractability of their self-state systems. Most psychotherapeutic work is directed toward these self-state systems and toward the “whole human being.” Much general information and many functions and abilities are shared across self-states, indicating the fundamental unity of the mental processes of the dissociative individual. This provides the foundation for therapeutic efforts directed at the development of a more consciously integrated sense of self.

Self-States in Possession-Form Dissociative Identity Disorder

Dissociative possession experience involves altered states; thus, DBS theory conceptualizes possession-form dissociative identity disorder in similar terms as the non–possession-form disorder. In societies where spiritual agents (e.g., deities, ghosts, animal spirits, demons, *jinn*) are considered ontologically “real,” these agents represent alternative sources of identity-state construction. Individuals who are maltreated may engage these supra-individual self-states as figures that represent aversive or supportive elements in their lives. Even in societies where supernatural forces are not generally endorsed, intracultural variation may encourage the assumption of spiritual self-states within certain families and social groupings (e.g., highly religious groups). This process would give rise to possessing agents that have a collective existence. For example, in India the same spirit may possess different family members down the generations (or different villagers in the same generation). In societies where the cultural conception of the self is more porous to other self-states depending on context (“dividual”), including spiritual self-states, the development of DBS may proceed along less individualistic paths, resulting in differently organized self/identity states. Similarly, clinical data on Western families with multigenerational dissociative identity disorder have found similar self-states across three generations of the same family.

The Four-Factor Theory of Dissociative Identity Disorder. Richard Kluft has proposed a Four-Factor Theory (FFT) of dissociative identity disorder that is a specific case of the DBS model. The first factor is the *capacity for dissociation,* likely due to a genetic diathesis (discussed in the section on Neurobiology). Per DBS theory, the next factor involves *experiences that overwhelm other nondissociative adaptive capacities of the child.* The child cannot fight or flee from perpetrators but freezes and flees inwardly into altered states of consciousness that provide encapsulation and relief from traumatic experiences. The third factor is the *lack of soothing and restorative experiences,* where the traumatized child must find solitary ways to calm and soothe themselves to attempt mitigation of traumatic experiences and attendant emotions.

The fourth and final factor of the FFT involves the many posttraumatic, psychodynamic, familial, social, and cultural factors leading to the *secondary structuring, subjective personification, and elaboration of the self-states into "identity states"* (also see section on Culture, Dissociative Disorders, and Sociocognitive Phenomena). These can include posttraumatic, attachment-based psychodynamic factors resulting in the first, very early psychological self-construction, the creation of an isomorphic state: one child who is betrayed through perpetration by attachment figures and one who is not. Further elaboration may include, for example, a female child thinking that, if she were a boy, sexual abuse would not be happening, leading to development of a subjectively male state. Other personified states may attempt to mitigate and sequester the impact of multiple ongoing traumas, leading to development of multiple traumatized self/identity-states that encapsulate traumatic memories or parts of one or more memories. Traumatic and/or overwhelming experiences in adolescence or adulthood may also lead to the development of additional traumatized self/identity-states. Other states may be experienced as without emotions or pain perception to mitigate the impact of traumatic memories and experiences. These states may be experienced as biographical variants of the individual (e.g., a numb child identity)—as is typical in non–possession-form dissociative identity disorder. Alternatively, in possession-form presentations they may be experienced as supernatural, collectively validated states that transcend death (e.g., the ghost of the suffering child, out for revenge) or as mythical, insensible, invulnerable presences seeking redress (e.g., an avenging spirit).

Attempts to feel strong and safe may lead to subjective creation of a tough, powerful adult male state or a powerful spiritual ally, such as a deity. The desire to control and anticipate at least some aspect of continual, inevitable, unpredictable trauma may lead to the development of a state that is an internalized representation of the abusive, shaming perpetrator(s), often termed an *introject self-state;* alternatively, this could take the form of a demon or punishing animal spirit. By contrast, people who have been kind to the individual may also be represented as introject states. For example, if a supportive, kind relative dies, the person with dissociative identity disorder may have difficulty grieving the loss of this person. An introject state develops instead as an internal representation of the deceased individual, forestalling the need for grieving. In addition, later traumatic or overwhelming events may lead to development of additional self-states.

Mr. T, a divorced musician in his 30s, was admitted to a psychiatry unit. He was suicidal after repeatedly finding himself dressed in women's clothes. He denied recall of buying or putting on these clothes, or for the periods of time that he cross-dressed. He had become panicked, mortified, and intensely suicidal when family members discovered him dressed in female clothes. He had always viewed himself as "a tough, macho kind of guy." Mr. T had been in recovery from drugs and alcohol for a decade and vigorously denied any recent substance use. His family and AA sponsor corroborated Mr. T's having remained sober and screening for drugs and alcohol was negative. He reported gaps in his recall for many periods of his life. He ascribed the memory problems to substance use, although memory gaps were present for long periods of sobriety as well, which he attributed to "dry blackouts." He described sudden disconcerting lapses in his musical abilities, which had created job problems for him. At other times, he described finding that he could play other instruments proficiently without knowing how he had learned them. He reported that his head was "filled with opinions that don't belong to me." On more careful history taking, he described a chaotic early life, being sent to live with violent, abusive relatives. The only consistently positive figure in his development was his maternal grandmother, who had supported and protected him as much as she could. She had died suddenly—just before the episodes of cross-dressing began. As he discussed this, Mr. T reported hearing his grandmother's voice and broke down in tears stating that he could not bear the idea that she was dead. On sequential evaluation, several self-states were identified, including one based on his grandmother, reportedly created as a comforting protector in childhood, experienced as reactivated after her death. This self-state acknowledged buying and dressing in the female clothes, "So I could always be with him." Work on grieving for his grandmother resolved the acute episodes of amnesia and pressure to cross-dress. Mr. T was referred to long-term outpatient psychotherapy for dissociative identity disorder.

Shame and Childhood Trauma. Feelings of shame are likely a concomitant of most, if not all, traumatic experiences. The intent of all perpetrations is power, control, and dominance over the victim. Humiliation is the delivery system through which these are created and enforced. Multiple early life traumas, often involving deliberate, sadistic humiliation and mocking of the victim, lead to an extreme burden of shame. Systematic focus on shame is critical in the treatment of all interpersonally traumatized people, as well as on other powerful emotions like anger, fear, anguish, disgust, and their combinations. The development of dissociative identity disorder in early childhood leaves emotions experienced either as entirely suppressed or overwhelmingly activated, often accompanied by alexithymia.

Richard Kluft, based on the work on affect theory by Donald Nathanson, pioneered the conceptualization of self-states as embodying specific shame "scripts" such as *self-attack* (i.e., repeated self-shaming); *avoidance of shame affect* and anything that might engender it (e.g., by drugs, alcohol, nonsuicidal self-injury); *withdrawal* from situations that engender shame (e.g., a student drops out of college because of shame about performance); and/or *attack the other* by humiliation. Also, for example the latter is almost universally experienced by victims of perpetration, wanting to "get back at" and make the bully experience humiliating retaliation.

Negatively valenced introject self-states may take the form of the humiliating perpetrator(s). These represent "paradoxical protection" to anticipate and gain some sense of control over unpredictable, *but inevitable*, mortifying childhood violence by attachment figures. The adult individual with dissociative identity disorder experiences profound relational mistrust, anticipating that all attachments will result in betrayal: abuse, abandonment, or both. These commonly misunderstood self-states represent a kind of early warning system —usually unheeded—about the dangers of attachment. Therapists and other self-states often conceptually reify these self-states as “malevolent” or “bad,” as they may drive self-destructive behavior and/or “sabotage” gains in therapy. In general, it is far more effective to interpret the protective functions of these self-states, while insisting on maintaining safety to find recovery-oriented methods for protection from potential betrayal.

Other Substrates Shaping Self-states. In possession-form presentations, self-states are usually perceived to be “real” and external by members of the social group who share the sufferer’s perspective regarding the ontologic validity of spiritual beings. In non–possession-form dissociative identity disorder, however, the experience of self-states as literally separate, “outside” people is termed *delusional separateness.* This usually occurs when, early in development, repeated traumatic events have been experienced as so overwhelming, intolerable, damaging, shameful, disgusting, dissonant, and ego-dystonic that the person may attempt to subjectively extrude them altogether, as “not-me” events. In assessment of possible psychosis, it is important for clinicians to consider the prevailing range of cultural views of the person and their community as well as processes such as hallucinatory phenomena related to self-states and delusional separateness that are best explained by a dissociative and posttraumatic etiology.

Other factors that lead to secondary structuring of dissociative identity disorder identity states can include contradictory parental demands (i.e., father wants a son and dresses the child as a boy, even using a different name; mother wants a girl and does the converse). Sociocultural factors that lead to further elaboration of self/identity-states include all forms of media, as well as religious beliefs (e.g., in demons). The intelligence and creativity of the person can influence the complexity of the subjective elaboration of self-states. Some individuals with dissociative identity disorder may retreat into highly complex, elaborated inner worlds. This is most likely to occur when the experience of outer reality is perceived as too painful and especially when the person has not only undergone severe life traumas but also has experienced a limited positive adaptation to the world in their adult life. This constellation may be associated with poorer treatment outcomes. Similarly, studies have shown that psychotic patients with grandiose delusions are less adherent to medications than those with non-grandiose delusions. In both these clinical situations, patients attempt to avoid experiencing their painfully limited lives by remaining lost in elaborate dissociative fantasy worlds or grandiose delusional beliefs, respectively.

Inappropriate psychotherapy can lead to the development of additional self-states, especially if the treatment is experienced as unsafe and/or overwhelming to the person. Therapists may unintentionally reinforce separateness by playing favorites among self-states; engaging in very different behavior with different states (e.g., playing with child self-states); viewing self-states as "good" or "bad" rather than psychological adaptations that need to be understood; refusing to speak to some self-states; and not encouraging a decrease in dissociative barriers by failing to promote internal communication, collaboration, and coordination among self-states—for example, by making ongoing, separate appointments with self-states as if with separate people. Sexual or other forms of exploitation by therapists are severe traumatic betrayals and, as all adult traumas, can also lead to the development of additional self-states.

Unfortunately, the *least important phenomena associated with Factor 4* are those that lead to the most fascination with the patient or the diagnosis and emphasized in media portrayals of dissociative identity disorder. These include the person's highly elaborated, differing presentation of identity-states, or fantasies about the states ("12th century monk"; "Samurai Warrior"); in the media, even these factors are depicted in an exaggerated way compared to their clinical presentations. In patients, there may be elaboration of the presentational and subjective characteristics of the self-states with shifts in wardrobe, grooming, accents, gait, and so on. It is a poor prognostic sign when the patient has significant narcissistic investment in having dissociative identity disorder and/or in the extensive elaboration of marked presentational differences between self-states.

Personality Organization, Personality Disorders, and Dissociative Identity Disorder

Borderline personality disorder has been reported in 31% to 83% of patients with non–possession-form dissociative identity disorder. Some early explanatory theories characterized dissociative identity disorder as a variant of borderline personality disorder and/or within the psychoanalytic construct of borderline personality organization. However, several studies have shown substantial differences between persons with dissociative identity disorder and individuals with Cluster B personality disorders. Using the Structured Clinical Interview for the DSM-IV Dissociative Disorders (SCID-D; see Table 20–1), patients with dissociative identity disorder and those with Cluster B personality disorder (primarily borderline personality disorder) did not differ significantly on measures of depersonalization and derealization. However, patients with dissociative identity disorder reported significantly more amnesia, identity alteration, and identity confusion. Also, the dissociative patients had significantly earlier and more extensive histories of trauma.

Psychological assessment of large numbers of individuals with dissociative identity disorder using the Rorschach test, and other assessment instruments such as the Millon Clinical Multiaxial Inventory-III (MCMI-III), did not show a personality profile typical of borderline or histrionic personality types or of patients with psychosis studied with the same measures. On the Rorschach, individuals with dissociative identity disorder showed a unique personality structure characterized by marked traumatic reactivity, obsessional features, introversion, and high complexity of response—far more complex than individuals with personality disorders or psychosis, Also, contrasted with individuals with personality disorder or psychosis, those with dissociative identity disorder showed notable ability for self-reflection, as exhibited by good insight (Rorschach variable FD), reality testing, and capacity for logical as opposed to distorted thinking. These patients tended to try to distance from posttraumatic stimuli through auto-hypnotic and dissociative defenses.

The posttraumatic reactivity during testing led to the development of a specific Rorschach variable, the Trauma Content Score (TC), the sum of morbid, anatomy, sex, and blood special scores divided by total responses. Depending on the study, DID individuals show 50% to 150% TC responses on the Rorschach (i.e., sometimes more than one TC response per card), which are significantly higher than all other patient or nonclinical groups. During testing, when starting to experience PTSD symptoms, many individuals with dissociative identity disorder attempted to avoid these symptoms through hyperfocus on small features of the blot, autohypnotic trance symptoms, and depersonalization/derealization. This may parallel the findings among individuals with dissociative PTSD in a study of unconscious and conscious responses to fearful faces. The dissociative PTSD group showed rapid, marked unconscious arousal of fear circuitry to the threatening stimuli. With conscious processing, however, frontal brain systems that regulate emotional arousal in the amygdala were immediately activated, accompanied by increased dissociative symptoms. With failure of dissociative avoidance of intrusive PTSD symptoms, patients with dissociative identity disorder went into flashback. They increasingly lost capacities for reality testing and relationship to reality, switching to often regressed self-states, and even perceiving the Rorschach card itself as a malign object.

Consistent with the conceptualization of dissociative identity disorder as a posttraumatic developmental disorder, psychological profile studies of persons with dissociative identity disorder suggest *that early-life dissociation serves as a protective, resiliency-inducing function* for predisposed children growing up in violent, unpredictably malevolent environments. The dissociative “sequestration” of traumatic memories and emotions allows for more normal development in capacity for logical and reality-based thought, executive function, mentalization and insight, artistic and intellectual accomplishments, a sense of humor, hope, and willingness to engage in relationships with others. The latter may underlie a capacity for the therapeutic alliance, a transtheoretical predictor of response to psychotherapy.

Similarly, across studies on the MCMI-III, compared to other patient groups, individuals with dissociative identity disorder showed highest scores on depression, PTSD, and anxiety. They frequently endorsed negative personality traits—especially avoidant, self-defeating, passive-aggressive, and borderline traits—like patterns found in individuals with chronic PTSD. Consistent with the Rorschach findings of obsessional personality elements in dissociative identity disorder, on the MCMI-III they scored the *lowest* on histrionic traits, followed by low scores on narcissistic, antisocial, and sadistic traits.

DISSOCIATIVE DISORDERS, DISSOCIATION, AND BORDERLINE PERSONALITY DISORDER. Psychological profile differences between borderline and dissociative identity disorder patients have been delineated above. International studies of borderline personality disorder have identified dissociative and nondissociative subgroups. Consistent with the transdiagnostic studies of dissociation, the former had significantly poorer response to unmodified dialectical behavior therapy (DBT). International studies with the DES, A-DES, DES-T, SCID-D, MID, DDIS, and DSM-IV SCID-BPD, among other measures, assessed the relationship between dissociative identity disorder, other dissociative disorders, and borderline personality disorder. In meta-analyses of the DES, borderline personality disorder showed lower mean scores than the dissociative disorders and PTSD. Other studies using the DES described several borderline personality disorder subgroups, including about 25% without a dissociative disorder, 29% with mild dissociation, and 54% meeting DSM-IV criteria for either dissociative identity disorder or dissociative disorder NOS, both with more severe “borderline” symptoms. Studies of dissociation and borderline personality disorder using standardized diagnostic instruments, commonly find that substantial subgroups of the borderline group meet DSM diagnostic criteria for dissociative identity disorder.

Other studies suggest that individuals with dissociative identity disorder experience a self-image that is relatively stable within self-states, with instability manifesting due to shifting and overlap/interference among states. In studies using the MID, “core” dissociative symptoms in dissociative identity disorder primarily were predicted by actions/interactions of self-states; in individuals with borderline personality disorder, dissociation was predicted by rapid, stress-driven disruptive state shifts affecting perceptual, affective, and neurocognitive functioning, often accompanied by depersonalization/derealization. Unlike patients with borderline personality disorder, the avoidant personality traits of individuals with dissociative identity disorder indicate that many of the latter *prefer* to be alone—as is common in other posttraumatic stress populations—and do not make desperate efforts to avoid being alone.

Several international studies found that, when individuals with dissociative identity disorder meet DSM (IV-TR, 5) diagnostic criteria for borderline personality disorder, it was a severity indicator of *decompensation*, not core personality disorder pathology. A DDIS diagnosis of borderline personality disorder significantly correlated with more extensive trauma histories, higher scores on the DES, more passive-influence symptoms, increased complex amnesia, and similar severe dissociative symptoms. After stabilization through appropriate treatment interventions, these “borderline” individuals mostly no longer meet diagnostic criteria for borderline personality disorder.

ANTISOCIAL PERSONALITY FEATURES AND DISSOCIATIVE IDENTITY DISORDER. Forensic case series and intensive forensic case studies document individuals with dissociative identity disorder who have pervasive antisocial, psychopathic, and/or manipulative personality features. This includes a group of imprisoned murderers, none of whom had used their dissociative diagnosis as a legal defense. An intensive clinical case study of an arsonist with dissociative identity disorder suggested that she used the adaptive psychological capacities described in prior sections in the service of criminal behavior. This included malingering more severe dissociative symptoms to manipulate others and to attempt to avoid being adjudged by clinicians and the courts as responsible for her behavior. This subgroup of patients may be relatively infrequent in clinical settings but may be found in the social service or law enforcement systems. As noted earlier, a subgroup of individuals with the disorder is involved in criminal activities, including the severe neglect, abuse, and/or trafficking of their own children, perpetuating the multigenerational transmission of family violence and trauma. Paranoid personality traits, well beyond the typical posttraumatic suspiciousness of the complex trauma survivor, may be found among this subgroup. Some individuals with dissociative identity disorder may engage in sporadic antisocial behavior or manipulation of others. The differential diagnostic issue is whether there are global antisocial/psychopathic personality features, rather than problematic behavior related to specific circumstances, decompensation, and/or experienced as driven only by specific self-states.

Forensically and clinically, individuals with dissociative identity disorder should be held responsible for all behavior, whether ascribed to self-states, amnesia, or passive influence. Human beings engage in behavior, not self-states. The personality profile of individuals with dissociative identity disorder indicates that as "a whole human being," typically, they have the capacity to plan, to delay, and they know the difference between right and wrong behavior.

Table 20–13.
Amnesia and Memory Symptoms in Dissociative Identity Disorder

Blackouts or time loss
Disremembered behavior
Dissociative fugues
Unexplained possessions
Inexplicable changes in relationships
Perplexing fluctuations in skills, habits, and knowledge
Fragmentary recall of entire life history
Chronic mistaken identity experiences
Microdissociations (brief interruptions of awareness with amnesia)

Mental status examination questions for dissociative amnesia

If answers are positive, ask the patient to describe the event.

Make sure to specify that the symptom does not occur during an episode of intoxication.

1. Do you ever have blackouts? Blank spells? Memory lapses?
2. Do you "lose" time? Have gaps in your experience of time?
3. Have you ever traveled a considerable distance without recollection of how you did this or where you went exactly? Find yourself somewhere else without knowing how you got there?
4. Do people tell you of things you have said and done that you do not recall?
5. Do you find objects in your possession (such as clothes, personal items, groceries in your grocery cart, books, tools, equipment, jewelry, vehicles, weapons, etc.) that you do not remember acquiring? Out-of-character items? Items that a child might have? Toys? Stuffed animals?
6. Have you ever been told or found evidence that you have talents and abilities that you did not know that you had? For example, vocational, musical, artistic, mechanical, culinary, literary, athletic, etc.? Do your tastes seem to fluctuate inexplicably? For example, food preferences, personal habits, taste in music or clothes, use of tobacco products, alcohol, drugs, etc.?
7. Have you ever found or been told that at times you cannot access well-learned skills, For example, vocational, musical, artistic, culinary, mechanical, literary, athletic, or other talents or

abilities? This is not the same as “having a bad day” at work, doing athletics, etc.

8. Do you have gaps in your memory of your life? Are you missing parts of your memory for your life history? Are you missing memories of some important events in your life? For example, weddings, birthdays, graduations, pregnancies, birth of children, etc.?
9. Do you lose track of or tune out conversations or therapy sessions as they are occurring? Do you find that, while you are listening to someone talk, you did not hear all or part of what was just said?
10. What is the longest time period that you have lost? Minutes? Hours? Days? Weeks? Months? Years? Describe.

Adapted from Loewenstein RJ. An office mental status examination for chronic complex dissociative symptoms and multiple personality disorder. *Psychiatr Clin North Am.* 1991;14(3):567. Copyright © 1991, with permission from Elsevier. doi:10.1016/S0193-953X(18)30290-9.

Diagnosis and Clinical Features

Tables 20–13 through 20–16 describe the symptom clusters that are commonly found in individuals with dissociative identity disorder and some of the mental status questions that may elicit these symptoms.

Memory and Amnesia Symptoms. Since the descriptions of dissociative identity disorder in the early 19th century, memory deficits have been a hallmark of the disorder (see Table 20–13). Studies of large cohorts of patients with dissociative identity disorder assessed with the MID found that the experience of amnesia in these individuals manifests both as an ongoing symptom as well as clinically significant lacunae in retrospective autobiographical memory. In contrast, amnesia in individuals with dissociative amnesia disorder is characterized only by retrospective autobiographical memory deficits.

As part of the general mental status examination, clinicians should routinely inquire about current or past experiences of “losing time,” black-out spells, and major gaps in the continuity of recall for personal information. Patients rarely spontaneously report these, requiring active inquiry by the interviewer. Positive responses should be documented with specific examples provided by the patient. In some instances, patients report “coming to” amid an activity with little or no recall of how they began doing so. Frequently, patients find evidence of having done or acquired things for which they have no recall (see DID Case 1). Friends and family members may tell them about significant things that they have said or done that they cannot remember. Patients may find that they have unknowingly traveled some distance (a fugue episode) or that days or even weeks have passed for which they cannot account. Dissociative experiences of time loss are too extensive to be explained by normal failures of episodic/autobiographical memory. The former typically are characterized by either sharply demarcated onsets and offsets and/or a jumble of memory “fragments,” often with impaired sequencing and temporal awareness. Interviewers should establish that time-loss experiences occur in the absence of intoxication or substance abuse, although high rates of drug and alcohol abuse in dissociative patients may complicate this determination and some severely dissociative patients would prefer to ascribe their memory problems to substance use (see DID Case 1).

Commonly, patients with dissociative identity disorder report perplexing fluctuations in skills, habits, or well-learned abilities, such as how to do one’s job, fluency in a foreign language, and/or athletic or artistic abilities. Patients may report drawing a complete blank for skills or knowledge at times whereas, at other times, they easily and reliably access the information in question. Conversely, combining different types of dissociative amnesia, they may be told of abilities or capacities of which they are unaware (e.g., artistic, academic, athletic). Uncannily, they find that “somehow” an ability appears and then disappears; for example, a failing student suddenly finds themselves inexplicably knowing the correct answers on an important test. These patients may describe perplexing handwriting fluctuations, vision changes, and/or oscillation of interest in and tolerance for alcohol, drugs, and/or medications.

At the same time, the quotidian amnesia in dissociative identity disorder is relative. All the symptoms of ongoing amnesia are explicable by the actions, interactions, and/or conflicts among self-states. The “everyday” self-states may have no recall of a lengthy episode of dissociative fugue. Self-states that take responsibility for the fugue can recall it as detailed first-hand autobiographical memory. Variability in skills, habits, and knowledge can be deconstructed by asking to communicate with self-states experienced as responsible for these, such as a self-state that reports, “that was me who gave him the answers to the test. I didn’t want him to fail another time.”

Just as for ongoing dissociative amnesia, individuals with dissociative identity disorder report significant difficulties with retrospective autobiographical memory. Memory deficits are commonly experienced as sharply demarcated gaps and/or fragmentary, jumbled recall, typically more densely for childhood and adolescence. Neither pattern is consistent with research on normal autobiographical memory, which describes a graded recall declension with more distant life epochs recalled in less detail than more recent ones. Notwithstanding methodologic limitations, individual case studies using autobiographical memory paradigms have found marked alterations in retrospective autobiographical memory in dissociative identity disorder. This includes reports of no available life memories before early adolescence and/or gaps in memory for specific periods, such as between 8 and 12 years old.

Table 20–14.
Dissociative Identity Disorder Process Symptoms

Presence of dissociative identity disorder self-states

Switching behaviors (identity alteration)

Identity confusion

Passive influence symptoms/interference phenomena between self-states
  - Made feelings from self-states
  - Made impulses from self-states
  - Made actions by self-states
  - Thought insertion from self-states
  - Thought withdrawal by self-states
  - Self-states’ voices commenting on behavior
  - Self-states’ voices arguing

Multimodal hallucinations or pseudohallucinations (may occur in a flashback)
  - Visual
  - Auditory
  - Tactile
  - Olfactory
  - Gustatory
  - Somatic

Mental status examination questions for dissociative identity disorder process symptoms

If answers are positive, ask the patient to describe the event.

Make sure to specify that the symptom does not occur during an episode of intoxication.

1. Do you act so differently in one situation compared to another situation that you feel almost like you were two different people?
2. Do you ever feel that there is more than one of you? More than one part/side/aspect of you? Side of you? Do they seem to be in conflict or in a struggle?
3. Does that part(s)/side(s)/aspect(s) of you have its (their) own independent way(s) of thinking, perceiving, and relating to the world and the self? Have its (their) own memories, thoughts, and feelings?
4. Does more than one of these entities take control of your behavior?
5. Do you ever have thoughts or feelings, or both, that come from inside you (outside you) that you cannot explain? That do not feel like thoughts or feelings that you would have? That seem like thoughts or feelings that are not under your control (passive influence)?
6. Have you ever felt that your body was engaged in behavior that did not

Dissociative or posttraumatic thought disorder

Posttraumatic stress disorder–based cognitive distortions

Trance logic (toleration/rationalization of logical inconsistency in the hypnotized and/or dissociated state)

Other thought process abnormalities

Disorganization due to switching, passive influence

Posttraumatic suspiciousness

Literal and concrete alternating with abstract

Linguistic usage: Refers to self as we, they, us, him, her, etc.

Depersonalization and derealization symptoms
  - Depersonalization
  - Out-of-body experiences (In DID, sometimes experienced as observing oneself from "inside" oneself.)
  - Body parts changing in size

seem to be under your control? For example, saying things, going places, buying things, writing things, drawing, or creating things, hurting yourself or others, etc.?

That your body does not seem to belong to you?

7. Do you ever feel you have to struggle against another part/side/aspect of you that seems to want to do or to say something that you do not wish to do or to say?
8. Do you ever feel that there is a force (pressure, part) inside you that tries to stop you from doing or saying something? That you struggle with?
9. Do you ever hear voices, sounds, or conversations in your mind? That seem to be discussing you? Commenting on what you do? Telling you to do or not do certain things? To hurt yourself or others? That seem to be warning you or trying to protect you? That try to comfort, support, or soothe you? That provide important information about things to you. That argue or say things that have nothing to do with you. That have names. Men? Women? Children?
10. I would like to talk with that part (side, aspect, facet) of you (of the mind) that is called the "angry one" (the Little Girl, Janie, that went to Atlantic City last weekend and spent

  (metamorphopsia)
- Emotional numbing
- Loss of connection to the body
- Loss of connection to one's own memories
- Derealization
- *Déjà vu*
- *Jamais vu*
- *Déjà vécu*
- Dream-like states

lots of money, etc.). Can that one come forward now, please? If that feels too distressing, can you listen inwardly and let that aspect of the mind speak to you, and you tell me what is said?

11. Do you frequently have the experience of feeling like you are outside yourself, inside yourself? Beside yourself, watching yourself as if you were another person?
12. Do you ever feel disconnected from yourself, your thoughts, or your body as if you (your body) were not real? Like a robot? Mechanical?
13. Do you frequently experience the world around you as unreal? As if you are in a fog or a daze? As if it were painted? Two-dimensional? Flat?
14. Do you ever look in the mirror and not recognize yourself? See someone else there?
15. Do you feel like you have lost the capacity for feeling? For emotions? For affection toward others? For hunger? Thirst? Pain?

Adapted from Loewenstein RJ. An office mental status examination for chronic complex dissociative symptoms and multiple personality disorder. *Psychiatr Clin North Am.* 1991;14(3):567. Copyright © 1991, with permission from Elsevier. doi:10.1016/S0193-953X(18)30290-9.

Autobiographical memories may have a depersonalized or derealized quality where recalled events seem to be memories of a dream; as if the patient had seen them happen to someone else; as if they had read about the events; or heard about the events from someone else. Some dissociative patients report that they only know about major parts of their past from discussions with friends or relatives or from looking at photographs of themselves in past activities. Other individuals may describe systematized amnesias with lack of memory for certain categories of experience, such as recall of school events, but normal memory for home events; inability to recall a specific family member; and/or failure to recall some everyday experiences (e.g., failure to recall snow despite growing up in a New England state).

Ms. B was seen in consultation in a hospital after becoming confused at work, unable to recall how to do her job, at which she was highly proficient. She reported lack of recall of other recent events, such as coming to the hospital. An extensive work-up for medical causes of memory pathology was entirely normal and the neurologic consultant opined that this was a case of “hysterical” memory loss. On closer questioning about memory, Ms. B noted that she had “always had a lousy memory.” For example, she reported that “my memory is like Swiss cheese.” She described gaps in her memory for years in grammar school, high school, and college. She could not recall her wedding, but “knew” it happened from pictures. She could not recall her school or teachers in first to fourth grades but recalled school and teachers in fourth to seventh grades. Despite many yearbook pages devoted to her accomplishments in high school, she was perplexed at her limited recall of those years. Her memory was reportedly more continuous once she left home for college. She stated, “I have no memory for my older brother, not at all. They [family members] said he was a holy terror, violent, out of control, using drugs. He’s in prison now for rape.” She attempted to be “perfect” but would have puzzling lapses of recall in school, for events friends reported to her, and embarrassing sexual situations that went against her moral values. These problems continue into the present with her being told of excellent presentations she has made at work, for which she denied any recall. She reported lack of any artistic ability but at times she can paint brilliantly (“I do it, I can just do it”), moving her hand in an elegant depiction of a brush stroke. Old friends make fun of her difficulties with recall of vacations and other events, even when there are photos of the events. She reported feeling “different ages at different times, sometimes more than one at the same time.” She frequently found herself having driven somewhere without recall of the drive. She discovered many items in her closet that she cannot recall purchasing. She has been approached by people she does not know, but insist that they have done things with her, gone places with her, and so on. She rationalizes this, and many other memory problems, as "well, isn't this just normal, to have a lousy memory? Maybe I just have a face that's familiar." After beginning psychotherapy, Ms. B identified self-states that accounted for many of these unrecalled situations. She also began to remember continual violence and chaos in her family as well as repeated rape beginning as a small child by her now imprisoned brother and his friends—reportedly ignored by her parents, despite finding evidence of the assaults (e.g., bloody underpants that her mother complained required repeated washing).

Table 20–15.
Autohypnotic Symptoms in Dissociative Identity Disorder

Spontaneous trance
Deep enthrallment
Spontaneous age regression
Negative hallucinations (not perceiving aspects of the self or environment that are present)
Hidden observer phenomenon (awareness of what is occurring when in the hypnotized state—e.g., experiencing pain in a study of hypnotic analgesia)
Trance logic (tolerance of logical inconsistency during a trance state)
Voluntary analgesia/anesthesia
Eye roll, eye blinking, etc., with switching

Mental status examination questions for spontaneous autohypnotic symptoms

1. Do you frequently space out, trance out, block out, or withdraw from the world around you? Perhaps by putting yourself in a pleasant scene or place in your mind? By focusing your attention on something inside or outside of you? (spontaneous trance)
2. Do you get so wrapped up in a book or a movie that you can completely block out everything else around you? As if the world could end, and you would still be completely engrossed in that activity? (enthrallment)
3. Do you feel that you are different ages at different times? Do you ever feel like you get little? Like you become a child or an adolescent again? When this happens, does it feel like your body changes in size? Do you experience it in your mind only, or does your whole perception of yourself and the world change also? (spontaneous age regression)
4. Do you ever feel like you are more than one age simultaneously? Child and adult? Teenager and adult? Older than your age and your age? Other combinations? More than two ages at once?

5. Do you ever *not* see or hear what is going on around you? Do you or can you block out people and things altogether? (negative hallucinations)
6. Are you able to block out (ignore) physical pain if you want to? Wholly? Partly? Always? Sometimes? (voluntary analgesia)
7. Is there a part of the mind that appears to be observing: aware of and recalls what is occurring when other parts do not? (hidden observer)

Adapted from Loewenstein RJ. An office mental status examination for chronic complex dissociative symptoms and multiple personality disorder. *Psychiatr Clin North Am.* 1991;14(3):567. Copyright © 1991, with permission from Elsevier. doi:10.1016/S0193-953X(18)30290-9.

INTER-IDENTITY AMNESIA. Since the 19th century, inter-identity amnesia has been studied as a paradigm for amnesia in dissociative identity disorder. Here, it is hypothesized that two specific self-states reliably and repeatedly report complete amnesia between themselves. The notion of relatively stable, fixed “two-way” amnestic states is based on the classical notion of dissociative identity disorder as made up of unchanging, structured “personalities” whose characteristics are reliably reproduced on repeated assessment. This contrasts with the current understanding of dissociative identity disorder as a posttraumatic developmental disorder involving dynamic self-state systems based on a variety of intrapsychic, interpersonal, and psychosocial needs over developmental time in a sociocultural context and a phenomenology based on overlap, interference, and intrusion as well as distinct state switching. As in the psychobiological studies, attempts to replicate the classic studies at the NIMH in the 1980s found that, compared to simulating controls, participants with dissociative identity disorder often could not maintain a stable state throughout testing and/or were disrupted by subtle switching and overlap/interference.

Table 20–16.
Comparison of Non–Possession-Form and Possession-Form Presentations of Dissociative Identity Disorder

Characteristic	Non–Possession-Form	Possession-Form
	Internal aspects of the person, usually representing idiosyncratic ages, genders, and functions related to the person's life and intrapsychic needs (e.g., early childhood, adolescence, protector)	External agents usually of supernatural or spiritual origin following collectively defined identities (e.g., the spirit of a dead person or demons, deities, spiritual beings, animals, or mythical figures)
Behavior of alternative identity	Patterned by characteristics of identity: childlike, aggressive, fearful, sexual	Patterned by collective representations of identity: deity, demon, ghost
Elaborateness of presentation	Typically reticent, inhibited, obsessional	May be dramatic (e.g., shrieking, thrashing) or reticent
Relationship with trauma	Linked systematically to extreme, repeated, and	Associated with traumatic events and nontraumatic stressors (e.g., romantic disappointments). However,

	early childhood maltreatment	systematic studies of childhood trauma are lacking
Onset and presentation to others	Typically gradual and hidden, upon accumulation of traumatic exposures	Gradual and nonspecific (e.g., diverse somatic complaints) or abrupt and specific (e.g., sudden transition to altered identity)
Course	Chronic; in relapsing and remitting pattern	May be transient and episodic, although chronic, relapsing course has been observed
Extent of misdiagnosis	In high-income countries, an average of 3.9 incorrect diagnoses and 7–12 years in care before correct diagnosis	Unknown
Treatment	Multiphasic treatment like that for other complex posttraumatic dissociative disorders	Often exorcism, initiation into spirit devotion cult, and/or specialized role in healing practice, typically supervised by indigenous healer. However, exorcism may fail, and may be experienced as traumatic to the person, with alienation from religion and spiritual life. Then, may respond to phasic treatment as per the non–possession-form,

and may experience self-state system phenomena more typical of non–possession-form.

Despite this, Rafaele Huntjens and her colleagues as well as other cognitive researchers have carried out a series of comparisons of simulators with individuals with dissociative identity disorder who report mutually amnestic self-states. Studies included tests of neutral episodic information, perceptual and conceptual priming, procedural memory, transfer of trauma-related words, and stimulus valence as shown by affective priming. They found no objective evidence of inter-identity amnesia on any of these tests and concluded that dissociative amnesia in dissociative identity disorder might have more to do with subjective appraisal and "metamemory" than with actual lack of accessibility of memory between self-states. Paradoxically, in another study based on implicit and explicit memory tasks, different researchers found that individuals with reported *co-consciousness* (i.e., the ability to completely share memory and information) *did* have inter-identity amnesia on memory tasks between self-states.

Using the same paradigm of purportedly mutually amnestic self-states, Huntjens and colleagues also studied autobiographical memory in dissociative identity disorder. Here, using constructs from structural dissociation theory, the mutually amnestic self-states were conceptualized as "traumatic" and "normal" identity states, with the former reporting recall of traumatic memories as first-hand autobiographical memory and the latter claiming complete amnesia for these. In two standard autobiographical memory paradigms, these researchers could not document inter-identity amnesia for autobiographical memory between the pairs of self-states. To perform these types of studies, as well as neurobiology studies using a similar paradigm based on structural dissociation theory, dissociative identity disorder individuals need to be stabilized in treatment sufficiently that they can more reliably switch among "traumatic" and "normal/neutral" self-states and, even with difficulty, are more able to maintain state stability during testing. As treatment outcome studies show, successful treatment response leads to significant decreases in dissociative amnesia barriers among self-states and could contribute to the findings.

In an additional study by Huntjens' group, individuals with dissociative identity disorder were compared with several control groups, including participants with PTSD resulting from multiple episodes of childhood sexual abuse and maltreatment but without dissociative identity disorder. The groups with PTSD and dissociative identity disorder both showed patterns of *overgeneral autobiographical memory* with difficulty retrieving a specific event from memory. This pattern may be related to avoidance of recalling specific information about traumatic memories to attenuate intrusive symptoms. This study concluded that overgeneral memory processes may maintain dissociated memory in traumatized individuals, especially those with severe, multiple early life traumas.

In a directed-forgetting paradigm between subjectively amnestic dissociative identity disorder self-states, findings by another group of researchers supported partial and state-dependent explicit memory difficulties, indicating possible encoding problems as well as retrieval inhibition between self-states. Amnesia between states was not rigidly compartmentalized, consistent with the view that, in dissociative identity disorder, access to memory between self-states may vary in complex ways depending on various factors. In this study, retrieval inhibition began to break down with more emotionally charged stimulus words. This confirmed the observations of both the psychological profile and fMRI studies, where subjects attempted to dissociate to attenuate posttraumatic intrusions but, when unable to do so, devolved into experiencing severe intrusive symptoms.

The latter two studies are consistent with the current view of dissociative identity disorder. More definitive, if more difficult, future studies of dissociative amnesia in dissociative identity disorder could be based on real-time experiential sampling methodology with assessment of psychobiological variables that may affect autobiographical memory accessibility (e.g., interpersonal stress, increased PTSD symptoms, etc.).

Process Symptoms. Dissociative process symptoms include depersonalization and derealization, dissociative hallucinations, passive influence and interference experiences, and dissociative cognition (see Table 20–14). Some authorities include dissociative alterations in self-state (shifting/switching) under this category. Depersonalization and derealization are commonly reported in dissociative identity disorder and may include profound out-of-body experiences as well as observing oneself from *within* one's mind. Patients frequently report feeling spaced out or disconnected from themselves and others. The world is perceived as distant or unreal, with a hazy or foggy quality. Patients may report feeling at times as if they exist in a waking dream state. Out-of-body experiences commonly take the form of watching oneself from a distance (inner, as well as outer), as if observing another person, with little or no ability to affect one's actions.

Dissociative Hallucinations. Based on data from the MID, dissociative hallucinations are reported by close to 100% of individuals with dissociative identity disorder; they are common in both possession-form and non–possession-form presentations. In non–possession-form states, they are typically experienced as voices originating from within the mind, as opposed to coming from the outside. Individual-hallucinated voices often represent a distinctive age and gender, and in possession-form presentations a type of agent, such as a demon or deity. These voices may comment negatively about the patient, argue with each other, command the patient to perform certain acts, discuss neutral topics, or sometimes provide useful information or comfort. The voices may be indistinct but described with a personified quality, such as being at a party but not being able to make out all the conversations around one.

Patients with non–possession-form dissociative identity disorder generally recognize that the voices are hallucinations and may be reluctant to reveal their existence for fear of being considered psychotic or "crazy." Possession-form patients may share the same fear, but usually attribute the voices to culturally accepted spiritual agents. Ultimately, dissociative voices often are identified with specific self-states or possessing agents. Visual hallucinations often take the form of detailed images with traumatic or frightening content. Other visual hallucinations may be understood as depicting self-states or may even have a complex artistic quality. Tactile, somatosensory, gustatory, and olfactory hallucinations may also occur, leading to misdiagnoses of seizure disorder or other brain disorders. Intrusive posttraumatic flashbacks and images may also be experienced as complex multimodal hallucinations. Musical hallucinations also may occur.

Studies using the MID found that, compared to SCID-diagnosed individuals with a psychotic disorder, those with dissociative identity disorder experience hallucinations as ego-alien and do not have delusional explanations for them or passive-influence experiences. Compared to individuals with psychosis, persons with dissociative identity disorder scored significantly higher on scales measuring child voices, angry voices, persecutory voices, and voices arguing and commenting on the individual's thoughts and behavior. Hallucinations in dissociative identity disorder, but not in psychosis, are related to posttraumatic responding.

Individuals with dissociative identity disorder also report *negative hallucinations*—in which external percepts and stimuli are not consciously registered. This can occur in any sensory modality: auditory, visual, tactile, gustatory, and olfactory, as well as pain perception. Most positive and negative hallucinations in individuals with dissociative identity disorder are not a manifestation of a psychotic process. They are similar to phenomena experienced in deep trance states among highly hypnotizable individuals who can experience hallucinations in all sensory modalities. As dissociative flashbacks in PTSD often involve multi-sensory hallucinosis, posttraumatic factors likely also contribute to multimodal hallucinations in dissociative identity disorder. Unfortunately, in current practice, hallucinatory phenomena of any kind, but especially voices commenting or arguing, are *automatically* diagnosed as psychotic, without considering their differential diagnosis and/or exploring the phenomenology more deeply. This ignores a large literature on nonpsychotic hallucinations in the general population and “hysterical hallucinations” commonly described in the older psychiatric literature.

Passive Influence Symptoms and Symptoms of Overlap, Intrusion, and Interference among Self-States. Passive influence and interference symptoms include many apparent “first-rank” Schneiderian symptoms (FRS), such as audible thoughts, voices arguing with each other, influences playing on the body, thought withdrawal and insertion, and “made” feelings, impulses, and actions. These symptoms were once considered pathognomonic of schizophrenia but they can also be found in patients with affective and other psychotic disorders. As noted in the section on Nosology, studies of passive influence symptoms in dissociative identity disorder find them more prevalent than in studies of psychotic disorders. As with hallucinations, individuals with dissociative identity disorder may feel confused, puzzled, or ashamed of these “crazy” symptoms. Typically, these individuals have evocative, often personified descriptions of passive influence, such as, “I feel like someone else wants to cry with my eyes”; “I keep getting these “ghost” feelings….” Unless suffering from actual comorbid psychosis, persons with dissociative identity disorder almost never endorse symptoms of thought broadcasting, audible thoughts, and delusional perception. As with hallucinations, studies comparing passive-influence symptoms in dissociative identity disorder and psychotic disorders found that a salient differential diagnostic factor was lack of delusional explanations by individuals with dissociative identity disorder. Patients may report strong affects or impulses that they experience without a sense of personal ownership but with a peremptory sense of intrusion and control. Complicating their differentiation from psychotic conditions, possession-form presentations of passive-influence are usually experienced as external and “real.”

Dissociative Thought Disorder. Distinctive responses to standardized projective testing can often be helpful in distinguishing dissociative patients from other diagnostic groups, such as patients with affective disorders, nondissociative forms of PTSD, personality disorders, psychotic disorders, and factitious disorders. Research using projective testing finds distinctive cognitive process markers, including evidence of confusing and contradictory responses to the same stimulus. Other forms of dissociative identity disorder thought abnormalities include profound trauma-based cognitive distortions, such as pervasive self-blame for abusive trauma. These posttraumatic and dissociative cognitive distortions may seem so intense as to be virtually delusional at times. However, in most cases, one or more self-states will report understanding that these attributions are incorrect or that they would be incorrect if someone else believed them. Individuals with dissociative identity disorder may also appear to have derailment and loosening of associations caused by rapid, dysfunctional shifting of states that overlap and interfere with each other.

Dissociative Alterations in Identity

NON–POSSESSION-FORM PRESENTATIONS. Clinically, *non–possession-form* dissociative alterations in identity may first manifest through odd first-person plural or third-person singular or plural self-references. Patients may refer to themselves using their own first names or make depersonalized self-references, such as “the body,” when describing themselves and others; for example, “The father hurt the body, so she was upset. We tried to protect her, but it didn’t work.” Patients often describe a profound sense of concretized internal division or personified internal conflicts between self-aspects. In some instances, these self-states may have proper names or may be designated by their predominant affect or function: for example, “the angry one” or “the wife.” Patients may suddenly change the way in which they refer to others: for example, “the son” instead of “my son.”

Clinicians may observe a group of behaviors collectively referred to as *state shifting or switching behaviors*. These may be accompanied by intra-interview amnesias, where the patient does not seem to recall or is confused about the process and content of an assessment or therapy session. These *microdissociative episodes* may manifest via abrupt shifts in train of thought or sudden inexplicable changes in affect or in rapport and may then quickly shift back again. A variety of physical signs, including pronounced upward eye rolls or bursts of rapid blinking and eyelid fluttering, may occur in conjunction with microdissociative episodes. The patient's tone of voice and manner of speaking, posture, and demeanor may show subtle to marked alteration. When clinicians encounter evidence of possible microdissociative episodes, using nondirective, open-ended questions, they should seek to clarify what the patient is experiencing and can recall. Dissociative amnesia is a common concomitant of these episodes; sometimes they may extend for a few minutes or even the entire psychotherapy session. These cognitive, behavioral, and physical shifts are manifestations of personality state shifting/switching or overlap and interference between states, or both.

POSSESSION-FORM PRESENTATIONS. *Possession-form* presentations of state switching may be subtle or dramatic. Onset can be gradual and nonspecific (e.g., diverse somatic complaints) or sudden and specific, in the form of an abrupt transition to the identity state. If the presentation is gradual and nonspecific, transition to the identity may occur spontaneously or after contact with an indigenous healer specializing in the treatment of pathologic possession. Dramatic onsets may be characterized by semi-purposeful movements (e.g., thrashing, falling to the ground, shrieking); aggressive or violent actions directed at self or others accompanied by derogatory verbalizations; and/or specific gestures, comments or requests denoting the appearance of a known possessing agent by (1) reference to standard attributes of culturally recognizable figures or (2) the name and degree of relation of deceased family members or acquaintances. At times the possessed individual "acts out" some narrative aspect of the possessing agent's story to communicate to their social network the suffering inflicted by the spiritual being or the desires of the possessing agent. For example, a young woman, experiencing herself as afflicted by the spirit of a dead adolescent girl who committed suicide by jumping in a well, engaged in suicidal behaviors that evoke these past events. Alternatively, individuals possessed by high-status spirits may demand deference from family members.

Possession by the spiritual agent is frequently episodic, resulting in alternation between the usual identity of the subject and the altered state. The subject in their usual identity appears dazed, exhausted, distressed, or confused about the situation. They may report visual or auditory hallucinations regarding the possessing agent, as well as “unconsciousness” and partial or total amnesia for the altered state. States of consciousness may alternate continuously within one episode over hours to weeks, with each state lasting minutes to hours. This pattern may persist intermittently over months or years.

Ms. C is a 23-year-old bilingual woman (Spanish-English) of Latinx Caribbean descent who was born and raised in a large eastern city of the United States. Her family was active in an Afro-Catholic religious cult involving normative dissociative possession states until, when Ms. B was 5 years old, all members converted to Evangelical Protestantism. Since early childhood, she was thought to have spiritual abilities allowing her to predict the future in her dreams and to identify other co-religionists suffering from demonic possession (by seeing "shadows" around them). These were considered spiritual gifts and were highly prized by her family and congregation. Since age 4, Ms. B was the victim of sexual abuse in the form of touching her genitals without penetration by older individuals outside the family. She also received physical abuse as "punishment' by her parents, reportedly resulting in severe bruising but no fractures. Conflicts with her family led to a potentially lethal suicide attempt at age 15 (medicine cabinet overdose). This led to a 3-week inpatient hospitalization and outpatient follow-up with antidepressants for a diagnosis of major depression with panic attacks. However, after 3 months, she self-discontinued treatment. At age 21, Ms. B began to question her faith and to ask God in prayer whether he and the demons she was seeing were "real." In this context, she experienced a prolonged possession episode that was involuntary and extremely distressing. Onset came during sleep, when she felt a demon attempt to enter her body but not fully succeed. This occurred for several consecutive nights and was followed during the day by the onset of voices mumbling and calling her name. Several days later, she felt "many spirits coming inside of me" and taking over control of her actions. Daily for 3 months, these "demons" spoke through her mouth in a deep and "not human" voice— "like a monster or several voices at once"—and communicated with her family and other church members who came to pray around her. During most of these episodes, Ms. B was aware of the surrounding events but was unable to control her speech or actions. She reported amnesia for some of the possession episodes but reported recall of most of them. She had no visual hallucinations of the main possessing demon. However, she heard him speaking to her, saying that he would “drive her crazy.” At those times, Ms. B felt that God sent her a message of reassurance. Once, she felt evil spirits were pulling her soul out of her body. She saw an angel standing at the foot of the bed, that she believed was sent by God to protect her, described as “invisibly white, see-through white” with wings and no face. This helped calm her. Multiple attempts at exorcism conducted by local church leaders were unsuccessful. A prominent pastor was called in from abroad to assist. The demon claimed to recognize the pastor and alluded to their past encounters during other cases of possession. His exorcisms were also unsuccessful. After 3 months, while she was participating in a religious retreat, she felt God sent her a message, “today is the day.” At that point, “the demon trembled and left.” Since this event, two years prior, she did not suffer another possession state. However, she remains intensely phobic of the dark. She reports auditory and visual illusions when hearing background noises in an otherwise silent room (“sound like voices”) or if sees herself in a mirror from a distance (“my face is scary, as if I had a demon inside again”). One year ago, she had an episode of conversion hemiparalysis that came on gradually over the course of a day and then disappeared. She continued to suffer from major depression and panic attacks.

Table 20–16 compares possession-form and non–possession-form presentations of dissociative identity disorder.

Symptoms Related to Spontaneous Autohypnotic Phenomena. Individuals with dissociative identity disorder commonly exhibit or describe symptoms of involuntary autohypnotic phenomena, consistent with their high scores on standardized hypnotizability scales (see Table 20–15). These include spontaneous trance states that can be clinically disabling; spontaneous or voluntary anesthesia and analgesia; negative hallucinations; spontaneous age regressions; and *trance logic*. The latter is defined as tolerance or rationalization of logical contradiction in the hypnotic state but with ongoing reality orientation. For example, an individual with dissociative identity disorder reported disorientation to time, stating that it was 15 years before the interview. When told the actual year, she said, “Oh. Wow! We’re in the future!” Other autohypnotic symptoms overlap with dissociative process symptoms (e.g., hallucinations, thought disorder, child self-states, passive influence) and PTSD symptoms (e.g., behavioral re-experiencing episodes with multimodal hallucinations and age regression to the time of the trauma).

Other Associated Symptoms. As noted above, non–possession-form dissociative identity disorder is conceptualized as a paradigmatic posttraumatic disorder, with most patients meeting DSM-5-TR diagnostic criteria for PTSD, major depression, current or past substance use disorders, and somatic symptom disorders, as well as commonly engaging in self-destructive and suicidal behavior. Table 20–17 enumerates associated symptoms of dissociative identity disorder.

Table 20–18 lists the most important differential diagnostic considerations for dissociative identity disorder. One must differentiate disorders that are comorbid with dissociative identity disorder from those that must be differentiated from dissociative identity disorder. For example, depression, PTSD, substance abuse, and somatic (previously somatoform) symptoms are very common comorbidities with dissociative identity disorder.

Numerous recent studies have shown a robust relationship between mood disorders and an antecedent history of trauma, particularly childhood sexual abuse. Major depression is increasingly understood as one of the outcomes following traumatic experiences. Most individuals with dissociative identity disorder meet criteria for a depressive disorder, typically characterized as “treatment resistant,” with little or no response to multiple adequate trials of antidepressants with augmentation strategies, mood stabilizers, ECT, and transcranial magnetic stimulation. Careful assessment of the affective state of the individual commonly finds a variety of emotions and cognitions as part of the “depression.” These may include intense shame, profound grief, posttraumatic dysphoria, demoralization, survivor guilt, and distress at one’s life situation (see section on Target Symptoms for Medications in the Treatment of Dissociative Identity Disorder-Mood Disorder Symptoms).

Frequent rapid “mood swings” are common, usually experienced as occurring within minutes or hours. These are usually due to posttraumatic and dissociative state alterations and are much more rapid than even the most rapid cycling documented in bipolar disorder. There may be considerable overlap between PTSD symptoms and those of depressive disorders. These include problems with attention and concentration, disturbed sleep, anxiety (fear in PTSD), and dysphoria/anhedonia. Unlike individuals with mood disorders, those with dissociative identity disorder typically describe a sleep disturbance dominated by nightmares and posttraumatic reactivity with specific phobias of night, bed, and sleep based on childhood nocturnal sexual assaults. Trauma-based sleep symptoms must be specifically inquired about as patients generally will not disclose them when asked generic questions about problems falling asleep, staying asleep, early morning awakening, and/or bipolar hypo/hypersomnia. Under the sway of some self-states, individuals with dissociative identity disorder may describe even several days of hyperactive behavior. However, this is not experienced “across the whole mind,” is not accompanied by typical bipolar sleep symptoms, and typically is not accompanied by manic grandiosity, although an apparent mixed state presentation may be harder to separate from bipolar disorder.

Table 20–17.
Associated Symptoms Commonly Found in Dissociative Identity Disorder

- Posttraumatic stress disorder symptoms
  - Intrusive symptoms
  - Severe dissociative flashbacks and nightmares
  - Avoidance and numbing symptoms
- Persistent and exaggerated negative beliefs/expectations and cognitions about self/world
- Persistent negative emotional state (especially shame/guilt, fear, sadness, anger)
- Detachment/estrangement
- Hypervigilance
- Reckless/self-destructive behavior
- Problems with sleep and concentration
- Symptoms of somatic symptom disorders
  - Conversion and pseudoneurologic symptoms, especially, seizure-like episodes
  - Somatic symptom disorder
  - With predominant pain symptoms such as:
    - Headache, abdominal, musculoskeletal, pelvic pain
    - Psychological factors affecting other medical conditions
    - Medical disorders associated with adverse childhood experiences (see Table 20–7)
    - Perimenstrual disorders (sexual abuse)
    - Irritable bowel syndrome (sexual abuse)
    - Gastroesophageal reflux disease (sexual abuse)
    - Morbid obesity (sexual abuse)
- Affective symptoms
  - Depressed mood, dysphoria, or anhedonia (typically, “depressed all my life”)
  - Brief mood swings or mood lability (within minutes/hours)
  - Suicidal thoughts, plans, intent, and behavior
  - Guilt and survivor guilt
  - Profound shame
  - Helpless and hopeless feelings

Obsessive-compulsive symptoms
  - Ruminations about trauma and its impact on the person
  - Obsessive counting, singing
  - Arranging
  - Washing
  - Checking

Compulsive body movements

Table 20–18.
Differential Diagnosis of Dissociative Identity Disorder

Comorbidity vs. differential diagnosis (see discussion in text of outcomes of childhood trauma and Table 20–7)
Affective disorders
Psychotic disorders
Anxiety disorders
Posttraumatic stress disorder
Personality disorders
Neurocognitive disorders
Neurologic and seizure disorders (e.g., multiple sclerosis, tic disorders, movement disorders)
Somatic symptom disorders
Other dissociative disorders
Factitious disorders
Malingering

OCD symptoms are also commonly found in individuals with PTSD. Obsessive-compulsive personality traits are common in dissociative identity disorder and intercurrent OCD symptoms are regularly found in individuals with the disorder, with a subgroup manifesting severe OCD symptoms. OCD symptoms commonly have a posttraumatic quality, for example, checking repeatedly to be sure that no one can enter the house or the bedroom, compulsive washing to relieve a feeling of being dirty because of abuse, and repetitive counting or singing in the mind to distract from anxiety over being abused.

Child and Adolescent Presentations (also, see section on Developmental Model of Dissociative Disorders). A growing clinical research literature documents the diagnosis and treatment of child and adolescent dissociative disorders, including dissociative identity disorder. In many respects, children and adolescents manifest similar core dissociative symptoms and secondary clinical phenomena as adults. Age-related differences in autonomy and lifestyle, however, may significantly influence the clinical expression of dissociative symptoms in youth. For example, dissociative amnesias and perplexing forgetfulness may be more apparent in school situations than in work or family life. Younger children have a less linear and continuous sense of time and are often unable to self-identify dissociative discontinuities in their behavior. Informants, such as teachers and relatives, may be useful in documenting dissociative behaviors.

Several normal childhood phenomena, such as imaginary companionship, elaborated daydreams, or participation in normal dissociative possession experiences as part of religious rituals, must be carefully differentiated from pathologic dissociation in younger children. For example, preadolescent children with dissociative identity disorder may manifest fewer obvious switching behaviors than adolescents or adults. The clinical presentation may be that of elaborated or autonomous imaginary companions, with these companions taking control of the child’s behavior. This is often experienced as passive influences or voices in the child’s mind commanding the child to behave in certain ways.

Ms. D had been in treatment for dissociative identity disorder for several months when she voiced concern that her 8-year-old daughter was “depressed.” One day, to her psychiatrist’s surprise, she brought her daughter to therapy. During the session, the child began to fidget, then got up and began to walk around the room. The mother suddenly appeared to shift in demeanor, appearance, and voice tone, shouted at the girl in a harsh and angry voice, grabbed the daughter’s arm tightly, and began to angrily slap the child. The little girl attempted to curl up in a fetal position, wailing like a much younger child, while the psychiatrist jumped into the fray, and literally physically separated the mother from beating the child. As he pulled the mother away, the little girl crawled under a table, and said, “I hate you! I hate you!” Both then appeared dazed, the mother wandered back to her chair, the little girl emerged from under the table, looking confused, but appearing more her actual age. When queried about what had happened, both claimed no recall of the events, even when the psychiatrist detailed for them what had just happened. The psychiatrist then arranged for another family member to pick up the little girl and stay with her. He probed the mother until an angry self-state emerged spitting, “the little s–t isn’t mine, it’s *hers*!” and took responsibility for the events in the office and admitted to beating her daughter on numerous occasions. The psychiatrist insisted that the patient report herself to the Department of Social Services (DSS) to investigate her ability to care for her daughter. He made the safety of the child the central part of the patient’s treatment and assisted with the DSS investigation, resulting in additional supervision for the mother and child but placement remaining in the home. He also interviewed the father, a passive man, whose work often took him away from home and who had abdicated child rearing to his wife. A colleague evaluated the daughter, who reported “people in my head,” whom she said talked to her and “told [her] what to do.” The little girl described “two ladies who are twins” in her mind; “one is a nice mommy, and the other is mean, and you cannot tell them apart, so you never know when one or the other is around.” The child said that “a lot of the people in my head are twins, one does it one way, the other the opposite. When that mean lady is around, my twin littler ones are there.” The psychiatrist learned that the mother’s self-state system also was based on many groups of subjectively experienced twin self-states, including the maternal and abusive ones, that reportedly called themselves by the same names as the little girl’s twin mother and abusive self-states. The mother told the psychiatrist that *her* mother also likely shares the diagnosis of dissociative identity disorder and was born with a twin sister who died shortly after birth. Work with the abuser self-states—actually based on the grandmother—led to cessation of violence toward the child. Another group of self-states acknowledged bringing the child to the mother’s therapy to expose the maltreatment they felt was impossible to stop without outside assistance. Careful monitoring over many years continued in both the mother’s and child’s psychotherapies.

Pathology and Laboratory Examination

A clinical mental status examination (Office Mental Status Examination) based on the various symptom clusters described in Tables 20-13 through 20-15 can be helpful in the diagnosis of dissociative identity disorder. Questioning about amnesic, depersonalization/derealization, autohypnotic, PTSD, mood, passive-influence, and somatic symptoms usually precedes detailed inquiry about symptoms related to the self-states. Careful in-depth questioning about these phenomena may readily bring forth information related to actions/interactions of self-states. For example, discussion of complex amnesia experiences may readily lead to questions about how purposive behaviors can occur without apparent memory. Discussion of PTSD symptoms may lead to manifestations of self-states related to traumatic experiences.

The development and increasing use of screening instruments and standardized diagnostic interviews (see Table 20–2) has contributed to the rising number of identified cases of dissociative identity disorder over the last decade. Before the development of these assessment tools, hypnotic or amobarbital interviews were used primarily for diagnosis of the disorder. Due to the current academic and forensic controversies surrounding dissociative disorders and trauma memory, as well as most clinicians' lack of familiarity with these modalities, it is prudent to reserve these interventions for emergency situations when other methods of assessment have failed; for example, in a female patient who is emergently hospitalized after engaging in repeated, dangerous nocturnal fugues and who cannot account for these behaviors despite intensive interviewing. These interventions should optimally be conducted by a clinician experienced in their use and in the differential diagnosis of dissociative disorders. Clear, informed consent should be obtained for use of these interventions for diagnosis of dissociative identity disorder and recall of traumatic experiences. Informed consent should include education that memory material reported under hypnotic or pharmacologic conditions should not be considered any more or less valid than memories recalled under other conditions. Pharmacologically facilitated interviews (e.g., amobarbital interviews) are now considered conscious sedation by The Joint Commission and can only be performed with the assistance of an anesthesiologist.

Differential Diagnosis

On average, more than 6 to 12 years pass in high-income countries between first psychiatric contact and the diagnosis of dissociative identity disorder. The time lag between symptom onset and referral for treatment in possession-form presentations may vary, depending on the dramatic or gradual nature of the symptoms. Treatment failure and apparent “treatment resistance” may be an important differential diagnostic indicator of dissociative identity disorder, especially after multiple, at best only partially successful, psychotherapy (e.g., CBT, DBT, psychoanalytic), medication, ECT, and rTMS trials with adequate dosages for appropriate time frames. Rather than displaying histrionic flamboyance, it is estimated that over 95% of individuals with non–possession-form dissociative identity disorder are reticent about revealing their dissociative symptoms, especially hallucinations, amnesia, and self/identity divisions. They most commonly present as being relatively inhibited and obsessional, with mood and somatic complaints, and typically receive an average of three or more psychiatric diagnoses before their dissociative identity disorder is recognized. Symptom patterns distinguishing dissociative identity disorder from mood disorders, psychotic disorders, and personality disorders have been discussed earlier in this section.

Table 20–18 lists the most common disorders that must be differentiated from dissociative identity disorder.

However, because many of these disorders can coexist with dissociative identity disorder, as well as be mimicked by dissociative identity disorder, differential diagnosis may be complex. A small subgroup of individuals with dissociative identity disorder has an intercurrent psychotic disorder with a mixture of psychotic thought derailment, both psychotic and dissociative hallucinations, and at times tangential, bizarre associations and persecutory delusions that may imbricate with posttraumatic mistrust and interpersonal hypersensitivity. Often, they can distinguish between psychotic hallucinations and those of the internalized voices of the self-states although this can be harder to differentiate in possession-form cases. Individuals with dissociative identity disorder with comorbid psychosis may describe thought broadcasting, audible thoughts, ideas of influence and reference, and delusional explanations for these phenomena. Even so, some of these patients may preserve insight that psychotic symptoms are different from posttraumatic and dissociative symptoms. Some of these patients may have a mixture of delusional beliefs that amplify posttraumatic hypersensitivity which can be difficult to entangle (e.g., “they all can see the shame that’s on my face,” as opposed to “everywhere I go, the victim people follow me. I hear them whispering about my bad face”).

Individuals with actual intercurrent bipolar disorder may experience sustained hypomanic or manic episodes—different in duration and quality from dissociative mood shifts—and more typical bipolar depression symptoms that are superimposed on the patient’s baseline chronic complex posttraumatic mood state. Stabilization of the actual intercurrent bipolar disorder symptoms with mood-stabilizing medication or ECT is necessary to allow the patient to participate fully in phasic trauma treatment for dissociative identity disorder. However, depressive symptoms usually are only returned to the patient’s more usual chronically dysphoric baseline.

Factitious, Imitative, and Malingered Dissociative Identity Disorder

Concerns about factitious and malingered dissociative identity disorder are common. Recently, reports are increasing of individuals claiming to have dissociative identity disorder who do not meet diagnostic criteria for the disorder when carefully assessed clinically or with structured interviews, such as the SCID-D-R. There may be a mixture of factors leading to this presentation, including misdiagnosis, factitiousness, and assumption of a social role by an abuse victim. Dutch researchers have named this *imitative dissociative identity disorder* when there does not appear to be conscious simulation. Patients may build their lives around their diagnosis and are commonly supported in so doing by concerned others and by their therapists.

Clinical indicators of falsified or imitative dissociative identity disorder may include those typical of other factitious or malingering presentations, such as symptom exaggeration, lies, use of symptoms to excuse antisocial behavior (e.g., amnesia only for bad behavior), amplification of symptoms when under observation, refusal to allow collateral contacts, legal problems, and *pseudologia fantastica.* In contrast, individuals with genuine dissociative identity disorder usually are reticent to endorse symptoms and appear confused, conflicted, ashamed, and distressed by symptoms and trauma history, alternating between endorsing then denying them. In forensic settings, genuine individuals with dissociative identity disorder may refuse to use an insanity defense, much as many paranoid patients do. The feigning patients frequently show little dysphoria about their disorder and posttraumatic triggers and appear to be gratified by having the diagnosis. These imitative and factitious patients fit the socio-cognitive model of dissociative identity disorder, with one exception: rigorous diagnostic efforts show that they do not meet diagnostic criteria for the disorder.

The SCID-D-R has been used to help distinguish genuine from feigned/imitative dissociative identity disorder in clinical and forensic contexts. On the SCID-D-R, individuals with fabricated dissociative identity disorder tend not to describe the typical comorbid conditions and symptoms associated with the disorder—such as amnesia, identity confusion, and identity alteration—although they may report high levels of depersonalization and derealization. Instead, they often describe symptoms in a jargon-filled way (“I’m dissociating” or “I’m switching”) and cannot readily explain their subjective experience on follow-up questions on the SCID-D-R or in clinical interviews. Most frequently, they show personality disorder traits on structured diagnostic interviews but do not meet diagnostic criteria for a dissociative disorder.

There are no absolute clinical indicators to differentiate imitative, factitious, or malingered dissociative identity disorder from the actual disorder. Some persons with the bona fide disorder become invested as their main social role in a dissociative identity disorder identity or that of being a trauma survivor. Some patients with possession-form presentations use or exaggerate symptoms for overt external gain. For example, a late adolescent newlywed had recently moved in with her husband and his family of origin. She reported being treated “badly” by her mother-in-law; this allowed her to return home for prolonged periods to “treat” her possession state. Just as individuals with non–possession-form presentations can be reticent and secretive about their illness, this may also be the case for many possession-form individuals. In some instances of the latter, possessing self-states may remain undisclosed until a specialist healer exposes them.

There are also patients with the actual disorder who report factitious histories, produce factitious crises, and create factitious self-states. Some patients report that their only happy childhood experiences were when they were hospitalized medically. Like more typical patients with factitious disorder, they seek out and try to remain hospitalized in psychiatric and medical settings. Münchhausen syndrome and Münchhausen syndrome-by-proxy have also been reported in patients with dissociative identity disorder and their first-degree relatives. In forensic contexts, individuals with dissociative identity disorder seeking exculpation have malingered greater severity of dissociative and posttraumatic symptoms.

Ms. F, a 48-year-old divorced mother of three, was referred to a trauma disorder inpatient unit by her managed care company for consultation about diagnosis and treatment. She had been diagnosed with dissociative identity disorder after doing “inner-child work” with her outpatient therapist, as well as during her outpatient therapist’s group for sexually abused women. Ms. F reported an extensive history of childhood neglect, as well as physical and sexual abuse outside her family from ages 8 to 12. She had numerous hospitalizations for suicidal ideation and nonsuicidal self-injury. She was on disability and had been investigated several times by her state division of child welfare, which had recently placed her children in foster care. On arrival in the trauma unit, Ms. F proclaimed to patients and staff that, “I have dissociative identity disorder. I was sexually abused.” She did not seem distressed by either of these ideas. She repeatedly stated, “I’m switching, I’m switching,” although staff members experienced with dissociative identity disorder patients did not observe any of the typical signs of dissociative identity disorder switches nor had they observed a patient with dissociative identity disorder announce switches in this way. Ms. F became irritable when staff attempted to dissuade her from speaking about this in the milieu. She was evaluated clinically, observed by nursing staff in the ward milieu, and given a battery of psychological tests, including the DES, the SCID-D, PTSD assessments, the Rorschach, and personality assessments. On the clinical evaluation, Ms. F did not endorse complex chronic amnesia symptoms, amnesia for her life history, autohypnotic or passive influence symptoms, or inner voices. She endorsed episodes of depersonalization and derealization and reported only “child and baby alters” but could describe little about them when questioned. She would say things like, “See, now I’m like a child. Can’t you tell? That’s one of my littles.” She did not endorse any sense of fear, shame, conflict, or inner struggle around “switching” nor were any of the typical phenomena of switching noted at these times. Ms. F became vague when discussing reported childhood sexual abuse. She became angry when confronted about contradictions in her accounts, the dramatic nature of her presentation, and discrepancies between her history and her emotional reactions. She appeared uninterested in individual or group psychotherapy that sought to teach her more effective ways to manage symptoms and increase coping. On psychological and diagnostic testing, Ms. F did not endorse most items consistent with clinical dissociation other than depersonalization and derealization. She did not display typical PTSD or dissociative reactivity to inquiries about the abuse history nor did she meet diagnostic criteria for PTSD. She did not show behaviors typical of individuals with dissociative identity disorder during a serendipitous event while being tested, remaining unaffected when another patient suddenly required aggression management on the ward near the testing room, whereas other patients reacted with severe PTSD and dissociative symptoms. Ms. F was diagnosed with DSM-IV-TR factitious disorder with psychological symptoms, mood disorder NOS, and personality disorder NOS with histrionic, borderline, and dependent features. However, she did not appear to be consciously feigning and seemed to believe in her self-reports of switching, self-states, and a history of childhood sexual abuse. Both Ms. F and her outpatient provider were resistant to the new diagnostic formulation.

Recently, a group of researchers has explored the standard inventories used in forensic settings for detection of malingering, such as the Structured Interview for Reported Symptoms (SIRS), developed by Richard Rogers. They found that the SIRS misclassified individuals with SCID-D-R–diagnosed dissociative identity disorder as feigning. Rogers replicated this finding and modified the SIRS, now as the SIRS-II, to classify dissociative trauma survivors more accurately. At this time, assessment of possible factitious or malingered dissociative identity disorder presentations—or factitious or malingered amplification of symptoms in persons with actual dissociative identity disorder—requires a complete clinical and psychometric evaluation, in addition to a review of old records, data from collateral informants, and other forms of assessment. Psychometric instruments, which usually must be administered by a psychologist familiar with how symptoms of dissociation present on standard measures, generally include the DES, MID, SCID-D-R, Rorschach, Thematic Apperception Test (TAT), Trauma Symptom Inventory (TSI), Minnesota Multiphasic Personality Inventory (MMPI)—although this also may misclassify dissociative individuals—Personality Assessment Inventory (PAI), SIRS-II, and additional assessment measures of malingering, and assessment measures for depression and PTSD, among others.

Course and Prognosis

Little is known about the natural history of untreated dissociative identity disorder. Small studies of patients with dissociative identity disorder diagnosed in middle age and in geriatric populations have shown that severe dissociative identity disorder symptoms can persist or can appear in older-age patients. A relapsing and remitting course is common and appears to be the modal pattern in possession-form cases. Also, patients can mask or suppress symptoms for periods of time. Both phenomena may be mistaken for complete spontaneous remission of the disorder. Possession-form presentations tend to be described in case series as more transient and episodic than non–possession-form cases. It is possible that the emergence of the possession-form type is facilitated by the same porosity of the self that patterns the external attribution of alternate self-states, leading to more transient states after a wider range of stressors. However, it is also possible that possession-form cases only seem transient and episodic, but that this is the overt presentation of phenomena that persist in more subtle form over time. Support for this possibility comes from data on relapse of possession-form presentations. In India, for example, relapse within 15 years was not infrequent, and was also recorded after two to three decades. Four of 15 subjects in one study were found to have suffered previous episodes, and one subject in another series reported three consecutive postpartum attacks. In China, individuals with chronic possession-form illness may be hospitalized after a history of recurrent episodes. Systematic research with larger samples over time is required to clarify this issue.

Some individuals with untreated dissociative identity disorder remain involved in abusive relationships or violent subcultures, resulting in their retraumatization and likely that of their children, with the potential for multigenerational transmission of trauma and the disorder (see Case 4). Experience with many dissociative identity disorder cases suggests that there are diverse subgroups of afflicted individuals. Some function very well for long periods while others have severely impaired and dysfunctional life trajectories often beginning early in development. The former may only show symptoms when severely stressed or overwhelmed; the latter may be continuously and severely symptomatic and come to clinical attention as children or adolescents. As described in the section on Nosology, there are high rates of disability in individuals with dissociative identity disorder in the community and in partial hospital and inpatient settings.

Patient presentations and prognosis vary somewhat across the life span. Children with non–possession-form dissociative identity disorder show many dissociative symptoms and behaviors but typically have fewer and less-crystallized self-states that are less invested in their individuality. If diagnosed early, and removed from abusive or neglectful environments, childhood dissociative identity disorder may resolve relatively rapidly with appropriate treatment. In adolescence, self-states become more distinct and more invested in their autonomy. Additional self-states may appear that are associated with life stresses, such as academic, athletic, social, or sexual challenges, and the self-sate system dynamics may become more complicated and polarized. In general, adolescents have a poorer prognosis than children or adults in part because, like other adolescents, they are often not invested in their treatment and are reticent to open up to adults. Better outcome has been reported in adolescents with dissociative identity disorder when their families were successfully engaged in treatment. Young adults typically present in crisis and have a layering of affective, somatic, posttraumatic, and personality disorder symptoms in addition to their core dissociative pathology.

First presentations of dissociative identity disorder in older patients frequently involve a life event, such as the loss of a job, death of a parent, or revictimization, such as a rape, that reactivates earlier conflicts and destabilizes the self-state system. Achieving sobriety in a substance-abusing patient with dissociative identity disorder may precipitate overt symptoms of the disorder and of PTSD. In adult patients, there appear to be treatment-responsive and treatment-refractory subgroups of individuals. Time course of improvement also may vary, with some patients improving relatively quickly and others requiring intensive treatment over long periods. Some healing traditions for possession-form presentations reformulate the person’s suffering into beneficent individual and communal practice via initiation into a spirit devotion cult—such as the Siri cult of South India—or education into the roles of oracle (diviner), exorcist, or, rarely, avatar (divine incarnation).

Prognosis is poorer in patients with comorbid organic mental disorders, psychotic disorders, and severe medical illnesses. Refractory substance abuse and eating disorders also suggest a poorer prognosis. Other factors that usually indicate a poorer outcome include significant antisocial personality features, current criminal activity, ongoing perpetration of abuse, and current victimization—including being trafficked, with refusal to leave abusive relationships or situations, whether out of fear or for other reasons. Repeated adult traumas with recurrent episodes of ASD may severely complicate the clinical course. Boundary violations and abuse by a prior treating therapist may prolong the treatment course because the treatment setting now feels unsafe and precipitates posttraumatic reactivity. Severe personality disorders are also generally associated with a poorer prognosis characterized by narcissistic overinvestment in multiplicity as a way of life, in having the disorder and/or the elaborate secondary structuring of self-states; engagement in psychotherapy primarily to seek gratification; and a refusal to take responsibility for behavior, behavior change, and/or symptom management. Number of self-states, however, has only a moderate effect on treatment course.

Treatment of Dissociative Identity Disorder

Current treatment approaches to non–possession-form dissociative identity disorder have evolved considerably. By contrast, the treatment of possession-form presentations remains largely unexplored except in anthropologically oriented observation studies. Available data suggest broad similarities between the treatment of both possession-form and non–possession-form presentations, including direct negotiation with the possessing self-states, allowing them to “give voice” to their concerns and assisting them to identify and redress their problems. Specialized indigenous practitioners and ritual therapies are generally available and widely utilized in many low-and-middle-income countries, but psychiatric treatment is typically avoided. Indigenous treatments include neutralization of the conflict or stress via the communal rituals involved in exorcism and/or other longer-term activities, such as integration into a healing community with a specialized function. The evolution of the self-states over time, and their eventual change to a more adaptive configuration, constitute one major basis of improvement that is common to both possession-form and non–possession-form presentations. Unlike non–possession-form cases, however, possession-form presentations typically do not resolve by unification into a subjectively singular self, given the supernatural attribution of most possessing agents.

The rest of the section on Treatment addresses the care of non–possession-form cases. An in-depth discussion of these current treatment approaches can be found in the *Guidelines for the Treatment of Dissociative Identity Disorder in Adults 3rd Revision* (2011) from the ISSTD (http://www.isst-d.org). In 2004, the ISSTD promulgated *Guidelines for the*

Evaluation and Treatment of Dissociative Symptoms in Children and Adolescents that can also be found on the ISSTD website. Both guidelines are in revision.

The current model of treatment of the individual with dissociative identity disorder follows a phasic model that is the current standard of care for complex posttraumatic disorders. The phases include (1) development of safety and symptom stabilization; (2) (optionally) focused, in-depth attention to traumatic material; and (3) integration or reintegration, in which the individual moves away from a life adaptation based on chronic traumatization and victimization. These phases are relatively heuristic, and aspects of each may be part of the others.

Treatment Outcome Studies in Complex PTSD and Dissociative PTSD. Convergent data from treatment outcome studies of patients with complex PTSD, the dissociative subtype of PTSD, and dissociative identity disorder have shown lack of response, or even clinical deterioration, if standard, unmodified Progressive Exposure/Cognitive Behavioral Treatment (PE/CBT) models for PTSD are used to treat these populations. Studies have shown better outcome in individuals with complex and dissociative PTSD from treatment paradigms that give attention to stabilization of dissociation, emotional dysregulation, and safety issues before, or even instead of, intensive processing of trauma memories. In some studies, highly modified exposure protocols—to “fractionate” the intensity of memories—were required for trauma memory processing in individuals with multiple early life traumas to prevent flooding by trauma memories.

Other studies report significantly higher-than-expected discontinuation rates and lower rates of abstinence after treatment for individuals with dissociative disorders who received standard substance abuse treatment. They also show lack of response to standard treatment for anxiety disorders and higher rates of non-retention in group therapy among sexual abuse survivors. In addition, elevated dissociative symptoms in patients with borderline personality disorder predict limited response to standard DBT.

A Norwegian study evaluated the response of dissociation and other symptoms among consecutive admissions to an inpatient program that provided stabilization treatment based on a phasic trauma treatment model for complex trauma survivors. Treatment primarily consisted of group and individual therapy, with adjunctive pharmacotherapy. Patients who met

DSM-IV-TR diagnostic criteria for dissociative identity disorder or dissociative disorder NOS (Example 1) only substantially improved when dissociative symptoms were directly addressed in treatment. None of their patients had previously been assessed or treated for a dissociative disorder and, at the time of this study, the program did not specifically target dissociative disorders in childhood trauma survivors.

Treatment Outcome Studies of Dissociative Identity Disorder

Single-case descriptions of successful treatments for dissociative identity disorder date back more than a century. Systematic outcome studies of non–possession-form cases, however, have only appeared recently. Reviews of clinical case series generally agree that patients with dissociative identity disorder improve with the phasic trauma treatment model. A meta-analysis of eight treatment outcome studies for dissociative identity disorder, including inpatient and outpatient treatments—with treatment provided by both nonexpert and expert clinicians—showed that the phasic model of dissociative identity disorder treatment was associated with improvement across a range of symptoms and comorbidities. These included fewer DSM-IV-TR diagnoses of comorbid axis I and II disorders; decreased dissociation, depression, anxiety, suicidality, and substance abuse symptoms; and reduced general distress. In studies from an inpatient specialty trauma disorder program, gains persisted at 2-year follow-up. Limitations of these studies include data from only a single clinician in some studies, small samples sizes, nonrandom inclusion of patients, high discontinuation rates, failure to account for nonspecific effects such as regression to the mean, and lack of control groups.

Cost Efficacy Studies of Dissociative Identity Disorder Treatment. As described above, individuals with dissociative identity disorder may spend years of unproductive treatment in the mental health system. Costs of their treatment have been hypothesized to be unnecessarily inflated due to lack of treatment response for incorrect diagnoses, frequent utilization of more restrictive levels of care (e.g., inpatient, residential, partial hospital, and intensive outpatient treatment), as well as costs for unnecessary medical procedures and work ups (e.g., repeated MRIs, EEG telemetry). Although costs were not estimated, a general population study in New York State found, compared to other diagnoses, the highest rate of disability in individuals with dissociative identity disorder. An international, naturalistic, longitudinal treatment outcome study of DSM-IV-TR dissociative identity disorder and dissociative disorder NOS (see next section) found significant reductions in inpatient utilization at 30 months, consistent with clinical improvement and reduction in treatment costs.

Data from 55,000 Massachusetts Medicaid patients from 1993 to 1996 showed that inpatient treatment for dissociative identity disorder cost over $3 million during the study period—33.5% of the total Medicaid inpatient costs—although patients with dissociative identity disorder made up only 2.6% of the sample. Cost of treatment for patients with dissociative identity disorder was higher than for patients with panic disorder, major depressive disorder, and bipolar disorder. In a large study of health care utilization, although less than 1% of the sample, wives of active-duty military personnel with a dissociative disorder had more outpatient visits than wives seeking treatment for any other psychiatric disorder. Individual case studies and small case series document substantial cost savings (e.g., ~50% to 67% reduction in costs) after correct diagnosis of dissociative identity disorder and initiation of phasic trauma treatment. Cost savings included both decreased utilization of more restrictive levels of care and reductions in medical costs.

A recent qualitative, systematic review of cost-efficacy studies for dissociative disorders demonstrated that these patients were costly to treat and a reduction in service utilization and associated costs with correct diagnosis and trauma/dissociation-focused treatment. However, studies varied in how costs were estimated and in other variables, making a meta-analysis impossible. The study authors suggested that better data are needed to rigorously assess treatment cost and cost savings in diagnosis and treatment of individuals with dissociative identity disorder.

Longitudinal Treatment Outcome Studies

Bethany Brand and colleagues developed the *Treatment of Patients with Dissociative Disorders Study* (TOPDD Study; http://www.towson.edu/topddstudy/index.asp) that included several treatment outcome study initiatives. The first was an international, prospective, longitudinal outcome study of 280 patients diagnosed with DSM-IV-TR dissociative identity disorder or Dissociative Disorder NOS Example 1 treated by 292 therapists from 19 countries. Treatment progress was assessed prospectively at study entry and at 6, 18, 30, and 72 months of treatment. Outcome measures were assessed using a five-stage model where stages 1, 3, and 5 corresponded to the three stages in the clinical Phasic Trauma Treatment model: establishing stability and safety (stage 1); processing of trauma memories (stage 2); and integration and reconnection within self and with others (stage 3).

At each time point, therapists completed a data sheet on the patient's stage of treatment (1 to 5), Global Assessment of Functioning (GAF), and incidence of self-harm, suicide attempts, and hospitalizations. Also at each time point, clinicians completed the Progress in Treatment Questionnaire (PITQ) that assesses management of impulse control; emotion regulation; capacity to use skills for symptom management; decreases in posttraumatic cognitive distortions and misperceptions; increased internal communication and cooperation between self-states; capacity to work on, and full awareness of, posttraumatic memories; integration of self-states; and increasing capacity to see oneself and others in a realistic, integrated way.

Patient measures were the DES, the Posttraumatic Check List-Civilian version (PCL-C), and the Symptom Checklist-90 (SCL-90-R). Patients also completed a behavioral checklist to assess the frequency over the last 30 days of self-injurious behavior, suicide attempts, "very impulsive" behaviors, alcohol and drug consumption, and high-risk behaviors that could have resulted in death. Also, patients reported the frequency of positive experiences and adaptive behaviors in the last 30 days including: volunteering or attending school, working for pay, using symptom management techniques, socializing, and feeling good "even if for a brief period." Patients also reported the days in inpatient and day treatment programs in the month prior to assessment.

Results from the 30-month follow-up are shown in Table 20–19.

The initial cross-sectional TOPDD results showed that patients in treatment stage 1 had higher levels of dissociation and PTSD, more hospitalizations, higher overall distress, and less adaptive functioning than patients in stage 5. As reported by both patients and therapists, at the 30-month follow-up there was decreased dissociation, PTSD, general distress, depression, suicide attempts, self-harm, dangerous behaviors, drug use, physical pain, and hospitalizations, as well as improved functioning as indicated by higher GAF scores. At the 30-month assessment, more patients were involved in positive activities in volunteer jobs and/or attending school and socializing and reported feeling well. Consistent with the phasic treatment model, more patients progressed from early stages of treatment to more advanced stages than those that moved in the opposite direction.

At 30-month follow-up, patients with the highest levels of dissociation, and with the most severe depression, also showed significant improvements in these symptoms. More patients showed “sudden improvement” than “sudden worsening” (i.e., 20% increase or decrease in symptoms) at one or more time point(s). Rates of revictimization showed a decreasing trend during the study. Therapists reported less frequent revictimization events and stressors among the patients who improved suddenly relative to those who worsened. Consistent with clinical experience, revictimization and/or severe external stressors were most likely to contribute to worsening in treatment. Only 1.1% of patients showed worsening over more than one data collection point, less than overall rates of worsening in treatment outcome studies of general psychiatric patients. As noted above, there were two deaths by suicide among study patients.

Younger patients, despite having elevated symptoms at baseline, were more likely to show improvement, indicating that early treatment of severely dissociative patients may lead to a more positive life trajectory. Patients reported a significant decrease in feeling like “different people” and hearing voices. A trend decrease in reports of dissociative amnesia supports the view that the phasic trauma treatment model helps to improve core dissociative process symptoms, not to increase them as critics of this model have suggested. At 72 months, 61 therapists from the original study provided data on their patients’ progress. These included continuing significant improvement in GAF, decline in inpatient utilization, decreased stressors, and fewer episodes of sexual revictimization. Patients maintained prior gains in decreased rates of nonsuicidal self-injury and suicidal behavior.

In summary, the TOPDD study documented that the phasic trauma treatment model improves a wide range of symptoms and adaptive functioning indices in patients with pathologic dissociative disorders and documented decreased utilization of more restrictive levels of care. Limitations of this study included the lack of a randomized control group, of fidelity compliance assessments for therapists, and of a specified psychotherapy protocol.

The TOPDD Network Study

The TOPDD Network Study is an ongoing study that assesses the impact of web-based interventions on a group of international patient–therapist dyads. Patients were clinically diagnosed with either DSM-IV-TR or DSM-5 dissociative identity disorder (76.5%) or dissociative disorder NOS/other specified dissociative disorder (20.6%/2.9%). (The study spanned the transition from DSM-IV-TR to DSM-5). Unlike most treatment outcome studies, patients were *not* excluded for current suicidality, nonsuicidal self-injury, recent hospitalization, comorbid disorders, substance abuse, social isolation, current stressors, and/or clinical instability. Evaluations at baseline and at 6, 12, 18 and 24 months included the PITQ (therapist and patient versions), DES, PCL-C, and the Difficulties in Emotional Regulation scale (DERS). Therapists reported on numbers of episodes of nonsuicidal self-injury, suicide attempts, and days of hospitalization for the 6 months prior to the study. Published data from the first 24 months of the study are described here (N = 111).

Patient–therapist dyads were provided 45 web-based psychoeducational videos, each 5 to 15 minutes in length, with additional written and behavioral practice exercises. These focused on establishing safety and reducing suicidal and high-risk behaviors as well as nonsuicidal self-injury. Other videos focused on symptom management (e.g., emotional dysregulation, uncontrolled PTSD and dissociative symptoms). The authors tracked how often videos were used, both in and outside of therapy. New videos could only be accessed 1 week after the previous one.

Table 20–19.

Findings at 30 Months from the Treatment of Patients with Dissociative Disorders Study (TOPDD Study)

Completion of Treatment, Drop out, and Termination Rates

Variable	Patients	Therapists	# Countries
Baseline	226	295	19 North America, Latin America, Europe, UK, Australia, New Zealand, China, Singapore, Israel, Slovakia, South Africa
N at 30-month follow-up	119	135	
% Drop out of study (of those who were still in treatment)	36%	48%	–
Successful treatment completers	3% of baseline sample (*n* = 6)	–	–
Terminate treatment for any reason	18% of baseline sample (*n* = 40)	–	–

Summary of significant treatment outcomes (Note: All variables changed in the hypothesized direction, so all significant linear changes indicate improvement over time. If the curvilinear term was significant, it means that the change in that variable flattened out over time; i.e., improvement slowed over time.)

Continuous Variables	Direction of Change	Significant Improvement over Time (Linear Change Significant)	Change Flattens over Time (Curvilinear Variable was Significant)
Patient Reported			

Dissociation	Decrease	X	
PTSD	Decrease	X	X
Distress overall	Decrease	X	X
Depression	Decrease	X	X
Therapist Reported			
GAF	Increase	X	
Adaptive Functioning (PITQ)	Increase	X	X

One Item Dichotomous Variables	**Direction of Change**	**Significant Improvement Over Time (Linear Change Significant)**	**Change Flattens over Time (Curvilinear Variable was Significant)**
Patient Reported: Symptoms			
Self-injurious behavior	Decrease	X	
Dangerous behavior	Decrease	X	
Getting high	Decrease	X	X
Impulsive behavior	Decrease	X	
Inpatient hospitalizations	Decrease		X (hospitalizations went up in the second follow-ups, but decreased by third and fourth follow-ups)
Patient Reported: Adaptive Functioning			
Volunteering/Attending School	Increase	X	
Socializing	Increase	X	
Feeling good	Increase	X	
Therapist Reported:			

Symptoms		
Self-injurious behavior	Decrease	X
Inpatient hospitalizations	Decrease	X

Not significant:

- Only single-item variables
- Patient reported: suicide attempts, drinking alcohol, using symptom management skills
- Therapist reported: day program attendance

Data from Brand BL, McNary SW, Classen CC, et al. A longitudinal naturalistic study of patients with dissociative disorders treated by community clinicians. *Psychological Trauma.* 2013;5(4):301–308. doi:10.1037/a0027654; and a paper presented as part of a symposium at the annual convention of the International Society for Traumatic Stress Studies, Montreal, Canada.

White female patients predominated; the sample was 10% male, 1.5% transgender; 5.6% Latinx, 2.8% Asian, 1.4% Black, and 7% other ethnoracial background. At baseline the patient sample was divided into a high-DES group (average = 59.89) and a low-DES group (average = 18.21). The former had significantly higher baseline PCL-C and DERS scores. At study completion, effect sizes showed large improvements in patients' adaptive functioning (PITQ) and emotion regulation (DERS); medium improvement in PTSD symptoms (PCL-C); and smaller improvement in DES scores. The high-dissociation group improved more than the low-dissociation group, with large effect sizes for improvement in adaptive capacities, emotion regulation, PTSD symptoms, and dissociation. The low-DES group demonstrated medium-to-large effect size improvements in postintervention adaptive capacities and emotion regulation but only small improvements in PTSD and dissociation. The effect sizes for change in therapist-rated adaptive function were of medium and low magnitude in the high and low-dissociation groups, respectively. The DES decrements were significantly related to time in the study. Over 24 months, the high-DES group showed a steeper slope of reduction in PCL-C and DERS scores. Over the 24 months, improvement in emotional regulation predicted improvement in dissociation scores.

Despite study limitations (e.g., clinical diagnoses instead of structured interview diagnoses, possible selection effects of patient–therapist dyads), the inclusion of a high proportion of patients with nonsuicidal self-injury, suicidality, and hospitalizations is a particular strength. The frequency of nonsuicidal self-injury substantially decreased over the 2 years of the study. Therapist reports of patient self-injury behavior in the 6 months prior to the study were not normally distributed in the sample, with 60.4% engaging in these behaviors (average = 13.75 episodes for those engaging in nonsuicidal self-injury; overall sample range 0 to 150 episodes). At study conclusion, for the group with nonsuicidal self-injury, about 70% showed steady reduction in this symptom (average = 1.74 episodes), about 25% increased, and the remainder stayed the same or increased and then decreased. The three patients with the highest number of episodes of nonsuicidal self-injury (100, 125, 150 events in the prior 6 months) showed reduction to 0, 10, 10 times, respectively, in the last 6 months of the study. There was a trend for a decreasing number of suicide attempts and hospitalization days.

In a subsequent analysis, greater difficulty with emotional awareness, clarity, and acceptance predicted nonsuicidal self-injury, as did impulsivity and struggle with goal-directed behavior. This finding confirms the expert consensus that it is essential to educate individuals with dissociative identity disorder that dangerousness to self is a maladaptive, trauma-based attempt at self-regulation. Interventions focused on the patient "stopping bad behavior" are doomed to failure. The authors' clinical experience is that the latter stance—basically shaming the individual for something experienced as uncontrollable—increases dangerousness to self, as shame is one of the main dysregulated emotions that drives these behaviors. Individuals with dissociative identity disorder can markedly decrease dangerousness to self when they can make sense of these behaviors as attempts at self-regulation. Reduction in these behaviors accompanies improvements in emotional regulation skills, challenges to trauma-based cognitive distortions about the necessity of self-destruction, and the increasing capacity to delay self-destructive actions and replace them with recovery-oriented symptom management and emotional regulation skills.

Expert Consensus Guidelines for Phasic Treatment of Dissociative Identity Disorder

Central to development of the TOP-DD studies, 36 international dissociative identity disorder experts were surveyed to identify treatment phases and evidence-based interventions. The experts rated 28 interventions on a 0 (never) to 4 (very often) scale across the stages of the phasic trauma treatment construct. They agreed upon a core set of beneficial interventions. The most frequently recommended interventions for each treatment stage are shown in Table 20–20.

Even in the stage where work on trauma memories is most predominant, the expert treatment recommendations still focus on safety, stabilization, the therapeutic alliance, cognitive work, and symptom management. In the final phase, the therapeutic work is directed toward “reintegration,” living well in the present, and relegating traumatic memories to the status of “bad memories”—as opposed to flashbacks, behavioral re-experiencing, and/or intense posttraumatic reactivity to current situations. These stages are heuristic since trauma memory may need to be addressed—if only in a more cognitive and distanced manner—in stage 1 and worked through again from a more integrated perspective in stage 3. Safety may be an issue at all stages of trauma treatment. Finally, the entirety of trauma treatment is directed toward helping patients adapt better to their current circumstances.

Table 20–20.
Expert Recommendations for Treatment of Dissociative Identity Disorder—Stages 1–5

Stage	1	2	3	4	5
Establish safety	✓	✓	✓		
Establish/repair alliance	✓	✓	✓	✓	✓
Teaching/practicing grounding	✓	✓	✓		
Educate about disorders and treatment options	✓	✓			
Diagnose/rule out comorbid psychiatric illnesses	✓				
Teaching/practicing self-care	✓	✓		✓	✓
Developing healthy relationships	✓			✓	
Affect tolerance and impulse control	✓	✓	✓	✓	
Stabilizing from current day stressors/crises	✓	✓	✓	✓	✓
Teaching/practicing containment	✓	✓	✓		
Developing healthy relationships		✓			✓
Cooperation with self-states		✓	✓	✓	
CBT focus on cognition distortions			✓		
Identify/work with self-states			✓		
Awareness of body sensation			✓	✓	✓
Discussing therapeutic relationship				✓	✓
Processing patient's reactions to therapy				✓	✓
Ego strengthening activities				✓	✓
Awareness of emotion					✓
Stabilizing following intrusions from reported perpetrators					✓

Note: Phasic Trauma Treatment Model has three Stages, corresponding to TOPDD stages 1, 3, 5—see text for discussion.

Data from Brand BL, Myrick AC, Loewenstein RJ, Classen CC, et al. A survey of practices and recommended treatment interventions among expert therapists treating patients with dissociative identity disorder and dissociative disorder not otherwise specified. *Psychol Res Pract Policy.* 2012;4(5):490–500.

The experts recommended interventions around safety most frequently in all but the last stage of treatment. Safety is a consistent focus because each stage can bring about emotionally dysregulated crises that patients may feel compelled to manage via unsafe behaviors. Developing and repairing the therapeutic alliance was recommended as a top intervention in every stage of treatment, indicating the centrality of the therapeutic alliance in successful treatment of dissociative identity disorder. Recent research suggests that the therapeutic alliance is important in predicting improvement across a range of symptoms as it is in most successful psychotherapies. Individuals with dissociative identity disorder have suffered betrayal and maltreatment in the context of early dependent relationships (see sections on Betrayal Trauma and Type D attachment). On the other hand, an aspect of their resiliency is belief in the possibility of a positive attachment, despite the cruel betrayals. This may underlie their capacity to develop the therapeutic alliance that, along with ability to use insight, may be crucial in their recovery. As individuals with dissociative identity disorder suffer maltreatment almost universally in the context of attachment, experts recommended ongoing attention to transference issues, particularly traumatic transference reactions, that are endemic to this population.

Other core interventions recommended across all stages of treatment include differential diagnosis to identify symptoms better conceptualized as aspects of the complex posttraumatic dissociative processes, as opposed to genuine comorbid disorders; psychoeducation about the role of medications in dissociative identity disorder treatment; strategies for improved awareness of emotions and development of emotional regulation skills and impulse control; management of daily functioning and current relationships; and help coping better with stressful life situations, losses, and additional trauma. The treatment recommendations in this chapter are based on these expert recommendations, the ISSTD Treatment Guidelines, and the interventions documented in the TOPDD studies that have demonstrated efficacy in these patients' improvement.

Stage 1 Phasic Trauma Treatment: Stabilization and Safety

Successful treatment of dissociative identity disorder *requires* development of safety from all aspects of dangerousness to self/others and stabilization of dyscontrolled dissociative, PTSD, and other comorbid symptoms. Before establishing safety, the patient actually and/or symbolically lives in a world of ongoing trauma, whether perpetrated by others and/or by the patient on themselves due to ongoing dangerousness to self/others. The clinician must logically prioritize basic life and health over other interventions. The clinician may need to resort to many different levels and types of interventions to protect the patient, including hospitalization, specialized substance use or eating disorders programs, police assistance and shelter for victims of intimate partner violence, social service interventions to protect minor children, and assistance with housing and access to medical care. To achieve safety, stability, and overall treatment progress, effective stabilization in patients with dissociative identity disorder requires psychotherapeutic work with self-states/self-state systems.

Direct Therapeutic Engagement with Self-States

Clinicians new to dissociative identity disorder are frequently uncomfortable or perplexed by the need to work directly with self-states and how to do this without producing a chaotic regression or reinforcing the patient's sense of separateness among self-states.

The model of dissociative identity disorder as a childhood failure of unification of self/identity across states, relationships, and contexts frames the psychotherapeutic approach to working with self-states. No self-state is any more or less "real" or better or worse than another. All are aspects of a single human being and have adaptive psychological importance that needs to be heard and respected. The clinician should not engage in interactions that reinforce notions of self-states as separate people. These can include "playing favorites," supporting the notion that some self-states are fundamentally "good" and others "bad," refusing to communicate with certain self-states, and giving ongoing, separate appointments to self-states as one would with individual patients. However, this does not mean that the clinician condones self-destructive, dangerous, suicidal, violent, or other high-risk behaviors. To help stop these behaviors and develop the person's repertoire of recovery skills; it is critical to understand the self-states that drive these self-regulatory behaviors and the trauma-based cognitive distortions that underlie them.

All self-states must be held accountable and responsible for the behavior of any other state, even if experienced with amnesia or lack of subjective ownership. Human beings engage in behavior, not self-states. One can attempt to understand the mental state of the person who does the behavior, but *explanations are not excuses*. Expert consensus and clinical experience postulates that to do otherwise is an invitation for the patient to regress. Because patients with dissociative identity disorder have both psychological strengths and liabilities, over time they improve when held to a standard that underscores their strengths. As described above, self-states can be understood as developmentally concretized forms of organizing, symbolizing, and mobilizing mental contents around psychological survival from repeated, unpredictable, attachment-based maltreatment. The fullest engagement of all mental symbols and sources of knowledge, skills, and adaptations is likely to produce the best outcome in any psychotherapy. This approach also forms the recommended basis for forensic standards around competence and not-guilty-by-reason-of-insanity (NGRI) judgments. It uses the psychological-structure findings of preservation of logical thinking and reality testing and the notion that patients with dissociative identity disorder can inhibit and conform behavior in many contexts.

Concerns about "iatrogenesis" are often raised in these circumstances. Understanding the self-states fundamentally as forms of intrapsychic symbolization reduces this concern to a more routine psychotherapeutic task of understanding the meaning of mental contents, rather than a suppression of some mental states as less authentic than others. At the same time, the clinician must also be able to appreciate the patient's subjective reality in which the self-states are experienced as distinct selves or even as separate persons. The therapist must be able to tolerate engaging with the self-states *as if* they were separate entities, while simultaneously maintaining the understanding that the self-states represent the structuring of psychological processes in personified form within a mind.

The continuity, stability, and respectful impartiality of the therapist toward the different self-states provides an important therapeutic experience that helps the patient experience, examine, and integrate these dissociated aspects of self. The processing of negative life events, often from the multiple perspectives of different self-states, is an important part of the narrative reorganization of the patient's divided self/identity into a more coherent whole. The therapist usually is actively involved in negotiating between and among the various self-states to achieve therapeutic goals. The therapist works to assist the self-states/system with improved communication, coordination, cooperation, and internal empathy. This stance helps with development of emotional regulatory skills and adaptive functioning. As shown by the TOPDD studies, rather than reinforcing dissociation, working in this way gradually diminishes dissociative barriers among self-states. This is associated with decreased dissociative amnesia, dissociative hallucinosis, passive-influence symptoms, and dysregulated state-switching/shifting, as well as better ability to manage PTSD symptoms and maintain safety from dangerousness to self/others.

Practical Management of Safety Issues

As described in the TOPDD network study data, patients with dissociative identity disorder use unsafe and high-risk behaviors as attempts to regulate, manage, and modulate severe dissociative, PTSD, emotional, somatic, and cognitive dysregulation. Work on safety generally includes discussion of the antecedents to, and functions of, these behaviors as well as safety agreements and crisis management plans. There is usually a logic behind the patient's self-destructive behavior. For example, there is a highly prevalent cognitive distortion in this population that, "I'm going to get hurt no matter what I do, so control of the timing and intensity of harm is all the control I get." These beliefs can be reframed as an attempt to achieve a sense of control to survive the helplessness and unpredictability of repeated maltreatment throughout development. The therapeutic goal then becomes to change this "survival skill" to a "recovery skill": developing adult, non–trauma-based strategies for appraising one's current life situation and level of external danger, gaining a repertoire of skills to keep oneself safe in the present, and creating overall safety and self-protection in one's current life.

Usually, in this patient group, self-destructive behavior has an addictive component and many of these patients have current or past problems with substance misuse. Using addiction metaphors can be helpful: "sobriety" from dangerousness to self involves a commitment to safety, the realization that symptoms likely will *worsen* once the behavior stops, and intensive work to find alternatives to help manage intolerable, often unnamable, affective states such as extreme shame, humiliation, rage, horror, helplessness, and disgust/revulsion. At times, self-injury is experienced as an attempt to "feel something" when in a numb state and/or to "put the pain on the outside" where it seems more real, or where (the patient believes) others may take it more seriously.

Safety Agreements. Safety agreements may be a useful clinical tool to assist the individual with dissociative identity disorder manage safety issues and to support the person's ambivalence about being unsafe. Safety agreements are no substitute for the clinician's judgment about the patient's safety from dangerousness to self or others. The clinician must continually assess the patient's safety and, based on clinical judgment, act to protect the patient or others, including initiating involuntary hospitalization, despite the patient's protestations that they will maintain the agreement.

It is useful to think of safety agreements as *delaying* agreements, where the patient has access to a hierarchy of interventions to use instead of self-destruction, up to and including going to the emergency department for stabilization. Individuals in recover from addiction/substance abuse are taught to think in "one-day-at-a-time" approaches to sobriety. In trauma recovery, patients with dissociative identity disorder may be unable to commit to safety "forever," but may be able to commit to a safety plan for a more manageable period (e.g., from therapy session to therapy session). Some clinicians prefer that the patient write down and sign this type of agreement. The authors' experience is that the crucial factor is the patient's honesty about safety and the ability to maintain themselves as a safe outpatient, following the safety plan. Many signed safety agreements are not honest and many verbal commitments are. Some patients feel that writing and signing an agreement reinforces the seriousness of and concretizes the commitment to it. It is not necessary for "all" self-states to sign the safety agreement. Instead, the clinician can ask all states, "known or unknown," if there is any objection to the agreement and to indicate so by a physical signal, such as a head nod or verbally. If there is an objection, the clinician can ask for some sort of communication in the patient's mind about the concern. Often, the issues that emerge are quite helpful to the patient in the process of working honestly on safety. It may help the patient to realize that the clinician must trust the *patient's* honesty regarding safety agreements and that the patient is not the only one in the therapy who has concerns about trust. If a signed agreement is created, the patient may sign several names and add that "and all [other] known or unknown" self-states agree as well. Occasionally, the patient with dissociative identity disorder will have an idiosyncratic posttraumatic reaction to the words, "safety" or "agreement." In this case, the patient may be asked to provide a reasonable alternative, such as the "no-immediate-action pact."

EDUCATION AND INFORMED CONSENT. It is important to educate the individual with dissociative identity disorder about the disorder, its actual and apparent comorbidities, and the course of treatment. Educative interventions help decrease anxiety about symptoms that are often frightening and overwhelming, build a therapeutic alliance, and provide information that is the basis for a meaningful consent for treatment. Furthermore, it is necessary to educate the patient about the contentious and divisive debates that surround the diagnosis and treatment of dissociative identity disorder in contemporary psychiatry and psychology.

Other major issues in informed consent include advising the patient of potential risks and benefits of the treatment and the current evidence base for the treatment model. In addition, patients should be informed about approaches that differ from the phasic trauma treatment model, such as the iatrogenic/sociocognitive/FM that advocates ignoring the self-states and the history of trauma, and the evidence base and potential risks and benefits of these alternatives.

The patient should also be informed about current controversies concerning traumatic memory and its retrieval and explication in therapy. Additional informed consent should be obtained for formal hypnosis and pharmacotherapy. The patient should be counseled that symptomatic worsening may occur temporarily during treatment, particularly during phases in which memory material is worked with in depth. Simultaneously, however, subjective amelioration of other symptoms may occur, with the patient achieving a better sense of control, mastery, and self-coherence as they successfully work through painful material.

Similarly, like many obsessional individuals, patients with dissociative identity disorder often respond positively to learning that their dissociative and PTSD symptoms have a name and an organizing framework. There is often relief at a diagnosis that fits the patient’s subjective life experience, especially if there has been a long prior history of unsuccessful psychiatric treatment for other disorders. At the same time, patients may respond with distress at the dissociative identity disorder diagnosis due to media and popular stereotypes, because it brings into sharper relief the meaning of painful symptoms, due to shame reactions such as feeling like a freak, and/or because it brings more focus to the distressing history of early trauma and problematic relationships with family members.

Individuals with dissociative identity disorder often need to be counseled that the purpose of treatment is not to suppress or destroy self-states but rather to allow therapeutic engagement with them as vital aspects of the mind. Self-states may express strong wishes to “get rid” of others and are simultaneously terrified that the therapist will attempt to do so. Most experienced clinicians educate patients that treatment is geared toward optimal function and adaptation and that there are usually too many variables at the beginning of treatment to predict whether it will necessarily ultimately result in unification of all self-divisions, a concept that the patient usually cannot really comprehend at the beginning of treatment. At the same time, most experienced clinicians see unification of all self-states as usually providing the best long-term outcome in dissociative identity disorder.

The education and informed consent process should be ongoing throughout treatment. Many of these issues need to be revisited as treatment proceeds and the patient changes in therapy. As with the safety agreement, if a signed informed consent is used, it is not necessary for “all” self-states to sign the agreement and self-states that are experienced as newly emerging later in treatment may report amnesia for prior psychoeducation. An overly rigid concern about how many states sign the agreement may reinforce subjective separateness by treating the self-states more as separate people than subjectively structured personified mental processes.

Boundaries and Treatment Frame

Most patients with dissociative identity disorder report that their traumatic life experiences occur over long periods of time in the context of close personal relationships with those who have hurt them. These patients often report boundary violations in subsequent relationships with teachers, therapists, medical professionals, and social service workers, among others. For these patients, the role description of significant others does not necessarily predict their behavior, and relationships are often seen as “up for grabs.” The patient expects and fears that the clinician will violate boundaries. There may be a subtle (or not so subtle) pull in the transference for the therapist to change the frame of treatment. Patients may experience this as trying to make the inevitable happen, so that they do not have to endure the agony of waiting for things to go wrong. Accordingly, clinicians working with these patients must pay careful attention to issues of the treatment frame and appropriate and consistent boundaries. These include firm limits on the length of therapy sessions; acceptable behavior during sessions (e.g., the therapy space is “not a hurting place,” it is one of safety for therapist and patient alike to talk about important issues); the extent of extra-session contacts; and payment of fees; among many others. Dissociative identity disorder individuals may express relief at the predictability of the treatment frame and boundaries, even as they may have other traumatic-transference based reactions to them. Calm, educative, non-blaming interpretation of the patient’s fears over the perceived inevitability of boundary violations is preferable to enacting them.

The ISSTD Guidelines, the expert survey, and the TOPDD study all converge to discourage boundary-crossing activities, even if the patient insists that only these interventions will help, and/or particularly help specific self-states, such as child states that are portrayed as needing concrete demonstration of love from someone. Particularly, the therapist may be asked to hold or hug the patient, hold the patient’s hand, or other types of physical contact. Rather than acting on this, the clinician should inquire if some self-states are distressed at the idea of this physical contact, as “nice” touching may have evolved into traumatic abuse in the patient’s reported history. Then, the issue is a therapeutic discussion of the meaning of the request and the patient’s profound conflict about it.

It is not recommended that the clinician accept more than a token gift such as a card or a small piece of artwork; give the patient gifts; routinely contact the patient while on vacation; engage in out-of-office contact with the patient, such as going for walks and/or going out for coffee; and screen the patient’s voice mailbox for calls from reported perpetrators, among many other potentially boundary-crossing behaviors. Clinicians who feel overwhelmed by the treatment of these patients are more likely to accede to the patient’s insistence on these sorts of actions. It is important to recognize that the patient’s reported history of abuse often involved being special to the abuser in some way. Thus, these boundary transgressions recreate this double-edged situation for the patient. Once again, education of the patient about boundaries is the most helpful clinical strategy.

Many of the clinical and conceptual issues that help make treatment workable with these patients are not taught in standard training programs for mental health professionals. The ISSTD has a series of courses on diagnosis and treatment of dissociative disorders that can be accessed at www.isst-d.org. Consultation with experienced clinicians should be considered by clinicians who are new to working with this patient population, and/or if stalemates or seemingly unsolvable predicaments develop in treatment.

Interventions for Safety and Stability: Containment, Grounding, and Imagery Techniques

The expert consensus recommends that use and teaching of grounding techniques begin in stage 1 and continue through subsequent stages.

Grounding may include a variety of interventions that begin in therapy but are practiced independently by the patient outside of sessions. Grounding may include having the patient change positions, open their eyes, use their five senses to become more aware of current circumstances: touch an object, such as the arms of the chair, keep their feet on the floor, and look around the room. If the patient enters a severe dissociative flashback, the therapist may literally have to orient the patient to the date, place, and circumstances, including the name and profession of the clinician. Grounding can also assist the patient with the confusion, fear, and dysphoria that occurs with overlap and interference among self-states as well as help the patient maintain state stability to prevent dysfunctional switching.

In all but the last stage of treatment, expert consensus suggests that the patient develop containment skills for traumatic material. These interventions help the patient use dissociative defenses more adaptively. Patients may oscillate from feeling “all shut down” to overwhelming intrusion of trauma material. The patient can be taught that containment is different than shutting down. It involves a voluntary psychological sequestration of trauma memories or overwhelmed self-states until these can be worked with in a gradual way later.

With the onset of treatment, most expert therapists rapidly introduce grounding and containment skills for symptom management. This necessitates active structuring of sessions to work on containment of severe symptoms and skills-building. Containment techniques include teaching self-hypnosis (only if the clinician is credentialed in the use of hypnosis) and imagery to control the intrusiveness of traumatic material (see discussion on hypnosis). These techniques use patients’ high dissociative and hypnotic capacity to assist their dissociative defenses to become more flexible and adaptive. This may involve imagery techniques to attenuate the intensity of PTSD, somatoform, and dissociative symptoms; successful stabilization of self-states that embody or are experienced as creating symptoms; and encouragement of collaboration, communication, and empathy among self-states. Experienced clinicians work to attenuate the impact of PTSD intrusions rather than opening this material prematurely. Imagery interventions also may include ego-strengthening (i.e., interventions to promote better overall functioning); calming imagery such as safe-place imagery; imagery for sequestration and/or attenuation of difficult material such as flashbacks; reaffirming statements; relaxation training; and cognitive-behavioral work to change trauma-based cognitions.

Given the consistency of strategies recommended by experts, the authors of the expert survey concluded that these interventions make up the core treatment processes and structure for treating dissociative identity disorder. The consistency of the recommendations in the expert survey and the ISSTD Treatment Guidelines, as well as results of the TOPDD studies, indicates that there is now an evidence-based standard of care for the treatment of dissociative identity disorder.

Responsibility for Recovery. The clinician should be clear that the ultimate responsibility for symptom and behavioral management outside of sessions lies with the patient, who is viewed as an active partner in the treatment. In general, if not immediately, patients with dissociative identity disorder respond well to taking more active responsibility for safety or behavioral control between sessions. The therapist may prescribe homework for the patient to work on symptom management and internal communication between sessions. These strategies are usually essential in countering the patient's tendency toward regression and demoralization. Patients often use journal writing as a helpful adjunctive treatment strategy that allows more controlled expression and communication among self-states, fulfillment of homework assignments, and a more attenuated and distanced way of handling overwhelming subjective experiences.

Trauma-Focused Cognitive-Behavioral Interventions. Individuals with dissociative identity disorder may experience many cognitive errors and distortions based on traumatic life experiences and the ability of dissociation to interfere with reality testing. These include basic misattribution of causality: "I caused my father to abuse me"; de-catastrophizing (minimizing violent physical discipline as appropriate), and catastrophizing (a minor disappointment becomes a matter of survival); trauma-based doublethink (e.g., "sex is love," "anger is violence," "self-assertion is mean," "hurting myself makes me safe," and many others).

Other cognitive distortions include the insistence on delusional separateness and delusional disorientation to time and place (i.e., self-states insist that they do not inhabit the same body and are unaffected by harm that comes to it; and that self-states literally exist in a different time or place, respectively); that the patient is helpless to control themselves and requires the clinician to manage all difficulties; that the clinician is completely untrustworthy and must not be allowed any access to the patient's mind; since love and sex are the same, and because trauma and abuse are inevitable, it is best to invite them or enact them oneself, so that at least their timing, intensity, and shamefulness can be better controlled.

Many cognitive distortions are only slowly responsive to cognitive therapy techniques, and successful cognitive interventions may lead to additional dysphoria and sometimes explosive, intrusive PTSD symptoms. Behind these cognitive distortions frequently lie exceptionally painful realizations and recollections. For example, if the patient gives up the idea of blameworthiness for childhood maltreatment, the reality must be faced that beloved relatives have done the patient grievous harm—have treated the patient as a hated object, not a person. This may lead to flood of flashbacks embodying this reality.

Transference and Countertransference. Individuals with dissociative identity disorder manifest a complex multilayered transference, as "whole human beings," system(s) of self-states, and individual self-states. Transference is dominated by themes of trauma and abuse, with the therapist seen as (in alternation) perpetrator, rescuer, victim (with the patient as a—typically disguised—perpetrator), or unconcerned bystander/witness. Individuals with dissociative identity disorder commonly suffer from the deep-seated, betrayal-based assumption that there are only three outcomes of attachment: exploitation, abandonment (if the person is not abusive), and/or both. Interactions may subtly recreate themes from the patient's history of attachment-based maltreatment. Individuals with dissociative identity disorder in a *flashback transference* may literally envision the clinician as the embodiment of an abusive figure from the past, requiring active interventions to help the patient separate the past from the present. Interactions in therapy may unconsciously reenact themes from the patient's experiences with malevolent, often sadistic, attachment figures.

Countertransference responses typically involve similar posttraumatic and dissociative responses with, at times, almost uncanny symbolic repetition of past interactions the patient describes with people from childhood. Classically, the inexperienced clinician may begin as an over-involved rescuer for the patient. When this inevitably fails, the clinician may feel exasperated, overwhelmed, de-skilled, and ashamed. The therapist may then develop a detached, hostile skepticism, and/or (often abrupt) discontinuation of treatment. In the best case, the clinician seeks expert consultation, training, and education about diagnosis and treatment of dissociative identity disorder. The therapist learns that these dynamics—reminiscent of intrusion and numbing in PTSD—are represented by self-states in conflict about these issues, and are best worked with by exploration of these conflicts.

Clinicians may experience countertransferential autohypnotic and dissociative phenomena during dissociative identity disorder treatment that need to be recognized and managed. Education about hypnosis may help the clinician identify and manage these experiences. Secondary PTSD and burnout have been reported for therapists and treatment teams working with patients with dissociative identity disorder unless adequate attention is paid to limit-setting, boundaries, and assistance to clinicians adversely experiencing reports of extreme abusive trauma or being overwhelmed by work with these patients.

Affect Regulation/Shame. As described above, innate affect theory is a helpful framework for viewing the experience of the individual with dissociative identity disorder as well as the structure of the inner world of the self-states. The four components of a "compass of shame" include "attack self," "attack other," "avoidance," and "withdrawal." Shame affects and "shame scripts" may dominate the psychic life of patients with dissociative identity disorder. As noted above, Richard Kluft has developed a typology of dissociative identity disorder self-states that links their characteristics to one or more of the poles of the shame compass. Kluft suggests that avoidance of difficult material in dissociative identity disorder is frequently related to hidden shame scripts. In many instances, the dissociative patient is not amnestic, but profoundly inhibited by shame avoidance from discussing important material. Expert consensus agrees that systematic attention to shame allows for improved negotiation of many difficulties and potential impasses in the treatment of dissociative identity disorder.

Stage 2 Phasic Trauma Treatment: Work on Traumatic Memories. For patients who stabilize and form a reasonable working alliance in treatment, longer-term treatment goals involve the detailed, affectively intense, psychotherapeutic processing of life experiences, especially traumatic experiences, and the transformation of the meaning of these experiences for the individual. Authorities emphasize that, in most cases, intensive, detailed psychotherapeutic work with traumatic memories should only be initiated after the patient has demonstrated the ability to consistently maintain safety, to use symptom management skills independently, that the self-state system can work together in a reasonably cooperative way, and after a solid therapeutic relationship has been established.

The patient should be able to give informed consent and have a realistic understanding of the potential risks and benefits of intensive focus on traumatic material. Potential risks may include temporary, acute worsening of PTSD; affective, somatoform, and self-destructive symptoms; and short-term interference with daily activities. Long-term benefits may include significant amelioration of dissociative and PTSD symptoms, decreases in subjective self-division, fusion of self-states, and freeing of psychological energy for daily life. The patient must be able to understand that the goal is integration of dissociated thoughts, feelings, recollections, and perceptions, not the exhumation or expulsion of memories per se. As with other traumatized patients, particularly during this treatment phase, grieving becomes a major issue. Grieving may focus on the necessity of viewing early relationships differently but also on the losses engendered by the individual's life trajectory. These include losses of a normal childhood and adolescent development; loss of potential educational, work, and family life precluded by the posttraumatic/dissociative life trajectory; loss of relationships with important others; and a sense of "wasted" years seeking and undergoing the rigors of this type of treatment.

Furthermore, the patient should not be in the midst of an acute life crisis or major life change, comorbid medical and psychiatric disorders should be stabilized, they must have the ego-strength and psychosocial resources to withstand the rigors of the process, and there must be adequate resources, such as support by significant others and, potentially, support for additional sessions. Many of these are listed in Table 20–21 (see also Table 20–22. Contraindications to Stage 2 Work). The clinician must also have the skills to work in this phase of treatment, and have the time, availability and resources for more intensive work and need for possible emergency contacts (e.g., not about to leave on a long vacation; not in a particularly stressful life-space, etc.).

Table 20–21.
Therapeutic Considerations for Stage 2 Phasic Trauma Treatment in Dissociative Identity Disorder (See also Table 20–22 on Contraindications for Stage 2 Memory Integration Work)

- Patient has achieved the goals of earlier stages, including
  - Basic acceptance of diagnoses, trauma issues
  - Ability to maintain safety despite stress
  - Basic understanding of the importance of safety
  - Reasonable level of empathy, internal communication, coordination, collaboration/noninterference among self-states
  - Demonstrated ability to independently use symptom management/containment strategies
  - Mapping of self-state system gives basic view of major self-state groups/types
  - Basic stabilization of paradoxically protective angry, “persecutory” self-states
  - Adequate therapeutic alliance
  - Severe dysregulated PTSD controlled
  - Uncontrolled, dysfunctional switching absent, infrequent, or reasonably easily contained
  - Rational motivation for engaging in the process
  - Reasonably stabilized life circumstances
  - Presence of interpersonal supports (and how to manage if limited or nonexistent)
  - Stable comorbidities, both psychiatric and medical
  - Adequate ego strength to undergo this process without major decompensation
  - The therapist is prepared in both availability, skills, and ego strength to work with/tolerate the material

Data from Kluft RP, Loewenstein RJ. Dissociative Disorders and Depersonalization. In: Gabbard GO, ed. *Gabbard’s Treatment of Psychiatric Disorders*. 4th ed. American Psychiatric Press; 2007.

Table 20–22.
Contraindications to the Memory—Integration Phase of Amnesia Treatment (see also Table 20–21)

Has not achieved stable safety from nonsuicidal self-injury, suicidality, high-risk behaviors; eating disorders; lack of adequate food, clothing, shelter, medical care, etc.
Current substance use disorders or dependence
Has not achieved symptom stabilization
Severe, uncontrolled, intrusive posttraumatic stress disorder symptoms
Extreme PTSD avoidance symptoms
Severe, uncontrolled dissociative symptoms (dysfunctional switching, intractable internal conflict, fugue episodes, etc.)
Dysregulated mood symptoms (e.g., mania, bipolar depression)
Self-state systems do not engage in adequate internal communication, cooperation, coordination; severe lack of internal empathy
Inadequate therapeutic alliance
Current abusive relationship
Acute life crisis or times of life transition (divorce, job change, completing Ph.D., etc.)
Other potentially problematic life circumstances (e.g., has small children at home, limited child care, etc.)
Severe physical illness or infirmity
Severe personality disorder symptoms (e.g., antisocial features)
Comorbid psychosis, severe regression
Unrealistic idea of purpose of Stage 2 work (e.g., “Then I’ll know the truth”; “So I can sue my parents”; “I’ll get rid of all the memories,” etc.)
Current involvement in litigation
Impending absence of therapist

Adapted with permission from the *Treatment of Psychiatric Disorders*. 3rd ed. Vol 2. American Psychiatric Association; 2001.

Experienced clinicians attempt to carefully structure affectively intense sessions focused on traumatic material, with attention being given to affect modulation, re-stabilization of the patient before concluding the session, and reasonable availability to assist the patient supportively between sessions. In addition, many sessions may be needed to explicate fully the cognitive and emotional meaning of traumatic events so they become integrated into a nondissociative, non-flashback autobiographical memory narrative.

TRAUMATIC MEMORIES. Media attention and legal cases have led to concerns over the authenticity of traumatic recollections of patients with dissociative identity disorder. Recent rigorous case series have supported the experience of clinicians, who have found corollary information to corroborate these patients' recollections or who have even unearthed data about traumas that the patient does not recall. Based on corollary information, however, some individuals with dissociative identity disorder can misinterpret and misrepresent contemporary information or confabulate aspects of their history. In addition, dissociative patients are hardly immune to ordinary human emotions, such as greed, envy, wishes for revenge, wishes to evade consequences for misbehavior, and desires to placate significant others. All of these may complicate the patient's veracity, especially in situations of potential financial gain, media attention, legal cases, and possible loss of family contacts, among others. Collateral informants, such as family members, may be subject to the same sources of unreliability as the index patient and their input should be weighed accordingly. The concern about "false" trauma memories does not extend to concern about "false" *non-trauma* memories, as perpetrators typically threaten victims to remain silent and/or, after an episode of rape, may tell them that "nothing happened… You won't remember what happened," and so on.

Typically, individuals with dissociative identity disorder oscillate from regarding their recollections as all true or all false. Self-states may take opposing positions. In the absence of reliable collateral information, the clinician is best served by maintaining a neutral stance toward the patient's recollections of their life by neither "believing" nor "disbelieving" a given memory report, other than questioning recollection of clearly impossible events (e.g., alien abduction). It is generally most helpful for the clinician to identify for the patient their internal conflicts over the veracity of recollections and to invite an open airing of all points of view. Repeated discussion of the complex factors that may impact autobiographical recall typically is part of an ongoing informed consent process as treatment progresses.

The therapist should respond respectfully and thoughtfully to all clinical material brought by the patient. Patients can be helped not to come to premature closure about their views of past events. Some individuals with dissociative identity disorder come to understand life events differently as they become less dissociative over the course of treatment. This is another reason that clinicians must avoid validating memories for the patient or dismissing them out of hand. Therapy is most successful for the patient if the clinician maintains the role of therapist to help the patient make sense of their autobiography, not personal advocate, detective, or derisive skeptic.

Stage 3 Phasic Trauma Treatment: Fusion, Integration, Resolution, and Recovery. Over the course of treatment, significant unification of dissociated mental processes may be observed. Self-states lose distinctness and decrease compartmentalization of thoughts, memories, and affects. The patient develops a more unified sense of self. Transference is modified consistent with these changes. Amnesia and switching become less apparent. Fusion of states results in psychological merging of two or more entities at a point in time, with a subjective experience of loss of all separateness. The term *integration* is sometimes used synonymously with *fusion* but is more generally defined as the process of undoing all forms of dissociative division during treatment. Some patients proceed to what appears to be a complete fusion of all states, with a shift in self-representation from that of an individual with dissociative identity disorder to one with a consistent and continuous sense of self across all behavioral states. Many patients never attain a complete fusion of their personalities but achieve a therapeutic *resolution:* improved communication, collaboration, and cooperation among self-states leading to relative stability and adequate function.

As this integrative process occurs, PTSD symptoms usually improve significantly. Patients often experience a freeing of energy toward everyday life and away from trauma-focused ways of living. Cognitive distortions frequently substantially subside. At the same time, nondissociative and integrative coping strategies must be identified and substituted for dissociative responses to life stressors. The patient must be helped to connect and cope with the larger world in a more functional manner.

Hypnosis. Hypnosis is a set of techniques that facilitate treatment, not a treatment in and of itself. Because of the high hypnotizability and autohypnotic states of the patient with dissociative identity disorder, all treatment of this disorder involves hypnotic phenomena. Clinicians working with this population should obtain training in hypnosis, if only to recognize spontaneous, autohypnotic processes that may be readily observable during treatment, as well as specialized training in hypnosis for traumatized and dissociative individuals. Hypnotherapeutic interventions can often alleviate self-destructive crises or reduce symptoms, such as flashbacks, dissociative hallucinations, and passive-influence experiences. These interventions usually act far more robustly and rapidly than medications, particularly as crisis interventions. It is important that patients be taught to use these techniques independently of the therapist, often by having several self-states become “internal” hypnotherapists. In addition, teaching patients self-hypnosis usually gives them a sense of control over lifelong autohypnotic experiences that have bewildered and overwhelmed them. Hypnosis is also used to create relaxed mental states in which negative life events can be examined without overwhelming the patient’s “window of tolerance.” Clinicians should be aware of current controversies over the impact of hypnosis on accurate reporting of recollections and should use appropriate informed consent for its use. Expert consensus on hypnotherapy for trauma disorders opines that the patient should receive psychoeducation that any memory material emerging during hypnotic interventions should not be viewed differently than any others that emerge during treatment: they may be essentially veridical, partially accurate, or completely confabulated.

Psychopharmacological Interventions. Guidelines for the use of medications with dissociative patients emphasize the need to identify specific treatment-responsive symptoms rather than attempting to treat symptoms of pathologic dissociation per se, as no medication(s) have been shown to be effective for this indication. Medications may be helpful in attenuating symptoms to assist the patient in stabilizing during treatment. Patients should be advised that medication response is likely to be partial, devising the best “shock-absorber” system for the patient at a given time. In general, success is more likely if the target symptoms for the medication are present across most or all self-states rather than confined to one or a few. Because many individuals with dissociative identity disorder lead stressful lives, and psychotherapy for the disorder may also be stressful, rapid symptom oscillations are common. In general, it is not advisable to chase each symptom shift with a medication change. Expert consensus states that it is preferable to change medications in response to shifts in the psychological “climate,” not the day-to-day alterations in the psychological “weather.” Medications and medication targets are listed in Table 20–23.

INFORMED CONSENT FOR PSYCHOPHARMACOLOGICAL INTERVENTIONS. Individuals with dissociative identity disorder require psychoeducation about which symptoms are likely to be medication responsive, the partial nature of medication effects for most comorbid disorders, and that even relatively robust medication effects may appear to be overwhelmed by severe stress—although without the medications, the patient may fare considerably worse during a crisis. These patients respond best if treated like a partner in evaluating the use of medications and medication response, with the physician as expert consultant, rather than in a more authoritarian stance.

Some individuals with dissociative identity disorder have an addiction-like affinity for medication that they use to attempt anesthesia for all psychological distress. These patients may be reluctant to accept the shock-absorber model of psychopharmacology and struggle against the psychiatrist’s attempt to keep the medication regimen as rational and simple as possible. They may come to treatment on complex, extensive polypharmacy regimens that may include a plethora of psychoactive medications including medications for psychiatric disorders, sleep, pain, and medical disorders. Much effort may be required to simplify these regimens, especially as the patient learns more effective psychotherapeutic and symptom-management interventions.

On the other hand, some individuals with dissociative identity disorder report being drugged during abuse and have self-states that are phobic of taking medications. Considerable effort may then be required to educate the patient that medications are tools that potentially could be used for either good or malign purposes; it is how and why they are used that is important to assess. The patient can be asked to risk taking medications as prescribed, and to report honestly about medication effects, so that medication response and adverse effects may be better distinguished from posttraumatic reactivity. Without this process, the psychiatrist may later come to learn that the patient has been partially or fully nonadherent. Individuals with dissociative identity disorder often fear that medications may be used coercively; now that the psychiatrist controls something the individual depends on the psychiatrist can now demand “something” in return. In addition, the “need” for medications may symbolize the patient’s sense of need and attachment as inherently humiliating and dangerous.

Psychoeducation may help alleviate the individual’s concern about the psychiatrist exerting “control” through medications. The psychiatrist only writes a prescription, they do not control all the next steps involved in the patient getting the medication from the pharmacy, taking it home, taking it as prescribed, and honestly reporting back to the psychiatrist.

If the psychiatrist is working as a medication manager alongside a nonphysician psychotherapist, it is important to maintain communication about the patient. The therapist should let the psychiatrist know about issues that may be affecting symptoms, such as an acute loss or life crisis. In turn, the psychiatrist should inform the therapist about medication changes that may have an impact on the psychotherapy and that the therapist can assist in assessing. Education of the therapist about the role of medications in dissociative identity disorder is important as well. It is key for the medicating psychiatrist to attend to issues of boundaries and treatment frame. It is vital that the patient have only *one* primary therapist and that the psychiatrist not become an additional therapist who sees the patient frequently (e.g., once weekly) unless major medication changes require more intensive contact. In addition, the psychiatrist should be clear that they are only available for medication/medical emergencies and not be seen as the designated emergency back-up if the patient cannot reach the therapist during a psychotherapy or safety crisis. Also, the medicating psychiatrist should be clear about their availability for back-up during the therapist’s absences, and only assume this role after specific agreement.

Table 20–23.
Medications for Associated Symptoms in Dissociative Identity Disorder

Medications and somatic treatments for PTSD, affective disorders, OCD, etc.

Alpha-Adrenergic Blocking Agents

Prazosin (Minipress) helpful for PTSD nightmares, and flashbacks, if patient can tolerate BP symptoms. Patients may require up to 20–25 mg daily for maximal effect with divided daily dosing for daytime flashbacks/intrusive symptoms. Time of administration can be adjusted, e.g., patient has a period with severe intrusive symptoms in the afternoon.

Doxazosin (Cardura) also useful for this indication and may help with daytime symptoms with single daily dosing—may require 10–16 mg daily.

Clonidine (Catapres) Less consistently helpful than prazosin/doxazosin.

β-Blockers (for PTSD hyperarousal symptoms; if giving both β-blockers and α-blockers, give α-blocker first to prevent reflex tachycardia)

- Selective Serotonin/Selective Serotonin/Norepinephrine Reuptake Inhibitors (no preferred agent, except for OCD symptoms); these medications mostly provide a partial, “shock-absorber” anti-depressant response; rarely effective for PTSD in this population. There are no data on antidepressant augmentation strategies in this population. The authors’ experience is that this patient population rarely responds to this; other experts disagree. Response likely to be small.
- Fluvoxamine (Luvox) (for OCD presentations)
- Clomipramine (Anafranil) (for OCD, if patient is safe enough to take a TCA)
- Other tricyclic antidepressants
- Monoamine oxidase inhibitors (if patient can reliably maintain diet safely)
- Mood stabilizers (*more useful for PTSD and anxiety than mood symptoms*)
    - Lamotrigine (Lamictal) (may help with depressive symptoms more than others)
    - Carbamazepine (Tegretol)
    - Divalproex (Depakote)
- Benzodiazepines
    - Clonazepam (Klonopin) and lorazepam (Ativan) have best track records

  - Atypical neuroleptics
    - Ziprasidone least likely to cause weight gain: a major issue for many dissociative patients.
    - Clozapine for patients with mixture of psychotic and dissociative symptoms; very severe almost delusional cognitive distortions
    - Typical Neuroleptics, e.g., Perphenazine
  - Medications for acute dyscontrol
    - Oral or intramuscular neuroleptics
    - Oral or intramuscular benzodiazepines
  - Medications for sleep problems
    - Trazodone (Desyrel) (high margin of safety; rareantidepressant effect at antidepressant doses, i.e., 1,000–1,500 mg. daily)
    - Low-dose mirtazapine (Remeron)
    - Low-dose tricyclic antidepressants
    - Low-dose neuroleptics
    - Benzodiazepines, Zolpidem (Ambien) (often less helpful for sleep problems in this population)
    - Anticholinergic agents (diphenhydramine [Benadryl], hydroxyzine [Vistaril])
    - Melatonin and congeners sometimes effective
  - Medications for nonsuicidal self-injury, addictions
    - Naltrexone (ReVia)
- Limited data on electroconvulsive treatment (ECT); best for “double depression” presentation, not for baseline, long-term “mood” issues (see text). No data on other brain stimulation techniques, ketamine infusion, etc.

OCD, obsessive-compulsive disorder; PTSD, posttraumatic stress disorder.

Target Symptoms for Medications in the Treatment of Dissociative Identity Disorder-PTSD Symptoms. The most robust medication treatment effect in this patient population is the response of nightmares to α_1-adrenergic antagonists, most commonly prazosin (Minipress) in a wide dose range, as documented in many studies, primarily of patients with combat-related PTSD. Clinical experience suggests that these medications are more effective in patients with more severe PTSD symptoms. If the patient tolerates the blood pressure effects, research in combat-related PTSD finds that dosages of prazosin as high as 18 to 25 mg daily (in women and men, respectively), in single or twice-daily doses, are effective for this indication. Daytime dosing may also be helpful for control of flashbacks and intrusive symptoms. Recent studies show that doxazosin (Cardura), a longer-acting α_1-adrenergic antagonist, may also be effective for nightmares and intrusive symptoms in PTSD. In addition to prazosin/doxazosin, a variety of medications are helpful as adjunctive agents to help with sleep problems. These are listed in Table 20–23.

Nocturnal PTSD symptoms in this population, and in other patients with complex trauma disorders, may be complicated by posttraumatic reactivity related to reported nocturnal physical and sexual abuse; then, nighttime, the bedroom, being in bed, and even sleep itself may be experienced as posttraumatic triggers. Patients may report being drugged during these attacks, compounding the phobia of bedtime medications in some self-states, while other self-states describe desperation to find an anodyne for sleep problems). Also, patients may report that some self-states “stay up” to protect the patient from possible abuse and/or emerge nocturnally, often in the early morning hours, as this was the only time the patient felt safe enough to allow awareness of internal experience. Because of these issues, some self-states may attempt, usually successfully, to psychologically override the beneficial effects of medications or to refuse to take medications at all. These issues require psychoeducation about nocturnal PTSD triggers; cognitive behavioral interventions (e.g., attempts to decrease posttraumatic responses to these phobic stimuli and to help the patient separate past from the present); internal communication around conflicts between self-state groups about medication; and symptom-management techniques (e.g., finding a “safe place” to sleep for traumatized self-states before the patient attempts go to bed; finding a more adaptive way for self-states to “guard” against feared nocturnal attack, etc.). Patients should begin work on these issues several hours before bedtime.

Other medications have shown to be efficacious for PTSD symptoms in double-blind studies of PTSD patients. These include the selective serotonin reuptake inhibitors (SSRIs) sertraline (Zoloft) and paroxetine (Paxil), the tricyclic antidepressant (TCA) desipramine (Norpramin), and the monoamine oxidase inhibitor (MAOI) phenelzine (Nardil). Unfortunately, most medication studies of PTSD do not include patients with suicidal ideation, multiple comorbid disorders, and substance abuse and thus may have limited ecologic validity for complex trauma patients.

Experts in psychopharmacology for this population have not found robust efficacy for the antidepressant medications for comorbid PTSD symptoms although, as discussed below, these medications may help ameliorate some mood disorder symptoms in these patients.

Other medications for PTSD symptoms that can have efficacy in reducing intrusive symptoms, hyperarousal, and anxiety in patients with dissociative identity disorder include β-blockers such as propranolol (Inderal), other α_1-adrenergic antagonists like clonidine (Catapres), anticonvulsants such as lamotrigine (Lamictal) and carbamazepine (Tegretol), and typical and atypical neuroleptics. One small study showed that risperidone was efficacious for intrusive PTSD symptoms in complex PTSD patients. However, since the response of PTSD, anxiety, and panic symptoms to these medications is usually relatively limited in patients with dissociative identity disorder, the psychiatrist and patient together should carefully balance the potential risks (e.g., metabolic syndrome) against the potential benefits of these medications.

Many patients with dissociative identity disorder report some response to benzodiazepines such as lorazepam (Ativan) or clonazepam (Klonopin) for posttraumatic fear, panic, anxiety, and sleep problems. These patients may tolerate higher doses of these medications than many nondissociative patients. The clinician should educate the patient that severe PTSD involves states of terror, not anxiety. Thus, limited response to anti-anxiety medications is the rule. Careful monitoring of patients on benzodiazepines is necessary to assess tolerance and compulsive drug-seeking behaviors. Gradual tapering of these medications is essential to prevent withdrawal symptoms.

Target Symptoms for Medications in the Treatment of Dissociative Identity Disorder-Mood Disorder Symptoms. As described above, patients with dissociative identity disorder commonly have a complex negative mood state, often described as depression, but with elements of many other negative affective states (e.g., profound shame, grief, distress over ongoing stressful life events, the severely dysphoric impact of PTSD). These patients commonly describe themselves as having been depressed for their “whole life.” Nonetheless, antidepressant medications, including SSRIs, selective serotonin/norepinephrine reuptake inhibitors (SNRIs), TCAs, and MAOIs are often important in the partial reduction of depression and stabilization of mood. The psychiatrist should take a careful history of the mood symptoms and question the patient about other affects such as shame, sadness, and grieving; in addition, the psychiatrist should educate the patient that complete remission of the depressive symptoms is unlikely and about the shock-absorber nature of medications. In general, expert consensus has shown that no antidepressant is likely to be any more effective than any other in this population. Somatic treatments for mood disorder symptoms in dissociative identity disorder patients may be most effective in a "double depression" clinical picture although somatic treatments typically return the depressed mood back to the patient's usual chronic negative mood state.

A subgroup of patients with dissociative identity disorder experiences marked obsessive-compulsive symptoms (OCD) and may preferentially respond to antidepressants such as fluvoxamine (Luvox) and clomipramine (Anafranil). During the clinical evaluation, clinicians should always assess patients for the presence of OCD symptoms.

In general, unless the patients describe more clear-cut, sustained, hypomanic or manic episodes, they infrequently respond to mood stabilizers, especially lithium. If they describe more typical bipolar depression symptoms, usually these too will follow the double depression pattern of treatment response.

OTHER TARGET SYMPTOMS FOR MEDICATIONS IN THE TREATMENT OF DISSOCIATIVE IDENTITY DISORDER. Case reports suggest that aggression may respond to carbamazepine in some individuals with dissociative identity disorder if EEG abnormalities are present. Open-label studies suggest that naltrexone (ReVia) may be helpful to ameliorate recurrent self-injurious behaviors and other compulsive, quasi-addictive high-risk behaviors in a subset of traumatized patients. This medication may be helpful in reducing relapse in patients with dissociative identity disorder who have significant substance abuse issues.

Questions are often raised about the efficacy of neuroleptic medications, particularly for symptoms such as hallucinations. Expert consensus is that the hallucinations in patients with dissociative identity disorder almost never respond robustly to antipsychotics, as they are hypothesized to originate in dissociative, posttraumatic, and autohypnotic processes, not psychosis. Patients with dissociative identity disorder with a true comorbid psychotic disorder can sometimes distinguish between neuroleptic response to the "outside" psychotic voices and report lack of effect on the voices of the self-states. Occasionally, thought-disorganized, overwhelmed, chronically ill patients with dissociative identity disorder who have not responded to trials of other neuroleptics respond favorably to a trial of clozapine (Clozaril).

Electroconvulsive Therapy, Repetitive Transcranial Magnetic Stimulation, and Ketamine Treatment. During their long psychiatric careers of misdiagnosis and treatment failure, many patients with dissociative identity disorder receive unsuccessful trials of ECT. This led to the view that ECT was not an effective treatment in this patient population. One small clinical study of severely depressed dissociative disorder patients, primarily with DSM-III-R dissociative disorder NOS, suggested that, for some patients, ECT was helpful in ameliorating refractory mood disorders and did not worsen dissociative memory problems as measured by the DES. Clinical experience in tertiary care settings for severely ill individuals with dissociative identity disorder suggests that a clinical picture of major depression with persistent, refractory melancholic features across all states may predict a positive response to ECT. As with medications, response to ECT is more likely in a double depression clinical presentation where melancholic symptoms develop as a distinct change from the patient's chronically depressed baseline. ECT response is usually only partial, as is typical for most successful somatic treatments in patients with dissociative identity disorder and returns the patient to their more usual mood state.

There are no data in this population on response to transcranial magnetic stimulation, ketamine infusion, or Esketamine treatment. The authors' clinical experience in a few cases is that none of these treatments were effective for refractory depression or suicidality in dissociative identity disorder individuals. A severely depressed, highly suicidal individual with dissociative identity disorder—who had failed trials of appropriate staged psychotherapy, multiple antidepressants and/or mood stabilizers, and multiple courses of ECT—died by suicide during a trial of ketamine infusions. However, the suicide could have been related, not to ketamine, but to the patient's despair over lack of response to this "last-resort" treatment, as well as ongoing, severe psychosocial stress. As in other cases, ketamine treatment was not associated with worsening of dissociative symptoms.

Inpatient Treatment. As with patients with other psychiatric diagnoses, inpatient hospitalization is most used for treatment of acute worsening of baseline unsafety, typically accompanying a major destabilization. Most dissociative identity disorder individuals experience self-destructive and suicidal ideation/impulses at baseline. Thus, hospitalization returns the patient to their usual outpatient safety status and should not be expected to definitively resolve all safety issues.

Less commonly, in the age of managed care, hospitalization provides a safe environment for intensive work with painful affects and memories that cannot be safely managed on an outpatient basis despite intensive efforts. For several reasons, patients with overt dissociative identity disorder can have disruptive effects on the milieu of general psychiatric units and often stimulate division and conflict among staff about the best way to relate to the patient and the self-states. Typically, the split is between the believers and the skeptics, often paralleling larger societal debates, as well as the subjective conflict within the patient about belief in the dissociative identity disorder diagnosis and the authenticity of traumatic memories. Inpatient treatment on general hospital units is most successful if there is an attempt to bridge different opinions about dissociative identity disorder to focus on a unified approach for ameliorating specific target symptoms and restabilizing the patient to allow safe resumption of outpatient treatment.

Inpatient programs specializing in the treatment of dissociative and PTSD patients are more successful. They provide a unified treatment with a consistent approach to the patient’s symptoms, prioritizing safety issues, often by resolution of internal conflicts and providing a safe environment to work with dangerous self-states; rigorous boundaries and firm limit setting, especially around responsibility for safety and behavioral dyscontrol; specific, active treatment interventions to assist with dissociative and PTSD symptoms; and an emphasis on helping the patient become proficient in symptom management skills, including skills for improved internal communication, cooperation, coordination, and empathy among self-states.

In both specialty and general hospital units, the patient should be expected to respond to their legal name, or at least one specific name, in the hospital milieu and in all therapeutic interactions, except individual psychotherapy/psychiatric rounds or individual interactions with designated hospital staff. The patient should be expected to be able to present in a reasonably functional adult mode in the hospital, with regressed, dysfunctional, or child states restricted to designated one-on-one encounters or private time in the patient’s room. The patient is urged to seek their “inner adult” in this regard. The patient’s activities in the hospital (e.g., group therapy) should be limited to those in which the patient can participate while following these guidelines.

Trauma-Informed Inpatient Care. The concept of trauma-informed care has led to initiatives to reduce to zero the use of inpatient seclusion and restraint. Evidence shows that use of seclusion and restraint engenders heightened risk of injury to both patients and staff; alternatives to these improve the safety of psychiatric inpatient units. Seclusion and restraints can be psychologically traumatizing to patients (and sometimes to staff) as well as retraumatizing to traumatized patients by revivifying—often uncannily—prior traumatic experiences, including those in psychiatric hospitals. Trauma-informed care concepts involve attempts to understand symptoms—particularly those of dyscontrol/unsafety in the hospital—as logically related to prior traumatic events where the patient (and staff) is unaware of the connection between trauma and symptoms. Thus, interventions directed at these issues may provide a safe resolution to the situation, as well as the patient and treatment team gaining a comprehensible, nonpejorative framework for working on many problematic behaviors.

For example, a dissociative identity disorder inpatient repeatedly became dyscontrolled and engaged in nonsuicidal self-injury around the 3 PM shift change on a psychiatric unit, typically resulting in manual and physical restraint. On inquiry about whether there were any trauma reminders about the 3 PM shift change, she responded, “Yes, when I was 16, they put me in a state hospital for 6 months. Most days at 3 PM shift change, some orderlies would tie me up and rape in the quiet room.” This understanding allowed for work with self-states on “separation of past/present,” and symptom management skills around shift change, as well as other hospital-based trauma reminders. The dyscontrol episodes rapidly decreased to zero.

Hospital staff working in the trauma-informed model can educate patients about these initiatives and the importance for patients and staff of reducing use of restraint. Staff and patients can work to identify trauma-reminders and triggers, and anticipate these with development of safe, recovery-oriented methods to manage anticipated dyscontrol. Unfortunately, dissociative identity disorder inpatients may still become so unsafe that, when less restrictive alternatives have failed, hospital staff may be required to use these restraints to prevent escalating dangerous behavior. If restraints are needed, they should be used for the minimum time needed to restore safety. After a restraint episode, processing the event is crucial (e.g., by using the DBT behavioral chain analysis or similar) to focus on making sense of these episodes, noticing early warning signs, and finding earlier interventions (e.g., relaxation techniques, “dialing down” and/or safe-place imagery, and oral prn medication, among others) to prevent symptom escalation.

Specialty units for dissociative identity disorder and related disorders use restraint much less frequently than general hospital units. Specialty unit staff is trained to help patients use symptom management and containment strategies to anticipate and manage potential crises. Inpatients with dissociative identity disorder in behavioral dyscontrol may also respond to as-needed oral dosages of atypical or typical neuroleptics with or without benzodiazepines.

If the dissociative inpatient shows a limited or transitory response to psychosocial interventions, and remains dangerously agitated and unsafe, oral and/or intramuscular prn medications may help decrease these symptoms and avoid more restrictive interventions. Single doses of oral or intramuscular neuroleptics may help manage acute dyscontrol by allowing a change in behavioral state. Options include 2 to 5 mg of haloperidol (Haldol) oral or intramuscular; 5 to 10 mg of oral, sublingual, or intramuscular olanzapine (Zyprexa); or 20 mg of oral or intramuscular ziprasidone (Geodon); with or without 1 to 2 mg of oral/intramuscular lorazepam (Ativan).

Discharge from hospital usually occurs when the patient with dissociative identity disorder has negotiated the acute crisis; has a realistic structure for outpatient safety consistent with discharge to the next level of care; has made adequate gains in symptom management skills and internal communication to handle problems more adaptively; medications have been adjusted to provide better treatment of comorbid disorders; and/or the outpatient psychosocial environment has restabilized sufficiently.

Partial Hospitalization. Partial hospital treatment can be an effective modality for the dissociative patient if clinical interventions target trauma-based issues and adaptations. Specialized partial hospital trauma programs with groups emphasizing symptom containment, cognitive and behavioral interventions, development of life skills, and psychoeducation can effectively help stabilize more severely and persistently ill dissociative patients as well as provide an intensive stabilization experience for higher-functioning individuals. Patients with dissociative identity disorder frequently do not respond positively to more usual partial hospital programs developed for patients with psychotic and/or mood disorders.

Adjunctive Treatments

GROUP THERAPY. Authorities agree that, in therapy groups including general psychiatric patients, the emergence of self-states can be disruptive to the group process by eliciting excess fascination or by frightening other patients. Therapy groups composed only of patients with dissociative identity disorder are more successful, although the groups must be carefully structured, must provide firm limits, and should generally focus only on here-and-now issues of coping and adaptation. Some stabilized patients may be able to work effectively in process-oriented psychotherapy groups. Even so, it is important that there are firm boundaries and limits on group structure and contacts by patients outside of sessions; the intensity of trauma material that the patients can discuss be monitored; and that patients who become destabilized and unsafe may need to return to a group that focuses on symptom management and stability.

FAMILY THERAPY. Family or couples therapy is often important for long-term stabilization and to address pathologic family and marital processes that are common in individuals with dissociative identity disorder and their family members. Education of family and concerned others about the disorder and its treatment may help family members cope more effectively with dissociative identity disorder and PTSD symptoms in their loved ones. Group interventions for education and support of family members have also been found helpful. Family members should be discouraged from interacting with individual self-states, calling them out, and relating to them as separate individuals. Family members should be helped to support the patient’s goal of becoming a functional adult in adult relationships and a functional parent to children. Sex therapy may be an important part of couples’ treatment when the patient is more stabilized, as the patient may be triggered into flashback by sexual interactions and thus intensely phobic of intimate contact. Clinicians who work with these patients and their spouses on sexual issues should be trained to work with survivors of childhood sexual abuse and on the specific complexities of these issues for patients with dissociative identity disorder, including the patient’s difficulty in excluding traumatized self-states from intimate interactions.

Family therapy with the individual’s family of origin is often extremely distressing to patients with dissociative identity disorder. Some interactions with these family members may lead to an acute decompensation in the patient. Contacts should be carefully structured for defined therapeutic purposes, and the patient should be well prepared for potential adverse reactions. At the same time, such meetings may help clarify or resolve the conflictual emotions that patients frequently experience toward family members. “Confrontation” of family members about past traumas in an accusatory manner almost invariably has a disastrous outcome for the patient and the relatives. Before any such meeting, it is most helpful for the patient with dissociative identity disorder to clarify hidden fantasies about these interactions (e.g., “they’re going to say they are sorry”), and how these are typically unlikely to occur.

Despite the description of intrafamilial maltreatment, some individuals with dissociative identity disorder maintain ongoing contact with members of the family of origin, including financial dependence. Clinical case series document ongoing incestuous abuse into adulthood—even middle age—by family members of individuals with dissociative identity disorder. Thus, the clinician should help identify possible lack of safety during these contacts and make this a major focus of treatment to help protect the patient. Typically, specific self-states are activated in this situation and other self-states are “off-line” and/or are unable to interfere. Apart from this or other lack of safety in contact with the family of origin, the clinician should focus on the individual’s conflicts about the family and contact with them, also typically parceled out among self-states. If the individual with dissociative identity disorder is a victim of intimate partner violence, the clinician needs to take a similar approach. In both these situations, clinicians must be patient as there is usually intense, fear-based attachment to perpetrators.

DIALECTICAL BEHAVIOR THERAPY. DBT is a treatment developed by the psychologist Marsha Linehan for borderline personality disorder. Its concepts and techniques may be helpful to structure aspects of stabilization throughout dissociative identity disorder treatment since many patients with borderline personality disorder report significant trauma histories and fit the complex trauma construct. Helpfully for the patient with dissociative identity disorder, DBT focuses on the primacy of safety, therapy-interfering behaviors, affect regulation strategies, distress tolerance, and radical acceptance.

However, DBT also may have notable limitations for the patient with dissociative identity disorder. Many individuals with dissociative identity disorder report—often repeated—failed trials of DBT, as DBT treatment outcome studies show that dissociative individuals do not respond to unmodified DBT. DBT does not focus on stabilization and management of dissociative symptoms such as depersonalization/derealization and dissociative amnesia, flashbacks, and unconscious reenactments, traumatic transference, and related trauma-based symptoms. DBT does not address stabilization of the self-states' contribution to symptoms since resolution of many symptoms in dissociative identity disorder treatment can only occur in the context of work with the patient's self-state system.

EYE MOVEMENT DESENSITIZATION AND REPROCESSING. Eye movement desensitization and reprocessing (EMDR) is a treatment for posttraumatic disorders developed by the psychologist Francine Shapiro. In treatment outcome studies for PTSD, EMDR has similar efficacy to cognitive-behavioral therapy. However, patients with suicidality, substance abuse, and multiple comorbidities have generally been excluded from EMDR outcome studies, as in other treatment studies of cognitive-behavioral therapy and medications for PTSD. Dismantling studies have shown that the eye movements that distinguish EMDR from other therapies do not add to its therapeutic effect, although EMDR practitioners dispute this.

There are no systematic studies of EMDR in patients with dissociative identity disorder, although some case reports have been published. Experts in both EMDR and phasic trauma treatment of dissociative identity disorder have written an appendix to the standard text on EMDR and a section of the ISSTD treatment guidelines cautioning use of this modality in dissociative identity disorder treatment unless the patient is stabilized in stage 1 treatment.

Because EMDR has the significant exposure and free association components, in unmodified form this modality can cause significant harm to individuals with dissociative identity disorder. The lead author has considerable experience with adverse outcomes of EMDR when used in these patients during stage 1 of treatment or before even beginning stabilization in treatment. These outcomes have included severe posttraumatic and dissociative crises, suicide attempts, self-destructive behavior, and acute worsening of PTSD, dissociative, and mood disorder symptoms, often resulting in emergency hospitalization as well as long-term impairment.

The ISSTD Treatment Guidelines, clinical case series, and the EMDR text appendix state that clinicians may choose EMDR as an *adjunctive* treatment in some cases to allow the patient to process memories during stage 2 treatment. Specifically, for the clinician to practice EMDR with patients with dissociative identity disorder, the clinician must be knowledgeable and skilled in phasic trauma treatment, have completed advanced training in EMDR, and have received supervision in this modality for the treatment of dissociative identity disorder.

SELF-HELP GROUPS. Individuals with dissociative identity disorder usually have a negative outcome to self-help groups or 12-step groups for incest survivors, now often moved online. Most experienced clinicians, as well as the ISSTD guidelines, strongly discourage these patients from participating in these modalities. Complications include intensification of PTSD symptoms due to discussion of trauma material without clinical safeguards, exploitation of the patient with dissociative identity disorder by predatory group members, contamination of the patient's recall by group discussions of trauma, and a feeling of alienation even from these other reputed sufferers of trauma and dissociation. Negative outcomes for individuals with dissociative identity disorder using social media have been described in the section on History.

EXPRESSIVE AND OCCUPATIONAL THERAPIES. Expressive and occupational therapies, such as art and movement therapy, have proven particularly helpful in treatment of patients with dissociative identity disorder. Art therapy may help patients contain and structure severe dissociative identity disorder and PTSD symptoms, as well as allow for safer expression of thoughts, feelings, mental images, and conflicts that may be difficult to verbalize. Movement therapy may facilitate normalization of body sense and body image for these severely traumatized patients. Occupational therapy may help the patient with focused, structured activities that can be completed successfully and may help with grounding and symptom management.

Ms. V was a 33-year-old married woman who presented to psychiatric attention after discovering her 5-year-old daughter “playing doctor” with several neighborhood children. The patient began to become fearful that her daughter would be molested. She became increasingly obsessed with this idea, much to the bafflement of her husband. The patient was seen by her internist and treated with antianxiety agents and antidepressants, but with little improvement. Ms. V became increasingly anxious, phobic, depressed, and preoccupied. She sought psychiatric consultation from several clinicians but repeated, adequate trials of antidepressants, antianxiety agents, and supportive psychotherapy resulted in limited improvement. After the death of her father from complications of alcohol use disorder, the patient became more symptomatic. He had been estranged from the family since the patient was approximately 12 years of age, owing to his drinking and associated antisocial behavior. Ms. A developed a variety of somatic complaints, including headaches, abdominal pain, menstrual and gastrointestinal problems, back pain, and sleep difficulties. Repeated medical workup was unrevealing, leading to diagnoses of fibromyalgia, irritable bowel syndrome, and premenstrual tension. Family and marital difficulties increased, and work function deteriorated. Psychiatric hospitalization was precipitated by the patient’s arrest for disorderly conduct in a nearby city. She was found in a hotel, in revealing clothing, engaged in an altercation with a man. She denied knowledge of how she had come to the hotel although the man insisted that she had come there under a different name for a voluntary sexual encounter. On psychiatric examination, the patient described dense amnesia for the first 12 years of her life with the feeling that her “life started at 12 years old.” She reported that, for as long as she could remember, she had an imaginary companion, an older adult African American woman, who advised her and kept her company. She reported hearing other voices in her head: several women and children as well as her father’s voice repeatedly speaking to her in a derogatory way. She reported that much of her life since age 12 was also punctuated by episodes of amnesia: for work, for her marriage, for the birth of her children, and for her sex life with her husband. She also reported perplexing changes in skills and repeated experiences of sudden “blasts” of emotion that would appear and suddenly disappear, primarily anger, fear, and sadness. She described feeling that her thoughts were being removed from her mind, especially when attempting to recall difficult issues in life. She reported experiencing saying things that felt out of her control: “My mouth has a mind of its’ own; my mind has a mouth of its’ own; my mind has a mind of its’ own!” At times, she reported automatic writing, as if her “hand is possessed” and different handwritings appeared without her control. Her husband noted that, at times, she would speak like a child; sometimes she would adopt a southern accent; and, at other times, she would be angry and provocative. She frequently had little recall of these episodes. Questioned more closely about her early life, the patient appeared to enter a trance and stated, “I just don’t want to be locked in the closet” in a child-like voice. Inquiry about this produced rapid shifts in state between self-states that differed in manifest age, facial expression, voice tone, and knowledge of the patient’s history. Gradually, with more communication from self-states, the patient recalled a history of family chaos, brutality, and neglect during the first 12 years of her life, until her mother, also alcoholic, achieved sobriety and fled her husband, taking her children with her. The patient began to recall episodes of physical abuse, sexual abuse, and emotional torment by the father, her siblings, and her mother. The self-states reported “piggy-backing” emotions, thoughts, impulses, and actions, leading to the experiences of passive-influence. This also explained “taking thoughts out of her head, that she isn’t supposed to know,” when the patient suddenly blanked on what was being discussed. The patient improved significantly with psychotherapy directed at stabilization of dissociative identity disorder and PTSD. She responded well to clomipramine (Anafranil) with a significant reduction in obsessive-compulsive and depressive symptoms. Family sessions with the patient’s mother and siblings confirmed much of the history, with the family recalling episodes of maltreatment that the patient did not recollect. The patient’s mother had bid the family never to speak of the earlier difficulties, hoping that everyone would “just forget the whole thing.” An older sister, who also had been severely abused, met diagnostic criteria for dissociative identity disorder. Other siblings had histories of substance use disorders, mood disorders, PTSD, and dyscontrol episodes. Family therapy was helpful in stabilizing the patient’s marriage and helping her husband with the aftermath of the hospitalization and its precipitants. The patient found adjunctive art therapy particularly effective. At 3-year follow-up, the patient reported fusion of most self-states and marked diminution in dissociative, somatoform, and PTSD symptoms, although she still required clomipramine for stabilization of mood and OCD symptoms.

Dissociative Amnesia

The DSM-5-TR diagnostic criteria for dissociative amnesia emphasize that the disorder represents a recall deficit in *autobiographical memory* (some authorities [e.g., Staniloiu and Markowitsch—see below], using Tulving's model, prefer the term episodic-autobiographical memory [EAM]), particularly for traumatic and stressful experiences. As described in the prior sections, cumulative adverse experiences during childhood, particularly the extent and severity of childhood sexual and/or physical maltreatment, lead to increased deficits in autobiographical memory before age 18. To qualify for a diagnosis of dissociative amnesia, the amnesia cannot be better explained by medical, neurological, or substance use disorders, including a seizure disorder, transient global amnesia, head injury, or another neurocognitive disorder. The amnesia also cannot be better explained by dissociative identity disorder, ASD, PTSD, somatic symptom disorder, a psychotic or cyclic mood disorder, or acute intoxication. Dissociative fugue is conceptualized as a specifier of dissociative amnesia, not a separate dissociative disorder. Typically, in the dissociative fugue presentation the individual comes to clinical attention because of apparently purposeful travel or bewildered wandering that is accompanied by dissociative amnesia for identity and/or other important autobiographical information, including details of the travel the person has engaged in.

Studies with the MID find that the symptom of amnesia in the dissociative amnesia *disorder* is experienced as retrospective gaps and/or fragmentary memory of the individual's history that is too extensive to be explained by normal changes in autobiographical memory over the lifespan. Conversely, in dissociative identity disorder, amnesia includes *both* retrospective autobiographical memory disturbances and ongoing dissociative amnesia for current life. These include perplexing acquisition (or disappearance) of possessions (e.g., clothes, tools, weapons, items in one's grocery cart); inexplicable fluctuations in well-learned skills, habits, and knowledge (e.g., including finding evidence or being told of skills or habits about which the individual has no awareness); description of disremembered behavior; and/or ongoing episodes of dissociative fugue, among others.

What Is Dissociative Amnesia?

Dissociative amnesia can be broadly defined as a reversible autobiographical memory impairment in which groups of memories for personal experience that would ordinarily be available for conscious recall cannot be retrieved or retained in autobiographical memory (or, if temporarily retrieved, cannot be wholly retained in consciousness). This disturbance may be based on neurobiologic changes in the brain caused by traumatic stress. However, the disorder manifests as a potentially reversible form of psychological inhibition or retrieval deficit. Amnesia due to brain injury or cognitive disorders usually cannot be reversed, as is the case with most episodes of amnesia due to substance intoxication. On the other hand, a variety of clinical and experimental studies have shown state-dependent effects for some forms of memory retrieval with a variety of types of intoxicants (e.g., alcohol, barbiturates, amphetamines, marijuana) and with some mood state shifts (e.g., mania vs. depression). In these state-dependent/mood-state–dependent memory studies, reinstatement of context is important in demonstrating the phenomenon. Reinstatement of context, as well as cued retrieval, may be a factor in the recall of memories that have been subject to dissociative amnesia although, as described below, this usually comes in the form of PTSD intrusive symptoms and flashbacks, or highly emotion-filled recollections, not normal autobiographical recall.

Discrete Behavioral States Theory and Dissociative Amnesia

From the perspective of DBSs theory, dissociative amnesia represents complex examples of state-dependent/mood-dependent learning, memory storage, and retrieval. As discussed previously research shows that childhood autobiographical memory deficits (CAMD) involve memories of trauma as well as recall deficits for major aspects of everyday life whilst growing up with chronic interpersonal trauma. Autobiographical memory is fundamental for developing coherent, adaptive sense of self and identity. Accordingly, a major concomitant of dissociative amnesia is disordered, often limited, constricted self-states/systems. Metacognitive, autonoetic functions are also disrupted, as individuals with dissociative amnesia typically are impaired in making sense of themselves and their behavior. Missing, jumbled, and/or fragmentary memory impairs the ability to connect current thoughts, emotions, and behavior with past experiences. Because of dissociative amnesia, there may be unconscious reenactments. For example, these individuals may be unable to recognize threat and terminate involvement with dangerous attachment figures, both from the childhood and/or adulthood. Individuals with dissociative amnesia often minimize the importance of their memory deficits (see case examples below) and avoid attempts to connect current life difficulties with their early life development.

During the "memory wars" of the 1990s and early 2000s, the debate polarized between a group that posited that "trauma memory is not forgotten, it is remembered all too well," and those who insisted that delayed recall for trauma occurred, often for years of an individual's life. Some of the debate can be understood as disagreement about the *mechanism* for delayed recall of traumatic experiences (e.g., repression, dissociation, ordinary forgetting, conscious withholding of information due to shame, or some or all of these). The DBS model, supported by a variety of clinical and experimental studies posits that *both* models contribute to the understanding of trauma memory. Overwhelming trauma can be encoded in memory, and recall can be inhibited/impaired by state-dependency. Trauma survivors commonly describe deliberate attempts to avoid thinking about trauma, and/or reluctance to share memories due to shame, attachment to perpetrators, failure to appreciate the abusiveness of interpersonal violence during childhood, etc. All these factors may be relevant for an individual growing up with ongoing intrafamilial violence. As one individual explained his improved recall as an adult of growing up with unpredictable, sadistic maltreatment, "I had to live there."

Conceptualizing Dissociative Amnesia

Conceptualizing dissociative amnesia generally involves four factors. First, relatively large groups of memories and associated affects have become unavailable, not only single memories, feelings, or thoughts. Second, the unavailable memories usually relate to day-to-day autobiographical information that would ordinarily be a routine part of conscious awareness: who a person is, what they did, where they went, what happened, with whom they interacted, what was said, what they thought and felt at the time, and so on. Third, the ability to remember new information, general cognitive functioning, and language capacity are usually intact, although in some cases the dissociative process can interfere with retrieval of perceptual, semantic, and procedural information (e.g., what did I see?, what is a contact lens?, what do I do with it?) and registration of new memories (anterograde or continuous dissociative amnesia). Finally, the dissociated memories often indirectly reveal their presence in variably disguised form, such as intrusive visual images, flashbacks, somatoform symptoms, nightmares, conversion symptoms, and behavioral reenactments, with the individual impaired in ability to conceptualize their behavior as a reenactment. Dissociative amnesia states may alternate with states of hyperamnesia, flashbacks and/or derealized awareness in which the person experiences detachment or estrangement from autobiographical memory.

Clinical Presentation of Dissociative Amnesia

There are two basic presentations of dissociative amnesia. First, the "classical presentation" is generalized dissociative amnesia, a sudden disturbance in which extensive aspects of autobiographical memory are unavailable to conscious verbal recall, often including loss of awareness of personal identity. These patients are frequently seen in emergency departments or general medical or neurologic services because the sudden development of memory loss requires medical assessment. In addition, during an acute amnestic state, some individuals demonstrate symptoms that are best conceptualized as symptoms of peritraumatic dissociation and/or ASD. These include disorientation, perplexity, alteration in consciousness, somatoform symptoms, purposeless or fugue-type wandering, or a combination of these.

The clinical consensus of dissociative disorder experts is that generalized dissociative amnesia is most commonly a manifestation of dissociative identity disorder; an amnestic self-state is created or activated to resolve seemingly irreconcilable stress and conflict, often in the context of acute trauma, and may also represent a symbolic suicide. Multiple episodes of generalized dissociative amnesia are strongly indicative of underlying dissociative identity disorder.

The classic disorder is an overt, sometimes dramatic presentation that frequently results in the patient being brought quickly to medical attention because of amnesia. The paradigmatic case is that of the individual who is found without memory for identity or life history, with or without a dissociative fugue, sometimes leading to media reporting of the case. This presentation of amnesia is frequently found in individuals who have experienced extreme acute trauma such as combat or rape. It also commonly develops in the context of profound intrapsychic conflict or emotional stress; for example, a terrified soldier in combat who experiences excruciating conflict over their urge to flee despite believing that “all cowards should be shot.” However, in the latter cases, there is almost always an antecedent history of severe traumatic experiences in childhood and/or adulthood. An illustrative case involves a World War II naval petty officer who presented with generalized dissociative amnesia. After his ship in the Pacific was hit, he was adrift for some days in a life raft with several other sailors. One was horribly wounded, screaming in agony, again and again begging them to kill him. Eventually, as he knew the man would die from his wounds, the petty officer shot him and pushed his body into the sea. All in the boat agreed to keep what happened secret. Generalized dissociative amnesia developed a year later when he was to be remobilized for the invasion of Normandy.

The second and far more prevalent form of dissociative amnesia is a deletion of significant portions of autobiographical memory and/or fragmentary experience of it. Ordinarily, patients do not complain of this and it is usually only discovered in taking a careful life history. Dissociative amnesia often has a clear-cut onset and offset, so that the person is subjectively aware of gaps in autobiographical memory. For example, a patient reports that she does not remember being in third grade despite having a clear memory of other school years. Individuals with dissociative amnesia also may report confused, jumbled, fragmentary autobiographical memory. Some may report both patterns for different segments of their history. Usually, such symptoms are associated with traumatic circumstances: for example, the patient reports being told that in third grade she was kidnapped by her estranged father during a contentious custody dispute, held hostage for months, and repeatedly raped by him. In some cases, the patient may deny recall for their entire childhood or other major life epochs. Unlike generalized dissociative amnesia, here there is no loss of personal identity or impairments in procedural and/or semantic memory.

A 45-year-old, married, left-handed, male bus dispatcher was seen in psychiatric consultation on a medical unit. He had been admitted with an episode of chest discomfort, light-headedness, and left-arm weakness. He had a history of hypertension and ischemic chest pain, without myocardial infarction. Psychiatric consultation was called, as the patient demonstrated memory loss for the previous 12 years, behaving and responding to the environment as if it were 12 years previously. He insisted that he was unmarried, denied he had a son, did not recall his current job, and denied recollection of current events (e.g., current president) but correctly recalled historical events from 12 years before (e.g., president 12 years earlier). Physical and laboratory examinations were unchanged from the patient's usual baseline. Imaging studies of the brain were normal. The patient displayed intact intellectual function other than the 12-year amnesia. He was perplexed by the contradiction between his memory and current circumstances. The patient described a childhood history of brutal beatings and physical discipline by his father, beginning when he was 3 years old. He was a decorated Vietnam-era combat veteran and described amnesia alternating with fragmentary memory for his combat experiences. In the military, he had been a left-handed golden gloves champion. He described amnesia for several of his boxing matches and, to his dismay, at one time "found" himself ferociously beating another soldier who had attempted to bully one of his friends. He was educated about his disorder and given the suggestion that his memory could return as he could tolerate it, perhaps overnight during sleep or perhaps over a longer time. If this strategy was unsuccessful, hypnosis or an amobarbital interview was proposed.

At the next examination, the patient reported that his memory had spontaneously returned. Before the amnestic episode, he described being stunned when his beloved wife said she wanted a separation and asked *him* to discuss this with his son. In addition, due to a variety of other family and work problems, he had come to feel completely responsible for his coworkers and for the care of his relatives. He was panicked, overwhelmed, and enraged. He had felt violently angry with his wife but had vowed in the past that he would "beat to death anybody who tried to hurt her." He stated that he would have attempted suicide but he "couldn't" because he had too many people relying on him. The amnesia developed after he felt a kind of "paralysis" in his left arm while talking with his wife about the separation. Insightfully, he reported that, “I went back to the time, 12 years ago, before I met my wife, had this job and all this family stress. Back then, none of these problems existed.”

Diagnosis and Clinical Features

As noted above, there are two major clinical presentations of dissociative amnesia.

Classic Presentation. The primary data base for this form of dissociative amnesia are clinical case series and case reports going back to World War I. For example, a systematic neuropsychological study assessed memory function in 10 patients with severe retrograde dissociative amnesia with a median duration of 14 months. They were assessed on tests of retrograde and anterograde amnesia as well as of autobiographical memory. One respondent admitted to malingering and two acknowledged symptoms consistent with dissociative identity disorder, with the acute dissociative amnesia a manifestation of an amnestic self-state. Other than the malingerer, every patient had a long psychiatric history; dissociative respondents evidenced symptoms that were not feigned and improved or remitted with treatment. All subjects reported difficulty recalling specific autobiographical memories. In addition, tests of famous faces and past public events revealed that individual patients showed unique but internally consistent, memory difficulties for this material. These were not thought to represent factitious processes but rather the patients’ commonsense, individualized notions of how memory works that informed the phenomenology of their presentations.

For many years, Angelica Staniloiu and Hans J. Markowitsch have studied cases and case series of individuals with generalized dissociative amnesia, with and without fugue. As discussed in the section on Neurobiology, they have proposed a neurobiologic model of “mnestic block syndrome” to describe the phenomenology and neurobiology of the retrieval deficits characterizing the complex memory disturbances in these individuals. Recently, they published an extensive evaluation of one of the largest modern case series of 28 dissociative amnesia/dissociative fugue individuals, ranging from ages 11 to 54 (average = 35.6 years), all with the chronic form of dissociative amnesia. Staniloiu and Markowitsch emphasize the prevalence of this chronic course and do not include cases of the sometimes spontaneous, rapid resolution of amnesia found in the older civilian and military literatures.

Staniloiu and Markowitsch performed an extensive neuropsychological test battery, including the DES and the DES-T, personality inventories, and forensic tests (e.g., the Test of Memory Malingering [TOMM]) to assess issues of feigning. Four cases involved criminal legal involvement, including embezzlement, reckless driving, and attempted blackmail in a foreign country. The latter individual developed dissociative fugue and generalized dissociative amnesia reportedly in response to the blackmail victim hiring someone to kill him. However, patients were not studied with lifetime trauma inventories, PTSD scales, or diagnostic assessments for dissociation like the MID or the SCID-D. Because of the individuals' amnesia for life history, an inventory of traumatic exposures was obtained via clinical interviews and collateral sources. A newspaper reporter acquired corroboration of some of the childhood trauma sustained by the blackmailer. As in prior studies, the researchers did not find a specific personality profile associated with generalized dissociative amnesia. They found extensive deficits in executive function, attention/concentration, and emotional processing, the latter characterized by emotional blunting and "*la belle indifference,"* as found in many classic cases from the 19th and 20th centuries. Confirming findings in prior studies, there were documented deficits in episodic, autobiographical, semantic, and procedural memory, as well as the rare form of anterograde (continuous).

Dissociative Amnesia

Staniliou and Markowitsch found no evidence supporting the sociocognitive/fantasy models of dissociation. Confirming the TM of dissociation they found evidence of childhood maltreatment in seven cases despite the fact that childhood history was often unavailable due to amnesia. Antecedent trauma history included childhood sexual, physical, and emotional abuse; placement in an orphanage; bullying by peers; unexpected severe illness or death of a loved one; death threats; surviving a terrorist bombing and witnessing numerous dead maimed bodies after the attack, etc.

Documented events that precipitated amnesia were either psychological stress or a mixture of psychological and physical stress (e.g., several cases with mild head injuries). As these authors have reported previously, a history of highly stressful/traumatic migration experiences was a factor in six generalized dissociative amnesia/fugue cases. Stresses included refugee status; lack of ability to speak the language across several migrations; social dislocation and marginalization in their new countries; being orphaned and sexually abused in a foreign country as a child, and other stressors. They documented three cases of child/adolescent dissociative amnesia, one with a 2-year chronic course. As with the adult cases, these children had documented stressful/traumatic events as precipitant and/or distal etiology of the amnesia.

Other studies and case series report that these patients may present with comorbid somatic and/or classical conversion symptoms, as in the bus dispatcher case described above. They also may display alterations in consciousness, depersonalization, derealization, trance states, spontaneous age regression, dissociative fugue, and ongoing anterograde dissociative amnesia. Depression and suicidal ideation are reported in many, but not all, cases. A personal past and/or family history of somatoform or dissociative symptoms predisposes individuals to develop acute amnesia during traumatic circumstances. As has been found in studies of Vietnam-era veterans, the World War II studies of amnesia during combat show that the severity, intensity, and frequency of combat are the most important factors in the development of combat-related posttraumatic disorders, including dissociative amnesia. In a study of British soldiers after the battle of Dunkirk, the point prevalence of dissociative amnesia for the battle was 14% to 35% in those on the front lines. Six percent developed dissociative amnesia with “trifling stress” (e.g., they were in training camp or waiting for deployment). Almost all in the latter group were described with the dissociative fugue subtype and one with generalized dissociative amnesia. Most had family histories of “neurosis, psychosis… [and] amnesic attacks,” as well as dissociative fugue.

Generalized dissociative amnesia/fugue may occur in individuals where avoidance of responsibility plays a role. Sexual indiscretions, legal difficulties, financial problems, or fear of anticipated combat may be part of the clinical matrix that surrounds the amnesia. There is a possible association in some cases with an antecedent history of head trauma with or without loss of consciousness. Most case series of combat-related dissociative amnesia/fugue, with or without conversion symptoms, specifically ruled out head trauma as a significant factor in development of these disorders.

If carefully questioned, some of these patients give a history of recurrent episodes of amnesia or fugue, typical of dissociative identity disorder. In the dissociative identity disorder cases, the amnesia occurs when—to cope with overwhelming or traumatic circumstances and/or profound intrapsychic conflict—the person begins to experience themselves as a new—or newly reactivated—self-state who is unaware of some or all their past.

In some acute dissociative amnesia cases, the amnesia resolves within hours to months, spontaneously or through psychotherapy, hypnotherapy, pharmacologically facilitated interviews, or combinations of these modalities. However, as described above, recent case series show that a subset of these patients develop a chronic, disabling course. Chronic dissociative amnesia may be severely impairing and associated with limited ability to return to family, social, or occupational roles, even when patient “learns” about the missing autobiographical information. This acute presentation should be regarded as a psychiatric emergency, as clinical experience suggests that the more time passes before resolution of the amnesia the more likely that a chronic course will occur.

Ms. M is a 55-year-old woman was seen in psychiatric consultation because of sudden, complete loss of memory for her identity and entire life history. She had been found by her family after participation with a female friend in a “new age” therapy workshop where, in a session involving picturing oneself going back in time and being reborn, the patient appeared to enter a deep trance state in a fetal position with difficulty awakening. When she did, she could not identify who she was, where she was, or anything else about herself or her situation. Several days later, when found by her family, she was dazed, confused, and disoriented, unable to recognize them or provide any history of recent events. She had difficulty learning new information, seemingly forgetting it as it was provided. She was panicked and complained of not “feeling in my body.” She had difficulty recalling how to dress herself, what were and how to use her contact lenses, tampons, and so on. She underwent a complete neurologic, neuropsychiatric, and neuropsychological evaluation with the conclusion that she was not suffering from a cognitive disorder. When referred for consultation, she had “relearned” information about her identity, family, and her life, but fundamentally had no available autobiographical memory to contextualize this information. She complained of ongoing amnesia, spontaneous trances, and periods of intense depersonalization and derealization. She was impaired in performing most daily functions, having no recall of these skills. She had been depressed, and this improved with antidepressant medication, but her memory remained unchanged. Her memory did not return with hypnotic interventions or after an amobarbital interview. In psychotherapy, she gradually was able to recall events around the loss of memory. The “friend” who had accompanied her to the conference had pressured her sexually and, with great distress, she finally acceded. This event, experienced with profound shame, had uncanny parallels to episodes of childhood sexual abuse that the patient had described in a prior psychotherapy. After this disclosure, the patient began to have gradual return of her autobiographical memory. She described and was observed to shift among several self-states, including bewildered, amnestic child states, and several “new” Ms. M states without memory of the past. She reported that her previously functional states had all “gone away,” refusing to return, citing how “devastated”—that is, ashamed—she felt about the events that precipitated the amnesia. She received a diagnosis of dissociative identity disorder. Gradually, she was more able to access her “functional” self-states and become better able to manage life issues. Still, the amnestic self-states denied recall of the sexual assault. These self-states “knew about” her history prior to the amnesia but denied first-hand memory of these. On reflection, she thought that, during several very stressful times in her life, she suffered at least two prior episodes of dissociative amnesia and conversion symptoms, albeit recalled without a sense of first-hand experience.

DISSOCIATIVE AMNESIA FOR VIOLENT/HOMICIDAL BEHAVIOR. Several international case series document dissociative amnesia for violent or homicidal behavior. These violent individuals described acute onset of extreme emotions, especially violent rage, precipitated by major interpersonal stress, such as unexpectedly finding evidence that a spouse had been unfaithful. Typically, they “came to” recognizing that they had savagely assaulted or killed someone. In general, these individuals did not attempt to use memory loss for exculpation. They called law enforcement and/or emergency services for assistance themselves after realizing what they had done, and accepted responsibility and legal consequences for the violent episodes. Unfortunately, none of the case series systematically studied these individuals with self-report or diagnostic inventories for dissociation, PTSD, other comorbid psychopathology (e.g., substance use disorders, mood disorders), history of trauma, history of violence, etc.

DISSOCIATIVE FUGUE SPECIFIER. The dissociative fugue specifier of dissociative amnesia is usually associated with generalized memory loss typical of the classic presentations of dissociative amnesia. Expert consensus views dissociative fugue as usually a manifestation of dissociative identity disorder, with amnestic, peripatetic self-states, either newly created or activated under stress, that flee from the individual’s current life, sometimes in an unconscious reenactment of a childhood attempts to escape from a violent family. The classic fugue cases with change in identify (e.g., Ansel Bourne) were considered by contemporaneous experts (e.g., William James) to be “double consciousness/dual personality.” During dissociative amnesia/fugue, individuals may engage in complex, purposive behavior, sometimes over weeks or months. Recurrent dissociative fugue states likely only occur in dissociative identity disorder.

In his case studies, Janet described that a single idea that symbolizes or condenses (or both) several important ideas and emotions frequently dominates the patient's thinking when in the fugue state. At some point, the individual becomes aware of the amnesia and/or loss of personal identity, resulting in their being brought for clinical evaluation. Sometimes, at this point, the individual returns to their baseline state, usually with amnesia for the fugue and identity alteration. As in case Example 1 in this section, there may be an age-regression to a chronologically earlier life period.

In some severe cases of PTSD, commonly comorbid with dissociative identity disorder, nightmares may be terminated by a waking fugue during which the patient hides in a closet, goes to another part of the house, or runs outside. Dissociative children and adolescents may be more limited than adults in their ability to travel. Thus, fugues in this population may be brief and involve short distances. Some cases of children or adolescents who precipitously run away actually may be instances of dissociative fugue, with the child escaping from an abusive or violent home situation. No systematic research exists on this point. All children and adults with the dissociative fugue subtype should be assessed for dissociative identity disorder.

During a fugue, patients often seem "normal" and tend not to attract attention. Some individuals, however, display overtly bizarre, disorganized, or dangerous behavior, such as a soldier who, during battle, began a fugue episode by standing up, exposing himself to intense enemy fire, and then dazedly walked away from the front lines. After the termination of a fugue, in addition to amnesia, the patient may experience perplexity, confusion, trance-like behaviors, depersonalization, derealization, and conversion symptoms. As the patient with dissociative fugue begins to become less dissociated, they may display mood, anxiety or PTSD symptoms, and/or intense suicidal ideation.

Nonclassic Presentation. Individuals with nonclassic dissociative amnesia can be said to have a *covert* dissociative syndrome, because their primary complaint infrequently relates directly to amnesia. Commonly, patients with the nonclassic presentations of amnesia do not reveal the presence of dissociative symptoms unless directly asked about them. These patients are often uncomfortable when amnesia is inquired about and may minimize the presence or rationalize the importance of the symptom.

In nonclassic presentations, the amnesia consists of a circumscribed memory gap or series of memory gaps for the life history, and/or fragmentary recall of the life history; often for times when traumatic events occurred, such as childhood or wartime. However, as described in prior sections, more extensive dissociative amnesia for life history also may occur, with the person describing large gaps in recall and/or fragmentary recall of years of early life. Like other traumatized individuals, these patients frequently come to treatment for other symptoms and disorders, such as depression, mood swings, substance abuse, sleep disturbances, somatoform symptoms, anxiety, panic, suicidal behavior, nonsuicidal self-injury, violent outbursts, eating disorder symptoms, and/or interpersonal problems. Amnesia may also occur for flashbacks or behavioral re-experiencing episodes related to trauma.

Ms. G is a 35-year-old, unmarried psychologist who was admitted to hospital because of increasingly severe nonsuicidal self-injury, a behavior she had engage in for years. She reported chronic depressed mood, anxiety, and interpersonal problems with inability to form intimate relations. Medications and intensive outpatient psychotherapy had helped moderate some of her symptoms but she had become increasingly demoralized by her overall failure to improve in treatment. She described virtually no recall for her life between ages 7 and 13 including family holidays, school events, and vacation trips. Her father had abused alcohol throughout her childhood, only achieving sobriety in her early adolescence. When she was 13, her older sister disclosed sexual abuse by the father to their minister, who "did a church intervention," but there was no social service or police referral. Her father acknowledged the abuse at the time and received treatment by church counselors for alcohol abuse and "sex addiction." He denied any other episodes of sexual abuse toward his children. Ms. G denied any feelings about these episodes stating that, "he took care of the problem. Now, he's a great support. I have no reason to be mad at him." She explained her amnesia by saying that "maybe nothing important happened and that's why I don't remember." Treatment focused on establishing safety from self-injury. Self-harm impulses increased after phone conversations with her parents. Also, she began to have nightmares, highly disturbing bodily sensations in her genitals and rectum, and would wake up vomiting. Realizing that these were symptoms of PTSD, she found this uncanny, stating, "Nothing bad ever happened to me." Finally, she began to have discrete flashbacks of sexual assaults by her father, together with her sister at times. In an emotional family meeting without the father present, the sister, long estranged from the family, confirmed that these events had occurred. She said, "I had hoped you would never remember, so I never told you."

Subsequently, Ms. G's father was arrested and pled guilty to federal charges of possession of videos of children being raped. At the state level, he pled guilty to charges of sexual abuse over many years of girls in the church, and was sentenced to a lengthy concurrent sentence in federal prison. The state investigated and found systematic child sexual abuse within the patient's family church, including ministers, prominent church members like her father, and individuals in the church hierarchy.

Several of these men were charged and convicted of child sexual abuse. The state obtained a conviction of a church elder for conspiracy to conceal child abuse in the church. The long-term psychotherapy of Ms. G and her sister were complex and difficult, resulting in estrangement from their mother. Ms. G stopped engaging in any form of self-destruction or high-risk behavior. She described grieving over many lost years of her life but reported freedom to enjoy her life now, with renewed energy in her work and her personal life.

Only a few studies have examined the clinical characteristics of nonclassic dissociative amnesia patients. In a small, highly selected sample referred for consultation at a specialty clinic for dissociative disorders, most dissociative amnesia patients were women. As in the data from the ACE study, almost all had a prior history of childhood or adult physical, sexual, or emotional abuse and neglect, or a combination of these. The trauma history was less severe than has been reported in patients with dissociative identity disorder. Duration of dissociative amnesia ranged from relatively brief recurrent episodes to years of the life history. Traumatic precipitants of the amnesia were most common, although in 30% of cases amnesia was present for problematic behaviors such as sexual indiscretions and nonsuicidal self-injury. The most common comorbid conditions were mood disorders, PTSD, and personality disorder symptoms. Patients differed from individuals with dissociative identity disorder in having lower rates of substance abuse, nonsuicidal self-injury, hallucinations, fugues, sexual dysfunction, and somatoform symptoms. Mean DES scores were lower on average than in cases of dissociative identity disorder. Family history was characterized by alcohol abuse and mood disorders.

A random community sample of 1,301 university students in Sivas, Turkey was assessed for dissociative disorders and borderline personality disorder. The authors validated and used a self-report inventory for dissociative amnesia, the Steinberg Dissociative Amnesia questionnaire (SDAQ) developed as a screening measure based on the SCID-D dissociative amnesia scale. Eighty respondents (6% of the sample) met DSM-IV-TR criteria for borderline personality disorder, 78 of whom (98%) also met DSM-IV-TR criteria for a dissociative disorder. Of these, 15 participants (18.8% of the subsample with borderline personality disorder, 1.2% of the total sample) met diagnostic criteria for DSM-IV-TR dissociative amnesia. As in all other studies, in dissociative identity disorder a diagnosis of borderline personality disorder was a severity measure, indicating a higher level of childhood trauma exposure and more severe, acute dissociative and PTSD symptoms.

Pathology and Laboratory Examination

Dissociative amnesia can be diagnosed with the DDIS, SCID-D-R, and/or MID. The DES and the SDAQ can be used as screening measures. Patients with dissociative amnesia have been reported to have high hypnotizability, as measured with standardized hypnosis scales.

Differential Diagnosis

Autobiographical Memory, Nonpathologic Amnesia, and Dissociative Amnesia. Nonpathologic forms of amnesia have been described, such as infantile and childhood amnesia, amnesia for sleep and dreaming, and hypnotic amnesia. Most forms of dissociative amnesia are thought primarily to involve difficulties with autobiographical memory—rather than implicit or general semantic memory—although rarely, as illustrated above, semantic and/or procedural memory may be disrupted. Autobiographical memory may have episodic and semantic components (e.g., “last year I bought a new car” vs. “every 5 years I buy a new car,” respectively).

Studies of autobiographical memory support the notion that infantile and childhood amnesia can be experimentally documented. However, data on preverbal learning and memory in young children suggest that later amnesia may result from difficulty translating preverbal memory into verbal form. Experiments in preverbal children designed to overcome this factor show that children can later report preverbal memories accurately in verbal form under suitable experimental conditions.

Studies demonstrate that normal adult autobiographical memory has a retention gradient for the last 20 to 30 years of the subject’s life, often with a subjective sense of wearing away of memories for the past, in contrast to the subjective gaps/fragmentation of memory typical of dissociative amnesia. Older adult patients in autobiographical memory studies have a relative decrease in recent autobiographical memories and a *reminiscence component* for the events of their youth. Older adults have more semantic autobiographical memories and fewer episodic autobiographical memories, anchored in specific time points. Consistent with this, there may be a shift from the first person (field) to the third person (observer) form of autobiographical memory.

Autobiographical memory has several complex aspects, including emotion, mood, and state-based differentials in memory. For example, negative events may be preferentially recalled in depressed mood states. Schema memory effects may result in the inclusion of a memory of something one has not experienced into memories of a sequence of experienced events. Due to a variety of factors, individuals may develop “pseudomemories” for events that did not occur. This can include false confessions, experimentally induced “false memories,” and failures to recall events that *did* happen (omissions or “false nonmemory”). It may be difficult to differentiate actual memory alteration in these situations from the *belief* that something did or did not occur.

Autobiographical memory for repeated routine events, such as going to work or school every day, may not be encoded as discrete memories for each day but rather as broad memory categories for the events that repeatedly reoccur (*regeneric memory*). This is also thought to occur with some traumatic experiences, such as repeated episodes of childhood sexual abuse that occur frequently during many years (e.g., a child being incestuously raped most weekends of every year from age 5 to 15).

Clinically, patients with dissociative amnesia may have a variety of autobiographical memory deficits. For example, an individual case study of a rape survivor with subsequent generalized dissociative amnesia was evaluated with cognitive testing while amnesic and after memory recovery. As compared to a control with a cognitive disorder, the dissociative amnesia patient was able to recall memories from various points in the life history without the temporal gradient that characterized the retrograde amnesia of the cognitive disorder patient. Recall of autobiographical information during dissociative amnesia seemed related to positive life events that were unconnected with the traumatic events precipitating the amnesia. In this patient, implicit autobiographical memory phenomena were also documented. Similar phenomena have been described in posthypnotic amnesia, with implicit demonstration that the memories for which amnesia has been suggested have been encoded and stored but without being directly accessible for retrieval. Amnesic patients may also have intense reactions to stimuli that are emotionally significant without being aware of the reason for the reaction or the significance of the stimulus, such as when a patient with PTSD has a flashback without consciously knowing what triggered it and often without clear recall later of the memory being evoked.

Several studies have confirmed the clinical observation that individuals with dissociative amnesia for their life history can demonstrate implicit autobiographical memory while amnesic. When asked to free associate, to imagine, or to make up a story, on projective tests, even with ectodermal responses, patients with dissociative amnesia include elements of autobiographical information without necessarily being consciously aware of this.

There is no single test or examination that can establish absolutely whether a memory disorder has a dissociative, neurocognitive, factitious, malingered, or mixed etiology. The clinician evaluating the amnesic patient must have a reasonable index of suspicion about any of these. In ambiguous cases, thorough reassessment on an ongoing basis of the clinical situation is helpful. However, most patients with dissociative amnesia present differently from other disorders with memory impairment (Table 20–24). The evaluation of acute dissociative amnesias is described in Table 20–25.

Cognitive Disorders. In patients with dementia, amnestic disorders, and delirium, the memory loss for personal information is embedded in a far more extensive set of cognitive, linguistic, attentional, behavioral, and memory problems. Loss of memory for personal identity is usually not found without evidence of marked disturbance in many domains of cognitive function. Confabulation may be present to various degrees and is usually implausible or bizarre. Causes of neurocognitive amnestic disorders include thiamine deficiency (Korsakoff psychosis), cerebral vascular accident (CVA), postoperative syndrome, infection, anoxia, and transient global amnesia. Electroconvulsive therapy (ECT) may also cause a marked temporary amnesia and, occasionally, persistent memory problems. In ECT-related amnesia, however, memory loss for autobiographical experience is unrelated to traumatic or overwhelming experiences and seems to involve multiple types of personal experience, most commonly those occurring just before or during the ECT treatments.

Table 20–24.
Differential Diagnosis of Dissociative Amnesia

Normal Autobiographical Memory Difficulties
Age-related cognitive decline
Nonpathologic forms of amnesia
Infantile and childhood amnesia
Amnesia for sleep and dreaming
Hypnotic amnesia
Dementia
Delirium
Amnestic disorders

Neurologic disorders with discrete memory loss episodes
Posttraumatic amnesia
Transient global amnesia
Amnesia related to seizure disorders

Substance-related amnesia
Alcohol
Sedative-hypnotics
Anticholinergic agents
Steroids
Marijuana
Narcotic analgesics
Psychedelics
Phencyclidine
Methyldopa (Aldomet)
Hypoglycemic agents
β-Blockers
Lithium carbonate
Many others

Other dissociative disorders
Dissociative fugue subtype
Dissociative identity disorder
Other dissociative disorders
Acute stress disorder
Posttraumatic stress disorder
Somatic symptom disorder

Psychotic episode/disorders

Lack of memory for psychotic episode when returns to nonpsychotic state

May be difficult to differentiate from thought disordered recall in schizophrenia

Mood disorder episode

Lack of memory for aspects of episode of mania when depressed and vice versa or when euthymic

Factitious disorder

Malingering

Table 20–25.
Evaluation of Acute Amnesia

- Complete history (to the extent that this can be gathered)
- Medical history
- Psychiatric history
- Trauma history (combat, violence, childhood maltreatment, etc.)
- Developmental history
- Collateral informants (if available)
  - Family and concerned others
  - Medical, military, police, and social service records
- Sequential clinical observation
- Physical and neurologic examination
- Baseline physical and laboratory examination
- Full mental status examination
- Mini-mental state examination
- Dementia workup
- Electrocardiogram
- Electroencephalogram
  - Telemetry in unusual cases
- Brain imaging
- Neuropsychological assessment

Posttraumatic Amnesia. In posttraumatic amnesia due to brain injury, there is usually a clear-cut physical trauma, a period of unconsciousness or amnesia—or both—and often objective clinical evidence of brain injury. In general, the length of the posttraumatic amnesia is a reasonable predictor of cognitive outcome. Retrograde amnesia may also occur. An extensive retrograde amnesia out of proportion to the head injury may warrant an investigation for dissociative factors. Head trauma may cause an anterograde amnesia in some cases.

Seizure Disorders. In most cases of epileptic seizures, the clinical presentation is different from that of dissociative amnesia, with clear-cut ictal events and sequelae. Patients with nonepileptic seizures, however, may resemble epileptic patients in some ways but may also have dissociative symptoms, such as amnesia, elevated scores on dissociation measures like the DES, and an antecedent history of psychological trauma. Both epileptic and nonepileptic seizures can worsen with stress. Rarely, patients with recurrent focal-onset impaired-awareness seizures (previously complex partial seizures) may present with ongoing bizarre behavior, memory problems, irritability, or violence, making them hard to distinguish from patients with dissociative amnesia. Some patients with epilepsy have exhibited wandering or semi-purposeful behavior, or both, during seizures and in postictal states, for which there is subsequent amnesia. However, seizure patients in an epileptic (as opposed to dissociative) fugue often exhibit abnormal behavior, including confusion, perseveration, and abnormal or repetitive movements as well as experiencing an aura, motor abnormalities, stereotyped behavior, perceptual alterations, incontinence, and a postictal state. Stressful life events may be associated with an increase in seizure frequency in some susceptible epileptic patients and thus this factor alone is insufficient for differential diagnosis. Serial or telemetric EEGs, or both, usually show abnormalities associated with epilepsy accompanying behavioral changes whereas the EEG is normal in nonepileptic seizures. However, both may occur in the same individual.

Substance-Related Amnesia. Several substances and intoxicants have been implicated in the production of amnesia. Common offending agents are listed in Table 20–24. In most cases, a careful history from the patient and ancillary sources, sequential clinical observation, and objective testing clarifies the substance-related nature of the amnesia. In some instances of pathologic intoxication, in which a small amount of alcohol or of another substance produces a major behavioral disinhibition, the alcohol may be producing its effect by facilitating the onset of a dissociative episode in a susceptible individual, possibly suffering from dissociative identity disorder.

This may be analogous to the disinhibition that occurs in a pharmacologically facilitated clinical interview. Individuals may report amnesia for violent or other out-of-character behavior during such an episode.

The most difficult differential diagnostic problem usually involves patients with a history of both substance-induced and dissociative memory problems. Some of these patients may minimize dissociative amnesia and vice versa. Clinically, the relative contribution of the substance abuse and the dissociation may only be fully clarified by sequential clinical observation once the patient has achieved sobriety.

WANDERING BEHAVIOR. Individuals with dissociative amnesia may engage in confused wandering. To qualify for the dissociative fugue subtype, the individual must exhibit *purposeful* travel away from home or a customary place of daily activities and is usually preoccupied by a single idea that is accompanied by the wish to escape.

Wandering behavior during a variety of general medical conditions, toxic and substance-related disorders, delirium, dementia, and amnestic syndromes could theoretically be confused with the dissociative fugue subtype of dissociative amnesia. However, in most cases, the somatic, toxic, neurologic, or substance-related disorder can be ruled in by the clinical history, physical examination, laboratory tests, or toxicologic and drug screening. Use of alcohol or other substances may be involved in precipitating an episode of dissociative fugue. Wandering and purposeful travel may occur during the manic states of bipolar disorder or schizoaffective disorder, with difficulty recalling behavior that occurred after return to the euthymic or depressed state. In purposeful travel due to mania, the patient is typically preoccupied with grandiose ideas and often calls attention to themselves owing to inappropriate behavior. Assumption of an alternative identity does not occur, although the subjective differences in sense of self, emotions, and behavior may lead the person to feel like “different people” during the various mood states. Similarly, peripatetic behavior may occur in some patients with psychotic disorders. Memory for events during wandering episodes in such patients may be difficult to ascertain owing to the patient’s thought disorder. However, in general, dissociative fugue patients do not demonstrate a psychotic thought disorder or other symptoms of psychosis.

Transient Global Amnesia. Transient global amnesia may be mistaken for dissociative amnesia, especially since stressful life events may precede either disorder. However, in transient global amnesia there is a sudden onset of complete anterograde amnesia and learning abilities; pronounced retrograde amnesia; preservation of memory for personal identity; anxious awareness of memory loss with repeated, often perseverative, questioning; overall normal behavior; lack of gross neurologic abnormalities in most cases; and rapid return of baseline cognitive function, with a persistent short retrograde amnesia. The patient is usually over 50 and shows risk factors for cerebrovascular disease, although epilepsy and migraine have been etiologically implicated in some cases.

Dissociative Disorders. As noted previously, individuals with dissociative identity disorder or dissociative amnesia experience autobiographical memory deficits for past experiences. Most patients with dissociative identity disorder report multiple forms of complex amnesia, including recurrent blackouts, disremembered behavior, unexplained objects in their possession, and fluctuations in skills, habits, and knowledge. Some patients with other dissociative disorders, particularly acute dissociative reactions to stressful events, may develop amnesia in the context of other dissociative symptoms (e.g., perceptual disturbances). Individuals with more extensive amnestic episodes may qualify for a dissociative amnesia diagnosis (see section on Other Dissociative Disorders).

Acute Stress Disorder and Posttraumatic Stress Disorder. Many individuals with dissociative amnesia meet partial or full diagnostic criteria for ASD or PTSD. The DSM-5-TR stipulates that, to be diagnosed, the dissociative amnesia must be distinct from the course of ASD or PTSD. In practice, clinical judgment usually determines whether the extent of the amnesia warrants a separate dissociative diagnosis. As described in the section on etiology, a large-scale, latent class analysis study of traumatized adolescents found that the best model fit for PTSD dissociative subtype included dissociative amnesia in addition to depersonalization and derealization symptoms.

Other Psychiatric Disorders. State-related memory problems that are observed in individuals with cyclic mood disorders and psychosis may be understood, following the DBST framework, as state-change disorders or exacerbations with state change (e.g., acute psychosis or exacerbation of psychotic disorder). Upon resolution of an acute psychotic episode, many individuals report lack of recall of their symptoms, behavior, or life events while psychotic. This lack of recall may also be observed in many individuals who recover from an episode of delirium. In contrast, a minority of individuals may fully recall their symptoms of psychosis or delirium and become tormented with confusion, fear, and shame regarding these experiences. State-related transitions between mania, depression, and euthymia may be accompanied by an autobiographical memory deficit for mood, behavior, perceptions, and sense of self during the different states. Manic individuals may deny recall of depressive episodes and vice versa. During euthymic states, the individual may have difficulty recalling details of being in the other mood states.

Malingering and Factitious Amnesia. There is no definitive test, battery of tests, or set of procedures that invariably distinguishes true dissociative symptoms from feigning. Malingerers/individuals with factitious disorders have been noted to continue their deception even during hypnotically or barbiturate-facilitated interviews. As noted previously, many of the classical dissociative amnesia cases occur in a context of financial, sexual, and legal problems or in soldiers who wished to escape from combat. Feigned amnesia is more common in individuals presenting with the acute, classic forms of dissociative amnesia. However, in one forensic case, an adult attempting to sue an admitted abuser using the delayed discovery rule was shown to have falsified delayed recall for trauma to overcome the statute of limitations. Investigations showed that the patient had discussed the abuse with others on many occasions before the purported delayed recall. On the other hand, some patients may have re-dissociated, and deny recalling and discussing traumatic experiences in the past.

Also, the legal theory of delayed discovery for trauma due to dissociative amnesia unrealistically requires complete amnesia (i.e., “repressed memory”) for childhood sexual abuse, with de novo recollection many years later. Although this does occur, most dissociative amnesia individuals experience variable recollection of childhood abuse in adulthood with fragmentary recall, derealized recall, emotionless recall, PTSD intrusive symptoms, and recall then re-dissociating memory. This can include often partial recollection of some abuse experiences but not others, even those corroborated by witnesses, and/or documented in social service or medical records. Some individuals display dissociative disconnection of the seriousness of maltreatment. For example, a man reported that, after lighting a fire in a wastebasket at age 4, his father burned his fingers with a lighter. When the clinician cited that as abusive, the man insisted it was legitimate discipline.

As some U.S. states have done, a more realistic legal standard to expand the statute of limitations for childhood abuse is that the individual was unable to appreciate the *impact* of trauma until later life, whether dissociative amnesia is present or not. For example, maltreated individuals may not connect or comprehend the damage done to them until later life; after failed marriages, damaging substance use disorders, loss of employment, psychiatric hospitalizations, nonsuicidal self-injury, suicide attempts, and other difficulties.

Malingering of general dissociative amnesia with/without fugue may occur in individuals who are attempting to avoid consequences of legal, financial, or personal difficulties and in soldiers attempting to avoid combat or unpleasant military duties. These precipitating factors may be present as well in bona fide dissociative amnesia/fugue (see above). In the forensic context, the examiner should always consider the diagnosis of malingering when dissociative amnesia/fugue is claimed. On the other hand, factitious forms of dissociative amnesia often resemble the CAMD clinical picture. This may be related to wishing to seem a more interesting patient and/or adopting the imitative role of a childhood trauma survivor, or that of a more severely harmed trauma survivor. In some complex cases both genuine and exaggerated amnesia/fugue may be reported. For example, a woman with dissociative identity disorder feigned experiencing a several-day dissociative fugue after reported revictimization by a childhood abuser. As discrepancies in her story emerged, she admitted that she had spent the weekend with a new lover, betraying her long-term partner and frightening her outpatient therapist.

On the other hand, in clinical case reports, many malingerers have quickly confessed their deceptions spontaneously or when confronted by an examiner. In these nonforensic reports, the malingered amnesiacs were frequently pathetic individuals whose deception was transparent. Nevertheless, it can often be unclear where the conscious deception began and the unconscious defenses ended. In forensic contexts, feigned dissociative amnesia often is conveniently located for events related to criminal behavior but not to other areas of life. Feigners may become angry and resistant when their memory issues are challenged. They may claim amnesia for some circumstances where their responsibility is at issue yet appear to have detailed recall of events and people with whom they are displeased and from whom they wish to extract some form of gain.

Bethany Brand and colleagues have developed psychological assessment batteries to attempt to discriminate feigned from genuine dissociative disorders and from dissociative disorders with factitious/malingered exaggeration. These may include standard screening and diagnostic inventories for dissociative disorders, personality assessment, MMPI, Rorschach, other measures to assess trauma history, PTSD, and complex trauma, as well measures used in forensic assessment of feigning (e.g., SIRS-II, TOMM). She and her colleagues have demonstrated that patients with actual dissociative disorder may be misclassified on some forensic measures for feigning. Interpretation of findings may require a psychologist with extensive clinical and assessment experience with dissociative disorders.

A patient who presents to psychiatric attention asking to “recover repressed memories” as a chief complaint is most likely to have a factitious disorder, and may display factitious or the imitative form of dissociative identity disorder described in prior sections. Most of these individuals do not describe bona fide amnesia when carefully questioned but they are often insistent that they must have been abused—or more extensively abused—in childhood to explain their unhappiness or life dysfunction. Unlike feigners, most dissociative individuals are fearful, ashamed, and avoidant of recall of traumatic memories and minimize and deny the dissociative amnesia and trauma history, even when corroborated. Memory recall in dissociative amnesia does not present as typical episodic, autobiographical memory recall. Rather, these usually are experienced as disturbing intrusive flashbacks with alarming and vivid perceptual, somatoform, and emotional experiences. They may display freezing or agitation, often with defensive dissociative symptoms like depersonalization/derealization, trance states, and attempts to “re-dissociate” the material.

Clinical Course and Prognosis

Little is known about the clinical course of dissociative amnesia. Acute dissociative amnesia can resolve spontaneously once the person is removed to safety from traumatic or overwhelming circumstances. At the other extreme, many patients develop chronic forms of generalized, continuous, or severe localized amnesia who may be so profoundly disabled to require social support, such as nursing home placement or intensive family caretaking.

Individuals with nonclassic dissociative amnesia may have a persistent form of the disorder that does not cause them overt distress and thus may not seek clinical attention. Some of these individuals can be understood as being in an episode of PTSD characterized primarily by avoidant symptoms. In some cases, a later traumatic event or stress—even a relatively minor one—precipitates a florid episode of PTSD with alternations of re-experiencing and avoidant or amnestic symptoms.

Studies have shown that recall of previously dissociated memory frequently occurs outside of therapeutic settings, triggered by a variety of stimuli, including the person's children reaching a specific age at which the individual was abused, abuse of their own children, media accounts of trauma or abuse, death of an abusive parent, a variety of sensory cues, and feeling safe enough from danger in one's life situation, among others. Meta-analytic studies show that there is no difference in accuracy of trauma memory with delayed recall compared with "continuous" trauma memory.

Treatment

The diagnosis and treatment of dissociative amnesia became controversial in the 1990s due to disputes, often in the media and the courts, over delayed recall of childhood sexual abuse (see section on History of Dissociation). Prior to this, dissociative (psychogenic) amnesia and delayed recall for combat or other forms of trauma had not been considered controversial, even written about by clinician/researchers who later claimed it could not occur. Authors from the World War I and II eras opined that some of these wartime cases were malingered. In this regard, particularly during and after World War I, punitive treatments such as electric shock were devised for wartime amnesia and conversion symptoms. Also, controversy exists not only for the existence of dissociative amnesia but also for how to treat it.

Critics of the dissociative amnesia construct invented a false, nonexistent treatment: *recovered memory therapy* (RMT) to characterize a procedure in which clinicians supposedly make aggressive efforts and pressure patients to recall reported hidden, long-forgotten traumas as the central focus of treatment. RMT does not represent a known school of therapy or of scholarly research. It may more accurately describe individual fringe practitioners and/or media or layperson views of trauma treatment.

The controversy over delayed recall of childhood sexual abuse is often discussed as if the clinical situation involves forensic determination of the “truth” of “accusations.” Episodic, autobiographical memory is adaptive, not forensic. The gist of trauma memory is more likely accurate than peripheral details. In most circumstances, clinicians do not “validate” information patients report in treatment (e.g., where they were born, who was in their family, where they live, etc.) Barring extensive documentation that may include perpetrator admissions, documented physical injuries, etc., the clinician must realize that they “were not there” to determine the “truth” of maltreatment reports. Traumatized patients often vacillate between recalling childhood maltreatment (e.g., “I know this happened to me”) and discounting their recollections (e.g., “I’m a liar, I made it all up”). As discussed in the section on dissociative identity disorder, just as in any other psychotherapy, the clinician’s role is to assist the patient in making sense of their autobiography, not to make a forensic determination. This allows patients to explore issues like screen memories, viewing events differently from a more integrated perspective, and/or becoming increasingly clear that they accurately recall experiencing maltreatment during childhood. Clinicians find that it is more common that traumatized patients struggle, not with trauma “accusations,” but with attachment to abusive caregivers, associated with dogged rationalization of or self-blame for maltreatment. Thus, reviews of the literature on trauma treatment emphasize that the predominant model for work on traumatic memories involves a focus on symptom stabilization and safety, integration of posttraumatic memories, beliefs, cognitions, affects, somatic representations, and object relations, not on memory recall as an end in itself.

Phase-Oriented Treatment. As for dissociative identity disorder, phase-oriented treatment is the current standard of care for the treatment of dissociative amnesia, although there are no systematic studies with large cohorts of patients. Treatment of the patient with acute, classic dissociative amnesia follows a similar phasic model. However, here, memory recall is a central issue, because loss of memory for personal identity and/or dense deletions in autobiographical memory is disabling and requires rapid intervention. On the other hand, case series of patients with generalized dissociative amnesia document serious suicide attempts and even death by suicide when amnestic barriers are removed precipitously without adequate preparation and patient stabilization. Clinicians should make sure that they and the patient are both prepared for the abreaction of extreme emotions (e.g., terror, rage, extreme shame), that accompany recollection of devastating traumatic experiences and/or extreme psychological conflict.

Safety in Acute Stress Disorder with Dissociative Amnesia. Studies of psychiatric combat casualties with acute dissociative amnesia—with or without fugue—provide useful practical principles for treating ASD with dissociative amnesia. Acute generalized dissociative amnesia/fugue may occur in civilian settings during natural disasters, mass casualty events, refugees exposed to terrorism, torture, and civilians affected by war with/without refugee status. Unfortunately, in many of the latter circumstances, resources for medical, let alone psychiatric, care are minimal. In refugee populations, individuals may suffer from additional physical and sexual violence and exploitation, be grieving over the violent deaths of relatives, and lack basic provision of food, water, clothing, housing, or sanitation.

To the extent possible, establishing the person's physical safety is the first concern. This involves removing the individual from the traumatizing environment (e.g., acute combat), evaluating and treating medical problems, and providing shelter, food, and the ability to sleep. The paradigm of *psychological first aid* may be helpful in the initial stabilization of the acutely traumatized, amnesic individual. Sedative medications, such as the benzodiazepines, may be indicated to assist the patient with sleep. Case series suggest that, in this acute context, many patients with acute amnesia may have a relatively rapid spontaneous remission of symptoms once brought to the safety of the hospital or other protected environment (e.g., removal from the battlefield, provision of sleep). If immediate spontaneous remission does not occur after removal from traumatic environments, symptoms may abate later simply during the clinician's taking a psychiatric history or merely with suggestions and reassurance.

Phasic Treatment. As described in the section on dissociative identity disorder, three basic phases are recognized for treatment of both types of dissociative amnesia. This structure is heuristic to some extent because aspects of each phase may be worked on during another. First, there is a stabilization phase, with a focus on safety, symptom control, containment of affects and impulses, and education about trauma treatment. Once adequate personal safety and clinical stability are established, if indicated, the individual may engage in a second phase that focuses on the integration of traumatic material in greater depth. This processing may involve attempts to gently interpret the fear that may underlie ongoing amnesia and to resolve material that is not dissociated or less completely dissociated. Finally, there is a third phase of *resolution* or *reintegration* in which the traumatized person is reconnected to ordinary life. In this phase, the focus is less on the trauma per se and more on the development of a renewed, reinvigorated life that overcomes the lack of freedom imposed by symptoms of the trauma disorder and the dominance of the person's psychology by issues related to traumatization.

In general, studies of treatment outcome in survivors of rape and childhood sexual abuse have shown better outcome when patients directly discuss dissociated trauma material in the context of a carefully designed, physically structured individual or group psychotherapy. International studies of individuals in specialized treatment for childhood sexual abuse show that dissociative trauma survivors do not improve without a direct focus on dissociation.

Safety in Nonclassic Dissociative Amnesia. Patients with nonclassic amnesia presentations generally should be managed within the framework of a longer-term psychotherapy directed at resolving the complex psychological sequelae of severe traumatization due to childhood maltreatment, combat, rape, intimate partner violence, trafficking, and/or other forms of adult victimization. Here, too, the first tasks of treatment are restoration of the patient's physical well-being and safety and establishment of a therapeutic alliance. The clinician must be prepared to intervene actively if the patient is acutely dangerous to self or others. Typically, these patients' difficulties involve suicide attempts, nonsuicidal self-injury, eating disorders, high-risk behaviors (e.g., reckless driving), alcohol or substance abuse, involvement in abusive or destructive relationships, episodes of rage or violence, abuse of the individual's own children or family members, and lack of adequate food, clothing, or shelter. Hospitalization may be necessary for stabilization, as may referral to specialty resources such as treatment for substance abuse or eating disorders.

In individuals with severe intrusive PTSD symptoms alternating with amnesia, containment, and management of intrusive recollections rather than detailed processing of the traumatic material are usually the main goals of the stabilization phase of treatment. This may be accomplished with supportive psychotherapy, pharmacotherapy, imagery or hypnotic techniques for containment and symptom control, and/or dialectical behavioral and cognitive therapy techniques, or a combination of these. Although there is no pharmacologic agent that specifically targets dissociative amnesia, specific psychopharmacological treatment of the patient’s PTSD, mood, impulsive, psychotic, obsessive-compulsive, or anxiety symptoms may help stabilize severe symptoms that prevent the patient’s meaningful participation in psychotherapy.

Long-term treatment for these patients is focused on the manifold dimensions of dysfunction that chronic trauma engenders. These include problems with emotional, mood, anxiety, and impulse regulation; disordered attachment schemas engendering troubled relationships and problems with interpersonal boundaries; spontaneous altered states of consciousness and dissociation; memory problems; cognitive distortions and disordered meaning systems; perceptual abnormalities; problems with the sense of self and body image; somatoform symptoms; and self-destructiveness.

Treatment of Amnesia. Treatment of patients with nonclassic forms of dissociative amnesia necessitates that, to provide ongoing informed consent, the clinician should familiarize themselves with the controversies about trauma and memory, as well as the standard of care for phasic treatment. Factors that usually contraindicate intensive memory integration work are listed in Table 20–22 (see also Table 20–23 in the section on Dissociative Identity Disorder).

Free Recall. Patients with both forms of dissociative amnesia may respond well to free-recall strategies in which they allow memory material to enter consciousness. The clinician is supportive and nondirective but focuses on conflict, psychological defenses, and/or conscious reluctance to free recall to take place. Classic free-association suggestions are often the most helpful in understanding factors that interfere with recall and in allowing recall to occur at a pace that the patient can tolerate. Authorities recommend that in all trauma processing work, to the extent possible, the clinician not ask leading questions, whether in a free-recall situation or with methods that facilitate memory recall (see below). This includes statements like, “what happened then?,” “what happened next?,” or “what did you think/feel at the time?” When exploring posttraumatic cognitive distortions like unwarranted self-blame and shame for aspects of the traumatic experiences, the clinician can inquire about what led to these beliefs, what would the person think if someone else described similar reactions in a traumatic context, and so on. These strategies gently challenge posttraumatic cognitive distortions. In general, it is most helpful to recontextualize these beliefs by locating them within the emerging history, as the patient may experience the all-too-common posttraumatic schema that self-blame is preferable to the experience of utter helplessness at the core of most traumatic experiences. Distress at recall may be titrated by techniques for distancing and fractioning emotion. For example, the patient may be asked to give a brief “headline” or two, as if reading a newspaper, but not to go into more detail (e.g., reading the newspaper article in full) until better able to do so.

Transference Interpretations. Studies of transference in patients with combat-related PTSD and severe dissociative disorders indicate that a traumatic transference is usually the predominant transference theme in these individuals. This is a set of unconscious perceptions of the clinician based on relationships formed under traumatic circumstances. For instance, the therapist becomes the buddy who was killed next to the patient in battle, the persecutory abusive parent, the incompetent officer who sent the patient into battlefield disaster, the neglectful relative unconcerned about the patient’s abuse, or the patient themselves who is subject to the sadistic, abusive behaviors of an implacable other. Identification of the overt patterns of traumatic transference observed by the therapist may be another route to undoing amnesia. A recent study of corroborated trauma memory found that transference-based recall was more accurate than other forms of facilitated recall.

Cognitive Therapy. Cognitive therapy may have specific benefits for individuals with trauma disorders. Identifying the specific cognitive distortions that are based in the trauma may also provide an entrée into autobiographical memory for which the patient experiences amnesia. As the patient is more able to correct cognitive distortions, particularly about the meaning of prior trauma, more detailed recall of traumatic events may occur, although the potential for flooding is a significant concern. Educating patients about the concept of hindsight bias may be helpful. Here, later information biases understanding of events as they unfolded in real time (e.g., the “Monday morning quarterback”). In posttraumatic disorders, hindsight bias is invariably *negative*, with shame, self-blame, and survivor guilt for what occurred.

Facilitated Recall of Trauma Material. A hierarchy of techniques to facilitate recall has been described for dissociative amnesia. The techniques include: (1) context reinstatement (i.e., attempts to focus the patient on time periods for surrounding the amnesia); (2) state-dependent recall (i.e., intensification and focus on affects or somatic sensations that appear to be related to trauma, such as terror, horror, confusion, rage, or sudden nausea or dizziness); and (3) specialized adjunctive techniques (e.g., hypnosis, relaxation, imagery, pharmacologically assisted interviews). Each of these techniques may facilitate recall of dissociated memories, particularly in classic forms of dissociative amnesia. It is particularly important that, when implementing these techniques, to the extent possible, the clinician use nonleading, nonsuggestive questions to minimize concern about inaccurate or confabulated recall. Some nonclinical research studies suggest that techniques for facilitation of recall may increase inaccurate recall. This remains a complex topic in terms of the ecologic validity of this research to treatment of traumatized individuals; what aspects of recall may be inaccurate (e.g., peripheral details vs. central aspects of the narrative), and what factors induce the inaccuracy (e.g., the techniques themselves or more general expectations and demand characteristics, such as hypnosis is somehow “special” in evoking accurate memories). The clinician’s overall stance about trauma treatment and approach to trauma memory is most important in managing these issues.

Hypnosis for Amnesia. Hypnosis has frequently played an important adjunctive role in the treatment of individuals with dissociative amnesia and dissociative fugue. Hypnosis is not a treatment in itself; rather, it is a set of adjunctive techniques that facilitate certain psychotherapeutic goals. The construct of hypnosis encompasses a wide and complex domain of phenomena and responses that have been applied to many therapies and diverse individuals. Concern about inaccuracies in hypnotically facilitated recall is based on research on memory for nontraumatic information showing that hypnosis increases a person’s confidence in what is recalled, even if it is erroneous. However, critical reviews have noted the complexity of this research problem, the inconsistency of findings among studies, and the many variables related to hypnosis and nonhypnotic factors that appear to influence this and related phenomena. These reviews have emphasized that memory confabulation is related to techniques of social influence on suggestible subjects rather than to hypnosis per se.

The use of hypnosis for memory recall in no way ensures the veracity, or lack of veracity, of the information produced. The clinician should be aware of the reconstructive nature of memory and should avoid suggestions to the contrary. Thus, in addition to obtaining the patient's informed consent for the use of hypnosis, it is currently considered prudent to also obtain informed consent concerning the reliability or possible lack thereof for autobiographical information revealed during hypnosis. Studies show that educating patients about the controversies regarding hypnosis and memory and informing them of the need to critically evaluate any memory material that they recall during hypnosis, or at any other time in treatment, reduces the likelihood of credulous acceptance of potentially inaccurate information.

Also, clinicians should be aware that, in some states of the United States, individuals who have been exposed to therapeutic hypnosis might be enjoined from giving testimony as a witness in court. This may be so even if the hypnosis had little relationship to the events at legal issue. The clinician should attempt to discuss fully these issues with the patient and their legal counsel.

Hypnosis can be used in several different ways in the treatment of dissociative amnesia. Hypnotic interventions can contain, modulate, and titrate the intensity of symptoms; facilitate controlled recall of dissociated memories; provide support and ego strengthening for the patient; and, finally, promote working through and integration of dissociated material. In addition, the patient can be taught self-hypnosis to apply containment and calming techniques in their everyday life. Successful use of containment techniques, whether hypnotically facilitated or not, also increases the patient's sense that they can more effectively be in control of alternations between intrusive symptoms and amnesia.

In the acutely amnestic patient, a few sessions to help the patient experience trance successfully and to explore containment and distancing techniques may be sufficient to allow focused hypnotic work on the material for which the patient is amnestic. Clinicians since World War I have recognized how valuable it is for patients to process dissociated material repeatedly over several sessions, often at different levels of affective intensity, to complete the process of integrating the material. In cases of acute amnesia, the first session for memory processing may necessarily be explorative. Here, again, the initial goal is to gain an overview of the information for which the patient is amnesic. As in cases of chronic amnesia, subsequent sessions then focus on reworking the material in greater detail. However, in nonclassic cases of severe, longstanding, posttraumatic disorders, the amnesia itself should be addressed only gradually in the context of a longer-term psychotherapy in which life stability and function are the basic foci.

It is useful to try to account systematically for different dimensions of the traumatic experience—sensory, affective, cognitive, and behavioral—to ensure that all key components have been identified and reconstructed. It is also useful to attempt to account systematically for a variety of painful affects that are commonly experienced during traumatic experiences such as despair, sorrow, grief, horror, shame, helplessness, rage, guilt, confusion, and anguish. Not all of these may be present in each patient; however, it is useful for the clinician to keep track of which affects the patient most readily describes and which seem less available. Inquiry about these other affects may be quite helpful in resolving the amnesia. Shame, horror, helplessness, and overwhelming confusion are emotions that patients may have the most trouble identifying without assistance from the therapist.

The treatment process is similar when the acute amnesia results not from traumatic experiences but from aspects of the person’s current behavior that conflict with deeply held moral values or behavioral standards. In these instances, treatment initially seeks to help the patient tolerate these affects and conflicts without resorting to dissociative defenses. Frequently, these patients describe a developmental history characterized by a rigid family moral code enforced with harsh physical discipline. These experiences often appear to be the traumatic underpinning of the dissociative diathesis. Psychotherapy and hypnotherapy in these patients are directed in part at reducing the patient’s brutally unreasonable, and often conflicting, expectations of themselves and the guilt and shame that so often accompany acute dissociation. Furthermore, these patients may have great difficulty tolerating anger or violent impulses because such affects tend to trigger recall of earlier experiences with physical abuse or similar traumas, often as a flashback experience. Thus, psychotherapy may serve not only to focus on the conflicts that led to the acute amnesia but also to explicate and work through the patient's thoughts, feelings, and self-perceptions related to the antecedent traumatic events.

Somatic Therapies. There is no known pharmacotherapy for dissociative amnesia other than pharmacologically facilitated interviews. A variety of agents have been used for this purpose, including sodium amobarbital, thiopental (Pentothal), oral benzodiazepines, and amphetamines. To date, no controlled studies have assessed the efficacy of any of these agents in comparison with one another or with other treatment methods. A single placebo-controlled study of a barbiturate-facilitated interview did not find superiority of sodium amobarbital over placebo in producing more clinically useful information. In more systematic World War II studies of barbiturate-facilitated interviews for amnesia and conversion disorder, this treatment was described as producing more rapid recovery, especially for amnesia, although little overall difference was found in recovery in comparison to subjects treated with psychotherapy alone or with hypnotherapy.

Amobarbital narcosynthesis is a term devised to underscore the need for the patient to process material uncovered in a pharmacologically facilitated interview in their usual conscious state. Pharmacologically facilitated interviews are used primarily with acute amnesias and conversion disorder, among other indications, in general hospital medical and psychiatric services. This procedure is also occasionally useful in refractory cases of chronic dissociative amnesia when patients are unresponsive to other interventions. The Joint Commission now considers pharmacologically facilitated interviews to be conscious sedation, requiring the presence of an anesthesiologist. This procedure must be performed in settings where resuscitation equipment is available in case of respiratory arrest, a possible, albeit rare, complication. The interview usually is audiotaped or videotaped to replay for the patient, because amnesia generally is present for the material produced. Although in some cases repeated procedures may be helpful, in others, repeated procedures have led to the patient's dependence on pharmacologically facilitated interviews.

The current controversies over delayed recall for trauma have also focused on the use of pharmacologically facilitated interviews for patients with reported amnesia for childhood maltreatment. No modern systematic data exist on memory fallibility or accuracy in individuals undergoing pharmacologically facilitated interviews. As discussed previously, pharmacologically facilitated interviews were used extensively in World

War II for the treatment of combat-related dissociation. Although wartime memories recovered with pharmacologically facilitated interviews were generally considered to be accurate, some subjects were reported to dissemble completely or withhold crucial information, or both, despite barbiturate treatment.

Given the current controversies, the clinician should obtain informed consent focusing on the nature of memory for pharmacologically facilitated interviews for amnesia as for patients considering hypnosis. The clinician should emphasize that these drugs are not a truth serum; whatever apparently new information emerges under the drug condition should be regarded no differently with respect to accuracy than any other material that emerges over the course of treatment.

Group Psychotherapy for Amnesia. During World War II, group psychotherapy and group hypnotherapy were among the treatments used to promote recovery in traumatic war-related amnesia. Highly supportive, structured, reassuring, and re-educative approaches were often used by therapists in such cases to accomplish return of the patient to functional status and to prevent chronic disability.

Time-limited and longer-term group psychotherapies have been reported to be helpful for combat veterans with PTSD and for survivors of childhood abuse. During group sessions, some authors report that patients recall memories for which they have had amnesia. Supportive interventions by the group members or the group therapist, or both, may facilitate integration and mastery of the dissociated material. On the other hand, concern has been raised about memory contamination in such group therapy settings, as it has for patients' involvement in 12-step or self-help groups for trauma survivors.

TREATMENT OF DISSOCIATIVE FUGUE SUBTYPE. Treatment for dissociative fugue follows the basic structure described above for treatment of dissociative amnesia. Once stabilization is achieved, subsequent therapy is focused on helping the patient regain memory for identity, life circumstances, and personal history. Some patients with dissociative fugue may resist uncovering their actual identity even with hypnosis or pharmacologically assisted interviews. Appeals through the local (sometimes regional, national, or international) media may not alert the patient's concerned others. In one case, a patient was asked to randomly select numbers on a phone keypad. The patient tapped a phone number—

apparently outside of conscious awareness—that allowed the treating clinician to find the patient’s family hundreds of miles away.

Family, sexual, occupational, or legal problems that were part of the original matrix that generated the fugue episode may be substantially exacerbated by the time the patient’s original identity and life situation are detected. Thus, family treatment, legal, and social service assistance may be necessary to help resolve such complex difficulties. Also, the patient may display other extreme or bizarre dissociative symptoms, such as the Ganser symptom of approximate answers, or attempt to engage in another fugue to escape the situation.

DEPERSONALIZATION/DEREALIZATION DISORDER

As described in the section on Nosology, the DSM-IV-TR diagnosis of depersonalization disorder was renamed depersonalization/derealization disorder in DSM-5. This was based on a review that found no differences between patients with predominant depersonalization and those with predominant derealization, confirming the ICD-10/11 conceptualization of either symptom as characteristic of the disorder.

For many years, the ubiquity of depersonalization/derealization as a psychiatric symptom obscured its broader recognition as a disorder. Consequently, patients with chronic and primary depersonalization/derealization are commonly misdiagnosed as having solely mood or anxiety disorders. These disorders may be comorbid with chronic depersonalization/derealization and/or may develop in response to untreated depersonalization/derealization disorder symptoms. Recent systematic research by Daphne Simeon and Anthony David on two large cohorts of well-characterized patients with depersonalization/derealization disorder yielded strikingly consistent findings in the United States and Europe. Clinical features, course, prognosis, and cognitive and neurobiologic correlates clearly distinguish this disorder from other psychiatric disorders in which depersonalization/derealization symptoms occur. Also, valid and reliable instruments exist for screening and diagnosis, such as the Cambridge Depersonalization/Derealization Scale developed by Mauricio Sierra and Germán Berrios (see section on Nosology and Table 20–2).

Etiology

Discrete Behavioral States Theory and Depersonalization/Derealization Disorder. Individuals with depersonalization/derealization disorder describe highly distressing, continuous or relapsing/remitting DBSs characterized by alienation from somatic, perceptual, cognitive, and affective processes. Unlike individuals with psychotic or delusional disorders, reality testing remains intact: for example, when queried these individuals acknowledge that they "feel" terrifyingly dead but they do not literally believe that they are dead. Individuals with the disorder usually have limited ability to modulate or moderate these states. Some describe states characterized by numbness and detachment. Typically, these individuals' emotionless, robotic outward presentation belies their substantial subjective distress at experiencing chronically anhedonic states of being. Conversely, many individuals with depersonalization/derealization disorder present with states of overt agitation and desperation about experiencing disconnection from themselves and/or the world.

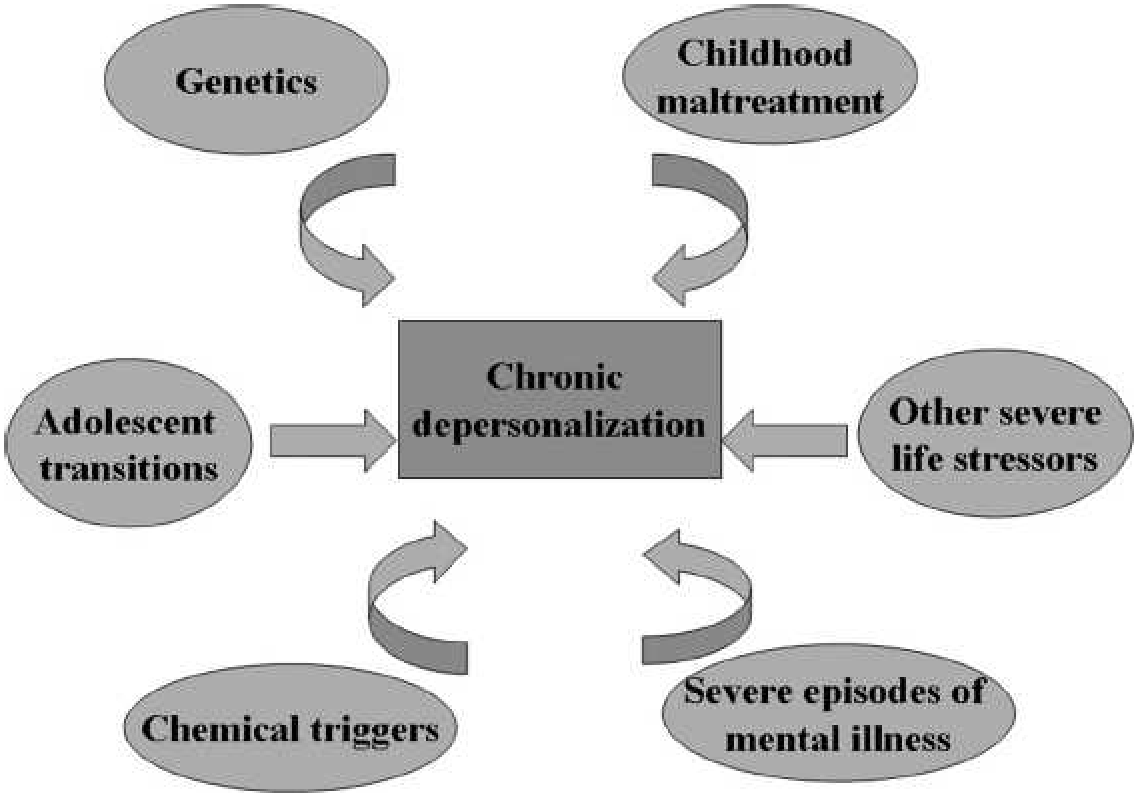
Genetics
Childhood maltreatment
Adolescent transitions
Chronic depersonalization
Other severe life stressors
Chemical triggers
Severe episodes of mental illness

FIGURE 20–2. Schematic integration of proposed models to explain depersonalization disorder.

Several models have been proposed to explain this disorder. A schematic integration is presented in Figure 20–2.

Psychodynamic. Traditional psychodynamic formulations have viewed depersonalization/derealization as a disintegration of the ego or as an response in defense of the ego. These explanations stress the role of overwhelming painful experiences or conflictual impulses as triggering events. The high rates of depersonalization/derealization symptoms in normal adolescents and in patients conceptualized as having borderline or narcissistic personality organizations are cited as evidence that ego immaturity, ego deficits, and/or self-self-system lability and/or deficits predispose to these conditions. More recently, attention has been drawn to the similarities between depersonalization/derealization and obsessive-compulsive symptoms. Patients with depersonalization/derealization disorder often display obsessive-like behaviors with respect to their symptoms. The split between an observing and a participating self is likened to the division between intellect and emotion in obsessive patients. Both groups respond to serotonin reuptake inhibitors, although the therapeutic response for patients with depersonalization/derealization disorder is usually much less robust.

Traumatic Stress. As described in prior sections, acute peritraumatic depersonalization/derealization states are common in accident victims and military recruits undergoing high-stress training, permitting peritraumatic, survival-based downregulation of autonomic functions (e.g., heart rate, blood pressure, rate of respiration, and of the HPA axis). In depersonalization/derealization disorder these states can be conceptualized as disconnecting from their earlier survival function and becoming persistently autonomous, activated by later stressful and/or traumatic experiences (e.g., financial, social, medical, psychopharmacological, and/or other stresses).

Other antecedent experiences to the development of depersonalization/derealization disorder include the unexpected, sudden death of a family member or close friend; sexual orientation struggles in adolescence; and major family/work stresses. In some cases of severe and/or prolonged mood or anxiety disorders with comorbid DSM-5-diagnosed depersonalization/derealization disorder, this latter disorder becomes chronic after remission of the other disorders.

A controlled clinical study found significantly more childhood trauma, especially emotional abuse perpetrated by parents, in well-diagnosed patients with depersonalization/derealization disorder compared with healthy controls. Unlike the control group, separation or loss of parents appeared protective of development of depersonalization/derealization symptoms. Overall dissociation scores were related to sexual and emotional abuse in the control group, but emotional abuse alone predicted depersonalization/derealization scores.

In support of the power of emotional abuse as a traumatic stressor and of its link to dissociation, Martin H. Teicher and colleagues have shown that parental verbal/emotional abuse was “a potent form of maltreatment” in a community sample of 554 young adults. In this study, verbal/emotional abuse experiences resulted in moderate to large increases in measures of dissociation, anxiety, depression, anger-hostility, somatoform symptoms, and symptoms of “limbic irritability,” such as paroxysmal somatic disturbances, brief hallucinatory events, automatisms, and dissociative experiences.

In a recent report, a Puerto Rican convenience sample of 40 women with a history of interpersonal trauma was compared with 40 control women without such history. Subjects completed Spanish versions of the DES, a factor-analytic derived depersonalization/derealization subscale of the DES (DES-D), the CDS, the Patient Health Questionnaire (PHQ-9), and a validated measure of types and severity of prior physical, sexual, and/or emotional abuse. Half the trauma sample reported a history of sexual abuse. Participants reporting more frequent abuse experiences also had significantly higher mean depersonalization/derealization scores on the DES, DES-D, and CDS. The largest effect sizes were for CDS and depression scores. One-quarter of the trauma group scored above the CDs cutoff score of ≥70, indicating high likelihood of a depersonalization/derealization disorder diagnosis. As in other studies (see section on Epidemiology), there was a significant correlation between PHQ-9 depressive and depersonalization/derealization scores. The authors concluded that, despite limitations of this convenience-sample study, the findings provide evidence of a significant linkage between antecedent trauma and dissociative symptoms in a Latin American cultural context.

In a recent study of 106 women with PTSD (80% PTSD dissociative subtype) in a specialized partial hospital trauma program were studied using a multivariate linear regression model of multiple childhood adversities. These included physical abuse, emotional abuse, sexual abuse, emotional neglect, and physical neglect. In this sample, emotional abuse was significantly associated with higher depersonalization/derealization scores and physical abuse was associated with lower dissociation scores. There was no interactive effect of multiple types of abuse.

However, some studies have not shown significantly higher rates of childhood interpersonal trauma in individuals with depersonalization/derealization disorder compared to control groups (e.g., with depressive disorders). Some of these studies have found significantly higher rates of a first-degree relative with severe psychiatric illness (e.g., psychotic disorder or bipolar disorder). One hypothesis is that the chronic fear engendered by the relationship with the psychotic relative was etiologic in the development of the depersonalization/derealization disorder. For example, one patient reported that throughout her childhood her father and older brother left her to manage on her own her mother's psychotic and violent episodes. Despite intense terror and dread about her own safety, the patient feared more for her mother's safety. She followed her mother through the house while the mother typically brandished a large kitchen knife, loudly verbigerating at invisible persecutors.

As in the case of individuals with dissociative amnesia, an earlier history of trauma and/or severe stress may be present but is not the initiating factor in the onset of depersonalization/ derealization. Symptom exacerbations are commonly related to negative emotions, high levels of sensory input, and threatening, stressful, and/or unfamiliar situations. Traumatic or stressful events may precipitate or exacerbate depersonalization/derealization disorder symptoms.

Many individuals with depersonalization/derealization disorder do not identify an immediate traumatic antecedent and do not report a clear precipitant to disorder onset. On the other hand, nontraumatic stressors—such as severe interpersonal, financial, or occupational losses—have been associated with the onset or exacerbation of the disorder. Psychopharmacological stressors, most commonly marijuana and hallucinogens, can sometimes precipitate chronic depersonalization/derealization. These individuals can be conceptualized as having a neurobiologic or genetic vulnerability to chronic depersonalization/derealization after drug use, although early life trauma can also play a role.

Mr. A, a successful 42-year-old actor, was referred for psychiatric evaluation. Since his late teens, he had abused alcohol and a variety of drugs, usually in binges lasting several weeks. After one of his brothers died of a heroin overdose, he engaged in a several-week binge on alcohol, marijuana, and inhaled and smoked cocaine. At the conclusion of the binge, after his first sober night "crash," he awakened with intense, highly distressing depersonalization/derealization experiences. He reported feeling completely alien and disconnected from himself, his emotions, and his awareness of the environment. He reported estrangement from his body, not experiencing it as connected to him, feeling as if his limbs were either too far away from him or too close. He reported that the world seemed flat and unreal, almost as if it were a painting. He tried to read scripts that he had been given to prepare for auditions but the words seemed disconnected from him. When he tried reading scripts aloud, his voice sounded alien, "tinny," far away, and unfamiliar. He denied being fully out of his body, observing himself as if he were another person. However, he reported being "just beside myself, not really in myself, like to one side. It's terrible; I can't get back to me. I don't feel connected to who I am. I know I am feeling things, but I don't know what they are. All I feel is desperate to get back to how I used to be."

He described his father as a dictatorial alcoholic who berated him throughout his youth, calling him names, humiliating, and mocking him. He denied a history of childhood sexual abuse but said that both parents would hit him "randomly, with whatever they got their hands on." The father terrorized the family, including repeatedly insulting and hitting the patient's mother in front of the patient and his younger brother. Substance abuse began in adolescence to numb himself from feeling terrified, impotently enraged, and humiliated. He described substance abuse, depression, and suicide attempts in first-degree relatives, including his deceased brother, no family members had sought mental health treatment and it was stigmatized in his family. A complete physical, laboratory, and neurologic work-up were all were within normal limits. Due to the symptoms, he stopped using alcohol and drugs completely, literally becoming phobic of drug use. However, several years later, he remained chronically depersonalized and only able to work in a very limited way. He reported distress that he cannot experience emotions about his brother's death or the family violence and dysfunction that led both of them to abuse substances with such adverse outcomes. He had little response to trials of antidepressants, mood stabilizers, and naltrexone. He reported mild improvement in mood and ruminations about his symptoms on fluvoxamine 300 mg daily and naltrexone 100 mg daily. Primarily, he was managed in a supportive psychotherapy, having had limited response to trials of cognitive-behavioral, dialectical behavioral, and psychodynamic psychotherapy.

Cognitive Theories. In the last decade, increasing attention has been paid to cognitive and behavioral aspects of chronic depersonalization/derealization, essentially positing that an initial, relatively benign, and possibly transient dissociative response is reinforced, maintained, and exacerbated by a vicious cycle of dysfunctional cognitions and behaviors. E.C.M. Hunter and colleagues in the United Kingdom have put forth such a cognitive-behavioral model, proposing that an initial trigger (e.g., trauma, anxiety, depression, stress, fatigue, intoxication) can induce transient symptoms of depersonalization/derealization which are subsequently cognitively processed by an individual as either situational or catastrophic. If the attributions are situational, and therefore more benign, the depersonalization/derealization symptoms will tend to fade as the situational factors alleviate. However, if the attributions are catastrophic, they may evoke overwhelming fears such as going mad, losing control, becoming invisible, or having a permanent brain dysfunction. In turn, these fears may lead to increased anxiety coupled with paradoxically decreased arousal, resulting in increased intensity of the depersonalization/derealization symptoms as the individual enters the maintenance phase. During this phase, individuals may start to avoid situations that they associate with symptom provocation; become preoccupied with safety behaviors (such as acting “normal”); develop cognitive biases resulting in symptom over-monitoring; and experience a reduced threshold for the perception of threat. Over time, these maintenance factors are thought to perpetuate or exacerbate the symptoms.

Diagnosis and Clinical Features

In clinical samples, the disorder is equally common in men and women but is often missed. Under-diagnosis is facilitated by several aspects of the disorder. Patients experiencing depersonalization/derealization symptoms have difficulty putting their ineffable experiences into words and may have problems with alexithymia. They often fear that they may be seen as “crazy” if they **do** describe their symptoms and may be reluctant to speak about them. They may describe their severe suffering in ways that do not adequately convey their distress to an examiner; for example, repeatedly stating that, “I feel dead,” “Nothing seems real,” or “I’m standing outside of myself.” Individuals with depersonalization/derealization disorder typically endure an intensely unpleasant, and often disabling, subjective experience. Many say that they would gladly exchange the depersonalization/derealization for physical pain that would at least reconnect them with their body. Others say they do not contemplate dying because they already are dead. Some patients may appear remarkably detached and unemotional, due to profound dissociation of affects. As the patient’s complaints do not match their affective presentation, clinicians may not take the severity of this disorder as seriously as they should.

There are several distinct components to experiences of depersonalization and derealization. Individuals may be reluctant or have difficulty describing their symptoms, and clinicians should familiarize themselves with the range of depersonalization/derealization symptoms for direct inquiry. *Depersonalization* may include a sense of bodily changes, duality of self as observer and actor, being cut off from others, and being cut off from one's own emotions. These individuals may suffer from teleopsia (objects appear further away), pelopsia (objects appear nearer than they are), metamorphopsia, micropsia, and/or macropsia. They may experience their body changing in size and shape; disintegration of time sense with little ability to assess how recently or remotely events occurred; *déjà vu*, *déjà vécu*, and *jamais vu*; numbing of emotions and perceptions, including the inability to perceive hunger or thirst; and feeling robotic, detached, and/or without thoughts or feelings. The individual may report out-of-body experiences or, as in the case example above, a highly dysphoric feeling of partial disconnection from oneself. On the other hand, in *derealization,* the world appears strange, foreign, or dreamlike. Objects may appear as if viewed from a great distance and as if they were two-dimensional, without depth or substance, or as if observed through a fog, mist, or hazy, frosted glass. Sounds come from a distance, muffled and distorted. Objects feel strange to the touch. Colors dim and lose their vitality. The faces of others change, becoming unfamiliar and frightening.

Two factor analytic studies of core symptoms of the disorder found convergent results. There were five discrete symptom clusters: unreality of self, defined as feeling detached from one's physical body, mind, and thoughts; perceptual alterations, encompassing visual, tactile, and somatosensory distortions; emotional numbing, characterized by blunted affect, detachment from pain, and abulia; anomalous subjective recall, consisting of disrupted experiences of memory and time; and alienation from surroundings, comprising symptoms of derealization.

A latent profile analysis of 303 individuals (57.4% male) with depersonalization/ derealization disorder at a British clinic for the disorder were assessed with the DES, CDS, Beck Anxiety Inventory (BAI), and questionnaires for past medical and psychiatric history, symptoms of depersonalization and derealization, triggers of episodes, and other aspects of the disorder.

The latent profile analysis resulted in five subgroups: three varied on low, moderate, or high symptom severity across measures. Two other subgroups diverged on dissociative symptoms: a high DES dissociation group with amnesia, absorption, and depersonalization/derealization; and a high depersonalization/derealization group with higher CDS scores, but lower DES scores, particularly amnesia and absorption. The authors did not assess trauma history or conduct diagnostic assessments for other dissociative disorders (e.g., MID, SCID-D, MDI) or for comorbid diagnoses (e.g., SCID-5 to rule out schizophrenia). They suggested that the high-dissociation group may be better characterized by a diagnosis of dissociative identity disorder or of dissociative amnesia as conditions "comorbid" with depersonalization/derealization disorder. They speculated that the high-severity group may include subjects with schizophrenia. The elevated DES absorption factor in the high-dissociation group suggests the high hypnotizability found in dissociative identity disorder and dissociative amnesia. Preliminary studies of hypnotizability in depersonalization/derealization disorder do not show elevated hypnotic capacity on standardized measures. Also, in terms of DSM-5-TR diagnostic rules, diagnoses of dissociative identity disorder or schizophrenia would supersede that of depersonalization/derealization disorder (see below in section on Differential Diagnosis).

Cross-Cultural Aspects of Depersonalization/Derealization Disorder

In the 1990s, anthropologist Richard Castillo described a set of depersonalization/derealization symptoms derived from research in multi-cultural populations that was intended to broaden the cross-cultural validity of depersonalization/derealization disorder. Castillo's cluster of symptoms is similar to the more recent data-derived factor analytic models in Western samples, suggesting that the original "Western" descriptions may have been based on a limited sample of individuals. Components of Castillo's description that coincide with these contemporary descriptions include: persistent or recurrent experiences of depersonalization and/or derealization in which familiar experiences of the self and environment become distorted, potentially resulting in physical and emotional numbing; alterations in body image; experiences of being an outside observer of one's mental processes or body; feeling like an automaton; time disturbances; feeling as if one's body or environment is "unreal" or dreamlike; and visual distortions (e.g., macropsia).

Relevant to diagnosis, the depersonalization/derealization experience must not occur exclusively during a culturally approved religious, spiritual, or ritual activity (e.g., meditative practice) and is not considered normal by the cultural standards of the patient's own social group(s). However, some individuals may initially induce these experiences intentionally but over time lose control over them and develop fear and aversion for related practices.

Differential Diagnosis

The variety of conditions comorbid with depersonalization/derealization disorder complicates the differential diagnosis. These include major depressive disorder (73%); any anxiety disorder (64%); obsessive-compulsive disorder (21%); avoidant personality disorder (23%); and borderline personality disorder (21%). Symptoms of depersonalization/derealization may be associated with PTSD, ASD, bipolar disorder, schizophrenia, dissociative identity disorder, and dissociative amnesia, among others. Depersonalization/derealization symptoms may also result from a medical or neurologic condition, particularly seizure disorders, intoxication or withdrawal from illicit drugs, or as a side effect of medications (e.g., over-the-counter cold medications containing dextromethorphan; prescription or over-the-counter medications with anticholinergic potency). A thorough medical and neurologic assessment is not routinely needed, except for patients whose symptoms or course appear atypical, or if there is reasonable suspicion or evidence of medical or neurologic disorders or toxicity. Expert consensus suggests consideration of medical, neurological, and toxicologic evaluation when the depersonalization/derealization begins after age 40; there are focal neurologic findings; symptoms of numbness or tingling in extremities or sudden onset of symptoms of head pressure; history of head injury, seizure, or brain disorder; or infectious diseases or other medical conditions that can affect the CNS (e.g., systemic lupus, Lyme). Other medical conditions that can generate depersonalization/derealization include migraine, sleep apnea, vertigo, Ménière disease, metabolic diseases, and delirium. In general, depersonalization/derealization symptoms with medical etiology have a predominantly sensory character without the elaborated descriptions and personalized meanings common to psychiatric etiologies.

Work up should include standard laboratory tests, specialized labs (e.g., for autoimmune disorders), and urine and/or serum drug screens. Sleep-deprived EEG, and sometimes ambulatory or inpatient telemetric EEG, may be necessary to rule out focal seizures with impaired awareness (partial complex seizures). Brain imaging may be necessary to assess for structural abnormalities or other processes, as well as specialized assessments (e.g., vestibular testing, polysomnogram).

Drug-related depersonalization/derealization is typically transient, but persistent depersonalization/derealization may follow an episode of intoxication with a variety of substances, notably marijuana, hallucinogens, MDMA (Ecstasy), ketamine, and more recently salvia.[1] When depersonalization/derealization disorder is precipitated by drug use, there is no difference in the clinical course from depersonalization/derealization disorder without this precipitant.

To establish a differential diagnosis, a thorough and detailed psychiatric history needs to be obtained, especially in the presence of comorbid mood and anxiety disorders. Current diagnostic criteria require that the chronic depersonalization/derealization not occur exclusively in the presence of other disorders. It is common to obtain a history whereby an earlier and frightening episode of severe anxiety or depression—which conceivably may have acted as an internal stressor and destabilized a person’s sense of self—may precipitate chronic depersonalization/derealization. Subsequently, the latter continues even after the initial anxiety or mood episode remits. In another common scenario, depersonalization/derealization disorder tends to be more refractory, often continuous for periods of many years or decades, while the comorbid disorders go into periods of remission, whether with treatment or spontaneously. Conversely, patients who are severely and chronically ill with depersonalization/derealization disorder may become unremittingly depressed in response to their primary dissociative disorder, leading to misdiagnosis of a primary affective disorder. As discussed above, significant depersonalization/derealization symptoms may be found in dissociative identity disorder, often with CDS total scores several times higher than the cutoff score of ≥70 (e.g., 150 to 220) considered diagnostic for depersonalization/derealization disorder. Individuals with generalized dissociative amnesia, as well as patients with the childhood autobiographical memory deficits (CAMD)-form of dissociative amnesia, may experience depersonalization/derealization as a manifestation of the dissociative disorder and/or other comorbidities (e.g., PTSD, depression, obsessive-compulsive disorder, conversion disorder). from other dissociative disorders.

As a defining differential-diagnostic point to discriminate depersonalization/derealization disorder from psychotic disorder, patients with the former condition generally have intact reality testing about perceptual alterations. Also, the presence of a psychotic thought process, overvalued ideas, ideas of influence/reference, and Schneiderian symptoms distinguish depersonalization/derealization disorder from psychotic spectrum disorders such as schizophrenia, prodromal schizophrenia, or schizotypal personality disorder. The diagnosis of depersonalization/derealization disorder is based on the “as if” experiential nature of depersonalization/derealization, without distorted, delusional, dereistic cognitive explanations. This is similar to the assessment of a psychotic process accompanying obsessive-compulsive symptoms. A study comparing perceptual alterations and magical ideation in individuals with depersonalization/derealization disorder and healthy volunteers found that people with the disorder only had elevated perceptual, but not cognitive, distortions.

Course and Prognosis

Depersonalization or derealization symptoms after traumatic experiences or intoxications commonly remit spontaneously. Depersonalization/derealization accompanying mood, psychotic, or other anxiety disorders commonly remit only with definitive treatment of these conditions.

One-third of individuals with depersonalization/derealization disorder have an episodic, relapsing, and remitting course; one-third a chronic course; and one-third an episodic course that becomes chronic. Many patients with chronic depersonalization/derealization disorder are characterized by severe impairment in occupational, social, and personal functioning. In most studies, mean age of onset appears be in late adolescence or early adulthood. Most patients with depersonalization/derealization disorder are initially treated for the secondary anxiety and mood disorder symptoms. The primary nature of the depersonalization/derealization disorder is often only recognized later, if at all. These individuals, like those with dissociative identity disorder, may spend many unproductive years in the mental health system before receiving a correct diagnosis, “failing” to respond to standard treatments and often characterized as “treatment resistant.” As in the case of individuals with dissociative identity disorder, expert clinical opinion suggests that persons with untreated depersonalization/derealization disorder may, after years of unproductive treatment, suffer demoralization and clinical deterioration. This may lead to a chronic and disabling course with impairment in family, social, and occupational functioning.

Treatment

Clinicians working with patients with depersonalization/derealization report them to be singularly refractory to treatment. Several factors may lead to this perception. Due to the typically substantial delay in diagnosis of the condition, clinicians often do not have the opportunity to intervene until years after its onset, at which point the symptoms tend to become constant, nonfluctuating, and less responsive to therapeutic intervention. Clinical reports suggest that individuals with depersonalization/derealization disorder seen earlier during their disorder, and those who continue to have fluctuating or episodic symptoms over the years, may be more responsive to treatment.

Psychotherapy. Many studies describe a range of psychotherapies to treat depersonalization/derealization disorder, including psychodynamic, cognitive, cognitive-behavioral, hypnotherapeutic, and supportive therapies. No systematic data exist that compare these modalities. There are few systematic studies of psychotherapy for depersonalization/derealization disorder.

Psychodynamic treatment focuses on underlying threats to self-constancy that give rise to intolerable affective states. According to psychoanalytic theories, a person in whom the cohesiveness or stability of self-representations is profoundly threatened may resort to depersonalization/derealization (i.e., a disconnection from the self) as a solution. Similarly, psychodynamic theories suggest that depersonalization/derealization can be linked to various levels of personality organization. In individuals with a psychotic-spectrum disorder, depersonalization/derealization may be triggered by impaired self-other differentiation. In borderline psychopathology, unstable and shifting self-representations may be associated with depersonalization/derealization. In narcissistic pathology, when self-constancy is threatened by the real or imagined loss of self-objects serving purposes of object constancy, depersonalization/derealization may arise. Finally, in neurotic psychopathology, de-repressed self-representations associated with overwhelming intrapsychic conflict may trigger depersonalization/derealization. In psychodynamic psychotherapy, the therapist can observe and analyze, often in a moment-to-moment fashion, these and related dynamics as they occur during psychotherapy sessions and/or are described by the patient.

More recently, Daphne Simeon has proposed that addressing the fear of affects is a crucial component of psychotherapy for depersonalization/derealization disorder. Dissociation of affect (“not feeling”) and alexithymia (difficulty identifying and verbalizing internal emotional states) are core components of the condition. Depersonalized/derealized patients will frequently recount traumatic or stressful life experiences in a highly robotic, matter-of-fact, and emotionless way. The goal of an affect-based therapy is to uncover, experience, label, own, and verbalize intolerable emotions and to process these emotions in the context of underlying conflicts and disavowed self-representations and self-states. Eventually, such emotions can be gradually integrated with the core sense of self so that the individual can transition from an “unreal” self to a more “real” owned self. Which affects need to be dealt with vary greatly depending on each person’s history and sense of self, ranging from the negative to the positive, and can include anger, grief, sorrow, shame, guilt, excitement, and love. In a simplistic sense, this approach is “opposite” to the affect regulation component of the treatment of trauma-related conditions such as PTSD or dissociative identity disorder, as the goal in the therapy of individuals with depersonalization/derealization disorder is to activate and experience rather than to better modulate and regulate intense affects. On the other hand, as noted in the section on dissociative identity disorder, patients with dissociative identity disorder and the dissociative subtype of PTSD may attempt to mobilize depersonalization/derealization, affect suppression, and state shifting to distance and modulate posttraumatic intrusions. Frequently, in patients with this PTSD subtype, this effort at dissociative distancing fails and patients experience severe flashbacks. They may report, and/or display, oscillations between being all “shut down” and experiencing overwhelming flashbacks that can completely overwhelm awareness of current reality when they do break through into consciousness.

In recent years, a disorder-specific cognitive-behavioral therapy approach to depersonalization/derealization disorder has been developed and tested, based on the cognitive-behavioral model outlined earlier. In this type of treatment, the focus is on developing techniques to cognitively reinterpret symptoms as less threatening; to reduce symptom monitoring; and to diminish avoidance and safety behaviors. In an uncontrolled, prospective treatment trial of 21 patients utilizing this approach, significant reductions in ratings of depersonalization/derealization, depression, and anxiety were found. In addition, by the end of the 12-week treatment, 29% of participants no longer met diagnostic criteria for the disorder. Recently, a small, uncontrolled study of a staged psychoeducational, mindfulness, and CBT group therapy intervention for adolescents with depersonalization/derealization disorder found that, at 6 months follow-up, about two-thirds of the patients decreased CDS scores below cutoff criteria.

A proportion of severely impaired patients with depersonalization/derealization disorder may require long-term supportive treatment. The clinician will need to be acutely aware of the patient's interpersonal sensitivity, distress, and sense of hopelessness about the condition. Some supportive psychotherapy elements are very important for the effective treatment of all patients with the disorder. These include psychoeducation, such as giving the condition a name and reassuring the patient that it will not evolve into a psychotic disorder and that there is no evidence of permanent brain damage. Also, psychoeducation and supportive interventions can help the patient lift the barrier of secrecy based on shame about symptoms in general, alleviate guilt or shame associated with a drug-induced onset, and provide hope for the future. In this educational/supportive/cognitive-behavioral model, the clinician challenges the patient's experience of the "physicality" of the experience. This can include emphasizing that depersonalization/derealization disorder is a psychological disorder, as physical as it may feel; providing a sense of control over the symptoms that may be perceived as continuous in intensity; using techniques that can help modulate the level of arousal and indirectly soften the symptoms; and reassuring women who are considering pregnancy that the heritability, as far as it is known, appears to be limited. There are no adequate data on the utility of either hypnosis or EMDR in treating chronic depersonalization/derealization. One study found that these patients had relatively low hypnotic capacity on standardized hypnotizability measures, suggesting that this may not be a helpful intervention for patients with depersonalization/derealization disorder.

Pharmacotherapy. Extensive retrospective treatment histories have not provided support for major symptom improvement with any of the known and widely used classes of psychiatric medications, although benzodiazepines, serotonin reuptake inhibitors, and stimulants appear to be partially efficacious for some patients. In past decades typical antipsychotics were widely used for chronic depersonalization/derealization, often as first-line treatment, as many clinicians viewed depersonalization/derealization symptoms as indicative of a psychotic-spectrum condition. There is no empirical evidence for the efficacy of typical or atypical antipsychotic medications in depersonalization/derealization disorder. They should only be used to target overwhelming anxiety or mood disturbance associated with the depersonalization/derealization. Common side effects of these medications may *exacerbate* an experience of emotional deadness and lack of emotional response to oneself or the world. Thus, in some individuals with depersonalization/derealization disorder these medications may *increase* the target symptoms. Typically, the numbing/deadening neuroleptic side effects set in gradually as dosage increases. Thus, neither patient nor clinician appreciates the source of symptoms worsening—sometimes leading to even higher neuroleptic dosage.

There is some anecdotal support for the use of serotonergic antidepressants (e.g., SSRIs, clomipramine [Anafranil]) in depersonalization/derealization disorder. However, a recent large randomized, controlled trial of fluoxetine found only a statistically marginal, and clinically minimal, improvement in depersonalization/derealization compared to placebo. Depersonalization/derealization and dissociation scores did not significantly decline in the fluoxetine-treated group, paralleling patients' subjective reports that their symptoms had not changed. On the other hand, patients reported feeling somewhat better and less distressed by the symptoms. Within the fluoxetine-treated group, depersonalization/derealization responders were more likely than nonresponders to show improvement in comorbid anxiety disorder symptoms. This may account for the experience of being less distressed by their symptoms, although the depersonalization/derealization symptoms remained unchanged.

Therefore, it appears that serotonergic antidepressants are unlikely to be of major benefit for depersonalization/derealization symptoms but may potentially ameliorate comorbid obsessive-compulsive, depressive, and/or anxiety symptoms. Lamotrigine—a mood-stabilizing anticonvulsant that promotes glutamate release—initially showed promise for depersonalization/derealization disorder symptoms in a small open-label study. However, a subsequent small crossover placebo-controlled trial found that all nine participants were nonresponders. Some have suggested that the combination of SRIs and lamotrigine may have additive benefit in treating the disorder, but no empirical support exists for this hypothesis.

Opioid antagonists such as naltrexone (Revia), cognitive enhancers, benzodiazepines, and stimulants all have had some benefit in clinical treatment of specific patients with depersonalization/derealization disorder. In uncontrolled and/or proof-of-concept studies, a trial of repetitive transcranial magnetic stimulation (rTMS) for depersonalization/derealization disorder has shown efficacy in some patients. rTMS models have involved stimulation of the TPJ, and the VLPFC, areas that have been implicated in the generation of depersonalization/derealization disorder symptoms.

DEPERSONALIZATION/DEREALIZATION DISORDER (DDPD) CASE 2

Ms. R a 27-year-old unmarried graduate student with a master's degree in biology came to clinical attention for intermittent episodes of "standing back," usually associated with anxiety-provoking social situations. When asked about a recent episode, she described presenting a paper in a seminar course. "All of a sudden, I was talking, but it didn't feel like it was me talking. It was very disconcerting. I had this feeling, 'Who's doing the talking?' I felt like I was just watching. Watching someone else talk. Listening to words come out of my mouth, but I wasn't saying them. It wasn't me. It went on for a while. I was calm, even sort of peaceful. It was as if I was very far away. In the back of the room somewhere—just watching myself. But the person talking didn't even seem like me really. It was like I was watching someone else." The feeling lasted the rest of that day and persisted into the next as it gradually dissipated.

She thought that she remembered having similar experiences during high school but was certain that they occurred at least once a year during college and graduate school. Although she said that she usually felt detached and sometimes peaceful during the experience, she was upset at the thought that she would likely have more episodes. She complained that the sudden onset, the eeriness of seemingly watching an almost unrecognizable version of herself from a distance, and the sense of not "being in the world" were almost unbearable in retrospect.

Ms. R, an only child, reported recurrent, intense childhood fear and distress when overhearing or witnessing the frequent violent arguments and periodic physical fights between her parents. She remembered lying awake in bed, listening to them, imagining the terrible things that were occurring, or about to occur, between them. In addition, the family was subject to many unpredictable dislocations and moves owing to the patient's father's intermittent difficulties with finances and employment. The patient's anxieties did not abate when the parents divorced when she was a late adolescent. Her father moved away and had little further contact with her. Her relationship with her mother became increasingly angry, critical, and contentious. She was unsure whether she experienced depersonalization during childhood while listening to her parents' fights.

After receiving a depersonalization/derealization disorder diagnosis, she began a multimodal treatment to address symptoms. She became aware of more chronic depersonalization/derealization symptoms with detachment from her emotions, distorted experiences of her body, metamorphopsia, and feeling that she was permanently separated from other people, as if behind a glass wall. She found psychoeducation about her symptoms relieving, as she had a name and concepts to explain her experiences not, as she put it, "just plain crazy." Also, she was helped to look at the impact the family violence and dislocations had on her development. She recalled ongoing terror that her parents would move and that she would be left behind, although at other times she guiltily fantasized about, after abandonment, being adopted by a "nice" family. She changed schools often and felt "different" from other children, fearful to not invite anyone to her house as she could not predict what her parents would do.

She found ongoing psychoeducation validating, allowing her more freedom to look at the chronic terror that characterized her childhood. She realized the extent of emotional neglect she experienced, along with witnessing the family violence. She responded as well to psychodynamically informed interventions that helped her notice increasing depersonalization/derealization in the moment, particularly at times of impending strong emotions like fear, anger, and shame. Cognitive interventions helped her become aware of many distorted beliefs that she harbored, feeling responsible for her parents' well-being, blaming herself for her difficulties in development, as well as associated symptoms of depression. On diagnostic exploration, she described a variety of obsessive-compulsive symptoms since childhood, including magical thinking that they would protect her from the chronic danger she experienced at home. She had a partial response to fluvoxamine 200 mg daily with decreases in panic, obsession-compulsive, and depressive symptoms. As treatment progressed, she was amazed at how strangely good it felt to be aware of her emotions, perceptions, body, and mind, as well as experiencing joy when engaged in music and art and when hiking in the countryside. She began to develop a group of friends and became involved with her first boyfriend. She was amazed at how accepting they all were of her and that they too had issues with dysfunction they experienced in their families of origin.

OTHER SPECIFIED DISSOCIATIVE DISORDERS

The category *other specified dissociative disorders* describes conditions characterized by a primary dissociative response that do not meet diagnostic criteria for one of the DSM-5-TR dissociative disorders. This category encompasses a clinically heterogeneous group of dissociative conditions, often with multiple comorbidities, and provides a diagnostic home for several Western and non-Western idioms of distress. Many patients with these conditions have a current or past history of severe trauma. To receive a diagnosis of other specified dissociative disorder, the dissociative presentation must not be better accounted for by ASD, PTSD, specific dissociative disorders, or another disorder.

The other specified dissociative disorder diagnosis in DSM-5-TR lists four examples to assist clinicians in recognizing conditions encompassed by this category. Research findings may support the future inclusion of some or all these as specified dissociative disorders and/or lead to modifications of the diagnostic criteria for the current DSM-5-TR dissociative disorders. Future studies should obtain additional data on each example to determine whether it represents a stable disorder with a typical set of symptoms as well as on its etiology, precipitants, neurobiology, course, treatment response, and other outcomes. Some patients who initially receive this diagnosis will ultimately be better conceptualized as having another dissociative disorder, such as dissociative identity disorder.

EXAMPLE 1: ATYPICAL PRESENTATIONS NOT MEETING FULL DIAGNOSTIC CRITERIA FOR DISSOCIATIVE IDENTITY DISORDER

Definition

Example 1 of other specified dissociative disorder covers presentations similar to dissociative identity disorder but with: (a) less-than-marked discontinuities of self and agency and/or (b) alterations of identity or episodes of pathologic possession in an individual who reports no dissociative amnesia. This is an analog of the ICD-11 partial dissociative identity disorder diagnosis.

Diagnosis and Clinical Features

As noted in prior sections, a major goal of the DSM-5-TR Work Group was to reduce the frequency of DSM-IV-TR dissociative disorder NOS diagnoses. Nonetheless, Example 1 is meant to indicate cases where the experience of self-states is not sufficiently distinct and/or the discontinuities of self insufficiently apparent or severe to warrant a diagnosis of dissociative identity disorder (see section on Changes in Diagnostic Criteria to the Dissociative Disorders).

Example 1 can be used as an initial placeholder for a putative diagnosis of dissociative identity disorder. This allows additional history-taking, clinical observation, use of diagnostic screening inventories and interviews, and acquisition of data from collateral informants. Over time, with further assessment, most individuals with this diagnosis will likely meet criteria for dissociative identity disorder. Rarely, clinicians describe patients who fulfill the DSM-5-TR Criterion A for dissociative identity disorder, but in whom dissociative amnesia is not reported or observed, even on careful questioning. Clinical experience suggests that, with careful sequential observation, most of these individuals ultimately meet the amnesia criterion for dissociative identity disorder. Ultimately, the clinician usually: (a) detects subtle dissociative amnesia symptoms; (b) realizes that the patient manifests dissociative amnesia for episodes of dissociative amnesia; (c) witnesses the patient not recalling several minutes of a therapy session, with the patient unaware that this has occurred); and/or (d) discovers subtle dissociative amnesia for aspects of the life history but virtually no contemporary dissociative amnesia symptoms such as ongoing blackouts, fugues, and unexplained belongings. Some of these patients may report continuity of memory inconsistent with dissociative amnesia. However, on careful evaluation, they report that they experience substantial parts of their autobiographical memory as depersonalized and derealized, without an affective, first-hand component: as if it happened to someone else, they heard about it, read about it, and/or have been told by friends and families about past events. Ultimately, the clinician and patient may realize that depersonalized/ derealized memory conceals typical forms of dissociative amnesia. These patients may be quite distressed, shocked, ashamed, and scared to acknowledge an extensive autobiographical memory deficit. They can go into a crisis when they realize that they may not be able to count on their memory to recall what happened during important parts of their lives.

Some patients with dissociative identity disorder who have achieved partial, but not complete, unification of self-states may continue to display these states, albeit usually in much more attenuated form, and may have significantly improved memory continuity. As treatment progresses, they may no longer meet Criterion A for dissociative identity disorder. Patients with dissociative identity disorder who achieve complete unification do not manifest self-states or dissociative amnesia. Under significant stress or with later life trauma, however, some of these individuals may display reactivated self-states but usually without the distinctness that manifested during earlier phases of treatment. Also, they may report dissociative amnesia symptoms for the stressful events that led to the disruption in unification. Technically, some of these patients meet diagnostic criteria for other specified dissociative disorder. As noted in the section on etiology, there are no studies that examine the question of change in taxon membership when individuals with dissociative identity disorder achieve stable unification.

Treatment

Based on research on the DSM-IV-TR category of dissociative disorder NOS, most patients with Example 1 of other specified dissociative disorder will require a treatment approach similar to the one recommended for dissociative identity disorder.

COERCIVE PERSUASION

Definition, Diagnosis, and Clinical Features

The concept of *brainwashing* or *thought reform* emerged from Western reactions to the Soviet show trials of the late 1940s and alarm over Communist Chinese methods of coercive political re-education. These fears intensified during the Korean conflict after many American prisoners of war made anti-American statements. In 1950, a journalist, Edward Hunter (later identified as a Central Intelligence Agency [CIA] operative) proposed that the Chinese Communists had discovered techniques to modify mental attitudes and beliefs, a process that he called *brainwashing.*

After the Korean conflict Armistice, a team of psychiatrists and psychologists, including Robert J. Lifton and Edgar Schein, interviewed the returning prisoners. Similar CIA studies obtained information from Communist interrogators and former Soviet and Chinese prisoners. These studies concluded that the apparent brainwashing of these prisoners, as well as of the victims of the Russian show trials, was the result of police interrogation techniques including torture and severe physical and psychological deprivation. Cooperation by the prisoners was rewarded with the possibility of ending or pausing the torture or the provision of food and warm clothing. The reports concluded that the basic attitudes of the American prisoners in Korea had not been altered.

Subsequent investigations with other Korean conflict prisoners suggested that some of these individuals may have actually been subject to “thought reform” procedures that resulted in their continuing to falsely insist that they were spies. Studies by Schein and others concluded that changes in an individual’s basic beliefs could occur with torture, particularly when fostering prisoners’ dependency on the captors. The captors/torturers became the prisoner’s only human contact and used this form of coercion to control the prisoners’ behavior and sometimes their attitudes and views. Judith Herman, in her classic books on trauma, noted the similarity of the captor/captive relationship across various traumatic contexts: intra-familial child abuse, intimate partner violence, trafficking, and political torture.

The CIA and “Mind Control”. In the 1970s, the legacy of the fear of Cold-War-era “brainwashing” was revealed in declassified government documents and during Congressional hearings, which documented misuses of behavioral science by the CIA and other U.S. government agencies since the late 1940s. These efforts attempted to develop methods of behavioral control to combat and/or preempt potential Communist mind control “weapons.” The CIA studied or supported experiments in the United States and Canada on volunteers as well as unwitting or coerced subjects, including prisoners, psychiatric patients, and/or military personnel. These included the use of psychedelic drugs, marijuana, other toxins and intoxicants, radiation, hypnosis, sensory deprivation/overstimulation, repeated ECT treatments, and/or bizarre conditioning experiments. At least one individual killed himself during an LSD experiment. In the 1980s, a few older adults, permanently psychiatrically damaged U.S. and Canadian victims of CIA-funded “psychiatric” experiments during the 1960s reached a small monetary settlement with the U.S. and Canadian governments—neither of which admitted responsibility. Also, in anticipation of the congressional inquiries, the CIA leadership mandated destruction of most documents related to these “programs.”

According to reviews of remaining CIA documents, some CIA officials apparently believed that hypnosis could be used to create assassins or spies with hypnotically induced dissociated identity states and/or amnesia who could resist interrogation and torture. Other CIA officials doubted that this could be done successfully. However, due to the destruction of documents, it is unclear if the CIA ever undertook these “multiple personality” experiments. Popularized accounts continue to suggest that the U.S. government did attempt to do so. However, a careful reading of these cases shows that, based on available information, these individuals’ symptoms had little similarity to those of dissociative identity disorder.

“Cult” Mind Control

Allegations of brainwashing or mind control resurfaced in the late 1960s and 1970s in the context of some counterculture religious, self-help/psychotherapy, and political groups. The complex topic of what constitutes a “cult” spans the fields of psychiatry, psychology, sociology, anthropology, religious studies, and political science. The modern use of the term “cult” is usually pejorative, denoting a subset of “unconventional” people within religious or secular groups who hold views characterized as “extremist and/or false” and who often live under the control of a charismatic, authoritarian leader. However, the religious history of the United States is characterized by myriad religious sects and/or spinoffs from traditional religions; political movements; and schools of psychotherapy/self-help/self-improvement, among others. Some have quipped that a cult designation simply means, “A religion I don’t like.” Psychiatrist Marc Galanter reframed the religious cult mind-control debate by advocating the more neutral term *charismatic religious sects*, pointing out that some individual’s behavior within such a social group may reflect psychological adaptation rather than psychopathology.

“Destructive” cults are defined more precisely, although in multiple ways. These definitions tend to include a rigid authoritarian, hierarchical structure; violence toward members/outsiders, including rape/sexual torture of children and/or adults; abusive psychological/emotional control tactics; extreme isolation; confinement abuse; physical assaults; forced suicide and/or murder; violent psychological/ physical/ homicidal retaliation against those who challenge the group from within and/or try to leave; retaliation against outside critics of the group; and a financial structure that primarily benefits the leader(s). These cults may center on religion/ spirituality/ mysticism, politics, psychotherapy/self-improvement, and economic promise, among others.

Dissociation and “Mind Control”

In February 1974, the heiress Patty Hearst was reportedly kidnapped, tortured, sexually abused, and held in prolonged solitary confinement by a radical “cult-like” group, the Symbionese Liberation Army. After months of this treatment, Hearst was said to have been transformed into a new dissociative self/identity state, that of the revolutionary “Tania,” who aided her captors in criminal acts, including bank robbery and murder. A group of defense psychiatric experts including Louis J. West, Lifton, Martin Orne, and Margaret T. Singer advanced the theory that the Tania state was developed in response to torture and mind control. In this view, Hearst was under duress for violent behavior in the Tania state and thus not criminally responsible. They cited the “Stockholm syndrome,” based on a bank robbery and hostage situation in Sweden in which hostages reportedly “traumatically bonded” with the robbers, complicating the hostage negotiations. The Patty Hearst jury was not convinced and convicted Hearst.

Recent meta-analyses and case studies (e.g., that of Jaycee Dugard who was kidnapped and lived with her captors over 18 years, bearing two children) have raised questions whether Stockholm syndrome has reliable/valid diagnostic/syndromal criteria, and is more than an unscientific media construction. Dugard and other survivors of long-term kidnapping have denied having Stockholm syndrome, stating that they were not subject to “traumatic bonding,” but were doing their best to survive under the control of dangerous, homicidal individuals who avoided law enforcement on multiple occasions. Feminist critics have posited that Stockholm syndrome is better conceptualized as a patriarchal theory allowing male law enforcement to rationalize their failures in difficult hostage situations by blaming suggestible, psychologically “weak” women.

Dissociative “Ego States” and Cults

Some researchers on destructive cults posit that cult membership produces an “ego state” or dissociated self-aspect that allows the person to function ego-syntonically within the cult milieu.[2] This entity is not conceptualized as meeting criteria for a dissociative identity disorder self-state but suggests a diagnosis that may be analogous to Example 1 in this section. Some forensic experts in cases of violent behavior by cult members have posited that the influence of the cult leader results in a “shared psychotic disorder” (now Example 4 in Other Specified Schizophrenia Spectrum and Other Psychotic Disorder). Other experts disagree. No research studies have examined either question.

In this regard, some have pointed out that destructive cult groups like the People’s Temple, Hare Krishna, and Scientology seek out people who are psychologically vulnerable and seeking “meaning” in life, typically feeling estranged from religion, families, and communities. It has been hypothesized that cult members, as in pedophilia and intimate partner violence, initially “love-bomb” (groom) the vulnerable person, who then gradually is psychologically enmeshed in the group and manipulated into feeling responsible and obligated to it. Ultimately, the entire life of the person and all their main relationships—friendships, marriage, child-rearing—may be centered on the group. In the authors’ experience, destructive cult groups value highly dissociative individuals. The latter can readily enter deep trance states, block out pain, appear to have uncanny telepathic states, and manifest transformational, possession, and/or seemingly lycanthropic states.

Individuals leaving these groups, as those involved in destructive religious-offshoot cults like the polygamous Fundamentalist Church of Latter-Day Saints, may feel completely anomic, adrift, and alone in the “regular” world, shunned by their family and community, now living in a culture with which they have little experience. Some are subject to direct retaliation from group members after attempts to leave. Although this has not been studied, symptoms of depersonalization/derealization and possibly identity alteration could accompany individuals in these situations.

Dissociation and Identity Alteration in Children Associated with Armed Groups. Systematic torture and psychological coercion leading to self/identity alteration may have an even greater effect in children and adolescents. It is estimated that at least 300,000 individuals under 18 in over 30 conflicts worldwide take part in combat or act in support of armed groups as spies, cooks, laborers, sex slaves, and other functions. Authorities view the commonly used term “child soldier” as problematic, preferring “children associated with armed groups.” Like the terms “child prostitution” and “child pornography,” “child soldier” implies agency by the child. Authorities on child trafficking suggest eliminating the former terms and substituting more accurate ones such as, for example, “children sold for rape by and to adults in person, on video, livestreaming.” Some children join armed groups without gross coercion (e.g., for security; training; joining a liberation struggle; health care; escape from abusive families; deaths of family during conflict; destruction of home community; lack of opportunities for education or employment). Also, in some cultural settings, the term “underage soldier” is culturally comprehensible, but not “child soldier.” Here, adulthood begins at ages 12 to 14 after completion of specific ritual actions by the individual and/or community-specific rituals of transition into adulthood.

Based on studies of individuals who have left them, some of the most destructive groups only abduct children and adolescents and use systematic techniques to attempt to create new self/identity-states, loyally bonded to the group. The Lord’s Resistance Army (LRA) began in the late 1980s in Northern Uganda as a rebel movement. The LRA has since spread into the Democratic Republic of Congo, South Sudan, and the Central Africa Republic. As of 2008, this group had abducted over 20,000 children and, through extreme, murderous violence, displaced over 1.9 million people in Northern Uganda. Qualitative studies describe a series of systematic techniques used by LRA to sever attachments and bonds to children's families and cultural backgrounds, and to create a bond to the LRA group. The latter is to ensure loyalty as the group spreads beyond Uganda, especially into new cultural areas with different languages.

Techniques included isolation from others with shared cultural or linguistic characteristics, forced silence, and violent responses if individuals broke these rules or failed to learn the ethnic language of the group leaders. Children caught leaving the group often were publicly beaten to death by other children from their own ethnic group. Children were segregated by gender roles; girls were assigned to families to become servants or one of the wives of a group leader. Boys would be assigned to families of LRA leaders and given increasingly difficult tasks and loyalty tests to earn the right to carry a gun and become a soldier. Ultimately, ritual witchcraft ("*gris gris*") was performed on the boys as a rite of transition to becoming a soldier. Other forms of violence to sever old community ties involved raiding a child's home village and forcing him to kill members of his own tribe or family.

Case studies of adolescents and young adults who left the LRA and similar groups describe dissociative self-states with switching between violent group–oriented states and pacific self-states that want to reintegrate with their communities. Other survivors of these groups are described as strongly identified with the group values, dominated by violent self/identity-states for whom killing is a syntonic, desirable, almost a peak experience. A case study from the Netherlands describes a 24-year-old man from Sierra Leone, abducted at age 8 by a violent group like the LRA. He had witnessed, and was subjected to, violent and murderous atrocities as well as being forced to participate in them. Escaping at age 13, he eventually found asylum in the Netherlands. At the time of treatment, he met DSM-IV-TR diagnostic criteria for dissociative identity disorder, PTSD, and major depressive disorder. He switched among self-states, including malevolent, violent, group-affiliated self-states. He responded to a transference-focused, trauma/dissociation-focused psychotherapy that tracked self-state alterations and reframed the violent self-states as protectors.

Diagnosis and Clinical Features

As described above, there are a variety of situations where torture and psychological coercion are brought to bear on children, adolescents, and adults. It is possible that some, but not all, individuals subjected to extreme torture and psychological coercion develop dissociative symptoms such as depersonalization/derealization, dissociative amnesia, dissociative identity disorder, or self-states consistent with a diagnosis of Example 1 in this section. They may exhibit reduced cognitive flexibility, behavioral regression, and profound changes in values, attitudes, beliefs, and sense of self.

Treatment

There are no empirical treatment studies of individuals subjected to extreme coercion with indoctrination into a belief system that alters their identity. The triphasic trauma/dissociation treatment structure described in other sections would be a logical treatment framework. It is beyond the scope of this chapter to describe the sequential psychosocial interventions for disarmament, demobilization, and reintegration of children associated with armed groups. Some youth have been treated with a group therapy model to help build skills, process trauma, and rejoin their communities, if possible. Clinical case series and studies of children returning from these groups suggest that cultural competence is essential for successful treatment. Local categories of distress—in addition to or instead of DSM-5-TR PTSD, mood or dissociative symptoms—emphasize survivors being haunted nightly by the ghost of someone they killed, with hallucinations of the deceased. A systematic study of spirit possession in Northern Uganda found that *cen*, a specific form of possession by the spirit of a murdered person, was associated with the highest rates of prior witnessing, experiencing, and/or perpetrating extreme violence in former child members of LRA. The authors conceptualized this form of pathologic possession as a dissociative disorder, also associated with the highest rates of PTSD and depression. Fear of the *cen* was not only focused on the former soldier but on the expectation that the *cen* would afflict the youth's family or village if they returned without being healed.

In some cases, successful treatment consisted of specific healing rituals to expunge the possessing entity and to focus the sufferer on the future. In other reports, former soldiers have gone directly to ask for forgiveness from family members of those they have killed to permit reintegration into their community. These communities had procedures to structure asking for forgiveness, granting it, making amends, and reintegrating the former combatant back into their community.

ACUTE DISSOCIATIVE REACTIONS TO STRESSFUL EVENTS

Definition

Example 3 of other specified dissociative disorder describes acute, transient dissociative symptoms that last 1 month or less, and frequently only hours or days, and occur in reaction to stressful events. Because many of the briefer episodes ended when the patient was brought rapidly for emergency treatment, they might have lasted longer without intervention. Acute dissociative reactions cover a spectrum of severity, ranging from milder presentations of mixed dissociative symptoms to more severe but transient presentations that combine dissociative and psychotic symptoms. The prominence of somatic presentations of dissociation (e.g., conversion symptoms such as analgesia or paralysis) and psychological presentations (e.g., derealization, constriction of consciousness) also varies; a mix of both presentations is typical.

The acute dissociative symptoms may be precipitated by traumatic or nontraumatic stressors and persist for up to a few weeks after the stressor dissipates. Although patients are not always able to identify and report a stressful event as a potential cause of their disorder, a triggering event is usually recognized as precipitating the dissociative reaction once collateral information is obtained. The severity of the stressor varies, including frankly traumatic events such as the unexpected, violent death of a relative or experiencing a natural disaster. However, less stressful events, such as the loss of a significant relationship, may also precipitate this presentation. Community-level stressors, such as high levels of neighborhood violence, may predispose individuals to experiencing Example 3 reactions. The Western research literature has tended to subsume immediate dissociative reactions to stress or trauma into the concept of peritraumatic dissociation, defined as dissociative reactions occurring during and after a traumatic event. However, Example 3 does not require the presence of traumatic exposure, which distinguishes this condition from ASD.

Some cultural syndromes and other folk illnesses may fulfill criteria for Example 3 when they are characterized by acute reactions to adversity with mixed but transient dissociative symptoms. For instance, some cases of *ataques de nervios* (“attacks of nerves”) in Latino populations would best be described using this diagnosis. However, cultural concepts of distress do not have a one-to-one relationship with psychiatric disorders; instead, they cut across diagnostic clusters and must be translated into psychiatric categories at the level of the individual. For example, some persons with *ataques de nervios* qualify for a dissociative diagnosis—including other specified dissociative disorder—whereas others do not, meeting criteria instead for panic disorder, PTSD, or no psychiatric disorder.

Diagnosis and Clinical Features

The literature identifies several overlapping symptom clusters characteristic of these acute, brief, reactive dissociative conditions.

DSM-5-TR describes symptoms of constricted consciousness, depersonalization/derealization, perceptual disturbances (such as time slowing, metamorphopsia, brief amnestic experiences, transient stupor, and/or functional neurologic (conversion) sensory–motor deficits. Uncomplicated cases of Example 3 generate disturbances in interpersonal relationships and work-related functioning that usually resolve within a few days. Nevertheless, crisis intervention during the acute episode is often required. This category is also associated with a risk of suicidal behavior.

Acute dissociative reactions may be codified culturally into stereotyped reactions with associated causal attributions, courses, and treatment responses. In Latino communities, for example, cases of *ataque de nervios* may present with prominent dissociative symptoms consistent with Example 3 of other specified dissociative disorder. Reports of this cultural syndrome are frequent in U.S. Latino communities, ranging from 5 % to 11% across Latino subgroups in a representative community-based sample. Patients with *ataques* describe acute fits of emotionality associated with a sense of loss of control, usually precipitated by family stressors. Severity of *ataques,* assessed as lifetime number of episodes, is positively and independently related to self-reported dissociative symptoms measured with the DES and to clinician-diagnosed dissociative disorders, assessed with the SCID-D-R. In a representative community-based sample of U.S. Latinos, as well as in an urban primary care Latino sample, endorsement of *ataque* on a standardized survey was associated with a higher risk of suicidal ideation and behavior as well has higher morbidity (e.g., greater impairment).

Ms. C, a 27-year-old woman who migrated to the United States at age 14 from the Dominican Republic, had an episode of acute dissociative and aggressive symptoms while living in her brother's apartment in a large urban center. She had lost her apartment 3 months before, after separating abruptly from her common-law husband, and had been taken in by her brother into his rented apartment, along with her 7-year-old daughter. The morning of her acute episode, her brother informed Ms. C that she could not stay in the apartment any longer because he was being threatened with legal action for violating his lease by taking in additional family members.

After receiving the news, Ms. C reacted initially with apparent calm. However, during a discussion with her brother 2 hours later —when he again raised the topic of her departure and inquired as to her timeline for leaving—she became agitated, started screaming and crying, and attacked her brother with her hands and nails, aiming for his face. The brother was able to restrain her, but she ripped part of his T-shirt off his chest. For approximately one minute, whimpering softly, Ms. C slowly and repeatedly tore into small pieces the section of the T-shirt.

The episode waned gradually, and Ms. C became coherent again and asked her brother, "Why are you holding me like this?" Helped to lie down on a bed, she denied any memory of the event, recalling only the initial part of the argument. She did recall "an enormous rage" as her episode began and seeing her brother as "unreal" and "very small and far away" and she herself as "immensely powerful and tall, my head nearly touching the ceiling." She felt she was moving "in slow motion" at first but reported the absence of recall from that point forward.

Ms. C did not want to seek help for this episode—which she labeled an "*ataque de nervios*" (attack of nerves)—attributing the onset of her acute symptoms of depersonalization, derealization, time slowing, micropsia/macropsia, and micro-amnesia to the impact of the sudden news of her "eviction" by her brother. She felt this was a "normal" reaction, proportionate to the stressfulness of the situation, and would resolve on its own. However, Ms. C's family was very concerned about her amnesia, questioning her ability to care for her daughter on her own.

Later that afternoon, Ms. C's relatives escorted her to the psychiatric Emergency Room. She described feeling exhausted after her acute episode and was irritable during the interview, especially about her family's bringing her to the hospital. She slept on her gurney during most of the next 14 hours. The next day in the ER, Ms. C had two more episodes of sudden agitation, crying, and screaming, during which she attempted to scratch her face and leave the room. Because she responded quickly to verbal intervention and reassuring physical contact, she was not given medications or placed in restraints but was kept on one-to-one observation. Laboratory test results were unremarkable, as were an electroencephalogram and a lumbar puncture. On examination, she was crying softly, her mood was sad, and she reported transient depersonalization and derealization and ongoing amnesia for her acute episode. She was oriented to time, person, and place; had no psychotic symptoms; and denied current suicidal ideation.

Considering her gradual improvement and lack of evidence of immediate danger to self or others, the ER staff did not feel Ms. C qualified for an involuntary hospitalization, and she refused a voluntary admission. After 36 hours of ER observation, she was discharged with an outpatient appointment for herself and her daughter. However, she did not keep the initial appointment and was lost to follow-up.

Acute Mixed Dissociative Psychotic States

Many of the patients with mixed and transient dissociative and psychotic symptoms would fit the venerable diagnostic construct of *hysterical psychosis,* a condition studied extensively in the 19th century. Janet focused particularly on the traumatic origins of these symptoms and described their resolution through symptom stabilization, work on traumatic memories and fixed ideas, and, finally, conscious synthesis of the traumatic memories. As described above (section on History), at the end of the 19th-century dissociation-based theories of hysteria waned, the Freudian theories of hysterical symptoms became ascendant, and the Bleulerian view of schizophrenia came into prominence, subsuming hysterical psychosis. One dominant view was that individuals with hysterical psychosis had histrionic personality traits, limited ego resources, and were overwhelmed by conflicts, often sexual in nature, and/or acute losses. Other authors maintained that these individuals manifested a type of schizophrenia and others saw them more in the factitious/malingering spectrum, especially in forensic contexts. More recently, with the rediscovery of Janet's work and the development of the TM of dissociation, individuals with hysterical psychosis have been described as reacting to severe traumatic events, leading to alternative terminology such as *reactive dissociative psychosis.*

These combined dissociation-psychosis presentations must be distinguished from brief psychotic reactions, usually to traumatic stressors, without dissociative symptoms. In the North American literature, many of these patients may be conceptualized under the DSM-5-TR diagnosis of *brief psychotic disorder,* previously known in DSM-III and DSM-III-R as *brief reactive psychosis*. Symptoms of the latter condition overlap considerably with that of hysterical psychosis, as the diagnostic criteria for brief reactive psychosis specified that precipitating events included severe stressors such as a major loss or traumatic experiences such as combat. In DSM-5-TR, brief psychotic disorder has specifiers for occurrence after severe stress/trauma and without stress/trauma, respectively. Studies of brief psychotic disorder have not applied measures of dissociation and systematic studies of stressful or traumatic antecedents to brief psychotic disorder have not been conducted.

Turkish researchers describe hysterical psychosis as a common, well-known syndrome that is regularly diagnosed by Turkish psychiatrists, often presenting in the emergency department. Some Turkish authorities have reported case series of patients where hysterical psychosis was the presenting symptom cluster of dissociative identity disorder. The latter was recognized as the patient responded to treatment of the acute symptoms.

Cases of dissociative psychosis usually have a sudden onset; duration is typically 1 to 3 weeks, but in some cases may be longer. The individual's affect is often labile. Due to rapid shifts among different behavioral states, the individual may exhibit extreme ambivalence—doing and undoing some act in a psychotic or perseverative fashion. The psychotic and dissociative symptoms may have prominent posttraumatic content, which can be embedded in complex perceptual and somatosensory phenomena that occur automatically and uncontrollably (e.g., dissociative flashbacks to the traumatic event). In some cases, functional neurologic (conversion) symptoms may be the most prominent presenting sign(s). This condition causes a temporary impairment of daily functioning that typically requires short-term hospitalization. It is also associated with risk of suicidal behavior. A high prevalence of childhood abuse and/or neglect is often associated with dissociative psychosis.

Nonpsychotic cases of Example 3 are differentiated from adjustment disorder, ASD, and PTSD by the predominantly dissociative nature of the presentation and the absence of other symptoms of these disorders (e.g., hyperarousal, avoidance). There is no dissociative subtype of adjustment disorder and ASD can occur in the absence of dissociative symptoms. Example 3 presentations do not require a traumatic stressor.

Case descriptions of dissociative psychosis have included bizarre behavior, such as acting like an animal, and catatonic-like symptoms. Speech patterns may be disjointed and intermittent, sometimes giving the clinical impression of psychotic thought disorder. However, when the patient can speak coherently, there is no evidence of a psychotic thought disorder.

Differential Diagnosis of Dissociative Identity Disorder and Dissociative Psychosis. Some case series of individuals with dissociative psychosis have described symptoms that appear consistent with dissociative identity disorder. For example, a woman diagnosed with hysterical psychosis repeatedly hallucinated a second woman, with more resilient characteristics than herself, who conversed with her supportively about the life difficulties that preceded the psychotic episode.

Reports describe the experience of dissociative psychosis as consisting of multimodal hallucinosis; age regression; extensive dissociative amnesia; oneroid states; pseudodelirum/ pseudodementia; and delusional phenomena. However, in these accounts, the apparent delusions may be understood as related to, and usually symbolically condensing, traumatic experiences and intrapsychic conflicts that have generated the episode. Clinical experience suggests that most of these patients do not have a psychotic disorder but, with more extensive evaluation, will meet diagnostic criteria for dissociative identity disorder. In general, in dissociative identity disorder, despite delusionally separate and/or disoriented self-states, other self-states maintain reality testing and disavow the delusional beliefs. In superimposed dissociative psychosis, reality testing and relationship to reality is impaired across all self-states, with loss of reality testing, until/unless the psychosis is lysed.

Pathology and Laboratory Examination. Research on hysterical psychosis has documented high scores on standard measures of hypnotic capacity in individuals with this disorder. High hypnotizability distinguishes this group of patients from those with psychotic disorders, who generally have low scores on standard measures of hypnotizability.

In contrast to acute organic mental disorder (delirium), persons with dissociative psychosis do not exhibit findings that are characteristic of delirium on neuropsychiatric examination, laboratory tests, EEG, or brain imaging. They are characterized by dissociative symptoms, including amnesia, dissociative flashbacks, and disruptions of consciousness rather than by the waxing and waning of symptoms, cognitive deficits, and interpersonal impoverishment typically found in patients with delirium.

Treatment

There are no systematic studies of treatment of cases of Example 3, and treatment principles are derived from case series. *Ataques de nervios* and other folk illnesses fitting Example 3 may be characterized by suicidal or aggressive behaviors amid the acute dissociative episode. Restraining the person may be temporarily necessary. In many cases, the patient or family do not seek professional help. Paradoxically, if the triggering stressors are deemed minor by significant others or there are concerns about chronicity, the individual may be brought for psychiatric treatment. While comorbid symptoms and disorders (e.g., depression) can respond to psychotherapeutic and/or psychopharmacological interventions, the dissociative symptoms are more likely to respond to supportive-expressive psychotherapy that identifies the traumatic and/or stressful events that preceded the dissociative reaction, as well as the intense psychological conflicts that, almost invariably, have been generated in the context of the stressful events. Usually, supporting the patient as they “tell the story” of the stressful event(s) is highly therapeutic. Recent research finds support for trauma-informed cognitive-behavioral therapy, with or without narrative and psychodynamic modalities for individuals whose reactions were precipitated by traumatic exposure.

Treatment of dissociative psychosis follows the triphasic model: education to help the patient understand the disorder; supportive interventions to help the patient develop better control over symptoms; expressive interventions to give voice to distress over traumas, conflicts, and losses; transformation of reenactments into a coherent narrative, and return of adequate reality testing. Following Janet, modern case series have demonstrated the effectiveness of hypnotic interventions in the treatment of patients with dissociative psychosis, many of whom demonstrate high hypnotizability on standardized measures. Clinicians trained in hypnosis may find that assessment of hypnotizability aids in differential diagnosis and hypnotic interventions may be effective as an adjunct to psychotherapy for these patients.

Hypnotic interventions give these patients a greater sense of control over symptoms (and autohypnotic states); help the patient to understand symptoms as logical in the context of trauma and/or stressful events; and can provide imagery that may help the patient with affect regulation, symptom control, grief-work, mastery of traumatic material, and integration of these experiences for healing. Patients may also helpfully use self-hypnotic techniques to manage symptoms outside of psychotherapy.

Patients with dissociative identity disorder in an acute episode of dissociative psychosis should be treated for the underlying disorder. The clinician should provide the patient psychoeducation about acute dissociative/psychotic symptoms to reassure them that these are part of the dissociative disorder and not necessarily a manifestation of a comorbid psychotic disorder. The clinician should emphasize that treatment of the episode will follow the phasic trauma treatment model to understand what has engendered the acute symptoms. Patients may benefit from learning that medication is unlikely to have a major direct impact on symptoms but may help decrease the intensity of severe terror states and assist with sleep (see discussion of psychopharmacological treatments for dissociative identity disorder). The clinician should have an index of suspicion that the patient with dissociative identity disorder in an acute dissociative psychotic episode may be responding to a recent traumatic event such as a sexual assault that they cannot process in verbal form without psychotherapeutic interventions.

Ms. M, a 29-year-old patient with dissociative identity disorder was admitted to a psychiatric hospital by her therapist. The therapist reported that the patient had been struggling with safety issues and did not attend an emergency appointment. The therapist called 911 and the police had to forcibly enter the patient's home. The patient was lying in her bed, mute except for moaning, rocking back and forth, and was otherwise unresponsive. She was taken to the emergency department where medical, toxicology, and laboratory examination were within normal limits. The patient had superficial self-inflicted cuts on her arms, bruising on her legs, and cuts on her feet. Examined by the ED psychiatrist, she spoke in a robotic voice, and said she was seeing "lights flashing, colors spinning" and was "hearing so many voices, I can't make them out at all;" some voices were experienced inside her mind but others were "bouncing around the room." She saw "lots of faces," felt "my body just floating around," but also would moan and say, "it hurts" but could not explicate this further. She said her thoughts were "being sucked out of my mind" but was perplexed about this, yet no delusional explanation was elicited. She denied thought broadcasting or audible thoughts and was "scared and frightened." She reported that her radiator at home "wants to kill me," but had no insight or distance from this idea.

At times her thoughts appeared disorganized with blocking; at other times, thought process was coherent and goal directed. She reported intense urges "from somewhere" to "cut myself to pieces and die." She was admitted to the hospital and treated for "acute psychosis" with olanzapine 10 mgs by mouth daily and lorazepam 2 mg by mouth every 6 hours as needed, with no change in symptoms over a week other than increased sedation. She was seen by a consultant who attempted to explore the origins of her recent symptom exacerbation. With consistent focus on the patient's making sense of her experience, she reported a recent sexual assault by an "old friend" whom she had encountered at a bar, who may have drugged her drink. Taken to his house, she found herself being assaulted, felt very confused, and attempted to run away, cutting her feet. She reported that his grabbing her legs as she tried to run out of the house caused the bruising. She eventually escaped and found her way home. Then, she cut herself “for punishment for being so stupid.” She reported a history of being drugged repeatedly in childhood, particularly while being trafficked as part of a ring based in her family of origin. The hallucinations and other perceptual abnormalities, as well as the thought disorganization, were related to specific flashbacks of these experiences, apparently stimulated by the drugging and the rape. After talking about this, the hallucinations abated and the patient became calmer. She was given psychoeducation about psychotic-like dissociative symptoms. At the time of discharge, antipsychotic medication was discontinued and the patient appeared grounded, insightful, and oriented, without perceptual disturbances or hallucinations, other than her usual self-state voices and PTSD intrusive symptoms.

DISSOCIATIVE TRANCE DISORDER

Definition

DSM-IV-TR included pathologic trance and pathologic possession trance states in Appendix B as separate elements of a combined category of dissociative trance disorder, one of the “Criteria Sets and Axes Provided for Further Study.” Categories in this section of the Manual fell short of full inclusion in DSM-IV-TR as specified diagnoses but were included in the Appendix to stimulate research for possible incorporation in DSM-5. Pathologic dissociative possession states are included in DSM-5-TR under dissociative identity disorder, leading to its redefinition as having two variants: possession-form and non–possession-form, both understood as culturally patterned forms of distress. By contrast, data on Example 4, dissociative trance disorder, was deemed insufficient to include it in DSM-5-TR as a specific dissociative disorder or to subsume it under another dissociative disorder. In ICD-11, trance disorder is a specified dissociative disorder, which is classified together with possession trance disorder as a combined category. Support for trance disorder as a distinct diagnosis is based on research from diverse international settings (e.g., Turkey, India, Mozambique, Uganda). A main distinction between Example 4/trance disorder and pathologic dissociative possession trance presentations in both classification systems is the absence of identity alteration in the former, which is ascribed in the latter to the influence of a spirit, power, deity, or other spiritual entity. Moreover, individuals with possession-form presentations exhibit a wider range of dissociative behaviors which are attributed to the possession agent. However, while this distinction is clear conceptually, in practice, the line between trance and possession trance conditions often becomes blurred. Most reports of trance conditions describe experiences that precede the emergence of identity alteration in the affected person and could therefore be considered prodromal elements of possession trance conditions, rather than a separate disorder.

Diagnosis and Clinical Features

DSM-5-TR describes Example 4, dissociative trance disorder, as a temporary, marked alteration in the state of consciousness or loss of the customary sense of personal identity without change of identity. There is often a narrowing or complete loss of awareness of the immediate surroundings or a selective focus on stimuli within the environment. This may manifest as a profound unresponsiveness or insensitivity to environmental stimuli. In addition, the individual may display minor stereotypical behaviors (e.g., finger movements) of which they are unaware and/or are experienced as beyond their control. Some patients experience transient paralysis and/or loss of consciousness.

The description in ICD-11 is similar: pathologic trance is defined as a recurrent or single and prolonged marked alteration in the person’s state of consciousness or a loss of the normal sense of personal identity without the experience of being replaced by an alternative identity. Attention and awareness may be limited to, or concentrated upon, unusually narrow and selective aspects of the immediate environment and there is often a limited but repeated set of movements, postures, and utterances that are experienced as outside the person’s control.

Both the DSM-5-TR and ICD-11 definitions exclude presentations characterized by identity alteration, including those attributed to an external possessing agent. Moreover, in both manuals, trance states should only be considered symptomatic of a mental disorder when they are involuntary and unwanted, not accepted as a part of a collective cultural or religious practice, and result in significant distress or significant impairment in personal, family, social, educational, occupational, or other important areas of functioning.

Research on trance conditions without identity alteration may combine individuals with trance states not associated with subsequent identity alteration (“ICD-11 trance disorder” or “DSM-5-TR Example 4”) and those whose trance states precede dissociative possession state experiences (possibly “prodromal” possession-form presentations). With this limitation in mind, non/pre-possession trance states in settings in Mozambique and Uganda have been associated with hearing voices, strange dreams or nightmares, restlessness, inability to speak, and elevated somatic symptoms (e.g., headaches, stomach pains), among others.

Pathology and Laboratory Examination. There are little systematic data on differential diagnosis of this example of other specified dissociative disorders. Patients with this condition may require neurologic evaluation, EEG, and brain imaging to rule out a seizure disorder, and cardiac workup to rule out hypotension and/or arrhythmias. The differential diagnosis of the abnormal movements in Example 4 may include essential tremor and Huntington disease. Although the latter may be associated with significant psychiatric psychopathology, it usually does not present with loss of awareness of surroundings until very late in the course when dementia is prominent. It runs in families and can be ruled out by genetic testing, if needed. Children with attention-deficit hyperactivity disorder may be hyperfocused on things that interest them, like video games or television, but do not manifest trance-like behavior apart from this. Many individuals can “tune out” parents, spouses, and others during family strife; however, this is not the same as the spontaneous, involuntary disorder of pathologic trance states. Patients with this condition may demonstrate high hypnotizability on standard hypnotizability testing, a finding which may be helpful in differential diagnosis.

Treatment

There are no systematic treatment studies of dissociative trance in the absence of pathologic dissociative possession states. The treatment of trance states in settings with developed concepts of possession illness usually follows healing practices for spirit possession. In cultural settings without a tradition of these possession trance healing practices, treatment may proceed by clarifying the nature of the trance states and their relationship to more ordinary phenomena such as daydreaming. Assessment of hypnotizability may also allow the clinician to produce the same symptoms the patient is experiencing with hypnotic induction. This may lead to an increased sense of control and may be sufficient to assist the patient in recognizing autohypnotic states and developing strategies to manage these. In more chronic cases, a supportive-expressive psychotherapy, similar to that for functional neurologic (conversion) disorder, may be most successful.

Ms. M, a married woman in her 50s, was seen in consultation for "passing-out attacks." She had had increasingly frequent episodes of grimacing, rapid eye-blinking, and piano-playing finger movements, accompanied by an "eye-roll," showing only the sclera, and falling to the ground in an unresponsive state, often for several hours at a time. Extensive work up showed no indication of seizures, other neurologic disorders, cardiac causes, and metabolic causes, among others. Repeated courses of antipsychotic medications and mood-stabilizers for "periodic catatonia" were ineffective and only caused problematic side effects. Anti-depressants helped with the patient's mood but the episodes persisted. The consultant performed a lengthy diagnostic interview and administered screening and diagnostic interviews for dissociative disorders, ruling out dissociative identity disorder. Although the patient reported "watching" herself as if at a distance during the episodes and experiencing them "as if something just comes over me that I cannot control," she did not meet criteria for depersonalization/derealization disorder.

The consultant discovered many current personal, marital, and family stresses and determined that frequency of episodes increased with higher levels of family discord. The patient reported a childhood history of severe verbal violence between her parents and directed at her and her siblings. She coped with this psychologically by "just leaving" and "tuning it all out." The patient was educated about autohypnosis and trance states and rapidly achieved deep hypnotic trance with a brief induction. Hypnotic interventions were able to reproduce each of the patient's symptoms as well as a "hidden observer" that remained aware of what was occurring. Over several sessions, the patient was able to learn self-hypnosis to reproduce symptoms and to alleviate them. She was delighted at this improved sense of control and returned to her outpatient providers for work on the family issues.

The patient returned to the consultant several years later, disabled by the same symptoms, with unresponsive episodes many times a day. She had stopped doing self-hypnosis and family problems had recrudesced. She was admitted to a specialized trauma disorders unit where, initially, she continued to have many episodes a day. Psychotherapy focused on viewing the trance episodes as a maladaptive coping strategy and challenging the patient to handle problems and difficult affects (anger, shame, sadness) without hypnotic evasions, likening this to an addiction to drugs or alcohol. Milieu interventions involved staff not engaging in efforts to “rescue” the patient from episodes and helping other patients also not attempt to rescue her by calling staff to intervene. Family therapy focused on psychoeducation and practical solutions to the many issues faced by the family. A strong recommendation was made for outpatient psychotherapy with a clinician proficient in hypnosis and outpatient family and marital therapy. At the time of discharge, the patient had not had a trance episode for over a week.

FURTHER READINGS

Adams-Clark AA, Gómez JM, Gobin RL, et al. Impact of interpersonal, family, cultural, and institutional betrayal on adult survivors of abuse. In: Geffner R, White JW, Hamberger LK, et al., eds. *Handbook of Interpersonal Violence and Abuse Across the Lifespan*. Springer Nature; 2022. pp. 4275–4301.

Aponte-Soto MR, Martinez-Taboas A, Velez-Pastrana MC, et al. The relationship between interpersonal abuse and depersonalization experiences. *Rev Puertorriquena Psicol*. 2019;30:48–59.

Brand BL, Schielke HJ, Putnam KT, et al. An online educational program for individuals with dissociative disorders and their clinicians: 1-year and 2-year follow-up. *J Trauma Stress*. 2019;32:156–166.

Calati R, Bensassi I, Courtet P. The link between dissociation and both suicide attempts and non-suicidal self-injury: Meta-analyses. *Psychiatry Res*. 2017;251:103–114.

Dalenberg CJ, Brand BL, Loewenstein RJ, et al. Inviting scientific discourse on traumatic dissociation: Progress made and obstacles to further resolution. *Psychol Inj Law*. 2020;13:135–154.

Ellickson-Larew S, Stasik-O’Brien SM, Stanton K, et al. Dissociation as a multidimensional transdiagnostic symptom. *Psychol Conscious*. 2020;7:126–

150.

Fung HW, Ross CA, Yu CK, Lau EK. Adverse childhood experiences and dissociation among Hong Kong mental health service users. *J Trauma Dissociation*. 2019;20:457–470.

Granqvist P, Sroufe LA, Dozier M, et al. Disorganized attachment in infancy: a review of the phenomenon and its implications for clinicians and policy-makers. *Hum Dev*. 2017;19:534–558.

Hecker T, Barnewitz E, Stenmark H, Iversen V. Pathological spirit possession as a cultural interpretation of trauma-related symptoms. *Psychol Trauma*. 2016;8:468–476.

Kate M-A, Hopwood T, Jamieson G. The prevalence of dissociative disorders and dissociative experiences in college populations: a meta-analysis of 98 studies. *J Trauma Dissociation*. 2020;21:16–61.

Kelly JT, Branham L, Decker MR. Abducted children and youth in Lord's Resistance Army in Northeastern Democratic Republic of the Congo (DRC): mechanisms of indoctrination and control. *Confl Health*. 2016;10:11.

Lebois LAM, Li M, Baker JT, et al. Large-scale functional brain network architecture changes associated with trauma-related dissociation. *Am J Psychiatry*. 2021;178(2):165–173.

Lewis-Fernandez R, Kirmayer LJ. Cultural concepts of distress and psychiatric disorders: Understanding symptom experience and expression in context. *Transcult Psychiatry*. 2019;56:786–803.

Loewenstein RJ. Dissociation debates: Everything you know is wrong. *Dialogues Clin Neurosci*. 2018;20:229–242.

Loewenstein RJ. Firebug! Dissociative identity disorder? Malingering? Or …? An intensive case study of an arsonist. *Psychol Inj Law*. 2020;13:187–224.

Lyssenko L, Schmahl C, Bockhacker L, Vonderlin R, Bohus M, Kleindienst N. Dissociation in psychiatric disorders: A meta-analysis of studies using the dissociative experiences scale. *Am J Psychiatry*. 2018;175(1):37–46.

Maraldi EO, Krippner S, Barros MCM, Cunha A. Dissociation from a cross-cultural perspective: Implications of studies in Brazil. *J Nerv Ment Dis*. 2017;205:558–567.

Myrick AC, Webermann AR, Loewenstein RJ, Lanius R, Putnam FW, Brand BL. Six-year follow-up of the treatment of patients with dissociative disorders study. *Eur J Psychotraumatol*. 2017;8:1344080.

Nester MS, Brand BL, Schielke HJ, Kumar S. An examination of the relations between emotion dysregulation, dissociation, and self-injury among dissociative disorder patients. *Eur J Psychotraumatol*. 2022;13(1):2031592.

Putnam FW. *The Way We Are: How States of Mind Influence Our Identities, Personality, and Potential for Change*. IUP; 2016.

Reinders A, Chalavi S, Schlumpf YR, et al. Neurodevelopmental origins of abnormal cortical morphology in dissociative identity disorder. *Acta Psychiatr Scand*. 2018;137:157–170.

Salter M. Organized abuse in adulthood: Survivor and professional perspectives. *J Trauma Dissociation*. 2017;18:441–453.

Sar V, Alioglu F, Akyuz G, Tayakisi E, Öğülmüş EF, Sönmez D. Awareness of identity alteration and diagnostic preference between borderline personality disorder and dissociative disorders. *J Trauma Dissociation*. 2017;18:693–709.

Sar V, Dorahy MJ, Kruger C. Revisiting the etiological aspects of dissociative identity disorder: a biopsychosocial perspective. *Psychol Res Behav Manag*. 2017;10:137–146.

Schaffler Y, Cardena E, Reijman S, Haluza D. Traumatic experience and somatoform dissociation among spirit possession practitioners in the Dominican Republic. *Cult Med Psychiatry*. 2015;40:74–99.

Staniloiu A, Markowitsch HJ, Kordon A. Psychological causes of autobiographical amnesia: A study of 28 cases. *Neuropsychologia*. 2018;110:134–147.

van Duijl M, Kleijn W, de Jong J. Are symptoms of spirit possessed patients covered by the DSM-IV or DSM-5 criteria for possession trance disorder? A mixed-method explorative study in Uganda. *Soc Psychiatry Psychiatr Epidemiol*. 2013;48:1417–1430.

Vissia EM, Giesen ME, Chalavi S, et al. Is it Trauma- or Fantasy-based? Comparing dissociative identity disorder, post-traumatic stress disorder, simulators, and controls. *Acta Psychiatr Scand*. 2016;134:111–128.

Vonderlin R, Kleindienst N, Alpers GW, Bohus M, Lyssenko L, Schmahl C. Dissociation in victims of childhood abuse or neglect: a meta-analytic review. *Psychol Med*. 2018;48:2467–2476.

Webermann AR, Myrick AC, Taylor CL, Chasson GS, Brand BL. Dissociative, depressive, and PTSD symptom severity as correlates of nonsuicidal self-injury and suicidality in dissociative disorder patients. *J Trauma Dissociation*. 2016;17:67–80.

---

[1] Salvia is member of the mint family found in Mexico. A legal, unregulated drug, it is a potent hallucinogen whose leaves, stems, and seeds can be purchased online and smoked, infused into drinks, or swallowed. The active ingredient, salvinorin-A, binds to the kappa opiate receptor, producing a hallucinogenic experience lasting from 30 minutes to hours.

[2] “Ego state” is a term originally coined by the psychoanalyst Paul Federn, later expanded by John and Helen Watkins, as a general theoretical construct about the mind, psychotherapy, and the phenomenology of hypnosis. Here, the term “ego” is used differently from the various ways it has been used in psychoanalytic theory,

particularly psychoanalytic ego psychology. For the purposes of this chapter, an ego state can be thought of as a discrete self-state, differing from those of dissociative identity disorder by not being a relatively separate center of information processing and without an ongoing sense of itself, a body image, and a capacity for dissociated action.

21

Normal Sexuality and Sexual Disorders

21.1 Normal Human Sexuality and Sexual Dysfunctions

Virginia A. Sadock, M.D.

Sexuality is a significant factor in a person's mental health. It is relevant to identity, to self-esteem, to the formation and quality of intimate relationships, and to a subjective sense of well-being. One's sexuality is enmeshed with other personality factors, biologic factors, physical health, and the prevailing mores of the society in which a person lives. It includes a personal perception of one's gender, and it reflects developmental experiences with sex throughout the life cycle.

Sexual behavior is diverse. Normality is defined differently in various cultures and at different times in the same culture. Normal sexual behavior brings pleasure to oneself and is frequently directed toward the pleasure of a partner. It involves stimulation of the primary sex organs, often including coitus, and frequently leads to climax; it is devoid of inappropriate feelings of guilt or anxiety and is not compulsive. Recreational, as opposed to relational sex, including casual, premarital, and extramarital activity and masturbation, is normal in some contexts. Sexual thoughts and fantasies are an integral part of sexuality.

Sexuality has been studied by physicians and researchers over time. In the classical era, Hippocrates cited the clitoris as the site of female sexual arousal, the first physician historically recorded to have made that assessment. In the Middle Ages, Islamic physicians recommended coitus interruptus as a form of birth control. At the end of the Renaissance and the beginning of the Reformation, a linen sheath was devised as a condom, not for purposes of birth control, but as protection against syphilis. During the Victorian era, sexologists such as Havelock Ellis (Fig. 21.1–1) and Richard von Krafft-Ebing (Fig. 21.1–2) presented diverging perspectives on sexual behavior. During that same period, Sigmund Freud developed his innovative theories on libido, childhood sexuality, and the effects of the sexual impulse on behavior. In the modern era, the research of Alfred Kinsey, the work of William Masters and Virginia Johnson, and the development of drugs that prevent contraception, aid erection, and replace hormones that decrease with menopause and aging contributed to the development of an era of sexual liberality. Sex has also been a consistent source of motivation and a focus of interest to humankind in general. Depictions of sexual behavior have existed from the time of prehistoric cave drawings through Leonardo da Vinci's anatomical illustrations of intercourse to current pornographic sites on the internet.

The most recent developments in the field of sexuality involve changes in cultural/legal acceptance of behaviors formerly considered taboo or abnormal. There is more acceptance of transgender behavior; homosexuality, first dropped by the APA as a sexual disorder in 1973, is no longer a barrier to joining the military, nor is transgender status. The majority of Americans currently are in favor of same-sex marriage, and self-definitions of sexuality are less binary (homosexual/heterosexual, male/female) and more fluid. Physicians have proposed moving gender designation on birth certificates below the line of demarcation, which would prevent using the designation as proof of gender in legal situations. The *Diagnostic and Statistical Manual of Mental Disorders, fifth edition* (DSM-5) finds fantasies insufficient for a diagnosis of paraphilia but requires acting on the fantasy or impulse. Recent clinical reports include studies of sexual behavior in older adults that have helped breach the taboo against sex in the older adults. Studies of gender differences in response to sexual stimulation indicate that the sequential/linear response model first postulated by Masters and Johnson seems more reflective of male response but appears to be an insufficient model of the experience of many women. Current research is examining the effect of technology and events such as the pandemic on sexual behavior. Conservative trends are apparent in the overturn of Roe vs. Wade and in legislation by some state legislatures severely limiting access to abortion and limiting access to medical care of transgender youth.

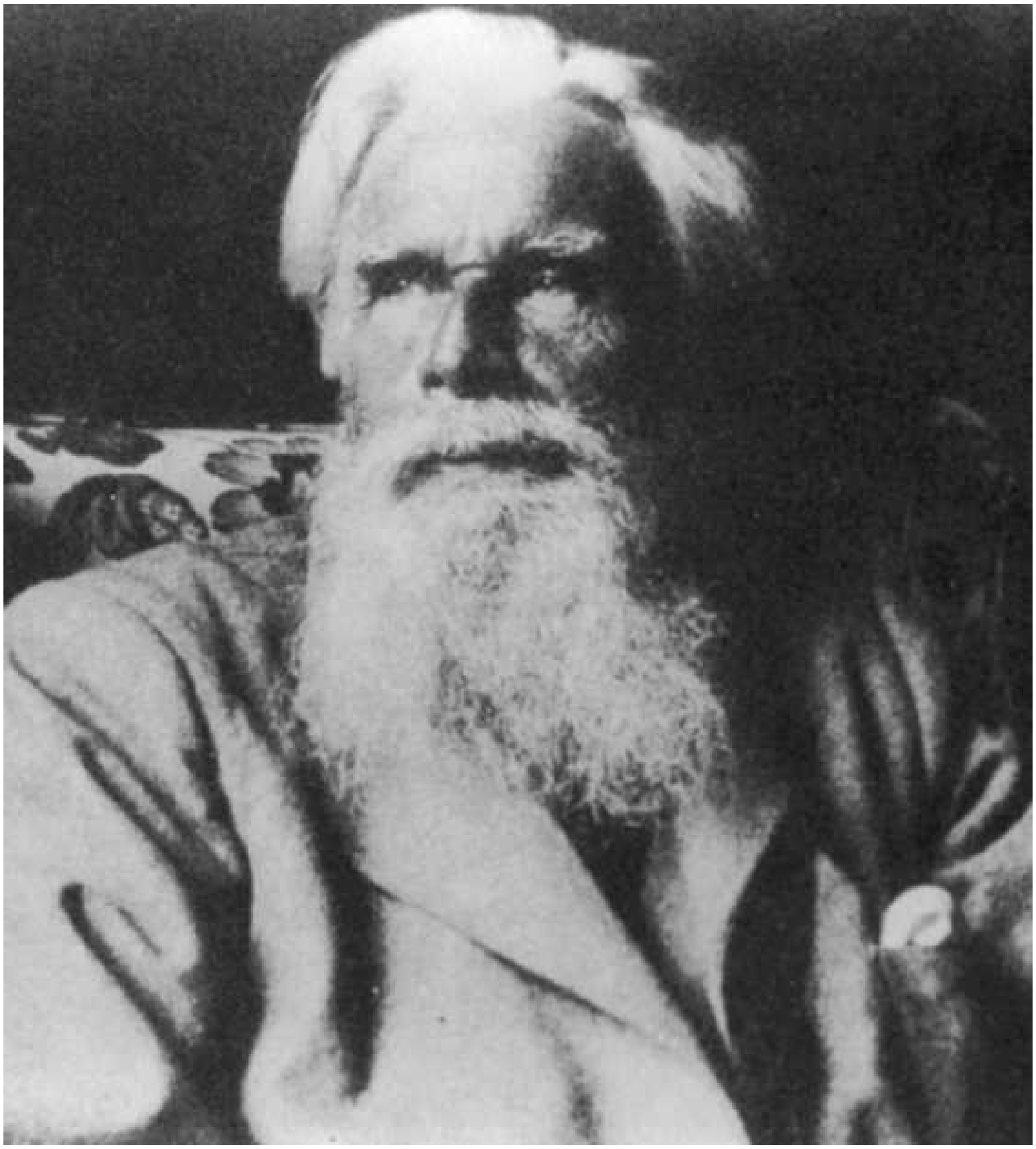

FIGURE 21.1–1. Havelock Ellis (1859–1939). In his book, *Studies in the Psychology of Sex* (1896), Ellis recorded examples of normal and abnormal sexuality. It remains a classic in the field of sexology. (Courtesy of New York Academy of Medicine.)

FIGURE 21.1–2. Richard von Krafft-Ebing (1840–1903), a psychiatrist who published a classic text, *Psyhopathia Sexualis* (1898), in which he documented every variation in sexuality, including zoophilia, necrophilia, urolagnia, and lust murder, among others. Case reports were so lurid and detailed that early editions were published in Latin. (Courtesy of New York Academy of Medicine.)

NORMAL SEXUALITY

Anatomical and Physiologic Bases

Knowledge about the organs of sexuality and the normal physiologic response of men and women is necessary for an informed understanding of sexual dysfunctions. In the 1960s, emphasis was placed on psychological and social contributions to dysfunction; in the 1990s, greater emphasis was placed on the genetic, neuroanatomical, and neurochemical model of human sexuality than on psychological and social factors. Currently, most clinicians believe an interplay of all these factors should be considered in treating sexual dysfunction.

Male Anatomy. The external genitalia of the normal adult man includes the penis, scrotum, testes, epididymis, and parts of the vas deferens. Internal components include the vas deferens, ejaculatory ducts, and prostate gland.

Freud referred to the penis as the executive organ of sexuality. In studies, male initiation of sexual interaction is usually found to be the determining factor in a couple's frequency of sexual connection. Since antiquity, culture has represented the penis in a variety of art forms. In ancient Greece, the cults of Dionysus, Priapus used the phallus as a recurrent symbol of fertility and rejuvenation. The word *penis* has been traced from the Latin, meaning variously "tail" or "to hang," and refers to the pendant position of the organ in its resting or flaccid state. The size of the penis varies within a fairly constant range, but sex researchers over the years have disagreed on the dimensions of the range. All agree, however, that concern over the size of the penis is practically universal among men. Masters and Johnson report a range of 7 to 11 cm in the flaccid state and 14 to 18 cm in the erect state. Of particular interest was their observation that the flaccid dimension bears little relation to the erect dimension, as a smaller penis erects proportionally more than a larger one.

Circumcision, a procedure in which the prepuce is surgically removed, has been practiced for centuries as a religious rite by Jews and Muslims and is a common medical procedure in the United States today. The circumcised penis, with its exposed glans, was once believed to be less sensitive because of the cornification of the epithelium. In laboratory studies, however, researchers have found no difference in tactile threshold between the circumcised penis and the uncircumcised penis. Intravaginally, the prepuce of the uncircumcised penis remains retracted behind the glans during penile thrusting, dispelling the myth that premature ejaculation may be more common in uncircumcised men because of increased stimulation caused by preputial movements. The most recent recommendation of the American Academy of Pediatrics states that there are clearer medical benefits to circumcision than previously recognized, including lessened ability to transmit cases of AIDS, syphilis, or gonorrhea, a lower incidence of penile cancer, and fewer genitourinary infections in the first year of life. They stop short of recommending circumcision and leave the decision of whether or not to circumcise to parents in consultation with their doctors. However, they encourage insurance companies to cover the procedure.

Ejaculation is the forceful propulsion of semen and seminal fluid from the epididymis, vas deferens, seminal vesicles, and prostate into the urethra. The dilation of the prostatic urethra and the passage of fluid into the penile urethra provides the man with a sensation of impending climax, the emission phase of the ejaculatory process. Indeed, once the prostate contracts, ejaculation is inevitable. The ejaculate is then propelled through the penile urethra by contractions of the striated pelvic and perineal muscles. This phase of ejaculation is essentially under somatic efferent control. The ejaculate consists of approximately one teaspoon (2.5 mL) of fluid and contains approximately 120 million sperm cells. It is believed that the larger the ejaculate, the more pleasurable the orgasm, but this belief is highly subjective. The sense of pleasure that accompanies orgasm is thought to be a cortical experience. Men who have had a prostatectomy and therefore no longer ejaculate report they still experience strong and satisfactory orgasms.

Female Anatomy. The external genitalia of the normal woman, also called the *vulva,* include the mons pubis, major and minor lips, clitoris, glans, vestibule of the vagina, and vaginal orifice. The internal system includes the ovaries, fallopian tubes, uterus, and vagina.

The clitoris is an erogenous female organ, only part of which is visible. Masters and Johnson described the clitoris as the primary female sexual organ because orgasm depends physiologically on adequate clitoral stimulation. Anatomically, the clitoris has a nerve net that is proportionally three times as large as that of the penis. The clitoris consists of the glans and the prepuce or clitoral hood. The clitoral prepuce is contiguous with the labia minora, and, during coitus, the penis does not stimulate the clitoris directly. Rather, penile thrusting exerts traction on the minor lips, which, in turn, stimulates the clitoris sufficiently for orgasm. During heightened excitement, just before orgasm, the clitoris retracts under the clitoral hood because of the contraction of the ischiocavernosus muscles.

Retracting, the clitoris moves away from the vaginal barrel, which makes clitoral–penile contact impossible. The size of the clitoris varies considerably and is unrelated to the sexual responsiveness of a particular woman.

Alfred Kinsey found that when women masturbate, most prefer clitoral stimulation. That finding was refined further by Masters and Johnson, who reported that women prefer the shaft of the clitoris to the glans because the glans is hypersensitive if stimulated excessively.

The word *vagina* comes from the Latin word meaning “sheath.” The vagina is usually collapsed, a potential rather than an actual space. Approximately 8 cm long, the vagina extends from the cervix of the uterus above to the vestibule of the vagina or vaginal opening below. In most virgins, a membranous fold, the hymen, separates the vestibule and opening from the rest of the vaginal canal. The mucous membrane lining the vaginal walls rests in numerous transverse folds. To accommodate the penis during sexual intercourse, the vagina expands in both length and width. After menopause, because circulating estrogen concentrations decrease, the vagina loses much of its elasticity

In 1950, Ernst Graefenberg described an area surrounding the female urethra in the anterior wall of the vagina that has come to be called the *G spot.* Approximately 0.5 to 1.0 cm in size, it becomes engorged during sexual stimulation. Many women report that stimulation of the area is highly pleasurable and, in some, can induce orgasm. Graefenberg believed that the tissue here was analogous to the prostate and might account for the spurt of fluid during orgasm reported by some women, similar to male ejaculation. Some women state that a vaginal orgasm is a different experience for them, providing different sensations than a clitoral orgasm. One explanation is that their pleasure is enhanced by the sense of closeness engendered by coitus. Another hypothesis is that these women climax in response to coital stimulation of the G spot.

Innervation of Sex Organs. Innervation of the sexual organs is mediated primarily through the autonomic nervous system (ANS). Penile tumescence occurs through the synergistic activity of two neurophysiologic pathways. A parasympathetic (cholinergic) component mediates reflexogenic erections via impulses that pass through the pelvic splanchnic nerves (S2, S3, and S4). A thoracolumbar, mainly sympathetic pathway transmits psychologically induced impulses. Both parasympathetic and sympathetic mechanisms are thought to play a part in relaxing the smooth muscles of the penile corpora cavernosa, which allows the penile arteries to dilate and causes the inflow of blood that results in penile erection. Relaxation of cavernosal smooth muscles is aided by the release of nitric oxide, an endothelium-derived relaxing factor. Clitoral engorgement and vaginal lubrication also result from parasympathetic stimulation that increases blood flow to genital tissue.

Table 21.1–1.
Responses of Sex Organs to Autonomic Nerve Impulses

Effector Organs	Adrenergic Impulses		Cholinergic Impulses
	Receptor Type	**Responses**	**Responses**
Urinary bladder			
Detrusor	β_2	Relaxation (usually)	Contraction
Trigone and sphincter	α_1	Contraction	Relaxation
Ureter			
Motility and tone	α_1	Increase	Increase (?)
Uterus	α_1, β_2	Pregnant: contraction (α_1), relaxation (β_2); nonpregnant: relaxation (β_2)	Variable
Sex organs, male	α_1	Ejaculation	Erection
Skin			
Pilomotor muscles	α_1	Contraction	
Sweat glands	α_1	Localized secretion	Generalized secretion

Adapted from Westfall TC, Macarthur H, Westfall DP. Neurotransmission: the autonomic and somatic motor nervous systems. In: Brunton L, Knollmann B, Hilal-Dandan R, eds. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*. 13th ed. McGraw-Hill Education; 2018.

Evidence indicates that the sympathetic (adrenergic) system is primarily responsible for ejaculation. Through the hypogastric plexus, adrenergic impulses innervate the urethral crest, the muscles of the epididymis, and the muscles of the vas deferens, seminal vesicles, and prostate. Stimulation of the plexus causes emission. In women, the sympathetic system facilitates the smooth muscle contraction of the vagina, urethra, and uterus that occurs during orgasm.

The ANS functions outside of voluntary control and is influenced by external events (e.g., stress, drugs) and internal events (hypothalamic, limbic, and cortical stimuli). Considering these influences, it is not surprising that erection and orgasm are so vulnerable to dysfunction (Table 21.1–1).

Endocrinology. From the time of conception, hormones play a major role in human sexual development. Unlike the fetal gonads, which are under chromosomal influence, the fetal external genitalia are very susceptible to hormones. Testosterone mediates the development of the undifferentiated mesodermal Wolffian ducts into the male vas deferens, epididymis, and seminal vesicles. Dihydrotestosterone, produced from testosterone, induces the penis and scrotum. The raphe in men corresponds to the anatomical location of the vaginal orifice in women (Fig. 21.1–3).

Exogenous hormonal administration can cause external genital development inconsistent with the fetal sex gland development. For instance, if the pregnant mother receives sufficient exogenous androgen, a female fetus possessing an ovary can develop external genitalia resembling those of a male fetus. Fetal, maternal, or exogenous hormones administered to a pregnant woman may all affect the development of the external genitalia of the fetus. Deprived of male and female gonads and the respective hormones, testosterone, and estrogen, the human adult does not develop normal secondary sexual characteristics, is incapable of reproduction, and, in the case of the woman, does not develop a menstrual cycle.

Testosterone is the hormone believed to be connected with libido in both men and women. In men, stress is inversely correlated with testosterone blood concentration. Other factors, such as sleep, mood, and lifestyle influence circulating levels of the hormone. The release of testosterone in men is under the control of the hypothalamic–gonadal–pituitary axis. The hormone is secreted in a pulsatile manner and in a diurnal rhythm, with the highest levels occurring in the morning and the lowest levels in the evening. Normal concentrations range from 270 to

1,100 ng/dL. Decreased testosterone concentrations are apparent by age 50 years and decrease at the rate of approximately 100 ng/dL per decade. However, many healthy, aging men never become hypogonadal. It has also been suggested that the sensitivity of androgen receptor sites decreases in aging men.

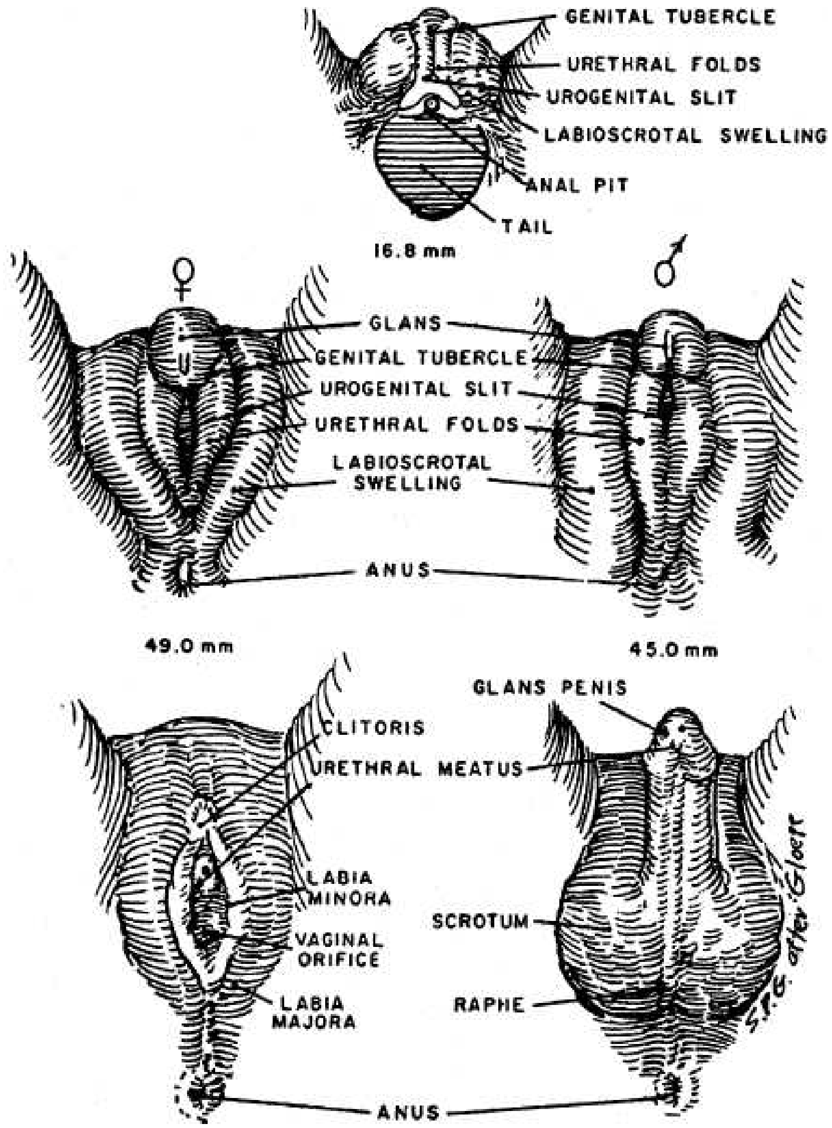
GENITAL TUBERCLE
URETHRAL FOLDS
UROGENITAL SLIT
LABIOSCROTAL SWELLING
ANAL PIT
TAIL
16.8 mm
♀
♂
GLANS
GENITAL TUBERCLE
UROGENITAL SLIT
URETHRAL FOLDS
LABIOSCROTAL SWELLING
ANUS
49.0 mm
45.0 mm
GLANS PENIS
CLITORIS
URETHRAL MEATUS
LABIA MINORA
SCROTUM
VAGINAL ORIFICE
RAPHE
LABIA MAJORA
ANUS

FIGURE 21.1–3. Differentiation of male and female external genitalia from indifferent primordial. Male differentiation occurs only in the presence of androgenic stimulation during the first 12 weeks of fetal life. (Reprinted with permission from Rhoades RA, Bell DR. *Medical Physiology: Principles for Clinical Medicine*. 6th ed. Wolters Kluwer; 2022.)

Recent studies indicate that estrogen is also a factor in the male sexual response and that a decrease in estrogen in middle-aged males results in greater fat accumulation, just as it does in women. Progesterone mildly depresses desire in men and women, as do excessive prolactin and cortisol.

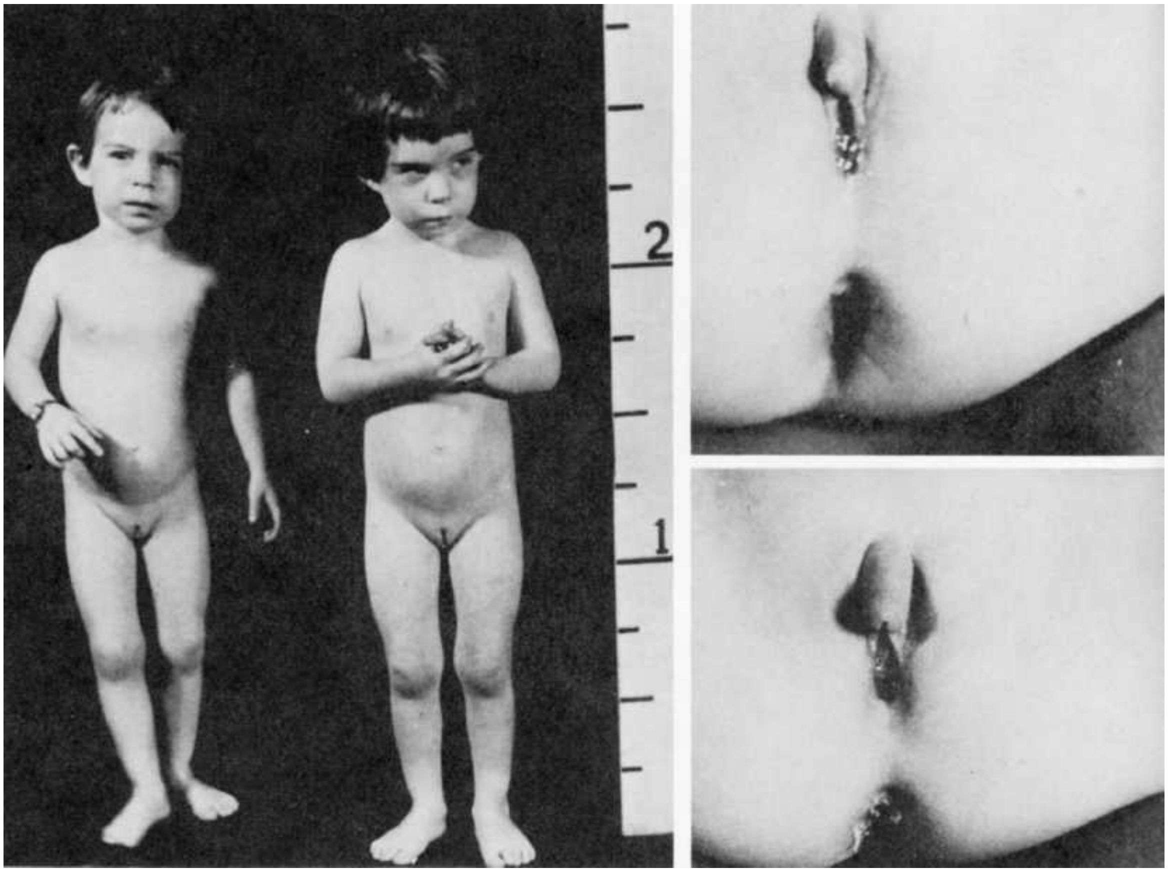

FIGURE 21.1–4. Twins born to a mother who received testosterone during pregnancy. Note the enlarged clitoris in each child. (Courtesy of Robert B. Greenblatt, M.D. and Virginia McNamarra, M.D.)

Androgen administered to men to address the loss of potency and loss of libido is usually unsuccessful unless testosterone concentrations are below normal, and administration to women may precipitate disturbing virilization. Many clinicians correct the hormone deficiency of the postmenopausal period with hormone replacement therapy (HRT). Testosterone has been used in combination with estrogen and progesterone in women who do not respond to estrogen alone. The combination is especially useful in treating headaches, depression, and reduced libido. Oxytocin is involved in pleasurable sensations during sex and is found in higher levels in men and women following an orgasm.

Diethylstilbestrol (DES), an androgenic steroid, was prescribed in the 1950s and 1960s for pregnant women with threatened abortion. However, the drug had untoward effects on the children (especially female children) born to these mothers. Reports of cervical and uterine abnormalities were reported in women, and reproductive tract abnormalities were reported in men. The children (called *DES daughters and sons*) organized a National DES Education Program sponsored by the National Cancer Institute to provide information on the potential medical problems confronting those born to DES mothers (Fig. 21.1–4).

Genetics. Genes are involved in gonadal differentiation, resulting in the formation of the bipotential gonad into either a testis or ovary. The best-defined gene in this process is *SRY* (sex-determining region of the Y chromosome), located on the short arm of the Y chromosome. Other genes also play a smaller role, such as *WT1* (Wilms tumor 1), *SF*-*1* (steroidogenic factor 1), *SOX9, DAX1,* and *MIS*-*12* (müllerian-inhibiting substance 12). Defects or mutations in these genes cause failures in gonadal differentiation that produce clinical syndromes known as *intersex disorders.* Figure 21.1–5 describes some genetic factors involved in the determination of male sex.

Central Nervous System and Sexual Behavior

Cortex. The cortex is involved both in controlling sexual impulses and in processing sexual stimuli that may lead to sexual activity. One study using positron emission tomography (PET) scans to monitor the brain activity of men in their 20s while they were shown various films, including a documentary, a comedy, and a sexually explicit film, found that some areas of the brain were more active during the sex film than they were during the other films. These areas included the orbitofrontal cortex, which is involved in emotions, the left anterior cingulate cortex, which is involved in hormone control and sexual arousal, and the right caudate nucleus, whose activity is a factor in whether sexual activity follows arousal.

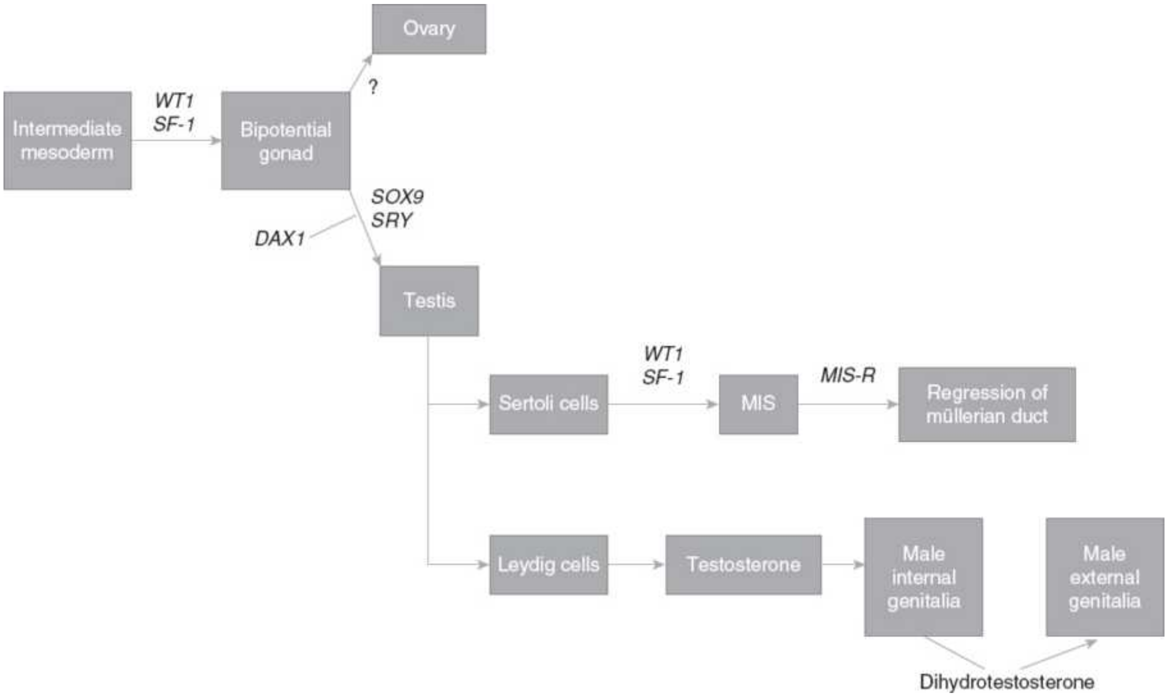
Ovary
?
Intermediate mesoderm
WT1
SF-1
Bipotential gonad
SOX9
SRY
DAX1
Testis
WT1
SF-1
Sertoli cells
MIS
MIS-R
Regression of müllerian duct
Leydig cells
Testosterone
Male internal genitalia
Male external genitalia
Dihydrotestosterone

FIGURE 21.1–5. A complex series of steps must occur in gonadal differentiation. A number of genes are critical to appropriate male genital development. *SRY* (sex-determining region of the Y chromosome), a gene on the short arm of the Y chromosome, is a testis-determining factor. The *SOX9* gene is also important in male sexual differentiation. *DAX1,* an orphan member of a nuclear hormone receptor family located on the X chromosome, interacts with steroidogenic factor 1 (*SF-1*). Other genes involved in male gonadal differentiation include the tumor-suppressor gene *WT1* (Wilms tumor 1), and the müllerian inhibiting substance gene (*MIS*) and its receptor, MIS-R.

Limbic System. Experimentation with animals has demonstrated that the limbic system is directly involved with elements of sexual functioning. In all mammals, the limbic system is involved in behavior required for self-preservation and the preservation of the species.

Chemical or electrical stimulation of various sites of the limbic system, the lower part of the septum and the contiguous medial preoptic area, the fimbria of the hippocampus, the mammillary bodies, and the anterior thalamic nuclei have all elicited penile erection.

The hippocampus is believed to influence genital tumescence and affect the regulation of the release of gonadotropins. Stimulation of the amygdala in primates initiates oral (chewing, lip-smacking) and then genital (penile erection) behavior. Researchers have stated that the closeness of these functions may derive from the evolutionary fact that the olfactory sense was strongly involved in both feeding and mating. They speculate that the evolution of the third subdivision of the limbic system may reflect a shift in importance from olfactory contact to visual communication in sociosexual behavior.

Studies of women have revealed that those areas activated by emotions of fear and anxiety are notably quiescent when the woman experiences an orgasm.

Brainstem. Brainstem sites exert inhibitory and excitatory control over spinal sexual reflexes. One site, the nucleus paragigantocellularis, has been identified as important in inhibitory control of climax-like responses in men and is suspected to be involved in female physiology as well. This nucleus projects directly to pelvic efferent neurons in the lumbosacral spinal cord, apparently causing them to secrete serotonin, which is known to inhibit orgasms. The lumbosacral cord also receives strong projections from other serotonergic nuclei in the brainstem, the raphe nuclei, pallidus, magnus, and parapyramidal region.

Brain Neurotransmitters. A vast array of neurotransmitters is produced by the brain, including dopamine, epinephrine, norepinephrine, and serotonin, among others. All affect sexual function. For example, an increase in dopamine is presumed to increase libido. Serotonin produced in the upper pons and midbrain is presumed to have an inhibitory effect on sexual function. Oxytocin, the neurohormone involved in the milk ejection reflex, is released with orgasm and is believed to reinforce pleasurable activities.

Spinal Cord. As described in the innervation of sex organs, sexual arousal and climax are ultimately organized at the spinal level. Sensory stimuli related to sexual function are conveyed via afferents from the pudendal, pelvic, and hypogastric nerves. These afferents terminate in the medial portions of the dorsal horn and in the medial central gray matter of the spinal cord. Several separate experiments suggest that sexual reflexes are mediated by spinal neurons in the central gray region of the lumbosacral segments.

BIOLOGIC DIFFERENCES BETWEEN THE SEXES

Hormonal Differences

The effects of estrogen in the female and androgen in the male are necessary for the development of ova and spermatozoa, respectively, and for potential fertility. They underlie and define the development of secondary sex characteristics at puberty. Both sexes initially synthesize androgens such as androstenedione or testosterone from acetate or cholesterol in the ovary and testes. Estrogen is then synthesized from the androgen. The two most potent forms of these hormones are estradiol and testosterone.

The defining hormonal difference between women and men is the quantity of these hormones present in the sexes: the testis makes approximately 7,000 ng of testosterone per day and converts one-quarter of 1% to estradiol. The ovary makes only 300 ng of testosterone per day and converts 50% of it to estradiol. Testosterone production in the ovary may vary during the menstrual cycle, but as testosterone production increases, so does the percentage of it, which is converted to estradiol. Overall, men make 20 times as much androgen as do women and women convert 200 times as much androgen to estradiol as do men. The contrasts are further emphasized because estrogen levels are measured in picograms, while androgen levels are measured in nanograms.

Intracellular hormone metabolism also contributes to sex differentiation. Many cells contain enzymes that can convert androgen to estrogen or testosterone to its more potent metabolite, dihydrotestosterone. Also present in certain cells are proteins, which influence the binding of sex hormones to elements affecting transcription sites in relevant DNA.

Ultimately, each tissue can develop an individual androgenic or estrogenic identity that might not be assumed, predicted, or distinguished on the basis of circulating hormone levels. These cellular identities may be particularly important in the brain and may have relevance to sexual behavior, sexual identity, partner choice, and brain dimorphism.

Genetic Differences

Human cells contain 23 pairs of chromosomes for a total of 46. There are 22 pairs of autosomes and 1 pair of sex chromosomes. The sex chromosomes are the X (female) chromosome and the Y (male) chromosome. These chromosomes determine biologic sex. Genetically, the normal male has a chromosomal 46XY karyotype, and the normal female has a 46XX karyotype. The X chromosome carries approximately 1,090 genes, while the human Y chromosome contains only about 80 genes. Further, the Y chromosome contains a unique class of genes in the male-specific region, particularly associated with testicular function. Also, the genes on the Y chromosome have twice the mutation frequency of those on the X chromosome.

The startling difference in the number of genes carried by a 2X woman as contrasted with a single X man is moderated by the fact that the second X chromosome in women is largely inactivated sometime before there are 20 cells in the blastocyst. However, 15% of the genes on the second X escape inactivation, and another 20% are only partially inactivated. The inactivated X chromosome is converted from a linear X into the Barr body, a little ball of genetically inert DNA.

Sex determination is induced by the different karyotypes. In the male, testis growth is induced by the sex-determining region of the Y chromosome. In the female, the follicle surrounding her oocytes is stabilized by the second X. Further differences in the ways the XX and XY genotypes influence early differentiation and later development are being studied.

PSYCHOSEXUALITY

Sexuality and total personality are so entwined that it is virtually impossible to speak of sexuality as a separate entity. The term *psychosexual* is therefore used to describe personality development and functioning, as these are affected by sexuality. *Psychosexual* applies to more than sexual feelings and behavior, and it is not synonymous with libido in the broad Freudian sense.

Freud’s generalization that all pleasurable impulses and activities are originally sexual has given lay people a somewhat distorted view of sexual concepts and has given psychiatrists a confused picture of motivation. For example, some oral activities are directed toward obtaining food, and others are directed toward achieving sexual gratification. Both activities are pleasure seeking and use the same organs, but they are not, as Freud contended, both necessarily sexual. Labeling all pleasure-seeking behaviors *sexual* obviates specifying precise motivation. People may also use sexual activities to gratify nonsexual needs, such as dependency, aggression, power, and status. Although sexual and nonsexual impulses may jointly motivate behavior, the analysis of behavior depends on understanding the underlying individual motivation and their interactions.

Sexuality is something more than physical sex, coital or noncoital, and something less than all behaviors directed toward gaining pleasure.

Psychosexual Factors

Sexuality depends on four interrelated psychosexual factors: sexual identity, gender identity, sexual orientation, and sexual behavior. These factors affect personality, development, and functioning.

Sexual identity is the pattern of a person's biologic sexual characteristics: chromosomes, external and internal genitalia, hormonal composition, gonads, and secondary sex characteristics. In normal development, these characteristics form a cohesive pattern that leaves individuals in no doubt about their sex.

Gender identity is an individual's sense of maleness or femaleness. By the age of 2 or 3 years, most children have a firm conviction that "I am a boy" or "I am a girl." Gender identity results from an almost infinite series of clues derived from experiences with family members, peers, and teachers and from cultural phenomena. For instance, male infants tend to be handled more vigorously, and female infants tend to be cuddled more. Fathers tend to be more aware of their sons' adolescent concerns than of their daughters' anxieties. Boys are more likely to be physically disciplined than girls are. A child's sex may affect parental tolerance for aggression and for reinforcement or extinction of activity and of intellectual, aesthetic, and athletic interests. Physical characteristics derived from a person's biologic sex (e.g., physique, body shape, and physical dimensions) interrelate with an intricate system of stimuli, including rewards, punishment, and parental gender labels, to establish gender goals. Studies of children with intersex conditions have drawn attention to a physiologic basis for gender identity. In particular, they have focused on the masculinization or feminization of the fetal brain. Some people claim a gender identity at odds with their biologic attributes. Others possess a fluid gender identity or one that does not fit into any recognized category, which they call "other."

Gender Role. Related to, and in part derived from, gender identity is gender role behavior. John Money and Anke Ehrhardt described gender role behavior as all those things that a person says or does to disclose themselves as having the status of boy or man, girl or woman, respectively. A gender role is not established at birth, but it is built up cumulatively through (1) experiences encountered and transacted through casual and unplanned learning, (2) explicit instruction and inculcation, and (3) spontaneously putting two and two together to sometimes make four and sometimes five. The usual outcome is a congruence of gender identity and gender role. Although biologic attributes are significant, the major factor in achieving the role the culture considers appropriate to a person's sex is learning.

Sexual orientation describes the object of a person's sexual impulses: heterosexual (opposite sex), homosexual (same-sex), or bisexual (both sexes). In some people, there is fluidity regarding objects of attraction. The majority of people have a heterosexual orientation. In the United States, 2.8% of men and 1.4% of women identify themselves as homosexual. These numbers are compatible with figures from Western European countries as well. However, a higher percentage of persons have had at least one same-sex experience in their lives. In addition, homosexuals congregate in urban areas, so the incidence of homosexuality in some large cities is 8% or 9%. A group of people have defined themselves as "asexual" and assert this as a positive identity. Some researchers believe this lack of attraction to any object is a manifestation of a desire disorder. Other people wish not to define their sexual orientation at all and avoid labels. Still, others describe themselves as polysexual or pansexual and claim a fluid identity. The percentage of people in the United States defining themselves as homosexual, bisexual, asexual, transgender or fluid has doubled in the past decade to 7%.

Sexual behavior includes desire, receptivity, fantasies, the pursuit of partners, autoeroticism, and all the activities engaged in to express and gratify sexual needs. It is an amalgam of psychological and physiologic responses to internal and external stimuli.

Table 21.1–2.
Female Sexual Response Cycle[a]

Organ	Excitement Phase	Orgasmic Phase	Resolution Phase
Mental	Lasts several minutes to several hours; heightened excitement before orgasm, 30 seconds to 3 minutes	3–15 seconds	10–15 minutes; if no orgasm, 0.5–1 day
Skin	Just before orgasm: sexual flush inconsistently appears; maculopapular rash originates on abdomen and spreads to anterior chest wall, face, and neck; can include shoulders and forearms	Well-developed flush	Flush disappears in reverse order of appearance; inconsistently appearing film of perspiration on soles of feet and palms of hands
Breasts	Nipple erection in two-thirds of women, venous congestion, and areolar enlargement; size increases to one-fourth more than normal	Breasts may become tremulous	Return to normal in approximately 0.5 hour
Clitoris	Enlargement in diameter of glans and shaft; just before orgasm, shaft retracts into prepuce	No change	Shaft returns to normal position in 5–10 seconds; detumescence in 5–30 minutes; if no orgasm, detumescence takes several hours

Labia majora	Nullipara: elevate and flatten against perineum Multipara: congestion and edema	No change	Nullipara: increase to normal size in 1–2 minutes Multipara: decrease to normal size in 10–15 minutes
Labia minora	Size increases two to three times more than normal; change to pink, red, deep red before orgasm	Contractions of proximal labia minora	Return to normal within 5 minutes
Vagina	Color change to dark purple; transudate appears 10–30 seconds after arousal; elongation and ballooning; lower third constricts before orgasm	3–15 contractions of lower third at intervals of 0.8 seconds	Ejaculate forms seminal pool in upper two-thirds; congestion disappears in seconds or, if no orgasm, in 20–30 minutes
Uterus	Ascends into false pelvis; labor-like contractions begin in heightened excitement just before orgasm	Contractions throughout orgasm	Contractions cease, and uterus descends to normal position
Other	Myotonia	Loss of voluntary muscular control	Return to baseline status in seconds to minutes
	A few drops of mucoid	Rectum: rhythmic	Cervix color and

secretion from Bartholin glands during heightened excitement	contractions of sphincter	size return to normal, and cervix descends into seminal pool
Cervix swells slightly and is passively elevated with uterus	Hyperventilation and tachycardia	

---

[a] A desire phase consisting of sex fantasies and desire to have sex often precedes the excitement phase.

Physiologic Responses

Sexual response is a true psychophysiological experience. Arousal is triggered by both psychological and physical stimuli; levels of tension are experienced both physiologically and emotionally, and with orgasm, normally a subjective perception of a peak of physical reaction and release occurs along with a feeling of well-being. Physical health, psychosexual development, psychological attitudes toward sexuality, and attitudes toward one's sexual partner are directly involved with and affect the physiology of human sexual response.

Normally, men and women experience a sequence of physiologic responses to sexual stimulation. In the first detailed description of these responses, Masters and Johnson observed that the physiologic process involves increasing levels of vasocongestion and myotonia (tumescence) and the subsequent release of the vascular activity and muscle tone as a result of orgasm (detumescence). In men, ejaculation is a normal part of the climax.

Refractory Period. Following orgasm, men experience a refractory period during which they cannot be restimulated to further arousal. The length of this period is age and situation dependent, with younger age and novel situations usually resulting in a shorter refractory period. Women do not have a refractory period and can experience multiple, successive orgasms.

Tables 21.1–2 and 21.1–3 describe the physiologic changes in the male and female that accompany sexual arousal, orgasm, and, subsequently, resolution back to the prestimulation baseline.

It is important to remember that the sequence of responses can overlap and fluctuate. In addition, a person's subjective experiences are as important to sexual satisfaction as the objective physiologic response. Figures 21.1–6 and 21.1–7 illustrate several possible patterns in the phases of the male sexual response and female sexual response, respectively.

Table 21.1–3.
Male Sexual Response Cycle[a]

Organ	Excitement Phase	Orgasmic Phase	Resolution Phase
Mental	Lasts several minutes to several hours; heightened excitement before orgasm, 30 seconds to 3 minutes	3–15 seconds	10–15 minutes; if no orgasm, 0.5 to 1 day
Skin	Just before orgasm: sexual flush inconsistently appears; maculopapular rash originates on abdomen and spreads to anterior chest wall, face, and neck, and can include shoulders and forearms	Well-developed flush	Flush disappears in reverse order of appearance; inconsistently appearing film of perspiration on soles of feet and palms of hands
Penis	Erection in 10–30 seconds caused by vasocongestion of erectile bodies of corpus cavernosa of shaft; loss of erection may occur with introduction of asexual stimulus—e.g., loud noise; with heightened excitement, size of glans and	Ejaculation: emission phase marked by three to four contractions of 0.8 seconds of vas, seminal vesicles, prostate; ejaculation proper marked by contractions of 0.8 seconds of urethra and ejaculatory spurt of 12–20 in at age 18, decreasing with	Erection: partial involution in 5–10 seconds with variable refractory period; full detumescence in 5–30 minutes

	diameter of penile shaft increase further	age to seepage at age 70	
Scrotum and testes	Tightening and lifting of scrotal sac and elevation of testes; with heightened excitement, 50% increase in size of testes over unstimulated state and flattening against perineum, signaling impending ejaculation	No change	Decrease to baseline size because of loss of vasocongestion; testicular and scrotal descent within 5–30 minutes after orgasm; involution may take several hours if no orgasmic release takes place
Cowper glands	2–3 drops of mucoid fluid that contain viable sperm are secreted during heightened excitement	No change	No change
Other	Breasts: inconsistent nipple erection with heightened excitement before orgasm	Loss of voluntary muscular control	Return to baseline state in 5–10 minutes
		Rectum: rhythmic contractions of sphincter	
	Myotonia: semispastic contractions of facial, abdominal,	Heart rate: up to 180 beats/min	

and intercostal muscles	Blood pressure: up to 40–100 mm systolic; 20–50 mm diastolic
Tachycardia: up to 175 beats/min	
Blood pressure: rises to 20–80 mm systolic; 10–40 mm diastolic	Respiration: up to 40 respirations/min
Respiration: increased	

[a] A desire phase consisting of sex fantasies and desire to have sex precedes the excitement phase.

Gender Differences in Desire and Erotic Stimuli

Sexual impulses and desires exist in men and women. In measuring desire by the frequency of spontaneous sexual thoughts, interest in participating in sexual activity, and alertness to sexual cues, males generally possess a higher baseline of desire than do women, which may be biologically determined. In the male, a desire phase consisting of sex fantasies and the desire to have sex usually motivates initiation of the sex act and precedes arousal. Motivations for having sex, other than desire, exist in both men and women but seem to be more varied and prevalent in women. In women, they may include a wish to reinforce the pair bond, the need for a feeling of closeness, a way of preventing the man from straying, or a desire to please the partner. Also, women, particularly those in long-term relationships, may experience desire synchronously with, or after the onset of, arousal rather than before it. Receptivity to the sex act seems to be as important a prelude to sexual behavior in women as desire. Recent studies have indicated variations in the male response pattern as well, primarily a conflation of the sensation of the experience of desire and arousal.

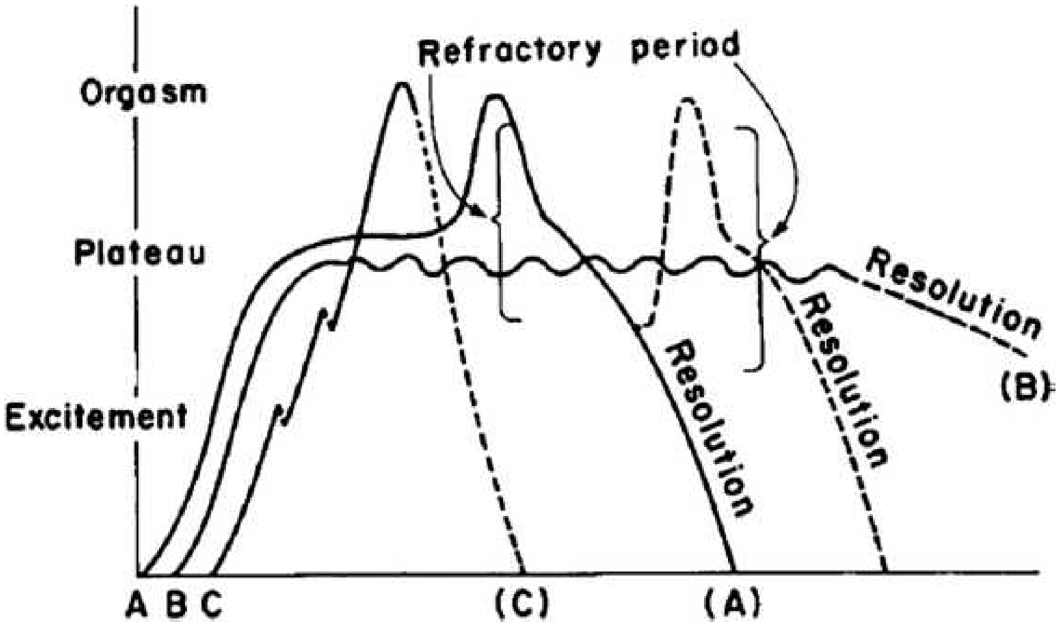

FIGURE 21.1–6. Male sexual response. An individual man may experience any of these three patterns (**A, B,** or **C**) during a particular sexual experience. (Reprinted with permission from Walker JI. *Essentials of Clinical Psychiatry*. JB Lippincott; 1985:276.)

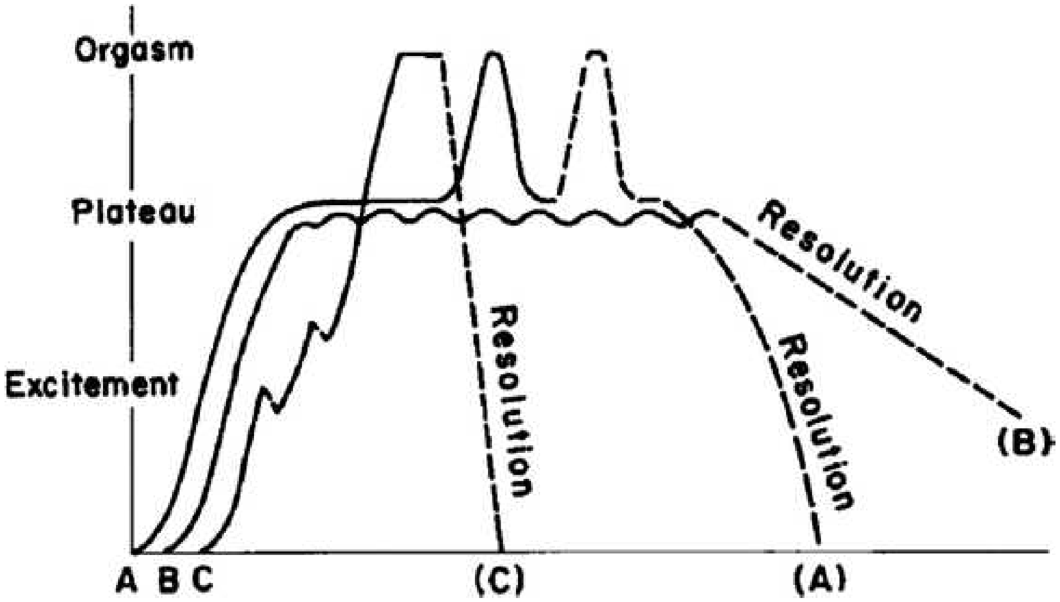

FIGURE 21.1–7. Female sexual response. An individual woman may experience any of these three patterns (**A, B,** or **C**) during a particular sexual experience. (Reprinted with permission from Walker JI. *Essentials of Clinical Psychiatry*. JB Lippincott; 1985:276.)

Although explicit sexual fantasies are common to both sexes, the external stimuli for the fantasies frequently differ for men and women.

Many men respond sexually to visual stimuli of nude or barely dressed women. Women report responding sexually to romantic stories with an erotic component and a hero whose passion for the heroine impels him toward a lifetime commitment to her. The popular erotic novel, Fifty Shades of Gray and the streaming series Bridgerton evolve with just such a plot.

A complex component of female sexuality is that a woman's subjective sense of arousal is not always congruent with her physiologic state of arousal. Specifically, her sense of excitement may reflect a readiness to be aroused rather than physiologic lubrication. Conversely, she may experience signs of arousal, including vaginal lubrication, without being aware of them or without wanting to act on them. These instances rarely occur in men.

Masturbation

Masturbation usually is a normal precursor of object-related sexual behavior and a form of sexual pleasure that generally lasts throughout a person's lifetime. No other form of sexual activity has been as universally practiced in spite of being severely condemned by many cultures for long periods of time.

In the classical period, Greco-Roman writers and medical authorities, such as Galen, recommended masturbation as a healthful practice for both men and women. The categorization of masturbation as sinful in Western culture derives from Judeo-Christian attitudes. Masturbation is sometimes called *onanism* after Onan in the Old Testament, who was slain for spilling his seed upon the ground. Many biblical scholars today posit that this punishment was not a result of his masturbation but because he did not obey Jehovah's commandment to take his brother's wife as his own. The prohibition against masturbation was reinforced by Christian Church fathers, particularly by St. Augustine, who preached celibacy and held that sex was appropriate only for purposes of procreation. Ambivalence about masturbation also existed in Eastern cultures. Some Hindus and the ancient Chinese believed that loss of semen resulted in a reduction of vital male essence. However, the prohibitive emphasis was on ejaculation and not on manipulation of the genitalia. Female masturbation was better tolerated or ignored.

For many years, science held attitudes toward masturbation that were as negative as many religious views. What had been viewed as sinful came to be seen as pathologic (Fig. 21.1–8). For example, in the 1800s, Krafft-Ebing believed that masturbation could lead to insanity. Currently, and in the second half of the 20th century, a liberal attitude toward sexual behavior, including masturbation, has prevailed but is consistently challenged by more restrictive thinking. Research by Kinsey in the 1940s into the prevalence of masturbation indicated that nearly all men and three-fourths of all women masturbate sometime during their lives.

Longitudinal studies of development show that sexual self-stimulation is common in infancy and childhood. Just as infants learn to explore the functions of their fingers and mouths, they learn to do the same with their genitalia. At about 15 to 19 months of age, both sexes begin genital self-stimulation. Pleasurable sensations result from any gentle touch to the genital region. Those sensations, coupled with an ordinary desire for exploration of the body, produce a normal interest in masturbatory pleasure at that time. Children also develop an increased interest in the genitalia of others—parents, children, and animals. As youngsters acquire playmates, the curiosity about their own and others' genitalia motivates episodes of exhibitionism or genital exploration. Such experiences, unless blocked by guilty fear or shaming, contribute to continued pleasure from sexual stimulation.

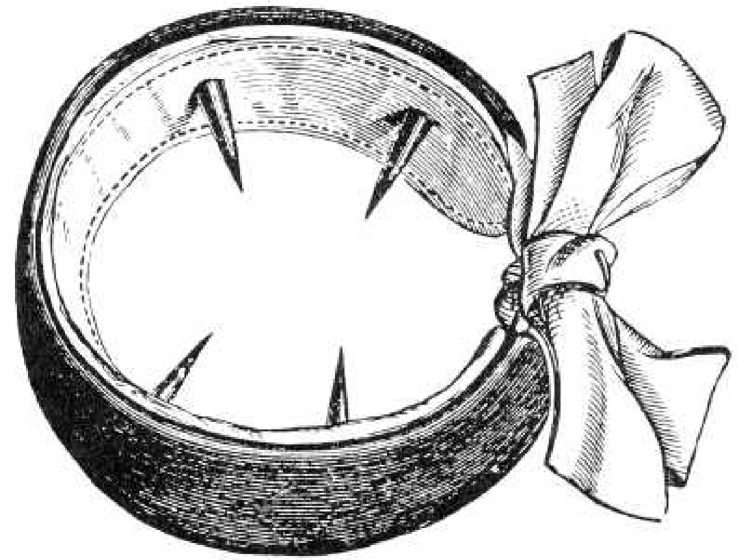

FIGURE 21.1–8. A four-pointed penile ring, which was used to prevent masturbation during the 19th century. This engraving shows how the ring was tied to the penis. (Reprinted with permission from Sadock BJ, Kaplan HI, Freedman AM. *The Sexual Experience*. Williams & Wilkins; 1976.)

With the approach of puberty, the upsurge of sex hormones, and the development of secondary sex characteristics, sexual curiosity intensifies, and masturbation increases. Adolescents are physically capable of coitus and orgasm but are usually inhibited to some degree by social restraints. The dual and often conflicting pressures of establishing their sexual identities and controlling their sexual impulses produce a strong physiologic sexual tension in teenagers. Masturbation is a normal way to reduce that tension. In general, males learn to masturbate to orgasm earlier than females and masturbate more frequently. An important emotional difference between the adolescent and the youngster of earlier years is the presence of coital fantasies during masturbation in the adolescent. These fantasies are an important adjunct to the development of sexual identity; in the comparative safety of the imagination, the adolescent learns to perform the adult sex role. This autoerotic activity is usually maintained into the young adult years when it is normally replaced by coitus.

Couples in a sexual relationship do not abandon masturbation entirely. When coitus is unsatisfactory or is unavailable—because of illness or other unavailability of a partner—self-stimulation can serve an adaptive purpose, combining sensual pleasure and tension release. Studies of masturbation in married individuals have found different motivations for the activity in men and women. Partnered men frequently masturbate if they are dissatisfied with the partnered sex. For partnered women, masturbation is not a partner substitute but seems to indicate comfort with their own desires, reflects a high level of sexual interest, and often leads to increased partnered sex. In general, masturbation has come to be seen as a separate source of sexual pleasure that is not in conflict with partnered sex.

[Several studies found that in men, orgasm raised the serum prostate-specific antigen (PSA) significantly. Males scheduled for PSA tests should be advised not to masturbate (or have coitus) for at least 7 days prior to the examination.]

Moral taboos against masturbation have generated myths that masturbation causes mental illness, decreased sexual potency, or infertility. No scientific evidence exists to support such claims. Masturbation is a psychopathologic symptom only when it becomes a compulsion beyond an individual’s willful control. Then, it is a symptom of disturbance, not because it is sexual, but because it is compulsive. It is also symptomatic of a sexual problem when it is the main sexual activity of a person with an available intimate partner. Masturbation is probably a universal aspect of psychosexual development, and, in most cases, it is adaptive.

SEXUAL LEARNING AND SEXUAL BEHAVIOR

The sex drive is innate and varies in intensity in different people, but much sexual behavior is learned. Early sexual experiences, particularly those during puberty and early adolescence, can have an imprinting effect. If they are strongly associated with pleasure and release of tension, they are likely to be repeated, and the person is conditioned to a particular form of sexual expression. In the normal person, sexual learning and experimentation continue throughout the life cycle, the repertoire of sexual behavior expands, and the behaviors are compatible with cultural norms.

Sexual learning begins in childhood. In a broad sense, learning occurs through parent–child interaction, including the meeting of the infant's needs, cuddling, and the reinforcement or discouragement of gender-associated activities. Cuddling and appropriate physical touching engender emotional security and positive feelings in infants toward their bodies. That physicality lays the groundwork for a healthy body image that is a component of sexual self-esteem. A good body image also derives from mastery of early physical activity and a positive parental approach to tasks such as toilet training.

Genital self-stimulation is a normal activity of babies. It is particularly pronounced between the ages of 15 and 19 months, and it is part of the general interest of children in their bodies. The activity is reinforced by the pleasurable sensations it produces. Children also learn by watching the interaction of their parents. They observe demonstrations of physical affection (although usually not the sex act itself), and they are sensitive to the sexual undertone of flirtation, bantering, or seductive interchange.

Sexual learning in childhood is adversely affected by exploitative, abusive adults. Incestuous activity, with or without penetration, or sexual abuse by other adults (babysitters, other older relatives, teachers, coaches, etc.) is damaging to the child. It may precipitate sexual dysfunction in adult life, as well as promiscuity and problems with intimacy. It may produce delinquent activities and behavioral problems in adolescence, as well as learning problems in school.

Some children are inappropriately stimulated short of outright incest or genital fondling. They are the objects of extremely seductive parents. Termed *eroticized children,* they frequently become sexually precocious.

Children who view the *primal scene*—the term Freud used to describe a child seeing sex between their parents—may be traumatized. This is particularly the case if they interpret coital sounds and movements as physical aggression between the parents.

Adolescence

With the approach of puberty, the upsurge of sex hormones, and the development of secondary sex characteristics, sexual curiosity is intensified.

In addition to masturbation, adolescents learn sexually through caressing and kissing with partners. In early adolescence, sex play may involve a partner of the same sex for a short period of time for heterosexuals as well as for homosexuals. Conversely, homosexuals may experiment with the opposite sex, although this occurs more frequently with lesbians than with homosexual men. Adolescence is also the time when one's body image becomes more definitive, and a sense of sexual adequacy and sexual desirability begins to develop. Peer acceptance by the same sex and by the opposite sex is of paramount importance. Engaging in sexual talk and jokes, kissing, stimulating genitalia manually or orally, experimenting with degrees of nudity, and experimenting with different partners or with one partner are part of the process of learning about sexuality. These experiences reinforce the adolescent's sense of being a sexual male or female. There is significant variability in the time of onset and degree of sexual experimentation in adolescence. It varies according to the physiology of the teen and the values of one's immediate community and peers.

In the recent past, a "hook-up" culture has been described among college students that is replacing traditional dating. It involves casual sex, that is, sex outside a committed or romantic relationship that may occur between friends ("friends with benefits") or as a one-night or occasional-night event. However, a survey contrasting student cohorts from 1988 to 1996 and 2004 to 2012 found respondents from the latter era did not report more sexual partners since age 18, more frequent sex, or more frequent sex partners during the previous year than respondents from the earlier era.

First Coitus

The first coitus is a rite of passage for both sexes. In the United States, the majority of people have experienced coitus by young adulthood, by their early 20s. In a study of people aged 18 to 59 years, over 95% had included coitus in their last sexual interaction.

The young man experiencing intercourse for the first time is vulnerable in his pride and self-esteem. Cultural myths still perpetuate the idea that he should be able to have an erection with no or little stimulation and that he should have an easy mastery over the situation, even though it is an act that he has never before experienced. Cultural pressure on the woman with her first coitus reflects the remaining cultural ambivalence about her loss of virginity, despite the current era of sexual liberality. This is demonstrated in the statistic that only 50% of young women use contraception during their first coitus, and of that 50%, an even smaller number use it consistently thereafter. Young women with a history of masturbation are more likely to approach intercourse with positive anticipation and confidence. Studies have shown that young women are more likely to enjoy their first coitus if it occurs in the context of a positive emotional relationship.

Adulthood

By their late 20s, 70% of men and 85% of women have formed a union, either exclusive cohabitation or marriage. The sexual impulse is catalytic in forming and maintaining adult love relationships. Ideally, a mature sexual relationship encompasses the capacity for intimacy and love for one’s partner.

As sexual access ceases to be problematic, more attention is focused on the activity itself. Interference with sexual activity arises from the time and energy required for the pursuit of careers, child-rearing, and other family and community obligations. Nonetheless, studies reveal a much higher frequency of sexual interaction among married persons than among single persons. The modal frequency for married persons is three times a month, with many couples relating sexually twice a week. The highest frequency of sexual interaction, four or more times a week, exists among cohabiting couples.

Kissing, intercourse, and masturbation are frequent sexual activities. Many adults have oral sex. Studies show that, although it is not a regular activity for most couples, many people are likely to engage in it occasionally throughout their lifetimes. The incidence of the practice of cunnilingus and fellatio are the same; the lifetime prevalence is 75% of sexually active people. Anal sex is not part of the regular sexual practices of most Americans. However, the incidence of the activity is rising. In one study, 35% of heterosexual men and 30% of heterosexual women reported having anal sex when asked about their sexual activities during the preceding year. In general, a graph of sexual practices does not form a bell curve. Most curves of sexual behavior are strongly skewed, with many persons indulging in a particular behavior or, conversely, very few people doing so.

Middle Age

During middle age, the frequency of marital intercourse may decline. The rates of interaction depend more on male interest, and the middle-aged man may devote much of his energy to his career at this point. The decline in sexual interaction often reflects the de-eroticization of the woman because of her wife–mother role. Familiarity also contributes to a decreased passion, and some men have difficulty connecting marital sexual activity with the erotic scripts they elaborated during their adolescence. Conversely, a couple's experience of each other makes them comfortable and augments their skill at mutual arousal. For women, interest in sexual competence typically increases at this time. A woman's erotic pleasure and sexual commitment are strongly connected to feelings of attachment toward her partner and the security of being in a loving relationship.

With late middle age, the biologic drive decreases in intensity. It takes the man longer to reach orgasm, he has a longer refractory period, and he requires more stimulation to achieve an erection. The woman must adjust to the hormonal fluctuations of her perimenopausal years, and she, too, requires more stimulation to become aroused. Although medications are available to address these specific physiologic changes, they do not eliminate all need to adjust to the aging process. In terms of family dynamics, children often leave the house for work or further education at this time. This fact plus freedom from concern regarding unwanted pregnancy may enhance sexual activity for persons who have more time to direct attention toward each other and renew their life as a couple.

However, middle age is the period of rising extramarital activity, although the frequencies of such occurrences are low—75% of men and

85% of women remain faithful throughout the lifetime of their marriages. Daniel Offer has explained that the patterning of extramarital sexual activity for both sexes continues to express earlier patterns of psychosexual development:

> For men, it predominantly has the capacity for detachment that in adolescence was directly related to the pursuit of sexual fantasies and the homosocial validation of masculinity. For women, on the other hand, it resembles a quest for circumstances that justify and confirm a romantic self-image rather than a quest for orgasms.

In the context of a loving, communicative, and committed relationship, a decreased frequency of sexual interaction does not herald the onset of extramarital affairs or threaten the stability of the marriage.

Old Age

In the last two decades, coitus has also been a part of the sexual repertoire of older adults due to the development of sildenafil-type drugs, which facilitate erections in men, and hormonally enhanced creams or hormonal pills, which counteract vaginal atrophy in postmenopausal women. Prior to the development of these drugs, and still today, many older adults enjoyed gratifying sex play, exclusive of coitus. An aversion to the idea of sex in the older adults persists in today’s culture. A man who had coitus with his wife while she was in a nursing home and suffered from a degree of dementia was accused by his stepchildren of sexual assault and brought to trial. He was found innocent.

An estimated 70% of men and 20% of women over age 60 years are sexually active. Recent studies have found that three in four people aged 57 to 64 years, two in four people aged 64 to 75 years, and one in four people aged 75 to 85 years had sexual intercourse in the year before the study. Of interest is that the reported frequency of intercourse, two or three times a month, is the same as the frequency reported by young and middle-aged adults. Expected physiologic changes in men include a longer time for an erection to occur, decreased penile turgidity, and ejaculatory seepage; in women, decreased vaginal lubrication and vaginal atrophy are associated with lower estrogen levels. Illness and medications can adversely affect sexual behavior. However, many older adults engage in sexual behavior exclusive of coitus which provides them with physical connection and satisfaction.

In studies, approximately half of older sexually active people aged 59 to 85 years report one or more sexual dysfunctions. Women report problems of low desire, difficulty with lubrication, and inability to climax. Among men, erectile disorder is the most commonly reported dysfunction. About 14% of older men report using medications or supplements to improve their erections. The factors contributing to continuing sex life in older adults include the presence of a partner, the absence of severe illness, and a history of sexual activity in young and middle adulthood. Physicians should include sexual behavior in their review of systems with older age patients since the majority of people in this age group will not bring up their concerns unless their doctor initiates the discussion. One study reported that over 85% of older adults would like doctors to include questions about sex during routine visits.

LOVE AND INTIMACY

Freud postulated that psychological health could be determined by a person's ability to function well in two spheres, work, and love. A person able to give and receive love with a minimum of fear and conflict has the capacity to develop genuinely intimate relationships with others. A desire to maintain closeness to the love object typifies being in love. Mature love is marked by intimacy that is a special attribute of the relationship between two persons. When involved in an intimate relationship, the person actively strives for the growth and happiness of the loved person. Sex frequently acts as a catalyst in forming and maintaining intimate relationships. The quality of intimacy in a mature sexual relationship is what Rollo May called "active receiving," in which a person while loving, permits themselves to be loved. May describes the value of sexual love as an expansion of self-awareness, the experience of tenderness, an increase of self-affirmation and pride, and sometimes, at the moment of orgasm, loss of feeling of separateness. In that setting, sex and love are reciprocally enhancing and healthily fused.

Some persons experience conflicts that prevent them from fusing tender and passionate impulses. This can inhibit the expression of sexuality in a relationship, interfere with feelings of closeness to another person, and diminish a person's sense of adequacy and self-esteem. When these problems are severe, they may prevent the formation of, or commitment to, an intimate relationship.

Sex and the Law

Medicine and the law both assess the impact of sexuality on the individual and society and determine what healthy or legal behavior is. Appropriateness or legality of sexual behavior, however, is not always viewed the same way by professionals in both disciplines. The issues at the interface of sexual science and the law often are emotionally charged and reflect cultural divisions about acceptable sexual mores. They include attitudes about abortion, pornography, prostitution, sex education, the homosexual and transgender community, the treatment of sex offenders, and the right to sexual privacy, among other issues. Laws regarding these issues (e.g., criminalization of oral or anal sex by consenting adults or the need for parental permission by minors who are requesting an abortion) vary from state to state.

Taking a Sex History

A sex history provides important information about patients, regardless of the presence of a sexual disorder or whether that is the patient's chief complaint. The information can be gradually obtained as therapy progresses and through open-ended questions. It is important for the interviewer to be nonjudgmental and sensitive to many patients' embarrassment in discussing sexual issues. As part of the sexual history, information regarding safe sex practices, sexually transmitted diseases, sexual orientation, gender identity, and history of any sexual abuse must be ascertained. The percentage of sexual abuse survivors seen in sex therapy programs is higher than that in the general population. The outline in Table 21.1–4 provides a guide to areas to be covered and a structure that can be used when time is limited.

Sexual Dysfunctions

The essential features of sexual dysfunctions are inhibition, specifically, an inability to respond to sexual stimulation physiologically or psychologically, or the experience of pain during the sexual act. The pain may have a physiologic, psychological, or mixed etiology. Dysfunction can be defined by a disturbance in the subjective sense of pleasure usually associated with sex or by objective performance.

According to the 11th revision of the *International Classification of Diseases* (ICD-11), sexual dysfunctions are syndromes that refer to a person's difficulty "experiencing personally satisfying, non-coercive sexual activities." The definition acknowledged the complex nature of the sexual response, incorporating psychological, interpersonal, social, cultural, and physiologic factors, and notes that any one of these factors can affect the response.

Definition

In the DSM-5, sexual dysfunctions include male hypoactive sexual desire disorder, female sexual female interest/arousal disorder, erectile disorder, female orgasmic disorder, delayed ejaculation, early (premature) ejaculation, genitopelvic pain/penetration disorder, substance/medication-induced dysfunction, other specified sexual dysfunction, and unspecified sexual dysfunction. Sexual dysfunctions are diagnosed only when they are a major part of the clinical picture. If more than one dysfunction exists, they should all be diagnosed. Sexual dysfunctions can be lifelong or acquired, generalized or situational, and result from psychological factors, physiologic factors, and combined factors, and numerous stressors, including prohibitive social mores, health and partner issues, and relationship conflicts. If the dysfunction is attributable entirely to a medical condition, substance use, or adverse effects of medication, then sexual dysfunction due to a general medical condition or substance-induced dysfunction should be diagnosed. In DSM-5, the specification of severity of the dysfunction is indicated by noting whether the symptomatology or the patient’s distress is mild, moderate, or severe.

Sexual dysfunctions are frequently associated with other mental disorders, such as depressive disorders, anxiety disorders, personality disorders, and schizophrenia. In many instances, the sexual dysfunction may be diagnosed in conjunction with another psychiatric disorder. If the dysfunction is largely attributable to an underlying psychiatric disorder, only the underlying disorder should be diagnosed. Sexual dysfunctions are usually self-perpetuating, with the patient increasingly subject to ongoing performance anxiety and a concomitant inability to experience pleasure. In relationships, the sexually functional partner may reacts with distress or anger due to feelings of deprivation or a fear that they are an insufficiently attractive or adequate sexual partner. In such cases, the clinician must consider whether the sexual problem preceded or arose from relationship difficulties and weigh whether a diagnosis of sexual dysfunction relevant to relationship issues is appropriate.

The sexual problem offers the psychiatrist a perspective on the person’s capacity to give and receive pleasure, to love and be loved, to be psychologically intimate, and to maintain some degree of sexuality through the vicissitudes of life.

Similar disorders exist among both homosexual and heterosexual partners, with variations imposed by anatomical differences. For example, although penile–vaginal dysfunction cannot be described among homosexuals, penile–anal dysfunction may occur. Regardless of sexual orientation, the physiologic response to sexual stimulation applies equally to same-sex and heterosexual partners, and the methods and principles for treatment are essentially similar.

Rating Scales

Several questionnaires are available to assess sexual function, in addition to the assessments made following the taking of a sexual history. They were developed primarily to evaluate illness or medication-related effects on sexual functioning. The most commonly used scales are the Arizona Sexual Experience Scale, the Brief Sexual Function Questionnaire, the Changes in Sexual Functioning Questionnaire, the Derogatis Sexual Function Inventory, and the Rush Sexual Inventory. The scales vary in length, reliability, validity, and method of administration. Some are rated by the patients themselves, others by the therapist. The formats include structured and semistructured approaches. The scales differ in the number of symptoms assessed, the dysfunctions they target, and the time frame they cover.

DESIRE, INTEREST, AND AROUSAL DISORDERS

Male Hypoactive Sexual Desire Disorder

This dysfunction is characterized by a deficiency or absence of sexual fantasies and desire for sexual activity for a minimum duration of approximately 6 months (Table 21.1–5). Men for whom this is a lifelong condition have never experienced many spontaneous erotic/sexual thoughts. Minimal spontaneous sexual thinking or minimal desire for sex ahead of sexual activity is not considered a diagnosable disorder in women, particularly if the desire is triggered during the sexual encounter.

Also, some men and women define themselves as being asexual. They are not distressed by a lack of desire and feel they are erroneously pathologized if this diagnosis is applied to them. A claim of asexuality occurs among women more frequently than among men. Some people define themselves as being on the “ace” (short for asexual) spectrum, which includes graysexuals, a person who rarely experiences sexual attraction, and demisexuals, one who experiences sexual attraction only after forming an intimate bond. There is controversy as to whether asexuals are likely to be aromantic—not interested in falling in love. The reported prevalence of low desire is greatest at the younger and older ends of the age spectrum, with only 2% of men aged 26 to 44 years affected by this disorder. A reported 6% of men aged 18 to 24 years and 40% of men aged 66 to 74 years have problems with sexual desire. Some men may confuse decreased desire with decreased activity. Their erotic thoughts and fantasies are undiminished, but they no longer act on them due to health issues, unavailability of a partner, or concern about another sexual dysfunction such as erectile disorder.

Table 21.1–4. Taking a Sex History

- I. Identifying data
  - A. Age
  - B. Sex
  - C. Occupation
  - D. Relationship status—single, married, number of times previously married, separated, divorced, cohabiting, serious involvement, casual dating (difficulty forming or keeping relationships should be assessed throughout the interview)
  - E. Sexual orientation—heterosexual, homosexual, or bisexual (this may also be ascertained later in the interview)
- II. Current functioning
  - A. Unsatisfactory to highly satisfactory
  - B. If unsatisfactory, why?
  - C. Feelings about partner satisfaction
  - D. Dysfunctions?—e.g., lack of desire, erectile disorder, inhibited female arousal, anorgasmia, premature ejaculation, retarded ejaculation, pain associated with intercourse (dysfunction discussed below)
  - 1. Onset—lifelong or acquired
    - a. If acquired, when?
    - b. Did onset coincide with drug use (medications or illegal recreational drugs), life stresses (e.g., loss of job, birth of child), interpersonal difficulties?
  - 2. Generalized—occurs in most situations or with most partners
  - 3. Situational
    - a. Only with current partner
    - b. In any committed relationship
    - c. Only with masturbation
    - d. In socially proscribed circumstance (e.g., affair)
    - e. In definable circumstance (e.g., very late at night, in parental home, when partner initiated sex play)
  - E. Frequency—partnered sex (coital and noncoital sex play)
  - F. Desire/libido—how often are sexual feelings, thoughts, fantasies, dreams, experienced (per day, week, etc.)?
  - G.

Description of typical sexual interaction

1. Manner of initiation or invitation (e.g., verbal or physical? Does same person always initiate?)
2. Presence, type, and extent of foreplay (e.g., kissing, caressing, manual or oral genital stimulation)
3. Coitus? Positions used?
4. Verbalization during sex? If so, what kind?
5. Afterplay? (whether sex act is completed or disrupted by dysfunction); typical activities (e.g., holding, talking, return to daily activities, sleeping)
6. Feeling after sex: relaxed, tense, angry, loving

H. Sexual compulsivity? (intrusion of sexual thoughts or participation in sexual activities to a degree that interferes with relationships or work, requires deception, and may endanger the patient)

III. Past sexual history

A. Childhood sexuality

1. Parental attitudes about sex—degree of openness or reserve (assess unusual prudery or seductiveness)
2. Parents' attitudes about nudity and modesty
3. Learning about sex
   a. From parents (Initiated by child's questions or parent volunteering information? Which parent? What was child's age?)? Subjects covered (e.g., pregnancy, birth, intercourse, menstruation, nocturnal emission, masturbation)?

b. From books, magazines, or friends at school or through religious group?

c. Significant misinformation

d. Feeling about information

4. Viewing or hearing primal scene—reaction?
5. Viewing sex play or intercourse of person other than parent
6. Viewing sex between pets or other animals

B. Childhood sex activities

1. Genital self-stimulation before adolescence—age? Reaction if apprehended?
2. Awareness of self as boy or girl—bathroom sensual activities (regarding urine, feces, odor, enemas)?
3. Sexual play or exploration with another child (playing doctor)—type of activity (e.g., looking, manual touching, genital touching)?

Reactions or consequences if apprehended (by whom?)?

IV. Adolescence
   A. Age of onset of puberty—development of secondary sex characteristics, age of menarche for girl, wet dreams or first ejaculation for boy (preparation for and reaction to)
   B. Sense of self as feminine or masculine—body image, acceptance by peers (opposite sex and same-sex), sense of sexual desirability, onset of coital fantasies
   C. Sex activities
   1. Masturbation—age begun? Ever punished or prohibited? Method used, accompanying fantasies, frequency (questions about masturbation and fantasies are among the most sensitive for patients to answer)?
2. Homosexual activities—ongoing or rare and experimental episodes? Approached by others? If homosexual, has there been any heterosexual experimentation?
3. Dating—casual or steady; description of first crush, infatuation, or first love
4. Experiences of kissing, necking, petting ("making out" or "fooling around"), age begun, frequency, number of partners, circumstances, type(s) of activity
5. Orgasm—when first experienced (may not be experienced during adolescence)? With masturbation, during sleep, or with partner? With intercourse or other sex play? Frequency?
6. First coitus—age, circumstances, partner, reactions (may not be experienced during adolescence), contraception and/or safe sex precautions used

V. Adult sexual activities (may be experienced by some adolescents)
   A. Premarital sex
   1. Types of sex play experiences—frequency of sexual interactions, types and number of partners
   2. Contraception and/or safe sex precautions used
   3. First coitus (if not experienced in adolescence)—age, circumstances, partner
   4. Cohabitation—age begun, duration, description of partner, sexual fidelity, types of sexual activity, frequency, satisfaction, number of cohabiting relationships, reasons for breakup(s)

    5. Engagement—age; activity during engagement period with fiancé(e), with others; length of engagement
  B. Marriage (if multiple marriages have occurred, explore sexual activity, reasons for marriage, and reasons for divorce in each marriage)
    1. Types and frequency of sexual interaction—describe typical sexual interaction (see above). Satisfaction with sex life? View of partner's feeling.
    2. First sexual experience with spouse—when? What were the circumstances? Was it satisfying? Disappointing?
    3. Honeymoon—setting, duration, pleasant or unpleasant, sexually active? Frequency? Problems? Compatibility?
    4. Effect of pregnancies and children on marital sex
    5. Extramarital sex—number of incidents, partner, emotional attachment to extramarital partners? Feelings about extramarital sex
    6. Postmarital masturbation—frequency? Effect on marital sex?
    7. Extramarital sex by partner—effect on interviewee
    8. Ménage à trois or multiple sex (swinging) (polyamony)
    9. Areas of conflict in marriage (e.g., parenting, finances, division of responsibilities, priorities)

VI. Sex after widowhood, separation, divorce—celibacy, orgasms in sleep, masturbation, noncoital sex play, intercourse (number of and relationship to partners), other

VII. Special issues
  A. History of rape, incest, sexual, or physical abuse
  B. Spousal abuse (current)
  C. Chronic illness (physical or psychiatric)
  D. History or presence of sexually transmitted diseases
  E. Fertility problems
  F. Abortions, miscarriages, or unwanted or illegitimate pregnancies
  G. Gender identity conflict—(e.g., transsexualism, wearing clothes of opposite sex)
  H. Paraphilias—(e.g., fetishes, voyeurism, sadomasochism)

Table 21.1–5.

DSM-5 Diagnostic Criteria for Male Hypoactive Sexual Desire Disorder

A. Persistently or recurrently deficient (or absent) sexual/erotic thoughts or fantasies and desire for sexual activity. The judgment of deficiency is made by the clinician, taking into account factors that affect sexual functioning, such as age and general and sociocultural contexts of the individual’s life.

B. The symptoms in criterion A have persisted for a minimum duration of approximately 6 months.

C. The symptoms in criterion A cause clinically significant distress in the individual.

D. The sexual dysfunction is not better explained by a nonsexual mental disorder or as a consequence of severe relationship distress or other significant stressors and is not attributable to the effects of a substance/medication or another medical condition.

Specify whether:

Lifelong: The disturbance has been present since the individual became sexually active.

Acquired: The disturbance began after a period of relatively normal sexual function.

Specify whether:

Generalized: Not limited to certain types of stimulation, situations, or partners.

Situational: Only occurs with certain types of stimulation, situations, or partners.

Specify current severity:

Mild: Evidence of mild distress over the symptoms in criterion A.

Moderate: Evidence of moderate distress over the symptoms in criterion A.

Severe: Evidence of severe or extreme distress over the symptoms in criterion A.

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

A variety of causative factors are associated with low sexual desire. Patients with desire problems often use inhibition of desire defensively to protect against unconscious fears about sex. Sigmund Freud conceptualized low sexual desire as the result of inhibition during the phallic psychosexual phase of development and of unresolved oedipal conflicts. Some men, fixated at the phallic stage of development, are fearful of the vagina and believe they will be castrated if they approach it. Freud called this concept vagina dentata. He theorized that men avoid contact with the vagina when they unconsciously believe that the vagina has teeth. Lack of desire can also result from chronic stress, anxiety, or depression.

The presence of desire depends on several factors: biologic drive, adequate self-esteem, the acceptance of sexuality as a positive attribute, previous good experiences with sex, the availability of an appropriate partner, and a good relationship in nonsexual areas with one's partner. Damage to any of those factors may result in diminished desire.

Hypoactive *sexual* desire disorders often become manifest during puberty and may be a lifelong condition. A general medical workup should be conducted to rule out a medical cause, which, if present, would be, according to DSM-5, diagnosed as hyposexual desire disorder due to a general medical condition.

A 20-year-old college student presented with a complaint of insufficient desire for sex. He stated that he was bisexual but complained, "I wish I could be more interested in women." During the course of the interview, he revealed that he had sexual fantasies about men dating from his adolescence and that he masturbated regularly to erotic visual stimuli involving men. He had fears of erectile dysfunction prior to coitus with women but did not experience these fears prior to sexual interaction with men. These interactions involved manual and oral sex play but no penetration. He connected with men online and had not disclosed his sexual interest in or contact with men to anyone. This patient did not suffer from hyposexual desire disorder. His problems arose from conflicted feelings about his sexual orientation.

A 44-year-old married man had a history of good sexual interaction with his wife of 18 years until 8 months prior to his consultation. Nine months ago, he had sought help for a depression triggered by the loss of employment, financial stress, and lack of immediate prospects for a new job. He had become lethargic and withdrawn and was doing little to pursue new work. He was treated with fluoxetine, 20 mg per day, which helped him, and currently, he was actively networking in pursuit of employment. However, his sexual interest remained diminished. His diminished libido appeared to be a side effect of his medication, and his diagnosis was pharmacologically induced hyposexual desire disorder.

Female Sexual Interest/Arousal Disorder

The combination of interest (or desire) and arousal into one dysfunction category reflects the recognition that women do not necessarily move stepwise from desire to arousal but often experience desire synchronously with or even following beginning feelings of arousal. This is particularly true for women in long-term relationships. As a corollary, women experiencing sexual dysfunction may experience either/or both inability to feel interest or arousal, and they may often have difficulty achieving orgasm or experience pain in addition. Some may experience dysfunction across the entire range of sexual response/pleasure. Complaints in this dysfunction category present variously as a decrease or paucity of erotic feelings, thoughts, or fantasies; a decreased impulse to initiate sex; a decreased or absent receptivity to partner overtures; or an inability to respond to partner stimulation (Table 21.1–6).

A complicating factor in this diagnosis is that a subjective sense of arousal is often poorly correlated with genital lubrication in both normal and dysfunctional women. Therefore, complaints of lack of pleasure are sufficient for these diagnoses even when vaginal lubrication and congestion are present. A woman complaining of lack of arousal may lubricate vaginally but may not experience a subjective sense of excitement. Some studies using functional magnetic resonance imaging (fMRI) have revealed a low correlation between brain activation in areas controlling genital response and simultaneous ratings of subjective arousal. Physiologic studies of sexual dysfunctions indicate that a hormonal pattern may contribute to responsiveness in women who have arousal dysfunction. Masters and Johnson found that women are particularly desirous of sex before the onset of the menses. Other women report feeling the greatest sexual excitement immediately after the menses or at the time of ovulation. Alterations in testosterone, estrogen, prolactin, and thyroxin levels have been implicated in female sexual arousal disorder. In addition, medication with antihistaminic or anticholinergic properties causes a decrease in vaginal lubrication.

Table 21.1–6.

DSM-5 Diagnostic Criteria for Female Sexual Interest/Arousal Disorder

A. Lack of, or significantly reduced, sexual interest/arousal, as manifested by at least three of the following:
    1. Absent/reduced interest in sexual activity.
    2. Absent/reduced sexual/erotic thoughts or fantasies.
    3. No/reduced initiation of sexual activity, and typically unreceptive to a partner's attempts to initiate.
    4. Absent/reduced sexual excitement/pleasure during sexual activity in almost all or all (approximately 75–100%) sexual encounters (in identified situational contexts or, if generalized, in all contexts).
    5. Absent/reduced sexual interest/arousal in response to any internal or external sexual/erotic cues (e.g., written, verbal, visual).
    6. Absent/reduced genital or nongenital sensations during sexual activity in almost all or all (approximately 75–100%) sexual encounters (in identified situational contexts or, if generalized, in all contexts).

B. The symptoms in criterion A have persisted for a minimum duration of approximately 6 months.

C. The symptoms in criterion A cause clinically significant distress in the individual.

D. The sexual dysfunction is not better explained by a nonsexual mental disorder or as a consequence of severe relationship distress (e.g., partner violence) or other significant stressors and is not attributable to the effects of a substance/medication or another medical condition.

Specify whether:

Lifelong: The disturbance has been present since the individual became sexually active.

Acquired: The disturbance began after a period of relatively normal sexual function.

Specify whether:

Generalized: Not limited to certain types of stimulation, situations, or partners.

Situational: Only occurs with certain types of stimulation, situations, or partners.

Specify current severity:

Mild: Evidence of mild distress over the symptoms in criterion A.

Moderate: Evidence of moderate distress over the symptoms in criterion A.

Severe: Evidence of severe or extreme distress over the symptoms in criterion A.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

Factors such as life stresses, menopause, presence of adequate sexual stimulation must be evaluated before making this diagnosis. Relationship problems are particularly relevant to acquired interest/arousal disorder. In one study of couples with markedly decreased sexual interaction, the most prevalent etiology was marital discord.

A couple presented for therapy, stating that they had not had sexual relations for 3 years. The man was aged 56 years, and the woman was 54. They had been married for 21 years and reported that their relationship was gratifying aside from the lack of sex which bothered both of them. Although they had sexual relations in the past, those interactions had been troublesome to each party. The man suffered from premature ejaculation, about which he was ashamed.

The woman complained of being frustrated with their sexual activity. In addition, she experienced his dysfunction as a personal rejection and a sign of lack of sufficient love for and consideration of her. Although she still experienced desire in the form of fantasies, her receptivity to sexual intimacy with him disappeared. She began rejecting him if he initiated sex, resulting in the cessation of any sexual interaction between them.

In addition to his early ejaculation, the couple was dealing with her acquired sexual interest/arousal disorder.

Male Erectile Disorder

The male erectile disorder was historically called *impotence.* The term was dropped for a more medical designation but also because it was considered derogatory and had negative connotations for the man affected with the problem. However, it describes with accuracy the feelings of powerlessness, helplessness, and resultant low self-esteem men with this dysfunction frequently suffer (Table 21.1–7). A man with a lifelong male erectile disorder has never been able to obtain an erection sufficient for insertion. In acquired male erectile disorder, a man has successfully achieved penetration at some time in his sexual life but is later unable to do so. In situational male erectile disorder, a man is able to have coitus in certain circumstances but not in others; for example, he may function effectively with a prostitute but be unable to have an erection when with his partner. Erectile disorder is also diagnosed when the man can obtain an erection but cannot maintain it for a reasonable amount of time necessary to complete the sex act.

Table 21.1–7.
DSM-5 Diagnostic Criteria for Male Erectile Disorder

A. At least one of the three following symptoms must be experienced on almost all or all (approximately 75–100%) occasions of sexual activity (in identified situational contexts or, if generalized, in all contexts):
   1. Marked difficulty in obtaining an erection during sexual activity.
   2. Marked difficulty in maintaining an erection until the completion of sexual activity.
   3. Marked decrease in erectile rigidity.

B. The symptoms in criterion A have persisted for a minimum duration of approximately 6 months.

C. The symptoms in criterion A cause clinically significant distress in the individual.

D. The sexual dysfunction is not better explained by a nonsexual mental disorder or as a consequence of severe relationship distress or other significant stressors and is not attributable to the effects of a substance/medication or another medical condition.

Specify whether:

Lifelong: The disturbance has been present since the individual became sexually active.

Acquired: The disturbance began after a period of relatively normal sexual function.

Specify whether:

Generalized: Not limited to certain types of stimulation, situations, or partners.

Situational: Only occurs with certain types of stimulation, situations, or partners.

Specify current severity:

Mild: Evidence of mild distress over the symptoms in criterion A.

Moderate: Evidence of moderate distress over the symptoms in criterion A.

Severe: Evidence of severe or extreme distress over the symptoms in criterion A.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

Acquired male erectile disorder has been reported in 10% to 20% of all men. Freud declared it common among his patients. Erectile disorder is the chief complaint of more than 50% of all men treated for sexual disorders. Lifelong male erectile disorder is rare; it occurs in about 1% of men under age 35. The incidence of erectile disorder increases with age. It has been reported variously as 2% to 8% of the young adult population. Alfred Kinsey reported that 75% of all men were impotent at age 80. There is a reported incidence of 40% to 50% in men between ages 60 and 70. All men over 40, Masters and Johnson claimed, have a fear of impotence, which the researchers believed reflected the masculine fear of loss of virility with advancing age. Male erectile disorder, however, is not universal in aging men; having an available sex partner is related to continuing potency, as is a history of consistent sexual activity and the absence of vascular, neurologic, or endocrine disease. Twenty percent of men fear erectile dysfunction prior to their first coitus; the reported incidence of actual erectile dysfunction during first coitus is 8%. As Stephen Levine has stated, the first coital encounter “is a horse race between excitement and anxiety.”

Male erectile disorder can be organic or psychological or a combination of both, but in young and middle-aged men, the cause is usually psychological. A good history is of primary importance in determining the cause of the dysfunction. If a man reports having spontaneous erections at times when he does not plan to have intercourse, having morning erections, or having good erections with masturbation or with partners other than his usual one, the organic causes of his erectile disorder can be considered negligible, and costly diagnostic procedures can be avoided. Male erectile disorder caused by a general medical condition or a pharmacologic substance is discussed later in this section.

Freud ascribed one type of erectile disorder to an inability to reconcile feelings of affection toward a woman with feelings of desire for her. Men with such conflicting feelings can function only with women whom they see as degraded (*Madonna–Putana complex*). Other factors that have been cited as contributing to impotence include a punitive superego, an inability to trust, and feelings of inadequacy or a sense of being undesirable as a partner. A man may be unable to express a sexual impulse because of fear, anxiety, anger, or moral prohibition. In an ongoing relationship, the disorder may reflect difficulties between the partners, particularly when a man cannot communicate his needs or his anger in a direct and constructive way. In addition, episodes of erectile disorder are reinforcing, with the man becoming increasingly anxious before each sexual encounter.

A cohabiting man and woman, law students, aged 25 years, presented for treatment with a chief complaint of limited sexual interaction due to erectile disorder on the part of the man. He experienced morning erections and had no difficulty when he masturbated, obviating a physiologic etiology for his problem. The woman evidenced no dysfunction and appeared loving and supportive of her partner.

His history revealed normal puberty and first coitus at age 17 with a high school girlfriend to whom he was emotionally attached. Sex with her was erratic. The relationship ended when both went to different colleges, and the ending was not traumatic. Of significance in the sexual history was his description of never initiating involvements but becoming involved with girls who “chose me.” The patient was an only child and related a family history of a tyrannical father who denigrated the patient for not attending an Ivy League law school. Of note, the patient was attending a highly regarded law school, had gone to a competitive entry high school, and had attended a well-respected college. The patient was aware of having a high anxiety level. At the time of treatment, he was particularly anxious about obtaining a summer internship with a “top” law firm.

The patient’s erectile disorder was another manifestation of his general anxiety and shaky self-esteem. Behavior therapy was initiated to treat his sexual anxiety. The couple was assigned a regimen of sexual exercises designed to be minimally demanding, thus allowing the man time to gain confidence in his ability to perform sexually. With the combination of nondemand exercises and the cooperation and acceptance of his partner, the patient was able to overcome his erectile disorder.

Orgasmic Disorders

Female orgasmic disorder (also known as *inhibited female orgasm,* or *anorgasmia*) is defined as the recurrent and persistent inhibition of the female orgasm, manifested by the absence or delay of orgasm after a normal sexual excitement phase, or a marked diminution of the intensity of orgasmic sensations. Women who can achieve orgasm with noncoital clitoral stimulation but cannot experience it during coitus in the absence of manual clitoral stimulation are not categorized as anorgasmic. The woman herself reports this complaint. However, some anorgasmic women are not distressed by the lack of climax and derive pleasure from sexual activity. In the latter instance, a woman may present with this complaint because her partner is troubled by her lack of orgasm.

Physiologic research on the female sexual response has demonstrated that orgasms caused by clitoral stimulation are physiologically identical to those caused by vaginal stimulation. Freud's theory that women must give up clitoral sensitivity for vaginal sensitivity to achieve sexual maturity is now considered misleading, although some women say that they gain a special sense of satisfaction from an orgasm precipitated by coitus. Some therapists attribute that to the psychological feeling of closeness engendered by the act of coitus, but others maintain that the coital orgasm is a physiologically different experience. Many women achieve orgasm during coitus by a combination of manual clitoral stimulation and penile–vaginal stimulation.

A woman with a lifelong female orgasmic disorder has never experienced orgasm by any kind of stimulation. A woman with an acquired orgasmic disorder has previously experienced at least one orgasm, regardless of the circumstances or means of stimulation, whether by masturbation or while dreaming during sleep. Studies have shown that women achieve orgasm more consistently with masturbation than with partnered sex. Kinsey found that 5% of married women over 35 years of age had never achieved orgasm by any means. The incidence of never having experienced orgasm is reported as 10% among all women. The incidence of orgasm increases with age. According to Kinsey, the first orgasm occurs during adolescence in about 50% of women as a result of masturbation or genital caressing with a partner; the rest usually experience orgasm as they get older. Lifelong female orgasmic disorder is more common among unmarried women than married women. Increased orgasmic potential in women over 35 years of age has been explained on the basis of less psychological inhibition, greater sexual experience, or both.

Acquired female orgasmic disorder is a common complaint in clinical populations. One clinical treatment facility reported having about four times as many nonorgasmic women in its practice as female patients with all other sexual disorders. In another study, 46% of women complained of difficulty reaching orgasm. Inhibition of arousal and orgasmic problems often occur together. The overall prevalence of female orgasmic disorder from all causes is estimated to be 30%. A recent twin study suggests that orgasmic dysfunction in some females has a genetic basis and cannot be attributed solely to cultural differences or developmental experiences. In that study population researchers found an estimated heritability for difficulty reaching orgasm with intercourse of 34% (similar to the population at large) and an estimated heritability for difficulty reaching orgasm with masturbation of 45% (a considerably greater percentage than the population at large).

Numerous psychological factors are associated with female orgasmic disorder. They include fears of impregnation, rejection by a sex partner, and damage to the vagina; hostility toward men; poor body image, and feelings of guilt about sexual impulses. Some women equate orgasm with loss of control or with aggressive, destructive, or violent impulses; their fear of these impulses may be expressed through inhibition of arousal or orgasm.

Prolonged marital discord may lead to acquired anorgasmia. Cultural expectations and social restrictions on women are also relevant. Many women have grown up to believe that sexual pleasure is not a natural entitlement for the so-called decent women. Nonorgasmic women may be otherwise symptom free or may experience frustration in a variety of ways; they may have such pelvic complaints as lower abdominal pain, itching, and vaginal discharge, as well as increased tension, irritability, and fatigue.

A 36-year-old woman presented with extreme distress over being anorgasmic. She was unmarried and had recently ended a relationship in which she had been very attracted to the man but was embarrassed by her lack of orgasm, about which she never told him. The patient could not bring herself to orgasm by masturbation and was reluctant to fantasize.

As a child, she had witnessed the primal scene three times that she could remember. Her parents made no attempt to close their door while having sex, and she found herself listening for sounds of coitus.

Treatment focused both on masturbatory exercises with encouragement to fantasize and material to stimulate fantasy and analytically oriented psychotherapy to deal with oedipal conflicts and inappropriate sexual stimulation during childhood.

Delayed Ejaculation

In male delayed ejaculation, sometimes called *retarded ejaculation,* a man achieves ejaculation during coitus with great difficulty, if at all. The problem is sometimes present with masturbation but appears as a problem primarily during partnered sex. A man with lifelong delayed ejaculation has never been able to ejaculate during partnered sexual activity. The problem is usually most pronounced during coital activity. The disorder is diagnosed as acquired if it develops after previously normal functioning. Some researchers think that orgasm and ejaculation should be differentiated, especially in the case of men who ejaculate but complain of a decreased or absent subjective sense of pleasure during the orgasmic experience (orgasmic anhedonia).

The incidence of male orgasmic disorder is much lower than the incidence of premature ejaculation or erectile disorder. Masters and Johnson reported an incidence of delayed ejaculation of only 3.8% in one group of 447 men with sexual dysfunctions. A general prevalence of 5% has been reported. However, an increase in the presentation of this disorder in sex therapy programs has been seen in the more recent past. This has been attributed to the increasing use of antidepressants which can have a side effect of delayed ejaculation and high use of internet pornography sites. These sites offer a level of stimulation involving such a variety of people and acts that they may inure the man to the stimulation of more typical partnered activity. One study of adolescent males who used these sites frequently, prior to live sexual interaction, reported that these teens do not develop neuronal synapses that will enable them to respond to usual partnered interactions with sufficient pleasure to allow them to achieve climax.

Lifelong delayed ejaculation indicates severe psychopathology. A man may come from a rigid, puritanical background; he may perceive sex as sinful and the genitals as dirty, and he may have conscious or unconscious incest wishes and guilt. He usually has difficulty with closeness in areas beyond those of sexual relations. In a few cases, the condition is aggravated by an attention deficit disorder. A man's distractibility prevents sufficient arousal for a climax to occur.

In an ongoing relationship, acquired male delayed ejaculation disorder frequently reflects interpersonal difficulties. The disorder may be a man's way of coping with real or fantasized changes in a relationship, such as plans for pregnancy about which the man is ambivalent, the loss of sexual attraction to the partner, or demands by the partner for greater commitment as expressed by sexual performance. In some men, the inability to ejaculate reflects unexpressed hostility toward a woman. The problem is more common among men with obsessive-compulsive disorder (OCD) than among others.

A 43-year-old married man presented with delayed ejaculation of 5 months duration. The onset of the dysfunction coincided with his agreeing to try to have a child with his wife. This was the patient's second marriage, and he had a 9-year-old daughter from his first marriage who lived primarily with his ex-wife. He and his current wife had discussed having children before they married. Although the patient did not want another child, he agreed to father one because his current wife, age 39, was childless, and he felt it would be selfish to deny her a baby. He was afraid of losing his wife if they did not try to conceive. Therapy involved behavioral exercises geared to restore the patient's pleasure in sexuality as well as introspective psychotherapy to address his ambivalence about conception.

A 31-year-old male presented with a history of lifelong delayed ejaculation intravaginally. He was engaged to be married, and his fiancée was increasingly distressed by his dysfunction. The patient demonstrated obsessive characteristics. For example, he refused to allow his fiancée to hang up his shirts because he was adamant that they be spaced apart to an exact degree. In addition, he had been traumatized in his adolescence by seeing his mother miscarry a pregnancy “all over the kitchen floor.” The couple was treated with behavioral sex therapy exercises.

Delayed ejaculation may have physiologic causes and can occur after surgery of the genitourinary tract, for example, with a prostatectomy. Delayed ejaculation may also be associated with Parkinson disease and other neurologic disorders involving the lumbar or sacral sections of the spinal cord. The antihypertensive drugs guanethidine monosulfate (Ismelin) and methyldopa (Aldomet) have been implicated in delayed ejaculation, as have the phenothiazines and almost all the antidepressants. Strictly organic cases are diagnosed as medically or pharmacologically induced sexual dysfunction.

Early Ejaculation

In early ejaculation, men persistently or recurrently achieve orgasm and ejaculation before they wish to. The diagnosis is made when a man regularly ejaculates before or within approximately 1 minute after penetration. DSM-5 refers only to “vaginal penetration” in its diagnostic criteria even though it is entirely possible for the disorder to occur in men who are homosexual and do not engage in vaginal penetration. DSM-5 defines the disorder as mild if ejaculation occurs within approximately 30 seconds to 1 minute of vaginal penetration, moderate if ejaculation occurs within approximately 15 to 30 seconds of vaginal penetration, and severe when ejaculation occurs at the start of sexual activity or within approximately 15 seconds of vaginal penetration. A difficulty with these specifiers involves time distortions which patients make in both overestimating and underestimating time from penetration to climax. Clinicians need to consider factors that affect the duration of the excitement phase of the sexual response, such as age, the novelty of the sex partner, and the frequency and duration of coitus. As with the other sexual dysfunctions, premature ejaculation is not diagnosed when it is caused exclusively by organic factors or when it is symptomatic of another clinical psychiatric syndrome.

Early ejaculation is more commonly reported among college-educated men than among men with less education. The complaint is thought to be related to their concern for partner satisfaction, but the true cause of this increased frequency has not been determined. Early ejaculation is the chief complaint of about 35% to 40% of men treated for sexual disorders. In DSM-5, the writers state that the disorder, with its newly defined time parameter, would now be an accurate diagnosis for only 1% to 3% of men. Some researchers divide men who experience early ejaculation into two groups: those who are physiologically predisposed to climax quickly because of shorter nerve latency time and those with a psychogenic or behaviorally conditioned cause. Difficulty in ejaculatory control can be associated with anxiety regarding the sex act, with unconscious fears about the vagina, or with negative cultural conditioning. Men whose early sexual contacts occurred largely with prostitutes who encourage a rapid conclusion of the sex act, or whose sexual contacts took place in situations in which discovery would be embarrassing (e.g., in a shared dormitory room or in the parental home) might have been conditioned to achieve orgasm rapidly. With young, inexperienced men, who have the problem, it may resolve in time. In ongoing relationships, the partner has a great influence on a premature ejaculator, and a stressful marriage exacerbates the disorder. The developmental background and the psychodynamics found in premature ejaculation and in erectile disorder are similar.

PAIN DISORDER

Genitopelvic Pain/Penetration Disorder

In DSM-5, this disorder refers to one or more of the following complaints, of which any two or more may occur together: difficulty having intercourse; genitopelvic pain; fear of pain or penetration; and tension of the pelvic floor muscles. Previously, these pain disorders were diagnosed as *dyspareunia* or *vaginismus.* These former diagnoses could coexist with each other or lead to one another and could understandably lead to fear of pain with sex. Thus, it is reasonable to gather them into one diagnostic category. For purposes of clinical discussion, however, the distinct categories of dyspareunia and vaginismus remain useful.

Dyspareunia. Dyspareunia is recurrent or persistent genital pain occurring before, during, or after intercourse. Dyspareunia is related to and often coincides with vaginismus. Repeated episodes of vaginismus can lead to dyspareunia and vice versa; in either case, somatic causes must be ruled out. DSM-5 cites that 15% of women in North America report recurrent pain during intercourse.

In most cases, dynamic factors are considered causative. Chronic pelvic pain is a common complaint in women with a history of rape or childhood sexual abuse. Painful coitus can result from tension and anxiety about the sex act that causes women to involuntarily contract their pelvic floor muscles. The pain is real and makes intercourse unpleasant or unbearable. Anticipation of further pain may cause women to avoid coitus altogether. If a partner proceeds with intercourse regardless of a woman's state of readiness, the condition is aggravated. There is an increase in reported dyspareunia postmenopausally due to hormonally induced physiologic changes in the vagina; however, specific complaints of difficulty having intercourse occur more often in premenopausal women. There is some increase in dyspareunia in the immediate postpartum population, but it is usually temporary. Dyspareunia may present as any of the four complaints listed under genitopelvic pain/penetration disorder and should be diagnosed as genitopelvic pain/penetration disorder.

A 20-year-old woman presented with pelvic pain/penetration disorder 6 months after her marriage. She and her husband had experienced manual sex play to which she responded with some arousal, but the marriage remained unconsummated. The woman had been brought up in an orthodox Jewish community, where virginity was an imperative and social and sexual interaction between adolescent and young adult men and women were limited. The patient had minimal experience dating, and her husband was similarly inexperienced sexually. Therapy consisted of sex education and behavioral sexual exercises.

Dyspareunia can occur in men, but it is uncommon and usually associated with a medical condition such as Peyronie disease, prostatitis, or gonorrheal or herpetic infections. Vasocongestion during sexual activity without orgasmic release also may lead to discomfort. Rarely, some men experience pain during ejaculation (postejaculatory pain disorder). This pain is caused by an involuntary spasm of the perineal muscles that may be due to psychological conflicts about the sex act or may be an adverse effect of some antidepressant medications. Sexual dysfunction caused by medication is the diagnosis used when a substance or medication is the sole or major causal factor for the pain.

Sexual Dysfunction Due to a General Medical Condition

The category of sexual dysfunction due to a general medical condition covers sexual dysfunction that results in marked distress and interpersonal difficulty when there is evidence from the history, physical examination, or laboratory findings of a general medical condition judged to be causally related to the sexual dysfunction.

Male Erectile Disorder Due to a General Medical Condition. Many studies have focused on the relative incidences of psychological and organic male erectile disorder. Statistics indicate that 20% to 50% of men with erectile disorder have an organic basis for the disorder. A physiologic etiology is more likely in men older than 50 and the most likely cause in men older than age 60. Physiologically, erectile dysfunction may be due to a variety of medical causes (Table 21.1–8). In the United States, it is estimated that two million men cannot gain erections because they suffer from diabetes mellitus; an additional 300,000 are dysfunctional because of other endocrine diseases; approximately 1.5 million are dysfunctional as a result of vascular disease; an estimated 180,000 because of multiple sclerosis; approximately 400,000 because of traumas and fractures leading to pelvic fractures or spinal cord injuries; and another 650,000 are estimated to be dysfunctional as a result of radical surgery, including prostatectomies, colostomies, and cystectomies. In addition, the clinician should be aware of the possible pharmacologic effects of medication on sexual functioning. The increased incidence of organic causes for erectile dysfunction may partly reflect the increased use of psychotropic and antihypertensive medications. Adverse effects of medication may impair male sexual functioning in a variety of ways. Castration (removal of the testes) does not always lead to sexual dysfunction, depending on the person. Erection may still occur after castration via a reflex arc that passes through the sacral cord erectile center. It is triggered when the inner thigh is stimulated.

Table 21.1–8.
Diseases and Other Medical Conditions Implicated in Erectile Dysfunction

- Infectious and parasitic diseases
    - Elephantiasis
    - Mumps
- Cardiovascular disease
    - Atherosclerotic disease
    - Aortic aneurysm
    - Leriche syndrome
    - Cardiac failure
- Renal and urologic disorders
    - Peyronie disease
    - Chronic renal failure
    - Hydrocele and varicocele
- Hepatic disorders
    - Cirrhosis (usually associated with alcohol dependence)
- Pulmonary disorders
    - Respiratory failure
- Genetic disorders
    - Klinefelter syndrome
    - Congenital penile vascular and structural abnormalities
- Nutritional disorders
    - Malnutrition
    - Vitamin deficiencies
- Endocrine disorders
    - Diabetes mellitus
    - Dysfunction of the pituitary–adrenal–testis axis
    - Acromegaly
    - Addison disease
    - Chromophobe adenoma
    - Adrenal neoplasia
    - Myxedema
    - Hyperthyroidism
- Neurologic disorders
    - Multiple sclerosis
    - Transverse myelitis
    - Parkinson disease
    - Temporal lobe epilepsy
    - Traumatic and neoplastic spinal cord diseases

  - Central nervous system tumor
  - Amyotrophic lateral sclerosis
  - Peripheral neuropathy
  - General paresis
  - Tabes dorsalis
- Pharmacologic contributants
  - Alcohol, tobacco, and other dependence-inducing substances (heroin, methadone, morphine, cocaine, amphetamines, and barbiturates)
  - Prescribed drugs (psychotropic drugs, antihypertensive drugs, estrogens, and antiandrogens)
- Poisoning
  - Lead (plumbism)
  - Herbicides
- Surgical procedures
  - Perineal prostatectomy
  - Abdominal–perineal colon resection
  - Sympathectomy (frequently interferes with ejaculation)
  - Aortoiliac surgery
  - Radical cystectomy
  - Retroperitoneal lymphadenectomy
- Miscellaneous
  - Radiation therapy
  - Pelvic fracture
  - Any severe systemic disease or debilitating condition

The mechanisms involved in the various diseases that lead to erectile disorder include general endothelial dysfunction or dysfunction of the penile endothelium specifically, atherosclerosis, dysregulation of nitric oxide synthase, a decrease of nitric oxide, low testosterone levels due to hyperprolactinemia-influenced changes in the hypothalamic–pituitary axis, vasculopathy, autonomic neuropathy, and disruption of neural proerectile processes.

PHYSIOLOGIC TESTS. A number of procedures, benign and invasive, are used to help differentiate psychogenic erectile dysfunction. The procedures include monitoring nocturnal penile tumescence (erection that occurs during sleep) normally associated with rapid eye movement; monitoring tumescence with a strain gauge; measuring blood pressure in the penis with a penile plethysmograph or an ultrasound (Doppler) flow meter, both of which assess blood flow in the internal pudendal artery; and measuring pudendal nerve latency time. Neurologic impairment of penile function may be indicated by a decreased vibratory perception in the penis. Other diagnostic tests that delineate organic bases for erectile disorder include glucose tolerance tests, plasma hormone assays, liver and thyroid function tests, prolactin and follicle-stimulating hormone (FSH) determinations, and cystometric examinations. On physical examination, evidence of hypogonadism such as small testes, gynecomastia, reduced or absent body, and facial hair warrant attention. Invasive diagnostic studies include penile arteriography, infusion cavernosonography, and radioactive xenon penography. Invasive procedures require expert interpretation and are used only for patients who are candidates for vascular reconstructive procedures.

MEDICAL VERSUS PSYCHOGENIC CAUSES. A good history is crucial to determining the cause of the male erectile disorder. If a man reports having spontaneous erections at times when he does not plan to have intercourse, having morning erections or only sporadic erectile dysfunction, or having good erections with masturbation or with partners other than his usual one, then organic causes for his disorder can be considered negligible, and costly diagnostic procedures can be avoided. When a medical basis for erectile dysfunction is found, psychological factors often contribute to the dysfunction, and psychiatric treatment may be helpful. Some diabetics, for instance, may experience psychogenic erectile dysfunction.

Dyspareunia Due to a General Medical Condition. An estimated 30% of all surgical procedures on the female genital area result in temporary dyspareunia. In addition, 30% to 40% of women with the complaint who are seen in sex therapy clinics have pelvic pathology. Organic abnormalities leading to dyspareunia and vaginismus include infected hymenal remnants or congenitally imperforate or unusually thick hymens; episiotomy scars; Bartholin glands infections; various forms of vaginitis and cervicitis; testosterone deficiency; vulvodynia; vaginal treatment, such as radiation or surgery that has caused scarring; dermatologic disorder, such as lichens sclerosis or lichens planus; and endometriosis. Postcoital pain reported by women with myomata and endometriosis is attributed to uterine contractions during orgasm. Postmenopausal women may have dyspareunia because of thinning of the vaginal mucosa and reduced lubrication. Women who have been on birth control pills for many years may have lessened lubrication and mild vaginal thinning similar to the postmenopausal woman.

Dyspareunia can also occur in men, but it is uncommon and is usually associated with an organic condition such as Peyronie disease, which consists of sclerotic plaques on the penis that cause penile curvature and make penetration difficult and painful.

Hypoactive Sexual Desire and Arousal/Interest Disorder Due to a General Medical Condition. Desire commonly decreases after major illness or surgery, particularly when body image is affected after such procedures as mastectomy, ileostomy, hysterectomy, and prostatectomy. Illnesses that deplete a person's energy, chronic conditions that require physical and psychological adaptation, and serious illnesses that may cause the person to become depressed can all result in a marked lessening of sexual desire or interest in both men and women. In some cases, biochemical correlates are associated with hypoactive sexual desire disorder (Table 21.1–9). A recent study found markedly lower serum testosterone concentrations in men complaining of low desire than in normal controls in a sleep laboratory situation. Drugs that depress the central nervous system (CNS) or decrease testosterone production can decrease desire.

Table 21.1–9.
Neurotransmitter Effects on Sex Function

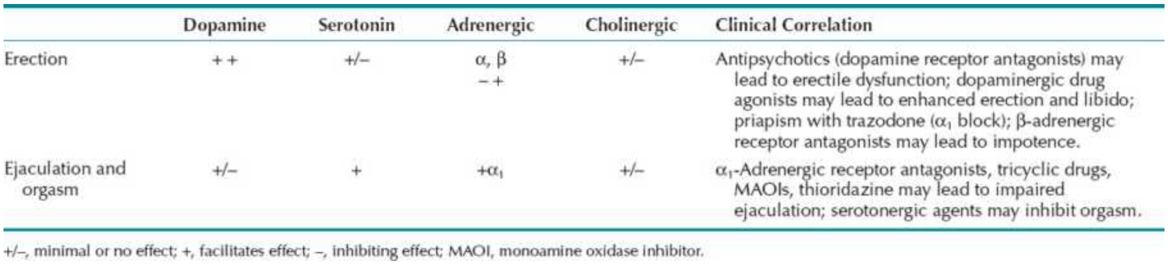

	Dopamine	Serotonin	Adrenergic	Cholinergic	Clinical Correlation
Erection	++	+/–	α, β – +	+/–	Antipsychotics (dopamine receptor antagonists) may lead to erectile dysfunction; dopaminergic drug agonists may lead to enhanced erection and libido; priapism with trazodone (α_1 block); β-adrenergic receptor antagonists may lead to impotence.
Ejaculation and orgasm	+/–	+	+α_1	+/–	α_1-Adrenergic receptor antagonists, tricyclic drugs, MAOIs, thioridazine may lead to impaired ejaculation; serotonergic agents may inhibit orgasm.

+/–, minimal or no effect; +, facilitates effect; –, inhibiting effect; MAOI, monoamine oxidase inhibitor.

Other Male Sexual Dysfunction Due to a General Medical Condition

Delayed ejaculation can have physiologic causes and can occur after surgery on the genitourinary tract, such as prostatectomy. It may also be associated with Parkinson disease and other neurologic disorders involving the lumbar or sacral sections of the spinal cord. The antihypertensive drug guanethidine monosulfate (Ismelin), methyldopa (Aldomet), the phenothiazines, the tricyclic drugs, and the selective serotonin reuptake inhibitors (SSRIs), among others, have been implicated in retarded ejaculation. Delayed ejaculation must also be differentiated from retrograde ejaculation, in which ejaculation occurs but the seminal fluid passes back into the bladder. Retrograde ejaculation always has an organic cause. It can develop after genitourinary surgery and is also associated with medications that have anticholinergic adverse effects, such as some phenothiazines.

Other Female Sexual Dysfunction Due to a General Medical Condition

Some medical conditions—specifically endocrine diseases such as hypothyroidism, diabetes mellitus, and primary hyperprolactinemia—can affect a woman's ability to have orgasms. Several drugs also affect some women's capacity to have orgasms. Antihypertensive medications, tricyclic drugs, SSRIs, and, frequently, monoamine oxidase inhibitors (MAOIs) have interfered with female orgasmic capacity. One study of women taking MAOIs, however, found that after 16 to 18 weeks of pharmacotherapy, the adverse effect of the medication disappeared, and the women were able to reexperience orgasms, although they continued taking an undiminished dosage of the drug.

SUBSTANCE/MEDICATION-INDUCED SEXUAL DYSFUNCTION

The diagnosis of substance-induced sexual dysfunction is used when evidence of substance intoxication or withdrawal is apparent from the history, physical examination, or laboratory findings. The disturbance in sexual function must be predominant in the clinical picture. Distressing sexual dysfunction occurs soon after significant substance intoxication or withdrawal or after exposure to a medication or a change in medication use. Specified substances include alcohol, amphetamines, or related substances, cocaine, opioids, sedatives, hypnotics, or anxiolytics, and other or unknown substances.

Abused recreational substances affect sexual function in various ways. In small doses, many substances enhance sexual performance by decreasing inhibition or anxiety or by causing a temporary elation of mood. With continued use, however, erectile engorgement and orgasmic and ejaculatory capacities become impaired. The abuse of sedatives, anxiolytics, hypnotics, and particularly opiates and opioids nearly always depresses desire. Alcohol may foster the initiation of sexual activity by removing inhibition, but it also impairs performance. Cocaine and amphetamines produce the following similar effects: although no direct evidence indicates that sexual drive is enhanced, users initially have feelings of increased energy and may become sexually active; ultimately, dysfunction occurs. Men usually go through two stages: an experience of prolonged erection without ejaculation, then a gradual loss of erectile capability.

Patients recovering from substance dependency may need therapy to regain sexual function, partly because of psychological readjustment to a nondependent state. Many substance abusers have always had difficulty with intimate interactions. Others who spent their crucial developmental years under the influence of a substance have missed the experiences that would have enabled them to learn social and sexual skills.

Pharmacologic Agents Implicated in Sex Dysfunction

Almost every pharmacologic agent, particularly those used in psychiatry, has been associated with an effect on sexuality. In men, these effects include decreased sex drive, erectile failure, decreased volume of ejaculate, and delayed or retrograde ejaculation. In women, decreased sex drive, decreased vaginal lubrication, inhibited or delayed orgasm, and decreased or absent vaginal contractions may occur. Drugs may also enhance the sexual responses and increase the sex drive, but this is less common than adverse effects. The effects of psychoactive drugs are detailed below.

Antipsychotic Drugs. Most antipsychotic drugs are dopamine receptor antagonists that also block adrenergic and cholinergic receptors, thus accounting for adverse sexual effects (see Table 21.1–9). Chlorpromazine (Thorazine) and trifluoperazine (Stelazine) are potent anticholinergics, and they impair erection and ejaculation. With some drugs, the seminal fluid backs up into the bladder rather than being propelled through the penile urethra. Patients still have a pleasurable sensation, but the orgasm is dry. When urinating after orgasm, the urine may be milky white because it contains the ejaculate. The condition is startling but harmless. Paradoxically, some rare cases of priapism have been reported with antipsychotics.

Antidepressant Drugs. The tricyclic and tetracyclic antidepressants have anticholinergic effects that interfere with erection and delay ejaculation. Because the anticholinergic effects vary among the cyclic antidepressants, those with the fewest effects (e.g., desipramine [Norpramin]) produce the fewest sexual adverse effects. The effects of the tricyclics and tetracyclics have not been documented sufficiently in women; however, few women seem to complain of any effects.

Some men report increased sensitivity of the glans that is pleasurable, and that does not interfere with erection, although it delays ejaculation. In some cases, however, the tricyclic causes painful ejaculation, perhaps as the result of interference with seminal propulsion caused by interference with, in turn, urethral, prostatic, vas, and epididymal smooth muscle contractions. Clomipramine (Anafranil) has been reported to increase sex drive in some persons. Selegiline (Deprenyl), a selective MAO type B (MAO_B) inhibitor, and bupropion (Wellbutrin) have also been reported to increase sex drive, possibly by dopaminergic activity and increased production of norepinephrine.

Venlafaxine (Effexor) and SSRIs most often have adverse effects because of the rise in serotonin levels. A lowering of the sex drive and difficulty reaching orgasm occur in both sexes. Reversal of those negative effects has been achieved with cyproheptadine (Periactin), an antihistamine with antiserotonergic effects, and with methylphenidate (Ritalin), which has adrenergic effects. Trazodone (Desyrel) is associated with the rare occurrence of priapism, the symptom of prolonged erection in the absence of sexual stimuli. That symptom appears to result from the α_2-adrenergic antagonism of trazodone.

The MAOIs affect biogenic amines broadly. Accordingly, they produce impaired erection, delayed or retrograde ejaculation, vaginal dryness, and inhibited orgasm. Tranylcypromine (Parnate) has a paradoxical sexually stimulating effect in some persons, possibly as a result of its amphetamine-like properties.

General Effects. Because depression is associated with a decreased libido, varying levels of sexual dysfunction and anhedonia are part of the disease process. Some patients report improved sexual functioning as their depression improves as a result of antidepressant medication. The phenomenon makes the evaluation of sexual side effects difficult; also, the side effects may disappear with time, perhaps because a biogenic amine homeostatic mechanism comes into play.

LITHIUM. Lithium (Eskalith) regulates mood and, in the manic state, may reduce hypersexuality, possibly by a dopamine antagonist activity. In some patients, impaired erection has been reported.

SYMPATHOMIMETICS. Psychostimulants, which are sometimes used in the treatment of depression, include amphetamines, methylphenidate, and pemoline (Cylert), which raise the plasma levels of norepinephrine and dopamine. Libido is increased; however, with prolonged use, men may experience a loss of desire and erections.

α-ADRENERGIC AND β-ADRENERGIC RECEPTOR ANTAGONISTS. α-Adrenergic and β-adrenergic receptor antagonists are used in the treatment of hypertension, angina, and certain cardiac arrhythmias. They diminish tonic sympathetic nerve outflow from vasomotor centers in the brain. As a result, they can cause impotence, decrease the volume of ejaculate, and produce retrograde ejaculation. Changes in libido have been reported in both sexes.

Suggestions have been made to use the side effects of drugs therapeutically. Thus, a drug that delays or interferes with ejaculation (e.g., fluoxetine [Prozac]) might be used to treat premature ejaculation.

ANTICHOLINERGICS. Anticholinergics block cholinergic receptors and include such drugs as amantadine (Symmetrel) and benztropine (Cogentin). They produce dryness of the mucous membranes (including those of the vagina) and erectile disorder. However, amantadine may reverse SSRI-induced orgasmic dysfunction through its dopaminergic effect.

ANTIHISTAMINES. Drugs such as diphenhydramine (Benadryl) have anticholinergic activity and are mildly hypnotic. They may inhibit sexual function as a result. Cyproheptadine, although an antihistamine, also has potent activity as a serotonin antagonist. It is used to block the serotonergic sexual adverse effects produced by SSRIs, such as delayed orgasm.

ANTIANXIETY AGENTS. The major class of anxiolytics is the benzodiazepines (e.g., diazepam [Valium]). They act on the γ-aminobutyric acid (GABA) receptors, which are believed to be involved in cognition, memory, and motor control. Because they decrease plasma epinephrine concentrations, they diminish anxiety, and as a result, they improve sexual function in persons inhibited by anxiety.

ALCOHOL. Alcohol suppresses CNS activity generally and can produce erectile disorders in men as a result. Alcohol has a direct gonadal effect that decreases testosterone levels in men; paradoxically, it can produce a slight rise in testosterone levels in women. The latter finding may account for women reporting increased libido after drinking small amounts of alcohol. The long-term use of alcohol reduces the ability of the liver to metabolize estrogenic compounds. In men, that produces signs of feminization (such as gynecomastia).

OPIOIDS. Opioids, such as heroin, have adverse sexual effects, such as erectile failure and decreased libido. The alteration of consciousness may enhance the sexual experience in occasional users.

HALLUCINOGENS. The hallucinogens include lysergic acid diethylamide (LSD), phencyclidine (PCP), psilocybin (from some mushrooms), and mescaline (from peyote cactus). In addition to inducing hallucinations, the drugs cause a loss of contact with reality and expand a sense of consciousness. Some users report that the sexual experience is similarly enhanced, but others experience anxiety, delirium, or psychosis, which clearly interferes with sexual function.

CANNABIS. The altered state of consciousness produced by cannabis may enhance sexual pleasure for some persons. Its prolonged use depresses testosterone levels.

BARBITURATES AND SIMILARLY ACTING DRUGS. Barbiturates and similarly acting sedative–hypnotic drugs may enhance sexual responsiveness in persons who are sexually unresponsive as a result of anxiety. They have no direct effect on the sex organs; however, they do produce an alteration in consciousness that some persons find pleasurable. They are subject to abuse and can be fatal when combined with alcohol or other CNS depressants.

Methaqualone (Quaalude) acquired a reputation as a sexual enhancer, which had no biologic basis in fact. It is no longer marketed in the United States.

Other Specified Sexual Dysfunction

Diagnoses in this category include dysfunctions that distress the patient but do not meet the criteria for the disorders discussed above. An example would be sexual aversion, in which the patients cannot enjoy the sex act because of a phobic aversion to touching the genitalia.

Unspecified Sexual Dysfunction

This category covers other sexual dysfunctions that do not fall under the diagnoses discussed above but that cause distress and interference with sexual connection and pleasure. In these cases, the clinician does not choose to elaborate on the differences from the diagnoses discussed above.

Further examples of other specified or unspecified disorders include individuals who experience the physiologic components of sexual excitement and orgasm but report no erotic sensation or even anesthesia and the male experience of orgasm with a flaccid penis. The orgasmic woman who desires but has not experienced multiple orgasms can be classified under this heading as well. Disorders of excessive rather than inhibited function, such as compulsive masturbation, might be diagnosed under unspecified dysfunction. These diagnoses also might be used to cover complaints engendered by a couple, rather than individual, dysfunction—for example, a couple in which one partner prefers morning sex and the other functions more readily at night or a couple with unequal frequencies of desire.

Compulsive Sexual Behavior. The concept of compulsive sexual behavior (Table 21.1–10), or sex addiction, was developed in the 1980s to describe persons who compulsively seek out sexual experiences and whose behavior becomes impaired if they cannot gratify their sexual impulses. The concept of sex addiction is derived from the model of addiction to drugs such as heroin or addiction to behavioral patterns such as gambling. Addiction implies psychological dependence, physical dependence, and a withdrawal symptom if the substance (e.g., the drug) is unavailable or the behavior (e.g., gambling) is frustrated.

DSM-5 does not have a diagnosis of *sex addiction* or *compulsive sexual behavior*. ICD 11 includes a diagnosis of compulsive sexual behavior. A person whose entire life revolves around sex-seeking behavior and activities, who spends an excessive amount of time in such behavior, and who often tries to stop such behavior but cannot do so is well known to clinicians. Such persons show repeated and increasingly frequent attempts to have a sexual experience and deprivation evokes symptoms of distress. Ten-step programs are usually a necessary part of the treatment of this behavior.

Persistent Genital Arousal Disorder. Persistent genital arousal disorder has previously been called persistent sexual arousal syndrome. It has been diagnosed in women who complain of a continual feeling of sexual arousal, which is uncomfortable, demands release, and interferes with life pleasures and activities. These women frequently masturbate, sometimes incessantly, because climax provides relief. However, the relief is temporary, and the sense of arousal returns rapidly and remains. The sense of arousal in these cases is neither pleasurable nor exciting, and the women are not interested in a sexual experience but in relief from their symptoms. Some women have reported masturbating so frequently to alleviate the arousal that they have irritated their genitalia severely. One case of attempted suicide has been reported with this syndrome, with the woman stating that she could no longer tolerate the sensations and that she had masturbated so often that her vulva was raw.

Table 21.1–10.
Signs of Sexual Addiction or Compulsive Sexual Behavior

1. Out-of-control behavior
2. Severe adverse consequences (medical, legal, interpersonal) due to sexual behavior
3. Persistent pursuit of self-destructive or high-risk sexual behavior
4. Repeated attempts to limit or stop sexual behavior
5. Sexual obsession and fantasy as a primary coping mechanism
6. Need for increasing amounts of sexual activity
7. Severe mood changes related to sexual activity (e.g., depression, euphoria)
8. Inordinate amount of time spent in obtaining sex, being sexual, or recovering from sexual experience
9. Interference of sexual behavior in social, occupational, or recreational activities

Data from Carnes P. Don’t Call It Love. New York: Bantam Books; 1991.

Krafft-Ebing reported a case of a young woman who masturbated almost incessantly and was unable to control her impulses. She had frequent coitus with many men, but neither coitus nor masturbation sufficed, and she was eventually placed in an institution. Krafft-Ebing referred to the condition as “sexual hyperaesthesia,” which he believed could occur in otherwise normal persons.

This case would have to be differentiated from sexual addiction. The differentiating factor would be whether the woman desired an orgasm for itself and looked forward to having sex or whether she was seeking relief from incessant and intolerable stimulation. There is some speculation that this disorder is due to nerve damage or anomaly, but the etiology is unknown.

Postcoital Dysphoria. Postcoital dysphoria is not listed in DSM-5. It occurs during the resolution phase of sexual activity when individuals normally experience a sense of general well-being and muscular and psychological relaxation. Some individuals, however, experience postcoital dysphoria—after an otherwise satisfactory sexual experience, they become depressed, tense, anxious, and irritable and show psychomotor agitation. They often want to get away from their partner and may become verbally or even physically abusive. The incidence of the disorder is unknown, but it is more common in men than in women. Its several causes relate to the person's attitude toward sex in general and the partner in particular. It may occur in adulterous sex and with prostitutes, when there is a profound fear of intimacy, or when individuals cannot experience sex without consequent strong feelings of guilt. The fear of sexually transmitted infection causes some persons to experience postcoital dysphoria. Treatment requires insight-oriented psychotherapy to help patients understand the unconscious antecedents to their behavior and attitudes.

Unconsummated Marriage. A couple involved in an unconsummated marriage has never had coitus and is typically uninformed and inhibited about sexuality. Their feelings of guilt, shame, or inadequacy are increased by their problems, and they experience conflict between their need to seek help and their need to conceal their difficulty. Couples present with the problem after having been married for several months or several years. Masters and Johnson reported an unconsummated marriage of 17 years duration.

Frequently, the couple does not seek help directly, but the woman may reveal the problem to her gynecologist on a visit ostensibly concerned with vague vaginal or other somatic complaints. On examining her, the gynecologist may find an intact hymen. In some cases, however, the wife may have undergone a hymenectomy to resolve the problem. Inquiry by a physician who is comfortable in dealing with sexual problems may be the first opening to a frank discussion of the couple's distress. Often, the pretext of the medical visit is a discussion of contraceptive methods or even a request for an infertility workup. Once presented, the complaint can often be successfully treated. The duration of the problem does not significantly affect the prognosis or the outcome of the case.

The causes of unconsummated marriage are varied: lack of sex education, sexual prohibitions overly stressed by parents or society, problems of an oedipal nature, immaturity in both partners, overdependence on primary families, and problems in sexual identification. Religious orthodoxy, with severe control of sexual and social development or the equation of sexuality with sin or uncleanliness, has also been cited as a dominant cause. Many women involved in an unconsummated marriage have distorted concepts about their vaginas. They may fear that the vagina is too small or too soft, or they may confuse the vagina with the rectum, leading to feelings of being unclean. The man may share in those distortions about the vagina and, in addition, perceive it as dangerous to him. Similarly, both partners may have distortions about the man's penis, perceiving it as a weapon, as too large, or as too small. Many patients can be helped by simple education about genital anatomy and physiology, by suggestions for self-exploration, and by correct information from a physician. The problem of the unconsummated marriage is best treated by seeing both members of the couple. Dual-sex therapy has been markedly effective. However, other forms of conjoint therapy, marital counseling, traditional psychotherapy on a one-to-one basis, and counseling from a sensitive family physician, gynecologist, or urologist are all helpful.

Body Image Problems. Some individuals are ashamed of their bodies and experience feelings of inadequacy related to self-imposed standards of masculinity or femininity. They may insist on sex only during total darkness, not allow certain body parts to be seen or touched, or seek unnecessary operative procedures to deal with their perceived inadequacies. Concerns about body image occur more frequently in women than in men. Certain people have difficulty adjusting to narcissistic injuries inflicted by age or illness and withdraw from sex as a result. Body dysmorphic disorder should be ruled out.

Don Juanism. Some men who appear to be hypersexual, as shown by their need to have many sexual encounters or conquests, use their sexual activities to mask deep feelings of inferiority. Some have unconscious homosexual impulses, which they deny by compulsive sexual contact with women. After having sex, most Don Juans are no longer interested in the woman. The condition is also referred to as *satyriasis* or as a form of sex addiction.

Nymphomania. Nymphomania signifies an excessive or pathologic desire for coitus in a woman. There have been few scientific studies of the condition. Those patients who have been studied usually have had one or more sexual disorders, usually including female orgasmic disorder. The woman often has an intense fear of loss of love or a continuing need to confirm her sexual desirability. She attempts to satisfy her dependency needs rather than to gratify her sexual impulses through her actions. It is sometimes classified as a form of sex addiction.

Fantasies. Other atypical disorders are found in individuals who have one or more sexual fantasies about which they obsess, feel guilty, or are otherwise dysphoric. As indicated in Table 21.1–11, however, the range of common sexual fantasies is broad.

Postcoital Headache. Postcoital headache is a headache immediately after coitus and may last for several hours. It is usually described as throbbing and is localized in the occipital or frontal area. The cause is unknown but may be vascular, due to muscle contraction (tension), or psychogenic. Coitus may precipitate migraines or cluster headaches in predisposed persons.

Table 21.1–11.
Common Sexual Fantasies

Men	Women
Heterosexual	
Replacement of established partner	Replacement of established partner
	Forced sexual encounters with men
Forced sexual encounters with women	Observing sexual activity Idyllic encounters with unknown men
Observing sexual activity	
Sexual encounters with men	Sexual encounters with women
Group sex	
Homosexual	
Images of male anatomy	Forced sexual encounters with women Idyllic encounters with established partner
Forced sexual encounters with men Sexual encounters with women	
Idyllic encounters with unknown men	Sexual encounters with men
	Memories of past sexual experiences
Group sex	Sadistic imagery

Adapted with permission from the Schwartz MF, Masters WH. The Masters and Johnson treatment program for dissatisfied homosexual men. *Am J Psychiatry*. 1984;141(2):180.

Orgasmic Anhedonia. Orgasmic anhedonia is a condition in which the person had no physical sensation of orgasm, even though the physiologic component (e.g., ejaculation) remains intact. Medical causes, such as sacral and cephalic lesions that interfere with afferent pathways from the genitalia to the cortex, must be ruled out. Psychological causes usually relate to extreme guilt about experiencing sexual pleasure—these feelings produce a type of dissociative response that isolates the affective component of the orgasmic experience from consciousness.

Female Premature Orgasm. Data on female premature orgasm are lacking; no separate category for premature orgasm in women is included in DSM-5. However, in one study, 10% of women felt they reached orgasm too quickly.

Masturbatory Pain. Some individuals may experience pain during masturbation. Organic causes should always be ruled out. A small vaginal tear or early Peyronie disease may produce a painful sensation. The condition should be differentiated from compulsive masturbation. People may masturbate to the extent that they do physical damage to their genitals and eventually experience pain during subsequent masturbatory acts.

Certain masturbatory practices have resulted in what has been called *autoerotic asphyxiation,* which is usually classified as a paraphilia (hypoxyphilia). The practices may involve masturbating while hanging by the neck to heighten erotic sensations and the intensity of the orgasm through the mechanism of mild hypoxia. Although they intend to release themselves from the noose after orgasm, an estimated 500 to 1,000 persons a year accidentally kill themselves by hanging. Most who indulge in the practice are men; transvestism is often associated with the habit, and most deaths occur among adolescents. Such masochistic practices are usually associated with severe mental disorders, such as schizophrenia and major mood disorders.

TREATMENT

The treatment of sexual disorders has evolved significantly since the 1970s when Masters and Johnson focused the attention of the psychiatric community on sexual disorders. Innovations in treatment reflect the results of research and changes in the patient population. For example, in the late 1960s, most cases of erectile dysfunction were considered psychological in origin, with approximately 20% of erectile problems having an organic cause. Currently, the numbers are reversed, with the majority of cases considered to result from a physiologic problem. Actually, many cases are of mixed origin. Similarly, early patients in sex therapy were assumed to suffer, in part, from lack of sexual information and culturally reinforced negative attitudes toward sex. In the past 50 years, the public has received a great deal of accurate information about sex through the media. Patients now have more knowledge and greater sexual sophistication. Their dysfunctions frequently have a complex etiology involving psychodynamic and relational issues. Problems of desire are seen with increasing frequency and are among the most challenging cases for therapists. Couples' issues involve problems of trust, intimacy, lack of sexual attraction, and struggles for dominance. Finally, along with the rest of psychiatry, sex therapy has experienced medicalization. Biologic treatment approaches have developed rapidly. An eclectic approach that allows the use of several techniques sequentially or in combination may be necessary.

Before entering therapy, the patient may need a thorough medical evaluation, including medical history, physical examination, and appropriate laboratory studies when necessary. If a medical cause for the disorder is found, treatment should be directed toward ameliorating that cause.

Before 1970, the most common treatment of sexual dysfunction was individual psychotherapy. Classic psychodynamic theory considers sexual inadequacy to have its roots in early developmental conflicts, and the sexual disorder is treated as part of a more pervasive emotional disturbance. Treatment focuses on the exploration of unconscious conflicts, motivation, fantasy, and various interpersonal difficulties. Therapy assumes that removal of the conflicts will allow the sexual impulse to become structurally acceptable to the patient's ego and thereby find appropriate means of satisfaction in the environment. Unfortunately, the symptom of sexual dysfunction frequently becomes secondarily autonomous and persists after the resolution of the other problems evolving from the patient's pathology. The addition of behavioral techniques is often necessary to cure the sexual problem.

Dual-Sex Therapy

The theoretical basis of the dual-sex therapy approach is the concept of the marital unit or dyad as the object of therapy. The method of dual-sex therapy was originated and developed by Masters and Johnson. Dual-sex therapy does not accept the idea of a dysfunctional half of a patient couple. Both individuals are involved in a relationship in which there is sexual distress, and, thus, both should participate in the therapy program.

The sexual problem often reflects other areas of disharmony or misunderstanding in the marriage. The marital relationship as a whole is treated with emphasis on sexual functioning as a part of that relationship. Improved communication in sexual and nonsexual areas is a specific goal of treatment. Psychological and physiologic aspects of sexual functioning are discussed with an educational attitude. Suggestions are followed in the privacy of the couple's home.

Initial histories are taken to determine suitability for this type of treatment. Evidence of major underlying psychopathology suggests further psychiatric evaluation, and participation in the program may be deferred until the patient seems better able to benefit from it. Concurrent psychotherapy with a psychiatrist while participating in dual-sex therapy is sometimes recommended.

Each patient is interviewed individually early in the course of treatment. A complete sexual history is obtained, which is later reflected back to the couple to help them understand their present problem. For example, a history of culturally or parentally transmitted inhibition can help relieve the partner of an anorgasmic woman of the idea that he is a bad lover. Similarly, the partner of a man with erectile disorder may be relieved of the concern that his problem is a result of her being insufficiently attractive. The individual sessions also help the therapist understand the patients' lifestyle and allow for suggestions that fit into that lifestyle.

Behavioral Exercises. Treatment is short term and behaviorally oriented. Specific exercises are prescribed to help the couple with their particular problem. Sexual dysfunction often involves a fear of inadequate performance; thus, couples are specifically prohibited from any sexual play other than that prescribed by the therapist. Initially, intercourse is interdicted, and couples learn to give and receive bodily pleasure without the pressure of performance. Beginning exercises usually focus on heightening sensory awareness to touch, sight, sound, and smell.

During these exercises, called *sensate focus exercises,* the couple is given much reinforcement to lessen anxiety. They are urged to use fantasies to distract them from obsessive concerns about performance, which is termed *spectatoring.* The needs of both the dysfunctional partner and the nondysfunctional partner are considered. If either partner becomes sexually excited by the exercises, the other is encouraged to bring them to orgasm by manual or oral means. This procedure is important to keep the nondysfunctional partner from sabotaging the treatment. Open communication between the partners is urged, and the expression of mutual needs is encouraged. Resistances, such as claims of fatigue or not enough time to complete the exercises, are common and must be dealt with by the therapist. Genital stimulation is eventually added to general body stimulation. The couple is sequentially taught to try various positions for intercourse without necessarily completing the act and using varieties of stimulating techniques before they are permitted to proceed with intercourse.

The specific exercises vary with differing presenting complaints, and special techniques are used to treat the various dysfunctions. In cases of vaginismus, for instance, the woman is advised to dilate her vaginal opening with her fingers or with size-graduated vaginal dilators as part of the therapy. She may also be referred to a physiotherapist specializing in pelvic disorders. In cases of premature ejaculation, an exercise known as the *squeeze technique* is used to raise the threshold of penile excitability. In this exercise, the man or the woman stimulates the erect penis until the earliest sensations of impending orgasm and ejaculation are felt. Penile stimulation is then stopped abruptly, and the coronal ridge of the penis is forcibly squeezed for several seconds. The technique is repeated several times. A variation is the *stop–start technique,* in which stimulation is interrupted for several seconds, but no squeeze is applied. Masturbation to the point of imminent orgasm raises the threshold of excitability to a more tolerant stimulation level. The man is encouraged to focus on sensations of excitement rather than distract himself from them. This makes him more familiar with his excitement pattern and lets him feel in control rather than overwhelmed by sensations of arousal. Communication between the partners is improved because the man must let his partner know his level of sexual excitement so that she can squeeze the penis before the ejaculatory process has started. Sex therapy has been successful with some premature ejaculators; however, a subgroup of dysfunctional men may need pharmacotherapy as well.

A man with sexual desire disorder or erectile disorder is sometimes told to masturbate to demonstrate that full erection and ejaculation are possible. A woman with a lifelong female orgasmic disorder is directed to masturbate, sometimes using a vibrator. Kegel exercises may be introduced to strengthen the pubococcygeal muscles—that is, the woman is encouraged to contract her abdominal and perineal muscles during masturbation and coitus. When a man has an erectile disorder, the woman may be instructed to stimulate or tease his penis. The same technique is used with men who suffer from retarded ejaculation, with stimulation sometimes involving a vibrator. Delayed ejaculation is managed by extravaginal ejaculation initially and gradual vaginal entry after stimulation to the point of near ejaculation.

Treatment Goals. The overall goal of treatment is to initiate an educational process, to diminish the fears of performance felt by both sexes, and to facilitate communication in sexual and nonsexual areas. Therapy sessions follow each new exercise period, and problems and satisfactions (both sexual and nonsexual) are discussed. Specific instructions and new exercises geared to the individual couple's progress are reviewed in each session. Gradually, the couple gains confidence and learns (or relearns) to communicate verbally and sexually. Dual-sex therapy is most effective when the sexual dysfunction exists apart from other psychopathology.

Hypnotherapy

Hypnotherapists focus specifically on the anxiety-producing symptom—that is, the particular sexual dysfunction. Successful use of hypnosis helps the patient gain control over the symptom that has been lowering self-esteem and disrupting psychological homeostasis. Patient cooperation is first obtained and encouraged during a series of nonhypnotic sessions with the therapist, designed to develop a secure doctor–patient relationship and a sense of physical and psychological comfort on the part of the patient and to establish mutually desired treatment goals. During that time, the therapist assesses the patient's capacity for the trance experience. The nonhypnotic sessions also permit the clinician to take a careful psychiatric history and do a mental status examination before beginning hypnotherapy. Treatment focuses on symptom removal and attitude alteration. In a trance state, patients can entertain ideas incongruent with their usual (nonhypnotized) perceptions of reality. Patients are instructed in developing alternative means of dealing with the anxiety-provoking situation (i.e., the sexual encounter).

For example, a woman with vaginismus is given the posthypnotic suggestion that she will feel no pain during intercourse and will be able to relax the muscles surrounding her vagina. If compliance with the suggestion is successful, she can deal with the anxiety produced by the sex act. She is also taught new attitudes, such as being entitled to sexual pleasure. Under hypnosis, her fear or anger at sexual contact can be examined, and she learns how her emotions are expressed by involuntary vaginal spasms.

Some patients respond particularly well to the use of self-hypnosis and indirect suggestion. These techniques allow them to retain a greater sense of control over their situation. Typically, patients are instructed to conjure up images and develop ideas antithetical to their dysfunctional responses. For example, a woman with an arousal disorder may first agree to concentrate on imagery that causes her to salivate. She is then told that, just as she has made her mouth water by focusing on stimulating images, she can affect the lubricating response of her vagina by focusing on images she finds erotic or romantic. At the same time, the therapist helps her deal with her anxieties about a positive sexual response. Patients are also taught relaxing techniques to use before sexual relations. With these methods to alleviate anxiety, the physiologic responses to sexual stimulation can more readily result in pleasurable excitation and discharge. Hypnosis may be added to a basic individual psychotherapy program to accelerate the impact of psychotherapeutic intervention.

Behavior Therapy

Behavior therapists assume that sexual dysfunction is learned, maladaptive behavior. Behavioral approaches were initially designed to treat phobias. In cases of sexual dysfunction, the therapist sees the patient as phobic of sexual interaction. Using traditional techniques, the therapist sets up a hierarchy of anxiety-provoking situations for the patient, ranging from the least threatening to the most threatening. Mild anxiety may be experienced at the thought of kissing, and massive anxiety may be felt when imagining penile penetration. The behavior therapist helps the patient master the anxiety through a standard program of systematic desensitization. The program is designed to inhibit the learned anxious response by encouraging behaviors antithetical to anxiety. The patient first deals with the least anxiety-producing situation in fantasy and progresses by steps to the most anxiety-producing situation. Medication, hypnosis, or special training in deep muscle relaxation is sometimes used to help with the initial mastery of anxiety.

Assertiveness training helps teach patients to express their sexual needs openly and without fear. Exercises in assertiveness are given in conjunction with sex therapy, and patients are encouraged both to make sexual requests and to refuse to comply with requests perceived as unreasonable. Sexual exercises may be prescribed for patients to perform at home, and a hierarchy may be established, starting with activities that proved most pleasurable and successful in the past.

One treatment variation involves the participation of the patient's sexual partner in the desensitization program. The partner, rather than the therapist, presents the hierarchical items to the patient. In such situations, a cooperative partner is necessary to help the patient carry gains made during treatment sessions to sexual activity at home.

Behavior therapy techniques have been particularly effective in treating women with severe inhibition of excitement and orgasm when such feelings were accompanied by strong feelings of anxiety, anger, or disgust.

Group Therapy

Methods of group therapy have been used to examine both intrapsychic and interpersonal problems in patients with sexual disorders. The therapy group provides a strong support system for patients who feel ashamed, anxious, or guilty about a particular sexual problem. It is a useful forum in which to counteract sexual myths, correct misconceptions, and provide accurate information regarding sexual anatomy, physiology, and varieties of behavior.

Groups for the treatment of sexual disorders can be organized in several ways. Members may all share the same problem, such as premature ejaculation; members may all be of the same sex and have different sexual problems, or groups may be composed of both men and women who are experiencing different sexual problems. Group therapy may be an adjunct to other forms of therapy or the prime mode of treatment. Groups organized to cure a particular dysfunction usually have a behavioral approach. For example, patients with anorgasmia may participate with others who have the same problem in a short-term, intensive group experience. Sexual histories, feelings of inadequacy, and concerns about body image are shared. Specific physiologic information, sometimes with the aid of audiovisual materials, is presented to the group members. Members are given homework assignments (e.g., they may be instructed to masturbate). A combination of group support and group pressure helps the participants complete assignments some might otherwise avoid. As the short-term group process nears termination, members are encouraged to talk about their experiences with their partners.

Groups have also been effective when composed of sexually dysfunctional married couples. The group provides an opportunity to gather accurate information, provides consensual validation of individual preferences, and enhances self-esteem and self-acceptance. Such techniques as role-playing and psychodrama may be used in treatment. These groups are not indicated for couples in which one partner is uncooperative, has severe depression or psychosis, or has a strong repugnance for explicit sexual audiovisual materials or a strong fear of groups.

Integrated Sex Therapy

One of the most effective treatment modalities is the use of sex therapy integrated with supportive, psychodynamic, or insight-orientated psychotherapy. Adding psychodynamic conceptualizations to the behavioral techniques used to treat sexual dysfunctions allows treatment of patients with sex disorders associated with other psychopathology. Also, this type of therapy is appropriate for patients with hypoactive desire disorders or interest/arousal disorder. Insight-oriented therapy helps them deal with problems in their interpersonal relationships or intrapsychic conflicts that frequently are at the root of the problem. The themes and dynamics that emerge in patients in analytically oriented sex therapy are the same as those that emerge in psychoanalytic therapy—relevant dreams, fear of punishment, aggressive feelings, difficulty with trusting the partner, fear of intimacy, oedipal feelings, and fear of genital mutilation.

Mr. K, a 62-year-old divorced man, presented for therapy with a chief complaint of erectile dysfunction with his current partner, Miss B. They had been living together for 4 years. At the beginning of their relationship, they had had an active and nondysfunctional sex life. He was a successful businessman who met Miss B at a business convention where she was a hostess/welcomer at the convention center. Mr. K was very attracted to her. He also liked her poise and social ease and felt he could be proud of her.

Shortly after they began living together, he began seeing other women as well. Miss B eventually found out and insisted that he stop seeing other women. She also demanded that they get married. Mr. K agreed to stop his infidelities and agreed to become engaged. His erectile difficulties with Miss B, which had started somewhat before this confrontation, became persistent and pronounced. Even though he was able to achieve and maintain erections if he used sildenafil, he began avoiding sex with her altogether. He did not think of leaving the relationship (although he broke off the engagement) because he felt he was aging, feared loneliness, and lacked confidence that he could find another desirable woman. Also, he had experienced similar problems in his marriage and had enough insight to realize that his pattern of behavior was destructive. (Miss B had her own set of issues that kept her in the relationship.) She had become depressed but was more hopeful since Mr. K had agreed to start therapy.

Mr. K had a poised, secure façade, but he was hypersensitive to criticism, frequently felt like an outsider in social groups, relied on reassurance from women for his self-worth, and dealt with his anxieties with denial and sexual impulsiveness.

Mr. K’s issues reflected experiential and developmental problems. It had been hard for him to establish peer relationships growing up due to the family’s frequent travels and relocations. His father was a moderately successful, demanding man who was competitive with Mr. K and would send him into inappropriate situations. For example, he would send him on errands in a new location where, as a boy, Mr. K did not know his way around, and be critical of Mr. K if he failed in the errand; or he would tell him to “stand up for yourself” with boys who were substantially older and stronger than the patient. At the same time, Mr. K was his mother’s “darling, her everything.” Resulting unresolved oedipal issues made it difficult for him to direct both tender and passionate feelings toward the same woman. Unresolved dependency needs made it hard for him to discuss differences he had with Miss B in a constructive, assertive way, and he denied cumulative resentments (appropriate or inappropriate).

These issues were addressed in introspective psychotherapy. Miss B was also referred for individual therapy, and the couple was seen together for behaviorally oriented sex therapy.

The combined approach of individual and sex therapy is used by the general psychiatrist, who carefully judges the optimal timing of sex therapy and the ability of patients to tolerate the directive approach that focuses on their sexual difficulties.

Biologic Treatment Methods

Biologic treatments, including pharmacotherapy and surgery, have applications in specific cases of sexual disorder. Advances in biologic treatment methods, particularly pharmacologic methods of treating sexual dysfunction, have significantly augmented the therapist's catalog of therapeutic approaches. Most of the recent advances involve male sexual dysfunctions. Current studies are under way to test pharmacologic treatment of sexual dysfunctions in women. Flibanserin is a drug approved by the U.S. Food and Drug Administration (FDA) to treat low sexual interest in women. Questions about the drug included a report of a minimal increase in desire during trials and concerns about side effects, particularly hypotension and syncope. Physicians must undergo a certification process to prescribe flibanserin.

Pharmacotherapy. A variety of drugs have been explored in the treatment of male sexual dysfunction. The major medications used are nitric oxide enhancers such as sildenafil (Viagra); oral prostaglandin (Vasomax); alprostadil (Idex), injectable phentolamine; and transurethral alprostadil (MUSE), all used in the treatment of erectile disorder.

NITRIC OXIDE ENHANCERS. Nitric oxide enhancers facilitate the inflow of blood to the penis necessary for an erection. The physiologic mechanism of penile erection involves the release of nitric oxide in the corpus cavernosum during sexual stimulation. Nitric oxide activates the enzyme guanylate cyclase, which increases cyclic guanosine monophosphate concentration and produces smooth muscle relaxation in the corpus cavernosum that allows the penile vessels to dilate and admit blood. The first drug developed, sildenafil enhances the effect of nitric oxide by inhibiting the enzyme that degrades cyclic guanosine monophosphate. Thus, sildenafil augments the natural process involved in gaining and maintaining an erection during sexual stimulation. The drug takes effect approximately 1 hour after ingestion, and its effect can last up to 4 hours. Sildenafil has no effect in the absence of sexual stimulation.

The most common adverse events associated with sildenafil are headaches, flushing, and dyspepsia. Some sildenafil users see things in a blue tint for several hours after taking the medication; because of this, airline pilots have been prohibited from taking the drug so that this visual artifact does not interfere with safe landings. A more serious visual side effect is the loss of visual acuity and nonarteritic anterior ischemic optic neuropathy resulting in blindness. Although the occurrence is extremely rare, phosphodiesterase type 5 inhibitors should not be used with men with serious optic pathology. The use is also contraindicated for people taking organic nitrates. The concomitant action of the two drugs can result in large, sudden, and sometimes fatal drops in systemic blood pressure. The U.S. FDA has posted 522 deaths as of 2014, in which sildenafil was listed as an associated medication and advises caution in prescribing sildenafil to men with a recent (6-month) history of myocardial infarction, stroke, life-threatening arrhythmia, significant hypotension or hypertension, cardiac failure, angina, or retinitis pigmentosa. Sildenafil has also been implicated in cases of deafness. One study of Swedish men using phosphodiesterase 5 inhibitors found a modestly significant correlation between use and the presence of malignant melanoma. The researchers stated that this association was not proved to be causal. In association with side-effect alerts, the FDA has reiterated that it is a safe drug. Sildenafil is not effective in all cases of erectile dysfunction. It fails to produce an erection rigid enough for penetration in approximately 50% of men who have had radical prostate surgery or in those with long-standing, insulin-dependent diabetes. It is also ineffective in certain cases of nerve damage.

Two other nitric oxide enhancers similar to sildenafil—vardenafil (Levitra) and tadalafil (Cialis)—have been developed. Tadalafil has an effective therapeutic window of 36 hours compared to sildenafil and vardenafil, which are effective for approximately 4 hours.

Oral phentolamine has proved effective as a potency enhancer in men with minimal erectile dysfunction. It may prove useful for men with cardiac problems, as sildenafil is contraindicated for men using organic nitrates, but it is not currently approved by the FDA. Apomorphine is also being tested as an oral remedy for erectile dysfunction.

ALPROSTADIL. In contrast to the oral medications, injectable and transurethral alprostadil act locally on the penis and can produce erections in the absence of sexual stimulation. Alprostadil contains a naturally occurring form of prostaglandin E. Prostaglandins are composed of complex hydroxy fatty acids, and they have wide biologic influences. Although some prostaglandins are vasoconstrictive, prostaglandin E_1, found in alprostadil, is a powerful vasodilating agent, especially in local vascular areas, such as the corpus cavernosum of the penis. The drug causes direct smooth muscle relaxation of penile vessels and erectile tissue; this reaction lowers the vascular resistance of the corpus and significantly increases blood flow to the penis. The firm erection produced within 2 to 3 minutes by increased blood flow may last as long as 1 hour. Treatment consists of the patient’s self-injection of alprostadil into the corpus before coitus. This technique is easily taught and relatively painless. Infrequent adverse effects include penile bruising and changes in liver function test results, which are readily reversible when a man stops the injections. However, possible hazardous sequelae exist, including priapism and sclerosis of the small veins of the penis. In some men for whom alprostadil was not effective, an injection of Trimix, a combination of prostaglandin E, phentolamine (an αl-adrenergic antagonist), and papaverine (a nonspecific phosphodiesterase inhibitor) produced an erection. Another substance being tried is a vasoactive intestinal polypeptide (VIP). Intracavernous injection of VIP causes erection and has a parasympathomimetic effect. Some researchers speculate that this substance, which has been found in the hypothalamus and the female genital organs, is the essential factor in male and female arousal. In Europe, phenoxybenzamine (Dibenzyline) is used to produce erections by injection into the penis. Serious adverse effects include priapism and pain accompanying the injection, and the drug is not allowed as a therapy in the United States.

Alprostadil can also be delivered via the urethra, eliminating the need for self-injection. Some men prefer the local, nonsystemic effects of alprostadil to oral sildenafil; others prefer sildenafil because it seems more like a nonpharmacologically aided response to them and their partners. A small trial found a topical cream effective in alleviating erectile dysfunction. The cream consists of three vasoactive substances that are a mixture of ergot alkaloids. Alprostadil is a useful treatment for patients in whom nitric oxide enhancers are contraindicated (e.g., patients who are taking nitrate-containing medication). It is also of use in patients who do not tolerate the side effects of other drugs and may work for some patients who have found sildenafil ineffective, for example, in some postprostatectomy patients.

The pharmacologic treatments described above are useful in the treatment of erectile dysfunction of various causes: psychogenic, neurogenic, arterial insufficiency, venous leakage, and mixed.

Mr. J presented for therapy with a chief complaint of erectile disorder. He was a 49-year-old man who had been divorced for 6 years after a marriage of 12 years duration. Although he suffered no erectile dysfunction in his marriage, and in fact, experienced strong desire, his wife refused to have sex with him in the last 2 years of their marriage. By that time, their relationship had seriously deteriorated. Although it was Mr. J's decision to end the marriage, he felt hurt and humiliated by his ex-wife's sexual rejection and was wary of ever committing to someone else. His sense of rejection evoked memories of a painful adolescence when he had felt very inadequate with girls and was fearful of being mocked by male peers. After his divorce, he had a series of dating relationships with women but suffered from recurrent erectile disorder with all of them.

When he came for therapy, he was in the midst of a new relationship. The patient was prescribed sildenafil, 50 mg, but he was reluctant to tell his girlfriend that he was taking a medication to help him function sexually. Instead, he avoided sex with her. Several therapy sessions were devoted to encouraging him to share his sexual problem with her. He finally did so, and his girlfriend, a woman with confidence in her sexuality, was accepting of his problem. In fact, she encouraged him to take the sildenafil. After 3 months of an active sex life without erectile problems, Mr. J was told to try intercourse without sildenafil. With the emotional support of his girlfriend, he made a successful attempt at coitus. The patient continues to use sildenafil occasionally. The knowledge that he has a "backup" to help him with his erection if he is feeling very anxious or stressed enables him to maintain a normal active sex life.

Erectile dysfunction of psychological or mixed origin should not be treated by medication alone, even though the dysfunction can be corrected pharmacologically. The drugs should be used in conjunction with, not as a replacement for, sex therapy. In some cases, erectile dysfunction serves as a defense against unconscious conflicts that must be faced when the sexual symptom is removed as a result of medication. In some cases, the patient deals with his conflicts by simply not using the medications prescribed for his erectile problem. Also, patients in long-standing or marital relationships require the cooperation of their partners for this treatment to be effective. Although most women are very accommodating to their partner's desire for treatment, several concerns occur with some frequency: the element of romance is an important part of sexual interaction for many women, and pharmacologic assistance of erections may eliminate that sense of romance; the woman, or the couple, may bemoan the lack of spontaneity when pharmacotherapy is part of the sex act; and some women feel deprived of feedback about their desirability. Joint sessions help the couple cope with these issues.

SILDENAFIL USE IN WOMEN. The physical sign of sexual excitement in women is vaginal lubrication. That lubrication is a transudate believed to result from increased vasocongestion of the extensive capillary net surrounding the vaginal walls. The same physiologic mechanism—vasocongestion—results in erection in the man. Researchers believe that sildenafil may facilitate blood flow in women just as it does in men and, thus, give women with inhibited excitement of psychological or physiologic origin a pharmacologic remedy for their dysfunction. Some reports suggest that women find it preferable to apply sildenafil in a cream base to the clitoris, labia, and vagina. However, as in men, pharmacologic treatment may need to be combined with psychological modalities to be effective.

Recognition of sexual excitement may be more complex for women than it is for men. For example, in studies of response to pornography, men and women underwent physiologic measurements of excitement after visual stimuli. In these studies, men were more accurate in correlating their subjective sense of arousal with physiologic arousal (i.e., erection) than were women (i.e., lubrication).

Other Pharmacologic Agents

ANTIDEPRESSANTS. The side effects of antidepressants, particularly the SSRIs, which include delayed orgasm, have been used to prolong the sexual response in patients with premature ejaculation. This approach is particularly useful in patients refractory to behavioral techniques or who may have physiologically determined premature ejaculation. Daily treatment is recommended. The SSRIs vary in their ability to delay ejaculation, with paroxetine (20 to 40 mg) being the most effective, followed by clomipramine (10 to 50 mg), sertraline (50 to 100 mg), fluoxetine (20 to 40 mg), citalopram (20 to 40 mg), and escitalopram (10 to 20 mg). On-demand use of SSRIs can be tried but is less effective. However, on-demand use of clomipramine (20 to 40 mg) may delay ejaculation after 4 to 6 hours, although it may lead to nausea on the same day or the next day. Long-term adverse effects of daily SSRI use include weight gain with its associated risk for type 2 diabetes mellitus. Also, depressed patients must be monitored for suicidal thoughts, and bipolar patients monitored for mania. Other on-demand treatments for premature ejaculation include topical anesthetics such as lidocaine or prilocaine cream or spray, which is applied to the glans of the penis. Use must be titrated to prevent sufficiently strong anesthesia that interferes with erection. Sometimes condom use is necessary to prevent vaginal numbness.

Other Pharmacologic Agents. Numerous other pharmacologic agents have been used to cure various sexual disorders. Intravenous methohexital sodium (Brevital) has been used in desensitization therapy. Antianxiety agents may have some application in tense patients, although these drugs can also interfere with the sexual response. The use of antidepressants has been advocated in the treatment of patients who are phobic of sex and in those with a post-traumatic stress disorder (PTSD) after rape. Bromocriptine (Parlodel) is used in the treatment of hyperprolactinemia, which is frequently associated with hypogonadism. Such cases are first worked up to rule out pituitary tumors. Bromocriptine, a dopamine agonist, may improve sexual function impaired by hyperprolactinemia.

Yohimbine is an α-adrenergic receptor antagonist that may cause dilation of the penile artery and improve erections. Recreational drugs, including cocaine, amphetamines, alcohol, and cannabis, are considered enhancers of sexual performance. Although they may provide the user with an initial benefit because of their tranquilizing, disinhibiting, or mood-elevating effects, consistent or prolonged use of any of these substances impairs sexual functioning.

Ginseng has been reported to have androgenic effects. One report described the case of a mother who ingested large amounts of ginseng during her pregnancy, resulting in androgenization of the neonate, who was born with pubic hair and enlarged testes.

Other drugs that have been used by women to alleviate arousal dysfunction include oral phentolamine, topical prostaglandin E, oral oxytocin, ginkgo biloba, and various psychostimulants, including caffeine. Many of these drugs have not been approved for the treatment of female sexual dysfunction and must be prescribed with caution.

Dopaminergic agents have been reported to increase libido and improve sex function. Those drugs include L-dopa, a dopamine precursor, and bromocriptine, a dopamine agonist. The antidepressant bupropion has dopaminergic effects and has increased sex drive in some patients. Selegiline, an MAOI, is selective for MAO_B and is dopaminergic. It improves sexual functioning in older persons.

HORMONE THERAPY. Androgens increase the sex drive in women and in men with low testosterone concentrations. Women may experience virilizing effects, some of which are irreversible (e.g., deepening of the voice). This problem is avoided if the woman applies a mild (compounded) testosterone ointment to her labia, inner thighs, or abdomen. Ongoing care by an endocrinologist is necessary for women who require more intensive testosterone treatment. Testosterone can be administered to men through injections, transdermal patches, or as gels, creams, or pellets. Oral preparations are associated with an increased risk of hepatotoxicity. Side effects may include hypertension and prostate enlargement, among other side effects. However, recent studies question effects on the prostate.

Clomiphene, usually used as a fertility drug for women, also stimulates testosterone production in men by action on the pituitary–gonad feedback cycle.

Gonadotrophin-releasing hormone (GnRH), also known as *luteinizing hormone–releasing hormone* (LHRH), stimulates the release of luteinizing hormone, which increases testosterone secretion in both sexes. GnRH is used as an inhalant in Europe. It stimulates desire and increases potency. Because GnRH is released normally in a pulsatile fashion, portable infusion pumps have been developed that simulate pulsatile delivery. An excess of GnRH suppresses estrogen and testosterone; thus, the therapeutic use of GnRH is limited by a narrow therapeutic window.

Women who use estrogens for HRT or for contraception may report decreased libido; in such cases, a combined preparation of estrogen and testosterone has been used effectively. Estrogen itself prevents thinning of the vaginal mucous membrane and facilitates lubrication. It may also increase tactile sensitivity. Vaginal estrogen creams have been helpful to postmenopausal women who cannot take HRT due to their medical or family history.

Testosterone is given to women around the world in the form of tablets, implants, patches, and creams. In the United States, methyltestosterone pills and compounded creams are the primary methods of administration. Testosterone is primarily administered to postmenopausal (occurring surgically or naturally) women, but testosterone deficiency also exists in premenopausal women.

Mrs. M presented for therapy alone with a chief complaint of pain on intercourse dating from the onset of menopause 14 months before coming for treatment. Her husband knew of her visit but did not accompany her, as he was away on one of many routine and frequent business trips.

Sexual intercourse was infrequent because of his heavy travel schedule, but their sex play was varied, and, before menopause, Mrs. M had enjoyed sex. Mrs. M found her sexual interactions with her husband gratifying. She had a strong libido and would masturbate when Mr. M was away.

Mrs. M was on oral HRT, which had helped but had not eliminated the pain she had with intercourse. In consultation with her gynecologist, the occasional use of hormone cream was added to her treatment regimen. In addition, she was instructed to use vaginal dilators routinely when Mr. M was away. This was not for purposes of masturbation, which for her involved clitoral stimulation, but for routine stretching of the vagina. Nonhormonal vaginal cream was prescribed for the couple to use with intercourse. The combination of oral HRT, vaginal creams, and the routine use of vaginal dilators served to eliminate Mrs. M’s pain. She was diagnosed with sexual dysfunction due to a general medical condition.

Pheromones are sexual scents that are found in animals and may be present in humans. They produce dramatic sex-seeking behavioral patterns in animals (e.g., male deer following female deer in estrus, mounting behavior in primates). Human pheromones are believed to be short-acting fatty acids present in vaginal secretions and male sweat. In one study, women were consistently more attracted to items impregnated with a chemical derived from male sweat (α-androstenol) than to control items. In another study, the sweat of women was preserved in underarm pads they changed daily, and the date each pad was worn was correlated with the women's menstrual cycle. A second group of women was asked to smell these pads as they were rubbed above their upper lips, without knowing what they were and recognizing no scent but the alcohol preservative used on the pads. Depending on whether they were exposed to pads from the early or late part of the first group's (the wearers) menstrual cycle, the second group (the sniffers) saw their own menstrual cycles shortened or lengthened. This area is still being researched.

ANTIANDROGENS AND ANTIESTROGENS. Estrogen and progesterone are antiandrogens that have been used to treat compulsive sexual behavior in men, usually in sex offenders.

Medroxyprogesterone acetate (Depo-Provera), used primarily as a contraceptive in women, inhibits the secretion of gonadotropins. It is used in men with compulsive sexual behavior to reduce libido by lowering testosterone levels. Cyproterone acetate is a strong antiandrogen used in Europe to treat sex offenders. At dosages of 100 to 200 mg a day, the sex drive disappears within 2 weeks.

Tamoxifen (Nolvadex) is used to treat breast cancer and has antiestrogenic properties. In some women, increased libido may result from unopposed testosterone levels. Clomiphene (Clomid) is an ovulatory stimulant used in women with anovulatory menstrual cycles desiring pregnancy. Clomiphene increases the release of pituitary gonadotropins, and some women may report increased libido. Neither of these drugs, however, is used as a treatment for decreased sexual interest in women.

Mechanical Treatment Approaches

STEAL SYNDROME. In male patients with arteriosclerosis (especially of the distal aorta, known as *Leriche syndrome*), the erection may be lost during active pelvic thrusting. The need for increased blood in the gluteal muscles and others served by the ilial or hypogastric arteries takes blood away (steals) from the pudendal artery and, thus, interferes with penile blood flow. Relief may be obtained by decreasing pelvic thrusting, which is also aided by the woman-superior coital position.

VACUUM PUMP. Vacuum pumps are mechanical devices that patients without vascular disease can use to obtain erections. The blood drawn into the penis after the creation of the vacuum is kept there by a ring placed around the base of the penis. This device has no adverse effects, but it is cumbersome, and partners must be willing to accept its use. Some women complain that the penis is redder and cooler than when erection is produced by natural circumstances, and they find the process and the result objectionable.

A vacuum pump with the marketing name *Eros* was devised for women to precipitate a clitoral erection. It works similarly to the vacuum pump used to create penile erections in men, although no band is placed around the clitoris once erection is obtained. The pump is not necessary for intercourse to occur; it was devised to enhance female excitement, but its effectiveness has not been studied rigorously.

Surgical Treatment

MALE PROSTHESES. Surgical treatment is infrequently advocated, but improved penile prosthetic devices are available for men with inadequate erectile response who are resistant to other treatment methods or who have medically caused deficiencies. For instance, it is used in some cases of Peyronie disease when more conservative treatments have been unsuccessful. There are two main types of prosthesis: a semirigid rod prosthesis that produces a permanent erection that can be positioned close to the body for concealment and an inflatable type that is implanted with its own reservoir and pump for inflation and deflation. The latter type is designed to mimic normal physiologic functioning. Placing a penile prosthesis in a man who has lost the ability to ejaculate or to have an orgasm as a result of medical causes will not restore those functions. Men with prosthetic devices have generally reported satisfaction with their subsequent sexual functioning, but their wives report much less satisfaction. Presurgical counseling is strongly recommended so that the couple has a realistic expectation of what the prosthesis can do for their sex lives. Postsurgical counseling may also be necessary to help the couple adapt to their rediscovered ability to have intercourse. They might experience a high level of anxiety if their sex life was inactive for a prolonged period before surgery. Prosthetic devices have been associated with severe adverse effects, including perforation, infection, urinary retention, and persistent pain.

Some surgeons are attempting revascularization of the penis as a direct approach to treating erectile dysfunction resulting from vascular disorders. Such surgical procedures are indicated in patients with corporal shunts that allow normally entrapped blood to leak from the corporal spaces, leading to inadequate erections (steal syndrome). Limited reports exist of prolonged success with the technique. Endarterectomy can be of benefit if aortoiliac occlusive disease is responsible for erectile dysfunction.

Another medical treatment being studied for erectile disorders is electrostimulation at the base of the penis. Initial reports indicate minimal physical discomfort in patients receiving this therapy. However, the response to treatment is inconsistent, and a problem exists in terms of maintaining erections. At the present time, the treatment seems to have no benefits.

FEMALE PROCEDURES. Surgical approaches to female dysfunctions include hymenectomy in the case of dyspareunia in an unconsummated marriage, vaginoplasty in multiparous women complaining of lessened vaginal sensations, or freeing clitoral adhesions in women with inhibited excitement. Such surgical treatments have not been carefully studied and should be considered cautiously.

Outcome. Demonstrating the effectiveness of the varieties of sex therapy is just as difficult as assessing the effectiveness of outpatient psychotherapy in treating other psychological problems. As with other disorders, the more severe the psychopathology associated with a problem of long duration, the more adverse the outcome is likely to be.

The most difficult treatment cases involve couples with severe marital discord. Cases involving problems of fear of intimacy, excessive dependency, or excessive hostility are also complex. Other challenges are posed by patients with impulse disorders, unresolved conflicts about sexual orientation, or fetishistic defenses. Patients with acquired dysfunctions have a better prognosis than do patients with lifelong dysfunctions. Disorders of desire or interest require longer, more intensive therapy than some other disorders, and their outcomes are very variable.

When behavioral approaches are used, empirical criteria that are supposed to predict outcomes are more easily isolated. Using these criteria, for instance, couples who regularly practice assigned exercises appear to have a much greater likelihood of successful outcome than do more resistant couples or those whose interaction involves sadomasochistic or depressive features or mechanisms of blame and projection. Flexibility of attitude is also a positive prognostic factor. Overall, younger couples tend to complete sex therapy more often than older couples. Couples whose interactional difficulties center on their sex problems, such as inhibition, frustration, or fear of performance failure, are also likely to respond well to therapy.

Methods that have proved effective singly or in combination include training in behavioral sexual skills, systematic desensitization, directive marital counseling, traditional psychodynamic approaches, group therapy, and pharmacotherapy. Although most clinicians prefer to treat a couple for sexual dysfunctions, treatment of individual persons has also been successful. The frequency of sessions is not a significant factor in treatment success. Thus, whether patients have intensive daily therapy over a period of 2 weeks, weekly therapy, or biweekly therapy appears to have little effect on the outcome of treatment. Also, the use of one therapist to treat a couple instead of a dual-sex cotherapy team is nearly as effective and certainly more practical. A multimodal regimen and an eclectic approach are frequently necessary to treat sexual disorders. Many couples who do not reach their ideal goal of sexual interaction, nonetheless, report considerable improvement and satisfaction in their sex lives and their general sense of well-being.

CULTURAL MORES AND SEXUALITY

Cultural mores affect the societal acceptance, legality, and self-perception of normal sexual behavior. An overview of the mores of Western civilization reveals a pendulum that oscillates between the liberal and the puritanical, between tolerance and restriction of behaviors, and between acceptance or repression of human sexuality. Mores are affected by knowledge, superstition, scientific advances, and religion, in addition to the outbreak of sexually transmitted infections and types of government. Proscription of sexual behavior goes back to the beginning of recorded time. The Book of the Dead in Egypt, 6,000 years ago, the Code of Hammurabi, and the Hittite Code, all had prohibitions against adultery.

Other sexual behaviors such as rape and sexual abuse of children have been condemned by law and public opinion across time.

In different eras, there are many examples of cultural swings in sexual attitudes. In the classical era, sensuality was characteristic of Greek life. There was a double standard, however. Women were not allowed the same sexual freedom permitted to men. Homosexuality was acceptable among men. By contrast, the Middle Ages was a time of sexual repression. Sexual behavior was restricted to marriage; contraception and abortion were condemned, and homosexuality was punishable by death. The Victorian era was a study in hypocrisy. Women and children were believed to have no sexual feelings, and the Victorians displayed a near obsession with masturbation inventing numerous outcomes such as blindness and insanity. At the same time, a sexual subculture existed, with prostitution and pornography prevalent. In the current era, sexual liberality has emerged. There are many contributing factors: the invention of the automobile, which allowed young adults to court without chaperones; the research of men like Kinsey bringing the discussion of sexual behavior into the public arena; the development of the birth control pill; the adjudication of Roe v. Wade, and societal movements including feminism and the civil rights movement that changed former societal attitudes.

Finally, it is apparent that a segment of society always challenges the prevalent cultural mores. If the attitudes are liberal, there are countervailing forces pushing for restrictive mores; if the prevailing attitudes are repressive, a segment in society will push for more sexual freedom. Above, I mentioned that changes regarding sexuality involved greater legal and societal acceptance of behaviors formerly considered abnormal. Currently, however, there are opponents of birth control, abortion, sex education in schools, sex outside of marriage, and the acceptance of transgendered persons. Support for those ideas is increased by the fear of sexually transmitted diseases, by cries that the family and society are disintegrating because of immoral sexual behavior, and when sexual freedom is used as a substitute for intimate relationships. Whatever the predominant values, different views challenge the status quo, exerting pressure on the pendulum of sexual mores to swing again in the opposite direction.

FURTHER READINGS

Bogaert AF. Asexuality: what it is and why it matters. *J Sex Res*. 2015;52(4):362.

Brassard A, Dupuy E, Gergeron S, Shaver PR. Attachment insecurities and women's sexual function and satisfaction: the mediating roles of sexual self-esteem, sexual anxiety, and sexual assertiveness. *J Sex Res*. 2015;52(1):110–119.

Busby DM, Leonhardt ND, Leavitt CE, Hanna-Walker V. Challenging the standard model of sexual response: evidence of a variable male sexual response cycle. *J Sex Res*. 2020;57(7):848–859.

Carvalho J, Veríssimo A, Nobre PJ. Psychological factors predicting the distress to female persistent genital arousal symptoms. *J Sex Marital Ther*. 2015;41(1):11–24.

Coady D, Fish N. *Healing Painful Sex*. Seal Press; 2011.

*Delgardo PL, McGauey CA, Moreno FA, Laukes C, Gelenberg AJ. Treatment strategies for depression and sexual dysfunction. *J Clin Psychiatry*. 1999;17:22.

Federman DD. Three facets of sexual differentiation. *N Engl J Med*. 2004;350:323–324.

Fischer N, Štulhofer A, Hald GM, Carvalheira A, Træen B. Sexual satisfaction in older heterosexual couples from four European countries: exploring the roles of actual and perceived discrepancy in sexual interest. *J Sex Res*. 2021;58(1):64–73.

Goldstein I, Lue TF, Padma-Nathan H, Rosen RC, Steers WD, Wicker PA. Oral sildenafil in the treatment of erectile dysfunction. Sildenafil Study Group. *N Engl J Med*. 1998;338:1397–1404.

Grubbs JB, Volk F, Exline JJ, Pargament KI. Internet pornography use: perceived addiction, psychological distress, and the validation of a brief measure. *J Sex Marital Ther*. 2015;41(1):83–106.

Handy AB, Freihart BK, Meston CM. The Relationship between subjective and physiologic arousal in women with and without arousal concerns. *J Sex Marital Ther*. 2020;46(5):447–459.

Hille JJ, Simmons MK, Sanders SA. Sex and the ace spectrum: definitions of sex, behavioral histories, and future interest for individuals who identify as asexual, graysexual, or demisexual. *J Sex Res*. 2020;57(7):813–823.

Krafft-Ebing R. *Psychopathia Sexual*. Matthes & Seitz; 1984.

*Laughman EO, Gagnon JH, Michael RT, Michaels S. *Sex in America*. University of Chicago Press; 1994.

Leitenberg H, Detzer MJ, Srebnik D. Gender differences in masturbation and the relation of masturbation experience in preadolescence and/or early adolescence to sexual behavior and sexual adjustment in young adulthood. *Arch Sex Behav*. 1993;22:87–98.

Levine SB. *Psychotherapeutic Approaches to Sexual Problems: An Essential Guide for Mental Health Professionals*. American Psychiatric Association Publishing; 2020.

Linet OI, Ogrinc FG. Efficacy and safety of intracavernosal alprostadil in men with erectile dysfunction. The Alprostadil Study Group. *N Engl J Med*. 1996;334:873–877.

*Masters WH, Johnson VE. *Human Sexual Response*. Little, Brown; 1970.

McCarthy B, Pierpaoli C. Sexual challenges with aging: integrating the GES approach in an elderly couple. *J Sex Marital Ther*. 2015;41(1):72–82.

Negash S, Van Ness Sheppard N, Lambert NM, Fincham FD. Trading later rewards for current pleasure: pornography consumption and delay discounting. *J Sex Res*. 2016;53(6):689–700.

Padma-Nathan H, Hellstrom WJ, Kaiser FE, et al. Treatment of men with erectile dysfunction with transurethral alprostadil. *N Engl J Med*. 1997;336:1–7.

Rhoden EL, Morgentaler A. Risks of testosterone-replacement therapy and recommendations for monitoring. *N Engl J Med*. 2004;350:482–492.

*Rosen RC, Lane RM, Menza M. Effects of SSRIs on sexual function: a critical review. *J Clin Psychopharmacol*. 1999;19:67–85.

*Sadock VA. The treatment of psychosexual dysfunctions: an overview. In: Grinspoon L, ed. Psychiatry 1982. *The American Psychiatric Association Annual Review*. American Psychiatric Press; 1982.

Sadock VA. Group psychotherapy of psychosexual dysfunctions. In: Kaplan HI, Sadock BJ, eds. *Comprehensive Group Psychotherapy*. Williams & Wilkins; 1983:286.

Sarin S, Amsel R, Binik YM. A streetcar named “Derousal”? A psychophysiological examination of the desire-arousal distinction in sexually functional and dysfunctional women. *J Sex Res*. 2016;53(6):711–729.

*Semans JH. Premature ejaculation: a new approach. *South Med J*. 1956;49:353–358.

Sternbach H. Age-associated testosterone decline in men: clinical issues for psychiatry. *Am J Psychiatry*. 1998;155:1310–1318.

Van Houdenlove E, Gijs L, T’Sjoen G, Enzlin P. Asexuality: a multidimensional approach. *J Sex Res*. 2015;52(6):669–678.

Vrangalova Z. Hooking up and psychological well-being in college students: short-term prospective links across different hookup definitions. *J Sex Res*. 2015;52(5):485–498.

Waldinger MD, Hengeveld WH, Zwinderman AH, Olivier B. Effect of SSRI antidepressants on ejaculation: a double-blind, randomized, placebo-controlled study with fluoxetine, fluvoxamine, paroxetine, and sertraline. *J Clin Psychopharmacol*. 1998;189:274–281.

21.2 Homosexuality, Gay and Lesbian Identities, and Homosexual Behavior

JACK DRESCHER, M.D., AND WILLIAM M. BYNE, M.D., PH.D.

Historically, homosexuality was considered to be a social problem whose origins were attributed to any number of problematic causes, including sin, degeneracy, poor parenting, social delinquency, and mental illness. Today, psychiatry and other mental health professions have adopted a model of sexual orientation that regards same-sex attractions, behaviors, and relationships as normal variations of human sexuality. This normal variation model informs positions taken by the American Psychiatric Association (APA), the American Psychological Association, and the American Psychoanalytic Association (APsaA) in support of equal rights for people who experience same-sex attractions, including marriage equality (same-sex marriage). In retrospect, these current positions were the inevitable outgrowth of the APA's 1973 conclusion that homosexuality is not a mental disorder and its decision to remove homosexuality per se from its *Diagnostic and Statistical Manual of Mental Disorders* (DSM). Not only did this action play a major role in the cultural normalization of same-sex relationships in the United States, but APA's diagnostic reclassification had an international impact as well. In 1990, the World Health Organization (WHO) removed the diagnosis of "homosexuality" *per se* from the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). Although "homosexuality" was replaced by F66 codes (other sexual disorders, including ego-dystonic sexual orientation), in 2019, WHO deleted these codes and any remaining diagnoses that pathologize homosexuality from ICD-11. ICD-11 officially comes into effect in January of 2022, at which time member nations may begin reporting morbidity and mortality statistics using the updated nosology.

In light of these changes in diagnostic systems, the focus of mental health interventions has shifted dramatically. Where psychiatry and medicine once focused on questions like “What causes homosexuality?” or “Can homosexuals be converted to heterosexuals?,” today’s mental health policy foci with respect to sexual orientation increasingly concern the mental health needs of sexual minority people, including those who identify as lesbian, gay, or bisexual (LGB). This shift from historical pathologic models of same-sex attractions and behavior to a normal variant model has been accompanied by dramatically increased social acceptance of openly LGB individuals in many, although not all, cultures. Acceptance has evolved rapidly in the West. For example, the 8th edition of this chapter published in 2004 noted that same-sex marriage was then only legal in Belgium and the Netherlands. At the time of this writing, in 2021, same-sex marriage is available in 29 countries: Argentina, Australia, Austria, Belgium, Brazil, Canada, Colombia, Costa Rica, Denmark, Ecuador, Finland, France, Germany, Iceland, Ireland, Luxembourg, Malta, Mexico, the Netherlands, New Zealand, Norway, Portugal, South Africa, Spain, Sweden, Taiwan, the United Kingdom, the United States, and Uruguay. In 2013, in *United States v. Windsor*, the U.S. Supreme Court overturned Section 3 of the discriminatory Defense of Marriage Act (DOMA) that prevented the federal government from recognizing same-sex marriages in states where such marriages were legal. The Court went a step further in 2015 in *Obergefell v. Hodges,* a ruling that overturned 13 remaining state constitutional bans and made same-sex marriage legal in all 50 states. Civil unions, although providing fewer legal protections than marriage, at the time of this writing, are available in Chile, Croatia, Cyprus, Czech Republic, Estonia, Greece, Hungary, Italy, Lichtenstein, Northern Ireland, Slovenia, and Switzerland.

Another indicator of increased acceptance of LGB people is decreased prohibitions against them serving openly in the militaries around the world. In 2011, the U.S. Congress repealed “Don’t Ask, Don’t Tell,” a policy that prevented LGB gay people from serving openly in the United States armed forces. Currently, LGB people can serve openly in the militaries of over 50 countries.

On the other hand, while there has been increased acceptance of same-sex attraction and behavior in an increasing number of countries, this is not the case throughout the world. In 2013, Russia passed a law criminalizing the distribution of “propaganda” to minors that support “nontraditional sexual relationships,” generally taken as meaning same-sex relationships and behavior. Other former Soviet states and countries in

Eastern Europe, Hungary, and Poland, for example, have passed similar laws. Engaging in homosexual behavior is still considered a crime in almost 70 countries, many of which are in Africa, and in some countries, homosexuality is punishable by imprisonment (i.e., Algeria, Bangladesh, Egypt, Malaysia) or death (i.e., Afghanistan, Iran, Iraq, Nigeria). Despite efforts by the United Nations and other international human rights groups to have such legislation repealed, some African countries such as Gambia, Nigeria, and Uganda recently passed laws increasing criminal penalties for expressions of or even support for homosexuality.

Despite the social gains noted above, relative to their heterosexual counterparts, LGB men and women may exhibit multiple mental health problems, including higher frequencies of psychiatric problems such as anxiety disorders, depression, suicidality, and alcohol and substance abuse. A preponderance of evidence suggests that the excess mental health burden is not due to homosexuality, per se, but instead derives from the chronic stresses that accompany sexual minorities' stigmatized social status and living in a society with deeply embedded heterosexist attitudes. Recent efforts have focused on informing mental health practice with LGB individuals based on empirical research. Based on a review of this evidence, psychologist John Panchankis identified the following as candidates to adapt for evidence-based practice with sexual minorities: identity concealment, rejection hypervigilance, and internalized stigma. Practices based on high-quality empirical research hold promise for optimizing the mental health care of LGB patients.

DEFINITIONS

Common terms used to describe human sexuality have a range of meanings, not only in common parlance but among medical professionals and authoritative texts as well. The definitions of these terms as they are used in this chapter are presented here. *Sexual orientation* refers to a person's erotic response tendencies or sexual attractions, be they directed toward individuals of the same-sex (*homosexual*), the other sex (*heterosexual*), both sexes (*bisexual*), or neither (asexual). The terms homosexual and heterosexual can lead to confusion when used to describe the sexual orientation of people whose gender identity differs from the sex they were assigned at birth (i.e., people often collectively subsumed under the umbrella term *transgender;* see chapter 21.4 on gender identity). Increasingly in the United States, sexual orientation is described relative to one's gender identity rather than the sex to which one was assigned at birth. For example, the sexual orientation of a transgender man (someone assigned female at birth who now identifies as a man) who is sexually attracted to women could be described as heterosexual or *straight.* Sexual orientation may be described unambiguously independent of one's gender identity as *androphilic* (attracted to men), *gynephilic* (attracted to women), bisexual (attracted to both men and women), or *analophilic* (attracted to neither). Importantly, in conversations with patients, clinicians should use the sexual orientation labels that the patient uses. However, it is helpful to clarify how the patient defines their self-label.

Sexual orientation and *sexual preference* are sometimes used interchangeably, but the former term has a wider currency in contemporary professional and popular usage. Sexual orientation is the term most experts use, while "sexual preference" is a less precise term as it implies a volitional or voluntary component. While the decision to engage in sexual behavior is volitional, the preponderance of evidence suggests that the direction(s) of one's erotic responsiveness is not. Collectively, individuals whose sexual orientation is other than heterosexual are increasingly referred to as *sexual minorities.* Opponents of LGB rights (historically referred to *as gay or gay and lesbian rights, etc.*) have claimed paraphilic desires and disorders (see Paraphilias section) are also "sexual orientations." However, this is neither a widely held view nor how the term, paraphilia, is usually interpreted in either science or in antidiscrimination legislation. Sexual orientation is often described as having three components—desire, behavior, and identity—which may or may not be congruent in an individual; however, while all three are related to sexual orientation, the object(s) of one's sexual desire (i.e., men, women, both, neither) correspond(s) most closely to the definition of sexual orientation above.

The Heterosexual-Homosexual Rating Scale, better known as the Kinsey Scale, is a 7-point continuous scale, with 0 representing exclusive heterosexuality, 6 representing exclusive homosexuality, and 3 representing equal amounts of both—has been widely used in the study of sexual orientation. The Kinsey scale has been criticized for its unidimensional conceptualization of sexual orientation, suggesting a diminished attraction toward one sex may be proportional to an increased attraction to the other sex. Because fantasies/desires are not always congruent with an individual’s behaviors, each can be assessed separately using the Kinsey scale. While sexual desire and sexual behavior can be described quantitatively (e.g., the proportion of times that sexual activities are with men vs. women, the relative strength of attraction to men vs. women, etc.), sexual identity tends to be described categorically (e.g., heterosexual, homosexual, bisexual, asexual, unsure) and does not lend itself to quantification with the Kinsey scale (see “Homosexuality and Nongay Identities,” below). Further, the Kinsey scale is a research instrument and has no clinical utility beyond illustrating the point that individuals’ sexual desires and behaviors are not necessarily exclusively either homo- or heterosexual. Finally, scores on the Kinsey scale are based on reported self-appraisals rather than objective measures.

Sexual identities sometimes referred to as *sexual orientation identities,* refer to how individuals view and label themselves based on their subjective experience of their sexuality and the meanings they attribute to their sexual desires and behaviors. In the case of *homosexual, lesbian, gay,* or *bisexual* identities, these usually involve some measure of self-acceptance of one’s homosexual desires and perhaps identification with a community of similar others by virtue of having the same pattern of sexual attractions. This community is often referred to as the LGB community. The LGB acronym is often extended to LGBT in recognition of the fact that sexual minorities and gender minorities (e.g., transgender/transsexual individuals) are subject to similar stigma. More recently, the acronym has been extended by adding “Q” to represent people who are questioning their sexual or gender identity or who identify as queer, and by adding “+” to acknowledge additional sexual or gender minority groups, including those who identify as asexual or intersex. The acquisition of an identity as a sexual minority person is often conceptualized as a developmental process that occurs over time. Models of the development of LGB identities usually portray a series of progressive stages involving tasks such as *coming out* (the process of recognizing that one experiences same-sex attractions and acknowledging that fact to oneself and others), involvement in LGB communities, perhaps establishing same-sex romantic relationships, and integrating one’s sexual minority identity into other aspects of the self.

Importantly, not all persons who experience homoerotic desire or participate in homosexual behavior follow such a linear path or develop an enduring LGB identity. In addition to sexual orientation, the terms *gay,* *lesbian, bisexual, etc.,* may be used to signify identities based on one's recognition of and identification with similar others on the basis of sexual orientation. Thus, some individuals may acknowledge their homosexual orientation without identifying as gay, lesbian, or bisexual. For example, some identify as nongay homosexuals. In other words, a *sexual orientation identity* (aka *sexual identity*) is not synonymous with *sexual orientation.* Years may elapse between one's first awareness of same-sex attractions and integration of that awareness into one's sexual orientation identity, that is, identifying as an LGB or other sexual minority person. Sometimes such integration does not occur.

For example, a woman or man with a homosexual or bisexual orientation may not accept having same-sex attractions or may believe that such feelings are transitory. Such an individual might marry a member of the other sex and maintain an ostensibly heterosexual lifestyle. Such persons may never lose their same-sex desires and, even if consciously aware of them, still may not identify as gay, lesbian, or bisexual. Such a person may have a history of same-sex sexual behaviors and may even form homosexual relationships while at the same time expressing disdain for the gay or lesbian "lifestyle" and the LGB community. Such an individual, if bothered by their same-sex attractions, may seek religious or professional help to rid themselves of them. Some such individuals refer to themselves as "ex-gays." Research on whether people can actually change their sexual orientation through willful effort (see "Sexual Orientation Conversion Efforts," below) is not persuasive, although the issue recurrently generates cultural controversy in some quarters. It is clear that individuals can project a public image as heterosexual individuals but not clear at all that one can change their sexual orientation by volition. Other individuals, some of whom refer to themselves as "nongay homosexuals," may acknowledge their homosexual orientation but resist acting upon their sexual urges and desires. Although one's sexual orientation is highly unlikely to change, sexual orientation identities are more flexible because they reflect how a person deals cognitively and psychologically with the awareness of their same-sex attractions. Consequently, a comprehensive clinical understanding of a patient's sexuality includes knowledge of all components of their sexual orientation: desire, behavior, and identity.

A *heterosexual identity* refers, in theory, to the subjective experience of those who are primarily aware of erotic attraction to persons of the other sex. As a social construct, heterosexual identity, has not been as well developed or defined as LGB identities. In today's society, unless one has ongoing contact with LGB individuals, a heterosexual identity may not be consciously experienced as anything other than what is expected or "normal." Where a heterosexual developmental outcome is assumed to be the norm, it is assumed that most people have a heterosexual identity unless they either disavow it or proclaim an alternative identity. Consequently, there is no parallel in the development of a heterosexual identity to either the experience of realizing that one is gay, lesbian, or bisexual or the process of revealing that identity to others. These processes are termed *coming out to oneself* and *coming out to others*, respectively. In childhood, most individuals assume that they will become heterosexual adults. Early resistance to accepting one's same-sex attractions and adopting a gay or lesbian identity is, therefore, believed to be a common and nonpathologic feature of the coming out process.

Homosexuality was first introduced as a medical term in the second half of the 19th century to describe erotic desire for persons of the same sex. *Heterosexuality* was coined later to describe erotic desire for persons of the other sex. A third category of sexual orientation, *bisexuality,* was used to describe an attraction to members of both sexes. The terms homosexuality, heterosexuality, and bisexuality are not applied consistently in theoretical, research, or popular discourse. Depending on the context, these terms have been used to refer to a wide range of differing constructs. These include the categories of sexual desire, gender role attributes, forms of sexual behavior, and personal and social identities. In the 19th century, bisexuality referred to a hypothetical bipotentiality of an organism to physiologically develop as either a male or female of its species. Upon discovery that the embryonic precursors of the external human genitalia did not differentiate as male or female until the 12th week of gestation, it was assumed that human beings carried a bisexual potential for psychic development in them as well. Sigmund Freud, among others, working prior to the elucidation of the biologic sexual differentiation of the mammalian brain, took this paradigm one step further and hypothesized that human beings are psychologically bisexual. Many theorists work from an underlying assumption that sexual orientation reflects some unitary quality or essence, rooted perhaps in biology, that transcends cultures and historical periods. Belief in this common essence has been termed *essentialism,* and those who hold this belief are referred to as *essentialists.* Essentialist assumptions have sometimes dismissed or minimized the personal and social meanings of sexual desire and relationships that contribute to sexual identities. The contrasting view that these phenomena are culturally and historically dependent is referred to as *constructionism.* Constructionist assumptions tend to dismiss the role of biology. Although essentialism and constructionism are often viewed as competing schools of thought, that is not necessarily the case. Essentialist theory, in general, deals with sexual orientations, while constructivist theory usually deals with sexual identities. It is certainly possible that individuals throughout history have had sexual orientations as described above, even in cultures where the concept of sexual orientation did not exist and in which sexual orientation had less effect on the individual's identity, or social role and labeling.

In this chapter, *homosexuality* refers to the characteristic of having erotic desire directed toward someone of one's own sex and gender identity and usually of the same birth-assigned sex. Its presence in an individual does not necessarily imply or dictate the concurrent presence or absence of any other characteristics such as heterosexual attractions or gender-typical/atypical interests and behaviors, including mannerisms. It should be further noted that respectful person-first language does not employ sexual orientation descriptors as nouns. Rather than referring to "heterosexuals, bisexuals, and homosexuals," person-first language employs these terms as adjectives (e.g., heterosexual individual, homosexual individual, etc.). When used as an adjective, *homosexual* can also refer to sexual ideation or activity involving members of the same-sex or gender identity.

Similarly*, bisexuality* can refer to an erotic propensity, an individual identity, or a pattern of sexual behaviors. Bisexuality may occur *sequentially* (expressed as an attraction to or relationship with both sexes, albeit at different times in one's life) or *concurrently* (attraction to or relationships with individuals of both sexes at the same time). Some theorize that bisexuality signifies a form of identity diffusion or a transitional state leading toward the development of a more stable heterosexual or gay or lesbian identity. However, some individuals maintain a stable and persistent bisexual identity and pattern of sexual behaviors and/or attractions across time. Limited research suggests that bisexual women are often perceived by nonbisexual people as "over-sexed" while bisexual men are perceived as being unable to accept their homosexuality. Thus, bisexual individuals may experience more social stigma than gay people and may be reluctant to declare themselves. In this chapter, references to bisexuality and bisexual persons are sometimes included because some parallels exist in the experiences of homosexual and bisexual individuals. A fuller discussion of the topic of bisexuality, however, is beyond the scope of this chapter.

In the latter part of the 20th century, the terms *gay, lesbian,* and *bisexual* emerged in popular usage as a way to refer to men and women whose sexual identity, to some degree, consciously acknowledged their sexual attractions to their own or both genders. *Gay* is sometimes also used as a politically inclusive term for both men and women whether they experience homosexual or bisexual attractions. In contrast to the medical term, *homosexual,* which focuses on one aspect of a person's identity, their sexual attractions, as if it were the sum of the person's entire identity, the terms *lesbian* and *gay* have the advantage of not necessarily connoting sexual acts.

In recent years, there has been an increasing proliferation of additional gender categories and simultaneous rejection of the gender binary, a system that recognizes only two (i.e., male and female) genders. Consequently, the term *bisexual,* which implies attraction to only two genders (i.e., male and female), has been joined by a more recent adjective, *pansexual*. This term was coined to indicate a pattern of sexual attraction to other people that is not limited by their sex, sex assigned at birth, current gender identity, or manner of gender expression. Some who do not experience sexual attraction to others may employ the label *asexual* to indicate they are not sexually attracted to others, while some may use this label to indicate that they do not engage in sexual activity with others. Others may describe themselves as *demisexual to* indicate they experience sexual attraction only in the context of an emotional relationship. *Questioning* is sometimes used as an identity label by individuals who are not sure of their sexual orientation. Individuals who are struggling to find the "right" label for their sexual orientation sometimes benefit from learning about *questioning* as an identity label. *Queer* is the preferred identity label of some who are members of either the sexual or gender minority community, or both. While *queer* was once universally considered a slur, it was reclaimed as an empowering label by some sexual and gender minorities through the academic discipline of queer theory and street-level activism. Nevertheless, it continues to be considered pejorative by many members of the gay community—again, it is always helpful to ask.

In much of the early psychologic and psychoanalytic literature, *homosexual* was used as a noun to designate a person with same-sex desire or behavior; however, the use of the term as a noun has decreased in both scientific and popular discourse. For example, recognizing the wide range of sexual identities in those populations at risk of transmitting human immunodeficiency virus (HIV) infection through sexual behavior, the literature on HIV and acquired immunodeficiency syndrome (AIDS) refers to *men who have sex with men* (MSM). MSM do not necessarily think of themselves as gay or bisexual. In some cultures, only a receptive sexual partner is considered gay or homosexual, while an insertive man who also has sex with men and women is not culturally defined as either homosexual or bisexual. Thus, efforts to categorize people on the basis of particular behaviors often come into conflict with both cultural and subjective experiences of the meaning of homosexuality.

Many historical discussions of homosexuality erroneously conflated sexual orientation with *gender identity* and linked homosexual or bisexual orientations with abnormal gender identity. Such theories presumed the cause of homosexual attractions and behaviors to reflect confusion about one's own gender identity, which was then thought to subsequently cause confusion about the sex of individuals to whom one was attracted. However, in contrast to *gender dysphoria,* homosexuality usually entails no desire to transition or to affirm one's gender as being different from that assigned at birth but instead signifies that one's erotic interests are by and large confined to members of one's own sex, gender identity, and gender role. Transgender phenomena are addressed comprehensively in the Gender Identity chapter.

Transvestitism refers to sexual urges and fantasies involving crossdressing. When the behavior is distressing to a patient and brought to the attention of a psychiatrist, it may be diagnosed as *transvestic disorder,* which DSM-5 classifies as a *paraphilic disorder* (refer to Paraphilias section). Crossdressing is strongly linked in the popular imagination with homosexuality. However, in DSM-IV-TR, the diagnosis of *transvestic fetishism* could only be made in heterosexual men as most fetishistic crossdressers are heterosexual men who crossdress in private. These men are often married to women who may or may not know about their crossdressing interests. DSM-5 removed the restriction to heterosexual men, opening the diagnosis to women or gay men who crossdress out of sexual interest. While the change could increase the number of people diagnosed with transvestic disorder, the diagnosis still requires that individuals experience significant distress or impairment because of their behavior and urges. In any event, most gay men and lesbian women do not crossdress. However, there are social venues within gay and lesbian communities that allow for public crossdressing. These may include either socially performative events (gay pride, Carnival, Halloween, or Mardi Gras parades) or crossdressing as a form of paid entertainment (drag shows).

THEORIES

Although heterosexuality is often regarded as nature's default setting—and therefore not requiring any explanation—theories of homosexuality's "causes" abound. One exception can be found in Plato's *Symposium,* which tells a creation myth explaining not only male and female homosexuality but heterosexuality as well. According to this myth, humanity originally consisted of three sexes. Its members were joined in pairs consisting of two men, two women, or a man and a woman (androgyne). To diminish their power and teach them to fear the gods, Zeus divided each pair in half. Thereafter, the split-off aspects were in constant search for their missing halves. Thus, men who were originally one-half of the androgyne seek women and vice versa. Women who seek women and men who seek men were instead part of the original male and female beings.

Although male homosexual behavior played an important and ritualized role in ancient Babylonia and classical Greece, homosexual acts were prohibited in other cultures. In the Abrahamic religions (Judaism, Christianity, and Islam), prohibitions against homosexual behavior as sinful can be found in religious texts. In the Old and New Testaments of the Bible, these include Genesis 19, Leviticus 18:7, 22, Leviticus 20:13, Judges 19, I Kings 22:46, II Kings 23:7, Romans 1:27, I Corinthians 6:9, and I Timothy 1:9–10. Historian John Boswell argued that contemporary intolerance of homosexuality was not an essential factor of Christianity itself but only became a prevailing church attitude in the second millennium ACE. Although Michel Foucault argued that the idea of identifying *a type or a category of person* based on either homosexual behavior or identity is a relatively new cultural phenomenon that coincides with the medicalization of homosexuality during the 19th century, Boswell believed that a category of homosexual individuals, although not entirely equivalent to a modern gay- or lesbian-identified individual, has always existed across time and cultures. Plato's androgyne myth, for example, accounts for the origins of such categories. Foucault and Boswell may each have been right in that recognition of a category of individuals defined by their sexual orientation (as opposed to engaging in homosexual behaviors) has been discontinuously present across history, and that at any given point in history it is recognized by some cultures and religions but not by others. Today, the study of the forms and meanings of homosexual behavior in different cultures and historical periods is a rich field primarily engaged in by scholars of *lesbian and gay studies* and *queer theory.* Nevertheless, the forms of socialization and identity development associated with homosexuality in contemporary Western society are generally considered to be unique. The history of the development of present forms of homosexual identity under the Western influence is the focus of the following review of theoretical perspectives about homosexuality.

History

The contemporary debates about the meaning of homosexuality can be traced to the middle of the 19th century, when European sexologists, including several prominent psychiatrists, began the study of homosexuality from a scientific and medical perspective. Through much of the 20th century, this intellectual enterprise was dominated by a predominance of biologic and psychological, particularly psychoanalytic, theories of homosexuality. Psychiatric theories of the first half of the 20th century, with a few exceptions, regarded homosexuality as either a form of psychopathology or as a developmental arrest. In the latter half of the 20th century, however, theories that characterized homosexuality as a normal variant of human sexuality began to gain ascendance. This view emerged from a wide range of social and scientific perspectives outside medicine and psychiatry and eventually supplanted earlier theories of homosexual pathology and immaturity. This modern *normal variant* view began gaining influence around the time of the publication of the two Kinsey volumes on male and female sexuality—in 1948 and 1953, respectively—and culminated in the APA's removal of homosexuality from its list of mental disorders, the DSM, in 1973. This action later prompted WHO to remove homosexuality *per se* from ICD-10 in 1990 and to later remove all references to sexual orientation from ICD-11 in 2019. Consequently, the normal variant model is accepted by today's mental health mainstream.

Medicalization of Homosexuality: From Morality to Pathology

19TH- AND 20TH-CENTURY BIOLOGIC VIEWS. The evolution of today's conceptualizations of homosexuality is often tracked from the mid-19th century. In 1864, in response to Bavaria's Paragraph 175, which criminalized male (but not female) homosexual acts, German lawyer and activist Karl Ulrichs began publishing treatises that criticized the law. Ulrichs theorized that some women were born with a man's spirit trapped in their bodies (*Urningins*), and some men were born with a woman's spirit trapped in theirs (*Urnings*). Although he said these people constituted a "third sex," the concept drew on a binary approach to gender that presumed an attraction to men is a feminine attribute and that an attraction to women is a masculine one.

In 1869, a Hungarian writer, Károli Mária Kertbeny, who also wrote critically about laws criminalizing same-sex behavior (*sodomy laws*), first coined the terms *homosexual* and *homosexuality.* The influential psychiatrist, Richard von Krafft-Ebing, later adopted and popularized the terms in his 1886 classic *Psychopathia Sexualis.*

Krafft-Ebing and the psychiatrist Carl Westphal adopted many of Ulrich's ideas; however, the former two deemed unnatural (*theories of pathology*) the very same traits that the latter had described as normal. Magnus Hirschfeld, a German physician and sexual emancipator was an openly homosexual psychiatrist who also believed there was a biologic basis for homosexuality. Although he considered homosexuality to be one of several "intermediate sexual types," presumed to have a number of different biologic etiologies, Hirschfeld, the standard-bearer of *third sex theory in his time,* nevertheless asserted that the condition was natural (*theory of normal variation*). Similarly, the English sex reformer Havelock Ellis thought of homosexuality as a biologic variation. And although he attributed it to "imperfect sexual differentiation," he nevertheless considered homosexuality to be nonpathologic since it did not cause any illness in the affected individual. Other biologic bases for homosexuality began to be examined during this period. These included studies of anatomical structures in the brains and nerve structures in the rectums of homosexual men. Endocrinologic hypotheses led to experiments in which the supposedly atrophied or abnormal testicles of homosexual men were removed and replaced with what were believed to be the normal testicles of heterosexual men. However, these experiments did not make homosexual men heterosexual.

Early biologic theories, with their emphasis on heredity, endocrinologic influences, and anatomical or structural differences between homosexual and heterosexual persons, heralded many of the recent efforts to find a biologic basis for homosexuality, some of which were interpreted as suggesting that homosexuality is biologically determined and, therefore, not derivative of childhood experiences or indicative of psychopathology. Biologic theories, however, also provided a rationalization for eugenic efforts in Germany to eradicate homosexual persons. As recently as the 1970s, East German surgeons unsuccessfully tried to eliminate homosexuality in men by operating on the hypothalamus where the “cause” of homosexuality was presumed to lie. Throughout the 20th century, interest in the origins of sexual orientation focused almost exclusively on men, a trend less evident today.

PSYCHOANALYTIC APPROACHES. Freud did not view homosexuality as a sign of illness, by which he meant a symptom arising from psychic conflict. Instead, he saw homosexuality as the unconflicted expression of an innate instinct. Although his conjectures were based on clinical observations rather than objective studies, he noted that homosexuality could occur in individuals who had no other signs of deviation and no impairment in their functioning. Freud believed in *constitutional bisexuality* that every individual had a certain component of *masculine* (active) as well as *feminine* (passive) *tendencies.* A homosexual tendency did not necessarily imply a homosexual behavior. Although he viewed bisexual tendencies as universal, Freud believed some people were constitutionally endowed with more of one tendency than the other. He believed life experiences, particularly traumatic ones (environmental factors), could have an impact on the development and expression of one’s innate instincts (biologic factors). Under normal and nontraumatic circumstances, the component instincts that determine the sex of one’s final object choice should be consistent with one’s anatomical sex. That is to say, an anatomic male should ideally express the masculine component instinct and obtain sexual satisfaction from women. However, Freud also conjectured that even adults who are ostensibly heterosexual individuals retain the homosexual component, albeit in sublimated form. For Freud, adult, nonsublimated homosexuality represented an arrest in development on the road from the instinctual bisexuality of childhood to mature heterosexuality. His *theory of psychic immaturity,* that is, a developmental arrest to account for adult homosexuality, comprised an alternative category to religious sin, medical theories of pathology, or normal variant third-sex theories. By maintaining that some degree homosexuality, at least developmentally, could be a normal part of everyone's experience, Freud's conjectures allowed for the possibility that the adult homosexual person might sufficiently mature and, if sufficiently motivated, become heterosexual. However, late in his life, Freud doubted the possibility of changing a homosexual orientation to a heterosexual one.

After Freud's death, Sandor Rado's theories came to hold greater sway on psychoanalytic theorizing about homosexuality. In the 1940s, Rado argued that the theory of bisexuality was based on a faulty analogy with anatomical bisexuality. That is, underlying Freud's theory was a 19th-century belief in embryonic hermaphroditism, the hypothesis that the potential to become an anatomical man or a woman was present in every embryo. Rado claimed this theory had been disproved, which led to his claim that heterosexuality is the only nonpathologic outcome of human sexual development. He viewed homosexuality as an illness, specifically a phobic avoidance of the other sex caused by parental prohibitions against childhood sexuality. Almost all of the mid-20th century psychoanalytic theorists who pathologized homosexuality followed Rado's theory in one form or another. The psychoanalytic community's shift from Freud's model of immaturity (homosexuality as a normal developmental step toward adult heterosexuality) to Rado's model of pathology (homosexuality as a sign of development gone awry) led some analysts to claim they could "cure" homosexuality.

The work of psychoanalyst Irving Bieber and his associates was particularly influential in its portrayal of a pathogenic family type—a detached and rejecting father and a close-binding and domineering mother—that presumably led to homosexuality in the 106 adult homosexual men they studied. This theory of familial etiology, however, has not found support in the research of larger, nonpatient gay male populations. Although about 70% of their subjects dropped out during treatment, Bieber and his colleagues claimed a 27% cure rate among 29 of the 32 men who remained in the study. The Bieber study was criticized for its methodology

—it was not peer-reviewed—and the authors were unable to provide any long-term follow-up data on their subjects or produce any patients to support their claims of change. Finally, while some discussion of the etiology of female homosexuality existed in the early psychoanalytic literature, the primary emphasis in psychoanalysis, as in the biologic sciences, was on male homosexuality; often, presentations of homosexuality in women were simply treated as gender-reversed images of those for male homosexuality.

EARLY THEORETICAL ASSUMPTIONS. Nineteenth and early 20th-century theories of homosexuality, be they biologic (e.g., neurologic or endocrinologic), developmental, or psychoanalytic, largely relied on a dichotomous model of categories: Male/female, masculine/feminine, and heterosexual/homosexual. Whether the theorist believed homosexuality to be a normal variant, a form of pathology, or immaturity, the theory usually relied on the assumption that the wide range of human sexuality could be reduced to the two-component parts of male and female and that some intrinsic quality of one gender had made its way into a person of another gender. In addition, scientific and medical studies served to remove the responsibility of defining homosexuality from the realms of morality and religion and secured that task within science and medicine. In doing so, science and medicine created a category of person, “the homosexual,” in contrast to earlier moral and religious beliefs that homosexuality was a type of sinful sexual behavior that could be engaged in by anyone, rather than an erotic response pattern directed toward members of one’s sex. Although science and medicine perpetuated the social stigma of homosexuality, taking it out of the realm of sin and immorality into the realm of pathology and immaturity, these developments would also eventually set the stage for the transition toward normalizing homosexuality that began to take shape in the middle of the 20th century.

Normalization of Homosexuality: From Diagnosis to Normality

INFLUENCE OF RESEARCH FINDINGS FROM SEXOLOGY, CROSS-CULTURAL, ANIMAL, AND PSYCHOLOGICAL STUDIES. A major factor in the shift away from pathologizing theories about homosexuality was the work of sex researchers. Unlike psychiatrists and psychoanalysts, who based their theories on clinical work with people seeking treatment in their offices or on evaluations of prison populations, sex researchers sampled larger, nonpatient populations. Thus, the publication of two studies on male and female sexuality by Alfred Kinsey and his colleagues, in 1948 and 1953, respectively, proved to be a scientific and cultural watershed. Hundreds of thousands of copies were sold of the two reports, and were translated into many languages. Although normal variant theories of homosexuality had been around for almost a century, at the time of its publication, Kinsey's work signaled the beginning of a cultural shift away from viewing homosexuality as pathology toward viewing it as a normal variant of sexual desire and behavior. Despite his studies' methodologic flaws, Kinsey's finding of a widespread presence of same-sex feelings and behaviors among his samples of several thousand American men and women changed the general impression that homosexuality was a rare and pathologic phenomenon. His homosexuality/heterosexuality scale, which conceptualized sexual orientation on a continuum, expanded on the binary model used by earlier theorists. As Kinsey put it in his 1948 report, "The world is not to be divided into sheep and goats. Not all things are black, nor all things white. It is a fundamental of taxonomy that nature rarely deals with discrete categories. Only the human mind invents categories and tries to force facts into separated pigeon-holes. The living world is a continuum in each and every one of its aspects."

Cleland Ford and Frank Beach's work, published in 1951, found homosexual behavior to be common across cultures and documented its existence in almost all nonhuman primate species. The interpretation of these findings supported the notion that homosexuality was both natural (as opposed to a traditional view that it was caused by civilized decadence) and widespread. In 1957, psychologist Evelyn Hooker published a ground-breaking study comparing projective test results from 30 nonpatient homosexual men with those of 30 nonpatient heterosexual men. Her study found that experienced judges, unaware of whose test results they were interpreting, could not distinguish between the two groups. Again, the medical and psychiatric literature of that time, which regarded homosexuality as pathologic, was mostly based on evaluations of individuals troubled about their sexuality. Hooker's study of *nonpatient subjects* was a serious challenge to the then-prevailing psychiatric and psychoanalytic belief that homosexuality was always associated with mental illness or serious psychopathology. One of the significant impacts of Hooker's finding was to draw attention to the a priori assumption of many clinical papers that homosexuality was itself pathologic. In the

1960s, psychiatrists like Judd Marmor—a psychoanalyst and future president of the APA—Robert Stoller (also a psychoanalyst) and Richard Green gave serious consideration to the nonpsychiatric perspectives offered in these newer research findings. Another psychiatric/psychoanalytic dissenter was Thomas Szasz, who criticized mental health professionals in general for labeling many unconventional behaviors as signs of illness. Rather than practicing medicine, Szasz accused his fellow psychiatrists of using diagnoses as a way to increase their own power and influence.

DECLASSIFICATION BY THE AMERICAN PSYCHIATRIC ASSOCIATION. Homosexuality had been officially classified as a mental disorder in APA's first edition of the DSM (DSM-I) in 1952. There it was designated as a type of "sociopathic personality disturbance." Classifying homosexuality as a mental illness was not controversial at the time of the DSM-I's publication; societal attitudes of that time coincided with prevailing medical views that homosexuality was an aberration. In 1968, the revised DSM-II still listed homosexuality as a "sexual deviation," but sexual deviations were no longer categorized as a sociopathic personality disturbance.

The DSM's 1968 revision (DSM-II) coincided with a time of great social foment in the United States. In addition to anti-Vietnam War activity, Black civil rights protests, and emerging feminist voices, gay rights activism also gained public notice. Gay activism coalesced around the 1969 Stonewall riots in New York City. Efforts by police to raid and shut down a gay bar (The Stonewall), a routine police activity of the time, led to 3 days of rioting in Greenwich Village. Having fought off police and government successfully, the gay rights movement was prepared to challenge psychiatric authority as well. This led to increasingly vocal opposition to psychiatry's classification of homosexuality as a mental disorder. Before that time, many homophile organizations, both in the United States and abroad, had accepted the medical view of homosexuality as a mental disorder. Accepting the view of homosexuality as a disease meant treating it as a disability, a category intended to replace moral and religious judgments with scientific, objective, and humane attitudes. However, in most instances, medical and psychiatric portrayals of homosexuality were as problematic as the religious views they supplanted. For example, a homosexual individual could be denied the right to immigrate to the United States on the grounds that the person suffered from a mental disorder. Gay men and lesbian women were discharged from the military as medically unfit. To gay and lesbian activists of that period, the psychiatric designation of homosexuality as a mental disorder had exacerbated antihomosexual societal prejudices rather than ameliorated them. These activists found a receptive audience among an increasing number of psychiatrists familiar with the nonpsychiatric research findings showing homosexuality occurred in large numbers of people, in persons who demonstrated normal psychological adjustment, and across cultures.

Gay activists chose to confront APA about its position on homosexuality —they disrupted APA's annual meetings in 1970 and 1971. In response to these protests, APA's leadership took several steps. First, they permitted two educational panels at the APA's 1971 and 1972 meetings. The 1971 panel, titled "Lifestyles of Non-Patient Homosexuals," featured nonpatient gay activists explaining to their psychiatric audience the stigma caused by the "homosexuality" diagnosis. At the 1972 meeting, the activists were on another panel, "Psychiatry: Friend or Foe to Homosexuals: A Dialogue." The latter panel presented the perspectives of the activists, Judd Marmor, and psychiatrist, John Fryer, M.D., who appeared as "Dr. H Anonymous," disguising his true identity from the audience with a rubber mask, a fright wig, and an oversized tuxedo. He spoke of the discrimination gay psychiatrists faced in their own profession, telling his psychiatric audience:

> *As psychiatrists who are homosexual, we must know our place and what we must do to be successful. If our goal is academic appointment, a level of earning capacity equal to our fellows, or admission to a psychoanalytic institute, we must make certain that no one in a position of power is aware of our sexual orientation or gender identity. Much like the black man with the light skin who chooses to live as a white man, we cannot be seen with our real friends—our real homosexual family—lest our secret be known and our dooms sealed.* (John Fryer Papers (Collection 3465), Historical Society of Pennsylvania. Finding aid. Accessed May 3, 2023. http://digitalhistory.hsp.org/pafrm/doc/speech-dr-henry-anonymous-john-fryer-american-psychiatric-association-125th-annual-meeting.)

The second step taken by APA, in addition to these public panels, was to embark upon an internal process of studying the scientific question of whether homosexuality should be considered a psychiatric disorder. APA's Nomenclature Committee was the scientific body charged with addressing this issue. The Committee members reviewed the psychiatric and psychoanalytic literature on the subject as well as the sexology literature. The latter, a subject not usually taught in psychiatric training programs at that time, was unfamiliar to most practicing psychiatrists. Following their study of the issue, which included interviews with psychiatrists who favored and objected to making any changes, the Nomenclature

Committee recommended removing homosexuality per se from the DSM. The proposal was approved by the APA's Council on Research and Development, its Reference Committee, and by the Assembly of District Branches (now known as the APA Assembly) before being accepted by APA's Board of Trustees in December 1973. This marked the end of the classification of homosexuality per se as an illness. Other major mental health professional organizations, including the American Psychological Association and the National Association of Social Workers, soon endorsed the APA action. The decision to declassify homosexuality was accompanied by passage and issuing of a position statement by the APA that supported the protection of the civil rights of homosexual persons and that read:

> *Whereas homosexuality in and of itself implies no impairment in judgment, stability, reliability, or vocational capabilities, therefore, be it resolved that the American Psychiatric Association deplores all public and private discrimination against homosexuals in such areas as employment, housing, public accommodations, and licensing, and declares that no burden of proof of such judgment, capacity, or reliability shall be placed on homosexuals greater than that imposed on any other persons. Further, the APA supports and urges the enactment of civil rights legislation at local, state, and federal levels that would insure homosexual citizens the same protections now guaranteed to others. Further, the APA supports and urges the repeal of all legislation making criminal offenses of sexual acts performed by consenting adults in private.* (Reproduced with permission from American Psychiatric Association. *American Psychiatric Association Position Statement on Civil Rights of Homosexual Persons*. American Psychiatric Association; 1973.)

Some psychiatrists, primarily psychoanalysts, who continued to adhere to pathologizing views of homosexuality, criticized the Board of Trustees and challenged their action by petitioning the organization to hold a referendum in which the entire APA membership could vote. The Board of Trustee's decision to remove homosexuality from the DSM was upheld by a 58% majority of voting APA members.

The original scientific and clinical rationale for removing homosexuality from the DSM was based, in part, on a new formulation of the definition of a mental disorder by the Nomenclature Sub-Committee's Chair, Robert Spitzer M.D. In order to be classified as a disorder, a condition had to meet the criteria of either an experience of subjective distress or of a generalized, objective impairment in social effectiveness or functioning resulting from the condition itself. Neither of these criteria applied to gay and lesbian individuals who were satisfied with their sexual orientation and who did not demonstrate impaired functioning. In recognition of the opposition to the diagnostic change, however, the APA did not fully embrace a normal variant model of homosexuality; instead, it made a compromise. The diagnosis of *homosexuality* in the original text of DSM-II was replaced by *sexual orientation disturbance* (SOD) in its seventh printing in 1974. As a result, individuals comfortable with their homosexuality were no longer considered mentally ill. According to SOD criteria, only those who were “bothered by,” “in conflict with,” or “wished to change” their homosexuality had a mental disorder.

The diagnostic category of SOD had two significant conceptual problems. First, the diagnosis could also apply to heterosexuals, a solution to APA’s internal debate that did not concur with clinical reality. There were no reported cases of unhappy heterosexual individuals seeking psychiatric treatment to become homosexual. This problem of diagnostic over-inclusiveness was resolved in the 1980s DSM-III, where SOD was replaced by *ego-dystonic homosexuality* (EDH). The name change, however, did not resolve a thornier conceptual issue, which was that of making patients’ subjective experiences of their own homosexuality *the determining factor* in labeling their homosexual feelings as an illness. To rely on patient subjectivity as the major determining factor was now incongruous with the new evidence-based approach that psychiatry had embraced with DSM-III. This ultimately led, in 1987, to EDH being removed from the revised DSM-III-R. In 1990, the WHO followed suit and removed homosexuality from the 10th revision of the ICD (ICD-10), replacing it with a diagnosis that resembled SOD: ego-dystonic sexual orientation. This diagnosis was removed from ICD-11 in 2019.

For two decades after the 1973 APA decision, the meetings and publications of the APsaA were among the few American mental health forums where pathologic views of homosexuality continued to be expressed. Eventually, however, normal variant views of homosexuality prevailed in the APsaA as well. In 1991, the organization issued a nondiscrimination statement to allow training of gay and lesbian psychoanalytic candidates and the promotion of qualified gay and lesbian individuals to the position of training and supervising analyst. In 1997, APsaA became the first major mental health organization to support marriage equality. In 2000, APsaA took a position opposing sexual conversion therapy. However, psychoanalytic views that pathologized homosexuality nevertheless could still be found in the international psychoanalytic community. Then, in 2002, APsaA's influence in the International Psychoanalytical Association (IPA) led the latter to issue a nondiscrimination statement allowing for the selection of openly gay and lesbian candidates and the promotion of gay and lesbian faculty. In 2019, coinciding with the 50th anniversary of the Stonewall uprising, APsaA President Lee Jaffe issued a public apology to the lesbian, gay, bisexual, and transgender (LGBT) community for the organization's role in perpetuating stigma and discrimination.

In 2019, the WHO Assembly approved a recommendation of its Working Group on Sexual Health and Sexual Disorders that recommended the removal of the F66 diagnoses, including ego-dystonic sexual orientation, from ICD-11. The Working Group, reviewing the evidence published since 1990, found no evidence that these diagnoses were clinically useful as they did not contribute to health service delivery or treatment selection nor provide essential information for public health surveillance.

All of these actions represent definitive acknowledgment by the psychiatric, psychoanalytic, medical, and mental health professions that homosexuality, in itself, is not a mental disorder. Not only does this change in the mental health professions' view of homosexuality represent a wider acceptance of an affirming and more scientific view of homosexuality, but it also symbolizes a dramatic conceptual shift in the cultural meaning and significance of homosexual behavior, particularly within, but not entirely limited to Western culture.

PREVALENCE

One must use caution when interpreting studies that address the prevalence of homosexuality in a population. First, one must consider the survey methods: What population did the survey aim to assess? Was a random sample of this population ascertained, and was the sample large enough to provide statistically meaningful results? One must clearly identify which aspects of sexual orientation were assessed (e.g., sexual desire for men, women, or both or neither; sexual activity with men women, both or neither; sexual identity [heterosexual/straight, homosexual/gay/lesbian, bisexual, pansexual, asexual, other]). Ideally, surveys should also assess the gender of participants by querying both their sex assigned at birth (e.g., what sex or gender was on your initial birth certificate [male, female, other–specify]) as well as how they identify their gender currently (male, female, transgender man, transgender woman, other, etc.). Additionally, one must determine whether surveys reflect the proportion of respondents who have ever experienced same-sex desire, who have ever engaged in same-sex behavior, or who have identified themselves as homosexual at any time in their lives, within a restricted period of time, at a particular stage of their lives, or in a stable and ongoing pattern.

Beyond the issues of survey design, the stigma associated with homosexuality may interfere with full disclosure. The impact of stigma on survey results may vary depending on the survey method. Although face-to-face interviews are generally considered preferable, logistical issues may limit the sample size or the geographic region(s) surveyed. Surveys based on random telephone numbers may achieve large representative nationwide samples but may result in underreporting of homosexuality if the subjects do not have privacy at the time of their telephone interview.

Regardless of the survey method employed, persons who are not accepting of or open about their sexual orientation are less likely to respond fully or accurately to questions about their sexual behavior or identity. Thus, even large-scale random surveys are more likely to result in underestimates rather than overestimates of homosexual phenomena. Here, the history of assessing the prevalence of homosexuality, primarily in the United States, is reviewed chronologically.

The Kinsey Survey of Male Homosexuality

The 1948 Kinsey study of male sexuality, based on a sample of over 5,000 men between the ages of 16 and 55 years, concluded that 4% of the men were exclusively homosexual throughout their lives after adolescence, 50% were exclusively heterosexual, and 46% were somewhere in between. The report also concluded 10% of males are “more or less exclusively homosexual for at least 3 years between the ages of 16 and 55.” The latter conclusion eventually became the basis of the common misperception that 10% of the population is homosexual.

The Kinsey study has subsequently been criticized for its sampling errors, including an overrepresentation of men with histories of incarceration, which are believed to have led to inflated estimates of the extent of homosexual and bisexual behavior. However, Paul Gebhard, Kinsey’s successor as director of the Kinsey Institute for Sex Research, dedicated years to culling purported contaminants from Kinsey’s data and concluded that his figures were off by no more than a few percentage points. In addition, Kinsey's numbers are not that different from recent large-scale surveys employing modern sampling methods. While absent in early studies, recent studies have included women.

Later Surveys

A large-scale study of sexuality in the United States published in 1994 by Edward Laumann and colleagues was long regarded as providing the most reliable estimates of the prevalence of homosexuality in the United States. This study inquired about individual desire, behavior, and identity and obtained a carefully stratified representative sampling of the general population. This study, based on a total sample of 3,159 men and 1,749 women, found that 2.7% of the men and 1.3% of the women had participated in same-sex sexual behavior during the preceding year, and 4.9% of the men and 4.1% of the women had done so since age 18; 7.7% of the men and 7.5% of the women reported having experienced sexual desire for someone of the same-sex, and 2.8% of the men and 1.4% of the women reported a homosexual or bisexual identity. These figures varied considerably across groups based on age, marital status, education, religion, race, and place of residence. Several surveys done over the past 20 years in Canada, Australia, and Europe have produced similar results.

A 2011 report by the Institute of Medicine called attention to marked disparities in sexual and gender minority populations in both physical and mental health. It called for all federally funded national health surveys to include queries regarding sexual orientation and gender identity so that health disparities could be characterized and addressed in different sexual and gender minority populations. Currently, 10 national data systems have begun to collect such data. Of these, two, the National Survey of Drug Use and Health (NSDUH) and the National Survey of Family Growth (NSFG), have assessed three domains of sexual orientation (attraction, behavior, and identity) in adults ages 18 to 44 and the Population Assessment of Tobacco and Health (PATH) has assessed these domains in adolescents ages 14 to 17 (Table 21.2–1). For men (ages 18 to 44), these surveys found that over 2.3% to 2.8% endorsed being mostly or exclusively attracted to men, while approximately 2% identified as gay, with approximately another 2% identifying as bisexual. For women (ages 18 to 44), 1.6% to 2.5% reported being mostly or only being attracted to the same gender, with 1.6% to 2.5% reporting being attracted to both genders, while 1.3% to 1.5% identified as lesbian or gay.

Table 21.2–1.
Sexual Attraction and Identity Reported by Adults and Adolescents in U.S. National Surveys

Sexual Attraction Reported by Adults

Males	**NSDUH**	**NSFG**
Only/mostly to other gender	93.8	95.3
Equally to both genders	1.1	0.8
Only/mostly to same gender	2.8	2.3
Unsure	1.3	0.7
Females		
Only/mostly to other gender	90.5	93.4
Equally to both genders	4.3	3.2
Only/mostly to same gender	2.5	1.6
Unsure	1.4	1.2

Sexual Identity Reported by Adults

Males	**NSDUH**	**NSFG**
Heterosexual	95.1	95.0
Gay	2.2	1.9
Bisexual	1.4	2.0
Don’t know	0.4	[a]
Refused to answer	0.8	[a]
No data/missing response	0.1	[a]
Females		
Heterosexual	92.9	92.2
Gay/Lesbian	1.5	1.3
Bisexual	3.5	5.5
Don’t know	0.8	[a]
Refused	1.2	[a]
No data/missing data	0.1	[a]

Attraction & Identity Reported by Adolescents in PATH Survey

Attraction	**Males**	**Females**

Other sex only	85.6	64.8
Mostly other sex	5.1	13.8
Equally to both	2.0	8.2
Mostly same sex	0.9	2.0
Only same sex	1.1	1.2
No sexual attraction	5.2	10.0
Identity	**Males**	**Females**
Straight	94.6	84.0
Lesbian/Gay	1.3	1.9
Bisexual	2.7	9.8
Something else	1.5	4.3

All data are in percentage.

NSDUH, National Survey on Drug Use and Health (conducted 2015; ages 18–44; *N* = 68,053).

NSFG, National Survey of Family Growth (conducted 2011–2013; ages 18–44; *N* = 10,416).

PATH, Population Assessment of Tobacco and Health Survey (conducted 2015–2016; ages 14–17; *N*= 7,222).

[a]Low precision estimate.

A report published by Rahman and colleagues in 2020 examined the prevalence of self-reported sexual identity and sexual attractions in 28 nations using data from 191,088 participants from a 2005 BBC Internet Survey. Sexual identity was reported categorically as heterosexual, gay/lesbian, or bisexual, and sexual attractions were reported using separate 7-point scales to indicate the degree of attraction to men and to women. Across nations, rates of endorsing a heterosexual identity were essentially the same for men (90.0%) and women (90.7%); however, a significantly greater proportion of men (4.9%) compared to women (2.1%) endorsed a homosexual identity ($P < .001$). The proportion of participants who endorsed a heterosexual identity was significantly ($P < .001$) higher than the proportion who endorsed being not predominately attracted to the same-sex (82.6% and 66.2% for men and women, respectively). Thus, participants who labeled themselves as “heterosexual” included a substantial number of men and women who also endorsed significant attraction to their own sex.

These later studies do not support the aforementioned popular belief that Kinsey's study showed that 10% of the population is gay. In the 12 largest U.S. cities, however, the number of gay and lesbian persons does approach the 10% figure, perhaps reflecting their migration to areas with greater tolerance for sexual diversity or sizable LGB communities. Yet, the 10% figure in the Rahman study for those endorsing something other than a heterosexual identity is similar to the 10% figure from the Kinsey study for men who reported being "more or less exclusively homosexual for at least three years between the ages of 16 and 55."

Given the limitations in existing studies, it is impossible to determine definitively the number of persons who have experienced same-sex attraction, have engaged in same-sex behavior, or identity as LGB or other sexual minority; however recent large-scale studies employing modern survey methods provide similar results.

Historically, there has been a political dimension surrounding the issue of homosexuality's prevalence in the population. Proponents and opponents of gay rights have attempted to respectively exaggerate or diminish the importance of the gay and lesbian presence in the population at large. Although the exact numbers and proportions remain somewhat elusive, it is nevertheless clear that a larger proportion of persons report same-sex desire than report same-sex behavior, and more report engaging in same-sex behavior than report a homosexual or bisexual identity. For example, a 2006 study of men in New York City found 12.4% reported having had sex only with men during the previous year; however, the majority (8.9%) identified as "straight," with only 3.3% identifying as gay and 0.2% identifying as bisexual. For the clinician, it is important to understand that prevalence is dependent on which component of homosexuality and which social or demographic groups are being considered. The validity of estimates of the extent of homosexuality in a given population is likely to be influenced by factors such as the degree of social repression of homosexuality and the cultural meaning of same-sex desire, behavior, and identity. Importantly, with respect to sexually transmitted infections, including HIV, sexual behavior is the risk factor and not sexual identity; therefore, questions to assess risk should address behaviors, not identity.

ORIGINS OF HOMOSEXUALITY

Historically, most research has been done from a heteronormative perspective that treats heterosexuality as normal and therefore not in need of explanation, whereas homosexuality has been viewed as an abnormality to be explained. Consequently, much research has focused on homosexuality in particular rather than on sexual orientation in general, and, until recently, often with the aim of elucidating the origins of homosexuality in order to inform efforts aimed at preventing or “curing” it. Four areas in which research on the origins of homosexuality has focused are biology, cross-cultural studies, psychological development, and the acquisition of gay and lesbian identities.

Biologic Studies

Since the emergence of the concept of sexual orientation in Western medical discourse, there has been a debate as to whether it is determined or primarily influenced by biologic as opposed to psychosocial factors. In recent years, most scholars have suggested that it emerges from an interaction between both sets of factors. The role of biology often enters into sociopolitical debates, sometimes referred to as culture wars, regarding whether or not homosexuality is mutable (perhaps even a chosen lifestyle) and whether gay people should be encouraged to change or be granted civil protections against discrimination. Weaknesses in the evidence base for a biologic contribution to sexual orientation are sometimes employed to support arguments for conversion therapies or discriminatory policies. As shown below, however, even behaviors acquired through experience can be immutable. Furthermore, biologic evidence is not needed to demonstrate that homosexuality is impervious to change. That is clearly evident in the psychological literature reporting the ineffectiveness of multiple forms of therapies designed to change it. While sexual behaviors and sexual orientation identities may shift in some individuals, few if any of even the most highly motivated have changed their homosexual orientation despite subjecting themselves to painful aversion therapies or investing emotional and financial resources in years of psychoanalysis or reparative therapies. Rather than issues of choice, biologic determinism, and immutability, the arguments for equal rights and civil protections for homosexual and bisexual persons, like those for religious and racial minorities, are better framed in terms of human rights, justice, privacy, and liberty.

Before reviewing the biologic research into sexual orientation, this section begins by describing three basic models for conceptualizing the possible contributions of biologic factors and their interactions with psychological and social factors.

Biologic Models to Explain Homosexuality

MODEL OF PERMISSIVE BIOLOGIC EFFECTS. In this model, biology plays a permissive role by providing the neural machinery through which sexual orientation is inscribed by formative experience. A permissive role could also include delimiting the developmental stage during which the relevant formative experiences must occur. For example, some songbirds must learn their species' songs by hearing them during a restricted period of early development. While the song is clearly acquired through experience, biology determines the sensitive period during which that experience must occur. Once a particular song has been acquired, that will be the bird's song for life; it will not be able to unlearn that song or acquire another. This is not meant to imply that sexual orientation might be acquired by simple mimicry. Instead, it shows that phenomena acquired through experience may, nonetheless, be immutable and, therefore, militates against the widespread notion that sexual orientation must be innate since it is impervious to change.

MODEL OF DIRECT BIOLOGIC EFFECTS. In this model, the underlying premise for most biologic research regarding sexual orientation, the actions of biologic factors such as genes and hormones would directly influence the organization or activity of brain circuits that mediate sexual orientation. The brain is viewed as possessing an intrinsic (i.e., constitutional) sexual orientation or predisposition, the expression of which could nonetheless be subsequently modified by experience.

MODEL OF INDIRECT BIOLOGIC EFFECTS. This model addresses the contribution of individual differences, some of which may be constitutionally influenced. In this model, biologic factors influence sexual orientation only indirectly as the more direct or immediate action of biologic factors is upon temperament or other personality traits. From birth, these personality or temperamental traits influence how an individual experiences, interacts with, and modifies the environment. This would include how an individual participates in shaping the relationships and experiences thought by some to influence the development of sexual orientation. Although similar to the model of permissive biologic effects, the model of indirect biologic effects goes further and suggests that the relevant formative experiences, including those that are subjectively experienced, may, themselves, be strongly affected by biologically influenced personality variables.

The existing biologic data relevant to the underpinnings of sexual orientation are equally compatible with both the direct and indirect models. The distinction between these models can be appreciated in their differing interpretations of three of the more robust findings in the sexual orientation literature as reviewed by William Byne and Bruce Parsons. The first of these findings is that the propensity to engage in rough-and-tumble play during the juvenile period appears to be influenced by prenatal exposure to androgenic (i.e., “male”) hormones. Second, compared to heterosexual men, more, but not all, homosexual men recall less childhood interest in competitive rough-and-tumble play. Third, compared to heterosexual men, homosexual men are more likely to recall their fathers as having been distant or rejecting.

In the direct model interpretation, the aversion to rough-and-tumble play represents the childhood expression of a brain that has been prewired for homosexuality. This was the position of Richard Isay, a psychoanalyst who suggested that biologic factors wire the brain for sexual orientation and consequently reverse the polarity of the Freudian Oedipus complex. According to this model, in addition to shunning rough-and-tumble play, prehomosexual boys are erotically interested in their fathers during the oedipal period. Fathers might recoil from their prehomosexual sons’ gender nonconformity or sexual inclinations. Even if the father did not recoil, Isay speculated that in adulthood, gay men might nevertheless recall their fathers as having been cold or distant in order to avoid conscious awareness of their earlier sexual attractions to them.

Alternatively, an indirect model interpretation postulates that a biologically influenced aversion to rough-and-tumble play does not necessarily imply prewiring for homosexuality. Instead, such an aversion becomes a potent factor predisposing to homosexual development in particular environments—perhaps where this behavior is stigmatized as “sissy” behavior and, therefore, causes the boy to see himself as different from his father and male peers. In this scenario, the father's withdrawal from his son would contribute to, rather than result from, his homosexuality. Significantly, an aversion to rough-and-tumble play would arguably have different consequences in environments in which such behavior is acceptable, perhaps making no contribution to sexual orientation at all. This example does not suggest that a distant or rejecting father (whether real or projected) could be a feature of all pathways to male homosexuality. Based on the indirect model, one might conjecture how any number of biologically influenced personality variants could influence sexual orientation. Theoretically, a given variant might predispose to homosexuality in one environment and to heterosexuality in another while making no contribution to sexual orientation in environments in which the variant is perceived to have no valence with respect to perceptions of gender-appropriate behavior.

Neuroendocrine Research. Historically, much biologic research has begun with the premise that the constitution of gay men and lesbian women reflects some intermediary state between that of their male and female heterosexual counterparts. This assumption led to investigations of various attributes for evidence of biologic sexual atypicality in gay men and lesbian women, whether in genes, chromosomes, hormones, hormone receptors, or brain structures. From the 1950s into the 1970s, considerable research examined the endocrine system for atypical levels of the so-called sex hormones in gay men and lesbian women. An overwhelming majority of such studies failed to demonstrate a significant correlation between sexual orientation and adult hormone levels. In fact, androgens, frequently referred to as "male hormones," although they are present in both sexes, have been found to increase sexual desire in both men and women and yet have no effect on sexual orientation. Moreover, sexual orientation does not change in adults when hormone levels are altered by gonadal malignancies, trauma, or surgical removal. Currently, however, there remains much speculation about the potential role of prenatal exposure to androgens and other hormones derived from them. Variations in exposure could be mediated by variations in hormone syntheses, transport mechanisms, blood or brain levels, enzymatic degradation or modification, or variations in hormone receptors or epigenetic modifications of DNA that affect any of the foregoing.

THE PRENATAL HORMONAL HYPOTHESIS. This hypothesis posits that (1) the brains of heterosexual men and women differ from each other both structurally and functionally; (2) these sex differences in the brain result, at least in part, from early, perhaps prenatal influences on brain development; and (3) some of these hormonally influenced differences determine whether one is sexually attracted to men and/or to women. This perspective, therefore, views sexual orientation as a derivative of a hormonally mediated developmental process leading to sexual differentiation of the brain. Consequently, the brains of homosexual individuals are expected to exhibit neurobehavioral characteristics considered to be more typical of the other sex. This supposition has sometimes been referred to as the *intersex hypothesis* of homosexuality, that is, that the brain of homosexual and bisexual individuals is intermediate in some ways between a prototypical male and female brain.

The prenatal hormonal hypothesis draws extensively on numerous studies and observations of rodents, in which the balance between male and female patterns of mating behaviors is strongly influenced by the amount and timing of early androgen exposure. The period of maximal sensitivity to these organizing effects of androgen varies in different species. The rat has been employed extensively in such research because its period of brain sexual differentiation extends into the early postnatal period (which corresponds approximately to the midtrimester of human gestation). Thus, hormonal exposure of the rat's brain can be experimentally manipulated during the period of sexual differentiation by perinatal gonadectomies and injections of various hormones as well as their receptor agonists and antagonists. Such research has revealed that different aspects of sexual differentiation of the rat brain are mediated by different pathways of androgen action in the brain. Adding to this complexity, the extent to which sexual differentiation of the brain relies on particular androgens and their metabolites appears to vary across species. Thus, laboratory findings cannot be automatically or easily extrapolated from one laboratory animal species to the next, let alone from animal models to humans. This section first reviews what is known about sexual differentiation of the brain in rats. Subsequent sections highlight particular aspects of the rat model that do not apply to primates, including humans. The reader is advised to always be cautious when reading materials about human sexuality that discuss sexual differentiation of the brain generically without identifying the species from which particular research findings were derived.

Normal development of the male rat brain requires completion of two processes: (1) masculinization—induction of the male phenotype, and (2) defeminization—removal of the female phenotype. Both processes, although distinct, are largely directed by the same steroid hormone, estradiol, which is derived from testosterone by the action of aromatase enzymes in the brain. Subsequently, the brain-derived estradiol interacts with estrogen receptors. Various hypotheses have been advanced to explain why the brains of female fetuses are not masculinized and defeminized by estrogens of maternal origin or those that originate from their own ovaries. A possible explanation is that peripheral estrogens are bound by proteins that hinder their crossing of the blood–brain barrier.

The binding of estrogens to their brain receptors activates a variety of downstream cascades, one of which involves the upregulation of prostaglandin E2. This pathway mediates the masculinization of sexual behavior in rats but does not affect defeminization of sexual or maternal behavior. Thus, the intracerebral administration of prostaglandin E2 to female rat pups produces animals that are behaviorally masculinized but not defeminized. That is, they exhibit male-typical mounting behavior in addition to the full range of female sexual and maternal behaviors. Conversely, administration of indomethacin, which blocks the synthesis of prostaglandin E2, in male pups, produces animals whose behavior is defeminized but not masculinized. Although estradiol simultaneously defeminizes and masculinizes behavior, those actions are mediated by divergent mechanisms downstream of the estrogen receptor.

Problems inevitably arise when one attempts to draw conclusions about human sexual psychology based on extrapolations from rodent behaviors caused by experimental, prenatal, or perinatal endocrine manipulation. For example, a neonatally castrated male rat that shows a female receptive posture called *lordosis,* when mounted by another male, is considered by some to be “homosexual,” as is the perinatally androgenized female rat that mounts other females. The male that mounts another male is sometimes considered to be exhibiting heterosexual behavior, as is the female that displays lordosis when mounted by another female. In this particular laboratory paradigm, sexual orientation is defined in terms of specific behaviors and postures. In contrast, human sexual orientation is defined not by the motor patterns of copulation but by one’s pattern of erotic responsiveness and the gender of one’s preferred sex partner.

Because of the problems in equating rodent behavior with human sexual orientation, researchers have used a variety of strategies to assess partner preference in animals. This is sometimes done by seeing whether a test animal chooses to approach a male or a female stimulus animal placed in opposite arms of a T-maze. Although some unaltered laboratory animals will spontaneously direct most of their sexual behaviors toward their own sex, animal studies of sexual orientation are usually carried out on animals that have been experimentally manipulated. For example, a genetically male rodent may either be castrated as a neonate (depriving his developing brain of androgens), or particular androgen-responsive regions of his brain may be destroyed. However, in order to activate the display of female-typical behaviors and preferences in such male animals, estrogen injections are also required in adulthood. Similarly, androgens must be given again after puberty in order to activate mounting behaviors in female animals that were exposed to androgens prenatally. Because adult homosexual men and women have hormonal profiles that are indistinguishable from those of their heterosexual counterparts, it remains unclear how findings based on such hormonally abnormal animals say anything informative about human sexual orientation. Studies are beginning to examine the neurobiologic underpinnings of sexual behavior in male sheep that, without any experimental manipulation, do not mount females but instead mount only males. The males that receive mounts are often weak or incapacitated and unable to fend off the mounting sheep. These so-called gay sheep have not been described as displaying the feminized patterns of behavior that in most animal models are associated with developmental androgen deficiency in males.

Ethical constraints make it impossible to test the prenatal hormonal hypothesis in humans. That is, one cannot experimentally manipulate the prenatal hormonal exposure of human fetuses to study the effects on subsequent sexual orientation. However, there are alternative strategies. These include assessing sexual orientation in individuals with known or suspected prenatal hormonal abnormalities and assessing presumed correlates of prenatal hormonal exposure in known homosexual and heterosexual individuals.

ORIENTATION FOLLOWING DOCUMENTED PRENATAL HORMONAL ABNORMALITIES: DIFFERENCES/DISORDERS OF SEX DEVELOPMENT/INTERSEX CONDITIONS. The prenatal hormonal hypothesis would predict that an increased proportion of men with medical conditions known to involve prenatal androgen deficiency or insensitivity would be homosexual or transsexual. The same expectation would hold for women exposed prenatally to excess androgens. Because androgens are required for the development of the external male genitalia, intersex individuals may be born with external genitals that are intermediate in morphology between those typical for males and females. Affected individuals, historically referred to as *hermaphrodites* or *pseudohermaphrodites* (depending on their gonadal status), were later referred to as individuals with *intersex conditions* and currently as individuals having *differences or disorders of sex development* (DSDs).

The incidence of homosexuality (relative to the sex assigned at birth) is increased in some intersex conditions, particularly among female-assigned individuals who were exposed to elevated androgens prenatally. Interpretation of these results, however, is complicated by several complex variables. For example, the rearing of individuals born with external genitalia that are not clearly male- or female-typical may not be unambiguous with regard to the assigned gender, despite the intentions of physicians and parental figures. This may occur when the affected individual has had multiple surgeries in an attempt to construct stereotypically normal-appearing genitalia. Even surgeries performed so early that the individual has no memory of them may leave physical scars or other anomalies that give rise to concerns about gender. Moreover, parents may remain ambivalent about the gender of the intersex child and communicate this unwittingly and nonverbally to the child. Thus, studies that detect an increased incidence of lesbian or bisexual identity, behavior, or fantasy in female-assigned individuals with somatic intersex conditions do not provide unequivocal support for the direct model of androgenic effects on sexual orientation.

Individuals with *complete androgen insensitivity syndrome* (CAIS) due to mutations of the androgen receptor are an important exception to the rule of increased homosexuality among female-assigned individuals with intersex conditions who were exposed to elevated prenatal androgens. These individuals were exposed to high androgen levels in utero; however, their tissues lack the capacity to respond to androgens. Therefore, their external genitalia follows the path of female differentiation, although they are genotypically male and have normally functioning testes. Because of their external appearance, they are assigned as females at birth and reared as such. Until relatively recently, they often remained undiagnosed until puberty, when they failed to menstruate. Despite having all of the known metabolic machinery and receptors necessary to act on the brain to both defeminize and masculinize sexual behavior in the rat, the vast majority of adult individuals with CAIS have been described as maintaining a female gender identity and exhibiting stereotypically female behavioral patterns and interests, including sexual interest in males. These observations led some to suggest that in contrast to rodents, male-typical differentiation of the human brain is mediated by the activation of androgen receptors and does not require that androgens be first converted to estrogens.

PRESUMED CORRELATES OF PRENATAL HORMONAL EXPOSURE IN HOMOSEXUAL INDIVIDUALS. Another strategy in attempting to establish a link between prenatal hormonal exposure and sexual orientation has been to examine presumed correlates of such exposure in known homosexual and heterosexual individuals. The most obvious correlates would be abnormalities of testicular or ovarian function and abnormalities in the differentiation of the external genitalia. Such variations are rare in homosexual individuals and lend no support to the prenatal hormonal hypothesis; however, androgens could influence genital development and sexual orientation at different periods of development. Alternatively, differentiation of the substrates that mediate sexual orientation could be orchestrated by an as-yet unidentified androgen-mediated pathway that does not influence the other variables studied. Animal research suggests two additional correlates of early androgen exposure that have been explored in humans as possible correlates of sexual orientation: The pattern of luteinizing hormone secretion by the pituitary gland elicited by estrogen and neuroanatomical sexual dimorphisms.

LUTEINIZING HORMONE RESPONSES. In rats, exposure of the developing hypothalamus to aromatizable androgens (i.e., androgens that can be converted to estrogens) determines the signal that the adult’s brain will relay to the pituitary gland in response to high levels of estrogen in the bloodstream. If rats do not experience high levels of aromatizable androgens at a certain early phase of development (as in normal males), the adult brain will respond to estrogen by directing the pituitary to increase its secretion of luteinizing hormone; a phenomenon referred to as *positive feedback.* However, if the developing rat brain is exposed to high androgen levels during this developmental period, it will not be able to support this positive feedback response. Thus, in rats, the presence or absence of this positive feedback response can be used to infer the pattern of the brain's early androgen exposure. Specifically, the positive feedback response can be elicited in adulthood from normal females and from males that were castrated during the sensitive period of sexual differentiation of the brain, but not from normal males or from females that were experimentally exposed to high androgen levels during that sensitive period of brain development.

These observations generated speculation that in humans, the response of luteinizing hormone to an injection of estrogen would also indicate the sexually differentiated state of the brain and correlate with sexual orientation. Empirical research in this area has produced conflicting results; however, the current consensus opinion is that this feedback response in humans does not correlate with sexual orientation. Moreover, in humans and other primates, convincing research suggests that the brain mechanism regulating luteinizing hormone is the same in both sexes, rather than taking two sexually distinct forms as it does in rodents. If there is no sex difference in the feedback mechanism, it cannot be argued logically that the mechanism should exhibit atypical sexual differentiation in gay men and lesbian women. Since the positive feedback response is required for cyclical ovarian function and fertility in females, menstruation and pregnancy in lesbian women also support this interpretation.

NEUROANATOMICAL SEXUAL DIMORPHISM. Over the past four decades, sex differences have been confirmed in the size of several brain structures in a variety of laboratory animals. These findings have generated speculation concerning the existence of parallel differences in the human brain associated not only with sex but also with sexual orientation.

Most of the structural sex differences identified involve specific cell groups within a broad region of the rodent hypothalamus that participates in regulating a variety of functions, including sexually dimorphic copulatory behaviors. Like the sex differences in copulatory behaviors, several structural sex differences develop in response to sex differences in early androgen exposure: High androgen levels at the appropriate time lead to male-typical anatomy, whereas low levels lead to female-typical anatomy.

Consequently, behavioral sex differences are thought to be mediated, at least in part, by structural differences.

The best-studied anatomical sex difference involves a cell group that straddles the medial preoptic and anterior regions of the hypothalamus—the sexually dimorphic nucleus of the preoptic area (SDN-POA). In the rat, this structure is five to eight times larger in males than in females. Damage to the preoptic region decreases mounting behavior in laboratory animals, whereas electrical stimulation elicits mounting behavior.

The belief that the SDN-POA participates in regulating male sex behavior in rats led to the search for a comparable nucleus in humans. A cell group called *nucleus intermedius* but also known as the *first interstitial nucleus of the anterior hypothalamus* (INAH1), was described by one laboratory as being larger in men than in women and was, therefore, prematurely designated as the *sexually dimorphic nucleus* (SDN) in atlases of the human hypothalamus. Of four laboratories that examined this nucleus for variation with sex and sexual orientation, only one found it to vary with sex, and none found variation with sexual orientation. Its designation as the SDN, therefore, appears to have been premature.

Another structure, the *third interstitial nucleus of the anterior hypothalamus* (INAH3), is perhaps a more promising candidate for homology with the rat's SDN-POA. Three independent laboratories found this nucleus to be much larger in presumed heterosexual men than in women, and one laboratory found the nucleus to contain more neurons in men. By extrapolation from animal work, this sex difference is widely believed to reflect sex differences in early hormone exposure, although, at present, that hypothesis cannot be directly tested in humans. The AIDS epidemic made it possible to study this nucleus from deceased individuals whose medical records indicated homosexual behavior as their HIV risk factor. (Unless someone dies from complications of AIDS, there is usually no documentation of sexual orientation in the medical records available for autopsy studies.) Presumed heterosexual control brains derived, in part, from individuals whose HIV risk factor was intravenous drug use. Those studies found the volume of INAH3 to be smaller in homosexual men than in heterosexual men, but the number of neurons within the nucleus did not vary with sexual orientation. The suggestion of volume reduction must be viewed skeptically for a variety of technical reasons. For example, tissue shrinks in the process of fixation for histologic analysis. This shrinkage influences measures of volume but not measures of cell number. Thus, the finding of equal numbers of neurons in homosexual and heterosexual men may be a more reliable finding than the suggestion of a difference in the volume of the nucleus. Alternatively, the presence of an equal number of cells in a smaller volume could reflect decreased neuropil (e.g., synapses, dendrites) in homosexual men. Decreased neuropil could conceivably reflect low androgens either prenatally or at the time of death due to the antiandrogenic effects of some medications used to combat the opportunistic infections of AIDS patients. A bias could have been introduced into these studies since the majority of heterosexual male subjects with AIDS were intravenous drug users who, compared to gay males, are likely to be less adherent with medication regimens. Additional studies would be required to clarify these issues.

Speculation regarding the function of INAH3 has been based on the assumption that it is the homolog of the rat's SDN-POA. The finding that the size of the SDN-POA correlates positively with the frequency of mounting behavior displayed by male rats established the belief that the SDN-POA participates in regulating male sex behavior. That interpretation, however, is at odds with the fact that the rat's nucleus can be destroyed by electrolytic lesions on both sides of the rat's brain without any discernible effect on mounting behavior. A better case can be made that the SDN-POA functions to inhibit the expression of a female mating posture by male rats. In rats in which intracerebral prostaglandin E2 was manipulated, the volume of the SDN-POA was found to correlate with defeminization and not with the masculinization of copulatory behaviors. Moreover, following lesions that include the SDN-POA, male rats can be induced to display the female copulatory response if they are also injected with estrogen and progesterone.

In addition to the hypothalamus, researchers have sought to identify variation with sex and sexual orientation in the brain commissures, the fiber bundles that connect the left and right hemispheres of the brain. The rationale underlying these studies was drawn from speculation regarding the neuroanatomical basis of statistical sex differences in the lateralization of particular cognitive functions within the brain. Several lines of investigation suggest that women, compared to men, are likely to exhibit more bilaterally symmetrical hemispheric activation while performing particular cognitive tasks. A popular hypothesis holds that this sex difference in brain activity reflects increased interhemispheric communication in women compared to men and that this increased communication between the hemispheres is due to more interconnecting fibers between the hemispheres. More than 50 studies have examined the width or area of particular portions of the corpus callosum, the largest of the brain's commissures, for sex differences. Popular accounts to the contrary, the majority of these studies have found no sexual dimorphism. Studies controlling for variation in the callosum as a function of handedness and aging, however, have found differences in the isthmus of the corpus callosum, which contains axons connecting the right and left parietotemporal cortical regions. These regions mediate functions of language and spatial cognition that are asymmetrically represented and that, according to some studies, vary with both sex and sexual orientation.

NEUROPHYSIOLOGIC MEASURES. A number of neurophysiologic measures have been reported to differ between homosexual men and women and their heterosexual counterparts. Several of these studies report that homosexual individuals exhibit neurophysiologic responses more typical of heterosexuals of the other sex, but to date, none have been adequately replicated by independent laboratories. These include studies of regional patterns of brain activation in response to exposure to putative pheromones. Although positive findings in this area should not be ignored and certainly merit attempts at replication, the existence of a functional vomeronasal organ in humans, thought necessary to respond to pheromones, is itself controversial. In addition, lesbian women have been reported to exhibit masculinized otoacoustic emissions, echo-like waves emitted by the inner ear response to brief sounds, which has been conjectured to reflect increased androgen exposure prenatally. No difference in these emissions has been detected between homosexual and heterosexual men.

ANTHROPOMETRIC CHARACTERISTICS. Several anthropometric characteristics, believed by some to reflect androgen exposure, have been explored in relation to sexual orientation. The characteristics examined include measures of height and weight as well as the amount and distribution of facial hair, the ratio of shoulder width to hip width, the size of the genitalia, and more recently, dermatoglyphic (fingerprint) characteristics, and finger-length ratios. Most of these studies have been flawed in one or more ways, making the occasional positive findings difficult to interpret. Some of these flaws include reliance on self-reports of small, self-selected samples or on measures obtained by raters who were not blind to the sexual orientation of the subjects.

Finger-length ratios have been subjected to considerable investigation for variation with both sex and sexual orientation. The ratio of the length of the second to fourth fingers (2D:4D) was first proposed as a marker for prenatal androgen action in 1998. Since then, over 100 studies have been published testing its association with prenatal androgens or using 2D:4D as a putative marker to investigate the association between prenatal androgens and a variety of outcomes, including sexual orientation. The validity of digit ratios to serve as a marker of prenatal androgen action, however, is highly controversial. Some but not all studies suggest that adult women tend to have longer index fingers (second digits) than adult men relative to the length of the ring finger (fourth digit). In adults, the correlation between the 2D:4D ratio is not high, with a 60% overlap between males and females. The study of congenital adrenal hyperplasia in individuals treated soon after birth has been the gold-standard method for studying the effects of prenatal or perinatal androgens. Examination of the 2D:4D ratio in affected individuals has produced equivocal results. Similarly, study of the 2D:4D ratio as a function of sexual orientation have produced discrepant results. A decreased ratio has been reported with more consistency in lesbian women relative to heterosexual women than has an increased ratio in gay relative to heterosexual men. Nevertheless, the inability to definitively link this ratio to either sexual orientation or prenatal hormone exposure due to inconsistent replication studies does not lend substantial support to the prenatal hormonal hypothesis of sexual orientation in either sex.

PROPOSED MECHANISMS OF ATYPICAL ANDROGEN EXPOSURE. As reviewed above, in men, clinical syndromes involving abnormal androgen function rarely have an impact on sexual orientation, whereas an association between elevated prenatal androgen levels and sexual orientation is statistically demonstrable in women in the context of particular clinical syndromes. Maternal stress during pregnancy and its putative impact on fetal hormone exposure has been examined in relation to homosexuality in men but is no longer widely believed to be a contributory factor.

In multiple samples investigated by one group of researchers, homosexual men have been found to have a greater number of older brothers than heterosexual men. According to these researchers, each older brother increases the “relative risk” of being homosexual by 33% to

48%, and this antecedent older brother effect accounts for the orientation of one of every seven gay men. The hypothesis advanced by these authors is that the late birth order, with more male siblings born earlier, could lead to a progressive immune response of the mother to products associated with a male, which could impair the defeminization and masculinization of subsequent male fetuses. The existence of such “antimale antibodies” is hypothetical, as is the male-associated factor that would trigger their production. Androgens have been hypothesized in this regard; however, under usual circumstances, they are not antigenic. The Y-linked minor histocompatibility locus has also been hypothesized as the triggering factor. Yet, how such mechanisms could selectively impair only certain, less well-established processes, such as the organization of the brain, and not others, like the formation of the genitalia, is not explained by this hypothesis. Nor does this theory explain why the majority of boys late in birth order do not become homosexual, even if their elder brother is homosexual. Finally, the effect has been reported in several studies as applying only to right-handed gay men, while left-handed gay men had fewer older brothers than expected. The biologic premise of the antecedent older brother effect remains questionable, and attempts at replication by independent groups of researchers are needed.

Genetic and epigenetic variations in relation to sexual orientation are areas of current research, as discussed below. Much research in this area considers the hypothesis that genetic or epigenetic mechanisms could affect an individual’s sexual orientation through an impact on hormonal function.

Genetic Research. As in neuroendocrinologic research, some genetic studies have also been premised on the intersex hypothesis of homosexuality. For example, there have been attempts to show that gay men and lesbian women have chromosomal material typical of the other sex in their cells and attempts to link homosexuality with genetically controlled differences in the process of sexual differentiation. To date, these studies have produced negative results. Other genetic approaches, not necessarily wedded to the intersex hypothesis, including heritability as well as genetic linkage and association analyses, have also been employed and have provided some leads.

FAMILY AND TWIN STUDIES. Although homosexuality has been found to run in families, such findings are not helpful in distinguishing between genetic and environmental influences because most related individuals share both environmental and genetic influences. A first step toward disentangling genetic and environmental influences can be made by studying sexual orientation in identical and fraternal twins. One such study assessed sexual orientation not only in identical and fraternal twins but also in the nontwin biologic brothers and the unrelated adopted brothers of homosexual men. The concordance rate for identical twins (52%) in that study was much higher than the rate for fraternal twins (22%). Assuming that environmental influences would be the same for all brothers reared together, the higher concordance rate in the identical twins is consistent with a genetic effect, as identical twins share all of their genes, while fraternal twins, on average, share only half of their genes.

It would be a mistake, however, to attribute the increased concordance rate in identical twins to increased gene sharing alone. If there were no environmental effects on sexual orientation, then the rate of homosexuality among the adopted brothers should have equaled the rate of homosexuality in the general population. Recent studies place the prevalence of homosexuality in men between 2% and 5%. That the concordance rate for homosexuality in adopted brothers was found to be 11% (much higher than in the general population) suggests an environmental contribution. The concordance rate for homosexuality among nontwin biologic brothers was only 9%, a figure statistically indistinguishable from the 11% recorded for adopted brothers. If the concordance rate for homosexuality among nontwin brothers is the same whether or not the brothers are genetically related, then the concordance rate cannot be explained exclusively by genetics.

When considered together, the data from the studies of twins and adopted brothers suggest that the increased concordance for homosexuality in identical twins may be due to a combination of both genetic and environmental influences. Acting via epigenetic mechanisms, environmental influences can induce lasting changes in gene expression without altering the nucleotide sequence of genes. Such environmental influences could theoretically be exerted by the prenatal intrauterine environment as well as by the postnatal environment. Perhaps the most interesting finding to emerge from twin studies is that approximately 50% of identical twins are discordant for sexual orientation even when they are reared together. The discordance in sexual orientation among genetically identical individuals reared in the same family has been consistent across studies and underscores just how little is actually known about the contributions of genes to sexual orientation.

LINKAGE STUDIES. The notion of a *gay gene* has captured the popular imagination. It is worth underscoring, however, that gay genes are not required for homosexuality to run in families or for researchers to determine that homosexuality is heritable. *Heritability* has a precise technical meaning and reflects only the *degree* to which a given trait is *associated* with genetic factors. It says nothing about the specific genetic factors involved or about the mechanisms through which they exert their influence. Furthermore, the concept of heritability provides no information about how a particular trait might change under different environmental conditions. Thus, homosexuality could be highly heritable even if genes influenced sexual orientation entirely through very indirect pathways. The question is often raised as to how genes predisposing to homosexuality could be maintained in the gene pool, assuming that reproduction would be decreased in individuals who carry them. The *Kin Selection Hypothesis* proposes that such genes could be maintained in the population if the costs of not directly reproducing are offset by enhancing the reproductive success of kin, perhaps through altruistic behavior or some other mechanism. For example, one study found increased fecundity among the female maternal relatives of male homosexual individuals and attributed this finding to genes predisposing to homosexuality in men.

Genes alone cannot directly determine any behavior or psychologic phenomenon. Instead, genes direct a particular pattern of ribonucleic acid (RNA) synthesis that, in turn, specifies the production of a particular protein that may then influence behavior. There are necessarily many intervening pathways between a gene and specific behaviors and even more intervening variables between a gene and a pattern that involves both thinking and behaving. The term *homosexual gene* is, therefore, without meaning unless one proposes that a particular gene, perhaps through a hormonal mechanism, directs the brain to develop in such a way that an individual is erotically attracted to others of the same gender. Evidence confirming such a simple and direct link between genes and sexual orientation is lacking.

In 1993, a highly publicized study of markers on the X chromosome presented statistical evidence that genes influencing sexual orientation reside on a portion of that chromosome known as the q28 region. A 1995 study by the same group corroborated the Xq28 linkage in an independent sample of men but found no evidence of an influence of the Xq28 region on sexual orientation in women. In 1999, an independent research team was unable to duplicate the Xq28 finding in men using a comparable experimental design. However, a 2005 genome-wide linkage study (GWLS) that included samples from the 1993 study concluded that genes in the Xq28 region might contribute to sexual orientation in men in some families but found no linkage to Xq28 in the full sample. A more recent GWLS implicated genes in a broader region of the X chromosome that includes the Xq28 region in male homosexuality and additionally implicated genes in a region of chromosome 8. The mechanisms by which genes in these regions might influence sexual orientation in men remain speculative. Such speculation includes the possibility that such genes increase sexual attraction to men in both men and women, leading to increased reproduction in women and, therefore, maintaining the genes in the population. Candidate genes, including the androgen receptor and aromatase enzyme, have also been investigated for variation in sexual orientation with negative results to date. These genes were chosen for examination because of their role in sexual differentiation of the brain in laboratory animals. As noted above, however, evidence for a role of the aromatase enzyme in sexual differentiation of the primate brain is currently lacking.

GENOME-WIDE ASSOCIATION STUDIES. While twin and family studies, including GWLS studies, suggest that sexual orientation is genetically influenced, most genetic studies lack the statistical power to implicate any specific genes in sexual orientation. Genome-wide association studies (GWAS), however, can employ much larger samples than GWLS in order to overcome the power limitation. A 2019 GWAS employed blood samples from 477,522 individuals in the United Kingdom Biobank and 23andMe with replication analyses in 15,142 samples from the United States and Sweden. Samples were coded by sex and sexual history (ever vs. never had a same-sex partner). Five autosomal loci were significantly associated with a history of reporting same-sex behavior and, in aggregate, accounted for 8% to 25% of same-sex behavior with imperfect correlations between sexes. The authors concluded that sexual behavior is influenced by not one but many genes. An important limitation of this study is that it addressed same-sex behavior and not sexual orientation, and there are other traits that might be genetically influenced and affect the odds of ever having had sex with a partner of one’s own sex. For example, openness to experience, one of the “big five” personality dimensions, is genetically influenced and shows a medium effect size difference between heterosexual and bisexual men and women. One might, therefore, hypothesize that openness to experience might have an impact on the lifetime probability of ever having sex with a person of one’s own sex, independent of one’s sexual identity.

Directions for Future Research. Historically, biologic research on sexual orientation has been driven largely by the prenatal hormonal hypothesis and has focused on the potential role of prenatal androgens. Early androgen exposure clearly organizes the mating behavioral repertoires of laboratory animals; however, it is not clear how the regulation of stereotyped reproductive behaviors in animals pertains to sexual orientation in humans. Nevertheless, it does not seem likely that early androgens influence the adult sexual behaviors of all mammals studied except humans. It is clear, however, that the role of aromatization of androgens to estrogens within the developing brain plays a lesser, if any, role in primates compared to rodents. Although early androgen exposure in human birth-assigned females increases the likelihood of some degree of gynephilia in adulthood, there is no compelling evidence for differences in prenatal androgen levels between heterosexual and homosexual men. One possibility that has yet to be explored is that some yet to be identified mechanism downstream of the androgen receptor plays a role in the development of gynephilia but no significant role in other aspects of androgen action such as external genital development. A variation of such a hypothetical mechanism might preclude the development of gynephilia. Females without this hypothetical variation would then be predisposed to gynephilia following relatively mild exposure to elevated androgens prenatally, whereas men with this variation would be predisposed to androphilia despite normal androgen levels prenatally. Currently, relatively little is known regarding the biochemical pathways that are influenced in the primate brain by the binding of androgen to its receptor. Microarray technologies for full transcriptome analysis offer a mechanism for elucidating these pathways in order to identify candidate genes for further study. As noted above, epigenetic variations, which can be environmentally influenced, could lead to such variations in the effects of androgens in the absence of genetic differences (i.e., differences in DNA sequences). Recent technologic innovations now allow high throughput examination of epigenetic modifications, which are heritable modifications to chromosomes that do not alter the genetic sequence. Large-scale heritability and linkage studies may also provide insights; however, the identification of genes associated with sexual orientation alone will be relatively uninformative in the absence of studies to identify their mechanisms of action. Such mechanisms could prove to be very indirect. The interpretation of such studies should not conflate sexual orientation with sexual behavior since, as noted above, genetically influenced personality dimensions could influence the likelihood of ever having engaged in sex with someone of the same sex, which is very different from experiencing predominant or exclusive sexual attractions to individuals of one's own sex. While GWAS have tremendous statistical power, they will not be informative with respect to the genetics of sexual orientation unless the samples upon which they are based are derived from individuals in which sexual orientation has been assessed, *per se,* instead of, or in addition to an assessment of a proxy variable such as ever having engaged in sex with a member of one's own sex as was assessed in the recent GWAS discussed above.

Cross-Cultural Studies

Ford and Beach's 1951 review of animal and cross-cultural studies, *Patterns of Sexual Behavior,* was ground-breaking in concluding that homosexual behavior (defined as "behavior involving stimulation and excitation of the sexual organs" was widespread within the animal kingdom and apparently universal in all human cultures. More specifically, they reviewed the status of homosexuality in 78 cultures of the mid-20th century and concluded that 48 approved some expressions of homosexual behavior but that 28, including the United States, did not. In disapproving cultures, homosexuality was stigmatized from childhood on into adulthood by ridicule and punishment, and sometimes even the death penalty.

Building on the work of Margaret Mead, Ruth Benedict, and other anthropologists, researchers classified structural types of sexual development and forms of homosexual organization in different cultures. Gilbert Herdt, in particular, explicated cross-cultural models based on his study of ritualized homosexuality among the Sambia tribe of New Guinea and his review of findings from other anthropologic studies of sexual practices. Although traditional developmental theory posits a presumed continuity in sexual development across the lifespan, Herdt instead suggested the idea of "discontinuity" in sexual development to explain variation in the psychologic experience and symbolic meaning of sexuality. He also explored the cultural aspects of fluidity in the experience and expression of sexuality that have defined much of the theoretical discourse on bisexuality.

Herdt also described a typology of homosexuality based on a cross-cultural organization of same-sex sexual behavior. The first type is *age-structured homosexuality,* usually involving older and younger males and generally including a sequential pattern of childhood same-sex practices that yield primarily heterosexual behaviors in adults. An example would be the male homosexual practices of classical Greece (*pederasty*). Manifestations more akin to current conceptualizations of transgender phenomena have previously been referred to as *gender-reversed homosexuality,* which refers to a reversal in the normative binary sex-role dress and behavior for a given culture. Examples of such “gender-reversed homosexuality” were considered to be the Native American *berdache* and the *fa’afafine* of modern Samoa, culturally accepted roles in which anatomically male individuals, assumed many aspects of the female gender role. Today, such phenomena are more commonly considered within the spectrum of gender identity rather than sexual orientation. The third type identified by Herdt is *role-specialized homosexuality,* in which same-sex sexual activity is restricted to certain social roles and positions. This type occurs among the shamans of some Native American tribes who, in following dreams or visions revealing their special status, crossdress and engage in homosexual behavior. Role-specialized homosexuality was also found among certain working-class women in 19th-century China. Finally, Herdt described the *modern gay movement,* consisting of egalitarian same-sex practices, as the fourth type of homosexual organization. Although acknowledging that hierarchical relationships still exist in parallel with traditional forms of heterosexual relationships, he concluded that homosexuality in the contemporary United States (and in other Western countries) is historically unique in that it comprises a sexual orientation, a social identity, and a political movement.

Psychosocial Studies of Development

Classical theories of psychological development posit the origins of adult personality and identity in childhood experience. These theories often lay out presumed linear pathways of development that progress sequentially from childhood to adulthood. In the later part of the 20th century, however, focus shifted to the distinct processes of development during adulthood itself. The development of sexual orientation might be better understood by examining psychological and interpersonal events throughout the lifespan that may be characterized by both continuous and discontinuous patterns of development. There is little empirical research to support the traditional view that early childhood behaviors, sexual awareness, and feelings are definitive precursors of adult behaviors, awareness, and feelings. Further, most psychosocial studies are severely limited by problems of sampling and have until very recently tended to focus on white, urban, openly identified, and community-affiliated gay men and lesbian women. It is not known whether one can generalize the development and experience of the latter group to the experience of individuals coming from other ethnic, racial, or culturally diverse groups. While much current research addresses intersectionality (the cumulative and interactive impacts of multiple social categories) on health disparities and how they affect sexual and gender minority populations, little research has addressed intersectional influences on the emergence of sexual orientation and identity developmentally.

Findings from psychosocial studies vary depending on the research methodology. For example, in the mid-20th century's retrospective studies of adult homosexual patients in psychoanalysis, the narratives these individuals told of their upbringing were often consistent with cultural notions of homosexuality. In other words, how these individuals remembered their early childhood may have resulted from social learning of popular theories of homosexuality, either before or during their psychoanalytic treatment, with subsequent internalization of the theory as part of their life scripts. In addition, many studies of purported childhood precursors of adult homosexuality often take a narrow focus on selected characteristics (i.e., gender role atypicality in boys) associated with stereotyped notions of adult homosexuality. Consequently, such studies tended to ignore the enormous individual variability in patterns of development toward and characteristics associated with adult sexual identity.

Historically, children with marked atypical gender behavior (who might meet criteria for today's diagnosis of gender dysphoria of childhood, or GDC), adults disturbed by their homosexual attractions and/or behaviors and voluntarily seeking psychiatric treatment, and incarcerated individuals served as the primary source of statistics about and models for the psychological development of sexual orientation and same-sex behavior. Actual empirical studies of prepubescent homoeroticism in American society are nonexistent, although there are increasing studies of gay, lesbian, and bisexual adolescents. The potential psychosocial understanding that might derive from ethnographic and other social science approaches to studying sexuality in children has been applied only in other cultures and has thus not been utilized as a tool for the study of the development of sexual orientation or sexual identity in American society. Nevertheless, some theoretical models hypothesize how adult gay, lesbian, and bisexual identities develop in Western society that are based on empirical studies of the specific behaviors and experiences associated with establishing these identities.

Antecedents of Homosexuality in Childhood and Adolescence. Early attempts to describe childhood psychosocial antecedents of adult homosexuality consisted primarily of psychoanalytic formulations derived from the histories of psychiatric patients. These usually depicted trauma, loss, and disturbed family relationships in childhood as causal mechanisms. In contrast, modern sexologists conduct field studies of larger samples of nonpatient populations. These studies are often guided by essentialist views that assume a particular personality characteristic or specific behavior is a reliable predictor of adult sexual orientation. To date, no single set of characteristics or unitary pathway leading to adult homosexuality (or heterosexuality) has been identified.

One robust finding in studies of childhood factors associated with adult homosexuality is the recollection of early gender nonconformity. That is to say, femininity in boys and masculinity in girls appear to be correlated with the development of adult homosexuality, but their predictive power is limited, and any specific relationships between childhood gender nonconformity and homosexuality remain unknown.

More recently, studies have focused on the adjustment, health, and mental health issues of young people who self-identify as gay, lesbian, or bisexual who either present to community centers providing services or who are in college. Over time, adolescents appear to be coming out at a younger age which may cause problems within the home due to a lack of parental acceptance. The U.S. Department of Health and Human Services now requires that all organizations serving runaway and homeless youth be equipped to serve sexual and gender minority/LGBTQ youth, who represent a disproportionate segment of this population.

Acquisition of Gay and Lesbian Identities. Mental health professionals should have some understanding of the theoretical description of gay, lesbian, and bisexual identity development in individuals, as well as the meanings of personal identities at a cultural level. From a social constructivist perspective, the appearance of LGB identities—and the detailed description of their emergence and multifaceted expression within both society and the individual—can be seen as an outcome of both the personal growth and development of the individual and of the cultural evolution of the concept of a "homosexual person."

Models of LGB identity acquisition describe an intrapsychic and interpersonal process whereby a subjective experience of oneself as LGB becomes increasingly congruent with societal perceptions of that identity. This process has been conceptualized as occurring over time, in sequential stages, and in a manner that leads to increasing integration of one's experience of being LGB into other aspects of one's sense of self. Vivian Cass conceptualized the first stage as a period of homosexual self-awareness and possible confusion about the presence and meaning of same-sex erotic attractions. In people who come to identify as gay, lesbian, or bisexual, the period of homosexual self-awareness is followed by a time during which the individual considers accepting these feelings and begins to incorporate them into experiences. In many cases, this phase is followed by an extended phase of growing acceptance of and perhaps even pride in the new identity.

The process of identity development leads to different outcomes for different individuals; it is influenced by interpersonal effects and consequences as well as feedback from others and may be interrupted or terminated at any point. Most models of gay, lesbian, and bisexual identity development incorporate the completion of several events or tasks: *Coming out* or becoming aware of same-sex attraction and disclosing it to others, forming relationships with other gay and lesbian persons, engaging in same-sex sexual behaviors; becoming involved in gay and lesbian activities and communities; establishing significant romantic relationships with persons of the same-sex, and learning how to integrate one's sexual orientation identity into other aspects of the self.

Models of gay and lesbian identity acquisition sometimes present an image of progressive elaboration and synthesis of homosexual orientation with other aspects of the self. However, individuals vary considerably in the degree to which they follow such models. In some persons, the development of a gay, lesbian, or bisexual identity may be inhibited in response to any variety of factors. These include a punitive or harsh internal or environmental response to openly expressing aspects of such an identity or a redirection of one's focus to other developmental tasks, such as schooling or career development. Individuals from certain religious communities may eschew adopting a gay or lesbian identity. As models of gay, lesbian, or bisexual identities historically developed based on experiences of urban White people, individuals who live in other types of communities (e.g., rural areas and communities of color) may participate in same-sex sexual activities yet eschew an identity based on these behaviors.

Individual variations in the extent of development of gay, lesbian, and bisexual identities may result both from prohibitive factors (including stigmatization of homosexuality or gender atypicality), either internal or external, that interfere with optimal personal growth and from normative psychosocial influences, reflecting diverse individual experiences and different adaptational needs. The tasks for any clinician working with these individuals are to appreciate the individuality of identity development in each patient, to understand their particular psychological and social resources and needs in this area, and to appreciate the risks and benefits associated with life's choices. Each individual will develop their own identity; the clinician can only help to optimize the understanding of the meaning of various options and to establish opportunities for maximizing freedom in making choices and better understanding the ranges of their potential significance.

Homosexuality and Nongay Identities. The complexities involved in establishing a gay, lesbian, bisexual, or other sexual identity require the clinician's tolerance, respect, patience, and sophistication about sexual matters. The different subjectivities one may encounter in working with LGB and other sexual minority patients can often be understood and appreciated in concepts like hiding/revealing, awareness/unawareness, and acceptance/nonacceptance. Patients with same-sex attractions can be helped to understand that their own sexuality will, in part, be shaped by the meanings attributed to homosexuality in a predominantly heterosexual world. The subjectivity of sexual identity is shaped by culture and language, often with little or no regard to prevailing scientific categories of sexuality. Some broad subjectivities, based on the work of Jack Drescher, are identified here to provide clinicians with a way to understand the range of ways in which patients may manage their same-sex feelings (Fig. 21.2–1).

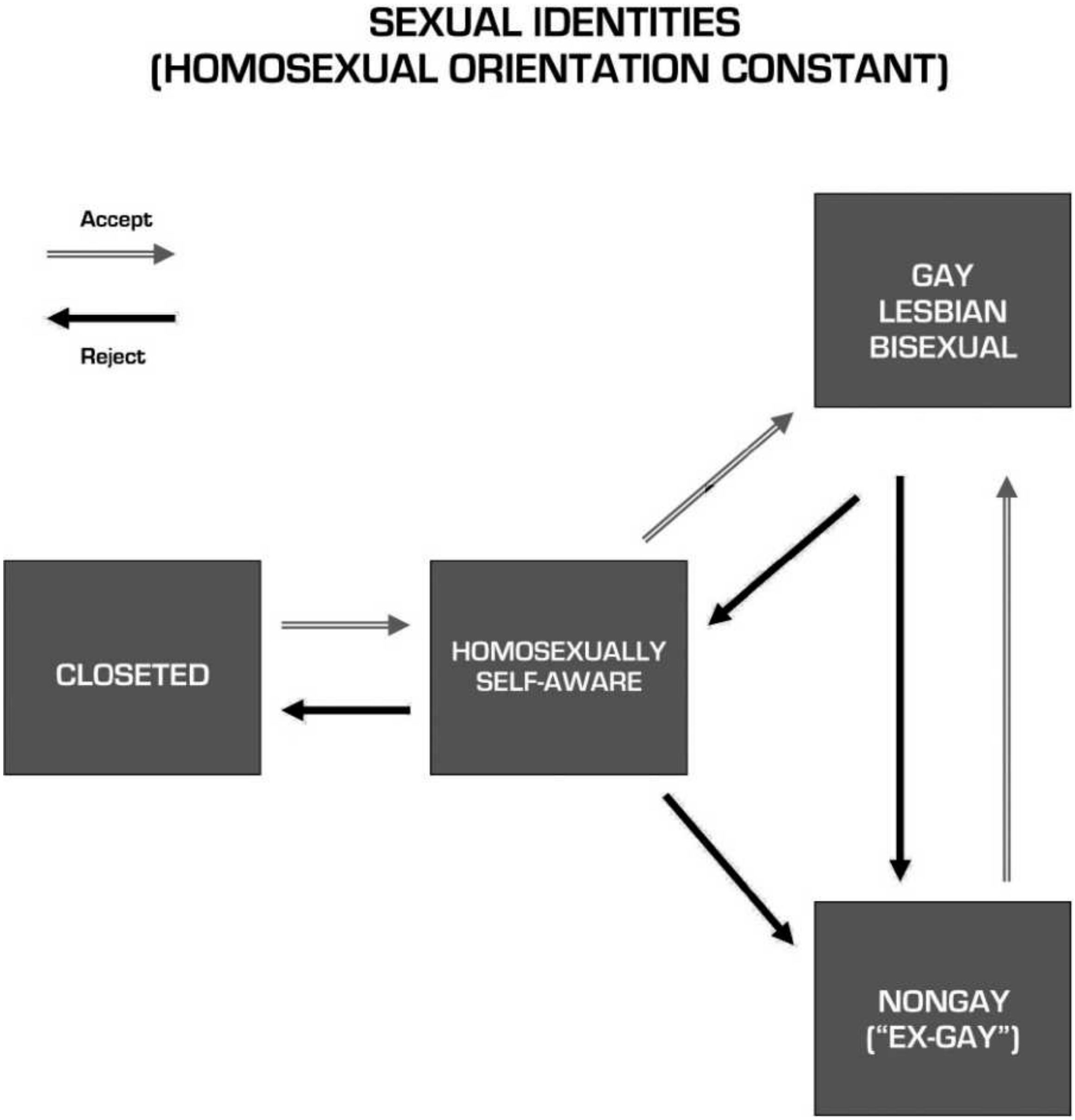
SEXUAL IDENTITIES
(HOMOSEXUAL ORIENTATION CONSTANT)
Accept
Reject
GAY
LESBIAN
BISEXUAL
CLOSETED
HOMOSEXUALLY SELF-AWARE
NONGAY ("EX-GAY")

FIGURE 21.2–1. Sexual identities (homosexual orientation constant).

Closeted individuals are unable to acknowledge to others, and sometimes, even to themselves that they have homoerotic feelings and fantasies. From a behavioral perspective, the popular terms *closeted* or *in the closet* may refer to a person who is actively engaging in homosexual acts but concealing them from others. However, from the retrospective accounts of many gay men and lesbian women, the experience of being in the closet is often more psychologically complex. The feeling of same-sex attraction is often unacceptable and, as a result, unavailable to direct awareness or integration into the individual's public persona. It may therefore be hidden from the self as well as from others. A closeted individual may not act on same-sex attractions at all or may do so only in a *dissociated state.* Dissociated feelings are not unconscious, as in the classical concept of latent homosexuality, but they are out of awareness and only sometimes accessible to consciousness at certain times or in certain situations. Subjectively, a closeted man may tell himself he really does not experience same-sex attractions or, if he is aware of the feelings, he hopes they have some other meaning besides homosexuality. This *psychological state of being closeted* is often reported retrospectively by individuals only *after* they have moved on to a later stage. It is also entirely possible for an individual never to leave this stage.

Homosexually self-aware individuals can acknowledge, at least to themselves, that they experience homoerotic feelings. However, these subjectivities and behaviors may vary from individual to individual. A person may choose not to act on the feelings or may act on them in a secretive way. Some behaviorally closeted individuals are homosexually self-aware, and some are not. Some homosexually self-aware individuals consider the possibility of accepting and integrating these feelings into their public persona. For many individuals, as in Cass's model, this subjectivity may be a normative phase in eventually coming out and affirmatively accepting a gay or lesbian identity. Others may not accept these feelings, as in the example of a homosexually self-aware individual seeking a celibate lifestyle as a way of binding and avoiding a gay or lesbian identity that might prove problematic, not only in the context of one's family, social circle, employment, or religion but may be in conflict with other aspects of an individual's identity. For a clinician to assume of a patient that "She is gay but doesn't know it" is to make an external judgment that may offer little insight into the individual's subjective experience. Such a phrase could describe a closeted woman who is truly unaware of her homoerotic feelings; it could just as well describe a homosexually self-aware woman who knows she has these feelings but is trying to prevent others from knowing she has them; and finally, it could describe a lesbian who is out to a limited circle of individuals, recognizes her own sexual feelings, but is selective about revealing them.

A third subjectivity is that of *nongay/non–lesbian-identified homosexual individuals.* These people are aware of same-sex feelings and may act on them but often identify as heterosexual. They may enter into heterosexual marriages, either hoping that marriage will “cure” their homosexuality or prevent others from questioning their sexual identity. “The Down-Low,” for example, refers to a way of living often associated with urban Black men who reject what they perceive to be an identity associated with White gay culture. While aware of and acting on their same-sex attractions, they are often behaviorally closeted outside the community of other men with whom they have sex and may enter into heterosexual marriages. Some men on the down-low who associate HIV risk with “homosexuality” may engage in high-risk behavior but not see the need for either pre-exposure prophylaxis (PrEP) or HIV testing. Because almost all gay men and lesbian women who come out have gone through a phase of rejecting their own homosexual feelings before accepting their own gay or lesbian identity, it is sometimes assumed that all sexual minority individuals will overcome this period of rejection. However, this is not always the case, and some individuals may never accept either their homosexual feelings or identify as gay, lesbian, or bisexual. These people may seek out ways to change their sexual orientation (see below). In an effort to disidentify with homoerotic feelings and activities, some refer to themselves as either “ex-homosexual” or “ex-gay.”

When conceptualizing sexual identities, they should not be thought of as existing on a developmental continuum or as being associated with more or less psychopathology. It should also be emphasized that this approach to understanding individual subjectivities does not view sexual identities as offering any diagnostic information about an individual. In addition, these subjectivities are not mutually exclusive; there is overlap between them and differing motivations within them. Nor are these sexual subjectivities immutable. Although it may be difficult, if not impossible, to change one’s sexual orientation, it is entirely possible to change attitudes about one’s sexuality or how one labels it. As the subjectivities outlined above are shaped by individual and cultural factors, there may be a wide range of psychosocially constructed attitudes and responses that individuals may develop toward their own homosexuality. Furthermore, the subjectivities outlined above have no exact heterosexual equivalents, as ambivalence about coming out either to oneself or publicly is not intrinsic to forming a heterosexual identity.

LIFE STAGE AND CULTURAL ISSUES

The day-to-day lives of lesbian women and gay men may often be similar to those of heterosexual cohorts. However, gay people face unique developmental tasks throughout the life cycle. For example, during childhood and adolescence, young people who become aware of significant attractions to someone of the same sex will need to understand what makes them different from the majority of people around them. Some will choose to disclose their homosexuality to others through the process of coming out, some will wait until they are older to do so, and some will never do so. During adulthood, and depending upon where they live and work, these men and women will confront unique challenges as they try to establish close relationships and create families. They may do so even without traditional, formal rituals that celebrate their relationships or laws to protect their families. However, they navigate life stages, gay men and lesbian women and bisexual men and women will need to contend both with the adversity of having a stigmatized sexual identity, as well as the potentially life-affirming factors associated with finding one's true self, finding one's true partner, and being part of a community where one can be oneself.

Antihomosexual Attitudes

The terms *antihomosexual attitudes* or *antigay and antilesbian attitudes* are used here to refer to a wide range of critical and disapproving beliefs and feelings about homosexuality, gay men, and lesbian women. Antihomosexual attitudes were once considered normative in most social and institutional settings, including in schools, among the police, in the armed services, and within religious organizations. However, as cultural attitudes have shifted, many traditionally antihomosexual groups and organizations have changed their positions. As a consequence of increasing cultural debates about homosexuality, many high schools and universities now offer clubs—gay–straight alliances—for students of all sexual identities to socialize. Many Western countries allow openly gay and lesbian personnel to serve in their armed forces and police departments. Some religious denominations appoint openly gay and lesbian priests, ministers, and rabbis to serve their congregations. Despite these changes, antihomosexual attitudes are still widespread, and it is helpful to identify their potential impact on gay, lesbian, and bisexual patients.

The term *homophobia* refers to fear or hatred of homosexuality and of gay and lesbian persons (*external homophobia*) or the self-hatred that gay people feel for themselves (*internal homophobia*). All LGB persons, to some extent, contend with *internalized homophobia*—they incorporate disapproving societal views of homosexuality and experience these feelings and beliefs as critical self-evaluation.

The effects of internalized homophobia on the development of one's sexual identity will differ widely at an individual level. Among other things, however, these feelings can lead to the suppression or closeting of awareness of one's own same-sex attraction, impair the acceptance of one's homosexuality, and interfere with the integration of one's homosexual orientation into other aspects of identity. For many gay men and lesbian women, internalized homophobia is experienced as a feeling of devaluation and limitation, not only of their homosexuality but of themselves; for some, it may play a role in the development of psychiatric and behavioral symptoms, including depression, anxiety, denial, and perhaps even suicide.

As the term homophobia gained widespread popular currency, it has been generally, if inexactly, applied to a range of other antihomosexual attitudes. The term has been criticized for suggesting that the existence of antihomosexual attitudes in an individual indicates the presence of a phobia or a formal psychiatric diagnosis. However, cultural attitudes are outside the realm of psychiatric classification. Thus, holding antihomosexual attitudes does not necessarily represent the existence of a mental disorder. Research on homophobic attitudes demonstrated that they are correlated with certain demographic and personality characteristics, including sex, geographical location, religious beliefs, lower education, and a high degree of authoritarianism.

Heterosexism is defined as an ideologic system that denies, denigrates, and stigmatizes any nonheterosexual form of behavior, identity, relationship, or community. This system regards heterosexuality as *normal* and *natural* and treats gay and lesbian people as outsiders who are neither. Heterosexism organizes experience from the point of view that naturalizes and idealizes heterosexuality and either dismisses or ignores a gay, lesbian, or bisexual subjectivity. Heterosexist beliefs can serve to enforce "compulsory heterosexuality" in society. Some common expressions of heterosexism include valuing heterosexual relationships over homosexual ones, an intensely felt opposition to gay and lesbian parenting—regardless of individuals' actual parenting qualities—and the belief that lesbian women and gay men are unsuited for certain professions. Heterosexist attitudes are pervasive in Western cultures where most books, plays, movies, and print and television advertising contain naturalized depictions of heterosexuality. Heterosexism is not necessarily motivated by fear or hatred; in many ways, it is primarily self-referential. However, in the experience of many LGB people, heterosexism is frequently experienced as antihomosexual, which is why it is sometimes confused with homophobia.

Heterosexism does not necessarily imply malice or ill will toward lesbian women and gay men. Nevertheless, in extremis, it may equate the nonheterosexual with the immoral and unnatural and serve to rationalize the condemnation of same-sex relationships and behaviors. *Moral condemnations of homosexuality* treat homosexual acts as intrinsically harmful to the individual, to the individual's spirit, and to the social fabric. Those who condemn same-sex activities believe that heterosexist traditions of philosophical, legal, and religious opposition to homosexuality are, in and of themselves, sufficient reasons to forbid its open expression in the modern world. This reasoning led a U.S. Supreme Court majority to uphold the State of Georgia's right to ban same-sex acts in the 1986 case of *Bowers v. Hardwick.* The justices stated that their decision did "not require a judgment on whether laws against sodomy are wise or desirable" but was based on the "ancient roots" of proscriptions against consensual homosexual activity and a longstanding condemnation of those practices in "Judeo–Christian" moral and ethical standards. In other words, their decision justified the censure of homosexuality in the present because it was censured in the past and argued that antihomosexual laws are permissible because they have always existed. Ironically, more than a decade later, that same statute was overturned by the Georgia Supreme Court in the case of a heterosexual male defendant accused of engaging in so-called unnatural sexual practices with a woman. Then, in 2003, the U.S. Supreme Court made a 6 to 3 historic ruling in *Lawrence and Garner v. Texas* when it found the rest of the country's sodomy laws unconstitutional.

Many contemporary, antihomosexual religious authorities speak of embracing the homosexual person, but not the person's homosexuality.

This antihomosexual attitude is sometimes referred to as “loving the sinner and hating the sin.” It derives from a belief that one *chooses* to be “a homosexual” despite prohibitions against it. As they believe all sinners can be saved, antihomosexual religionists believe they have a religious duty to care about those who make immoral choices. It should be underscored that *choice* is a charged word in moral debates about homosexuality. Similarly, at the other end of this cultural divide, gay people believe they *choose* to act on their homoerotic feelings in order to avoid the choice of suffering from the lack of emotional and sexual fulfillment.

Antihomosexual attitudes—in the form of homophobia, heterosexism, and moral condemnations of homosexuality—shape important aspects of development and day-to-day experience in the lives of lesbian women and gay men. Further, antihomosexual attitudes may lead to *antigay or antilesbian violence.* Even in the absence of overt violence, however, antihomosexual attitudes impinge on the completion of normative developmental tasks in some gay men and lesbian women and may delay or prevent successful adaptation and achievement of an integrated sense of self. If a gay person has been assaulted, the experience may be internalized. Even if they have not been physically assaulted themselves, gay men and lesbian women are sensitized to the ways in which members of the dominant culture, either in positions of authority or on its margins, feel they can make threats and do as they please. In addition, many individuals have had verbal assaults directed against them for being gay or lesbian. There is a growing literature of population studies showing that stigmatization of homosexuality can increase mental health problems in communities.

Coming Out and Meanings of Disclosure

Coming out as a gay, lesbian, bisexual, or other sexual minority person has two components: *coming* out to oneself, which entails recognizing and identifying oneself as a sexual minority person, and *coming out to others.* Coming out to oneself is the subjective experience of inner recognition that homosexual desires, which one may have previously viewed as unacceptable, are part of one’s self. Coming out to others entails disclosing one’s same-sex attractions or sexual minority identity to others and is an ongoing process because people are generally assumed to be heterosexual unless they declare themselves to be otherwise. Thus, one may selectively choose to whom they will come out, and under what circumstances, throughout life: when meeting new people at school, at work, or in day-to-day activities. Initially, one may come out to a select few in the hope of finding social reinforcement and support. Some sexual minority men and women may only come out in certain spheres of their lives, for example, within friendship groups, but remain closeted in other areas, such as in the work setting or within families.

Usually, disclosure of one's sexual minority identity involves either *telling* someone else about that identity or *displaying* some commonly accepted signal of that identity. Adoption of certain styles of dress or behavior, residing in neighborhoods where other gay men and lesbian women live, entering into gay and lesbian communities, and involvement in relationships with persons of the same sex may serve to confirm the identity of a sexual minority person. For some, these activities are signifiers of coming out, performed to signal their sexual minority identity to others. However, it should also be noted that many young people with stable, heterosexual identities often play with gender ambiguity. Nontraditional gender performances are not necessarily a sign of homosexuality, and interpreting them as such may lead to confusion about or an inaccurate assessment of a person's actual sexual identity.

Most empirical studies of coming out describe the ages at which gay men have reported events associated with coming out; fewer studies report on the process for lesbian women. However, those studies show, in general, that women tend to have their first sexual experiences and to come out later than men. Younger lesbian women may experience stronger prohibitions against coming out because of intense expectations of heterosexual dating for women. More recent studies have also suggested that coming out as a sexual minority, much like the age of the first sexual experience, is occurring at progressively earlier ages among adolescents and that there is significant variation in the process based on a variety of factors in addition to sex, such as socioeconomic status, education, and ethnicity. The study of various cultures also demonstrates variation in the coming out process based on the degree of sexual restrictiveness of society at large and the individual's family.

A. is a 25-year-old deeply religious man raised in conservative evangelical traditions. He entered treatment complaining of panic attacks and depression, worried that he might have contracted HIV after a recent, first homosexual experience several months earlier. Although his sexual behavior had not been high-risk, A. was certain he had contracted the virus, and repeated negative HIV tests did not assuage his anxiety. However, he gradually responded to a combination of antidepressant medication, anxiolytics, and twice-a-week psychotherapy.

In the first year of treatment, A.'s sexual identity was rarely discussed as treatment mainly focused on his anxieties about HIV, his job, and interpersonal relationships. Gradually, therapy did address the question of his identity. A. declared he had had sexual relations with a college girlfriend and questioned how could he possibly be gay? He also found it difficult to reconcile the idea of being gay with his religious beliefs. For months, he would go back and forth on the question of whether or not he was gay, and if he were, how could he possibly accept that and stay true to his faith. He would cite moral condemnations of homosexual behavior from scripture and modern religious authorities in an effort to dissuade himself (and the therapist) from the possibility of his accepting a gay identity.

A year into treatment, A. casually told the surprised therapist that he had already decided *he was gay* several months earlier. This private, intrapsychic shift in self-identification represented A.'s coming out to himself. He then offered some personal history he had previously omitted, acknowledging not only childhood sexual fantasies about men but that he had fantasized about sex with men while having sex with his college girlfriend.

Following this, a significant part of treatment focused on A.'s anxieties about coming out to his father, whom he described as a religious, conservative man who would *never* accept a gay son. A. had come out to his mother, who told him not to tell his father, with the warning that "it would kill him." Yet each time he had an argument with his father, A. would contemplate writing a "coming out letter." Asked how his father would react to such a letter, A. acknowledged his wish to use it as a weapon to hurt him. When the therapist asked whether A.'s wish to come out reflected his anger and hostility toward his father rather than a wish for openness and dialogue, he decided not to write the letter.

Eventually, A.'s mother told her husband that their son was gay. He then called A. to say nothing anyone told him would make him stop loving his son. This surprised the patient. The therapist pointed out that if his father was not as unaccepting as he had imagined, perhaps he had projected his own lack of acceptance onto his father. This interpretation allowed the patient to locate the feelings of nonacceptance within himself with which he struggled as he tried to reconcile a growing sense of himself as a gay man.

Coming out is a complicated process that involves both internal and external dimensions, occurs across a lifetime, and may vary in different domains of a person's life. Coming out can mean increased comfort with one's own feelings, and this is integral to social and psychological development. Greater ease in expressing one's feelings, both to oneself and to others, can lead to enrichment of work and relationships. However, even persons who "appear" to be gay or lesbian may have never verbally disclosed their sexual orientation, and others who provide no stereotypical indications of being gay or lesbian may have disclosed their identity to most of their family, friends, and coworkers. Some persons who are fully aware of their homosexuality may never disclose this fact to others. Furthermore, the time frame between awareness of same-sex attraction and self-labeling as gay or lesbian may range from less than a year to decades. In other words, each individual's coming out process is unique and multidimensional.

Race and Ethnicity

Intersectionality refers to how intersections of an individual's social characteristics, such as race, ethnicity, social class, gender identity, and sexual orientation, create interdependent systems of discrimination and advantage. It is increasingly considered part of good clinical practice to view each patient through an intersectional lens. That is, one should consider the impact of each of the patient's multiple social characteristics and identities on their experience, including the types of advantage or discrimination directed toward them by the larger society.

Multiple intersectional factors may interact to compound *sexual minority stress,* that is, the stigma and discrimination experienced by sexual and gender minority individuals with multiple marginalized identities. Theirs is the experience of being a *multiple minority.* For example, the cumulative impact of interactions between racism in the gay community and antihomosexual attitudes within segments of the African-American community may outstrip the coping mechanisms of some Black sexual minority individuals. Finding support within either the Black or gay and lesbian communities may be difficult, and integration of potentially conflicting identities based on race, religion, and sexual orientation may be an important need for persons from these groups. Similar issues may arise for persons from other marginalized racial and ethnic groups. Latinx men and women are often influenced by antigay and antilesbian religious attitudes as well as by traditional, heterosexist gender role, and family expectations. Native peoples' concepts of gender, sexuality, and sexual orientation often diverge considerably from European notions of these constructs and may incorporate beliefs about spirituality as well. Asian Americans come from many different cultural groups with diverse attitudes toward homosexuality and gay, lesbian, and bisexual persons. Thus, although membership in a particular group may assist in predicting family, religious, and social reactions to homosexuality; the diversity of backgrounds and intersecting identities within all groups, whether minority or majority, requires careful assessment to better understand each individual's experience.

The expression of sexual attraction and identity may be strongly influenced by religious, racial, and ethnic identities. A range of identities may be experienced as being in conflict with each other, resulting in such strategies as either maintaining complete secrecy about homoerotic feelings, adopting a bisexual rather than a gay or lesbian identity, or even feeling a need to choose between cultural and sexual orientation identities. For some persons, a conflict between identities may lead to profound confusion and alienation from both groups. The integration of potentially oppositional identities may be extremely difficult to achieve, and the difficulty of this struggle must be appreciated and understood by psychiatrists who work among multicultural gay and lesbian populations.

Religion

Individuals from different religious backgrounds may have had exposure to a wide range of attitudes toward homosexuality, and they may or may not maintain ties to the religion of their families of origin. At a social level, religion influences sexual minority men and women who belong to religious groups that present supportive, tolerant, or disapproving views of homosexuality. Religion also has an impact on all gay and lesbian persons, even those who currently identify as atheists and agnostics, because it contributes to shaping cultural and political attitudes toward homosexuality. Many sexual minority individuals attempt to avoid disapproving or condemnatory religious teachings about homosexuality by eschewing any contact with organized religion; others work within their religious community to try to change traditional, unaccepting attitudes about homosexuality; and still, others stay attached to, while perhaps feeling powerless to change, religions that reinforce their own antihomosexual attitudes. Many gay men and lesbian women have actively sought to reconcile their sexual identity and spirituality within traditional religious groups; some have been successful in doing so. Others have moved to create new churches and spiritual settings that affirm their sexual identities and their relationships. For example, the Metropolitan Community Church (www.mccchurch.org) is an international church that does specific outreach to sexual and gender minority people.

Increasingly, religious organizations within American society have been forced to grapple with the meanings of homosexuality, both as a personal issue for their members and their families and as an issue of ongoing theologic contention. A growing number of religious groups have moved toward positions that either tolerate or actively embrace gay men and lesbian women; some are ordaining gay and lesbian priests, ministers, and rabbis. Pope Francis, when asked about his view of gay priests in 2013, responded, “If someone is gay and he searches for the Lord and has good will, who am I to judge?” Other socially conservative religious organizations have vociferously reaffirmed their traditional opposition to open expressions of homosexuality and criticize the morality of individuals who adopt gay, lesbian, and bisexual identities; they are also critical of heterosexual religious authorities who accept those identities. Regardless of one’s personal relationship to religion or one’s involvement with a specific religious group, sexual minority men and women can be affected by religious teachings and beliefs about homosexuality.

Disapproving religious views of homosexuality often become internalized in sexual minority men and women and may have an effect on their psychosexual development, particularly inhibiting the sexual expression and the development of relationships. Religious beliefs that disapprove of homosexuality constitute substantive, ongoing themes in psychotherapy with some gay, lesbian, and bisexual patients. Many of these themes are intertwined with issues regarding family relationships, particularly in sexual minority men and women who have had strict, fundamentalist, or orthodox religious upbringings. Sexual minority men and women who grow up in communities where social and religious activities are deeply intertwined may stop participating in religious activities and become estranged from their families. They may be deeply concerned about the moral implications of being LGB and about their alienation from their coreligionists. Spiritual development itself may be affected by an inability to reconcile the conflicting internal and external messages concerning the erotic and spiritual aspects of the self. Sexual minority individuals who are able to overcome disapproving, religiously based self-evaluations may nevertheless find it difficult to be comfortable with their sexual identity. Even when faced with denigrating judgments of their sexual orientation, some individuals may not be able to, nor do they wish to, disengage from their religions. Many others struggle to establish a connection to a different, more affirming spiritual community. Conflicts between religion and sexuality are an important focus in clinical work with religious, sexual minority individuals. Followers of the Abrahamic religions may benefit from direction to scholarly works examining the historical contexts and meanings of the putatively antihomosexual scriptures.

Interaction of Sexual Orientation with Stage of Development

The 2011 Institute of Medicine report, *The Health of Lesbian, Gay, Bisexual, and Transgender People,* endorses a life-course framework that acknowledges that events at each stage of life influence subsequent stages and recognizes that experiences are shaped by one’s age cohort and historical context. In general, many of the psychosocial issues confronting sexual minority individuals during different developmental stages are the same as those for heterosexual persons. However, issues specific to being gay or lesbian occur during childhood, adolescence, young adulthood, midlife, and old age. While developmental tasks across the life stages do not take place at the same age for all persons, most sexual minority individuals will have to confront certain tasks during each stage of development.

Childhood. Based on retrospective accounts from adults, children who grow up to be LGB often struggle with feelings of being different from an early age. For some of these children, there is an additional and perhaps related issue of atypical gender role behavior. Feeling different can lead to a sense of alienation that may result in social isolation. Boys who are not stereotypically masculine may be shunned, humiliated, derided, and bullied by their peers in school and other social settings, leading to future difficulties with self-esteem. Girls may be less subject to this process, at least until adolescence, as “tomboy behaviors” in girls are usually better tolerated than so-called effeminacy in boys. Nevertheless, children with gender-atypical behavior are often criticized or otherwise devalued within their own families and sometimes even by mental health professionals. Some parents, believing they might prevent their child from growing up to be LGB, may try to shape developmental outcomes toward more typical gender role behaviors in their children. There is no research supporting the theory that homosexuality can be prevented by attempts to force children to behave in a more gender-typical fashion. Other parents may ignore or distance themselves from children who fail to conform to parental expectations of conventional, heterosexual development. These experiences can lead children who grow up to be LGB to internalize rejecting reactions of others as they come to devalue aspects of their selves that they come to view as unacceptable to others and which they ultimately come to associate with atypical sexuality. When this process is accompanied by incidents of either overt discrimination or violence, the child may suffer lasting impairment in the capacity for psychological and social adaptation.

Adolescence. Adolescence is a particularly vulnerable time for most young people. It is during this period that they must establish a coherent identity, separate from their families of origin, and practice patterns of relating and working that will be further developed in early adulthood. Many LGB youth first become aware of their same-sex attraction during puberty when they begin to mature physically. From retrospective accounts, LGB adults often recall a budding cognitive awareness of their same-sex attractions during adolescence. Many youth, who will subsequently identify as LGB, are involved in heterosexual relationships during this time. However, some of them may enter the early stages of establishing an LGB identity, as they begin to have affectional and physical relationships with persons of the same sex. Some adolescents may go on to develop connections with community structures, such as gay–straight alliances, LGB youth groups, and LGBT community centers. These activities can be either inhibited or facilitated by the reactions of families, relatives, teachers, religious leaders, and peers.

A 2009 adolescent study by Ryan et al. found family rejection of their LGB children increased rates of depression, substance abuse, suicidal ideation, and suicide attempts. Fear of parental rejection may explain why many gay and lesbian adolescents, either consciously or unconsciously, delay some aspect of the development of their sexual identity until they are older. Often, they obtain a greater degree of safety and support by going away to school, moving to an urban environment that allows more anonymity and acceptance, or creating a hometown network of friends who will be accepting of a sexual minority identity. Some adolescents may suppress any awareness of their same-sex attractions; others may defensively attempt to establish a heterosexual identity and relationships, while still others may consciously avoid any outward display of an LGB identity. Less often, but with increased frequency in urban settings, LGB youth may openly identify their sexual orientation at an early age, disclose their feelings to others, and complete significant aspects of LGB identity formation during adolescence. Even outside urban settings, this process has also been greatly facilitated by access to the Internet, where LGB youth can more easily find one another. It remains to be seen whether future generations of young people will follow in the footsteps of an older generation whose members often entered into unsatisfactory heterosexual relationships and eschewed intimate relationships with persons of the same sex out of fear of disclosure of being LGB.

Despite the rapid social changes taking place, there is still an ongoing stigma associated with being LGB. Where there are few adult role models and inadequate resources for support, discrimination, and violence are often directed against visible LGB youth in schools. In the absence of recognition and affirmation from parents and other adult figures in their lives, most pregay and prelesbian adolescents undoubtedly continue to experience significant difficulty in nonaffirming environments. Recent studies show that sexual minority adolescents are at increased risk for anxiety, depression, suicidality, and substance abuse and that bullying between the ages of 12 and 15 years of age was associated with the subsequent presence of anxiety disorders.

Young Adulthood. The period of transition from adolescence to adulthood is a time when many sexual minority individuals have their first opportunity to come out and establish sexual minority identities independent of their families of origin. Sexual minority individuals in their 20s and 30s confront the same developmental tasks as heterosexual peers. These include developing a career, establishing a social identity, developing intimate relationships, and, increasingly for some, raising children. However, for LGB individuals, compared to heterosexual cohorts, these tasks may be accomplished in different ways and at different times than for their heterosexual cohorts. Although heterosexual men and women may be seriously dating, getting engaged, and getting married, closeted sexual minority men and women may date and marry members of the other sex as they struggle to avoid the stigma of homosexuality. Such hiding activities may cause a delay in the consolidation of an integrated sexual identity. Others may sublimate their sexual energies entirely and invest greater efforts in careers. The delays and omissions that follow the secrecy surrounding LGB adolescence can have developmental consequences. For example, some gay men and lesbian women in their 20s and 30s may appear, in some respects, to behave from a developmental perspective somewhat like adolescents. In some urban settings, such individuals form cliques, in-groups, and out-groups with a strong emphasis on style, conformity to standards of dress, a hierarchy of popularity based on looks, athleticism, and affability, as well as opportunities for experimentation with sex and drugs. For some individuals, this “delayed gay adolescence” provides opportunities denied to them when they actually were adolescents: To learn peer and social skills they may not have yet acquired and to experiment with public enactments of their sexual identities. The psychiatrist might view these relational patterns as understandable for those who, in adolescence, had to hide their sexual feelings, behaviors, and identities. Psychiatrists may find it helpful to contextualize risky behaviors related to sexuality, alcohol, and other substance abuse within a developmental framework. During early adulthood, many heterosexual persons marry and have children, options that have historically been much less available to sexual minority couples—although that is beginning to change in places where same-sex marriage is available.

In previous generations, a significant proportion of young adult men and women with predominately same-sex attractions adopted a heterosexual identity and entered into heterosexual marriages; many also had children and raised families. On June 26, 2015, in its ruling on the case *Obergefell v. Hodges*, the U.S. Supreme Court struck down all state bans on same-sex marriage, made same-sex marriage legal in all U.S. states, and required all states to recognize marriage licenses issued in other states. With advances in reproductive technologies and access to marriage equality, the number of same-sex couples having children has grown, resulting in what has been called a *gay baby or gayby boom* (see below).

Many sexual minority men and women have chosen careers that tend to ensure greater acceptance of their sexual identity. Although LGB individuals are employed, whether openly or not, in every type of profession, often they may feel more comfortable working in fields where they can feel comfortable being out. Nonetheless, discrimination in hiring and promotion has been a problem for gay men and lesbian women, while a lack of legal protections created considerable employment-related stress for many of them. Prior to 2020, only 21 U.S. states had legislation prohibiting employment discrimination on the basis of sexual orientation, while the majority of states offered no such protections. In a landmark ruling on June 15, 2020 (*Bostock v. Clayton County*), the U.S. Supreme Court ruled that an employer who fires or discriminates against an individual for being gay or transgender violates Title VII of the Civil Rights Act of 1964, which prohibits discrimination of the basis of sex. The Court said, “If the employer fires the male employee for no reason other than the fact he is attracted to men, the employer discriminates against him for traits or actions it tolerates in his female colleague.” This ruling thus prohibits employment discrimination on the basis of sexual orientation throughout the United States.

In the past, some sexual minority couples may have delayed raising children until midlife due to multiple barriers. In some cases, their children may be from a previous heterosexual marriage, or they may be adopted or conceived by a single sexual minority parent or two people in a committed same-sex relationship. Although attitudes appear to be changing, some of these parents do not receive the same support from their extended families as their heterosexual counterparts. They may often rely much more heavily on friendship networks and paid childcare. Recent access to legal same-sex marriage, together with increased reproductive options, may shift reproduction in sexual minority couples toward younger ages (see below for a discussion of Gay and Lesbian Parents and their Children).

Midlife. For sexual minority men and women, the midlife decades of the 40s and 50s are characterized, as they are for heterosexual men and women, by the consolidation of career identity, concerns about generativity, emerging problems with their own health, and aging and caring for aging parents. LGB baby boomers who are currently in late middle age and older represent the first generation to have openly come out in significant numbers.

Sexual minority individuals share the same midlife health concerns as their heterosexual cohorts. Some, however, may avoid healthcare or receive suboptimal health care if they fear disclosing their sexual orientation to health care personnel. Due to the AIDS epidemic, from the 1970s into the mid-1990s, many gay and bisexual men were confronted during midlife with issues of death and dying associated usually with later stages of life. Increasingly, as HIV treatments have become more effective, many gay and bisexual men find themselves living with a chronic illness that creates its own ongoing uncertainties about one’s health status.

The Older Adults. Individuals 60 years or older have a wide range of experiences in relation to aging; some will develop serious illnesses and become incapacitated, while others will live healthy lives for several more decades. In addition, the commercial worship of youth tends to stigmatize older adults. All older persons need to deal with the possibility of incapacitating illness and the approach of death. Older LGB individuals face additional concerns. For example, they may worry about how they will be treated in assisted living facilities and skilled nursing facilities. In addition, they must confront a widespread stereotype that older LGB people are always lonely and unhappy. Studies, however, have shown older gay men and lesbian women to be as well-adjusted as their heterosexual counterparts. In addition, they may sometimes be better able than younger gay people to contend with the critical reactions to their sexual identities as they may be less concerned about the opinions of others, or they may have come to be more accepting of their sexuality.

The best predictors of psychological adjustment in older LGB individuals appear to be a stable sexual identity and integration into an LGB community. A large majority of older gay men and lesbian women report being happy, deny significant concerns about loneliness or fear of death, and describe good integration into social networks.

Specific problems for older LGB individuals include a lack of adequate resources; reluctance to disclose their sexual orientation to health care and other providers; and an almost total invisibility regarding the needs of older LGB persons, whether single or in relationships, within institutions and agencies that provide care for the aged including nursing homes, supportive living facilities, and hospice care. Failing to recognize LGB people is a function of both the denial of sexuality in older people in general and of an unwillingness to recognize relationships that are not legally sanctioned. Nevertheless, there are a growing number of retirement communities that either cater to older adult LGB people or are "gay-friendly." Social service agencies like Services and Advocacy for GLBT Elders (SAGE, formerly Senior Action in a Gay Environment, www.sageusa.org) have also emerged to provide services for older adult LGBT individuals.

Families and Relationships

Traditionally, gay, lesbian, and bisexual persons usually grew up in families that did not acknowledge or affirm their emerging homosexual attractions or provide them with positive, adult sexual-minority role models. This is gradually changing, as evidenced by LGB adolescents coming out who are living at home and the development of organizations like PFLAG (formerly Parents, Families, and Friends of Lesbians and Gays, https://pflag.org). Nevertheless, not everyone who comes out finds acceptance from their family of origin. In some cases, gay people in relationships do not come out to their families at all, particularly if familial nonacceptance poses a threat to the viability of the couple. Alternatively, many LGB individuals create what has been called *families of choice* or friendship networks. These provide interpersonal bonds, acceptance, adult friendships, and resources unobtainable from their families of origin. Many LGB people go on to form new family structures in their ongoing, romantic same-sex relationships, long-term cohabiting relationships, and single or coparent households with children. Emerging professional literature has begun to describe these couples and families.

Families of Choice. Many sexual minority individuals grew up in families that either ignored, disapproved, or openly denigrated their homosexual feelings and emerging identities; some report significant physical and emotional harm in families that condemned their homosexual orientation. The reactions of parents and other family members to a young person coming out, whether deliberate or inadvertent, can significantly influence the self-esteem and self-acceptance of sexual minority youth. It may irreparably shape the young person's expectations of acceptance or rejection by others outside the family. In the past, LGB youth could expect no more than tolerance within their families; few would receive an active acceptance or empathic support for the experiences and concerns that occur in relation to their emerging sexualities. Greater social acceptance of homosexuality appears to be changing some parents' attitudes toward an increasing acceptance of their LGB children, but this is still a distinct minority position.

Consequently, once most LGB young people achieve sufficient separation from their families of origin, they often create networks of friends and acquaintances who provide them with the support and respect they either lacked or were denied within their families. For a single LGB person, *families of choice* are extremely important; they serve the functions of an extended family. They may help a person pass through developmental crises, are present at celebrations and rituals, and provide a source of comfort and intimacy. In a community that until recently has had little recognition of their primary affectional relationships, these friendships may even be more enduring than romantic relationships, and some even take on greater day-to-day importance than a lover or partner in the life of a particular LGB individual. Understanding and accepting the importance of these friendship networks is essential for the clinician working with LGB people.

Couples. For heterosexual couples, family sanction of marriage provides support and stability. Until recently, same-sex couples often did not find support from their families and, in many cases, even found opposition. This phenomenon is currently shifting toward greater acceptance.

At the time of this writing, same-sex marriage is legal in the United States, Canada, 16 European countries, 7 Latin American countries, Australia, New Zealand, South Africa, and Taiwan. An increasing number of countries and municipalities around the world now legally recognize same-sex relationships through civil unions or domestic partnerships. The 2020 U.S. census reported that 0.8% of all U.S. households are same-sex and that of the 980,000 same-sex households, 52% comprised married partners. Among both married and unmarried same-sex households, there were more female than male couples.

Gay and Lesbian Parents and Their Children. Many lesbian women have conceived, and many gay men have fathered children. In the past, this often took place within a heterosexual marriage that ended when a closeted gay or lesbian partner came out. However, the 1990s saw the start of a *gay baby boom* in which single or coupled gay men and lesbian women began raising children they had either adopted or conceived using reproductive technologies that included donor insemination (using known or unknown donors) or surrogacy. Consequently, in recent years there has been greater visibility of families with gay and lesbian parents, and many younger gay men and lesbian women today seriously consider the option of having children. It is estimated that anywhere between one and fourteen million children live in households headed by same-sex couples.

Numerous studies describe the characteristics of gay and lesbian families and the impact on children of having a gay or lesbian parent. This research was initially undertaken in the late 1970s to evaluate societal concerns and judicial beliefs about the presumed harm done to children who had been raised by gay and lesbian parents. For example, it was not uncommon—and in some states, it is still common—to award custody of children to a nongay parent in a divorce if the other parent comes out as gay, lesbian, or bisexual. These custody decisions were not based on the results of empirical studies but on the belief that LGB parents were either by nature unfit or would influence the eventual sexual identity of their own children. However, the vast preponderance of relevant research strongly suggests that having a lesbian or gay parent has no adverse psychological consequences for children. Nor does having a gay or lesbian parent influence the eventual sexual orientation of children when compared with the children of heterosexual parents. What studies do show is that children with two parents do better than children with one parent, regardless of the sexual orientation of the parents.

LGB parents and their children also have to confront the fact of being different from heterosexual families and may experience stigma or antihomosexual bias. Historically, the ability of both same-sex parents to be legally recognized as parents was prohibited in many states. Consequently, the children in families with same-sex parents were often denied parental benefits from the partner with no legal standing, including clear inheritance rights, death benefits, or the right to be covered by a nonbiologic parent’s health insurance. In response to the needs of these children, in 2001, the American Academy of Pediatrics issued a position statement in support of legal adoption by a second same-sex parent. The APA issued a similar position statement in 2002.

Currently, when a legally married couple has a child, they are both automatically presumed to be legal parents. This presumption applies to same-sex parents when children are born to couples who are legally married or where their state recognizes their civil union or domestic partnership at the time of birth. While most states have not addressed this issue directly, every state that has addressed the issue in its highest court has held that the marital presumption and assisted reproductive provisions apply equally to opposite and same-sex couples. There has been considerable debate regarding the psychological and other quality of life outcomes of the children of same-sex parents. A recent review of 79 scholarly articles on the topic concluded, “Taken together, this research forms an overwhelming scholarly consensus, based on over three decades of peer-reviewed research, that having a gay or lesbian parent does not harm children.” Of the 79 studies reviewed, only four reported that offspring of same-sex couples fared worse.

CLINICAL APPROACHES

Although the psychiatric issues of LGB individuals often resemble those of the general population, there are unique aspects to the clinical work with this patient population.

General Considerations

As many sexual minority patients report having encountered disapproval, discrimination, insensitivity to their concerns, and even abusive treatment within clinical settings, it is not uncommon for them to avoid medical care and to be reluctant about disclosing their sexual orientation identities to clinicians. They may be unwilling to enter treatment without some clear demonstration by the clinician of a nonjudgmental and accepting attitude.

Inclusive and Special Clinical Settings. Sexual minority men and women present for treatment in the same inpatient, outpatient, and residential clinical settings as other patients. In recent years, specialized clinical units and services have been developed, both in the United States and other countries, for working with sexual minority persons. These may include community health centers intended to serve the needs of the local LGBT communities, not-for-profit and for-profit clinics, state and municipal government-supported clinics, and specialized services geared to particular problems like HIV infection or substance abuse. Whether services are provided in general or specialized settings, it is always to the patients' benefit when staff members are trained to deliver culturally sensitive care to sexual minority patients. In fact, multiple studies show that lack of cultural sensitivity in taking a medical history together with lack of familiarity with health matters of particular concern to sexual minority patients contribute to their avoidance of care.

The clinical setting and staff with whom the patient comes in contact are often important determinants in creating a climate of inclusiveness for sexual minority patients. Explicit and implicit communications of acceptance in these settings can ensure successful psychiatric treatment. For example, the language and behavior of psychiatrists and other staff should convey that they do not automatically presume that all individuals, couples, or heads of families are heterosexual (*heterosexism* or *heteronormativity*). Questions about relationships should be open-ended and nonjudgmental when inquiring about the gender of a patient's partner. It should not be assumed that all marriages are heterosexual ones or that married patients are sexually exclusive with their spouses. Questions regarding sexuality should distinguish between desire, behavior, and identity, with the recognition that, especially in younger patients, desires, and behaviors may be more easily described than identity. Many sexual minority men and women report a history of negative experiences in the health care system and may, therefore, enter the clinical setting with a sense of fear and trepidation. The presence of gay-friendly publications or even a rainbow flag in a waiting area can help put sexual minority patients at ease. In addition, self-reporting forms for patients to fill out that neutrally inquire about diverse sexual orientations and family arrangements can reassure LGB persons—as well as those uncertain about their sexual orientation—that they will be treated with respect in this clinical setting and that they can talk openly about their concerns. Since 2011, The Joint Commission (TJC) has required that health care organizations have an LGBT nondiscrimination statement for accreditation to receive Medicare reimbursement. TJC encourages LGBT-inclusiveness in its publication, *Advancing Effective Communication, Cultural Competence, and Patient- and Family-Centered Care for the Lesbian, Gay, Bisexual, and Transgender (LGBT) Community: A Field Guide* which also describes how to create an LGBT-inclusive clinical environment.

Another important strategy in clinical work with LGB patients is to routinely include, when appropriate, references to significant friends and members of nontraditional families during evaluation and treatment. When necessary, same-sex partners and children in these families should be involved in the care of LGB patients. For example, it may be important to obtain corroborative information from partners, to identify potential sources of a patient's problems within the context of the couple or family, or to include a partner in maintaining adherence. In addition, issues of confidentiality and decision-making in working with heterosexual families should also apply to working with gay and lesbian families.

In general, physicians should be conversant with advance directives such as the durable health care power of attorney and the appointment of health care proxies within the states in which they are licensed to practice. These issues may be particularly relevant when working with sexual minorities and alternative families. Historically, same-sex couples did not have marriage rights and found other ways to protect their relationships and families of choice. Even today, some committed couples have not availed themselves of the right to marry. In cases where one of the partners becomes incapacitated, courts may grant important medical decision-making rights and privileges to biologic family members and exclude the wishes of long-term partners. Clinicians should address this important issue with their sexual minority patients.

The Sexual Orientation of the Psychiatrist. The issues of the psychiatrist's sexual orientation and of disclosing this orientation to the patient are complex. The heterosexist assumption that everyone is heterosexual usually means that a psychiatrist's sexual orientation does not come up in treating most patients. However, this assumption does not hold when treating sexual minority patients. A heterosexual psychiatrist trained to be nondisclosing may refuse to answer questions about their sexual orientation and turn the question back on the patient. In general, this might be an error for many sexual minority patients. Not disclosing could be interpreted by a patient as meaning that the psychiatrist is not entirely comfortable speaking about themselves or about homosexuality in general, that the psychiatrist might be closeted, or some other meaning.

In the cases of sexual minority patients working with psychiatrists who are openly LGB, there are several possible benefits and risks to self-disclosure. The benefits of working with a psychiatrist of the same-sexual orientation may include an easier rapport; greater efficiency resulting from shared knowledge about experiences associated with being LGB; if the therapist is older, an ability to share knowledge about the gay community with a younger patient; and affirmative role modeling if the therapist has established a secure LGB identity. Possible risks for such therapeutic dyads include unwarranted assumptions that the patient and psychiatrist share experiences or personal characteristics based on sexual orientation identity alone; reducing the focus of treatment to sexual orientation identity issues alone; transferential or counter-transferential overidentification of the therapist or patient with each other; or shared, unconscious collusion to avoid discussing painful events or affects associated with LGB development.

How an LGB psychiatrist's sexual orientation identity is disclosed to a patient may vary. Sometimes the patient has requested to work with an openly LGB therapist and has obtained this knowledge before beginning treatment. In other cases, the patient learns this information after treatment has begun, either as the result of self-disclosure by the psychiatrist or through some outside source. When self-disclosure is involved, the timing of the disclosure may have a significant impact on facilitating or inhibiting the work in therapy. In some cases, it may have no effect at all. It is useful for the LGB psychiatrist to be able to identify those individuals and the clinical situations in which working with an openly LGB therapist might be most helpful.

In general, the psychiatrist's sexual orientation identity should not be the most significant factor in determining clinical outcomes. Any therapist working with LGB patients needs to maintain a conscious awareness of their own antigay attitudes and heterosexist assumptions; possess a reasonable knowledge base about LGB issues; apply gay-sensitive and gay-affirmative approaches to psychotherapy; and obtain adequate supervision, peer supervision, or outside consultation when necessary. Having an openly LGB identity is no substitute for good training, and LGB mental health professionals should constantly strive to improve their knowledge base and clinical skills.

Sexual orientation may be far less significant than other personal characteristics in determining the effectiveness of a psychiatrist in working with sexual minority patients. A gay or lesbian therapist who has not fully come out or established a comfortable identity may have trouble working with patients struggling with similar concerns. Regardless of sexual orientation identity, a therapist who is uncomfortable talking about sexual issues will have more trouble being helpful to LGB patients with such concerns. Ultimately, the degree of empathy experienced and expressed by the therapist for the patient will, in most cases, be an important factor in determining the effectiveness of treatment.

Specific Clinical Issues

There are clinical concerns that are unique to sexual minority individuals, some of which are particular to their sexual orientation identity. The presentation of other concerns that do not directly relate to sexual orientation identity may also be shaped by the fact that a person is gay, lesbian, or bisexual, and as noted above for some persons being a member of a sexual minority group may interfere with, and even prevent, interactions with the health care system.

Societal and Domestic Violence. The problem of *antigay* or *antilesbian violence* is widespread. In 1990, the U.S. government passed a law ordering a study of hate crimes, including attacks on gay men and lesbian women. Because of concerns about antigay violence at the time, the law's passage was the first time the federal government made any affirmative mention of sexual orientation.

One of the most common forms of violence is *gay-bashing:* Gangs of young men descend on neighborhoods or gathering places where gay people meet. Armed with bats, clubs, or other weapons, they attack anyone they believe to be gay. The most sensational case of antigay violence in the United States was the 1998 death of Matthew Shepard, a young man from Wyoming, robbed and killed by two other young men because they perceived him to be gay. The trial of his killers, and the increasing publicity surrounding gay-bashing incidents in general, provoked wider discussion of previously unspoken cultural beliefs. For example, some perpetrators of violence have put forward a *gay panic defense,* arguing that violence is an excusable response by a heterosexual man to an unwanted sexual advance from a gay man. To date, no court has accepted this defense, and in 2014, California became the first U.S. state to ban the use of this defense (as well as the *trans panic defense*) in murder trials. Since that time, 15 other U.S. states have legislated similar bans, and efforts have been made to introduce a similar ban at the federal level.

LGB individuals who have been the objects of violence, or sexual or domestic abuse, may have difficulties in reporting their experiences and seeking help; they often fear disclosing their sexual identities or receiving judgmental responses—and even harassment—from police or social service workers. In other words, antigay violence is presumed to be an under-reported phenomenon because its victims are often reluctant to come forward and face reprisals from the very authorities to whom such incidents are reported. It has also been argued that antigay violence is the inevitable outgrowth of the unexamined beliefs associated with homophobia, heterosexism, and moral condemnations of homosexuality. According to this argument, those who marginalize homosexuality may be either directly or indirectly responsible for the brutality aimed at LGB people. Inevitably, LGB people may have legitimate concerns about potential physical attacks. The problem of violence directed against them is widespread and well-documented and ranges from verbal abuse to homicide. Violence against LGB people may be committed by strangers, family members, gangs, fellow soldiers within the military, and sometimes by authority figures, such as the police.

Although some therapists believe that childhood sexual abuse causes homosexuality, studies show no causal connection between the phenomenon and the adult development of a gay, lesbian, or bisexual identity. Studies show that sexual abuse of sexual minority individuals during childhood appears to occur at about the same rate as in the general population. Studies of domestic violence and abuse in lesbian and gay relationships demonstrate that physical abuse occurs at about the same rate as in heterosexual relationships.

The impact of violence on sexual minority individuals, as on other victims of violence, includes post-traumatic stress disorder, depression, and a range of other psychological and emotional problems. These include feelings of inadequacy, embarrassment, or shame. In addition, there may be specific symptoms associated with being part of a stigmatized minority group. For example, the victims of antigay or antilesbian violence may hold themselves responsible and believe that their homosexuality caused the assault. The experience of violence can lead an individual to deny a gay or lesbian identity or to avoid other LGB individuals. Depending on the circumstances, the treating psychiatrist should be prepared to respond to the immediate crisis following an incident of abuse or violence, to the potential long-term effects of these events on the individual; and to the ongoing concerns of all sexual minority individuals about potential violence being directed against them.

The psychiatrist can be helpful by recognizing the influence of exposure to antigay violence on the personalities and experiences of gay, lesbian, and bisexual persons and carefully identifying the intrapsychic and interpersonal wounds they may produce. It is also reasonable for psychiatrists and other clinicians to acknowledge the everyday dangers that sexual minority individuals face or may have faced. A psychiatrist can be most helpful by being sensitive to patient anxieties regarding antihomosexual attitudes, as LGB patients can be retraumatized when clinicians deny or minimize the extent to which the phenomenon colors their lives and affects their self-esteem. It should be noted that in many cultures, antihomosexual attitudes are ubiquitous, and no one is free of them, not even psychiatrists. This obliges clinicians to understand and be aware of the ways in which these attitudes may have affected the development of their own identities and belief systems. Such awareness can help psychiatrists to come to a greater empathic appreciation of the impact of antihomosexual attitudes on the gay, lesbian, or bisexual patient they are treating.

Suicide. Determining the rate of suicide attempts and completed suicide among LGB persons is difficult. Factors complicating this research include the reluctance of many individuals distressed about their same-sex feelings to discuss them with mental health professionals, inadequate assessment of sexual orientation in most health care settings, and the absence of information identifying sexual orientation on death certificates. As noted in the aforementioned report of the Institute of Medicine, lack of disclosure of homosexuality poses a greater problem for older compared to more recent studies as social acceptance of homosexuality has increased.

In spite of methodologic difficulties, studies consistently suggest that gay, lesbian, or bisexual youth have an increased risk for attempted suicide. With few exceptions, the increased rate of suicidality among LGB youth in comparison with heterosexual youth is consistent across age groups (i.e., middle school, high school, and young adult populations), gender, race/ethnicity, and differing definitions of sexual orientation (i.e., same-sex attraction, self-identification, and behavior). Suicide rates among gay, lesbian, and bisexual youth may vary on the basis of other risk factors, such as mental illness, substance abuse, rejection by family, and perhaps, degree of gender role nonconformity. For all age groups, an increased risk of suicide may accompany struggles with internalized homophobia, problems with coming out, and difficulty in establishing a firm gay, lesbian, or bisexual identity. At least one study also found suicidal ideation and behavior to vary in sexual minority individuals as a function of race and ethnicity. Compared to Whites, Blacks, and Latinos were more likely to report a history of a suicide attempt; however, they did not evidence increased rates of mental illness. Most of the suicide attempts came at an early age and were interpreted as probably coinciding with the coming out period and likely to have been associated with increased social disapprobation within the Black and Latino communities.

Determination of the exact numbers of gay, lesbian, bisexual, or uncertain people who attempt or complete suicide may be less important than understanding the characteristics that contribute to a risk for suicide. For gay or lesbian youth, enforced invisibility and a lack of adequate resources to address their health and mental health needs can prevent adequate assessment of their concerns or interfere with developing specialized programs for prevention and treatment. Unlike other minority youth, gay, lesbian, and bisexual youth usually do not share their minority status with either parent and consequently may not readily turn to their parents for support. In the United States, this problem has been exacerbated by a socially conservative political climate in which efforts to identify the mental health needs of gay, lesbian, and bisexual individuals are sometimes labeled as an “endorsement” of homosexuality. This has hindered public funding and research efforts to study suicide among gay, lesbian, and bisexual young people, as well as efforts to provide them with social support.

In general, any young person in distress should be carefully evaluated for concerns about sexuality and sexual identity. The psychiatrist should take pains to ensure that assessment and treatment do not place young people at further risk of suicide if and when they disclose their sexual orientation to others. For example, premature coming out, sometimes precipitated by a conversation with a mental health professional, can increase the possibility of rejection by peers and family. Moreover, all physicians should be aware that many parents are hostile to sexual minorities and should assess the risk that the family's attitudes may pose to the patient. These risks may include verbal and physical abuse, subjection to damaging sexual orientation conversion efforts (SOCE), rejection of the child, and even homelessness. Appropriate steps must be taken if the parents cannot provide a safe environment for their child. The potential risk of suicide for persons of any age who are struggling with coming out and establishing a gay, lesbian, or bisexual identity should be recognized. Clinical services appropriate for managing suicidal feelings and behaviors and for providing support to suicidal persons should be available as needed.

Alcohol and Substance Abuse. In response to the 2011 call for the inclusion of sexual orientation and gender identity data in federal surveys, there is now data on alcohol use for sexual minority men and women ages 18 and older from the National Survey of Drug Use and Health (2015 to 2017). The survey includes data from 126,463 individuals, including 8,241 (6.5%) participants identifying as sexual minorities. With respect to alcohol, a disparity emerged by age 25 when alcohol use declined in heterosexual participants but increased in sexual minority men and women until age 50. Gay men ages 18 to 23 had the highest prevalence of alcohol use disorder, while bisexual women had the highest disparity ratio relative to the heterosexual women comparison group.

Another study also employed NSDUH to assess past-year cocaine use, crack cocaine use, heroin use, methamphetamine use, OxyContin misuse, and lifetime use of chemsex drugs (*chemsex* refers to the use of sex-enhancing drugs among MSM). Sexual minorities showed higher odds of past-year substance use and lifetime chemsex drug use. Bisexual women had higher odds for the use of all analyzed substances relative to heterosexual women. Otherwise, being older and being a woman were protective factors. Urbanity, being uninsured, and an unmet need for mental health treatment were associated with significantly higher odds of substance use, chemsex drug use, and binge drinking, established between drug use and health indicators with higher odds of drug use for lower health ratings. Sexual minorities experienced significantly higher levels of socioeconomic vulnerability, which was, in turn, associated with increased odds of drug use. Other recent studies have reported that all sexual minority subgroups are more likely to be tobacco smokers than their heterosexual peers.

Some attribute this increased usage to the psychosocial stressors associated with being LGB. In many cities, bars and clubs where one can enter freely and anonymously are often the first and only point of entry for some individuals into the wider gay and lesbian community.

Recreational or “club drugs” are popular among some segments of the gay and lesbian community. Among these are MDMA (3,4-methylenedioxy-*N*-methylamphetamine), colloquially referred to as Ecstasy, which can induce depression and panic attacks in those who use it. Ketamine, also known as K or Special K, is a dissociative anesthetic that induces feelings of unreality and may cause catatonia. Gamma-hydroxybutyrate, known as G or GHB, is a naturally occurring biologic substance whose effects mimic alcohol. In modest doses, it can cause sleep or coma, and some cases have been fatal in overdose. The psychiatrist working with gay men and lesbian women involved in the club scene or circuit parties should be aware of these particular substances, the lifestyle choices associated with their use, and their potential risks, including potential interactions with prescribed medications.

There are numerous specialized programs, both inpatient and outpatient, that simultaneously address substance use problems while respecting the sexual identities of gay, lesbian, and bisexual persons. There are also 12-step and other self-help programs that provide special groups for gay men and lesbian women and that allow for a more comfortable and open discussion of both sexual identity and substance use issues.

The psychiatrist should be prepared to identify a potential overlap between problems with acceptance of one’s sexual orientation identity and substance abuse as well as the association of substance use with many sexual and gender minority venues. Patients with addiction problems, gay, lesbian, bisexual, or otherwise, sometimes show signs of either a poorly integrated or a self-loathing identity. Such persons may exhibit denial, fear, anger, guilt, helplessness, hopelessness, dishonesty, low self-esteem, self-hatred, and social isolation and alienation. Assessment of these traits and an awareness of how addiction and internalized homophobia may reinforce each other are helpful in providing adequate treatment. Furthermore, it may sometimes be difficult to treat substance abuse problems in gay, lesbian, and bisexual patients without addressing concomitant sexual identity problems. As a result, substance abuse programs that are both gay-sensitive and gay-affirming work best for treating substance abuse disorders among gay men and lesbian women. Prevention or reduction of substance abuse among this population must include the development of settings for socializing without alcohol or drugs.

Talking Sex. As psychiatrists and other health care professionals are often uncomfortable talking about sexuality in general, dealing with the specific sexual behaviors and concerns of gay men and lesbian women may be particularly difficult. Problems may result from either a psychiatrist's unfamiliarity with or disapproval of same-sex behaviors or both. Given the inadequate training in human sexuality in most training programs, it may be difficult for the psychiatrist to discuss the sexual concerns of LGB patients in an objective and comprehensive manner, to be aware of special diagnostic issues, and to approach sex therapy with these men and women. These concerns might include difficulties performing anal intercourse in a committed couple or concerns about HIV infection among gay men, and orgasmic difficulties among lesbian women. Some gay men with HIV develop low serum testosterone and may seek treatment for loss of sexual drive. Some gay men may struggle with issues of sexual compulsivity, and bisexual men and women may have difficulty mastering safer sex practices with one or both sexes.

Patterns of same-sex sexuality have been shown to be different from those in heterosexual couples, with female couples having decreased sexual interaction and male couples initially having increased amounts of sexual interaction in comparison with heterosexual couples. Typologies of male couples have sometimes described them in terms of their degree of sexual exclusivity and openness, and some male couples present with problems related to sexual exclusivity. In contrast, female couples present more often with problems related to decreased sexual desire. There also appear to be generational differences in openness of relationships, which may be attributable to the growing accessibility to legal protections for same-sex couples via marriage or civil unions.

The psychiatrist working with the sexual problems of LGB individuals and couples should consider organic, developmental, interpersonal, and social causes. Assessment and treatment can adapt traditional, heterosexually oriented paradigms of sexuality to include consideration of issues related to LGB development and to the dynamics of same-sex relationships. Special attention should be given to cultural factors that may contribute to sexual problems among gay men and lesbian women. These include a lack of readily available and accurate information about same-sex practices as well as the inevitable internalization of judgmental societal attitudes about homosexuality.

Medical Illness. Although often presented with the same medical problems as the general population, the Institute of Medicine found that lesbian women and gay men experience unique health disparities as well as barriers to equitable health care. One such barrier is psychosocial: Gay men and lesbian women often have to deal with judgmental attitudes from health care personnel which may lead some to avoid or delay seeking care. This may not be as great a problem in some urban settings with large LGBT communities where clinicians, either in private practice or public settings, have experience in caring for lesbian and gay patients. However, even in large urban settings, not all LGB patients have access to health care personnel who are sensitive and knowledgeable about their concerns or their increased risk for some diseases.

There are no gynecologic problems that are unique among lesbian women as opposed to among heterosexual women. Lesbian women, however, may be at greater risk for undetected disease as studies show they may undergo fewer pelvic examinations and Papanicolaou (Pap) smears and mammograms. Sexually transmitted diseases are less common among women who are sexually active only with other women, and there are few cases of HIV infection in these women. Safer sex counseling and screenings must be tailored to the individual’s sexual practices as revealed in taking a complete sexual history. The risk for some types of cancer in women may be affected by factors such as a history of sexual intercourse with men, the number of pregnancies, age at first giving birth, and breastfeeding. Women who have not had children and have not breastfed, have not used oral contraceptives, and are older when they first give birth are at increased risk of breast cancer. Historically, lesbian women have had a lower number of pregnancies than heterosexual women. However, an increasing number of lesbian women have children using reproductive technologies, including anonymous or known donor insemination.

In general, when compared to heterosexual men, gay men who engage in unsafe sexual practices have increased rates of sexually transmitted diseases and HIV infection. Since the 1980s, the influence of HIV infection and AIDS on individual lives and the communities of gay men has been enormous and has affected almost every aspect of the personal and public experience of being gay. Public health campaigns to reduce infection among gay male populations appeared to work for some time in reducing the rate of infection. However, as younger gay men enter the community, educational strategies tailored to older gay men may not reach these new arrivals, and public health responses to their needs have not always been effective. For example, some young gay men may believe that only older gay men are infected with HIV and that they are safe from HIV infection if they only have sex with men their own age. There is also the phenomenon among some gay men of engaging in frequent *barebacking* or *unprotected anal intercourse*, and the rare phenomenon of *bugchasing* when an HIV-negative individual knowingly barebacks with a partner known to be positive. Ignoring public health warnings about HIV has been attributed to different causes, depending on the individual, such as lack of information, taking calculated risks about who is and is not infected (*serosorting*) "AIDS fatigue," and an unwillingness to continue using condoms. Public health measures in this area remain an issue of great concern.

HIV infection—either being infected or knowing someone with the virus—is a potential problem in the life of any gay man who enters the health care system. Health issues related to HIV include testing for infection; ensuring early treatment intervention in infected individuals; preventing further HIV infection; responding to reasonable and irrational fears about contracting the disease; dealing with the chronic stress resulting from either multiple losses, illness, or both; and caring for someone else who is infected. For many gay men, an association of HIV infection with one's sexual orientation identity can exacerbate one's own internalized homophobia.

In July 2012, the Food and Drug Administration (FDA) approved a combination of emtricitabine and tenofovir (trade name Truvada) for PrEP against HIV infection for individuals at high risk. The Centers for Disease Control and the WHO have published guidelines for PrEP in at-risk populations. These guidelines include the use of condoms in addition to PrEP as some men in clinical studies seroconverted despite apparent adherence to the drug regimen. The message to always use condoms has not proved effective in ending the AIDS epidemic; however, programs including PrEP have the potential to do so. In 2014, New York's governor issued a plan to end the epidemic in his state by 2020, specifically to reduce the number of annual new infections from 3,000 to 750. Based on the number of new infections in 2019 (1,700), the goal for 2020 was adjusted upward to 825. The plan includes PrEP, universal HIV testing, and a connection to care for people living with HIV since modern treatments have the capacity to lower viral loads to the point that the risk of transmitting the virus is low. Data from the National HIV Surveillance System suggest that treatment as prevention is being grossly underutilized. Surveillance data suggest that less than half of MSM who are living with HIV are receiving PrEP, although 90% are aware of it. The goal of universal HIV testing and connecting those who test positive with treatment is clearly a public health priority. Integrating antiretroviral therapy for HIV-positive partners and PrEP for HIV-negative partners nearly eliminates the risk of HIV transmission in serodiscordant couples. A newer treatment, postexposure prophylaxis (PEP) antiretroviral regimen initiated within 72 hours of known exposure to HIV, can prevent infection and seroconversion.

There are generational differences. Gay men under 30 years may be untouched by HIV, having never known anyone infected. Those gay men who were out in the 1980s and 1990s have been affected by the epidemic. Among other things, the psychiatrist can play an important role in differentiating functional and organic symptoms that occur in HIV-infected patients, for example, diagnosing the clinical depression that may occur as a result of low testosterone levels, which occur in 30% to 50% of symptomatic HIV infections. Psychiatrists may also be called on to treat any psychiatric sequelae of the disease, including depression, anxiety disorders, psychosis, and dementia. Bipolar disorder is also seen in HIV infection. Control of coexisting psychiatric disorders is crucial to the patient’s adherence to their HIV medication regimen.

Recent advances in treating HIV infection have changed its course from an almost universally fatal diagnosis to a chronic medical condition. However, there are some small, inconclusive studies showing aging-related health problems associated with HIV, such as osteoporosis and chronic obstructive pulmonary disease.

Physical illness in gay men and lesbian women may interact with experiences associated with their sexual identity. In particular, avoidance of healthcare due to fear of stigma in healthcare settings is believed to be a major contributor to sexual minority health disparities. Lifestyle factors, family structures, and social networks can influence the extent and type of support available. The stress of being ill may evoke psychological reactions that affect capacities for intimacy, affection, and sexuality.

Psychotherapy

Gay and Lesbian Affirmative Psychotherapies. The 1970s saw the introduction of literature on *gay and lesbian affirmative psychotherapy.* This approach, a response to the historical pathologizing of homosexuality, was described by Alan Malyon as a "theoretical disposition [that] regards homosexuality as a nonpathologic human potential. The goals of gay-affirmative psychotherapy are similar to those of most traditional approaches to psychological treatment and include both conflict resolution and self-actualization." LGB-affirmative psychotherapy recognizes the centrality of sexual attraction, behaviors, and identity in a person's life but also appreciates that sexual orientation is only one aspect of personality and, depending on the patient's needs and goals, may or may not be a central focus of therapeutic work.

Acceptance, acknowledgment, and affirmation of gay, lesbian, and bisexual persons are essential elements in psychotherapeutic work with sexual minority men and women. These therapeutic stances reflect the therapist's regard and empathy for the patient's sexual feelings and for their same-sex relationships. The therapist's attitudes are meant to offer the patient a contrast to the social denial and denigration that the patient may have experienced in the past. The psychiatrist working with LGB patients should maintain a respectful position that will both help to contain the struggles of the patient and at the same time serve to convey a recognition of the ongoing adverse impact of social prejudice against gay and lesbian persons. The clinician's lack of awareness of or inability to explicitly acknowledge the effects of discriminatory interpersonal and social forces working against the expression of same-sex feelings and desires may suggest to the patient that the psychiatrist agrees with such attitudes.

As the sexual, romantic, and affectionate expressions of sexual minority patients have often been either ignored, ridiculed, or rejected, for many gay, lesbian, and bisexual individuals, speaking about them for the first time with a therapist may represent a unique opportunity. Feeling listened to without judgment, having one's feelings acknowledged as real by a person in authority, and having them affirmed as commonly felt and normal are powerful affirmations. The therapist's ability to tolerate, accept, and validate the patient's feelings allows the patient to express their own antihomosexual feelings and doubts, including fears about the critical reactions of others, disapproving feelings about being gay or lesbian, and regret over losses the patient may have sustained after coming out as gay or lesbian.

Contemporary Psychoanalytic Approaches to Gay and Lesbian Patients. Until the late 20th century, psychoanalytic approaches to LGB individuals focused on homosexuality first as a developmental arrest and then as a “neurotic” symptom. Starting in the late 1970s, newer psychoanalytic approaches emerged. Rather than trying to get homosexual patients to conform to heterosexual norms, these newer approaches see the therapist as an agent of the sexual minority patient.

Analytic work with sexual minority patients has drawn attention to aspects of the therapy process sometimes overlooked in doing psychotherapy with nongay patients. As a result, this approach offers insights into some general principles of psychoanalysis, as well as basic psychoanalytic beliefs and practices, including the nature of the psychotherapeutic frame; the values and risks of therapist self-disclosure; the limitations of psychoanalytic data to support theories of etiology; how adherence to theoretical preconceptions will restrict or inhibit a therapist’s clinical listening; experience-near versus experience-distant responsiveness from the therapist; the therapist’s embeddedness in cultural preconceptions and how they have an impact upon treatment; the role of the patient’s and the therapist’s subcultural identities and how they coconstruct narratives in treatment; and the meanings and therapeutic uses of countertransference. Developing one’s own therapeutic stance in treating sexual minority patients depends on specialized training in psychotherapy, continuing education, the therapist’s own personal analysis, and ongoing self-analysis. Furthermore, the therapist’s stance does not just entail what one knows; it also includes what the therapist does not know and requires an ability to allow a dynamic interplay of knowing and not knowing in the patient, in the therapist, and in the transitional space between them.

An analytically oriented therapist aims to create a therapeutic holding environment, which is a space where all of a patient’s feelings and ideas are allowed to emerge. As in sexual minority affirmative psychotherapy, the analyst’s job is to remain open to hearing the patient’s complaints, sexual or otherwise. In the atmosphere of the holding environment, meanings are discovered, created, and deciphered. All patients, not just sexual minorities, can benefit from a therapeutic holding environment based on respectful principles. Respect for the patient is essential. But by itself, however, it is not enough. The subject of homosexuality often evokes uncomfortable feelings and denigrated meanings. When these emerge in treatment, they certainly need to be tolerated and respected by the therapist. In addition, a prerequisite for doing psychotherapy with sexual minority individuals is for therapists themselves to be able to accept their patients' homosexuality as a normal variation of human sexuality and to value and respect same-sex feelings and behaviors as well. The clinician who automatically regards a homosexual or bisexual orientation as a symptom of some other clinical problem is not engaging in respectful behavior toward the patient. It should again be noted that for some LGB individuals, being treated with such respect by a therapist is a novel experience.

In addition to respect, a contemporary psychoanalytic approach to LGB patients focuses on the meanings, rather than the origins, of human sexuality. The therapist who searches for the causes of homosexuality or bisexuality will ultimately mistake its meaning to the patient for something else. Alternatively, this approach shifts its focus to the affective meanings of the language of the patient, as well as those of the therapist, and to their transferential and countertransferential implications. From this perspective, it is important to understand there are consequences to a therapist's taking sides in a sexual minority patient's interpersonal struggles with family and friends. Everyone a patient discusses should be treated respectfully, including those of whom the patient expresses disapproval. This therapeutic stance allows for a wider exploration of a patient's feelings and identifications with those who accept the patient's sexual identity and those who do not. Patients may try to resolve inner conflicts about being gay or lesbian by selective inattention to (dissociating from) their own antihomosexual identifications. This is sometimes seen in sexual minority individuals who preach rigid doctrine to themselves in order to affirm their sexual orientation identity. Unable to tolerate conflictual feelings about their sexuality, these individuals sound as if they are trying to convince themselves that “It is okay to be gay.” However, this strategy only reverses the feelings and identifications of a former state of mind. In the subjectivity of a closeted identity, heterosexuality is idealized, and homosexuality dissociated. After coming out, being gay can become idealized, while disapproving feelings are denied. For example, a patient's internalized religious beliefs may not be limited to condemnations of homosexuality. They also include moral and ethical beliefs that are integral to the adult self. Patients cannot simply rid themselves of self-condemnatory attitudes without severing their attachments to other important identifications. The psychotherapeutic challenge is to integrate patients' adult feelings and understanding of sexuality, ethics, and morality with their internalized childhood beliefs. Therapeutic holding entails being able to contain all those aspects of the patient so that greater integration of all the patient's feelings can take place.

Sexual Orientation Conversion Efforts. In recent years, there has been an increase in public attention paid to organizations and individuals who make the claim that sexual orientation is mutable. However, most of the arguments for the position that "change is possible" are not taking place in scientific journals but on religious websites, in the popular press, and in social media. It appears that the goal of these efforts has not been to convince the mental health professions that homosexuality can be changed but to market to the public the idea that it can.

The audience for this advertising approach is individuals with same-sex feelings or attractions who do not wish a homosexual identification. These individuals, often for religious reasons, although not always, are unable or unwilling to accept their homoerotic feelings as normal, natural, or moral. Toward that end, they may seek support from family members and friends, coreligionists, and professionals who also disapprove of homosexuality. Such individuals may seek clinicians who claim to be able to change sexual orientation through *sexual orientation conversion efforts (SOCE).* These can include approaches that go by the name of "reparative therapy," "sexual reorientation therapy," or "sexual conversion therapy." Referring to homosexuality as "unwanted same-sex" attractions is a more recent marketing tool. Those unhappy about their homosexuality may seek religious fellowship among the so-called "ex-gay groups."

APA issued its most recent position statement on the subject in 2020:

> APA condemns any practice that aims to change one's sexual orientation or gender expression in the form of conversion therapy, or any other similar type of therapy, as ethically and morally wrong and, additionally, these practices represent a significant risk of harm by subjecting individuals to forms of treatment that have not been scientifically validated. (Reproduced with permission from the American Psychiatric Association. *American Psychiatric Association Position Statement on Same-Sex Orientation*. American Psychiatric Association; 2020.)

In taking such a position on SOCE, organized psychiatry has acted to protect patients who may be harmed by those procedures. Anecdotal reports indicate that unsuccessful conversion therapies may cause depression, intimacy avoidance, and sexual dysfunction. In 2009, the American Psychological Association issued a report of its *Task Force on the Appropriate Therapeutic Response to Sexual Orientation.* The Task Force conducted a systematic review of the peer-reviewed journal literature on SOCE and concluded that efforts to change sexual orientation are unlikely to be successful and involve some risk of harm. Starting with California in 2012, more than 20 U.S. states and Washington DC have passed laws banning licensed clinicians from the practice of SOCE with minors, and similar bans have been introduced in the legislatures of other states. In 2016, Malta became the first country to ban SOCE for individuals of all ages, and in 2020, Albania and Germany issued similar bans. Most telling, in 2013, one of the largest ex-gay organizations in the world, Exodus International, closed operations, at the same time apologizing for any harm their message of “cure” may have inflicted upon the gay community. In 2020, a United Nations Independent Expert on Sexual Orientation and Gender Identity appointed by the UN Office of the High Commissioner for Human Rights issued a report calling for a “global ban on conversion therapy.”

Nevertheless, some individuals may still wish to change their sexual orientation for religious or other reasons. Clinical exploration of the irrational aspects of patients’ internalized antihomosexual attitudes does not always lead to acceptance of one’s homosexuality. Thus, even if homosexuality per se is not a mental disorder, psychiatry, and other mental health professions may wish to find a way to help individuals who wish to rid themselves of same-sex feelings. However, this does not and should not require redefining homosexuality as an illness. If psychiatry and other mental health professions are to play a role in assisting such patients, perhaps the field of plastic surgery might serve as a model: Cosmetic plastic surgeons routinely treat nonpathologic but socially stigmatized physical conditions.

Plastic surgeons, however, employ standards of care that are not matched by those of sexual conversion therapists, who have yet to develop anything approaching scientific and clinically sound selection criteria for patients. It also is not certain that contemporary conversion therapists can develop more exacting standards of care. For example, after the mental health mainstream endorsed a normal variant model, social acceptance of homosexuality increased. Concurrently, the professional training, credentials, and standing of sexual conversion therapists, in general, have diminished. For example, one study showed that conversion therapists regularly violated professional codes of conduct regarding informed consent, confidentiality, pretermination counseling, and provision of referrals after treatment failure. A field dominated by medically trained practitioners half a century ago is now primarily the province of less rigorously trained and sparsely regulated clinicians, pastoral counselors, and self-help groups. In addition, as many conversion therapists work primarily within a faith-healing model, it remains to be seen whether they can successfully develop scientific and clinical selection criteria to identify individuals who may be harmed by SOCE. Until this can be accomplished, the dictum to "first, do no harm" should be kept in mind for those concerned about the well-being and the quality of care for all patients, regardless of their eventual sexual orientation.

ETHICAL AND LEGAL ISSUES

Principle I of *The Principles of Medical Ethics: With Annotations Especially Applicable to Psychiatry* states that:

> *A physician shall be dedicated to providing competent medical service with compassion and respect for human dignity.*

The annotation goes on to say,

> *A psychiatrist should not be a party to any type of policy that excludes, segregates, or demeans the dignity of any patient because of ethnic origin, race, sex, creed, age, socioeconomic status, or sexual orientation.*

There are several ethical and legal issues that may come up in doing psychiatric work with sexual minority individuals. Among these are confidentiality of medical records, a potential for abuse in settings that discriminate against gay, lesbian, and bisexual patients, and maintenance of standards of care.

Information about sexual feelings, behaviors, identity, and relationships should be entered into the medical record only when it is relevant to the evaluation and treatment of the patient and must always be treated in a confidential manner. Court decisions have upheld the status of confidentiality between a therapist and a patient to be as important as a lawyer-client privilege. The U.S. Supreme Court's 1996 *Jaffe v. Redmond* decision, in particular, holds that while medical records may be part of a patient's public medical record, process notes of private psychotherapy conversations remain privileged. In any event, maintaining the confidentiality of an individual's sexual orientation may be an ongoing concern in some cases, particularly where patients may face discrimination if the medical records reveal their sexual identity in the workplace, a health care setting, or from insurers.

Although a clinician's knowledge of a patient's sexual identity is essential to delivering quality of care, the recording of this information in an unprotected medical record (as in a public clinic or in a managed care setting) without informing the patient of this fact may become problematic. Nonclinical communications to others (a clinic's clerical staff, managed care case managers, school officials) about a patient's sexual identity or sexual behavior without the informed consent of the patient are always unethical. When treating patients in settings where total confidentiality cannot be ensured, it is usually best if the psychiatrist discusses with the patient what can and cannot be kept confidential before any personal data are entered into the chart. If or when the patient does not wish personal information to be recorded and if the psychiatrist does not believe this information should be withheld, a referral to another clinician is highly recommended.

There are institutional settings, such as prisons or religious schools and universities, where public knowledge of a person's sexual minority orientation could have adverse consequences for a patient. As in the public sector, the psychiatrist in these settings is employed by a third party and not directly by the patient. Nevertheless, the ethical constraints on the psychiatrist are similar to those in the public mental health sector: The confidentiality of the patient, with the exception of potential harm to self or others, takes precedence over other institutional needs.

Appropriate standards of care in treating gay, lesbian, and bisexual patients require that psychiatrists monitor their own antihomosexual biases when providing assessment and treatment. In U.S. culture, antihomosexual attitudes are ubiquitous, and, consequently, no one is free of them, not even gay and lesbian psychiatrists. This obliges the psychiatrist to understand and be aware of the ways in which these attitudes have had an impact on the development of their own identities. This can help a psychiatrist come to a greater empathic appreciation of the impact of such attitudes on a gay, lesbian, or bisexual patient. Psychiatrists and other practitioners who continue to treat homosexuality as if it were a mental illness engage in such practices outside the mental health mainstream. It is possible that they may risk ethical sanctions, particularly if they violate professional codes of conduct regarding informed consent, confidentiality, pretermination counseling, and provision of referrals after treatment failure. To date, there has been at least one case of a psychiatrist sanctioned by the APA for offering a patient unwanted SOCE. Closeted gay and lesbian psychiatrists struggling with their own identity issues may either need to refrain for a period of time from working with patients who have similar issues or seek supervision or treatment themselves. All psychiatrists, regardless of their own sexual identity, have an ethical obligation to be up-to-date and informed within their field and should regularly undertake continuing education about the mental health concerns and needs of gay, lesbian, and bisexual patients.

TRAINING AND RESEARCH NEEDS

Conceptions of sexuality, sexual identity, and sexual orientation have shifted significantly in the second half of the 20th century and continue to evolve and differentiate as awareness of sexual diversity increases. Increasing awareness of sexual diversity necessitates an expanding lexicon to provide descriptors for newly recognized sexual subgroups and identities. It also necessitates taxonomical and nosologic revisions to provide coherent and meaningful systems of classification and nomenclature. A solid grounding in contemporary theories of sexuality will help psychiatry residents gain the skills and confidence necessary to interact sensitively and competently with persons of all sexual orientations. Competence in working with sexual minorities also entails an awareness that changing meanings in the vernacular and professional terminology may lead to miscommunication unless unambiguous descriptions are elicited to ensure that the mental health professional understands the meanings that patients intend to communicate with their self-labels of identities and behaviors. To facilitate unambiguous communication, it is often helpful for clinicians to review with patients the range of meanings attributed to terms associated with sexual orientation in order to arrive at a mutual understanding of how those terms will be used in sessions.

In addition to training in contemporary sexology, medical school, and psychiatric residency education should review the origins and evolution of the concept of sexual orientation within Western medical discourse. The mechanism of change and the thinking that went into homosexuality's removal from the DSM-II deserves special emphasis. It should be emphasized that the first DSM workgroup to exclude homosexuality was also the first to grapple with the issue of what actually constitutes a mental disorder. After careful study, the scientific committee concluded that homosexuality did not meet the emerging criteria for what constituted a mental disorder that would shape the subsequent DSM-III. Teaching residents this history is not only important to learning about the development of scientific thought, but it is also essential to countering the erroneous but common misconception that homosexuality was deleted by a simple vote of APA members. That misconception contributes to the erroneous perception that the decision was primarily politically rather than scientifically motivated.

Although homosexuality is no longer in the DSM as a disorder, it is important that faculty and trainees learn about the real mental health consequences of being a sexual minority individual, many of which reflect individual responses to antihomosexual attitudes and social prejudice, aka *sexual minority stress.* Although accepting attitudes toward gay, lesbian, and bisexual individuals are increasing, antihomosexual bias remains widespread in society at large and persists within health care systems. The 2011 report by the Institute of Medicine, The Health of Lesbian, Gay, Bisexual, and Transgender People, concluded that the stress due to the stigma associated with being a sexual minority person contributes to the physical and mental health disparities of gay, lesbian and bisexual people As addressed in chapter 21.4, the stigma associated with being transgender contributes significantly to the health disparities seen in that community. The mental health disparities include higher rates of depression, anxiety disorders, suicidality, and substance abuse. Following that report, the National Institutes of Health established the Sexual and Gender Minority Research Office, and numerous rigorously conducted studies have further characterized the impacts of SGM minority stress on health and identified points of interventions. In particular, the report concluded that instruction about the lives and mental health needs of gay men, lesbian women, and bisexual persons should not be restricted to classes on sexual minority issues but should be fully integrated into the cultural competence curricula as well as the broader curriculum on sexuality in psychiatry and other health disciplines. Such cultural competence training should foster an appreciation of the diversity within the sexual minority community and familiarity with the common issues and concerns within sexual minority communities. At the same time, cultural competence training should discourage stereotyping individual patients as each one presents with a unique constellation of issues. Recent surveys report that many psychiatry residents neglect to routinely take adequate psychosexual histories and report discomfort in doing so or not knowing the essentials that are required in particular clinical situations. It is the responsibility of residency training programs to ensure that these training needs are met.

Past and present theories on the origins of sexual orientation and their role in patient care also need to be addressed. Understanding the history of pathologizing etiologic theories and the "therapies" based on them gives trainees some insight into the trepidation with which sexual minority individuals may approach mental health professionals. This knowledge also prepares residents to appropriately respond when patients bring concerns about this history into treatment. It is important that residents understand that sexual orientation is rarely subject to change and appreciate the possible harm caused by efforts to alter a person's sexual orientation. Residents should master the fundamentals of therapies that are affirmative and respectful of an individual's sexual orientation and identity. Given the prominence of biologically deterministic theories pertaining to sexual orientation, residents should also know that the decision to declassify homosexuality was not based on reductive presumptions about etiology or resistance to change. Contemporary gay-affirmative therapies entail acceptance of homosexuality, regardless of any presumptions about its etiology, as a normal variation of human sexuality. Their focus is on identifying and addressing a patient's life problems and conflicts, not revealing why the individual is gay or lesbian. Fear that a heterosexual therapist would focus on their sexual orientation rather than their chief complaints motivates many sexual minority individuals to seek sexual minority therapists who are publicly open about their sexual identities.

Table 21.2–2.

Internet Resources for Professionals and for GLB Individuals and Their Families

American Medical Association Advisory Committee on LGBTQ Issues (public education for patients and clinicians)	https://www.ama-assn.org/member-groups-sections/advisory-committee-lgbtq-issues
AGLP—Association of LGBTQ+ Psychiatrists (education and professional support for psychiatrists, residents, and medical students)	http://www.aglp.org/
Gay, Lesbian, and Straight Education Network (GLSEN) (support for LGBT youth in the school system)	https://www.glsen.org/resources
GLMA (education and professional support for physicians and other health professionals)	http://www.glma.org/
Group for the Advancement of	http://www.aglp.org/gap/

Psychiatry (GAP): LGBT Mental Health Syllabus (for medical students and residents)	
HealthyPeople.gov (information on lesbian, gay, bisexual, and transgender health, including mental health)	https://www.healthypeople.gov/2020/topics-objectives/topic/lesbian-gay-bisexual-and-transgender-health
Hetrick–Martin Institute (social support for LGBT youth)	http://www.hmi.org/
Lambda Legal (advocacy group for legal protections)	http://www.lambdalegal.org/
PFLAG (support, information, and resources for LGBTQ people, their parents, families, and allies)	http://community.pflag.org/
SAGE (advocacy and services for LGBT elders)	http://www.sageusa.org/issues/
Society for the Psychological Study of Lesbian, Gay,	http://www.apa.org/about/division/div44.html

and Bisexual Issues—Division 44 of the American Psychological Association (public education and professional support)	
The Trevor Project (resources and information aimed at preventing suicide in LGBTQ youth)	https://www.thetrevorproject.org/resources/
What We Know (review and links to scholarly research on well-being of children with gay or lesbian parents)	https://whatweknow.inequality.cornell.edu/topics/lgbt-equality/what-does-the-scholarly-research-say-about-the-wellbeing-of-children-with-gay-or-lesbian-parents/

Instruction about the mental health needs of sexual minorities must include education concerning obstacles to their health care, many of which are being diminished by recent legislation and policy changes. Nevertheless, sexual minorities, due in part to negative past experiences, may avoid healthcare settings. They may also be guarded about discussing their sexual behavior with healthcare providers, fearing that it will lead to discrimination. Moreover, the use of heteronormative language and failure to inquire about sexuality on the part of health care providers may further discourage patients from disclosing their sexual minority status. Yet, such disclosure is a prerequisite to discussing how that status affects their lives as well as their particular sexual concerns. Alternatively, an atmosphere that fosters disclosure can be perceived as more welcoming and is likely to enhance adherence with follow-up appointments and engagement in patient-centered care.

Although all residents will benefit from contact with respected psychiatrists who are openly LGB, sexual minority residents may particularly benefit from access to sexual minority faculty as role models. National organizations like AGLP: The Association of LGBTQ+ Psychiatrists (www.aglp.org) and the Caucus of Gay, Lesbian, and Bisexual Psychiatrists of the APA can serve as such a resource. Medical schools and professional organizations should increase the opportunities available for training about the clinical concerns of sexual minorities and encourage gay and lesbian faculty and other leaders to be visible and active in a wide range of professional activities.

There is still much to be understood about sexual orientation and how the needs of sexual minority individuals vary demographically and across the life cycle. Training should engender curiosity about these unknowns and awareness of areas in need of further research. Preliminary research has described variations in sexual desire, behavior, and identity associated with a range of demographic and personal characteristics such as sex, age, race, class, geographic location, and religion. Until recently, most studies have focused on White, urban, middle-class men. However, over the past decade, an increasing number of studies have begun to examine the impact of intersectional factors, including race, ethnicity, and socioeconomic factors, on sexual minority physical and mental well-being. Because most studies have been cross-sectional, there is a need for longitudinal studies to investigate the development and experience of sexual orientation and sexual identity over time and across the lifespan. Compared to cross-sectional studies, longitudinal studies may be more helpful in elucidating the factors that shape individual expressions of desire and behavior, as well as contributors to sexual minority disparities in physical and emotional health. Of note, in addition to rigorously assessing sexual orientation and gender identity, many federal health surveys are currently collecting data longitudinally.

Current research has moved beyond traditional biologic and psychosocial approaches to incorporate constructivist approaches that interpret the role of fluctuating cultural and other historical forces in shaping individual lives and experiences.

INTERNET RESOURCES

The websites listed in Table 21.2–2 offer useful resources for gay, lesbian, and bisexual patients, their families, and mental health professionals themselves.

FURTHER READINGS

Adelson SL, Dowshen NL, Makadon HJ, Garofalo R. Introduction to lesbian, gay, bisexual, and transgender youth health. *Pediatr Clin North Am*. 2016;63(6):xvii–xxi.

Advancing Effective Communication, Cultural Competence, and Patient- and Family-Centered Care for the Lesbian, Gay, Bisexual, and Transgender (LGBT) Community: A Field Guide. The Joint Commission; 2011.

American Psychiatric Association. Position statement on issues related to homosexuality. 2013. Accessed August 15, 2021. http://www.psychiatry.org/File%20Library/About-APA/Organization-Documents-Policies/Policies/Position-2013-Homosexuality.pdf

American Psychological Association Task Force on the Appropriate Therapeutic Response to Sexual Orientation. *Report of the Task Force on the Appropriate Therapeutic Response to Sexual Orientation*. Author; 2009.

*Bailey MJ, Vasey PL, Diamond LM, Breedlove SM, Vilain E, Epprecht M. Sexual orientation, controversy, and science. *Psychol Sci Public Interest*. 2016;17(2):45–101.

*Bayer R. *Homosexuality and American Psychiatry: The Politics of Diagnosis*. Princeton University Press; 1987.

Breedlove MS. Prenatal influences on human sexual orientation: expectations versus data. *Arch Sex Behav*. 2017;46:1583–1592.

Cochran SD, Drescher J, Kismodi E, Giami A, García-Moreno C, Reed GM. Proposed declassification of disease categories related to sexual orientation in ICD-11: rationale and evidence from the Working Group on Sexual Disorders and Sexual Health. *Bull World Health Organ*. 2014;92:672–679.

Drescher J. *Psychoanalytic Therapy and the Gay Man*. Routledge; 1998.

*Drescher J, Merlino JP. *American Psychiatry and Homosexuality: An Oral History*. Routledge; 2007.

Drescher J, Schwartz A, Casoy F, et al. The growing regulation of conversion therapy. *J Med Regul*. 2016;102(2):7–12.

Ganna A, Verweij KJH, Nivard MG, et al. Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior. *Science*. 2019;365(6456):eaat7693.

Gartrell, NK, Bos HM, Goldberg NG. Adolescents of the U.S. National Longitudinal Lesbian Family Study: sexual orientation, sexual behavior, and sexual risk exposure. *Arch Sex Behav*. 2011;40:1199–1209.

Gooren L. The biology of human psychosexual differentiation. *Horm Behav*. 2006;50:589–601.

Hooker E. The adjustment of the male overt homosexual. *J Proj Tech*. 1957;21:18–31.

Institute of Medicine. *The Health of Lesbian, Gay, Bisexual and Transgender People: Building a Foundation for Better Understanding*. The National Academies Press; 2011.

LeVay S. *Gay, Straight, and the Reason Why: The Science of Sexual Orientation*. Oxford University Press; 2011.

Levounis P, Drescher J, Barber ME. *The LGBT Casebook*. American Psychiatric Publishing; 2012.

Magee M, Miller DC. *Lesbian Lives: Psychoanalytic Narratives Old and New*. Analytic Press; 1997.

Malyon AK. Psychotherapeutic implications of internalized homophobia in gay men. *J Homosex*. 1981;7:59–69.

Medley G, Lipari RN, Bose J, Cribb DS, Kroutil LA, McHenry G. Sexual orientation and estimates of adult substance use and mental health: results from the 2015 National Survey on Drug Use and Health. *NSDUH Data Rev*. 2016. Accessed August 15, 2021. https://www.samhsa.gov/data/sites/default/files/NSDUH-SexualOrientation-2015/NSDUH-SexualOrientation-2015/NSDUH-SexualOrientation-2015.htm

Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. *Psychol Bull*. 2003;129(5):674–697.

Pachankis JE. The scientific pursuit of sexual and gender minority mental health treatments: toward evidence-based affirmative practice. *Am Psychol*. 2018;73(9):1207–1219.

Pathela P, Hajat A, Schillinger J, Blank S, Sell R, Mostashari F. Discordance between sexual behavior and self-reported sexual identity: a population-based survey of New York City men. *Ann Int Med*. 2006;145:416–425.

Peralta RL, Victory E, Thompson CL. Alcohol use disorder in sexual minority adults: age- and sex-specific prevalence estimates from a national survey, 2015–2017. *Drug Alcohol Depend*. 2019;205:107673.

Plöderl M, Tremblay P. Mental health of sexual minorities: a systematic review. *Int Rev Psychiatry*. 2015;27:367–385.

Reed GM, Drescher J, Krueger RB, et al. Revising the ICD-10 mental and behavioural disorders classification of sexuality and gender identity based on current scientific evidence, best clinical practices, and human rights considerations. *World Psychiatry*. 2016;15:205–221.

Roselli CE. Neurobiology of gender identity and sexual orientation. *J Neuroendocrinol*. 2018;30(7):e12562.

Ryan C, Huebner D, Diaz RM, Sanchez J. Family rejection as a predictor of negative health outcomes in white and Latino lesbian, gay, and bisexual young adults. *Pediatrics*. 2009;123:346–352.

Schwartz A, Yarbrough E, McIntosh CA. Sex and relationship issues in work with the LGBTQ community. *Focus (Am Psychiatr Publ)*. 2020;18(3):277–284.

Stall R, Dodge B, Bauermeister JA, Poteat T, Breyrer C. *LGBTQ Health Research: Theory Methods Practice*. Johns Hopkins University Press; 2020.

United Nations. Practices of so-called “conversion therapy”: Report of the Independent Expert on Protection against Violence and Discrimination based on Sexual Orientation and Gender Identity. 2020. Accessed online July 16, 2023 at https://digitallibrary.un.org/record/3870697?ln=en

21.3 Paraphilic Disorders

RENEÉ M. SORRENTINO, M.D.

DSM-5 defines *paraphilia* as “any intense and persistent sexual interest other than sexual interest in genital stimulation or preparatory fondling with phenotypically normal physically mature, consenting human partners.” The term *paraphilic disorder,* as defined by the *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition (DSM-5) and its text revision (DSM-5-TR), is used to refer to a paraphilia that is currently causing distress or impairment to the individual or a paraphilia whose satisfaction has entailed personal harm, or risk of harm, to others. Paraphilic interests are more common in men, although not well researched in women, and include a wide variety of behaviors typically seen in clinical populations seeking evaluation and treatment for paraphilias. Those who most frequently seek psychiatric treatment for sex offenses are those accused of or involved in child molestation, voyeurism, exhibitionism, fetishism, frottage, and public masturbation. The term *sex offender* is used to refer to an individual who has been legally convicted of a sex offense. The terminology sex offender does not provide any information regarding the offender’s reason for committing the sexual offense, nor does it indicate the presence or absence of a paraphilia. The psychiatric treatment of sex offenders is distinct from other clinical populations. In many jurisdictions in the United States, sex offender legislation has mandated treatment for offenders deemed “sexually dangerous” or “sexually violent” in psychiatric hospitals where they are civilly committed. Such legislation has generated longstanding debates on the diagnosis of paraphilias, the nature of mental illness, and the treatability of sex offenders. The changes to the paraphilia diagnoses in DSM-5 and DSM-5-TR generated debate about how to conceptualize paraphilias and the adverse legal consequences implicit in identifying certain sexual behaviors as pathologic.

DEFINITION

The definition of sexual deviance or sexual perversities depends on the particular society and culture in which the behavior exists. A behavior may be considered deviant by one society and accepted, albeit unusual, by another. The term *paraphilia* is derived from the Greek words “*para*” meaning next to, and “*philia,*” meaning love. The etymologic definition of paraphilia is “next to or alongside love.” The term paraphilia, therefore, refers to a love that is perverted or “alongside the norm.” In addition to the presence of a deviant sexual interest, paraphilias are characterized by difficulty resisting an impulse for the sexual acts.

DSM-5 made a distinction between paraphilias and paraphilic disorders. The term paraphilic disorders refers to a paraphilia that is currently causing distress or impairment to the individual or a paraphilia whose satisfaction has entailed personal harm, or risk of harm, to others. The definition of paraphilias in DSM-5 changed. Previously, paraphilias were defined as intense, recurring sexual fantasies, sexual urges, or behaviors that involve nonhuman objects, children, or nonconsenting adults, suffering or humiliation (to self or to others), lasting for at least 6 months. DSM-5 defined paraphilia as “any intense and persistent sexual interest other than sexual interest in genital stimulation or preparatory fondling with phenotypically normal, physically mature, consenting human partners.”

DSM-5 and DSM-5-TR include eight specific paraphilic disorders: exhibitionistic disorder (sexual arousal from exposure of genitals to a stranger), pedophilic disorder (sexual arousal to prepubescent individual, generally 13 years of age or younger), voyeuristic disorder (observing others’ sexual activities), fetishistic disorder (use of inert objects, such as female undergarments), transvestic disorder (cross-dressing), sexual sadism disorder (inflicting suffering or humiliation), sexual masochism (being humiliated, beaten, bound, or made to suffer), and frotteuristic disorder (touching, rubbing against a nonconsenting person). Paraphilia not otherwise specified (NOS) has been replaced in DSM-5 with specified paraphilic disorder and unspecified paraphilic disorder as the result of a system-wide standardized DSM-5 change to each residual DSM-IV NOS category.

HISTORY

Societies define normal sexual behavior as that behavior that results in healthy, nonharming sexual contact. Those sexual practices that do not result in healthy, nonharming sexual contact are judged as undesirable and in need of change. Those individuals who engage in sexual behaviors outside this definition are subject to criminal action. Historically, society has created sexual prohibitions that, if breached, have serious legal and social consequences. Of all criminal behaviors, those who breach sexual prohibitions are viewed as the most deplorable. Those behaviors that cause obvious harm to others, such as sexual contact with children or bodily punishment to another, are universally defined as abnormal or sexually deviant. However, those sexual practices that do not harm others are not as clearly defined. Some societies have further defined sexual deviance as sexual practices that are maladaptive or deviations from the social norm. This definition would then include those behaviors that are considered evolutionarily maladaptive and uncommon. Richard von Krafft-Ebing, a German psychiatrist, credited with formally introducing the study of sexology as a psychiatric phenomenon, identified paraphilias first in his 1886 *Psychopathia Sexualis* (Sexual Psychopathy). This highly influential psychiatric text laid the groundwork for the development of research and treatment in this area that has taken place over the past century.

The term *paraphilia* was coined by psychologists in the early 1900s to refer to the sexual variations that were just starting to be researched. Because of a lack of understanding, it included transvestism and other nonsexual variations as well. Previously, these were called *perversions.* Paraphiliacs treated by psychologists were seeking help for behavior they could not control. The word retained this meaning and did not change to include those who were sexually curious.

Sigmund Freud held that human sexuality advanced through stages. He theorized that individuals would progress toward “normal” heterosexuality unless prevented. He saw all paraphilias as infantile, that is, not mature sexual behavior. Because the paraphilias that involved diapers fit well into this theory, they were well-documented. Wilhelm Stekel’s *Patterns of Psychosexual Infantilism* came from this perspective. Homosexuality, sadism, zoophilia, and many others were included in his examples.

As research advanced, Freud’s notion of psychosexual infantilism was replaced with different views. These views dealt individually with the major paraphilias, such as fetishism, masochism, and transvestism. The diaper paraphilias were generally classified into one of these major categories as a result of the author’s viewpoints. In 1987, the third revised edition of DSM (DSM-III-R) used *infantilism* to refer to those who had a “desire to be treated as a helpless infant and clothed in diapers” and categorized infantilism under sexual masochism.

COMPARATIVE NOSOLOGY

DSM originally classified sexual deviations with psychopathic personality disturbances. This classification was based on the belief that sexual deviations were unlawful, and thereby, those individuals engaging in sexual deviations were psychopathic. In DSM-II, sexual deviations were classified as personality disorders. DSM-III changed the nomenclature of these behaviors from sexual deviation to paraphilia. Paraphilias were classified as psychosexual disorders, which included gender identity disorders, psychosexual dysfunctions, and egodystonic homosexuality. DSM-IV-TR reclassified transvestism from a disorder of gender identity to a paraphilia called *transvestic fetishism.* DSM-5 redefined paraphilias, made a distinction between paraphilias and paraphilic disorders, and added course specifiers. The paraphilic disorders are presented in DSM-5 and DSM-5-TR according to the following classification scheme: anomalous activity preferences, which include voyeuristic disorder, exhibitionistic disorder, and frotteuristic disorder; algolagnic disorders, which include sexual masochism disorder and sexual sadism disorder; anomalous target preferences, which include pedophilic disorder, fetishistic disorder, and transvestic disorder.

The question of whether paraphilias are mental disorders was directly answered in DSM-5. In DSM-5, paraphilias are not considered mental disorders unless they cause distress or impairment (criterion B) to the individual, or the paraphilia causes personal harm or risk of harm to others. If a paraphilia causes impairment or harm, it is a mental disorder termed "paraphilic disorder." This change in nomenclature represents an attempt by DSM-5 to preserve the distinction between normative and nonnormative sexual behavior without labeling nonnormative sexual behavior a mental disorder.

DSM-5 was the first edition to add course specifiers to paraphilic disorders. Similar to the course specifiers in other DSM diagnoses, the course specifiers indicate a change in symptoms. The course specifiers are included in all of the paraphilic disorders, with the exception of other specified paraphilic disorder, unspecified paraphilic disorder, and pedophilic disorder. The course specifier named "in a controlled environment" was introduced to indicate the absence of paraphilic urges in settings where opportunities to engage in paraphilic behaviors are restricted. An individual "in full remission" has not acted on the urges with a nonconsenting person, and there has been no distress or impairment in social, occupational, or other areas of functioning for at least 5 years in an uncontrolled environment. The question of whether a paraphilic disorder can be in remission independent of the environment is not known.

Previous editions of DSM attempted to address the diagnostic difficulties and assumptions inherent in the diagnostic category of paraphilias. There is a second group of sexual impulse disorders not currently classified as paraphilias because the particular sexual behaviors affected are not considered "deviant" with respect to contemporary cultural norms. These disorders have been designated as paraphilia-related disorders by Kafka based on the following clinical data:

- The boundary for social as well as sexual deviance is largely determined by cultural and historical context. As such, sexual disorders once considered paraphilias (e.g., homosexuality) are now regarded as variants of normal sexuality; so too, sexual behaviors currently considered normal (e.g., masturbation) were once culturally proscribed.
- Paraphilia-related disorders have been diagnosed in male paraphiliacs, and men selected for studies of paraphilia-related disorders are reported to have paraphilias as well.
- Paraphilia-related disorders can produce a pattern of sexual frequency and intensity with concomitant psychosocial impairment that does not readily distinguish them from nonviolent paraphilic disorders.
- Analogous with paraphilias, the presence of a single paraphilia-related disorder increases the likelihood of occurrence of a multiplicity of these behaviors in an affected person.
- Paraphilias and paraphilia-related disorders can both be ameliorated by antiandrogens and serotonergic antidepressants, such as SSRIs. Disorders of sexual impulsivity can be pleomorphic and may include sexual arousal to behaviors that are socially “deviant” as well as “normal.” In fact, inasmuch as paraphilia-related disorders may be common in paraphiliacs and also occur without the companion of the latter, paraphilia-related disorders may be the more prevalent form of sexual impulsivity. In addition, although the estimated sex difference ratio for paraphilia-related disorders is unknown, protracted promiscuity (e.g., “nymphomania”) and compulsive masturbation are not uncommon behaviors described by women.

The paraphilic-related disorders were not incorporated in DSM. However, DSM-5 considered including portions of these disorders by proposing a new disorder, hypersexual disorder. Hypersexual Disorder was defined as nonparaphilic sexual interests with an impulsivity component. Hypersexual disorder was rejected from DSM due to concerns regarding its diagnostic validity.

DSM-5 proposed broadening the pedophilic disorder diagnosis to include attraction to pubescent as well as prepubescent children. The proposed modified diagnosis, pedophebephilic disorder, referred to a distinct group of individuals who are sexually aroused to adolescents. Ultimately pedophebephilic disorder was not incorporated in the DSM-5.

There has been a longstanding disagreement regarding whether rapists meet the criteria for paraphilia. The category of paraphilic coercive disorder was proposed in DSM-5 for individuals with intense, repetitive urges of 6 months' duration to commit rape, who had either acted on these urges or were disturbed by the urges. DSM has never included rape or paraphilic coercive disorder as a diagnosis.

The changes to paraphilic disorders in DSM-5 represent an attempt to answer the central philosophical question of what constitutes disordered sexual behavior. They include a change in nomenclature and the addition of courses. Previous additions to DSM have attempted to make a distinction between normal and abnormal sexual behaviors. DSM-5 made a distinction between paraphilias and paraphilic disorders, indicating that paraphilias are not ipso facto mental disorders. The basic structure of the diagnostic criteria in DSM-5 is unchanged from DSM-IV-TR.

The World Health Organization approved the International Classification of Diseases, 11th version (ICD-11), which includes the replacement of the section of the "Disorders of Sexual Preference" with "Paraphilic Disorders." Similar to the DSM-5, this change represented a move from stigmatizing unconventional or unusual sexual behaviors to defining pathologic or disordered sexual behaviors.

EPIDEMIOLOGY

Gender Ratio

Paraphilic disorders are predominantly thought to be male sexuality disorders because women rarely present in clinical or forensic settings, and research is limited. In nonclinical samples, however, both men and women report a variety of paraphilic interests. In these studies, men report a higher incidence of sexual interests in sadism, exhibitionism, and fetishism, whereas women report a higher interest in masochism. The reason for this gender disparity is unknown, but the role of gender bias in the reporting and evaluation of sexual offenses likely is contributory.

Paraphilic Populations. The frequency of paraphilic behavior is unknown. The majority of epidemiologic data relating to paraphilias derive from clinical populations. It is difficult to obtain the frequency of paraphilic behaviors in the nonclinical population, as individuals are reluctant to report sexual fantasies and behaviors. In one study of the general population, 61.7% of men reported fantasies of initiating a young girl into sexuality; 33.0% described fantasies of raping adult women; 11.7% described masochistic fantasies; 5.3% described fantasies of having sex with an animal; and 3.2% described fantasies of initiating a young boy into sexuality. A study of college students' sexual behavior revealed that 42% reported having participated in voyeurism; 35% in frottage; 8% in making obscene telephone calls; 5% in coercive sexual activity; 3% in sexual contact with girls under age 12; and 2% in exhibitionism. A total of 65% reported having participated in some variant of paraphilic behavior. Another general population study, including men and women, found similar levels of interest in fetishism and masochism.

In a university study of both men and women, 50.6% of men and 41.5% of women identified a paraphilic interest. Interests in voyeurism, exhibitionism, sadism, and frotteurism were higher in men, whereas interests in fetishism and masochism were greater in women.

Paraphilia Disordered Populations. Those individuals who seek evaluation and treatment are a distinct group. It is unlikely that individuals seeking treatment represent the heterogeneous population of individuals with paraphilias. Some paraphilias are more likely to come to the attention of others because they involve another person (such as pedophilia or exhibitionism). However, those paraphilias that do not involve another person, such as the use of fetish objects, are less likely to be treated. Those paraphilias whose behavior is criminal tend to be overrepresented in clinical populations because they frequently lead to mandatory reporting or court-ordered treatment.

Data gathered from more than 90 treatment programs throughout North America regarding 2,129 cases of individuals seeking assessment indicate that 37.1% had been involved in child molestation; 20.2% in voyeurism; 13.8% in exhibitionism; 13.3% in fetishism; 11.2% in frottage; and 10.7% in public masturbation. A study of 561 paraphilic persons who had committed more than 291,000 paraphilic acts against more than 195,000 victims indicated that, of all victims, 37.3% were victims of exhibitionism; 28.6% of frottage; 13.6% of voyeurism; 11.8% of molestation against boys outside the home; 3.5% of public masturbation; 2.3% of molestation of girls outside the home; 1% of obscene telephone calls; 0.9% of bestiality; 0.5% of rape of adult women; 0.2% of molestation of girls in the home; 0.2% of urolagnia; and 0.1% or less of sadism, masochism, fetishism, molestation of boys within the home, obscene mailings, coprophilia, or attraction to odors.

Frequency of Paraphilic Behavior

The presence of a paraphilia does not indicate whether an individual engages in paraphilic behavior or sexual fantasy. Some individuals never act on paraphilic urges. Rather, they engage in sexual gratification from sexual fantasies. Other individuals carry out paraphilic fantasies. In a study of individuals seeking treatment, the median number of paraphilic acts by paraphilic category are as follows: Nonincestuous molestation of a female victim 1.4; nonincestuous molestation of a male victim 10.1; incestuous molestation of a female victim 4.4; incestuous molestation of a male victim 5.2; rape of adult women 0.9; exhibitionism 50.5; voyeurism 16.5; frottage 29.5; transvestism 25.0; fetishism 3.3; sadism 3.0;

masochism 36.0; obscene telephone calls 30.0; public masturbation 50.0; and bestiality 2.2.

Crossing Over

Initially, individuals suffering from a paraphilia were thought to be obsessed with one type of paraphilic behavior. Recent studies suggest that individuals with paraphilias may “crossover” from one paraphilia behavior to another. Of paraphilic persons primarily involved with actual touching of victims (frottage, rape, and pedophilia), 30.6% had previously or concomitantly been involved in paraphilic acts not involving touching of the victim (voyeurism and exhibitionism). Sixty-four percent of paraphilic individuals primarily involved with nontouching had previously or concomitantly carried out paraphilic acts that involved touching. Of paraphilic persons primarily victimizing outside the family, 29.4% had previously or concomitantly carried out paraphilic acts against family victims. Of those who primarily had been involved in paraphilic acts with family members, 65.8% had previously or concomitantly carried out paraphilic sexual acts against nonfamiliar victims. Of paraphilic persons who primarily committed acts against female victims, 22.9% had previously or concomitantly carried out paraphilic acts against male victims. Of those who had primarily carried out paraphilic acts against male victims, 62.6% had previously or concomitantly carried out paraphilic acts against female victims. Those paraphilias with the highest incidence of crossover were sadism, public masturbation, male incest pedophilia, fetishism, masochism, voyeurism, female nonincest pedophilia, exhibitionism, male nonincest pedophilia, transvestism, frottage, female incest pedophile, and rape of adult women. These studies suggest that paraphiliacs tend to cross over between touching and nontouching of their victims, between family and nonfamily members, between female and male victims, and to victims of various ages.

Comorbid Psychiatric Disorders

Studies have demonstrated high-comorbidity of general psychiatric disorders in paraphilic sexual offenders. The general psychiatric disorders comorbid with paraphilic sexual offenders include personality disorders, mood disorders, social anxiety disorders, substance use disorders, attention deficit hyperactivity disorder, and other neurodevelopmental disorders. The prevalence of general psychiatric disorders in paraphilic disordered, nonoffending patients is unknown. One study of 600 undergraduates with paraphilic interests by Watts and colleagues found narcissism, antisocial, and psychopathic features were most significantly associated with paraphilic interests.

ETIOLOGY

The etiology of paraphilias is unknown. Biologic theories, including neuroanatomical and neuroimaging studies of sex offenders, indicate that congenital or acquired brain damage is overrepresented. Some pedophilic perpetrators show structural impairment of brain regions critical for sexual development. Subtle defects of the right amygdala and closely-related structures might be implicated in the pathogenesis of pedophilia and might reflect developmental disturbances or environmental insults at critical periods. Neurologic hypotheses of paraphilia propose that sexual deviance is associated with frontal and/or temporal lobe damage. This damage may translate into an individual’s inability to control sexual impulses or directly cause paraphilic behavior because of tissue damage.

A monoamine hypothesis for the pathophysiology of paraphilic disorders was first articulated in 1997 by Kafka. This hypothesis was based on four converging lines of empirical evidence. First, the monoamine neurotransmitters dopamine, norepinephrine, and serotonin serve a modulatory role in human and mammalian sexual motivation, appetite, and consummatory behavior. Second, the sexual effects of pharmacologic agents that affect monoamine neurotransmitters can have both significant facilitative and inhibitory effects on sexual behavior. Third, paraphilic disorders appear to have Axis I comorbid associations with nonsexual psychopathologies that are associated with monoaminergic dysregulation. Last, pharmacologic agents that enhance central serotonergic function, in particular, have been reported to ameliorate paraphilic sexual arousal and behavior.

Psychoanalytical Theory

Psychoanalysts generally theorize that these conditions represent a regression to or a fixation at an earlier level of psychosexual development, resulting in a repetitive pattern of sexual behavior that is not mature in its application and expression. In other words, an individual repeats or reverts to a sexual habit arising early in life. Another psychoanalytic theory holds that these conditions are all expressions of hostility in which sexual fantasies or unusual sexual acts become a means of obtaining revenge for childhood trauma. The persistent, repetitive nature of the paraphilia is caused by an inability to erase the underlying trauma completely. Indeed, a history of childhood sexual abuse is sometimes seen in individuals with paraphilias.

Behaviorist Theory

Behaviorists suggest that paraphilias begin via a process of conditioning. Nonsexual objects can become sexually arousing if they are frequently and repeatedly associated with a pleasurable sexual activity. The development of paraphilia is not usually a matter of conditioning alone; there must usually be some predisposing factor, such as difficulty forming person-to-person sexual relationships or poor self-esteem.

The following are situations or causes that might lead someone in a paraphiliac direction:

- parents who humiliate and punish a small boy for strutting around with an erect penis;
- a young boy who is sexually abused;
- an individual who is dressed in a woman’s clothes as a form of parental punishment;
- fear of sexual performance or intimacy;

- inadequate counseling;
- excessive alcohol intake;
- physiologic problems;
- sociocultural factors;
- psychosexual trauma.

DIAGNOSIS AND CLINICAL FEATURES

DSM identifies eight categories of paraphilic disorders. All of the paraphilic disorders meet the following DSM criteria: The activity must be the sole means of sexual gratification for a period of 6 months and cause “marked distress or interpersonal difficulty” or “whose satisfaction has entailed personal harm, or risk of harm, to others” to be considered such. The categories of paraphilic disorders are exhibitionism, fetishism, frotteurism, masochism, pedophilia, sadism, transvestic fetishism, and voyeurism. Paraphilias that do not meet the guidelines of these categories are called a specific paraphilic disorder or unspecified paraphilic disorder. These terms replaced paraphilia NOS that existed in DSM-IV-TR. Paraphiliacs may choose an occupation or volunteer in an environment that places them in close contact with the desired object. For example, a pedophile may choose to volunteer with young children in an effort to make contact with children.

Diagnostic Considerations

The diagnosis of a paraphilic disorder is based on the individual’s report or demonstration of sexual arousal to a specified object or behavior with resultant impairment or risk of harm to self or others. For example, an individual with pedophilia may meet the diagnostic criterion by reporting sexual arousal to prepubescent children with a resultant psychological impairment such as anxiety or depression or having acted on the interests. Alternatively, a pedophile may meet the diagnostic criterion for pedophilia by demonstrating sexual arousal to stimuli of prepubescent children (e.g., erection to pictures of children in sexual poses) with a history of sexual contact with children. Paraphilic interests generally develop shortly after puberty and follow a chronic course with periods of variability in the degree of sexual interests, fantasies, or behaviors.

The diagnosis of a paraphilic disorder should be made after a comprehensive psychiatric evaluation, often referred to as a psychosexual evaluation. The clinical interview should include a general medical and psychiatric history to rule out medical conditions which may mimic paraphilias, such as treatment with dopaminergic agents or a history of traumatic brain injury. The sexual history should include a history of childhood exposure to sexual acts, sexual experiences, and sexual functioning, including masturbation patterns. A review of systems for each of the paraphilias is important as some individuals may present with multiple paraphilias.

Many individuals are reluctant to share sexual interests, fantasies, or behaviors due to shame or fears of legal consequences. Like all psychiatric disorders, it is important to obtain collateral data to inform diagnoses. When possible, obtaining a history of sexual behaviors from significant others is important. In addition, when an individual has legal involvement, reviewing police reports is imperative.

Exclusive and Nonexclusive Subtypes. Individuals who identify a paraphilic interest to be the only way to achieve sexual gratification are classified as “exclusive” paraphiliacs. A man who is only sexually aroused to prepubescent children is a pedophile, exclusive subtype. Individuals who have paraphilic interest but are able to derive sexual gratification from other experiences are called “nonexclusive” paraphiliacs. A nonexclusive pedophile experiences sexual gratification from both prepubescent children and adults. Nonexclusive paraphiliacs are more common than exclusive paraphiliacs.

Exhibitionistic Disorder

Exhibitionistic disorder, also known as “flashing” or “indecent exposure,” is qualified by recurrent and intense sexual arousal from the exposure of one’s genitals to an unsuspecting person persisting for a period of at least 6 months. In order to meet the criteria for a disorder, the individual must act on these sexual urges with a nonconsenting person, or the sexual urges or fantasies cause clinically significant distress or impairment in social, occupational, or other important areas of functioning. The diagnosis of exhibitionistic disorder describes individuals, who for at least 6 months, experience sexually gratifying urges, fantasies, or behaviors that are related to displaying nudity to an unsuspecting individual. Sexual arousal derives from the showing of one’s genitals, breasts, or buttocks to a nonconsenting individual. The diagnosis may be made if the sexual urges or fantasies lead to significant distress, impairment in functioning, or actual behavior.

Exhibitionism occurs predominately in men, with a peak age of onset in the 20s. In the United States, most exhibitionists are Caucasian males. About half of exhibitionists are married. The true prevalence is unknown, but the literature estimates a prevalence of 2% and 4%. Generally, no additional contact with the observer is sought; the individual is stimulated sexually by gaining the attention of and startling the observer. The exhibitionist often masturbates as part of the exposure. Exposing oneself while driving or in a public place is the most common expressions of exhibitionism. DSM-5 added specifiers to identify the source of sexual arousal. The specifiers include the following: sexual arousal by exposing genitals to prepubertal children, sexual arousal by exposing one’s genitals to physically mature individuals, and sexual arousal by exposing genitals to prepubertal children, and to physically mature individuals.

Exhibitionism is one of the three most common sexual offenses in police records (the other two are voyeurism and pedophilia). Exhibitionists have been found to have a high-rate of comorbidity among the paraphilias as a group and with other mental disorders. In other words, a patient in treatment for exhibitionism is highly likely to engage in other forms of deviant sexual behavior and to suffer from depression or an anxiety or substance abuse disorder. It should be noted, however, that exhibitionists are a heterogeneous group of individuals. The risk factors established in the literature include insecure attachment, childhood sexual abuse, substance abuse, and sexual dysfunction, which may not apply to all exhibitionists.

Fetishistic Disorder

Individuals who achieve sexual gratification from the use of nonliving objects or a highly specific focus on nongenital body parts over a period of at least 6 months with resultant clinically significant distress or impairment in social, occupational, or other important areas of functioning meet the criteria for fetishistic disorder. DSM-5-TR specifies that the fetish objects are not limited to articles of clothing used in cross-dressing or devices specifically designed for the purpose of tactile genital stimulation. Fetishistic disorder is further defined by specifiers indicating the source of sexual arousal such as “body part(s), nonliving object(s), or other.” The use of objects, most commonly women’s undergarments, shoes, stockings, or other clothing items, suffer from fetishism. Fetishists often collect the object of sexual gratification. Some of the more common objects are women’s lingerie or specific materials like silk, leather, or fur. Partialism, an older term, refers to fetishes specifically involving nonsexual parts of the body. Fetishism has predominantly been described in men and often exists with other paraphilias.

Fetishistic disorder is used to refer to at least a 6-month period of time in which an individual experiences sexual urges, fantasies, or behaviors related to objects, situations, or body parts not traditionally viewed as sexual. The sexual arousal must cause significant distress or impairment in functioning in order to be considered a paraphilic disorder. The true prevalence of fetishistic disorder is unknown because the literature is limited to case series. Most individuals with fetishistic disorder are male, and the most common fetish interest is feet. The etiology of fetishistic interests is unknown but likely includes biologic, developmental, and sociocultural factors.

Frotteuristic Disorder

The word “frotteur” derives from the French word “frotter,” meaning to rub. Frotteurism refers to the sexual gratification achieved by touching or rubbing against a nonconsenting person. In order to meet the criteria for frotteuristic disorder, the individual must experience recurrent and intense sexual arousal over a 6-month period, and either act on the behavior or the sexual urges or fantasies cause clinically significant distress or impairment in social, occupational, or other important areas of functioning. This behavior often occurs in busy, crowded places, such as on busy streets or on crowded buses or subways. Frotteurism is predominantly described in men. A frotteur may fantasize that the woman he is rubbing up against is mutually aroused by the behavior.

The diagnosis of frotteuristic disorder is based on a greater than 6-month period of time in which an individual experiences gratification from urges, fantasies, or behaviors-related to brushing up against an unsuspecting individual. An individual meets the criteria for the diagnosis if he experiences significant distress or impairment in functioning as a result of the fantasies or has acted on the fantasies.

The true prevalence of frotteuristic disorder is unknown. The literature is limited to small samples without clear diagnostic criteria. The prevalence rates among these small studies range from 8% to 35%. The most common presentation is a male rubbing up against the breasts or genital area of a female in a crowded setting. The etiology of the disorder is not well-understood. Possible explanations for the behavior include psychoanalytic perspectives that the individual has unmet childhood needs met by rubbing up against an individual.

Pedophilic Disorder

Pedophilia has been included in the American Psychiatric Association DSM as a mental disorder since 1968. In DSM-III and DSM-III-R, pedophilia was defined as involving “recurrent intense sexual urges and sexually arousing fantasies involving sexual activity with a prepubescent child or children.” This criterion excluded individuals who molested children but did not experience recurrent intense sexual urges or fantasies involving sexual activity with a prepubescent child. DSM-IV-TR defined the diagnostic criteria for pedophilia as “recurrent, intense sexually arousing, fantasies, sexual urges, or behaviors involving sexual activity with a prepubescent child or children.” These fantasies, sexual urges, or behaviors are pedophilic if they “cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.” The pedophile also has to be 16 years of age or older and at least 5 years older than the victim child, although, for late adolescents, this age discrepancy is not specified. DSM-5 redefined pedophilia as pedophilic disorder, referring to the following criteria: over a period of at least 6 months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving sexual activity with a prepubescent child or children (generally age 13 years or younger). The individual has acted on these sexual urges, or the sexual urges or fantasies cause marked distress or interpersonal difficulty. The individual is at least 16 years of age and at least 5 years older than the child or children. The disorder is further defined by specifying whether the sexual fantasies, sexual urges, or behaviors are the exclusive type, attracted only to children, or nonexclusive type, attracted to children and adults. Pedophilic arousal is further identified as sexually attracted to males, females, both, or and identified based on whether it is limited to incest.

The term *child molester* is a descriptive, nonclinical term. Some child molesters are driven by pedophilic interests, while others molest children as a result of nonpedophilic interests, such as the unavailability of an appropriate sexual partner. Pedophiles typically seek treatment as a result of legal consequences for illegal behavior.

For most pedophiles, the onset of attraction to children begins at the time of puberty or adolescence. The course of pedophilia is long-term. Individuals who engage in sexual activities with pubescent teenagers under the legal age of consent (ages 16 to 18) are known as *hebephiles* (attracted to individuals at the early stages of puberty) or *ephebophiles* (attracted to pubertal individuals). Most pedophiles do not use physical force to engage a victim in sexual contact; rather, they use manipulation and cajoling to elicit the desired response from the child. Pedophiles commonly justify and minimize the actions by stating that the child enjoyed the experience or that the experience was of educational value.

The majority of pedophiles are male. From federal data on sexual crimes, females were reported to be the molester in 6% of all juvenile cases. A study of 4,007 child molesters found 1% to be female, but the authors of this study suggested the number was low because of the systematic underreporting of women for molestation. Most pedophiles are nonexclusive, meaning that they have a sexual attraction to adults in addition to children. Pedophiles are categorized according to the type of exclusivity, gender of the child they are attracted to, and whether the child is familial (e.g., incest) or nonfamilial. A larger percentage of pedophiles are homosexual or bisexual in orientation to children. Individuals attracted to females usually prefer children between the ages of 8 and 10 years. Individuals attracted to males usually prefer boys between the ages of 10 and 13 years.

Fifty percent to seventy percent of pedophiles can be diagnosed as having another paraphilia, such as frotteurism, exhibitionism, voyeurism, or sadism. Pedophiles are approximately two and a half times more likely to engage in physical contact with a child than simply voyeuristic or exhibitionist activities. Typically, pedophiles engage in fondling and genital manipulation more than intercourse, with these three exceptions: intercourse occurring in cases of incest, occurring in pedophiles with a preference for older children or adolescents, and occurring when children are physically coerced.

The question of whether individuals who view child pornography meet the criteria for pedophilic disorder and are at risk for committing hands-on sexual offenses has been the focus of research. Research has not established a clear relationship between viewing child pornography and child sexual offenses. Studies show the majority of men convicted or charged with child pornography offenses have demonstrated sexual arousal to prepubescent stimuli on phallometric testing.

Although preliminary, imaging techniques such as functional magnetic resonance imaging (fMRI) have attempted to identify variations in brain morphology in individuals with pedophilia. fMRI in combination with emotional and/or sexual stimuli showed pedophilia is likely associated with altered brain activity, particularly in the frontal brain areas and the temporal lobe.

Research into the etiology of pedophilia has been inconclusive. Reported differences between pedophiles and the general population include lower intelligence, a slight increase in the prominence of left-handed individuals, impaired cognitive abilities, neuroendocrine differences, and brain abnormalities, particularly frontocortical irregularities or differences.

Pedophilic disorder refers to the sexual gratification for at least a 6-month period of time from sexual fantasies or behavior related to sexual activity with a prepubescent individual. In order to be considered pedophilic, the individual must be at least 16 years old, and the prepubescent child must be at least 5 years older than the individual. The diagnosis is made when the individual's ongoing sexual urges or fantasies have led to significant distress, impairment in functioning, or actual behavior. The true prevalence of pedophilic disorder is unknown. Studies examining pedophilic interest, not the disorder, show prevalence rates ranging from 2% to 24%, depending on the sample. Research of men convicted of child sexual abuse found that half did not meet the criteria for pedophilic disorder.

Sexual Masochism Disorder

Sexual masochism is a term used to refer to sexual gratification produced by pain that is inflicted by others or by one's self. Sexual masochistic disorder refers to, over a period of at least 6 months, recurrent and intense sexual arousal from the act of being humiliated, beaten, bound, or otherwise made to suffer from resultant clinically significant distress or impairment in social, occupational, or other important areas of functioning. Studies suggest that an estimated 5% to 10% of the population has engaged in some masochistic sex play. Research suggests that sexual masochism is a fairly benign activity. Studies of practicing masochists have found them to be reasonably well-adjusted, successful individuals. If studies describe sexual masochism as a benign activity, is sexual masochism a mental disorder? Proponents of sexual masochism as a mental disorder theorize that an individual may become dependent on masochistic activity or may engage in dangerous or injurious behavior warranting a pathologic aspect of the behavior. In most cases, masochism does not cause harm to individuals. However, sexual masochism disorder with asphyxiophilia, which refers to the practice of achieving sexual arousal derived from the autoerotic restriction of airflow, can cause accidental death. There is no scientific evidence to support the diagnosis of sexual masochism as pathologic. Therefore, if an individual seeks treatment for masochism and there is no harm caused by the behavior, the individual will not benefit from treatment.

Sexual masochism is one of the few paraphilias described in women. As discussed previously, the majority of paraphilias are almost never diagnosed in women. The three main characteristics of masochism are pain, loss of control, and humiliation. One of the psychodynamic theories postulated about masochism is Freud's writings that masochism is derived from sadism: More precisely, the person wishes to dominate and hurt others, but this desire gives rise to guilt, so the person defensively converts it into its opposite, namely a desire to submit to domination by others. However, this theory is not entirely supported by empirical evidence.

Sexual masochism disorder refers to at least a 6-month period of time with ongoing sexual arousal from fantasies, urges, or behaviors related to the physical or psychological abuse one is made to suffer. The sexual urges or fantasies cause significant impairment in functioning, significant distress, or actual behaviors. The true prevalence of sexual masochism disorder is unknown. Studies examining sexual, sadomasochistic interests in nonclinical populations reveal the behavior as fairly common. The prevalence, however, of masochistic disorder has not been adequately studied.

Sexual Sadism Disorder

Sexual sadism is sexual gratification achieved from the infliction of psychological or physical suffering or both on a partner. Sexual sadism disorder is a period of at least 6 months with recurrent and intense sexual arousal from the physical or psychological suffering of another person, and the individual has acted on these sexual urges with a nonconsenting person or the sexual urges or fantasies cause clinically significant distress or impairment in social, occupational, or other important areas of functioning. Sadists experience sexual fantasies involving the suffering of another as well as compulsions to enact fantasies. The sadist's partner may consent to the behavior in an effort to satisfy their own masochistic fantasies, or the partner may be the victim. Sadism, like masochism, has been described in both men and women. Similarly, both sadism and masochism are sexual behaviors that are reported in the context of normal sexuality.

Krafft-Ebing subclassified sexual sadism into several categories, including:

1. Lust-murder;
2. Mutilation of corpses or necrophilia;
3. Injury to females (stabbing);
4. Defilement of women;
5. Other kinds of assaults on women—symbolic sadism in which, for example, the perpetrator cuts the hair of his victims rather than harming them directly;
6. Ideal sadism or sadistic fantasies alone without acts;
7. Sadism with other objects, for example, whipping boys;
8. Sadistic acts with animals.

In some individuals, sadistic fantasies and compulsions lead to rape or homicide. However, not all rapes or sexual homicides are perpetrated by sadists. Sadistic rapists are aroused by the use of gratuitous violence, whereas the preferential rapist is thought not to use greater force than is necessary for the submission of the victim. The sadistic rapist is detected by demonstrating a higher level of arousal to a description of physical but nonsexual assault on female victims when subject to phallometric assessment. One confounder, however, is that normal men acknowledge sexual arousal to the idea of raping or hurting a woman. Among the general male population, 30% report fantasies of tying up and raping a woman.

The majority of sadists report the onset of fantasies in their late teen years. Some sadists' accounts of sexual arousal to violence are based on an association between excitement and punishment in childhood. However, the majority of sadists are unable to delineate a historical marker for the sadistic sexual appetite.

The diagnosis of sexual sadism disorder is used to define a period of at least 6 months of sexual gratification from fantasies, urges, or behaviors related to the harm of a nonconsenting individual by psychological abuse or physical abuse. The diagnosis is made when the individual has acted on the fantasies or experienced sexual fantasies or urges that cause significant distress or impairment in functioning. The true prevalence of sexual sadism disorder is unknown and varies depending on the sample studied. A prevalence rate ranging from 2% to 30% has been reported in forensic samples.

Transvestic Disorder

The literal meaning of the word transvestism is to wear the clothing of the opposite sex. DSM-5-TR uses the term transvestism to refer to a male who dresses in women's clothing with resultant sexual arousal. Transvestic disorder is defined as the recurrent and intense sexual arousal from cross-dressing for over a 6-month period of time in which the fantasies, sexual urges, or behaviors cause clinically significant distress or impairment in social, occupational, or other important areas of functioning. If the cross-dresser is specifically sexually aroused by fabrics, materials, or garments, the specifier "with fetishism" should be added. Cross-dressing typically begins in prepubertal children but does not appear to have a sexual component until late adolescence or early adulthood. Individuals with transvestic disorder are usually heterosexual. However, there is a subgroup of men with transvestic disorder who report gender dysphoria, which can be defined as discontent with one's biologic sex, the desire to possess the body of the opposite sex, and the desire to be regarded by others as a member of the opposite sex. Initially, the presence of gender dysphoria ruled out the diagnosis of transvestic disorder according to DSM-5-TR. However, additional research suggests the presence of gender dysphoria is not exclusionary in the diagnosis of transvestic disorder. DSM-5-TR specifier "with autogynephilia" refers to individuals who are sexually aroused by thoughts or images of self as a female. Autoandrophilia is used to refer to sexual arousal to the thought or image of oneself as male.

It is unclear whether transvestism causes the level of dysfunction typical of mental disorders. Clinical evidence suggests that transvestic disorder impairs the ability to form intimate pair bonds with other persons in the sense that relationships tend to break down because the individual with transvestic fetishism often has more interest in the nonhuman aspects of the partner (e.g., clothing). However, not all cases of transvestism exhibit dysfunctional relationships or distress.

The diagnosis of transvestic disorder describes a period of at least 6 months of ongoing sexual arousal from sexual fantasies, urges, or behaviors-related to dressing in the clothing of the opposite gender. Sexual arousal results in significant distress or impairment in functioning. The true prevalence of transvestic disorder is unknown. In nonclinical samples, 3% of men and 0.4% of women describe one or more incidents of transvestic fetishism. The prevalence of transvestic disorder in clinical populations is unknown.

Voyeuristic Disorder

Voyeurism is sexual gratification achieved from watching people undressing or nude or observing them during sexual acts without their knowledge or consent. Voyeuristic disorder is qualified by recurrent and intense sexual arousal from observing an unsuspecting person who is naked, in the process of disrobing, or engaging in sexual activity over a period of at least 6 months. In addition, to meet the criteria for voyeuristic disorder, the individual is at least 18 years of age and has acted on these sexual urges with a nonconsenting person, or the sexual urges or fantasies cause clinically significant distress or impairment in social, occupational, or other important areas of functioning. Voyeurs are usually men in their 20s or 30s who prefer to watch unknown women. They may masturbate while observing the victims but typically seek no contact with the victim. Some voyeurs report sexual fantasies that the woman they are watching will solicit sex from them. The onset of voyeuristic behavior is usually before the age of 15 years. Voyeurs tend to have multiple paraphilias and to cross over from one paraphilia to another. Although some voyeurs do not commit more serious paraphilic acts, there are some suggestions in the literature that serial rapists and murderers began with histories of voyeurism. Voyeurism is one of the least understood behaviors of paraphilias.

The diagnosis of voyeuristic disorder refers to a 6-month period of ongoing sexual gratification from observing an unsuspecting individual who is undressing, naked, or engaging in sexual activity. In order to meet the diagnostic criterion for the diagnosis of voyeuristic disorder, an individual must be at least 18 years old and either engaged in the actual behavior or experienced significant distress or impairment in functioning as a result of the specific sexual arousal. The true prevalence of voyeuristic disorder is unknown. Nonclinical studies of paraphilic interests rate voyeurism and frotteurism as the most popular. In these studies, both men and women report voyeuristic interests at the rate of 12% in men and 4% in women. Lower satisfaction with life, substance use, high sexual interest, more sexual partners, and psychological problems are associated with voyeuristic behaviors.

PATHOLOGY AND LABORATORY EXAMINATION

Diagnostic Tests

The diagnosis of a paraphilic disorder is made based on data gathered during the psychiatric evaluation. There is no definitive test for paraphilic disorders. Psychological testing is not helpful in such diagnoses. Objective tests of sexual interests may be helpful in identifying sexual interests. Objective tests include sexual history polygraphs, visual reaction time tests, and penile plethysmography (phallometric testing). Polygraphs are not routinely used in the clinical setting but are more common in mandated sex offender treatment. Visual reaction time tests identify sexual interest based on the amount of time an individual views images of males and females of various ages.

Plethysmography is the phallometric assessment of physiologic sexual arousal by use of equipment that measures penile circumference in males. The test is conducted by placing a small circular strain gauge around the penis. This gauge is calibrated to measure changes in the circumference of the penis. The test subject is shown both audio and visual stimuli depicting various sexual vignettes. As the individual becomes aroused by the stimuli, the erect penis displaces the gauge, and a measurement is recorded. Phallometric testing helps elucidate an individual's sexual interest without depending on self-disclosure. One of the difficulties working with individuals who have paraphilias is the individual's reluctance to report his sexual interests.

Objective testing is one way of addressing the difficulty of self-disclosure in paraphiliacs. In addition to diagnostic evaluations, objective testing is helpful in determining an individual's response to treatment.

DIFFERENTIAL DIAGNOSIS

The diagnosis of a paraphilia begins with the identification of the individual's sexual interest. It is important to understand an individual's sexual preference before the diagnosis of paraphilia can be made. One of the common errors made in the area of paraphilias is to make a diagnosis based on an individual's behavior rather than on the individual's sexual preference. For example, the differential diagnosis for a child molester may include pedophilia, personality disorder, mental retardation, or substance abuse. In order to determine what diagnosis fits the child molester, one has to understand what motivated the molester to have sexual contact with the child. If the molester describes sexual urges and fantasies of sexual contact with a prepubescent individual, they meet the criteria for pedophilia. However, if the molester reports a history of learning difficulties consistent with intellectual disability and explains his molesting behavior as his desire to "have sex with anyone," the diagnosis is an intellectual disability and possibly personality disorder. The differential diagnosis of paraphilias is simple if one understands what the sexual motivations of the subject are.

Like all psychiatric symptoms, it is important to determine if sexual behaviors or fantasies are better explained by a general, nonparaphilic disorder. For example, an individual who presents with variable sexual interests in the setting of psychosis or mood episodes may not meet the criteria for a paraphilic disorder. The clue to a nonparaphilic presentation is the emergence of sexual interests later in life in settings of psychiatric decompensation. Paraphilic interests may wax and wane but rarely develop late in life. Medical conditions such as traumatic brain injuries, cerebrovascular accidents, and neurodegenerative syndromes such as Huntington disease may present with new-onset sexual symptoms. Dopaminergic agents have also been associated with hypersexuality and other paraphilic symptoms.

COURSE AND PROGNOSIS

Paraphilic behaviors emerge in adolescence and early adulthood. In general paraphilic behaviors are chronic. Although it has been shown that individuals may cross over from one paraphilia to another, the overall course of paraphilic behavior is chronic. Because of the early onset and repudiated biologic priming of the paraphilias, they are often difficult to treat.

TREATMENT BACKGROUND

The efficacy of treatment in paraphiliacs is unknown. Empirical evidence for the treatment of individuals with paraphilias is derived from studies of sex offenders. Although studies comparing treated and untreated sex offenders have been done, measurement of outcome is flawed, with recidivism rates underestimating actual recurrence of the problematic behavior. Methodologic challenges such as sample selection, study design, and assessment of outcome represent significant problems in treatment efficacy studies. The majority of treatment efficacy studies are conducted on incarcerated or civilly committed offenders. Those individuals who are incarcerated or civilly committed represent a unique population with treatment outcomes that may not generalize to the nonincarcerated or community-residing sex offender. Many studies do not differentiate between types of offenders. For example, a study that evaluates child molesters may not differentiate between intra- and extrafamilial offenders. Research suggests that incest offenders recidivate at approximately half the rate of extrafamilial child molesters.

A well-designed study of treatment efficacy in sex offenders would compare treated offenders with a matched, randomly assigned control group. However, the ethical and public safety implications of such a study impede its design. Sex offenders are extremely heterogeneous, and when this is not recognized, it is very difficult to make comparisons among treatment efficacy studies.

Biologic Treatment

The scientific basis of biologic treatment in paraphiliacs is the reduction of sexual behaviors by decreasing testosterone levels. Testosterone is a steroid hormone produced by the testes that are responsible for the development of secondary sex characteristics in men. Testosterone helps maintain sex drive, the production of sperm cells, male hair patterns, muscle mass, and bone mass. Testosterone strongly influences both male and female sexual drive and the resultant sexual behavior. Animal studies show that male copulatory behavior is almost entirely dependent on circulating levels of plasma testosterone.

The production of testosterone is mediated by FSH and luteinizing hormone (LH), two hormones produced by the pituitary gland. LH signals the testes to produce testosterone. The hypothalamus senses whether the production of testosterone is too much or too little. If the testes are producing too little testosterone, the hypothalamus secrets a hormone called GnRH, which signals the pituitary to increase production of LH, which results in an increase in testosterone production. If the testes are producing too much testosterone, the hypothalamus secretion of GnRH is inhibited. This process, whereby the hypothalamus modulates the production of testosterone, is called negative inhibition.

Surgical Treatment

Surgical treatment for sex offenders consists of two types: neurosurgery and castration. The neurosurgical procedure involves stereotaxic removal of parts of the hypothalamus to disrupt the production of male hormones and decrease sexual arousal and impulsive behaviors. This procedure had significant adverse effects and was considered largely ineffective. Surgical castration is the removal of the testes. The effect of surgical castration is to globally reduce available androgens by the removal of the testes. Sturup followed 107 castrated sex offenders and compared them to 58 who were not castrated over a period of 18 years. The castrated individuals recidivated at a rate of 4.3%. The 58 uncastrated individuals recidivated at a 43% rate. Postcastration follow-up studies provide the most comprehensive outcome data on the effect of reducing plasma testosterone and the resultant suppression of the sexual drive and deviant sexual behavior. These studies have reported recidivism rates of less than 5% with follow-up periods of up to 20 years in a large sample size.

The introduction of antiandrogen and hormonal medications has rendered surgical castration nearly obsolete. The effects of surgical castration can be achieved through chemical castration (i.e., the use of medications to decrease testosterone production) without the invasiveness and irreversibility of surgery. Some states in the United States, such as Texas and California, mandate chemical or surgical treatment of dangerous sexual offenders. Controversy exists as to whether surgical castration should be offered as a treatment, as chemical castration achieves the same results and spares the procedure.

Pharmacologic Treatment

Antiandrogens. Androgens are male sex hormones that promote the development and maintenance of male sex characteristics. Testosterone is the most common androgen in man. Antiandrogen treatment refers to treatment with drugs used to block the production or interfere with the action of male sex hormones. Cyproterone acetate (CPA; Androcur) and medroxyprogesterone (MPA; Depo-Provera) are the two most commonly used antiandrogen medications. CPA is not available in the United States but is widely available in

Canada and Europe. Both CPA and MPA are synthetic progesterone that reduce the serum level of testosterone. Reduction of testosterone has been shown to reduce libido, erections, ejaculations, and spermatogenesis. The World Federation of Societies of Biological Psychiatry (WFSBP) concluded in their guidelines for the biologic treatment of paraphilias with MPA that there is minimal research-based evidence to support this recommendation. They also reported severe side effects observed with MPA, concluding that the benefit/risk ratio did not favor the use of MPA. WFSBP similarly concluded that there is minimal research-based evidence to support treatment with CPA.

The risks associated with the use of antiandrogen agents include side effects such as weight gain, hyperglycemia, hot, and cold flashes, liver dysfunction, hypertension, muscle cramps, phlebitis, gastrointestinal complaints, and feminization. In addition, there is little known about the long-term sequelae of antiandrogen treatment. The majority of antiandrogen studies follow patients for 8 years or less. The patient who agrees to antiandrogen treatment should be educated about the absence of long-term data.

Hormonal Agents

Leuprolide (Eligard) and triptorelin (Trelstar) are hormonal agents referred to as long-acting GnRH agonists. These agents inhibit the secretion of LH with a resulting decrease in plasma testosterone levels and libido. They produce a chemical castration in that the hypothalamic–pituitary axis is exhausted, and there is a potent inhibition of gonadotropin. A meta-analysis conducted by WFSBP concluded that there were no controlled studies, and some biases were observed. However, the efficacy observed in these open studies was very high, and in most cases, subjects were previously treated with psychotherapy or other antiandrogens without efficacy. The WFSBP reported that despite the lack of controlled studies, the efficacy of hormonal agents was observed in open studies in patients with treatment-refractory and or severe paraphilic disorders. A Cochrane analysis of testosterone-lowering medications concluded that the efficacy of such medications was uncertain and that further research was necessary.

The side effects associated with GnRH agonists include decreased bone mineral density or osteopenia, weight gain, hyperglycemia, diabetes, hypertension, and insomnia. The most commonly reported side effects are erectile/ejaculatory problems and gynecomastia. In general, patients reported fewer side effects on GnRH agonists when compared to antiandrogens. The preliminary studies of GnRH agonists suggest that leuprolide and triptorelin may be more effective, better-tolerated alternatives to antiandrogen treatment.

Selective Serotonin Reuptake Inhibitors

Although the pharmacologic mechanism is poorly understood, SSRIs have been shown to be effective in reducing paraphilic symptoms. Animal models showed that decreased serotonin levels increased sexual appetitive behaviors while enhanced central serotonin activity reduced them. Clinical observations of the similarities between paraphiliac fantasy urges or behaviors with obsessive-compulsive symptoms led to hypotheses that SSRIs may be efficacious in paraphilias as they have been in OCD. Case reports have reported the positive effects of SSRIs in the treatment of paraphilias. Most of the studies examining the efficacy of SSRIs in paraphilias are case reports or small sample sizes and do not include placebo-controlled double blinds studies. The WFSBP published the Guidelines for the biologic treatment of paraphilias in 2010 and again in 2020. These guidelines conclude that “a critical analysis of all studies that involved the use of SSRIs in the treatment of paraphilias concluded that the results of psychotropic drug interventions are not favorable.” In summary, although data exist to suggest that SSRIs are effective in the treatment of paraphilias, presently, there are insufficient data to conclude that SSRIs are equally efficacious as antiandrogens or hormonal agents.

Other psychotropic medications, including anticonvulsants, lithium, and tricyclic antidepressants, have not been shown to be efficacious in paraphilic disorders.

The WFSBP has proposed an algorithm for the pharmacologic treatment of paraphilic disorders. This algorithm assigns treatment based on the level of severity of the paraphilia. The less severe paraphilias are treated with psychotherapy, while the more severe paraphilias add biologic treatment starting with SSRIs and moving to testosterone-lowering medications.

Psychological Treatment

Psychological approaches to the treatment of sexual offenders focus on direct treatment approaches to modify offenders' cognitions and attitudes. Cognitive approaches described the concept of cognitive distortions (minimizations, excuses, justifications) as one factor that may maintain offender behavior. Relapse prevention and the sexual assault cycle are theoretical models purporting to describe the cognitions and behaviors of sex offenders before, during, and after their abusive acts. Relapse prevention models of treatment were the mainstay of treatment for many years. However, in the past decade, these models have largely been abandoned for strength-based rehabilitation programs. The two most common strength-based rehabilitation programs that are currently employed in sex offender treatment are the Good Lives Model, and the Risk Need Responsivity Model. Research has consistently demonstrated that structured behavioral or cognitive-behavioral treatment designed to address offending behavior is more effective in reducing reoffending than didactic, experimental, nondirective, or psychodynamic modes of treatment.

Current sex offender treatment programs have focused on identified dynamic risk factors for sexual offending. These treatment programs are adopted for the treatment of paraphilic disorders since evidenced-based psychological treatments for nonoffending paraphiliacs do not exist. These risk factors, which serve as the focus of treatment, include the following: insecure attachments, lack of intimacy, emotional loneliness, poor self-regulation, sexual preoccupation, deviant sexual interest, emotional congruence with children, lack of concern for others, attitudes supportive of offending women, or hostility toward them. Treatment efficacy has also been measured as it relates to the qualities of the therapist. Therapists who are nonjudgmental and empathic deliver the most effective treatment regardless of the orientation of the treatment program.

Recidivism rates from cognitive and behavioral treatment programs range from 3% to 31%, depending on the study. A meta-analysis found that cognitive-behavioral treatment and antiandrogen treatment were comparable in their treatment effects and significantly more effective than behavioral treatment alone.

MODELS OF TREATMENT

The treatment of individuals with paraphilias and sex offenders has primarily focused on two areas: Cognitive-behavior group treatment and pharmacologic treatment. Treatment programs target risk factors empirically associated with sexual offending and recidivism. Meta-analysis has shown that cognitive-behavioral treatment is most effective in reducing sex offender recidivism. Treatment models have changed over the years as research in the field of sex offender management has evolved. Currently, cognitive-behavioral treatment remains the most widely accepted and empirically supported model of sex offender treatment. Best practices include the use of actuarial tools and data-driven instruments in the determination of risk. Research demonstrates that actuarial methods are consistent in predicting the relative risk of sex-offending behavior and are considered more accurate than unstructured professional judgment. A research query asking sex offenders in treatment about acceptable types of treatment revealed the preferred therapies were individual psychotherapy, social skills training, and group therapy, whereas aversion conditioning, castration, and sex drive-reducing drug therapy were the least acceptable forms of treatment.

ETHICAL AND LEGAL CONSIDERATIONS

The legal and ethical issues that arise in the course of sex offender treatment are unique to this population. As such, the traditional codes of ethics employed in medical treatment are not applicable to the treatment of sex offenders. A significant percentage of sex offenders receiving treatment have been mandated by the courts to do so as part of sentences of incarceration or release into the community. The concept of mandated or involuntary treatment raises the issue of whether informed consent is possible in sex offender treatment. A condition of informed consent for medical treatment is that the consent must be voluntary. Obviously, court-ordered sex offender treatment is not voluntary. Individuals who reject treatment are subject to punishment imposed by the courts. Individuals who have been court-ordered to receive treatment do not have a choice regarding the type of treatment or the treatment provider. Sex offenders are required to complete particular programs, irrespective of any other treatment they might be receiving, in order to gain community release or avoid imprisonment.

In the medical model, patients have a right to refuse various types of treatment. The right to refuse treatment is based on an individual's constitutional right to privacy. The court views sex offenders as incompetent patients, and as such, the court becomes the decision-maker regarding treatment.

Confidentiality and Privilege

Confidentiality refers to the physician's obligation to keep information learned in a professional relationship private from other parties. Privilege refers to the patient's right to prevent a physician from providing testimony about personal medical information. Both confidentiality and privilege are routinely breached in sex offender treatment. When individuals enter treatment, they are required to give permission for their cases to be discussed with both clinical and nonclinical personnel, correctional officers, members of their family, past and potential victims, and those associated with them and fellow offenders.

In response to the deviations from traditional ethical codes inherent in sex offender treatment, ethical and practice guidelines were developed by the association for the treatment and prevention of sexual abuse (ATSA). ATSA's code of ethics endorses standards of professional conduct that promote competent practice, and as such, they represent a public commitment to clients and society toward the goal of preventing sexual violence. The ATSA guidelines state that the ethical care of sex offenders is achieved by encouraging individuals to take responsibility for their behavior, that is, admission of guilt. ATSA maintains that the identification and collaborative management of risk and safety factors are indeed in the best interests of both sex offender patients and potential victims due to the grave consequences incurred by sexual offender recidivism.

Castration of Sex Offenders

In 1996, California became the first state to authorize the use of either chemical or surgical castration for certain sex offenders who were being released from prison into the community. To date, an additional eight states authorize surgical or chemical castration for sex offenders, including Georgia, Montana, Oregon, Wisconsin, Florida, Iowa, Louisiana, and Texas. The APA stated that “these laws, which predicate release from prison on chemical castration by surgery or antiandrogenic agents, are objectionable because they are not based on adequate diagnostic and treatment considerations. They also improperly link medical treatment with punishment and social control.” In some states, informed consent for the procedure is not required. In other states, informed consent requires only that the offender be informed regarding the side effects.

FUTURE DIRECTIONS

Further Research

Although studies have demonstrated the effectiveness of some modalities of sex offender treatment, the scientific literature does not identify a definitive effective treatment. The literature suggests that both cognitive-behavioral therapies and antiandrogen medication decrease sex offender recidivism. A decrease in sex offender recidivism, however, does not translate to a decrease in paraphilic behaviors. However, the current sex offender legislation in the United States is not based on the literature. Rather the legislation focuses on the preventive detention of sex offenders without concurrent evidence-based sex offender treatment. The future of research in the treatment of sex offenders and paraphilias lies in the development of evidence-based knowledge that will inform both clinical decisions and public policy. This effort may be afforded by interdisciplinary collaboration between the scientific community, correctional professionals, and policymakers.

Specialized Training

Many psychiatrists are unfamiliar with the fundamentals of the assessment and treatment of individuals with paraphilias. The advent of laws for the civil commitment of sex offenders has resulted in the need for psychiatrists trained in paraphilias. As a result of sex offender legislation, the treatment and release decisions regarding sex offenders have become a focus of attention for psychiatry. Forensic psychiatrists or psychologists must render an opinion as to whether the sex offender has a diagnosed mental disorder and, as such, represents a risk to public safety if released from custody into the community. In order to provide competent care to individuals with paraphilias, psychiatrists should be educated about the evidence-based approach to the treatment and evaluation of this population.

Education of the Public

The general public perceives sex offenders as a homogeneous group of offenders who are not treatable. However, recent meta-analyses have demonstrated that sex offenders are a highly heterogeneous population with different rates of reoffending and responsiveness to treatment. The successful treatment of sex offenders is dependent, in part, on the support of the public and the courts. In order to implement an evidence-based treatment, the offender must be afforded the opportunity for treatment. The opportunity for treatment is created by an informed society that legislates treatment that is informed by scientific evidence. The education of the public is integral to the successful treatment of sex offenders.

FURTHER READINGS

*American Psychiatric Association Task Force on Sexually Dangerous Offenders. *Dangerous Sex Offenders: A Task Force Report of the American Psychiatric Association*. American Psychiatric Association; 1999.

Bancroft J. *Human Sexuality and Its Problems*. Churchill Livingstone; 1989.

Birgden A, Cucolo H. The treatment of sex offenders: evidence, ethics, and human rights. *Sex Abuse*. 2011;23(3):295–313.

Bradford JM. The neurobiology, neuropharmacology, and pharmacological treatment of the paraphilias and compulsive sexual behavior. *Can J Psychiatry*. 2001;46(1):26–34.

Briken P, Bourget D, Dufour M. Sexual sadism in sexual offenders and sexually motivated homicide. *Psychiatr Clin N Am*. 2014;37:215–230.

Briken P, Hill A, Berner W. Pharmacotherapy of paraphilias with long-acting agonists of luteinizing hormone-releasing hormone: a systematic review. *J Clin Psychiatry*. 2003;64(8):890–897.

Briken P, Kafka MP. Pharmacological treatments for paraphilic patients and sexual offenders. *Curr Opin Psychiatry*. 2007;20(6):609–613.

Brooks-Gordon B, Bilby C, Wells H. A systematic review of psychological intervention for sexual offenders. I. Randomised control trials. *J Forensic Psychiatry*. 2006;17(3):442–466.

Castellini G, Rellini AH, Appignanesi C, et al. Deviance of normalcy? The Relationship among paraphilic thoughts and behaviors, hypersexuality, and psychopathology in a sample of university students. *J Sex Me*. 2019;15(9):1322–1335.

Dawson SJ, Bannerman BA, Lalumière ML. Paraphilic interests: an examination of sex differences in nonclinical sample. *Sex Abuse*. 2016;28(1):20–45.

Donovan BT. *Hormones and Human Behavior*. Cambridge University Press; 1984.

Grossman LS, Martis B, Fichtner CG. Are sex offenders treatable? A research overview. *Psychiatr Serv*. 1999;50(3):349–361.

Hanson RK, Bussière MT. Predicting relapse: a meta-analysis of sexual offender recidivism studies. *J Consult Clin Psychol*. 1998;66(2):348–362.

Johnson RS, Ostermeyer B, Sikes KA, Nelsen AJ, Coverdale JH. Prevalence and treatment of frotteurism in the community: a systemic review. *J AM Acad Psychiatry Law*. 2014;42:478–83.

Joyal CC, Carpentier J. The prevalence of paraphilic interests and behaviors in the general population: a provincial survey. *J Sex Res*. 2017;54(2):161–171.

Kafka M. Axis I psychiatric disorders, paraphilic sexual offending and implications for pharmacological treatment. *Isr J Psychiatry Relat Sci*. 2012;49(4):255–261.

Krueger RB, Kaplan MS. Depot-leuprolide acetate for treatment of paraphilias: a report of twelve cases. *Arch Sex Behav*. 2001;30:409–422.

Långström N. The DSM Diagnostic criteria for exhibitionism, voyeurism, and frotteurism. *Arch Sex Behav*. 2010;39:317–324.

Långström N, Zucker KJ. Transvestic fetishism in the general population: prevalence and correlates. *J Sex Marital Ther*. 2005;31(2):87–95.

Laws R, O'Donohue W. *Sexual Deviance: Theory Assessment and Treatment*. Guilford; 1997.

Levenson J, D'Amora J. An ethical paradigm for sex offender treatment: response to Glaser. *West Criminol Rev*. 2003;4(2):143.

Makhlouf Y, Kerbage H, Khauli N, Ramadan Z, Khoury B. Legal and policy considerations in Lebanon related to proposals for paraphilic disorders in the World Health Organization's International Classification of Diseases Manual, 11th version. *J Sex Med*. 2020;17:491–504.

Rosler A, Witztum E. Treatment of men with paraphilia with a long-acting analogue of gonadotropin-releasing hormone. *N Engl J Med*. 1998;338:416–422.

Saleh FM. Serotonin reuptake inhibitors and the paraphilias. *Am Acad Psychiatry Law Newsletter*. 2004;29(3):12–13.

Savoie V, Quayle E, Flynn E. Prevalence and correlates for individuals with sexual interest in children: a systemic review. *Child Abuse Negl*. 2020;115:1–28.

Scott CL, Holmberg T. Castration of sex offenders: prisoners' rights versus public safety. *J Am Acad Psychiatry Law*. 2003;31:502–509.

Sreenivasan S, Weinberger LE, Garrick T. Expert testimony in sexually violent predator commitments: conceptualizing legal standards of "mental disorder" and "likely to reoffend." *J Am Acad Psychiatry Law*. 2003;31(4):471–485.

Seeman MV. Portrait of Exhibitionist. *Psychiatr Q*. 2020;91:1249–1263.

Seto MC, Cantor JM, Blanchard R. Child pornography offenses are a valid diagnostic indicator of pedophilia. *J Abnorm Psychology*. 2006;115:610–615.

Sturup GK. Sexual offenders and their treatment in Denmark and other Scandinavian countries. *Int Rev Criminal Policy*. 1953;4:1–19.

Thibaut F, De La Barra F, Gordon H, Cosyns P, Bradford JM, WFSBP Task Force on Sexual Disorders. The World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for the biological treatment of paraphilias. *World J Biol Psychiatry*. 2010;11:604–655.

Ventriglio A, Bhat PS, Torales J, Bhugra D. Sexuality in the 21st century: Leather or rubber? Fetishism explained. *Medical J Armed Forces India*. 2019;75:121–124.

Ward T, Salmon K. The ethics of care and treatment of sex offenders. *Sex Abuse*. 2011;23(3):397–413.

Watts AL, Nagel MG, Latzman RD, Lilienfeld SO. Personality Disorder features and paraphilic interests among undergraduates: differential relations and potential antecedents. *J Pers Disord*. 2019;33(1):22–48.

Zurbriggen EL, Yost MR. Power, desire, and pleasure in sexual fantasies. *J Sex Res*. 2004;41:288–300.

21.4 Gender Identity, Gender Diversity, Gender Dysphoria/Incongruence

Jack Drescher, M.D., and William M. Byne, M.D., Ph.D.

Variations of gender expression have been documented throughout history, in myth, literature, and in actual practice. Starting in the 19th century, gender-diverse presentations became a subject of medical and psychiatric interest. Nonconforming expressions of gender were initially conflated with and labeled as “homosexuality.” However, by the mid-20th century, the modern concept of gender identity was teased out from that of sexual orientation (see homosexuality 21.2). After that happened, diagnoses of “transsexualism” and “gender identity disorder” began to appear in diagnostic manuals, along with homosexuality.

With the publication of the fifth edition of the Diagnostic and Statistical Manual (DSM-5) in 2013 and its text revision (DSM-5-TR) in 2022, a less stigmatizing diagnosis of *gender dysphoria* (GD) replaced *gender identity disorder* (GID). In the 11th edition of the International Classification of Diseases and Related Health Problems (ICD-11), *transsexualism* was replaced by *gender incongruence* (GI), and, for the first time, the diagnosis was moved out of the ICD Mental Disorders section. These diagnostic changes came amidst a growing social acceptance of gender-diverse and transgender individuals, which led to a rethinking of early psychopathologizing theories and formulations about them.

In recent years, however, the subject of what exactly constitutes socially acceptable expressions of a person’s gender identity has become a subject of popular interest, debate, and concern. Most troubling, the issue has been increasingly politicized in the forms of “bathroom bills,” laws intended to prevent transgender individuals from using restroom facilities that accord with their gender identity; proposed and enacted legislation making it a felony for clinicians to provide medical treatment to transgender minors; and laws prohibiting transgender youth from playing on high school sports teams that accord with their gender identity and expression.

In other words, questions have been raised not only about what constitutes appropriate expressions of gender but also who ultimately is responsible for deciding what is and is not appropriate: Psychiatrists? Religious authorities? Civil rights and human rights advocates? Complicating matters further, not all cultures endorse *binary* notions of gender constituted simply of "men" and "women," Native American "two-spirit people" *(berdache)* or *hijras* of India being notable examples. In the context of cultural shifts in contemporary discussions of gender, this chapter attempts to provide a more expansive view of gender diversity and gender identity in addition to discussing psychiatric aspects of diagnosing GD/GI and its treatment.

DEFINITIONS

Although *sex* and *gender* are often used interchangeably, *sex* refers to the status of biologic variables that can be described as either male-typical or female-typical in normatively developed individuals (e.g., genes, chromosomes, gonads, internal and external genital structures). Prior to the mid-1950s, the word *gender* had been employed primarily to refer to classes of nouns in languages; however, John Money and his colleagues introduced the concept of *gender* in reference to attributes of people; specifically, those attributes that are associated with the sex assigned at birth but not necessarily determined by it. For example, they observed that most individuals born with somatic variations of sex development (historically referred to as hermaphroditism, pseudohermaphroditism, and intersex conditions but now referred to as disorders/differences of sex development (DSD) often live and think of themselves as males or females according to the sex they were assigned by physicians and parents in infancy and independent of the status of any particular genetic, endocrinologic, or physical variable of sex.

Gender, therefore, came to refer to the social categories of males (i.e., boys, men) and females (girls, women) as well as to factors related to living in the social role of a man or a woman. More recently, however, sex and gender tend to be interpreted synonymously in medicolegal contexts. The sex initially assigned to individuals with a DSD is referred to as their *sex assigned at birth* (or *gender assigned at birth*), a term which has increasingly been used to refer to the sex/gender initially assigned to all individuals, but particularly in discussing the developmental histories of transgender individuals. Another common term in the literature, *natal gender,* whose usage has fallen out of favor in recent years, was interchangeable with both *sex* and *gender assigned at birth.*

Gender identity, a term first coined by psychoanalyst Robert Stoller in the mid-20th century, came to signify one's persistent inner sense of belonging to either the male or female gender category. Today, there are other recognized categories of gender identity that are neither male nor female. In earlier times, from a more binary perspective, John Money defined *gender role* as those things a person says or does to disclose themselves as having the status of boy or man, girl or woman, respectively (e.g., general mannerisms, mode of dress, interests, pastimes, deportment, and demeanor; spontaneous topics of talk in unprompted conversation and casual comment). *Gender role* was conceptualized as the outward expression of the inner sense of *gender identity.* However, the terms *experienced gender* and *expressed gender,* respectively, have increasingly replaced gender identity and gender role, particularly in the legal and political arenas concerned with antidiscrimination protections. Again, although initial descriptions of gender role and identity implied two discrete gender categories, gender is increasingly regarded as more *fluid,* reflecting either a continuous dimension or even a mosaic of masculinity/femininity as opposed to a dichotomy of masculine and feminine. Thus, gender identity/experienced gender may be described as male (boy/man), female (girl/woman), somewhere in between, some combination of both, or neither. Gender role/expressed gender may be described as masculine, feminine, or nonbinary. Complicating matters further, notions of what constitutes *masculine* or *feminine* vary across individuals, cultures, and times.

Transgender is a popular term coined by Virginia Prince in the early 1970s, referring to people who lived full time in a gender that was not the one that usually went with their genitals. In recent decades, it has become an umbrella term that usually refers to someone whose gender identity or gender expression and gender assigned at birth are discordant or not conforming to stereotypical social norms. Being transgender may be associated with a desire to transition to or live as a gender different from the gender assigned at birth. Other identities could include *genderqueer, gender nonconforming (GNC), bigender, nonbinary, two-spirit,* or *eunuch*. In recent years, there has been a public proliferation of gender identities. In addition, members of the transgender community have coined the term *cisgender* (based on the concept of cis/trans isomerism in organic chemistry) as a way to describe nontransgender individuals whose gender identity and birth-assigned sex are in alignment. Some cisgender people publicly embrace this identity, but many simply think of being cisgender as "normal."

The concepts of changing the sexual characteristics of the body and transitioning gender role date back several millennia, notably found in the Greek myth of Tiresias, who was born with the body of a man but changed by the goddess Hera into a woman for 7 years. Of more recent historical vintage, the protagonist of Virginia Woolf's novel *Orlando* goes to sleep as a man and inexplicably wakes up as a woman. Today, modern medical and surgical practices, as well as changing social mores, have shifted the prospect of changing the sexual appearance of one's body from a subject of mythology and fiction to a highly specialized clinical reality. Historically, the term *transsexual* referred to an individual who employed hormonal and/or surgical means to modify the body, so it conformed to their experienced gender. In that nosology, someone who had both gender-affirming hormonal as well as bottom and top surgical treatments was said to have made a *complete transition.* An individual who had only hormones or only surgery was said to have had a *partial transition.* An individual assigned male at birth who transitioned to live as a woman was called a male-to-female (MTF) individual, an MTF transsexual or, in contemporary language, a *transgender woman, a transwoman, or woman of transgender experience.* An individual assigned female at birth who transitioned to live as a man was called a female-to-male (FTM) individual, an FTM transsexual, or, again in today's terminology, *a transgender man, transman,* or *man of transgender experience.*

Historically referred to as *sex reassignment surgery,* more recently, these surgical transition procedures have been referred to as *gender reassignment surgery*, *gender confirmation surgery, and most gender-affirming surgery (GAS).* Because this chapter addresses surgeries done in the context of affirming the experienced gender identities of adolescents and adults, this chapter uses GAS in reference to these procedures. Surgical options include *top surgery* (breast augmentation for transgender women, mastectomy, and male breast construction for transgender men) and bottom surgery (gonadectomy and genitoplasty for transgender women, genitoplasty, gonadectomy, and hysterectomy for transgender men). Transgender women often undergo facial electrolysis, while some transgender individuals elect surgeries for facial and voice feminization or masculinization. Some individuals may not desire to undergo all of these surgical procedures, while financial or medical contraindications may prevent others from obtaining any or all of the procedures they desire.

Similarly, not all individuals who identify as transgender desire, seek, or obtain hormonal or surgical treatments for transition. To find authentic gender expression, some may choose to transition in some respects but not others. That is, the transition may just be *social,* wearing the clothing or accessories of their experienced gender, or it may entail taking gender-affirming hormones and/or undergoing surgery to acquire secondary sexual characteristics of their experienced gender. Some individuals who identify as transsexual object to being classified under the transgender umbrella, believing that those who experience a need to transition hormonally and/or surgically belong to a category distinct from those who do not. It is good clinical practice to ask patients which terms they use and how they wish to be addressed.

There is a great deal of popular confusion between being transgender and being homosexual. However, transgender identities, in and of themselves, do not necessarily reveal anything about one's sexual orientation. Whether one thinks of oneself as a man or woman does not predict whether they are attracted to men or to women. Additionally, the terms *homosexual* and *heterosexual* can lead to confusion when there are discrepancies between an individual's birth-assigned sex and experienced/expressed gender. Should the terms apply to an individual's pre- or post-transition identity? In such situations, sexual orientation may be better described, in conversations among professionals, as *androphilic* (attracted to men), *gynephilic* (attracted to women), bisexual, or *analophilic* (attracted to neither). For example, many transgender women who are sexually attracted to women might be described to professional colleagues as gynephilic or attracted to women; however, in conversations with patients, their preferred sexual orientation labels should be used.

The presence of marked incongruence between one's experienced/expressed gender and birth-assigned gender is listed among the criteria for the DSM-5 Text Revision (DSM-5-TR) diagnosis of *GD in adolescents and adults.* Importantly, *GD,* historically defined as unhappiness or a sense of inappropriateness with one's birth-assigned sex or the gender role associated with that sex, must cause clinically significant distress or impairment in order to satisfy the DSM diagnostic criteria of a disorder. There is also a DSM-5-TR diagnosis of *gender dysphoria in childhood* (GDC). Longitudinal studies have found that many individuals meeting criteria for GDC do not meet criteria for GD as adults. Instead, they may have a range of gender identities different from their sex assigned at birth but no longer meet criteria for GD (e.g., they may no longer meet the distress criterion or seek transition services), and others will identify as cisgender heterosexual adults, while many will identify as gay, "homosexual," or bisexual in adolescence or adulthood. However, most homosexual adults would not have met the criteria for GD as children. In the ICD-11, the gender diagnoses are *GI of adolescents* *and adults* and *GI of childhood.* To make these diagnoses does not require any distress or dysfunction criteria. For that and other reasons, the diagnoses were moved out of the Mental Disorders section of ICD into a new parent category: *conditions related to sexual health* (see this chapter's section on Diagnoses).

Transvestism refers to sexual urges and fantasies involving cross-dressing. When the behavior is distressing to a patient and brought to the attention of a psychiatrist, it may be diagnosed as *transvestic disorder,* which DSM-5-TR classifies as a *paraphilic disorder*. Cross-dressing is strongly linked in the popular imagination with homosexuality. However, in DSM-IV-TR, the diagnosis of *transvestic fetishism* could only be made in heterosexual men as most fetishistic cross-dressers are heterosexual men who cross-dress in private. These men are often married to women who may or may not know about their cross-dressing interests. DSM-5 removed the restriction to heterosexual men, opening the diagnosis to women or gay men who cross-dress out of erotic sexual interest. However, the diagnosis still requires that cross-dressing be a source of sexual arousal and that individuals experience significant distress or impairment because of their associated fantasies, urges, or behaviors. In any event, most gay men and lesbians do not cross-dress. However, there are social venues within gay and lesbian communities that allow for public cross-dressing. These may include either socially performative events (gay pride, Carnival, Halloween, or Mardi Gras parades) or cross-dressing as a form of paid entertainment (drag shows). Some, although not all, cross-dressers may identify as transgender individuals.

THEORIES

Scientific and social theorizing about gender diversity began in the middle of the 19th century. Although there are theories purporting to explain the origins of any gender identity, at present, it should be emphasized that the manner in which either transgender or cisgender identities develop is unknown.

History

19th and Early 20th Century Theories. The early history of theorizing about gender diversity was conflated with theories of what is today referred to as homosexuality. Many theorists routinely conflated homosexuality with transgender identities, relying upon binary beliefs that used conventional heterosexuality and cisgender identities as a frame of reference. For example, in the mid-19th century, German lawyer and activist Karl Ulrichs hypothesized that some men were born with a woman's spirit trapped in their bodies. He believed these men constituted a *third sex* and named them *urnings.* While historians of homosexuality unremarkably and routinely seem to regard Ulrichs' urnings as homosexual men, a female spirit in a male body closely resembles the narratives of 20th-century theories of transsexuality.

Until the early 21st century, with rare exceptions, transgender presentations were usually classified as "psychopathologic." One early exception was Carl Westphal's 1870 account of *Contrary Sexual Feeling,* the case of a young woman who cross-dressed as a boy and who was attracted to women with whom she formed relationships. Richard von Krafft-Ebing, on the other hand, reinforced the psychopathologic view in his landmark publication, *Psychopathia Sexualis,* an early psychiatric diagnostic manual. There he documented cases of individuals who desired to live as members of the other sex and those who had been born to one sex and were living as members of the other. In a chapter entitled "General Pathology," he presented what today would be thought of as transgender representations as cases of what he called *Metamorphosis Sexualis Paranoia.* In these examples, there was a significant conflation between homosexuality and transgender presentations.

German psychiatrist Magnus Hirschfeld is credited with being the first to distinguish the desires of homosexuality (to have partners of the same sex) from what he called *transvestism* (to live as the other sex). However, those distinctions were not broadly accepted until decades later. Hirschfeld also assisted surgeons in Europe who were already experimenting with GCS in the 1920s.

Mid-20th Century. The possibility of medical and surgical gender transition seized the popular imagination when American George Jorgensen went to Denmark as a birth-assigned male and returned to the United States in 1952 as transgender woman, Christine Jorgensen. In the wake of much public and professional controversy about Jorgensen's transition, Christian Hamburger and the other Danish physician who performed her surgery published a report of their treatment of her "transvestism" in the *Journal of the American Medical Association.* The publicity surrounding Jorgensen's transition eventually led to greater popular, medical, and psychiatric awareness of a then little-known concept that would eventually come to be known as *gender identity,* as well as recognition of an increasing number of people wishing to "crossover." For those who eventually would come to identify as transsexual and later transgender, increased public discussions of gender reassignment and gender identity would provide a way to put a name to their feelings and desires. As a result, an experience and presentation of gender once considered exceedingly rare gradually became more commonplace.

However, at the time of Jorgensen's 1950s transformation and for several decades thereafter, many professionals were critical of offering hormonal and surgical gender-affirming procedures. Many physicians and psychiatrists criticized using surgery and hormones to irreversibly—and in their view incorrectly—treat people suffering from what they perceived to be either a severe neurotic or psychotic, delusional condition in need of psychotherapy and "reality testing." It was in this cultural context that the first two editions of the DSM published in 1952 and 1968, respectively, did not include any gender diagnoses. The diagnosis of trans-sexualism (sic) first appeared in 1975 in the ICD-9. In 1980, a newly revamped DSM-III offered new diagnoses based on a growing body of research on child and adult gender identity. Table 21.4–1 outlines the placement of gender diagnoses in the ICD and DSM manuals.

The decisions to include *transsexualism* in the ICD and DSM were based on a growing body of research and clinical observations, most notably the work of John Money, Harry Benjamin, Robert Stoller, and Richard Green.

John Money, a psychologist and sexologist, first began publishing theories regarding gender identity development in the 1950s. Based on studies of children born with DSDs, he theorized that one's sense of being male or female—what eventually came to be known as one's gender identity—was acquired and that acquisition was primarily determined by external environmental factors. Citing cases of gender assignment in intersex children born with ambiguous genitalia, Money believed that parental attitudes and rearing practices had a strong effect on whether a child would accept the gender category to which it had been surgically assigned. Although Money believed that biology also played a role, he believed that the role of the psychosocial environment was critical and that any biologic predisposition could be overridden if parents unambiguously reinforced the gender assignment. Based on his observations, he suggested that a person's gender identity was fixed by 3 years of age and considered efforts to change a person's gender identity difficult if not impossible in anyone older. Pessimism about changing an adult's gender identity increased interest in an alternative for those dysphoric in their assigned gender: gender reassignment.

In the mid-1960s, in the wake of Money's theoretical work and his clinical and research findings, Johns Hopkins opened the first university-affiliated, multidisciplinary gender clinic offering gender reassignment to those seeking treatment. More than 40 academic centers in the United States would later open gender clinics as well. More recently, observations of gender identity outcomes in individuals with particular DSD syndromes have suggested that the influence of biologic factors on gender identity is stronger than that suggested by Money.

Table 21.4–1.
Gender Diagnoses in the ICD and DSM

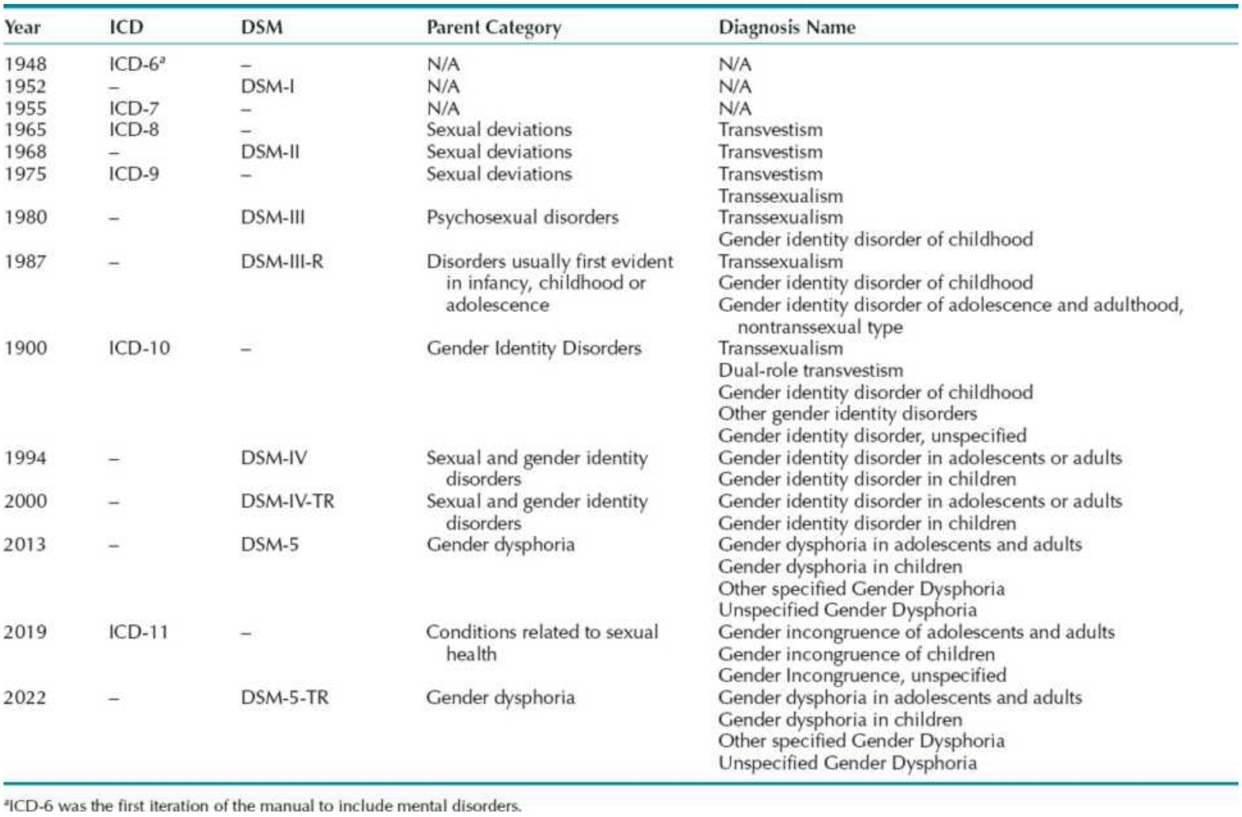

Year	ICD	DSM	Parent Category	Diagnosis Name
1948	ICD-6[a]	–	N/A	N/A
1952	–	DSM-I	N/A	N/A
1955	ICD-7	–	N/A	N/A
1965	ICD-8	–	Sexual deviations	Transvestism
1968	–	DSM-II	Sexual deviations	Transvestism
1975	ICD-9	–	Sexual deviations	Transvestism Transsexualism
1980	–	DSM-III	Psychosexual disorders	Transsexualism Gender identity disorder of childhood
1987	–	DSM-III-R	Disorders usually first evident in infancy, childhood or adolescence	Transsexualism Gender identity disorder of childhood Gender identity disorder of adolescence and adulthood, nontranssexual type
1900	ICD-10	–	Gender Identity Disorders	Transsexualism Dual-role transvestism Gender identity disorder of childhood Other gender identity disorders Gender identity disorder, unspecified
1994	–	DSM-IV	Sexual and gender identity disorders	Gender identity disorder in adolescents or adults Gender identity disorder in children
2000	–	DSM-IV-TR	Sexual and gender identity disorders	Gender identity disorder in adolescents or adults Gender identity disorder in children
2013	–	DSM-5	Gender dysphoria	Gender dysphoria in adolescents and adults Gender dysphoria in children Other specified Gender Dysphoria Unspecified Gender Dysphoria
2019	ICD-11	–	Conditions related to sexual health	Gender incongruence of adolescents and adults Gender incongruence of children Gender Incongruence, unspecified
2022	–	DSM-5-TR	Gender dysphoria	Gender dysphoria in adolescents and adults Gender dysphoria in children Other specified Gender Dysphoria Unspecified Gender Dysphoria

[a] ICD-6 was the first iteration of the manual to include mental disorders.

Harry Benjamin, a physician, is credited with both popularizing the term *transsexual* and for raising awareness about trans individuals within the medical profession. Benjamin offered transgender individuals hormonal treatment at a time when mainstream psychiatry and medicine regarded gender-incongruent individuals as confused homosexuals, neurotics, transvestites, people with schizophrenia, or some combination thereof. Benjamin believed that transsexuality was a biologic disorder arising, for example, in birth-assigned males due to brain feminization in utero. Consistent with his essentialist view, he believed psychotherapeutic attempts to change a gender identity were futile. Among the first physicians to experiment with hormonal and surgical treatments for aging, he eventually pioneered the treatment of gender dysphoric individuals using sex hormones. Notably, he accomplished this in a private practice setting without either university or academic support. In acknowledgment of his early advocacy, in 1979, the newly formed Harry Benjamin International

Gender Dysphoria Association (HBIGDA) was named in his honor. HBIGDA, which later changed its name to World Professional Association for Transgender Health (WPATH), would go on to develop *Standards of Care (SOC)* for treating individuals with gender dysphoria. The WPATH SOC are periodically revised and, at the time of this writing, are presently in their eighth version (SOC v8).

Robert Stoller was an American psychiatrist and psychoanalyst whose theorizing, like Money's, grew out of clinical experience with both transsexual individuals and individuals with somatic intersex conditions. Stoller is credited with introducing the concept of *gender identity* into both the psychoanalytic and psychiatric literature. In contrast to Benjamin's essentialist views, Stoller believed that in some cases, childhood family dynamics were responsible for "causing" adult transsexualism. Stoller, influenced by the separation-individuation theories of Margaret Mahler, opined that GD in prepubescent boys was a developmental arrest due to an excessively close and gratifying mother–infant symbiosis that prevented the child from adequately separating himself from his mother's female body and feminine behavior.

Richard Green, who at different stages of his career trained with Money, Benjamin, and Stoller, most notably studied gender diversity in prepubescent children. His book, "*The Sissy Boy Syndrome" and the Development of Homosexuality,* was a prospective study that tracked into adulthood the development of 66 gender-atypical boys who stated a wish to be a girl. Seventy-five percent of the children Green studied grew up to be gay men. This finding was consistent with studies done over the years at specialized gender clinics where the majority of children, mostly assigned male at birth, grew up to be gay and not transgender. Green, who served on the DSM-III Subcommittee on Psychosexual Disorders, recommended including the diagnoses of *transsexualism* and *GID of childhood* in the 1980s DSM-III. The childhood diagnosis would subsequently make an appearance in 1990's ICD-10 as well.

Transgender Emergence. Following the publication of a 1979 John Hopkins study by Jon Meyer and Donna Reter citing "no objective improvement" in transsexual women who had received GAS, Hopkins closed its gender clinic. Importantly, the conclusion of "no objective improvement" was based on a comparison of an "adjustment score" before and after transition. This score was a composite of variables that seem of questionable relevance to the resolution of GD today, including legal and economic considerations as well as marriage or "gender-appropriate" cohabitation. The study did, however, conclude that GCS was subjectively satisfying to recipients. Nevertheless, following the publication of this study, all but 3 of the more than 40 university-affiliated U.S. gender programs closed. This dramatically changed the delivery of medical care for GD in the United States. Prior to the closing of these clinics, in the 1960s and 70s, gender reassignment performed at university-based gender clinics was usually conducted with the aim that the recipient blend into a male/female, heteronormative binary mainstream. In the absence of university protocols, those seeking treatment had what transgender activist and writer Dallas Denny calls an "open market" for the treatment of their GD, as those desiring surgical and hormonal transition services were no longer screened for the potential to blend into the heterosexual, cisgender mainstream. Further, transgender individuals began to seek each other out as they were no longer constrained by the privacy requirements of gender clinics that kept them from meeting each other. The mid-1980s saw the creation of support groups and conferences at which they would meet and exchange ideas and experiences. By the 1990s, transgender individuals began to produce scholarly literature of their own. Also, starting in the 1990s, many of the existing lesbian, gay, and bisexual (LGB) civil rights advocacy groups began a process of *transgender inclusion* in recognition of the similarities in types of prejudice and discrimination faced by the entire LGBT community. Today openly transgender individuals from all walks of life are increasingly in the public eye. In recent years, antidiscrimination protections and/or hate crime laws for gender identity and gender expression have been enacted at the statewide level in 22 U.S. states and Washington DC. In 2012, the American Psychiatric Association (APA) issued position statements in support of transgender access to care and opposing discrimination against transgender individuals. Over the past 20 years, outcome studies following surgical gender reassignment have focused on subjective variables such as relief of GD, subjective sense of well-being, and satisfaction with transition as opposed to regret. Such studies have found the process of gender reassignment to be effective in lessening suffering and increasing subjective satisfaction, with major regrets being rare.

The 2012 APA Task Force on GID reviewed the evidence on satisfaction versus regret following GAS. Taken together, the studies suggest that major regrets are rare. For example, a review of 79 follow-up studies conducted between 1961 and 1991, including more than 1,000 MTF and 400 FTM patients, found that major regrets such as reversion to the gender assigned at birth occurred in only 1% to 1.15% of patients. Identified correlates of regret included an inadequate diagnosis of major mental health comorbidity, misdiagnosis, absence of adequate preparation, and poor family support. In 2018, similar results were reported based on the Amsterdam Cohort of Gender Dysphoria Study (1972 to 2015). Specifically, the percentage of those who underwent gonadectomy within 5 years of starting gender-affirming hormones remained stable over time (74.7% for transgender women and 83.8% for transgender men), and only 0.6% of transgender women and 0.3% of transgender men who underwent gonadectomy experienced regret. More recently, there have been anecdotal reports of *detransitioning* by individuals who may have successfully transitioned earlier in life but who now, for a variety of reasons, have returned to living in their gender assigned at birth. A few recent qualitative studies of detransitioned individuals indicate many reasons for making this decision.

Recent decades have also seen growing public awareness and acceptance of gender-diverse minors, be they prepubescent children or adolescents. Where once the predominant psychiatric approach for prepubescent children was to dissuade them from their gender-atypical interests, alternative treatment approaches have evolved (discussed later in this chapter). There has also been an increase not only in families allowing their children to socially transition before puberty but in schools and school boards willing to accommodate those transitions.

Mental Disorder or Normal Gender Diversity? Following greater visibility and paralleling gay activist protests of the mid-20th century, an increasing number of transgender-identified individuals and their allies challenge the characterization of transgender and transsexual identities, as well as GD, as mental disorders. Just as the gay community once argued for removal from an earlier diagnostic nosology, the transgender community has adopted similar normalizing arguments to make a case for removal. These include adopting normalizing etiologic theories, such as the belief that one is born transgender; adopting a transhistorical approach that connects modern transgender identities to historical figures and cultures; using modern cross-cultural studies to show that antitransgender attitudes are culture bound; using statistics regarding prevalence to refute the notion that transgender expressions are rare; underscoring the difficult, if not impossible, task of changing a gender identity; and adopting normative language to replace medical terminology.

In response to these social changes, there was much public debate about the retention of gender diagnoses in the DSM-5, starting with the appointment of the DSM-5 Workgroup on Sexual and Gender Identity Disorders in 2008 up until the publication of DSM-5 in 2013. The decision to retain the diagnoses in DSM-5 was in part based upon the need to maintain access to care for individuals seeking medical and surgical treatment; however, the name was changed to GD, which is more descriptive and less stigmatizing. While the use of V-codes was suggested and considered, this solution was ultimately rejected as limiting access to care since third-party payers do not reimburse for V-codes.

Similar concerns about perpetuating the stigma of a mental disorder were raised as the World Health Organization (WHO) revised the ICD, with the ICD-11 approved by the WHO Assembly in 2019. The Working Group on the Classification of Sexual Disorders and Sexual Health (WGSDSH) was charged with reviewing and evaluating clinical and research data informing gender diagnoses. As

WHO is part of the United Nations, in addition to its medical mission, it also has a human rights mission. Consequently, many advocates, several countries, the Council of Europe Commissioner for Human Rights, and the European Parliament advocated strong positions that issues related to transgender identity should not be classified as mental disorders in the ICD-11.

While reducing stigmatization associated with mental disorders is important, the argument for removing a diagnostic category from the Mental Disorders section of the ICD simply because mental disorders are stigmatized was neither compelling nor persuasive. Psychiatric illnesses, like “social deviance,” often create conflicts between individuals and society. Consequently, both the psychiatrically ill and minority groups are subject to stigma. Stigmatization of individuals with psychiatric disorders is a social problem across cultures. WHO, health care policy makers, psychiatrists, and other mental health professionals are actively engaged in trying to reduce the stigma of mental disorders in order to increase access to care and to encourage people to avail themselves of mental health services. Further, mental disorders are health conditions which is why they are in the ICD. One unintended consequence of belaboring distinctions between medicine and psychiatry is the perpetuation of existing stigma and prejudices against the mentally ill.

Nevertheless, the combined stigmatization of being transgender and of having a mental disorder diagnosis created a doubly burdensome situation for this population that contributed adversely to their health status, as well as to their attainment of human rights. For example, transgender people are much more likely to be denied care in general medical or community-based settings, given the perception that they must be treated by psychiatric specialists, even for conditions that have nothing to do with being transgender. Further, there are unique circumstances in the case of this particular diagnosis that relate to the ability of a person to be viewed as competent to make certain legal decisions. Government agencies in many countries have demonstrated a lack of understanding of the need to extend recognition of a change in legal gender status on identity documents such as passports, birth certificates, and drivers’ licenses. Courts may demonstrate prejudice by ruling against transgender people in child custody decisions in cases where being diagnosed with a mental disorder is used by an ex-spouse to call their competence as parents into question.

It should be underscored that from a historical perspective, the classification of gender identity diagnoses as mental disorders was serendipitous. It appears to have been based more on prevailing social attitudes at the time than on available scientific evidence. For example, if the social narrative at that time had been that transsexuality was related to a “sex atypical prenatal hormone exposure of the brain” rather than being a “sexual deviation,” the result could very well have been that “transsexualism” would have been placed in the ICD-10 chapter on “endocrine, nutritional, and metabolic diseases.” In fact, the etiology of the condition was unknown when placement decisions were made in the past, and it remains unknown now. There are no scientifically based criteria to differentiate normal and pathologic gender identity, and the manner in which *any* gender identity develops remains unknown and a matter of theoretical speculation. The extant scientific database cannot empirically answer the question of whether this diagnosis is purely a “mental” phenomenon as opposed to a biologic one. A number of studies posit physiologic rather than “mental” causes of transgender identification. In addition, many of the treatments employed today (e.g., hormonal, surgical, voice therapy, depilation) could not be construed as conventional mental health treatments. That is, today’s approaches to GI involve a focus on the body and social role rather than changing the individual’s experienced gender. There is no reason to expect that this approach will change in the foreseeable future.

Historically, based on the conceptualization of GI as a mental disorder, as well as the need to distinguish it from other mental disorders, many countries required psychiatrists and other mental health professionals to act as *gatekeepers* to transition services for transgender people. In some countries, transgender people wishing to transition must prove to psychiatrists that they are the “right” kind of transgender and that they have been so for a sufficiently long period of time in order to qualify for services. This has created some controversy, as many individuals seeking transition do not meet criteria for a psychiatric diagnosis nor do they desire mental health treatment, and the gatekeeping function may be seen as an unnecessarily burdensome requirement. How the gatekeeping role is performed has contributed to considerable tension between mental health professionals and transgender advocacy groups. In particular, mental health professionals may be viewed as gatekeepers rather than professionals who have been placed in a gatekeeping role. Being viewed as gatekeepers by those seeking services can interfere with the positive contributions that psychiatrists and other mental health professionals can make to the transition process. It can also be disheartening to those mental health professionals who see themselves as advocates for the transgender community to find themselves labeled as gatekeepers. In addition to ensuring that patients understand the potential benefits, risks, and limitations of their desired gender-affirming treatments, counseling and therapeutic approaches can be very important in helping an individual prepare for and cope with the personal and social effects of transition. In this regard, the role of mental health professionals is not fundamentally different from their participation as a part of assessment and treatment for individuals undergoing other complex, life-changing medical treatments, such as organ transplantation.

Consequently, in addition to recommending a name change to GI, WHO’s Working Group on Sexual Disorders and Sexual Health recommended that the diagnoses be removed from the ICD-11’s section on mental and behavioral disorders. This recommendation was accepted, and the diagnosis was moved out of the Mental and Behavioral Disorders section into a new chapter called *Conditions Related to Sexual Health.* A recommendation from advocacy groups that the diagnoses be listed as Z-codes (similar to DSM’s V-codes) was rejected for similar reasons: third-party payers, including national health services, do not reimburse for the treatment of Z-codes.

The retention of an ICD diagnosis for adolescents and adults has not been as controversial an issue as the retention of a diagnosis for prepubescent children. Even removing the diagnosis from ICD’s

Mental and Behavioral Disorders section has not satisfied some advocacy groups who argue that gender-diverse children are being stigmatized with a *medical diagnosis* where previously they complained of the stigma of a *psychiatric diagnosis.* These arguments have not been persuasive. Further, children with GI require access to care that is often complex and involves treatment of the family and social environment, as well as the child. Consequently, the ICD-11 diagnosis of GI in childhood narrowly defines the category so that gender-atypical presentations, in and of themselves, are not considered diagnosable conditions.

FREQUENCY

The prevalence of GD specifically, or transgender and gender-diverse identities in general, is hard to determine exactly for a variety of reasons, including concealment of these identities due to stigma as well as methodologic problems in defining the populations of interest and designing adequate sampling strategies. Researchers who study incidence and prevalence have tended to focus on individuals who presented to specialized gender clinics and received a diagnosis of GD or an earlier DSM or ICD equivalent. Most such prevalence studies have been conducted in European countries, and the prevalence for any GD diagnosis across populations has been assessed to be 0.0007% to 0.04% for those assigned male at birth and 0.007% to 0.02% for those assigned female at birth. Since not all who seek hormonal and surgical reassignment attend specialty clinics, the above rates are likely underestimates of transsexual individuals and clearly underestimates of transgender individuals who either do not desire or do not seek transition services. Studies that employ more inclusive sampling methods not restricted to specialty clinics tend to yield higher prevalence figures. Recent prevalence estimates based on surveys of self-reporting general population samples in the United States and Europe reveal much higher numbers. Responses vary, depending on how the question is asked. For example, self-identification as transgender ranges from 0.1% to 0.6 %, experiencing oneself as having an incongruent gender identity range from 0.6% to 1.1 %; feeling one is a person of different sex ranges from 0.4% to 0.6%; and the desire to undergo medical treatment ranges from 0.2% to 0.6 %.

In nonreferred children and adolescents, prevalence is believed to be less than 1%. In children, sex ratios of individuals assigned male at birth to individuals assigned female at birth range from 1.25:1 to 4.3:1. Recent studies show increasing numbers of children and adolescents presenting to specialty clinics, presentation at younger ages, more frequent early social transition, and a shift in sex ratio with a greater number of individuals assigned female at birth in adolescents and young adults.

DIAGNOSIS

DSM-5 Text Revision (DSM-5-TR)

The DSM-5 revision involved a diagnostic name change from GID and the creation of a new overarching parent category known as GD. One rationale for this change was to remove the stigmatizing term “disorder” from the diagnostic label. The term “dysphoria” also captures the aversive emotional component of the clinical presentation, as well as shifting clinical focus from “identity” per se to patient distress.

In addition, GD was moved out of an earlier parent category of sexual and gender identity disorders (where it was clustered with paraphilias and sexual dysfunctions) and into a separate category of its own. Other DSM-5 revisions included a new specifier, “with a disorder of sex development (DSD),” for individuals who have intersex conditions and develop symptoms of GD. This is in contrast to previous iterations of the DSM, where the presence of a somatic intersex condition was considered an exclusionary criterion for making a GID diagnosis. Any issues regarding the particular DSD condition and/or need for engagement in ongoing medical care should be clarified and appropriately addressed. The DSM-5 also included a new “post-transition” specifier for older adolescents and adults who have transitioned to full-time living in the desired gender.

Unlike the DSM-III and IV, DSM-5 removed the sexual orientation specifier as it no longer had any clinical relevance in the treatment of GD. The diagnostic criteria are unchanged in the DSM-5-TR.

The DSM-5-TR includes four GD diagnoses: GD in adolescents and adults, GD in childhood, other specified GD, and unspecified GD. The first two are described in greater detail later in this chapter.

Gender Dysphoria in Children. The diagnosis can only be made in children who meet two criteria, A and B, in the DSM-5-TR. Eight potential manifestations of criterion A are listed, of which at least six must be met. One of these six criteria must be “a strong desire to be of the other gender or an insistence that he or she is the other gender (or some alternative gender [that is neither conventionally male nor female] different from one’s assigned gender).” The other A-symptom criteria include preferring cross-dressing and rejecting conventional attire of one’s assigned gender; preferring gender roles of the other gender in make-believe play; preferring toys, games, or activities stereotypically associated with the other gender; and preferring play with children of the other gender. As children, those assigned male at birth typically reject masculine toys, games, and activities and avoid rough-and-tumble play; those assigned female typically reject feminine toys, games, and activities. These children frequently express a dislike of their sexual anatomy and may express the desire for the sex characteristics, primary and/or secondary, of the experienced gender. In addition to meeting six A criteria, Criterion B must be met, which requires the presence of clinically significant distress or impairment in social, school, or other important areas of functioning. While the presence of a DSD was an exclusionary criterion for diagnosing GID in previous editions of DSM, the DSM-5 added a new specifier: “with a DSD.” When this specifier is used, both the specific DSD, as well as the GD diagnosis should be coded. As some DSDs require ongoing medical care (e.g., salt-wasting congenital adrenal hyperplasia and individuals rendered agonadal by treatment), appropriate medical referrals should be offered, if there are any questions about the particular DSD diagnosis, as well as engagement in necessary medical care.

The differential diagnosis of GD in childhood includes simple atypical gender behavior in the absence of clinically significant distress or impairment. The diagnosis is not meant to merely describe nonconformity to stereotypical gender role and behavior such as “tomboyism” in girls or “girly” or “sissy” behavior in boys. The majority of such children are comfortable with their bodies and assigned genders and remain so in adolescence or adulthood. GD must also be distinguished from gender concerns that may arise as epiphenomena of coexisting disorders, including autism spectrum disorders (ASDs), personality disorders, or psychotic disorders, as well as a stated desire to be the other gender due to some perceived social advantage of being that gender or to escape a perceived disadvantage of the assigned gender (e.g., following sexual trauma). In cases where a child experiences GD during periods when psychotic symptoms are in remission, a psychotic disorder and GD may both be diagnosed. GD in childhood has different criteria sets than GD in adolescents and adults, and most children with GD in childhood do not subsequently meet diagnostic criteria for GD in adolescence and adulthood.

Gender Dysphoria in Adolescents and Adults. The DSM-5-TR diagnosis of GD in adolescents and adults is unchanged from the DSM-5. To establish the diagnosis, at least two of any of the six are required to be present for at least 6 months. These include a pronounced *incongruence* between one’s experienced or expressed gender and the primary/secondary sex characteristics of the *assigned* or *natal gender.* For prepubescent adolescents or those in puberty’s early stages, this symptom is expressed as aversive anxiety to *anticipated* secondary sex characteristics. Another criterion is a strong desire to be rid of the birth-assigned gender’s primary and/or secondary sex characteristics due to incongruence with one’s experienced/expressed gender as opposed to other motives. For prepubescent adolescents or those in puberty’s early stages, there is a strong desire to prevent further development of anticipated secondary sex characteristics. A third criterion is a strong desire for primary and/or secondary sex characteristics of the other gender. A fourth is a strong desire to be either of the other gender or some alternative gender, neither conventionally male nor female, rather than one's assigned gender. A fifth is a strong desire to be treated by others as a member of the other gender or some alternative gender. The sixth and final criterion is a strong conviction that one's feelings and reactions are more typical of the other gender or some alternative gender to the assigned gender.

There is also a B criterion of clinically significant distress or impairment in function in social, occupational, or other areas. As in the child diagnosis, there is a specifier, "with a DSD," that requires coding a specific DSD condition, if present, in addition to the GD diagnosis. As with the child diagnosis, any issues regarding the particular DSD condition and/or need for engagement in ongoing medical care should be clarified and appropriately addressed. There is also a "post-transition" specifier for individuals who have transitioned to living full time in the desired gender, regardless of a legal name change, who have undergone or plan to have at least one cross-sex medical or surgical procedure. This specifier allows individuals to maintain the diagnosis needed to access ongoing gender-affirming care (e.g., gender-affirming hormones), even if they no longer meet either the A or B criteria for the diagnosis after the transition.

The differential diagnosis of GD in adolescents and adults includes nondiagnosable *atypical gender behavior; transvestic disorder* (a paraphilia in which individuals are sexually aroused by cross-dressing—some individuals may meet criteria for both transvestic disorder and GD); *body dysmorphic disorder* (involving a wish to alter or remove a specific body part perceived to be unattractive or abnormally formed); and *schizophrenia* and *other psychotic disorders* (in which delusions of being another gender may present). Importantly, if the full criteria for GD are met in an individual with a psychotic disorder when the psychosis is in remission, the diagnosis of GD can still be made, as is the case if the GD preceded the onset of the psychotic illness. Some individuals assigned male at birth may seek castration to ameliorate psychological or physical effects of androgens yet still may identify as male or employ a different gender identity label such as male *eunuch*, *agender,* or *nonbinary*.

The DSM-5-TR contains two additional gender diagnoses. The first is *other specified GD*. This diagnosis is given when the patient has manifestations characteristic of GD, and the clinician wishes to specify why the full diagnostic criteria are not met. This typically occurs when a sufficient number of the A criteria are met, but the duration criterion is not met at the time the diagnosis is given. The second is *unspecified GD*. This diagnosis may be used when the clinician chooses not to specify why the full criteria are not met and may be appropriate, for example, when the clinician has spent insufficient time with the patient to clarify whether sufficient symptoms are present to meet diagnostic criteria.

ICD-11

The WHO's ICD gender diagnoses differ from those of DSM-5-TR in several ways. First, GID of childhood was renamed *GI of children,* while transsexualism was renamed *GI of adolescents and adults.* Second, both diagnoses were removed from the Mental and Behavioral Disorders section of ICD and are part of a new section entitled Conditions Related to Sexual Health.

Gender Incongruence of Childhood. ICD-11's GI of Childhood is characterized by a marked incongruence between an individual's experienced/expressed gender and the assigned sex in prepubertal children. A marked incongruence between the child's experienced/expressed gender and the child's assigned sex must be manifested by the following three indicators: (1) a strong desire on the child's part to be a different gender than the assigned sex or insistence that they are a gender different from one's assigned gender; (2) a strong dislike on the child's part of their sexual anatomy or anticipated secondary sex characteristics and/or a strong desire for the primary and/or anticipated secondary sex characteristics that match the experienced gender; and (3) make-believe or fantasy play, toys, games, or activities and playmates that are typical of their experience rather than their assigned sex. In addition, the incongruence must have persisted for about 2 years, so the diagnosis cannot be made before approximately age 5. The diagnosis can only be assigned to children before puberty.

Gender Incongruence of Adolescence and Adulthood. GI of adolescence and adulthood (GIAA) is characterized by a marked and persistent incongruence between an individual's experienced gender and the assigned sex. The diagnosis cannot be assigned prior to the onset of puberty. GIAA often leads to a desire to transition in order to live and be accepted as a person of the experienced gender. Establishing congruence may include hormonal treatment, surgery, or other health care services to make the individual's body align, as much as desired and to the extent possible, with the experienced gender.

Diagnostic guidelines include a marked incongruence between the individual's experienced gender and the assigned sex, as manifested by at least two of the following four criteria: (1) a strong dislike or discomfort with one's primary and/or secondary sex characteristics (in adolescents, anticipated secondary sex characteristics) due to their incongruity with the experienced gender; (2) a strong desire to be rid of some of all of one's primary and/or secondary sex characteristics (in adolescents, anticipated secondary sex characteristics) due to their incongruity with the experienced gender; (3) a strong desire to have the primary or secondary sex characteristics of the experienced gender; and (4) a desire to be treated (to live and be accepted) as a person of the experienced gender. The diagnosis cannot be assigned prior to the onset of puberty, and the GI must have been continuously present for at least several months.

Reconciling DSM-5-TR and ICD-11 Diagnostic Codes

While the DSM has its own coding system, which at one time was accepted by most insurers, in 2015, the United States began requiring the use of ICD-10 codes for insurance reimbursement. Consequently, DSM-5-TR also includes equivalent ICD-10 codes for all its diagnostic categories. The DSM-5-TR diagnosis of GD in adolescents and adults (302.85) also lists the ICD-10 diagnosis code for transsexualism (F64.0) as an equivalent code.

However, the diagnostic codes of ICD-11 are not equivalent to those of ICD-10. The code for ICD-11's GIAA is HA60. This means that although GIAA is no longer categorized as a mental disorder in ICD-11, the United States is still using an ICD-10 code that continues to code the condition as a mental disorder (F codes). As it took 25 years for the United States to adopt ICD-10, which was first approved by WHO in 1990, it is not clear when ICD-11 codes will be adopted in the United States. At the time of this writing, the APA is exploring ways to integrate the ICD-11 changes into the U.S. coding system (ICD-10 Clinical Modification or ICD-10-CM). However, given the bureaucratic obstacles involved in making such changes, the adoption of ICD-11's nonpsychopathologizing code is not imminent.

ORIGINS OF GENDER DYSPHORIA

Because GI does not necessarily lead to clinically significant distress, understanding its etiology requires understanding both the acquisition of gender identity and how a mismatch between assigned and experienced gender lead to clinically significant distress in some individuals. This section examines theories about the acquisition of gender identity, both congruent and incongruent with the sex assigned at birth, and then about the origins of distress resulting from GI in different age groups.

Acquisition of Gender Identity

Studies focusing separately on biologic and psychosocial factors suggest that both contribute to gender identity, but the precise role of each, their interactions, and the contribution of the individual to the process remain incompletely understood. Early interest in gender identity focused on the infant/toddler's cognitive abilities during the period of gender identity formation and elucidating the limits of the developmental window of this period. The studies of John Money and his collaborators briefly discussed previously in this chapter suggested a period of maximal sensitivity for the acquisition of gender identity that ends at approximately age 2. They found that among children who experienced gender reassignment (usually in the context of a DSD or penile injury) after the neonatal period, 11 of 14 adjusted to the reassignment without complications if the reassignment occurred prior to 27 months of age compared to only 1 of 4 who adjusted to the reassignment if it occurred after 27 months. Subsequent work of Money and others suggested that gender identity may be influenced beyond this period but that after some imprecisely defined time, it becomes more or less impervious to change. In other words, they concluded that gender identity is determined primarily by rearing practices during the first 2 years or so after birth.

Subsequent research by others concluded that infants begin to categorize gender-relevant attributes prior to the development of expressive language. For example, at 9 months, infants can associate male and female faces with male and female voices, respectively, looking longer at a female than a male face when a female voice was presented and vice versa. By 18 months, infants associate gender labels such as “lady” or “man” with female and male faces. As language develops, by 15 to 24 months, toddlers can label adults as male or female and, by 30 months, can label peers and self appropriately by gender. As gender labels emerge, in the majority of individuals, a preference for play with peers of the same gender emerges, and boys and girls tend to ascribe more positive attributes to their own gender and more negative attributes to the other. Some toddlers, however, show a preference for peers of the other gender and may even label themselves as the other gender as early as between the ages of 2 and 3.

By age 2, children can recognize their own genitals when shown anatomically correct dolls. Although by age 3 they may be aware that boys have penises and girls do not, less than half were found to understand this anatomic difference as the basis for gender categorization even by age 5. Between the age of 4 and 6, many children have yet to master the concept of *gender constancy,* an understanding that external changes in appearance or activity do not change one’s sex/gender (in the context of these studies, sex and gender are used interchangeably). Thus, they show confusion when asked whether a change in behavior or clothing (such as a boy playing with a doll and wearing a dress) would change their sex/gender. Similarly, they may have yet to master the concept of *gender stability,* the understanding that for most individuals, sex/gender remains constant throughout life. It is not until age 6 or 7 that nearly all children understand that genitals are the primary basis for sex/gender categorization and that sex/gender for most people is constant and stable. These cognitive limitations leave open the possibility of a world for young children in which the feeling that one is a sex/gender other than that assigned at birth would not be immediately countered by the awareness that, for most people, one's sex/gender assigned at birth is retained for life. For example, until the age of 6 or 7, a child assigned female at birth who does not understand the association between external genitalia and sex/gender has not mastered the concept of gender constancy and may believe that she could grow up to be a man merely by dressing and acting like a man.

Role of Psychosocial and Biologic Factors. The permissive, direct, and indirect/interactional models for conceptualizing the potential roles of biologic and psychosocial factors outlined for the development of sexual orientation may apply to gender identity as well.

MODEL OF PERMISSIVE BIOLOGIC EFFECTS. In this model, biology plays a permissive role by providing the neural machinery through which gender identity is inscribed by formative experience. A permissive role could also include delimiting the developmental stage during which the relevant formative experiences must occur. For example, some songbirds must learn their species' song by hearing it during a restricted period of early development. While the song is clearly acquired through experience, biology determines the sensitive period during which that experience must occur. Once a particular song has been acquired, that will be the bird's song for life; it will not be able to unlearn that song or to acquire another. This is not meant to imply that gender identity might be acquired by simple mimicry. Instead, it shows that phenomena acquired through experience may, nonetheless, be immutable.

MODEL OF DIRECT BIOLOGIC EFFECTS. In this model, the actions of biologic factors such as genes or hormones would directly influence the organization or activity of brain circuits that mediate gender identity. The brain is viewed as possessing an intrinsic (i.e., constitutional) gender identity or predisposition, the expression of which could nonetheless be subsequently modified by experience.

MODEL OF INDIRECT BIOLOGIC EFFECTS. This model addresses the contribution of individual differences, some of which may be constitutionally influenced. In this model, biologic factors influence gender identity only indirectly as the more direct or immediate action of biologic factors is upon temperament or other personality traits. From birth, these personality or temperamental traits influence how an individual experiences, interacts with, and modifies the environment. This would include how an individual participates in shaping the relationships and experiences thought by some to influence the development of gender identity. Although similar to the model of permissive biologic effects, the model of indirect biologic effects goes further and suggests that the relevant formative experiences may, themselves, be strongly affected by biologically influenced personality variables.

An impact of psychosocial factors in the acquisition of gender identity has been documented in a large body of research in developmental psychology on normative development. These factors include verbal labeling (e.g., as a boy or girl) and nonverbal gender cueing (e.g., mannerisms, clothing, hair length, and style) of children by parents and others in their social environment, the shaping of gendered behavior by positive and negative reinforcement and later by explicit statements of gender-role expectations. Problems in parent–child attachment interacting with temperamental dispositions in the child have been suggested to be causally implicated in GD and have been cited as a focus for psychodynamic therapy of the child. Child psychologist, Susan Coates, observed higher levels of psychopathology in clinical samples of primary caregivers and suggested that parental psychological abnormalities may contribute to GD. For example, a child assigned male at birth with a clinically depressed mother may identify with her in an attempt to elevate her mood. These observations, however, do not distinguish between cause and effect.

A review of the clinical data on individuals with somatic intersex conditions (aka DSDs) suggests that some individuals with functional androgen receptors and prenatal exposure to male-typical levels of androgens appear to have the capacity at birth to develop either a male or female gender identity depending on the gender of rearing. When assigned female, these individuals nevertheless have an increased likelihood of exhibiting masculinized play preferences and GD, with some ultimately rejecting the female identity and role. The proportion that ultimately rejects the female assignment is unknown because this can occur quite late in life, varies across DSD syndromes, and relatively few studies have followed and adequately assessed affected individuals in adulthood. The probability of switching to a male gender identity and role is perhaps highest in androgen-sensitive individuals whose testes are left in place until puberty. These data are open to a variety of interpretations, including the one that androgens act by modifying hypothetical brain circuits that mediate gender identity (direct model) or that androgens bias behavioral or temperamental propensities in a manner that facilitates the acquisition of a male identity (indirect model). In addition to acting prenatally, the elevated androgen secretion that occurs in the neonatal period and again at puberty might reinforce the effects of prenatal androgens. At puberty, the psychological impact of somatic virilization in response to elevated androgens must be considered in addition to the possibility of physiologic effects on the brain. Since the mid-1950s, the majority of XY individuals with somatic intersex conditions and markedly incomplete masculinization of the external genitalia have been assigned female. Thus, there are inadequate data upon which to base speculations about gender development in individuals with low prenatal androgen exposure who were assigned male.

A hypothesis that appears to be growing in popularity is that cross-gender identification results from a brain-limited form of intersexuality, that is, in transgender individuals, some brain regions or circuits that are involved in gender identity differ between cisgender men and women and are sexually discordant with the rest of the body in transgender individuals. This hypothesis is derivative of Ulrichs' 19th-century notion of a spirit of one sex trapped in the body of the other sex. Consistent with this hypothesis, based on their work with postmortem material, Dick Swaab and collaborators in the Netherlands have reported that a subcomponent of a brain structure called the bed nucleus of the terminalis varies in size as a function of gender identity rather than sex assigned at birth, a study that has yet to be replicated. Additional studies have described other brain differences, both anatomical and neurophysiologic, but the evidence at this point is insufficient for categorizing transgender people as having a form of CNS-limited intersex condition. Similarly, the literature on concordance for GD in monozygotic and dizygotic twins lends some support for a role of genetic factors, especially in individuals assigned male at birth. Due to small sample sizes and the absence of genotype studies, however, definitive conclusions cannot be drawn.

Controversial Issues

AUTOGYNEPHILIA IN BIRTH-ASSIGNED MALES. Psychologist and sex researcher Ray Blanchard has hypothesized two major developmental pathways leading to GD in transgender adults. In this theory, "homosexual transsexuals" are effeminate, gender-atypical birth-assigned boys who are attracted to partners of their own birth-assigned gender. That is to say, they are *androphilic* or attracted to men. The other group consists of what he terms "heterosexual transsexuals" who are *gynephilic* (attracted to women) and who are usually not overtly gender atypical as children and who often spend much of their lives in relationships with women. The latter group often has a history of cross-dressing in secret that first starts out as a means of sexual arousal and may go on to develop *autogynephilia,* in which a birth-assigned male finds the fantasy of being a woman to be sexually arousing. Although also attracted to women, these individuals are often aroused at the fantasy of having sex or actually engaging in sex with men who treat them as women. Cross-dressing, arousal to cross-dressing, or fantasies of being female are not classified as psychiatric disorders unless they cause clinically significant distress or impairment in important areas of functioning. In some cases, individuals with autogynephilia later develop GD and seek gender reassignment. After transition, they may prefer to engage in relationships with cisgender or transgender women.

Although the theory of an autogynephilic pathway to transsexualism has been in the research literature for decades, in recent years, it evoked controversy. While some members of the transgender community do identify as autogynephilic, others find offensive the linkage of transgender identity with paraphilia as well as the linkage of the desire to transition gender to sexual arousal. Some see it as reductionistic to hypothesize only two pathways to an identity that, in clinical practice, appears to be quite heterogeneous in its development. One scholarly critic of the concept is Charles Moser, a sexual medicine expert, who surveyed 29 birth-assigned women and found 93% of them met Blanchard’s criteria for autogynephilia. This finding is at odds with Blanchard’s theory that the condition does not occur in birth-assigned women and is a type of “male sexuality.” Moser argues that the meaning of a sexual interest in one’s own body is not well understood for either cisgender or transgender individuals of either sex assigned at birth. Transgender activist and biologist Julia Serano further notes that a correlation between autogynephilia and GD does not imply causation. She notes (1) that there are exceptions to the two subtype taxonomy (e.g., large numbers of androphilic MTF individuals who experience arousal by presenting female); (2) that some transgender people experience such arousal as a passing phase rather than a pressing paraphilic sexual drive leading to GD; and (3) that many nonandrophilic MTF transsexual individuals experienced the desire to be the other gender long before they experienced arousal to presenting as women or fantasizing of doing so.

RAPID-ONSET GENDER DYSPHORIA. In 2018, Lisa Littman published a study of 256 parental reports that their adolescent children who were assigned female at birth reported an onset of GD either during puberty or after its completion. The GD was reported to occur in the context of belonging to a peer group where one, multiple, or even all of the child's friends reported being gender dysphoric or transgender identified at the same time. Parents also reported their children demonstrated increased immersion in social media use prior to disclosing a transgender identity. The study defined "rapid-onset gender dysphoria" (ROGD) as GD beginning in adolescence or young adulthood in individuals who would not have met the criteria for GD in childhood. The study claimed these findings were inconsistent with existing research literature about the onset of GD/GI. It also raised questions of whether GD of this type, as with eating disorders in adolescents, could be a product of social contagion.

The study's publication provoked immediate criticism and controversy, leading the journal to request a postpublication review. Activists saw the study's publication as an effort to discredit the authenticity of transgender identities. A methodologic criticism was the study's selection bias in interviewing parents rather than the children or the clinicians evaluating the children. Further, the onset of GD without any features of GD before childhood is not a novel finding since prepubescent children with GD have long been viewed as a separate patient group from those who have GD onset in adolescence or adulthood. The study was republished in 2019 with some minor changes and caveats, emphasizing that ROGD was not a formal diagnosis and addressing the study's limitations as well as its potential utility in hypothesis building.

Origins of Distress Associated with Gender Incongruence

An objection to retaining GD as a psychiatric diagnosis in DSM is that it identifies gender-diverse identities and expression as pathologic when the distress required for the diagnosis may be related to the reaction of the social environment to the child's gender diversity. Since not all who experience a mismatch between their assigned and experienced gender experience clinically significant distress, the origins of distress should be considered separately from the origins of the cross-gender identification. Psychologist, Diane Ehrensaft, suggests that in most children with GD, the distress arises from stigma and prohibitions within the child’s social environment. She bases this suggestion on her observation that distress often resolves or diminishes when gender dysphoric children are allowed to freely express their experienced gender. Another hypothesis, for which there is very limited support, is that for some individuals, the transgender identification might, itself, be an expression of distress due to inadequate parenting or loss of a parent. For example, a boy who loses his mother through death or emotional withdrawal may suddenly show a preference for stereotypically feminine attire and state that he is or wants to be a girl. Some have interpreted such cases as the boy replacing the lost mother through identification and imitation. The ICD diagnosis of GI bypasses this controversy altogether as it requires no distress criterion to make a diagnosis.

Parents may act to limit social stigma and prohibitions in preschool-aged children in the home; however, unless preventive measures are taken, starting school and increasing interactions with the outside world may prove traumatic and exacerbate dysphoria. For some, the onset of puberty has been associated with the resolution of GD, with some reporting that pubertal changes forced them to accept that they were not the gender they had experienced themselves to be. For others, the onset of puberty and development of the secondary sex characteristics of the gender assigned at birth exacerbate dysphoria and may precipitate an emotional crisis. The foregoing pertains to those in whom GD began in childhood and persisted into adolescence. The origins of GD with onset in later adolescence or adulthood are more poorly understood, especially when there was no history of gender nonconformity observed in childhood. In some individuals, as mentioned previously in this chapter, late-onset GD has been hypothesized to be erotically driven (*autogynephilia*) and may entail a history of arousal to cross-dressing prior to the onset of anatomical dysphoria. Alternatively, the authors of a study of military veterans with late-onset GD suggested that the apparent late onset was due to other psychological factors. Specifically, many transgender veterans who were assigned male at birth reported that they joined the military in an attempt to purge their transgender feelings (*suppression*), believing that the military environment would make men of them, while others arrived at this interpretation after the fact (*repression*).

Directions for Future Research

Extensive research is needed on GD, especially with respect to identifying, understanding, and addressing the needs of individuals with GD in a manner that avoids stigma. This is particularly important in light of the considerable body of evidence linking stigma and *gender minority stress* to pronounced mental and physical health disparities seen in the transgender population. *Gender minority stress* refers to the high levels of chronic stress experienced by transgender people. Stressors include ridicule and bullying on the basis of gender nonconformity, intentional misgendering, exclusion from gendered spaces consistent with one's gender identity or gender presentation, as well as discrimination in multiple arenas, including employment, housing, and even health care. Much of the research linking minority stress to health disparities has been conducted in the arena of public health. Mental health disparities linked to minority stress in transgender populations include elevated rates of depression, suicidality, post-traumatic stress disorder, and substance abuse. In addition, many transgender individuals report avoiding health care due to past experiences of stigma and/or anticipation of stigma, and this avoidance exacerbates health disparities. Thus, transgender populations also exhibit marked general health disparities.

Considerable literature on GD addresses the identification of predictors of desistence versus persistence of GD in youth but, to date, no such reliable predictive system has been found. Recommendations for the minimization of stigma and its adverse impacts on self-esteem should be prioritized. This will entail considering the home and social environments in developing interventions aimed at minimizing stigma and optimizing long-term outcomes. Research is also needed on the selection of adolescents for pubertal suppression as well as the short- and long-term benefits and risk profiles of the treatment with respect to social, health, and psychological endpoints. The same holds true regarding greater individualization for the timing of the initiation of gender-affirming hormones. The range of needs of individuals with both GD and a DSD also requires greater attention.

There is also considerable interest in identifying genetic, endocrine, and brain correlates of GD. Studies of the development of gender identity in individuals with some DSDs have yielded some insights into hormonal pathways that may affect gender identity in individuals with DSDs; however, their relevance to GD in the absence of a DSD precludes broad generalizations.

There has been considerable speculation that the hormonal exposure of the developing brain, whether due to gene-controlled mechanisms (e.g., hormone production, hormone receptor mechanisms) within the fetus or exogenous factors (e.g., hormonal exposure or stress to the mother during pregnancy), may strongly influence or determine gender identity. In other words, GD could represent a brain-limited intersex condition. While there is some support for such hypotheses, definitive data are currently lacking. For example, the structure of some brain regions has been reported to vary between cisgender men and cisgender women and to be sex reversed in transgender individuals; however, such studies lack consistent replication. In addition, other sexually dimorphic attributes of the brain have been reported to shift on receipt of gender-affirming hormones. Again, however, replication is insufficient to date to allow strong conclusions.

Clinical research is much needed to better understand the substrates of the distress and impairment that are required for the diagnosis of GD. In particular, is distress intrinsic to one’s gender identification and expression, per se, in response to a perceived discrepancy between one’s sense of self and their body or sex assigned at birth, or in response to the disapproving reactions of others to the perception of one’s gender atypicality?

Research is also needed to better inform clinically appropriate interventions and their optimal timing and contexts, for example, addressing GD in the context of entering school, subsequent decisions about pubertal suppression and starting gender-affirming hormones, as well as decisions about whether to transition gender at school and, if so, exactly how and when.

The origins of gender identity are of broad general and theoretical interest, while the causes of the distress and impairment in GD demand research because of their implications for clinical interventions. Due to the low prevalence of GD and the heterogeneity of presentations, the research could be facilitated best if clinics pool their population samples and employ standardized assessments and interventions. Longitudinal studies from childhood through adulthood of such pooled samples have the potential to not only inform best clinical practices but to identify the range of factors that influence gender identity and its stability or change across development.

SOCIAL AND CULTURAL ISSUES

Delineating the full extent of social and cultural issues affecting transgender individuals is beyond the scope of this chapter. Among the *stigmatizing* challenges some transgender individuals face are rejection by family (leading to homelessness for some adolescents), estrangement from children produced in pretransition heterosexual relationships, rejection by peers, rejection by religious authorities, lack of acceptance from school authorities, bullying, harassment, job discrimination, lack of adequate and affirming medical and mental health care, and being subjected to violence. Members of the transgender community have produced or contributed to a sizable literature detailing personal stories of coming out as transgender that continues to grow. These include Jennifer Finney Boylan's *She's Not There: A Life in Two Genders,* Jamison Green's *Becoming a Visible Man,* and the multiauthored volume, *Trans Bodies, Trans Selves,* edited by Laura Erickson-Schroth.

CLINICAL APPROACHES

The most widely recognized guidance for treating persons with GD comes from the World Professional Association for Transgender

Health Standards of Care (WPATH SOC), now in their eighth revision. The SOC v8, which uses the ICD-11 diagnosis of Gender Incongruence (GI) endorse psychological evaluation and a flexible staged transition in which fully reversible steps (e.g., presenting as one's experienced gender or social transition) usually precede partially reversible steps (hormonal treatments to suppress development of unwanted secondary sex characteristics or to bring about the secondary sex characteristics consistent with one's experienced gender), which precede irreversible surgical procedures (e.g., gonadectomy, vaginoplasty in individuals assigned male at births, hysterectomy, salpingo-oophorectomy and phalloplasty in individuals assigned female at birth). Mastectomy and male breast construction in individuals assigned female at birth can be initiated prior to gender-affirming hormone treatment. Hormone therapy can be initiated with a referral from a qualified mental health professional or another health professional appropriately trained in behavioral health and able to assess GI and the eligibility and readiness of the patient for such therapy. WPATH SOC v8 stresses the importance of multidisciplinary evaluations prior to somatic gender-affirming treatments in order to rule out any conditions that would call the diagnosis of GD into question or indicate risks associated with the desired gender-affirming treatments. If any other such condition is present, it should be adequately addressed so that the desired treatments may be safely initiated. Previous versions of the SOC specified minimum durations of living in the experienced gender role prior to the initiation of hormonal therapy, but beginning with version 8 makes no specification in this regard. The WPATH SOC v8 suggests 1 year of continuous gender-affirming hormone therapy (unless contraindicated in an individual) and two referral letters prior to genital surgery. The letters must document persistent GI and the capacity to make a fully informed decision and give consent for treatment. If any medical or mental health concerns are present that may compromise the safety or safe adherence with the desired treatments, they must be adequately assessed and treated.

The WPATH SOC provides flexible guidelines, not rigid rules. For example, individuals assigned female at birth may elect mastectomy and perhaps male breast construction prior to presenting socially as male. Further, not all clinicians follow the WPATH SOC. Consequently, adults who have the capacity to give informed consent may receive gender transition treatments because they satisfy the qualifying criteria of the treating clinicians. These criteria vary among clinicians and clinics. Reviews that have graded the quality of evidence relating particular components of the WPATH SOC to outcomes have concluded that psychotherapy prior to transition is associated with a good outcome. A majority of insurers and providers of gender-affirming treatments may state that they follow WPATH SOC guidelines; however, they may not follow them precisely, creating uncertainty about which gender-affirming treatments are actually covered. Mental health professionals should address this possibility with patients preparing for gender-affirming treatments and encourage them to review this matter with their medical and surgical providers and insurers.

Treatment of Prepubescent Children

Working with children with GD can be considered to follow one of three approaches, each of which will be discussed in turn. None of these approaches entail hormonal or surgical interventions. The first approach focuses on working with the child and caregivers to lessen GD and to decrease cross-gender behaviors and identification. This approach is based upon the assumption that the child's self-esteem can be best served by improved social integration, including positive relationships with same-sex peers, expected to come with decreased gender atypicality. Since GD that persists into adolescence is more likely to persist into adulthood and manifest as transsexuality, this approach has been pursued by some with the goal of inducing desistence prior to adolescence; however, no empirical evidence suggests that is a realistic goal. Persistence is considered to be an undesirable outcome by some but not all clinicians who work in this area. A variety of reasons are cited for considering persistence undesirable, including social stigma as well as the likelihood persistence will eventually entail hormonal and surgical procedures with their associated risks and costs. Critics of this approach have likened it to so-called *sexual orientation conversion efforts (SOCE)* or *reparative therapies* of homosexuality, labeling it as *gender identity conversion efforts*. However, unlike SOCE practices, this clinical approach to children with GD is accepting of adult homosexual outcomes. It has, nevertheless, been legislatively defined as conversion therapy and banned in many jurisdictions.

A second approach, often referred to as *the Dutch approach* because it was developed in the Netherlands, is to remain neutral with respect to gender identity outcome and to not target gender identity and gender-atypical behaviors for change. Also referred to as *watchful waiting,* the goal is to allow the developmental trajectory of gender identity to unfold naturally without pursuing or encouraging a specific outcome. This approach entails combined child, parent, and community-based interventions to support the child in navigating the potential social risks. This approach is premised on evidence that GD diagnosed in childhood usually does not persist into adolescence and beyond and on the lack of reliable markers to predict in whom it will or will not persist. This approach aims to avoid damaging the child's self-esteem by risking conveying to them that their likes and dislikes, behaviors, and mannerisms are somehow intrinsically wrong. This approach may also entail assisting the child in mastering the cognitive, emotional, and behavioral tools for coping with the social stigma of gender diversity.

A third model, referred to by its practitioners as the *gender affirmative approach,* entails affirmation of the child's expressed-gender identification by mental health professionals, family members, and others. Thus, the child may be supported in socially transitioning to a gender role that affirms their experienced gender, with the option of age-appropriate endocrinologic and surgical interventions if GD persists. The rationale for supporting transition before puberty is the belief that a transgender outcome is to be expected in some children and that these children can be identified so that primary caregivers and clinicians may opt to support early social transition. Since childhood transition is done solely at a social level without medical intervention until puberty, some assert that children who transition this way can revert to their originally assigned gender if necessary.

Reservations regarding this approach include that the peer-reviewed literature does not support the view that reliable clinical markers exist to distinguish children whose GD continues into adolescence and adulthood from those where it does not; as well as some studies showing that a social transition in childhood may increase the likelihood of GD's persistence and the need for hormonal and or surgical interventions. In addition, individual case reports suggest that after transitioning in childhood, reverting to the birth-assigned gender may entail stressful social complications.

An APA Task Force on Treatment of GID published in 2012 identified the major tasks for mental health professionals working with children referred for gender concerns. These include accurately evaluating the gender concerns that precipitated the referral; accurately diagnosing any gender identity–related concerns in the child according to the criteria of the most current DSM; accurately diagnosing any coexisting psychiatric conditions in the child, as well as problems in the parent–child relationship, and recommending appropriate treatment; providing psychoeducation and counseling to the caregivers about the range of treatment options and their implications; providing psychoeducation and counseling to the child appropriate to their level of cognitive development; when indicated, engaging in psychotherapy with the appropriate persons, such as the child and/or primary caregivers, or to make appropriate referrals for these services; educating family members and institutions (e.g., daycare and preschools, kindergartens, schools, churches) about gender diversity and GD; and assessing the safety of the family, school, and community environments in terms of bullying and stigmatization related to gender atypicality and to address suitable protective measures.

Kim and Jan are identical twins who were assigned male at birth. Their mother's pregnancy and delivery were unremarkable. Starting at age 2, both children began to exhibit strong atypical gender interests. They preferred playing with girls' dolls and rejected stereotypical boys' toys. As they got older, they liked to play dress-up with their mother's clothing, sometimes wrapping their heads with towels to simulate long hair. Both insisted that they were girls. They preferred to urinate sitting down. Although the parents initially thought their feminine interests would fade away with time, when the children were 5 years old, they sought professional consultation at a gender clinic in the major metropolitan center where they lived.

Both children were given a DSM-5-TR diagnosis of GD in children. The clinic discussed treatment options that ran the gamut from discouraging all cross-gender behavior at one extreme to the possibility of socially transitioning the children to live and be publicly identified as girls. The parents felt that inhibiting the atypical gender interests of their children was too harsh but were not at all comfortable with the idea of social transition, given that it was not clear at the time whether the GD of either child would desist or persist. They opted for what is known as the Dutch approach or watchful waiting. They would be identified publicly as boys with gender-atypical interests, and an effort was made not to interfere with the unfolding of either child's gender outcome. This entailed much psychological support for the parents, some members of their extended family, and occasional consultation with school officials.

At age 10, Kim told the parents that he no longer thought of himself as a girl. At age 14, he came out to his parents as gay. Jan, on the other hand, continued to feel like a girl. At age 12, as puberty began to approach, Jan became very panicky about anticipated bodily changes, and her parents approved of treating her with puberty-suppressing drugs. Her GD never remitted, and when she turned 16, Jan was treated with feminizing hormones.

At age 18, she began seeking consultation to explore surgical treatments.

Treatment of Adolescents

Unlike previous versions of the SOC, v8 includes a chapter specifically devoted to adolescents. In addition, it has removed minimum age requirements for gender-affirming medical care but strengthened the criteria for treatments with the aim of allowing each adolescent the opportunity to receive individualized gender-affirming treatment at the appropriate time. Management of GD that has persisted from childhood is more straightforward than management of GD with later onset, especially when late onset occurs in the context of recent psychological trauma or co-occurring psychopathology. Adolescents with recent onset should be screened carefully to detect the emergence of the desire for gender transition in the context of trauma as well as for any disorders such as schizophrenia, mania, or psychotic depression that may produce gender confusion. When present, such psychopathology must be taken into account in assessing the adolescent’s desire to transition as well as how supportive to be of that desire. The presence of significant psychopathology does not preclude transition but suggests that one should proceed more slowly in establishing the diagnosis of GD as well as ensuring that the psychopathology is treated and controlled to the extent that it will not interfere with adherence to a safe transition plan. Even the presence of a disorder with psychotic features does not exclude a GD diagnosis if the gender concerns persist when the psychosis remits. In addition, ASDs may occur with increased frequency in individuals with GD. Impaired abilities to articulate inner experiences and to assess how they are perceived by others, together with cognitive rigidity and an intolerance of ambiguity, may present special difficulties in gender identity formation in individuals with ASDs. While these limitations do not preclude gender transition in those with ASDs, they, nevertheless, present challenges both in the assessment and treatment of GD.

With the beginning of puberty, the development of the secondary sex characteristics of the gender assigned at birth often triggers or exacerbates the anatomical dysphoria of adolescents with GD. In the last three decades, the option has become available for patients approaching puberty with severe GD and minimal, if any, coexisting psychopathology to have puberty suspended medically in order to prevent or to minimize the development of unwanted secondary sex characteristics, some of which cannot be fully reversed with subsequent hormonal or surgical therapies. Many adults who transitioned gender without the benefit of pubertal suppression express regret over the irreversible body changes that occurred during puberty. Withholding puberty suppression for adolescents is not a neutral option as it can prolong and exacerbate distress, and by allowing pubertal changes of the gender assigned at birth to proceed, contribute to a gender-atypical appearance that could provoke abuse and stigmatization. Current evidence suggests that pubertal suspension can be done for a period of up to several years, during which time the patient, in consultation with clinicians, can decide whether it is preferable for the adolescent to revert to living in the role of the gender assigned at birth or to continue gender transition with gender-affirming feminizing or masculinizing hormone therapy. There are currently few data regarding the best timing for initiating hormone treatment in adolescents, and no studies comparing outcomes when such treatment is initiated in adolescence as opposed to adulthood. In those with DSDs who have been gonadectomized, puberty consistent with the adolescent’s experienced gender is generally induced at the time their peers are entering puberty. One study of carefully selected individuals in the Netherlands, however, found that, as assessed by satisfaction with surgery and lack of regrets, the outcome was generally better in individuals who initiated reassignment as adolescents compared to those who initiated reassignment as adults. A longitudinal follow-up study by the same group found that when pubertal suppression was employed in very carefully selected minors, following age-appropriate pubertal induction and GCS in adulthood, subjects scored as well or better than the general population on measures of psychological and social adjustment.

The WPATH SOC v8 requirements for puberty-suppressing hormones are that the adolescent reached Tanner stage 2 of puberty; the diagnostic criteria of GI as per the ICD-11 have been met in situations where a diagnosis is necessary to access care; the experience of GI is marked and sustained over time; the adolescent demonstrates the emotional and cognitive maturity required to provide informed consent/assent; and any mental health concerns that could interfere with diagnosis or capacity to consent/assent have been addressed. Parents/guardians should be involved in the process unless their involvement is determined to be harmful to the adolescent or not feasible. According to the WPATH SOC v8, genital surgery should not be carried out until patients demonstrate the emotional and cognitive maturity required to provide informed consent/assent for treatment and have had least 12 months of gender-affirming hormone therapy or longer, if required, to achieve the desired surgical result for gender-affirming procedures. They must have also been sufficiently informed of the potential benefits, limitations, and risks of the desired treatment in order to give informed consent/assent. Chest surgery in adolescents assigned female at birth can be carried out earlier than genital surgery, but preferably after “ample time of living in the desired gender role” and after 1 year of testosterone treatment.

Treatment of Adults

Transgender people are sufficiently commonplace that even adult psychiatrists whose practice does not focus on gender diversity may encounter patients who are transitioning gender or contemplating gender transition. Thus, even psychiatrists who are not gender specialists should be prepared to provide informed and respectful care to gender-diverse patients, including accurate diagnostic assessments and culturally attuned treatment plans that include not only the provision of immediate care but the arrangement of necessary referrals for those who are, or who are contemplating, transitioning gender. All clinicians should be aware of the federal, state, local, and regulatory agency policies that pertain to the provision of care to transgender and gender-diverse patients in their workplace. Adult patients coming to a gender identity clinic often present with straightforward requests for hormonal and surgical gender reassignment. When GD is severe and intractable, gender reassignment is often the best solution.

Mental health professionals who work with individuals who plan to transition using hormonal and/or surgical treatments or who are in the process of doing so need to be knowledgeable about these treatments, particularly their mental health implications. All should be aware that hormonal regimens may lead to infertility and should receive appropriate counseling and referrals if preservation of reproductive potential is desired. Those assigned male at birth should be aware that estrogen treatment will cause erectile dysfunction as well as atrophy of the testes and penis.

As discussed previously in this chapter, which, if any, medical and surgical procedures to employ in one's transition plan varies from one individual to the next. Similarly, some seek mental health services for psychological support or to assist in decision making, while others do not.

FTM transition usually includes the use of androgens, which produce male secondary sex characteristics, including beard growth and male distribution of body hair, possible alopecia, deepening of the voice, and coarsening of the facial features and skin. Androgen administration also enlarges the clitoris, often to the extent that metoidioplasty becomes feasible. MTF transition often consists of both estrogen administration in conjunction with agents that either suppress androgen levels or interfere with androgens' actions. Estrogen effects include breast development and feminizing changes to skin and hair, although, for many who transition after completion of male pubertal development, depilation will be needed. Surgical reduction of the laryngeal cartilage or facial surgeries may be needed for a convincing feminine appearance. The use of hormonal preparations is much more effective in "adding" physical sexual characteristics than in "subtracting" those that have already developed with natural puberty. Utilization of either androgens or estrogens carries the potential for both added health risks and, in some cases, physiologic benefits. Just as cisgender men and women have different profiles of health risks, some of which are related to differences in sex hormones, gender-affirming hormonal regimens may shift some risks in the direction of an individual's affirmed gender. Similarly, normal ranges for some laboratory bloodwork measures also vary slightly between cisgender men and women and are subject to shift with gender-affirming hormone treatment. Psychiatrists may prescribe and monitor the effects of hormonal preparations used in gender transition, but this aspect of transition is best managed by endocrinologists or primary care physicians with training in this area. All of the treating clinicians should be aware of potential psychological effects, including the emotional changes that may occur with the use of either androgen or estrogen supplementation. Increased libido usually occurs with androgen use. Some who transition MTF also experience a stronger interest in sex, perhaps due to increased self-confidence after physical feminization. Often, individuals who are transitioning benefit from ongoing psychiatric care. In addition to the psychotherapeutic work involved in the context of life-changing experiences, psychiatric care may include monitoring the psychiatric effects of hormone use along with the prescribing physician and monitoring adherence with medical monitoring. Safer sex information, and instruction in self-protective negotiations in sexual settings, should be provided by the psychiatrist if this has not been done by the prescribing clinician. This information should be tailored to the needs and experiences of transgender persons.

Surgeries for purposes of gender transition include breast and chest ("top") surgeries as well as genital and gonadal ("bottom") procedures. The goal of masculinizing top surgery is not mastectomy, as would be performed for the treatment of cancer of the breast, but the creation of a natural-appearing male chest. Some transgender women have breast augmentation surgery due to achieving minimal breast development with hormonal treatment alone. Many adults undergoing feminizing genital surgery receive penile inversion vaginoplasty with clitoroplasty, labiaplasty, and orchiectomy.

Feminizing genital surgery can consist of either metoidioplasty with limited scrotoplasty or more extensive surgery, including phalloplasty with grafted tissue from another body site, urethral extension, scrotoplasty, and vaginectomy. Hysterectomy and oophorectomy are usually performed in either case. Transgender masculine patients may also seek oophorectomy as a gender-affirming procedure or for medical indications. In addition to electrolysis or laser treatments for removing unwanted facial hair, many transwomen also benefit from voice training that focuses on five parts of speech: pitch, intonation, volume, articulation, and resonance to improve their female gender expression. Some transgender adults may seek out facial feminization or masculinization surgery as well as surgeries to masculinize or feminize the voice. Voice coaching is usually recommended prior to surgeries. Transgender women may also desire a tracheal shave to reduce the prominence of the Adam’s apple. Feminizing voice surgery is sometimes performed at the same time as the tracheal shave.

In order to give truly informed consent, patients should be able to describe the procedures they will receive in reasonable detail, including realistic outcomes and risks, as well as what, if any, convalescent period and medical follow-up will be involved. To demonstrate readiness for each step of the treatment, they must also demonstrate the capacity to adhere to its demands, including necessary medical monitoring and postsurgical care.

Maria, who was assigned female at birth, reported being a tomboy as a child. In childhood, she had both boys and girls as playmates, but her best friends were boys. She enjoyed sports and boy's games and had no interest in dolls. She resisted wearing skirts and dresses and preferred boyish clothing. She reported feeling different from other girls, a feeling that intensified when she entered puberty, at which time she wore loose-fitting shirts to conceal breast development and became markedly distressed with the onset of menses. She recalled wishing she had been born male. From her earliest awareness of sexual attractions, they were directed toward girls, and by the age of 15, she began to label herself a lesbian and began to socialize with lesbian and gay teens. Nevertheless, she felt different from the other lesbians in a way she could not quite describe. She wore masculine clothes, baggy shirts, concealed her breasts by binding them, and kept her hair cropped short.

In her second year of college, she began dating a woman with whom she fell in love, and the two began living together as a couple. In the relationship, she realized she experienced herself as a man rather than as a lesbian, began to explore transgender websites, and enrolled in a human sexuality course at college. Now using male pronouns, *he* found the identity of *transgender man* aptly labeled his experienced gender, began to think of transitioning, and discussed the possibility with his partner, who was supportive. He changed his name to Mark, began to wear only male clothing, and requested that family, friends, and colleagues address him as a man. He experienced a level of authenticity living as a man that he had not previously experienced as a lesbian. He referred himself to a medical school–affiliated clinic that offered hormonal and surgical gender confirmation services. After careful evaluation, a transition plan was formulated, which included a mastectomy and male breast construction. Testosterone therapy for lowering the voice, beard growth, and masculinizing the body was temporally delayed to allow time for Mark's ovaries to be stimulated to harvest eggs for cryoprotection to allow future reproduction (with his partner expected to carry the pregnancy after fertilization with donated sperm). He also entered into individual and couples psychotherapy. Upon starting testosterone therapy, his menses stopped, his voice deepened, facial and body hair grew, his sex drive increased, and his clitoris enlarged. As eggs had been harvested, Mark elected to undergo hysterectomy and oophorectomy sooner rather than later. He also elected metoidioplasty. Four years later, Mark had graduated college, was working full time, and the couple had their first child.

The APA Task Force on GD identified the following tasks of mental health professionals working with adults with GD: assessing and diagnosing patients' gender concerns according to DSM criteria and assuring that these are appropriately addressed; assessing and correctly diagnosing any coexisting psychopathology and assuring that it is addressed adequately; distinguishing between GD with concurrent psychiatric illness and gender manifestations that are not part of GD but epiphenomena of psychopathology; engaging in psychotherapy with gender-diverse individuals as indicated. This includes identifying the elements that should be addressed in therapy, including the impact of discrimination and stereotyping; ensuring that individuals who are in the process of transitioning, or who are considering or planning to do so, receive counseling from a qualified professional about the full range of treatment options and their physical, psychological, and social implications including both their potential benefits and the full range of potential limitations (e.g., loss of reproductive potential and possibilities for preservation of gametes and assisted reproduction), risks and complications; ascertaining eligibility and readiness for hormone and surgical therapy or locating professionals capable of making these ascertainments to whom the patient may be referred; and educating family members, employers, and institutions about GD.

Treatment of Individuals with DSDs Including Somatic Intersex Conditions

Some individuals with DSDs, in a proportion that varies greatly with the syndrome and assigned gender, become dysphoric in their assigned gender and may desire to transition. In DSM-5-TR, these individuals can meet criteria for the diagnosis of GD, although they were categorically excluded from the diagnosis of GID in previous versions of the DSM (but could be given the diagnosis of GID not otherwise specified). A variety of issues in the clinical care of individuals with DSDs require the expertise of mental health professionals. The clinical options and decision-making processes that bear on gender transition and reassignment overlap to some extent regardless of the presence or absence of a DSD. When a DSD is present, however, there have been fewer barriers to legal gender reassignment, and the barriers to hormonal and surgical treatments in conjunction with gender reassignment are lower. This is because some individuals with DSDs, particularly those involving somatic intersex conditions, will have been gonadectomized prior to adolescence as part of routine treatment (e.g., to reduce the risk of malignancy that is associated with some gonadal abnormalities) and thus will have been rendered sterile. They will also require the age-appropriate administration of exogenous hormones to induce puberty. Thus, concerns of gonadectomy, pubertal suppression, lifelong dependence on exogenous hormones, and infertility differ from those patients without a DSD.

Major areas of involvement of psychiatrists and other mental health professionals in the care of individuals with DSDs and GD include the evaluation of gender identity and the assessment of incongruences, if present, between gender identity and assigned gender; decision making regarding gender reassignment; psychotherapy to address significant GD in individuals with a DSD who do not transition gender; psychological aspects of the endocrine management of puberty in the context of gender reassignment; and psychological/psychiatric aspects of genital surgery in the context of gender reassignment. Whether or not GD is present, individuals with DSDs have often been subjected to a variety of significant stigma due to their condition and may benefit from psychotherapy to address it.

Some individuals may identify as being intersex in the absence of any physical or medical evidence of a DSD. As some DSDs require ongoing medical treatment or screenings, sometimes as a result of their treatment history, it is important that clinicians clarify whether they are receiving needed care. For example, some DSDs increase the risk of particular malignancies, and affected individuals may require ongoing monitoring, while some individuals with DSDs have been gonadectomized and require hormone replacement therapy to prevent osteoporosis. Others with some forms of congenital adrenal hyperplasia require ongoing treatment to prevent salt-wasting and hypotensive crises. It is, therefore, important to clarify if an individual who identifies as intersex has a DSD that requires ongoing medical treatment and if they are engaged in such treatment. Medical referrals should be offered in cases of uncertainty. Importantly, referrals can be made without challenging a patient’s intersex identity by stating that a referral is needed to determine whether they have a condition that requires ongoing treatment. In addition, if a mental health provider employs the DSM-5-TR GD specifier, “with a disorder of sex development,” the patient’s specific DSD diagnosis must be recorded and coded.

Other Clinical Issues

The 2011 IOM report on the health of LGBT people drew increased attention to the fact that compared to the general population, LGBT populations experience increased violence, barriers to health care, and poorer physical and mental health. The report further pointed out the tendency of studies to lump LGBT populations together rather than studying them separately, greatly limiting knowledge as to how disparities affect each population individually. This is particularly true for the transgender population, the smallest segment of the LGBT community. The report called for all federally funded national health surveys to include queries regarding sexual orientation and gender identity (SOGI) so that these disparities could be characterized and addressed in different sexual and gender minority populations. Reports based on these surveys consistently confirm marked health disparities in gender minority populations.

To date, however, with 27,715 respondents, the 2015 U.S. Transgender Survey (USTS) conducted by the National Center for Transgender Equality and the National LGBTQ Task Force is the largest survey devoted to the lives and experiences of transgender individuals. Findings from the online survey are overall similar to the U.S. National Transgender Discrimination Survey (NTDS) conducted in 2010 with 6,450 participants. Participants in both surveys endorsed a high rate of avoiding health care due to a history of discrimination in health care settings. For example, in the USTS, a third who had seen a provider reported negative experiences in the receipt of care, such as being verbally harassed or denied treatment due to being transgender, while 23% reported they had avoided seeking health care in the prior year due to fear of being mistreated and 33% said they did not see a health care provider during the past year because they could not afford to do so. These barriers to care greatly exacerbate the health disparities seen in the transgender community.

Violence. Transgender people are often targeted for violence based on their nonconformity with gender norms and/or their perceived sexual orientation. Assaults against transgender people, particularly transgender women of color, tend to be particularly violent, and the Human Rights Campaign has shown that a high proportion of homicides in the United States are antitransgender hate crimes. In the USTS, approximately 10% who were out to their immediate family reported having experienced violence within the family because they were transgender, and many were forced to leave home. Over half who reported being perceived as transgender while in school reported abuse, including verbal harassment, being physically or sexually assaulted because they were transgender, and 17% reported dropping out of school to escape abuse. Nearly half reported having been sexually assaulted at some time in their life, and, in the year prior to the survey, 30% of those who were employed reported having experienced difficulties at work due to being transgender, including being fired, denied promotion, or being verbally harassed. In addition, 31% reported having experienced mistreatment in the past year in a place of public accommodation, with 2% reporting physical assault.

As adults, transgender individuals report increased rates of childhood maltreatment and abuse which are predisposing factors for the later onset of psychiatric symptoms and disorders, including depression, post-traumatic stress disorder, and substance abuse. In the NTDS, those who endorsed a transgender or gender-nonconforming identity in grades K-12 reported extremely high rates of harassment (78%), physical assault (35%), and sexual violence (12%). Harassment was so severe that 15% left a school in K-12 settings or in higher education. Perpetrators of harassment and abuse were reported to be teachers as well as peers, with teacher perpetrators being associated with worse outcomes with respect to HIV infection and substance abuse. Transgender individuals are also among the groups most likely to be impacted by intimate partner violence.

Suicide. There is evidence indicating that transgender populations are at increased risk for suicide. Because gender identity is not systematically recorded at the time of death, the number of transgender deaths from suicide (or any other cause) is not known. However, high rates of suicide have been reported among individuals receiving medical interventions for gender transition. Nondisclosure of transgender status is associated with a lower risk of suicide attempts, possibly because hiding protects against rejection, discrimination, and victimization, which are otherwise strongly linked to suicide attempts in transgender populations. In surveys, 25% to 43% of transgender adults have reported lifetime suicide attempts.

In the NTDS, 41% of respondents reported attempting suicide compared to 1.6% of the general population, with rates rising for those who lost a job due to bias, had a history of being harassed/bullied in school, had low household income, or were the victim of physical assault or sexual assault. Discrimination was pervasive throughout the entire sample; however, the impact of antitransgender bias was particularly marked when it was combined with racial discrimination. People of color, in general, fared worse than White participants, with African American transgender respondents faring worse than all others in many areas examined. In the USTS, 39% endorsed having experienced severe psychological distress in the month prior to the survey, and 40% reported a lifetime history of attempted suicide, approximately nine times the national average.

Assessment and treatment of suicide risk in transgender individuals requires practitioners to be familiar with the cultural milieu of their lives. Not only should clinicians be cognizant of the impact that antitransgender or *transphobic* attitudes can have on normal psychosocial development, but they must also recognize extreme situations that may arise. This requires awareness of and asking about events that a patient—perhaps in the closet or fearing lack of support from a medical/psychiatric authority—may not offer spontaneously. For example, is a child, adolescent, or adult being bullied or harassed for gender-atypical behavior? Has an adolescent or adult been rejected after coming out to a family member or friend? Has an adolescent been thrown out of their home by unaccepting parents? Is a desperate transgender patient unable to access needed medical care for transition? These are just some examples of the kind of culturally sensitive knowledge clinicians need to access if they are to provide support to this at-risk patient population.

Mental Health. Survey studies have consistently found transgender men and women to report higher rates of substance abuse, suicidal ideation, suicide attempts, and stress-related psychiatric disorders (e.g., anxiety, depression, and somatization) compared to national norms. As noted, many report misusing substances to cope with stigma and discrimination due to their gender expression. Similarly, self-reported social stigma and discrimination have been found to be correlated with the degree of psychological distress. These findings based on self-report are similar to findings based on examining the electronic health records of transgender U.S. veterans who received care through the Veterans Health Administration. Specifically, transgender veterans were found to have an increased prevalence of post-traumatic stress disorder, depression, alcohol abuse, and suicidality. Examination of mortality data revealed that veterans with transgender-related ICD-9 diagnoses have an increased suicide rate, with the average age of suicide being 49.4 years.

Transgender individuals and those with GD may also have psychiatric disorders that are independent of their gender concerns, and while even serious mental illnesses such as major depression, bipolar disorder, borderline personality disorder, and schizophrenia are not absolute contraindications to gender transition, they must be adequately treated so as not to interfere with adherence with the transition regimen. Very importantly, the manifestations of coexisting disorders should not be mistaken for manifestations of GD that will resolve with the transition. Those in transition should understand that coexisting psychiatric disorders are very likely to persist and require treatment after transition.

Medical Illness. Little research has focused specifically on physical health disparities in transgender populations beyond sexually transmitted infections among transgender women and the effects of hormone therapy on physical health. There have been reports of increased risk of venous thromboembolic disease associated with feminizing hormone therapy, but this appears to have been primarily associated with the use of ethinyl estrogen, found in birth control pills and no longer used for gender-affirming treatment in the United States. There have also been reports of elevated prolactin levels associated with feminizing hormonal therapy; however, this may be specifically linked to the use of cyproterone acetate as an adjunctive antiandrogen and has not been reported when spironolactone is used to block androgen action or when gonadotropin-releasing hormones are employed to decrease androgen production. As gonadal hormones are needed for the maintenance of bone mineral density, it is important, especially while taking antiandrogen therapy or after gonadectomy, for transgender patients to adhere to their gender-affirming hormone replacement regimens. A case-control study of disparities in over 5,000 transgender veterans found an increased prevalence of 17 out of 18 medical diagnoses, including chronic obstructive pulmonary disease, diabetes mellitus, and HIV infection. Many of the observed disparities are likely the result of barriers to health care in transgender individuals and not their gender-affirming regimens. Childhood and adolescence are times of good physical health. Thus, studies exploring the physical health of children with GD are lacking.

Children. Primary modes of therapy utilized in working with children with GD include individual insight-oriented psychoanalytic or psychodynamic psychotherapy; protocol-driven psychotherapy such as behavior modification; parent and peer relations-focused therapy; and parent and child therapeutic groups. Additional interventions include support groups for primary caregivers, community education through websites and conferences, school-based curricula, and specialized youth summer camps. The primary focus of intervention is often the primary caregivers. Depending on the treatment approach chosen, work may include parenting support and psychoeducation, careful guidance regarding behavior modification efforts and their potential to damage self-esteem, and instruction in techniques for building self-acceptance and resilience in the child. Some interventions are multifaceted and involve the school and community as well as the child and family. These include diversity education and steps to prevent bullying. As stated, psychotherapeutic efforts to change the gender identification in prepubescent children with GD have provoked angry responses in the LGBT community, with some arguing that they represent a form of conversion or reparative therapy and with some jurisdictions legislatively defining them as such. Given that most children with GD may experience spontaneous remission of their symptoms, it is not clear how, if at all, these therapeutic interventions influence the rate of desistence.

Adolescents and Adults. While GD in adolescents or adults sometimes remits, there is no convincing empirical evidence for remission in response to psychotherapeutic interventions, although there are anecdotal reports. Consequently, as such approaches have not been well studied in the last century, let alone proven, they are currently not considered effective by mainstream psychiatry and other mental health professions. This lack of psychotherapeutic responsiveness in part underlies the longstanding approach of assisting gender dysphoric individuals with transition rather than trying to change their gender identity. This entails supportive therapy and psychoeducation to ensure that the individual understands the long and often difficult process of a physical gender transition and has fully explored other options, including not transitioning (and living as a homosexual person if one is sexually attracted to members of their sex assigned at birth), transitioning socially without medical or surgical intervention, or selecting which endocrine and surgical alternatives to pursue. Psychotherapy can assist patients in exploring options, including social, hormonal, and surgical, for finding authenticity in their experience and expression of gender.

ETHICAL AND LEGAL ISSUES

Legal Change of Gender

Depending on what country and jurisdiction they live in, transgender individuals face a range of different regulations if and when they seek a change of gender on legal documents. Some U.S. states (Kansas, Tennessee) currently refuse to amend the gender on birth certificates; Tennessee's statute specifically forbids amending the sex designation on birth certificates. A majority of states permit one's name and sex to be changed on birth certificates, either through amending an existing birth certificate or issuing a new one; however, GAS is a prerequisite in many states, some of which require documented infertility. All U.S. states allow a legal change of gender on drivers' licenses, although requirements for doing so vary from state to state. Countries without marriage equality may not recognize the marriage of two individuals whose identity documents show they are the same gender. For example, a transgender woman who had not changed the gender marker on her identity documents might not be able to marry a man, whereas, if she had changed her identity documents, she would not be able to marry a woman. Surviving transgender spouses have been denied inheritances and surviving spousal rights in many cases.

In the United States, the gender designation on federally issued documents such as passports can now be changed with a single physician's letter that can be issued without requiring either hormonal or surgical treatments. However, some countries will not allow legal changes of gender unless the individual has been rendered infertile via surgical means.

Antidiscrimination Protections

Title VII of the U.S. Civil Rights Act prohibits discrimination on the basis of sex which has been interpreted in several federal cases since 2004 as extending to claims of discrimination based on gender identity or failure to conform to stereotypical notions of masculinity or femininity. In 2012, the Equal Employment Opportunity Commission, the federal agency in charge of enforcing employment discrimination laws, declared that antitransgender bias was sex discrimination under Title VII. In 2014, Attorney General Eric Holder issued a memo clarifying that the Justice Department will no longer assert that Title VII's prohibition of gender discrimination does not encompass gender identity, per se. In 2020, in the case of Bostock v. Clayton County, Georgia, the U.S. Supreme Court decision declared that Title VII workplace protections applied to transgender (and sexual orientation) status. Currently 25 U.S. states, the District of Columbia, Guam, Puerto Rico, and the U.S. Virgin Islands have statutes that protect against discrimination on the basis of sexual orientation or gender identity in both the public and private sector.

In 2021, President Biden issued Executive Order 13988 on Preventing and Combating Discrimination on the Basis of Gender Identity or Sexual Orientation requiring each federal agency to fully implement statutes that prohibit sex discrimination and policy set forth in the order. The Department of Housing and Urban Development then issued a memorandum clarifying that the Fair Housing Act's prohibition against sex discrimination in housing extends to discrimination on the basis of actual or perceived sexual orientation or gender identity. Currently, the majority of U.S. states do not provide state-level antidiscrimination protections for transgender individuals.

Gatekeeping and Informed Consent Models

Many countries with national health care systems require mental health professionals to act as *gatekeepers* of transition services for transgender people. This gatekeeping function is ostensibly to reduce the incidence of post-transition regrets, particularly for those undergoing irreversible surgical procedures. Yet there is a somewhat irrational *gender policing* quality to this requirement since few other surgical procedures, including those considered high risk, require psychiatric evaluation and approval. Finally, there is no empirical evidence that all prepubescent children who socially transition but then desist can simply transition back to the gender assigned at birth without complication. Given the complexity involved in the first social transition, it seems unlikely that transition back to the original gender would not carry its own set of risks. Definitive answers are unlikely to be forthcoming promptly given both the small numbers of children who transition gender and the ethical considerations that preclude randomized controlled trials. The informed consent model provides multiple opportunities to prepare parents and patients for the choices they must make with a realistic understanding of the range of outcomes. Some clinics in the United States now employ an "informed consent model" for transgender endocrine care that differs from the gatekeeping model by a focus on informed consent as the threshold for the initiation of gender-affirming hormonal therapy. While less emphasis is placed on mental health care unless a patient requests it, when significant mental health concerns are present that may diminish the patient's ability to safely adhere to treatment, they need to be addressed before hormones are prescribed, even in this model.

Access to Care for Incarcerated Individuals

The Prison Rape Elimination Act (PREA) was passed by the U.S. Congress in 2003 and became effective in 2012. This act specifies that individuals entering prison who identify as transgender should not be placed in gender-specific facilities, housing units, or programs based solely on their external genital anatomy but that their own views with respect to safety will be taken into consideration in housing. How PREA is enforced outside of federal prisons varies from state to state. Similarly, access to gender-affirming hormones and surgeries varies across jurisdictions. In recent years, transgender advocacy groups have had some success in lawsuits against prison facilities for denying access to hormone treatment, arguing that denial of such needed medical treatment constitutes “cruel and unusual punishment.” Many carceral facilities now have policies to continue gender-affirming hormones that were started prior to incarceration, while others arrange for inmates to be assessed for GD and started on affirming hormones after incarceration. In 2017, California became the first state to provide GAS for an inmate, and in 2019, the 9th Circuit Court of Appeals ruled that the state of Idaho must provide GAS for an inmate, and the surgery was done in 2020. Other prison systems now have programs for assessing inmates’ eligibility and readiness for gender-affirming surgeries. In 2023 the first inmate in the custody of the federal Bureau of Prisons, received GCS.

Ethical and Legal Issues in Child Treatment Controversies

Clinical approaches to gender-diverse children raise many as yet unanswered ethical questions. For example, some transgender advocates maintain that clinical attempts to prevent transsexuality in prepubescent gender-diverse children, no matter how well meaning, are always unethical. There are presently no controlled studies that demonstrate that discouraging intense, atypical gender behavior and interests in childhood does, in fact, reduce persistence or prevent transsexuality. Further, rigorous studies are not feasible due to the ethical concerns they would raise together with the fact that parents would be unlikely to allow the treatment of their child to be assigned at random. As previously stated, there are no reliable indicators to distinguish children whose GD will continue into adolescence from those whose GD will not. Since no clinician can accurately predict the future gender identity of any particular child, it is reasonable to assume that efforts to discourage gender-atypical play and identifications may be experienced as hurtful and possibly even traumatic by children whose GD continues into adolescence and adulthood. Also, it is possible that opposing a wish to explore cross-gender expression is harmful to some children, and not just persisters. Whatever the outcome regarding their transgender behavior or identity, these children may internalize disapproving attitudes toward GNC behavior and expression (*transphobia*) with possible consequences for adult development. Is it ethical to delay the social transition of children in whom GD will persist in order to prevent possible psychological harm to those who will desist? Put another way, are the children who will grow up to be transgender being subjected to unnecessary stress in order to preserve the well-being of those who will not? Finally, there is no empirical evidence demonstrating that a prepubescent child who is permitted to socially transition but then desists can simply and harmlessly transition back to the gender assigned at birth. Given the complexity involved in the first social transition, should the claim that transition back to the original gender is entirely without risks and pitfalls be questioned? In the absence of empirical studies, is permitting early social transition without a verifiable system of distinguishing future persisters from desisters ethical? Further, would conducting such empirical studies even be ethical?

While clinicians debate some of the ethical questions outlined above, forces outside the clinical setting are weighing in. In 2020, a U.K. lawsuit against the Tavistock Clinic's Gender Identity Development Service (GIDS) in London was heard by the Royal Court of Justice. The Court decided that any medical treatment, mentioning puberty blockers in particular, of a child under the age of 16 would require court authorization. That decision was overturned on appeal in 2021, with the Appeals Court noting "it was for clinicians rather than the court to decide on competence." Nevertheless, in response to complaints about long waiting lists for evaluating new patients, the GIDS is scheduled to be closed in 2024 to be replaced by two regional centers at existing children's hospitals in the United Kingdom.

Also, at the time of this writing, in the United States, at least 16 state legislatures have introduced bills to make it a crime, and in some states a felony, for health care professionals to provide medical or surgical treatment to minors before the age of 18. The rationale for introducing such legislation has been couched in names like “The Vulnerable Child Protection Act” (Florida) or “The Save Adolescents from Experimentation (SAFE) Act” (Arkansas). In 2021, the Arkansas legislature sent their bill criminalizing gender transition services to the Governor, who subsequently vetoed it, calling the legislation an unwarranted intrusion of government into the relationships between health care professionals and their patients. However, his veto was overridden by both houses of the legislature, making Arkansas the first state to pass such legislation. The law was challenged in federal court and, after a trial in 2023, the judge declared “The Court finds that the State has failed to prove that its interests in the safety of Arkansas adolescents from gender transitioning procedures or the medical community’s ethical decline are compelling, genuine, or even rational” and in violation of the U.S. Constitution’s First and Fourteenth Amendments. At the time of this writing, it is likely that the State of Arkansas will appeal.

TRAINING AND RESEARCH NEEDS

Training

Transgender presentations are sufficiently common that even adult psychiatrists whose practice does not focus on gender diversity encounter patients who are transitioning gender or contemplating gender transition. However, GD and gender transition receive little attention in medical school or psychiatric residency curricula. In a survey of deans of medical education at over 100 medical schools across North America, 34% rated coverage of SOGI minority health care as poor to very poor, with only 30% rating it as better than fair. Further, on average, only 5 hours were devoted to training in caring for LGBT patients though out the 4-year medical school curriculum, and only 30% of schools provided any training on gender transition.

In a survey of psychiatrists, many reported difficulty distinguishing GD from “gender confusion” that may occur as an epiphenomenon of psychiatric illnesses, notably personality, mood, dissociative, and psychotic disorders. The authors suggested that GD is likely to be underdiagnosed, leading to inappropriate treatment. These surveys of medical professionals’ training are consistent with the NTDS, in which 50% of transgender respondents believed that clinicians were not aware of their unique needs. In recognition of these deficiencies, calls for increased clinician training in the provision of care to transgender individuals have been made by several distinguished and authoritative bodies, including the Institute of Medicine, the Department of Health and Human Services, and The Joint Commission. Ideally, training in SOGI minority issues, including gender diversity transgender, and GD, should be embedded within the broader curriculum of human sexuality and integrated across all levels of medical school and residency training. SOC v8 recommends all health care professionals working with adolescents have expertise in gender identity development and gender diversity in children and adolescents and the ability to assess capacity for assent and consent, and possess general knowledge of gender diversity across the lifespan. Psychiatric training should include (1) the evolution of concepts of gender, gender diversity, and associated terminology; (2) the evolution of medical approaches to gender diversity, including the changing roles of mental health professionals in transgender health care as reflected in successive versions of the WPATH SOC and the emergence of informed consent models; (3) common child and adolescent presentations; evaluation of gender-diverse youth; assessment and management of any coexisting psychopathology in minors; treatment options, including pubertal suppression; persistence and desistence of GD of childhood; and family concerns; (4) common adult presentations, the process of gender transition and other options for authentic gender expression; assessment and management of concurrent psychiatric illness, and stage of life concerns; (5) mental health assessments for gender-affirming hormones and surgery; (6) presentations, evaluation, and management of GD in patients with somatic intersex conditions; (7)

complex presentations, nonbinary gender identities, and the role of the mental health professional in alternative models of treatment; and (8) relevant jurisdictional and institutional policies including those pertaining to respectful treatment of gender diverse individuals as well as those that affect access to gender transition services. While the goal of psychiatric residency training is not to produce gender specialists, the general psychiatrist, regardless of specialty or work setting, should be prepared to perform a culturally competent initial assessment of gender-diverse individuals, including a formulation based on the current version of the DSM, provide initial treatment, and provide appropriate referrals.

Research

Some future research directions on the origins of gender identity and GD are addressed previously in this chapter. Other areas of research include the following:

1. Identification of health disparities and barriers to care for transgender individuals is a first step toward addressing them. Much of what is presently known comes from convenience samples which are not fully reliable. The U.S. Department of HHS is now gathering systematic data. Larger-scale studies will become possible if both sex assigned at birth and current gender identity are routinely gathered as demographic variables in health care settings and recorded in EHR as recommended by HHS, IOM, and Joint Commission.
2. Mental health disparities, including mood and anxiety disorders, substance abuse, and suicidality, have been detected in studies to date. It is important to research the causes of these disparities and to find effective means of intervention. Considerable research suggests minority stress and discrimination, including harassment and bullying, play a significant role. Emerging research suggests family acceptance and peer support as means of ameliorating the effects of minority stress. More studies are needed to elucidate not only effective means for overcoming the effects of minority stress but to identify the causes of intolerance of gender diversity and to address them.

3. There is a need to compare the quality-of-life outcomes associated with diverse treatment approaches to gender diversity in childhood. Robust experimental designs, however, such as randomized trials in this area, are neither ethical nor feasible. Very few parents would allow the treatment of their child to be randomly assigned. Concern has been expressed that attempts to “normalize” a child’s gender expression may have long-term harmful effects on self-esteem. Longitudinal follow-up studies without randomization could potentially address this issue.
4. More longitudinal studies following pubertal suppression are needed. Recent studies by Dutch researchers found that when puberty suppression was done in very carefully selected minors, following GCS as adults, they scored as well or better than the general population on measures of psychological and social adjustment. More studies are needed to determine if these results will generalize to those receiving this treatment in other clinical settings. Since pubertal suppression is a relatively recent intervention, long-term follow-up studies are needed to rule out long-term risks.
5. What are the roles of mental health professionals in optimally addressing the needs of gender diverse individuals, including those considering, planning, or in the process of transition? Follow-up studies are needed to examine how particular mental health interventions influence long-term adjustment and subjective satisfaction. Matching comparison groups—those treated with particular mental health interventions versus those who were not—for potentially confounding variables, however, would be difficult.

Table 21.4–2. Internet Resources for Professionals and for Transgender Individuals and Their Families

American Medical Association LGBTQ Advisory Committee on LGBTQ Issues (public education for patients and clinicians)	http://www.ama-assn.org/ama/pub/about-ama/our-people/member-groups-sections/glbt-advisory-committee.page? https://www.ama-assn.org/member-groups-sections/advisory-committee-lgbtq-issues
AGLP (education and professional support for psychiatrists, residents and medical students)	http://www.aglp.org/
Gay, Lesbian, and Straight Education Network (GLSEN) (support for LGBT youth in the school system)	http://www.glsen.org/students/tsr
GLAAD: Transgender 101 (public education)	http://www.glaad.org/transgender/trans101
GLMA (education and professional	http://www.glma.org/

support for physicians and other health professionals)	
Group for the Advancement of Psychiatry (GAP): LGBT Mental Health Syllabus (for medical students and residents)	http://www.aglp.org/gap/
Hetrick-Martin Institute (social support for LGBT youth)	http://www.hmi.org/
Lambda Legal (advocacy group for legal protections)	http://www.lambdalegal.org/issues/transgender-rights
PFLAG (support groups for families)	http://community.pflag.org/transgender
SAGE (service and advocacy group for LGBTQ+ elders)	http://www.sageusa.org/issues/transgender.cfm
VHA LGBTQ+ Health Program (advocacy and links to LGBTQ+	https://www.patientcare.va.gov/lgbt/

patient care coordinator at every VA)	
Society for the Psychological Study of Lesbian, Gay and Bisexual Issues—Division 44 of the American Psychological Association (public education and professional support)	http://www.apa.org/about/division/div44.html
World Professional Association for Transgender Health (WPATH) (professional support and standards of care)	http://www.wpath.org/

INTERNET RESOURCES

The websites listed in Table 21.4–2 offer useful resources for gay men, lesbian and bisexual patients, their families, and mental health professionals themselves.

FURTHER READINGS

Benjamin H. *The Transsexual Phenomenon*. Julian Press; 1966.

Blanchard R. Early history of the concept of autogynephilia. *Arch Sex Behav*. 2005;34:439–446.

Bockting W, Coleman E, Deutsch MB, et al. Adult development and quality of life of transgender and gender nonconforming people. *Curr Opin Endocrinol Diabetes Obes*. 2016;23:188–197.

Brown GR, Jones KT. Mental health and medical health disparities in 5135 transgender veterans receiving healthcare in the Veterans Health Administration: a case-control study. *LGBT Health*. 2016;3:122–131.

Byne W, Bradley SJ, Coleman E, et al; American Psychiatric Association Task Force on Treatment of Gender Identity Disorder. Report of the American Psychiatric Association Task Force on Treatment of Gender Identity Disorder. *Arch Sex Behav*. 2012;41(4):759–796.

Coates S, Wolfe S. Gender identity disorder in boys: the interface of constitution and early experience. *Psychoanal Inq*. 1995;51:6.

Dhejne C, Öberg K, Arver S, Landén M. An analysis of all applications for sex reassignment surgery in Sweden, 1960–2010: prevalence, incidence, and regrets. *Arch Sex Behav*. 2014;43:1535–1545.

Drescher J. Queer diagnoses: parallels and contrasts in the history of homosexuality, gender variance, and the diagnostic and statistical manual. *Arch Sex Behav*. 2010;39:427–460.

Drescher J. Queer diagnoses revisited: the past and future of homosexuality and gender diagnoses in DSM and ICD. *Int Rev of Psychiatry.* 2015;27:386–395.

Drescher J. Controversies in gender diagnoses. *LGBT Health*. 2014;1(1):10–14.

Drescher J, Byne W, eds. *Treating Transgender Children and Adolescents: An Interdisciplinary Discussion*. Routledge; 2013.

Drescher J, Cohen-Kettenis PT, Reed GM. Gender incongruence of childhood in the ICD-11: controversies, proposal, and rationale. *Lancet Psychiatry*. 2016;3(3):297–304.

Drescher J, Cohen-Kettenis P, Winter S. Minding the body: situating gender identity diagnoses in the ICD-11. *Int Rev Psychiatry*. 2012;24(6):568–577.

Drescher J, Pula J. Ethical issues raised by the treatment of gender variant prepubescent children. *Hastings Cent Rep*. 2014;44(Suppl 4):S17–S22.

Erickson-Schroth L, ed. *Trans Bodies, Trans Selves: A Resource for the Transgender Community*, 2nd Edition. Oxford University Press; 2022.

Fisher AD, Castellini G, Risotri J, et al. Cross-sex hormone treatment and psychobiological changes in transsexual persons: two-year follow-up data. *J Clin Endocrinol Metab*. 2016;101:4260–4269.

Goodman M, Adams N, Corneil T, Kreukels B, Motmans J, Coleman E. Size and distribution of transgender and gender nonconforming populations: a narrative review. *Endocrinol Metab Clin North Am.* 2019;48:303–321.

Grant JM, Mottet LA, Tanis J, Harrison J, Herman JL, Keisling M. *Injustice at Every Turn: A Report of the National Transgender Discrimination Survey*. National

Center for Transgender Equality and National Gay and Lesbian Task Force; 2011.

Institute of Medicine. *The Health of Lesbian, Gay, Bisexual and Transgender People: Building a Foundation for Better Understanding*. The National Academies Press; 2011.

James SE, Herman J, Keisling M, Mottet L, Anafi M. *U.S. Transgender Survey (USTS)*. Inter-university Consortium for Political and Social Research; 2015 [distributor], 2019-05-22.

The Joint Commission. *Advancing Effective Communication, Cultural Competence, and Patient- and Family-Centered Care for the Lesbian, Gay, Bisexual, and Transgender (LGBT) Community: A Field Guide*. Joint Commission Resources; 2011.

Kreukels BPC, Steensma TD, de Vries ALC, eds. *Gender Dysphoria and Disorders of Sex Development*. Springer; 2013.

Moser C. Autogynephilia in women. *J Homosex*. 2009;56:539–547.

Mueller SC, De Cuypere G, T'Sjoen G. Transgender research in the 21st century: a selective critical review for a neurocognitive perspective. *Am J Psychiatry*. 2017;174:1155–1162.

Reed GM, Drescher J, Krueger RB, et al. Disorders related to sexuality and gender identity in the ICD-11: revising the ICD-10 classification based on current scientific evidence, best clinical practices, and human rights considerations. *World Psychiatry*. 2016;15:205–221.

Serano JM. The case against autogynephilia. *Int J Transgenderism*. 2010;12:176.

Steensma TD, McGuire JK, Kreukels BPC, Beekman AJ, Cohen-Kettenis PT. Factors associated with desistence and persistence of childhood gender dysphoria: a quantitative follow-up study. *J Am Acad Child Adolesc Psychiatry*. 2013;52(6):582–590.

Wiepjes CM, Nota NM, de Blok CJM, et al. The Amsterdam Cohort of Gender Dysphoria Study (1972–2015): trends in prevalence, treatment, and regrets. *J Sex Med* 2018; 15:582–590.

World Professional Association for Transgender Health (WPATH). Standards of Care for the Health of Transsexual, Transgender and Gender Non-Conforming People, 8th Version, 2011. http://www.wpath.org/

Wylie K, Knudson G, Khan SI, Bonierbale M, Watanyusakul S, Baral S. Serving transgender people: clinical care considerations and service delivery models in transgender health. *Lancet*. 2016;388:401–411.

Yarbrough E. *Transgender Mental Health*. American Psychiatric Association Press; 2018.

▲ 21.5 Sexual Addiction

MARCY VERDUIN, M.D.

A disorder in which a person's driven sexual behavior is poorly controlled and continues despite harmful consequences has a documented history of more than two millennia. The Pali Canon, the standard scriptures of Theravada Buddhism that were compiled from oral tradition and written in 29 BCE, includes the following passage:

> Then early in the morning the Blessed One, having put on his robes and carrying his bowl and outer robe, went into Savatthi for alms. He saw that most of the people in Savatthi were excessively attached to sensual pleasures, that they lived delighting in, addicted to, infatuated with, fastened to, absorbed in sensual pleasures. Then, on realizing the significance of that, the Blessed One on that occasion exclaimed:
>
> Blinded by sensuality, covered by the net, veiled with the veil of craving, bound by the bond of heedlessness like fish in the mouth of a trap, they go to aging and death like a suckling calf to its mother. [Udana VII.4 Kāmesu Satta Sutta, tr. Thanissaro Bhikkhu]

Psychiatrist Richard von Krafft-Ebing, who published on this topic in the late 1800s, identified this condition as *hypersexuality*, which is generally defined as excessive sexual interest (or urges) or activity. It was an advance over its predecessors, *satyriasis* and *nymphomania*, which are defined as uncontrollable or excessive sexual desire in a man and in a woman, respectively. The definition of *hypersexuality* is more closely related to sexual behavior, which is more objectively observable and quantifiable than sexual desire. In addition, hypersexuality is gender-neutral.

In 1945, the psychoanalyst Otto Fenichel coined the term *sexual addiction* to designate the condition of driven sexual behavior that is under discussion. Whereas excessiveness in the definition of hypersexuality represents a quantitative dimension, the term *sexual addiction* directs attention toward indications of a sexual behavior's adverse consequences and a person's impaired control over the behavior. It also implies a relationship between this disorder and other disorders, the names of which also include the word *addiction*.

The next section addresses terminologic and diagnostic matters that pertain to this condition. Before proceeding, three characters are introduced whose stories give human faces to the clinical abstractions, provide examples of how a condition of driven sexual behavior may look in a person's life, and illustrate the potentially tragic seriousness of the condition. (The characters are fictional and are presented as illustrative examples, not case reports.)

Ronnie, an electrician in his mid-20s, had used alcohol and other drugs heavily in high school. He dropped out of college in his freshman year and went to work in his uncle's electrical contracting business. After a serious accident on the job, while he was intoxicated, he quit recreational use of psychotropic substances.

Over the next several months, he found that his sexual fantasies and urges to masturbate were becoming more frequent and more intense than they previously had been. Since his middle teens, Ronnie had masturbated nearly every night before going to sleep, unless he was having sex with a girlfriend or a woman he had picked up. Some mornings during his first year of being drug-free, Ronnie would get strong urges to masturbate. If he did not act on these urges, he would remain "horned up" all day and would feel restless, distracted, and irritable. Consequently, he began to masturbate in the mornings as well, even though doing so sometimes resulted in his arriving to work late.

Some months later, Ronnie began to get urges to masturbate when he was at work, particularly if he saw a woman at the job site. Again, he found that if he did not masturbate, he would become uncomfortable and unable to focus on his work. No longer sufficiently excited by his fantasies, Ronnie began to watch free online pornographic videos while he masturbated. He also discovered peer-to-peer file-sharing sites from which he could at no charge (but illegally) download better quality, longer, and more varied porn videos.

Before long, Ronnie was masturbating three or four times a day and even more when he felt under stress. He was chronically tired from lack of sleep, had received a second warning about tardiness and inattentiveness at work, and felt disgusted with himself. He tried several times to stop masturbating. But each time, sexual fantasies accompanied by arousal would intrude into his consciousness throughout the day, and he would feel as though he were going to blow up. The only way he knew to free himself from the building pressure was to masturbate. At the beginning of a new sexual relationship, he would delete his collection of pornography and resolve to quit masturbating. Within a few months, though, masturbation would creep in and take over, his sexual interest in his girlfriend would fade, and the relationship would fall apart.

As Ronnie became increasingly bored with familiar porn videos and new ones with the same old plot formulas, he began to accompany his masturbation with increasingly sadomasochistic videos. He also started using telephone sex and online sex-cam services, neither of which were free of charge. When he had “maxed out” his credit cards, he obtained new ones that he soon ran up to their limits. He got behind on his rent, and the power company threatened to cut him off unless he paid his bill.

Socially isolated, deeply in debt, and about to lose his job, Ronnie realized that his preoccupation with masturbation and pornography was ruining his life. But he felt helpless. Meanwhile, he also was beginning to sense that pornography, phone sex, and sex-cams no longer excited him as much as they used to. He started going to shopping malls specifically for the purpose of looking at women and girls, whom he would later fantasize about raping while he masturbated. When he then began to feel compelled to study the patterns of the women who worked in the malls and to compile schedules of their comings and goings, he became scared that he was in danger of acting on his fantasies. At that point, he sought help.

Lori was an MBA graduate and a highly effective manager. Yet she was unable to hold a job for more than a year. When she began a new job, her performance was exemplary. After about 2 to 6 months, she would begin to act as though she were distracted and less engaged. The quality of her work would gradually decline, and her work relationships would become strained. Eventually, she would be asked to resign and would comply.

Lori was in love with being in love. She loved the boost to her self-esteem and self-confidence; the soft ongoing sexual arousal that made her feel fully alive; the sexual excitement that thrilled her when she just thought about her lover; and the “mind-blowing” orgasms she experienced during sex with him. Being in love filled her emptiness, made her whole, kept the pain and terror out of her awareness, and gave her a reason to live.

When Lori was in love with a man, nothing else in her life felt as important to her as having sex with him. She would want sex desperately. If circumstances prevented her from having sex with him, she would become restless, then agitated, then frantic. She could interrupt this sequence by masturbating during the restless stage before she became agitated. Otherwise, she could terminate such episodes only by drinking herself into a stupor. When she was in love, she gradually neglected everything else in her life, including her work. When she was not in love, she would search desperately for a man and typically would find someone with whom to fall in love and have sex. Often, her feeling of being in love did not last. Often, the man lost interest before she did.

On several occasions, Lori was reminded that her trouble with employment was related to her love life. When she was in love, each time she tried to hold back or defer sexual activity—for example, to complete a work assignment or attend a meeting—she experienced the aforementioned sequence of restlessness, agitation, and frenzied panic. When she was not in love and refrained from her desperate quest—primarily by staying home during evenings and weekends—she experienced abandonment feelings that intensified, threatened to overwhelm her, and were increasingly accompanied by suicidal thoughts. Twice she had attempted to kill herself.

After Lori's 30th birthday, she found herself increasingly drawn to "bad boys," men with an air of danger who did not follow rules. One evening, some friends of a boyfriend who had moved in with her joined them for dinner. An hour later, a dozen FBI agents stormed in and placed everyone under arrest. Two kilos of uncut cocaine were later found in Lori's house. Lori and her boyfriend were arrested for possession with intent to distribute and operating a continuing criminal enterprise.

Lori was allowed to plead guilty to a minor charge in exchange for her cooperation with prosecution. A year or so after the trial, she began to feel hints of the familiar craving to fall in love and be swept into a delicious maelstrom of sex and passion. She had by then come to realize that many of the misfortunes in her life had resulted from following where that craving led. When the hints became stronger, she sought help.

Wayne, a college dropout in his middle 40s, had difficulty holding a job despite being a capable computer programmer and a gifted musician. He had no close friends and spent much of his free time by himself, either working on his computer or playing his electric piano through headphones. Sometimes he would go to a casino, which he said that he did in order to be around other people.

Wayne had been involved in two brief relationships with women, both of whom he drove away by his demanding possessiveness and paranoid jealousy. He tended to develop obsessive crushes on women whom he hardly knew and then would feel driven to watch and follow them. Complaints of harassment by female coworkers had led to his being fired from two jobs, and once he was arrested for stalking.

Wayne masturbated several times a day, even at work. At night, he sometimes called random telephone numbers until he found a woman who would answer his questions about her sex life while he masturbated. When he had an urge to masturbate and did not act on it, he experienced a disorganizing panic that would intensify until he felt that he was about to burst. He found that he could not resist these urges, even when succumbing could jeopardize his job or get him into trouble with the law.

The strongest urge that Wayne experienced was to expose his penis to a woman and masturbate while she looked on in shock and fear. For this reason, he particularly liked to expose himself in elevators, when he and a woman were the only passengers. Even after two arrests, one conviction, and one occasion on which a woman’s boyfriend beat him severely, Wayne felt driven to continue riding elevators and hunting for unaccompanied women. One time, Wayne targeted a woman who had once been raped. Having vowed that she would never be victimized like that again, she carried a handgun. When Wayne turned toward her with his penis exposed, she panicked and shot him. He survived, thanks in part to the woman’s immediate 911 call. Six months after being discharged from the hospital, he was arrested again for exposing his penis to a woman in an elevator.

While the three persons who are described in the case examples differed in many respects, they had one trait in common. All engaged in some form of sexual behavior in a pattern that was characterized by two key features: recurrent failure to control the behavior and continuation of the behavior despite significant harmful consequences.

DIAGNOSIS AND RELATED MATTERS

Terminology

Agreement on the name of this disorder of driven sexual behavior is not universal. The primary contenders have been *sexual compulsivity* (or *compulsive sexual behavior*) and *sex addiction* (or *sexual addiction*). *Sexual compulsivity* is the name given to this condition by investigators who argue that the symptomatic sexual behavior is compulsive because it functions to reduce anxiety and other painful affects. While phenomenologic characteristics and treatment research support classification of this condition as an addictive disorder, the clearest argument against describing its symptomatic behavior as compulsive appears in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5-TR) and also in previous editions of the DSM. The definition of compulsion (compulsive behavior) in DSM-5-TR includes its function of providing relief from anxiety and other painful affects. However, it also specifies that compulsions are not performed to produce pleasure. Although reduction of anxiety or distress contributes to a person's motivation to engage in the sexual behavior that characterizes this disorder, pleasure or gratification also contributes. The relationship between compulsion and the condition being considered here is explicitly clarified in the differential diagnosis of OCD:

Other compulsive-like behaviors. Certain behaviors are sometimes described as “compulsive,” including sexual behavior (in the case of paraphilias), gambling (i.e., gambling disorder), and substance use (e.g., alcohol use disorder). However, these behaviors differ from the compulsions of OCD in that the person usually derives pleasure from the activity and may wish to resist it only because of its deleterious consequences. (Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 270, 271 [Copyright © 2022]. American Psychiatric Association. All Rights Reserved.)

A suitable name for the condition under discussion would designate behavior that, like compulsive behavior, (1) feels driven, (2) entails harmful or unpleasant consequences, and (3) functions to reduce anxiety or other painful affects; and also that—unlike compulsive behavior—(4) functions to produce pleasure or gratification. *Addiction* is the term that best fits this description. Drug addiction, for example, has the characteristics of feeling driven, entailing harmful consequences, and functioning both to relieve painful affects and to produce pleasure.

However, determination of whether “sex addiction” is the most appropriate designation for the condition of driven sexual behavior that is being considered here requires a clear and meaningful definition of addiction.

Definition of Addiction

As noted in DSM-5-TR, neither tolerance nor withdrawal is necessary nor are both together sufficient for a diagnosis of substance use disorder. The characteristics that seem to be both necessary and sufficient for identifying a pattern of psychoactive substance use as a substance use disorder are (1) recurrent failure to control the use of one or more psychoactive substances and (2) continuation of psychoactive substance use despite significant harmful consequences. (“Recurrent failure to control” does not mean that addicted individuals invariably lose control when they use drugs. It means that their predictions that they would refrain from or limit their drug use have repeatedly proven to be unreliable.)

The preceding discussion of terminology identified two critical characteristics of addictive behavior. It can reduce anxiety and other painful affects, as does compulsive behavior. In addition, it produces pleasure or gratification (that may diminish with repetition). Combining this distinctive dual motivational function—positive as well as negative reinforcement—with the key features that distinguish drug addiction from simple drug use yields a workable, behaviorally nonspecific definition of addiction, as follows. *Addiction* is a condition in which a behavior that can function both to produce pleasure and to reduce anxiety or other painful affects is employed in a pattern that is characterized by two key features: (1) recurrent failure to control the behavior and (2) continuation of the behavior despite significant harmful consequences.

The definition of a specific addictive disorder can be derived from the definition of addiction by replacing the phrase, “a behavior that can function both to produce pleasure and to reduce painful affects” with the name of a behavior that can perform both of these functions (e.g., alcohol consumption, gambling, eating, sexual behavior). Accordingly, sex addiction is defined as a condition in which some form of sexual behavior is employed in a pattern that is characterized by two key features: (1) recurrent failure to control the behavior and (2) continuation of the behavior despite significant harmful consequences. In other words, sex addiction is a condition in which some form of sexual behavior relates to and affects a person’s life in such a manner as to accord with the definition of addiction. A sex addict is a person who engages in some form of sexual behavior addictively—that is, in a pattern that is characterized by recurrent failure to control and continuation despite significant harmful consequences.

In a reprise of terminology matters, the definition of addiction as a pattern of behavior indicates that addiction is accurately identified not as something that a person has, like cancer, but as something that a person does. Also inaccurate is the characterization of a person as addicted to a substance or behavior, which would imply that the person is a passive victim over whom the substance or behavior has control. The engine of addiction is not in a substance or behavior. It is in the person. The link between a person and an addictive disorder or pattern of behavior is most accurately described by a statement to the effect that a person engages in one or more behaviors addictively. Some common phrases such as “addicted to porn” are inaccurate not only for the reasons just mentioned but additionally because they misidentify the pathognomonic behavior. In an addictive disorder that involves pornography, the behavior that is used addictively is not the viewing of pornography. It is masturbation, for which looking at pornography serves as a facilitator or enhancer. Terminology was considered worthy of further discussion because imprecise use of “addiction” and related terms is sometimes exploited as an opportunity to discredit the term, however much the imprecision resides not in the term but in its user.

Diagnostic Criteria for Sex Addiction

DSM-5-TR excludes several behavioral addictions, including sex addiction, on grounds that peer-reviewed evidence is insufficient to establish their diagnostic criteria. Meanwhile, the generation of peer-reviewed evidence depends on investigators of this disorder having a valid and reliable set of provisional diagnostic criteria by which to identify subjects who are appropriate for their studies.

Provisional diagnostic criteria for sex addiction that follow the DSM format can be derived from the DSM-5-TR criteria for any substance use disorder, the paradigmatic addictive disorder, by replacing the name of the substance and its use—for example, *alcohol* and *alcohol use*—with the term *sexual behavior*, replacing *taken* with *engaged in*, and replacing “a characteristic withdrawal syndrome for the substance” with “a characteristic biopsychological withdrawal syndrome of physiologically and/or psychologically described changes upon discontinuation of the sexual behavior.” The set of criteria that is presented here demonstrates that diagnostic criteria for sex addiction can be developed that is consistent with the form and content of the current DSM, are expressed in the same descriptive language, and are likely to be as reliable and valid as are the DSM-5-TR diagnostic criteria for substance use disorder. To increase specificity and reliability, a diagnosis of sex addiction requires that at least four of the diagnostic criteria be met. (A diagnosis of substance use disorder requires only that two criteria be met.)

Sex Addiction

A maladaptive pattern of sexual behavior that has led to *clinically significant impairment or distress*, as manifested by the occurrence within a 12-month period of four (or more) of the following:

1. the sexual behavior is often engaged in over a longer period, in greater quantity, or at a higher level of intensity than was intended
2. efforts to cut down or control sexual behavior have repeatedly been unsuccessful
3. a great deal of time is spent in activities necessary to prepare for the sexual behavior, engage in the behavior, or recover from its effects
4. the subject experiences a craving (strong desire or urge) to engage in the sexual behavior
5. engaging in sexual behavior has resulted in failure to fulfill significant responsibilities at work, school, or home
6. important social, occupational, or recreational activities are given up or reduced as a result of engaging in the sexual behavior
7. the sexual behavior continues despite persistent or recurrent social or interpersonal problems that its effects have caused or exacerbated
8. the sexual behavior continues despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the behavior
9. the sexual behavior is recurrently performed in situations in which doing so is physically hazardous
10. tolerance, as defined by either of the following:
    a. producing the same levels of desired effects as on earlier occasions requires a frequency or intensity of sexual behavior that is markedly greater than had been required in the past

   b. engaging in sexual behavior at the same frequency and intensity as on earlier occasions produces levels of desired effects that are markedly lower than had been produced in the past
11. withdrawal, as manifested by either of the following:
   a. a characteristic biopsychological withdrawal syndrome of physiologically and/or psychologically described changes upon discontinuation of the sexual behavior
   b. the same (or a closely related) sexual behavior is engaged in to relieve or avoid withdrawal symptoms

These diagnostic criteria are far from the final word on diagnosing sex addiction. They are provisional and are likely to need revision in the light of further developments in research or theory. For example, investigators might find that the number of criteria required (four) is too low or that the required time frame (12 months) is unnecessarily long. They might determine that adding (or substituting) items from the DSM-5-TR diagnostic criteria for gambling disorder—such as preoccupation, engaging in the behavior to relieve dysphoric moods, restlessness or irritability when attempting to decrease or discontinue the behavior, or lying to conceal the extent of the behavior—enhances their discriminative utility. At the very least, these criteria provide a starting point for research. The research function of diagnostic criteria is critically important, particularly for newly recognized conditions. The inclusion of a condition in the next DSM depends on scientific research that documents its legitimacy. Meanwhile, a condition can be researched only if clear, valid, and reliable diagnostic criteria enable researchers to recognize it. Multiple varied sets of diagnostic criteria limit the generalizability of research results as well as undermine claims that the condition is a valid diagnostic entity.

Significantly, no form of sexual behavior in itself constitutes sex addiction. Whether a pattern of sexual behavior qualifies as sex addiction is determined not by the type of behavior, its object, its frequency, or its social acceptability, but by how the behavior relates to and affects a person's life. Any sexual behavior has the potential to be engaged in addictively but constitutes an addictive disorder only to the extent that it occurs in a pattern that meets the diagnostic criteria or accords with the definition. The key features that distinguish sex addiction from other patterns of sexual behavior are: (1) the person is not reliably able to control the sexual behavior and (2) the sexual behavior has significant harmful consequences and continues nonetheless.

Sex addiction can now be distinguished from the diagnostic entity in DSM-5-TR that is most similar to it, paraphilic disorder. DSM-5-TR defines *paraphilia* as “any intense and persistent sexual interest other than sexual interest [sic] in genital stimulation or preparatory fondling with phenotypically normal, physically mature, consenting human partners.” It defines *paraphilic disorder* as “a paraphilia that is currently causing distress or impairment to the individual or a paraphilia whose [sic] satisfaction has entailed personal harm or risk of harm to others.” The most apparent distinction is that sex addiction can be diagnosed when its diagnostic criteria are met whether the symptomatic behavior is paraphilic or nonparaphilic, while paraphilic disorder is limited to paraphilias. A second distinction is that impaired ability to control the symptomatic behavior is prominent in the diagnosis of sex addiction but is mentioned in neither the definition of paraphilic disorder nor the diagnostic criteria for any of the paraphilic disorders. Third, risk of harm to others as a result of symptomatic behavior is sufficient to warrant a diagnosis of paraphilic disorder but not a diagnosis of sex addiction. Fourth, DSM-5-TR defines paraphilia as an interest, not a behavior. Diagnostic criterion B for most of the paraphilic disorders is “The individual has acted on these sexual urges with a nonconsenting person, *or the sexual urges or fantasies* (author’s italics) cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.” The definition of and diagnostic criteria for sex addiction depend on objectively observable events: sexual behavior, comparison between addicts’ predictions that they would refrain from or limit their sexual behavior and their subsequent sexual behavior, and consequences of the sexual behavior. Sexual interests, urges, or fantasies per se are not relevant to a diagnosis of sex addiction.

Clinical Examples. A few clinical examples are presented to illustrate the considerations that are involved in making a diagnosis of sex addiction.

Dale, a dealer of fine art, began psychoanalysis to help him get past his lifelong depression and inability to sustain an intimate relationship. Ever since he quit drinking 6 years earlier, he had been desperately lonely. Consciously he wanted nothing more than to find a partner, someone with whom he could share his life. However, he was very anxious in social interactions. He feared that he would not be able to hold up his end of a conversation and would then be humiliated. He avoided situations in which conversation was likely. Instead, he would go to parks or adult theaters where he could meet other men for anonymous sex. After these sexual encounters, he felt dirty and disgusted with himself. Often, he would suspect that he had picked up HIV. He also had been beaten several times and robbed twice. But when he was alone or even when he anticipated being alone—for example, when he was driving home after work or after an evening activity—he would experience intense anxiety that felt like he was coming apart into a billion pieces. He could not tolerate this feeling, and sexual activity relieved it. Masturbating to pornography at home took the edge off his anxiety and usually enabled him to sleep. However, he increasingly was finding that it did not help as much as it once had done. On numerous occasions, he tried to get involved in cultural or religious activities where he would be likely to meet educated, stable gay men. Every time he soon found himself back at a park or a theater.

Steve was a software developer in his 40s who was in treatment for depression, anxiety, and marital difficulties. Sex was among the sources of tension in his marriage. He wanted to have sex with his wife every day, sometimes two or even three times a day, and he became upset when she did not comply. When his wife declined to have sex with him, he felt desperate. He feared that she did not love him, that he was not good enough for her, and that she was tired of him and was preparing to leave him. On such occasions, he usually withdrew and immersed himself in work. He rarely masturbated, and he did not seek sex elsewhere. When his wife's lupus flared up, Steve cared for her sensitively, and she expressed appreciation and gratitude. At such times, he felt needed and valued, and he rarely thought of sex. Steve's desire for sex occasionally offended his wife, who felt then that he would rather have sex than talk with her. Sometimes when she complied with Steve's requests for sex, she resented him. Further inquiry did not reveal other harmful consequences from Steve's sexual behavior, nor did it suggest that he had any difficulty controlling his sexual behavior.

Jolene was a married emergency department nurse in her early 30s who was “in love” with two married men with whom she was sexually involved and having extramarital affairs. Meanwhile, she loved her husband Ed, and could not stand the thought of leaving him. She trusted him and felt that he understood her. He knew of Jolene’s recurrent pattern of intense infatuation with a new man, resisting, giving in to overpowering sexual attraction, and then becoming bored. When she was not sexually involved with at least one man outside her marriage she felt desperate, panicky, and empty. Typically, she would begin a new involvement as soon as she sensed that she was beginning to become bored with her current extramarital affair. Sure, she noted, she had tried to rein it in, many times. After all, she continued, it had cost her two marriages, her kids, her place in medical school, a whole lot of time, and a whole lot more emotional hell. She had tried meditation, counseling, and women’s support groups. She had joined a fundamentalist church. She even had taken excessive doses of medroxyprogesterone to decrease her sex drive until she developed a pulmonary embolus, from which she almost died. Jolene had become involved with Ed while she still was married to her second husband. She lived in constant fear that Ed, too, would get fed up with her and throw her out. Now one of her lovers was talking about killing himself if Jolene did not marry him. The other lover was talking about confronting Ed. Jolene felt overwhelmed and had developed symptoms of major depression and panic disorder.

Charla, a hairstylist and fashion model in her 20s, agreed to accompany her boyfriend Mike to a session with his psychotherapist. Mike was upset that Charla was having sex with other men and thought she might be a sex addict. Charla said that she had told Mike shortly after they met that sex was part of her career, and if he did not think he could stand it, they should not get involved. Having sex with wealthy and powerful men advanced her success as a model, she explained, and brought her money and expensive gifts. She acknowledged that what she was doing was like prostitution, but it was clean and legal, and she was in charge. She stated that she certainly did not do it for the sex, which did nothing for her. She was not orgasmic with these men. She likened having sex to cutting hair. Her rewards were the finished product, a satisfied customer, and a nice tip. She also enjoyed her sense of control over these wealthy and powerful men. Charla stated that she had never engaged in more sexual behavior than she intended and never tried or wanted to decrease her sexual behavior. She also denied having experienced withdrawal symptoms or the desire to engage in substitute behavior when she was not sexually active. She recognized that her sexual behavior was bothering Mike, but she said that it was his problem. She was making ten times as much money as her father ever did and was not about to give that up because Mike’s fragile male ego could not deal with it. Mike acknowledged that she seemed to be in control of her sexual behavior and that he was not aware of any problems that resulted from her sexual behavior other than its effect on him.

The clinical example of Dale seems to meet all the criteria except 1. The example of Steve might meet criterion 3 but does not seem to meet any of the other criteria. The example of Jolene meets criteria 2, 3, 6, 8, and probably 11. Charla probably meets criterion 5. She might or might not meet criterion 6, depending on the importance (to her) of the activities with Mike that she gives up to have sex with other men. Accordingly, Dale and Jolene merit diagnoses of sex addiction according to the diagnostic criteria, while Steve and Charla do not. (The conclusion is not that Charla's behavior is acceptable nor that she is psychiatrically healthy, only that she does not merit a diagnosis of sex addiction.) An informal assessment based on the definition of sex addiction reaches the same conclusions. Dale and Jolene demonstrate patterns of sexual behavior characterized by recurrent failure to control and continuation despite significant harmful consequences, while Steve and Charla do not.

Differential Diagnosis

A number of other conditions can at first glance look like sex addiction. The hypersexual activity and paraphilic behaviors that often characterize sex addiction can occur also as manifestations of underlying brain pathology and occasionally are its earliest or most prominent symptoms. Paraphilic or hypersexual behavior can occur in the context of a seizure disorder, especially in association with temporal lobe epilepsy. More broadly, any disorder that is associated with impairment of cerebral functioning can weaken normal inhibitory controls and thereby allow the expression of sexual behaviors that ordinarily are suppressed. Disinhibited sexual behavior has been reported particularly in patients with frontal lobe lesions and in older adult patients with dementia. Hypersexual behavior can occur also as a side effect of medication, particularly antiparkinsonian agents. Finally, sexually aggressive behavior has been associated with elevated levels of testosterone. Distinguishing these conditions from sex addiction is usually facilitated by the presence of additional symptoms or circumstances that suggest the underlying organic disturbance, although altered sexual behavior may be the earliest manifestation in some cases of brain pathology. Clues that invite an organic evaluation include onset in middle age or later, regression from previously normal sexuality, excessive aggression, report of auras or seizure-like symptoms prior to or during the sexual behavior, and impaired perceptual or motor skills. The most useful instrument for determining whether a case of paraphilic or hypersexual behavior represents sex addiction is the set of diagnostic criteria for sex addiction. Patterns of paraphilic or hypersexual behavior that are not symptoms of sex addiction are generally characterized neither by a persistent desire to cut down or control the behavior nor by tolerance or biopsychological withdrawal symptoms on discontinuation of the behavior (usually affective discomfort, irritability, or restlessness). The diagnostic criteria for sex addiction are useful also in distinguishing sex addiction from nonaddictive patterns of exploitative or aggressive sexual behavior that can occur with antisocial personality disorder.

Obsessions and compulsions with sexual content can occur in OCD. Sexual obsessions were reported in 32% of OCD patients in a large study. However, the content of these obsessions consisted most often not of sexual fantasies but of fears of acting on sexual impulses or fears of being a pervert. Sexual fantasies are typical of sex addiction. As noted earlier, the symptoms of sex addiction differ from sexual obsessions and compulsions primarily in that the former is associated with sexual arousal and sexual pleasure, while the latter are typically accompanied not by sexual arousal but by anxiety.

A syndrome that meets the diagnostic criteria for sex addiction can occur in the context of other psychiatric disorders, including bipolar disorder, schizophrenia, personality disorders, and psychoactive substance use disorder. When the diagnostic criteria for both sex addiction and another psychiatric disorder are met, both diagnoses are warranted, regardless of whether sex addiction might be secondary to the other psychiatric disorder. The diagnosis of sex addiction is a descriptive designation of how a pattern of sexual behavior relates to and affects a person’s life. It does not presume anything about how the pattern of sexual behavior developed nor is it precluded by the presence of other conditions that may have influenced its development.

Epidemiology

Two early studies reported that 3% to 6% of the American population suffer from sexual addiction, and 5% meet diagnostic criteria for sexual compulsivity. Slightly higher percentages of sexually compulsive respondents were reported by Swedish studies of individuals who used the Internet for sexual purposes. A smaller Canadian study found that 5 of 40 community subjects (12.5%) were sex addicts. A more recent and larger Canadian study reported that 50 of 114 (43.9%) incarcerated sex offenders and 21 of 117 (17.9%) socioeconomically matched community males reached the addiction criterion in the sexual addiction screening test, which was used also in the study that reported a prevalence of 3% to 6%. The authors attributed the difference to the higher socioeconomic class of the subjects in the earlier study.

About 80% of individuals who use sexual behavior addictively are male. Gender differences in prevalence are greater for paraphilic than for nonparaphilic sex addictions. Within the nonparaphilic arena, males tend to focus more on physical sexual gratification and thus more often engage addictively in masturbation, use of prostitutes, and impersonal sex. Females tend to focus more on the emotional aspects of sexuality and thus more often engage addictively in romantic relationships. Sex addiction is usually a chronic disorder with remissions and exacerbations.

Addictive use of sexual behavior typically begins in the teens, peaks between the ages of 20 and 40, and then gradually declines. The “modal subject” in a cohort that responded to an advertisement for evaluation and treatment of sex addictions and compulsive sexual behaviors was characterized as a 34-year-old, white, married, Catholic college graduate earning a middle-class income.

Comorbidity and Familial Patterns

A review of comorbidity findings is prefaced by the observation that little research has actually been conducted specifically with individuals who use sexual behavior addictively. For a variety of reasons, most of the reviewed research studied paraphilic subjects and sex offenders. While the group of sex addicts overlaps with the group of sex offenders and more so with the group of paraphilic subjects, these groups are not identical.

Researchers have noted significant comorbidity among different forms of sex addiction: individuals who engage addictively in one form of sexual behavior are likely also to engage addictively in other forms of sexual behavior. High frequencies of comorbidity have been observed not only within the category of paraphilic sex addictions but also between paraphilic and nonparaphilic sex addictions. A variety of psychiatric disorders have been reported to be comorbid with sex addiction at a rate that is higher than their prevalence in the general population. The strongest trend that the research reveals is significant comorbidity of sex addiction with substance use disorder, gambling disorder, bulimia nervosa, and compulsive buying. Significant comorbidity between sex addiction and mood disorders, anxiety disorders, attention-deficit hyperactivity disorder (ADHD), and personality disorders also has been reported. The one family history study that has been published found a significantly increased prevalence of sex addiction and other addictive disorders (substance use disorder, eating disorders, and gambling disorder) in the parents of sex addict probands. More research is needed in the areas of sex addiction comorbidity and family history, particularly in light of the usefulness of such findings in understanding etiology and in formulating treatment plans.

Course of Illness

A recent search of PubMed found no publications that reported on the course of illness for sex addiction (or for sexual addiction, sexual compulsivity, compulsive sexual behavior, or paraphilia). In DSM-5-TR, the course of illness descriptions for the paraphilic disorders mention possibilities but provide no information.

That sex addiction shares the addictive process with other addictive disorders suggests that its course of illness may be similar to those of the others. Thus far, the preceding statement is but a hypothesis in search of empirical investigation.

THE ADDICTIVE PROCESS

Other disorders in addition to sex addiction and substance use disorders fit the definition of addiction and fulfill the diagnostic criteria for addictive disorder, including gambling disorder, binge eating disorder, bulimia nervosa, and kleptomania. In addition, they share a number of clinical and phenomenologic features that are characteristic of addiction: (1) characteristic course—the disorder typically begins in adolescence or early adulthood and follows a chronic course with remissions and exacerbations; (2) behavioral features—narrowing of behavioral repertoire, continuation of the behavior despite harmful consequences; (3) individuals' subjective experience of the condition—sense of craving, preoccupation, excitement during preparatory activity, mood-altering effects of the behavior, sense of loss of control; (4) progressive development of the condition—craving, loss of control, narrowing of behavioral repertoire, and harmfulness of consequences all tending to increase as the duration of the condition increases; (5) experience of tolerance—as the behavior is repeated, its potency to produce reinforcing effects tends to diminish; (6) experience of withdrawal phenomena—psychological or physical discomfort when the behavior is discontinued; (7) tendency to "relapse"—to return to harmful patterns of behavior after a period of abstinence or control has been achieved; (8) propensity for behavioral substitution—when the behavioral symptoms of the disorder have come under control, tendency for addictive engagement in other behaviors to emerge or intensify; (9) relationship between the condition and other aspects of affected individuals' lives—neglect of other areas of life as the behavior assumes priority; and (10) recurrent themes in the ways individuals with these conditions relate to others and to themselves—including low self-esteem, self-centeredness, denial, rationalization, and conflicts over dependency and control.

That these six conditions with diverse symptomatic behaviors—sex addiction, substance use disorders, gambling disorder, binge eating disorder, bulimia nervosa, and kleptomania—meet the same set of diagnostic criteria for addictive disorder and share such an array of clinical features suggested a hypothesis that they—and, by extension, addictive disorders in general—share an underlying biopsychological process. This hypothesis has been supported by extensive research findings that demonstrated (a) significantly elevated lifetime comorbidity among these conditions and between them and other psychiatric disorders, (b) significant family history relationships, (c) similar responses to psychiatric medications, (d) similar results on psychometric assessments, and (e) similar neurobiologic correlates. Comprehensive reviews have concluded that variation in the liability to a substance use disorder correlates with personality phenotypic variations that predate the initiation of substance use. Prospective studies have reported the temporal and predictive primacy of anxiety, mood, and ADHD when one of them is found to be comorbid with a substance use disorder. Subsequent studies found that deficits in affect- and self-regulatory functioning usually precede and increase the risk for development of substance use disorders, bulimia nervosa, gambling disorder, and sex addiction. That the documented abnormalities of psychological and neurobiologic functioning are temporally primary to and predictive of the behavioral syndromes that have been under consideration here suggests that the underlying biopsychological process that these conditions share precedes their onset and is not simply a consequence of the behavior or lifestyle that characterizes them.

Definition

Addictive process is the term that designates the core biopsychological process that addictive disorders are hypothesized to share. It develops prior to the emergence of addictive patterns of behavior and constitutes a biopsychological predisposition to develop an addictive disorder. It can be defined descriptively as an enduring, inordinately strong tendency to engage in some form of pleasure-producing behavior in a pattern that is characterized by impaired control and continuation despite harmful consequences. It is not what makes cocaine, ice cream, or sex pleasurable for people in general but what makes the drive for cocaine, ice cream, or sex so much more inexorable for those who use it addictively. The class of addictive disorders includes substance use disorders, gambling disorder, binge eating disorder, bulimia nervosa, kleptomania, sex addiction, and other enduring conditions in which a behavior that can function both to produce pleasure and to reduce painful affects is employed in a pattern that is characterized by two key features: (1) recurrent failure to control the behavior and (2) continuation of the behavior despite significant harmful consequences. The group of addictive disorders is most usefully considered as not a collection of distinct disorders but an underlying process that can be expressed in one or more of various behavioral manifestations.

Thus, the development of an addictive disorder is shaped by two sets of factors: those that contribute to the underlying addictive process and those that guide the selection of a particular behavior as the one preferred for addictive use. Whereas the type of behavior that is being used addictively is the most readily noticeable manifestation of an addictive disorder, the addictive process is the component of the disorder that leads to the behavior being used in ways that are pathologic. In other words, the addictive process is the underlying disease, and an instance of addictive behavior is a symptom of the disease. Addictive behavior is behavior that the person under consideration is using or is likely to use addictively: that is, with impaired control and despite harmful consequences.

That the addictive process is a biopsychological process does not mean that it represents an interaction between biologic and psychological, nor that it has psychological properties that emerge out of relationships within its biologic system. Biologic and psychological are not distinct forms of reality but discrete sets of terms that represent distinct ways of experiencing and understanding a unitary human reality. The addictive process is a unitary biopsychological process that can be described and understood in both biologic and psychological terms. Brief neurobiologic and psychological formulations of the addictive process follow.

Neurobiologic Formulation

From the neurobiologic standpoint, the underlying vulnerability that all addictive disorders share can be understood to involve impairments in three inter-related functional systems: motivation-reward, affect regulation, and behavioral inhibition.

Impaired motivation-reward exposes addicts to unsatisfied states of restless anhedonia, irritable tension, and subjective emptiness. In the context of impaired motivation-reward function, behaviors that are associated with activation of the reward system are more strongly reinforced (via both positive and negative reinforcement) than they otherwise would have been. Motivation-reward impairment may additionally distort the relative reward values among varieties of experiences.

Impaired affect regulation renders addicts chronically vulnerable to painful affects, affective hyper-responsivity, and emotional instability. In the context of impaired affect regulation, behaviors that are associated with escape from or avoidance of painful affects are more strongly reinforced (via negative reinforcement) than they otherwise would have been. Combined impairments of motivation-reward and affect regulation particularly enhance the salience of drug self-administration and other behaviors that are associated with both (a) activation of the reward system and (b) escape from or avoidance of painful affects.

Impaired behavioral inhibition increases the likelihood that urges for some form of reinforcement in the short term (negative, positive, or both) will override the consideration of longer-term consequences, both negative and positive. The combination of impaired motivation-reward, impaired affect regulation, and impaired behavioral inhibition produces a condition in which urges to engage in behaviors that both provide pleasure or gratification and reduce painful affects are extraordinarily difficult to resist despite the harmful consequences that they might entail.

Impairments in motivation-reward, affect regulation, and behavioral inhibition result from interactions among genetic and environmental influences. They can be characterized as vulnerability factors, biologic markers, or endophenotypic dimensions of addiction. The diathesis-stress model provides an etiologic schema that guides the organization of research findings. Genetic predispositions interact with adverse experiences during critical phases of development to result in a phenotype that is neurobiologically vulnerable to the effects of stress later in life. This phenotype thus lowers an individual's threshold for developing an addictive disorder upon further stress exposure. Factors that have been found to contribute to the development of an addictive process include genetics, maternal gestational (prenatal) stress, caregiving deficiencies during infancy, and adverse childhood events or conditions.

Genetics. Researchers now generally believe that the predisposition to develop an addictive disorder results from the interaction of multiple genes. A large number of genes contribute to the risk for substance use disorder, but no single gene displays such a large magnitude of effect that it accounts for a major fraction of the genetic influence. The same holds for gambling disorder. Meanwhile, most of the genetic liability to developing addictive use of any one psychoactive substance is shared among the substances, as distinct from liability that is specific to a particular substance or class of substances. Along similar lines, the genetic risk for alcohol use disorder accounts for much of the risk for gambling disorder, and genes that increase the risk of developing gambling disorder increase the risk of developing other impulsive-compulsive and addictive patterns of behavior. Many addiction-related genetic variants (polymorphisms) do not fit into categories that are defined by substances or behaviors and are more usefully organized into categories that are defined by the brain functions that they alter. One such set of categories is motivation-reward, affect regulation, and behavioral inhibition.

MOTIVATION-REWARD. The homozygous *11* genotype of the dopamine (DA) D_1 receptor is associated with increased risk of alcohol use, cigarette smoking, use of other drugs, gambling, compulsive shopping, and compulsive eating. The Taq *A1* allele of the D_2 receptor gene predicts alcoholism, cigarette smoking, addictive use of psychoactive substances, pathologic gambling, and exaggerated reward value of food. The Taq *A1* allele is associated with reduced D_2 receptor density, hypersensitive presynaptic D_2 receptors, and decreased general responsiveness of the reward system to rewarding stimuli, along with heightened responsiveness after events that increase levels of DA in the synapse between the ventral tegmental area and the ventral striatum. Such an increased reward effect against a baseline of diminished reward could promote the development of an addictive disorder by intensifying a person's motivation to repeat behaviors that increase intrasynaptic DA in the midbrain motivation-reward system. Behaviors that do so include self-administration of a hedonic substance, eating (especially sweets), gambling, and engaging in sexual behavior. The Val(158) allele of the COMT gene that codes for catechol-*O*-methyl transferase (an enzyme that inactivates DA and norepinephrine [NE]) is associated with alcoholism, methamphetamine abuse, heroin addiction, and polysubstance abuse.

Variants of the gene that codes for the κ-opioid ligand prodynorphin occur more frequently in alcoholics, cocaine abusers, and methamphetamine addicts than in controls. Polymorphisms of the gene for the cannabinoid CB_1 receptor are associated with increased risk for addiction to alcohol, cocaine, amphetamines, cannabis, and multiple substances.

AFFECT REGULATION. The gene that encodes the serotonin (5-HT) transporter protein has a variant site (*5-HTTLPR,* for 5-HT transporter-linked polymorphic region) that is characterized by 2 alleles, 1 short (*s*) and the other long (*l*). The *s* allele, and especially the homozygous *s*/*s* genotype, is associated with alcoholism, heroin addiction, cigarette smoking, pathologic gambling, bulimia, and binge eating disorder. It is associated also with affective instability, anxiety-related personality traits, and heightened sensitivity to mild stressors.

The *A* allele of the gene that encodes brain-derived neurotrophic factor (BDNF) is associated with alcoholism, methamphetamine abuse, heroin abuse, tobacco smoking, and all subtypes of eating disorders. It is specifically associated with binge eating behavior, a phenotypic trait that occurs in both bulimia and binge-eating disorder. The long allele of another BDNF gene variant, in the gene's promoter region, is associated with vulnerability to polysubstance use disorder.

Alcoholism is associated with three polymorphisms of the gene that codes for neuropeptide Y (NPY) and with two haplotypes of the galanin gene. The mediating link in both cases is thought to be impaired regulation of anxiety.

BEHAVIORAL INHIBITION. Whereas DA D_1, D_2, and D_3 receptors play central roles in the motivation-reward system, D_4 and D_5 receptors are involved primarily in behavioral inhibition and attentional processing. The long allele of a polymorphism in the D_4 receptor gene is associated with impulsive personality traits and is a risk factor for adolescent alcohol abuse, adolescent "hard drug" use, heroin abuse, cue-elicited heroin craving, greater severity of alcohol and opiate addiction, pathologic gambling, binge eating, and cue-elicited craving for food.

The common 148 bp variant of the D_5 receptor gene is correlated with substance abuse and also with ADHD and novelty seeking. By itself, the 10-repeat allele of the DA transporter (DAT) gene *DAT1* is not associated with increased consumption of alcohol. However, the correlation between novelty seeking and increased alcohol consumption is significantly greater with this allele than with the normal genotype.

Studies have reported associations between variants of the gene *GABRA2*, which encodes the α-2 subunit of the $GABA_A$ receptor, and both alcoholism and polysubstance use disorder. Risk alleles of this gene are associated also with conduct disorder in children and with antisocial personality disorder in adults. Polymorphisms of the gene that encodes the α-3 subunit of the $GABA_A$ receptor are associated with alcoholism.

Although impaired behavioral inhibition or impulsivity is a key feature of addiction, research has linked variants of genes that code for 5-HT_{1B} and 5-HT_{2A} receptors with greater impulsivity among alcoholics and bulimics, respectively, but not with increased risk for any addictive disorder. In at least some instances, genotypic variations seem to map onto disorder-relevant intermediate phenotypes (i.e., trait-level variations) within a population that is affected by a particular disorder more closely than they map onto gross phenotypes, such as the disorders themselves.

Maternal Gestational (prenatal) Stress. Research with rodents and monkeys has indicated that maternal gestational stress (MGS) amplifies the subsequent reinforcing effects of ethanol, psychostimulants, and morphine on the offspring. It accelerates sensitization to amphetamine, increases motor-responsiveness to cocaine, and potentiates cocaine-primed reinstatement of cocaine-seeking behavior. In addition, it intensifies physiologic-emotional responses to postnatal stress, retards the learning of both active and passive means of managing stress, and increases impulsivity. Prospective studies with human subjects have found MGS to be associated with high affective/behavioral reactivity, negative affect, difficulty being consoled, and disinhibitory traits in the offspring. The consensus among investigators is that exposure to MGS results in a general susceptibility to psychopathology rather than increased risk for any particular form of psychopathology.

Caregiving Deficiencies During Infancy. Considerable development of the human central nervous system occurs after birth, particularly during the first 2 years of life. Postnatal brain development is highly responsive to conditions in the environment. Epidemiologic, clinical, and preclinical data indicate that early exposure to adverse environments engenders a vulnerability to distortion of the physiologic stress response and subsequent impairments in affect regulation, motivation-reward, and behavioral inhibition. With respect to postnatal brain development in mammals, the most important constituent of the environment is the mother (or mother-substitute). Research has investigated the potential consequences in adulthood of two mother-related variables that may occur during infancy: maternal separation (MS) or deprivation and deficiencies in the quality of maternal care.

MATERNAL SEPARATION. Compared to normally reared rats, adult rats that were exposed to MS during infancy are characterized by increased ethanol and sucrose consumption; amplified sensitivity to the neurochemical and behavioral effects of psychostimulants and opiates; increased susceptibility to the rewarding properties of psychostimulants, opiates, and food; enhanced acquisition and maintenance of operant responding for drug or food reinforcer; elevated risk of developing addiction-like patterns of consuming or self-administering ethanol, cocaine, morphine, and sucrose solution; and greater tendency to overeat and gain weight when given free access to food. On a fasting/refeeding schedule, compensatory hyperphagia during refeeding days diminishes gradually in rats that were raised normally but continues in MS rats. A prospective study with human subjects found that compared to offspring who had been breastfed for more than 1 month, offspring who had been weaned at age 1 month or earlier had a 1.47-fold greater risk of being hospitalized in a psychiatric unit with an ethanol-related diagnosis.

MS during infancy is associated also with other behavioral abnormalities. Most common are behaviors that suggest impaired affect regulation: increased anxiety-like behaviors in response to aversive environmental stimuli, intensified behavioral and visceral reactivity to stress, increased startle response, increased fearfulness in the face of novelty, increased depression-like behavior, decreased social interaction, reduced coping behavior in stressful environments, and increased vulnerability to developing anxiety-like or depression-like behavior when subsequently exposed to chronic stress. The amount of ethanol consumed by MS animals is highly correlated with the magnitude of the glucocorticoid (GC) response to a mild stressor and with measures of anxiety-like and fearful behaviors in both rats and monkeys.

Behaviors that are consistent with impaired motivation-reward function have been reported in MS research with rats and with monkeys. These include reduced social motivation, blunted perception of differences in reward value, and reduced operant responses on progressive ratio reinforcement schedules, which suggest either low motivation or diminished sensitivity to reward. MS rats' low motivation and diminished sensitivity to reward have been normalized by administration of fluoxetine.

Other studies of MS animals found increased impulsive behavior, locomotor hyperactivity, reduced prepulse inhibition of the startle response, deficits in shifting the attentional set, lower-risk assessment and higher-risk taking, and impaired retention of inhibitory avoidance. Children who were adopted from deprived institutional settings before age 43 months were found to display high degrees of inattention and overactivity at 6 and 11 years of age. They were found also to demonstrate deficits in executive functions of planning, inhibition, set-shifting, working memory, generativity, and action monitoring, which were most severe in children who had experienced more than 6 months of institutional maternal deprivation.

MATERNAL CARE DEFICIENCIES. A group of studies has explored the relationship between naturally occurring variations in maternal care of infant rats and the latter's capacity for affect regulation as they develop into adulthood. Maternal care is represented by mother rats' licking or grooming and arched-back nursing (LG-ABN) of their newborn progeny—simply licking and grooming (LG) in some recent studies. These studies have provided a credible experimental model for the relationship between the maternal care of human infants and their ensuing development.

Differences in the frequency of these maternal care behaviors have been found to be predictive of individual differences in affect regulation in the offspring when they become adults. Compared to the offspring of mother rats that show high levels of LG-ABN, the offspring of low–LG-ABN dams exhibit heightened anxiety-like behavior, exaggerated responses to stress, and less time spent in social interaction. Female offspring of low–LG-ABN mothers exhibit earlier puberty, increased sexual receptivity, and greater vulnerability to stress-induced binge-eating than do their high–LG-ABN counterparts.

To test the hypothesis that the observed characteristics of low–LG-ABN mothers' offspring had resulted from lower levels of maternal care, as distinct from genetic factors that also produce low–LG-ABN behavior, rat pups were "cross-fostered"—raised from birth by mothers whose LG-ABN frequency was opposite from that of the mothers to whom they were born. Pups that were born to high–LG-ABN dams and raised by low–LG-ABN mothers become hyperemotional, high-strung adult rats. Meanwhile, pups that were born to low–LG-ABN dams and raised from birth by high–LG-ABN mothers become mellow, well-adjusted adult rats. These findings support the hypothesis that the lower levels of maternal care that these rats received during their early postnatal lives was a critical determinant of the observed differences in adult rat functioning.

Variations in maternal licking/grooming influence the development of stress-responsiveness primarily through central processes that stimulate—via amygdalar corticotropin-releasing hormone (CRH)—or inhibit—via gamma-aminobutyric acid (GABA)—the noradrenergic systems that activate endocrine and behavioral responses to stress. Relative to offspring of high–LG-ABN mothers, adult offspring of low–LG-ABN mothers exhibit increased CRH expression in the central amygdala (CeA), increased CRH receptor density in the locus ceruleus (LC, primary site of NE synthesis), decreased density of hippocampal GC receptors (and hence diminished GC feedback sensitivity), decreased expression of glutamic acid decarboxylase (the rate-limiting enzyme in GABA synthesis) in the medial prefrontal cortex (PFC), reduced GABA affinity of benzodiazepine (BZD)-$GABA_A$ receptor complexes in the LC and nucleus tractus solitarii, and decreased BZD-$GABA_A$ receptor complex densities in the basolateral amygdala, CeA, and LC. Consequently, the offspring of low-LG-ABN mothers display higher stress–induced increases in paraventricular nucleus levels of NE than do the offspring of high–LG-ABN mothers. Noradrenergic input to the paraventricular nucleus provides the primary source of activation for synthesis and release of the CRH that initiates that hypothalamic–pituitary–adrenal (HPA) axis cascade. Compared to adult offspring of high–LG-ABN mothers, adult offspring of low–LG-ABN mothers also show decreased mGluR1 mRNA and protein in the hippocampus, decreased type I metabotropic glutamate receptor (mGluR1)-induced hippocampal long-term depression, and increased paired-pulse depression (PPD). PPD is an inhibitory feedback process that prevents excessive glutamate release during high-frequency stimulation.

At the heart of these differences in neurotransmitter/neuromodulator and receptor functioning are differences in gene expression. Maternal care of infants regulates the expression of the hippocampal glucocorticoid receptor (GR) gene by influencing two processes: acetylation of histones H3-K9 and methylation of the NGFI-A consensus sequence on the exon 1(7) promoter. Adult offspring of low–LG-ABN mothers were found to have lower levels of DNA methylation than do adult offspring of high–LG-ABN mothers. These differences emerged over the first week of life, were reversed with cross-fostering, persisted into adulthood, and were associated with altered histone acetylation and transcription factor (NGFI-A) binding to the GR promoter.

Central infusion of either a histone deacetylase inhibitor or a methyl donor removed the group differences in NGFI-A binding, GR expression, and HPA and behavioral responses to stress. These findings suggest a causal relation among epigenomic state (the sum of all epigenetic changes [DNA methylations and histone modifications] to an organism’s genome), GR expression, and the effect of maternal-infant care on subsequent stress responses in the offspring. Moreover, they demonstrate that an epigenomic state of a gene can be established through behavioral programming and is potentially reversible.

The demonstration by this body of studies that naturally occurring differences in maternal care of infants are correlated with subsequent differences in the behavior and neurobiology of adults constitutes an important evidence-based bridge between neuroscience and psychodynamic/psychoanalytic understanding.

Adverse Childhood Events or Conditions. Children who have been sexually or physically abused manifest abnormal baseline and stressor-responsive cortisol levels, either abnormally elevated or abnormally flat. Childhood abuse may sensitize the stress response system, thus rendering persons who were abused during childhood particularly vulnerable to stress and increasing their risk for stress-related disorders. This vulnerability may result in hypersecretion of CRH when these individuals are stressed.

Upon exposure to a standardized (laboratory) psychosocial stressor, women with a history of childhood sexual or physical abuse demonstrated markedly increased pituitary–adrenal and autonomic responses. The mean adrenocorticotropic hormone response was highest in women with a history of childhood abuse and current major depression, who also showed significantly increased cortisol responses to psychosocial stress. These findings suggest that HPA axis hyper-reactivity, probably due to hypersecretion of CRH, is a persistent consequence of childhood abuse that may contribute to a diathesis for later development of addiction and other psychiatric disorders.

A study screened 120 healthy college students for quality of parental care during childhood. Five from the top end of the distribution and five from the bottom end were invited to participate in a PET study during which they were asked to complete a stressful psychosocial task. The scans indicated that the psychosocial stressor elicited a significant release of DA in the ventral striatum of subjects in the low but not the high parental care group. The low parental care group also showed higher baseline cortisol levels and higher increases in cortisol during the stressful task. The magnitude of the cortisol response to stress was highly correlated with the magnitude of the ventral striatum DA response. These findings suggest that the chronic stress of neglectful or abusive parental care during childhood results in an HPA axis that is hypersensitive to psychosocial stressors and a midbrain that is hyper-responsive to triggers of DA release—conditions that are highly conducive to the development of addictive disorders.

In preclinical studies, reinstatement of drug-seeking behavior after extinction of its association with drug delivery can be elicited by brief exposure to a stressor. In human subjects, exposure to stressors and presentation of stress-related imagery have been identified as potent triggers for drug-seeking and recurrence of drug use. Most studies on drug self-administration have reported increases in responding after repeated or prolonged exposure to stress levels of GCs. GCs are understood to enhance drug responding by selectively facilitating DA transmission in the nucleus accumbens shell. During chronic stress, repeated increases in GCs and DA result in sensitization of the reward system. This sensitized state, which can persist after the end of the stress response, renders subjects more responsive to pleasurable substances or behaviors that are associated with release of mesolimbic DA and consequently more vulnerable to the development of addiction.

Thus, addicts' vulnerability to stress-induced reinstatement of drug-seeking and thence to drug use depends on the sensitivity of their HPA axis and the responsiveness of their mesolimbic dopaminergic system, in addition to the frequency and intensity of stressor exposures. Both elevated baseline cortisol levels and higher increases of cortisol in response to stressors are common sequelae of adverse conditions or recurrent adverse events during childhood.

Stress and disorders that are related to chronic stress tend to increase dendritic atrophy, accelerate neuronal degeneration, and subvert neuronal regeneration in the hippocampus and PFC. Mediating factors may include chronically elevated levels of GC, decreased expression of BDNF, stunted sprouting of serotonergic axons from insufficient availability of BDNF, glial cell loss, and decreased arborization and density of noradrenergic axons. In the developing brain, elevated levels of catecholamines and cortisol may lead to structural deviations or deficits through accelerated loss of neurons, delays in myelination, abnormalities in developmentally appropriate synaptic pruning, and inhibition of neurogenesis. Sustained GC exposure truncates dendrites and impairs neurogenesis in the hippocampus and PFC. Reduced baseline and hyporesponsive cortisol levels also can cause neuronal damage.

Together, these processes feed the downward spiral by potentiating hippocampal and cortical atrophy. Neuronal degeneration in the hippocampus and PFC diminishes the capacity of these regions to modulate or inhibit amygdalar stress/fear pathways and undermines hippocampal inhibitory control over cortisol release by the HPA system. Evidence of neuronal loss in the anterior cingulate cortex (ACC) has been reported in children, adolescents, and nonhuman primates with histories of adverse experiences early in life. In addition, adult survivors of early abuse have been found to have changes in hippocampal structure and function. Thus, these stress-induced processes may lead to compromised executive function, impaired affect regulation, and a greater incidence of impulsive behaviors.

Psychological Formulation

The central component of the addictive process as psychologically formulated is impaired self-regulation. Associated character pathology emerges from the same genetic-environmental matrix in which the self-regulatory impairments originate and develop in response to the interactive effects of impaired self-regulation and adverse personal history.

Impaired Self-Regulation. The descriptive definition of the addictive process that was presented earlier identifies objectively confirmable features that observers of the person in question might similarly apprehend. The addictive process also has a subjective definition; that is, what it feels like from the inside. Defined subjectively, the addictive process is an enduring tendency to feel driven to use some form of pleasure-producing behavior as a consciously or unconsciously intended means of relieving or avoiding painful affects and regulating one’s sense of self. A driven need to relieve painful affects and to regulate one’s sense of self suggests that addicts’ inherent systems for regulating affects and sense of self are impaired. Dependence on externally directed behavior to relieve painful affects and regulate one’s sense of self suggests a relative incapacity to utilize other, more adaptive means of managing. The subjective definition of the addictive process suggests that impaired self-regulation is at the nucleus of the addictive process.

Self-regulation is a comprehensive term for the biopsychological processes that regulate one's subjective states and behavior. The subjective states that are most affected when self-regulation is impaired are affects and sense of self. Volition, sensory experience, and thought processes are often also affected but less dramatically. Impaired internal regulation of people's subjective states leads them to depend on externally directed actions to regulate their affects and sense of self. Under such circumstances, intense and potentially traumatic affects may be experienced not as emotions but as urges to act. Impaired behavioral regulation limits people's capacity to resist urges to engage in actions that modify their affects or sense of self and that seem to promise, however illusorily, to protect them from emotional trauma.

The self-regulation system is a set of processes that are interrelated both functionally and developmentally. It can be understood to consist of three groups of functions: affect regulation, self-care, and self-governance.

Affect regulation functions are the psychobiologic processes that regulate the subjective, physiologic, behavioral, and interpersonal aspects of one's emotional states. These functions include self-soothing, self-enlivening, keeping affects' intensity within a manageable range, maintaining a balance when affects conflict, and tolerating affects without being overwhelmed.

Self-care functions concern the organization and regulation of one's perceptual, cognitive, executive, and behavioral processes in the service of self-protection and self-nurturing. Self-protection involves recognizing, assessing, warning oneself of, and defending oneself against dangers. Self-nurturing involves recognizing, assessing, setting priorities among, and taking care of one's needs. A critical component of self-care is the ability to use affects as signals or guides for appropriate action.

Self-governance functions include values, standards, self-esteem, self-punishment, and maintenance of a stable and cohesive sense of self. Self-governance thus subsumes (1) what psychoanalytic theory identifies as the superego—in its encouraging, approving, and rewarding aspects as well as its prohibiting, condemning, and punishing aspects; (2) the ego ideal or self-ideal, a fantasied representation of oneself that embodies all one's ideals; and (3) the processes that are involved in regulating one's sense of self.

The core self-regulation system develops during the first years of life out of the reciprocal interactions between the expression of infants' genes and infants' relationships with their primary caregivers. Considerable development of the human brain occurs after birth, particularly during the first 2 years of life, and its postnatal development is highly responsive to conditions in the caregiving environment. In healthy development, "good enough" genetic endowment interacts with caregivers' more or less adequate provision of functions that regulate the infants' somato-affective inner world. As their physiologies and prepsychological anlagen are challenged but not overwhelmed, infants gradually become able to provide for themselves the regulatory functions for the provision of which they had been depending on their caregivers. The metaphorical shorthand for this developmental process is *internalization*. However, various combinations of genetic and environmental factors can interfere with internalization and disrupt the development of the self-regulation system. Later in life, adaptive self-regulatory functioning, particularly in the areas of self-care and self-governance, depends on effective management of one's inner conflicts—which is predicated on awareness of them. When awareness of inner conflicts is blocked, their operation in the dark can distort the functioning of the self-regulation system—as can the processes that block such awareness.

To the extent that individuals' affect regulation is impaired, their emotional responses are easily provoked, unusually intense, and slow to recede. Their affect states tend to be unstable, intense, and disorganizing. Especially threatening are abandonment-separation anxiety, claustrophobic anxiety (anxiety associated with the general idea of being closed in, trapped, or smothered), depressive affects, the “superego affects” of guilt and shame, and affects that are associated with narcissistic injury. But almost any intense affect, including excitement or love, can be experienced as threatening when affect regulation is impaired.

To the extent that individuals’ self-care is impaired, they are deficient in their abilities to avoid danger, protect themselves, and take care of their needs. The dangers to which they expose themselves include physical, emotional, and narcissistic dangers. (“Narcissistic” means related to self-esteem and sense of self.) Similarly, the needs to which they do not adequately attend include physical, emotional, and narcissistic needs. As a result, they are particularly susceptible to getting hurt and being chronically dissatisfied on all three levels.

To the extent that individuals’ self-governance is impaired, they lack a consistent sense of self, a stable self-esteem, and a clear sense of direction. They may have skewed or fragmented values, ideals, and sense of meaning. Any threat to their sense of being in control can potentially trigger intense anxiety about being overwhelmed by helplessness. Their behavior tends to oscillate between extremes, and they may swing from actions that violate their values to actions that are self-injurious or invite punishment from the environment.

When individuals’ self-regulation is impaired, they live in a world of primitive pain, terrifying anxieties, and an unstable sense of self. They tend to respond with overwhelmingly intense affects to events that most others seem to take in stride. They experience the risk of being thus affectively overwhelmed and losing their self-coherence as threats to their survival, so they do what they can to avoid affective states that are potentially traumatic. Many seek desperately for something that can perform for them the self-regulatory functions that people with intact self-regulation systems are able to do autonomously. Often, they develop fantasies (usually unconscious) that the solution for their self-regulatory problems lies in something outside their selves: a substance, a person, a nonhuman object, a bodily state, or an institution. Many of them learn to ward off traumatic affects and self-states by engaging in a rewarding activity such as eating, taking a mood-altering substance into their bodies, gambling, or engaging in some form of sexual behavior. They then tend to become dependent on this activity to provide sham versions of self-regulatory functions that they have not been sufficiently able to provide for themselves.

When external or internal cues that are associated with intense affects or with loss of self-coherence become conditioned to particular craving sensations or discriminative stimuli for particular addictive-behavioral responses, these cues can trigger or be experienced as urges to engage in symptomatic behavior. Since these behaviors are addicts' desperate attempts to ward off traumatic affects and self-states that threaten their survival, their urges to enact the behavior typically are so imperative that their patterns of engaging in it are likely to be characterized by impaired control and by relative imperviousness to the restraining influence of harmful consequences.

Associated Character Pathology. *Character pathology* is an inclusive term that refers to both (a) enduring, pervasive, relatively inflexible patterns of inner experience and behavior that begin in adolescence or early adulthood and lead to significant functional impairment or subjective distress, and (b) the underlying psychobiologic processes of which the inner experience and behavior are observable manifestations (inner experience being observable to the subject). Its inclusion of (b) distinguishes it from the more widely used term *personality disorder*.

The character pathology that tends to be associated with addictive disorders originates under the same genetic and environmental circumstances as do the just discussed self-regulatory impairments. Its development is then shaped by infants' and children's needs to protect themselves from the overwhelming affects and loss of sense of self to which their personal history and impaired self-regulation expose them. Accompanying character pathology that tends to be associated with addiction (each to varying degrees in different individuals) includes: characteristic self-protective processes, impaired ability to symbolize affects, impaired management of the basic conflict, and reliance on the primal fantasy.

CHARACTERISTIC SELF-PROTECTIVE PROCESSES. The self-protective processes (often referred to as defenses) that people who engage addictively in one or more forms of behavior most typically employ are denial or disavowal, externalization, and the processes that constitute the narcissistic character/personality system.

Denial or disavowal is an unconsciously motivated failure either to recognize the reality of what is perceived or to appreciate the personal significance or implications of what is perceived. Its function is to support dissociation, an unconsciously intended removal from the realm of awareness something that if consciously experienced could precipitate overwhelming emotional pain or self-fragmentation. In addictive disorders, denial or disavowal most often occurs to support dissociation of: (1) painful affects, particularly guilt, shame, and affects that evoke guilt or shame; (2) one side or the other of an inner conflict, particularly the basic conflict and its derivatives; (3) narcissistic vulnerability and limitations; and (4) impulses that have been associated with severe trauma.

Externalization is an unconscious process whereby an aspect of one's self is experienced as an aspect not of oneself but of something external to oneself. It is the self-protective process that is most characteristic of addiction. The basis for people's transition from simply engaging in a rewarding activity to engaging in that activity addictively is their externalization of their self-regulatory functions to that rewarding activity or to the object of that activity. Their mental representations of their capacity to regulate their affects and sense of self shifts from their mental representations of their selves to their mental representations of that activity or object. In the course of this shift, their mental representations of their self-regulatory functioning are infiltrated by fantasies of omnipotence and become more effective in people's mental representations of the rewarding activities or objects than they ever could have been in real life.

The scope of externalization in addicts' lives extends beyond their self-regulatory functions. For example, to support the denial of an inner conflict, addicts (or anyone else) can assign one side of the conflict to an external agent and thereby transform the inner conflict into an external conflict. Or, in an attempt to take magical omnipotent control over their inner lives, they may externalize a passively experienced, traumatic inner state to a tangible external problem that they believe can be resolved through effort and action. A consistent feature of externalization is its direction of attention and energy toward action in the external world.

The DSM-5-TR diagnostic criteria for narcissistic personality disorder (lightly abridged) are: (1) grandiose sense of self-importance; (2) preoccupation with fantasies of unlimited success, power, brilliance, beauty, or ideal love; (3) belief that one is unique and special; (4) abnormal need for admiration; (5) sense of entitlement; (6) interpersonal exploitativeness; (7) lack of empathy; (8) envy of others or belief that one is envied by others; and (9) arrogant behaviors or attitudes. The use in common parlance of *narcissistic* and *narcissist* as terms of disparagement reflects the magnitude of people's antipathy toward narcissistic behaviors and attitudes (or perhaps of their motivation to disavow their own narcissistic problems). It also reflects people's failure to appreciate that these behaviors and attitudes arise as desperate measures to avoid annihilation and persist because they are unconsciously believed to be required as protection against that same danger.

The narcissistic character system is a variable set of self-protective processes that function to preserve one's sense of self from annihilation or fragmentation and to sustain an artificial self-esteem that supports the dissociation of overwhelming shame, humiliation, vulnerability, sense of inadequacy, and helplessness. Grandiosity is the most overt aspect of the artificial self-esteem. It supports dissociation of the aforementioned affective experiences by trumpeting their opposite. Deficient empathy (ability to sense the emotions that another person is consciously or unconsciously experiencing) can result from genetic factors or from isolation in early infancy. However, it most commonly originates early in life as a self-protective deadening of one's empathic sense under circumstances in which experiencing a primary caregiver's emotions entails a threat of annihilation (e.g., by becoming affectively overwhelmed, or by losing a sense of self-other boundary and subjectively ceasing to exist as a distinct self). One's capacity for empathy can be further undermined by impaired ability to recognize or name affects.

A sense of entitlement can originate in grandiose fantasies. More often, it is based on reality. During the first years of life, most if not all persons who later develop narcissistic character pathology did not receive enough of the caring, nurturing, or soothing that they needed for healthy development—and to which they as mammals were biopsychologically entitled by virtue of having been born. In the context of that history, lingering feelings of need, emptiness, deprivation, and entitlement are understandable. Grandiosity may then seep into and swell the entitlement to drive away feelings of need, vulnerability, and helplessness. In general, the problem with entitlement is not entitlement as such, but people's unconscious fantasies that they will receive restitution from people in their current lives for what they were entitled to but did not get in their early childhoods. The main function of this fantasy, like that of the primal fantasy from which it is an offshoot, is to support the dissociation of traumatic early experiences of privation or abandonment.

Individuals whose functioning is structured by the narcissistic character system are unconsciously terrified by the ongoing threats to their psychological survival. They desperately seek and try to control potential sources of narcissistic supplies—whatever might help preserve their beleaguered sense of self or sustain their prosthetic self-esteem. Since they unconsciously experience their survival to be at stake, they tend to relate to other persons primarily in terms of the latter's capacity to provide or interfere with access to such supplies. Addictive behavior patterns can be understood as a natural expression of the characteristic modes of adaptation that individuals with narcissistic character pathology employ to deal with their underlying vulnerabilities and impairments.

In addictive disorders, denial, externalization, and narcissistic self-protective processes support and promote one another. In effect, they operate together as one system of character or personality.

IMPAIRED ABILITY TO SYMBOLIZE AFFECTS. People who struggle with addictive disorders share with psychosomatic disorder patients and post-traumatic stress disorder (PTSD) patients an impaired ability to symbolize affects. They have difficulty recognizing, naming, and verbalizing their affects, which they typically experience as physical bodily states rather than meaningful emotions. These disturbances often are accompanied by a lowered capacity for empathy and diminished emotional involvement in interpersonal relationships. The ability to recognize and identify emotions is a critical component of affect tolerance, which accordingly is diminished when the ability to symbolize affects is impaired. Affects that arise under these circumstances consequently are more often experienced as potentially overwhelming, as incipient threats to the stability of the self. The experience of affects as physical bodily states moreover disposes people to address them by physical means, by action—and particularly by action that directly affects the body.

Episodes of addictive behavior are usually precipitated by painful affects that addicts at some level perceive as threatening to overwhelm them. However, they often are consciously aware of neither the sense of being threatened nor even the affects themselves. Preceding an episode of addictive behavior, they typically experience a nonspecific sense of dysphoria or discomfort—a vague restlessness, uneasiness, and unpleasant sense of tension. Some addicts do not experience the emotional discomfort at all but are aware only of an urge or a craving to engage in symptomatic behavior.

An impaired ability to symbolize affects can result from one or more of the following factors: (1) a primary developmental deficit; (2) regression to a more primitive kind of symbolic and/or affective functioning (regression is reversion to a developmentally earlier mode of being that usually occurs as an unconsciously intended means of self-protection from traumatic emotional pain); (3) a broad-spectrum self-protective process that is directed against affects or symbolic processes in general; and (4) a “walling off” of affect regulation functions, unconsciously reserving them for one’s caregiver or, more accurately, for the caregiver as represented in one’s mind.

THE BASIC CONFLICT. The basic conflict is between two fundamental sets of drives and fears that emerge during the first 2 years of life. On one side is a drive for attachment, connection, or merger with the caregiver, which is combined with fear of being abandoned by or isolated from the caregiver. On the other is a drive for separation, autonomy, and mastery, which is combined with fear of being engulfed, swallowed up, or controlled by the caregiver. In short, the basic conflict can be understood as being between a basic pull toward attachment-union with another person and a basic push toward being a separate, autonomous self.

The importance of the basic conflict derives first of all from its ubiquity. It is not a form of psychopathology but a part of the human condition that pulsates in the background throughout people’s lives. How it manifests depends on the life context and on where a person is developmentally. Examples of forms it can take include conflicts between wishes to be taken care of and fears of being controlled, and between wishes to be competent and fears of being abandoned. In addition to being ubiquitous, the basic conflict is perpetual and unavoidably painful. Both connections with other persons and existence as a separate self are necessary for an individual’s psychological survival, and a move toward one side can evoke painful feelings of threat or loss related to the other. When the basic conflict is not effectively managed, it intensifies any inner conflict that can be mapped onto it. (Management of a conflict does not mean that it has been solved and settled but that relatively stable dynamic equilibrium is being maintained.)

A number of factors can make the basic conflict difficult to manage. Most often, people have difficulty resolving the basic conflict when they unconsciously believe that it is associated with more anxiety than they can tolerate. Perhaps their primary caregivers gave them messages that their connections with their caregivers were contingent on their not being separate persons and that being separate would result in being abandoned. Or their history with their primary caregivers led them to expect that connecting with another person would be too likely to lead to being devastatingly hurt, humiliated, or losing their identity. Or they unconsciously refuse to become separate and masterful because to do so would imply recognition that their unmet early childhood needs would never be met, which would expose them to traumatic depressive pain. One or more of these patterns characterize the early history of most addicts. A more general factor is impaired self-regulation, which can intensify anxiety or emotional pain to the point of feeling as though it were threatening to overwhelm and destroy the subject. The final common pathway is that affected persons unconsciously believe that awareness of the basic conflict would be accompanied by overwhelming, shattering anxiety. Consequently, they avoid that awareness by keeping one side of the conflict out of conscious view. This process of dissociation or splitting is often sustained by unconsciously assigning (i.e., externalizing) the dissociated side of the conflict to another person. The other person is then experienced as though they were pushing for that side of the conflict.

For example, a fictional patient Blake unconsciously wants his psychotherapist to take care of him totally but also fears being taken over and losing his sense of self. Assigning the former side of the conflict to the therapist could lead Blake to experience his therapist as trying to seduce or trick him into becoming dependent while Blake experiences himself as trying to hold onto his autonomy. Conversely, assigning the latter side to the therapist could lead Blake to experience himself as trying to develop a closer relationship with his therapist while the therapist is holding back or pushing him away.

THE PRIMAL FANTASY. The primal fantasy is an unconscious fantasy that can develop during the first 2 years of life when some combination of deficient caregiver responsiveness and a child's genetically or epigenetically based impairments interfere with the fulfillment of the child's needs to be cared for, held, nurtured, and soothed. The fantasy is that those fundamental needs that were not met during early childhood will someday be met. While it is defined by anticipated gratification, its power derives primarily from its capacity to sustain the dissociation of overwhelmingly painful early experiences of privation or abandonment and thus to provide protection from traumatic pain. The belief that unmet needs of early childhood might yet be met enables avoidance of the realization that the opportunity to have those needs met is, along with early childhood, lost and gone forever.

A common form of this fantasy begins with children's implicit belief that the reason their needs were not met was that they were bad or not good enough and thus did not deserve to have their needs met. Their minds gravitate toward this explanation in preference to others that locate the cause of their privation in the primary caregiver or an unpredictable universe. Believing that they themselves are to blame for their needs not being met is painful but believing that they are helpless and can do nothing to influence this condition is even more so. Self-blame sustains illusions of having control. The fantasy then takes over with the belief that if they were to be good enough or do well enough, they would get their needs met. In addition to fostering a sense of efficacy, this belief inoculates the primal fantasy against empirical test. Individuals' failure to get their needs met after being good or doing well does not disprove the fantasy. It just indicates that they were not good enough or did not do well enough.

When people rely on the primal fantasy to protect themselves from the pain of early privation or abandonment, they have a hard time separating from their parents and becoming authentic, self-responsible persons. The fantasy is about meeting the needs of early childhood, so the subject of the fantasy must in significant ways remain in early childhood. As long as people do not grow up, they can maintain the fantasy that the unmet needs of their early childhood still can be met. Growing up, becoming separate and authentic persons, means letting go of the fantasy. The consequent annihilation anxiety leads some individuals to avoid separating from their parents, taking care of themselves, or assuming adult responsibilities.

People typically externalize the primal fantasy onto another person. They at some level believe that important needs that were not met during their early childhood will be met by this person. During their subsequent childhood, this fantasy can increase their vulnerability to being sexually abused. Later in life, it can lead them repeatedly into dysfunctional relationships. Sexualization of the fantasy can turn them toward addictive involvement in romantic relationships or other kinds of sexual behavior. The psychodynamics of addictive disorders in general usually involve the primal fantasy. In such cases, people unconsciously perceive the drug or food, or other objects of the addictive behavior as the fantasy object that will relieve their pain, fill their emptiness, make them whole, and meet in distorted form the unmet needs of their early childhoods.

Development of the Addictive Process. The development of the addictive process in psychological terms can be condensed to: (1) due to some combination of genetic and environmental factors, individuals' self-regulatory functions are impaired; (2) impaired self-regulation leaves them abnormally vulnerable to being overwhelmed by intense emotional states and loss of self-coherence; (3) in attempts to protect themselves from the danger of being thus overwhelmed, they externalize fantasy-tinged mental representations of their self-regulatory functions onto nonself activities or objects over which they believe that they have control. When the behavior that results occurs in a pattern that is characterized by recurrent loss of control and continuation of the behavior despite significant harmful consequences, they can be characterized as “addicts,” whatever the behaviors in which they addictively engage.

SELECTION OF SEXUAL BEHAVIOR

The addictive process represents a predisposition to engage in one or more behaviors that can provide both pleasure and relief in a pattern that is characterized by recurrent failure to control and continuation despite harmful consequences. Given a sufficient degree of predisposition, what determines the specific behavior(s) in which a person addictively engages? More particularly, what might lead a person to choose (unconsciously) a form of sexual behavior rather than, for example, using cocaine or eating?

The development of a particular addictive disorder seems to have three components: (1) by virtue of the addictive process, individuals are predisposed to depend on some kind of pleasure-producing behavior to regulate their (dysregulated) affects and sense of self; (2) they take a mood-altering substance or engage in some other rewarding behavior; and (3) the effects of the substance-taking or other behavior—positively reinforcing, negatively reinforcing, and aversive—and the individuals' expectations about its effects in the future combine to determine the likelihood that they will select to take that substance or engage in that behavior as their addictive behavior of choice. These effects and expectations are determined by both genetic and experiential factors.

The foregoing observation suggests that sex addiction develops in addictively predisposed individuals who experience some form of sexual stimulation and learn to expect that sexual behavior will not only provide pleasure but also relieve the affects that they experience as most dangerous, through means that are congruent with their characteristic modes of adaptation.

Experiences that sexual stimulation provides both pleasure and relief begin early in life. Newborn infants' first and most fundamental experience of pleasure and soothing is not suckling a breast or ingesting milk. It is skin-to-skin contact. This bodily contact supplies warmth, which relieves the first and (usually) most extreme distress of early extrauterine life. It also soothes somatic-affective discomfort and thereby provides a foundation for the subsequent development of affect-regulation. Skin precedes the mouth as the first erogenous zone. As infants develop, their responses to skin-to-skin contact begin to differentiate and become specific to particular areas of contact. Physiologic and behavioral changes that indicate intense pleasure in response to bodily contact in the genital region become increasingly distinct from the pleasure responses to skin-to-skin contact with other areas of the body.

Infant observation studies have reported that at some time between 15 and 19 months of age, both male and female infants normally demonstrate an upsurge in genital self-manipulation. In time this takes on a more recognizably masturbatory pattern. Infants actively produce warm, pleasurable feelings in their own bodies which earlier they only passively experienced through nurturing maternal care. The genital self-stimulation seems from the outset to be mother-directed. During masturbation, children often engage their mothers visually and smile at them. Later they use mother-connected objects, such as nursing bottles or transitional objects, for masturbatory activities. Children’s early sexual interest and activity not only provide pleasure and tension release but moreover (in healthy development) facilitate the consolidation of self and object representations.

From the beginning of life, skin-to-skin sensations and then sexual sensations are intimately associated with survival, mother, and soothing, the last of which is the foundation of affect-regulation. Genital self-manipulation in the second year of life is mother-oriented and influences the development of not only mental representations of emotionally significant others but also sense of self and self-regulation. Significantly, infants are able to engage in sexual self-stimulation and provide the associated pleasurable and soothing feelings for themselves, independent of others’ availability and willingness to soothe them. Not until they are older are they able to obtain and consume food on their own. A subjective drive to provide regulation of one’s inner states by external means that seem to be within one’s own control is at the core of addiction. Freud characterized masturbation as “the primal addiction” from which all other addictive disorders derive.

Early experiences of sexual stimulation that provide both pleasure and relief are practically universal among neurologically intact infants. Yet not all individuals who are addictively predisposed become sex addicts. The eventual development of sex addiction in addictively predisposed individuals typically is associated with either seduction (overt or covert) or maternal sexualization. Sexualization is the infiltration of unconscious sexual fantasy into ordinarily nonsexual behaviors that hence can be used as vehicles for fantasied sexual gratification, which functions not so much to provide pleasure as to protect oneself from potentially overwhelming affect or self-fragmentation.

(In this paragraph and the next, "mother" and "maternal" are intended to refer to the mothering person, whether she be the child's mother or a nonmaternal primary caregiver.) When a woman who unconsciously relies on sexualization for psychological survival becomes a mother, sexualization is a constituent of the matrix in which the mother–child relationship begins. The influence of maternal sexualization on that relationship and hence on the child's early development is typically greater when the child is the same sex as the woman's fantasized sexual object. During infancy, the emotionally and psychologically deprived child experiences the gratifying aspects of the mother's sexualization and overt or covert seduction as though they were the vitally needed but missing nurturance, holding, and soothing. The child comes to identify with the mother's sexualization and seductiveness as a way of preserving the tie to her. Sexual seductiveness may eventually become the child's predominant mode for relating to others and for expressing their desperate hunger for a fulfilling connection with another person.

In short, if the etiology of sex addiction were to be summed up in one brief equation, it would be: sex addiction = the addictive process + maternal sexualization.

TREATMENT

A core element of the medical way of thinking is the distinction between a disease and its symptoms and signs—observable indications of a disease that often result from the body's response to it. According to the understanding of addiction that is presented here, the addictive process is the underlying disease, and instances of addictive behavior are symptoms and signs of the disease. Among the corollaries of the disease-symptom distinction is that treating the symptoms–in this case, instances of addictive behavior–without treating the underlying disease–the addictive process–is likely to be of limited effectiveness. Meanwhile, a distinctive feature of addictive disorders is the frequency with which external and internal consequences of the symptoms exacerbate the disease or interfere with its treatment. An implication of these conditions is that treatment for an addictive disorder is most likely to be effective when it addresses both the symptomatic behavior and the underlying addictive process. Treatment modalities that are directed primarily toward addressing symptomatic behavior are discussed below under the heading, Symptomatic Behavior Management. Those that are directed primarily toward addressing the addictive process are discussed subsequently under the heading, Healing of the Addictive Process.

Symptomatic Behavior Management

A broad range of sexual behaviors can be used addictively, from those that in themselves are natural and benign aspects of healthy life, such as masturbation and consensual adult sex, to those that violate other people and inflict deep and lasting damage, such as sexual abuse of children and rape. The goals of symptomatic behavior management in the treatment of sex addiction vary with the degrees of violation and damage that are associated with the behavior.

At the more violating-damaging end of the continuum, the goal is to prevent any occurrence of symptomatic sexual behavior. At the less violating-damaging end, the ultimate goal is to facilitate addicts' development of capacities to distinguish between healthy and unhealthy sexual behaviors (the latter being those that, if engaged in by this particular person, are likely to be engaged in addictively and lead to harmful consequences) and between healthy and unhealthy occasions of engagement in a sexual behavior that is not in itself unhealthy for this person. In general, engagement in a healthy sexual behavior is unhealthy when its primary motivation (usually unconscious) is to keep some heretofore unconscious mental content from entering awareness.

The middle region is occupied by sexual behaviors that have been identified as unhealthy for this person, either initially or as a result of their having learned how to identify unhealthy behaviors. In this broad area, the goals of symptomatic behavior management are twofold: (1) to prevent the occurrence of symptomatic sexual behavior; and (2) if symptomatic behavior does occur, to bring about its discontinuation before it leads to significant harmful consequences. Symptomatic behavior management consists primarily of behavior mastery training, other cognitive-behavioral techniques, and dialectical behavior therapy (DBT). On occasion, it may also require behavioral modification or endocrinologic medication. While support groups do not constitute treatment, they can be valuable adjuncts that sustain people’s efforts to refrain from engaging in symptomatic behavior.

While discontinuation of all sexual behavior is not a goal of treatment, an initial period of 1 to 3 months during which addicts refrain from any kind of sexual behavior is often a useful launch into treatment. Few if any are able yet to reliably distinguish healthy from unhealthy with respect to either sexual behavior or occasions of engagement in it. During this period of refrainment, self-perpetuating behavioral, neurobiologic, and psychodynamic patterns will have a chance to cool down, and subjective states that symptomatic behavior had functioned to exclude from awareness may begin to bubble in. During such a period of complete sexual refrainment, any instance of sexual behavior can be explored to identify vulnerability factors, triggers, and what the trigger evoked internally that overwhelmed or undermined the person’s control system.

Behavior Mastery Training. Behavior mastery training is a set of techniques that sex addicts can use on their own to help themselves refrain from enacting symptomatic behavior. The three primary techniques are risk management, urge handling, and symptom containment.

RISK MANAGEMENT. Symptomatic sexual behavior is more likely to occur under some conditions than others. Risk management teaches sex addicts to identify factors that increase the likelihood of symptomatic behavior. They then are trained to initiate protection against the influence of these factors before urges to engage in the behavior overwhelm their ability to organize an effective defense.

The first step is to review specific instances of symptomatic behavior in detail while focusing on the thoughts, feelings, and circumstances that led up to the act. Such reviews often can illuminate sequences of steps that culminate in episodes of symptomatic behavior, key features or patterns that different sequences share, and triggers that precipitate the sequences. Once identified, high-risk situations can be avoided or entered only when accompanied. Behaviors that increase susceptibility can be discontinued or restricted. Recognition of risk factors is also facilitated when the addict practices self-monitoring: for example, by keeping a daily journal of situations, times of day, moods, people, activities, and thoughts that are associated with urges to engage in symptomatic behavior.

The earlier in the process leading up to symptomatic behavior that a person's healthy resources can be mobilized, the less likely they are to engage in symptomatic behavior. Consequently, a critical component of risk management is identification of behavioral warning signs. These are behaviors that occur early in a sequence that culminates in symptomatic behavior and thus can be read as signals that such a sequence has already begun (and must be interrupted).

URGE HANDLING. During the course of healing from sex addiction, urges to engage in symptomatic behavior are inevitable. Urge-handling comprises behavioral and cognitive skills that help people who use sexual behavior addictively manage these urges without reverting to symptomatic behavior.

Behavioral skills are actions that can be put into practice. Beyond avoiding an expanded range of risk factors, they include specific protective responses that people can enact when they are exposed to a risk factor—for example, physically leaving a situation, engaging in an activity that precludes the next step toward sexually acting out, or contacting a peer and talking about the situation. Behavioral skills are most effective when planned in advance and rehearsed, since the concentration that formulation and performance of an unfamiliar behavioral plan require may not be available in the midst of a strong addictive urge.

Cognitive skills are methods that people can employ to modify their own thoughts. They include (1) accepting urges to engage in symptomatic sexual behavior as natural accompaniments of the healing process and hence not as signs of inadequacy; (2) recognizing that urges are not imperatives to act but surface manifestations of affects; (3) reviewing the benefits of refraining from symptomatic behavior and the potential negative consequences of engaging in that behavior; and (4) when fantasies of engaging in the behavior persist, prolonging the fantasies beyond the pleasurable experiences and imagining the harmful consequences and accompanying guilt, shame, or humiliation that could ensue. Particularly in the early phases of healing, cognitive skills are most likely to be effective in practice when they have been verbally rehearsed with another person.

SYMPTOM CONTAINMENT. An episode of symptom recurrence begins with an instance of symptomatic sexual behavior and ends when the addict resumes refraining from such behavior before it produces significant harmful consequences. If the behavior continues to the point that it generates harmful consequences, the event is not a recurrence (of symptoms) but a reversion (of health status). Symptom-containment skills are developed in order to prevent a recurrence from becoming a reversion. The goal is to reinstate refrainment before the symptomatic behavior generates significant harmful consequences.

A discussion of symptom-containment skills presupposes that symptomatic behavior can recur and thus seems to imply that recurrences can to some degree be tolerated. Such tolerance may make sense when the symptomatic behavior is visiting strip clubs or masturbating to pornography. But some sexual behaviors, such as rape and sexual contact with children, are so harmful to others that zero tolerance for the behavior is the only reasonable treatment stance. When people who use sexual behavior addictively are not reliably able to restrain themselves from acting on urges to engage in such seriously harmful behavior, treatment begins with behavior modification or endocrinologic pharmacotherapy, or both. These interventions are continued until such time as the patients are reliably able to restrain themselves. If a sex-addicted person who is not reliably able to restrain him- or herself from acting on urges to engage in seriously harmful sexual behavior does not accept a recommendation for behavior modification and/or endocrinologic pharmacotherapy, a clinician would be well-advised not to undertake the treatment of this person. The risks are considerable, and if the patient is not sufficiently motivated to cooperate with a recommendation of this type, the likelihood of significant progress in treatment is too small to justify such risks. The discussion of symptom-containment that follows thus applies only to the treatment of sex-addicted individuals whose sexual behaviors, while far from benign, are not so seriously harmful to others.

Even when motivation is genuine and therapeutic management is competent, symptomatic behavior can recur. However much a sex addict may consciously wish to refrain from symptomatic behavior, a strong unconscious motivation to resume the behavior persists—fueled by the same motivations and self-regulatory impairments that originally drove the addictive use of the behavior. Some combination of emotional stress, narcissistic injury, weakening of internal behavioral controls, and gaps in the external support structure can converge momentarily to shift the motivational balance in favor of resuming the behavior. When such a motivational shift happens to coincide with the availability of an opportunity to engage in the behavior, symptom recurrence is likely.

Like urge-handling skills, symptom-containment skills consist of behavioral and cognitive skills. Behavioral skills are actions that limit the progression of symptomatic behavior and reinstitute healing, such as communicating with one's therapist, getting together with a supportive friend, or attending a support group meeting. Similar to the behavioral skills of urge-coping, those of symptom containment are most effective when they are planned in advance and rehearsed.

Cognitive skills for handling symptom recurrence involve reframing the recurrence or reconceptualizing what having engaged in symptomatic behavior means. A recurrence is most usefully understood not as a sign of failure or the beginning of reversion but as a signal that a change in the addict's self-management program is needed. It represents a combination of affect and availability for which the addict had not yet developed an effective coping strategy. A recurrence thus reframed becomes an opportunity to learn and to improve the therapeutic program. A problem-solving approach is ultimately more effective in containing symptomatic recurrences and promoting healing than is judgmental reproof or self-reproach. In the case of both behavioral and cognitive skills for symptom containment, the sooner after a recurrence that an Intervention occurs, the greater the probability of preventing reversion.

Behavior mastery training conceptualizes urges to engage in symptomatic sexual behavior as signals of disruptive affects that addicts need to develop healthier, more adaptive means to manage. In thus shifting the focus from controlling the behavior to understanding the affects, behavior mastery training provides a natural bridge from behavior management to psychodynamic psychotherapy. It can be imparted in therapeutic groups or integrated into individual psychotherapy.

Other Cognitive-Behavioral Techniques. Cognitive-behavioral therapy (CBT) is a psychotherapeutic approach that conceptualizes psychological problems mainly in terms of maladaptive learning. Treatment focuses on identifying and correcting inaccurate or distorted thoughts as a means of changing maladaptive behaviors and emotional states. Interventions typically are directive, practical, task-oriented, and educative in nature. Through CBT, people acquire new information, develop new ways of thinking and reacting, and learn specific skills that they can use for the rest of their lives.

A variety of cognitive-behavioral techniques in addition to behavior mastery training can be helpful in the treatment of sex addiction. Patients can be helped to identify distorted thoughts that enable them to dehumanize people who are hurt by their symptomatic sexual behavior or to avoid becoming aware of the emotional, relational, and existential ways in which their sexual behavior harms themselves.

Victim empathy training helps patients become aware of how their behavior affects other people's emotional states. The training focuses not only on direct victims on whom the sexual behavior was perpetrated (when it was inflicted on others) but also on indirect victims: people in the patients' lives who have been hurt or damaged as a result of the patients' sexual behavior, other behavior that prepared for or concealed the sexual behavior, and (eventually, perhaps) behavioral symptoms of the patients' character pathology. Patients are helped to become aware of how their symptomatic behavior has caused people about whom they care to have felt or to feel painful emotional states: abandonment feelings, intense anxiety, loss of self-esteem, and so on. The empathically experienced pain and associated feelings of remorse, guilt, or self-hatred can then function as the aversive affects that are associated with symptomatic sexual behavior in the aversive affect induction procedure. The potential usefulness of this variant is limited to patients who are capable of consciously accessible empathic perception.

Most other cognitive-behavioral techniques focus not on the symptomatic behavior itself but on other aspects of a person's life that predispose to reliance on symptomatic behavior as a means of coping with painful affects and unmet needs. Examples of skills training that helps patients to more effectively manage their affects and meet their needs include anger management, assertiveness training, problem-solving, communications training (e.g., learning to speak in the format “When you did *x*, I felt *y*” rather than accusing or blaming), and lessons in how to initiate and maintain healthy sexual relationships.

Dialectical Behavior Therapy. DBT is a variant of CBT that was developed to enable people who suffer from borderline personality disorder (BPD) to manage their emotions more effectively and thereby gain control over their self-harming and self-endangering behaviors. It has been applied also in the treatment of substance use disorders, eating disorders, and sex addiction. DBT understands the core vulnerability in BPD to be emotional dysregulation. Its fundamental objective is to help patients develop skills that enhance emotional regulation and implement them in their daily lives. These skills are organized into four modules: mindfulness skills, distress tolerance skills, emotion regulation skills, and interpersonal effectiveness skills.

Mindfulness skills in DBT are cognitive-behavioral adaptations of Eastern meditation practice and philosophy. They are identified as observation, description, and participation in the present moment effectively and without judgment. Patients are instructed to nonjudgmentally focus their attention on their inner and outer environments and then describe their observations. They also are taught to act with focused attention and a nonjudgmental mind. A version of meditation typically is taught as a means of developing mindfulness skills.

Distress tolerance skills enable patients to experience painful and disruptive feelings that they cannot immediately change without impulsively engaging in maladaptive, self-harming actions in a desperate reach for relief. At their core is learning to practice nonevaluative acceptance of experienced reality: one’s current emotional state(s), the current situation, oneself, and the world in which one lives. The goal is to help patients become capable of calmly recognizing painful or threatening situations and their impact rather than hiding from them or being overwhelmed. They then are more able to make wise decisions about whether to act and, if so, how.

Emotion regulation skills enable patients to reduce their maladaptive emotional reactions to situations in day-to-day life and their tendency to respond to such emotional reactions with impulsive behavior. Patients are taught to (1) identify and describe their emotions, (2) understand the functions of their emotions, (3) reduce their vulnerability to the triggering of impulsive actions by intense emotions, (4) identify and challenge cognitions that lead to maladaptive emotional reactions, (5) identify obstacles to modifying their maladaptive emotional reactions, and (6) increase the frequency of positive emotional experiences in their lives.

Interpersonal effectiveness skills in DBT range from basic social skills to assertiveness and goal-oriented interpersonal problem-solving. They include effective strategies for asking for what one wants, declining requests to do something one does not want to do, and negotiating interpersonal conflict. The skills taught are intended to maximize the chances that a person’s goals in a specific situation will be met while at the same time not damaging either the relationship or the person’s self-respect.

Patients’ histories usually indicate general areas where skills are deficient and thus where skills training may be helpful or necessary. More specific skill deficits and their consequences can be elucidated by chain analysis, a CBT technique that facilitates identification of the factors that promoted or permitted an occurrence of maladaptive behavior. When an instance of self-harming or self-endangering behavior has occurred, the sequence of cues, emotional responses, and cognitions that led up to and followed the behavior are reviewed in detail. The therapist and patient then attempt to determine the specific vulnerability factors and antecedent events that increased the likelihood that the behavior would occur and the consequences that reinforced the behavior. The goal of chain analysis is to identify the patterns of variables that tend to control the patient’s dysfunctional behaviors. This information then guides training in skills that reduce vulnerabilities in the future and lower the risk that precipitating events will recur. Chain analyses in DBT often focus on the various ways in which emotions are related to the initiation and maintenance of dysfunctional behaviors. Attention is directed toward the intense or aversive emotional states that may precede target behaviors as well as the potential emotion regulatory function that such behaviors may serve.

To the extent that DBT helps patients develop skills that enhance their emotional regulation, it contributes not only to symptomatic behavior management but also to the healing of the addictive process.

Treatments to Modify Paraphilic Behavior. As noted earlier, paraphilic disorders and sex addictions are different kinds of disorders that share an area of overlap. For many of the individuals who occupy this area of overlap, failure to address the paraphilic nature of their arousal patterns, urges, and sexual behavior can undermine the effectiveness of treatment for sex addiction. In some of these individuals, those who are paraphilic predators, treatment for sex addiction outside a locked facility may entail too much risk of harm to others unless their paraphilic behavior and sometimes their capacity for sexual arousal are shut down. Worthy of note is that the class of paraphilias includes sexual behavior in which one person coerces another into participation. Such behavior is motivated more by sadism than by craving for sexual gratification. In DSM-5-TR, sexual sadism disorder is included among the paraphilic disorders.

BEHAVIOR MODIFICATION. Behavioral methods are oriented toward helping paraphilic individuals reduce their erotic arousal to paraphilic objects/subjects or shift the balance of erotic arousal potential from paraphilic to nonparaphilic interests. The primary forms of behavioral treatment for paraphilic disorders are aversion conditioning, covert sensitization, masturbatory training, imaginal desensitization, and aversive affect induction. All rely on classical or Pavlovian conditioning, in which the repeated pairing of two stimuli increasingly enables one of them to elicit physiologic and affective responses that previously had been elicited only by the other.

In aversion conditioning, unwanted patterns of sexual behavior are linked repeatedly with unpleasant stimuli, most often electric shock or foul odors. This straightforward classical conditioning procedure is intended to transform stimulus features of the sexual behavior into conditioned stimuli or triggers for the aversive responses to the unpleasant stimuli. A number of uncontrolled studies found that aversion conditioning yielded positive results, while one found no significant effect. Another study found that the success of aversion conditioning depended on the availability of appropriate sexual outlets and the absence of severe nonsexual pathology.

Covert sensitization was developed as an imaginary form of aversion conditioning, in which fantasies of paraphilic arousal are paired with fantasies of aversive events in order to promote a learned association between paraphilic themes and unpleasant feelings. In covert sensitization training, patients are first led through a relaxation exercise, which enhances their receptiveness to vivid mental imagery and helps them to focus their attention on the imagery rather than on external events. They then are instructed to visualize themselves engaging in some aspect of their paraphilic behaviors and are encouraged to feel personally involved. At that point, patients are told to imagine either an unpleasant event, such as nausea and vomiting, or an aversive consequence that could follow the behavior in reality, such as public humiliation or shame at being discovered by someone whose esteem deeply matters to them. The aversive valence of the imagined noxious scene can be reinforced by the presentation of a nauseating odor. Studies have found covert sensitization to be effective in the treatment of exhibitionism and pedophilia. It neither inflicts physical discomfort nor requires expensive technology, and patients can be taught to self-administer the treatment. Consequently, many behavior therapists prefer to begin with covert sensitization, turning to aversion therapy only for patients who do not respond to covert sensitization.

Masturbatory training attempts to shift paraphilic patients’ arousal patterns in the conventional direction by controlling the visual stimuli they view or the fantasies they conjure when they masturbate. The two main forms of masturbatory training are satiation and fading. Satiation involves instructing male patients to masturbate to orgasm with conventional sexual stimuli or fantasies and then to continue masturbating while visualizing their deviant objects. The idea is that masturbating after orgasm will extinguish erotic arousal to paraphilic fantasies. The aim of the fading technique is to shift patients' sexual fantasies from deviant to conventional during periods of sexual arousal. Patients may be provided with a video or sequence of slides on which to focus while they masturbate. As they masturbate, the visual presentation progressively fades from nonparaphilic to paraphilic sexual images. Alternatively, patients may be instructed to begin masturbating to paraphilic fantasies and then gradually evolve the fantasies to nonparaphilic sexual scenarios. One study of masturbatory satiation reported mild success, but other therapeutic modalities were provided at the same time and no post-treatment follow-up was mentioned.

The technique of imaginal desensitization is based on the theory that once an individual has developed a sequence of sexual activity that begins with a sexual thought or fantasy and ends with overt behavior, interruption of the sequence results in anxiety, which then motivates completion of the sequence. Imaginal desensitization uses the methods of systematic desensitization to diminish the anxiety that is aroused by interruption of that sequence. The procedure begins with the patient describing a scene in which they are stimulated to carry out the paraphilic sexual activities for which they sought treatment. After relaxation is induced, the patient visualizes performing a sequence of behaviors that leads to noncompletion of the act. Relaxation is allowed to develop with each visualized behavior before the patient proceeds to the next behavior in the sequence. The patient is thus trained to visualize not completing the sexual act while remaining relaxed. One study reported imaginal desensitization to be more effective than covert sensitization.

Aversive affect induction involves focusing on the negative emotional consequences of symptomatic sexual behavior while the urge to engage in such behavior is active. Specific cues that patients have identified as being associated in their minds with symptomatic sexual behavior (e.g., the home page of a website that features sadistic pornography, the sounds of children playing) are presented to evoke an urge to engage in that behavior. Patients are then asked to describe the emotions that they experience before, during, and after an episode of symptomatic sexual behavior—for example, excitement, power, sexual release, shame, self-loathing. The intent is that associations develop in the patient's mind between urges to engage in symptomatic sexual behavior and narcissistically gashing emotions, such as shame and self-loathing. Aversive affect induction has the potential to be helpful in nonparaphilic sex addiction, but its use in that capacity has not been studied.

ENDOCRINOLOGIC AGENTS. Endocrinologic agents are used primarily to reduce sex drive in paraphilic men. These include progestogens and gonadotropin-releasing hormone (GnRH) agonists. Estrogens have been studied for use in controlling sexual behavior but were found to have too many undesirable effects.

The progestogens used in the treatment of paraphilias include cyproterone acetate (CPA) and medroxyprogesterone acetate (MPA). Their action as progestogens blocks the secretion of the pituitary gonadotropins, luteinizing hormone (LH), and follicle-stimulating hormone (FSH). CPA has strong progestational and antigonadotropic activity. In addition, it competes with androgens for androgen receptors and thus is an antiandrogen. CPA has been reported to decrease sexually driven aggression and to reduce arousal responses to pedophilic and coercive sexual stimuli but not to mutually consenting adult sex scenes. Side effects of weight gain, depression, and feminization have been noted. MPA exhibits no antiandrogenic activity. Its efficacy in treating paraphilias results primarily from its induction of the enzyme testosterone reductase in the liver, which decreases circulating levels of testosterone. A number of clinical studies have demonstrated that MPA can reduce sex drive, sexual thoughts, sexual behavior, and aggressiveness. Unfortunately, its administration is frequently associated with unpleasant side effects and occasionally with serious adverse events such as deep vein thrombosis, pulmonary embolism, hepatic dysfunction, and precipitation or exacerbation of diabetes mellitus. Noncompliance also is a significant problem. One study reported a dropout rate of greater than 50%. Both progestogens are available in oral and IM formulations.

Analogs of GnRH have been developed that have higher potency and longer duration of action than naturally occurring GnRH. Triptorelin is injected once a month and leuprolide acetate is injected once every 3 months. Initial administration of these agents raises serum testosterone by stimulating pituitary gonadotrope cells to increase production of LH and FSH. LH then stimulates testicular Leydig cells to synthesize more androgens. During the first 2 to 4 weeks after initiation of treatment, sex drive may actually increase. However, continuous administration of long-acting GnRH analogs produces downregulation of GnRH receptors on the pituitary gonadotropes. GnRH receptor downregulation leads to decreased secretion of LH and FSH, a consequent decrease in the synthesis of testosterone, and thus a substantial diminution of sex drive. Several reports of uncontrolled open-label trials of GnRH agonists in the treatment of paraphilias and hypersexual disorders have been published, and all demonstrated significant positive effects. The main side effects were erectile dysfunction, hot flashes, and a decrease in bone density. These results suggest that GnRH agonists could prove to be a more effective, safer, and less noxious alternative to the progestogenic agents. More study of these promising agents is warranted.

A review article recommended that endocrinologic agents be used in combination with serotonergic antidepressants, not as monotherapies, and that progestogens be tried before LHRH (GnRH) agonists. Another review reported that patients who had taken CPA, MPA, or selective serotonin reuptake inhibitors (SSRIs) and subsequently took a GnRH agonist reported greater therapeutic benefits from the latter.

Although these medications do not change the direction of the paraphilic person’s sexual interest, they decrease the intensity of sexual drive, so the person is more in control and less likely to engage in paraphilic behavior. The primary therapeutic function of these medications is to reduce sex drive to manageable levels in those individuals whose ability to control their behavioral impulses is so impaired as to put them at risk either to injure themselves or others or to render them unresponsive to psychological interventions. Endocrinologic agents can lower the risk of problematic sexual behavior during the interval between the initiation of treatment and the consolidation of the changes that affect-regulating agents, behavior modification, DBT, and/or psychodynamic psychotherapy can induce. A review article on the treatment of paraphilias concluded that the combination of pharmacologic therapy and psychotherapy is associated with greater efficacy than either treatment alone.

A number of other medications that were developed for other purposes have side effects of decreased sex drive or erectile dysfunction and are on occasion used informally—often by patients on their own initiative and sometimes without informed medical direction—in attempts to reduce the symptoms of sex addiction. These include antihypertensive agents, anticholinergics, antihistamines, disulfiram, antipsychotic medications, and hair growth stimulants. None of these have been rigorously studied for use in controlling sexual fantasies, urges, or behavior.

Support Groups. Support groups are distinguished from treatment groups in that the effective agent is neither a therapeutic method/technique nor a person who leads the group but the support that group members provide for one another. Support groups for sex addicts typically provide a safe, minimally judgmental sanctuary where they can talk about shame- and guilt-laden matters with others whose presence in the group and self-identification as sex addicts indicates that they too struggle with comparable shame- and guilt-laden matters. In such groups, sex addicts can learn from one another’s experiences, identify with others who manage their addiction more effectively or make good decisions more often, benefit from others’ encouragement and positive responses to their successes, and feel compassion for others’ stumbles when they are not yet ready to begin feeling compassion for themselves. When they or others in the group engage in symptomatic behavior (recurrence or reversion) and are welcomed back into the group with neither reprimand nor willingness to hear excuses, they can learn to be more self-accepting and also more honest with themselves. More broadly, sex addiction support groups provide opportunities to experience commonality (the “I’m not the only one” experience), empathic bonding, group cohesiveness, and sense of belonging. They thereby can facilitate the development of abilities to make meaningful connections with others and to turn to people in times of need instead of turning to addictive behavior.

The most widespread and easy-to-access support groups for sex addicts are the 12-step groups that have adapted the model of Alcoholics Anonymous to sex addiction. They offer membership at no charge for as long as a person wishes, around-the-clock support, and connection with a worldwide support network. They present a coherent framework for approaching addictive problems and life in general that includes self-awareness, acceptance of what is, and the centrality of human relationships. In addition, 12-step groups offer (but do not impose) a nonjudgmental, nondogmatic spiritual foundation. Meanwhile, a consequence of the lack of formal leadership structures and internal controls in 12-step groups is that the characteristics of a particular group are shaped primarily by the individuals who attend it. How much a person benefits from and enjoys being in a group is a function of their goodness of fit with the other participants in the group.

Healing of the Addictive Process

The addictive process can be addressed through psychiatric medications that improve affect regulation (which includes regulation of anxiety) or reduce impulsivity and psychodynamic psychotherapy. Meditation and yoga can be helpful adjuncts to the primary treatment modalities. Since the addictive process underlies all addictive disorders, the material in this section is not specific to sex addiction but applies to other addictive disorders as well.

Psychiatric Pharmacotherapy. Psychiatric pharmacotherapy is direct intervention to improve emotional and behavioral self-regulation as well as to treat other symptoms of comorbid psychiatric disorders. In sex addiction, craving and urges to act out are expressions of dysregulated emotional states. Such urges are more likely to be acted out when behavioral regulation is impaired. Consequently, enhancement of affect regulation tends to diminish the frequency and intensity of addictive urges while enhancement of behavioral regulation reduces the likelihood that urges will lead to an episode of symptomatic behavior. Medications that improve affect regulation often also reduce impulsivity, either directly or through their effects on affect regulation.

AFFECT-REGULATING AGENTS. A number of case reports and open-label studies have provided evidence for the efficacy of affect-regulating agents, particularly the SSRIs, in the treatment of paraphilias and nonparaphilic sex addictions, even in the absence of major depression. Some studies found that paraphilic patients with comorbid depression showed a concurrent decrease in paraphilic behavior when their depressive symptoms improved. Agents that have been found to be effective include fluoxetine, sertraline, citalopram, paroxetine, fluvoxamine, nefazodone, imipramine, desipramine, clomipramine, lithium carbonate, topiramate, carbamazepine, divalproex, buspirone, and electroconvulsive therapy. Most of these studies reported a positive response rate in the range of 50% to 90%. One investigator reported anecdotally that augmentation of an SSRI with bupropion or a psychostimulant can further reduce sexual fantasies, urges, and behavior, particularly when symptoms of ADHD are present. Others have found that divalproex or lamotrigine can be helpful for sex addiction symptoms that arise in the context of atypical bipolar conditions or “emotionally unstable character disorders.” Antidepressant medications, especially the SSRIs, can produce diminished libido as a side effect, but a number of studies noted that antidepressants reduced the drive for symptomatic sexual behavior without decreasing the drive for healthy sexual behavior. The only double-blind, placebo-controlled study in this series found that citalopram reduced the frequency of sex addicts’ symptomatic sexual urges, masturbation, and use of pornography while having no significant effect on partnered sexual behaviors. The large number of positive findings is encouraging, but the confidence with which conclusions can be drawn from them is limited by the paucity of controlled studies.

That the symptoms of sex addiction respond to affect-regulating medications in a pattern that resembles the response pattern of affective disorder symptoms is consistent with the earlier description of the addictive process that underlies sex addiction. As formulated neurobiologically, the addictive process originates in impaired affect regulation, impaired behavioral inhibition, and impaired functioning of the motivational-reward system. Addictive urges can be recognized to represent both (1) expressions of dysregulated affect, often a nonspecific tense or irritable dysphoria that is experienced consciously less as an affect than as an urge to do something, usually to engage in symptomatic behavior; and (2) conditioned stimuli for coping responses (responses to cope with dysregulated affect) that have been learned in the context of impaired behavioral inhibition and impaired motivational-reward function. Interventions that enhance affect regulation and behavioral inhibition can thus be expected to diminish addictive craving and symptomatic behavior.

In nontechnical terms, the impairment of affect regulation that underlies addiction can be seen as an imbalance between the intensity of patients' affects, particularly their painful affects, and the capacity of their built-in coping systems. Painful affects, such as anxiety and dysphoria, are not in themselves pathologic or "negative." They are signals that people send to themselves to call their attention to something that is important for their physical or psychological survival. Painful affects become a problem when their intensity overwhelms or threatens to overwhelm a person's coping system. They are then experienced as extreme dangers that must be escaped or avoided and are no longer used as signals. Often patients will not consciously experience the painful affects or even the threats of these affects but will feel only a vague unease. Or they may be closed off to their affects and may experience only bodily states that they interpret as urges to engage in some kind of symptomatic behavior. People who have learned that certain self-initiated behaviors can quell such dangerous affects and prevent the system from being overwhelmed will experience strong urges to engage in these behaviors whenever such affects threaten. This process forms the core of addictive and compulsive patterns of behavior.

Psychiatric medication works by restoring or reinforcing the brain chemistry that is associated with patients' built-in coping systems. Although psychodynamic psychotherapy also can enhance the capacity of their coping systems, patients' accessibility to psychotherapy and to what it can offer may be limited when they are emotionally overwhelmed. To the extent that they shut down or shut out subjective experience because they unconsciously perceive that it threatens their psychological survival, their access to their inner lives will be compromised. Moreover, states of being overwhelmed can disorganize or disrupt cognitive functioning, so they not only are less aware of material with which they can work, but they also are less able to work with the material of which they are aware. When psychiatric medication is effective in establishing a more stable balance between patients' affects and the capacity of their coping systems, the coping systems are less likely to be overwhelmed and patients can better tolerate experiencing their painful affects as affects (rather than as states of emergency). Consequently, they will have more conscious access to their inner states, and their cognitive functions will be more organized and free to operate. These developments provide them with more opportunity to work with their inner states in psychotherapy. Subsequent psychotherapeutic work can then help them to develop healthier, more adaptive ways of regulating and responding to their inner states, which may eventually enable them to manage without psychiatric medication.

MU-OPIOID RECEPTOR ANTAGONISTS. The mu-opioid receptor antagonist naltrexone has been reported to reduce symptoms of “compulsive sexual behavior” as an adjunct to affect-regulating medications, primarily SSRIs. Other studies have found naltrexone by itself to reduce urges and symptomatic behavior in gambling disorder and kleptomania as well as in substance use disorders.

Some reports indicate that it reduces the pleasure that is associated with symptomatic behavior but not the pleasure that occurs in the course of normal living.

Psychodynamic Psychotherapy. Psychodynamic (or psychoanalytic) psychotherapy is a circumscribed dialogue in the context of an intimate relationship between two people, the explicit goal of which is to help one of the participants become more fully self-aware in order to master emotional distress and remedy impairments in social or occupational functioning. The central role of developing self-awareness distinguishes psychodynamic psychotherapy from other forms of psychological treatment. Mastering emotional distress and remedying functional impairment identify this process as treatment, above and beyond the educational and growth-promoting functions that it serves. In short, psychodynamic psychotherapy is a dialogue that fosters self-awareness in order to heal the personality. Its functional core is constituted by three interacting processes: understanding, integration, and internalization.

UNDERSTANDING. Understanding as a component of psychodynamic psychotherapy is verbally mediated knowledge that the psychotherapist imparts in the course of treatment. It is a primarily cognitive process that is not specific to psychodynamic psychotherapy. What is more specific (in addition to the content) is the recognition that patients’ difficulties in understanding or in the practical application of what has been understood may reflect motivational factors that operate outside their conscious awareness. Moreover, psychodynamic psychotherapy regards these difficulties not as obstacles but as opportunities to explore the nature, function, and history of the underlying motivational factors. Both general and specific understanding are useful.

General understanding concerns information about a condition from which a person may be suffering and factors that characteristically contribute to its development. At the beginning of treatment, this process often enhances a person’s capacity and motivation to participate in psychodynamic psychotherapy.

For example, early in the treatment of sex addicts, the psychotherapist can briefly describe in simple terms the psychodynamic processes that underlie sex addiction and other addictive disorders. Such information, first of all, provides an organized array of concepts that can help people name their experiences, find order in them, and integrate them into their cognitive networks. Having words to describe inner states and processes gives the conscious mind a handle by which to apprehend them. Moreover, naming inner realities provides perspective. The part of the mind that names in so doing steps back from the part of the mind that feels. Painful emotions and self-states no longer occupy the totality of a person’s subjective psychic space, so they become less likely to be overwhelmed by them.

As noted earlier, emotions are important self-generated signals. But when they become or threaten to become overwhelming, people experience them as threats to their survival. Their survival instincts then drive them to ward off or shut down their emotions. Insofar as they do so, they are unable to benefit from their emotions’ signal function and the information that it conveys. To the extent that people get perspective and become less overwhelmed, they can begin to experience their emotions as signals rather than as states of emergency. Significant progress down this path usually depends on the integration and internalization components of psychodynamic psychotherapy.

General understanding also offers benefits that are more tangibly practical and specific to addiction. Learning that addictive behavior represents an attempt to ward off painful or disorganizing inner states that threaten to overwhelm the person attenuates the shame and self-loathing that addicts feel in relation to their symptomatic behavior and reduces associated self-destructive inclinations. Hearing that urges to engage in symptomatic behavior are conscious manifestations of anxiety that surges when such threats arise shifts addicts’ focus from fighting the urge to understanding the threat. In consequence, behavioral symptom management becomes less effortful, and addicts become more capable of using psychodynamic psychotherapy effectively.

General information about developmental factors that may have contributed to the genesis of sex addiction provides a framework within which the condition becomes comprehensible, often thereby diminishing addicts' anxiety. Moreover, it reduces the self-hatred or self-punishment that people typically experience when they blame themselves for the emotional pain and functional impairment that are associated with sex addiction.

Specific understanding concerns the patient him- or herself and what actually is going on in their life. The primary focus of specific understanding is the dynamic process that has led up to a particular instance of symptomatic sexual behavior. This process is typically of the following form: (1) impaired regulation of emotional states develops out of the interaction between genetic predispositions and the infant's relationship with their early caregiving environment; (2) specific vulnerabilities develop as the emotions that are evoked by painful or scary childhood events (or less commonly, postchildhood events) are hyperintense to the point of being traumatic due to impaired affect regulation; (3) later in life a current event taps into one of those vulnerabilities in the context of deficient affect regulation and thereby triggers a disproportionately intense emotional response that threatens to overwhelm the person; and (4) a desperate reflexive attempt to protect oneself from this threat precipitates an occurrence of symptomatic behavior.

In general, behavioral symptoms arise as (1) expressions of painful or disturbing emotional states that an impaired affect regulation system is not effectively managing, (2) unconsciously directed self-protective processes that operate to keep such emotional states and inner conflicts out of one's awareness, or (3) a combination of both. In addictive disorders, symptomatic behavior may be the only conscious manifestation of dysregulated emotional states that at an unconscious level threaten to burst into a patient's awareness and overwhelm them. Thus, in addition to being distressful or debilitating, symptomatic behaviors are conveyors of information about patients' emotional states, affect regulation systems, and self-protective processes.

The likelihood of accessing this information is greatest when inquiry focuses on the timing of changes in the symptomatic behavior—when it emerged out of quiescence, got worse, got better, or changed shape. The emergence or exacerbation of symptomatic behavior indicates that some event activated a painful or disturbing emotional state that the affect regulation system was unable to manage. The point of information access is temporal contiguity. Unless the timing of symptoms is random, the occurrence of an instance of symptomatic behavior implies an antecedent change in contextual conditions that (by evoking a potentially traumatic emotional state) led the behavior to occur when it did rather than at some other time. Specification of when the symptomatic behavior occurred guides an investigation of what was going on in the person's life just before the symptomatic act, with an eye toward identifying the event(s) that may have precipitated it.

The process of specific understanding typically is initiated when a patient reports an instance of symptomatic behavior. In the treatment of sex addiction, the initial focus is on symptomatic sexual behavior. As treatment progresses and as the frequency of symptomatic sexual behavior decreases, the focus may expand to other symptomatic or maladaptive behaviors. The process begins with a review of circumstances and interactions that preceded the episode of symptomatic behavior. During this review, antennae are tuned for any event that might have triggered an intense emotional response. Identification of the triggering event usually leads to identification of the emotional response, though on occasion the emotional response is identified first. The goal is to help patients increase their capacity to be aware of disproportionate intensity in their emotional responses before they act, so they can factor it out of their decisions about what actions to take.

Guided by the hypothesis that symptomatic behavior characteristically occurs when an event taps into an area of emotional vulnerability, the investigation readily expands in the direction of exploring the contributory factors that were shaped by personal history and intensified beyond the range of ordinary manageability by impaired affect regulation. As it does so, the therapeutic process shifts beyond understanding to include integration, the next component of psychodynamic psychotherapy to be discussed.

INTEGRATION. Integration refers to a process through which the patient's personality or self becomes increasingly integrated as its various aspects coalesce into a coherent whole. Psychologically, many people are characterized by deficits in integration on several levels. Without conscious intent or awareness, they reflexively engage in a variety of mental processes that function to protect them from emotional trauma. Most of these self-protective processes can be understood as ways of keeping out of people's awareness material that they unconsciously imagine would lead to overwhelming pain or self-fragmentation were they to become aware of it. Such material can include emotions, needs, wishes, fears, inner conflicts, core beliefs, or the self-protective processes themselves. Ultimately, what they are protecting themselves from most of the time are emotions that they fear would overwhelm and psychologically destroy them were they allowed to enter their consciousness.

For example, when people automatically push a need or a wish out of their awareness, they do so usually not because the need or wish is dangerous in itself, but because past experiences of other people's responses to their expressing that need or wish led them to associate it with potentially overwhelming emotions such as disappointment, loss, shame, guilt, rage, or humiliation. Similarly, when people automatically keep an inner conflict unconscious—for example, by allowing only one of its components into awareness at a time—they do so because the conflict is fraught with unbearable anxiety, which in turn develops because choosing one of the alternatives and giving up the other is associated with painful emotions that could be traumatic whichever alternative is chosen. Finally, the self-protective processes that people keep out of their awareness typically are not themselves infused with overwhelming emotion. But if someone were aware that a self-protective process was operating, it would be less likely to succeed in concealing the emotion-laden material that it was unconsciously intended to conceal. Thus, people also (unconsciously) hide their self-protective processes from their inner view.

Self-protective processes are not obstacles to self-awareness that must be torn down for healing to occur. Nor do they constitute a “false self” that must be dismantled or peeled away to liberate and reveal a person’s “true self.” People’s means of protecting themselves from emotional trauma are no less true than the mental processes and material that they keep from consciousness and no less important as potential or actual objects of self-awareness. Exploration of self-protective processes often yields treasure troves of information. In addition, a person’s psychological survival can at times depend on inner processes that protect them from emotional trauma. The goal of treatment is not to get rid of them but to increase the influence of patients’ conscious reason on when and how they operate.

People’s subjective experience and behavior are influenced by conscious thoughts, feelings, and reasoning, and also by emotions, needs, wishes, fears, inner conflicts, core beliefs, and self-protective processes of which they are not consciously aware. When such influences operate outside a person’s awareness, they are not susceptible to being modulated or contextualized by their conscious reason.

The process of integration occurs through consciousness. The critical step is becoming aware of feelings, thoughts, and fantasies that heretofore had remained unavailable to conscious awareness. Consciousness is primarily a verbally mediated process, so becoming aware of what had been going on in one’s mind unconsciously involves the development of associative links between unconscious mental processes and their verbal representations. Such links are the basic elements of integration. They typically develop as a result of temporal contiguity between the occurrence of a heretofore unconscious mental process and the conscious registration of its verbal designation or description.

Psychodynamic psychotherapy promotes integration (1) by providing an intimate interpersonal situation that tends to stir up people’s emotions, needs, wishes, fears, inner conflicts, core beliefs, and self-protective processes; (2) by emphasizing verbal communication and meaning, with both patient and therapist working together to create verbal renditions of the patient’s inner world; and (3) by facilitating identification and working through of self-protective processes that interfere with the flow of experience or communication. It thus provides optimal conditions for patients to link their unconscious mental processes with the cocreated verbal descriptions of them and thus develop the self-awareness that leads to integration.

Psychodynamic psychotherapy is less structured and less directed by the therapist than other forms of psychotherapy. Either from the very beginning or after an initial period during which background information is gathered, the steering wheel is left in or returned to the patient’s hands. Meanwhile, the therapist has at hand an auxiliary set of controls, as in a driver’s education car, so they can take over if for whatever reason the patient becomes temporarily incapable of driving safely. Recognition of this safety feature, which often remains implicit, helps enable patients to take more risks in exploring realms of their mental world with their therapists than they would have done on their own.

The patient’s primary instruction (which may or may not be explicitly stated) is to say whatever occurs to their mind without directing or censoring it, even if it seems irrelevant, trivial, offensive, or embarrassing. Psychodynamic factors outside people’s awareness operate all the time, influencing what they think, say, and do. During most of the time that they are awake, their conscious mental processes also are operating, and they dominate the conscious field of view. The influences of unconscious psychodynamic processes are difficult to discern except perhaps when indirectly embodied in slips of the tongue or symptomatic acts. To the extent that the dominance of conscious reason over patients’ verbal output is softened, the influence of unconscious psychodynamic factors over the content and form of what they say becomes more discernable. As patients' words increasingly reflect the operation of psychodynamic factors, inferences about these factors that are based on what patients say become more productive.

In general, psychodynamic psychotherapists do not set agendas for therapy sessions or direct the flow of the dialogue but try to discover and bring to light their patients' underlying emotions, needs, wishes, fears, inner conflicts, core beliefs, and self-protective processes, as inferred from what they observe in the patients and in themselves. Their primary charge is to offer comments and questions that are likely to expand their patients' self-awareness, directly or indirectly. Comments or questions that can expand patients' self-awareness indirectly include those that identify ways in which patients unconsciously constrain or sabotage the expansion of their self-awareness—usually to protect themselves from emotions that are associated with something of which they (unconsciously) fear to become aware.

Whereas the aspects of a person's self that can become integrated in treatment extend over many years of history, the process of integration occurs in the present and operates with present experience, whatever subject matter is being discussed. The closer the subject matter under discussion is to present experience, the more active and salient the associated emotions are, the more available are underlying psychodynamics to conscious exploration, and the more likely is integration to be potentiated by the discussion. Consequently, psychotherapeutic dialogue tends to be most effective in promoting integration when it addresses issues as they arise in the here-and-now. In the psychotherapeutic situation, the most here-and-now issues are those that concern the relationship between the patient and the therapist. Hence, these issues often are the most productive to address.

Early in life, people begin to develop inner models of self-and-other-in-relationship that organize how they perceive, experience, and respond to other people and also how they perceive and experience themselves in relation to others. These models or mental representations are more than audiovisual recordings. They encompass the somatic and behavioral residues of the complex emotions, needs, wishes, fears, inner conflicts, core beliefs, and self-protective processes that individuals experienced during their interactions with other people or developed in response to such interactions. Their experiences with others in the first years of their lives have the most profound and enduring influence on their models for two reasons that probably are related. Humans are helpless at birth and during their early years depend on their caregivers for their survival. Since what they learn about their caregivers and how to respond to them is literally a matter of life or death, natural selection has favored those who learned quickly and held on tightly to what they learned. Meanwhile, humans are born with central nervous systems that are only partially developed. Most of the remaining development occurs epigenetically during the first two years of life and is shaped by the young humans' interactions with their environment, which during those early years consists primarily of their caregivers. The developmental plasticity of human central nervous systems enables them to be much more specifically adapted to each individual's particular environment than would anything that the genetic code could engineer by itself. As a corollary, the people who constitute a person's early environment have a profound and enduring influence on how the central nervous system develops and subsequently functions.

People's inner models of self-and-other-in-relationship are usually of great value in their early relationships with their primary caregivers, on whom they depend for their survival. Their inner models guide them to act in ways that increase the likelihood that their needs will be met, decrease the likelihood that they will experience traumatic emotional-somatic pain, and in general enhance the chances that they will survive. As people's interpersonal world expands, their inner models continue to be adaptive to the extent that (1) the individuals with whom they are now interacting are similar to those who were around when the models initially developed, and (2) the models are sensitive to the differences between past and present people and can evolve to accommodate to the latter.

Sometimes, however, people’s models of self-and-other-in-relationship are not flexible enough to accommodate the differences between the caregivers whose interactions with them shaped the models and the individuals with whom they now interact—particularly when the models are suffused with anxiety about threats to survival. Under these conditions, people’s inner models may guide them to perceive, experience, and respond to significant individuals in ways that are maladaptive—that cause them emotional distress or impair their social or occupational functioning. Often, such emotional distress and functional impairment are what lead people to seek help.

In psychodynamic psychotherapy, the relationship between the patient and therapist is the stage in which the patient’s inner models reveal themselves. As patient and therapist spend time together, interacting at a level more intimate than ordinary discourse, the emotional significance of their relationship deepens, and the patient’s inner models of self-and-other-in-relationship become increasingly activated. A process develops through which the emotions, needs, wishes, fears, inner conflicts, core beliefs, and self-protective processes that are associated with these inner models and that unconsciously influence the patient’s other significant relationships increasingly affect how the patient perceives, experiences, and responds to the therapist.

The therapeutic relationship, however, differs from other close relationships in a crucial respect. To the extent that their own personal vulnerabilities permit, psychodynamic psychotherapists refrain from reacting with defensiveness, retaliation, withdrawal, sadism, or exploitation of their patients’ words and actions. They recognize that patients’ responses to them always reflect some degree of accurate perception of the therapist and some degree of distortion or disproportion due to the activation of patients’ inner models that are incongruous with the current situation. To the extent that patients’ inner models have participated, opportunities become available for both patient and therapist to learn how those models operate. The therapeutic relationship then provides a safe environment in which the two of them together can explore in real-time the meaning of the patient's responses in terms of their emotions, needs, wishes, fears, inner conflicts, core beliefs, and means of self-protection. This collaborative exploration gives words to the patient's subjective, not-quite-conscious experience of these processes while they still are fresh, thus expanding the patient's awareness of them and promoting their integration into the patient's living narrative. Bringing together in consciousness the various disintegrated aspects of a person's psychic processes gradually heals the personality and enables it increasingly to function as an integrated whole. The net result is a more conscious choice, more flexibility, and more freedom.

The interaction between a patient's inner models and their relationship with the therapist is more complex than was its description in the preceding paragraphs. The activation of the patient's inner models that manifests as distorted or disproportionate responses not only provides information and opportunities to expand awareness. It also opens the inner models to modification through the patient–therapist relationship. This process is an aspect of the third component of psychodynamic psychotherapy, internalization, which is discussed in the next section.

INTERNALIZATION. Internalization is the least understood component of psychodynamic psychotherapy. Originally it is the process whereby young children gradually become able to provide for themselves the regulatory functions that had been provided for them by their caregivers. The development of self-regulatory functions occurs gradually over the course of numerous child–caregiver interactions in which the caregiver responds more or less appropriately to the child's needs for nurturing, soothing, and holding. What is internalized over the course of these interactions is not only self-regulatory functions but moreover a relationship in which a caregiver provides regulatory functions for another who needs regulation. The term *internalization* is a metaphor that represents a combination of procedural learning and classical conditioning that interacts dialectically with people's genetically guided (and epigenetically adjusted) maturation. The capacity for developmental internalization of self-regulatory functions is greatest during early childhood but it continues throughout life.

The function of internalization in psychodynamic psychotherapy derives from its role in the development of self-regulation. A primary means by which psychotherapy promotes the healing of impaired self-regulation is by providing new opportunities for people to internalize self-regulatory functions that they did not adequately internalize during their childhood. Development of affect regulation functions can be facilitated by internalization of therapists' provision of a holding environment, their soothing and enlivening functions, and their treatment of emotions as indications of inner dynamics that deserve attention. Development of self-care functions can be facilitated by internalization of the therapist's nurturing and protective functions, accompanied by internalization of the message that the patient is valued as a person and believed to be worth taking care of. Development of healthy self-governance functions can be facilitated by internalization of the therapist's integrity, respectfulness, nonjudgmental acceptance, and general stabilizing function. Development of self-mindfulness—a central factor in affect regulation, self-care, and self-governance—can be facilitated by internalization of the therapist's actively receptive attention to the patient's words, actions, and empathically discernable emotions. Raw material for internalization is provided not only by the therapists' relationships with their patients but also by how they manage or respond to everything else, including their own affects, intrusions of external circumstances into a session, minor mishaps, medical problems, and personal crises. Worthy of note is that patients' perception of potentially internalizeable material occurs through empathy as well as through the five primary senses.

Psychodynamic psychotherapy is not the only context in which self-regulatory functions can be internalized. Internalization can occur in any significant relationship and also in some group situations. What distinguishes psychodynamic psychotherapy is (1) the depth of the relationship; (2) the therapeutic frame of a dyadic relationship with a person who combines caregiving attributes and authority attributes; (3) the therapist's ability to facilitate the patient's personality integration and thus to minimize the extent to which the patient's projections distort the internalization process; and (4) (as patients have a right to expect) the therapist's level of mental-emotional health.

In psychodynamic psychotherapy with people who suffer from addictive disorders, internalization can be facilitated by these individuals' tendencies to experience their self-regulatory functions as being located in the external world and to seek them there. The very process that drives their maladaptive externalizing behavior may thus be an important facilitator of therapeutic internalization in psychotherapy.

After a detailed discussion of the neurobiologic development of the addictive process, a section on treatment that devotes much more space to talk therapies than to pharmacology may seem to be a non sequitur. However, psychodynamic psychotherapy, DBT, and behavior mastery training are neurobiologic as well as psychological processes. Every psychological event is also a neurobiologic event—a unitary event that can be known and described in either of two ways, psychological/subjective or neurobiologic/objective. An early study reported that fluoxetine and CBT were almost identical in their effects on both symptoms and caudate glucose metabolism in OCD patients. Advances in neuroimaging technology have enabled more recent studies to report more comprehensive and detailed findings on the neural correlates of psychotherapy and of changes that occur in the course of psychotherapy.

Meditation and Yoga. A few decades ago, meditation and yoga were generally regarded as culturally alien fringe activities. Since then, they have become almost mainstream. More importantly, they have been scientifically investigated for potentially beneficial effects on the symptoms of a variety of psychiatric and nonpsychiatric medical conditions. Thus far, rigorous research has not been conducted on the effects of meditation or yoga on sex addiction. However, their well-documented effectiveness as adjuncts in the treatment of other addictive disorders and related psychiatric conditions suggests that they would likely be similarly effective as adjuncts in the treatment of sex addiction. Similar to psychodynamic psychotherapy, the primary goal of yoga and of the meditation described below is self-awareness. All three modalities also share a set of accompanying values: mindfulness, acceptance (of self, of what is), integration, emotional balance, and resilience.

MEDITATION. Published studies have reported meditation to be associated with decreased use of alcohol, cocaine, cannabis, heroin, and tobacco; reduced craving and withdrawal symptoms; prevention of substance use disorders; and decreased binge-eating and emotional eating. It has been reported to be an effective adjunct also in the treatment of depression, anxiety disorders, PTSD, and insomnia. Meditation has been found to correlate with improved self-esteem and increased scores on self-related character scales, high scores on which are associated with a lower risk of personality disorder. It has been reported to improve self-regulation over the life span and especially to improve affect regulation. Research findings indicate that it also enhances executive function, insightful problem-solving, cognitive flexibility, and attentional processing. While meditation is most often practiced individually, one study found that group meditation was more effective than individual meditation in increasing mindfulness and reducing cognitive-behavioral avoidance and impulsivity.

Neuroimaging studies have revealed that meditation improves activation and connectivity in brain areas that are related to self-regulation. It has been reported to increase activation of the ventrolateral PFC, functional connectivity between the PFC and the amygdala, and fractional anisotropy in white matter pathways that connect the ACC to other brain areas. Long-term meditators were found to have larger hippocampal volumes than did nonmeditators, particularly in the subiculum, which plays a significant role in stress regulation. In beginning meditators, meditation was found to induce down-regulation of the left amygdala during emotional processing, which indicates a “top-down” form of affect regulation. In experienced meditators, it was found to induce deactivation of default mode network areas (DMN; medial prefrontal and posterior cingulate cortices, precuneus, medial temporal and inferior parietal lobes) and to weaken functional connectivity between them, but not to influence responses in brain regions that are involved in emotional reactivity during emotional processing—which indicates a “bottom-up” form of affect regulation. Meditation has been reported to stimulate GABA release by the thalamic reticular nucleus and thereby promote GABAergic cortical inhibition, which has been implicated in improved cognitive performance and enhanced emotional regulation.

Meditation is practiced in a variety of forms. The majority involve an intentional focus of attention—for example, one’s own breathing, one’s own movements (e.g., walking meditation), a word or idea, a sound of internal or external origin, or a particular sensation. Patients who have practiced some form of meditation and found it to be helpful are encouraged to continue or resume it. Others can be taught the form of meditation that the Buddha practiced: *Anapana-sati* Yoga, the yoga of breath awareness. Practitioners begin by finding a comfortable upright seated position, usually with their legs crossed. Once settled, they try not to move at all during the entire meditation period (usually 20 to 60 minutes) other than to breathe. The suggested focus of attention is the sensation of their own breath coming in and going out through their noses—more specifically, what the skin between their upper lip and nostrils feels as each inhale or exhale brushes across it. Thoughts and emotions are not banished. When they arise, meditators notice but do not name or describe them, in a manner analogous to noticing clouds that pass across the sky while one’s attention is focused on something closer. When attention wanders to such thoughts or emotions, meditators gently bring it back to their breath. This form of meditation is suggested here because it is simple, direct, and unencumbered. Moreover, it facilitates an experience in which the mind can let go of naming or holding on to what is noticed and remain in the immediacy of the moment.

YOGA. Research findings that concern the effects of yoga on psychiatric disorders and mental functioning are remarkably similar to those that concern the effects of meditation. Yoga has been found to decrease the use of alcohol, tobacco, and cannabis; reduce the risk of developing a substance use disorder; reduce the use of opioids for chronic pain; decrease binge-eating; and decrease eating disorder symptoms in a mixed cohort with diagnoses of anorexia, bulimia, or eating disorder NOS. A review article concluded that, as an adjunct in the treatment of eating disorders, yoga can be an effective method for increasing self-awareness, reflection, and the ability to self-soothe. It has been reported also to be an effective adjunct in the treatment of anxiety, depression, obsessive-compulsive disorder (OCD), PTSD, and ADHD, and to improve quality of life in schizophrenia. Yoga practice has been reported to be associated with increased plasma levels of oxytocin in schizophrenia patients, larger hippocampal volumes, and increased serum levels of BDNF.

That physical exercise in itself is beneficial raises a question about whether the benefits of yoga are due to a particular characteristic of yoga or more generally to the exercise that it provides. One study found that 12 weeks of yoga was associated with greater improvements in mood and anxiety than was a metabolically matched walking exercise. In the yoga group, improved mood and decreased anxiety were positively correlated with thalamic GABA levels. Another study reported that following 8 weeks of yoga practice, participants in the yoga intervention group showed significantly greater improvement in performance on the executive function measures of working memory capacity and efficiency of mental set shifting and flexibility compared with their stretching-strengthening counterparts.

Yoga as a whole consists of physical yoga, the active or intentional component, and mental yoga, the body of yogic knowledge and principles. The latter includes philosophy, ethics, science, and spirituality. Physical yoga includes the poses (*asanas*) that are yoga’s most visible manifestation. In a yoga practice, movements into and out of poses are guided by the breath–each is initiated with either an inhalation or an exhalation. The sequence of poses and movements promotes relaxation, inner integration, and receptivity as well as improving flexibility, balance, and strength. The developmental goal of a yoga practice is to increase the practitioner's bodily self-awareness, breath awareness, and sense of body–mind integration. Its immediate goal is to bring the practitioner to a body-mind state that is conducive to meditation. (The origin of meditation is often ascribed to Buddhism. However, Gautama Buddha was a yogi, not a Buddhist. Prior to his awakening [*bodhi*], he had practiced intensively for years under two consecutive hermit teachers of yogic meditation.)

The form of meditation that was described above can be learned and practiced without much instruction. Some people might find even the foregoing brief description of the process sufficient as a starting point. Learning and practicing yoga require considerably more instruction. Classes and individual teaching are widely available. In addition, considerable information about yoga can be obtained from books or online, and audio–visual instruction is available online and via DVD.

EFFECTS ON THE ADDICTIVE PROCESS. The documented effectiveness of meditation and (separately) yoga in decreasing the use of a variety of psychoactive substances and the frequency of eating disorder symptoms (primarily binge eating) suggests that both of these adjunctive treatment modalities are effecting positive change at the level of the addictive process. Reported changes that address addictive process-related impairments include improved self-regulation (meditation), increased self-awareness and ability to self-soothe (yoga), and improved affect regulation (both). Both meditation and yoga also have been associated with improved executive function, which would support healing of impairments in self-care and self-governance. The beneficial effect of yoga as an adjunct to the treatment of ADHD and OCD suggests that it may improve behavioral inhibition.

Development of a Healthier Lifestyle. A lifestyle is constituted by a profusion of elements, each of which varies along a number of dimensions. Among the more prominent elements are nutrition, sleep, exercise, love relationships, social relationships, work (activity that is motivated by anticipation of its wanted consequences), play (activity that is motivated by enjoyment or pleasure that is inherent to the activity), and creative activity. Many would add spirituality. To illustrate how the term *dimension* is being used here, examples in the area of nutrition include caloric intake, protein intake, frequency of meals, and eliminated foods such as gluten or (for vegetarians and vegans) meat. The most relevant dimensions and the optimal values for each dimension vary from person to person. Complex lifestyle elements concern balances between simple elements, such as the balance between work and play or between social time and solitary time. What constitutes being in balance also tends to vary from one person to another.

People can begin to work on reducing stress in their lifestyles by scanning a span of time for activities, situations, and people that are associated with increased levels of stress. Once a “stressogen” has been identified, they can decide whether to eliminate it, modify it, or leave it as is (when according to their own calculations its benefits outweigh its costs).

Sex addicts’ symptomatic behavior has continued because it was reinforced by its consequences, both gratification and alleviation or evasion of painful affects. Their ongoing capacity to refrain from such behavior is enhanced when their lifestyles include nonharmful activities that produce similarly reinforcing consequences or strongly reinforced activities that preclude symptomatic behavior (or both).

Developing a healthier lifestyle requires more than information about changes that would promote physical or mental-emotional health. It requires deciding to act, taking action, and continuing the action. When this sequence breaks down, the interference may have originated in either maladaptive thoughts and ways of thinking or emotional dynamics that were stirred up by something related to the potential change. Under such circumstances, a brief cognitive-behavioral or psychodynamic psychotherapeutic intervention (respectively) may help patients resume their paths toward more health and balance in their lives.

Prognosis

To date, no outcome studies have been published that evaluate the effectiveness of integrated treatment for sex addiction. Empirical research on almost every aspect of sex addiction is lacking: neurobiology, psychometrics, family history, diagnostic criteria (reliability, coverage, and predictive validity), course of illness, and response to treatments. This deficit may have been due to the unavailability until recently of clear and meaningful diagnostic criteria for sex addiction. It also may have been due to reluctance by many to consider sex addiction as a fit subject for scientific study. Perhaps this chapter and similar publications will stimulate the empirical research that this field so desperately needs.

Although the long-term response of sex addiction to various types of treatment cannot at this time be reliably predicted, a number of factors have been identified that seem to influence the prognosis. These factors can be clustered in three groups: (1) illness factors (that tend to worsen the prognosis), which include early age of onset, high frequency of symptomatic sexual behavior, concomitant use of alcohol or other drugs, and absence of anxiety or guilt about the behavior; (2) healing support factors (that tend to improve the prognosis), which include a stable job, a stable primary relationship, a supportive social network, and availability of appropriate sexual outlets; and (3) personality factors (that tend to improve the prognosis), which include intelligence, creativity, self-observatory capacity, sense of humor, capacity to form interpersonal connections, and motivation for change—which is a function of people's ability to experience themselves as responsible as well as their degree of subjective distress. Prognosis is also influenced by comorbidity with other psychiatric disorders, by the degree of associated character pathology, and by the kind of relationship that develops between an addict and their therapist (or whoever else is most involved in the addict's treatment).

In general, treatment for sex addiction is likely to be most effective when the number of different clinicians (and hence opportunities for splitting) is kept to a minimum. When the psychodynamic psychotherapist is also the clinician who conducts DBT or symptomatic behavior management and (when indicated) prescribes psychiatric medication, patients' psychodynamically based (conscious or unconscious) undermining of the latter modalities is more likely to be effectively managed. At another level, assigning different therapeutic modalities to different clinicians is inconsistent with the integrative intent of the treatment.

Significant healing from sex addiction, as from any other addictive disorder, takes time. Addiction interferes with normal development and corrodes interpersonal relationships. The resulting problems remain after symptomatic behavior has ceased and only then can they begin to be effectively addressed. More fundamentally, as described in the preceding section, the principal constituent of healing from an addictive disorder is healing of the addictive process. The initial phase of this healing, which in addition to the overall goals of treatment focuses specifically on affect regulation and stabilization of sense of self, is estimated to require 3 to 5 years.

Healing is not cure. Sex addiction—like other addictive disorders, most psychiatric conditions, and the majority of diseases in medicine—is not an acute disease that effective treatment can cure but a chronic disease that effective treatment can, with patients' active collaboration, bring into remission. And remission in chronic disease is a provisional condition, the continuation of which depends on patients' maintenance of a healthy lifestyle. However much addicts (of any behavioral type) may have benefited from treatment, most will to some extent remain vulnerable to being overwhelmed by intense affects and loss of self-coherence in situations that would be neither traumatic nor disorganizing for individuals whose self-regulation systems are intact. Sex addicts are particularly likely to experience threats of being overwhelmed by these internal states as urges to engage in sexual behavior. Even after stable remission has been achieved, ongoing health-promoting activity may be required indefinitely to prevent recurrence of symptomatic behavior and preserve the progress that has been made. Healing does not make addicts' underlying pain disappear. It helps them become increasingly able to accept and manage their pain while freeing them to more fully and joyfully live their lives.

ACKNOWLEDGMENT

This chapter represents an update and revision of work done by Dr. Aviel Goodman. We are indebted to him for his work and retain much of his prose in this chapter.

Further Readings

Champagne DL, Bagot RC, van Hasselt F, et al. Maternal care and hippocampal plasticity: evidence for experience-dependent structural plasticity, altered synaptic functioning, and differential responsiveness to glucocorticoids and stress. *J Neurosci*. 2008;28:6037.

Coen SJ. Sexualization as a predominant mode of defense. *J Am Psychoanal Assoc*. 1981;29:893.

Feldstein Ewing SW, Chung T. Neuroimaging mechanisms of change in psychotherapy for addictive behaviors: emerging translational approaches that bridge biology and behavior. *Psychol Addict Behav*. 2013;27:329.

Frascella J, Potenza MN, Brown LL, Childress AR. Shared brain vulnerabilities open the way for nonsubstance addictions: carving addiction at a new joint? *Ann N Y Acad Sci*. 2010;1187:294–315.

Garcia FD, Thibaut F. Sexual addictions. *Am J Drug Alcohol Abuse*. 2010;36:254–260.

*Goodman A. Organic unity theory: an integrative mind-body theory for psychiatry. *Theor Med*. 1997;18:357.

*Goodman A. Neurobiology of addiction: an integrative review. *Biochem Pharmacol*. 2008;75:266.

*Goodman A. *Sexual Addiction: An Integrated Approach*. International Universities Press; 1998.

*Grant JE, Potenza MN, Weinstein A, Gorelick DA. Introduction to behavioral addictions. *Am J Drug Alcohol Abuse*. 2010;36:233.

Guay DR. Drug treatment of paraphilic and nonparaphilic sexual disorders. *Clin Ther*. 2009;31:1.

Gutman DA, Nemeroff CB. Persistent central nervous system effects of an adverse early environment: clinical and preclinical studies. *Physiol Behav*. 2003;79:471.

Harned MS, Banawan SF, Lynch TR. Dialectical behavior therapy: an emotion-focused treatment for borderline personality disorder. *J Contemp Psychother*. 2006;36:67.

Hyman SE, Malenka RC, Nestler EJ. Neural mechanisms of addiction: the role of reward-related learning and memory. *Ann Rev Neurosci*. 2006;29:565.

Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *Am J Psychiatry*. 2003;160:687.

Khan MMR. *Alienation in Perversions*. Hogarth; 1979.

Khantzian EJ. Self-regulation and self-medication factors in alcoholism and the addictions: similarities and differences. *Rec Dev Alcohol*. 1990;8:255.

Koob GF, LeMoal M. *Neurobiology of Addiction*. Academic Press; 2006.

Marchand WR. Neural mechanisms of mindfulness and meditation: evidence from neuroimaging studies. *World J Radiol*. 2014;6:471.

Messina I, Sambin M, Palmieri A, Viviani R. Neural correlates of psychotherapy in anxiety and depression: a meta-analysis. *PLoS One*. 2013;8:e74657.

Moffett MC, Vicentic A, Kozel M, Plotsky P, Francis DD. Maternal separation alters drug intake patterns in adulthood in rats. *Biochem Pharmacol*. 2007;73:321.

*Myers W. Sexual addiction. In: Dowling S, ed. *The Psychology and Treatment of Addictive Behavior*. International Universities Press; 1995.

Nestler EJ. Is there a common molecular pathway for addiction? *Nature Neurosci*. 2005;8:1445.

Reid RC, Carpenter BN, Spackman M, Willes DL. Alexithymia, emotional instability, and vulnerability to stress proneness in patients seeking help for hypersexual behavior. *J Sex Marital Ther*. 2008;34:133.

Scharfman MA, Vogel SA. Current concepts of the development of sexuality. *J Amer Psychoanal Assn*. 1989;37:787.

Skowronek IB, Mounsey A, Handler L. Clinical inquiry: can yoga reduce symptoms of anxiety and depression? *J Fam Pract*. 2014;63:398.

Stein DJ, Hollander E, Anthony DT, Schneier FR, Fallon BA. Serotonergic medications for sexual obsessions, sexual addictions, and paraphilias. *J Clin Psychiatry*. 1992;53:267.

Volkow ND, Fowler JS, Wang G-J. The addicted human brain viewed in the light of imaging studies: brain circuits and treatment strategies. *Neuropharmacology*. 2004;47[Suppl 1]:3.

Volkow ND, Wise RA. How can drug addiction help us understand obesity? *Nature Neurosci*. 2005;8:555.

Weaver IGC, Cervoni N, Champagne FA, D’Alessio AC, Sharma S. Epigenetic programming by maternal behavior. *Nature Neurosci*. 2004;7:847.

Zhang TY, Labonté B, Wen XL, Turecki G, Meaney MJ. Epigenetic mechanisms for the early environmental regulation of hippocampal glucocorticoid receptor gene expression in rodents and humans. *Neuropsychopharmacology*. 2013;38:111.

22

Feeding and Eating Disorders

CHRISTINE C. CALL, PH.D., EVELYN ATTIA, M.D., AND B. TIMOTHY WALSH, M.D.

As their name suggests, feeding and eating disorders are characterized by persistent disturbances in feeding and eating behaviors that significantly interfere with the afflicted individual's life, often by negatively and severely impacting physical health or medical status, impairing psychosocial functioning, and causing intense distress. The psychopathology defining these illnesses manifests most saliently in disordered behaviors, but disturbances in thoughts, attitudes, and emotions are also prominent. Because of the medical complications inherent to feeding and eating pathology, these illnesses are serious and often dangerous, making early identification and intervention particularly important, although, unfortunately, it is not uncommon for them to go unrecognized and untreated.

While the syndromes described within the feeding and eating disorders category center around disordered feeding and eating behaviors, they are also each sufficiently different in etiology, clinical presentation, course, prognosis, and treatment needs to warrant distinct diagnostic criteria. The "Feeding and Eating Disorders" chapter of the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (DSM-5) and its text revision (DSM-5-TR) includes six diagnoses: anorexia nervosa, bulimia nervosa, binge eating disorder (BED), avoidant/restrictive food intake disorder (ARFID), pica, and rumination disorder. The chapter also includes two broad heterogeneous categories meant to capture presentations that are consistent with, but do not meet full criteria for, a formally recognized feeding and eating disorder: other specified feeding and eating disorders (OSFED) and unspecified feeding and eating disorders (USFED). Feeding and eating disorders appeared in one comprehensive section for the first time in DSM-5, as they were previously described in separate sections. Feeding disorders have received significantly less research attention than eating disorders, although since the merging of the two sections in DSM-5, feeding disorders (in particular ARFID) have received increased research attention. Because of the limited data available, this chapter will not discuss aspects of feeding disorders that remain poorly understood, including etiology and course.

DEFINITIONS

The definitions of feeding and eating disorders provided below are meant to introduce the reader to each syndrome, but diagnostic criteria and clinical characteristics are discussed in detail in the "Diagnosis and Clinical Features" section of this chapter.

Anorexia Nervosa

Anorexia nervosa is marked by significantly low weight based on developmental trajectory and height. Individuals with this disorder exhibit a persistent and relentless drive for thinness, leading to behaviors that cause weight loss or (particularly in later phases of the illness and for adolescents who are still growing) prevent weight gain. Most prominently, individuals with this illness significantly restrict their food intake, but may also engage in other behaviors that impact weight, including purging (e.g., self-induced vomiting, laxative, diuretic, or enema use), and intense and often compulsive exercise. Individuals with anorexia nervosa demonstrate a strong fear of gaining weight or becoming fat and typically describe body image distortions, believing that they are fat or large even at a low body weight. Weight and shape tend to become essential in self-evaluation, taking over the majority of the individual's self-schema. In addition, individuals with this illness often fail to recognize the medical implications of their current weight status.

Bulimia Nervosa

While sharing certain core features with anorexia nervosa, bulimia nervosa is defined by recurrent episodes of binge eating followed by compensatory behaviors, which might include self-induced vomiting, laxative, diuretic, or enema use, intense exercise, fasting, and medication misuse. Central to the conceptualization of bulimia nervosa is the definition of binge eating, a term that refers to the consumption of an unusually large amount of food accompanied by a sense of lack of control, or the feeling that one cannot stop eating. Individuals with bulimia nervosa maintain their weight at or above normal range for age and height but, similar to those with anorexia nervosa, exhibit intense preoccupation with shape or weight. They often rely almost entirely on these factors in self-evaluation, which typically leads to harsh self-judgment.

Binge Eating Disorder

Individuals with BED, like those with bulimia nervosa, experience repeated episodes of binge eating. However, different from the behavioral pattern of bulimia nervosa, these episodes are *not* followed by compensatory behaviors. While individuals with BED may, and often do, attempt some level of dietary restriction outside of binge eating, they do not engage in inappropriate behaviors meant specifically to compensate for an episode of overeating, like purging or fasting. Individuals with this disorder also maintain their weight at or above the normal range for age and height (and, often, weight is in the overweight or obese category).

Avoidant/Restrictive Food Intake Disorder

ARFID is characterized by persistent disturbances in eating behavior, leading to clinically significant consequences (e.g., nutritional deficiencies, psychosocial impairment, etc.). The restrictive eating that marks this disorder may arise for a variety of reasons (e.g., concern or disgust with texture/consistency/smell of specific food items, anxiety around eating due to a history of choking, vomiting, or a pattern of extremely stressful mealtimes). Notably, shape or weight concern must not motivate the restrictive behaviors as it does in anorexia nervosa and bulimia nervosa. Likewise, a diagnosis of ARFID should not be made if the relevant behaviors occur solely in the context of a separate disorder (e.g., a neurodevelopmental disorder), unless the behaviors are severe enough to warrant specific clinical attention. Finally, feeding or eating behaviors that are considered culturally or socially appropriate should not be considered in making a diagnosis of ARFID.

Pica

Individuals with pica repeatedly consume nonfood items (e.g., paper). As with ARFID, this disturbance must occur outside of the context of culturally sanctioned practices or pre-existing medical or psychiatric conditions that would cause this behavior (e.g., a neurodevelopmental disorder) unless the behavior is so impairing that it requires additional, specific clinical attention.

Rumination Disorder

Individuals with rumination disorder repeatedly regurgitate food and then spit it out, rechew, or reswallow it. Rumination disorder cannot be diagnosed in the presence of anorexia nervosa, bulimia nervosa, BED, or ARFID, although the behavior of rumination often occurs in the context of these syndromes. Likewise, individuals with other medical or psychiatric conditions should not receive a diagnosis of rumination disorder unless the behaviors cause substantial impairment sufficient to necessitate specialized intervention.

Other Specified and Unspecified Feeding and Eating Disorders

Finally, DSM-5 includes two broad, heterogeneous categories meant to capture clinically significant feeding or eating disturbances that do not clearly fall within one of the diagnostic categories outlined above. OSFED provides descriptions of disorders that are not yet formally recognized due to insufficient research describing and characterizing them, or those that are similar to the formal eating disorder categories without meeting the frequency or duration criteria needed for these diagnoses. For example, purging disorder, in which an individual engages in repeated purging behavior without the binge episodes that mark bulimia nervosa, is briefly outlined within the category OSFED. Also outlined within this category are atypical anorexia nervosa, bulimia nervosa of low frequency and/or limited duration, BED of low frequency and/or limited duration, and night eating syndrome. USFED, on the other hand, exists to capture individuals whose behaviors and symptoms are not adequately described by a formal feeding or eating disorder, or by the examples provided in OSFED.

History

Although the specific diagnostic criteria describing feeding and eating disorders continue to shift over time as research findings help guide the current understanding of these illnesses, historical accounts of disordered eating date back as far as the 4th century. Some of these early descriptions appear remarkably similar to the currently accepted definitions of feeding and eating disorders. Still, several of the syndromes now classified as feeding and eating disorders were not officially recognized until much more recently, with BED, only formally accepted as an eating disorder in DSM-5 in 2013, being an excellent example.

Anorexia Nervosa

Anorexia nervosa is the oldest recognized feeding or eating disorder, perhaps because the characteristic low weight that defines the illness has made it more easily discernible over history. The disorder was first formally characterized and named in the 1860s and 1870s by Sir William Gull in London, who coined the name anorexia nervosa, and Dr. Charles Laségue in France, who referred to it as anorexia hysterique. Both physicians described cases of individuals presenting at markedly low body weights and engaging in self-starvation behaviors.

Although Sir William Gull and Dr. Charles Laségue were the first to clearly document reports of the illness, instances of anorexia nervosa likely occurred well before the mid-1800s. In

1689, Richard Morton described a young woman with amenorrhea and significant weight loss, which he ascribed to her mental state. Evidence of self-imposed starvation, often for religious purposes but nonetheless quite severe, dates back to the fourth century, when the gnostic sect of Christian men, believing that material objects including food were evil, engaged in strict and intense fasting practices. Similarly, the term "Holy Anorexia" refers to the practices of numerous women during the Middle Ages (many canonized by the Catholic Church as saints in the 14th and 15th centuries) who strove for holiness by denying themselves food, the self-starvation seeming to become self-sustaining and often resulting in premature death. While it is unclear whether the motives behind these accounts of self-starvation were at all related to body image or feelings about shape or weight, it is nonetheless interesting to note that the most prominent behavior in the current conceptualization of anorexia nervosa, severe self-imposed food restriction, has occurred throughout history.

Since Sir William Gull and Dr. Charles Laségue presented the first clear clinical descriptions of the syndrome, the search to understand this pernicious illness has taken a variety of turns. In the late 1800s, anorexia nervosa was temporarily confused with postpartum pituitary necrosis, leaving a lasting, even if disproved, assumption that some primary abnormality of endocrine function was implicated in the illness's origin. At the beginning of the 20th century, early psychoanalytic hypotheses proposed that the illness might develop from fears of oral impregnation and related themes. Starting in the 1950s, Hilda Bruch formulated a more contemporary psychodynamically oriented hypothesis. Finally, in the 1960s, researchers including Gerald Russell, Arthur Crisp, and Pierre Beaumont initiated the first empirical studies of anorexia nervosa, launching the modern-day understanding of the syndrome.

As research on anorexia nervosa has grown, modifications to diagnostic criteria have been made to better capture the illness. Most recently, DSM-5 featured minor clarifications and modifications to the diagnostic criteria. Perhaps most notably, the DSM-IV criterion that requires an individual to exhibit amenorrhea for at least 3 months to qualify for a diagnosis of anorexia nervosa was removed in DSM-5. The current wording now allows women who report regular or irregular menstrual activity to receive a diagnosis. This change came after several research studies found that women who met all other diagnostic criteria for anorexia nervosa but continued to menstruate regularly did not differ significantly from women meeting full DSM-IV criteria for the syndrome in terms of course, prognosis, or illness severity.

In addition to the deletion of the amenorrhea criterion, DSM-5 and DSM-5-TR feature a clarification to the language describing weight status in anorexia nervosa. This small change aims to eliminate the misunderstanding that "low weight" is defined by one particular body mass index (BMI) or percent ideal body weight (IBW) cutoff. Because assessing weight status is crucial to making a diagnosis of anorexia nervosa, DSM-IV provided 85% of IBW as an example of low weight within the diagnostic criteria for anorexia nervosa. Unfortunately, this specific number was often interpreted as the definition rather than an example of low weight, leading individuals who may have qualified for a diagnosis of anorexia nervosa but presented at greater than 85% of IBW to instead receive a diagnosis of EDNOS. DSM-5 has therefore removed the reference to 85% IBW and has added additional information to the text about determining whether the individual's weight is significantly low by taking into consideration factors including age, developmental trajectory (i.e., where a child or adolescent falls on their individual growth curve), weight history, and medical complications.

Finally, the diagnostic criterion describing fear of weight gain was modified so that individuals who do not endorse or articulate this facet of the illness, but nonetheless demonstrate persistent behaviors that interfere with weight gain or maintenance, may receive a diagnosis of anorexia nervosa. This clarification is meant to capture children and adolescents who may not express this fear due to their age and developmental status, or individuals who, for other reasons (including cultural differences), deny fat-phobia.

Bulimia Nervosa

Bulimia nervosa was formally delineated much later than anorexia nervosa, but it, too, was likely present well before its first official documentation. Although historical accounts of the behaviors associated with bulimia nervosa are not nearly as common as those of the self-starvation inherent to anorexia nervosa, a number of ancient medical texts refer to overeating and subsequent purging. Moreover, given the secrecy that surrounds the syndrome and the lack of easily observable physical indications of the illness (as in anorexia nervosa), it would not be surprising if cases went unnoticed for quite some time before the disorder was formally recognized. Even today many cases of bulimia nervosa go unrecognized because the afflicted individuals maintain normal weight throughout the duration of the disturbance and often hide their disordered behaviors out of shame.

In 1979, Gerald Russell gave a name to the behavioral pattern that was being reported with increasing frequency at the time—episodes of binge eating followed by inappropriate compensatory behaviors. At first, bulimia nervosa, as he named it, was viewed as a variant of anorexia nervosa, but it was quickly recognized that, although the disorders overlap in certain ways, the two are distinct. In 1980, bulimia nervosa was included alongside anorexia nervosa in DSM-III.

Since its formal inclusion in DSM-III, diagnostic criteria for bulimia nervosa have remained largely unchanged, although, notably, there were minor changes made in DSM-5 that aim to reassign individuals classified as EDNOS in DSM-IV into a more appropriate and descriptive category. The primary modification involves the frequency of binge eating and purging behaviors. To meet DSM-IV criteria for bulimia nervosa, individuals must engage in binge eating and inappropriate compensatory behaviors at least twice per week on average for 3 or more months. Since the publication of DSM-IV, research has shown that individuals who exhibit these behaviors once per week on average do not differ clinically from those who experience them more frequently. Thus, in DSM-5, the threshold frequency of binge eating and compensatory behaviors was reduced to once per week for a diagnosis of bulimia nervosa.

Binge Eating Disorder

Interestingly, the behavioral pattern that defines BED, recurrent episodes of binge eating without any inappropriate compensatory behaviors, was clearly characterized in 1959 by Albert Stunkard, two decades before Russell described bulimia nervosa. Despite being documented much earlier than bulimia nervosa, BED began to receive attention and was formally recognized much later. In fact, BED did not appear in the DSM until its fourth edition (DSM-IV) in 1994, when it was described in the appendix as an example of an eating disorder not otherwise specified (EDNOS). Significant research focus on the disorder over the following two decades led to its formal recognition as an eating disorder in DSM-5. From DSM-IV to DSM-5, diagnostic criteria for BED also shifted in parallel to the shifting criteria for bulimia nervosa; individuals who engage in binge eating at least once per week (instead of 2 or more times per week) for 3 or more months now meet full diagnostic criteria for BED.

Avoidant/Restrictive Food Intake Disorder, Rumination Disorder, and Pica

The predecessor to ARFID (feeding disorder of infancy or early childhood), along with rumination disorder and pica, were categorized as feeding and eating disorders of infancy and early childhood in DSM-IV. Their shift into the DSM-5 “Feeding and Eating Disorders” chapter demonstrates the growing awareness that these disorders can occur throughout the lifespan, not just in infancy and early childhood. Besides the removal of the age requirement for these disorders, the diagnostic criteria for pica and rumination disorder remain essentially unchanged from DSM-IV to DSM-5. The diagnostic criteria for ARFID, on the other hand, were substantially expanded in DSM-5 to account for evolving research. In DSM-5, an individual may receive a diagnosis of ARFID even outside the context of weight loss or growth impairment (diagnostic criteria in DSM-IV) provided that the problematic eating behaviors significantly interfere in some other way. (In DSM-5-TR, criterion A was slightly reworded to make clear that interference with psychosocial functioning is sufficient to satisfy this criterion.) DSM-5 also expanded the definition of this disorder by allowing individuals with a comorbid medical condition to receive a diagnosis of ARFID if their eating behavior and subsequent impairment warrant it. Finally, the DSM-5/DSM-5-TR criteria for ARFID recognize motivations for the restrictive eating behavior not previously delineated in DSM-IV’s definition of feeding disorder in infancy or early childhood. In addition to the reasons for food avoidance or restriction described in DSM-IV, such as staying away from foods with certain sensory characteristics or demonstrating a lack of interest in eating in general, DSM-5/DSM-5-TR describes concerns with aversive consequences of eating as another possible explanation of the behaviors observed in ARFID. This concern may stem from experiencing a negative food-related event (e.g., choking or observing someone choke while eating) or from an emotional disturbance surrounding eating (e.g., a history of particularly problematic and difficult mealtimes due to a poor relationship with the caregiver).

COMPARATIVE NOSOLOGY

Currently, two major systems are widely utilized in the diagnosis of mental disorders: The DSM, published by the American Psychiatric Association (APA) and considered the standard classification system in the United States, and the International Classification of Diseases (ICD), published by the World Health Organization (WHO) and used by a significant number of countries worldwide. The most recent edition of the DSM (DSM-5-TR) was released in 2022, following DSM-5 in 2013, and the most recent edition of the ICD (ICD-11) was presented at the World Health Assembly in 2019 and formally replaced ICD-10 starting January 1, 2022. The diagnostic criteria for feeding and eating disorders are largely consistent across DSM-5-TR and ICD-11, although there is a notable difference in the conceptualization of binge eating, as described later in this chapter. (Notably, when the United States will officially adopt the ICD-11 coding scheme is unclear. The ICD-10 codes were adopted only in 2015, more than 20 years after their approval by WHO.)

EPIDEMIOLOGY

Feeding and eating disorders cut across demographic divides, affecting males and females of all ages, races, and cultural backgrounds, although certain demographic characteristics impart higher risk for each syndrome. With the broadening of diagnostic criteria in DSM-5, more individuals fall into a formal feeding or eating disorder category than before, as a notable portion has been reassigned from the broad heterogeneous other and unspecified categories that previously accounted for over 50% of eating disorder cases. Nonetheless, feeding and eating disorders are relatively rare when compared to more common psychiatric conditions, like major depressive disorder. Although a large proportion of the population will, at some point, experience body dissatisfaction and engage in dieting behaviors, a much smaller percentage will struggle with clinically relevant problematic eating and weight-related cognitions and behaviors, and an even smaller subset will meet full criteria for a feeding or eating disorder. Recent estimates suggest that the lifetime prevalence of any eating disorder is approximately 8% in women and 2% in men. Table 22–1 provides an overview of the epidemiology of anorexia nervosa, bulimia nervosa, and BED. Data on the epidemiology of ARFID, pica, and rumination disorder are not included in Table 22–1 as these diagnoses have not yet been assessed in large-scale population-based studies. Preliminary data about their prevalence in specific subpopulations are included in the sections below.

Anorexia Nervosa

Although a substantial proportion of those diagnosed with subthreshold anorexia nervosa in DSM-IV now receives a formal diagnosis in DSM-5-TR, combined rates of subthreshold and threshold anorexia nervosa have remained relatively stable since the 1970s. Various studies have, however, documented a significant increase in incidence rates (the number of new cases in the population over a set period of time) in the high-risk group of 15- to 19-year-old females in recent years. It remains unclear if this increase signifies an earlier age of illness onset or earlier recognition and intervention. Across all age groups and genders, recent epidemiologic studies suggest that lifetime prevalence (the proportion of individuals who have the illness at any point in their life) of DSM-5 anorexia nervosa is about 1.4% in women and 0.2% in men. Among girls and women, point prevalence (prevalence at a specific point in time) of DSM-5 anorexia nervosa is approximately 0.5% to 1%. In a recent study, the 12-month prevalence estimate for anorexia nervosa was 0.05%, which is much lower than prior estimates for reasons that are still unclear.

Although epidemiologic studies indicate that it may occur approximately 10 times more frequently in females, anorexia nervosa does affect males. Recent research suggests that the illness, which continues to be perceived by many as a female disorder, may be even more common in men and boys than previously believed. It may go unrecognized in men for a variety of reasons. In particular, men may be reluctant to seek treatment out of shame, and clinicians may be less likely to recognize the syndrome in male versus female patients. Men may also present with different symptoms than women, such as a drive for both muscularity and leanness (i.e., the muscular ideal) and the inappropriate use of performance-enhancing drugs (e.g., anabolic steroids) to achieve muscularity, which may contribute to underdetection by clinicians. Additionally, while anorexia nervosa has been thought of historically as a disease of white, wealthy women, it impacts individuals of all racial and socioeconomic backgrounds.

Bulimia Nervosa

As with anorexia nervosa, rates of bulimia nervosa have increased with changes to diagnostic criteria. However, overall, studies suggest that the incidence of the illness has decreased in recent years. It is estimated that lifetime prevalence of bulimia nervosa is approximately 2% in women and 0.6% in men, and point prevalence is about 1.5% in women and 0.1% in men. 12-month prevalence was recently estimated at 0.14%, which is lower than prior estimates for reasons that are unclear. Average age of onset also seems to have decreased, although this finding may be an artifact of earlier detection. Like anorexia nervosa, the disorder is more common in women than in men.

Binge Eating Disorder

BED is the most common eating disorder and the least gender-divided of the three. Lifetime prevalence of BED has been estimated at 1% to 3% for women and 0.5% to 1.0% for men, while point prevalence is approximately 0.5% in women and roughly half of that in men. Twelve-month prevalence was recently estimated at 0.44%. Rates of BED are particularly high in individuals with BMIs in the overweight or obese range.

Table 22–1.
Epidemiologic Features of DSM-5-TR Eating Disorders

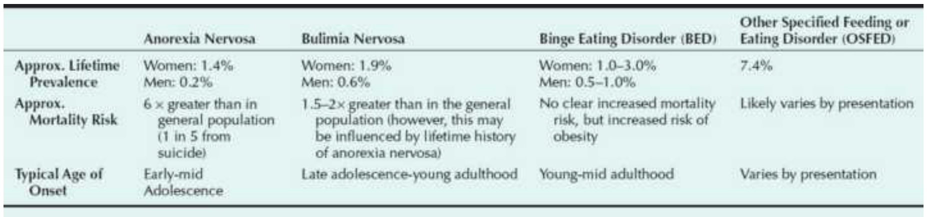

	Anorexia Nervosa	Bulimia Nervosa	Binge Eating Disorder (BED)	Other Specified Feeding or Eating Disorder (OSFED)
Approx. Lifetime Prevalence	Women: 1.4% Men: 0.2%	Women: 1.9% Men: 0.6%	Women: 1.0–3.0% Men: 0.5–1.0%	7.4%
Approx. Mortality Risk	6 × greater than in general population (1 in 5 from suicide)	1.5–2× greater than in the general population (however, this may be influenced by lifetime history of anorexia nervosa)	No clear increased mortality risk, but increased risk of obesity	Likely varies by presentation
Typical Age of Onset	Early-mid Adolescence	Late adolescence-young adulthood	Young-mid adulthood	Varies by presentation

Avoidant/Restrictive Food Intake Disorder

Less is known about the epidemiology of ARFID. Extant data suggest that prevalence rates range from 0.3% (3-month prevalence) to 3.2% (point prevalence). Within neurogastroenterology clinics, prevalence rates are higher (e.g., among adults, over 6% with a threshold ARFID diagnosis, and over 17% with subthreshold or unconfirmed ARFID diagnosis). ARFID appears to onset at an earlier age than anorexia nervosa, bulimia nervosa, or BED. Additionally, individuals with ARFID, compared to those with an eating disorder, are more likely to present with a co-occurring medical problem and to have a longer illness duration prior to presenting for treatment. Some studies have also reported that, unlike in eating disorders, ARFID occurs more frequently in males than females.

Rumination Disorder and Pica

Epidemiologic data on rumination disorder and pica are particularly limited and results are variable. Self-report measures have found point or 3-month prevalence estimates of rumination behavior ranging from less than 1% to over 10% in adults, and between approximately 1.5% and 5% in children. Prevalence estimates of pica are even more limited, with studies suggesting that this disorder is present in approximately 1.7% to 5% of children. Rumination disorder and pica appear to commonly co-occur with ARFID and with eating disorder symptoms, but additional research is needed on the epidemiologic features of these disorders.

ETIOLOGY

Our current understanding of the specific causes of feeding and eating disorders is limited. Research regarding the etiology of eating disorders suggests that a confluence of factors, including genetic, biologic, psychological, and environmental factors, likely contributes to the onset of these illnesses. However, the precise mechanisms of causation remain elusive. Most studies investigating the etiology of feeding and eating disorders examine individuals who are already afflicted. Yet, the disorders themselves have such a profound impact on biology and psychology that separating cause from effect proves difficult. Moreover, the relatively low frequency of these disorders makes longitudinal studies that examine individuals before and after illness onset challenging, as a very small portion of the original sample will go on to develop a feeding or eating disorder. Despite these obstacles, studies have allowed investigators to identify risk factors and form research-informed theories about the cause of these syndromes.

Current theories posit that individuals with eating disorders possess a set of predisposing traits (e.g., biologic, genetic, and/or personality vulnerability) which, when triggered by a precipitating event (e.g., the stress of puberty, the decision to go on a diet), result in the illness. Once the illness sets in, several factors work to maintain it (e.g., the social rewards of weight loss, the effects of the starvation state in anorexia nervosa). For example, a young woman with a family history of anorexia nervosa (predisposing factor) may decide to go on a diet (precipitating event) with her friend. While the friend, with no genetic predisposition to the illness, may successfully lose a few pounds before her weight stabilizes, the young woman's initial weight loss may quickly spiral into a state of severe, continued restriction and marked weight loss. Once at a low weight, the obsessional thinking common to the starvation state, along with other factors, may serve to sustain the cycle of restrictive eating (maintaining factor). A discussion of the current understanding of the etiology of feeding and eating disorders, and the gaps within, is detailed below.

Anorexia Nervosa

Genetic and Biologic Factors. Several lines of evidence demonstrate that genetic vulnerability plays an important role in the development of anorexia nervosa. In recent years, the Psychiatric Genetics Consortium has developed genome-wide association studies (GWAS) in eating disorders and, to date, has identified eight loci that contribute to the genetic risk for anorexia nervosa. These studies have also established genetic correlations between anorexia nervosa and other disorders, like anxiety disorders. Relatedly, studies of the serotonin transporter offer preliminary evidence that this gene may interact with environmental stressors to play a role in the development of anorexia, emphasizing that a convergence of elements is necessary for illness development. Prior to the emergence of GWAS in eating disorders, family studies provided useful information about genetic susceptibility to anorexia nervosa, demonstrating that individuals with a family history of the illness are much more likely than those with no family history to receive a diagnosis during their lifetime. In fact, the risk of developing the illness is up to 11 times greater in individuals with a first-degree relative who has experienced the illness. Risk is also increased if a relative has a history of a different eating disorder. Twin studies have demonstrated that concordance rates of anorexia nervosa are substantially higher in monozygotic (identical) twins as compared to dizygotic (fraternal) twins, suggesting that something more than family environment is at play in increasing the odds of illness development. Finally, once the illness sets in, the biologic and psychological changes to the body that occur in the starvation state, including low mood and increased obsessionality, may help to maintain the illness.

Developmental Factors. Anorexia nervosa tends to develop during adolescence, suggesting that factors particular to this period may put individuals at risk for the syndrome. Adolescence represents a time of increased biologic, psychological, and social change. The experience of going through puberty and experiencing changes to body shape and weight may serve as a major stressor for some, triggering or worsening body dissatisfaction and low self-esteem. Research suggests that this trend may be particularly true for individuals who enter puberty earlier than their peers. Weight-related teasing and bullying may become increasingly prevalent at this time, as well, exacerbating the issue. Some studies also suggest that hormonal changes during puberty, specifically in estrogen levels, may play a role in eating disorder development, although this has not been firmly established. Finally, a number of major social and psychological transitions occur throughout adolescence, including identity and role formation, increasing independence from parents, and the initiation of romantic relationships. These stressors and others may work to catalyze the eating disorder.

Psychological and Behavioral Factors. Several psychological factors appear to confer added risk for developing anorexia nervosa. Certain personality traits including high levels of perfectionism, self-discipline, harm-avoidance, and self-criticism are common in individuals with the illness. Individuals with the restricting subtype, in particular, exhibit low impulsivity and are much more likely to delay rewards than individuals without the illness. Cognitive inflexibility is usually prominent, as well. In a subset of individuals, mood and anxiety disorders or symptoms precede the development of the anorexia nervosa. Obsessive-compulsive disorder (OCD) and obsessive-compulsive personality traits also appear to serve as vulnerability factors. In recent years, the habit hypothesis of anorexia nervosa has also garnered initial support. This hypothesis posits that when the behaviors responsible for extreme weight loss in anorexia nervosa emerge, they are typically goal-directed and reward-driven. However, over time they become automatic, fixed responses to certain cues (i.e., habitual). Neuroimaging data support this hypothesis by demonstrating that in patients with anorexia nervosa (compared to healthy controls), food choice is mediated by the dorsal striatum, an area that has long been implicated in habitual behaviors.

Environmental and Social Factors. Environmental and social factors small and large, from experiences in the family or school setting to cultural ideals, may play a role in the development of anorexia nervosa. The influence of family functioning style as a potential predisposing factor remains controversial. No single specific family functioning style appears to be either a necessary or sufficient requirement for developing an eating disorder. As in most psychiatric disorders, various family dysfunctional styles appear to act as nonspecific vulnerability factors and also hamper recovery.

Participation in certain activities in which weight is emphasized may increase the probability of developing anorexia nervosa or another eating disorder. Ballet, gymnastics, modeling, and weight-restricted sports like wrestling or light-weight rowing may lead to preoccupation with body form and unhealthy attempts to control weight. As many of these activities likely select for individuals who are higher in perfectionism, another risk factor for anorexia nervosa, they may be even more likely to contribute to illness onset.

Although it is a common misconception that Western media causes eating disorders, the cultural ideal of thinness, which is certainly perpetuated by the media, may fuel the overvaluation of shape and weight that marks both anorexia nervosa and bulimia nervosa. Indeed, studies in Fiji, where Western media was not available until the 1990s, showed a significant increase in cases of eating disorders following the introduction of Western television shows to the country. Nonetheless, classic anorexia nervosa remains rare, and research has firmly established that media alone does not cause eating disorders.

Bulimia Nervosa

Like anorexia nervosa, there is still much to learn regarding the etiology of bulimia nervosa. This syndrome likely results from the confluence of several factors, as well, many of which overlap with those implicated in the development of anorexia nervosa, but others that differ substantially.

Genetic and Biologic Factors. As in anorexia nervosa, genetics likely play a role in the development of bulimia nervosa, as evidenced by twin and family studies. Individuals with a family history of bulimia nervosa, mood disorder, substance use disorders, or obesity are at higher risk for developing the syndrome. Although GWAS are underway in eating disorders, they have primarily focused on anorexia nervosa and, thus, our knowledge of specific genetic factors implicated in bulimia nervosa remains limited.

Neurobiologic disturbances are also present in individuals with bulimia nervosa and may increase the probability of binge eating. In particular, individuals with the disorder tend to display delayed gastric emptying, enlarged stomach capacity, and reduced secretion of cholecystokinin (CCK), a peptide hormone released by the small intestine that helps to signal satiety during food consumption. Together, disruptions to these neurobiologic processes may put individuals at elevated risk for overeating.

Developmental Factors. The same psychological and social stressors of puberty and adolescence that put individuals at risk for anorexia nervosa also confer risk for bulimia nervosa. Research suggests that hormonal changes during this time, especially shifts in estrogen levels, may be an even greater risk factor in bulimia nervosa than in anorexia nervosa.

Psychological Factors. Unlike individuals with anorexia nervosa, those with bulimia nervosa tend to exhibit high levels of novelty seeking and impulsivity. They also tend to display elevated levels of harm avoidance, negative emotionality, and stress reactivity. Afflicted individuals are more likely than others to experience a comorbid substance use disorder and engage in self-harm, leading to speculation that a subset of individuals with bulimia nervosa may have a propensity toward impulsivity across a range of problematic behaviors. Research findings regarding the possibility of multi-impulsivity as a risk factor for bulimia nervosa remain mixed.

Mood and anxiety disturbances appear to serve as risk factors for bulimia nervosa, as well. In a related vein, research suggests that strong negative emotions may play a role in triggering individual binge episodes in individuals with the disorder, as well, thus acting as maintenance factors in the illness.

Just as the starvation state works to maintain anorexia nervosa, the cycle of binge eating and purging often becomes self-sustaining in bulimia nervosa. Individuals with bulimia nervosa tend to restrict their food intake outside of binge episodes, which puts them at increased risk for episodes of overeating. In turn, after experiencing a binge episode and the resulting compensatory behaviors, individuals often renew their commitment to restrictive eating in an attempt to avoid future overeating episodes. Thus, this cycle tends to become self-perpetuating and works to maintain the disorder.

Environmental and Social Factors. Individuals with bulimia nervosa are likely affected by similar environmental and social stressors as those with anorexia nervosa.

Binge Eating Disorder

Even less is known about the etiology of BED than of anorexia or bulimia nervosa. Twin studies suggest that BED, like other eating disorders, is heritable. Genetic correlations also exist between binge eating and obesity, symptoms of bulimia nervosa, and alcohol dependence. Research thus far has suggested that a history of childhood obesity, mood disorder, and certain negative family dynamics may put individuals at higher risk for developing the illness. In some individuals, dietary restriction appears to play a role in precipitating episodes of binge eating, as in bulimia nervosa. Many individuals with BED, however, seem to experience binge eating outside of the context of dietary restriction. For these individuals, binge episodes may instead be provoked by intense emotions.

DIAGNOSIS AND CLINICAL FEATURES

In DSM-5-TR, the diagnostic criteria for anorexia nervosa, bulimia nervosa, and BED lead to a diagnostic hierarchy so that only one of these diagnoses can be assigned to an individual at a single time. In order of precedence, a diagnosis of anorexia nervosa trumps the other disorders, followed by bulimia nervosa. Finally, rumination disorder cannot be assigned in the presence of any feeding or eating disorder (other than pica). Of note, while these diagnoses may, in general, not occur simultaneously, diagnostic crossover is relatively common, with an individual who previously met criteria for one feeding or eating disorder (e.g., anorexia nervosa) later meeting criteria for a different feeding or eating disorder (e.g., bulimia nervosa). Despite diagnostic crossover and certain shared features, feeding and eating disorders differ substantially from one another in terms of symptoms, behaviors, course, and treatment, as described throughout this chapter.

Anorexia Nervosa

Anorexia nervosa is characterized by three core features, the most salient being restriction of calorie intake resulting in significantly low body weight based on age and height. No specific weight-loss standard defines the illness, so it is important to consider several factors in addition to age and height in determining what constitutes significantly low body weight for an individual. These factors include sex, developmental trajectory, weight history, and current physical health. Nonetheless, drawing from CDC recommendations, DSM-5-TR notes that the usual lower limit of normal body weight in adults is a BMI of 18.5 kg/m^2. Of note, this standard does not apply to children and adolescents who are still undergoing development; for such young people, it is critical to assess their growth trajectory.

Per DSM-5 standards, individuals with anorexia nervosa must demonstrate an intense fear of gaining weight or becoming fat. For individuals who do not articulate or explicitly endorse a fear of weight gain, a diagnosis of anorexia nervosa may still be made in the presence of persistent behaviors that interfere with achieving adequate weight. The behavioral component of this criterion is particularly relevant to children and adolescents, whose age and developmental status may lead to difficulties in verbalizing motivations for engaging in eating disordered behaviors. Additionally, this aspect of the criterion allows individuals who may not endorse fear of fat for cultural reasons to nonetheless obtain a diagnosis of anorexia nervosa.

Finally, anorexia nervosa is marked by disturbances in the experience of body shape or weight. This criterion typically manifests in the way in which an individual with anorexia nervosa perceives their body, believing that their shape is too large or the number on the scale too high despite being at a significantly low weight. Individuals with anorexia nervosa may also place a disproportionate emphasis on thinness in self-schema, such that body shape or weight becomes among the most important aspects in evaluating themselves. Often individuals with this illness do not recognize the severe and dangerous medical risks associated with their current low body weight, even after family, friends, and treatment providers repeatedly emphasize their concern.

Within anorexia nervosa, two subtypes, restricting and binge eating/purging subtype, exist. Individuals with the restricting subtype keep weight off by limiting their daily food intake, often by counting calories, controlling portion size by measuring or weighing food, avoiding specific food items (e.g., cake, red meat) or classes of foods (e.g., fats, desserts), only eating at certain times of the day or a certain amount of time after the last meal, and following other rigid rules that govern eating behaviors. They may also engage in intense and compulsive exercise to lose or maintain weight.

Individuals with the binge eating/purging subtype typically engage in a similar pattern of restrictive eating and may also exercise to keep weight off. However, in addition to these behaviors, individuals who fall under this subtype exhibit recurrent eating binges and/or purging through self-induced vomiting, laxative, or diuretic use. Individuals with these clinical features differ from those with bulimia nervosa because of their low weight status.

Bulimia Nervosa

Bulimia nervosa is marked by repeated episodes of binge eating followed by compensatory behaviors intended to prevent weight gain. Additional details about the definition and assessment of binge eating episodes and compensatory behaviors are provided in the following paragraph. Based on DSM-5-TR criteria, binge eating, and compensatory behaviors must occur on average at least once per week for 3 or more months to warrant a diagnosis of bulimia nervosa. As mentioned in the previous section, unlike individuals with anorexia nervosa, those with bulimia nervosa must be at or above the normal weight during the occurrence of these behaviors. Finally, individuals with bulimia nervosa place undue influence on shape and weight during self-evaluation.

Binge Eating. Central to the diagnosis of bulimia nervosa is the definition of a binge eating episode. In general, according to the DSM-5-TR, an episode of eating is considered a binge when two factors are present: (1) the individual consumes an objectively large amount of food; that is, an amount that most would deem unusually large taking into account the context, and (2) the individual experiences a sense of loss of control, such as feeling that they cannot stop or resist eating at some point during the episode.

While this definition provides a useful framework for thinking about an episode of eating, determining whether an amount of food is objectively large can be difficult. It is useful for the clinician or researcher to ask the patient to describe a typical episode in which eating feels out of control, generate a list of specific foods consumed, and then determine the usual quantity of each food item eaten. Because context may impact whether an amount is considered objectively large, several questions that probe the specific circumstances surrounding the episode should be considered. For example: Was the individual eating alone or with others? If eating with others, how much was everyone else eating? Did the eating occur at a time when overeating might be expected, like at Thanksgiving or a buffet? How much time was spent eating? Did the individual stop eating for an extended period of time during the episode? Table 22–2 provides an overview of questions that may be useful in determining the presence of binge eating and eliciting a typical binge episode.

Although the current DSM-5-TR conceptualization of bulimia nervosa requires clinicians and researchers to assess whether the individual is regularly consuming unusually large amounts of food, as described above, some controversy exists surrounding the importance of the size of the binge. Some experts believe that the amount of food consumed may be less important than the experience of loss of control over eating in making a diagnosis of bulimia nervosa. Indeed, ICD-11 does *not* require that a loss of control eating episode be objectively large for it to be counted as a binge episode. That is, individuals may receive a diagnosis of bulimia nervosa (or BED) if they are having recurrent and frequent subjective and/or objective binge episodes. Research has yet to determine whether individuals who regularly experience loss of control while consuming a normal or small amount of food differ clinically from those who experience objectively large binge episodes, although some studies suggest that binge size does not correlate to psychopathology or impairment. Nevertheless, DSM-5-TR requires that individuals have objectively large binge episodes to qualify for a diagnosis of bulimia nervosa, while ICD-11 has eliminated this requirement.

Table 22–2.
Assessing for Binge Eating

Determining Size of Binge: Is It Objectively Large?

1. Ask about the specific food items typically consumed (e.g., donuts, pizza, chocolate bars)
2. Determine amount of food consumed (e.g., 5 large donuts, 3 slices pizza, 5 fun size chocolate bars)
3. Assess duration of episode, asking about any breaks in eating to determine whether example is better characterized as one or multiple episodes (e.g., 1 hour with no breaks vs. 3 hours with hour-long break in eating)
4. Ask about context:
    a. Was individual alone or with others?
        - If with others, what were they eating at the time?
    b. Did eating occur in a context in which overeating might be expected (e.g., buffet, holiday gathering)?

Probing for Loss of Control

Did you feel like you were unable to stop eating, even if you wanted?

How strong was the drive to continue eating?

If you were interrupted by an external event, would you be able to stop eating?

Did you feel like you could not stop eating until the episode reached its own, natural end point?

Compensatory Behaviors. Individuals with bulimia nervosa engage in inappropriate behaviors to compensate for binge episodes and to control shape and weight. In clinical samples, these individuals most commonly report self-induced vomiting or laxative use. Other compensatory behaviors may include diuretic or enema use, intense exercise, fasting, and medication misuse (e.g., misusing thyroid medication or, in type I diabetes, purposefully restricting insulin use to aid in weight loss). In addition to these behaviors, which must be specifically intended to compensate for episodes of overeating, individuals with bulimia nervosa also tend to restrict or limit their food intake outside of binge episodes. This regular restriction likely plays an important role in the cycle of engaging in binge eating and compensatory behaviors, as described in the previous section on etiology. Specifically, strict rules regarding eating often become impossible to maintain, making the individual more vulnerable to binge eating episodes. Binge episodes, in turn, lead to compensatory behaviors and a renewed desire to restrict food intake to prevent future binge eating episodes, although this logic typically fails.

Binge Eating Disorder

Similar to bulimia nervosa, a diagnosis of BED requires that an individual experiences binge eating episodes at least once per week for 3 or more months. Unlike those with bulimia nervosa, individuals with BED do not engage in regular inappropriate compensatory behaviors. Moreover, in addition to the subjective sense of loss of control that marks binge episodes, behavioral evidence of loss of control must characterize these episodes as evidenced by three or more of the following behavioral criteria: eating much more quickly than usual, continuing eating until uncomfortably full, consuming large quantities of food when not hungry, eating alone due to embarrassment about the amount of food consumed, and feeling depressed, disgusted, and/or very guilty after the eating episode. To warrant a diagnosis of BED, the individual must also report extreme distress about the binge episodes.

Avoidant/Restrictive Food Intake Disorder

ARFID manifests as continual disturbances in eating behavior, which result in clinically significant impairment, such as failure to maintain appropriate weight for age and height, significant nutritional deficiencies, reliance on nasogastric feeding or nutritional supplements to maintain developmentally appropriate weight and/or stable nutritional status, or psychosocial impairment (e.g., inability to eat in public). Individuals with this disorder may exhibit restrictive eating patterns for a variety of reasons, including intense discomfort or dislike of food with certain sensory characteristics (e.g., texture, consistency, smell), fear of eating following an anxiety-provoking, eating-related event (e.g., fear of swallowing after choking or watching someone else choke; fear of eating after an episode of vomiting), or inability to eat normally due to emotional disturbances surrounding eating (e.g., extreme stress surrounding mealtimes due to unpleasant interactions with caregiver). Of note, the reason motivating the restrictive or avoidant food intake must not be weight or shape related, as in anorexia or bulimia nervosa, culturally sanctioned, due to insufficient access to food, or caused by another disorder (e.g., developmental delay).

An 11-year-old K was referred to an eating disorder specialist after her weight fell below the 7th percentile for her age and height. Although K had received treatment for anxiety throughout much of her childhood, she did not manifest worrisome eating until 18 months prior to presentation when she refused to eat during a family vacation because “all the foods tasted weird.” When she returned home, she began to refuse food, citing an intense fear of vomiting. K’s parents were encouraged to wait out the period of change since children often go through phases of picky eating. Yet, K’s eating behaviors became increasingly extreme, and K continued to eat very small amounts. She refused to eat in restaurants and avoided going to friend’s houses or even eating at school, stating that she was less likely to vomit when at home. When she did eat, K would push her food around on the plate, crying when asked to try more. K had been thin throughout childhood, but her weight had hovered around the 25th percentile prior to age 9. When her weight dropped below the 10th percentile and the pediatrician explained the dangerous consequences associated with low weight, K grew tearful and stated that she wanted to gain weight. When asked more about her food restriction, she spoke of an event that occurred years earlier, when she had eaten a meal, grown nauseous, and vomited. She adamantly denied any concern with her appearance, restating her desire to be at a healthy weight.

Pica

Pica refers to the recurrent eating of nonfood, nonnutritive items for at least 1 month by individuals who are at a developmental stage at which this behavior is inappropriate. Individuals who consume nonfood items as part of a culturally sanctioned practice should not receive a diagnosis of pica. Pica most often occurs in individuals with developmental delays or in pregnant individuals, but when a medical or psychiatric condition such as this is present, a diagnosis should only be made if the behavior is so impairing or harmful that it requires separate clinical attention. As mentioned previously, a diagnosis of pica may be assigned in addition to another feeding or eating disorder.

Rumination Disorder

In rumination disorder, individuals repeatedly regurgitate the food that they have consumed before spitting it out, rechewing, or reswallowing it. These behaviors must last at least 1 month for a diagnosis to be made. Although the behavior of rumination may occur in the context of other eating disorders, specifically anorexia or bulimia nervosa, a diagnosis of rumination disorder cannot be given in addition to these illnesses. In other words, to warrant a diagnosis of rumination disorder, the individual must not meet criteria for another feeding or eating disorder (other than pica).

PATHOLOGY AND LABORATORY EXAMINATION

Medical complications are common and often quite serious in both feeding and eating disorders and require close monitoring by the clinical team. Depending on the nature of the disorder, pathophysiologic disturbances may result from starvation and malnutrition, the methods utilized to lose or maintain weight (e.g., restrictive eating, overexercising, self-induced vomiting, laxatives, diet pills, diuretics, or other medication misuse), binge eating, and (in pica) eating of nonfood items. Mortality rates are elevated in both anorexia nervosa and bulimia nervosa due in large part to medical consequences of the disordered behaviors, making medical monitoring particularly important.

When a clinician suspects the presence of a feeding or eating disorder, a full physical examination should be conducted, including height, postvoiding weight in a hospital gown (observing for hidden weights), vital signs (with particular attention to heart rate and orthostatic blood pressure changes), a complete examination of skin, muscle, and subcutaneous fat (noting the degree of starvation), and a neurologic examination (which is, in general, completely normal). Laboratory tests are helpful for assessing the severity of feeding and eating disorders and their medical consequences, although it should be noted that, in many cases, laboratory tests are normal even in the presence of a serious feeding or eating disorder. Abnormal laboratory findings may also be helpful in enhancing the patient's motivation, particularly for individuals who minimize their disorder and its medical implications.

While many of the medical consequences associated with feeding and eating disorders are not imminently dangerous, some require urgent medical evaluation, such as a prolonged QT or QTc (rate corrected) interval on electrocardiogram (ECG) or marked hypokalemia. Medical specialist consultation should be regularly utilized in the management of patients with such abnormalities. Table 22–3 lists the suggested laboratory tests for patients with anorexia nervosa or bulimia nervosa, with judgment regarding the full extent of the workup for individual patients based on weight loss, severity of illness, subtype, and comorbidity. For example, for individuals with anorexia nervosa binge eating/purging type and bulimia nervosa, laboratory tests often include determination of electrolytes and serum amylase, which are more likely to be abnormal in those who purge than in the restricting type of anorexia nervosa. Table 22–4 summarizes common medical consequences of anorexia nervosa, bulimia nervosa, and BED. The following sections provide additional details regarding pathophysiology that may be specific to each eating disorder.

Table 22–3.
Suggested Laboratory Workup for Patients With Anorexia Nervosa and Bulimia Nervosa

For all patients:

- Complete blood count
- Electrolytes
- Blood urea nitrogen, creatinine
- Thyroid-stimulating hormone, free thyroxine
- Total protein and prealbumin
- Fasting glucose
- Amylase if purging occurs
- Serum phosphate
- Electrocardiogram

If underweight or experiencing amenorrhea:

- Bone mineral density (dual-energy x-ray absorptiometry)

Table 22–4.
Common Medical and Physical Complications Associated With Eating Disorders

	Anorexia Nervosa	Bulimia Nervosa	BED
General	▸ Low weight ▸ Loss of body fat ▸ Muscle atrophy ▸ Hypothermia	▸ Weight within normal-overweight/obese range	▸ Weight often in overweight/obese range (with associated complications not listed here)
Skin	▸ Lanugo (fine hair growth on body) ▸ Hair loss or thinning ▸ Carotenemia (yellowing of skin)	▸ Russell sign: Calluses on hands from teeth scraping during self-induced vomiting (not extremely common)	
Cardiovascular	▸ Arrhythmias ▸ Bradycardia ▸ Hypotension ▸ Orthostasis ▸ Prolonged QTc ▸ Peripheral edema	▸ Prolonged QTc ▸ Myocardial toxicity from emetine (ipecac)	
Gastrointestinal	▸ Delayed gastric emptying ▸ Constipation	▸ Gastric distention ▸ Swollen parotid and salivary glands ▸ Elevated serum amylase levels ▸ Erosion of dental enamel, multiple caries	
Electrolyte/Metabolic	▸ Hypokalemia (low potassium) ▸ Hyponatremia (low sodium) ▸ Hypochloremia (low chloride) ▸ Alkalosis ▸ Hypercholesterolemia (high cholesterol) ▸ Hypoglycemia (low blood sugar) ▸ Elevated liver enzymes	▸ Hypokalemia (low potassium) ▸ Hyponatremia (low sodium) ▸ Hypochloremia (low chloride) ▸ Alkalosis	
Endocrine	▸ Low luteinizing hormone ▸ Low follicle-stimulating hormone	▸ Oligomenorrhea or amenorrhea	

	▶ Low estrogen or testosterone ▶ Low thyroxine ▶ Elevated cortisol ▶ Amenorrhea ▶ Decreased bone mineral density (can lead to osteopenia or osteoporosis)
Hematologic	▶ Leukopenia ▶ Anemia

Anorexia Nervosa

Patients with anorexia nervosa may present with certain physiologic abnormalities that are readily apparent even without an extensive evaluation. These may include significantly low body weight, lanugo (soft, fine hair that grows on the body in the starvation state), hair loss or thinning, and carotenemia (yellowing of the skin). Patients may also present with medical complaints that are often confirmed by further examination. They may, for example, report constantly feeling cold and, indeed, their body temperature is often lower than normal, secondary to hypometabolic activity associated with starvation. Reports of weakness may be due in part to muscle atrophy. Many patients report lightheadedness or fainting and are often orthostatic by vital signs, in addition to exhibiting bradycardia. Reports of gastrointestinal discomfort may be related to delayed gastric emptying and constipation, which commonly accompany anorexia nervosa.

Many malnourished patients with anorexia nervosa develop the euthyroid sick syndrome, in which decreased levels of thyroid-stimulating hormone (TSH), total thyroxine (T_4), and total triiodothyronine (T_3) may be seen, free T_4 and free T_3 are usually unchanged, and reverse T_3 (rT_3) is elevated. Generally, these endocrine abnormalities represent energy-conserving consequences of starvation that improve spontaneously with refeeding. They do not require immediate treatment but should be followed to determine if improvement occurs with refeeding.

The hypothalamic–pituitary–gonadal axis is virtually always affected in anorexia nervosa. Low estrogen and progesterone levels are common in women, and low testosterone levels are common in men. Additionally, low levels of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) are seen. Patients with anorexia nervosa have low levels of serum leptin that correlate with low body weight and low body fat, and mediate changes in the hypothalamic–pituitary–gonadal axis.

Related to these abnormalities, bone mineral density is typically low and patients are at risk for osteopenia and osteoporosis, likely due to a combination of low levels of estrogen, high levels of cortisol, and malnutrition. Dual-energy x-ray absorptiometry (DEXA) scans can be helpful in alerting patients to the realistic medical consequences of their disorders. The mechanisms of osteopenia and osteoporosis in anorexia nervosa are more complex than in postmenopausal osteoporosis, and treatments useful in postmenopausal women do not necessarily help in anorexia nervosa. Orally administered estrogen, calcium, and vitamin D supplementation in anorexia nervosa have not been shown to offer benefits in bone mineral density while the patient is still low in weight. While almost all other medical consequences of anorexia nervosa resolve with weight restoration, bone health may not fully recover.

The structural and functional neuropathology of anorexia nervosa is increasingly well documented by magnetic resonance imaging (MRI), computed tomography (CT), fMRI, positron emission tomography (PET), and single photon emission CT (SPECT) scans. Decreases in total brain volume and sulcal complexity and increases in brain ventricular size are usually seen. The neuropathology of anorexia nervosa reverses substantially with refeeding and weight gain. Approximately 40% of typical anorexia nervosa patients score in an abnormal range on two or more neuropsychological tests in standard batteries, suggesting that these patients may have difficulties benefiting from cognitive-behavioral therapies and other psychotherapies in these states. The extent to which long-term improvement in neuropsychological abnormalities may occur in proportion to weight restoration and cessation of binge/purge activity over a prolonged period of stability is not yet clear, and persistent abnormalities in neuropsychological performance following nutritional rehabilitation have been associated with poor prognosis. In usual clinical practice, neuroimaging is not necessary, but in patients with atypical features, especially if other findings suggest that a brain lesion may be present, imaging studies may be of value.

Bulimia Nervosa

Individuals with bulimia nervosa are mainly at risk for medical consequences related to inappropriate compensatory behaviors. Self-induced vomiting, laxative or diuretic misuse, and other compensatory behaviors can cause electrolyte disturbances with serious health risks. Of note, individuals with the binge eating/purging subtype of anorexia nervosa are also at risk for many of the medical sequelae discussed below.

Hypokalemia (low serum potassium), a particularly dangerous electrolyte imbalance that can cause cardiac arrhythmias, commonly occurs in individuals with recurrent vomiting. Thus, potassium should be regularly monitored, especially in individuals with self-reported purging. Likewise, ECGs are recommended to assess for arrhythmias and for prolongation in QT and QTc, which may even occur in individuals without electrolyte abnormalities. Hypomagnesemia (low magnesium) may co-occur with hypokalemia and must also be addressed as it inhibits prolonged normalization of potassium.

Hyponatremia (low serum sodium) also commonly occurs in patients with bulimia nervosa (and often in anorexia nervosa) due either to increased water intake, which dilutes sodium concentration in the blood, or to the development of a syndrome of inappropriate antidiuretic hormone release (SIADH). Individuals with eating disorders may consume excessive amounts of water to aid in appetite suppression and subsequent food restriction or, in the context of treatment that involves weigh-ins, to falsely inflate weight (“water loading”). Hyponatremia increases seizure risk and thus must be addressed quickly. Regardless of the etiology of the hyponatremia, fluid restriction is the treatment of choice. When fluid restriction does not sufficiently improve sodium levels, intranasal arginine vasopressin (ddAVP) should be considered.

As in anorexia nervosa, many individuals with bulimia nervosa may present with dehydration (often due to fluid loss from inappropriate compensatory behaviors). Dehydration may result in tachycardia, orthostatic hypotension, and laboratory abnormalities related to prerenal azotemia, including elevated creatinine and blood urea nitrogen (BUN) levels. Although these issues typically resolve with the normalization of food and fluid intake, intravenous hydration may be necessary in the short term for speedier symptom relief and the reversal of laboratory abnormalities.

Binge Eating Disorder

Obesity and its consequences are the primary health concerns for individuals with BED. Whether individuals with obesity and BED are at greater risk for medical complications than individuals with obesity without an eating disorder remains unclear.

DIFFERENTIAL DIAGNOSIS

Differentiating feeding and eating disorders from other disorders is typically not difficult when a clinician is able to assess current symptoms and behaviors and obtain a full history. Eating disorders in particular have been studied extensively and are therefore more easily separated from other syndromes. Because much less research describing the presentation, course, prognosis, and treatment response of feeding disorders exists at the moment, differential diagnosis for these syndromes may prove more complicated.

Anorexia Nervosa

Because of the prominent physical and behavioral symptoms (i.e., significantly low body weight and restrictive eating) that mark anorexia nervosa, this disorder is typically easy to identify. In certain cases, however, differentiating the disorder from other syndromes may prove more difficult, and clinical judgment may become key in making a diagnosis.

Differential diagnosis may be particularly difficult for individuals who have reached a low body weight but deny or do not articulate fear of gaining weight or becoming fat. By DSM-5-TR criteria, these individuals may still warrant a diagnosis of anorexia nervosa provided that they demonstrate behaviors that persistently interfere with weight gain. Children and individuals from certain cultural backgrounds, in particular, may not acknowledge a fear of becoming fat, but may demonstrate behavioral manifestations of the illness. In these cases, especially in the evaluation of children, collateral information from parents and others may assist in creating a clear picture of the behaviors that the afflicted individual engages in, thus elucidating the diagnosis. Obtaining a full history of the weight loss and any related behaviors often helps to clarify whether the individual should receive a diagnosis of anorexia nervosa, as well.

Throughout the evaluation process, particularly when certain aspects of the illness (such as fear of weight gain) are not readily apparent, clinicians should consider whether a different disorder better explains the low weight status. Medical conditions that may interfere with weight gain, including gastrointestinal problems and thyroid conditions, should be included in the differential diagnosis. When considering these syndromes, however, clinicians should bear in mind that medical sequelae, including gastrointestinal and thyroid abnormalities, are common in eating disorders and their presence does not necessarily exclude a diagnosis of anorexia nervosa.

Individuals with certain psychiatric conditions may also present with weight loss, suppressed appetite, or even fear of eating which can be distinguished from anorexia nervosa. Individuals with psychotic disorders may demonstrate delusions relating to food (e.g., that it has been poisoned) that lead to restricted intake. Mood and anxiety disorders may also include symptoms of weight loss or restrictive eating. For some individuals, a major depressive episode may be accompanied by weight loss and lack of appetite. In certain cases (i.e., when the diagnostic criteria become present), an additional diagnosis of anorexia nervosa may be warranted, but, for many, the weight-related symptoms remain attributable to the episode of depressed mood. Likewise, certain anxiety disorders may lead to weight loss or restrictive behaviors. For example, individuals with a specific phobia of vomit may restrict their eating out of concern that eating could lead to an episode of vomiting, and individuals with

OCD may develop compulsions related to eating behavior (e.g., only taking a bite of food every minute, or eating foods in specific order depending on color). Unless they also begin to manifest other characteristics of anorexia nervosa (including low weight and preoccupation with shape and weight), these individuals should not receive a diagnosis of anorexia nervosa. Clinicians should also consider a diagnosis of ARFID, which is distinct from anorexia nervosa in that psychopathology related to shape and weight concern is absent in this illness.

Under most circumstances, distinguishing anorexia nervosa from bulimia nervosa should be relatively straightforward. The two are differentiated largely on the basis of weight status, with low weight individuals qualifying for a diagnosis of anorexia nervosa regardless of whether or not they engage in episodes of binge eating and compensatory behaviors. The distinction may become more complicated when an individual with recurrent binge eating and compensatory behaviors presents at a body weight that borders on low but is not clearly in the normal or low range (e.g., a BMI of 18.5 kg/m^2). In these cases, obtaining a detailed weight history may help determine if the individual has lost weight in the context of the disordered eating behaviors (making a diagnosis of anorexia nervosa more likely), or if the individual has continued to maintain preillness weight throughout the course of the eating disorder (making a diagnosis of bulimia nervosa more likely).

Individuals who have recently lost a significant amount of weight and demonstrate other characteristics of anorexia nervosa, including fear of weight gain, but who remain within a normal weight range may meet criteria for atypical anorexia nervosa, an example of OSFED in DSM-5. Diagnostically, atypical anorexia nervosa essentially mirrors classic anorexia nervosa, with the exception of the low weight criteria. However, little is known about course, prognosis, or effective intervention strategies for individuals with this diagnosis.

Bulimia Nervosa

When able to obtain a report of current eating behaviors, differentiating bulimia nervosa from other disorders should be relatively straightforward. As discussed earlier in the chapter, perhaps the most challenging aspect inherent to the assessment of bulimia nervosa revolves around determining whether an individual experiences objective binge episodes. This factor is key, as individuals who do not experience objective binge episodes should not, according to DSM-5-TR, receive a diagnosis of bulimia nervosa. Determining whether an individual is experiencing objective binge episodes and compensatory behaviors at the threshold required for a diagnosis of bulimia nervosa (an average of one time per week for 3 or more months) may be particularly important in differentiating bulimia nervosa from other eating disorders. Specifically, individuals who engage in regular compensatory behaviors (e.g., self-induced vomiting, laxative, diuretic, or enema use) but do not report recurrent, objective binge eating episodes may fall under the OSFED category, purging disorder. Alternatively, those who engage in both objective binge eating episodes and compensatory behaviors, but do so less than once per week or for fewer than 3 months should be assigned to the OSFED category, bulimia nervosa of low frequency and/or limited duration. Individuals who report recurrent binge eating without compensatory behaviors should receive a diagnosis of BED. Finally, as mentioned above, bulimia nervosa is distinguished from anorexia nervosa, binge eating/purging type based on weight status.

Some individuals experiencing a major depressive episode may present with episodes of overeating secondary to their mood diagnosis. If, however, their episodes are objectively large, accompanied by a sense of loss of control, followed by compensatory behaviors, and have occurred at or above the required frequency and duration thresholds, an additional diagnosis of bulimia nervosa may be warranted.

Binge Eating Disorder

BED should be relatively easy to recognize based on the patient's self-reported pattern of eating. As described above, BED can be distinguished from bulimia nervosa in that recurrent compensatory behaviors are absent in this diagnosis. Mood should also be assessed as individuals experiencing a major depressive episode may report overeating episodes that could be attributable to the depressive episode or may warrant a separate diagnosis of BED.

Distinguishing BED from the patterns of eating seen in individuals with overweight or obese BMIs without an eating disorder may be particularly difficult. Many individuals presenting above the normal weight range may describe consuming large amounts of food on a regular basis. Only a subset of this population, however, will meet full criteria for BED. In individuals with episodes of overeating who have BMIs consistent with overweight or obesity, it is particularly important to probe for behavioral evidence of lack of control while eating. Individuals with BED typically describe feeling as if they are unable to stop or resist eating, even if interrupted by an external cue, like their phone ringing. Additionally, individuals with BED must demonstrate marked distress over their binge eating.

Avoidant/Restrictive Food Intake Disorder

ARFID should be relatively easy to distinguish from other feeding or eating disorders. While individuals with this disorder may present at low weight and often exhibit eating behaviors that appear quite similar to those seen in anorexia nervosa, individuals with ARFID lack any excessive concern with shape or weight. As mentioned in the discussion of anorexia nervosa, differentiating between these disorders may require particular thought when evaluating children, young adolescents, and individuals from certain cultures who may not explicitly endorse fear of weight gain but who may, nonetheless, demonstrate repeated attempts to avoid weight gain. For this reason, clinicians should not immediately exclude anorexia nervosa from the differential upon hearing an individual deny shape or weight concern. If evidence of attempts to lose or maintain low weight is present, the individual may warrant a diagnosis of anorexia nervosa rather than ARFID.

Importantly, ARFID should not be diagnosed in the presence of another medical or psychiatric condition that explains the restrictive food intake unless the behaviors or symptoms are severe enough to require separate, specialized clinical attention. If eating behaviors are so restrictive that the individual is at risk for medical or nutritional complications, can only maintain health with the assistance of nutritional supplements or nasogastric feeding, or experiences significant psychosocial impairment specifically due to problems with food intake, a diagnosis of ARFID should be made even in the presence of another pre-existing condition. Pre-existing conditions may include medical issues that impact food intake, like gastrointestinal problems that make eating uncomfortable, or psychiatric illnesses, particularly anxiety disorders (e.g., OCD, specific phobia of choking or vomiting), autism spectrum disorders, and developmental delays.

Pica

Pica is the sole feeding or eating disorder that may be diagnosed in addition to any other feeding or eating disorder. For this reason, regular consumption of nonfood items should be assessed even when the presence of anorexia nervosa, bulimia nervosa, BED, ARFID, or rumination disorder has already been established. Like ARFID, pica should not be diagnosed in addition to a pre-existing medical or psychiatric condition unless the feeding or eating behaviors cause substantial impairment to require specialized intervention. Because pica as a behavior often occurs in the context of developmental delays or certain medical conditions (e.g., pregnancy), it is particularly important to determine the severity of the symptoms and whether they necessitate additional care.

Rumination Disorder

Although the behavior of rumination sometimes occurs within an episode of another feeding or eating disorder, particularly anorexia nervosa or bulimia nervosa, rumination disorder as a syndrome cannot be diagnosed according to DSM-5-TR in the presence of anorexia nervosa, bulimia nervosa, BED, or ARFID. Rumination disorder should also be distinguished from the OSFED category, purging disorder, as the behaviors present in these disorders, while similar to some degree, are fundamentally distinct. Rumination refers to the regurgitation of food that has been consumed. While individuals who engage in this behavior may spit out food, this pattern is distinct from purging through self-induced vomiting.

As with ARFID and pica, rumination disorder should not be diagnosed in addition to another disorder unless the behaviors necessitate specialized treatment. Individuals with certain medical conditions, particularly gastrointestinal issues like acid reflux, may report the symptom of rumination, but will usually describe that this behavior occurs involuntarily and that it resolves with treatment of the underlying medical condition. Individuals with developmental delays may also engage in rumination, but only warrant a separate diagnosis if clinically indicated based on symptom severity.

COURSE AND PROGNOSIS

Anorexia Nervosa

Anorexia nervosa typically develops in adolescence, although illness onset may occur at almost any point during the lifespan, with increasingly frequent accounts of it preceding puberty. Most often, the disorder begins when an individual in the normal- or slightly overweight range (commonly an adolescent girl or young woman) starts a diet with the intention of losing a few pounds. The decision to begin dieting usually signifies some level of pre-existing body dissatisfaction and is often spurred by a recent stressor like a breakup, transition into a new school, or a negative comment about shape or weight made by a peer, family member, coach, or dance instructor. For some individuals, initial weight loss is caused not by a diet but rather by an incident entirely outside of the individual’s control, like an acute gastrointestinal condition. Either way, after losing a few pounds, individuals tend to report experiencing positive reinforcement, often in the form of compliments from those around them, that encourage continued weight-loss efforts.

Through this point in the illness trajectory, these individuals may resemble any other teen or adult who loses a few pounds and finds the experience rewarding. However, for individuals vulnerable to developing anorexia nervosa, the illness typically intensifies rather quickly after the initial weight loss. Dietary restriction becomes increasingly severe, with individuals limiting portion size or restricting calories to dangerously low levels, avoiding entire food groups (particularly high-fat foods), and skipping meals. Some individuals also begin to limit fluid intake. Exercise may begin to acquire a compulsive and intense quality, and binge eating and purging (through various means) may emerge. As individuals reach a state of malnourishment and low weight, idiosyncratic ritualistic eating behaviors (e.g., cutting food into small pieces, only eating at set times, concocting strange mixtures of foods) become prevalent. Anxiety and depressive symptoms worsen, preoccupation with body shape and weight intensifies, and individuals tend to become increasingly withdrawn from friends and family.

The precise course of the illness varies substantially, although certain patterns have emerged in the literature. Follow-up studies of patients with anorexia nervosa reveal that, at the time of assessment, approximately 30% to 50% have achieved full recovery, 10% to 20% remain chronically ill, and the remainder have improved but continue to struggle with certain disordered behaviors. The chronically ill group often requires multiple hospitalizations, with a small number of patients being in and out of the hospital for the majority of their lives. Because research on anorexia nervosa tends to examine individuals who have presented for specialized treatment, little is known about the frequency with which individuals experience a brief illness and achieve recovery without eating disorder care.

Of particular note, anorexia nervosa has a mortality rate as high as that associated with any psychiatric illness. Compared to the general population, individuals with the illness are up to six times more likely to die. The majority of deaths are attributable to medical complications of low weight and malnourishment but a smaller, yet significant, proportion of deaths (approximately 1 in 5) are due to suicide.

Although additional research is needed to identify predictors of outcome in anorexia nervosa, studies have found that adolescents with a shorter duration of illness tend to have a better prognosis, emphasizing the importance of early detection and intervention. Studies examining inpatients with anorexia nervosa just prior to and following discharge have also identified certain predictors of post-treatment outcome. In particular, individuals who achieve full weight restoration on an inpatient unit and maintain their weight in the first month after discharge are, perhaps unsurprisingly, more likely to remain at a healthy weight up to a year after treatment. Lower BMI at discharge and weight loss in the first month after treatment, on the other hand, predict poorer long-term outcome. Additionally, individuals who demonstrate an ability to consume a diet that is high in variety and energy density (i.e., a greater concentration of kcal/g) prior to discharge seem to do better after treatment. Finally, a comorbid diagnosis of depression predicts continuing to meet diagnostic criteria for anorexia nervosa decades later.

Bulimia Nervosa

Bulimia nervosa usually develops slightly later than anorexia nervosa, in the late teens or early 20s. Although a substantial portion of individuals report binge eating as their first behavioral symptom of bulimia nervosa, in just under half of all cases restrictive eating and dieting appear to precede the onset of binge eating. Diagnostic crossover from anorexia nervosa to bulimia nervosa is relatively common and, in some cases, restriction may lead to low weight initially, with individuals passing through an episode of anorexia nervosa prior to the development of the bulimia nervosa. In other cases, individuals will remain at normal weight throughout the duration of the dieting. Either way, for those whose illness begins with restrictive eating, this behavior eventually becomes unsustainable, or the individual begins to experience binge episodes. In some cases, the individual may experiment with or even regularly engage in purging before the binge episodes emerge, but more commonly purging starts as a way to compensate for food binges. The cycle of restriction, binge eating, and purging tends to become self-maintaining, as discussed previously.

In general, individuals with bulimia nervosa are more likely than those with anorexia nervosa to attain full or partial recovery, particularly if they have received treatment for the illness. Recent research suggests that, on average, up to 35% of patients with bulimia nervosa who complete an evidence-based treatment achieve full abstinence of core illness symptoms following treatment, and that this rate of abstinence is maintained at follow-up. A larger proportion of patients achieve partial remission. Notably, research finds that bulimia nervosa often goes undetected by health care providers. Several reports suggest an increased mortality risk in this illness. However, study samples typically include a subgroup with a lifetime history of anorexia nervosa who may falsely elevate the mortality risk specifically associated with bulimia nervosa, therefore limiting the interpretation of these data.

Research has yet to identify any clear predictors of outcome in bulimia nervosa. Studies have examined a variety of potential prognostic factors, including duration of illness, age of onset, illness severity, comorbid diagnoses, and personality characteristics, with mixed results. In individuals who receive treatment for bulimia nervosa, studies have found that rapid symptom reduction predicts better outcome.

Binge Eating Disorder

Compared to anorexia nervosa and bulimia nervosa, less is known about the course of BED. In general, the disorder appears to affect an older age group, with patients often presenting for treatment well into adulthood. In most cases, it appears that episodes of binge eating emerge outside of the context of dietary restriction.

Individuals with BED have a much better prognosis than those with anorexia or bulimia nervosa. A substantial number of afflicted individuals appear to experience spontaneous remission without any treatment. One long-term follow-up study found that over 80% of individuals who initially met criteria for BED achieved full recovery 5 years later. The vast majority of individuals in this sample did not receive any treatment over the course of the study. Shorter-term follow-up studies have also found impressive recovery rates, with over 50% of the sample experiencing full remission after just 6 months. Nonetheless, rates of obesity and overweight remain high in individuals with BED, putting these individuals at risk for weight-related health complications.

TREATMENT

Treatment of feeding and eating disorders remains challenging, with empirically supported interventions still lacking for certain populations, like individuals with ARFID, and rates of treatment nonresponse or relapse remaining relatively high even in populations for which evidence-based treatments exist, like bulimia nervosa. Interventions for feeding disorders, in particular, have received very little research attention, although the merging of feeding and eating disorders into one cohesive section in DSM-5 has prompted further investigation into treatment strategies for ARFID, in particular. Although additional research efforts are needed to identify effective interventions for feeding and eating disorders, the sections below will discuss commonly utilized treatments that represent the current understanding of best clinical practices.

Treatment Goals

Although each feeding and eating disorder is substantially different from the others and therefore responds well to different interventions, certain goals of treatment remain the same across the board. As behavioral disorders, one of the primary goals of treatment for any feeding or eating disorder is to disrupt disordered behaviors, specifically, problematic patterns of feeding or eating. For disorders that put individuals at risk for weight problems (i.e., low weight in anorexia nervosa or ARFID), another key goal of treatment is to normalize and stabilize weight. For eating disorders in particular, these two initial goals (normalization of eating behaviors and weight) complement each other, as the normalization of one often promotes the normalization of the other. For example, in anorexia nervosa, weight gain and nutritional rehabilitation often lead to a decrease in ritualistic eating behaviors, and as ritualistic eating behaviors diminish, weight gain and maintenance become more attainable.

A secondary, but often equally important, goal of treatment for eating disorders in particular consists of altering the problematic cognitions, attitudes, and beliefs that may help maintain the disordered behaviors. In anorexia nervosa and bulimia nervosa especially, it is common for many distorted thoughts to decrease in intensity as weight and/or behaviors normalize. In anorexia nervosa, weight stabilization often leads to reduction in obsessive thinking. In bulimia nervosa, breaking the cycle of binge eating, purging, and restriction often allows individuals to recognize the irrational nature of their fears about weight gain. In either case, after disrupting problematic behaviors, additional therapy is typically necessary to work through some of the more deeply ingrained beliefs, such as those that tie weight and shape to self-worth.

Finally, eating disorders have high rates of psychiatric comorbidity, and addressing these problems often becomes important after the eating disorder is stabilized (although, in some cases, the comorbidities may take precedence). For many individuals, certain comorbid symptoms, particularly depressive and anxiety symptoms, substantially diminish following the stabilization of the eating disorder. As the state of starvation is associated with depressed mood and higher levels of anxiety, this is particularly true for individuals with anorexia nervosa who restore their weight. Likewise, as discussed in greater detail below, certain treatments (especially the use of SSRIs and other antidepressants) for bulimia nervosa and BED can simultaneously improve eating disorder symptoms and mood. For other individuals, however, comorbidities require additional, separate treatment even after eating disorder remission.

Level of Care

Important to the discussion of treatment of eating disorders, in particular, is the concept of level of care. Level of care refers to the range of treatment settings that exist (in this case, for eating disorders), from less restrictive options like outpatient therapy, to more intensive interventions like residential or hospital-based treatment programs. In choosing an appropriate option for an individual with an eating disorder, illness severity, including medical status, psychosocial impairment, and the degree of behavioral disturbance, should be considered. Treatment setting should assure safety without being any more restrictive than necessary. Because eating disorders are serious psychiatric conditions that may be associated with medical or psychiatric comorbidity, intensive treatment options often become appropriate or even necessary.

Anorexia Nervosa

The egosyntonic nature of anorexia nervosa (i.e., affected individuals often view some of their symptoms as aspects they wish to continue, or as parts of themselves) can make treatment of this syndrome particularly difficult. Individuals are often ambivalent about recovery. While they may identify certain thoughts or behaviors that they would like to resolve (e.g., distressing obsessive thoughts about food, shape, or weight; objective or subjective binge eating episodes; weakness and fatigue), or certain areas of their life that might improve without the eating disorder (e.g., relationships, concentration), they typically remain very afraid of and resistant to the idea of gaining weight. This ambivalence can make engaging patients in treatment particularly challenging. Therefore, an initial goal of treatment should be for the clinician to establish a positive, supportive relationship with the patient and to begin to involve him or her in treatment planning. Psychoeducation, including a full discussion of the medical risks associated with low weight, often plays an important role in motivating the patient, particularly in the initial phases of treatment.

Because of the complex nature of the illness and the high frequency of serious medical and psychiatric comorbidities, individuals with anorexia nervosa typically require a treatment team, or a group of clinicians with different specializations who work together to help the individual make progress. This team may deliver care on an outpatient basis or in a more structured setting and typically consists of a group of multidisciplinary clinicians, including a therapist and/or a psychiatrist, nutritionist, and general medical doctor. In more structured settings, the team often includes additional members, including nursing staff, mental health aids, recreational or occupational therapists, and social workers. For children and adolescents, in particular, parents or other family members may take on an active role in refeeding and weight restoration. Each member plays a different specialized role, but it is imperative that the team regularly communicates regarding treatment goals, progress, and plan.

For adults with anorexia nervosa, no particular treatment has emerged as clearly superior based on empirical studies. For adolescents and children with anorexia nervosa, certain treatments, particularly family-based therapies (discussed below), have demonstrated efficacy in research trials. Despite the limited research support for existing interventions, a variety of treatment types and modalities can promote recovery for individuals with the illness.

Nutritional Rehabilitation and Weight Restoration. Nutritional rehabilitation and weight restoration are primary treatment goals that may be achieved at any level of care, although inpatient hospitalization typically allows for the most efficient refeeding process because of the close supervision with medical monitoring that is available. The following section will discuss weight restoration and nutritional rehabilitation in a structured setting, although many of these principles may be applied to less structured treatment settings, as well.

Structured treatments are indicated for individuals who have not achieved weight gain in outpatient settings, or whose behaviors or symptoms are severe enough to require a higher level of monitoring or care than is possible on an outpatient basis. Individuals presenting with serious medical complications or suicidal ideation or behavior are typically candidates for inpatient treatment. Several options that vary in intensity exist under the umbrella of structured treatment, but all share the primary goal of weight restoration. Most structured settings employ behavioral protocols aimed at normalizing weight through calorie prescriptions, meal plans, structured and supportive meals, disruption of compensatory behaviors (including purging and exercise), and expectations of regular weight gain. A treatment team also provides a range of individual and group therapies, some of which are discussed later in this section.

A privilege or level system is often employed, such that individuals enter the program at the lowest and most restrictive level but can increase their level and subsequently gain privileges and autonomy as they progress through treatment. Prorecovery behaviors (e.g., eating 100% of food on the meal plan, abstaining from compensatory behaviors, meeting weight gain goals) are rewarded with positive reinforcements, which may mean gaining access to different group activities, receiving less monitoring and supervision, taking a more active role in meal planning, etc. Behaviors that cater to the eating disorder (e.g., restricting food intake, purging or exercising, failing to meet weight gain expectations) are met with negative reinforcements, or a reduction in privileges and an increase in supervision. For individuals who continuously fail to gain weight, activity is restricted to decrease calorie expenditure, and bed rest may be employed.

In most structured facilities, calorie prescription starts between 1,500 and 1,800 kcal/day (and emerging evidence suggests that higher rates, around 2,000 kcal/day, are safe and more efficacious) consumed across three to four supervised meals and snacks. Macronutrient content is also prescribed so that patients receive the necessary daily requirements. Because this diet represents an increase in food intake compared to pretreatment consumption for almost all patients with anorexia nervosa, medical status should be monitored closely for signs of refeeding syndrome, a rare but potentially fatal medical complication that involves fluid, electrolyte, and mineral shifts in the body (see Table 22–5 for common signs of this syndrome). Despite the need for careful medical monitoring during this time, a diet of 1,500 to 1,800 kcal/day has been shown to be a safe starting point for patients with anorexia nervosa across various research studies. Additionally, due to the medical complications associated with low weight and starvation, it is typically viewed as safer and more ethical to increase calories (and therefore, weight) relatively quickly.

Most programs will maintain calories around 1,500 to 1,800 for several days until the patient is clearly medically stable. At this point, calories are typically increased in a predetermined sequence, often by several hundred calories every other day, until a diet consisting of approximately 3,500 to 4,000 kcal/day is achieved. To meet these calorie needs, patients may consume high-calorie nutritional supplements, like Ensure, as part of the meal plan. For individuals who refuse to consume solid foods or nutritional supplements, some facilities utilize nasogastric feeding (i.e., tube feeding) to achieve weight gain, although this option is generally viewed as less ideal as it is important for patients to practice normal eating as part of the recovery process.

While calorie prescriptions on structured units are typically quite high, it is important to consider that in order to gain 1 lb, an individual must exceed his or her energy requirements by roughly 3,500 kcal. Thus, the calorie prescriptions described above typically result in gaining between 2 to 5 lb per week. Goal weights are determined at the start of the treatment, but vary across facilities and are often impacted by insurance coverage. Generally speaking, for an individual to achieve full weight restoration, BMI should be increased to greater than 18.5 kg/m^2 or 85% IBW, and most clinicians agree that a BMI above 20 kg/m^2 or 90% IBW is healthier and more likely to prevent relapse. For children or adolescents, goal weight should be determined based on pediatric growth curves and, for adults, it is often useful to consider preillness weight history to determine if a goal weight substantially higher than a BMI of 18.5 kg/m^2 or 85% IBW should be recommended. Despite these clinical goals, health insurance companies may limit reimbursement, covering weight restoration to weights or BMI ranges that are often below ideal.

Table 22–5.
Signs of Refeeding Syndrome

Lab Abnormalities
Hypophosphatemia
Hypokalemia
Hypomagnesemia
Thiamine deficiency
Glucose intolerance

EKG Abnormalities
Dysrhythmias
Torsades de pointes

Physical Symptoms
Edema
Respiratory difficulty
Muscle pain and weakness
Gastrointestinal disturbances (e.g., constipation, diarrhea)

While weight restoration often resolves many of the depressive and anxiety symptoms associated with the starvation state, additional therapy is almost always indicated after weight has been normalized to prevent relapse, continue to focus on disordered thoughts and behaviors and work through any underlying issues or comorbidities. Some structured treatments utilize stepped care systems, in which individuals can transition from higher levels of care (e.g., inpatient or residential) to lower levels of care (e.g., day treatment, evening program, outpatient) within the same system as they progress through treatment. Other programs depend on case managers or other staff members to set up discharge plans that aim to help individuals maintain treatment gains. Throughout structured treatment and after, a variety of psychotherapy options are available.

Psychotherapy. Several psychotherapies have been evaluated for the treatment of anorexia nervosa, both in structured and unstructured settings. For adults, several therapies are regularly used and recommended for clinical management of the illness. Some consistent research findings regarding modest but significant clinical improvements following these therapies support their use; however, upon careful inspection it appears that no particular therapy is better than the others for the treatment of adults with anorexia nervosa, and none is dramatically superior to treatment as usual.

Perhaps the most extensively studied and utilized treatment for anorexia nervosa is cognitive-behavioral therapy (CBT), which often focuses on helping the individual alter disordered behaviors regarding eating, exercise, and other compensatory behaviors, challenge related distorted thoughts, and identify negative core beliefs. Interpersonal psychotherapy (IPT), which focuses on dysfunction in interpersonal relationships, social roles, and communication, is also often used for the treatment of anorexia nervosa with similar outcome to that of CBT. Treatment that targets maladaptive habits in anorexia nervosa has also garnered initial support. Research has also begun to explore third-wave cognitive behavioral therapy approaches such as acceptance and commitment therapy (ACT), which aims to increase psychological flexibility or the ability to live life according to one’s personal values despite psychological distress. The effect of these therapies on outcomes in anorexia nervosa requires additional study. While most psychotherapies for anorexia nervosa that have been empirically evaluated are rooted in cognitive-behavioral principles, focal psychodynamic therapy has been examined in a controlled trial, and again, seems to fare similarly to other therapies.

For children and adolescents with anorexia nervosa, family-based therapies have garnered substantial research support. As their name suggests, these therapies focus on involving parents and other family members in the treatment of the eating disorder. The "Maudsley Method," in particular, has demonstrated effectiveness across a number of research studies. In this approach, parents take on the responsibility of refeeding their child with the therapist's coaching and support. The parental role often mirrors that of staff members on an inpatient unit who are helping a patient achieve weight restoration. Parents help to enforce meal plans, caloric intake, and weight gain by providing empathic but firm support and supervision, particularly during mealtimes. As the child progresses through treatment, control over eating is ultimately transitioned back to them. The final stage of treatment focuses on ways in which the eating disorder has impacted the child or adolescent's development, with the goal of ultimately establishing a healthy adolescent identity. More recently, a related therapy, parent-focused treatment, has emerged. Parent-focused treatment is nearly identical to family-based therapy, but parents meet alone with a clinician, which may improve access to and acceptability of care. Preliminary studies support the efficacy of parent-focused treatment. It is also worth noting that a Maudsley anorexia treatment for adults (MANTRA), which targets interpersonal factors among other things, has recently been developed and shows promise.

Psychopharmacology. Although many medications have been studied for the treatment of anorexia nervosa, empirical support for most is lacking. The treatment of anorexia nervosa using antidepressants, particularly selective serotonin reuptake inhibitors (SSRIs), has received significant research attention. Nonetheless, studies have consistently demonstrated that, when compared to placebo, antidepressants do not significantly aid in weight gain, alter eating pathology, or reduce depressive or anxiety symptoms in underweight individuals with anorexia nervosa. Rather, weight restoration itself has been shown to reduce symptoms of depression and anxiety. A small number of studies have also examined the efficacy of SSRIs after weight restoration and similarly found no significant impact on relapse prevention or psychological outcome. Despite these findings, some individuals whose depressive symptoms do not diminish with weight restoration may benefit from antidepressant treatment following weight restoration.

Atypical antipsychotic medications have also been studied in the treatment of anorexia nervosa. In normal-weight individuals without eating pathology, atypical antipsychotic medication often results in extreme and unwanted weight gain, although the precise mechanism of this change remains unknown. In patients with anorexia nervosa, one atypical antipsychotic, olanzapine, has demonstrated efficacy in promoting weight gain with few side effects at a dose of 5 to 10 mg/day. Of note, treatment studies of olanzapine in patients with anorexia nervosa have demonstrated much more modest weight gains than are observed in noneating disorder psychotic patients taking the medication. Additionally, olanzapine does not appear to improve psychological symptoms of anorexia nervosa.

Ms. T, a 20-year-old woman, was admitted to an inpatient eating disorder unit at a height of 5 ft 7 in and a weight of 105 lb (BMI: 16.4 kg/m^2; 75% IBW). At the time of her admission, Ms. T described a history of disordered eating starting 2 years earlier, soon after entering college. The transition to college was stressful for Ms. T who felt homesick and struggled to make friends. She ran cross-country throughout high school, maintaining a BMI of around 19.5 kg/m^2 until college, when, in the context of discontinuing her exercise regimen and adjusting to buffet meals in the cafeteria, she gained approximately 10 lb. Ms. T experienced significant distress surrounding her weight gain and began eliminating snacks and desserts from her diet, successfully losing 5 lb. She enjoyed compliments from acquaintances regarding her weight loss and decided to resume regular exercise, eventually increasing her running to 5 to 10 miles daily. Her exercise quickly began to feel driven, and she noted feelings of intense guilt and increased dietary restriction when she missed a day. She also began to skip meals and avoided eating in the cafeteria. Within 2 months, she lost 20 lb (BMI: 18.0 kg/m^2; 82% IBW), her menstrual periods became irregular, and she struggled with low mood. She became intensely afraid of becoming fat and weighed herself twice daily to ensure that her weight had not increased. At this point, Ms. T was subsisting on two meals per day totaling approximately 500 to 700 kcal and consisting of foods that she deemed “safe” (e.g., fat-free Greek yogurt, grilled chicken, vegetables). She ate at a set time each day and weighed and measured each food item before consuming it, keeping a detailed food log. When her parents became aware of the eating disorder, Ms. T began to see a therapist, nutritionist, and medical doctor on an outpatient basis. She increased her weight by several pounds and her exercise and food rituals loosened. However, the following year, an ankle injury prevented her from running and Ms. T began to restrict her food intake more severely, losing to a BMI of 16.4 kg/m^2 (75% IBW). Subsequently, her treatment team recommended inpatient care with the goal of restoring her weight and becoming reacquainted with healthy eating habits.

Bulimia Nervosa

Interventions with strong empirical support exist for bulimia nervosa, although there is still substantial room for improvement as a significant portion of individuals continues to experience symptoms after receiving evidence-based treatment. Many individuals with bulimia nervosa respond well to and remain medically stable during the course of established outpatient treatments and, thus, inpatient care is indicated far less often than it is for anorexia nervosa.

Psychotherapy. Several psychotherapy approaches for bulimia nervosa have received research attention, with CBT emerging as the treatment of choice for the illness. CBT has demonstrated efficacy in both individual and group formats, with 30% to 50% of individuals reporting symptom abstinence following treatment and a higher percentage reporting symptom reduction. CBT for bulimia nervosa has also been adapted into pure and guided self-help formats, with both showing greater effectiveness than a waitlist control condition (although research has yet to determine how CBT self-help formats compare to the standard face-to-face modality).

CBT for bulimia nervosa consists of several phases, each meant to disrupt problematic behaviors (e.g., binge eating, purging, restricting) and identify and restructure distorted thoughts and beliefs that may be maintaining the unhealthy behaviors (e.g., “Eating sugary foods will make me fat"). In general, disordered behaviors are tackled first and, once they have improved, cognitive distortions become the focus of treatment. The initial phase of treatment often involves psychoeducation and working with the patient to increase the regularity of eating and resist urges to binge or purge. Self-monitoring and recording are key strategies in CBT and may rely on pen and paper or utilize web-based applications, which have become increasingly accessible and seemingly useful. In addition to self-monitoring and recording, patients receive various homework assignments to help identify and alter unhelpful behaviors and dysfunctional attitudes and beliefs. Patients are also taught to identify stressors or triggers and to employ more adaptive coping strategies. Finally, after symptoms have abated, relapse prevention strategies are used to reduce the likelihood of a return of symptoms by anticipating and preparing for stressful situations and setbacks likely to be encountered in the future.

In addition to CBT, several other psychotherapies have emerged as helpful in the treatment of bulimia nervosa. IPT, in particular, has been shown to reduce binge eating and purging, although research suggests that CBT does so more rapidly. Specifically, one study found that, at the end of treatment, a significantly higher percentage of individuals in the CBT group experienced symptom abstinence as compared to the IPT group (45% in CBT vs. 8% in IPT). However, individuals in the IPT group make additional gains after treatment and, 1 year later, essentially catch up to the CBT group in terms of rate of symptom abstinence (40% in CBT vs. 27% in IPT). As in anorexia nervosa, IPT for patients with bulimia nervosa focuses on dysfunctional interpersonal relationships.

While they have received less research attention, dialectical behavioral therapy (DBT), ACT, and integrative cognitive-affective therapy (ICAT) also show promise for the treatment of bulimia nervosa. Each draws from principles of CBT, but differs in certain ways. DBT focuses on helping patients develop mindfulness, emotion regulation, and distress tolerance in an attempt to regulate distressing emotions that may contribute to maladaptive eating behaviors. As described above, ACT helps individuals develop psychological flexibility by identifying personal values and learning skills for accepting uncomfortable internal experiences (e.g., anxiety) in the service of committing to living a values-driven life. ICAT, in addition to focusing on distorted cognitions and problematic behaviors as in CBT, emphasizes the role of dysfunctional interpersonal interactions on the eating disorder. Additional research is needed to further clarify the efficacy of these therapies in the treatment of bulimia nervosa.

Psychopharmacology. Research has identified several medications that effectively reduce symptoms in many patients with bulimia nervosa, particularly when administered in conjunction with evidence-based psychotherapy. Notably, research suggests that a course of drug therapy alone is inferior to a course of CBT alone, but some studies suggest that a combination of the two interventions may be more helpful than either alone. Thus, medication as the sole therapy for bulimia nervosa is typically not the first choice recommendation, but medication as an adjunctive treatment for bulimia nervosa has a strong evidence base.

Antidepressants in particular have amassed substantial research support for the treatment of bulimia nervosa and, of note, are equally effective for patients with and without depressive symptoms. Although several antidepressants appear to significantly reduce binge eating and purging, the SSRI fluoxetine has been studied the most extensively and is viewed as the medication of choice for this illness. Although the medication is traditionally prescribed at a dose of 20 to 40 mg/day for the treatment of depressive disorders, studies have found that a higher dose (60 mg/day) is more effective in disrupting binge eating and purging. Even at this higher dose, fluoxetine has fewer side effects than some of the other medications described below that have demonstrated efficacy in treating bulimia nervosa. Research has yet to determine the optimal length of fluoxetine treatment for bulimia nervosa as most trials have been relatively short.

Tricyclic antidepressant medications (TCAs), monoamine oxidase inhibitors (MAOIs), and atypical antiepileptic agents such as topiramate have also demonstrated efficacy in randomized controlled trials of bulimia nervosa. TCAs and MAOIs, however, are not viewed as a first-line treatment due to significant side effects and the risk of fatal overdose. Across several studies, topiramate has resulted in significant weight loss in patients with bulimia nervosa, a complicated side effect for an eating disorder patient that should be carefully considered prior to treatment initiation. Finally, bupropion is contraindicated for bulimia nervosa due to concerns regarding an increased risk for seizures in patients who are actively purging.

Binge Eating Disorder

Because both disorders are marked by recurrent episodes of binge eating, much of the treatment research for BED has focused on interventions that are known to effectively treat bulimia nervosa. However, the substantial portion of patients with BED who are overweight or have obesity may require additional treatment to achieve a healthier body weight, although clinicians tend to agree that binge eating behavior should be addressed before weight reduction is attempted. Additional studies examining interventions for BED are currently underway.

Psychotherapy. As in bulimia nervosa, CBT has emerged as a highly effective treatment for BED with over 50% of individuals attaining abstinence from binge eating following treatment. CBT for BED is structured quite similarly to CBT for bulimia nervosa, without the focus on inappropriate compensatory behaviors. Despite its efficacy in reducing binge eating in BED, studies have found that CBT does not effectively result in weight loss for individuals with overweight or obesity.

IPT may be even more useful for BED than it is for bulimia nervosa. In fact, research examining IPT versus CBT for BED suggests that the two interventions may be equally efficacious both at the end of treatment and in follow-up. DBT, too, may be effective in the treatment of BED, but additional studies are needed to determine this therapy’s impact on this population.

Psychopharmacology. Many of the medications that are useful for bulimia nervosa also reduce binge eating in BED, although long-term studies have yet to examine whether these treatment gains persist over time. Specifically, SSRIs and TCAs compare favorably to placebo in eliminating binge eating but do not aid in weight loss. Because it both reduces binge eating and causes weight loss, topiramate appears to serve as an effective treatment for BED, as well. Weight loss medications have not been extensively studied in the treatment of BED, although one, orlistat, seems to promote weight loss but does not reduce binge eating. Other appetite-suppressing and weight-reducing medications have been considered for this population, including stimulants. Most recently, lisdexamfetamine has been approved by the FDA for short-term treatment of BED.

Weight Loss Treatment. Treatment of the obesity associated with BED requires both nutritional modification with calorie reduction and an increase in exercise. For severe obesity, especially class III (BMI ≥40 kg/m^2), medical supervision and, possibly, bariatric surgery may be recommended.

ARFID

Prior to the inclusion of ARFID in the Feeding and Eating Disorders section of DSM-5, very little research had investigated effective treatments for this disorder. Since the publication of DSM-5, preliminary treatment research has been conducted, primarily in children and adolescents. Overall, these initial studies suggest that CBT shows promise for children, adolescents, and adults with ARFID. Some patients with ARFID, similar to those with anorexia nervosa, may be medically compromised and require a higher level of care than outpatient treatment to achieve nutritional rehabilitation and/or weight restoration. In other cases, when medical sequelae are less severe, an outpatient medical team can appropriately manage the needs of the patient. There are not yet consensus guidelines regarding the medical management of patients with ARFID, but many of the practices described above for patients with anorexia nervosa can be effectively used for this population.

Pica and Rumination Disorder

Very little research has investigated effective treatments for pica or rumination disorder. Behavioral strategies that have been identified as useful for other disturbances of eating behavior are often "borrowed" for treatments that aim to normalize behaviors in individuals with these disorders.

FURTHER READINGS

Arcelus J, Mitchell AJ, Wales J, Nielsen S. Mortality rates in patients with anorexia nervosa and other eating disorders: a meta-analysis of 36 studies. *Arch Gen Psych*. 2011;68(7):724–731.

*Attia E, Walsh BT. Behavioral management for anorexia nervosa. *NEJM*. 2009;360(5):500–506.

Becker AE, Burwell RA, Gilman SE, Herzog DB, Hamburg P. Eating behaviours and attitudes following prolonged exposure to television among ethnic Fijian adolescent girls. *Br J Psychiatry*. 2002;180(6):509–514.

Brigham KS, Manzo LD, Eddy KT, Thomas JJ. Evaluation and treatment of avoidant/restrictive food intake disorder (ARFID) in adolescents. *Curr Pediatr Rep*. 2018;6(2):107–113.

Bulik CM, Blake L, Austin J. Genetics of eating disorders: what the clinician needs to know. *Psychiatr Clin N Am*. 2019;42(1):59–73.

*Culbert KM, Racine SE, Klump KL. Research review: what we have learned about the causes of eating disorders—a synthesis of sociocultural, psychological, and biological research. *J Child Psychol Psychiatry*. 2015;56(11):1141–1164.

Franko DL, Tabri N, Keshaviah A, et al. Predictors of long-term recovery in anorexia nervosa and bulimia nervosa: data from a 22-year longitudinal study. *J Psychiatr Res*. 2018;96:183–188.

*Galmiche M, Déchelotte P, Lambert G, Tavolacci MP. Prevalence of eating disorders over the 2000–2018 period: a systematic literature review. *Am J Clin Nutr*. 2019;109(5):1402–1413.

Garber AK, Cheng J, Accurso EC, et al. Short-term outcomes of the study of refeeding to optimize inpatient gains for patients with anorexia nervosa: a multicenter randomized clinical trial. *JAMA Pediatr*. 2021;175(1):19–27.

Garber AK, Sawyer SM, Golden NH, et al. A systematic review of approaches to refeeding in patients with anorexia nervosa. *Int J Eat Disord*. 2016;49(3):293–310.

Gorrell S, Loeb KL, Le Grange D. Family-based treatment of eating disorders: a narrative review. *Psych Clin*. 2019;42(2):193–204.

Gorrell S, Murray SB. Eating disorders in males. *Child Adolesc Psychiatr Clin N Am*. 2019;28(4):641–651.

*Hay P. A systematic review of evidence for psychological treatments in eating disorders: 2005–2012. *Int J Eat Disord*. 2013;46(5):462–469.

Heaner MK, Walsh BT. A history of the identification of the characteristic eating disturbances of bulimia nervosa, binge eating disorder and anorexia nervosa. *Appetite*. 2013;65:185–188.

*Himmerich H, Kan C, Au K, Treasure J. Pharmacological treatment of eating disorders, comorbid mental health problems, malnutrition and physical health consequences. *Pharmacol Ther*. 2021;217:107667.

Kaplan AS, Walsh BT, Olmsted M, et al. The slippery slope: prediction of successful weight maintenance in anorexia nervosa. *Psychol Med*. 2009;39(6):1037–1045.

*Klein DA, Walsh BT. Eating disorders: clinical features and pathophysiology. *Physiol Behav*. 2004;81(2):359–374.

Klump KL. Puberty as a critical risk period for eating disorders: a review of human and animal studies. *Horm Behav*. 2013;64(2):399–410.

Murray HB, Bailey AP, Keshishian AC, et al. Prevalence and characteristics of avoidant/restrictive food intake disorder in adult neurogastroenterology patients. *Clin Gastroenterol Hepatol*. 2020;18(9):1995–2002.

Murray HB, Thomas JJ, Hinz A, Munsch S, Hilbert A. Prevalence in primary school youth of pica and rumination behavior: the understudied feeding disorders. *Int J Eat Disord*. 2018;51(8):994–998.

*Pisetsky EM, Schaefer LM, Wonderlich SA, Peterson CB. Emerging psychological treatments in eating disorders. *Psychiatr Clin N Am*. 2019;42(2):219.

Schebendach J, Mayer LE, Devlin MJ, Attia E, Walsh BT. Dietary energy density and diet variety as risk factors for relapse in anorexia nervosa: a replication. *Int J Eat Disord*. 2012;45(1):79–84.

Silén Y, Sipilä PN, Raevuori A, et al. DSM-5 eating disorders among adolescents and young adults in Finland: a public health concern. *Int J Eat Disord*. 2020;53(5):520–531.

Smink FRE, van Hoeken D, Hoek HW. Epidemiology of eating disorders: incidence, prevalence and mortality rates. *Curr Psychiatry Rep*. 2012;14(4):406–414.

*Solmi M, Wade TD, Byrne S, et al. Comparative efficacy and acceptability of psychological interventions for the treatment of adult outpatients with anorexia nervosa: a systematic review and network meta-analysis. *Lancet Psychiatry*. 2021;8(3):215–224.

*Svaldi J, Schmitz F, Baur J, et al. Efficacy of psychotherapies and pharmacotherapies for bulimia nervosa. *Psychol Med*. 2019;49(6):898–910.

Udo T, Grilo CM. Prevalence and correlates of DSM-5–defined eating disorders in a nationally representative sample of US adults. *Biol Psychiatry*. 2018;84(5):345–354.

van Elburg A, Treasure J. Advances in the neurobiology of eating disorders. *Curr Opin Psychiatry*. 2013;26(6):556–561.

van Hoeken D, Hoek HW. Review of the burden of eating disorders: mortality, disability, costs, quality of life, and family burden. *Curr Opin Psychiatry*. 2020;33(6):521–527.

*Voderholzer U, Haas V, Correll CU, Körner T. Medical management of eating disorders: an update. *Curr Opin Psychiatry*. 2020;33(6):542–553.

23

Sleep Disorders

Amir Sharafkhaneh, M.D., Ph.D., Habibolah Khazaie, M.D., and Max Hirshkowitz, Ph.D.

OVERVIEW

To understand sleep and its disorders, it helps to begin with sleep's three basic characteristics. (1) Sleep is a process required for proper brain function. Failure to sleep impairs thought processes, mood regulation, and a host of normal physiologic functions. (2) Sleep is not a single process; there are several distinct types of sleep. These different types of sleep differ both qualitatively and quantitatively. Each type of sleep has unique characteristics, functional importance, and regulatory mechanisms. Selectively depriving one particular type of sleep produces compensatory rebound when an individual is allowed to sleep *ad lib*. (3) Sleep is not a passive process; sleep can be associated with a high degree of brain activation and metabolism.

Several physiologic mechanisms regulate sleep, and when these systems go awry, sleep disorders occur. Initial interest in sleep disorders emanated from psychiatrists, psychologists, and neurologists. During the past five decades, medical research expanded sleep medicine into a multidisciplinary field. Research confirming sleep-disordered breathing's medical consequences attracted many pulmonary, internal medicine specialists, and cardiologists to the field. Perioperative risks associated with patients with sleep-disordered breathing attracted the attention of anesthesiologists, and treatment options for breathing disorders now involve surgeons and dentists. Sleep–wake cycle related endocrinology and circadian rhythm research migrated from the laboratory bench to the bedside. Nonetheless, the seriousness of sleep disorders remains poorly recognized by the general public and many clinical practitioners. Progress, however, is being made. *Sleep Medicine* is now a recognized medical specialty, and ACGME accredited Sleep Medicine Fellowship Training programs exist.

Some sleep disorders are both dangerous and their consequences can be costly. Hypersomnolence is a serious, potentially life-threatening condition affecting not only the sleepy individual but also their family, coworkers, and society in general. In fact, sleep-related motor vehicle accidents represent a major public safety concern, with some states enacting criminal statutes to deter sleepy driving. Investigations link many major industrial catastrophes to sleepiness. Research shows that sleep-disordered breathing contributes to hypertension, heart failure, and stroke. Sleep disorders' direct cost per annum in the United States is estimated at 16 billion dollars, with indirect costs ranging upward to more than 100 billion dollars.

SLEEP DISORDERS CLASSIFICATIONS

Three different nosologies provide classification systems for sleep disorders: (1) the Diagnostic and Statistical Manual of Mental Disorders (DSM), (2) the International Classification of Sleep Disorders (ICSD), and the International Classification of Diseases (ICD).

The American Psychiatric Association's DSM-5 and its text revision (DSM-5-TR) represents the standard diagnostic system used by most physicians who are not Sleep Medicine Specialists. Sleep Medicine specialists often prefer using the ICSD-3 (which is cross-referenced to the required DSM-5-TR coding). The World Health Organization's ICD-10 is not as popular (in the United States). The three systems' organizational schemas differ. Some lump multiple nosologic entities into a single diagnostic classification while others split disorders, providing separate designations. In the present chapter, we will use DSM-5-TR as an organizational guideline because (1) it represents a diagnostic standard, (2) its coding system is widely used, and (3) it is the parsimonious classification system describing essential sleep–wake disorders. ICSD-3 is more comprehensive; however, diagnostic differentiation sometimes resembles research diagnostic criteria rather than clinical practice criteria. Where appropriate, ICSD-3 diagnostic entities will be discussed as clinical subtypes. ICSD-3 is the direct descendent of the first comprehensive sleep disorders nosology commissioned in 1979 by the Association of Sleep Disorders Centers (now the American Academy of Sleep Medicine) (see Table 23–1). Diagnostic criteria for DSM-5-TR and ICSD-3 are quite similar (and sometimes identical) for the major sleep disorders. Finally, we chose DSM-5-TR to organize this chapter because, after all, this is the Comprehensive Textbook of Psychiatry.

INSOMNIA DISORDER

Previously, DSM-IV defined insomnia as difficulty initiating sleep, maintaining sleep, or nonrestorative sleep for 1 month or more. ***The overall conceptualization of insomnia has changed dramatically, beginning with the dissatisfaction with sleep quantity or quality serving as the primary element.*** Three other paradigm shifts in thinking also include the following. First, nonrestorative sleep and its consequent sleepiness no longer represent a key feature for insomnia. Nonrestorative sleep is now considered an attribute of hypersomnolence. Second, whereas insomnia was previously considered a symptom secondary to some other underlying problem, it is now recognized as a possible independent condition. Third, the hitherto insomnias considered secondary are now considered comorbid. In the past, practitioners were admonished to treat insomnia's cause rather than the symptoms. There was an implicit notion that by doing so, the sleep problems would improve. Clinical experience suggested otherwise. Consequently, current therapeutics favor providing relief and managing symptoms. To some extent, this change in thinking parallels our getting away from monotherapy. In the past, it was argued that if insomnia was related to depression, treating the insomnia would mask the depression and thereby interfere with antidepressant treatment regimens. This does not appear to happen.

Diagnostic Criteria

Insomnia should be evaluated according to eight criteria. DSM-5-TR sets forth these as follows: (1) how it adversely affects sleep; (2) how it influences subsequent wakefulness; (3) its frequency; (4) its chronicity; (5) whether it occurs in conjunction with some other sleep disorders, (6) medication or substance use, or (7) some other psychiatric, neurologic, or medical conditions. Also, (8) the sleep difficulty must occur notwithstanding an adequate opportunity to sleep.

Table 23–1.
Outline of the International Classification of Sleep Disorders, Third Edition

I. Insomnia
    A. Chronic Insomnia Disorder
    B. Short-Term Insomnia Disorder
    C. Other Insomnia Disorder
    D. Isolate Symptoms and Normal Variants
        1. Excessive Time in Bed
        2. Short Sleeper

II. Sleep-Related Breathing Disorders
    A. Obstructive Sleep Apnea Disorders
        1. Obstructive Sleep Apnea, Adult
        2. Obstructive Sleep Apnea Disorder, Pediatric
    B. Central Sleep Apnea Syndromes
        1. Central Sleep Apnea with Cheyne–Stokes Breathing
        2. Central Apnea Due to a Medical Disorder without Cheyne–Stokes Breathing
        3. Central Sleep Apnea Due to High-Altitude Periodic Breathing
        4. Central Sleep Apnea Due to a Medication or Substance
        5. Primary Central Sleep Apnea
        6. Primary Central Sleep Apnea of Infancy
        7. Primary Central Sleep Apnea of Prematurity
        8. Treatment-Emergent Central Sleep Apnea
    C. Sleep-Related Hypoventilation Disorders
        1. Obesity Hypoventilation Syndrome
        2. Congenital Central Alveolar Hypoventilation Syndrome
        3. Late-Onset Central Hypoventilation and Hypothalamic Dysfunction
        4. Idiopathic Central Alveolar Hypoventilation
        5. Sleep-Related Hypoventilation Due to a Medication of Substance
        6. Sleep-Related Hypoventilation Due to a Medical Disorder
    D. Sleep-Related Hypoxemia Disorder
        1. Sleep-Related Hypoxemia
    E. Isolated Symptoms and Normal Variants
        1. Snoring
        2. Catathrenia

III. Central Disorders of Hypersomnolence

  - A. Narcolepsy Type 1
  - B. Narcolepsy Type 2
  - C. Idiopathic Hypersomnia
  - D. Kleine–Levin Syndrome
  - E. Hypersomnia Due to a Medical Disorder
  - F. Hypersomnia Due to a Medication or Substance
  - G. Hypersomnia Associated with a Psychiatric Disorder
  - H. Insufficient Sleep Syndrome
  - I. Isolated Symptoms and Normal Variants
    - 1. Long Sleeper
- IV. Circadian Rhythm Sleep–Wake Disorders
  - A. Delayed Sleep–Wake Phase Disorder
  - B. Advanced Sleep–Wake Phase Disorder
  - C. Irregular Sleep–Wake Rhythm Disorder
  - D. Non–24-Hour Sleep–Wake Rhythm Disorder
  - E. Shift Work Disorder
  - F. Jet Lag Disorder
  - G. Circadian Sleep–Wake Disorder Not Otherwise Specified
- V. Parasomnias
  - A. NREM-Related Parasomnias
    - 1. Confusional Arousals
    - 2. Sleepwalking
    - 3. Sleep Terrors
    - 4. Sleep-Related Eating Disorder
  - B. REM-Related Parasomnias
    - 1. REM sleep behavior disorder
    - 2. Recurrent Isolated Sleep Paralysis
    - 3. Nightmare Disorder
  - C. Other Parasomnias
    - 1. Exploding Head Syndrome
    - 2. Sleep-Related Hallucinations
    - 3. Sleep Enuresis
    - 4. Parasomnia Due to Medical Disorder
    - 5. Parasomnia Due to a Medication or Substance
    - 6. Parasomnia, Unspecified

        7. Isolated Symptoms and Normal Variants
            a. Sleep Talking
- VI. Sleep-Related Movement Disorders
    - A. Restless Legs Syndrome
    - B. Periodic Limb Movement Disorder
    - C. Sleep-Related Leg Cramps
    - D. Sleep-Related Bruxism
    - E. Sleep-Related Rhythmic Movement Disorder
    - F. Benign Sleep Myoclonus of Infancy
    - G. Propriospinal Myoclonus at Sleep Onset
    - H. Sleep-Related Movement Disorder Due to Medical Disorder
    - I. Sleep-Related Movement Disorder Due to a Medication or Substance
    - J. Sleep-Related Movement Disorder, Unspecified
    - K. Isolated Symptoms and Normal Variants
        - 1. Excessive Fragmentary Myoclonus
        - 2. Hypnagogic Foot Tremor and Alternating Leg Muscle Activation
        - 3. Sleep Starts (Hypnic Jerks)
- VII. Other Sleep Disorder

From American Academy of Sleep Medicine. *International Classification of Sleep Disorders*. 3rd ed. American Academy of Sleep Medicine; 2014.

The primary criteria descriptively relate to how insomnia affects sleep. Specifically, one or more of the following must be present: (a) difficulty initiating sleep, (b) difficulty maintaining sleep, or (c) early-morning awakening with an inability to resume sleep. In the case of children, insomnia can manifest as resistance to the caregiver’s designation of bedtime and/or difficulty sleeping without caregiver intervention. The influences on subsequent wakefulness in DSM-5-TR include “significant distress or impairment of social, occupational, educational, academic, behavioral, or other important areas of functioning.” ICSD-3 also considers fatigue; malaise; attention, concentration, or memory impairment; mood disturbance; irritability; behavioral problems (especially in children), including hyperactivity, impulsivity, or aggression; reduced motivation, energy, or initiative; and error or accident proneness.

Frequency and chronicity criteria stipulate sleep disturbance (with associated symptoms) three times per week (or more) for 3 months or longer. Based on DSM-5-TR, the insomnia must not be better explained by another sleep, psychiatric, neurologic, or medical disorder; must not be attributable to a medication or substance use; and must persist when an adequate opportunity for sleep occurs.

When the diagnosis of Insomnia Disorder is made, the clinician should also specify if there are comorbid conditions (be they sleep disorder, psychiatric, neurologic, or medical) and whether the insomnia occurs episodically (symptoms last one or more months but less than 3 months), persistently (3 months or more), or recurrently (two, or more, episodes in a 12-month period).

Insomnia Subtypes

Neither DSM-5-TR nor ICSD-3 formalizes diagnostic criteria for clinical and pathophysiologic insomnia subtypes. Nonetheless, ICSD-3 describes each briefly. These include psychophysiological insomnia, idiopathic insomnia, paradoxical insomnia, inadequate sleep hygiene, behavioral insomnia of childhood, insomnia comorbid with mental disorder, insomnia comorbid with medical condition, and insomnia due to drug or substance. ICSD-3 authors caution that “Despite the heuristic appeal of these various subtypes, it is often difficult to discriminate among them in clinical practice.” While this may be the case, paying attention to and understanding the overlapping complications of these subtypes can be essential for creating an effective treatment plan. The insomnia subtypes are described here to alert the clinician of possible insomnia nuances and complications.

Psychophysiological Insomnia. Psychophysiological insomnia involves conditioned arousal associated with the thought of sleeping. Objects related to sleep (e.g., the bed, the bedroom) likewise have become conditioned stimuli that evoke insomnia. Daytime adaptation is usually good (in contrast to impaired function in patients with psychiatrically related insomnia); however, there can be extreme tiredness, and the individual can become desperate. Psychophysiological insomnia often occurs in combination with stress and anxiety disorders, delayed sleep phase syndrome, and hypnotic drug use and withdrawal. Other features characteristic of psychophysiological insomnia include: (1)

excessive worry about not being able to sleep, (2) trying too hard to sleep, (3) rumination or inability to clear one's mind while trying to sleep, (4) increased muscle tension when getting into bed, (5) other somatic manifestations of anxiety, (6) ability to fall asleep when not trying to (e.g., when watching television), and (7) sleeping better away from own bedroom (including the sleep laboratory). In one sense, the individual has developed a performance anxiety concerning the ability to sleep, and autonomic nervous system conditioning reinforces the situation. Stimulus Control Therapy is well suited for treating psychophysiological insomnia (see behavioral treatments for insomnia). This nonpharmacologic approach is designed to improve the association between going to bed and being able to fall asleep. Because many patients with psychophysiological insomnia have developed poor sleep habits, improving sleep hygiene is usually beneficial. If muscle tension and rumination at bedtime are prominent features, relaxation therapy is a useful ancillary treatment.

Idiopathic Insomnia. Idiopathic Insomnia characterizes patients with a lifelong inability to obtain adequate sleep. The insomnia predates any psychiatric condition, and other etiologies must be ruled out or treated, including psychophysiological insomnia, environmental sleep disturbances, and practices constituting poor sleep hygiene. If the insomnia persists, then it is assumed that there is a defect in the neurologic mechanisms that govern the sleep–wake system. That is, the sleep homeostatic process is dysfunctional. Sleep restriction therapy may provide some benefits and will also give the clinician a clue to the limits of the patient's homeostatic process. Sleep restriction is a technique by which homeostatic drive is maximized by sleep schedule compression (see behavioral treatments). Unfortunately, patients with idiopathic insomnia will likely never have normal sleep. Treatment is difficult; however, improved sleep hygiene, relaxation therapy, and the use of hypnotic medicines are reportedly helpful. Although controversial, this group may require chronic pharmacotherapeutic intervention.

Paradoxical Insomnia. Paradoxical Insomnia, at its core, involves a dissociation between sleep and its usual attendant unconsciousness. In Paradoxical Insomnia, individuals think they are awake and have insomnia even though brain electrophysiologic activity pattern is consistent with the correlates of normal sleep. The discrepancy between subjective and objective sleep parameters is common in patients with various forms of insomnia; however, it is more extreme in paradoxical insomnia. That is, Paradoxical Insomnia is diagnosed when a patient complains of difficulty initiating or maintaining sleep, and no objective evidence of sleep disruption is found. For example, a patient sleeping in the laboratory reports taking more than an hour to fall asleep, awakening more than 30 times, and sleeping less than 2 hours the entire night. By contrast, the polysomnographic tracing showed sleep onset occurring within 15 minutes, few awakenings, a 90% sleep efficiency, and total sleep time greater than 7 hours. On many occasions, the patients report total lack of sleep with no or minimal daytime consequences. Paradoxical Insomnia can occur in individuals who are apparently free from psychopathology; however, it may represent a somatic delusion or hypochondriasis. Some patients with Paradoxical Insomnia have obsessional features with respect to somatic functions. Cognitive relabeling and/or diffusing the worry about being unable to sleep can help. Interestingly, anxiolytics can profoundly reduce the perception of sleeplessness without remarkably changing sleep physiologically.

In the past, paradoxical insomnia was called *pseudoinsomnia, sleep state misperception,* and *insomnia without objective findings*. These designations can be problematic because they imply invalidation of the complaint, thereby placing the clinician at odds with the patient. Making matters worse, this dissociation between the perception and physiology is very common in patients with schizophrenia. Interestingly, a study has shown an increased metabolic rate and maximum O_2 uptake in patients with *Paradoxical Insomnia* compared to normal controls. It is ironic that while most individuals readily acknowledge dreaming and understand that dreams constitute mental activity, they nonetheless are under the general mistaken impression that mental activity ceases during sleep. One might arrive at this conclusion based on the inability to recall anything between bedtime and morning arising. However, this confuses the amnesic property of sleep with lack of mentation. While it is true that, under normal circumstances, sleep and unconsciousness are coupled, in some cases these phenomena dissociate. Anxiety can provoke sleep-state misperception, and ruminative worry about not sleeping adds fuel to the fire. Thus if one lies in bed attempting to sleep and has ruminative thoughts, the individual may perceive themselves as awake, even though brain activity is consistent with the electrophysiologic correlates associated with normal sleep. A psychoeducational approach to informing the patient about this dissociation between brain activity and the perception of sleep can be helpful. Furthermore, the ruminations (which seem so urgent at the time but seem trivial in the morning, if remembered at all) may actually be a useful marker of sleep onset. If an individual can interpret the rumination as an indicator that they are asleep and it belays their worrying about not being able to sleep, the paradoxical insomnia often diminishes or even disappears.

Inadequate Sleep Hygiene. Inadequate Sleep Hygiene refers to insomnia produced by behaviors that are not conducive to good sleep. ICSD-1 defined inadequate sleep hygiene as “a sleep disorder due to the performance of daily living activities that are inconsistent with the maintenance of good quality sleep and full daytime alertness.” Many behaviors can interfere with sleep. Some of these behaviors increase arousal. For example, consuming caffeine or nicotine at night or engaging in excessive emotional or physical stimulation within a few hours of bedtime. Other behaviors interfere with sleep architecture, including daytime naps and a large variation of the daily sleep–wake schedule. It is widely believed that improving sleep hygiene can help relieve insomnia (see Table 23–2). It must be realized, however, that many activities related to sleep hygiene are lifestyle issues. Making a lifestyle change is difficult. Treatment should focus on only one or two problem areas at a time. Overwhelming the patient by attempting too many lifestyle changes or a complex regimen seldom succeeds.

Table 23–2.
Do's and Don'ts for Good Sleep Hygiene

	DO	DON'T
Maintain regular hours of bedtime and arising		
If you are hungry, have a light snack before bedtime		
Maintain a regular exercise schedule		
Give yourself an approximately an hour to wind down before going to bed		
If you are preoccupied or worried about something at bedtime, write it down and deal with it in the morning		
Keep the bedroom cool		
Keep the bedroom dark		
Keep the bedroom quiet		
Take naps		
Watch the clock so you know how bad your insomnia actually is		
Exercise right before going to bed in order wear yourself out		
Watch television in bed when you cannot sleep		
Eat a heavy meal before bedtime to help you sleep		
Drink coffee in the afternoon and evening		
If you cannot sleep, smoke a cigarette		
Use alcohol to help in going to sleep		
Read in bed when you cannot sleep		
Eat in bed		
Exercise in bed		
Talk on the phone in bed		

These recommendations are made when an individual experiences difficulty initiating sleep, maintaining sleep, or has early morning awakenings with difficulty resuming sleep. Additionally, most of these recommendations qualify as general good sleep health practices.

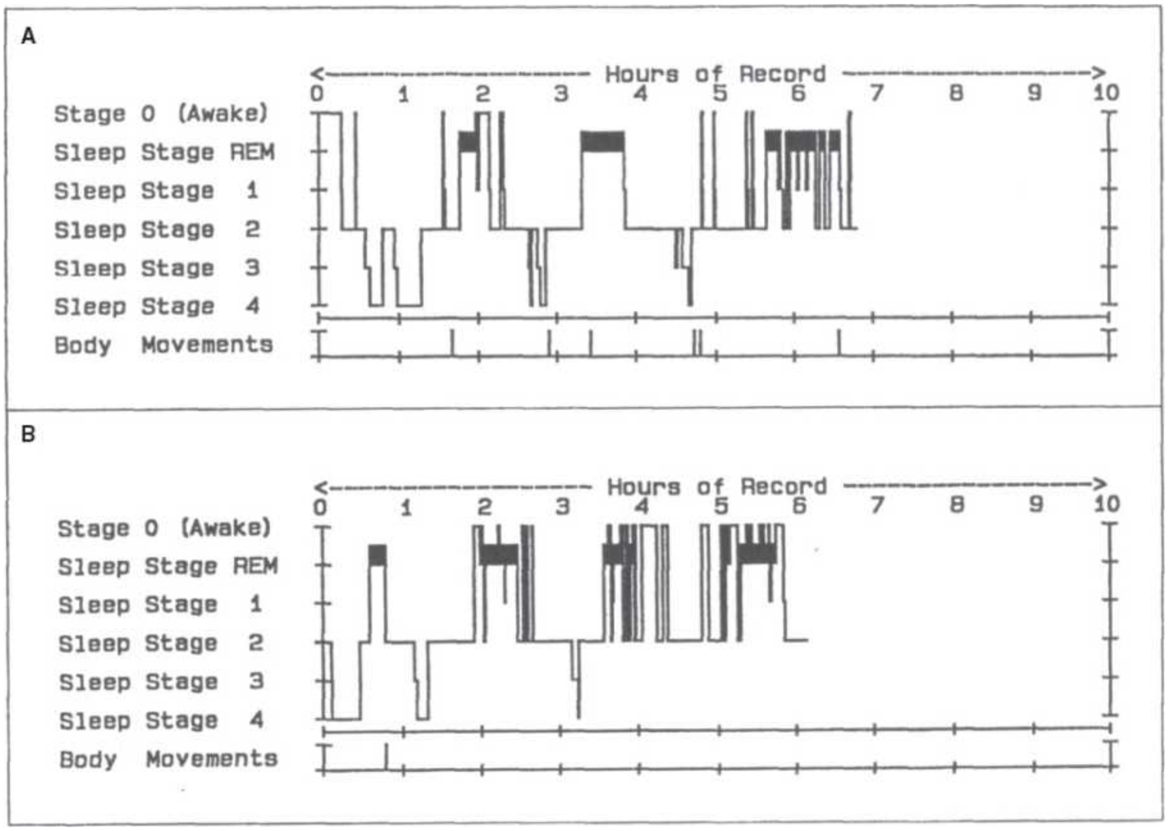

FIGURE 23–1. Sleep histograms showing normal sleep and insomnia. Sleep stage histograms comparing normal sleep (**Panel A**) with that found in a patient with major depressive disorders (**Panel B**). Difficulty maintaining sleep and early morning awakenings are common complaints in patients with depression. **Panel B** illustrates the electrophysiologic correlates of these complaints beginning, in this case, after approximately 2 hours of sleep. Sleep continuity becomes disrupted as morning approaches. Also present is a markedly reduced latency to REM sleep. This sleep feature is characteristic of this patient population and is thought by some to reflect cholinergic–aminergic imbalance.

Behavioral Insomnia of Childhood. Behavioral Insomnia of Childhood results from the child’s dependency on specific stimulation, objects, or setting for initiating sleep or returning to sleep. Without the presence of the teddy bear or parent, sleep onset is significantly delayed. Alternatively, without adequate limit setting by the caregiver, bedtime stalling (“Dad, I’m thirsty – can I get some water?”), or bedtime refusal (“I’m not tired! I don’t want to go to sleep!”) can ensue. *Limit-setting Sleep Disorder* occurs when a child takes advantage of an inadequately enforced bedtime schedule. Childhood insomnia, as it is sometimes called, involves a child refusing to or delaying going to bed when the caretaker stipulates. If unresolved, when the child reaches adolescence and can set their own schedule, inadequate sleep hygiene, insufficient sleep syndrome, or irregular sleep–wake pattern may become the problem. Delay strategies employed by the child include getting up to use the bathroom, requests for another drink, wanting another bedtime story, or complaints of hunger. Sometimes children will claim that they are afraid of the dark, monsters under the bed, or some other fictitious creation. The well-intentioned parent may be manipulated out of emotional concern. The problem often begins when the child can climb out of the crib or has been moved to a bed. The caregivers' knowledge and ability to properly set limits is critical; therefore, the education of caregivers can provide effective treatment.

Insomnia Comorbid With Mental Disorder. Insomnia Comorbid with Mental Disorder is the most common form of disorders of initiating and maintaining sleep. Thirty-five percent of patients seen in sleep disorders centers with insomnia as their chief complaint had a mental disorder, and half of these patients had a mood disorder.

Ninety percent of patients with **Major Depressive Disorder (MDD)** have insomnia. Insomnia is a risk factor (or marker) for depression on 3-year follow-up, conferring an odds ratio of 3.95 compared to an individual without insomnia. Insomnia is also an independent risk factor for suicide in patients with MDD, and 1 month after a trauma, insomnia is predictive of who will develop PTSD 1 year after trauma. It has been suggested that treating insomnia may help prevent mental disorders, especially depression. Sleep disturbances using EEG criteria in MDD include: (1) generalized sleep disturbance (increased sleep latency, increased nocturnal awakenings, and early morning awakenings), (2) slow wave sleep decreased in the first NREM-REM cycle with delta activity shifts to the second NREM period, (3) latency to REM sleep shortened, (4) REM sleep occurring earlier in the night, and (5) REM density increased (especially early in the night) (see Fig. 23–1). Selective REM sleep deprivation alleviates depression. Furthermore, this antidepressant action can persist for several days or more. Interestingly, most antidepressant medication suppresses REM sleep (including selective serotonin reuptake inhibitors, tricyclic antidepressants, and monoamine oxidase inhibitors). Arecoline (a cholinergic agonist) infusion will induce REM sleep in depressed patients, and scopolamine withdrawal mimics depression in normal subjects (including EEG sleep changes). These data support the

Cholinergic-Aminergic Imbalance theory of Major Depression. While it still holds that any drug capable of suppressing REM sleep has potential antidepressant properties, a couple of antidepressants do not suppress REM sleep (e.g., bupropion and nefazodone). Thus, the theory needs reformulation. Restated, REM suppression is sufficient but not necessary for antidepressant action. Treating insomnia while treating MDD may improve antidepressant effect. In a study on patients meeting DSM-IV criteria for both MDD and insomnia (*n* = 545), each received a morning dose of fluoxetine and were randomized to nightly eszopiclone 3 mg (ESZ+FLX) or placebo (PBO+FLX) for 8 weeks. Subjective sleep and daytime function were assessed weekly. Patients in the ESZ+FLX group had significantly decreased sleep latency, wake time after sleep onset, increased total sleep time, sleep quality, and depth of sleep. Eszopiclone co-therapy also produced significant and progressive improvement in mood when compared to fluoxetine alone. And the improvements of adjunctive insomnia therapy persisted for 2 weeks after the eszopiclone treatment was discontinued; fluoxetine was continued.

Bipolar disorder profoundly adversely affects sleep. Patients with mania and hypomania seldom complain of sleep problems even though they sleep only for a short duration (2 to 4 hours per night), have very prolonged latency to sleep, and sometimes have reduced slow wave sleep. When they enter the depressed phase, they are more likely to be seen at a sleep disorders center. Clinically, it is also noted that patients with bipolar disorder, if they break from their sleep–wake routine and stay up very late or all night, may provoke a manic episode.

Patients with **Schizophrenia** have no consistent EEG sleep changes except that they do not have REM sleep rebound in response to REM sleep deprivation. The only other consistent finding is that patients with schizophrenia frequently deny having slept even though EEG-EOG-EMG sleep appears normal. Patients with anxiety and personality disorders often have sleep-onset and sleep maintenance insomnia. Short REM sleep latencies and reduced slow wave sleep are reported in some studies; however, the finding is not specific to schizophrenia and is not consistent.

Patients with **Generalized Anxiety Disorder** worry excessively about multiple events, for more days than not, for at least 6 months; have difficulty controlling the worry; and the worrying causes at least three of the following: disturbed sleep, muscle tension, irritability, difficulty concentrating, restlessness, being easily fatigued. Although insomnia is one of the key symptoms, data addressing nonsedating anxiolytic efficacy for treating disturbed sleep are sparse. In a couple of randomized, placebo-controlled studies, the benzodiazepine receptor agonist (BZRA) zopiclone improved sleep onset and daytime anxiety when compared to benzodiazepines (BZDs). However, zopiclone is not available in the United States.

Insomnia Comorbid With Medical Condition. Insomnia accompanies many medical and neurologic conditions. Given pain's potential for disturbing sleep, all medical conditions producing pain can (and usually do) disturb sleep. Unfortunately, a synergy exists between pain and sleep such that poor sleep lowers the pain threshold. This vicious cycle can present a difficult treatment challenge. However, reducing pain can also improve sleep, and improving sleep can reduce pain. Thus, a two-pronged approach to pain and sleep can prove beneficial. In other medical conditions, the sleep disturbance appears to be secondary. For example, sleep-related gastroesophageal reflux disease (GERD) is often associated with insomnia. Treating the reflux results in sleep improvement but insomnia treatments seldom, if ever, relieve nocturnal GERD. Patients with chronic obstructive lung disease commonly suffer from both sleep-onset and sleep maintenance insomnia. Treating insomnia does not appear to ameliorate COPD. Neurodegenerative diseases are frequent associates of sleep disorders, including insomnia, hypersomnolence, parasomnias, and circadian rhythm disorders. Discussing specific medical and neurologic disorder associations with sleep disorders is beyond this chapter's scope, and the reader is referred to Chokroverty (2017) for details.

Insomnia Due to Drug or Substance. Many prescription drugs, even when taken properly, can disturb sleep. This includes antidepressants (i.e., fluoxetine), antihypertensives, corticosteroids, hypolipidemic medications, antiparkinsonian drugs, decongestants (i.e., pseudoephedrine), anorectics, stimulants, and antiepileptic medications. In addition, some sedating medications cause rebound insomnia when discontinued.

Alcohol and hypnotic use initially promote sleep onset because of their sedating properties. A problem occurs when sleep quality is adversely affected, tolerance develops after chronic use, or withdrawal begins.

Alcohol may relax a tense person and thereby decrease latency to sleep; however, sleep later in the night will usually be fragmented by arousals. As tolerance develops to the alcohol, greater amounts or more frequent dosing are needed to sustain the effects. Furthermore, during withdrawal or after tolerance develops, insomnia may rebound to a level more severe than the initial disturbance. If sleep is considered a behavior, and this behavior becomes dependent on the use of alcohol, then this is a form of alcoholism. Interestingly, for many patients with alcohol dependence, insomnia predates alcohol use and persists even during prolonged periods of abstinence.

Caffeine (the active ingredient in coffee) and theobromine (the active ingredient in chocolate) are methylxanthines, and they act as psychostimulants in the mammalian central nervous system (CNS). Psychostimulants increase sleep latency, reduce sleep efficiency, and decrease total sleep time. Caffeine’s half-life is 3 to 7 hours and may interfere with sleep when consumed in large quantities throughout the day or even in smaller portions closer to bedtime. Some individuals are hypersensitive to methylxanthine, and any coffee or chocolate can trigger difficulty falling asleep or awakening after a couple of hours of sleep with difficulty getting back to sleep. Finally, abuse of illicit substances, particularly stimulants (such as cocaine and amphetamines), interferes with sleep onset and sleep maintenance. Unlike with alcohol, discontinuation of these substances will cause hypersomnolence.

Short Sleeper. People who fall under this category require less than 5 hours of sleep per 24-hour period to maintain normal daytime functioning and mood. The patient may be distressed about not sleeping. However, if the short sleep duration does not provoke adverse daytime consequences, functional impairment, or is associated with relevant comorbid conditions, full Insomnia Disorder criteria are not met. Being a short sleeper appears to run in families, and mutations in two specific genes (DEC2 and ADRB1) and a variant of the BHLHE41 gene are seen in short sleeper families.

Diagnostic Workup of Insomnia

Diagnostic workup of insomnia may include thyroid function tests, glycosylated hemoglobin level, complete blood count, serum iron studies, ferritin level, liver function tests, renal functions tests, and other tests that may be indicated depending on the clinical presentations and comorbid conditions.

Sleep diaries are a cost-effective and reliable way to assess the sleep–wake cycle in individuals with insomnia. The most widely used self-evaluating questionnaires are the Insomnia Severity Index (ISI) and Epworth Sleepiness Scale. Polysomnography is not routinely recommended for the initial assessment of individuals with insomnia but is used in the assessment of patients suspected of suffering from other sleep disorders (such as sleep apnea, periodic limb movement disorder) associated with insomnia, and individuals with chronic insomnia disorders that fail to respond to treatment.

Treating Insomnia

In general, nonpharmacologic interventions are the first line for insomnia treatment. However, the use of pharmacologic agents is frequently requested by patients due to their faster effect, and therefore the clinician should be aware of the risk of dependency and try to shorten the duration of drug use, and ensure the focus of treatment is not removed from nonpharmacologic therapy. If nonpharmacologic therapy is successful, then the patient should start to gradually discontinue drug treatment at the first opportunity.

Nonpharmacologic Treatment With Cognitive Behavioral Therapy for Insomnia (CBTi). This treatment modality combines behavioral and cognitive techniques to overcome dysfunctional sleep behaviors and misperceptions, and distorted, disruptive thoughts about sleep. The therapist begins with a careful clinical interview to assess the insomnia’s etiology, chronicity, severity, associations, and comorbid conditions. A treatment plan is then designed using relevant and appropriate cognitive and behavioral techniques. These may include universal sleep hygiene, stimulus control therapy, sleep restriction therapy, relaxation therapies and biofeedback, cognitive therapy, and occasionally, paradoxical intention. With recent advances in technology and telehealth, various forms of CBTi can be delivered with no need for in-person visits.

Studies repeatedly show significant, sustained improvement in sleep symptoms with CBTi. Improvements include reduced wakefulness duration, decreased number of awakenings, and shorter latency to sleep onset. Short-term benefits are similar to that of medication, but CBTi tends to have lasting benefits, even 36 months after treatment. With the cessation of the medication, insomnia frequently returns and is sometimes accompanied by rebound insomnia. CBTi has not been shown to produce any adverse effects. There are no established “best practice” guidelines for the length or quantity of sessions. CBTi, however, is not without limitations. Most data do not compare the efficacy of the individual components (described below) of CBTi. However, sleep hygiene education alone does not appear to improve sleep. Intuitively, the multicomponent approach addresses many of the variables contributing to insomnia. Other limitations of CBTi include lack or unavailability of trained specialists, cost of sessions, and noncoverage by insurance carriers.

The effects of CBTi take longer to emerge than the effects of medications. Individuals are often desperate when they finally seek treatment for their insomnia. This makes it difficult to convince them to try a therapy that may require several weeks before it provides relief. Also, patients must be active participants in this type of therapy. Many individuals not only want a “quick fix,” but they also want to undergo a procedure or have something administered rather than be involved in the therapeutic process. For optimal CBTi, patients must commit to multiple sessions and be open to the idea that modifying thoughts and behaviors about sleep can improve insomnia symptoms. The “quick fix” model is common in primary care; by contrast, psychiatrists are accustomed to a delayed response from their experience with antidepressants and other psychotropic medicines. Consequently, psychiatrists may be more comfortable recommending CBTi than other practitioners.

Although firmly focused on cognitive and behavioral issues, it helps to extend CBTi just slightly into the psychodynamic sphere. For some patients with longstanding difficulty sleeping, being an insomniac becomes an important part of their identity. There may be primary or secondary gain to such identification. It is the negative emotional response (i.e., anger at the inability to control one’s sleep, feeling like a failure because one can’t sleep) to insomnia that contributes to its chronicity. In general, these individuals tend to internalize rather than express emotion, feel a heightened need for control, experience interpersonal difficulties, and have significant discontent with past events. For this subset of people, if the emotional response is not addressed, there is more likely to be a limited response to CBTi or a relapse of insomnia over time. The clinician who is attuned (i.e., to the patient’s tendency to view something as a failure rather than a challenge) will be better able to intercept barriers to treatment. “As chronic insomnia is an extremely psychologically painful condition, a high level of motivation and willingness to explore oneself, at least with a specific focus, is often present.”

UNIVERSAL SLEEP HYGIENE. The focus of universal sleep hygiene is on modifiable environmental and lifestyle components that may interfere with sleep as well as behaviors that may improve sleep. Especially because some of these behaviors are difficult to change, only one or two items that the patient and clinician collaboratively choose should be addressed at a time. This will give the patient the best chance at a successful intervention. Sleep enhancing directives are enumerated in Table 23–2. Often a few simple alterations in a patient’s habits or sleep environment can be effective. The clinician, however, needs to spend time reviewing both the patient’s routine and its irregularity. In some respects, the essence of insomnia is its variability. The day-to-day changes in behavior and the changing severity of sleeplessness can obscure the factors responsible for the problem. A carefully explained sleep hygiene program, with follow-up, represents a fairly inexpensive but effective intervention. Furthermore, improving sleep habits can enhance sleep even when the major cause of insomnia is physical.

STIMULUS CONTROL THERAPY. Stimulus control therapy is a deconditioning paradigm developed by Richard Bootzin and colleagues at the University of Arizona. This treatment aims to break the cycle of problems commonly associated with difficulty initiating sleep. By attempting to undo conditioning that undermines sleep, stimulus control therapy helps reduce both primary and reactive factors involved in insomnia. The rules attempt to enhance stimulus cues for sleeping and diminish associations with sleeplessness. The instructions are simple; however, they must be followed consistently. The first rule is to go to bed only when sleepy in order to maximize success. Second, use the bed only for sleeping. Do not watch television in bed, do not read, do not eat, and do not talk on the telephone while in bed. Third, do not lie in bed and become frustrated if unable to sleep. After a few minutes (do not watch the clock), get up, go to another room, and do something nonarousing until sleepiness returns. The goal is to associate the bed with rapid sleep onset. Rule three should be repeated as often as needed. The fourth and final instruction attempts to enhance the mechanisms underlying the circadian and sleep-wake cycles; that is, awaken at the same time every morning (regardless of bedtime, total sleep time, or day of the week) and totally avoid napping. Stimulus control therapy does work; however, you may not see results during the first few weeks or month. If continually practiced, the bouts of insomnia lessen in both frequency and severity.

SLEEP RESTRICTION THERAPY. Sleep restriction therapy is designed to increase sleep efficiency by decreasing the amount of time spent awake while lying in bed. Developed by Arthur Spielman, this therapy specifically targets those patients who lay awake in bed, unable to sleep. Restricting time in bed can help consolidate sleep. If the patient reports sleeping only 5 hours of a scheduled 8-hour time in bed, reduce the time in bed. However, it is advised not to reduce bedtime to less than 4 hours per night and to warn the patient about the hazards of daytime sleepiness. Sleep at other times during the day must be avoided, except for the older adults, who may take a 30-minute nap. The clinician then monitors sleep efficiency (time asleep as a percentage of the time in bed). When sleep efficiency reaches 85% (averaged over five nights), time in bed is increased by 15 minutes. Sleep restriction therapy produces a gradual and steady decline in nocturnal wakefulness.

RELAXATION THERAPY AND BIOFEEDBACK. The most important aspects of relaxation therapy are that it be performed properly and over-learned. Self-hypnosis, progressive relaxation, guided imagery, and deep breathing exercises are all effective if they produce relaxation. The goal is to find the optimal technique for each patient, but not all patients need help relaxing. **Progressive muscle relaxation** is especially useful for patients who experience muscle tension. Purposefully tense (5 to 6 seconds) and then relax (20 to 30 seconds) muscle groups beginning at the head and ending at the feet. Appreciate the difference between tension and relaxation. **Guided imagery** has the patient visualize a pleasant, restful scene, engaging all of their senses. **Breathing exercises** are practiced for at least 20 minutes per day for 2 weeks. Once mastered, use the technique once at bedtime for 30 minutes. If it does not work, try again another night. It is important that the technique not become associated with failure to fall asleep. **Abdominal breathing** is described below. The patient must become comfortable with each step before moving on to the next. First, in the supine position, breathe normally through your mouth or nose, whichever is more comfortable, and attend to your breathing pattern. Second, while maintaining that rhythm, begin to breathe more with your abdomen and less with your chest. Third, pause for a half-second after each breath cycle (in and out) and evaluate the breath. How did it feel? Was it smooth? Eventually, each breath will become uniform and smooth. Fourth, find a place where you can best feel the air move in and out. Concentrate on that spot and on the air moving in and out. Fifth, visualize intrusive thoughts as floating away; stop practicing and try again later if there are too many thoughts. **Biofeedback** provides stimulus cues for physiologic markers of relaxation and can increase self-awareness. The machine can measure muscle tension in your forehead or finger temperature. Finger temperature will rise when you become more relaxed. Patients require careful and adequate training; simply giving them an instruction tape is not especially helpful. Techniques are ideally mastered during the day for several weeks before application to the sleep problem; this is best achieved outside of the bed. By the time the techniques are applied in bed, the skill should be automatic. Relaxation techniques readily lend themselves to being combined with sleep hygiene and stimulus control therapies. Sometimes, they make for good distractions from a person thinking about their inability to sleep. The ruminations fuel insomnia and if the ruminator can be distracted, they may sleep better.

COGNITIVE THERAPY. This effective, validated treatment for a variety of psychiatric conditions, including major depression and generalized anxiety, has been adapted for use with insomnia. The cognitive aspect of insomnia treatment targets the negative emotional response to an appraisal of a sleep-related situation. The negative emotional response is thought to produce emotional arousal, which in turn contributes to or perpetuates insomnia. People who have maladaptive cognitions tend to exaggerate the negative consequences of insomnia: “there must be something really wrong with me if I can’t fall asleep in 40 minutes.” They also tend to have unrealistic expectations about their sleep requirements: “if I don’t sleep eight hours a night, then my whole day will be ruined.” The first step is to identify these cognitions, next challenge their validity, and finally substitute them with more adaptive cognitions (see Table 23–3). This is most easily illustrated by example, as in the box below.

Table 23–3.
Catastrophizing and Alternative Thoughts Scenario

Situation	Alarm clock goes off at 6:02 am Monday
Thought	“I hardly slept at all last night. I should be able to get more sleep. There must be something wrong with me. Great, I’ll be too tired to concentrate on anything today.”
Behavior/Feeling	Hit snooze to try to get 9 more minutes of sleep. Despair 70% Irritated 75%
Evidence	“Too tired to concentrate” For: I usually feel tired on Mondays. I lost my keys last week. Against: I just got a promotion, so I must be doing something right
Alternative balanced thought	“I did get some sleep last night and I usually can concentrate on my work.”
Feeling	Despair 50% Irritated 60%

A 42-year-old male presents to the office with a 5-year history of insomnia. He identifies being fired from his job and the birth of a colicky baby as precipitating factors in his inability to sleep. However, even after finding a new position with better hours and pay, and with the child now sleeping through the night, he continues to experience difficulty falling and staying asleep. Perpetuating factors include low back pain and a spouse with periodic limb movement disorder. He reports spending 8 to 9 hours in bed each night and sleeping only 4 to 5 hours intermittently. He watches 1 hour of TV in bed before turning out the light for bedtime. He spends hours watching his minutes tick away. He does not wake feeling rested and, when his alarm goes off, he is frequently already awake and has the thoughts, "I hardly slept at all last night. I should be able to get more sleep. There must be something wrong with me. Great, I'll be too tired to concentrate on anything today."

Examples of his maladaptive thoughts: ***"I should be able to get more sleep."*** This faulty appraisal of sleeping ability may relate to a need for control over sleep. This need for control interferes with having a more laissez-faire attitude about a few missed hours of rest. Such thoughts also lead to feelings of frustration and anger. ***"Great, I'll be too tired to concentrate on anything today."*** This is a misattribution of daytime impairment due to poor sleep. The patient is also magnifying the negative/ discounting the positive and engaging in black-and-white or all-or-nothing thinking. Could he be too tired to concentrate on some things but not all things? Might his inability to concentrate be due to a myriad of other factors? ***"There must be something wrong with me [if I can't get enough sleep]."*** Catastrophizing and emotional reasoning are at play here: just because you have a feeling, doesn't mean that the thought/feeling is true. A strongly held belief that sleeplessness negatively impacts physical and mental health can set off catastrophizing.

ACCEPTANCE AND COMMITMENT THERAPY (ACT). Acceptance and Commitment Therapy (ACT) includes six processes or skills and is another modality for the management of insomnia. These processes include Acceptance, Cognitive Defusion, Being Present, Self as Context, Values, and Committed Action. In fact, after receiving these skills, the patient increases psychological flexibility. Psychological flexibility is the ability to behave adaptively and in accordance with values despite having painful experiences. ACT has promising effects on insomnia, anxiety, and depression management.

PARADOXICAL INTENTION. This is a cognitive technique with conflicting evidence regarding its efficacy. In clinical practice, compliance is often a barrier, but it does work for a limited number of patients. The theory is that performance anxiety interferes with sleep onset. So when the patient tries to stay awake for as long as possible, rather than fall asleep, performance anxiety will be reduced, and sleep latency will improve.

Pharmacologic Treatment. For many years, BZDs were the most commonly prescribed sedative hypnotic medications for treating insomnia. BZRAs have replaced BZDs as the standard for sedative hypnotic medications used to treat insomnia. BZRAs include zolpidem, zaleplon, and eszopiclone. Subsequent development of a low-dose antihistamine (doxepin), melatonin receptor agonists (ramelteon and tasimelteon), and orexin antagonists (suvorexant and lemborexant) have more recently augmented the armamentarium of medications approved for treating insomnia. However, unapproved off-label (and not recommended) use of sedating antidepressants is common. Low-dose atypical antipsychotics (olanzapine, risperidone, and quetiapine) are effective in the treatment of paradoxical insomnia in patients who manifest with delusional features and do not respond to psychoeducation. These agents may also be helpful in patients with other subtypes of insomnia who have not responded to other sedative hypnotic agents. It seems they can be used to treat insomnia; however, the risk of adverse side effects such as metabolic syndrome should not be overlooked. Additionally, various over-the-counter (OTC) sleep aids are available. Nonprescription formulas include sedating antihistamines, protein precursors, and other substances. L-tryptophan was popular and readily available at health food stores until an outbreak of eosinophilia led to it being pulled off the shelves. Melatonin is the leader among self-administered food additives believed by some to alleviate sleeplessness. Essentially, low-dose melatonin can improve sleep when used to align or re-entrain the circadian rhythm with the desired sleep–wake schedule.

Prescription medicines are rigorously tested in clinical trials; therefore, they hold an advantage over the virtually untested OTCs. To attain FDA approval as a hypnotic, a medication must be safe and effective. Most hypnotic medications are approved for short, not long-term, use. Exceptions include zolpidem-modified release, eszopiclone, and ramelteon, all of which are approved for long-term therapy. When properly used, hypnotics can provide immediate and adequate relief from sleeplessness. Insomnia, however, usually returns upon discontinuation of dosing.

HYPERSOMNOLENCE DISORDER

The term hypersomnolence broadly refers to excessive sleepiness, an excessive quantity of sleep, and drowsiness with reduced vigilance during usual periods of wakefulness. Excessive sleepiness can be a serious, debilitating, potentially life-threatening condition. It affects not only the afflicted individual but also their family, coworkers, and the public at large. Although excessive sleepiness can arise from insufficient sleep, disrupted sleep, or desynchrony with one's circadian rhythm, Hypersomnolence Disorder is thought to stem from some basic neurologic sleep regulation dysfunction.

Hypersomnolence adversely affects attention, concentration, memory, and higher-order cognitive processes. Serious results of sleepiness include failure at school, loss of employment, motor vehicle accidents, and industrial disasters. The transportation industry is particularly prone to sleep-related accidents, including trucking, railroad, marine, and aviation. There are many sleep disorders associated with excessive daytime sleepiness; however, Obstructive Sleep Apnea Hypopnea (OSAH) is by far the most common dyssomnia seen in sleep disorder centers. By contrast, Hypersomnolence Disorder is a relatively rare condition.

Diagnostic Criteria

Hypersomnolence Disorder should be evaluated according to six criteria. DSM-5-TR sets these forth as follows: (1) how it influences wakefulness notwithstanding adequate sleep time (7 hours or more for the adult); (2) its frequency; (3) its adverse effect on functioning or emotional stability; (4) that the sleepiness is not better explained by whether it occurs in conjunction with some other sleep disorder, (5) medication or substance use, or (6) some other psychiatric, neurologic, or medical condition. Other conditions may be present (and should be specified); however, they must not better explain the hypersomnolence. All criteria must be met.

The primary criteria descriptively relate to how hypersomnolence affects wakefulness. Specifically, one or more of the following must be present: (a) recurrent sleep episodes or lapses into sleep within the same day; (b) a prolonged (9 hours or more) sleep episode that is not refreshing; (c) difficulty achieving full wakefulness upon abrupt awaking (sometimes called, *sleep inertia*). The frequency criterion stipulates that hypersomnolence must occur three times (or more) weekly for at least 3 months. The influence on wakefulness must be associated with "significant distress or impairment in cognitive, social, occupational, or other important areas of functioning." To diagnose Hypersomnolence Disorder, the symptoms must not be attributable to the use of a medication or substance.

When the diagnosis of Hypersomnolence Disorder is made, the clinician should specify if there are comorbid conditions (be they sleep disorder, psychiatric, neurologic, or medical) and whether the hypersomnolence duration is acute (less than 1 month), subacute (1 to 3 months' duration), or persistent (more than 3 months). The diagnosis should also include a specification for severity based on the patient's difficulty maintaining alertness while socializing, inactively sitting, reading, watching television, working, driving, or engaging in other activities. The difficulty can result from extreme drowsiness, nearly irresistible sleepiness, or actual lapses into sleep. The clinician should specify "mild" if difficulties occur 1 to 2 days per week, "moderate" if 3 to 4 days per week, and "severe" if 5 to 7 days per week.

Characteristics and Hypersomnolence Subtypes

Hypersomnolence Disorder presents in several forms. It may be associated with very long sleep periods after which the individual remains sleepy (see Fig. 23–2). The disorder can also occur with adequate but not long sleep periods. Sleep in patients with Hypersomnolence Disorder is usually well preserved, and sleep efficiency remains high even in forms associated with extended sleep schedules (12 hours or more). Furthermore, the patient readily falls asleep if given an opportunity to nap the following day. Sleep sometimes has elevated slow wave sleep; however, the EEG sleep pattern is essentially the same as that found in normal individuals who are sleep deprived. Unlike a sleep-deprived individual, the sleep pattern continues in this profile even after several nights of extended sleep. The etiology of Hypersomnolence Disorder is not known; however, a CNS cause is presumed. Some individuals with Hypersomnolence Disorder are HLA Cw2 positive, have autonomic nervous system dysfunctions, and have other affected family members. Some cases (approximately 10%) begin with a viral infection and quickly evolve into more permanent hypersomnolence (e.g., infectious mononucleosis, viral pneumonia, Guillain-Barre Syndrome). There also appears to be a familial form of Hypersomnolence Disorder with autosomal-dominance inheritance.

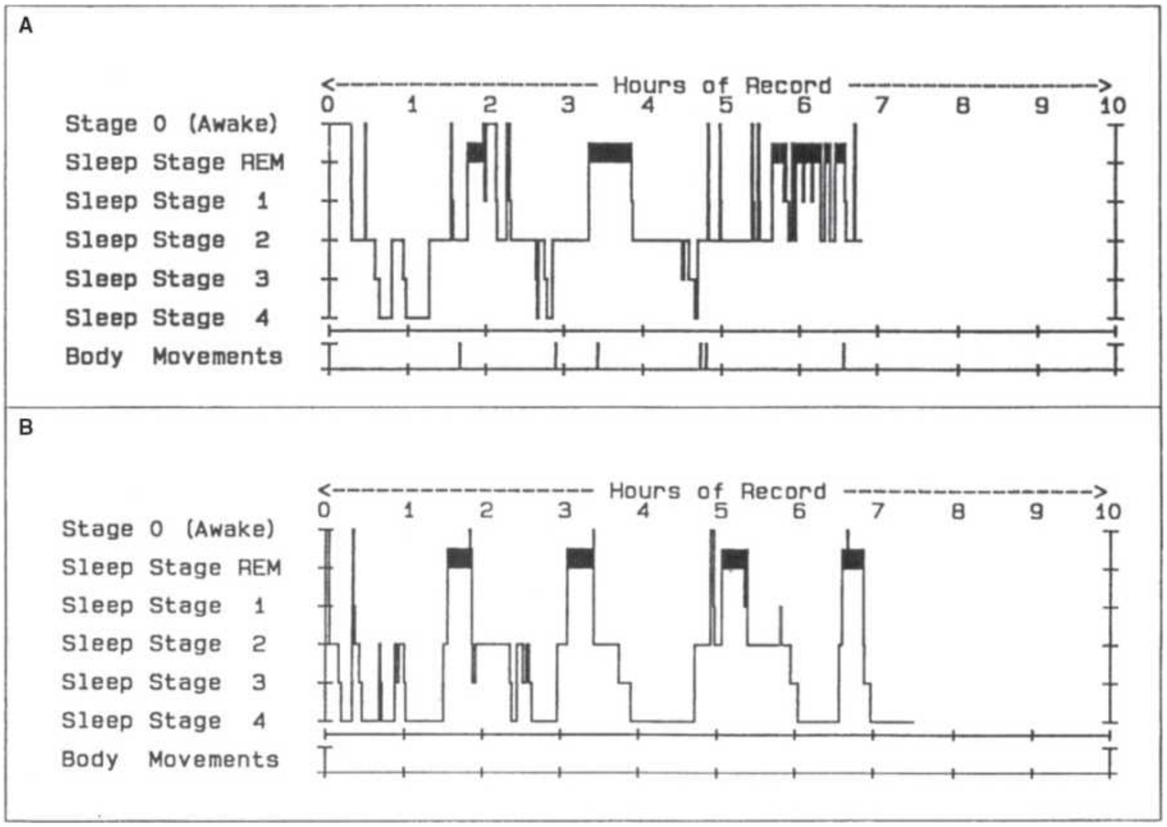

FIGURE 23–2. Sleep stage histograms comparing a normal sleep profile with that of a patient with hypersomnolence. **Panel A** illustrates the sleep pattern recorded in a healthy young adult. His sleep was interrupted several times by awakenings associated with other concurrent laboratory procedures; however, all parameters were within normal limits. **Panel B** depicts the sleep pattern of a patient with severe daytime sleepiness. The patient did not have narcolepsy or apnea. In the laboratory, sleep onset was rapid, sleep continued virtually without interruption, and he was extremely difficult to arouse. This recording was terminated after 7.5 hours to begin daytime testing; however, the patient's sleep revealed little indication of progressing toward a lighter state. In contrast to the normal sleep pattern, slow-wave sleep duration was elevated and it continued to emerge in alternation with REM sleep throughout the night. The patient awoke feeling unrefreshed and not satisfied with his sleep.

DSM-5-TR does not include descriptions of Hypersomnolence Disorder subtypes; however, ICSD-2 and ICSD-3 discuss some of these variants. Perhaps the intriguing of these is Kleine–Levin syndrome. Kleine–Levin predominantly afflicts males in early adolescence; however, it can occur later in life and in females. In its classic form, the recurrent episodes are associated with extreme sleepiness (18- to 20-hour sleep periods), voracious eating, hypersexuality, and disinhibition (e.g., aggression). Episodes typically last for a few days up to several weeks and appear once to 10 times per year. A monosymptomatic hypersomnolence form can occur. The frequency of the HLA DQB1*02 is increased in patients with this syndrome.

Medical conditions known to cause hypersomnolence include head trauma, stroke, encephalitis, Parkinson disease, inflammatory conditions, tumors, genetic diseases, and neurodegenerative diseases. Recurrent hypersomnolence can also occur related to the menstrual cycle and appears in months of menarche. The symptoms typically last for 1 week and resolve with menstruation. Treatment with oral contraceptives is effective, and therefore the disorder is thought to be secondary to a hormone imbalance.

Of course, sleepiness can be provoked by many prescription medications and recreational substances. Use or abuse of sedative hypnotics, sedating antihistamines, sedating antidepressants, antiepileptics, neuroleptics, and opioid analgesics can cause sleepiness. Hypersomnia may also be provoked by withdrawal from cocaine, amphetamines, caffeine, and nicotine.

Treating Hypersomnolence Disorder

Hypersomnolence Disorder is usually managed pharmacologically. Although there are no medications indicated for its treatment, stimulants are commonly used with varying degrees of success. The wake-promoting substances (modafinil, armodafinil, solriamfetol, and pitolisant) or traditional psychostimulants (amphetamines and their derivatives and methylphenidate) sometimes help manage the sleepiness. In addition to drug therapies, the overall therapeutic approach usually includes lifestyle adjustment, psychological counseling, drug holidays to reduce tolerance (if stimulants are used), and careful monitoring of refills, general health, and cardiac status.

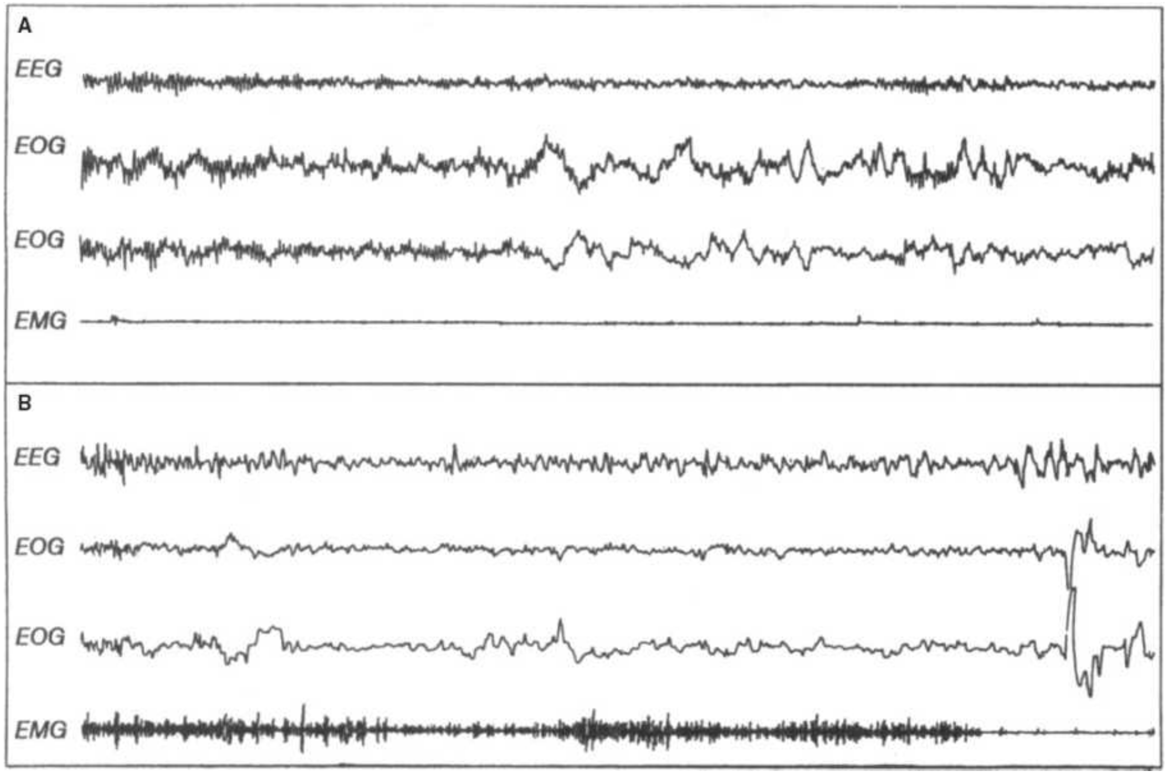

FIGURE 23–3. Polysomnographic tracing comparing normal sleep onset with that of a patient with narcolepsy. Each panel illustrates approximately 30 seconds of polysomnographic recording beginning with relaxed wakefulness. **Panel A** (normal sleep progression) shows a reduction of EEG alpha activity and development of slow rolling eye movements. **Panel B** shows the normally expected abatement of EEG alpha activity associated with increased theta activity and the appearance of a few slow eye movements. However, within 25 seconds (far right of figure) a swift loss of muscle tone occurs accompanied by rapid eye movements. This appearance of sleep-onset REM sleep (SOREM) characterizes narcolepsy and is part of the diagnostic criteria.

NARCOLEPSY

For decades, Narcolepsy was characterized by a tetrad of symptoms. These included (1) excessive sleepiness, (2) cataplexy, (3) sleep paralysis, and (4) hypnagogic hallucinations. Hypersomnolence and cataplexy always served as the key symptoms. Cataplexy involves the sudden loss of muscle tone, usually with continuing full or partial consciousness. Cataplexy is commonly precipitated by laughter. However, other strong emotions can "trigger" cataplexy. Cataplexy severity ranges widely from transient weakness in the knees to total paralysis while the patient is fully conscious. Episodes may last from several seconds to minutes. Usually, the patient cannot speak and may fall to the floor. According to DSM-5-TR, in children (or during the first 6 months of onset), cataplexy can be recognized by “spontaneous grimaces or jaw-opening episodes with tongue thrusting or a global hypotonia, without any obvious emotional triggers.”

Sleep paralysis and hypnagogic (or hypnopompic) hallucinations are nonspecific symptoms and may occur in isolation or in conjunction with other sleep disorders. Polysomnographic sleep studies typically revealed abnormal sleep architecture in which REM sleep occurs soon after sleep onset both at night and during daytime naps (see Fig. 23–3). This, in connection with the symptom tetrad, suggested narcolepsy-involved REM sleep intrusion, presumably resultant from dysfunction of REM sleep generator gating mechanisms. The tetrad features resembled REM sleep characteristics. Sleep paralysis is similar to the muscle atonia occurring during REM sleep and hypnagogic hallucinations resembling dreaming. Nocturnal sleep is often fragmented, and there can be considerable sleep disturbance. Patients may experience depression associated with narcolepsy, especially when not treated. Social isolation, difficulty with academics and employment, and fear of driving contribute to a sense of loss experienced by patients with narcolepsy.

The discovery that Narcolepsy is strongly associated with a hypocretin (orexin) deficit radically changed diagnostic practice. Recent works suggest that selective loss of hypocretin-producing neurons in narcolepsy with cataplexy may have an autoimmune nature. Narcolepsy diagnostic criteria now include variants without cataplexy. ICSD-3 has separate criteria for Narcolepsy with and without cataplexy. DSM-5-TR handles these variants with specific stipulations.

Diagnostic Criteria

The diagnostic criteria for narcolepsy presented in this section are based on the criteria set forth in DSM-5-TR (Table 23–4).

DSM-5-TR also indicates that the clinician should specify whether the diagnosis is: (a) Narcolepsy without cataplexy but with hypocretin deficiency, (b) Narcolepsy with cataplexy but without hypocretin deficiency, autosomal dominant cerebellar ataxia, deafness, and narcolepsy (because these conditions can be associated with low or intermediate CSF hypocretin levels), (c) Autosomal dominant narcolepsy, obesity, and type 2 diabetes (because patients with obesity and type 2 diabetes can have low CSF hypocretin levels), or (d) Narcolepsy secondary to another medical disorder (because some diseases or head trauma can be associated with hypocretin neuron destruction).

Table 23–4.
DSM-5-TR Diagnostic Criteria for Narcolepsy

A. Recurrent periods of an irrepressible need to sleep, lapsing into sleep, or napping occurring within the same day. These must have been occurring at least three times per week over the past 3 months.

B. The presence of at least one of the following:
   1. Episodes of cataplexy, defined as either (a) or (b), occurring at least a few times per month:
      a. In individuals with long-standing disease, brief (seconds to minutes) episodes of sudden bilateral loss of muscle tone with maintained consciousness that are precipitated by laughter or joking.
      b. In children or in individuals within 6 months of onset, spontaneous grimaces or jaw-opening episodes with tongue thrusting or a global hypotonia, without any obvious emotional triggers.
   2. Hypocretin deficiency, as measured using cerebrospinal fluid (CSF) hypocretin-1 immunoreactivity values (less than or equal to one-third of values obtained in healthy subjects tested using the same assay, or less than or equal to 110 pg/mL). Low CSF levels of hypocretin-1 must not be observed in the context of acute brain injury, inflammation, or infection.
   3. Nocturnal sleep polysomnography showing rapid eye movement (REM) sleep latency less than or equal to 15 minutes, or a multiple sleep latency test showing a mean sleep latency less than or equal to 8 minutes and two or more sleep-onset REM periods.

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 422, 423 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

The variant of narcolepsy, previously called *Narcolepsy Due to a Medical Condition,* often exhibits cataplexy. The cataplexy can be provoked by tumors or sarcoidosis of the hypothalamus; multiple sclerosis plaques affecting the hypothalamus; paraneoplastic syndrome with anti-Ma2 antibodies; Neimann–Pick type C disease (especially in children younger than 5); and possibly Coffin–Lowry syndrome. Cataplexy is usually absent in narcolepsy associated with multiple sclerosis, myotonic dystrophy, Prader–Willi syndrome, Parkinson disease, and multiple system atrophy. Narcolepsy due to a medical condition included what used to be called Posttraumatic Hypersomnia, which may or may not present with cataplexy.

Treating Narcolepsy

Currently, Narcolepsy cannot be cured; however, symptoms can be managed with either the wake-promoting substance modafinil (first-line treatment) or traditional psychostimulants like amphetamines and their derivatives (if modafinil fails). Two new additions to the stimulant group include pitolisant (a selective histamine H_3 receptor reverse agonist) and solriamfetol (a selective dopamine and norepinephrine reuptake inhibitor). Sodium oxybate has proven extremely effective for reducing cataplexy, even in cases where the cataplexy was thought to be intractable. Studies also suggest that sodium oxybate helps improve sleep and relieves some of the sleepiness associated with narcolepsy. In addition, REM sleep suppressing drugs (e.g., many antidepressants) may help cataplexy. This approach capitalizes on the anticholinergic REM sleep suppressant properties of these drugs. Since cataplexy is presumably an intrusion of REM sleep phenomena into the awake state, the rationale is clear. Many reports indicate that imipramine and protriptyline are quite effective for reducing or eliminating cataplexy. Selective serotonin reuptake inhibitors have gained popularity because they are associated with fewer side effects than the tricyclic antidepressants. The American Academy of Sleep Medicine recently released an updated guideline on treatment of central disorders of hypersomnolence (see Maski et al., 2021). Although drug therapies are the treatment of choice, the overall therapeutic approach should include scheduled naps, lifestyle adjustment, psychological counseling, drug holidays to reduce tolerance (if stimulants are used), and careful monitoring of refills, general health, and cardiac status.

SLEEP-RELATED BREATHING DISORDERS

Overview

Sleep-related breathing disorders include disorders along several different dimensions, including those involving (1) large airway obstruction during sleep, (2) breathing cessation resulting from central respiratory mechanisms, and (3) hypoventilation without breathing cessation. A sleep-related breathing cessation (**apnea**) is defined as a cessation of breathing for 10 seconds or more during sleep. Hypopnea is a reduction in breathing for 10 seconds or more during sleep. The 10 seconds duration is in adult patients. In the pediatric population, a respiratory event during sleep lasting longer than two times the respiratory cycle time is considered abnormal. These sleep-related breathing impairments are most often caused by airway obstruction; however, sometimes respiratory reduction results from central (brainstem) changes in ventilatory control, heart failure, or metabolic factors.

Cessation or reduced breathing during wakefulness is not intrinsically pathophysiologic. One can voluntarily hold one’s breath to avoid noxious odors. When speaking, an individual reduces airflow to the lungs. Neither of these activities constitutes dysfunction. However, when asleep, breathing cessations or reductions typically provoke CNS arousal, significant oxyhemoglobin desaturation, or both. Thus, the apnea or hypopnea ***induces*** a pathophysiologic event.

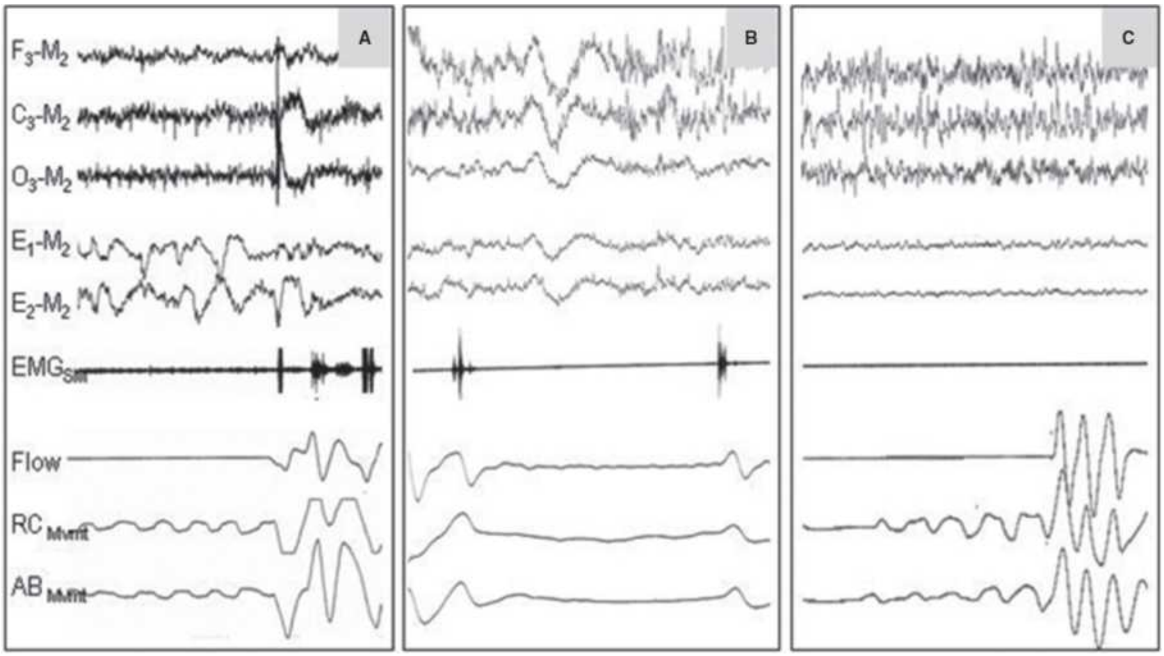

FIGURE 23–4. Polysomnographic tracings illustrating sleep apnea episodes. **Panel A** illustrates an obstructive sleep apnea, **Panel B** shows a central sleep apnea, and **Panel C** depicts a mixed sleep apnea. Note the presence of respiratory effort (rib cage and abdominal movement), notwithstanding cessation of airflow, throughout the obstructive apnea and during the latter portion of the mixed apnea. By contrast, respiratory effort is absent during the flow cessation throughout the central apnea and during the early portion of the mixed apnea. Notation: F3 (left frontal scalp derivation); M2 (right mastoid electrode location); C3 (left central scalp derivation); O3 (left occipital scalp derivation); E1 (left outer canthus eye electrode location); E2 (right outer cantus eye electrode location); EMGSM (surface electrode electromyogram from submentalis muscle); Flow (airflow measured with nasal–nasal–oral thermistors); RC Mvmt (rib cage movement measured with inductive pleythsmography); AB Mvmt (abdominal movement measured with inductive pleythsmography). (Reprinted from Hirshkowitz M, Sharafkhaneh A. Diagnostic assessment methods in adults. In: Kryger MH, Avidan AY, Berry RB, eds. *Atlas of Clinical Sleep Medicine*. 2nd ed. Elsevier Saunders; 2014:378.)

Each sleep apnea can be classified as obstructive, central, or mixed. Obstructive apnea results from airway collapse, notwithstanding continued respiratory effort. In central apnea, the breathing cessation results from decreased or absent respiratory effort. Mixed apnea typically begins with a central apnea, but when respiratory effort resumes, airway collapse prevents ventilation. Figure 23–4 illustrates obstructive, central, and mixed apnea episodes. A partial airway obstruction that produces a 4% (or more) oxyhemoglobin desaturation is scored as an obstructive hypopnea, according to Central Medicare Services (CMS) criteria. By contrast, partial airway obstruction not meeting the CMS desaturation criteria but provoking a brief CNS arousal can be scored as a respiratory effort–related arousal (RERA) according to AASM criteria. Figure 23–5 shows a polysomnographic tracing for obstructive hypopnea and RERA.

Obstructive Sleep Apnea Hypopnea (OSAH)

Diagnostic Criteria. Obstructive sleep-related breathing ranges across the spectrum from upper airway resistance syndrome to severe OSAH. **Adult Diagnostic Criteria:** An adult meets the criteria if an overnight polysomnographic sleep study or home sleep apnea testing (HSAT) shows a combined total of predominantly obstructive apnea and hypopnea of 15 (or more) per hour of sleep. However, if the patient has relevant symptoms or comorbid conditions, only 5 (or more) events per hour are required. Relevant symptoms include any of the following: hypersomnolence (given that the individual has adequate opportunity to sleep); witnessed cessation of breathing during sleep; gasping or snorting; or habitual snoring. Medicare and/or ICSD-3 also considers relevant comorbid conditions such as hypertension, mood disorder, cognitive dysfunction, coronary artery disease, stroke, congestive heart failure, atrial fibrillation, or type 2 diabetes as reasons to use the lower apnea and hypopnea criteria. ICSD also defines **Pediatric Diagnostic Criteria** as a child with overnight polysomnographic sleep showing a combined total of one (or more) obstructive apneas, mixed apneas, and hypopneas per hour of sleep. An alternative polysomnographic finding of obstructive hypoventilation (25% [or more] of total sleep time with hypercapnia [$PaCO_2$ > 50 mmHg]) in association with flattened nasal pressure tracings during inspiration. Either of these two polysomnographic findings, when associated with snoring, paradoxical movement of the chest wall and abdomen during sleep, sleepiness, behavioral or learning problems, or hyperactivity, makes the diagnosis. Recent years have seen a significant shift toward using HSAT with or without measurement of EEG and sleep stages for diagnosis of OSAH.

Comorbidities. In adults, obstructive sleep apnea is associated with many serious and chronic illnesses. A review of over four million records from the Veterans Health Administration (VHA) from 1998 to 2001 found a

2.91% (118,105) prevalence of sleep apnea in that population. Comorbid diagnoses included hypertension (60.1%), obesity (30.5%), diabetes mellitus (32.9%), cardiovascular disease including angina and myocardial infarction (27.6%), heart failure (13.5%), and stroke including transient ischemic attacks (5.7%). Psychiatric comorbidity in the sleep apnea group was significantly higher ($P < .0001$) than in the non-sleep apnea group for mood disorders, anxiety and posttraumatic stress disorder, psychosis, and dementia. There are several theories about the reason for this association. Psychiatric disorders may be a consequence of sleep apnea (and disturbed sleep and hypoxemia). Conversely, psychiatric disorders may predispose people to develop sleep disturbances such as sleep apnea. Pediatric obstructive sleep apnea is associated with neurobehavioral and cardiovascular problems. Obesity and growth abnormalities are also reported.

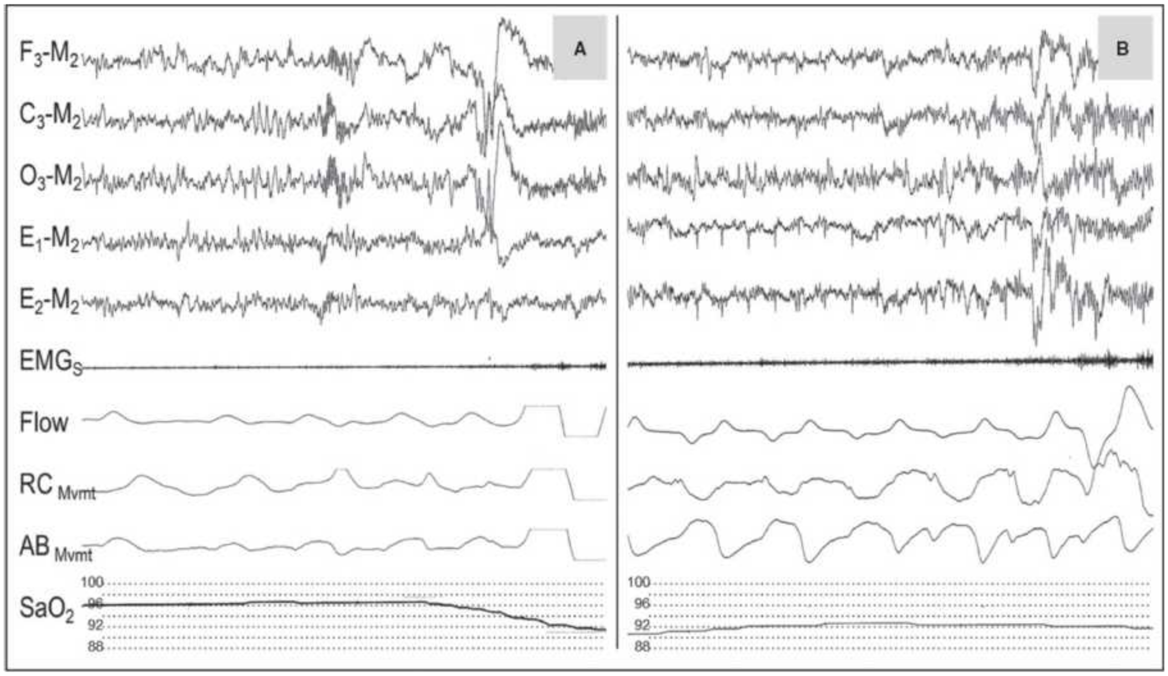

FIGURE 23–5. Sleep hypopnea and respiratory effort–related arousal (RERA) episodes. **Panel A** illustrates an obstructive hypopnea, **Panel B** shows a RERA. Note the presence of respiratory effort (rib cage and abdominal movement), notwithstanding reduced of airflow. The hypopnea episode is accompanied by significant oxyhemoglobin desaturation as the event progresses. By contrast, the RERA does not provoke desaturation; however, a CNS arousal occurs followed by increasing ventilation. Notation: F3 (left frontal scalp derivation); M2 (right mastoid electrode location); C3 (left central scalp derivation); O3 (left occipital scalp derivation); E1 (left outer canthus eye electrode location); E2 (right outer cantus eye electrode location); EMGSM (surface electrode electromyogram from submentalis muscle); Flow (airflow measured with nasal–nasal–oral thermistors); RC Mvmt (rib cage movement) measured with inductive pleythsmography; AB Mvmt (abdominal movement) measured with inductive pleythsmography; SaO_2 (pulse oximetry) shows oxygen saturation percentage. (Reprinted from Hirshkowitz M, Sharafkhaneh A. Polysomnography and home sleep test assessment methods in adults. In: Kryger MH, Avidan AY, Berry RB, eds. *Atlas of Clinical Sleep Medicine*, 3rd ed. Elsevier; 2024:447.)

Treatments. Interventions available for treating OSAH include (a) positive airway pressure (PAP), (b) oral appliances, (c) positional therapy, (d) surgical intervention, and (e) weight loss. Medication therapies have been tried for OSA but without success. Medroxyprogesterone acetate was originally thought to be helpful but is seldom used now. Likewise, tricyclic antidepressants sometimes decrease apnea severity by reducing REM sleep, the stage of sleep in which obstructive apnea is usually more frequent. Theophylline was reported to reduce apnea, but further study is needed. There were also animal studies suggesting mirtazapine and similar serotonin presynaptic affecting compounds improved breathing; however, studies in humans were disappointing. The only medications approved for use in patients with OSAH are the wake promoting substances including modafinil, armodafinil, and solriamfetol. These stimulant agents, however, do nothing to treat the pathophysiology of airway occlusion but rather are used as an adjunct for treating the residual sleepiness that persists in about 8% to 12% of patients who are otherwise well treated with and adequately use PAP therapy.

POSITIVE AIRWAY PRESSURE (PAP). Currently, PAP is the preferred treatment for OSAH. The PAP apparatus consists of a fan-driven blower, a nasal mask or oronasal interface, and tubing connecting the two. The airflow through the interface (usually a mask or nasal pillows) provides a positive pressure that offsets oropharyngeal collapse produced by inspiratory negative thoracic pressure. In this manner, it acts as a pneumatic splint, thereby maintaining the airway. Even the most severe sleep apnea can usually be alleviated when the pressure is properly titrated. Results are often dramatic. PAP devices come in several varieties. The most common systems provide a pre-set continuous positive airway pressure (CPAP). For individuals who find it difficult to exhale against continuous pressure, bi-level positive airway pressure (BPAP) may provide a solution. BPAP devices have different inspiratory and expiratory pressure settings. More recently, systems that sense the patient’s changes in airway resistance and automatically adjust the positive airway pressure (APAP) have been gaining popularity. Such APAP systems should theoretically be able to adapt to changes in pressure requirements produced by sleep deprivation, medications, weight change, sleep depth, illness, and aging. Finally, timed bi-level and servo-ventilation systems have also been developed but fall into the category of noninvasive positive pressure ventilation (NIPPV) systems that are more appropriate for treating other pulmonary diseases and breathing problems in neuromuscular diseases. PAP therapies’ tremendous efficacy and remarkable safety have made it the standard of care for patients who can tolerate sleeping with the machine (see Fig. 23–6). The major therapeutic challenge is utilization. Patient education and systematic follow-up are crucial. When problems with the interface, the pressure, nasal stuffiness, and other barriers to routine, nightly use arise, they must be remedied quickly to assure therapeutic adherence. The use of reminders improves PAP utilization. Current PAP technology allows telemonitoring of PAP use and adjustment of therapy if needed. When used properly, PAP therapy's success has rendered surgical intervention a secondary option, resorted to mainly after PAP failure, rejection, or nonadherence.

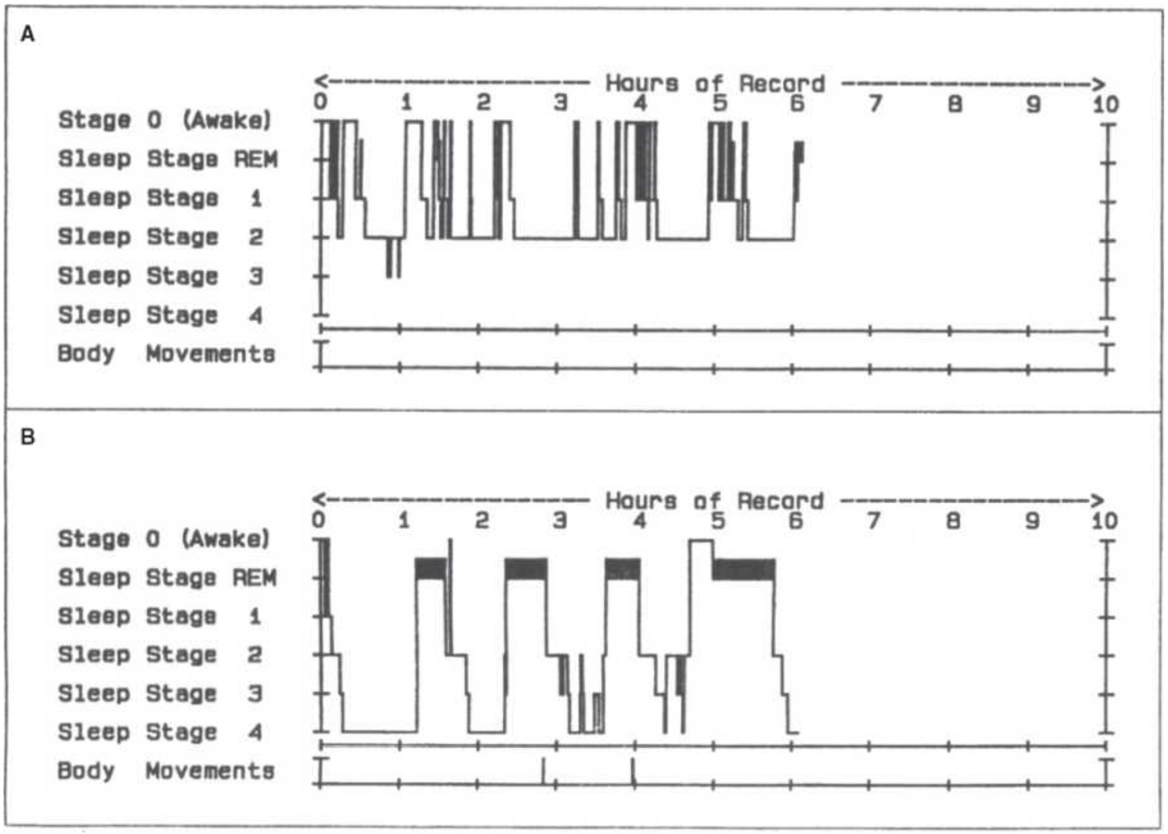

FIGURE 23–6. Sleep histogram showing CPAP-related sleep improvement. Sleep stage histogram illustrating the immediate, dramatic improvement in sleep architecture produced by treating obstructive sleep apnea with continuous positive airway pressure (CPAP) therapy. **Panel A** illustrates the abnormal sleep pattern on a night when the patient had more than 200 episodes of obstructive sleep apnea. Sleep is disturbed by frequent awakenings while REM and slow-wave (stages 3 and 4) sleep are nearly absent. **Panel B** shows data from the same patient being treated with CPAP on the next night. Normalization of sleep continuity with a massive rebound in REM and slow-wave sleep is evident.

ORAL APPLIANCE. Oral appliances represent another therapeutic option that is gaining popularity. A variety of oral appliances have also been developed to treat snoring and can benefit patients with mild-to-moderate

OSAH. The general approach is to manipulate the position of the mandible, lift the palate, and/or retain the tongue. Randomized trials indicate that some oral appliances sufficiently improve airway patency. Use of a titratable appliance is recommended over non–custom-made devices. However, in patients with severe OSAH, improvement does not always reach satisfactory levels; therefore, follow-up evaluations are critically important.

POSITIONAL THERAPY. In some cases, obstructed breathing occurs only when the patient sleeps supine. In such situations, preventing patients from sleeping on their backs may produce beneficial results. Tennis balls sewn onto or placed into pockets on the back of the nightshirt, foam wedges, or devices that alarm when the patient is supine may accomplish this goal. Although such interventions are found useful in clinical practice, large-scale systematic clinical trials are needed.

SURGICAL INTERVENTION. Aggressive surgical treatments evolved soon after OSAH's pathophysiologic and potentially life-threatening consequences were recognized. The earliest surgical intervention was designed to create a patent airway; thus, in the late 1970s tracheostomies were performed on individuals with severe apnea. There is little doubt that tracheostomy succeeds in creating an airway. Although no longer the preferred treatment, it remains a standard against which newer, more refined therapies are judged. Second-generation surgical approaches attempted to correct airway obstructions and malformations. Early studies of uvuolopalatopharyngoplasty (UPPP) suggested that modification of the soft palate effectively relieved most sleep apnea. Later follow-up results were less impressive. Approximately 30% to 50% of patients with sleep apnea benefit from UPPP. These patients are likely those with oropharyngeal obstruction; thus, careful attention to selection criteria presumably improves outcome. However, if obstruction occurs in the posterior airway space (PAS), maxillomandibular surgery may be appropriate. In retrognathic patients or patients with cephalometrics revealing compromised PAS, moving the jaw forward can achieve impressive normalization of breathing during sleep. Drug-induced sleep endoscopy (DISE) may assist in better identifying the site(s) of obstruction before planning surgical intervention, including multilevel or stepwise upper airway surgery. Upper airway stimulation by implanting a unilateral hypoglossal nerve stimulator is a new addition to surgical intervention for the management of OSAH.

A recently approved implantable device acts by stimulating the hypoglossal nerve, which produces tongue protrusion, opens the upper airways, and reduces the severity of sleep apnea. The device, called Inspire, is FDA approved and is placed by an ENT specialist.

WEIGHT LOSS. Weight loss is known to help many patients. However, because losing weight and maintaining weight loss is difficult and unreliably achieved, the prudent clinician should recommend weight loss but rely on other therapies.

Central Sleep Apnea

The key characteristic linking the various central sleep apnea syndromes is that breathing cessations are not provoked by upper airway flow limitations or obstructions. Breathing diminishes or stops due to a reduction or lack of respiratory effort. Several different etiologies can produce central respiratory events, including heart failure, high altitude, brainstem lesions, metabolic conditions, CNS depressants, congenital abnormalities, and PAP treatment.

Diagnostic Criteria. For diagnosing Central Sleep Apnea, DSM-5-TR requires polysomnographic evidence of five or more central apneas per hour of sleep. Also, the disorder must not be better explained by another concurrent sleep disorder. Clinicians should specify whether the central apnea is associated with (a) Cheyne–Stokes breathing, (b) comorbid opioid use, or (c) no evidence of airway obstruction. ICSD-3 describes eight diagnostic entities, as follows: (1) Central Sleep Apnea with Cheyne–Stokes Breathing, (2) Central Apnea Due to a Medical Disorder without Cheyne–Stokes Breathing, (3) Central Sleep Apnea Due to High Altitude Periodic Breathing, (4) Central Sleep Apnea Due to Medication or Substance, (5) Primary Central Sleep Apnea, (6) Primary Central Sleep Apnea of Infancy, (7) Primary Central Sleep Apnea of Prematurity, and (8) Treatment-Emergent Central Sleep Apnea. These conditions are described in the Central Apnea Subtypes section below.

Central Apnea Subtypes Described in ICSD-3

CENTRAL SLEEP APNEA WITH CHEYNE–STOKES BREATHING. Cheyne–Stokes Breathing is characterized by a unique breathing pattern consisting of prolonged hyperpnea events during which tidal volume gradually waxes and wanes in a crescendo–decrescendo fashion. The hyperpnea events alternate with apnea and/or hypopnea episodes that are associated with reduced ventilatory effort. This pattern is most common in older men with congestive heart failure or stroke. As with Primary Central Sleep Apnea, the patients present with daytime sleepiness or fatigue, insomnia, and awakening with shortness of breath. Polysomnography reveals five or more central apnea and hypopnea episodes per hour of sleep.

CENTRAL APNEA DUE TO A MEDICAL DISORDER WITHOUT CHEYNE–STOKES BREATHING. This form of central sleep apnea is usually caused by a brainstem lesion associated with a wide range of variable etiologies. Cardiac and renal disorders can also cause central apnea. Diagnostic criteria require a polysomnographically verified rate of five or more central apneas and hypopneas per hour of sleep.

CENTRAL SLEEP APNEA DUE TO HIGH ALTITUDE PERIODIC BREATHING. Central apnea at sleep onset is universal at elevations above 7,600 m but can occur at 1,500 m in some individuals (especially with a recent or rapid ascent). Periods of central apnea alternate with periods of hyperpnea in a 12- to 34-second cycle. This is an extension of normal respiratory control at sleep onset, where medullary pH receptors diminish their set-point and require lower pH to respond. At high altitudes, hyperventilation causes a hypocapnic alkalosis that reduces ventilation during sleep. Sleep architecture may suffer with increased duration in stages one and two and less slow wave sleep. REM sleep may not be affected. This condition can be treated with acetazolamide which lowers serum pH and increases the respiratory drive. Acetazolamide side effects include metabolic acidosis, electrolyte imbalance, anaphylaxis, Stevens–Johnson syndrome, toxic epidermal necrolysis, and agranulocytosis. Common reactions include but are not limited to fatigue, anorexia, taste changes, polyuria, diarrhea, melena, tinnitus, and photosensitivity.

CENTRAL SLEEP APNEA DUE TO MEDICATION OR SUBSTANCE. Central apnea episodes can be provoked by various drugs or drug combinations, most notably, long-acting opiates. An essential objective diagnostic criterion is a central apnea index (number of episodes per hour) of five or more per hour of sleep.

PRIMARY CENTRAL SLEEP APNEA. This is an idiopathic form of central sleep apnea. Patients typically have low normal $PaCO_2$ while awake and have a high ventilatory response to CO_2. They may present with daytime sleepiness, insomnia, and/or awakening with shortness of breath. Respiratory cessations during sleep occur independently of ventilatory effort. Polysomnography reveals five or more central apneas per hour of sleep.

PRIMARY CENTRAL SLEEP APNEA OF INFANCY. This form of central sleep apnea involves prolonged apneas or hypopneas with concomitant hypoxemia, bradycardia, or both. This condition afflicts preterm neonates presumably because their brainstems are not fully developed. The condition may be exacerbated by other medical problems that further compromise the infant's physiologic and developmental status.

PRIMARY CENTRAL SLEEP APNEA OF PREMATURITY. Central apnea, cyanosis, significant oxyhemoglobin desaturation, or bradycardia may be detected during postnatal hospital monitoring in infants less than 37 weeks conceptual age. Polysomnography or home monitoring must show either (i) recurrent 20 second or longer central apneas or (ii) periodic breathing for 5%, or more, of total sleep time.

TREATMENT-EMERGENT CENTRAL SLEEP APNEA. In patients with obstructive sleep apnea who are treated with PAP, central apnea episodes sometimes emerge. If the central apnea persists over time and lowering the pressure permits obstructions to re-occur (i.e., lower pressures are not adequate to treat the obstructive apnea), the patient is considered to have treatment-emergent central sleep apnea.

Sleep-Related Hypoventilation

Sleep-related hypoventilation is commonly associated with chronic obstructive pulmonary disease, obesity, and neuromuscular disease. However, other comorbid medical conditions may contribute. ICSD-3 differentiates hypoventilation disorders from hypoxemia disorders and further subdivides each. In hypoventilation disorders, ICSD-3 includes (1) obesity hypoventilation syndrome, (2) congenital central alveolar hypoventilation syndrome, (3) late-onset central hypoventilation and hypothalamic dysfunction, (4) idiopathic central alveolar hypoventilation, (5) sleep-related hypoventilation due to a medication or substance, and (6) sleep-related hypoventilation due to a medical disorder. ICSD-3 also lists (7) sleep-related hypoxemia as a separate nosologic entity. Snoring and catathrenia (sleep-related groaning) are considered isolated symptoms and normal variants.

Diagnostic Criteria. Polysomnography must demonstrate either (a) episodes with elevated CO_2 levels and decreased respirations and/or (b), if CO_2 is not monitored, episodes of persistently low oxyhemoglobin levels that are not provoked by apnea or hypopnea. Clinicians should specify if the hypoventilation is attributable to (i) congenital central alveolar hypoventilation, (ii) comorbid sleep-related hypoventilation (e.g., COPD), or if it is (iii) idiopathic hypoventilation. Severity should also be specified related to evidence of relevant diseases (e.g., right heart failure) or blood gasses (if available).

CIRCADIAN RHYTHM SLEEP–WAKE DISORDERS

This collection of sleep disorders shares the same basic underlying etiology: desynchrony between an individual’s internal circadian biologic clock and the desired sleep–wake cycle. The circadian (circa+dias, approximately one day) pacemaker is located in the suprachiasmatic nucleus (SCN). SCN firing oscillates with an almost sinusoidal pattern, the period of which is 24 hours, and the output correlates with the daily fluctuations in core body temperature. SCN firing patterns persist even in *isole* preparations. Mismatched circadian clock and desired schedules can arise from improper phase relationships between the two, travel across time zones, or dysfunctions in the basic biologic rhythm. DSM-5-TR provides six specific subtype coding for disorders listed as separate ICSD-3 diagnostic entities. Jet Lag Disorder is also listed in ICSD-3 but not in DSM-5-TR.

Diagnostic Criteria

To diagnose a Circadian Sleep–Wake Disorder, an individual must (1) have an alteration in or misalignment between their endogenous circadian rhythm and their desired sleep–wake schedule; (2) the alteration or misalignment must disrupt sleep-producing insomnia and/or hypersomnolence, and (3) this circumstance must either elicit significant distress or social, occupational, or other functional impairment. The clinician should specify if the Circadian Sleep–Wake Disorder subtype constitutes (a) Delayed sleep phase type, (b) Advanced sleep phase type, (c) Irregular sleep–wake type, (d) Non–24-hour sleep–wake type, (e) Shift work type, or (f) Unspecified type. The diagnosis should also indicate whether the problem is (a) **episodic** with symptoms lasting 1 month or more but less than 3 months; (b) **persistent** with symptoms lasting 3 months or more; or (c) **recurrent** with two (or more) episodes within a 1-year period.

Circadian Rhythm Sleep–Wake Disorder Subtypes

Delayed Sleep Phase Type. The Delayed Sleep Phase circadian disorder occurs when the biologic clock runs slower than 24 hours or is shifted later than the desired schedule. This produces a phase delay in the sleepiness-alertness cycle. Individuals with a delayed sleep phase are more alert in the evening and early nighttime, stay up later, and are more tired in the morning. These individuals are commonly referred to as *night owls*.

Advanced Sleep Phase Type. The Advanced sleep phase is when the circadian rhythm cycle is shifted earlier. Therefore, the sleepiness cycle is advanced with respect to clock time. Individuals with advanced sleep phase are drowsy in the evening, want to retire to bed earlier, awaken earlier, and are more alert in the early morning. Individuals with this pattern of advanced sleep phase are sometimes called *early birds or larks* (sometimes jokingly called *surgeons* in the medical center).

Irregular Sleep–Wake Type. The irregular sleep–wake pattern occurs when the circadian rhythm is absent or pathologically diminished. The sleep–wake pattern is temporally disorganized, and the timing of sleep and wakefulness is unpredictable. Individuals with this condition will have a normal amount of sleep during a 24-hour period; however, it will be fragmented into three or more episodes that occur irregularly. Long daytime naps and inappropriate nocturnal wakefulness will occur. Except in unusual circumstances, activities of daily life are significantly impaired.

Non–24-Hour Sleep–Wake Type. A patient may develop this type of circadian rhythm disorder when the circadian sleep–wake pacemaker has a cycle length greater or less than 24 hours and is not reset each morning. Under normal circumstances, resynchronization of our circadian rhythm occurs daily in response to the light-dark cycle. Problems occur incrementally when internal and environmental clocks become more and more out of phase. If the circadian clock's period is longer than 24 hours and does not reset each day, the patient experiences progressively worsening sleep-onset insomnia and daytime sleepiness. Sleep problems peak when circadian and environmental clocks are 12 hours out of phase and then begin to lessen, emulating progressively resolving advanced sleep phase. Eventually, the clocks correlate, and the sleep–wake cycle is normal for a few days, after which the insomnia–hypersomnia cycle begins again. For this reason, non–24-hour sleep–wake disorder has been called periodic insomnia and periodic excessive sleepiness.

Shift Work Type. Many service industries require 24-hour operation (e.g., transportation, health care). Similarly, as our cultures became more capital intensive, mining and manufacturing became around-the-clock enterprises. The number of individuals doing shift work has been increasing steadily for decades. Shift workers commonly suffer from insomnia, excessive sleepiness, or both. Some individuals require only a short time to adjust to a shift change, while others have great difficulty. Frequent shift rotation adds to the problems. Also, to meet social demands, shift workers will often adopt a non-shifted sleep–wake schedule on weekends and holidays. Even those individuals who try to stay shifted usually retain an unshifted circadian rhythm. The result can be severe insomnia when attempting to sleep and excessive sleepiness when attempting to remain awake. The result is profound sleep deprivation as circadian rhythm continues to be poorly entrained to the sleep–wake schedule. The natural low point in the normal sleep–wake rhythm occurs at approximately 3:00 to 5:00 A.M. Interestingly, this is precisely the time frame during which transportation and industrial accidents commonly occur as a direct consequence of sleepiness.

Jet Lag (not included in DSM-5-TR). With the advent of high-speed air travel, an induced desynchrony between circadian and environmental clocks became possible. Thus, the term jet lag came into use. When an individual rapidly travels across many time zones, they induce either a circadian phase advance or a phase delay, depending upon the direction of travel. Typically, one or two time zone translocations will not produce a sustained problem; however, overseas travel can be marked by great difficulty in adjusting one's sleep–wake routine. Individuals who frequently travel for business can find themselves quite impaired at the time they need to make important decisions. Furthermore, "owls" will experience greater difficulty adjusting to eastward travel because resynchronization requires phase advance. Similarly, larks theoretically will have more difficulty with westward travel. The number of time zones crossed is a critical factor. Normally, healthy individuals can easily adapt to 1- to 2-time zone changes per day; therefore, natural adjustment to an 8-hour translocation may take 4 or more days.

Circadian Rhythm Sleep–Wake Disorder Unspecified Type. During illnesses that keep patients bedridden, during hospitalizations, and in some forms of dementia, individuals often sleep *ad lib.* The resulting chaotic sleep–wake pattern adversely affects the circadian rhythm. The breakdown in the sleep–wake cycle may be further exacerbated by medication with sedative properties. Sleep in patients in the ICU is disturbed by noise, light, and the therapeutic and monitoring procedures being performed. The resulting disorganized sleep–wake pattern can produce a significant sleep disorder. Additionally, abuse of recreational street drugs (e.g., methamphetamine and ecstasy) is associated with individuals remaining awake overnight or continuously for several days at a time. These episodes of prolonged wakefulness ultimately produce periods of profound hypersomnia.

Treating Circadian Rhythm Sleep–Wake Disorders

Under normal circumstances, the internal circadian pacemaker is reset each day by bright light, social cues, stimulants, and activity. In cases where these factors fail to re-entrain the circadian rhythm, circadian sleep–wake disorders occur.

Light Therapy. Research indicates that light exposure can reset the circadian pacemaker. This seems especially the case when the light is bright (greater than ten thousand lux) or in the blue spectrum. The biologic clock can be reset with precise timing of bright light exposure. Exposure to light modifies the set-point of the biologic clock. Using core body temperature as a physiologic marker, one can use bright lights to produce phase delay when presented before the temperature nadir. By contrast, light exposure after temperature nadir evokes phase advance. The closer one presents light to the point of inflection (temperature nadir), the more robust the response in altering the cycle. Thus, early morning bright light therapy can be used to phase advance individuals with delayed sleep phase syndrome. Similarly, exposure to bright light in the evening can help patients with advanced sleep phase syndrome. More recently, it was discovered that the blue spectrum of light is the crucial ingredient in phase setting and shifting. Therefore, blue light exposure (or limitation with blue-blocking eyewear) at specific times can provide therapeutic benefits. Clinicians increasingly rely on light therapy to shift circadian sleep–wake rhythms of individuals with delayed and advanced sleep phases, shift workers, astronauts, and individuals experiencing jet lag.

Medication. Administering melatonin to individuals who are blind from birth successfully entrains their circadian sleep–wake rhythm. Researchers posit that melatonin secretion acts as the biologic substrate for the internal circadian oscillator. Under normal circumstances, melatonin levels begin to rise at dusk and remain elevated until dawn. Bright light suppresses the release of melatonin. Melatonin, in a sense, is the signal of darkness in the brain. As such, it can be used clinically to manage sighted patients with disturbed sleep-wake cycles. Melatonin is available over the counter in the United States. Prescription synthetic melatonin agonists (ramelteon and tasimelteon) are also available (FDA approved for treating patients with sleep-onset insomnia). A 2-mg sustained-release melatonin (Circadin) is available in many European countries, Australia, and some places in Asia. Interestingly, the wake-promoting compound modafinil is FDA approved for treating sleepiness in patients with shift work type circadian rhythm sleep–wake disorder.

Chronotherapy. Chronotherapy involves progressively phase delaying a person until the circadian oscillator is synchronized with the desired sleep-wake schedule. Most young and middle-aged adults have a propensity to progressively phase delay their circadian sleep–wake rhythm anyway. Thus, phase delaying each night by 2 or 3 hours is thought to be easier than phase advancing because it capitalizes on a natural tendency. Halting the phase delay at the appropriate moment and maintaining the desired synchrony can be challenging. The patient also has to cope with an odd sleep-wake schedule for the better part of a week during therapy (which can interfere with school or work). For these reasons, the development of light therapy and pharmacologic interventions has overshadowed chronotherapy in the past two decades.

PARASOMNIAS (AND SLEEP-RELATED MOVEMENT DISORDERS)

DSM-5-TR includes only three of the ten specific parasomnias described in ICSD-3. Furthermore, DSM-5-TR includes Restless Legs Syndrome (RLS) as a parasomnia, while in ICSD-3, it is classified as one of the seven specific Sleep-Related Movement Disorders.

Parasomnias are sometimes referred to as disorders of partial arousal. In general, parasomnias are a diverse collection of sleep disorders characterized by physiologic or behavioral phenomena that occur during or are potentiated by sleep. One conceptual framework posits many parasomnias as overlap or intrusion of one basic sleep–wake state into another. Wakefulness, NREM sleep, and REM sleep can be characterized as the three basic states that differ in their neurologic organization. During wakefulness, both the body and brain are active. In NREM sleep, both the body and brain are inactive. REM sleep, however, pairs an atonic body with an active brain (capable of creating elaborate dream fantasies). Regional cerebral blood flow, MRI, and other imaging studies confirm increased brain activation during REM sleep. It certainly appears that in some parasomnias, there are state boundary violations. For example, arousal disorders (sleepwalking and sleep terrors) involve momentary or partial wakeful behaviors suddenly occurring in NREM sleep. Similarly, the ICSD-3 nosologic entity Isolated Sleep Paralysis is the persistence of REM sleep atonia into the wakefulness transition, while REM sleep behavior disorder is the failure of the mechanism creating paralytic atonia such that individuals literally act out their dreams.

The occurrence frequency of significant parasomnias varies. Clinical significance relates more to the medical consequences or the patient’s evoked level of distress than to how often the abnormal events occur. Parasomnias’ irregularities in occurrence make them difficult to document in the sleep laboratory. Sleep studies, however, are indicated for ruling out whether the unusual behavior is secondary to a seizure disorder or a sleep-related breathing disorder.

Non-Rapid Eye Movement (NREM) Sleep Arousal Disorders

Sleepwalking. Sleepwalking in its classic form, as the name implies, is a condition in which an individual arises from bed and ambulates without fully awakening. Sometimes called somnambulism, individuals can engage in a variety of complex behaviors while unconscious. Sleepwalking usually occurs during NREM slow wave sleep and characteristically begins toward the end of the first or second slow wave sleep episodes. Sleep deprivation and interruption of slow wave sleep appear to exacerbate, or even provoke, sleepwalking in susceptible individuals. Sleepwalking episodes may range from sitting-up and attempting to walk to conducting an involved sequence of semi-purposeful actions. The sleepwalker can often successfully interact with the environment (e.g., avoiding tripping over objects in his path). However, the sleepwalker will often interact with the environment inappropriately, which sometimes results in injury (e.g., stepping out of an upstairs window or walking into the roadway). There are cases in which sleepwalkers have committed acts of violence. An individual who is sleepwalking is difficult to awaken. Once awake, the sleepwalker will usually appear confused. It is best to gently attempt to lead sleepwalkers back to bed rather than attempting to awaken them by grabbing, shaking, or shouting. In their confused state, the sleepwalker may think they are being attacked and may react violently to defend themselves. Sleepwalking in adults is rare, has a familial pattern, and may occur as a primary parasomnia or secondary to another sleep disorder (e.g., sleep apnea). By contrast, sleepwalking is very common in children, with peak prevalence between 4 and 8 years of age. After adolescence, it usually disappears spontaneously. Nightly to weekly sleepwalking episodes associated with physical injury to the patient and others are considered severe.

Sleep Terrors. A sleep terror is characterized by a sudden arousal with intense fearfulness. They may begin with a piercing scream or cry (see Fig. 23–7). Autonomic and behavioral correlates of fright typically mark the experience. An individual experiencing a sleep terror usually sits up in bed, is unresponsive to stimuli, and is confused or disoriented if awakened. Vocalizations may occur, but they usually are incoherent. Notwithstanding the intensity of these events, amnesia for the episodes usually occurs. Like sleepwalking, these episodes usually arise from slow wave sleep. Fever and CNS depressant withdrawal potentiate sleep terror episodes. Unlike nightmares in which an elaborate dream sequence unfolds, sleep terrors may be devoid of images or contain only fragments of very brief but frighteningly vivid and sometimes static images. Sometimes called *Pavor nocturnus*, incubus, or night terror, a familial pattern has been reported. Like other slow wave sleep parasomnias, sleep deprivation can provoke or exacerbate sleep terrors. Psychopathology is seldom associated with sleep terrors in children; however, a history of a traumatic experience or frank psychiatric problems is often comorbid in adults with this disorder. Severity ranges from less than once per month to an almost nightly occurrence (with injury to the patient or others).

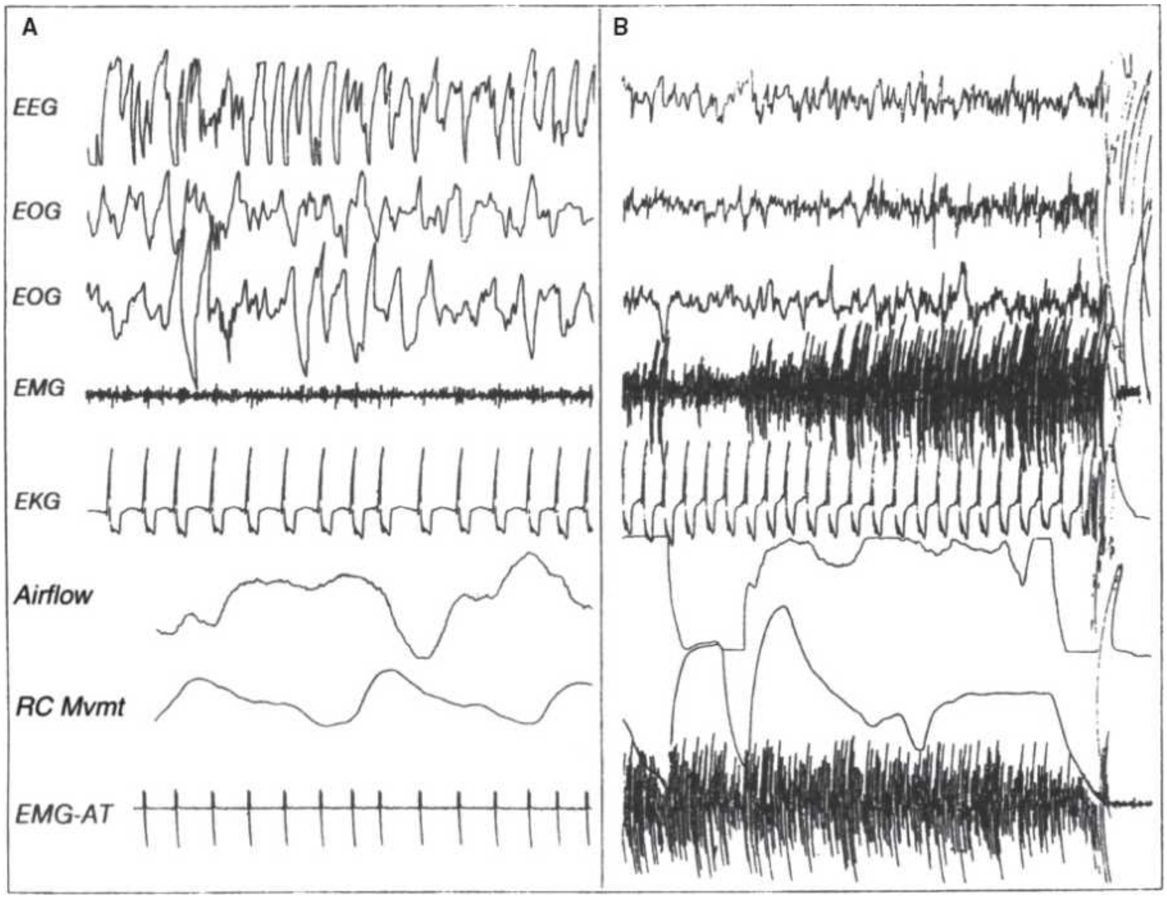

FIGURE 23–7. Polysomnographic tracing illustrating a sleep terror. **Panel A** shows approximately 14 seconds of tracing occurring immediately before the sleep terror. Prominent EEG slow-wave activity and other characteristics of stage 4 sleep are seen. **Panel B** shows the awakening, accompanied by tachycardia and movement. EEG activity is ambiguous and the patient eventually disconnected his electrodes as he thrashed about in bed (visible at far right of figure). Although the patient was screaming and greatly agitated, there was no report of dreaming. In the morning, there was little recollection of anything having occurred during the night.

Diagnostic Criteria. For Sleepwalking, DSM-5-TR requires repeated episodes of rising from bed and walking. The sleepwalker usually has a blank expression, is generally nonresponsive, and is difficult to awaken. For Sleep Terror, there must also be recurrent episodes where the individual awakens with intense fear and autonomic signs of distress. The patient is typically so panicked as to be inconsolable.

The clinician should specify if the NREM Parasomnia is: (a) sleepwalking, (b) sleepwalking type with sleep-related eating, (c) sleepwalking type with sleep-related sexual behavior (sexsomnia), or (d) sleep terror type.

Nightmare Disorder

Nightmares are frightening or terrifying dreams. Sometimes called dream anxiety attacks, they can produce sympathetic activation and ultimately awaken the dreamer. Nightmares occur in REM sleep and usually evolve from a long, complicated dream that becomes increasingly frightening. Having aroused to wakefulness, the dream content is typically remembered (in contrast to sleep terrors). Some nightmares are recurrent, and reportedly when occurring in association with posttraumatic stress disorder, they may be recollections of actual events. Common in children ages 3 to 6 (prevalence estimates range from 10% to 50%), nightmares are rare in adults (1% or less). Frequent and distressing nightmares are sometimes responsible for insomnia because the individual is afraid to sleep. In Freudian terms, the nightmare is an example of the failure of the dream process that defuses the emotional content of the dream by disguising it symbolically, thus preserving sleep. Most patients afflicted with nightmares are free from psychiatric conditions. Nonetheless, individuals at risk for nightmares include those with schizotypal, borderline, and schizoid personality disorders, as well as schizophrenia. Hartmann (1998) posits that nightmares are more common in individuals with “thin boundaries” who are open, trusting, and often have creative or artistic inclinations. Having thin boundaries makes these individuals more vulnerable; furthermore, they may be at risk for schizophrenia. Traumatic events are known to induce nightmares, sometimes immediately but at other times delayed. The nightmares can persist for many years. Several medications are known to sometimes provoke nightmares, including L-DOPA, beta-adrenergic blockers, and withdrawal from REM suppressant medications. Finally, drug or alcohol abuse is associated with nightmares.

Frequently occurring nightmares often produce a “fear of sleeping” type of insomnia. In turn, the insomnia may provoke sleep deprivation, which is known to exacerbate nightmares. In this manner, a vicious cycle is created. Treatment using behavioral techniques can be helpful. Universal sleep hygiene, stimulus control therapy, lucid dream therapy, and cognitive therapy reportedly improve sleep and reduce nightmares. BZD and BZRAs are sometimes used as treatment. Additionally, evidence for the use of prazosin, a central nervous system alpha-1 receptor antagonist, in the treatment of PTSD-related nightmares is growing. Prazosin significantly increased total sleep time and REM sleep time and significantly reduced trauma-related nightmares and distressed awakenings.

Rapid Eye Movement Sleep Behavior Disorder

REM sleep behavior disorder involves a failure of the patient to have atonia (sleep paralysis) during stage REM sleep. The result is that the patient literally enacts their dreams. Under normal circumstances, the dreamer is immobilized by REM-related hypopolarization of alpha and gamma motor neurons. Without this paralysis or with intermittent atonia, punching, kicking, leaping, and running from bed during attempted dream enactment occur. The activity has been correlated with dream imagery, and unlike during sleepwalking, the individual seems unaware of the actual environment but rather is acting on their dream sensorium. Thus, a sleepwalker may calmly go to a bedroom window, open it, and step out. By contrast, a person with REM sleep behavior disorder would more likely dive through the window thinking it is a dream-visualized lake. Patients and bed partners frequently sustain an injury, sometimes serious (e.g., lacerations, fractures). A wide variety of drugs and comorbid conditions can precipitate or worsen RBD (see the section on Parasomnia Due to Drug or Substance Use and Parasomnia Due to Medical Conditions). In animals, presumed RBD can be produced with bilateral perilocus coeruleus lesions. In humans, there is a suggestion that RBD may result from diffuse hemispheric lesions, bilateral thalamic abnormalities, or brainstem lesions. Clonazepam has been used successfully to treat RBD.

Restless Legs Syndrome

RLS is characterized by the irresistible urge to move the legs when at rest or while trying to fall asleep. Patients often report crawling feelings in their legs. Moving the legs or walking around helps alleviate the discomfort (see Fig. 23–8). Thus, as individuals lie in bed and relax, they are disturbed by these sensations. They then move their legs and again try to fall asleep. This cycle sometimes continues for hours and results in profound insomnia.

Diagnostic Criteria. For the diagnosis of RLS, DSM-5-TR requires the criteria shown in Table 23–5.

An NIH workshop established criteria for the diagnosis of RLS (see Table 23–6). Uremia, neuropathies, and iron and folic acid deficiency anemias can produce secondary RLS. RLS is also associated with fibromyalgia, rheumatoid arthritis, diabetes, thyroid diseases, and chronic obstructive pulmonary disease. Detailed history and physical exam are an important part of the RLS work up. In addition, ferritin levels should be checked.

Pharmacologically, the dopaminergic agonists pramipexole, rotigotine, and ropinirole and gabapentin encarbil (extended-release form) are FDA approved and represent the treatments of choice. These four compounds have evidence-based support. Nonetheless, other agents used to treat RLS include dopamine precursors (e.g., levedopa), BZDs, opiates, and anti-epileptic drugs (e.g., gabapentin). Sometimes the alpha-2 agonist clonidine is also used. Nonpharmacologic treatments include avoiding alcohol or nicotine use close to bedtime, massaging the affected parts of the legs, taking hot baths, applying hot or cold to the affected areas, and moderate exercise.

Other ICSD-3 Parasomnias and Sleep-Related Movement Disorders

Confusional Arousals. ICSD-3 defines Confusional Arousals as a milder form of NREM Sleep Parasomnias. It is common in young children. The child will typically partially awaken from sleep and sit up. The episodes are marked by confusion, but usually the child lies back down and resumes sleep. It is thought that confusional arousals, sleepwalking, and sleep terrors lie on a continuum.

Sleep-Related Eating Disorder. This diagnostic entity is subsumed under the DSM-5-TR classification NREM Sleep Parasomnias and specified as “sleepwalking type with sleep-related eating.” In adults, nocturnal eating can be conditioned to awakening. Eating may become obsessional, and several small meals may be eaten during the course of a night. The individual may be unaware of the activity, and weight gain can become a problem. Essentially, this syndrome’s ICSD-3 diagnostic criteria are as follows:

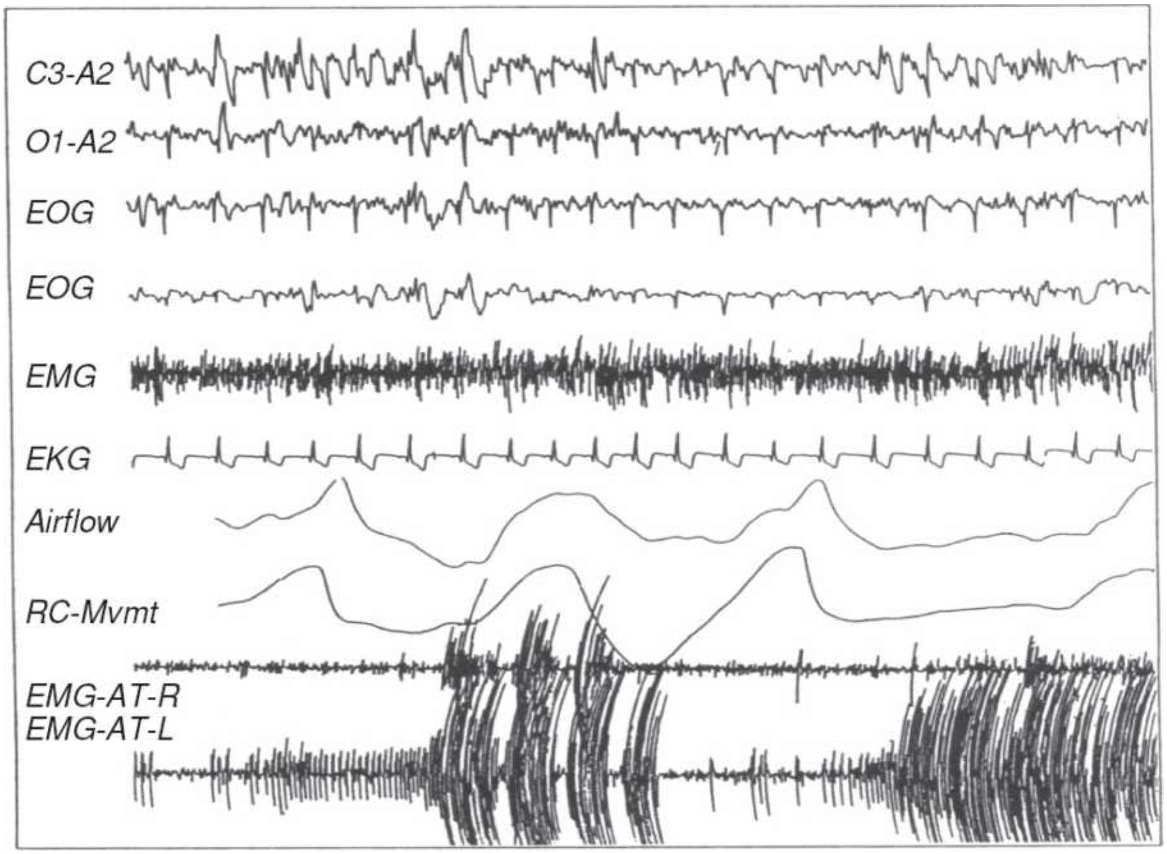

FIGURE 23–8. Restless legs syndrome. This patient presented with complaints of uncomfortable, crawling sensations in the legs when trying to fall asleep. Patients commonly report an urge to move the leg in order to dispel the sensation. This figure shows a bilateral pattern of leg EMG activity; however, the discharge is more pronounced in the left anterior tibialis (EMG-AT-L) than the right (EMG-AT-R). This pattern continued for more than an hour as the patient attempted to fall asleep; note that the sharp activity in central and occipital EEG (C3-A2 and O1-A2, respectively) and EOG is EKG artifact and not an EEG abnormality.

Table 23–5.
DSM-5 Diagnostic Criteria for Restless Legs Syndrome

A. An urge to move the legs, usually accompanied by or in response to uncomfortable and unpleasant sensations in the legs, characterized by all of the following:
   1. The urge to move the legs begins or worsens during periods of rest or inactivity.
   2. The urge to move the legs is partially or totally relieved by movement.
   3. The urge to move the legs is worse in the evening or at night than during the day, or occurs only in the evening or at night.

B. The symptoms in Criterion A occur at least three times per week and have persisted for at least 3 months.

C. The symptoms in Criterion A are accompanied by significant distress or impairment in social, occupational, educational, academic, behavioral, or other important areas of functioning.

D. The symptoms in Criterion A are not attributable to another mental disorder or medical condition (e.g., arthritis, leg edema, peripheral ischemia, leg cramps) and are not better explained by a behavioral condition (e.g., positional discomfort, habitual foot tapping).

E. The symptoms are not attributable to the physiologic effects of a drug of abuse or medication (e.g., akathisia).

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 464, 465 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Table 23–6.
International Study Group Diagnostic Criteria for Restless Legs Syndrome

Essential features	An urge to move legs, usually accomplished or caused by uncontrollable or unpleasant sensations in the legs
	The urge to mover, or unpleasant sensations that begin or worsen during periods of rest or inactivity such as lying or sitting
	The urge to move or unpleasant sensations are partially or totally relieved by movement, such as walking or stretching, at least as long as the activity continues
	The urge to move or unpleasant sensations are worse in evening or night than during the day or only occur in the evening or night
Common features	Positive family history: 3–5 times higher prevalence of RLS among first-degree relatives
	Response to dopaminergic therapy
	PLM during sleep or wakefulness
	Clinical course: middle-age onset and usually with progressive course
	Sleep disturbance resulting in sleep-onset insomnia or daytime sleepiness
	No abnormality in primary form of the syndrome

Reprinted from Allen RP, Picchietti DL, Garcia-Borreguero D, et al; International Restless Legs Syndrome Study Group. Restless legs syndrome/Willis-Ekbom disease diagnostic criteria: updated International Restless Legs Syndrome Study Group (IRLSSG) consensus criteria–history, rationale, description, and significance. *Sleep Med.* 2014;15(8):860–873. Copyright © 2014, with permission from Elsevier. doi:10.1016/j.sleep.2014.03.025.

a. The patient has recurrent sleep-related eating episodes.
b. The patient consumes peculiar foods, food combinations, or toxic substances; creates a hazard or is injured during food preparation (e.g., starting a fire on the stove); or the patient suffers adverse health consequences from sleep-related eating (e.g., substantial weight gain).
c. The eating is associated with partial or complete amnesia and general lowering or loss of conscious awareness.

d. The eating is not better explained by another disorder, medications, or other substance use.

Recurrent Isolated Sleep Paralysis. Sleep paralysis is, as the name implies, an inability to make voluntary movements during sleep. ICSD-3 lists it as a parasomnia when it occurs at sleep onset or upon awakening, a time when the individual is partially conscious and aware of the surroundings. This inability to move can be extremely distressing, especially when coupled with the feeling that there is an intruder in the house or when hypnagogic hallucinations are occurring. Sleep paralysis is one of the tetrad of symptoms associated with narcolepsy; however, it is known to occur (with or without hypnagogia) in individuals with neither cataplexy nor excessive daytime sleepiness. Although sometimes frightening, sleep paralysis is a feature of normal REM sleep briefly intruding into wakefulness. The paralysis may last from one to several minutes. Interestingly, sleep paralysis with hypnagogia may account for a variety of experiences where the sleeper is confronted or attacked by some sort of creature. The common description is that a “presence” was felt to be near, the individual was paralyzed, the creature talks, attacks, or sits on the sleeper’s chest, and then vanishes. Whether it is called incubus, “Old Hag,” vampire, ghost oppression (kanashibari), witch riding, or alien encounter, elements common to sleep paralysis are seen. Irregular sleep, sleep deprivation, psychological stress, and shift work are thought to increase the likelihood of sleep paralysis occurring. Occasional sleep paralysis occurs in 7% to 8% of young adults. Estimates of at least one experience of sleep paralysis during the lifetime range from 25% to 50%. Improved sleep hygiene and assuring sufficient sleep are first-line therapies. Sometimes, if the individual voluntarily makes very rapid eye movements or is touched by another person, the episode will terminate.

Exploding Head Syndrome. Individuals with this parasomnia “hear” a loud imagined noise or a sense of a violent explosion in the head just as they are about to fall asleep or during a nocturnal awakening. The experience can occur just once or recurrently. There is no pain associated with the noise, but the individual may be concerned that they are having a stroke or that something is very wrong. Even a single episode can trigger severe insomnia. There are no known neurologic consequences to this syndrome.

Sleep-Related Hallucinations. Sleep-Related Hallucinations are typically visual images occurring at sleep onset (hypnagogic) or on awakening (hypnopompic) from sleep. Sometimes difficult to differentiate from dreams, they are common in patients with narcolepsy. Complex hallucinations are rare and usually happen with abrupt awakening and without remembering dreaming. Images tend to be vivid and immobile and persist for several minutes (usually disappearing when a light is turned on). The images can be frightening.

Sleep Enuresis. Sleep enuresis is a disorder in which the individual urinates during sleep while in bed. In adults, sleep enuresis often accompanies severe obstructive sleep apnea, and in such cases, it would fall into the category of parasomnia related to a medical disorder. In children, bedwetting, as it is commonly called, has primary and secondary forms. In children, primary sleep enuresis is the continuance of bedwetting since infancy. Secondary enuresis refers to relapse after toilet training was complete and there was a period where the child remained dry. Usually, after toilet training, bedwetting spontaneously resolves before age 6 years. Prevalence progressively declines from 30% at age 4 to 10% at age 6 to 5% at age 10, and 3% at age 12. Parental primary enuresis increases the likelihood that the children will also have enuresis. A single recessive gene is suspected. Secondary enuresis in children may occur with the birth of a sibling and represent a “cry for attention.” Secondary enuresis can also be associated with nocturnal seizures, sleep deprivation, and urologic anomalies. In adults, sleep enuresis is sometimes seen in patients with sleep-disordered breathing. In most cases, embarrassment, shame, and guilt are the most serious consequence. Nonetheless, if sleep enuresis is not addressed, it may leave psychosocial scars.

Treating sleep enuresis involves medication, behavioral interventions, or both. Among medications, desmopressin acetate is the preferred choice. An evidence-based medicine review of 47 randomized trials found efficacy with desmopressin. Other medications used to treat sleep enuresis include imipramine and oxybutynin chloride. Behavioral treatments, including bladder training, using conditioning devices (bell and pad), diet modification, and fluid restriction, have reportedly had good success when properly administered. Other treatments include psychotherapy, motivational strategies, and hypnotherapy.

Parasomnia Due to a Medical Disorder. Seizure disorder should always be on the top of a differential diagnoses list for most parasomnias. In fact, the AASM practice guidelines concerning the indications for polysomnography include using sleep testing to rule out seizures when diagnosing sleep terror, sleepwalking, RBD, nightmares, and other parasomnias (see Fig. 23–9). Sleep-related breathing disorders are also known to trigger sleepwalking, enuresis, sleep terror, confusional arousal, and nightmares. RBD is associated with a variety of neurologic conditions, including Parkinson disease, dementia, progressive supranuclear palsy, Shy–Drager syndrome, narcolepsy, and others.

Parasomnia Due to Drug or Substance Use. Many drugs and substances can trigger parasomnias, particularly those compounds that lighten sleep; however, alcohol is notorious for producing sleepwalking (especially in individuals who have taken sleeping pills). RBD can be provoked or worsened by biperiden, tricyclic antidepressants, monoamine oxidase inhibitors, caffeine, venlafaxine, selegiline, and serotonin agonists. RBD may also occur during withdrawal from alcohol, meprobamate, pentazocine, and nitrazepam. Medications known to provoke nightmares include L-DOPA and beta-blockers. Nightmares can also be caused by drug-induced REM sleep rebound (e.g., withdrawal from REM suppressing drugs like methamphetamine), beta-blockers, and alcohol abuse or withdrawal.

Periodic Limb Movement Disorder. Periodic Limb Movement Disorder (PLMD), previously called nocturnal myoclonus, involves brief, stereotypic, repetitive, nonepileptiform movements of the limbs, usually the legs. It occurs primarily in NREM sleep and involves an extension of the big toe. A partial flexion of the ankle, knee, and hip may also occur. These movements range from 0.5 to 5 seconds in duration and occur every 20 to 40 seconds. The leg movements are frequently associated with brief arousals from sleep and, as a result, can (but do not always) disturb sleep architecture (see Fig. 23–10). The prevalence of PLMD increases with aging and can occur in association with folate deficiency, renal disease, anemia, and the use of antidepressants. Pharmacotherapy for PLMD associated with RLS is the same as RLS. Clinical trials of pharmacotherapy for other forms of PLMD are lacking. However, BZDs, especially clonazepam, reportedly improve sleep in patients with PLMD.

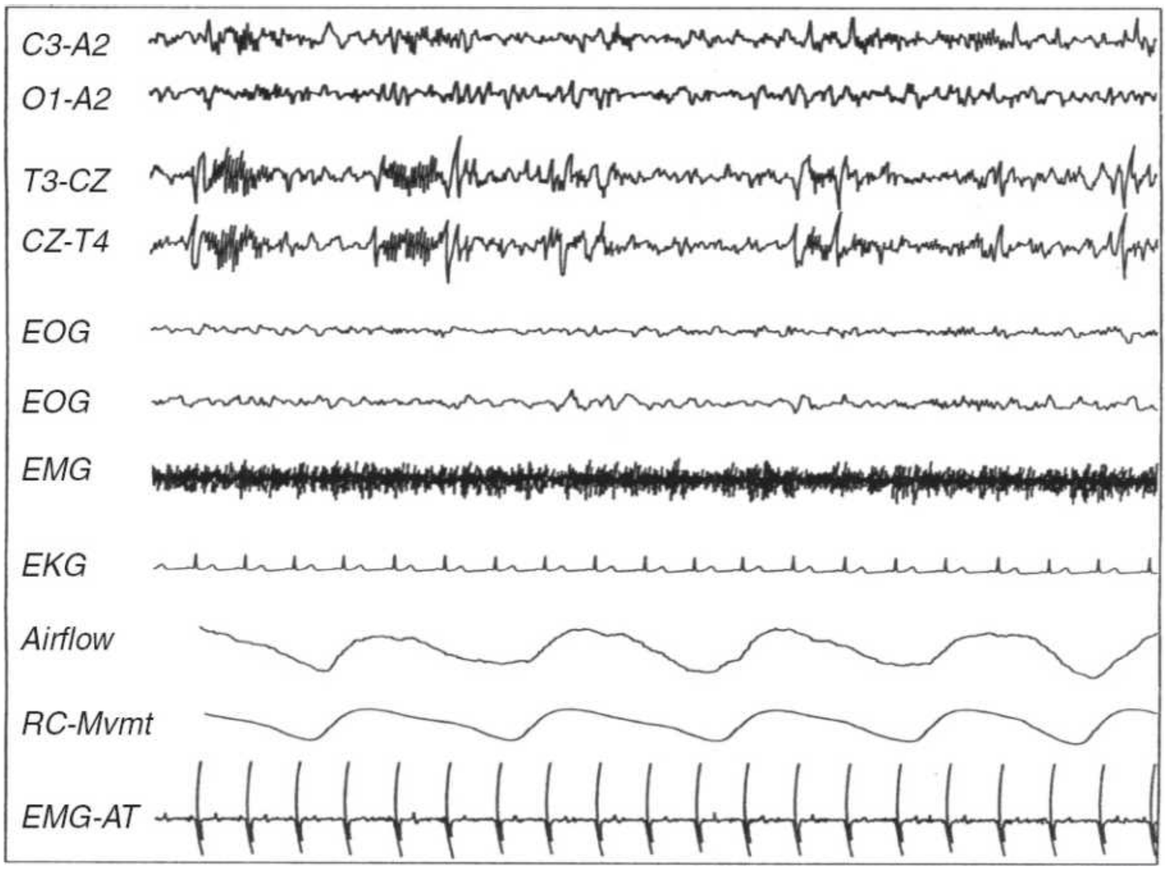

FIGURE 23–9. Polysomnographic tracing from a patient with complex partial seizures (CPS). EEG sleep spindles of excessive magnitude and duration mixed with sharp activity are visible, especially in temporal lobe derivations. The slow paper speed used in sleep studies makes it difficult to classify this activity precisely; however, the abnormality is apparent. Such a finding should be confirmed with standard clinical EEG techniques both during wakefulness and sleep.

Sleep-Related Leg Cramps. Nocturnal leg cramps are much like leg cramps that occur during wakefulness. They usually affect the calf and are painful muscle contractions. The pain awakens the sleeper and thereby disrupts sleep. Metabolic disorders, mineral deficiencies, electrolytic imbalances, diabetes, and pregnancy are known precipitators. The reason that some individuals have repeated leg cramps during sleep and not during the day is not known.

Sleep-Related Bruxism. Sleep-Related Bruxism is diagnosed when an individual grinds or clenches their teeth during sleep (see Fig. 23–11). ICSD formerly classified sleep bruxism as a parasomnia but now lists it as a sleep-related movement disorder. Sleep bruxism can produce abnormal wear on the teeth, damage teeth, provoke tooth and jaw pain, and/or make loud, unpleasant sounds that disturb the bed-partner. Sometimes atypical facial pain and headache also result. More than 85% of the population may brux at one time or another; however, it is clinically significant in only about 5%. Teeth grinding occurs in any sleep stage but appears most common at the transition to sleep, in stage 2 sleep, and during REM sleep. Some evidence indicates that teeth-grinding during REM sleep is more commonly associated with dental wear or damage. Sleep bruxism does not appear to be exacerbated by dental malocclusion but rather worsens during periods of stress. Researchers studying sleep bruxism find that many patients have less frequent teeth-grinding when sleeping in the laboratory; therefore, repeated studies may be needed to document the disorder. By contrast, bruxism frequently appears on polysomnographic recordings made for other purposes and may occur secondary to (a) sleep-related breathing disorders, (b) the use of psychostimulants (e.g., amphetamine, cocaine), (c) alcohol ingestion, and (d) treatment with some serotonin reuptake inhibitors. Differential diagnosis should rule out nocturnal seizures. Sleep bruxism can occur infrequently (monthly), regularly (weekly), or frequently (nightly). Severity is judged on the basis of sleep disruption, consequent pain, and dental damage. The usual treatment involves having the patient wear an oral appliance to protect the teeth during sleep. There are two basic types of appliances used. The soft one (mouth guard) is typically used in the short term, whereas the hard acrylic one (bite splint) is used longer term and requires regular follow-up. Relaxation, biofeedback, hypnosis, physical therapy, and stress management are also used to treat sleep bruxism. A variety of drug therapies have been tried (BZDs, muscle relaxants, dopaminergic agonists, and propranolol); however, reliable outcome data are not available.

Sleep-Related Rhythmic Movement Disorder. This sleep disorder is marked by repetitive, rhythmic movements usually involving the head and neck. Usually occurring during the transition from wakefulness to sleep, this movement disorder may also continue during light sleep. ICSD formerly classified this nosologic entity as a parasomnia, but it is now listed as a sleep-related movement disorder. The Sleep-Related Rhythmic Movement Disorder has been called by many names, including *jactatio capitis nocturna*, head banging, head rolling, body rocking, and *rhythmie* *du sommeil*. A majority of infants will body rock. Some clinicians think body rocking develops from the soothing effect of vestibular stimulation. If the rhythmic movement persists into childhood and involves head banging, the risk of injury increases. Male to female ratio is 4:1. Severity ranges from less than one episode weekly to nightly episodes producing injury.

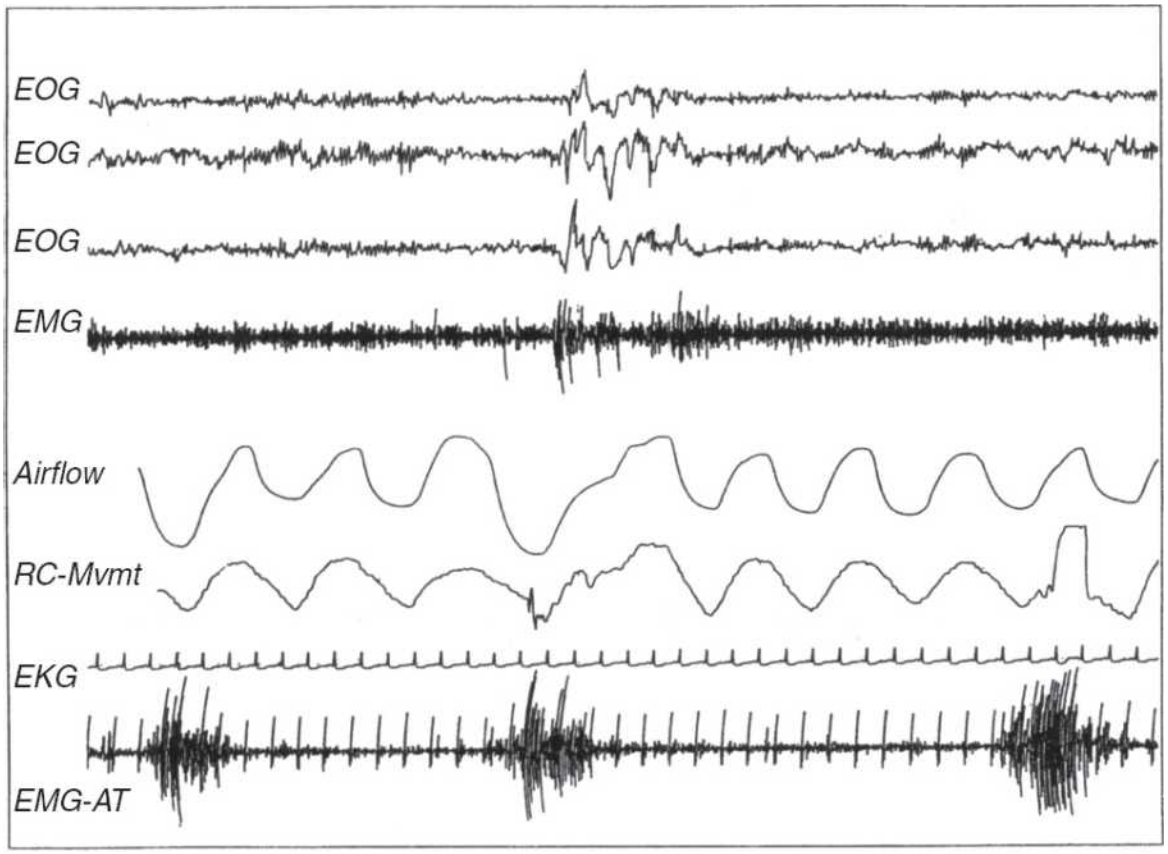

FIGURE 23–10. Polysomnographic tracing illustrating periodic leg movements in sleep. This figure shows periodic leg movements during sleep. Approximately 33 seconds of recording shown here clearly reveal a pattern of electromyographic activity in anterior tibialis (EMG-AT). The activity pattern repeats at regular intervals and sometimes produces transient arousals (note EEG-EOG-EMG) after movement at the center of the figure. This patient, whose chief complaint was nonrefreshing sleep with sleep maintenance insomnia, had more than 250 such leg movements during each night he slept in the laboratory.

Benign Sleep Myoclonus of Infancy. Previously called benign neonatal sleep myoclonus, this disorder is characterized by asynchronous jerking of limbs and trunk during quiet sleep in neonates. This benign, apparently rare parasomnia usually begins within the first week of life and may last a few days or several months. No treatment is recommended.

Propriospinal Myoclonus at Sleep Onset. This is a spinal cord mediated movement disorder, sometimes associated with spinal cord lesions. Movements appear during times of wakeful relaxation, and they may interfere with sleep onset. They start in the abdominal and truncal muscles and then progress to the neck and proximal muscles of the limbs. Treatment with clonazepam or anticonvulsants may be effective.

Sleep-Related Movement Disorder Due to a Medical Disorder. Neurologic diseases associated with daytime movement disorders can likewise be associated with sleep-related movement disorders. Stress, anxiety, and sleep deprivation may contribute to bruxism.

Sleep-Related Movement Disorder Due to Medication or Substance. A variety of medications and substances can produce or exacerbate sleep-related movement disorders. Stimulants can produce rhythmic movement disorders and bruxism. Antidepressants (including most tricyclics and selective serotonin reuptake inhibitors), antiemetics, lithium, calcium channel blockers, antihistamines, and neuroleptics can provoke Restless Legs symptoms and Periodic Limb Movement Disorder.

Sleep-Related Movement Disorder, Unspecified. ICSD-3 provides this nosologic classification for coding other unusual behaviors and otherwise unexplained movement disorders occurring during sleep.

Isolated Symptoms and Normal Variants

Sleep Talking. As the name implies, sleep talking in classic form involves unconscious speech during sleep. It is seldom recognized in an individual unless it annoys the bed partner. It can be induced by fever, stress, or conversing with the sleeper. Somniloquy may accompany sleep terror, sleepwalking, confusional arousals, OSA, and REM sleep behavior disorder.

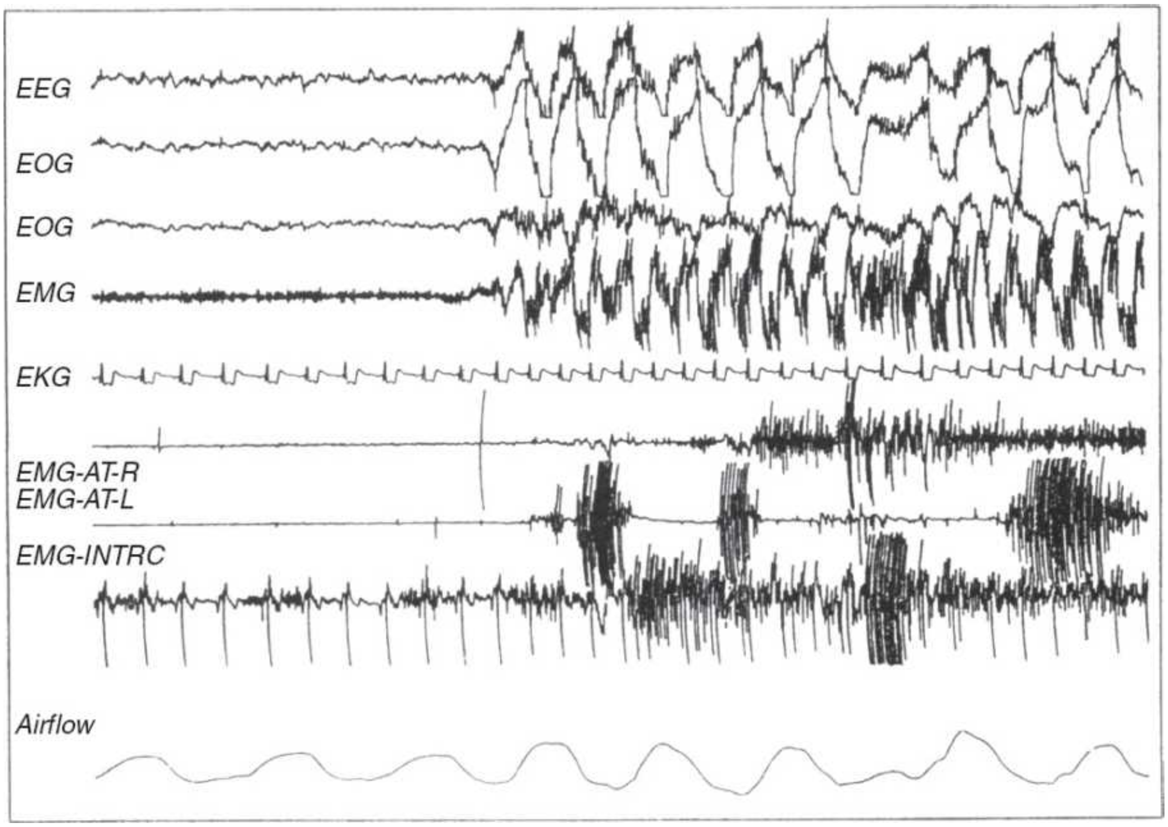

FIGURE 23–11. Polysomnographic tracing showing sleep-related bruxism. This figure showing approximately 25 seconds of tracing was obtained from a patient during an episode of bruxism. Bruxism can occur during any stage of sleep or wakefulness. The interference pattern of EEG-EOG-EMG channels is typical and reflects the rhythmic jaw movement and grinding of the teeth. This patient had many such episodes, some that caused awakening. Readily observable tooth damage and jaw pain were noted.

Excessive Fragmentary Myoclonus. These small movements or muscle fasciculations of the fingers, toes, or corners of the mouth are involuntary and can occur during wakefulness or sleep. Although little or no visible movement is present, the patient is typically aware of the twitching. In patients with apneas, twitching may worsen during periods of hypoxemia.

Hypnagogic Foot Tremor and Alternating Leg Muscle Activation. Hypnagogic foot tremor occurs at sleep onset or during stages 1 and 2 of sleep. It consists of a rhythmic movement of the toes or feet for seconds to minutes. Alternating Leg Muscle Activation consists of brief activation of the anterior tibialis in one leg and then the other.

Sleep Starts (Hypnic Jerks). Sleep starts are sudden, brief muscle contractions that occur during the transition between wakefulness and sleep in 60% to 70% of adults. The contractions commonly involve the legs; however, there is sometimes movement in the arms and head. This "hypnic jerk," as it is sometimes called, is usually benign. The sleep start, however, can interfere with the ability to fall asleep and may be accompanied by sensations of falling, a hallucinated flash of light, or a loud crackling sound. In severe cases, the sleep start produces profound sleep-onset insomnia.

SUBSTANCE/MEDICATION-INDUCED SLEEP DISORDER

Medications and substances associated with or thought to provoke sleep disorders are discussed in sections: Insomnia Due to Drug or Substance; Characteristics and Hypersomnolence Disorder Subtypes; Central Sleep Apnea Due to Medication or Substance; Parasomnia Due to Drug or Substance; and Sleep Related Movement Disorders Due to Medication or Substance.

TOOLS IN SLEEP MEDICINE

Clinical Interview

A careful and thorough clinical interview remains one of the most informative parts of a patient workup for sleep disorders. The habitual bedtime and arising times for both weekdays and weekends, the frequency, duration, and how refreshing naps are, and overall levels of sleepiness and satisfaction with sleep are good places to begin. Specific sleep problems relating to difficulty initiating and maintaining sleep are important, including whether there is rumination at bedtime, fear of not being able to sleep, or excessive worry when attempting to sleep. Leg movements, leg sensations, leg cramps, teeth grinding, dream enactments (with or without injury), and other movements should be queried. Morning headaches, morning dry mouth, nocturnal reflux, hyperhidrosis, nocturia, enuresis, nocturnal tongue biting, nightmares, sleep terrors, and other sleep-related problems should be reviewed. The presence of family pets and whether they sleep in the bedroom (or bed) can be important in some cases.

Polysomnography

Polysomnography is the continuous, attended, comprehensive recording of physiologic activity during sleep. A polysomnogram is typically recorded at night and is 6 to 8 hours in duration. Brain wave activity, eye movements, submentalis electromyography activity, nasal-oral airflow, nasal pressure, respiratory effort, oxyhemoglobin saturation, heart rhythm, and leg movements during sleep are recorded. Body position is usually noted, and snoring sounds may be recorded. Brain wave activity, eye movements, and submentalis electromyogram are important for identifying sleep stages and central nervous system arousals. Muscle tension and movements subside with deeper sleep and can also be useful in diagnosing periodic limb movement disorder and RLS. Nasal airflow, nasal pressure, respiratory effort, and oxyhemoglobin saturation are instrumental in diagnosing sleep apnea and other sleep-related breathing disorders.

Indications for polysomnography include:

1. diagnosis of sleep-related breathing disorders
2. PAP titration and assessment of treatment efficacy
3. evaluation of sleep-related behaviors that are violent or may potentially harm the patient or bed partner
4. differentiating narcolepsy from other hypersomnolence disorders
5. differentiating parasomnias from nocturnal seizures.

Polysomnography can also be used to diagnose atypical parasomnias, sleep-related problems secondary to neuromuscular disorders, periodic limb movement disorder, and arousals secondary to a seizure disorder. Also, patients with excessive daytime sleepiness, or those who wake up gasping or choking, should be referred for polysomnography. A sleep study is not needed to diagnose restless leg syndrome.

Polysomnography is also recommended to assess sleep quality and quantity on the night prior to a Multiple Sleep Latency Test (MSLT) being conducted to diagnose narcolepsy.

Home Sleep Apnea Testing (HSAT)

HSAT involves recording a limited number of cardiopulmonary parameters to assess patients for sleep-related breathing disorders. HSAT may include limited channels of EEG recording for sleep staging. HSAT is much less expensive than polysomnography. Usually, airflow, respiratory effort, heart rhythm, snoring sounds, and oximetry are recorded. HSAT devices will detect sleep apnea in patients with moderate-to-severe pathophysiology. Negative studies are problematic because HSAT is less sensitive than full laboratory polysomnography. Patients with negative tests should be scheduled for laboratory polysomnography, notwithstanding obvious symptoms or comorbidity. Additionally, HSAT does not check for the full spectrum of sleep disorders; therefore, after a breathing disorder is diagnosed and treated, residual symptoms need careful follow-up.

Multiple Sleep Latency Test

The MSLT is indicated for diagnosing narcolepsy. Every 2 hours, beginning 2 hours after morning awakening, a 20-minute nap opportunity is provided. The patient is instructed to let themself fall asleep; that is, to not resist falling asleep. Electroencephalographic, electrooculographic, and submentalis electromyography activity are recorded to determine sleep stage. The latency to sleep is used to assess hypersomnolence, and the appearance of REM sleep on two or more nap opportunities confirms narcolepsy (or on one nap if the previous night's polysomnography revealed REM sleep within 15 minutes of sleep onset). If the patient falls asleep on a given nap opportunity, the nap is terminated 15 minutes after initial sleep onset. If the patient does not fall asleep, the session is terminated after 20 minutes of recording. A total of five nap opportunities are provided at 2-hour intervals across the day.

Maintenance of Wakefulness Test

Similar procedurally to MSLT, the maintenance of wakefulness test (MWT) provides 40-minute test sessions at 2-hour intervals across the day. However, the patient is instructed to try and remain awake and resist sleep rather than not resist sleep. This technique is used to assess treatment outcomes and is sometimes required as part of “fit for duty” testing. Patients sit in a comfortable chair or on the bed with a bolster pillow in a darkened room while recordings are made. The first epoch of stages 2, 3, 4, or REM or three consecutive epochs of stage 1 mark unequivocal sleep onset. Falling asleep on MWT indicates some level of sleepiness. 59% of volunteers (presumably normal) remain awake for the entire 40 minutes across all four trials (using unequivocal sleep criteria). Any sleep latency <8 minutes is abnormal. Sleep latency ranging from 8 to 40 minutes is of unknown significance. The mean sleep latency (first sleep epoch) is 30.4 ± 11.2 minutes, and the upper 95% confidence interval is 40 minutes.

Actigraphy and Sleep Diary

An actigraph is a device that measures and records movement. It is usually worn on the wrist (like a watch) and can be used to measure the sleep–wake cycle. Depending on the model and the settings, it can make continuous recordings for days or weeks. It can be especially useful for assessing insomnia, circadian rhythm disorders, movement disorders, or an assortment of rare events. Research grade actigraph recorders are validated test instruments. Recently, the public's increased awareness and focus on fitness has spawned a market for consumer electronic devices that are essentially actigraphic recording devices. Many of these products purport to measure sleep and wakefulness; however, the precision and accuracy levels are unknown. A sleep diary can be used alone or with actigraphy to better identify the sleep–wake cycles of individuals with insomnia or circadian rhythm sleep–wake disorders. A sleep diary can be completed daily for a week or more, showing the bed and rise times, sleep-onset latency, wake after sleep onset, and sleep duration. It can also show the intake of CNS acting agents, including medications, caffeine, and alcohol intake in relation to the sleep–wake cycle.

Subjective Tools

STOPBANG. ***S***noring, ***T***iredness, ***O***bserved apnea, high blood ***P***ressure, ***B***MI, ***A***ge, ***N***eck circumference, and male ***G***ender is an eight-item yes/no sleep apnea risk screening tool with a minimum of zero and a maximum of 8, with the higher number indicating a higher risk of sleep apnea. The tool was initially developed to screen patients who undergo general anesthesia. However, it has gained widespread use as a sensitive screening tool for sleep apnea since then.

Epworth Sleepiness Scale (ESS). ESS is an eight-item subjective tool that asks the respondent to rate the probability of falling asleep in eight different circumstances, even if the responder has not been in that situation. Each condition will be scored from zero to three, with the total score from zero to 24. A higher score indicates higher sleepiness. Scores of 12 and above are considered abnormal. ESS is a measure of subjective sleepiness in contrast to MSLT, which measures physiologic sleepiness.

Insomnia Severity Index. The ISI is a seven-item subjective tool that measures a patient's perception of their insomnia. It not only measures the insomnia-related symptoms but also the effects of insomnia and how much the individual is concerned or distressed because of insomnia. Each item can be scored from 0 to 4 with maximum of 28. An ISI of 10 or higher identifies insomnia with high sensitivity and specificity. ISS is mostly used in research, but it can be easily used in a sleep practice.

SLEEP DISORDERS CASES

CASE STUDY 1 (HYPERSOMNIA)

DESCRIPTION

Mr. J is a 28-year-old single male with an approximately 10-year history of fatigue and sleepiness in the daytime. He began to recognize daytime sleepiness as a problem in his first year of college when he would fall asleep in class or the dormitory. He admits that his sleep–wake schedule was disrupted during college due to taking long naps and then staying up until 1:00 or 2:00 A.M. to complete his studies. His grades and social life suffered, and he describes himself as depressed, isolated, and hopeless about his planned future as a certified public accountant (CPA).

As a child, Mr. J. says he slept "normally." In high school, he felt best with 10 hours of sleep per night and functioned well in the daytime. Mr. J denies abuse of alcohol or drugs. He does not use tobacco and drinks about 8 to 10 cups of coffee per day. Family history is negative for known sleep or psychiatric disorders. Physical examination findings are noncontributory except for a body mass index of 0.29. Routine labs are normal, including TSH.

Mr. J's excessive sleepiness has continued to the present day, notwithstanding some improved sleep hygiene. Improvements include a more consistent bedtime, trying not to nap, and a torturous month-long trial without caffeine. He remains dysphoric and discouraged about his future, blaming his chronic sleepiness as the continuing impediment to his life plans. "I'm just tired of being tired," he says.

Currently, his bedtime is between 10:00 and 10:30 P.M.; his wake-up alarm is set for 6:30 A.M. He oversleeps at least once a week on workdays and sleeps from 10:30 P.M. until 10:00 A.M. on weekends in an attempt to "catch up." He has difficulty awakening and feels unrefreshed or mildly refreshed. By drinking six to eight cups of coffee in the morning, he can usually avoid dozing during the morning. Luckily, he is an independent CPA and can schedule client appointments during this relatively alert time. After lunch, he routinely falls asleep at the computer while working. He sleeps for 20 to 60 minutes and is usually awakened by his secretary. He then drinks another two cups of coffee and continues with his work. Unexpected napping can also occur later in the afternoon or evening, and he has "nodded off" while driving. He sleeps alone; however, he has been told that he snores loudly. He does not awaken gasping or choking. He denies hypnagogic hallucinations and sleep paralysis but thinks he may feel weak after the rare occasions when he participates in a heated argument.

DISCUSSION OF CASE 1

Mr. J has one of the hypersomnias. His history is most consistent with obstructive sleep apnea syndrome, idiopathic hypersomnia, sleep deprivation in a long sleeper, dyssomnia associated with mood disorder, and narcolepsy. The ancillary symptoms of narcolepsy are absent, with the possible exception of cataplexy. When cataplexy occurs clearly, the diagnosis of narcolepsy is strongly indicated. However, Mr. J's possible infrequent weakness during heated arguments is equivocal for cataplexy. His persistent desire for a 10-hour sleep period would be unusual for a patient with narcolepsy.

A long sleeper or an individual with idiopathic hypersomnia requires prolonged sleep periods and may awaken groggy, as does Mr. J. The main differentiating feature is that whenever consistently given a chance to have a full nightly sleep period (usually 10 to 12 hours), the long sleeper does not experience excessive daytime sleepiness. Furthermore, there may be associated autonomic nervous system dysfunction or polysomnographic evidence of elevated slow wave sleep percentage in patients with idiopathic hypersomnia.

Sleepiness associated with a low-grade chronic depression may be difficult to differentiate from other causes of hypersomnolence. Polysomnography, psychiatric interview, and psychometric testing can be helpful. Mr. J relates his dysphoria to sleepiness and not vice versa; nonetheless, a dyssomnia associated with mood disorder should be considered.

Obstructive sleep apnea syndrome is a strong possibility. Mr. J is overweight (BMI = 0.29) and snores loudly. Many patients are unaware of gasping or choking for breath. Often, family members witness cessation of breathing during sleep and urge patients to seek treatment. However, Mr. J lives and sleeps alone.

Polysomnography is recommended in patients suspected of obstructive sleep apnea syndrome, narcolepsy, or idiopathic hypersomnia. These disorders usually require lifelong treatment and have significant morbidity and mortality if untreated.

CASE STUDY 2 (PARASOMNIA)

DESCRIPTION

Ms. R is a 20-year-old woman who was referred with symptoms of talking, mumbling, and crying out during sleep. At least twice per week, she screams in her sleep. She is bothered with excessive sleepiness and falling asleep inappropriately, such as during a conversation. When inactive, she is tired and sleepy, even after a full 8-hour night of sleep. However, she has energy when motivated and leads a vigorous life. Once, she awakened outside of her apartment, and her roommate had to let her back in because she had locked herself out. She does not recall the sleepwalking episode or other nocturnal wanderings but sometimes remembers yelling. From the history, crying seems to occur in light sleep, but she rarely recalls any sleep-related thoughts or dreams. However, there is a history of occasional nightmares and bruxism. The patient uses an oral appliance to protect her teeth. Leg-kicking and mild snoring without gasping or choking are noted. The patient also complains of leg-kicking during sleep. Her sleep–wake schedule is irregular, and she averages between five and seven hours of sleep per night. She occasionally awakens with a headache in the morning.

Previous health history includes a hospitalization for febrile convulsions during infancy, ophthalmologic surgery for strabismus during childhood, and tonsillectomy as a teenager. Health is otherwise excellent. The patient does not smoke tobacco or drink alcohol.

DISCUSSION OF CASE 2

By history, Ms. R has one or more of the parasomnias. Sleep talking alone does not require a sleep study, but this patient has nocturnal wanderings. Polysomnography with clinical EEG is indicated to rule out unrecognized nocturnal seizure disorder or other organic factors inducing sleepwalking. Sleepwalking is common and not necessarily abnormal in young children; however, it is rare in adults and merits careful evaluation. Ms. R’s excessive daytime sleepiness is likely due to insufficient sleep (5 to 7 hours per night) and possibly parasomnia-related disruption. Interestingly, many parasomnias are exacerbated by sleep deprivation, as is nocturnal seizure disorder.

Sleep studies were performed using comprehensive, attended, laboratory polysomnography. Prior to the overnight study, a clinical EEG was performed. The clinical EEG study did not reveal any significant abnormal EEG activity during baseline, photic stimulation, or hyperventilation. An extended EEG montage was used during the sleep study. Overall sleep quality was within the normal range. Sleep efficiency was 96%, and latency to sleep was 1 minute. REM sleep percentage was elevated (31%), and latency to REM sleep was less than normal (57 minutes). Slow wave sleep was normal in percentage, but EEG delta activity was very high amplitude. The overall macroarchitectural sleep pattern suggested rebound from sleep deprivation.

By contrast, sleep microarchitecture contained many abnormal features. We observed high-amplitude paroxysmal EEG bursts. Excessively prolonged sleep spindles were noted, and rhythmic K complexes were observed (see Fig. 23–12). There was one arousal out of slow wave sleep with rhythmic EEG discharges alternating with sharp waves. Sharps and spikes occurred several times; however, the focus was difficult to localize (possibly right temporal lobe). There were frequent body movements and full body jerks, most of which occurred during NREM sleep. There were episodes of moaning during slow wave sleep and laughing during stage 2 sleep, followed by high-amplitude theta bursts and REM sleep. We observed frequent movements and arousals from REM sleep but no REM-related spikes or sharp waves. Seizure-like EEG activity was noted during the night and occurred predominantly during slow wave sleep. However, the patient did not attempt to sleepwalk. Sharp wave and spike activity increased during the final 45 minutes of the sleep study.

The patient did not have sleep-related breathing impairment, and SaO_2 nadir was 90%. She had no periodic limb movements during sleep, and polygraphic features associated with restless legs syndrome were absent.

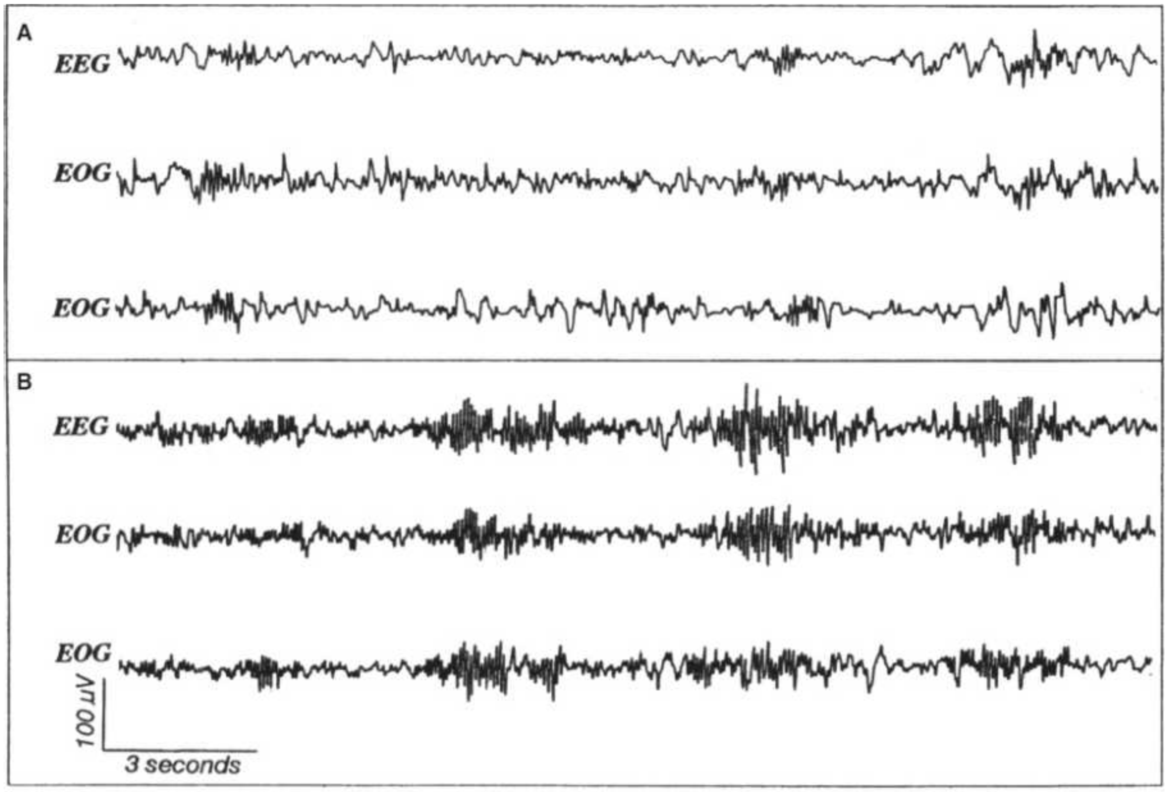

FIGURE 23–12. Polysomnographic tracings illustrating normal and excessive spindle activity. Polysomnographic tracings comparing normal sleep spindle EEG activity during stage 2 sleep (**Panel A**) with that of a patient chronically using a benzodiazepine (**Panel B**). This patient, who had been treated with benzodiazepines for more than a decade (and currently taking an extremely high dose) before being seen at the sleep disorders center. Sleep was grossly abnormal. The most obvious aberration was the tremendous increase in the frequency, magnitude, and duration of EEG spindle activity (see **Panel B**). Additionally, slow-wave sleep was absent, stage 2 was grossly elevated, and spindles even intruded into REM sleep.

CASE STUDY 3 (INSOMNIA)

DESCRIPTION

Ms. W is a 41-year-old female presenting with a two and one half year complaint of sleeplessness. She has difficulty falling asleep (30 to 45 minutes sleep onset latency) and awakens every hour or 2 after sleep onset. These awakenings may last 15 minutes to several hours, and she estimates approximately 4.5 hours of sleep on an average night. She rarely takes daytime naps, notwithstanding feeling tired and edgy. The patient describes her sleep problem with the following words: “It seems like I never get into a deep sleep. I have never been a heavy sleeper, but now the slightest noise wakes me up. Sometimes I have a hard time getting my mind to shut down.” She views the bedroom as an unpleasant place of sleeplessness and states, “I tried staying at a friend’s house where it is quiet, but then I couldn’t sleep because of the silence.”

At times, Ms. W is unsure whether she is asleep or awake. She has a history of clock-watching (to time her wakefulness) but stopped doing this when she realized it was contributing to the problem. The insomnia is unrelated to seasonal changes, menstrual cycle, or time-zone translocation. Her basic sleep hygiene is good. Appetite and libido are unchanged. She denies mood disturbance, except that she is quite frustrated and concerned about sleeplessness and its effect on her work. Her work involves sitting at a microscope for six of a nine-hour working day and meticulously documenting her findings. Her final output hasn’t suffered, but she must now “double-check” for accuracy.

She describes herself as a worrier and a Type-A personality. The patient does not know how to relax. For example, on vacation, she continually worries about things that could go wrong. She will not even begin to unwind until she has arrived at the destination, checked in, and unpacked. Even then, she is unable to relax.

Medical history is unremarkable except for tonsillectomy (age 16 years), migraine headaches (current), and diet-controlled hypercholesterolemia. She takes naproxen as needed for headaches. She does not currently drink caffeinated beverages, smoke tobacco, or drink alcoholic beverages. She does not use recreational drugs.

The problem with insomnia began after relocation to a new city and place of employment. She attributes her insomnia to the noisy neighborhood in which she now lives. She first sought treatment 18 months ago. Her family practice physician diagnosed depression, and she was started on fluoxetine, which made her “climb the walls.” Antihistamines were tried next with similar results. She was then switched to low-dose trazodone (for sleep) and developed nausea. After these medical interventions, she sought medical care elsewhere. Zolpidem 5 mg was prescribed, but it made her feel drugged, and upon discontinuation, she had withdrawal effects. Another family practice physician diagnosed “nonspecific anxiety disorder” and began buspirone, an experience she describes as “having an alien try to climb out of my skin.” Buspirone was discontinued. Paroxetine was tried for 8 weeks with no effect. Finally, a psychiatrist was consulted, who diagnosed adult attention deficit disorder (without hyperactivity) and suggested treatment with methylphenidate. At this point, the patient was convinced that a stimulant would not help her insomnia and demanded a referral to a sleep disorders center.

DISCUSSION OF CASE 3

Ms. W’s symptoms fall into the broad category of insomnia, and the symptoms began after moving from one city to another. Environmental sleep disorder (noise) and adjustment sleep disorder (new job, city, and apartment) are likely initial diagnoses. However, a more chronic, endogenous problem has become operative. What is it? Ms. W is a “worrier” and meticulous but doesn’t reach diagnostic criteria for personality or anxiety disorders. Dyssomnia associated with mood disorder should be considered in any patient with sleep maintenance problems and early morning awakening insomnia. However, this patient does not have other significant signs of depression. Unfortunately, many patients are misdiagnosed with depression or “masked depression” on the sole basis of an insomnia complaint and unsuccessfully treated with antidepressant medication. Ms. W’s job demands long hours with focused concentration. Her job performance has been superior for many years, notwithstanding insomnia. Thus, a diagnosis of attention deficit disorder is unlikely. Idiopathic insomnia implies a childhood complaint, which Ms. W denies.

The likely working diagnosis is psychophysiological insomnia (PPI). There may be some sleep state misperception (sometimes unclear of whether she is awake or asleep), but this cannot adequately account for the constellation of symptoms. An initial treatment plan should include further documentation of the sleep pattern using a sleep log. Behavioral treatments will likely benefit this patient. Medications with sedative effects are sometimes useful during the initial treatment of PPI. However, thus far, they have done more harm than good in this patient. She is likely to be a challenging patient to treat.

CASE STUDY 4 (PTSD AND INSOMNIA)

DESCRIPTION

Mr. M., age 68, is a retired army officer who with a history of combat. He was referred to the sleep disorders center to evaluate chronic complaints of insomnia. Mr. M was in active service on the battlefield for 8 years and suffered what some describe as “blast-wave” or “shell shock,” leading to PTSD from frequent exposure to explosions. After a full evaluation in the psychiatry ward, he was diagnosed as suffering from chronic war–induced PTSD.

Following his retirement from the military, Mr. M was unable to work due to mental illness. Twenty-five years following his wartime experiences, he still complained of disabling insomnia that was resistant to treatment. He described the insomnia as “an umbrella that covers all of his other symptoms,” which included severe aggression and poor relationships with his family, including his wife. He said that he had a good day after a good night’s sleep and that after suffering a bad night, he experienced severe symptoms. He begged the psychiatrist to help him to get better sleep, and he emphasized, “I cannot sleep a wink.” He feels guilty following frequent outbursts of verbal aggression toward his wife. He also feels isolated due to the insomnia and does not participate in any social or leisure activities. Both his wife and his referring physician feel that Mr. M. is malingering; in fact, his wife states that he sleeps well and even upsets her sleep due to his snoring.

A psychiatric evaluation in our clinic yielded a score of 10 on the Pittsburgh Sleep Quality Index (PSQI), which implies poor sleep quality. Two consecutive nights of assessment by polysomnography yielded normal sleep patterns. After excluding significant secondary gain, he was diagnosed with paradoxical insomnia and treated with nonpharmacologic treatment, mainly Cognitive Behavior Therapy for insomnia (CBTi). The therapy was relatively successful: following 8 weeks of therapy, his PSQI score decreased to 6, and he reported subjective satisfaction and improvement in his sleep.

DISCUSSION OF CASE 4

Difficulty falling or staying asleep and nightmares are symptoms included in the DSM-5-TR diagnostic criteria for PTSD. Patients with chronic war-induced PTSD with insomnia complaints often have an insomnia disorder. Similarly, Mr. M. had both chronic war-induced PTSD and comorbid chronic insomnia disorder. This case highlights four points. First, subjective complaints about insomnia should not be ignored, as they reveal the unique personal experience and perspective of a given patient even if they mismatch with objective data. Some might even consider overlooking of subjective perspectives as maltreatment; at least, inaccurate labels of malingering can invoke great distress for patients. Second, when there is a discrepancy between subjective and objective measurements of sleep, with subjective measures indicating greater sleep problems than objective measures, psychotherapeutic interventions such as CBT can be effective. We found that to be the case with Mr. M. Third, careful gathering of subjective sleep complaints is critical. Sometimes patients will express concern about particular aspects of their sleep health, and treatment of those subjective complaints through CBT can improve sleep. We found this also with Mr. M. Fourth, objective measurements should not be ignored. They can be valuable to complete diagnostic processes, evaluate progress, and exclude other sleep disorders such as sleep-related breathing disorders.

In conclusion, subjective and objective measurements are two faces of “the sleep assessment coin” in patients with PTSD. Each perspective has unique value in diagnosing and treating patients.

SUMMARY

Sleep medicine is an important, rapidly evolving discipline. There is great diversity in presenting complaints, etiology, course, and treatment of sleep disorders. Several nosologic systems exist for classifying sleep disorders; however, the ICSD-3 appears to be the most comprehensive. Sleep disorders significantly alter an individual’s health, mood, quality of life, and cognitive ability. Sleep disorders may be primary; however, many are associated with other medical, neurologic, and psychiatric conditions. Disturbances in sleep or wakefulness may also result from substance use or abuse. Sleep disorders are particularly important for psychiatrists because insomnia and hypersomnia commonly occur in a wide range of psychiatric conditions. With appropriate treatment, most sleep disorders can be effectively managed. The extent of improvement in a patient’s general well-being after treatment is usually dramatic. A good night of sleep is as important to general health as nutrition and exercise.

FURTHER READINGS

American Academy of Sleep Medicine. *International Classification of Sleep Disorders*. 3rd ed. American Academy of Sleep Medicine; 2014.

American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications*. WWW. AASM. ORG. American Academy of Sleep Medicine; 2014.

Ancoli-Israel S, Cole R, Alessi CA, Chambers M, Moorcroft WH, Pollak C. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*. 2003;26(3):342–392.

Chokroverty S. *Sleep Disorders Medicine: Basic Science, Technical Considerations, and Clinical Aspects*. 4th ed. Saunders Elsevier; 2017.

Hirshkowitz M, Rose MW, Sharafkhaneh A. Neurotransmitters, neurochemistry, and the clinical pharmacology of sleep. In: Chokroverty S, ed. *Sleep Disorders Medicine*. 3rd ed. Saunders Elsevier; 2009:67–79.

Hirshkowitz M, Sharafkhaneh A. Diagnostic assessment methods. In: Kryger MH, ed. *Atlas of Clinical Sleep Medicine*. Saunders Elsevier; 2010:261–282.

Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine*. 6th ed. Saunders; 2015.

Kushida CA, Littner MR, Morgenthaler T, et al. Practice parameters for the indications for polysomnography and related procedures: An update for 2005. *Sleep*. 2005;28(4):499–521.

Kyle SD, Aquino MR, Miller CB, et al. Towards standardisation and improved understanding of sleep restriction therapy for insomnia disorder: a systematic examination of CBT-I trial content. *Sleep Med Rev*. 2015;23:83–88.

Maquet P. Positron emission tomography studies of sleep and sleep disorders. *J Neuro*. 1997;244(4 Suppl 1):S23–S28.

Morin CM, Bootzin RR, Buysse DJ, Edinger JD, Espie CA, Lichstein KL. Psychological and behavioral treatment of insomnia:update of the recent evidence (1998-2004). *Sleep*. 2006;29(11):1398–1414.

Nishino S, Mignot E. Pharmacological aspects of human and canine narcolepsy. *Prog Neurobiol*. 1997;52:27–78.

Ohayon MM, Dauvilliers Y, Reynolds CF III. Operational definitions and algorithms for excessive sleepiness in the general population: implications for DSM-5 nosology. *Arch Gen Psychiatry*. 2012;69(1):71–79.

Ohayon MM, Reynolds CF III. Epidemiological and clinical relevance of insomnia diagnosis algorithms according to the DSM-IV and the International Classification of Sleep Disorders (ICSD). *Sleep Med*. 2009;10:952–960.

Maski K, Trotti LM, Kotagal S, et al. Treatment of central disorders of hypersomnolence: an American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. *J Clin Sleep Med*. 2021;17(9):1895–1945.

24

Impulse-Control Disorders

24.1 Intermittent Explosive Disorder

Taylor A. Ochalek, Ph.D., and F. Gerard Moeller, M.D.

DIAGNOSIS AND EPIDEMIOLOGY

Intermittent explosive disorder (IED) in DSM-5 and its text revision (DSM-5-TR) represents a disorder of recurrent, problematic, reactive (i.e., affective or impulsive), aggressive behavior that, over the lifetime, affects about 5% to 8% of individuals in the United States. Key diagnostic features of IED of DSM-5-TR include: (1) aggressive outbursts have a rapid onset, typically without any premeditation, last for less than 30 minutes, and usually occur in response to provocation by a known individual; (2) less severe episodes (verbal or noninjurious physical) often intersperse between more severe episodes; (3) the disproportionate impulsive-aggressive response to a psychosocial stressor; (4) aggressive outbursts are impulsive or provoked by anger and cause significant distress or impairment in psychosocial functioning; (5) the individual is above the age of 6 years or comparable developmental level; (6) the aggressive acts are “not better accounted for” by other DSM-5 disorders or drugs or a documented general medical condition (see Table 24.1–1). As described in DSM-5-TR, differential diagnosis section, a diagnosis of IED should not be given to individuals if the aggressive symptoms are better explained by another disorder such as disruptive mood regulation disorder, antisocial personality disorder (ASPD), or borderline personality disorder, delirium, major neurocognitive disorder, personality change due to another medical condition (aggressive type), substance intoxication or withdrawal, ADHD, conduct disorder, oppositional defiant disorder or autism spectrum disorder. “In addition, children ages 6 to 18 years should not receive this diagnosis when impulsive aggressive outbursts occur in context of an adjustment disorder.” However, The DSM-5-TR also suggests that a diagnosis of IED can be made in addition to the diagnosis of ADHD, conduct disorder, oppositional defiant disorder, or autism spectrum disorder if the recurrent impulsive aggressive outbursts are in excess of those usually seen in these disorders and require independent clinical evaluation. Clinical history that is helpful in determining whether the aggressive behavior is better accounted for by another disorder includes whether the aggressive behavior is primarily impulsive and not for secondary gain, does not only occur in the context of a mood disorder, and does not occur only when the individual is intoxicated. If this is not the case, the aggressive behavior is more likely related to another condition.

The DSM-5-TR criteria incorporates criteria that had been considered research criteria previously because studies show that these criteria better identify individuals with elevated levels of aggression, impulsivity, familial risk of aggression, and abnormalities in neurobiologic markers of aggression. Another significant change in DSM-5 IED criteria is the type of aggressive outbursts that should be considered; physical aggression was required in DSM-IV, whereas verbal aggression and nondestructive/noninjurious physical aggression also meet criteria in DSM-5 and DSM-5-TR. In addition, DSM-5 and DSM-5-TR also provides more specific criteria defining frequency needed to meet criteria and specifies that the aggressive outbursts are impulsive and/or anger based in nature and must cause marked distress, impairment in occupational or interpersonal functioning, or be associated with negative financial or legal consequences. Furthermore, because of the lack of research on this disorder in young children and the potential difficulty of distinguishing these outbursts from normal temper tantrums in young children, a minimum age of 6 years (or equivalent developmental level) is required. Finally, especially for youth, the relationship of this disorder to other disorders (e.g., ADHD, disruptive mood dysregulation disorder) has been further clarified.

Table 24.1–1.
DSM-5-TR Diagnostic Criteria for Intermittent Explosive Disorder (American Psychiatric Association, 2013)

A. Recurrent Behavioral outbursts representing a failure to control aggressive behavior, as indicated by one of the following:
   1. Verbal aggression (e.g., temper tantrums, tirades, verbal arguments or fights) or physical aggression toward property, animals, or other individuals, occurring twice weekly, on average, for a period of 3 months. The physical aggression does not result in damage or destruction of property and/or physical assault involving physical injury to animals or other individuals.
   2. Three behavioral outbursts involving damage or destruction of property and/or physical assault involving physical injury to animals or other individuals occurring within a 12-month period.

B. The magnitude of aggressiveness expressed during the recurrent outbursts is grossly out of proportion to the provocation or to any precipitating psychosocial stressors.

C. The recurrent aggressive outbursts are not premeditated (i.e., they are impulsive and or anger-based) and are not committed to achieve some tangible objective (e.g., money, power or intimidation).

D. The recurrent aggressive outbursts cause either marked distress in the individual or impairment in occupational or interpersonal functioning or are associated with financial or legal consequences.

E. Chronological age is at least 6 years (or equivalent developmental level).

F. The recurrent aggressive outbursts are not better explained by another mental disorder (e.g., major depressive disorder, antisocial personality disorder, borderline personality disorder) and are not attributable to another medical condition (e.g., head trauma, Alzheimer’s disease) or to the physiological effects of a substance (e.g., a drug of abuse, a medication). For children ages 6–18 years, aggressive behavior that occurs as a part of an adjustment disorder should not be considered with this diagnosis.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

The age of onset for IED is in adolescence to early adulthood, ranging from 12 to 21 years of age, and is more common in males than females. There is a substantial comorbidity between IED and other psychiatric disorders, with 81% of individuals with IED meeting criteria for at least one other psychiatric disorder, with the most common comorbid disorders being substance use disorders, disruptive behavior disorders, PTSD, ASPD, borderline personality disorder, anxiety disorders, and major depression. In general, the age of onset of IED precedes that of the comorbid disorders, and aggression scores tend to be highest among those diagnosed with IED and a comorbid disorder versus those diagnosed with one of the above disorders alone or IED alone. Studies have also shown that those with IED are at increased risk for engagement in self-harm behaviors and global emotion regulation deficits. For instance, one study demonstrated greater alexithymia and anger rumination among patients with IED compared to psychiatric and healthy control groups, but interestingly also found higher affective empathy in IED versus healthy controls.

ETIOLOGY

The well-documented case of Phineas Gage points out the importance of the prefrontal cortex in impulsive and aggressive behavior. Gage was a railroad worker who survived after a metal rod was blown through his skull in an accident. Prior to the injury he was an upstanding member of the community and a reliable worker. After the injury he was "…fitful, irreverent, indulging at times in the grossest profanity (which was not previously his custom), manifesting but little deference for his fellows, impatient of restraint or advice when it conflicts with his desires,…."

In a review of the literature on frontal lobe dysfunction and violent behavior, Brower and Price concluded that the literature supports an association between frontal lobe dysfunction and increased aggressive behavior, with focal orbitofrontal injury especially being associated with increased aggression. For an individual patient, if an injury were clearly associated with the aggressive behavior, then by DSM-5 criteria the diagnosis of IED would not be made.

However, many violent patients have a history of head trauma due to the nature of their behavior that is not clearly associated with the onset of aggression. It is also unclear what degree of injury is necessary to increase aggressive behavior, making the diagnosis of personality change associated with a general medical condition (head trauma) challenging to separate from IED. A functional magnetic resonance imaging (fMRI) study in patients with IED showed higher error activity and increased activation in a dorsolateral prefrontal cortex subregion, during a color-word Stroop task, possibly indicating a specific neural correlate of clinically high anger expression. Although more research is needed, the literature clearly supports the association between prefrontal cortex brain dysfunction and impulsive aggressive behavior as seen in IED.

Further biologic evidence indicates that the etiology of aggression and IED may be associated with biologic processes. In one study, inflammatory cytokines were higher among individuals with IED versus healthy controls. Consistent with these findings, a genome-wide methylation study found genetic evidence that biologic processes mediating inflammatory response in the immune system, such as cytokine signaling through the interleukin-3 pathway and GABAergic neuronal differentiation were associated with IED.

A third line of biologic evidence supports alteration of neurotransmitter function and impulsive aggression, with the largest body of evidence supporting a role for the neurotransmitter serotonin in aggression. A number of animal studies associate reduced serotonin function and increased aggression. Reduction in the serotonin precursor tryptophan, lesions of the serotonin cell bodies in the dorsal raphe, and serotonin gene knock-out in mice all increase aggression in animal models. In humans, lower levels of the serotonin metabolite 5-hydroxyindole acetic acid (5-HIAA) are associated with violent suicide attempts and impulsive aggression;

however, other studies have found no association between 5-HIAA levels and aggression. Further evidence supporting a role for reduced serotonin function leading to aggression in humans comes from studies using behavioral laboratory aggression paradigms. Studies using these paradigms have examined the role of serotonin using a drink deficient in the serotonin precursor tryptophan, which produces a significant reduction in plasma tryptophan levels and a reduction in brain serotonin. Several studies using this method of reducing serotonin have shown that there is an increase in aggression in human subjects after tryptophan depletion. The role of central serotonin function in IED has been examined in a study in IED patients where the authors found that physiologic responses to central 5-HT stimulation with dexfenfluramine (d-FEN) were reduced as a function of aggression (but not generalized impulsivity) in human subjects. The central 5-HT deficits in these individuals have been attributed to abnormalities involving 5-$HT_{2A/2C}$ postsynaptic receptors in the limbic hypothalamus. Likewise, several studies have found a reduction in impulsive aggressive behavior after treatment with serotonin reuptake inhibitors in patients with prominent histories of impulsive aggressive behavior. Another line of evidence suggests a significant inverse relationship between platelet 5-HT transporter levels and aggression in individuals with IED, further supporting the notion that reductions in brain 5-HT function are associated with aggressive behavior, particularly impulsive aggression.

Other factors that have been associated with impulsive aggression include childhood exposure to violence, maltreatment, and neglect. A recent study found that there was significant association between IED and aversive childhood parenting environment comparable to other psychopathologies. Another study by the same group found that physical abuse in childhood was independently associated with IED, with impulsivity and aggression as the mediators of the relationship between physical abuse and IED. Others have shown that IED is associated with childhood exposure to interpersonal traumatic events. It has been suggested that it is possible that a gene-environment interaction may exist where individuals who have certain serotonin gene polymorphisms may be more prone to impulsive aggressive behavior after exposure to childhood maltreatment.

TREATMENT

Pharmacotherapy

Pharmacotherapy has been a mainstay of the treatment of aggressive behavior in hospital emergency rooms for many years. However, this type of pharmacotherapy targets the immediate threat of aggression, where the issue of specificity of the treatment is not as important. For the treatment of IED, where impulsive aggression is a longstanding pattern of behavior, it is not clinically acceptable to sedate the patient in order to reduce the aggressive behavior over the long term. For pharmacotherapy for IED, the target is reduction in aggression without reduction in other behaviors.

Several medications are known to reduce aggression and prevent rage outbursts, including antidepressants (namely selective serotonin reuptake inhibitors, or SSRIs), mood stabilizers (lithium and anticonvulsants), and antipsychotic drugs. Placebo-controlled studies targeting long-term treatment of impulsive aggression with pharmacotherapy date to the work of Sheard, who found that violent young inmates showed a significant reduction in aggressive behavior after treatment with lithium compared to inmates treated with placebo. There is also evidence that lithium has an effect of reducing suicide, as well as impulsive aggression directed at others. The primary difficulty with the use of lithium for the treatment of impulsive aggression is its relatively narrow therapeutic window and side effects, including nausea, vomiting, and polyuria.

A number of anticonvulsants have been shown to reduce impulsive aggression in patients without seizure disorders. Phenytoin has been shown to reduce aggressive acts in inmates with impulsive aggression in placebo-controlled studies at blood levels that would be considered low for the treatment of epilepsy. Valproic acid in various preparations has also been shown to reduce impulsive aggression. Placebo-controlled trials showed that valproate and divalproex significantly reduce aggressive acts in individuals with personality disorders and conduct disorder. For individuals with personality disorders, the effect on aggression was greatest in individuals with highest trait impulsivity, pointing out the importance of impulsivity in the pharmacotherapy of aggression. Other anticonvulsants with some evidence to support an effect on impulsive aggression include carbamazepine and topiramate. To date, the evidence supporting carbamazepine as an antiaggressive agent is mixed, with some positive studies and a larger negative trial. Topiramate has also been shown to reduce aggression in several studies carried out by the same research group.

Because of the link between lower serotonin function and aggression, a number of trials of SSRIs for impulsive aggression have been undertaken, with five double-blind studies showing a significant reduction in impulse aggression in nonbipolar, nonpsychotic patients treated with SSRIs compared to placebo; however, less than a third of patients showed a full remission of impulsive aggressive behaviors. In another placebo-controlled pilot study, administration of lorcaserin, a 5-HT_{2c} receptor agonist, was associated with significant reductions in provoked aggressive responses among individuals diagnosed with IED. However, lorcaserin was voluntarily withdrawn from the U.S. market given evidence of association of long-term administration with increased cancer risk.

As mentioned previously, the use of high-dose antipsychotic agents as a sedative is not clinically acceptable for chronic treatment of aggression due to the lack of specificity and side effects. However, lower doses of newer-generation antipsychotic agents appear to have a specific antiaggressive effect that is independent of the antipsychotic or sedative effect. Although no controlled studies have specifically examined the efficacy of newer-generation antipsychotic medications in IED, a number of studies have shown a reduction in aggression in children and adolescents with disruptive behavior disorders and autism, as well as in adults with dementia. In children and adolescents with disruptive behavior disorders and subaverage intelligence, doses of 0.02 to 0.06 mg/kg/day of risperidone are associated with reduction in a variety of symptoms including irritability, lethargy/social withdrawal, hyperactivity/noncompliance, and aggression. However, studies in this population have also shown higher side effects in the risperidone-treated subjects, including sedation, extrapyramidal symptoms, and weight gain. A contrasting case report of an adolescent with IED observed no significant improvements in aggression with risperidone and sertraline treatment, but following discontinuation of these medications, observed positive changes when carbamazepine was added to an ongoing regimen of aripiprazole.

In older adult patients with dementia at doses of 0.8 to 2.0 mg per day, risperidone has shown efficacy in reducing aggression and agitation. However, the U.S. Food and Drug Administration has determined that there is an increased risk of death in older adult patients with dementia-related psychosis treated with risperidone of between 1.6 and 1.7 times that seen in placebo-treated patients due to a variety of causes, including cardiovascular and infectious etiologies. This has led to a “black box” warning for risperidone related to these potential side effects. The risk of risperidone compared to other medications should be kept in mind in light of a recent study that found no difference in improvement in agitation and aggression between risperidone and citalopram in patients with dementia.

Psychostimulants, particularly methylphenidate, have been shown to be effective at reducing impulsive aggression in conduct disorder and ADHD. However, their use for the treatment of impulsive aggression outside conduct disorder or ADHD is limited due to the abuse potential of these medications, as well as the potential for inducing mania in subjects with bipolar disorder.

Psychotherapy

Cognitive behavioral therapy (CBT) and contingency management have both shown evidence of reducing aggressive behavior in both adults and children. However, the literature on the treatment of IED per se with psychotherapy is limited. A randomized controlled trial compared group and individual CBT for the treatment of IED (n = 45) with a wait-list control group. After 12 weekly sessions, group and individual CBT reduced aggression, anger, hostile thinking, and depressive symptoms, while improving anger control relative to wait-list participant patients, with improvements persisting at the 3-month follow-up. Those who attended individual therapy sessions also reported an improvement in their overall quality of life.

FURTHER READINGS

American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision. Washington, DC, American Psychiatric Association, 2022.

Brower MC, Price BH. Neuropsychiatry of frontal lobe dysfunction in violent and criminal behaviour: a critical review. *J Neurol Neurosurg Psychiatry*. 2001;71(6):720–726.

*Coccaro EF. Intermittent explosive disorder as a disorder of impulsive aggression for DSM-5. *Am J Psychiatry*. 2012;169(6):577–588.

*Coccaro EF. Psychiatric comorbidity in intermittent explosive disorder. *J Psychiatr Res*. 2019;118:38–43.

Coccaro EF, Fanning JR, Phan KL, Lee R. Serotonin and impulsive aggression. *CNS Spectr*. 2015;20(3):295–302.

Coccaro EF, Lee RJ. 5-HT_2c agonist, lorcaserin, reduces aggressive responding in intermittent explosive disorder: a pilot study. *Hum Psychopharmacol*. 2019;34(6):e2714.

Coccaro EF, Lee R, Coussons-Read M. Elevated plasma inflammatory markers in individuals with intermittent explosive disorder and correlation with aggression in humans. *JAMA Psychiatry*. 2014;71:158–165.

Coccaro EF, Lee RJ, Kavoussi RJ. A double-blind, randomized, placebo-controlled trial of fluoxetine in patients with intermittent explosive disorder. *J Clin Psychiatry*. 2009;70(5):653–662.

Coccaro EF, Lee R, Kavoussi RJ. Aggression, suicidality, and intermittent explosive disorder: serotonergic correlates in personality disorder and healthy control subjects. *Neuropsychopharmacology*. 2010;35(2):435–444.

Coskun F, Akca OF. Treatment of intermittent explosive disorder with carbamazepine. *Clin Neuropharmacol*. 2018;41:82–83.

De Deyn PP, Buitelaar J. Risperidone in the management of agitation and aggression associated with psychiatric disorders. *Eur Psychiatry*.

2006;21(1):21–28.

Fahlgren MK, Puhalla AA, Sorgi KM, McCloskey MS. Emotion processing in intermittent explosive disorder. *Psychiatry Res*. 2019;273:544–550.

Fettich KC, McCloskey MS, Look AE, Coccaro EF. Emotion regulation deficits in intermittent explosive disorder. *Aggress Behav*. 2015;41(1):25–33.

Jenkins AL, McCloskey MS, Kulper D, Berman ME, Coccaro EF. Self-harm behavior among individuals with intermittent explosive disorder and personality disorders. *J Psychiatr Res*. 2015;60:125–131.

*Jones RM, Arlidge J, Gillham R, Reagu S, van den Bree M, Taylor PJ. Efficacy of mood stabilisers in the treatment of impulsive or repetitive aggression: systematic review and meta-analysis. *Br J Psychiatry*. 2011;198(2):93–98.

Lee R, Meyerhoff J, Coccaro EF. Intermittent explosive disorder and aversive parental care. *Psychiatry Res*. 2014;220(1–2):477–482.

Lewitzka U, Severus E, Bauer R, Ritter P, Muller-Oerlinghausen B, Bauer M. The suicide prevention effect of lithium: more than 20 years of evidence-a narrative review. *Int J Bipolar Disord*. 2015;3(1):32.

Malone RP, Delaney MA, Luebbert JF, Cater J, Campbell M. A double-blind placebo-controlled study of lithium in hospitalized aggressive children and adolescents with conduct disorder. *Arch Gen Psychiatry*. 2000;57(7):649–654.

McCloskey MS, Noblett KL, Deffenbacher JL, Gollan JK, Coccaro EF. Cognitive-behavioral therapy for intermittent explosive disorder: a pilot randomized clinical trial. *J Consult Clin Psychol*. 2008;76(5):876–886.

*Moeller FG, Barratt ES, Dougherty DM, Schmitz JM, Swann AC. Psychiatric aspects of impulsivity. *Am J Psychiatry*. 2001;158(11):1783–1793.

Moeller SJ, Frobose MI, Konova AB, et al. Common and distinct neural correlates of inhibitory dysregulation: stroop fMRI study of cocaine addiction and intermittent explosive disorder. *J Psychiatr Res*. 2014;58:55–62.

Montalvo-Ortiz JL, Zhang H, Chen C, Liu C, Coccaro EF. Genome-wide DNA methylation changes associated with intermittent explosive disorder: a gene-based functional enrichment analysis. *Int J Neuropsychopharmacol*. 2018;21:12–20.

Nickel MK, Nickel C, Mitterlehner FO, et al. Topiramate treatment of aggression in female borderline personality disorder patients: a double-blind, placebo-controlled study. *J Clin Psychiatry*. 2004;65(11):1515–1519.

Nickerson A, Aderka IM, Bryant RA, Hofmann SG. The relationship between childhood exposure to trauma and intermittent explosive disorder. *Psychiatry Res*. 2012;197(1–2):128–134.

Pandina GJ, Aman MG, Findling RL. Risperidone in the management of disruptive behavior disorders. *J Child Adolesc Psychopharmacol*. 2006;16(4):379–392.

Pollock BG, Mulsant BH, Rosen J, et al. A double-blind comparison of citalopram and risperidone for the treatment of behavioral and psychotic symptoms associated with dementia. *Am J Geriatr Psychiatry*. 2007;15(11):942–952.

Stanford MS, Helfritz LE, Conklin SM, et al. A comparison of anticonvulsants in the treatment of impulsive aggression. *Exp Clin Psychopharmacol*. 2005;13(1):72–77.

Sheard MH, Marini JL, Bridges CI, Wagner E. The effect of lithium on impulsive aggressive behavior in man. *Am J Psychiatry*. 1976;133(12):1409–1413.

Sinzig J, Dopfner M, Lehmkuhl G. Long-acting methylphenidate has an effect on aggressive behavior in children with attention-deficit/hyperactivity disorder. *J Child Adolesc Psychopharmacol*. 2007;17(4):421–432.

Steiner H, Petersen ML, Saxena K, Ford S, Matthews Z. Divalproex sodium for the treatment of conduct disorder: a randomized controlled clinical trial. *J Clin Psychiatry*. 2003;64(10):1183–1191.

24.2 Pyromania

Taylor A. Ochalek, Ph.D., and F. Gerard Moeller, M.D.

DIAGNOSIS AND EPIDEMIOLOGY

The diagnosis of pyromania requires a pattern of behavior, with multiple episodes of deliberate and purposeful fire setting on multiple occasions. The fire-setting behavior can cause shame to the individual, with one study finding that 38.1% of individuals with pyromania considered committing suicide to stop their fire-setting behavior. There is tension or emotional arousal before the act and relief after the act, similar to OCD and kleptomania. Individuals with pyromania are curious about fires and are fascinated with fire paraphernalia and institutions and people affiliated with the fire departments. There is pleasure or relief when setting fires, or when watching or engaging in its aftermath. In comparison to arson, the fire setting is not done for monetary reasons, to hide a crime, to express a sociopolitical idea, for vengeance, or in response to psychosis or judgment that is impaired by major neurocognitive disorder, intellectual disability, or substance intoxication. The diagnosis of pyromania is not better explained by conduct disorder, a manic episode, or antisocial personality disorder. Pyromania is highly comorbid with other disruptive, impulse-control, and conduct disorders, substance use disorders, gambling disorders, and bipolar disorders.

It is important to differentiate pyromania from other causes of fire setting. Fire setting is a behavior that can be accidental or intentional and is not always a symptom of an underlying psychiatric illness. Fire setting (e.g., playing with matches, lighters) may also occur during childhood due to curiosity and may be part of the developmental process. Assessments for fire setting in children and adolescents include the Children's Firesetting Inventory (CFI), the Firesetting Risk Interview (FRI), the Firesetting Incident Analysis Parent Version (FIA-P), and Firesetting Incident Analysis Child Version (FIA-C). A diagnosis of pyromania is excluded if the fire setting is better explained by conduct disorder, mania, antisocial personality disorder, major neurocognitive impartment, intellectual disability, substance intoxication, or in response to delusions, hallucinations, or a medical condition such as epilepsy.

There is a lack of epidemiologic data on pyromania, but there is data on fire setting without the diagnostic restrictions of pyromania. According to DSM-5, the prevalence of pyromania is unknown, but lifetime fire setting in a sample of U.S. residents 18 and older was found to be about 1.0%. In this sample, Caucasian males aged 18 to 35 years living in the western region of the United States had higher rates of fire setting. The most common comorbidities with fire setting include lifetime alcohol and marijuana use disorders, conduct disorder, antisocial and obsessive-compulsive personality disorders, and the most common motive for fire setting, in general, is revenge and anger. In contrast, pyromania, for which revenge and anger are excluded as causes of fire setting, appears to be quite rare. A study found that only 3.3% of arson recidivists met DSM-IV criteria for pyromania, largely because the majority of the arsonists had committed the act under the influence of alcohol, which is an exclusion criterion in DSM-IV, DSM-5 and DSM-5-TR and in fact, most arsonists do not meet criteria for pyromania. A study of 204 inpatient adults on a psychiatric unit found that 3.4% endorsed current pyromania symptoms, and 5.9% had lifetime pyromania symptoms. What literature does exist about the prevalence of pyromania is in children and adolescents, where prevalence estimates range from 2.4% to 3.5%, with 60% of all fires in large U.S. cities being lit by individuals between 11 and 18 years of age.

ETIOLOGY

Pyromania has been described in the literature for over 200 years, however, the etiology of pyromania is poorly understood. Freud described pyromania as consequence of abnormal psychosexual development. Literature on the origins of pyromania has suggested factors ranging from temperament to environmental factors and parental psychopathology. Other authors have suggested pyromania and the other impulse-control disorders may share a neurobiologic link with substance use disorders. One case describes the onset of pyromania with the use of escitalopram. A single case report found a perfusion deficit in the left inferior frontal lobe on single photon emission computed tomography (SPECT), which resolved after treatment with cognitive behavioral therapy (CBT) and topiramate. Other support for frontal inhibitory dysfunction in pyromania comes from neuropsychologic testing of one subject with pyromania, which found attentional, visual memory, and executive function impairments. However, evidence from a case report suggests that impulsivity and decision-making may not be characteristic of all pyromania diagnoses. In this report, compared to 19 age-matched healthy controls, an adult diagnosed with pyromania demonstrated greater cognitive inflexibility but similar performance on measures of impulsivity and decision-making. These deficits in cognitive inflexibility share features of compulsivity and obsessive-compulsive disorder. Earlier limited studies suggest that pyromania may be associated with lower concentration of norepinephrine and serotonin metabolites 3-methoxy-4-hydroxyphenylglycol (MHPG) and 5-hydroxyindoleacetic acid (5-HIAA).

In a recent case study of a 28-year-old with young onset Parkinson disease, a pathogenic variant of the FBXO7 gene was associated with impulse control disorder symptoms, including firesetting behavior, following dopamine agonist and levodopa treatment. This suggests that some individuals may be genetically susceptible to the effects of dopamine on increasing impulse control disorders, including pyromania; however, further research is needed to characterize the genetic underpinnings of pyromania to better inform novel treatment approaches.

There is a lack of data on the development and course of pyromania. Most individuals with pyromania start setting fires in adolescence or young adulthood. The frequency and intensity of fire setting may increase over time or may wax and wane. The course of pyromania is unknown but smaller studies suggest it may be chronic. Pyromania occurs more frequently in males with poorer social skills and learning difficulties.

TREATMENT

There are no controlled clinical trials in the treatment of pyromania, and there are no FDA approved medications for the disorder. Case reports on pharmacologic management of pyromania have been sparse and results have been inconclusive. Some case reports have shown a reduction in pyromania urges and behaviors with SSRIs, lithium, naltrexone, stimulants, topiramate, valproic acid, carbamazepine, antiandrogen medication, clonazepam, and olanzapine. Other reports have shown no improvement in pyromania with pharmacotherapy.

What little literature does exist suggests CBT and other psychotherapies may be helpful. Other successful behavioral therapies for fire setting include aversive therapy, education, relaxation techniques, positive reinforcement, graphing therapy (i.e., a visual representation to enhance understanding of temporal correlations between a fire-setting event and antecedent emotions and events), and fire safety education. One study found that CBT, fire safety education (FSE), and a home visit from a firefighter (HVF) improved fire-setting behaviors in children who had recently set a fire. However, CBT was the most efficacious in reducing fire setting, interest in playing with matches, and fire-related behaviors. Another study, which examined a group CBT intervention (Firesetting Intervention Program for Mentally Disordered Offenders; FIP-MO) among deliberate fire setters demonstrated that the intervention was associated with significant improvements in several treatment areas, such as attitudes and interests related to fire, anger expression, and a better understanding regarding fire safety and understanding personal fire-setting behavior, compared to the treatment as usual group.

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders,* Fifth Edition, Text Revision. American Psychiatric Association; 2022, Washington, DC.

Bennett BK, Gamelli RL, Duchene RC, Atkocaitis D, Plunkett JA. Burn education awareness recognition and support (BEARS): a community-based juvenile firesetters assessment and treatment program. *J Burn Care Rehabil*. 2004;25(3):324–327.

Blum AW, Odlaug BL, Grant JE. Cognitive inflexibility in a young woman with pyromania. *J Behav Addict*. 2018;7(1):189–191.

Burton PRS, McNiel DE, Binder RL. Firesetting, arson, pyromania, and the forensic mental health expert. *J Am Acad Psychiatry Law*. 2012;40(3):355–365.

Ceylan MF, Durukan I, Turkbay T, Akca OF, Kara K. Pyromania associated with escitalopram in a child. *J Child Adolesc Psychopharmacol*. 2011;21(4):381–382.

Dell’Osso B, Altamura AC, Allen A, Marazziti D, Hollander E. Epidemiologic and clinical updates on impulse control disorders: a critical review. *Eur Arch Psychiatry Clin Neurosci*. 2006;256(8):464–475.

Franklin GA, Pucci PS, Arbabi S, Brandt MM, Wahl WL, Taheri PA. Decreased juvenile arson and firesetting recidivism after implementation of a multidisciplinary prevention program. *J Trauma*. 2002;53(2):260–264.

Freud S. The acquisition of power over fire. *Int J Psychoanal*. 1932;13:405–410.

*Grant JE. SPECT imaging and treatment of pyromania. *J Clin Psychiatry*. 2006;67(6):998.

Grant JE, Levine L, Kim D, Potenza MN. Impulse control disorders in adult psychiatric inpatients. *Am J Psychiatry*. 2005;162(11):2184–2188.

*Grant JE, Kim SW. Clinical characteristics and psychiatric comorbidity of pyromania. *J Clin Psychiatry*. 2007;68(11):1717–1722.

Koles MR, Jenson WR. Comprehensive treatment of chronic fire setting in a severely disordered boy. *J Behav Ther Exp Psychiatry*. 1985;16(1):81–85.

Kolko DJ. Efficacy of cognitive-behavioral treatment and fire safety education for children who set fires: Initial and follow-up outcomes. *J Child Psychol*

Psychiatry. 2001;42(3):359–369.

Kolko D. Firesetting in children and youth. In: Van Hasselt VB, Hersen M, eds. *Handbook of Psychological Approaches with Violent Offenders: Contemporary Strategies and Issues*. Kluwer Academic; 1999.

Lindberg N, Holi MM, Tani P, Virkkunen M. Looking for pyromania: characteristics of a consecutive sample of Finnish male criminals with histories of recidivist firesetting between 1973 and 1993. *BMC Psychiatry*. 2005;5:47.

Milrod LM, Urion DK. Juvenile fire setting and the photoparoxysmal response. *Ann Neurol*. 1992;32(2):222–223.

Parks RW, Green RDJ, Girgis S, Hunter MD, Woodruff PWR, Spence SA. Response of pyromania to biological treatment in a homeless person. *Neuropsychiatr Dis Treat*. 2005;1(3):277–280.

Rix KJ. A psychiatric study of adult arsonists. *Med Sci Law*. 1994;34(1):21–34.

Roncero C, Rodriguez-Urrutia A, Grau-Lopez L, Casas M. [Antiepileptic drugs in the control of the impulses disorders]. *Actas Esp Psiquiatr*. 2009;37(4):205–212.

Tyler N, Gannon TA, Lockerbie L, Ó Ciardha C. An evaluation of a specialist firesetting treatment programme for male and female mentally disordered offenders (the FIP-MO). *Clin Psychol Psychother*. 2018;25:388–400.

*Vaughn MG, Fu Q, Delisi M, et al. Prevalence and correlates of fire-setting in the United States: results from the National Epidemiological Survey on Alcohol and Related Conditions. *Compr Psychiatry*. 2010;51(3):217–223.

Virkkunen M, Nuutila A, Goodwin FK, Linnoila M. Cerebrospinal fluid monoamine metabolite levels in male arsonists. *Arch Gen Psychiatry*. 1987;44(3):241–247.

*Yoo D, Choi J-H, Im J-H, et al. Young-onset Parkinson's disease with impulse control disorder due to novel variants of F-box only protein 7. *JMD*. 2020;13:225–228.

24.3 Kleptomania

TAYLOR A. OCHALEK, PH.D., AND F. GERARD MOELLER, M.D.

DIAGNOSIS AND EPIDEMIOLOGY

Kleptomania is one of the least studied of the disruptive, impulse-control, and conduct disorders. As conceptualized in DSM-5 and its text revision (DSM-5-TR), the key diagnostic feature of kleptomania is that the items stolen are not needed for personal use or for their monetary value and are not in response to anger, vengeance, a delusion, or a hallucination (see Table 24.3–1). The diagnosis of kleptomania also includes a sense of tension immediately before the theft and a sense of relief at the time of the theft, similar in structure to obsessive-compulsive disorder (OCD). The individual may hoard or return the stolen objects. Usually, the thefts are not preplanned, and very little attention is given to the consequences of the theft. The urge to steal is ego dystonic and distressing. Subsequently, people with kleptomania experience guilt, depression, and suicide. The emotional stress, jail time, and marital and relationship conflict from shoplifting have been shown to decrease quality of life in individuals with kleptomania. As with other impulse control disorders, the diagnosis of kleptomania is not better explained by conduct disorder, a manic episode, or antisocial personality disorder.

DSM-5-TR lists the prevalence of kleptomania in the general population to be 0.3% to 0.6%. The prevalence of kleptomania in those arrested for shoplifting ranges from 3.8% to 24%. A study of 204 consecutive admissions of psychiatric inpatients found a current prevalence of 7.8% and a lifetime prevalence of 9.3%, implying that the disorder may be more common than previously thought. Some of the underdiagnosis of kleptomania may be related to the fact that the disorder is ego dystonic and patients are embarrassed to discuss their symptoms. The course of kleptomania is chronic with waxing and waning of symptoms. Kleptomania may begin in childhood, adolescence, or adulthood, but commonly begins during adolescence, with an average age of onset of 18.75 years. Kleptomania occurs more frequently in women, with about two-thirds of kleptomania patients being women. Females with kleptomania are more likely to be married, shoplift household items, store stolen items, shoplift at a later age, and have a comorbid eating disorder or other impulse-control disorder. Kleptomania is highly comorbid with compulsive buying, mood disorders, anxiety disorders, eating disorders, personality disorders, substance use disorders, as well as other disruptive, impulse-control, and conduct disorders. The rate of OCD in kleptomania has ranged from 6.5% to 60%. Due to the high comorbidity between OCD and kleptomania, it has been suggested that kleptomania may be a variant of OCD.

Table 24.3–1.
DSM-5-TR Diagnostic Criteria for Kleptomania (American Psychiatric Association, 2013)

A. Recurrent failure to resist impulses to steal objects that are not needed for personal use or for their monetary value.
B. Increasing sense of tension immediately before committing the theft.
C. Pleasure, gratification, or relief at the time of committing the theft.
D. The stealing is not committed to express anger or vengeance and is not in response to a delusion or a hallucination.
E. The stealing is not better accounted for by Conduct Disorder, a Manic Episode, or Antisocial Personality Disorder.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

Kleptomania is associated with increased impulsivity, with a study showing that patients with kleptomania had higher scores on the Barratt Impulsiveness Scale, greater cognitive impulsivity, and greater sensation seeking in comparison to a psychiatric group without impulse control or substance use disorders. Furthermore, individuals with greater severity of kleptomania symptoms have been shown to demonstrate more executive functioning impairment on the Wisconsin Card Sorting Test (WCST), greater cognitive impulsivity, and greater sensation seeking.

Higher impulsivity may also be associated with shoplifting-related legal consequences among individuals with kleptomania. One study conducted on 107 adults with kleptomania reported higher impulsivity scores on the Eysenck Impulsiveness Questionnaire (EIQ) among those with at least one shoplifting-related arrest compared to those without a shoplifting-related arrest history. In contrast, another study did not observe an association between impulsivity and kleptomania symptom severity on the EIQ and found that lower ADHD symptoms were actually associated with greater symptom severity.

In a study that examined the clinical correlates of kleptomania symptom severity among 112 adults with kleptomania, comorbid diagnosis of OCD or eating disorder (i.e., anorexia nervosa, bulimia nervosa), a higher frequency of stealing-related urges and feelings of excitement, stealing from relatives, and a lower prevalence of family history of substance use disorder(s) were significantly associated with greater symptom severity.

ETIOLOGY

The Swiss physician Mathey described kleptomania as early as the 19th century, but little is known about the etiology of kleptomania. There have been case reports of kleptomania being associated with head trauma, frontal lobe lesions, cortical atrophy, dementia, and hypoglycemia secondary to insulinoma. On diffusion tensor imaging, there is reduced fractional anisotropy in the inferior frontal white matter in patients with kleptomania, implying that the disorder may be associated with subtle white matter pathology. Similarly, in a neuroimaging case study, an adult with overall normal neurocognitive functioning diagnosed with kleptomania, OCD, and CREST (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia) exhibited hyperactivity in the right ventral striatum region, and CREST-related white matter damage in the orbitofrontal region. This indicates that white matter lesions in the orbitofrontal brain region may play a nuanced role in the etiology of kleptomania. Another study found that although as a group patients with kleptomania did not show deficits on neuropsychological testing, the severity of illness was associated with impairment in executive functioning.

Data is lacking and limited on the genetics of kleptomania. A study found that individuals with kleptomania had more first-degree relatives with alcohol use disorders compared to controls. Treatment studies (discussed below) have focused on medications that target serotonin and dopamine, as well as the opioid and glutamatergic neurotransmitter systems; however, there is not a large body of data showing these neurotransmitters play a role in kleptomania, although there are higher rates of OCD in relatives of individuals with kleptomania compared to the general population, suggesting a potential common neurobiology.

TREATMENT

Limited data exist to guide treatment options. To monitor results of treatment, various scales have been used. Although the Yale-Brown Obsessive Compulsive Scale (Y-BOCS) is not a validated rating scale for kleptomania, it has been used to monitor treatment response in kleptomania. The Kleptomania Symptom Assessment Scale (K-SAS), a self-report measure, has been shown to have adequate psychometric properties in assessing changes in behaviors, urges, and cognition during treatment.

Pharmacotherapy

Due to the guilt associated with the behavior, most individuals with kleptomania seek treatment after being arrested. Pharmacologic studies for kleptomania have been limited to mostly case reports and case series. Selective serotonin reuptake inhibitors have been tried based on the notion that kleptomania may be an obsessive-compulsive spectrum disorder. However, case reports have shown mixed results for SSRIs, with some reports of improvement, but at least one report described patients who developed kleptomania during treatment with SSRIs. There have also been case reports of lithium and topiramate reducing urges to steal in kleptomania. An open-label trial found that the mu opiate receptor antagonist naltrexone reduced urges to steal and stealing behavior in kleptomania. In an individual case study of a patient with intellectual disability, aggressive behavior, and kleptomania, naltrexone administration decreased both stealing behavior and aggression. In another open-label trial, the N-methyl-D-aspartate (NMDA) receptor antagonist memantine reduced urges to shoplift and shoplifting behavior, along with improving impulsivity, mood, anxiety, and psychosocial functioning. In one of the few placebo-controlled trials in the treatment of kleptomania, there was a significant effect of naltrexone on kleptomania symptoms, with 66% of the patients that received naltrexone achieving remission, whereas 7.7% of those receiving placebo reported symptom remission. The only controlled trial of antidepressants in the treatment of kleptomania found no benefit of escitalopram in the treatment of kleptomania.

Psychotherapy

There is little literature regarding the treatment efficacy of psychotherapy for kleptomania. Case reports and case series suggest that cognitive behavioral therapy (CBT) approaches may be helpful. One recent case report demonstrated the initial feasibility of combining exposure and prevention therapy (ERP) with CBT and biofeedback in reducing stealing behavior and improving kleptomania symptoms. Findings demonstrated that ERP + CBT reduced kleptomania symptoms and sympathetic nervous system arousal (measured via galvanic skin response) from baseline; however, the addition of naltrexone did not provide further clinical benefit. Although biofeedback offers an innovative and individualized approach to support psychotherapy, larger-scale studies are needed to examine its reliability and generality.

FURTHER READINGS

Aizer A, Lowengrub K, Dannon PN. Kleptomania after head trauma: two case reports and combination treatment strategies. *Clin Neuropharmacol*. 2004;27(5):211–215.

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition*. American Psychiatric Publishing Inc.; 2013.

*Bayle FJ, Caci H, Millet B, Richa S, Olie JP. Psychopathology and comorbidity of psychiatric disorders in patients with kleptomania. *Am J Psychiatry*. 2003;160(8):1509–1513.

Blum AW, Odlaug BL, Redden SA, Grant JE. Stealing behavior and impulsivity in individuals with kleptomania who have been arrested for shoplifting. *Compr Psychiatry*. 2018;80:186–191.

Christianini AR, Conti MA, Hearst N, Cordás TA, de Abreu CN, Tavares H. Treating kleptomania: cross-cultural adaptation of the Kleptomania Symptom Assessment Scale and assessment of an outpatient program. *Compr Psychiatry*. 2015;56:289–294.

25

Adjustment Disorders

Jeffrey W. Katzman, M.D., and Cynthia Geppert, M.D., Ph.D., D.P.S., M.A., M.P.H., M.S.B., M.S.J., F.A.C.L.P., D.F.A.P.A., F.A.S.A.M., H.E.C-C.

The diagnostic category of adjustment disorders (ADs) is widely used among clinicians in practice. The construct is often criticized for lack of scientific support, though research in the area has increased in the last 5 years, and world events have demonstrated its clinical utility. The DSM-5 classification of AD among a group of stress response syndromes underscores that the central heuristic feature of AD is the linkage between the development of symptoms to an external stressful event. The event often involves financial, legal, or relationship difficulties or a new medical diagnosis. The adjustment disorder construct is often criticized for lack of scientific support, though interest in the study of this diagnosis appears to have grown over the past 5 years. The diagnosis of AD strays from the general phenomenologic approach of the DSM. The fifth edition of the diagnostic and statistical manual does not provide observable symptom criteria, such as the familiar list of signs and symptoms that characterize the other major diagnostic categories.

The link between an environmental stressor and symptom formation is consistent with paradigms many clinicians utilize and with the health beliefs of the general public about the adverse effects of stress. This may account somewhat for the appeal of this diagnostic category for patients and practitioners alike. The development of psychophysical symptoms in response to distressing experiences is a recognized part of the human condition, making the diagnosis often less socially stigmatizing than other psychiatric disorders. Furthermore, medical insurance carriers historically have not considered the diagnosis as a preexisting condition in the same way as they have labeled other psychiatric disorders such as major depression. This distinction frequently leads clinicians to use the diagnosis to protect patients from future discrimination in employment matters and medical, life, and disability insurance applications.

Ultimately, it is important to note that, despite the historically problematic nature of this clinical entity, the AD diagnosis was ranked seventh of 44 diagnostic categories by psychiatrists in a 2018 study by the World Health Organization, inquiring about the most frequent diagnoses used within a 1-week period. In a similar worldwide study in 2013, thousands of psychologists worldwide asked the same question and ranked the AD diagnosis ninth. Yet, in these same studies, the AD diagnosis ranked in the lowest percentile of ease of use or goodness-of-fit in clinical practice.

COMPARATIVE NOSOLOGY: DSM

Despite the clinical appeal of this category of diagnoses, ADs as a diagnostic entity have been seen as problematic for a number of reasons, and the reconfiguration of DSM-5 has done little to address these problems. The diagnostic criteria describe a syndrome in which a stressful event leads to the development of a symptom complex. However, within the diagnostic construct, there are no qualitative or quantitative criteria to determine what stressor meets the diagnostic threshold as there is for post-traumatic stress disorder (PTSD). Similarly, the symptom complex that develops has been criticized as lacking specificity, which is the obvious drawback of its generalizability. Unlike DSM-5 (and its text revision, DSM-5-TR), ICD-11 has to some degree addressed this problem, thereby creating a notable difference in nosology between the two primary classification systems. This shift in definition in ICD-11 raises some particular concerns regarding research on AD, as the DSM and ICD-11 are describing potentially different psychiatric entities at this point. Critics also contend that the temporal course between the stressor and the development of symptoms lacks rigorous scientific evidence. Furthermore, it is difficult in clinical practice to directly link an event to the development of a symptom complex. These issues have raised questions about the diagnosis in general and have no doubt contributed to the lack of academic investigation. Perhaps the greatest dilemma in considering this diagnosis like others in the trauma and stress-related DSM-5 category involves the question as to why some individuals develop symptoms in response to a stressful event while others do not, making AD a perfect vehicle for studies of the often-underappreciated role of resilience in the face of stressful life circumstances.

The entire formulation of an AD has been impugned with disparaging designations such as a “wastebasket” or “afterthought” diagnosis. Despite recognition among experts working on DSM-5 of the need to clean up the diagnosis of AD, the final version left the wastebasket full. Academic literature on AD identifies three main nosologic weaknesses. The first and most serious objection is that the diagnosis is yet another manifestation of the increased medicalization of life problems, as well as the hegemony of biologic models in psychiatry at the expense of social, psychological, cultural, and spiritual conceptualizations and responses to illness. The second critique points to the diagnosis’ failure to present a consistent, coherent, and clear explication of the essential relationship between an external stressor and internal diathesis. While the criteria require the stressor to precede the development of symptoms in a causative manner, no specific symptoms, duration, nature, or magnitude of stressor that would trigger the development of AD has been delineated. It is extremely difficult in clinical practice to infer causality from the stressor to the development of symptoms. A clinician must then assess whether an individual’s symptoms are above and beyond what would be considered normal for that particular stressor. This raises one of the central philosophical issues in postmodern psychiatry: where normality ends, and pathology begins.

The third “wastebasket” criticism refers to the use of AD for clinical presentations that fail to meet the criteria of other psychiatric diagnoses—particularly major or minor depressions. Similarly, the predictive validity of the diagnosis has been criticized in that many initial diagnoses of AD are later changed to another primary affective, substance use, or personality disorder. Frequently the main factor in that change is the duration of the disturbance and not the degree of severity. Once the temporal threshold has been met for another anxiety mood or even trauma disorder, the diagnosis is changed, yet the arbitrariness of these time periods is well-recognized. Claiming AD is transformed suddenly into another entity after 2 weeks or 2 months is not sound science. However, defenders of the diagnosis argue that the nosologic problem lies more with the vague, abstract, and over-general formulation of the criteria than the actual concept of a time-limited maladaptive reaction to stress. This has begun to be addressed in the most current version of ICD-11. Despite these deficiencies, the diagnosis has had an enduring clinical and heuristic value that warrants a more nuanced understanding of its utility and meaning.

HISTORY

ADs as a personal reaction to an environmental or social stressor are likely universal phenomena, which have occurred in all historical epochs and cultures. A recent author suggests that the first recognizable clinical description of an AD and its appropriate treatment is in the 11th-century writings of the Islamic physician–philosopher Avicenna. Historically, DSM has always included some version of a diagnostic category involving an acute psychological response to an environmental stressor. The initial impetus for this came from documentation of severe wartime stress seen in World War II, as well as through the evolution of crisis intervention theory and practice. DSM-I, published in 1952, described the category of Transient Situational Personality Disorder. Within this category were the subtypes of gross stress reaction, adult situational reaction, adjustment reaction of infancy, adjustment reaction of childhood, adjustment reaction of adolescence, and adjustment reaction of late-life. It is notable that even the first appearance of this diagnosis displays an overlap in the diagnostic symptoms between the concepts of a stressor, depressive and anxious affect, and one’s personality.

The diagnosis was modified somewhat in DSM-II, which changed the concept to transient situational disorder. This category was reserved for “more or less transient disorders of any severity (including those of psychotic proportions) that occur in individuals without any apparent underlying mental disorders and that represent an acute reaction to overwhelming environmental stress.” DSM-II provided similar developmental subtypes to DSM-I, with the exception that the subtypes of gross stress reaction and adult situational reaction were eliminated.

DSM-III introduced the formal diagnosis of AD. The developmental periods of the earlier diagnostic systems were removed. Instead, the subtypes of AD were categorized based on the predominant affective experience. These included AD with depressed mood, anxious mood, mixed emotional features, disturbance of conduct, mixed disturbance of emotions and conduct, work inhibition, withdrawal, and atypical features. DSM-III-R retained these diagnostic subtypes and added an additional one involving physical complaints. DSM-III-R also specified that symptoms of an AD could not exceed 6 months.

DSM-IV modified the diagnosis of adjustment disorder in several ways. The subtypes of mixed emotional features, work inhibition, withdrawal, and physical complaints were eliminated. The stressor was allowed to persist for an indefinite period of time. A descriptor of chronicity was specified, whereby symptoms persisting for greater than 6 months were deemed chronic.

DSM-5

DSM-5 did not make substantial changes to the actual diagnostic criteria for AD. A key change in classification did occur that, in the view of one author, “represents a major shift in thinking and is the most significant change since its introduction in the DSM-III.” AD is now classified with reactive attachment disorder, disinhibited social engagement disorder, PTSD, acute stress disorder, other specified trauma- and stressor-related disorders, and unspecified trauma- and stressor-related disorders in a newly delineated category of Trauma- and Stressor-Related Disorders. In some measure, this reorganization settles, at least for the time being, the long historical debate about whether AD belongs in the family of mood, anxiety, or stress disorders. The authors of DSM-5 chose to emphasize most strongly the nature of AD as one of a range of human responses to stress and trauma and to conceptualize the disturbances of mood, anxiety, and conduct that constitute the established subtypes of AD as varied aspects of the stress–trauma response spectrum.

In prior versions of DSM with a multiaxial diagnostic system, AD was coded on Axis I, with the specific stressor coded on AXIS III or IV on the problematic axial system that DSM-5 has eliminated. Under that classification schema, the diagnosis of AD could not be assigned if an Axis I disorder had already been diagnosed. This restriction generated further obfuscation of the diagnostic picture, which in turn could lead to inappropriate treatment recommendations. In this way, AD was and continues to be considered a subthreshold diagnosis within DSM, as it is not coded in the presence of another Axis I diagnosis.

DEFINITION

ADs are a diagnostic category characterized by an emotional response to a stressful event. The stressors involved for the other diagnoses in the DSM-5 category of Trauma and Stress-Related disorders have criteria, which are fairly well-delineated. As discussed above, there are currently no such criteria for the stressor of an AD. Typically, the stressor involves financial issues, a medical illness, or a relationship problem. The symptom complex that develops may involve anxious or depressive affect or may present with a disturbance of conduct. There is little specificity to qualify these symptoms other than an excess of what would be expected from the particular stressor or significant social or occupational impairment. By definition, the symptoms must begin within 3 months of the stressor and must remit within 6 months of removal of the stressor. The symptom complex must not qualify for another Axis I condition. Bereavement is an exclusion. The subtypes of AD did not significantly change in DSM-5: AD with depressed mood, anxious mood, mixed anxiety and depressed mood, disturbance of conduct, mixed disturbance of emotions and conduct, and unspecified type. Recent research has questioned the value of maintaining these different categories of AD, as there appears to be more in common between individual subtypes than evidence for distinct diagnostic entities. Glaesmer et al. demonstrated that the correlation between different subtypes was extremely high, suggesting that these subtypes were not adequately distinguishable. This finding has been replicated in more recent studies, which all found that subtypes of DSM-5 were highly intercorrelated, suggesting that there is insufficient evidence at present to substantiate the existence of distinct subtypes of AD. In an outpatient sample, no differences in sociodemographics, comorbid diagnoses, or lifetime history of the disorder could be found in groups diagnosed with the anxious and depressive subtypes of AD. The author raised the question of collapsing these subtypes into one diagnostic entity.

STRESSOR CRITERIA

DSM-5-TR criteria for AD rely first and foremost on the presence of a stressor. “The presence of emotional or behavioral symptoms in response to an identifiable stressor is the essential feature of adjustment disorder.” The explanatory material accompanying AD criteria in DSM-5-TR attempts to sketch the nature of the stressors, yet the contours remain unclear, especially when compared with the relatively sharper profile of the trauma of PTSD or the obvious stressor in bereavement. The text contains multiple examples of this ambiguity. For instance, the stressor can be single, like the loss of a significant relationship, or multiple, such as combined employment and marital problems. The stressor can be recurrent, such as a construction worker who cannot find work during the winter, or continuous, most commonly with AD from a chronic illness. In contrast with the individualistic focus of most of the disorders in DSM, AD can be familial or communal in its impact, as can PTSD. The implicit developmental triggers of AD—well known to clinicians—are made explicit, such as the "empty nest syndrome" or a child going off to college. This leaves the clinician to determine whether any particular stressor could lead to the development of a stress response for a specific person. The vast range and variance in putative stressors are quite large, and it is difficult to reasonably believe that such a divergent group of human experiences would lead to a similar psychological experience.

SYMPTOMS

The description of AD symptoms that develop in response to these diverse stressors is fairly vague in DSM-5. They are either a reaction that would be greater than expected given cultural and contextual factors, or they significantly impair social, occupational, or other domains of functioning. The implication of the current DSM-5 criteria is that there is an expectable reaction to a particular stressor. Clinicians are left to determine what response might be "expectable" to any identifiable event. Any such assessment of the response would be influenced by the personal history, values, and culture of the individual examiner and the patient.

TEMPORAL COURSE

DSM-5 did not change the time course of AD found in prior versions of the manual. The onset of symptoms of AD must be within 3 months of a stressor and the resolution of symptoms within 6 months of the termination of the stressor unless another major psychiatric diagnosis preempts AD. There is some evidence to support the idea of symptom development closely following the time course of the stressor. Two studies have demonstrated the close temporal relationship between a stressor and symptom formation. In one study of children with a new diagnosis of insulin-dependent diabetes, nearly one-third developed ADs. The highest risk period was in the first month following diagnosis, and 31 of 33 children had developed symptoms within the first 3 months following the diagnosis of diabetes. The current criteria for AD in ICD-11 capture this idea, in that, unlike DSM-5, symptoms must develop within 1 month of the onset of the stressor.

It is likely that symptom formation in some individuals might come with a considerable time lag. The "with delayed expression" specifier for PTSD allows for the development of symptoms 6 months or later after the time of the stressor. DSM-5 includes a similar exception under the "Other Specified Trauma- and Stressor-Related Disorders" that lists a presentation of AD with delayed onset 3 months after a stressor. The literature confirms the potential temporal variation in symptom formation from the onset of the stressor. One study examined the timing and number of life events occurring in the year before the onset of emotional disorders in school-aged children in an outpatient clinic. Thirty percent of cases did not develop symptoms until 6 months to 1 year following the undesirable event. Children who experienced multiple negative life events were more likely to have less time between a stressor and symptom formation.

The diagnosis of AD specifies that the symptoms do not persist for more than 6 months once the stressor—or its consequences—has terminated. There is a DSM-5 diagnosis included in the Other Specified Trauma- and Stressor-Related Disorder classification for an AD-like disorder with a duration of symptoms prolonged for greater than 6 months past the point where the stressor is no longer active. It is often extremely difficult for a clinician to gauge when a stressor is no longer impacting an individual. In many cases, stressors leading to this diagnosis, such as chronic illness, retirement, or marital tensions, may last indefinitely. Rather, the meaning of the stressor to the individual may change over time, for instance, with a support group or couples' therapy. Scholarship supports the notion that emotional symptoms can linger for quite some time. A study compared the long-term mental health effects in adults suffering from two different community stressors (the Three Mile Island accident and widespread unemployment due to layoffs). Symptoms remained elevated for as long as 42 months after both stressors. It is unclear, however, if this data shows that an event can lead to persistent emotional responses or whether the event itself is ongoing and thus continues to impact an individual. More recent studies suggest that symptoms of AD may become chronic for nearly one-third of individuals.

A study at the University of Michigan captures the enduring nature of stress. The research identified increased stressors, including marital breakups and stress, and financial issues as lasting symptoms following the September 11, 2001, terrorist attacks in a city 600 miles from the actual events. In addition, there was an increased rate of medically serious suicide attempts in the 2-year period following the attacks compared to the previous 2 years. These results speak to the potential chronicity of stress following a single traumatic event and may be applicable to the development of chronic AD as well.

ICD

The *International Statistical Classification of Diseases and Related Health Problems* (ICD-11) also contains a category of ADs. ICD-11 revised its ADs nosology with this version. With his change, ICD-11 aims to make the AD criteria more specific and amenable to academic research, though it now describes a syndrome fairly distinct from that described in DSM-5. The diagnosis had long been similar to the DSM entity in outlining the development of psychological symptoms following a stressor. However, in ICD-11, the symptoms must appear within 1 month of the stressor rather than the 3-month temporal course of DSM-5. Like DSM-5, the ICD-11 criteria require that symptoms must not persist for greater than 6 months after the removal of the stressor. While DSM-5 removed the acute and chronic specifiers for the subtypes of AD, ICD-11 includes a category of Prolonged Depressive Reaction for AD-like symptoms that persist for up to 2 years.

The ICD-11 makes a major shift from ICD-10 and from DSM-5 in the description of symptoms. These are no longer simply vague emotional or behavioral symptoms but rather involve the specific category of preoccupation with the stressor. This may take the form of either excessive worry about the stressor or recurrent and distressing thoughts about the event reminiscent of illness anxiety disorder and PTSD in DSM-5. The ICD-11 symptoms must also include the idea of a “failure to adapt.” This involves significant impairment in function in at least one important life domain. This is different from DSM-5, which suggests that symptoms may cause either a significant impairment in function *or* disproportionate distress given the severity of the stressor. Interestingly, thus far, it is unclear if this addition of more specific symptoms of AD in ICD-11 stands up to rigorous research. An analysis by Kazlaukas et al. of 10 studies using the Adjustment Disorders New Module (ADNM), based on the ICD-11 definition, suggests that empirical data do not provide sufficient evidence for the current ICD-11 symptom structure. While ICD-11 provides greater symptom specificity, it retains the status of a subsyndromal diagnosis as it is not made in the presence of another psychiatric condition, relegating it generally to the category of a minor psychiatric condition. Nevertheless, the presence of greater symptoms specificity has led to more European research in the area of AD, with the emergence of 10 new studies in the 3 years following the release of ICD-11 criteria in 2016. It is likely that this encouraging trend will continue to facilitate additional research in the area of AD specificity.

RELIABILITY

The literature contains mixed information about the reliability of the AD construct. The results of the UK-World Health Organization (UK-WHO) study of the reliability of the ninth revision of International Statistical Classification of Disease categories in children and adolescents were consistent with this. The interrater reliability for

ADs was 0.23; this was lower than for many other categories. The clinical relevance of poor reliability was underscored in a study that examined the concordance rate between an initial diagnosis of AD given by a house officer and a subsequent diagnosis by a consulting psychiatrist. Fifty percent of the inpatients with AD actually had delirium.

VALIDITY

There is a growing body of research that supports the content and predictive validity of the diagnosis. Content validity has been demonstrated through several studies that differentiated AD from major depression on the one hand and normal reactions to stress on the other. It should be noted that results from the European Outcome of Depression International Network (ODIN) research failed to demonstrate content validity in that it identified no variables that might distinguish AD from depression.

For instance, a study of consultation cases in the medical setting found that patients given the AD diagnosis also had higher ratings on Axis IV, lower illness severity, and were more likely to be improved at the time of discharge.

Predictive validity was demonstrated through a study involving a group of 2,078 male Navy-enlisted personnel diagnosed with transient situational disturbance. This proved to be a less disabling diagnosis in terms of chronicity, length of hospitalization, and ability to return to a previous job than any other psychiatric condition diagnosed. The predictive validity of the diagnosis was also demonstrated in a large study of adult inpatients. The investigators sought to determine whether patients with AD shared a similar prognosis at 5 years. Seventy-nine percent of adults in the study were well at 5-year follow-up. The diagnosis was not as predictive for adolescents with 57% well at 5 years. The authors suggest that the adult cohort’s good prognosis implies specific characteristics of the AD group, and that indicates the diagnosis carries with it a prediction of recovery at 5 years. The cohort of children in the sample, however, does not exhibit this same predictive validity.

A study of Danish conscripts supports both content and predictive validity. Subjects diagnosed with AD showed a steady increase in symptoms of depression, sleep, personality problems, and somatization from the time of entering the service until discharge. Fourteen days after discharge, scores on all symptom domains were comparable to those at entrance into the military, and 1 year later, except for a higher depression score, subjects were comparable to controls. Subjects with AD also differed from general conscripts on personality profiles at enrollment, having more adverse family events, depression, and personality problems.

Historically, AD was not included as a potential diagnosis on most structured diagnostic measurements, including The Schedule for Clinical Assessment in Neuropsychiatry (SCAN; Wing et al., 1990), Structured Clinical Interview for DSM-IV (SCID; First et al., 2002), and the Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). However, new formal assessment measures for AD have been developed. The Diagnostic Interview for Adjustment Disorder (DIAD) is a structured interview based on DSM AD criteria. It includes 29 items that assess symptoms associated with a stressor and delineates the distress and impairment experienced as a result of these. Initial data demonstrates “moderate to good” construct validity. The ADNM developed by Glaesmer et al. in 2015 is based on three groups of symptoms originally proposed for the ICD-11 diagnostic revision. These include symptoms of intrusion, avoidance, and failure to adapt. While this scale likely needs further exploration, it has been applied more recently to community samples to identify the prevalence of AD. Validation studies of this instrument have demonstrated good levels of diagnostic specificity and sensitivity, with shorter forms also demonstrating high-construct validity.

CULTURAL PERSPECTIVES

An interesting discussion is posited regarding the utility of the AD diagnosis through the lens of cultural anthropology. DSM-5 cautions the clinician attempting to decide whether or not a given response is disproportional to the expected reaction to a stressor to "take into account the external context and the cultural factors that might influence symptom severity and presentation." This perspective considers the evidence that most cultures have identified a psychological phenomenon whereby an individual is stressed to the point of symptom development. This parallels the lay concept of a "nervous breakdown" within a particular cultural setting. In this sense, the ADs share features with other culture-bound syndromes such as Susto, Koro, and Arctic Hysteria in that they all describe the development of a set of symptoms following a particular stressor. DSM-5 includes these and other culture-bound syndromes in the category of Other Specified Trauma- and Stressor-Related Disorders and specifically mentions *Ataque de Nervios.* The particular cultural presentation of an AD will be affected by: (1) the nature, intensity, and meaning of the stressor in question; (2) the nature of the modal personality configuration of the people involved—which includes style and/or rules about behavior and emotional expression; (3) idiosyncratic features of the self in question; and (4) the meaning that AD has in the culture. The individual is first treated within the family and other social support networks. When this fails, the event is inevitably medicalized in some fashion.

The universality of a maladaptive reaction to disturbing but less-than-catastrophic traumas is shown in studies in non-Western countries.

Studies with refugees, immigrants, victims of torture, and expatriate populations suggest that cultural conflict, problematic acculturation, social or economic displacement, or political disenfranchisement or persecution are all contemporary stressors that may lead to the development of AD in a global context.

Experiences of black persons and people of color, as well as marginalized populations in societies, put them at greater risk for the development of AD. Around the world, individuals of diverse religious, cultural, ethnic affiliations, as well as LGBTQ individuals, face discrimination and often violence. A long history of police brutality toward African Americans in the United States reached a nadir in 2020 with the tragic deaths of George Floyd, Breonna Taylor, and Ahmaud Arbery. They, among many others, are the victims of 400 years of mistreatment of black people, reflecting rampant structural racism in America that has led to protests and cries to defund the police. Such cultural inequities can lead to an experience of grief and anger that understandably may impact other areas of one's life. There is a heavy burden of economic and political injustice and health disparities that persons of color have long experienced in American society. It becomes critical for persons of privilege to begin to understand the experience of these disenfranchised members of their own communities and begin to take action to curb such inequities.

Modern medical culture often utilizes the AD diagnosis as a means of simultaneously destigmatizing and legitimatizing psychic distress. No other diagnostic category in DSM is so well-suited to communicating both to patients and third parties that the dysfunction in questions is the result of an understandable, if an exaggerated reaction to external events rather than an internal dysfunction, whether biologic or psychosocial. Third parties, including health and life insurance carriers, employers, and family members, often look at this diagnosis as a nonrecurring phenomenon with a generally good outcome. AD enables a clinician to use the current language of health care delivery to indicate that the patient is significantly impaired to warrant clinical attention and hence payment for services rendered. It also underscores the assumption in Western medical culture that maladaptive forms of distress necessarily represent clinical diagnostic entities and, as such, can be seen as yet another form of the medicalization of distress.

A 46-year-old second-year pediatric resident Dr. S receives a negative evaluation from her faculty supervisor and is put on academic probation pending her ability to show significant improvement in her clinical performance. Dr. S had been very successful in her medical career until she sustained a back injury in a biking accident, and one of her sons was diagnosed with bipolar disorder. These added pressures significantly impaired Dr. S's ability to function as a resident physician. After the evaluation, she was unable to get out of bed and often called in sick. Dr. S's mood was irritable and depressed, and she found herself unable to concentrate or make decisions either at work or home. When her children found her crying one afternoon, Dr. S realized she needed help and sought assistance from employee health. The therapist who assessed her indicated she felt the patient was suffering from major depression. The patient was quite concerned that this diagnosis would suggest to her program director that she was not fit for duty. She thus requested the therapist diagnose her with a less serious condition. The therapist agreed to use the DSM-5 criteria for AD in her clinical notes and for billing purposes.

EPIDEMIOLOGY

Historically, few studies examined the prevalence of AD in community samples, underscoring its marginal and ambiguous status in the outpatient diagnostic repertoire. Neither the Epidemiologic Catchment Area nor the National Comorbidity Study included AD in the survey instruments. One exception to the exclusion of AD in epidemiologic research is the European ODIN study, which was a cross-sectional two-phase community assessment using first the Beck Depression Inventory and then the SCAN, one of the few structured instruments to include AD. Less than 1% of the subjects were diagnosed with AD, and the rate varied from country to country. The discrepancy between these findings and almost all previous studies raises further questions regarding the reliability and validity of current diagnostic criteria, not just for AD but also for all the depressive disorders bordering AD. Further epidemiologic information on AD can be gleaned from an examination of specific populations exposed to stressful events. One of the largest studies looking at the prevalence of ADs in a clinical population examined patients on entry to the clinical services of the Western Psychiatric Institute. Intake involved a diagnostic assessment of over 11,000 individuals of all ages using a semistructured assessment form and DSM-III criteria sheets. Ten percent of the sample was found to have ADs, making it the second-largest diagnostic category. In children and adolescents under 18 years of age, over 16% had ADs. In adults, the female-to-male ratio was approximately two-to-one. In children, there was only a slight preponderance of females with the diagnosis.

However, more recent studies employing more sophisticated scales looking for the presence of AD have estimated a point prevalence closer to 2% of the general population. It is highly likely that the unprecedented global stressors of the last decade, including the COVID-19 pandemic, the resultant global economic collapse, a rising consciousness regarding climate change and the imminent threat to our environment, and increasing political polarization threatening democracy will lead to an increase in the prevalence in 2020 and the years following. In a population over 65, the prevalence rate was found to be 2.3% using this scale. In a German study using the new ICD-11-based ADNM, rates ranged from 9% to 1.4%, varying on the version of the scale used. Rates of AD have been found to be much higher using this scale in specific populations. In postconflict regions, the prevalence of AD using the ICD-11 concept ranged from 6% to 40%. This lends further support for the current move to include ADs in the DSM-5 category of Stress Response Syndromes.

Bachem and Casey describe various settings in which the rates of AD are much higher than the general population. While no studies have been conducted in primary care settings or consult-liaison environments using the revised ICD-11 criteria, these areas have historically seen more frequent presentations of AD. In consult-liaison services, the diagnosis of AD using DSM has historically been approximately 12%, though this number appears to be declining in favor of the diagnosis of Major Depressive Disorder in this cohort, perhaps due to the greater intensity of study of the latter. A multisite study in consultation psychiatry services in the United States, Canada, and Australia found a diagnosis of AD among 12% of all consultations, with another 11% identified as having the potential of an AD diagnosis. Other C/L service studies have described a prevalence of nearly 30%.

Decades of research have found a higher prevalence of adjustment disorder in patients in the medical setting—especially patients with cancer and those receiving palliative care. One meta-analysis of 70 studies showed a rate of 19.4% of patients with AD, compared with 16.3% with MDD, and the diagnosis of AD was nearly three times as common as MDD in the medically ill.

In a study of 107 patients with head and neck cancers, 14 (12.1%) had a diagnosis of AD. In another study of 55 women following a first recurrence of breast cancer, 19 (35%) were diagnosed with AD.

Other specific unfortunate life events put individuals at particular risk for the development of an AD. One such group is described in a Danish study examining the prevalence of adjustment disorder in women who have experienced a stillbirth—loss of a fetus anytime following 22 weeks. This large-scale study showed that women had a transient increased risk for psychiatric hospitalization within the first month following fetal death. The most likely diagnosis in this group was an AD. Another such group involves individuals who have been incarcerated.

Finally, one must consider the emergence of two dramatic global public health events, climate change, and the COVID-19 pandemic, as triggers for the development of AD on a far wider scale than has ever been seen in clinical care. Research is already suggesting that worldwide climate change is increasing rates of mental health disorders, as well as exacerbating existing conditions. Events such as uncontrolled wildfires in Australia and California with the devastation of wildlife and ecosystems, as well as human habitation, flooding, and earthquakes have left many physically displaced and facing the loss of loved ones and their health and livelihood. The destruction of farms due to droughts around the world has led to famine and disease. For many, a stable sense of self comes through connection with their environment and its predictable rhythm of the seasons. These climate disasters not only result in food scarcity: They also rob especially rural populations of the very environmental conditions that promote a sense of self and instill purpose. The literature describes presentations of PTSD, depression, suicidality, anxiety disorder, and substance use disorders related to these catastrophic changes in climate. AD is generally not screened for such studies, yet it is probable that these stressors have similarly driven up the incidence of this diagnosis. Such mental health issues may be short-lived if the stressor is resolved, or they may lead to a chronic form of adjustment disorder in the presence of ongoing environmental distress or even devolve into a more serious diagnosis. Pathways to mental health problems from climate-related events include economic, migratory, and community structure changes leading to shame, uncertainty, and potentially depression. Climate change may displace as many as 150 million people, and the maintenance of social supports may outweigh fiscal support in enabling postdisaster recovery. There is also emerging evidence that suicidality itself may be influenced by climate-related change, including linkages between suicide and air pollution, drought, and rising temperatures.

A 58-year-old male, Mr. J, presents to his local internist in Northern Canada explaining that his farm has been nonfunctional in the past 3 years due to alternating drought and floods from climate change. Mr. J is facing severe economic hardships, having difficulty paying the mortgage on the property that has been in his family for three generations. He is concerned he and his wife will need to sell the farm for whatever they can get and move to a city. Neither Mr. J nor his wife has any easily marketable skills for an urban economy. He has already had to let go of workers who have been laboring on the land beside him for years. Family traditions have been discontinued, as all of his adult children have already gone to the city to find a source of enhanced income, realizing there was no longer a future on the family farm. Mr. J feels lonely and isolated, with a sense of shame and demoralization. He has recently purchased a gun as he does not know if he can continue living this way.

Kazlauskas and Quero identify “AD as one of the most likely causes of consequences of the pandemic on mental health.” Based on early European and American data, as well as the experience of past pandemics, experts anticipate an increase in AD worldwide as a result of risks of contracting and spreading the virus, caring for those who are ill, dealing with multiple episodes of grief, loss of employment and financial security, and social impact of public health restrictions. European studies in the first months of the pandemic have found that both health and work-related stressors constituted risk factors for the development of AD. A web-based survey in Poland conducted in March-April when the country first went into lockdown found that 49% of respondents reported symptoms of AD-related to COVID-19, with women and those without a full-time job being most affected. When the ADNM-20 was administered, 14% of this sample met ICD-11 criteria. British researchers used the Internet to conduct a cross-sectional survey of over 1,200 citizens at two different times in the pandemic, March-April, and June. Probable AD was found in approximately 16% of the respondents with the diagnosis, and the study hypothesis of a dose-response effect was confirmed with those who experienced four or more COVID-19-related stressors being more likely to meet diagnostic criteria. Participants who were younger or in isolation had higher rates of AD. Health care professionals are at particularly high risk of developing AD during the pandemic due to long hours caring for very sick patients often without adequate resources; seeing intense suffering and death; conflicts between professional and personal duties; and fear of contracting the virus and exposing friends and family. Swiss researchers administered the ADNM to practitioners working in a hospital during COVID-19 and diagnosed 8% with AD, with nursing staff and women having higher rates of AD and depression. The authors emphasize this group may benefit from psychotherapeutic techniques that have benefited health care professionals in China, such as cognitive restructuring and efforts to reduce the use of substances as a coping mechanism.

These pervasive negative effects on mental health are likely to endure even once a vaccine becomes available. This, combined with shortages of overstretched mental health professionals, likely results in delayed diagnosis and treatment of AD that may well become chronic.

The magnitude of the impact of COVID-19-related ADs makes it imperative that public mental health organizations and practitioners identify and deploy effective diagnostic and treatment options. The nature of the new ADNM as a self-report measure enables it to be potentially utilized to screen large populations. Prior research with virtual treatment modalities has been promising, and there is an urgent need to expand and implement the use of telehealth and online therapies, especially as outreach to the most vulnerable populations.

One contributor to this discrepant epidemiologic data has been the lack of a structured diagnostic instrument for AD similar to those available for other psychiatric disorders, although a new instrument has been developed. The instrument is a 29-item self-assessment—the DIAD—that has been used in two sample populations of patients already diagnosed with adjustment disorder. The assessment tool identified three major characteristics of patients with AD: intrusive memories of the event, avoidance of feelings, and a failure to adapt. The authors conclude that this tool will likely facilitate future research in the area of ADs. The DIAD covers the domains of stressful life events, onset and duration of the stressor, symptoms of distress, and impairment from the stressor, along with the relationship between these and the stressor. Further studies will, of course, be needed to establish psychometric validity, but the development of the DIAD is certainly a step forward in diagnostic rigor for AD.

ETIOLOGY

Current knowledge regarding the relationship between a stressful life experience and the development of symptoms remains incomplete. Unresolved questions persist in the literature regarding the etiology of AD. Many authors in the field criticize the idea of linking a single stressor to a symptom complex. As discussed earlier, if a stressor does impact an individual, the model of linear causality has less utility than other complex models of causality. Recent literature has considered symptom formation following a stressor from a variety of paradigms. These include psychological models of processing stressful life experiences; psychodynamic models of affect attunement; the impact of early life experiences and ongoing psychiatric issues on the vulnerability to stress; and biologic vulnerability to stressful life experiences.

Israelachvili provides a compelling description from the social psychology literature of processes that may underly the idea of adjustment. He explains that adjustment often results from the experience of transition and that an individual must adapt to that transition. Transitional periods in an individual’s life might be considered an adjustment to a new set of circumstances. He compares this to the more humanistic idea of adjustment in which one might bring their own resources to a situation in the transitional process, to both change the self and perceived circumstances or environment. This is to be distinguished from the process of socialization, where the transition only involves an understanding of a new set of circumstances, such as the culture of a new place of employment or the background of a spouse's heritage. He also underscores our lack of understanding of what constitutes a good adjustment: Describing that adjustment is often considered through the presence of maladaptation to a transition, and it is through these negative markers that we define adjustment.

One investigator describes the integration of stressful life experiences as a series of responses. The first of these involves the realization that a stressful event has occurred. A period in which the information is suppressed then follows. There is then alternating intrusion of the event and attempts at suppressing it, along with working through the information, which results in an integration of the material into one's cognitive schema or in symptom formation.

Another author has elaborated a theory of associative network formation following a significant stressful experience. The network consists of three elements: stimulus information about the stressor; information about cognitive, behavioral, and physiologic reactions to the stressor; and meaning elements representing basic assumptions and their violation. Symptoms occur when one is re-exposed to an element of the associative network. An example the literature offers is how meeting a former employer may trigger thoughts about the unexpected dismissal from current employment and why this happened. Psychotherapy would be aimed at modifying this associative network.

Another model proposes that the etiology of symptoms comes not only from a representation of the event in memory but also from individual differences in the meaning and understanding of the stressor and its sequelae. If the meaning of a stressor to an individual greatly disturbs their autobiographical narrative, it will lead to more disturbing psychological consequences. In such cases, it becomes more difficult to reorganize an understanding of previous experiences to reestablish a stable sense of self.

Contemporary psychodynamic theory focuses on the context in which an event occurs and how this leads to symptom formation. The intersubjective paradigm delineates the critical nature of attunement from others following a significant stressor. A stressor leads to the generation of an affective experience, which in turn leads to the desire for understanding of this affect. If others fail to attune to this affect but rather criticize, reject, or neglect the individual, he is left to cope on his own, leading to symptom formation. Though generally supported, this idea has been questioned in some studies, which infer that excessive social support can, in fact, lead one to feel incapable of coping unaided, in turn leading to lower self-esteem and distress.

Psychodynamic theory also posits that it is the lack of an experience of affect or feelings about the stressor that generates problems. Individuals may not have the opportunity to experience their feelings about an event or toward a person because the emergence of these feelings represented a danger or threat for whatever reason. Perhaps it threatened a caregiver, perhaps there was not sufficient time, or perhaps this was already a learned style. The emergence of these feelings signals danger and a sense of anxiety. Individuals develop psychological defenses to mitigate this anxiety which in themselves can cause problems. For example, anger at an individual may be turned against the self, resulting in symptoms of depression and demoralization.

Some evidence exists that individuals with greater deprivation early in life are at greater risk for developing symptoms following a life stressor. Current literature points to the importance of childhood experiences in the vulnerability to symptom formation in response to later life stress. Evidence from a variety of sources suggests that early experience with diminished control may foster a cognitive style characterized by an increased probability of interpreting or processing subsequent events as out of one’s control. This may represent a psychological vulnerability to the development of anxiety. Research in the area of adverse childhood events further underscores the contribution of early life experiences in the generation of depressive affect later in life. For many, the experience of adversity in childhood is associated with the development of further stressful life events over the course of development and adulthood. Both early adverse childhood events, as well as these associated later adult experiences of adversity, seem to contribute partially and significantly to the development of the experience of the effect of depression.

Recent literature has furthermore examined the question of resilience to stressful life experiences. Most of this research has focused on resilience in the face of adversity during childhood and adolescence. Studies of children exposed to stressors including war, family violence, poverty, and natural disasters have revealed a consistent pattern of individual characteristics associated with successful adaptation. These include good intellectual functioning, effective self-regulation of emotions and attachment behaviors, a positive self-concept, optimism, altruism, a capacity to convert traumatic helplessness into learned helpfulness, and an active coping style in confronting a stressor.

Biologic theorists have postulated the role of numerous neurotransmitters in the generation of pathology in individuals most at risk at the time of a stressful life event. Following a review of eleven neurochemical, neuropeptide, and hormonal systems, one author postulates that changes following acute stressful experiences may lead to long-term risk factors for the future capacity to cope with stress. This “allostatic load” of neurochemical networks may relate to individuals’ resilience in the face of stressful life events. One prediction is that individuals in the highest quartile for measures of HPA axis, corticotrophin-releasing hormone (CRH), locus coeruleus-norepinephrine, dopamine, and estrogen activity, and the lowest quartile for dehydroepiandrosterone (DHEA), neuropeptide Y, galanin, testosterone, 5HT1A receptor, and benzodiazepine receptor function will have the highest index for psychobiological allostatic load and increased risk for psychopathology after exposure to stress. These neurotransmitters act through brain regions, including the amygdala, hippocampus, locus coeruleus, and prefrontal cortex, deemed critical in the modulation of a stressful experience.

There are few actual imaging studies looking at the brain changes associated with AD. However, one imaging study of 25 patients with AD found that patients with AD had decreased gray matter volume in the right medial frontal gyrus compared with healthy controls. This brain region is involved in the extinction of learned fear, and a deficit in this area may result therefore in maladaptive behaviors. It is also associated with the default mode network, which is associated with various psychiatric disorders, including depression. Evidence suggests a role for the default state network in the development of anxiety and depression after a stressful event. The involvement of this brain region suggests problematic emotional regulation following a stressful event.

DIAGNOSIS AND CLINICAL FEATURES

A precipitating experience in the form of a stressful event(s) or life situation is the sine qua non for the diagnosis of AD. It is characterized by the development of emotional or behavioral symptoms in the context of an identified psychosocial stressor(s). The symptoms vary widely in form and severity, and the person may or may not be aware of the stressor as being significant. In contrast to other DSM-5 disorders, the AD criteria have no clear and specific symptom profile that defines the condition.

It is deemed clinically significant if accompanied by marked distress “out of proportion to the severity or intensity of the stressor” or by significant impairment in social and occupational functioning. The challenge for the clinician in these situations is to differentiate a reasonable and expected response to psychosocial stressors from that which might indicate a diagnosis of AD. Thus, a reaction within the normal range of human experience must be included in the differential.

Many individuals who seek psychiatric care following a significant life event may desire help in the form of psychotherapy or psychopharmacology, yet their response may not, in fact, be outside of what would be expected. One author notes that there is no way to categorize “problems of daily living” in the current system, yet, with more reliance on mental health practitioners for a role that may have once belonged to families, community support, and religious affiliations, it is critical to understand that such reactions may not be pathologic.

AD may occur in any age group. Some studies have identified school problems as the most frequent precipitant in adolescents, while marital, financial, and health issues are more common in adults. Several studies have described the presenting features of ADs in clinical samples. Common symptoms of AD identified in a large study of AD included depressed mood, low self-esteem, suicidal behavior, increased motor activity, hypervigilance, impulsivity, and substance use. In most of these samples, depressive symptoms have been the most prominent.

Studies of ADs using structured diagnostic instruments have reported high levels of comorbidity. The degree of comorbidity may be an important determinant of the level of impairment and risk for poorer outcomes n persons with adjustment disorder. In one study of Medicare patients, 54% of patients diagnosed with AD also displayed abnormal illness behavior when using the structured interview for Diagnostic Criteria for Psychosomatic Research (DCPR). Symptoms of illness behavior documented included disease phobia, health anxiety, thanatophobia, nosophobia, and illness denial. One-third also met the criteria for somatization. Demoralization, alexithymia, and unstable mood were diagnosed in nearly 30% of patients.

Several studies have reported a significant association of AD with suicidal ideation that beliefs the general clinical impression that is it a less serious mental health condition. In one study of patients who had attempted self-harm and presented to the emergency room, those with AD had more impulsive acts, were more often associated with alcohol abuse, and attempts at self-harm occurred earlier in the course of the disorder than for those patients with a diagnosis of major depression. A 2019 literature review of AD in emergency departments and medical units found a high prevalence of AD in patients presenting with suicidal ideation and behaviors, with poisoning being the most common method of self-harm. However, there was a lack of high-quality studies indicating the need for more solid research to be conducted in this critical area of consultation psychiatry. These findings underscore the seriousness of AD in a subset of individuals; subthreshold conditions may carry high morbidity, including the risk of self-harm and death.

Subtypes

Historically, six subtypes of ADs have been described based on the predominant symptom presentations. DSM-5 continues the subtypes described in DSM-IV-TR. These do not identify specific symptom profiles for each of the subtypes; however, attempts have been made to more accurately characterize them. Notable symptoms for individual ADs subtypes are (1) depressed mood: depression, hyposomnia, low self-esteem, and suicidal indicators; (2) anxiety: generalized anxiety, increased motor activity, and situational anxiety; (3) disturbance of conduct: impulsivity, lack of insight, and violent behavior; and (4) mixed disturbance of emotions and conduct: excessive alcohol ingestion, suspiciousness, hostility, defrauding behavior, and homicidal ideation. A study in a population of primary care patients found that the anxious subtype was the most prevalent of the ADs, twice as prevalent as the depressive subtype, and far more prevalent than other forms.

There was a movement to include two additional subtypes in DSM-5. The first of these involved abnormal grief, but this was ultimately included in a separate section of the diagnostic manual. The second involved a subtype of PTSD/acute stress disorder. This was to include individuals with symptoms that do not rise to the full diagnostic criteria of either of these disorders but met many of the criteria. Both subtypes underscore the relationship between these diagnoses, as all involve a stressor and ensuing symptomatology. Ultimately, however, the PTSD/ASD subtype was not specifically included in the DSM-5 but was acknowledged through the inclusion of the category of Other Trauma- and Stressor-Related Disorders (Table 25–1).

A 45-year-old nurse Ms. D had been detailed 4 months earlier from her usual primary care assignment to the acute medical unit of an academic medical center to help manage a COVID-19 surge in her state. She presented to the center's emergency department after having ingested 10 tablets of the 5-mg zolpidem tablets her family nurse practitioner prescribed her for occasional insomnia. She has no previous psychiatric history or diagnosis. After taking the medications, she told her wife, who called an ambulance. She felt that the exigencies of the pandemic had constrained her to practice outside her scope yet also felt guilty that she was not more altruistic about being given the opportunity to assist her colleagues and care for very sick patients. Her children, 9 and 14, have been attending school online, and this has resulted in tension between her and her wife about who will monitor their activity. Ms. D is torn between visiting her aged parents and placing them at risk, and worrying about how the effects of prolonged isolation on their chronic health problems. The constant fear of contracting the virus and the even greater anxiety about exposing her wife, who is immunosuppressed, have left her unable to sleep, periodically tearful at work and at home, retreating to her bedroom rather than joining in family meals. She is not sure if she just took the pills because she wanted to not wake up for a while and forget everything or really wanted to die. During the interview, after acute medical stabilization, she was tearful at times and alternated anger at the hospital and her wife with shame about her overdose and the extra stress it placed on her family. Following crises intervention in the emergency setting, she responded well to individual psychotherapy over a 6-month period, as well as a peer-support group of other health care professionals dealing with the strains of the pandemic. The therapist helped Ms. S negotiate an assignment to a new vaccination clinic being set up outpatient, which helped her feel she was making an important contribution to the effort to combat the pandemic in a way that took advantage of her competencies. Ms. S and her wife were able, with the help of a shared therapy session, to develop a schedule for child care responsibilities that reduce the friction in their relationship. Ms. S was started on an antidepressant with the expectation that it could be tapered after 6 to 9 months, depending on the state of the pandemic. When she left the mental health care she had, she was not only in full remission but felt she had developed new resilience in the face of challenges.

Table 25–1.
DSM-5-TR Subtypes of Adjustment Disorders

With depressed mood
With anxiety
With mixed anxiety and depressed mood
With disturbance of conduct
With mixed disturbance of emotions and conduct
Unspecified

DIFFERENTIAL DIAGNOSIS

Stressful life events frequently trigger or exacerbate a wide range of psychiatric disorders. The largest number of studies investigating the differential diagnosis of AD has usually compared the depressed mood subtype to the upper threshold of major depression and the lower threshold of patients without a diagnosable illness according to DSM. This research has supported the middle or transitional position of AD in terms of symptom severity, impairment of function, course of illness, and prognosis. However, the resemblance of patients with AD to “normal” patients raises the interesting question of whether the syndrome is most appropriately understood as a psychiatric illness or as “problems of living” that become overwhelming as a result of an individual’s lack of internal resources and external supports.

The presence of a stressor in the AD diagnostic criteria places acute and post-traumatic stress in the logical differential diagnosis.

The nature of the stressor and the accompanying constellation of affective and autonomic symptoms have been far better defined and characterized for both stress disorders. In contrast, the stressor in AD can be of any severity, with a wide range of possible symptoms. When the response to an extreme stressor does not meet the acute stress or post-traumatic disorder threshold, the AD diagnosis would be appropriate, reinforcing the rationale of placing them under one diagnostic heading.

Some authors have suggested that AD within a stress context may be a *forme fruste* of PTSD for patients lacking either the severity of the stressor or the required number and type of symptoms for post-traumatic diagnosis. Acute stress disorder is even more closely allied to AD in that it also has a limited time course and often a better outcome. Identifying the factors of genetics, environment, or their combination, which account for the different developmental course of stress-related disorders is an area warranting further research. With the placement of the AD in the category of Stress Response Syndromes, this is more likely to happen in future years. The precise nature of the relationship between the stress disorders remains unclear, though there appears to be a correlation between these entities. One recent study of 544 bereaved Israelis identified a close relationship in the new ICD category of prolonged grief disorder (PGD), PTSD, and AD. The severe symptoms of PGD in patients appeared to be mediated both by the severity of the traumatic loss, serious life events, as well as by the concurrent symptoms of PTSD and AD.

The AD subtypes need to be distinguished from subthreshold types of the major mental disorders and the so-called not otherwise specified (NOS). Affective disturbance, anxiety, or conduct problems associated with an identifiable stressor render the diagnosis AD rather than the NOS categories in DSM-V. There is some evidence that these diagnoses do, in fact, represent distinct entities. In one study at the Rhode Island Hospital involving 3,400 patients diagnosed with either depression NOS or AD, important differences were identified. The patients with depressive disorder NOS were more often diagnosed with social phobia and a personality disorder.

The patients with depressive disorder NOS reported more anhedonia, increased appetite, increased sleep, and indecisiveness, whereas the patients with AD reported more weight loss, reduced appetite, and insomnia. There was no significant difference between the groups in the overall level of severity of depression or impaired functioning. The patients with depressive disorder NOS had a nonsignificantly elevated morbid risk of depression in their first-degree relatives. An additional review comparing ADs and depressive episodes in clinical practice focused on the role of personality disorders. Though the rate of personality disorders in patients with ADs is fairly high, the rate in patients with depressive disorders is substantially higher. Interestingly, if the symptoms of AD persist for over 6 months after the stressor or its consequences have resolved, the diagnosis would have to change to the appropriate NOS category, if symptoms remain subthreshold or to a more specific psychiatric diagnosis.

Consultation-liaison psychiatrists have long noted the difficulty of diagnosing a mood disorder in hospitalized or medically ill patients. The differentiation of AD from a mood disorder due to a general medical condition or demoralization is complex. While the medical condition may be the obvious stressor, affective symptoms may arise as a direct physiologic consequence of some medical conditions such as neurologic or rheumatologic conditions or the stress of prolonged hospitalization or disability. In medical and other settings, care must be taken to distinguish between the reasonable and the expected response to psychological stressors, which many authors believe includes demoralization and the inordinate response that may suggest an AD. For the clinician, it may be most useful to see the disorders along a dynamic continuum of affective stress responses where the intensity of distress and nature or degree of functional impairment are more determinative of the need for psychiatric intervention than the actual diagnosis. Normative data on stressor response patterns that differentiate pathologic from abnormal responses are not available in any rigorous fashion, and, as the case below illustrates, clinical and value judgments will continue to play a significant role in the diagnosis of ADs.

A 59-year-old unmarried man has been hospitalized for over 3 months with complications from a perforated gallbladder that required repeated surgeries. He has spent most of the time in the intensive care unit with a nasogastric tube and in severe pain. At times, the surgeons tell the patient that he can recover and return to his job as a security guard while, at other points, the surgical team informs him that the best they can hope for is to discharge the patient to a nursing home. As the patient goes into his fourth month of hospitalization, the nurses notice that for several days a week, the patient is less communicative, seems alternatively sad and then irritable, sometimes refuses to work with physical therapy, and expresses doubts he should ever have agreed to the initial surgery. The surgeon is concerned the patient is “giving up” and requests a psychiatric consultation “for depression and whether the patient would benefit from medication.” After reviewing the chart and interviewing the patient, the resident on the consultation-liaison psychiatry service tells the attending, “I would have bad days too if I had been this sick in the hospital this long.” The attending agrees and informs the surgeons that “the patient’s behaviors are within the normal range of human experience when confronted with protracted serious illness.” The consult team recommends continued clear and consistent communications with the patient about his prognosis and additional support from the patient’s care team, chaplains, and volunteers.

COURSE AND PROGNOSIS

Classic conceptualizations of AD see it as a transitional diagnostic category and thus, by definition, not as an enduring diagnosis: this is one of the reasons clinicians and patients alike have felt it to be less stigmatizing and more hopeful than a major depressive disorder or PTSD diagnoses. The proposed etiology of the disorder assumes that it would not occur without a stressor and that symptoms do not persist beyond 6 months after the stressor or its consequences have terminated. If they do persist, another psychiatric diagnosis would presumably be assigned. The symptoms, therefore, have two prognostic courses: either they resolve or progress to a more serious illness. In some subpopulations, this later course is often the case. In a study by O'Donnell et al., trauma survivors with a diagnosis of AD at 3 months were 2.67 times more likely to go on to receive a more serious diagnosis at 12 months when compared to those without a diagnosis at the 3-month mark. These later diagnoses included PTSD, major depressive disorder, and generalized anxiety disorder. Additionally, over 34% of those in the study continued to meet the criteria for AD at the 1-year time period. This study, among others, challenges the idea of AD as a minor diagnosis, demonstrating that in certain populations, symptoms may well persist or even evolve into a more specific diagnostic entity. An additional study with children and adolescents showed that up to one-third had symptoms of AD which persisted over 6 months, and one-fourth still remained symptomatic after a year. Furthermore, the youth had significant morbidity and poor outcome in contrast to the benign profile of AD found in another research. The majority of studies in diverse settings do support the understanding that AD generally has a shorter duration of illness and hospitalization of patients with AD compared to other Axis I conditions.

Attempts have been made by a number of investigators to address the question of whether an initial episode of adjustment disorder predicts the subsequent development of more specific and severe pathology. Multiple studies have shown a fairly benign course for adults, with a lower likelihood of recovery in adolescents. A 5-year follow-up study at the University of Iowa showed a recovery rate of 71% in adults versus 44% in adolescents. An additional 8% of adults were well at follow-up but had intervening problems, compared to 13% of the adolescent group. Almost 70% of the adolescent group was ill during the 5-year follow period, while less than 30% of the adult group experienced psychiatric disorders in the study period. While most of the adults developed major depressive disorder and alcohol use disorder, the adolescents developed a wider range of major psychiatric disorders, including schizophrenia, bipolar disorder, antisocial personality disorder, substance use disorder, and major depression. The chronicity of symptoms and behavioral disturbance were the best predictor of a poor outcome. The specific adjustment disorder alone did not predict the follow-up diagnosis. In studies following disasters such as Hurricane Katrina in the Gulf, specific factors have been found to decrease the risk of symptoms. These included excellent support systems, high income, and education level, successful mastery of past traumatic events, timely information, availability of recovery services, compassionate understanding from health care providers, and good follow-up.

Research has disclosed a risk for suicide, especially in adolescent patients with AD, not previously fully appreciated. A recent study of 119 patients with AD indicated that 60% had documented suicide attempts in the past, and 96% had been suicidal during their admission to the hospital. Fifty percent had attempted suicide immediately prior to their hospital admission. Comorbid diagnoses of substance use disorder and personality disorder contributed to the suicide risk profile.

Psychological autopsies of 19 adolescent suicides found that one-fifth of the victims had an AD, but over half had a personality disorder as well. The patients who completed suicide were dealing with multiple psychosocial pressures and stressful life events, and few were receiving treatment, suggesting that inadequate coping and lack of professional treatment may have played a role in the deaths. Autopsy results also discovered that the suicidal process was much shorter and more rapid in its evolution in adolescents with AD than in those with other diagnoses and that most of the victims did not have a history of emotional disturbance. Together, these and other studies serve as a cautionary tale to clinicians to reconsider the conceptualization of AD as one of significant suicidal risks deserving of careful monitoring of symptoms and prompt intervention.

A 17-year-old high school senior experienced rejection in his first serious relationship. In the weeks after the end of the relationship, he began to exhibit dysphoric mood accompanied by anxiety and psychomotor agitation. His stress was exacerbated by his inability to play sports and socialize with his friends due to the restrictions of the current COVID-19 pandemic. His parents both had pre-existing medical conditions and did not allow him to socialize with anyone outside of his own family–themselves and his two much younger brothers. He had received counseling in junior high school when his parents considered divorce, and in high school, he began using alcohol and marijuana, contributing to a suspension in his sophomore year for fighting. A month after his breakup, he began to tell his parents that life was no longer worth living without his former girlfriend and that he could not continue with this high degree of social isolation. Two months later, his parents came home from a long walk and found him hanging in the garage with a note stating he could not go on alone.

TREATMENT

The AD diagnosis is common in many specific populations, though given its status as a potential minor and self-limited diagnosis, the clinician is often faced with the question as to whether there is sufficient psychiatric morbidity to warrant treatment intervention. The treatment of individuals with AD requires the careful assessment of the nature and severity of symptoms with consideration of the risk factors associated with poor outcomes, such as poor premorbid functioning and persistent stressors, as well as the risks of treatment, particularly with pharmacotherapy. It is essential to understand the meaning of the stressor(s) to the patient and why it is associated with the presenting symptoms. A paucity of solid formal studies confounds and complicates clinical decision-making. The first formal systematic review of treatment strategies was not published until 2018–this despite its clinical use for decades. This study, however, ranked most all interventions as “low-” or “very low-” quality based on the lack of formal assessment tools for AD, the use of very small sample sizes, the lack of clinician assessments prior to the intervention, the lack of clinician-administered assessments, the lack of follow-up measures, and the lack of control for ongoing medication use.

The primary goals of treatment for AD, as with most mental health conditions, are to relieve symptoms and enable the patient to obtain a level of adaptive functioning that is comparable or, in some situations, better than the level of premorbid functioning. It is important to understand and facilitate those factors that mitigate the pathologic response to stress, as well as taking into account the patient’s level of vulnerability, personal strengths, and history of adaptation to challenges and capacity for resilience. Treatment interventions should be designed to minimize the impact of the stressors on day-to-day function and to mobilize adaptive stress-coping mechanisms.

Specific psychotherapeutic treatment interventions that have been considered in individuals with AD include supportive psychological approaches, behavioral and cognitive-behavioral approaches, problem-solving techniques, relaxation techniques, and psychodynamic interventions. Short-term treatment may be adequate for many patients; however, more extended treatment may be appropriate in situations where individual characteristics predispose the patient to stress intolerance. Depending on the level of acuity and existing supports, crisis intervention and case management may be necessary on a short-term basis to preclude rapid progression, suicide attempts, and perhaps avoid hospitalization. Individual psychotherapy tends to be supportive in nature rather than exploratory, though emerging literature challenges this assumption. It is important to help the patient to better understand the meaning of the stressor and why it may have overwhelmed their coping mechanisms.

A study demonstrated large effect sizes in terms of therapeutic change in a group of patients primarily diagnosed with AD who participated in psychotherapy. Though the proposed mechanism of change involved the development of reliance on more mature defense mechanisms, only a moderate change in defensive functioning was observed. The authors suggested that longer-term psychodynamic psychotherapy or psychoanalysis might be helpful in these responsive patients to further develop a less defensive style. Another study shows promising results for the use of meditation training for patients with both anxious and depressive subtypes of ADs. One group has demonstrated the use of an online virtual reality cognitive–behavioral therapy for use at home as homework assignments for patients with ADs. Interestingly, in a Cochrane meta-analysis, only problem-solving therapies showed a reduction in the time for patients with the diagnosis of AD to return to part- or full-time work. The preliminary findings of internet-based self-help intervention for ICD-11 AD showed promise as well, with a moderate effect size demonstrated for a population of 284 study patients with AD diagnoses using the BADI formal instrument, though the dropout rate in the study was high. Participants in this study reported an experience of an average of four life stressors. Among the most common were too much or not enough work, conflicts in the family, financial difficulties, difficulties due to deadlines, and conflicts at work. Another ongoing study demonstrates the use of multiple modalities tailored for a group of cancer patients. This includes a module of psychoeducation, followed by two modules of group interventions, mindfulness, and eHealth.

In the model, Israelachvili proposed, the therapist might consider the situation of the patient more specifically. If the issue involves a required adaptation, then such models as Acceptance and Commitment Therapy (ACT) or Morita Therapy might be most helpful in adapting to new circumstances. However, in the case of adjusting to new transitional experiences, humanistic client-focused approaches might be most useful. In cases of socialization to new circumstances, either cognitive behavioral therapy or developmental therapies might be more useful, as these promote skills development.

Family, couples, and group therapies may have utility depending on the presenting circumstances. Many individual modalities may well have utility in a group. One study demonstrated good efficacy in the capacity to regulate emotions as a result of group therapy with multiple modalities for women recently diagnosed with breast cancer.

The Dutch have actually developed practice guidelines for managing ADs in occupational and primary health care which aim to reduce suffering, sick leave, disability, and economic losses. The guideline recommendations are based on cognitive–behavioral therapy, stress inoculation training, and graded activity with the goal of improving the problem-solving ability of workers. Recommendations have also been made for the management of patients with AD who are actively deployed. These recommendations can be generally taken as minimizing the stressor, teaching psychological skills, and working with the meaning of the stressor with ongoing suicide risk assessment. Medications should be used for symptoms of insomnia and anxiety in the short run, and antidepressants for more moderate-to-severe forms of depression.

The last 5 years have seen a bit more attention to pharmacotherapy for AD, often with complementary and alternative agents. Although there have been few methodologically sound primary studies, several well-done reviews of the small body of literature on AD psychopharmacology have been published in 2018. Stein’s systematic review of the pharmacology literature could only identify 11 randomized clinical trials. Ironically, one reason for the paucity of rigorous research is the impression that AD is a minor and mental health problem that usually remits without treatment. Thus, pharmacologic decision-making must balance the potential side effects of medications with the considerable symptom burden and risk of suicide that characterize AD. The optimal approach will prioritize patient preferences and medical judgment with contextual features of the specific AD case presentation.

An RCT of the plant extract Euphytose (EUP) involving 182 French outpatients with AD showed a decrease in anxiety for the treatment group. Patients received two tablets of EUP for 28 days. Fifty-seven percent of patients treated with EUP compared to 39%

treated with placebo were considered responders post-treatment with a HAM-A score less than 14. However, the study lacked a specific scale for AD. A randomized placebo-controlled trial of Gingko biloba in patients with generalized anxiety disorder, the anxious subtype of AD, and cognitive decline for 4 weeks found that anxiety, aggression, and global clinical impression scales all showed statistically significant benefits.

A naturalistic study favored antidepressants as first-line pharmacotherapy for AD. Fifty-two adjustment disorder patients treated over 56 days with trazodone (Desyrel) found that nearly 65% obtained remission, and another 26% responded to treatment. The medication was well-tolerated, and its efficacy was due to the amelioration of insomnia, anxiety, and depressed mood. Obviously, the results of these small and early trials require rigorous replication, but they do offer some pharmacologic options to complement the primarily psychosocial treatment of the disorder.

For the anxiety subtype of AD, prospective comparison studies with etifoxine, benzodiazepines, and buspirone have shown the drug is better tolerated and more efficacious than either of the other two agents. Etifoxine is a nonbenzodiazepine drug that modulates GABA-A receptors available in Europe. A 2006 protocol compared lorazepam and etifoxine in 191 patients with DSM-IV AD over 4 weeks. On the Hamilton Anxiety Rating Scale used as the primary outcome measure, etifoxine was not inferior to lorazepam, and the novel agent was better tolerated and had a higher number of responders, as well as fewer patients with rebound anxiety. Another comparator study that used alprazolam had similar results but also found that after medication discontinuation, HAM-A ratings in the etifoxine group continued to decrease while those in the alprazolam group increased and also had more central nervous system adverse events.

There is plenty of less-rigorous evidence from psychosomatic medicine and clinic settings suggesting it is reasonable to consider the judicious use of medications to treat the specific symptoms of ADs. Typically, these are antidepressants and anxiolytics. The growing literature on the adverse effects of benzodiazepines, especially in the older adults and those with a substance use disorder, require reconsidering their use even in consultation-liaison and emergency department settings. Several studies have shown that even short-term use of benzodiazepines for AD can result in withdrawal. It is often helpful and important to treat such issues as insomnia and severe anxiety with pharmacologic interventions in the short run. The treatment of depressive symptoms with pharmacotherapy is less clear. ADs represent the greatest growth area in the prescription of antidepressants, yet the evidence for the utility in improving depressive symptoms in ADs is scant. Selective serotonin reuptake inhibitors have been found to be useful in treating some subthreshold depressive symptoms and may benefit certain subtypes of ADs. It should, however, be emphasized that psychosocial strategies remain the mainstay of treatment, with pharmacologic intervention being a complementary treatment modality. There is a great need for more robust research into the psychopharmacology of AD, with particular attention to the identification of novel neurobiologic strategies. Research that adopts a trauma-stress neuroscience paradigm may prove to be the most promising pathway to identify useful agents for AD.

FUTURE DIRECTIONS

Creating an interactive model that takes into account both stress and resiliency factors that are responsive to individual and cultural differences remains challenging. Questions abound regarding AD. For instance, what is the relationship between AD and other stress-related disorders such as PTSD and acute stress disorder? Are subtypes of ADs more appropriately conceptualized as minor disorders of mood, anxiety, and behavior?

Some authors have suggested a better understanding of the AD might require a radical choice. Either the diagnosis of AD is not based on a definition that lends itself to rigorous investigation—thus, it should be reconsidered—or its pragmatic utility, despite methodologic problems, needs to be recognized as a reason for further research, particularly study of its epidemiologic and therapeutic aspects. The latter option would require continued efforts to improve the specificity, reliability, and validity of the diagnostic category. The placement of the ADs in the Stress Response section of DSM represents the first step in this direction. Further research must also be directed toward developing optimal treatment approaches using randomized controlled trials and rigorous cost-benefit analyses.

The onset of unparalleled planetary crises in the form of climate change, the COVID-19 pandemic, political violence, and widening social injustice will likely result in an increase in AD. Patients, families, communities, and professionals would all benefit from a scholarly strengthening of this clinically valuable diagnostic category so that a more robust understanding of AD may improve prognostication and, most importantly, the discovery of beneficial treatments.

FURTHER READINGS

Arbus C, Hergueta T, Duburcq A, et al. Adjustment disorder with anxiety in old age: comparing prevalence and clinical management in primary care and mental health care. *Eur Psychiatry*. 2014;29(4):233–238.

Bachem R, Casey P. Adjustment disorder: a diagnosis whose time has come. *J Affect Disord*. 2018;227:243–253.

Ben Ezra M, Hou WK, Goodwin R. Investigating the relationship between COVID-19-related and distress and ICD-11 adjustment disorder: two cross-sectional studies. *BJPsych Open*. 2020;7(1):e21.

Caneo C, Marston L, Bellon JA, King M. Examining the relationship between physical illness and depression: is there a difference between inflammatory and non inflammatory diseases? A cohort study. *Gen Hosp Psychiatry*. 2016;43:71–77.

Casey P. Adjustment disorders: new developments. *Curr Psychiatry Rep*. 2014;16(6):451.

Cornelius LR, Brouwer S, de Boer MR, Groothoff JW, van der Klink JJL. Development and validation of the Diagnostic Interview Adjustment Disorder (DIAD). *Int J Methods Psychiatr Res*. 2014;23(2):192–207.

Eimontas J, Rimsaite Z, Gegieckaite G, Zelviene P, Kazlauskas E. Internet-based self-help intervention for ICD-11 adjustment disorder: preliminary findings. *Psychiatr Q*. 2018;89(2):451–460.

Esterwood E, Saeed SA. Past epidemics, national disasters, COVID-19, and mental health: learning from history as we deal with the present and prepare for the future. *Psychiatr Q*. 2020;91:1121–1133.

Fegan J, Doherty AM. Adjustment disorder and suicidal behaviors presenting in the general medical setting: a systematic review. *Int J Environ Res Public Health*. 2019;16(16):2967.

Glaesmer H, Romppel M, Brähler E, Hinz A, Maercker A. Adjustment disorder as proposed for ICD-11: dimensionality and symptom differentiation. *Psychiatry Res*. 2015;229(3):940–948.

Hrabok M, Delorme A, Agyapong VIO. Threats to mental health and well-being associated with climate change. *J Anxiety Disord*. 2020;76:102295.

Jacobson JC, Maytal G, Stern TA. Demoralization in medical practice. *Prim Care Companion J Clin Psychiatry*. 2007;9(2):139–143.

Kazlauskas E, Quero S. Adjustment and coronavirus: how to prepare for COVID-19 pandemic-related adjustment disorder worldwide? *Psychol Trauma*. 2020;12(S1):S22–S24.

Kazlauskas E, Zelviene P, Lorenz L, Quero S, Maercker A. A scoping review of ICD-11 adjustment disorder research. *Eur J Psychotraumatol*. 2018;8(sup7):1421819.

Killikelly C, Lorenz L, Bauer S, Mahat-Shamir M, Ben-Ezra M, Maercker A. Prolonged grief disorder: its co-occurrence with adjustment disorder and post-traumatic stress disorder in a bereaved Israeli general-population sample. *J Affect Disord*. 2019;249:307–314.

Kramer U, Despland JN, Michel L, Drapeau M, de Roten Y. Change in defense mechanisms and coping over the course of short-term dynamic psychotherapy for adjustment disorder. *J Clin Psychol*. 2010;66(12):1232–1241.

Lorenz L, Bachem RC, Maercker A. The adjustment disorder–new module 20 as screening instrument: cluster analysis and cut-off values. *Int J Occup Environ Med*. 2016;7(4):215–220.

Maercker A, Lorenz L. Adjustment disorder diagnosis: improving clinical utility. *World J Biol Psychiatry*. 2018;19(sup1):S3–S13.

Myung W, Na KS, Ham BJ, Oh SJ, Ahn HW, Jung HY. Decreased medial frontal gyrus in patients with adjustment disorder. *J Affect Disord*. 2016;191:36–40.

O’Donnell ML, Agathos JA, Metcalf O, Gibson K, Lau W. Adjustment disorder: current developments and future directions. *Int J Environ Res Public Health*. 2019;16(14):2537.

O’Donnell ML, Alkemade N, Creamer M, et al. A longitudinal study of adjustment disorder after trauma exposure. *Am J Psychiatry*. 2016;173(12):1231–1238.

O’Donnell ML, Metcalf O, Watson L, Phelps A, Varker T. A systematic review of psychological and pharmacological treatments for adjustment disorder in adults. *J Trauma Stress*. 2018;31(3):321–331.

Pelkonen M, Marttunen M, Henriksson M, Lonnqvist J. Suicidality in adjustment disorder: clinical characteristics of adolescent outpatients. *Eur Child Adolesc*

Psychiatry. 2005;14:174–180.

Perkonigg A, Lorenz L, Maercker A. Prevalence and correlates of ICD-11 adjustment disorder: findings from the Zurich Adjustment Disorder Study. *Int J Clin Health Psychol*. 2018;18(3):209–217.

Polyakova I, Knobler HY, Ambrumova A, Lerner V. Characteristics of suicidal attempts in major depression versus adjustment reactions. *J Affect Disord*. 1998;47:159–167.

Skruibis P, Eimontas J, Dovydaitiene M, Mazulyte E, Zelviene P, Kazlauskas E. Internet based modular program BADI for adjustment disorder: protocol of a randomized controlled trial. *BMC Psychiatry*. 2016;16:264.

Starkman MN. The terrorist attacks of September 11, 2001 as psychological toxin: increase in suicide attempts. *J Nerv Ment Dis*. 2006;194:547–550.

Stein DJ. Pharmacotherapy of adjustment disorder: a review. *World J Biol Psychiatry*. 2018;19:S46–S52.

Strain JJ, Friedman MJ. Considering adjustment disorders as stress response syndromes for DSM-5. *Depress Anxiety*. 2011;28:818–823.

van Beek FE, Wijnhoven LMA, Jansen F, et al. Prevalence of adjustment disorder among cancer patients, and the reach, effectiveness, cost-utility and budget impact of tailored psychological treatment: study protocol of a randomized controlled trial. *BMC Psychol*. 2019;7(1):89.

26

Personality Disorders

C. ROBERT CLONINGER, M.D., PH.D., AND NIGEL LESTER, M.D.

Progress in understanding the psychobiology of human personality has continued during the past 5 years in ways that now reveal its complex multi-component structure, dynamic functioning, and even the evolutionary changes that account for the capacities for self-awareness and intersubjectivity that distinguish modern humans from other animals. Specifically, modern humans have a unique capacity to integrate self-control and self-awareness so that they can self-regulate their habits, emotions, goals, and values coherently, which promotes their health and well-being through self-actualization. Furthermore, the intersubjective awareness of *Homo sapiens* (literally, wise man) allows them to palpably feel their connection with other people and nature, which allows them to distinguish between self and other while consciously identifying with other people and nature. Together these capacities allow a person to develop in their health and well-being by cultivating self-awareness and intersubjectivity with flexibility, creativity, and prosocial values. The genetic basis of the capacity for self-actualization and self-transcendence is not present in the genomes of other hominoids, such as chimpanzees or Neanderthals, even though we share nearly all the same protein-coding genes with them.

The genetic basis for modern human awareness depends on RNA genes that regulate the expression of protein-coding genes in three distinct systems of learning and memory: habits, intentionality, and self-awareness. Nearly all the heritability of human personality is explained by 972 genes, which are organized in these three nearly disjoint networks. A comparison of the genomes of modern humans, Neanderthals, and chimpanzees revealed that 267 genes of these genes for personality are found only in *Homo sapiens.* Most of the genes found only in modern humans are active in the self-awareness network and under positive selection. Ninety-five percent of these 267 genes were RNA regulatory genes, mostly long RNA genes that do not code for proteins, and that, instead regulate the expression of protein-coding genes. The RNA regulatory genes also organize the co-expression of other genes in functional clusters, including their overexpression in brain regions that specify the components of the brain network for self-awareness previously identified by functional brain imaging.

The three genotypic systems for human personality evolved in a staggered manner. The most primitive arose in monkeys and apes about 40 million years ago and is responsible for emotional reactivity; that is, it regulates impulses, learning habits, social attachment, and conflict resolution. Less than 2 million years ago, the second network emerged in a common ancestor of Neanderthals and modern humans, enabling the regulation of intentional self-control, that is, self-direction and cooperation for mutual benefit. Finally, about 100,000 years ago, the network of creative self-awareness emerged in behaviorally modern humans with a capacity for narrative art, science, and spirituality.

These new findings uncover the basic mechanisms required for promoting healthy personality development and, remarkably, physical, mental, and social well-being. It turns out that a healthy personality is actually the foundation for all aspects of well-being in general because it provides the regulatory processes by which we learn to integrate our habits, goals, and values so that they are congruent, reasonable, and beneficial for ourselves, others, and the world in which we live. The authors will briefly summarize a few major findings in order to orient readers before going into more depth about personality and its disorders. In this way, readers can update what they already know or assume to be true with new knowledge that will provide a more comprehensive and empirically sound approach to the assessment and treatment of personality disorder (PD).

THREE INTERACTIVE SYSTEMS OF LEARNING AND ADAPTABILITY

In addition to general intelligence, modern humans have three distinct systems of learning: behavioral conditioning, intentionality, and self-awareness. These are partially dissociable but normally interact to allow people to adapt to changing internal and external conditions in ways that are realistic and health-promoting. Integration of these three systems is essential for the development of human personality and for human health and well-being.

Consequently, the strengths of earlier theories and treatment methods that address only one or two of these three learning processes can now be incorporated into a more comprehensive biopsychosocial paradigm, thereby enabling more consistent and complete benefits for a broader range of people. Our new knowledge provides a solid foundation from which to deepen our understanding of the mechanisms of development and change that are needed for an effective and positive way to promote well-being and treat PDs in ways that are appropriate for the challenges of modern times rather than just managing them interminably.

FLUIDITY AND CREATIVITY IN THE DYNAMICS OF PERSONALITY DEVELOPMENT

The underlying mechanisms of human personality are systems of learning, so personality is not fixed or determined at birth. When we are healthy, we are constantly learning and creating our health in each moment. This fluidity and creativity allow us to both shape and adapt to changing internal and external conditions. The emergence of self-awareness in modern humans brought with it cognitive skills for vivid recollection of the past, assessment of the present, and imaginative anticipation of alternative futures. Self-awareness also provides humans with the capacity to develop social norms and cultural values to maintain large communities in which people can normally cooperate respectfully with one another and, optimally, identify with one another and the world with a harmonizing sense of interdependence and intersubjective awareness. For example, shared reciprocal interactions between a caretaker and child allow each to experience their having an impact on the experiences of the other. Through such intersubjective relationships, children develop trust and self-respect when they feel loved and loveable, valued and valuable. Unfortunately, some conditions (e.g., abuse and neglect of children, and, more generally, fear, violence, and social injustice in society) impair the development of self-awareness, impairing progress in self-actualization, intersubjectivity, and self-transcendence. As a consequence of such impairments, both individual and cultural perspectives about the goals and values that guide the way people live can become maladaptive and unsustainable. Fortunately, adolescence brings with it changes in personality that stimulates reassessment of cultural traditions, thereby facilitating a creative process of identity formation and re-evaluation of a society’s goals and values that may no longer be sustainable or beneficial.

For example, the World Values Survey and other prospective studies have identified three clusters of people whose values correspond to the dominance of one of the three systems of human learning: (i) “traditionals” who learn mostly by behavioral conditioning and are highly emotionally reactive and resistant to changing their habits and customs, (ii) “materialists” who rely mainly on self-control and are highly self-interested but may cooperate for mutual benefit, and (iii) “cultural creatives” who are highly aware of themselves and their connections with other people and nature so that they are highly egalitarian, self-transcendent, and altruistic. Younger birth cohorts are shifting toward a majority being “cultural creatives” so that they are much more egalitarian, less materialistic, and more self-transcendent than their predecessors. Cultural creatives are more socially and ecologically aware, and so they are concerned about climate change, excessive militarism, and the focus on competition and profit that leads to social inequity and frequent ill health in developed countries, including mental distress and PD in addition to chronic physical diseases. To the detriment of individual and collective well-being, such shifts in paradigm are often opposed by older adult authorities in order to conserve their traditions and authority by delaying changes that the more self-transcendent members of society perceive to be needed for a healthier and sustainable society. For example, in materialistic societies, the expression of self-transcendence is suppressed by social norms during young adulthood, but this is not sustainable as people inevitably must face ultimate situations that they cannot control, such as unpredictable climatic change, suffering, and death. Empirically, self-transcendence is a crucial component of people's ability to live in harmony with one another and the natural world.

IMPORTANCE OF SELF-TRANSCENDENCE FOR PERSONALITY DEVELOPMENT AND HEALTH PROMOTION

People and their cultures have a high frequency of personality dysfunction and related impairment in physical, mental, and social well-being unless they have clear insight and understanding of how humans can integrate their three systems of learning and memory. In brief, the path to well-being is a path of increasing awareness involving increasing character strengths for Self-directedness (i.e., intrapersonal character strengths, such as being resourceful and responsible), Cooperativeness (i.e., interpersonal character strengths, such as being tolerant and helpful), and Self-transcendence (i.e., transpersonal character strengths, such as being altruistic and contemplative), which guide and direct the regulation of emotional desires and attachments, prioritize goals and brings habits and goals in congruence with values. These three strengths of character create a dynamic cycle in which self-transcendent values inspire creative thinking, and plasticity allows its free expression. The same process produces personal well-being through self-actualization and collective well-being through intersubjective (transpersonal) identification with others and the natural world. Empirically, in prospective studies, the initial personality characteristic that leads to the most growth and maturation in personality is high Self-transcendence, but there are substantial additional contributions from high Persistence and Novelty Seeking. Self-Transcendence and Persistence are also positively correlated with nearly all character strengths and virtues. Large-scale randomized controlled trials show significant benefits from activities that promote the development of each of the three dimensions of character, as well as substantial synergistic effects from their joint activation that help to moderate emotional distress.

There is much that we can learn from the strengths and limitations of earlier theories and treatment methods for cultivating Self-directedness and Cooperativeness, which have long been recognized as important contributors to character and health. However, nearly all prior models of human personality have been materialistic (e.g., behaviorist, somatic, and pharmacologic approaches) or dualistic (mind–body approaches assuming that people are separate from one another and the rest of nature). As a result, prior models, including DSM-5 and ICD-11, have inadequately recognized the importance of the fact that high Self-transcendence is the major predictor of healthy personality development, as is true even in highly materialistic societies because it is a fundamental characteristic of human nature.

Self-transcendence is the process by which a person becomes aware of their participation in something greater than their individual self. The palpably felt experience of identification with someone or something greater than the individual self can transform a person's perspective incrementally from a sense of separation to intersubjectivity and eventually to participate in the unity of all beings. More specifically, the conscious realization of participation in something beyond yourself, such as union with humanity, nature, or God, evokes positive emotions, particularly joy and awe, which are often combined with attitudes of humility and reverence. Prospective studies indicate that the cultivation of self-transcendence and well-being involves the dynamic interplay of three processes: the awakening of plasticity (i.e., being able and willing to change), virtue (i.e., having intuitive insight into what is good for a person's self and others), and creative functioning (i.e., being innovative, purposeful and responsible so that our habits are congruent with our goals and values).

The cultivation of self-transcendence by these three processes describes the essential features of healthy personality development, which leads to a life that is healthy, happy, and good. Even when a role for self-transcendence in developing a healthy personality has been partially recognized, emphasis was initially placed on the conditioning of habits by behaviorists, or intellectual analysis of dysfunctional thoughts and emotions by cognitive therapists, or struggle with unconscious irrational forces by psychoanalysts. There was little consideration of the need to awaken a more self-transcendent perspective until the development of "third-wave" psychotherapies and positive approaches to psychiatry and psychology. The addition of tools for cultivating self-transcendence through meditative practices, nature experiences, and virtuous acts have reduced drop-out from cognitive-behavioral and psychodynamic therapies and improved outcomes.

It is to be expected that people reared in materialistic cultures are likely to be resistant to the evidence that self-transcendence is essential for healthy living. Nevertheless, most people feel empty and discontented until they directly experience the interdependence of all things as a result of the intrinsic human need for self-transcendence. The interconnectedness and inseparability of living things is not simply a matter of opinion or belief. An increasingly large body of evidence from research in biology and ecologic systems has shown that symbiosis and cooperation are more involved in the healthy functioning of ecosystems than the competition. Research on the microbiome in humans and in soils has proven that the deep interconnections within the biosphere are necessary for the body and communities of organisms to function properly. These interactions with the microbiome function at the level of genes and epigenetics, and the microbiome seems capable of interacting with the human nervous system and endocrine system. The cells of living organisms are only partially separated from one another and the world by the semi-permeable boundaries of cells that are actively engaged in the exchange of materials in and out of themselves. More deeply, according to quantum physics, quanta are constantly arising from, and returning to, a universal field of energy and information, which is called the Higgs field.

Human experiences of transcendence also speak to the interconnectedness of living things. It is natural for people to feel their connection with other people and the world. Individuals who have had transcendent experiences, including many physicists, biologists, and cognitive scientists, frequently regard the belief that we are separate from other aspects of the world as illusory. They would argue that this illusion is caused by self-preoccupation and inadequate development of an allocentric or transpersonal perspective, as is seen in various forms of psychopathology, such as autism or antisocial PD. In reality, they suggest that all living organisms are actively engaged in relationships with their surrounding world in which we are embodied, embedded, enactive, and extended. According to Albert Einstein, for example,

> “A human being is a part of the whole, called by us “Universe,” a part limited in time and space. He experiences himself, his thoughts and feelings as something separated from the rest – a kind of optical delusion of his consciousness. This delusion is a kind of prison for us, restricting us to our personal desires and to affection for a few persons nearest to us. Our task must be to free ourselves from this prison by widening our circle of compassion to embrace all living creatures and the whole of nature in its beauty.” (Einstein A. Letter to Norman Salit. March 4, 1956. Princeton University Press.)

In fact, more than three out of every four persons experience moments of self-transcendence at least occasionally. The capacity for adopting allocentric and transpersonal perspectives develops in human children around age 7 with subsequent development of ethical values and virtues such as kindness and fairness, which are based on awareness of connections beyond our individual self. In materialistic societies, however, the level of self-transcendent traits decreases during early adulthood when people are encouraged to focus on work and marriage, and then rises again as people must face ultimate situations that they cannot control. Prospective studies indicate that the proportion of people who are highly self-transcendent (variously described as post-materialists, cultural creatives, or creative characters) is increasing from approximately one-third in older birth cohorts to nearly half of younger birth cohorts of the general population, with higher rates in egalitarian countries and lower rates where self-interested materialists dominate social institutions and limit freedom of expression.

Despite how common intersubjective and self-transcendent experiences are in the general population, the scientific investigation of these human phenomena has always been controversial outside of domains like Anthropology and Religious Studies. Scientists and philosophers who identify as agnostic or atheist often consider self-transcendence as an outdated perspective from a bygone era or a relic of our evolutionary history related to the human brain meaning-making functions looking for meaning where there is none (see, e.g., Richard Dawkins, Daniel Dennett). However, all individuals experience some type of intersubjective and transcendent experience, even if that experience is not related to the divine. Many quotidian experiences like ethics, art, culture, beauty, nature experiences, and the search for truth and social justice are experiences of intersubjective and self-transcendent nature. Consequently, it is self-defeating to ignore them when considering improvements in child development, education, health promotion, or therapeutics at any age. Research has demonstrated that nature experiences lead to strong feelings of joy and connection and improve physical and mental health. Similar to nature experiences, experiences of mindfulness are being used the world over to help people experience a bit of peace and self-transcendence in the workday or in our daily lives during stress or isolation, which was prevalent during the coronavirus pandemic.

These studies and experiences suggest that we are not only healthiest but also most realistic when we are aware of our extended relationships and act in ways that reflect the deeper reality of our interconnectedness with all things. It is illusory to regard other people and nature as mere objects from which we are separate. In reality, we are interconnected aspects of a greater whole, and we harm and disturb ourselves when we abuse or neglect others. Given the contemporary threats to human survival from our disharmony with one another and the planet that we share, the time has come for psychiatry to contribute all that it can to promote the development of healthy personalities and the transformation of the resources and values of our cultural institutions that are needed to provide the milieu to support healthy personality development.

Why Has the Importance of Self-Transcendence Been Neglected So Often?

In view of the vast evidence for the importance of Self-transcendence for healthy personality development and for well-being generally, it is useful to consider why its role has been neglected so often during the past century. The importance of self-awareness, and particularly self-transcendence, is actually prominent in wisdom literature that extends back to the earliest descriptions of the path to well-being in antiquity by Homer, Plato, and other positive philosophers, as summarized by Cloninger (2004). Even at the beginning of the 20th century, there was substantial scientific and popular interest in investigating intersubjective and self-transcendent phenomena, such as the benefits of virtue or the possibility of extrasensory perception, that could not be explained by materialistic or nominalist theories of reality. However, fraud was detected often enough that most scientists became reluctant to attempt rigorous research to test the assumptions of local realism until the emergence of theories of relativity and quantum mechanics in physics.

Likewise, psychology and psychiatry were handicapped by the absence of techniques to uncover the underlying processes of motivation and learning in humans empirically until the tools of functional brain imaging and genetics were developed. Much research in psychophysiology was conducted on experimental animals who do not possess human self-awareness. Psychiatrists, such as Freud, recognized the inadequacy of tools for psychobiological research but speculated about the underlying structure of human motivation and learning using imagination and astute observations of their patients. Of course, their speculations were biased by their own uncertain assumptions about the nature of reality.

The recent identification of the genetic and neural basis for the three complex systems of learning and memory in humans reveals that the clinical assessment of Self-transcendence is also vulnerable to a serious error that had been made by early personality psychologists and that persists in the minds of most of their successors to this day. For many decades most psychologists have assumed that personality is best measured in terms of quantitative dimensions, such as Eysenck’s measures of Neuroticism, Extraversion, and Psychoticism. They also assumed that the genetics of apparently uncorrelated dimensions could be studied independently of one another. However, these dimensions (“factors”) were artificially created by the uncertain assumptions of restrictive statistical models, including that the factors they extracted were homogeneous and largely independent. During the last few years, the authors were surprised to observe that the genetic basis for personality was related to configurations of multiple personality traits rather than to the individual traits (such as Self-directedness) that emerge in particular person–situation interactions as a result of heterogeneous processes in different people. Put another way, the natural measurement units for the genetics of personality are multi-trait configurations, not individual dimensions. Dimensions, such as Self-directedness, Cooperativeness, and Self-transcendence, are useful to measure because they describe person–situation interactions, but they are actually composite effects of multiple underlying adaptive networks specified by different genes and brain circuitry in different people. There are many molecular and brain pathways to behaviors that look similar superficially or that overlap partially.

The assessment of human character traits illustrates the difficulty of understanding human personality structure and development prior to the availability of modern methods for genotyping, brain imaging, and unrestricted analysis of complex adaptive systems using artificial intelligence. For example, using these modern methods, a multidisciplinary team of experts led by Igor Zwir and Robert Cloninger (2018–2021) identified five clusters of people with different genotypic profiles who differed in Self-transcendence and other character traits. People with a genotypic predisposition for creative self-awareness were high in Self-transcendence accompanied by high Cooperativeness and Self-directedness; they were often idealistic, altruistic, and contemplative with frequent self-transcendent experiences. People with the genotypic network for intentional self-control were low in Self-transcendence and highly developed in both Self-directedness and Cooperativeness; they were often materialistic and individualistic and would do things for mutual benefit, but they would not make sacrifices for others. Other clusters with different genotypic backgrounds were low in Self-directedness and/or Cooperativeness, even though some of these reported that they were high in Self-transcendence. These latter individuals had PD with schizotypical features and were highly emotionally reactive and cognitively disorganized despite reporting that they were high in Self-transcendence. In other words, people who report that they are high in Self-transcendence may be either highly self-aware (i.e., high in all three character traits), or they may have magical thinking and perceptual aberrations and not be truly self-transcendent, as is seen in schizotypy. This heterogeneity became clear once multi-trait profiles were studied and was confirmed once the different molecular and brain mechanisms could be measured. Until this heterogeneity was uncovered, however, many clinicians, unfortunately, neglected to distinguish between individuals who were highly self-aware and those with magical thinking. Studies that used the individual measure of high Self-transcendence to predict well-being found highly variable outcomes, which some falsely concluded implied that self-transcendence was not a beneficial trait. Today, we are fortunate that modern tools have allowed us to show that the creative profile (i.e., high scores on all three character traits) is a strong and consistent predictor of physical, mental, and social well-being. In other words, self-awareness leads to all three character traits being well-developed, so a person cannot be truly self-transcendent unless they are also self-directed and cooperative. Likewise, a person cannot be fully hopeful and self-actualizing (spiritual aspects of self-directedness) or loving and compassionate (spiritual aspects of Cooperativeness) unless they are truly self-transcendent.

GENERAL IMPORTANCE OF PERSONALITY ASSESSMENT IN MEDICINE

The understanding of personality and its disorders is what has most often distinguished psychiatry fundamentally from other branches of medicine. A person is a self-aware human being, not a machine-like object that lacks self-awareness. Personality refers to all the characteristics that distinguish a constantly developing, self-organizing human being from a predictable machine-like object. In other words, personality refers to all the ways by which someone shapes and adapts in a unique way to an ever-changing internal and external environment.

Diverse therapeutic techniques, such as diet, exercise, mood self-regulation, meditation, or acts of kindness, influence health, and personality development in ways that are largely indistinguishable from one another or from effective allopathic treatments. Hence, according to the principles of person-centered therapeutics described by the Cloningers, the development of well-being appears to be the result of activating a synergistic set of mechanisms of well-being, which are expressed as greater functioning, plasticity, and virtue in adapting to life's challenges. The conditions that promote well-being and recovery from illness involve a set of common features of health that include hope, empathy, and respect for both self and others. These common characteristics of well-being emerge from a self-transcendent outlook on life, which is typical of healthy and creative people in all human cultures. An outlook of separation predisposes a person to feelings of fear, alienation, and individual pride or shame, thereby predisposing them to ill health and dissatisfaction with life. Person-centered care promotes health by providing the experience of an outlook of unity in the therapeutic alliance, which can later be generalized beyond the alliance. An outlook of unity fosters well-being by activation of a synergistic spiral of increasing self-directedness, cooperativeness, and self-transcendence, as will be described in more detail later.

Only about 15% of the variance in treatment outcomes is attributable to specific techniques of different psychotherapeutic schools, whereas about 85% of the variance in psychotherapy outcomes is explained by common factors shared by different approaches. What is attributed to the strong placebo effects observed in most drug or psychotherapy trials is substantially determined by common psychosocial factors, which can be as large or larger than putatively specific treatments. The common factors in all treatments include the patient's characteristics, the therapist's qualities of respect (i.e., prizing, unconditional positive regard, acceptance, trust), empathic understanding, and genuineness (i.e., realness, authenticity), and the quality of the therapeutic alliance between them (shared goals, emotional engagement, and exchange). In other words, the common factors for promoting well-being by positive personality development depends on promoting self-transcendence in a human relationship or therapeutic alliance that is intersubjective and coactive. These common factors are characteristic of all truly healthy human relationships, and they are important for the maintenance and recovery of all aspects of well-being, whether physical, mental, social, or spiritual.

Knowledge of the personalities of people provides the biopsychosocial context in which their behavior and relationships can be adequately understood as a complex adaptive process, which is crucial for all diagnosis and treatment. As a result, no psychiatric assessment is adequate without a description of the patient's personality and its development across the lifespan. In fact, it can be argued that person-centered approaches are preferred in all branches of medicine because the neglect of thorough personality assessment and its implications for health leads to people developing many chronic diseases, which are the source of most of the burden of disease in affluent societies. More than 70% of the burden of disease overall could be prevented by universal health care guided by person-centered principles in the individual care and public health and, in particular, effective personality assessment and treatment.

Even if a psychiatrist takes a reductive, rather than an integrative, view of psychological medicine, the assessment of PD is still essential in clinical practice because PD is such a common condition in contemporary society. PD occurs in 10% to 20% of the general population and in about half of psychiatric inpatients or outpatients. People with PD have chronic impairments in their ability to work and love, tend to be less educated, drug dependent, single, unemployed, and have marital difficulties. They consume a large portion of community services, social welfare benefits, public health, and prison resources.

Assessment and Treatment in Clinical Practice

Despite the great importance of personality and its disorders, clinicians seldom have the time or inclination to ask all the questions specified in the DSM-5 criteria for PD. Most psychiatrists find the official criteria of little utility except for rare and extreme prototypic cases. Official criteria have failed to provide the practicing clinician a tool to address personality assessment in an efficient and systematic way that is needed for optimal patient care.

The limitations of DSM for assessing PD are largely the result of its reliance on categorical criteria despite the absence of evidence that PD is composed of discrete entities. Also, most dimensional systems for personality describe the differences between people rather than the intrapsychic processes by which an individual can grow and develop in mental health. As a result, most categorical and dimensional systems for personality have become stigmatizing and are not useful prescriptions for positive personality development. Persons with PD are frequently labeled as "aggravating," "demanding," or "parasitic" and are generally considered to have a poor prognosis for emotional growth and maturation. Alternatively, they may be seductive or dependent and elicit inappropriate blurring of professional boundaries, such as sexual interest or the urge to rescue. These patients challenge the limits of physicians' theoretical knowledge, clinical skills, and even personal maturity. In general, practitioners with low narcissism and high levels of energy, tolerance, flexibility, creativity, and compassion are optimal for treating patients with PD because their own self-actualization allows them to maintain an encouraging and resilient working alliance without the patient dropping out or the practitioner burning out. Recognizing and treating disorders of personality helps clinicians to develop their personal maturity and clinical skills.

Fortunately, it is now possible to carry out a brief assessment of personality and its disorders as part of routine psychiatric examinations in an efficient way that may actually make assessments of both personality and other psychopathology more reliable and valid by identifying risk and protective factors for psychopathology. Such assessments also serve to establish a therapeutic alliance because trust can only be established by intersubjective interactions. Everyone likes to feel that they are appreciated and understood as a person who is respected and valued, not a separate object or diagnostic case. People feel appreciated when a clinician can predict their motivation and behavior with understanding and accuracy. Therefore, the next part of this chapter discusses some basic definitions and then provides the essentials for clinical assessment of personality as an integrated part of the mental status and psychiatric history. Later sections expand upon the essentials to provide a fuller understanding of disorders of personality and their measurement, etiology, development, and treatment.

BASIC DEFINITIONS AND TERMINOLOGY

The terms personality, temperament, motivation, character, and psyche are often used interchangeably. This is both misleading and unjustified. In what follows, these concepts are reviewed to distinguish them with more clarity.

Personality

Personality is both complex and unique: on the one hand, people differ greatly from one another in multiple components of behavior, and, on the other hand, each person expresses only one of their many potential lifestyles. The common characteristic of all existing definitions of personality is that they are functional. That is, they focus on questions related to motivation and the mental adaptation of the organism. Specifically, Allport defined personality as the “dynamic organization within the individual of those psychophysical systems that determine his/her unique adjustment to his/her environment.”

Allport elaborated on this definition by explaining that the expression “dynamic organization” emphasizes that personality is an organized system (“unitas multiplex”) that is constantly evolving and changing. The phrase “within the individual” means that personality refers to intrapsychic processes, not judgmental comparisons of one person to another. The term “psychophysical” means that personality is neither exclusively mental nor exclusively neural but a combination of the two. The expression “unique adjustment to the environment” has both functional and evolutionary significance pointing to personality as a mode of survival and, more generally, learning and adaptation, which is unique to each individual. Allport's use of the verb "determine" suggests that personality traits are determining dispositions that limit a person's behavioral repertoire.

It is now known that people are self-aware and have enough freedom of will to change their behavior in directions that are not necessarily predictable or determined by their past or genetic predisposition. Even monozygotic twins change the way their genes are expressed over the course of their lifetime as a result of epigenetic processes. Human self-awareness permits plasticity in the genome, the brain, and other organs to be shaped by the deliberate choices we make about diet, activity, social relationships, and other intentional, intersubjective, and creative activities. Accordingly, Allport's definition has been modified as follows: human personality is *the dynamic organization of the biopsychosocial systems by which a person shapes and adapts in a unique way to a changing internal and external environment.* In other words, personality is a way of describing the dynamic processes that occur within the person to shape and adapt to life experiences by self-aware learning and proactive planning. The maturation and integration of human personality involves one's growing self-awareness through experiences across a wide range of situations.

It has been widely accepted that personality develops through the interaction of hereditary dispositions and environmental influences. Waddington's notion of genetic canalization (or "epigenetic landscape") has been revised to include reciprocal interactions among genetic endowment, environmental stimulation, and changes in self-awareness across the lifespan. Genetic differences account for about half of the variance in differences between people for most normally distributed temperament traits. The remaining 50% of the variance in differences between people is attributed to environmental influences. Most of the environmental influences are unique to each individual, particularly for temperament traits, but parental rearing and sociocultural norms have a substantial influence on a person's character development. Of note, adoption studies suggest a somewhat lower heritability of about 30% for personality traits. The higher heritability estimates from twins are the result of nonadditive genetic influences (e.g., higher-order interaction among alleles at each locus or among loci), which is the same for monozygotic twins but contributes little to the resemblance of other relatives. Epistasis (i.e., interactions among multiple gene loci) is so strong for both temperament and character that it was difficult to identify specific genes for personality until methods were developed to take epistasis into account.

However, heritability estimates refer only to genetic influences on the differences between people at one point in time. People are able to change from one time to another as a result of growth in self-awareness in ways that are only partially constrained by their past genetic expression. For example, the intelligence quotient (IQ) is highly heritable, but there has been an increase of about 3.3 IQ points per decade that requires frequent restandardization of IQ tests. The increase in IQ scores in successive birth cohorts is called the “Flynn effect.” The increase in IQ is attributable to skill in problem solving on the spot (“fluid intelligence”), such as recognition of similarities, rather than skills dependent on past individual education, such as vocabulary size (“crystallized intelligence”). Likewise, the differences between people in Self-Transcendence at one point in time is 49% heritable, but the prevalence of self-transcendent experiences increased from 48% in 1987 to 76% in 2000 in English people according to studies of the zoologist David Hay. These rapid changes in fluid intelligence and self-transcendence cannot be explained by genetic evolution. Character and fluid intelligence can develop rapidly within individuals in response to cultural developments that influence human society as a whole and/or in response to changes in self-awareness.

From the structural standpoint, personality can be decomposed into temperament, character, and psyche. Roughly speaking, temperament involves basic emotions, character involves rational concepts and propositions of a person and their relationships, and the psyche involves intuitive self-awareness and fluid intelligence. Identity formation is the process of growing in self-awareness of all aspects of being, as discussed later. Self-awareness and fluid intelligence influence personality development substantially, so measures of temperament and character provide an incomplete understanding of personality development unless the distinctive role of self-transcendence as a process and value that motivates growth and self-actualization is recognized. The basic functions of personality are to feel, think, and perceive and to incorporate these into purposeful behaviors that flexibly and creatively express our identity under changing conditions.

Temperament

Temperament refers to the body’s biases in the modulation of conditioned behavioral responses to prescriptive physical stimuli. Cloninger and colleagues (2019) define a person’s temperament as their disposition to learn how to behave, react emotionally, and form attachments automatically by associative conditioning. This definition provides necessary and sufficient criteria to distinguish temperament from other aspects of personality that become integrated with it across the lifespan. Recent genetic work has found over 700 genes that account for nearly all the heritability of human temperament and has confirmed that they are consistently enriched in pathways activated by behavioral conditioning in animals. These pathways include the two major molecular pathways for response to extracellular stimuli, the Ras-MEK-ERK, and the PI3K-AKT-mTOR cascades. These pathways are activated by a wide variety of physiologic and psychosocial stimuli that vary in positive and negative valence and in consequences for health and survival. Changes in these pathways are orchestrated to maintain cellular homeostasis despite changing conditions by modulating temperament and its circadian and seasonal rhythms.

Behavioral conditioning (i.e., procedural learning) involves presemantic sensations that elicit basic emotions, such as fear or anger, independent of conscious recognition, descriptive observation, reflection, or reasoning. Pioneering work by A. Thomas and S. Chess conceptualized temperament as the stylistic component (HOW) of behavior, as differentiated from motivation (WHY) and content (WHAT) of behavior. Modern concepts of temperament, however, emphasize its emotional, motivational, and adaptive aspects. Specifically, four major temperament traits have been identified and subjected to extensive neurobiologic, psychosocial, and clinical investigation in humans and other animals: Harm Avoidance, Novelty Seeking, Reward Dependence, and Persistence. It is remarkable that this four-factor model of temperament is roughly similar to the ancient four temperaments: individuals differ in the degree to which they are melancholic (Harm Avoidance), choleric (Novelty Seeking), sanguine (Reward Dependence), and phlegmatic (Persistence). However, the four temperaments are now understood to be quantitative traits that occur in all possible combinations within the same individual rather than as mutually exclusive categories.

Psychobiology of Temperament. Temperament traits of Harm Avoidance, Novelty Seeking, Reward Dependence, and Persistence are defined as heritable differences underlying one's automatic response to danger, novelty, social approval, and intermittent reward, respectively. These four temperament traits are closely associated with the four basic emotions of fear (Harm Avoidance), anger (Novelty Seeking), attachment (Reward Dependence), and ambition (Persistence).

Individual differences in temperament and basic emotions modify the processing of sensory information and shape early learning characteristics, especially associative conditioning of unconscious behavior responses and preattentive components of perception. Temperament is conceptualized as heritable biases in emotionality and learning that underlie the acquisition of emotion-based, automatic behavioral traits and habits observable early in life and relatively stable over one's lifespan.

Each of the four major dimensions is a normally distributed quantitative trait, moderately heritable, observable early in childhood, relatively stable in time, and moderately predictive of adolescent and adult behavior. Temperamental differences, which are not very stable initially, tend to stabilize during the second and third years of life. Accordingly, ratings of these four temperament traits at age 10 to 11 years were moderately predictive of personality traits at ages 15, 18, and 27 years in a large sample of Swedish children. Whereas character traits develop in a specific direction toward socially favored cultural norms, the number of people who increase in Harm Avoidance, Reward Dependence, and Persistence is about the same as the number who decrease, so that the average of these traits remains the same from ages 20 to 45 years. Novelty Seeking decreases about 10% per decade as a person's impulse control matures.

Table 26–1.
Descriptors of Individuals Who Score High and Low on the Four Temperament Dimensions

Temperament Dimension	Descriptors of Extreme Variants	
	High	**Low**
Harm avoidance	Pessimistic	Optimistic
	Fearful	Daring
	Shy	Outgoing
	Fatigable	Energetic
Novelty seeking	Exploratory	Reserved
	Impulsive	Deliberate
	Extravagant	Thrifty
	Disorderly	Orderly
Reward dependence	Sentimental	Detached
	Open	Aloof
	Warm	Cold
	Affectionate	Independent
Persistence	Industrious	Lazy
	Determined	Spoiled
	Ambitious	Underachiever
	Perfectionist	Pragmatist

The four dimensions have been repeatedly shown to be universal across different cultures, ethnic groups, and political systems on every inhabited continent. In summary, these aspects of personality are called temperament because they are heritable, manifest early in life, are developmentally stable, and are consistent in different cultures. Temperament traits are like crystallized intelligence in that they do not show the rapid changes within increasing age or across birth cohorts that are observed for fluid intelligence and character traits. Table 26–1 summarizes contrasting sets of behaviors that distinguish extreme scorers on the four dimensions of temperament. Note that each extreme of these dimensions has specific adaptive advantages and disadvantages, so neither high nor low scores inherently mean better adaptation.

The component traits ("facets") for each of the four temperament dimensions have distinct learning characteristics and correlate more strongly with one another than with other components of temperament. They share a common source of covariation that is strong and invariant regardless of changes in the environment and past experience. Each of the four temperament dimensions has unique genetic determinants according to family and twin studies, as well as studies of genetic associations with specific DNA markers. However, genome-wide association studies have uncovered groups of genes underlying configurations of multiple temperament traits corresponding to those described in children by Chess and Thomas: (i) "easy" children who have "reliable" temperament profiles with low Harm Avoidance, low Novelty Seeking, high Reward Dependence, and high Persistence, (ii) "difficult" children who have a "sensitive temperament profile" with high Harm Avoidance, Novelty Seeking, and Reward Dependence, and "slow to warm up" who have an "antisocial temperament profile" with high Novelty Seeking, low Reward Dependence, and low Persistence. Despite the value of multi-trait profiles, most research and clinical work has preferred to use quantitative dimensional measures for the assessment of brain mechanisms in research and for clinical assessment and treatment. In practice, configurations can be derived from profiles of their individual constituents, and both approaches have utility.

The four dimensions of human temperament correspond closely to those observed in other mammals. Multiple levels in the phylogeny of learning abilities in animals, from invertebrates to man, indicate that learning has multiple component processes that are hierarchically organized and interact extensively throughout development (Fig. 26–1).

In Figure 26–1, temperament corresponds to the processes of sensation, association, and motivation that underlie the integration of skills and habits based on emotion. Figure 26–1 is useful to specify the hierarchical phylogenetic and ontogenetic organization of learning and its relevance to the concepts of temperament and character, not to compare learning in humans at a particular age to learning in lower animals. Specifically, temperament and character are conceptualized on the basis of three types of memory and learning that have been described in humans. These are complex brain systems for behavioral conditioning (also called procedural learning of habits and skills), intentionality (including executive self-control for attaining goals, semantic learning of facts, and abstract symbolization), and self-awareness. Self-awareness includes constructive learning (i.e., recollection of the past and imagining the future), prospective learning (i.e., the encoding, storing, and retrieval of possible intentions), and the related capacities for a theory of mind (i.e., the ability to attribute mental states to ourself and others to facilitate empathic social interaction), the default mode (i.e., awareness of internal milieu without focus on external tasks), autobiographical memory (i.e., vivid recollection of past experiences with contextual awareness of when and where facts were learned), and story-telling (i.e., the meaningful composition of narrative figural art and language), which are all aspects of self-awareness and share largely overlapping brain circuitry. Temperament (the “emotional core” of personality) involves procedural memory regulated by complex distributed circuits in the cortico–striato–limbic system, primarily the cortical sensory areas, amygdala, and the caudate, and the putamen. Procedural memory underlies associative learning and involves presemantic perceptual processing of visuospatial information and affective valence that can operate independently of abstract conceptual and/or volitional processes. In contrast, intentionality involves higher cognitive functions of self-control, abstraction, and symbolization. These three systems of learning and memory can be dissociated functionally from one another but normally interact cooperatively so that habits, goals, and values can be integrated.

Neurobiologic studies of animals have been highly informative about the functional organization of brain systems underlying procedural learning and temperament. The explicit animal models that were originally described on the basis of extensive work in rodents and other nonprimates provided hypotheses that have recently been confirmed by extensive work in humans using modern techniques of brain imaging, neurochemistry, and neurogenetics. The most comprehensive neurobiologic model of learning in animals that has been systematically related to the structure of human temperament is summarized in Table 26–2. This model distinguishes four dissociable brain systems for behavioral inhibition (Harm Avoidance), behavioral activation (Novelty Seeking), social attachment (Reward Dependence), and partial reinforcement (Persistence).

The psychobiology of Harm Avoidance, Novelty Seeking, Reward Dependence, and Persistence is briefly described later.

HARM AVOIDANCE. Harm Avoidance involves a heritable bias in the inhibition of behavior in response to signals of punishment and frustrative non-reward. It is observed as fear of uncertainty, shyness, social inhibition, passive avoidance of problems/danger, rapid fatigability, and pessimistic worry in anticipation of problems even in situations that do not worry other people. Adaptive advantages of high Harm Avoidance are cautiousness and careful planning when hazard is likely. The disadvantages occur when hazard is unlikely but still is anticipated, which leads to maladaptive inhibition and anxiety. People low in Harm Avoidance are carefree, courageous, energetic, outgoing, and optimistic, even in situations that worry most people. The advantages of low Harm Avoidance are confidence in the face of danger and uncertainty, leading to optimistic and energetic efforts with little or no distress. The disadvantages are related to unresponsiveness to danger or unrealistic optimism with potentially severe consequences when hazard is likely.

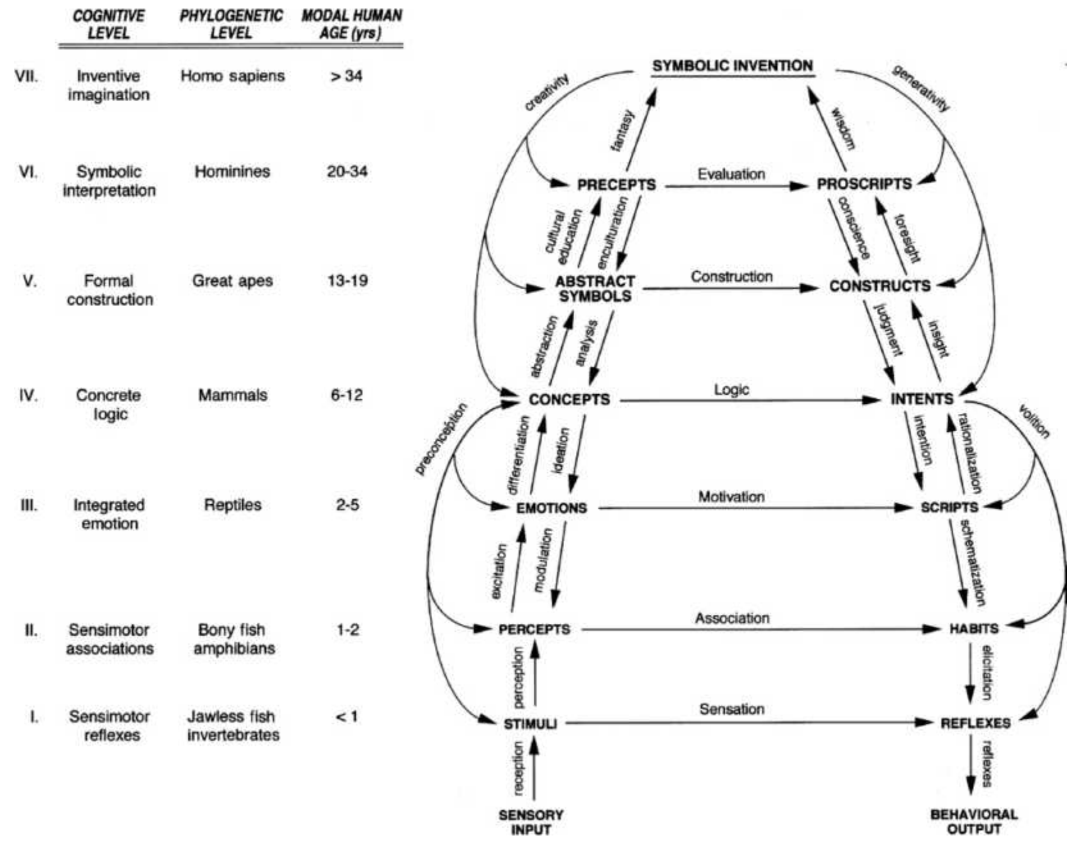

Table 26–2.
Four Dissociable Brain Systems Influencing Stimulus–Response Patterns Underlying Temperament

Brain System (Related Personality Dimension)	Principal Neuromodulators	Relevant Stimuli	Behavioral Response
Behavioral inhibition (Harm avoidance)	GABA Serotonin (dorsal raphe)	Aversive conditioning (pairing CS and UCS) Conditioned signals for punishment and frustrative non-reward	Formation of aversive CS Passive avoidance Extinction
Behavioral activation (Novelty seeking)	Dopamine	Novelty CS of reward CS or UCS of relief of monotony or punishment	Exploratory pursuit Appetitive approach Active avoidance Escape
Social attachment (Reward dependence)	Norepinephrine Serotonin (median raphe)	Reward conditioning (pairing CS and UCS)	Formation of appetitive CS
Partial reinforcement (Persistence)	Glutamate Serotonin (dorsal raphe)	Intermittent (partial) reinforcement	Resistance to extinction

CS, conditioned stimulus; GABA, γ-aminobutyric acid; UCS, unconditioned stimulus.

Adapted from *Cloninger CR. A systematic method for clinical description and classification of personality variants. A proposal. *Arch Gen Psychiatry*. 1987;44:573–588.

The psychobiology of Harm Avoidance is complex. In animal studies, ascending serotonergic projections from the dorsal raphe nuclei to the substantia nigra inhibit nigrostriatal dopaminergic neurons and are essential for conditioned inhibition of activity by signals of punishment and frustrative non-reward. Benzodiazepines disinhibit passive avoidance conditioning by gamma-aminobutyric acid (GABA)-ergic inhibition of serotonergic neurons originating in the dorsal raphe nuclei. The anterior serotonergic cells in the dorsal raphe nucleus intermingle with the dopaminergic cells of the ventral tegmental area, and both groups innervate the same structures (e.g., basal ganglia, accumbens, amygdala), providing opposing dopaminergic–serotonergic influences in the modulation of approach and avoidance behavior. The anterior serotonergic projections from the dorsal raphe to striatum, accumbens, amygdala and frontal cortex are usually associated with 5HT2 receptors. Individuals who are high in both Harm Avoidance and Novelty Seeking are expected to have frequent approach-avoidance conflicts, as seen in cycles of binging and purging in bulimia. More generally, excessive behavioral inhibition (i.e., high Harm Avoidance) predisposes individuals to anxiety, depression, and low self-esteem. Effective antidepressant treatment lowers scores in Harm Avoidance, but higher scores in Harm Avoidance predict poorer responses to antidepressants, including tricyclics and selective serotonin reuptake inhibitors.

Recent neuropsychological studies confirm that Harm Avoidance is associated with individual differences in classical aversive conditioning, whereas other dimensions of personality are uncorrelated (Table 26–3). Harm Avoidance, and not other dimensions of temperament, is also reliably associated with individual differences in startle responses: individuals who are high in Harm Avoidance have a greater startle response when they are viewing unpleasant stimuli, whereas those who are low in Harm Avoidance have a reduced startle response when they are viewing pleasant stimuli. Other neuropsychological tests also confirm that Harm Avoidance is associated with behavioral inhibition, as measured by Mismatch Negativity, an indicator of early preattentive sensory processing and automatic attention in orienting responses. Hence, the average effects of Harm Avoidance on startle obscure the important nonlinear interactions between predisposition and mental state. Likewise, Harm Avoidance is associated with the Posner validity effect (the Posner task uses a detection reaction time paradigm with the discrete presentation of simple visual stimuli in the periphery). Cues that correctly direct attention to the spatial location where the target stimulus will appear are called valid cues, whereas those that direct attention to incorrect locations are called invalid cues. In three large samples of healthy college students, those higher in Harm Avoidance had greater slowing of their reactions after invalid cues, or less benefit from valid cues, than others. Harm Avoidance is associated with greater sensitivity to the perception of fearful stimuli, which is mediated by strong functional connectivity between the amygdala and the anterior cingulate cortex. Even at rest, sustained fearful reactivity is sustained in people with the stable trait of Harm Avoidance, which is associated with reductions of slow oscillatory activity of the dorsal part of the medial prefrontal cortex in several independent studies.

Table 26–3.
Psychobiological Findings about Harm Avoidance

Variables	Observed Effects of High Harm Avoidance
Neuropsychology	
Aversive conditioning	Greater associative conditioning with punishment (*r* = 0.4)
Startle reflex	Potentiation of eye-blink response to aversive stimulus when HA is high and reduction when low
Fear perception	Greater sensitivity to fearful stimuli from stronger Amgydalar–Anterior Cingulate connectivity (*r* = 0.5)
Fear sustainment	Reduced slow low-frequency oscillations at rest in dorsomedial prefrontal cortex associated with greater amygdalar connectivity
Mismatch negativity	Greater amplitude of inhibiting early orienting response to deviant sounds in frontal cortex (*r* = –0.4)
Posner validity effect	Greater slowing of responses after invalid cues (*r* = 0.3)
Neuroimaging	
Insular salience network	Increased connectivity of right anterior insula with anterior cingulate cortex and dorsomedial prefrontal cortex
Amygdala activation	Increased sensitivity to fearful stimuli mediated by activation of amygdala and its connectivity with perigenual cingulate and dorsomedial prefrontal cortices
Volume of anterior cingulate cortex	Anterior cingulate cortex volume is larger in right and smaller in left hemisphere
Volume of cerebellar cortex	White matter and cortex smaller bilaterally
Mean diffusivity of striatum	Increased in pallidum bilaterally
Neurochemistry	
Platelet 5HT-2 receptor	Fewer receptors (*r* = –0.6)
Plasma GABA	Lower level (*r* = –0.5)
Neurogenetics	
5HT Transporter promoter gene	Greater reuptake activity
Tryptophan	Slower serotonin synthesis, especially in women reared in a

hydroxylase	hostile parental environment

Carriers of a gene that influences the activity of the serotonin transporter have increased Harm Avoidance, increased reactivity of the amygdala to fear, and increased risk of depression compared to others. Harm Avoidance scores explain 30% of the variance in the functional connectivity between the amygdala and perigenual anterior cingulate, which are involved in the perceptual processing of fearful stimuli. There is also a correspondence between structure and function: healthy human volunteers who are high in Harm Avoidance tend to have a greater volume of their right anterior cingulate gyrus and smaller volume of cerebellar white matter and cortex bilaterally (Table 26–3). Likewise, individual differences in Harm Avoidance are significantly negatively correlated with functional differences in activity in the dorsal part of the medial prefrontal network at rest in several independent studies. This includes significant negative correlations between regional blood flow and the score for Harm Avoidance in paralimbic regions, such as the left parahippocampal gyrus and the left orbitoinsular junction, and various neocortical regions in the frontal, parietal, and temporal cortex. Positron emission tomography at NIMH with 18-FDG in 31 healthy adult volunteers during a simple continuous Performance Task showed that Harm Avoidance was associated with increased activity in the anterior paralimbic circuit, specifically the right amygdala and insula, the right orbitofrontal cortex, and the left medial prefrontal cortex. Subsequent fMRI studies have confirmed that individuals with high levels of Harm Avoidance have increased sensitivity to fearful stimuli mediated by activation of the amygdala and its connectivity with the perigenual cingulate cortex and the dorsal part of the medial prefrontal cortex. This stress sensitivity network also interacts with the Insular Salience Network, which involves the right anterior insula and its connections with the anterior cingulate and the dorsomedial prefrontal cortex. This activation pattern corresponds well to the 5HT2 terminal projections of the dorsal raphe. However, 5HT2 has been measured only in platelets (see Table 26–3).

Higher plasma GABA levels have also been correlated with low Harm Avoidance. Plasma GABA has also been correlated with other measures of anxiety proneness and correlates highly with brain GABA levels. Finally, a gene on chromosome 17q12 that regulates the expression of the serotonin transporter has been found to account for 4% to 9% of the total variance in Harm Avoidance in most, but not all, of the tests of this relationship. These findings support a role for both GABA and serotonergic projections from the dorsal raphe underlying individual differences in behavioral inhibition as measured by the trait of Harm Avoidance.

NOVELTY SEEKING. Novelty Seeking reflects a heritable bias in the initiation or activation of appetitive approach in response to novelty, approach to signals of reward, active avoidance of conditioned signals of punishment, and escape from unconditioned punishment (all of which are hypothesized to covary as part of one heritable system of learning). This trait is called Novelty Seeking because its most distinctive feature is exploratory approach behavior, as confirmed by fMRI studies showing increased striatal dopamine release activated by novel stimuli regardless of whether they predict reward or not in people high in Novelty Seeking (Table 26–4). Novelty Seeking is observed as exploratory activity in response to novelty, impulsiveness, extravagance in approach to cues of reward, and active avoidance of frustration. Individuals high in Novelty Seeking are quick-tempered, curious, easily bored, impulsive, extravagant, and disorderly. Adaptive advantages of high Novelty Seeking are an enthusiastic exploration of new and unfamiliar stimuli, potentially leading to originality, discoveries, and reward. The disadvantages are boredom susceptibility, impulsivity, angry outbursts, potential fickleness in relationships, and impressionism in efforts. Persons low in Novelty Seeking are slow tempered, uninquiring, stoical, reflective, frugal, reserved, tolerant of monotony, and orderly. Their reflectiveness, systematic effort, and meticulous approach are clearly advantageous when these features are adaptively needed. The disadvantages reflect tolerance of monotony and lack of enthusiasm, potentially leading to prosaic routinization of activities.

Mesolimbic and mesofrontal dopaminergic projections have been shown to have a crucial role in incentive activation of each aspect of Novelty Seeking in animals. For example, dopamine-depleting lesions in the nucleus accumbens or the ventral tegmentum of rats and mice lead to neglect of novel environmental stimuli and reduce both spontaneous activity and investigative behavior. Behavioral activation by dopaminergic agonists is dependent on the integrity of the nucleus accumbens but not the caudate nucleus. In human studies, individuals at risk for Parkinson disease have low premorbid scores in Novelty Seeking but not other dimensions of personality, supporting the importance of dopamine in incentive activation of pleasurable behavior. The initiation and frequency of hyperactivity, binge eating, sexual hedonism, pathologic gambling, Internet addiction, drinking, smoking, and other substance abuse, especially stimulants, are each associated with high scores in Novelty Seeking.

The psychobiological correlates reported for Novelty Seeking are summarized in Table 26–4. High scores correlate with greater capacity for synthesis of dopamine in the ventral striatum and reduced D2/D3 autoreceptor availability in the substantia nigra and ventral tegmental area in the midbrain, thereby leading to accentuated dopaminergic responses to novelty and other conditions that induce dopamine release. Human functional neuroimaging studies also show that the high Novelty Seekers have greater functional connectivity with the hippocampus, amygdala, and orbitofrontal cortex in response to novel stimuli than other people. Structural imaging studies indicate that people with higher Novelty Seeking have larger left posterior cingulate gray matter volumes and smaller inferior frontal gray matter volumes.

Evoked potential studies confirm that Novelty Seeking is associated with greater cortical reactivity and augmentation of stimulus intensity in response to somatosensory and novel stimuli. Novelty Seekers are also sensitive to both sedatives and stimulants: they are easily oversedated by benzodiazepines and overstimulated by amphetamines, leading to deterioration in their information processing. Their reaction times are slow to neutral stimuli, but they show a greater approach to novel stimuli, whether these predict reward or not.

Table 26–4.
Psychobiological Findings about Novelty Seeking

Variables	Observed Effects of High Novelty Seeking
Neuropsychology	
Reinforcement learning	More approach to novel stimuli and slower to react to neutral stimuli
Neocortical reactivity	Greater cortical reactivity to somatosensory (MEG) and novel stimuli (ERP)
Striatal–mediated behavioral activation	Greater striatal dopamine release activated by novel stimuli that do not predict reward
Impulsive decision making	Greater ventromedial prefrontal cortical activity at rest leads to reduced delay of gratification
Cortisol response to stress	Reduced at rest and after social stress
Neuroimaging	
Connectivity of ventral striatum	Greater connectivity with hippocampus, amygdala, and orbitofrontal cortex is activated by novel stimuli
Insular salience network	Reduced activity
Posterior cingulate volume	Greater gray matter volume in left hemisphere
Inferior frontal cortex volume	Smaller gray matter volume bilaterally
Neurochemistry	
Ventral striatal dopamine synthesis	Greater synthesis capacity ($r = 0.8$)
Midbrain dopamine autoreceptors	Reduced autoreceptor availability in substantia nigra/ventral tegmental area of midbrain, leading to more rapid sensitization to stimulants
Dopamine DRD2 receptor	Negative correlation of Novelty Seeking with DRD2 levels in insula bilaterally in parkinsonism
Neurogenetics	
Dopamine transporter gene	Greater reuptake activity
Dopamine DRD2 receptor gene	Moderates development of higher adult Novelty Seeking in response to harsh discipline in childhood

The association of increased striatal dopamine release in response to novel stimuli with high Novelty Seeking is more specifically associated with a higher density of the dopamine transporter, suggesting that Novelty Seeking involves increased reuptake of dopamine at presynaptic terminals, thereby requiring frequent stimulation to maintain optimal levels of postsynaptic dopaminergic stimulation. Novelty Seeking leads to impulsive decision making and various pleasure-seeking behaviors, including cigarette smoking, which may explain the frequent observation of low platelet MAO B activity because cigarette smoking has the effect of inhibiting MAO B activity in platelets and in the brain.

Studies of candidate genes involved in dopamine neurotransmission (e.g., dopamine transporter and D4DR) have provided evidence of association with Novelty Seeking and no other dimension of temperament. The dopamine transporter, which is responsible for the presynaptic reuptake of dopamine and a major site of action of drugs, including stimulants such as methylphenidate, is encoded by a gene locus on chromosome 5p. Polymorphisms at this gene locus are associated with Attention-Deficit Disorder and other disorders related to variation in Novelty Seeking. Likewise, polymorphisms at the DRD4 locus have been associated with Attention-Deficit Disorder, opioid dependence, and other traits related to Novelty Seeking.

REWARD DEPENDENCE. Reward Dependence reflects a heritable bias in the maintenance of behavior in response to cues of social reward. Reward dependence is characterized by sentimentality, social sensitivity, attachment, and dependence on approval by others. Individuals high in Reward Dependence are tender-hearted, sensitive, dedicated, dependent, and warmly sociable. One of the major adaptive advantages of high Reward Dependence is the sensitivity to social cues, which facilitates affectionate, social relations, and genuine care for others. The disadvantage is related to suggestibility and loss of objectivity frequently encountered with people who are excessively socially dependent. Individuals low in Reward Dependence are practical, tough-minded, cold, socially insensitive, irresolute, and indifferent if alone. The advantages of low Reward Dependence are personal independence and objectivity not corrupted by efforts to please others. Its adaptive disadvantage is related to social withdrawal, detachment, and coldness in social attitudes.

Noradrenergic projections from the locus coeruleus and serotonergic projections from the median raphe are proposed to influence such reward conditioning (Table 26–2). In animals, stimulation of the noradrenergic locus coeruleus or its dorsal bundle, or direct application of norepinephrine, decreases the firing rate of terminal neurons and increases their sensitivity to other afferents so that targeted stimuli can stand out from nontargeted stimuli. In humans, short-term reduction of norepinephrine release by acute infusion of the alpha-2 presynaptic agonist clonidine selectively impairs paired-associate learning, particularly the acquisition of novel associations. The noradrenergic locus coeruleus is located at the same posterior level of the brainstem as the serotonergic median raphe, and both of these posterior monoamine cells innervate structures important to the formation of paired associations, such as the thalamus, neocortex, and hippocampus. Neurophysiologic studies show that the anterior temporal lobe decodes social signals, such as facial images and social gestures. Consequently, individuals high in Reward Dependence are expected to be particularly sensitive in their social communication, whereas those low in Reward Dependence are expected to be socially aloof. The “love hormone” oxytocin is released in response to sexual orgasm, and its plasma level is positively correlated ($r = 0.5$) with individual differences in Reward Dependence but not other dimensions of personality.

The psychobiological correlates reported about Reward Dependence are summarized in Table 26–5. As predicted, Reward Dependence is associated with individual differences in the formation of conditioned signals of reward. This is also supported by its association with individual differences in paired-associate learning. People who are high in Reward Dependence also show greater discrimination of facial emotions and social cues, which is mediated by their having larger oxytocinergic regions in the hypothalamus. In individuals with high Reward Dependence, the oxytocinergic regions of the hypothalamus have reduced connectivity with the amygdala, septum, and dorsal anterior cingulate in response to perceptual processing of facial emotions and social cues. Reward Dependence is also associated with greater opioid receptor availability in the ventral striatum bilaterally. Consequently, people who are high in Reward Dependence have readily form warm and caring attachments to other people.

Table 26–5.
Psychobiological Findings about Reward Dependence

Variables	Observed Effects of High Reward Dependence
Neuropsychology	
Reward conditioning	Enhanced associative pairing with rewards and not novelty without reward
Paired associates	Better learning of paired associations
Posner validity effect	Faster responses after valid cues
Perceptual processing of facial emotions and social cues	Better discrimination of facial emotions mediated by variability of regional oxytocin receptor activity
Mismatch negativity	Greater latency of preattentive MEG response to changes in tonal duration in left hemisphere
Neuroimaging	
Corticostriatal system	Greater gray matter volume and connectivity of striatum (caudate) with orbitofrontal cortex and temporal lobes
Caudate volume and activity	Greater size and activity
Oxytocinergic system	Larger oxytocinergic regions in hypothalamus with reduced connectivity to amygdala, septum, and dorsal anterior cingulate, which is a network activated by perceptual processing of facial emotions and social cues
Neurochemistry	
Striatal opioid availability	Greater opioid receptor availability in ventral striatum bilaterally
Plasma oxytocin	Higher oxytocin levels ($r = 0.5$)
Plasma cortisol	Higher morning cortisol when depressed ($r = 0.3$)
Neurogenetics	
5HT2 c receptor	Allelic association (effect size 2.0)
Epidermal growth factor	Epidermal growth factor A61G G/G polymorphism associated with higher RD in men

PERSISTENCE. Persistence reflects a heritable bias in the maintenance of behavior despite the frustration, fatigue, and intermittent reinforcement. It is observed as industriousness, determination, ambitiousness, and perfectionism. Highly Persistent people are hardworking, perseverant, and ambitious overachievers who tend to intensify their effort in response to anticipated reward and perceive frustration and fatigue as a personal challenge. High Persistence is an adaptive behavioral strategy when rewards are intermittent, and contingencies remain stable. When the contingencies change rapidly, perseveration becomes maladaptive. Individuals low in Persistence are indolent, inactive, unstable, and erratic; they tend to give up easily when faced with frustration, rarely strive for higher accomplishments, and manifest a low level of perseverance even in response to intermittent reward. Accordingly, low persistence is an adaptive strategy when reward contingencies change rapidly and may be maladaptive when rewards are infrequent but occur in the long run.

Persistence can be objectively measured by the partial reinforcement extinction effect (PREE), in which persistent individuals are more resistant to the extinction of previously intermittently rewarded behavior than other individuals who have been continuously reinforced. Earlier work in rodents showed that the integrity of the PREE depends on projections from the hippocampal subiculum to the nucleus accumbens. This glutaminergic projection may be considered as a short circuit from the behavioral inhibition system to the behavioral activation system, thereby converting a conditioned signal of punishment into a conditioned signal of anticipated reward. This connection is probably disrupted in humans by lesions of the orbitomedial cortex that may have a specific antipersistence effect of therapeutic benefit to some patients with severe obsessive-compulsive PD. Bilateral cingulotomy, which reduces Harm Avoidance only, is less effective in reducing persistent compulsive behavior than cingulotomy combined with orbitomedial lesions.

These findings in animals suggested that the regulation of Persistence involved the ventral striatum, which includes the nucleus accumbens. This has been tested in humans and confirmed (Table 26–6). Individual differences in persistence are strongly correlated (r = 0.8) with responses measured by fMRI in a circuit involving the ventral striatum, orbitofrontal cortex/rostral insula, and dorsal anterior cingulate cortex (BA 32). Subjects low in persistence exhibited relative decreases in activity within this circuit, whereas those high in persistence exhibited relative increases. Persistence scores also correlated with apparent selection bias, such that subjects with high persistence scores made relatively more pleasant judgments at the expense of neutral judgments when viewing pictures from the International Affective Picture System.

Motivation

According to materialistic theories of personality, survival and reproduction are the basic drives for all animal species, including human beings. These are expressed in humans primarily through affects and emotions, which are derived from their experiences. In contrast to the limited motivational spectrum of the basic drives, emotions have an independent motivational power that makes them the primary motivational system for many people, including the personality processes that materialists use to seek satisfaction and meaning, albeit without enduring success because seeking satisfaction from what you cannot possess or control is ultimately disappointing (e.g., greed is insatiable, and nobody lives forever or without suffering).

Nevertheless, it is useful to recognize that the temperament traits of Harm Avoidance, Novelty Seeking, Reward Dependence, and Persistence, with their associated emotions of fear, anger, attachment, and ambition, are observable early in development. Research in children has demonstrated that, during the first several months of life, fear and anger differentiate from the disposition to distress (a tendency to become upset and autonomically aroused easily and intensely). During this period, which is characterized by active regressive changes in the organization of neuronal circuits and density of synapses, especially in frontal, temporal, and limbic cortical areas, many other complex and social functions in human behavior emerge. These new functions are, however, organized around an early set of moderately stable but modifiable behavioral and emotional dispositions referred to in this work as temperament traits.

A relatively limited spectrum of basic emotions is associated with the four temperament traits (Table 26–7). Depending on whether a particular temperament trait is high or low, certain emotions will tend to dominate one’s motivation, perception, and behavior, particularly when a person’s character is not yet highly developed.

Table 26–6.
Psychobiological Findings about Persistence

Variables	Observed Effects of High Persistence
Neuropsychology	
Partial reinforcement extinction effect (PREE)	Greater resistance to extinction (quitting effort to obtain a goal) despite frustration from infrequent reward with high persistence
Positively biased reward perception	Neutral stimuli perceived as rewards
Cortical reactivity	Reduced cortical excitability at rest
Phonemic mismatch negativity	Greater strength of preattentive response to infrequent vowel sounds
Rey word list learning	Better learning without reinforcement
Intelligence and achievement	Higher academic achievement and slightly higher IQ
Neuroimaging	
Reward-seeking system	Greater activation of circuit involving ventral striatum, lateral orbitofrontal cortex (BA 47), and anterior cingulate in response to intermittent rewards, giving rise to the PREE
Striatal-mOFC connectivity	Greater connectivity of striatum and medial orbitofrontal cortex (mOFC)
Hippocampal–accumbens pathway	Glutaminergic projection from subiculum to nucleus accumbens is essential to PREE
Neurochemistry	
Salivary alpha-amylase activity	Greater noradrenergic arousal on average (whether at rest or stressed)
Neurogenetics	
Interaction of DAT1 and 5HT2a	Greater persistence is associated with interaction of dopamine transporter (DAT1) and serotonin receptor 5HT2A
Brain-derived neurotrophic factor	Association with Val66Met polymorphism modulates the effects of parental rearing

Table 26–7.
Effects of Positive (+) and Negative (–) Reinforcement on Emotional State of Four Temperaments

Temperament Dimension	High Scorers		Low Scorers	
	+	–	+	–
Harm avoidance	Anxious (agitated)	Depressed (retarded)	Cheerful	Fearless
Novelty seeking	Euphoric	Angry	Placid	Stoical
Reward dependence	Sympathetic	Disgusted	Aloof	Indifferent
Persistence	Enthusiastic	Steadfast	Unstable	Discouraged

As can be inferred from Table 26–7, the same external stimulus is likely to elicit responses via activation of multiple temperament dimensions. For example, novel or unfamiliar stimuli elicit interest in approach in proportion to Novelty Seeking, as well as inhibition of approach in proportion to Harm Avoidance. Consequently, people who are high in both Novelty Seeking and Harm Avoidance experience approach-avoidance conflicts, as seen in bulimics who binge and purge. Each temperament trait clearly involves an integration of multiple emotional drives that may be conflicting (competitive) or facilitatory (synergistic), particularly when a person is not fully self-actualized (i.e., when their habits are not yet congruent with their goals and values).

In contrast, dualistic theories of personality recognize that both the analytical mind (the intentional self) and temperament have an important influence on a person's motivation (i.e., what they want to do in life). According to this perspective, temperament involves a relatively small set of emotions associated with one's basic needs, for example, safety, which are called primary motives. Excessive fear and anger, associated with temperament, are then regarded as motivationally monopolistic and thought to take over the personality by altering perception, learning, and behavior in a biased way. However, under normal circumstances, after survival needs are met, the goals of normally developing personality change to include not only the integrity of the physical self but also the strength and integrity of the analytical mind or mental self (e.g., self-esteem). The mental self also allows adaptation to numerous social goals (e.g., education, occupation, family) and a rich spectrum of secondary (social) emotions (such as shame and compassion). These "secondary," "social," or "growth" motives are closely functionally related to character development according to non-materialistic (i.e., dualistic and holistic) models. Specifically, basic emotions of fear, anger, and excitement are transformed into more complex secondary emotions, such as carefulness, assertiveness, and joy, through the interaction with increasingly more complex internalized concepts associated with character. Although some basic character components develop early in life, such as trust and confidence, it is the completion of self-object differentiation (“Me” vs. “Not me”) between 18 months and 3 years of age that sets the stage for the later development of character traits and secondary emotions with an allocentric perspective, such as empathy, which emerges around age 7 years.

The secondary emotions may take over as primary motivators of further character development and maturation. Of note, they are not as monopolistic as the basic emotions and thus motivate the development of more flexible and adaptable personality traits. As shown in Table 26–8, each of the three character traits is associated with a typical pattern of secondary emotions.

Table 26–8.
Effects of Positive (+) and Negative (–) Reinforcement on Emotional State of Three Characters

Character Dimension	High Scorers		Low Scorers	
	+	–	+	–
Self-directed	Hopeful	Resourceful	Vain	Shameful
Cooperative	Loving	Forgiving	Scornful	Revengeful
Self-transcendent	Joyful	Peaceful	Greedy	Miserable

According to materialistic and dualistic models of motivation, abnormal (deviant, immature) motivation derives from two or three dominant, monopolistic elementary emotional needs associated with survival. In contrast, more mature motivation develops after basic needs are met, and the person is freed to experience numerous secondary motives for growth in character and in self-awareness. This could explain the motivational inflexibility and poverty of deviant personality and account for some of the rich motivational diversity and flexibility of people with mature self-control. However, executive functions and analytical reasoning do not explain the early emergence of self-transcendent features in children, such as trust. Nor does it really explain the personal motive to serve others or to make sacrifices that do not benefit the individual. It is important to recognize the need for cultivating self-transcendence throughout life for two reasons: (i) If self-transcendent motives are not basic, there is no potential for self-transcendence because something cannot arise from nothing; and (ii) If self-transcendent motives are basic but are not nurtured from the beginning, they are likely to wither and be ignored later, just as happens when the gifts of child prodigies are not cultivated. Fortunately, self-transcendent motives such as trust and love are basic aspects of life along with basic drives for purpose, meaning, and coherence in modern humans. Much research shows that symbiosis and cooperation are more basic motives in all life forms than fear and competition.

Empirically, any temperament profile can lead to well-being with healthy character development. Temperaments are only monopolistic when they are not integrated with regulatory character strengths.

Character

Character refers to the mind, that is, the conceptual core of personality. It involves individual differences in self-concepts and human relations that reflect personal goals and values. In other words, character is what individuals make of themselves intentionally and/or creatively. Character is rational and volitional. Whereas temperament involves basic emotions such as fear and anger, character involves secondary emotions such as empathy and shame, and, in self-transcendent individuals, virtues in action such as practical wisdom, hope, and compassion. As a result, character can be described as one's mental self-government, which involves executive, legislative, and judicial functions. In what follows, the concept of character is outlined on the basis of the seven-factor psychobiological model of personality, which is also called the biopsychosocial model, to emphasize the three processes of human learning that comprise personality. The psychodynamic concept of character, which has contributed many illuminating clinical and theoretical formulations from a dualistic perspective, is discussed in some detail as well.

Psychobiology of Character. Character (or the "conceptual core" of personality) involves higher cognitive functions, which are organized as two of the three human systems of learning and memory. Whereas temperament involves behavioral conditioning and procedural memory, character involves the brain networks for intentionality and self-awareness. The human system for intentionality including executive self-control for attaining goals, semantic learning of facts, and abstract symbolization. The human system for self-awareness includes constructive learning (i.e., recollection of the past and imagining the future), prospective learning (i.e., the encoding, storing, and evaluation of possible intentions), and the related capacities for the theory of mind (i.e., the ability to attribute mental states to ourself and others to facilitate empathic social interaction), the default mode (i.e., awareness of internal milieu without focus on external tasks), autobiographical memory (i.e., vivid recollection of past experiences with contextual awareness of when and where facts were learned), and story-telling (i.e., the meaningful composition of narrative figural art and language), which are all aspects of self-awareness and share largely overlapping brain circuitry.

All three systems are strongly related to personality, and they naturally interact to allow a person to adapt to changing circumstances and emerging opportunities. Such synergistic interactions mean that healthy human function cannot be reduced to materialistic models, as advocated by some molecular biologists, or to dualistic models, as advocated by some cognitive-behavioral or psychodynamic therapists. The brain networks for emotional reactivity, intentionality, and self-awareness function interdependently, not separately. They integrate information from all parts of the brain into a more-or-less coherent whole in self-awareness, and to do so requires insight into what is really valuable combined with the plasticity to change in order to moderate their appetites and habits. Put another way, materialistic approaches that treat a person’s brain and other parts of the body with medications may transiently mask symptoms to stabilize but do not help a person to learn how to live well. Likewise, dualistic approaches, which treat both behavioral and intellectual aspects of PD, do not allow people to let go of their struggles with themself and others. Healthy awareness depends on synergistic interactions among brain systems for self-awareness, intentionality, and behavioral conditioning. Well-being requires the wisdom to know what is good for yourself and others, the willingness to change flexibly, and the creativity and perseverance to find a way to achieve your valued goals. To understand this, it is useful first to describe the phenotypic components of character.

The executive, legislative, and judicial functions of mental self-government can be measured as three distinct character traits, which are called Self-Directedness, Cooperativeness, and Self-Transcendence, respectively. These character traits are adaptive, but their low ends are less advantageous because there are few circumstances in which low character development is adaptive, even temporarily. Self-Directedness quantifies differences in the executive competence of individuals. A highly self-directed person is self-sufficient, responsible, reliable, resourceful, goal-oriented, and self-accepting. The most advantageous summary feature of self-directed individuals is that they are realistic and effective; that is, they are able to adapt their behavior in accord with individually chosen, voluntary goals. Individuals low in Self-Directedness are blaming, helpless, irresponsible, unreliable, reactive, and unable to define, set, and pursue internally meaningful goals. Such poor executive function, manifest as unrealistic behavior and lack of internal guidance, is rarely advantageous to the individual.

Cooperativeness quantifies differences in the legislative functions of individuals. That is, Cooperativeness allows people to make the rules and follow the principles by which they get along with other people. Highly cooperative people conceptualize themselves as integral parts of human society. Such highly cooperative persons are described as tolerant, helpful, empathic, principled, and compassionate. These features are advantageous in teamwork and social groups but not for individuals who must live in a solitary manner. People who are low in Cooperativeness are intolerant, unhelpful, self-absorbed, opportunistic, unforgiving, and revengeful. They primarily look out for themselves and tend to be hostile and inconsiderate of other peoples’ rights or feelings.

Self-Transcendence quantifies individual differences in the judicial functions of people. In other words, Self-Transcendence allows one to recognize when various rules apply in a particular context (i.e., in the narrative of intersecting lives). Self-Transcendence reflects the extent to which people conceptualize themselves as an integral part of the universe as a whole. Self-transcendent individuals are described as judicious, insightful, idealistic, contemplative, and spiritual. These traits are adaptively advantageous when people are confronted with ultimate situations such as suffering, illness, or death, which is inevitable with advancing age. They may appear disadvantageous in secular, materialistic societies where idealism, modesty, and meditative search for meaning might interfere with the acquisition of wealth and power. People low in Self-Transcendence tend to be pragmatic, objective, materialistic, controlling, and pretentious. Such individuals appear to fit in well in most Western societies because of their pragmatism and materialistic ambitions. However, they consistently have difficulty accepting suffering, failures, personal and material losses, and death, which leads to a lack of serenity and adjustment problems, particularly with advancing age.

Contrasting sets of descriptors that distinguish high and low scorers are shown in Table 26–9. High and low scorers in each character dimension are distinguished by behavior traits that arise from differences in concepts that are each internally consistent but not logically falsifiable. For example, people low in Self-Transcendence live in the material world, skeptical of whatever they cannot prove objectively and use practically. In contrast, for highly Self-Transcendent individuals, the meaning of life goes beyond material things and includes intuitive awareness of what is beautiful, true, and good, to which materialists may be insensitive. Each set of beliefs appears to be internally consistent to those who hold them, who also regard their approach to life as realistic.

However, firm beliefs may lead to an identity that is long-lasting, but they are not a sufficient basis for well-being or the coherence of our habits, goals, and values with our human nature and the reality of all-being. The development of an authentic sense of identity cannot be determined by accepting or rejecting what others want us to believe and do. Each person must look within themselves to perceive and understand their own nature, what satisfies them, and what they find meaningful. We can only cultivate a personality that promotes well-being by learning the answers to these questions about life for ourselves. Each of us may differ in our initial intuitions about what is real, satisfying, and meaningful, but we can also learn by comparing our initial understanding to the scientific observations about the three aspects of human personality and the development of well-being. That way, we can learn from the objective and subjective experiences of others while relying ultimately on our own insight and judgment. Life is a wonderful opportunity to explore and discover what truly exists, what we can understand, and what is valuable and meaningful. Character develops in a stepwise manner in incremental shifts from infancy through late adulthood. Under health-promoting conditions, it develops along a path of increasing awareness leading to maturity and coherence. The timing and rate of transitions between levels of maturity are nonlinear functions of antecedent temperament configurations, systematic cultural biases, and experiences unique to each individual, which depend on individual differences in episodic memory or intuitive self-awareness that enables human beings to recollect past experiences and imagine future activities. Dualistic theories regard developing character traits as merely newly internalized concepts about one's self and the external world designed to optimize adaptation of temperament (i.e., early emotionality) to the environment by reducing discrepancies between one's emotional needs and norm-favoring social pressures. In contrast, research and treatment with the biopsychosocial model now recognizes that developing character strengths are inspired by our understanding of virtue and by our creativity in finding ways to express that insight in action flexibly so that our habits, goals, and values are congruent with our understanding of reality as a whole.

Table 26–9.
Descriptors of Individuals Who Score High and Low on the Three Character Dimensions

Character Dimension	Descriptors of Extreme Variants	
	High	**Low**
Self-Directedness	Responsible	Blaming
	Purposeful	Aimless
	Self-accepting	Vain
	Resourceful	Passive
	Spontaneous	Conflicted
Cooperative	Tolerant	Intolerant
	Empathic	Insensitive
	Helpful	Selfish
	Compassionate	Revengeful
	Principled	Opportunistic
Self-Transcendent	Self-forgetful	Self-preoccupied
	Transpersonal	Self-centered
	Spiritual	Skeptical
	Contemplative	Objective
	Idealistic	Pragmatic

In Figure 26–1, character corresponds to the processes of logic, construction, and evaluation of abstract symbols that are based on a conceptual representation of information and are well developed only in some mature humans. These processes are related to executive functions (predominantly logic), legislative functions (predominantly construction), and judicial functions (predominantly evaluation). Early fMRI research in healthy human volunteers by Debra Gusnard and Marc Raichle demonstrated a strong correlation between Self-Directedness and a cortical circuit involving the medial prefrontal cortex, which is known to regulate executive function (see Cloninger, 2004). Earlier work also showed that psychophysiological markers of neocortical processing, such as the P300 event-related potential and Contingent Negative Variation, are correlated with measures of character but not temperament. Namely, Self-Directedness, but not temperament traits, correlates moderately with the evoked potential P300 ($r = 0.4$, $P < .002$). Likewise, Cooperativeness correlates with the Contingent Negative Variation (this was particularly obvious for the Empathy subscale, $r = 0.4$). Individuals with Parkinson disorder differ in temperament from others (they are lower in Novelty Seeking premorbidly) but not in character. These early fMRI and ERP findings have been confirmed and extended by recent structural and functioning neuroimaging findings demonstrating that character involves higher cortical functions instantiated in specific neocortical networks of learning and memory (Table 26–10). The executive and legislative functions related to Self-Directedness and Cooperativeness involve conceptual processing in components of the intentional learning system, which recognize and appraise the feelings and intentions of one's self and of others. In contrast, Self-Transcendence involves the functioning of the autobiographical memory system that allows self-aware (autobiographical) learning and memory functions with a holistic narrative perspective, including mental time-travel, vivid recollection, and imagination of the past and future as a continuity, and identity formation as a creative process.

Table 26–10.
The Relationship between Character Traits, Higher Cognitive Functions, and Brain Structures Based on Voxel-Based Morphometry

Character Dimension	Higher Cognitive Function	Associated Brain Structures
Self-Directedness	Executive functioning Evaluation of internal cues Self-reflection Appraisal of the self and personal intentions	Executive brain network including Anterior cingulate cortex, Posterior cingulate cortex Medial prefrontal cortex Temporal cortex (for self-appraisal)
Cooperativeness	Legislative functioning Empathy Metacognition Appraisal of others' feelings and intentions	Legislative brain network including Medial frontal cortex (white matter) Precentral gyrus (white matter) Superior temporal gyrus (gray matter, BA22)
Self-Transcendence	Judicial functioning Autobiographical recollection Mental time-travel Holistic perspective shifting (time, place, person) Imagination of possible future events in vivid detail in place and time Transpersonal identification	Judicial brain network involves autobiographical memory and learning, including Middle temporal gyrus (greater GMV of BA 21) Inferior parietal gyrus (greater GMV of BA 40) Superior frontal gyrus in medial prefrontal cortex (smaller GMV of BA 10)

Data from Gardini S, Cloninger CR, Venneri A. Individual differences in personality traits reflect structural variance in specific brain regions. *Brain Res Bull.* 2009;79(5):265–270; and Van Schuerbeek P, Baeken C, De Raedt R, De Mey J, Luypaert R. Individual differences in local gray and white matter volumes reflect differences in temperament and character: a voxel-based morphometry study in healthy young females. *Brain Res.* 2011;1371:32–42.

Consequently, it is useful to distinguish three major aspects of human personality: temperament, character, and identity. The characteristics of temperament, character, and identity are summarized in Table 26–11. A person's identity is cultivated and formed in autobiographical memory, which integrates a person's emotional drives (temperament) with their emerging goals and values (character). Identity is the narrative self in the context of a self-aware individual developing over time in a variety of situations. Identity, character, and temperament are continuously influencing one another as personality develops over the lifespan and are also responsive in more or less adaptive ways to changes in life situations and events, including changing social roles with age, marriage, and work, as well as traumatic and therapeutic experiences.

Table 26–11.
Key Differences between Temperament, Character, and Identity

Variable Properties	Temperament	Character	Identity
Awareness level	Automatic	Intentional	Self-aware
Memory form	Percepts Procedures	Concepts Propositions	Intuitions Narratives
Learning principle	Associative conditioning	Logical/conceptual analysis	Automatic Intelligence
Role of subject in mental activity	Passive Reproductive	Active effort and logical construction	Effortless and fluid creativity
Key brain system	Limbic systems Corticostriatal habit system	Semantic memory (frontotemporal neocortex)	Autobiographic/autonoetic memory system
Form of mental representation	Stimulus-response sequences varying additively in strength	Logical hierarchies following from assumptions, prior goals, and values	Narrative schema and interactive networks varying qualitatively in configuration

Genomic Approaches to Understanding Temperament and Character. Personality development, including identity formation, in particular, involves the adaptive development and integration of temperament and character in the changing situations and roles of a person's life. How do temperament and character develop? What are the mechanisms that allow adaptive learning across the lifespan in relation to temperament and character? These basic questions have been matters of speculation in the past, but, as described briefly in the introduction, Zwir, Cloninger, and their associates have made much progress in uncovering the hidden mechanisms by carrying out person-centered Genome-wide Association Studies (GWAS) of temperament and character to learn what genes are associated with temperament only, character only, and the interactions of both temperament and character together. The method of person-centered GWAS allows for gene–gene and gene–environment interactions. Different profiles of temperament and of character were associated with different sets of functionally interacting genes detected through single-nucleotide polymorphisms. The broad heritability of the various traits of personality are all around 50%, but this depends on substantial interactions among multiple genes and environmental variables. The average effects of individual genes explain less than 1% of the variability in personality, but gene–gene interactions and gene–environmental interactions account well for the broad heritability of temperament and character. The whole genome was screened for specific genes using about 1 million single-nucleotide polymorphisms in more than 4,000 people representative of the general populations of Finland, Germany, and Korea. Personality was found to be highly polygenic, with the associated genes distributed across all the human chromosomes. Figure 26–2 shows the number of genes per chromosome that are strongly associated with clusters of people with particular temperament profiles (i.e., temperament only), clusters of people with particular character profiles (i.e., character only), and clusters of people with profiles of both temperament and character traits (i.e., temperament and character). Sets of multiple genes were identified independent of information about the personality traits, and the clusters of people were identified without information about the genetic profiles. The gene sets, therefore, represent naturally occurring combinations of genes with distinct functional interactions. The clusters of personality traits represent attractor states in a complex adaptive system, allowing people to function in a particular situation more or less well, as illustrated in the figures that follow next. The sets of genes and the personality profiles were very strongly associated, but the relationships were complex in the sense that the same gene set could be expressed in multiple ways (i.e., multifinality, pleiotropy), and different sets of genes could be expressed in the same way (i.e., equifinality, heterogeneity).

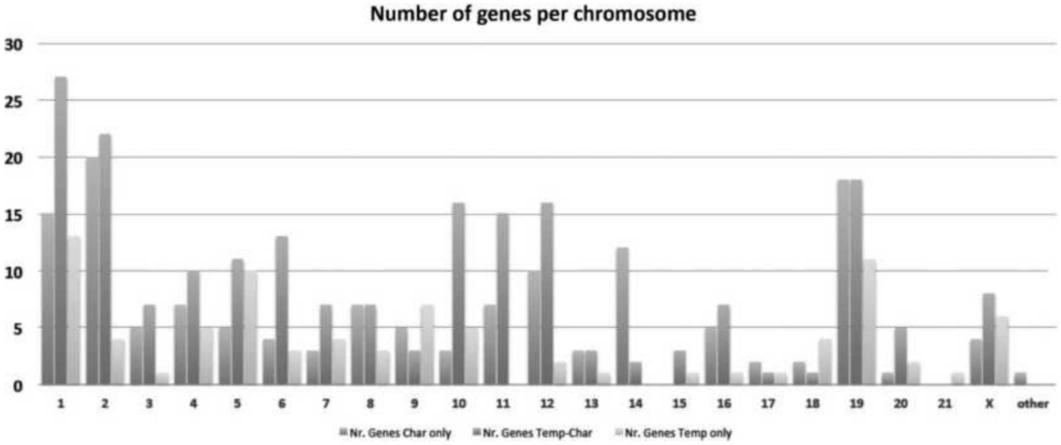

FIGURE 26–2. Number of genes per chromosome strongly associated with temperament profiles only (temp only), character profiles only (char only), and the interaction of both temperament and character (temp and char) in person-centered GWAS of the temperament and character inventory. (Used with permission of the Anthropedia Foundation.)

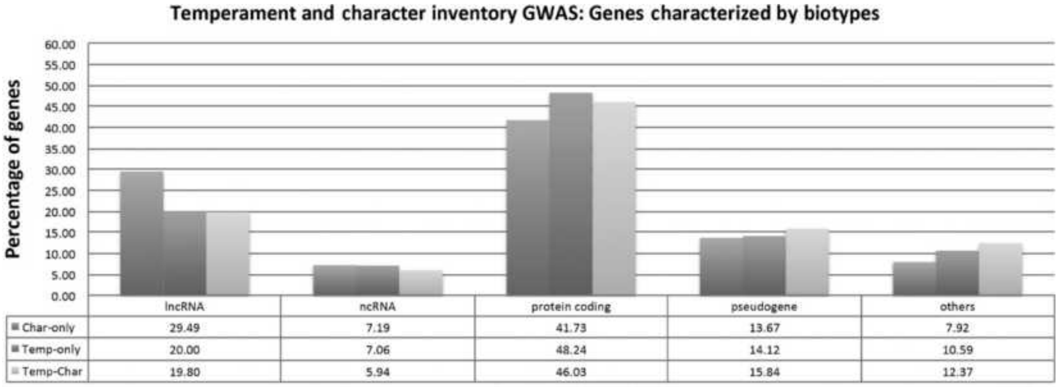

	lncRNA	ncRNA	protein coding	pseudogene	others
Char-only	29.49	7.19	41.73	13.67	7.92
Temp-only	20.00	7.06	48.24	14.12	10.59
Temp-Char	19.80	5.94	46.03	15.84	12.37

FIGURE 26–3. Percentage of different types of genes strongly associated with temperament profiles only (temp only), character profiles only (char only), and the interaction of both temperament and character (temp and char) in three independent samples (Finland, Germany, and Korea). (Used with permission of the Anthropedia Foundation.)

The complex development of personality profiles from interacting sets of multiple genes and environmental variables will be described in more detail later. Now it is fundamentally important to recognize that most of the genes that influence personality have regulatory functions rather than being protein-coding genes (Fig. 26–3). In older genetic studies of personality, nearly all attention has been focused on protein-coding genes related to neurotransmitter function or neurotropin functions like those mentioned for temperament traits in tables. However, it is the regulatory genes that actually help to understand the change and development of human personality. In fact, 95% of the 267 genes that are found in modern Homo sapiens, but not in Neanderthals or chimpanzees are RNA genes that do not encode proteins, as shown in Figure 26–4. These regulatory

RNA genes account for the emergence of human self-awareness, creativity, and prosocial behavior that allowed modern humans to be more adaptable than other hominids. These findings confirm that nearly all differences among Homo sapiens, Neanderthals, and chimpanzees lie in the regulation of the expression of genes that encode proteins.

Most of the regulatory genes unique to modern humans are long non-coding RNAs (lncRNAs), other shorter non-coding RNAs, and pseudogenes. These non-coding RNAs (ncRNAs) are functional RNA molecules that are transcribed from DNA but are not translated into proteins. They can regulate gene expression at both the transcriptional and posttranscriptional levels. They are involved in epigenetic processes that can be divided into two main groups: the short ncRNAs of less than 30 nucleotides and the long ncRNAs of more than 200 nucleotides. The major classes of short ncRNAs are microRNAs (miRNAs), short interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs). Both long and short RNAs are known to play a role in heterochromatin formation, histone modification, DNA methylation targeting, and gene silencing. However, lncRNAs and the shorter ncRNAs act differently.

Nearly 30% of the genes associated with character only involve lncRNAs, whereas the percentage is only 20% in the other two groups, as shown in Figure 26–3. In contrast, protein-coding genes are more often associated with temperament only than with the groups involving character. The more prominent role of lncRNAs in character than for temperament is consistent with the greater importance of epigenetic change for the development and maintenance of all aspects of well-being across the human lifespan. The human capacity for healthy longevity permits humans to have extended periods of learning early in life in order to accumulate knowledge and understanding, which is a remarkable feature of modern humans and their complex social relationships. This genomic finding supports findings from longitudinal studies that character shows much greater change during young adulthood than does temperament. LncRNAs form extensive networks of ribonucleoprotein (RNP) complexes with numerous chromatin regulators. These chromatin regulators target these enzymatic activities to the appropriate locations in the genome. Apparently, lncRNAs can function as modular scaffolds to specify higher-order organization in RNP complexes and in chromatin states. LncRNAs have several functions in the regulation of gene transcription, including the regulation of specific gene transcription (either activation or repression), posttranscriptional regulation, differential spicing, and regulation of epigenetic modifications, including histone and DNA methylation, histone acetylation and sumoylation, and remodeling of chromatin domains. These findings suggest the hypothesis that character development involves epigenetic processes involving the regulation of gene expression and chromatin modification. LncRNA dysfunction is also often observed in common chronic diseases related to stress and lifestyle regulation, including cancer, cardiovascular disease, neurologic diseases, and immune-mediated diseases. Fortunately, people can create their own health by prudent self-regulation of their lifestyle, including diet, exercise, recreation, and social and occupational activities.

In contrast, the shorter ncRNAs are mostly involved in distinct temporal expression patterns during embryogenesis, acting in specific tissues to regulate expression at a posttranscriptional level. MicroRNAs usually bind to a specific target messenger RNA with a complementary sequence to induce cleavage, degradation, or block translation, and they have crucial roles in epigenesis. Silent-interfering RNA functions in ways similar to miRNAs to mediate posttranscriptional gene silencing as a result of messenger RNA degradation and to induce heterochromatin formation. Piwi-interacting RNAs are so named due to their interaction with the piwi family of proteins. They are involved in chromatin regulation and suppression of transposon activity during embryogenesis. The percentage of ncRNA genes associated with personality is only 5% and does not distinguish the temperament clusters from the character clusters.

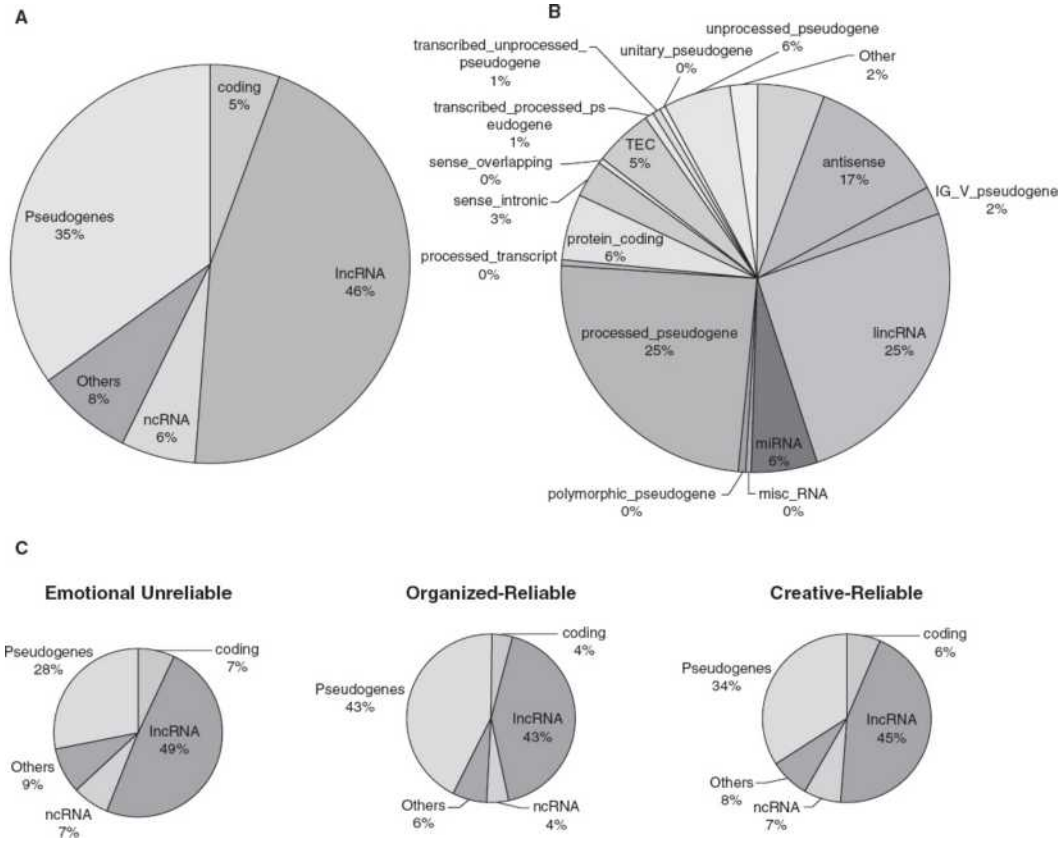

FIGURE 26–4. The types of genes for personality found only in humans. (**A**) Broad gene categories, (**B**) Fine-grained gene categories, and (**C**) Analysis of the genes from (**A**) in the genotypic networks associated with the Emotional-unreliable, Organized-reliable, and Creative-reliable personality networks of modern humans. (Reprinted from Zwir I, Del-Val C, Hintsanen M, et al. Evolution of genetic networks for human creativity. *Mol Psychiatry.* 2021;27:354–376. doi:10.1038/s41380-021-01097-y.)

As shown in Figure 26–3, the distribution of the types of genes associated with character differs from that of other groups. This adds support to findings from longitudinal studies of development and from functional brain imaging that temperament and character are distinct domains of personality with distinct functions in different brain regions. The prominence of regulatory functions for genes associated with personality provides a mechanism for understanding the way learning and memory processes influence personality development as a complex adaptive process. It is clear from the genomic results that personality involves dynamic self-organization that is more or less adaptive depending on both nature and nurture. The convergent findings from neurogenetics and psychosocial development provide a foundation for renewed appreciation of earlier clinical work on personality from diverse fields, including psychodynamics.

Psychodynamic Concept of Character. The psychodynamic concept of character derives from the concept of defense mechanisms. The latter are defined as automatic, unconscious psychological processes (both cognitive and emotional), which protect against anxiety generated by intrapsychic conflicts and external stressors. All defense mechanisms are grouped into three levels, for example, mature (sublimation, anticipation, humor, altruism), neurotic (centered around suppression), and immature (centered around splitting). Psychodynamic concepts (Anna Freud, Otto Fenichel) define mature character traits as residues of previous mature defense mechanisms that have become stable behavior patterns and conscious coping mechanisms. In other words, the mature character reflects one’s capacity to postpone immediate gratification and process internal needs through sublimation, anticipation, altruism, and humor. Behaviors reflecting these particular defense mechanisms correspond partially to the description of Self-Directedness (especially sublimation), Cooperativeness (especially altruism), and Self-Transcendence (especially anticipation and cheerful humor). Thus, psychodynamic theories regarded self-transcendence as a potentially constructive illusion motivated by the demands of the Super-ego to satisfy social requirements. According to psychodynamic theories, the benefits of self-transcendence can be recognized as mature defenses, which develop from unconscious defenses against internal conflicts and later may become conscious coping mechanisms and character traits with mutual benefits for self and others. However, the notion that people are motivated at best by self-interest that may or may not extend to others has never adequately explained the frequent spontaneous display of virtuous behaviors, such as sacrifice for others and other expressions of love, hope, and faith. Today, genetic and neuroimaging research confirms that self-awareness evolved as a third system of human learning in *Homo sapiens* with distinctive molecular and neurobiologic mechanisms that provided humankind with the means to perceive their union with others. As described in the introduction, our inseparability from others is a reality recognized by most people, particularly those who are healthy, happy, kind, responsible, and wise. This ability to be aware of our interconnection with all things involves a perception of a deeper level of reality than is recognized by people who feel the self is separate from other people and things. The failure to recognize our inseparability from others is unfortunate because self-interest is insatiable and ultimately self-defeating; striving for what is unattainable always diminishes health, happiness, and meaning in life.

Psychodynamic theories suggest that the Ego not only protects from but shifts, organizes, and reacts to internal impulses and external stimuli. According to this dualistic perspective, the pattern of the integrative functions attributed to the Ego and the ways in which a person combines various functions to satisfy multiple external and internal pressures constitute a person's character. In other words, the psychodynamic description of Ego-strength corresponds to Self-Directedness, which in the biopsychosocial model of personality is a person's degree of understanding of the self-transcendent virtue of hope. Likewise, in psychodynamic models, Cooperation is motivated by a person seeking mutual benefit to reduce tensions arising from social demands, whereas in the biopsychosocial model, cooperation is a person's degree of understanding of love. Similarly, Self-transcendence is a person's understanding of faith, which Freud considered an illusion, apparently at least in part, because of his own disappointment with the failures of his highly religious father.

Psychodynamic theories describe two forms of character disorders: neurotic character and character neurosis. Neurotic character traits are postulated to derive from neurotic defenses (e.g., suppression, reaction formation, projection, repression, undoing), which have dissociated from their original conflict and have become inflexible, pervasive, and ego-syntonic traits of everyday behavior. Anna Freud called this inflexibility "the armor plating" of character. As extensions of previous conflicts between one's emotions and opposing social pressures, neurotic character traits protect a person from being involved in such situations again. For example, excessive scrupulousness, a frequently observed character trait of obsessive personalities, is a "reaction formation" against their usually strong aggressive impulses. By being rigidly hyperscrupulous, obsessive persons are supposed to protect themselves from conscious experience of aggression and from intrapsychic conflicts that would have been triggered by such an experience. In contrast, according to the biopsychosocial model of personality, such neurotic character traits arise because of inadequate self-actualization (i.e., the development of self-awareness sufficient to integrate a person's habits with their goals and values). Put another way, neurotic character traits are the result of weak self-regulation of temperament by character, which occurs when there has been impaired development of self-awareness due to genetic predisposition and/or environmental influences.

Character neurosis is observed when an inflexible neurotic character trait interferes with healthy parts of the personality. For example, character neurosis is present when excessive cleanness (an ego-syntonic character trait observed with obsessive individuals) frequently interferes with the need to interact freely with others (generated by healthy parts of obsessive personality); instead of being perceived as natural, cleanness becomes so excessive that it is perceived as frustrating. According to the biopsychosocial model of personality, character neurosis occurs when extreme temperament drives inhibit the healthy development of character, as is particularly common with extreme anxiety-proneness (i.e., high Harm Avoidance) and impulsivity-aggressivity (i.e., high Novelty Seeking) in combination with reinforcing rearing experiences, such as parental abuse and neglect. Of note, character neurosis has been included in the *International Classification of Diseases* (ICD)*,* 10th revision, but not in ICD-11, DSM-IV, or -5.

Individual differences in the maturity of defense mechanisms (i.e., normal, neurotic, immature) tend to covary with the severity of psychopathology (i.e., normality, anxiety disorders, PD, respectively). Put another way, the level of a person’s well-being and overall personality development mainly reflects the development of their character (i.e., how high are they on all three character traits of Self-directedness, Cooperativeness, and Self-transcendence). In addition, the subtype of PD depends on their temperament profile, so patients with PD, who in general have immature defenses, also tend to “favor” certain immature mechanisms that express their particular temperament profile and its influence on their character profile, which together, in turn, shape their self-awareness, identity formation, behavioral style, and clinical presentation. These dominant defenses are, for example, the projection for paranoid PD, acting out for antisocial PD, fantasy for schizoid PD, and so forth. Note, however, that in addition to these favorite defenses, PDs also use more general immature defenses that contribute to their irresponsibility, such as their tendency to blame others and act opportunistically in a selfish manner. Reflecting these and similar observations about the importance of defense mechanisms for the overall understanding of deviant personality, the DSM-IV classification enabled clinicians to classify the use of a particular, dominant defense mechanism on Axis II, but the multiaxial system and description of defense mechanisms have been omitted in DSM-5 as psychiatrists operating in managed care settings have increasingly limited the practice to pharmacotherapy without a systematic approach to psychotherapy.

Psyche and Identity

Psychology is literally the study of the psyche, but this word is frequently misunderstood. Psyche refers to a person’s self-aware consciousness, intuitive insight, or spirit. Intuitive insight, such as the inspiration derived from introspection or contemplation, is described by some as silently listening to one’s inner voice, soul, or intuitive Reason. Plato described the process of listening to intuitive reason as recollection or auto-noesis. The contemporary philosopher David Chalmers distinguishes “Mind 1” (i.e., the analytical mind, which is algorithmic and hierarchical in form and uses convergent logic to specific conclusions based on initial assumptions) from “Mind 2” (i.e., the psyche or thought aspect of the soul, which is intuitive and holographic in form and uses intuitive insight and divergent reasoning to explore alternative perspectives and future outcomes as a creative narrative that tolerates ambiguity). Cognitive scientist Endel Tulving showed that the episodic learning system of apes is the foundation for “auto-noetic learning” in modern humans which enables the unique human capacities for self-awareness described in the introduction of this chapter, including human capacities for creativity, narrative art, and language, science, and spirituality. More precisely, in measurable scientific terms, individual variability in the psyche can be measured in terms of differences in a person’s level of intuitive self-awareness, which is observable in terms of the scope of a person’s identity, particularly their degree of self-transcendence. The word “identity” is derived from the Latin *identitas,* which is based on the Latin word *idem* for “same.” The material contents of a person’s body change over time, but a person’s identity has a nonmaterial basis that can be the same throughout the narrative of a person’s life and can even extend beyond the local reality of their self at a particular time and place. Intuitive self-awareness is unique to human beings, but humans do show variation in their level of self-awareness in ways that have a strong influence on susceptibility to PDs or the capacity for wisdom and well-being. In Figure 26–1, the psyche involves the unique human process of symbolic invention, which is based on intuitive self-awareness, leading to the human invention of art, science, and spirituality. The growth of self-awareness is crucial to the development of full coherence of personality, which is manifest as creativity, well-being, and wisdom. The scope of a person’s identity can be quantified in terms of the extent to which a person views their self as a separate localized object or as a self-transcendent being through the process of transpersonal identification.

Psychobiology of the Psyche. Detailed studies of human learning and memory have distinguished procedural, semantic, and episodic memory. As described in relation to temperament, procedural memory underlies associative learning and involves presemantic perceptual processing of information from the physical senses that can operate independently of abstract conceptual and/or volitional processes. In contrast, intuitive learning is based on immediate recognition from a single episode of observation. One type of memory based on immediate recognition is knowledge of facts, which has been called “semantic memory” or “long-term declarative memory,” as discussed in relation to the registration of cognitive schema underlying personality traits. The second type of recognition memory is awareness of one’s own intentions in a spatiotemporal context, which has been called “episodic memory.” In other words, episodic memory involves self-awareness and recall of events in a context that gives personal meaning to the “when” and “where” of life experiences. Initially, this type of recognition memory was defined in terms of materials and tasks, such as intentions (i.e., one’s self in a spatiotemporal context). Subsequently, the concept of semantic memory has been elaborated in terms of intentionality, that is, not only learning of facts but also abstract reasoning, propositions, and intentions about goals and object relationships. Likewise, episodic memory has been refined and elaborated in terms of awareness of one’s self, the sense of subjective time, and a conscious state of self-awareness of the continuity in one’s past, present, and future, which is different from dreaming or imagination. Tulving now describes episodic memory as self-awareness or “autonoetic consciousness,” which is consistent with other descriptions of ordinary human self-awareness as based on immediate recognition by the intuitive senses. Episodic memory is also called “recollection” to distinguish it from factual knowledge that has no personal context in space and time. Tulving notes that there is no evidence that such intuitive self-awareness is present in any animal species except human beings.

Procedural, semantic/intentional, and episodic/self-aware memories can be functionally dissociated from one another. For example, individuals with Parkinson disorder, characterized by striatal lesions, exhibit deficits in procedural learning but not in semantic or episodic memory. Individuals with an amnestic syndrome, characterized by lesions in the medial temporal lobe, have deficits in semantic learning of new facts but not in procedural learning or episodic consciousness. Individuals with amnestic syndrome have anterograde amnesia for new facts but may have no deficit in their immediate sense of self-awareness. Bilateral lesions in the human hippocampus lead to deficits in self-awareness (i.e., episodic consciousness or the intuitive recognition of one’s self in a spatiotemporal context). Episodic memory is subserved by a widely distributed network of cortical and subcortical brain regions that overlaps with, but also extends beyond, the networks subserving other memory systems. Key regions in this network include the hippocampus, regions of the prefrontal cortex (including the frontal poles and medial prefrontal cortex), and the anterior cingulate cortex.

Essentially, the human episodic memory system acts like a mirror that provides spontaneous self-awareness to humans, whereas other primates require the provision of an external mirror to even recognize themselves. This mirroring function depends on hemispheric specialization in the encoding and retrieval of information. The left prefrontal cortex is more involved in encoding episodic memories and retrieval of semantic memories, whereas the right prefrontal cortex is more involved in the retrieval of episodic memories. The episodic retrieval mode is a state of consciousness (in which a person is actively recollecting or remembering their past) that involves the strong activation of sites in the right prefrontal cortex and the weak activation of sites in the left prefrontal cortex and medial anterior cingulate. The ability to recall past events in vivid detail is closely related to the ability to imagine future events in future detail. Hence, the human capacity for recollection and imagination is a component of self-aware narrative learning. People who have been abused or traumatized are often unable to recall events in their life around the time of the trauma and are likely to be unable to imagine their future with sufficient detail to function adaptively in creating a satisfying way of living.

The experimental evidence shows that human beings have an intuitive form of learning and memory as the basis for their unique capacity for self-awareness is now strong, but its significance has yet to be widely appreciated. Current evidence suggests that the hippocampus and other parts of the episodic memory system, such as the prefrontal cortex, are necessary for encoding episodic memories but not for the acquisition of factual knowledge. The hippocampus processes all sensory information available to a person about both one's external and one's internal milieu. Moreover, this pan-sensory information is constantly updated, allowing the registration and recall of ongoing life experiences in a personal context that is unique to each individual. The hippocampus is also known to maintain its plasticity throughout life, including the ongoing generation of new neurons as it adapts to ongoing experience. In fact, genes associated with impaired neurotrophin functions that promote healthy neuronal formation in the hippocampus (e.g., *NTRK3*) are associated with disturbances of self-awareness, as is seen in PDs and the schizophrenias. In contrast, long-term declarative memories, such as factual knowledge and self-reported character traits, are stored in brain networks that show greater stability and less plasticity than the hippocampus.

The plasticity of the self-awareness learning system has important clinical implications for the treatment of PD. Individuals in whom self-awareness is well developed are characterized by a creative character profile in which Self-transcendence is highly developed along with Self-directedness and Cooperativeness. In contrast, individuals in whom the intentional learning system is dominant are characterized by an organized character profile in which Self-directedness and Cooperativeness are highly developed but Self-Transcendence is not. Self-Transcendence is the strongest predictor of the rate of growth in personality in prospective studies, presumably because it is an indicator of both the ability and willingness to change and the ability to evaluate conduct and theories about intangible forces in life, such as love, hope, and faith. In contrast, people who are low in Self-transcendence are less flexible because they are guided by traditions and fixed beliefs. The character strengths of people with organized characters are merely static concepts, which are abstractions from which the spatiotemporal context has been reduced to a factual statement by a fixed choice that eliminates the intentional flexibility inherent in self-aware consciousness. They depend largely on the operation of the intentionality brain network for executive control, symbolization, and storage of facts, which depends at a molecular level on the calcium-second messenger system processing information hierarchically within neurons. Essentially, self-aware consciousness is “living” with many possible outcomes that can be voluntarily influenced, whereas factual knowledge is “dead” or “frozen” like a collapsed quantum wave function. In contrast to the intentionality brain network, the network for creative self-awareness, as described by Cloninger in 2004, depends on distinctive molecular mechanisms that give rise to human consciousness with the quantum-like properties as expected in a scale-free dynamic network in which all components are interdependent and function optimally in coordination. These findings about self-aware consciousness suggest a possible explanation of the importance of intersubjective awareness in the development of self-transcendence and of the substantial resistance to change of character and related cognitive schema when cognitive-behavioral or psychodynamic strategies are used. The plasticity of the episodic memory system and its role in self-awareness also suggests a crucial role for growth in self-awareness as a means of experiential transformation of personality.

Development and Assessment of the Psyche and Identity. Self-awareness is a recently evolved system of learning and memory that appears to be unique to human beings, just as components of the network, such as the prefrontal cortex, are recent evolutionary developments. As expected for a recently evolved function, self-aware learning develops later than procedural or intentional learning. Children younger than 4 years do not yet have a mature, self-aware memory system, so people generally have no direct recollection of their past before 4 years of age. In contrast, basic emotions are present at birth, and self-object differentiation occurs between 18 months and 3 years of age.

Cloninger has observed five distinct levels of identity resulting from varying degrees of intuitive awareness in human beings. There is a hierarchy of levels of awareness that can be measured in terms of what are sometimes called innate human ideas or intuitions about being and human nature. Assessment requires an empathic understanding of the way another person is conscious of being, freedom of will, beauty, truth, and goodness, which provide markers of a hierarchy of levels of awareness. These five aspects of consciousness depend on one’s intuitive awareness of the world. The awareness of facts and intentions are only two of the empirically observable levels of awareness.

For example, some patients with severe PDs complain of “emptiness,” which is a fearful feeling of isolation, separateness, annihilation, or lack of being. Such patients lack a stable awareness of their being, so they are sometimes described as having a borderline level of personality organization. This is characterized by difficulty in sublimation, intolerance of distress, distortions of reality testing, and poor self-acceptance. Their unstable sense of being (i.e., lack of spiritual fullness, identity, and vitality) makes them so emotionally unstable and hopeless that they often mutilate themselves or attempt suicide. About 1% of people are deficient in their intuitive sense of being and experience life with a borderline organization chronically. Another 1% of people experience such disorganization and emptiness intermittently under stress. In contrast, the vast majority of people have a well-developed awareness of being all the time. This level of awareness involves memory of existence with a stable sense of being, that is, without fear of annihilation. This level of awareness is based on the recognition and memory of “facts,” or what has traditionally been called “semantic memory.”

Second, other people usually have hopeful confidence in their being but feel that they are slaves to extrinsic influences and trapped in time like a hell or prison from which they have no means of escape. In other words, such people have a lack of flexibility in their actions or little freedom of will. The lack of awareness of freedom of intention (i.e., will) also distorts the sense of a free flow in time. The lack of consciousness of free will in time has been described as “chronesthesia” (mental time travel) in patients who have brain lesions in their episodic memory system. Even in people with no brain lesions, the lack of awareness of free will can be pervasive. For example, some people are fatalistic materialists and claim that freedom of will is an illusion or something in which they are lacking. Their own intuition contradicts that of humanists and dualists who feel that rational responsibility and freedom of will are what gives human beings their dignity. About 10% of people experience a lack of freedom and the urgency of time commitments to external influences chronically throughout their lives. This corresponds closely to traditional definitions of PD, which involve a lack of freedom or flexibility in adaptive behavior. In contrast, most people have a stable intuitive sense of their freedom and responsibility for their intentions. More generally, this level of awareness involves a person’s awareness of their intentions in a personal spatiotemporal context. This is the second of Tulving’s types of intuition-based types of memories, but it does not exhaust the levels of human awareness.

Third, some people have no esthetic sensibility. That is, some human beings have no intuitive sense of what is lovely or beautiful. Cloninger has observed that those who had attained each level always had an awareness of the lower levels but not the higher levels of consciousness. For example, those with an awareness of beauty also had an awareness of being, freedom, but not necessarily an awareness of truth or goodness. In other words, these indicators of consciousness define a hierarchy of levels of awareness. It is shocking to recognize that some patients with PDs have no recollection at any time in their lives of feeling a sense of wonder and admiration for the beauty of any poem, painting, or nature scenes! Try to understand what that means for the quality of their lives. When such unaware individuals see a rainbow, they do not feel it is beautiful or wonderful! They just see physical colors and shapes with no conscious recognition of beauty. The complete absence of any sense of love and beauty is rarely acknowledged, occurring in only about 3% of individuals, but many more show little intensity in their appreciation of beauty. Although they know what opinions are fashionable, it has little or no impact on their ordinary awareness. Individuals with an unstable sense of beauty are consequently unstable in their capacity for love and emotional intimacy in relationships. Approximately half of marriages in the United States end in divorce, which is an objective indicator of instability in the sense of what is beautiful and loved. In contrast, the stable intuitive sense of what is beautiful and loved is characteristic of the other half of the general population, who have sometimes been designated as “romantics.” In particular, esthetic sensitivity is a highly developed intuitive sense in romantic poets and novelists, such as Emerson and Thoreau, and in romantic musicians, such as Schubert.

Fourth, still other individuals have a sense of being, freedom, and beauty at least some times in their lives but never have any intuition of absolute truth. This is typical of individuals who describe themselves as pragmatists or objectivists. Such individuals also are usually agnostic or atheists because they have no intuition of their participation in a unity of consciousness, which gives rise to faith. About 10% of individuals in the United States are atheists or agnostics, which is consistent with recent nationwide Gallup polls. According to Gallup polls (2017), about 9 of 10 (87%) Americans believe in God, but the majority do not feel that institutional religions are adequately inspiring and satisfying their spiritual needs. Eight percent of Americans do not believe in God but believe in a universal spirit or higher power. Only 5% of people are atheists who also deny belief in a universal spirit or higher power. However, it is important to note that many people who describe themselves as atheists or agnostics actually have a well-developed spiritual sensitivity, as has been described by David Hay. In fact, some atheists may be more spiritual in their dedication to truth and tolerance than people who regularly attend religious services out of custom and convention.

Individuals who have attained each level usually have had an awareness of the lower levels but not the higher levels of consciousness. For example, individuals with a sense of emptiness also lacked awareness of freedom, beauty, truth, and goodness. Those with an awareness of truth also had an awareness of being, freedom, and beauty but not necessarily an awareness of goodness. Hence, awareness of truth is an intuitive sense that is well developed in elevated levels of consciousness. The intuitive sense of truth is prominent in many scientists, such as Teilhard de Chardin and Carl Rogers. It is even more fully developed in highly self-aware sages, such as Hegel, Leonardo da Vinci, and Gandhi, who viewed their lives as experiments in truth.

Fifth, still other individuals may have had a sense of being, freedom, beauty, and truth at times in their lives but had never had any intuition (i.e., direct awareness) of the intrinsic goodness of all things despite the perversions and corruption that exist in the world. Unfortunately, this group includes nearly everyone at times when they wonder about the existence and nature of evil or seek retribution and revenge against other human beings who threaten or attack us or perhaps only frustrate us. Deep awareness of the universal unity of being and consciousness is required to love one's enemies. The intuitive sense of goodness is well developed only rarely. There is a partial awareness of it in individuals who are highly developed in all three character traits, including Self-Transcendence, such as the American transcendentalists such as Emerson and Thoreau. However, a well-developed awareness of goodness occurs only rarely, as in wise people such as Plato and Gandhi.

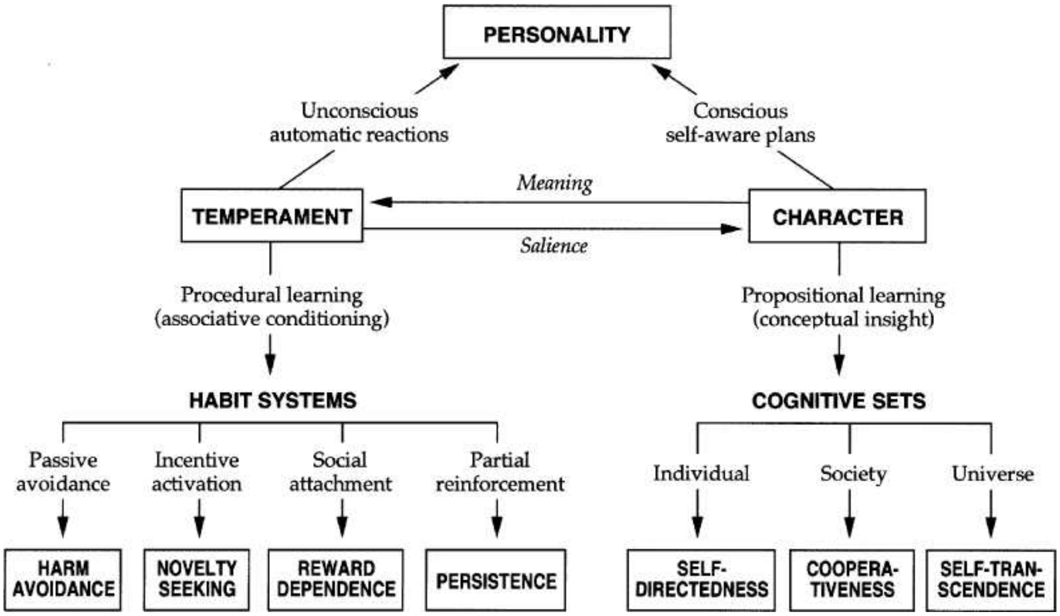
PERSONALITY
Unconscious automatic reactions
Conscious self-aware plans
Meaning
TEMPERAMENT
CHARACTER
Salience
Procedural learning (associative conditioning)
Propositional learning (conceptual insight)
HABIT SYSTEMS
COGNITIVE SETS
Passive avoidance
Incentive activation
Social attachment
Partial reinforcement
Individual
Society
Universe
HARM AVOIDANCE
NOVELTY SEEKING
REWARD DEPENDENCE
PERSISTENCE
SELF-DIRECTEDNESS
COOPERA-TIVENESS
SELF-TRAN-SCENDENCE

FIGURE 26–5. Personality: An integrated psychobiological schema.

Finally, it is important for clinicians to realize that character development is modular and stepwise, and there can be blind spots or blockages in any particular module. As a result, there are infrequent exceptions to the rule that high levels of awareness imply awareness at all the lower levels. Such exceptions are particularly striking in high-functioning patients, such as alcoholic physicians and other professionals. Such patients may have mature character development overall but have specific blind spots in their insight and judgment that lead to serious impairment in particular aspects of their lives, such as romantic relationships or various forms of impulse control.

PERSONALITY: AN INTEGRATED VIEW

Recent reports about complex and fundamental processes underlying neurophysiology, development, and phenomenology of personality provide powerful guidelines for the formulation of a comprehensive biopsychosocial model of personality as the coherent organization of the human body, mind, and spirit. The model integrates earlier phenotypic observations about behavior with contemporary concepts based on neuroimaging, the biology of behavior, genetics, psychometric advances, and nonlinear modeling of normal and deviant personality development. Personality is conceptualized as a complex adaptive system involving a multidimensional interaction among temperament, character, and psyche. The coherent maturation of personality requires awareness of the natural unity of being, which, in turn, depends on the interplay of three systems of learning and memory for behavioral conditioning of habits and skills, intentional learning of facts, goals, and logical propositions, and self-aware insight and appraisal of values and theories (Fig. 26–5).

Through the interaction of temperament and character, different aspects of meta-cognition of the self and the external world modify the significance and the salience of sensory percepts and affects regulated by temperament and vice versa. In other words, temperament regulates what one notices and, in turn, character modifies its meaning so that the salience and significance of all experience depend on both one's temperament and character. In turn, the development of character is derivative of individual differences in intuitive self-awareness. Essentially, personality development depends on the joint interactions among three systems of learning and memory. These are the systems for behavioral conditioning, intentionality, and self-awareness, which are related to temperament, character, and identity, respectively. Reliable measures of the functioning of these three systems are provided by temperament (procedural learning of emotional drives), Self-Directedness and Cooperativeness (intentional self-control of goals and object relationships), and Self-Transcendence (self-aware learning in autobiographical memory with a creative imagination and theory of mind). The least plasticity is in temperament (behavioral conditioning), and the most plasticity is in self-awareness, with intentionality intermediate to the other two in its malleability.

Consequently, treatment efficacy depends substantially on whether treatment is focused on temperament, character, or identity. Somatic therapies, medications, and behavioral conditioning target the temperament, which usually changes little beyond specifically conditioned situations. Cognitive and psychodynamic approaches seldom eliminate or transform maladaptive cognitive schemas (i.e., transform intentional goals and propositions) but may improve the management of conflicts by means of prolonged treatment of more than a year. Only treatment focused on intuitive awareness is likely to lead to an actual cure of PD by the transformation of the level of intuitive awareness (i.e., insight and judgment). Randomized controlled trials now do show the effectiveness of such triple-process approaches.

Any effective treatment has to allow the patient to be ready to see the world more coherently. Extreme temperament traits make PD patients inflexible, biased, and reactive. These “armor-like” features make it difficult for these patients to elevate their self-awareness so that they can see or feel or intuitively recognize the world of beauty and truth that seems “hidden” to them. Everyone has become able to recognize how beautiful the truth is only after they are able to achieve a level of personal acceptance. Medication and cognitive therapies can help a person to be “ready” for higher levels of maturity, but the transformation always involves a change in perspective. Therapists must be flexible and alert to opportunities to help patients see through the fog of their fears and other negative emotions, which cloud their perspective. The key to change is often quite unique and unexpected. Accordingly, there is no recipe or algorithm by which individuals always progress, and effective therapy requires ingenuity and flexibility because of the uniqueness of each person’s strengths and weaknesses.

ESSENTIALS OF PERSONALITY ASSESSMENT IN CLINICAL PRACTICE

The assessment of personality can be made during routine assessment combining information from mental status and the lifetime psychiatric history. Integrating this information means that the clinician must recognize and understand the invariance of the organization of human thought regardless of timescale: mental status observations span milliseconds (e.g., a transient smile, fear, or anger) to minutes (e.g., perceptual awareness, fluidity of speech, resourcefulness, and speed in problem solving) whereas the lifetime history spans months (e.g., episodes of depression, gambling, substance abuse) to decades (e.g., repeated divorces, achievements). Human thought is dynamic with an average of 10 different thoughts per second, involving sudden shifts in attention, emotion, and intention, but it varies in ways that can also be observed to develop over the entire lifespan of the same individual. To be able to describe and understand this variability coherently, it is helpful for a clinician to have an integrated model of the structure of thought and personality.

Human personality has five layers or “planes of being,” according to Cloninger. The five layers of personality are concerned with human adaptations in situations that are perceived to be predominantly concerned with reproduction and sexuality (“sexual plane”), practical everyday activities concerned with power and possessions (“material plane”), emotional bonds and social attachments (“emotional plane”), communication and culture (“intellectual plane”), and understanding what is beyond individual human existence (“spiritual plane”). The brain systems that make it possible for human beings to adopt these different perspectives have evolved in a hierarchy of discrete steps over time to allow the survival of progressively more flexible, aware, and creative organisms.

The development of well-being and the regulation of potential conflicts among emotional drives in self-aware consciousness involves three branches of mental self-government. These branches of self-government correspond to the higher cognitive processes related to the three character traits and should be systematically assessed during the mental status examination. The assessment of insight and judgment is often trivialized to a subjective judgment about whether patients agree with their doctor ("insight") and do what they are told to do ("judgment"). It is more useful to evaluate insight as a combination of reasoning ability (self-directedness) and intuition (self-transcendence), which corresponds well to concepts of insight as rational intuition that date back at least to the writings of Plato. Judgment has come to mean something like the legislative function measured by Cooperativeness, providing a capacity for empathy, teamwork, and principled action. However, the rating of these processes can be made much more informative and precise by recognizing how the three character traits are expressed in different situations. Each of these three cognitive functions has modules for adapting to each type of situation or layer of personality, which can be measured reliably by specific facets of character.

The three branches of mental self-government and their components are summarized in Table 26–12, along with the specific subscales of the Temperament and Character Inventory (TCI) designed to measure these functions. Self-Directedness in the TCI measures the executive functions, which can be considered the strength component of insight. It is expressed in various situations, including a person being self-directed in the sexual plane (i.e., resourceful), in the material plane (i.e., purposeful), in the emotional plane (i.e., self-accepting), in the intellectual plane (i.e., responsible), and in the spiritual plane (i.e., self-actualizing with few bad habits that interfere with achieving one's goals). On mental status examination, a person may be impartial and realistic in their thinking, high in self-esteem, and resourceful and far-sighted in actualizing their plans if they are highly self-directed, or they may be emotionally vulnerable, biased and unrealistic in their thinking, low in self-esteem, and short-sighted in planning. Remember that the rating of this one characteristic (executive functioning expressed as self-directedness) reveals more about the level of the person's maturity, their vulnerability to psychopathology, and their prognosis for recovery with a positive quality of life than any other feature of the mental status. Self-directedness is the expression of a life guided by a sense of hope that allows a realistic acceptance of reality as a path to understanding the wonders and mysteries of life, even if they appear inconsistent with past expectations. Self-directedness is not a fixed trait—it can be temporarily reduced during stress, trauma, intoxication, or psychosis, and it can be enhanced by psychotherapy or life experiences such as opportunities to gain in self-directedness by one's accomplishments from responsible work. Everyone can grow in self-directedness under appropriate therapeutic conditions, such as reduced stress, nonviolence, and sobriety, especially when encouraged by the hope and compassion of others.

Cooperativeness measures judgment as the legislative function, which involves making rules that allow people to get along in a reasonable and flexible manner with one another in each plane of their life. Cooperativeness can be recognized in the sexual plane as tolerance instead of prejudice, in the material plane as being helpful instead of hostile, in the emotional plane as empathy instead of self-centeredness, in the intellectual plane as being principled instead of opportunistic, and in the spiritual plane as a sense of compassion and fairness instead of vengefulness. In brief, good judgment is shown by a person's flexibility and ability to get along with others with values guided by compassion and principle. Poor judgment is shown by prejudice, lack of impulse control, insecure or hostile attachments, difficulty with teamwork, and opportunism. Comparison of a person's interaction during mental status examination with their life history is important to understand the range of cooperativeness they have demonstrated at different times during their life. For example, a person who is usually cooperative may temporarily become inflexible and hostile under stress or the aftermath of violent trauma. Alternatively, a person who has been uncooperative in the past may develop better judgment. A person may also have a blind spot in their awareness that elicits uncooperativeness in a recurrent but specific type of situation but allows good judgment in most other situations.

Table 26–12.
Descriptions of Three Branches of Mental Self-Government (Executive, Legislative, and Judicial Cognitive Functions) That Regulate the Emotional Drives Relevant to Each of the Five Layers of Human Personality, Which Are Defined by the Predominant Focus of the Person's Perspective in a Situation[a]

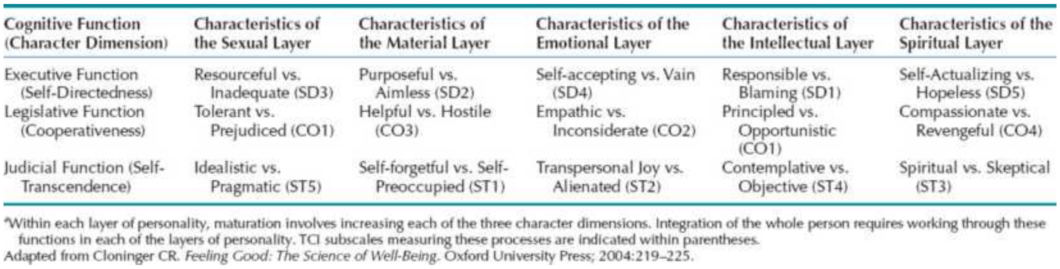

Cognitive Function (Character Dimension)	Characteristics of the Sexual Layer	Characteristics of the Material Layer	Characteristics of the Emotional Layer	Characteristics of the Intellectual Layer	Characteristics of the Spiritual Layer
Executive Function (Self-Directedness)	Resourceful vs. Inadequate (SD3)	Purposeful vs. Aimless (SD2)	Self-accepting vs. Vain (SD4)	Responsible vs. Blaming (SD1)	Self-Actualizing vs. Hopeless (SD5)
Legislative Function (Cooperativeness)	Tolerant vs. Prejudiced (CO1)	Helpful vs. Hostile (CO3)	Empathic vs. Inconsiderate (CO2)	Principled vs. Opportunistic (CO1)	Compassionate vs. Revengeful (CO4)
Judicial Function (Self-Transcendence)	Idealistic vs. Pragmatic (ST5)	Self-forgetful vs. Self-Preoccupied (ST1)	Transpersonal Joy vs. Alienated (ST2)	Contemplative vs. Objective (ST4)	Spiritual vs. Skeptical (ST3)

[a] Within each layer of personality, maturation involves increasing each of the three character dimensions. Integration of the whole person requires working through these functions in each of the layers of personality. TCI subscales measuring these processes are indicated within parentheses.
Adapted from Cloninger CR. *Feeling Good: The Science of Well-Being*. Oxford University Press; 2004:219–225.

Self-Transcendence measures the depth component of insight. Depth of insight allows people to know intuitively when legislative rules apply in a particular situation. The depth of a person's insight is shown in the sexual plane by a person's being idealistic rather than selfishly pragmatic. Depth of insight is shown in the material plane by being self-forgetfully absorbed in action rather than self-preoccupied. In the emotional plane, insight is shown by awareness of one's transpersonal connections with nature and other people, which leads to a sense of abundance, joyful exuberance, and being altruistic and gracious. In the intellectual plane, depth of insight is shown by being contemplative rather than materialistically objective. In the spiritual plane, insight is shown by a sense of connectedness to what is divine or beyond human existence, expressed as faith or spiritual acceptance rather than skepticism or nominalism. In general, poor depth of insight is shown by the shallowness of awareness or magical ideation, whereas the depth of insight is shown by wisdom, creativity, faith, and depth of understanding. In most people, the depth of insight is indicated by the frequency of peak experiences of transpersonal union, inseparable connectedness with nature, or oceanic feelings. However, reports of such peak experiences are sometimes the product of magical (schizotypal) thinking in people who are very low in self-directedness. Hence, it is important to evaluate both the strength and depth components of insight (i.e., both self-directedness and self-transcendence).

All personality modules involve person-by-situation interactions that are regulated by a set of dynamical nonlinear systems, which allows human beings to be purposeful, flexible, and self-aware in their adaptation to life. *Human personality is not adequately characterized as a set of linear traits because the components of personality are nonlinear in their functional effects and relationships with one another.* Human thought involves the frequent rapid adaptation of a person to different situations or different perspectives on the same experience. The connections between different brain modules change suddenly and synchronously with changes in thought.

The full three-dimensional structure of personality can be visualized as a three-dimensional spiral with five planes and increasing in height, width, and depth from the sexual to the spiritual plane. The height of the spiral indicates a person’s strength of insight, as measured by their level of TCI Self-Directedness, which indicates a person’s strength to face and adapt to reality rather than to blame, deny, or repress unpleasant facts. The width of the spiral indicates a person’s broadness of social concern and flexibility of judgment, as measured by their level of TCI Cooperativeness. The depth of the spiral indicates a person’s depth or shallowness of insight, as measured by their level of TCI Self-Transcendence. Essentially, a person’s thoughts can shift rapidly throughout their possible range, spiraling down with increasing negative emotions and other indicators of ill health when stressed (e.g., preoccupied with fears, threats, or desires) or spiraling up with increasing positive emotions and other indicators of well-being when calm and secure (e.g., feeling hopeful, kind, and aware). A psychiatrist is often presented with a patient in a state of distress in which their thoughts have fallen to a low point in the range. The first step in treatment is to create conditions that allow the patient’s thoughts to ascend in the spiral of self-aware consciousness through validation, encouragement, medication, and mental or breathing exercises that allow a person to regain their more hopeful, kind, and coherent perspective toward life. Later work can help understand and reduce their vulnerabilities for relapse and recurrence.

Although human thought is highly dynamic with a wide range, one can use averages or modes of one's thoughts and relationships as central indicators of states of being that are metastable, as is typical of equilibrium states in nonlinear dynamical systems. Frequent metastable states of being are often described as personality traits, even though they are neither fixed nor linear in their functions and relationships. Accordingly, personality is a way of describing the dynamic processes that occur within the person to shape and adapt to life experiences by learning and planning. The maturation and integration of human personality involves one's growing in self-awareness through experiences across this wide range of situations.

As a result, personality is a complex expression of nonlinear interactions among a whole hierarchy of learning systems that develop over time as a complex adaptive process. Each personality module has three aspects, each regulated by dissociable but overlapping systems of learning and memory: a somatic or emotional aspect regulated by learning of habits and skills, a cognitive aspect regulated by learning of facts and propositions, and a spiritual aspect regulated by self-aware learning of one's continuity with and participation in a whole greater than oneself. These three systems can be dissociated but normally interact through their extensive interconnections. For example, a person can know how to do something (using procedural memory) but not know the fact that he knows the skill (a deficit in semantic memory). A person can know a particular fact and report it to others (using semantic memory) but not have a memory of when and where they learned the fact (a deficit in episodic memory). The systems evolve and mature in sequence but operate simultaneously from early in childhood. As a result, the same symptom (e.g., feeling rejected) can be the result of different vulnerabilities (e.g., temperaments with high Reward Dependence, character with low Cooperativeness, poor intrapsychic integration leading to low well-being). Therefore, the effective clinician must look beyond presenting symptoms to understand and treat the underlying processes that produce symptoms, which is not possible with categorical diagnoses or multidimensional systems designed to compare and judge one person against another. Nevertheless, it is possible to use information about clinical syndromes to help identify the underlying processes that are most likely.

Table 26–13 lists the personality traits that are most consistently associated with different clinical disorders and states of mental health. This is not meant to be an exhaustive listing of all the personality correlates of different mental disorders but only the traits that are so strongly related to different disorders that a clinician can use the information in the differential diagnostic process. For example, high Harm Avoidance is commonly associated with most forms of psychopathology, but it has a particularly strong relationship as a cause of mood and anxiety disorders.

Therefore, high Harm Avoidance should be suspected as a predisposing trait in mood and anxiety disorders. The presence of Harm Avoidance can be quickly evaluated by inquiring about child and adolescent traits such as being a worrier, pessimist, or shy. Likewise, high Novelty Seeking should be suspected if a person is alcoholic, drug abuser, gambler, or bulimic. Novelty Seeking can be quickly evaluated by inquiring about a person having a quick temper in their youth, being impulsive, extravagant, or rule-breaking, instead of slow-tempered and liking order and structure in what they do. On the contrary, mature character development confers well-being and resilience to psychopathology. In addition, positive character development facilitates recovery from mental illness, promoting the development of a sense of coherence, personal mastery, self-esteem, and life satisfaction. Thus, the distinction between temperament traits, which are value-neutral, and character traits, whose development can facilitate recovery and well-being, provides a perspective that is encouraging and not pejorative.

Table 26–13.
Personality Traits[a] That Are Highly Likely to Be High (+) or Low (–) in Different Mental Disorders and States of Mental Health

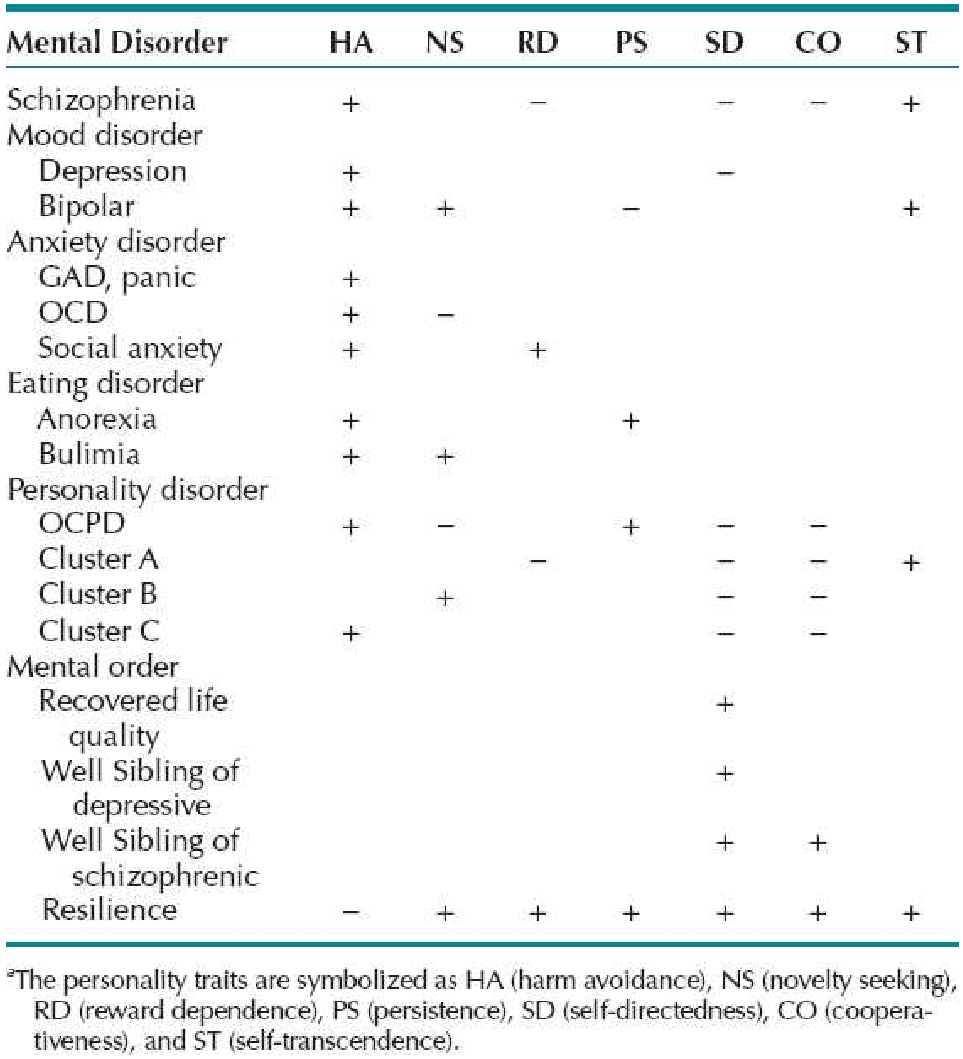

Mental Disorder	HA	NS	RD	PS	SD	CO	ST
Schizophrenia	+		−		−	−	+
Mood disorder							
Depression	+				−		
Bipolar	+	+		−			+
Anxiety disorder							
GAD, panic	+						
OCD	+	−					
Social anxiety	+		+				
Eating disorder							
Anorexia	+			+			
Bulimia	+	+					
Personality disorder							
OCPD	+	−		+	−	−	
Cluster A			−		−	−	+
Cluster B		+			−	−	
Cluster C	+				−	−	
Mental order							
Recovered life quality					+		
Well Sibling of depressive					+		
Well Sibling of schizophrenic					+	+	
Resilience	−	+	+	+	+	+	+

[a] The personality traits are symbolized as HA (harm avoidance), NS (novelty seeking), RD (reward dependence), PS (persistence), SD (self-directedness), CO (cooperativeness), and ST (self-transcendence).

CONCEPTUAL AND PSYCHOMETRIC ISSUES

Personality Traits: Person versus Situation Debate

Personality traits are neuropsychic structures with the capacity to render many stimuli functionally equivalent and to initiate and guide equivalent (consistent) forms of adaptive and expressive behavior. Accordingly, the DSM defines personality traits as “enduring patterns of perceiving, relating to, and thinking about oneself and the environment,” that is, other people and the world as a whole. The major value of personality traits, therefore, lies in their usefulness in identifying predictable regularities in an individual’s behavior.

The stability of personality traits in time and across situations has been the central issue in personality theory for decades, as it bears direct relevance to the question of whether internal dispositions or external situations determine behavior. The "person vs. situation" debate was initiated in the early 1900s by two groups of scientists ("situationists" vs. "personologists"); the former viewed behavior as highly situationally specific, the latter as centrally organized and purposive. As noted by Epstein and O'Brien, the dilemma remained unresolved for decades because it was not realized that behavior can be situationally specific at the item level and cross-situationally general at the aggregate level. Single items of behavior have limited reliability and generality. They must be aggregated over situations and occasions in order to reveal broad and stable traits, which allow moderately accurate predictions of behavior tendencies (without having to specify the eliciting situations) but only weak predictions of behavioral acts in particular situations.

On the contrary, global personality traits also do not explain behavior completely because people do not exhibit unmodulated consistencies in behavior across time and situations (complete invariance in behavior is associated more with psychopathology than with normality). Biogenetic factors influence how an individual adapts to experience, and the latter modifies adaptive tendencies. Hence, personality traits are not fixed. This current understanding of traits differs from the conservative one of 25 years ago. The emerging consensus is an interactionist position allowing room for both situational (external) and dispositional (internal) determinants of behavior in a variety of complex combinations, as described in Tables 26–11 and 26–12 for temperament and character. Fortunately, it is now possible to specify facets of personality traits that are relevant to particular situations and also to aggregate these facets into aggregate scales that are moderately predictive across a broad range of situations. In comparisons among many popular personality models, the TCI predicts behavior and clinical disorders as well or better than other models.

Phenotypic and Developmental Trait Personality Models

The major difficulty in relating observed variation in personality to its underlying biologic processes is that the observed phenotypic structure that distinguishes one person from another differs from the underlying psychobiological organization of adaptive processes that occur within the individual. Factor analysis describes only the differences between people, not the number or content of the processes that occur within individuals as they learn to adapt in time. Hence, the *phenotypic trait personality models,* which are usually factor analytically derived (e.g., Eysenck's tridimensional model, Costa and McCrae's Five-Factor Model), account for much of the observed phenotypic variance but may not correspond to the underlying intrapsychic processes. For example, Neuroticism is a composite of both high Harm Avoidance and low Self-Directedness. However, Harm Avoidance and Self-Directedness have unique genetic determinants and are regulated by different brain circuits. For example, Harm Avoidance, and not Self-Directedness, is moderately correlated with Mismatch Negativity in evoked potentials, which is a measure of behavioral inhibition related to the orienting response, automatic attention, and early unconscious sensory processing. Furthermore, Self-Directedness, but not Harm Avoidance, is strongly correlated with activity with the late-evoked potential P300 and with activation of the medial prefrontal cortex in fMRI studies during executive tasks. This double dissociation of Harm Avoidance and Self-Directedness with early- and late-occurring evoked potentials is consistent with the phylogenetic history of Harm Avoidance evolving early and Self-Directedness evolving late.

In contrast, *developmental trait personality models* take into account both the underlying biologic dispositions to observable behaviors and individual differences in responses to experience during personality development. The most recognized dimensional models are those of Sjobring, Zuckerman, and Cloninger. Despite some limitations, Sjobring's developmental model in the late 1940s is a heuristic attempt to take personality research beyond the level of phenotypic observations and factor analysis. The model defines three dimensions of latent variation predisposing to personality: "validity" (i.e., the degree of energy available in adaptation to experience), "solidity" (the consistency of adaptation), and "stability" (referring to the maximum potential of a person to develop skillful habits in adapting to routine). These three genetic dispositions were postulated to be independent of one another and from the factor of general intelligence (referred to as "capacity") and also to define one's susceptibility to behavior disorders. Validity is similar in description to low Harm Avoidance, Solidity to low Novelty Seeking, and Stability to low Reward Dependence.

Personality has often been described in terms of a linear sequence of qualitatively discrete developmental stages and structural types. However, the frequent occurrence of heterotypic individuals with intermediate or mixed features calls into question the basic assumption that personality structures are limited to a finite number of types that are qualitatively discrete and homogeneous. Likewise, it is becoming increasingly clear that there is not only variation in the rate and direction of personality development as a whole but also asynchrony and inconsistency among its major components that are reorganized through time. The following section outlines a more general nonlinear quantitative theory of personality development that does not make the restrictive assumptions about a finite number of qualitatively discrete structural types.

PERSONALITY DEVELOPMENT: A SELF-ORGANIZING BIOPSYCHOSOCIAL COMPLEX

The development of normal and abnormal personality traits can be quantified as a complex dynamic system that is self-organizing and partly molded by familial and sociocultural influences. In order to formulate a realistic model, Svrakic and colleagues (1996) had to allow for the reciprocal interactions among multiple and etiologically distinct components of personality that account for its complex patterns of development across a person's lifespan. This biopsychosocial model also accounts for the frequent, but not invariant, development of stage-like periods of moderately stable personality configurations (similar to personality types) punctuated by abrupt transitions in which there is structural reorganization and emergence of qualitatively new features. The model is based on a sophisticated mathematical framework that has been implemented and tested with prospective empirical data. More recently, the developmental interactions among temperament, character, and identity in terms of interactions among the systems of learning for behavioral conditioning, intentionality, and self-awareness have been tested and interpreted.

The central ideas behind this model are the following:

- Personality is a multidimensional dynamical system comprising more elementary operating components (traits) that are organized in an interdependent way critical to carrying out a particular function. Such a system is characterized by particular rules of operation, for example, the principles of associative learning within the temperament function, long-term intentional learning within the system for self-control of goals and relationships, and the development of insight by self-awareness in identity formation. The satisfaction of such multiple constraints results in nonlinear dynamics, which is characteristic of all systems involving growth and development in biology, neuroscience, psychology, and sociology. Such multidimensional dynamical systems are usually called complex adaptive systems.
- The correlations observed among multiple personality traits can be used to explain the spontaneous organization of stable multidimensional configurations (i.e., personality types); these types develop in a stage-like fashion with successive periods of prolonged stability punctuated by rapid transitions. However, the progression of each individual is unique because differences in intuitive awareness result in substantial variability that is unique to each individual.
- The model accounts for functional interactions among multiple types of influence, including genetic factors, family environment, sociocultural norms, and experiences unique to each individual.

- Based on the specified basic differences between temperament, character, and identity (Table 26–10), this model takes into account all aspects of information processing, from associative conditioning of habits and skills to enculturation about goals and values within a unique biopsychosocial context.

Personality development is presented as a walk on an adaptive (or fitness) landscape with two or more hills (representing high adaptive values) separated by valleys (representing low adaptive values). Fitness is defined as the ability to produce a change in personality. In general, a complex system subjected to constraints responds to these constraints by optimizing fitness, that is, by adaptive changes in personality. As change in personality is motivated by optimization of fitness, people will most likely and most rapidly move in the direction of the greatest increase in adaptive value (i.e., the nearest hill). Once they have reached the peak of the hill, they will stay there for a relatively long time because they would first have to decrease fitness (i.e., descend into the valley) in order to find a hill with a higher peak (i.e., better adaptation). Such adaptive development is called U-shaped.

As noted, temperament and character traits are etiologically distinct but functionally related. Such an interactive system has the property of being self-organizing as a result of the collective dynamics among its multiple components. In other words, the organism is spontaneously driven to find patterns of behavior that result in the coherence of all information about both the external and the internal milieu. For example, Harm Avoidance interacts with both Novelty Seeking and Reward Dependence, inhibiting approach to novel stimuli and inhibiting social attachments by increasing fear of the unfamiliar and sensitivity to social criticism. Accordingly, consistent negative correlations of Harm Avoidance with Novelty Seeking and Reward Dependence have been found. In contrast, Novelty Seeking and Reward Dependence are positively synergistic, facilitating sociability and seeking social approval. Such interactions influence the stability of each of the eight possible multidimensional configurations or “profiles” of temperament traits (Fig. 26–6), making some more stable than others.

In the fitness landscape, the valleys (or the adaptive minima) are unstable because even small perturbations, such as random events, will drive the system away from these points of the lowest fitness. In contrast, hills (i.e., adaptive maxima) are stable because they act as “attractors” for all lower points in the neighborhood. Consequently, personality interactions tend to self-organize into such an attractor state and to remain in that configuration in the absence of external pressure or maturational processes. For the understanding of both the chronicity and treatment resistance of PDs, it is of critical importance to realize that a particular personality configuration may be highly stable, even though it may not be the most adaptive behavior possible for that individual (to be discussed).

We are mainly concerned with patterns of change of character traits as a function of relatively stable temperament traits and variation in the level of self-awareness, as discussed later. The three character traits interact to produce eight possible character configurations or profiles (Fig. 26–7).

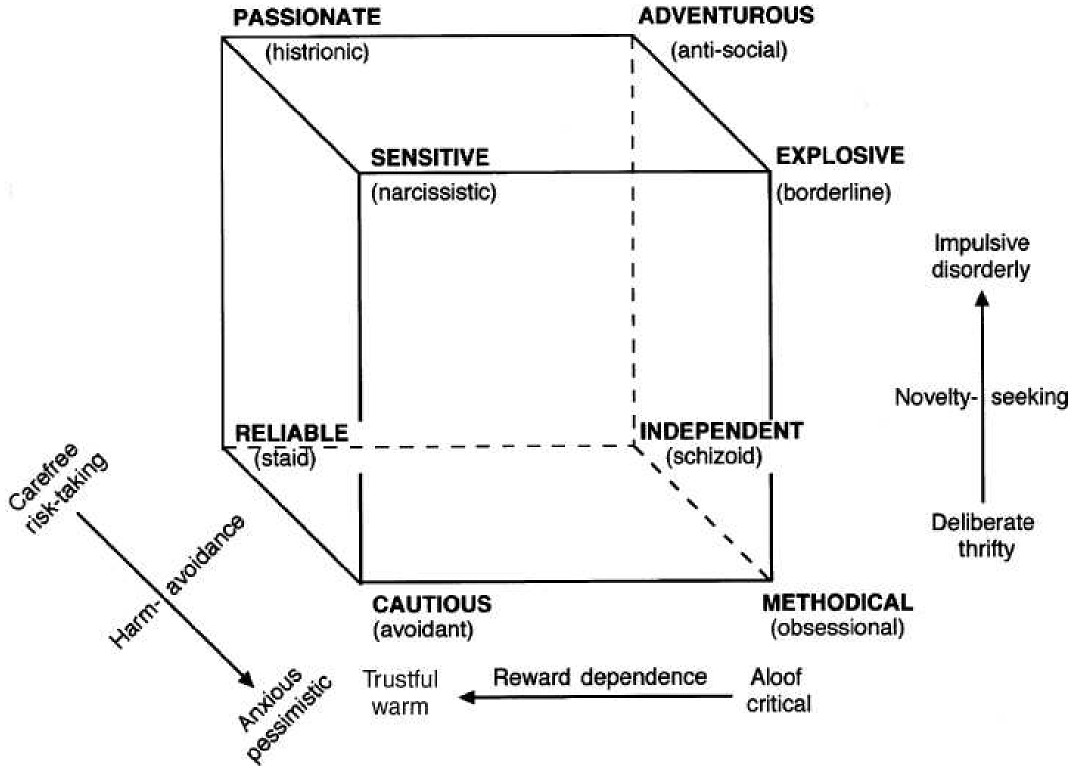
PASSIONATE
(histrionic)
ADVENTUROUS
(anti-social)
SENSITIVE
(narcissistic)
EXPLOSIVE
(borderline)
Impulsive
disorderly
Novelty- seeking
Deliberate
thrifty
RELIABLE
(staid)
INDEPENDENT
(schizoid)
Carefree
risk-taking
Harm- avoidance
CAUTIOUS
(avoidant)
METHODICAL
(obsessional)
Anxious
pessimistic
Trustful
warm
Reward dependence
Aloof
critical

FIGURE 26–6. Temperament configurations.

Any change in these character configurations can be predicted from the initial temperament configuration taking into account temperament–character interactions and the effects of sociocultural norms, random events unique to the individual, social learning within the family, and genetic factors. However, whether there is change depends on self-awareness, which is unique to each individual and unpredictable, that is, underdetermined or free. All these sources of individual differences influence the likelihood for different stable character configurations in the outcome. The sociocultural norms can be additionally accounted for as the systematic bias in development associated with demographic variables, such as subculture, gender, race, age, and occupation. These norms and social education in the family create environments that are supportive or oppressive of specific character development. With all this taken into account, it turns out that a single initial temperament configuration may lead to several different stable character configurations; this aspect of development is referred to as multifinality. These results are graphically represented in

Figures 26–8 through 26–10, with different initial temperament configurations and their possible character outcomes.

Three dimensions of character (Self-Directedness—SD, Cooperativeness—CO, and Self-Transcendence—ST) are drawn along three orthogonal axes, and each character configuration is represented by a point in this abstract space. Eight possible high–low combinations of three character dimensions are represented by the corners of the cube. Observe that certain temperament types may and do give rise to several possible stable character outcomes, each with a certain probability, a feature characteristic of nonlinear systems. The size of the bulging at a particular character configuration is proportional to the probability of that outcome. Figures 26–8 through 26–10 illustrate this for adventurous (Hi NS, Lo HA, Lo RD), explosive (Hi NS, Hi HA, Lo RD), and sensitive (Hi NS, Hi HA, Hi RD) temperament types. Similar results can be obtained for the remaining five temperament configurations.

Individuals are born with one temperament profile with several possible character outcomes. This multifinality reflects the fact that the fitness landscape in personality development has both objective and subjective components. The sense of identity, feelings, and intentions unique to the individual do influence the evaluation of the “objectively” worthwhile goals and values. The observed discrepancies between the natural directions influenced by antecedent temperament traits and actually achieved character configurations point to the importance of external events such as social learning and random fluctuations in the final character outcome.

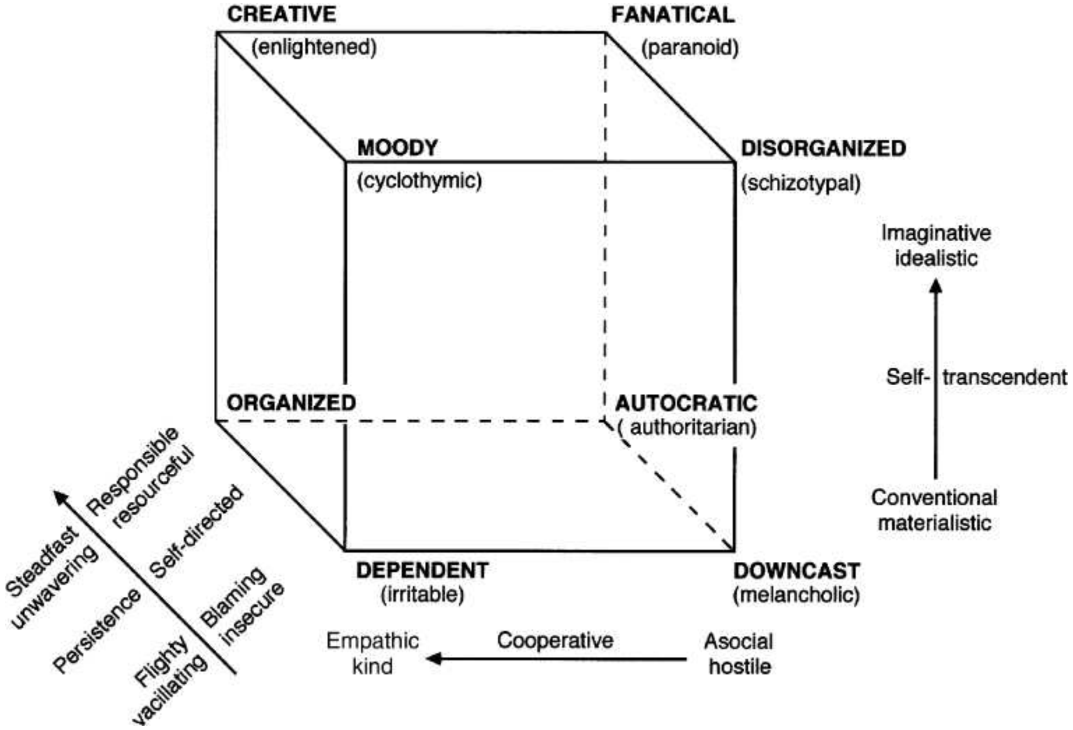
CREATIVE
(enlightened)
FANATICAL
(paranoid)
MOODY
(cyclothymic)
DISORGANIZED
(schizotypal)
Imaginative
idealistic
Self- transcendent
Conventional
materialistic
ORGANIZED
AUTOCRATIC
(authoritarian)
Responsible
resourceful
Self-directed
Blaming
insecure
Steadfast
unwavering
Persistence
Flighty
vacillating
DEPENDENT
(irritable)
DOWNCAST
(melancholic)
Empathic
kind
Cooperative
Asocial
hostile

FIGURE 26–7. Character configurations.

As a result of the presence of points of high- and low-adaptive values in the fitness landscape, personality development is characterized by periods of relative stability alternating with more rapid transitions to new adaptive levels. The subscale structure of the TCI was formulated to specify character in terms of 15 component steps of its development, as shown in Table 26–14.

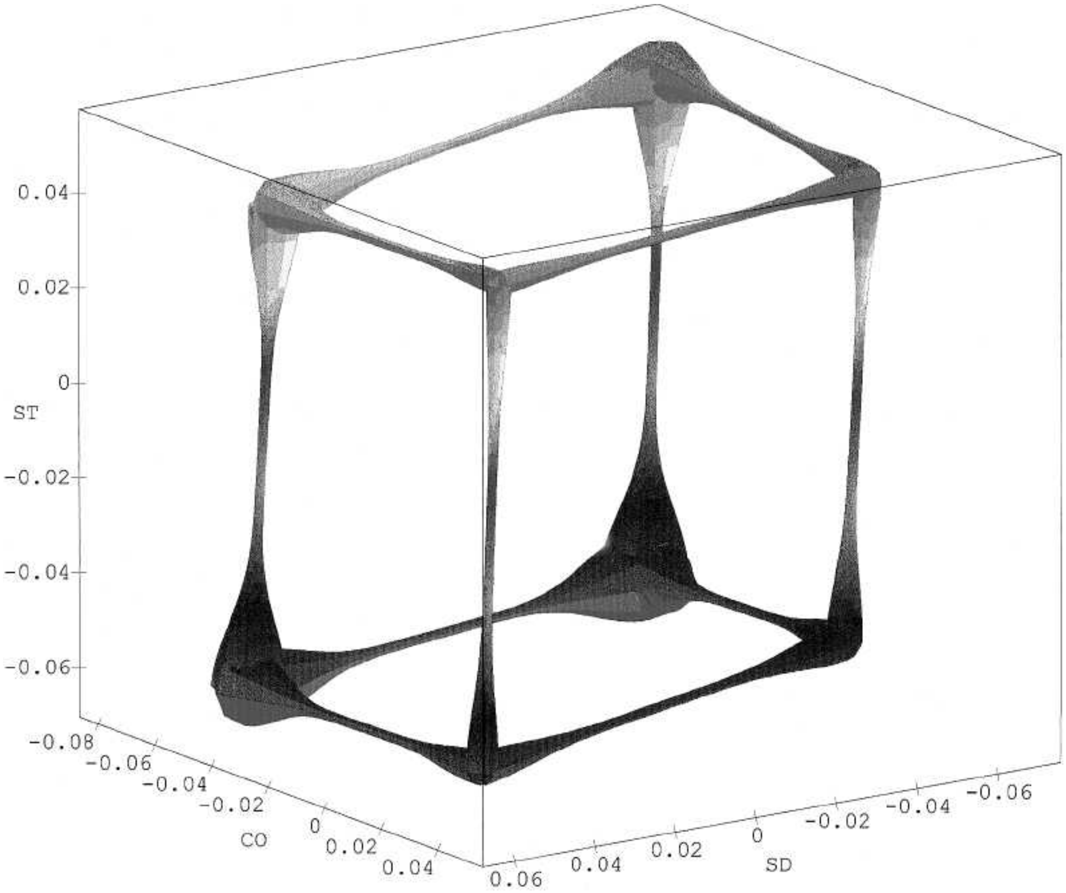

FIGURE 26–8. Most probable character outcomes for the adventurous antecedent temperament traits. CO, cooperativeness; SD, self-directedness; ST, self-transcendence.

The 15 steps in character development are a theoretical ideal, corresponding to the modal pathway that leads to full character development with high scores on all three character dimensions, which has been found to be optimal for healthy longevity, satisfaction, and meaning in life. This 15-step developmental sequence is consistent with prior qualitative descriptions of developmental stages by Piaget, Freud, and Erikson (Table 26–14) but allows for the actual nonlinearity in development that depends uniquely on each individual's identity. It also corresponds to recent findings on the stage-like development of intersubjectivity through reciprocal interactions among subjects. For example, imitation and mirroring of infants with their caretakers give rise to a basic sense of social connectedness and mutual acknowledgment with others. The first stage of intersubjectivity, after 2 months of age, involves the initial signs of intersubjective reciprocity and trust rather than automatic imitation. Next, there is the emergence of the sense of freedom in choices, such as joint attention to material objects. With the maturation of autobiographical memory around 4 years of age, there is an emergence of values that are jointly represented and negotiated with others.

By age 7 years, there is an initial capacity for allocentric perspective-taking that provides a basis for developing more self-transcendent values such as kindness and fairness. The conditions for cultivating intersubjective awareness correspond closely to the common factors for health promotion identified in person-centered therapeutics, as discussed in the introduction. More work is needed to synthesize the early work on personality development with recent work in neurobiology and the psychosocial sciences that recognizes the importance of interactions among multiple complex systems of learning and memory in humans.

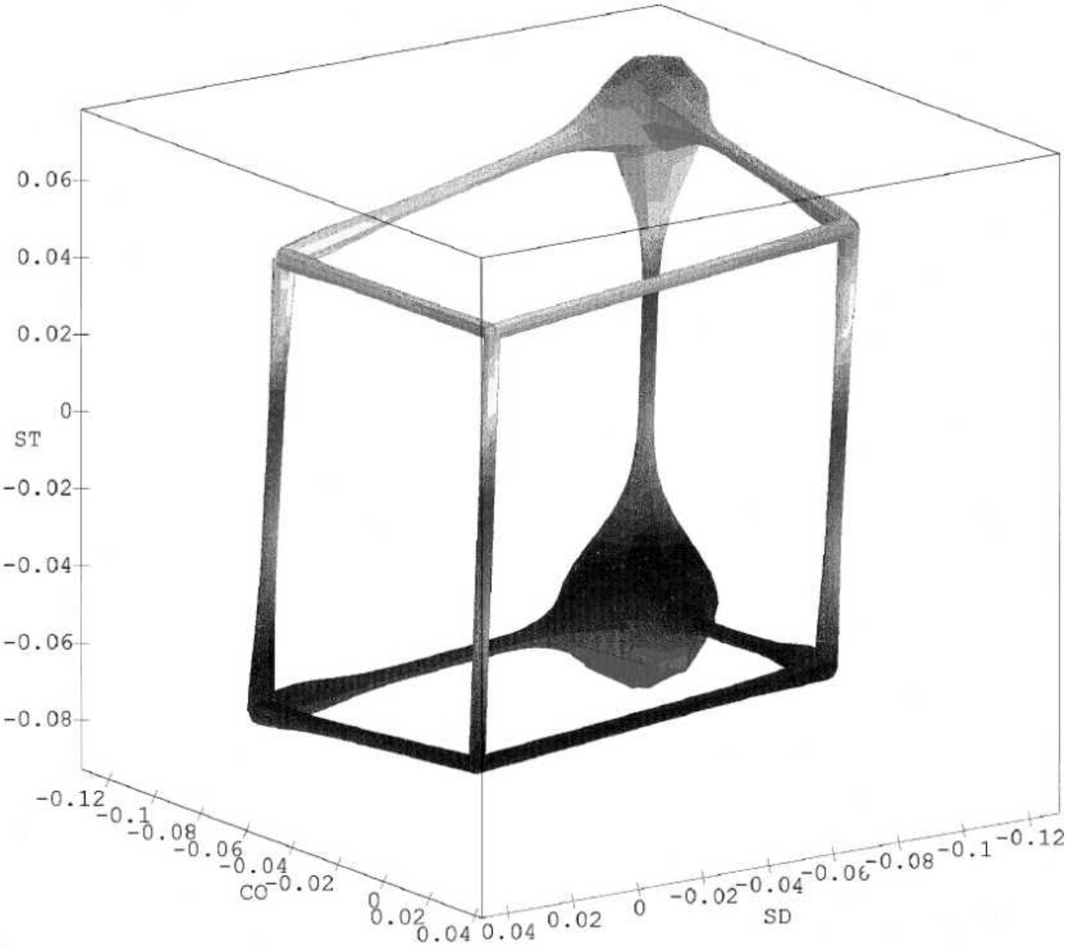

FIGURE 26–9. Most probable character outcomes for the explosive (borderline) antecedent temperament traits. CO, cooperativeness; SD, self-directedness; ST, self-transcendence.

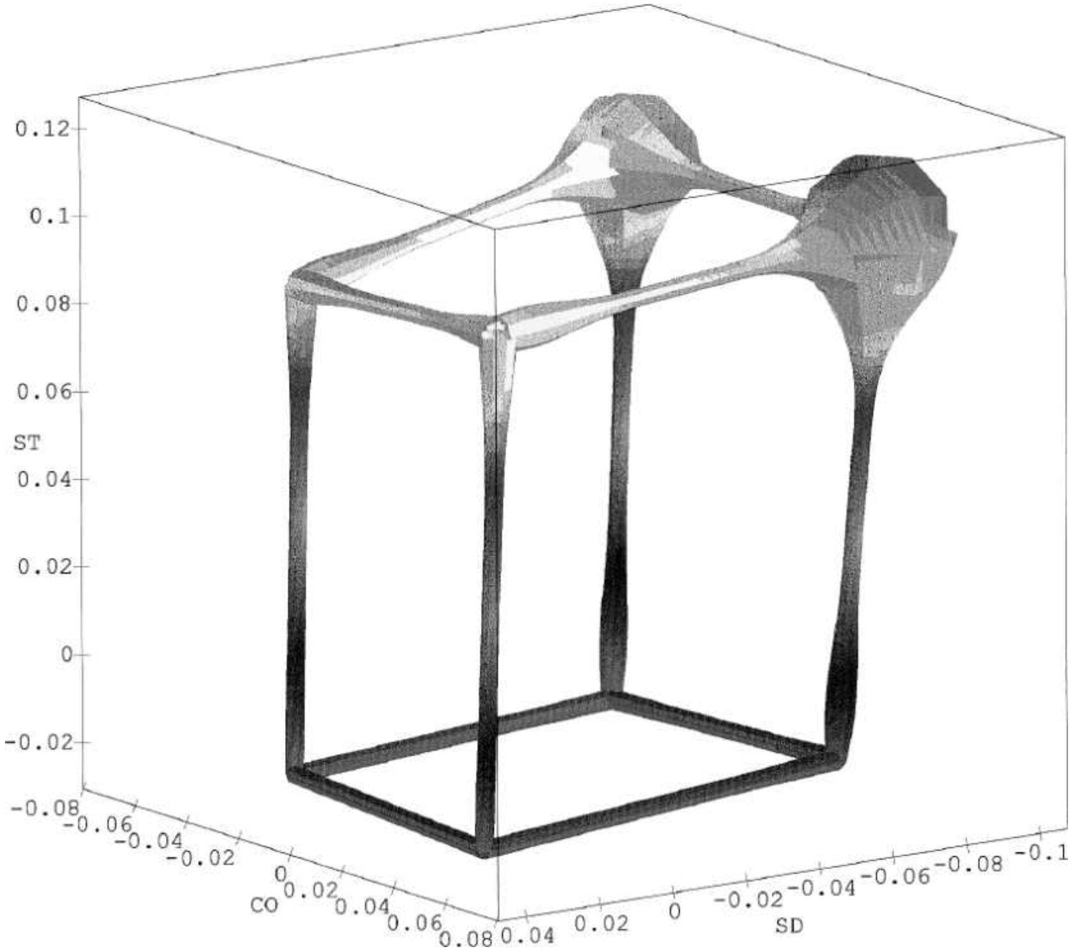

FIGURE 26–10. Most probable character outcomes for the sensitive antecedent temperament traits. CO, cooperativeness; SD, self-directedness; ST, self-transcendence.

Table 26–14. Comparison of Different Descriptions of Character Development and Its Spiral Path

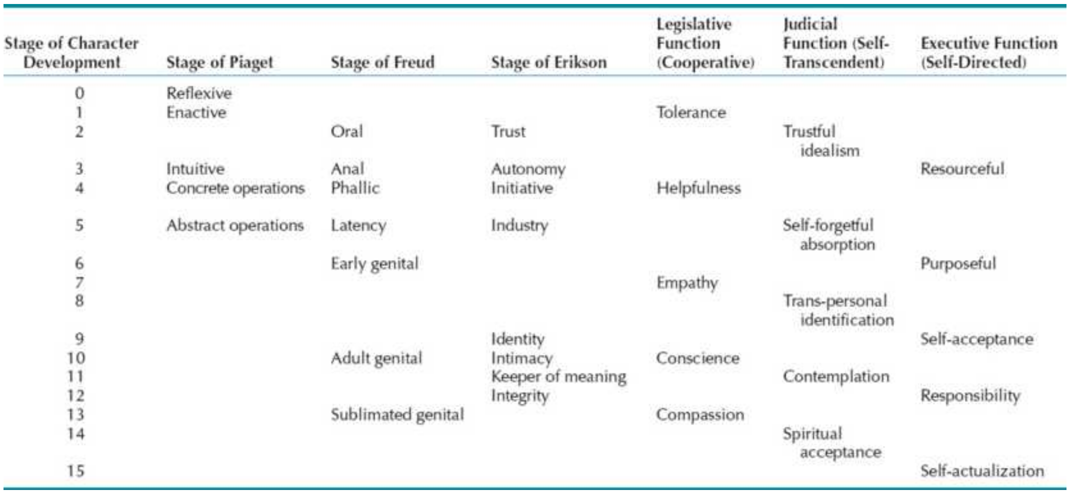

Stage of Character Development	Stage of Piaget	Stage of Freud	Stage of Erikson	Legislative Function (Cooperative)	Judicial Function (Self-Transcendent)	Executive Function (Self-Directed)
0	Reflexive					
1	Enactive			Tolerance		
2		Oral	Trust		Trustful idealism	
3	Intuitive	Anal	Autonomy			Resourceful
4	Concrete operations	Phallic	Initiative	Helpfulness		
5	Abstract operations	Latency	Industry		Self-forgetful absorption	
6		Early genital				Purposeful
7				Empathy		
8					Trans-personal identification	
9			Identity			Self-acceptance
10		Adult genital	Intimacy	Conscience		
11			Keeper of meaning		Contemplation	
12			Integrity			Responsibility
13		Sublimated genital		Compassion		
14					Spiritual acceptance	
15						Self-actualization

A person's identity can extend beyond the present to influence the trajectory of the development along a life project that can be creative if they develop strengths in all three dimensions of character or self-defeating if they do not. Observe that successive steps in character developmental form a spiral pattern, where each revolution around the spiral (presented as each of the five TCI tiers in Table 26–14) introduces a new set of developmental issues associated with the new set of component facets of the three major character traits. Furthermore, each revolution of the spiral depends on different levels in a hierarchy of intuitive awareness, as described earlier. For example, in early character development, during the first revolution around the spiral, a person encounters problems of trust versus mistrust, self-respect versus shame, and moderation versus indulgence; these are associated with Tolerance or Social Acceptance (facet 1 for Cooperativeness), Resourcefulness (facet 1 for Self-Directedness), and Idealism (facet 1 for Self-Transcendence). This depends substantially on a person's intuitive sense of being, whereas later developmental tasks depend on the intuitive senses of freedom and still later beauty, truth, and goodness. As can be inferred from Table 26–14, as they move along the spiral, people successively face new developmental tasks associated with new component facets in Self-Directedness, Cooperativeness, and Self-Transcendence. This spiral pattern of character development provides an opportunity to correct "developmental errors" that were made at any of the previous steps within the same trait. For example, as character develops around the spiral to reach a new step of Self-Directedness (e.g., responsibility), one gets "in line" historically with issues of self-acceptance, purposefulness, and resourcefulness (also steps of Self-Directedness but encountered during earlier revolutions) (Table 26–14). This alignment facilitates retrospective revisions of maladaptive responses to challenges that occurred at these earlier developmental levels. The possibility of "retrospective healing" of old character problems is important in planning the psychotherapy of patients with PDs. Namely, when addressing problems related to a certain developmental step, these patients are more susceptible to changes related to previous steps within the same character traits. In other words, learning about character in any one of the five planes of being can facilitate growth in self-awareness of similar processes in the other planes of being. This process of increasing maturation and integration of character continues throughout life because of the intrinsic human drive for coherence that is expressed as hope, love, and joy. Fortunately, vivid recollections in self-awareness allow healing by accepting the reality of what was given in experience and simultaneously allowing better understanding and imagination of creative solutions in the present and future.

Computer simulations of the self-organized development of character have been carried out, beginning with average initial temperament traits and taking into account all of the above contributing factors. These simulations found that children first increase in Self-Transcendence (i.e., become imaginative and enjoy fantasy), then increase in cooperativeness (i.e., become conforming and rule-based), and only later increase in self-directedness (i.e., behavior becomes increasingly self-reliant and autonomous). This prediction is consistent with the description of the stages of Ego development by Loevinger and her colleagues. This early sequence does not exclude the subsequent further development of other character dimensions in response to demands of changing social roles with age and/or changes in self-awareness. It is this subsequent development in response to external social pressures and ultimate existential questions that may explain adult self-actualization and moral development as described by Erikson, Kohlberg, and others. Consequently, intellectual components of the character traits are conscience (ethical principles such as fairness), contemplation, and responsibility (Table 26–14). Likewise, the ultimate spiritual components of the character traits are compassion, spiritual acceptance, and self-actualization (also called enlightened second nature).

PERSONALITY DISORDER

Normal versus Deviant Personality

Normal personality is usually defined (i) directly, using criteria of health ideals, or (ii) indirectly, as the opposite to deviant personality, or most frequently (iii) statistically, by behaviors that are most common in the given environment. However, the distinction between "normal" and "abnormal" personality is inherently relative as it relies on arbitrary cutoff points on the continuum between two extremes (very low and very high) of any behavior. This distinction is also context-dependent as the same behavior, manifested in different situations, could be viewed as normal or maladaptive (e.g., invariant cautiousness when danger is unlikely and the same trait when danger is likely). However, relying solely on the social or situational context to establish the diagnosis is problematic because PD involves many noninterpersonal traits as well (e.g., narcissistic persons satisfy many aspects of their grandiosity in fantasy). Here, personal deviance alone also does not reliably distinguish between normal behavior and PD (e.g., some individuals are very socially withdrawn without impairment in professional functioning or signs of personal suffering and distress). In other words, both personal and social aspects are needed to fully account for the symptoms of PD.

Classification of Personality Disorders

The leading classifications of PDs are the ICD of the World Health Organization and the DSM of the American Psychiatric Association. The ICD classification of PDs, 9th edition, relied extensively on Kurt Schneider's book "Psychopathic Personalities" and his description of 10 discrete socially deviant personality types. The 10th revision of ICD (ICD-10), published in 1987, corresponds more closely to the DSM system, however. Anglo-American concepts of PD originate in Pritchard’s description of moral insanity, which was later termed “sociopathy,” and finally became antisocial PD in DSM-II in 1968. However, other classified PDs in the DSM system can be traced to the work of Kurt Schneider. This chapter describes the DSM-5 classification of PD, which is the same as DSM-IV, originally published in 1994. It is, however, descriptively similar to ICD-10 and, for practical purposes, much of the following section on DSM-5 holds for ICD-10 as well.

It is also useful to be aware of changes introduced in the 11th edition of the ICD (ICD-11), which was approved by the WHO in 2019, and those proposed as alternative DSM-5 criteria that were rejected by the APA because of their unwieldy content and lack of scientific consensus. Both of these interesting criteria sets involve two major steps, as proposed by Cloninger in 2000: first, rating the presence of PD and its severity in terms of levels of adaptive functioning, and second, describing the specific traits that are prominent in the person.

In the ICD-11, personality is first rated as mild, moderate, or severe PD based on disturbed functioning of aspects of self or relations with others. Likewise, in the unofficial alternative DSM-5 criteria for PD, the presence of PD is based on mild, moderate, or severe personality dysfunction similar to Self-directedness (viz., identity and self-direction) and Cooperativeness (viz, empathy and intimacy), which are empirically are strong general predictors of official DSM PDs.

The ICD-11 and DSM-5 classifications then differ from one another in the second step. The separate classification of 10 PDs from ICD-10 is omitted in ICD-11 (except that borderline may be coded as a qualifier) but are retained in the official criteria of DSM-5, which are unchanged from DSM-IV. However, both official ICD-11 and unofficial research criteria for PD listed as an appendix to DSM-5 permit clinicians to code for five trait dimensions of personality dysfunction. However, empirical research shows that these two sets of trait domains actually describe four analogous trait domains: Negative Affectivity, Detachment, Dissociality/Antagonism, and Disinhibition. The fifth dimension proposed by these models is different: DSM-5 suggests Psychoticism (vs. Lucidity), whereas ICD-11 Anankastia (i.e., obsessional orderliness and perfectionism). The four descriptive qualifiers of PD correspond well to the four dimensions of temperament that distinguish subtypes of PD and have distinctive molecular biology as described by Cloninger and colleagues. These traits have been previously described by “the four As” by Mulder and Joyce: asthenic/anxious (i.e., negative affective features such as high Harm Avoidance, as in the DSM anxious cluster C and Galen’s melancholic temperament), adventurous/antisocial (i.e., dissocial and antagonistic features such as high Novelty Seeking, as in the DSM impulsive cluster B and Galen’s choleric temperament), asocial (i.e., detached features such as low Reward Dependence as in the DSM aloof cluster A and unlike Galen's sanguine temperament), and anankastic (i.e., perseverative and perfectionistic features such as high Persistence as in obsessional personalities and Galen's phlegmatic temperament). Empirical studies also show that four dimensions are adequate because Disinhibition and Anankastia may be bipolar extremes of a single dimension in self-reports and clinician ratings. In any case, it is difficult to choose between alternative descriptive models because factor analysis is a statistically indeterminate method and falsely assumes that personality is made up of independent traits rather than dynamically interactive systems of learning. For a reasonable scientific choice among alternative models, it is essential to take into account data about the genetics, neurobiology, and development of personality, as is done in the psychobiological model of personality but not the descriptive models of ICD-11 and DSM-5.

The individual personality traits used to subtype or qualify PD can also occur in combination with one another, so the produced categories are really not mutually exclusive, just as is also true of the corresponding DSM-IV clusters. More fine-grained descriptions can be used as in DSM, but the international classification needs to be efficiently applied around the world in places with limited resources, so ICD-11 requires only ratings of the severity of PD. Even more lengthy criteria like the DSM-5 may be too unwieldy and burdensome. The alternative DSM-5 criteria also involve categorical ratings that are not systematically derived from psychobiologically validated dimensions of personality, so the American Psychiatric Association judged it premature to endorse them. The official DSM-5 criteria, which are the same as in DSM-IV, will be described in detail here because of a wealth of clinical experience and research that is available about them.

DSM-5-TR Classification

According to DSM-5 and DSM-5-TR, the critical criterion for distinguishing deviant personality traits is the presence (evidence) of long-term maladaptation and inflexibility that are manifested as either subjective distress or socio-occupational functional impairment. According to the DSM-5, disorders of personality involve enduring patterns of inner experience and behavior that are manifest by at least two areas of cultural deviance involving perceptions and thoughts, emotional reactivity, interpersonal functioning, and impulse control.

The pattern is stable and of long duration, and its onset can be traced back at least to adolescence or early adulthood. It is inflexible and pervasive across a broad range of personal and social situations and leads to clinically significant distress or impairment in social, occupational, or other important areas of functioning.

It is unfortunate that these general criteria are rather vague because the most important assessment about personality is the level of maturity of character functions. The most commonly diagnosed PD is “not otherwise specified.” As a result, Cloninger has proposed that clinicians first evaluate the dimensions of character to determine whether there is any PD or not. As a second step, the features that distinguish subtypes can be described, but this is of secondary importance because the main focus of treatment is character development in every case.

PD subtypes classified in DSM-5 are Schizotypal, Schizoid, Paranoid (Cluster A), Narcissistic, Borderline, Antisocial, Histrionic (Cluster B), and Obsessive-Compulsive, Dependent, and Avoidant (Cluster C). In addition to these 10 “standard” disorders, the DSM-5 recognizes “personality change due to another medical condition,” “other specified PD” such as people with mixed features of more than one personality category, and “unspecified PD” for presentations in which the clinician chooses not to be specific or has insufficient information to be specific. Two other disorders, Passive-aggressive and Depressive that were recognized for further study in DSM-IV are omitted in DSM-5 but can be diagnosed as: “other specified PD.” The behaviors associated with borderline, passive-aggressive and depressive adaptations are regarded by many experts as nonspecific expressions of severe character dysfunction.

The DSM-5 arranges categorical PDs into three clusters, each sharing some clinical features (see Table 26–15): Cluster A includes three disorders with odd, aloof features, such as Paranoid, Schizoid, and Schizotypal. Cluster B includes four disorders with dramatic, impulsive, and erratic features, such as Borderline, Antisocial, Narcissistic, and Histrionic. Cluster C includes three disorders sharing anxious and fearful features, such as Avoidant, Dependent, and Obsessive-Compulsive. The ICD-11 may further separate the obsessive-compulsive cases from the avoidant and dependent cases. Several studies have supported the construct validity of these clusters except that the symptoms for compulsive disorder tend to form a separate fourth cluster, as proposed for ICD-11 and indicated in Table 26–15. Note that the three dimensions underlying Clusters A, B, and C (i.e., detachment, impulsivity, and fearfulness) correspond closely to normal temperament traits (i.e., low Reward dependence, high Novelty Seeking, and high Harm Avoidance, respectively), suggesting that variation in these temperament traits might be significant in distinguishing among the three clusters of disorders.

Table 26–15.
Qualitative Clusters and Subtypes of Personality Disorders According to the American Psychiatric Association (DSM-IV, 1994)

Cluster	Subtype	Discriminating Features
Odd/eccentric		
	Schizoid	Socially indifferent
	Paranoid	Suspicious
	Schizotypal	Eccentric
Erratic/impulsive		
	Antisocial	Disagreeable
	Borderline	Unstable
	Histrionic	Attention seeking
	Narcissistic	Self-centered
Anxious/fearful		
	Avoidant	Inhibited
	Dependent	Submissive
	Obsessive	Perfectionistic
Other specified		
	Passive–aggressive	Negativistic
	Depressive	Pessimistic

INDIVIDUAL PERSONALITY DISORDERS

General Diagnostic Guidelines

DSM-5 introduces the following general guidelines for the diagnosis of PD:

Many of the features of various PDs may also be seen during an episode of another mental disorder. The diagnosis of PD is only made when the features are typical of long-term functioning and are not limited to a discrete episode of another mental disorder. Likewise, when maladaptive behavior features are due to the direct psychological effects of another mental disorder, including substance use (e.g., various psychoactive substances including medication), diagnosis of PD is not warranted. However, it is possible to diagnose personality change due to another medical condition such as temporal lobe epilepsy; in this case, the medical condition must be specified along with the associated diagnosis of personality change.

Judgments about personality functioning must take into account one's ethnic, social, and cultural background. PD should not be confused with acculturation following immigration or with the expression of customs and religious and political values characteristic of one's culture of origin.

Diagnosis of specific PD may be made in children or adolescents when observed maladaptive personality traits are pervasive, persistent, and unlikely to be limited to a particular developmental stage or other major psychiatric illness. The features must be present for more than 1 year to diagnose a PD in an individual younger than 18 years. The only exception to this is Antisocial PD, which cannot be diagnosed in individuals younger than 18 years.

Clinical experience points to a potential sex bias in diagnosing PDs. Certain PDs are diagnosed more frequently in men (e.g., Antisocial and Schizoid), whereas some disorders are more frequently diagnosed in women (e.g., Borderline, Histrionic, and Dependent). Although it is likely that there exist real gender differences in the prevalence of these disorders, clinicians are cautioned not to diagnose certain PDs in males and females because of social stereotypes about typical gender roles and behaviors.

Observed maladaptive behavior traits must be pervasive, that is, manifest in a wide range of personal and social contexts (i.e., at home, at work, with family and friends), not isolated aspects of the person's life. The collection of data from collateral informants is thus considered critical to ensuring the high-quality personality assessment, diagnostic reliability, and validity.

In addition to "official" diagnostic criteria, DSM-5 specifies a group of "associated features." These features are not given in the form of explicit criteria (as their "official" counterparts) but rather as loose descriptions of optional but fairly frequent behaviors intended to help clinicians in cases in which the diagnosis is not certain.

Cluster A Personality Disorders

Paranoid Personality Disorder

CLINICAL CRITERIA. The hallmarks of Paranoid PD are excessive suspiciousness, and distrust of others expressed as a pervasive tendency to interpret actions of others as deliberately demeaning, malevolent, threatening, exploiting, or deceiving. Diagnostic features also include at least four of the following:

- Hypersensitivity to and unforgiveness of insults, slights, and rebuffs
- Unwarranted tendency to question the loyalty of friends or fidelity of spouse or sexual partners

- Reluctance to confide in others because of unwarranted fear that the information will be used against them
- Reads hidden demeaning or threatening meanings into benign remarks or events
- Preoccupation with “conspirational” explanations of and hidden demeaning or threatening meanings into benign events or remarks
- Persistently bears grudges (i.e., is unforgiving)
- Unwarranted tendency to perceive attacks on their character or reputation with angry reactions or counterattacks

Some of the associated features include the following:

- Excessive need to be self-sufficient and strong sense of autonomy
- Pervasive inability to relax and compromise
- Frequent involvement in legal disputes
- Frequently impress others as “fanatics”
- Tendency to form closed groups or “cults” consisting of people with similar beliefs
- Peculiar ability to generate fear in others

COMPLICATIONS. Brief reactive psychosis, particularly in response to stress.

COMORBIDITY. These patients are at increased risk for Major Depression, Obsessive-Compulsive Disorder, Agoraphobia, and Substance Abuse or Dependence. The most common co-occurring PDs are Schizotypal, Schizoid, Narcissistic, Avoidant, and Borderline.

Paranoid PD has been postulated to be a premorbid antecedent of Delusional Disorder, paranoid type.

IMPAIRMENT. Frequently only mild and typically includes occupational and social difficulties.

SEX RATIO. According to DSM-5, this disorder is more commonly diagnosed in males.

EPIDEMIOLOGY. Prevalence rates of 0.5% to 4.4% in the general population, 10% to 30% for psychiatric inpatients, and 2% to 10% for psychiatric outpatients are reported in DSM-IV-TR and DSM-5.

FAMILIAL PATTERN AND GENETICS. Some studies have demonstrated an increased prevalence of this PD among relatives of probands with chronic Schizophrenia and Delusional Disorder, paranoid type.

DIFFERENTIAL DIAGNOSIS. Paranoid PD is distinguished from Schizophrenia (especially paranoid type), Delusional Disorder, paranoid type, and a mood disorder with psychotic features based on periods with positive psychotic symptoms, such as delusions and hallucinations, in the latter. When a brief reactive psychosis with delusions complicates the clinical picture of Paranoid PD, this distinction is far more difficult. The duration of the latter and its frequent association with stress is usually sufficient for differential diagnosis.

Paranoid PD is sometimes difficult to distinguish from the following PDs:

- Schizotypal (which includes magical thinking, unusual perceptual experiences, oddities in speech, appearance, and thought processes)
- Obsessive-Compulsive, Schizoid, Borderline, and Histrionic (all with no prominent paranoid ideation)
- Avoidant (which includes fear of embarrassment)
- Antisocial (which includes personal gains in antisocial behavior)
- Narcissistic (which includes fear of having their “hidden” imperfections and flaws revealed)

Schizoid Personality Disorder

CLINICAL CRITERIA. The hallmarks of schizoid PD are a pervasive pattern of social detachment and a restricted range of expressed emotions in interpersonal settings beginning by early adulthood and present in a variety of contexts, as indicated by four or more of the following:

- Indifference to praise and criticism
- Preference for solitary activities and fantasy (“loners”)
- Lack of interest in sexual interactions
- Lack of desire or pleasure in close relationships
- Emotional coldness, detachment, or flattened affectivity
- No close friends or confidants other than family members
- Pleasure experienced in few if any activities

Some of the associated features include the following:

- Difficulty in expressing anger, even in response to direct provocation, which contributes to the impression of flattened affect
- Passivity in adverse circumstances
- Severe lack of social skills

COMPLICATIONS. Very brief reactive psychosis, particularly in response to stress.

COMORBIDITY. This PD sometimes appears as the premorbid antecedent of Delusional Disorder, Schizophrenia, or rarely Major Depression. The most common co-occurring PDs are Paranoid, Schizotypal, and Avoidant.

IMPAIRMENT. Frequently severe problems in social relations; occupational problems develop when interpersonal involvement is required, solitary work sometimes favorably affects overall performance.

SEX RATIO. According to DSM-5, this disorder is more commonly diagnosed in males and may cause more impairment in them.

EPIDEMIOLOGY. Prevalence rates varying from 3.1% to 4.9% have been reported (DSM-5).

FAMILIAL PATTERN AND GENETICS. An increased prevalence among the relatives of probands with Schizophrenia or Schizotypal PD has been reported.

DIFFERENTIAL DIAGNOSIS. Schizoid PD is distinguished from Schizophrenia, Delusional Disorder, and mood disorders with psychotic features based on periods with positive psychotic symptoms, such as delusions and hallucinations, in the latter. When a brief reactive psychosis complicates the clinical picture of Schizoid PD, this distinction is far more difficult. The duration of the latter and its frequent association with stress is usually sufficient for differential diagnosis.

Schizoid PD is distinguished from Autistic Disorder and Asperger’s Disorder by more severely impaired social interactions and stereotypic behaviors and interests in the latter two disorders. Schizoid disorder is distinguished from the following PDs:

- Schizotypal (which includes magical thinking, unusual perceptual experiences, oddities in speech, appearance, and thought processes)
- Avoidant (adequate emotionality in the latter, also social isolation due to the fear of embarrassment, not to indifference)
- Obsessive-Compulsive (adequate capacity for intimacy, despite sometimes excessive isolation due to perfectionism and workaholic attitudes)
- Paranoid (which includes suspiciousness, ideas of reference, and guarded facade)

Schizotypal Personality Disorder

CLINICAL CRITERIA. Schizotypal PD is characterized by social and interpersonal deficits as indicated by pervasive discomfort with reduced capacity for close relationships as well as cognitive and perceptual distortions and eccentric behavior (not severe enough to meet criteria for Schizophrenia). Diagnostic features also include at least five of the following:

- Ideas of reference (not delusions)
- Odd beliefs and magical thinking (superstitiousness, beliefs in clairvoyance, telepathy, “sixth sense”; in children and adolescents, bizarre fantasies or preoccupation)
- Unusual perceptual disturbances (illusions including bodily illusions, sensing a presence of a person nearby)
- Paranoid ideation and suspiciousness
- Odd, eccentric, peculiar behavior
- Lack of close friends, except family members
- Odd thinking and speech without coherence (e.g., vague, metaphorical, overelaborated, stereotypic)
- Inappropriate or constricted affect
- Social anxiety that does not diminish with familiarity and is associated with paranoid fears

PREDISPOSING FACTORS. Empirical findings about this disorder are difficult to interpret as there appear to be several clinical subtypes of this disorder, with potentially differential relationships to the schizophrenic spectrum. Adoption, family, and twin studies demonstrate an increased prevalence of schizotypal features in the families of schizophrenic patients, especially when schizotypal features were not associated with comorbid affective symptoms. Small anticipatory saccades that disrupt Smooth Pursuit Eye Movement, hypothesized to be a marker of genetic vulnerability to schizophrenia, have also been detected in schizotypal and introverted personalities. The nature of the relationship of this PD and Schizophrenia is still controversial, with some genetic studies not finding the relationship and some studies finding a pattern of gradation in multifactorial liability from schizotypal personality (mild) to broad schizophrenia (moderate) to narrow schizophrenia (severe).

COMPLICATIONS. Transient psychotic episodes, particularly in response to stress. Symptoms sometimes become so significant that they meet the criteria for Schizophreniform Disorder, Delusional Disorder, and Brief Psychotic Disorder.

COMORBIDITY. More than half of these patients have had at least one episode of Major Depression, and 30% to 50% have Major Depression concurrent with this PD. The most common co-occurring PDs are Schizoid, Paranoid, Avoidant, and Borderline.

IMPAIRMENT. Typically includes occupational and social difficulties.

SEX RATIO. Unknown; frequently diagnosed in females with fragile X syndrome.

EPIDEMIOLOGY. Prevalence rates of 3.9% to 4.6% in the general population are reported in DSM-5. Earlier reports suggested a range between 1% and 6%. It may be less common in clinical populations than in the general population, according to DSM-5.

FAMILIAL PATTERN AND GENETICS. An increased prevalence of this PD among the first-degree relatives of probands with Schizophrenia. Also, an increased prevalence of Schizophrenia and other psychoses in the relatives of probands with Schizotypal PD. The disorder itself tends to aggregate in families (DSM-5).

DIFFERENTIAL DIAGNOSIS. Schizotypal PD is distinguished from Schizophrenia, Delusional Disorder, and mood disorders with psychotic features based on periods with positive psychotic symptoms, such as delusions and hallucinations, in the latter. When a brief reactive psychosis with delusions complicates the clinical picture of Schizotypal Personality, this distinction is far more difficult.

Schizotypal PD is difficult to distinguish from the heterogeneous group of solitary, odd children whose behavior is characterized by social isolation, eccentricity, and peculiarities in language seen in Autistic Disorder, Asperger's Disorder, and Expressive and Mixed Receptive–Expressive Language Disorder. Communication disorders might be distinguished by the primacy and the severity of the disorder in language accompanied by compensatory efforts of the child to communicate by other means and also by specialized language assessment. Autistic Disorder and Asperger's Disorder are distinguished on the basis of more severely impaired social interactions and stereotype behaviors and interests in the latter two disorders. Schizotypal disorder is distinguished from the following PDs:

- Schizoid and Paranoid (which rarely have magical thinking, unusual perceptual experiences, or oddities in speech, appearance, and thought processes)
- Narcissistic (with predominant sense of grandiosity, fragile self-esteem, and fear of having their "hidden" imperfections or flaws revealed)
- Avoidant (which rarely has oddities in appearance and behavior, and here fear of embarrassment, not disinterest and detachment, causes social avoidance and isolation)
- Borderline (characterized by affective instability and stormy relationships as well as impulsive and manipulative behavior)

Cluster B Personality Disorders

Antisocial Personality Disorder

CLINICAL CRITERIA. The hallmarks of antisocial PD are a pervasive disregard for and violation of rights of others occurring since the age of 15 years and continuing into adulthood. A person has to be 18 years of age or older, and there has to be evidence of conduct disorder before the age of 15 years (conduct disorder involves a repetitive and persistent pattern of behavior in which the basic rights of others or major age-appropriate social rules are violated; the examples include aggression to people and/or animals, destruction of property, deceitfulness or theft, serious violation of rules). Diagnostic features also include at least three of the following:

- Failure to conform to social norms (resulting in frequent arrests)
- Deceitfulness, including lying and conning others for personal profit or pleasure
- Impulsivity or failure to plan ahead
- Irritability and aggressiveness, including repeated physical fights or assaults
- Recklessness, with disregard for the safety of self and others
- Irresponsibility, indicated by the failure to honor financial obligations or to sustain consistent work behavior
- Lack of remorse, indicated by indifference or rationalizing having hurt, mistreated, or stolen from others

Some of the associated features include the following:

- Promiscuity and inability to sustain a monogamous relationship
- Lack of empathy, cynicism, contempt for feelings, rights, or suffering of others
- Inflated and arrogant self-appraisal
- Abusiveness and irresponsibility toward children

COMPLICATIONS. Dysphoria, tension, low tolerance for boredom, depressed mood, and premature, violent death.

COMORBIDITY. These patients are at increased risk for Impulse Control Disorders, Major Depression, Substance Abuse or Dependence, Pathologic Gambling, Anxiety Disorders, and Somatization Disorder. The most common co-occurring PDs are Narcissistic, Borderline, and Histrionic.

IMPAIRMENT. Extremely variable typically includes social difficulties.

SEX RATIO. According to DSM-5, this disorder is more commonly (3:1) diagnosed in males.

EPIDEMIOLOGY. Prevalence rates of 3% for males and 1% for females in the general population, 3% to 30% in clinical settings with even higher rates for forensic samples and substance abusers, have been reported. As noted in DSM-5, there has been some concern that this disorder may be underdiagnosed in females, given the emphasis on aggressive items in diagnosing Conduct Disorder.

High frequency of Antisocial PD is associated with low socioeconomic status and urban settings.

COURSE. After the age of 30 years, both the most flagrant antisocial behaviors (promiscuity, crime) and the less severe behaviors and substance use tend to decrease. Even after the severe antisocial behavior “burns out,” people diagnosed with antisocial PD usually continue to be irritable, impulsive, and detached.

FAMILIAL PATTERN AND GENETICS. Antisocial PD is more frequent among the first-degree biologic relatives of probands with this disorder than in other relations. Biologic relatives of females with Antisocial PD are at increased risk for the same disorder than biologic relatives of males with Antisocial PD. Genetic studies have suggested familial transmission of Antisocial PD, Substance Use, and Somatization Disorder, the former two being characteristic of males, the latter of females in the same family.

Adoption studies have shown that both genetic and environmental factors contribute to the risk for this disorder.

Both adopted, and biologic children of parents with Antisocial PD are at increased risk for this disorder. Conduct Disorder (before the age of 10 years) and accompanying Attention-Deficit Hyperactivity Disorder increase the likelihood of developing antisocial personality in adult life. Conduct Disorder is more likely to develop into Antisocial Disorder with erratic parenting, neglect, or inconsistent parental discipline.

DIFFERENTIAL DIAGNOSIS. Antisocial PD is distinguished from Bipolar Disorder, manic on the basis of episodic course, and euphoric mood of the latter. Antisocial PD is distinguished from the following PDs:

- Narcissistic (which rarely manifests serious criminality, aggression, and deceit and is characterized by an excessive need for admiration and envy of others)
- Histrionic (which includes seductiveness, attention-seeking, superficiality, and rarely serious criminality and aggressiveness)
- Borderline (which includes manipulativeness to gain nurturance and affective instability). Contrary to the common belief that borderline personalities rarely commit crimes, individuals with explosive or borderline temperaments (high Harm Avoidance, high Novelty Seeking, and low Reward Dependence) frequently manifest antisocial behaviors (“secondary” psychopathy). Secondary psychopathy is distinguished from antisocial personality proper (or

"primary" psychopathy) as the latter has a different temperament profile: high Novelty Seeking with low Harm Avoidance and low Reward Dependence, corresponding to the adventurous temperament profile.
- Paranoid (which includes suspiciousness, guarded attitude, rarely serious antisocial behaviors)
- Adult Antisocial Behavior (with no personality pathology in the background)

Narcissistic Personality Disorder

CLINICAL CRITERIA. The hallmarks of Narcissistic PD are a pervasive sense of grandiosity (in fantasy or in behavior), need for admiration, lack of empathy, and chronic intense envy. Diagnostic features also include at least five of the following:

- Grandiose sense of self-importance and specialness
- Preoccupation with fantasies of unlimited success, power, brilliance, beauty, or ideal love
- Sense of entitlement
- Regards self as "special" and unique
- Interpersonal exploitativeness such as taking advantage of others to achieve own needs
- Lack of empathy
- Excessive need for admiration and acclaims
- Intensive and chronic envy
- Arrogant and haughty attitude

Some of the associated features include the following:

- Fragile self-esteem (which exclusively depends on external admiration) with hypersensitivity to criticism
- High achievements more frequent than in any other PD
- Strong feelings of shame and humiliation
- Exhibitionism (behavior motivated by the pleasure of being looked at)
- Fear of having their "hidden" imperfections and flaws revealed

COMPLICATIONS. Social withdrawal, depressed mood, and Dysthymic or Major Depressive Disorder in reaction to criticism or failure.

COMORBIDITY. These patients are at increased risk for Major Depression and Substance Abuse or Dependence (especially cocaine use). The most common co-occurring PDs are Borderline, Antisocial, Histrionic, and Paranoid.

IMPAIRMENT. Frequently severe, typically includes marital problems and interpersonal relationships in general.

SEX RATIO. According to DSM-5, this disorder is more commonly diagnosed in males (50% to 75% of diagnosed cases are males).

COURSE. Chronic. However, narcissistic symptoms tend to diminish after the age of 40 years, when pessimism usually develops.

EPIDEMIOLOGY. Prevalence rates of 2% to 16% in the clinical population, less than 1% in the general population are reported in DSM-IV-TR and DSM-5.

PREDISPOSING FEATURES. There may be a higher risk for this PD in the offspring of narcissistic parents who impart on their children an unrealistic sense of grandiosity. In addition, most narcissistic persons are realistically talented, beautiful, or highly intelligent, as these features serve as the nucleus around which the sense of specialness is further organized.

DIFFERENTIAL DIAGNOSIS. Narcissistic PD is distinguished from Manic or Hypomanic episodes by the episodic course, euphoria, and functional impairments in the latter two. Narcissistic personality is distinguished from the following PDs:

- Antisocial (narcissistic exploitativeness is more driven by the wish to establish one’s dominance than by material gains, history of Conduct Disorder, no excessive need for admiration)
- Borderline (which includes unstable self-concept, chaotic behaviors, self-destructive gestures, chronic anxiety, and rarely high achievements)
- Histrionic (which includes capacity for empathy, emotional display, and rarely unscrupulousness and exploitation of others)
- Obsessive-Compulsive (which includes inflexibility, detail orientedness, and social isolation in addition to perfectionism and belief that others cannot do things as well)
- Paranoid and Schizotypal (which include suspiciousness and social withdrawal).

Histrionic Personality Disorder

CLINICAL CRITERIA. The hallmarks of histrionic PD are pervasive and excessive self-dramatization, excessive emotionality, and attention-seeking. Diagnostic features also include at least five of the following:

- Inappropriate sexual seductiveness or provocativeness

- Excessive need to be in the center of attention
- Rapidly shifting and shallow expression of emotions
- Suggestibility
- Physical appearance used for attention-seeking purposes
- Impressionistic speech lacking detail
- Self-dramatization, theatricality, exaggerated expression of emotions
- Relationships are considered more intimate than they really are

Some of the associated features include the following:

- Difficulties in achieving emotional intimacy in romantic or sexual relationships
- Craving for excitement and stimulation
- Promiscuity or complete sexual naiveness
- Low tolerance for delayed gratification

COMPLICATIONS. Frequent suicidal gestures and threats to coerce better caregiving. Interpersonal relations are unstable, shallow, and generally ungratifying. Frequent marital problems secondary to the tendency to neglect long-term relationships for the excitement of new relationships.

COMORBIDITY. These patients are at increased risk for Major Depression, Somatization Disorder, and Conversion Disorder. The most common co-occurring disorders are Narcissistic, Borderline, Antisocial, and Dependent.

IMPAIRMENT. Frequently only mild and typically includes personal, romantic relationships.

SEX RATIO. There seems to be a general agreement that this disorder occurs far more frequently among women. According to DSM-5, the disorder might be equally frequent among men and women.

EPIDEMIOLOGY. Prevalence rates of about 2% in the general population, 10% to 15% for psychiatric inpatients and outpatients are reported in DSM-IV-TR and DSM-5.

FAMILIAL PATTERN AND GENETICS. This disorder tends to run in families. A genetic “link” between Histrionic and Antisocial PD and Alcoholism has been suggested.

DIFFERENTIAL DIAGNOSIS. Histrionic PD is distinguished from the following PDs:

- Antisocial (which includes antisocial behaviors and crime to gain profit, power, or some other material gratification, history of Conduct Disorder, no excessive self-dramatization, no exaggerated emotional expression)

- Borderline (which includes unstable self-concept, chaotic behaviors, self-destructive gestures, chronic anxiety, identity disturbance)
- Narcissistic (which includes fear of having their “hidden” imperfections and flaws revealed, sense of grandiosity and specialness)

ANTISOCIAL, NARCISSISTIC, AND HISTRIONIC SPECTRUM. Narcissistic, Antisocial, and Histrionic PDs qualify for the so-called spectrum disorders. The three disorders have been shown to aggregate in the same family and to co-occur in the same person. Phenotypically distinguishable disorders may be referred to as spectrum disorders if they meet the above two conditions. Spectrum disorders sometimes reflect a differential expression of the same liability. Antisocial, Narcissistic, and Histrionic PDs exemplify the spectrum in which proneness to impulsivity and aggression (associated with high Novelty Seeking) interferes with character development and maturity. Empirical data show that symptoms of the three disorders tend to group around impulsivity, aggression, and dramatic affects. A spectrum disorder may also reflect a less deviant form of the other on an underlying liability scale. In that regard, the underlying antisocial behaviors increase in severity as one proceeds from Histrionic Personality (manipulation) via Narcissistic Personality (exploitation) and Antisocial Personality (violent and property crime).

Borderline Personality Disorder

CLINICAL CRITERIA. The hallmarks of Borderline PD are pervasive and excessive instability of affects, self-image, and interpersonal relationships, as well as marked impulsivity. Diagnostic features also include at least five of the following:

- Frantic efforts to avoid real or imagined abandonment. **Note:** Do not include suicidal or self-mutilating behavior covered in Criterion 5
- Unstable and intense interpersonal relationships, with alternating between idealization and devaluation
- Markedly and persistently unstable self-image or sense of self
- Impulsivity in at least two potentially self-damaging areas (spending, sex, substance abuse, binge eating, reckless driving). **Note:** Do not include suicidal or self-mutilating behavior covered in Criterion 5
- Recurrent suicidal behavior, gestures, threats, or self-mutilating behaviors
- Instability of affect due to marked reactivity of mood
- Chronic feelings of emptiness
- Inappropriately intense anger or difficulty controlling anger
- Stress-related, transient paranoid ideation or dissociative symptoms

Some of the associated features include the following:

- Tendency to undermine self when close to realizing a goal
- Feeling more secure with nonhuman objects (pets, inanimate objects) than in interpersonal relationships

PREDISPOSING FACTORS. Numerous studies have pointed to early traumatic experiences in the etiology of this PD. Recently, a tripartite etiopathogenetic model, including childhood trauma, vulnerable temperament, and a series of triggering events, has been formulated. Both dynamic and biologic psychiatry agree that a combination of early traumatic events and certain biologic vulnerabilities (mostly in the emotional domain) represent primary etiologic factors for this disorder. Familial aggregation of Borderline PD has been repeatedly demonstrated.

COMPLICATIONS. Psychotic-like symptoms (hallucinations, body image distortions, hypnagogic phenomena, ideas of reference) in response to stress; premature death or physical handicaps from suicide and suicidal gestures, failed suicide, and self-injurious behavior.

COMORBIDITY. These patients are at increased risk for Major Depression, Substance Abuse or Dependence, Eating Disorder (notably Bulimia), Post-Traumatic Stress Disorder, and Attention-Deficit Hyperactivity Disorder. Borderline PD co-occurs with most other PDs.

IMPAIRMENT. Frequent and severe; frequent job losses interrupted education, broken marriages.

SEX RATIO. According to DSM-5, this disorder is more commonly diagnosed in females (75% of diagnosed cases are females).

COURSE. Variable. Most commonly follows a pattern of chronic instability in early adulthood, with episodes of serious affective and impulsive dyscontrol. The impairment and the risk of suicide are the greatest at the young adult years and gradually wane with advancing age. In the fourth and fifth decades, these individuals tend to attain greater stability in their relationships and functioning.

EPIDEMIOLOGY. Prevalence rates of about 2% in the general population, 10% for psychiatric outpatients, 20% for psychiatric inpatients, and 30% to 60% among patients with PDs are reported in DSM-5. The diagnosis is more common in younger than in older samples, suggesting a natural tendency toward maturation and remission.

FAMILIAL PATTERN AND GENETICS. Physical and sexual abuse, neglect, hostile conflict, and early parental loss or separation are common in the childhood histories of patients with this disorder. Borderline PD is five times more common among relatives of probands with this disorder than in the general population. It also increases the familial risk for Antisocial PD, Substance Abuse, and Mood Disorders.

DIFFERENTIAL DIAGNOSIS. Borderline PD is distinguished from Mood Disorder, Dysthymic Disorder, and Cyclothymia (with depression and mood swings mimicking borderline affective problems) based on efforts to avoid abandonment, unstable relationships with alternating between idealization and devaluation, identity disturbance, impulsivity in potentially self-damaging areas, chronic feelings of emptiness, and inappropriately intensive anger or difficulty controlling anger (these symptoms are rarely observed in Mood Disorder, Dysthymia, or Cyclothymia). Borderline disorder shares many features and is difficult to distinguish from all other PDs, frequently as an exclusion diagnosis based on typical clinical symptoms for other PDs. Borderline disorder is distinguished from Identity Problems (the latter is limited to a developmental stage).

Cluster C Personality Disorders

Avoidant Personality Disorder

CLINICAL CRITERIA. The hallmarks of avoidant PD are pervasive and excessive hypersensitivity to negative evaluation, social inhibition, and feelings of inadequacy. Diagnostic features also include at least four of the following:

- Avoidance of occupational activities that involve significant interpersonal contact because of fears of criticism, rejection, or disapproval
- Unwillingness to be involved with others unless certain of being liked
- Restraint in intimate relationships because of the fear of being shamed or ridiculed
- Preoccupation with being criticized or rejected in social situations
- Inhibition in new social situations because of feelings of inadequacy
- Reluctance to take personal risks or to engage in any new activities because they may prove embarrassing
- Views self as socially inept, personally unappealing, or inferior to others

Some of the associated features include the following:

- Fearful and tense demeanor
- Fear of blushing or crying in front of others in response to criticism

- Social isolation is accompanied by craving social relations and fantasizing about ideal relationships with others

COMPLICATIONS. Social phobia.

COMORBIDITY. These patients are at increased risk for Mood and Anxiety Disorders (especially Social Phobia, generalized type). The most common co-occurring disorders are Schizotypal, Schizoid, Paranoid, Dependent, and Borderline.

IMPAIRMENT. Can be severe and typically includes occupational and social difficulties.

SEX RATIO. According to DSM-5, this disorder is equally frequent in males and in females.

EPIDEMIOLOGY. Prevalence rates of 0.5% to 2% in the general population and 10% for psychiatric outpatients are reported in DSM-IV-TR and DSM-5.

COURSE. Frequently begins in childhood with shyness and fear of strangers and new situations. Disfiguring illness and shyness in childhood predispose children for this PD.

DIFFERENTIAL DIAGNOSIS. Avoidant PD is very difficult to distinguish from Social Phobia (many authors believe that these are alternative labels for the same or similar condition). In Social Phobia, specific situations rather than interpersonal contact are avoided. Panic Disorder with Agoraphobia also manifests avoidance but usually after the onset of panic attacks. Avoidant disorder is distinguished from the following PDs:

- Schizotypal and Schizoid (social isolation of avoidant personalities is accompanied by the desire for social relations, which is not observed in Schizoid and Schizotypal PD)
- Paranoid (which includes guarded attitude, preoccupation with hidden meanings, and conspirational explanations of being events)
- Dependent (which is focused on being taken care of rather than on the fear of negative evaluation)

Dependent Personality Disorder

CLINICAL CRITERIA. The hallmarks of Dependent PD are pervasive and excessive need to be taken care of that leads to clinging behavior, submissiveness, fear of separation, and interpersonal dependency.

Diagnostic features also include at least five of the following:

- Difficulty in making everyday decisions without excessive reassurance and advice from others
- Needs others to assume responsibility for major areas of their life
- Difficulties expressing disagreement with others because of fear of loss of support or approval. **Note:** Do not include realistic fears of retribution
- Lack of initiative
- Unrealistic preoccupation with fears of being left to take care of self
- Urgent search for another relationship as a source of care and support when a close relationship ends
- Extensive efforts to obtain nurturance and support from others (to the point of volunteering to do unpleasant things)
- Uncomfortable and helpless feelings when alone because of exaggerated fears of being unable to take care of self. Pessimism, self-doubt, self-devaluation.

Some of the associated features include the following:

- Prominent pessimism and self-doubt since childhood
- Hypersensitive to criticism and rejection
- Anxious avoidance of responsibility

COMPLICATIONS. Mood Disorders, Anxiety Disorders, Adjustment Disorder, Social Phobia. Low socioeconomic status, poor family, and marital functioning.

COMORBIDITY. These patients are at increased risk for Major Depression, Anxiety Disorders, and Adjustment Disorder. The most common co-occurring disorders are Histrionic, Avoidant, and Borderline.

IMPAIRMENT. Frequently only mild and typically includes interpersonal relationships and occupational functioning if independence is required.

SEX RATIO. Equally frequent in males and females (DSM-5).

EPIDEMIOLOGY. Prevalence rates of 0.5% to 0.6% are reported in DSM-5 for this specific diagnosis, even though dependency is a prominent feature of many PDs.

FAMILIAL PATTERN AND GENETICS. Unknown. Chronic physical illness or Separation Anxiety Disorder may predispose to Dependent PD.

DIFFERENTIAL DIAGNOSIS. Dependent disorder is distinguished from dependency seen in Mood Disorders, Panic Disorder, Agoraphobia, and as a result of a general medical condition. Dependent disorder is distinguished from the following PDs:

- Borderline (which includes unstable, stormy relationships and the reaction to abandonment with rage, emptiness, and demands as opposed to increasing appeasement and submissiveness seen with dependent personalities)
- Histrionic (which includes gregarious flamboyance with active demands for attention)
- Avoidant (which includes social isolation because of the fear of negative evaluation as opposed to clinging and submissive behavior of dependent personalities)

Obsessive-Compulsive Personality Disorder

CLINICAL CRITERIA. The hallmarks of Schizoid PD are pervasive and preoccupation with orderliness, perfectionism, and mental and interpersonal control at the expense of flexibility, openness, and efficiency. Diagnostic features also include at least four of the following:

- Preoccupation with details, rules, lists, order, procedures, organization, or schedules to the extent that the major point of the activity is lost
- Perfectionism that interferes with task completion
- Excessive devotion to work and productivity to the exclusion of leisure activities and friendships (not accounted for obvious economic necessity)
- Overconscientiousness, scrupulousness, and inflexibility about matters of morality, ethics, or values (not accounted for by religion or culture)
- Inability to discard worn off or worthless objects with no sentimental value
- Reluctance to delegate tasks or work with others unless they submit exactly to their way of doing things
- Stinginess (money viewed as something to be hoarded for future catastrophes)
- Rigidity and stubbornness

Some of the associated features include the following:

- Decision-making difficulties when no strict rules or established procedures dictate correct action
- Anger and frustration when not able to maintain control of the physical or interpersonal environment (anger typically not expressed directly)
- Excessive attentiveness to their relative status in dominance–submission relationships
- Controlled and restricted emotional expression, reserved personal style
- Formal and serious quality of everyday relationships

PREDISPOSING FACTORS. It has been repeatedly demonstrated that Obsessive-Compulsive PD and Obsessive–Compulsive Disorder frequently coexist. Obsessions and compulsions have been repeatedly linked to high central serotonergic function, which is associated with anxiety in general, supporting the hypothesis that obsessions and compulsions represent psychological and behavioral mechanisms reflecting underlying anxiety.

COMPLICATIONS. Distress and difficulties when confronted with new situations that require flexibility and compromise. Myocardial infarction (secondary to features typical of type A personalities, such as time urgency, hostility, competitiveness).

COMORBIDITY. These patients are at increased risk for Major Depression and Anxiety Disorder. Equivocal evidence for increased risk of Obsessive-Compulsive Disorder.

IMPAIRMENT. Frequently severe, it typically includes occupational and social difficulties.

SEX RATIO. According to DSM-5, this disorder is twice as common in males than in females.

EPIDEMIOLOGY. Prevalence rates of 2% to 8% in the general population and 3% to 10% for psychiatric outpatients are reported in DSM-IV-TR and DSM-5.

FAMILIAL PATTERN AND GENETICS. Some studies have demonstrated familial aggregation of this disorder.

DIFFERENTIAL DIAGNOSIS. Obsessive-compulsive PD is distinguished Obsessive–Compulsive Disorder based on true obsessions and compulsions in the latter. This PD is distinguished from the following PDs:

- Schizoid (which includes lack of capacity for intimacy and social isolation secondary to emotional detachment, as opposed to devotion to work, discomfort with emotions)
- Antisocial (which includes material goals in antisocial behavior and criminality as opposed to the hypermorality of obsessive personalities)
- Narcissistic (which includes a sense of grandiosity, self-aggrandizement, exhibitionism, and fear of having their “hidden” imperfections and flaws revealed)

Other Personality Disorders

Personality Change due to Another Medical Condition. This diagnosis is made when a persistent change in someone's personality is due to another medical condition, such as temporal lobe epilepsy. The disturbance must be marked and the direct result of the other medical condition, which must be specified. The disturbance cannot be better explained by another mental disorder and does not occur exclusively during the course of a delirium. It must meet the general criteria for a PD in terms of subjective distress or impairment in functioning. Several subtypes can be specified on the basis of the predominant feature:

- Labile type
- Disinhibited type
- Aggressive type
- Apathetic type
- Paranoid
- Other
- Combined (more than one of the above features)
- Unspecified (e.g., if the change is due to substance abuse and features are not specified)

Other Specified Personality Disorder

This disorder is diagnosed when the DSM-5 criteria for dysfunctional personality (i.e., general criteria for a PD) are met, but no specific criteria for any of the subtypes are observed. For example, a patient may not meet the criteria for any of the classified disorders but still manifest features from more than one PD ("mixed PD"), which, together, cause distress or impairments in social or professional functioning. This category is also used when the clinician judges that a specific PD that is not included in DSM-5 is appropriate (e.g., depressive personality or passive-aggressive personality).

Unspecified Personality Disorder

This diagnosis is applied when the general criteria for a PD are met, and the clinician has insufficient information to make a specific diagnosis or chooses not to communicate the specific reason that the presentation does not meet the criteria for a specific diagnosis.

Table 26–16.
Quantifiable (Dimensional) Features of Personality Disorder

- Consistent features
  - Low self-directedness
    - Irresponsible, blaming
    - No mature goals
    - Resourceless, helpless
    - Poor self-esteem
    - Undisciplined
  - Low cooperativeness
    - Intolerant of others
    - Lack of empathy
    - Unhelpful
    - Revengeful
    - Unprincipled
- Variable features
  - High persistence (obsessive-compulsive symptoms only)
  - Low reward dependence (odd cluster only)
  - High novelty-seeking (erratic cluster only
  - High harm avoidance (anxious cluster only)

Summary Description of the DSM-5-TR Personality Disorders

Qualitative features that are diagnostic of a PD as classified in DSM-5 are summarized in Table 26–16. Cloninger has also provided practical criteria to detail these general features so that a clinician may first decide whether a person has any PD before proceeding to diagnose a specific cluster or category of illness. The general features of all PDs are the character traits, whereas the clusters and categories are determined by additional temperament features.

Disadvantages of DSM-5-TR Classification

In addition to problems shared by all categorical models of personality (discussed later), the DSM classification has its specific limitations.

First, DSM-5 is an etiologically “atheoretical” classification of PD based solely on descriptive criteria derived from observations of prototypical cases. This “atheoretical” approach was expected to stimulate work on the etiopathogenetic understanding of deviant behaviors classified as PDs. Unfortunately, this has not been the case. The introduction of descriptive criteria in DSM-III has inspired extensive research of some basic psychometric aspects of the PD diagnosis but rarely studies of etiology, underlying dynamics, or pathogenesis. While this emphasis on psychometrics obviously reflects an attempt of this new field to delineate its subject, it is also reflective of the purely descriptive approach, which provided no etiopathogenetic models or hypotheses for testing. For example, the DSM criteria have been of little or no help in developing the psychobiological model described here. In contrast, the experimental animal and twin studies on which the psychobiological model of personality was built have now been confirmed and extended by extensive work in human beings using functional brain imaging and person-centered genomic–phenomic analyses.

Second, the DSM-IV describes as pathognomonic for PD maladaptive response patterns that are “enduring,” that is, appearing before adulthood with long-term duration. In practice, however, it can be difficult to distinguish long-term maladaptation indicative of PD from chronic personality changes caused by other factors, such as other mental disorders (e.g., chronic depression), long-term situational factors (e.g., job-related timidity), and by other medical disorders (e.g., irritability associated with hyperthyroidism).

Third, the DSM definition of PDs is inherently imprecise, as it requires that maladaptive behaviors cause **clinically significant** subjective distress or **clinically significant** impairment in social and occupational function, or both. **“Subjective distress”** usually presents as low self-esteem, anxiety, guilt, depression, and hypochondriasis. These symptoms are frequently diagnosed as another mental disorder while the background PD is overlooked, especially now that the multiaxial system of DSM-IV that distinguished personality variation from other disorders has been dropped in DSM-5 and DSM-5-TR. The quantifier “significant” is an arbitrary diagnostic element that can hardly be made objective, as shown by allowing a clinician’s judgment without qualification or justification to determine what is significant.

CLINICAL AND PSYCHOMETRIC ISSUES

Egosyntonic and Alloplastic Nature of Personality Disorder

Patients with PD typically blame other people or unfavorable circumstances for their own problems. This externalizing of responsibility is a result of the following two characteristics. First, most of these patients perceive their own deviant behaviors as appropriate and adequate; in other words, their symptoms are egosyntonic. The exceptions to this are patients with Dependent and Avoidant PD. Both disorders are characterized by prominent anxiety associated with maladaptive behaviors, and this frequently causes the patients to perceive their symptoms as disturbing, that is, egodystonic.

Second, patients with PD try to change others, not themselves; in other words, their attitude is alloplastic. The exceptions to this rule are patients with Avoidant and Schizoid PD. Avoidant persons usually try to improve their own performance to avoid negative evaluations by others. Schizoid personalities are socially detached to the extent that they are usually indifferent to what others might do, think, or feel.

Personality Disorder: Social or Clinical Diagnosis?

One of the general DSM-5 requirements for PD is that maladaptive behavior “deviates markedly from the expectations of the individual’s culture.” This requirement introduces a significant social connotation to what was meant to be a clinical classification. Indeed, clinicians frequently diagnose PD based solely on the extent to which certain behavior is acceptable within the local society.

Sociocultural pressures are always norm-favoring, promoting phenotypes that fall within the range of accepted behavioral norms. As shown later in the text, most temperament profiles with high Reward Dependence have a low incidence of diagnosed PD, whereas those with high Harm Avoidance and/or high Novelty Seeking increase this risk. In most Western societies, the normative phenotype is not the one with average personality traits (as one would expect given the adaptive flexibility of such configuration) but the one with high Reward Dependence (high Reward Dependence also means much easier conditioning of socially accepted behaviors than is the case with other temperament traits).

In a similar way, child psychiatry classifies three broad groups of mental disorders in children: one group is characterized by high Novelty Seeking (e.g., externalizing disorders such as Attention-Deficit Hyperactivity Disorder, Conduct Disorder), one group by high harm Avoidance (e.g., internalizing disorders such as Depressive and Anxiety Disorders), and one group by low Reward Dependence (asocial disorders such as Autism). Children with these extreme temperaments are diagnosed and treated to change their behavior, whereas children with high Reward Dependence (sociable, attached, dependent) are selectively permitted to develop with no corrective intervention. This raises a question about the role of medicine and psychiatry in “engineering” psychological profiles of young children by “cherry-picking” the favorable profiles and suppressing the unfavorable ones, which is beyond the scope of this chapter. Suffice here to say, biologic adaptation appears to have become less critical for human survival than behavioral and spiritual adaptation. Natural selection is not targeting human morphologic features anymore but rather human personality features and level of awareness. Through its suppression of certain temperament traits and its norm-favoring influence on character development, modern human society appears to have an important role in this process.

Categorical versus Dimensional Approach to Personality Disorder

Medical diagnosis has historically been categorical, and most clinicians tend to “think” categorically for two reasons. First, one of the most important functions of diagnosis is to prescribe appropriate corrective action, that is, treatment. The decision of whether a person is “affected” or “unaffected” (“sick” or “not sick”) is functionally equivalent to the decision whether or not that person needs treatment (hence, categorical diagnosis facilitates treatment decisions in practical work). Second, categorical systems simplify professional communication as they describe prototypical cases that meet all or most of the diagnostic criteria. Just one category, for example, antisocial personality, can communicate a great deal of vivid clinical information about the patient. However, categorical descriptions convey useful information only about prototypical and severe cases but little or no information about atypical, mixed, or mild cases (see later).

Categorical Approach. The categorical approach adopts the medical diagnostic model that a PD is either “present” or “absent” and individuals either “affected” or “unaffected.” Categorical models are optimal for discontinuous variables (e.g., marital status, fractured humerus). However, when distinguishing characteristics of a disorder are variants of normality, such as personality traits, any categorical decision about the presence or absence of a disorder is arbitrary. In other words, one of the major disadvantages of categorical models of PD is that they establish arbitrary cutoff points or thresholds for continuous behavior traits. This becomes clear, especially in mild cases close to the cutoff point, which cannot be well classified neither as “affected” or “unaffected.” In an attempt to minimize this problem, DSM-5 now allows clinicians to classify “maladaptive personality traits” that do not meet the threshold for a PD. Nevertheless, the categorical approach is not optimal for patients who do not perfectly fit their “pigeonhole” in the classification, so an additional wastebasket category for “atypical” or “mixed” cases had to be established, and this often is used to describe most cases of PD. DSM-5 calls this wastebasket “Other specified personality disorders,” but that does not make it more than the wastebasket it was when it was frankly acknowledged to be “not otherwise specified.” Finally, the categorical approach to personality does not even establish a prescriptive relationship between diagnosis and treatment because of the significant overlap of diagnostic criteria for PD subtypes, and categorical systems usually yield multiple personality diagnoses for individual patients. In such cases, treatment decisions and priorities are not easy to establish.

Dimensional Approach. Dimensional models define a number of graded and continuous behavior dimensions and specify individual differences as quantitative variations along these dimensions. These models are optimal for features that vary quantitatively (e.g., as personality does) and have different adaptive significance in different situations. These models account for the fact that everyone has multiple personality traits, more or less prominent and adaptive, rather than being simply present or absent. This eliminates the need for a description of multiple and overlapping categorical diagnoses. By quantifying one's position on a continuous personality trait, dimensional models conserve information about individual patients more than categorical models. Personality dimensions are as easily applied to the common atypical and mild cases as they are to the rare prototypical and severe cases. Finally, by using a dimensional system, one can easily manipulate complex data mathematically and include many variables.

Research has shown that normal personality traits and deviant personality traits observed in PDs represent a continuum where deviant traits are extreme variants of normal, adaptive behaviors. Note, however, that dimensional models do not answer the question of what is it about some traits that make them disorders. Natural breaks or points of rarity on the continuum of normal and maladaptive personality traits have not been detected. The extreme standing alone is not sufficient to be dysfunctional (e.g., people with low Reward Dependence can be well adapted in their environment). Dimensional models have to introduce cutoff points for continuous traits to make them more compatible with categorical diagnoses and thus more useful for treatment decisions or, as proposed here, to use probability estimates for categorical PD diagnosis based on the severity of measured temperament and character traits.

Integrated View of the Importance of Personality Subtypes

There is widespread dissatisfaction with the current official criteria for PD subtypes. The most common clinical diagnosis of PD using DSM-IV was "not otherwise specified." There is no evidence that the putative categories are truly discrete entities that are either present or absent. Should clinicians rely on dimensional or prototypical approaches? Dimensional approaches can consider configurations of scores on multiple dimensions that correspond to particular subtypes of PD. Prototypical approaches allow clinicians to rate the degree to which a particular case matches with the criteria of a particular subtype, now regarded as a prototype or model. What these suggestions ignore is that complex adaptive systems, such as personality, do have moderately stable configurations and that such configurations have modest separation from neighboring conditions. Furthermore, these metastable, quasi-discrete configurations are often maintained by particular gene–environmental interactions that are most predictable when a specific configuration of traits is specified. What is most important is for the clinician to realize that personality subtypes are configurations of component processes that interact as expressions of a nonlinear dynamical system. They are neither discrete taxons nor arbitrary configurations of dimensions. However, they tend to be stable over time which is mostly reflecting the fact that most (but not all) people with PD do not "spontaneously" mature but need professional intervention or some major life events (either extremely good or extremely bad or tragic) to "jump-start" the process. It is useful to recognize traditional subtypes of PD as long as you realize they are metastable, quasi-discrete expressions of complex biopsychosocial processes that are adaptive and dynamic.

In other words, the advantages of categories and dimensions can be combined by a recognition that in complex adaptive systems, it is the configuration of multiple dimensions into a personal profile that determines the functionality of the system. Such multidimensional profiles identify specific individuals with specific characteristics so that they have the advantages of categorical diagnoses without falsely implying that they represent a discrete disorder. For example, a person with antisocial personality is specified in terms of high Novelty Seeking combined with low Harm Avoidance and low Reward Dependence in people with immature character. The person-centered approach to profile analysis provides clusters of people whose biopsychosocial characteristics and behaviors are relatively stable but can be modified by changing circumstances, age, or treatment. Consequently, clinicians are not misled to slip into the fallacies of thinking that PDs are discrete disease entities or that people with a PD cannot change.

Measurement of Personality Disorders

Self-Report or Interview. Interviews are often considered by psychiatrists to be more reliable and valid diagnostic tools than self-reports for measuring PD, whereas psychologists often advocate psychometric testing. In fact, no data for a definite advantage of one of these formats have been established. People, in general, are biased when self-reporting deviant behaviors (in PD, these behaviors are ego-syntonic, perceived as “OK”), so interviewing the patient reduces this problem because unclear, inconsistent, or defensive responses can be clarified, and the patient’s demeanor and appearance observed. On the negative side, interviews tend to be affected by systematic biases, ideologic orientation, experience, and rating idiosyncrasies of interviewers. In self-reports, various response sets (e.g., careless inconsistency, defensiveness, and exaggeration) and validity items can be measured to help in interpretation. Recently, a set of performance scales, independent of the item content, based on the observed regularities in the technical pattern of responding to questions has been defined for validity analysis of the TCI. These performance scales are used to test the validity of self-reported personality traits. Specifically, self-report to questions, regardless of the content of the questions, tends to follow specific patterns that correlate with personality traits. For example, the number of endorsed items that are otherwise rarely found in the general population correlates with low Reward Dependence. These performance scales are used to predict scores on each of the temperament and character dimensions, which are then compared with self-reported, content-dependent scores on these same traits. Substantial score differences between “subjective-” and “performance”-

based scores on personality traits indicate either significant bias or intentional misrepresentation. Training in the use and interpretation of the TCI is available online (https://anthropedia.org).

Both self-reports and interviews depend on the patient's accuracy, honesty, and level of insightfulness. Hence, the collection of data from collateral informants and expert ratings is usually considered critical to ensuring a high-quality personality diagnosis.

State-Trait Effect. Most major clinical disorders have their less severe variants in the form of a PD (Table 26–17). The "state-trait" distinction between symptoms reflecting long-term personality traits and those reflective of transient clinical states can be difficult. Indeed, transient variation in one's emotional state can influence both self-report and clinical evaluation of long-term personality characteristics. The susceptibility to such state effects has been demonstrated for some categorical measures of PD. The state effect can be reduced, although not eliminated when subjects are repeatedly reminded to describe their usual, not their current, behavior, feelings, and attitudes. In contrast to categorical diagnosis, only those dimensions of personality that underlie susceptibility to depression and anxiety, such as neuroticism and Harm Avoidance, tend to covary with mood and anxiety states. Also, increases in anxiety or depression lead to a mild and transient decrease in Self-Directedness and Cooperativeness. Consequently, a person with no enduring PD may act immaturely when depressed or under stress. Self-reports of other dimensional traits, Novelty Seeking, Reward Dependence, and Persistence, in particular, are largely independent of and unaffected by comorbid mood and anxiety states.

Table 26–17.
Phenotypic Similarity between Major Clinical Syndromes and Personality Disorders

Clinical Syndromes: Major Variant	Personality Disorders: Minor Variants	Shared Symptoms
Schizophrenia spectrum	Schizoid PD Schizotypal PD	Negative and positive
Delusional disorder	Paranoid PD	Suspiciousness
Major depression	Depressive PD Borderline PD	Proneness to depressed mood; proneness to self-destruction
Cyclothymia Bipolar disorder Mania	Narcissistic PD Histrionic PD Antisocial PD	Mood swings; impulsivity
Obsessive-compulsive disorder	Obsessive-compulsive PD	Hypochondriasis; inflexibility
Social phobia	Avoidant PD	Shyness; avoidant behavior
Panic disorder with agoraphobia	Dependent PD	Dependency

Categorical Tests for Personality Disorder. The majority of categorical tests for PDs are based on the DSM criteria and share all the advantages and disadvantages of that categorical approach. In the DSM-III field trials, the interrater reliability was 0.61, and the test–retest reliability was 0.54 for any PD. The reliability is usually lower for individual subtypes than for the presence versus absence of any PD because of the difficulty distinguishing among the subtypes. In the DSM-5 field trials, reliability was questionable for most diagnoses, and the reliability for Borderline PD was only 0.34. As a result, the ICD-11 is considering a simpler approach that clinicians can implement more reliably.

Among the **categorical self-reports**, the two most frequently used are The Personality Diagnostic Questionnaire (PDQ R) and the MMPI scales for PD (MMPI PD). The PDQ R is a self-report version of the DSM criteria for PDs, along with the “Impairment/Distress” scale to quantify the significance of impairment or subjective distress required by DSM criteria. In practice, this instrument has been shown to overdiagnose PDs when compared to expert clinical diagnoses.

The MMPI PD scales were derived from the MMPI item pool to evaluate symptoms for DSM-III PDs. The MMPI PD scales are incorporated into MMPI so that, in addition to evaluating PD and other clinical syndromes, clinicians and researchers can also use other MMPI scales (e.g., Lie, Validity, Defensiveness) for validity and response set analyses.

Several semistructured interviews have been developed to diagnose PDs. These typically take more than an hour to conduct, so they have limited use in clinical practice. Among the **interviews that categorically diagnose** PDs, the most commonly used is The Structured Interview for DSM-IV PDs (SIDP IV). The original SIDP was developed for PDs described in DSM-III, and it has been regularly updated with each new DSM revision and edition (SIDP R for DSM-III-R, SIDP IV for DSM-IV). To increase its diagnostic validity, the SIDP-IV requires that answers for selected questions be obtained from a collateral informant. This test can be combined with a 10-minute SIDP-IV screen for PD, which can considerably reduce testing time in populations with many individuals who do not have PD. The screening instrument alone tends to overdiagnose cases but is useful to screen out people with no PD.

Other popular interviews for categorical classification of PDs include the Structured Clinical Interview for DSM-IV PDs (SCID II), The Diagnostic Interview for Personality Disorders (DIPD), and the Personality Disorder Examination (PDE). The PDE is available in an international version from the World Health Organization in several languages and makes diagnoses according to the ICD-10 in addition to DSM-IV. The interviews for DSM-IV PD remain the same for DSM-5 because the official criteria have not changed. There is a questionnaire for the personality trait domains of the alternative DSM-5 criteria called the Personality Inventory for DSM-5 (PID-5), but this involves questions about features that are not sanctioned by either ICD-11 or DSM-5 officially and is based on a restrictive statistical model with assumptions that are no longer plausible.

Dimensional Assessment of Personality Disorders. Historically, personality has been studied and evaluated using dimensions or traits and PD categorically as discrete entities of psychopathology. Hence, conceptual advances in the field of normal behavior and personality have had little or no effect on research in the field of PD and vice versa. For example, if the features associated with Eysenck's model of personality, that is, neuroticism, extraversion, and tough-mindedness, are combined, the resulting combinations do not correspond closely to traditional categories of PD.

The popular Five-Factor Model of normal personality has been recently advocated to account for underlying dimensions of PD as well. However, the ability of the Five-Factor Model to account for the underlying structure of PD has been tested primarily in nonclinical samples and normal individuals. The few studies that tested the Five-Factor Model in clinical samples showed that neuroticism, extraversion, and low agreeableness predict PD symptoms. After controlling for age and depression, only high neuroticism and low agreeableness remained as significant predictors of PD symptoms. These two dimensions define PD in a nonspecific way, namely, as a general predisposition to psychopathology (i.e., high neuroticism) accompanied by an antagonistic behavior facade (i.e., low agreeableness).

Neuroticism also tends to be confounded with nonspecific factors (such as depression or anxiety) that reduce its ability to distinguish PD, other psychopathology (e.g., mood and anxiety disorders), and well-adjusted individuals with high neuroticism, leaving open the question of its specificity for PDs.

In the authors' work, they found that some common core features, distinct from other forms of psychopathology that cause personal dysfunction, characterize PD subtypes classified in DSM-IV and DSM-5 as overlapping syndromes. For example, DSM-IV and DSM-5 arrange all PDs into three Clusters (A, B, and C), each comprising a number of PD subtypes with similar clinical features. Cluster A includes subtypes with predominantly odd and eccentric symptoms (e.g., paranoid and schizoid PD), Cluster B includes subtypes with dramatic, erratic, and impulsive symptoms (e.g., Histrionic, Antisocial, Borderline PD), and Cluster C includes subtypes with fearful and anxious symptoms (Dependent, Avoidant, Obsessive PD). These clusters are suggested to represent three independent clusters of PD subtypes, but empirically the clusters overlap with one another, much as do the subtypes within clusters.

Dimensional tests of PDs are formatted as self-reports (except for Tyrer's Personality Assessment Schedule). The most frequently used are the Millon Clinical Multiaxial Inventory (MCMI), the earlier mentioned NEO PI that evaluates the Five-Factor Personality Model and the TCI that evaluates the Seven-Factor Psychobiological Personality Model. These tests have been originally designed to evaluate normal personality (the NEO PI), psychiatric patients (the MCMI), and both normal personality and deviant personality (the TCI). The NEO PI is discussed previously vis-a-vis the homogeneity of its component traits and its usefulness in clinical work with PD. The other two tests are briefly discussed later.

The MCMI III (designed to match DSM-IV PD categories) evaluates long-term behavior traits systematized as 10 Basic Personality Patterns (dependent, avoidant, schizoid, passive-aggressive, narcissistic, antisocial, hysterical, compulsive, aggressive, and self-defeating) and three Pathologic PDs (i.e., Borderline, Schizotypal, and Paranoid). The latter is used to determine the severity of the 10 Basic Personality Patterns. The test also evaluates nine clinical syndromes. Some reports suggest that MCMI III tends to overdiagnose PD in comparison to expert clinicians.

In the literature, the number of proposed dimensions that underlie deviant personality ranges from 2 to 24 in the literature, although, in most models, it can be reduced to four dimensions (aloof, anxious, anankastic, and adventurous). The optimal number of underlying dimensions and the content of dimensions that purport to describe personality and its disorders most efficiently are still vigorously debated. This ambiguity arises because factor analyses of the behavior of unrelated individuals are always compatible with an infinite number of rotations of the inferred phenotypic factors. Fortunately, recent GWAS have been replicated in multiple independent samples with results that identify unique sets of genes underlying each of the facets of the TCI using a generalized clustering method that does not make arbitrary assumptions about the number of personality factors or their relationships with one another in individuals. In what follows, the distinction between temperament and character is used to improve both the etiologic understanding and the diagnosis of PD. In the treatment section of the chapter, the concepts of temperament and character are used to outline treatment guidelines for patients with PD. First, the TCI is described briefly.

The Temperament and Character Inventory. The TCI evaluates four major temperament dimensions (Harm Avoidance, Novelty Seeking, Reward Dependence, and Persistence) and three major character dimensions (Self-Directedness, Cooperativeness, and Self-Transcendence). Each of these main temperaments and character dimensions is composed of component facets (i.e., subscales) to evaluate response patterns elicited by specific stimuli (e.g., Harm Avoidance is observed in different settings as worry and pessimism, and/or fear of uncertainty, shyness, or excessive fatigability). The seven major dimensions and their correlations with the number of symptoms of PDs are shown in Table 26–18. The TCI is a family of tests with several specialized versions designed for varying types of informants (self-report, peer ratings, interviewers), varying age groups (for 3 to 5 years the preschool TCI, 7 to 14 years the Junior TCI, for 12 to 18 years the adolescent TCI, and 15+ years the Adult TCI), scope of information (temperament, character, both), and level of clinical detail (from 60 items for major dimensions only to 240 items for multiple facets of each dimension). The tests for various age groups overlap in their range so that exactly the same test can be administered to a person as they develop across the lifespan. The test measuring the temperament dimensions only was originally called the Tridimensional Personality Questionnaire (TPQ). In the TPQ, Novelty Seeking, Harm Avoidance, Reward Dependence, and Persistence were all measured, but Persistence was originally scored as a component of Reward Dependence. The name of the test was changed to TCI when the character scales were added, and Persistence was recognized as a fourth, separately inherited temperament dimension in twin studies in Australia and United States.

The self-report version of the TCI is a 240-item test available in both true/false and five-point Likert formats. Its psychometric properties are presented in the TCI Manual and online training (anthropedia.org).

Table 26–18.
Correlations between Temperament and Character Inventory Scales and the Total Number of Symptoms for Personality Disorders, Cluster A, Cluster B, and Cluster C

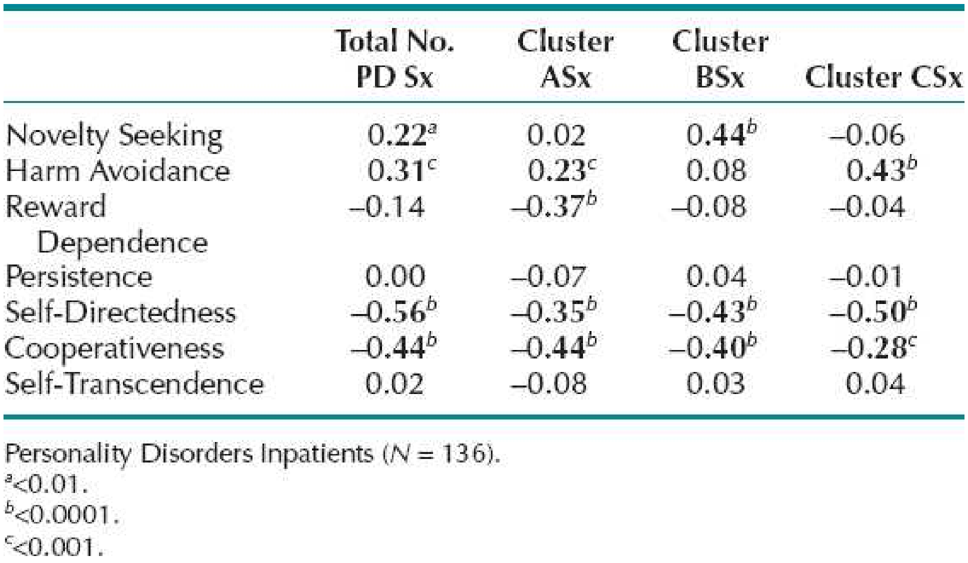

	Total No. PD Sx	Cluster ASx	Cluster BSx	Cluster CSx
Novelty Seeking	0.22[a]	0.02	0.44[b]	–0.06
Harm Avoidance	0.31[c]	0.23[c]	0.08	0.43[b]
Reward Dependence	–0.14	–0.37[b]	–0.08	–0.04
Persistence	0.00	–0.07	0.04	–0.01
Self-Directedness	–0.56[b]	–0.35[b]	–0.43[b]	–0.50[b]
Cooperativeness	–0.44[b]	–0.44[b]	–0.40[b]	–0.28[c]
Self-Transcendence	0.02	–0.08	0.03	0.04

Personality Disorders Inpatients (*N* = 136).

[a] <0.01.

[b] <0.0001.

[c] <0.001.

The ability of the TCI to predict categorical diagnoses of PDs was clinically tested in samples of psychiatric inpatients and outpatients with and without PDs and varying mood and anxiety states. Correlation and regression analyses consistently demonstrate that low scores on character dimensions, especially Self-Directedness and Cooperativeness, are associated with high symptom counts for any PD, for each of the DSM clusters of PD (Tables 26–18 and 26–19), and for each clinical subtype of PD, with or without comorbid clinical syndromes and independent of age. As shown in Table 26–19, the TCI character scales of Self-Directedness, Cooperativeness, and Self Transcendence predict the number of PD symptoms after controlling for age, anxiety, and depression.

Clearly, low scores on character traits represent a core feature (common denominator) of all PD clusters and subtypes.

In differential diagnosis, PD subtypes can be distinguished on the basis of their configuration of temperament scores with little overlap because each subtype correlates with the TCI temperament dimensions in a unique way (Fig. 26–11).

The aforementioned results have important diagnostic and treatment ramifications for PD. Diagnosis is discussed later; the treatment aspects are reviewed in the treatment section.

Table 26–19.
Stepwise Multiple Regression

Total Number of PD Symptoms		
	Partial R^2	***P***
Total number of interview PD symptoms predicted by TCI (controlling for age and depression; inpatients [*N* = 136])		
Age and Depression	0.35	0.0001
Novelty Seeking	—	NS
Harm Avoidance	—	NS
Reward Dependence	—	NS
Persistence	—	NS
Self-Directedness	0.13	0.0001
Cooperativeness	0.03	0.0076
Self-Transcendence	0.02	0.0376
Cumulative R^2	0.18	0.0001
Model R^2	0.53	0.0001
Total number of self-report PD symptoms predicted by TCI (controlling for age, depression and anxiety; outpatients [*N* = 109])		
Age, Anxiety, and Depression	0.45	0.0001
Novelty Seeking	—	NS
Harm Avoidance	—	NS
Reward Dependence	—	NS
Persistence	—	NS
Self-Directedness	0.20	0.0001
Cooperativeness	0.05	0.0001
Self-Transcendence	0.05	0.0001
Cumulative R^2	0.30	0.0001
Model R^2	0.75	0.0001

DIAGNOSIS OF PERSONALITY DISORDERS. Character and temperament traits delineate both the core features and the distinguishing features of PDs (Tables 26–18 and 26–19; Fig. 26–11). Poorly developed character traits, especially Self-Directedness and Cooperativeness, are a common denominator extending across all subtypes of PD as defined in DSM and are used in diagnosis to predict categorical PD symptoms (Tables 26–18 and 26–19). Clinically, low character scores explain chronic difficulties in accepting responsibility, setting long-term goals, fragile self-esteem, and other features associated with low Self-Directedness that are so characteristic of individuals with PD as described in DSM-5. Usually, but not always, these patients are also uncooperative as well (i.e., revengeful, opportunistic, self-centered, socially intolerant, unhelpful to others, and lacking in empathy, compassion, or principles), as is also described in DSM-5.

The combination of high Self-Transcendence with low scores on Self-directedness and/or Cooperativeness correlates with schizotypal and paranoid symptoms (depicting primary process thinking and fanaticism, respectively) and with borderline, histrionic, and narcissistic symptoms (depicting proneness to dissociation), much as predicted in Figure 26–5. In addition, recent data indicate that high Self-Transcendence might be an important component in one's susceptibility to psychosis (when other two character traits are low or, on the contrary, it may predispose to high creativity and well-being (when other two character traits are high). As previously discussed, recent psychobiological studies suggest that the healthiest and happiest people are those who are high in all three character dimensions. Now that the human population's use of resources is exceeding what the planet can replenish, high Self-Transcendence has become crucial if the human species is to avert mass extinction. Human health and survival depend on the cultivation of self-transcendent values and behavior that are altruistic and creative. In addition, the maturity and integration of all three character traits promote physiologic health, happiness, and fulfillment in ways that are indicated objectively by measures of psychophysiological coherence and markers of longevity, such as optimal heart rate variability, vagal tone, and DHEA/cortisol ratio.

High Self-Directedness alone is not always protective against PD. Some narcissistic and antisocial persons can be self-sufficient, resourceful, and successful but still maladapted because of low Cooperativeness (i.e., intolerance of others, revengefulness, low empathy, etc.).

Once the probability for the presence of PD is established on the basis of character, temperament traits are used for differential diagnosis of subtypes. Membership in Cluster A (aloof, eccentric personalities) is most clearly determined by low Reward Dependence (described as social detachment). Membership in Cluster B (dramatic, erratic personalities) is most clearly determined by high Novelty Seeking (described as impulsivity). Membership in Cluster C (anxious, fearful personalities) is determined by high Harm Avoidance (described as fearfulness) (Table 26–20). Other data indicate that Persistence predicts Obsessive–Compulsive traits.

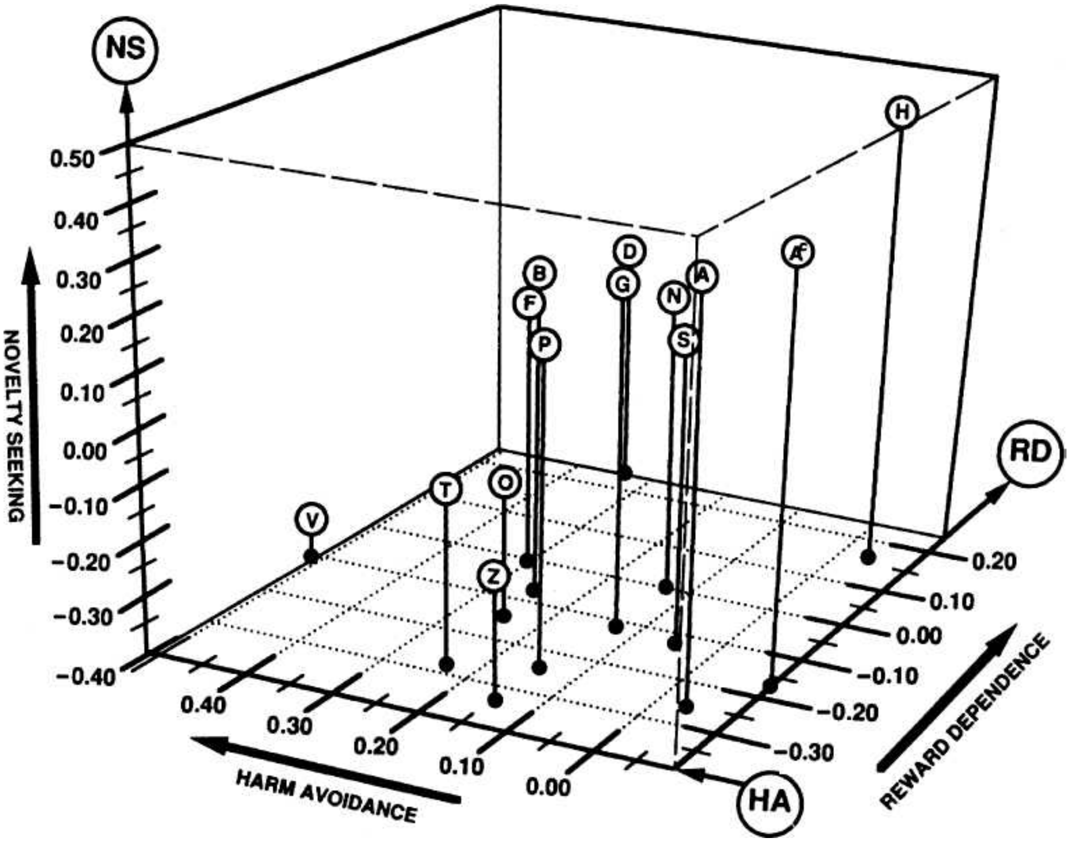

FIGURE 26–11. Correlations between individual PD subtypes and temperament traits of harm avoidance (HA), novelty seeking (NS), and reward dependence (RD). A, antisocial; Ac, antisocial children; B, borderline; D, dependent; F, self-defeating; G, passive-aggressive; H, histrionic; N, narcissistic; O, obsessive; P, paranoid; S, sadistic; GT, schizotypal; V, avoidant; Z, schizoid.

Table 26–20.
TCI Scales and Subscales

I. Temperament:

Harm Avoidance (HA)
- HA1: Worry and pessimism vs. uninhibited optimism
- HA2: Fear of uncertainty
- HA3: Shyness with strangers
- HA4: Fatigability and asthenia

Novelty Seeking (NS)
- NS1: Exploratory excitability vs. stoic rigidity
- NS2: Impulsiveness vs. reflection
- NS3: Extravagance vs. reserve
- NS4: Disorderliness vs. orderliness

Reward Dependence (RD)
- RD1: Sentimentality
- RD2: Sociability vs. aloofness
- RD3: Attachment vs. detachment
- RD4: Dependence vs. independence

Persistence (PS)
- PS1: Eagerness of effort vs. laziness
- PS2: Work hardened vs. spoiled
- PS3: Ambitiousness vs. underachieving
- PS4: Perfectionism vs. pragmatism

II. Character:

Self-Directedness (SD)
- SD1: Responsibility vs. blaming
- SD2: Purposefulness vs. lack of goal direction
- SD3: Resourcefulness vs. helplessness
- SD4: Self-acceptance vs. self-striving
- SD5: Spontaneity vs. lack of congruent second nature

Cooperativeness (CO)
- C1: Social acceptance vs. social intolerance
- C2: Empathy vs. social disinterest
- C3: Helpfulness vs. unhelpfulness
- C4: Compassion vs. revengefulness
- C5: Fairness/principles vs. self-serving opportunism

Self-Transcendence (ST)
- ST1: Self-forgetfulness vs. self-preoccupation
- ST2: Transpersonal identification vs. alienation
- ST3: Spiritual acceptance vs. skepticism
- ST4: Contemplation vs. objectivity
- ST5: Idealism vs. practicality

Individual subtypes of PD are distinguished on the basis of their unique pattern of correlations with temperament traits without any overlap (Fig. 26–11). In other words, the clinical presentation of PD varies along four underlying dimensions, corresponding to the four temperament traits of Harm Avoidance, Novelty Seeking, Reward Dependence, and Persistence (Table 26–16).

The authors' model does not require arbitrary cutoffs for the presence or absence of PD. PD can be quantified in terms of the degree of severity or probability of diagnosis can be estimated. As shown in Table 26–21, the probability for the presence or absence of PD based on the Self-Directedness scores alone ranges from very low (@ 10%) to very high (@ 95%) in clinical samples.

The authors' model "reconciles" some aspects of the categorical and dimensional approach to PD. It specifies seven temperament and character ***dimensions*** to account for individual differences in behavior. These dimensions interact to create temperament and character composites (types, profiles, or configurations), which correspond to **categorical** DSM subtypes of PD (Figs. 26–6 and 26–7). These multidimensional profiles also correspond to the clusters of people with particular profiles of traits associated with sets of genes identified in person-centered GWAS, as noted earlier. In other words, this model is dimensional but still can be used for categorical clinical diagnosis because each temperament and character composite profile corresponds to one clinical category of PD (in practice, this is useful for purposes of description and prediction, of course). These multidimensional profiles are related to underlying causal variables strongly but in complex ways.

Table 26–21.
Predictions of Personality Disorders from Self-Directedness Scores[a]

Self-Directedness Scores	Predicted with PD (%)	Observed with PD	
		f/N	%
44–39	13	4/18	22 ± 0.10
38–34	23	5/17	29 ± 0.11
33–30	33	4/16	25 ± 0.11
29–27	44	8/20	40 ± 0.11
26–23	55	8/19	42 ± 0.11
22–20	66	6/12	50 ± 0.14
19–16	70	14/16	88 ± 0.08
<16	85	17/18	94 ± 0.05

[a] Predicted % = 100 $e^y/(1 + e^y)$, where y = –3.37 +.13 (SD score).

PATHOGENESIS OF PERSONALITY DISORDER

Only a few developmental models of PD are available. Of those, the psychodynamic model focuses on the effects of early traumatic experiences and a constitutionally high disposition to anxiety and anger, whereas the spectrum hypothesis suggests that PDs may be a minor form of particular major clinical syndromes. A growing body of evidence points to nonlinear pathogenesis with multiple factors, including genetic, physiologic, psychological, social, and spiritual influences, which are fully incorporated into the psychobiological model of PD.

The Spectrum Disorder Hypothesis

Echoing Kretchmer’s work from many decades ago, some experts still consider subtypes of PDs as “minor” variants of “major” clinical syndromes. Indeed, some clinical disorders have variants in PDs. For example, Schizophrenia is represented by Schizoid PD, Delusional Disorder by Paranoid PD, Social Phobia by Avoidant PD, Cyclothymia by the group of emotionally unstable Cluster B PDs, and so forth (see Table 26–17).

These disorders are symptomatically similar and frequently differentiated from PDs only by the severity of their clinical presentation. The etiopathogenetic nature of this major-minor relationship has not been established. PDs might be predisposing factors, premorbid features, or subclinical expression of other major clinical disorders. Their common etiology, however, is still an open question. An etiologic relationship has only been supported for Schizotypal personality and the spectrum of Schizophrenia.

Psychodynamic Model of Personality Disorder

The central idea behind this model is that PD subtypes, classified as separate nosologic units, reflect different behavioral expressions of the same core deficit in personality. The so-called “borderline level of personality organization” described by Kernberg has been postulated to represent the nucleus shared by most subtypes of PD.

As defined by Kernberg (1975), the borderline level of personality organization is characterized by ***nonspecific manifestations of Ego weakness*** (such as lack of impulse control, lack of anxiety tolerance, and lack of sublimatory potentials), ***specific Ego defects*** (manifested as partially blurred self-object boundaries, mild to moderate primary process thinking, periodically distorted reality testing), ***partial object relations*** (manifested as alternations between “all good” and “all bad” perception of the self and external objects), ***primitive defense mechanisms*** (centered around splitting), and ***identity disturbance.*** In addition, an inadequately developed Super-Ego is either sadistic (e.g., obsessive personalities) or rigid in some but completely absent in other areas, permitting conflict-free expression of impulses and behaviors.

Immature defenses and a fragmentary self-concept (which are very similar to the above-described borderline personality organization) are a normal early phase in mental development. In the case of PD, however, this early developmental phase pathologically persists in later periods. Etiologically, this pathologic persistence of immaturity is postulated to reflect either constitutional factors (e.g., poor anxiety tolerance, high aggressivity, genetic vulnerability to certain affects) or traumatic environmental factors (e.g., early separation, loss of a parent, physical or mental abuse, neglect).

According to this dualistic theory, because of such traumatic etiologic factors, aggressively charged, negative representations of the self and external objects are predominately incorporated into the internal world. This, in turn, interferes with the crystallization of the early Ego, which normally occurs around positive, libidinally cathected representations of the self and external objects. In other words, normal early motivation and growth are centered around positive primary emotions, particularly interest and joy. The relative predominance of strong negative emotions, particularly anger and fear, interferes with normal character development. As a consequence, primitive defenses (splitting and the related defenses), which are normally predominant in this early phase of life, pathologically persist in the inner world and interfere with normal development. More mature defenses, such as repression, require more energy for their operation than splitting (e.g., repression requires at least as much energy for its operation as is contained in unacceptable impulses). In contrast, splitting keeps opposite “good” and “bad” self and object representations apart from each other. As Kernberg pointed out, splitting and early Ego weakness perpetuate each other. By protecting positive self and object representations, that is, keeping them apart from their negative analogs, primitive defenses ensure at least some development around positive experiences but interfere with neutralization of aggression and fear. These “pregenital” emotions and their related impulses (or a chaotic combination of pregenital and genital ones) persist as dominant motivators of behavior and significantly interfere with normal personality development.

The Borderline Core of Personality Disorder. Kernberg describes the borderline personality organization as follows:

1. Chronic free-floating anxiety
2. Polysymptomatic neurosis (e.g., multiple phobias related to one’s body or appearance, bizarre conversion symptoms, dissociative reactions, hypochondriasis)

3. Polymorphous perverse sexual trends (e.g., the coexistence of genital and pregenital elements, bizarre forms of perversions involving aggression)
4. Poor impulse control and addictions
5. Shift toward primary process thinking (“magical” thinking, primitive fantasies)
6. Partially impaired reality testing

Note, however, that the aforementioned symptoms are not unique to individuals with borderline personality but are shared by other subtypes of PD. Reich and Frances have shown that “the diagnosis of borderline personality organization is essentially equivalent to the diagnosis of PD (any subtype).” Likewise, a hierarchical cluster analysis of identical and fraternal twins, their siblings and parents, showed that 12 DSM PD subtypes can be grouped into eight heritable syndromes, which can be further aggregated into one group closely corresponding to the above-described borderline personality organization.

Perry and Vaillant wrote: “Just as fever, pus, and callus formation are the body’s physiologic reactions to insults of disease, personality disorders reflect persons’ efforts to heal themselves.” Indeed, a PD reflects a person’s attempt to overcome its underlying fragility through purposeful but maladaptive behaviors. These maladaptive behaviors are referred to as “purposeful” because they compensate for the fragility and proneness to decompensation associated with the borderline nucleus. Persons with PD become less fragile because the personality deviation (e.g., narcissism) takes over as the principal motivator of behaviors but more abnormal because an already immature personality is further impaired by additional distortions of self- and object representations.

These compensatory behaviors can have different types of symptomatic setups (e.g., avoidant, paranoid, antisocial, schizoid, dependent). Note that compensatory behavior styles help to classify a syndrome regarding its descriptive label (e.g., avoidant, antisocial) but tell little about the real character of the disorder underlying the descriptive label. Compensatory behavior styles develop as a function of both internal factors (e.g., fearfulness causes avoidant symptoms) and external factors (e.g., social class shapes antisocial behaviors). In other words, depending on certain constitutional and environmental factors, persons with borderline level of personality organization “create” the clinical picture of their PD.

A subgroup of borderline patients manifests the core symptoms, for example, stormy affects, vague identity, unstable relationships in a rather stable way (“stable instability”). These patients are fragile and prone to fragmentation but do not develop any compensatory behavior facade to protect their fragility. Some of them can periodically “mimic” behaviors of any PD subtype (avoidant, antisocial, obsessive) as a temporary, “chameleon-like” solution to their chronic internal problems.

The described compensatory process generates a two-level personality structure with the dominant compensatory self-concept and the split-off real self-concept. Consequently, the clinical expression of PD is dominated by overt, prototypical behaviors, but a set of subtle and covert symptoms and signs associated with the underlying borderline nucleus are regularly observed in the clinical picture. These difficult-to-detect symptoms are overlooked in the DSM system as diagnostic criteria of PD. For example, the clinical presentation of the narcissistic personality is dominated by the grandiose self, a deviant self-concept that generates an unrealistic sense of one's specialness, entitlement, arrogance, and other prototypical narcissistic behaviors (included in the DSM description of this disorder). The hidden real self of a narcissistic person is recognizable through chronic feelings of inferiority, hypochondriasis, envy, and pessimism that are not included in the DSM description of this disorder, even though they regularly accompany and sometimes even dominate the clinical presentation of pathologic narcissism.

Some Conceptual and Methodologic Issues. The borderline level of personality organization can be reliably distinguished from the schizophrenic spectrum disorders but less reliably from affective syndromes. The same immature personality processes that increase the susceptibility to deviant behaviors may also increase the susceptibility to mood disorders. In many cases, mood disorders and PD are comorbid and can be interwoven to the extent that no meaningful distinction between them can be made (as implied in the concepts of depressive personality, hysteroid dysphoria, or characterologic depression).

An increasing number of psychometric tests to improve the reliable measurement of psychodynamic constructs are available (e.g., the Defense Style Questionnaire and the Structural Interview of Personality Organization). Satisfactory diagnostic stability and discriminant validity in patients diagnosed as "borderline personality organization" has been demonstrated.

Both the psychodynamic model and the psychobiological model define a common denominator ("the core feature"), which extends across discrete subtypes of personality syndromes. This core feature is the borderline personality organization in the psychodynamic model and poorly developed character traits, particularly Self-Directedness, in the psychobiological model. Symptomatically, low Self-Directedness and borderline personality organization are for all practical purposes identical. For example, individuals with low Self-Directedness are described as blaming, helpless, irresponsible, unreliable, with no internal direction, and with difficulties in defining and pursuing stable goals (rather, they tend to experience numerous short-term, frequently mutually exclusive motives, none of which can develop to the point of real personal significance). It has been shown that low Self-Directedness correlates highly with the use of immature defense mechanisms that are central in the psychodynamic concept of borderline personality. Etiologically, both concepts postulate that excessive negative emotionality, such as fear and anger (or high Harm Avoidance and high Novelty Seeking), interferes with developmental processes and increases the risk of PD. The psychodynamic concepts can be incorporated into the more comprehensive biopsychosocial understanding of PD based on a solid objective foundation from genetics and functional brain imaging that were not available when the psychodynamic concepts were initially developed.

Psychobiological Summary of Personality Disorder

Psychodynamic understanding of PD, as shown previously, was a useful first step in the understanding of human behaviors. However, the psychodynamic model neither accounts for nor provides directions for expanding the current understanding of the full etiologic complexity of PD. In other words, as an etiologic theory, the psychodynamic model addresses only a portion of the iceberg representing deviant behaviors. As a treatment tool, manualized psychotherapies using different techniques tend to yield remarkably similar benefits, as expected from nearly all the variance in the outcome being attributed to the intersubjective and person-centered qualities of the therapeutic alliance, as discussed earlier. Psychodynamic strategies are an integral part of the comprehensive psychobiological approach to the treatment of PD presented in this chapter, but optimal treatment outcomes depend on effectively addressing all three human systems of learning and memory in order to help people to integrate their habits and skills with meaningful goals and self-transcendent values.

Modern psychobiology of behaviors provides an integrative, multifactorial, and etiologic model of the development of personality and its disorders. The biopsychosocial approach is based on four fundamental perspectives: genetic, neurophysiologic, learning, and phenomenologic that interact in a nonlinear way to produce phenotypic differences in behavior traits, cognitive style, and self-awareness. The psychobiology and sociocultural influences on temperament and character traits are discussed throughout this chapter (see Tables 26–2 through 26–10 for review). The adaptive self-organization of character is discussed in the section on development. Diagnostic and treatment implications are discussed in the corresponding sections. The following section summarizes the biopsychosocial concept of PD.

Multiple genetic and environmental factors interact in complex ways to influence the risk of PD. As a reminder, roughly half of the variance in temperament is explained by over 700 genes, which are nearly all protein-coding genes that mediate behavioral conditioning. It is the complex interactions of multiple genes with one another and with environmental variables, not the average effects of individual genes, that account for variability in personality. These temperament traits determine one's susceptibility to specific neurochemical processes that in turn underlie individual differences in basic emotions and influence early learning characteristics. These temperament traits have distinct genetic, neurobiologic, and phenomenologic correlates (Tables 26–2 through 26–6).

Character represents the higher cognitive processes that allow self-regulation of temperament and is also influenced by many genes, including many regulatory RNA genes that influence intentional self-control and creative self-awareness. These RNA genes regulate the expression of genes that encode proteins so that changes in self-awareness can actually allow a person to modify the expression of the genes the influence their physical, mental, and social well-being. Temperament and character traits reciprocally influence each other, as a person adapts to changes in their internal and external milieu and learns how to express their identity in satisfying and meaningful ways. As shown in Table 26–22, various temperament types affect differentially one's risk of immature character and PD. Some configurations, mostly those with high Reward Dependence, are protective against PD, whereas some increase this risk (e.g., the explosive or borderline profile with low Reward Dependence, and high Novelty Seeking and Harm Avoidance). Contrary to the common belief, average scores on the temperament dimensions do not protect against maladaptation and immaturity. People with average temperament traits have an average (not a decreased) risk of PD (Table 26–22).

Extreme temperament variants are not necessarily indicative of PD. They are expected to be associated with long-term personal, social, and/or occupational impairments that warrant the PD diagnosis ***only when accompanied by inadequate development of the three character traits of Self-directedness, Cooperativeness, and Self-transcendence.*** In other words, poorly developed character is what makes some behavior traits maladaptive and increases the risk of PD. An individual high in Novelty Seeking and low in Harm Avoidance may have an impulsive PD if they are low in Self-Directedness and Cooperativeness, or may be an inquisitive scientist without PD if they develop a creative character (i.e., are well developed in all three character traits). A creative character profile provides the intuitive insight and judgment to have a unifying outlook on life characterized by understanding the virtues of hope (i.e., Self-directedness), love (i.e., Cooperativeness), and faith (i.e., Self-transcendence), which leads to physical, mental, social, and spiritual well-being. In contrast, individuals who are emotionally reactive have an unconscious outlook of separation characterized by fear, unmet desires, self-doubt, and shame (sometimes compensated by vanity). Intermediate between these extremes, individuals with organized character profiles (i.e., high Self-directedness and Cooperativeness but low Self-transcendence) are individualistic and materialistic, seeking to maximize their wealth, power, and fame without recognition that their lack of identification with other people and the world is ultimately self-defeating because their struggles and desires are insatiable.

Table 26–22.
Relative Risk of Personality Disorder as a Function of Temperament Type in a Sample from the General Community

Temperament Type	Configuration	*N*	Immature (%)
High risk			
Borderline	NHr	39	72
Obsessional	nHr	44	59
Antisocial	Nhr	25	48
Passive–aggressive	NHR	30	40
Average risk			
Average profile	—	15	33
Low risk			
Avoidant	nHR	30	17
Schizoid	nhr	31	16
Histrionic	NhR	50	12
Reliable	nhR	36	6
Total		300	33

Lower case letters (h, n, r) and capital letters (H, N, R) indicate low and high values for temperament traits.

The concepts of temperament and character are essential to decompose the symptoms of PD into the ***common features*** shared by all PDs, and ***distinguishing features*** that permit the discrimination among PD clusters and individual subtypes. This is not possible for models that confound temperament and character. This also indicates that PD reflects deviations in both temperament and character, that is, in both emotional and conceptual personality processes.

Emotionally reactive individuals with little character development are often diagnosed with PDs. Such individuals have low physical, mental, and social well-being, but this may still represent a point of developmental stability for many individuals, even though it is a suboptimal level of adaptation. Such suboptimal outcomes, which are clearly maladaptive relative to the possible maximum for that individual, tend to be stable. The resistance to change of suboptimal personality configurations occurs because substantial improvement in well-being often requires a person to give up immediate gratification in order to achieve something of greater ultimate value (e.g., a cocaine addict giving up cocaine in order to establish a responsible and sober life). Therefore, growth in personality and well-being requires much more than the conditioning of habits and skills or advice about healthy goals; personal growth in well-being depends strongly on growth in self-awareness and particularly in self-transcendence so that a person can imagine their future in vivid detail and commit to achieving their valued goals with perseverance.

Behavior genetic research has demonstrated the importance of gene–environment interactions in personality development. For example, prospective studies of birth cohorts in Finland have demonstrated that specific genetic variants of the dopamine DRD2 and DRD4 receptor influence the development of Novelty Seeking but only in individuals exposed to harsh and inconsistent discipline by emotionally distant parents. Behavior geneticists have demonstrated that the effect of sociocultural factors on personality appears less specific than that of genetic factors, influencing success in adaptation rather than its form or personality style. Hence, psychosocial disorganization in the rearing environment of a child has a substantial influence on the risk of PDs. This is essential for preventive strategies as even temperament configurations with a high risk of PD might be overcome in homes and communities that provide security and limits on behavior in a warm, compassionate manner, as well as encouraging self-directed choice and the value of respect for other people.

PD is associated with younger age. This indicates maturation (i.e., remission of deviant behaviors) with increasing age. In general, three dimensions of personality change substantially with age. Novelty Seeking decreases with age by approximately 18% so that older individuals become less impulsive (more reflective), less rule-breaking (more orderly), and less quick-tempered (more stoical). Cooperativeness increases markedly in most children during school age and then increases by 12% on average after the age of 18 years. Self-Directedness increases markedly in most people during adolescence and young adulthood, increasing on average by 9% after the age of 18 years. Prospective studies with four measurements from the ages of 20 years to 45 years show that Self-Directedness and Cooperativeness increase substantially with increasing age, whereas Self-Transcendence decreases between the ages of 20 and 45 years before increasing again as people must face ultimate situations like their own mortality.

The decreasing prevalence of PD with age is attributable to the increased development of both Self-Directedness and Cooperativeness with age. The additional tendency for Novelty Seeking to decrease with age explains the finding that patients with impulsive PDs show more improvement than those with anxious or eccentric PDs. The best-documented finding of a change in deviant behaviors is the remission of criminal behavior in individuals with Antisocial PD, which usually occurs if and when they get a good job, find a loving spouse, or experience a spiritual conversion. These individuals nearly always remained impulsive (high Novelty Seeking), risk-taking (low Harm Avoidance), and aloof (low Reward Dependence) but became mature enough to maintain work and family life in a stable and sober manner.

TREATMENT

Stages in the Development of Self-Awareness and Well-Being

A full assessment of personality requires consideration of a person's level of self-awareness and well-being, not just their impairments. Health and well-being are more than the absence of deviant traits. Well-being depends on a person's level of self-awareness and leads to the expression of human virtues and positive emotions that go beyond what is average in contemporary society.

There are three major stages of self-awareness along the path to well-being, as summarized in Table 26–23, based on extensive work by many people. The absence of self-awareness occurs in severe PDs and psychoses in which there is little or no insightful awareness of the preverbal outlook or beliefs and interpretations that automatically lead to emotional drives and actions. Lacking self-awareness, people act on their immediate likes and dislikes, which is usually described as an immature or "childlike" ego state.

The first stage of self-awareness is typical of most adults most of the time. Ordinary adult cognition involves a capacity to delay gratification in order to attain personal goals but remains egocentric and defensive. Ordinary adult cognition is associated with frequent distress when attachments and desires are frustrated. Hence, the average person can function well under good conditions but may frequently experience problems under stress. At this stage of self-awareness, a person is able to make a choice to relax and let go of their negative emotions, thereby setting the stage for acceptance of reality and movement to higher stages of coherent understanding. Therefore, patients who are disturbed by negative emotions and unmet desires can be taught to practice various skills for relaxation to help them enter the first stage of self-awareness. There are a wide variety of effective relaxation techniques, including deep and slow abdominal breathing to activate the vagal nerve, one-pointed concentrated meditation increasing vigilance, nature walks or listening to harmonious music.

Table 26–23.
Three Stages of Self-Awareness on the Path to Well-Being

Stage	Description	Psychological Characteristics
0	Unaware	Immature, seeking immediate gratification (“childlike” ego state)
1	Average adult cognition	Purposeful but egocentric; able to delay gratification but has frequent negative emotions (anxiety, anger, disgust) (“adult” ego state)
2	Metacognition	Mature and allocentric; aware of own subconscious thinking; calm and patient, so able to supervise conflicts and relationships (“parental” ego state, “mindfulness”)
3	Contemplation	Effortless calm; impartial awareness; wise, creative, and loving; able to access what was previously unconscious as needed without effort or distress (“state of well-being,” “soulfulness”)

Adapted from Cloninger CR. ***Feeling Good: The Science of Well-Being***. Oxford University Press; 2004:219–225.

The second stage of self-aware consciousness is typical of adults when they operate like a “good parent.” Good parents are allocentric in perspective—that is, they are “other-centered” and capable of calmly considering the perspective and needs of their children and other people in a balanced way that leads to satisfaction and harmony. This state is experienced when a person is able to observe their own subconscious thoughts and consider the thought processes of others in a similar way to observing their own thoughts. In this state, a person is interested in understanding their thoughts without judging or blaming, which requires that they are already in a state of relaxation and acceptance. Hence, the second stage is described as “metacognitive” awareness, mindfulness, or “mentalizing.” The ability of the mind to observe itself allows for more flexibility in action by reducing dichotomous thinking. At this stage, a person is able to observe himself and others for understanding, without judging or blaming. However, in a mindful state, people still experience the emotions that emerge from a dualistic perspective, so mindfulness is only moderately effective in improving well-being.

The third stage of self-awareness is called contemplation because it is the direct perception of one’s initial perspective—that is, the preverbal outlook or schemas that direct one’s attention and provide the frame that organizes one’s expectations, attitudes, and interpretation of events. Direct awareness of one’s outlook allows the enlarging of consciousness by accessing previously unconscious material, thereby letting go of wishful thinking and the impartial questioning of basic assumptions and core beliefs about life, such as “I am helpless,” “I am unlovable,” or “faith is an illusion.” In other words, contemplation is a reversal of the process of repression: contemplation enlarges consciousness, whereas repression reduces it. The third stage of self-awareness can be described as contemplative because, in this state, a person becomes aware of deep preverbal feelings that emerge spontaneously from a unitive perspective, such as hope, compassion, and reverence. To enter this state from mindfulness, a person begins to recognize that their perceptions may be biased by unconscious information so that they can begin to examine and question what they had previously assumed to be an impartial perspective. Contemplation is much more powerful in transforming personality than is mindfulness, which often fails to reduce feelings of hopelessness because it allows enlargement of consciousness to provide a broader perspective for flexible, prosocial, and creative adaptation.

Extensive empirical work has shown that movement through these stages of development can be described and quantified in terms of steps in character development or psychosocial development, as in the work of Vaillant on Erikson’s stages of ego development. As previously mentioned, such development can be visualized as a spiral of expanding height, width, and depth as a person matures or increases in the coherence of personality. Likewise, the movement of thought from week to week or month to month has the same spiral form regardless of the timescale. Such “self-similarity” in form regardless of timescale is a property characteristic of complex adaptive systems, which are typical of living processes in general, whether they are psychosocial or neurobiologic. The clinical utility of this property is that therapists can teach people to exercise their capacity for self-awareness, moving through each of the stages of awareness just described. Their ability to do so, and the difficulties they have, reveals the way they are able to face challenges in life. Cloninger has developed an exercise called the “Silence of the Mind” meditation, with explicit instructions to take people through each of the stages of awareness as well as they can. The first phase of this meditation results in a relaxed state in the first stage of self-awareness. The second phase facilitates entry into the second stage of self-awareness, and the third phase into the third stage of self-awareness, if the person is able to do so. Using this and a way of observing thought during the mental status examination, mental health professionals can assess a person’s thought and its level of coherence in a way that is constructive, easy, and precise without being judgmental.

Each step of growth in self-awareness and well-being requires particular psychobiological conditions. The emotional dynamics of growth in well-being are depicted in Figure 26–12 as a transition within a nonlinear dynamical system. The transition from one state of lower well-being (state A) to a state of higher well-being (state B) requires growth in self-awareness. Each such transition requires a person to accept the limitations of the assumptions they have been making about what is satisfying and valuable in order to begin adapting in ways that are more coherent and flexible. However, such growth requires a person to let go of what is familiar that so there is a temporary increase in negative emotions before there is further growth in positive emotions and other aspects of well-being. Therefore, an effective clinician must help provide appropriate conditions to facilitate this transition ("the energy of activation") with their own compassion, hope, patience, and awareness, realizing that the patient must ultimately be the one who does the work to grow in self-understanding from an outlook of separation with frequent negative emotions and self-defeating behaviors toward an outlook of unity with frequent positive emotions and self-actualizing behaviors. Each step along the pathway to well-being requires different procedures to facilitate growth because different aspects of the being are involved in the different steps in character development. These specific techniques for the various steps in development are described in what follows.

General Principles of Treating Personality Disorder

At the outset of treatment of PD, and as a general principle, the psychiatrist should establish compassionate but professional boundaries while seeking a healthy and strong working alliance with the patient. The psychiatrist should endeavor, as often as possible, to be clear, concise, and consistent in the face of the many unconscious and conscious attempts to derail a healthy treatment process that can occur when working with individuals struggling with PD. Less is often more unless responding to an acute crisis situation (discussed later in this chapter) and, even in the acute crisis situation, the psychiatrist must guard against overreaction and catastrophizing, creating an appearance to the individual of the potential for "giving-up" on the patient. The psychiatrist working alone will have to navigate the challenges of being both a "good object" and "bad object" while retaining a consistency of approach, reality testing where appropriate, and maintaining the therapeutic alliance. The risk of boundary issues and burn-out can be mitigated by the use of a consultant or supervisor to reflect regularly on the case. Alternatively, creating a team approach can also help to avoid these pitfalls.

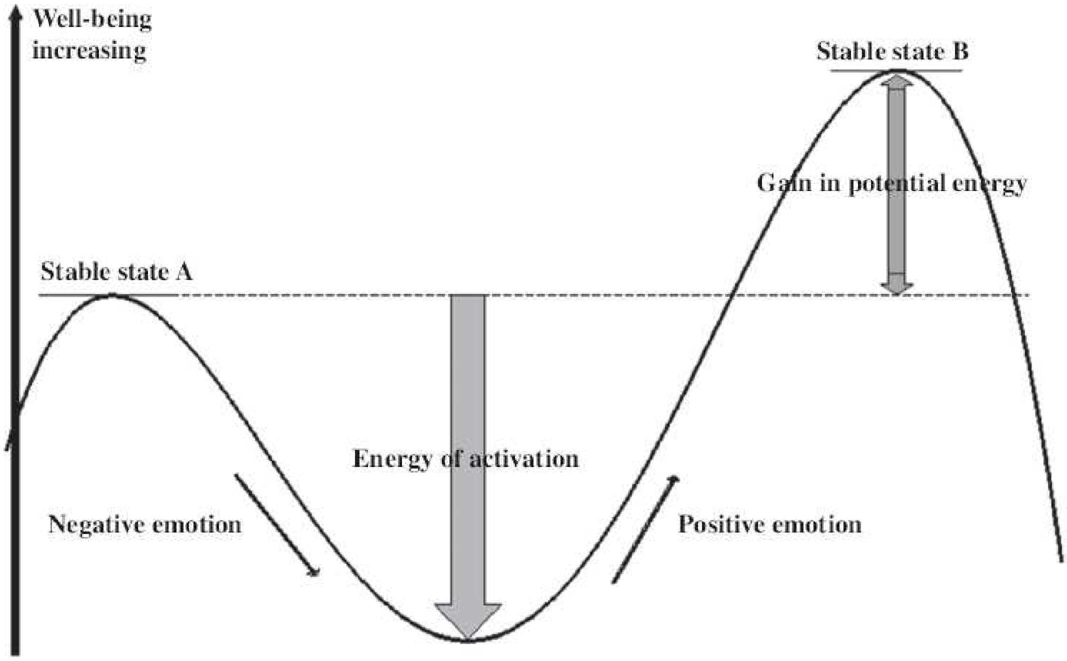

FIGURE 26–12. The emotional dynamics of well-being as successive transitions in a complex adaptive system. (Used with permission of Anthropedia Foundation.)

In general, the undertaking of the treatment of PD is most effective when the psychiatrist is part of a highly coordinated team. This may include therapists, primary care physicians, a well-being coach, close family members, and other specialist practitioners. This provides the surest way to avoid making the most common errors such as those that arise from countertransference issues (discussed later in this chapter), supports consistency in reinforcing boundaries, and aids in a creative person-centered treatment response by having “many pairs of eyes” on the situation. Treatment of PD should be approached as a “marathon,” not a “sprint,” and building a care team and support network allows for greater longevity in the treatment approach. A strong emphasis on coordination and communication among the team allows for the consistency of this approach. Therefore, at times it may require that up to 80% of the time and energy devoted to helping an individual with PD is spent “backstage” coordinating with the support network, with the remaining 20% in front of the individual. However, this will change over time as the team grows in its confidence to provide clear, concise, and consistent communication with the patient, who is at the center of this treatment process. Third-party payer systems often do not reimburse for time spent in care coordination, and thus increasing numbers of psychiatrists choose to opt out of insurance systems in the interest of optimal care.

Medication is only one of many tools that will be used by the psychiatrist in the treatment of PD. The array of symptoms and comorbid disorders means that there are multiple pharmacologic options that may be considered (discussed in more detail later in this chapter). However, it is important to keep some general principles in mind when prescribing for patients with PD. Medication can often become a kind of “transitional object” for the immature personality, and as such, the psychiatrist must try to avoid making medication the entire focus of treatment or building the therapeutic relationship solely around prescribing. This only enables the avoidance of addressing and prioritizing what the patient (and clinician) may see as the more difficult challenges, such as self-regulation, goal setting, and expansion of self-aware consciousness. Psychiatrists can all too easily be seduced into focusing on what they see as the “easier” or concrete issue of dosages and medication compliance, unintentionally cultivating dependence and undermining the development of autonomy and responsibility, which are at the foundation of long-term sustainable change in PD. On the other hand, by holding the patient accountable for the reasonable and responsible use of a medication regimen, one that they can take ownership of, medication will be part of the overall maturation of the individual.

Another common challenge in the use of medication in PD, resulting from the presence of multiple comorbidities, is the frequent use of poly-pharmacy and all the inherent risks it presents. Again, if approached in a person-centered way, this can be an opportunity for building responsibility by both complex reconciling regimens and creating ways for the patient to be accountable for compliance.

The work of Zwir, Cloninger, and colleagues on the underlying genetic architecture of personality helps to orient the treatment process around some general principles of character maturation involving stages of learning. During the “emotional-regulation” phase, the patient may need “containment” and strong boundaries but will also benefit from the consistency of the individual treatment setting and/or the unity of the treatment team. By receiving clear and consistent messages, with “everyone on the same page,” the patient is better placed to experience the satisfaction of moderating their own emotional responses and diminishing the use of primitive defenses such as splitting. While the patient may initially try to project their own outlook of separation onto the psychiatrist or the clinical team (e.g., all nurses are treated as “good” and all social workers are treated as “bad,” or vice versa), diligence and/or team work, presenting a concise unified message, will extinguish much of this behavior and often remarkably quickly. The psychiatrist, sometimes viewed as an “authority figure,” is critical in the process of building team cohesiveness and coherence. The so-called “third-wave” cognitive therapies, such as Dialectical Behavioral Therapy (DBT), may also be incorporated as part of the treatment plan for this phase of development.

During the “self-control” phase, the use of coaching techniques and a person-centered approach will help the patient take ownership of their treatment plan through accountability, goal setting and following through on daily objectives, working in the service of others, building empathy, and cultivating a more allocentric outlook rather than egocentricity. With this, the “victim” tendency of the patient begins to diminish, and with frequent positive feedback, the patient beings to grow in hope and motivation for recovery.

During the "self-awareness" phase, the patient is helped to gain perspective. This can come from a wide variety of sources. The use of narrative therapy, time in nature, meditation, and acting in awareness of our interconnectedness with the universe all help to give the patient a greater sense of purpose and helps the patient to grow in free will. The well-being coaching program created by the Anthropedia Foundation is a rare and unique example of an internationally validated, cross-cultural program that specifically addresses self-transcendence and offers concrete tools to cultivate growth in awareness as well as growth in character more generally. Such a program could be usefully integrated into a team approach when addressing this phase.

It is of utmost importance for the psychiatrist to recognize that these treatment "phases" are stages of learning and, like the stages of change (Transtheoretical Model of Prochaska and DiClemente) or the stages of self-awareness discussed earlier (Biopsychosocial Model of Cloninger), are not a simple linear progression. Rather, the psychiatrist must see them as part of an upwards "spiral" of growth. Each phase is interconnected, and the treatment emphasis may move from one to the other and back again as the patient experiences maturation of character and growth in self-control. An upwards spiral means a return to the "same place" will be a frequent and inevitable feature of progress. It is crucial to point out to the patient, and the clinical team, that this is not back to "square-one" but a return to familiar challenges with an incrementally elevated perspective and an opportunity to make new and more satisfying choices. The feelings of emptiness and powerlessness, so common in PDs, can markedly diminish when the patient understands and appreciates this upwards journey and the incremental improvements in problem-solving abilities that accompany it.

To partially summarize the general principles of treating PD, we can turn to an inspiring quote sometimes attributed to Plato. "Be kind, for everyone you meet is fighting a hard battle." The psychiatrist needs to seek all possible ways to remain kind in the face of what can sometimes be overwhelming despair on the part of the patient. Team support, consistent boundaries, clear communication, and a person-centered approach using the multiple available therapeutic tools will all help the clinician to remain calm and compassionate and to stay the course. Above all, however, the psychiatrist must retain perspective; everyone is struggling, and it is a hard battlefield. The psychiatrist is on their own path to well-being and can use the challenges of therapy with patients who have disordered personalities as a stimulus for their own growth in humility, compassion, and expansion of awareness. The rise in the prevalence of psychopathology of personality, which has become perhaps the dominant clinical challenge of our time, cannot be separated from the rise of an individualist consumerist society that encourages unlimited desire and celebrates the hubris of unregulated self-expression that treats others as objects to be controlled. When considering the maturation of a personality in such a chaotic environment, it seems perhaps remarkable that humans are able to find joy and wisdom at all, and thus we can marvel at the dignity and inspiration of the natural tendency for growth and self-actualization present in every human being.

Primary Pathology and Secondary Symptoms of PDs

A common assumption is that pharmacotherapy impacts only the biologic aspect and psychotherapy only the psychological aspect of the human being. However, psychotherapy can modulate structural CNS characteristics, such as receptor activity and expression, through epigenetic mechanisms. Likewise, pharmacologic treatment of depressed mood or symptoms of psychosis can restore complex psychological functions, such as integration of identity. Some might consider it ironic that the strongest evidence for the effectiveness of psychotherapy comes from biologic neurosciences.

However, pharmacotherapy is not a substitute for psychotherapy because taking medication does not involve learning, although it may provide stabilized conditions that facilitate learning. For example, medications can be helpful in modulating anxiety or aggression but not in enhancing self-awareness or inter-subjective awareness. Unstable mood, intensive but unstable interpersonal relationships or risky and self-destructive behavior can be prominent symptoms observed in patients with PDs, but they are non-specific and secondary complications of the underlying problem that needs to be treated. Fortunately, the recent advances in deconstructing the complex structure of learning systems underlying human personality can now inform and guide treatment for PD.

Preventable Errors in Treating Individuals with Personality Disorders

There are three major barriers to effective treatment of PD, but all are preventable errors within the control of the health care provider. The first is the frequent loss of professional objectivity, signaled by the development of strong emotions (positive or negative) and called positive or negative countertransference. This inappropriate personal involvement is a red flag to reassess diagnosis and treatment. Frequent discussions and counseling with colleagues are useful because even strong countertransference feelings sometimes persist unrecognized.

The second preventable error in PD management is to believe the myth that PDs cannot be treated effectively. This myth is partly generated by countertransference problems of some professionals and then sustained by a failure to consider facts showing the effectiveness of treatment. Belief in the untreatability of a patient creates a self-fulfilling prophecy. For example, many controlled studies indicate that even severe PDs, such as borderline or antisocial

PDs, can be effectively treated with an appropriate condition, such as a cooperative therapeutic alliance.

The third preventable error in PD management is to give direct advice on personal and social problems. This is counterproductive in patients with PD because they usually become dependent, noncompliant, or resentful. Occasionally, direct advice may be offered to some antisocial, narcissistic, and schizoid patients who are at low risk of developing dependency and need precise structure and direction initially. However, it is most beneficial to provide guidance and support without giving direct advice. All that is usually required in supportive therapy is compassionate attention, respect for the dignity of other human beings despite their flaws, and reinforcement of the patient’s existing coping mechanisms that are adaptive.

When tempted to give direct advice to patients, remember that change in personality requires more than common sense and logic. If the relationship leads to frequent advice-giving, then referral to a psychiatrist or a psychologist may be indicated. People change if they become self-aware, usually from personal recognition of dissatisfaction with themselves and their relationships. Personal growth thus arises from new insights about oneself and the environment. Direct advice robs the patient of the opportunity to develop new insights and to learn from their mistakes. In summary, supporting these patients involves joint evaluation of options and encouragement to practice skills in solving problems.

Substantial personality change, which is invariably needed by people with PDs, involves an extensive growth in self-awareness and character development, which requires precise diagnostic analyses, specific treatment strategies, and expert training. The expert treatment may include any of the several available psychotherapy approaches and is usually combined with pharmacotherapy. The major points relevant to psychotherapy and pharmacotherapy of PDs are summarized later.

As already mentioned, individuals with PD have a peculiar capacity to elicit strong emotions from other people. They are often described as aggravating, unlikable, difficult, or bad. Alternatively, they may be seductive or dependent and elicit inappropriate emotions or actions, such as sexual interest or the urge to rescue. Even professionals may have difficulty treating them with respectful objectivity because of a blurring of personal boundaries. Such loss of objectivity occurs because the patient’s deeply felt assumptions about other people may often elicit interpersonal responses that are appropriate to the patient’s assumptions. People’s assumptions about themselves and others often become self-fulfilling prophecies because of automatic mechanisms of affect transfer. If someone smiles at you, communicating appreciation, it is natural to experience feelings of social attachment and to smile back automatically. Likewise, if someone frowns, communicating anger, it is natural to feel defensive in preparation for their angry attack. For example, many patients with PD are suspicious and hostile about others' motives. This distrustful attitude is communicated in many verbal and nonverbal ways and often elicits disagreement or frank hostility from others. These uncooperative responses reinforce the original negative assumptions of the patient, which in turn leads to further alienation.

This vicious cycle of affect transfer can be interrupted only by professional objectivity combined with patience and compassionate respect for the patients' disability. Such objectivity arises from recognizing the overall meaning and implications of their pattern of interpersonal signals so that their verbal and nonverbal communication takes on diagnostic and therapeutic, rather than personal, significance. However, objectivity and empathy are not sufficient and leave the therapist vulnerable to burnout with difficult patients. The necessary patience and compassion require that the therapist has developed their own awareness, including self-transcendence so that they can identify with the other without suffering when the other acts in maladaptive ways. In optimal therapeutic relationships, physicians should be compassionate, hopeful, and patient, and the person in treatment need only be able to identify something of value to seek from their treatment. Whenever professionals become aware of strong positive or negative emotions toward a patient (the so-called "countertransference" reactions), this should help alert them to the possibility that the patient has a PD and/or that the therapist has an opportunity to develop greater insight too, which can be extremely valuable. It is important to remember that both patient and therapist are learning in therapy because of the inter-subjective nature of an effective therapeutic alliance.

As many patients with PD do not recognize or admit their psychopathology, they resist and resent psychiatric diagnoses and any form of mental health treatment. Accordingly, it is prudent to let the patient define their treatment goals and then jointly evaluate the likelihood of successful outcome until treatment goals that both patient and the therapist agree upon can be identified. Initially, these goals should be as simple and concrete as possible (e.g., "to develop social skills" or "to reduce alcohol use"). In many but not all cases, successful completion of this initial phase will motivate the patient to define other, more complex treatment goals and to continue treatment.

A psychiatrist should keep in mind that there is a natural succession of stages in the treatment of patients with PDs. Each has different goals and requires different methods. The complete psychiatrist should be prepared to guide the patient along these stages, ever ready to advance to the next stage if the patient is interested and prepared to do so.

Four Major Stages in Treatment of Personality Disorders

The four stages in the treatment of a patient with PD can be described as (1) crisis management and stabilization, (2) awakening of a positive perspective and personal values in life, (3) other-centered awareness, and (4) integrated intelligence. The initial stage of crisis management and stabilization deals with the presenting problem and stressors in order to help the patient get into a calm enough state and a working alliance with the psychiatrist. The second stage involves elevating a person's outlook on life so that they can experience things they enjoy and value under relaxed conditions. In addition to establishing a trusting therapeutic alliance and a safe context in which a person can face their fears and other difficulties with realistic acceptance, the second phase involves an initial spiritual awakening that has often been neglected in strictly cognitive-behavioral or psychodynamic approaches but without which there is little capacity for fundamental change in the quality of life. The third stage of metacognitive (other-centered) awareness involves increases in self-awareness and capacity for contemplation that elevate a person's usual thoughts, feelings, and relationships in a wide range of conditions. The fourth stage of integration of reason and love in action allows a person to be mature and happy even under conditions that were previously stressful.

During most of the 20th century, many people were doubtful that people with PD could really get well. It was assumed that genetic predispositions and memories of past neglect or trauma were rarely transformed, so the improvement was limited to better defenses and coping strategies. Now research and treatment document that fundamental and lasting growth in insight and well-being is possible when a person is able to recollect their past and imagine options for their future vividly in self-awareness. The past can be reprocessed from a perspective of deeper insight, which gives reality a fresh and more hopeful meaning. People can and must create their own health in each moment, just as our memories are not immutable. Rather they are constantly reconstructed and reimagined as a living interpretation of reality. What we remember or imagine can be felt and reexperienced from different points of view without distorting reality, as occurs when listening to an unfolding narrative or a joke's punchline. Both humor ("ha-ha") and insight ("ah-ha") involve shifts in perspective that evoke a reintegration of feelings and thoughts with new meaning. Likewise, patients with PDs can pass through the natural stages of increasing awareness and insight on their own (i.e., remit spontaneously) or be guided through these stages in treatment facilitated by a scientifically designed set of physical, personal, social, cognitive, and spiritual exercises.

A General Approach to Psychotherapy for Personality Disorder

A useful approach at the initial stage of treatment is to focus primarily on the chief complaints that bring them to treatment. These complaints are often related to a person's work, relationships, and/or general health, which often offer an adequate basis for the diagnosis and initial treatment of PD. For example, a systematic focus on a person's work, school performance, or relationships can identify attitudes, feelings, and behaviors that help them recognize the barriers to their success and happiness. However, this can bring up many sensitive issues that they prefer to minimize or ignore or that lead them to lie and distort information about their real situation.

An alternative initial approach, particularly if they are entering treatment at the request or demand of someone else, is to focus on healthy lifestyle choices, which provides a non-threatening basis for the evaluation of a patient's goals, values, habits, and skills (i.e., their personality). Choices about diet, weight control, exercise, smoking, drinking, and ways of relaxing and managing stress are appropriate for discussion with one's physician and do not threaten or stigmatize the patient. Discussion of these choices with a patient can provide a guiding stimulus for developing more self-direction and constructive planning about life. Discipline in working toward chosen goals is an indicator of maturity. Success in accomplishing even simple goals builds self-confidence and often generalizes to other forms of self-directedness. Lack of success stimulates learning about goal setting and personal growth by, for example, breaking a problem into smaller steps to be taken one at a time. In this process, patients have the opportunity to learn from experience, which is essential for patients with PDs. In many cases, such goal setting and problem solving lead to the ability to admit faults and to recognize one's strengths and limitations. Respect for one's self from accomplishment requires acceptance of responsibility and leads to a trust of others. Self-respect and respect for others usually progress hand in hand. Such patient guidance is a simplified and non-threatening form of what is usually called cognitive-behavioral therapy (CBT) and can be safely practiced in a busy office practice with short but regular sessions.

Growth in self-awareness (i.e., increase in insight and judgment) is possible only once a person is calm enough to face and accept unpleasant realities about themselves. As a result, therapy usually proceeds by a series of small steps that lead to progressively more mature and integrated ways of viewing and coping with the challenges of life. Figuratively speaking, a person who feels fearful, hungry, and rejected cannot appreciate beauty and wonders while struggling to gratify basic needs for safety, food, and comfort. However, if they attempt to meet their most salient in ways that are inconsistent with health-promoting values, thoughts, and habits, then they are acting in a self-defeating manner. Consequently, it is difficult to specify the necessary or sufficient condition for the progress of any particular person because both internal and external resources need to be considered to identify a way to work toward personally valued goals that will promote increasing plasticity, virtue, and health.

There are no procedural recipes that assure treatment outcomes because growth in self-understanding depends on the dynamics of nonalgorithmic processes such as insight, judgment, and creative foresight. What a person desires depends as strongly on their worldview as on their external circumstances. A person’s worldview influences what they find important, meaningful, and satisfying. For example, one person may react to a disaster with posttraumatic stress, whereas another finds meaning in the opportunity to help others, thereby growing in self-awareness and well-being. In fact, internal psychological resources, rather than external material and social supports, satisfy a person’s needs to an increasing degree as they mature. People with PDs have limited internal psychological resources, so their psychotherapy must begin to awakening values that will promote well-being and a willingness to work for greater self-understanding.

What clinicians can do to guide their practice is to understand the three processes that provide the means to healthy personality development and the activities that activate those processes. As mentioned in the introduction, the three basic processes that interact in increasing well-being in everyone are (i) plasticity (i.e., the ability and willingness to change), (ii) virtue (i.e., having intuitive insight into what is good for yourself and others), and (iii) creative functioning (i.e., being innovative, resourceful, purposeful, and responsible so that our habits become congruent with our goals and values). The cultivation of self-awareness and eventually self-transcendence by these three processes provides a pathway that people can follow with freedom and flexibility with available resources and help at their own pace in order to develop a life that is healthy, happy, and good.

People are unlikely to work to achieve goals that they are uninterested in personally. Therefore we encourage people to explore what they value through reflection and meditation on the subject and to consider how much they act on what they value most in their daily lives. Modern lifestyles have put far too great an emphasis on things that are not satisfying or beneficial to the detriment of our health and well-being. Consequently, it is crucial for our own well-being to practice activities that most, if not all people, find effective for cultivating the expression of virtue in their lives. When people are uncertain about what they find satisfying and valuable, they can be invited, as an initial experiment in understanding well-being, to practice living in ways characteristic of people with high levels of well-being. People with creative characters (i.e., who are high in Self-directedness, Cooperativeness, and Self-Transcendence) have the highest levels of well-being. Regardless of their initial personality configuration, most people find it satisfying and beneficial to practice acts of love, hope, and faith, much as is described in positive psychology with acts of kindness or gratitude, because this stimulates people to realize what they personally value. These virtuous practices are broadly described as (1) working in the service of others (e.g., practicing acts of kindness, which expresses an understanding of love), letting go (e.g., letting go of resentments and fights with one’s self and others, which expresses an understanding of hope), and growing in awareness (e.g., acts of faith about what is not tangible or objectively proven, such as meditation to cultivate serenity, mindfulness, insight and impartial judgment, which express an understanding of faith and humble acceptance of the imperfection of the human condition). In addition, it is helpful to teach most people about the processes of thought so that they can be more aware of what biases our perceptions, intentions, and values, and thereby interferes with their functioning.

The Role of Medication in the Treatment of Personality Disorder

Medications are often helpful, but they always carry some risk of side effects, and the advantages and disadvantages must be carefully weighed to respect the patient's wishes and to help them be calm and organized enough for further growth in self-awareness. Medication never makes people more self-aware, which requires learning but can help create more calm and stable conditions for doing psychotherapy. In general, the use of medications should be considered temporary to stabilize the patient and then tapered as awareness and self-regulation improve. The complexity of psychopathology in patients with PD requires the psychiatrist to remain vigilant to idiosyncratic and unpredictable reactions, a high prevalence of side effects, a tendency toward impatience for benefits, and the loss of calm judgment resulting in impulsive decisions or prescribing in desperation.

During the initial stage of treating patients with PDs, medications are often used to target specific symptoms with the goals of relieving subjective distress, risky or self-destructive behaviors, and/or conflict with others, thereby paving the way for later stages of treatment that require calmness, safety, and a reduction in defensiveness, to facilitate growth in self-awareness.

Symptomatic pharmacotherapy is not focused on the individual subtypes of PD but rather on the following four symptom domains shared by all subtypes: (i) mood and anxiety dysregulation, related most strongly to Harm Avoidance, (ii) aggression and impulse control, related most strongly to Novelty Seeking, (iii) social and emotional detachment, related most strongly to Reward Dependence, and (iv) psychotic symptoms and cognitive distortions, related most strongly to intellectual reasoning and Persistence. Recommendations about pharmacotherapy of symptom domains shared by most PDs are summarized in Table 26–24.

Table 26–24.

Commonly Used Medications in Clinical Practice According to Target Symptoms of Personality Disorder

Target Symptom	First-Line Medication	Other Options (Carefully evaluate of risk–benefit ratio and use low doses where appropriate)	Not Recommended
I. Mood dysregulation and anxiety			
Chronic cognitive anxiety[a]	SSRI SNRI GABA Analog (e.g., valproates, pregabalin) Buspirone	MAOI Second Generation Antipsychotic (e.g., quetiapine) Third Generation Antipsychotics (aripiprazole) Clonazepam (short term only)	Benzodiazepine: Generally, not recommended because of the high risk of dependence, abuse and addiction If used, consider a benzodiazepine with a long half-life and use for short trials only: limited to a few days, occasional or intermittent use, or courses not exceeding 2 weeks.
Chronic somatic anxiety	SNRI (e.g., Duloxetine, venlafaxine) GABA Analog Beta-blocker	MAOI TCA	
Obsessions	SSRIs	Second Generation Antipsychotics (e.g., quetiapine) TCA (esp. clomipramine)	

Acute and severe anxiety	Clonazepam, (short term only) Mirtazapine Valproates	Lithium Second Generation Antipsychotics (e.g., quetiapine) Third Generation Antipsychotics (aripiprazole)	
Atypical depression, dysphoria, Classical depression	SSRI SNRI	MAOIs Second Generation Antipsychotics (e.g., quetiapine) Third Generation Antipsychotics (aripiprazole)	TCA: Generally, not recommended as depression may worsen in some patients and due to the risks of overdose.
Emotional lability, rapid cycling	Lithium Lamotrigine Valproates	Second Generation Antipsychotics (e.g., quetiapine) Third Generation Antipsychotics (aripiprazole)	Antidepressants: There is a risk of precipitating mania.

II. Aggression and impulsivity

Affective aggression (Quick temper with normal EEG)	Lithium SSRI Anticonvulsants (e.g., valproates, lamotrigine, carbamazepine)	Second Generation Antipsychotics (e.g., olanzapine, quetiapine)	Benzodiazepine: Generally, not recommended as it frequently induces disinhibition.
Predatory aggression (cold-blooded revenge/cruelty)	No effective pharmacologic treatment.	Second Generation Antipsychotics Lithium	

		Valproates Beta-blocker	
Organic-like aggression (traumatic brain injury)	Beta-blocker Carbamazepine Cholinesterase inhibitor (e.g., donepezil)	Second Generation Antipsychotics (e.g., quetiapine) TCA	
Ictal aggression (abnormal EEG)	Carbamazepine Valproates Diphenylhydantoin Clonazepam		TCA and Antipsychotics: Generally, not recommended as they increase the risk of seizures.
III. Social and emotional detachment			
Chronic asocial disinterest and/or blunted affect	Second Generation Antipsychotics (e.g., quetiapine, clozapine) Third Generation Antipsychotics (aripiprazole)		
IV. Cognitive-perceptual distortions and psychotic symptoms			
Acute and brief psychotic episodes	Second Generation Antipsychotics (e.g., risperidone, quetiapine) Note: Only give for the duration of the symptoms		
Chronic low-level psychosis	Second Generation Antipsychotic (e.g., quetiapine) Third Generation Antipsychotics (aripiprazole)		

---

[a] Psychotherapy is the usual treatment of choice.

(i) Symptom domain of mood dysregulation and anxiety

This symptom domain includes mood instability, typical and atypical depression and/or dysphoria, and chronic cognitive and somatic anxiety.

Emotional instability and mood swings may respond to lithium or valproates (for patients with frequent episodes of euphoria) or lamotrigine (for patients with more frequent depressive episodes). Low-dose atypical psychotropics may be used to stabilize mood or improve depression as the second-line medications. As a general rule, antidepressants may risk an episode of mania in persons with emotional lability or cyclothymia. The risk may be reduced by the combined use of a mood stabilizer. TCAs are extremely dangerous in an overdose, so these drugs ought to be used with caution in patients with PD.

Atypical depression and dysphoria are very frequently observed in patients with any subtype of PD. Serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine reuptake inhibitors (SNRIs) are commonly used as first-line treatments. Monoamine oxidase inhibitors (MAOIs) or low-dose atypical psychotropics, such as aripiprazole, ziprasidone, lurasidone, or clozapine, may also be used after careful consideration of the risk–benefit ratio. TCAs are not recommended as at least half of the patients with PD who suffer from atypical depression worsen on TCAs.

Typical depression, including Major Depression, may complicate any PD. These syndromes are treated with antidepressants, including heterocyclics, in doses suggested for primary major depression.

Patients with PD often present with both cognitive anxiety (anticipatory worry) and somatic anxiety (concerns about bodily pains and psychophysiological reactions). While the treatment of choice for chronic cognitive anxiety usually is psychotherapy, especially various forms of CBT, chronic cognitive anxiety may respond to SSRIs, MAOIs, SNRIs, GABA analogs (such as valproates, pregabalin, gabapentin), and (with caution) long half-life benzodiazepines. Chronic somatic anxiety is more responsive to MAOIs, SNRIs (e.g., venlafaxine or duloxetine), and buspirone. Avoidant traits can also be effectively treated with either SSRIs or MAOIs. Some components of somatic anxiety, such as sweating, palpitations, diarrhea, and tremor, can be treated with beta-blockers. Severe, psychotic-like anxiety may respond to low-dose neuroleptics, especially drugs with low D2 affinity (e.g., quetiapine). Despite the relative safety of novel atypical drugs, caution in prolonged use is essential because there is a substantial risk of serious side effects. For example, the use of risperidone carries a risk per person per year of 0.6% for tardive dyskinesia, 0.35% for sudden cardiac death, and 0.02% to 2.44% for neuroleptic malignant syndrome. Therefore, the risk of serious and irreversible side effects is about 1% per person per year and requires careful consideration of the risk–benefit ratio with informed consent.

(ii) Aggression

It is useful, though sometimes difficult, to distinguish different types of aggression. However, diagnosis of different types of aggressive behaviors has important diagnostic implications as well as crucial guidelines for pharmacotherapy. The most common form of aggression occurs when a quick-tempered person is provoked by frustration or threats. Here, the very act of aggression does not have a discernible long-term goal or secondary gain. This is often called "affective aggression" and is frequent in Cluster B impulsive–aggressive individuals who are high in novelty seeking and low in harm avoidance. Aggression that appears to be unprovoked sometimes occurs in patients with cerebral instability documented by an abnormal electroencephalogram (EEG) and is often called "ictal aggression" regardless of any associated personality traits. Predatory aggression or "cruelty" is also known as "organized aggression" because it involves planning, social motives, and/or secondary gains. This type of aggression involves hostile revengefulness and taking pleasure in victimizing others, usually with intact impulse control. Such predatory aggression is most frequent in individuals who are very low in cooperativeness and who are also emotionally dissociated and disengaged, such as Antisocial and Schizoid PD. Finally, "organic-like" aggression is often accompanied by poor social judgment and disinhibition; it is best distinguished from other impulsive–aggressive syndromes by prominent distractibility, inattention, and emotional lability, all characteristic of patients with frontal lobe lesions. These patients also manifest high somatic anxiety with panic and cardiorespiratory symptoms, muscular tension, and motor restlessness.

Multiple double-blind trials have shown the efficacy of lithium carbonate in the treatment of affective aggression. Lithium salts help impulsive–aggressive individuals to be more reflective, that is, to think about consequences before acting on impulse. To a lesser extent, it may be helpful in reducing cruelty and lack of cooperativeness, but this may be an indirect result of reducing impulsivity, which often is a predisposing influence in the development of hostility and revengefulness. Likewise, low-dose atypical neuroleptics may be useful in setting the stage for the patient to modify old habits and assist them in reducing affective aggression. The decision to use long-term neuroleptics requires consideration of potential side effects, such as tardive dyskinesia, and should be made with the carefully informed consent of the patient. Anticonvulsants, such as valproates, lamotrigine, carbamazepine, and oxcarbazepine (to mention only those most frequently used), reduce both the intensity and the frequency of unprovoked angry outbursts in many patients regardless of the normality of their EEG. Double-blind trials have shown that psychostimulants and catecholamine agonists, such as methylphenidate, are often beneficial in the treatment of inattentive and hyperactive adults who are impulsive and aggressive, especially when the symptoms have begun in early childhood.

In accord with postulated serotonergic mechanisms in aggressive behaviors, antidepressants (particularly SSRIs and SNRIs) are considered by many to be beneficial for certain subtypes of impulsive PD (e.g., borderline, histrionic). Finally, MAOIs are effective in some dysphoric states with somatic anxiety and hostility.

There is no effective pharmacotherapy for predatory, premeditated, and cold-blooded aggression. One treatment with some chances of success is psychotherapy, usually CBT. Low-dose traditional neuroleptics (haloperidol) or low-dose novel psychotropics (Risperdal, quetiapine) may help tone down predatory aggression. Of note, antipsychotic-level doses have not been shown more effective here and are not recommended. When accompanied by impulsivity, predatory aggression may improve if treated with beta-blockers.

There are some relative contraindications for the aforementioned listed drugs. Lithium should not be given to antisocial persons without aggression and impulsivity because it does not diminish nonaggressive antisocial behaviors (such as lying, cheating, and stealing) and is poorly tolerated by anxious schizoid individuals. Thus, lithium exposes such patients to the risk of neurologic, renal, and thyroid toxicity without providing much benefit. Likewise, benzodiazepines and alcohol have disinhibiting effects on violence, reduce conditioned avoidance behavior (“loosen inhibitions”), and further impair passive avoidance learning in impulsive antisocial persons. The use of benzodiazepines seems appropriate only in nonaggressive asocial behaviors, for example, patients with schizoid PD. As a general rule, typical and novel psychotropics are used with extreme caution, and the patient is closely monitored so that early signs of complications and adverse effects are recognized and prevented, usually by terminating the offending agent.

(iii) Symptom domain of social and emotional detachment

Emotional detachment, cold and aloof emotions, and disinterest in social relations (“chronic asociality”) are most frequently observed with schizoid and paranoid PD and, to a lesser extent, with antisocial PD. In some cases, emotional detachment or disinterest may respond to novel psychotropics, such as aripiprazole, risperidone, quetiapine, clozapine, olanzapine, or ziprasidone. These medications may help reduce social withdrawal and other features of “aloof” PDs with less risk of extrapyramidal symptoms than with typical neuroleptics. However, dose adjustment is crucial to maintain compliance because of frequent side effects. In cases in which emotional disinterest reflects an underlying depression, antidepressants (SSRIs or MAOIs) frequently help. One should be cautious with TCAs in schizotypal PD because they may worsen and/or trigger psychosis.

(iv) Cognitive-perceptual distortions; psychotic symptoms

Cognitive-perceptual distortions observed in a patient with PD primarily include mild cognitive symptoms, such as bizarre fantasies, eccentric attitudes, suspiciousness, egocentric perception of reality ("everything revolves around me"), nonpsychotic thought disorder (such as ideas of reference and magical thinking), unusual perceptual experiences (such as illusions), and eccentric and bizarre behaviors, among many others. Dissociative phenomena are also frequently observed, such as déjà vu, depersonalization, and derealization. These and similar symptoms can be occasionally observed with all subtypes of PDs but are more frequent in Cluster A and to a lesser extent in Cluster B PDs. These chronic, low-level, psychosis-like but essentially nonpsychotic symptoms may improve indirectly, after pharmacologic stabilization of disturbing affects, and/or after psychotherapeutic correction of underlying cognitive beliefs, and/or after reduction of realistic external stress. For example, a narcissistic person can be "paranoid" that other people have conspired to prevent their success. Psychotherapy is the treatment of choice for this kind of nonpsychotic paranoia. Reflecting the underlying operation of splitting (a primitive defense mechanism that prevents a realistic perception of self and others), the same narcissistic person may at other times perceive these same "enemies" in a diametrically opposite way, as friends or supporters, and even idealize them. In cases in which these symptoms persist and interfere with everyday functioning despite psychotherapy, medications are occasionally needed, usually novel atypical psychotropics (such as aripiprazole, quetiapine).

Table 26–25.
Awakening a Positive Perspective: Four Major Principles of the Science of Well-Being

Positive Approaches to Well-Being	Experiences and Activities	
	Recommended	**Not Recommended**
Letting go of struggles	Acts of hope and self-direction; accepting responsibility Silence of mind, phase 1[a] (calm reflection)	Any violence, fighting, complaining or blaming others consumerism feeding greed/addiction
Working in the service of others	Acts of kindness/cooperation; purposeful giving of oneself Union in nature[a] Silence of mind, phase 2[a] (mindful meditation)	Divisions feeding fear and hate; possessiveness and hoarding; seeking power and dominance
Growing in awareness	Acts of faith and humility; transcendent problem solving Union in nature[a] Listening to the heart[a] Silence of mind, phase 3[a] (contemplation)	Criticism or praise of self and others, depending on external direction, denial of what one does not understand
Understanding thought processes	Calm reflection, mindful meditation, contemplation (silence of mind[a], all three phases)	Seeking justification, depending on external advice, depending on intellect alone

[a] Further literature and programs for the cultivation of well-being are available through a nonprofit educational foundation (http://anthropedia.org).

Used with permission of the Anthropedia Foundation.

Some chronic cognitive disturbances, such as mild ideas of reference or suspiciousness, tend to subside when the background emotional tension is reduced. For example, alprazolam has been found to be beneficial in patients with borderline personality, particularly those with a history of drug abuse and suspiciousness. However, long-term use of benzodiazepines is associated with a high risk of drug dependence, particularly in patients with PDs, so benzodiazepines like alprazolam should only be prescribed after careful consideration of the risk–benefit ratio and their use carefully monitored for evidence of abuse or dependence.

Chronic cognitive and behavior oddities, such as eccentric attitudes, bizarre behavior, odd beliefs, or extrasensory perceptions, are seen with Schizotypal PD. There is substantial empirical evidence that Schizotypal symptoms ought to be classified among other Schizophrenia Spectrum Disorders. Acute, brief reactive psychoses may complicate most subtypes of PD. These are treated symptomatically, according to accepted pharmacologic practices, most frequently with second-generation atypical psychotropics due to much better safety and tolerability.

Again, acute psychotic symptoms requiring medication may subside when environmental stressors are brought under control; thus, one should be ready to lower the dose or discontinue the medication.

An Integrated Approach to the Treatment of Personality Disorder

Neither medications nor cognitive-behavioral approaches, alone or in combination, are usually adequate to transform a person's personality in a fundamental way. A systematic and integrated approach to personality transformation without tension or conflict has been described by Cloninger as coherence therapy in his book on the Science of Well-Being. The second stage of treating PD involves the awakening of the positive outlooks on life that are needed for well-being, as described in Table 26–25. The basic principles of well-being are summarized in Table 26–25, along with therapeutic experiences and activities for patients that are designed to help them value the dignity of their life and that of others as human beings as a result of self-awareness of each person's body, mind, and psyche. Psychiatry literally means the "healing of the psyche," and the psyche refers to a creative, self-aware consciousness that is more than analytical reasoning. Recovery from mental illness most consistently involves a transformation of a person's spirit or basic perspective about life. However, psychiatrists often become narrowly focused on the reduction of symptoms or risk of harm and forget the importance of helping the patient to maintain self-respect and hope. So often, the insight that psychiatry is the "healing of the psyche" has been neglected or denied as a result of the errors of materialistic and reductionistic thinking. What is meant here by an awakening and healing of the psyche is that the patient becomes directly aware that their worldview (i.e., their outlook on life) has an impact on their thoughts, emotions, and actions. The human outlook on life is what makes people vulnerable to mental disorders, so people must become aware of the assumptions implicit in their initial perspectives. Without some degree of awareness of the consequences of one's initial outlook, it is not possible to transcend the conflicts and contradictions in the thoughts and emotions that derive from these more or less coherent outlooks. As a result, psychotherapies that neglect the stage of spiritual awakening inevitably leave their patients locked in an inner struggle among parts of themselves from which there is no escape, as described poignantly by Freud about himself in ***Civilization and Its Discontents.*** Further information about coaching or therapy to cultivate well-being is available for interested clinicians (http://anthropedia.org).

Once there is an awakening of self-awareness to a metacognitive level, then a patient can proceed to the advanced stages of treatment that are briefly summarized in Table 26–26. Notice now that the patient is behaviorally stable and self-aware, then it is possible to focus on the ***causes*** of the symptoms that were targets of initial intervention during the stabilization phase. For example, the causes of anxiety and dysregulated mood are rooted in a patient's lack of self-respect and trust in others, as expressed in their feelings of separateness, catastrophe, and victimization (see Table 26–26). The therapeutic strategies for addressing each of these sets of causes include psychoeducation, physical and other nonverbal therapies (e.g., psychomotor activities, acupressure), emotional skills training, cognitive skills training, and spiritual exercises. As a result, no one form of therapy (such as behavioral, cognitive, interpersonal, or psychodynamic) is really comprehensive, even with the addition of modules from positive psychology or mindfulness training.

Table 26–26.

Physical, Personal, Social, Cognitive, and Spiritual Therapeutic Procedures for Advanced Character Development (Stages 3 and 4 of Coherence Therapy for Personality Disorders)

Target Problems	Character Goals	Recommended Experiential Activities
1. Mood dysregulation (feeling separate, intolerant/hateful, self-critical, catastrophizing/distressed, victimized)	Trust and self-respect	Psychoeducation about health, sex, stress Physical exercise for relaxation and fitness Reconciliation of anxiety versus risk-taking Recognition of immature automatic thoughts Teamwork to build trust and respect
2. Aggression/impulsivity (being violent, angry, frustrated, greedy/jealous, proud, selfish)	Impulse control and self-mastery	Psychoeducation about nutrition, biorhythms Self-efficacy about cravings/substance abuse Reconciliation of impulsivity versus rigidity Nonviolent assertive communication Social service and acts of forgiveness
3. Social dysregulation (feeling rejected, detached, insecure or cold, unappreciated, unfriendly)	Empathy and secure attachments	Psychoeducation about social signals Empathy training and active listening skills Reconciliation of approval-seeking versus aloofness

		Social problem solving Therapeutic touch and appeasement
4. Cognitive distortion (feeling meaningless, judgmental, dualistic, afraid of death, lacking faith)	Mindfulness and meaning	Psychoeducation about self-transcendence Practice of mindfulness meditation Reconciliation of persistence by virtue Reflection on mysteries of God, life/death Creative works (art, music, writing, etc.)
5. Emptiness (seeking fulfillment, satisfaction, positive emotions, well-being coaching[a], virtue, well-being)	Virtue and well-being	Psychoeducation about well-being Recognizing triggers of negative emotion

[a] Literature and training available through http://anthropedia.org.

Used with permission of the Anthropedia Foundation.

Tables 26–24 and 26–26 highlight that the symptomatic targets for psychobiological treatment correspond closely to the skill training modules that have been shown to be moderately effective in the treatment of severe PDs by Marsha Linehan and others, except that the focus is shifted from a behavioral approach to the cognitive perspectives or schemas that lead to the behavioral problems. In addition, a nondualistic approach is taken along a stepwise path of character development to well-being, and emphasis is placed on reconciliation of emotional conflicts that can be measured by the extremes of each of temperament dimensions. In essence, the extremes of each temperament are transcended by the development of particular forms of spiritually elevated thoughts—namely, self-respect reconciles the extremes of Harm Avoidance, self-mastery reconciles the extremes of Novelty Seeking, secure attachments reconcile the extremes of Reward Dependence, and the cultivation of virtues and transcendent meaning reconciles the extremes of Persistence and the limits of the finite human intellect. Both extremes of each temperament have advantages and disadvantages, and transcendence of the underlying conflict resulting from these disadvantages allows a person to live without tension or conflict about these issues as a result of a more holistic initial perspective.

A major practical advantage of therapy modules that focus on the reconciliation of both extremes of each temperament is that heterogeneous groups of patients with widely different personality profiles can be treated together. The focus is on transcending emotional conflicts, coherent character development, and well-being for everyone, not particular personality subtypes. The stigma linked to PD and mental disorders, in general, is mitigated by facing the fact that everyone is imperfect but can learn to live without fear along the path to well-being. The design of therapies has often failed to recognize the path of development of character and well-being and the crucial role of spirituality in transcending and sublimating emotional conflicts. Psychotherapies have almost entirely focused on thought and have generally ignored the body as well as the psyche as a result of the influence of errors in dualistic, behavioral, materialistic, or psychoanalytical thinking. Such an intellectual stance is arrogant and ultimately self-defeating. Alternative therapies have focused on methods that emphasize the body and ignore the benefits of medications or psychodynamics, which is equally arrogant and self-defeating. The coherence therapy described here is based on an integrative psychobiological approach that attends in a balanced way to all three aspects of one’s being—body, mind, and psyche. There is no incompatibility between biologic, psychological, or spiritual aspects of the treatment of PD. In fact, there is a crucial synergy between the components of treatment directed at each aspect because rational brains do not function well when the emotional brain (i.e., limbic system) is distressed. There is much to be learned about the treatment of PD and related mental disorders, but the approach outlined here provides a paradigm that will allow clinicians to integrate what they know about people and bring that to help and guide their treatment of every patient they encounter. It should be emphasized again that advanced psychotherapy or the use of psychotropic medications for the treatment of PD requires expert training or close supervision by an experienced clinician. This is a rapidly advancing area of psychiatry and psychology. Continuing education about the treatment of PD is likely to be especially rewarding for both psychiatrists and their patients because it offers a crucial gateway into the path to well-being and reduced disability from the full range of mental disorders.

Further Readings

Allport G. ***Personality: A Psychological Interpretation***. 1937.

American Psychiatric Association. ***Diagnostic and Statistical Mental of Mental Disorders*** (IIIrd ed, III-R, IV, IV-TR, 5). APA Press; 1980, 1987, 1994, 2000, 2013.

Bach B, Christensen S, Kongerslev MT, Sellbom M, Simonsen E. Structure of clinician-reported ICD-11 personality trait qualifiers. ***Psychol Assess***. 2020;32(1):50–59.

*Cloninger CR. A systematic method for clinical description and classification of personality variants. A proposal. ***Arch Gen Psychiatry***. 1987;44:573–588.

*Cloninger CR. ***Feeling Good: The Science of Well-Being***. Oxford University Press; 2004.

Cloninger CR. Evolution of human brain functions: the functional structure of human consciousness. ***Aust NZ J Psychiatry***. 2009;43(11):994–1006.

*Cloninger CR, Cloninger KM. Person-centered therapeutics. ***Int J Pers Cent Med***. 2011;1(1):43–52.

*Cloninger CR, Cloninger KM, Zwir I, Keltigangas-Jarvinen L. The complex genetics and biology of human temperament: A review of traditional concepts in relation to new molecular findings. ***Transl Psychiatry***. 2019;9(1):290.

*Cloninger CR, Svrakic DM, Przybeck TR. A psychobiological model of temperament and character. ***Arch Gen Psychiatry***. 1993;50:975–990.

Cloninger CR, Zohar AH. Personality and the perception of health and happiness. ***J Affect Disord***. 2011;129(1–2):24–32.

Cloninger CR, Zohar AH, Hirschmann S, Dahan D. The psychological costs and benefits of being highly persistent: personality profiles distinguish mood disorders from anxiety disorders. ***J Affect Disord***. 2012;136(3):758–766.

Hansenne M, Pinto E, Scantamburlo G, et al. Harm avoidance is related to mismatch negativity (MMN) amplitude in healthy subjects. ***Pers Individ Dif***. 2003;34:1039–1048.

Josefsson K, Jokela M, Cloninger CR, et al. Maturity and change in personality: developmental trends in temperament and character in adulthood. ***Dev Psychopathol***. 2013;25:713–727.

Kernberg O. ***Borderline Conditions and Pathological Narcissism***. Aronson; 1975.

Markett S, Weber B, Voigt G, et al. Intrinsic connectivity networks and personality: the temperament dimension harm avoidance moderates functional connectivity in the resting brain. ***Neuroscience***. 2013;240:98–105.

*Mezzich JE, Botbol M, Christodoulou GN, Cloninger CR, Salloum IS, eds. ***Person-centered Psychiatry***. Springer International Publications; 2016.

Mulder R, Joyce P. Temperament and structure of PD symptoms. ***Psychol Med***. 1997;27(1):99–106.

Perry C, Vaillant G. Personality disorders. In: Kaplan HI, Sadock BJ, eds. ***Comprehensive Textbook of Psychiatry***. 5th ed. Williams and Wilkins; 1989:1352–1387.

Quilty LC, Godfrey KM, Kennedy SH, Bagby RM. Harm avoidance as a mediator of treatment response to antidepressant treatment of patients with major depression. ***Psychother Psychosom***. 2010;79:116–122.

Reich J, Frances A. The structural interview method for diagnosing borderline disorders. ***Psychiatr Q***. 1984;56:229–235.

Rochat P. ***Others in Mind: The Origins of Self-consciousness***. CambridgeUniversity Press; 2009.

Schneider K. ***Psychopathic Personalities***. Cassell; 1956.

Spreng RN, Mar RA, Kim ASN. The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. ***J Cogn Neurosci***. 2008;21(3):489–510.

Svrakic NM, Svrakic DM, Cloninger CR. A general quantitative theory of personality development: fundamentals of a self-organizing psychobiological complex, development and psychopathology. ***Dev Psychopath***.1996;8:247–272.

Svrakic DM, Whitehead C, Przybeck TR, Cloninger CR. Dimensional diagnosis of personality disorders by the seven factor model of temperament and character. ***Arch Gen Psychiatry***. 1993;50:991–999.

Tulving E. Episodic memory: from mind to brain. ***Ann Rev Psychol***. 2002;53:1–25.

*Tyrer P, Reed GM, Crawford M. Classification, assessment, prevalence, and effect of personality disorder. ***Lancet***. 2015;385(9969):717–726.

*Zachar P, St. Stoyanova D, Aragona M, Jablensky A. ***Alternative perspectives on psychiatric validation: DSM, ICD, RDoC and Beyond***. Oxford University Press; 2015.

Zwir I, Arnedo J, Del-Val C, et al. Uncovering the complex genetics of human character. ***Mol Psychiatry***. 2020;25(10):2295–2312.

Zwir I, Arnedo J, Del-Val C, et al. Uncovering the complex genetics of human temperament. ***Mol Psychiatry***. 2020;25(10):2275–2294.

*Zwir I, Arnedo J, Del-Val C, et al. Three gene-environment networks for human personality. ***Mol Psychiatry***. 2021;26(8):3858-3875.

*Zwir I, Del-Val C, Hintsanen M, et al. Evolution of genetic networks for human creativity. ***Mol Psychiatry***. 2022;27:354-376.

27

Consultation-Liaison Psychiatry and Psychosomatic Medicine

27.1 History and Current Trends

Carol L. Alter, M.D., and Steven A. Epstein, M.D.

Psychosomatic medicine has been a specific area of focus within the field of psychiatry for more than 50 years. The term *psychosomatic* is derived from the Greek words *psyche* (soul) and *soma* (body). The term literally refers to how the mind affects the body. Unfortunately, it has come to be used, at least by the lay public, to describe an individual with medical complaints that have no physical cause and are “all in your head.” In part due to this misconceptualization, the American Psychiatric Association’s (APA’s) *Diagnostic and Statistical Manual of Mental Disorders* (DSM), in 1980, deleted the nosologic term psychophysiological (or psychosomatic) disorders and replaced it with “Psychological Factors Affecting Other Medical Conditions” in DSM-5. Nonetheless, the term continues to be used by researchers and is in the title of major journals in the field (e.g., *Psychosomatic Medicine* and *Journal of Psychosomatic Research*). It is also used by one of the two major national organizations in the field (the American Psychosomatic Society) as well as international organizations (e.g., the European Association of Psychosomatic Medicine). In an effort to better align the name of the field with the practice of the subspecialty, in 2017, the name “consultation-liaison (C-L) psychiatry” was adopted in coordinated efforts resulting in a change in the name of the subspecialty by the American Board of Medical Specialties (ABMS) and the American Board of Psychiatry and Neurology. Similarly, the former Academy of Psychosomatic Medicine became the Academy of Consultation-Liaison Psychiatry, and its journal was renamed the Journal of the Academy of Consultation-Liaison Psychiatry.

ANCIENT HISTORY

According to Hippocrates (460 to 370 BCE), who was influenced by Egyptian and Greek traditions, disease originated within the body and was due to an imbalance in fluid matter. That imbalance could be related to or even caused by a similar imbalance in the patient's external environment. Hippocrates warned the doctor to consider the altitude, the wind direction, the purity of water supply, and the season of the year before making any diagnosis. He pointed out that citizens of cities in particular environments would run to phlegmatic or sanguine temperaments as a reflection of the city's coldness or heat. Physical or fluid imbalance could be caused by emotional upset, too. Hippocrates reported that fear produced sweat and that shame brought on palpitations of the heart; he urged young physicians to look at patients with kindly expressions and never with impatience because impatience could inhibit the return of health. He is said to have cured a king of an intestinal lesion by analyzing a dream. It was Hippocrates who wrote, "in order to cure the human body, it is necessary to have a knowledge of the whole of things."

In *Timaeus*, Plato remarked that trouble in the soul could bring trouble to the body, and, in *Charmides*, Plato quoted Socrates, who had himself attributed the words to a Thracian king: "As it is not proper to cure the eyes without the head, nor the head without the body, so neither is it proper to cure the body without the soul." Aristotle (384 to 322 BC) observed that the emotions of anger, fear, courage, and joy affect the body, and Aretaeus (first century AD) wrote that "the vehement affections of the soul" could cause paralysis.

MODERN HISTORY

As Edward Shorter discusses in detail in his history of psychosomatic illness, ways of presenting illness have varied over history since patients unconsciously manifest symptoms that are thought to represent somatic illnesses. Prior to 1800, physicians did not conduct clinical evaluations and could not distinguish somatic from psychogenic illness. As a result, the diagnoses of hysteria and hypochondriasis could easily be made erroneously in the presence of medical illnesses and did not suggest any specific disease presentations. According to Shorter, “cultural shaping” of patients’ symptoms began in the middle of the 19th century with the popular diagnosis of spinal irritation. That diagnosis was made when a physician believed that spinal cord irritation caused symptoms, and the patient complained of spine sensitivity as well as peripheral symptoms. As was the case for previous explanations of illness, spinal irritation was either an invalid diagnosis in a patient with another medical illness or a “code word” for psychosomatic symptoms. The term “psychosomatic” was first used by Johann Christian Heinroth in 1818.

In the mid-19th century, after much popularity in Europe and the United States, the diagnosis of spinal irritation went out of favor. It was replaced by “reflex neurosis” as an explanation for somatization. That diagnosis was based on the view that any affected organ could cause irritation in any other organ in the body. The conceptualization was based on reflex theory, the notion that nervous connections throughout the body regulate all bodily organs, independent of free will. The Berlin neurologist Moritz Romberg was an influential proponent of the view that the uterus was the cause of hysteria, a reflex neurosis. In this view, hysteria, which encompassed globus hystericus, hysterical seizures, and other psychogenic disorders, was due to irritation of the uterus and the ovaries. He believed that male genitals could also cause hysteria but that they do so less commonly because they are less frequently irritated. In the late 19th century, reflex theory was one of the leading models of nervous illness. Since physicians promulgated the theory to their patients, it is not surprising that they found many of their patients to be preoccupied with their genital organs. Gynecologist August Rheinstadter of Cologne commented skeptically in 1884:

> But now, a true hysteromania, a Furor uterinus, has unfortunately arisen from the previous reluctance of women to undergo a gynecological examination and from the indifference which doctors once displayed towards gynecological disease. So that every woman who suffers from migraine, stomach cramps, or palpitations believes herself to have a uterine illness and indeed will find some physician willing to indulge her in the treatment of this presumed cause.

In the late 19th century, many physicians in Europe and the United States believed that gynecologic problems produced hysteria and other forms of mental illness. As a result, gynecologists performed surgery to treat hysteria and insanity, including the removal of healthy ovaries. The widespread availability of gynecologic surgery in the United States also led women to request such surgery to relieve psychosomatic symptoms. In his history of psychosomatic illness, Shorter notes that some Americans' "addiction to surgery" was due to practitioners convincing the public that vague physical symptoms were due to reflex phenomena from peripheral organs. The nose was the final organ addressed by reflex theory in the late 19th century. One of its chief proponents was Wilhelm Fliess, a Berlin general practitioner, who greatly influenced Freud.

In the 19th century, hysterical fits were the most common form of motor hysteria. There were two primary forms: catalepsy and uncontrolled motor activity such as thrashing about. The psychiatrist Pierre Briquet described over 400 women with hysterical conversion disorders whom he had evaluated at the Pitié-Salpêtrière Hospital in Paris from 1849 to 1859. Many had multiple unexplained somatic complaints; the diagnosis of Briquet syndrome ultimately developed into the diagnosis of somatization disorder which is most similar to the current DSM-5 diagnosis of somatic symptom disorder.

Hysterical paralysis, often due to stressful life events, was apparently quite common through the 19th and early 20th centuries. Louis Verhaeghe, a spa physician in Belgium, wrote in 1850:

> How often do we see patients with paralyzed arms, legs, the side of the face, or even a more limited area such as one or two fingers, problems attributable to an unequal distribution of nervous fluid. It is individuals of a pronounced nervous temperament, subject to hysterical fits, somnambules, persons subject to shameful habits or who abuse the pleasures of love, who are most likely to experience these paralyses.

Astasia-abasia was also seen frequently; Silas Wier Mitchell in Philadelphia referred to it as "hysterical motor ataxia." One of the most prominent physicians of the late 19th century, Jean-Martin Charcot, promulgated a highly influential theory of hysteria. He believed that it was a chronic, inherited functional disease of the nervous system. "In the [hysterical] fit," he said in 1882, "nothing is left to chance, that to the contrary everything unfolds according to the rules, which are always the same and characterize what we see in our outpatients as well as inpatients; they are valid for all countries, for all epochs, for all races, and are, in short, universal.” Although his conception of hysteria was widely disseminated, some physicians who were aware of his ability to suggest symptoms to vulnerable patients were dubious of his theory. The Swiss neurologist Paul Dubois wrote in 1904: “Endowed with the spirit of authority, [Charcot] . . . suggested to them their attitudes and their gestures. Example is contagious even in sickness, and in the great hospitals of Paris, at La Salpêtrière, all cases of hysteria resemble each other. At the command of the chief of the staff, or of the interns, they begin to act like marionettes, or like circus horses accustomed to repeat the same evolutions.” It was not until near the end of his life that Charcot’s theories became more psychologically focused.

Within a decade after he died in 1893, Charcot hysteria had virtually disappeared. Many physicians began to believe that some of his patients most likely had an organic illness. Joseph Babinski’s discovery in 1896 of the Babinski sign was extremely helpful in distinguishing organic disease from hysteria. In 1901, he put forward a new conceptualization of hysteria: It was defined as any symptom that could be induced by suggestion and eliminated by persuasion.

As the reflex model faded in prominence in the late 19th century, the most dominant model became one in which central nervous disease was held responsible for nervous symptoms. In the early 20th century, psychological theories of somatization became popular, but they became dominant only in the latter half of the 20th century. “Neurasthenia,” or tired nerves, was a new diagnosis at the end of the 19th century that was brought to prominence by George Beard, a New York neurologist. A term applied broadly, the diagnosis encompassed depression, anxiety, somatization, male hysteria, and chronic fatigue. Wilhelm Stekel, a Viennese psychoanalyst, coined the term “somatization” in the 1920s.

Sigmund Freud was the principal theoretician to bring psyche and soma back together. He demonstrated the importance of emotions in producing mental disturbances and somatic disorders. His early psychoanalytic formulations detailed the role of psychic determinism in somatic conversion reactions. Picking up where the Greeks and Romans had left off, Freud redignified the study of the emotions as a separate study and pointed to their relationship with the soma in the new field of psychiatry. Using Freud’s insight, a number of workers in the early decades of the 20th century tried to expand the understanding of the interrelationship of psyche and soma. The influence on adult organ tissue of various unresolved pregenital impulses was proposed by Karl Abraham in 1927, the application of the idea of conversion reaction to organs under the control of the autonomic nervous systems was described by Sandor Ferenczi in 1926, and the attaching of symbolic meaning to fever and hemorrhage was suggested by George Groddeck in 1929.

In the 20th century, somatization symptoms changed from predominantly neurologic (e.g., hysterical paralysis) to other symptoms such as fatigue and chronic pain. Edward Shorter attributes this change to three causes: (1) improvements in medical diagnostic techniques made it easier to rule out organic causes for neurologic disease; (2) the central nervous system (CNS) paradigm faded; and (3) social roles changed (e.g., the disappearance of the historical notion that “weak” women would be expected to have fainting spells and paralysis).

Although hysterical neurologic symptoms have remained relatively less common in the 21st century, CNS explanations of chronic pain and fatigue are gaining prominence. For example, functional brain research has demonstrated brain dysfunction and possibly genetic contributions among some individuals with fibromyalgia and chronic fatigue syndrome. Those syndromes, while still thought by some to represent somatization variants, are currently established medical diagnoses.

HISTORY OF PSYCHOSOMATIC MEDICINE RESEARCH

Since the mid-20th century, research in psychosomatic medicine and C-L psychiatry has taken two interconnected paths. Psychosomatic medicine research has generally focused on understanding psychophysiological mechanisms underlying mind–body relationships. C-L research, however, has largely been directed at understanding psychiatric problems among clinical populations with medical illnesses.

In the early 20th century, two theoretical trends developed, one suggesting that specific emotions led to specific cell and tissue damage, and the second that generalized anxiety created the preconditions for a number of not-necessarily predetermined diseases. Cases of shell shock during World War I and new endocrine studies provided observations for the swell of interest in theory evidenced in the 1930s.

In the early 1930s, Franz Alexander, who had left Berlin to move to Chicago, where he founded the Chicago Institute for Psychoanalysis, described how he believed specific psychological conflicts caused specific somatoform symptoms. Alexander saw conflict as a stress and suggested that, when conflict presents itself, the individual may suppress the stress and produce, through the voluntary nervous system, a reaction such as the conversion reaction described by Freud. Alternatively, after suppressing stress, the individual's sympathetic responses may remain alert for heightened aggression or flight, or parasympathetic responses may remain alert for heightened vegetative activity. According to these theories, prolonged alertness and tension can produce physiologic disorders and eventual pathology of visceral organs.

For instance, Alexander postulated that in a passive-dependent person without someone to satisfy their dependence, stress is created. That particular stress may stimulate and keep alert the parasympathetic nervous system, which means that too much gastric acid is secreted, and gastric hypermotility results, all of which may lead to a peptic ulcer. Another dependent person with a different genetic profile may, in repressing conflict, stimulate parasympathetic overfunctioning through pathways leading to colitis or asthma. Still, other dependent persons' efforts to move beyond dependency result in overstimulation of the sympathetic system, and the resulting chronic alertness produces migraine, hypertension, or arthritis.

Although there were multiple case reports supporting Alexander's views, it has not been demonstrated that the same specific conflicts are present in all cases of the same disease. Early methodologically flawed reports were rooted in psychoanalytic explanations of "psychosomatic" illnesses. This work had numerous problems, including selection bias, small sample size, and lack of control groups.

A number of investigators developed the pathway concept into other theories involving the whole personality. In 1954, Flanders Dunbar proposed that the ambitious, hard-driving person would be prone to coronary occlusion. That personality type is similar to the competitive, restless, time-haunted, and coronary-bound type A person proposed by M. Friedman and R. H. Rosenman in 1959.

A second trend in psychosomatic medicine that began in the middle of the 20th century has been to investigate what happens to a person in a nonspecific way when faced with stress. In the 1950s, Harold Wolff and Stewart Wolf observed that chronic hyperfunction or chronic hypofunction in the vascular and secretory activities of the mucosa of the gastrointestinal (GI) and respiratory systems could produce pathology. Overfunctioning of the mucosa was correlated with hostility, and underfunctioning was correlated with fear or sadness. The patient's entire reactive patterning and their life history account for whether they react to stress by hyperfunctioning or hypofunctioning.

Other workers in the nonspecific group described various possible mechanisms by which psychologically induced stress may cause organic disease in humans and animals. In 1950, Hans Selye thought that the hypophyseal–adrenocortical axis responded to various types of physical and psychic stress with hormonal changes that can ultimately cause a variety of organic diseases, such as rheumatoid arthritis and peptic ulcer disease. Selye viewed such diseases as a byproduct of the body's attempt to adapt to stress from any source.

Experimental psychologists with a learning theory conception of behavior studied the effects of chronic unrelieved anxiety in humans and animals and found that gastric hydrochloric acid production increases under such circumstances. Because such acidity is a precursor of peptic ulcer, they concluded that chronic anxiety is the variable intervening between the behavioral and physical events involved in psychosomatic illness. Modern research provides some support for the role of stress as a predisposing factor for the development of peptic ulcer disease (e.g., the work of Levenstein).

Beginning in the late 1960s, T. H. Holmes and R. H. Rahe showed that life crises often precede illness and that there is some correlation between the intensity of the crisis and the severity of the illness that follows; the crisis ranked as the most intense by subjects was the death of a spouse. George Engel suggested a giving-up–given-up complex, in which individuals feel powerless to change their environments or themselves and eventually lose the will to try, becoming mentally or physically ill, perhaps because of changes in the immunologic and neuroendocrine systems.

As exemplified by the long history of psychosomatic research in pain and GI disease, the quality of research in the field has evolved significantly over the past 70 years. Early research efforts were psychoanalytically oriented case studies of conditions such as ulcerative colitis. Such research often led to invalid conclusions. For example, one report by A. Johnson and colleagues in 1947 of a series of 33 patients with rheumatoid arthritis led to the preliminary conclusion that women with this disorder have “a chronic inhibited hostile aggressive state as a reaction to the earliest masochistic dependence on the mother that is carried over to the father and all human relationships, including the sexual.” Nonetheless, even those authors were aware of potential methodologic limitations of their work: “but these factors . . . are found so commonly in patients who do not suffer from arthritis that additional etiologic factors, still unknown, must be postulated.”

Psychophysiological research began in the first half of the 20th century. Researchers used techniques such as electromyogram (EMG), electroencephalogram (EEG), and galvanic skin response to examine physiologic parameters among individuals with a variety of psychosomatic illnesses. Research has evolved to the present day to include areas such as heart rate variability and stress and the role of the CNS in illnesses such as functional movement disorders, fibromyalgia, and chronic fatigue syndrome. Pain treatment research has evolved from less-controlled studies to randomized-controlled studies.

As early as the 1960s, researchers reported on the prevalence of various psychiatric problems in inpatient and outpatient medical settings. Methodologic variability led to wide variation in estimates of comorbidity. Early work also reported on the characteristics of patients referred for psychiatric consultation. R. J. Kahana’s and G. L. Bibring’s classic work in 1964 emphasized the importance of understanding personality types to the effective treatment of one’s patients.

C-L psychiatry research in the mid-20th century also elucidated the most common reasons for psychiatric consultation. While many of those reasons hold true to this day, psychoanalytic influences on physician consulting behavior have clearly waned in recent years. Consider Zbigniew Lipowski’s summary in 1967 of the two most common diagnostic questions asked of the consulting psychiatrist:

Table 27.1–1.
Typical Reasons for a Psychiatric Consultation (After Z. Lipowski)

1. Suicidal attempt or threat
2. Grossly disturbed behavior, e.g., delirium and psychosis
3. Excessive emotional reactions, e.g., depression and anger
4. Refusal to cooperate
5. Delayed convalescence
6. Conflict between patient and personnel
7. Patient with a psychiatric history
8. Psychiatric side effects of drugs
9. Selection and/or preparation of patients, e.g., for surgery
10. Disposition

1. Is this particular bodily complaint (or set of complaints) explicable as an expression of psychological ill-health, that is, is it partly or wholly psychogenic?
2. Is the inner experience of psychic change reported by the patient or his observable behavior explicable as the direct results of, or a psychological reaction to, organic disease?

However, reasons for requests for assistance in patient management have seemingly changed very little since Lipowski summarized them in 1967 (Table 27.1–1) and will be familiar to all currently practicing consultation-liaison psychiatrists.

Table 27.1–2 outlines the history of psychosomatic medicine.

CURRENT TRENDS

The practice of C-L psychiatry has evolved considerably since its early clinical origins and has come to focus on psychiatric illnesses that occur in the setting of physical health care. In large part, this evolution has occurred as a result of the increased complexity of medicine, the increased understanding of the relationship of medical illness to psychiatric illness, and the greater appreciation of mind and body as one. A key outcome of this has been the granting of subspecialty status for C-L psychiatry. Clinical care is now delivered in a variety of health care settings and utilizes an ever-expanding set of diagnostic tools as well as many effective somatic and psychotherapeutic interventions. Research in the area has progressed to include a greater understanding of the relationship between chronic medical conditions and psychiatric disorders and has examined the pathophysiologic relationships, the epidemiology of comorbid medical, and psychiatric disorders, and the role that specific interventions play in physiologic, clinical, and economic outcomes.

Table 27.1–2.

Major Conceptual Trends in the History of Psychosomatic Medicine

I. Psychoanalytic

Sigmund Freud (1900): Somatic involvement occurs in conversion hysteria, which is psychogenic in origin—e.g., paralysis of an extremity. Conversion hysteria always has a primary psychic cause and meaning; i.e., it represents the symbolic substitutive expression of an unconscious conflict. It involves organs innervated only by the voluntary neuromuscular or the sensorimotor nervous system. Psychic energy that is dammed up is discharged through physiologic outlets.

Sandor Ferenczi (1910): The concept of conversion hysteria is applied to organs innervated by the autonomic nervous system; e.g., the bleeding of ulcerative colitis may be described as representing a specific psychic fantasy.

George Groddeck (1910): Clearly organic diseases, such as fever and hemorrhage, are held to have primary psychic meanings; i.e., they are interpreted as conversion symptoms that represent the expression of unconscious fantasies.

Franz Alexander (1934, 1968): Psychosomatic symptoms occur only in organs innervated by the autonomic nervous system and have no specific psychic meaning (in contrast to conversion hysteria) but are end results of physiologic accompaniments of certain specific unconscious repressed conflicts. Presented first conceptualization of the biopsychosocial model.

Helen Flanders Dunbar (1936): Specific personality types are associated with specific psychosomatic diseases, an idea similar to Meyer Friedman’s 1959 theory of the type A coronary type.

Peter Sifneos, John C. Nemiah (1970): Elaborated the concept of alexithymia. Developmental arrests in the capacity to identify and express conflict-related affect result in psychosomatic symptom formation.

II. Psychophysiologic

Walter Cannon (1927): Demonstrated the physiologic concomitants of some emotions and the important role of the autonomic nervous system in producing those reactions. The concept is based on Pavlovian behavioral experimental designs.

Harold Wolff (1943): Attempted to correlate life stress to physiologic response, using objective laboratory tests. Physiologic change, if prolonged, may lead to structural change. He established the basic research paradigm for the fields of psychoimmunology, psychocardiology, and psychoneuroendocrinology.

Hans Selye (1945): Under stress a general adaptation syndrome develops. Adrenal cortical hormones are responsible for the physiologic reaction.

Meyer Friedman (1959): Theory of type A personality as a risk factor for cardiovascular disease.

Robert Ader (2007): Beginning in the 1970s, established the basic concepts and the research methods for the field of psychoneuroimmunology.

III. Sociocultural

Karen Horney (1939), James Halliday (1948): Emphasized the influence of culture in the development of psychosomatic illness. They thought that culture influences the mother, who, in turn, affects the child through her relationship with the child—e.g., nursing, child-rearing, and anxiety transmission.

Thomas Holmes, Richard Rahe (1975): Correlated the severity and the number of recent stressful life events with the likelihood of disease.

IV. Systems theory

Adolph Meyer (1958): Formulated the psychobiologic approach to patient assessment that emphasizes the integrated assessment of developmental, psychological, social, environmental, and biologic aspects of the patient’s condition. Basic concept of the biopsychosocial model is implicit in his approach.

Zbigniew Lipowski (1970): A total approach to psychosomatic disease is necessary. External (ecologic, infectious, cultural, environmental), internal (emotional), genetic, somatic, and constitutional factors, as well as past and present history, are important and should be studied by investigators working in the various fields in which they are trained.

George Engel (1977): Coined the term “biopsychosocial” derived from general systems theory and based on conceptual ideas introduced much earlier by Alexander and Meyer.

Leon Eisenberg (1995): Contemporary psychiatric research demonstrates that the mind–brain responds to biologic and social

> vectors while being jointly constructed of both. Major brain pathways are specified in the genome; detailed connections are fashioned by, and consequently reflect, socially mediated experience in the world.

Psychiatric morbidity is very common in patients with medical conditions, with a prevalence ranging from approximately 20% to 65%, depending on the illness. Patients in the general hospital have the highest rate of psychiatric disorders when compared to community samples or patients in ambulatory primary care. For example, compared to community samples, depressive disorders in the general hospital are more than twice as common, and substance abuse is two to three times as common. Delirium occurs in approximately 20% of inpatients. Psychiatric morbidity has serious effects on medically ill patients and is often a risk factor for exacerbation of their medical conditions. It is well established that depression is both a risk factor and a poor prognostic indicator in coronary artery disease. Psychiatric illness worsens cardiac morbidity and mortality in patients with a history of myocardial infarction, diminishes glycemic control in patients with diabetes, and decreases return to functioning in patients experiencing a stroke. Depressive and anxiety disorders compound the disability associated with stroke. In the context of neurodegenerative diseases such as Parkinson disease or Alzheimer disease, depression, psychosis, and behavioral disturbances are significant predictors of functional decline, institutionalization, and caregiver burden. Hospitalized patients with delirium are significantly less likely to improve in function compared to patients without delirium. Delirium is associated with worse outcomes after surgery, even after controlling for the severity of medical illness.

In addition, depression and other mental disorders significantly impact the quality of life and the ability of patients to adhere to treatment regimens (e.g., in patients with diabetes mellitus). Psychiatric disorders are linked to nonadherence with antiretroviral therapy, adversely affecting the survival of human immunodeficiency virus (HIV)-infected patients, and they worsen the prognosis and quality of life of cancer patients. Patients with mental disorders also have a much higher chance of being homeless, living in poverty, or experiencing other adverse social determinants of health, such as financial and nutritional insecurity. All of these factors substantially increase health risks and outcomes.

Failure to identify, evaluate, diagnose, treat, or achieve symptom resolution of psychiatric morbidity in medical care settings results in significantly increased service utilization. For example, depression, dementia, and delirium are associated with higher utilization of medical care, both in the hospital (longer lengths of stay) and after discharge. Untreated depression is associated with higher medical utilization after hospital discharge and with higher mortality and morbidity in coronary disease, hypertension, diabetes, and stroke.

THE CLINICAL PRACTICE OF CONSULTATION-LIAISON PSYCHIATRY

The primary focus of C-L psychiatry is the diagnosis and treatment of psychiatric disorders in patients with complex medical conditions. Thus, C-L psychiatry addresses a variety of clinical problems that occur in patients being treated in medical settings, including psychiatric disorders such as delirium and dementia, which occur as a *result* of a medical condition; anxiety or depression, which occur *in the setting* of chronic medical conditions but may also be due to medical causes; and somatic symptom disorders. In addition to the care of specific conditions, C-L psychiatry encompasses a variety of other clinical activities that occur in the medical setting, including evaluation of decision-making capacity, attention to quality of life and symptom management in chronic and terminally ill individuals, provision of liaison services to medical staff, and assistance with management of primary psychiatric conditions such as bipolar disorder or schizophrenia. Table 27.1–3 summarizes the types of clinical problems addressed in C-L Psychiatry.

Table 27.1–3.
Summary of Clinical Problems in C-L Psychiatry

Type of Clinical Problem	Example
Psychiatric symptoms secondary to a medical condition	Delirium, dementia
Psychiatric symptoms as a reaction to medical condition or treatments	Anxiety related to chemotherapy; depression related to limb amputation
Psychiatric complications of medical conditions and treatments	Mania secondary to corticosteroid treatment
Psychological factors contributing to medical symptoms	Somatic symptom disorders
Medical complications of psychiatric conditions or treatment	Neuroleptic malignant syndrome; acute withdrawal from alcohol or other substance
Co-occurring medical and psychiatric conditions	Recurrence of depressive disorder in the setting of cancer treatment (conditions occur independently); schizophrenia in a patient with end-stage renal disease.
Psychiatric/psychosocial assessment	Decision-making capacity evaluation; evaluation prior to organ transplantation

C-L psychiatry is practiced in a variety of settings and encompasses a combination of consultative, treatment, and liaison activities. Historically, the majority of this work occurred within the general hospital setting. Largely this has been due to the fact that the greatest risk factor for requiring psychiatric intervention has been the severity of the medical illness. However, as more illnesses become chronic in nature (e.g., cancer and HIV), there has been an increase in the delivery of C-L services in outpatient settings as well. Regardless of where care is delivered, the basic approaches are similar: To conduct a psychiatric assessment, provide psychotherapeutic, behavioral, or pharmacologic interventions to patients, and work closely with other medical professionals in a liaison capacity.

EVALUATION PROCESS IN CONSULTATION-LIAISON PSYCHIATRY

Evaluation in the medical setting includes a standard psychiatric assessment as well as a particular focus on the medical history and context of physical health care. In addition to obtaining a complete psychiatric history, including a past history, family history, developmental history, and a review of systems, the medical history, and current treatments should be reviewed and documented. A full mental status examination should be completed, and focused neurologic and physical examinations are often indicated depending on the nature of the presenting problem.

Several components of the process should be emphasized. First, it is critical to communicate with the referring clinician. In many cases, the reason for consultation is not clear, and it is often helpful to understand why the treating team believes the patient might have a psychiatric problem. Such information can guide the assessment and may provide information about other concerns the treatment team might have, including the patient's behavior and over or underuse of treatment, such as pain medication. It may also reveal the team's frustration with a particular patient. Second, collateral information from family members and other health care providers, including mental health professionals, is important in understanding any prior history of psychiatric problems and helping the team plan for ongoing care. Third, while the interview of the patient should be comprehensive, it also needs to be sensitive to the patient's medical status. In many cases, patients in the general medical setting are experiencing pain or discomfort, distracted by medical problems, or cognitively compromised. Thus, the interview may need to be focused on the reason for consultation.

A full mental status examination, including a specific focus on cognitive functioning, should be completed. The mental status should address the level of consciousness, attention, concentration, memory, executive function, language, praxis, mood, affect, perception, judgment, and insight. The Mini-Mental State Examination or Montreal Cognitive Assessment is often a helpful component because it provides an objective score that can serve as a baseline for retesting, especially when delirium or dementia is a concern.

The psychiatric consultant should also ensure that appropriate laboratory and other diagnostic evaluations have been conducted. Basic chemistry, hematologic, and cardiovascular workup can be supplemented with other tests, including thyroid function, liver function, vitamin B_{12}, folate, toxicology, serologic tests for syphilis, and HIV testing. Additional diagnostic studies, including computed tomography (CT) or magnetic resonance imaging (MRI), scans of the brain, EEG, or lumbar puncture, may also be indicated.

Another important objective of the psychiatric evaluation is to gain an understanding of the patient's experience of their illness. In many cases, this becomes the central focus for both the psychiatric assessment and interventions. It is often helpful to develop an understanding of the patient's developmental and personal history as well as key dynamic conflicts, which in turn may help to make the patient's experience with illness more comprehensible. Such an evaluation can include the use of the concepts of stress, personality traits, coping strategies, and defense mechanisms. Observations and hypotheses that are developed can help to guide psychotherapy with patients aimed at diminishing distress and may also be helpful for the primary medical team in their interactions with the patient.

Finally, a full report synthesizing the information should be completed and should include specific recommendations for additional evaluations and intervention. The psychiatrist should limit information in the report to that which is medically necessary. Ideally, the report should be accompanied by a discussion with the referring physician.

TREATMENTS USED IN CONSULTATION-LIAISON PSYCHIATRY

A host of interventions have been successfully utilized in C-L psychiatry. Specific consideration must be given to medical illness and treatments when making recommendations for psychotropic medications. Psychotherapy also plays an important role and may vary in its structure and outcomes as compared to therapy that occurs in mental health practice.

Psychopharmacologic recommendations need to consider several important factors. In addition to targeting a patient's active symptoms, considering the history of illness and treatments, and weighing the particular side effect profile of a particular medication, there are several other factors that must be considered that relate to the patient's medical illness and treatment. It is critical to evaluate potential drug–drug interactions and contraindications to the use of potential psychotropic agents. Since the majority of psychotropic medications are metabolized in the liver, awareness of liver function is important. General appreciation of side effects, such as weight gain, risk of development of diabetes, and cardiovascular risk, must be considered in the choice of medications. In addition, it is also important to incorporate knowledge of recent data that outline the effectiveness and specific risks involved for patients with co-occurring psychiatric and physical disorders. For example, a greater understanding of the side effects of antipsychotic medications has raised concerns about the use of these medications in patients with dementia.

Psychosocial interventions also require adaptation when used in this population. The methods used and the goals of psychosocial interventions in the medically ill are often determined by the consideration of disease onset, etiology, course, prognosis, treatment, and understanding of the nature of the presenting psychiatric symptoms, in addition to an understanding of the patient's existing coping skills and social support networks.

Every major psychosocial intervention has been used in C-L psychiatry, including dynamic, supportive, and cognitive behavioral interventions, delivered in family, group, and individual settings. The goals for psychosocial interventions vary widely depending on the nature of both the psychiatric issues and the medical illness. In addition, the specific intervention will vary based on the illnesses and goals. For instance, while a goal generally present for all medical illnesses may be addressing changes in role and function resulting from illness, addressing concerns regarding death and dying will be a key concern for those in the terminal phases of an illness such as cancer. Coping with changes in physical function and pain are key issues that can be addressed for patients with arthritis. Addressing barriers to adherence to treatment becomes critical to psychosocial interventions in patients with diabetes.

An extensive body of research has demonstrated positive clinical outcomes associated with a variety of interventions. At a minimum, psychosocial interventions have been shown to improve quality of life and psychiatric conditions; in addition, there is some evidence that psychosocial interventions may positively impact treatment adherence and medical outcomes. Interventions aimed at depression in patients with diabetes have shown not only decreases in depressive symptoms but also reduction in medical costs of care.

MENTAL DISORDERS IN PRIMARY CARE

An area of considerable focus has been the successful identification and management of patients with mental disorders treated in the primary care setting. Mental health conditions are highly prevalent, ranging from 30% to 60% depending on the chronic health condition being treated, but are often not recognized and not effectively treated in most cases. In addition, patients with mental health conditions and other chronic health conditions are two to three times more expensive than those without psychiatric comorbidities. Although a variety of interventions, including education of primary care physicians (PCPs), screening initiatives, and implementation of quality improvement programs, may have a short-term benefit, they have not demonstrated a long-term impact on the detection and treatment of mental health and substance use conditions.

Programs that incorporate access to mental health services within primary care have resulted in improved detection and treatment of these common conditions. Colocation programs, often described as primary care behavioral health (PCBH), increase interaction between mental health providers and PCPs and may be more convenient for patients receiving care in primary care settings. PCBH has been shown to improve outcomes for those patients participating in the programs but is limited in its ability to improve screening, diagnosis, or treatment outcomes over larger populations.

The Collaborative Care Model (CoCM) provides a systematic approach to the care of mental health conditions in the primary care setting. In this approach, a behavioral health care manager and a psychiatrist (who serves as a consultant) work with the PCP to screen, diagnose, and treat patients with mental health conditions. Randomized-controlled trials of CoCMs for treating depression and other common mental health and substance use conditions in primary care settings have shown that they significantly improve the quality of care and outcomes for mental health disorders (such as depression and anxiety), improve patient and provider satisfaction, and increase patient adherence to the treatment regimen. Evidence supports its use for most mental health and substance use conditions. Reimbursement for the service began in 2017, and since that time, there has been increasing availability of the model.

RESEARCH IN PSYCHOSOMATIC MEDICINE

Research in psychosomatic medicine has focused on several key areas that have contributed to the current understanding of the burdens of co-occurring psychiatric and medical disorders. Important advances in psychosomatic medicine research have identified common pathways between psychiatric illness and physical illnesses, such as cardiovascular disease or diabetes, a bidirectional risk associated with the presence of psychiatric disorders or medical conditions such as stroke, and the identification of clinical interventions that can impact both medical and psychiatric outcomes. However, research in psychosomatic medicine has been limited by the complexity of research questions that require investigation of both medical and psychiatric outcomes and the existing “division” of the research enterprise into separate psychiatric and medical domains.

Table 27.1–4.

Current Research in Psychosomatic Medicine

Types of Studies	Examples of Current Research
Epidemiologic Studies	
Occurrence, nosology, and impact of different types of psychiatric morbidity over the course of general medical conditions	PTSD and other psychiatric comorbidities in individuals with traumatic brain injury
Mechanistic Studies	
Understanding the role of psychosocial, biologic, and genetic factors in the causation of psychiatric and medical morbidity	Affective circuitry in chronic pain Social support and biologic changes such as inflammation Social determinants of health
Examining the social, psychological, biologic, and genetic factors involved in somatic symptom disorders, chronic pain, and fatigue	Depression and chronic pain Functional brain imaging abnormalities in chronic fatigue and conversion
The effects of medical illnesses and their treatments on the brain and on the development of secondary psychiatric disorders	Medications that may cause delirium Paraneoplastic limbic encephalitis
Understanding how psychiatric morbidity affects medical outcomes	Depression and adherence to medical treatments
Treatment Studies	
Efficacy and safety of psychiatric treatments as applied to the complex medically ill	Cognitive behavioral therapy in somatic symptom disorders The efficacy and safety of α-2 adrenoreceptor agonists in treating delirium
How treating psychiatric conditions may affect physical	The effect of treating depression on hemoglobin A_{1c} levels in diabetic

health outcomes	patients. The impact of brief psychotherapy interventions on immune response across different disease states. The effect of mindfulness therapies on a patient with degenerative neurologic diseases.
Service Delivery Studies	
The delivery of efficacious psychiatric treatments for the complex medically ill across different service settings, such as primary care, home-based care, and long-term care settings	Use of collaborative care strategies in primary and specialty care settings
Development of population health models that incorporate risk for psychiatric disease into the overall management of health and economic outcomes	Use of risk modeling to identify patients who can benefit from psychiatric consultation in inpatient settings. Use of implementation science to maximize the use of population-based strategies

Various conceptual models of the relationship between medical and physical illness underlie the core research areas. For example, a large body of epidemiologic work has examined the relationship of depressive disorders to poor outcomes among patients with cardiovascular disease. In contrast, in the field of psychoneuroimmunology, basic science techniques are used to examine the relationship between psychosocial stressors and immune functioning. Table 27.1–4 presents an overview of conceptual areas and specific current research areas.

CURRENT PSYCHOSOMATIC MEDICINE RESEARCH

Psychosomatic medicine is now involved with a wide, ever-expanding spectrum of investigations examining the medical illness–psychiatry relationship. Important contributions have occurred in acquired immunodeficiency syndrome (AIDS), cancer, transplantation, cardiac, neurologic, pulmonary, renal, GI disease, and obstetrics and gynecology. In each of these areas, first-generation studies identified the extent and nature of psychiatric morbidity associated with the most common diseases. More sophisticated second-generation cross-sectional epidemiologic studies have established the prevalence rates of a broader range of psychiatric disorders in several medical conditions, as well as newer psychiatric syndromes, such as mild traumatic brain injury.

In some cases, a new body of research has extended the knowledge about genetic, neurochemical, and behavioral factors contributing to the development of psychiatric disorders among complex medically ill populations. This work has shown that many different mechanisms are involved. Examples include the relationships between depression and diabetes and between specific autoantibodies and limbic encephalitis. This research has led to some intervention studies focused on reducing the rate of occurrence of these disorders, for example, hospital programs that reduce the incidence of delirium.

Of long-standing interest to psychosomatic medicine has been the relationship between emotional factors and disease development. Recent studies have demonstrated the dramatic impact of coexisting psychiatric morbidity on the course or outcome of specific medical illnesses. Studies have also been extended to examine the effect of psychiatric comorbidity on health costs. Comorbid psychiatric disorders have been shown to increase the length of hospital stay and rehospitalization rates in numerous illnesses. Several cost-effectiveness studies have demonstrated the value of C-L psychiatry interventions in hospital and outpatient populations. In addition, psychobiologic research continues to focus on emotional factors and illness development in individuals who do not necessarily have psychiatric illnesses. For example, there continues to be important work in the relationship between emotional stressors and the functioning of the immune system.

THE FIELD OF PSYCHOSOMATIC MEDICINE AND C-L PSYCHIATRY

The term “Psychosomatic Medicine” was first used by Felix Deutsch in the early 1920s. Organized psychosomatic medicine and C-L psychiatry had their origins in the publication of the journal *Psychosomatic Medicine* in 1939, the founding of the American Psychosomatic Society in 1947, and the journal *Psychosomatics* in 1960. In 1968, John Schwab published the first C-L psychiatry textbook. The Academy of Psychosomatic Medicine was founded in 1953 and has been the national organization for C-L psychiatrists. In 2017, the organization adopted the new name Academy of Consultation-Liaison Psychiatry.

The origins of C-L psychiatry fellowships can be traced to George Engel's and John Romano's fellowship program in psychosocial skills at Rochester University College of Medicine. This program, first offered in 1946, was offered to nonpsychiatric physicians as well as psychiatrists. The Massachusetts General Hospital established its first consultation service in 1954.

In 2003, the ABMS approved psychosomatic medicine as a subspecialty field of psychiatry. The approval of the subspecialty was both a response to and an opportunity for the field to highlight and expand the foundation in clinical care, research, teaching, and other scholarly activities related to the care of this special patient population. Although C-L specialists do not provide all, or even most, of the care to this patient population, the presence of the subspecialty clearly has had a positive impact on general training, research, and academic psychiatry. In 2017, the ABMS approved a name change of the field to C-L psychiatry.

C-L psychiatry, as defined and characterized by the subspecialty, is practiced by psychiatrists. The Academy of Consultation-Liaison Psychiatry has approximately 2,000 members, and about 3,000 psychiatrist members of the APA state that they are interested in or spend a substantial amount of time providing consultation in a general hospital setting. Other mental health professionals also provide care for this population. Psychologists have long served in the general medical setting providing assessments, psychotherapeutic, and behavioral interventions, and conducting research. The health psychology section of the American Psychological Association has approximately 3,000 members. In addition, social workers and nurses participate in care for these patients, including in collaborative care programs. In many settings, C-L psychiatrists function as part of multidisciplinary teams, which include psychiatry, psychology, nursing, and social work. In most cases, these teams work within medical specialty settings that provide care to special populations such as those with HIV, cancer, cardiovascular disease, transplant, and diabetes.

The APA formed the Council on Psychosomatic Medicine (now Council on Consultation-Liaison Psychiatry) in 2004 in response to the approval of the subspecialty. The APA Council and the Academy of Consultation-Liaison Psychiatry, as well as other organizations, have participated actively in the development of clinical standards, addressed issues related to access to care and reimbursement, and provided education to members, trainees, medical professionals, and the public. Care of special populations such as those with cancer, HIV, transplant, and obstetrics and gynecology is the focus of other organizations (e.g., American Psychosocial Oncology Society and North American Society for Psychosocial Obstetrics and Gynecology). Research is a core interest for some organizations, including the American Psychosomatic Society. In addition to the organizations in the United States, there has been continued strong interest in the field internationally (e.g., The European Association of Psychosomatic Medicine).

Education and training in C-L psychiatry are integrated throughout the psychiatry curriculum in medical school and general psychiatric training. Starting in 1961, the Accreditation Council for Graduate Medical Education (ACGME) addressed the need for psychiatric residents to collaborate with other services besides psychiatry, including medicine, pediatrics, and neurology. By the 1980s, the ACGME's Residency Review Committee (RRC) required all psychiatry residency programs to provide teaching and experience in the specialty. The current RRC requirements require at least 2 months of full-time work in which residents consult under supervision on other medical and surgical services.

Currently, there are 62 ACGME-accredited C-L fellowship programs in the United States. The ACGME program requirements for C-L psychiatry fellowships can be found on their website (www.acgme.org). Table 27.1–5 lists the clinical activities in which C-L psychiatry fellows must demonstrate competence. In addition, there are over 30 residencies that combine psychiatry with internal medicine, family practice, pediatrics, or neurology.

Table 27.1–5.
Patient Care Competencies Required of All Consultation-Liaison Psychiatry Fellows

1. Establishing rapport with all medical patients.
2. Diagnosing and treating psychiatric disturbances that occur among the physically ill.
3. Conducting psychiatric evaluations of individuals who have psychiatric complications of medical illnesses; psychiatric complications of medical treatments, including medications, traditional and new surgical or medical procedures, transplantation, and experimental therapies; and typical and atypical presentations of psychiatric disorders due to medical, neurologic, and surgical illnesses.
4. Evaluating and managing individuals with acute and chronic pain; delirium, dementia, and psychiatric disorders due to medical illness; somatic symptom disorders; palliative care and end-of-life issues; and issues in adjusting to the emotional stresses of medical illness.
5. Assessing the capacity of individuals to give informed consent for medical and surgical procedures in the presence of cognitive impairment.
6. Providing psychosocial interventions, including psychotherapeutic interventions appropriate for the medically ill.
7. The appropriate use of psychoactive medication in medical, neurologic, obstetrical, and surgical conditions.
8. Assessing and managing suicidality and other high-risk behavior in the medical setting.

Adapted from ACGME Program Requirements for Graduate Medical Education in Consultation-Liaison Psychiatry.
https://acgme.org/Portals/0/PFAssets/ProgramRequirements/409_ConsultationLiaisonPsychiatry_2020.pdf?ver=2020-06-19-132534-420.

ACCESS TO CONSULTATION-LIAISON PSYCHIATRIC CARE: IMPLICATIONS FOR HEALTH POLICY

Despite an abundance of evidence that psychiatric and psychosocial interventions have positive economic and clinical outcomes for patients with co-occurring psychiatric and medical problems, access to and reimbursement for psychiatric and other mental health services are severely limited. In the United States, reimbursement for these services is complex due to the separation of mental health and physical health benefits in most private and public insurance coverage. Even in areas where there is reimbursement for mental health services, it is often administered through a mental health carve-out that separates reimbursement for these services from the medical care reimbursement, resulting in patients not being able to receive care in the medical setting or have access to professionals with the appropriate expertise to treat psychiatric conditions in the context of complex medical illnesses.

Over the past several years, health care reimbursement has been shifting from paying for services delivered (fee for service) to paying for value. Value-based payments range from increasing payment based on provider or health system improvement in outcomes or other measures of quality to putting health systems fully at financial risk for the delivery of care to a population. Given the high cost of care for patients with comorbid mental health and other nonmental health conditions, treatment provided by C-L psychiatrists can help to improve both clinical and economic outcomes. C-L psychiatrists are participating in efforts to define processes of care, including proactive inpatient consultations, depression screening and treatment, use of validated rating scales in treatment of mental disorders in all care settings, and measurement of outcomes of care. These measures can support the roles of psychiatry in the medical setting in value-based care.

ACKNOWLEDGMENT

The founding editor of this textbook, Harold I. Kaplan MD, had a long-standing interest in psychosomatic medicine and the history of psychiatry. He wrote the section on the history of psychosomatic medicine for the first edition of this textbook and for each subsequent edition thereafter until his death in 1998. This chapter includes some of his original work.

FURTHER READINGS

American Psychiatric Association, Academy of Psychosomatic Medicine. Dissemination of integrated care within adult primary care settings: the collaborative care model. Report, 2016. https://www.psychiatry.org/psychiatrists/practice/professional-interests/integrated-care/collaborative-care-model

Alexander F. *Psychosomatic Medicine: Its Principles and Application*. W.W. Norton; 1950.

Archer J, Bower P, Bilbody S. et al. Collaborative care for depression and anxiety problems. *Cochrane Database Syst Rev*. 2012;10CD006525.

Cannon WB. *The Wisdom of the Body*. W.W. Norton; 1932.

Deutsch F. The choice of organ in organ neurosis. *Int J Psychoanal*. 1939;20:1.

Dunbar F. *Emotions and Bodily Changes*. Columbia University Press; 1954.

Eisenberg L. The social construction of the human brain. *Am J Psychiatry*. 1995;152:1563.

Engel GL. The need for a new medical model: a challenge for biomedicine. *Science*. 1977;196:129.

Freud S. Fragment of an analysis of a case of hysteria. In: Strachey J, ed. *Standard Edition of the Complete Psychological Works of Sigmund Freud*. Vol 7. Hogarth Press; 1953:40.

Friedman M, Byers SO, Rosenman RH, Elwitch FR. Coronary-prone individuals (type A behavior pattern): some biochemical characteristics. *JAMA*. 1970;212:1030.

Gitlin DF, Levenson JL, Lyketsos CG. Psychosomatic medicine: a new psychiatric specialty. *Acad Psychiatry*. 2004;28:4.

Grinker RR, Robbins FP. *Psychosomatic Case Book*. Blakiston; 1954.

Holmes TH, Rahe RH. The social readjustment rating scale. *J Psychosom Res*. 1967;11:213.

Horney K. *The Neurotic Personality of Our Time*. W.W. Norton; 1937.

*Kaplan HI. History of psychosomatic medicine. In: Sadock BJ, Sadock VA, eds. *Kaplan and Sadock's Comprehensive Textbook of Psychiatry*. 8th ed. Lippincott Williams & Wilkins; 2005:2105.

Katon WJ, Lin EHB, Von Korff M, et al. Collaborative care for patients with depression and chronic illnesses. *N Engl J Med*. 2010;363:2611–2620.

Keefe FJ, Lumley MA, Buffington ALH, Carson JW, Studts JL. Changing face of pain: evolution of pain research in psychosomatic medicine. *Psychosom Med*. 2002;64:921.

Levenson JL. *The American Psychiatric Publishing Textbook of Psychosomatic Medicine: Psychiatric Care of the Medically Ill*. 3rd ed. American Psychiatric Publishing; 2019.

Levenstein S, Rosenstock S, Jacobsen RK, Jorgensen T. Psychological stress increases risk for peptic ulcer, regardless of Helicobacter pylori infection or use of nonsteroidal anti-inflammatory drugs. *Clin Gastroenterol Hepatol*. 2015;13(3):498.

Lichtman JH, Froelicher ES, Blumenthal JA, et al; American Heart Association Statistics Committee of the Council on Epidemiology and Prevention and the Council on Cardiovascular and Stroke Nursing. Depression as a risk factor for poor prognosis among patients with acute coronary syndrome: systematic review and recommendations: a scientific statement from the American Heart Association. *Circulation*. 2014;129:1350–1369.

Lipowski ZJ. Review of consultation psychiatry and psychosomatic medicine. I. General principles. *Psychosom Med*. 1967;29:153.

*Lipsitt DR. Consultation-liaison psychiatry and psychosomatic medicine: the company they keep. *Psychosom Med*. 2001;63:896.

Melek SP, Norris DT, Paulus J, Matthews K, Weaver A, Davenport S. *Potential Economic Impact of Integrated Medical-Behavioral Healthcare: Updated Projections for 2017*. Milliman, Inc.; 2018, monograph.

Novack DH. Realizing Engel's vision: psychosomatic medicine and the education of physician-healers. *Psychosom Med*. 2003;65:925.

Pence BW, O'Donnell JK, Gaynes BN. The Depression treatment cascade in primary care: a public health perspective. *Curr Psychiatry Rep*. 2012;14(4):328–335.

Possemato K, Johnson EM, Beehler G, et al. Patient outcomes associated with primary care behavioral health services: a systematic review. *Gen Hosp Psychiatry*. 2018;53:1–11.

Ridley M, Rao G, Schilbach F, Patel V. Poverty, depression, and anxiety: causal evidence and mechanisms. *Science*. 2020;370:eaay0214.

*Shorter E. *From Paralysis to Fatigue: A History of Psychosomatic Illness in the Modern Era*. Free Press; 1992.

Stern TA, Freudenreich O, Smith FA, Fricchione GL, Rosenbaum JF. *Massachusetts General Hospital Handbook of General Hospital Psychiatry*. 7th ed. Elsevier; 2018.

27.2 Cardiovascular Disorders

JUDITH RESTREPO, M.D., MARGO C. FUNK, M.D., M.A., F.A.C.L.P., AND SCOTT BEACH, M.D.

Cardiovascular disorders are the leading cause of death in the United States and the industrialized world. Coronary artery disease (CAD) is one of the leading causes of worldwide disability burden. Psychiatric disorders are widely associated with cardiovascular disease, both in the development of cardiovascular disease itself and as a complication. Comorbidity poses a treatment challenge for cardiology and psychiatry. Behavioral risk factors, such as smoking, sedentary lifestyle, and failure to adhere to medical treatment recommendations are clearly exacerbated by psychological stress and psychiatric illness. Moreover, psychiatric disorders are common in many cardiac patients, and intervention not only improves adherence to and outcomes of medical therapy but also relieves suffering due to the psychiatric illnesses themselves. From an epidemiologic standpoint, the most important categories of cardiovascular disease include CAD, heart failure (HF), arrhythmias, and valvular and myocardial disease. This chapter provides a brief presentation of background information on these cardiovascular conditions and is accompanied by a discussion of psychological conditions and psychiatric disorders as risk factors for their development and course. The chapter also includes some discussion of the cardiac impacts of psychopharmacology and other conditions of interest before turning to psychiatric comorbidity with established cardiovascular disease and relevant treatment issues.

CAD AND HTN

Brief Review of Symptoms Treatment and Epidemiology

Atherosclerotic CAD is a progressive physiologic process characterized by the formation of atherosclerotic plaque in one or more of the coronary arteries, which then can have progressive effects on the heart. Acute episodes of illness occur owing to the formation of thrombus on the surface of a disrupted atherosclerotic plaque, leading to total or subtotal occlusion of the affected coronary artery, abrupt myocardial ischemia, and, if not rapidly corrected, myocardial cell death. These acute coronary syndromes (ACS)—unstable angina, myocardial infarction (MI), and cardiac arrest—lead to death or progressive disability due to recurrent pain, cardiac arrhythmias, and HF. Approximately 18.2 million adults older than 20 years of age in the United States have CAD, making it the most common type of heart disease. In the United States, the incidence of CAD is more than 5 million cases per year, and more than 600,000 persons per year have a first MI. Established risk factors for CAD include family history, male sex, hypertension (HTN), hyperlipidemia (HLD), diabetes (DM), sedentary lifestyle, obesity, and smoking. Premenopausal women have reduced risk compared to men, but the incidence of clinically significant disease rises after menopause to match that of men. Atherosclerosis generally begins early in life but is not clinically evident until midadult life. Behavioral and psychological factors have increasingly been recognized to contribute to the risk of incident CAD and disease progression. Notably, depression is independently associated with an almost doubled risk of the development of CAD. Current practice in the treatment of ACS attempts urgent thrombolysis or revascularization by a percutaneous coronary intervention (PCI) procedure (angioplasty, stent placement, atherectomy) or coronary artery bypass graft (CABG) surgery. Preventive measures and medical management include β-adrenergic blockade; statin therapy to stabilize plaque; and aspirin or antiplatelet drugs, exercise, abstinence from or cessation of smoking, blood pressure control, and maintenance of normoglycemia. Complications of acute or progressive disease, such as arrhythmias and HF, must also be treated.

Role of Depression, Anxiety, and Post-Traumatic Stress Disorder

Depression, anxiety, post-traumatic stress disorder (PTSD), and acute and chronic mental stress have been evaluated as risk factors for the development and progression of CAD. Structural impacts of mental health such as low socioeconomic status and low social support have been shown to have significant relationships with each of these individual psychological factors, and some investigators have proposed these latter characteristics as more promising indices of psychological risk. Data from the Normative Aging Study demonstrate a dose–response relationship between negative emotions (a combination of anxiety and depression symptoms) and incident CAD.

Depression. Depression is strongly associated with an increased risk for the development and progression of CAD. Numerous large-scale, prospective, epidemiologic studies reflect an increased relative risk of first MI and cardiac-related mortality in those with depression. This finding holds even after controlling for smoking, a potent risk factor for cardiovascular morbidity and mortality. Smoking as an additive risk remains far more prevalent in those with depression than in the population at large. The effect of depression on risk for MI remains in follow-up over 10 to 20 years. Meta-analyses have also demonstrated patients with pre-existing CAD and depression have doubled risk of adverse coronary disease–related outcomes, including MI, revascularization procedures for unstable angina, and death, as compared to nondepressed counterparts. Depression also has been shown to more strongly predict reduced perceived quality of life (QOL) and physical functioning than physiologic cardiac limitations such as reduced ejection fraction or exercise-induced myocardial ischemia. A widely replicated relationship appears to exist between the severity of depression symptoms after acute MI or unstable angina and the risk of death over a 5-year follow-up, even after controlling for other significant prognostic factors (Fig. 27.2–1). Depression is also associated with an increased risk of death at 5 and 10 years of follow-up as related to the intervention of CABG, including immediately prior to or at 6 months postintervention. In addition, poor response to antidepressant treatment after an MI is associated with a higher risk of subsequent cardiac events and death.

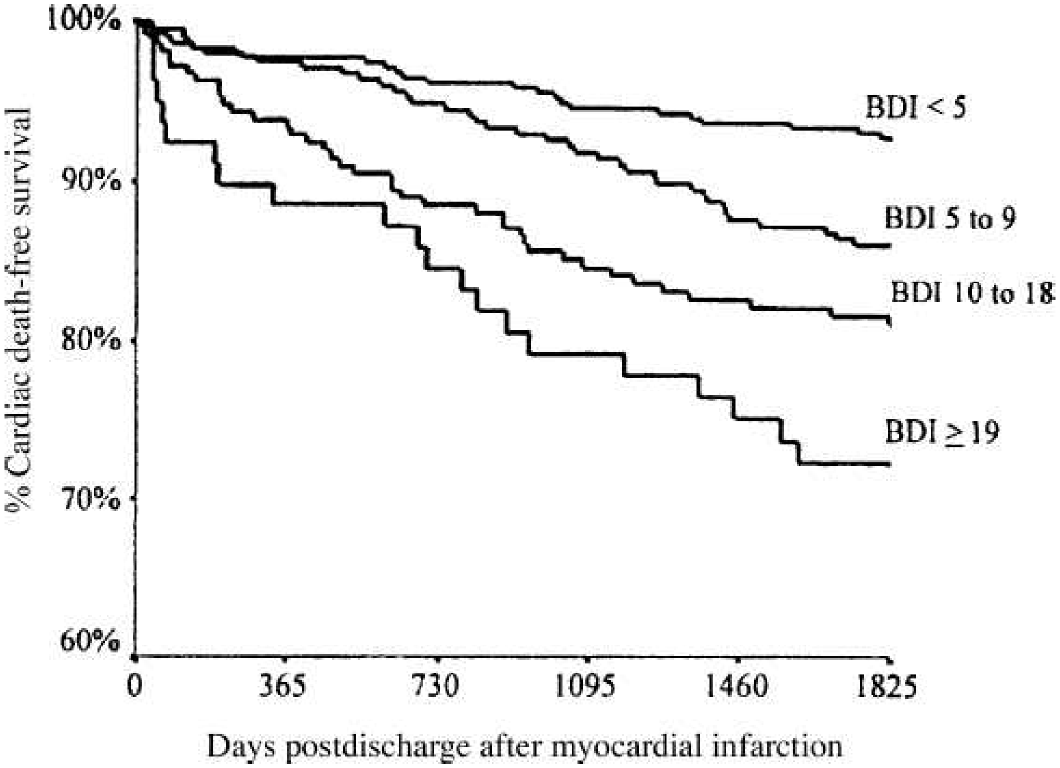

FIGURE 27.2–1. Long-term survival after myocardial infarction in relation to Beck Depression Index (BDI) score during hospitalization. (Reprinted with permission from Lespérance F, Frasure-Smith N, Talajic M, Bourassa MG. Five-year risk of cardiac mortality in relation to initial severity and one-year changes in depression symptoms after myocardial infarction. *Circulation*. 2002;105(9):1049.)

The mechanisms by which depression may increase coronary disease risk have not been established, but potential mechanisms may be organized into behavioral and biologic pathways. Behavioral mechanisms include social isolation, physical inactivity, poor adherence, and smoking. The relationship between cigarette smoking and mental illness, particularly depression, has been well documented. History of depression is associated with a lifetime risk of smoking, current smoking, and a lower successful quit rate. As the prevalence of smoking trends downward in the United States, these associations have strengthened.

Biologic pathways include links of depression with abnormal neuroendocrine and autonomic function, smoking, and platelet dysfunction. Hypothalamic–pituitary–adrenal axis dysregulation in depression is manifest by elevated levels of circulating cortisol and loss of the usual diurnal variation in cortisol level. Elevated cortisol has toxic effects on the coronary artery endothelium and plays a role in plaque development. Disordered platelet aggregation, leading to increased thrombus formation, may also play a role in increasing the risk of coronary events in depression. Measures of in vivo platelet activation and platelet aggregation show that depressed patients exhibit greater prothrombotic activity at baseline and greater platelet activation by an orthostatic challenge. Serotonergic antidepressants, especially those with a high affinity for the serotonin reuptake transporter, reduce platelet activation, and several studies indicate that they reduce the likelihood of first MI.

Anxiety Disorders and Post-Traumatic Stress Disorder. Anxiety is common in those with CAD. The estimated 12-month prevalence of any anxiety disorder in those diagnosed with CAD is estimated to be over 10% in men and almost 25% in women. Most studies have focused on the role of depression as an independent risk for the development of CAD; however, anxiety has also been shown to independently impact the prognosis of those with CAD outside of depression. Panic disorder, in particular, has been associated with CAD. Anxiety has been associated with the progression of pre-existing CAD. Studies have demonstrated a specific increase in the risk of cardiac death in those with unmanaged anxiety. Like depression, the behavioral impact of anxiety remains a major consideration as a mechanism for its impact on CAD. Anxiety, like depression, is associated with smoking and a sedentary lifestyle. Given the correlation between depression and anxiety, it is difficult to fully differentiate the impacts of anxiety in isolation without more longitudinal and large population studies. Less data are available regarding whether treatment of anxiety is beneficial in impacting the incidence and progression of cardiac disease.

The impact of PTSD on the development and course of cardiovascular and cerebrovascular disease has become a topic of considerable recent investigation. Military combat veterans with PTSD have an increased risk of the development of CAD, HF, and peripheral vascular and cerebrovascular disease. The estimated prevalence of PTSD in those with CAD has varied from 4% to 24%. PTSD has been induced by cardiac disease and its interventions at a prevalence rate estimated to be 12% on average. The effects remain significant after adjustment for smoking, depression, and other CAD risk factors. Both behavioral and physiologic processes are likely to be involved as mediators of risk of CAD, including smoking, difficulty adhering to a healthy lifestyle, and autonomic dysfunction with abnormal sympathetic nervous system activation. Even after adjustment for medication adherence, PTSD is associated with increased morbidity and mortality as related to cardiac disease.

Stress and Behavioral Risk Factors

The focus of study regarding correlations of stress and CAD has largely been related to the impact of stress on blood pressure, given the clear link of HTN to the development of CAD. Laboratory and ambulatory monitoring studies clearly demonstrate that acute mental stress results in transient elevation of systolic and diastolic blood pressure. Some studies have found that these transient hemodynamic responses predict increased blood pressure and incident HTN over long-term follow-up. Some population-based cohort studies, examining stress, negative affect, and anxiety as predictors of incident HTN, have found a positive association, with 1.5- to 3-fold increased risk, while others have found none or even a negative association.

HTN is an asymptomatic chronic condition that eventually results from and is highly associated with the risk of CAD, other vascular diseases and their complications, MI, and cardiomyopathy. In the United States, 20% to 50% of adults have a diagnosis of HTN. Epidemiologic risk factors for the development of HTN include male sex, family history, non-White race, and high dietary sodium intake. The role of psychosocial risk factors in the development of HTN is controversial. Anxiety, acute mental stress, chronic anger, and job strain have been associated with the risk of HTN; however, epidemiologic studies have not yielded consistent associations.

Acute mental stress induces arterial endothelial dysfunction and paradoxical vasoconstriction in atherosclerotic arterial segments. Acute mental stress has a significant effect on coronary artery blood flow that may be of significance in patients with pre-existing CAD. States of fear, excitement, and, especially, acute anger reduce blood flow through atherosclerotic coronary segments, provoke coronary spasm, and are associated with abnormal left ventricular wall motion and electrocardiogram (ECG) evidence of myocardial ischemia. Mental stress–induced ischemia is associated with an increased risk of subsequent cardiac events in patients with known CAD and may occur even in patients who do not demonstrate evidence of ischemia during exercise stress testing. Relaxation training can alter autonomic activation during mental stress, and a potential therapeutic role for such training may exist in stress-induced ischemia. There is also evidence that sertraline (Zoloft) treatment reduces myocardial ischemia induced by mental stress.

Studies also suggest chronic mental or emotional stress may also play a role in the development and progression of CAD. Longitudinal community-based studies have found that those who reported persistently difficult close social relationships or reported job strain had an independent elevated risk of new-onset CAD or recurrent cardiac events. In the largest case-control study of predictors of MI, the INTERHEART study found that psychosocial factors, defined as a combination of stress and depression, ranked as the third-highest predictor of MI, raising the risk of MI to comparable levels of smoking or DM. Almost half of adults under 75 report experiencing daily interpersonal stress, which has a substantial impact on health behaviors and mental and physical health. For example, caring for an older adult parent is associated with increased tobacco use, increased alcohol consumption, and sedentary behavior, all of which are risk factors for CAD. A meta-analysis found that subjects who have greater feelings of loneliness and social isolation have a greater risk of new-onset CAD. The INTERHEART study found that, compared to healthy controls, those suffering a first MI self-reported greater stress at the workplace in the prior year. Unemployment and involuntary job loss also have been associated with a nearly twofold increase in rates of both stroke and MI, along with HF. The substantial impact of stress on health is also likely due to major effects on health behaviors, including weight gain, decreased physical activity, decreased sleep time, eating fewer healthy foods, and utilizing increased negative coping behaviors such as alcohol or tobacco use. However, studies of those with CAD found that, even after adjustment for health behaviors, low social support was associated with increased mortality, suggesting that health behaviors may not explain this relationship entirely.

Psychosomatic cardiology research in the 1970s and 1980s was dominated by interest in the relationship between coronary heart disease and a behavior pattern characterized by easily aroused anger, impatience, aggression, competitive striving, and time urgency (the “type A personality”). In the setting of evolving research, hostility is associated with several physiologic processes such as increased circulating catecholamines, increased coronary calcification, and increased lipid levels during an interpersonal conflict. Such processes could, in theory, contribute to CAD when such emotional states are frequent and chronic; however, data remains mixed. More recent attention to traits of negative affectivity (tendency to more commonly experience negative emotions) and social inhibition (the tendency to suppress the expression of emotion) often when combined referred to type D personality. This personality type has been associated with an increased rate of depressive symptoms, lack of positive emotions, and a substantially increased risk of cardiac events in patients with CAD. This personality risk has been more steadily replicated in the literature regarding cardiac risk and has been connected to perceived stress and low self-rated QOL.

Impact of Severe Mental Illness

Persons with severe mental illness (SMI) such as bipolar disorder (BPAD), schizophrenia, and schizoaffective disorder (SCAD) on average have a life expectancy that is 10 to 17.5 years shorter than the general population. The risk of mortality is doubled due to CAD. Even when accounting for the higher rate of death by suicide and accidents, CAD remains the leading cause of death in those with SMI accounting for up to one-quarter of deaths. CAD occurs earlier in people with SMI as compared to the general population and is thought to be partly related to the use of antipsychotics. Atypical antipsychotics impact glucose processing and are associated with rapid weight gain that can lead to metabolic syndrome (HTN, HLD, increased abdominal fat, insulin resistance, and hyperglycemia). Though significant limitations are present in obtaining comparative data between different antipsychotics and risk, research has suggested clozapine (Clozaril), and olanzapine (Zyprexa) carry the highest cardiac–metabolic risk while aripiprazole, ziprasidone, and lurasidone carry the lowest. Mood stabilizers such as valproic acid and lithium are also associated with weight gain.

Those with SMI also have higher rates of sedentary behaviors, smoking, medication side effects, and problems with access to medical care. Smoking is highly prevalent in individuals with SMI, which leads to vascular endothelial damage and systemic inflammation. Smoking cessation can improve psychiatric symptoms, QOL, and if quitting occurs by age 50, cut mortality in half. Smoking cessation can also have an impact on antipsychotic levels. Tobacco smoke increases cytochrome

P450 activity which alternate metabolism of antipsychotics, and levels may increase with cessation. Negative symptoms in patients with schizophrenia are associated with physical inactivity that itself raises CAD and is related to poor diet and obesity. Many studies have demonstrated that the use of metformin with atypical agents such as clozapine or olanzapine reduces medication-related weight gain; however, the long-term impact on CAD outcomes is not clear. For patients with BPAD and SCAD, hypomanic and manic episodes are associated with the increased risk of substance use, which in turn increases CAD risk.

The risk of CAD in SMI remains complex, and many physiologic features suggest that SMI and CAD function in parallel. Both are associated with disruptions in the hypothalamic–pituitary–adrenal axis, inflammation, and immune activation. Lipid metabolism in individuals with BPAD is an area of ongoing research into a potential link between the pathophysiology of BPAD and CAD. Though theories exist, these pathways continue to remain elusive, and further investigation is required regarding the physiologic link between CAD and SMI.

It has been noted that patients with SMI who experience an acute MI are less likely than the general population to receive appropriate treatment, including both medications and revascularization procedures. Those with serious mental illness are at 85% higher risk of death related to complications of CAD as compared to a regionally matched general population. Those with SMI are less likely to have a primary care physician and more likely to have difficulty maintaining continuous insurance coverage. Among patients with schizophrenia entering the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study, 88% of patients with HLD, 62% of patients with HTN, and 30% of patients with DM were not being treated with medications for these conditions. Baseline assessments for these patients should include a thorough medical history, thyroid testing, fasting plasma glucose, fasting blood lipid panel, baseline ECG, and resting blood pressure.

MYOCARDIAL DISEASE

Heart Failure

Brief Review of Symptoms Treatment and Epidemiology. HF (formally referred to as congestive HF) is a result of many forms of heart disease and represents the most common final pathway toward disability and mortality. More than 6.2 million adults in the United States have HF, and HF was listed on 13.4% of death certificates in 2018. Half a million people develop HF per year in the United States. The leading risk factors for HF include CAD, DM, HTN, MI, and mitral and aortic valve disease.

Shortness of breath, reduced exercise tolerance and peripheral edema are typical symptoms of HF. Downstream effects of hepatic congestion and renal dysfunction lead patients to struggle with oscillation between the extremes of being “too wet” or “too dry,” often resulting in complex and, at times, varying medication regimens based on symptoms. ECG can demonstrate a shift in axis and conduction changes. Echocardiography or cardiopulmonary exercise stress testing can confirm the clinical diagnosis by formally assessing the heart’s ability to contract. Angiography and positron emission tomography may be used to identify patients who will experience improved cardiac function after a revascularization procedure. B-type natriuretic peptide levels can also be used to mark the severity of physiologic derangement in exacerbation.

HF is a progressive condition with an approximately 50% 5-year mortality. Classification schemes for the staging of HF are shown in Table 27.2–1. To some degree, prognosis depends on the possibility of reversing the underlying cause. Some forms of infectious and inflammatory myocarditis are completely reversible. For most patients, however, the course tends to be one of stable or gradually downhill functional status punctuated by acute episodes of fluid overload or cardiac arrhythmia, precipitating hospitalization.

Treatment of HF patients usually includes angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARB), β-adrenergic blocking agents, spironolactone, and diuretics. Heart transplantation and left ventricular assist devices (LVADs) are options for refractory cases. Intravenous (IV) inotropic agents that increase the force and rate of heart contraction, such as dobutamine or milrinone, are sometimes prescribed for severely ill patients who would otherwise succumb to low-output states. Many trials of agents with positive inotropic effects have failed to establish a survival benefit. Cardiac arrhythmias are common in patients with HF, often leading to the need for further medication management.

Table 27.2–1.
Heart Failure Classification

New York Heart Association	American Heart Association
Functional classification	Prognostic classification
Class I: Asymptomatic stage	Stage A: Presence of risk factors
Class II: Mild symptoms with moderate exertion	Stage B: Structural heart disease without symptoms
Class III: Mild exertion limited by symptoms	Stage C: Structural heart disease with current symptoms or previous symptoms that have responded to therapy
Class IV: Symptoms at rest	Stage D: Symptoms are refractory to conventional treatment and require assist device, transplant, or palliative care

Psychological Factors Affecting HF. Relatively little attention has been paid to the role of psychological factors in HF. Poor self-care related to depression has been highlighted as a factor in poor outcomes in HF, along with fatigue, denial, lack of knowledge, and impaired cognitive function. Clinically, it appears that adjustment disorders, depression, and anxiety are relatively common problems in this population. Disorders of alertness and cognition occur in patients with hypotension, hyponatremia, hepatic failure, and renal failure secondary to HF. However, symptoms of HF may confound the psychiatric presentation. Dyspnea and orthopnea, for example, may resemble panic, and particularly orthopnea may interfere with sleep. Hopelessness, decreased self-esteem, suicidal ideation, and other cognitive aspects of affective illness are not normally present in HF patients and may help make the distinction. Comorbid depression is associated with higher morbidity and mortality.

DEPRESSION. Depression has been recognized as a risk factor in HF incidence and mortality, and the role of psychological factors in outcome after heart transplantation has been explored. Longitudinal studies of those with HTN have shown patients with high depression symptoms have higher rates of HF development versus patients with low depressive symptoms. Adjusting for demographic and medical variables, depression was associated with an almost threefold increased risk of the development of HF. Several studies have demonstrated the substantial prevalence of depression symptoms in patients with HF and higher rates of mortality when depression is present. Management of depression symptoms with antidepressants or cognitive behavioral therapy (CBT) has not clearly demonstrated an improved prognosis or reduction in mortality in those with HF.

Surgical Management of HF

HEART TRANSPLANT. Heart transplantation is available to between 2,000 and 2,500 patients annually in the United States. It provides approximately 75% 5-year survival for patients with severe HF, who would otherwise have a less than 50% 2-year survival. Candidates for heart transplantation typically experience a series of adaptive challenges as they proceed through the process of evaluation, waiting, perioperative management, postoperative recuperation, and long-term adaptation to life with a transplant. These stages of adaptation typically are accompanied by anxiety, depression, elation, and working through grief. Mood disorders are common in transplant recipients.

Transplant Medication Interactions. Psychiatric disorders can arise post heart transplantation and can include mood disorders, anxiety, PTSD, sleep disorders, and at times cognitive dysfunction. In addition, medications used for immunosuppression, such as chronic steroids or calcineurin inhibitors (cyclosporin or tacrolimus), are associated with depression, anxiety, insomnia and can have cognitive effects themselves. When managing psychiatric concerns after transplantation considering potential immunosuppressant toxicities or side effects is an important aspect of evaluation. In addition, pharmacologic interactions need to be considered when choosing treatment options.

Selective serotonin reuptake inhibitors (SSRIs) are the most widely used psychotropic medication in the heart transplant population and generally have little effect on the CYPSA4 system with limited impact on immunosuppressant interactions. In-vitro studies have suggested sertraline is the least likely to have medication–medication interactions followed by fluoxetine (Prozac) and then citalopram (Celexa). However, some studies have suggested sertraline can affect cyclosporine concentration. When starting a new SSRI, it has been recommended that monitoring should increase in regards to calcineurin inhibitors until a steady dose is reached. Nefazodone (Serzone) should be avoided in the transplant population because of its impact on CYP3A4 and the risk of altering immunosuppressant levels. The serotonin–norepinephrine reuptake inhibitor (SNRI) venlafaxine (Effexor) has also been proven to have little effect on the CYP3A4 isoenzyme system and is generally considered to have limited interaction with post-transplant regiments. Other SNRIs have limited data and can be used with caution. Tricyclic antidepressants (TCAs) are generally avoided, given the impact they have on cardiac conduction. St. John's wort, an herbal supplement sometimes used to address mood or anxiety, has been associated with altering immunosuppressant levels, suddenly lowering them, and potentially could increase the risk of rejection.

Other psychotropic medications can also have an impact on a heart transplant. Calcineurin inhibitors can amplify the nephrotoxicity of lithium. The risk of QTc prolongation should be considered in patients that require additional antimicrobial management as part of prophylaxis of high-dose immunosuppression. Beyond medication–medication interactions, other impacts should be considered on patient health. Patients on chronic steroids may have an increased impact of weight gain or metabolic syndrome when given agents such as mirtazapine (Remeron) or atypical antipsychotics.

VENTRICULAR ASSIST DEVICES. The number of patients with end-stage HF utilizing mechanical circulatory support (MCS) has dramatically increased, with over 10,000 people in the United States living with an LVAD for treatment of end-stage HF. Implantable LVADs are the most common type of MCS used in patients with HF. LVADs are used to maintain adequate systemic blood flow in three clinical scenarios: as a bridge to transplantation until a transplant organ becomes available; as a bridge to recovery in patients whose ventricular function is expected to improve; and as destination therapy (DT) in patients requiring life-long MCS due to inability to be listed for cardiac transplant.

ELEMENTS OF PSYCHIATRIC EVALUATION FOR TRANSPLANT OR MCS. In most settings, psychiatric comorbidity is a contraindication for heart transplantation due to the limited supply of organs and concern that patients with psychiatric comorbidity may have difficulty with post-transplant adherence. Though LVADs also require complex care, the selection of recipients, in theory, should be less restricted due to the device availability. In 2018, a collaboration between the International Society for Heart and Lung Transplantation, Academy of Psychosomatic Medicine, American Society of Transplantation, International Consortium of Circulatory Assist Clinicians, and the Society for Transplant Social Workers convened an expert writing group to develop a set of consensus recommendations for the psychosocial evaluation of candidates for both adult cardiothoracic transplant and long-term MCS. Primary evaluation domains include risk factors for poor outcomes after transplantation/implantation; factors related to patients' knowledge, understanding, and capacity to engage in decision making; factors specific to patients' personal, social, and environmental resources, and circumstances; and factors specific to patients under consideration for MCS. The latter domain includes evaluation of understanding of basic device operation and device malfunction; evidence of cognitive limitations or physical limitations that may compromise the capacity to operate the device or appropriately respond to device alarms; and safety of the home environment for device operation, including a reliable source of electricity and accessibility by emergency personnel.

An emerging body of evidence has demonstrated that preoperatively assessed psychosocial variables predict not only psychiatric but also medical outcomes after heart transplantation or MCS. Personality disorders, substance abuse disorders, dimensional measures of coping skills and social support, and clinically assessed compliance history have been linked to the increased risk of poor postoperative compliance, rejection episodes, and increased mortality in heart transplants. Depression and PTSD after heart transplantation are associated with heightened mortality. The largest national study to date of psychiatric comorbidity in continuous-flow LVAD recipients reviewed 15,024 patients enrolled in the Interagency Registry for Mechanically Assisted Circulatory Support (INTERMACS) between 2008 and 2017. The prevalence of psychiatric comorbidities was substantially lower than in the general population. The study also found no increased risk of mortality in patients with a history of drug or alcohol abuse, suggesting that a less restrictive selection process may be warranted for MCS.

Implantation of an LVAD introduces a unique and highly accessible means of suicide, as the device can be actively disconnected or passively allowed to run out of battery power. A 2020 French study of 494 patients who were part of the ASSIST-ICD study found a 2% rate of attempted (*n* = 2) or completed (*n* = 8) death by suicide within the 19-month period following LVAD implantation. Suicide means included LVAD driveline disconnection, driveline section, or drug intoxication. Patients that attempted or completed suicide were significantly more likely to be treated with DT or at a center without an LVAD coordinator. Acute inpatient psychiatric treatment for LVAD patients may prove challenging as most psychiatric units do not have staff who are specially trained in the care and management of a patient with an LVAD. In such cases, close collaboration of the multidisciplinary team (psychiatrist, psychologist, social worker, nursing, LVAD coordinator, cardiologist and surgeon, pharmacist, financial navigator, and primary care provider) is essential. For patients with LVADs, it is important to differentiate suicidality from the desire to stop life-saving treatment for end-stage HF due to impaired QOL. LVAD disconnection may be an appropriate choice once a psychiatric evaluation has ruled out the possibility of treatable psychiatric illness and ensuring there is a consensus of the medical team, patient, and patient’s family.

Cardiomyopathies

Takotsubo Syndrome. Takotsubo syndrome (TTS) is an acute and usually reversible cardiac syndrome, first described in Japan in 1990, characterized by transient, often circumferential, ventricular wall motion abnormalities, which when visualized on echocardiography closely resembles the Japanese octopus trap “Tako-Tsubo.” TTS has been known by a variety of names, including *Takotsubo cardiomyopathy, stress-provoked cardiomyopathy, and broken heart syndrome*. TTS represents 2% of patients presenting to the hospital with acute chest pain. In addition to chest pain, TTS typically presents with ECG changes, significant elevation of serum natriuretic peptide, a small troponin elevation, and normal coronary arteries. There is often partial to full recovery of systolic function at 3- to 6-month follow-up.

TTS is often precipitated by significant emotional stress or physical illness, or trauma. The pathophysiology of TTS is incompletely understood; however, catecholamines appear to have a central role in stunning the ventricular myocardium. A predominance of cases occur in postmenopausal women and though there have been some statements in the literature that this may be due to the coping style of women, a more rigorous hypothesis involves the role of estrogen via several protective mechanisms, including as a sympatholytic and a stabilizer of the hypothalamic–pituitary–adrenal stress response. Furthermore, estrogen modulates the expression of myocardial β-adrenoreceptors, which are thought to play a key role in the development of TTS. Removal of estrogen may increase the susceptibility of postmenopausal women both to the catecholamine storm itself and the downstream deleterious effects on the heart.

Treatment of TTS typically involves the addition of antiadrenergic agents. Emotional or psychiatric triggers should be evaluated and treated. There are no discrete guidelines about psychopharmacologic treatment for comorbid psychiatric illness; however, it should be noted that certain antidepressant classes, including SNRIs, bupropion, TCAs, and monoamine oxidase inhibitors (MAOIs), effectively increase norepinephrine. A few cases of TTS in the setting of SNRIs, atomoxetine, and lithium have been reported. TTS has also been described in the setting of electroconvulsive therapy (ECT). There is a paucity of data to guide psychopharmacologic treatment post-TTS; until there is better-guiding data, it is reasonable to avoid medication classes that increase catecholamine levels.

Hypertrophic Cardiomyopathy. Hypertrophic cardiomyopathy (HCM) is an autosomal dominant cardiac condition characterized by hypertrophy of the left ventricle with or without right ventricular hypertrophy, the systolic anterior motion of the mitral valve, left ventricular outflow tract obstruction, mitral regurgitation, diastolic dysfunction, myocardial ischemia, and fibrosis. HCM is the most common cause of sudden nontraumatic death in adolescent competitive athletes and is associated with a 1% average annual risk of sudden cardiac death. The diagnosis of HCM often follows the sudden death of a close relative, prompting genetic testing and cardiac imaging of related family members. People with HCM have a higher prevalence of mood, anxiety, and substance use disorders than the general population. In patients with HCM, heart-focused anxiety, characterized by repeated complaints of chest pain, palpitations, and other somatic sensations with concurrent fears of having a heart attack or dying, is associated with a shorter length of time since diagnosis and younger age.

Clozapine-Associated Myocarditis, Cardiomyopathy, and Cardiotoxicity. Clozapine has been associated with myocarditis, cardiomyopathy, and subclinical cardiotoxicity, typically in patients without premorbid cardiovascular disease or other medical risk factors. Clozapine-associated myocarditis (CAM) presents with dyspnea, palpitations, tachycardia, low to moderate fever, flu-like illness, nausea, dizziness, and occasional chest pain. The etiology of CAM is unclear but may be due to a type 1, IgE-mediated acute hypersensitivity reaction to clozapine. It is best diagnosed through detection of clinical signs and symptoms, ECG changes (ST-segment depression >1 mm or T-wave inversion in ≥ two contiguous leads other than aVR), abnormal cardiac biomarkers, and results of cardiac imaging (e.g., echocardiography and cardiac magnetic resonance imaging). Endomyocardial biopsy is the gold standard in the diagnosis of CAM, but its use is limited due to the invasive nature of the procedure.

Clozapine-associated cardiomyopathy is a disorder of myocardial contractile dysfunction that may present with dyspnea, tachycardia, palpitations, chest pain, and fatigue, though patients may be asymptomatic. Diagnosis is supported by increased serum BNP and defined by echocardiographic evidence of left ventricular ejection fraction (LVEF) <50% or by a 10% reduction of LVEF from baseline.

Subclinical clozapine-associated cardiotoxicity (SCAC) is typically asymptomatic but may present with paroxysmal atrial fibrillation (AF) and mild left ventricular/biventricular functional impairment. Echocardiography and elevation of NT-proBNP are helpful in the diagnosis of SCAC.

There is no specific requirement for ECG monitoring when initiating clozapine. When possible and where resources allow, a conservative approach involving a baseline ECG and a weekly ECG in addition to laboratory monitoring (troponins, C-reactive protein) and vital signs during the first 4 weeks of treatment may be considered. Alternatively, a symptom-triggered approach to monitoring may be used if concerning signs and symptoms arise. Regardless of the chosen approach, the absence of cardiac monitoring resources should not preclude treatment with clozapine when the cardiac risk is outweighed by the risk of psychiatric morbidity in high-risk psychiatric populations. In general, a clozapine rechallenge is not advised for patients who experience CAM, clozapine-associated cardiomyopathy, or SCAC.

VALVULAR HEART DISEASE

The relationship between valvular heart disease and psychiatric disorders has been a matter of considerable interest over the past several decades.

Aortic Valve Disease

Aortic stenosis (AS) is a narrowing of the aortic valve that can result in chest pain, dyspnea, presyncope or syncope, fatigue, and palpitations. Treatments for AS include medical management and aortic valve replacement (AVR), either surgical AVR (SAVR) or transcatheter aortic valve implantation (TAVI). Primary indications for AVR, per the 2020 American College of Cardiology (ACC)/American Heart Association (AHA) guideline for the management of patients with valvular heart disease, include severe high-gradient AS with symptoms of exertional dyspnea, HF, angina, syncope, or presyncope; symptomatic patients with low-flow, low-gradient severe AS with reduced LVEF; and asymptomatic patients with severe AS and LVEF <50% or who are undergoing cardiac surgery for other indications. The type of valve prosthesis (mechanical vs. bioprosthetic) chosen is based on multiple factors, including patient age, long-term anticoagulation risk (required with mechanical valves), surgical risk, and patient preferences. Replacement by any modality is considered superior to medical treatment.

Current scoring algorithms used to determine the best modality of valve replacement include the European System for Cardiac Operative Risk Evaluation Score (EuroSCORE) II and the Society of Thoracic Surgeons Predictive Risk of Mortality Score (score ≥8% associated with increased mortality). These algorithms do not include assessments of cognition, depression, or other psychiatric illness or frailty, though depression and anxiety are risk factors for adverse effects after cardiac surgery. A recently developed screening tool, SMARTIE (screening for and managing risk factors in TAVI), which uses the Mini-Cog, PHQ-2, and measures of frailty, found that cognitive deficits were an independent predictor of both postoperative delirium and 30-day mortality. Another study found that measures of frailty, specifically the Clinical Frailty Scale, when combined with traditional risk models, improves the predictive value for mortality after TAVI. A recent analysis from the FRAILTY-AVR study, the first to examine the interplay of depression, frailty, and mortality in patients undergoing TAVI and SAVR, found that depression is often underdiagnosed in this population unless specifically screened. Postprocedural changes in depression modulated the effect of depression on mortality such that patients with preprocedural depression that persisted were at greater risk than patients with preprocedural depression that resolved and in patients with new-onset depression postprocedure.

Infective Endocarditis

Infective endocarditis (IE) is an infection of the heart valves and/or endocardium (lining of the heart) caused by bacteria entering the bloodstream from a variety of sources, including dental procedures, cardiac implantable electronic devices (CIEDs), chronic skin disorders, burns, and injection drug use (IDU). The incidence of injection drug use-related infective endocarditis (IDU-IE) in people who inject drugs is 1.5 to 4 cases per 1,000 years of IDU. Surgical intervention (usually valve repair or replacement) is required in 20% to 40% of IDU-IE patients at the time of first-episode IE. More than half of these patients will require repeat surgical intervention due to ongoing IDU. Types of injected drugs vary regionally, but in general, opioids make up most patients with IDU-IE (67%), followed by methamphetamine (26.3%), bath salts (5.3%), and cocaine (16%), with the remainder unknown.

Optimal management of IDU-IE requires collaboration between cardiology, cardiac surgery, psychiatry/addictions with a dual focus on the infection and the substance use disorder. The ACC/AHA guidelines for the management of patients with valvular heart disease recommend the involvement of addiction specialists to guide treatment in patients with IDU-IE, including the assessment of futility when repeat valve surgery is needed in the case of ongoing IDU. Notably, policies requiring abstinence as a requirement for surgery, including the use of signed contracts, are ineffective, especially during the critical period of acute IE when the management of substance withdrawal may not be prioritized, and patients may present as disengaged in the treatment planning process. For patients with opioid use disorder, medications including buprenorphine, methadone, or naltrexone are a critical element of in-hospital treatment and have been shown to reduce mortality, morbidity, hospital length of stay, and use of illicit drugs.

Other Valvular Heart Disease and Psychiatric Disorder Relationships

Mitral Valve Prolapse. The association between mitral valve prolapse (MVP) and a patient's psychological–emotional state, specifically depression, anxiety, panic, and PTSD, has been disputed for decades. Though most patients with MVP are asymptomatic, a small subset experience dyspnea, palpitations, and fatigue. Due to the overlap of such symptoms with some anxiety disorders, a causal link between MVP and anxiety disorders has been suggested, even in the absence of reliable evidence. Several recent studies have disputed these prior claims, suggesting that MVP is not a determinant of a patient's psychoemotional status. Rather it is the patient's perception of the severity of illness that shapes psychiatric symptoms and QOL. It is essential for physicians to present a reassuring stance when discussing MVP with patients, especially in the accurate description of MVP prognosis and treatment.

Obsessive-Compulsive Disorder and Rheumatic Heart Disease. Acute rheumatic fever and its sequela, rheumatic heart disease (RHD), is a preventable inflammatory condition that results from untreated group A β-hemolytic streptococcal (GAS) infection. In 2015, the global incidence and prevalence of RHD, a condition that poses a high burden in low-income countries and among indigenous and other socioeconomically disadvantaged populations, was >300,000 and 33 million, respectively. RHD has been loosely associated with an increased prevalence of obsessive-compulsive spectrum disorders on the whole, including obsessive-compulsive disorder (OCD), subclinical OCD, tic disorders (including Tourette syndrome), trichotillomania, and body dysmorphic disorder. Pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections (PANDAS) are characterized by the acute onset of OCD, verbal or motor tics, or cognitive disorders in children with a history of GAS infection. The antineuronal antibodies related to GAS infection cross-react with heart valves through molecular mimicry and are identical in both PANDAS and Sydenham chorea. It has been suggested that a relationship between RHD and PANDAS may be mediated by this mechanism. Cardiologic screening may be considered in patients with PANDAS; however, more studies are needed to improve knowledge of this relationship.

ELECTRICAL CONDUCTION DISEASE

Electrophysiology and ECG Interpretation

There are many texts, papers, and guidebooks available regarding in-depth ECG interpretation. In general, most approaches support following a systematic approach of interpretation that includes assessment of the major components of rate, rhythm, intervals, and axis in varying orders. Interpretation of an ECG should begin with an assessment of the patient and the clinical context. The basics of ECG interpretation are outlined below, with the focus of first gathering basic data from the ECG to then guide the interpretation of the rhythm. This review is certainly not comprehensive and is not meant to teach proficiency in atypical rhythm diagnosis. However, we have included a brief reference and refresher on the major components of ECG interpretation.

Terminology and ECG Basics

Rate. The heart rate (HR) can be calculated by a variety of methods. ECGs are composed of boxes of two different sizes. On the x-axis, the boxes represent time. On the y-axis, the boxes represent amplitude or voltage. When using boxes to denote time, the small boxes are equal to 0.04 seconds, and the large boxes represent 0.2 seconds. A standard ECG measures 10 seconds of the heartbeat. As such, there are 5 small boxes per large box and 50 large boxes per 10 seconds of ECG. Since common convention dictates that HR be measured in beats per minute (bpm), 300 large boxes represent 1 minute on the ECG. The “rate by squares” method for determining HR on ECG consists of counting the number of large boxes between R waves and dividing 300 by this number. Another option to calculate HR from ECG interpretation is the “QRS complexes per ECG” method which is often useful in irregular rhythms. This approach consists of counting the amount of QRS complexes on the standard ECG and multiplying by six because the run time of a standard ECG is 10 seconds.

Bradycardia, defined by an HR of fewer than 60 bpm, may occur in normal subjects, especially trained athletes, but also occurs pathologically with medications, hypothyroidism, cardiac ischemia, degenerative conduction disease, excess vagal tone, and during a variety of electrolyte disturbances. Idiopathic primary sinoatrial node degeneration and pharmacologically induced dysfunction are the most common causes of bradycardia. Symptomatic patients typically present with dizziness or syncope. SSRIs have been associated with minor degrees of bradycardia that can reduce cardiac workload and benefit the diseased heart. Lithium is also associated with bradycardia and has been connected to sinus node dysfunction and prolonged sinus node recovery time. HR in patients on lithium has been shown to be lower at rest as well as with exercise though the specific effect of lithium on HR remains unclear as most studies have not controlled for lithium-induced hypothyroidism or patients being treated with other medications which affect HR. Lithium has also been associated with sinus block and sinus arrest. Atrioventricular (AV) node block has also been associated with lithium therapy, but again studies have not clearly controlled for other causes, including cardiotoxic medication.

Tachycardia is defined by an HR of greater than 100 bpm. Tachycardia itself is often asymptomatic unless marked and can be associated with palpitations, weakness, fatigue, dizziness, and near syncope. Psychotropic medication with antimuscarinic and anticholinergic effects (i.e., all atypical antipsychotics, TCAs, MAOIs, benztropine, diphenhydramine) can cause sinus tachycardia which, in healthy patients without cardiac disease, usually resolves with time. Up to 90% of patients taking clozapine have sinus tachycardia.

Axis. The axis of an ECG represents the net vector of depolarization of the heart. It is useful in diagnosing enlargement, rotation of the heart, infarction, and certain arrhythmias. The normal axis of the heart lies between -30 degrees and 90 degrees. Axis can be determined by looking at two leads specifically: Lead 1 (0 degrees) and aVF (90 degrees). These two leads are chosen because they represent perpendicular vantage points from which the net electrical depolarization of the heart can be deciphered. Axis deviation is most commonly a result of ventricular hypertrophy shifting the position of the heart in the chest. Left axis deviation is more common as it can be seen in patients who have had sustained HTN leading to left ventricle hypertrophy. Right axis deviation is relatively rare as it is typically the result of congenital cardiac abnormalities or right ventricle hypertrophy seen in pulmonary HTN or pulmonary thromboembolism (PTE).

Segments and Intervals. The intervals on an ECG are used to assess the amount of time required for a certain portion of the cardiac cycle to complete and thus are a surrogate for the efficiency of conduction through the electrical system of the heart. Common convention dictates that flat areas of the ECG without a wave in them are known as a “segment,” and those with a wave in them are called “intervals.” An ECG tracing is composed of a P wave which is then followed by the QRS complex, and lastly, a T wave. The different waves are created by electrical activity in the areas of the heart (P = atrial depolarization, QRS = ventricular depolarization, and T = ventricular repolarization). Intervals can be pathologically long or short, and changes in their duration can be clues to localize the lesion causing the conduction abnormality.

The ST segment and T wave correspond to ventricular repolarization within the cardiac cycle. As opposed to depolarization which occurs passively by triggering voltage-gated ion channels, repolarization is an active process that requires adenosine triphosphate. Thus, it is especially susceptible to ischemia and hypoxia. ST elevations are most associated with transmural infarction; however, careful attention to the clinical scenario is critical. Entities associated with ST elevations include STEMI, myocarditis and pericarditis, stress cardiomyopathy, left ventricular aneurysm, hyperkalemia, and early repolarization, among others. T-wave inversions are another electrocardiographic finding which can instill fear due to their association with ischemia. It is also important to note that T-wave changes and ST depressions cannot localize to a vascular territory. The differential diagnosis for T-wave inversions is broad and ranges from benign to medical emergency. This includes ischemia, electrolytic derangements, myocarditis and pericarditis, PTE, subarachnoid hemorrhage, and effect from antipsychotic or lithium toxicity. Other T-wave abnormalities such as flattening can be seen with normal antipsychotic doses and electrolyte changes such as hypokalemia and hypercalcemia. Peaked T waves are seen in electrolyte disturbances such as hyperkalemia; however, hyperacute T-wave peaking is the first electrocardiographic sign of ischemia.

Widening of the QRS complex artificially prolongs the QT interval without prolonging repolarization. There is no standardized method to correct the QT interval for a wide QRS. However, the consensus of the AHA, ACC, and Heart Rhythm Society (HRS), recommends adjustment for both QRS duration as well as HR or the use of an alternative interval, such as the JT. Psychiatrists who frequently treat patients with cardiac disease should be familiar with methods to correct the QT interval for a wide QRS. When performing a risk assessment for torsades de pointes (TdP), the presence of ventricular conduction delay (VCD) or VP should be considered evidence of structural or functional heart disease and trigger a correction of the QT-interval duration for the wide QRS.

Rhythms. Identifying rhythms can be an important tool to understanding the electrical pathway of the heart. Rhythms can be considered narrow or wide and regular or irregular. Narrow rhythms are identified by having a narrow QRS and are, by definition, supraventricular. They are narrow because they follow the normal conductive system of the heart. A wide QRS signifies either a ventricular beat or conduction delay, which is sometimes referred to as “aberrancy” or a bundle branch block (BBB). A QRS is wide when the complex is greater than 100 ms. When the HR is fast, and the QRS is narrow, this is referred to as narrow complex tachycardia. When the HR is fast, and the QRS is wide, this is referred to as a wide complex tachycardia. Table 27.2–2 lists the different narrow and wide complex tachycardias.

A normal sinus rhythm will have a P wave preceding every QRS and a QRS coming after every P wave. A total absence of P waves indicates a lack of an organized electrical activity in the atria (such as that seen in AF) or a rhythm generated from below the AV node (e.g., “junctional” rhythm, which happens when the pacemaker cells proximal to the AV node fail to fire). If there are more P waves than there are QRS complexes, an AV block exists.

Table 27.2–2.
Types of Rhythms with Tachycardia

Regular, narrow complex tachycardias	Sinus tachycardia Atrial flutter Atrioventricular nodal reentrant tachycardia (AVNRT) Atrioventricular reciprocating tachycardia (AVRT)
Irregular narrow complex tachycardias	Atrial fibrillation Multifocal atrial tachycardia (MAT)
Wide complex, regular tachycardias	Monomorphic VT SVT with aberrancy
Irregular, wide complex tachycardias	Polymorphic VT (of which torsades de pointes is an example) Pre-excited AF (AKA AF with aberrancy referring to Atrial fibrillation with an accessory pathway)

AF, atrial fibrillation; SVT, supraventricular tachycardia; VT, ventricular tachycardia.

Ventricular Arrhythmia and Sudden Cardiac Death

Malignant ventricular arrhythmias may occur as an electrical accident in individuals with no structural heart disease but more often occur in those with underlying heart disease. Because conduction of myocardial depolarization may be delayed in regions of myocardial tissue damage, regions distal to the affected area may be stimulated to depolarization while adjacent tissue is undergoing repolarization, setting off a premature depolarization of the adjacent tissue that may precipitate ventricular fibrillation.

Lethal ventricular tachyarrhythmias are more likely to occur in the setting of structural heart disease (e.g., after an MI) but also are related to genetic disorders regulating ion channels that control ventricular depolarization and repolarization, such as congenital long QT syndrome and Brugada syndrome. These latter disorders are sometimes clinically silent until “unmasked” by medications including lithium, TCAs, and phenothiazines.

Because autonomic cardiac modulation is profoundly sensitive to acute emotional stress, such as intense anger, fear, or sadness, it is not surprising that acute emotions can stimulate arrhythmias. Instances of sudden cardiac death related to sudden emotional distress have been noted throughout history in all cultures. In community samples, depression, phobic anxiety, and mixed anxiety symptoms increase the risk of sudden cardiac death. After the destruction of the World Trade Center in New York in 2001, patients with implanted defibrillators had sharply increased frequency of ventricular tachycardia (VT) events. In post-MI patients with ventricular ectopy, comorbid depression is associated with significantly increased rates of VT–ventricular fibrillation. The effect of psychotropic drugs on the QT interval and arrhythmia risk is discussed below.

Atrial Fibrillation

AF is the most common cardiac arrhythmia in clinical cardiology practice; it is an irregularly irregular heart rhythm characterized by disorganized depolarization of atrial myocytes. The result is ineffective atrial contraction and rapid HR. The ventricular response can follow, becoming extremely rapid, with palpitations and symptoms of reduced cardiac output. The most common risk factors are older age, male gender, HTN, and valve disease. Cardiac surgery, hyperthyroidism, renal disease, DM, sleep apnea, smoking, alcohol consumption, and psychological stress are also associated with the development of AF. AF is associated with increased morbidity and mortality primarily due to CAD, HF, and stroke. Independent of these factors, it is associated with cognitive impairment and psychological distress. Treatment of AF should be focused on mitigating stroke risk as well as rate or rhythm control using β-blockers, calcium channel blockers, cardioversion, or ablation. Anticoagulation reduces the risk of embolic events but increases the risk of bleeding.

AF is linked with depression, anxiety, and mood disorders, as well as poor QOL. A year-long prospective study of 95 patients with AF monitored emotional states via daily electronic diaries found negative affectivity increased the likelihood of an AF episode two- to fivefold, whereas happiness decreased the likelihood. Successfully treating AF has been shown to improve QOL and reduce depression symptoms. Patients with successful ablation had improvement on major depression inventory scores compared to patients with unsuccessful ablations at long-term follow-up. Using antidepressants in patients with AF to control symptoms of depression and anxiety in an effort to prevent AF has not been well studied. However, some studies have shown other methods of management of anxiety and depression, such as yoga, decrease both asymptomatic and symptomatic AF episodes and improve scores of anxiety, depression, and perceived QOL. Management of depression and anxiety should take into consideration the drug interactions between psychotropic and cardiovascular medications. Notably, antidepressants may impact bleeding risk in those who are anticoagulated, as addressed elsewhere.

Depolarization and Repolarization Abnormalities

Repolarization Abnormalities

TORSADES DE POINTES AND QTC. Over the past decade, considerable attention has been paid to the risk of QT prolongation with certain psychiatric medications. QT prolongation is considered a marker of risk for TdP, a lethal polymorphic VT that can present with sudden cardiac death. Medications that prolong the QT interval via effects at the delayed potassium rectifier current (Ik_r) or its encoding gene (hERG) may increase the risk for TdP. Nonmodifiable risk factors for TdP include female sex (between puberty and menopause), increasing age, congenital or acquired long QT syndrome (caused by genetic abnormalities), personal history of structural or functional cardiac disease, or a family history of sudden cardiac death. Besides certain medications, other modifiable risk factors include acute medical illness (particularly cardiac disease), hepatic or renal impairment, hypokalemia, hypomagnesemia, hypocalcemia, pharmacokinetic drug–drug interactions, and bradycardia.

Approach to QTc with Psychotropic Drugs. The QT interval represents the duration of ventricular depolarization and repolarization. Because the QT decreases as HR increases, it must be corrected for HR (QTc). The most common formula used to derive the QTc is the Bazett formula ($QTc = QT/\sqrt{RR}$), a square-root formula used as the default formula in most ECG machines. While the Bazett formula accurately estimates the QTc when the HR approximates 60 bpm, it tends to undercorrect at lower HRs and overcorrect at higher HRs. The AHA recommends using a linear correction formula, such as the Framingham

($QTc = QT + 0.154[1000 - RR]$) or Hodges ($QTc = QT + 1.75[HR - 60]$) correction formulas. The QTc interval is normally less than 450 ms in women and 430 ms in men. Prolongation of the QTc interval is associated with progressively increasing risk of TdP. The risk of ventricular arrhythmias increases by 1.052^x, where x is every 10 ms above a QTc of 400 ms. The QTc interval is an imperfect index of risk of TdP and sudden cardiac death: some drugs that prolong the QTc interval have never been associated with cases of TdP.

ANTIPSYCHOTICS. As a class, most antipsychotic agents prolong the QTc interval and may therefore increase the risk for TdP. Among first-generation antipsychotics, low-potency phenothiazines such as chlorpromazine (Thorazine) and thioridazine (Mellaril) cause the greatest degree of QTc prolongation. Although haloperidol, particularly the IV formulation, has historically been associated with QTc prolongation, recent evidence suggests that the risk has been exaggerated, and there is actually no evidence suggesting that any form of haloperidol causes greater QTc prolongation than most other antipsychotics. Ten of 11 prospective studies involving >1,500 patients and doses up to 20 mg of IV haloperidol found no evidence of QTc prolongation.

Among second-generation antipsychotics, aripiprazole (Abilify) and lurasidone (Latuda) are the only agents that have not been associated with QTc prolongation, with some evidence suggesting that aripiprazole may even shorten the QTc interval slightly. Ziprasidone (Geodon) is associated with the greatest QTc prolongation of the second-generation antipsychotics, with iloperidone (Fanapt) a close second, though TdP is rare with either. Other atypical antipsychotics, including olanzapine, risperidone (Risperdal), and clozapine, prolong the QTc to some degree but are not thought to convey substantial risk. Notably, quetiapine (Seroquel) does have an additional warning from the U.S. Food and Drug Administration (FDA) regarding the potential for QTc prolongation. It is generally not possible to risk stratify most antipsychotic agents, aside from concluding that thioridazine, ziprasidone, and iloperidone carry increased risk, while aripiprazole and lurasidone carry lower risk. Other agents cannot be separated out from one another.

For patients who do not have heart disease, none of the second-generation antipsychotics has a well-established risk of TdP when taken at recommended doses. If feasible, ECG monitoring should be considered for any patients prescribed thioridazine, ziprasidone, or iloperidone, or for patients with multiple other risk factors for QTc prolongation and TdP being prescribed any antipsychotic agent. For patients in the hospital prescribed IV haloperidol, daily ECG monitoring may only be needed for cumulative doses >25 mg or in patients with multiple risk factors.

An increase of 30 ms should trigger increased monitoring, and an increase of 60 ms is a cause for concern. A particular time patients are at elevated risk for TdP is during a restraint. During this time, there are increased circulating catecholamines increase cardiac oxygen consumption and lower K^+ uptake into the muscle (through adrenergic stimulation), leading to low serum K^+ levels. This metabolic risk, combined with the inherent antipsychotic risk, leads to increased myocardial instability and a high risk for arrhythmia. Other risk factors for drug-induced TdP include older age, female sex, electrolyte abnormalities, renal or hepatic dysfunction, pre-existing heart disease (e.g., infarction), bradycardia, and treatment with more than one QT-prolonging agent as well as genetic predisposition. Table 27.2–3 outlines other substances that have been shown to prolong the QT interval.

ANTIDEPRESSANTS

Serotonin Modifiers. For many years, SSRI antidepressants were largely considered safe with regard to QTc prolongation, though case reports existed for all agents except for paroxetine. In 2011, the FDA issued a warning regarding the risk of QTc prolongation with citalopram, stating that it should not be used at doses greater than 40 mg. This warning was based on a thorough QT study, which demonstrated QTc prolongation of 18.5 ms at 60 mg. Subsequent studies have suggested that citalopram does separate out from other SSRIs and typically causes a dose-dependent QTc prolongation of 10 to 20 ms, though the risk of TdP appears minimal. In the setting of the FDA warning, a large percentage of patients had their doses of citalopram reflexively reduced, leading to increased psychiatric morbidity without impact on cardiac mortality.

Table 27.2–3.
Substances that Prolong the QT Interval

- **Alcohol**
- Amiodarone
- Azithromycin
- Chloroquine
- Cisapride (withdrawn due to TdP)
- Citalopram
- Clarithromycin
- Cocaine
- Disopyramide
- Dofetilide
- Erythromycin
- Fluconazole
- Furosemide
- **Halofantrine**
- Ketoconazole
- Levofloxacin
- Methadone
- Moxifloxacin
- Pentamidine
- Procainamide
- Quinidine
- Quinolones
- Sotalol
- Tamoxifen
- Terfenadine (withdrawn due to TdP)
- Vandetanib

TdP, torsades de pointes.

Escitalopram (Lexapro) has also been shown to separate out from other SSRIs in a few studies, though the associated QTc prolongation is minimal, and no warnings exist for escitalopram. Other SSRIs are generally considered safe, with sertraline having been studied the most in cardiac populations.

Among other antidepressants, venlafaxine has been associated with QTc prolongation in some studies. Mirtazapine also has mixed evidence, with some studies suggesting no risk of QTc prolongation even in overdose, but one study suggesting some association with an increased risk of arrhythmias. Bupropion (Wellbutrin) was associated with QTc prolongation in overdose in one study, but this was likely driven by tachycardia and the use of the Bazett formula to derive the QTc. Other antidepressants, including duloxetine (Cymbalta), have not been associated with QTc prolongation.

Trazodone (Desyrel) exerts a modest dose-dependent prolongation of the QT interval, which is likely clinically insignificant, but warrants caution in cases of overdose. Neither benzodiazepines nor buspirone (Buspar) has an appreciable effect on the QT interval. Multiple cases of QTc prolongation and TdP have been reported with donepezil (Aricept). However, the association is unclear, considering all reports have been in patients older than 80 years with multiple other risk factors for QTc prolongation. Amantadine (Gocovri) has been found to trigger TdP in the setting of overdose and co-ingestion with other QTc prolonging agents.

Based on the available evidence, there is no indication for ECG monitoring in most patients prescribed antidepressants. For patients with significant risk factors, a baseline and follow-up ECG is recommended when feasible. For patients being started on citalopram, even without other risk factors, this may be a reasonable approach as well. It is important to not reflexively reduce a patient's citalopram dose, however, and compelling arguments can be made for the use of citalopram at doses greater than 60 mg in select patients and with judicious monitoring.

Tricyclic Antidepressants. As opposed to SSRI and other antidepressants, which can cause QTc prolongation via effects on repolarization, TCAs have long been known to affect depolarization, leading to QRS widening. The effect is dose dependent but can happen at both therapeutic and toxic doses. The change is unlikely to be clinically relevant in a normal heart. However, complete heart block can occur in the presence of pre-existing BBB, second and higher degrees of AV block, or with toxic levels (i.e., overdose). Ventricular pacemakers may also be suppressed by TCA, causing ventricular asystole. They should therefore be avoided in all cases of conduction disease beyond a first-degree block.

OTHER PSYCHIATRIC MEDICATIONS

Mood Stabilizers. Antiepileptic medications, including valproate (Depakote), lamotrigine (Lamictal), carbamazepine (Tegretol), oxcarbazepine (Trileptal), topiramate (Topamax), and gabapentin (Neurontin), have not been found to prolong the QT interval. Valproate may have some protective effects via stabilization of cardiac conduction. In 2021 the FDA issued a drug safety communication that lamotrigine carries the potential for increased risk of arrhythmia in people with clinically important structural or functional heart disorders. The mechanism of risk was not defined except to warn that risk may be further increased when used in combination with other sodium channel blockers, including carbamazepine and oxcarbazepine (about which the FDA stated should not be considered safer than lamotrigine until more data is available). Notably, sodium channel blockade affects the ventricular *depolarization* phase of cardiac conduction, not the repolarization phase represented by the QT interval.

The effect of lithium on QTc prolongation is uncertain, as only a few small studies have examined the relationship. Serum lithium concentrations within the therapeutic range of 0.6 to 1.2 mEq/L do not appear to prolong the QT interval. There does appear to be a positive correlation between increased serum lithium concentrations and QTc prolongation; however, QTc increases are modest and unlikely to be clinically significant.

Stimulant and Stimulant-Like Medications. No statistically or clinically significant changes in the QT interval have been found in clinical trials of short- and long-term treatment with methylphenidate (Ritalin) and amphetamine drugs. Atomoxetine (Strattera) has been linked to QTc prolongation by a few case studies. However, these were complicated by other variables, including the concurrent use of other QTc prolonging medications. Furthermore, a pooled analysis of five randomized, double-blind trials showed no clinically significant QTc prolongation with atomoxetine. Atomoxetine has been shown to inhibit the IK_r channel in vitro in a dose-dependent fashion, which may have implications for QTc prolongation in cases of overdose. Modafinil has no clinically significant effect on the QT interval. Clonidine and guanfacine have not been associated with statistically or clinically significant increases in the QT interval.

Methadone and Buprenorphine. Methadone has a strong association with QTc prolongation and TdP. Cardiac monitoring recommendations for methadone have been controversial, given competing priorities to sustain a high degree of vigilance through ECG monitoring while maintaining a harm reduction approach that aims not to restrict access to life-saving treatment for opioid use disorder in settings where ECG monitoring is not available. Methadone safety guidelines proposed by the American Pain Society and HRS in 2013 compromised on the recommendation for baseline ECG monitoring, suggesting that practitioners “consider obtaining an ECG prior to methadone initiation in patients known to be at higher risk” for QTc prolongation and TdP. For patients with known risk factors for QTc prolongation, a prior QTc >450 ms, or a history suggestive of prior ventricular arrhythmia, an ECG should be obtained prior to methadone initiation. Alternatives to methadone, such as buprenorphine (Suboxone), should be considered in patients with a baseline QTc between 450 and 500 ms. It is recommended that methadone not be used in patients with a QTc >500 ms. All identified modifiable risk factors should be evaluated and corrected, if possible, prior to methadone initiation. Buprenorphine is a safer alternative to methadone with no known association with QTc prolongation or TdP. Buprenorphine is known to block hERG Ik_r channels; however, it is a hundred times less potent of an inhibitor than methadone.

Impact with Cardiac Medications. The prescribed prevalence of nonpsychotropic medications associated with QTc prolongation or TdP is high. One in five intensive care unit patients is coprescribed two or more QTc prolonging medications. The top nine medications prescribed by the emergency department have effects on the QT interval. In the outpatient setting, a quarter of all prescriptions are associated with QTc prolongation. Commonly used medications that prolong the QT interval include antiarrhythmics, antiemetics, antibiotics (macrolides, fluoroquinolones), antifungals, antivirals, and anticancer medications.

Depolarization Abnormalities

VENTRICULAR CONDUCTION DELAY. VCD refers to impaired electrical conduction through the ventricles due to slowed, cell-to-cell depolarization of the myocardium rather than rapid conduction through the His–Purkinje system. VCD manifests on the ECG as widening of the QRS complex and includes interventricular conduction delay, BBB, and ventricular pacing.

TCAs have long been known to affect depolarization, leading to VCD. At therapeutic doses, these agents only pose a risk for patients with pre-existing cardiac disease, but in overdose, they can lead to arrhythmias even in otherwise healthy individuals. The use of TCAs in the period following acute MI is contraindicated by FDA. Although no guidelines exist for the use of electrocardiographic screening and monitoring when prescribing TCAs, it is standard practice to obtain an ECG before prescribing a TCA to rule out unrecognized VCD or ischemic heart disease, especially in individuals older than 40 years.

Cardiac Implantable Electronic Devices (CIEDs)

CIEDs detect cardiac conduction and, when the need arises, deliver appropriate electrotherapy, whether it be pacing or shocks. The three main types of CIED are permanent pacemakers (PPMs), implanted cardioverter defibrillator (ICD), and cardiac resynchronization therapy (CRT).

Permanent Pacemakers. PPMs are primarily used for the treatment of bradyarrhythmias, with the goal of sustaining an adequate HR for daily activity. Pacemakers are programmed to monitor HR and initiate pacing if certain criteria are met. Typically, PPMs are set to pace when the HR dips below 50 to 60 bpm. With a growing focus on QTc prolongation and psychopharmacology, it is important to understand how PPMs influence the electrophysiology of the heart and how this is translated onto the ECG. With ventricular pacing, instead of the typical rapid conduction from the AV node through the His bundle and bundle branches, the electrical signal originates from the implanted lead and travels from cell to cell through the myocardium. This cell-to-cell transmission is much slower than the intrinsic conduction system, and as a result, the QRS complex is widened. This QRS widening must be addressed for proper QTc interpretation (see ECG interpretation).

Also relevant to PPMs is the concept of overdrive pacing, which can be used to mitigate the risk of TdP in some high-risk patients. There is evidence to support the reduction of early afterdepolarizations, a common trigger of TdP when the heart is paced at rates of 70 bpm and above. In situations where a patient has numerous TdP risk factors, has a CIED, and requires high-risk psychotropic medications, overdrive pacing may be used to bridge the patient through the high-risk period.

Implanted Cardioverter Defibrillators. ICDs provide high-voltage shocks with the goal of terminating VT or ventricular fibrillation (VF), followed by rapid bursts of pacing to keep the patient alive after or in between shocks. Common clinical indications for an ICD include survival from sudden cardiac arrest due to VF, syncope with clinically relevant sustained VT, long QT syndrome with the experience of syncope (presumed TdP), and HF. All ICDs have a backup pacing function, typically set to pace at 30 to 40 bpm.

Idiopathic Lead Migration. Lead migration is a rare complication of CIEDs that occurs when newly implanted leads dislodge and retract into the device pocket. The phenomenon was first described as “twiddler’s syndrome,” a term suggesting purposeful or subconscious manipulation by the patient, observed on chest radiographs by twisting of the leads along the long axis of the pulse generator. Later terms include “reel syndrome” and “ratchet syndrome.” These phenomena have often been attributed to risk factors, including older adult age, compulsive behavior, or psychiatric history. However, a pooled series of 87 case reports demonstrated no correlation of lead migration with manual manipulation, older adult age, obesity, oversized pocket, or psychiatric history. It is important for psychiatrists to be aware of idiopathic lead migration as patients with an unrelated psychiatric diagnosis who experience this phenomenon may be victims of stigma and incorrectly labeled as “device manipulators.”

MASQUERADERS

Cardiac Masqueraders of Psychiatric Illness

Some cardiac conditions may be “masqueraders” of psychiatric illness, either due to prominent symptoms misidentified as anxiety, mood-related, trauma-based, psychosis, or personality disorder, or conversely, the absence of an easily identifiable “medical” cause.

Syncope and Torsades de Pointes Masquerading as Nonepileptic Seizures. The frequency of epilepsy misdiagnosis has been estimated at 20% to 30%. Cardiovascular conditions, most commonly syncope, and nonepileptic seizures are the most common imitators. The reflex syncopes (vasovagal syncope, carotid sinus syncope) may cause blackouts complicated by abnormal movements (myoclonic jerking, convulsions) secondary to generalized cerebral hypoxia. TdP may also present with seizure-like activity and should always be considered in the differential, especially if the patient is prescribed a QTc-prolonging medication.

A comprehensive cardiology workup is suggested in all patients with symptoms suggestive of epilepsy who are not found to have electroencephalogram correlates of typical paroxysmal events. As nonepileptic seizures (neurologic symptom disorder/conversion disorder) are a diagnosis of exclusion, it is imperative for cardiac causes of blackouts, spells, and paroxysms to be ruled out. A study of 1,567 patients receiving formal evaluation at an epilepsy center, including video-EEG, found that of 78 cases not receiving a diagnosis of epilepsy who were referred for additional cardiologic assessment and syncope was formally diagnosed in 50 patients.

Pulmonary Thromboembolism. PTE may present with symptoms of dyspnea, chest pain, cough, anxiety, palpitations/racing heart, fainting, and lightheadedness. Many of these symptoms can also be seen in anxiety disorders and panic attacks, leading to the potential for misdiagnosis of this life-threatening condition. Patients with a premorbid psychiatric diagnosis may be at higher risk for misdiagnosis or delayed diagnosis of PTE due to “diagnostic overshadowing,” a phenomenon that can impair the diagnostician’s ability to look beyond a psychiatric etiology. Risk factors for PTE include malignancy, fractures of the leg or hip, hypercoagulable state or clotting disorder, history of MI or stroke, recent major surgery or travel, obesity, sedentary lifestyle, immobilization, pregnancy, older age, and hormone replacement. PTE should always be considered in the differential of a patient presenting with new-onset anxiety or panic, especially if symptoms are accompanied by hemodynamic changes or if PTE risk factors are present.

Psychiatric Masqueraders of Cardiac Illness

Atypical Chest Pain and Functional Cardiac Symptoms. Somatic symptom disorder, panic disorder, anxiety, PTSD, and depression can all present with somatic complaints and represent a substantial issue in ambulatory and emergency cardiology practice. In studies of patients presenting with the chief complaint of palpitations, these diagnoses account for approximately 30% of cases. In this population, the psychiatric disorder is associated with more frequent recurrent symptoms, emergency room visits, hypochondriacal concerns, and impairment in activities of daily living. Similarly, chest pain is a common presentation for panic disorder. In a study of 1,999 consecutive patients presenting with chest pain, a panic disorder was diagnosed in 38%. Patients without CAD were twice as likely to receive a panic disorder diagnosis. Panic patients were younger, more likely to be female, and more likely to be unemployed. In another study of individuals presenting to the emergency room with noncardiac chest pain (NCCP), variables associated with panic disorder included the absence of CAD, female sex, atypical quality of chest pain, younger age, and high level of self-reported anxiety. Recommendations for patients presenting with NCCP and found to have psychiatric morbidity include CBT, mindfulness-based stress reduction, hypnosis, breathing training and guided physical activity, pharmacotherapy, and care management approaches.

PSYCHIATRIC SYMPTOM MANAGEMENT OF PATIENTS WITH CARDIAC DISEASE

Depression

The mainstay of pharmacotherapy in patients with depression and cardiac disease is the SSRI antidepressants. Sertraline, in particular, has been the most studied agent in cardiac populations and is often considered the first-line agent. In patients with ACS, the sertraline antidepressant heart attack randomized trial (SADHART) demonstrated that sertraline was a safe and effective treatment for major depression, indistinguishable from placebo with respect to hemodynamic, ECG, and inotropic effects. Sertraline was associated with a higher rate of response to treatment, better QOL, and a trend toward fewer adverse cardiac events. The antidepressant effect was most pronounced in patients with more severe depression, with prior episodes of depression before the index cardiac event, and with the onset of the current episode before the index event. Several other small trials have tended to show a beneficial effect of sertraline on depression after MI. In contrast, however, a randomized placebo-controlled trial of sertraline for depression in HF patients did not show benefit with respect to either cardiovascular or mood outcomes.

The CREATE trial showed that citalopram is effective when combined with clinical management for the treatment of major depression in patients with CAD. However, given emerging concerns with citalopram and QTc prolongation, citalopram is no longer recommended as a first-line agent in patients with cardiac disease. See previous section on repolarization abnormalities for detailed discussion and consideration of QTc impact.

SSRIs inhibit platelet storage of serotonin and may prolong bleeding in patients via an effect on platelets. They do not have a systematic effect on coagulation, as measured by the prothrombin time test, even in patients treated with warfarin (Coumadin). Bleeding risk is increased in patients treated with concomitant SSRIs, antiplatelet agents, and warfarin, as well as those treated with SSRIs and nonsteroidal anti-inflammatory drugs. The use of acid-reducing medications may mitigate the risk of gastrointestinal bleeding in patients on long-term SSRIs. There is about a 1.5- to 2-fold increased risk of hospitalization for bleeding events when SSRIs are combined with antiplatelet agents or warfarin, and an estimated 36 years of treatment are needed to cause one additional hospitalization. The benefit of SSRI therapy likely outweighs this risk in many patients. Additional cardiovascular effects of psychotropic drugs are listed in Table 27.2–4.

Table 27.2–4.
Cardiovascular Effects of Psychotropic Drugs

Agent	Effects
Tricyclic antidepressants	Orthostatic hypotension; delayed cardiac conduction, heart block; type 1A antiarrhythmic effect; ventricular arrhythmias in overdose; tachycardia
Selective serotonin reuptake inhibitors	Bradycardia Citalopram: Dose-related QTc prolongation
Monoamine oxidase inhibitors	Hypotension; peripheral edema
Bupropion (Wellbutrin)	Hypertension
Nefazodone (Serzone), trazodone (Desyrel)	Hypotension
Venlafaxine (Effexor)	Hypertension, especially at doses >300 mg/day
Lithium (Eskalith)	Sinus node dysfunction, including sinus arrest
First-generation antipsychotic agents	QT interval prolongation; torsades de pointes
Second-generation antipsychotic agents—except aripiprazole (Abilify)	QT interval prolongation

TCAs are no longer recommended because of the risk of orthostatic hypotension, cardiac conduction disturbances, and arrhythmias. See previous section Repolarization Abnormalities for details regarding conduction changes in TCAs. Among other antidepressants, mirtazapine has been found to have some benefit in depressed patients with cardiac disease, but metabolic side effects make it a less ideal medication. Experienced psychiatrists often use stimulants as (off-label) primary or adjunctive treatment of medically ill patients with depressed mood and psychomotor retardation. Please see the attention-deficit hyperactivity disorder (ADHD) section below for information about stimulant use.

Few studies have systematically controlled concurrent treatment and cardiac diagnoses in assessing effects of ECT in cardiac patients, but available information suggests that patients with cardiac disease are more likely than comparison patients without heart disease to experience cardiovascular complications during ECT. During ECT, initial vagally mediated bradycardia is followed by substantial increases in HR and blood pressure. Patients with ischemic heart disease or ventricular arrhythmias are at increased risk of MI or ventricular tachyarrhythmias. Most complications are transient and do not necessarily preclude the use of this modality. ECT has been used successfully in patients with heart transplants.

Psychotherapeutic interventions, particularly interpersonal therapy (IPT) and CBT, have also been shown to be beneficial for depression in patients with cardiac disease. Notably, whereas pharmacotherapy has not been demonstrated to be helpful for depression in patients with HF, CBT has. Several collaborative care model studies, employing stepped-care interventions and incorporating patient preferences in the choice of initial treatment, have demonstrated good results with respect to patient satisfaction, mood, and QOL outcomes, in patients with depression following admission for CAD or HF, in patients with recent ACS, and in post-CABG patients (Bypassing the Blues Trial). Aerobic exercise has also been demonstrated to significantly reduce depression symptoms in patients with depression and CAD or HF, and rehabilitation and stress management interventions that include exercise have demonstrated beneficial effects on recurrent cardiovascular morbidity and mortality in patients with CAD and HF.

Anxiety

Anxiety is common in patients with cardiac disease, with 13% of HF patients meeting diagnostic criteria for an anxiety disorder and up to 30% endorsing significant anxiety. Following an ACS, 20% to 30% of patients experience elevated levels of anxiety, and 10% to 14% have anxiety levels higher than the average psychiatric inpatient. Similarly, clinically significant anxiety is present in up to 25% of patients awaiting CABG, though in most cases, this anxiety resolves in the 3 months postprocedure. Psychological interventions to reduce anxiety symptoms in patients with cardiovascular disease have been found to be beneficial. In the absence of randomized trials for anxiety that are specific to cardiovascular disorder, it is reasonable to utilize treatments of anxiety established in other clinical settings, including SSRIs. Benzodiazepines may also be particularly effective for short-term management of anxiety in the cardiac patient, without detrimental cardiac effects.

PTSD

Acute coronary events and their treatment can be traumatic. Up to 15% of patients are likely to develop PTSD in the year after an ACS or cardiac surgery, most commonly between 3 and 6 months after the cardiac event. Patients with Automated Implantable Cardioverter Defibrillators (AICDs) may also experience PTSD following the firing of the device. In one study, having more than five shocks from an ICD predicted higher post-traumatic stress symptoms at follow-up.

Risk factors for the development of cardiac-related PTSD include prior history of PTSD, depression, or generalized anxiety disorder. Cardiac events such as MI differ from other traumatic events in important ways. The life-threatening experience comes from within rather than outside the body, so that instead of directing vigilance to external sources of danger, the cardiac patient may instead become hypervigilant about and preoccupied with somatic signals of disease, resembling patients with the somatosensory amplification experience seen in patients with panic and illness anxiety.

Treatment of PTSD in cardiac patients follows the general guidelines for the treatment of PTSD with psychotherapy, focusing on a review of the traumatic events and the use of SSRIs. A trial found telephone-based CBT reduced PTSD symptoms in ICD patients but did not result in an effect on the rate of ICD discharge.

Delirium

Delirium is a common problem in severely ill cardiac patients. Patients most at risk for delirium include those undergoing cardiac surgery and those with HF. The incidence of delirium is 17% among those admitted to a cardiac floor, 20% among those admitted for ACS, and up to 25% among those undergoing cardiac surgery. In patients with HF, delirium commonly results from hypoxia, hyponatremia, liver failure, or azotemia, as pulmonary congestion and poor end-organ perfusion progress. In postoperative patients, the clinical picture may include all of these elements; in addition, cerebrovascular hypoperfusion during surgery, infection, sedatives, and narcotics may contribute to delirium. Management relies on correcting the underlying abnormality while managing psychosis or agitation with antipsychotic agents. IV haloperidol (Haldol) has been commonly used and well studied in delirious cardiac patients and can be administered frequently, with appropriate hemodynamic and ECG monitoring. Other sedatives should generally be avoided, although the agitated patient may benefit from concurrent administration of lorazepam (Ativan) with haloperidol. It should be noted that no antipsychotic drug has an indication for the treatment of delirium and that the FDA has mandated a warning label about the risk of sudden death associated with antipsychotic drug use in the management of agitation in older adults. The use of dexmedetomidine for sedation and pain control, sedative interruption, and early mobilization appears to reduce the incidence of delirium after cardiac surgery, although it can cause bradycardia.

Persistent, subtle memory and cognitive impairment may also occur after CABG. Predictors of adverse neuropsychiatric outcomes include older age, proximal aortic atherosclerosis, prior history of neurologic disease, systolic HTN, pulmonary disease, and excessive alcohol use. Adverse cerebral outcomes substantially increase hospitalization duration, in-hospital mortality, and the need for discharge to institutional supportive care. Although embolic events associated with cardiopulmonary bypass have long been associated with adverse cerebral outcomes, cardiac surgery performed “off-pump,” that is, without cardiopulmonary bypass, has not been found to result in significant improvement in the rate of cognitive impairment or neuropsychiatric adverse outcome.

ADHD

Stimulants have been associated with an increased risk of cardiovascular events and death. All stimulants are associated with increased HR and elevated blood pressure. Patients should be evaluated for CAD before beginning therapy and be periodically monitored, primarily with blood pressure checks, while they are treated. Despite concerns regarding cardiovascular risk, these medications are generally safe. Recent studies have shown that among young and middle-aged adults, current or new use of ADHD medications, compared with nonuse or remote use, is not associated with an increased risk of serious cardiovascular events regardless of CAD history. In general, prescreening of HTN and arrhythmia is conducted. However, routine pretreatment screening with ECGs is not currently recommended by the American Academy of Pediatrics or the American Academy of Child and Adolescent Psychiatry.

PSYCHIATRIC SIDE EFFECTS OF CARDIOVASCULAR MEDICATIONS

A few drugs used in the treatment of cardiovascular disorders have psychiatric side effects that should be kept in mind when assessing differential diagnosis of psychopathology in patients with cardiovascular disease. Treatment of cardiac rhythm disorders can engender psychiatric complications. Many antiarrhythmic agents have psychiatric side effects, including delirium, hallucinations, paranoid ideation, and mood disturbance. Some of these effects are listed in Table 27.2–5. Amiodarone deserves special mention because it is widely used for patients with arrhythmia control and can cause profound hypothyroidism in just a few months. Symptoms can include cognitive dulling and depressed mood and maybe misattributed to depression or underlying HF; correction of the hypothyroid state may be neglected. In addition, amiodarone inhibits hepatic drug metabolism via cytochromes P450 3A4, 2D6, and 2C9, potentially increasing blood levels of many psychotropic agents metabolized through these enzyme systems and increasing the likelihood of toxicities, including QT prolongation and arrhythmic events. The calcium channel blocker diltiazem is also a strong inhibitor of the 3A4 system with the potential for toxic drug interactions. Whereas β-blockers were previously thought to contribute to depression, that association remains controversial, with the majority of studies finding no correlation. Finally, digoxin is sometimes associated with delirium.

Table 27.2–5.
Some Psychiatric Side Effects of Cardiovascular Drugs

Medication	Effects
Digoxin (Lanoxin)	Visual hallucinations
Angiotensin-converting enzyme inhibitors	Mood elevation
Thiazide diuretics	Mood elevation, erectile dysfunction
β-Adrenergic blockers	Fatigue, sexual dysfunction
Lidocaine (Dalcaine)	Delirium
Mexiletine (Mexitil)	Delirium
Amiodarone (Cordarone)	Hypothyroid state, with associated

ACKNOWLEDGMENT

This chapter is an update of a previous chapter by Peter Shapiro, MD and Adam Critchfield, MD and includes text from their original chapter.

FURTHER READINGS

Barocas JA, Morgan JR, Wang J, McLoone D, Wurcel A, Stein MD. Outcomes associated with medications for opioid use disorder among persons hospitalized for infective endocarditis. *Clin Infect Dis*. 2021;72(3):472–478.

Beach SR, Celano CM, Noseworthy PA, Januzzi JL, Huffman JC. QTc prolongation, torsades de pointes, and psychotropic medications. *Psychosomatics*. 2013;54(1):1–13.

Bixby AL, VandenBerg A, Bostwick JR. Clinical management of bleeding risk with antidepressants. *Ann Pharmacother*. 2019;53(2):186–194.

Brojmohun A, Lou JY, Zardkoohi O, Funk MC. Protected from torsades de pointes? What psychiatrists need to know about pacemakers and defibrillators. *Psychosomatics*. 2013;54(5):407–417.

Campbell KA, Madva EN, Villegas AC, et al. Non-cardiac chest pain: a review for the consultation-liaison psychiatrist. *Psychosomatics*. 2017;58(3):252–265.

Caro MA, Rosenthal JL, Kendall K, Pozuelo L, Funk MC. What the psychiatrist needs to know about ventricular assist devices: a comprehensive review. *Psychosomatics*. 2016;57(3):229–237.

Charton M, Flécher E, Leclercq C, et al. Suicide attempts among LVAD recipients: real-life data from the ASSIST-ICD study. *Circulation*. 2020;141(11):934–936.

Chernyak Y, Teh L, Henderson DR, Patel A. Practice issues for evaluation and management of the suicidal left ventricular assist device patient. *Prog Transplant*. 2020;30(1):63–66.

Celano CM, Millstein RA, Bedoya CA, Healy BC, Roest AM, Huffman JC. Association between anxiety and mortality in patients with coronary artery disease: a meta-analysis. *Am Heart J*. 2015;170(6):1105–1115.

Celano CM, Villegas AC, Albanese AM, Gaggin HK, Huffman JC. Depression and anxiety in heart failure: a review. *Harv Rev Psychiatry*. 2018;26(4):175–184.

Chou R, Cruciani RA, Fiellin DA, et al. Methadone safety: a clinical practice guideline from the American Pain Society and College on Problems of Drug Dependence, in collaboration with the Heart Rhythm Society. *J Pain*. 2014;15(4):321–337.

Dew MA, DiMartini AF, Dobbels F, et al. The 2018 ISHLT/APM/AST/ICCAC/STSW recommendations for the psychosocial evaluation of adult cardiothoracic transplant candidates and candidates for long-term mechanical circulatory support. *Psychosomatics*. 2018;59(5):415–440.

Dietze CS, Ekosso-Ejangue L, Israel CW, Bien CG, Fauser S. Benefits of additional cardiologic examination in patients admitted for differential diagnosis to the Epilepsy Center Bethel. *Epilepsy Res*. 2018;148:44–47.

Drudi LM, Ades M, Turkdogan S, et al. Association of depression with mortality in older adults undergoing transcatheter or surgical aortic valve replacement. *JAMA Cardiol*. 2018;3(3):191–197.

Elbatarny M, Bahji A, Bisleri G, Hamilton A. Management of endocarditis among persons who inject drugs: a narrative review of surgical and psychiatric approaches and controversies. *Gen Hosp Psychiatry*. 2019;57:44–49.

Funk MC, Beach SR, Bostwick JR, et al. QTc prolongation and psychotropic medications. *Am J Psychiatry*. 2020;177(3):273–274.

Gomez-Bernal F, Madva EN, Puckett J, Amonoo HL, Millstein RA, Huffman JC. Relationships between life stressors, health behaviors, and chronic medical conditions in mid-life adults: a narrative review. *Psychosomatics*. 2019;60(2):153–163.

Gray ME, McQuade ETR, Scheld WM, Dillingham RA. Rising rates of injection drug use associated infective endocarditis in Virginia with missed opportunities for addiction treatment referral: a retrospective cohort study. *BMC Infect Dis*. 2018;18(1):532.

Huffman JC, Mastromauro CA, Beach SR, et al. Collaborative care for depression and anxiety disorders in patients with recent cardiac events: the Management of Sadness and Anxiety in Cardiology (MOSAIC) randomized clinical trial. *JAMA Intern Med*. 2014;174(6):927–935.

Ivkovic A, Puckett J. The psychological and pharmacological correlations of atrial fibrillation. *Psychiatr Ann*. 2016;46(12):702–711.

Lichtman JH, Froelicher ES, Blumenthal JA, et al. Depression as a risk factor for poor prognosis among patients with acute coronary syndrome: systematic review and recommendations: a scientific statement from the American Heart Association. *Circulation*. 2014;129(12):1350–1369.

Lyon AR, Citro R, Schneider B, et al. Pathophysiology of Takotsubo syndrome: JACC state-of-the-art review. *J Am Coll Cardiol*. 2021;77(7):902–921.

Mullan C, Caraballo C, Ravindra NG, et al. Psychiatric comorbidity and outcomes after left ventricular assist device implantation for end-stage heart failure. *JACC Heart Fail*. 2020;8(7):569–577.

Rautaharju PM, Surawicz B, Gettes LS, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part IV: the ST segment, T and U waves, and the QT interval a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society Endorsed by the International Society for Computerized Electrocardiology. *J Am Coll Cardiol*. 2009;53(11):982–991.

See I, Gokhale RH, Geller A, et al. National public health burden estimates of endocarditis and skin and soft-tissue infections related to injection drug use: a review. *J Infect Dis*. 2020;222(Suppl 5):S429–S436.

Surawicz B, Childers R, Deal BJ, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part III: intraventricular conduction disturbances a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology. *J Am Coll Cardiol*. 2009;53(11):976–981.

Swetz KM, Ottenberg AL, Freeman MR, Mueller PS. Palliative care and end-of-life issues in patients treated with left ventricular assist devices as destination therapy. *Curr Heart Fail Rep*. 2011;8(3):212–218.

Wilson SR, Givertz MM, Stewart GC, Mudge GH Jr. Ventricular assist devices: the challenges of outpatient management. *J Am Coll Cardiol*. 2009;54(18):1647–1659.

Writing Committee Members; Otto CM, Nishimura RA, Bonow RO, et al. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Joint committee on clinical practice guidelines. *J Am Coll Cardiol*. 2021;77(4):e25–e197.

Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2013;62(16):e147–e239.

27.3 Gastrointestinal Disorders

ASHWINI NADKARNI, M.D.

The relationship between the gastrointestinal (GI) tract and the brain has been apparent for centuries, given such idioms as “butterflies in my stomach” or “gut-wrenching.” However, it was not until the 19th century that such observations began to be objectively examined. William Beaumont (1785 to 1853), an early pioneer of GI physiology, isolated hydrochloric acid from gastric secretions and described the influence of emotions on gastric acid secretion. While the techniques used to interrogate the anatomy and physiology of the GI tract function have advanced, the relationship between the psyche and GI function, both in health and disease, remains prominent in modern clinical practice and research. Research from a diverse array of disciplines, coupled with technologic advances, particularly in functional neuroimaging, has provided an increasing body of evidence for the dynamic role of the intestinal microbiome in influencing bidirectional communication between the brain and the gut or the central nervous system (CNS) and enteric nervous system. Such neural, immune, and hormonal interactions which comprise the brain–gut–microbiome axis have gained widespread acceptance as the germane construct to explain the normal, acute, and chronic alterations in digestive tract function. Moreover, this model of circuitous communication has provided a biologic construct to underpin the biopsychosocial concept of GI disorders by facilitating the integration of many contributing factors, whether they are biologic, psychological, or social in nature.

In many ways, considering the essence of the construct of the brain–gut–microbiome axis, it should come as no surprise that many GI disorders are comorbid with psychiatric disorders. As an example, major depressive disorder, a disease associated with immune activation and cytokine release, is closely associated with inflammatory bowel disease (IBD). The aims of this section are to provide the reader with a summary of both IBD and functional GI disorders (FGIDs), itself comprised of a number of distinct disorders, as many of these patients are routinely referred to the psychiatrist. There has been an appreciable interest in developing a mechanistic-based understanding of the pathophysiology of FGIDs over the recent past. Within secondary care, more than one-third of new patients are diagnosed as having one or more FGIDs. The FGIDs are a heterogeneous group consisting of in excess of 25 separate disorders, defined according to the Rome multinational consensus as “variable combinations of chronic or recurrent GI symptoms which are not explained by structural or biochemical abnormalities.” FGIDs are characterized by impairment in intestinal movement and nerve sensitivity or brain–gut dysfunction, resulting in pain and discomfort in the absence of any physical, biologic, or anatomical abnormalities. The most prevalent functional GI disorder is irritable bowel syndrome (IBS), with a reported global prevalence in excess of 10% of the adult population, peaking in the second, third, and fourth decades of life with a female preponderance.

The Rome process, which has developed in an iterative fashion in a manner not dissimilar to the *Diagnostic and Statistical Manual of Mental Disorders,* has operationalized diagnosing FGIDs. In particular, Rome IV utilizes updated diagnostic algorithms and integrates innovations in research on the contributions of the brain–gut–microbiome axis and pharmacogenomics to functional disorder pathophysiology. In addition, Rome IV incorporates information on the patient illness experience and biopsychosocial, gender, and cross-cultural understandings of FGIDs. The use of imprecise and stigmatizing terms is therefore diminishing, facilitating the pivotal role of the psychiatrist to not only treat FGID, particularly in those who are refractory to standard medical therapy but also reinforce psychoeducation on the neurobiologic underpinnings of functional disease. Thus, modern nomenclature has integrated the value of the “liaison” role of the psychiatrist in collaborating with the patient as well as the medical team. The chapter will be structured around examining both “organic” and “functional” disorders of the digestive tract. The seemingly arbitrary split of gastroenterologic disorders in “organic” or “functional” based upon the presence or absence of a particular biomarker remains the dominant model in specialist practice, although such dualism has marked limitations. For instance, peptic ulcer disease was considered to be a psychosomatic disorder prior to the identification of *Helicobacter pylori*. Furthermore, while the identification of colonic inflammation may be consistent with a diagnosis of gastroenteritis or IBD, many patients continue to remain symptomatic even in the context of mucosal healing and resolution of inflammation, pointing to the possibility of functional GI symptoms which overlap with IBD. This has led to the proposal that IBD and IBS, in reality, represent a similar disease at differing ends of a spectrum where central and peripheral factors are involved in symptom generation. This chapter will be structured around describing the epidemiology, putative pathogenesis, and broad management of both “organic” and “functional” GI disorders. Each disorder will be further illustrated with a clinical case to illustrate important practice points as well as a focus on the pivotal role of the psychiatrist in successfully managing such patients.

INFLAMMATORY BOWEL DISEASE

IBD occurs when inappropriate inflammation occurs in the gut in genetically susceptible individuals. The two major types of IBD are Crohn disease (CD) and ulcerative colitis (UC), with the former affecting any portion of the GI tract from mouth to anus, whereas the latter is confined solely to the colon. The classification of indeterminate colitis describes patients in whom a diagnosis of UC or CD cannot be made based on standard clinical testing, including colonoscopy, imaging, laboratory tests, and biopsy. Patients with IBD present with abdominal pain, diarrhea, fatigue, malnutrition, weight loss, anemia, joint pain, anemia, and skin lesions. While recent data on the prevalence of IBD was reported to range from 252 to 439 cases per 100,000 population in the United States, a shift in the epidemiologic pattern has been reported, indicating a stabilizing incidence in high-income countries with a high burden and prevalence and a rapid rise in newly industrialized countries in South America, Eastern Europe, Asia, and Africa.

Although the etiopathogenesis of IBD is incompletely understood, evidence suggests that this disorder is caused by a complex interaction between an exaggerated immunologic response to a proinflammatory microbiome in a genetically susceptible host. Severe impairment in health-related quality of life results not only from changes in bowel habits, abdominal pain, and rectal bleeding but also fatigue, cognitive disorders, and extraintestinal manifestations such as inflammatory arthritis or primary biliary cirrhosis. Moreover, medical interventions such as systemic corticosteroids or immunosuppressive drugs may result in secondary health problems compounding the severity.

Mood disorders affect up to one in three patients with IBD; however, the precise relationship between psychiatric disorders and IBD is incompletely understood. For instance, the question of whether depression has a temporal relationship with IBD or the role of antidepressants in protecting against IBD remains to be clarified. Evidence suggests that anxiety and depression are more common in patients with IBD, particularly severe and active IBD, and that the symptoms of these affective disorders are more prominent during periods of disease relapse. Nevertheless, there is a paucity of studies that objectively investigate the hypothesis that anxiety and depression are risk factors for IBD. In one large and robust cohort study (n = 403,665), depression was found to increase the risk of developing both CD and UC; moreover, the use of selective serotonin reuptake inhibitors (SSRIs) and tricyclic antidepressants (TCAs) was found to protect against both CD and UC, thereby mitigating the risk of IBD.

Population-based studies examining the incidence and prevalence of psychiatric disorders after a diagnosis of IBD have found that prevalence rates of depression, anxiety, and bipolar disorder are higher in IBD patients than in matched patient cohorts without IBD, while the prevalence rates of schizophrenia do not differ. One cohort study of pediatric patients ($n = 226$) found three depression phenotypes: mild depression, associated with fatigue, irritability, and depressed feelings; somatic depression, associated with anhedonia, change in appetite, fatigue, physical complaints, irritability, and depressed feelings; and a third group exhibiting cognitive despair and more traditional depressive symptoms, such as morbid and suicidal ideation, weeping, hopelessness, fatigue, and feelings of sadness.

Depression and anxiety may affect symptoms and functioning through the development of heightened pain sensitivity of the intra-abdominal organs (termed visceral hypersensitivity), exacerbation of bowel symptoms, and reduced self-motivation to overcome the difficulties associated with IBD. An increased frequency of IBS overlay, visceral hypersensitivity, fatigue, cognitive dysfunction, and insomnia are also correlated to depression. Cognitive dysfunction includes impairment in signal processing speed, short-term memory, and verbal IQ deficits linked to the impact of systemic inflammation.

Among anxiety disorders, generalized anxiety disorder, panic disorder, and social anxiety disorder occur at high rates. Phobias include fear of nausea, vomiting, fecal incontinence, and food and its impact on IBD symptoms. Fear of incontinence can also precipitate agoraphobia. Post-traumatic stress disorder (PTSD) is particularly prominent, given the susceptibility to hospitalization, surgery, and symptom severity.

The relationship between IBD and depression is bidirectional: Within the evolving area of psychoneuroimmunology, the biologic mechanisms by which the nervous system can influence immune function, both at the level of the GI mucosa and systemically, are increasingly being appreciated, particularly with animal models. However, such mechanisms are likely to be numerous and interactional with one another. For example, the contents of mast cells, various cytokines, and chemokines may activate primary afferent neurons, thereby sensitizing them and facilitating increased afferent input into central pain centers. In addition, depression may adversely impact health-related quality of life because of social withdrawal, irritability toward others, as well as rumination, negativistic thinking, and hopelessness.

Psychological and Psychiatric Interventions in IBD

The treatment of depression advances outcomes for IBD, impacting self-management behaviors, maintenance of social support networks, and relationships with the treatment team or caregivers. Improvement of depression also reduces rates of IBD relapse and time to relapse as well as optimizes the response to biologic immunomodulators. A recent study reported by Szigethy et al. compared the efficacy of cognitive behavioral therapy (CBT) to supportive nondirective therapy (SNDT) in treating adolescents with comorbid IBD and depression. While both psychotherapies resulted in significant reductions in depression scales, SNDT was associated with a greater reduction in IBD activity in the absence of health-related quality-of-life improvements. Hypnotherapy may also have a role in treating IBD. Miller and Whorwell administered gut-focused hypnotherapy to 15 patients with severe IBD and demonstrated improvements in quality of life and a decreased reliance on corticosteroids (more than half of the patients stopped corticosteroid therapy completely). They concluded that hypnotherapy might represent a promising adjunctive treatment of IBD with steroid-sparing effects, although there is a need for well-designed controlled trials in the area. A recent meta-analysis also identified seven randomized controlled trials (RCTs) for mind/body interventions such as lifestyle modification, hypnotherapy, relaxation training, and mindfulness, which may display a degree of efficacy in treatment, although the results were heterogeneous and at significant risk of bias.

Psychotropic medications are empirically supported in the treatment of patients with IBD. Reduced appetite, nausea, bowel-movement frequency, and weight changes are responsive to atypical antidepressants such as mirtazapine. Comorbid IBS, as well as visceral hypersensitivity and insomnia, are useful targets for TCAs. The serotonin–norepinephrine reuptake inhibitors (SNRIs) can be effective for refractory depression, anxiety, and chronic pain syndromes with a side-effect profile similar to SSRIs and more mild manifestations of TCAs. Bupropion has been demonstrated to lower circulating levels of tumor necrosis factor-α (TNF-α) and can also have an effect on fatigue and cognitive symptoms. In one prospective cohort study (n = 67), 6 months of antidepressant drug treatment, including SSRIs, SNRIs, and atypical antidepressants, was associated with an improvement in depression, anxiety, quality of life, and sexual functioning scores, as well as an improvement in CD activity index.

A 62-year-old woman with a known history of UC presented to the gastroenterology outpatient with a 2-week history of rectal bleeding, abdominal discomfort, and profuse diarrhea. On clinical examination, she was tachycardic, and her left iliac fossa was tender in the absence of guarding. Apart from raised inflammatory markers, her hematologic, biochemical, and microbiologic markers were normal. She was admitted to the clinic, and a colonoscopy was undertaken, whose appearances are shown in Figure 27.3–1. Following the colonoscopy, a diagnosis of an acute exacerbation of UC was made, and the patient was commenced on oral 5-aminosalicylates and intravenous hydrocortisone 100 mg a day. After 48 hours of therapy, her GI symptoms began to improve rapidly. On the third day of her admission, she began to become confused and agitated. Hematologic and biochemical parameters were normal, and an infection screen did not demonstrate any abnormalities. A computerized tomography (CT) scan of her head was normal. Over the next 3 days, she began to become increasingly anxious and suspicious of staff members, suggesting that they belonged to the secret service. A liaison psychiatrist was called, and a diagnosis of steroid-induced psychosis was made. She was commenced on regular haloperidol, and the intravenous steroids were rapidly tapered and discontinued. She was commenced on the immunosuppressive agent azathioprine. Her psychiatric symptoms improved over the next 10 days, and she was discharged home.

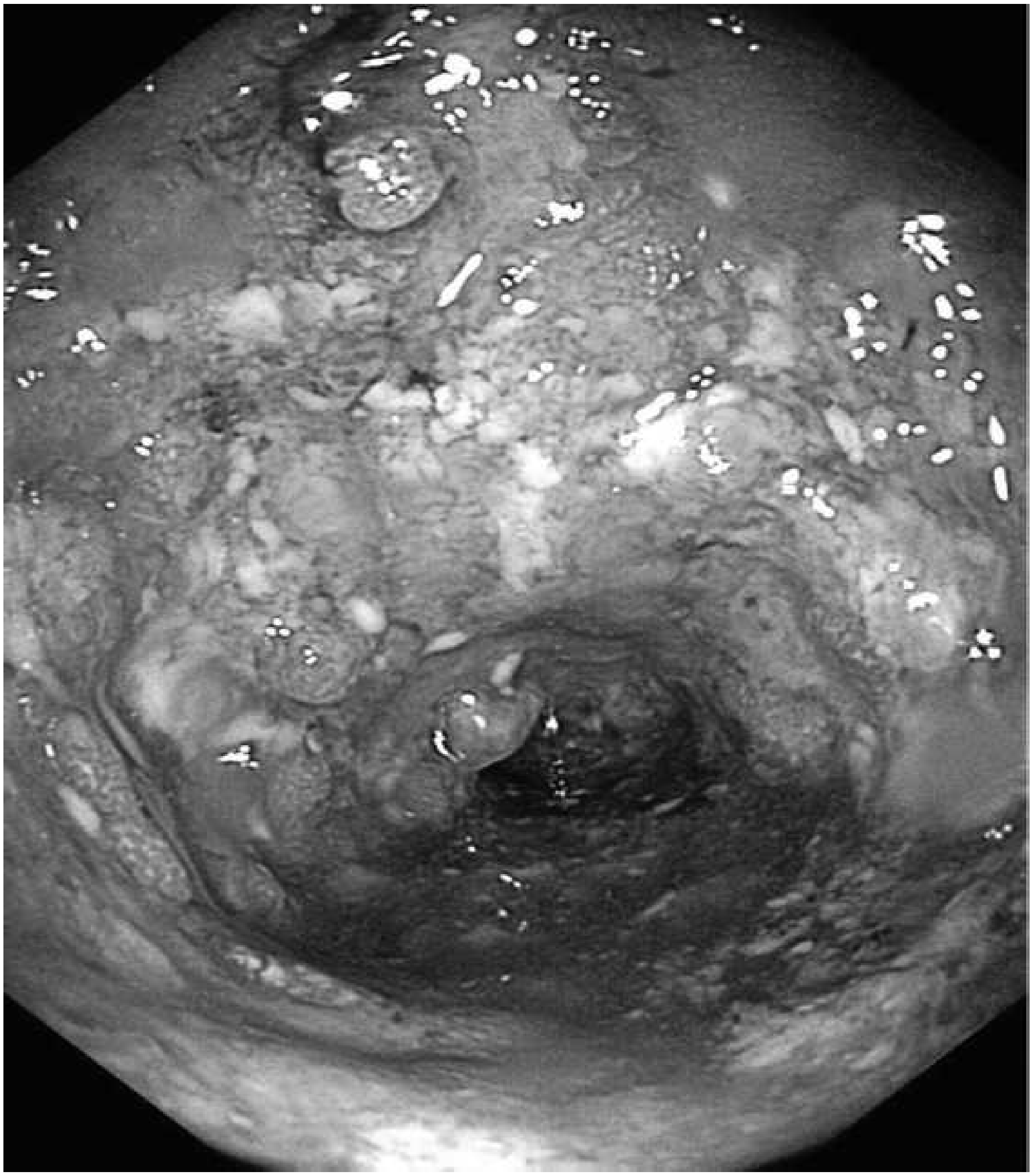

FIGURE 27.3–1. Appearances of the rectum at colonoscopy. The mucosa is friable with marked erythema, absent vascular pattern, and deep ulceration. These features are consistent with severe disease.

Learning Points

1. Steroid-induced psychosis is uncommon but can present in patients with IBD.
2. In such cases, discontinuation of steroids and appropriate management of psychosis is needed.
3. A number of factors have been postulated to confer a higher risk of steroid-induced psychosis, including female gender and higher doses

of steroids.

FUNCTIONAL GI DISORDERS

Visceral pain is common and remains a leading worldwide cause of health care usage. Moreover, unexplained abdominal pain is the 6th most common cause of acute presentation to secondary care service in women and the 10th most common cause in men. In the United Kingdom, it has been estimated that nonspecific abdominal pain exerts an annual cost in excess of £100 million. FGIDs are a group of heterogeneous disorders that represent one of the great unmet needs in gastroenterology and are characterized by recurrent unexplained abdominal pain. In contrast to organic GI disorders consisting of impairment in organ morphology or motility GI disorders resulting in impairment of function, FGIDs relate solely to the patient's interpretation and reporting of an illness experience.

Although all FGIDs are associated with a reduction in health-related quality of life, they are not associated with any mortality per se. For the sake of brevity, this chapter will discuss three particularly prevalent classes of FGIDs categorized according to the updated Rome-IV criteria and most likely to be encountered by the psychiatrist. These are IBS (a functional bowel disorder), centrally mediated abdominal pain syndrome (CAPS) (a functional GI pain disorder), and functional chest pain (FCP) (a functional esophageal disorder).

FUNCTIONAL BOWEL DISORDERS: IRRITABLE BOWEL SYNDROME

The most frequently encountered functional GI disorder is IBS, with a reported global prevalence of >10% and a marked socioeconomic burden. IBS is also associated with a significant diminution in work productivity and health-related quality of life. The main clinical features of IBS are recurrent abdominal pain and/or discomfort, associated with an alteration in stool frequency or consistency, and relief with defecation (Table 27.3–1). Extra-GI features are also prevalent and include lassitude, headache, dysmenorrhea, and dyspareunia. While symptoms of IBS characteristically wax and wane, Rome-IV criteria additional require patient reporting of pain or discomfort at least 4 days per month over the past 2 months. IBS can be usefully subdivided based on predominant bowel habits and is grouped as IBS-C, IBS with diarrhea, IBS with mixed bowel habits, or IBS unclassified (see Fig. 27.3–2).

Table 27.3–1.
The Rome IV Definition of Irritable Bowel Syndrome

Rome IV Diagnostic Criteria[a] for Irritable Bowel Syndrome

Recurrent abdominal pain on average at least 1 day/week in the last 3 months, associated with two or more of the following criteria:

1. Related to defecation
2. Associated with a change in the frequency of stool
3. Associated with a change in the form (appearance of stool)

[a] Criteria fulfilled for the last 3 months with symptom onset at least 6 months prior to diagnosis.

Modified from the original Rome Foundation Rome IV Criteria—Reprinted with permission from the Rome Foundation; all rights reserved.

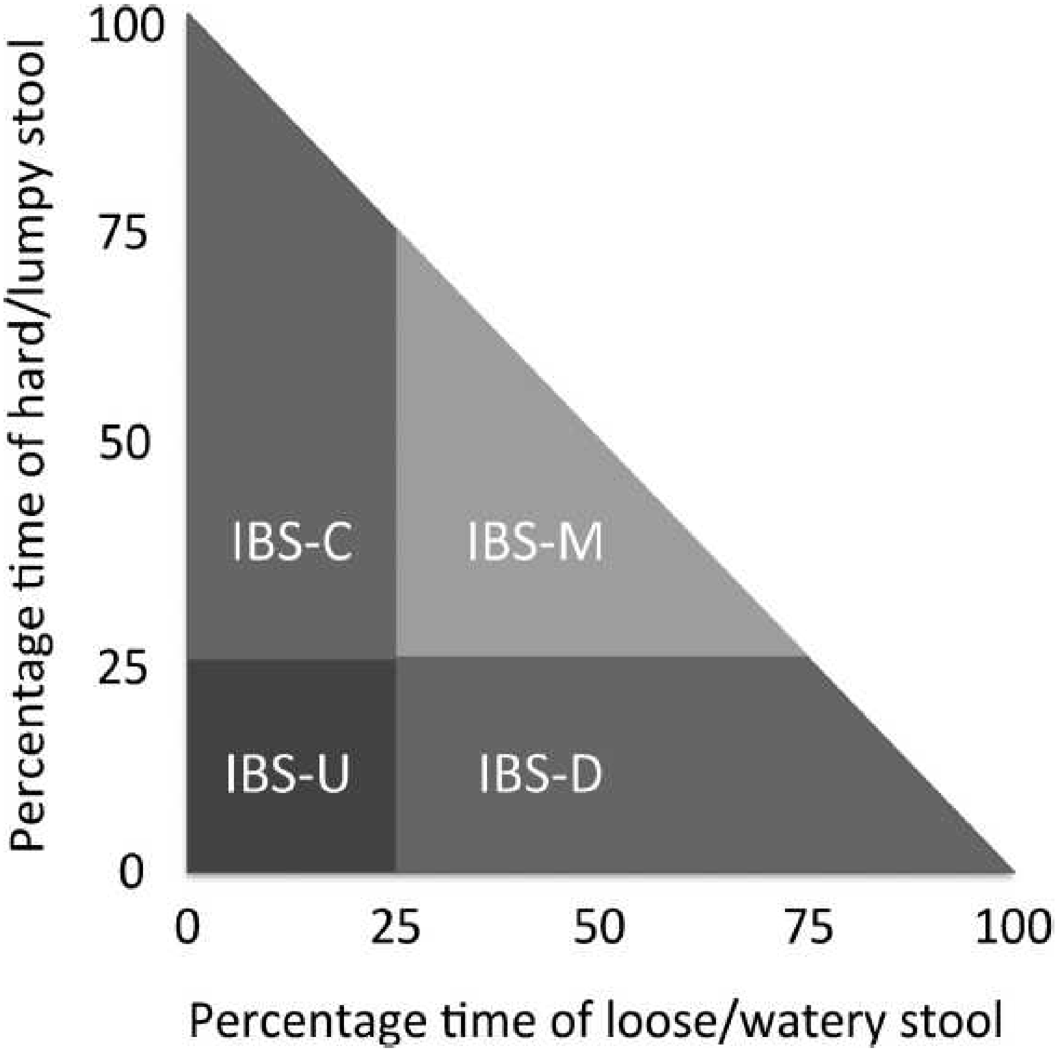

FIGURE 27.3–2. The different subtypes of IBS are classified according to the predominant bowel habit: IBS with constipation (IBS-C), IBS with diarrhea (IBS-D), IBS with mixed bowel habit (IBS-M), and IBS-unclassified (IBS-U).

Psychiatric comorbidities such as major depressive disorder, panic disorder, generalized anxiety disorder, PTSD, and somatic symptom disorder occur at the highest frequencies: and it has been estimated that over half of patients with IBS suffer from these disorders. One cohort study (n = 3,934 IBS subjects; n = 6,127 non-IBS control group) found that patients with IBS had a 3.6-fold risk of developing psychiatric disorders. In animal models, studies have shown that adverse early life events, such as maternal separation, represent independent risk factors for the development of IBS and visceral hypersensitivity in adulthood. In humans, there is evidence that a history of sexual abuse, especially in childhood, can alter visceral pain sensitivity. Furthermore, the psychological context in which GI symptoms are interpreted by an individual may predict the development of IBS following an episode of gastroenteritis. Thus, psychiatric disorders and IBS appear to have bidirectional comorbidities; just as the evolution of IBS symptoms is influenced by affective factors and psychosocial stressors, those psychiatric comorbidities further contribute to the cause and maintenance of IBS symptoms, thereby influencing clinical outcomes. The etiology of such bidirectionality includes dysregulation of the brain–gut axis, hypothalamic-pituitary-adrenal (HPA) axis, altered GI motility, visceral hypersensitivity, infectious factors, enhanced immunologic reactivity, genetic susceptibility, and psychosocial factors.

A 35-year-old woman television producer presented with a 3-year history of increasing abdominal pain, distension, and alternating bowel habit. She described her pain as intermittent and cramping in nature. The intensity was 8/10 (10, being the worst pain possible), to the degree that it was influencing her ability to perform socially and at work. Her pain was temporarily relieved by defecation. She gave no past psychiatric history and denied any affective symptoms. She stressed that her only problem was her GI symptoms and her not being able to effectively cope with them at work. Her clinical examination was unremarkable, as were standard hematologic and biochemical parameters. Her antitissue transglutaminase antibodies were normal. Colonoscopic examination of the lower GI tract was macroscopically and microscopically normal, although she found the procedure very painful. Standard medical therapies did not improve her symptoms. Her gastroenterologist referred her to the local liaison psychiatrist service. She was not anticipating benefit from this referral, as she believed it was not relevant to her GI symptoms.

After taking time to reassure her that the consultation was primarily aimed at dealing with her GI symptoms and sufficiently validating her physical problems, she relaxed, and the rapport improved. On account of this deepening in rapport, she then felt confident enough to volunteer more information about the degree that her physical symptoms affected her. She described easy fatigability, anhedonia, sleep disturbance, weight loss, and low mood. She then discussed how she was experiencing marital and work stress as a result of her physical problems of late and how her “perfectionism and self-criticism” had always been a problem for her in similar situations in the past. The consultation then logically continued to make connections between stress, personality, and somatic symptoms and how best to treat these.

Learning Points

1. Patients with IBS often report discomfort during lower GI endoscopic investigations as a result of visceral hypersensitivity.
2. Standard medical therapies often fail to improve symptoms in IBS.

3. In many cases, patients are often unwilling to engage with psychiatric services.
4. Chronic functional GI symptoms are frequently associated with psychopathology.

Functional Gastrointestinal Pain Disorders: Centrally Mediated Abdominal Pain Syndrome

CAPS (formerly known as functional abdominal pain syndrome) is characterized by frequently recurrent or nearly continuous abdominal pain associated with a degree of loss of daily activity in the absence of a change in bowel habit (see Table 27.3–2). CAPS is notable for the predominance of pain with the exclusion of other symptoms characteristics of FGID. The reported prevalence ranges from 0.5% to 2% with a female preponderance and occurs less commonly than other FGIDs. Associated with the pathophysiology are altered central sensory processing, genetic and environmental vulnerability, and structural brain changes, including increased cortical thickness in the somatosensory cortex and decreased cortical thickness in regions of pain processing such as the insula and anterior cingulate cortex.

Table 27.3–2.
The Rome IV Diagnostic Criteria for the Diagnosis of Central Mediated Abdominal Pain Syndrome

Rome IV Diagnostic Criteria for Centrally Mediated Abdominal Pain Syndrome

*Must include **all** of the following:*

1. Continuous or nearly continuous abdominal pain
2. No or only occasional relationship of pain with physiologic events (e.g., eating, defecation, or menses)[a]
3. Pain limits some aspect of daily functioning[b]
4. The pain is not feigned
5. Pain is not explained by another structural or functional gastrointestinal disorder or other medical condition
6. CAPS is typically associated with psychosocial comorbidity, but there is no specific profile that can be used for diagnosis.
7. Criteria fulfilled for the last 3 months with symptom onset at least 6 months prior to diagnosis.

[a] Some degree of gastrointestinal dysfunction may be present.

[b] Daily function could include impairments in work, intimacy, social/leisure, family life, and caregiving for self or others.

Modified from the original Rome Foundation Rome IV Criteria—Reprinted with permission from the Rome Foundation; all rights reserved.

There is an absence of international consensus with respect to the optimal clinical management of CAPS, largely due to the paucity of well-designed randomized-controlled trials for therapeutic interventions. Anxiety, depression, and somatization are common, but patients are often reluctant to accept the contribution of these to their pain symptom profile. Establishing an effective rapport with the patient, which incorporates empathy, validation, and education, facilitates realistic treatment goals of adaptation to the symptoms rather than exhaustive identification of a cure. Thus, within a contemporaneous practice, psychiatric interventions are focused on providing pharmacologic treatments which minimize emotional distress and leverage advances in analgesia.

A 19-year-old woman was referred to the gastroenterology clinic with long-term chronic abdominal pain that had initially begun during her early teenage years. She had seen five different gastroenterologists in the past. She had abdominal pain most of the time and described the pain as burning in nature. The pain was not associated with a change in bowel habits and was not relieved by defecation. There was no association with her menstrual cycle. She also had multiple tender points across her shoulders and back, which her general practitioner had diagnosed as fibromyalgia. She has been admitted to the hospital in excess of 30 times over the past 12 months. These acute presentations had been in the context of multiple and repeated normal hematologic, biochemical, metabolic, immunologic, endoscopic, and radiologic investigations. Of specific note, she had had a normal CT scan of her abdomen and pelvis and ileocolonoscopy with histology in the last 12 months. The gastroenterologist made a clinical diagnosis of Central Mediated Abdominal Pain Syndrome (CMAP).

Learning Points

1. Patients with one functional GI disorder often suffer from other chronic somatic pain syndromes, such as fibromyalgia.
2. There is often a delay in making the diagnosis of a functional GI disorder.
3. Patients with chronic unexplained symptoms are often disenfranchised from the medical profession. The lack of a specific diagnosis often leads to multiple investigations, which in itself may heighten a patient's anxiety. Frequently, the consequences of this are poor outcomes to treatment and onward referral to another physician, and thus the cycle begins again (see Fig. 27.3–3). This cycle can be broken with time, empathy, and validation of the patient's symptoms.
4. Team-based care with collaboration between the psychiatrist, therapist, gastroenterologist, primary care physician, and nutritionist assist both the patient and clinician in achieving improved outcomes.

FUNCTIONAL ESOPHAGEAL DISORDERS: FUNCTIONAL CHEST PAIN

FCP is characterized by recurrent unexplained retrosternal chest pain in the absence of associated esophageal symptoms such as heartburn and dysphagia. FCP accounts for about one-third of patients with noncardiac chest pain. The Rome IV diagnostic criteria include at least 3 months of symptoms, with onset at least 6 months prior to diagnosis. A change from Rome III to Rome IV is that gastroesophageal reflux disease (GERD), eosinophilic esophagitis, a major esophageal motor disorder, must be excluded to make the diagnosis requiring workup with upper endoscopy and biopsy, reflux testing, and esophageal manometry. Notwithstanding significant reductions in quality of life, patients often have disproportionately high health care utilization and often manifest in recurrent negative investigations across medical specialties, such as cardiology, and are frequently recalcitrant to standard pharmacotherapies.

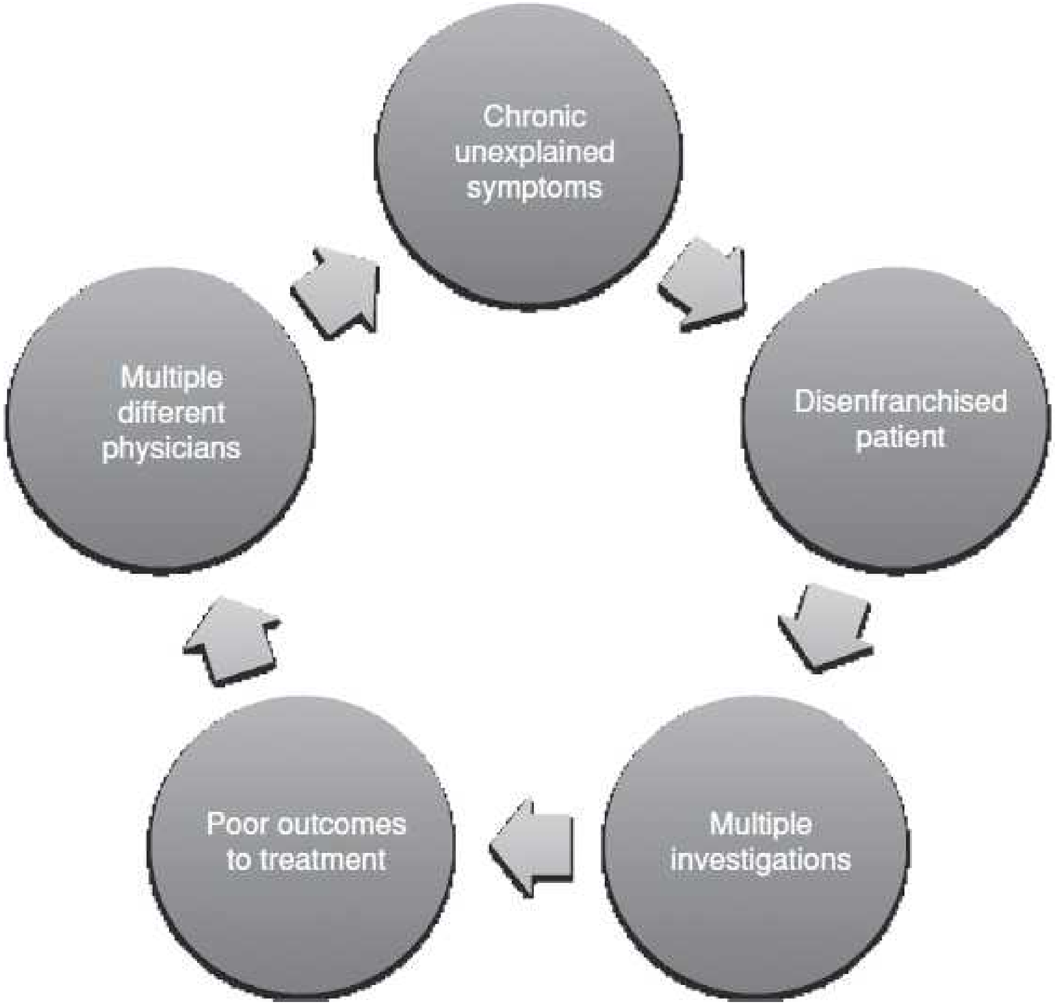
Chronic unexplained symptoms
Disenfranchised patient
Multiple investigations
Poor outcomes to treatment
Multiple different physicians

FIGURE 27.3–3. A patient presenting with chronic unexplained symptoms is frequently investigated multiple times, across multiple physicians leading to poor outcomes. Such a cycle can only be interrupted with time, empathy, and validation of the patient's symptoms.

The pathophysiologic mechanisms that underpin FCP are incompletely understood; however, several broad mechanisms have been suggested.

First, heightened mechano- or chemoreceptor esophageal hypersensitivity occurs in response to esophageal distension. Second, the stress-responsive physiologic systems, encompassing the autonomic nervous system (ANS) and the HPA axis, are implicated in a number of functional disorders, including FCP. These stress-responsive systems act as brain–body interfaces and are critical components of the physiologic adaption in response to stress. Third, altered central processing of intraesophageal stimuli is another proposed mechanism. Finally, up to 61% of patients with FCP have a demonstrable degree of psychological comorbidity, which itself may enhance esophageal perception and hypervigilance.

Depressive disorders, anxiety disorders and somatic symptom and related disorders have been described in patients with FCP; psychiatric interventions not only address such comorbidities but also improves the patient's response to the primary therapeutic intervention and reduces the perception of the burden and intensity of the disease. Antidepressant classes such as TCAs, SSRIs, and SNRIs serve as neuromodulators and are the mainstay of treatment.

A 47-year-old man who had had a myocardial infarction and coronary artery bypass graft 17 years prior presented with recurrent episodes of severe onset central chest pain. He had presented on multiple occasions to his local accident and emergency department, where standard testing did not demonstrate a cardiac cause. Onward referral to a cardiologist, who had performed a coronary angiogram, did not demonstrate recurrent ischemic heart disease. The cardiologist arranged a referral to a gastroenterologist who performed an upper GI endoscopy with biopsies, barium meal, and abdominal ultrasound, which were all found to be within normal limits. The further assessment included esophageal manometry and 24-hour pH testing that again was normal. A therapeutic trial of proton pump inhibitors did not improve the symptoms. A diagnosis of FCP of presumed esophageal origin was made. After a period of 15 months, the gastroenterologist, frustrated with the lack of clinical progress, referred him to a psychiatrist. The psychiatrist found him depressed and tearful. He had also developed panic attacks. The psychiatrist diagnosed major depressive disorder with panic attacks. After a detailed discussion with a cardiologist about the risks and benefits of antidepressants, he was treated with an SSRI antidepressant and improved rapidly after 4 or 5 weeks.

Learning Points

1. FGIDs may occur in males.
2. FGIDs frequently occur concurrently with organic disease.

3. Numerous bodily symptoms may attract numerous investigations for physical diseases. This led to a profound delay in the detection of this patient’s depressive disorder and, therefore, prolonged suffering and impairment of health-related quality of life.
4. Treatment of this patient’s depressive disorder was relatively straightforward, but there needed to be a joint decision with the cardiologist regarding the safety of antidepressant treatment in patients with severe heart disease.

THE ROLE OF THE PSYCHIATRIST IN MANAGING FUNCTIONAL GI DISORDERS

Psychiatrists have a vital role in the clinical management of patients with FGIDs, both directly treating the patient and nurturing a team-based framework for care with the primary care physician, gastroenterologist, and other supporting clinicians. This section provides background information on the FGIDs for psychiatrists, discusses recent research relating to the biopsychosocial mechanisms that contribute to the illness experience, provides practical aspects of diagnosis, and describes a rationale for the treatment of patients with FGIDs for psychiatrists. Most of the treatment options for these conditions involve the use of psychotropic medications and psychotherapeutic methods familiar to psychiatrists. There is hence considerable benefit for psychiatrists in engaging more actively in the care of this often misunderstood and undertreated patient group.

The Somatic Presentations of Depression in Functional GI Disorders and the Failure to Diagnose it Because of “Presentation Bias”

When a patient primarily presents with physical symptoms and the apparent absence of psychiatric symptoms, any subsequent failure to identify and diagnose psychiatric symptoms may be attributed to “presentation bias.” A consultation that is empathic, supportive, and appropriately scheduled pays considerable dividends for both patient and clinician. More often than not, the maxims of “going slow, is going fast” or “don’t do something, stand there” are helpful strategies to help the patient feel trusting enough to disclose their psychosocial associations. In particular, the validation of physical symptoms and vigilance not to be blinded by presentation bias are of critical importance in achieving mutually satisfactory outcomes for both patient and clinician. In an international study of 1,146 patients with major depression, only 69% reported physical symptoms. A similar study showed that in patients presenting with solely somatic complaints, only 22% were correctly diagnosed with anxiety or depression. This is in contrast to patients who presented with psychosocial problems, where 77% were correctly diagnosed with psychiatric conditions. Kroenke et al. observed that with an increase in the number of physical symptoms, there is a marked rise in affective problems, although this frequently goes unnoticed (see Table 27.3–3).

Table 27.3–3.
Physical Symptoms Are a Potential Surrogate Marker of Psychological Distress. The Higher the Number of Physical Symptoms, the Greater the Degree of Psychological Distress

Symptom Counts	Number of Patients	Mood or Anxiety (%)
0–1	106	4
2–3	131	18
4–5	129	31
6–8	96	52
9+	37	78

Adapted from Kroenke K, Jackson JL, Chamberlin J. Depressive and anxiety disorders in patients presenting with physical complaints: clinical predictors and outcome. *Am J Med.* 1997;103(5):339–347. Copyright © 1997 with permission from Elsevier Science Inc.

Largely as a consequence of furthered understanding of FGIDs and their pathophysiologic determinants, there has been a significant improvement in available treatment options. Previously, abnormal motility within the GI tract was highly influential in treatment choices. However, dysmotility fails to fully explain the full symptomology and pathogenesis of FGIDs. It is now widely accepted that dysregulation at any level of the gut–brain axis can lead to chronic symptoms. It is within the construct of dysregulation within the gut–brain axis that one can better appreciate the symptom complex as a whole. Novel neurotransmitter and functional brain imaging research implicate dysregulation of the gut–brain axis at the peripheral, spinal, and cerebral levels, all of which may contribute toward the symptoms of functional GI disorder. Neurotransmitters such as serotonin (5-HT), norepinephrine (NE), corticotropin-releasing factor (CRF), and opioids modify both motility and sensation in the gut. It is for this reason that centrally acting treatments such as psychotropic and psychotherapeutic agents have become more commonly used. Due to their effect on the 5-HT and NE pathways, these treatments directly modulate all levels of the gut–brain axis, in addition to their traditional use of managing any psychological and psychiatric disturbances that are commonly associated with these disorders.

The use of pharmacotherapeutic agents for FGIDs has grown considerably over the last two decades. Recent evidence suggests that up to one in eight patients with a functional GI disorder are offered an antidepressant. It is relatively common that when a patient presents to tertiary care, secondary care gastroenterologists have already commenced an antidepressant.

Psychosocial Factors and the Biopsychosocial Model in Functional GI Disorders

One of the most clinically useful ways to develop a broadened understanding of the pathophysiology underpinning FGIDs is via the biopsychosocial model, first conceptualized by the late George Engel in 1977. In this model, the influence of the CNS (both at the spinal and cerebral level), ANS, and HPA axis may culminate in sensory and motor dysfunction of the GI tract. Furthermore, via the autonomic and peripheral nervous system, the GI tract can communicate with the CNS and thus provide a means of bidirectional communication, termed the gut–brain axis (see Fig. 27.3–4). Threats to homeostasis, whether it be peripheral (e.g., infection and abdominal surgery), central (e.g., history of abuse), or indeed psychosocial (e.g., alexithymia, catastrophization, and ongoing work stress and life events), frequently hold an important role in the perpetuation and clinical manifestation of these disorders through centrally mediated pathways. There is proportionality between the importance of psychosocial factors and the severity of the functional disorder, to the degree that these factors are of major significance at the severe end of the functional spectrum. The inability to incorporate and successfully deal with these factors at present leads to more gastroenterology referrals and needless special investigations at great cost, both financially and in quality of life.

Psychosocial factors have been demonstrated to significantly influence symptom generation and illness behavior during all stages of the natural history of FGIDs: predisposing, precipitating, and perpetuating. It has been shown that up to 75% of patients seeking care at a tertiary referral center also fulfilled the psychiatric diagnostic criteria for anxiety or depression. This represents a marked increase in prevalence compared to patients seen in primary care or even in other general gastroenterology practices. Patients with FGIDs commonly give personal histories of major stressful life events, including sexual abuse, personal losses, separation, or early histories of deprivation. Patients, particularly those at the severe end of the spectrum, may also perpetuate their symptoms by means of maladaptive illness behavior such as catastrophizing that often "drives" and maintains hypersensitivity within the viscera (see Fig. 27.3–5).

Independent of pre-existing psychiatric diagnoses, upsetting feelings and distress with regard to chronic GI conditions have a negative effect on a patient's mental health and physical state. In the biopsychosocial conceptual model, bidirectional influences constantly occur. For instance, psychiatric disorders and stress can enable functional GI disorder symptoms. Likewise, chronic functional GI disorder symptoms can lead to physiologic effects. It is known that stress aggravates motility, lowers pain thresholds, and increases gut inflammation. Gut distension in rats activates the locus coeruleus and leads to anxiety behavior; in humans, it can activate the anterior midcingulate leading to a greater affective pain experience. In this manner, the process both creates and perpetuates symptoms and worsens overall cognition, leading to further examinations and referrals. Repeated physical examinations, poor patient understanding, and incomplete and often minimal or overtechnical explanations from physicians are in many cases a further cause of anxiety and fear that perpetuates the severity of symptoms. This kind of chronic stress, which modulates the ANS and activates the HPA axis, can have a sustained effect on inflammation and the gut–brain axis as a whole and explains conditions such as postinfective (PI)-IBS. Psychosocial distress at the time of infection has been shown to be an independent predictor of later PI-IBS development.

It has been suggested that 15% to 45% of patients with FGIDs additionally fulfill the diagnostic criteria for somatization disorder. In these patients, typically those afflicted with severe and refractory symptoms, non-GI symptoms and comorbid conditions such as fibromyalgia, chronic fatigue, and chronic generalized pain are also reported. These patients are typically the “nonresponders” in a number of pharmacologic intervention research studies, and it is possible that these individuals set a lower threshold for symptom reporting. It is postulated that these patients have central dysregulation of their pain regulatory pathways (central sensitization). As many of these pathways are activated by neurotransmitters in common with psychiatric conditions (such as 5-HT, NE, and opiates), this understanding has important implications for using centrally acting treatments. This can be done as monotherapy or in combinations with other pain medications and/or psychotherapeutic treatments (e.g., CBT and hypnotherapy). It is interesting to note that this could be effective even in the apparent “absence of psychiatric symptoms,” as the pathways are now the target of treatment and not the psychiatric syndrome.

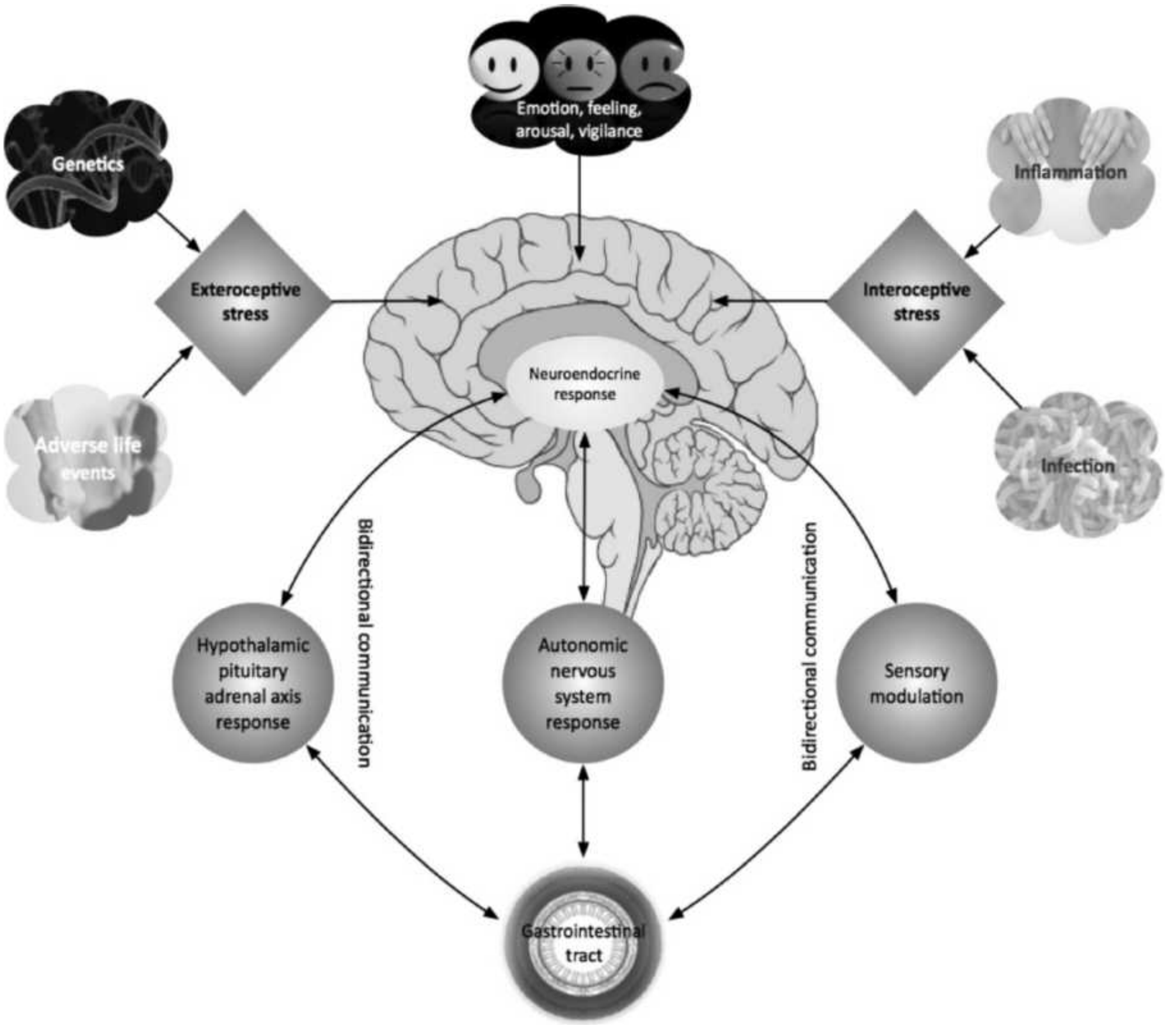
Emotion, feeling, arousal, vigilance
Genetics
Inflammation
Exteroceptive stress
Interoceptive stress
Neuroendocrine response
Adverse life events
Infection
Bidirectional communication
Bidirectional communication
Hypothalamic pituitary adrenal axis response
Autonomic nervous system response
Sensory modulation
Gastrointestinal tract

FIGURE 27.3–4. A highly schematic representation of the gut–brain axis. Exteroceptive or interoceptive stressors may activate a complex neuroendocrine response, which may influence the stress responsive systems, thereby influence GI tract motility and sensory sensitivity. This response feeds back into the central nervous system giving to the concept of bidirectional communication.

GENERAL PSYCHIATRIC APPROACH AND MANAGEMENT FOR FUNCTIONAL GI DISORDERS

Therapeutic Relationship

This can be developed when the psychiatrist (a) elicits and acknowledges the patient's beliefs, concerns, and expectations; (b) offers empathy when needed; (c) clarifies misunderstandings; (d) provides education; and (e) negotiates a plan of treatment with the patient. Some patients are initially unwilling to accept the role of psychosocial factors in the illness and the need for psychiatric intervention; this difficulty is commonly seen in patients who have suffered severe developmental trauma such as sexual abuse.

Associating Bowel Symptoms with Psychosocial Factors

In order to adequately assess the link between bowel symptoms and psychosocial factors, the authors advocate that the patient is asked to keep a daily record of their symptoms along with the frequency and consistency of bowel movements and the timing of menstruation. Such a diary can be usefully complemented with records concerning dietary and lifestyle changes and psychosocial stressors. This information can provide the opportunity for psychoeducation and laying the basis for cognitive behavioral strategies. The psychiatrist’s role in reinforcing the medical condition, namely, an FGID, is also essential since the naming of the disease, alongside explaining the development, pathophysiology, prognosis, and various treatment options, further build rapport with the patient.

Reassurance

Reassurance can only be achieved after the referring physician fully understands the patient’s particular concerns, and it is here where the psychiatrist can act as a mediator, especially with emotionally complicated or aroused patients. The overall favorable prognosis for the functional GI disorder can then be stressed.

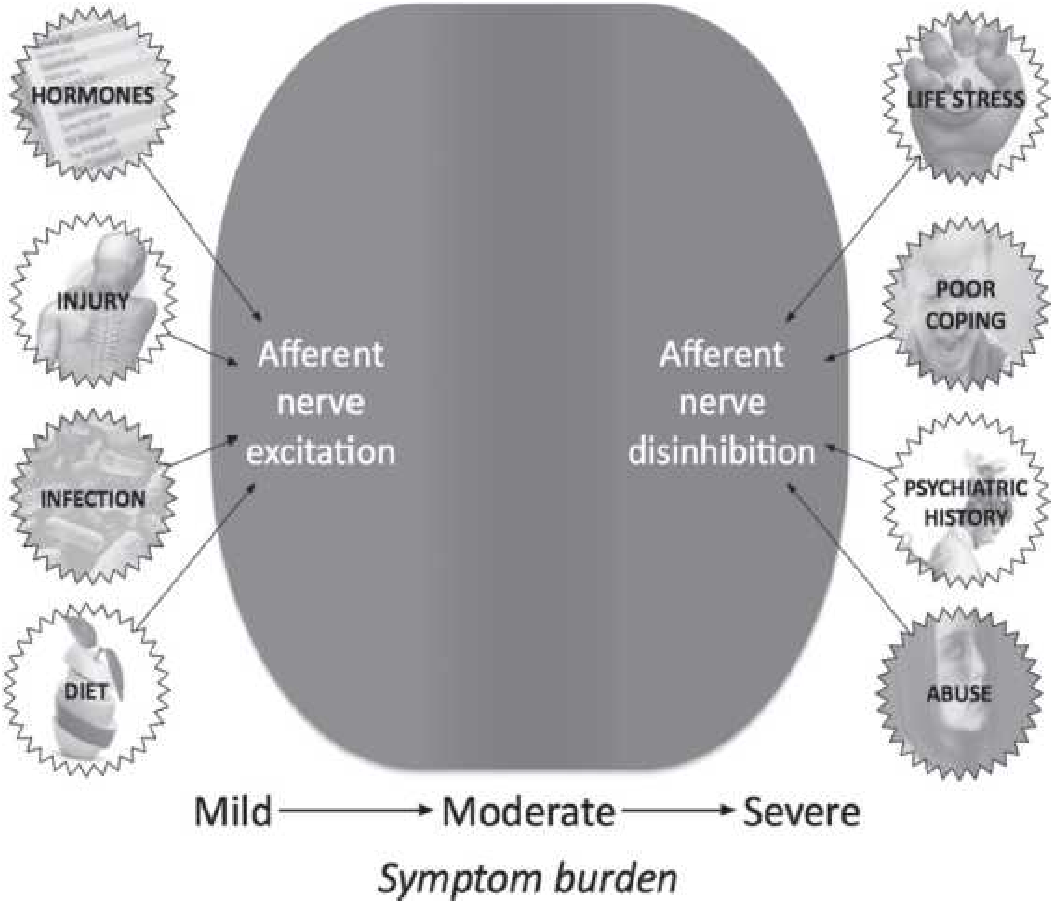
HORMONES
INJURY
INFECTION
DIET
LIFE STRESS
POOR COPING
PSYCHIATRIC HISTORY
ABUSE
Afferent nerve excitation
Afferent nerve disinhibition
Mild
Moderate
Severe
Symptom burden

FIGURE 27.3–5. A number of physiologic factors such as hormones, injury, infection, and diet can alter and enhance visceral afferent nerve signaling. Chronic life stress, poor coping, a psychiatric history, or abuse can cause long-term dysregulation or disinhibition of visceral afferent nerve leading to hyperexcitability. (Data from Drossman DA, Chang L, Bellamy N, et al. Severity in irritable bowel syndrome: a Rome Foundation Working Team report. *Am J Gastroenterol.* 2011;106(10):1749–1759. doi:10.1038/ajg.2011.201.)

Accept the Adaptations of Chronic Illness

When the symptoms are chronic, the patient should be educated as to the nature of the illness as a troublesome cluster of symptoms rather than as an indication of an undiagnosed underlying pathology. In some patients, chronic illness may provide attention from others, release from usual responsibilities, and social and financial compensation. In these situations, appreciable clinical improvement may take a long time but can be advanced if the physician places more focus on improving the patient's function in the presence of illness rather than attempting to "cure." The psychiatrist must help to minimize diagnostic studies and symptomatic treatments and work toward reinforcing health-promoting behavior in the patient.

SPECIFIC PSYCHIATRIC APPROACH AND MANAGEMENT FOR FUNCTIONAL GI DISORDERS

The specific psychiatric approach consists primarily of the diagnosis and treatment of psychiatric comorbidity and includes (1) psychiatric disorders (e.g., major depression and panic disorder) that require specific treatments (e.g., antidepressants, CBT, or other psychotherapy); (2) a history of abuse that comes to light during the consultation and may be interfering with adjustment to the current illness; (3) serious impairment in the daily function that requires specific treatment to improve coping skills; and (4) somatization, where multiple symptoms lead to numerous consultations across specialties. In order for the gastroenterologist to refer a patient to a mental health professional, they must acknowledge the relevance of the psychosocial aspects to the patient's presenting problem. Otherwise, the patient may resist the referral because of perceived stigma or a rejection ("the workup is negative, it must be in your head") by the gastroenterologist. Continuing care by the gastroenterologist may also be necessary.

PSYCHOPHARMACOLOGICAL PRINCIPLES AND MANAGEMENT FOR FUNCTIONAL GI DISORDERS

It is particularly with the use and clinical management of psychotropic medication that the role of the psychiatrist becomes crucial as part of the multidisciplinary team in the treatment of FGIDs. The rationale for using psychotropic agents lies in the high comorbidity: roughly half of the patients with FGIDs also suffer from depression and/or anxiety disorders, which may respond to psychopharmacological intervention. In addition, data continues to support both the efficacy of antidepressants in the relief of chronic pain and FGIDs: Empirical supports exist for efficacy in FGIDs independent of a change in mood.

Antidepressants

The most commonly used psychotropic medications in FGIDs are antidepressants. Specifically, TCAs are commonly used for their dual central and peripheral pain-controlling effects, as well as being well tolerated and, for some, effective at low doses. TCAs additionally slow GI transit, of value for IBS-C patients. Numerous reviews and meta-analyses indicate the potential benefits of antidepressants in FGIDs, producing both a reduction in pain and global improvements in quality of life. They also indicate that the treatment with antidepressants is associated with a significantly decreased relative risk of persistent IBS symptoms. An odds ratio of between 2.6 and 4.2 has been estimated for overall improvement in IBS, translating into a number needed to treat (NNT) for antidepressants of 3.2.

In an open-label physician-rated study, up to 80% of patients with IBS showed improvement in antidepressants, and adherence to treatment was higher than for any other treatments. However, the therapeutic effects of antidepressants on visceral hypersensitivity remain unclear as study results to date have been mixed, although it appears that the peripheral effects on the gut may be of secondary importance as most of these patients responding to antidepressants have failed treatment with conventional gut-acting agents. TCAs and SSRIs may improve global measures of well-being without having a direct correlation with pain ratings. Especially with the use of SSRIs, it seems that improvements in global measures of well-being have the most benefit to patients, as opposed to the direct effect on pain ratings. The antidepressant and anxiolytic effects of SSRIs are mediated by a reduction in the sensitivity of 5HT-1A autoreceptors and heteroceceptors. SSRIs can also stimulate GI motility, thereby assisting with constipation-predominant IBS. Because morbidity associated with FGIDs is linked to the global distress of the patient, as seen with the impairment in social interaction, work absenteeism, and functional limitations, the improvements brought about by SSRIs are still deemed therapeutic, and their use in this context is seen as good practice. Furthermore, the classes of antidepressants have added differential benefits in the treatment of FGIDs. A generally accepted guideline is that the TCAs and SNRIs are used for pain regulation, particularly where diarrhea is the predominant bowel habit, while the SSRIs should be used for somatization associated with anxiety and phobias where constipation predominates. Furthermore, mirtazapine is often helpful in the context of nausea and for boosting appetite.

Antipsychotics and Mood Stabilizers

Antipsychotics like risperidone, olanzapine, and quetiapine have been found to be useful at low doses for visceral pain conditions associated with high arousal states and agitation, for instance, in conditions such as globus, aerophagia, regurgitation, and postprandial distress syndrome. Mood stabilizers such as carbamazepine, sodium valproate, pregabalin, and gabapentin have been found to be helpful in chronic visceral pain disorders. On the whole, the general approach is to first use monotherapies at doses that can be tolerated by the patient. Should this be unsuccessful to remediate symptoms, combinations of the same or different classes could be used. It has been reported that a combination at lower doses could have a better side-effect profile than, for instance, monotherapy at a higher dose.

For patients with more severe symptoms refractory to antidepressants, low doses of quetiapine mitigate anxiety and sleep disturbances, augment the effect of antidepressants, and provide an independent analgesic effect.

Anxiolytics

Anxiolytic agents are effective for reducing anxiety in the short term, but their CNS depressant effect, including mild transient cognitive dysfunction and the risk of addiction with long-term benzodiazepines, limits their use in FGIDs as they are mostly chronic in nature. Prescribing benzodiazepines beyond initial short-term stabilization should therefore only be considered after persistent refractory responses to other pharmacologic options (e.g., psychotropic or mood stabilizers). Alternative strategies for the treatment of anxiety should be adopted.

The antihypertensive clonidine, often employed in treating anxiety disorders, increases colonic compliance and has been found to be effective for symptoms of diarrhea. Patients should be monitored closely for orthostatic hypotension and drops in blood pressure.

Psychotropic Drugs and Neuroregeneration

The rationale for the use of centrally acting treatments in recent years is rooted in the concept of neuroplasticity. This regards the observed loss of cortical neurons following psychiatric trauma, but subsequent neurogenesis (i.e., the regrowth of neurons) following treatment with antidepressants and possibly psychotherapy. Functional neuroimaging studies have shown reduced neuronal density in cortical brain regions involved in emotion and pain regulation in patients with pain disorders and IBS, just as with major depression and PTSD. Therefore, pain and psychological trauma (and especially their combination) can be neurodegenerative, much like Alzheimer or Parkinson disease. It is in these psychological and pain conditions that the use of antidepressants and other centrally targeted agents and methods might offer some remedy by stimulating an increase in the levels of brain-derived neurotrophic factor (BDNF) following treatment. BDNF stimulates neurogenesis, and this is observed with prolonged periods of treatment that correlate with the degree of recovery from depression. Finally, it is observed that the duration of antidepressant treatment correlates with a decreased frequency of relapse and recurrence in depression. These findings provide insight into neuronal growth regulation in key areas of the central pain matrix and provide new and important opportunities for research and patient care using antidepressants for the treatment of FGIDs.

PSYCHOLOGICAL THERAPIES FOR FUNCTIONAL GI DISORDER

Concerning the use of psychological therapies, experience and empirical research suggests that no single particular treatment is superior for treating FGIDs. The most important aspect of treatment is the patient’s acceptance of the need for treatment and their motivation to engage in it. This can be enhanced if the gastroenterologist and psychiatrist (psychologist/therapist) help the patient accept the treatment as a necessary part of an overall plan of care. These range in their depth and ambition, from a limited attempt to control symptoms with simple behavior therapy to hypnosis or insight-oriented psychotherapy. It should be recognized that the availability of the more time-consuming techniques is limited in the public health sector, and their use should therefore be restricted to only the most difficult cases.

Relaxation Therapy

Relaxation therapy is the simplest form of “psychotherapy” that can easily be taught to patients. This is usually accessed on a self-help basis through a number of apps, audio recordings, or classes at local community centers or gyms. It is thought that if stress causes FGIDs, reducing autonomic arousal by relaxation may reduce symptoms and induce a sense of well-being, allowing the patient to feel more confident and greatening their autonomy. Patients are taught to exclude sources of tension and to increase the degree of relaxation. Arousal reduction training includes a variety of different methods to teach patients to counteract the physiologic sequelae of stress or anxiety. The most widely used arousal reduction techniques include (1) progressive muscle relaxation training; (2) biofeedback for striated muscle tension; (3) autogenic training; (4) mindfulness-based stress reduction (MBSR); and (5) Tai-Chi, transcendental, or Yogic meditation. Such therapies have been shown to have an NNT of 6, however, with a broad range (95% CI 3–60). Unfortunately, to date, there are few well-designed controlled studies concerning short- and long-term efficacy.

Cognitive Behavioral Therapy

CBT is based on the assumption that FGIDs in some patients are behavioral diseases generated by responses to life events. In addition, patients who may benefit from this type of treatment may have (a) a high prevalence of anxiety, depression, and assertion difficulties; (b) a high need for social approval; and (c) perfectionistic attitudes. CBT consists of a wide range of strategies and procedures designed to bring about alterations in patients' perceptions of their situation. It facilitates their ability to control GI symptoms by learning new ways of thinking and behaving through personal experience and practice. The treatment essentially involves helping the patient to recognize maladaptive patterns of thinking and behavior, which produces a greater sense of control and autonomy in the patient. It insists that the patient takes some responsibility for the illness and helps them find a more healthy way of dealing with the underlying problem. Empirically, CBT has been demonstrated to produce good results. For instance, one meta-analysis which examined nine RCTs compared to control groups (n = 610) found that in 145 of 349 (41.5%) patients undergoing CBT, the symptoms did not improve, compared to 166 of 261 (63.6%) patients in the control groups. Despite the impact of CBT, access can be problematic, and the expertise of providers varies. Given this, internet-based CBT may be an opportunity for psychiatrists to consider for their patients, especially in light of empirical outcomes from small studies showing promise for such treatments.

One adjunctive therapy to CBT is mindfulness-based cognitive therapy (MBCT), which adds elements of CBT to stress reduction. Studies evaluating outcomes have demonstrated substantial improvements in symptom severity and quality of life as compared to a control group following an 8-week trial of mindfulness-based therapy. Moreover, MBCT has been demonstrated to impact visceral sensitivity.

Interpersonal Therapy

Interpersonal therapy focuses on resolving difficulties in interpersonal relationships that underlie or exacerbate abdominal symptoms in FGIDs. It consists of the identification and mental processing of key problem areas that include unresolved grief or loss, role transitions, and relationship discord. The initial focus is on the patient's abdominal symptoms, which are explored in detail. Emotional distress and abnormal feeling states arising from or linked to physical symptoms are then identified. Key problem areas in relationships and their link to physical and psychological symptoms are understood. This leads to the targeting of maladaptive relationship patterns, which may have developed after key childhood experiences such as trauma or sexual abuse. Solutions to these interpersonal difficulties are first tested out in therapy before subsequently being implemented outside of the therapy environment. Treatment is essentially an exercise in identifying and solving interpersonal problems, which facilitates a greater sense of control and autonomy and therefore reduces the overall degree of sustained arousal in interpersonal difficulties, producing an improvement in symptoms.

Psychodynamic Therapy

Psychodynamic psychotherapy is similar to interpersonal psychotherapy and requires a close relationship between the patient and the therapist, in which difficulties in interpersonal relationships can be highlighted. As the patient understands these problems, they may act upon these insights that may lead to a reduction in symptoms. Analytical psychotherapy attempts to provide the patient with an insight into why particular symptoms have developed and what they might mean or represent in the light of changes in key relationships. There is an implicit assumption that this insight will cause long-lasting changes in attitude and behavior. Symptoms often seem to stem from significant life changes (often a loss of a relationship) set against the background of a fragile personality and difficulty coping with separation. The patient works through their interpersonal difficulties through a safe and contained relationship with the therapist. Psychodynamic psychotherapy has been demonstrated to have an NNT of 4 (95% CI 2–20), and in fact, the world gastroenterology organization has recommended it as a form of therapy for patients.

Hypnotherapy

Hypnosis is used to induce a state of relaxation with heightened suggestibility, in which the hypnotherapist uses progressive muscular relaxation with suggestions of relaxation to reduce striated muscle tension and to relax GI smooth muscle. This then alters the underlying abnormalities of gut motility or sensation and impacts the patient's perception of their symptoms. This is firstly initiated in the presence of the therapist, with the ultimate aim of enabling patients to control symptoms on their own. Patients are asked to practice autohypnosis at home using an audiotape, with the goal of being able to self-administer suggestions of relaxation. Unfortunately, the success of therapy can be dependent on the enthusiasm and skill of the therapist and is both time-consuming and expensive to provide. Studies, however, confirm that long-term efficacy is cost-effective in severe refractory cases and suggest that treatment is more likely to be successful with younger patients and those without serious psychopathology.

Biofeedback Therapy

Biofeedback is a therapeutic approach that depends on feedback as a measure of visceral function to show the patient a disturbance in physiology so that they themselves can learn to correct it. Heart rate variability feedback in relaxation and stress reduction is a growing application in FGIDs, but the most common use of biofeedback is in the treatment of incontinence and constipation. The therapy aims to make the patient more sensitive to rectal sensation and avoid inappropriate straining. It also provides a detailed explanation of normal physiology and a reeducation concerning the optimum defecatory patterns. A supportive relationship that the therapist can offer such patients undoubtedly contributes to the overall positive effect.

CONCLUDING REMARKS

The gut and the brain are intimately linked, and any disorder in one can manifest in the other. Therefore, psychiatry can play a pivotal therapeutic role in the management of patients with GI disorders.

ACKNOWLEDGMENTS

This chapter represents an update and revision of previous chapters in this series, most recently by Drs. Claude Andrew Botha, James K. Ruffle, and Adam D. Farmer from the last edition. We are indebted to Drs. Botha, Ruffle, and Farmer for their work and retain much of their prose in this current edition.

Further Readings

Abreu MT. The genetics and pathogenesis of inflammatory bowel disease. *Gastroenterol Hepatol (NY)*. 2013;9:521–523.

GBD 2017 Inflammatory Bowel Disease Collaborators. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol*. 2020;5(1):17–30.

Bernstein CN, Hitchon CA, Walld R, et al. Increased burden of psychiatric disorders in inflammatory bowel disease. *Inflamm Bowel Dis*. 2019;25(2):360–368.

Canavan C, West J, Card T. Review article: the economic impact of the irritable bowel syndrome. *Aliment Pharmacol Ther*. 2014;40:1023–1034.

Chang L, Lembo A, Sultan S. American Gastroenterological Association Institute Technical Review on the pharmacological management of irritable bowel syndrome. *Gastroenterology*. 2014;147:1149–1172e2.

Drossman DA. Functional gastrointestinal disorders: history, pathophysiology, clinical features, and Rome IV. *Gastroenterology*. 2016;150(6):1262–1279.

Fadgyas-Stanculete M, Buga AM, Popa-Wagner A, Dumitrascu DL. The relationship between irritable bowel syndrome and psychiatric disorders: from molecular changes to clinical manifestations. *J Mol Psychiatry*. 2014;2(1):4.

Fass R, Shibli F, Tawil J. Diagnosis and management of functional chest pain in the Rome IV era. *J Neurogastroenterol Motil*. 2019;25(4):487–498.

Ford AC, Talley NJ, Schoenfeld PS, Quigley EMM, Moayyedi P. Efficacy of antidepressants and psychological therapies in irritable bowel syndrome: systematic review and meta-analysis. *Gut*. 2009;58:367–378.

Frolkis AD, Vallerand IA, Shaheen AA, et al. Depression increases the risk of inflammatory bowel disease, which may be mitigated by the use of antidepressants in the treatment of depression. *Gut*. 2019;68(9):1606–1612.

Hetterich L, Stengel A. Psychotherapeutic interventions in irritable bowel syndrome. *Front Psychiatry*. 2020;11:286.

Keefer L, Drossman DA, Guthrie E, et al. Centrally mediated disorders of gastrointestinal pain. *Gastroenterology*. 2016;150(6):1408–1419.

Keefer L, Kane SV. Considering the bidirectional pathways between depression and IBD: recommendations for comprehensive IBD care. *Gastroenterol Hepatol (NY)*. 2017;13(3):164–169.

Kuiken SD, Tytgat GN, Boeckxstaens GEE. The selective serotonin reuptake inhibitor fluoxetine does not change rectal sensitivity and symptoms in patients with irritable bowel syndrome: a double blind, randomized, placebo-controlled study. *Clin Gastroenterol Hepatol*. 2003;1:219–228.

Mikocka-Walus A, Turnbull D, Moulding N, Wilson I, Andrews JM, Holtmann G. Psychological comorbidity and complexity of gastrointestinal symptoms in clinically diagnosed irritable bowel syndrome patients. *J Gastroenterol Hepatol*. 2008;23:1137–1143.

Miller V, Whorwell PJ. Treatment of inflammatory bowel disease: a role for hypnotherapy? *Int J Clin Exp Hypn*. 2008;56:306–317.

Nadkarni A, Silbersweig DA. Gastroenterological disease in patients with psychiatric disorders. In: Summergrad P, Silbersweig DA, Muskin PR, Querques J. *Textbook of Medical Psychiatry*. American Psychiatric Association Publishing; 2020.

Nahon S, Lahmek P, Durance C, et al. Risk factors of anxiety and depression in inflammatory bowel disease. *Inflamm Bowel Dis*. 2012;18:2086–2091.

Shah E, Kim S, Chong K, Lembo A, Pimentel M. Evaluation of harm in the pharmacotherapy of irritable bowel syndrome. *Am J Med*. 2012;125(4):381–393.

Shivaji N, Ford AC. Prevalence of functional gastrointestinal disorders among consecutive new patient referrals to a gastroenterology clinic. *Frontline Gastroenterol*. 2014;5:266–271.

Simon GE, VonKorff M, Piccinelli M, Fullerton C, Ormel J. An international study of the relation between somatic symptoms and depression. *N Engl J Med*. 1999;341:1329–1335.

Sinagra E, Romano C, Cottone M. Psychopharmacological treatment and psychological interventions in irritable bowel syndrome. *Gastroenterol Res Pract*. 2012;2012:486067.

Sirois F. Steroid psychosis: a review. *Gen Hosp Psychiatry*. 2003;25:27–33.

Spiller R, Aziz Q, Creed F, et al; Clinical Services Committee of The British Society of Gastroenterology. Guidelines on the irritable bowel syndrome: mechanisms and practical management. *Gut*. 2007;56(12):1770–1798.

Szigethy E, Bujoreanu SI, Youk AO, et al. Randomized efficacy trial of two psychotherapies for depression in youth with inflammatory bowel disease. *J Am Acad Child Adolesc Psychiatry*. 2014;53:726–735.

Szigethy EM, Youk AO, Benhayon D, et al. Depression subtypes in pediatric inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2014;58(5):574–581.

Taft TH, Bedell A, Craven MR, Guadagnoli L, Quinton S, Hanauer SB. Initial assessment of post-traumatic stress in a US cohort of inflammatory bowel disease patients. *Inflamm Bowel Dis*. 2019;25(9):1577–1585.

Thorkelson G, Bielefeldt K, Szigethy E. Empirically supported use of psychiatric medications in adolescents and adults with IBD. *Inflamm Bowel Dis*. 2016;22(6):1509–1522.

Tremaine WJ. Diagnosis and treatment of indeterminate colitis. *Gastroenterol Hepatol (NY)*. 2011;7(12):826–828.

Tyler K, Moriceau S, Sullivan RM, Greenwood-van Meerveld B. Long-term colonic hypersensitivity in adult rats induced by neonatal unpredictable vs predictable shock. *Neurogastroenterol Motil*. 2007;19:761–768.

Valet M, Gundel H, Sprenger T, et al. Patients with pain disorder show gray-matter loss in pain-processing structures: a voxel-based morphometric study. *Psychosom Med*. 2009;71:49–56.

Van Langenberg DR, Yelland GW, Robinson SR, Gibson PR. Cognitive impairment in Crohn's disease is associated with systemic inflammation, symptom burden and sleep disturbance. *United European Gastroenterol J*. 2017;5(4):579–587.

Yanartas O, Kani HT, Bicakci E, et al. The effects of psychiatric treatment on depression, anxiety, quality of life, and sexual dysfunction in patients with inflammatory bowel disease. *Neuropsychiatr Dis Treat*. 2016;12:673–683.

Yeh HW, Chien WC, Chung CH, Hu JM, Tzeng NS. Risk of psychiatric disorders in irritable bowel syndrome—a nationwide, population-based, cohort study. *Int J Clin Pract*. 2018;72(7):e13212.

27.4 Psychosomatic Medicine: Obesity

MEGAN E. LAVERY, PSY.D., AND NINOSKA D. PETERSON, PH.D.

DEFINITION, NOSOLOGY, AND EPIDEMIOLOGY

Obesity is a global epidemic with physical, economic, and psychological costs. The prevalence of obesity in the United States continues to rise, with 42.4% of U.S. adults meeting the criteria for obesity according to the 2017 to 2018 National Health and Nutrition Examination Survey (NHANES) (Table 27.4–1). In the United States, approximately 18.5% of children ages 2 to 19 have obesity. In both adults and children, there is a higher rate of obesity among minority groups, including Hispanics and African Americans. Rates of obesity have shown an alarming trend upward since the 1980s, with rates in the United States leveling off over the past decade. However, unlike the flattening of the obesity curve overall, severe obesity continues to increase.

Obesity is defined primarily by the body mass index (BMI). BMI is measured as weight in kilograms divided by height in meters squared (kg/m^2). A normal BMI is 18.5 to 24.5, and individuals are considered overweight with BMIs from 25 to 30. BMIs above 30 are in the obesity range, with a BMI above 40 considered to have severe obesity (Table 27.4–2).

Obesity raises the risk of morbidity from hypertension, dyslipidemia, type 2 diabetes, coronary heart disease (CHD), stroke, gallbladder disease, osteoarthritis, sleep apnea and respiratory problems, and some cancers. Obesity is also associated with increased risk in all-cause and cardiovascular disease mortality. In addition, obesity has a negative impact on quality of life, mobility, mental health, and stress level.

The economic costs of obesity and its associated health problems place a heavy burden on the US healthcare system. The costs of preventing, diagnosing, and treating obesity are directly observable, whereas the indirect costs may include loss of income, decreased productivity, restricted activity, and absenteeism. Costs may also include the value of future income lost by premature death. According to the Centers for Disease Control, the estimated medical cost of obesity in the United States is $147 billion yearly in 2008 U.S. dollars.

DIAGNOSIS, CLINICAL FEATURES

BMI, as described above, has been the most widely used measure of obesity. While it is useful for population-level measurements and has utility as an estimate of health risk, BMI has also been criticized for failing to account for bone density, muscle mass versus fat mass, fitness, and other biometric measures of health. Another indicator of obesity is the measurement of waist circumference. Waist circumference greater than 35 inches for women or greater than 40 inches for men is considered obesity. High-waist circumference, or a hip-to-waist ratio, may elevate risks for obesity-related disease as well. Obesity was recognized by the American Medical Association as a disease in 2013. While obesity is not included as a diagnosis in the Diagnostic and Statistical Manual-5 (DSM-5), binge eating disorder (BED) has been formally accepted as a diagnosis, as discussed later.

Obesity in childhood is diagnosed using an age- and sex-specific percentile for BMI due to the variations in body composition with age and gender during childhood. Children with a BMI between the 85th and 95th percentile for children of the same age and gender are considered overweight, and children with a BMI above the 95th percentile are considered to have obesity.

Table 27.4–1.
Prevalence of Obesity and Severe Obesity Among U.S. Adults, Age 20 Years and Over, NHANES 2017–2018

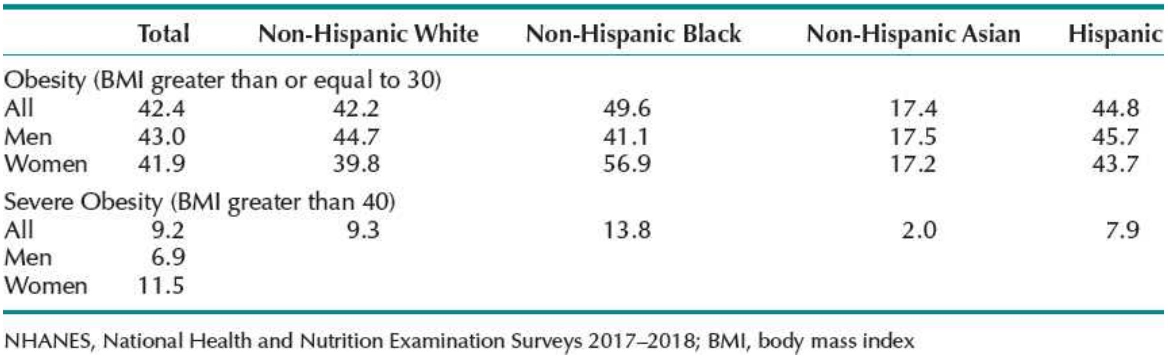

	Total	Non-Hispanic White	Non-Hispanic Black	Non-Hispanic Asian	Hispanic
Obesity (BMI greater than or equal to 30)					
All	42.4	42.2	49.6	17.4	44.8
Men	43.0	44.7	41.1	17.5	45.7
Women	41.9	39.8	56.9	17.2	43.7
Severe Obesity (BMI greater than 40)					
All	9.2	9.3	13.8	2.0	7.9
Men	6.9				
Women	11.5				

NHANES, National Health and Nutrition Examination Surveys 2017–2018; BMI, body mass index

Table 27.4–2.
Classification of Obesity and Recommended Treatment

	Obesity Class	Body Mass Index (kg/m^2)	Recommended Treatment
Underweight		<18.5	Nutrition/eating disorder treatment
Normal		18.5–24.9	Maintenance
Overweight		25–29.9	BWM
Obese	I Mild	30–34.9	BWM + MWM
	II Moderate	35–39.9 (plus comorbidities)	BWM + MWM or BWM + WLS
	III Severe	>40	BWM + WLS

BWM, behavioral weight management; MWM, medical weight management; WLS, weight loss surgery.

COURSE AND PROGNOSIS

Childhood obesity is a risk factor for adult obesity and is related to the severity of adult obesity. Children with obesity are more likely to develop cardiovascular risk factors such as high-blood pressure and high-cholesterol; impaired glucose tolerance, insulin resistance, and type 2 diabetes; breathing problems such as sleep apnea and asthma; joint and pain problems; fatty liver disease, gallstones, and gastroesophageal reflux; and social and psychological problems, such as stigma and poor self-esteem (Fig. 27.4–1).

In adults, associated medical comorbidities of obesity include congestive heart failure, coronary artery disease, venous state ulcers, insulin resistance, polycystic ovarian syndrome, diabetes mellitus, hypertension, abdominal hernia, gall stones, gastroesophageal reflux disease, nonalcoholic fatty liver disease, stress urinary incontinence, deep venous thrombosis, pulmonary embolism, gout, plantar fasciitis, depression, anxiety, pseudotumor cerebri, stroke, infertility, asthma, obstructive sleep apnea, hypercholesterolemia, osteoarthritis, cancers, and others (Fig. 27.4–2). The risk of mortality and morbidity increases with a BMI above 35 and rises exponentially with the severity of BMI. Together, obesity and overweight are the second leading cause of preventable death in the United States. An estimated 300,000 deaths per year are due to the obesity epidemic.

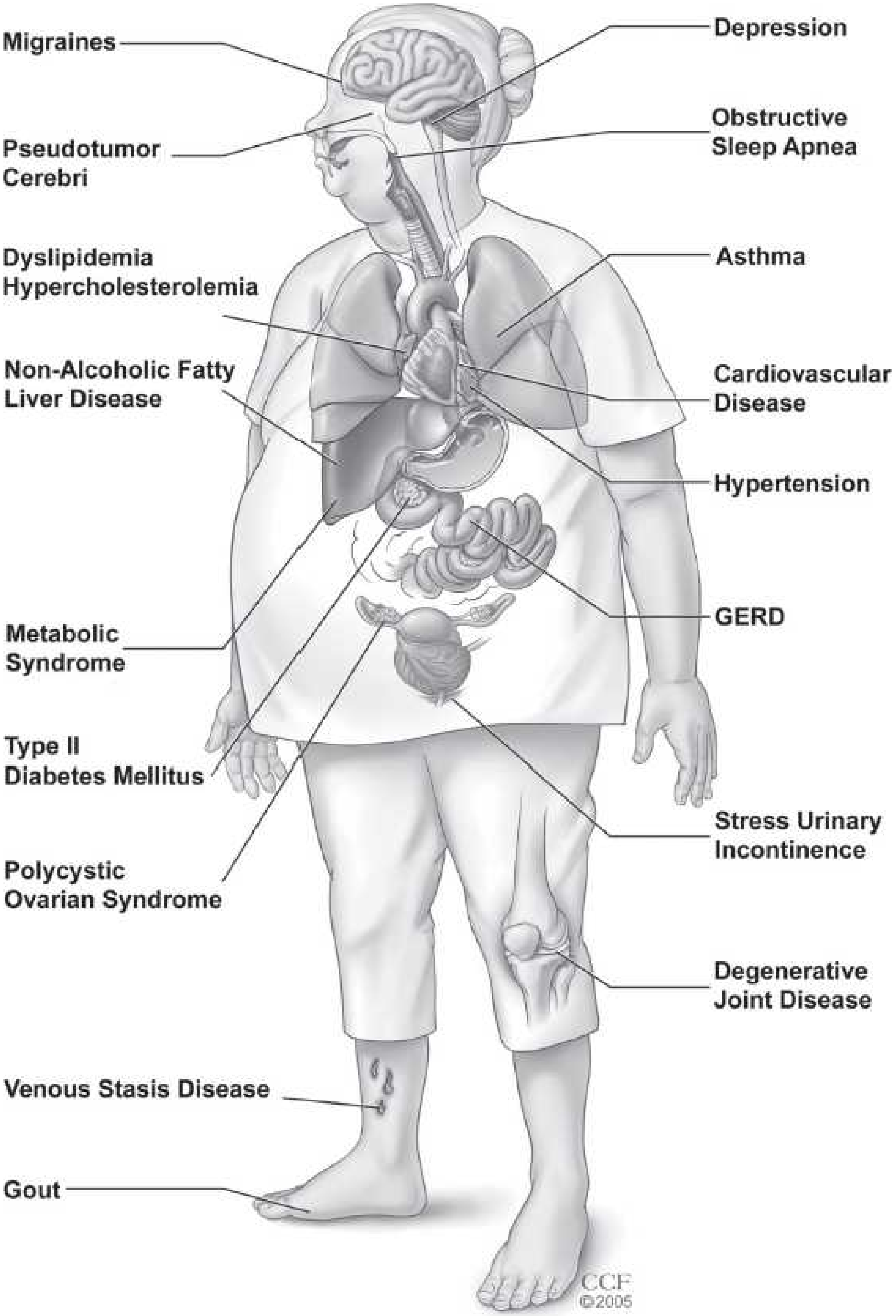
Migraines
Pseudotumor Cerebri
Dyslipidemia Hypercholesterolemia
Non-Alcoholic Fatty Liver Disease
Metabolic Syndrome
Type II Diabetes Mellitus
Polycystic Ovarian Syndrome
Venous Stasis Disease
Gout
Depression
Obstructive Sleep Apnea
Asthma
Cardiovascular Disease
Hypertension
GERD
Stress Urinary Incontinence
Degenerative Joint Disease
CCF
©2005

FIGURE 27.4–1. Medical comorbidities of obesity in adults. (Reprinted with permission, Cleveland Clinic Center for Medical Art & Photography © 2006–2015. All Rights Reserved.)

Behavioral and lifestyle interventions can effectively result in modest weight loss, which has a positive effect on comorbidities. However, the rate of recidivism is high, and weight loss is difficult to maintain for a variety of reasons. For patients with severe obesity, weight loss surgery appears to be an effective long-term intervention that results in significant weight loss and resolution of comorbidities.

PATHOLOGY AND LABORATORY CONSIDERATIONS

When working with patients with obesity, clinicians may want to consider the following adjunctive information: routine blood work for hyperlipidemia, hypertension, hypothyroid function, and hyperglucosemia; sleep studies to identify obstructive sleep apnea, and vitamin level panel to identify vitamin deficiencies.

ETIOLOGY

Fundamentally, obesity is the result of an imbalance between energy intake and expenditure. However, the development of obesity is far more complex than this simplified view portrays. Given the alarming rise in obesity rates, researchers have increasingly sought to elucidate the multifaceted causes of obesity in hopes of finding strategies to combat this epidemic.

Genetics, Biology, and Obesity

From an evolutionary perspective, the ability to store energy as adipose tissue would have yielded a selective advantage for survival and reproduction. Using this concept, Neel (1962) proposed the notion of thrifty genotypes. This theory hypothesizes that thrifty genotypes code for effective energy storage. Through natural selection, genes with the ability to save and store energy likely overrode other genes, resulting in many thrifty genes undergoing such selection. In fact, researchers have found over 600 single-gene mutations and genomic and chromosomal regions potentially related to obesity. Thus, due to natural selection, many individuals may be genetically susceptible to developing obesity.

In addition, twin and adoption studies have underscored the role of genes in obesity. This body of literature suggests that heritable genetic differences account for 55% to 85% of weight variance. Interestingly, several twin and adoption studies also highlight how most of the nongenetic factors influencing weight may stem from nonshared environmental contexts unique to a given person (e.g., school environment and peer relationships). Contrary to common assumptions, the evidence suggests that the shared family environment may have a minimal effect on BMI. Therefore, genes and individual nonshared experiences may have a more significant impact on weight than the family environment.

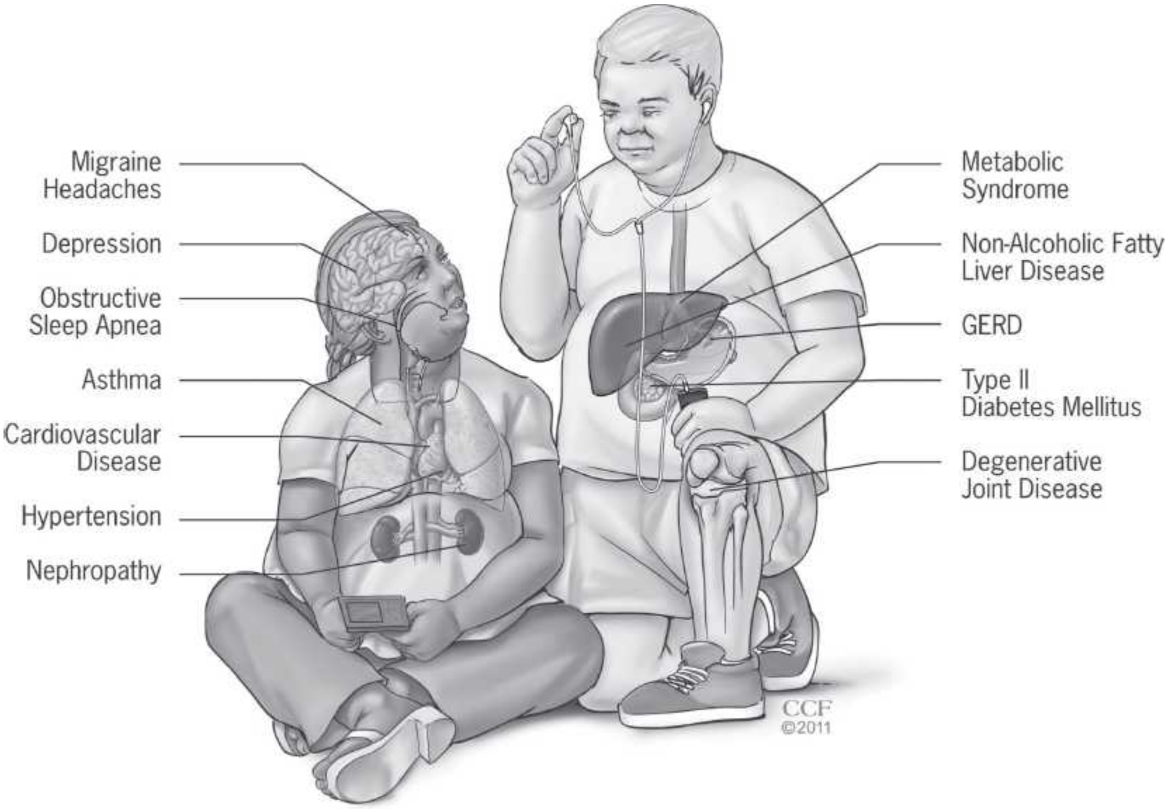
Migraine Headaches
Depression
Obstructive Sleep Apnea
Asthma
Cardiovascular Disease
Hypertension
Nephropathy
Metabolic Syndrome
Non-Alcoholic Fatty Liver Disease
GERD
Type II Diabetes Mellitus
Degenerative Joint Disease
CCF
©2011

FIGURE 27.4–2. Medical comorbidities of obesity in adolescents. (Reprinted with permission, Cleveland Clinic Center for Medical Art & Photography © 2006–2015. All Rights Reserved.)

Increasingly, adipose tissue and hormones have also been implicated as biologic contributors to obesity. More specifically, there is growing evidence that enlarged adipocytes in individuals with obesity create inflammation and result in the dysregulation of metabolic homeostasis, impacting levels of leptin, insulin, and adiponectin, thus creating metabolic syndrome. Researchers have just begun to understand how such changes can impact weight, as well as exercise, food intake, and overall hormonal control. In particular, the role of the gut microbiome in weight gain and weight loss has been the target of increasing investigations. A body of evidence suggests that the gastrointestinal microbiota can impact dietary energy utilization and influences genes that regulate energy expenditure and storage.

Environment and Obesity

However, genetics in and of themselves do not explain the rising rates of obesity globally. A large body of research has underscored the importance of global environmental influences on eating behavior. For most of history, humans have struggled to combat food shortage, disease, and hostile surroundings. However, the modern industrialized environment stands in stark contrast to this type of setting. Instead, the current environment is overflowing with calorically dense foods, along with energy-saving devices. To capture these changes, Swinburn developed the term “obesogenic environment.” Researchers contend that the environments of most industrialized nations promote a positive energy balance through the availability and affordability of highly palatable, energy-dense foods, as well as numerous opportunities for inactivity (e.g., car ownership, television, sedentary job environments, and the supermarket system). In the current environment, expending an ample amount of energy typically requires individuals to engage in deliberate, consistent healthy lifestyle choices. Unfortunately, making a purposeful choice to be active is at odds with humans’ biologic propensity to conserve energy. Therefore, humans’ biologic predisposition makes them extremely vulnerable to overeating and inactivity, particularly in an environment with endless food and inactivity cues. The obesogenic environment has thus created an ideal setting for the growth of obesity. Many researchers and lobbyists have argued for national and global food reform, including taxation on sugary foods and drinks, as a primary means of combating obesity.

Food Addiction and Obesity

Despite a common food environment, research indicates that the susceptibility to food cues varies across individuals and serves as a predictor of BMI, the ability to resist cravings, and dietary compliance. Within this framework, researchers contend that food can become an addiction akin to other substance use disorders. More specifically, studies suggest that recurrent exposures to highly palatable foods result in diminished dopamine levels in response to eating. As a result, individuals must consume larger amounts of food to feel satisfied, resulting in a cycle of overeating. The term *food addiction* (FA) has been used to capture the resulting pattern of excessive consumption. It is thought that FA to highly palatable, processed foods could be a key contributor to the development of obesity. Therefore, the availability of calorically dense foods, combined with a potential for physiologic addiction to these foods, may play a central role in the increasing rates of obesity.

The Yale Food Addiction Scale (YFAS) was specifically developed to measure FA by applying DSM-IV criteria for substance dependence to eating behaviors. A 2014 systematic metanalysis examined studies using the YFAST to assess for FA. The researchers found that the weighted mean prevalence of FA was 19.9% across adult samples. Higher BMI was associated with higher rates of FA diagnosis and number of symptoms. In fact, rates of FA were doubled in populations with overweight or obesity compared to healthy controls (24.9% vs. 11.1%, respectively). Additionally, in disordered eating samples, the mean prevalence of FA was extremely high (57.6%) compared to samples with no eating disorder diagnoses (16.2%). These results highlight the relationships between FA, weight, and disordered eating.

Behaviors, Cognitions, and Obesity

Many individuals report eating in response to negative emotions. Certain internal experiences, such as anxiety, sadness, negative thoughts, and cravings, are frequently felt to be unpleasant. Individuals vary in their ability to tolerate such internal distress. Some individuals may make a great effort to escape these experiences via psychological (i.e., suppression) or behavioral (i.e., isolation) avoidance. It is common for individuals to use food as a means of temporarily suppressing emotional distress. In fact, a significant number of patients report regularly eating in response to stress, boredom, fatigue, anxiety, and depression. For many, this pattern may begin in childhood and sustain throughout adulthood. Eating can become a learned response to emotional distress and, for some, may become a means of coping. Research suggests emotional eating is positively associated with weight and BMI. Thus, emotional eating can develop into a primary maladaptive affect regulation tool, resulting in weight gain.

Dimensions of Personality and Obesity

There is also a growing body of research focusing on the influence of personality traits on weight. In particular, many studies suggest that neuroticism (manifested in greater anxiety, depression, impulsivity, emotional vulnerability, hostility, and anger) is a risk factor for overweight or obesity, particularly for women. It is hypothesized that individuals with higher levels of neuroticism may be more susceptible to using food as a dysfunctional coping strategy. In addition, the literature indicates that for patients with obesity, there are higher rates of impulsivity with and without disordered eating (unhealthy eating patterns often characterized by distress and/or loss of control (see Tables 27.4–3 and 27.4–4). Similarly, high-reward sensitivity appears to be another risk factor for obesity. Researchers postulate that individuals with higher sensitivity to reward have a proclivity toward uncontrolled eating, resulting in sustained positive energy balance. However, higher levels of conscientiousness (regulation of internal urges) and self-control have been shown to be protective in regard to weight gain. Given this literature, it may be important to consider personality traits when working with patients struggling with obesity.

Table 27.4–3.
DSM-5 Diagnostic Criteria for Binge Eating Disorder

Recurrent episodes of binge eating (eating a larger than normal amount with loss of control at least once per week for 3 months) accompanied by at least three of the following:

- Eating very rapidly
- Eating until uncomfortably full
- Eating when not hungry
- Eating alone or hiding eating from others
- Feeling guilty or disgusted after a binge

No recurrent compensatory purging, exercising, or fasting

Marked distress

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 392, 393 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Table 27.4–4.
Proposed Criteria for Night Eating Syndrome (NES)

1. **Core criteria (One of the following):**
   A. Evening hyperphagia and/or
   B. Nocturnal awakenings with food ingestion (at least 2 × /week)
2. **Three of five associated symptoms:**
   A. Insomnia: Sleep initiation and maintenance insomnia
   B. Morning anorexia: Either in the form of lack of desire to eat in the morning or skipping breakfast more days than not
   C. Urges to eat at night
   D. Depressed evening mood
   E. A belief that one must eat to sleep

Symptoms must be associated with distress and occur for 3 or more months

Data from Allison KC, Lundgren JD, O'Reardon JP, et al. Proposed diagnostic criteria for night eating syndrome. *Int J Eat Disord*. 2010;43(3):241–247.

PSYCHIATRIC COMORBIDITIES OF OBESITY

Eating Disorders

Eating disorders are common psychiatric comorbidities of obesity and include BED, bulimia, and syndromes such as night eating, graze eating, and loss of control consumption of high-calorie liquids.

BED has been included in the DSM-5 as a stand-alone disorder and was previously included in the DSM-IV under the category Eating Disorder Not Otherwise Specified (Table 27.4–3). BED is characterized by eating large amounts in a short period of time with a feeling of loss of control. Associated symptoms include feelings of guilt, eating quickly, eating when not hungry, hiding eating patterns from others, and eating until uncomfortably full. The binge episodes must be distressing and occur at least once per week for 3 months to meet the criteria for BED. The prevalence of BED among patients with obesity is estimated to be 30%, and patients with BED tend to have obesity. BED appears to be the most common eating disorder, with estimated population rates of 2%.

Other eating syndromes sometimes associated with obesity include graze eating and night eating syndrome (NES). Graze eating syndrome is less well-defined than NES and typically involves eating large amounts slowly over the day (e.g., nibbling or picking) or continuous eating. Graze eating is often associated with distress and eating for reasons other than hunger. However, unlike BED, there are limited consensuses around whether graze eating is characterized by feelings of loss of control. In fact, a 2016 study examining bariatric surgery candidates who graze eat found that a majority of patients who reported graze eating did not experience loss of control, instead endorsing a greater degree of overall distress and psychopathology. The minority of patients evidencing graze eating with loss of control reported an even greater risk of psychopathology. Further, a high-calorie liquid type of grazing has also been identified as a pattern of drinking high-calorie liquids continuously throughout the day. In addition, NES is a relatively understudied disorder involving a circadian delay in food consumption. According to the 2010 proposed criteria, the diagnostic criterion for NES requires the presence of evening hyperphagia or nocturnal awakenings with food ingestion. Additional symptoms include morning anorexia, urges to eat at night, insomnia, low evening mood, and a belief that one must eat to sleep (Table 27.4–4). There appears to be a relationship between NES and weight in some, but not all, studies. Within the research, NES has been associated with several correlates of distress, including depressed mood, anxiety, low self-esteem, poorer body image, and decreased quality of life. Not surprisingly, there are higher rates of mood and anxiety disorders in NES populations, including current and lifetime major depression. It is also important to note that more distress has been associated with patients who wake up and eat in the middle of the night versus patients with only evening hyperphagia. Given these findings, some consider patients with nocturnal eating episodes to be more severe than patients without nocturnal ingestions.

Bulimia nervosa is characterized by binge episodes followed by compensatory behavior such as vomiting, laxative use, or overexercise. Binge/purge episodes must occur at least once per week for 3 months to meet the criteria for the disorder, consistent with BED. Patients with bulimia may have obesity but are more typically in the normal BMI weight range.

Depression

Depression and obesity often overlap as disease processes. This is perhaps unsurprising as increased appetite and low-energy are among the diagnostic symptoms of depression (DSM-5). There is a high-incidence of depression among patients with obesity, with estimates ranging from 19% to 66%. The relationship between obesity and depression appears to be reciprocal, with 80% of studies suggesting obesity contributes to depression and 53% of studies suggesting that depression leads to obesity. Obesity may lead to depression by negatively affecting health and body image, increasing functional impairment, decreasing quality of life and physical activity, and adding to the stigma. Depression may lead to obesity by contributing to an elevated stress reaction, impairing immunity, negative cognitions, eating behaviors, and poor adherence to treatment. Factors moderating the relationship between depression and obesity include gender, age, socioeconomic status, and severity of obesity. Women, those with increasing age or those with lower income, are more likely to experience depression with obesity. Other factors that increase the likelihood of depression with obesity include binge eating behaviors and experiencing stigma.

Anxiety

Anxiety disorders are the most common psychological disorders affecting 25% of the population. Obesity may lead to anxiety by causing distress from weight-related stigma. In addition, the physical comorbidities and negative effects of obesity on quality of life may lead to stress and cause anxiety disorders. However, the prospective evidence for obesity causing anxiety is weak. A review of the literature suggests there is a moderate amount of evidence for a weak but positive association between anxiety and obesity. There appears to be a stronger relationship between severe obesity and anxiety. Among anxiety disorders, specific phobia, social phobia may have the strongest associations with obesity. Although no longer in the anxiety category, post-traumatic stress disorder may be as well. Anxiety disorders may contribute to obesity through the hypothalamic pituitary adrenal axis dysfunction, and anxiety may increase cravings for high-sugar and high-fat foods. Anxiety may also be related to obesity through negative childhood events such as abuse and eating disorders, with patients developing anxiety in response to trauma and those with anxiety having a higher risk of eating disorder behaviors.

Severe Mental Illness

Patients with severe mental illness (SMI), such as schizophrenia and bipolar disorder, are at increased risk of obesity, with identified rates up to 60%. Although weight gain side effects of psychotropic medications appear to play a major role, there is some evidence of being overweight among medication naïve patients. Patients with SMI also have an increased risk of medical conditions related to obesity, including type 2 diabetes, cardiovascular disease, and metabolic syndrome. Individuals with schizophrenia and bipolar disorder have a shorter lifespan and an increased risk of death from medical causes. Weight gain side effects of psychotropic medications are a cause for obesity in persons with SMI, with atypical antipsychotics associated with more weight gain than typical agents, and patients treated with a combination of atypicals and mood stabilizers showing the highest weight gain. In persons with SMI, binge eating behaviors and female gender are also associated with higher rates of obesity. Clinicians should closely monitor weight in patients with SMI and include regular measurement of BMI, waist circumference, lipid profile, and plasma glucose level. Consideration of adjunctive therapy such as cognitive-behavioral weight management and diet/exercise/lifestyle changes can mitigate weight gain effects in persons with SMI. Other options include choosing the medication that best treats the patient's mental disorder with the least weight gain profile, switching antipsychotics to promote weight loss in patients who experience significant weight gain, and considering adjunctive treatment with medications that promote weight loss.

Psychiatric Medications and Weight Gain. Weight gain may contribute to nonadherence to psychotropic medications, and the majority of antidepressants, antipsychotics, and mood stabilizers contribute to weight gain (Table 27.4–5). Antipsychotics appear to have some of the highest rates of weight gain, including olanzapine (2.3 kg/month), quetiapine (1.8 kg/month), and clozapine (1.7 kg/month).

Table 27.4–5.
Psychotropic Medications and Weight Gain

Psychotropic Medication	Weight Gain Profile
Antipsychotics	
Olanzapine (Zyprexa)	+++
Quetiapine (Seroquel)	+++
Risperidone (Risperdal)	++
Ziprasidone (Geodon)	+
Asenapine (Saphris)	+
Iloperidone (Fanapt)	++
Paliperidone (Invega)	0
Aripiprazole (Abilify)	+
Clozapine (Clozaril)	+++
Antidepressants	
Imipramine (Tofranil)	+
Amitriptyline (Elavil)	+
Trazodone (Oleptro)	+
Fluoxetine (Prozac)	0/+
Citalopram (Celexa)	+
Fluvoxamine (Luvox)	0
Duloxetine (Cymbalta)	+
Escitalopram (Lexapro)	+
Bupropion (Wellbutrin)	–
Mirtazapine (Remeron)	+
Paroxetine (Paxil)	++
Sertraline (Zoloft)	+
Venlafaxine (Effexor)	0
Mood Stabilizers/Anticonvulsants	
Divalproex (Depakote)	++
Lamotrigine (Lamictal)	0
Carbamazepine (Tegretol)	++

Lithium (Lithobid)	++
Zonisamide (Zonegran)	–
Topiramate (Topamax)	–

–, Weight loss; 0, Weight neutral; +, Weight gain (mild); ++, Weight gain (moderate); +++, Weight gain (significant).

Weight gain due to antidepressants appears to be under-recognized yet an important consideration, given the high prescribing volume of these medications. Tricyclic antidepressants (TCAs) and monoamine oxidase inhibitors (MAOIs) appear to have higher rates of weight gain than serotonin selective reuptake inhibitors (SSRIs). Weight gain is associated with mirtazapine, whereas bupropion may be the only antidepressant agent that contributes to weight loss. On average, antidepressants result in a 1 to 3 kg weight gain. The short-term weight loss seen with fluoxetine appears to be transient, with later weight gain offsetting this effect. Weight gain is a common side effect of mood stabilizers, with divalproex, carbamazepine, and lithium appearing to cause the most gain (0.9 to 15 kg). Lamotrigine appears to be more weight-neutral, and zonisamide may promote weight loss.

The causes of weight gain with psychotropic use may include increased calorie intake, slowing of the metabolism, and increased appetite. Patients should be alerted to weight gain as a potential side effect and closely monitored with routine blood work for hyperlipidemia, hypertension, and hyperglucosemia. Patients being prescribed psychotropics with weight gain side effects should be given recommendations simultaneously for diet and exercise as a preventative measure. Careful consideration of the weight gain profile of medication should be given, especially in patients already with increased BMI and comorbidities. In cases where the benefits of the medication on the psychiatric symptoms appear to outweigh the risks of weight gain, consideration of adjunctive weight loss medications may be considered. Of note, there is an emerging literature base looking at the effects of metformin in mitigating antipsychotic-induced metabolic dysfunction via changes in the gut-brain axis.

Substance Use Disorders

Researchers have found similarities in the neurobiology between obesity and substance use disorders, suggesting that these common paths might also have treatment indications. Some conceptualize obesity as an FA with compulsive overeating, as previously discussed. Drugs and addictive foods may activate similar brain regions involved in reward processing and decision-making.

There appears to be a positive relationship between nicotine dependence and obesity. Epidemiologic studies also point to a relationship between alcohol use disorders and obesity, although no consistent relationships between cocaine/stimulant use and obesity have been found. The cooccurrence of substance use disorders and obesity suggests that treatments such as cognitive behavioral therapy (CBT), diet/exercise, and medications targeting underlying biologic mechanisms in the reward circuitry of the brain may be useful for both conditions. Considering that substance use remission often results in weight gain, awareness of weight management in treating substance use disorders is important.

Table 27.4–6.
Types of Diets

Type of Diet	Calorie Deficit	Weight Loss	Summary
Low-calorie diet (LCD)	800–1,500 cals/day	0.5–1 kg/wk	Carbohydrates 55% Protein 15% Fat <30%
Very low-calorie diet (VLCD)	500–800 cals/day	1.5–2.5 kg/wk	70–100 g of protein/day via liquid meal replacements
South Beach diet/Atkins/Zone diet	<30 g of carbohydrate/day	Rapid initial weight loss followed by steady weight loss	Low carbohydrate Some promote high fiber, lean meats, and low-glycemic carbohydrates
Weight Watchers	Promotes balanced food choices via a points system based on the amount of fats, carbohydrates, protein, and fiber in foods.	1–2 lb/wk	Mod fat, balanced nutrition Focus on activity Group/online support
Jenny Craig/Nutrisystem	Prepackaged low-calorie food	1–2 lb/wk	Meal replacements (1,200–1,500 calories/day) with gradual return to grocery foods in maintenance phase Individual consultant

and/or online support

TREATMENT

Treatments for obesity are tailored to the risk level and individual needs of the patient (Table 27.4–2). The basis for all weight management treatment is behavioral lifestyle intervention, including behavioral modification, diet, and exercise. Patients in the overweight range (BMI between 25 and 30) are encouraged to engage in behavioral weight management. Patients with mild obesity (BMI >30) are directed to add medical weight management, and patients with severe obesity (BMI >40) are candidates for weight loss surgery.

Benefits of Weight Loss

A large body of literature calls attention to the physical and psychological benefits of even small weight reductions. For instance, a weight loss of 5% to 10% of initial body weight has demonstrated clinically significant improvements across physical domains, including blood pressure and lipid levels. Similarly, a 3% decrease in weight has been associated with decreased metabolic risk factors. Weight loss of only 1 to 9 lb has even been correlated with decreased mortality in individuals who are overweight and have diabetes. In addition, studies have consistently shown a positive relationship between weight loss efforts and psychological variables, including body image, health-related quality of life, self-esteem, and depression. Arguably, this body of evidence supports the concept that a weight loss goal is never too small to result in benefit.

Commercial Diets

Despite the complex etiology of obesity, individual lifestyle and behavioral change can play a critical role in addressing factors influencing weight. Many individuals may initially attempt weight loss through commercial diets. For individuals who struggle with their weight, there are numerous commercial diets available with various food and activity goals (see Table 27.4–6 for details and comparisons of diets). Significant weight loss has been observed with many low-carbohydrate or low-fat diets. Generally, weight loss differences between diet types are minimal. Studies suggest that behavioral support and exercise may play key roles in weight loss differences versus the diet regimen itself. Similarly, in comparing several popular diet programs (Atkins, Weight Watchers [WW], Ornish, and Zone diet), research indicates that adherence is the biggest predictor of weight loss, regardless of the type of diet. Therefore, adherence appears to be the critical factor in positive outcomes. Unfortunately, due to gradual nonadherence, it is not uncommon for patients who have attempted numerous weight loss programs to be at their starting weight or higher.

Behavioral Weight Loss

Although some individuals may experience success with commercial diets, many patients may need additional specialized support. In considering psychological treatment options for obesity, numerous studies demonstrate the efficacy of lifestyle and behavioral weight loss (BWL) interventions. In fact, based on a systematic review of the literature, the updated 2013 Guidelines for Managing Overweight and Obesity in Adults encourage clinicians to offer a rigorous, multicomponent behavioral intervention to patients with a BMI >30 or >25 with at least one comorbidity or risk factor. The basic tenets of BWL stem from Learning Theory, which assumes that behavior can be altered by modifying the context in which the behavior occurs. Thus, treatment aims to reduce maladaptive behaviors, such as the consumption of calorically dense food and sedentary behaviors, while reinforcing healthy alternatives (e.g., the selection of healthy food choices and implementation of exercise). Through psychoeducation and skill-based interventions, such treatments encourage patients to make permanent changes to their food intake, activity level, and environment. In an effort to promote new, enduring habits, patients are encouraged to take a balanced approach to food choices and exercise. For example, primary dietary goals often included reducing calorie and saturated fat intake while increasing the consumption of complex carbs, vegetables, and other nutritionally dense foods. Thus, BWL supports flexibility over rigid adherence.

In addition to general dietary and activity goals, BWL treatment focuses on learning behavioral strategies to facilitate weight loss. For example, self-monitoring and goal setting are incorporated into treatment to help patients observe their calorie targets and exercise objectives. It can be particularly important for clinicians to help patients set short-term, reasonable behavioral goals that can be monitored over time. In addition, BWL protocols typically focus on teaching stimulus control techniques; modifying cues in the environment to increase the potential that a healthy behavior will occur. For example, to decrease the likelihood of an overeating episode, individuals may limit their access to trigger foods by purchasing a cone of ice cream versus a gallon container. Similarly, to increase the probability of being active, individuals may keep their workout clothes in their car so that they go straight to the gym after work before returning home. Behavioral strategies, such as stimulus control, have consistently been shown to be an effective component for decreasing sedentary behavior in the context of weight loss.

Commonly, BWL treatment is disseminated in a group format with 4 to 6 months of weekly sessions. Due to the skill-based nature of the treatment, information is typically delivered in a psychoeducational format with topics presented as stand-alone modules. The LEARN (Lifestyle, Exercise, Attitudes, Relationships, and Nutrition) Program is considered the gold standard for BWL group-based treatment. The protocol has been utilized in numerous clinical and research settings. Similar to standard BWL programs, LEARN encourages patients to implement gradual lifestyle changes. Following the federal nutritional and activity guidelines, the program focuses on moderate caloric restriction and a gradual increase in physical activity. The program is structured in a series of additive steps with some repetition to facilitate mastery.

On average, BWL interventions produce a 5% to 10% weight reduction over a 6-month period. An evaluation of the literature conducted by the American College of Cardiology, the American Heart Association in Dallas, Texas, and the Obesity Society found that comprehensive behavioral interventions involving a modification of diet, increased physical activity, and behavioral therapy skills resulted in an average weight loss of 8 kg; with most patients’ weight stabilizing after 6 months of treatment. Despite these typical outcomes, both clinical experience and research suggest that many individuals with overweight and obesity seeking weight loss begin treatment with goals that are two to three times greater than the expected weight loss range. While some research suggests no association or a positive relationship between initial weight loss goals and results, several studies have found that greater weight loss expectations are related to poorer weight loss outcomes. Thus, it may be important to discuss realistic expectations with patients prior to beginning BWL.

Weight Loss Maintenance

Although many patients can successfully lose weight during the initial phase of treatment, obesity treatments demonstrate limited effectiveness in regard to long-term weight loss maintenance. Unfortunately, a large body of literature suggests that following the completion of weight loss programs, most individuals regain a majority of their weight 3 to 5 years after treatment. In fact, it is estimated that less than 5% of patients who lose weight will have maintained their losses after 4 to 5 years. While some research suggests that metabolic changes may account for a small percentage of weight regain, the disappointing long-term results of standard behavior treatment are often attributed to decreased compliance with behavioral, dietary, and physical activity guidelines. In fact, it is posited that at the time of greatest weight loss, many individuals have already displayed marked deviations from their initial healthy behavior changes. Thus, the trend for individuals to regain weight underscores how difficult it can be to alter habits long-term.

In understanding the difficulties with sustained weight loss, it can be helpful to understand how some individuals are able to achieve weight loss maintenance. Across studies, when comparing individuals who have maintained weight loss long-term to individuals who have regained weight, maintainers typically report lower dietary fat consumption, higher levels of physical activity (specifically vigorous activity), increased use of general behavioral strategies, and more frequent self-weighing. Taken together, this body of literature indicates that sustained behavioral changes are associated with more long-term positive weight outcomes. Moreover, the National Weight Control Registry is a particularly useful database of individuals followed prospectively over time who have on average lost more than 65 lb and sustained their weight losses for more than 5.5 years. One study used this registry and found that several factors predicted weight regain versus maintenance over a one-year period. Risk factors for weight gain included a larger initial weight loss, heavier maximum weight, more significant history of weight cycling, shorter period of maintaining current weight, greater depressive symptoms, more disinhibition, and a greater desire to continue to lose weight. These results indicate that even among individuals who have lost a significant amount of weight and kept the weight off for a period of time, there are still several risk factors that make some individuals more vulnerable to regaining weight.

Numerous studies have shown that dietary self-monitoring is associated with both greater initial weight loss and sustained maintenance of lost weight. Unfortunately, akin to other health behaviors, compliance with self-monitoring often decreases over time. Given this trend, one study sought to identify what aspects of self-monitoring are essential for long-term success. Researchers examined the effects of frequency (total number of food records), consistency (number of weeks with >2 records), and comprehensiveness of self-monitoring on weight loss maintenance in 220 women at 12 months into a weight loss journey. Results revealed that individuals who continued to self-monitor frequently and consistently at 12 months sustained weight losses more effectively. At the same time, the findings demonstrated that increased comprehensiveness of monitoring did not result in any added benefit above and beyond the effect of frequency. These findings indicate that continuing to complete food diaries frequently and consistently may help individuals to sustain a level of awareness and adherence, though the detail of self-monitoring appears to have less of an impact on outcomes.

Psychological Treatment of Obesity

Cognitive Behavioral Therapy for Weight Loss. In addition to general lifestyle and behavioral treatments, CBT is an effective, empirically supported psychological intervention for weight management in both youth and adults. CBT for obesity typically includes psychoeducation (especially around dietary guidelines and physical activity), behavioral skills (e.g., self-monitoring and stimulus control), and cognitive strategies. CBT can be utilized as a stand-alone treatment or can be used to improve behavioral lifestyle treatment components. Many of the tenets and skills taught in CBT treatment overlap with standard weight-loss interventions, including the use of psychoeducation, self-monitoring, stimulus control, and problem-solving. The literature suggests that CBT programs result in similar, if not slightly higher, weight losses than BWL, with most patients losing 5% to 20% of their initial weight. CBT may enhance traditional treatments by focusing on cognitive interventions targeting the maladaptive thinking that often triggers overeating, poor body image, and inactivity.

More specifically, cognitive restructuring is an important component of CBT that involves identifying and challenging negative thought patterns. For example, after eating a cookie, an individual may think, “I blew it. I might as well give up and eat the whole package.” As a result of this thought, the person may binge on the box of cookies. In an attempt to reduce such problematic behavioral patterns, patients are trained to identify cognitive distortions (irrational thoughts) and learn techniques to challenge beliefs that lead to unhealthy behaviors, such as identifying evidence for and against their thoughts. Clinicians often utilize Socratic questioning to help patients learn to challenge automatic thoughts and underlying beliefs. For instance, in the example above, clinicians may ask, “Why do you think it was difficult for you to accomplish your eating goals this week?” Similarly, clinicians may recommend that patients engage in behavioral experiments to challenge their negative assumptions. For example, patients can remain inactive as a result of thinking, “I’m too tired after work to do anything.” Such patients are asked to directly confront this thought by taking a walk despite fatigue. By the end of treatment, it is hoped that patients can elucidate and challenge dysfunctional thoughts and generate alternative thoughts at the moment.

Emotional distress often serves as a trigger for these negative thought patterns. Therefore, when using CBT, clinicians may also focus on stress-reduction methods outside of food. In particular, clinicians may teach relaxation strategies, such as guided imagery, diaphragmatic breathing, and progressive muscle training to help patients more effectively manage problematic emotional states. In high-stress situations, relaxation techniques are intended to directly decrease emotional distress while indirectly facilitating patients’ ability to identify negative thoughts. Ultimately, it is anticipated that relaxation skills and other stress-management techniques can help patients reduce emotional distress and focus on actions consistent with weight loss goals.

Mindfulness-Based Approaches. Developments in mindfulness and acceptance-based interventions have led to increased attention to alternatives to more traditional BWL approaches. Mindfulness and acceptance-based interventions focus on changing patients’ relationships with unpleasant thoughts, feelings, and bodily sensations. For example, acceptance and commitment therapy (ACT) targets experiential avoidance (the avoidance or attempt to control internal events, such as difficult thoughts, emotions, or bodily sensations), with the primary aim of helping patients increase their willingness to experience negative emotional experiences through mindfulness, cognitive defusion, and values techniques. Presumably, by becoming more willing to observe the present moment, patients are able to reduce maladaptive patterns used for avoidance (i.e., eating) and focus on value-driven behaviors (i.e., engaging in activities in the service of health and quality of life). Within this framework, ACT contends that experiential avoidance is the core of both emotional eating and eating in response to environmental stimuli. More specifically, emotional eating typically occurs in response to emotional distress and serves as an attempt to reduce internal negative experiences. Although eating in response to external triggers does not necessarily stem from negative emotions, individuals often experience uncomfortable thoughts and feelings when trying to resist overeating in response to these cues. ACT seeks to promote healthy behaviors consistent with patients' values while reducing avoidance through mindfulness and acceptance techniques.

The research examining mindfulness-based interventions for weight loss supports the use of ACT and general mindfulness techniques to address disordered eating patterns, physical activity, food cravings, and body image dissatisfaction. In addition, a growing body of research demonstrates the positive effect of ACT on behavioral change in the context of weight loss and weight maintenance. One 2008 exploratory study examined the effectiveness of a brief ACT-based group treatment on 68 females randomized to ACT treatment or a control condition. Although no significant differences were initially found in BMI and mental health, intervention participants reported significantly increased physical activity levels at 6 months compared to controls. When participants who reported never using skills were excluded from the analysis, the ACT intervention participants showed significantly greater reductions in BMI. Further, a 2020 systematic review and meta-analysis of third-wave cognitive-behavioral therapies found that ACT-based interventions consistently demonstrated greater weight loss outcomes than standard behavioral treatment at 12- and 24-month follow-up; with other timepoints favoring ACT or demonstrating no significant differences between ACT and behavioral treatment. There is also a growing body of research suggesting that the combination of standard behavioral treatment and ACT may result in improved weight loss and weight maintenance. It is postulated that standard behavioral treatment provides patients with tools to alter their eating patterns, while ACT targets psychological flexibility, cultivating patient’s ability to effectively cope with new challenges as they arise.

Motivational Interviewing. As the weight loss maintenance research suggests, not all patients seeking weight loss are successful. Some patients may struggle to maintain behavioral change long-term, while others may never even begin to change before dropping out of treatment. The Transtheoretical Model (TTM), first proposed by Prochaska and DiClemente, helps to explain why some patients experience difficulty initiating small changes while others keep food diaries and make numerous behavioral modifications from the start. Although the TTM was initially developed to explain failure in smoking cessation interventions, it is now used to conceptualize a wide range of health behaviors, from condom use to exercise adoption and overeating. Regardless of the specific behavior, the TTM postulates that adopting a new behavior involves an evolution through six distinct stages of readiness to change (precontemplation, contemplation, preparation, action, maintenance, and termination) (Table 27.4–7). Within this framework, behavior change is theorized to be a process that develops over time.

Table 27.4–7.
Stages of Change

Stage of Change	**Description**
Precontemplation	▶ No intention to change in foreseeable future ▶ May be unaware of problem ▶ Often present to therapy because of pressure from others ▶ Usually feel coerced or threatened by treatment
Contemplation	▶ Aware problem exists ▶ Seriously considering change but no commitment to action ▶ Struggle with positive evaluation of problem behavior and effort it might take to overcome it
Preparation	▶ Intending to take action in the next month ▶ May have unsuccessfully taken action in the past year ▶ May report small changes (“baby steps”) ▶ Some reductions may be evident
Action	▶ Modifications in behavior, experience, and/or environment ▶ Overt behavior change ▶ One day to 6 months
Maintenance	▶ Prevent relapse and consolidate the gains ▶ Remaining free of the problem behavior and/or consistently engaging in incompatible behavior for >6 months ▶ Can be considered to last a lifetime for some behaviors
Termination	▶ Change process complete ▶ No longer working to prevent relapse ▶ Total confidence and self-efficacy across all high-risk situations ▶ Exited from the cycle of change

Data from Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. *Am Psychol*. 1992;47(9):1102–1114.

Prior to initiating lifestyle interventions for obesity, it is critical to assess and understand a presenting patient’s readiness for change. Patients in the precontemplation stage of change may have a limited appreciation that their obesity is a problem and only come to treatment due to pressure from others. It is not uncommon for patients in this stage to have little intention to change their lifestyle and therefore feeling threatened by treatment. Although patients in the contemplation stage are aware that a problem exists (i.e., “I need to lose weight in order to improve my diabetes”), they often remain stuck between a desire to change versus the perceived obstacles of change. Patients in this stage are sincerely contemplating change but have yet to make a commitment to action. One study indicates that patients commonly remain in the contemplation stage for at least 2 years prior to proceeding to the next stage of change. In the subsequent stage (preparation stage), patients are typically intending to take measurable action within the month and may have even unsuccessfully attempted changes over the past year. In addition, a patient in the preparation stage may describe currently making small changes in preparation for broader change. For example, patients may report that they have obtained a gym membership in anticipation of starting to exercise in the near future. Arguably, many patients coming for a medical weight management evaluation are in this stage of change. Often patients may be seeking professional support in order to take overt steps toward a lifestyle change.

When patients proceed into the action stage, they will typically begin to alter their behavior, environment, and/or experience in an attempt to address their obesity. For instance, patients may change their home environments by throwing away trigger foods and replacing these items with healthier purchases. Within the TTM, individuals are considered to be in the action stage if they have changed a dysfunctional behavior (e.g., inactivity and high-calorie food choices) from as little as one day up to 6 months. In the addictions literature, behavior change is typically defined as the absence of a given behavior (i.e., remaining sober). In weight management, the notion of elimination is less applicable, and instead, behavioral modification or reshaping may be more relevant. Overall, changes during this period are often the most visible. Subsequently, individuals frequently receive the most external recognition during this stage. In particular, it is not uncommon for patients to receive compliments about their weight loss, which can often serve to reinforce lifestyle changes. Following 6 months of consistent change, patients enter the maintenance stage, where they work to consolidate gains and prevent relapse. Individuals are thought to be in the maintenance stage when they remain free of a behavior or implement a new, incompatible behavior for more than 6 months. In regard to lifestyle changes related to weight management, many consider the maintenance stage to be lifelong. However, technically, patients can enter the termination stage, where they have exited the cycle of change. In order to achieve this stage, patients must no longer be working to prevent relapse and have complete confidence in all high-risk situations.

Most individuals engaging in behavior change proceed through the stages of change in a spiral versus linear pattern. Relapse is the rule rather than the exception. Therefore, many patients will not maintain changes on their first attempt. In fact, as they modify their behavior, it is quite common for patients to regress back to an earlier stage and consequently cycle back through the stages. For instance, patients engaging in BWL may have consistently engaged in healthy eating patterns for a few months. However, while at a holiday party, they may have a binge eating episode and subsequently feel ambivalent about continuing to make positive behavior changes. At the beginning of treatment, it is essential to educate patients about the spiral nature of change and to normalize these common experiences.

Furthermore, given the varying levels of readiness for change, when presented with a patient seeking treatment for their obesity, it is important to consider stage-matched treatment. For instance, if patients are in the contemplation stage, giving them action tasks will likely result in failure or drop-out. Unfortunately, most treatments, including weight management programs, are action-oriented and assume that patients are in the action stage. Treatment is the most effective when clinicians match treatment goals to patients’ stages of change. In fact, treatments focused on helping patients move just one stage a month can double the chances that patients will engage in change on their own in the future.

Movement through the stages of change has been shown to be influenced by mediating variables, including the decisional balance and self-efficacy. More specifically, the decisional balance is conceptualized as the balance between the perceived comparative gains and losses (pros and cons) of making a change versus the status quo. Typically, patients in the earlier stages (precontemplation and contemplation) identify more disadvantages than advantages to behavior change. The decisional balance tends to shift as patients move through the stages of change, with patients often perceiving more advantages than disadvantages to change in the preparation, action, and maintenance stages. For example, in the contemplation stage, patients likely perceive many barriers to engaging in consistent activity, while those in the preparation stage may begin to recognize that increased activity results in increased energy and overall improved well-being. Similarly, self-efficacy (individuals’ confidence in their ability to change) appears to be a key variable in movement through the stages. Studies suggest that self-efficacy typically increases linearly across the stages of change.

Table 27.4–8.
Strategies Using Motivational Interviewing Based on Stage of Change

Stage of Change	Potential Clinical Strategies
General techniques	*Express empathy *Increase self-efficacy *Roll with resistance *Develop discrepancy
Contemplation stage	*Identify pros and cons of change versus status quo *Normalize ambivalence *Explore barriers to change *Increase commitment *Support self-efficacy
Preparation stage	*Clarify goals *Explore options *Reinforce sense of personal choice *Practice skills *Develop a plan
Action stage	*Support self-efficacy *Engage in problem-solving *Reinforce commitment to change and success thus far

Adapted from Erol S, Erdogan S. Application of a stage based motivational interviewing approach to adolescent smoking cessation: the Transtheoretical Model-based study. *Patient Educ Couns.* 2008;72(1):42–48. doi: 10.1016/j.pec.2008.01.011.

Motivational interviewing (MI) is a treatment approach based on the stages of change model. MI is intended to enhance patients' motivation for change and thus support patients' movement through the stages of change. MI is considered a patient-centered counseling style with a heavy emphasis on collaboration and partnership between patients and clinicians. The treatment postulates that each patient holds an intrinsic motivation for change. In line with a collaborative approach, clinicians seek to enhance patients' motivation for change through eliciting personal goals, values, and perceptions. Within this framework, the primary goal of MI is to help patients explore and resolve ambivalence around behavior change (see Table 27.4–8 for details regarding MI principles and techniques).

Given the high-rates of weight regain, MI has increasingly been examined and utilized as an alternative or addition to more traditional skill-based weight management techniques. Two large-scale weight loss studies (the Diabetes Prevention Program and the Look AHEAD trial) utilized MI as a part of their lifestyle intervention. In addition, a 2011 meta-analysis of randomized controlled trials using MI found that MI techniques moderately reduced BMI above and beyond control interventions. Across studies using weight as a primary outcome, MI significantly enhanced weight loss outcomes, with participants who received MI losing 1.47 kg more on average than participants receiving control treatments. Of note, in studies tracking change in BMI, MI produced improved but nonsignificant reductions in BMI over controls. Taken together, this body of literature underscores the potential utility of MI in addressing patients' motivation for change when engaging in weight loss treatments.

Case Example

The case of “Robin” exemplified some of the obstacles often seen in BWL treatment. Robin was a 24-year-old Caucasian female self-referred for weight and occasional binge eating episodes. At a BMI of 40, Robin identified several ways obesity negatively impacted her life. In particular, she discussed how she often avoided social gatherings as a result of poor body image and social anxiety related to her weight. Robin further acknowledged that she often used food to cope with stress and anxiety. At the initial session, Robin verbally endorsed a strong desire to lose weight and change her behaviors. However, across treatment sessions, Robin did not complete food monitoring forms and experienced difficulty meeting her exercise goals. Her clinician initially tried to address these difficulties by helping her identify and problem-solve barriers to change. Although barriers may have played a role in Robin’s noncompliance, it is likely that she was struggling with ambivalence around change and may have had low self-efficacy regarding her ability to engage in consistent healthy behaviors. Instead of continuing to focus on action, the clinician decided it was important to explore Robin’s ambivalence further. In particular, the clinician highlighted to Robin the discrepancies between her words and her actions. In addition, they spent time identifying the pros and cons to change. Once ready, Robin, along with her clinician, clarified goals and discussed several options moving forward. Through using MI techniques and increasing her intrinsic motivation for change, Robin was able to move forward into the action stage of change. Fortunately, the clinician realized by using MI that continuing to focus on action-oriented tasks, as most treatments promote, was likely to result in drop-out.

Psychological Treatment of Eating Disorders. Several psychological treatments are supported in treating eating disorders typically associated with obesity. The research examining inventions for BED highlights the various treatment options currently available for several disordered eating patterns. Given a strong evidence base, CBT is currently the gold standard treatment of BED. The treatment is typically disseminated in 12 to 19 weekly group sessions. Across studies, CBT has demonstrated marked improvements in binge eating episodes, symptoms of psychopathology, psychosocial functioning, and modest reductions in weight. In fact, research indicates that patients may experience a clinically significant response by the fourth session. Similar outcomes have also been shown in research examining interpersonal psychotherapy (IPT), though results typically are not observed as quickly. While CBT focuses on dysfunctional behaviors and cognitions related to eating and body image, IPT aims to treat the interpersonal difficulties related to disordered eating patterns. IPT identifies four potential focus areas, including grief, interpersonal role conflicts, role transitions, and interpersonal deficits. In a collaborative, supportive environment, clinicians often utilize clarification, exploration, emotional expression, behavioral modification, and communication analysis to help patients address their interpersonal difficulties. Long-term follow-up in randomized control trials has further supported the sustained treatment effects of both CBT and IPT. For example, one study compared the efficacy of CBT and IPT on BED and associated symptoms 4 years after the cessation of treatment. Across treatments, 64.4% of patients continued to demonstrate full remission from binge eating, supporting the lasting efficacy of both interventions. While the IPT group showed improvements in binge eating symptoms over time, the CBT group demonstrated worsening of symptoms, though treatment outcomes were not significantly different at any time point. The authors propose that these differences in recovery trajectories likely reflect leveling-off between-treatment effects, given the rapid benefits typically observed with CBT. Further, for both treatments, patients continued to report significant improvements in associated eating disorder symptoms and overall psychopathology at four years, though a few variables, including binge episode days, restraint, and anxiety, trended toward relapse. Overall, these results support the continued efficacy of CBT and IPT several years following treatment of BED.

In addition to CBT and IPT, clinicians have utilized dialectical behavior therapy for BED (DBT-BED). DBT-BED focuses on treating emotion dysregulation associated with disordered eating patterns. In a group setting, DBT typically covers mindfulness, distress tolerance, and emotional regulation skills. Patients are further encouraged to use mindful eating and disordered eating-specific diary cards. Increasingly, research has supported the efficacy of DBT for binge eating. One study examined potential variables moderating the effect of DBT-BED treatment outcomes. Interestingly, the researchers found that higher baseline psychopathology was associated with better treatment results. Findings suggested that avoidant personality disorder and the early onset of overweight and dieting (<15 years old) moderated the effect of treatment on post-treatment binge eating episodes. It is possible that patients who have a longstanding history of using food as an avoidant coping mechanism may specifically benefit from a treatment focused on adaptive emotional regulation. However, additional research is needed to elucidate these findings. In addition, CBT and IPT hold the strongest support for the psychological treatment of bulimia nervosa. Similar to outcomes with BED, patients appear to have a quicker resolution of binging and purging behaviors with CBT, though outcomes between CBT and IPT are typically similar around one year.

Less is known regarding the behavioral treatment of NES. However, there is strong emerging evidence for the use of CBT for NES. Treatment typically focuses on behavioral techniques (e.g., closing the kitchen at a specific time and delaying getting out of bed), increased awareness of thoughts and behaviors associated with night eating (i.e., “If I don’t eat, I won’t fall back to sleep”), cognitive restructuring, and sleep hygiene behaviors. A pilot study of a 10-session CBT intervention for NES demonstrated significant decreases in evening caloric intake, nocturnal ingestions, number of awakenings per week, and weight. Improvements were also seen in mood and quality of life. In addition, a small randomized clinical trial underscored the potential utility of providing NES patients with educational material and relaxation techniques as a means of helping patients to reduce the percentage of food eaten after the evening meal. Although this initial research is promising, continued investigation is needed in this specialized area of eating disorder treatment.

Medical Treatments

Medications for weight loss should be considered in patients with a BMI >30 or a BMI >27 with serious comorbidities (Table 27.4–2). Behavioral and lifestyle interventions, including a diet and exercise plan, are essential adjunctive therapies when using medications for weight loss. Medication treatment of obesity has a varied history, including concerns about addiction and serious side effects. Amphetamines were used for weight loss in the early 20th century but led to substance abuse. In the late 20th century, the combination treatment of di-fenfluramine and phentermine (“Fen-Phen”) became popular until valvular heart disease side effects were discovered. Over-the-counter (OTC) treatments such as ephedrine with caffeine were linked with reports of death and pulled from the market. Research on the use of leptin to treat weight problems has been disappointing, and mood changes associated with potential weight loss drugs such as ecopipam and rimonabant have prevented their approval for use. Sibutramine (Meridia) was approved for several years but withdrawn after increased blood pressure, and adverse cardiovascular events were discovered.

Current medications approved for weight loss include orlistat (Xenical, Alli), phentermine/topiramate (Qysmia), naltrexone/bupropion (Contrave), phentermine (Adipex), and liraglutide (Saxenda) (Table 27.4–9). Orlistat acts as a lipase inhibitor, thereby reducing caloric absorption, and is available in prescription formulation (120 mg three times daily before meals) and an OTC formulation (60 mg). Orlistat typically results in an 11% weight loss, although oily stools and GI distress are common side effects. Orlistat may also cause a reduction in absorption of fat-soluble vitamins and, in rare cases, liver disease.

Phentermine/topiramate is approved for long-term use in doses of 3.75 to 15 mg phentermine and 23 to 96 mg extended-release topiramate and appears to work through a reduction in appetite. Weight loss with the medication appears to be 9% to 10%. Side effects may include paraesthesias, dizziness, altered taste, insomnia, constipation, and dry mouth. This medication, as with all weight loss medications, is contraindicated in pregnancy. Of note, other potential risks include suicide, mood, and sleep disorders.

Naltrexone/bupropion was brought to market in 2014 and appears to reduce food intake by acting on adrenergic and dopaminergic receptors in the hypothalamus. Significant weight loss and reduction in comorbidities were achieved on the medication. Side effects include nausea and potential interactions with SSRIs and MAOIs.

Table 27.4–9.
Medications Approved for Weight Loss Treatment

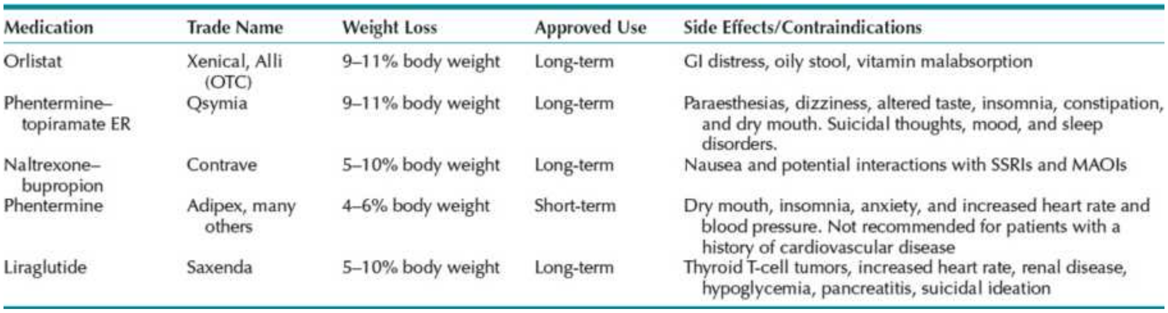

Medication	Trade Name	Weight Loss	Approved Use	Side Effects/Contraindications
Orlistat	Xenical, Alli (OTC)	9–11% body weight	Long-term	GI distress, oily stool, vitamin malabsorption
Phentermine–topiramate ER	Qsymia	9–11% body weight	Long-term	Paraesthesias, dizziness, altered taste, insomnia, constipation, and dry mouth. Suicidal thoughts, mood, and sleep disorders.
Naltrexone–bupropion	Contrave	5–10% body weight	Long-term	Nausea and potential interactions with SSRIs and MAOIs
Phentermine	Adipex, many others	4–6% body weight	Short-term	Dry mouth, insomnia, anxiety, and increased heart rate and blood pressure. Not recommended for patients with a history of cardiovascular disease
Liraglutide	Saxenda	5–10% body weight	Long-term	Thyroid T-cell tumors, increased heart rate, renal disease, hypoglycemia, pancreatitis, suicidal ideation

Phentermine is approved for short-term use in treating patients with obesity. Because of the potential for abuse, its use is typically limited to 12 weeks. Phentermine is the most often prescribed medication for weight loss in the United States, but there is little data to determine its long-term efficacy. Effects of the medication appear to be dose-related, and side effects can include dry mouth, insomnia, anxiety, and increased heart rate and blood pressure. Phentermine is not recommended for patients with a history of cardiovascular disease.

Liraglutide is the newest addition to the drugs approved for weight loss. Its action on GLP-1 may regulate areas of the brain involved in appetite. Weight loss results are promising, but side effects/risks include thyroid t-cell tumors, acute pancreatitis, increased heart rate, and suicidal ideation. The most recent studies on GLP-1 agonists are promising and have suggested that patients taking this class of medication have the potential to lose more weight when compared to other weight loss medications; though weight-gain has been shown if patients discontinue the medication.

Further evaluation of antidiabetic drugs for weight loss indications is an area of future research in the field. There are concerns about the extended use of weight loss medications, as they typically do not produce permanent weight loss (i.e., when the drug is stopped, the weight is regained.) In addition, the history of approved weight loss medications with later emerging serious side effects is important to consider for future research.

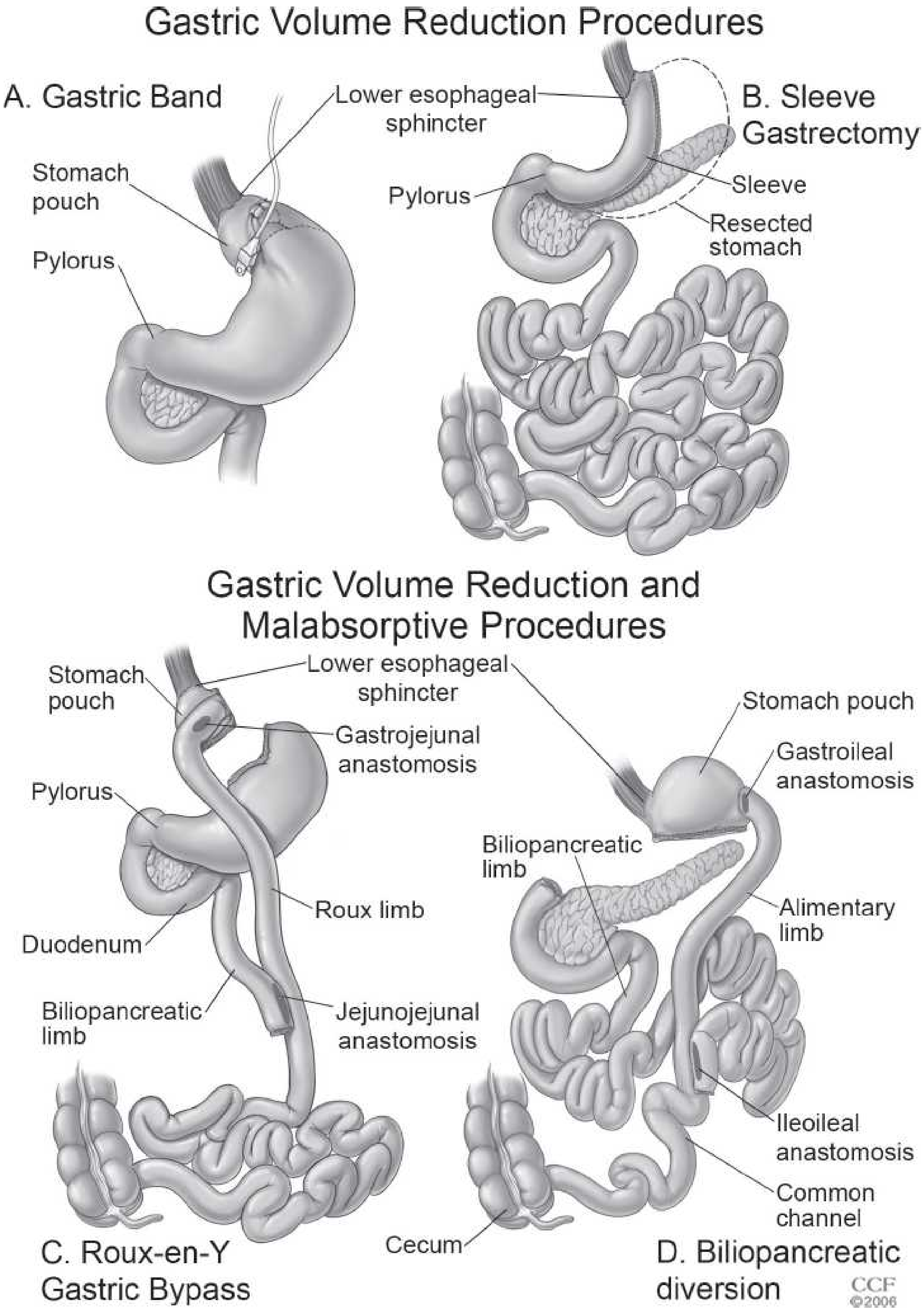
Gastric Volume Reduction Procedures
A. Gastric Band
Lower esophageal sphincter
Stomach pouch
Pylorus
B. Sleeve Gastrectomy
Pylorus
Sleeve
Resected stomach
Gastric Volume Reduction and Malabsorptive Procedures
Stomach pouch
Lower esophageal sphincter
Gastrojejunal anastomosis
Pylorus
Roux limb
Duodenum
Biliopancreatic limb
Jejunojejunal anastomosis
Stomach pouch
Gastroileal anastomosis
Biliopancreatic limb
Alimentary limb
Ileoileal anastomosis
Common channel
Cecum
C. Roux-en-Y Gastric Bypass
D. Biliopancreatic diversion
CCF
©2006

FIGURE 27.4–3. Restrictive, malabsorptive, and combination weight loss surgery procedures. (Reprinted with permission, Cleveland Clinic Center for Medical Art & Photography © 2006–2015. All Rights Reserved.)

Surgical Treatments

Weight loss surgery is currently the most effective long-term treatment of severe obesity (Table 27.4–2). Patients with severe obesity (BMI >40) or moderate obesity (BMI >35) with associated obesity-related disease are candidates for weight loss surgery. Candidates should be well-informed, motivated, psychologically stable, have good social support, have no substance abuse, and be an acceptable operative risk. Patients seeking weight loss surgery should have documented nonsurgical weight loss attempts and be willing to make long-term lifestyle changes.

The current surgical treatments include both restrictive and malabsorptive procedures (Fig. 27.4–3). Laparoscopic adjustable gastric banding (LAGB) is a procedure that involves an adjustable device that creates a small gastric pouch and narrow outlet, thus restricting the amount of intake and inducing satiety. LAGB is reversible and has low early postoperative risks, but the weight loss tends to be lower than with other procedures. In addition, late postoperative risks such as erosion and slippage are common. Laparoscopic sleeve gastrectomy (LSG) involves the removal of 70% to 80% of the stomach, restricting intake, and appears to result in greater weight loss than the LAGB. Malabsorptive procedures, including the biliopancreatic diversion and duodenal switch (BPD-DS), bypass a segment of the small intestine, so less food is absorbed. These procedures are rarely performed because of the high-risk of nutritional deficiencies. The Roux-en-Y gastric bypass (RYGB) is a combination procedure using both restriction and malabsorption to promote weight loss and results in a 60% to 80% loss of excess body weight. Data from the American Society for Metabolic and Bariatric Surgery reported 2,56,000 surgeries were performed in the United States in 2019, with the most common primary procedure cited as the LSG (59.4%), followed by RYGB (17.8%), and LAGB and BPD-DS (both 0.9%).

The majority of bariatric surgeries are performed laparoscopically, which reduces the risk of infection, incisional hernia, postoperative pain, and wound complications. Benefits of weight loss surgery include weight loss of 40% to 80% excess body weight and resolution of comorbidities such as metabolic syndrome, hypertension, sleep apnea, liver steatosis, hypercholesterolemia, and gastroesophageal reflux disease. Other benefits include increased life expectancy and improved quality of life. Potential risks of weight loss surgery include bleeding, anastomotic leak, wound infection, thromboembolism, anastomotic strictures, marginal ulcers, bowel obstruction, cholelithiasis, and nutritional deficiencies. The risk of perioperative death for RYGB is approximately 0.5%, with risks increasing for older patients, male patients, and lower surgeon volume.

Psychological Considerations for Bariatric Surgery. Psychological risks after weight loss surgery include increased risk of alcohol abuse, increased suicide rates, and possible problems with absorption of psychiatric medication. Researchers have established an increased risk of the development of alcohol use disorders in patients following RYGB and LSG that may be related to increased absorption. There appears to be a three-fold increase in suicide risk for patients following weight loss surgery, with patients in the first 3 years after surgery and men being at a higher risk of suicide. Another possible concern for mental health providers is the apparent decrease in absorption of SSRIs after Roux-en-Y. Given the high number of bariatric patients on antidepressants, recommendations include close psychiatric monitoring after surgery and movement to shorter-acting as opposed to extended-release formulations.

Other considerations are the stresses of adjusting to a new lifestyle, changes in family relationships due to role changes from weight loss, body image issues, and coping with excess skin. Research also suggests concern regarding eating disorder behaviors returning after surgery, such as loss of control eating, bulimia, and anorexia. Adjunctive counseling and psychotherapy can be helpful in monitoring psychological symptoms after surgery and helping patients adjust to a new lifestyle.

Psychological treatment of the bariatric candidate ideally involves a behavioral health professional such as a psychologist or psychiatrist who is fully integrated into the bariatric team. The rates of psychiatric disorders, such as depression, bipolar depression, BED, and anxiety disorders, are high among bariatric surgery patients. For example, one study of presurgical bariatric patients found a 68% lifetime incidence of Axis I disorders, and 38% of patients currently met the criteria for an Axis I diagnosis. Presurgical evaluation of the bariatric patient is standard and typically includes psychological testing and a psychosocial interview. Elements of the evaluation should include assessment of the patient’s motivation, knowledge of surgery, and lifestyle changes needed, weight loss and other expectations, mental health, substance use, social supports, stressors, coping strategies, and eating patterns. Psychological contraindications for surgery include unstable depression or psychosis, suicidal ideation, active bulimia, and active substance abuse.

Behavioral health specialists may also play a role in optimizing patients for surgery by engaging them in mental health treatment to address instability and treating eating disorder behaviors presurgically. The psychological team may work with inpatient bariatric patients presenting with anxiety or depressed mood in the perioperative period, and follow-up after surgery may help to monitor mood and behavior change as the patients adjust to their new lifestyle. The behavioral health specialist often plays a leading role in providing support groups as part of the bariatric program.

FUTURE DIRECTIONS

There is a well-established relationship between obesity and psychiatric disorders, but more needs to be understood about the common causes of obesity and psychological disorders and how best to provide treatment to patients with both obesity and mental health disorders. The field would benefit from additional clinical trials that utilize combination treatments such as medications with behavioral weight management/CBT to help mitigate weight gain from psychotropic medications and optimally treat mental health disorders with the lowest side effect profile for patients with obesity. Improving adherence to treatment and addressing factors related to weight regain will continue to be areas of study. The role of the behavioral health specialist in the treatment of obesity is likely to grow as health care moves toward a more integrated approach to patient treatment and understanding mind-body health. There are benefits to adjunctive psychological treatment at all stages of obesity treatment, including prevention, diagnosis, behavioral, medical, and surgical treatment. Future research should establish data to support the integrated treatment of obesity in the psychiatric field.

FURTHER READINGS

Allison KC, Lundren JD, Moore RH, Stunkard AJ. Cognitive behavior therapy for night eating syndrome: a pilot study. *Am J Psychother*. 2010;64(1):91–106.

*Amstrong MJ, Mottershead TA, Ronksley PE, Sigal RJ, Camobell TS, Hemmelgarn BR. Motivational interviewing to improve weight loss in overweight and/or obese patients: a systematic review and meta-analysis of randomized controlled trials. *Obes Manag*. 2011;12:709–723.

Anastasiou CA, Karfopoulou E, Yannakoulia M. Weight regaining: from statistics and behaviors to physiology and metabolism. *Metabolism*. 2015;64(11):1395–1407.

*Bray GA. Medical treatment of obesity: the past, the present and the future. *Best Pract Res Clin Gastroenterol*. 2014;28:665–684.

Centers for Disease Control and Prevention. Overweight and obesity. http://www.cdc.gov/obesity/index.html

Erol S, Erdogan S. Application of a stage based motivational interviewing approach to adolescent smoking cessation: the transtheoretical model-based study. *Patient Educ Couns*. 2008;72(1):42–48.

Gariepy G, Nikita D, Schmitz N. The association between obesity and anxiety disorders in the population: a systematic review and meta-analysis. *Int J Obesity*. 2010;34:407–419.

Gerlach G, Herpertz S, Loeber S. Personality traits and obesity: a systematic review. *Obes Rev*. 2015;32(16):32–63.

Heinberg LJ, Ashton K, Windover A. Moving beyond dichotomous psychological evaluation: the Cleveland Clinic Behavioral Rating System for weight loss surgery. *Surg Obes Relat Dis*. 2010;6(2):185–190.

Hilbert A, Bishop ME, Stein RI, et al. Long term efficacy of psychological treatments for binge eating disorder. *Br J Psychiatry*. 2012;200:232–237.

King WC, Chen JY, Mitchell JE, et al. Prevalence of alcohol use disorders before and after bariatric surgery. *JAMA*. 2012;307:2516–2525.

Lasikiewicz N, Myrissa K, Hoyland A, Lawton CL. Psychological benefits of weight loss following behavioral and/or dietary weight loss interventions. A systematic research review. *Appetite*. 2014;72:123–137.

Lee WJ, Almulaifi A. Recent advances in bariatric/metabolic surgery: appraisal of clinical evidence. *J Biomed Res*. 2015;29(2):98–104.

Lillis J, Kendra KE. Acceptance and commitment therapy for weight control: model, evidence, and future directions. *J Contextual Behav Sci*. 2014;3(1):1–7.

MacLean PS, Wing RR, Davidson T, et al. NIH working group report: innovative research to improve maintenance of weight loss. *Obesity*. 2015;23(1):7–15.

Marek RJ, Lavery M, Heinberg L, Rish J, Ashton K. A review of psychological assessment instruments for use in bariatric surgery evaluations. *Psychol Assessment*. 2016;28(9):1142–1157.

*McElroy S. Obesity in patients with severe mental illness: overview and management. *J Clin Psychiatry*. 2009;70:12–21.

Michael JC, Belendiuk KA, Kuchera AM, Rofey DL. Cognitive behavioral therapy for weight management. *Behav Therapist*. 2013;36(2):28.

*Nihalani N, Schwartz TL, Siddiqui UA, Megna JL. Obesity and psychotropics. *CNS Neurosci Ther*. 2012;18:57–63.

Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011–2012. *JAMA*. 2014;311:806–814.

Peterhänsel C, Petroff D, Klinitzke G, Kersting A, Wagner B. Risk of completed suicide after bariatric surgery: a systematic review. *Obes Rev*. 2013;14(5):369–382.

Peterson ND, Middleton KR, Nackers LM, Medina KE, Milsom VA, Perri MG. Dietary self-monitoring and long-term success with weight management. *Obesity (Silver Spring)*. 2014;22(9):1962–1967.

Preiss K, Brennan L, Clark D. A systematic review of variables associated with the relationship between obesity and depression. *Obes Rev*. 2013;14:906–918.

Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. *Am Psychol*. 1992;47(9):1102–1114.

Pursey KM, Stanwell P, Gearhardt AN, Collins CE, Burrows TL. The prevalence of food addiction as assessed by the Yale Food Addiction Scale: a systematic review. *Nutrients*. 2014;6:4552–4590.

Robinson AH, Safer DL. Moderators of dialectical behavior therapy for binge eating disorder: results from a randomized controlled trial. *Int J Eat Disord*. 2012;45(4):597–602.

Tapper K, Shaw C, Ilsley J, Hill AJ, Bond FW, Moore L. Exploratory randomized controlled trial of a mindfulness-based weight loss intervention for women. *Appetite*. 2009;52:396–404.

VanBuskirk KA, Potenza MN. The treatment of obesity and its co-occurrence with substance use disorders. *J Addict Med*. 2010;4:1–10.

Yen YC, Huang CK, Tai CM. Psychiatric aspects of bariatric surgery. *Curr Opin Psychiatry*. 2014;27(5):374–379.

Yska JP, van der Linde S, Tapper VV, et al. Influence of bariatric surgery on the use and pharmacokinetics of some major drug classes. *Obes Surg*. 2013;23:819–825.

27.5 Respiratory Disorders

Thomas Ritz, Ph.D., Hannah O. Nordberg, M.A., Ph.D., Chelsey A. Werchan, Ph.D., and E. Sherwood Brown, M.D., Ph.D., M.B.A.

DEFINITION AND COMPARATIVE NOSOLOGY

Psychiatric symptoms and syndromes in respiratory disorders are less extensively investigated than those in other common medical illnesses. However, numerous clinical features of these illnesses may increase the risk of developing a psychiatric disorder. In individuals with comorbid psychiatric illness, clinical features may also influence psychiatric symptoms. Respiratory disorders are frequently associated with dyspnea, hypoxia, poor sleep quality, and insomnia, which all can increase the risk for a variety of psychiatric illnesses. Furthermore, severe respiratory disorders are associated with substantial functional impairment that can increase the risk of anxiety and depression. Comorbid respiratory and psychiatric conditions contribute to worse health outcomes including increased emergency care utilization. Many of the medications given for lung diseases can have significant central and peripheral nervous system effects.

The psychiatrist's role in respiratory disease includes (1) diagnosis and treatment of comorbid psychopathology; (2) addressing individual psychosocial triggers of exacerbation; (3) assessment and treatment of emotional, neurocognitive, and psychosocial consequences of chronic respiratory disease and its management; and (4) supporting the management of the disease, including adherence, motivation for rehabilitation, devise management strategies, lifestyle interventions, and improvement in comanagement by families and partners.

ASTHMA

Asthma is a very common respiratory illness in both children and adults. Global data indicates an increase in asthma prevalence in recent years. Improvement in diagnosing the condition may contribute to rising prevalence rates, however, asthma remains poorly controlled in many individuals who receive medical care for their disease. Below is a case vignette illustrating the complex relationship between disease-related factors and vulnerability to psychiatric disorders in asthma, a relationship that also applies to other respiratory conditions.

A 34-year-old woman presents with a history of episodes of severe and prolonged coughing since childhood. Symptoms typically started following upper respiratory infections and resulted in persistent coughs that made it difficult for the patient to sleep through the night. Exercise also left her wheezing and coughing at the end of the day, particularly during summer. Asthma was diagnosed at 6 years of age, after a particularly severe exacerbation that required an emergency room visit. The patient was prescribed a low dose of inhaled corticosteroids as maintenance medication and a β-adrenergic bronchodilator for as-needed use. During her teens, the patient became highly self-conscious about her asthma symptoms and her medication adherence greatly decreased, resulting in poorer symptom control and increased bronchodilator use. During an emergency room visit at age 18, she received a burst of high-dose oral corticosteroids for a week. Her energy greatly increased, and she developed severe insomnia. She spent one sleepless night cleaning her room. She was also very irritable and had a verbal altercation with a classmate at school. These symptoms resolved within a week of stopping the corticosteroids, but similar symptoms emerged with subsequent corticosteroid courses prescribed for asthma exacerbations.

During college, the patient was provided with a brief asthma management training, which included a plan to avoid overexposure to outdoor air pollutants during critical days in summer and improvement in indoor environment by mold and house dust mite eradication. Because she also reported anxiety and depression, a psychiatric consultation was obtained, which revealed major depressive disorder, generalized anxiety disorder, and occasional panic attacks. Feelings of helplessness were accompanied by hyperventilation, which led to strong symptoms of shortness of breath, chest tightness, and dizziness. Symptoms sometimes spiraled out of control as the patient feared suffering from an asthma attack. Bronchodilators were only partly effective in relieving these symptoms. The treating psychiatrist initiated pharmacologic treatment, with an SSRI and cognitive behavioral training that also included slow and shallow breathing exercises to reduce hyperventilation.

Historical Perspective

Observations of labored breathing and asthma have a long history and can be traced back to ancient texts from Egypt and Greece. Already at that time, a connection between the brain and lungs was suspected. Hippocrates viewed health as a balance of healthy quantities of humors and an excess of humors was linked to illness. If one of these, phlegm, was produced excessively, it would condense in the brain and travel down to the lower organs, among these the lungs. Hence, the term *catarrh*, translated from Greece as “flowing” or “flowing down.” The psychoanalytic tradition viewed asthma as a vegetative manifestation of passive–regressive impulses that stem from unresolved conflicts related to maternal attachment, famously culminating in the interpretation of asthma as “suppressed cry for the mother.” With progress in elucidating molecular and cellular aspects of asthma pathophysiology, the psychological study of asthma became more heavily oriented toward illness management behavior in the 1970s and 1980s. Large-scale epidemiologic studies and psychophysiological laboratory experimentation have recently produced more convincing empirical evidence for influences of psychosocial factors and psychopathology on both etiology and pathophysiology of asthma.

Disease Manifestation and Pathophysiology

Asthma is a chronic, inflammatory airway disease that leads to episodic symptom exacerbations, which extoll a substantial burden on the patients’ well-being and can be life-threatening if not controlled adequately. Pathophysiologic changes include obstruction, inflammation, hyper-responsiveness, and remodeling of the airways. The airways are hyper-reactive in asthma, with stronger than normal narrowing of the bronchi in response to a range of triggers, including allergens, physical exercise, air pollutants, odors, smells and irritants, cold air, respiratory tract infections, certain drugs (e.g., aspirin), gastric reflux, or psychological stress, and emotions. Patients experience unpleasant breathing sensations (dyspnea), which are characterized by shortness of breath, chest tightness, coughing, and/or wheezing. Additionally, mood and energy levels are affected in symptomatic episodes, commonly leading to fatigue, irritability, anxiety, fear, and/or panic. In fact, there has been evidence to suggest that although breathing symptoms are often the primary focus of visits for patients with asthma, hyperventilation, and fatigue symptoms are endorsed quite often in adults with asthma, suggesting the importance of screening and identification of extrapulmonary symptoms in patients with asthma.

There is no single diagnostic indicator for asthma. Evidence comes from multiple sources, such as disease history and manifestations (including symptom characteristics and precipitating factors); physical examination focused on upper respiratory tract, chest, and skin; mechanical lung function testing by spirometry; demonstration of reversibility of airway obstruction by bronchodilator application; assessment of airway inflammation; airway hyper-reactivity testing with methacholine, histamine, cold air, or exercise; and skin or blood tests for allergic sensitization. Similarly, asthma severity and control is determined by numerous factors including medication use and type of medications taken, lung function, symptoms recently experienced, functional limitations, and number of exacerbations requiring care.

Prevalence and Societal Burden

Over 334 million people worldwide suffer from asthma, with prevalence rates anticipated to increase by 100 million cases in 2025. Over 30 million individuals have asthma in the United States. Although a clear diagnosis is difficult in the first 5 years of life, both children and adults experience asthma. In children, asthma is the most common chronic disease, affecting more than 7 million children in the United States. Children from minority populations and socioeconomically disadvantaged backgrounds are disproportionally affected by asthma, in particular African American and Puerto Rican communities. With close to 250,000 deaths from asthma per year worldwide, asthma-related morbidity and mortality highlight the need for better control and prevention methods. The socio-economic burden of asthma is high in adult and child populations, with contributions from direct costs (e.g., outpatient care), indirect costs (e.g., missed school days), and intangible costs (e.g., increased limitations to physical activity).

Pharmacologic Treatment and Self-Management

The goal of asthma pharmacologic management is to achieve optimal control of asthma with the right type and dose of agents. Anti-inflammatory medication, in particular inhaled corticosteroids, is the first line of treatment and is required as daily maintenance medication. Short-acting bronchodilators (typically β-adrenergic) are used as rescue medication on an as-need basis and provide relief from symptoms for approximately 6 hours. Long-acting β-adrenergic bronchodilators help prevent bronchoconstriction for approximately 12 hours, particularly throughout the night. Other anti-inflammatory agents are sometimes used as maintenance medication, such as leukotriene modifiers, which target specific inflammatory mediators and also have bronchodilatory properties. Biologics including immunoglobulin-E inhibitors are sometimes prescribed for severe allergic asthma. Severe cases or disease exacerbations typically require high doses of systemic corticosteroids.

Consistent efforts are required by patients with asthma to self-manage their condition, which includes adherence to medication schedules, action plan-guided self-adjustments of medication in exacerbations, self-monitoring symptoms or lung function, and control of critical environmental and endogenous asthma triggers. Typically, around half of patients show suboptimal medication adherence, which contributes to low asthma control and frequent exacerbations. Suboptimal medication adherence is seen more frequently in minority groups that are disproportionally affected by asthma. Barriers to medication adherence include disease belief of “no symptoms, no asthma.” Patients with this belief report poor adherence to daily maintenance medications and rely on rescue medications only when symptomatic. In addition, high costs of outpatient care can restrict individuals from lower income groups from affording daily medications and reaching optimal medication adherence. Comprehensive educational programs in self-management improve asthma outcomes and are generally recommended by guidelines. Designing and implementing interventions based on theories of behavior change, learning theory, or motivational interviewing techniques can enhance their effects. Involving families in self-management of children and young adolescents is essential and additional training in family problem-solving skills or parent coaching can benefit asthma-related quality of life. In addition, programs providing support to patients who more frequently report dysfunctional beliefs or have heavier trigger exposure (e.g., cockroaches and dust-mites in the home) can help optimize asthma care.

Stress and Emotion in Asthma

A long history of clinical and epidemiologic literature links asthma symptoms and exacerbations to patients’ reports of a variety of emotional precipitants, typically various negative affective states, but also ambivalence or conflicts over emotional expression, or even positive states of excitement and mirth. Emotional triggers of asthma symptoms are identified by around one-fourth of patients as their top triggers. They are uniquely associated with lower asthma control and quality of life, as well as higher rates of exacerbations and emergency treatments, over and above other asthma triggers. Laboratory studies have established that bronchoconstriction can be elicited by stress and emotion-induction techniques, with an estimated 20% to 40% of patients responding with clinically relevant airway obstruction. Major physiologic pathways for obstruction due to psychological stimuli are parasympathetic excitation, hyperpnea (increased depth of breathing), and hypocapnia. In addition, distinct central nervous system pathways have been identified for effects of stress and emotion on the airways. Emotional triggers are specifically associated with hyperventilation symptoms, which in turn are related to lower perceived health, likely through reduced effectiveness of conventional asthma medication.

Observational studies have also expanded the current knowledge, showing that psychosocial stressors, negative life events, and adversity, such as low socioeconomic status, are linked to lung function decline, airway inflammation, and changes in immune status favoring asthma exacerbations. Recent theoretical models also emphasize a role in stress-induced infections, epithelial dysfunction, microbiome (dysbiosis), and epigenetic changes. In addition, studies indicate a role of family emotional health, including parent and caregiver emotions, in asthma symptomology, with disease control associated with parental mood states including depression. Consequently, emotions and stress are now listed among the well-established triggers of asthma symptoms, as reflected in current asthma management guidelines. Elevated anxiety and problems communicating emotions, a facet of alexithymia, have been found in patients who had experienced a near-fatal asthma exacerbation.

Evidence is also strong for an association between anxiety, depression, and patients’ perception and report of asthma symptoms. Given that asthma symptoms are typically episodic, the patient’s ability to detect and communicate symptoms effectively are key to successful management of the disease. Greater symptom exacerbations and worse lung function have been observed in children whose parents report high-anxiety levels, indicating an importance of family emotional environment in relation to asthma. Anxious patients tend to overestimate or over-report airway obstruction. Consequently, they use more rescue inhalers and access more professional care than necessary, including emergency room treatments. Patients who develop an intense fear of asthma exacerbations may be more emotionally reactive to physical symptoms they experience, perceiving any somatic change as threatening. In turn, these patients can experience respiratory symptoms of anxiety, such as dyspnea. These patients may have difficulty delineating between asthma and anxiety symptoms, which can bidirectionally influence one another, and be treated with more oral corticosteroids than less fearful patients.

Neuroimaging studies have demonstrated that brain structures and pathways typically involved in emotional processing and psychopathology, in particular limbic structures, including right anterior insular, are also involved in the perception of dyspnea, thus improving the mechanistic understanding of emotional influences on airway symptoms. Furthermore, many patients with asthma have deficits in recognizing airway obstructions and thus underestimate the severity of their disease state. As a consequence, they show insufficient adherence to maintenance medication and delay necessary treatments, thereby risking life-threatening exacerbations.

Comorbidity with Psychiatric Disorders

Epidemiologic studies have shown that asthma is associated with an elevated risk of psychiatric conditions, including anxiety disorders (agoraphobia/panic, generalized anxiety disorders, post-traumatic stress disorder, and social phobia), depressive disorders, alcohol use disorders, schizophrenia, and suicide. Sleep disorders are experienced at higher rates in asthma, with individuals with asthma experiencing insomnia and obstructive sleep apnea 2.5 times more often than individuals without asthma. Sleep disorders can exacerbate already present mood problems or lead to the development of additional psychiatric conditions. Symptoms patients with asthma experience during exacerbations can include anxiety, panic, irritability, frustration, and fatigue, closely mimicking some of the psychiatric symptomatologies. The unsettling experience of the respiratory disease may lead to secondary psychiatric problems, or conversely, psychopathology may facilitate neuroendocrine and/or inflammatory/immunologic pathways involved in asthma pathogenesis. Thus, asthma symptoms experienced in childhood are associated with greater levels of anxiety and shyness, and they predict the future development of anxiety disorders. Subsequent onset of panic disorder has been interpreted as evidence for a suffocation-alarm theory of panic, which views panic as an exaggerated response to CO_2 increases and posits that sensitization to rising CO_2 levels due to respiratory disease is a critical factor in the development of panic. Beyond direct pathways, third variables, such as shared genetic risk factors or specific childhood adversities leading to disease manifestations at different time points, also need to be considered. In support of such assumptions, elevated levels of a range of psychiatric conditions are typically found in first-degree relatives of adolescents with asthma, although these could also be secondary to the burden of asthma care in the family.

Existing psychiatric disorders and asthma can influence each other in a number of ways. Overlap of symptoms can cause challenges in the diagnosis of asthma. Prominent symptoms of a psychiatric condition, such as schizophrenia, can distract from comorbid asthma and lead to its underdiagnosis and undertreatment. Other complications can affect the course and treatment of asthma. Strong emotional states associated with psychopathology can affect asthma pathophysiology directly through pathways discussed earlier. In the case of psychiatric disorders, these states occur repeatedly over a prolonged period and thereby constitute stressors that affect organ function chronically. Thus, hyperventilation associated with panic disorder can worsen lung function and potentially exacerbate inflammation. Anxious or depressed mood can also affect disease management, for example, by over- or underuse of rescue inhalers, lack of adherence to prescribed maintenance medication, neglect of trigger control, problematic health behaviors affecting the airways (e.g., smoking), or delayed treatment seeking. Additionally, adoption of a lifestyle that is less health-promoting, such as increased alcohol use and poor diet can worsen disease-related outcomes.

Symptoms of asthma can be a source of panic or contribute to the fatigue and exhaustion experienced in mood disorders. Internalizing disorders in children seeking cognitive behavioral treatment are more severe in those with comorbid asthma. Earlier research has highlighted problems in patients who are prone to intense states of “panic-fear,” most likely indicating clinically significant anxiety disorders. Such patients were found to have excessive rates of hospitalization for asthma exacerbations. Asthma-related quality of life is typically reduced with comorbid psychopathology. Conversely, asthma can also amplify the symptomatic appearance of psychiatric conditions and complicate their treatment. For example, nocturnal symptoms of asthma can impede sleep quality in asthma and worsen insomnia symptoms. Asthma-related factors influencing sleep when treating insomnia and other sleep disorders, such as nighttime wheezing, are thus important to consider. Additionally, nocturnal asthma symptoms may distract from comorbid sleep conditions (e.g., sleep apnea), leading to underdiagnosis of sleep conditions and exacerbation of psychiatric symptoms.

Asthma may also be associated with an increased risk of cognitive disorders. Large studies suggest that asthma is associated with an increased risk of mild cognitive impairment, and that an asthma diagnosis increases the risk of later development of dementia. The reasons for this association between asthma and cognitive disorders are not known, but could be related to reduction in respiratory function, inflammatory processes, or medication side effects. Recent findings suggest that greater rates of cognitive decline in older adults with asthma are associated with poorer lung function and earlier onset of the disease. Asthma is also associated with regional brain volume changes. A smaller hippocampus, a brain region involved in memory and emotion, was reported in people with asthma. Furthermore, asthma duration, control, and fear of exacerbations have been negatively associated with putamen and pallidum volume.

Treatment of Comorbid Psychopathology in Asthma. The psychiatric treatment of comorbid asthma patients requires specific considerations and precautions. However, despite findings of high comorbidity rates, the development of specifically adapted treatments for comorbid asthma patients has been lagging. Medication treatments for both conditions can have mutual adverse side effects. Psychotropic medication for anxiety can affect respiration adversely and thus may complicate asthma symptomatology. High-doses of benzodiazepines or hypnotics can lead to respiratory depression and are related to a greater risk of death from asthma. Overall, evidence to support pharmacologic treatment of anxiety and depression in asthma remains inconsistent.

However, there is evidence that cognitive behavior therapy (CBT) for anxiety disorders, and perhaps also depression, adapted for asthma, is beneficial in reducing anxiety and bronchodilator use and improving quality of life. Treatments typically combined generic anxiety-related treatment elements, such as cognitive restructuring, interoceptive exposure to feared bodily sensations, progressive muscle relaxation, and psychoeducation, with asthma-relevant elements such as asthma education and breathing retraining and eliminated treatment aspects potentially detrimental to patients with asthma, such as voluntary hyperventilation as interoceptive exposure exercise. Empirical support from published psychological intervention studies is still largely missing, and there is a clear need for targeting asthma-related depressive deficits, including low self-esteem, social isolation, and reduced medication adherence and management behaviors. Treatment with CBT for insomnia, including web-based applications that may help reduce barriers to treatment, appears to be beneficial for adults with comorbid insomnia and asthma.

Adjunctive Cognitive and Behavioral Treatment Modalities for Asthma

Adjunctive cognitive and behavioral treatment modalities have been developed or adapted for asthma, which are not specifically directed at treating comorbid psychiatric conditions; but address psychiatrically relevant deficits in patients with asthma, such as maladaptive breathing patterns, inaccurate perception of airway obstruction, or suffering from excessive stress. They can significantly add to pharmacotherapy. Certain habitual breathing patterns that support hypocapnia or hyperpnea are targeted by dedicated training modules of slow, abdominal, shallow, more regular, and/or nasal breathing. Although breathing is under voluntary control, characteristics of the person's breathing pattern are typically quite stable and systematic, and instructed efforts are needed to change it permanently. Training in slow, abdominal breathing has been shown to reduce symptoms and improve patients' quality of life. Mechanisms underlying such improvements are largely unexplored but may include a better gas exchange efficiency or reduction in high-flow rates that can irritate and dry out the hyper-reactive airways. However, exaggerated deep breaths in slow breathing can also lower PCO_2 levels, thus making hypoventilation training a safer approach. Successful interventions train prolonged respiratory pauses or slower and shallower breathing with capnometry-assisted biofeedback. More sustained symptomatic improvements have been noted. Feedback can also be provided to increase respiratory sinus arrhythmia by slow breathing, which has been associated with some of the same benefits. Although various forms of relaxation training, meditation, or yoga have considerable face validity for addressing the distress associated with a chronic condition, findings on benefits for patients with asthma have remained quite variable, and high-quality large-scale clinical trials are lacking. This is not surprising given that the rationale for such techniques is unclear. Relaxation aims to reduce sympathetic activity and increase parasympathetic activity, both of which support bronchoconstriction, and may therefore be contraindicated.

Interoception training can address a potentially dangerous lack of awareness of airway obstruction in patients with asthma. It is administered in protocols of systematic exposure to added resistive loads that vary in the extent to which they impede airflow. Patients learn to better distinguish different loads or to lower their perceptual thresholds for just noticeable loads.

Lifestyle Interventions in Asthma

Lifestyle interventions promote health-related behaviors including eating a nutritious diet, improving sleep hygiene habits, and regularly exercising for asthma. Dietary interventions influencing inflammatory pathways in patients with asthma are promising, with emerging research suggesting associations between nutrient-rich food intake and better asthma control. Obesity is known for causing complications in asthma, making weight control interventions an important treatment modality for affected patients. Sleep interventions have been found to benefit sleep quality, but little is known about effects on asthma symptomology. Physical activity interventions suggest some improvements in reduced asthma symptomology, with some studies suggesting reduction in exercise-induced asthma and systemic inflammation, although studies vary regarding type of exercise, outcome measures, and lengths of interventions.

CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD)

Chronic obstructive pulmonary disease (COPD) is a group of progressive and irreversible lung diseases (e.g., emphysema and chronic bronchitis) in adults that lead to progressive reduction in airflow in the lungs.

Common symptoms include chronic cough and dyspnea, breathlessness that is continuous, often frightening, and is associated with considerable limitations on daily activities. COPD is strongly associated with tobacco smoking in approximately 80% of patients, but environmental exposures, in particular fumes from biomass fuel stoves and open fireplaces, are prominent risk factors in developing countries. Genetics may also play a role in the development of COPD. The prevalence of COPD in adults is approximately 13%. COPD is currently a leading cause of death worldwide with an increasing trend in prevalence for adults over 40 years of age.

Diagnosing COPD involves obtaining patient history, symptoms, current and past tobacco use, information on exposure to fumes and other irritants, and pulmonary function testing. A diagnosis of COPD is typically made around 40 years of age or later when symptoms from exposure begin to develop.

Pharmacologic treatment of COPD involves bronchodilator medication, antibiotics, or inhaled corticosteroids. Oxygen therapy, lung volume reduction surgery, or lung transplantation may be necessary in severe cases or later stages. Individuals with COPD experience increased rates of psychiatric problems, and experiencing psychiatric problems is also linked to greater rates of a later COPD diagnosis. Multidisciplinary rehabilitation is not only focused on physical exercise, but also includes education and self-management skills, nutritional components, and psychosocial support. It is associated with improved exercise tolerance, reduced exacerbations, and improved dyspnea and quality of life.

Comorbidity with Psychiatric Disorders

The prevalence of depression in COPD ranges from about 25% to 50%. Greater disease severity, dyspnea, female sex, and disability appear to be risk factors. When depression is present in COPD, it is associated with reduced quality of life and poorer adherence to pulmonary rehabilitation programs. The presence of depression may be more strongly associated with functional impairment than lung disease severity.

Although the results are somewhat inconsistent, cognitive behavioral treatment with or without pharmacologic therapy appears to improve depressive symptoms in people with COPD but has modest effects on disease-related outcomes. Depressive symptoms are related to poorer disease management and reductions in other health behaviors, including physical activity in individuals with COPD. Pulmonary rehabilitation, particularly if it includes stress reduction or psychotherapy, appears to be associated with improvement in depression.

Anxiety disorders, particularly panic disorders, are much more common in patients with COPD than in the general population. As with depression, anxiety appears to be associated with the poor functional status. The treatment of anxiety in COPD has been the topic of few studies. A potential concern with the use of benzodiazepines is decreased respiratory drive.

Sleep disorders are increasingly common among individuals with COPD. Insufficient sleep has been related to worse health outcomes for patients, including hospitalization, in particular sleep-disordered breathing. Comorbidities, such as metabolic and cardiovascular disease, are seen in patients with comorbid COPD. Obstructive sleep apnea also occurs at higher rates, which, if untreated, exacerbates health issues associated with COPD. Insomnia symptoms are present in around 3 of 10 patients. Worse sleep quality in COPD has been linked to greater anxiety and depression in a community sample of adults, however, relations between improvements in sleep, psychiatric presentation, and COPD outcomes remain to be studied.

Cognitive impairment (e.g., working memory and attention) is also common in COPD. The cognitive deficits appear to worsen with pulmonary disease progression and comorbid obstructive sleep apnea, emphasizing the importance of early intervention and targeting comorbid sleep disorders to preserve function. Some data suggest improvement in cognition with pulmonary rehabilitation.

In psychiatric assessments of COPD patients, it is important to consider that the majority have a significant smoking history. Tobacco use is associated with increased rates of a variety of psychiatric disorders. The increased risk of both mood and anxiety disorders appears in large part to be explained by nicotine dependence. Furthermore, rates of tobacco use are much higher patients with major psychiatric disorders, particularly schizophrenia, than in the general population. Thus, people with psychiatric illnesses are at an increased risk of developing COPD. Therefore, the etiology of the elevated rates of psychiatric disorders in COPD may be particularly complex and may reflect effects of hypoxia, functional impairment, feelings of breathlessness, as well as preexisting vulnerability related to tobacco use.

Adjunctive Cognitive and Behavioral Treatment Modalities for COPD

CBT for anxiety has been shown to reduce anxiety and healthcare utilization of patients with COPD. Short-term CBT has shown promise in helping to reduce psychological and physical symptoms, and depressive symptoms have been reduced in fewer sessions than anxiety symptoms. Learning CBT skills to manage anxiety and dyspnea symptoms as part of treatment may have the potential to improve COPD self-monitoring and disease management. Given the mostly older age of patients, with some approaching end of life prematurely, qualitative research has also emphasized the need for therapies revolving around acceptance and existential themes.

Lifestyle Interventions in COPD

Regular physical activity is an encouraged practice for those suffering from COPD. Exercise programs remain limited primarily to pulmonary rehabilitation, but physical activity interventions should also be embedded in daily life, although longitudinal improvements in health and mood are yet unknown. Although patients are encouraged by physicians to be more physically active, specific exercise prescriptions or individualized training programs are rare. Pulmonary rehabilitation including exercise components can reduce anxiety, depression, and cognitive function. Yoga and water-based activity have also been explored as promising exercise programs for COPD.

Although smoking is the major cause of COPD, many patients do not quit smoking after diagnosis and continue despite exacerbation of their symptoms. Smoking cessation programs thus remain a critical avenue for intervention, with counseling and nicotine replacement therapy as options as supportive elements.

Although comorbid sleep conditions can worsen COPD and other health outcomes, little research has assessed sleep interventions in this population. Continuous positive airway pressure treatment to manage obstructive sleep apnea can lead to improvements in health status, suggesting that a referral to a sleep specialist should be prioritized when indicated in psychiatric care. The effectiveness of psychoeducation on sleep hygiene and interventions for insomnia in COPD remain to be explored.

INTERSTITIAL LUNG DISEASE

Interstitial lung disease includes a diverse group of illnesses of the interstitium surrounding the alveoli of the lungs. The illnesses are characterized by progressive scarring and fibrosis of the lungs. Examples include sarcoidosis, asbestosis, hypersensitivity pneumonitis, drug reactions, and even infections including COVID-19. Progressive worsening of dyspnea is a characteristic feature. These illnesses can be idiopathic, familial, or result from a variety of environmental exposures or certain medications. The prognosis is poor. No treatments are presently available that substantially alter the course of interstitial lung diseases.

The literature on the nature or prevalence of psychiatric disorders in interstitial lung disease is modest. High rates of depression and anxiety that are related to the severity of dyspnea are reported in patients with idiopathic pulmonary fibrosis.

CYSTIC FIBROSIS

Physiology and Manifestations

Cystic fibrosis (CF) is an autosomal recessive disease that is characterized by lung and digestive symptoms and affects approximately 75,000 individuals worldwide. It is the most common genetic disease among White individuals, and it is caused by a mutation on chromosome 7 that leads to malfunction of the cystic fibrosis transmembrane regulator (CFTR) protein, which facilitates movement and concentration of chloride ions across the cell membranes of the sweat glands, sinuses, lungs, pancreas, gastrointestinal tract, and reproductive organs.

Individuals with CF experience a wide range of varying symptoms, including recurrent cough and lung infections, wheezing or shortness of breath, salty skin, sinus infections, pancreatic insufficiency, poor nutrition and growth, and difficulty with digestion and bowel movements, resulting in significant variations in individual presentations of the disease. Over time, recurrent respiratory infections and bacterial colonization result in gradual loss of lung function, and respiratory failure is the most common cause of mortality. In adulthood, individuals are at increased risk of the development of CF-related diabetes due to deterioration of the pancreas and osteopenia or osteoporosis as a result of poor nutrition and bone density.

Diagnosis and Treatment

Diagnosis of CF is commonly made through the results of a sweat test that examines the salt concentration of an individual's sweat. Due to the malfunction of the CFTR protein, individuals with CF have abnormally high concentrations of chloride in their sweat. Additionally, genetic testing can be used to make a definitive diagnosis and to identify the individual's specific mutations.

Currently, there is no cure for CF. Treatment is aimed at slowing the progression of the disease and preventing infection. Since discovery of the CFTR protein in the late 1980s, medical treatment of CF has progressed rapidly and life expectancy for those with the disease has likewise increased dramatically over the years, with treatment including CFTR modulators that improve functioning of the protein and lung transplantation when severity progresses. CFTR modulators are effective for specific mutations. White individuals appear to have the mutation susceptible to treatment at higher rates than non-White patients, contributing to a health disparity in care across racial groups. Non-White individuals with CF report worse health outcomes, including worse lung function in Black individuals.

Treatment of CF involves an extremely complex and time-consuming daily regimen that may include airway clearance, aerosolized medication, nutritional support, dietary supplements, and other medications, such as antibiotics and anti-inflammatories. Patients in North America are often followed by a CF center, accredited by the Cystic Fibrosis Foundation, which involves care from a team of specialists, including pulmonologists, nurse practitioners, respiratory therapists, dieticians, physical therapists, and social workers. They are routinely monitored every 3 to 4 months through their clinics in order to adjust treatment approaches as necessary.

Self-Management and Psychosocial Considerations

As with many chronic illnesses, self-management in CF is a vital component of medical maintenance of the disease and the patient's overall quality of life. Effective self-management involves many skills, including adherence to airway clearance and other respiratory treatments, monitoring of symptoms of exacerbations, and regular coordination with the CF care team. Research suggests that adherence to different aspects of self-management in CF varies, with the lowest rates of adherence often found with chest physiotherapy recommendations.

Psychosocial considerations for health care professionals treating patients with CF will include a variety of issues, depending on the patient's overall level of functioning, medical status, and phase of life. Common psychosocial issues impacting quality of life in CF include pain management, lost time at work or school due to doctor's appointments and attending to disease management, management of pain, transitions from child to adult CF care, fertility concerns, coping with disease management (including consideration of end-of-life care and lung transplantation during severe progression of the disease), and psychiatric conditions. Coordination of care during childhood will necessitate consideration of family dynamics and the mental health of adult caregivers. Overall family functioning has been found to significantly impact quality of life and health status in patients with CF.

Emerging Adulthood/Transition

The growing adult population has necessitated the development of adequate protocols for transition of patients to adult CF care. This transitional phase has yet to be thoroughly addressed by existing research and available recommendations are uncommon and often vague. This transition is an adjustment for many patients who have had prolonged care with a single multidisciplinary team for several years. Continuous and effective communication between the medical providers and the patient is recommended during the stage of transition in order to ensure proper continuation of care and well-being of the patient.

Fertility Issues. Fertility issues and family planning may become matter of concern as patients with CF enter adulthood. Most males with CF lack a vas deferens, causing infertility, and a significant portion of females may also experience fertility problems. Additionally, genetic counseling may be warranted as patients consider whether their offspring will be carriers of the disease or potentially have the diagnosis itself.

Psychiatric Comorbidities

Rates of depression in CF vary by age and type of screening tool being used. On average, 5% to 19% of adolescents and 13% to 29% of adults with CF reported symptoms of depression. Overall, prevalence rates for symptoms of anxiety range from 14% to 35% across different countries, with an average of 30% of patients reporting symptoms. Additionally, a high degree of comorbidity occurs between anxiety and depression, with those reporting anxiety symptoms being from 13.6 times (adults) to 15.0 times (adolescents) more likely to also report symptoms of depression than those without elevations in anxiety.

Elevated rates of depression and anxiety are also reported in caregivers of patients with CF, with prevalence rates ranging from 48% (mothers) to 36% (fathers) for anxiety and from 37% (mothers) to 31% (fathers) for depression. When caregivers are experiencing elevated symptoms of depression or anxiety, adolescents with CF are more likely to report symptoms.

In patients with CF, higher rates of depression and anxiety are negatively associated with health status and may have several negative outcomes, such as poor treatment adherence, increased health care costs, lower lung function, and decreased quality of life. Sleep disorders and higher BMI have also been linked to higher levels of depression and anxiety in CF. Careful and regular monitoring of psychological functioning is recommended in order to provide appropriate intervention or mental health referrals as needed. Prescription of psychiatric medications would require careful consideration, particularly with regard to side effects, dosages, and potential interactions with other medications that patients are prescribed. The need to cope with symptoms, treatment, and progression of the disease can also increase the likelihood of a psychiatric presentation.

Psychological Interventions

Psychological interventions for patients with CF may target several different outcomes, including education about their disease, emotional functioning, quality of life, treatment adherence, airway clearance techniques, and dietary intake. Doubts about the effectiveness of medication for management of CF are related to depressive symptoms in patients. Previous psychological interventions have primarily been educational or behavioral in nature and target specific treatment concerns rather than broader psychosocial functioning. There were limited significant effects, including CBT for nutrition adherence and lung transplant planning.

Few studies have examined the importance of factors such as social support. There is limited evidence that social support can encourage more mental wellness in CF. Potential cross-infection is a barrier that limits in-person interaction between individuals with CF; online social support interventions could be promising in relation to psychiatric functioning and health outcomes. Interventional research in this area has lagged in comparison to the medical advances in treatment and disease management. Increased attention is needed to develop effective interventions that improve treatment adherence, quality of life, and independent self-management in this population, given that better treatment options have prolonged life expectancy significantly beyond adolescence.

Lifestyle Interventions in CF

Limited findings suggest that physical activity in CF may provide benefits to psychological functioning and lung function. Physical activity appears to provide benefits for mental health in CF when beneficial to patients' overall health, indicating a need to also explore contraindications (e.g., severe CF symptoms). Exercise programs can also alter patients' perceptions of their health. Given gastrointestinal tract complications in CF, maintaining a healthy diet may be important, although the effects of dietary intake on psychiatric symptoms remain to be investigated.

RESPIRATORY FAILURE

Over one million persons in the United States are admitted to intensive care units annually due to respiratory failure. Respiratory failure can be a consequence of a variety of lung diseases. As with other serious illnesses, delirium is common with respiratory failure. Both hypoxia and hypercapnia can occur in respiratory failure and in severe cases, mechanical ventilation is needed. Therefore, the psychiatric syndromes observed in this illness include those of severe lung disease as well as depression and post-traumatic stress disorder or other anxiety disorders related to mechanical ventilator support and the intensive care unit. Some data suggest that the administration of corticosteroids may decrease the risk of developing post-traumatic stress disorder during an intensive care unit admission.

SOMATOFORM RESPIRATORY DISORDERS

A variety of dysfunctional and distressing respiratory regulation abnormalities have been identified in adults and children under various terms in past decades. Typically, diagnosis has relied heavily on exclusion of somatic diseases and a lack of demonstrable pathophysiologic processes, which can produce most of the same symptoms. The adequacy of these disorders as diagnostic entities has often been debated. Efforts have been made more recently to distill common characteristics of these conditions and formulate positive diagnostic criteria under the umbrella of *somatoform respiratory disorders*. These would distinguish subgroups of functional disorders of respiratory regulation (e.g., vocal cord dysfunction), stereotypic respiratory disorders (e.g., habit cough), and psychogenic respiratory disorders (e.g., suffocation anxiety). However, the label “psychogenic” has become increasingly obsolete as contemporary neuroscience demonstrates tight associations between mental, neurophysiologic, peripheral autonomic, and somatic processes in health and disease. Common to all conditions is that a psychiatric background and specific behavioral interventions and/or psychotherapy are often suggested, but the available evidence is mostly limited to case reports and uncontrolled studies. Although the patients’ quality of life is often reduced, the prognosis is good to excellent with adequate treatment. Reassurance that the symptoms are not indicative of a somatic disease and that the symptoms are treatable already provides significant relief for patients and their families.

Chronic Hyperventilation

Hyperventilation is defined as breathing in excess of metabolic demand leading to a fall in arterial PCO_2 levels. The respiratory pattern is often deep and/or fast, leading to exaggerated minute ventilation relative to demands from physical activity. As a consequence, patients complain about symptoms of dizziness, headaches, chest pain, peripheral paresthesia (lips and limbs), and anxiety. Despite overbreathing, patients have paradoxical sensations of air hunger and shortness of breath. Levels of PCO_2 <30 mm Hg have been proposed as an indicator of unequivocal chronic hypocapnia, but many of patients’ symptoms are less tightly linked to PCO_2 levels than previously thought, which is not surprising given the complexity of regulation of respiratory sensations. Whereas acute respiratory hyperventilation leads to disturbances in acid–base balance with high pH levels (acute respiratory alkalosis), pH levels in chronic hyperventilation are often near-normal due to compensatory bicarbonate changes.

Triggers of hyperventilation are often emotional, and hyperventilation is most frequently observed in panic disorder. However, chronic hyperventilation is also found without evidence of anxiety disorders (idiopathic hyperventilation). Breathing training focused on reducing hypocapnia (hypoventilation training) is the mainstay of therapy. Slow, shallow, abdominal, and nasal breathing is ideally supported by measurement and feedback of actual PCO_2 values. Rebreathing techniques using bags are discouraged because they reduce oxygenation and may trigger panic by raising CO_2 levels. Patients profit from identifying individual triggers, decatastrophying, and reattribution of benign symptoms, and education about hyperventilation and anxiety.

Vocal Cord Dysfunction

Patients with vocal cord dysfunction suffer from sudden, intermittent adduction (narrowing or closure) of the vocal cords during inspiration. The inspiratory occurrence is most common and paradoxical, but obstruction can also be observed during expiration or both phases of the breathing cycle. Proposed positive diagnostic criteria are presentation with sudden onset of dyspnea, problems with inspiration, intermittent symptoms that resolve spontaneously after a few minutes, and symptoms self-localized in the throat. Diagnostic confirmation in the acute stage can be obtained by flexible laryngoscopy (closure of vocal cords >50%) or abnormal inspiratory and/or expiratory features of the flow-volume loops in spirometry.

Episodes are often associated with considerable fear of suffocation and panic, and anticipatory anxiety of an additional attack is common. A variety of triggers can be identified, including stressful life events, anxiety, exercise, irritants, gastro-esophageal reflux, respiratory infections, postnasal drip, or olfactory stimuli. A psychiatric background can be found in up to 75% of patients and anxiety disorders, a history of early life traumatic experiences, or abuse are not uncommon. The condition can co-occur with asthma or mimic it, complicating differential diagnosis and treatment and leading to extensive and costly measures. Unresponsiveness to bronchodilator therapy and lack of airway hyper-reactivity can help exclude asthma. The recommended treatment is ideally multidisciplinary and may include speech and behavioral therapy. Underlying somatic conditions such as reflux, infections, or nasal allergies require adequate medication treatment.

Chronic Cough Disorder (Habit Cough and Psychogenic Cough)

Cough is an integrated evoked, reflexive, or voluntary motor pattern that has important functions in airway protection and clearance. It is elicited by irritation of the vocal cords or airways and is a nonspecific symptom of a range of diseases and conditions, including respiratory infections, asthma, COPD, gastroesophageal reflux, allergic rhinitis, and postnasal drip. Chronic cough that is persistent, without evidence for any of these causes, has often been labeled as habit cough or psychogenic cough. It is most frequently reported in children and adolescents. The disruptive nature of the phenomenon can have negative social consequences. Reduction or cessation during the night or with distraction has also been frequently observed, although use of these clinical presentations as defining criteria has been debated. A recent expert panel advocated replacing habit cough with “tic cough,” aligning it with tic disorders as it shares their central features of suppressibility, distractibility, suggestibility, and variability, as well as the presence of a sensation of urge (premonitory sensation), regardless of whether the cough is the only tic or one of many. The panel also favored the term *somatic cough disorder* to replace psychogenic cough, placing it within the DSM-*5* framework of somatization disorder, in which a psychological distress is expressed as a physical symptom. The diagnosis should only be given if a tic disorder and other causes can be excluded, and the patient meets criteria for DSM-5 somatization symptom disorder. Somatic cough disorder and other psychogenic and functional breathing disorders appear to be associated with high rates of psychiatric comorbidity including ADHD, tic disorders, and phobias. Hypnosis and suggestion therapy are the most successful treatment modalities for somatic cough disorder and tic cough, in particular in children.

Sighing Dyspnea

A sigh breath is characterized by a sudden increase in depth, and often also length, of the breathing cycle. It occurs intermittently and is usually limited to a single breath. There is no clear convention as to the magnitude of volume change necessary, with some sources requiring a full inspiration to inspiratory capacity and others requiring 2 to 2.5 times the regular tidal volume. The clinical manifestations of sighing dyspnea include a more frequent rate of sighing, which is often episodic and accompanied by sensations of breathlessness or chest pain, and sometimes also yawning. Most observations have been in children, adolescents, or young adults. Proposed positive criteria are self-limited sighing with intensified inspiration, normal respiration rate, perceived respiratory effort greater than sensation of respiratory distress, and no nighttime awakening from sighing. Elevated sigh breathing has also been observed in adult anxiety patients, particularly in panic, and is thought to maintain hypocapnia. Breathing training focused on regularity of breathing (without suppressing occasional sighs) and reducing hyperventilation appears as the most relevant primary treatment. For psychotherapy and anxiolytic drug therapy, a psychiatric background should be clearly established in the individual case.

PSYCHIATRIC SYMPTOMS ASSOCIATED WITH RESPIRATORY DISEASE MEDICATIONS

As illustrated by the representative case vignette at the beginning of this section, medications used to treat respiratory illnesses can have effects on the brain and the patients' experience. Oral corticosteroids such as prednisone are frequently prescribed for patients with lung diseases. The use of these medications ranges from brief "bursts" for asthma exacerbations or other acute and reversible pulmonary changes to chronic therapy that can last for years for several lung diseases. Long-term use of corticosteroids is associated with a variety of systemic side effects, including suppression of the hypothalamic–pituitary–adrenal axis, glaucoma, cataracts, weight gain, infections, and osteoporosis. These side effects are sometimes termed iatrogenic Cushing syndrome because of the frequent development of truncal obesity, moon faces, and buffalo hump. Inhaled corticosteroids are also frequently used for some lung diseases. While these corticosteroid preparations are designed to be locally acting, when used at high doses for extended periods, they can be associated with many of the side effects observed with oral corticosteroids.

Shortly after the widespread use of corticosteroids began in the early 1950s, reports of psychiatric symptoms from their use began to appear in the literature. A Hollywood movie, “Bigger than Life,” was even released that provided a fictionalized account of the potential for corticosteroids to induce psychiatric side effects and even lead to abuse.

The use of corticosteroids is associated with an increased risk of mood symptoms that include depression, irritability, mania, and mixed features. Acute, higher-dose corticosteroid therapy, as with bursts for asthma exacerbations, may be more strongly associated with mania or hypomania while depression may be more common with chronic, lower-dose therapy. Psychotic features, such as delusions and hallucinations, can also occur with corticosteroids.

Perhaps even more common than mood symptoms are changes in cognition. In animal models, the hippocampus appears to be particularly sensitive to the effects of corticosteroids. Consistent with effects on the hippocampus, memory appears to be a cognitive domain that shows changes with corticosteroid therapy. In healthy controls, even a single high-dose exposure to corticosteroids is associated with a decline in declarative and working memory performance. Structural brain changes, including a reduction in hippocampal volume, appear to accompany memory changes, even during very brief corticosteroid administration. More generalized cognition changes, as with delirium, also occur during corticosteroid therapy, especially in older persons.

The most clearly established risk factor for psychiatric side effects with corticosteroids is dose. Thus, the lowest effective dose of these medications should be used when treating lung diseases or other conditions. If the dose cannot be reduced or if this approach is ineffective, the treatment, like the symptoms, is similar to that of bipolar disorder. Although placebo-controlled trials have not been conducted, mood stabilizers such as lithium, anticonvulsants, and antipsychotics seem to ameliorate mood changes and psychosis with corticosteroids.

Other medications used for lung diseases may also have psychiatric side effects. Anxiety, agitation, and depression have been reported with theophylline and beta-2 adrenergic agonists. Theophylline may also be associated with an increased risk of suicidal ideation. Postmarketing reports suggested a possible link between leukotriene inhibitors and a variety of psychiatric symptoms, including agitation, insomnia, anxiety, depression, and suicidal ideation. However, the data do not appear to support an association of these medications with completed suicide with a recent report even suggesting a decreased risk of depression with montelukast as compared to inhaled corticosteroids.

FUTURE DIRECTIONS

The psychiatric symptoms associated with respiratory illnesses have been less extensively studied than those in other common medication conditions, such as coronary artery disease. The underlying mechanisms are not well understood but are likely multifactorial. Furthermore, little is known about the treatment for depression, anxiety, and other disorders in this population. The increasing prevalence of respiratory illnesses, in particular, COPD and asthma, and current population trends in aging, pose considerable challenges to scientists and clinicians. In addition to psychological interventions for improving mental health symptoms, lifestyle interventions that may aid in managing psychiatric health along with symptoms of respiratory disease remain to be investigated. Potential underlying processes involved such as systemic inflammation and gut microbiota in dietary interventions, for example, are critical to explore and remain understudied.

Interdisciplinary collaboration will be needed to further elucidate the interaction of psychosocial factors with respiratory illness and to develop and implement adequate pharmacologic, cognitive, and behavioral interventions. Consideration of electronic methods of intervention including mobile applications and telemedicine for mental health will need to be explored, with the potential benefit of increasing accessibility to treatments aimed at improving patient care.

FURTHER READINGS

Abbott J, Elborn JS, Georgiopoulos AM, et al. Cystic Fibrosis Foundation and European Cystic Fibrosis Society survey of cystic fibrosis mental health care delivery. *J Cyst Fibros*. 2015;14:533–539.

Akula M, Kulikova A, Khan DA, Brown ES. The relationship between asthma and depression in a community-based sample. *J Asthma*. 2018;55:1271–1277.

Brown ES, Sayed N, Van Enkevort E, et al. A randomized, double-blind, placebo-controlled trial of escitalopram in patients with asthma and major depressive disorder. *J Allergy Clin Immunol Pract*. 2018;6:1604–1612.

Caldera-Alvarado G, Khan DA, Defina LF, Pieper A, Brown ES. Relationship between asthma and cognition: the Cooper Center Longitudinal Study. *Allergy*. 2013;68:545–548.

Capel LH. A history of the pulmonary catarrhs and airways obstruction. *Br J Dis Chest*. 1966;60:113–119.

Carlson SM, Kim J, Khan DA, et al. Hippocampal volume in patients with asthma: results from the Dallas Heart Study. *J Asthma*. 2017;54:9–16.

Cooley C, Park Y, Ajilore O, Leow A, Nyenhuis SM. Impact of interventions targeting anxiety and depression in adults with asthma. *J Asthma*. 2020;28:1–24.

Drake K, Palmer AD, Schindler JS, Tilles SA. Functional outcomes after behavioral treatment of paradoxical vocal fold motion in adults. *Folia Phoniatr Logop*. 2017;69:154–168.

Gada E, Khan DA, DeFina LF, Brown ES. The relationship between asthma and self-reported anxiety in a predominantly healthy adult population. *Ann Allergy Asthma Immunol*. 2014;112:329–332.

Grüber C, Lehmann C, Weiss C, Niggemann B. Somatoform respiratory disorders in children and adolescents-proposals for a practical approach to definition and classification. *Pediatr Pulmonol*. 2012;47:199–205.

Havermans T, Willem L. Prevention of anxiety and depression in cystic fibrosis. *Curr Opin Pulm Med*. 2019;25(6):654–659.

*Holland AE, Fiore JF Jr, Bell EC, et al. Dyspnoea and comorbidity contribute to anxiety and depression in interstitial lung disease. *Respirology*. 2014;19:1215–1221.

*Jain A, Lolak S. Psychiatric aspects of chronic lung disease. *Curr Psychiatry Rep*. 2009; 11:219–225.

*Judd LL, Schettler PJ, Brown ES, et al. Adverse consequences of glucocorticoid medication: psychological, cognitive, and behavioral effects. *Am J Psychiatry*. 2014;171:1045–1051.

Kaptein AA, Klok T, Moss-Morris R, Brand PL. Illness perceptions: impact on self-management and control in asthma. *Curr Opinion Allergy Clin Immunol*. 2010;10:194–199.

Kavanagh J, Jackson DJ, Kent BD. Sleep and asthma. *Curr Opin Pulm Me*. 2018;24:569–573.

Meuret AE, Ritz T. Hypocapnia in anxiety and respiratory disease: empirical evidence and clinical strategies. *Int J Psychophysiol*. 2010;78:68–79.

Nyenhuis SM, Dixon AE, Ma J. Impact of lifestyle interventions targeting healthy diet, physical activity, and weight loss on asthma in adults: what is the evidence? *J Allergy Clin Immunol Pract*. 2018;6:751–763.

Pakhale S, Luks V, Burkett A, Turner L. Effect of physical training on airway inflammation in bronchial asthma: a systematic review. *BMC Pulm Med*. 2013;13:38.

Quittner AL, Saez-Flores E, Barton JD. The psychological burden of cystic fibrosis. *Curr Opin Pulm Med*. 2016;22(2):187–191.

Ritz T, Kroll JL, Aslan S, et al. Subcortical gray matter volumes in asthma: associations with asthma duration, control, and anxiety. *Brain Imaging Behav*. 2020;14:2341–2350.

Ritz T, Meuret AE, Trueba AF, Fritzsche A, von Leupoldt A. Psychosocial factors and behavioral medicine interventions in asthma. *J Consult Clin Psychol*. 2013;81:231–250.

Sansing-Foster V, Haug N, Mosholder A, et al. Risk of psychiatric adverse events among montelukast users. *J Allergy Clin Immunol Pract*. 2021;9:385–393.e12.

Trueba AF, Ritz T. Stress, asthma, and respiratory infections: pathways involving airway immunology and microbial endocrinology. *Brain Behav Immun*. 2013;29:11–27.

Vanfleteren LE, Beghe B, Andersson A, Hansson D, Fabbri LM, Grote L. Multimorbidity in COPD, does sleep matter? *Eur J Intern Med*. 2020;73:7–15.

Verma S, Cardenas-Garcia J, Mohapatra PR, Talwar A. Depression in pulmonary arterial hypertension and interstitial lung diseases. *N Am J Med Sci*. 2014;6:240–249.

Vertigan AE, Murad MH, Pringsheim T, et al. Somatic cough syndrome (previously referred to as psychogenic cough) and tic cough (previously referred to as habit cough) in adults and children: Chest guideline and expert panel report. *Chest*. 2015;148(1):24–31.

von Leupoldt A, Fritzsche A, Trueba AF, Meuret AE, Ritz T. Behavioral medicine approaches to chronic obstructive pulmonary disease. *Ann Behav Med*. 2012;44:52–65.

Ye G, Baldwin DS, Hou R. Anxiety in asthma: a systematic review and meta-analysis. *Psychol Med*. 2021;51:11–20.

Zeng Y, Jiang F, Chen Y, Chen P, Cai S. Exercise assessments and trainings of pulmonary rehabilitation in COPD: a literature review. *Int J Chron Obstruct Pulmon Dis*. 2018;13:2013–2023.

27.6 Diabetes: Psychosocial Issues and Psychiatric Disorders

Paul Ciechanowski, M.D., M.P.H., and Robert J. Boland, M.D.

Diabetes is one of the most complex chronic medical conditions, requiring a high degree of patient self-management within a supportive social network to ensure optimal outcomes. Psychiatric illness can adversely affect patients' self-care behaviors and ability to interact with others. Psychiatric illnesses, such as depression, may also lead to neurohormonal changes, such as changes in the hypothalamic–pituitary axis, which may adversely affect glucose control.

DIABETES: THE NATURE OF THE ILLNESS

Diabetes results from insufficient insulin secretion or resistance to insulin. Persons with type 1 diabetes produce no or minimal insulin; the onset is typically early in life. Type 1 diabetes accounts for 5% to 10% of all patients with diabetes. Type 1 diabetes is due to an autoimmune, predominantly T-cell process that selectively destroys insulin-producing pancreatic β cells. This autoimmune process likely has a multifactorial origin with environmental and genetic factors. The onset of type 2 diabetes generally occurs during adulthood, although childhood onset is increasing. Individuals with type 2 diabetes usually have insulin resistance, but as they age, they have decreased insulin production, with many eventually becoming insulin dependent. Genetic vulnerability and obesity are the most common causes of this condition, and dietary restriction and exercise are often sufficient early-stage treatments. In later stages, patients usually require oral hypoglycemic medications and insulin for adequate control.

Individuals with diabetes are at significantly increased risk of heart disease, stroke, peripheral vascular disease, kidney disease, lower extremity ulcers and amputations, retinal disease, neuropathy, infections, digestive disease, and periodontal disease. The Centers for Disease Control and Prevention (CDC) estimates that the cost of diabetes in the United States is $327 billion annually due to medical costs and lost work and wages. Diabetes is the seventh leading cause of death in the United States.

Obesity Epidemic

There is an epidemic of obesity and type 2 diabetes worldwide, including in the United States. The CDC estimates that as many as 37 million Americans have diabetes, which is more than triple the prevalence in 1980. They also report that from 1999 to 2000 through 2017 to 2020, the obesity prevalence increased from 30.5% to 41.9%. In addition, there are differences in the genetic predisposition for type 2 diabetes among different ethnicities. For example, per the CDC, non-Hispanic Black adults had the highest age-adjusted prevalence of obesity (49.9%), followed by Hispanic adults (45.6%), non-Hispanic White adults (41.4%), and non-Hispanic Asian adults (16.1%).

The epidemic of obesity in the United States has also caused a marked increase in the prevalence of metabolic syndrome. This prediabetic syndrome includes central obesity, elevated blood glucose, elevated blood pressure, low–high-density lipoproteins (HDL), and elevated triglycerides. Elevated blood glucose is related to increased insulin resistance, and in the United States, over 96 million U.S. adults (1 in 3 adults 18 or older) have “prediabetes.” Prospective studies have shown that metabolic syndrome raises the risk of developing coronary vascular disease by over 50% and the risk for type 2 diabetes by approximately threefold compared to individuals without metabolic syndrome.

PSYCHIATRIC COMORBIDITY

Thomas Willis, an early British physician, stated in the 17th century that diabetes is caused by “sadness or long sorrow and other depressions and disorders.” However, despite the early recognition of this association, it was not until the latter part of the 20th century that several epidemiologic studies began to address this problem.

Epidemiology

These studies demonstrated that depression is more likely to occur in people with diabetes regardless of whether the person is aware of having diabetes. Many others who do not have a depressive disorder experience considerable distress, putting them at risk for many psychiatric disorders. Multiple studies report a prevalence of depressive disorders in people with diabetes at about 10% to 15%, thus, roughly twice that of the average population. Perhaps as many as 25% of persons with diabetes experience subsyndromal depressive symptoms. Of note is that some recent studies suggest that this rate is increasing. This possibility is especially alarming as many cases remain undiagnosed. Specific populations may be harder hit. For example, in the United States, one review suggested that a quarter of African Americans with diabetes have depression.

A meta-analysis of 39 studies reported a 12% prevalence of major depression in patients with diabetes. This meta-analysis also showed that in controlled studies, there was a twofold increased prevalence of depression in patients with diabetes compared to controls. A large health maintenance organization-based study found that 12% of approximately 4,800 adult patients with predominantly type 2 diabetes (96%) met the criteria for major depression and 8.5% for minor depression. The group with major depression and diabetes compared to those with diabetes alone were younger, more likely to be female, were less educated, had significantly higher medical comorbidity, had more diabetes complications, had a longer duration of diabetes, had higher hemoglobin A1c (HbA1c) and BMI values, and were more likely to be smokers. In addition, the patients with depression and diabetes developed diabetes an average of 5 years earlier than those with diabetes alone.

Depression tends to be either a chronic or recurrent condition in patients with diabetes. For example, data from the large HMO study described above showed that 63% of patients with diabetes and major depression had 2 years or longer of significant depressive symptoms. Another study found that 79% of patients with diabetes with major depression relapsed over a 5-year follow-up period with a mean of four episodes per patient.

Diabetes and Depression in Non-Western Countries. Most of the data on the prevalence of diabetes and depression is from Western and higher-income studies. However, some data from other parts of the world suggest an even higher prevalence of depression in patients with diabetes. For example, a review of studies done in South Asia reported that the prevalence of type 2 diabetes and elevated depression symptoms ranged from 11.6% to 67.5%. A meta-analysis of studies done in Iran estimated that the prevalence of depression in patients with diabetes was 54%. In Saudi Arabia, a study of children with type 1 diabetes reported a depression prevalence of 27%. Investigators in India found a similarly high rate of depression in patients with diabetes. In addition, the investigators found that depression was associated with younger age, female sex, low socioeconomic status, and poor self-management.

Diabetes and Depression in Youth. Less research has focused on depression among youth with diabetes. The prevalence of elevated depressive symptoms in youth with type 2 diabetes is approximately 20% (this includes syndromal and subsyndromal cases). Some studies suggest that these symptoms adversely affect glycemic control, most likely because of their adverse effects on medication adherence and diet. Fewer studies have examined the effect of comorbid depression on youth activity or physiology.

A study examined the rate of depression in 2,672 youth (aged 10 to 21 years), most of whom had type 1 diabetes. Based on a depression screening measure, 14% were mildly depressed, and 8.6% were moderately to severely depressed. In addition, participants with depression and diabetes had significantly higher HbA1c levels and a higher number of emergency room visits. Another study from an extensive national database found that youth hospitalized with type 1 diabetes with comorbid depression were twice as likely to be rehospitalized in the following 12-month period compared to those with a diabetes diagnosis alone.

Depression and Diabetes May Share Similar Biologic Mechanisms

The close relationship between diabetes and depression suggests that the two disorders may share common mechanisms. Most commonly, experts suggest that the link between the two is psychological: the burden of a chronic disease such as diabetes predisposes to depression, or depression affects motivation and causes poorer self-management, for example. However, there may also be shared biologic mechanisms. Hyperglycemia, for example, can affect mood states, and high plasma glucose in the brain can increase oxidative stress, ultimately leading to neuronal damage, which causes depression. Alternatively, the microvascular changes associated with diabetes and identifiable in the brain may also predispose to depression. Finally, we increasingly appreciate the role of inflammation in psychiatric disorders, including depression. Type 2 diabetes is associated with low-grade systemic inflammation, and this same inflammation may underly the depression seen in diabetes. Others have suggested calcium/cAMP signaling as the underlying cellular mechanism linking the two disorders. Unfortunately, most of the work investigating this to date is cross-sectional and limited to correlations, and we need longitudinal research to understand potential shared mechanisms better. One example is an intriguing but preliminary study using mice with type 2 diabetes that demonstrated that a swimming exercise decreased depressive-type behaviors and glucose and inflammatory cytokines in the mice.

Psychiatric Symptoms Often Occur at Key Developmental Stages in Patients with Diabetes

Patients and their family members are most likely to exhibit psychiatric symptoms at critical stages of diabetes management, including reactions to initial diagnosis, when a complication (i.e., neuropathy, retinopathy) develops, or when intensifying a treatment regimen is needed, such as starting insulin. Conversely, chronic stressors and psychiatric symptoms may markedly interfere with adherence to self-care regimens (e.g., diet, exercise, taking medications, cessation of smoking) and diabetic disease control. In addition, family conflict regarding diabetes tasks, such as following a new dietary regimen or increasing exercise, may also interfere with treatment.

Opportunities for clinicians to screen for psychosocial difficulties and specific psychiatric disorders include at the time of initial diagnosis of diabetes, during regularly scheduled diabetes visits, during visits with diabetes nurses or educators, during hospitalization or development of a diabetes complication, and when discussing problems with glucose control or adherence to lifestyle or medication regimens. In addition, clinicians may want to consider using validated brief screening tools for depression and anxiety, such as the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder (GAD-7), respectively, or use the Problem Areas in Diabetes (PAID) instrument as a way of better understanding diabetes-specific emotional distress.

Patient–Provider Relationship: Effect of Psychiatric Comorbidity and Maladaptive Attachment Styles

Effective self-management of either type 1 or type 2 diabetes requires a close working relationship with a physician, other health care providers, and family members. This partnership with the health care provider allows the patient and their providers to work in an interactive, iterative fashion to achieve ideal medication management that maximizes benefits and minimizes side effects while implementing essential lifestyle changes (e.g., losing weight, increasing exercise, cessation of smoking).

Satisfaction with the quality of care of a patient’s physician is associated with enhanced adherence to medical regimens in patients with diabetes. Depressive disorders and maladaptive attachment styles are associated with decreased ratings of satisfaction with care in patients with diabetes. Moreover, primary care effectiveness studies testing interventions that improve major depressive outcomes showed that enhanced depressive outcomes were associated with increased satisfaction with care.

Depression and anxiety disorders in adulthood are also associated with higher rates of having experienced childhood adversity (abuse and neglect) and with maladaptive attachment styles, which are associated with lower levels of trust in others. Maladaptive attachment styles may make it difficult for patients to develop trusting relationships with others, including health care providers. Two styles, dismissing and fearful attachment, are associated with difficulty in trusting others and are the predominant relationship styles in 48% of adult primary care patients with diabetes. Adults with a dismissing attachment style develop strategies early in life in which they become highly independent and uncomfortable trusting others. In contrast, adults with a predominantly fearful attachment style may have an initial desire for social contact at times of distress that is inhibited by fear of rejection, resulting in approach-avoidance behavior.

Among over 4,000 primary care patients with type 2 diabetes, those with dismissing and fearful attachment styles had 46% and 64% more missed primary care appointments, respectively, and both styles were associated with less satisfaction with care compared to those with secure attachment styles. In the same sample, compared to patients with a secure attachment style, those with dismissing attachment style had lower levels of exercise, foot care, diet adherence, oral hypoglycemic medications, and higher rates of smoking. Other studies have also shown that in individuals with diabetes, dismissing and fearful attachment styles had decreased glucose monitoring, nonadherence with insulin injections, higher glucose levels, and higher mortality over 5 years. The adverse impact of dismissing and fearful attachment described above was independent of the effect of depression.

Depression may also increase the severity and impact of maladaptive attachment style traits. One study demonstrated that decreased depressive symptoms over 10 months were associated with transitioning to more adaptive attachment behaviors and greater trust in others. This study suggests that affective illness increases maladaptive interpersonal behaviors, potentially making collaboration with the physician more difficult and contributing to patients with these disorders receiving suboptimal preventive medical care and having poorer diabetes outcomes.

Type 1 Diabetes: Developmental and Family Issues

Having a child with a chronic illness often causes great stress for families and may interfere with normal childhood development. For example, research suggests higher divorce rates in families where one child has a significant chronic illness or developmental disability, and children with chronic illnesses such as diabetes have twofold higher rates of affective illness and anxiety disorders.

Type 1 diabetes may be particularly taxing for children and their families. Family-based care for diabetes requires checking blood glucose several times a day and altering insulin doses depending on blood glucose levels. This illness also requires following careful dietary regimens and getting consistent levels of exercise. When children develop this illness early in life, much of the responsibility for diabetes care depends on parents and may add significantly to parental stress. Recurrent hospitalization, inability to participate in usual childhood outings (particularly those away from home), and parental overprotectiveness can all delay normal childhood progression toward mastery and autonomy.

Moreover, as the child moves into adolescence, there may be more significant conflict as the child tries to separate and individuate. Acting out behavior such as dietary indiscretions, missing insulin doses, and avoiding checking blood sugars can all lead to poor glucose control during adolescence. Individual and family therapy approaches may be necessary to improve the quality of diabetes care and to help adolescents and their families adjust to new developmental phases.

Effects of Depression on Diabetes Symptom Burden

Comorbidity between depression and diabetes increases a patient's risk for morbidity from both illnesses, including diabetes-related complications, disease management, quality of life, and mortality.

Several studies have examined the association of comorbid depression and anxiety in patients with diabetes on physical symptoms characteristic of poor glucose control (i.e., polyuria, polydipsia). These studies have shown that depression and anxiety disorders were more highly associated with self-reported diabetes symptoms than were HbA1c levels or the number of diabetes complications. For example, a study by Ludman and colleagues found that, after controlling for potential confounders, patients with diabetes and depression were significantly more likely to report 10 diabetes symptoms than patients with diabetes alone (odds ratios between 2 and 5 for each symptom). Interestingly, patients with two or more diabetes complications compared to those with fewer than two complications had a significantly higher likelihood of having 9 of 10 diabetes symptoms, but the odds ratios were only between 1.3 and 2.0. These findings demonstrate that depression is more highly associated with diabetes symptom burden than the number of diabetes complications. In addition, a study that compared collaborative care versus usual primary care in treating depression in 387 patients with diabetes found that not only was collaborative care associated with better depression outcomes but also significantly decreased diabetes symptom burden compared to usual care despite not having a significant effect on glycated hemoglobin levels.

Some symptoms of diabetes, such as chronic pain due to neuropathy, may lead to bidirectional adverse experiences with depression. Depression can amplify pain complaints, thus preventing habituation to aversive symptoms. Biologic evidence suggests that depression-associated serotonin and norepinephrine depletion may lead to the dysregulation of descending pain pathways that typically dampen aversive symptom perception. Longitudinal studies have also shown that chronic pain is associated with a higher incidence of depression. Antidepressant medications, such as venlafaxine and duloxetine, that increase norepinephrine and serotonin are effective in treating diabetic neuropathy, even in nondepressed patients.

Effect of Depression and Anxiety on Functioning

Multiple studies have shown that depression and anxiety lead to additive social and vocational disability in patients with chronic medical illnesses. For example, the Pathways Epidemiologic Study has longitudinally surveyed over 4,000 patients in a large health maintenance organization with diabetes and shown that comorbid major depression in patients with diabetes was associated with a 10-fold increase in the likelihood of functional impairment on the World Health Organization (WHO) Disability Scale and a fourfold increased odds of having 20 or more days of reduced ability to do housework in the past 90 days after controlling for diabetes complications, medical comorbidity, and glycemic control. Among 1,642 nonelderly persons in the Pathways Epidemiologic Study (excluding homemakers and retired persons), those with comorbid depression had a fourfold increased likelihood of missing more than 5 days of work in the prior month, more difficulty performing work tasks, and greater levels of unemployment after controlling for comorbid chronic medical illness, diabetes complications, and glycemic control.

A large longitudinal epidemiologic study that followed approximately 2,800 older adult Hispanic Americans in the Southwest over 7 years found that the interaction of depression and diabetes was synergistic in predicting the greater incidence of decrements in activities of daily living, even after controlling for sociodemographic characteristics. Both depression and diabetes independently predicted increased incidence of decrements in function, and when they were comorbid, there was more than an additive adverse effect on functioning.

A study that compared a collaborative care intervention to usual primary care in enhancing the quality of depression care in over 400 older adult patients with diabetes found that the intervention patients not only had improved depressive outcomes but also had improved physical functioning over 12 months.

THE RELATIONSHIP BETWEEN DIABETES, DEPRESSION, AND OBESITY

The relationship between diabetes, depression, and obesity is complex and not necessarily linear. Patients with obesity and depression are more likely to have other chronic diseases, including diabetes. The relationship may be, in part, genetic. Genetic studies estimate that 12% of the variants linked to obesity are shared by depression, and the associations between obesity and type 2 diabetes are much higher in large genome-wide association studies (GWAS). Serotonin is critical to many functions that involve one or more of the disorders, including mood regulation, satiety, insulin and leptin action, carbohydrate metabolism, and inflammatory cytokines, and it may represent a pathway linking these disorders. The same could be valid for the HPA axis, which modulates stress, metabolic rate, appetite, and inflammation.

Adverse health habits such as smoking, poor diet, and sedentary lifestyle also occur in a higher prevalence in patients with mental illness than in those without, which may also increase the incidence of obesity, metabolic syndrome, and type 2 diabetes. However, the most important may be decreased physical activity associated with depression.

EFFECT OF DEPRESSION AND OTHER PSYCHIATRIC AND MEDICAL COMORBIDITIES

Patients with diabetes must optimally follow complex self-care regimens, including changing their diet, increasing exercise, checking blood glucose and altering medication dosages based on glucose fluctuations, stopping smoking, and taking multiple disease control medications. Depression may adversely affect these self-management activities by decreasing energy, concentration, memory, and motivation. A meta-analysis of over

20 studies of patients with chronic medical illness found that depression and anxiety were associated with a threefold higher rate of nonadherence to medical regimens.

In the Pathways Epidemiologic Study, patients with major depression and diabetes, compared to those with diabetes alone, were less likely to eat a healthy diet, more likely to eat high-fat foods, less likely to exercise, and more likely to smoke. Depressed patients were also more likely to have lapses in refilling prescriptions for oral hypoglycemic, lipid-lowering, and antihypertensive medications. These medication lapses may lead to a higher risk of diabetes complications.

Other psychiatric illnesses such as bulimia, night eating syndrome, binge eating syndrome, and chronic psychiatric illnesses such as schizophrenia and bipolar illness are associated with poor self-management of chronic illnesses such as diabetes and poor quality of medical care. For example, patients with schizophrenia and bipolar disorder are twice as likely to die from cardiovascular disease as the general population. In addition, research has shown that patients with diabetes and schizophrenia are treated less aggressively for their cardiovascular risk than patients with diabetes without a mental disorder. To worsen matters, patients with diabetes and schizophrenia have an elevated rate of depression compared to patients with diabetes who do not have schizophrenia.

A meta-analysis examined the prevalence of eating disorders among patients with type 1 diabetes versus controls. The prevalence of anorexia nervosa was not significantly different in patients with type 1 diabetes versus controls (0.27% vs. 0.06%). In comparison, the prevalence of bulimia nervosa (1.73% vs. 0.69%) and anorexia nervosa and bulimia nervosa combined (2.00% vs. 0.75%) were significantly higher versus controls in patients with type 1 diabetes. Some patients with insulin-dependent diabetes also try to lose weight by skipping or reducing their insulin injections—a phenomenon called “insulin withholding”—predisposing them to poor glucose control and future complications.

Psychiatric Illness and Diabetes Complications

A meta-analysis of 22 studies involving more than a million participants showed that depression was associated with an increased risk of macrovascular (HR = 1.38; 95% CI: 1.30–1.47) and microvascular (HR = 1.33; 95% CI: 1.25–1.41) complications in patients with diabetes.

Furthermore, the relationship was bidirectional in that diabetes complications also increased the risk for depression.

Similarly, a meta-analysis of 11 studies with 34,185 subjects found that patients with diabetes were at significantly elevated risk of developing diabetic retinopathy, if they had comorbid depression (fixed effects OR = 1.50, 95% CI 1.39–1.63; random effects OR = 1.58, 95% CI 1.35–1.84). The risk was somewhat higher for individuals greater than those aged 60.

The mechanisms of this association are not entirely understood but likely involve poor self-care and, perhaps, shared biologic mechanisms, such as increased cortisol levels and inflammation.

Several longitudinal studies have also shown that depression was associated with increased complications. For example, in a large longitudinal study of Hispanic older adult respondents in the southwestern part of the United States, older adult patients with diabetes who were initially depressed versus those with diabetes alone were three times as likely to develop macrovascular and microvascular complications over 7 years. In the Pathways Epidemiologic Study that followed over 4,000 patients with diabetes over 5 years, those with depression and diabetes versus those with diabetes alone were 36% more likely to develop microvascular complications and 24% more likely to develop macrovascular complications. A large longitudinal study of children with type 1 diabetes also showed that the risk for developing retinopathy was independently associated with three risk factors: duration of diabetes, time spent with poor glucose control, and time spent in major depression.

One study demonstrated that antidepressant use decreased macrovascular complications and mortality risk. However, it did not affect microvascular disease.

Depression and Diabetes: Cardiac Risk Factors. Approximately 50% of mortality in patients with diabetes is related to myocardial infarction or stroke. In addition, a recent meta-analysis of 17 studies including over a million patients found that, for patients with type 2 diabetes, depression was associated with a higher risk of both nonfatal (relative risk 1.35, 95% confidence interval [CI] 1.20 to 1.53) and fatal (relative risk 1.47, 95% CI 1.21–1.77) cardiovascular events.

Several mechanisms might explain this increased risk. Depression can adversely affect autonomic activity, heart rate variability, catecholamine levels, inflammation, and endothelial and platelet dysfunction, all of which are risk factors for cardiovascular events. More work must be done to understand this relation and identify interventions to lessen the risk.

Diabetes, Depression, and Dementia. Type 2 diabetes is associated with a greater prevalence of cognitive impairment, cognitive decline, and Alzheimer disease. A review of 25 prospective studies found that people with diabetes had a 1.2 to 1.5 greater odds of developing cognitive decline than those without diabetes. Patients with type 2 diabetes have reduced sensitivity to insulin (i.e., increased insulin resistance) and reduced neuronal sensitivity to insulin. Reduced neuronal insulin sensitivity may exacerbate neurodegeneration, causing cognitive decline and dementia. Diabetes is also associated with a higher risk of vascular disease, including stroke, which may increase the risk of dementia, particularly vascular dementia. Hypoglycemia is a complication of treatment with glucose-lowering drugs, particularly insulin. Severe hypoglycemic episodes associated with emergency room visits or hospitalizations have increased the risk of dementia in patients with diabetes. Comorbid depression can further contribute to the association between diabetes and cognitive decline. The Diabetes and Aging Study, a cohort investigation examining risk factors for dementia in individuals with diabetes, examined more than 19,000 patients over a 3- to 5-year period. Patients with comorbid depression had a 100% increased risk of dementia during the observation period.

Cognitive decline is likely to interfere with the self-management of diabetes, potentially worsening insulin resistance and hyperglycemia. In addition, both depression and diabetes can adversely affect executive functioning. Therefore, screening patients with diabetes using tools like the Mini-Mental Status Examination or the Montreal Cognitive Assessment (MoCA) is essential to ensure early recognition. In addition, enlisting family support for patients with cognitive impairment and diabetes is essential to slow the progression of diabetes.

Two studies have also found that patients with comorbid depression and type 2 diabetes, compared to those with diabetes alone, had an approximately twofold greater risk of developing dementia. A study of approximately 29,000 patients with type 2 diabetes developed a risk prediction model for dementia and identified depression as an important clinical risk factor. Another study that analyzed data from a large randomized trial designed to control glucose and cardiovascular risk factors in patients with diabetes found that those with comorbid depression and diabetes versus those with diabetes alone had a greater risk of cognitive decline in neuropsychological tests over 40 months. Most recently, a study of 3,742 individuals with type 1 diabetes, 50 years or older, found that depression was associated with a 72% increase in dementia (fully adjusted HR = 1.72; 95% CI:1.12–2.65). This study followed subjects for 25 years and estimated that the presence of depression doubled the incidence of dementia. Similarly, the Edinburgh type 2 diabetes study followed 1066 subjects for an average of 10 years and found that depression was associated with a more than 2.5-fold risk for dementia (HR 2.59 [95% CI 1.62, 4.15]) after adjusting for age and sex. In their study, systemic inflammation did not seem to explain this risk.

Effect of Depression on Medical Costs

Given the adverse effect of depression on somatic symptom burden, functional impairment, poor adherence to self-care, and increased complications, it is not surprising that several studies have shown that total health service costs of patients with diabetes and depression are significantly higher than those with diabetes alone. Studies that have carefully controlled for demographic factors, the severity of diabetes, and medical comorbidity have found that total health service costs are approximately 70% higher for individuals with major depression and diabetes than for diabetes alone. This finding is crucial for health care systems because patients with diabetes have approximately two- to fourfold higher medical costs than those without diabetes. Therefore, a 70% increase in the costs of these already expensive patients adds a disproportionate amount of total costs to a health care system. Other studies have also shown that patients with depression and diabetes, compared to those with diabetes alone, have more than twice the risk of needing an intensive care unit (ICU) admission after controlling for sociodemographic factors and the severity of diabetes. ICU admission costs are about double the costs of regular hospital admissions. Another study showed that those with depression and diabetes, compared to those with depression alone, were about 40% more likely to have an ambulatory care-sensitive hospitalization (a potentially avoidable hospitalization if patients received a higher quality of primary care).

Studies have also shown that the increase in mental health care costs does not explain the increased cost for patients with diabetes and depression compared to those with diabetes alone. Instead, there is an increase in costs in every category that is measured, including primary care visits, medical specialty visits, nonpsychiatric prescriptions, x-ray and lab costs, emergency room and urgent care visits, inpatient medical care and mental health visits, and prescriptions. In addition, two large primary care-based studies (the Pathways and IMPACT [Improving Mood—Promoting Access to Collaborative Treatment for Late Life Depression] studies) testing collaborative care interventions showed that improving the quality of care and outcomes of patients with depression results in cost savings in medical costs that offset the additional costs of the collaborative care mental health intervention.

Psychiatric Illness and Mortality in Patients with Diabetes

A systematic review of 10 longitudinal studies found that comorbid depression in patients with diabetes was associated with a 50% increased mortality risk. One study showed that depression or diabetes in older adult patients increased mortality risk by almost one-third, and respondents with both disorders had a fourfold increase in mortality compared to older adult with neither disorder.

Among patients with diabetes in the Pathways Epidemiology Study, after controlling for age, gender, and socioeconomic factors, comorbid major depression was associated with a twofold increase in mortality over 3 years compared to nondepressed patients. Minor depression also increased the risk of mortality by 67%. Even when controlling for diabetes, medical disease severity, and adverse health behaviors (smoking, obesity, sedentary lifestyle, HbA1c), major and minor depression were still associated with an approximately 50% increase in mortality.

Similarly, a study of 3,923 patients in Spain with type 2 diabetes found that depression was associated with higher mortality risk (HR = 1.40, 95% CI:1.20–1.65, *P* <.001). The strongest predictors were age, insulin use, amputation, heart failure, and male gender.

TREATING DIABETES AND DEPRESSION

One systematic review of treatment studies by van der Feltz-Cornelis and colleagues identified 43 studies with 4,602 patients with diabetes mellitus and either a depressive disorder or subthreshold depression. Most studies reported that depression treatment had a positive effect on depressive symptoms. This finding might seem obvious; however, this is encouraging given that psychiatric disorders comorbid with medical illnesses tend to be less responsive to treatment. Whether depression treatment also improves diabetes outcomes is less clear. A meta-analysis of the 43 studies showed a small but significant positive effect on glycemic control. The effect size was greatest for pharmacotherapy (0.987, 95% CI 0.127; 1.846, *P* = .024) and group-based therapy (0.953, 95% CI 0.185; 1.722, *P* = .015) with a lesser (although significant) effect reported for psychotherapy (0.607, 95% CI 0.147; 1.066, *P* = .010) and collaborative care (0.207, 95% CI 0.050; 0.364, *P* = .010) approaches. The effect sizes for exercise, online, and phone treatment did not differ from usual care.

As with all meta-analyses, combining various studies, often with different protocols and aims, is complex and coming to meaningful conclusions is difficult. Nevertheless, some individual studies have had more positive results.

Three efficacy studies that randomized patients with major depression and diabetes to antidepressant versus placebo or CBT versus a diabetes education group found significant improvement in depression in all studies (effect sizes in antidepressant trials between 0.62 and 0.63, and 1.16 in the CBT trial) but only one study showed a beneficial effect on HbA1c levels.

Among patients with depression and diabetes in the above three large trials, 60% to 80% were diagnosed with double depression (i.e., concurrent major depression and persistent depressive disorder, i.e., dysthymic disorder). Double depression predicts treatment resistance and recurrence or relapse. Comorbid chronic medical illness in patients with depression is associated with a more chronic course, slower recovery, and less treatment response. Given the high rates of relapse and chronicity in patients with depression and diabetes, researchers have begun to test the effect of maintenance antidepressant treatment. Several antidepressant trials that tested the effect of maintenance treatment have suggested that long-term improvements in treatment that resulted in enhanced rates of recovery and remission are associated with improved glycemic outcomes. One maintenance trial randomized 152 patients who had recovered with an open-label trial of an SSRI to either SSRI or placebo for up to 52 weeks. Patients who received maintenance treatment with the SSRI had a significantly greater depression-free interval than those treated with a placebo. In addition, patients using an SSRI who recovered from depression and sustained remission for at least 1 year had improved HbA1c levels.

A second maintenance trial that treated 93 patients with type 2 diabetes with acute phase bupropion (Wellbutrin) offered maintenance-phase bupropion treatment to patients who remitted (N = 63). BMI, total fat mass, and HbA1c values decreased significantly, while composite diabetes care improved over the initial acute phase, and these effects persisted through the maintenance phase. In addition, a reduction in BMI and depression severity with bupropion treatment predicted lower HbA1c levels after acute phase therapy, but only a reduction in depression predicted lower HbA1c levels during the maintenance phase of treatment.

How Can Depression Treatment Improve Diabetes Care?

Piette et al. have proposed a model to explain how depression treatment may improve the outcomes of comorbid chronic disease, which they applied to diabetes. In this model, effective treatment of depression may be more difficult because of the adverse health behaviors (smoking, obesity, sedentary lifestyle) and higher symptom burden and functional disability associated with depression. Adverse health behaviors do not necessarily improve with effective depression treatment alone and are negative prognostic factors for depression outcomes (e.g., obesity is associated with negative social feedback, lower self-esteem, and less exercise, which all affect mood). New treatment models that provide interventions aimed at improving the quality of depression care, optimizing positive health behaviors (exercise, diet, self-management), and treating chronic medical illness by optimizing medication treatment have shown improvement in depression and medical disease control outcomes. Furthermore, studies show that a significant contributor to suboptimal outcomes in primary care is *clinical inertia*, that is, lack of timely treatment intensification by the primary care physician or endocrinologist, such as titrating medications for reducing blood glucose, blood pressure, and cholesterol. For example, among 161,697 Kaiser Permanente patients with diabetes, 50% were above target for A1c, blood pressure, and LDL cholesterol. Lack of treatment intensification by health care teams was a more significant contributor to out-of-target values than poor patient treatment adherence.

Collaborative Care

Three large primary care-based randomized trials tested collaborative care treatment of depression among patients with diabetes: 418 patients with diabetes and depression age 60 and over in the IMPACT trial; 329 mixed-age patients with diabetes and depression with a mean age of 59 years in the Pathways trial, and 387 predominantly Hispanic patients in which 52% of patients were 55+ years old. These trials tested a stepped-care nurse intervention for depression versus usual primary care. All three trials showed marked intervention versus usual care effects in quality of depression care and significant improvements in depression outcomes (effect sizes between 0.47 and 0.68), but no significant changes in self-care such as diet or checking blood glucose and no significant difference in HbA1c levels. In addition, approximately half of the usual primary care controls were treated with antidepressants by their primary care physician during the 12-month trials.

Using Multicondition Collaborative Care to Reduce Clinical Inertia. The TEAM care clinical trial was a landmark study that expanded the collaborative care model from primarily focusing on treating depression to addressing multiple chronic conditions simultaneously using one case manager per patient, who focused on “A-B-C-D,” that is, A1c, blood pressure, cholesterol (LDL) and depression. This clinical trial investigated 214 patients with depression and with poorly controlled diabetes or heart disease (i.e., patients had to have either an HbA1c level of at least 8.5%, a systolic blood pressure >140 mm Hg, or a low-density cholesterol level (LDL) of >130 mg/dL). The patients were randomized to a *multicondition collaborative care intervention* to improve the quality of medication treatment for depression and these medical illnesses and improve self-care (exercise, diet, cessation of smoking, and taking medications as prescribed). In addition to treat-to-target and systematic case review methods, this intervention included motivational enhancement, problem-solving treatment, and behavioral activation clinical methods by front-line nurse case managers. The intervention was associated with significant improvements in depression, HbA1c, SBP, LDL control, and exercise.

Another randomized trial of 180 patients prescribed medications for diabetes and depression in primary care tested an intervention that utilized care managers integrated into primary care to offer education, guideline-based treatment recommendations, and monitoring of adherence and clinical status. This intervention was significantly more likely than usual primary care to achieve HbA1c levels <7% and remission of depression. Another study evaluated the impact of a telephone-delivered cognitive behavioral therapy intervention targeting depressive symptoms, a pedometer-based walking program aimed at increasing physical activity levels and improving diabetes-related outcomes in 291 patients with type 2 diabetes and significant depressive symptoms. Although the investigators found no differences between the intervention and usual care patients on the primary outcome of HbA1c levels, the intervention group had more significant reductions in depressive symptoms, systolic blood pressure, and step counts.

These trials show that to improve depression and medical outcomes, programs need to address clinical inertia by combining interventions for depression, as well as interventions aimed at improving adherence and intensifying diabetes disease control medications and self-care activities (e.g., providing pedometers to encourage increased exercise).

Summary

The meta-analysis and the individual studies discussed above suggest three conclusions. First, organized depression treatment programs such as collaborative depression care are effective for patients with diabetes. Second, the effects of depression treatment may be smaller for patients with chronic medical illness than in general primary care samples. Third, short-term treatment programs limited to improving depression may not have clear or consistent effects on medical outcomes such as HbA1c values. Finally, the maintenance trials tend to show that patients who have remitted in acute phase treatment and remain on antidepressants over the next year have improved glycemic control, whereas acute phase trials of antidepressants have inconsistent effects on HbA1c levels.

PSYCHIATRIC MEDICATIONS, OBESITY, AND TYPE 2 DIABETES

The widespread use of atypical antipsychotic medications for patients with bipolar illness, schizophrenia, and to a lesser extent, treatment-resistant depression has led to concerns about these psychiatric medications causing obesity, metabolic syndrome, and type 2 diabetes. Table 27.6–1 lists the antipsychotics and their associated weight gain. Clozapine (Clozaril) and olanzapine (Zyprexa) have the highest risk. In addition, numerous case reports describe the less common occurrence of acute hyperglycemia after initiating these medications before a substantial weight gain. In addition, some but not all cases report hyperglycemia resolved after discontinuing the medication. These data suggest a direct role of atypical antipsychotic medications in increasing insulin resistance and causing type 2 diabetes, in addition to an indirect role in causing obesity.

Data have shown two- to fourfold increased rates of metabolic syndrome in patients with schizophrenia treated with typical or atypical antipsychotic agents. An estimated 16% to 25% of patients with schizophrenia have type 2 diabetes. However, the risk of developing diabetes with atypical antipsychotic agents appears to be higher than with typical antipsychotic agents. Data from a large Veterans Administration study of over 38,000 patients treated with atypical versus typical antipsychotic agents show an approximately 10% increased rate of diabetes associated with treatment with atypical antipsychotic agents. The large-scale CATIE trial also showed significantly more weight gain associated with atypical antipsychotic medications such as olanzapine and clozapine than those treated with perphenazine (Trilafon), a typical antipsychotic agent.

Table 27.6–1.
Antipsychotics and Associated Weight Gain

Likelihood of Weight Gain	Antipsychotic
High	Clozapine Olanzapine
Moderate	Chlorpromazine Quetiapine Risperidone Paliperidone
Low	Aripiprazole Amisulpride Asenapine Haliperidol Ziprasidone Lurasidone

Given the increasing concerns about the atypical antipsychotic agents causing obesity, metabolic syndrome, and type 2 diabetes, experts recommend using monitoring protocols. These protocols include eliciting a personal and family history of diabetes, heart disease, and other risk factors such as smoking and regularly monitoring weight, waist circumference, blood pressure, fasting blood glucose, and lipid profiles.

It is important to emphasize that other psychiatric medications can also lead to weight gain. The use of anticonvulsant agents such as valproic acid (Depakene, Depakote) and carbamazepine (Tegretol), as well as lithium carbonate (Eskalith, Lithobid) in patients with bipolar disorder, are also frequently associated with weight gain. Lamotrigine (Lamictal), another anticonvulsant medication increasingly used in bipolar disorder and treatment-resistant depression, is generally not associated with weight gain. Although topiramate (Topamax) has not been FDA-approved for use in bipolar disorder, psychiatrists often use lower dosages of this medication as an adjunctive medication due to its effect on decreasing appetite and inducing weight loss. Metformin (Glucophage) can also induce weight loss in patients taking atypical antipsychotics.

Among antidepressant agents, tricyclic antidepressant medications and mirtazapine (Remeron) cause weight gain in many patients. However, SSRIs appear to induce weight gain in only 5% to 10% of patients compared to placebo. Bupropion is the only antidepressant showing no weight gain or slight weight loss in placebo-controlled trials.

Using Technology to Reach Larger Populations

Telehealth and other approaches that harness readily available technology can help clinicians reach many otherwise difficult-to-access populations. For example, the pharmacist-led telehealth management program significantly improved medication adherence (for diabetes and depression). This finding is promising since collaborative care approaches also rely on similar technology. However, there are a growing number of web-based or app-based interventions for diabetes and depression, and there is a great need for more research to test the efficacy of these approaches.

A Clinical Approach to Treating Depression in Patients with Diabetes

Table 27.6–2 outlines a diabetes-focused clinical approach to treating depression and other psychiatric problems in patients with diabetes. Because of the increased prevalence of depression in patients with diabetes, clinicians should use screening tools such as the PHQ-9 to recognize depression in diabetes settings. Frequently, patients with diabetes and comorbid depression presenting to a health care provider are unaware of their depressive symptoms. Instead, they may experience physical symptoms of diabetes, such as polyuria and fatigue. In addition, they commonly present with high levels of frustration, resignation, or guilt related to increased difficulties in managing glucose and other diabetes-specific tasks. These symptoms may result from depression’s effects on the HPA axis or the aforementioned negative effects of depression in patients with diabetes, such as amotivation, fatigue, anhedonia, lack of concentration, and hopelessness, which make diabetes self-management a problematic task.

Table 27.6–2.
Diabetes-Focused Depression Care

In addition to standard evaluation of current (and past) episodes of depression:

1. Assume depressed patient is experiencing “loss of control” of disease self-management or outcomes resulting in frustration, bewilderment, resignation, etc.
2. Assess for understanding of bidirectional link between stress and suboptimal disease self-management and outcomes.
3. Define depression and how it overlaps with (and is distinct from) “stress.”
4. Review symptoms of depression and how these symptoms overlap with or mimic diabetes symptoms. Discuss depression-related symptom amplification.
5. Consider “biologic” sequelae resulting from comorbid depression:
    - e.g., increased glucose levels (increased glucometer mean + SD, higher HbA1c)
    - e.g., increased neuropathy pain, i.e., increased intensity, functional impairment
6. Consider “behavioral” sequelae resulting from comorbid depression:
    - e.g., decreased self-care (refilling prescriptions for medications and monitoring strips; glucose monitoring; exercise; diet). Look for helplessness/ “giving up” or sense of being overwhelmed around disease self-care
    - e.g., missed appointments or high health care utilization (labs, visits, urgent care)
    - e.g., decreased satisfaction with care
    - e.g., decreased ability to rely on others
7. Screen for comorbid anxiety ± panic attacks and look for inability to differentiate anxiety symptoms from diabetes symptoms (e.g., hypoglycemia).
8. Screen for associated eating concerns (e.g., emotional eating in response to sadness/loneliness/anger, etc.; binge eating ± purging; night eating).
9. Break down tasks (self-management of diabetes, depression, heart disease, other illnesses) and help patient prioritize in order of importance
10. For patients with comorbid anxiety, consider using SSRI or SNRI
11. For patients with sexual dysfunction, consider using bupropion or mirtazapine (consider weight gain with mirtazapine)

12. For patients with significant neuropathy, consider bupropion, venlafaxine, or duloxetine
13. Consider adjunctive brief psychotherapy for:
    - emotional eating (CBT)
    - breaking down problems (problem-solving therapy)
    - improving treatment adherence (motivational interviewing)
    - improving effective patient–provider interactions (patient coaching)

SD, standard deviation; HbA1c, hemoglobin A1c; SSRIs, selective serotonin reuptake inhibitors; SNRI, serotonin and norepinephrine reuptake inhibitor; CBT, cognitive-behavioral therapy.

The clinician treating individuals with diabetes and depression may benefit from the concept of *diapression,* which recognizes that individuals with diabetes who develop depressive symptoms may experience specific characteristics.

- Depression can mimic and amplify diabetes symptoms
- Depression is associated with poor diabetes self-management and treatment adherence
- Depression can lead to adverse lifestyle habits
- Depression is associated with changes in health care utilization patterns
- Symptoms of depression reduce levels of trust and satisfaction with care

A powerful way of engaging patients experiencing depressive symptoms may be to nonjudgmentally validate their experience of a recent “loss of control” of their diabetes. Furthermore, this approach helps provide opportunities to educate patients about the interaction between stress, depression, and anxiety on one hand and diabetes outcomes on the other. Patients are often relieved to hear that it is not their fault that their glucose control has become suboptimal.

Screening for comorbid anxiety is critical since symptoms of anxiety and panic disorder may worsen the control of diabetes. Furthermore, it is often difficult to disentangle anxiety symptoms from hypoglycemic symptoms for patients on insulin who may have frequent hypoglycemic episodes because of poor self-management. As a result, patients may eventually ignore significant hypoglycemic symptoms or catastrophize when anxious because they misinterpret anxiety as hypoglycemic symptoms. Using screening tools for anxiety, such as the GAD-7, can help when working with patients with diabetes.

Asking about eating behaviors and achieving emotional regulation through eating is vital since patients with diabetes and depression have a higher risk of bulimia, binge eating disorder, and night eating syndrome. A subgroup of patients eat a significant proportion (e.g., more than 25%) of their caloric intake after supper time, and such “night eating” behaviors are associated with higher HbA1c levels, more complications, and higher BMI. A nonjudgmental approach that optimally treats underlying anxiety or depression and potentially uses CBT techniques for problematic eating behaviors can effectively relieve emotional eating in patients with diabetes.

Other practical strategies for patients with depression include problem-solving treatment, behavioral activation, CBT for depression, and the use of medications, with care to choose medications that may add benefits for pain control in patients with neuropathy pain (about one-third of patients with diabetes have neuropathy).

FUTURE DIRECTIONS

The United States is facing an epidemic of type 2 diabetes, and psychiatric illness comorbidity with this medical illness is very high. Despite the cost of comorbidity, depression screening is not a routine part of most diabetes care, and many cases of depression in patients with diabetes remain undiagnosed. Therefore, it is important to screen patients with diabetes for psychiatric illnesses and to screen patients with psychiatric illnesses for diabetes since comorbidity is associated with adverse medical and psychiatric outcomes.

ACKNOWLEDGMENT

This chapter represents an update and revision of previous chapters in this series, authored by Wayne Katon, MD., and Dr. Ciechanowski. Dr. Katon, a pioneer in collaborative care mental health care and leader in consultation-liaison psychiatry, died in 2015. The authors are indebted to Dr. Katon for his work on the Kaplan and Sadock series and mental health and primary care fields. This edition includes much of his work and ideas and retains much of his excellent prose.

FURTHER READINGS

Ajuwon AM, Love R. Type 2 diabetes and depression in the African American population. *J Am Assoc Nurse Pract*. 2020;32(2):120–127.

Alaqeel A, Almijmaj M, Almushaigeh A, Aldakheel Y, Almesned R, Al Ahmadi H. High rate of depression among Saudi children with type 1 diabetes. *Int J Environ Res Public Health*. 2021;18(21):11714.

Black S, Kraemer K, Shah A, Simpson G, Scogin F, Smith A. Diabetes, depression, and cognition: a recursive cycle of cognitive dysfunction and glycemic dysregulation. *Curr Diab Rep*. 2018;18(11):118.

Black SA, Markides KS, Ray LA. Depression predicts increased incidence of adverse health outcomes in older Mexican Americans with type 2 diabetes. *Diabetes Care*. 2003;26(10):2822–2828.

Bogner HR, Morales KH, de Vries HF, Cappola AR. Integrated management of type 2 diabetes mellitus and depression treatment to improve medication adherence: a randomized controlled trial. *Ann Fam Med*. 2012;10(1):15–22.

Carr AL, Sluiman AJ, Grecian SM, et al. Depression as a risk factor for dementia in older people with type 2 diabetes and the mediating effect of inflammation. *Diabetologia*. 2021;64(2):448–457.

Ciechanowski P. Diapression: an integrated model for understanding the experience of individuals with co-occurring diabetes and depression. *Clin Diabetes*. 2011;29(2):43–49.

Ciechanowski P, Russo J, Katon W, et al. Influence of patient attachment style on self-care and outcomes in diabetes. *Psychosom Med*. 2004;66(5):720–728.

Cohen LB, Taveira TH, Wu WC, Pirraglia PA. Pharmacist-led telehealth disease management program for patients with diabetes and depression. *J Telemed Telecare*. 2020;26(5):294–302.

Foley DL, Mackinnon A, Morgan VA, Castle DJ, Waterreus A, Galletly CA. Comorbid diabetes and depression in a national sample of adults with psychosis. *Schizophr Bull*. 2018;44(1):84–92.

Gilak-Dalasm M, Peeri M, Azarbayjani MA. Swimming exercise decreases depression-like behaviour and inflammatory cytokines in a mouse model of type 2 diabetes. *Exp Physiol*. 2021;106(9):1981–1891.

Gilsanz P, Schnaider Beeri M, Karter AJ, Quesenberry CP Jr, Adams AS, Whitmer RA. Depression in type 1 diabetes and risk of dementia. *Aging Ment Health*. 2019;23(7):880–886.

Gulley LD, Shomaker LB. Depression in youth-onset type 2 diabetes. *Curr Diab Rep*. 2020;20(10):51.

Inoue K, Beekley J, Goto A, Jeon CY, Ritz BR. Depression and cardiovascular disease events among patients with type 2 diabetes: a systematic review and meta-analysis with bias analysis. *J Diabetes Complications*. 2020;34(12):107710.

Kalra G, Gill S, Tang TS. Depression and diabetes distress in South Asian adults living in low- and middle-income countries: a scoping review. *Can J Diabetes*. 2020;44(6):521–529.e1.

Katon W, Lin EH, Kroenke K. The association of depression and anxiety with medical symptom burden in patients with chronic medical illness. *Gen Hosp Psychiatry*.

2007;29(2):147–155.

Katon WJ, Lin EH, Von Korff M, et al. Collaborative care for patients with depression and chronic illnesses. *N Engl J Med*. 2010;363(27):2611–2620.

Lustman PJ, Williams MM, Sayuk GS, Nix BD, Clouse RE. Factors influencing glycemic control in type 2 diabetes during acute- and maintenance-phase treatment of major depressive disorder with bupropion. *Diabetes Care*. 2007;30(3):459–466.

Majumdar S, Sinha B, Dastidar BG, et al. Assessing prevalence and predictors of depression in type 2 diabetes mellitus (DM) patients–The DEPDIAB study. *Diabetes Res Clin Pract*. 2021;178:108980.

Nasrallah HA, Meyer JM, Goff DC, et al. Low rates of treatment for hypertension, dyslipidemia and diabetes in schizophrenia: data from the CATIE schizophrenia trial sample at baseline. *Schizophr Res*. 2006;86(1–3):15–22.

Nouwen A, Adriaanse MC, van Dam K, et al. Longitudinal associations between depression and diabetes complications: a systematic review and meta-analysis. *Diabet Med*. 2019;36(12):1562–1572.

Owens-Gary MD, Zhang X, Jawanda S, Bullard KM, Allweiss P, Smith BD. The importance of addressing depression and diabetes distress in adults with type 2 diabetes. *J Gen Intern Med*. 2019;34(2):320–324.

Park M, Katon WJ, Wolf FM. Depression and risk of mortality in individuals with diabetes: a meta-analysis and systematic review. *Gen Hosp Psychiatry*. 2013;35(3):217–225.

Piette JD, Richardson C, Valenstein M. Addressing the needs of patients with multiple chronic illnesses: the case of diabetes and depression. *Am J Manag Care*. 2004;10(2 Pt 2):152–162.

Salinero-Fort MA, Gómez-Campelo P, Cárdenas-Valladolid J, San Andrés-Rebollo FJ, de Miguel-Yanes JM, de Burgos-Lunar C. Effect of depression on mortality in type 2 diabetes mellitus after 8 years of follow-up. The DIADEMA study. *Diabetes Res Clin Pract*. 2021;176:108863.

Sevilla-González MDR, Quintana-Mendoza BM, Aguilar-Salinas CA. Interaction between depression, obesity, and type 2 diabetes: a complex picture. *Arch Med Res*. 2017;48(7):582–591.

Siegel KR, Ali MK, Zhou X, et al. Cost-effectiveness of interventions to manage diabetes: has the evidence changed since 2008? *Diabetes Care*. 2020;43(7):1557–1592.

Sullivan MD, Katon WJ, Lovato LC, et al. Association of depression with accelerated cognitive decline among patients with type 2 diabetes in the ACCORD-MIND trial. *JAMA Psychiatry*. 2013;70(10):1041–1047.

van der Feltz-Cornelis C, Allen SF, Holt RIG, Roberts R, Nouwen A, Sartorius N. Treatment for comorbid depressive disorder or subthreshold depression in diabetes mellitus: systematic review and meta-analysis. *Brain Behav*. 2021;11(2):e01981.

van Sloten T, Schram M. Understanding depression in type 2 diabetes: a biological approach in observational studies. *F1000Res*. 2018;7:F1000 Faculty Rev-1283.

Wu CS, Hsu LY, Pan YJ, Wang SH. Associations between antidepressant use and advanced diabetes outcomes in patients with depression and diabetes mellitus. *J Clin Endocrinol Metab*. 2021;106(12):e5136–e5146.

Zou Y, You W, Wang J, et al. Depression and retinopathy in patients with type 2 diabetes mellitus: a meta-analysis. *Psychosom Med*. 2021;83(3):239–246.

27.7 Endocrine and Metabolic Disorders

KATHERINE E. WILLIAMS, M.D., KATHLEEN WATSON, PH.D., AND NATALIE RASGON, M.D., PH.D.

Endocrine disorders are prevalent illnesses that a psychiatrist is likely to see in clinical practice regularly. This chapter will focus on the most common endocrine illnesses with neuropsychiatric symptoms and those commonly associated with primary psychiatric disorders. Recently, there has been considerable progress in understanding the intimate and complex relationships between endocrine and psychiatric disorders, including increasing evidence for a bidirectional relationship between the two classes of illness. For instance, insulin resistance has been associated with major depression and vice versa. Treatment with insulin-sensitizing agents has been shown to treat insulin resistance and depressive symptoms. Research is currently underway to elucidate the complex relationship between endocrinopathies and psychiatric illnesses.

GENERAL PRINCIPLES

Endocrinology is the study of hormones or biochemical mediators that travel from their production site to influence the function of other tissues. They exist in the form of peptides, steroids, and amines. Endocrinopathies result from excessive or deficient amounts of a hormone or resistance to a hormone's action. For example, target tissues may not respond to a normal level of a hormone. These latter cases are primarily hereditary and are characterized by normal or elevated hormone levels, despite an absence or deficiency of the hormone's action. Elevated levels result from the lack of negative feedback on the hormone’s production, which results from the failure or deficiency of the hormone’s action.

Hormone production is primarily regulated by the levels and activity of the hormones themselves and additionally by hormones from other sites. For example, hypothalamic-releasing factors regulate the production of pituitary hormones. The pituitary gland, situated in the sella tursica at the base of the skull, is divided into the anterior and posterior lobes.

The anterior pituitary produces six hormones

- **Thyroid-stimulating hormone (TSH):** regulates the production of thyroid hormones
- **Growth hormone (GH):** promotes growth and development
- **Prolactin (PRL):** stimulates lactation
- **Luteinizing hormone (LH)** and **follicle-stimulating hormone (FSH)**: necessary for gonadal function in men and women
- **Adrenocorticotropic hormone (ACTH):** controls adrenal cortex production of glucocorticoid.

The posterior pituitary gland stores two hormones produced by the hypothalamus

- **Antidiuretic arginine vasopressin (AVPJ):** regulates fluid balance
- **Oxytocin:** necessary for milk let-down during breastfeeding

The hypothalamus additionally produces releasing factors that influence pituitary hormone production.

- **Growth hormone-releasing hormone (GHRH) and growth hormone release inhibiting factor (GIH; also known as somatostatin)**: influence production of GH
- **Gonadotropin-releasing hormone (GnRH)**: influences the production of LH and FSH
- **Prolactin-inhibiting factor (PIF)** (i.e., dopamine)
- **Prolactin-releasing factor (PRF)**: influences PRL production
- **Thyrotropin-releasing hormone (TRH)**: influences TSH release

- **Corticotropin-releasing hormone (CRH)**: influences ACTH release

In addition to the hypothalamus and the pituitary, the major endocrine organs include the adrenal, thyroid, parathyroid glands, ovaries, and testes. Endocrinopathies may lead to heterogeneous psychiatric symptoms, as hormones affect various organ systems. Patients are more likely to present primarily with psychiatric symptoms early in the course of endocrine disease and experience a worsening of their psychiatric condition and the emergence of physical symptoms with the disorder's progression. Endocrinopathies should also be considered as a possible cause of treatment-resistant psychiatric disorders. Because endocrine disorders are common medical conditions, an evaluation of endocrine function is an essential part of the initial diagnostic workup. A laboratory evaluation should include serum electrolytes, fasting blood glucose, TSH, and calcium concentrations to identify potential endocrine etiologies. Psychiatric symptoms will not resolve until the underlying endocrine disturbance is corrected and may worsen with psychotropic agents. When psychiatric disorders appear to result from a medical condition, they are categorized accordingly in the fifth edition of the Diagnostic and Statistical Manual (DSM-V), e.g., “depressive disorder due to another medical condition.”

THYROID DISORDERS

The thyroid hormones triiodothyronine (T3) and thyroxine (T4) are produced from the thyroid gland and are necessary to regulate metabolic activity and protein synthesis. They are formed from a common precursor, thyroglobulin (TG), stored in the thyroid follicles. T4 is converted to T3 in extrathyroidal tissues, the metabolically active hormone. TSH (thyrotropin), produced by the anterior pituitary, promotes the synthesis and release of these hormones and is regulated by TRH, produced in the hypothalamus. Low serum levels of thyroid hormones stimulate the release of TRH and TSH. Notably, central nervous system (CNS) dopamine decreases TSH secretion and glucocorticoids.

Thyroid function depends partly on varying serum, androgen, estrogen, and TSH levels. In the case of T3 and T4, they are tightly protein bound to thyroxine-binding globulin (TBG) so that only 0.3% of T3 and 0.02% of T4 are free. Total serum T4 levels vary with TBG levels; for example, exposure to estrogens (pregnancy and oral contraceptive) raises TBG and, therefore, serum T4, whereas androgens have the opposite effect. The free T4 index (FT4I), which measures the percentage of free T4 (FT4), produces a more accurate determination of thyroid function. FT4I is elevated in hyperthyroid and low in hypothyroid patients and is an essential test for assessing thyroid status. Measures of TSH levels are also necessary for the workup of thyroid disorders. Levels are elevated in thyroidal hypothyroidism and are usually low in pituitary–hypothalamic hypothyroidism. The range of normal levels of TSH is 0.3 to 4.7 mIU/L, and free T3 index (FT3I) ranges from 5 to 11 mIU/L and 75 to 175 mIU/L, respectively. Varying hormone production levels are indicative (or predictive) of thyroid function and play an essential role in our assessment of TRH.

The TRH stimulation test (TRH infusion test) is sometimes performed to evaluate thyroid function. After the patient’s baseline TSH level is obtained, TRH is administered intravenously (IV), and TRH levels are then measured at regular intervals. A normal response is a rise in TSH of at least 5 mIU/mL above the baseline TSH value within 30 minutes of an injection of TRH. In hypothyroid patients, this response is exaggerated, whereas, in hyperthyroid patients, it is blunted. With the development in the 1990s of highly sensitive TSH testing, the TRH stimulation test is seldom used clinically. Tests of radioactive iodine (RAI) uptake are used to determine the activity of the thyroid gland. A patient receives a small amount of oral RAI, and a scan then determines the amount taken up by the thyroid gland. These tests can help us understand the influence of thyroid hormones on larger bodily systems, including the CNS.

Thyroid hormones have widespread actions on the CNS, including genomic and nongenomic effects that underlie the changes in mood and cognition seen in hypothyroidism. T3 is the primary thyroid hormone available to peripheral organs, whereas T4 is the primary hormone available to the brain. The symptoms of hypothyroidism may result from alterations in receptor functioning, neurotrophin production, or changes in intracellular signaling that occur due to alternations in thyroid hormone levels. For instance, decreased postsynaptic B receptor number and function are associated with low levels of thyroid hormones. Brain synthesis of serotonin (5HT) is reduced in hypothyroidism. PRL and cortisol responses to the 5HT agonist fenfluramine are blunted compared to euthyroid controls, suggesting reduced central 5HT activity.

Serum 5HT levels correlate with serum T3 levels. Thyroid hormone is associated with reduced sensitivity of 5HT1A autoreceptors in raphe nuclei and increased 5H2 sensitivity. Studies of regional cerebral blood flow have found decreased regional blood flow in the bilateral amygdala, hippocampus, and anterior cingulate cortex in hypothyroid patients. Treatment with thyroid hormone reduced neuropsychiatric complaints and restored regional blood flow, especially in the ventral anterior cingulate cortex. Many, but not all, studies of major depressive disorder have reported that the hypothalamic-pituitary-thyroid (HPT) axis is perturbed at multiple levels, including increased cerebrospinal fluid (CSF), TRH, TRH-stimulated TSH blunting, and nocturnal TSH blunting. Neurobiologically, this thyroid activation has been proposed to represent a compensatory mechanism to a pathologic process, i.e., depression or the primary pathology itself.

Hypothyroidism

Definition and Comparative Nosology and Epidemiology. Hypothyroidism results from inadequate production of thyroid hormone. In overt hypothyroidism, serum TSH levels are elevated, and serum T4 levels are low compared to normal levels within the general population. Subclinical hypothyroidism (SCH) is diagnosed when TSH levels are elevated, but T4 levels are normal. Hypothyroidism must be distinguished from the euthyroid sick syndrome, which is associated with several medical illnesses, including myocardial infarction, chronic kidney disease, diabetic ketoacidosis, anorexia nervosa, and cirrhosis. In this condition, conversion from T4 to T3 is decreased, leading to normal TSH but lower T3 levels.

Etiology. The most common cause of hypothyroidism in adults worldwide is iodine deficiency. In countries where iodine is added to salt, Hashimoto thyroiditis (chronic lymphocytic thyroiditis) is the most common cause. In this autoimmune condition, high serum concentrations of autoantibodies to the enzymes thyroglobulin (TGAb) and thyroid peroxidase (TPOAb) are found in the serum.

Lithium is another common cause of hypothyroidism. After long-term lithium use, 20% of patients develop overt hypothyroidism with a goiter, and another 20% develop SCH. This risk is elevated in women and patients with antithyroid antibodies. The longer the exposure to lithium, the greater the risk of developing either overt or SCH. Other causes of hypothyroidism are hypopituitarism due to tumor or postpartum hemorrhage, previous thyroid surgery, or antithyroid drugs (ATDs).

Epidemiology

GENERAL POPULATION. Overt hypothyroidism is more common in women than men in the general population, affecting 2% of women compared to less than 0.1% of men. SCH also predominates in women, occurring in approximately 7% of women compared to 3% of men. The prevalence of hypothyroidism increases with age, and SCH is prevalent in postmenopausal women, occurring in approximately 20% of this population. The progression of SCH to overt hypothyroidism is approximately 5% to 8% per year. Risk factors for conversion include higher baseline TSH, female sex, age, the presence of antithyroid antibodies, and a family history of autoimmune disorders.

PREGNANCY. Overt hypothyroidism is estimated to occur in 0.3% to 0.5% of pregnancies and SCH in 3% to 5%. The presence of elevated antithyroid peroxidase (TPO) and anti-TG titers are also common in euthyroid pregnant women, and higher titers are associated with an increased risk of developing postpartum thyroiditis that occurs in 7% of postpartum women. Twenty to forty percent of women with this condition develop permanent hypothyroidism. Women with recurrent spontaneous abortion also have higher overt and SCH rates.

PSYCHIATRIC POPULATIONS. Hypothyroidism is even more common in psychiatric patient populations, and the gender difference persists. Bipolar disorder patients who are not treated with lithium have a high prevalence of hypothyroidism, that is, 25% in women and 12% in men. Hypothyroidism is especially common in rapid-cycling bipolar disorder. It is estimated that 20% of patients treated with lithium develop hypothyroidism, and up to 50% may develop a goiter.

With lithium treatment, females are five times more likely than men to develop hypothyroidism. Older age and the pretreatment presence of thyroid antibodies are also risk factors for hypothyroidism and lithium use. Due to the high prevalence of postpartum thyroiditis in women with postpartum depression, the American Thyroid Association recommends that women with postpartum depression have TSH, free T4, and anti-TPO screening.

Diagnosis and Clinical Features

PHYSICAL SYMPTOMS. Physical symptoms associated with overt and SCH are diverse and reflect the widespread influence of the thyroid hormones on metabolic, gastrointestinal, cardiac, neurologic, reproductive, and even skin functioning. Symptom onset is usually insidious, and classic symptoms include weight gain, constipation,treatment may not lead to full remission cold intolerance, muscle cramps, dry skin, brittle and thinning hair, especially lateral eyebrows, slowed deep tendon reflexes (DTRs), paresthesias, and menstrual disturbances, including amenorrhea and menorrhagia.

Hypothyroidism may also be associated with bradycardia and diastolic or systolic hypertension. Laboratory abnormalities associated with hypothyroidism include elevated cholesterol and triglycerides and normochromic, normocytic anemia.

NEUROPSYCHIATRIC SYMPTOMS. Neuropsychiatric symptoms are classically associated with overt hypothyroidism; however, mood and cognitive symptoms have also been found in patients with SCH, as well as euthyroid antibody-positive patients. The ongoing autoimmune process underlying the majority of cases of hypothyroidism may be associated with nonspecific symptoms of decreased energy and decreased motivation, as seen in other autoimmune conditions. Depressed mood, apathy, psychomotor retardation, slowed cognition with memory and concentration complaints occur in overt hypothyroidism. Neuropsychiatric testing confirms the subjective complaints, as hypothyroid patients demonstrate decreased scores on complex attention tasks, memory, especially verbal memory, information processing speed, impaired working memory, and implicit learning. SCH is also associated with cognitive impairment, primarily in patients younger than 75 years old. Both overt and SCH are associated with poorer antidepressant treatment response in patients with major depressive disorder.

A dramatically altered mental status can occur in patients with chronic autoimmune thyroiditis who develop encephalitis or vasculitis. This rare condition, Hashimoto encephalopathy, is associated with altered sensorium, delusions, auditory hallucinations, and catatonia, as well as cognitive dysfunction and even coma. Acute episodes of focal neurologic deficits suggesting a stroke may occur, as well as generalized tonic–clonic seizures and status epilepticus have been reported.

Postpartum thyroiditis may be present with hypothyroidism alone (43%), with hyperthyroidism alone (32%), followed by hypothyroidism (25%). The most common symptoms of the hyperthyroid state are anxiety, palpitations, and fatigue. The hypothyroid state is associated with decreased energy, depressed mood, and impaired concentration. Patients who are euthyroid, but are thyroid antibody positive, have been found to have a higher prevalence of depressive symptoms in some studies.

Ms. A, a 54-year-old woman with no previous psychiatric history, experienced depressed mood, lethargy, apathy, weight gain, and problems with memory and concentration over the past year. Initially, she attributed her symptoms to menopause, but her symptoms have not improved after six months of hormone replacement therapy. Her TSH level was elevated at 14 mIU/L. The patient was started on levothyroxine (Levoxyl) 0.05 mg/day, and within 8 weeks, she experienced an improvement in energy, interest in her activities, and subjective memory. A TSH drawn 8 weeks after initiation of thyroid replacement was normal at 3.3 mIU/L.

Ms. K, a 32-year-old woman with recurrent spontaneous abortion and recurrent major depressive disorder presented to the ER at 14 weeks postpartum for major depression with passive suicidal ideation. Ms. K had discontinued her antidepressant in the first trimester of pregnancy and remained euthymic until she became increasingly anxious a few weeks before delivery. Her pregnancy was complicated by gestational diabetes, and she delivered prematurely at 36 weeks. She restarted her serotonin reuptake inhibitor (SSRI) at her prepregnancy dose within a week of delivery, but her mood continued to deteriorate. She experienced the new onset of tachycardia, increasing insomnia, agitation, irritability, inability to focus, and complaints of panic.

After approximately a month, her symptoms developed into severe depression with lethargy and apathy. Lithium 600 mg at bedtime was added for treatment-resistant depression, but her symptoms did not improve. When she experienced a decrease in milk supply and increasing problems with memory and concentration, she began to feel she could not care for her infant, experienced suicidal ideation, and was voluntarily admitted to the hospital. Laboratory tests were as follows: TSH 8 mIU/L, free T4 0.5 ng/dL, T3 normal. Anti-TPO and antithyroid antibody titers were elevated.

Physical examination revealed a painless goiter and ultrasound hypoechogenicity. Ms. K was diagnosed with postpartum thyroiditis, and Ms. K and her husband responded well to psychoeducation about autoimmune thyroiditis and its role in her treatment resistance. Synthroid was initiated, and lithium was discontinued. Buproprion was also added to her SSRI, and within 4 weeks, she experienced significant improvement in her mood and cognition. At 6 weeks, Ms. K’s TSH and T4 levels were normal, and she continued on Synthroid with plans for thyroid function tests (TFTs) again in 6 months, a trial taper off the Synthroid after 12 months, and TFTs a third time 1 to 2 months after that.

Pathology and Lab Examination. Serum TSH is used to screen for hypothyroidism, and normal values range from 0.3 to 7.5 mIU/L. The upper limits of normal increase with age and obesity. Elevated TSH levels with decreased FT4 and T3 levels indicate hypothyroidism.

When the TSH is elevated, but the thyroid hormone levels are in the normal range or lower limit of normal, this is consistent with SCH. When the FT4 is low, but the TSH level is normal, this suggests the rarer central hypothyroidism caused by hypothalamic or pituitary dysfunction.

Differential Diagnosis. The differential diagnosis of hypothyroidism includes affective disorders, including major depression and dysthymia. Since autoimmune thyroiditis may have periods of acute thyrotoxicosis en route to hypothyroidism, the patients may appear to suffer from a primary panic disorder or generalized anxiety disorder. The cognitive complaints in hypothyroidism should be differentiated from primary dementia disorders such as Alzheimer disease or multi-infarct dementia.

Course and Prognosis. Rates of conversion from SCH to overt hypothyroidism vary from 4.3% to 17.7%. Risk increases with higher baseline TSH levels and the presence of TPOAb. Measurement of TPO and thyroglobulin antibodies (Tg) antibodies is useful for confirming hypothyroidism's immunologic cause and evaluating conversion risk. While 70% of patients with postpartum thyroiditis are euthyroid 12 to 18 months postpartum, a subset of patients will progress to permanent hypothyroidism and require long-term thyroid replacement. Furthermore, the recurrence risk is high at 42%.

Treatment. Oral synthetic thyroxine replacement therapy is used to treat overt central hypothyroidism. Starting doses are adjusted to the patient's age, comorbid medical conditions, body weight, and the severity and duration of hypothyroidism. Doses are adjusted until T4 concentrations have increased and TSH serum levels decrease to normal values.

Despite normalization of TSH and FT4 levels, primary subclinical and overt hypothyroidism treatment may not lead to full remission of neuropsychiatric symptoms. Treatment-resistant depression is frequently associated with SCH, and treatment with T3 is beneficial for treatment response and remission when added to either selective SSRIs or tricycle antidepressants (TCAs). Hashimoto encephalopathy is treated with corticosteroids, and 90% to 98% of patients respond to this treatment.

Hyperthyroidism

Definition and Comparative Nosology. Hyperthyroidism, or thyrotoxicosis, results from the overproduction of thyroid hormone by the thyroid gland. The most common cause is Graves disease. Toxic nodular goiter causes another 10% of cases among middle-aged and older adult patients. Other causes include hyperfunctioning solitary thyroid adenomas, thyroiditis, use of exogenous thyroid hormone, TSH-producing pituitary adenoma, pituitary resistance of suppression of TSH secretion by thyroid hormone, thyroid carcinoma, choriocarcinoma, hydatiform moles, and struma ovarii.

Subclinical hyperthyroidism describes a condition in which circulating thyroid hormone levels are normal, but the TSH response to TRH is blunted or absent. Patients may experience symptoms of hyperthyroidism, such as nervousness, irritability, fatigue, and tachycardia. Subclinical hyperthyroidism may progress to overt hyperthyroidism.

A thyroid storm (thyrotoxic crisis) is a life-threatening syndrome that is usually precipitated by illness or injury but may also develop after withdrawal of ATDs or after therapy with RAI. Symptoms include marked tachycardia, weakness, fever, and altered mental status.

Epidemiology. In the general population, the prevalence of hyperthyroidism is approximately 0.5%. Graves disease is the most common cause, accounting for 80% of cases. Graves disease occurs four times more frequently in women than men and occurs primarily in the third and fourth decades of life. The prevalence of thyroiditis is estimated to occur in 5% to 9% of women in the postpartum period, usually resolving spontaneously by one year after delivery.

Etiology. Patients with Graves disease produce thyroid-stimulating antibodies that activate the TSH receptor, producing increases in thyroid hormone. These patients frequently have a family history of other autoimmune disorders, such as systemic lupus erythematosus and myasthenia gravis. Hyperthyroidism from toxic nodular goiter results from the lack of response of thyroid nodules to normal feedback regulation. In the case of thyroiditis, hyperthyroidism results from inflammation-induced damage to the thyroid gland, which then spills the stored colloidal thyroid hormones. Other causes of hyperthyroidism include excessive thyroid hormone intake and decreases in estrogen levels. Pregnancy or estrogen replacement therapy may increase thyroid-binding globulin, leading to decreased free T4. Thyroid doses need to be adjusted downward to avoid iatrogenic hyperthyroidism when estrogen levels fall.

Diagnosis and Clinical Features. Signs and symptoms of hyperthyroidism include increased pulse, arrhythmias, elevated blood pressure, fine tremor, heat intolerance, excessive sweating, increased appetite, weight loss, palpitations, tachycardia, frequent bowel movements, menstrual irregularities, muscle weakness from the catabolism of muscle protein, exophthalmos, lid lag, infrequent blinking, and hyperactive DTRs. In cases of thyroiditis, the patient may experience pain and tenderness in the thyroid region. Psychiatric features include nervousness, fatigue, insomnia, mood instability, and dysphoria. Speech may be pressured, and patients may demonstrate a heightened activity level.

Neuropsychiatric symptoms include a short attention span, impaired recent memory, and an exaggerated startle response. There may be visual hallucinations, paranoid ideation, and delirium in severe cases. Although some symptoms of hyperthyroidism resemble those of a manic episode or mixed manic episode, an association between hyperthyroidism and mania has rarely been observed.

Pathology and Laboratory Examination. Laboratory examination reveals elevated levels of total T4 and FT4. Levels of total and free T3 are usually elevated but are not necessary to diagnose hyperthyroidism. TSH level is abnormally low, as elevated thyroid hormones suppress its production, and the TSH response to TRH is blunted. The uptake of RAI by thyroid hormone is increased, except in cases of hyperthyroidism due to thyroiditis or ingestion of thyroid hormone.

Differential Diagnosis. Panic disorder, generalized anxiety disorder, and social and simple phobias present with symptoms that resemble hyperthyroidism, including tachycardia, diaphoresis, and tremulousness. Other conditions that produce symptoms of autonomic hyperarousal include hypoglycemia, diabetes, dehydration, and pheochromocytoma. Chronic obstructive pulmonary disease may produce anxiety, diaphoresis, dyspnea, and tachycardia. In some cases, hyperthyroidism may produce severe agitation resembling a manic state. Certain medications, for example, appetite suppressants, stimulants, and street drugs, such as cocaine and amphetamines, may also produce these symptoms. Laboratory testing distinguishes between hyperthyroidism and other medical conditions.

Course and Prognosis. Rates of remission of Graves disease depend on the type and length of treatment. Patients treated with radioiodine rarely experience a recurrence of hyperthyroidism. However, permanent hypothyroidism is a common complication of RAI.

Patients treated with ATDs for six months have approximately a 30% remission rate compared to 80% in patients treated for two years. Patients who undergo surgical ablation have a relapse rate of approximately 10%; these patients are also at risk for permanent hypothyroidism. Hyperthyroidism due to thyroid nodular disease usually remits with RAI therapy. Thyroiditis is usually a short-lived condition, lasting only a few weeks to months and resolving spontaneously. Treatment of thyroiditis includes supportive measures such as B-adrenergic agents for tachycardia and close follow-up of TFTs since patients with thyroiditis, especially postpartum thyroiditis, are at an increased risk of hypothyroidism.

Treatment. Treatments for Graves disease include ATDs, RAI, and surgical thyroidectomy. The ATDs, propylthiouracil (PTU), and methimazole (Tapazole) inhibit thyroid hormone synthesis. PTU additionally inhibits the monodeiodination of T4 to T3. Methimazole has the advantage of single-day dosing. Initial doses of methimazole range from 10 to 40 mg and, for PTU, from 75 to 100 mg three times a day (t.i.d.). The doses should be adjusted as the patient returns to a euthyroid state. Usual maintenance doses of methimazole are 5 to 10 mg a day, and doses of PTU are 50 to 100 mg a day. A rare and serious side effect of these medications is agranulocytosis. Other side effects include pruritus, rash, and fever. Length of treatment varies from 6 months to 2 years. These medications may not produce a therapeutic effect for several weeks because of the long half-life of T4. Some clinicians combine the ATDs with T4 replacement to prevent iatrogenic hypothyroidism.

Although some patients with Graves disease respond to ATDs, most require RAI. An RAI uptake test done before administration of iodine-131 allows for determining the appropriate dose. Hypothyroidism may be a complication of RAI, and patients may subsequently require thyroid hormone replacement.

A subtotal thyroidectomy is an option for patients who are allergic to ATDs or are unwilling to undergo RAI therapy. Patients should receive ATDs or potassium iodide or b-adrenergic antagonists to be euthyroid before surgery and reduce the risk of postoperative thyrotoxic crisis. Subtotal thyroidectomy has several potential complications, such as hypoparathyroidism and recurrent laryngeal nerve injury. Patients may have recurrent hyperthyroidism after surgery if insufficient thyroid tissue is removed. In general, thyroidectomy is used only if other forms of treatment are not possible.

Beta-blockers, such as propranolol (Inderal), metoprolol (Lopressor), atenolol (Tenormin), or nadolol (Corgard), can provide symptomatic relief while treatment is in progress. Atenolol at 50 to 100 mg per day and nadolol at 80 mg per day have the advantage of once-daily dosing. These medications should be used cautiously in patients with asthma or heart failure. Iodine (given as a potassium iodide solution or as Lugol solution) inhibits thyroid hormone. As the duration of its antithyroid action is brief, its use is limited to pretreatment before surgery, management of thyroid storm, and after radioiodine therapy to speed the reduction in thyroid hormone levels. Treatment for thyroid nodular goiter consists of beta-adrenergic antagonists and radioiodine. Treatment of thyroiditis consists of a brief course (a few weeks) of beta-adrenergic antagonists, as this condition is short-lived. For patients with psychotic symptoms, antipsychotic agents can be used. However, specific antipsychotic agents are best avoided. For example, those most likely to produce tachycardia, like low-potency antipsychotics or QT interval prolongation. Similarly, TCAs should be used with caution, if at all, in these patients. In general, the psychiatric symptoms resolve with successful treatment of hyperthyroidism.

ADRENAL DISORDERS

The adrenal gland produces three major classes of steroid hormones: glucocorticoids, mineralocorticoids, and sex steroids.

The main steroid hormones produced by the adrenal gland are cortisol, aldosterone, testosterone, estrone, and estradiol. The function of glucocorticoids is to promote gluconeogenesis, lipolysis, and glycogenesis, increase catecholamine activity during stress, and inhibit inflammatory reactions. Aldosterone’s effect is primarily to increase renal sodium reabsorption and potassium excretion. Only a small proportion of circulating sex steroids are produced by the adrenal glands, as their primary production site is the gonads.

Production of cortisol is under control of ACTH from the anterior pituitary, which, in turn, is regulated by CRH from the hypothalamus. Negative cortisol feedback occurs at pituitary, hypothalamic, and suprahypothalamic levels. CRH secretion follows a circadian pattern, subsequently producing circadian patterns of ACTH and cortisol levels, with their levels peaking in the early morning and reaching a nadir in the evening. CRH release is also influenced by emotional and physical stress.

Hypercortisolism

Definition and Comparative Nosology. Spontaneous Cushing syndrome results from excess cortisol from the adrenal glands, such as a cortisol-producing adenoma, or from excessive secretion of ACTH, which stimulates the adrenal gland to produce cortisol. Cushing disease, the most common form of spontaneous Cushing syndrome, results from excessive pituitary secretion of ACTH, usually from a pituitary adenoma or rarely ectopically from neuroendocrine tumors. Cushing syndrome is ACTH dependent if the syndrome results from the effect of ACTH on the adrenal gland, and non-ACTH dependent, in cases of excessive cortisol regardless of ACTH levels, for example, as a result of an adrenal adenoma. The hypothalamic–pituitary–adrenal (HPA) axis responsivity to stressors may be influenced by stress in infancy and even in utero. Studies in primates and rodents show that prenatal stress may induce long-lasting effects on the HPA axis regulation. Offspring whose mothers were stressed during pregnancy show greater HPA axis response to stressors than control offspring.

Epidemiology and Etiology. The most common cause of hypercortisolemia is iatrogenic since corticosteroids are used for various medical conditions. The prevalence of Cushing syndrome is approximately 40 per million, and it occurs far more frequently in women, particularly in the third and fourth decades of life. The most common etiology of endogenous Cushing syndrome is an ACTH-producing pituitary adenoma at 70%, while adrenal adenomas are estimated to comprise 20% of cases. Other causes are autonomous ACTH-producing tumors, such as oat cell carcinoma of the lung, islet-cell pancreatic tumors, or thymomas.

The relationship between stressful life events and the onset of Cushing syndrome is controversial, with some studies finding a greater frequency of adverse life events before the development of the disorder, whereas others have found no such association. Similarly, the relationship between depressive symptoms and the onset of Cushing syndrome remains inconclusive. Some studies have reported a high frequency of depression preceding the physical manifestations of the illness, whereas others have found no greater prevalence of depressive symptoms than noted in control groups.

The etiology of cognitive impairment in hypercortisolemic patients is believed to result from the effect of corticosteroids on hippocampal neuronal damage. Prolonged exposure to elevated corticosteroid levels is associated with hippocampal damage and smaller hippocampal volume. The hippocampus is involved in selective attention, learning, and memory, and reduced hippocampal volume is associated with lower scores on memory tests. In electrophysiologic studies, acute cortisol administration reduces the response to relevant but not irrelevant stimuli. This finding suggests that cortisol can impair selective attention and discriminate important information. Several researchers have suggested that the cognitive impairments in patients with clinical depression, many of whom have HPA axis dysregulation, may result, in part, from cortisol effects on the hippocampus. Furthermore, animal models have shown that corticosteroids downregulate brain-derived nerve growth factor (BDNF). BDNF is essential for neuronal growth and plasticity and can facilitate learning and memory.

The exogenous use of steroids may also produce psychiatric sequelae. The likelihood of psychiatric side effects increases in direct proportion to prednisone dosage. A dosage of fewer than 40 mg per day is associated with less than a 2% incidence of psychiatric symptoms, but this rate rises to almost 20% when the dosage exceeds 80 mg per day.

Diagnosis and Clinical Features. Weight gain is a common nonspecific symptom in hypercortisolemia. Other more unique features of Cushing syndrome include a characteristic moon facies or rounded face from the accumulation of adipose tissue around the zygomatic arch, and truncal obesity, as well as the “buffalo hump” appearance from cervicodorsal adipose tissue deposition. The catabolic effect of cortisol on protein produces muscle wasting, slow wound healing, easy bruising, and thinning of the skin, leading to abdominal striae. Bones become osteoporotic, sometimes resulting in pathologic fractures and loss of height. Diabetes may occur in some patients due to gluconeogenesis stimulated by the high level of glucocorticoids. Chronic ACTH stimulation of the adrenal glands may lead to excess androgen production, with secondary hirsutism, acne, and amenorrhea. Hyperpigmentation may also occur, as ACTH is a precursor of a melanocyte-stimulating hormone, which produces a skin-darkening effect. Although cortisol’s mineralocorticoid effects are mild, it produces sodium retention, potassium loss, metabolic alkalosis, and edema at high concentrations. Hypertension occurs in the majority of patients.

Psychiatric symptoms in Cushing syndrome patients are frequent, diverse, and often severe. Approximately 75% of patients report depressed mood. Of these, approximately 60% experience moderate or severe depression. However, the severity of depression does not appear to be influenced by the etiology underlying Cushing syndrome. Depressive symptoms occur more commonly in female patients than male patients with Cushing syndrome.

Hypersensitivity to stimuli and emotional lability are common, especially irritability, which is increased in 86% of patients and is often an early manifestation of the illness. Anxiety is also prominent and usually characterized as generalized anxiety, although panic attacks can occur. Appetite can be dysregulated and may be either increased or decreased, and most patients experience fatigue. Over 65% of patients will complain of sleep problems, and early morning awakening is common. Patients with Cushing syndrome have disrupted sleep patterns similar to patients with major depression: lower sleep efficiency, less total sleep time, shortened time to onset of rapid eye movement (REM) sleep. Somatic symptoms and elevated neuroticism scores on Eysenck Personality Inventory have also been reported with significant improvements after reducing cortisol levels. Social withdrawal may develop due to shame regarding one's physical appearance. Paranoia, hallucinations, and depersonalization are estimated to occur in 5% to 15% of cases.

Manic and psychotic symptoms occur much less frequently than depression, at a rate of approximately 3% to 8% of patients but rising to as much as 40% in patients with adrenal carcinomas and iatrogenic steroids. In cases of iatrogenic hypercortisolemia, the risk of mania and psychosis appears to be dose-dependent, with doses over 80 mg/day of prednisone resulting in 18% of patients suffering severe psychiatric symptoms. The psychiatric disturbances in prednisone-treated patients often appear within the first few days to two weeks of treatment and occur more commonly in women than men.

Cognitive changes are common in Cushing syndrome, with approximately 83% of patients experiencing deficits in concentration and memory. Declarative and verbal memory deficits predominate. The severity of these deficits correlates with plasma cortisol and ACTH levels. However, the severity of cognitive changes does not correlate with the severity of depression; hypercortisolemia appears to have independent effects.

Cognitive deficits are also common in patients exposed to high doses of exogenous steroid hormones, and steroid-induced reversible dementia syndromes have been reported. Neuropsychiatric studies in healthy volunteers have confirmed that acute high-dose steroids may be associated with reversible declarative memory, episodic and working memory deficits, while chronic hypercortisolemia is associated with impaired spatial memory, arousal, attention, and executive functioning.

The withdrawal of steroids can also produce psychiatric disturbances, particularly depression, weakness, anorexia, and arthralgia. Other steroid-induced withdrawal symptoms include emotional lability, memory impairment, and delirium. Withdrawal symptoms have been noted to persist for as long as eight weeks after corticosteroid withdrawal. Rapid withdrawal of steroids can precipitate an adrenal crisis, a potentially life-threatening condition.

Hypercortisolemia is associated with glucose intolerance in 30% to 60% of patients, and diabetes develops in 20% to 50%. Dyslipidemia is also common; patients have high triglyceride levels and elevated cholesterol.

Ms. A, a 42-year-old marathon runner, was diagnosed with tendinitis prior to her 10K race and began prednisone at 60 mg/day. Within a day of starting the medication, she developed insomnia, and sleep decreased from 8 to 6 hours per night. Her speech was pressured, and her energy level increased. She reported that even though she had been advised to decrease her activity, she was unable to resist cleaning her house "from top to bottom." She also described increased irritability and anxiety. Within two weeks of discontinuing the prednisone, her mental state returned to baseline.

Pathology and Laboratory Examination. In Cushing syndrome, plasma and urine levels of cortisol are elevated, and normal circadian rhythms of cortisol secretion are blunted or absent. Administration of metyrapone (Metopirone), which blocks cortisol synthesis, is usually followed by increased cortisol production. It produces no such increase in cases of adrenal or ectopic origin, whereas, in cases of pituitary origin, postmetyrapone response is augmented.

Another test that distinguishes the pituitary from adrenal etiologies of Cushing syndrome is the dexamethasone suppression test (DST). In cases of Cushing from an adrenal origin, administration of 1 mg of the synthetic steroid dexamethasone (Decadron) does not suppress cortisol. In contrast, it suppresses cortisol in the pituitary Cushing syndrome. The overnight DST is considered the procedure of choice for screening for Cushing syndrome. It involves administering 1 mg dexamethasone at midnight, and a plasma cortisol level is obtained the following morning. A plasma cortisol level of greater than 200 nmol per liter indicates a high likelihood of Cushing syndrome. The overnight 1 mg DST is simple to administer and produces greater specificity and sensitivity than other screening procedures, such as measuring urinary-free cortisol. As plasma cortisol levels vary considerably across the day, isolated measurement of plasma cortisol does not provide adequate screening for Cushing syndrome. DST may also show nonsuppression in major depression, alcoholism, pregnancy during acute physical or emotional stress.

The cognitive problems classically associated with hypercortisolemia appear to be related to the hippocampal neuron damage as neuroimaging studies show smaller hippocampal, cingulate gyrus, and medial frontal gyrus volumes in patients. Computed tomography (CT), magnetic resonance imaging (MRI) scans, and autopsy studies show global cerebellar and cerebral volume loss in endogenously and exogenously corticosteroid-exposed patients. A quarter of patients continue to have residual mood and cognitive complaints after treatment, and brain volume often does not completely return to the previous volume.

Differential Diagnosis. Patients with Cushing syndrome commonly have mood changes similar to those with a major depressive episode. Both groups of patients have a shortened REM latency, elevated levels of cortisol, and abnormal findings on DSTs. The physical manifestations of Cushing syndrome help distinguish between the two diagnoses but may be absent early in the illness. Other distinguishing symptoms of Cushing syndrome are greater irritability and mood lability than in most patients with unipolar depression and a greater tendency to feel best in the morning hours, in contrast to patients with major depression, melancholic subtype, who tend to feel worse at that time.

Treatment and Course. The treatment of choice for pituitary ACTH-producing tumors in Cushing disease involves surgical resection, transsphenoidal adenotomy, and if unsuccessful, pituitary irradiation. Adrenal adenomas and carcinoma are removed surgically, and in the case of carcinomas, chemotherapy is subsequently instituted. Medications that inhibit cortisol production, such as ketoconazole or metyrapone, are used when patients are awaiting surgery or radiation or have unsuccessful first-line treatments. These drugs are limited by side effects, such as hepatotoxicity (ketoconazole).

Mifepristone (RU486), a progesterone and glucocorticoid receptor antagonist, is a new addition to the medical treatment of Cushing syndrome. It received FDA approval in 2012 for the treatment of hyperglycemia in patients who are not surgical candidates or who have not achieved remission after surgery. Mifepristone has also been studied in patients with psychotic depression, a condition associated with HPA axis dysregulation and abnormalities in glucocorticoid receptors.

After treatment for hypercortisolemia, either with surgery or medication, most patients experience improved hypertension and corrected electrolyte imbalances. The physical stigmata and subtle cognitive deficits may take months or years to resolve. After successful treatment of Cushing syndrome, most patients experience significant improvement in depressive and anxiety symptoms. For patients undergoing prednisone treatment, the simultaneous use of lithium and typical (haloperidol) or atypical antipsychotic medications help prevent the development of manic or psychotic symptoms. Steroid psychosis resolves after discontinuation of the medication. Prednisone should be tapered, rather than discontinued abruptly, to allow the pituitary gland to resume its production of ACTH.

Adrenocortical Insufficiency

Definition and Comparative Nosology. Adrenocortical insufficiency occurs when the adrenal gland inadequately produces three major steroid hormones—glucocorticoids, mineralocorticoids, and sex steroids. In cases of primary adrenal insufficiency, adrenal hypofunction results from adrenal gland disease (e.g., autoimmune disease, infection, idiopathic atrophy, and metastatic tumor).

ACTH and CRH levels rise in response to low levels of adrenal steroids. Secondary adrenal insufficiency results from deficient ACTH secretion because of pituitary disease. Tertiary adrenal insufficiency refers to the deficient hypothalamic secretion of CRH. Adrenal aldosterone production tends to be less affected by secondary or tertiary forms of adrenal insufficiency.

Epidemiology. The prevalence of primary adrenal insufficiency (Addison disease) is estimated at 110 cases per 1 million population, with an incidence of 5.6 cases per 1 million per year.

Etiology. Autoimmune destruction of the adrenal cortex produces more than 90% of primary adrenal insufficiency disease cases. Secondary and tertiary adrenal insufficiency result primarily from the withdrawal of exogenous corticosteroids or ACTH, as their administration suppresses the HPA axis. Adrenal insufficiency secondary to HPA axis suppression is now the most common cause of adrenal insufficiency as these medications are now so widely used in multiple medical conditions, from asthma to organ transplantation. Less common etiologies include pituitary tumors, trauma, infectious disease (e.g., human immunodeficiency virus [HIV]), infarction, and surgical ablation.

Diagnosis and Clinical Features. Clinical symptoms usually do not become apparent until more than 90% of the adrenal cortex is damaged. Adrenal insufficiency often goes misdiagnosed for years due to its insidious onset and generally generic early symptoms. Patients often undergo extensive medical testing before receiving a diagnosis; many are initially diagnosed with primary psychiatric or gastrointestinal problems. Adrenal insufficiency can be life-threatening, particularly in cases of acute adrenal insufficiency, such as after adrenal hemorrhage from an infectious process or the use of anticoagulants. Physical signs and symptoms include weakness, hypoglycemia, hyponatremia, hyperkalemia, nausea, diarrhea, fever, and hypotension.

Chronic adrenal insufficiency produces more subtle symptoms, including loss of libido, fatigability, salt craving, weight loss, vitiligo, nausea, loss of body hair, muscle cramps. ACTH is metabolized to alpha-melanocyte–stimulating hormone, which promotes melanocyte activity, so hyperpigmentation of the skin may develop as ACTH levels rise.

Psychiatric symptoms of adrenal insufficiency include apathy, irritability, anxiety, and depression, ranging from mild to severe. Symptoms may develop slowly over months or years. Even patients receiving corticosteroid replacement therapy are at an increased risk of developing severe affective disorders. While psychosis is rare in adrenal insufficiency compared to hypercortisolemia, case reports exist.

The suspicion of adrenal insufficiency occurs if plasma cortisol levels from the early morning are low since this is the time when they should be peaking. A diagnosis of adrenocortical insufficiency is primarily based on plasma cortisol measures during the ACTH stimulation test. After intravenous administration of 250 micrograms (ug) cosyntropin, cortisol levels usually peak in 30 to 90 minutes. A normal test is defined as a baseline cortisol level of at least 5 ug/dL, a rise between baseline and stimulated cortisol levels of greater than 7 ug/dL, and a stimulated cortisol level of greater than 18 ug/dL. Low cortisol responses to this test indicate adrenal insufficiency. Adrenal insufficiency due to hypothalamic disease is associated with an adequate response of ACTH to CRH administration, but cortisol levels may fail to rise because of adrenal atrophy after prolonged ACTH deficiency.

Pathology and Laboratory Examination. Common laboratory findings include low serum levels of sodium and elevated levels of potassium and low or normal plasma cortisol levels. The diagnosis of adrenal insufficiency is particularly likely if plasma cortisol levels are drawn in the early morning (when cortisol levels should be peaking) and are low (less than 3 ug/dL). In cases of primary adrenal insufficiency, ACTH levels are elevated, and cortisol levels do not rise after ACTH stimulation. In adrenal insufficiency resulting from pituitary disease, ACTH levels are low or normal, and the response to CRH is blunted. If the adrenal insufficiency results from hypothalamic disease, the ACTH response to CRH administration is adequate, but cortisol levels may fail to rise because of adrenal atrophy after prolonged ACTH deficiency.

Differential Diagnosis. The symptoms of chronic adrenal insufficiency significantly overlap with those of depression. The symptoms of easy fatigability, muscle cramps, anorexia, and lethargy may also lead to chronic fatigue syndrome. In cases of acute adrenal insufficiency, such as after abrupt discontinuation of steroid administration, patients may develop a psychotic or delirious syndrome.

Course and Prognosis. Physical and mental symptoms respond quickly to steroid replacement. A patient's prognosis depends largely on the underlying cause of the adrenal dysfunction. In cases of chronic adrenal insufficiency, patients have an unremitting illness requiring lifelong steroid treatment.

Treatment. Acute adrenal insufficiency requires immediate treatment with IV hydrocortisone (Hydrocortone), in addition to fluid replacement with saline solution and potassium supplementation. In cases of primary adrenal insufficiency, a mineralocorticoid (e.g., fludrocortisone) is also required. In cases of chronic adrenal insufficiency, prednisone or hydrocortisone is administered orally as maintenance treatment. Patients should be encouraged to carry a card or bracelet identifying their diagnosis, medications, dosages, and physician.

Recently, treatment of adrenal insufficiency has focused on the prevention of adrenal insufficiency due to steroid withdrawal. Discontinuation of steroids should be slow to prevent adrenal insufficiency associated with abrupt steroid withdrawal. When adrenal insufficiency is primary, a mineralocorticoid is required. In cases of chronic adrenal insufficiency, prednisone or hydrocortisone is administered orally as maintenance treatment.

DIABETES MELLITUS AND INSULIN RESISTANCE

Diabetes is a significant health problem affecting over 34.2 million Americans, and 20% do not know they have diabetes. Twenty-six percent of all Americans are prediabetic, whereas 80% do not know they are prediabetic. It currently ranks as the eighth most common cause of mortality in the Unites States. Most of the morbidity and mortality of diabetes are from long-term complications. The increasing economic and medical burdens of diabetes currently exceed 327 billion dollars annually, and the medical burdens of diabetes have made this condition a focus of intense public health concern.

Furthermore, the socioeconomic cost of diabetes is expected to escalate with the increase in its prevalence due to the aging population, the increase in visceral obesity in American society, a trend toward physical inactivity, and a growing population of susceptible African Americans, Native Americans, and Asian Americans. Furthermore, diabetes is of great concern in the psychiatric population, as recent research has shown an increased prevalence of noninsulin-dependent diabetes mellitus (NIDDM) in psychiatric patients compared to the general population. Many psychiatric medications increase the risk of the development of this disorder.

Diabetes mellitus is a syndrome that includes a heterogeneous group of disorders characterized by hyperglycemia and absolute or relative insulin deficiency. Type I diabetes (DMI) accounts for approximately 10% of the cases, whereas type II diabetes (DMII) accounts for more than 85%.

Type I Diabetes

Definition and Comparative Nosology. DMI (insulin-dependent diabetes), previously known as juvenile-onset diabetes, occurs because the pancreatic Islets of Langerhans beta cells are damaged due to an autoimmune attack. Pancreatic islet cell antibodies are present in 95% of new-onset DMI patients. Over time, insulin deficiency develops, and glucose cannot be mobilized into the body’s cells for use as energy, resulting in hyperglycemia. While the damage to insulin-producing cells in DMI occurs over years, the symptoms may occur over days to weeks, and patients may present in a hyperglycemic crisis.

There is considerable evidence that genetic and environmental factors are significant determinants of DMI. Although the general public has a 0.4% chance of developing DMI, the risk increases to approximately 2% if the mother has diabetes and 6% if the father or siblings have DMI. Concordance rates for DMI reach 50% in monozygotic twins. In terms of genetic markers, DMI is primarily determined by genes in the HLA region of chromosome 6. The HLA-DQ locus appears to be the best single marker of susceptibility, particularly among whites. However, recent genome screens have identified at least 50 additional loci that may also contribute to disease risk. Although DMI is likely a polygenic disorder, epidemiologic patterns, including seasonal and temporal changes in incidence, suggest that environmental factors are involved. However, except for the possible role of viruses and infant nutrition, the environmental factors that precipitate DMI have not been well established.

Epidemiology and Etiology. DMI is the most common form of diabetes in people younger than 20 years of age, but it can occur at any age. Most cases occur before 35 years of age, with peak onset during the second decade of life. DMI affects approximately 1.7 to 3.4 million people in the United States and is the most common chronic metabolic disorder to affect children. It is most prevalent in White populations, especially those in Scandinavia, and is rare in people of Asian or African descent.

Type II Diabetes

Definition and Comparative Nosology. Unlike people with DMI, people with DMII produce insulin. DMII is generally preceded by a clinically silent phase of impaired fasting glucose or impaired glucose tolerance. Insulin resistance, defined as a defect in the ability to drive glucose to its main target tissues, is the primary pathophysiologic feature in most cases. Despite intensive research, the pathologic mechanisms behind insulin resistance are still not fully understood. A variety of cellular and molecular defects have been implicated in the condition. These include dysfunctional insulin receptors, aberrant receptor signaling pathways, and abnormalities in glucose transport or glucose metabolism. Despite a continuing decline in insulin sensitivity, hyperinsulinemic individuals maintain normal or near-normal blood glucose levels. However, as insulin resistance gradually increases over time, the pancreas cannot compensate by increasing insulin secretion, and progressive hyperglycemia develops. Not all people with insulin resistance develop diabetes. Beta-cell function also plays an essential role. Whether beta-cell dysfunction is a primary or secondary factor remains unclear. Impaired beta-cell function may result from exhaustion—long-term exposure to insulin resistance—and prolonged, increased demand for insulin secretion that eventually exceeds beta-cell capacity. Minimal elevation of plasma glucose during the deterioration of glucose tolerance may also impair beta-cell function—glucose toxicity. Some people may maintain high levels of sustained insulin secretion, whereas others may have acquired or inherited an abnormality that makes beta cells susceptible to exhaustion of glucose toxicity. Such defects could include reduced islet-cell mass, limited replicative capacity, and more rapid age-related functional or structural abnormalities.

Epidemiology and Etiology. DMII affects approximately 10% of the U.S. population. It is most common in people older than 45 years of age who are overweight, although the incidence of DMII among young obese individuals is rising.

Obesity appears to be the strongest risk factor for the development of DMII, especially among women. Among the psychiatric population, obesity may be iatrogenic, specifically with antipsychotic use. Lack of exercise, smoking, and a poor diet are risk factors for developing DMII. DMII is most prevalent among African Americans, Latinos, Asian-Pacific Islanders, and Native Americans.

Diagnosis, Clinical Features, and Laboratory Examinations

Classic symptoms of DMI and DMII include polyphagia, polyuria, and polydipsia. Other symptoms include blurred vision, fatigue, weight loss, poor wound healing, dry mouth, dry or itchy skin, or recurrent infections (e.g., vaginal yeast infections, groin rash, and external ear infections). Patients with either type of diabetes may also present with acute metabolic compensation and dehydration, nausea, vomiting, abdominal pain, confusion, or coma. Individuals with DMI often experience weight loss developing over several days associated with the early stages of the disease. The onset of these first symptoms may be fairly abrupt or more gradual. Many patients with DMII remain asymptomatic for years with subtle signs and symptoms, such as neuropathy.

Three clinical measures may be used to diagnose diabetes:

1. A fasting plasma glucose test and a value of ≥126 mg/dL.
2. A nonfasting plasma glucose value of ≥200 mg/dL in people with diabetes symptoms.
3. An abnormal oral glucose tolerance test with a 2-hour glucose value of ≥200 mg/dL.

The hemoglobin A1c test (HbA1c) has become a standard method of diagnosing DMI and DMII and monitoring treatment over time. The HbA1c reflects glycemic control over the past 2 to 3 months; two values over 6.5% are diagnostic for diabetes.

Insulin Resistance, Diabetes, and Psychiatric Illness

An emerging body of literature indicates associations between DMI, DMII, and insulin resistance (a frequent precursor to DMII) and psychiatric illnesses.

DMI. Eating disorders (ED) occur reasonably often in adolescents with DMI and present challenges in diagnosis and management. A meta-analysis of recent studies reports that ED occurs in 8.2% of DMI patients compared to 2.8% of controls. Insulin omission is a unique purging behavior found in up to 40% of DMI patients; risk factors include female sex and higher BMI, as insulin is omitted to lose weight. EDs should be considered in patients with poor glycemic control or multiple episodes of diabetic ketoacidosis.

DMII. DMII and major depression appear to have a bidirectional relationship in disease incidence. For instance, major depression has been demonstrated to be a significant factor in causing hospital admissions and death in persons with DMII, and antidepressant therapy with selective 5HT reuptake has been reported to ameliorate insulin sensitivity in DMII patients.

Conversely, while the risk estimates vary depending on populations studied, measures, and diagnostic criteria, studies consistently report that DMII is associated with an increased risk for major depression. Depression is associated with an increased risk for insulin resistance and DMII.

Shared lifestyle factors and histories, and biologic mechanisms underlie this complex relationship. For instance, there may be shared genetic mechanisms, and several genes have been identified that may play a role in both conditions, including genes that encode glycogen synthase kinase 3 and the dopamine 2 receptor. Another potential shared pathology is insulin resistance. Insulin resistance is a proinflammatory state associated with elevations in cytokines. Inflammation has been associated with the development of symptoms of depression in both animal models and humans, and patients with major depression with elevated inflammatory markers are less responsive to antidepressants and lithium. In addition, shared risk factors such as a sedentary lifestyle, or a diet rich in saturated fat and low in vitamin D, also contribute to the risk for both diseases.

Finally, psychotropic medication use, particularly atypical antipsychotics (SGAs), increases the risk of DMII. Olanzapine and clozapine have been associated with the highest risk for incident DMII, while aripiprazole, ziprasidone, and lurasidone appear to present a lower risk. While the weight gain and risk for DMII differs among the SGAs, the American Diabetes Association, the American Psychiatric Association, and the International Society for Bipolar Disorders each have published guidelines that call for a thorough screening of patients before initiation of SGAs that include weight, blood pressure, fasting glucose and lipids, and documentation of the family history of diabetes and cardiovascular disease. Consensus guidelines recommend metabolic monitoring at regular intervals to intervene early in developing a metabolic syndrome.

Insulin Resistance. Insulin resistance, even in the absence of a DMI or DMII diagnosis is increasingly recognized as an important regulator of mood and cognition. Insulin resistance has recently been linked to increased mood instability in unipolar and bipolar patients. A meta-analysis of 20-plus studies showed significantly increased insulin resistance in depressed patients compared to control subjects. A Dutch cohort study found that individuals with insulin resistance had 51% greater odds of nonremitted major depression than nondepressed controls. In the same study, insulin resistance was also positively associated with depressive symptoms and chronicity. A subsequent study among people with no history of mood disorder found that a moderate increase in insulin resistance was associated with an 89% increased rate of incident major depressive disorder over a nine-year follow-up.

A meta-analysis of adjuvant pioglitazone, an insulin-sensitizing agent, improved nonremitted major depression in nondiabetic patients. However, more research into the mechanisms that underpin the link between insulin resistance and psychiatric disease and how this affects treatment is required.

Treatment. DMI patients require insulin treatment. While patients with DMII can partially manage it by controlling their weight, watching their diet, and exercising regularly, the American Diabetes Association recommends treatment with diabetes oral glucose-lowering medication (e.g., sulfonylureas, meglitinides, biguanides, thiazolidinediones [TZDs], and alpha-glucosidase inhibitors) upon diagnosis. Generally, metformin (Glucophage) is first-line and titrated to the maximum dose unless side effects develop. If one agent is ineffective in reaching the HbA1c goal, a second agent is added. Some patients require therapy with three oral hypoglycemics. Insulin treatment is indicated for people with DMII when lifestyle changes and oral medications fail. Insulin can be combined with oral agents.

At the time of diagnosis, 20% of people with DMII already have other risk factors that contribute to the high incidence of chronic complications, including blindness, renal failure, neuropathy, peripheral vascular disease, stroke, and myocardial infarction. Therefore, it is recommended that patients with DMII undergo a complete assessment of the adequacy of their metabolic control, risk factors for vascular complications, and extensive education on self-management. Because 80% of patients with DMII are overweight, particular emphasis should be placed on nutritional education and exercise. Because of the complications associated with diabetes, frequent follow-up visits should be scheduled, dependent on the severity of complications that a patient may be experiencing. At a minimum, patients with DMII should have an annual complete physical examination, including a foot examination and examination of any systems found to be abnormal previously. A dilated eye examination should be performed annually by an ophthalmologist. The Center for Disease Control recommends that HbA1c be checked at least twice a year. A fasting lipid profile and testing for microalbuminuria should be performed annually.

More data are needed on the effect of oral hypoglycemics on psychiatric outcomes among patients with comorbid psychiatric illness and DMII. Furthermore, psychiatric illness, particularly major depression, is associated with reduced medication compliance and self-care behavior among people with DMII. Therefore, coordinated care for psychiatric illness and DMII may be beneficial.

POLYCYSTIC OVARY SYNDROME

Definition and Clinical Features

Polycystic ovary syndrome (PCOS) is the most widespread endocrine disorder among women. Based on the Rotterdam criteria adopted by the National Institute of Health in 2012, PCOS occurs in up to 15% of women in the general population of reproductive-age women. The prevalence of PCOS is 30% to 75% among women presenting for menstrual irregularities or hirsutism. First, the clinician should exclude other causes of hyperandrogenism, such as congenital hyperplasia, Cushing syndrome, and androgen-secreting tumors. PCOS is then diagnosed if two of three conditions or symptoms are present: oligo-ovulation or anovulation, clinical or biochemical signs of hyperandrogenism, or ultrasound-documented polycystic ovaries. Notably, normal-appearing ovaries no longer rule out this diagnosis if the other two criteria are met, as PCOS is now recognized as primarily an endocrine disorder. Clinical features of PCOS include hirsutism, coarse hair distributed in a male pattern such as the chin, upper lip, periareolar breast area, male pattern alopecia, and acne. Irregular menstrual cycles and infertility may be the presenting symptom in some women.

BMIs over 25 kg/m (overweight) and 30 kg/m (obese) are common in PCOS, affecting 20% to 80% of patients. However, some women with PCOS are normal or underweight, and weight is not a diagnostic criterion. Sleep apnea is also increased in PCOS.

Etiology

The fundamental defect in PCOS remains unclear, but theories implicate dysfunction of the hypothalamic–pituitary–gonadal axis (HPG) (Fig. 27.7–1). GnRH appears to be insensitive to ovarian steroids, leading to a normal negative feedback mechanism failure. Increased LH pulse amplitude and frequency result in increased thecal cell androgen synthesis.

FSH synthesis is reduced in the granulosa cells, resulting in impaired follicular development. An elevated LH to FSH ratio is one of the most common laboratory results in PCOS. Normally, FSH regulates aromatase activity, leading to estrogen production from androgen precursors in the ovary, but in PCOS, the impaired FSH secretion leads to less estrogen production and more androgen production.

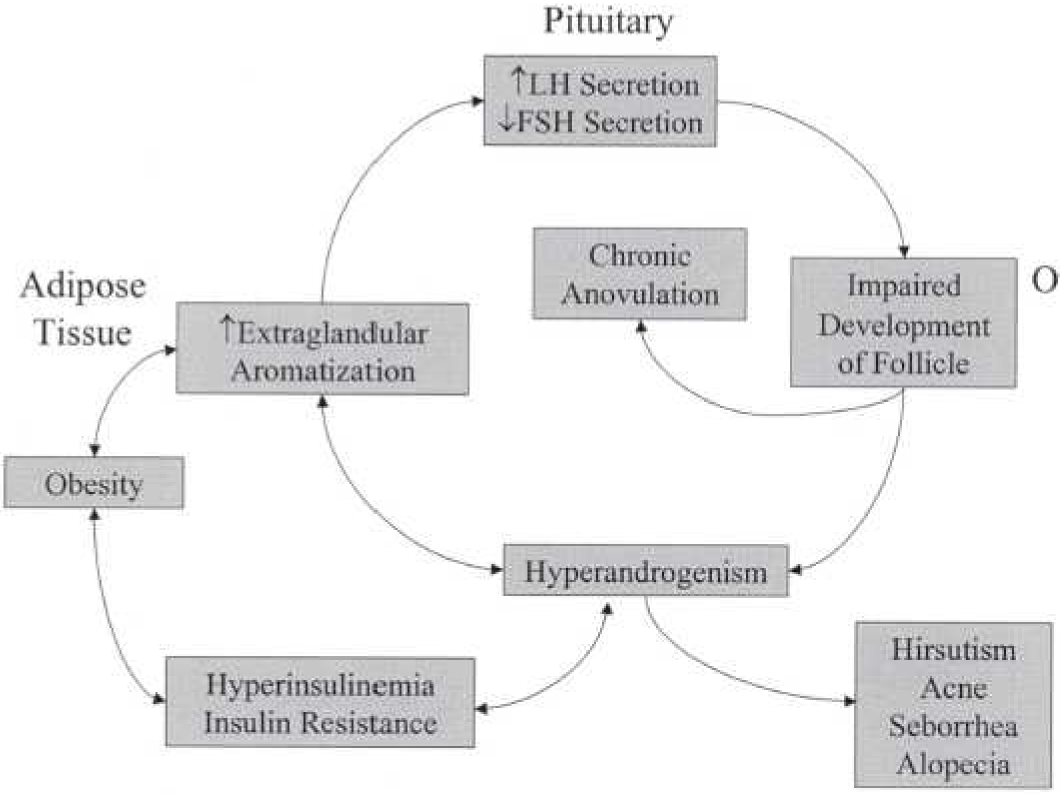
Pituitary
↑LH Secretion
↓FSH Secretion
Adipose Tissue
↑Extraglandular Aromatization
Chronic Anovulation
Impaired Development of Follicle
Ovary
Obesity
Hyperandrogenism
Hyperinsulinemia
Insulin Resistance
Hirsutism
Acne
Seborrhea
Alopecia

FIGURE 27.7–1. Proposed mechanism for perpetuating peripheral endocrine disorders. FSH, follicle-stimulating hormone; LH, luteinizing hormone.

Insulin resistance is also a critical pathogenic mechanism in PCOS, and women with PCOS have high rates of glucose intolerance regardless of weight. Hyperinsulinemia is a compensatory increase in insulin secretion due to peripheral insulin resistance. Approximately 50% of women with PCOS are obese, which may contribute to insulin resistance and other symptoms of PCOS. A strong association between insulin resistance and obesity has been recognized for years, yet the mechanism by which insulin resistance may provoke obesity, and vice versa, are still unclear. Obesity itself increases the risk of insulin resistance, although insulin resistance can also occur in the absence of obesity. Insulin sensitivity is decreased by 35% to 40% in PCOS patients, independent of obesity.

Furthermore, increased androgen levels are associated with insulin resistance and obesity. Insulin is known to increase ovarian androgen production in women with PCOS via stimulation of the thecal cell of the ovaries. Furthermore, increased insulin levels lead to decreased sex hormone-binding globulin production by the liver, which leads to more free-circulating testosterone.

Clinical and Laboratory Examination

Central obesity is common, but many women with PCOS are normal weight. On ultrasonography, the ovaries are commonly enlarged alongside multiple cystic follicles. Normal-appearing ovaries, however, do not rule this condition. Because of the absence of ovulation, estrogen is unopposed by progesterone, and vaginal ultrasound may reveal endometrial hyperplasia and carcinoma.

LH levels are usually elevated, whereas FSH levels are low or normal, and androstenedione, testosterone, and estrone levels are elevated. Fasting blood sugar, HbA1c may be elevated later in the illness due to insulin resistance. Undiagnosed DMII is 7 to 10 times more common in women with PCOS than women without PCOS of similar age.

Differential Diagnosis

Rapid onset of marked virilization, such as low voice or clitoromegaly, should raise the suspicion of an adrenal or ovarian androgen-secreting tumor. Serum testosterone and dehydroepiandrosterones (DHEAs) are significantly increased. Hyperprolactinemia and thyroid disorders are associated with menstrual irregularities and should always be ruled out through serum PRL levels and TFTs. Congenital adrenal hyperplasia caused by 21-hydroxylase deficiency is a rare cause of virilization and menstrual abnormality.

PCOS and Psychiatric Illness

Women with PCOS have an increased prevalence of major depression compared to age-matched controls, and a meta-analysis controlling for BMI found rates as high as four times greater than controls.

The etiology of this higher rate of depression is controversial since studies are conflicting, and total and free testosterone and androstenedione levels do not consistently differ between women with PCOS with depression compared to PCOS women without depression.

Insulin resistance may be a common pathologic link between depression and PCOS (see Fig. 27.7–1). Insulin resistance was found to be higher in PCOS women with higher depression scores than PCOS women with lower scores. Metformin has been shown to decrease depression and insulin resistance in PCOS patients. However, improvement in insulin resistance does not appear to be the only mechanism by which insulin-sensitizing drugs improve mood in PCOS patients. Treatment with pioglitazone, a TZD insulin sensitizer, was associated with a more significant reduction in depression symptoms than metformin in patients with major depression and PCOS. However, the homeostatic model of insulin resistance, HOMA-IR, did not differ.

Anxiety symptoms are also higher in women with PCOS compared to controls. However, few studies have examined the prevalence of anxiety disorders in this population using standardized measures. Psychological reactions to the physical changes associated with PCOS, including virilization and obesity, have been proposed to contribute to depressive and anxiety symptoms.

Bipolar disorder has been associated with PCOS, but controversy remains whether the higher prevalence of PCOS is associated with mood-stabilizing agents, underlying comorbidity, or shared pathologic mechanisms. For instance, genetic predisposition likely plays a role in developing insulin resistance in women with bipolar disorder. These women sometimes develop insulin resistance with exposure to mood stabilizers. At the same time, menstrual abnormalities have been found in some bipolar disorder patients to precede medication initiation, suggesting an underlying predisposition to abnormal menstrual cycles in this population.

The antiepileptic drug valproic acid is associated with an increased risk of PCOS both in patients with epilepsy and bipolar disorder. Rates of incident PCOS vary across studies; in the most extensive study to date of bipolar disorder patients, 10.5% of women on valproic acid developed PCOS compared to 1.4% of women on other mood stabilizers.

Ms. K.B. was a 30-year-old white woman diagnosed with PCOS who had a history of treatment-resistant major depression, most recently treated with venlafaxine (Effexor XR). Ms. K. had a serum glucose level of 83 mg/dL and an insulin level of 19.0 µIU/mL. HOMA-IR measured her insulin resistance at 3.9. HOMA-IR values greater than 2.3 are indicative of insulin resistance.

Ms. K. B. subsequently began metformin (850 mg per day, titrated to 2,550 mg per day) and spironolactone (Aldactone) (100 mg per day). After two weeks of these medications, Ms. K.B.'s mood had improved, and at a follow-up visit 3 months later, Ms. K.B. reported feeling euthymic, had lost 10 pounds in the last 3 months, and she began menstruating approximately every 33 days. She continued to experience low mood, decreased energy, and irritability during the week before menses. However, these symptoms resolved with the onset of menses. Because these symptoms did not significantly impair Ms. K.B., no antidepressant medications were prescribed. Laboratory tests showed that her serum glucose level was 89 mg/dL, and her serum insulin level at this time was 15.0 µIU/mL. Her insulin resistance as measured by HOMA-IR had decreased to 2.0.

Treatment

In PCOS treatment, oral diabetic medicines are utilized to improve insulin resistance. Metformin inhibits hepatic glucose production, increasing peripheral insulin sensitivity, although it does not modify pancreatic insulin secretion. The decrease in insulin resistance may be achieved via decreasing gut absorption of glucose, improving glucose uptake by tissues, increasing the number of insulin receptors, or a combination of these. Metformin is widely used to treat insulin resistance in patients with DMII. In treating PCOS, metformin has been shown to decrease insulin resistance, acne, hirsutism, and total and bioavailable testosterone. A significant decrease in body mass index and waist-to-hip ratio, as a result of treatment with metformin, has also been observed as restoration of ovulation.

Several agents can treat hyperandrogenism associated with PCOS. Low-dose oral contraceptives (containing estrogen/progesterone) are frequently first-line treatments for acne and hirsutism. They decrease LH secretion, which leads to decreased thecal cell androgen synthesis.

Some women cannot use oral contraceptives because of the increased risk of deep venous thrombosis. Instead, they may use spironolactone, an antimineralocorticoid, diuretic, and antiandrogen. It was originally used to treat hypertension but is also a weak inhibitor of testosterone biosynthesis. Spironolactone is an effective medication to reduce the hirsutism often seen in PCOS women, spironolactone has also been shown to decrease insulin resistance and fasting insulin levels in PCOS patients, and there has been an observed decline in testosterone levels. Other treatments include antiestrogens, such as clomiphene citrate (Clomid), the first-line treatment for PCOS women who desire pregnancy, and laparoscopic surgery. Weight reduction is strongly recommended in the over 50% of PCOS women who are obese. Hirsutism may respond to spironolactone, but it may be difficult to tolerate, as it may produce mood changes, decreased libido, and fatigue.

HYPERPROLACTINEMIA

Definition and Comparative Nosology

The anterior pituitary produces PRL, which stimulates milk production from the breast and modulates maternal behavior. Its production is inhibited by dopamine, produced by the tuberoinfundibular neurons in the arcuate nucleus of the hypothalamus. Normal levels in women vary from 5 to 25 ng/mL and from 5 to 15 ng/mL in men. Levels fluctuate during the day and peak during sleep. Exercise and emotional stress can elevate PRL levels.

Epidemiology and Etiology

Hyperprolactinemia occurs most frequently due to PRL-secreting pituitary adenomas (prolactinomas) and idiopathic hyperprolactinemia, two conditions more common in women. Typical neuroleptic medications can raise PRL levels as much as 20-fold and do so in a dose-dependent fashion, with PRL levels reaching a plateau. While all typical neuroleptics appear equally likely to raise PRL levels, haloperidol (Haldol) appears to be especially effective because of its strong dopamine receptor D2 (D2) blockade.

The atypical antipsychotics also increase PRL levels, but in general, less frequently than the typical antipsychotics, and the rise is much less, except in the case of risperidone (Risperdal).

The rate of hyperprolactinemia in patients on risperidone is very high, estimated at 70% to 100% in recent studies. Quetiapine (Seroquel) and olanzapine (Zyprexa) are less likely to raise PRL levels, and studies report 10% to 40%, while lurasidone (Latuda) and asenapine (Saphris) have not been found to raise PRL levels. Studies have shown that switching patients with risperidone-induced prolactinemia to aripiprazole, a partial D2 agonist, decreases PRL levels. Clinicians should be aware of the rarer hyperprolactinemia associated with antidepressants, especially when working with infertility patients. All the SSRIs have been associated with case reports. Only one study has been completed, which reported that in patients with normal prolactin levels, 4.5% of men and 22% of women developed hyperprolactinemia after 12 weeks of fluoxetine treatment. Other medications that may elevate PRL levels include oral contraceptives, estrogens, TCAs, metoclopramide, verapamil, and propranolol. Hypothyroidism results in elevated levels of PRL since TRH stimulates PRL release. Acromegaly is also associated with hyperprolactinemia as a result of the lactogenic properties of GH. Hypothalamic disease, such as craniopharyngiomas, is also associated with hyperprolactinemia, as the hypothalamus exerts an inhibitory influence on PRL secretion. Physiologic hyperprolactinemia occurs in pregnant and breastfeeding women; nipple stimulation also elevates PRL levels.

Traumatic childhood experiences, such as separation from parents or living with an alcoholic father, have been reported to produce a greater predisposition to hyperprolactinemia. Stressful life events are also associated with galactorrhea, even in the absence of elevated PRL levels.

Diagnosis, Clinical Features, and Laboratory Examinations

Patients with elevated PRL levels may be asymptomatic, although symptoms usually develop when PRL levels reach 60 to 100 ng/mL. Women may experience amenorrhea, as PRL suppresses normal menstrual cycling, and men may develop gynecomastia. Women and men may experience low libido, decreased fertility, and galactorrhea (inappropriate lactation), although the latter occurs primarily in women. Headache and visual impairment may result in cases of pituitary adenoma. Women with pathologic hyperprolactinemia have been reported to experience elevated levels of depression, anxiety, and hostility compared to controls, which lessen with normalization of PRL levels.

Treatment

A laboratory workup is the initial step in treatment and should include TFTs, a pregnancy test, and a chemistry panel. A brain imaging study should be obtained, either using an MRI with gadolinium or CT with contrast. Bromocriptine (Parlodel), a dopamine agonist, reduces the synthesis and secretion of PRL. In approximately 80% of cases, bromocriptine at dosages of 2.5 to 5 mg twice a day (b.i.d.) or t.i.d. is an effective treatment. However, many patients cannot tolerate the side effects of bromocriptine, which can include severe nausea and dizziness. Alternative medications include quinagolide (Norprolac), pergolide (Permax), and cabergoline (Dostinex). Medication use must continue indefinitely, as medication discontinuation usually leads to a relapse of hyperprolactinemia. Radiotherapy is another treatment strategy for patients who are refractory to medication treatment. Surgical removal through transsphenoidal surgery is used far less often than before the 1990s because of the high recurrence rate of prolactinomas.

FURTHER READINGS

Ajmal A, Joffe H, Nachtigall LB. Psychotropic-induced hyperprolactinemia: a clinical review. *Psychosomatics*. 2014;55(1):29–36.

Altshuler LL, Bauer M, Frye MA, et al. Does thyroid supplementation accelerate antidepressant response? A review and meta-analysis of the literature. *Am J Psychiatry*. 2001;158(10):1617–1622.

Aronson R, Offman HJ, Joffe RT, Naylor CD. Triiodothyronine augmentation in the treatment of refractory depression. A meta-analysis. *Arch Gen Psychiatry*. 1996;53(9):842–848.

Baethge C, Müller-Oerlinghausen B, Bauer M, et al. Supraphysiological doses of L-thyroxine in the maintenance treatment or prophylaxis-resistant affective disorders. *Neuropsychopharmacology*. 2002;27(4):620–628.

Bauduin SEEC, van der Wee NJA, van der Werff SJA. Structural brain abnormalities in Cushing’s syndrome. *Curr Opin Endocrinol Diabetes Obes*. 2018;25(4):285–289.

Bauer M, Glenn T, Pilhatsch M, Pfennig A, Whybrow PC. Gender differences in thyroid system function: relevance to bipolar disorder and its treatment. *Bipolar Disord*. 2014;16(1):58–71.

Bauer M, Silverman DHS, Schlagenhauf F, et al. Brain glucose metabolism in hypothyroidism: a positron emission tomography study before and after thyroid hormone replacement therapy. *J Clin Endocrinol Metab*. 2009;94(8):2922–2929.

Bruce-Jones W, Zolese G, White P. Polycystic ovary syndrome and psychiatric morbidity. *J Psychosom Obstet Gynaecol*. 1993;14(2):111–116.

Cohen ST, Welch G, Jacobson AM, De Groot M, Samson J. The association of lifetime psychiatric illness and increased retinopathy in patients with type I diabetes mellitus. *Psychosomatics*. 1997;38(2):98–108.

Davis JD, Tremont G. Neuropsychiatric aspects of hypothyroidism and treatment reversibility. *Minerva Endocrinol*. 2007;32(1):49–65.

Degner D, Haust M, Meller J, Rüther E, Reulbach U. Association between autoimmune thyroiditis and depressive disorder in psychiatric outpatients. *Eur*

Arch Psychiatry Clin Neurosci. 2015;265(1):67–72.

Feldman AZ, Shrestha RT, Hennessey JV. Neuropsychiatric manifestations of thyroid disease. *Endocrinol Metab Clin North Am*. 2013;42(3):453–476.

Joffe RT, Sokolov ST. Thyroid hormones, the brain, and affective disorders. *Crit Rev Neurobiol*. 1994;8(1–2):45–63.

Kashani L, Omidvar T, Farazmand B, et al. Does pioglitazone improve depression through insulin-sensitization? Results of a randomized double-blind metformin-controlled trial in patients with polycstic ovarian syndrome and comorbid depression. *Psychoneuroendocrinology*. 2013;38(6):767–776.

Kornegay CJ, Vasilakis-Scaramozza C, Jick H. Incident diabetes associated with antipsychotic use in the United Kingdom general practice research database. *J Clin Psychiatry*. 2002;63(9):758–762.

Mcintyre RS, Mancini DA, McCann S, Srinivasan J, Kennedy SH. Valproate, bipolar disorder and polycystic ovarian syndrome. *Bipolar Disord*. 2003;5(1):28–35.

O’Donovan C, Kusumakar V, Graves GR, Bird DC. Menstrual abnormalities and polycystic ovary syndrome in women taking valproate for bipolar mood disorder. *J Clin Psychiatry*. 2002;63(4):322–330.

Oomen HA, Schipperijn AJ, Drexhage HA. The prevalence of affective disorder and in particular of a rapid cycling of bipolar disorder in patients with abnormal thyroid function tests. *Clin Endocrinol (Oxf)*. 1996;45(2):215–223.

Özerdem A, Tunca C, Çımrın D, Hıdıroğlu C, Ergör G. Female vulnerability for thyroid function abnormality in bipolar disorder: role of lithium treatment. *Bipolar Disord*. 2014;16(1):72–82.

Pinhas-Hamiel O, Hamiel U, Levy-Shraga Y. Eating disorders in adolescents with type 1 diabetes: challenges in diagnosis and treatment. *World J Diabetes*. 2015;6(3):517–526.

Rasgon NL, Altshuler LL, Gudeman D, et al. Medication status and polycystic ovary syndrome in women with bipolar disorder: a preliminary report. *J Clin Psychiatry*. 2000:61(3):173–178.

Reynolds-May MF, Kenna HA, Marsh W, et al. Evaluation of reproductive function in women treated for bipolar disorder compared to healthy controls. *Bipolar Disord*. 2014:16(1):37–47.

Samuels MH. Psychiatric and cognitive manifestations of hypothyroidism. *Curr Opin Endocrinol Diabetes Obes*. 2014;21(5):377–383.

Schwartz SA, Weissberg-Benchell J, Perlmuter LC. Personal control and disordered eating in female adolescents with type I diabetes. *Diabetes Care*. 2002;25(11):1987–1991.

Sobrinho LG. The psychogenic effects of prolactin. *Acta Endocrino l(Copenh)*. 1993;129 Suppl 1:38–40.

Starkman MN. Neuropsychiatric findings in Cushing syndrome and exogenous glucocorticoid administration. *Endocrinol Metab Clin North Am*. 2013;42(3):477–488.

Thomsen AF, Kvist TK, Andersen PK, Kessing LV. The risk of affective disorders in patients with adrenocortical insufficiency. *Psychoneuroendocrinology*. 2006;31(5):614–622.

Watson KT, Simard JF, Henderson VW, et al. Incident major depressive disorder predicted by three measures of insulin resistance: a dutch cohort study. *Am J Psychiatry*. 2021;178(10):914–920.

Whybrow PC, Grof P, Gyulai L, Rasgon N, Glenn T, Bauer M. The electronic assessment of the longitudinal course of bipolar disorder: the ChronoRecord software. *Pharmacopsychiatry*. 2003;36 Suppl 3:S244–S249.

Winston AP. Eating disorders and diabetes. *Curr Diab Rep*. 2020;20(8):32.

27.8 Psycho-Oncology

DANIEL MCFARLAND, D.O., M.S., YESNE ALICI, M.D., AND WILLIAM S. BREITBART, M.D., F.A.C.L.P., D.F.A.P.A., F.A.P.O.S.

Psycho-oncology is a sub-discipline of consultation-liaison psychiatry that requires specialized knowledge of cancer and psychiatry. It deals with diagnosing and treating the psychological aspects of cancer along a continuum from prevention to survivorship to the end of life. While the origins of psycho-oncology can be attributed to Donald Sutherland’s study of women after mastectomy, the field did not begin to form until the mid-1970s alongside an emerging concern for patient outcomes in oncology. The field is concerned with the psychological response of patients, their families, and health care providers to the effects of cancer and its treatments. The psychological, behavioral, and social factors that influence cancer risk, detection, adaptation, and survival are encompassed by the field of psycho-oncology and applied to its practice. Psycho-oncology is inherently inter-disciplinary and reflects ideal oncologic approaches to patient-centered care as a multi-disciplinary endeavor. Many fields touch psycho-oncology such as consult-liaison psychiatry, nursing, social work, chaplaincy, psychology, behavioral medicine, integrative medicine, along with complementary and alternative approaches to cancer care. Along with an active proliferation of clinical psycho-oncology programs attached to cancer centers, research centers devoted to psycho-oncology form a vital part of psycho-oncology whose researchers come from many interrelated disciplines, including palliative and integrative medicine oncologic sciences, health, and social psychology, as well as other psychological disciplines. Psycho-oncology training programs, clinical researchers, and practitioners exist throughout the world and are represented by over 50 national organizations as well as an international organization. The American Psychosocial Oncology Society (APOS) is the multidisciplinary national organization representing the subspecialty of psycho-oncology in the United States. The International Psycho-Oncology Society (IPOS) is an international organization that was formed in 1984 in recognition of the need for communication among professionals with interests in psycho-oncology worldwide. Psycho-oncology principles and practice have been established and are part of clinical practice guidelines in oncology. The National Comprehensive Cancer Network (NCCN) publishes distress screening guidelines, which continue to serve as foundational for the intercalation of psychosocial care into oncology. The Institute of Medicine (IOM) published the first policy statement regarding psycho-oncology in 2008. The IOM reported that routine oncology practice must integrate psycho-oncology to achieve quality cancer care. This was further reinforced by the American College of Surgeons Commission on Cancer. The Commission on Cancer now requires accredited cancer centers to have instituted and established psychosocial distress screening programs. In addition to research efforts, psycho-oncology offers evidence-based treatments for common psychological issues experienced in oncology settings. The future of psycho-oncology rests in the implementation and dissemination of evidence-based treatments along with improving recognition of psychosocial issues in oncology.

This chapter will describe the history of psycho-oncology, the psychosocial challenges experienced by cancer patients, the common psychiatric disorders in cancer patients, the psychiatric syndromes specific to tumor site and cancer treatment modality, the therapeutic interventions available, pediatric and geriatric psycho-oncology, issues specific to cancer survivors, patients at genetic risk for cancer, and psychiatric aspects of cancer screening.

HISTORICAL BACKGROUND

The field of psycho-oncology would not be possible without being able to discuss cancer. Unfortunately, cancer was an off-limits discussion for centuries, and patients generally suffered silently. Over the 20th century, descriptions of stigma and cancer-related experiences as well as disclosure of cancer diagnoses emerged as socially accepted. These changes paralleled improvements in oncologic care through surgery, radiation, and systemic therapies. Also, the latter half of the 20th century saw the advent of social sciences where patient self-report was deemed scientifically valid. Thus, psycho-oncology began as an empirically based science of psychological phenomena related to cancer. Yet, the shame, stigma, and superstition of cancer continue to lurk in common parlance and vernacular. Beliefs about cancer and what it signifies continue to haunt patients, families, and societies despite technologic and treatment advances.

For centuries, truth-telling about cancer was seen as cruel and lacking in compassion. This idea is still prevalent in distinct cultures and parts of the world where cancer treatments are not accessible. This paradigmatic approach prevailed as a manifestation of patriarchal medical care. The 20th century also saw advances in informed consent requiring both disclosure and an assessment of patient understanding. The American Cancer Society (ACS) was formed in 1913 to educate the public on the warning signs of cancer and to fight fatalism that interfered with the early presentation for treatment. In 1937 the National Cancer Institute (NCI) was established to seek etiologies and treatments for cancer. In the 1950s, the addition of chemotherapy to the combined treatment modalities resulted in the cure of several childhood tumors. By the early 1970s, with improvements in survival, oncologists became more comfortable in discussing cancer diagnoses with their patients, and the patients’ reluctance to identify themselves as cancer patients diminished. Alongside, hospice programs were developed to improve pain management and palliative care with increasing interest in delivering the best supportive care to patients at the end of life. Clinicians’ enhanced comfort with communicating a diagnosis of cancer, increased concern for palliative symptom control, and the growing interest in quality of life and patient rights emphasized the need for supportive and psychological aspects of care. In the 1980s, psycho-oncology units began to develop in larger cancer centers. First, prevalence studies of psychiatric and psychological sequelae in cancer were reported. Special units were developed throughout Europe and the United States during the 1970s, followed by societies such as the British Psychosocial Oncology Group (1983), the IPOS (1984), and the Japanese Psycho-Oncology Society (1985). Conferences, journals (*Journal of Psychosocial Oncology*, 1982; *Psycho-Oncology*, 1992), textbooks, and training programs followed thereafter. In the 1990s, behavioral research in changing habits (e.g., smoking), diet, and lifestyle improved the education of the public on cancer prevention. Health-related quality-of-life assessments and, more recently, patient-reported outcomes have become part of outcome measures in clinical trials. In the early 21st century, psycho-oncology, a discipline almost 40 years old, continues to grow, with the development of novel psychotherapy modalities for advanced cancer patients; intervention trials to improve symptom control in the terminally ill; increased awareness of the role of communication skills training; researchers' efforts to understand and conceptualize the effects of cancer therapies on the CNS; and recognition of the special needs of older adult cancer patients with the rapidly growing older adult population worldwide.

It is important to emphasize that despite all the developments outlined, historical attitudes toward cancer have contributed to the reluctance of patients and families to identify their emotional problems to the clinicians, even today, especially in underserved populations, different cultures, and several parts of the world.

INCIDENCE OF CANCER AND CANCER MORTALITY RATES IN THE UNITED STATES

In the United States, the lifetime risk of developing cancer is slightly less than one in two for men and slightly more than one in three for women. Cancer is the second most common cause of death after heart disease, accounting for one in every four deaths. The 5-year relative survival rate for all cancers diagnosed between 2007 and 2013 is 67%; this has improved from 50% between 1975 and 1977. Prostate and breast cancers are the most common cancer type in men and women, respectively, but lung cancer is responsible for the highest rates of mortality in both groups. Lung cancer accounts for about 15% of cancer diagnoses. The incidence rates of lung cancer have been declining in men since the 1980s; the incidence rates in women are approaching a plateau after a long period of increase. Incidence rates of prostate cancer have changed substantially over the past 20 years, rapidly increasing from 1988 to 1992, declining sharply from 1992 through 1995, remaining stable from 1995 to 2000, and decreasing since 2000. These trends largely reflect increased prostate cancer screening with prostate-specific antigen (PSA) blood testing. Death rates from breast cancer in women have decreased since the 1990s due to a combination of earlier detection and improved treatment. Colorectal cancer is the third most common cancer in both men and women. The incidence of colorectal cancer has decreased over the past two decades with screening colonoscopies, through the removal of polyps before they become cancerous, and also by detecting a higher proportion of cancers at early and more treatable stages. Given these incidence and mortality figures, it is rare that a family is spared the experience of cancer in one of its members, and many live through it repeatedly, given shared environmental or genetic factors.

COMMON PSYCHOLOGICAL RESPONSES TO CANCER

The diagnosis of cancer creates a crisis requiring the patient to adapt quickly to catastrophic news. Major concerns are fear of death, dependency, disfigurement, disability, and abandonment, as well as disruptions in relationships, role functioning, and financial status. The diagnosis of cancer induces distress, and individuals show an initial characteristic response of shock and disbelief that usually lasts less than 1 week. Patients try to control the level of emotional distress while making crucial treatment decisions. The presence of a relative or friend can help with the processing of important information. Research has shown that the way the news is conveyed by health care professionals can influence a patient's beliefs, emotions, and attitudes toward the future as well as the medical staff. There are guidelines and recommendations on how to deliver the diagnosis, treatment plan, and prognosis. The patient's response to diagnosis is often followed by a period of commotion with mixed symptoms of psychological distress, anxiety, and possibly depression, irritability, insomnia, poor concentration, and inability to function. These symptoms, if transient, begin to resolve with support from family and friends and from the physician who outlines a treatment plan that offers hope. At some point, the patient adapts to the diagnosis, its treatment, and what is often termed "a new normal." The patient may incorporate new and additional coping skills or expand upon previously used coping strategies. Patients who are diagnosed with cancer have varying levels of psychological maturation and the ability to adapt to the threat of cancer. All patients employ some form of coping strategy to integrate the new information and reality that cancer brings to their lives.

Factors in Adaptation to Cancer

Patient responses are based on their previous experiences, societal and cultural expectations, cancer type, stage, and treatment, and the meaning the patient gives to this diagnostic event. Most patients seek to understand what caused the cancer, how they may have contributed to its development through genetics or behavior, and if there's anything they can do to correct the situation. For many patients, the role of society and stigma has been lessened in contemporary science-driven culture, yet the residue of moral culpability remains, which may lead to feelings of personal failings, hopelessness, or demoralization. Patients are immediately struck by a potential change in their family structures and adapted familial and professional roles. The patient’s sense of self is challenged as the ‘new normal’ is integrated.

Thankfully, public knowledge and awareness of cancer as a variable threat to life and liberty as opposed to an immediate “death sentence” has helped ease this transition from a societal standpoint, but this outlook and worldview are not necessarily applicable or useful for all patients and families. Multicultural issues are paramount for patients who are attempting to integrate this information. The threat of cancer is perhaps more nuanced, and society has begun to see it that way as well. This has created less of a threat from alienation and stigma as many patient advocacy groups provide patient education, support, and community. As a starting point, the legal requirement of informed consent has improved communication between the physician and the patient about illness, treatment options, prognosis, and participation as part of the treatment. However, this may be experienced as an additional burden to some patients who may not be ready or able to cope with the severity of their illness. Even following curative-intent cancer treatment, patients are constantly monitored for recurrence and the development of new cancers, which may continue to precipitate fear and anxiety, even if just periodically (e.g., fear of cancer recurrence or progression).

The patient-derived factors that modulate adaptation to cancer originate from three sources: intrapersonal factors, interpersonal factors, and socioeconomic and social class. The *intrapersonal factors* include pre-existing character style, coping ability, ego strength, developmental stage of life, and the impact and meaning of the cancer at that stage. *Interpersonal factors* refer to the social support derived from others. *Socioeconomic status* influences the third set of patient-derived factors. Lower socioeconomic status can be a barrier to health literacy and access to health care due to lack of education and fewer resources. The cancer-derived factors that affect adaptation to cancer are related to characteristics of the disease itself, such as stage of the disease, symptoms, site, prognosis, type of treatment, and the impact in functionality.

Of note, historical literature evaluated the role of stress, personality, and psychiatric disorders in the development or initiation of cancer. While some large epidemiologic studies suggest a small but potential association, well-designed studies do not confirm that depression, personality, stress, or major life events cause cancer despite the popularity of this concept. Personality factors, traumas, and other life stressors should be distinguished from behaviors that do underlie the development of many cancers (e.g., tobacco consumption, obesity).

EPIDEMIOLOGY OF PSYCHIATRIC DISORDERS IN CANCER PATIENTS

Psychiatric disorders are either present premorbidly before cancer or develop concomitantly with cancer. Although most psychiatric disorders develop following a diagnosis of cancer, attention to patients with cancer and serious mental illness is increasing. Most psychiatric disturbances in patients with cancer relate to the cancer itself or treatment side effects. The prevalence is highest among patients with advanced cancer and poor prognosis. The prevalence of any psychiatric disorder in patients with cancer is approximately 50%. More than two-thirds of those diagnoses are adjustment disorders, followed by major depression (10% to 15%) and delirium (10%). Inpatient studies show a higher incidence of both depression (20% to 45%) and delirium (up to 75% with advanced disease). Studies of psychiatric consultation data reveal that treatable syndromes, such as major depression and delirium, continue to be underdiagnosed and undertreated, despite their known prevalence in patients with cancer.

Table 27.8–1 lists the most common symptoms found in a group of cancer patients with advanced cancer. Five of the top seven symptoms are in the psychological realm and are found in more than 50% of patients. Three of the next 10 symptoms are present in 12% to 45% of patients. Other studies have given comparable results. Recognizing the magnitude of the symptom burden of cancer patients has led researchers to an increased interest in identifying patients by their symptoms (e.g., symptom clusters) rather than by their psychiatric diagnosis.

In order to capture the prevalence of psychiatric disorders in patients with cancer, screening instruments and measures are employed. In place of screening for anxiety and depression directly, psychological distress screening is mandated as part of the oncology standard of care by numerous professional societies, including the NCCN. Psychological distress is seen as less stigmatizing and provides for a multitude of reasons for distress beyond psychiatric symptoms. The NCCN developed the Distress Thermometer and Problem List for this purpose. It is a one-item measure of distress attached to several problem domains. It has been translated into numerous languages and is used globally. In addition, many studies have used a wide range of validated instruments, and more focused subpopulations have documented the high prevalence of moderate to severe distress in different subtypes of cancer. Breast cancer remains the cancer type with the most available data regarding psychological distress. The impact on caregivers is receiving increasing attention as caregiver distress is compounded by patient distress.

Table 27.8–1.
Symptoms in Patients with Cancer

Symptom	% Positive
Fatigue	74
Worry	71
Sadness	66
Pain	63
Drowsiness	61
Dry mouth	56
Insomnia	54
Poor appetite	45
Nausea	44
Bloating	39
Difficulty in concentration	38
Change in taste	36
Constipation	33
Cough	30
Sexual dysfunction	24
Incontinence	12
Nightmares	11

GENERAL PRINCIPLES OF ASSESSMENT AND TREATMENT IN PSYCHO-ONCOLOGY

The hallmark of treatment in psycho-oncology is therapeutic activism. In addition to its focus on symptom management, psycho-oncology seeks to reintegrate personhood in the face of cancer. A growing understanding of how cancer and its treatments may be traumatic events informs this integrated, present-focused approach.

The physician must also address the caregiver and the family, as well as the caregiver and/or family–staff interface. The psycho-oncologist should remain available and present through all stages of the illness. Psychiatrists must be prepared for the existential plight of the cancer patient, which intensifies the expectable countertransference reactions. They must resist the urge to distance themselves from a painful situation while guarding against overinvolvement. They must be aware of the patient’s physical condition and yet must not allow acute physical or emotional distress, feelings of despair, and barriers to communication (e.g., tracheostomies) to prevent a thorough evaluation. This requires a clear assessment of the clinician’s personal feelings that should be addressed prior to extensively visiting with a patient. Feelings of fear, guilt and existential plight are quite common but are also barriers to engaging with the patient in an attempt to understand their psychological ailments.

Areas of Assessment

The standard psychiatric history should also include an exploration of the patient’s understanding of the illness and its meaning. What is the patient’s understanding of their disease? Can the patient review what they’ve been told? Are there discrepancies in understanding between patients, family members, and staff, which may lead to unnecessary stress and disruption? The next area is the impact of medical factors on mental state, and the last area is the role of the environment, especially the patient’s family and their interface with the health care system. The recurrent demands of cancer treatments promote intense relationships between the patient and family and their medical caregivers, which, if dysfunctional, can lead to psychiatric symptoms and impairment of care.

Treatment Modalities

Psychopharmacology. Because there is frequently a physiologic cause or component to psychological symptoms, the psycho-oncologist must encourage an etiologic work-up and must recommend medical treatments, such as correction of electrolyte abnormalities and removal of offending drugs. However, symptomatic relief must be provided without waiting for the results. Relief from pain and insomnia goes a long way toward relieving psychological distress. Insomnia is recognized but is often mistreated with hypnotics or antihistamines, which cause habituation or daytime sedation and may worsen symptoms of depression and delirium. The depressed patient may benefit from a low-dose sedating antidepressant, such as mirtazapine (Remeron) or trazodone (Desyrel); or the mildly confused or anxious patient may benefit from a low dose of antipsychotic, such as olanzapine (Zyprexa), risperidone (Risperdal), or quetiapine (Seroquel) at bedtime. Short- to medium-acting benzodiazepines are good hypnotics for intermittent use in the absence of complicating factors, such as delirium.

Fatigue is the most common symptom in patients with cancer. It can be caused by a psychiatric disorder, the cancer itself, more often, the cancer treatments, or, most often, a combination of the three. Cancer-related fatigue is extensively reviewed later in this section.

Given the interrelatedness of physical and psychological symptoms, it is more important than ever that psychiatrists be comfortable discussing the treatment of symptoms such as nausea, vomiting, diarrhea, dry mouth, and shortness of breath.

Medication management is discussed in detail under specific psychiatric conditions later in this section.

Psychotherapeutic Interventions. Telepsychiatry is being increasingly used as an effective form of intervention, particularly for isolated homebound patients, patients in isolation, or children who need to maintain a sense of connection to a hospitalized parent. Video instruction as a form of psychoeducation is used frequently but is more effective when used in combination with face-to-face counseling rather than as standalone tools.

Opportunities for patient education, counseling, group interventions, and informal patient-to-patient connections are facilitated over the Internet. However, the dissemination of false or harmful information over the Internet is an increasingly important concern. Unfortunately, many people and organizations post biased and unreliable information to which patients are susceptible. These days, patients are generally well informed about their cancers, but the information gathered on the Internet may need to be changed or modified. It is increasingly common for patients to come into the physician's office having done research on their disease. Exploring the patient's Internet experiences can be a productive interaction, but the psychiatrist should maintain an open mind without relinquishing their professional stance.

Reviews of the many available psychotherapeutic modalities, such as cognitive-behavioral therapy, crisis intervention, problem-solving techniques, group psychotherapy, and supportive psychotherapy with an educational and a psychodynamic focus, describe a beneficial effect for all of them. Structured interventions are efficient, showing benefit after a brief course of treatment. Some studies have demonstrated persistent effects after 1 year. Working with individual patients or groups, the psychotherapies lower distress, increase self-esteem and self-image, decrease anxiety, and give patients a better sense of control. Dynamically oriented therapy still has a place in individuals whose personal issues overshadow the cancer, a situation that is more frequent than one might imagine. However, it should remain flexible to reflect the patient's shifting condition.

Referrals often begin with a crisis intervention but rapidly move into other modes. Many cancer patients are happy to leave psychotherapy when they feel better and return only episodically, when in relapse or crisis. This is not treatment resistance; it is inherent in the original treatment contract. Two major issues for patients are the reality of living a shortened life span and the coming of death. Despite their close connection, these are often dealt with separately, the first as a negotiable issue on a continuum with other life problems, the second as the mysterious and frightening absolute that it is. Psychiatrists must overcome their own reflexes to stay with the first issue and to avoid the second. Later in their course, many patients come to an ambivalent acceptance of and even a yearning for death. It is important to recognize both sides of the ambivalence and to help families and staff tolerate and accept it as well. Families may be feeling ambivalent at this time and are also in need of support. If the patient dies, one is in a unique position to offer bereavement counseling.

Families should be routinely involved in the treatment of ill or physically dependent patients; hence, family-focused therapy and family groups are now more common and are under intensive study. Caregiver groups have developed in recognition of their heavy burden. Cancer survivors are being studied and treated in the family context, understanding that the entire family is permanently altered and that the welfare of several individuals is at stake.

Patient groups provide education, problem-solving exercises, mutual support, ventilation, validation of feelings, reassurance and help about side-effect management, and help in negotiating the doctor–patient relationship and the interface with the medical system. These effects are so positive that groups have multiplied. When a patient describes a negative response to a support group, the reaction should be respected but explored because it usually reveals important psychological concerns.

Groups can be led by mental health professionals who facilitate psychological work and oncology staff members who represent an authoritative source of information. Patients appreciate both kinds of help, and combination groups with a mix of structured and unstructured meetings satisfy the greatest number of patients. Self-help groups increase patients’ sense of empowerment and can be helpful, but, over time, they may be overtaken by individual members and make some patients feel excluded or minimized. Putting in democratically defined structures and guidelines can increase their longevity. Fifteen years after its publication, efforts to duplicate a group intervention that prolonged survival in breast cancer patients have been unsuccessful. However, even if the effect on survival is in doubt, the improved quality of life, coping skills, self-esteem, pain management, and interpersonal relations were never questioned and remain as important as ever.

Meaning-Centered Psychotherapy (MCP) is a novel modality that is designed to help patients with advanced cancer to sustain or enhance a sense of meaning, peace, and purpose in their lives even as they approach the end of life. Inspired primarily by the works of Viktor Frankl and further informed by the contributions of Irvin Yalom, Breitbart and colleagues adapted Frankl's concepts involving the importance of meaning in human existence and created MCP. While MCP relies heavily on Frankl's concepts of meaning, it also incorporates other fundamental existential concepts related to the search, connection, and creation of meaning. Using didactics and experiential exercises in each session, the therapist and the patient(s) work together to help patients understand the importance and relevance of sustaining, reconnecting with, and creating meaning in their lives through common and reliable sources of meaning. Furthermore, patients are taught that these various sources of meaning may serve as *resources* during especially difficult times to help diminish despair near the end of life. Both individual and group formats of MCP have been developed and tested in randomized controlled studies. These studies provide strong support for the efficacy of MCP as a treatment for psychological and existential/spiritual distress among patients with advanced cancer.

Educational and support programs can be built into the medical routine, as in pre-surgery education groups, post-surgery rehabilitation groups, and even waiting room groups. The growing number of survivors has prompted the development of innovative posttreatment group programs. Behavioral techniques, such as hypnosis, self-hypnosis, active and passive relaxation, desensitization, and distraction, are useful in a wide range of situations in oncology, including relief of acute and chronic anxiety related to surgery and painful procedures; relief or prevention of conditioned symptoms, such as anticipatory nausea and vomiting (ANV) associated with chemotherapy; and, with age-appropriate modification, improved management of almost all pediatric problems. Their spectrum of use can be as varied as the expertise and initiative of available practitioners.

Social Services. Social workers who are available in large centers or community institutions are the first line of psychological support. They may provide extensive psychotherapeutic support, given the training and the opportunity, but, at the very least, they perform a triage function, alerting staff to patients with serious psychopathology. Their classic role is to help patients and families find their way to available resources, which is often the most helpful intervention that patients identify. If social workers are not available, or it is not part of their mandate, psychiatrists must do their best to fulfill this function by at least having a good file of local resources and a readiness to refer patients to them.

Chaplaincy Services and the Interface of Religion and Psychiatry. As many as 80% of cancer patients describe themselves as believing in God, and a large proportion of those state that faith is an important, if not primary, contributor to their coping ability. There is a large and growing literature on the role of spiritual belief in health behavior. Most of the studies are small and methodologically weak, and there may be a publication bias against negative results. Many more studies have found an association between religious beliefs and psychological variables. Several studies have shown that religious beliefs are associated with mature and active coping methods. A few studies have focused on negative effects, such as an irreducible sense of betrayal, a sense of being punished for personal failings, or strong feelings of guilt about anger and loss of faith.

There is some controversy in the pastoral community about the usefulness, or even the propriety, of doing research on religious phenomena and the appropriateness of doctors intruding on religious practices, especially in prescriptive ways. However, it is important for psychiatrists to become knowledgeable about the interface between spiritual beliefs and psychological experiences without a negative bias toward the former, especially with ill patients in whom existential concerns are universally present.

The psychiatric consultant should inquire about faith, its meaning, associated religious practices, and its impact on the coping response. It can be a source of strength or guilt at all stages of the disease, ranging from the earliest “What did I do to cause this?” to “Will God give me only what I can carry?” in midcourse to the poignant life review of the late stage. It is often a primary modifier of attitudes toward suicide and terminal care decisions. Mental health professionals should be prepared to deal with these areas in an unself-conscious, uncondescending manner, enabling the patient to fully integrate this aspect of their personality into the therapeutic relationship and the current crisis. However, this should never be mistaken for spiritual counseling. Some centers are fortunate to have chaplains available to patients on request. Psychiatrists should work with them, should let the patient know about their availability, and should arrange a visit if the patient shows interest. They should encourage the use of other spiritual guides in the community as well. Where specialized chaplains are not available, the psychiatrist should become familiar with selected religious leaders from the major faiths that appear in their practice to obtain advice and to make referrals.

The Wellness Movement. Wellness programs are a growing group of programs that demonstrate multiple and original uses of group modalities to help cancer patients and their families. The Wellness Community (TWC) and Cancer Support Community are good prototypes in addition to many locally based programs throughout the country, some of which provide additional personal services, such as transportation or accompaniment to doctors’ appointments. These organizations are generally not-for-profit organizations, are free of charge, and are devoted to helping cancer patients and their families cope and grow through the illness. They create an upbeat, welcoming environment for the whole family and position themselves as complementing standard medical treatment. They encourage patients to take a more active role in their care and coach them to interact more effectively with their medical caregivers. Although volunteers play a large role, and patients themselves are welcome to help, the specialized offerings are all run by trained professionals.

Volunteerism. The programs described previously, created and maintained by dedicated local and national volunteer workers and philanthropists, are only a small part of the support provided to cancer patients through volunteer activities. The ACS, the Leukemia Lymphoma Society of America (LLSA), and numerous institutions and family foundations have developed patient services along with their major research funding. "I Can Cope" is one of the largest. Cancer Care runs numerous groups and has established telephone and Internet groups. One organization fulfills the wishes of extremely ill children (e.g., a trip to Disneyland or a visit from a famous sports hero). Another provides free transportation to medical centers in executive jets. Most of them originated with grateful survivors or bereaved families trying to give meaning to their loss. This does not include the hundreds of thousands of individuals who volunteer their time in hospitals and clinics and are willing to do whatever is asked to make the patient's experience more tolerable. For the volunteers who provide the help, it is a source of increased self-esteem and life purpose, whether they have a personal cancer experience or not. The magnitude and variety of offerings are encouraging for medical and familial caregivers and a reminder that selfless initiatives are still an important aspect of society.

DIAGNOSIS AND MANAGEMENT OF PSYCHIATRIC DISORDERS IN CANCER PATIENTS

Delirium, Dementia, and Other Cognitive Disorders

Cognitive disorders due to chemotherapy, radiation, cancer, and paraneoplastic syndromes will be reviewed separately later in this section. Dementia will be reviewed under the geriatric psycho-oncology section.

Delirium. Delirium is the most common and often serious neuropsychiatric complication in the management of cancer patients characterized by an abrupt onset of disturbances in the level of consciousness, awareness, attention, cognition, and perception that tend to fluctuate over the course of the day, precipitated by an underlying medical condition. Delirium is a medical emergency that needs to be identified and treated vigorously. Unfortunately, delirium is frequently underdiagnosed and untreated in the medical setting, which leads to increased morbidity and mortality, interference in the management of symptoms such as pain, an increased length of hospitalization, increased health care costs, as well as distress for patients and their caregivers. Delirium is associated with increased morbidity in the terminally ill, causing distress in patients, family members, and staff. Fifty-four percent of patients were found to have recalled their delirium experience after recovery from delirium in a study of hospitalized cancer patients. Factors predicting delirium recall included the degree of short-term memory impairment, delirium severity, and the presence of perceptual disturbances. The most significant factor predicting distress for patients was the presence of delusions. Patients with hypoactive delirium were just as distressed as patients with hyperactive delirium.

EPIDEMIOLOGY. In cancer patients, delirium is present in 25% to 85% depending on the stage of illness. Prevalence rates of delirium range from 15% to 30% in hospitalized cancer patients; they are higher in the last weeks of life, at 40% to 85%. The reported prevalence of delirium in hospice may approach as high as 85%. Older adult patients who develop delirium during a hospital stay have an estimated mortality rate of 22% to 76% during that hospitalization.

ETIOLOGY. Frequently, a combination of factors (three or more etiologies) contributes to the development of delirium. Delirium is irreversible 30% to 40% of the time, especially in the terminally ill. If the etiology of the delirium is unidentifiable or irreversible, relief from the distressing symptoms of delirium may be provided to the patient. Predisposing factors including age, functional impairment, and the nature and severity of illness increase the risk of experiencing delirium during hospitalization. Dementia is an important independent risk factor for developing delirium. Physical impairment, sensory impairment, low BMI, cachexia, and malnutrition have been associated with increased risk for delirium. With an increase from zero to three risk factors, the incidence of delirium rises from 2% to 50%, bringing longer hospital stays and poorer outcomes. In patients with advanced cancer, delirium can be due either to the direct effects of cancer on the CNS or to indirect CNS effects of the disease or treatments (e.g., medications, electrolyte imbalance, dehydration, major organ failure, infection, vascular complications, paraneoplastic syndromes). Given the large numbers of drugs, cancer patients require and the fragile state of their physiologic functioning, even routinely ordered hypnotics are enough to tip patients over into a delirium. Narcotic analgesics such as levorphanol (Levo-Dromoran), morphine sulfate (Roxanol), and meperidine (Demerol) are common causes of confusional states, particularly in the older adults and terminally ill. Chemotherapeutic agents are known to cause delirium including methotrexate (Rheumatrex), ifosfamide (Ifex), fluorouracil (Adrucil), vincristine (Oncovin), vinblastine (Velban), bleomycin (Blenoxane), BCNU (carmustine, Gliadel), cisplatin (Platinol-AQ), asparaginase (Elspar), procarbazine (Procanbid), and corticosteroids. Most patients receiving chemotherapeutics will not develop prominent CNS effects unless dexamethasone (Decadron) or prednisone (Deltasone) is included in the regimen. In addition, drug withdrawal may be a common cause of delirium, especially in postoperative patients whose drugs have been interrupted for surgery and have not been promptly restarted (Table 27.8–2).

CLINICAL FEATURES AND DIAGNOSIS. Clinically, the diagnostic gold standard for delirium is the clinician's assessment utilizing the *DSM-5* criteria. Clinical features of delirium include abrupt onset, fluctuating course, attention disturbances, altered level of alertness, awareness, and arousal, increased or decreased psychomotor activity, disturbance of sleep–wake cycle, affective symptoms, altered perceptions, disorganized thinking, and incoherent speech, disorientation, and memory impairment. Neurologic abnormalities including cortical (aphasia, apraxia, dyssomnia, dysgraphia) and motor symptoms (tremor, asterixis, myoclonus, changes in muscle tone, and deep tendon reflexes) can also be present. It is this protean nature of the symptoms and the fluctuation of clinical findings that make delirium so difficult to diagnose and treat.

Delirium is classified into three clinical subtypes, based on psychomotor behavior, including the hyperactive, the hypoactive, and the mixed subtype. Approximately two-thirds are either of the hypoactive or mixed subtypes. The hyperactive form is most often characterized by hallucinations, delusions, agitation, and disorientation, while the hypoactive form is characterized by confusion and sedation but is rarely accompanied by hallucinations or delusions.

Table 27.8–2.
Causes of Delirium in Cancer Patients

Direct
- Primary brain tumor
- Metastatic spread to the brain

Indirect
- Hypoxia
- Metabolic encephalopathy due to organ failure
- Electrolyte imbalance
- Alcohol and drug withdrawal states
- Treatment side effects from
  - Chemotherapeutic agents, steroids, biologic response modifiers
  - Radiation
  - Opioids
  - Anticholinergics
  - Antiemetics
  - Benzodiazepines
- Infection
- Hematologic abnormalities
- Nutritional deficiencies
- Vitamin deficiencies (thiamine, folate, vitamin B_{12}, and niacin)
- Paraneoplastic syndromes
- Endocrine disorders
- Tumors of the pituitary, thyroid, parathyroid, and adrenal
- Multiple endocrine neoplasia syndromes
- Carcinoid tumors, mainly of the ovary and lung

Many of the clinical features and symptoms of delirium can also be associated with other psychiatric disorders such as depression, mania, psychosis, and dementia. For instance, delirious patients, not uncommonly, exhibit mood disturbances such as anxiety, fear, depression, irritability, anger, euphoria, apathy, or lability. Delirium, particularly the hypoactive subtype, is often initially misdiagnosed as depression. In distinguishing delirium from depression, particularly in the context of advanced cancer, an evaluation of the onset and temporal sequencing of depressive and cognitive symptoms is particularly helpful.

Many scales or instruments have been developed that can aid the clinician in rapidly screening for cognitive disorders or in establishing a diagnosis of delirium, including the Delirium Rating Scale-Revised 98, the Confusion Assessment Method (CAM), the Abbreviated Cognitive Test for Delirium, and the Memorial Delirium Assessment Scale (MDAS).

The MDAS is a 10-item delirium assessment tool validated among hospitalized patients with advanced cancer and acquired immunodeficiency syndrome (AIDS). The MDAS is a delirium diagnostic screening tool as well as a reliable tool for assessing delirium severity among patients with advanced disease. A cutoff score of 13 is diagnostic of delirium. Recently, MDAS was further examined for clinical utility and validation in a population of advanced cancer patients in a palliative care unit. These investigators found the MDAS to be useful in this population and found that a cutoff score of 7 of 30 yielded the highest sensitivity (98%) and specificity (76%) for a delirium diagnosis in this palliative care population.

ASSESSMENT AND MANAGEMENT OF DELIRIUM. The standard approach to managing delirium in the medically ill, and even in those with advanced disease, includes a search for underlying causes, correction of those factors, and management of the symptoms of delirium. The desired and often achievable outcome is a patient who is awake, alert, calm, cognitively intact, not psychotic, and communicating coherently. Studies in patients with earlier stages of cancer have demonstrated the potential utility of a thorough diagnostic assessment. When such diagnostic information is available, specific therapy may be able to reverse delirium. The diagnostic evaluation includes an assessment of potentially reversible causes of delirium. A full physical examination should assess for evidence of sepsis, dehydration, or major organ failure. Medications that could contribute to delirium should be reviewed. A screen of laboratory parameters will allow assessment of the possible role of metabolic abnormalities. Imaging studies of the brain and assessment of the cerebrospinal fluid may be appropriate in some instances.

In the terminally ill patient who develops delirium in the last days of life (referred to as terminal delirium), the management of delirium is unique, presenting a number of dilemmas, and the desired clinical outcome may be significantly altered by the dying process. When confronted with delirium in the terminally ill or dying patient, a differential diagnosis should always be formulated as to the likely etiology or etiologies. There is an ongoing debate as to the appropriate extent of diagnostic evaluation that should be pursued in a dying patient with terminal delirium. Most palliative care clinicians would undertake diagnostic studies only when a clinically suspected etiology can be identified easily, with minimal use of invasive procedures, and treated effectively with simple interventions that carry minimal burden or risk of causing further distress. Diagnostic work-up in pursuit of an etiology for delirium may be limited by either practical constraints such as the setting (home, hospice) or the focus on patient comfort, so that unpleasant or painful diagnostics may be avoided. Most often, however, the etiology of terminal delirium is multifactorial or may not be determined. Etiology is discovered in less than 50% of terminally ill patients with delirium. When a distinct cause is found for delirium in the terminally ill, it is often irreversible or difficult to treat. One study found that 68% of delirious cancer patients could be improved, despite 30-day mortality of 31%. Another found that one-third of the episodes of cognitive failure improved following an evaluation that yielded a cause for these episodes in 43%. In a recent prospective study of delirium in patients on a palliative care unit, investigators reported that the etiology of delirium was multifactorial in the great majority of cases. Even though delirium occurred in 88% of dying patients in the last week of life, delirium was reversible in approximately 50% of episodes. Causes of delirium that were most associated with reversibility included dehydration and medications. Hypoxic and metabolic encephalopathies were less likely to be reversed in terminal delirium.

TREATMENT. The standard approach to the management of delirium includes the identification and elimination of factors contributing to delirium, in addition to nonpharmacologic and pharmacologic treatment interventions. Nonpharmacologic approaches include creating a calm, comfortable, well-lit environment, using orienting objects such as calendars and clocks, early mobilization, having family members around, limiting room and staff changes, and allowing patients to have uninterrupted periods of rest at nights to improve their sleep–wake cycle. One-to-one nursing observation may also be necessary and useful. Fluid and electrolyte balance, nutrition, and vitamins may be useful.

Supportive techniques alone are often not effective in controlling the symptoms of delirium, and pharmacologic treatment is necessary in most cases. Antipsychotics constitute the primary pharmacologic intervention. Antipsychotics such as haloperidol (Haldol) are frequently utilized as antiemetics in the medical setting; however, only 0.5% to 2% of hospitalized cancer patients receive haloperidol for the management of the symptoms of delirium. In terminally ill populations, only 17% receive an antipsychotic for agitation or psychological distress, despite an estimated prevalence of delirium ranging from 25% in the hospitalized cancer patient to 85% in the terminally ill.

Haloperidol (a “typical” antipsychotic) is the gold-standard medication for the treatment of delirium among cancer patients due to its efficacy and safety (e.g., few anticholinergic effects, lack of active metabolites, and availability in different routes of administration). Haloperidol in low doses (1 to 3 mg per day) is usually effective in targeting agitation and psychotic symptoms. In general, doses of haloperidol need not exceed 20 mg in a 24-hour period; however, some clinicians advocate higher doses in selected cases. In severe agitation related to delirium, clinicians may add lorazepam to haloperidol. This combination may be more effective in rapidly sedating patients and may help minimize any extrapyramidal adverse effects of haloperidol. However, there is evidence that benzodiazepines can worsen delirium, and switching to a more sedating antipsychotic is another option. The FDA has issued a warning against the risk of QTc prolongation and *torsades de pointes* with IV haloperidol; thus, monitoring QTc intervals daily among medically ill patients receiving IV haloperidol has become the standard clinical practice.

Oral or intravenous (IV) chlorpromazine is considered to be an effective alternative to haloperidol (with or without lorazepam) when increased sedation is required, especially in the ICU setting where close blood pressure monitoring is feasible, and for severe agitation in terminally ill patients to decrease distress for the patient, family, and staff. It is important to monitor chlorpromazine's anticholinergic and hypotensive side effects, particularly in older adult patients.

Atypical antipsychotics, such as olanzapine (Zyprexa), risperidone (Risperdal), quetiapine (Seroquel), and aripiprazole (Abilify), are increasingly used in the treatment of symptoms of delirium. Several studies have been conducted with atypical antipsychotics, including olanzapine, risperidone, quetiapine, aripiprazole, and ziprasidone, in the treatment of delirium.

A Cochrane review comparing the efficacy and the incidence of adverse effects between haloperidol and atypical antipsychotics concluded that, like haloperidol, selected newer atypical antipsychotics (risperidone, olanzapine) were effective in managing symptoms of delirium. Haloperidol doses greater than 4.5 mg per day were more likely to result in increased rates of EPS compared with the atypical antipsychotics, but low-dose haloperidol (i.e., less than 3.5 mg per day) did not result in a greater frequency of extrapyramidal adverse effects.

APA practice guidelines provide directions for the use of antipsychotics in the treatment of delirium, and growing evidence supports their use. The APA guidelines for the treatment of delirium recommend the use of low-dose haloperidol (i.e., 1 to 2 mg PO every 4 hours as needed, or 0.25 to 0.5 mg PO every 4 hours for the older adults) as the treatment of choice in cases where medications are necessary.

In the light of existing literature, risperidone may be used in the treatment of delirium, starting at doses ranging from 0.25 to 1 mg and titrated up as necessary, with particular attention to the risks of EPS, orthostatic hypotension, and sedation at higher doses. Olanzapine can be started between 2.5 and 5 mg nightly and titrated up, with sedation being the major limiting factor, which may be favorable in the treatment of hyperactive delirium. The current literature on the use of quetiapine suggests a starting dose of 25 to 50 mg and a titration up to 100 to 200 mg per day (usually at twice-daily divided doses). Sedation and orthostatic hypotension are the main dose-limiting factors.

Important considerations in starting treatment with any antipsychotic for delirium may include EPS risk, sedation, anticholinergic side effects, cardiac arrhythmias, and possible drug–drug interactions. The FDA has issued a “black box” warning of the increased risk of death associated with the use of typical and atypical antipsychotics in older adult patients with dementia-related psychoses. Clinicians should attempt to use low doses only for short periods of time, especially when treating older adult patients with delirium. However, leaving delirium untreated may impose a greater risk of morbidity and mortality (Table 27.8–3).

Table 27.8–3.
Antipsychotics in the Treatment of Delirium

Generic Name	Approximate Daily Dose Range (mg)	Route[a]
Haloperidol	0.5–2 mg every 2–12 hrs	po, iv, sc, im
Chlorpromazine	12.5–50 mg every 4–12 hrs	po, iv, im
Olanzapine	2.5–10 mg every 12–24 hrs	po, ODT[b]
Risperidone	0.25–2 mg every 12–24 hrs	po, ODT[b]
Quetiapine	12.5–200 mg every 12–24 hrs	po
Ziprasidone	10–40 mg every 12–24 hrs	po
Aripiprazole	5–30 mg every 24 hrs	po, ODT[b]

po, orally; iv, intravenous; sc, subcutaneous; im, intramuscular.

[a] Risperidone and aripiprazole are available in liquid formulations.

[b] Orally disintegrating tablet (ODT) forms are available for olanzapine, risperidone, and aripiprazole.

Several researchers have studied both pharmacologic and nonpharmacologic interventions in the prevention of delirium among older patient populations, particularly in surgical settings, with mixed results. The applicability of these interventions to the prevention of delirium in cancer patients is limited.

Anxiety Disorders

Anxiety is the most common response in the setting of cancer. It is a natural human response mechanism when confronted by a threat, but it can become maladaptive when the response is excessive. For example, during treatment, a patient sometimes will learn that their treatment has failed or that the disease has recurred. These situations are often met with disbelief and denial followed by emotional upheaval, characterized by mixed symptoms of depression and anxiety, together with loss of appetite, insomnia, and lack of concentration. These symptoms typically decrease gradually and resolve over a 7- to 10-day period. Anxiety may represent an acute reaction to cancer; however, some patients have premorbid psychological issues that are exacerbated by the diagnosis of cancer. This premorbid psychological functioning is influenced by the developmental state, personality, and prior psychiatric history.

Symptoms and Classification of Anxiety Disorders. Anxiety is manifested by a broad array of physical signs and measures of autonomic activation, changes in thinking (i.e., intrusive thoughts), and behaviors. The most common anxiety disorders in cancer patients are standardized in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5).

Generalized anxiety disorder is characterized by excessive anxiety across different settings. The worrying is difficult to control, evoking fatigue, irritability, muscle tension, and sleep dysfunction.

Panic disorder entails recurrent unexpected panic attacks, marked by intense fear or discomfort associated with palpitations, diaphoresis, dyspnea, sense of choking, nausea, dizziness, or derealization. The attacks may occur from a calm or anxious state and are followed by worry, concern, and behavior changes related to the attack.

Specific phobias are characterized by persistent and excessive fear elicited by the presence or anticipation of a specific object or situation. Phobias of blood, needles, hospitals, magnetic resonance imaging machines and radiation simulators may complicate treatment adherence. Patients undergoing chemotherapy have nausea and vomiting, and about 25% to 30% develop ANV despite adequate antiemetic regimens. Patients experience nausea when they are exposed to stimuli that remind them about drug infusions.

The classification “anxiety disorder due to a general medical condition” is meant to describe cases in which the symptoms are a direct physiologic consequence of the illness. Anxiety is ubiquitous and can be found with many of the conditions that cause other psychiatric disorders. Anxiety in cancer can be caused by different medical conditions, including pulmonary embolism, hypoxia, hypoglycemia, and cardiac disorders. Hormone-secreting neoplasms and paraneoplastic syndromes can also cause anxiety (Table 27.8–4).

Anxiety due to substance withdrawal is recognized and treated by most clinicians. Nicotine withdrawal is less commonly recognized. Many patients stop smoking abruptly on receiving their diagnosis and may experience severe anxiety symptoms, sometimes associated with dissociative symptoms and bizarre behaviors. Behavioral smoking cessation programs are an indispensable adjunct and are commonly offered to family members as well, taking advantage of what is known as a window of opportunity when defenses are shaken and motivation is high.

Substance-induced anxiety is a common cause of anxiety in cancer patients. They include bronchodilators, frequently used in lung cancer patients in whom they compound the anxiety caused by air hunger, corticosteroids, the mainstay of many protocols; anti-inflammatory cytokines, which are being increasingly used for treatment, and antipsychotics, including some of the ones that are used as antiemetics, such as metoclopramide (Reglan). Akathisia is often mistaken for anxiety by patients and staff; thyroxine, psychostimulants, sympathomimetic agents, serotonergic agents, and antihistamines may produce symptoms of anxiety.

Table 27.8–4.
Medical Conditions Associated with Anxiety

Metabolic	Hyperkalemia
	Hyperthermia
	Hypoglycemia
	Hyponatremia
	Vitamin deficiencies
Cancer	Central nervous system neoplasms
	Carcinoid syndrome
	Lung cancer
Endocrine	Adrenal abnormalities
	Thyroid abnormalities
	Parathyroid abnormalities
	Pituitary abnormalities
	Pheochromocytoma
Cardiovascular	Arrhythmia
	Congestive heart failure
	Myocardial infarction
	Angina pectoris
	Valvular disease
Pulmonary	Pulmonary embolism
	Asthma
	Chronic obstructive pulmonary disease
	Pneumothorax
	Pulmonary edema
Medication/toxic conditions	Corticosteroids
	Neuroleptics
	Thyroxine
	Sympathomimetic agents
	Serotonergic agents
	Antihistamines
	Withdrawal states (alcohol, opioid analgesics, sedative-hypnotics, caffeine)

	Antibiotics (cephalosporins, acyclovir, isoniazid)
Other symptoms	Pain

Treatment. Treatment aims include reducing the patient’s overall level of emotional distress as well as reducing specific target symptoms that may impair social or occupational function. Treatment of anxiety in cancer patients depends on the etiology and the timing of the onset of symptoms. In some situations, anxiety symptoms can represent disorders that started before the onset of a medical illness. In that case, the treatment approach should be directed to the premorbid condition. In other instances, anxiety can be the consequence of a medical condition or its treatment. In such cases, the treatment should target the underlying condition.

Current or past delirium can contribute to anxiety in cancer patients. Delirious patients can present with early symptoms of restlessness, irritability, altered sleep patterns, and impaired concentration. Patients who have been delirious may have persistent worry related to their memory of this experience or their behavior during episodes of delirium.

PHARMACOLOGIC TREATMENT. Medications used to treat anxiety in cancer patients are also used in primary anxiety disorders. These medications are benzodiazepines, antidepressants, and atypical antipsychotics.

NONPHARMACOLOGIC TREATMENT. Nonpharmacologic treatment involves several approaches: psychotherapy, education, and supportive measures. There are different types of psychotherapy, including but not limited to interpersonal psychotherapy, supportive psychotherapy, and cognitive-behavioral therapy. MCP should be considered in advanced cancer patients with anxiety symptoms. Several behavioral methods have been used, such as progressive muscle relaxation, breathing exercises, meditation, biofeedback, and guided imagery.

Trauma- and Stressor-Related Disorders

Trauma- and stressor-related disorders are a group of diagnoses that entail both a traumatic or stressful exposure and a subsequent change in emotions and behavior. These include adjustment disorders, acute stress disorder (ASD), and PTSD.

Adjustment disorders with depressed mood, anxious mood, and mixed emotional features represent the largest group of psychiatric diagnoses found in cancer patients. Diagnosis of cancer may invoke a normal reaction to stress where mood can be affected in cancer patients. They may use their coping mechanisms and continue with their life without it affecting their functioning. Patients with cancer are on a continuum, with acute stress responses at one end and mood and anxiety disorders at the other. Adjustment disorders represent a maladaptation in one or more of the psychosocial areas. This disorder may resolve when the stressor is over, but it may become chronic, requiring medications along with counseling. The prevalence of adjustment disorder in oncologic care settings is around 20% but often co-occurs with anxiety and mood disorders.

ASD follows exposure to a traumatic event. Having cancer is perceived as a life-threatening event. The symptoms of ASD include the combination of one or more intrusion, dissociative, avoidance, or arousal symptoms, and negative mood. Dissociative symptoms include an altered sense of one's self or surroundings or an inability to remember aspects of the event. These symptoms typically begin immediately after the trauma and persist for 3 days to 1 month after exposure; PTSD should be considered for symptoms that have persisted for more than 1 month after such an event. The prevalence of ASD has not been established in cancer patients.

The essential feature of PTSD is the development of problematic symptoms following the experience of a traumatic event that involves the threat of death or injury. The syndromal presentation includes intrusion symptoms, negative cognitions and mood, persistent avoidance of stimuli associated with the trauma, and marked alteration in arousal and reactivity associated with the trauma. PTSD differs from ASD in that the disturbance lasts for longer than 1 month. Prevalence of PTSD has been reported as high as 37% among survivors of non-Hodgkin lymphomas at a median of 12.9 years after diagnosis. It is difficult to assess classic PTSD symptoms in cancer patients. Evaluation of intrusive symptoms is nonspecific since cancer patients often report fears related to the future and rarely flashbacks or intrusive memories. Avoidance behavior is difficult to ascertain, as patients are frequently confronted with the disease and its treatment with potential trauma-related stressors, and there have been reports of patients who felt worse because of premature interventions directed toward re-experiencing the trauma so as to integrate it. Acute posttraumatic denial needs to be approached cautiously, if at all. PTSD can be acute (immediate and lasting less than 3 months), chronic (lasting more than 3 months), delayed (if starting more than 6 months after the event), and recurrent (when a recurrence of symptoms is precipitated by another event that is reminiscent of but not necessarily as severe as the initial event). This latter type is more frequent than reported and should be sought out because patients themselves do not make the connection. In a paradoxical case, a Vietnam veteran was unaware of the striking parallels between his severe depressive response to being told that he did *not* have lung cancer and his reaction on returning stateside many years earlier. He experienced both events as undeserved reprieves from death that aroused profound survivor guilt about his dead buddies and the living ones that he had left behind.

Because avoidance of people and places associated with cancer, such as a clinic, is part of PTSD, this syndrome may interfere with optimal medical care delivery. Therapies used to treat PTSD include crisis intervention to lessen symptoms and improve functioning, cognitive-behavioral therapy to change thinking and behavior, and support groups for further support and to normalize the presentation. Patients with severe symptoms may benefit from medications, such as antidepressants, anxiolytics, and neuroleptics.

Cancer-related posttraumatic stress (PTS) is a syndrome much like PTSD, though the presentation is generally less severe. Patients with a new diagnosis of cancer may experience recurrent frightening thoughts, distractibility, insomnia, and a sense of detachment or depersonalization. Helplessness is also a common symptom. Cancer-related PTS can occur anytime during or after treatment. Risk factors for cancer-related PTS can include diagnosis of cancer that recurs, a longer course of treatment, and a history of anxiety or trauma. Good social support and clear communication are thought to be protective against PTS symptoms. Treatments used for PTS are similar to those for PTSD, including early identification, crisis intervention, and relaxation training. Support groups may be helpful. For patients with severe symptoms of PTS, antidepressant and anxiolytic medicines may be used, especially in cases of comorbid mood and anxiety disorders.

Depressive Disorders

In DSM-5, depressive disorders include major depressive disorder, depressive disorder due to another medical condition, and substance/medication-induced depressive disorders. This section will principally review depressive disorders in cancer patients. Depressive symptoms are common in the oncology setting. Depression is a common psychiatric complication of cancer and is an important risk factor for suicide. Cancer patients are vulnerable to these symptoms at all stages of the illness, from the appearance of the first symptoms of cancer to the time of diagnosis, during treatment, palliative care, and even after remission or cure. Evaluating the nature and intensity of depressive symptoms is important. For the clinician, the relevant clinical question is how to identify the point when normal sadness or distress associated with cancer has become a depressive disorder. Diagnosis of depression in the context of cancer is challenging. Recognizing the risk factors for depression, with careful attention to the signs and symptoms of depression, leads to improved recognition and treatment of depressive disorders, thus increasing adherence to cancer treatment, improving quality of life, and reducing serious consequences such as a desire for hastened death and suicide.

Prevalence of Depression in Cancer Patients. Many studies have documented the prevalence of depression in cancer patients. The prevalence estimates of major depressive disorder in cancer patients have varied widely in different studies, from a low of 1% to a high of 50%. The use of different diagnostic measures and cutoff criteria impacts prevalence estimates. The site of cancer, the physical symptoms, and the stage of cancer are additional factors that contribute to differences in prevalence. In a study of terminally ill cancer patients, a symptom threshold consistent with DSM-IV-TR criteria was associated with a depression diagnosis in 13% of patients. A relatively minor reduction in the symptom severity threshold elevated the depression diagnosis to 26.1% of patients. A 2011 meta-analysis that stratified for various diagnostic systems and stages of disease reported a 25% prevalence rate for depressive or adjustment disorders and 29% for all types of mood disorders among cancer patients.

Risk Factors for Depressive Disorders in Cancer Patients. There are several risk factors for depression in cancer patients. The site of cancer and several chemotherapy regimens tend to increase the risk. Vinblastine, vincristine, interferon, procarbazine, asparaginase, tamoxifen, cyproterone, and corticosteroids are associated with greater risk. A higher prevalence of depressive disorders has been found among patients with pancreatic, head and neck, breast, and lung cancers, with relatively lower rates observed among those patients with lymphoma, colon, and gynecologic cancers. Depression is a common symptom of pancreatic cancer, with some early data suggesting that the mood disturbance is mediated by alteration of brain serotonergic function through proinflammatory cytokines (Table 27.8–5).

Depression may also develop secondary to organ failure or from nutritional, endocrine, and neurologic complications of cancer. Other risk factors for depressive disorders are advanced disease stage and physical disability, presence of other chronic medical illnesses, a previous history of depression, family history of depression, uncontrolled pain, low social support, social isolation, recent experience of a significant loss, and low self-esteem. Although it is well established that depression is more prevalent among women than men in the general population, the gender difference is not evident among cancer patients. Older individuals are at greater risk for depression even in the absence of medical illness.

Evaluation and Diagnosis of Depression in Cancer Patients. Diagnosis of depression is challenging in cancer patients due to the neurovegetative symptoms that mimic many symptoms caused by cancer or its treatment, such as loss of appetite, fatigue, sleep disturbances, psychomotor retardation, apathy, and poor concentration. The assessment of depressive symptoms in cancer patients should focus on the presence of dysphoria, anhedonia, hopelessness, worthlessness, excessive or inappropriate guilt, and suicidal ideation. The presence of these symptoms helps distinguish depression from cancer-related symptoms. Changes in appetite, sleep disturbances, fatigue, and loss of energy can still be used to diagnose the cancer patient who is physically well, although one must bear in mind that chemotherapy- and radiation therapy–induced fatigue and other vegetative disturbances continue for a long time. Poor memory and impaired concentration are more likely to be the initial chief complaints in older adult depressed patients. Although rare, delusions or hallucinations may accompany depression. In medically ill depressed patients, the presence of delusions and hallucinations might be reflective of a diagnosis of delirium, which should be ruled out first. If present, the diagnosis of delirium precludes the diagnosis of a mood disorder with depressive features.

Table 27.8–5.
Medical Causes of Mood Disorders in Cancer Patients

- Electrolyte disturbances
- Vitamin B_{12}, folate, or other vitamin deficiencies
- Cancer-related
  - Primary CNS tumors
  - CNS metastasis
  - Paraneoplastic syndromes
  - Pancreatic cancer, small cell lung cancer, breast cancer, carcinoid tumor, adrenocortical tumors
  - Neurologic diseases
  - Cerebrovascular disease
  - Dementia
  - Cerebral trauma
  - CNS infections
- Systemic disorders
  - Autoimmune disorders
  - Inflammatory disorders
  - Infections
- Endocrine abnormalities
  - Hyper- or hypothyroidism
  - Adrenal insufficiency
- Medications
  - Corticosteroids
  - Interferon and interleukin-2
  - Cardiac and antiarrhythmic drugs
  - Antibiotics, antivirals
  - Psychotropic drugs (antipsychotics, benzodiazepines, other sedative/hypnotics)
  - Opiates
  - Cimetidine
  - Levodopa, methyldopa
  - Pentazocine
  - Tamoxifen
  - Analgesics, and anti-inflammatory drugs
  - Some chemotherapeutic agents: vincristine, vinblastine, procarbazine, asparaginase, tamoxifen, cyproterone,

- mithramycin, L-asparaginase
- Other conditions
  - Cardiorespiratory disease
  - Renal disease, anemia, uncontrolled pain

CNS, central nervous system.

Four conceptual approaches have been proposed to evaluate depression in the medically ill: inclusive, exclusive, etiologic, and substitutive approaches. No single approach is inherently superior to the others. When selecting a conceptual approach, it is important to consider whether a diagnosis of depression is being made for clinical or research purposes. The inclusive approach, in which all symptoms of depression are counted regardless of their suspected etiology, is highly sensitive, making its use more appropriate for diagnoses in clinical settings. The substitutive approach replaces somatic symptoms of depression (i.e., fatigue, sleep disturbance, change in weight or appetite, difficulty concentrating) with psychological symptoms such as depressed appearance, social withdrawal, brooding, self-pity, pessimism, and anhedonia. This approach is conceptually reasonable, but there is little evidence for its superiority. The etiologic approach requires that the clinician include a symptom as part of a clinical depression only if it is clearly not a result of medical illness. It is theoretically sound; however, it is difficult to apply this reliably in oncology settings due to the overlap of symptoms. The exclusive approach eliminates the somatic symptoms from the diagnostic criteria. The high specificity associated with the exclusive approach makes it valuable for research purposes, but it is not desirable in clinical practice due to the risk of missing a diagnosis of depression.

Several subcategories of DSM-5 depressive disorders diagnoses are found in the context of cancer. A diagnosis of depressive disorder due to another medical condition is the appropriate diagnosis when the depressive disorder is due to underlying cancer, such as pancreatic cancer. When a medication (such as interferon) is the underlying cause of a depressive disorder, the diagnosis of substance/medication-induced depressive disorder is used. Corticosteroids are used in a wide range of situations and can result in depression to the extent that patients may become suicidal or psychotic. Establishing the etiology of symptoms of depression in cancer patients is difficult, and usually, multiple factors contribute to depressive symptoms in most of these patients. Self-report symptoms inventories, such as the Patient Health Questionnaire-9 (PHQ-9) or the Beck Depression Inventory (BDI), are useful for screening purposes but do not discriminate between a major depressive disorder and depressive disorder due to another medical condition. Major depressive disorder is defined as an episode of clinically significant persistent and pervasive depressed mood or anhedonia, accompanied by cognitive and behavioral symptoms. The DSM-5 criteria for major depression are consistent with the etiologic approach in which symptoms that are deemed to be due to a medical condition or medication do not count toward the diagnosis. Patients who have had a major depressive episode in the past are at risk of a recurrence in the context of cancer. Patients may have a pre-existing dysthymic disorder, which may become more manifest or worse in cancer patients.

Management of Depression in Cancer Patients. Management of depression in cancer patients requires a comprehensive approach that addresses the evaluation, treatment, and follow-up. The initial management begins with the establishment of a therapeutic alliance with the clinician and recruiting support from family or friends. It is important to note that a relationship with the oncologist is the key component of support for cancer patients. Maintaining ongoing contact with depressed cancer patients, especially terminally ill patients, ensures that continual evaluation and caring are available.

The NCCN has applied the APA practice guidelines for the treatment of depressive disorders in physically healthy individuals to the treatment of depression in cancer patients. There are several pharmacologic and psychotherapeutic strategies available. Prior to selecting an appropriate treatment, the site of cancer, current cancer treatment, comorbid medical conditions, and medications should be taken into consideration, any of which may contribute to depressive symptoms. Reversible conditions such as thyroid function abnormalities should be ruled out. If the depressive disorder is thought to be secondary to a medical condition or caused by a drug, the clinician should treat the underlying condition or change the drug; however, antidepressants are usually started concurrently to relieve the patient's suffering as quickly as possible.

PHARMACOLOGIC TREATMENTS. The use of antidepressant medications in cancer patients creates unique challenges. Rapid onset of action is preferable in cancer patients, especially in the terminally ill; however, antidepressants may take several weeks to have a therapeutic effect due to their delayed onset of action. An appropriate antidepressant should be selected based on the potential side effects of each antidepressant, drug–drug interactions, the patient's prognosis, primary symptoms of depression, and comorbid condition. Antidepressants should be started at low doses and titrated up slowly in medically frail cancer patients, especially in the older adults.

SSRIs have become the first line of treatment for depressive disorders in the medically ill. They are efficacious, generally well tolerated, and are not as toxic in overdose as TCAs. Some SSRIs, such as fluoxetine (Prozac) and fluvoxamine (Luvox), are inhibitors of cytochrome P450 isoenzymes. It is therefore important to monitor for the possibility of drug–drug interactions. Sertraline (Zoloft), citalopram (Celexa), and escitalopram (Lexapro) are less protein-bound and may have a lower risk of drug interactions with the P450 system. Citalopram (Celexa) causes dose-dependent QT interval prolongation and has an FDA-restricted maximum dose of 20 mg daily for subgroups of patients, including those taking a cytochrome P450 2C19 inhibitor and those older than 60 years. Many of the SSRIs are available in liquid forms, making it easier for patients who cannot swallow pills.

SNRIs are a growing class of antidepressants, which include venlafaxine (Effexor), desvenlafaxine (Pristiq), duloxetine (Cymbalta), and levomilnacipran (Fetzima). They are generally well tolerated, with a benign side-effect profile similar to SSRIs. However, norepinephrine reuptake inhibition may result in palpitations and hypertension. Therefore, blood pressure monitoring is recommended for patients on an SNRI. It is important to note that venlafaxine mainly inhibits serotonin reuptake at low doses; its effect on norepinephrine reuptake inhibition is seen at doses higher than 150 mg per day. Duloxetine shows serotonin and norepinephrine reuptake inhibition at starting doses. All four medications have a low P450 effect and moderate plasma protein binding, which minimize drug–drug interactions and allow for their common use in depressed cancer patients. These medications have proven effective for the treatment of neuropathic pain and are commonly prescribed as adjunct pain medications.

TCAs have been around for many years and are less expensive than many of the SSRIs. Because of their anticholinergic, antiadrenergic, and antihistaminergic side effects, they are less frequently used in cancer patients. Anticholinergic side effects should be especially avoided in older adult patients at risk for anticholinergic delirium or patients at risk for GI dysmotility or obstruction. Their role as adjunct pain medications, especially for neuropathic pain, has become their most common indication for use in cancer patients.

MAOIs are rarely used in the treatment of cancer patients with depression due to the risk of fatal hypertensive crisis when concurrently used with tyramine-rich food or sympathomimetic drugs.

Of the other antidepressants, bupropion (Wellbutrin) acts primarily on the dopamine system and may have a mild stimulant effect, which can be beneficial for cancer patients with fatigue or psychomotor retardation. It is generally tolerated well in the medically ill. Bupropion is associated with an increased risk of seizures at higher doses and should be used with extreme caution in individuals with CNS tumors or seizure disorders. Trazodone is often used in lower doses as a sleep aid rather than a primary antidepressant because its most common side-effect is sedation. Vilazodone (Viibryd) is a serotonin reuptake inhibitor with 5-HT$_{1A}$ receptor partial agonism; it has a less adverse impact on sexual function than other antidepressants. Mirtazapine (Remeron) acts by blocking the 5-HT_2, 5-HT_3, and α-2 adrenergic receptor sites. Its side effects of sedation and weight gain are beneficial for many cancer patients with insomnia and weight loss. It also has antiemetic properties. It is also available in a dissolvable tablet form, which is particularly useful for patients who cannot swallow or who have difficulty with nausea and vomiting. Mirtazapine has minimal drug–drug interactions. Vortioxetine (Trintellix) is a serotonin modulator and stimulator that is approved for the treatment of MDD. It is rarely used in patients with cancer due to the high risk of serotonin syndrome when used alone or in combination with other antidepressants, opioids, or CYP2D6 inhibitors.

Psychostimulants and wakefulness-promoting agents are helpful to treat depressed cancer patients' symptoms of fatigue, psychomotor retardation, and poor concentration. Not only can they induce an antidepressant effect, but they can quicken the response to concomitant antidepressants. Psychostimulants exert dopaminergic effects. Psychostimulants used in cancer patients include dextroamphetamine (Dexedrine), methylphenidate (Ritalin or Concerta), and lisdexamfetamine (Vyvanse). They have a major advantage over antidepressants due to their rapid onset of action and their role in relieving fatigue, promoting wakefulness, and countering opioid-related sedation. Side effects may be anorexia, anxiety, insomnia, euphoria, irritability, and mood lability. However, side effects are not common at low doses and can be avoided by slow titration. Hypertension and cardiac complications can occur; thus, it is advisable to monitor cardiac function. For depressed cancer patients with short survival expectancy, psychostimulants provide rapid relief from distressing depressive symptoms. Modafinil (Provigil) and armodafinil (Nuvigil) are known as wakefulness-promoting agents. They produce increased alertness, wakefulness, and energy. They are presumably better tolerated than psychostimulants but may also cause anxiety, restlessness, and insomnia. They should be used with caution in patients with poorly controlled hypertension.

PSYCHOTHERAPY. Several psychotherapeutic techniques have been successfully employed with depressed cancer patients. Psychotherapy is often combined with pharmacologic intervention. Supportive psychotherapy and cognitive-behavioral therapy are commonly used in cancer patients. MCP is a novel therapeutic modality that has been shown to improve depressive symptoms among advanced cancer patients both in individual and group therapy settings.

Group therapy can be helpful to improve social networks, connecting the patient with others who have the same diagnosis or treatment, decreasing the patient's sense of isolation. Supportive–expressive and cognitive-existential group psychotherapies have been studied and used successfully for cancer patients.

ELECTROCONVULSIVE THERAPY (ECT). ECT is an effective treatment modality for depressed patients. ECT should be considered in patients who are refractory to psychopharmacological treatment (i.e., at least two trials with antidepressant medicines from different pharmacologic classes that are adequate in dose, duration, and adherence), have severe weight loss secondary to depression, exhibit acute psychosis, or have a high suicide risk. Although CNS tumors and cardiac problems are not absolute contraindications, ECT is used with caution in these individuals.

Suicide Assessment and Management in Cancer Patients. The incidence of suicide is higher in cancer patients compared with the general population. Studies suggest that although a small number of cancer patients commit suicide, the relative risk of suicide in this population far exceeds that of the general population, particularly early after diagnosis. Compared to cancer-free persons, the relative risk of suicide after a cancer diagnosis is 12.6 in the first week and 3.1 during the first year; the risk elevations lower rapidly in subsequent years after diagnosis. Suicide may also occur in advanced cancer patients with escalating depression, hopelessness, and the presence of poorly controlled symptoms, particularly pain. Suicidal thoughts in patients with advanced disease, poor prognosis, or poorly controlled symptoms should not be viewed as rational. It is important to note that those patients may have a treatable major depressive episode influencing their suicidal ideation. Clinicians should evaluate for hopelessness and a diagnosis of depression in terminally ill patients with a persistent desire for death or suicidal intention (Table 27.8–6).

Table 27.8–6.
Suicide Risk Factors in Cancer Patients

Depression
Hopelessness
Uncontrolled pain
Extreme fatigue
Anxiety
Delirium
Substance abuse or dependence
Feeling alone in the world
Objective lack of social support
Stressful family relations
Intolerance of loss of control, dependence, and loss of dignity
Severe financial problems
Positive family history
History of severe trauma
Presence of active statements and plans
Personality disorder
Previous history of depression and suicide attempts

Prior history of psychiatric illness, previous history of depression or suicide attempts, recent bereavement, history of alcohol or other substance abuse or dependence, male gender, family history of depression or suicide and lack of family or social support, and recent losses are common risk factors for suicide. Untreated delirium may lead to unpredictable suicide attempts due to impaired judgment and impulse control. Older patients and individuals with head and neck, lung, breast, urogenital, GI cancers, and myeloma seem to have an increased risk of suicide. An international population-based study from Denmark, Finland, Norway, Sweden, and the United States has shown a small but statistically significant, increased long-term risk of suicide, 25 or more years after a breast cancer diagnosis.

Evaluation of suicidal thoughts should take into account the disease stage and prognosis. It is helpful to consider the issue of suicidality from four perspectives: (1) the suicidal thoughts that occur transiently in many patients with cancer, (2) suicidal thoughts in patients who are in remission with a good prognosis, (3) suicidal thoughts in patients with poor prognosis or poor symptom control, and (4) patients in terminal stages. Suicidal ideation or plans to commit suicide in cancer patients with a good prognosis or those in remission also require careful assessment. A Finnish survey of suicides in 1 year revealed that 4.3% had cancer, half of whom were in remission at the time of suicide. Patients were noted to have had a history of psychiatric illness prior to cancer diagnosis, in particular substance abuse.

It is important to recognize and aggressively treat high-risk patients for depression and address suicidal risk with acute psychiatric hospitalization, if necessary. Maintaining a supportive relationship, symptom control (e.g., pain, nausea, depression), and involving the family or friends are the initial steps in the management of a suicidal patient. A study examining the suicidal ideation and past suicidal attempt in adult survivors of childhood cancer found a strong correlation between physical health and suicidality, which underscores the importance of symptom control in cancer patients. Early psychiatric involvement with high-risk individuals can often avert suicide in cancer patients. A careful evaluation includes an exploration of the reasons for suicidal thoughts and the seriousness of the risk. The clinician should listen empathically, without appearing critical or judgmental. Allowing the patient to discuss suicidal thoughts often decreases the risk of suicide despite common belief to the contrary. Patients often reconsider and reject the idea of suicide when the physician acknowledges the legitimacy of their option and the need to retain a sense of control over aspects of their death. Rarely a competent patient presents a difficult problem by continuing to threaten suicide or to ask for help in dying. Patients must be given intense ongoing support, including open discussions about treatment options. Medical staff and family may need to be reminded of the competent patient’s right to refuse all treatments, even life-saving ones.

Bipolar Disorder

Cancer patients with bipolar disorder are most likely brought to the attention of consultation psychiatrists due to the complexity of their management. Acutely manic patients must be brought under control as rapidly as possible. The rigors and dangers of chemotherapy and radiation treatment require a level of consistent cooperation that manic patients cannot be trusted to sustain, even when they understand their situation and earnestly desire treatment. These are difficult cases that require patience and multidisciplinary cooperation. Psychiatrists should carefully assess for safety and level of functioning to determine the best treatment setting. It is also important to screen for substance abuse or dependence and to assess whether the patient is suicidal or a risk to others (homicidal or violent ideation). When the patient cannot be rapidly controlled, a brief psychiatric admission may be necessary, preferably followed by a direct transfer from the psychiatric ward to the oncology ward.

In an acutely manic or hypomanic cancer patient (especially in the older adults), it is important to complete a comprehensive medical assessment to rule out medical conditions or medications as the possible culprit of the mood disturbance. Corticosteroids can cause hypomania or even mania with psychotic features, especially in patients with prior history of mood disorder or a family history of bipolar disorder. The spectrum of mental disturbances related to steroids includes minor mood lability, affective disorders (mania or depression), cognitive impairment, and delirium. The incidence of these disorders ranges from 28% to 55% in noncancer populations, and they occur most commonly on higher doses. Symptoms usually develop within the first 2 weeks on steroids but, in fact, can occur at any time, on any dose, even during the tapering phase. These disorders are often rapidly reversible upon dose reduction or discontinuation. Manic symptoms can also be a part of postictal psychosis, most likely in patients with temporal lobe epilepsy or after temporal lobectomy. An EEG may be helpful for diagnostic clarity. Thyroid function abnormalities should also be ruled out, especially in patients with thyroid carcinomas or those who have undergone radiation to the neck.

Cancer patients with bipolar disorder in remission can be maintained on lithium (Eskalith) or other mood-stabilizing medications. Renal insufficiency is a contraindication to lithium use, but the end-stage renal disease can be managed with low-dose lithium after each dialysis. Medically ill patients must be monitored for lithium toxicity in the presence of vomiting and dehydration (as during chemotherapy), hypercalcemia, renal impairment (as with cisplatin [Platinol] treatment), possible hypothyroidism, or when there is an increased risk of cardiac arrhythmia (as in pre-existing heart disease or with the use of digoxin [Lanoxin], bronchodilators, or cardiotoxic chemotherapy agents, such as doxorubicin [Adriamycin]). Radiation therapy is not a contraindication, except for cranial irradiation, whose side effects mimic lithium toxicity, making monitoring difficult. Patients should be monitored carefully during the postoperative period for the development of nephrogenic diabetes insipidus (NDI), an uncommon but severe side effect. Patients may not report their symptoms of polyuria, polydipsia, or nocturia, as the insidious onset of the condition leads patients and families to believe that the increased intake and output are normal. The problem may go unrecognized for long periods if physicians do not specifically question their patients about the existence of polyuria, polydipsia, or nocturia. Serum lithium levels might be in the therapeutic range despite polyuria. Patients may present with severe hypernatremia during acute illness or after fluid deprivation in the perioperative period. Annual measurement of 24-hour urine volume may help prevent this potentially fatal condition in cancer patients using lithium for the treatment of bipolar disorder.

Cancer-Related Fatigue

Cancer-related fatigue is a persistent, subjective sense of tiredness related to cancer or cancer treatment that interferes with usual functioning. Fatigue is a common and highly distressing symptom of cancer, associated with reduced quality of life, as well as considerable psychological and functional morbidity. Fatigue is more distressing than pain, nausea, or vomiting, which are manageable by medications in most patients. Cancer-related fatigue is usually refractory to sleep and rest. The reported prevalence of cancer-related fatigue ranges from 4% to 91%, depending on the specific cancer population studied and the methods of assessment. Fatigue is present at the time of diagnosis in about 50% of cancer patients. It occurs in up to 75% of patients with bone metastases. Fatigue is also prevalent in long-term cancer survivors and has a serious impact on a person's quality of life.

A group of expert clinicians proposed a set of diagnostic criteria that have been included in the 10th edition of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). A standardized interview guide has been designed and studied for its reliability and validity to assess the proposed clinical syndrome, with results suggesting its utility in identifying patients experiencing clinically significant cancer-related fatigue. Fatigue is difficult to quantify. There are a variety of standardized self-report scales, most of which have been developed in the context of cancer. Different scales may measure fundamentally different aspects or even potentially distinct conceptions of fatigue. The challenge facing the clinician or researcher is to choose a reliable and valid tool for the measurement of fatigue that is most adequately suited to their purposes. Given the multifactorial nature of fatigue, accessory scales (e.g., depression scales) and measurements of certain biologic parameters should be used in addition to fatigue assessment tools in order to obtain a complete evaluation of a patient's fatigue.

Fatigue and depression may coexist in cancer patients, and there is a considerable overlap of symptoms of these two conditions. It is necessary to clarify the relationship between depression and fatigue in order to effectively evaluate and treat cancer-related fatigue. There are many symptoms common to both fatigue and depression, such as decreased energy and motivation, sleep disruptions, diminished concentration and attention, and problems with short-term memory. The clinician seeking to differentiate between depression and fatigue should remember that depressive symptoms due to fatigue are typically less severe and that patients tend to attribute such symptoms to the consequences of fatigue. Depression, on the other hand, is more likely in the presence of hopelessness, feelings of worthlessness or guilt, suicidal ideation, and a family history of depression. The nature of any causal relationship between cancer-related fatigue and depression is unclear.

Cancer-related fatigue is caused by various physical and psychosocial mechanisms, including tumor by-products, opioids, or other drugs (such as antidepressants, β-blockers, benzodiazepines, antihistamines), hypogonadism, hypothyroidism, cachexia, anemia, chemotherapy, radiation therapy, bone marrow transplantation, and treatment with biologic response modifiers. Cancer-related fatigue has been related to pain, depression, emotional distress, sleep deprivation, and reduced physical activity. Cytokines (interleukin-1 and -6 [IL-1, IL-6], tumor necrosis factor-α [TNF-α]) have a role in the development of cancer-related fatigue, which has led researchers to consider cytokine-antagonist drugs such as TNF receptor etanercept (Enbrel) and TNF-α antagonist thalidomide (Thalomid) to improve the tolerability of chemotherapy regimens and potentially to treat cancer-related fatigue and cachexia. Genetic variables have also been implicated to play a role in the development of fatigue among cancer patients. Advanced colorectal cancer patients with two variant forms of the dihydropyrimidine dehydrogenase (*DPYD)* gene —an enzyme that is involved in the breakdown of excess nucleotides involved in DNA, RNA, and ATP production—were significantly less likely than those patients with a form of the gene known as *DPYD*5* to report fatigue following treatment with a chemotherapy regimen of 5-fluorouracil, irinotecan, and oxaliplatin.

All patients should be screened for fatigue at their initial visit, at regular intervals during and following cancer treatment, and as clinically indicated. The NCCN practice guidelines on cancer-related fatigue recommend the use of numerical self-report scales or verbal scales to assess the severity of fatigue. If moderate or severe (a score of 4 to 10 on a scale of 0 to 10 with higher numbers indicating increased severity), a focused history and clinical examination with evaluation of the pattern of fatigue, associated symptoms, and interference with normal functioning are recommended. Potentially reversible causes of fatigue (such as pain, emotional distress, sleep disturbance, anemia, hypothyroidism) should be identified and treated, and nonessential centrally acting drugs should be eliminated. If anemia is the main cause of fatigue, the physician should determine the necessity of a transfusion in severely symptomatic patients. Clinical trials have shown that anemic patients have improved energy and less fatigue after erythropoietin treatment.

Nonpharmacologic approaches have been recommended by NCCN guidelines for the treatment of cancer-related fatigue. Activity enhancement and psychosocial interventions (i.e., education, support groups, individual counseling, stress management training) have been well supported by research. Dietary management, attention-restoring therapy, sleep therapy, and mind–body interventions have also been recommended in the treatment of cancer-related fatigue.

Many pharmacologic interventions are being evaluated in clinical trials for the treatment of fatigue in cancer patients (http://clinicaltrials.gov/). Psychostimulants, wakefulness-promoting agents, and antidepressants are the psychotropic medications used in the treatment of cancer-related fatigue.

Psychostimulants have also been used in the treatment of fatigue-related conditions such as pain, depression, and cognitive impairment. Methylphenidate has been studied most and appears to be relatively effective and well-tolerated despite relatively common side effects such as insomnia and agitation. Dexmethylphenidate is the d-isomer of methylphenidate, with a longer duration of action (about 6 hours) compared to methylphenidate. It has been found to be well tolerated and more effective than a placebo for improving fatigue in adult cancer patients.

Dextroamphetamine is the d-isomer of amphetamine and is more potent than methylphenidate, is commonly prescribed by clinicians for the treatment of cancer-related fatigue.

A meta-analysis of pharmacologic treatment options for cancer-related fatigue has concluded that methylphenidate might be effective for treating fatigue. A significant placebo response has been observed in randomized trials. There was also evidence that treatment with hematopoietic agents relieved fatigue due to chemotherapy-induced anemia.

Modafinil is a wakefulness-promoting agent, approved by the FDA in December 1998 for the treatment of excessive daytime sleepiness in patients with narcolepsy. Compared to other psychostimulants, modafinil has a novel mechanism of action and less abuse potential. It presumably enhances activity in the hypothalamic wakefulness center (i.e., tuberomammillary nucleus), activates tuberomammillary nucleus neurons that release histamine, and activates other hypothalamic neurons that release orexin/hypocretin. Modafinil is commonly used for the treatment of severe fatigue in multiple sclerosis and has been studied as a treatment option for cancer-related fatigue with an improvement of fatigue. The results are supportive of its use among cancer patients with severe fatigue. Further well-designed, randomized, controlled clinical trials are necessary to clarify the role of modafinil and armodafinil in the treatment of cancer-related fatigue.

Underlying depression should be treated with SSRIs, which are generally better tolerated than TCAs in cancer patients. When prescribing such medications, clinicians should remember that, as fatigued patients are generally more sensitive to antidepressant side effects, treatment should be initiated with small doses. Drug–drug interactions should also be carefully monitored. Studies have examined the role of paroxetine, sertraline, and bupropion in the treatment of cancer-related fatigue. Paroxetine and sertraline were effective in improving fatigue among cancer patients with comorbid depressive symptoms. Bupropion was found to be effective and well tolerated in both depressed and nondepressed cancer patients in open-label trials. However, controlled studies are required to determine whether the effect of bupropion on fatigue is independent of its antidepressant effects.

Corticosteroids have been used in the treatment of cancer-related fatigue. Current evidence on the use of corticosteroids is anecdotal, and responses to corticosteroids appear to be temporary. Due to concerns for toxicity associated with long-term administration, the use of corticosteroids should be restricted to the terminally ill, with concomitant fatigue and anorexia, and/or those with pain related to bone or brain metastases.

Megestrol acetate (Megace) is a progestational agent that has been found to improve appetite in cancer-related cachexia and has also been suggested to have a role in the treatment of cancer-related fatigue. The effects of megestrol acetate on fatigue are not clear, but they probably involve anticytokine and corticosteroid-type effects (Table 27.8–7).

Personality Disorders

Personality disorders are chronic character styles that lead to clinically significant interpersonal and behavioral problems. Personality affects the way an individual copes with different life events. This aspect is important in the medical setting because it will influence how the patient will cope with cancer and its concurrent stressors such as hospitalization and treatment. During the evaluation, it is important for the consultant psychiatrist to identify maladaptive responses to stressors as defense mechanisms that the patient might be using to deal with cancer. Patients’ symptoms are chronic and ego-syntonic, and it is the family and staff who feel stressed by a worsening of the patient’s usual dysfunctional modes of adaptation and who ask for help because of increased conflict and disrupted treatment. Personality-disordered patients do have acute stress syndromes, which may be overlooked because of the chronic issues. Evaluation should search for other psychiatric disorders. Comorbid anxiety or depression should be actively treated. Environmental measures should be implemented to decrease conflict and to improve patient care, including family and staff involved in the intervention. Patients with personality disorders are also distressed by unbearable emotional conditions and might require attention to basic needs such as sleep deprivation, hunger, and necessary medical and comfort care. Medications such as benzodiazepines or antipsychotics can be used to stabilize agitation or emotional upheaval. The psychiatrist should establish a working relationship by being empathetic with the patient and by being aware of transference and countertransference issues. The dynamics of difficult patients and ways of dealing with them have been well described in the psychiatric literature. The feelings of anger and guilt that patients with personality disorders routinely induce in mental health staff also occur in oncology staff. The latter, especially when inexperienced, may find these feelings harder to deal with because they are not trained to recognize these feelings as part of the patient’s pathology and are disarmed by the genuine pathos of the patient’s condition. Explanations, support, and guidance are helpful.

Table 27.8–7.
Psychotropic Medications Used in Cancer-Related Fatigue

Medication	Starting Dose	Dose Range	Comments
Psychostimulants			
Methylphenidate	2.5–5 mg daily or twice daily	5–30 mg/day, usually divided as twice daily	Longer-acting forms are available. Capsule forms can be sprinkled in food.
Dextroamphetamine	2.5–5 mg daily or twice daily	5–30 mg/day, usually divided as twice daily	Longer-acting formulations are available. Capsule forms can be sprinkled in food.
Wakefulness promoting agents			
Modafinil	50–100 mg daily	50–400 mg daily, may be divided as twice daily	Favorable side-effect profile
Antidepressants[a]			
Selective serotonin reuptake inhibitors			
Well-tolerated. Citalopram, escitalopram, and sertraline seem to have the least drug–drug interactions.			
Fluoxetine[b]	10–20 mg/day	10–60 mg/day	
Paroxetine	10–20 mg/day	10–40 mg/day	
Citalopram[b]	10–20 mg/day	10–40 mg/day	
Escitalopram[b]	5–10 mg/day	5–20 mg/day	

Sertraline[b]	25–50 mg/day	25–200 mg/day
Serotonin-norepinephrine reuptake inhibitors		
Well-tolerated. Monitor blood pressure, especially at higher doses.		
Venlafaxine	37.5–75 mg/day	37.5–225 mg/day
Duloxetine	20–30 mg/day	20–60 mg/day
Norepinephrine-dopamine reuptake inhibitor		
Doses higher than 300 mg a day should be used twice daily to minimize the risk of seizures.		
Bupropion	75 mg/day	75–450 mg/day
Tricyclic antidepressants		
Drug–drug interactions are more common compared to SSRIs.		
Secondary Amines		
Desipramine	25–50 mg/day	25–200 mg/day
Nortriptyline[b]	10–25 mg/day	10–150 mg/day
Tertiary Amines		
Amitriptyline	10–25 mg/day	10–150 mg/day
Doxepin[b]	25–50 mg/day	25–300 mg/day
Imipramine	10–25 mg/day	10–200 mg/day
α-2 antagonist/5-HT$_2$/5-HT$_3$ antagonist		
Mirtazapine	7.5–15 mg/day	7.5–45 mg/day

[a] Antidepressant medications have not been proven to be efficacious in the treatment of cancer-related fatigue in the absence of a diagnosis of clinical depression.

[b] Available in liquid formulations.

Schizophrenia and Other Psychotic Disorders

Patients with schizophrenia may appear less often in doctors’ offices because their disease interferes with the early and consistent seeking of care. They give idiosyncratic meanings to their symptoms and may tolerate them without complaint, often because of what is called a *high pain threshold*, although a more accurate description would be a different central processing mechanism. The patients’ unreliability continues throughout the treatment course and may complicate postoperative and long-term medical management. When treating hospitalized patients with schizophrenia, it is important to ensure continuity of care. Ongoing medications can usually be continued, except for a brief operative period. The medical staff needs help in handling bizarre behaviors and reassurance about intimidating mannerisms. They need a precise protocol and written orders for the management of aggressive or disruptive episodes, as well as a clear understanding of the legal and ethical considerations governing the patients’ right to refuse medical treatment and to leave the hospital. Decisional capacity is situation-specific and independent of the presence of a psychiatric disorder; hence, psychiatric consultation should be called to help establish its presence or absence in any conflictual situation.

Substance Use Disorders

In 2018, an estimated 164.8 million people aged 12 or older in the United States (60.2%) were past-month substance users (i.e., tobacco, alcohol, or illicit drugs). The 164.8 million past-month substance users in 2018 include 139.8 million people who drank alcohol, 58.8 million people who used a tobacco product, and 31.9 million people who used an illicit drug. While prevalence data for substance use disorders in cancer patients is limited, one can infer that a significant proportion of patients with a new cancer diagnosis have problematic drug and alcohol use.

Oral, pharyngeal, laryngeal, esophageal, breast, liver, and rectal cancers are more common in alcohol users than in non-alcohol users. Smokers who also drink are at much higher risk. Alcohol withdrawal syndrome can greatly affect a patient's cancer treatment. During acute hospitalization, alcohol withdrawal may commence on the second or third day and elicit a psychiatric consultation for anxiety, poor coping, or postoperative delirium. Thoracic and head and neck cancer patients are at high risk for this and for nicotine withdrawal, given their high incidence of smoking. The symptoms of nicotine withdrawal include anxiety, irritability, and insomnia. Transdermal nicotine patches are commonly used to ease the distress of patients forced to stop smoking suddenly. A clinician must be mindful that cigarette smoke includes polycyclic aromatic hydrocarbons that contribute to the induction of several cytochrome P450 enzymes (specifically CYP1A1, CYP1A2, and CYP2E1) and induction of glucuronidation of some drugs. Abrupt cessation of smoking can affect medication levels; this effect is independent of nicotine and not countered with nicotine patches.

Managing cancer pain in addicted patients requires a multidisciplinary approach with the involvement of a psychiatrist, a palliative care physician, social workers, and nurses. The psychiatrist should manage psychiatric comorbidities and the countertransference of the team. It is also important to set limits to control harm with substantial emotional support. Pain management in addicted cancer patients is challenging and should be focused on the level of pain, with attention to the addiction. In treating patients with past or current opioid use disorders, the clinician must be mindful of the effect of tolerance and opioid-induced hyperesthesia, in which cases the patient will require higher doses for pain control. Patients will benefit from long-acting opioids such as methadone (Dolophine) and by limiting the use of short-acting or breakthrough doses. Furthermore, adjuvant medications such as SNRIs or antiepileptics may be helpful. Scheduled urine toxicology screens with clear consequences and contingency plans are used to monitor patients' illicit drug usage.

Somatoform Disorders

The prevalence of somatoform disorders in cancer patients is unknown. The development of cancer in patients with a somatoform disorder can create difficult diagnostic problems because patients may incorporate the presenting signs and symptoms of cancer into their psychiatric disorder–associated illness behaviors. This results in complex syndromes with both oncologic and psychogenic components. Physicians may feel manipulated and frustrated by these presentations, which may lead to delays in treatment or suboptimal care through over- or under-treatment. A psychiatric evaluation and collaborative effort may allow the psychiatrist to work with the oncology staff to treat the patient appropriately through clarification of presentation, addressing psychiatric comorbidity, and setting limits on maladaptive behaviors.

PSYCHIATRIC SYNDROMES AND TUMOR SITE

CNS Cancers

Brain tumors are classified as primary or metastatic tumors. Brain metastases occur up to 10 times more frequently than primary brain tumors. Seventy percent of brain tumors are supratentorial. Primary brain and CNS tumors are relatively uncommon, accounting for approximately 23,380 (1.4%) new cancer diagnoses and 14,320 (2.4%) cancer-related deaths annually. Meningiomas, which are usually nonmalignant, are the most common tumor type. Glioblastomas are the most common primary malignant brain tumors, accounting for approximately 16% of all new primary brain tumors and 45% of primary malignant brain tumors. Unfortunately, they are generally associated with dismal prognoses, with 5-year survival rates of <5%.

Psychiatric symptoms of brain tumors are influenced by their anatomic location, interconnection of different areas of the brain, rate of tumor growth, premorbid level of functioning, tumor type, and increased intracranial pressure. Psychiatric symptoms are present in almost all patients with supratentorial tumors and are the presenting symptom in 25% of the patients. Frontal lobe tumors are associated with behavioral symptoms in 90% of patients. Mood symptoms can include anxiety, irritability, depression, apathy, and euphoria. There have been reported changes in personality in 70% of patients. These changes include irresponsibility, disinhibition, and inappropriate sexual behavior. Psychotic symptoms, including delusions and hallucinations, have been reported in 10% of patients. Bifrontal disease typically causes impairment of memory, lability of mood, urinary incontinence, and frank dementia, with prominent grasp, suck, and snout reflexes. Psychotic symptoms are common in patients with tumors of the temporal lobe and pituitary gland. Tumors affecting the temporal lobe can present with auditory and olfactory hallucinations as well as abnormal behavior (apathy, irritability, expansive mood with mania or hypomania). Nondominant temporal lobe tumors can cause minor perceptual problems and spatial disorientation. Dominant temporal lobe tumors can present with an impaired perception of verbal commands and sensory aphasia. Parietal lobe tumors affect sensory and perceptual functions more than motor modalities. Tumors in either parietal lobe can produce sensory disorders ranging from mild sensory extinction or astereognosis to a more severe sensory loss such as hemianesthesia. Poor proprioception in the affected limb is common. Homonymous inferior quadrantanopsia, incongruent hemianopsia, or visual inattention also may occur. Nondominant parietal lobe tumors cause contralateral neglect and, in severe cases, anosognosia and apraxia. Dominant parietal lobe tumors lead to alexia, dysgraphia, and certain types of apraxia. Occipital lobe tumors can produce contralateral homonymous hemianopsia or complex visual aberrations affecting the perception of color, object size, or location. Bilateral occipital disease, rarely seen at presentation, can produce cortical blindness.

Delirium is a common psychiatric complication in patients with CNS tumors. The main causes of delirium in this population are CNS insults, medications (corticosteroids), and metabolic changes. Dementia and other cognitive impairment are also symptoms and complications of the treatment of CNS tumors.

Management of Psychiatric Symptoms in Patients with Brain Tumors. Brain tumor patients will benefit from psychopharmacological, psychotherapeutic, and psychosocial interventions. The psychiatrist should work closely with the medical and surgical team to help patients function and improve their quality of life.

The data that support the pharmacologic treatment of patients with psychiatric symptoms come from clinical experience and case reports. Medications must be initiated at low doses due to frequent brain insults caused by the cancer and its treatment. Antipsychotics are indicated for treating delirium, anxiety, and mood disorders. The psychiatrist should be cognizant of several side effects in this population. Antipsychotics can lower the seizure threshold. Antipsychotics such as clozapine (Clozaril) and chlorpromazine have the greatest potential to lower seizure threshold; thus, they should be avoided. High-potency antipsychotics are indicated if sedation is not desired. Atypical antipsychotics should also be considered in treating psychotic and mood symptoms.

Psychostimulants such as methylphenidate and modafinil can improve attention, mood, and cognitive function. These medications should also be started at low doses and be used with caution due to the potential of causing hallucinations and decreasing the seizure threshold.

Antidepressant medications are effective in treating depression and anxiety. The SSRIs are frequently used due to their low anticholinergic activity compared to TCAs. Benzodiazepines for treating anxiety symptoms can induce disinhibition and delirium in patients with brain tumors.

Mood lability can be treated with lithium, carbamazepine (Tegretol), valproic acid (Depakene), and atypical antipsychotics.

Head and Neck Cancers

Head and neck cancers account for almost 3% of all cancer diagnoses. Head and neck cancers are diagnosed almost twice as often in men as in women. Time of diagnosis is more often after age 50, except for salivary gland and nasopharyngeal tumors, which may occur in younger age groups. Cigarette smokers have an increased risk for multiple head and neck primary cancers as well as lung cancer. Alcohol use has also been implicated as a causative factor for certain head and neck cancers, and the effects of alcohol and tobacco use seem to be additive. Infection with some types of human papillomavirus (HPV), particularly HPV-16, is a risk factor for head and neck cancers, especially oropharyngeal cancers that involve the base of the tongue and tonsils. In the United States, the incidence of oropharyngeal cancer associated with HPV infection is increasing. Patients with head and neck cancers have an increased risk of developing esophageal and lung cancers.

Surgery and radiation therapy are necessary to induce a cure. Chemotherapy and other systemic treatments (e.g., immunotherapy) are useful in the (neo)adjuvant setting or along with radiation, but when used alone, systemic treatments are not curative. Most head and neck cancers present with early or locally advanced disease that can be treated for cure. But the treatments are quite long and intense generally.

The diagnosis of head and neck cancer brings many psychosocial challenges. Patients are confronted with fears of death, pain, and loss. Patients also are concerned with disfigurement and disability from cancer treatment. In anticipation of surgery, patients experience distress and anxiety. An incidence of adjustment disorders as high as 36% to 50% has been reported. Recent studies have also demonstrated prevalence rates of clinically significant depression diagnosed by various self-reporting and structured interview techniques of between 17% and 30%, specifically in head and neck cancer patients. Both pharmacologic and psychological interventions are effective in treating depression in these patients.

Another challenge is managing the sensitivity to the stigma associated with HPV-positive oropharyngeal cancers as a sexually transmitted disease. Established risk factors for developing HPV-positive oropharyngeal cancers include having greater than six lifetime oral sex partners, concurrent anogenital squamous cell carcinoma, and HIV infection. Furthermore, husbands of women with cervical cancer are at a threefold increased risk of HPV-positive oropharyngeal cancer. Psychotherapy may focus on addressing the perceived shame of the transmission method.

In a suicide study among cancer patients, unmarried men with head and neck cancers were found to have the highest incidence of suicide. Other risks factors for suicide are personality disorders, alcoholism, and treatment sequelae such as facial disfigurement. Postoperative patients are at risk of developing delirium due to alcohol withdrawal, metabolic disturbances, and other medical complications. Extensive reconstructive operations affect swallowing physiology, leading to aspiration events and situational anxiety.

GI Cancer

Newly diagnosed cancers of the oral cavity, pharynx, and digestive system account for 21% of total cancer incidence and 25% of total cancer mortality in the United States.

Patients who suffer from GI tumors must adjust to several psychosocial changes. Patients have to deal with severe eating-related problems, weight loss, abdominal discomfort, diarrhea or constipation, as well as other disease-related issues such as ostomy and repeated operations.

The psychological issues in patients with colorectal cancer are related to the cancer itself and the impact of treatment of bowel function in relation to social and sexual activity. Patients who require ostomies may interpret this procedure as a significant loss. Different strategies have been recommended to patients with ostomies at different phases of recovery. Patients may feel anxious, hopeless, and helpless when an ostomy is recommended. It is important to involve the family and provide adequate teaching of ostomy procedural management and troubleshooting to minimize distress. Patients may experience shame, low self-esteem, anxiety, or sexual undesirability following their operation. Community ostomy support groups can provide guidance and counseling to patients and their families.

Surgical procedures involving the pelvic area may lead to sexual dysfunction through the removal of the posterior wall of the vagina, weakening of muscles of the pelvic floor, or damage to the pudendal and associated nerves. Determining the etiology of sexual dysfunction is challenging but dictates the potential treatments.

Lung Cancer

Lung cancer is the leading cause of cancer mortality in men and women in the United States, accounting for an estimated 154,050 deaths in 2018. Lung cancer accounts for 20% of all cancer deaths worldwide. In the United States, lung cancer is the leading cause of cancer-related death and exceeds deaths from breast, colon, and prostate cancers combined. Although lung cancer is increasingly more common in patients who have never smoked (especially for females), tobacco use is still associated with approximately 85% of lung cancer prevalence and is the primary risk factor for lung cancer. The overall 5-year survival for lung cancer is approximately 18% despite offering many new targeted and immunologically based treatments. Since the 5-year overall survival for localized lung cancer is 56%, the major reason for poor survival is the lack of early detection. Low-dose CT scans are now offered for patients with a significant smoking history (over 20 pack-years), but not enough patients receive screening. The most common histologic type of cancer is non–small cell (NSCLC), followed by small cell lung cancer (SCLC), which accounts for about 15% of all lung cancers.

Patients with lung cancer experience high levels of emotional distress following diagnosis and persistently through treatment. Disease-related symptoms such as pain, fatigue, and dyspnea are detrimental to the emotional well-being of patients. Studies of distress in patients with lung cancer reveal significant problems in the areas of emotional functioning, family relationships, and lack of information about the diagnosis, physical functioning, and cognitive functioning. In a study that included 987 patients, the prevalence of depression in lung cancer patients was found to be as high as 33%, and in half of those, the depression persisted at the first follow-up visit, 14 to 35 days after beginning chemotherapy for SCLC and 21 to 56 days after starting radiotherapy for NSCLC. An epidemiologic study by Sullivan and colleagues found that depressed lung cancer patients had inferior survival outcomes when compared to nondepressed patients with lung cancer. However, the survival disadvantage was only observed in patients who were persistently depressed on follow-up. Survival returned to baseline (nondepressed) levels for patients whose depression had remitted on follow-up. This longitudinal study not only demonstrates the importance of identifying and treating depression but demonstrates longitudinal variability based on depression status.

Given the overwhelming physical and emotional symptom burden for patients with lung cancer, cognitive disorders are also highly prevalent in patients with lung cancer, especially SCLC. These disorders may be related to brain metastases or other comorbidities and complications of the illness and its treatment. Paraneoplastic syndromes, brain radiation, and chemotherapy have been associated with cognitive impairment. Depression can also lead to cognitive impairment, and patients should be screened and treated for this condition.

Smoking is common in lung cancer patients, and some studies show that smoking cessation at the time of diagnosis of lung cancer may reduce the rate of development of tumors. In other words, patients who continue to smoke once diagnosed with lung cancer have inferior survival. Also, the relative risk of a second lung cancer developing following curative intent therapy of SCLC was lower for those who stopped smoking. Although the incidence of lung cancer never returns to exactly never smoker lung cancer incident, it begins to approximate it by about 15 years of smoking cessation. Active and recently quit smokers should be offered intensive tobacco cessation programs, including counseling, behavioral therapy, use of sustained-release bupropion and nicotine replacement, and telephone follow-up, to increase successful abstinence. Smoking cessation should be an active part of lung cancer treatment.

Genitourinary Cancers

Prostate Cancer. Prostate cancer is one of the leading causes of cancer death among men in the United States and is the most common cancer in men. Prostate cancer screening is a controversial issue, with many arguing that digital rectal examination (DRE) and PSA may lead to unnecessary and possibly harmful consequences. A mortality benefit has not been established from DRE or PSA screening. The U.S. Preventive Services Task Force and the Centers for Disease Control and Prevention recommend against the PSA. The American Society of Clinical Oncology and the American College of Physicians discourage screening in those expected to live less than 15 years and recommend others decide after weighing their potential risks and benefits. Treatment options for prostate cancer are surgery and external beam radiation or radioactive seed implants (brachytherapy). Hormonal therapy, chemotherapy, and radiation are offered for metastatic disease and as supplemental therapies for early-stage disease.

Men who have undergone curative-intent treatment with hormonal therapy plus radiation or surgery may experience mood liability associated with antihormonal therapies as well psychological issues regarding the risk and reality of impotence. A large component of therapy with men with prostate cancer is learning to adjust to this chronic disease state and find meaning and fulfillment outside these areas.

Testicular Cancer. Testicular cancer comprises a morphologically and clinically diverse group of neoplasms, most of which are germ cell tumors (GCTs) and are the most common solid tumors in men between the ages of 15 and 35 years. The pathognomonic presentation of a primary testicular tumor is a painless testicular mass that may range in size from a few millimeters to several centimeters. A radical inguinal orchiectomy minimizes local tumor recurrence and aberrant lymphatic spread. It is the only acceptable therapeutic and diagnostic procedure. GCT is classified into two major subgroups, seminomas and nonseminomas. Treatment for an early stage of seminomatous tumors is orchiectomy and radiation therapy. For moderate and advanced disease, the treatment is combined radiation and chemotherapy. Nonseminomatous tumors are treated with orchiectomy and combined chemotherapy for more advanced disease.

Bladder Cancer. Bladder cancer is two and a half times more common in males than in females and more common in whites than in blacks. Approximately 53,000 new cases occur per year in the United States, which is a 20% increase from 20 years ago. The incidence increases with age and peaks in the sixth and seventh decades of life. More than half of men and one-third of women with bladder cancer have a history of smoking cigarettes. The clinical stage of the primary tumor is determined by transurethral resection of the bladder tumor (TURBT). Surgery, alone or in combination with other treatments, is used in more than 90% of cases. Superficial, localized cancers may also be treated by chemotherapy or immunotherapy directly into the bladder. Chemotherapy alone or with radiation before cystectomy has improved some treatment results.

Kidney Cancer. Each year in the United States, there are approximately 31,900 cases of kidney and upper urinary tract cancers, resulting in more than 11,900 deaths. Renal carcinoma may remain clinically occult for most of its course. Approximately 30% of patients with renal carcinoma present with metastatic disease, 25% with locally advanced renal carcinoma, and 45% with localized disease. Surgery is the only known effective therapy for localized renal carcinoma. Adjuvant or palliative nephrectomy is frequently performed in patients with metastatic renal carcinoma. Many new targeted and immunotherapies are being offered to patients with localized and metastatic kidney cancer.

Management of Psychological Issues in Patients with Genitourinary Cancers. The causes of psychological distress in men with genitourinary cancers are related to the diagnosis of cancer, treatment options, potential sexual dysfunction, urinary incontinence, fatigue, weakness, pain, and other side effects from cancer and its treatment. Patients in this population have difficulties coping with changes in sexuality, relationships, and lifestyle. Management of these psychological issues includes education, psychotherapy, couple and family therapy, behavioral interventions, and medications.

GCTs, bladder cancer, and kidney cancers are some of the most common primary cancer sites that can spread to the brain. Metastasis must be on the clinician's differential diagnosis if a patient with one of these cancers presents with delirium or altered mood, personality, or behavior. Genitourinary cancer survivor men may be left feeling an insult to their well-being and identification with their gender. They may also have psychological decompensation related to their physical symptom burden.

Breast Cancer

Breast cancer ranks second among cancer deaths in women. Breast cancer accounts for about a third of all new cancer cases in the United States. Death rates have declined by an average of 2.3% per year since the 1990s in all women, with larger decreases in younger (less than 50 years) women. Risk factors for breast cancer include age, inherited genetic mutations (*BRCA1* and *BRCA2*), personal or family history of breast cancer, high breast tissue density, and radiation to the chest. Reproductive factors include a long menstrual history, use of oral contraceptives, never having children, and having a first child after age 30. Other factors are obesity, postmenopausal hormone therapy, physical inactivity, and consumption of one or more alcoholic beverages per day. Treatment of breast cancer involves lumpectomy or mastectomy with removal of axillary lymph nodes, radiation therapy, chemotherapy, or hormone therapy such as aromatase inhibitors. These treatments are often used in combination. Breast cancer has a complex psychological impact on women. Three factors have been reported to contribute to this psychological impact: the sociocultural context in which treatment options are offered, the psychosocial and psychological factors, and the medical factors. Breast cancer also has an impact on lifestyle due to physical discomfort, marital or sexual disruption, and disturbed body image. Women with breast cancer are likely to suffer from anxiety and depression. Prevalence of depression among early breast cancer women occurs twice as often as in the general female population, especially during the first year after diagnosis. In a study that included 4,496, the prevalence of depression among patients with breast cancer was estimated to be around 52.65%. Management of psychological distress includes providing patients with the skills to cope with their illness and to improve their quality of life. Antidepressant medications and psychotherapy are the gold standard treatments for mood and anxiety disorders in women with breast cancer. Antidepressants are also beneficial in the treatment of hot flashes in breast cancer. Approximately two-thirds of women with a history of breast cancer experience hot flashes due to cytotoxic or hormonal therapies that can cause premature menopause.

The current pathophysiologic hypothesis for hot flashes is the dysfunction of hypothalamic thermoregulatory processes due to low estrogen levels. Norepinephrine and serotonin are the two most recognized neurotransmitters involved in hypothalamic thermoregulation. In a study that included breast cancer survivors and menopausal women, each group was assigned to receive one of three different dose levels of venlafaxine (37.5 mg, 75 mg, and 150 mg daily) or placebo for 4 weeks. A dose-related diminution in average hot flash scores from baseline was noted (27% in the placebo group versus 37%, 61%, and 61% for the three venlafaxine groups, respectively).

Consideration should be given to potential drug interactions, including the potential for an antidepressant to affect hepatic metabolism of cancer drugs via cytochrome p450 enzyme induction or inhibition in all cancer patients. In breast cancer patients, tamoxifen has been widely studied in this context because of early reports that indicated an association between antidepressants with strong 2D6 inhibition (such as fluoxetine and paroxetine) and reduced efficacy of tamoxifen in maintaining remission in breast cancer. While several studies have confirmed that antidepressant drugs that have strong 2D6 inhibition can reduce the conversion of tamoxifen to its active metabolite, more recent studies indicate that this effect does not appear to have an impact on breast cancer recurrence rates or survival. However, until further research resolves the debate, it remains prudent to avoid potent CYP2D6 inhibitors in preference for antidepressant options that have weak 2D6 inhibition (such as venlafaxine or citalopram).

Gynecologic Cancers

The group of gynecologic cancers includes those of the ovary, fallopian tube, uterine corpus, cervix, vagina, and vulva. Cancer of the endometrium is the most common gynecologic cancer. Ovarian cancer has the highest mortality of gynecologic cancers. Over the last 30 years, the cervical cancer death rate has gone down by more than 50% due to increased use of the Pap test. Furthermore, the HPV vaccine will continue to decrease the number of women affected by the cancer-causing pathogen. Vaginal and vulvar cancers are less common. They are often detected earlier and thus more responsive to treatment. Women may sometimes delay seeking evaluation because the symptoms are mild or are attributed to other problems. For instance, abdominal bloating may be mistaken for premenstrual syndrome.

Treatment for these cancers includes surgery, chemotherapy, and/or radiation. The cancer and its treatment have significant impacts on hormones, sexuality, fertility, and bladder and bowel function, causing psychological distress.

Hematologic Malignancies

Hematologic malignancies are associated with significant psychosocial distress in many patients, particularly those undergoing stem cell transplantation. Leukemia, lymphoma, and multiple myeloma (MM) account for approximately 9% of all cancers, with an estimated annual combined incidence in the United States of about 150,000. Blood dyscrasias can occur during childhood or early adulthood, thus impacting patients during a normally productive and active time in their lives. Complicated treatment regimens with intensive, multicomponent protocols have greatly increased cure rates and have prolonged survival.

Distress is common before, during, and after treatment for hematologic malignancies and occurs along a continuum of severity that is elevated compared to the general population. Adjustment, mood, and anxiety disorders are particularly common. Individuals undergoing HSCT encounter major disruptions to social, family, and work life. They are separated from supports during transplant or chemotherapy induction and hospitalization. Challenges prior to allogeneic transplant may emerge as a result of stem cell donor selection, family stressors, and financial, housing, and other practical concerns. Patients are more vulnerable to distress during periods when prognosis or treatment course change (e.g., graft failure), severe adverse effects develop (e.g., GVHD), rehospitalization occurs, or changes occur to clinician follow-up (e.g., discharge from transplant center). Periods of immunologic compromise limit social contact, even after discharge. Illness-associated stigma and job loss can occur. Patients who have psychiatric issues such as anxiety or depression are at risk for poorer health outcomes, longer hospital lengths of stay, and higher mortality. Insomnia, delirium, fatigue, depression, and anxiety are common psychiatric symptoms and syndromes encountered during the treatment course. Survivors commonly experience PTSD, depression, anxiety, and cognitive impairment long after treatment completion. Psychosocial support is an essential component of the treatment of patients with hematologic malignancies at all stages of illness.

Paraneoplastic Syndromes

The term *paraneoplastic* refers to a group of disorders defined as indirect, remote, nonmetastatic effects of cancer. In 50% of the cases, symptoms precede the diagnosis of cancer and run an independent course. The syndromes may be due to (1) production of substances by the tumor that directly or indirectly cause distant symptoms, (2) depletion of normal substances that lead to a paraneoplastic manifestation, or (3) host response to the tumor that results in the syndrome.

Paraneoplastic Neurologic Disorders. Any part of the CNS and peripheral nervous system can be affected. Most paraneoplastic neurologic disorders begin acutely or subacutely, progress over weeks to months, and then stabilize. Many of these disorders are associated with serum and cerebrospinal fluid antibodies formed against antigens that are also expressed within the tumor. Dementia, particularly memory loss (amnesia), is a common feature in several variants, which are grouped under the name *carcinomatous encephalomyelitis*. Most of these patients have SCLC and anti-Hu antibodies.

Limbic encephalitis is another paraneoplastic syndrome that usually presents with abrupt changes in mood, personality and with severe impairment in recent memory. Patients may appear agitated or confused and may experience hallucinations as part of complex partial seizures. The neoplasms associated with limbic encephalitis include testicular cancer, SCLC, thymoma, breast cancer, and Hodgkin lymphoma. The antibodies associated with limbic encephalitis are anti-Hu, anti-Ma2, CRMP5, and amphiphysin.

Opsoclonus-myoclonus ataxia syndrome consists of uncontrolled arrhythmic eye movements. It is also associated with diffuse or focal myoclonus, truncal titubation with or without ataxia, or other cerebellar signs. Ataxia, irritability, insomnia, and dementia often accompany the opsoclonus. The incidence of this disorder is less than 2%. In adults, it occurs in the presence of breast, gynecologic, or other types of cancers. In children, it can be due to neuroblastomas or viral infections. It is associated with anti-Ri antibody. Paraneoplastic syndromes that affect the motor neurons can also present with psychiatric symptoms in patients with degeneration of cortical neurons, resulting in frontotemporal dementias.

Neuromyotonia consists of several disorders that affect the neuromuscular junction. These include neuromyotonia, the cramp fasciculation syndrome, and Morvan syndrome. The disorder is autoimmune and affects voltage-gated potassium channels. Morvan syndrome consists of memory loss, insomnia, changes in mood, and hallucinations. These disorders are associated with thymoma, SCLC, and Hodgkin lymphoma.

The treatment for paraneoplastic syndromes consists of treating the underlying tumor and suppression of the immune response using corticosteroids, plasma exchange, and intravenous immunoglobulin.

Metabolic and Endocrine Paraneoplastic Syndromes. Production of cytokines, hormones, or hormone precursors by the tumor may result in the development of endocrine syndromes in cancer patients. Tumors of endocrine glands may be physiologically active and produce anxiety, mood disorders, confusional states, and psychosis. In addition, glucocorticoid steroid medications are frequently used for almost all cancer as part of therapeutic regimens to reduce inflammation or as prophylactic antiemetic agents. These frequently cause insomnia, irritability, and depression.

Hypercortisolism (Cushing syndrome) can be caused by a pituitary adenoma, hypothalamic tumors, ectopic adrenocorticotropic hormone (ACTH) syndrome, and exogenous corticosteroids. Hypercortisolism presents with mood instability, depression, insomnia, fatigue, cognitive deficits, psychosis, and delirium.

A syndrome of inappropriate antidiuretic hormone secretion can be associated with SCLC, cerebral metastases, medications, and infections. This syndrome results in hyponatremia; the severity of the symptoms depends on the sodium level and the rate of drop in sodium levels. Patients can present with symptoms of nausea, vomiting, lethargy, stupor, delirium, cognitive impairment, seizures, agitation, focal neurologic deficits, and coma.

Carcinoid tumors cause anxiety, depression, and delirium. Carcinoma of the pancreas has been reported to be associated with an elevated incidence of depression, sometimes as the presenting symptom. Metabolic disturbances, including hypoglycemia, hyperglycemia, hypophosphatemia, and hypomagnesemia due to different types of cancers and can cause delirium, mood disturbances, and anxiety.

PSYCHIATRIC PROBLEMS DUE TO CANCER TREATMENT

Radiation and most of the drugs used in cancer treatment have significant side effects and are even carcinogenic in tumoricidal doses. Patients may become transiently more ill from the treatment than from the disease and develop side effects such as hair loss, all of which have a great emotional impact. Steady compliance requires a level of courage and endurance of which the medical staff is not fully aware. Patients may become upset about the side effects, convinced that they represent disease progression, depressed about the paradoxical nature of the treatment, and exhausted to the point at which they develop clinical depression and anxiety. If untreated, they delay or stop cancer treatment, moving to alternative treatments and jeopardizing their survival. Later adaptation is also impaired. Many patients, however, come to feel dependent on continued treatment as a means of holding the cancer at bay. There is a rise in anxiety level at the end of treatment associated with this fear and with the difficulty of making a transition to the survivor status. This can be assuaged by acknowledging it, exploring its cause, and making concrete plans for continued medical supervision and psychological support.

Radiation Therapy

About half of all the patients with cancer are treated with radiation therapy, either alone or in combination with other types of cancer treatment. Radiation therapy may be external or internal and may be curative or palliative. External radiation comes from a machine outside the body and is usually given on an outpatient basis. Internal radiation is implanted into or near the tumor in small capsules or containers. It may require a hospital stay. If the patient will have external radiation, the radiation oncologist uses a process called simulation to define where to aim the radiation. During simulation, the patient lies still on an examining table while the radiation therapist uses a special x-ray machine to define the treatment port or field. Simulation may also involve imaging studies (CT scans) to help the radiation therapist plan on how to direct the radiation. Patients with anxiety disorders, claustrophobia, or pain have difficulties with simulation and radiation treatment. The early effects of radiation may be seen a few days or weeks after treatments have started and may continue for several weeks after treatments are completed. Late effects may not be seen until months, or even years, later. Side effects depend on the site, dose, and size of the radiation fields and the duration of treatment. The skin may be red, tender, and sensitive during the first 2 weeks. Dryness and peeling may occur in 3 to 4 weeks. Patients may suffer from fatigue, sore throat, and anorexia. Patients who receive total body irradiation are at the greatest risk for having nausea. Without effective prophylaxis or treatment, more than 90% will experience nausea and vomiting acutely. Affected patients may suffer from various adverse complications, including dehydration, electrolyte disturbances, malnutrition, and aspiration pneumonia. Treatment of nausea includes adequate nutritional support and hydration. Medications to treat nausea include treatment with 5-HT$_3$ receptor antagonists (5-HTRAs) with or without corticosteroids, dopamine antagonists, antihistamines, or anticholinergics. Sexual impairment is common after radiation to the pelvic area, in the form of impotence in men and dyspareunia in women. The latter, related to vaginal fibrosis, is improved by vaginal dilatation. Acute, transient GI problems, such as diarrhea and food intolerance, are common; severe, irreversible GI problems are rare. Fatigue is reported in 80% of patients and is sometimes the most debilitating and demoralizing. It overlaps with symptoms of depression but exists independently of it. The normal response to brain radiation has been divided into three phases based on the time of initiation of symptoms. Acute effects occur during the first weeks of treatment. Symptoms include headache, drowsiness, nausea, vomiting, and worsening of pre-existing focal symptoms. Treatment with dexamethasone usually relieves symptoms. Early delayed or subacute encephalopathy appears at 1 to 6 months after completion of radiation therapy with headache, somnolence, fatigue, and deterioration of previous deficits. Subacute encephalopathy is caused by demyelination. It responds to steroids and is usually self-limited. The third is an irreversible late, delayed (months to years) form that is localized to the white matter and is probably secondary to vascular injury, demyelination, and necrosis. Late effects are also known as focal radiation necrosis, diffuse white matter injury, and combined therapy leukoencephalopathy. Symptoms of focal radiation necrosis include seizures, signs of increased intracranial pressure, and symptoms of localized lesions. Symptoms of diffuse white matter changes vary from lassitude to significant memory loss to severe dementia. Symptoms of combined therapy leukoencephalopathy include ataxia, confusion, dysarthria, seizures, dementia, or death. In long-term survivors (mainly from Hodgkin lymphoma and childhood leukemias) who have undergone whole brain radiation, a short-term memory deficit has been observed, as well as a decrease in intelligence quotient (IQ) of 10 to 25 points. Even in the presence of a normal IQ, neuropsychological problems and learning disabilities have been documented. The older adults and the young are the most affected.

Chemotherapy

Nausea and Vomiting. The immediate effects of chemotherapy include nausea and vomiting, which occur within 2 to 6 hours of treatment and may last up to 24 hours. The most emetogenic drugs cause symptoms in virtually all patients; the less emetogenic ones are variable. Some result in delayed vomiting, usually 8 to 24 hours after treatment. Chemotherapeutic agents and radiation cause vomiting by different mechanisms. Chemotherapy stimulates the chemoreceptor trigger zone directly, which is functionally outside the blood–brain barrier. This area, when exposed to toxins in the bloodstream and cerebrospinal fluid, can stimulate vomiting. This effect is mediated by 5-HT_3 and neurokinin type 1 receptors. Chemotherapy damages the GI mucosa and causes the release of neurotransmitters, including 5-HT_3, stimulating peripheral pathways via splanchnic and vagal nerves. It has also been reported that there is a neurohormonal component mediated by arginine vasopressin and prostaglandins.

Classical conditioning has been shown to be an important mechanism in the causation of several reactions to chemotherapy: ANV, which had been documented in 25% to 75% of patients before the use of better antiemetics; anticipatory immune suppression (AIS), the clinical significance of which has not been established; and learned food aversions (LFAs), which can be minimized by avoidance of staples and favorites. Patients develop a conditioned response to sights and smells reminiscent of the chemotherapy experience. After termination of treatment, these symptoms decrease; however, one study found that 6 to 14 months after treatment, 63% of patients still reported nausea, 5% reported vomiting, and 80% still reported anxiety with specific stimuli.

These reactions are presumed to represent conditioned responses, and treatment approaches have been formulated accordingly. Prophylactic antiemetic and antianxiety treatments are used to prevent or decrease the initial episode or conditioned response and have greatly minimized what used to be a dreaded side effect. With the advent of highly specific, centrally acting antiemetic agents, such as the 5-HT_3 receptor blockers, ondansetron (Zofran), and granisetron (Kytril), the success of antiemetic regimens has greatly increased, and new ones with different mechanisms continue to appear. The evidence support that these agents are effective in chemo- and radiation-induced nausea and postoperative nausea. Other medications that are indicated for the treatment of nausea are dopamine 2 (D_2) receptor antagonists that block the chemoreceptor trigger zone. The literature has shown that for the most common etiologies of nausea and vomiting at the end of life, 5-HT_3 antagonists are no more effective than the less expensive D_2 antagonists. Tardive dyskinesia is rare due to the short-term use of these medications. Akathisia occurs in 10% to 20% of patients on metoclopramide, but the coincident use of lorazepam lowers the incidence to less than 3%. Mild sedation is acceptable or even welcome. Hypotension may limit the use of prochlorperazine

(Compazine) and chlorpromazine. Since some of the nausea is also anxiety-related, benzodiazepines are effective in preventing anticipatory nausea. In an open-label trial, mirtazapine, a potent antagonist of 5-HT_2 and 5-HT_3 receptors, was found to decrease nausea. The commercially available cannabinoids, nabilone (Cesamet) and dronabinol (Marinol), target ubiquitous cannabinoid (CB) receptors in the central (CB1) and peripheral (CB1 and CB2) nervous systems. Dronabinol is no more effective than other drugs and has a similar side-effect profile, plus a significant incidence of dysphoric reactions in older patients. Corticosteroids, such as dexamethasone, are generally included in most antiemetic regimens for their potential effect on tumor-associated inflammation.

Behavioral methods are important, effective ways to lessen the reactions to chemotherapy. They are most effective when instituted before the conditioned response is established. They include hypnosis and progressive muscular relaxation with guided imagery or relaxation with cognitive distraction to block the conditioned stimulus and the conditioned response; systemic desensitization, which extinguishes the conditioned response; and cognitive distraction, which blocks perception of the conditioned stimulus. Patients are also helped by reassurance about the frequency and normalcy of their responses.

CNS Toxicity. Chemotherapeutic agents may cause neuropsychiatric problems in one of the following ways: Indirectly by causing metabolic abnormalities or other organ failure and directly by causing CNS toxicity. The CNS, comprised of the brain and spinal cord, is responsible for the neurologic functions of mental status, level of consciousness, sensory and motor functions, and cerebellar and cranial nerve functions. The blood–brain barrier determines whether a cytotoxic agent can reach the nervous system. Chemotherapeutic agents that can penetrate the blood–brain barrier can alter the cranial nerves and cognitive, cerebellar, motor, and sensory functions. Patients themselves are keenly aware of the cognitive impairment and talk to each other about the so-called *chemo-brain* in the waiting room and on the Internet.

Initially, mood changes, personality changes, and subtle cognitive changes may predominate, leading to psychiatric consultation. Only later does the encephalopathic nature of the condition becomes manifest. Since the mid-1970s, cross-sectional and longitudinal studies have found an association between chemotherapy and cognitive changes. Cognitive functioning is objectively assessed by performing a battery of neuropsychological tests that allow comparison to a control group or normative data. The research data have shown that there is a discrepancy between objective measures of cognitive impairment and the patients' self-report of their cognitive functioning. Patients' perception of the dysfunction appears to be more deteriorated than what is detected by the objective evaluation and is associated with anxiety or depression. This discrepancy between patients' self-report and objective measures of cognitive function might be due to the inability of the neuropsychological tests to evaluate cognitive functioning in real-life situations and the fact they are performed under optimal conditions. Another explanation for the discrepancy is that patients tend to attribute their premorbid cognitive impairment to cancer or to depression or anxiety. It is imperative to evaluate premorbid cognitive functioning that serves as a parameter for comparison with cognitive impairment posttreatment. Although acute cognitive changes during chemotherapy are common, studies in survivors have shown long-term posttreatment cognitive changes in 17% to 34% of patients. The cognitive changes occur in different areas of cognition, including working memory, attention/concentration, executive function, and information processing speed. The cognitive functioning is reduced but often remains in the normal range.

Studies that have used imaging techniques have shown a reduction in the volume of brain structures. Changes in the integrity of white matter tracks that connect brain structures have been implicated in decreased cognitive functioning.

Possible mechanisms by which chemotherapy can cause cognitive deterioration include direct neurotoxic effects, oxidative damage, immune dysregulation with the release of cytokines, decreased delivery of oxygen secondary to anemia and blood clotting, and hormonal changes. Genetic predisposition with a variability of genes that regulate neural repair or plasticity, such apolipoprotein E (APO E) and BDNF, may increase patients' vulnerability to chemotherapy-induced cognitive side effects. Certain drugs, such as the vinca alkaloids, especially vincristine, cause neurotoxicity at all doses; others, such as cisplatin, cause neurotoxicity only at higher doses. In some drugs, toxicity is mild and transient; in others, it is delayed, irreversible, or both. Toxicity is also usually, but not always, dose-related and cumulative. Certain drugs, such as methotrexate, when given intrathecally, may cause several CNS side effects, such as reversible encephalopathy, delayed irreversible encephalopathy, dementia, or developmental delay when coadministered with radiation therapy. Several studies have found an association between cognitive impairment and the number of cycles and higher doses of chemotherapy. A chemotherapy regimen that consists of methotrexate, cyclophosphamide (Cytoxan), and 5-fluorouracil (Adrucil) appears to be highly associated with cognitive impairment.

Several treatment interventions have been initiated and evaluated with methylphenidate, erythropoietin, Ginkgo biloba, and cholinesterase inhibitors, but the data are inconclusive, and further research is needed. Large longitudinal studies are needed to evaluate the duration of cognitive impairment, which remains uncertain (Table 27.8–8).

Table 27.8–9 lists agents that predominantly cause mood or psychotic symptoms. These chemotherapy agents and many more antineoplastic agents can cause delirium. Delirium due to chemotherapy agents is often reversible and remits spontaneously if the drug is stopped. This may take only a few days or up to several weeks. If the offending drug cannot be discontinued, symptomatic treatment should be instituted.

Disorders of special senses include ototoxicity, associated with cisplatin and several commonly used antibiotics; visual disturbances, associated with mitotane (Lysodren) and tamoxifen; gustatory and olfactory hallucinations, described with vinblastine. These disorders may occur slowly or suddenly and reversibly or permanently and often result in acute psychological distress requiring psychiatric consultation. Long-term effects of ototoxicity add considerably to the educational and neuropsychiatric impairment of these patients. The diagnosis is often missed in younger children, in whom early intervention is essential to minimize later developmental impairment.

Ifosfamide is a highly active chemotherapy agent, but its neuropsychiatric toxicity is a challenging problem, occurring in 5% to 30% of patients with cancer. Clinical symptoms can range from mild confusion to stupor, coma, and death. Somnolence, confusion, disorientation, fatigue, hallucinations, agitation, blurred vision, aphasia, seizures, asterixis, extrapyramidal symptoms, mutism, stupor, coma, and death, have all been described. Delirium is the most common neuropsychiatric toxicity. Psychotic disorders with delusions or hallucinations in the presence of a clear sensorium and mood disorders with psychotic features induced by ifosfamide have also been described. Chloroacetaldehyde (CAA) is a cytotoxic ifosfamide metabolite. The principal effect of CAA is the depletion of intracellular glutathione, which results in neurotoxicity due to impaired detoxification of neurotoxic substances. CAA is structurally similar to metabolites of ethanol (acetaldehyde) and chloral hydrate (trichloroacetaldehyde) with CNS depressant effects. Discontinuation of ifosfamide, and intravenous hydration, are part of the initial supportive care. Methylene blue, an agent FDA approved for the treatment of methemoglobinemia, and lower genitourinary tract discomfort, was first reported for treatment and prophylaxis of ifosfamide neuropsychiatric toxicity in 1994. Intravenous methylene blue 50 mg every 4 hours is recommended until the symptoms of ifosfamide neuropsychiatric toxicity resolve. Methylene blue is hypothesized to reverse ifosfamide neuropsychiatric toxicity through several different mechanisms, namely restoring mitochondrial respiratory chain function by serving as an alternative electron acceptor, restoring hepatic gluconeogenesis by reversing nicotinamide adenine dinucleotide plus hydrogen (NADH) inhibition, preventing the formation of CAA. Psychotropic medications were not found to be helpful in controlling psychiatric symptoms induced by ifosfamide in the authors’ experience, as also reported by other researchers.

Antineoplastic Hormones and Biologic Modifiers

The hormonal agents and biologic modifier agents are used in an increasing number of cancers. They are associated with depressive symptoms but also with many adjustment disorders related to their sexual side effects, such as feminization, masculinization, and disorders of sexual functioning. Besides the reversible effects already described, interferon can cause acute reversible encephalopathy, with confusion, lethargy, and hallucinations. One case of delayed progressive encephalopathy has been reported. IL-2 (Proleukin) causes headache, agitation, disorientation, and confusion, which are dose-related and reversible. A rare severe encephalopathy has also been reported. These drugs have become frightening to patients who hear of the multiple side effects. However, symptomatic treatment is effective for all but the rarest severe reactions. A randomized prospective study of melanoma patients given prophylactic paroxetine has raised some controversy, with some researchers decrying the indiscriminate use of an SSRI in a group of patients who may not all need it. Previous psychiatric history was a strong risk factor, as was the presence of baseline symptoms. It is easy for psychiatric consultants to help oncologists put in place a screening program asking patients about their psychiatric history and their baseline mood symptoms. Simple questionnaires, such as the Hospital Anxiety and Depression Scale (HADS), which has been developed for medically ill patients, have identified a 10% to 15% sample that would qualify for prophylactic treatment. Ongoing monitoring for depressive symptoms would allow early treatment and would spare patients from unnecessary suffering. Once a medical team has been educated, and a routine has been set in place, ongoing psychiatric consultation is not necessary for any but the new-onset or complicated cases.

Table 27.8–8.
Neurologic Complications of Selected Chemotherapeutic Drugs

Acute encephalopathy (delirium)

- Corticosteroids
- Methotrexate (Folex, Mexate) (high dose IV, IT)
- *Cis*-platinum
- Vincristine (Oncovin)
- Asparaginase (Elspar)
- Procarbazine (Matulane)
- 5-Fluorouracil (Adrucil) (± levamisole [Ergamisol])
- Ara-C (Cytosar-U)
- Nitrosoureas (high dose or arterial)
- Cyclosporine (Neoral)
- Interleukin-2 (Proleukin)
- Ifosfamide (IFEX)/mesna (MESNEX)
- Interferons
- Tamoxifen (Nolvadex)
- VP-16 (Toposar) (high dose)
- *N*-phosphoracetyl-l-aspartate (PALA)

Chronic encephalopathy (dementia)

- Methotrexate
- BCNU (BiCNU)
- Ara-C
- 1-Hexyl-5-fluozouracil (Carmofur; Mifurol)
- Fludarabine (Fludara)

Visual loss

- Tamoxifen
- Calcium nitrate
- Nitrosoureas (intra-arterial)
- *Cis*-platinum

Cerebellar dysfunction/ataxia

- 5-Fluorouracil (± levamisole)
- Ara-C
- Phenytoin (Dilantin)
- Procarbazine
- Hexamethylmelamine (Hexalen)
- Vincristine

Cyclosporine

Aseptic meningitis

Trimethoprim and sulfamethoxazole (Co-trimoxazole) (Bactrim)
IV immunoglobulin (Iveegam)
NSAIDs
Levamisole
Monoclonal antibodies
Metrizamide (Amipaque)
OKT3 (anti-C3 monoclonal antibody)
Ara-C
Carbamazepine (Tegretol)
Methotrexate

Headaches without meningitis

Retinoic acid (Avita)
Trimethoprim and sulfamethoxazole
Cimetidine (Tagamet)
Corticosteroids
Tamoxifen

Seizures

Methotrexate
VP-16 (high dose)
Cis-platinum
Vincristine
Asparaginase
Nitrogen mustard (Mustargen)
BCNU
Dacarbazine (DTIC-Dome) (intraarterial or high dose)
PALA
m-AMSA (AMSA P-D; Amsacrine)
Busulfan (Busulfex) (high dose)
Cyclosporine
Metronidazole (Flagyl)
Misonidazole
β-Lactam antibiotics
Iodinated contrast material (IV or IT)

Myelopathy (IT drugs)

- Methotrexate
- Ara-C
- Thiotepa (Thioplex)

Peripheral neuropathy

- *Vinca* alkaloids
- *Cis*-platinum
- Hexamethylmelamine
- Procarbazine
- 5-Aza12cytidine (5-Aza-C)
- VP-16
- VM-26 (Vumon)
- Misonidazole
- Methyl-C
- Ara-C
- Paclitaxel (Taxol)
- Suramin (Bayer 205)
- Mitotane (Lysodren)

BCNU, bischloroethylnitrosourea; IT, intrathecal; IV, intravenous; NSAIDs, nonsteroidal anti-inflammatory drugs.

Table 27.8–9.
Chemotherapy Agents Associated with Mood and Psychotic Symptoms

Dacarbazine (DTIC-Dome): depression and suicide reported, especially when used with hexamethylmelamine (Hexalen)
Vinblastine (Velban): frequent reversible depression
Vincristine (Oncovin): 5% incidence of hallucinations; depression noted
L-Asparaginase (Elspar): reversible depression noted
Procarbazine (Matulane): monoamine oxidase inhibitor; concurrent tricyclics and selective serotonin reuptake inhibitors are contraindicated; associated with mania and depression; potentiates alcohol, barbiturates, and phenothiazines
Hydroxyurea (Hydrea): hallucinations reported
Interferon: anxiety, depression with suicidal ideation common at high doses
Steroids: frequent alterations of mental state, ranging from emotional lability through mania or severe, suicidal depression to frank psychosis
Tamoxifen (Nolvadex): rare depression or delusional syndrome, or both
Leuprolide (Lupron): depression

Tailor-made vaccines against various parts of the malignant cell are being developed in increasingly sophisticated ways, in particular, for melanoma. They remain difficult to make, the responses are usually measured in months, and they are limited to research-based programs, but they are of great scientific interest and have much unexplored potential.

Immune-Mediated Checkpoint Inhibitor Immunotherapy

This new class of antineoplastic agents targeted key regulatory proteins that are designed to shut down the production of immune antibodies. As part of normal homeostasis, checkpoint inhibitor proteins prevent autoimmune disease and ensure that once an immune-mediated reaction is induced, it will be properly shut down. Checkpoint inhibitor therapy prevents the shutdown and allows for a greater and more sustained immunologic response. These therapies count on the immune system to have created its own unique immune response to the cancer, which appears foreign and helps to bolster the response by not allowing it to shut down. Of course, the checkpoint inhibitor is not specific to any one particular antibody response. However, side effects in patients with autoimmune diseases have not been as substantial as it was originally thought to be. Cancers induce immune responses at a variable rate, namely depending on how foreign the cell appears. Therefore, the number of mutation per cancer cell (appears more foreign) can predict response to checkpoint inhibitor therapy.

Surgery

Surgery continues to play a central role in the treatment of cancer. Several psychological and psychiatric issues arise around the time of surgery, such as an inability to give consent, preoperative anxiety, exacerbation of pre-existing psychiatric disorders, postoperative delirium, drug–drug interactions, drug withdrawal concerns, especially for patients on benzodiazepines, poor compliance with postsurgical rehabilitation recommendations, and many others. In addition to perioperative reactions common to different types of surgical procedures, psychological reactions specific to the surgery site and the related functional loss should be taken into consideration. Many studies on the psychological effects of mastectomy, hysterectomy, colostomy, and prostatectomy report the high cost paid in quality of life and recommend that more emphasis be placed on this dimension. Laryngectomy patients and those who undergo other extensive head and neck procedures experience considerable psychiatric morbidity; the loss of voice is catastrophic for many, and the blow to self-image is severe. Patients who have had curative limb amputations can have a good psychological outcome. There is little systematic information about patients who undergo unusual and extreme procedures, such as pelvic exenteration.

Stem Cell and Bone Marrow Transplants

Stem cell transplantation is being used for an increasing number of childhood and adult cancers. It is commonly recommended for leukemias, lymphomas, MM, and myelodysplastic syndromes. It is a potentially curative treatment for otherwise lethal disorders and succeeds best when given to patients in remission, with as little previous treatment as possible. Yet, it has an uncomfortable course and a significant incidence of graft-versus-host disease. Undertaking this type of treatment, especially when the patient is doing relatively well, can be stressful for patients, families, and caregivers. However, successful transplants are indeed lifesaving. Comprehensive psychiatric assessment has become a part of the routine pretransplant evaluation in major cancer centers to identify patients at high risk. Patients sometimes need psychological intervention before the procedure and often need it if their posttransplant course is complicated. Delirium is common during acute periods. Anxiety and depression symptoms are common due to the rigors of the experience. Nevertheless, there are enough patients who leave the hospital after a few weeks, with variable degrees of emotional and physical depletion, but essentially cured of an incurable disease.

New Supportive Technologies

The use of life support and nutritional support in critically ill patients has created a spate of psychological, legal, and ethical issues that are widely discussed in the lay press. Less well known are developments in the nutritional support of cancer patients who can no longer take in food because of physical obstruction, as in head and neck cancers, or because of physiologic failure, as in radiation enteritis. In the former, percutaneous endoscopic gastrostomy (PEG) or percutaneous endoscopic jejunostomy (PEJ) and, in the latter, home parenteral nutrition (HPN) allows a significant improvement in quality of life. These home-based technologies create major changes in family-structuring routines, such as the preparing and sharing of meals, and are time and labor-intensive, causing adaptational strains that may bring them to psychiatric attention.

PSYCHOLOGICAL CONCERNS IN PATIENTS WITH CHRONIC CANCER

Cancer has become curable in a growing number of cases, especially with improved screening techniques and early diagnosis. However, in many cases, it becomes a chronic disease, at which point treatment goals focus on management. While these efforts may be effective for many years, the patient's experience is nonetheless challenging and can involve navigating secondary prevention guidelines, episodes of acute disease, and sometimes coping with indolent tumors.

Patients vary in how they deal with these difficult life situations, and doctors vary in the style and content of their communications. Communication skills training has been identified as a neglected part of medical education and is being addressed by researchers and educators. All physicians communicate not only through their words but also with voluntary and involuntary nonverbal behaviors, which are similarly subject to interpretation, scrutiny, and sometimes distortion. In the case of cancer care, oncologists face the challenge of balancing the details of diagnosis, prognosis, and realistic expectations with the installation of hope. When treating patients with cancer, the psychiatrist must have an understanding of the medical reality, including the burden of illness and staging, treatment options and plan, and long-term ramifications of both the disease and treatment. Patients with cancer frequently ask their psychiatrists to clarify details of their treatment. This may happen for several reasons, ranging from having a better therapeutic relationship with their psychiatrist than their oncologist to limitations during oncology appointments such as anxiety, surprise/shock, or time.

Cancer Pain

The overall incidence of pain in cancer patients is approximately 50% but increases in frequency and severity with advancing disease. There are many pain syndromes that have been defined in cancer patients. Their variety illustrates the importance of evaluating and diagnosing the sources of pain in cancer patients because the treatments for each of them are different and, in many cases, specific. The psychiatrist has a special role to play throughout the course of the illness, especially as pain specialists can be difficult to access, and physicians still display reluctance to treat the pain of cancer patients aggressively enough. The primary goal of cancer pain management is effective symptom control to allow patients to participate in the necessary diagnostic and therapeutic approaches to treat their cancer and to achieve an acceptable quality of life at all stages of the disease.

The debilitating and demoralizing effects of pain give rise to an increased incidence of psychiatric symptoms and disorders at any stage of the disease. Pharmacologic treatment is essential. Patients and families may be comforted to know that pain treatment with opioids very rarely results in addiction, despite the emergence of physiologic dependence evident in tolerance and withdrawal. Addiction, or psychological dependence, is a pattern of behavioral and cognitive symptoms that indicate continued use of a substance despite significant substance-related problems. These symptoms include impaired control, social impairment, failure to fulfill major role obligations, and risky use. Patients' and physicians' fears of addiction are a major barrier to adequate cancer pain management. Patients with poorly managed or undertreated pain may mimic the signs of addiction, such as requesting higher and more frequent doses or exaggerating pain complaints; this behavioral presentation is sometimes referred to as pseudo-addiction.

Identification of the nature of the pain (somatic, visceral, and neuropathic) and the specific pain syndrome can facilitate the development of an approach for pain relief in patients with pain and cancer. It is well recognized that pharmacologic approaches represent the mainstay of therapy, but any therapeutic strategy must also implement the use of psychological, behavioral, mechanical, anesthetic, and neurosurgical approaches as needed.

The guiding principles of a therapeutic strategy for cancer pain must include

1. detailed assessment of the patient’s pain
2. individualization of the therapeutic approach
3. the assurance of available expertise to provide therapeutic strategies to patients
4. continual reassessment of the degree of pain relief and impact on mood, functional status, patient and family acceptance, and the patient’s overall quality of life
5. choosing the simplest approach prior to the use of complicated and extensive techniques
6. ongoing communication between the physician and patient in defining the options for therapy and the potential risk and benefit ratios of any of the therapeutic approaches
7. defining the goals of pain management at or near the end of life

Critical to the development of a successful therapeutic strategy is the physician’s ability to establish a trusting relationship with the patient. This requires a complete history of the pain complaint, including the patient’s description of the site of pain, its quality, exacerbating and relieving factors, its radiation, if any, its exact onset and temporal pattern, and to what extent the pain interferes with activities of daily living, work, and social life, as well as the degree to which the pain has had a psychological impact. Multiple pain complaints, common in patients particularly with advanced disease, need to be prioritized and classified. In some instances, verifying the history from a family member may be necessary because the patient is either unable or unwilling to detail the pain symptom appropriately. Some patients may deny or minimize their pain; in these cases, family members may be more objective in assessing the disability of patients who underreport their symptoms. At the time of the initial assessment, the patient’s psychological state should be evaluated, including the patient’s psychiatric history, the current level of anxiety or depression, suicidal ideation, and degree of functional incapacity. Numerous studies support the significance of psychological factors in accounting for differences in pain experiences in patients with cancer. In a study of the incidence of psychiatric disorders in patients with cancer, 39% of those patients who received a psychiatric diagnosis had significant pain. In contrast, only 19% of patients who did not receive a psychiatric diagnosis had significant pain. In this group with pain, the psychiatric diagnoses included adjustment disorders with depressed or mixed mood and major depression. Uncontrolled pain is a major factor in cancer-related suicide. Understanding the meaning of pain to the cancer patient is a critical component of a successful assessment strategy. It is not uncommon for patients during active therapy to endure significant pain for the promise of a successful outcome. With advanced disease, when curative therapy is no longer an option, it is common for patients and families to request that, if nothing else can be done, at least the patients' pain should be adequately managed.

Chronic cancer patients may have a constant source of pain or one that worsens slowly. The patient may become habituated to a given level of opiate and need increasing levels of opioids at a given level of disease and as the disease progresses. Persistent patient complaints should be heeded, even if scans and MRI do not reveal a new etiology. Chronic pain management modalities are important. Nonpharmacologic methods such as meditation, meditation-based stress reduction (MBSR), behavioral methods, group support, and supportive therapy focused on improving quality of life all have a role to play.

The principles of acute postoperative pain relief emphasize aggressive, immediate treatment, often with the use of patient-controlled analgesia (PCA) pumps. They should be carefully used in confused patients, and a bowel regimen should always be instituted at the same time that any opiate is prescribed. Sufficient initial relief is associated with an easier postoperative course and, in some studies, a shorter hospital stay. Pre-emptive anesthesia is known to prevent phantom pain, and there is one report of its beneficial effect in prostate cancer surgical patients.

Psychiatrists should be familiar with drug–drug interactions related to the use of opioids analgesics, and adjuvant pain medications (i.e., antidepressants, anticonvulsants). Tramadol (Ultram) is a commonly prescribed, weak mu-opioid receptor agonist that also inhibits serotonin and norepinephrine reuptake. As such, patients taking tramadol concurrently with antidepressants should be monitored for serotonin syndrome. Opioid analgesics, especially those with long elimination half-lives, increase the risk of delirium in cancer patients. Rotation of opioids, continual assessment, and prescribing the lowest effective dose are recommended approaches to minimize adverse effects.

Physicians should also remain alert to the possibility that cancer patients may experience pain from other causes. Through longitudinal evaluation of quality and persistence of complaints, psychiatrists play an important role in a collaborative approach to pain management.

Sexuality

Many patients hide their sexual concerns because they are embarrassed and perceive that their doctors are embarrassed. However, they are usually eager for information that is seldom provided. These complaints may vary from reports of sexual dysfunction and decreased libido to sudden onset of menopause symptoms.

Concerns for sexual function vary in younger and older cancer patients. Younger patients are more sexually active, and fertility remains an important issue. Research efforts have revealed conflicting results regarding the incidence of sexual dysfunction based strictly on the site of the primary cancer. A large sample of men with testicular cancer found no significant differences in dysfunction from population-based controls. However, a second study found that long-term male survivors of lymphoma had significantly more sexual dysfunction than matched controls. The sexual impairments associated with prostate, bladder, and rectal cancers match the amount of postsurgical sympathetic nerve damage rather than gross anatomical change. Both radical prostatectomy and radiotherapy for prostate cancer, as well as adjuvant hormone treatment, are often followed by severe long-term erectile dysfunction; sexual recovery is seldom achieved. Any course of radiation or chemotherapy can result in some sexual dysfunction, with radiation causing a high rate of dyspareunia.

Young female breast cancer patients often experience a long-term lack of sexual interest. The attitude of their partners toward their body and femininity is very important for their sexual well-being. Women treated for gynecologic cancers may develop a lack of interest, vaginal dryness, and dyspareunia. Treatment-induced menopause is common to surgical, radiologic, and chemotherapeutic modalities of treatment. The symptoms are the classically described ones, but anecdotal evidence points to higher severe symptoms when menopause is induced by treatment. Hot flashes are among the most troubling to treat because estrogens are contraindicated in many cancer patients, especially those with estrogen receptor–sensitive tumors.

The incidence of sexual dysfunction increases over time, and spontaneous remissions do not occur, even in the presence of stable or improved physical states. Both men and women treated for colorectal cancer can have persistent sexual dysfunction 4 years after diagnosis. Dysfunction is greater in younger patients across all diagnoses and is particularly high for cancers of breast, head, and neck, and other physically disfiguring cancers. In an ideal world, sexual concerns should be routinely addressed before treatment, with a baseline assessment that models open communication from the start. A detailed sexual history can be used to explore all possibilities and make subsequent discussions easier. Patients need detailed information about what problems to expect and about their reversibility. The information should be combined with reassurance that some form of sexual expression is always possible. If there is a spouse or sexual partner, early joint involvement may be useful.

Early discussions are particularly necessary because sexual concerns play an important role in patient decision-making. Both men and women may decide against the optimal treatment because of the fear of sexual impairment, which is not always justified. Then again, they may select very aggressive treatment in ambiguous situations without realizing the full impact it will have on their sexual functioning.

During the treatment, patients may experience direct pain and injury about which they may not grieve openly. Even in the absence of direct pain and injury, there is usually an interruption of sexual activity, which may be very threatening for patients who are uncertain about their partners' tolerance for the interruption. The extent of recovery needs to be clarified so that realistic coping can begin.

After the initial treatment period, patients should again be queried about sexual concerns. Patient workshops, support groups, and information sessions may be useful tools to complement individual treatment. A flexible format with a mix of information, emotional ventilation, and focused problem solving have been well received in several types of cancer.

The period after the end of active treatment is a critical one for future adjustment. The patient usually cannot resume premorbid functioning even in the absence of any direct injury, but fatigue, pain, disappointment, and fear about the future are maximal. It is also a time of uncertainty and, hence, a window of opportunity. Dysfunctional patterns have not yet crystallized, and given the opportunity, couples can work together toward a new equilibrium.

Recent studies reinforce the observation that psychological and relational factors play a large (often the largest) role in determining the degree of posttreatment recovery of sexual functioning. Given the lack of attention paid to sexual issues that is characteristic of much of medical treatment, psychiatrists and other mental health professionals can occupy an important role in the evaluation of sexual issues and facilitating treatment. They need to know the basic anatomy and physiologic potential of a patient's condition and should refer the patient for further work-up and interventions if needed. Psychological treatment includes couple psychotherapy and sex therapy. Sexual rehabilitation involves an adaptation of sex therapy principles and can be helpful. It is based on the desire-arousal-orgasm model of the sexual response and requires a phase-specific diagnosis. Late-phase dysfunction is quite effective if there is no anatomical interdiction. Early-phase dysfunction is more difficult to treat and benefits from a specialized referral.

Cancer Survivors

There are more than 14 million Americans living with a history of cancer, according to the ACS. Almost two-thirds of cancer survivors were diagnosed 5 or more years ago, and nearly half of all cancer survivors are older than 70 years. The rising number of cancer survivors is a tribute to the human intellect and resiliency; however, it is important to recognize that successful treatment does not come without times of distress and often long-lasting problems. In fact, there is a movement to consider oneself a cancer survivor from the day of diagnosis onwards.

According to the recent estimates from the ACS, the most frequent sites for invasive primary cancer among survivors were

- breast cancer among women (22%)
- prostate cancer among men (20%)
- and colorectal cancer (9%)

More than half of cancer survivors had a history of one of these cancers.

Among male cancer survivors, the three leading types of cancer are

- prostate cancer (43%)
- colorectal cancer (9%)
- melanoma (8%)

Among female cancer survivors, the three leading types of cancer are

- breast cancer (41%)
- cancer of the uterus (excluding the cervix, 8%)
- colorectal cancer (8%)

Survivors of childhood cancer who are under age 20 make up a small fraction of all cancer survivors (1%). The survival rates are higher for white patients and lowest for African Americans, Native American/Alaska Native, and Hispanics. Cultural factors, including language, values, traditions, and trust in providers, can influence underlying risk factors, as can access to treatment, health behaviors, beliefs about illness, and approaches to medical care.

In early self-report studies, survivors were believed to be free of psychological residue because many of them used denial and emphasized positive beliefs, such as having learned deeper values and being grateful for life. However, studies of leukemia/lymphoma survivors and breast cancer patients, using the Brief Psychiatric Symptom Inventory, showed a level of psychiatric symptomatology one standard deviation greater than the norm, or halfway between normal controls and psychiatric patients. Many subsequent studies confirm this association. Data from an American cohort of 9,535 long-term survivors of childhood cancer showed a 43.6% incidence of self-reported impairment in one or more of six domains: general health, mental health, functional impairment, activity limitations, pain resulting from cancer treatment, and anxiety or fears directly related to cancer and cancer treatment. Mental health concerns were the most frequently mentioned, and risk factors included female gender, low educational attainment, and low household income. A Dutch study found lower educational achievement, a lower rate of employment, a lower rate of marriage, and a higher incidence of men living with their family of origin as compared with a healthy control group. A history of a brain tumor or cranial irradiation was the strongest independent predictor but did not account for all of the variation. A Danish study confirmed this finding; although long-term survivors had more psychological symptoms, they did not have an increased rate of psychiatric hospitalizations unless they had had brain tumors or brain irradiation.

Delayed Consequences of Treatment

Four categories of delayed consequences of treatment have major psychological impacts. *Infertility problems, premature ovarian failure, and interference with sexual response* are particularly distressing if patients have not been educated and prepared to face them. However, there is evidence of persistent patient denial even in the face of preparation, such as sperm banking. Intervention should include endocrinologic, gynecologic, or urologic evaluation and treatment. Psychotherapy, couples therapy, group therapy, veteran patient programs, and self-help programs can all be used to good effect.

Pituitary, thyroid, and other endocrine failures occur as a consequence of radiation therapy. Patients with these problems often present with psychiatric symptoms. Patients treated before or during puberty have short stature due to a lack of growth hormone. Chronic fatigue is a common symptom that calls for patient and family education because it is often given an unjustified psychological label.

Increased vulnerability to late organ failure and secondary malignancies and to the chronic anxiety that this vulnerability generates is another challenge for survivors. If such a complication comes to pass, it finds the survivor less able to cope and more prone to a posttraumatic response.

Social rejection, employment discrimination, and serious insurance problems may be the lot of survivors in a frightened, suspicious world. Survivors who are well enough to return to work may plateau in their careers, being passed over for advancement by employers who fear a recurrence. Similarly, the survivors may be averse to moving due to the risk of being uninsurable. Despite the Americans with Disabilities Act (ADA), which bars employers from firing patients, there are cases of summary dismissal, and the appeal process is complex and uncertain. However, there are also cases in which sick employees are given extensive sick leave, warm support, and reassurance about their position. Also, the stigma of cancer is slowly diminishing as there are more examples of public figures undergoing treatment and continuing to lead productive lives.

Predictors of Poor Coping. Predictors of poor coping include previous psychiatric problems, lack of social support, low educational and economic status, and stage of disease at diagnosis. In treating the cancer survivors psychotherapeutically, one should strive for a detailed understanding of the patient’s experience. When the content is anxiety-provoking to the therapist, it is sometimes overlooked, leading to poorer outcomes even in the hands of seasoned therapists. Full knowledge is required to understand where the trauma response is primary and where it is a shield for deeper psychodynamic issues. Lingering too sympathetically on the trauma may deprive the patient of the opportunity to grow beyond it. Yet, failing to be sufficiently empathic in understanding its impact is obviously unproductive. There are many survivor programs that offer workshops, lectures, and groups for patients, couples, and families. The National Coalition of Cancer Survivors (NCCS) is emblematic of the possibilities for self-awareness and psychological and political empowerment.

Patients at Genetic Risk for Cancer

The advent of deoxyribonucleic acid (DNA) technology and sophisticated epidemiologic studies has permitted the identification of many healthy individuals at high genetic risk for developing cancer. Sources of genetic information include information from the patient’s family medical history, physical examination, and medical records. There is a wide availability of information for the public on the Internet, which can cause concerns among patients and may lead them to consult with their doctors. Once patients are identified as being at risk of developing cancer, they should be referred for genetic counseling. After an individual is found to have an increased risk of having cancer, there are potential interventions with the goal of reducing the risk, such as having a colonoscopy to detect colon cancer. Information about familial cancer has psychological effects. High-risk individuals have chronic anxiety, lowered self-esteem, and problems surrounding marriage and childbearing. Self-estimates of risk are usually much higher than their real risk, even for those in the highest risk bracket. Previously unsuspecting carriers do show increased distress, whereas decliners of testing remain anxious throughout. High risk and high anxiety often go hand in hand with denial and poor health maintenance behaviors, such as avoiding breast self-examination and surveillance testing. A group therapy intervention with educational and expressive components has been effective in reducing psychological distress, normalizing risk estimation, and improving health behaviors. Individual educational and supportive interventions have shown promise. Some form of psychological intervention should be a regular part of the surveillance programs that have been developed in many centers. Similarly, psychological support is crucial to deciding whether to enter surveillance trials, to have a prophylactic mastectomy, and, above all, to have testing. One study has shown that although only 71% of high-risk women would undergo testing for breast cancer, as many as 42% of average-risk women wished to undergo it as well. The consequences of testing go far beyond the psychological distress of the individual. Cooperation of unwilling family members may be required, and unwelcome knowledge may be forced on them. Marital, sibling and parent–child relations may suffer, at which point family-focused interventions come into play.

PSYCHOLOGICAL ASPECTS OF CANCER SCREENING

Screening for genetic risk is not the only area in which scientific and technologic advances have created ambiguous situations.

There are now many tumor markers, molecules associated with cancer, that can be used to

1. screen for the presence of cancer,
2. diagnose specific cancers,
3. determine prognosis in a given patient, and
4. monitor the course of the disease.

No test meets all requirements, but there is a tendency to overuse them. PSA for men is the most reliable. It is specific to the prostate, but an elevated level can signal prostatitis as well as cancer. It is extensively used both for screening and for monitoring prostate cancer and has resulted in many early diagnoses for a type of cancer that is found at autopsy in a large percentage of older adult men in whom it never grew to cause problems. The new patients are faced with difficult choices between watchful waiting, which some cannot tolerate, and treatment, which usually consists of either radiation or surgery preceded or followed by antiandrogenic hormonal therapy. All of these have significant and emotionally loaded side effects, such as impotence, incontinence, feminization, or depression. Cancer antigen 125 (CA125) is useful for following a known ovarian cancer but is not specific enough to be used for diagnosis, although it is often drawn for that purpose. Carcinoembryonic antigen (CEA) is even less specific for colon cancer.

Despite the potential for false positives, a recent study found that patients did not regret being tested. The author concluded that this enthusiasm would lead to the overuse of other screening methods, especially powerful, expensive techniques such as CT scans, MRIs, and positron emission tomography (PET) scans unless physicians have a higher threshold before ordering them without clear-cut criteria.

These patients may request counseling to cope with the many ramifications of test results and to help them decide on treatment. Asymptomatic patients who are told that they have cancer have an added burden when they undertake an arduous treatment. A part of them is never quite sure that there was really anything wrong with them. They continue to experience their illness as due to the treatment, and the more fragile patients with a tendency to denial, projection, mistrust, or paranoia may begin to doubt the reality of the cancer and the integrity of their physicians.

ETHICAL ISSUES

Psychiatry–Ethics Interface

In today’s contentious social climate, ethical and legal dilemmas multiply with increasing stress, often resulting in psychiatric consultation. If inappropriate psychiatric labeling has occurred, psychiatrists must remove the label, redefine the situation, bring out the underlying ethical issues, and set a decision-making process in motion. To be effective, they must be clear about the difference between psychiatric issues on the one hand and ethical and legal issues on the other hand. They must diagnose and act decisively in the former, be restrained and humble in the latter, and educate others about the distinction. If they understand the ethical issues, have a precise knowledge of the legal constraints operating in the case, and call for appropriate consultation, their interviewing skills and experience in conflict reduction can make them valuable participants. Selected ethical issues in the care of cancer patients are discussed in the following sections, with emphasis on their psychological correlates.

Truth-Telling

At the extremes, cultures that put a high value on autonomy consider the need for truth-telling as self-evident, whereas cultures in which the patient is tightly embedded in a wider family unit consider truth-telling as an exercise in maleficence. The spreading breakdown in close kinship ties and the steady improvement of modern cancer treatments are combining to increase truth-telling in growing segments of the world. However, local cultural deviance continues to arouse strong objections to it and may generate the need for outside consultation for cultural and religious guidance.

Informed Consent

In the United States, informed consent is legally required for conventional and experimental treatment. Given the toxicity and noncurative nature of much cancer treatment, this does not seem unreasonable and is spreading to other Western countries. However, it comes at some psychological cost. Anxiety reactions and occasional psychiatric decompensations are seen when patients feel overburdened by demands to make decisions. These often occur in predisposed individuals, even those with a good premorbid adjustment. Patients respond to active interventions, including support, medication, behavioral techniques, and active environmental manipulation of home and treatment settings. These interventions are usually sufficient to help patients make decisions on their own behalf and to undergo treatment, but posttreatment relapses may occur and are resistant to therapy.

Investigational Treatment

Investigational treatment is common in cancer. Most patients in academic centers will probably be considered for eligibility in an investigational protocol. Since the Nuremberg trials and the Helsinki Declaration, the moral requirement for informed consent from research subjects has been clearly spelled out and has been buttressed by regulations and institutions in the United States. Many critics feel it is honored in letter more than in spirit, but competent patients are known to systematically suppress information given to them to create the illusion of an old-fashioned paternalistic relationship; hence, the situation remains ambiguous, despite the efforts of lawyers, bioethicists, and well-intentioned physicians to tame it.

Randomized Clinical Trials. Critics of clinical trials claim randomization is incompatible with the doctor’s basic responsibility to the patient. Overall, patient reactions are determined less by the protocol than by their relationship to the physician. They have an irreducible need to believe that their physician thinks that the study is good for them. This is not completely false because no treatment reaches human trials if it has not shown some promise, and a physician-researcher is morally obligated and scientifically well advised to remove a patient from a protocol that will harm them. However, career pressures and local power relationships may mold physician behaviors, so institutional ethics are just as important as physician ethics but are much more complex and difficult to change. A good institutional review board and transparency in the presence of large medical teams are additional safeguards.

Some patients find randomization intolerable and refuse to participate or later demand to break the code. Most patients find trials with placebo versus treatment design emotionally more difficult to understand and tolerate than one with two treatment arms. Many patients seek experimental treatment with desperate intensity, peddling influence to enter limited protocols, or pleading to use untested agents that have not shown enough efficacy to warrant clinical trials. A few patients do participate in clinical trials out of strong altruistic motives, but, most of the time, these are secondary to the hope of personal benefit.

Alternatives to randomized clinical trials (RCTs) have been suggested, but they are unwieldy and less powerful than the trials. The results of large, multicenter RCTs are supposed to establish the standard of treatment, but there are inevitable intercenter differences even in the absence of any misconduct. In the case of the mastectomy-lumpectomy controversy of the 1970s and 1980s, RCTs from several countries gave consistent results that did indeed change standard medical practice. However, in the case of combined hormone replacement therapy (HRT), there have been contradictory trials, and only recently has there been a preponderance of results to establish that HRT not only lacks the cardiac and osteoblastic benefits that had been believed to offset their moderate but well-established increased breast cancer risk but also, in fact, comes at a higher risk for all of them.

PSYCHOLOGICAL PROBLEMS OF FAMILIES

Research on the family of the cancer patient is an active and exciting field, especially due to the integration of family therapy concepts. Some findings are disease-specific, but many of them can serve as a paradigm for understanding the impact of any major illness on the family.

Psychological Reciprocity

Awareness of the powerful reciprocity of emotional states and behaviors between a cancer patient and family members is well established from numerous studies of parents, spouses, and siblings that document levels of distress and functional impairment that can rival or even exceed that of the patient. More disturbingly, primary caregivers have been noted to worsen over time in the face of stabilization or improvement in the patient. Finally, studies of families after bereavement or cancer survival show a significant incidence of impaired functioning over extended periods of time and, again, a worsening over time in the distressed families. These findings have some urgency, given that most families experience cancer at some time and that it will probably be a drawn-out, chronic ordeal. High-risk families can be identified early and can be significantly helped by professional intervention.

Family Evaluation

Families can be evaluated from several points of view. The most traditional one is to examine how well the family is taking care of the patient. Medical staffs are quite sensitive to this and often call for help when families fail in this area. Another approach is to scan for psychopathology in family members. If the family member is acutely symptomatic and visible to the treatment team, a consultation follows, but absent members go untreated.

Areas for which the family is responsible include:

1. *Providing emotional support and containment*. Family members are expected and expect themselves to contain their own reactions and support the patient.
2. *Sharing responsibility for decision-making*. Family members often carry the lion’s share in researching and selecting options, with or without patient approval. This goes hand in hand with an enormous and universally reported *need for information*.
3. *Concrete caretaking*. In today’s budget-conscious climate and vanishingly short hospital stays, an ever-increasing amount of

care is being thrown onto the family. This is not acknowledged; hence, families receive little help or training.

4. *Meeting financial costs*. Even under socialized medicine, families absorb massive expenses and suffer significant downward mobility.
5. *Absorbing social costs*. In one study, 25% to 30% of primary caregivers lost their jobs, and more than half of the remaining primary caregivers moved to lesser-paying jobs to accommodate their need for flexibility.
6. *Maintaining stability*. The patient's role and contribution must be replaced, a difficult task in the case of young mothers or breadwinners. The new demands must be folded into ongoing functioning. Some members develop increased emotional needs that must be met. All of the standard family functions, such as emotional nurturing, feeding, clothing, sheltering, educating, and socializing, must continue unabated.
7. *Adapting to change*. It has been said that change may be the only constant during a cancer illness. However, it is useful to divide events into acute and chronic phases:
    a. *Characteristics of the acute phases*: An acute event is met with effective mobilization of family members but psychologically, they may be reacting quite asynchronously, with the denial of some and the pessimism of others causing significant friction. Members may protect themselves, under the guise of protecting each other, by colluding in a conspiracy of silence, which can rigidify and can be hard to change later. There is a universal need for information to decrease the uncertainty, to help the family problem-solve effectively, and to serve as a conduit for ensuring better communication, cohesion, and emotional containment.
    b. *Characteristics of the chronic phases*: Now, the family must return to developmental and maintenance tasks, which may expose conflicts between the patients' continuing needs and the needs of others. Initial mobilization is replaced with a paradoxical increase in psychological symptoms. Families isolate themselves, even when outside support is available, and may go into neutral gear, putting major decisions on hold.

Family-Centered Approaches

New and powerful methods of evaluation use family-centered approaches to examine how the family sustains itself and copes with the illness while still meeting the needs of all its members. Four concepts are important to understanding the impact of the illness and to devising interventions that are not instantly neutralized for the sake of family homeostasis.

Developmental Stages. Families go through developmental stages with specific tasks at each stage. Cancer disrupts the tasks at any point, sometimes permanently, if no help is offered. Any intervention must be compatible with helping the family carry out its phase-specific tasks.

Family Structure. One must understand how a family is organized to understand the caregiving dynamics and to intervene in ways that reinforce the family rather than threaten and destabilize it.

Family History, Beliefs, and Myths. Every family is permeated by stories that dictate unconscious responses to illness, dependency, loss, and other fundamental life experiences. They underlie behaviors that seem idiosyncratic or bizarre to untrained observers and must be factored into treatment. Explaining them to the treatment team minimizes scapegoating and polarization.

Family Style of Interaction. In past studies, when family researchers were studying family functioning outside the medical setting, good communication among family members was found to be the best predictor of outcome across several theoretical models. Other factors that were studied include cohesiveness, expressiveness, conflict, hostility, distance versus enmeshment, and adaptability, to name the most common. Cohesiveness in earlier models was thought to have a bimodal distribution, with too much cohesiveness having a negative impact on the development of individual members. Yet, that was not found to be the case when high cohesiveness was culturally determined. The study of families undergoing severe illness or death of a member has found cohesiveness to be an even better predictor of good functioning than communication.

Communication in the medical setting has very specific content and functions. The truth-telling debates bear witness to the fact that medical information is not viewed as an unalloyed blessing. Perhaps cohesive families can put medical information to better use and become even closer as they process it, whereas it may further stress and disrupt noncohesive families. Cohesiveness has also been found to be the common factor in families that show strain during crises but heal themselves over time.

Nontraditional Families. Divorced or blended families, single-parent families, homosexual families, and families that are ethnically, culturally, or religiously discrepant from the ambient medical culture experience many of the problems described previously but also have specific characteristics that are beginning to be systematically recognized.

Staff–Family Interface

During a life-threatening illness like cancer, the medical staff plays an enormous role in the family dynamics, invading the family system to create a new ad hoc hybrid system. That hybrid is the milieu in which consultation requests are generated. It must be understood and used for any intervention to be effective.

Guidelines for Intervention

Although most families show generosity, courage, and resiliency, as many as 25% to 30% would clearly benefit from intervention. Of those, a much smaller number requires formal family therapy, which must be done with a realistic understanding of the medical facts and the medical milieu. Most families manage with short-term crisis interventions designed with an accurate understanding of the family factors described previously and a commitment to the family, not just to selected members. In family-focused grief therapy, the patient and family meet to process the events and anticipate the upcoming loss. It continues for a few sessions after the patient's death, helping the family to consolidate the positive achievements made while the patient was still with them. The staff–family conference deserves special mention. It has the potential for being a powerful mutative event and has been known to resolve what had seemed to be knotty problems.

PSYCHOLOGICAL ISSUES FOR STAFF

Anatomy of Stress

The stress on caregivers has been documented in studies across many settings and many countries. Communication problems between doctors and patients lead not only to patient dissatisfaction but also to lower job satisfaction and self-esteem for the physician. The following stressors are more particular to the oncology setting:

1. *The negative aura* that still surrounds it.
2. *The toxic nature of treatment* and its inevitable side effects.
3. *The occupational hazards* surrounding chemotherapeutic and radioactive agents and the resistant infections to which immunocompromised patients are prone.
4. *The ambiguity of many treatment decisions*, which often breeds conflicts between staff members.
5. *The unsettled social and ethical climate* that surrounds palliative care and terminal care decisions.
6. *The extreme emotional reactions* engendered by cancer and projected onto staff in the form of overvaluation, ingratiation, magical expectation, and covert bargaining, followed by disillusionment and anger.
7. *Conscious and unconscious responses of staff to cancer*. At a conscious level, caregivers depend on their intellectual grasp of the situation and make a point of rejecting myth and superstition. They are proud of their ability to function under pressure with reliable emotional controls and to offer something in all situations,

even hopeless ones. Yet, at an unconscious level, there are covert adaptations that are idiosyncratic and irrational, often determined by historical precedent and resistant to change. They may include superstitious beliefs and stubborn fears and feelings. Their eruption into awareness makes the worker feel embarrassed and inadequate. Hence, the feelings are kept secret; the caregivers may believe that they are alone in having them. Thus, staff workers isolate themselves from one another.

8. *Social isolation*. Outside of the workplace, staff isolation is reinforced by family, friends, and a society that shuns reminders of a frightening topic. Caregivers are not singled out personally, but they may be discouraged from talking about their work and risk being shunned if they do not accept this taboo.

Clinical Spectrum of Staff Responses

Normal Adaptation and Coping. Staff undergo a developmental process when they begin to work intensively with cancer patients. A high initial level of dysphoria, anxiety, sadness and numbing recedes over the first few months, displaced by the need to demonstrate competence and the ability to survive. Over the next few months, existential issues are allowed to surface, and staff members shape a deeper adaptation that engages their whole personality. Formal support is not necessary, but informal support and peer group interactions are important to a successful resolution.

Adjustment Reaction. Adjustment reactions may occur during the formative stages of providing cancer care, but they also overtake staff members at any point in their careers when a personal loss resonates with their work (cancer in self or loved ones, in particular). These reactions are acute but time-limited and respond well to support or a brief change of venue. Rarely is psychiatric intervention needed. Sometimes adjustment does not improve with time, exposing an underlying psychiatric disorder, often pre-existing and precipitated by work stresses. Psychiatric evaluation is then indicated and may result in transfer to a less emotionally stressful situation.

Supportive Interventions for Staff

Staff support interventions generally fall into three categories: Organizational, individual, and group. Organizational issues include leaders who model realistic expectations and monitor overinvolvement in themselves, as well as others. A team approach encourages the sharing of difficult tasks, prevents a sense of being indispensable, encourages staff cohesion, and eases time off and vacations. Open communication and good orientation procedures are important. Individual responses include peer support, supervisor support, access to an employee assistance program, and other confidential psychological support or psychiatric evaluation. There are many published reports about the benefits of staff support groups in many forms and many settings, but they are anecdotal. These groups are difficult to run, but when they work, participants value them. The group must not be perceived as undermining unit norms or leadership. The structure must be provided, as needed, to alleviate staff anxiety while still allowing unscheduled topics and feelings to be aired. Therapy issues must be avoided, and discussions should clearly flow from work-based topics. An active area of new and productive intervention is the development of group-based workshop formats for improving communication. Many have been described and found to be helpful; more are under development.

CANCER IN THE PSYCHOTHERAPIST

Unlike degenerative diseases, with their linear course and associated impairments, cancer may have an episodic course with no loss of cognitive functioning. It inevitably forces a blurring of psychological and, later, overt boundaries in therapy. An all-but-intolerable burden is placed on doctors and patients when therapists who have cancer continue treatment too long because of their own need to deny their fate. This suggests that, difficult as it is, the therapist's most ethical course of action is to set a fixed termination date that allows the patients to transfer to someone else while the therapist is still alive and well enough to negotiate the termination. This may be asking too much of an individual in whom denial is both the main coping mechanism and the only impairment. The help of colleagues is invaluable at this time.

PEDIATRIC PSYCHO-ONCOLOGY

Magnitude of the Problem

The new cases of cancer in children under 15 years old continue to increase. Leukemias are the most common cancer in children, accounting for 30% of all new cases. Brain and CNS tumors are the second most common, making up 26% of childhood cancer cases. Caucasian children are more likely to develop cancer than any other ethnic group. The combined 5-year survival rates for all childhood cancers improved from 56% in the 1970s to 80% currently.

Child Patients

Sick children are benefiting from social and legal trends toward empowering them. They are being informed more honestly, and their cooperation is being thoughtfully enlisted. It has become evident that, like adults, children know or intuit a great deal without being told and are relieved by open communication. Its provision is complicated by the need to be age-appropriate and is more painful because of adult vulnerability to suffering in the young. As a result, adult denial often prevails, and children respond by increased behavioral symptoms or withdrawal.

Discussion about Diagnosis and Prognosis

A child can hardly remain oblivious to their diagnosis, given the endless clinic visits, repeated chemotherapies, flagrant symptoms such as hair loss, repeated exposure to other patients, and statements of people around them. Studies note that immediate perception outweighs distant reality; hence, intrusive procedures can loom larger than prognosis, especially in younger children. The developmental stages of death awareness in healthy children are well described but are probably not applicable to children with cancer or to other children who have been exposed to death in an untimely and personally threatening way that results in earlier and keener awareness. Thus, discussions are clearly indicated, but they must follow the child's lead, looking for and responding to their sources of anxiety and remaining reassuring about care and comfort if they cannot be so about the outcome. It has been found that children who received truthful information had less subsequent anxiety than those who did not. However, parents often request and benefit from psychological guidance about how to proceed because truth-telling to young children is strongly molded by the child's circumstances and level of possible understanding. This is not always obvious to even the most loving parents. But they are usually eager to learn and very responsive to guidance.

Psychiatric Diagnosis and Treatment

In approaching the psychiatric diagnosis of a child, it is important to know the child's previous developmental profile and to assess current symptomatology, arrests, or regressions. Pre-existing disorders are aggravated by cancer and must be monitored and treated throughout. The stress of cancer can make a mild disorder more severe.

Anxiety, depression, and delirium occur routinely in children and respond well to treatment. Behavioral interventions are the first line of treatment for anxiety, but some situations may require the use of anxiolytic medications. Benzodiazepines are sometimes used but can cause a disinhibition syndrome that may be mistaken for more anxiety. Mild depressive symptoms are common and often complicated by demoralization. After paroxetine was reported to predispose children to suicide, warnings were issued against the use of SSRIs in children. But experts point out that the evidence is not strong enough to warrant a categorical prohibition and suggest cautious use with close monitoring. Desipramine (Norpramin) is contraindicated owing to several reports of sudden death.

Children suffer from delirium during critical illness, similar to adults. Pediatric delirium occurs in at least 30% of critically ill children. As with adults, children with delirium have worse clinical outcomes than those unaffected by delirium, including days admitted to the pediatric intense care unit (PICU), memory deficits, and longer school absences. Haloperidol has been used for behavioral symptoms associated with delirium in children, such as agitation. There are no randomized placebo-controlled trials of atypical antipsychotics in children with delirium; however, risperidone (Risperdal), olanzapine (Zyprexa), and quetiapine (Seroquel) appear to be effective tools for safe management of behavioral symptoms during delirium. Pediatric psycho-oncologists should be quicker to use medications for behavioral symptoms than child psychiatrists. Starting with small doses, using the lowest effective dose for the shortest possible time, and diligent monitoring for side effects are advised. Further studies in this area are certainly needed.

Pain Management

Children are at high risk of receiving insufficient pain management. Contributing factors include misplaced beliefs about childhood pain perception, inappropriate use of as-needed dosing, unclear communication of pain experience from young children, and complaints made in the context of normal regression, which is sometimes used to disqualify the complaints. Exhausted, critically ill children, given relief of suffering, may enjoy opiates in a way that upsets some parents and caretakers, leading to lowering of dosing.

Recent findings indicate that the experience of pain in early life may lead to long-term behavioral and psychological consequences. It has also been found that chronic pain, including neuropathic pain, is far more common in children than was previously believed. These findings still wait to be extensively disseminated and translated into proactive and even pre-emptive pain treatment in infants and children.

Management of Stressful and Painful Procedures. Recent work on PTSD in children reinforces the principle that aggressive prophylactic relief during stressful procedures is important not only for current comfort but also for long-term adaptation. Poorly managed painful procedures not only brutalize the child but also result in anticipatory fear of pain, loss of trust and cooperativeness, phobias, and posttraumatic symptoms that do not improve over time. Unlike adults, children are anesthetized for bone marrow biopsies. Multimodal behavioral methods may be taught to parents, which can be extremely helpful during procedures.

Family Involvement

The families of pediatric patients are even more vulnerable than those of adults because of the child’s total dependency and the biologic, emotional, and social factors that make families uniquely responsible. No intervention has any hope of success if it does not address the family as much as the child. In one study, mothers had a greater rate of posttraumatic symptoms than their children.

Childhood Cancer Survivors

Child survivors face daunting tasks. They must move beyond lengthy hospital stays or repeated day hospital visits, painful procedures, and severe side effects. They may have endured separation from their family, incurred resentment from their siblings, and missed as much as a year of school. They may emerge with physical handicaps or disfigurements, but the most insidious problems are the cognitive and verbal deficits associated with chemotherapy and radiation therapy. These may be compounded by drug-induced hearing losses. Children need to be monitored and given remedial help promptly.

It has been noted that the child’s own perception of risk and the intensity of cancer treatment is more predictive of PTS than the objective ratings of health professionals. Hence, it is useful to monitor perceived health status. There is a role for psychological evaluation and support throughout the illness as well as during the post-illness phase.

Children and their families may also feel a sense of victory and relief that helps the children's self-esteem. Nonetheless, it is still important for families to be supported and queried for difficulties when they return for the follow-up appointments, as they have been found to function less effectively than nonaffected families. Contacting the child's school counselor and teachers is an effective intervention because the latter are eager to be educated about special needs and reassured about how to help the child. The ability to reintegrate comfortably into school life and school friendships is a major factor in the child's adjustment.

Consent in Pediatrics

All pediatric consents are third-party consents and carry the possibility of conflicts between parent and child. The parent may make a decision that is not in the interests of the child, as in the refusal of life-saving treatment. The courts have tended to overrule parental authority in such cases. Parent and child may disagree, especially adolescents, although disagreements occur with school-age children as well, presenting staff with painful problems. Psychiatric intervention may help resolve the conflict if the consultant does not impose personal beliefs. Legal recourse is used only after all attempts at conflict resolution have been unsuccessful.

Genetic Testing of Minors

Decisions about whether to offer genetic testing or screening should be determined in the best interest of the child. Genetic counselors discourage the use of genetic testing to abate parental curiosity and anxiety.

Staff Problems

The maintenance of staff morale, the prevention of dysfunctional norms, the support of faltering members, and the defusing of unconscious competitive feelings toward parents are important ways of contributing to the well-being of all participants. The need for staff members to acknowledge their grief with the repeated loss of children whom they came to know well has given impetus to the custom of memorial services for one pediatric oncology ward, where the staff members gather every few months to remember each child that died in the interim. There is also more attention being paid to families during the mourning process, with outreach offers of support and provision of support groups. Staff members may derive indirect benefit from this greater acknowledgment.

Team Approach

The multidisciplinary approach to patient care is most widely implemented in pediatric cancer patients. Psychiatrists who specialize in child psychology can play a useful role and, where the model is not in place, are in a good position to help organize it. Local teachers and clergy should be involved, and community resources should be used to help families reintegrate when they return home.

GERIATRIC PSYCHO-ONCOLOGY

The population of older people is growing rapidly throughout the world. Currently, there are about 35 million people in the United States over the age of 65. That number is expected to double by 2030 and comprise 20% of the population, according to the U.S. Census Bureau. Cancer is a leading cause of morbidity and mortality in the aging population, with approximately 60% of all cancers and 70% of cancer mortality occurring in people aged 65 years or older. Of the 24 million cancer survivors worldwide, 61% are 65 or older. Psychiatric care of older persons with cancer presents a unique challenge for clinicians due to differences in phenomenology, assessment, and treatment in this population. Differences in psychological sequelae or specific psychological needs of older cancer survivors have not been well studied, despite their growing numbers. Capacity assessments, informed consent, advanced directives, and surrogate decision-making are commonly encountered issues in caring for older cancer patients and should be handled based on the specific needs of the older adults.

Several cross-sectional studies have found lower or similar rates of psychological distress in older adult cancer patients compared to the younger adult population. Longitudinal studies either found no difference or higher psychological distress in younger adults. It is important to note that older adult cancer patients are at risk for underdetection and undertreatment of their psychiatric diagnoses due to decreased self-report of psychiatric symptoms, cognitive deficits, comorbid medical conditions, poor access to mental health care, and differences in phenomenology. In palliative care settings, there is a need to appreciate the individual differences in physical and psychosocial needs in older adults. The older adult with cancer at the end of their lives are at greater risk for untreated pain and depression and are more likely to commit suicide compared to their younger counterparts.

The goal of identifying and treating these distressing problems is to improve the overall quality of life and the functional status of older people living with cancer. This section will review the assessment and management of common psychiatric conditions in older adult patients with cancer.

Depression in Older Adult Cancer Patients

The prevalence of depression is 3% in older adult people living in the community and 10% to 15% for those living in nursing homes. Estimates of depression in cancer patients vary widely, with averages between 17% and 25%. Although many early studies of depression in cancer patients included older patients, there was not often a description of findings based on demographic data such as age. Several studies have reported that older patients with cancer have lower rates of depressive disorders or self-reported psychological distress than younger patients. Risk factors for depression in older adult cancer patients include loss of a spouse, functional disability, inadequate social support, uncontrolled pain, poor physical condition, advanced illness, previous history of depression, other life stresses or losses, family history of depression or suicide, and medications known to cause depression. A study of 420 older adult cancer patients showed that higher levels of symptom severity, lower levels of prior physical functioning, and greater physical functioning deficits predicted higher levels of depressive symptomatology. The prevalence rate for depression was 18%. Cancer site, gender, stage of disease, and treatment type had no significant effect in predicting depressive symptomatology. In a longitudinal analysis of the course of depressive symptomatology in older adult breast, prostate, colon, and lung patients, the prevalence of depression was 26.1% 2 to 4 weeks after diagnosis, compared to 15.8% 1 year after the diagnosis. Lung cancer patients had higher rates of depression, likely due to higher symptom severity. In a group of older adult colorectal cancer patients, a longitudinal study examining depressive symptomatology over the first year following cancer diagnosis has shown that gender, race, comorbid conditions, physical functioning, social functioning, and symptoms were significant predictors of depressive symptomatology. Female patients, African Americans, and patients with two or more comorbid conditions exhibited more depressive symptomatology. The highest prevalence rates for depression were at the time of initial assessment, 18%. A prospective study in head and neck cancer patients comparing older adult and younger cancer patients has shown that 28% of younger and 30% of older adult patients had depression. Both groups of patients had higher rates of depression 1 year after diagnosis. There was no significant difference between younger and older adult cancer patients regarding the prevalence of depressive symptoms.

Assessing depressed mood in older adult cancer patients is difficult. Neurovegetative symptoms of depression such as poor sleep, poor appetite, and decreased energy can be related to the cancer or the treatment; however, the older adult tend to have more somatic symptoms of depression than younger patients. Assessing the cognitive symptoms of depression such as hopelessness, worthlessness, guilt, and suicidal thoughts in addition to the somatic symptoms is important in geriatric cancer patients. Geriatric Depression Scale (GDS), the Center for Epidemiologic Studies Depression Scale (CES-D), and the HADS are commonly used in screening for depression in older adult patients. These scales need to be validated in older adult cancer patients. The NCCN Older Adult Oncology Guidelines have recommended the use of the Distress Thermometer to assess distress in cancer patients.

Differential diagnosis is an important part of depression assessment in older adult cancer patients. Although a cancer patient may respond with an appropriately sad mood in the face of initial diagnosis or setbacks such as recurrence, it is important to note that depression is not a normal part of aging or of having cancer. Late-onset first episode depression presents a diagnostic challenge to the clinician as some of these patients may be in the preclinical stages of Alzheimer disease or vascular dementia. Poor memory or impaired attention may be the initial presenting symptom of depression in older persons. Delirium should be ruled out in the presence of disturbance of level of alertness, attention deficits, and psychotic symptoms. Medical conditions should be considered as possible causes for depression.

Several psychotherapy techniques (i.e., interpersonal, supportive, cognitive-behavioral, group, psychodynamic psychotherapy, hypnotherapy, relaxation, and biofeedback) have been used effectively to help older adult cancer patients who are suffering from depression. The efficacy of antidepressants for the treatment of depression in cancer patients has been well established. There is also good evidence that depression in older adult patients with chronic medical illnesses can be treated successfully. Older adult patients should be started at lower doses of the medications, and the dose titration should be slow. SSRIs are usually the first line of pharmacologic treatment in depression. Studies in geriatric patients have found SSRIs to be equally effective as TCAs in the treatment of depression. Although TCAs have been the most extensively studied medications for the treatment of geriatric patients, they have been replaced by the newer antidepressants because of a preferable side-effect profile. In treating older adult cancer patients who have depression, it is important to carefully review medications for possible drug–drug interactions.

Of the newer antidepressants, bupropion can have a stimulant-like effect that may be helpful for older adult cancer patients with significant fatigue. Mirtazapine is associated with sedation and weight gain, which may be a beneficial effect in older adult cancer patients. Psychostimulants may be most helpful in older adult cancer patients where the depression is accompanied by psychomotor slowing, decreased energy, and apathy due to their rapid onset of action. ECT is a safe and effective treatment for depression in the older adult. However, it has not been specifically studied in older adult cancer patients.

Suicide in Older Adult Cancer Patients

Older persons account for 18% of all suicide deaths. The ratio of attempted to completed suicide is 4 to 1 among the older adult, compared to a ratio of 200 to 1 in a younger population. The incidence of suicide is also higher in cancer patients compared with the general population. Identification and prevention of suicide require the definition of risk factors. Physical illness is a common risk factor for suicide in late life. Older adult cancer patients, and individuals with head and neck, lung, breast, urogenital, GI cancers, and myelomas, seem to have an increased risk of suicide. It is important to recognize and aggressively treat high-risk patients for depression and address suicidal risk.

Anxiety in Older Adult Cancer Patients

Anxiety is not well studied in the geriatric cancer population. In patients with cancer, the prevalence of anxiety ranges from 21% to 28%. Anxiety may be a component of pain, delirium, or depression. The differential diagnosis of anxiety in older adult cancer patients can be difficult due to several factors contributing to anxiety.

Steroids, antiemetics such as prochlorperazine and metoclopramide, alcohol, or benzodiazepine withdrawal are associated with anxiety.

As in younger patients, older adult patients may benefit from meditation, guided imagery, progressive relaxation, and reframing of negative, irrational thoughts. Other psychotherapeutic techniques such as insight-oriented, supportive, and group psychotherapy may also help reduce anxiety. Benzodiazepines are generally useful for short-term relief of anxiety symptoms. The clinician should be cautious in the older adult population. Lorazepam, oxazepam (Serax), and temazepam (Restoril) are better tolerated by the older adult. However, with all benzodiazepines, side effects such as drowsiness, confusion, and motor incoordination must be monitored. One of the greatest morbidities is their potential to cause an imbalance in the older adult patient, leading to falls. Patients with dementia or brain injury may have a paradoxical reaction to benzodiazepines and become disinhibited, irritable, or agitated. Many of the antidepressants listed above are also useful in the treatment of anxiety disorders. Stimulants and stimulating antidepressants (i.e., bupropion) should be avoided in patients suffering from severe anxiety. Buspirone (BuSpar) is a nonbenzodiazepine antianxiety medication that can be a helpful anxiolytic in the older adult.

When patients cannot tolerate benzodiazepines or when their anxiety is not well controlled, the use of a low-dose atypical antipsychotic can be considered. Atypical antipsychotics do not have adverse effects on respiration as do benzodiazepines. None of the antipsychotic medications have been FDA approved for use in the treatment of anxiety disorders in the general adult or older adult population. It is also important to note that the FDA has recently issued a warning against the use of antipsychotics for behavioral disorders in older adult patients with dementia due to increased mortality from cardiac and cerebrovascular events.

Fatigue in Older Adult Cancer Patients

Fatigue is a common symptom in older cancer patients. A recent study of fatigue prevalence has found this symptom to be universal in the older adult age group. In 84% of the cases, patients' fatigue interfered with the general level of activity. Fatigue was the most reported symptom in a quality-of-life study among older adult cancer patients. Fatigue remained the most common symptom both at the time of diagnosis and 3 months after diagnosis.

In older adult patients, the common etiologies of cancer-related fatigue are immobility, pain, depression, sleep disorders, anemia, metabolic derangements, undernutrition, infections, side effects of cancer treatment, centrally acting drugs, and direct effects of the tumor. All of these conditions should be considered in the evaluation of fatigue. There are several fatigue assessment measures, such as the Brief Fatigue Inventory, the Multidimensional Fatigue Inventory, and Piper Fatigue Scale, none of which have been validated specifically for use in the older adult cancer population. The relationship between pain, fatigue, and depression is important. A statistically significant relationship was found between fatigue and depression, pain, number of medications, and physical function in a study of fatigue in the older adult. Another study has shown that 70% of patients with fatigue had psychiatric comorbidity. Due to the significance of this relationship, patients reporting fatigue should be assessed for pain and comorbid psychiatric conditions, particularly depression.

Treatment of fatigue in older adult cancer patients has not been well studied. The first approach should be to address any potentially identifiable etiologies, such as anemia, depression, insomnia, or any metabolic abnormalities. Exercise has been used as a treatment for fatigue in cancer patients receiving chemotherapy but has not been well studied as a treatment for fatigue in older adult cancer patients. Medications have been used to treat fatigue in cancer patients. Studies on the treatment of fatigue in older adult cancer patients have not been published. Psychostimulants such as methylphenidate or wakefulness-promoting agents such as modafinil have been used successfully for cancer-related fatigue. Modafinil may be a gentler, better-tolerated medication for the older adult.

Antidepressants with some stimulating effects such as bupropion or fluoxetine may be helpful to improve energy, especially when there are also concomitant symptoms of depression.

Delirium in Older Adult Cancer Patients

Delirium is a disturbance in the level of alertness, awareness, attention, cognition, or perception that develops over a short period of time (hours to days) and fluctuates during the course of the day. The older adult are particularly susceptible to the development of delirium. An estimated 30% to 50% of older adult patients develop delirium during medical hospitalization. Older adult patients who develop delirium during hospitalization have an estimated 22% to 76% risk of mortality during that admission. Major risk factors for delirium appear to be pre-existing cognitive impairment, advanced age, and severity of comorbid medical illness.

The etiology of delirium in older adult cancer patients is often multifactorial. Mini-Mental State Examination (MMSE) is a useful, commonly used screening tool for cognitive impairment, but it does not distinguish between delirium and dementia. The CAM, the DRS, and the MDAS are other tools that can be used in the assessment of delirium, as detailed above.

In the assessment of delirium in older adult cancer patients, one of the most challenging problems is to distinguish delirium from dementia. Further complicating the issue, patients with dementia can develop superimposed delirium. Patients with dementia appear relatively alert and have more stable cognitive deficits, as opposed to the fluctuating symptoms in delirium. The onset of symptoms in dementia is slower, with a chronically progressive decline in cognitive function rather than the relatively rapid onset of symptoms seen in delirium. The management of delirium in older adult cancer patients involves addressing the symptoms of the delirium while investigating and ultimately treating its underlying causes. The treatment of delirium in cancer patients, including the older adult, is extensively reviewed above.

Dementia in Older Adult Cancer Patients

There are an estimated 1.8 million older adult individuals in the United States who have advanced dementia requiring extensive care. DSM-5 classifies dementia syndromes under major neurocognitive disorders and allows for continued use of the term dementia. Patients with severe dementia often exhibit emotional and behavioral disturbances that make their care more difficult. Studies have shown that geriatric patients with cancer who are cognitively impaired have a worse performance status, higher incidence of depression, and are at greater risk for death. Alzheimer dementia is the most common type of dementia. It accounts for 60% to 70% of all patients with dementia. Another 15% to 25% of patients have vascular dementia, and approximately 10% to 15% will have mixed dementia. The remaining patients have other diagnoses such as Parkinson disease, dementia of Lewy body disease, and frontotemporal dementia. The recognition rates of cognitive impairment range from 25% to 90%, depending on the health care setting and the degree of impairment. Recognition of cognitive impairment in cancer patients is especially important due to issues of capacity and decision-making in older adults. A diagnosis of dementia does not automatically determine a lack of capacity. However, when cognitive impairment is identified, it should prompt further evaluation. Often a psychiatric evaluation will be necessary to determine the patient’s ability to participate in the decision-making process. When a patient is deemed not to have capacity, a health care proxy must be determined. Given the potentially poor quality of life for an older cancer patient with dementia, the extent of aggressive treatment should be carefully determined in a patient with severe cognitive deficits. Cognitive impairment may present special problems in the control of pain and other symptoms due to limitations in the patient’s self-report or active participation in assessment tools.

Dementia has also been noted to develop secondary to cancer treatment, including chemotherapy, radiation treatment, immunotherapy, hormonal therapy, and supportive medications. It is not well understood how chemotherapy or radiation therapy might affect cognition and whether older patients are at greater risk. A study of lung cancer patients has shown that the treatment of chemotherapy-related anemia could increase the cognitive capacity of older adult patients. A recent review of the literature concluded that there was not sufficient information on cancer treatment-related cognitive impairment in the older adult due to the small number of individuals studied, the variety of cancer types and treatments surveyed, the low number of longitudinal studies, and the limited battery of neuropsychological tests employed.

CANCER RISK AND SURVIVAL

The role of psychosocial factors in altering cancer risk and survival fascinates the public, but the discussion is often blighted by wishful thinking and genuine confusion. Cancer risk and survival are often conflated, and great emphasis is placed on the direct role of the individual's psychoneuroimmune mechanisms in preventing or curing cancer while ignoring many psychologically mediated effects that are far stronger and better established. There are many distorted beliefs and overinterpretations on the subject that psychiatrists need to know the current state of the art to better understand and educate their patients.

Cancer Biology

The natural history of a given cancer begins when changes in a cell's DNA establish the potential for this cell to become malignant. This *cancer initiation* precedes by many years (anywhere from 3 to 30 years) the second event, known as *cancer promotion*, in which growth suppression is lost and malignant multiplication begins. This again precedes by a period of months to years the appearance of clinical symptoms. Finally, there is a variable period between the appearance of any clinical symptoms and the seeking of care. Cancer risk can relate to the risk of initiation or the risk of promotion. Cancer survival is measured from the time of diagnosis, but this ignores the important factor of delayed diagnosis, for which initial disease staging is a partial but incomplete solution.

Psychosocial Factors

One attempt at clarifying the multifaceted impact of psychosocial factors on cancer is to think in terms of

1. The *external loop*, which refers to ways in which psychological, behavioral, and socioenvironmental factors affect the person's behavior and lifestyle and, hence, cancer risk. Even there, the actual behaviors must be teased out from their psychological determinants.
2. The *internal loop*, which refers to the ways in which the person's psychological status affects cancer risk and survival via changes in the internal milieu. This effect can be mediated in many ways, such as neurologic, immune, and endocrinologic mechanisms, individually and in varying relations to each other.

Psychologically Mediated Environmental Factors

Cancer Risk. By far, the largest role in the development of cancer is played by inescapable environmental insults, such as radiation, viruses, and chemicals, acting cumulatively over time on each individual's predetermined complement of potentially cancer-causing genes. The individual's exposure to injury is in part out of their control and in part mediated by social and psychological factors over which they have variable control. By far, the greatest volitionally determined contributors of the psyche to cancer incidence are personal lifestyle and habits. Smoking cessation, abstinence from alcohol, more exercise, and better eating habits would probably lead to a substantial decrease in the incidence of cancer in the United States.

Cancer Survival. The greatest contributor to decreased survival is a poor use of screening because early diagnosis is still the best predictor of prolonged survival for most cancer types. Many psychologically determined factors can play a role in delayed diagnosis. Patients with cognitive impairment, mental retardation, and psychotic disorders are especially at-risk populations for delayed presentation to health care professionals, which makes cancer screening efforts immensely important in those groups.

Research Review

Methodologic Considerations. To be reliable, the studies of the role of psychosocial variables should involve large populations over long periods of time. In commonly used case-control studies, the index population is usually unrepresentative of broader populations, and the matching is incomplete; hence, it is unlikely that one can generalize the findings to larger unselected populations unless the effects are so large as to be inescapable. In retrospective studies, the index population is equally unrepresentative, but these are essential pilot studies pointing the way to later larger ones. In prospective cohorts, researchers can control for the confounding variables, but the task remains methodologically daunting.

Current Findings. All this notwithstanding, more than 50 psychosocial factors have been claimed to play a role in cancer initiation or promotion. A wave of recent reviews has shown that few of them are valid, in many cases because the quality of the studies is poor, but also because the question is difficult to answer, given that cancer initiation begins 20 years before there is any clinical evidence. The lack of impact of bereavement has been demonstrated with recent, large, population-based studies. The role of external stress had been thought to be significant, but later studies have disproved this.

On the other hand, several recent studies have shown that social support is a significant positive survival factor, a fact that has important implications for mental health professionals. Social class has a variable effect on cancer risk, with some cancers labeled *cancers of poverty* and others labeled *cancers of affluence*. However, low social class has an unequivocal negative effect on cancer survival. One must take into account the fact that these individuals are less knowledgeable about and less able to avoid toxic exposures and unhealthy lifestyles and, more to the point, may lack access to decent, not to mention excellent, medical care.

Race is also a strong predictor of decreased survival, but the effect is much decreased or even disappears after correction for social class and shows great variability within racial groups. Two recent studies of race and ethnicity were published back to back, one showing significantly poorer results for nonwhite children with acute lymphoblastic leukemia, and the other showing that, given equal access to effective state-of-the-art therapy, children of both races could expect the same rate of cure. The experiment was possible because acute lymphoblastic leukemia has an acute onset and a well-defined course of treatment. Given access to the treatment, parents could be depended on to bring the child for life-saving treatment.

The effectiveness of interventions aimed at improving lifestyle and health behaviors has often been disappointing, usually because the complexity of the problem is underestimated or because unrealistically short-term results are being demanded, or both.

Internally Mediated Risk and Survival Factors

Recent developments in psychoneuroendocrinology or psychoneuroimmunology have confirmed the feedback loop between feelings and perceptions, on the one hand, and neurologic, endocrine, and immune changes, on the other. These are, in small part, an updating of classic psychosomatic concepts. More urgently, this is the area of research in which recent developments lend theoretical support to current popular beliefs about the power of the mind over the body.

Personality Traits, Feeling States, and Coping Styles. There are weak and conflicting results on the role of personality in cancer prevalence and survival beyond what is mediated through risk-avoidant behavior. Studies are handicapped by the absence of comprehensive prognostic indicator profiles. Meanwhile, no strong relationship has been established between personality style and disease outcome. Feeling states, such as level of distress or anxiety, have not been found to have any predictive value. A weakly significant association between passive coping styles and poor disease outcomes has been demonstrated in several studies using measures of the so-called fighting spirit, helplessness and hopelessness, stoic acceptance, and anxious preoccupation. The role of denial, which can be viewed as the operative factor, remains an open question and points again to the complexity of what is being measured.

Immune Mechanisms. Although crucial discoveries of the immune system describe several associations with mood and environmental changes, direct links to personality and coping styles have not been forthcoming. Importantly, epigenetic mechanisms postulate how early life stress and trauma change the physiologic reaction to stress later in life. These findings are relevant for patients with newly diagnosed or recurrent cancers who are experiencing repeated chronic stressors.

Endocrine Mechanisms. The HPA axis mediates several downstream effects based on negative and positive feedback loops. Mood states and chronic stress has been shown repeatedly to alter these feedback loop, thus resulting in higher levels of anti-inflammatory cortisol and other mediators. Unfortunately, stress hormones have long-term consequences that are also associated with stress and sequelae from psychiatric disorders. The dysregulation of hormonal feedback loops influences and is influenced by the environment, attitude, personality, and coping.

Effect of Psychological Interventions

Inasmuch as psychological interventions improve mood function and prevent or treat depression, there remains a possible survival benefit. It may not be directly from the intervention but from reversal depression which is associated with inferior survival based on numerous studies to date. Many psychological interventions have been studied and demonstrated level 1 evidence of efficacy in various cancer settings. These are most commonly cognitive-behavioral therapies, psychoeducation, expressive-supportive therapy, mindfulness-based stress reduction, and other mindfulness-based therapies.

ALTERNATIVE TREATMENTS

Complementary and alternative medicine (CAM) is often defined as any medical system, practice, or product that is not thought of as standard therapy for the treatment of cancer or cancer-related symptoms. Many of these therapies are incorporated into practice based on the field of Integrative Medicine. Of note, complementary treatments refer to adding a non-standard therapy, and alternative treatments replace the standard therapy. These interventions may be as essential as diet and understanding the roles of specific diets based on disease presentation. Patients and clinicians should be cautiously aware of the dangers of treatments that lack an evidence base and are used to replace standard treatment. The risk depends on the disease or symptom in question and the potential side-effect profile of the alternative agent if it is a medicine (e.g., herbal formulation). Of note, adherence to unproven therapies may reveal a distrust of the medical establishment, may feel consistent with the patient’s value system or may be influenced by societal or cultural factors. Unfortunately, the stigma of cancer propels people into magical thinking, which can have deleterious consequences. The toxicity of traditional treatments also makes patients so anxious that other options become compelling. The amount of out-of-pocket expenses spent on nontraditional treatment is in the billions. In one study, a large percentage of patients at a tertiary medical center was seeking alternative treatment while still in traditional treatment. They did so early in their course of treatment and took both simultaneously, dropping the traditional treatment before they dropped the alternative one. Educated and wealthy patients were best represented. The main treatments were physical ones to cleanse and to strengthen the body; special diets to purify and fortify the body, as well as avoiding harmful substances; immune therapies; faith healing; psychological methods to improve mental attitudes and to reinforce the body's defenses; and many Eastern modalities such as yoga, tai chi, and meditation that promote improved discipline and decreased anxiety.

Many such treatments give patients a sense of purpose and control. The last category, improving attitudes, is most troubling to the psychiatrist because it leads so readily to blaming the victim. Patients are told that they can influence their cancer by mental efforts. The unspoken corollary is that further disease will be due to inadequate efforts and, worse yet, that they caused the cancer in the first place. This plays directly on attribution mechanisms to which cancer patients are already universally prone. One can also see desperate family members grasping at these explanations and trapping the patient ever deeper in a web of guilt and responsibility as a means of coping with their own grief and frustration or an angry patient blaming a family member for causing the cancer.

The psychiatrist must remember that many patients are secretly involved in alternative care that makes them embarrassed and distant if they cannot share it. As time passes, if their condition worsens, the burden of failure and guilt makes a greater contribution to psychic distress. Hence, clinicians must encourage disclosure and must help patients extract something positive from the experience. Active cooperation from physicians in encouraging their patients to use the nonharmful modalities, especially the ones for which there is a consensus about the benefit, increases the patient's trust and adds to the placebo effect and synergy of all the patient's treatments. This is difficult to do if patients and families retreat into anger at the establishment and reduce cognitive dissonance by clinging desperately to their alternate beliefs.

Given how much cachet these therapies had achieved, recent surveys in breast cancer patients suggested that between 48% and 66% of patients are trying some form of CAM. A recent review showed that the current data available from RCTs fail to demonstrate the efficacy of CAM therapy as curative. A small number of CAM therapies may help patients feel better and recover faster. Aromatherapy and massage enhance the well-being of cancer patients.

One way in which the positive effects of complementary medicine are being brought into the mainstream while discouraging the dubious offerings is through the creation of complementary or integrative medicine centers within traditional medical centers. They shun herbal remedies but offer acupuncture, nutrition counseling, massage, meditation, gentle yoga, music therapy, and other disciplines geared to help patients with stress and pain. Patients are generally pleased that the help they had been surreptitiously seeking is now accepted by their physician. Herbs and the so-called food supplements remain unregulated and can be dangerous. *Aristolochia fangchi* was found to cause urothelial cancer in a group of Belgian patients, and patients who use *St. John's wort* to medicate their depression without telling their physician are, in effect, taking a tricyclic antidepressant drug without supervision, and one whose side effects have been better demonstrated than its efficacy. The NCI and the National Center for Complementary and Alternative Medicine (NCCAM) are sponsoring clinical trials at medical centers to evaluate CAM therapies for cancer.

FUTURE DIRECTIONS

The study of the psychological aspects of cancer continues to lead to many fundamental human issues, starting with the basic existential confrontations with death and the sense of self. Despite continuing progress, the public continues its fascination with the role of psychological factors in altering the course of cancer while ignoring the 50% decrease in cancer occurrence that increased exercise, better-eating habits, and stopping smoking and drinking would achieve. The issues associated with genetic advances illustrate the double edge of scientific progress, great promise, and great psychological stresses for which physicians are not prepared.

Psycho-oncology is one of the disciplines whose vitality and scientific productivity encouraged the American Board of Psychiatry and Neurology to make Consultation-Liaison Psychiatry a formal subspecialty. Integration of a psychiatrist into an oncology team that is familiar with the psychological aspects of cancer, which can identify vulnerable cancer patients at risk, and which can recognize and treat the common psychiatric disorders has become fundamental. Psycho-oncologists can look forward to further research to improve the treatment of psychiatric syndromes in cancer patients, to maximize the ability to identify high-risk patients and to offer them support, and to come closer to translating basic science insights into clinical practice.

WEBSITES

The APOS has a website at http://www.apos-society.org/clinical/default.asp. There is also an APOS hotline at 1-866-APOS-HELP, where individuals can call with cancer-related psychological problems and questions and are given appropriate referrals.

The IPOS has a website at http://www.ipos-society.org/index.htm. A multilingual core curriculum on psycho-oncology for professionals is available at the website.

The NCCN has a website at http://www.nccn.org and provides the complete library of NCCN Clinical Practice Guidelines in Oncology. Besides treatment guidelines, they include detailed supportive care guidelines (including distress) and the latest detection, prevention, and risk guidelines.

The NCI maintains an extensive website at http://www.cancer.gov, and the ACS has one at http://www.cancer.org. Both of these sites contain reliable information for the public, patients, families, survivors, and professionals. The NCI maintains a list of current research protocols.

FURTHER READINGS

*Ahles TA, Root JC, Ryan EL. Cancer- and cancer treatment-associated cognitive changes: an update on the state of the science. *J Clin Oncol*. 2012;30:3675–

3686.

American Cancer Society. *Cancer Facts and Figures 2012*. American Cancer Society; 2015.

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

*Breitbart W, Alici Y. Agitation and delirium at the end of life: “We couldn’t manage him.” *JAMA*. 2008;300(24):2898–2910.

Breitbart W, Alici Y. Evidence-based treatment of delirium in patients with cancer. *J Clin Oncol*. 2012;30(11):1206–1214.

*Breitbart W, Alici Y. *Psychosocial Palliative Care*. Oxford University Press; 2014.

*Breitbart WS, Butow P, Jacobsen P, Lam W, Lazenby M, Loscalzo M. *Textbook of Psychooncology*. 4th ed. Oxford University Press; 2021.

Breitbart W, Marotta R, Platt MM, Weisman H, Derevenco M. A double-blind trial of haloperidol, chlorpromazine, and lorazepam in the treatment of delirium in hospitalized AIDS patients. *Am J Psychiatry*. 1996;153(2):231.

*Breitbart W, Poppito S. *Individual Meaning-Centered Psychotherapy for Patients with Advanced Cancer*. Oxford University Press; 2014.

Breitbart W, Poppito S, Rosenfeld B, et al. Pilot randomized controlled trial of individual meaning-centered psychotherapy for patients with advanced cancer. *J Clin Oncol*. 2012;30(12):1304–1309.

Breitbart W, Rosenfeld B, Gibson C, et al. Meaning-centered group psychotherapy for patients with advanced cancer: a pilot randomized controlled trial. *Psycho-Oncology*. 2010;19(1):21–28.

Breitbart W, Rosenfeld B, Pessin H, et al. Depression, hopelessness, and desire for hastened death in terminally ill patients with cancer. *JAMA*. 2000;284(22):2907–2911.

Breitbart W, Rosenfeld B, Pessin H, Applebaum A, Kulikowski J, Lichtenthal WG. Meaning-centered group psychotherapy: an effective intervention for improving psychological well-being in patients with advanced cancer. *J Clin Oncol*. 2015;33(7):749–754.

*Chochinov HM, Wilson KG, Enns M, Lander S. Depression, hopelessness, and suicidal ideation in the terminally ill. *Psychosomatics*. 1998;39(4):336.

*Dalton SO, Boesen EH, Ross L, Schapiro IR, Johansen C. Mind and cancer: do psychological factors cause cancer? *Eur J Cancer*. 2002;38:1313.

Den Ousten BL, Traa MJ, Thong MS, et al. Higher prevalence of sexual dysfunction in colon and rectal cancer survivors compared with the normative population: a population-based study. *Eur J Cancer*. 2012;48:3161–3170.

Fang F, Fall K, Mittleman MA, et al. Suicide and cardiovascular death after a cancer diagnosis. *N Engl J Med*. 2012;366:1310–1318.

Haque R, Shi J, Schottinger JE, et al. Tamoxifen and antidepressant drug interaction in a cohort of 16,887 breast cancer survivors. *J Natl Cancer Inst*. 2015;108(3):djv337.

*Holland JC, Golant M, Greenberg DB, et al., eds. *Psycho-Oncology: A Quick Reference on the Psychosocial Dimensions of Cancer Symptom Management*. 2nd ed. Oxford University Press; 2015.

*Holland JC, Weiss T, Nelson CJ, Roth AJ, Alici Y. *Geriatric Psycho-Oncology*. Oxford University Press; 2015.

Hudson MM, Mertens AC, Yasui Y, Hobbie W, Chen H. Health status of adult long-term survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. *JAMA*. 2003;290:1583.

*Kangas M, Henry JL, Bryant RA. Posttraumatic stress disorder following cancer: a conceptual and empirical review. *Clin Psychol Rev*. 2002;22:499.

Kissane DW, Love A, Hatton A, et al. Effect of cognitive-existential group therapy on survival in early-stage breast cancer. *J Clin Oncol*. 2004;22(21):4255–4260.

Members NC-RFP. National Comprehensive Cancer Network (v.1.2014). Cancer-related fatigue. http://www.nccn.org/professionals/physician_gls/pdf/fatigue.pdf. Accessed May 15, 2015.

Mitchell AJ, Chan M, Bhatti H, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative care settings: a meta-analysis of 94 interview-based studies. *Lancet Oncol*. 2011;12:160–174.

National Cancer Institute. PDQ® Cancer-Related Post-traumatic Stress. National Cancer Institute; 2015. http://cancer.gov/cancertopics/pdq/supportivecare/post-traumatic-stress/Patient.

*National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology: Older Adult Oncology, version 1.2021. http://www.nccn.org/professionals/physician_gls/PDF/senior.pdf

*National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology: Adult Cancer Pain, version 2.2021. https://www.nccn.org/professionals/physician_gls/f_guidelines.asp#pain

*National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology: Distress Management, version 2.2021. https://www.nccn.org/professionals/physician_gls/f_guidelines.asp#distress

Silver G, Kearney J, Traube C, Atkinson TM, Wyka KE, Walkup J. Pediatric delirium: evaluating the gold standard. *Palliat Support Care*. 2014;13(3):513–516.

Smith SK, Zimmerman S, Williams CS, et al. Post-traumatic stress symptoms in long-term non-Hodgkin's lymphoma survivors: does time heal? *J Clin Oncol*. 2011;29:4526–4533.

Teo I, Krishnan A, Lee GL. Psychosocial interventions for advanced cancer patients: A systematic review. *Psychooncology*. 2019;28(7):1394–1407.

△ 27.9 End-of-Life and Palliative Care

XIOMARA ROCHA-CADMAN, M.D., AND MINDY GREENSTEIN, PH.D.

When a patient's death is close at hand, it is the physician's task is not only to ease it but also to remember that *a patient who is dying is a patient who is still alive!* For physicians, who work and train for many years, specifically learning how to forestall death, this crucial task is sometimes easier said than done.

THE CHANGING WORLD

In the 20th century, especially after World War II, science developed beyond our imagination, adding some 20 years to the life span in developed countries. Medicine today is a far cry from the doctors carrying black bags but armed mostly with compassionately shared explanations and predictions. Their reliable presence earned them their patients' trust and gratitude. While payment was necessary (sometimes in the form of bartering), it was really the shared feelings that sustained their sense of meaning.

In contrast, medicine today requires mastering more sophisticated knowledge, skills, and procedures than ever. While medical science is still driven by a fundamental desire to help, its relentless expansion makes demands on caregivers, even as it arms them for the fight. This tension can be particularly difficult to navigate when caring for patients at the end of life, as Atul Gawande describes so passionately in his acclaimed book, *Being Mortal*. He particularly emphasizes the importance of caring about patients' *total well-being* rather than focusing primarily on illness.

In an unexpected shock, the 21st century has reminded physicians that there are sometimes relentless and deadly illnesses against which they haven't been adequately prepared to do battle. For example, rare pandemics which they must fight before being armed with scientific data or even known treatments. While pandemics have often been thought of as once-in-a-century events, many fear that the modern global era has sped up that time clock.

This chapter will discuss the many ways in which physicians and other health care providers have used their ingenuity to improve this situation.

DEATH, LIVING, AND DYING

Death follows many trajectories, among which peaceful old age has never been the most common. It erupts in the midst of health, war, and peace or creeps along an episodic or deteriorating course. Different trajectories bring different experiences, and any death radically alters the future *life* trajectory of the bereaved. Death from chronic illness forces sufferers and their loved ones to live in its presence for a prolonged period. *Untimely* deaths take their own special toll on the survivors, especially when suicide was the method chosen.

The concept of *timely* deaths is problematic. Patients of any age can undergo such deterioration and suffering that death is viewed as liberation. Some older adults accept death, but many older adult patients find it premature, with many families wishing desperately to extend the life of a beloved parent. In such cases, only the medical caregivers think in terms of timeliness, and care must be taken lest the concept becomes a back door through which ageism creeps into geriatric care or, more commonly and painfully, widens the empathy gap that so troubles patients and families.

There is also a distinction of great psychological consequence between physical and social death. While the bearer endures physical death, the entire community experiences social death. In a car accident, the person dies an immediate physical death but can be said to have died a social death only after being buried and mourned and after the family begins to cope with the void left behind. Contrast this with a neurodegenerative disease, whose sufferers lose more and more of the roles that defined their social existence before dying a physical death, having been watched for years by grieving caregivers.

Thus, good end-of-life care must address the physical, psychological, spiritual, familial, social, and cultural characteristics that define the unique life that will be extinguished. The term *dignity* has been suggested as an organizing principle, perhaps because it points to the respect that it should induce in every witness. Developed further under the name "dignity therapy," its goal is to preserve the dying person's sense of integrity while coping with unavoidable physical insults and losses along the way. This requires medical caregivers who derive satisfaction from working toward this goal. A smile, a few moments of peace, scenes of reconciliation, and loving farewells—these are precious gifts from those who are cherished.

AWARENESS OF DEATH

Dying has been described as a physical process that begins at birth, but, more importantly, living is a state of mind that endures to the last breath. Thus, the goal of good palliative care should be the same as any care, namely helping the patient *live* to the fullest under any condition. Yet, there is an emotional chasm between thinking about one's death and seeing its actual contours. It has been said, most notably by Ernest Becker, that denial of death is at the root of many, if not all, social and cultural accomplishments, and it is undeniable that the strength of many religious beliefs is buttressed by their effectiveness in helping humans cope with suffering and death. Yet, Sigmund Freud believed that death anxiety represented the fear of abandonment and loss while fear of death arose because the human psyche could not conceive of its own non-being. That is technically true but incomplete. A number of people accept their death with equanimity. Some are heard to say they feel comforted by the thought of joining Mother Earth or the better life they believe will follow, bypassing the question of being or non-being altogether. Difficulty coping with it may be associated with long-standing psychological issues that had never been tackled effectively, but these are not necessarily major diagnoses. Some existential therapists believe that helping the living grow in the face of their mortality is at the core of strength. This may be useful to some patients but not likely for the ones whose mortality is already their relentless and increasingly intimate companion.

Theories aside, the involved emotions and grief can be so difficult to manage that many patients and families have a shifting perspective in which awareness alternates or coexists with denial. This has been seen in a particularly complex and painful way during the COVID-19 pandemic, where family members' and even patients' perspectives ranged from terror to denial so great that some referred to the pandemic as a hoax and anywhere in between.

This paradox is characteristic of ill patients and may confuse medical staff trying to meet the patient's needs in different ways. It is also seen when families make decisions for a permanently unaware relative. Ultimately, confrontations with death generate over-determined responses in *everyone*: patients, families (especially those members who fear [whether reasonably or not] that their behavior may have contributed to the patient's predicament), and sometimes health professionals, each playing out a personal script on a personal timetable. Discrepancies and misunderstandings are difficult to avoid and often become intense under such stress, all of which are the physician's responsibility to manage! One would be hard put to find discussions of what this responsibility really entails in medical texts of 30 years ago, but recent decades have seen major changes.

HUMANISM IN MEDICINE

Medicine used to be a tightly closed shop and still is from the strictly medical point of view. It now looks appreciatively on many other disciplines that were eager to help patients and that both studies and doctors' own experience have proven to be helpful. In addition to nurses, psychologists, social workers, and physical therapists, all of whom have always been there, there are now art therapists, music therapists, and writers who volunteer to help patients write their stories, to name a few. There are also support groups of all stripes to help patients and family members. Participants have voted with their feet and described improved mood and enriched lives. Groups have also been helpful to residents and fellows in many guises, including T-groups and journal clubs.

Humanism has extended further into medical education, which now includes expanded communication skills and ethics education, as well as cultural and religious sensitivity. In a large number of major hospitals and medical schools, senior faculty members have spontaneously produced films to help doctors understand illness from the patient's perspective. The most moving are the ones that describe a young doctor's terminal illness. The community is also exposed to expert panels of patients, former patients, and family members, as well as relevant films and plays. The questions-and-answer sessions that follow can provide meaningful exchanges between caregivers, patients, and family members. Even medical students are eager to participate. Stanford students, feeling that sexuality was not adequately covered in their schedule, arranged an evening course taught by a professor from a science department on the university campus. It was well attended and subsequently became part of the curriculum. Yale second-year students, eager to help the nearby poor community, organized a Saturday clinic backed by professors impressed by their desire to help. It went well, and by year's end, they presented the dean with a full roster of students ready to take over the following year.

Palliative Care

"Palliative care" once was thought of as a consolation prize when there was no more treatment to offer. It was more an announcement than a treatment. It has, however, undergone a welcome change. Like many important social changes, the current interest in palliative care goes back a long way. Two popular innovators contributed to a sea change. In 1969, Elisabeth Kübler-Ross published a very popular book, *On Death and Dying*, which vividly described the psychological experience of dying patients. Suddenly, they became people with an inner life, a meaningful journey, and important intrapersonal and interpersonal experiences. Psychiatrist Viktor Frankl argued that people can experience a sense of meaning in life until the last breath. Next, the work of Dame Cecily Saunders at St. Christopher's brought a fully elaborated model of good palliative care to world attention in the 1970s. Her insistence on the use of opiates at whatever dose was needed to relieve patient suffering met with widespread public approval (despite frequent medical disapproval), leading to the expansion of the hospice movement in the United Kingdom and a lively grassroots movement in the United States.

Since then, there has been a steady stream of dedicated physicians who devoted themselves to developing palliative care within hospitals, such as Katherine Foley, who pointed out that "competence by itself is not enough, it must be paired with compassion." The value of palliative care was quickly recognized, especially since it enhanced the quality of life for patients of any age, with any kind of diagnosis, and during curative treatment as well as terminal care. It became established as a subspecialty of medicine in 2006 and trained physicians, psychologists, advance care nurses, social workers, and clergy, as well as many other dedicated specialists and volunteers. The demand for them has become so great that there is a shortage of them. The need became particularly acute during the COVID-19 epidemic. In one study of New York City area hospitals, palliative care consults increased by 59% compared with pre-pandemic requests 2 months earlier.

The Introduction of Medical Ethics into Hospitals

The initial rapprochement between ethicists and hospital physicians was difficult without at least one of them already able to function as a bridge. However, in keeping with the flexibility of the era, ethics committees did become well integrated into their hospitals and are now mandated by law. A great strength of these committees is their multidisciplinary nature. Committee chairs are selected for extensive and appropriate medical or nursing experience, a thorough knowledge of the ethical issues, and a high level of experience. They continue to keep the members up to date and bring in speakers as needed.

Most important, they staff a highly valued consultation service that can support a physician in demanding conditions. The ability to prolong both life and the dying process in the context of individual patients' and family members' disparate values, goals, and levels of capacity can put physicians into an ethical bind, which also has psychological consequences for their own well-being. Doctors must recognize that managing them is a collaborative effort. In addition to nurses, there are also ethics committees that can mentor them about the various ethical issues involved.

The ethics consultation consists of a set of services provided by a group of individuals assigned by their institution who seek to resolve conflicts regarding specific ethical concerns. This group can identify and analyze conflicting values when it comes to decision making. One of the most common issues requiring consultation is end-of-life care, for example, whether the decision to withhold cardiopulmonary (CPR) attempts may be made without involving a patient surrogate.

Since terminal situations are the most tragic, the consultation service is often called and may support both family members and members of the team. That support may be a greater factor in resolution than any ethical argument and confirms the value of pairing competence with compassion.

PSYCHOLOGICAL REACTIONS OF PHYSICIANS

All humans are equal in the face of death, but physicians are expected to put aside their personal reactions to remain focused on the care of any dying patient, as well as the family. Yet, there is evidence that doctors may have more death anxiety than others and that some become physicians to conquer that fear. Regardless, long exposures to disease and death can breed anxiety, despite the appearance of familiarity, and, since patients and families cannot see their doctor falter, the practice of medicine demands high intellectual coping skills. Fortunately, this work is profoundly rewarding for most seasoned physicians, who had enough resilience and compassion to get through all the hard moments and benefit from the learning opportunities. Time and experience also facilitate resilience.

However, today's physicians, especially those still in graduate training, are dealing with a society with increasing expectations. Patients and families read a wide range of medical articles—sometimes useful, sometimes inflammatory—and may demand more ineffective treatments. Others may manifest uncontrolled grief or anger, sometimes becoming threatening. In the middle of the hectic pace and complex demands typical of today's hospitals, facing many such confrontations with death can arouse threatening emotions, especially in younger physicians. Rare, but perhaps deepest, is an inchoate fear that can be so rapidly submerged that it is barely experienced, much less explored! Basic self-protective behaviors such as tamping down one's emotional responses are inevitable. If recurrent, they will eventually lead to feelings of inadequacy and anxiety, which may feed upon themselves.

Also, when facing “bedrock” human pain, tired physicians may suddenly realize that some personal experience has been awakened. They tamp that down as well and correctly focus on tackling the situation at hand. However, recurrent experiences like this can ultimately impede their ability to be truly present with their patients. In all these cases, if physicians in this position avail themselves of supportive resources, constructive discussions can liberate them and help them feel strong.

This kind of denial is not universal. Many rise to the demands of their life choices to the benefit of patients, trainees, and colleagues. However, until recently, many of today's students have not previously been exposed to this kind of suffering since death had moved out of the home and into the hospital. This is why it is particularly important for institutions to support young physicians as they navigate this difficult terrain. Even those who usually navigate it well might need help from time to time. Fortunately, the last 25 years have seen welcome developments in the profession's awareness of the need for support and training of staff.

INSIDE THE MAELSTROM: PSYCHOLOGICAL REACTIONS OF PHYSICIANS DURING AN EPIDEMIC OR PANDEMIC

As of this writing, today's students, trainees, doctors, and all hospital staff are receiving the kind of education in death and dying that comes once in a lifetime, perhaps once in a century. This education is forced on them—as it is forced on everyone around them—during times of epidemics and pandemics, an education that is relentless, particularly in its special attack on our common humanity. The fight against COVID-19 has rightly been described as a war, turning doctors, nurses, and all health care staff into frontline soldiers. The psychological toll on them has been great, personally, professionally, even morally. Some physicians describe the pandemic as the greatest threat to the intimacy of the doctor-patient relationship they have ever known. They complain about how little time they have to grieve their losses when there are so many very sick people who need their attention. In the early days of a novel pandemic, an added burden is the fact that the health care team is still learning what helps and hurts their patients, who serve as involuntary subjects in an experiment for which they hadn't signed up. Themes common to health care workers' descriptions of this toll can be found in Table 27.9–1.

As of this writing, 3,000,000 people have already died around the world of this disease, with many more to come before vaccinations are able to subdue it. Medical staff can be overwhelmed by the sheer massiveness of the scale, sometimes to the point of having to turn away patients who need their help and who may be going home to a very lonely and terrifying death. This may also be the case for patients dying of other illnesses, for whom there is no room.

Fear is everywhere during such times, in grocery stores, in dentists' offices, but it is greatest on the hospital floor, where every colleague and every patient is both utterly vulnerable and a potentially lethal threat to everyone else. This situation gives rise to a feeling of dehumanized interactions between patients and the staff caring for them, who must be shielded from them by multiple layers of protective clothing, masks, and face shields so tightly sealed that they leave lines etched into faces long after they are removed. And yet, these same health care personnel must act as the human go-between between dying patients and the family members who would wish to be with them at the end. And they also inform the family when their loved one has died.

Table 27.9–1.
Common Themes in Health Care Workers' Reactions to the COVID-19 Pandemic

- Lack of control and a sense of futility
- Relentlessness
- Unresolved grief over patients dying at an unprecedented pace
- Emotional and physical exhaustion
- Guilt:
  - Learning on the job how to best treat patients for a new disease
  - The feeling of failing their patients morally when forced to make triage decisions they wouldn't make under normal circumstances, like sending people home to die because there is no room for them in the intensive care unit
  - Being afraid of patients and colleagues who might infect them
- Feeling isolated from the community, particularly the feeling of working in a "war zone" while many members of the outside community don't heed public health advice
- Fear of contracting and potentially spreading disease, particularly to family members and loved ones
- Anger
  - Shortages of personal protective equipment
  - Denial by others of the seriousness of the disease
- Dehumanization in the doctor–patient relationship
  - The need for personal protective equipment hides faces
  - Fear of contagion and the intense pace of work lead to spending less time with individual patients
- Feeling emotionally numb or fearing becoming so

The fear stays with medical staff as they return home, sometimes isolating themselves from the family in order to protect them from themselves until they return to work the next day to start it all over again, sometimes while mourning patients they have lost, but also friends, family, or colleagues as well.

Helping the Helpers

Help for medical professionals and staff must address environmental, interpersonal, and intrapsychic dimensions, especially in wards with prolonged illness and high mortality.

A supportive environment acknowledges the stressors and provides:

- Leaders who validate emotional realities and make clear but understandable demands
- A team approach that provides backup prevents isolation and precludes feeling too irreplaceable
- The use of practice interviews for a range of difficult situations, with a trained leader using actors or real patients
- The presentation of seminars and case conferences to discuss difficult cases, the ethical and existential issues that they arouse, or any reaction that a trainee wishes to discuss
- The use of a group for first-year trainees with an outside leader familiar with the hospital environment they face
- The availability of confidential psychological support without prejudice

Doctors can be more authentic with patients if they are allowed to be authentic with themselves and each other within the proper bounds of the workplace. This extends across wider groups as long as they have shared goals, as proven by some of the many humanistic programs offered to the wider hospital community in many medical centers, such as Memorial Sloan Kettering's *Art of Medicine* series. The result is a more comfortable atmosphere, easier relationships, and, above all, patients and families who benefit from the staff's ability to affirm shared goals.

The trainee, or any physician who is troubled or under-functioning in any way, should be approached for an individual discussion with a trusted supervisor and, if indicated, referred to a seasoned psychiatrist who is familiar with the setting.

One approach, originally introduced by hospices, is to hold staff memorial services at regular intervals, during which all patients who had died in the interim are named and remembered with stories, tears, and chuckles. These have now been instituted on many hospital services and are highly valued by staff. Some wards invite family participation, which is deeply valued.

Helping the Helpers During an Epidemic or Pandemic. A supportive environment during an epidemic or pandemic must remain particularly flexible and attuned to the many types and intensities of emotional assault on health care workers and provide:

- Different modalities of support as needed, including one-on-one sessions, support group sessions, or others using videoconferencing and telephone support when face-to-face interactions are dangerous.
- A chance for health care workers to choose which modality best suits them at a given time. For example, using a teleconferencing method such as Zoom, but allowing participants to mute their own computers or turn off their cameras if they want. They may prefer to listen than to speak. In one group for New York City area hospitals, for instance, nurses all preferred a telephone-based intervention rather than one mediated by videoconference.
- An environment in which facilitators remember that what they say to participants—whether individually or in a support group—is often less important than how well they listen to them.
- A chance for medical staff to grieve for all the losses they experience day in and day out.
- Consistent reminders to the medical staff regarding the importance of self-care. Proper self-care helps not only the staff members but also helps them to help their patients and colleagues and decrease the chance for burnout.
- Giving health care workers room to express their feelings while listening to others express theirs offers validation and respite, and sometimes, even helpful advice about how others get through their day. Perhaps the most important element of a support group or session is the reminder to medical staff that, even in the midst of so much isolation and intense feelings of dehumanization, they are all in it together, expressing their

common humanity by taking care of their patients and of each other.

COMMUNICATION BETWEEN DOCTOR AND PATIENT

Good communication skills are essential to the success of any clinical encounter. Poor skills can create vicious cycles; when physicians cannot identify the patient and family values that should influence medical decisions, it becomes hard to develop a solid treatment plan or to assuage uncertainties. This generates chronic anxiety and dissatisfaction in patients and family and increases the likelihood of inconsistent information from other members of the team. Finally, without effective communication, the patient and family experience the inconsistencies that often accompany their sense—right or wrong—of vague treatment goals, which lead them to demand yet more end-of-life health resources.

Guidelines to Good Communication

Communication skills can be learned in many ways, including experiential exercises, retreats, and workshops. Physicians who undergo training in this area report greater comfort and increased feelings of self-worth as they provide terminal care. Interactive teaching modules with role-plays are particularly effective. The experiential, interactive, and group aspects help clinicians detoxify their own insecurities and find support in one another. There are also didactic teaching modules and articles with clear and useful guidelines about breaking bad news. Despite the challenges posed by these tasks, the literature indicates that the courses increase confidence and improve communication. For an immediate example, a scenario follows:

What does the patient know? Always inquire about what the patient and family know before providing new information. “We are here to discuss the further plans for your care, but, before that, do you have any concerns or questions about your illness and the way things are going." Physicians are often startled to see gaps shaped by lack of information as well as culture, education, socioeconomic class, and religion. When such gaps are ignored, serious misunderstandings are likely to follow. Clarify cultural expectations if indicated.

How much does the patient want to know? Historically, families have generally "protected" loved ones from hearing bad news. This practice has been slowly changing in some parts of the world. Most western countries now view it as abuse. Patients have the full right to receive all information but are not deprived of the right *not* to know. The corresponding legal terms in the United States are *informed consent* and *informed refusal.* The language used in medical care is more detailed than for the legal system:

Full Disclosure is for alert, expectant patients, who may pay a psychological price for their knowledge, but will keep up an active dialog and maintain good access to the doctor.

Nondisclosure: Total nondisclosure applies to the very frightened, dependent patient or an incapacitated one; both will need various supports, management, and a proxy.

Individualized disclosure is for patients who only want to know so much. They should be given the right to *informed refusal,* which should be documented, and they should be asked to name a surrogate or a health proxy. "Are you saying that you would rather that I did not discuss the details of your illness with you?"

Sharing the information: The physician should make no more than a few selected points, as patients have difficulty processing too much information when stunned by painful news. "I have given you a lot to process all at once. We can talk tomorrow when you have had a chance to think more about it." When continuing earlier discussions, the focus should be on adapting the goals of care to a changing situation. "We cannot cure the cancer, but we have some treatments that may slow its growth." It is helpful to include familiar staff members. The tone must be one of openness and realism but must always contain a message of hope, even when the situation is so dire that hope lies in the promise of continued caring, support, and symptomatic care. "It is true that the treatment is no longer working, but there are many other measures to discuss, and, no matter what you choose, we will work closely with you to keep you comfortable." Feedback must always be obtained to ensure that the patient has understood and agrees. Patients should be encouraged to express their feelings and explore their level of understanding. "Before we end, I want to make sure that we all agree about the next steps."

Responding to patient and family feelings: The physician should be prepared for possible outbursts of intense emotion, especially in the wake of previous rosy promises. Alternatively, patients may have been given realistic information but subsequently comforted themselves with unrealistic hopes. The opposite is also heard: "You have taken away hope!" This can be minimized by an affirmative, engaged attitude about the many ways in which care will be continued to the end. Ambivalence is always present, and disbelief may predominate. It can be gently confronted, but there is no need to insist because the conversation will be resumed many times in the days to follow, and the passage of time will effect a change. Physicians must acknowledge the emotions, validate their normalcy, and reflect on the patient's and family's observed strengths while indicating an awareness of the difficulties that they are facing. "This is difficult news to hear, even with all of the support and comfort that I have seen you give each other."

Planning and follow-up: These are the keys to a successful interchange. Patient and family must be left with a clear sense of what care is being provided and why. Further meetings must be planned; end-of-life decisions are psychologically and morally so stressful that some repetition is essential. Their understanding and acceptance must be ensured, and multiple support sources must be explored. However, under so much emotional fear and pain, and in complex families, the responses may impair patient care, and psychiatrists may be consulted.

The tenor of these discussions will vary with each physician, but its understanding and respectful tenor should always be there.

Consultation: Communication is always enhanced by consulting with other members of the medical team. It is to everyone's benefit—

whether senior attendings or young physicians—to maintain good relationships with seasoned nurses, whose keen understanding of the patient's family situation helps greatly. Social workers are often sophisticated family therapists. Other resources available include psychologists, who are both therapists and have specialized skills, patient representatives, physical therapists, or any other staff playing a role in patient care.

Psychiatric Contributions

When asked for an inpatient consultation in a new ward, the psychiatric consultant must recognize that the ward is a complex interactive system that affects every patient case. Taking some time to learn the ropes while working on a case pays off for that day and on future consults. Staying in touch with staff is essential.

The consultant should evaluate a responsive patient alone first, then with input from family members. Next, sharing one's findings with the physicians and involved nurses will probably include additional questions and information exchanges on both sides. The results should be a manageable plan. With very difficult cases, the extra time is well rewarded. The psychiatrists must reframe family dynamics in ways the staff finds more understandable, or, if not more pointed, and open to different management.

The consultant is obliged to avoid unnecessary details, even—or especially—if they might become gossip. However, an anecdote that brings shared pleasure into the picture, like the patient's tennis prize, should definitely be shared. Turning to staff members, if psychiatric consultants observe something like exaggerated negative judgments, they can interpret some of the patient and family dynamics in a more understanding way. If consultants identify particular individuals who have become too involved positively or negatively, they must note this and review it privately with the individual. Over time, trusted psychiatrists may be called on by staff members for personal reasons, and obtaining the best referral for them is an important contribution to their work.

PHYSICAL AND PSYCHOLOGICAL SYMPTOMS IN PATIENTS WITH ADVANCED DISEASE

Studies have shown that dying patients experience more symptoms than they can describe. In a cancer patient study, outpatients averaged more than 9 symptoms, and inpatients had more than 13 symptoms, many of them unrecognized or undertreated. Unexpectedly, when psychological symptoms and existential concerns—such as nervousness, sadness, insomnia, fear of dependence, or loss of dignity—are included, they are significantly more prevalent than the physical ones. Many recent studies confirm that physical, psychological, neurologic, hormonal, and existential concerns interact with each other in ways that are only beginning to be understood. What this relationship indicates is that all symptoms must be attended to. Sometimes this needs to be emphasized with physicians untrained in palliative care.

Symptom Assessment

Any intervention must be examined in light of the patient's total illness and treatment goals. How does it relate to the general quality of life and the goals of care? For example, metastatic bone pain responds to radiation, but this should only be done if the benefit outweighs the discomfort of moving the patient and if the patient prefers this to intensified opiate treatment.

The use of validated instruments, such as the Memorial Pain Assessment Scale (MPAS), allows more reliable evaluations, making it harder to dismiss symptoms. It asks for the level of pain, from 0 to 10. Over time, patients serve as their own controls because their scores are internally consistent. Pain evaluation has become a vital sign used in routine reassessments, ensuring that the patient's voice is heard by overworked caregivers. Other scales that can measure distress and quality of life include the Hamilton Depression Rating Scale (HDRS), the Hamilton Anxiety Rating Scale (HADS), the Distress Thermometer, and the Quality of Life questionnaire (QOL). Internal consistency is true for them as well.

Assessment and Treatment of Pain

Incidence. Pain is common in dying patients but can be relieved or at least managed until the very end. However, the prevalence of pain at the end of life remains extremely high.

In cancer, its overall prevalence of 51% rises to 75% in advanced disease, of which 64% is moderate to severe, and 20% to 30% is said to be excruciating if untreated.

Unfortunately, cancer pain is commonly undertreated; about one-third receive suboptimal analgesia in proportion to their pain. Pain management should be individually tailored according to its pathophysiology as well as the patient's subjective experience. Cancer pain can be classified as either nociceptive or neuropathic and can be manifested as acute or chronic.

Pain is also frequent in AIDS and advanced multiple sclerosis (MS) and can be expected in almost any advancing illness unless patients slip into a coma, as seen in hepatic failure or untreated hypercalcemia. Pain specialists are widely available. In the hands of a skilled physician, adequate relief should be obtainable in the majority of cases, and some alleviation obtainable in the rest, but there are often significant side effects. Many of them can be corrected or minimized, but patients and physicians must decide together which symptoms or side effects are most acceptable.

The psychiatrist can work with psychological symptoms now known to be intimately connected to the experience of pain.

Types of Patients. The management of pain should be addressed first by distinguishing its intensity using a 0 to 10 numerical scale, and it must be individualized to address the patient's specific goals in terms of comfort and functioning. It is useful to recognize patients in several categories.

Acute pain patients respond well to treatment but can develop tolerance to any analgesic, requiring more medication if the pain lasts beyond a few days or weeks. This needs to be explored without prejudice or judgment in very ill or dying patients. Nor should patients who will live long enough to improve and continue to use the medications be viewed prematurely as *addicted* or *drug-seeking*. While the very success of recent treatments (bringing more cures or longer survival) has been accompanied by increased addiction and all its problems, this does not warrant decreasing pain treatment. Patients presenting with an acute pain crisis may require hospitalization for appropriate management or may be an appropriate candidate for hospice care.

Chronic pain patients: The treating physician should be careful not to minimize the severity of patients' pain over time or lose sight of the progression of the disease, even if it is slow. The use of opiates in chronic pain is still controversial, especially in nonmalignant pain. However, regardless of controversy, there is no justification for withholding relief in a patient with a predictably limited future, even one that may extend over many months.

Chronic pain patients in palliative care are already dependent and may begin to use opiates as a release from existential suffering, as well as from physical pain. This should not be denied to a dying patient. However, if indicated, a psychiatric consultant can help explore and differentiate between these conditions and can recommend specific interventions. Improved spiritual and psychological well-being occasionally results in less opiate use, but the guiding principle should always be effective relief of distress.

Palliative care patients with a substance abuse problem: If actively abusing or on methadone (Dolophine), these patients are already tolerant and must be given a correspondingly larger amount of medication for adequate pain control. Prejudices about such individuals must be put aside, and they should receive the same compassionate pain relief as others.

Multiple Pains. Patients with advanced disease commonly experience several different kinds of pain (Table 27.9–2). The distinctions are important because they call for different treatment strategies: somatic and visceral pains are more responsive to opiates, while neuropathic and sympathetically maintained pain often responds better to nonopioid adjuvant medications. If neuropathic pain is an integral part of the condition, it may respond poorly to treatment. Treatments to consider are glucocorticoids (dexamethasone) or specific anticonvulsants and antidepressants such as pregabalin (Lyrica) or gabapentin (Neurontin). A comorbid depression may be helped by analgesic antidepressants, for example, a tricyclic drug such as desipramine (Norpramin), or a serotonin-norepinephrine reuptake inhibitor such as duloxetine (Cymbalta). Therefore, the psychiatric consultant should seek out a sophisticated pain history and ensure that all of the patient's different pains receive individual attention until the last breath. Again, pain consultants should be consulted as needed.

Diagnosis and Prognostication. A careful evaluation may require an etiologic work-up; certain pains can be eradicated even in terminally ill patients. Even so, it is critical to always weigh the benefits of better pain relief against the pain and intrusiveness of diagnostic procedures. Clinical decisions are hindered by the inaccuracy of medical prognostication at the end of life. Studies show that the majority of DNR orders were written within less than a week of death. This over-optimism leads doctors to be more proactive with treatment and, often, less generous with opiates. It also worsens the family's ordeal and increases their later regrets, which can create lasting wounds. A consultation with the Palliative Care team may help to disentangle these thorny treatment issues.

Table 27.9–2.
Types of Pain

Nociceptive Pain	
Somatic pain	Usually, but not always, constant, aching, gnawing, and well localized: for example, bone metastases
Visceral pain	Usually, but not always, constant, deep, squeezing, and poorly localized, with possible cutaneous referral: for example, pleural effusion leading to (1) deep chest pain and (2) diaphragmatic irritation referred to the shoulder
Neuropathic pain	Burning dysesthetic pain with shock-like paroxysms associated with direct damage to peripheral receptors, afferent fibers, or the central nervous system, leading to loss of central inhibitory modulation and spontaneous firing: for example, phantom limb pain; can involve sympathetic somatic afferents
Psychogenic pain	Variable characteristics, secondary to psychological factors in the absence of medical factors: vanishingly rare as a pure phenomenon in patients with advanced conditions, in which pain is often present, but often is an additional factor in the presence of organic pain

Treatment of Pain. Aggressive multimodal treatment should be the norm. Drugs are the mainstay but must be part of a broader approach, including ancillary programs such as relaxation, guided imagery, massage, hypnosis, acupuncture, use of splints, and exercises. Music has always been soothing, to the point of helping regulate breathing in semicomatose patients. Unfortunately, there are multiple significant barriers to pain relief in cancer stemming from system obstacles, patients themselves, family members, and clinicians.

PHARMACOLOGIC TREATMENT OF PAIN. The World Health Organization's (WHO) three-step ladder for cancer pain control is a drug-treatment method known to be effective in many cultural and national settings. The use of drugs begins with nonnarcotic analgesics, moves to the opiates used for moderate pain, then to opiates used for severe pain, culminating in a continuous drip with adequate *rescue* doses for breakthrough pain. Many opiates can be escalated without any upper limit, as long as side effects are well managed. Nonnarcotic analgesics and other adjuvant medications should be used as adjuncts to narcotics at any stage.

CLASSES OF PAIN MEDICATION

Nonopioid Analgesics. Nonnarcotic analgesics include nonsteroidal anti-inflammatory drugs (NSAIDs) and acetaminophen. The latter is a good analgesic for mild-to-moderate pain, with few side effects, except rare hepatotoxicity, especially in the presence of alcohol abuse or hepatic disease. All NSAIDs have a ceiling effect; hence, their use in fixed combination with a narcotic means that the narcotic cannot be used to maximum effect unless more of it is added separately. NSAIDs produce analgesia by blocking the biosynthesis of prostaglandins, inflammatory mediators that initiate pain. They are said to help bone pain and are good anti-inflammatory agents, but side effects often limit their use in patients with advanced disease. Contraindications include renal disease, a history of peptic ulceration, hypertension, bleeding diathesis, and encephalopathy. When appropriate and available, use patient-controlled analgesia.

Opioid Analgesics. Some narcotics are effective for moderate to moderately severe pain and should be used until they are no longer effective. Opioids include morphine, fentanyl oxycodone, hydromorphone, oxymorphone levorphanol, and methadone. Morphine, hydromorphone, hydrocodone, oxymorphone, and codeine should be used with caution in patients with fluctuating renal function. Their main side effects are sedation, dizziness, and nausea, which can be dose-limiting, but confusion, delirium, and hallucinosis are also seen. The narcotics that work for severe to very severe pain are divided into pure agonists and mixed agonist–antagonists. Pure agonists have no ceiling effect and can be increased indefinitely, as long as side effects are controlled. Somnolence responds to stimulating agents such as dextroamphetamine methylphenidate and amphetamine and dextroamphetamine in both immediate- and delayed-release preparations. Modafinil, a non–stimulant-activating drug developed for use in narcolepsy, is effective in medically ill patients. Confusion, delirium, and organic hallucinosis respond to antipsychotic agents, and nausea can be helped by a variety of antiemetics (e.g., prochlorperazine). Mixed agonists–antagonists should be left in the hands of skilled practitioners because they have a ceiling effect, can reverse positive effects, or can precipitate an abstinence syndrome in patients already dependent on pure agonists. They also have a higher incidence of psychotomimetic effects, such as confusion and hallucinations. A frequent mechanism of psychotoxicity is the accumulation of drugs or metabolites whose duration of analgesia is shorter than their plasma half-life, notably morphine levorphanol and methadone. Using drugs such as hydromorphone with half-lives closer to their analgesic duration can relieve the problem without loss of pain control. Cross-tolerance is incomplete between opiates, so patients should be tried on several drugs, and the dosage should be lowered when switching drugs. Knowing equianalgesic doses of drugs and different routes of administration (Table 27.9–3) is important so as to avoid accidental under medication or overmedication. For example, to change a patient from intramuscular (IM) to oral morphine, one should multiply the IM dose by two to six times (depending on the chronicity of usage) to avoid causing the patient pain and provoking what others will call drug-seeking behavior. Dying patients who must choose between pain and alertness face a poignant dilemma, but somnolence can be countered by several safe psychostimulants and can be used in ways that meet shifting needs.

Dosing Instructions. The need for regular *by-the-clock* administration, as opposed to *as-needed* (PRN) administration, cannot be overemphasized. It is counterintuitive, but a maintenance schedule achieves faster and better pain control with a smaller total dose because it is an early and often pre-emptive response. PRN orders only provide a slightly or very delayed response. The patient may wait until the pain is significant before calling for medication and then wait again for it to be given, during which time increasing pain is worsened by anxiety and growing resentment. The patient starts to ring repeatedly, often annoying overburdened nurses. The patient has been set up! To some extent, so have the nurses.

Rescue doses should be directly available to the patient for breakthrough pain. Their repeated use should signal the need to raise the maintenance dose. Any patient on breakthrough doses should always be sent home or to a hospice with the same prescriptions since new doctors may not realize the patient's condition.

Adjuvant Drugs. Adjuvant drugs should be used simultaneously throughout. Many of them are psychotropics well known to the psychiatrist, but their analgesic effect has often been shown to be separate from their primary psychotropic effect. They are particularly important in neuropathic and sympathetically maintained pain, where they can be the mainstay of treatment, but they also introduce new side effects that need to be recognized.

Surgical and Anesthetic Modalities. Other modalities include palliative radiation and anesthetic and neurosurgical techniques that can all give regional relief where other methods have failed. Their intrusiveness is only warranted if they are painless once installed and can provide significant relief, as they often do. Other methods should be considered in imminently dying patients.

Coexistence of Pain and Psychiatric Symptoms. Patients with pain have a significantly higher incidence of depression and anxiety. It had been suggested that: (1) chronic pain may be a depressive equivalent, (2) pain may cause psychiatric syndromes, and (3) pain may coexist with psychopathology in vulnerable subjects. Given the steady stream of interactions now recognized as integral to the human organism, it can be said that *all* suffering patients are vulnerable, and treatment should proceed from *any* angle. Psychiatric interventions, along with primary medical ones, selected integrative medicine options and thoughtful, personal care all play a role. The result will be more than the sum of its parts.

Table 27.9–3.
Routes of Opioid Administration

Route	Comment
Oral	Preferred in cancer pain management.
Sublingual	Buprenorphine (Buprenex) is effective but SL buprenorphine is not indicated for pain. Efficacy of highly lipid-soluble drugs, such as fentanyl (Actiq), is likely, but there are no studies and little clinical experience. The efficacy of morphine (Duramorph) is controversial.
Rectal	Available for morphine, oxymorphone (Numorphan), and hydromorphone (Dilaudid). Customarily used as if the dose were equianalgesic to oral dose. Absorption is variable, however, and relative potency may be higher than expected.
Transdermal	Available for fentanyl. Dosing interval is 2–3 days. Empirical indications include problems with oral drugs, desire to offer a trial of fentanyl, compliance problems with oral dosing, or the possibility of improved quality of life if oral therapy is avoided.
Oral transmucosal	Formulation using fentanyl is currently undergoing trials for breakthrough pain.
Subcutaneous	
Repetitive bolus	Ambulatory infusion pumps can

Continuous infusion Continuous infusion with patient-controlled analgesia	provide continuous infusion with any parenteral opioid formulation. More advanced pumps can also provide patient-controlled analgesia. The clearest indication is an inability to tolerate the oral route.
Intravenous	
Repetitive bolus Continuous infusion Continuous infusion with patient-controlled analgesia	Continuous infusion is possible if a permanent venous access device is available.
Epidural	
Repetitive bolus Continuous infusion using percutaneous or implanted system	The clearest indication is pain below the arms and dose-limiting side effects from systemic opioids. Often co-administered with a local anesthetic.
Intrathecal	Usually administered via a totally implanted infusion pump. May be cost-effective for those patients with a clear indication for intraspinal therapy and long life expectancy.

Adapted with permission from Portenoy RK, Osta BE, Bruera E. Physical symptom management in the terminally ill. In: Chochinov HM, Breitbart W, eds. *Handbook of Psychiatry in Palliative Medicine*. 2nd ed. Oxford University Press; 2009.

Furthermore, in the face of advanced disease, the role of personality factors and pre-existing diagnoses becomes less and less significant. While psychiatrists mention them in their notes, they must also comment that *these issues are no longer in play*. This is important in preventing the staff from psychologizing the patient’s delirium and related behaviors.

However, there are some patients who manifest clear psychopathology, are more difficult to evaluate, and whose complaints are more easily disqualified. Hence, the consultant must remain attentive and help the staff treat their pain as aggressively as that of other patients. The psychiatric diagnosis and its impact on illness must be explained and treatment recommendations made while generating as much genuine sympathy for the patient as possible.

Undertreatment of Pain. The incidence of psychiatric complications is particularly high when pain is undertreated by the staff, which happens for many reasons. The daily use of the pain scale was driven by studies showing that when doctors and patients were separately asked to rate pain, their responses matched well for mild and moderate pain, but the higher the patient's pain rating went, the slower the physician raised theirs. Inadequate management and unawareness of resources are both possible reasons, but often it appears to be an unconscious avoidance of the treater's own helplessness in the face of the patient's complaints. Nurses and aides might also share it, especially if the patient is genuinely unlikable, but they may tend to recognize it sooner because of their more frequent and intimate interactions.

However, it can also be patients and/or their families who have false beliefs about medications, such as destructive effects of drugs or side effects, or addiction. These misconceptions are often a metaphor for fear of death. In its most severe form, a dependent spouse whose mate is suffering but not yet moribund, or the mother of a young child who is still conscious but exhausted, cannot tolerate the loss of reactivity that comes with pain relief. They cannot allow the patient much-needed relaxation or sleep, as they experience this separation as the beginning of their final loss. Psychiatrists can be called in, but familiar trusted physicians can often handle similar situations.

Medical Staff Vulnerabilities. Medical staff barriers to pain treatment can be better managed if better understood. At an unconscious level, doctors, and nurses, living through many patient experiences, develop an incompletely acknowledged, internal standard of how much pain is *expectable* in a given situation. The standard is personal, not universal, but tends to lean toward the so-called "good patient," that is, the brave and uncomplaining one. Each patient's complaints are unconsciously measured against both other patients' behaviors and the caregiver's own standard; many come up wanting. Nurses are more likely to make a correction for each other, but young, untested doctors who may still feel immortal may have more difficulty. They will not have the same understanding as a mature physician who has had a hard turn at being a patient or a senior physician who just lost a spouse after a long illness. In the final analysis, the tendency to underestimate pain is intimately related to every physician's personal history and stage of life, as well as their deepest motivation and coping mechanisms in doing this difficult work. They are capable of learning, but attention should be paid to it. They must also protect themselves against too much empathic identification, as well as guilt. An occasional lapse into ignoring or blaming the victim is an all-but-irresistible response to so many difficult situations and another good reason for close teamwork and the support and insight that come with it.

PSYCHIATRIC INTERVENTIONS. Psychiatrists can help optimize pain management by reorienting staff responses in a supportive, nonjudgmental way, emphasizing the need to respect the patient's statements, and reaffirming the now well-accepted importance of daily review of patients' pain scales. Psychiatrists can give the team insight into the patient's psychological state, emphasizing the patient's problems in ways that undermine a team's unconscious pain biases. They can sympathize with the team's difficulties and uncertainties, normalize their feelings, ensure adequate pain management, and re-educate them about the low addiction risk and its now irrelevant role in dying patients.

Table 27.9–4.
Therapy in the Terminally Ill Patient

Psychotherapy (individual, family, or group)
Crisis intervention
Family-focused grief therapy
Group supportive psychotherapy
Meaning centered psychotherapy
Dignity therapy
Cognitive-behavioral therapy
Behavioral therapy
Supportive–expressive therapy
Provision of information
Life review
Facilitating communication
Addressing unfinished business
Leaving a legacy
Provision of information
Reframing
Focusing
Problem-solving
Relaxation
Meditation
Mindfulness-based stress reduction
Hypnosis
Desensitization
Rehearsal
Music therapy

Psychiatrists' role in bringing about optimal pain relief is a powerful intervention in itself, but it is not complete unless integrated into a multipronged program that includes supportive psychotherapy and some of the many focused therapies that deal with terminal care (Table 27.9–4). Patients using behavioral interventions report a greater sense of control and decreased anxiety and pain. In one study, doctors underestimated the value of these methods compared to the patient’s own estimate. Hence, it is useful for psychiatrists to recommend them. Psychotherapy must help the patient differentiate between pain and suffering (which includes psychological symptoms but also prominent religious and existential concerns!) and must facilitate appropriate responses to each. Most hospitals have chaplains, but a trusted religious familiar is also desirable.

Fatigue

Fatigue is one of the most common symptoms reported by patients, and its effect on the quality of life is profound. In one study, 62% of cancer patients rated it as worse than pain, 12% said that it made them want to die. Yet only 27% reported that their doctor had suggested a treatment. Unfortunately, fatigue is hard to define, even harder to measure, and often goes unnoticed. It is a multidimensional concept with physical, cognitive, affective, and motivational components that differentiate it from symptoms such as weakness, somnolence, and tiredness after exertion. It is integral to acquired immunodeficiency syndrome (AIDS), cardiac, respiratory, hepatic, renal, and neurologic end-stage disease.

If internists accept fatigue as an inevitable correlate of major illness, psychiatrists may over-psychologize it. Fatigue is, after all, a defining criterion of the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* diagnosis of depressive and anxiety disorders. Psychologically driven fatigue may be incapacitating and is known to amplify somatic symptoms. Furthermore, the incidence of psychological disorders in very ill patients is high. However, as mentioned before, these distinctions are no longer clear; hence, teasing out psychologically from physically determined fatigue may not be possible because the two coexist with such regularity.

Table 27.9–5.
Possible Etiologies of Fatigue in the Medically Ill

I. Medical/physical conditions

- Associated with the underlying disease itself
- Associated with treatment for the disease
- Chemotherapy
- Radiation
- Surgery
- Biologic response modifiers
- Associated with inter-current systemic disorders
- Anemia
- Infection
- Pulmonary disorders
- Hepatic failures
- Renal insufficiency
- Malnutrition
- Neuromuscular disorders
- Hypothyroidism
- Hypogonadism
- Hypoadrenalism
- Associated with sleep disorders
- Associated with immobility and lack of exercise
- Associated with chronic pain
- Associated with the use of centrally acting drugs (e.g., opioids)

II. Psychosocial factors

- Associated with anxiety disorders
- Associated with depressive disorders
- Stress-related
- Related to environmental reinforcers

Adapted with permission from Portenoy RK. Physical symptom management in the terminally ill: An overview for mental health professionals. In: Chochinov HM, Breitbart W, eds. *Handbook of Psychiatry in Palliative Medicine*. 2nd ed. Oxford University Press; 2009.

Table 27.9–5 outlines some causes and interventions for end-of-life fatigue. It should be evaluated in multiple dimensions before deciding on treatment. It can be cruel to require a dying patient to move, but this can improve the quality of life in a patient who has more time left. Both patients benefit from being taught energy-conserving techniques. Psychostimulants used purely for fatigue have been observed to be helpful and can interrupt a vicious cycle of fatigue and despair, although different patients will have different risks from side effects.

Non–stimulant-activating drugs, such as modafinil and dextroamphetamine/amphetamine, are increasingly used in ill patients, although they are not, at this time, FDA-approved indications.

Insomnia

Sleep hygiene is important to the healthy and the sick, and yet hospitals can be blatant offenders in their lack of concern for patients' sleep. Dying patients may be desperate for the blessed release of "getting away" for a few hours. Of course, very ill patients may require care and procedures at any time of day, but such simple measures as minimizing noise and maintaining optimum light are neglected in environments that are otherwise attentive to patient needs. Patients and their caregivers need education about maintaining as close to a normal circadian rhythm as possible and preceding sleep time with a quiet routine. Long-acting pain medication is critical to prevent middle-of-the-night breakthroughs, and many patients need help in understanding the use of sustained-release preparations, such as morphine or oxycodone. Despite recent publicity about abuse, these long-acting pain medications are not as readily abused by very ill patients as by patients who may get better for some time. They have an important role to play in late-stage patients and should not be avoided because they are too long-lasting. Sedatives are equally important. Zolpidem and zaleplon are well tolerated, and the latter is short-acting enough that it can be taken in the middle of the night. Benzodiazepines have an antianxiety effect as well, which often makes them the drugs of choice in nondelirious patients but are contraindicated in delirious ones. Several antidepressants are so predictably sedating that they are often used as hypnotics even in the absence of a diagnosis of depression. They include trazodone, nefazodone, mirtazapine, amitriptyline, imipramine, and other tricyclics. Antipsychotics, such as olanzapine, quetiapine, and risperidone, can be used the same way. The side effects of all the psychotropics should be evaluated, but generally, the habituation that accompanies the use of many hypnotics does not occur.

Terminal Anorexia and Cachexia

Anorexia and the wasting state known as cachexia are common near the end of life, but, in some conditions, such as cancer and AIDS, they appear well before the end of life, causing great distress to family and friends. If they have not been educated about its involuntary nature, unprepared family members become angry that the medical staff does not share their intense concern. They may try to relieve their sense of helplessness by fixating on making the patient “try a little harder” to eat. This generates guilt and anguish in the patient at a time when peaceful human contacts are the most precious. Families have been known to request a psychiatry consultation for an eating disorder, wanting to believe, against all evidence, that better nutrition would change the inevitable. This example of making patients responsible for their illness is one of the most poignant and frequent demonstrations of love and desperation going awry. Families, and indeed the public as a whole, need to understand the involuntary and overwhelming nature of terminal anorexia and, even more importantly, that the patient is not experiencing hunger and will not benefit from forced nutrition.

In fact, the opposite is often true. When anorexia is part of the active dying process, total parenteral nutrition (TPN) has been shown to worsen the patient’s quality of life. However, in cases in which the patient can still experience hunger—for example, in ovarian cancer in which a nonfunctional GI tract may exist in the presence of functional liver and kidneys—TPN is available and has been known to sustain patients for long periods. (Its wisdom can be disputed, but that is a different issue.) Thirst, another family concern, is more related to dry mucous membranes and can be relieved with frequent small sips of water. Intravenous (IV) hydration and transfusions are counterproductive in the face of failing hearts or kidneys. Table 27.9–6 outlines some recognizable causes and the treatment of anorexia for patients who are not yet actively dying. The treatment plan should address multiple etiologies.

Other GI Symptoms

Nausea and Vomiting. Nausea is near the top of lists of unpleasant symptoms, and unfortunately, is one of the most common. Many patients say, “I would be okay if it were not for nausea!” The act of vomiting itself, although it gives temporary relief from nausea, is unpleasant and can be painful in certain conditions. It can be difficult to eradicate even when its etiology is known and all the more so when one cannot pinpoint the cause—a not infrequent event.

Available antiemetic medications are helpful in controlling more vomiting and nausea. A careful assessment may help to establish the best approach and most effective treatment for both. Asking the patient-specific questions, such as onset, frequency, time of the day, quality of content, and its association with other symptoms, will help the clinician to choose a more effective medication. For instance, escalation of pain medication could be associated with nausea as well as medical interventions, procedures, and certain treatments such as chemotherapy or radiation. Common causes of nausea and vomiting in the medically ill are related to opioids, NSAIDs, and centrally acting drugs. Intra-abdominal neoplasm and antineoplastic medications are common causes of nausea and vomiting, as are intracranial neoplasm and leptomeningeal disease. Other causes include psychological processes and metabolic disturbances such as hepatic failure or renal failure and electrolytes abnormalities.

Table 27.9–6.
Causes and Treatment of Anorexia

Causes

Disease-related

Nausea, fever, anemia, weakness, dehydration, pain
Sepsis, other acute malaise, electrolytes abnormalities
Weakness or damage in muscles of deglutition
Altered cognitive states
Altered metabolism: Uremia, hepatic failure
Abnormal response to nutritional intake

Treatment-related

Chemotherapy or other drug effect (nausea, diarrhea, altered taste or smell)
Surgery
Radiation therapy effects
Dialysis side effects
Transplant rejection, graft-versus-host disease
Cancer-related symptoms

Psychological

Anxiety syndromes, depression, delirium, specific phobias
Pre-existing or relapsed psychiatric disorders

Nonpharmacologic interventions

Education
Dietary management
Discuss etiology (often multifocal) and characteristics (involuntary) and with patient and family
Reassurance
Increase physical activity and psychosocial intervention
Alternative therapies (massage, muscle relaxation, mindfulness-based therapy, yoga)

Pharmacologic intervention

Megestrol acetate (Megace)
Steroids
Appetite-inducing psychotropics (olanzapine, quetiapine)

Psychostimulants (methylphenidate, dextroamphetamine)
Wakefulness-promoting agents (modafinil)
Selective-serotonin reuptake inhibitors (fluoxetine, paroxetine, citalopram, escitalopram, sertraline)
Selective-norepinephrine reuptake inhibitors (venlafaxine and duloxetine)
Norepinephrine-dopamine reuptake inhibitor (bupropion)
Alpha-2 antagonist/5HT2/5HT3 antagonist (mirtazapine)

Diarrhea and Constipation. Diarrhea is common in AIDS and in certain cancers. It is usually associated with local pain, burning, and often abdominal cramping. It also leads to fecal incontinence, one of the most humiliating and unpleasant symptoms a patient must endure. Rapid cleaning and good skin care by a relaxed, well-related person is important to shorten the experience and minimize pain, skin irritation, and breakdown. Drugs that slow down intestinal motility can be helpful, so the anticholinergic side effects of many psychotropics are welcome. However, it can be a very resistant symptom when caused by opportunistic organisms and/or *Clostridium difficile*.

Constipation is also common and can cause significant abdominal cramping, bloating, distention, nausea, and vomiting. Ideally, it should be prevented or aborted by adequate hydration, optimal mobilization when possible, and, most of all, early institution of an aggressive and consistent bowel regimen. The use of pain medications could contribute to constipation if not given with stool softeners. Enemas or disimpaction may be instituted if the patient desires. However, in someone close to death, increased opiates may be preferable, especially as normal bowel function is frequently absent during the terminal phase.

Respiratory Symptoms

Respiratory distress is one of the most difficult symptoms to endure—and to watch. Although measures have been developed to assess dyspnea, assessment of the ill patient should be based on self-report because objective measures, such as oxygen saturation and arterial blood gas measurements, do not often correlate with subjective experience. It is easy to treat because opioids and benzodiazepines decrease the anxiety and the urge to breathe within minutes, allowing the patient to rest and relax. However, in the actively dying patient, a significant decrease in the effort to breathe is associated with increased hypoxia and eventual carbon dioxide (CO_2) narcosis. Parenthetically, this is one of the most painless ways to die and is the subject of debates about the “double effect,” in which an action has an intended (patient comfort) and an unintended effect (hastening of death). If dose escalation is too sudden, it may hasten the moment of death, but when the drugs are prescribed appropriately, respiratory depression is gradual; death still comes gradually but more comfortably.

As with pain patients, the dyspneic patient is anxious and hyper-alert and may have already been exposed to opiates or benzodiazepines (the latter is now contraindicated). Therefore, the gradual escalation of drugs to give consistent relief of dyspnea does not result in sudden death. In fact, family members who have waited at a dying patient’s side comment on “how long it takes for a heart to give up.” One of the more common signals that the end is near is agonal breathing, where one hears only acute gasping breaths that become less and less frequent. If family members are keeping a long watch, it is useful to check on them and explain what is and will be happening. The dying patient can still display involuntary movements that are experienced as communication by a watching grieving spouse or child. These last moments should be managed with kindness and delicacy because they will never be forgotten.

Distress due to persistent coughing is exhausting and interferes with communication. Opioids turn off the cough reflex promptly, and if the patient is strong enough to cough repeatedly, opioids can and should be used without risk of respiratory depression.

Sedation of Patients. Neuromuscular blocking agents are used routinely in intubated patients to prevent them from fighting the respirator. Agitation can precipitate accidental extubation, leaving a conscious patient without any ability to breathe, a terrible experience. For that reason, patients in critical care units (CCUs) are often kept sedated while on the ventilator. When the patient is receiving only comfort measures, such a condition must be re-evaluated, but, whatever the case, *the planned withdrawal of life supports must always include adequate pre-sedation.*

This needs affirmation because, currently, there are more CCU patients stabilized on respirators who must be transferred to intermediate units to allow the treatment of newly acute patients. This is where final controversies may play out. There have been times when end-of-life management occupied the time of politicians whose duties lay elsewhere or when individual caregivers did not yet understand that the rule of care took priority over their personal convictions while it was being argued in appropriate venues. All of these scenarios with patients in pain and respiratory distress involve the double effect. This partly describes every drug ever given to any patient, but if the side effect is hastened death, the controversy should be brought up and carefully reviewed before proceeding. This will be discussed further in the section on ethical issues.

PRIMARY PSYCHOLOGICAL SYNDROMES AND SYMPTOMS IN TERMINALLY ILL PATIENTS

Psychological symptoms are nearly universal at the end of life. Psychiatric syndromes occur with an increased, definable frequency and may have different age and gender distribution. For example, anxiety and depression are as common in men as in women. The DSM-5-TR classification remains an important framework on which to base clinical observations, but it was not designed with dying patients in mind. Hence, for terminally ill patients, it is pragmatically useful to focus on three major conditions: anxiety, depression, and confusion—which, in fact, frequently coexist and overlap. Rarely, specific phobias of needles, enclosures, and the like may interfere with comfort into the terminal period and should be addressed to ease the patient’s experience. The consultant should remain alert to any precipitant factors that can cause new discomforts and work to defuse them. Major psychotic disorders may become submerged under the increasing symptomatology of the active dying process. However, they require specific attention when the patient is not actively dying and when the psychotic symptoms are clearly separate from and superimposed on the symptoms of either the disease or the treatment.

Prevalence

Much of the research on prevalence has been done in cancer and AIDS and shows a marked increase of psychiatric conditions near the end of life. Severe depressive symptoms rose from 25% to 77% in a sample of hospitalized cancer patients, although stricter criteria find 15% with major depression and another 15% having severe depressive symptoms. The prevalence of delirium rises from a range from 25% to 40% to as high as 85% with increasingly advanced disease. The association of psychiatric symptoms with pain was demonstrated in one of several consistent studies in which 39% of patients with a psychiatric diagnosis reported significant pain, compared to 19% for those without a psychiatric diagnosis. In a sample of AIDS inpatients, 65% had an organic mental disorder, and 27% had major depression. The financial cost of psychiatric disorders can be inferred from a study in which patients with a psychiatric diagnosis remained in the hospital 60 days longer than those without a psychiatric diagnosis.

General Treatment Principles

Improved quality of life is the *immediate* goal, so all available modalities that can bring symptom relief should be instituted rapidly while a plan for psychological and family interventions is designed and set into motion. Since psychiatric syndromes are intertwined with medical conditions, understanding the medical condition is necessary. Good communication with the treating physician yields a fuller analysis and better coordination with medical treatment goals.

Most commonly, terminally ill patients demonstrate an intertwining of anxiety and depression. These are difficult to tease apart, and the term *negative affect* has been suggested to define them as a symptom complex. It is hard not to feel grief at so much loss and fear about the dark unknown ahead. Individuals whose deep faith in an afterlife that animates their spirit are an exception, but, even among them, many describe coexisting regret at the loss of their temporal life with its rich palette of life events.

However, current symptoms may be an exacerbation of a pre-existing disorder, in which case, symptom control still remains the primary response. Other than medical complications, a personal or family history is the best predictor of psychiatric problems in advanced disease.

Death Trajectories

There are many different death trajectories with different conditions, of which a few will be briefly mentioned. Death from pulmonary insufficiency is slow and is terrible for both the patient and the family. Mild sedation eases the pain and may let the patient rest while slightly shortening the time to death (the so-called double effect). Psychiatrists may be needed when there is debate among participants, as might the Ethics Committee.

Neurologic disorders are multiple. Many cause a slow decline until, far into the disease, only shreds of the patient's sense of self remain intact. In such cases, both patient and family caregivers should be offered early support, the patient some tranquilization, and the family a chance to speak about their reactions. Under such conditions, there may be no previous psychopathology, but the late events can create it, as happens in accidents and disasters.

Every cancer is an unpredictable disease and can have a long and complex course, hence an unpredictable mode of death. It is also characterized by presenting several possible options, not only throughout the course but often during the final months. Families live through this unpredictability at a cost. Complications and disagreements cause stress for everyone, including medical staff. In addition to the usual chart review and examinations, it is useful for psychiatric consultants to also learn from the staff the “story” of the course of treatment. This often opens routes for improvement.

Anxiety in Patients with Advanced Disease

Characteristics. Anxiety can be the presenting symptom for almost all medical disorders and is often a side effect of medications. However, in patients with advanced disease, it usually presents somatically, such as restlessness, hyperactivity, tachycardia, GI distress, nausea, insomnia, shortness of breath, numbness, or tremor. It lowers the threshold for pain, worsens functional impairment, and increases distress. Anxiety often interferes with the patient’s ability to cooperate with other treatments or to relate optimally with loved ones. Patients are more likely to refer to fear, apprehension, or ruminations rather than to anxiety itself. It is also associated with deep existential concerns, as outlined in Table 27.9–7. Individuals with pre-existing anxiety disorders commonly have an exacerbation.

Mr. S, a 50-year-old physical therapist with newly diagnosed unresectable lung cancer, had panic symptoms when his wife left his bedside to attend to other needs. He hyperventilated, felt short of breath, became restless, was unable to concentrate, and felt overwhelmed with morbid ruminations about his future. His symptoms were exacerbated by his discomfort about his dependence on her. Mr. S. was taught special breathing and relaxation exercises and given clonazepam, which brought about a marked resolution of his anxiety. He felt calmer, more resilient, and able to handle periods of solitude with less difficulty.

Treatment. Anxiety is so common that it is sometimes discounted, but an etiology should be sought, and treatment should be instituted promptly. Anxious patients are more likely to be aroused enough for supportive therapy, cognitive-behavioral techniques, and complementary modalities, such as deep relaxation, meditation, and adapted yoga. If a patient brings up an issue that had never been resolved, discussing it is not only appropriate but a major gift, contrary to some therapists' feeling that "the work won't go anywhere." Bearing witness to one's life, especially in the company of another person, is a poignant and meaningful act. Patients' inner lives are their last personal possession; sharing and enriching them is a powerful intervention, one of the last times they will reflect upon themselves in a meaningful dialogue with another human being. It is perhaps the most meaningful of all.

Table 27.9–7.
Causes and Treatment of Anxiety

Causes	Treatment
Medical causes (psychiatric diagnosis: Anxiety secondary to a medical condition) Untreated delirium Central nervous system pathology Hormone secreting tumors Metabolic or electrolyte abnormalities Pulmonary embolus Respiratory insufficiency, hypoxia, dyspnea Sepsis Poor pain control Side effect of bronchodilators, steroids, β-adrenergic drugs, and other medications Substance withdrawal **Psychiatric disorders** Acute stress response Adjustment disorder with anxiety Adjustment disorder with mixed anxiety and depressive mood Panic disorder[a] Phobic disorder[a] Generalized anxiety syndrome[a] Depressive disorders[a]	**Search for an etiology**. If a medical etiology is present, treat it or minimize it. **Use a multimodal approach**. Should always include the provision of immediate relief along with other interventions. **Anxiolytic drugs** Benzodiazepines Lorazepam (Ativan), oxazepam (Serax), and temazepam (Restoril) are metabolized through conjugation and hence are more usable in hepatic insufficiency. Alprazolam (Xanax), clonazepam (Klonopin), diazepam (Valium), chlordiazepoxide (Librium) are benzodiazepines that are metabolized through oxidative pathways and can be used with renal impairment. Nonbenzodiazepines (beware of drug interactions) Selective serotonin reuptake inhibitors Escitalopram (Lexapro) Serotonin and norepinephrine reuptake inhibitors Venlafaxine (Effexor) and duloxetine (Cymbalta) Neuroleptics (alone or with a benzodiazepine)

Existential causes

- Awareness of oncoming death
- Fear of death itself
- Fear of the dying process
- Mourning of losses
- Interpersonal regrets
- Continuing conflicts
- Guilt over abandonment

- Haloperidol (Haldol), olanzapine (Zyprexa), quetiapine (Seroquel)
- Tricyclics
- Hydroxyzine (Atarax)
- Buspirone (BuSpar)

Nonpharmacologic interventions

- Behavioral methods
- Hypnosis, meditation, deep relaxation, with or without guided imagery (selected for induction of peacefulness)
- Supportive psychotherapy
  - Explores patient fears, reduces cognitive uncertainty, and reassures about accompaniment
- Family sessions
  - To improve communications, to reduce tensions, and to say goodbye
- Spiritual counseling

---

[a] Often exacerbations of pre-existing symptoms.

MEDICATIONS. Several classes of effective antianxiety medications can and should be used even in the presence of hepatic or renal impairment, given careful drug selection, dosage adaptation, and ongoing monitoring. Benzodiazepines add to the depressant effects of other sedating drugs and can worsen the symptoms of confusional states, making them a poor long-term choice for the older adults. However, if the patient has been taking them for a long time with good effect, continuing them is desirable because their action is immediate and, with careful dosage, generally beneficial. They must be chosen with the patient’s hepatic function in mind. All are metabolized by the liver, except for lorazepam, oxazepam, and temazepam, which are metabolized by conjugation and, hence, less dependent on liver function. Lorazepam and diazepam can be given parenterally and sublingually. The first is short-acting and is preferred by patients who want more control over their ability to interact. The second is long-acting and tends to accumulate unless monitored carefully, but it is effective and has the added benefit of being an excellent muscle relaxant, which, for some patients, gives significant pain relief in itself. Clonazepam lasts 12 hours and does not accumulate. Habituation can require dosage escalation of all benzodiazepines, but this is acceptable if the drug improves the quality of these patients’ limited life.

Table 27.9–8.
Causes and Treatment of Depression in Terminally Ill Patients

Causes	**Treatment**
Medical causes: Resulting in mood disorder secondary to a general medical condition	Search for an etiology; if a medical etiology is present, treat it, or minimize it
Medications	Psychopharmacological management
Steroids	
Interferon and interleukin-2 (Proleukin)	Choose an antidepressant based on the patient's review of symptoms
Methyldopa (Aldomet)	
Reserpine (Serpalan)	Tricyclic antidepressants in depressed patients with pain and insomnia
Barbiturates	Monoamine oxidase inhibitors
Propranolol (Inderal)	
Some antibiotics (amphotericin B [Amphocin])	Mirtazapine (Remeron) in depressed patients with insomnia and anorexia
Some chemotherapeutic agents	
Disease conditions	Bupropion (Wellbutrin) in sedated depressed patients
Metabolic and electrolyte abnormalities	
Endocrine abnormalities	Psychostimulants effective in treatment-resistant depression and sedated patients
Hormone secreting tumors	
Nutritional deficiencies	Selective serotonin reuptake inhibitors in depressed patients

	with cardiac risk factors
Anemia	
Uncontrolled pain	Serotonin norepinephrine reuptake inhibitors
Psychological causes	
Adjustment disorder with depressive mood	Mood stabilizers Nonpharmacologic interventions
Adjustment disorder with mixed anxiety and depressive mood	Individual psychotherapy
Comorbid with various anxiety disorders	Supportive psychotherapy Cognitive-behavioral therapy
Depressive disorder not otherwise specified	Family therapy
Existential causes (all held with morbid intensity impervious to reassurance)	Group therapy
Awareness of oncoming death	Spiritual counseling
Mourning of losses	
Interpersonal regrets	
Continuing conflicts	
Guilt about burdening family	
Guilt about betraying loved ones	
Decreased self-worth	
Self-criticism about perceived failed relationships and life achievements	

Depression in Patients with Advanced Disease

Prevalence. Depressive symptoms are also common in advanced disease and associated with the same existential factors as anxiety. Studies have found a prevalence ranging from 9% to 58%, depending on criteria strictness. Risk factors again include previous episodes or family history. Table 27.9–8 outlines the common causes and treatments for major depression.

Diagnosis. Differentiating somatic effects of disease from the neurovegetative criteria of major depression is especially daunting in the terminally ill. Since the DSM-5-TR criteria (developed for primary psychiatric patients) include somatic symptoms that may be due to their medical disease, a *substitutive approach* has been developed, which replaces physical symptoms with nonsomatic symptoms. Table 27.9–9 describes the Endicott Substitution Criteria, which is the substitutive approach in primary use because it has been found to perform as well as the DSM-IV-TR criteria and the stricter Research Diagnostic Criteria. It also highlights the important clinical observation that classically described depressive thoughts and feelings are not universal in terminally ill patients and, when present, reflect depression just as they do in physically healthy patients.

Treatment. Current studies and a large body of clinical experience with medically ill patients have shown that the pharmacologic treatment of depression can be useful through the last days of life, even in the face of definable medical causes.

Table 27.9–9.
Endicott Substitution Criteria for Depression

Physical-Somatic Symptoms	Psychological Symptom Substitute
Change in appetite or weight, or both	Tearfulness, depressed appearance
Sleep disturbance	Social withdrawal, decreased talkativeness
Fatigue, loss of energy	Brooding, self-pity, pessimism
Memory and concentration deficits, indecisiveness	Lack of reactivity

> Mrs. V, a diabetic patient with end-stage renal disease, was on dialysis for 2 years and was diagnosed with depression and severe insomnia. She started on a daily oral dose of 15-mg mirtazapine and immediately slept better and felt less depressed within 3 weeks. Four weeks later, she was admitted to the hospital with congestive heart failure for what would be her terminal admission. She did not wish to give up the antidepressant but now felt oversedated and wished to be more alert during the day to interact with her family. Twice-daily 5-mg doses of methylphenidate made her more engaged and communicative. After her death, her family expressed deep gratitude that they could continue to feel so connected to her until the end.

Antidepressants are the first line of treatment in patients who can be expected to survive long enough for the delayed onset to take effect. If there is no time or if increased energy and ability to interact with loved ones are urgently desired, patients can be treated with psychostimulants. Because the side effects of antidepressants can be very variable, including sedation and excess energy, they should be used empirically, and their side effects monitored carefully.

Cognitive-behavioral techniques are often less relevant because very ill patients often cannot concentrate well enough. However, for those who can, deep relaxation associated with guided imagery can be powerful enough to transport them, even briefly, to a better place.

Ms. D was a 17-year-old high school student with recurrent, unresponsive acute lymphocytic leukemia (ALL). Her loving parents encouraged her to do more than she could and were counseled separately to lower their expectations. When seen alone the first time, she was very aware, thoughtful, and saddened by her parents' requests, yet showed remarkable understanding and maturity. She gladly selected deep relaxation and guided imagery instead of further "talking" and mentioned that she loved hiking. The therapist saw her several times before she died, "taking her" on several favorite hikes. This created safe moments for Ms. D, who described them as "really peaceful." Relaxing this way enabled her to stay connected to her own strengths.

Supportive therapy of depressed terminally ill patients is one of the more difficult therapeutic tasks. Every therapist has moments of frustration instead of empathy with the imperviousness of depressive thinking and affect. When found at the door of death itself, the therapist's own feelings of helplessness and fear may generate the urge to flee. Taking an active role in symptom control, psychopharmacological treatment, and behavioral interventions can relieve the doctor and help the patient. Beyond that, if the patient seems to welcome the therapist's presence, then staying with the patient is paramount; a faithful presence at the end is profoundly meaningful to someone who is dying. Some therapists experience such moments as gifts to them as well.

Confusional States in Patients with Advanced Disease

Prevalence. The prevalence of delirium rises to 85% in patients with advanced disease. If one includes the last hours of life, it nears 100% unless the patient lapses rapidly into a coma or dies of an acute event, like pulmonary embolus. It is always unexpected and traumatizes the survivors, despite knowing full well that the patient was near death. Some patients slip into an irreversible coma, leaving families in a bedside vigil, which, while they can be exhausting, may actually give them a period of adaptation before the final moment of death. However, these events only occur 15% to 25% of the time.

Table 27.9–10.
Subtypes of Delirium

	Hyperactive	**Hypoactive**
Type	Hyperalert, agitated	Hypoalert, lethargic
Symptoms	Hallucinations, delusions	Sleepy, withdrawn
	Hyperarousal	Slowed
Examples	Withdrawal syndromes: alcohol, benzodiazepine	Hepatic, renal encephalopathies; benzodiazepine intoxication
Pathophysiology	Elevated or normal cerebral metabolism	Decreased global cerebral metabolism
	EEG: fast or normal	EEG: diffuse slowing
	Reduced activity in GABA systems	Overstimulation of GABA systems

EEG, electroencephalogram; GABA, γ-aminobutyric acid.

Reprinted with permission from Breitbart W, Lawlor PG, Friedlander M. Delirium in the terminally ill. In: Chochinov HM, Breitbart W, eds. *Handbook of Psychiatry in Palliative Medicine*. 2nd ed. Oxford University Press; 2009.

Diagnosis. Patients frequently experience some disorientation, impaired memory, concentration, and altered arousal as they become increasingly ill, especially if they are older adults. These symptoms may remain mild or may be the harbingers of a full delirium, as described in Tables 27.9–10 and 27.9–11 summarize the causes and treatment of delirium.

The DSM-5-TR diagnosis excludes the presence of dementia as an etiologic factor, but, in the growing older adult population, not only is pre-existing dementia common, it is a strong risk factor for a superimposed delirium and makes the delirium much more resistant to treatment. While it is harder to tease out how much symptomatology is reversible, this does not alter the management.

Since Paula Trzepacz's important paper, delirium is now divided into:

- A *hyperactive hallucinatory delirium*, easily recognized by both physicians and families, although the early symptoms are often mistaken for depression, anxiety, and poor coping. They may not get the diagnosis right, but they do bring a real problem to medical attention.
- A *hypoactive delirium* must be differentiated from the more easily recognized hyperactive state (Table 27.9–10). These patients are hypoactive, withdrawn, with varying levels of somnolence and unresponsiveness. They are quiet and undemanding, making it easier for this reversible condition to be long overlooked by the team. This is highly regrettable; one recent group of patients who recovered after treatment described their continued awareness and had acutely unpleasant memories of the experience. Hence, physicians should remain alert to diagnosing hypoactive delirium and treat it as aggressively as the hyperactive form.
- A *mixed delirium* includes patients with fluctuating activity levels. They may also have normal levels of psychomotor activity, but their attention and awareness are disturbed.

A psychiatric consultation was sought to evaluate depression in Mr. K, a 56-year-old practicing attorney with pancreatic cancer. His moderately severe back pain was apparently responding to morphine, but he was noted by the inpatient staff to be more withdrawn, disengaged, quiet, with poor eye contact, and somnolent throughout the day. The consultant found Mr. K to have disturbed arousal, with mild confusion and disorientation, slow speech, and disorganized thought process. Mr. K admitted to intermittently experiencing visual hallucinations that he had been too embarrassed to report to nursing staff. Diagnosed with a hypoactive delirium secondary to opioid medications, he was treated with a low dose of olanzapine, 2.5 mg at bedtime, improving his sensorium dramatically. He became alert, fully oriented, better related, and without perceptual disturbances or thought disorder, all without decreasing his much-needed pain medications.

Table 27.9–11.
Causes and Treatment of Delirium in Terminally Ill Patients

Causes	Treatment
Opioids	Medications commonly used in managing delirium
Some hallucinatory drugs such as LSD and PCP	Typical neuroleptics
Benzodiazepines	Haloperidol (Haldol)
Anticholinergic drugs	Chlorpromazine (Thorazine)
Sedative-hypnotics	
Steroids	Atypical neuroleptics[a]
Cardiac glycosides, antihypertensives	Olanzapine (Zyprexa)
Anticonvulsants	Risperidone (Risperdal)
Antihistamines	Quetiapine (Seroquel)
Bronchodilators	Ziprasidone (Geodon)
Benzodiazepine and alcohol withdrawal	Aripripazole (Abilify)
Metabolic and nutritional deficiencies	
Hypoxia	Benzodiazepines may be combined with neuroleptics to calm severely agitated delirious patients.
Fluid, electrolyte, osmolarity imbalance	
Vitamin deficiencies (thiamine, folate, vitamin B_{12}, niacin)	Lorazepam (Ativan)
Infections	Midazolam (Versed)
Liver and renal failure	
Vascular disorders	Anesthetics

Endocrine and hormonal disorders	Propofol
Endocrine tumors	
Carcinoid tumors: mainly ovary and lung	
CNS disorders, metastatic disease, and neoplasm	

Management. In a preterminal delirium, multiple etiologies are the rule and should be pursued. However, when a patient develops a terminal delirium in the last days of life, symptom control must again be the primary goal. Its general management includes security and orientation measures, pharmacologic management, and a search for treatable etiologies. Eighty to 90% of delirious patients respond to symptomatic treatment, even when close to death.

Haloperidol and chlorpromazine are the main IV medications, but the atypical antipsychotics are effective in patients who use other routes of administration. Drug companies are actively devising new routes of administration for several of them. If given alone, benzodiazepines worsen delirium and are contraindicated, but they are useful adjuncts to antipsychotic drugs, providing sedation for persistently agitated patients. When cognitive impairment and early delirium are present, antipsychotics alone can be effective in controlling symptoms. The weight gain and type 2 diabetes seen in chronic psychiatric patients are not a problem here, but dyskinesia, akathisia, tremors, parkinsonian rigidity, and, rarely, tardive dyskinesia does occur. Parkinsonian symptoms can occasionally affect muscles of deglutition and intercostal muscles, leading to impaired swallowing and respiration that are mistakenly ascribed to other causes.

Delirium is so common that medical staff deal with it quite matter-of-factly. However, a hyperactive one can frighten family members, who see it as a drastic change for the worse. They fear a permanent change in the patient, an especially upsetting one, given the content of hallucinations, their commonly paranoid cast, the appearance of anguish and terror, the frequent aggressiveness, and the loss of dignity associated with many delirious behaviors. In such situations, family education is critical; in fact, explanations and reassurances should be routinely given at the first hint of any symptoms to minimize their trauma. It usually falls to the nurses, who are well acquainted with it, closest to the patient’s condition, and well aware of the needs of the family and visitors. Hypoactive deliria, however, are sometimes mistaken for depressions.

PSYCHOLOGICAL ACCOMPANIMENT OF THE DYING PATIENT

When facing a bedrock human experience with an imminently dying patient, the usual language of therapy should be altered so as to avoid pathologizing tragedy. Accompaniment is the goal of interactions with dying patients, even those who are less distressed, and even while aggressive efforts are made to treat all treatable diagnoses and symptoms. The psychiatric consultant must be familiar with the patient’s medical condition. The patient–psychiatrist boundary should not be as well defined with terminally ill patients as with healthier ones. The doctor should be alert and responsive to signs of increased discomfort. Touching bedridden terminal patients calls for judgment but need not have negative connotations. If a patient is weak and sometimes unheard and unable to call for help, then helping the patient call for assistance is itself a strong intervention. Untangling or pulling up a sheet the patient cannot reach can be offered casually, as is helping to contact the nurse for more complex needs. Psychiatrists are well trained to respect boundaries, but this is a time when some flexibility is indicated.

Family involvement must be explored and encouraged, and staff members queried for their views. The latter group is, especially at this time, a part of the operant group and must be included in understanding and resolution of difficult situations.

The special case of the epidemic/pandemic amplifies and complicates all of these interpersonal dynamics. In such cases, every person is a possible vector for everyone else. This may also be true for patients dying of other illnesses, whose visitors could be vectors.

The predominance of protective clothing and equipment among both staff and visitors serves to highlight any sense of isolation already being experienced. As do extreme restrictions on visitors, as may happen at some points during epidemics. At such times, it is not uncommon for family members and patients to exchange last words through an iPad or other electronic means, a wrenching experience for all, including medical staff who serve as the family's stand-ins for physically accompanying the patient until death.

Every avenue to lessen isolation and increase the feeling of human connectedness should be explored, for example, having staff wear photos of themselves on their protective garments in case patients can't see their faces or making a point of holding a patient's hand through gloves. Family members must be given clear information about what to expect, including an understanding of the level of uncertainty about those expectations. And, while visitation restrictions might be necessary, the team must take all steps possible to enable family members to physically stay with their loved ones in some manner, as safely as possible. This would be a kindness to everyone, patient, family, and staff.

Sharing the Patient's Subjective Experience

Dying patients should be engaged for as long as they are able to respond. Giving them a chance to share feelings, memories, and achievements honors their existence and is a precious gift to them. A patient who is talking with an empathic listener feels themselves as *still living*, not dying, during that conversation. Eliciting and listening are more important than anything that the interviewer has to say. Although supportive responses are appreciated, unrealistic reassurance aborts authentic conversation and is cruelly isolating.

Many dying patients are deeply troubled by unresolved personal issues and often keep silent in response to covert messages from those around them. The therapist should try to help them explore these issues with as much care as with any other patient. Besides the painful presence of their disease, they may be still full of thoughts about the meaning of their lives, the values by which they have lived, the relationships that they feel they have honored (or betrayed), about death, about life after death, about their faith or lack of it, and about their relationship to God. The conversation opens the door to sharing burdens and shedding tears. The relief it can bring is like letting fresh air into the room.

Patients may also be grateful for help reconnecting to happy memories, meaningful achievements, joys, and laughter. Consultants may ask for permission to share some achievements with the team, who will be glad to discuss them with the patient, rather than only drug doses. Patients close to death can still enjoy mundane activities that bring them closer to loved ones or to treasured aspects of themselves. A late-life review may help patients see the totality of their lives, revealing positive meanings and connections and giving them a chance to value their lives with more depth, solidity, and self-esteem.

Some patients may have late emotional changes of desires or feelings. It has been shown in several settings that, as patients become sicker (than at the time of their decision making), they accept an increasingly limited quality of life as worth living and opt for more onerous, less promising treatments if they offer even a small chance of help. It is reassuring that IRB committees make sure that only procedures with some potential are allowed at such desperate times.

Disease-Specific Considerations

Different illnesses routinely bring special issues. For example, a patient on dialysis can opt for death rather than treatment three times a week and is more vulnerable to acting on feelings of depression, anger, hopelessness, and reactions to family neglect or conflict. The cancer patient must acknowledge possible death during complex treatments while hoping for a cure or remission. Treatment decisions in the face of a disease that is less and less treatable become increasingly difficult and can cause acute anxiety. Stem cell transplant patients experience high levels of anxiety and depression because they are getting a last chance with high stakes and high risks. The availability of organ transplants has created a large population of patients who know they may die while waiting for a cure that could be just around the corner. A patient who ignores medical advice to amputate a limb, and is later diagnosed with metastasis, suffers a great deal and benefits from added support and understanding. Other patients who have lost limbs may have lost a very important part of their identity. Neurodegenerative disorders are associated with increased physical disability and dependence. If cognition is maintained, a number of patients and their caregivers can manage well until the end. When loss of cognitive faculties is present, the problems can be difficult for the caregivers.

When a patient is dying during an epidemic, in addition to any family issues predating the onset of the illness or epidemic, are the myriad responses each person in the family unit may have toward both the illness itself as well as toward the patient dying from it. These may include guilt, either reasonable or unreasonable, over possibly having infected the patient, or anger at another family member for perhaps having done so, especially if reckless behavior was involved, or even anger at the patient themselves. A further complication is the large range of possible attitudes toward the disease itself. In the case of COVID-19, for example, attitudes may range from panic all the way to denial and referring to it as a hoax; behaviors may range from going to great lengths to avoid the disease—such as washing one’s grocery bags for fear of fomites—to refusing to wear masks, physically distance, or avoid crowds as recommended by public health officials. In such cases, family members must be helped to focus squarely on the patient’s experience and comfort, regardless of who did what to whom and leave family quarrels outside the hospital room door.

Group Therapies

Given enough survival time, group interventions for patients with advanced disease are helpful. They provide a forum in which patients can “detoxify death” by relating to each other’s concerns, talking about dying without fearing that they are upsetting loved ones, and practicing getting acquainted with dying. Many other kinds of approaches have been developed, including groups that use rational–emotive therapy, *meaning-based* groups based on Viktor Frankl’s work, *spiritually based* groups that use a nonsectarian but more faith-based vocabulary, groups that use mindfulness-based meditation techniques, and many free support groups run by nonprofit organizations, each focused on a particular condition. Various kinds of family involvement and support have also been given of benefit to both family and patient. While this subject deserves further attention, it goes beyond the scope of this chapter and is well documented elsewhere.

Other Modalities

Hypnosis, guided meditation, deep relaxation techniques, meditation, minimal movement yoga, music therapy, and techniques that address body and spirit, such as soothing the body and mind with forms of touch and gentle massage, are all helpful to patients who can participate. The consultant can make a personalized tape for the patient during which gentle relaxation and breathing are followed by a detailed evocation of a place or experience chosen by the patient. If the result is relaxation or peaceful sleep, the patient will continue to use it at home. Music therapy in skilled hands that follow the patient, not the score, has been observed to ease respirations and calm agitation in very ill patients. All of these methods deserve the growing attention they have been getting.

SPIRITUAL CONCERNS

More than 70% of patients turn to religious beliefs for comfort, and many others yearn for spiritual sources of meaning. The therapist must engage this side of the patient’s inner life in a comfortable, respectful way. The consultant should not presume to replace a chaplain but should integrate this aspect into psychological concerns, as most patients have learned that their religious or spiritual concerns do not find a home in the medical environment. Does the patient have religious beliefs? Did they have them in youth? What is their relationship to them? To a personal God? To a sense of a transcendent reality? Ned Cassem suggests that the patient's relationship to God be viewed as one of the patient's many important relationships. Some patients describe themselves as not religious but spiritual. If so, can they describe how they experience their spirituality? Can it be expressed through the values by which they have defined themselves? Through belief in a greater power or life force? For people with environmentalist convictions, the unbroken chain of life in nature may be a link to immortality.

The psychiatric consultant—and everyone else—should facilitate a connection to appropriate religious guides for any willing patient. Their presence can offer healing that no layperson can duplicate. Patients may lack the energy or the will to ask for them, and many do not know that they are available.

As mentioned earlier, patients with a well-elaborated faith, especially one that includes a reunion with God in the afterlife, tend to do better at the end. They may experience the illness as a test of their faith, may view suffering as having redemptive value, and may gain strength from a belief that God will not send them more than they can handle. However, they also often ask, “Why me?” and may struggle with anger, fear, a sense of betrayal or abandonment, disappointment, self-imposed guilt, and, at worst, a loss of faith that leaves them truly desolate. If the patient wants to continue to speak with the therapist, it is appropriate, but referral to an experienced religious authority should also be on the table.

Decision-Making Issues

Many patients base medical decisions about life-sustaining interventions on religious beliefs. The belief that life belongs to God and must be nurtured as long as possible, or a belief in miracles, or both, are often the reason given for demanding aggressive life-prolonging measures. This may be quite sincere, but patients and families may use religion as camouflage for dynamic conflicts or as a way to control end-of-life treatment decisions because they lack trust in the staff or are unable to face death. Mistrust is a frequent reason for rejecting medical recommendations and often crystallizes around inadequate DNR discussions. Today’s clinicians are trained to respect religious beliefs and can also recognize multiple agendas, but it may still take psychiatric input to untangle them and promote conflict resolution.

PATIENT–FAMILY UNIT, THE FUTURE SURVIVORS

The intensity of family relationships becomes even greater during the terminal period. Yet, that is also when one may see conspiracies of silence. With an unconscious patient, it is less necessary to intervene, but the saddest silence is that of a loving family, tense and quiet at the bedside because they are protecting the patient by not talking about dying, while the patient remains quiet for fear of upsetting them. Instead of closeness, expressions of gratitude, apologies, reminiscences, and farewells, there is distance. Although surrounded, such patients feel as though they are dying alone.

Family-Focused Grief Therapy

The stakes go beyond the patient. The last days of a dying patient are indelibly imprinted on the memories of the survivors. Events perceived to have gone wrong can resonate for years. Research has shown that families with good communication endure the experience better and have a more normal grieving process. Later, they are better able to function as a family, with better outcomes for individual members too. In one study, families that had openly discussed dying were initially more upset but, in the end, were less distressed than families that had been more protective. It is likely that the patient died still feeling loved and valued.

Psychiatrists can use family sessions to open patient–family dialogues that can identify discrepant views of the illness or treatment and explore concerns regarding an absent member, which may be undermining patient support and medical management. A major crisis, such as the imminent loss of a member, destabilizes the family structure, creating an opportunity for the psychiatrist to promote adaptive change and reconstitution. For example, family-centered grief therapy, which includes the patient and continues after the patient's death, provides a natural setting for such interventions.

When an older adolescent dropped out of college to take over her dying mother's duties with greater effectiveness, the family subtly discouraged her return to school after her mother's death. Psychiatric intervention while the mother was still alive allowed the patient to play a part in reorganizing family roles in a way that made it more possible for her daughter to continue her studies without facing too much hostility. The patient herself had wanted it settled, and the family responded lovingly.

The Last Vigil

As patients sicken, their consciousness waxes and wanes. It cannot be known how aware they are of their surroundings, but families usually keep watch, and the patient's comfort measures should not be changed unless it is clear they have become counterproductive. As patients fade, family members often feel as though they are being abandoned. This can be unexpectedly painful, even in relatives who had felt prepared for the event. Involuntary movements are often overinterpreted as last messages. The consultant should appreciate these dynamics, help the family grasp the reality, and prevent inappropriate requests for ever more life-sustaining measures. Comatose or moribund patients may look uncomfortable during agonal breathing, but their level of awareness is minimal. It is difficult for exhausted family members, who may feel guilty for wishing it was over. *These negative feelings should be recognized, validated, and* *normalized*. Taking turns can be encouraged. As death becomes imminent, the nursing staff will be attentive, calling for a physician and supporting the family.

During times of epidemic, the last vigil may, in fact, occur at home, punctuated by electronic visits via tablets or other technologic devices, often mediated by a nurse, who may, for instance, touch the patient as directed by the family. Such moments can be filled with guilt for not being able to be with one’s loved one. Medical staff may be able to help families make the most of these “visits” by helping them think ahead of time what activities might be most comforting to their loved one, whether music, prayer, reminiscences, or, if their loved one can speak, by simply being present and listening.

After Death: The Final Farewell

Responses range from dramatic to subdued. The family is given time alone, crying and comforting each other, sometimes singing hymns or songs. Sometimes a spouse or parent clings to the body and refuses to leave. Private leave-taking should be respected, and the grievers are usually ready to separate before institutional post-death routines are interfered with. Some surveillance always continues by nursing staff, who will call psychiatry if an occasional acutely distraught member refuses to leave, if suicide is mentioned, or if anything else concerns them.

All of the patient’s doctors should be informed about the death so that they can respond personally. A phone call or a handwritten note is highly valued by family members. The consultant should offer bereavement counseling as needed. Family members sometimes call back years later. Hospices automatically provide bereavement counseling. Some hospitals also have bereavement programs, sending letters to the family at intervals during the first year, with condolences and information about bereavement groups. The response is usually very positive.

Unattended Deaths

Many traumatic or unexpected deaths become known from a catastrophic phone call. The patient is gone, but the traumatized family needs help to absorb the loss, cope with the circumstances, and come to some kind of closure. An unexpected death, even an understandable one, such as a heart attack in a cardiac patient, leaves an aching sense of unfinished business over and above the expected grief. If the person was the victim of a crime, obsessive thoughts might be difficult to repress, and grief may turn to rage, with profound psychological disruption. It is also difficult, if not impossible, for survivors to fully make peace with suicide.

The absence of a body or a locale is also troubling. Soldiers take enormous risks to retrieve bodies. Families clamor for remains or a shred of irrefutable knowledge or proof. Emergency rooms, police departments, and religious and community institutions should be equipped with a list of resources to help survivors of traumatic deaths. Psychiatric input can include program development and consultation to a wide range of professionals and individuals. Families must be helped to construct their own ritual whereby they acknowledge to themselves and others the finality of their loss and perhaps create a place, if not a grave, around which to center their memorializing. In the case of a pandemic, the family may live stream services. If they prefer a more intimate way to say final goodbyes, they might wish to also plan for a memorial service to be held after the dangers have passed.

Attended Deaths

Patients can die in an acute hospital, in nursing homes, in hospices, and at home, with or without hospice support. Most patients still die in acute care hospitals, having received active care until shortly before they die. This may occur because death is sudden or because the family or patient needed to be in a place where “everything is being done.” Fortunately, a growing number of hospitals have palliative care teams that provide appropriate care in the acute hospital setting, and home care programs are still increasing. Many patients die in nursing homes without the benefit of special care.

This unfortunate situation could be remedied by bringing formal hospice care into nursing homes, but funding sources and turf issues need to be settled before it can if ever, become routine practice. Inpatient hospice care was the first model of care to be developed and is warmly remembered by grateful families whose multiple human needs and those of their loved ones had been met in ways that they had not been trained to expect. As hospice care has gained acceptance, the insufficiency of inpatient beds has encouraged the development of home hospice care. This has, in turn, encouraged more families to elect to keep dying loved ones at home.

In a home hospice program, the patient is evaluated and accepted in the usual manner but stays at home. The patient and family receive extensive instruction about what to expect, including the possibility of readmission to the hospital if they become truly overwhelmed. They are helped to obtain necessary materials, taught how to use them, and get as much home help as they need. All the while, they receive medical supervision, nursing care, and emotional support with 24-hour phone availability and routine daily contacts.

Without this kind of help, a good death at home can be difficult to achieve. With it, patients feel safe from the abandonment they commonly fear, and families feel safe from the terror of an unmanageable event. The families of these patients work hard but are more likely to feel competent and in control.

However, this rosy picture should not become a guilt-provoking expectation. Some families feel unable to manage at home, and, indeed, many do lack the physical and emotional resources. When a patient wishes to die at home, the family may concur while harboring serious qualms. When a patient pleads and a family refuses, the clinician must resolve a painful conflict. Sometimes a visit to a hospice or an interview with a hospice worker eases the impasse, leaving the patient more willing to go into a hospice or the family more willing to take a chance at home, as long as transfer to inpatient hospice is a guilt-free option.

SURROGATE BURDENS AND FAMILY COPING DURING THE END OF LIFE

The difficulties faced by family surrogates are sometimes discounted by the medical staff because the need for their decisions may displace other concerns. However, end-of-life decisions are momentous, and their memory remains as permanently vivid as classic traumatic memories. Laypeople severely overestimate the practical consequences of end-of-life decisions. Proxies feel that they had decided the “when” of a patient’s death, while all they did was to minimally alter the “how.” Health professionals should be sensitive to these distortions and should educate and support surrogates well before the actual moment of decision. Early, extensive communication can positively impact the surrogate’s ability to grieve and to reach a healthy sense of resolution.

Three months after her 83-year-old demented father's death, Mrs. X consulted the hospital psychiatrist because she was haunted by the decision that she had been asked to make about stopping antibiotics during his terminal pneumonia. His baseline pulmonary function was getting poorer, which led the physician to ask the question in good conscience, but she experienced her final act on her father's behalf as a betrayal.

Grief counseling focused on two areas. The first area was her relationship with her father before and during his years of dementia and how they led to the final decision. Had her family relationships been seriously dysfunctional, or had she been more vulnerable, the counseling might have required a longer course. This bifurcation is a common technical issue for psychiatrists dealing with cases of complicated bereavement. If the physician wishes to make the option available, it should be openly discussed, and the goals of treatment should be altered through a mutual decision process. The second area was her relationship with his caregivers, who, while generally supportive, had underestimated her vulnerability and her need for their affirmation of any decision she made.

Medical caregivers may not realize how little some families know and how much their words and behaviors are overvalued. Family members bring intense feelings and possibly contradictory agendas to the bedside. There may be jostling for authority or determination to replay an important family scenario one last time (e.g., favorite child, supportive sibling, or patriarchal authority) or confrontations previously held at bay. Formal proxy or informal surrogates are at added risk for being either criticized by family members or pressured for decisions by the staff. Psychiatric consultation should be called as early as possible when this constellation is identified.

Primary Caregivers

Primary caregivers are a special case. Their physical and psychological distress has been found to be, in some cases, as great or greater than that of the ill, but not dying, patient. Studies have shown that they experience a 25% to 30% incidence of job loss, and more than half move to lower-paying jobs to accommodate the need for flexibility. When the patient nears death, some caregivers feel ready or even eager to let go, which may cause some guilt. Sometimes they cling desperately to an individual whose continued existence has become the raison d'être for their own. This should not deny the finding that almost all caregivers also describe feelings of joy and transcendence while carrying out their task, as described in a study of HIV patients and their partner caregivers. However, they still come to an end in a state of depletion, and their welfare should be of concern to clinicians.

Ms. T and Ms. M were sisters who had lived together after they were widowed, although both had children. Ms. T had cared for Ms. M as the latter's coronary artery disease had worsened and was the health care proxy. When Ms. M suffered a stroke and became comatose but not brain dead, her son began to discuss goals of care with the physician, who had to remind him that it was not his decision. At the same time, she had to urge Ms. T to overcome her reluctance to confront the severity of her sister's condition and to protect her from her nephew's growing resentment. Aware that the closeness between the sisters had created a distance from the next generation, the physician-focused on fostering rapprochement between them. As the son acknowledged his appreciation for the care that his mother had received, his sister felt more supported and could begin processing her imminent loss.

LEGAL ISSUES: ADVANCE DIRECTIVES, PROXIES, AND SURROGATES

This section reviews information that does not always affect an end-of-life case but needs to be familiar to physicians and other staff.

Advance Directives

Advance directives (ADs) are instructions left by patients about their wishes regarding medical interventions once they can no longer make their own decisions. Legally binding in all 50 states, they are dictated by the Federal Patient Self-Determination Act of 1992. ADs refer to the right to *refuse*, not to demand, treatment. This distinction is sometimes lost on patients and even more on families, opening the door to stressful confrontations. Although the use of ADs is increasing, the majority of patients still do not have them, leaving family members to make decisions.

Living Will. Patients may leave instructions about what care they do or do not wish to receive after they become incapacitated. These can be extremely helpful to family members and medical caregivers. However, they are often vaguely worded and cannot predict the exact form that terminal events will take, so there is often an element of uncertainty about their interpretation. This can be lessened if there have been family discussions and/or legal advice.

Health Care Proxy. Patients may formally designate any individual to function as a health care proxy and make any and all health decisions on their behalf. The law dictates that choices be made according to the following hierarchy: (1) the patient's own wishes; (2) if the patient's wishes are unknown, the patient's best interests; (3) if that is ambiguous, the decision can be justified as being the most reasonable one; and (4) last, state interest can be invoked; the latter is currently applied mainly to pregnant women, in which case the welfare of the fetus has increasingly been seen as competing and having priority over that of the mother, but that is another story. The proxy cannot overrule any patient's written instructions, especially with respect to life-sustaining measures, such as feeding, hydration, and life supports.

Family members are believed to have the most reliable knowledge of patients' wishes and are deemed natural surrogates. Yet, numerous studies have found that agreement between patients and their family members about end-of-life decisions is no better than random. Indeed, their intimacy with the patient creates ambiguous motivations. Some proxies approach the task without any previous discussions with the patient and do not feel confident about understanding the patient's wishes. Others harbor different opinions and carry out their own wishes, despite legal instructions to the contrary. Both scenarios engender significant guilt, which endures long after the patient's death. In general, the psychological and ethical burden of proxies, as well as their need for ongoing education and support, is profoundly underestimated.

Surrogate Decision Maker. When patients have left no formal instructions, the nearest relative becomes the decision-maker by custom and by default. It is usually the one who is most involved, but, occasionally, the lot falls to one who is emotionally or geographically distant, uninformed, or unwilling. This leads to the problems mentioned above.

Durable Power of Attorney. A durable power of attorney is a broader legal power that includes the health proxy along with financial powers. It has been legally available for a long time but was less widely used.

Brain Death Determination. Brain death is defined as two successive determinations of a flat EEG at least 6 hours apart and is accepted as a legal alternative to death by cardiac arrest. Many families have difficulty accepting that a loved one with warm flesh and a beating heart is truly dead. Often, life supports are continued for a period of time while the family is helped to come to terms with the situation. Some families invoke religious beliefs to refuse it altogether, creating stalemates that often go to institutional ethics committees with unpredictable results. The growing use of organ transplantation has been a factor in the legal acceptance of brain death because it enables a better chance for the transplanted organ to work well.

Table 27.9–12. Advanced Directives

Directives	Patient	Family
Medical decisions		
Living will	Often causes problems if lacking details about withholding and withdrawal of life supports, feeding, nutrition, dialysis, total parenteral nutrition, blood products	Changing circumstances are associated with ambiguities; the family may deny the current applicability of the living will
Health proxy (or durable power of attorney)	Same problems as with living wills	
Legal guidelines		
Substituted judgment		If there are no written instructions, the proxy may act freely but can be opposed by the medical staff
Best interests		
Reasonable decision		
State interest		
De facto proxy behaviors		Conflicts may be referred to:
Proxy desires		An ethics committee
Staff desires		The courts (rare)
Brain death determination	Legally defined	Families may still demur on vaguely worded religious grounds
Autopsy	Usually no patient	Family demands are

	instructions	respected
Organ donation and transplantation	Patient’s desires should be honored	Patient’s choice may be overridden by family, absent formal directives;
Financial and legal decisions	Formal documents are binding	none of these decisions can be overturned without recourse to legal intervention

Autopsy and Organ Donation. A patient’s advance interdiction of autopsy or organ donation will not be overridden. However, in the absence of instructions, families are asked to make the decision. It can be upsetting for some families, especially the process of harvesting organs from a body that is brain dead and still appears alive. Even without pre-existing religious objections, some families find it difficult to allow. Ethicists have weighed concerns about the moral acceptability of pre-death procurement as decided by surrogates in the absence of clear patient instructions. Although there is an appropriate firewall between those with a stake in organ transplantation and those with direct responsibility to the patient, more and more people are aware that refusal deprives unknown individuals of a chance to live. A few sensitized family survivors may experience painful conflicts and may harbor residual guilt.

Table 27.9–12 summarizes this information and points out the extent to which all of these categories are subject to interpretation and argument. True to current society, only the instructions about financial and legal matters are unequivocally binding. The thicket of decisions confronting families today would bewilder a time traveler from as little as 50 years ago, who would glance at the patient and know it was time to mourn—nothing more, nothing less.

CULTURAL, SPIRITUAL, AND RELIGIOUS ISSUES

Culture and religion help shape the ways in which people make meaning out of suffering, loss, and death, and the ways in which they understand the sick role, the caregiver role, and every other basic role, such as a parent, partner, child, or sibling. When clinicians share the same unconscious assumptions, decisions can be negotiated more easily. Major differences still occur but play out against the same background music. The consultant should be attuned to the differences—not only intercultural differences but also the ways in which given patients, or family members deviate from their intracultural norms. It is a tall order but well worth the effort in decreased psychological stress, smoother decision making, and improved care. Furthermore, the knowledge gained in one case later eases the management of other cases with the same background. Many institutions have formal seminars or protocols to educate staff about recurrent cultural or religious scenarios. Table 27.9–13 describes a model for the evaluation of cultural issues.

Cross-Cultural Barriers

Cross-cultural problems can be manifested in multiple ways, the most important ones of which are discussed here.

Trust. It is more difficult to establish trust across wide cultural barriers. Some minority groups are thought to want more aggressive care, use fewer DNR orders, and participate less in organ donation and research because they do not trust medical caregivers to treat them as well as they treat the majority group. Some mistrust is justified. There are significant differences in quality of communication, access to care, and quality of care that result in well-documented poorer health outcomes for minority populations. However, when trust is present, the differences are often minimized.

Table 27.9–13.
Evaluation of Cultural Issues: The ABCDE Model

*A*ttitudes: Elicit *a*ttitudes about illness, family care, and responsibilities.
*B*eliefs: Explore *b*eliefs and ritual practices.
Ask direct questions about faith and religious experience; follow up on cues. “What does that mean to you right now?”
Do not respond to statements of religious and spiritual values with facts.
Reflect them respectfully.
Do not deny or pretend to go along with expectations such as miracles but use “I hope” or “I wish” statements that indicate a shared desire for the best outcome that can be arranged.
*C*ontext: Inquire about the cultural *c*ontext in which the patient and family operate.
*D*ecisions: Determine their *d*ecision-making style.
*E*nvironment: Learn about their current *e*nvironment.

Data from Koenig BA, Gates William J. Understanding cultural difference in caring for dying patients. *West J Med.* 1995;163:244.

Truth Telling. The requirement for truth-telling is so embedded in U.S. law and custom today that it may be hard to remember how foreign it can be to other cultures. Surveys of Chinese, Korean, Japanese, Ethiopian, Mexican American, and Native American groups found that only 35% to 52% saw end-of-life truth-telling as desirable. Furthermore, truth in prognostication is not well practiced even in western medicine. Not all physicians prognosticate truthfully, and some may be using good judgment by respecting a strong family culture. It is also useful to remember that nonverbal communications are more highly valued in other cultures. The ability to intuit what the other person is thinking, and to address concerns without too direct a confrontation, is appreciated.

Family Involvement in Decision Making. Families are always involved in decision making but, in some cultures, there may be a strong duty to completely relieve the patient of that burden. This custom is independent of socioeconomic status. In a study of Japanese American families, the custom was independent of the degree of acculturation as well. A family whose request for nondisclosure is ignored may feel upset, angry, and disrespected. If the patient gives informed refusal, the family can be accommodated. However, if the patient is ambivalent or appears interested, psychiatric consultation should be requested and must include an evaluation of cultural and religious factors as well as patient and family dynamics. The family should be educated about the patient's situation, and their concerns about disclosure should be explored nonconfrontationally with indirect questions about the consequences. "How will you feel if you cannot answer your father's questions?" "What will your mother think as she becomes sicker?" When the physician feels strongly that the patient must be asked to participate, it may be wise to also call an ethics committee consultation, for which the consultant's analysis is a useful document.

Mrs. M came from a traditional Chinese family and did not speak English. When she was found to have advanced carcinomatosis, her son and husband refused to tell her about it and were indignant at the thought that anyone would upset her with this information. Consultation with an acculturated Chinese physician from the same subgroup confirmed that their reaction was the standard one. He interviewed the patient and found that she had decisional capacity, had no interest in her medical condition, and shared her son and husband's views on the nature of family support. This was considered an adequate informed refusal and, although she was periodically revisited by the same physician, her demeanor never altered.

Such cases arouse discomfort in medical staff well indoctrinated in the rules of informed consent. However, assuming that a westernized moral climate applies to everyone has its dangers. It is sometimes wiser to respect family mores.

Religious Issues

Every religion has beliefs about death, and every culture has highly valued rituals to honor them. They play a major role in family expectations and reactions. Every hospital or hospice staff should be familiar with such customs in the major religions that they will encounter. They should have access to consulting appropriate clergy to enable them to understand and respect family reactions. The humanizing trends described above have made this more commonplace.

STAFF–FAMILY INTERFACE

In any inpatient setting, medical caregivers are an integral part of the terminal scenario. Their words and nonverbal cues are scrutinized, giving their conscious and unconscious biases an exaggerated impact on events. Staff members can become overinvolved with a patient and family, can scapegoat them, or can overestimate or underestimate their coping abilities, all of which negatively impact care.

In the hospice setting, patients, families, and staff are already all aware of the outcome and do not bear the burden of the many transitions and secrets that characterize the modern death trajectory in hospitals. Nevertheless, continued motivation without burnout and without developing dysfunctional norms requires vigilant attention, and hospices have had built-in psychological support from their inception, which probably explains their lower turnover rate of personnel.

In hospital settings, however, staff members must make constant transitions from day to day and from patient to patient on a given day. They must repeatedly shift set and help patients and families struggle through painful changes. Fortunately, in today's setting, psychological support is available to staff as well as patients, often within case discussions that now include psychological as well as medical issues.

PEDIATRIC END-OF-LIFE CARE

The death of a child always feels like a crime against nature. The biologically determined instincts that drive some adults to care for sick children may endow pediatric wards with a particular intensity that must be channeled with experience and training, but it also provides the children with the responsiveness they need. Medical caregivers must also be aware of the child's level of development and request psychiatric help early if there is a question of cognitive or emotional problems since they will interfere with the child's ability to handle the treatment and further course.

Their parents will remain distraught, although some may acquire strength in the process, while others will require ongoing attention. In all cases, *a cooperative, supportive connection between caregivers, the family members, and the child must be maintained throughout.* This tripartite connection helps the child feel more secure, while the family appreciates the consideration and support, and the caregivers can be, and feel, more effective.

ETHICAL RAMIFICATIONS OF THE TRIPARTITE CONNECTION

The requirements of the tripartite connection define the entire course of treatment and hinge on three values that must always be present: *trust, honesty, and confidentiality*. Their importance revolves mostly, though not exclusively, around three main issues: (1) The parents are facing the loss of a child to whom they gave life, which, along with grief, induces a sense of irrational guilt that cannot be fully erased and may drive them to make extreme demands against the child's interests. (2) Legally, pediatric patients are always third parties, ranging from helpless infants to 18-year-olds who may be wiser than their parents yet not free to make their own decisions. (3) The medical world had, until recently, not focused much on the fact that adolescent maturation is as relentless as breathing and continues its struggles to survive under a dark sky. Parents able to help younger children may find it more difficult to honor the adolescent's needs and require more information, explanations, and support.

Trust and Honesty

This creates the need for what was well characterized in this quote: "The goal from the outset is to establish a dual treatment relationship with both parent and child with a common goal of understanding the child's thoughts and feelings and improving the child's well-being. *It is necessary to establish a three-way message of trust."* (see Kearney J.) Physicians who regularly interact with the parents and child carry a special burden. The situation demands that they earn everyone's trust and honesty when meeting with children and parents, separately or together.

Confidentiality

Negotiating confidentiality is a delicate matter and is dependent on doctors' communication skills. Having addressed the need for full honesty, they are obligated to begin by explaining to the parents the legal limits to confidentiality when safety is at risk. Next, they must show the kind of understanding and support that convinces parents to trust that they will not lose touch with their child, even if they are not told all the thoughts the child has shared. Physicians must further explain that they *need* that trust because the child will only trust them if they can promise to use appropriate discretion when talking to the child's parents. Finally, doctors should describe their own determination to always keep in mind the priority of sustaining the parent–child relationship.

These meaningful issues must be addressed from the beginning of treatment, as gradually as necessary, when there is time for the parents to adapt incrementally to the more difficult times that lie ahead. It may require a psychiatrist since parents are such an important part of the patient's treatment that they need the same level of attention for the good of the child. Another potentially difficult position arises when parents, who have the legal right to make many decisions, are obstructing the best course. This is one more reason why it is critical to create a trusting collaboration from the beginning.

DEVELOPMENTAL STAGES

Infants and Toddlers

It is hard for anyone to watch the suffering of infants and toddlers, but while staff can move on more easily, the pain of parents is inherent in their parenthood. Giving birth implied giving life, and they are now failing and helpless. They need support to remain stable and keep providing the loving presence the child expects and needs. The medical team will do their part for the child's comfort and will continue to sustain the parents, although further help is often needed.

School-Age Children through Preadolescence

Within this age limitation, children vary widely in their capacity for absorbing what is happening. Adults often underestimate children's understanding, particularly in a crisis that draws out their most protective responses. However, the reality is that, like sponges, children listen intently to what they hear, soak up the emotions that flow around them, and subject everything to an active process of interpretation, organized around finding out why they are being left out. Punishment is often their frame of reference, which they may elaborate on in various ways. When this happens, the child has unwittingly been made to carry what should be the parents' burden.

These suffering parents should be helped gently from the beginning to understand more about their child's response to their situation and should learn to speak to the child in age-appropriate ways, surrounded by reassuring affection. A knowledgeable individual must review the child's level of understanding so as to acquaint parents with appropriate ways of responding, explaining to them that these interactions are as fundamental to parenting as the care they had given their infants. On today's wards, parents must be given special support, which is one reason why many pediatric wards have their own support staff, including psychiatrists and psychologists.

The Child's Cognitive Understanding of Death and Dying

Understanding children's developmental stages and their level of insight about terminal illness has practical implications for the clinician. It provides a basis for effective communication with children about their illness and guidelines for improving their coping skills. Although children's reactions depend on their personality and level of development, they are also strongly influenced by their emotional family milieu and must be understood within that context.

Children's reactions to their own dying or to the death of others follow a developmentally determined conceptual framework, enriched at times by their own interpretations of familial conversations. Accurate comprehension of death requires the mastery of four basic concepts: irreversibility, finality, causality, and inevitability (or universality). These are detailed in Table 27.9–14 and should be familiar to adults dealing with children facing severe illness and death because the children’s attempts at self-explanation make their course worse than it needs to be.

Table 27.9–14.
Concepts of Death and Implications of Incomplete Understanding for Adjustment to Loss

Irreversibility: Death is seen as a permanent phenomenon from which there is no recovery or return.

Example of incomplete understanding: The child expects the deceased to return as if from a trip.

Implication: Failure to comprehend this concept prevents the child from engaging in the process of detachment from the deceased, a necessary first step in successful mourning.

Finality (nonfunctionality): Death is seen as a state in which *all* life functions cease completely.

Examples of incomplete understanding: The child worries about a buried relative being cold or in pain; the child wishes to bury food with the deceased.

Implication: May lead to preoccupation with the physical suffering of the deceased and may impair readjustment.

Inevitability (universality): Death is seen as a natural phenomenon that no living being can escape indefinitely.

Example of incomplete understanding: The child views significant individuals (i.e., self, parents) as immortal.

Implication: If the child does not view death as inevitable, they are likely to view death as a punishment (either for actions or thoughts of the child or the deceased), leading to excessive guilt or shame.

Causality: The child develops a realistic understanding of the causes of death.

Examples of incomplete understanding: The child who relies on magical thinking is apt to assume responsibility for the death of a loved one by assuming bad thoughts or unrelated actions were causative.

Implications: Tends to lead to excessive guilt that is difficult for the child to resolve.

Reprinted from Lewis M, Lewis DO, Schonfeld DJ. Dying and death in childhood and adolescence. In: Lewis M, ed. *Child and Adolescent Psychiatry: A Comprehensive Textbook*. 2nd ed. Lippincott Williams & Wilkins; 1996, with permission.

Adapted from Schonfeld DJ. Crisis intervention for bereavement support. A model of intervention in the children's school. *Clin Pediatr (Phila).* 1989;28(1):29. doi: 10.1177/000992288902800105.

During the first few years, children have almost no concept of death other than a disappearance or separation. Most children under the age of 5 years do not think of death as final and believe that the departed person is somewhere else. The most commonly offered explanation is heaven. Worried about being separated from their caregivers, young children may ask, "When I die, will Mommy come with me?" However, even children under the age of 5 years develop a precocious understanding of death if they experience a traumatic event like the death of a parent or sibling or any event exposing them to death. Children between the ages of 6 and 9 years understand that death is final and entertain concrete thoughts about the process of death and dying. For example, an 8-year-old child may conceptualize death in terms of damage to a particular injured body part. Stories of devils and ghosts may frighten children, and they may have vague ideas about the soul continuing to live on after death. A 7-year-old child suffering from terminal Ewing's sarcoma asked his mother, "When I die, will my leg fall off and stop hurting?" Children aged 10 years and older have a fairly good understanding of death. They begin to see patterns in nature, recognize life cycles, and often become aware of their own mortality. Some are quite precocious and could qualify as young adolescents.

Younger children may resort to themes of retribution and magical thinking when reacting to their illness. Perceiving their suffering as a punishment for real or imagined misdeeds, they accept the burden of personal shame or guilt to account for it. They view their illness in an egocentric manner and express their fears accordingly: "Did I cause it? Who is going to look after me? Am I going to be abandoned?" An older child or adolescent, aware of the irreversibility of death, may exhibit anger, anxiety, and depression while trying to cope with the terminal illness. However, some children inspire respect by their courage and stoicism in the face of death.

Communication with Severely Ill Children

Children quickly learn to whom they can trust and confide. Whether or not individual children should be told about their prognosis is not a simple matter. Parents have psychological, moral, and legal authority. Their wishes must be explored earlier rather than later. Some parents do not want their children to be told they are going to die. It is useful to ask them *how they would respond if the child asked them if they were dying.* It should be the opening of an ongoing conversation, aiming to help them understand more about their particular child's response, the fears, the sadness, and the need for *authentic* reassurance. When some families are willing to tell the child, one individual should be selected to talk to the child, but only after discussions with the physician. Sometimes, the physician will be asked to be present during such conversations or might even be asked to be the one to speak to the child. The same information must then be conveyed to everyone close to the child because inconsistent explanations are confusing and even emotionally crippling. In such cases, an experienced child psychiatrist or psychologist can be helpful early in the process.

When the child asks a question about death, it is very important to first clarify what the child already knows and the reasons behind the question since they may be very different from what adults expect. Some children are responding to the anxiety and grief projected by their parents; others are preoccupied with pain, discomfort, mutilation, abandonment, and loneliness. The child should be given ample time and opportunity to express concerns, and any frightening misconceptions should be clarified. Even when children are told their illness is fatal, they should be comforted and reassured that the physicians will do everything possible to keep them comfortable and that they will always be loved and cared for. If the physician expects a difficult session, it may be useful to ask a child therapist to attend. In fact, early referrals are frequent and desirable.

Communication with Parents

Communicating with the parents of a very sick child follows the principles used in families of adult patients, save for one crucial difference: the parents' deep sense of responsibility for the child's welfare and the attendant feeling of failure. Some parents are too overwhelmed to function. Others throw themselves into research, taking the child for multiple opinions and treatments. Still, others never leave the child's bedside, which is helpful for younger ones but may dominate the child's relationships with other caregivers. Constant tasks leave them little time for thinking or feeling. A few may actually remove themselves. Fortunately, pediatric wards still witness remarkable feats of loving accompaniment in the face of increasing grief and exhaustion.

Parents often need guidance and support, particularly about whether they should treat the child normally or with special attention, especially when they have other children. Intensely protecting a young child beyond what the illness requires confuses the child, who has no framework to explain this novel behavior. The older child and the adolescent understand its basis; some welcome this protection, and some crave and require it. Others prefer more freedom and privacy, often acting out a fantasy of being well or finding a means of denying the fatality of their disease. The advice and support of an experienced child psychiatrist can help parents strike the right balance.

Parents may demand aggressive medical treatment because they cannot bear the thought of future regret over any omission. These feelings may not be openly voiced to the physician. They should be sought for and discussed openly if any other option is to be considered.

Finally, as parents learn to accept that their child will die, they reach a state of resignation that spares them the previous bouts of activity, ushering in a different kind of suffering and anticipatory grieving. In this exhausted state, one or both parents may find themselves wishing the child dead. These unacceptable secret thoughts burden them with excruciating guilt. They should be *sought* *for and clearly normalized.* It is a major gift to relieve them frequently of these attacks against themselves.

PSYCHIATRIC SYNDROMES IN THE TERMINALLY ILL CHILD

Although psychiatric problems in dying children are similar to those seen in adults, their presentation varies depending on the child's stage of development. The young child's ability to report symptoms of physical or emotional discomfort is generally undeveloped or unreliable. Other young children may not wish to disclose their suffering out of unwarranted shame or guilt from real or imagined wrongdoing. Clinging to magical thinking, these children may fear that voicing their complaints will worsen their symptoms. Others may be the opposite and will use a lot of energy on loud complaints, in which the line between physical and emotional is hard to distinguish. A few children may assume that if they wish to be healthy and think "good thoughts," the problems will go away. All this leads to withholding of information and compromised care. Older children may not be much more forthcoming. An astute physician knows to look for symptoms even if the child is not vocalizing them. Consulting with child therapists is useful in such cases.

Symptoms of delirium, depression, and anxiety occur in children with a terminal illness, and their causes are physiologically the same as in adult patients. Children with pre-existing psychopathology are more vulnerable to developing psychiatric syndromes as they confront terminal illness, but the symptomatic treatment is still evaluated the same way as with adults.

Delirium in Terminally Ill Children

As in adults, symptoms of mild delirium may be overlooked or mistaken for regressive behavior in a young child. Because a young child's world normally fluctuates between fantasy and reality, symptoms of visual or auditory hallucinations may be viewed by the parents as the child's fantasy rather than as a delirious state of altered perceptions. Disoriented states may be overlooked or minimized because children's attention to the items measured is variable at best. In response to this, Paula Trzepacz has adapted her delirium scale to children and adolescents.

As with adults, treatment of delirium requires an immediate institution of symptom management, along with a search for the etiology and correction of the underlying cause if possible. Neuroleptics, such as haloperidol or droperidol (Inapsine), as well as the newer antipsychotics, may be used to treat the symptoms. Benzodiazepines are usually avoided because of the associated disinhibition. In addition to this treatment, reassuring and reorienting the child by placing familiar objects and people in the room is helpful. Psychoeducation of parents can relieve fears and misconceptions and allow them to better support the child.

Depression in Terminally Ill Children

Depressive symptoms are expected in children battling a life-threatening illness. One may see anticipatory grief and mourning for the perceived loss and impending death or behavioral problems, such as regression, opposition, aggression, withdrawal from family and friends, and noncompliance with medical care. Very few children develop a full-blown major depressive disorder that requires pharmacologic intervention. As with adults, vegetative symptoms, such as fatigue, anorexia, and insomnia, are so common in terminal illnesses and major depression that it is impossible to tease them apart.

With adolescents, it is important to explore suicidal ideations and fantasies overtly. In their struggle with a bleak and uncertain future, some harbor suicidal thoughts as a means of gaining control over their destiny. To judge the level of risk, the consultant must explore the youth's understanding of what suicide means, the personal motivation for it, and the presence or absence of a plan. Known risk factors include the presence of previous or comorbid psychiatric disorders, persistent and intractable pain, and poor social supports.

Treatment of depressive symptoms begins with helping the families make the children feel confident that they will not be abandoned and that their comfort will always be a top priority. Maladaptive behavior patterns can be explained to the family as the child’s way of coping with anger and grief. Some parents can help the child process the myriad emotions that assail them in the face of impending death. Exploring the child’s feelings through play or drawings can be liberating. Providing children with an open forum to talk about their concerns, beliefs, and fears about death helps them emotionally process what is happening and feel less alone. Some courageous children decide to whom they want to leave their belongings and say personal goodbyes to friends and family in a variety of ways. This authentic and generous engagement can be empowering, but it *must* come entirely from the child, suggesting it could be very upsetting to an unprepared child.

Antidepressants may be necessary for a minority of children with major depressive disorder, but children should also be receiving active treatment for pain, fatigue, insomnia, and any other treatable symptoms. The etiology should be sought while treatment is instituted.

Anxiety in Terminally Ill Children

Anxiety in the face of death is ubiquitous. It may stem from physiologic issues such as intractable pain, compromised respiratory state, or medications, as well as from psychological issues such as separation from parents, the unfamiliarity of the hospital environment, and fear of death. One should always treat physiologic symptoms while helping parents treat the symptoms due to the child’s own observations, experiences, and interpretations. Ensuring consistent contact with parents, avoiding prolonged separation, addressing the emotional states of caregivers, and making the hospital milieu more child-friendly go a long way toward assuaging the child’s anxieties. More severe symptoms may require the use of anxiolytics. Short-acting benzodiazepines are preferred to minimize sedation, respiratory compromise, and disinhibition. Child psychiatrists and other child therapists should be effectively used.

Pain Management in Terminally Ill Children

Pain assessment in children is more complicated than in adults. Children who lack verbal skills cannot communicate pain to caregivers. Young verbal children may distract themselves from pain with play, some recoil into sleep, and some remain withdrawn and anxious. Adolescents may not be forthcoming about pain due to the belief that they are invincible or due to a misplaced fear of addiction.

Yet, with the use of appropriate rating scales, children can be reliable reporters. Age-appropriate and validated scales do exist, and more physicians are now familiar with them. As with adults, it has been found that professionals using clinical judgment assess the child's pain level as lower than do both the parents and the child. Worse yet, the greater the child's pain, the greater the discrepancy.

Most parents recognize their child's pain and are eager to see it relieved, but a few mistrust the medication to the point of forbidding it. Some are worried about addiction but accept education and support. A few are unable to tolerate seeing their child relax and become less related because they do not experience it as relief for the child but oncoming loss for themselves. With PCA, they express fear of overmedication. Yet, it has been demonstrated that children old enough to understand the technique showed significantly diminished use of opioids with PCA. Family education around these issues is important, but often a consultant must ease the parents into realizing how much their reactions are a way of avoiding the truth of their child's condition. Only after they confront the reality can they wholeheartedly cooperate with optimal pain management.

COMMUNICATION WITH ADOLESCENTS AND YOUNG ADULTS (AYA)

Introduction

One of medicine’s strengths is its focus on the individual and the positive attention it brings to each patient’s care. The earlier discussion of the close interaction between medical and psychological management for children and adults need not be repeated here because adolescence overlaps both categories, as will be discussed below. This section will discuss recent developments that specifically focus on adolescents and young adults.

Current Interface of Hospitals and Adolescents

Only in the last few decades have medical hospitals thought about special arrangements for adolescents since they could still be assigned to either pediatric or adult wards. Yet, human societies have always been aware of the adolescent trajectory as a meaningful period for them and society. There have been increasing numbers of specialists over time, but only in the last two decades have hospitals introduced dedicated spaces and trained individuals to work with them. They have been well utilized by terminally ill adolescents who suddenly have to integrate the most profoundly threatening experiences while still making meaningful relationships in the process.

AYAs’ Development during Their Illness

Adolescents and young adults, now referred to as “AYAs” by researchers, are a group of patients that should not be defined solely by age but by the level of development. It could include a precocious 12-year-old or a solitary 25-year-old. An early hospital study found that parents’ protective instincts were as powerfully aroused with older as with younger children, which made them cling more strongly to their parental decision making. This observation encouraged caregivers to help parents understand and respect their adolescents’ growth. It confirmed what was being increasingly noted, that relentless exploration toward adulthood continued throughout their illness to the very end. This means recognizing that their grim situation does not stop them from thinking about making new peer relationships in the hospital, still focusing on sexuality or at least intimacy, evaluating the life goals lost, maybe finding an appropriate new one, and last, despite—or because of—their misfortune, sometimes achieving unexpected levels of intellectual and moral development.

Adolescent psychiatry now plays a well-established role in medicine. Large medical centers provide gathering places where only adolescents, including gravely ill ones, can share their experiences with each other and where adolescent psychiatrists, psychologists, and teachers have roles to play. The main purpose is to allow adolescents, including dying ones, to still experience adolescent development. This is an excellent change, but it still is not always available in small hospitals and rural areas, where the smaller number of patients might require a different approach.

It is best to begin as early as possible to help AYA parents help their children to remain themselves. This help includes assessing AYA patients' readiness to discuss their disease, with a plan to help them understand and participate more and more—to the extent that they wish. Over time, their ability or willingness to confront their impending death should be assessed and explored, and if any of them show any ability to deal with it, they should be helped and encouraged. For some adolescents, it is deeply meaningful to write letters, leave gifts, or give instructions about future events such as funerals and posthumous remembrances of their birthday. Others simply hang on to close connections, leaving life on a note of fundamental love and attachment.

A group supported by the National Institute of Mental Health has been studying this area at length. They published the document *Voicing My CHOiCES* based on their research as well as feedback from AYA patients themselves. It guides health care providers to help patients and their families navigate this painful and complex process, using individual sections which AYAs can fill out with the support of health care providers, according to their comfort level. *Voicing My CHOiCES* provides a thoughtful and thorough discussion of how AYA patients can be helped to explore the important issues, although it also makes them aware of options at every point. When queried, over 90% of patients reported feeling positive about participating in the process.

PSYCHOLOGICAL ISSUES OF HEALTHY SIBLINGS

Studies of healthy siblings have found increased levels of regressed behaviors, somatic complaints, altered peer relations, and a drop in scholastic achievement. They have also found higher anxiety levels and lower self-esteem. Children are at the mercy of their environment, and it is easy for anguished adults to forget that siblings still require the same nurturance and now also have additional needs. Caregivers who notice any problems should help parents develop constructive responses. Professional consultation may be required for complicated situations or dysfunctional families.

Parental Response to Healthy Siblings

The illness and death of a sibling is a world-transforming event. It permanently changes family dynamics and sometimes leaves the family in financial straits that diminish the surviving child's future opportunities. Acutely, it profoundly disrupts the family routines and captures the adults' time and attention, leaving them emotionally depleted. Some parents are strong enough to remain available to their other children, but many react by protecting them (i.e., keeping them uninformed and away from the family tragedy). Sadly, the patient's siblings are marginalized at a time when they most need to feel like integral members of the family. Keeping healthy children engaged in the family drama requires an understanding of the child's developmental level and emotional state so as to know what will help them feel loved, trusted, and included. Optimally, the staff caring for the patient will help parents maintain a closer and more open relationship with their other children throughout the illness.

Children's Understanding of a Sibling's Terminal Illness

Siblings 5 years old and younger do not understand the terminal nature of the illness and may react negatively to the parent's emotional state. If they sense withdrawal and sadness, they become frightened and may regress or act out, hoping to gain parental attention. Children between the ages of 5 and 10 usually understand their sibling's condition but may become anxious about being abandoned by a distraught parent as well as fearing that harm will befall them. Some children are plagued with guilt because they have been jealous and angry about the attention given to the patient and remember past times when they harbored angry feelings and even wished the sick child was dead. Younger children may believe that this wish was magically fulfilled and thus blame themselves for the course of events. On the positive side, given both reassurance about being loved and age-appropriate information such as well-related options on how to relate to the ailing sibling, many children become eager to show their support by making drawings, bringing special toys and treats to the hospital. With parents' approval and affection, this fruitful participation in family events significantly enhances the sibling's self-esteem and sense of security and may help the sick child as well.

Adolescent siblings can become supportive, to the point of assuming parental roles toward younger siblings and taking over household tasks. When the family members begin to take it as a given, it may put these youngsters in a position that will interfere with their development. However, in today's world, it is increasingly likely for adolescents to withdraw from the family, acting out or devoting more time to being with their friends or just spending more time with their eyes glued to a screen. In any case, they too will need help, whether to prevent them from continuing their parentified role beyond the death or to help them return into the family fold in time to manage their feelings toward their dying siblings (with help, if needed), and experience the shared affection as well as the grief.

STAFF DYNAMICS IN END-OF-LIFE CARE

The family–staff interface in pediatrics and the associated trust and communication needs are the same as with adult patients, but there are a few special issues associated with the staff. Pediatricians and nurses choose to work with children because they love them. During an increasingly severe illness, they come to know the child well and may become very attached. The expression of affection toward a pediatric patient by caregivers need not have the inappropriate implications it has for adults, as long as the medical judgment is unimpaired. Ideally, the staff and the family remain united by their common dedication to the child, but occasionally the attachment is so strong that it creates conflicts. For example, a nurse's maternal instincts can engender an unstated competition with the child's mother, which makes her less supportive and affects other staff as well. The mother is usually keenly aware of it but sometimes afraid to mention it. This situation requires a change of assignments for the nurse, with special support and training to prevent recurrences.

Sometimes, staff members are quick to pass judgment on families that are not meeting their expectations. Once crystallized, these judgments can be very hard to revise. Negative stereotypes and competitive feelings are very destructive. They cannot be hidden from parents, who are exquisitely sensitive to staff behaviors and utterances. It is critical that team leaders set the tone and create a culture in which children and parents are *both* treated with equal empathy. Consultants must remind the staff how much the children want to feel securely embedded in their families at all stages of illness. Children's profound dependency on their families is such that they are enslaved by faithfulness, even in the face of parental indifference or abuse. *Any staff behavior that threatens family unity is unacceptable, no matter how well-intended*.

Staff overinvolvement also creates Intra staff rifts. In one case, an oncology fellow became so involved with a child that he argued heatedly over the attending's choice of treatment. In another case, two nurses became so attached to a patient that they arranged their work schedule to have one of them on duty every day because they felt their colleagues would not care for him well enough. It is not that unusual for young pediatricians to share some of the parents' feelings and, occasionally, to become more involved than they had intended. These doctors must learn to exercise control with the parents and would benefit greatly by sharing their feelings with a trusted colleague. In general, this is why group work and open discussions have become a larger part of the climate when working with terminally ill patients.

END-OF-LIFE DECISIONS IN PEDIATRICS

Assent and Consent

As children mature and are better able to process information about their illness, they can play a more active role in decision making. However, with the exception of emancipated minors, the child can only assent, while it is the parent who can give legally binding consent. This is usually not a problem because most parents respect the wishes of the dying child, but there may be times when discrepancies arise.

One scenario is that of adolescents refusing treatment that their parents, and sometimes their physicians, would like to administer. This situation calls for psychiatric consultation to allow all participants to explore their feelings and to help achieve consensus about the care plan. Consultation with an ethics committee may be necessary when an agreement cannot be reached.

Another scenario is that of an exhausted or frightened parent who refuses treatment that the staff believes a child would accept were they not afraid to disagree with the parent. It is more likely to happen if the staff has not achieved an easy, supportive relationship.

The single mother of a 9-year-old boy with relapsed leukemia refused a bone marrow transplant on religious grounds. The nursing staff dealing with the child felt that, at age 9, he was aware and would not have minded the treatment but was afraid to disagree with his mother. The consulting psychiatrist found that the mother felt overwhelmed and unsupported by the staff. Her religious feelings, although genuine, were also a way for her to escape the situation. The consultant urged the physicians and nurses to hold off on requesting any decision while being very supportive of the mother and validating her apprehensions and pain. Within 3 days, she was ready to proceed with the treatment, which the child accepted readily.

Neonatal and Infant Issues

Neonatologists can now save premature babies who weigh little more than 1 lb and are at high risk for intracranial hemorrhage or respiratory insufficiency. They can also salvage severely handicapped newborns who will always need intrusive medical care and whose quality of life will remain poor. Some people believe life-sustaining interventions should always be used, others believe they should never be used, and a large group thinks that it depends on circumstances. Among physicians who are not absolutely opposed to withholding life-sustaining measures, the relevant quality-of-life issues are: (1) the extent of physical impairment, such as severe neurologic damage; (2) the burden the child will place on the family; and (3) the ability of the child to derive pleasure from existence, such as having an awareness of being alive or being able to form relationships. From a philosophical point of view, these are very disparate arguments, but they derive from the pragmatic and casuistic method that has characterized valid medical thinking.

WHERE WE STAND NOW

Decision Making

Truth-telling and obtaining informed consent are legal obligations. The intellectual requirements for having the capacity to consent are well elaborated. First, the patient must be provided with information about the diagnosis, prognosis, and treatment options, as well as a review of the risks, benefits, alternatives to treatment, and the consequences of no treatment at all. Processing this information can be difficult, especially under the dual burden of severe disease and the nearness of death. Given the high incidence of cognitive clouding in very ill patients, every dying patient's right to make decisions might be imperiled if one focused only on cognitive criteria when establishing decisional capacity. Consultants must come when patients are most alert and lucid and *must look for an internally consistent sense of self and general awareness of their condition*. If present, these should weigh more heavily than the patient's ability to perform a standard mental status examination. This is not to imply that the consultant should be less aggressive in correcting the patient's cognitive state by all the usual means.

Informed Consent for Experimental Treatment. It is difficult for patients in relatively good health to absorb the basic facts about clinical trials. They may agree to the trial to please their physician or may believe that the physician is offering them a place in the trial because it is good for them. Reality #1: doctors continue to have a personal duty to the patient even if they enroll them in a trial; they must remove the patient from the trial if the patient's welfare is incompatible with the demands of the protocol. Reality #2: patients need to keep believing their doctor is serving their needs whether they are in a trial or not; this is even more intense for terminally ill patients who have become known as *desperate volunteers,* and who enter trials "because" it is their only chance—and indeed, their desperation cannot be assuaged! Its presence places an even greater ethical burden on both physician–researcher and Institutional Review Board (IRB). If both parties do their jobs, the psychological and ethical ambiguities inherent in working with desperate volunteers should not deprive the patients or society of the potential benefits of valid research.

An article on the ethics of phase I oncology trials observed that the risk–benefit ratio of phase I trials is not significantly worse than those approved by the FDA for clinical trials of chemotherapy. It identified three important issues relating to informed consent in terminally ill volunteers: (1) disclosure, which available data show to be adequate; (2) understanding, which also has been found acceptable in more than 70% of patients; and (3) voluntariness, which was initially strongly denied by moral philosophers. The authors argued convincingly that patients with advanced cancer have a different perspective than the healthy bioethicists who worry about them and feel they are being abused. Many dying patients are known to be truly yearning to try more low-yield treatments the closer they are to death.

Ethical Issues about End-of-Life Treatment

There are possibilities of management that are unique to end-of-life care. They include withholding treatment, withdrawing treatment, and invoking the double effect.

All of these were legitimized in 1982 when the Belmont Report, produced by a presidential commission and supported by major Catholic and Protestant doctrinal statements, affirmed the moral acceptability of withholding extraordinary measures or of stopping them later. It also described acceptable uses of the double effect.

However, it has proved to be surprisingly difficult to define *extraordinary* versus *ordinary* measures in the context of withholding or withdrawing treatment. If food and hydration are *ordinary*, what of nasopharyngeal tube feedings? What of IV fluids or percutaneous feedings? Perhaps everyone can agree that TPN is extraordinary. Cardiopulmonary life supports are generally accepted as *extraordinary*, but what of intubating a patient in worsening respiratory distress without cardiac arrest? What of raising vasopressors? What of the so-called temporary addition of dialysis? What of blood products? The persistent controversies can be understood better if one asks two fundamental questions:

1. *From what do patients die?* Do they die from their disease? Or because dialysis was discontinued? Or because they received morphine to relieve respiratory distress? Or because they were not resuscitated? Or because life supports were discontinued in the face of irreversible organ failures?
2. *To whom does life belong?* If it belongs to a higher being, many believers feel that it is held in trust and cannot be shortened. The most extreme of these vitalists feel that all available methods of treatment must be used, even when survival would be miraculous. However, if it belongs to the individual, or even if the family, tribe, or state has a claim, decisions can be made according to personal preferences, social mores, or legal criteria.

Table 27.9–15.
Range of Terminal Care Treatment Decisions

Never diminish comfort measures unless they have become uncomfortable!
Stopping
Curative treatment
Invasive procedures and treatments
Withholding
All diagnostic procedures
Blood products
Antibiotics
Total parenteral nutrition or dialysis, or both
Cardiopulmonary resuscitation
Invoking the double effect, if indicated
Withdrawing
The previously mentioned treatments
Life supports
Tube feedings
Intravenous hydration
Oral feedings

The fundamental incompatibility between these views is illustrated by looking at the range of end-of-life measures in Table 27.9–15. The vitalists may object to every measure, whereas those who believe that life belongs to the bearer may view them all as a matter of individual choice. A dedicated vitalist will believe that death resulted from the immediate action, such as discontinuation of dialysis, rather than from the irreversible disease process. They would refer to withholding treatment as *passive euthanasia* and to withdrawing it as *active euthanasia*. On the other hand, the nonvitalist would be comfortable with the use of slowly escalating doses of morphine to relieve terminal distress, even if it hastens death.

As the debate has become more heated, loaded terms have been used, such as *slow euthanasia*, which has been used to describe opiates for effective pain control, or *passive euthanasia*, which has been used to describe the withholding of invasive life-sustaining treatments. Thus, the vitalist position blurs the distinction between assisted death practices and good palliative care as it is currently understood. On the other side, dedicated proponents of assisted death have supported actions and policies that cause uneasiness even in those who are sympathetic to their point of view. Because physician aid in dying is being legalized in a growing number of jurisdictions, it is crucial to avoid polarization of a debate in which good palliative care is ultimately the casualty.

Physician Aid in Dying

The nature of patients' desire for death, outlined in Table 27.9–16, ranges from transient passive thoughts of suicide, through the suicidal preoccupations of major depressive episodes, to active requests for physician aid in dying. Physician aid in dying addresses only the last category, patients whose pain is uncontrolled or whose suffering is intolerable and who ask for help in ending their lives. However, it has been demonstrated that, indeed, many patients requesting help are clinically depressed and must be given the benefit of consultation and attempt at treatment.

Table 27.9–16.
Desire for Death in Terminally Ill Patients

Transient passive thoughts of death

Easily dissolved with improved quality of life

“It would be OK if I did not wake up”

Suicidal ideation in the context of major depression

Persistent, obsessive thoughts of dying and desire for death, associated with excessive guilt, feelings of self-hatred, and worthlessness

Not necessarily associated with a plan

Not responsive to medical and environmental measures

Request for physician aid in dying: assisted suicide and euthanasia

Commonly associated with fear and with hopelessness, with and without depression

Often has a strong component of more or less conscious anger

Physician aid in dying includes physician-assisted suicide (the provision of enough medication to enable a patient to commit suicide) and active voluntary euthanasia (bringing about the death of another person with their consent). Involuntary euthanasia (bringing about the death of another person without their consent) is murder in the United States.

Physician-assisted suicide and euthanasia were legalized in the Netherlands after a long period of having been openly tolerated and regulated. Opponents have pointed to longitudinal studies that show the incidence of euthanasia rising in proportion to physician-assisted suicide and found some cases of euthanasia of unconscious patients in which overt patient requests were not documented, leaving one to wonder about the dangers of the so-called slippery slope.

In the United States, two federal Circuit Courts of Appeals supported assisted suicide, one on the basis of equal protection under the law, equating it to the available right to discontinue life-preserving treatment, and the other on the basis of liberty and privacy rights. Both decisions were halted by injunctions and were reviewed by the U.S. Supreme Court. In June 1997, the Court refused to affirm a constitutional right to assisted suicide but stated a need for further social exploration. In particular, Justice Sandra Day O'Connor emphasized a positive right to excellent pain and symptom management, giving hope to many that palliative care would gain impetus, as indeed it has since then.

While euthanasia remains unacceptable, physician-assisted suicide was legalized in Oregon in 1997 after angry voters and attorneys general faced off over this issue. In response to physicians' reasonable fear of harassment, the California Board of Professional Medical Conduct was one of the first states to publish guidelines to protect physicians. Other professional bodies have done the same. Subsequent history in Oregon has not borne out the fears of a slippery slope. As of this writing, there are at least eight states and Washington, D.C. that have enacted *Death with Dignity* laws, with more potentially on the way. A list of current *Death with Dignity* laws can be found in Table 27.9–17.

However, it is important for physicians to be aware of psychiatric and medical issues which are sometimes overlooked. Delirium, for example, is associated with impulsivity and must be treated. Also at risk are patients with poor pain control, a sense of helplessness and loss of control, and a previous psychiatric diagnosis. Hopelessness as part of a pessimistic coping style has been found to be a better predictor of suicidal wishes than depression without hopelessness. To complicate matters, the *will to live*, as measured with a visual analog scale, was found to fluctuate rapidly, correlating with anxiety, depression, shortness of breath, and other painful or frightening symptoms. Treatment of pain and other symptoms often diminishes or dissolves the desire for help in dying. Provision of social and spiritual support is also powerful. Understanding these issues is more important to the patient's welfare than public hand wringing about the slippery slope.

Table 27.9–17.
Current Death Dignity Laws

States	Death with Dignity Statutes	Year Approved/in Effect
California	End of Life Option Act	2015/2016
Colorado	End of Life Options Act	2016
District of Columbia	D.C. Death with Dignity Act	2016/2017
Hawaii	Our Care, Our Choice Act	2018/2019
Maine	Death with Dignity Act	2019
New Jersey	Aid in Dying for the terminally Ill Act	2019
Oregon	Death with Dignity Act	1994/1997
Vermont	Patient Choice and Control at the End of Life Act	2013
Washington	Death with Dignity Act	2018

Voluntary Refusal of Food and Water

An Oregon report from Dr. Linda Ganzini found that almost one in three nurses had cared for patients who died a peaceful death after refusing food and water. Compared to the assisted suicide group, these patients were less likely to have cancer, in less acute pain, less preoccupied with control, a decade older, and more ready to die. This interesting study highlights a mode of death that is universally available to any competent patient. However, many patients requesting aid in dying feel out of control and are afraid to die in acute pain and solitude. They are anguished and yearn for loving care to heal both pain and suffering. Lacking that, they demand an ultimate act of nurturance that heals by bringing oblivion. By placing a heavy burden on others, they may express both yearning and resentment. Others, however, plan ahead, discussing those plans with loved ones, and finding physicians with whom they negotiate their request in a cooperative mode. Additionally, one should not ignore cases in which patients are driven by the realistic fear of losing the ability to express their wishes.

Caregiver Responses

In the same study, nurses stated that all staff members found it difficult to care for patients who elect to forgo fluids and nutrition. Similarly, Oregon physicians who had responded to a request for physician-assisted suicide described feeling very uncomfortable. They had discussed technical issues with colleagues but kept their conflicted feelings to themselves. They usually concluded that they had no regrets and believed that they had become more comfortable in dealing with end-of-life issues. Some may wonder whether this represents improved palliative care skills or desensitization to death. In any case, the development of palliative care has given physicians an option to deal with such patients to their professional and ethical satisfaction.

Societal Implications

Proponents of legalization emphasize dignity and autonomy, pointing to Dutch studies that name this as the most frequent reason for help-in-dying requests. Opponents point to the lack of good pain management and other palliative care, citing American and Canadian studies that found depression and poor symptom control as the major factors underlying patient suicidal tendency. Given severe retrenchment of health care expenditures, with politicians, insurance companies, and managed care corporations setting the policies, there is cause for concern even among individuals who have no philosophical or moral objection to assisted suicide. African Americans and other minorities—groups that have historically known inferior medical treatment or inferior representation in medical studies—tend to be far more opposed to legalizing assisted suicide than the rest of the population. Associations of disabled individuals also oppose it because they believe that many of them would not be alive had it been acceptable to terminate lives deemed not worth living. All of these groups insist that moral opprobrium and legal sanctions are essential to hold dark human tendencies in check. Independent of religious or philosophical convictions, historical and contemporary examples demonstrate that they cannot be ignored. But the authors do experience relief at the wholehearted adoption of palliative care and the gentle invasion of humanizing trends in medicine today.

The Moral Burden of Physicians

Medical training is inherently casuistic and teaches that each case must be decided on its own merits. Primary care physicians and hospitalists have depended on their professional values to guide them and to prevent abuse more often than not. However, this process worked as well as it did because it was buttressed by a social climate that protected individual life. Many of the choices physicians made were transparent. Decisions about withdrawing treatments and about nutrition and hydration were de facto shared by family members and other caregivers, giving social morality an opportunity to mold physician behavior. Medicating dying patients with drugs that invoke the rule of double effect is less controllable and presents doctors with the easiest opportunity for assisted suicide. Contrary to public myths about doctors' enjoyment of power, most physicians have not welcomed it and are delighted with an unambiguous mandate for good palliative care, free of moral and political shoals.

Given the complex motives of terminally ill patients and the fact that palliative care specialists are still in short supply, it seems dangerous to risk desensitizing medical caregivers in particular, and society in general, to the inadequacies of palliative care and, ultimately, to the complexities of helping a patient die. Yet, the legalization of physician-assisted suicide could lead to such a development. Palliative care specialists *are still in a race against time*, as they try to educate the medical profession *and* the public about the provision of state-of-the-art care before the legalization of aid in dying provides excuses for the avoidance of faithful attention to the needs of dying patients. The fact that state-of-the-art palliative care for suffering patients is not yet routine and that, even when it is available, can become mired in political controversy is tragic.

Whatever the outcome of the debate, psychiatrists should remember that there is no room for political or philosophical controversy at the bedside. There is just one reality that they and their medical colleagues should keep uppermost in their minds: a dying human being whose physical, social, emotional, and spiritual needs are all being effectively attended is far less likely to request help to commit suicide, and the family members—given proper help and support—derive a deep sense of peace from having helped their loved one to die feeling loved and secure.

ACKNOWLEDGMENTS

We'd like to pay tribute to the original author of this work, our colleague, Dr. Marguerite S. Lederberg, who has since died. At the time we prepared this chapter for the 50th-anniversary edition, Dr. Lederberg knew it would be her last work, and, in addition to her prodigious intellect and unparalleled clinical acumen, she put heart and soul into every page. She did so not only for the edification of the psychiatric community but also to leave a legacy for her co-authors to continue after she was gone. We are honored both to receive that legacy and to continue the work she started.

Further Readings

Agrawal M, Emanuel EJ. Ethics of phase 1 oncology studies: re-examining the arguments and data. *JAMA*. 2003;290(8):1075–1082.

American Academy of Hospice and Palliative Medicine. Position statement on physician-assisted death. *J Pain Palliat Care Pharmacother*. 2007;21(4):55.

Bakitas M, Ahles TA, Skalla K, Brokaw FC, Byock I. Proxy perspectives regarding end-of-life care for persons with cancer. *Cancer*. 2008;112(8):1854.

*Cherny N, Fallon M, Kaasa S, Portenoy R, Currow DC. *Oxford Textbook of Palliative Medicine*. Oxford University Press; 2015.

*Chochinov HM. *Dignity Therapy: Final Words for Final Days*. Oxford University Press; 2012.

*Chochinov HM, Breitbart W. *Handbook of Psychiatry in Palliative Medicine*. 2nd ed. Oxford University Press; 2009.

Connolly S, Galvin M, Hardiman O. End-of-life management in patients with amyotrophic lateral sclerosis. *Lancet Neurol*. 2015;14(4):435–442.

Coyne P, Mulvenon C, Paice J. American Society for Pain Management Nursing and Hospice and Palliative Nurses Association Position Statement: Pain Management at the End of Life. *Pain Manag Nurs*. 2018;19(1):3–7.

Crawley LM. Racial, cultural, and ethnic factors influencing end-of-life care. *J Palliat Med*. 2005;8(Suppl 1):58–69.

Ferrell BR, Coyle N, Paice J. *Oxford Textbook of Palliative Nursing*. 4th ed. Oxford University Press; 2014.

Fraser J, Harris N, Berringer AJ, Prescott H, Findlay F. Advance care planning in children with life-limiting conditions—the Wishes Document. *Arch Dis Child*. 2010;95:79.

Ganzini L, Goy ER, Miller LL, Harvath TA, Jackson A. Nurses' experience with hospice patients who deliberately hasten death by voluntary refusal of food and fluids. *N Engl J Med*. 2003;349:359.

*Gawande A. *Being Mortal: Medicine and What Matters in the End*. Henry Holt; 2014.

Hammes BJ, Briggs L. *Respecting Choices: Building a Systems Approach to Advance Care Planning*. Gundersen Health Systems; 2011.

Institute of Medicine. *Dying in America: Improving Quality and Honoring Individual Preferences near the End of Life*. National Academies Press; 2014.

*Kearney JA, Lederberg MS. Ethical issues in pediatric oncology. In: Wiener LS, Pao M, Kazak AE, Kupst MJ, Patenaude AF, eds. *Pediatric Psycho-Oncology: A Quick Reference on the Psychosocial Dimensions of Cancer Symptom Management*. Oxford University Press; 2015:313–324.

Kissane DW, Bloch S. *Family Focused Grief Therapy: A Model of Family-Centered Care during Palliative Care and Bereavement*. Open University Press; 2002.

LeRoy AS, Robles B, Kilpela LS, Garcini LM. Dying in the face of the COVID-19 pandemic: Contextual considerations and clinical recommendations. *Psychol Trauma*. 2020;12(S1):S98–S99.

Lo B, Ruston D, Kates LW, Arnold RM, Cohen CB. Discussing religious and spiritual issues at the end of life: a practical guide for physicians. *JAMA*. 2002;287:749–754.

McCurdy D. Ethical spiritual care at the end of life. *Am J Nurs*. 2008;108(5):11.

Meier DE, Isaacs SL, Hughes R. *Palliative Care: Transforming the Care of Serious Illness*. Jossey-Bass; 2010.

Morita T, Hirai K, Sakaguchi Y, Tsuneto S, Shima Y. Family-perceived distress from delirium-related symptoms in terminally ill cancer patients. *Psychosomatics*. 2004;45:107–113.

Moriyama, D, Scherer, JS, Sullivan, R, Lowy, J, Berger JT. The impact of COVID-19 surge on clinical palliative care: A descriptive study from a New York hospital system. *J Pain Symptom Manag*. 2021;61:(3):e1–e5.

Old JL. Discussing end-of-life care with your patients. *Fam Pract Manag*. 2008;15(3):18.

*Postovsky S, Ben Arush MW. Care of a child dying of cancer: the role of the palliative care team in pediatric oncology. *Pediatr Hematol Oncol*. 2004;21:67.

Qaseem A, Snow V, Shekelle P, Casey DE Jr, Cross JT Jr. Evidence-based interventions to improve the palliative care of pain, dyspnea, and depression at the end of life: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2008;148(2):141.

Turkel SB, Braslow K, Tavare CJ, Trzepacz PT. The delirium rating scale in children and adolescents. *Psychosomatics*. 2003;44:126.

Veira O, Sharfstein JM. ‘This level of death, it’s new’: Health care workers in their own words. *JAMA*. 2021;325(4):331–332.

Viswanathan R, Myers MF, Fanous AH. Support groups and individual mental health care via video conferencing for frontline clinicians during the COVID-19 pandemic. *Psychosomatics*. 2020;61(5):538–543.

*Zadeh S, Pao M, Wiener L. Opening end-of-life discussions: how to introduce Voicing My CHOiCES™, an advance care planning guide for adolescents and young adults. *Palliat Support Care*. 2015;(3):591–599.

27.10 Stress and Psychiatry

Traci J. Speed, M.D., Ph.D., Paul S. Nestadt, M.D., and Elizabeth J. Prince, D.O.

Before the 1930s, the term “stress” was primarily used in engineering and simply referred to the physical strain on a material object. The modern psychological construct of stress was formulated in the 20th century but has rapidly become a fundamental concept of psychiatry. Patients will readily describe their chronic and acute stressors without much prompting, and much can be learned from their particular way of recounting, prioritizing, and reportedly overcoming (or succumbing to) each instance of stress. All people share an existential vulnerability to life’s difficulties; as the classical Greeks would have it, “we are all but a heartbeat away from disaster.” Stress has been invoked as a cause of major psychopathology, a precipitator or trigger of psychiatric illness, and a contributor to considerable mental anguish. Similar roles have been cast for stress due to illness throughout medicine. Many of the suggested links are intuitively appealing. After all, who has not experienced somatic distress in the face of unpleasant events and thoughts? Such associations between stress and diverse symptoms are obvious. Whether stress causes, an illness is much more difficult to demonstrate.

The slipperiness of the term stress is both an advantage and a disadvantage. It is an advantageous term because it is all-embracing and synthesizes many aspects of life’s experience. It is disadvantageous because its referents are diverse. Stress refers both to a subject and a predicate, an event, and the consequences of that event. For clarity, it is probably wise to eschew the word stress altogether and use different terms for its various meanings. Thus, it is much clearer to refer to “stressful” events or “stressors” and the resulting “distress” or “stress responses” incurred. Such terms will be utilized throughout this section, but occasionally, the broader term of stress will be referenced.

HISTORY

The term *stress* was coined by Hans Selye (1907–1982). He observed that many highly diverse ways of perturbing the organism resulted in common physiologic responses. The concept was unusual because the prevailing ideology of medical research emphasized specificity, that is, that specific pathogens elicited unique pathologic responses. Selye found that highly diverse stimuli seemed to result in common changes. From animal studies, he reported that there is a "general adaptation system," a systemic well-orchestrated adaptive response to diverse stressors. Selye invoked the adrenocortical system as the crucial responder to stressful stimulation. He observed that any novelty or perturbation was associated with an elevation of adrenocortical activity, at least transiently, and that if stressors were unremitting, diverse pathologic changes would be evident.

Like all pioneers, Selye was guilty of oversimplification, a weakness that has become abundantly clear with the accumulation of additional empirical data. Not all organisms respond to all perturbations with the general adaptation system. However, his insights were important in shaping the field. As a result of his prolific scholarship, the field of psychology has been blessed (or cursed) with the term stress, which has become so popular that it is used constantly, even when researchers would prefer more precision. Stress is a term like neurosis or hysteria, ripe with years of reference, conveying both information and vagueness.

The other groundbreaker in stress research was Walter Cannon (1871–1945). Although he made major contributions in gastroenterology and roentgenology, his long-abiding interest was in the physiology of emotion and the autonomic nervous system. If Selye is known for his explorations of the adrenocortical system, Cannon is acclaimed for his methodical investigations of the other great pathway of stressful responses, the sympathetic nervous system (SNS). Cannon also focused on more immediate or short-term responses to stressors.

Whereas the term stress and the accumulating empirical research reflect the 20th century and the intellectual progeny of Selye and Cannon, there are earlier referents for the concept. For centuries physicians and patients have associated adverse life events and illness. In law, for example, there is a long history of tort cases alleging that emotional arousal secondary to a stressful event causes miscarriage. The links between emotion and sudden cardiac death have been noted in sources as diverse as the Bible, anthropology, and clinical experience. These types of links were repeatedly observed by crucial figures in the history of medicine. William Harvey in the 17th century and William Osler in the 19th century frequently alluded to the relationship between adverse life events and illness onset. Many contemporary cultures regard illness as the outcome of being “out of balance” with the environment and its demands, a manifestation of inadequate coping with diverse stressors.

The confusion between stress as a “cause” versus “trigger” of illness has hindered progress. The allegation that a stressor caused an illness is difficult to demonstrate in most human studies, which impose a short-term stressor and examine the resultant physiologic changes. Illnesses such as myocardial infarction usually become evident only after a long accumulation of coronary artery disease. As a result, how can one assert that “stress caused a myocardial infarction”? However, there is an alternative hypothesis that provides very strong evidence for a link between stressors and illness. That argument emphasizes that stressors trigger illness in individuals whose underlying illness is latent or subclinical. This sort of case is demonstrated routinely in psychophysiology and psychosomatic medicine studies. For instance, studies of the physiologic response to stressors in normal individuals demonstrate large amplitude changes in physiologic levels of blood pressure, gastric motility, and hormonal levels. When such psychophysiological studies are extended to patient groups, one sees breathtaking evidence for the role of stressors as trigger events in eliciting pathophysiologic phenomena. For example, even discussing emotional conflicts can elicit life-threatening arrhythmias in vulnerable individuals, and exposure to a catastrophic stressor can be accompanied by unique and reversible cardiomyopathy (Fig. 27.10–1). Similar effects of stressors on blood pressure have been repeatedly demonstrated in warfare and natural disasters.

A 54-year-old man whose office was in the vicinity of the World Trade Center had his blood pressure recorded twice a day from September 4 to September 21, 2001. His blood pressure dramatically increased during the week after the World Trade Center attack (Fig. 27.10–2).

The COVID-19 pandemic provided the 21st century’s first worldwide stress test. Studies into the differential impact of this universal stressor have found increased levels of self-reported distress and depressive symptoms in the context of infection risks and the mitigation measures required to stem the pandemic. However, more objective measures of severe psychiatric illness such as hospitalizations and suicides have demonstrated no such increase and, in some cases, have decreased outside of specific vulnerable populations. Studying the responses to both natural and contrived experiments will further elucidate the complex relationship between stress and psychiatric sequelae. In the future, we will continue to broaden our understanding of the human stress response at the cellular, organ system, individual, community, and population levels to best understand stress's role in mental life.

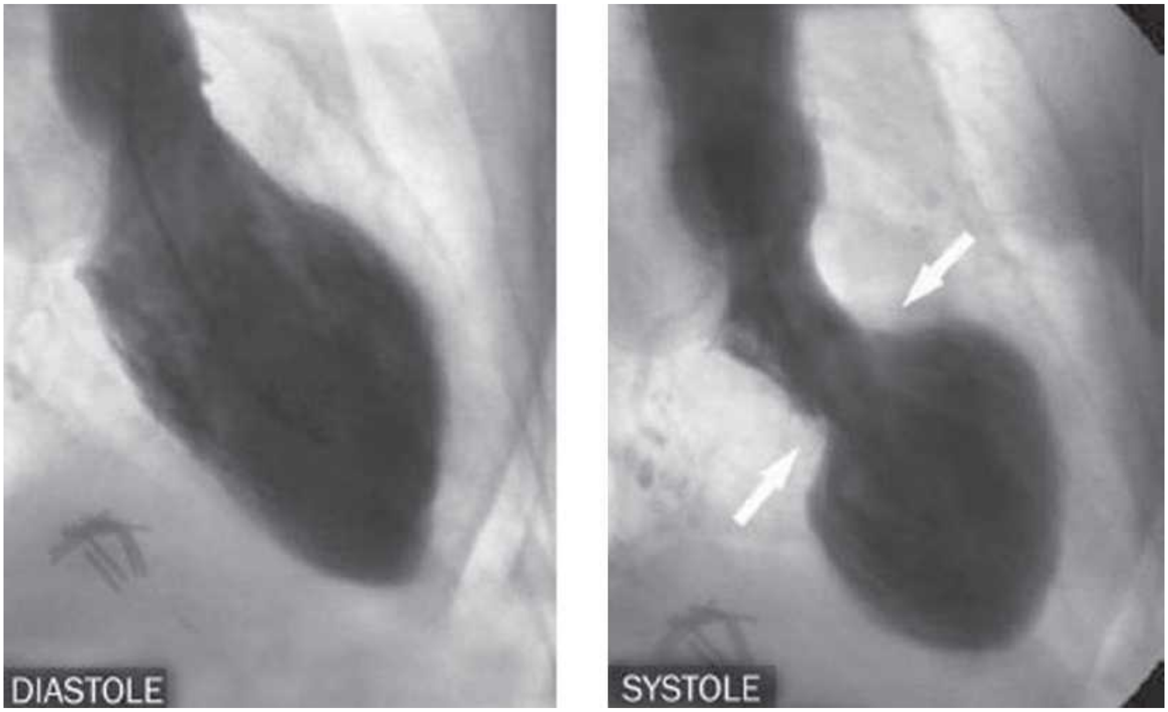

FIGURE 27.10–1. Impact of stress on cardiac function. A reversible cardiomyopathy can be elicited by psychologically stressful events. This is characterized by systolic dysfunction affecting the distal left ventricular chamber. (Reprinted with permission from Sharkey SW, Lesser JR, Zenovich AG, et al. Acute and reversible cardiomyopathy provoked by stress in women from the United States. *Circulation.* 2005;111(4):472–479. doi: 10.1161/01.CIR.0000153801.51470.EB.)

VULNERABILITY AND RESILIENCE

Parallel to confusion between stress as a trigger and stress as a response, another key conceptual problem revolves around the issue of vulnerability versus resilience or stress versus coping. The conceptual origins of resilience, like stress, are outside psychiatry and medical science in diverse fields from engineering to ecology. Resilience focuses on the recovery, adaptation, or even transformation of a system following exposure to stress. In psychological resilience, the system at the center of the process is a person.

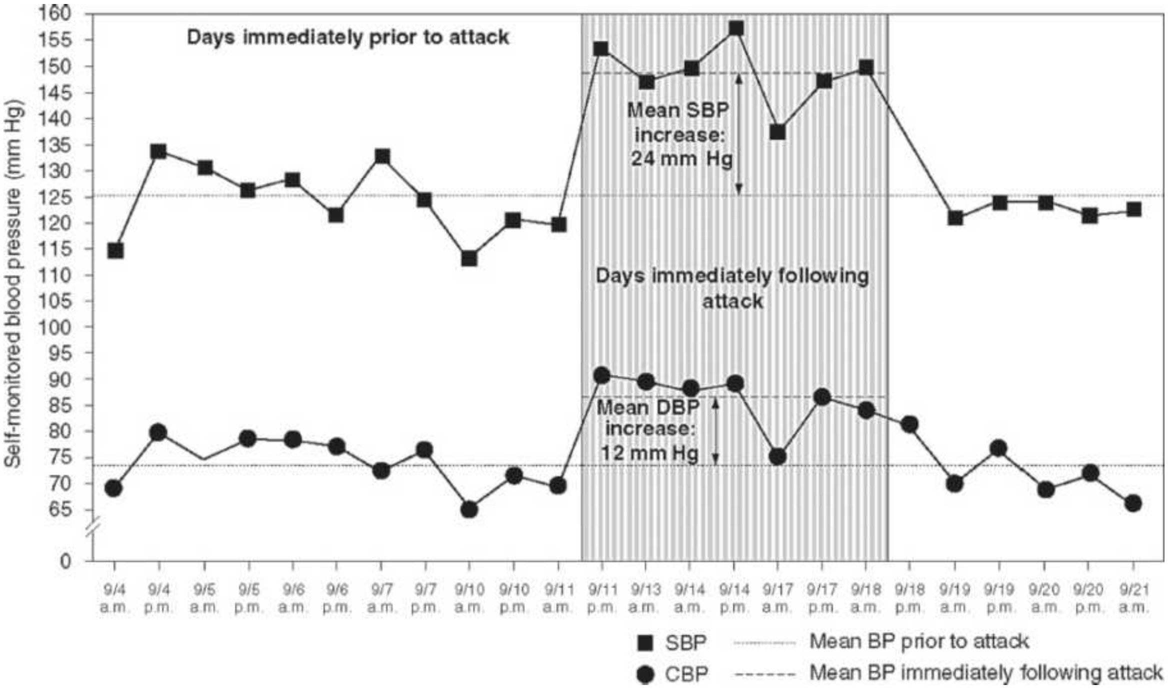

FIGURE 27.10–2. Impact of stress on blood pressure (BP). *DBP*, diastolic blood pressure; *SBP*, systolic blood pressure. (Reprinted with permission from Lipsky SI, Pickering TG, Gerin W. World Trade Center disaster effect on blood pressure. *Blood Press Monit.* 2002;7(4):249. doi:10.1097/00126097-200208000-00008.)

Researchers initially focused on the stressor, treating it like a virulent pathogen that would infect all individuals despite varied life experiences. Thus, the initial stress research primarily focused on the physiologic consequences of stressor imposition. The variability in response patterns then motivated investigators to study the subtleties of how people interpreted the stressful challenges. These interpretations were framed as "defensive" styles, how individuals guard themselves against painful emotional stressors. Subsequently, however, the inquiry broadened, and investigators examined individuals' defensive behaviors, information-seeking behaviors, affiliative behaviors, and all-around problem-solving behaviors. This heterogeneous combination is referred to as *coping behavior*.

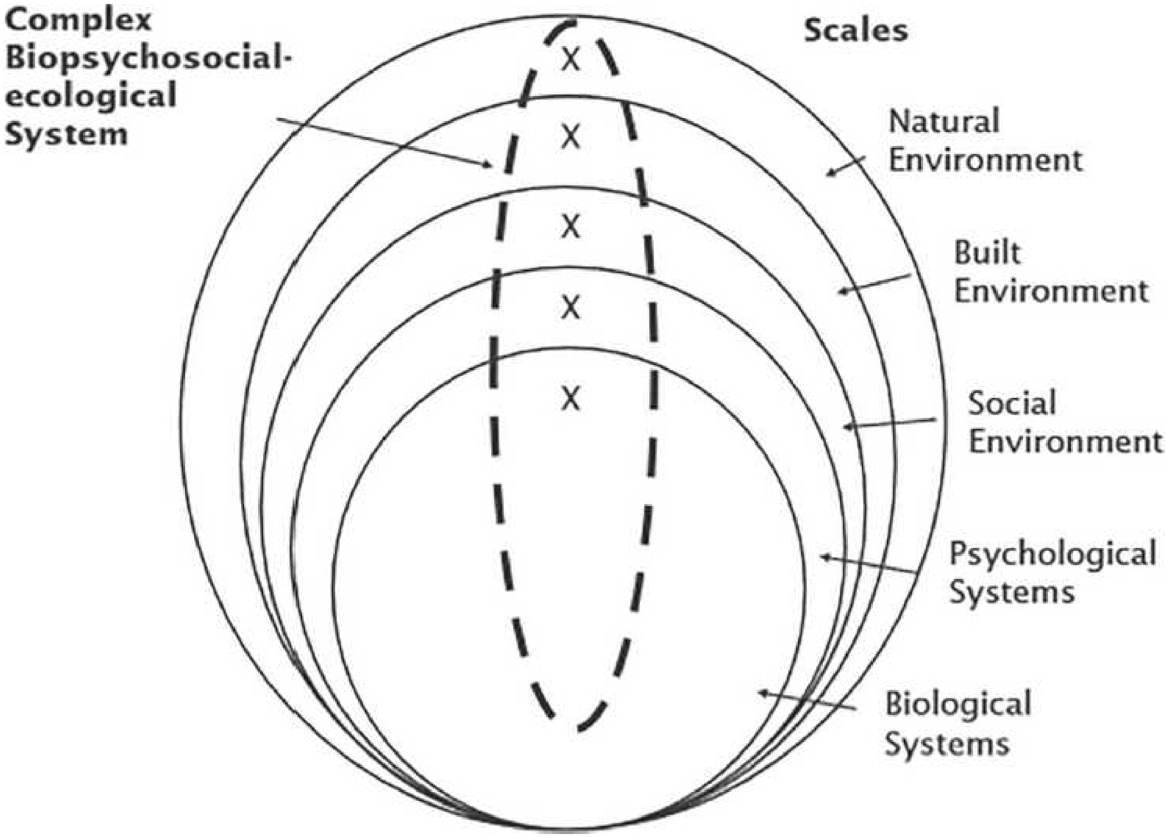
Complex Biopsychosocial-ecological System
Scales
Natural Environment
Built Environment
Social Environment
Psychological Systems
Biological Systems
X
X
X
X
X

FIGURE 27.10–3. A systemic model of resilience comprised of interacting scales. The Xs represent promotive and protective factors at different scales. The dashed ellipse represents a single multilevel system. (Reprinted from Ungar M, Theron L. Resilience and mental health: how multisystemic processes contribute to positive outcomes. *Lancet Psychiatry.* 2020;7(5):441–448. doi: 10.1016/S2215-0366(19)30434-1.)

Beyond behavior, genetic and environmental factors are also being studied for their impact on response to stress. For instance, a New Zealand study prospectively followed approximately 1,000 subjects aged 3 to 26 to learn more about stress effects on the onset of major depression. The promoter region of the serotonin transporter gene had a powerful effect on vulnerability to stress. Individuals who were homozygous for the short allele of the promoter region of the serotonin transporter gene were more likely to become depressed and suicidal in response to stressors. The role of the environment in stress is also being studied. Further research demonstrated an association between the short allele and increased vulnerability to stress in studies of childhood maltreatment and certain medical conditions.

As stress-related psychiatric disorders are influenced by genetic predisposition and environmental factors, researchers have begun studying epigenetic mechanisms that may account for environmental influences on the development of stress-related disorders in recent decades. Recent research shows that prenatal factors such as stress or mood disorders in mothers are risk factors for psychiatric disorders in offspring. These different influences on trajectories of response to stress suggest resilience is a process of multiple interacting scales rather than simply a trait possessed by an individual (Fig. 27.10–3).

MEASURING STRESS PSYCHOLOGICALLY

Regardless of whether stress is a cause or trigger, or whether one focuses on stress or factors that affect resilience to stress, the key step is to recognize and in some way quantify its extent. Doctors have always asked about the stressors in their patients' lives. Obtaining such information is, after all, at the core of history taking. "What types of things aggravate your symptoms? Are you facing any unusual burdens or challenges"? With certain symptoms such as angina, the link between stressor and illness is immediately obvious. However, it appears that stressful conditions aggravate many illnesses not traditionally thought of as "stress-sensitive." As will be discussed later, the mechanisms for this link are complex.

In the early 20th century, Adolf Meyer suggested that psychiatrists employ a "life chart" in evaluating their patients. The idea is elegant in its simplicity. The psychiatrist plots a patient's life on a line, beginning with birth and ending in the present. He or she then pencils in when various important life events have occurred and indicates the onset or worsening of health problems. The life chart turns out to be a very effective way of eliciting a clinical history. Sometimes, recurrent patterns of association between life events and illness episodes are suggested (Fig. 27.10–4).

The next step in the evolving field of life events was developed by the fruitful collaboration of Thomas Holmes and Richard Rahe. They felt that a list of commonly encountered stressors would be more useful than taking an unstructured history. After considerable research, they developed a list of roughly 40 life events ranging in severity from the death of a spouse to receipt of a traffic ticket. The checklist or schedule of recent events can be self-administered and requires very little time to complete. It is noteworthy that the form tallies both positive and negative life events. The underlying assumption is that all life events, be they positive or negative, impose a demand for adaptation and that such demands are stressful. While at first blush this may seem counterintuitive, positive life events are not totally benign; getting married brings with it a host of stressors. Another characteristic of the Holmes–Rahe approach is that the items are weighted in their adaptational demand. Most people rank order stressors similarly, and thus it is possible to assign a specific "number" of life-change units to the occurrence of each specific stressor. This technique has been used in hundreds of studies and has demonstrated that numerous illnesses follow on the heels of increases in life stress. It is not that the high-scoring individual is at risk, but rather that the individual with an *increase* in life stress units is at risk for any number of health consequences.

There are many criticisms of this approach and, of course, alternative ways of measuring stressful life events. Some investigators feel that negative life events carry with them *far* greater anguish and adaptational pressure and that the positive life events may be ignored. Some question the utility of adding all the life-change units together and coming up with one number to represent the total stressful exposure of the individual. Whereas the Holmes–Rahe approach may be extremely useful for epidemiologic research, for clinical use, it may be more helpful to have subscales that list, for example, the stressor scores associated with job pressure, home pressure, and so forth into separate categories. Some theorists feel that inventories of relatively rarely occurring stressful events are of limited utility and that more is to be gained by measuring the accumulation of the more common microstressors that individuals encounter frequently. Other investigators doubt the validity of a universal weighting of stressful events and favor allowing individuals to apply their idiosyncratic assessment of emotional travail elicited by events. Finally, common to all scales that measure stressful events is the problem of retrospective distortion. Patients forget, disavow, deny, and repress—all words that reflect a splitting off of accurate recall. As a result, when patients are asked to retrospectively report life events, they tend to forget events shrouded in the past. On the other hand, patients' beliefs sometimes cause them to attribute symptoms to stressful events instead of disorders.

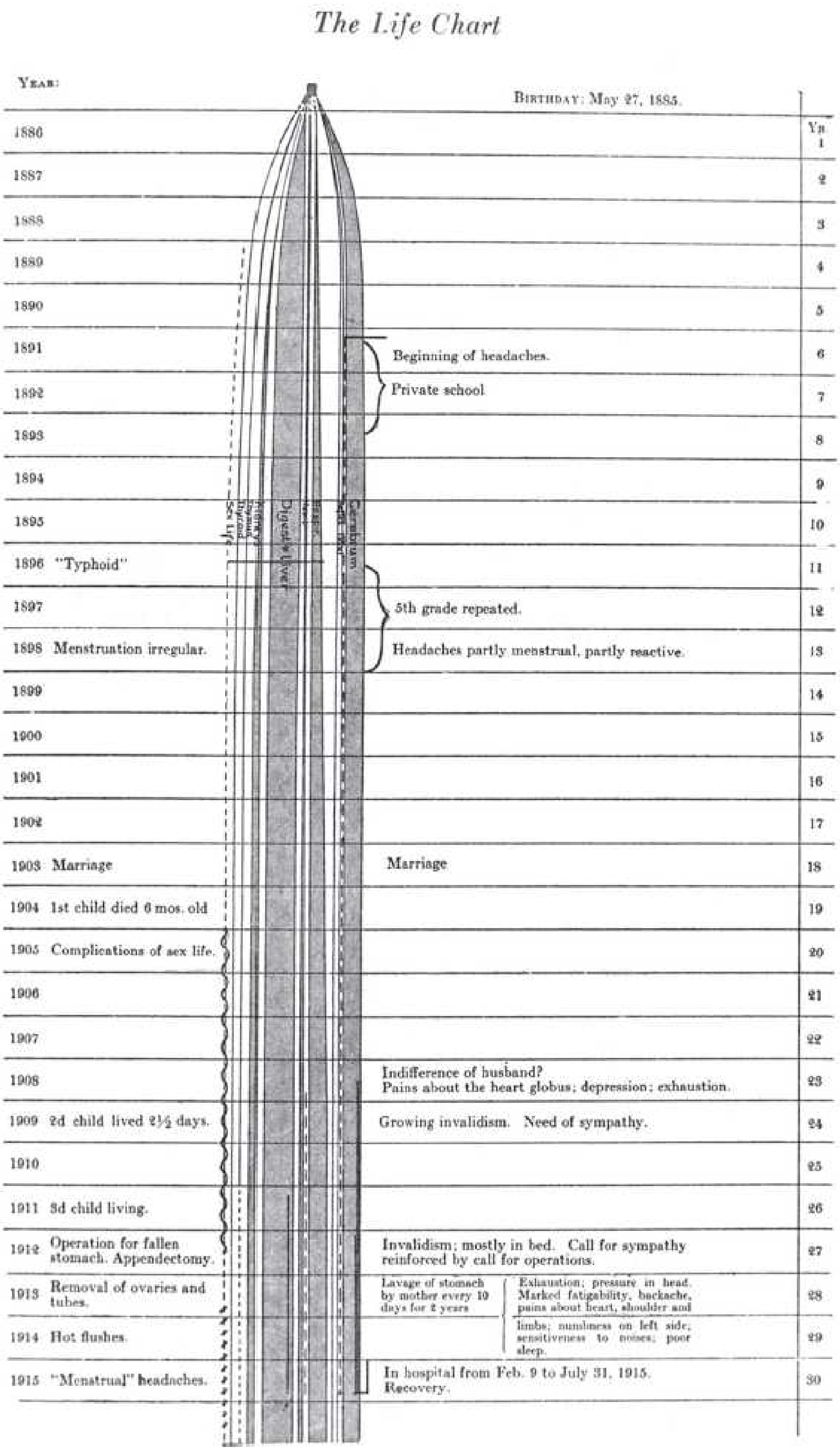

FIGURE 27.10–4. Line drawing of the life chart by Adolf Meyer. (From *Contributions to Medical and Biological Research, Dedicated to Sir William Osler, bart. M.D., F.R.S., in Honour of his Seventieth Birthday, June 12, 1919,*

by His pupils and Co-workers. Paul B. Hoeber, New York, 1919. Credit to Wellcome Library, London. Creative Commons License.)

SELECTED PHYSIOLOGIC RESPONSES TO STRESSORS

The body's reaction to a stressor (be it real, symbolic, or imagined) is to initiate a set of responses that seek to diminish the impact of the stressor. A stressor disrupts an organism's equilibrium, and the *stress response* consists of initiating physiologic adjustments that seek to react to the stressor, bring about an adaptive response, and restore homeostasis. Much is known about the physiologic response to *acute stress*, and considerably less is known about *chronic stress*. This represents a major gap in physicians' ability to relate the effects of stress to the occurrence of psychiatric illness in humans. One such example is disparate exposures among racial/ethnic minority populations and those with low socioeconomic status who may experience more lifelong stressors (e.g., discrimination, finances) than their more well-off counterparts. Many stressors occur over a prolonged period or have long-lasting repercussions (e.g., losing a spouse is followed by months or years of loneliness; a sexual assault is followed by years of apprehension and worry). Animal models that can capture this reverberation or "echo" of stress are limited. Still, what is known about the neuroendocrine and immune response to laboratory stressors is expanding and will ultimately inform the understanding of why and how stress harms humans.

Neurotransmitter Responses to Stress

Stressors activate a release of catecholamines from the autonomic nervous system, including noradrenaline, dopamine, and serotonin. The stressor type influences the brain regions and mediators that engage in the adaptive response. Physical stressors (e.g., blood loss, trauma, hypothermia) recruit the brainstem and hypothalamic brain regions. Psychological stressors (e.g., embarrassment, work-related stress) engage brain regions involved in emotion (amygdala), learning and memory (hippocampus), and decision-making (prefrontal cortex). The timing of stress affects neuronal pathways differently. For instance, early life stress affects neurotransmitter release in the prefrontal cortex, producing persistent changes through adulthood. Furthermore, emerging research demonstrates sex-specific effects of early life stress on the prefrontal cortex catecholamine response. While catecholamine response varies depending on the time of stress relative to developmental stage and the type of stressor, consistently male rats, have greater variability in neurotransmitter response to stress than females despite conventional views that females are more vulnerable to life stressors. At present, we have a rudimentary understanding that the nature of the interaction between sex and stress is complex. The sexually dimorphic responses to stress as seen in animal models may have implications for the pathophysiology of stress-related disorders, including major depression and anxiety. Further understanding of the molecular underpinnings of sex differences in response to stress may advance the development of efficacious treatments for these disorders. Duration of stress also plays a role in catecholamine signaling. Compared with acute stress, which evokes rapid transient alterations in catecholamine release, chronic stress induces sustained effects on neuronal firing and gene activation. Although this area of research remains novel, epigenetics provides insights into the pathogenesis of response to chronic stress, as DNA methylation, histone modifications, and microRNA activity mediate dynamic molecular adaptations of the central nervous system (CNS) to chronic stressors

Endocrine Responses to Stress

It has become clear that amino acid and peptidergic neurotransmitters are also intricately involved in the stress response. Endocrine response to stress is secreted from the hypothalamus into the hypophysial–pituitary–portal system in response to stress. Corticotropin-releasing hormone (CRH) acts at the anterior pituitary to trigger the release of adrenocorticotrophic hormone (ACTH). Once ACTH is released, it acts at the adrenal cortex to stimulate the synthesis and release of glucocorticoids. Glucocorticoids themselves have myriad effects within the body. However, their actions can be briefly summarized as promoting energy utilization, increasing cardiovascular activity (in the service of the flight-or-fight response), and inhibiting functions such as growth, reproduction, and immunity.

This hypothalamic–pituitary–adrenal (HPA) axis is subject to tight negative feedback control by its end products (i.e., ACTH and cortisol) at multiple levels, including the anterior pituitary, the hypothalamus, and the hippocampus. It is now known that the HPA axis is also influenced by multiple extrinsic systems, which should not be surprising given the coordinated nature of the stress response. For example, in addition to CRH, numerous secretagogues (i.e., substances that elicit ACTH release) exist that can bypass CRH release and act directly to initiate the glucocorticoid cascade. Examples of such secretagogues include catecholamines, vasopressin, oxytocin, and orexin. Interestingly, different stressors (e.g., cold stress versus hypotension) trigger different patterns of secretagogue release, again demonstrating that the notion of a uniform stress response to a generic stressor is an oversimplification.

A relevant feature of the HPA axis stress response is that it varies depending on whether the organism (most of these data come from rats and zebrafish) is exposed to the stressor on a single occasion, more than once, or over a prolonged period. Interestingly, it has not yet been possible to rely on a model that can predict whether “sensitization” (enhanced response to the same stress over time) or “desensitization” (a reduced response to the same stress over time) will occur for a given stressor. Nonetheless, the fact that the stress response can change dramatically depending on the frequency and duration of exposure highlights the necessity of specifying these elements when trying to understand the role of stress in clinical psychiatric illness.

PSYCHONEUROIMMUNOLOGY

Researchers have realized that the immune system is also a key player in stress physiology. There are numerous bidirectional interactions between the brain, behavior, and immune system, and these interactions are studied in psychoneuroimmunology. The immune system provides the body’s defense against invading external pathogens such as viruses and bacteria and abnormal internal cells such as tumors. Innate immunity refers to the body’s resistance to pathogens; this form of immunity operates in a nonspecific way without recognition of the different natures of various pathogens. Macrophages and granulocytes are nonspecific immune cells that induce immune activation by releasing proinflammatory cytokines (TNF-α; IL-6) and react to tissue damage by consuming debris and invading organisms. Specific immunity is acquired in response to identifying nonself-molecules called antigens. Each B-cell or T-cell is genetically programmed to attack a specific target by secreting antibodies (B-cell) or by killing cells of the body that harbor a virus (T-cell).

Interleukins or cytokine release from immune cells orchestrates both innate and specific immunity, and this cytokine network helps differentiate the immune response and the co-ordination of its magnitude and duration. However, the immunoregulatory processes cannot be fully understood without considering the internal and external milieu in which innate and specific immune responses occur, as well as the interactions of the brain, autonomic nervous system, neuroendocrine axis, and gastrointestinal microbiome with the immune system (Fig. 27.10–5).

Biologic Connections between the CNS and Immune System

Autonomic Nervous System. The CNS and the immune system are linked by two major physiologic systems, the HPA, and the autonomic nervous system, composed of sympathetic and parasympathetic branches. The SNS is a network of nerve cells running from the brainstem down the spinal cord and out into the body to contact various organs, including organs where the immune system cells develop and respond to pathogens. The sympathetic release of norepinephrine and neuropeptide Y, together with receptor binding of these neurotransmitters by immune cells, serves as the signal in this hard-wired connection between the brain and the immune system. In addition, sympathetic nerves penetrate the adrenal gland and cause the release of epinephrine into the bloodstream, which circulates to immune cells as another sympathetic regulatory signal.

Under both laboratory and naturalistic conditions, sympathetic activation suppresses the activity of diverse populations of immune cells. Sympathetic activation can also enhance other aspects of the immune response, such as the production of antibodies by B-cells and the ability of macrophages to release cytokines and thereby signal the presence of a pathogen. Additional studies indicate that sympathetic activation can also shunt some immune system cells out of circulating blood and into lymphoid organs (e.g., spleen) while recruiting other types of immune cells into circulation (e.g., natural killer [NK] cells). In general, SNS activation can reduce the immune system's ability to destroy pathogens that live inside cells (e.g., viruses) via decreases in innate and cellular immune responses while sparing or enhancing the humoral immune response to pathogens that live outside cells (e.g., bacteria). Together, these observations are a cornerstone for understanding fundamental neuroanatomic signaling between the autonomic nervous and immune systems.

Neuroendocrine Axis. Another way the brain can communicate with the immune system is via the HPA system. Cortisol influences the actions of various cells involved in an immune response by suppressing the cellular immune response. Indeed, a synthetic analog of cortisol is often used to suppress excessive immune system responses. However, cortisol can also prompt some immune cells to move from circulating blood into lymphoid organs or peripheral tissues such as the skin. It is important to note that immune cells can produce neuroendocrine peptides (e.g., endorphin, ACTH), which suggests that the brain, neuroendocrine axis, and immune system use the same molecular signals to communicate with one another.

Gastrointestinal Microbiome. Understanding the bidirectional link between gut physiology and brain function and how this link influences the immune system under normal and stressful conditions remains in its infancy despite wide acceptance of the “gut-brain” interaction (e.g., gut feelings, gut instinct, butterflies in one’s stomach). The gut itself is an important immune organ that provides a defensive barrier against external pathogens, with gut-associated lymphoid tissues comprising 70% of the immune system. The brain acts on gastrointestinal and immune functions to shape the gut microbiome, while gut microbes produce neurotransmitters and metabolites that interact with the neuroendocrine and immune systems and control leakage through the intestinal lining and blood–brain barrier (Fig. 27.10–6). Animal studies show that exposure to acute and chronic stress alters the relative proportions of the main microbiota phyla, which influences host immune and stress responsiveness; peripheral administration of proinflammatory cytokines induces depressive- and anxiety-like behaviors. In humans, chronic gut disorders are associated with intestinal microbiome dysbiosis and behavioral changes, including increased responsiveness to stress and increased depressive and anxiety symptoms.

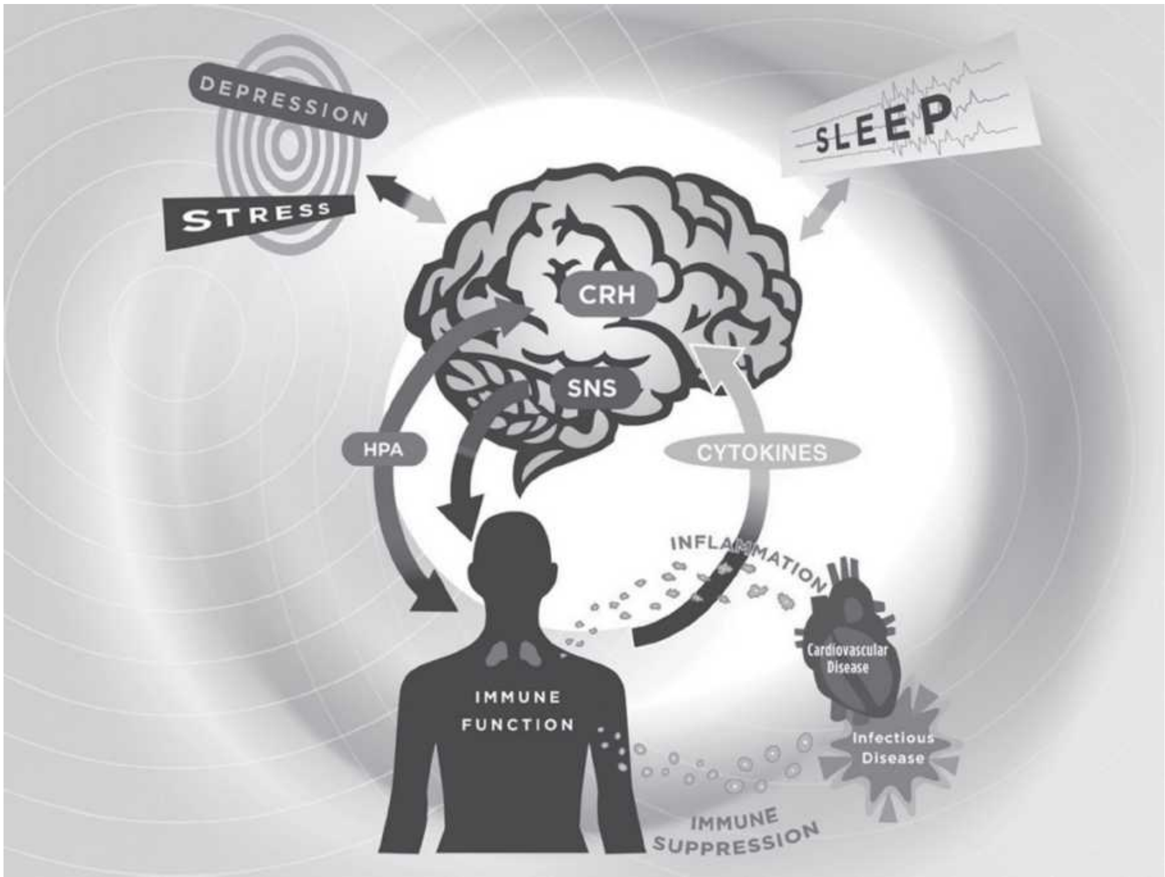
DEPRESSION
STRESS
SLEEP
CRH
SNS
HPA
CYTOKINES
INFLAMMATION
Cardiovascular Disease
IMMUNE FUNCTION
Infectious Disease
IMMUNE SUPPRESSION

FIGURE 27.10–5. Model of human psychoneuroimmunology interactions and clinical implications. The model depicts the bidirectional interactions between the brain and the immune system, and the clinical implications of immune alterations due to stress, depression, or sleep disturbance. *CRH*, corticotropin-releasing factor; *SNS*, sympathetic nervous system; *HPA*, hypothalamic–pituitary–adrenal axis. (Reprinted from Irwin MR. Human psychoneuroimmunology: 20 years of discovery. *Brain Behav Immun*. 2008;22(2):129–139. doi: 10.1016/j.bbi.2007.07.013.)

Central Modulation of Immunity. This converging evidence of brain–immune system interactions suggests that the brain has a physiologic role in regulating immunity. Indeed, the release of CRH in the brain alters various immune processes, including aspects of innate immunity, cellular immunity, and in vivo measures of antibody production. Peripheral immune measures also change following lesioning of the brain (e.g., hypothalamus) or in response to the stimulation of certain brain regions. The brain controls immune cells in lymphoid tissue just as it controls other visceral organs through a complex series of coordinated autonomic and neuroendocrine pathways.

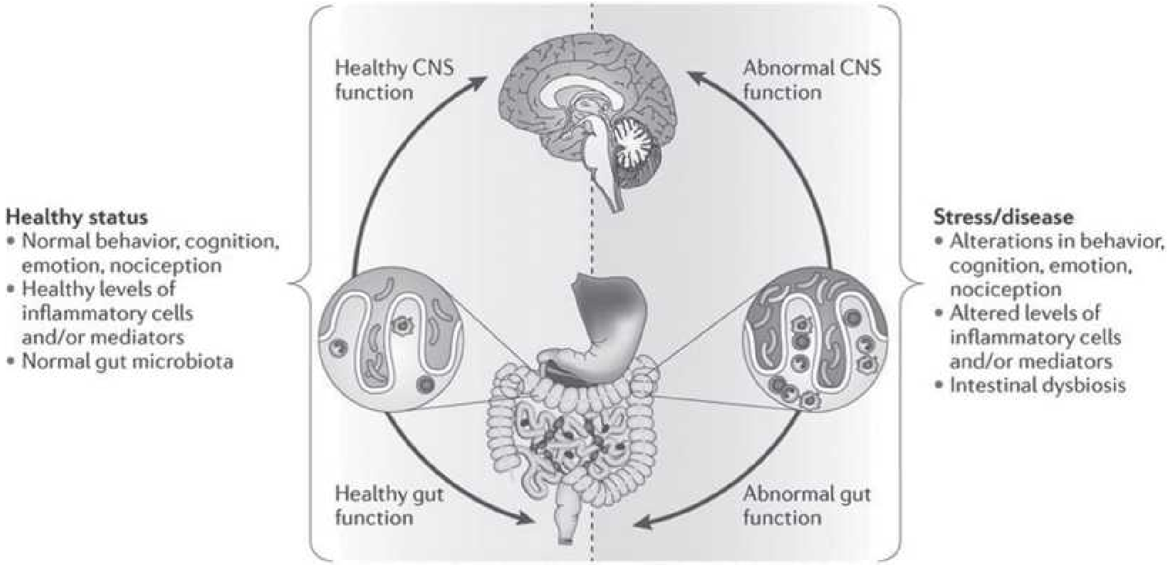
Healthy CNS function
Abnormal CNS function
Healthy status
• Normal behavior, cognition, emotion, nociception
• Healthy levels of inflammatory cells and/or mediators
• Normal gut microbiota
Stress/disease
• Alterations in behavior, cognition, emotion, nociception
• Altered levels of inflammatory cells and/or mediators
• Intestinal dysbiosis
Healthy gut function
Abnormal gut function

FIGURE 27.10–6. Model of the gut–brain axis. Normal gut physiology contributes to appropriate signaling to the central nervous system (CNS) and helps regulate healthy physiologic interactions. Intestinal dysbiosis alters gut physiology and leads to adverse consequences for CNS functions and results in disease states (*shaded area*). *CNS*, central nervous system. (Reprinted with permission from Cryan JF, Dinan TG. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat Rev Neurosci*. 2012;13(10):701. doi: 10.1038/nrn3346.)

Behavioral and Psychological Influences on Immunity

Given that psychological responses are expressed in neural activity with accompanying changes in neuroendocrine and autonomic function, it is not surprising that behaviors and emotions can alter immunity. One seminal example is classical conditioning of immune responses. Conditioned immune responses have also been found to retard disease progression in an animal model of autoimmune disease. In the clinical setting, immunosuppressive conditioning occurs in cancer patients who receive chemotherapeutic drugs, and conditioning processes are thought to contribute to the biologic effects of a placebo.

Stress and Immunity in Animals. In animals, acute exposure to stressors produces alterations of immunity that cannot be accounted for by the physical effects of the stress. Psychological components such as conditioned fear and uncontrollability induce altered immunity, and these effects are related in a dose-dependent manner to the severity of the stressor. Almost every immune system component has been found to respond to stress.

Acute Stress and Immunity in Humans. Acute laboratory stressors (e.g., mental arithmetic) produce profound and rapid changes in the immune system due to redistributing immunoregulatory cells from lymphoid organs such as the spleen into the vascular space. Such acute stressors elicit decreases in cellular immune responses and increases in markers of inflammation (e.g., IL-6), which are thought to be mediated by the release of sympathetic neurotransmitters and β-adrenergic receptor activation. Individuals who are older adults or undergoing chronic stress show exaggerated immune responses to acute stress and are likely to take longer to recover from the administration of stress, which may put them at increased risk for infectious disease. Finally, among depressed patients, acute psychological stress leads to exaggerated increases in inflammatory cytokine activity and greater activation of the inflammatory signaling pathway, nuclear transcription factor-κB.

Chronic Stress, Depression, and Immunity. In contrast to the effects of laboratory stress, chronic or naturalistic stressors such as bereavement or caregiving and depression are associated with reliable decreases of cellular and innate immunity, along with increases in proinflammatory cytokine activity, due possibly to downregulation of glucocorticoid receptor signaling. When the stress remits or depressive symptoms resolve, natural and cellular immune function normalization occurs.

Proinflammatory cytokine activity is also produced in excess by senescent cells. Telomeres are repeats of DNA sequences at the ends of chromosomes bound by specialized protective proteins. Telomere shortening in immune cells, which is a sign of cell senescence and aging, predicts earlier onset of the diseases of aging even after controlling for confounders. Mounting evidence supports an association between chronic psychological stress, stress-related disorders (e.g., depression), and compromised telomere maintenance. These findings highlight one of the many mechanisms of immune dysfunction in response to chronic stress. However, while the association of stress, psychiatric disorders, and telomere length is well-established, a better understanding of the link between telomeres and stress in the development of stress-related disorders is necessary to improve diagnosis and treatments.

Heterogeneity in the effects of stress and depression on immunity can also be accounted for by factors such as age, gender, ethnicity, health behaviors (e.g., smoking, alcohol consumption), and coping. Depressed patients who have comorbid alcohol abuse or tobacco smoking show exaggerated declines in natural and cellular immune responses. Personal characteristics such as coping and personality (e.g., positive affect), which moderate neuroendocrine and sympathetic activity, also contribute to individual differences in immune responses to psychological stress and vaccine responses.

Sleep, Cytokines, and Immunity

Disordered sleep and loss of sleep are thought to adversely affect resistance to infectious disease and alter inflammatory disease progression. Studies show that sleep deprivation impairs influenza viral clearance and increases bacteremia rates. In humans, normal sleep is associated with a redistribution of circulating lymphocyte subsets, increases of NK activity, IL-2, IL-6, and the trans-signaling IL-6 receptor agonist (IL-6R), and a relative shift toward Th1 cytokine expression, which is independent of circadian processes. Conversely, sleep deprivation suppresses NK activity and IL-2 production and induces decreases in specific antibody production to infectious challenge (e.g., vaccination against influenza or hepatitis B virus). In contrast, innate immunity and proinflammatory cytokine expression measures, including cellular and genomic markers of inflammation, are enhanced following sleep loss. In clinical populations who show disordered sleep (e.g., as a result of depression, bereavement, alcoholism), natural and cellular immune function alterations coincide with sleep loss and disturbances of sleep architecture.

Furthermore, decreases in sleep continuity or increases in rapid eye movement (REM) sleep are associated with increases in the nocturnal and daytime expression of IL-6, possibly with consequences for daytime fatigue. Laboratory pain testing methods are now being used to understand individual differences in proinflammatory cytokines reactivity. One recent study showed greater pain-evoked IL-6 among older adults with insomnia disorder than among those without insomnia. Bidirectional actions of cytokines on sleep have also been identified. In animals, cytokines have both somatogenic and inhibitory effects on sleep, depending on the cytokine, plasma level, and circadian phase. In humans, less is known about the sleep-regulatory effects of cytokines. Expression of the anti-inflammatory cytokine IL-10 prior to sleep predicts delta sleep amounts during the nocturnal period. In contrast, peripheral administration of the proinflammatory cytokine IL-6 reduces delta sleep, and nocturnal levels of both IL-6 and TNF temporally correlate with increases in REM sleep, particularly during the late part of the night.

Cytokine Influences on the CNS and Behavior

The nervous and immune systems convey information to each other, ultimately regulating neuroendocrine and autonomic outflow and the immune response. With peripheral immune activation, the inflammatory cytokines, TNF, IL-1, and IL-6, activate the HPA system via the central release of CRH. Immune activation leads to changes in peripheral physiology that are similar to a stress response. Coincident with these physiologic changes, animals show reductions in activity, exploration of novel objects, social interactions, food and water intake, and sexual behaviors, a response pattern known as *sickness behaviors.* Human studies have shown similar links between peripheral cytokines and behavioral changes. Healthy individuals develop anxiety and depression in response to low-dose cytokines. Even patients with the common cold show sickness behaviors such as fatigue, lack of interest, low mood, and cognitive impairment. Patients with chronic inflammatory diseases such as hepatitis C infection have increased rates of fatigue and depression with associated changes in the metabolism of central neurotransmitters in discrete brain regions. Large doses of cytokines, given as immunotherapy for cancer or hepatitis C, frequently worsen depressive symptoms, which can be attenuated by pretreatment with antidepressant medications.

Clinical Implications of Psychoneuroimmunology and the Cytokine–Brain Link

The links between stress and disease are bidirectional. Increasing evidence suggests that abnormalities in cytokine expression and regulation that occur during depression or stress might exacerbate the course of many chronic diseases. Research has primarily focused on the influence of stress exposures on the development and progression of disease, yet diseases can also serve as stressors. The following paragraphs provide several pertinent examples in which cytokines and psychoneuroimmunology processes are thought to contribute to disease processes and how the disease process influences stress.

Cardiovascular Disease. Depression is associated with an increased risk of coronary artery disease and increased morbidity and mortality after myocardial infarction, congestive heart failure, or mitral valve replacement. Atherosclerosis is now thought to be an inflammatory process involving a series of steps, each of which can be impacted by stress or depression. Activated macrophages within the vasculature secrete proinflammatory cytokines such as IL-1, IL-6, and TNF, leading to the expression of cellular adhesion molecules. With the recruitment of immune cells to the vascular endothelium and the release of inflammatory cytokines, endothelial activation occurs, facilitating further binding of immune cells. Importantly, psychological and physical stressors lead to an increased release of proinflammatory cytokines and the expression of adhesion molecules that tether and bind immune cells to the vascular endothelium. Depressed mood, sleep disturbance, inflammation, and arterial repair might form a vicious cycle promoting cardiovascular disease. In a study of 1,794 Canadians, depression and elevated inflammatory markers predicted an increased risk of coronary artery disease. In another study of patients with coronary heart disease, the severity of depression was associated with elevated CRP and IL-6. This link was driven by behavioral health factors, including physical inactivity, smoking, and obesity, suggesting other mechanisms including epigenetic factors may promote the link between depression and cardiovascular disease.

Infectious Disease Risk. Compelling evidence has shown that inescapable stress, a putative animal model of depression, increases susceptibility to viral diseases such as herpes simplex and infection via alterations in immune function. In humans, prospective epidemiologic studies and experimental viral challenge studies show that persons reporting more psychological stress have a higher incidence and greater severity of certain infectious illnesses such as Epstein–Barr virus infections and the common cold. Moreover, experimental vaccinations have probed the disease-specific and integrated in vivo immune system action to psychological stress.

Atopic Dermatitis. Atopic dermatitis is a chronic and relapsing skin disease characterized by impaired epidermal skin barrier function, inflammatory infiltrates, and excessive pruritus. Psychological stress contributes to the course of atopic dermatitis through its direct and indirect effects on the immune response. Skin biopsies from patients with atopic dermatitis have elevated Th2 cells, cytokines, and eosinophils with elevated IgE production in response to stress. Patients with atopic dermatitis and high trait anxiety also have elevated levels of IgE. Numerous studies demonstrate that psychological stress aggravates pruritus. Both antidepressants and psychological interventions, including cognitive-behavioral therapy (CBT), improve pruritus and psychogenic itch.

Stress and Autoimmune Disorders. Stress has been found to precede autoimmune disorders and exacerbate symptoms. In a negative feedback loop, proinflammatory cytokines stimulate the HPA axis, resulting in the secretion of glucocorticoids, which suppress the immune response. However, in autoimmune disorders such as rheumatoid arthritis, it is thought that the counter-regulatory glucocorticoid response is not fully achieved. Rheumatoid arthritis patients show a relative hypofunctioning of the HPA axis despite the degree of inflammation. Data suggest that chronic stress provokes increases in the production of the proinflammatory cytokine IL-6, which correlates with disease symptoms, including fatigue, pain, and functional limitations. Moreover, the presence of depression in rheumatoid arthritis patients undergoing stress is associated with exaggerated increases in IL-6, a biomarker predictive of disease progression.

Cancer and Psychoneuroimmunology. In recent history, cancer has been reconceptualized from an acute disease model to a chronic disease with interest shifting towards intraindividual differences that affect clinical outcomes. In particular, researchers are beginning to conceptualize the psychoneuroimmunological pathways that affect the quality of life of cancer patients. Elevated levels of cytokines have been associated with cancer stage in melanoma, breast cancer, bone sarcoma, and post-treatment fatigue in breast cancer survivors. Recent data suggest that memory deficits in newly diagnosed breast cancer patients are related to an activated immunologic response. Stressful life events and depression are associated with higher mortality rates across various cancers. In a study of ovarian cancer patients, those with a greater sense of social attachment were less likely to die. Furthermore, perceived social isolation or loneliness predicted morbidity and mortality risk across different ages. While data remains limited on psychoneuroimmunology's relationship to cancer, there is continued interest in discovering immune mechanisms to improve clinical cancer care.

Stress and Metabolic Syndrome. Stress during development, in the form of early life adversity, has been linked to subsequent development of metabolic diseases including obesity, insulin resistance, and type 2 diabetes. There are two primary proposed mechanisms for the association. First, direct biologic effects include stress-induced elevated blood levels of triglycerides, free fatty acids, glucose, and insulin, as well as permanent alterations in the HPA axis. Second, indirect effects through mental health and risk behaviors can lead to the development of the metabolic syndrome.

STRESS AND PSYCHIATRIC ILLNESS

Historical Conceptualization

Mental disorders have been ascribed to an imbalance in bodily humors, to the influence of external spiritual or other supernatural forces, and moral or somatic deficiency. In the early-to-mid-1800s, a school of thought emerged, led by Philippe Pinel (1745 to 1826) of the Salpêtrière in Paris and Amariah Brigham (1798 to 1849) in America, which held that the expression of mental illness was affected by life circumstances and, more broadly, by societal factors. These various influences converge in the biopsychosocial model proposed by Meyer. This orientation held that an individual's response to stress is modified by several intrinsic factors (e.g., genes, premorbid personality) and extrinsic factors (e.g., social support). The model incorporated individual temperamental and experiential characteristics, such as potential vulnerability (or resiliency) factors or stressful life events, as initiating or exacerbating factors, and a support network as a potentially modifying factor for the occurrence of mental illness. For the most part, this multidetermined model of mental illness dominates modern psychiatric thinking.

It is interesting to see how stress is handled in the diagnostic nomenclature of the fifth edition of the *Diagnostic and Statistical Manual of*

Mental Disorders (DSM-5). This classification does not emphasize the biopsychosocial model of the nature and genesis of psychiatric disorders because the decision to have the DSM-5 (similar to previous DSM editions) expound a *descriptive*, phenomenologic typology, one that theoretical or other subjective interpretations would least govern. Thus, in DSM-5, “major depression” is “major depression,” whether it appears in the context of job loss, divorce, war trauma, or for no apparent reason. Historically, there have been inconsistencies in this approach. Until the 2013 release of the DSM-5, major depression was not considered major depression if it occurred within 2 months of the death of a loved one. Such grief was somehow considered mutually exclusive from major depression, making it difficult to diagnose depression during bereavement, barring extreme symptoms such as suicidality or psychosis. The newest edition removes this bereavement exemption and replaces it with recommendations to consider true major depression even in the context of significant loss if the reaction is inappropriate to the stressor. In a welcome call for thoughtfulness, the text further allows the reader to exercise clinical judgment in assessing the appropriateness of reactions.

Stress can clearly contribute to pathology, so which diagnostic options are afforded by the DSM-5 to explain a depressive syndrome following a major stressor? For example, what if an individual becomes depressed following their parents’ divorce? According to DSM-5, if such a person develops a depressive syndrome, it might be considered an adjustment disorder (if short-lived), but if it is persistent, then a diagnosis of major depression applies. Following a traumatic stressor (e.g., sudden loss of a spouse in a motor vehicle collision), might one better diagnose post-traumatic stress disorder (PTSD)? Not necessarily. Although PTSD is another exception to the DSM-5 rule that the nomenclature should be atheoretical (it is one of the very few disorders in DSM-5 for which the occurrence of a stressful life event is pivotal for diagnosis), it has characteristic symptoms (e.g., re-experiencing symptoms) that are not equivalent to the depressive syndrome. However, comorbidity with major depression is extremely common in PTSD.

Further complicating this diagnostic dilemma is the suggestion that traumatic loss can result in a bereavement syndrome (suggested for further study in the DSM-5 as “persistent complex bereavement disorder”) that is not well captured by either the major depressive disorder (MDD) or PTSD diagnoses although it has elements of both. Thus, like those who struggle to understand the influence of life events on the emotional functioning of their patients, the DSM-5 continues to struggle with the best way to classify and incorporate stressful life events into diagnosis. In the face of this ongoing turbulence, however, much research has been conducted to sharpen the thinking about the relationship between stress and psychiatric illness.

Stress and Psychotic Disorders

There is little reason to suspect that stress plays a role in the pathogenesis of schizophrenia. Most current etiologic theories of schizophrenia focus on genetic and prenatal environmental factors or a combination thereof. There is, however, ample evidence that adverse life events and stressful social or familial milieu play an important role in determining the course of illness in general and episodes of relapse in particular.

Chronic interpersonal stress, usually studied in the family context, is an important risk factor for relapse in schizophrenia. Studies of “expressed emotion,” a proxy for familial tumult, have demonstrated that persons with schizophrenia who reside in homes with high-expressed emotion tend to relapse at twice the rate of those who live with families who have low-expressed emotion. Other stressors that plague the lives of many patients with schizophrenia (e.g., poverty, homelessness, and criminal victimization) also powerfully affect the clinical course of their illness. Combined with evidence that schizophrenia renders the individual more susceptible and sensitive to the negative effects of even minor stressors, it is obvious that stress plays a major role in the course of schizophrenia. Numerous studies have examined the effects on relapse rates of psychosocial interventions that seek to enhance schizophrenic patients’ ability to cope with stressful events.

Affective Disorders

Stressful life events and affective illness share a complicated relationship. As discussed earlier, studies of stress face common methodologic challenges. This challenge is particularly clear in studies of stress and depression because there is a tendency for depressed patients to selectively recall negative life events and associate these events with their current difficulties. Nonetheless, there is evidence that depressed patients experience more negative life events than nondepressed patients. This difference may be accounted for predominantly by an excess of “dependent” events (i.e., events that may conceivably have been *caused* by the individual’s actions while depressed, such as losing a job due to poor performance or failure in school), as opposed to those independent of the depression (e.g., illness of a family member).

Another way of thinking about the relationship between stress and depression is to focus not on specific life events, per se, but rather on the effects of chronic stress. Studies in this genre are few (e.g., examining rates of depression in mothers of chronically ill children or in Alzheimer caregivers compared to control subjects), but those that exist tend to support a relationship between chronic stress and the propensity for major depression. To the extent that most animal models of depression (e.g., chronic restraint, forced swim) are subchronic or chronic stress paradigms, the study of chronic stress effects on depression in humans offers the possibility for convergence and translation between the preclinical, epidemiologic, and clinical knowledge bases.

There is a fundamental similarity between the neuroendocrine effects of stress on humans and the neuroendocrine abnormalities in major depression. For 50 years, major depression has been recognized as associated with overactivity of the HPA axis, as inferred by elevated cerebrospinal CRH levels, increased plasma cortisol and corticotropin (ACTH) levels, increased urinary free cortisol, increased cerebrospinal fluid cortisol, and cortisol resistance to dexamethasone suppression. There is also evidence that some features of psychotic depression, in particular, may be mediated by hypercortisolemia. Some studies suggest that childhood maltreatment, a strong risk factor for affective and other forms of psychopathology, leaves a neurohormonal “scar” that leads to CRH hypersecretion under stress and maybe a vulnerability factor for mood (and anxiety) disorders.

Biopsychosocial models posit that risk factors on different levels (e.g., experiential and biologic) could influence susceptibility to major depression. Although these models have held sway for many years, only recently has supporting data been reported. Researchers are beginning to discover genetic variants within the stress hormone system that increase the risk for depression and other psychiatric disorders. Dysfunction of the glucocorticoid receptor pathway is reported in MDD and individuals exposed to early stressors. Several studies have shown that genetic variants alter transcriptional responsiveness of glucocorticoid receptor target genes to the glucocorticoid receptor and have further suggested that the risk of developing MDD after stressful life events may be influenced by an individual's sensitivity to cortisol release.

Increasingly, studies suggest that neither the gene nor the life stress alone significantly influenced risk; rather, the interaction of these factors explained the variation in depressive outcomes. Recent advances in epigenetics may specifically elucidate the mechanisms linking external stresses to affective psychopathology. Stress may affect gene expression through chromatin remodeling, histone acetylation, methylation, or DNA methylation. For instance, it has been noted that chronic social defeat stress in mice will increase the transcription of a particular DNA methyltransferase (*Dnmt3a*) in the nucleus accumbens, which in turn will increase depressive behaviors after mild social defeat exposures. This effect can be reversed by selectively blocking DNA methylation in that region.

Thus, contemporary genetic and epigenetic research yields empirical evidence of what has been viewed as a truism in contemporary psychiatric teachings: Genes and stressful life events operate in concert to elicit depressive episodes. However, it must also be considered that these types of gene-by-environment interactions may explain risk not only for major depression but also for an array of other stress-influenced mental disorders (e.g., anxiety disorders).

Anxiety Disorders (other than PTSD)

Compared to the voluminous literature on stress and depression, there is less empirical research on the relationship between stress and anxiety disorders. It is nonetheless apparent that panic disorder frequently has its onset or recrudescence in the context of stressful life events. In particular, there is evidence to suggest that either interpersonal conflict or serious illness (in a significant other) may trigger the onset of panic disorder in susceptible individuals. Given that patients with panic disorder are often somatically preoccupied and likely to interpret physical sensations as life-threatening, it is not surprising that the occurrence of illness in someone close to these patients might precipitate the onset or an exacerbation of their panic.

One difficulty in assessing the relationship between stressors, stress responses, and anxiety disorders may be that generalized anxiety disorder itself may include a pathologic prioritizing of stressors, distorting reported stress burdens for study. That said, at least one major long-term study has identified recent adverse life events as a risk factor for late-onset generalized anxiety disorder. Another found that stressors in the domains of loss and danger were particularly prone to predict the onset of anxiety syndromes. In terms of the effects of early life stressors, there is growing evidence that certain adverse early life events such as childhood emotional, physical, or sexual abuse are risk factors for the later development of panic disorder, particularly in women.

Post-Traumatic Stress Disorder

DSM-5 recognizes the existence of a group of disorders that are, by definition, stress-related. Included here are PTSD and autism spectrum disorder (ASD). ASD has its onset following particularly traumatic (often life-threatening) events, such as violent assault or serious accidents, and is denoted by the presence of prominent dissociative symptoms (e.g., derealization, numbing). When ASD occurs following trauma, it identifies a subset of individuals at several-fold increased risk for the subsequent development of PTSD (and major depression).

Whereas PTSD was not long ago considered to be a disorder of war veterans (15% to 20% of U.S. Vietnam veterans and U.S. soldiers deployed to the more recent Afghanistan and Iraqi conflicts may suffer from this disorder), it is now recognized as a highly prevalent disorder affecting a much larger segment of the population. Among the traumatic stressors that can elicit PTSD are relatively rare events such as natural disasters and frighteningly common events such as criminal victimization. In prior versions of the DSM, it was noted that to qualify for a diagnosis of PTSD, the traumatic event needed to be “beyond the range of normal human experience.” Epidemiologic research has shown that, as much as it is hoped it could be otherwise, horrible things happen to many people and must be considered part of the human experience. Consequently, although the debate continues, it is now generally accepted that any life-threatening (or, perhaps, life-altering) event, however common (e.g., a serious motor vehicle collision), can be considered sufficiently traumatic that it is capable of eliciting PTSD. With this definition in place, recent surveys place the lifetime prevalence of PTSD in the general U.S. population at around 8% to 10%, with a recognition that certain subpopulations (e.g., soldiers, immigrants from war-torn countries, Native Americans living on reservations) may have several-fold higher rates.

The occurrence of PTSD following severe psychological trauma provides the most persuasive evidence in favor of certain kinds of stress-causing psychiatric illnesses. However, it is important to note that the neuroendocrine profile of patients with PTSD is not what one might expect to see following chronic stress, nor is it akin to what is observed in major depression. Whereas the latter two conditions would be expected to be associated with hypercortisolism, many studies of patients with PTSD have instead demonstrated *hypocortisolism*. Although the neurobiologic implications of this finding remain to be elucidated, the message for stress researchers is clear: it is unlikely that a uniform neuroendocrine "stress response" explains the heterogeneity of psychiatric disorders encountered in the wake of psychological stress.

Most severe psychological stressors *do not* result in the development of PTSD (or other forms of psychiatric illness). Even the terror of the attacks on September 11, 2001, resulted in PTSD in only approximately 1 in 10 residents of Manhattan, New York; these data stand as testimony to the resilience of the human psyche. In light of the natural course of timely recovery, it is important for mental health workers to avoid rapidly pathologizing normal reactions to trauma and to debrief cautiously. Certain characteristics of the traumatic stressor or the immediate postexposure response seem to increase the propensity to develop PTSD following exposure. Some risk factors include the magnitude of and proximity to the stressor, the occurrence of physical injury (including even mild traumatic brain injury), and the occurrence of peritraumatic dissociation. Individual differences also contribute to the variance in PTSD expression following trauma exposure; these include female gender, genetic vulnerability, and a history of exposure to early childhood trauma. The following chapter provides a more detailed review of PTSD following critical illness and the postintensive care syndrome (PICS).

There has also been a surge of interest in recent years in the phenomenon of psychological resilience. Whereas serious traumatic stressors almost inevitably result in short-term symptoms and a decline in functioning, some individuals rebound especially quickly from such insults and go on to function extremely well, even in the face of ongoing chronic stress. Determination of individual differences in development and personality, extraindividual factors including social networks and environment, and neurobiologic mechanisms that underlie these observed differences are a subject of intense research scrutiny.

Stress, Personality Traits, and Disorders

The personality trait of neuroticism is directly tied to stress reactivity and the experience of negative, distressing emotions. High neuroticism is a feature of cluster B personality disorders, including borderline personality disorder (BPD). Alterations of stress perception and response have been observed in patients with personality disorders. Exposure to traumatic experiences, particularly early in development, has been associated with subsequent diagnosis of BPD. This association suggests a bidirectional connection between personality and stress.

PHARMACOLOGIC APPROACHES TO STRESS MANAGEMENT

As discussed above, stressors worsen most psychiatric conditions. When major psychiatric conditions like depression or psychosis are present, the most appropriate medications are those for the primary psychiatric diagnosis. However, more typically, the picture is primarily one of anxiety or insomnia, perhaps not of as great a magnitude as that found typically in anxiety disorder, but still very distressing to the patient. In such settings, pharmacologic therapy with anxiolytic medications or short-term hypnotics may be helpful but carries risks. The field is of two minds on this topic and has historically oscillated between extreme viewpoints. Besides the risk for drug dependence, some feel that benzodiazepines may impair the efficacy of psychotherapy by offering a temporary alternative to developing coping skills.

On the other hand, depriving patients of such medication when they are in crisis is unconscionable. Short-term use of low-dose benzodiazepines considerably helps patients struggle with unusual stressors. This intervention in association with supportive psychotherapy is extremely effective in crisis intervention.

COGNITIVE-BEHAVIORAL APPROACHES TO STRESS MANAGEMENT

William James famously suggested that the greatest weapon against stress is one’s ability to choose one thought over another. CBT methods are increasingly used to help individuals better manage their responses to stressful life events. These treatment methods are based on the notion that cognitive appraisals about stressful events and the coping efforts related to these appraisals play a major role in determining the stress response.

Lazarus et al. have developed stress and coping models that serve as the conceptual foundation for most CBT stress management protocols. These models maintain that two types of cognitive appraisal are especially important in mediating reactions to stress. The first, *primary appraisal*, refers to how a person evaluates the significance or meaning of a given event. When events (e.g., being diagnosed with a chronic illness such as rheumatoid arthritis) are appraised as harmful, threatening, or a loss (e.g., “Because of my rheumatoid arthritis, I will never be able to do anything I really care about”), the individual is more likely to become anxious, depressed, and withdrawn. However, if the same event is viewed as challenging (e.g., “Although my arthritis prevents me from doing some physically demanding activities, there are many other important and meaningful activities that I can do and enjoy”), then more positive outcomes are more likely to occur. *Secondary appraisal* refers to the process of evaluating what can be done about the stressful event and whether or not what can be done will work. Several factors can influence such secondary appraisals, including their learning history, breadth of individuals’ repertoire of coping skills, mastery of specific coping skills, and expectation that their skills will be effective.

CBT approaches to stress management have three major aims. The first aim is to help individuals become more aware of their cognitive appraisals of stressful events. The second aim is to educate individuals about how their appraisals of stressful events can influence negative emotional and behavioral responses and help them reconceptualize their abilities to alter these appraisals. The third aim is to teach individuals how to develop and maintain a variety of effective cognitive and behavioral stress management skills.

The Efficacy of CBT in Stress Management

Several controlled studies have documented the efficacy of CBT stress management protocols. Three such controlled studies provide good examples of this type of research.

Individuals who are in the early stages of coping with recovery from breast cancer surgery must cope with treatment-related symptoms and major disruptions in their work and family life. Investigators conducted a study of newly surgically treated breast cancer patients. Approximately 200 patients were randomly assigned to one of two groups: (1) a comprehensive CBT program focusing on stress management skills, or (2) an educational seminar control condition. Patients in the CBT condition met in groups weekly for 10 weeks and received relaxation, problem-solving, and assertion training. At the end of 3 months, patients in the CBT condition showed significantly greater improvements in intrusive thoughts about cancer, cancer-related anxiety, and emotional distress than patients in the control condition. Recently these patients were recontacted and found to have maintained significantly lower depressive symptoms and better quality of life up to 15 years postintervention.

Human immunodeficiency virus (HIV) is another incurable, high-mortality, highly stigmatized disease that carries a heavy stress burden both from the disease itself and the imperfect treatment options. Several studies have applied similar cognitive-behavioral stress management interventions to HIV-positive patients worldwide. One recent study randomized 71 low-income, racial-minority, HIV-positive, and HPV-positive women to either CBT stress management or a psychoeducation control group. The investigators found that the CBT group demonstrated lower depression scores and stronger immune function postintervention than the control group. Several studies of HIV-positive patients in the United States, Iran, Africa, and elsewhere have similarly demonstrated the effect of CBT stress management to significantly decrease depressive symptoms, increase well-being and sleep quality, and often improve biomarkers such as viral load and rates of cervical intraepithelial neoplasia. The therapy has also been found to alleviate functional dyspepsia symptoms, improve well-being in prostate cancer survivors, and decrease symptom severity and healthcare utilization costs in patients with chronically unexplained medical symptoms.

Stress Management Training

When CBT is used for stress management, training is provided in a wide range of stress management skills. Although space limitations preclude a discussion of all of these skills, it is important to highlight five skills that form the core of almost all CBT stress management programs: Self-observation, cognitive restructuring, relaxation training, time management, and problem-solving.

Table 27.10–1.
A Daily Diary Format for Recording Stress-Related Behaviors

Antecedents (Briefly describe the stressful event)	Stress Rating (Rate how stressful the event was: 0 = not at all, 100 = most stressful event ever experienced)	Behaviors (Indicate how you responded to the event)	Consequences (What was the outcome?)
A minor traffic accident: I was hit from behind at a stop light. Hurt my neck slightly, no damage to the car.	Rating = 70	Thoughts: This kind of thing always happens to me, my boss is really going to be angry at me for missing the morning meeting. Feelings: Frustrated, anxious. Physiologic response: Clenching my teeth, stomach churning. Behaviors: Started shouting at the driver who hit me and then went back to my car to calm myself down and wait for the policeman.	I was able to regain control over my emotions. The policeman came quickly and I was able to make the morning meeting on time.

Self-Observation. One of the most effective ways of helping individuals become more aware of how they respond to problem situations is to have them keep a daily record of their behavior. A daily diary format (Table 27.10–1) is often used with patients who are asked to record how they responded to challenging or stressful events that occurred each day. Entries are made in three columns: antecedents, behaviors, and consequences. In the antecedents column, patients record a specific environmental event that they perceived as stressful. The event might be an interpersonal stressor such as arguing with a spouse, confronting a coworker, or an intrapersonal stressor such as a flare in pain symptoms or a major natural disaster. To assess appraisals of the stressor, evaluators ask patients to rate how stressful they found this event on a 0 to 100 scale (e.g., 0 = not at all stressful, to 100 = the most stressful event ever experienced). Patients are instructed to record their behavioral reactions to the stressful event in the column marked behavior. The term *behavior* is broadly defined so that it encompasses not only overt behavioral responses but also more covert cognitive, affective, and physiologic responses. Thus, in the behavioral response column, individuals might indicate specific coping behaviors they engaged in (e.g., avoidance), cognitive responses to the stressor (e.g., “I am overwhelmed and can’t deal with this right now”), negative emotions (e.g., anxiety, depression), and aversive physiologic responses (e.g., increasing sweating, heart rate, or nausea). In the column marked consequences, patients record the outcomes resulting from their behavior in this situation.

Self-observation is effective for several reasons. First, it makes individuals more aware of behaviors they usually fail to notice. Individuals whose deal with social anxiety by avoidance, for example, may avoid potentially stressful situations in a more or less automatic fashion (routinely refuse invitations), or use very subtle forms of avoidance (arriving or leaving a party early). They may arrange their lifestyle (work alone) to have very few chances to confront their fears. By keeping a record of their behavior, individuals often begin to recognize these behavioral patterns. A second important feature of self-observation is that it helps individuals see how their cognitive-behavioral responses are linked to situational antecedents and consequences. Participants who keep daily diaries are often struck by the variations in their reactions across different events. For example, while they may cope effectively with one stressful event (e.g., meeting an important work deadline), they find themselves unable to deal with another stressful event (e.g., an argument with their teenage son or daughter). Self-observation highlights the role that coping efforts and cognitive appraisals and emotional and physiologic responses play in explaining these very different responses to stressful situations.

With the ubiquity of smartphones, the authors have seen growing interest in daily record-keeping of both behavioral and biologic data collected for individual patients. A wide range of apps have made it easy to document symptoms and situational factors (precipitating events, social context, etc.) in real-time. This approach, called *ecologic momentary assessment*, enables a person to capture stress and coping processes much closer to their occurrence and can increase the accuracy and validity of self-observation. Automatically graphing the timing of anxiety exacerbations, for instance, can aid patients in identifying patterns in order to better understand and change their responses to daily stressors.

Cognitive Restructuring. A hallmark of CBT is its insistence that cognition plays a central role in the stress and coping process. In CBT, cognitive appraisals about stressful events are considered the key factor in determining stress-related responses. Given this emphasis on cognition, it is not surprising that a major thrust of CBT approaches to stress management is helping participants become aware of and change their maladaptive thoughts, beliefs, and expectations.

CBTs have developed protocols for helping distressed individuals restructure dysfunctional or maladaptive thinking patterns. Such cognitive restructuring begins by educating patients about the cognitive model and that underlying cognitions are major determinants of emotional responding. According to the cognitive model, emotional reactions that are excessive or prolonged are often the result of cognitions that are distorted or dysfunctional.

The second step in cognitive restructuring is monitoring and analyzing dysfunctional thoughts. Table 27.10–2 lists several stress-related thoughts that might be identified through daily diary records. As can be seen, each of these thoughts can be linked to a negative emotion. A careful analysis of each thought also reveals an inherent error in logic. Certain types of errors, or cognitive distortions, occur quite frequently, and individuals can, thus, be taught to recognize them.

The third step in cognitive restructuring is to challenge and change cognitive distortions. Participants are often provided with a list of “key questions” (Table 27.10–3) they can apply to their negative thoughts. They are asked to select one or two questions that are useful in identifying the underlying logical problems in their thinking. Finally, participants develop a rational response that represents a more accurate and helpful cognitive response to the situation.

Table 27.10–4 illustrates how one might change various overly negative stress-related thoughts to more adaptive rational responses. If the rational response is to be effective, it must be used frequently, ideally each time the negative thought occurs. To ensure this occurs, therapists give patients home practice assignments in which they are asked to expose themselves to stressful events, monitor their negative thoughts, and respond with a more adaptive thought pattern. They are also asked to keep records of the resulting consequences, both changes in emotions and behaviors.

Table 27.10–2.
Stress-Related Negative Thoughts

Stressor	Cognition	Emotion
Painful medical procedure	I know I won’t be able to tolerate the pain.	Fear
Argument with a friend	She will never speak to me again.	Discouragement
Coworker injured on the job	It is my fault. I should have warned him about the danger.	Guilt
Late for a deadline	It’s hopeless, I’m always messing things up.	Depression
Spouse criticizes you	He just doesn’t get it. He will never understand what I am trying to do.	Anger

Cognitive restructuring is receiving increasing attention among researchers interested in stress management. Some of the most interesting topics receiving recent attention include how cognitive restructuring alters brain processes related to stress and coping and how cognition changes relate to sudden gains achieved during treatment.

Relaxation Training. Relaxation skills can be very helpful in managing stress. When individuals learn to relax, their overall muscle tension is reduced, as does their overall level of autonomic arousal. Individuals who can relax are also more likely to be able to think more rationally and restructure negative cognitions when faced with stressful events. They are also more likely to rest and sleep better, reducing the fatigue that often accompanies stress. Finally, relaxation skills may help interrupt maladaptive behavior patterns. Progressive relaxation training usually consists of a series of structured exercises involving tensing and relaxing major muscle groups slowly. The exercises are introduced during treatment sessions, and patients are encouraged to practice them daily at home. One of the most important tasks in relaxation training is helping patients learn how to generalize their skills in relaxation from home practice sessions to stressful and demanding daily life events. Several techniques are used to enhance generalization. The first is a method called *differential relaxation*. In differential relaxation, the trainee is instructed to engage in a daily task and only use the necessary muscles for carrying out the task. For example, while writing one's name, there needs to be activity in the hands and arms muscles, but the muscles in the face, lower trunk, legs, and feet can be deeply relaxed.

Differential relaxation is initially practiced in relatively easy situations (e.g., sitting or standing) where the demands are minimal. Over time, however, patients learn how to use differential relaxation in challenging and stressful situations. A second generalization method is the *mini-practice*. This exercise consists of a brief relaxation procedure in which the trainee takes a deep breath and then, while slowly exhaling, focuses on sensations of relaxation flowing downward from the face muscles to those of the neck, shoulders, trunk, and legs. The mini-practice takes approximately 30 seconds and thus can be done repeatedly over a day. Individuals are encouraged to develop innovative reminders to cue themselves to perform frequent mini-practices every day. As individuals increase this frequency, they often report that relaxation becomes more or less an automatic and habitual response.

Table 27.10–3.
Key Questions for Challenging Negative Automatic Thoughts

Is there an alternative explanation?
What is the evidence that this thought is true?
What will be the effect of continuing to think this way?
What is the best outcome, worst outcome, and most realistic outcome?
What is the likelihood that this will happen?

Adapted with permission of Beck JS. *Cognitive Therapy: Basics and Beyond*. Guilford Press; 1995.

Relaxation training is frequently used in managing stress in persons suffering from chronic medical disorders such as cancer, cardiovascular disease, and chronic pain. For some people, relaxation training and its variant of meditation can have powerful effects on physiologic and psychological responses to stressors. The heart rate response to intravenous isoproterenol (Isuprel) is a venerable tool of autonomic physiology with highly predictable results. A serendipitous case observation found that the heart rate response to pharmacologic doses of isoproterenol was blocked in a patient who spontaneously started to meditate during the infusion (Fig. 27.10–7).

Time Management. Time management methods are designed to help individuals restore a sense of balance in their lives. The first step in training in time management skills is designed to enhance awareness of current time use patterns. To accomplish this goal, therapists might ask individuals to record how they spend their time each day, noting the amount of time spent in important categories such as work, family, exercise, or leisure activities.

The second step in time management is designed to help individuals set their priorities. An important distinction can be drawn between urgency and importance in setting priorities. To explore this distinction, it is helpful to envision a table with the dimension of importance and the dimension of urgency along the other axis. Patients are then encouraged to enter into each quadrant of the table examples of activities from their own lives and to reflect on the patterns they observe. Individuals may note that much of their time is spent doing activities in the urgent-not-important quadrant or the not-urgent-not-important quadrant. Alternatively, they may report spending little or no time engaging in activities in the not-urgent-but-important quadrant. Patients are encouraged to discuss the benefits of increasing the priority they give to such important but less urgent activities such as exercising, resting, relaxing, or doing volunteer work. They are also encouraged to try to match the time spent engaging in such meaningful activities with the priority they place on the activity.

Table 27.10–4.
Negative Thoughts and Rational Responses

Negative Thought	Rational Response
I have no control over this situation.	It's true that I can't change what happened, but I can control how I respond.
I can't return to my old job, I am worthless.	Even though I may not be able to do the same kind of work, there are a number of other jobs that I may be able to do.
It's not fair that I am being treated this way.	Life is not always fair. Sometimes things go my way, other times they don't.
I am a burden on my family.	My family has had to give me more support lately, but I do a lot of things for them as well.
No one really cares about me.	There are people in my life (e.g., spouse or child) that I can turn to for support and love.

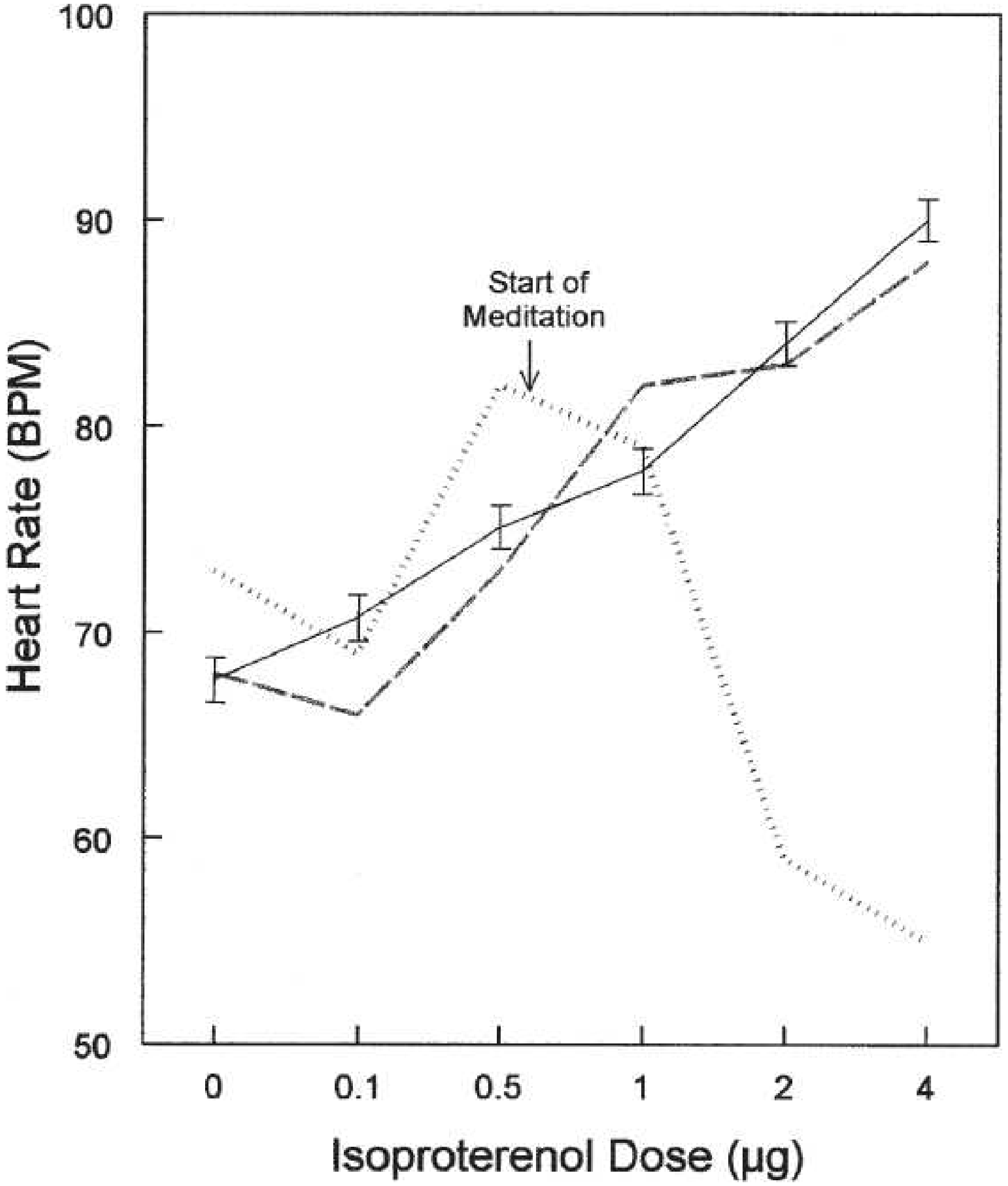

FIGURE 27.10–7. The effect of meditation on heart rate response to isoproterenol. The patient (*dotted line*) began spontaneously meditating and her heart rate dropped. The solid line presents the expected heart rate response. The patient was retested and instructed not to meditate (*dashed line*) and her response was then typical. BPM, beat per minute. (Reprinted from Dimsdale JE, Mills PJ. An unanticipated effect of meditation on cardiovascular pharmacology and physiology. *Am J Cardiol*. 2002;90(8):908–909. doi: 10.1016/s0002-9149(02)02726-1.)

Goal setting focuses on specific behavior change targets that fit an individual’s long-term priorities. First, the individual is asked to identify a goal to be accomplished in the next week. The goal, for example, might be to lose 2 lb or to sign up for an exercise program. Second, patients are asked to critically analyze the goal to ensure it is reasonable, specific, and personally meaningful. In the third step of goal setting, individuals are asked to reconfirm a time frame (e.g., 1 week) in which they will meet the goal. This step is important in reducing procrastination, a common obstacle to effective time management. Finally, the individual's efforts in reaching the goal are reviewed. Successful completion of a goal that has been long neglected can be very rewarding.

The evidence base for time management is growing. One meta-analysis of 32 empirical studies of time management found that the increased use of time management behaviors was significantly related to better health, lower levels of stress, and higher job satisfaction.

Problem-Solving. Problem-solving involves several basic steps. The first is problem identification. In this step, an individual tries to identify the key problematic aspects of a stressful event, such as problematic behaviors, thoughts, feelings, and physiologic responses. For example, dealing with the aftermath of a hurricane may be perceived as stressful because it is associated with (1) particular behavioral problems, for example, having to clean up while simultaneously caring for one's children; (2) overly negative thoughts, for example, "We will never recover from this event"; (3) negative feelings, for example, fear, anger, and depression; and (4) physiologic responses, for example, muscle tension, fatigue, and sleep problems. Helping individuals understand the various facets of their problematic responses is often helpful because it suggests what skills might be particularly useful. For example, if the problem is mainly a somatic one (increased tension), relaxation-based strategies may be the most effective.

The second step in problem-solving is generating alternatives. Brainstorming is used to generate a wide range of alternative problem solutions. The rules of brainstorming include: (1) individuals offer as many solutions as they can, even if the solutions do not immediately seem reasonable, (2) no criticism is allowed, and (3) there should be an attempt to mix and match solutions in creative ways.

The third step in problem-solving involves evaluating the alternatives and selecting the best solution. To begin this process, therapists can ask a participant to rate each solution's likelihood of a positive outcome. It is often helpful to discuss some factors that influence how the participant makes these ratings (e.g., cost, availability of resources, effects of the solution on others). Typically, one solution stands out as the best and most practical.

The final step of problem-solving involves implementing the solution. Participants try to apply the best solution in the problem situation and then review their progress with the therapist.

There is growing recognition that the development of problem-solving skills is crucial to preventing relapse and maintaining the effects of CBT. Participants who learn how to approach a problem situation and formulate and implement effective coping options benefit in several important ways. First, they can often anticipate problems and make plans to avoid them. Second, they cope with setbacks more effectively and can get their coping efforts back on track after a setback to prevent a major relapse. Finally, they are better prepared for termination of therapy because they have learned to engage in a form of "self-therapy" in which they can pinpoint and solve their problems.

Psychological First Aid. Like traditional first-aid, psychological first aid (PFA) is designed to respond to disasters focused on safety and stabilization. It was developed to reduce the incidence of PTSD following trauma and is intended for use by nonmental health professionals and lay responders. See the chapter on Disaster Psychiatry for a more thorough review of PFA.

FUTURE DIRECTIONS

A 51-year-old single woman was seen in psychiatric consultation for help with severe anxiety and occasional panic 8 months after being stabbed by a stranger in an apparent gang initiation assault. She survived without permanent injury but required a chest tube and a short ICU stay at the time and had been afraid to leave her home unaccompanied since the assault. Over the next few months, she began to experience nightmares, insomnia, loss of appetite, and episodes of severe shortness of breath, flushing, and chest pain lasting 15 to 20 minutes about once a week, sometimes leading to emergency department visits. An emergency physician prescribed her low-dose benzodiazepines, which helped her sleep but had begun to create a dependency.

In the consultation clinic, she was jumpy, fretful, and underweight. She reported feeling "stupid" for having been out alone at night. An SSRI was started, and her appetite, hypervigilance, and insomnia improved. She no longer required benzodiazepines. However, her anxiety regarding leaving the house remained, as did her fear of panic attacks and self-blame.

Cognitive-behavioral therapy helped her see this self-blame as a cognitive distortion, and she practiced countering this with affirmations that reassigned the guilt to her attacker. She practiced breathing and muscle relaxation exercises that reduced the fear associated with her panic attacks, and they eventually abated. Her psychiatrist helped her develop a gradual exposure plan to get her out of the house unaccompanied, first briefly in the daytime, then ultimately after dark alone. After 3 months of treatment, she had essentially returned to her full level of functioning.

By way of conclusion, the case study above demonstrates many of the issues discussed in this section—the confluence of adverse life events, physical illness, and emotional distress, and the pharmacologic and psychotherapeutic treatment of stress-related psychopathology.

How would this case be changed 10 years from now? How would it be conceptualized in future taxonomies of psychiatric illness? What sorts of additional information would tomorrow's clinician try to obtain to treat this patient? In other words, how will "stress" be evaluated and treated by psychiatrists in the future? Stressful life events have been a fixture throughout human history, though the nature of the stressors transforms over time. Crime and violence, for example, are not going to go away. Indeed, with increased population density, climate change, and volatile economies, such transgressions may afflict ever-greater populations. Inequity and racism are increasingly recognized sources of stress that impact disparities in health, including mental health and cognitive outcomes. Psychiatry has an important role in understanding the mechanisms through which determinants of health impact mental life across generations.

Perhaps, in the future, physicians may be better able to forecast individuals' vulnerabilities to such stressors. The extensive epidemiologic literature from military psychiatry and disaster psychiatry are consistent in finding a spectrum of risk among those exposed to major stressors. This finding has led to an emphasis on intervening in those at higher risk. Typically, this has meant focusing on the relatively socially isolated individuals, as social isolation is repeatedly observed as a risk for poorer outcomes after trauma exposure. With increasing research, physicians may also find that interventions are particularly important for individuals with certain life experiences or genetic legacies. With advances in understanding individual vulnerabilities comes the opportunity for early identification and intervention. Future research will no doubt test the efficacy of early treatment with psychopharmacologic, cognitive-behavioral, or combined treatments.

Greater understanding and standardization of the concept of resilience will allow for measurement and rigorous study of factors that predict adaptive outcomes. This will guide efforts to develop systems that foster resilience. Considering resilience as a process involving multiple systems emphasizes the key role of systems and networks beyond the individual (family, community, environment) in weathering challenges. To date, most approaches to managing stress have focused solely on the individual. However, with changes in the way health care is being delivered, there will be a greater need to involve family members, caregivers, and communities in the treatment process. Stress occurs in a social context and, for this reason, involving significant others in the process of both assessment and treatment makes sense. Significant others provide an important role as informants about the patient's status; they also can play an active role in helping patients better understand their adaptation to stress and encouraging the adoption and continued use of effective stress management interventions.

As medical science advances, new treatments and management approaches are emerging. Psychiatry will continue to play a significant role in coping with both illness and treatment stresses. Psychiatrists will also contribute to further research measuring how stress contributes to the development and trajectory of the disease.

Psychopharmacological interventions will assuredly also be different in the future. The days of prescribing barbiturates to help patients deal with stress are long gone. These dated interventions have disappeared, along with telegraphs and slide rules, to be replaced by more effective and safer agents. As more is learned about the fundamentals of stress, anxiety, and depression, it is only a matter of time before benzodiazepines and SSRIs will be supplanted by pharmacologic agents with different actions and (hopefully) stronger benefits and fewer risks. Moreover, if the promise of personalized medicine is fulfilled, then the ability to match treatments to patients will also improve.

As the current understanding advances and interventions evolve, psychiatrists will continue to provide patients with the tools and therapies required not just to overcome but actually to understand their stress. Looking ahead and to the future, it is important to not forget the insights already recognized at the dawn of stress research. While stress can be a hazard or even a risk factor, it can also be a potent constructive force in patients' lives.

ACKNOWLEDGMENTS

The authors would like to thank Joel Dimsdale, Francis Keefe, and all authors of previous editions of this chapter.

Further Readings

Antoni MH, Dhabhar FS. The impact of psychosocial stress and stress management on immune responses in patients with cancer. *Cancer*. 2019;125(9):1417–1431.

Beck AT, Weishaar M. Cognitive therapy. In: Wedding D, Corsini RJ, eds. *Current Psychotherapies*. 10th ed. Brooks/Cole Cengage Learning; 2013:231.

Bourvis N, Aouidad A, Cabelguen C, Cohen D, Xavier J. How do stress exposure and stress regulation relate to borderline personality disorder? *Front Psychol*. 2017;8:2054.

*Cathomas F, Murrough JW, Nestler EJ, Han MH, Russo SJ. Neurobiology of resilience: interface between mind and body. *Biol Psychiatry*. 2019;86(6):410–420.

Daviu N, Bruchas MR, Moghaddam B, Sandi C, Beyeler A. Neurobiological links between stress and anxiety. *Neurobiol Stress*. 2019;11:100191.

Epel ES, Blackburn EH, Lin J, et al. Accelerated telomere shortening in response to life stress. *Proc Natl Acad Sci USA*. 2004;101(49):17312–17315.

*Gold PW. The organization of the stress system and its dysregulation in depressive illness. *Mol Psychiatry*. 2015;20(1):32–47.

Goyal M, Singh S, Sibinga EMS, et al. Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. *JAMA Intern Med*. 2014;174(3):357–368.

*Haglund MEM, Nestadt PS, Cooper NS, Southwick SM, Charney DS. Psychobiological mechanisms of resilience: relevance to prevention and treatment of stress-related psychopathology. *Dev Psychopathol*. 2007;19(3):889–920.

Hirohata S, Abe A, Murasawa A, Kanamono T, Tomita T, Yoshikawa H. Differential effects of IL-6 blockade tocilizumab and TNF inhibitors on angiogenesis in synovial tissues from patients with rheumatoid arthritis. *Mod Rheumatol*. 2017;27(5):766–772.

Huang F, Wu X. Brain neurotransmitter modulation by gut microbiota in anxiety and depression. *Front Cell Dev Biol*. 2021;9:649103.

Huang H, Yan P, Shan Z, et al. Adverse childhood experiences and risk of type 2 diabetes: a systematic review and meta-analysis. *Metabolism*. 2015;64(11):1408–1418.

Irwin MR, Opp MR. Sleep health: reciprocal regulation of sleep and innate immunity. *Neuropsychopharmacology*. 2017;42(1):129–155.

Karg K, Burmeister M, Shedden K, Sen S. The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch Gen Psychiatry*. 2011;68(5):444–454.

Kaul D, Schwab SG, Mechawar N, Matosin N. How stress physically re-shapes the brain: impact on brain cell shapes, numbers and connections in psychiatric disorders. *Neurosci Biobehav Rev*. 2021;124:193–215.

Kendler KS, Hettema JM, Butera F, Gardner CO, Prescott CA. Life event dimensions of loss, humiliation, entrapment, and danger in the prediction of onsets of major depression and generalized anxiety. *Arch Gen Psychiatry*. 2003;60(8):789–796.

McGinty EE, Presskreischer R, Han H, Barry CL. Psychological distress and loneliness reported by US adults in 2018 and April 2020. *JAMA*. 2020;324(1):93–94.

Murray C, Griffin ÉW, O'Loughlin E, et al. Interdependent and independent roles of type I interferons and IL-6 in innate immune, neuroinflammatory and sickness behaviour responses to systemic poly I:C. *Brain Behav Immun*. 2015;48:274–286.

Paradies Y, Ben J, Denson N, et al. Racism as a determinant of health: a systematic review and meta-analysis. *PLoS One*. 2015;10(9):e0138511.

Patel SK, Wong AL, Wong FL, et al. Inflammatory biomarkers, comorbidity, and neurocognition in women with newly diagnosed breast cancer. *J Natl Cancer Inst*. 2015;107(8):djv131.

Pirkis J, Gunnell D, Shin S, et al. Suicide numbers during the first 9–15 months of the COVID-19 pandemic compared with pre-existing trends: an interrupted time series analysis in 33 countries. *EClinicalMedicine*. 2022;51:101573.

Réus GZ, Fries GR, Stertz L, et al. The role of inflammation and microglial activation in the pathophysiology of psychiatric disorders. *Neuroscience*. 2015;300:141–154.

Sampson TR, Mazmanian SK. Control of brain development, function, and behavior by the microbiome. *Cell Host Microbe*. 2015;17(5):565–576.

Schneider KE, Dayton L, Wilson D, Nestadt PS, Latkin CA. Distress in the time of COVID-19: understanding the distinction between COVID-19 specific mental distress and depression among United States adults. *J Affect Disord*. 2021;294:949–956.

Shaw W, Dimsdale J, Patterson T. Stress and life events measures. In: John Rush A Jr. ed. *Handbook of Psychiatric Measures*. American Psychiatric Association; 2000:221.

Song H, Fang F, Tomasson G, et al. Association of stress-related disorders with subsequent autoimmune disease. *JAMA*. 2018;319(23):2388–2400.

Stagl JM, Bouchard LC, Lechner SC, et al. Long-term psychological benefits of cognitive-behavioral stress management for women with breast cancer: 11-year follow-up of a randomized controlled trial. *Cancer*. 2015;121(11):1873–1881.

Ulrich-Lai YM, Ryan KK. Neuroendocrine circuits governing energy balance and stress regulation: functional overlap and therapeutic implications. *Cell Metab*. 2014;19(6):910–925.

Ungar M, Theron L. Resilience and mental health: how multisystemic processes contribute to positive outcomes. *Lancet Psychiatry*.2020;7(5):441–448.

Vaccarino V, Shah AJ, Mehta PK, et al. Brain-heart connections in stress and cardiovascular disease: implications for the cardiac patient. *Atherosclerosis*. 2021;328:74–82.

*Zannas AS, West AE. Epigenetics and the regulation of stress vulnerability and resilience. *Neuroscience*. 2014;264:157–170.

Zefferino R, Di Gioia S, Conese M. Molecular links between endocrine, nervous and immune system during chronic stress. *Brain Behav*. 2021;11:e01960.

27.11 Psychiatric Morbidity Following Critical Illness

CHRISTIAN ROMANCHEK, M.D., DIMITRY S. DAVYDOW, M.D., M.P.H., F.A.C.L.P., AND O. JOSEPH BIENVENU, M.D., PH.D.

Mr. Williams was a healthy, busy 35-year-old family man when he developed a cough with fever, diaphoresis, and pleuritic chest pain. At a local emergency department, he was started on oral antibiotics for “walking pneumonia.” However, his wife had to bring him back 2 days later as he had increasing shortness of breath and weakness, and he felt like he was dying. Because of inadequate oxygenation with noninvasive methods (respiratory failure), he was intubated and started on mechanical ventilation. He was diagnosed with acute respiratory distress syndrome (ARDS) and required intensive care for 12 days.

While in intensive care, Mr. Williams did not know where he was and had visual hallucinations of children with animal heads and circular mazes dripping with blood. He was quite agitated and required physical restraints, and his endotracheal tube and Foley catheter were apparently quite uncomfortable. He developed a belief that his penis had been excised and that he would have to pay a large sum to have it reattached. He thought his nurses were variously ignoring and trying to kill him.

After he was extubated and transferred to a general hospital ward, Mr. Williams required physical therapy until he could move around better. However, he remained easily fatigued for months, and his strength returned quite gradually, with substantial effort. He also had trouble getting things done, with his expectations greatly exceeding his abilities. He had trouble with his memory (e.g., where he had put his keys) and difficulty completing sequential tasks (e.g., putting furniture together). The loss of his fast-paced, high-responsibility job had a profound demoralizing effect on him.

Mr. Williams also became very fearful, particularly of infections, as he was told his critical illness stemmed from an infection. He used a mask when cleaning, changed his sheets frequently, and took vitamins and minerals meant to boost his immune system. He feared overdoing it at the gym, and his wife became embarrassed when he frequently spoke of illness to others. When he became ill with a cough or a runny nose, he stayed home or made unnecessary emergency room visits, and he avoided playing with his young son when his son had a cold. He had intrusive memories of wearing an oxygen mask in the emergency department and the sound of the ventilator when he was critically ill. He felt distant from his loved ones, unsupported, and that no one understood what he had been through. He no longer felt he would have a long life. He felt like a burden, particularly because he could not do much around the house. He wondered if he had done something wrong, why this had happened to him. He was frustrated by his slow recovery and avoided others, not wanting to be bothered. Sometimes he had difficulty falling asleep and would anxiously check his pulse and breathing, especially when concerned he was coming down with something. He was also irritable, especially when people did not show respect or empathy.

With advances in intensive care medicine, many more patients are surviving critical illnesses like severe sepsis and ARDS Unfortunately, many critical illness survivors suffer from post-traumatic stress disorder (PTSD), anxiety, and depressive phenomena, as well as cognitive impairment and physical weakness, and their family members are also often quite distressed. To increase awareness of these common problems, a group of clinical outcomes researchers and other stakeholders coined the term “postintensive care syndrome.” This section will touch on all these problems, as they are often relevant to critical illness/intensive care unit (ICU) survivors and their families, but this section will focus particularly on neuropsychiatric problems.

PSYCHIATRIC MORBIDITY IN CRITICAL ILLNESS SURVIVORS AND THEIR FAMILY MEMBERS

What is Stressful about Critical Illness?

In many ways, critical illness–associated distress is easy to understand. Patients who are critically ill with acute respiratory failure experience pain with endotracheal intubation and suctioning, as well as invasive monitoring and therapeutic interventions such as arterial and central intravenous lines, Foley catheters, and chest tubes. In addition, severe sepsis is associated with systemic inflammation and dysfunction of many organs, including the brain and the adrenal glands (leading to a need for endogenous and exogenous catecholamines to maintain adequate blood pressure). With acute brain dysfunction from infection, organ failure, and sedative medications, patients are often unable to process what they are experiencing (e.g., manipulation of a Foley catheter can be misperceived or misinterpreted as rape), and frightening hallucinations and delusions are common. All these stresses occur while patients have lost their autonomy and ability to communicate normally. If patients survive, they are often weak and cognitively impaired; thus, their work, social, and family lives are additionally disrupted by these factors. Similarly, family members face a number of stressors, including uncertainty about their loved ones' survival and ultimate functioning, financial difficulties with disrupted jobs and the need to provide informal care, and additional disruptions in work and social lives, including changing roles within the family.

Adult Critical Illness Survivors

Adult critical illness survivors have a substantially higher prevalence of psychiatric morbidity than expected compared to the general population. For example, around 20% of critical illness survivors suffer from clinically significant PTSD symptoms, approximately 30% with clinically significant depressive symptoms, and approximately 35% with clinically significant nonspecific anxiety symptoms. Studies employing clinician interviews suggest lower but still substantial prevalence of PTSD and depression-related diagnoses. Importantly, psychiatric morbidity may explain, at least in part, the frequently low health-related quality of life (HRQOL) in survivors.

Risk factors for psychiatric morbidity in critical illness survivors include prior psychiatric illness, in-ICU benzodiazepine/sedative doses, mechanical ventilation, and post-ICU memories of frightening, often psychotic, in-ICU experiences. In addition, early post-ICU distress strongly predicts later psychiatric morbidity.

Pediatric Critical Illness Survivors

Pediatric critical illness survivors also have higher rates of psychiatric disorders than their general population counterparts. Patient and parent reports indicate that substantial PTSD symptoms occur in 10% to 28% of child and adolescent critical illness survivors, and clinician interview results suggest frank PTSD in 0% to 21%. Patient and parent reports indicate that substantial depressive symptoms occur in 7% to 13% of this patient population, although, as in adults, major depressive disorder appears less common. Also, as in adults, clinically important psychiatric symptoms are associated with worse HRQOL.

As in adult critical illness survivors, a precritical illness history of psychiatric or developmental problems appears to be the most potent predictor of substantial psychiatric symptoms in children after intensive care. Also, in keeping with adult critical illness survivors, early postillness psychiatric symptoms appear to be predictors of later, more persistent PTSD and depressive symptoms. Notably, parental PTSD, anxiety, and/or depressive symptoms after a child's critical illness and the child's post-ICU psychiatric morbidity appear correlated.

Family Members of Critically Ill Patients

As mentioned, critical illness survivors are not alone in facing substantial distress. Family members of critically ill patients also have high prevalence of clinically important psychiatric symptoms, with concomitant poorer functioning and quality of life. Approximately one-third of parents of critically ill children experience substantial acute stress symptoms. Studies using questionnaires suggest that anywhere from 13% to 56% of relatives of critically ill patients have substantial PTSD symptoms, while 8% to 42% have substantial depressive symptoms. Furthermore, 21% to 56% of family members of critically ill patients have substantial nonspecific anxiety symptoms. As noted above, substantial PTSD, depressive, and nonspecific anxiety symptoms tend to co-occur in family members of critically ill patients.

Women with critically ill family members may be at greater risk of developing substantial PTSD and/or depressive symptoms than men. In addition, family members of patients who die or almost die in the ICU, and those family members who have additional stressors beyond their relative's critical illness, appear at greater risk of clinically important psychiatric symptoms. For family members of patients who die in the ICU, discordance between their preferred medical decision-making role and their actual role and being present at the time of the loved one's death may increase risk. Finally, wives of men with substantial functional limitations after a severe illness may be at increased risk of developing substantial distress compared to husbands of women with similar postillness limitations.

COGNITIVE DYSFUNCTION IN CRITICAL ILLNESS SURVIVORS

Prevalence

A growing body of literature suggests that critical illnesses can affect cognitive functioning substantially and persistently. In studies of ARDS survivors, 70% to 100% had evidence of cognitive impairment at hospital discharge, and 25% to 47% had clinically meaningful cognitive impairment at 2-year follow-up. In studies of general ICU survivors, a wider range (10% to 71%) showed evidence of cognitive impairment 1 year after hospital discharge. Notably, the degree of long-term cognitive impairment present in these patients can be comparable to that seen in patients with moderate traumatic brain injury or mild Alzheimer disease.

Affected Cognitive Domains and Neuroimaging Findings

While critical illness survivors can exhibit impairment in all domains of cognitive functioning, the domains most severely affected are attention and concentration, memory, and executive functioning. Neuroimaging studies of critical illness survivors have correlated these phenomena with cerebral atrophy, ventricular enlargement, and hippocampal volume reductions, as well as connectivity disturbances.

Potential Risk Factors

Although older age is a risk factor for persistent cognitive impairment after critical illness, younger and middle-aged adult survivors can also be affected. Preexisting cognitive impairment and depression appear to increase risk.

Several clinical factors appear to be causally related to, or at least markers of risk of, long-term cognitive dysfunction in critical illness survivors. For example, severe sepsis appears causally related, and delirium in the ICU is atleast a marker of risk for persistent cognitive dysfunction. In-ICU hypoxia, hypotension, and glucose dysregulation have all been associated with persistent cognitive impairment, and other clinical, genetic, and treatment factors may be relevant.

PHYSICAL DYSFUNCTION IN CRITICAL ILLNESS SURVIVORS

Neuromuscular weakness after critical illness or intensive care is very common, particularly early in the course of recovery (during the first year).

“ICU-acquired weakness” encompasses critical illness polyneuropathy and myopathy, as well as disuse atrophy. Polyneuropathy can recover particularly slowly (over several years).

Apparent risk factors for critical illness polyneuropathy and myopathy include hyperglycemia, systemic inflammatory response syndrome, sepsis, and multiorgan dysfunction. The causal factor for disuse atrophy is immobility (bed rest).

Importantly, physical functioning (including the ability to perform instrumental activities of daily living) can remain impaired longer than muscle weakness. Some apparent risk factors for reduced physical functioning include ICU administration of systemic corticosteroids, ICU-acquired illnesses, slow resolution of lung injury, depression, and older age.

INTERVENTIONS FOR PSYCHIATRIC DISORDERS IN CRITICAL ILLNESS SURVIVORS

Adult Critical Illness Survivors

Some experts advocate for psychological interventions to begin while patients are still critically ill in the ICU. Indeed, psychologists administering a simplified cognitive behavioral therapy (CBT) intervention in a trauma ICU appeared to substantially reduce the risk of survivors’ long-term emotional distress and need for further psychiatric treatment. Unfortunately, psychologists are not commonly employed in ICUs, and bedside nurses administering a similar intervention did not seem to have a similar salutary effect in a recent large randomized trial.

Another promising intervention to reduce long-term distress in critical illness survivors is ICU diaries. Unlike normal diaries, ICU diaries are written for patients who are critically ill. Clinicians, often intensive care nurses, write in plain language about the patient’s condition throughout the ICU stay, including diagnostic and therapeutic procedures. Family members are also encouraged to write about their observations, thoughts, feelings, and occurrences outside of the hospital. Sometimes the diaries include photographs of the ICU room (± the patient) and of staff. The rationale for ICU diaries is to help patients create a more coherent autobiographical narrative than they may otherwise have due to acute brain dysfunction while critically ill. Patients can process this information when they are ready and able, at their own pace, sometimes sharing with loved ones. Early studies suggested that ICU diaries are acceptable to patients and families and that they reduce long-term distress. ICU diary programs are becoming fairly common, especially in Europe. Nevertheless, results of recent large clinical trials suggest that ICU diaries may be most useful for patients with substantial emotional distress early in recovery.

Although ICU follow-up clinics are relatively common in parts of Europe, they are not common in North America, and patients face many logistical challenges attending such clinics in any case. Thus, distance interventions employing technology (telephones and computers) deserve consideration and show promise to deliver effective CBT-based interventions. Like ICU diaries, such interventions appear most promising for patients who have substantial distress post-ICU.

Virtual reality interventions also show promise in educating survivors about critical illnesses and intensive care. These interventions will likely help patients comprehend and process their experiences and improve long-term psychiatric outcomes.

Intriguingly, stress doses of hydrocortisone in the ICU also appear to reduce the incidence of post-ICU PTSD phenomena, though glucocorticoids are not prescribed routinely for this purpose.

Children

To date, there has been limited research on interventions to prevent psychiatric morbidity in pediatric critical illness survivors. A randomized controlled trial investigated the efficacy of an emotional regulation and anxiety reduction program for mothers of critically ill children; the children whose mothers received the intervention exhibited fewer behavioral problems following their critical illness. Notably, ICU diaries have also been adapted for use in children.

Families

Results of some studies suggest that ICU diaries prevent long-term emotional distress in relatives. In addition, studies suggest that enhanced communication strategies (e.g., informative brochures, longer care conferences) between critical care health care workers (e.g., physicians and nurses) and family members of dying critically ill patients reduce PTSD, depressive, and general anxiety symptoms in family members. In a recent study, trained facilitators acting as medical liaisons appear to benefit communication quality and end-of-life care decision-making (with reduced length of stay in patients that would ultimately die), although family members showed no benefits in depression, anxiety, or post-traumatic stress symptoms. Flexible family visitation policies have reduced anxiety and depression scores among family members of critically ill patients without increasing burnout rates among the primary clinical team.

Interventions for Post–Critical Illness Cognitive Dysfunction

There have been few studies of interventions to prevent persistent cognitive dysfunction in critical illness survivors. Preliminary evidence supports a role for in-home cognitive and physical rehabilitation, though larger randomized controlled trials are needed.

Several interventions to reduce or prevent delirium in the critically ill have been evaluated. Dexmedetomidine, a central alpha-2 adrenergic agonist, used for sedation with mechanical ventilation, is less deliriogenic than benzodiazepines. Early physical rehabilitation in the ICU, including during mechanical ventilation, reduces the incidence and duration of delirium. However, it remains unclear whether such delirium reductions will lead to better long-term cognitive outcomes.

Interventions for Post–Critical Illness Physical Dysfunction

The most promising interventions to reduce post-ICU weakness involve minimizing sedation and maximizing physical rehabilitation as early as possible, including while patients are critically ill. Such interventions reduce the need for prolonged mechanical ventilation and ICU stays and, as noted, delirium. Although to the authors' knowledge, the effects of these interventions on long-term neuropsychiatric outcomes are unknown, they seem unlikely to be harmful and likely to be beneficial.

FURTHER READINGS

Bienvenu OJ, Colantuoni E, Mendez-Tellez PA, et al. Depressive symptoms and impaired physical function after acute lung injury: a 2-year longitudinal study. *Am J Respir Crit Care Med*. 2012;185:517–524.

Bienvenu OJ, Gellar J, Althouse BM, et al. Post-traumatic stress disorder symptoms after acute lung injury: a 2-year prospective longitudinal study. *Psychol Med*. 2013;43:2657–2671.

Colville GA, Pierce CM. Children's self-reported quality of life after intensive care treatment. *Pediatr Crit Care Med*. 2014;14:e85–e92.

Cox CE, Hough CL, Carson SS, et al. Effects of a telephone- and web-based coping skills training program compared with an education program for survivors of critical illness and their family members: a randomized clinical trial. *Am J Respir Crit Care Med*. 2018;197:66–78.

Cox CE, Porter LS, Hough CL, et al. Development and preliminary evaluation of a telephone-based coping skills training intervention for survivors of acute lung injury and their informal caregivers. *Intensive Care Med*. 2012;38:1289–1297.

Davidson JE, Jones C, Bienvenu OJ. Family response to critical illness: post-intensive care syndrome—family. *Crit Care Med*. 2012;40:618–624.

Davydow DS, Hough CL, Langa KM, Iwashyna TJ. Depressive symptoms in spouses of older patients with severe sepsis. *Crit Care Med*. 2012;40:2335–2341.

Davydow DS, Hough CL, Langa KM, Iwashyna TJ. Presepsis depressive symptoms are associated with incident cognitive impairment in survivors of severe sepsis: a prospective cohort study of older Americans. *J Am Geriatr Soc*. 2012;60:2290–2296.

Davydow DS, Richardson LP, Zatzick DF, Katon WJ. Psychiatric morbidity in pediatric critical illness survivors: a comprehensive review of the literature. *Arch Pediatr Adolesc Med*. 2010;164:377–385.

Fan E, Dowdy DW, Colantuoni E, et al. Physical complications in acute lung injury survivors: a two-year longitudinal prospective study. *Crit Care Med*. 2014;42:849–859.

Friedman JI, Soleimani L, McGonigle DP, Egol C, Silverstein JH. Pharmacological treatments of non-substance-withdrawal delirium: a systematic review of prospective trials. *Am J Psychiatry*. 2014;171:151–159.

Garrouste-Orgeas M, Coquet I, Périer A, et al. Impact of an intensive care unit diary on psychological distress in patients and relatives. *Crit Care Med*. 2012;40:2033–2040.

Garrouste-Orgeas M, Flahault C, Vinatier I, et al. Effect of an ICU diary on posttraumatic stress disorder symptoms among patients receiving mechanical ventilation: a randomized clinical trial. *JAMA*. 2019;322:229–239.

Gries CJ, Engelberg RA, Kross EK, et al. Predictors of symptoms of posttraumatic stress and depression in family members after patient death in the ICU. *Chest*. 2010;137:280–287.

Iwashyna TJ, Ely EW, Smith DM, Langa KM. Long-term cognitive impairment and functional disability among survivors of severe sepsis. *JAMA*. 2010;304:1787–1794.

Jones C, Bäckman C, Capuzzo M, et al. Intensive care diaries reduce new-onset posttraumatic stress disorder following critical illness: a randomised, controlled trial. *Crit Care*. 2010;14:R168.

Jones C, Bäckman C, Griffiths RD. Intensive care diaries and relatives' symptoms of posttraumatic stress disorder after critical illness: a pilot study. *Am J Crit Care*. 2012;21:172–176.

Kross EK, Engelberg RA, Gries CJ, Nielsen EL, Zatzick D, Curtis JR. ICU care associated with symptoms of depression and posttraumatic stress disorder among family members of patients who die in the ICU. *Chest*. 2011;139:795–801.

Needham DM, Davidson J, Cohen H, et al. Improving long-term outcomes after discharge from intensive care unit: report from a stakeholders' conference. *Crit Care Med*. 2012;40:502–509.

Nikayin S, Rabiee A, Hashem MD, et al. Anxiety symptoms in survivors of critical illness: a systematic review and meta-analysis. *Gen Hosp Psychiatry*. 2016;43:23–29.

Pandharipande PP, Girard TD, Jackson JC, et al. Long-term cognitive impairment after critical illness. *NEJM*. 2013;369:1306–1316.

Parker AM, Sricharoenchai T, Raparla S, Schneck KW, Bienvenu OJ, Needham DM. Posttraumatic stress disorder in critical illness survivors: a meta-analysis. *Crit Care Med*. 2015;43:1121–1129.

Peris A, Bonizzoli M, Iozzeli D, et al. Early intra-intensive care unit psychological intervention promotes recovery from posttraumatic stress disorders, anxiety, and depression symptoms in critically ill patients. *Crit Care*. 2011;15:R41.

Rabiee A, Nikayin S, Hashem MD, et al. Depressive symptoms after critical illness: a systematic review and meta-analysis. *Crit Care Med*. 2016;44:1744–1753.

Rosa RG, Falavigna M, Barbosa da Silva D, et al. Effect of flexible family visitation on delirium among patients in the intensive care unit: the ICU Visits randomized clinical trial. *JAMA*. 2019;322:216–228.

Schelling G, Roozendaal B, Krauseneck T, Schmoelz M, De Quervain DJ, Briegel J. Efficacy of hydrocortisone in preventing posttraumatic stress disorder following critical illness and major surgery. *Ann N Y Acad Sci*. 2006;1071:46–53.

Vlake JH, von Bommel J, Hellemons ME, Wils EJ, Gommers D, van Genderen ME. Intensive care unit-specific virtual reality for psychological recovery after ICU treatment for COVID-19: a brief case report. *Front Med (Lausanne)*. 2021;7:629086.

Wade DM, Mouncey PR, Richards-Belle A, et al. Effect of a nurse-led preventive psychological intervention on symptoms of posttraumatic stress disorder among critically ill patients: a randomized clinical trial. *JAMA*. 2019;321:665–675.

White DB, Angus DC, Shields AM, et al. A randomized trial of a family-support intervention in intensive care units. *N Engl J Med*. 2018;378:2365–2375.

Wilcox ME, Brummel NE, Archer K, Ely EW, Jackson JC, Hopkins RO. Cognitive dysfunction in ICU patients: risk factors, predictors, and rehabilitation interventions. *Crit Care Med*. 2013;41:S81–S98.

27.12 Psychocutaneous Disorder

RAVENNA KHANNA, M.D., AND JASON P. CAPLAN, M.D.

EDITOR'S NOTE

Much of this chapter is derived from the exemplary and influential work of Drs. Madhulika A. Gupta and Aditya K. Gupta. Their expertise in the area of psychocutaneous disorders is well known in this country and abroad, especially in Canada, where Dr. M.A. Gupta is a Professor of Psychiatry at the University of Western Ontario and where Dr. A.K. Gupta is a Professor of Medicine at the University of Toronto Medical School. Interested readers are directed to the following source articles for further reading:

Fried RG, Gupta MA, Gupta AK. Depression and skin disease. *Dermatol Clin*. 2005;4:657–664. Gupta MA, Gupta AK. Current concepts in psychodermatology. *Curr Psychiatry Rep*. 2014;16:449–457.

While any organ system is subject to the complex interactions of psyche and soma and their potential to manifest as psychosomatic illness, the skin is unique in that it is the largest organ, the organ most immediately observable to others, and the most immediately accessible to oneself. Thus, the skin may be particularly prone to manifest psychosomatic illness influenced either by disturbances of neurotransmitters, hormones, or cytokines; by psychosocial relationships or disturbances of self; by (conscious or unconscious) self-harm; or by complex interactions of all three.

The field of psychodermatology or psychocutaneous medicine focuses on these interactions between the mind, the brain, and the skin. The interface between psychiatric and dermatologic conditions highlights the common origin of the brain and skin from the embryonic ectoderm. As such, both organ systems are influenced by common hormones and neurotransmitters that impact how they respond to internal and external stimuli. It is important to recognize that interactions between the mind, brain, and skin are multidirectional: psychopathologic factors can play an etiologic role in the development of skin disorders and exacerbate existing skin disorders, while dermatologic conditions may result in psychosocial consequences due to stigmatization and disfigurement. Significant psychiatric and psychosocial comorbidity has been reported in at least 30% of patients with dermatologic illness. Moreover, the vast majority of patients with psychodermatoses do not receive psychiatric care. As such, psychodermatology as a field encourages an interdisciplinary and humanistic approach to the management of patients with co-occurring dermatologic and psychiatric pathology.

PSYCHOSOCIAL IMPACT OF SKIN DISORDERS

While most individuals with dermatologic illness are able to appropriately adjust to their skin disorder and avoid undue distress, some patients with chronic skin illnesses and appearance-altering conditions experience a significant degree of social, psychological, and physical distress. The patient's experience of psychosocial distress is variable and depends on (1) the characteristics of skin disorder itself, (2) characteristics and demographics of the patient, and (3) cultural attitudes related to skin disease (often expressed as stigma).

In patients with dermatologic disorders, the analogy of "skin failure" (comparable to failure of other organ systems in the body like heart or kidney failure) implies a failure of skin as an organ of physical expression, thus leading to rejection, shame, isolation, and psychosocial impairment. Feelings of stigmatization in patients with psoriasis have been grouped into six dimensions: the *anticipation of rejection* (e.g., "I feel physically and sexually unattractive when my psoriasis is bad"); *sensitivity to opinions of others* (e.g., "Sometimes I feel outcast because of my psoriasis"); *feeling of being flawed* (e.g., "I believe other people think that psoriasis patients are dirty"); *guilt and shame* (e.g., "My psoriasis is a source of deep humiliation and embarrassment for my family and me"); *secretiveness* (e.g., "I try not to share with family members away from me, that I have psoriasis"); and *a more positive attitude which is unaffected by negative reactions of others* (e.g., "If my child has to develop psoriasis, he could develop his potential as though he did not have it"). A previous experience of rejection was considered to be a strong predictor of current beliefs. Some individuals with skin disorders noted that their main difficulties arose not from the disease itself but from others' reactions to their disease.

The desire to be "looked at and admired" may be formulated as a normal narcissistic trait. However, in individuals with skin lesions, to be looked at with surprise, fear, or disdain evokes feelings of painful shame and humiliation. People with normal skin are commonly inclined to turn away physically or psychologically from those suffering from skin diseases–a reaction that has been suggested to involve “a fleeting identification” with the afflicted person, followed by a negative empathic response. The visible skin lesion symbolizes how defective and vulnerable one can be and when coupled with observed feelings of shame and embarrassment, invokes a momentary flash of empathic identification. The healthy person proceeds to block out and dismiss from their awareness any compassion for the sufferer’s pain and then turns away and distances themselves from the person with skin disease. A similar phenomenon has been described in individuals with psoriasis, who typically report that they would avoid other people with psoriasis if they did not have active disease.

The emotional reaction to a particular skin condition is often affected by the patient’s understanding of its *origin*. A congenitally acquired condition such as a port-wine stain may invoke a different reaction versus an acquired skin disorder such as a sexually transmitted infection, lichen simplex, or prurigo nodularis (PN). The latter two conditions, for example, are induced by scratching, and diagnosis may invoke a sense of guilt and shame. In conditions with a genetic predisposition such as atopic dermatitis (AD) or psoriasis, the patient’s response may include blaming their parents and feelings of frustration or helplessness.

Associated symptoms of itching, burning, or pain may add to distress, insomnia, anxiety, and depression. The location or appearance of skin lesions might affect the patient’s self-image and how they are perceived by others. Lesions directly visible on the face and hands usually cause the greatest concern. Lesions present on the trunk and limbs can result in avoidance of activities that would require exposure, such as the gym, spas, swimming pools, or the beach. Genital lesions can be particularly embarrassing and limit sexual intimacy. Hair loss can diminish self-confidence and affect social poise. Lesions involving the nails and hands may be disabling in occupations involving transactions with others such as banks or retail stores. Congenital conditions like port wine stains or childhood AD can influence a child’s self-image and general emotional development. In adolescents, chronic skin conditions can significantly impact self-esteem and self-image, affecting the maturation of self, sexual identity, and personality. The course of skin illnesses with a chronic relapsing pattern or treatment-related side effects can have a major impact on patients and their families in an enduring fashion.

Age and gender may be important predictive factors when considering the psychiatric sequelae of a chronic dermatologic illness. A young woman suffering from psoriasis may be more psychosocially impacted than a man in the eighth decade of life with a similar condition. Personality traits can similarly influence a patient’s reactions and subjective experience of the illness. Narcissistic personality disorder is marked by “an expectation of admiration and attention, sense of entitlement, and excessive concern about how others regard them,” and individuals with that diagnosis may be particularly humiliated by a skin disease that impacts their personal appearance. Those diagnosed with a borderline personality disorder may consider skin disease as a specific threat to self-image and autonomy and react with mood instability, severe anxiety, and fear of abandonment when faced with a dermatologic diagnosis. For those with prominent obsessive-compulsive traits, skin disorders may invoke a loss of control over their body and feelings of being unclean or dirty due to skin lesions and topical treatments. The presence of a primary psychiatric disorder can similarly influence the patient’s thoughts and beliefs about the skin lesions. For example, depression may increase the sensation of itching in pruritic skin disorders such as AD, psoriasis, and chronic idiopathic urticaria (Fig. 27.12–1).

KOO CLASSIFICATION OF PSYCHOCUTANEOUS DISORDERS

Although there is no single universally accepted classification system of psychocutaneous disorders, the one most widely accepted was devised by John Koo. It includes five different groupings listed in Table 27.12–1, which also includes the corresponding diagnoses and diagnostic codes from the revised fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) and the diagnostic codes from the International Classification of Diseases (ICD) 10. *Psychophysiologic disorders* are bona fide skin disorders exacerbated by stress. The term *primary psychiatric disorders* refers to cases without “real” skin disease, but where patients present with significant psychopathology and visible skin lesions are self-induced. The term *secondary psychiatric disorders* refers to cases in which patients develop psychological problems as a result of skin disease and associated disfigurement. Some skin conditions, such as acne and alopecia areata (AA), although not life-threatening, may have a significant psychosocial impact on a patient’s quality of life. The term *cutaneous sensory disorders* refers to cases in which patients have unpleasant sensations on the skin, such as itching, stinging, burning, or crawling, with no proven skin-based etiology, and in which a psychiatric diagnosis may or may not be evident. These are considered to be equivalents of chronic pain syndromes. Subcategories of psychocutaneous disorders are not mutually exclusive. For example, acne can be precipitated by stress (a psychophysiological condition), but acne can also cause social anxiety and depression due to disfigurement (a secondary psychiatric condition).

Koo classification includes a separate category to describe the *use of psychotropic medications* in treating certain skin conditions, which may be more efficacious than traditional dermatologic treatments. For example, many antidepressants, such as doxepin, have shown greater efficacy as antipruritic agents when compared to traditional antihistamines. Psychotropic medications commonly used in the treatment of psychocutaneous disorders are summarized in Table 27.12–2.

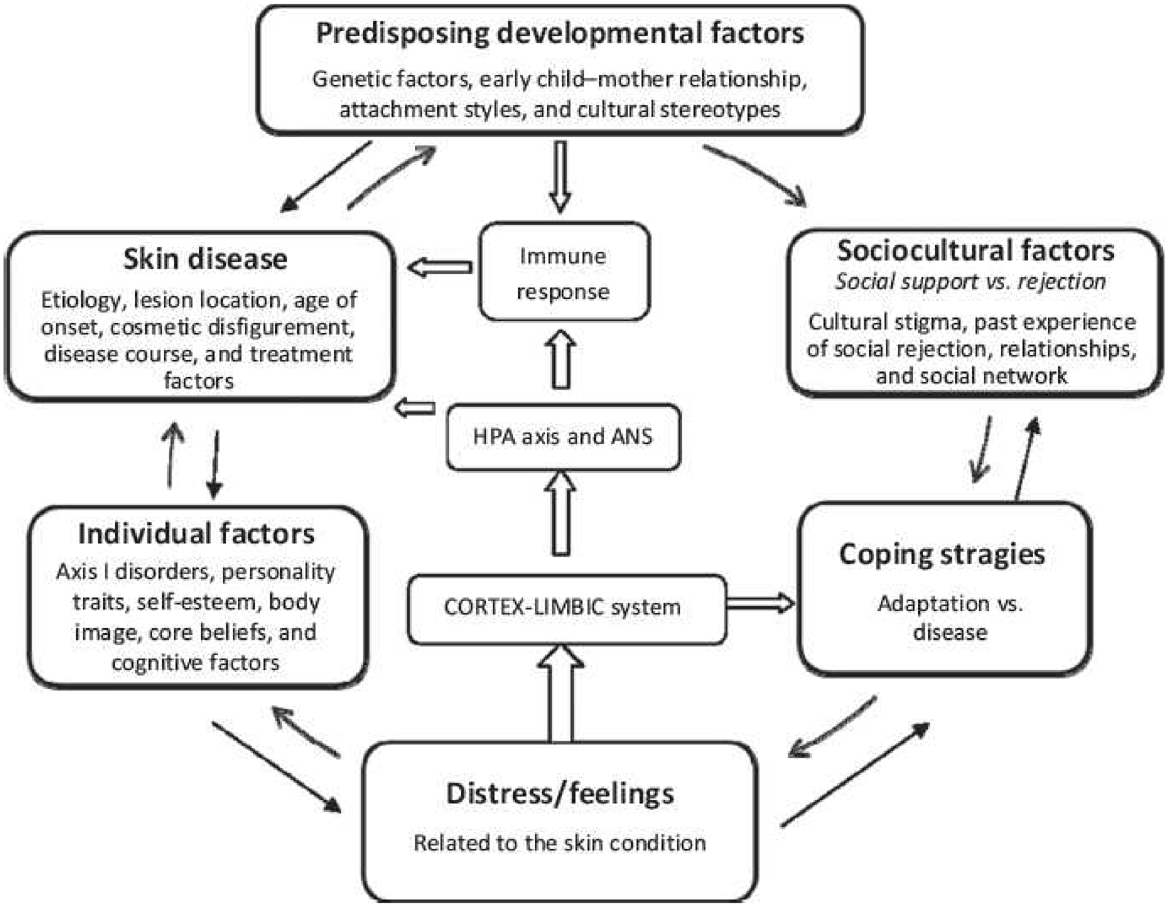
Predisposing developmental factors
Genetic factors, early child–mother relationship, attachment styles, and cultural stereotypes
Skin disease
Etiology, lesion location, age of onset, cosmetic disfigurement, disease course, and treatment factors
Immune response
Sociocultural factors
Social support vs. rejection
Cultural stigma, past experience of social rejection, relationships, and social network
HPA axis and ANS
Individual factors
Axis I disorders, personality traits, self-esteem, body image, core beliefs, and cognitive factors
CORTEX-LIMBIC system
Coping stragies
Adaptation vs. disease
Distress/feelings
Related to the skin condition

FIGURE 27.12–1. Model of psychosocial factors implicated in origin and maintenance of skin-specific affect/distress, and systemic adaptation response. (Adapted and modified from Thompson A. Coping with chronic skin conditions. In: Walker C, Papadopoulos L, eds. *Psychodermatology: The Psychological Impact of Skin Disorders*. Cambridge University Press; 2005:61.)

ESTABLISHED PSYCHODERMATOSES

These diagnoses encompass primary psychiatric disorders and cutaneous disorders, all of which present with both psychopathologic and dermatopathologic features.

Delusional Disorder, Somatic Type: Delusional Parasitosis

Clinical Features. Patients suffering from delusional parasitosis (DP) have the fixed belief that they are infested with parasites. G. Thibiérge first described the condition in the 1890s and called the syndrome *acarophobia*. Other terms used include parasitophobia, entomophobia, monosymptomatic hypochondriacal psychosis, and Ekbom syndrome. DSM-5 uses the term “delusional disorder (parasitosis),” which presents as a single somatic delusion with no impairment of thought processes. Patients often describe the sensation of bugs crawling on or inside the skin; some report seeing bugs or hearing buzzing sounds; and some “bring the parasites” for further examination (usually pieces of hair, skin, or cloth). This behavior is so characteristic that it has been referred to as “the matchbox sign” since patients will often bring this material to clinical encounters in a small container. Often a patient with DP may report a specific precipitant, such as a history of infection with a contagious organism or an actual infestation. Their attempts for self-treatment include repeated washing, checking, and cleaning; excoriation of skin with needles, knives, or fingernails; discarding household items; and excessive use of insect repellents or insecticides (Fig. 27.12–2). Relatives may share the delusion, and fear of contaminating others is common. Most patients seek dermatologic treatment because their delusion involves skin invasion.

Table 27.12–1.
Classification of Psychocutaneous Disorders

Koo's Classification	**DSM-5 Classification (codes listed as DSM-5, ICD-10)**
Psychophysiologic disorders	**(1) Psychological factors affecting other medical condition (316, F54**) Atopic dermatitis, psoriasis, alopecia areata, urticaria and angioedema, acne vulgaris
Primary psychiatric disorders	**(2) Delusional disorder, somatic type (297.1, F22)** Delusional parasitosis, delusions of a defect in appearance, delusions of a foul body odor
	(3) Obsessive-compulsive and related disorders (OCRD) Excoriation (skin-picking) disorder (698.4, F42.2), trichotillomania (312.39, F63.3), body dysmorphic disorder (300.7, F45.22)
	(4) Unspecified OCRD (body-focused repetitive behavior disorder) (300.3, F42.9) Onychophagia
	(5) Factitious disorder (300.19, F68.10): Dermatitis artefacta, psychogenic purpura
Secondary psychiatric	(6) Anxiety disorder due to another medical condition (293.84, F06.4)
	(7) Generalized anxiety disorder (300.02, F41.1)
	(8) Major depressive disorder (296.3, F33 plus modifiers)
Cutaneous sensory disorders	**(9) Unspecified somatic symptom and related disorder (300.82, F45.9)** Chronic idiopathic pruritus, idiopathic pruritus ani/vulvae/scroti

	(10) Somatic Symptom Disorder, with predominant pain (300.82, F45.1)
	Idiopathic glossodynia, dysesthetic (essential) vulvodynia
Psychotropic medication for nonpsychiatric indications	

Data from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, (Copyright © 2022). American Psychiatric Association; Koo J. Psychodermatology: a practical manual for clinicians. *Curr Probl Dermatol*. 1995;7:203–225.

Table 27.12–2.

Psychotropic Medications Used in Treatment of Psychocutaneous Disorders

PRURITUS	**SSRIs and TCAs** (amitriptyline and doxepin): commonly useful in pruritus **Doxepin 5% topical cream**: FDA-approved for short-term management of pruritus in AD **Gabapentin, pregabalin**: effective in prurigo nodularis, neuropathic pain, and fibromyalgia **Aprepitant** (neurokinin receptor [NKR]-1 antagonist): effective in chronic refractory pruritus. NKR-1 is a receptor for substance P, an important itch mediator. **Naltrexone** (opioid receptor antagonist): used in Rx of OCRD
BODY-FOCUSED REPETITIVE BEHAVIORS (under OCRD): SPD, TTM, BDD, dermatitis artefacta	**SSRIs** (fluoxetine, paroxetine, sertraline, citalopram, escitalopram, and fluvoxamine): considered first-line treatment for OCRD **Clomipramine** (TCA): found to be effective in TTM, BDD, OCD; use caution with TCAs (side effects). **Naltrexone** (opioid receptor antagonist): in one study reduced pruritus by 50% in OCRD, body repetitive disorders **N-Acetylcysteine** (NAC): treatment efficacy comparable to CBT in one study for TTM **Olanzapine and aripiprazole**: positive case reports for treatment of SPD and TTM **Mood stabilizers** (lamotrigine, lithium, valproic acid): few positive studies; indicated for Rx of emotional dysregulation/dissociation and comorbid mood disorder

CUTANEOUS DYSESTHESIAS: Glossodynia, vulvodynia, trichodynia/scalp dysesthesia urogenital and rectal pain syndrome	**TCAs** (doxepin, nortriptyline, imipramine, desipramine, and particularly amitriptyline): long h/o use, effective **SSRIs**: considered first-line **Mirtazapine**: useful when anticholinergic side effects of TCAs cannot be tolerated **SNRIs**: have efficacy in pain, neuropathic disorders, and fibromyalgia. **Gabapentin**: useful for neuropathic pain and pruritus, safe, minimal drug interaction, and no abuse potential.
DELUSIONAL PARASITOSIS	**Pimozide**: a potent neuroleptic, most commonly advocated, 60–100% response rate; monitor risk of QT prolongation. **Risperidone, olanzapine, amisulpride**: 75% response rate; preferred over pimozide due to improved side effect profile.

Rx, treatment; AD, atopic dermatitis; SSRIs, selective serotonin reuptake inhibitors; TCA, tricyclic antidepressants; TTM, trichotillomania; SPD, skin picking disorder; OCRD, obsessive-compulsive and related disorders; BDD, body dysmorphic disorder.

Diagnosis. Symptoms of DP can be confused with a specific phobia or obsessive-compulsive disorder (OCD), and the distinction between these conditions has obvious relevance to treatment. The differential diagnosis of DP includes a number of medical, neurologic, and psychiatric disorders, which are listed below in the section on psychogenic pruritus. The effects of certain drugs can mimic DP; cocaine and amphetamines especially may produce a tactile hallucination referred to as “cocaine bugs” or “Magnan sign,” where the affected person believes they have or feel parasites crawling under the skin. Patients with severe alcohol withdrawal may experience visual hallucinations of insects or tactile hallucinations that bugs are crawling over their skin (formication).

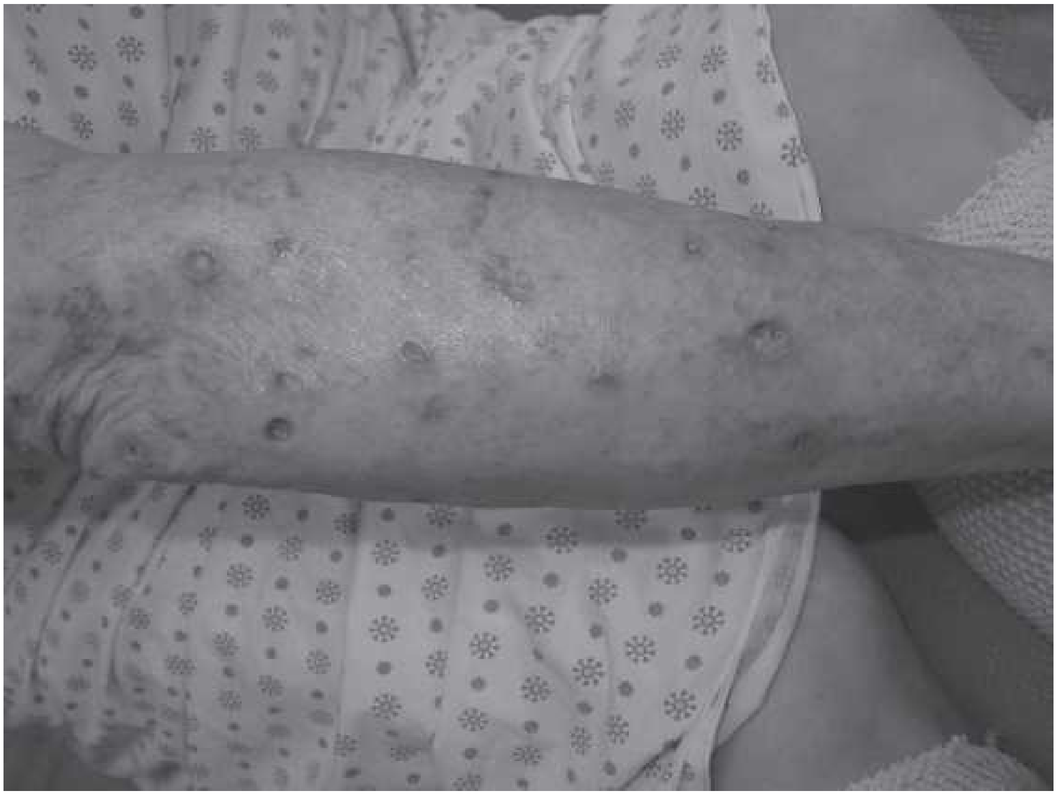

FIGURE 27.12–2. Delusions of parasitosis. Lesions are mostly excoriations. Patient picks on skin with nails and needles. (The Barankin Collection. Wolters Kluwer, 2005.)

Pathogenesis. The psychopathology of DP involving specificity of content and fixed belief has been the focus of many empirical studies. The pathogenesis of DP has been considered a multifactorial process in which psychic, cognitive, and social circumstances act as disorder-predisposing, triggering, and maintaining factors, respectively. The psychic factors may add predisposition to a particular delusional theme and involve complex interactions of age, gender, social situation, and other key experiences. Delusions are considered a dynamic process that only persists if disorder-maintaining social factors (isolation, alienation, avoidance behaviors) are operative. Magnetic resonance imaging (MRI) in patients with DP is suggestive of disturbed functioning in brain regions associated with the interpretation of perceptions, involving the putamen and the associated dorsal striato–thalamo–cortical loop (a circuit linked to somatic delusions and tactile misperceptions) along with frontal dysfunction (which is hypothesized to be responsible for disturbed reality testing). The therapeutic efficacy of D2-dopamine antagonists in DP implies a role in dopaminergic dysfunction in these regions. Patients with intoxication from substances influencing dopamine transmission (e.g., cocaine, amphetamines, etc.) may experience delusional infestation, further validating this hypothesis.

Patients with DP can be further subgrouped as:

1. Patients with *predominantly hypochondriacal traits* are convinced that they have an incurable physical illness. This has been described as a hypochondriacal delusion of parasitosis or simply DP. In some cases, those not suffering from organic psychosis, schizophrenia, or an affective disorder may be considered in the group of monosymptomatic hypochondriacal psychosis.
2. Patients with *paranoid delusions* and without hypochondriacal traits. These individuals tend to fight against parasites that are perceived as impairing their existence. This has been referred to as delusions of infestation. Such patients consult parasitologists, entomologists, and even pest control companies but seldom see dermatologists.
3. Patients with signs of both *hypochondriacal as well as paranoid delusions,* are referred to as hypochondriacal delusions of infestation. Such patients usually present for repeated dermatologic consultation, similar to those in the first group.

Treatment. Patients with DP are reluctant to seek psychiatric treatment. It is important to establish rapport and build a stable relationship based on trust. This is accomplished by listening to patients' stories, showing empathic understanding, asking how the condition affects their lives, and trying to reduce their sense of isolation. It is often counterproductive to inform patients in the first meeting that they are suffering from a delusional disorder. It is best to acknowledge complaints as an experienced reality. Treatment strategies include psychotropic medications as well as psychotherapy. With the implementation of such an integrative approach, improvement in delusional symptomatology can be expected in about two-thirds of patients.

Psychopharmacologic treatment of DP has typically included the use of pimozide, a potent neuroleptic. In one double-blind, crossover-controlled study of treatment of DP with pimozide, 10 of 11 patients improved after 6 weeks of treatment ranging from 1 to 5 mg per day. A follow-up study demonstrated that treatment with pimozide for 5 months resulted in long-term remission in many patients. Full remission of DP has been reported in 50% of those receiving pimozide. Other antipsychotic medications (risperidone, haloperidol, trifluoperazine, chlorpromazine) have also demonstrated efficacy in uncontrolled trials, and there have been no head-to-head studies to indicate that pimozide is superior to other antipsychotics. Unlike other antipsychotics, however, pimozide is also a potent opiate antagonist and has an antipruritic action which may contribute to the improvement seen in DP patients. As with other neuroleptics, pimozide carries a risk of prolongation of the QT interval, which may limit its use in patients with cardiac dysrhythmia. There are anecdotal case reports of treatment response from selective serotonin reuptake inhibitors (SSRIs), tricyclic antidepressants (TCAs), and electroconvulsive therapy (ECT). One case report of successful treatment with naloxone and naltrexone lends further weight to a pathophysiologic contribution from the endogenous opiate system.

Morgellons Disease

Morgellons disease (MD) is a pattern of dermatologic symptoms that may present in a manner very similar to DP. The condition was first described many centuries ago but has recently received much attention in the mass media. It has been postulated that patients exposed to information about MD on the internet may develop the strong belief that they too have the condition as a form of shared delusion. As such, many medical practitioners still consider MD a delusional disorder. More recent studies have, however, reported findings to suggest that MD may be a dermopathy associated with tick-borne illness. The diagnosis thus remains controversial.

Clinical Features. Patients with MD experience a crawling or stinging sensation on the skin and often also believe that they are infested with insects. Studies initially suggested that this diagnosis was a purely delusional disorder. More recent investigations, however, have demonstrated an association with spirochetal infection. MD more commonly affects females, with a mean age of 50 to 55 years and a prevalence of 6%. Psychiatric comorbidities (including depression, anxiety, bipolar disorder, ADHD, OCD, and schizophrenia) are noted in 68% to 82% of patients with MD.

Diagnosis. Diagnostic features include ulcerative cutaneous lesions with embedded or protruding microscopic filaments (colored, white, or black) that resemble textile fibers. Patients experience formication, described as biting, stinging, itching, or crawling sensations on the skin alongside systemic manifestations that mirror those of Lyme disease (LD); these include neuropathy, joint pain, cardiac complications, fatigue, and changes in cognition. A key step to diagnosis is to determine whether the patient presents with characteristics commonly associated with delusions —such as agitation, confusion, and mood or perceptual disturbances. PCR testing and confirmatory DNA sequencing have detected Borrelia DNA in specimens from MD patients, establishing a potential link between LD and MD, though other associated pathogens (such as *Helicobacter pylori*, *Treponema denticola*, and *Bartonella henselae*) have also been implicated in disease pathogenesis. It is plausible that the effects of microbial infection result in prolonged inflammatory processes in the brain, thereby causing neurodegenerative symptoms. While some MD patients exhibit neuropsychiatric symptoms, others do not.

Treatment. There are currently no universally accepted criteria for MD, and the potential physiologic underpinnings of the disease are contested, leaving treatment a challenge for providers. Trials have shown earlier treatment as associated with better outcomes for patients with MD. Intervention should be primarily directed at the underlying tickborne illness if one can be identified. Though there is some evidence suggesting that MD responds to antibiotic treatment, these findings have not been confirmed by controlled trials. Antipsychotic agents have been proposed for patients with neuropsychiatric symptoms. Pimozide, for instance, has been known to show improvement in symptoms such as formication associated with MD. Other studies suggest a combined regimen of antibiotics, antipruritic agents, and antipsychotic medications, though it is unclear which of the three resulted in improvement.

Obsessive-Compulsive and Related Disorders

Trichotillomania (Hair-Pulling Disorder). While François Hallopeau first coined the term trichotillomania (TTM) in 1889, from the Greek for *tricho* (hair), *tillo* (pull), and *mania* (fury), hair-pulling has been described since antiquity. Hippocrates advised physicians to assess hair-pulling as part of a routine mental status examination and described a case of hair-pulling in the apparent context of depression. The phenomenon of a *trichobezoar* (gastric hair ball) was first described in the medical literature in the 18th century

CLINICAL FEATURES. Trichotillomania (hair-pulling disorder) is classified in DSM-5 under obsessive-compulsive and related disorders (OCRD). The DSM-5 diagnostic criteria are outlined in Table 27.12–3.

Clinically, patients present with areas of differing hair lengths with blunt or tapered ends or areas of black dots on the surface of the scalp (Fig. 27.12–3). On exam, hair does not pull out easily (negative “pull test”). TTM affects women ten times more commonly than men.

SUBTYPES. Hair-pulling most typically involves the scalp but may also be seen in the eyebrows and eyelashes. Two types of hair-pulling have been delineated. *Focused pulling* describes the use of an intentional act to control unpleasant personal experiences, such as an urge, bodily sensation (e.g., itching or burning), or thought. This behavior is thought to be a compulsion in that the patient resists the act of pulling and feels relieved after pulling. In contrast, *automatic pulling* occurs outside the person’s awareness and most often during sedentary activities and has been formulated as more akin to an impulse control disorder. Many patients have a combination of these types of hair-pulling. Lack of awareness of hair-pulling has been postulated to be associated with possible “dissociation and emotional dysregulation” in patients with TTM.

Table 27.12–3.
The DSM-5 Diagnostic Criteria for Trichotillomania (Hair-Pulling Disorder)

A. Recurrent pulling of one’s hair resulting in hair loss.
B. Repeated attempts to decrease or stop hair pulling.
C. The hair pulling causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The hair pulling or hair loss is not attributable to another medical condition (e.g., a dermatological condition).
E. The hair pulling is not better explained by the symptoms of another mental disorder (e.g., attempts to improve a perceived defect or flaw in appearance due to body dysmorphic disorder).

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, p. 281 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

PATHOPHYSIOLOGY. Patients with TTM may have comorbid mood, anxiety, and eating disorders. Major depression was the most common mood disorder, and generalized anxiety disorder was the most common anxiety disorder in a study of a large series of patients with TTM. Comorbid cluster B (histrionic 26%, borderline 18%) and cluster C (obsessive-compulsive 27%, avoidant 14%) personality disorders are also commonly seen in patients with TTM. Familial association between TTM and OCD has been found in some studies, and patients with TTM will often report a family history of OCD or other anxiety disorders. OCD is thought to be related to the overactivity of cortico–striatal–thalamic–cortical (CSTC) circuits, which is also hypothesized to play a role in a range of dysfunctional repetitive behaviors, including TTM. One MRI study reported decreased putamen volume (with normal caudate volume) in patients with TTM compared to healthy controls, a finding suggestive of CSTC circuit involvement. A PET study of patients with TTM showed that treatment response to clomipramine was inversely correlated with anterior cingulate and orbitofrontal metabolism. In another study of patients with TTM, high baseline CSF 5-hydroxyindoleacetic acid (5-HIAA) levels predicted a better response to SSRI treatment, consistent with similar studies in OCD.

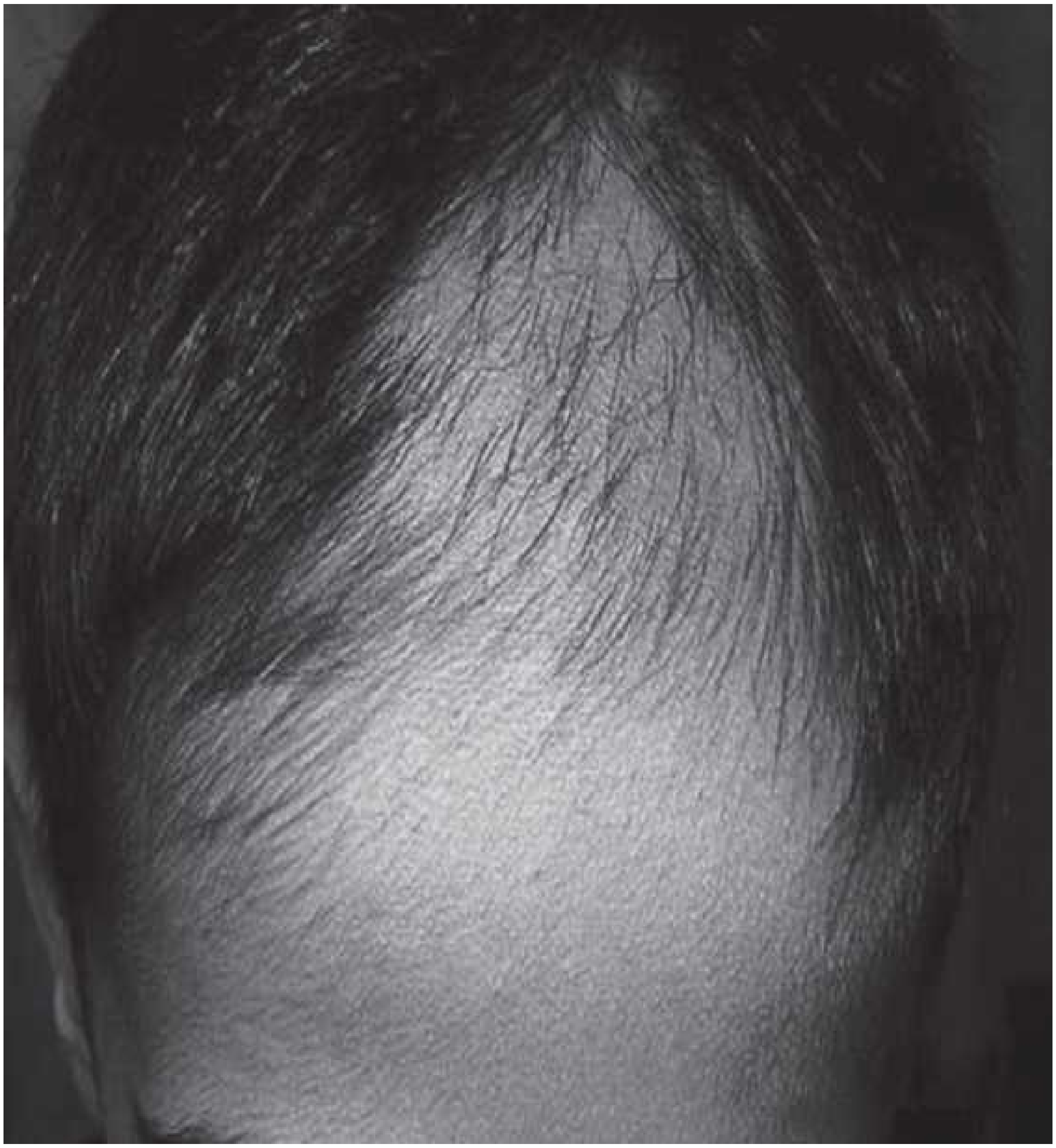

FIGURE 27.12–3. Picture of patient with Trichotillomania with alopecia of frontal scalp. Areas of different hair lengths with blunt or tapered ends, areas of black dots on the surface of the scalp. The hair does not pull out easily (negative “pull test”).

TREATMENT. The primary intervention is often heavily dependent on the age of onset of symptoms. In preschool children, where TTM is more often of the automatic type, parental guidance and increased awareness are most effective. Preadolescents and young adults are often not aware of their hair-pulling and require the establishment of a diagnosis and reassurance that their hair is normal and will regrow when it is left alone. Behavioral modification treatments are recommended if the condition becomes chronic. Adult-onset TTM is usually associated with psychiatric disorders and requires psychopharmacologic and behavioral interventions. In a double-blind crossover study, clomipramine was found to be more effective than desipramine for the treatment of TTM. More recent open trials of SSRIs (fluoxetine, fluvoxamine, paroxetine, sertraline, citalopram) and venlafaxine for the treatment of TTM suggest overall positive treatment response with both short-term and long-term benefits. Given their superior tolerability compared to TCAs, the SSRIs arguably remain the first-line choice medication treatment for TTM. Various case reports have also documented the potential efficacy of atypical antipsychotics (olanzapine, risperidone, and aripiprazole) in the treatment of TTM either as monotherapy or as augmentation of SSRIs. Mood stabilizers (lithium, oxcarbazepine, topiramate, and lamotrigine) have also been used in TTM patients with mood disorders.

Some studies have examined the effects of naltrexone on hair-pulling. Blocking opioid binding may decrease the positive reinforcement from hair-pulling. An open-label study of 14 children with TTM found naltrexone reduced hair-pulling; another study involving adults reported naltrexone had some improvement in hair-pulling behavior. Another pharmacologic approach to TTM involves the use of the glutamate modulator *N*-acetylcysteine (NAC). In a 12-week double-blind placebo-controlled trial, NAC was used in a dosage of 1,200 to 2,400 mg per day, with 56% of TTM patients showing significant improvement. Other studies using NAC reported treatment outcomes comparable to those seen with behavioral interventions like cognitive-behavioral therapy (CBT).

In terms of psychotherapeutic approaches, CBT which includes “habit reversal training (HRT)” is considered the first-line treatment option. *HRT* has three core elements: first, “awareness training” (self-monitoring), to become aware of urge that precedes pulling; second, “stimulus control,” to reduce the likelihood that pulling behavior begins; and third, “competing response training,” to engage in a behavior, at the earliest sign of pulling, that is physically incompatible with hair-pulling. There is evidence to suggest that combined use of behavioral therapies and medications may be more effective than either treatment alone.

Psychogenic Excoriation: Skin Picking Disorder. The term neurotic excoriation was introduced by Erasmus Wilson in 1875. In 1889, M. L. Brocq used the term *acne excoriée des jeunes filles* to describe the excoriations from self-inflicted lesions in young “neurotic” women with pre-existing acne. A variety of synonyms have been used to describe skin-picking syndrome, including pathologic skin-picking, neurotic excoriation, dermatillomania, or psychogenic excoriation. DSM-5 uses the term *excoriation (skin picking) disorder* (SPD). The diagnostic criteria for SPD are listed in Table 27.12–4.

Table 27.12–4.
DSM-5 Diagnostic Criteria for Excoriation (Skin-Picking) Disorder

A. Recurrent skin picking resulting in skin lesions.
B. Repeated attempts to decrease or stop skin picking.
C. The skin picking causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The skin picking is not attributable to the physiological effects of a substance (e.g., cocaine) or another medical condition (e.g., scabies).
E. The skin picking is not better explained by symptoms of another mental disorder (e.g., delusions or tactile hallucinations in a psychotic disorder, attempts to improve a perceived defect or flaw in appearance in body dysmorphic disorder, stereotypies in stereotypic movement disorder, or intention to harm oneself in nonsuicidal self-injury).

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, p. 284 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

CLINICAL FEATURES. SPD is a common skin disorder characterized by excessive picking, scratching, digging, rubbing, gouging, or squeezing normal skin or previously diseased skin (typically juvenile acne). Patients experience urges to excoriate their normal-appearing skin or visible pre-existing abnormalities like acne, papules, scabs, scars, or insect bites. The classic lesions of psychogenic excoriation are characterized by clean, linear erosions, scabs, and scars that can be hypopigmented or hyperpigmented. The distribution of excoriations is mostly confined to accessible areas like face, arms (mostly extensor surface), and anterior thigh, with sparing of middle and upper lateral back (resulting in the "butterfly sign" of preserved normal skin in areas the patient is unable to reach). This distribution of skin lesions can give the clinician a valuable clue, but it is not diagnostic of psychogenic excoriation because any condition that has pruritus as a primary symptom (such as hepatic, uremic, or idiopathic pruritus) can result in a similar distribution of skin lesions. Once the patient begins the process of self-picking, a focal itch sensation may initiate scratching, or the tissue damage caused by scratching may trigger further itching. The itch-scratch cycle may become part of the clinical picture and exacerbate the severity and chronicity of psychogenic excoriation. *Lichen simplex chronicus* is a chronic pruritic lichenoid plaque with skin thickening provoked and maintained by scratching or rubbing. Before arriving at a diagnosis of psychogenic excoriation, other conditions that can cause pruritus or excoriation must be ruled out. In the general population, the lifetime prevalence of SPD in adults is 1.4% or higher. Three-quarters or more of those diagnosed are female. Psychogenic excoriation is reported in 9% of dermatology outpatients presenting with pruritus. Most studies report a mean onset of age between 30 and 45 years, but acne excoriée can begin in adolescence and continue into adulthood.

PSYCHIATRIC COMORBIDITY. Some individuals engage in skin picking that is more focused, that is, with preceding tension and subsequent relief, consistent with obsessive-compulsive symptoms. Some individuals may engage in more automatic picking, occurring without full awareness and without preceding tension. Although the DSM-5 does not specifically classify such patients as dissociative, it has been postulated that patients engaging in skin picking without preceding tension or full awareness are likely to have higher levels of emotional dysregulation and dissociation without necessarily meeting the criteria for a dissociative disorder. Recognition of the dissociative component may thus be of clinical importance, as the standard approaches to treating obsessive-compulsive skin picking (i.e., habit reversal therapy and use of an SSRI) are not sufficient in those with high levels of dissociation.

A lifetime occurrence of other primary psychiatric disorders has been found in 54.5% to 100% of individuals with SPD, with OCD, TTM, generalized anxiety disorder, and major depressive disorder among the most common. Generalized anxiety disorder, panic disorder, social phobia, and posttraumatic disorder are present in 41% to 65% of patients. Mood disorders, including major depressive disorder, dysthymia, and bipolar disorder, are present in 48% to 68% of these patients. OCRD disorders that commonly co-occur with SPD include body dysmorphic disorder (BDD), compulsive nail-biting, and eating disorders. In one study, 45% of patients with BDD also met diagnostic criteria for SPD. Patients with SPD may have obsessions about contamination and skin abnormalities or may be preoccupied with having smooth skin, flawless complexion, and cleanliness. This obsessiveness and potential resultant compulsions might be severe enough to meet the criteria for OCD. OCRD due to another medical condition refers to "obsessive-compulsive symptoms best explained by the direct pathophysiologic consequence of another medical condition." Excoriation may occur comorbid with another dermatologic disorder. Whether SPD should be diagnosed in addition depends on the clinical assessment of to what point the skin picking is independent of other disorder.

TREATMENT. There are few controlled data on the pharmacologic treatment of psychogenic excoriation. One randomized controlled trial found that NAC 1,200 to 3,000 mg/day significantly reduced skin-picking symptoms. This suggests that the glutamatergic system may prove a useful target in treating SPD. Most of the available data comes from case studies, open trials, and small double-blind studies, which largely support treatment with SSRIs. Fluoxetine is the most studied SSRI in the treatment of psychogenic excoriation. In one double-blind, 10-week study, at a mean dose of 55 mg per day, fluoxetine was superior to placebo in reducing psychogenic excoriation. Other open trials and case studies report benefit from treatment with other SSRIs, such as sertraline and fluvoxamine, and the TCAs doxepin and clomipramine. SSRIs may be particularly effective in patients with prominent compulsive features, comorbid mood disorders, and anxiety disorders. Naltrexone at a dose of 50 mg per day has shown efficacy in reducing the psychogenic excoriation associated with pruritus. However, many patients with excoriation experienced increased pain with naltrexone treatment-limiting its value. Other case reports suggest neuroleptics such as pimozide, aripiprazole, and olanzapine may be of use. Patients with predominantly impulsive features or co-occurring bipolar disorder may benefit from anticonvulsants such as lamotrigine or divalproex or from atypical neuroleptics such as olanzapine. In one study, lamotrigine was associated with improvements in two-thirds of subjects with pathologic skin picking. A systematic review and meta-analysis of psychiatric treatments for SPD found a strong effect size for SSRIs, lamotrigine, and behavioral therapy. CBT and HRT have shown benefits in this cohort.

Body Dysmorphic Disorder. Prior to DSM-5, BDD had been categorized as a somatoform disorder but is now grouped under OCRD because of evidence of a relationship of BDD to obsessive-compulsive spectrum disorders.

Table 27.12–5.
DSM-5 Diagnostic Criteria for Body Dysmorphic Disorder

A. Preoccupation with one or more perceived defects or flaws in physical appearance that are not observable or appear only slight to others.

B. At some point during the course of the disorder, the individual has performed “repetitive behaviors (e.g., mirror checking, excessive grooming, skin picking, or reassurance seeking) or mental acts (e.g., comparing his or her appearance with that of others) in response to the appearance concerns.” (Gupta and Gupta, 2014, p. 453)

C. The preoccupation causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

D. The appearance preoccupation is not better explained by concerns with body fat or weight in an individual whose symptoms meet diagnostic criteria for an eating disorder.

Specify if:

With muscle dysmorphia: The individual is preoccupied with the idea that his or her body build is too small or insufficiently muscular. This specifier is used even if the individual is preoccupied with other body areas, which is often the case.

Specify if:

Indicate degree of insight regarding body dysmorphic disorder beliefs (e.g., “I look ugly” or “I look deformed”).

With good or fair insight: The individual recognizes that the body dysmorphic disorder beliefs are definitely or probably not true or that they may or may not be true.

With poor insight: The individual thinks that the body dysmorphic disorder beliefs are probably true.

With absent insight/delusional beliefs: The individual is completely convinced that the body dysmorphic disorder beliefs are true.

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 271, 272 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

DIAGNOSIS. Table 27.12–5 lists the DSM-5 criteria for BDD. DSM-5-TR, allows for a further specifier indicating the degree of insight regarding

BDD: “with good or fair insight,” “with poor insight,” and “with absent insight/delusional beliefs.”

CLINICAL FEATURES. When BDD involves concern about skin, patients typically present with imagined or slight flaws in one or more body areas. Typically these include the skin, hair, or nose. In the case of the skin, concerns can range from perceived acne or scars, to winkles or other skin lines or a pale complexion. BDD usually manifests during adolescence and has equal sex distribution. Patients with BDD engage in various time-consuming activities to improve the perceived defect, such as camouflaging, mirror checking, comparing with others, questioning, and reassurance seeking. Some patients go through several cosmetic and surgical procedures for their perceived defects. The functional impairment and distress expressed by the patients with BDD include severe depression and low self-esteem, anxiety, social isolation, and work-related problems. BDD is associated with a high rate of suicidal ideation and suicide attempts, especially among adolescents with severe acne. The shame and humiliation associated with BDD play a major role in the underreporting of this disorder.

A largely unexplored phenomenon of disturbed CBI that may represent a specific subtype of BDD is the longstanding practice of skin-bleaching. The desire for fair skin is not uncommon in men and women of Asian and African descent (especially in Fitzpatrick skin types IV-VI) who may be influenced by a society flooded with billboards of predominantly “light-skinned” Bollywood stars and widespread use of skin-lightening creams and cosmetics. Children with darker skin are subject to ridicule—contributing to a disturbed sense of self due to an unrealistic, Eurocentric beauty standard. With the use of skin-lightening practices—from bleaching creams to off-label IV glutathione treatments—this focus on skin tone perpetuates both dangerous skincare practice and mental illness. Beyond the physical consequences associated with unregulated skin-bleaching agents, there is a need to address the cultural, psychiatric, and ethical considerations of skin-lightening practices. As part of a multidisciplinary approach to care, dermatologists should continue to assess CBI as an important component of patient history while screening for potential comorbid psychiatric diagnoses in patients concerned with skin tone. Studies have reported an increased incidence of BDD and major depressive disorder with seasonal pattern in individuals who engage in frequent tanning practices.

TREATMENT. The prevalence of BDD has been reported as 9% to 15% among dermatology patients and 7% to 8% among cosmetic surgery patients in the United States. Studies suggest that more than 60% of individuals with DSM-defined BDD seek nonpsychiatric medical/surgical treatment, most often resulting in no change in their preoccupation with their perceived flaws. This has important implications in dermatology, as clinical treatment guidelines for dermatologic disorders increasingly emphasize the importance of assessing psychosocial comorbidity, including aspects of CBI (cutaneous body image), in order to achieve optimal treatment outcomes. Convincing patients with BDD to engage in psychiatric treatment can be challenging, and an empathic understanding and establishment of good therapeutic rapport on the part of the clinician are crucial. The efficacy of CBT with a focus on cognitive restructuring has been well-demonstrated. SSRIs are the drug of choice in the treatment of BDD, though studies suggest that individuals with BDD require longer trial durations (mean response time of 6 to 9 weeks) with relatively higher doses of SSRIs. A large study of TCAs has shown a significantly better efficacy for clomipramine when compared to desipramine.

Unspecified OCRDs: Body-Focused Repetitive Behaviors

CLINICAL FEATURES AND EPIDEMIOLOGY. Onychophagia (or chronic nail-biting) is a disorder affecting 20% to 30% of the general population and is common among children, beginning as early as 4 years of age and peaking during the teenage years. Females are more commonly affected than males. A progressive decline in nail-biting after the age of 40 has been described. Patients with onychophagia have clinical features of repetition, resistance, and relief, comparable to hair-pulling and skin picking. Patients with severe and chronic onychophagia can experience significant dental and medical problems, such as hand infections and mandibular deformities. Co-occurring thumb sucking is also frequent. Onychophagia behavior is usually increased during periods of conflict or stress.

DIAGNOSIS. The DSM-5 provides limited diagnostic criteria for unspecified OCRDs, describing instead presentations of symptoms characteristic of an OCD but not meeting full criteria for a more specific diagnosis. Thus, diagnostic features might be considered to include repetitive behaviors that are refractory to attempts to stop or cut down and produce some degree of functional impairment.

TREATMENT. Behavioral treatments, like habit reversal interventions, have demonstrated the greatest efficacy in treating this disorder. Other trials have seen some success with mild aversive techniques (such as the use of distasteful lacquer) and utilizing a nonremovable reminder (such as a wristband), though significant improvement was only seen when performed in conjunction with self-recording and physician visits.

Factitious Disorders

Dermatitis Artefacta. Dermatitis artefacta (DA) or factitious dermatitis is a disorder in which skin is the target of self-inflicted injury, and the patient uses more elaborate methods than simple excoriation to self-induce skin lesions.

CLINICAL FEATURES. DA can present as an aggravation of dermatosis, with a variety of skin lesions, including blisters, ulcers, erythema, edema, purpura, and sinuses. The morphology of DA lesions is often bizarre and linear, with clear-cut, angulated, or geometric edges. The presence of completely normal, unaffected skin adjacent to horrific-looking lesions is a valuable clue in the diagnosis of factitious dermatitis. Skin damage can be extensive, with full-thickness skin loss and severe scarring requiring plastic surgery (Fig. 27.12–4). The patient's description of the history of the skin lesions is usually vague, lacking detail about their appearance and evolution over time. The natural progression of lesions through different stages of development (as one might expect in a typical dermatosis) is typically absent. DA lesions are usually located in areas easily reached by the dominant hand. Patients with DA are seemingly unmoved by disfiguring and painful lesions and typically deny any role in their causation. This condition has been reported in 3% of dermatology patients with a female-to-male ratio as high as 8 to 1. It can present at any age but occurs most frequently in adolescents and young adults. Those with DA often have a history of other psychosomatic illnesses, have had multiple consultations with physicians, and have trialed multiple "ineffective" treatments. The course of this disorder varies from a brief episode to a lifelong behavior; one study reported that 30% of factitious dermatitis patients continued to present with lesions over 12 years or more.

PSYCHIATRIC COMORBIDITY. DA has been associated with several psychiatric disorders (besides a primary factitious etiology), with borderline and histrionic disorders being the most commonly reported. Other reported co-occurring conditions include schizoaffective disorder, adjustment disorder, major depressive disorders, PTSD, DP, intellectual disability, pervasive developmental disorders, and use of alcohol, cocaine, or lysergic acid diethylamide (LSD). Patients with factious disorders may have a history of contact with multiple health care services and carry a volume of medical notes. These patients are often aware of their impulsive acts in the context of emotional and relational instability, which may resemble symptoms of obsessive-compulsive spectrum disorders. Factitious dermatitis may have a connection with body image disturbance as a sequela of BDD. A patient with severe BDD focused on the skin may pursue requests for incisive dermatologic treatments or cosmetic surgery and, when denied access to such treatments, may take the initiative to “self-treat,” resulting in now “real” pathology.

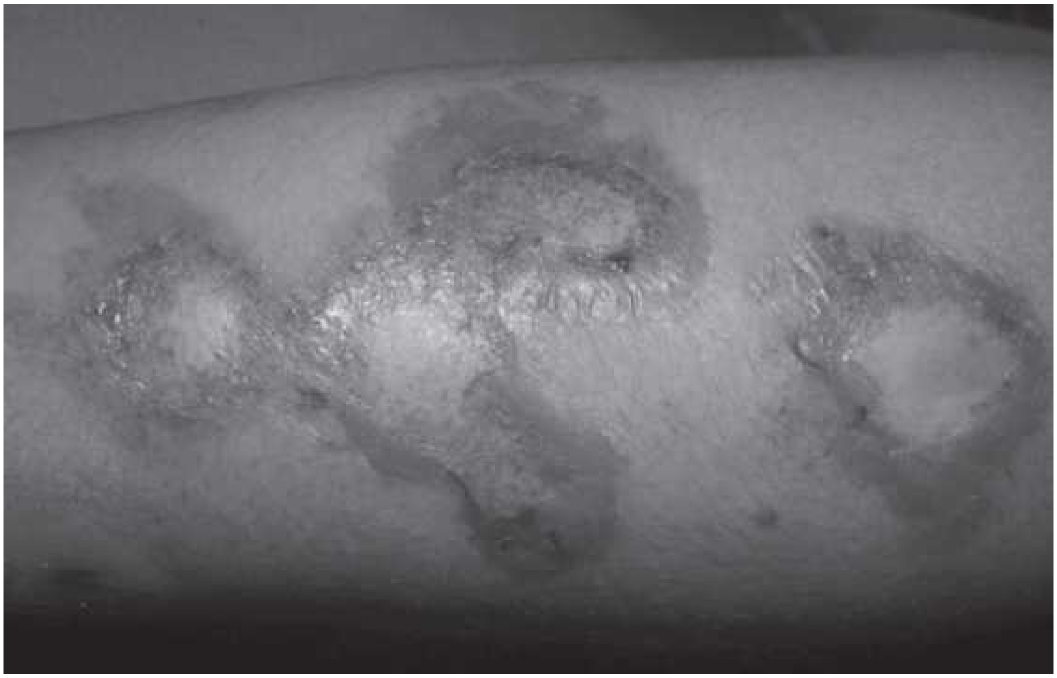

FIGURE 27.12–4. Picture of a patient with dermatitis artefacta: self-inflicted sharp, rounded lesions on the forearm made with a razor during times of stress and emotional dysregulation.

TREATMENT. There are no controlled trials for the treatment of DA. Patients commonly refuse referral to psychiatry, and direct confrontation may jeopardize the doctor–patient relationship. An effort to develop an empathic therapeutic relationship between the dermatologist and the patient is critical to the management of these patients. Once such a therapeutic relationship is established, some patients may accept appropriate therapeutic approaches and psychiatric referrals.

Psychogenic Purpura

Psychogenic purpura (Gardner–Diamond syndrome), also known as autoerythrocyte sensitization syndrome, is an extremely rare condition typically noted in women with other psychiatric diagnoses.

Clinical Features. Psychogenic purpura is regarded primarily as an autoimmune vasculopathy with sensitization to phosphatidylserine, a component of erythrocyte stroma. Presentation typically involves spontaneous development of painful edematous skin lesions generally following severe stress or emotional trauma. These predominantly occur in the extremities, may be isolated or multiple, and progress to ecchymoses over the next 24 hours. Associated initial symptoms include burning and pain followed by erythema, swelling, and pruritus. Medical work-up, including evaluation of hemostatic and blood coagulation function, is typically normal. The age of onset is variable, and women are affected more frequently.

Pathophysiology. A variety of mechanisms of psychogenic purpura have been proposed, including autoerythrocyte sensitization, conversion reaction, and factitious disorder. The evidence for autoerythrocyte sensitization is very limited. The psychogenic cause of purpura has been supported by some studies because of the association between psychosocial stress and lesion occurrence. Most patients also have multiple other unexplained somatic symptoms and significant psychiatric comorbidity–with generalized anxiety disorder, major depressive disorder, and borderline personality disorder among the most common diagnoses.

Treatment. There is no specific treatment for psychogenic purpura, though symptomatic treatment with antihistamines, corticosteroids, antidepressants, or hormones may be provided for severe general symptoms. Psychotherapy has shown some benefit in reducing stress-associated lesions.

Unspecified Somatic Symptom and Related Disorder

Pruritus and Chronic Idiopathic Pruritus (also known as Chronic Pruritus of Unknown Origin). Pruritus is a common unpleasant sensation that provokes a desire to scratch. It is a common observation that "itch" can be induced simply by thinking about it. Many patients with chronic pruritus experience scratching as an automatic response to the sensation of itch. Scratching provides immediate relief of discomfort, perhaps even serving to reduce inner tension, and this negative reinforcement leads to conditioned scratching. This culminates in a self-perpetuating cycle of itch-and-scratch, which may prompt a sense of loss of control and feelings of helplessness often accompanied by despondency and guilt.

PATHOPHYSIOLOGY. The epidermis is heavily innervated by axonal nerve endings with nociceptive (pain-sensing) unmyelinated C-fibers (Fig. 27.12–5). The sensation of an itch is mediated by specific nociceptive unmyelinated C-fibers modulated by itch-specific C-receptors. D. Andrew described a separate itch pathway by identifying a "second-order histamine-specific itch neuron" in the dorsal horn of the spinal cord implicated in neuropathic itch syndromes. Signal from the dorsal horn is carried by spinothalamic tract projecting to the thalamus and on to the anterior cingulate cortex, and the insula. The cingulate cortex is an important area in the processing of itch and became significantly activated in patients with AD after administration of histamine. Moreover, the degree of its activation correlated with disease severity and deactivation occurred after scratching in healthy controls. The insula is the seat of *interoceptive representation* and underpins the basis of subjective feelings from all sensory input from the body. Activation of the insular cortex has been reported with skin sensations including pain, itch, touch, temperature, and prickliness. The anterior cortex is known to be involved in the modulation of emotional and cognitive activity (such as reward anticipation). This may provide the neurophysiologic basis for why individuals with difficulties in emotional regulation and interoceptive awareness can engage in body-based repetitive behaviors and thus how emotional factors can affect the perception and processing of itch.

A number of chemicals can provoke or enhance pruritus, including histamine, prostaglandins, proteinases, substance P (SP), neuropeptides (NPs), cytokines, and bile salts. Some of these substances act directly on the free nerve endings, while others act through their action on mast or other cells. Among prostaglandins, PGE-2 does not itself cause itch but lowers the threshold and potentiates the itch provoked by histamine. Endogenous opioids, such as endorphin, met-enkephalin, and leu-enkephalin, can play a role in causing itch by either central or peripheral action. One PET imaging study demonstrated that intracutaneous injection of histamine activated the premotor area and inferior parietal lobule–regions involved in planned movement. One large population study demonstrated a significant correlation between mental stress and itch. Psychological factors such as depression can modulate histamine-induced pruritus by lowering the itch threshold, and the severity of pruritus has also been correlated with the degree of depressive symptoms. Studies further suggest that the presence of marked circulating blood or tissue eosinophilia among patients specifically diagnosed with chronic pruritus of unknown origin should determine treatment choice between immunomodulation or gabapentinoids.

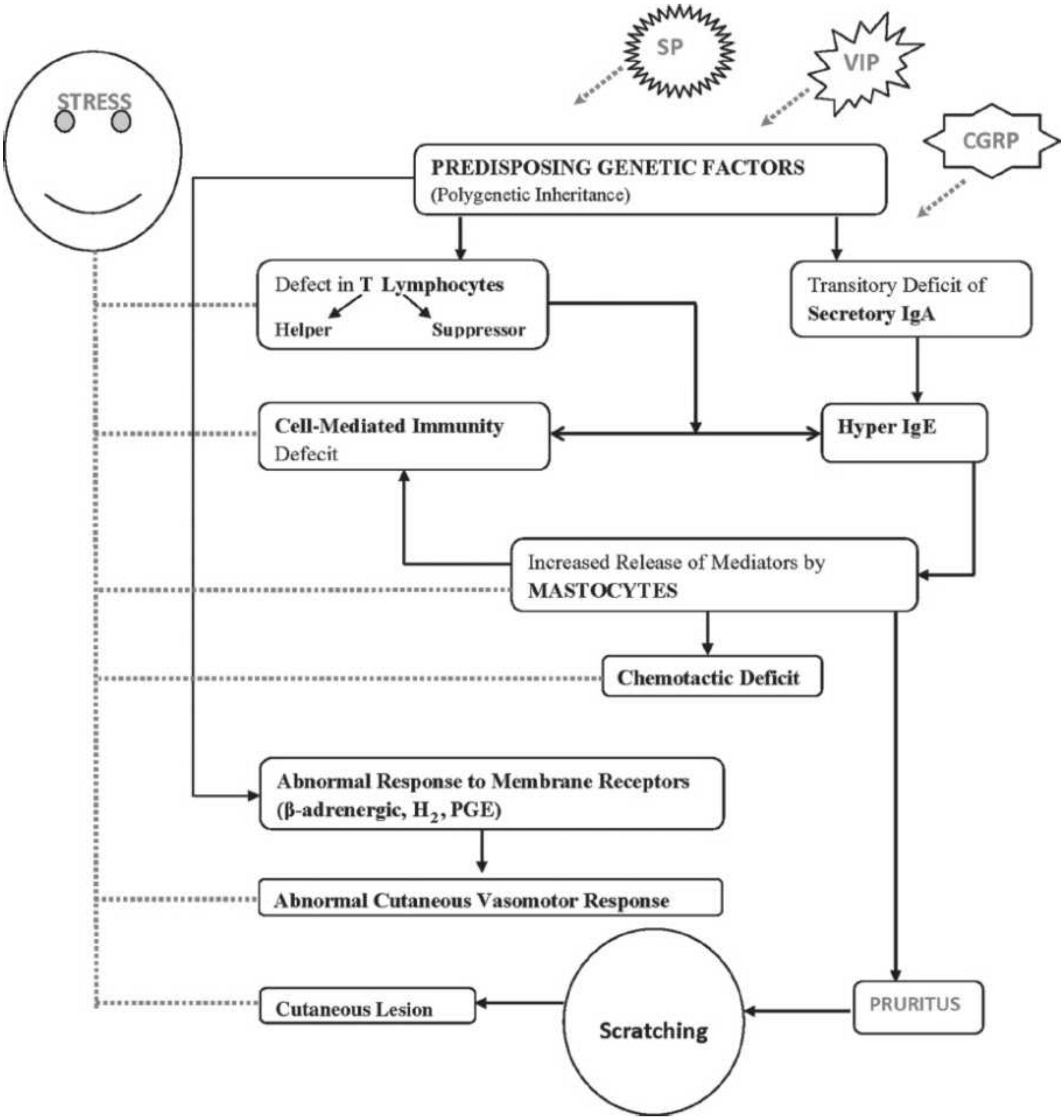
SP
VIP
CGRP
STRESS
PREDISPOSING GENETIC FACTORS
(Polygenetic Inheritance)
Defect in T Lymphocytes
Helper
Suppressor
Transitory Deficit of
Secretory IgA
Cell-Mediated Immunity
Defecit
Hyper IgE
Increased Release of Mediators by
MASTOCYTES
Chemotactic Deficit
Abnormal Response to Membrane Receptors
(β-adrenergic, H2, PGE)
Abnormal Cutaneous Vasomotor Response
Cutaneous Lesion
Scratching
PRURITUS

FIGURE 27.12–5. Hypothetical outline of the pathogenesis of atopic dermatitis (AD) correlated with the possible influence of emotional stress (*dotted lines*). SP, substance P; VIP, vasoactive intestinal peptide; CGRP, calcitonin gene-related peptide; Ig, immunoglobulin; H2, type 2 receptors for histamine; PGE, prostaglandin E. (Adapted from Panconesi E, Hautmann G. Psychophysiology of stress in dermatology. The psychobiologic pattern of psychosomatics. *Dermatol Clin.* 1996;14(3):399–421. doi:10.1016/S0733-8635(05)70368-5.)

Chronic idiopathic pruritus has been linked to a wide range of etiologies, including cutaneous, systemic, neurologic, and psychiatric causes. Evaluation of a patient with pruritus should always involve a careful examination and medical work-up to rule out any condition that may be treatable.

- *Medical conditions associated with pruritus* include chronic lymphocytic leukemia, polycythemia vera, malignant melanoma, syphilis, congestive heart failure, diabetes mellitus, vitamin B_{12} deficiency, pellagra, myxedema, hyperparathyroidism, chronic renal failure, hepatic disease, acquired immunodeficiency virus (AIDS), drug toxicity, and old age.
- *Associated neurologic conditions with pruritus* include dementia, multiple sclerosis, Parkinson disease, Huntington disease, neurofibromatosis, aneurysm, cerebral infarctions, and CNS tumors.
- *Psychogenic pruritus* is often associated with anxiety disorders, OCD, major depression, or psychosis. Chronic idiopathic pruritus, idiopathic pruritus ani (itching of the anal area), pruritus vulvae (itching of the vaginal area), and pruritus scroti (itching of the scrotum) may be the result of a central mechanism. Itch in patients with psychiatric diseases is not uncommon. Idiopathic itch was found to affect 36% to 42% of psychiatric inpatients.

DIAGNOSIS. Several neurologic causes of pruritus may occur in the absence of primary skin lesions and should be considered in the evaluation of potential psychogenic pruritus. Psychogenic pruritus is characterized by: (1) temporal association with psychiatric symptoms; (2) absence of awakening from sleep or insomnia due to pruritus; and (3) paroxysmal nature with sudden onset and resolution of increased severity and intervening symptom-free periods. Both psychogenic and organic pruritus may coexist, resulting in a mixed clinical presentation.

TREATMENT. Doxepin has been used effectively in chronic idiopathic pruritus. Amitriptyline may work well in itching associated with pain. Case reports have indicated some efficacy for the SSRIs and mirtazapine in the treatment of psychogenic pruritus. Gabapentin has been shown to be effective for PN, neuropathic itch, and uremic pruritus. Aprepitant, a neurokinin receptor (NKR)-1 (a receptor for SP—an important itch mediator) antagonist that has been used as an antiemetic, has been reported to be effective for chronic refractory pruritus. CBT and HBT can prevent the long-term complications of itching (such as lichen simplex chronicus) by interrupting the itch-scratch cycle. Some reports document the successful use of naltrexone as an antipruritic agent and demonstrated success with intranasal butorphanol (an opioid receptor agonist-antagonist) as a rescue therapy for chronic intractable itch.

Somatic Symptom Disorder, with Predominant Pain

Glossodynia. Patients with *glossodynia* present with chronic pain or burning affecting the tongue (or other areas inside the oral cavity). Other associated sensory changes may include paresthesia, changes in taste and smell, and mouth dryness. Glossodynia typically affects women, is most frequently reported in the sixth decade of life, and is observed in over 5% of patients seen by dentists. The condition is usually idiopathic, but glossodynia may be caused by vitamin B deficiencies, diabetes mellitus, candida glossitis, hormonal changes around menopause, salivary disturbances, and problems related to dentures or dental fillings. While higher rates of anxiety and depression are reported in patients with glossodynia, any causal relationship remains undefined. The duration of glossodynia is variable and may last for years.

Vulvodynia. The term vulvodynia is defined as “chronic vulvar discomfort with abnormal sensations” in the absence of skin disease or infection. A prevalence of 15% among patients of gynecologic outpatient practices has been reported. Most women with this disorder are between 20 and 50 years of age and have had pain for several years by the time they present for treatment. Women with vulvodynia are usually sensitive to touch or light pressure in the vestibule of the vagina, have pelvic floor musculature with altered contractile characteristics, and may also have heightened pain sensitivity in their arms or legs. There is growing consensus that vulvodynia is a chronic pain disorder. Vulvodynia patients experience sexual dysfunction as a result of their primary symptoms. The associated symptoms of allodynia (pain elicited by a nonpainful stimulus) and *hyperdynia* (when a stimulus causes greater pain than expected) suggest a neuropathic cause. The term *dysesthetic vulvodynia* is used to describe increased pain perception in the vulvar area without a known cause and may be of central or nerve root origin. This condition often occurs in perimenopausal or postmenopausal women, and they have fewer symptoms of dyspareunia and less point tenderness on examination than the “classic” vulvodynia cohort.

DIAGNOSIS. While somatization appears to be an important factor in cutaneous sensory syndrome patients and higher rates of anxiety and depression are found in this population, cutaneous sensory disorders are not specified in DSM-5. A comprehensive medical, neurologic, and psychiatric work-up should be performed to exclude any organic pathology.

TREATMENT. Once the diagnosis is made, the physician should acknowledge and validate the patient's experiences and symptoms. Pharmacologic interventions target the sensations of pain both peripherally and centrally. TCAs (amitriptyline, desipramine), SSRIs, gabapentin, and pregabalin have been effective in therapeutic trials. There are several case reports and uncontrolled trials suggesting the efficacy of the SNRIs venlafaxine and duloxetine. Augmentation with antipsychotics may be used in refractory cases. Controlled studies of CBT and biofeedback training to reduce pelvic floor spasms have indicated that these treatments are effective in symptom reduction.

PRIMARY DERMATOLOGIC DISORDERS ASSOCIATED WITH PSYCHIATRIC SYMPTOMS

The following diagnoses encompass primary dermatologic conditions with secondary psychiatric disorders.

Alopecia Areata

AA is a common disease involving localized loss of hair in round or oval areas, without visible inflammation of the skin (Fig. 27.12–6).

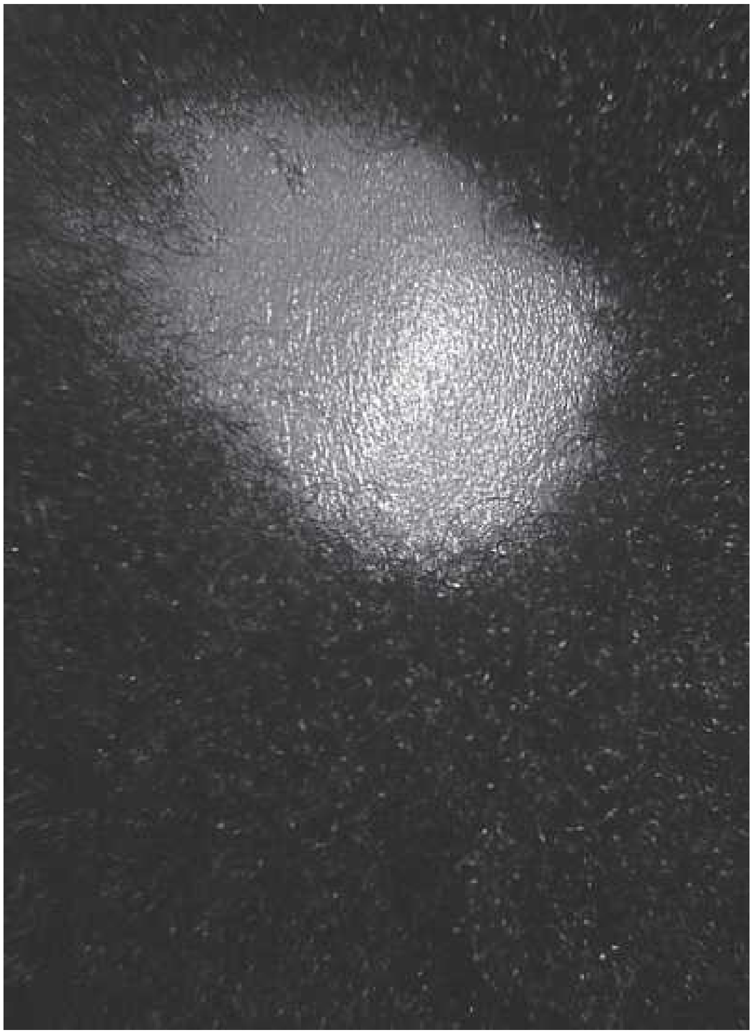

FIGURE 27.12–6. Alopecia areata of scalp. Sharp outline, without scaling, erythema, atrophy, or exclamation point hair. (From Fleisher GR, Ludwig S, Baskin MN. *Atlas of Pediatric Emergency Medicine*. Lippincott Williams & Wilkins; 2004.)

Clinical Features. Breakage of the hair shaft results in characteristic "exclamation-mark" hairs. Any hair-bearing surface may be affected, but AA most frequently occurs on the scalp (70% to 90%), eyebrows (4%), eyelashes, and beard (28% in males). About 1% of the U.S. population has at least one episode of AA by age 50, with 60% of cases presenting before 20 years of age. AA accounts for 2% of new dermatology outpatient visits in the United States. Incidence is equal across the sexes. AA is a partial loss of hair; alopecia totalis is a total loss of scalp hair, and alopecia universalis is a total loss of hair on the scalp and body. The course of AA is variable; 30% of cases end in complete recovery while 20% to 30% never recover. Poor prognosis is associated with prepubertal onset, atopy, rapid evolution of hair loss, nail dystrophy, and extensive hair loss.

Pathophysiology. The causes of AA remain unknown. Some evidence suggests that AA is a T-cell–mediated autoimmune condition that is most likely to occur in genetically predisposed individuals. There are associations between AA and autoimmune disorders such as Hashimoto thyroiditis (prevalence varies from 8% to 11.8%). The autoantibody response is heterogeneous and targets multiple structures of the anagen phase hair follicle. Whether these antibodies play a direct role in the pathogenesis or are epiphenomena is not yet known. There has been some investigation of how stress might precipitate AA. Acute stress leads to the release of corticotropin-releasing hormone (CRH) and activation of overexpressed type 2-β CRH receptors around the hair follicles, resulting in intense local inflammation. M. Toyoda and colleagues postulated a role for NPs in the pathogenesis of AA. SP is released by dermal nerve fibers around hair follicles, and their study revealed that SP and its degrading enzyme neural endopeptidase (NEP) were strongly expressed in affected hair follicles in the acute progressive as well as chronic stable phase of the disorder. It is possible that different pathogenic mechanisms may be involved in different subtypes of AA, leading to a variable prognosis.

Psychiatric Comorbidity. In a study of 294 patients with AA, the prevalence of major depression was 8.8%. Another study of 31 patients with AA found a 39% prevalence of major depression and a 79% lifetime prevalence for any psychiatric disorder. A study of 32 patients with AA reported a 66% prevalence of psychiatric disorders, including generalized anxiety disorder, adjustment disorder, and major depressive episodes. Patients with highly stress-reactive AA report more depressive symptoms than patients who deny an association of stress with their AA.

Treatment. The standard treatment of AA involves topical and intralesional steroid injections as well as systemic glucocorticoids. There are few studies involving psychiatric approaches to AA. In one randomized, double-blind, placebo-controlled trial, patients given imipramine at a dose of 75 mg per day had significantly more hair regrowth than those in the placebo arm. AA patients with comorbid major depression or anxiety disorders showed more improvement in AA symptoms when they were treated with SSRIs. Use of relaxation techniques, stress management, and psychotherapy has been recommended for AA, but there are no data from controlled studies to support their efficacy.

Prurigo Nodularis

PN is a chronic, pruritic skin disease marked by numerous hyperkeratotic nodules symmetrically distributed across the trunk and extensor surfaces of the extremities. The severity can range from presentations of a few to more than one hundred nodules that develop in the setting of chronic, refractory pruritus.

Clinical Features. Though nodules can erupt across any region of the body, patients present to the clinic with "the butterfly sign" with sparing of skin on the upper back. Most PN lesions are anywhere from several millimeters to 2 cm in diameter and the result of a cycle of chronic itch followed by repeated scratching. PN most commonly affects adults in middle age and is more frequently diagnosed in females. It is a relatively rare condition with a prevalence of 72 per 100,000 and is 3.4 times more likely to present in Black as compared to White patients. It is common with concomitant AD and has been associated with both HIV and hepatitis C infection. Of note, PN can be the presenting symptom of hematologic malignancies, in which case improvement or the resolution of lesions is seen after cancer treatment.

Pathophysiology. Though the exact pathogenesis of PN remains unknown, studies have demonstrated that immune cells and altered neural circuitry are both involved in the pruritus associated with PN. Perivascular infiltrates in the dermis of PN lesions are marked by an increase in T lymphocytes, mast cells, and eosinophilic granulocytes. Levels of both IL-31 and IL-4 are increased in PN skin lesions. Several NPs, specifically calcitonin gene-related peptide (CGRP) and SP, have been implicated in disease pathogenesis. CGRP causes neurogenic inflammation and dysregulated expression of mu and kappa opioid receptors, likely contributing to the pruritus associated with PN.

Psychiatric Comorbidity. PN is commonly associated with increased rates of depression and anxiety. Increased rates of eating disorders, self-harm, attention-deficit/hyperactivity disorder (ADHD), and schizophrenia have also been described in patients with PN. A national inpatient sample database study found longer inpatient stays and increased rates of psychiatric hospitalization in this patient population.

Treatment. Treatment of PN centers on breaking the itch-scratch cycle. Intralesional injections of triamcinolone acetonide (10 mg/mL) have been shown to improve PN lesions. High-potency topical corticosteroids can reduce pruritus and flatten nodules, while topical anesthetics such as 1% pramoxine lidocaine spray can provide modest itch relief. Medicated occlusive dressings also act as a barrier to excoriation. Other nonsteroid topical agents, including calcineurin inhibitors and topical capsaicin (0.025% to 0.3%), have demonstrated efficacy in the relief of itching and healing of nodules. Phototherapy, gabapentanoids, and systemic immunosuppressants (such as methotrexate and cyclosporine) also have reported benefits in the treatment of PN. Paroxetine, fluvoxamine, and duloxetine may also be successful in reducing pruritus associated with PN.

Vitiligo

Vitiligo is an acquired, autoimmune skin condition characterized by chronic depigmentation of the skin and mucosa. It is a disfiguring pigmentary disorder with a profound impact on quality of life due to social stigmatization. Misconceptions persist in some cultures that vitiligo is contagious and a sign of leprosy or sexually transmitted infection.

Clinical Features. Vitiligo has a global prevalence of 0.2% to 1.8%. Half of all cases have a childhood onset, but vitiligo may present throughout the lifespan. Females with vitiligo tend to have their first presentation earlier in life than their male counterparts. The condition is characterized by the development of well-circumscribed white macules and patches commonly seen on the face and acral regions. Vitiligo is typically otherwise asymptomatic, but lesions may itch and have a propensity to sunburn. Vitiligo is typically more noticeable in darker skin, and lesions commonly develop in areas of friction (Koebner phenomenon). Examination with a Wood lamp can be helpful in diagnosis.

Pathophysiology. Vitiligo results from the autoimmune destruction of melanocytes. Though the exact pathogenesis remains unknown, it is thought that disease development may be associated with increased expression of interferon-gamma and tumor necrosis factor (TNF) alpha and decreased expression of tumor growth factor (TGF) beta. Antibodies against anti-inflammatory enzymes (like tyrosine hydroxylase) may also play a role. Increased levels of homocysteine (associated with increased inflammation) may predispose patients to the disease.

Psychiatric Comorbidity. Approximately 75% of vitiligo patients are reported to have a co-occurring psychiatric disorder, most frequently depression. One large-scale cohort study found that patients with vitiligo had a 2.9-fold increased risk for psychiatric disorders, with increased rates of OCD, anxiety, MDD, bipolar disorder, and schizophrenia. Acral (facial) vitiligo has the greatest association with psychiatric morbidity (87%), followed by generalized vitiligo (68%) and mucosal vitiligo (63%).

Treatment. Standard treatment for vitiligo consists of topical corticosteroids, calcineurin inhibitors, and phototherapy. Given the limited efficacy of these measures, treatment should focus on optimizing patient quality of life. CBT, peer support groups, and interventions targeting co-occurring psychiatric disorders can play a substantial role in the management of individuals with vitiligo.

PSYCHOPHYSIOLOGIC DISORDERS

These diagnoses describe skin diseases precipitated or exacerbated by psychological stress.

Acne

Clinical Features. Acne is an inflammation of the pilosebaceous glands mainly involving the face and trunk that most frequently occurs in adolescence and manifests as comedones (comedonal acne), papulopustules (papulopustular acne), or nodules and cysts (nodulocystic acne and acne conglobata) (Fig. 27.12–7). Possible complications such as pitted, depressed, or hypertrophic scars may result from all types, especially in nodulocystic and conglobata acne. It is a very common skin disorder among young people and usually begins at 10 to 17 years of age in females and 14 to 19 years in males. The course is usually self-limiting, and most individuals improve over a period of months to years. Acne may persist into the third or fourth decade of life, especially in women. Genetic predisposition is multifactorial, and most patients with cystic acne have a family history of severe acne. Severe acne may be associated with XYY syndrome.

Pathophysiology. Acne results from changes in the keratinization pattern of sebaceous glands followed by dense keratin blocking the secretion of sebum. The keratin plugs are called *comedos,* and they cause collection of sebum within pilosebaceous units prompting complex interactions of androgens and bacteria (*Propionibacterium acne*) within the plugged sebaceous glands. Androgens stimulate the sebaceous glands to produce more sebum. Bacteria containing lipase breaks down lipids into fatty acids and produce proinflammatory mediators like IL-1 and TNF-α, which combine to cause a sterile inflammatory response. Ultimately, the distended follicle wall breaks, releasing the contents (sebum, lipids, fatty acids, keratin, and bacteria) into the dermis, invoking an inflammatory foreign body response resulting in the formation of pustules, nodules, and scarring. Patients with acne vulgaris frequently experience a stress-induced exacerbation of their acne lesions. This association is postulated to involve stress-induced activation of the hypothalamic–pituitary–adrenal (HPA) axis followed by increased secretion of glucocorticoids and androgens, exacerbating the inflammatory response within the pilosebaceous unit. Increased levels of anxiety in patients with cystic acne have been associated with higher blood catecholamine levels which then decrease with acne treatment.

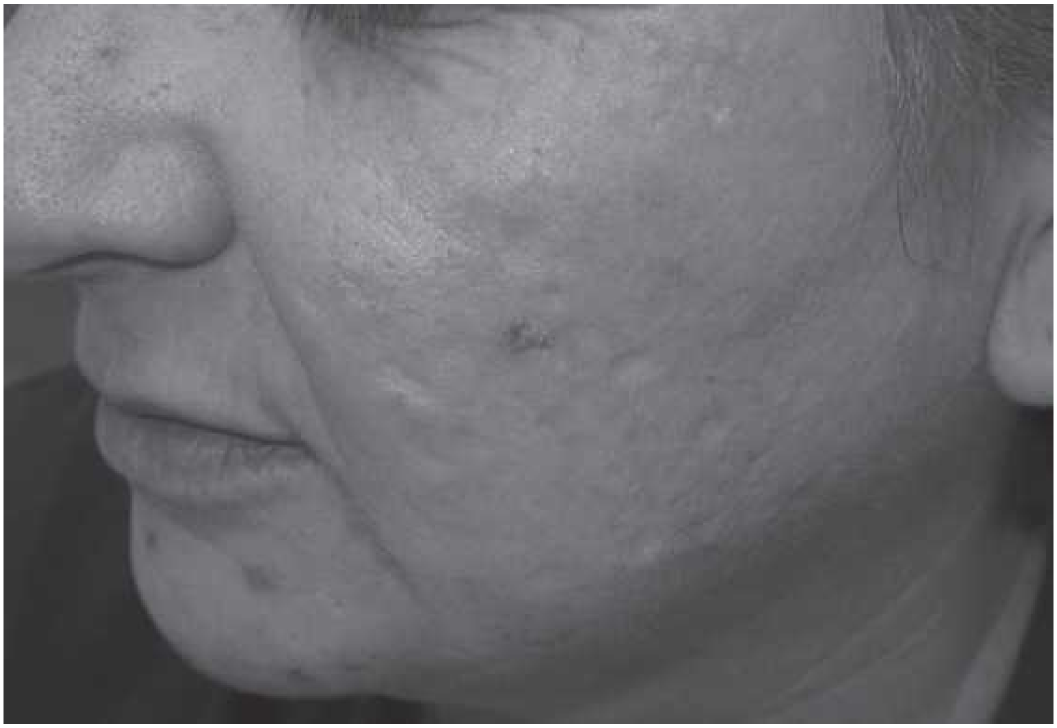

FIGURE 27.12–7. Acne vulgaris, nodulocystic type, with nodules, cysts, and pustules. (The Barankin Collection. Wolters Kluwer, 2005.)

Psychiatric Comorbidity. Disfigurement from acne can lead to depression, social phobia, anger, and low self-esteem. One study reported a 5.6% prevalence of suicidal ideation in adolescents and young adults with noncystic facial acne. In a large study of 4,376 patients with dermatologic disorders, 53% of those with acne reported a close chronologic association between emotional stress and exacerbation of their acne lesions (with an average latency of 2 days). In a study of 317 students aged 14 to 16 years, those with significant (i.e., more than 12) lesions had higher levels of emotional and behavioral difficulties than those with minimal acne. In adolescence, acne can significantly interfere with social interactions, sports, academic performance, and dating behavior. There is a highly significant positive association between acne and poor self-image. In one study, patients' ratings of their acne were consistently more severe than ratings provided by a clinician, and the patients rated their treatment outcomes significantly less favorably than their treating dermatologist. Even mild-to-moderate acne has been associated with depression, suicidal ideation, and completed suicide. Often, the presence of a primary psychiatric disorder such as OCD, delusional disorder, or eating disorder can add to the severity and chronicity of acne, as demonstrated in cases of acne *excoriée*, where the urge to persistently pick acne lesions results in severe disfigurement from scarring.

Treatment. Since younger patients with acne vulgaris have a higher risk of developing depression and suicidal ideation, it is important to educate and monitor each patient carefully. Treatment options for depression, anxiety, and social phobia in patients with acne include CBT, relaxation training, self-hypnosis, and pharmacotherapy if indicated. One study reported a greater degree of improvement in acne when the use of cognitive imagery and biofeedback relaxation was added to typical medical treatment, and discontinuation of such interventions led to a relapse in acne severity. There are case reports of improvement of acne after treatment of depression with paroxetine.

There has been extensive attention to the antiacne agent isotretinoin and its association with depression, suicidal ideation, and suicide attempts. An additional warning has been added to the psychiatric section of the package insert of the drug regarding aggressive or violent behaviors. A large epidemiologic study, however, failed to demonstrate a higher prevalence of depression and suicide in patients treated for acne with isotretinoin versus those treated with antibiotics. Design limitations of this and other studies preclude the establishment of firm conclusions about a possible causal relationship between isotretinoin and psychiatric pathology. A careful history will reveal whether or not an exacerbation of depressive symptoms is temporally related to initiating isotretinoin treatment. The acne patients with symptoms of depression should routinely be assessed for suicide risk regardless of the treatment employed.

Atopic Dermatitis/Eczema

Clinical Features. AD is a chronic inflammatory disease of the skin, characterized by pruritus, mild to severe erythema, scaling, excoriation, and lichenification of the skin (Fig. 27.12–8). AD typically begins in infancy or young childhood, presenting as erythematous and pruritic maculopapular eruptions. Subsequent rubbing and scratching leads to lichenification and further itching and scratching cycle (itch–scratch cycle). The pruriginous lichenoid papules and lichenification in flexor areas and hands become more prominent in adolescence and adulthood. Individuals with AD may have a personal or family history of atopic disease (including allergic rhinitis, allergic conjunctivitis, bronchial asthma, or seasonal allergies). It is estimated that 1% to 3% of adults and 5% to 20% of children have AD. AD manifests in the first year of life in about 60% of cases and 30% in the subsequent 1 year. The clinical course of AD improves by puberty, and initial presentation is rare after puberty. Both sexes are affected, but boys are somewhat more afflicted during the infantile phase. The hygiene hypothesis proposes that a Western lifestyle with early exposure to psychosocial stress and reduced microbial exposure (from near-sterile food preparations and large-scale immunizations) affects the differentiation of T cells toward an atopic profile of predominant TH2 cells. A combination of early-onset severe dermatitis, comorbid asthma, or allergic rhinitis is often associated with disease chronicity.

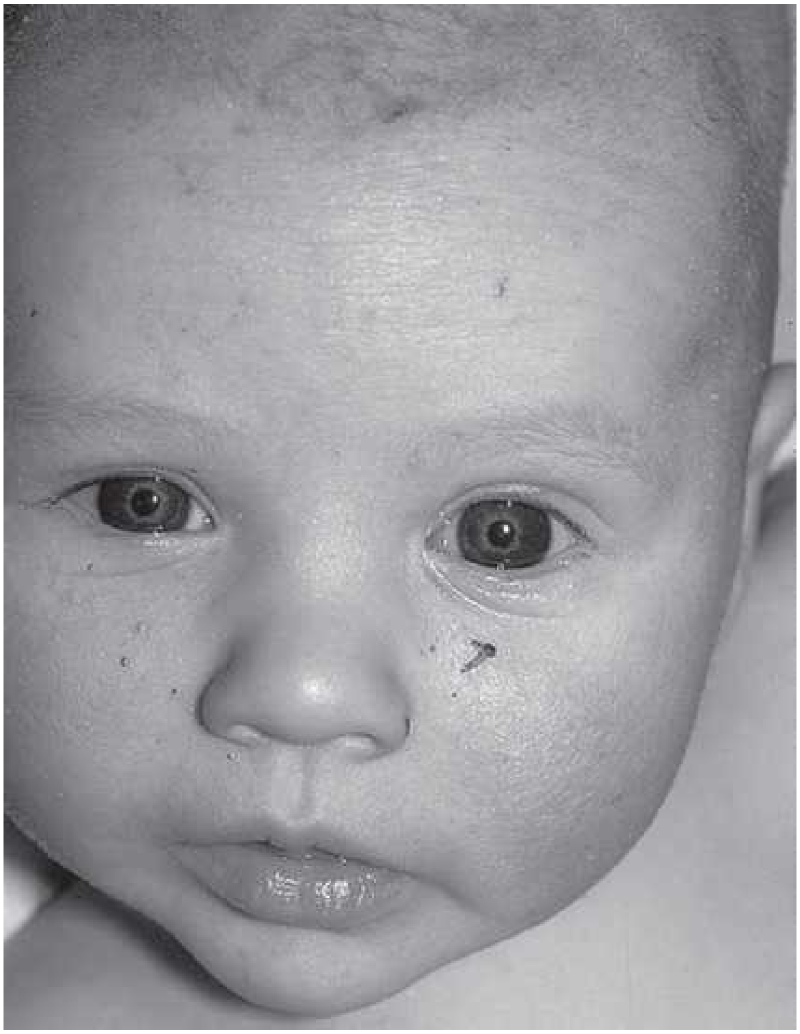

FIGURE 27.12–8. Atopic dermatitis. Infantile-confluent erythema, microvesiculation, scaling, and crusting on face. (Reprinted with permission from Goodheart HP. *Goodheart's Photoguide of Common Skin Disorders*. 2nd ed. Lippincott Williams & Wilkins; 2003.)

Pathophysiology. AD's pathophysiology involves multiple factors, including genetic, environmental, and psychosocial factors, which interact with underlying psychoneuroimmunologic mechanisms.

A genetic predisposition is implicated in the development of AD, and about two-thirds of patients have a positive family history of AD. The concordance rate for monozygotic twins (0.75) is higher than dizygotic (0.25) twins. Inheritance is considered to be polygenic and multifactorial. The inheritable traits of AD are thought to involve HPA axis abnormalities and an imbalance of the immunoregulatory T-lymphocyte system, leading to defective cell-mediated immunity and increased IgE production.

The psychosocial sequelae of AD have been described as including perceived maternal rejection and emotional separation from the mother from lack of early cutaneous stimulation, altered psychodynamics of the mother–child relationship, temperamental changes in AD infants, problems with self-esteem and relationships in childhood, and development of conflicted personality styles.

Psychiatric Comorbidity. Individuals with AD have been found to have increased rates of anxiety and depression. Scheich and colleagues reported that AD patients have higher traits of excitability and inadequate stress coping mechanisms, with these traits being particularly prominent in those with serum IgE levels greater than 100 IU/mL. M. Hashiro and M. Okumura examined immunologic markers and found that NK-cell activity and IL-4 levels are lower in AD patients with depression or anxiety, and IFN-γ levels are higher in AD patients than in controls. Decreased NK-cell activity is more strongly correlated with higher levels of anxiety. The severity of pruritus was directly correlated with the severity of depressive symptoms in AD patients, which is consistent with findings that depression and anxiety can magnify the itch perception and enhance scratching behavior. Higher anxiety levels in AD patients are correlated with ease of conditioning to develop an itch-scratch response. The emotional state of children with AD varies closely with the severity of the dermatitis. Adult patients with AD have been reported to have chronic anxiety disorders and a tendency to internalize anger.

Treatment. The role of psychiatric treatment in patients with AD includes reduction of itching and scratching and treatment of associated anxiety and depressive symptoms. Behavioral modalities focus on interrupting the itch–scratch cycle (HRT). Controlled trials found that topical 5% doxepin cream was effective in reducing pruritus because of its potent antihistaminic action. Trimipramine, which also features histamine antagonism, has been shown to improve sleep quality and reduce nighttime scratching in AD patients.

One meta-analysis examined various psychotherapeutic approaches (including CBT, relaxation training, meditation, and stress management) with respect to skin condition and subjective well-being. These treatments were significantly more effective than standard somatic treatments used alone and were effective in reducing anxiety, depression, and neuroticism. Combinations of psychotherapeutic interventions resulted in a reduction of the itch–scratch cycle, less use of local and systemic steroids, and longer periods of remission.

Psoriasis

The writer John Updike, who had psoriasis, poignantly wrote: “I am silvery, scaly. Puddles of flakes form wherever I rest my flesh. Lusty, though we are loathsome to love. Keen-sighted, though we hate to look upon ourselves. The name of the disease, spiritually speaking, is Humiliation.”

Clinical Features. Psoriasis is a chronic, relapsing skin disease presenting with erythematous, scaling papules and indurated plaques, arising preferentially on the elbows, knees, and scalp (Fig. 27.12–9). The clinical presentation takes various forms, sometimes causing nail dystrophies and arthritis. It affects 1.5% to 2% of the population in Western countries, with 3 to 5 million people with psoriasis in the United States. Psoriasis is hereditary, with an age of onset mostly in the early 20s. Childhood-onset predicts a severe and chronic course and is usually associated with a positive family history. Common trigger factors include physical trauma (Koebner phenomenon: rubbing and scratching stimulate the psoriatic proliferative process), acute bacterial or viral infections, and drugs like lithium, antimalarials, interferon, oral alcohol, β-adrenergic blockers, and the withdrawal of corticosteroids. Psychological stress is another reported trigger of exacerbations of psoriasis, with one large study reporting that 33% of patients noticed new lesions at times of distress. Another investigation of 536 patients found that 42% of them reported worrying was associated with exacerbation of their lesions, with the majority of those specifically noting the disease-related stress from cosmetic disfigurement and social stigma as more impactful than stressful major life events.

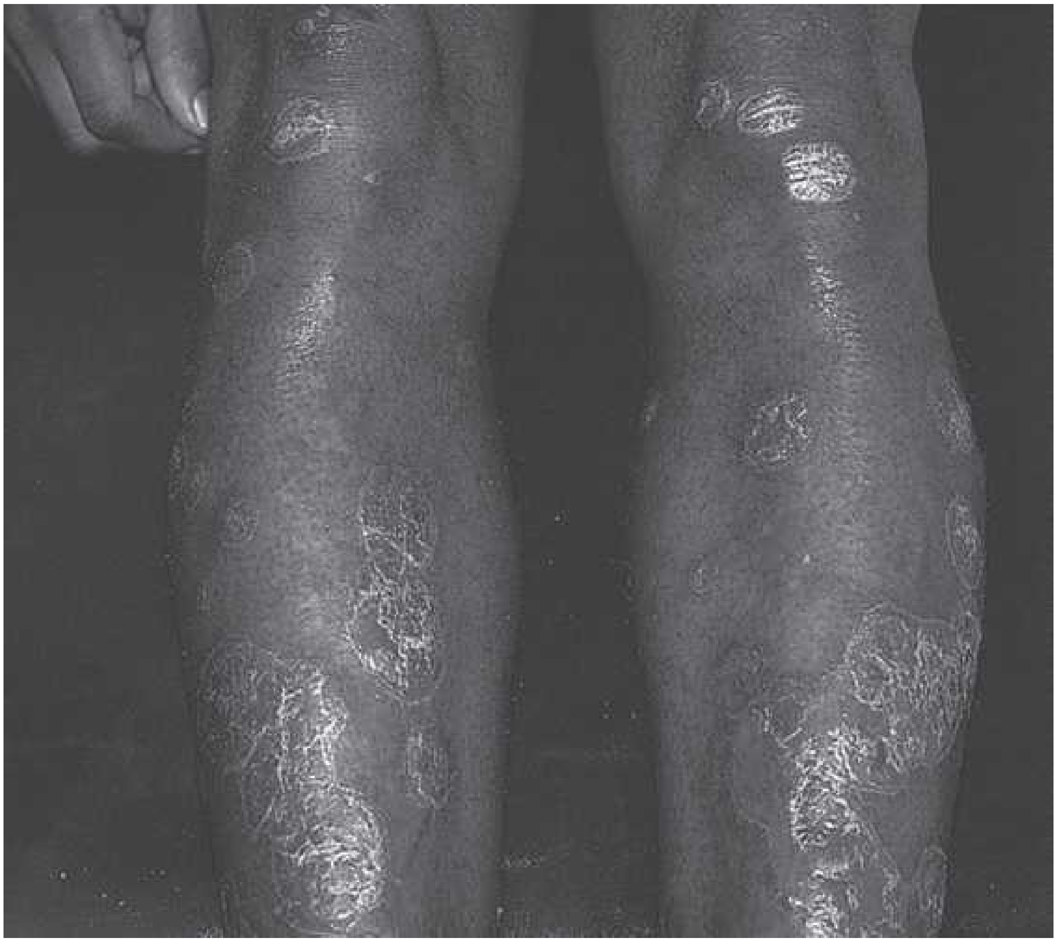

FIGURE 27.12–9. Psoriasis vulgaris, chronic stable type. Multiple large scaling plaques on the legs. (Reprinted with permission from Goodheart HP. *Goodheart's Photoguide of Common Skin Disorders*. 2nd ed. Lippincott Williams & Wilkins; 2003.)

Pathophysiology. Psoriasis plaques are produced largely by the activity of hypertrophic keratinocytes (multiplying approximately 28 times faster than normal skin) and inflammatory changes in the dermis. Histologic markers include capillary dilatation, sterile microabscesses, parakeratosis, and hyperproliferation of cutaneous nerves. The *neurogenic inflammation* hypothesis of psoriasis describes central roles for NPs like SP and nerve growth factor (NGF). SP and other NPs are released from the unmyelinated terminals of sensory fibers in the skin and generate local neurogenic inflammation in genetically predisposed individuals. Both lesional and symptomless psoriatic skin have been found to have increased numbers of neural filaments containing NPs like SP, vasoactive intestinal peptide (VIP), and CGRP. The specific effects of NPs on target cells in psoriasis are listed in Table 27.12–6.

Increased levels of SP both centrally and peripherally have been associated with increased levels of stress. Descending autonomic pathways via the dorsal root ganglia in the spinal cord can stimulate the further release of cutaneous NPs through SP-containing neurons. There is evidence for higher NGF levels in response to stress, which plays a central role in regulating skin innervation and upregulating NPs. Several studies indicate that increased expression of NGF in keratinocytes may be an early event in the pathogenesis of psoriasis. NGF stimulates T-lymphocyte proliferation and mast cell degranulation and induces expression of a potent chemokine RANTES, which has chemotactic effects on resting $CD4^+$ memory T cells and memory T cells.

There has been significant attention to the role of T lymphocytes in the pathogenesis of psoriasis, but the role of autoimmune inflammation in the pathogenesis of psoriasis remains controversial. The T-cell paradigm alone does not explain some of the important clinical features of psoriasis, including the Koebner phenomenon, the symmetrical distribution of lesions, and increased nerve density and NP expression in psoriatic lesions.

Table 27.12–6.
Neuropeptides and Their Effects on the Various Skin Cells

Target Cells	Neuropeptides		
	SP	**CGRP**	**VIP**
Endothelial	Proliferation, ↑ permeability	Proliferation, IL-8 secretion	Proliferation
Keratinocyte	Stimulates IL-1, IL-8, comitogen with CGRP, LTB4	Proliferation in association with SP	Mitogenic proliferation
Lymphocyte	Stimulates proliferation and synthesis of IL-2, ↑ IgA and IgM, B-cell differentiation	Chemotactic T-cell proliferation	Inhibits NK-cell activity
Macrophage	Stimulates IL-1, IL-6 secretion	Impairs antigen presentation	—
Mast cell	Proliferation, histamine and TNF-α release	Degranulation: histamine release	Histamine release
Monocyte	↑ Chemotaxis, phagocytosis, arachidonic acid synthesis	—	Enhanced migration
Neutrophil	Stimulates chemotaxis, phagocytosis	Enhances chemotaxis	—

↑, increase; SP, substance P; VIP, vasoactive intestinal peptide; CGRP, calcitonin gene-related peptide; Ig, immunoglobulin; IL, interleukin; NK, natural killer; TNF-α, tumor necrosis factor α.

Psychiatric Comorbidity. Patients with psoriasis have been reported to have higher prevalences of generalized anxiety disorder, major depressive disorder, and personality disorders. Suicidal ideation was reported in 7.2% of severely affected psoriasis inpatients versus 2.5% among less severely affected outpatients. Psoriasis-related stress has been associated with greater psychiatric comorbidity. Social deprivation and the experience of stigmatization were found to be correlated with higher depression scores. Adult psoriasis patients who experienced touch deprivation in social interactions had higher depression scores compared to those who did not perceive touch deprivation. The severity of pruritus is also associated with higher depression scores and a greater risk for suicide, with amelioration of pruritus associated with a decrease in depression scores. Early-onset psoriasis has been associated with greater difficulties with the expression of anger, a trait that may result in increased vulnerability to stress and depression.

Treatment. The therapeutic modalities with evidence for significant clinical improvement in patients with psoriasis include arousal reduction techniques like hypnosis, mindfulness meditation, and guided imagery training. CBT may be of specific value in addressing illness perceptions, including beliefs about the control and curability of illness. Data from controlled trials investigating the effects of psychopharmacologic interventions in psoriasis are quite limited. TNF-α antagonists are safe and effective in improving both the physical severity of psoriasis and quality of life. Depression is also associated with increased levels of TNF-α, and it is speculated that it may have a role in the fatigue and sleep disturbances seen in depression. The TNF-α antagonist adalimumab has been shown to improve depressive symptoms and fatigue in patients with psoriasis. Reduction in anxiety and depression scores has also been reported with the use of ustekinumab (an anti-IL-12/23 agent used to treat psoriasis). Psychotropic medications should be used with some caution in patients with psoriasis due to several conflicting case reports of induction or exacerbation of psoriasis with the use of fluoxetine and bupropion. Lithium is well-associated with precipitation or aggravation of psoriasis and should be avoided.

Telogen Effluvium

Clinical Features. Telogen effluvium (TE) is a scalp disorder characterized by the excessive shedding of hair in response to a variety of triggers, including drugs, trauma, and emotional or physiologic stress. The time between trigger exposure and hair loss is typically 8 to 12 weeks, differentiating this diagnosis from anagen effluvium, which describes the more immediate hair loss seen in response to some drugs, including cancer chemotherapeutic agents. It is also one of the most common causes of alopecia with trichodynia, a painful sensation of the scalp without underlying cutaneous disease. Diagnostic tests include the “hair wash test” (where after a period of abstaining from hair washing, the hair is washed and any shed hairs are trapped and assessed for quantity and length), trichograms (hair is plucked and microscopically examined), phototrichograms (hair in a 2 cm^2 area is clipped to a length of 1 mm and serial weekly photographs are used to monitor growth) and scalp biopsy.

Pathophysiology. Hair shedding in TE can be attributed to a variety of potential changes in the hair follicle growth cycle. In some instances, hair follicles leave the anagen phase prematurely, resulting in increased shedding 2 to 3 months later. In other cases, the anagen stage becomes delayed, resulting in heavier than normal telogen shedding. Drugs such as oral contraceptive pills, retinoids, β-blockers, ACE inhibitors, heparin, anticonvulsants, and antidepressants have all been associated with TE—usually 12 weeks after starting the regimen. A variety of stressors have been linked to TE, including psychological stress, surgery, childbirth, and jet lag. Severe deficiencies of zinc or vitamin D decreased iron stores, and calorie restriction have also been reported to trigger TE.

Psychiatric Comorbidity. One study found that 76.5% of patients with TE had symptoms of psychiatric illness. Trichodynia is commonly associated with TE and is also associated with depression, anxiety disorders, dysthymic disorder, and OCD. Disturbances in sleep have also been linked to TE.

Treatment. Effective treatment of TE hinges on the identification and elimination of the trigger. Minoxidil and finasteride are commonly used but have minimal efficacy. Topical corticosteroids can be effective in decreasing trichodynia associated with TE, and systemic corticosteroids can be employed in cases of TE associated with underlying systemic disorders. If a measurable nutritional deficiency is detected, it should be corrected with supplementation, a balanced diet, and stabilization of body weight. Clinicians may recommend meditation, yoga, or mindfulness exercises to address psychological stress, but there are not yet controlled data to support therapeutic efficacy vis-à-vis hair loss. TE is ultimately self-limiting and will not result in complete baldness.

Urticaria and Angioedema

Clinical Features. Urticaria is composed of wheals (hives) which are characterized by transient, well-circumscribed edematous papules and plaques that are usually pruritic and involve the superficial (papillary) dermis. Angioedema describes a large, deep, ill-defined edematous area involving the dermis and subcutaneous tissue. Thus, urticaria and angioedema are the same edematous process involving different levels of the cutaneous vascular plexus. Fifteen to 23% of the population may have a lifetime occurrence of this condition, with a chronic course in about 25% of the cases. *Acute urticaria* has a sudden onset with frequent recurrence within 30 days. It usually presents with large wheals, often associated with angioedema. Acute urticaria is IgE-dependent and associated with common triggers such as alimentary parasites, infection, and intolerance to some drugs or food. Urticaria reactions are mediated by vasoactive mediators in the skin, such as histamine from mast cells and basophils. *Chronic urticaria* has a recurring pattern over more than 30 days and presents with both small and large wheals. These wheals are *rarely* IgE-dependent but are often due to anti-FcεR and/or anti-IgE autoantibodies, suggesting autoimmune pathogenesis. Since an etiology is unknown in 80% of cases, the condition is considered idiopathic. Chronic urticaria predominantly affects adults and is twice as common in women as in men. About 40% of patients with chronic urticaria persisting for more than 6 months will continue to have urticaria for 10 years or longer with a serious compromise of quality of life.

Pathophysiology. Urticaria and angioedema can be subcategorized on the basis of etiology, including *immunologic* (IgE-mediated, complement-mediated, autoimmune, immune contact type), *physical* (friction-induced “dermographism,” cold urticaria, solar, cholinergic, pressure angioedema, vibratory angioedema), *idiopathic*, and *hereditary angioedema*. Of particular interest is adrenergic urticaria, presenting as typical “halo-hives” (papules surrounded by a white halo), which develop in response to acute emotional stress and are usually associated with increased plasma levels of epinephrine and norepinephrine. Adrenergic urticaria improves after treatment with β-blockers such as propranolol and can be induced by intradermal injection of norepinephrine.

Psychiatric Comorbidity. The role of psychological factors in chronic idiopathic urticaria is unclear. One study reported higher prevalences of depression and anxiety in women with chronic urticaria. Historically, angioedema has also been referred to as "angioneurotic edema" because of the frequent clinical observation that stressful events preceded its onset. The physiologic basis for stress-induced urticaria and angioedema is likely to involve secretion of NPs, such as VIP and SP, resulting in vasodilatation and urticarial wheal formation.

Treatment. In urticaria with a known trigger, treatment includes the prevention of future attacks by avoiding exposure to that agent. Controlled trials have indicated that antidepressants like doxepin (10 mg three times a day) are more effective than diphenhydramine (25 mg three times a day) in managing chronic angioedema. There are anecdotal reports of improvement in urticaria symptoms with the use of SSRIs in patients with urticaria and co-occurring panic disorder. One controlled study found that hypnosis and relaxation reduced the subjective sense of pruritus in cases of urticaria but did not affect the number of wheals.

SLEEP, CIRCADIAN RHYTHMS, AND SKIN

Sleep is vital to the growth and renewal of multiple physiologic systems. From a dermatologic perspective, chronic poor sleep quality is associated with increased signs of intrinsic skin aging, diminished skin barrier function, and lower satisfaction with appearance. Slow-wave sleep stimulates the release of growth hormone and promotes normal tissue repair.

The human sleep–wake cycle is tightly coupled to the circadian course of *core body temperature*. Heat loss in the evening, via distal skin regions, is intimately associated with sleepiness and ease of sleep initiation. The preoptic anterior hypothalamus (POAH) region of the brain is the main integrator of sleep and thermoregulation. If thermoregulatory function is impaired as a result of a skin disorder and heat is not dissipated normally, sleep onset may be prolonged, resulting in a decline in restorative sleep and subsequent risk of poor health outcomes.

The stratum corneum is the outermost layer of the epidermis. It undergoes circadian changes in its permeability, with higher permeability in the evening. Aquaporin-3 is protein found in the collecting duct cells of the epidermis, allowing water to exit cells. It is regulated by molecular clocks, and more active in the evening, resulting in increased pruritus at night.

In addition to Aquaporin-3, several other factors may contribute to nocturnal pruritus, including peripherial vasodilation at night with an consequential increase in skin temperature. Corticosteroids also have a circadian pattern and release of these is decreased at night; the result is a decreased anti-inflammatory response and associated pruritus. Prostaglandin D2 and E2 are elevated at night, and may also exacerbate pruritus. Many other substances, such as cytokines, NGF, and endogenous opioids, all which have circadian patterns, may also influence pruritus.

DERMATOLOGIC SIDE EFFECTS OF PSYCHOTROPIC MEDICATIONS

Dermatologic side effects of psychotropic medications are fairly common though the majority of them are benign and easily treatable. Most drug reactions are hypersensitivity reactions with an underlying immunologic mechanism; however, some reactions are attributed to cumulative dose, drug overdose, and drug–drug reactions. Dermatologic side effects of psychotropic medications are summarized in Table 27.12–7.

When categorizing dermatologic side effects, it is valuable to consider potential mechanisms responsible for the cutaneous pathology. Most common are Type A drug reactions (e.g., phototoxic reactions, anticholinergic effects, or extrapyramidal effects), which are due to the pharmacologic or toxic effects of the drug. They are well-defined adverse reactions that can be predicted to occur in a segment of the population exposed to the drug and are frequently dose-dependent. Type B drug reactions, on the other hand, are far less common and less predictable. Those experiencing them may have an innate predisposition for the reaction due to their genetic or metabolic makeup. Type B reactions are immunologically mediated (usually type I or IV hypersensitivities) and are thus largely unpredictable and not dose-dependent.

Risk factors associated with cutaneous drug reactions include female sex, HLA subtype, African American ethnicity, winter season, increased age, active substance abuse, and a high-loading dose on initiation of the drug. The amount and the rate of administration may also influence the likelihood of an adverse cutaneous reaction. Up to 95% of cutaneous drug reactions present with erythematous macules and papules (an exanthematous eruption) within 10 days of starting treatment. Steven–Johnson syndrome (SJS), toxic epidermal necrolysis (TEN), and drug-induced hypersensitivity syndrome (DIHS) are relatively rare reactions to psychotropics but are true dermatologic emergencies that require immediate recognition to allow for timely discontinuation of the offending agent. Other less important but noteworthy reactions include photosensitivity, pigmentation reactions, acneiform eruptions, lichenoid reactions, alopecia, erythema multiforme, and cutaneous pseudolymphoma.

Table 27.12–7.
Dermatologic Adverse Effects of Psychotropic Medications

Drugs	Skin Adverse Effects
MOOD STABILIZERS	
Lithium	Hair loss, acne, nail pigmentation, exacerbation of psoriasis
Valproic acid	Hair loss, hair color change, lupus, scleroderma, skin vasculitis
Lamotrigine	Stevens–Johnson syndrome, toxic epidermal necrolysis, angioedema
Carbamazepine	Pruritic rash, hair loss, Steven–Johnson syndrome, hypersensitivity reaction
ANTIDEPRESSANTS	
SSRIs	Allergic reaction (hives, urticaria), excessive sweating, pruritus; hair loss reported with fluoxetine
Bupropion	Erythroderma, erythema multiforme (rare)
Venlafaxine	Erythroderma, erythema nodosum
Duloxetine	Erythema multiforme
Mirtazapine	Erythroderma
TCAs	Photosensitivity, erythroderma
Vortioxetine	Pruritus
Vilazodone	Drug rash
ANTIPSYCHOTICS	
Haloperidol	Drug hypersensitivity
Phenothiazine	Skin pigmentation (blue-gray discoloration)
Clozapine	Erythema multiforme, Steven—Johnson syndrome, drug hypersensitivity
Risperidone	Erythema multiforme, erythroderma
Olanzapine	Drug rash with eosinophilia
Ziprasidone	Erythroderma
Aripiprazole	Drug rash (rare)
Lurasidone	Drug rash, pruritus, angioedema (rare)
Paliperidone	Drug rash
ANXIOLYTICS	
Alprazolam	Photosensitivity

Steven–Johnson Syndrome (SJS) and Toxic Epidermal Necrolysis (TEN)

Diagnosis and Clinical Features. Diagnosis is based on clinical assessment in conjunction with histopathologic findings of subepidermal blisters. Both conditions manifest with an “influenza-like” prodromal phase marked by malaise fevers with subsequent painful mucosal and cutaneous lesions. The two conditions are classified according to the degree of skin detachment: SJS involves <10% skin involvement, whereas TEN encompasses >30% skin involvement. SJS/TEN overlap refers to anywhere from 10% to 30% skin detachment. All forms of the disease are characterized by blisters (either partially or completely detached) with Nikolsky positive, detachable areas. Nasopharyngeal, esophageal, and genital mucosal lesions or strictures may occur, along with pulmonary, ocular, renal, and hepatic manifestations. Sepsis, however, is the leading cause of morbidity and fatality in the acute phase of the disease. A majority of cases of SJS/TEN due to psychotropic exposure can be attributed to the use of carbamazepine (67.8%) though the reaction has also been associated with exposure to lamotrigine oxcarbazepine and phenobarbital.

Treatment. The acute management of SJS/TEN requires a multidisciplinary approach with immediate withdrawal of causative agent and prompt referral to a medical center for supportive treatment with systemic corticosteroids, immunoglobulins, and cyclosporine A. Cyclophosphamide and thalidomide have previously been used in the treatment of SJS/TEN but are no longer recommended due to high mortality rates and little efficacy as a treatment.

Most Common Associated Cutaneous Adverse Events

Mood Stabilizers. Mood stabilizers are the class of psychotropics most likely to produce cutaneous adverse events, and neuroleptics are the least common. Of the mood stabilizers, carbamazepine and lamotrigine are associated with the highest rates of cutaneous drug reactions. There is evidence that concurrent use of valproic acid considerably raises the risk of a cutaneous drug reaction with lamotrigine administration. The overall incidence of cutaneous effects with lithium is low. The most frequent cutaneous reactions reported with lithium use are psoriasis, acneiform eruptions, folliculitis, alopecia, and exanthems.

Antidepressants. Of the antidepressants, TCAs have the highest reported incidence of cutaneous reactions, though these reactions often resolve with discontinuation of the offending agent. Of all the SSRIs, fluoxetine has been noted to have the highest association with purpura and ecchymosis, whereas alopecia has been reported with the use of all the SSRIs. Although there is conflicting data regarding hyperhidrosis with the use of SSRIs, it seems to be a relatively common issue that negatively impacts patient adherence to treatment with this drug class. The most common cutaneous effect of the SNRIs is hyperhidrosis which is thought to be induced by the inhibition of norepinephrine uptake resulting in the stimulation of peripheral adrenergic receptors.

Neuroleptics. The most common cutaneous side effects associated with the neuroleptics are pigmentation changes and photosensitivity, with chlorpromazine being the agent most frequently associated with either of these reactions. Discontinuation of the drug and avoidance of sunlight are essential in the treatment of this cutaneous reaction. Loxapine may be used as an alternative to phenothiazines to avoid skin pigmentation.

Treatment

Simple discontinuation of the offending drug often resolves cutaneous adverse effects, but this decision should be made with careful attention to the risk/benefit calculus. In cases of lithium and carbamazepine, lowering the dose can result in a resolution of the skin condition. Specifically for carbamazepine, lower loading doses on initiation and a slow increase in dosage can help avoid cutaneous eruptions altogether. Mild cases of hyperhidrosis can be managed with topical aluminum chloride hexahydrate. For the treatment of cutaneous drug reactions, antihistamines are used approximately 50% of the time, topical steroids 39% of the time, and oral steroids 37% of the time. If fever, arthralgia, palpable purpura, cutaneous sloughing, or mucus membrane erosions are noted, the drug should immediately be discontinued and an emergent dermatology consultation obtained. Patients who manifest with lesions over a large surface area should be hospitalized as these lesions may result in dehydration or infection, similar to that seen in patients with burns.

Prevention may soon become the focus of drug-induced cutaneous eruptions. There are compelling data that show a strong association between certain HLA subtypes and drug-induced cutaneous hypersensitivities. The FDA recommends genetic testing before initiating carbamazepine therapy and warns against using the drug in those with Asian ancestry or those known to have the HLA-B 1502 genotype on the product label. The label for carbamazepine also includes a warning regarding the risk of SJS/TEN in those positive for HLA-A 3101. Labeling for oxcarbazepine carries a similar warning regarding HLA-B 1502.

FUTURE DIRECTIONS

Psychocutaneous disorders span the interface between dermatology and psychiatry, encompassing a wide range of conditions with mixed etiologies with a significant impact on quality of life. Despite this, psychodermatoses remain largely understudied, with few large-scale trials to assess effective treatment. There is also a need for a better understanding of the impact of skin disorders on development in childhood and adolescence, self-image (CBI), early relationships, psychosocial functioning, and the roles of stigma and stress in the exacerbation of skin disorders. A growing understanding of psychoneuroimmunology is underscoring the role of emotional stress in the pathogenesis of some skin disorders and elucidating potential mechanisms that may mediate the bidirectional influences of stress and skin and provide targets for clinical interventions that may dampen or eliminate stress-induced exacerbation of dermatologic illness.

Further Readings

Asghar F, Shamim N, Farooque U, Sheikh H, Aqeel R. Telogen effluvium: a review of the literature. *Cureus*. 2020;12(5):e8320.

Blashill AJ, Oleski JL, Hayes R, et al. The association between psychiatric disorders and frequent indoor tanning. *JAMA Dermatol*. 2016;152(5):577–579.

Craig AD. How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci*. 2009;10(1):59–70.

Dhabhar FS. Psychological stress and immunoprotection vs. immunopathology in the skin. *Clin Dermatol*. 2013;31:18–30.

Dixon LJ, Lee AA, Viana AG, McCowan NK, Brodell RT, Tull MT. Anxiety sensitivity in dermatological patients. *Psychosomatics*. 2016;57(5):498–504.

Fesler MC, Middelveen MJ, Stricker RB. Clinical evaluation of Morgellons disease in cohort of North American patients. *Dermatol Reports*. 2018;10(1):7660.

Fried RG, Gupta MA, Gupta AK. Depression and skin disease. *Dermatol Clin*. 2005;4:657–664.

Grant JE, Chamberlain SR, Redden SA, Leppink EW, Odlaug BL, Kim SW. N-acetylcysteine in the treatment of excoriation disorder: a randomized clinical trial. *JAMA Psychiatry*. 2016;73(5):490–496.

Grant JE, Odlaug BL, Chamberlain SR, Kim SW. A double-blind, placebo-controlled trial of lamotrigine for pathological skin picking. Treatment efficacy and neurocognitive predictors of response. *J Clin Psychopharmacol*. 2010;30:396–403.

*Gupta MA. Emotional regulation, dissociation, and the self-induced dermatoses: clinical features and implications for treatment with mood stabilizers. *Clin Dermatol*. 2013;31(1):110–117.

Gupta MA, Gupta AK. Current concepts in psychodermatology. *Curr Psychiatry Rep*. 2014;16:449–457.

Gupta MA, Gupta AK. Evaluation of cutaneous body image dissatisfaction in the dermatology patient. *Clin Dermatol*. 2013;31:72–77.

Hashiro M, Okumura M. The relationship between the psychological and immunological state in patients with atopic dermatitis. *J Derm Sci*. 1998;16:231–235.

Hong TL, Wallergren MD, Yosipovitch MD. Psychosomatic factors in pruritus. *Clin Dermatol*. 2013;31:31–40.

Huang AH, William KA, Kwatra SG. Prurigo nodularis: epidemiology and clinical features. *J Am Acad Dermatol*. 2020;83(6):1559–1565.

Hundley JL, Yosipovitch G. Mirtazapine for reducing nocturnal itch in patients with chronic pruritus: a pilot study. *J Am Acad Dermatol*. 2004;50:889–891.

Khanna R, Kwon CD, Patel SP, et al. Intranasal butorphanol rescue therapy for the treatment of intractable pruritus: a case series from the Johns Hopkins Itch Clinic. *J Am Acad Dermatol*. 2020;83(5):1529–1533.

Koo J. Psychodermatology: a practical manual for clinicians. *Curr Probl Dermatol*. 1995;7:203–225.

Koo JY, Lee CS. *Psychocutaneous Medicine*. Marcel Dekker; 2003.

Kuhn H, Mennella C, Magid M, Stamu-O’Brien C, Kroumpouzos G. Psychocutaneous disease. *J Am Acad Derm*. 2017;76(5):779–791.

Lavery MJ, Stull C, Kinney MO, Yosipovitch G. Nocturnal pruritus: the battle for a peaceful night’s sleep. *Int J Mol Sci*. 2016;17(3):425.

Lerch M, Mainetti C, Beretta-Piccoli BT, Harr T. Current perspectives on Steven-Johnson syndrome and toxic epidermal necrolysis. *Clinic Rev Allerg Immunol*. 2018;54(1):147–176.

Lyons AB, Moy L, Moy R, Tung R. Circadian rhythm and the skin: a review of the literature. *J Clin Aesthet Dermatol*. 2019;12(9):42–45.

Mitkov MV, Trowbridge RM, Lockshin BN, Caplan JP. Dermatologic side effects of psychotropic medications. *Psychosomatics*. 2014;55(1):1–20.

Nowak DA, Wong SM. DSM-5 update in psychodermatology. *Skin Therapy Lett*. 2016;21(3):4–7.

Patel SP, Khanna R, Kwatra SG. Proposing a stratification scheme for Chronic Pruritus of Unknown Origin nomenclature based on presence of eosinophilia: implications for therapeutics and cohort homogeneity for clinical trials. *J Am Acad Dermatol*. 2019;(19):32474.

Patel T, Ishiuji Y, Yosipovitch G. Nocturnal itch: why do we itch at night? *Acta Derm Venereol*. 2007;87(4):295–298.

Picari A, Abeni D. Stressful life events and skin diseases: disentangling evidence from myth. *Psychother Psychosom*. 2001;70:118–136.

Selles RR, McGuire JF, Small BJ, Storch EA. A systematic review and meta-analysis of psychiatric treatments for excoriation (skin-picking) disorder. *Gen Hosp Psychiatry*. 2016;41:29–37.

Simons RE, Zevy DL, Jafferany M. Psychodermatology of vitiligo: psychological impact and consequences. *Dermatol Ther*. 2020;33:e13418.

Singh L, Pang X, Alexacos N, Letourneau R, Theoharides TC. Acute immobilization stress triggers skin mast cell degranulation via corticotropin releasing hormone, neurotensin, and substance P: a link to neurogenic skin disorders. *Brain Behav Immune*. 1999;13(3):225–239.

Tey HL, Wallengren J, Yosipovitch G. Psychosomatic factors in pruritus. *Clin Dermatol*. 2013;31(1):31–40.

Walker C, Papadopoulos L. *Psychodermatology: The Psychological Impact of Skin Disorders*. Cambridge University Press; 2005.

Yosipovitch G, Ishiuji Y, Patel TS, et al. The brain processing of scratching. *J Invest Dermatol*. 2008;128(7):1806–1811.

27.13 Organ Transplantation

ANDREA FRANCES DIMARTINI, M.D., MARY SCHWAB SELIGMAN, M.D., M.A.R., PH.D., CATHERINE CHANG CRONE, M.D., AND GERALD SCOTT WINDER, M.D., M.SC.

Although solid organ and other tissue transplants had been attempted since the early 1900s, it was not until 1954 that the first successful organ transplant was accomplished by Joseph Murray using an identical twin brother as a kidney donor. Despite this surgical breakthrough, routine transplantation of organs would not be possible until the challenge of adequate suppression of the immune system was resolved. In the early 1960s, the development of chemical immunosuppression allowed the first deceased donor kidney, liver, lung, and pancreas transplants to be performed, followed by the first heart transplant in the late 1960s. Two decades later, the discovery and development of cyclosporine, a much more powerful and specific immunosuppressant drug, provided selective immunoregulation of T cells without excessive toxicity. Cyclosporine's approval by the FDA in 1983 for use in solid organ transplantation led to a dramatic increase in the numbers of transplants performed, decreases in recipient morbidity, and increases in posttransplant survival rates. This drug and others that followed changed organ transplantation from an experimental procedure to the standard of care for many types of advanced organ disease.

In 1984, the U.S. Congress enacted the National Organ Transplant Act (NOTA), establishing a regionalized system of equitable organ distribution, and the U.S. Department of Health and Human Services contracted with the United Network of Organ Sharing (UNOS) to administer the nation's Organ Procurement and Transplantation Network (OPTN). A central feature of NOTA was a prohibition on the buying or selling of human organs for transplantation, heralding some of the controversies to come in this complex field. In addition to the OPTN's primary goals of increasing the efficiency of organ sharing and the supply of donated organs, UNOS collects extensive data on all solid organ transplants in the United States (e.g., waitlist counts, survival rates), serves as an educational resource, and guides policymaking through the OPTN. OPTN-mandated data collected from transplant programs, maintained by the Scientific Registry of Transplant Recipients (SRTR), are available for public access on the SRTR website (http://srtr.transplant.hrsa.gov). Although transplantation has become a life-sustaining procedure for those with advanced organ disease, the stresses inherent in the process—including the severe medical illness of the candidate, lack of available organs, and the uncertainty of surviving transplantation, risk of surgery, and the challenges of postoperative recovery—place enormous demands on transplant candidates, recipients, and their caregivers. In addition, the organ shortage—as well as extensive health care resources and the degree of personal responsibility required of transplant patients for successful outcomes—has created an environment where transplant programs attempt to select candidates deemed to have the best chance of optimal posttransplant outcomes. Due to these issues, psychosomatic medicine specialists and other mental health professionals are frequently involved in the evaluation, selection, treatment, and long-term care of transplant candidates and recipients. Knowledge of the organ-specific disease process is important, but one should be additionally familiar with transplant-specific issues in order to care for these complex and unique patients. This section will review the essential elements of adult solid organ transplantation, including the epidemiology of transplantation, the common types of organ disease leading to transplantation, and the course and prognosis for the transplant patient. Within the course and prognosis sections, the emphasis will be on the assessment of candidates and the specific phases of the transplantation process, as these areas must be understood in order for the mental health professional to provide appropriate patient care. In addition to the more common psychological factors and psychiatric disorders affecting the process and outcomes of organ transplantation, this section will also review issues particular to transplantation, including bridge-to-transplant technology, living organ donation, treatment issues specific to transplantation, and immunosuppressive medication neuropsychiatric side effects.

ORGAN TRANSPLANTATION

Definition and Comparative Nosology

Organ transplantation includes most of the solid organs either singly or in combination: kidney, pancreas, kidney–pancreas, liver, heart, lung, heart–lung, intestine, and multivisceral (e.g., multiple abdominal organs or heart–liver). A new area of organ transplantation recognized by the U.S. Department of Health and Human Services in 2014 is vascular composite allograft (VCA) transplantation. VCAs are a heterogeneous category of “organs” that include the face, limb, long bones, and soft tissues. Because they are rare in the United States, the remainder of this section will focus on other types of organ transplantation.

Transplantation can be performed with deceased donor organs, using either complete or partial organs (i.e., split liver, single lung, partial intestine, or pancreatic islet cells), or can otherwise involve a living donor who provides either a complete organ (i.e., single kidney) or partial organ (i.e., portions of the liver and intestine). For the past 10 years, living donors have accounted for 18% to 20% of all transplants performed. Even a heart can be donated through a domino procedure in which a recipient who requires a heart–lung transplant but has a sufficiently healthy heart can donate their heart to another. Although rarely performed in the United States, domino hearts have been predominantly from patients with cystic fibrosis or pulmonary hypertension. Whole-organ xenotransplantation has been unsuccessfully attempted in the past using pig and baboon organs. Currently, only pig islet cells are being investigated for transplantation in patients with pancreatic disease. However, xenogeneic rejection due to immune incompatibility and concerns over the risk for cross-species infections remain significant challenges.

Epidemiology

The greatest obstacle to receiving a transplant is the shortage of donated organs. The number of wait-listed individuals greatly exceeds the availability of donated organs. However, there has been some increase in organ availability as well. In 2003, the Department of Health and Human Services created the Organ Donation Breakthrough Collaborative designed to replicate and promote best practices by national organ procurement organizations for deceased donor consent that would increase access to available organs. The efforts of the collaborative are credited with an estimated 4,000 additional annual transplants per year. Unfortunately, the increasing number of deaths from the opioid crisis has also been credited with recent increases in organ donation. In addition, due to the ability to manage donor-derived infections, organ utilization has increased to include the use of less-than-perfect organs (i.e., those coming from a high-risk profile or Hepatitis C virus [HCV]-infected donors). A transplant candidate has to opt-in to accept such an organ. Over 50% of transplant candidates are willing to consider such organs. The limited organ availability has also increasingly led to the consideration and use of living organ donors. In kidney donation, to deal with incompatible living donor–recipient pairs, paired donation exchanges and, later in 2005, domino chains (beginning with a nondirected donor who initiates a sequence of kidney transplants) using computer algorithms to match successive donor–recipient pairs were initiated to efficiently maximize the available living donor pool. While performed for nearly a decade in Asia, where the living liver donor experience is widespread, paired exchange liver donor transplants were not done in North America until 2019. However, while these efforts have increased the numbers of donor organs, the numbers of wait-listed patients continue to steadily rise while the number of transplants performed each year is fairly stable (Fig. 27.13–1). Although the average amount of time on the waitlist depends on the type of organ required, the severity of illness, and the blood type of the candidate, for all major organ types, over 40% of U.S. wait-listed candidates waited 2 years or more for an organ. Unfortunately, 10% to 18% are removed (die or are too ill for the transplant) while on the waitlist.

Following transplantation, recipients of living kidney and living liver grafts show the highest long-term survival rates 10 years posttransplantation (79% and 74%, respectively) (Fig. 27.13–2). Recipients of deceased donor kidneys, deceased donor livers, and heart recipients have somewhat lower 10-year survival (62%, 63%, and 64%, respectively), and lung and intestine recipients have the poorest 10-year survival (33% and 45%, respectively). Even so, these data are from transplants performed over 10 years ago, and with advances in technology, immunosuppression, and medical care, the survival rates continue to improve. Graft survival rates are generally lower than patient survival rates, indicating that many transplant recipients eventually require retransplantation. For example, non–high-risk kidneys are expected to function between 9 to 12 years. In addition, many immunosuppressive medications are nephrotoxic, and liver, heart, lung, and intestine transplant recipients may eventually develop renal failure requiring kidney transplantation.

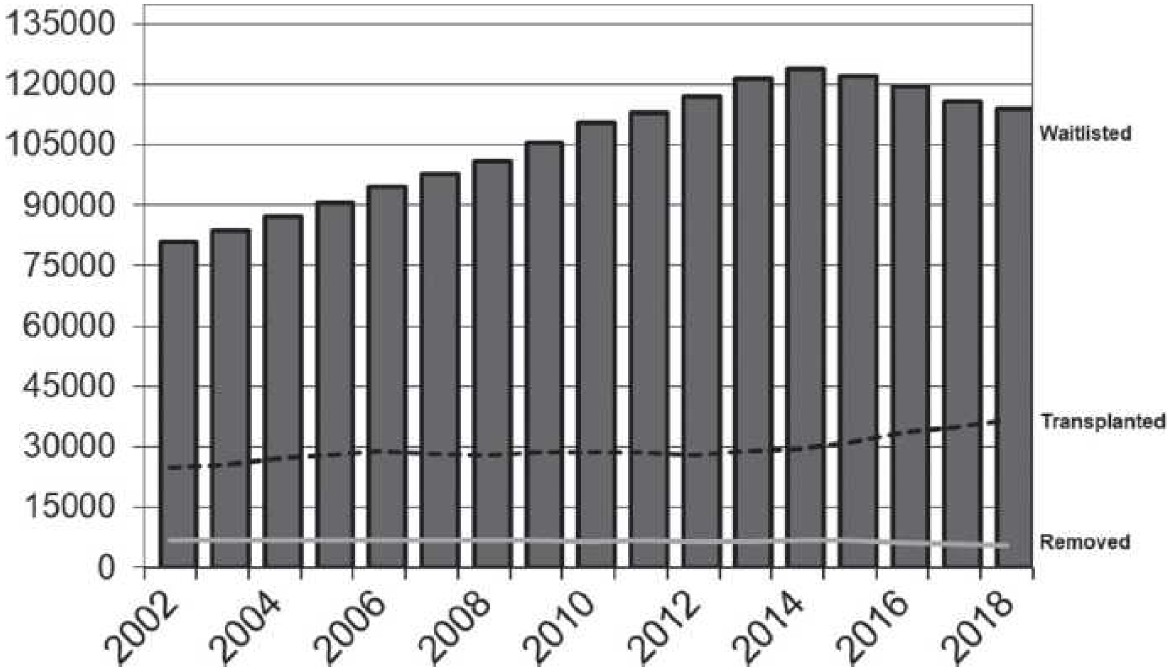

FIGURE 27.13–1. Waiting list statistics in the United States: 2002 to 2018. (Organ Procurement and Transplantation Network [OPTN] and Scientific Registry of Transplant Recipients [SRTR]. OPTN/SRTR 2021 Annual Data Report. 2023. Accessed June 2, 2023. http://srtr.transplant.hrsa.gov/annual_reports/Default.aspx.)

Etiology

The diseases that typically result in solid organ transplantation are listed in Table 27.13–1 by organ type and frequency. Advanced organ disease can occur acutely as in fulminant hepatic failure (e.g., acetaminophen overdose, acute viral hepatitis) or can be a chronic, progressive disease process (e.g., primary sclerosing cholangitis, chronic obstructive pulmonary disease). The disease could involve the organ parenchyma (e.g., diabetic nephritis), the vascular flow to the organ (e.g., Budd–Chiari syndrome), or transplantation could be required due to cancer in the organ (e.g., hepatocellular or renal cell carcinoma) or traumatic injury (e.g., blunt force trauma, inhalation/burns injury). In liver transplantation, recent shifts in the epidemiology of liver diseases, from an increase in the prevalence of nonalcoholic fatty liver disease (often from morbid obesity) and curative treatments for HCV (often incurred through injected drug use), are changing the referral diseases and reasons for pretransplant psychiatric consultation.

Course and Prognosis

Transplantation can be envisioned as a continuum demarcated by specific phases. The development of advanced organ disease heralds the beginning of the process. From evaluation at the transplant program to waiting on the national list, then early and longer-term recovery and adaptation to life as an organ recipient, each phase is associated with different stressors and requires different skills and resources from patients, caregivers, and mental health clinicians caring for them (Fig. 27.13–3). Mental health professionals are perhaps most frequently called on to aid in the evaluation of the prospective transplant candidate. Because these evaluations require a focus on transplant-specific areas somewhat different from a typical psychiatric consultation, they are addressed in detail below.

Clinical Evaluation for Transplantation. Due to the limited availability of donor organs, the considerable medical, social, and financial resources necessary for successful transplantation, the degree of personal responsibility of recipients, and the needed involvement of their family and other caregivers, careful assessment of transplant candidates and their caregivers is required. Pretransplant evaluations typically involve two components: a medical and a psychosocial assessment. During the medical assessment, patients are evaluated to determine whether there is a clear physiologic indication for transplantation based on OPTN criteria and to rule out significant medical comorbidities that pose additional risks or diminish the possibility of successful transplantation. For example, concurrent diabetes mellitus, morbid obesity, or a history of malignancy are conditions that can considerably raise the risk for posttransplant morbidity and mortality. Whether these conditions or others make a patient ineligible for transplantation will partly depend on the severity of the comorbid diseases but also on the transplant team's level of experience, ability, and comfort in working with high-risk patients.

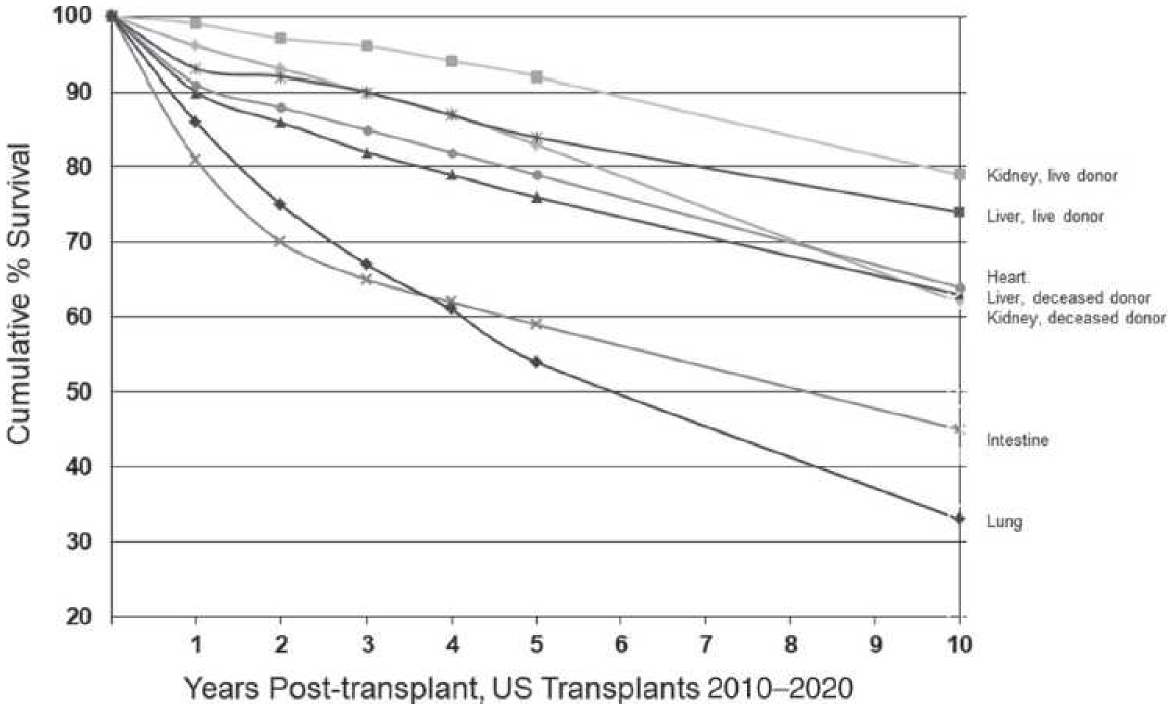

FIGURE 27.13–2. Transplant patient survival by organ type. (Organ Procurement and Transplantation Network [OPTN] and Scientific Registry of Transplant Recipients [SRTR]. OPTN/SRTR 2021 Annual Data Report. 2023. Accessed June 2, 2023. http://srtr.transplant.hrsa.gov/annual_reports/Default.aspx)

Table 27.13–1.

Diseases Resulting in Listing for Transplantation in the United States as of December 2019

Organ Type	Percentage	Organ Type	Percentage
Kidney, *n* = 101,337		**Liver, *n* = 12,555**	
Diabetes mellitus type I and II	38%	Alcoholic	31%
Hypertensive nephrosclerosis and malignant hypertension	21%	Hepatitis C	14%
Chronic glomerulonephritis	14%	Cholestatic disease	9%
Cystic kidney disease	9%	Acute liver failure	1%
Other	18%	Other[a]	35%
Heart, *n* = 3,385		**Lung, *n* = 1,423**	
Cardiomyopathy	59%	Obstructive lung disease (e.g., COPD/emphysema)	38%
Coronary artery disease	29%	Restrictive lung disease	44%
Congenital	6%	Pulmonary vascular disease	11%
Valvular disease	1%	Cystic fibrosis and immunodeficiency disorders	7%
Other	5%		
Intestine, *n* = 226		**Pancreas or Kidney Pancreas, *n* = 2,372**	
Other short gut syndrome	27%	Diabetes mellitus	
Congenital short gut syndrome	21%	Type I	76%
Necrotizing	10%	Type II	16%

enterocolitis			
Pseudo-obstruction	7%	Other	8%
Enteropathies	3%		
Other	32%		

[a] Other category presumed mostly nonalcoholic fatty liver disease according to 2021 OPTN annual report.

Organ Procurement and Transplantation Network (OPTN) and Scientific Registry of Transplant Recipients (SRTR). OPTN/SRTR 2021 Annual Data Report. 2023. Accessed June 2, 2023. http://srtr.transplant.hrsa.gov/annual_reports/Default.aspx.

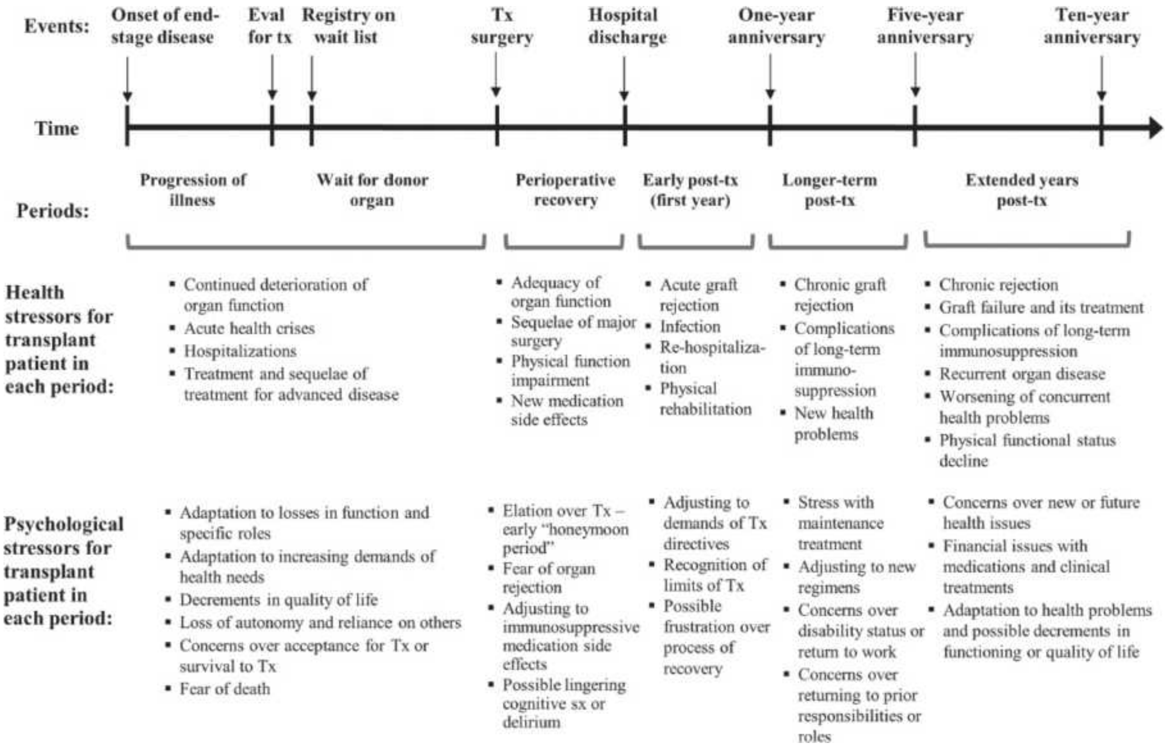
Events:
Onset of end-stage disease
Eval for tx
Registry on wait list
Tx surgery
Hospital discharge
One-year anniversary
Five-year anniversary
Ten-year anniversary
Time
Periods:
Progression of illness
Wait for donor organ
Perioperative recovery
Early post-tx (first year)
Longer-term post-tx
Extended years post-tx
Health stressors for transplant patient in each period:
• Continued deterioration of organ function
• Acute health crises
• Hospitalizations
• Treatment and sequelae of treatment for advanced disease
• Adequacy of organ function
• Sequelae of major surgery
• Physical function impairment
• New medication side effects
• Acute graft rejection
• Infection
• Re-hospitalization
• Physical rehabilitation
• Chronic graft rejection
• Complications of long-term immuno-suppression
• New health problems
• Chronic rejection
• Graft failure and its treatment
• Complications of long-term immunosuppression
• Recurrent organ disease
• Worsening of concurrent health problems
• Physical functional status decline
Psychological stressors for transplant patient in each period:
• Adaptation to losses in function and specific roles
• Adaptation to increasing demands of health needs
• Decrements in quality of life
• Loss of autonomy and reliance on others
• Concerns over acceptance for Tx or survival to Tx
• Fear of death
• Elation over Tx – early "honeymoon period"
• Fear of organ rejection
• Adjusting to immunosuppressive medication side effects
• Possible lingering cognitive sx or delirium
• Adjusting to demands of Tx directives
• Recognition of limits of Tx
• Possible frustration over process of recovery
• Stress with maintenance treatment
• Adjusting to new regimens
• Concerns over disability status or return to work
• Concerns over returning to prior responsibilities or roles
• Concerns over new or future health issues
• Financial issues with medications and clinical treatments
• Adaptation to health problems and possible decrements in functioning or quality of life

FIGURE 27.13–3. Organ transplant timeline: Critical events, time periods, and health and psychosocial issues. **Tx**, transplant; **QOL**, quality of life. (Adapted from Dew MA, Manzetti J, Goycoolea JR, et al. Psychosocial aspects of transplantation. In: Medscape. ***Organ Transplantation: Concepts, Issues, Practice, and Outcomes***. © 2002, WebMD, Inc. https://www.medscape.com/viewarticle/436541.)

Table 27.13–2.
Goals of the Pretransplant Psychosocial Evaluation

1. Foster a working relationship between patients and transplant team
2. Educate and prepare patients for transplantation
3. Reinforce the informed consent process for patients
4. Evaluate patients' coping strengths and weaknesses/vulnerabilities
5. Assess treatment adherence and ability to make lifestyle/behavior changes
6. Diagnose and recommend treatment options for active psychiatric disorders
7. Assess amount, stability, and availability of social support system
8. Develop individualized treatment plans for patients including requirements for listing
9. Provide input to team regarding transplant candidacy

RELATIVE IMPORTANCE OF PSYCHOSOCIAL CRITERIA. Although there is general consensus about the medical assessment and the criteria used for transplant candidacy, there are considerable differences among teams regarding the purpose of psychosocial evaluations, who performs these evaluations, the content of the evaluation, and the use of the psychosocial information obtained. This may be because transplant teams have differing opinions about the relevance and significance of psychosocial criteria. In addition, there are no national psychosocial criteria for candidate listing in the United States; although there are published consensus guidelines on the basic components of the psychosocial evaluation (see Tables 27.13–2 and 27.13–3 for the goals and content of the psychosocial evaluation). As a result, each transplant program must decide on the pertinent psychosocial criteria and the relevance of these criteria for individual patients, often with input from mental health professionals. In general, heart transplant teams are more likely to view psychosocial issues as potential contraindications, while kidney programs tend to be the most lenient. This difference may be partly due to the complexities of a heart transplant with poorer survival than in a kidney transplant combined with the greater availability of kidneys for transplantation and favorable cost benefits of kidney transplantation compared to dialysis. Differences in approaches toward substance use and addictive disorders and whether a specified period of abstinence is required prior to evaluation or listing vary substantially, and not all substances are considered equally across different types of organ transplant programs.

Table 27.13–3.

Components of a Comprehensive Pretransplant Psychosocial Evaluation

Current medical history and knowledge of current illness	Course of advanced organ disease (acute vs. chronic), prior hospitalizations, treatments, complications, level of functioning/disability, quality of life, current symptoms, understanding of illness and limits of currently received treatments
Attitude toward transplantation	Circumstances leading to referral for transplantation, understanding of reasons for referral, receptiveness to transplant, prior experience or knowledge about transplantation, concerns/fears/reservations, expectations about transplantation
Knowledge about transplantation	Knowledge about risks and benefits of treatment options, given patients severity of illness; understanding the reasons that transplantation is being considered; e.g., patient’s initial education about the pre- to posttransplant process, additional research/reading of their own on transplant, understanding of the risks and benefits involved with transplant, understanding of the long-term commitment to ongoing medical follow-up, immunosuppressants, monitoring, and lifestyle/behavioral changes required
Treatment adherence	Past and current medical adherence; knowledge about current treatment regimen and willingness to adherence to the posttransplant regimen; e.g., self-monitoring (e.g., checking blood sugar, blood pressure), diagnostic testing, other

	treatment interventions (e.g., operations, supplemental oxygen, pulmonary rehabilitation), diet, fluid restrictions, exercise, abstinence from alcohol/drugs/tobacco
Coping with illness	Response to illness and coping skills and strategies used to manage the illness and its daily impact; expression of acceptance or denial about prognosis and treatment options; e.g., utilization of active vs. passive coping skills for dealing with his or her illness (e.g., self-education, adjusts activity level vs. avoidance, heavy reliance on spouse, etc.), expressions of distress or feeling overwhelmed with his or her illness, methods for managing these feelings
Past medical history	Prior illnesses, hospitalizations, operations, experiences with health care system
Psychiatric history	Past and current history of mood disorders, anxiety disorders, or other disorders symptom severity, course of symptoms, chronicity of symptoms, receipt and response to past or ongoing treatment and willingness to undertake additional treatment hospitalizations, treatment adherence, thoughts of suicide or self-injury or prior suicide attempts
Substance use/abuse	Drugs/alcohol/tobacco frequency of use, amount, duration, diagnosable disorder, level of impairment to health/work/relationships, legal issues, knowledge of and insight into their addiction, commitment and

	ability to remain abstinent including prior attempts to reduce usage or prior periods of abstinence, prior and number of attempts at formal rehabilitation/counseling, attendance at 12-step meetings, sponsors, willingness to get therapy when needed, supports for abstinence, family history of substance abuse
Current medications	Knowledge of current medications including dose and frequency, understanding the purpose of each medication
Family history	Current and past history of medical, psychiatric, and substance abuse disorders
Social history	Developmental history, childhood experiences, educational background, literacy and health literacy, history of trauma, significant relationships, occupation/work experience, current living arrangements, financial status, dependents, concurrent life stressors (e.g., family dysfunction), cultural background including religious/faith practices, legal issues
Support system	Availability, stability, willingness and extent of support available to help care for patient before and after transplantation, including family members, friends, neighbors, religious organizations, and coworkers, understanding of these individuals about illness and treatment options and their role in assisting patient

Mental status examination and capacity	Level of consciousness, appearance, behavior, speech, presence of psychomotor agitation or retardation, mood, affect, thought process/content, insight into organ disease/transplant, judgment, attention, concentration, memory, and executive functioning, capacity to comprehend information and engage in decision making

Adapted with permission of Springer Nature. Dew MA, DiMartini AF, Dobbels F, et al. The approach to the psychosocial evaluation of cardiac transplant and mechanical circulatory support candidates. ***Curr Heart Fail Rep.*** 2019;16(6):201–211. doi: 10.1007/s11897-019-00443-0.

UNIQUE ROLE OF THE PSYCHOSOCIAL EVALUATOR. Due to these differences, the mental health professional must be knowledgeable about their team's approach to psychosocial issues and may need to educate the team on the prognosis and treatment of psychiatric disorders. Thus, mental health professionals have dual functions in the transplant evaluation process, serving the needs of both the transplant team and the transplant patient. Balancing these two roles may be complicated, but the mental health professional can both advocate for the patient and communicate and interpret the transplant team's requirements back to the patient. Information gathered during the psychosocial evaluation is essential to the transplant team's complete understanding of the patient. The communication back to the patient provides education on specific transplant-related expectations, the team's requirements for listing the patient as a transplant candidate, and the importance of these requirements to the success of transplantation. Assisting with an appropriate referral for recommended psychiatric or psychological treatment can be provided along with ongoing monitoring of the patient's progress in this treatment arena. In these ways, the mental health professional can facilitate the development of a long-term collaborative relationship between the patient and the transplant team.

SCREENING AND CONTENT OF THE PSYCHOSOCIAL EVALUATION. At most centers, psychosocial evaluations are performed by clinical social workers. Typically, this evaluation serves as a screening evaluation, with a more comprehensive psychiatric assessment requested for patients identified as having prior or current psychiatric histories or psychosocial issues requiring additional attention. This two-stage approach can be an efficient use of resources because the sheer number of candidates may preclude a comprehensive psychiatric evaluation of all patients, and additionally, with thorough screening, an evaluation may not be indicated.

A comprehensive pretransplant psychosocial evaluation contains all the areas of a typical comprehensive psychiatric assessment, with the addition of transplant-specific areas (e.g., coping with illness, adherence, and other health or illness behaviors) (Tables 27.13–2 and 27.13–3). Standardized transplant-specific assessment tools (i.e., Psychosocial Assessment of Candidates for Transplantation, Transplant Evaluation Rating Scale, Stanford Integrated Psychosocial Assessment for Transplantation, and Structured Interview for Renal Transplantation) can be used alongside the clinical interview to ensure that relevant topics are covered. These instruments should be completed by a trained clinician and provide a useful organization of information pertinent to transplantation. Non–transplant-specific instruments assessing psychiatric symptoms or cognitive status are sometimes used in the evaluation. Collateral information from family members, partners, medical and pharmacy records, toxicology screening, attendance at clinic appointments, and current or past health care providers to determine adherence and treatment engagement is often essential to substantiate or clarify information obtained from patient interviews. If a candidate has undergone a prior transplant evaluation, this information should be reviewed. Expanded medical record review is being facilitated by the increasing use of electronic medical records connecting both intra- and intermedical system data.

SPECIAL AREAS OF TRANSPLANT ASSESSMENT

Social Supports. Assessment of the presence of an adequate, stable, and available support system to help the patient through transplantation is another area requiring careful evaluation. Patients often require assistance with activities of daily living, medications, dressing changes, and transportation both before and after transplantation. Particularly while awaiting a transplant, some patients may be too ill to physically care for themselves. Others may develop cognitive impairments that require constant monitoring for safety. Caregivers often feel the need to present themselves to the transplant team as being a capable and concerned support system for the patient. However, the best evidence of adequate support is consistent behavior demonstrating this ability. Thus, the patient's caregivers should be present during the evaluation process so that their knowledge, ability (cognitive and physical), and willingness to take on these responsibilities can be assessed.

Treatment Adherence. Nonadherence to the posttransplant regimen and immunosuppressive medications, in particular, has been associated with poor clinical outcomes, contributing substantially to late acute rejection and graft loss. For these reasons, transplant teams are highly invested in a patient's ability to adhere to the long-term medical, behavioral, and lifestyle requirements necessary to optimize outcomes and survival. Following transplantation, patients must take daily medications; monitor critical parameters such as vital signs, blood sugar, and pulmonary functions; come to clinic appointments; undergo diagnostic tests; follow dietary restrictions; participate in physical rehabilitation or exercise programs; and adhere to restrictions on the use of illicit drugs, tobacco, and often alcohol as well. Lifelong adherence to immunosuppressants is necessary, as is adherence to other medications such as antihypertensives, lipid-lowering drugs, hypoglycemic agents, antibiotics, and antifungals. Despite this, immunosuppressant nonadherence alone is well recognized as a common cause of late-onset acute rejection, antibody-mediated rejection, graft dysfunction, and ultimately, transplant loss. Furthermore, a comprehensive meta-analysis of adherence studies across all organ types identified a rate of 23 cases of immunosuppressant nonadherence per 100 transplant recipients per year. Rates of nonadherence to other posttransplant requirements vary, with higher rates for home self-care behaviors (e.g., diet, exercise, home monitoring of vital signs) and lower rates of nonadherence in the areas of attending clinic appointments and use of restricted or banned substances.

Identification of the reasons and risk factors for nonadherence is one goal of the pretransplant evaluation. Pretransplant factors such as a history of nonadherence, being an adolescent or young adult, substance misuse, poor perception of health, and poor social supports appear to be most commonly associated with posttransplant nonadherence. In addition to commonly assumed behavioral reasons for poor adherence (e.g., poor organizational skills, failure to accept personal responsibility for one’s health, etc.), other social issues such as type of insurance coverage or transportation availability or even ambivalence about treatment should be considered. These factors and their impact on current adherence should be assessed during the psychosocial evaluation. Patient education on the importance of adherence and plans for pretransplant monitoring may be required. In addition, a posttransplant plan for monitoring and education should be considered because patient adherence often worsens with time since transplantation.

Substance Use. Substance use and substance use disorder is an area of special concern for transplant teams, especially in liver and heart transplantation, where the use of drugs or alcohol often leads to the need for transplantation. Inquiries about substance use or substance use disorder should be made with all patients, regardless of the primary cause of their advanced organ disease, and transplant programs may perform toxicology screening on all candidates at evaluation. Substance use disorder may not be assessed by other transplant team members for patients whose organ failure was not directly linked to alcohol or drugs, and the mental health professional may be relied on to address this area (see Table 27.13–3 for areas of substance use assessment). In the United States, cannabinoid use has increased with the passing of medical cannabis prescribing laws and legalization in many states. While smoked and edible cannabis has known health risks, there is no consensus in the transplant community regarding medical or recreational cannabis. There are also no guidelines for the use of electronic cigarettes containing tetrahydrocannabinol though their use has been linked to lung injuries and subsequent transplant. The issue of other medicinal or recreational cannabinoid use, like cannabidiol, is even more unclear. Many transplant programs exclude patients who are actively using recreational cannabis. In addition to drugs and alcohol, the cessation of tobacco use (both smoked and smokeless) has been a requirement for heart and lung transplant listing and is increasingly seen as a contraindication to listing for abdominal transplantation as well (see tobacco section below). There is no consensus on the use of nicotine-containing electronic cigarettes. Some programs may not allow listing while a patient is using nicotine replacement therapy for smoking cessation, so the specific allowances of the transplant team must be known. Transplant programs attempt to select patients deemed unlikely to relapse posttransplant due to concerns that substance use will diminish patient or graft survival (e.g., patients and transplanted organs at risk from direct toxicity of the substance, increased risk of infection due to risky behaviors, or increased morbidities) or that substance use relapse would lead patients to become nonadherent (see substance use outcomes below).

Informed Consent in Terminally Ill Transplant Candidates. The importance of informed consent for organ transplantation cannot be overemphasized, given the need for patient commitment to long-term responsibilities along with the potentially serious risks and complications that can arise in both the early and late term following transplantation. Psychosocial evaluations help to address how well patients understand their illness, prognosis, treatment options, and the transplant process itself, thereby facilitating the informed consent process. However, evaluating patients' understanding of their options may be problematic under the circumstances in which transplant patients are evaluated. Often, patients with advanced organ disease referred for transplant evaluation are facing life-or-death decisions, with transplantation being the only means to extend their survival. For other patients, alternative treatments such as hemodialysis have become too difficult to tolerate, or their quality of life has become unacceptable. Whether their circumstances are truly desperate or only perceived to be so, patients may have limited ability to fully appreciate, comprehend, and consider all aspects of transplantation. Their illness may have resulted in acute or chronic cognitive impairments that also reduce their ability to comprehend and make informed decisions (see additional discussion of impairments below). Given these factors, a careful psychosocial assessment with attention to informed consent should be performed, balanced with realistic expectations about a patient's behavior under such emotionally difficult conditions. The evaluation can identify areas in which the provision of additional information on risks versus benefits of transplantation is required before an informed decision can truly be made by the patient.

Emergency Evaluations. Requests for emergency psychosocial transplant evaluations may occur for patients with fulminant liver failure or acute ischemic or idiopathic cardiomyopathy who have not been previously evaluated by the transplant team. These patients may be stuporous, on mechanical ventilation, or be in a primary or drug-induced coma. Under these circumstances, the majority, if not all the pretransplant evaluation information gathered will depend on thorough interviews with family, friends, and health care providers and attempts to obtain previous medical and psychiatric records. The informed consent process remains a concern, but the focus will be on the ability of the family or surrogate decision maker to understand the patient's medical condition and the transplant process and maintain respect for the patient's wishes. Establishing whether a stable support system exists for these patients is especially important as neither patient nor family may have had the opportunity to prepare themselves for transplant or the posttransplant regimen.

Patients with fulminant hepatic failure due to acetaminophen-induced liver injury require special evaluation for psychopathology. Determining whether the ingestion was a mistake made by a patient attempting to treat pain or cold symptoms or an intentional act of suicide will lead to different recommendations and care plans. As with any suicide attempt, the nature of the attempt (e.g., impulsive vs. planned), current stressors, presence of active psychiatric illness, and other risk factors for subsequent suicide attempts or gestures need to be assessed. Whether specific patient characteristics could interfere with the patient's ability to develop a long-term collaborative relationship with the transplant team and consistently follow medical directives must be considered.

Living Donors. As noted earlier, organ transplantation increasingly often relies on living donation, particularly in kidney transplantation. Donor safety is of utmost concern, given the altruistic nature of living donation and the fact that donors have no medical need for surgery. As donation has expanded beyond biologically or emotionally close relatives of transplant patients to donors with little if any pre-existing relationship to the recipient, the medical and psychosocial evaluation of prospective donors has become more extensive. Currently, the evaluation process at U.S. centers is specified by OPTN policy, which includes requirements for the donor psychosocial evaluation and components that must be reviewed (see https://optn.transplant.hrsa.gov/media/1200/optn_policies.pdf). The evaluation is as detailed and exhaustive in prospective donors as in prospective transplant candidates (Table 27.13–4). Mental health professionals must assess the psychiatric history of prospective donors and determine their current level of functioning and their cognitive capacity to make an informed decision about donating. Consensus opinions and both OPTN and Centers for Medicare and Medicaid Services policies also require that all potential living donors be provided with an independent living donor advocate (ILDA) who is not responsible for the transplant candidate’s care or the decision to list the candidate for a transplant. The ILDA also evaluates the donor, assists them in gaining additional education about possible donation, and advocates on their behalf.

Table 27.13–4.
Areas of Assessment for Living Donor Evaluation

- Relationship between donor and recipient
- Reasons for donation
- Donor’s knowledge about the surgery including:
  - Donation and transplantation procedures
  - Risks to self
  - Risks to recipient
  - Expectations about donation/recovery
- Motivations including:
  - Urgency of the medical condition of recipient
  - Lack of other donors
  - Guilt
  - Psychological expectations (e.g., desire to be hero, special, etc.)
  - Expectations about postdonation relationship with recipient
  - Ambivalence
  - Evidence of coercion/inducement
  - Attitudes of significant others and family toward the donation
  - Availability of support
  - Financial resources and donor’s understanding of financial implications of donation
  - Work- and/or school-related issues (if applicable)
  - Timing of the donation relative to work, school, or other important life events
- Donor’s psychological health and history, including the following:
  - Psychiatric disorders
  - Personality disorders
  - Cognitive capacity
  - Coping resources/style
  - Pain syndromes
  - Prior psychological trauma/abuse
  - Substance use
  - Presence of any behaviors that may increase risk for disease transmission to the organ recipient

Need for any educational or therapeutic interventions before a decision can be made regarding suitability to donate.

Adapted with permission from Levenson JL. ***The American Psychiatric Publishing Textbook of Psychosomatic Medicine***. American Psychiatric Publishing; 2005. All Rights Reserved and from OPTN Policy 14 (http://optn.transplant.hrsa.gov/governance/policies/).

Under most circumstances, current active psychiatric or substance use disorders are contraindications to donation. However, exceptions may arise under certain circumstances, such as a parent donating to a critically ill child. Therefore, psychiatric management of prospective donors and follow-up to continue psychiatric treatment after donation may be required.

Outcomes following living donation are given special considerations later in this section.

A 30-year-old woman seeks to be evaluated as a living–related kidney donor for her older brother, who has advanced renal failure secondary to diabetes mellitus. She states that she first learned about her brother's need for transplantation 6 months ago when she was contacted by their mother. Before this, she had been estranged from the family for several years due to their disapproval of her lifestyle, which included drug abuse and unstable relationships with various men. She reports that she has tried to re-establish ties with her family before, but they have not been receptive to her efforts even when she has tried to show them she is leading a more stable lifestyle. When asked about her motives for wanting to donate to a brother with whom she has had no contact for years, she emphasizes that no one is pressuring her to donate. However, she admits feeling that her brother has been more responsible than she has been and a "model son" when it came to helping their parents through difficult financial times. She expresses hope that donating a kidney will help her family to see that she is now acting like a responsible adult and family member. In this case, the evaluator should be concerned over her motives for desiring to donate. Issues to explore would include her expectations of her relationship with her family and especially her brother following donation, feelings of guilt, and a desire to restore her status in the family.

COMMUNICATION REGARDING TRANSPLANT CANDIDATE AND LIVING DONOR EVALUATION RESULTS. Communication about patient suitability for transplantation or living donation and recommendations to the team should be plainly stated with minimal psychiatric jargon, providing straightforward guidance. However, caution must be taken regarding comments on a patient's suitability for transplant, partly because there is a lack of consensus on what psychosocial issues would make a patient unacceptable for transplant and because third-party payors may review the psychosocial assessment as part of their process to determine whether to financially approve a patient for transplant coverage. In some cases, health insurers may use information from the psychosocial evaluation to set forth more restrictive psychosocial requirements for the patient to fulfill before they agree to cover transplantation costs. Fortunately, most transplant teams do not reject candidates outright but allow an opportunity for psychosocial issues to be addressed or resolved, followed by a re-evaluation for candidacy. Areas of preparation that are of particular importance to transplantation are addressed below.

Preparation for Transplantation. In addition to addressing the transplant team's concerns about psychiatric or psychosocial issues and aiding in candidate eligibility determinations, the pretransplant psychosocial evaluation can be used to develop an individualized treatment plan to educate and prepare candidates for transplantation and optimize posttransplant outcomes (Table 27.13–2). For example, the concerns, fears, or even ambivalence about the surgery that patients express during the evaluation can be useful to allow the evaluator to target appropriate interventions. Optimally, candidates should be referred for any psychotherapeutic interventions early in their illness before medical or cognitive deterioration (e.g., encephalopathy) that will limit their ability to participate in or benefit from such treatment.

A 54-year-old woman with familial cardiomyopathy is referred for cardiac transplantation. Despite agreeing with her cardiologist about his referral, she fails to contact the transplant center until he reminds her during a follow-up appointment. When she begins her evaluation, she cancels some of the appointments and tests she is asked to complete. The transplant coordinator asks about these cancellations and the patient states they were due to scheduling difficulties at her job. When seen by the team's social worker, she admits to being hesitant about transplant due to her cousin's experience several years prior. Her cousin developed recurrent episodes of rejection and died 5 years after her transplant. The patient is afraid the same fate awaits her and only agreed to an evaluation upon her family's insistence. Once her fears are uncovered, the transplant team assures her that she would not be pressured to be listed for transplant. The coordinator meets with the patient and her family to better educate them about transplantation. During this meeting, the patient is able to gather information about her specific situation, the risks of surgery, and life as a transplant recipient. The patient also meets with the team's psychiatrist for several sessions to address her fears in greater depth and to reinforce her coping skills. Her anxiety lessens over time, and she eventually decides she wants to pursue transplantation.

BEHAVIORAL AND CONTINGENCY CONTRACTING. For patients with significant psychosocial risk factors such as substance abuse, treatment nonadherence, or other psychiatric disorders requiring ongoing treatment, transplant teams may decide to use behavioral contracts to stipulate requirements for listing or monitoring during the waiting period for transplantation. These contracts can positively encourage behavioral change, minimize miscommunication, foster collaboration between patient and team, and reinforce the need for patients to take active responsibility for their health care. Although not a legal document, behavioral contracts clearly define what is expected of patients, how their adherence to the contract will be assessed, and what consequences they will face if they cooperate with or fail to cooperate with listed expectations. Successful adherence to these requirements would result in listing for transplantation; nonadherence would lead to removal from the waiting list. For example, patients with psychiatric disorders may be required to receive ongoing outpatient psychiatric care and be periodically reassessed by the team to determine whether the psychiatric disorder is under adequate control. Patients with a history of treatment nonadherence may be asked to demonstrate adherence to various aspects of their care (e.g., medications, appointments, tests, dialysis, dietary, or fluid restrictions) for a period of time prior to listing and continue to be monitored during the waiting period leading up to transplantation.

A 58-year-old man is referred for cardiac transplantation due to a long history of ischemic cardiomyopathy. He suffered his first myocardial infarction in his mid-30s but was able to return to an active lifestyle after several months. He followed up with a cardiologist for a few years but eventually stopped as he was feeling well, had stopped smoking, and was exercising regularly and watching his diet. Despite his efforts, he suffered a second infarction in his early 40s and a third at age 50. Since his last infarction, he has developed steadily worsening symptoms of congestive heart failure. He has followed up with a cardiologist regularly and is on medications, but he argues with his physician whenever a new medication is added, or a dose increase is advised. He has occasionally tapered the doses of some of his medicines to see if he could get by with less. He complains that each medication adds new side effects and he hates being a "walking pharmacy." He is very interested in transplantation as he feels his quality of life is poor, but he knows he will have to be on a large number of medications posttransplant. In preparation for transplant listing, he is required to demonstrate consistent medication compliance and begin psychotherapy. In counseling, he is able to identify previously unresolved issues about his vulnerability, lack of control over his condition, and anger toward health care providers as they remind him of his debility. He is able to address these issues and develops a realistic approach to his health and health care.

SUBSTANCE USE—DISORDERED PATIENT. Requirements for the length of sobriety and addiction rehabilitation vary greatly from program to program and among different organ types even within a transplant program. Although stable abstinence is measured in years, not months, many liver transplant programs—for example—focus on the length of abstinence as the main criterion, with 6 months of abstinence being a benchmark for listing. Except for short abstinence, no single risk factor strongly predicts future use, although poor social support, family history of addiction, polysubstance use, and failed rehabilitation have been associated with increased risk. Addiction counseling before and after transplant remains the best treatment option for providing education about the illness and strategies for monitoring and avoiding relapse, lifestyle, and behavioral changes, as well as for identifying appropriate supports to maintain abstinence. For those with short abstinence (less than 1 year) and no prior therapy, addiction counseling is highly recommended as a prerequisite to transplant listing. Patients with some period of abstinence but no addiction rehabilitation may not perceive a need for addiction treatment. Many have stopped actively using substances due to being ill but have not incorporated the skills or strategies to stay sober and prepare for continued abstinence posttransplant. Motivational interviewing may be needed to assist those reluctant to engage in treatment. However, for some, outright mandating of addiction treatment as a requirement for transplant listing may be needed. The intensity of addiction treatment may need to be tailored or adjusted over time to the medical debility of the patient. An example is acute alcohol-related hepatitis, in which patients with very short abstinence (if any) experience a rapidly deteriorating course and programs must quickly consider whether to transplant such patients and how to address their addiction after transplant. Mental health professionals should be prepared to evaluate and provide treatment recommendations for such short-abstinence patients, keeping in mind that limited time and potential encephalopathy and/or physical debility could impede engagement in treatment and shorten the time frame to make listing decisions. As mentioned above, some centers use behavioral contracts stipulating the frequency and duration of addiction counseling. Others require ongoing, random toxicology screening to demonstrate that substances are not being used. Even after wait-listing, ongoing periodic clinical and toxicologic reassessment and monitoring are recommended because of the continued risk for relapse.

Transplant programs may also require patients on methadone maintenance therapy (MMT) to discontinue methadone either as a prerequisite for transplant listing or in the immediate postoperative period. However, the process of transplantation is very stressful for patients and is not an optimal time to attempt the additional challenge of a methadone taper. Although progressive medical illness may necessitate a decrease in the methadone dose, requiring a patient to discontinue MMT specifically for transplant listing could be considered unethical because a premature taper off methadone may precipitate a relapse, which would then prohibit the patient from transplantation. Patients who are taking buprenorphine may require supervised dose reductions preceding transplant to enable adequate perioperative pain management.

REASSESSMENT OF WAIT-LISTED CANDIDATES. Periodic psychosocial follow-up with wait-listed candidates, either in the hospital or outpatient setting, is desirable because the candidate’s status may change with respect to affective symptoms, stress, coping, cognitive functioning, adherence, substance use, social support, and financial issues. In some cases (i.e., resumption of substance use), reassessment may lead a candidate to be removed from the transplant list. More often, however, ongoing pretransplant reassessment not only provides continuity of care but can also improve the team’s understanding of the candidate, help identify problems that need to be resolved pretransplant, foster a collaborative relationship with the team, and enhance adherence. Some candidates may not return until the point of transplantation, and ongoing reassessment at the transplant program may not be geographically or financially feasible for the candidate and family. Appropriate psychiatric services should be identified for those at a distance from the transplant program if issues arise, and information on the treatment or resolution of the problem should be forwarded to the transplant team. Psychiatric clinicians and other physicians in the community, who may be providing services to transplant patients but are not familiar with transplantation, may benefit from a discussion with the psychosocial evaluator regarding the transplant team requirements for treatment goals or from the receipt of available written educational materials on transplantation.

Patient Adaptation through the Transplant Time Course

EVALUATION AND WAITING PERIOD. Patients face significant adaptive challenges as they proceed through the pre- to posttransplant course. Successful completion of the transplant evaluation and the individualized requirements to be accepted for listing on the national organ transplant candidate list are only the first hurdles. Many patients and their families perceive the waiting period to be even more psychologically stressful. This is especially the case if patients are newly adapting to their diagnosis of terminal advanced organ disease and considering for the first time the transplant surgery and future life as a transplant recipient. The patients' level of adjustment at the point of transplant evaluation may depend on how sick they are, how long they have lived with their illness (acute vs. chronic), and how long they have been anticipating (either positively or negatively) the issue of eventually requiring organ replacement. With chronic illness, there can be progressive debility, loss of vitality and physical and social functioning, changes in or loss of family or social roles, loss of independence, and financial hardships. Adapting to these changes can elicit anxiety, depression, avoidance, and denial and may require working through the grief of these losses. Unless candidates have a living donor and can therefore choose the timing of their surgery, they must endure the uncertainty of the wait time, the extent of their medical deterioration, and loss of functioning before the transplant surgery occurs. In addition, wait-listed candidates may develop medical contraindications to transplantation (i.e., infection, serious stroke, progressive organ dysfunction), and both patient and family should be made aware that their eligibility might change for many reasons over time. These inherent stresses in the transplant process will tax the character style, coping, and social support resources of patients and caregivers, and these stresses can increase over time. Psychological counseling for patients and families can aid in successfully negotiating these transitions and prepare them for either transplantation or death.

PERIOPERATIVE PERIOD. For patients fortunate enough to survive the waiting period and undergo transplantation, the immediate perioperative period is a time of recovery and adjustment to new challenges. In some cases, recovery will be rapid and uncomplicated, while in others, it may be prolonged and marked with multiple complications. Except for renal transplantation, patients begin their postoperative recovery in the intensive care unit. Common postoperative complications such as bleeding, infections, rejection episodes, immunosuppressive medication side effects, and perfusion injury to the donor organ delay recovery may be perceived as setbacks by the recipient and can lead to growing discouragement and frustration. Assisting patients through this period typically involves a combination of adequate pain control, education, reassurance, and careful monitoring for early psychiatric and psychological complications. Delirium is common during this period (see cognitive disorders and neuropsychiatric side effects of immunosuppressant sections below). Without complications, a typical posttransplant length of hospital stay can be as short as 3 to 4 days for kidney recipients, 7 to 10 days for liver and heart recipients, and approximately 2 weeks for lung recipients.

EARLY POSTTRANSPLANT ADAPTATION. The immediate postdischarge phase places tremendous burdens on recipients and their families and caregivers now responsible for following complex medical regimens posttransplantation. In addition to new combinations of multiple medications, patients and caregivers must attend clinic appointments and get laboratory tests frequently, sometimes several times a week for at least the first several months after hospital discharge. Complications that arise may necessitate additional procedures and tests at the hospital or emergency room. At home, patients and caregivers must monitor vital signs, follow specific diets, and adhere to restrictions on physical activity (e.g., lifting weight, driving) and on the use of alcohol, drugs, and tobacco. As recovery progresses, the regimen may become simpler, and clinic visits and laboratory tests less frequent, and some restrictions may be removed. Eventually, recipients are encouraged to return to physical activity and even employment if indicated.

TRANSPLANT-RELATED ILLNESS BEHAVIOR. Maladaptive illness behavior can interfere with successful psychological adaptation to the demands of transplantation. Experience working with adult transplant recipients suggests two major types of such illness behavior. First, for patients with a long history of living with chronic disease, transplantation may represent a transition from one state of illness to another, with the transplant process being a familiar part of life as a patient. However, some chronically ill patients may psychologically adapt to the sick role, developing coping strategies that perpetuate their dependency on being ill. These patients can have difficulty adjusting to or transitioning into a healthier state of being. They may feel that the transplant team is expecting too fast a recovery or pressuring them to get better. They may develop unexplained somatic complaints, such as chronic pain, or become nonadherent with posttransplant care recommendations. They may experience concern, anxiety, or fear over being seen by the transplant team less frequently as their postoperative course improves. The second pattern of maladaptive behavior can be observed in patients who become acutely ill and undergo emergency transplantation (e.g., due to fulminant hepatic failure or acute drug toxicity). For these individuals, with only the briefest period of pretransplant infirmity, the transplant can be an unwelcomed event. Awareness of their organ failure, their life-and-death circumstances, and their need for a donor organ may not occur until the perioperative period. These patients may feel a heightened sense of vulnerability and may have difficulty accepting or be in denial over their medical situation. They may attempt to return to normal functioning as quickly as possible posttransplant, recovering faster than the transplant team expects, yet possibly suffering later from pushing themselves too fast (i.e., returning to work before they are physically ready). They may resent having received a transplant with all the inherent restrictions, lifestyle changes, and regimens, and they may act out their anger or denial in episodes of nonadherence.

LONG-TERM OUTCOMES. Long-term physical functioning depends in part on recipients' initial state of health, postoperative complications, other medical comorbidities, the degree of adherence to medical directives, and their motivation to stay fit. Most transplant recipients experience a multiyear period following the first anniversary of the transplant in which they achieve maximal functioning of the transplanted organ, with a relatively low level and severity of complications and infrequent episodes of acute organ rejection. Their relatively good health during this period appears to contribute to improved emotional well-being. With time, however, other problems with the organ will develop (e.g., chronic graft rejection), and patients will develop long-term immunosuppressive medication complications, including renal insufficiency or failure, diabetes, hypertension, and cancers. In fact, 10% to 20% of nonrenal transplant recipients will experience renal insufficiency or failure in the 5 years following transplantation. In addition, some patients will have a recurrence of or ongoing organ damage from their primary disease that can eventually destroy their transplant. Overall, transplantation extends life and improves health and quality of life compared to pretransplant functioning, but for most transplant recipients, their health and quality of life remain somewhat lower than that of a normal healthy person (see

Figs. 27.13–2 and 27.13–3 on survival and long-term health issues posttransplant).

PSYCHIATRIC DISORDERS AFFECTING ORGAN TRANSPLANTATION

Similar to other populations with chronic physical illnesses, organ transplant candidates and recipients are at elevated risk for elevated psychiatric symptom levels and diagnosable psychiatric disorders. Mood- and anxiety-related disorders are the most common psychiatric illnesses observed both pre- and posttransplant, although acute cognitive impairment and delirium are often seen shortly pretransplant and perioperatively due to metabolic perturbations and the effects of surgery. Some residual cognitive impairments may remain in the long-term posttransplant, although whether transplant recipients are at increased risk for dementias is unknown. In subpopulations of transplant recipients with histories of substance use disorders (e.g., patients with alcohol-related liver disease [ALD]), the risk for relapse remains a concern both pre- and posttransplant. Psychosis in transplant populations, although rare, is a significant concern because of its potential impact on health and well-being if not carefully managed. There is mounting evidence that, if not well managed, each of these classes of psychiatric disorders can affect patient health outcomes after organ transplantation.

Mood Disorders

Up to 20% of kidney recipients, 30% of liver recipients, and 63% of heart recipients experience mood and anxiety disorders during the first several years after transplant. The most common mood disorders observed in transplant populations are major depressive disorder, dysthymia, and adjustment disorders. Elevated levels of subclinical depressive symptomatology have also been noted in most studies. The risk for depression appears to be greatest during the first year posttransplant. Indeed, rates of mood disorders have been observed to be higher during the first year posttransplant than during either the waiting period for transplant or during patients' lifetimes before they became transplant candidates. The heightened risk during the first year after the transplant has been linked to stressors associated with physical recovery and rehabilitation; adaptation to immunosuppressive medications and a new medical regimen; and psychosocial and role changes as the transplant recipient transitions from critical illness to a less illness-focused lifestyle. After the first year, the likelihood of additional episodes of mood disorders appears to decline for a period of several years. Although there is little empirical evidence on the prevalence of mood or other disorders beyond about 5 years posttransplant, anecdotal accounts suggest that the prevalence of these psychiatric disorders, depression, in particular, may begin to rise in the face of the late-term transplant-related health changes noted earlier.

A growing number of studies Indicate that the presence of either diagnosable mood disorders or elevated depressive symptomatology increases the risk of poor health outcomes in transplant candidates and recipients. In a recent meta-analysis of the transplant literature, Dew et al. demonstrated that depression, either pre- or posttransplant, has a strong negative impact on posttransplant survival, increasing the relative risk of mortality by 65%. Risk does not appear to differ by type of transplant. Although morbidities are infrequently studied, depression also appears to increase the risk of death-censored graft loss in kidney recipients. To date, however, no studies have examined the mechanism by which depression may lead to poor posttransplant health outcomes, although it is speculated that depression might result in poor health behaviors (e.g., nonadherence) and/or pathophysiologic abnormalities (e.g., neurovegetative symptoms or immune mechanisms) that could contribute to morbidities and mortality. Regardless of the mechanism, the potential for depression to at least predict morbidity and mortality, combined with the fact that depression contributes to poor quality of life and functional disability in both patient and nonpatient community populations, suggests the importance of timely treatment of this condition in transplant recipients. Screening for depression, especially in the early posttransplant phase, merits consideration.

Anxiety Disorders

Anxiety disorders appear to be less prevalent than mood disorders both before and after transplantation. They have been observed in up to one-third of transplant recipients, although there have been too few studies to specify prevalence rates of each type of anxiety disorder. Some anxiety disorders (e.g., panic disorder) appear to be more common in patients with advanced lung disease than patients with other types of advanced organ diseases, and these disorders may be more common after transplantation in these patients as well. As with mood disorders, the prevalence of anxiety disorders appears to decline beyond the first year posttransplant, although stressors associated with the development of medical complications in the very late years posttransplant may precipitate new episodes of anxiety disorders.

Anxiety disorders may have the capacity to predict health outcomes in transplant recipients, although research remains sparse, and the findings are mixed. Moreover, other studies have found little impact of anxiety disorders such as generalized anxiety disorder or panic disorder. However, the potential quality-of-life benefits of treating such conditions emphasize the importance of monitoring transplant recipients for the development of anxiety.

Posttraumatic Stress Disorder

Posttraumatic stress disorder (PTSD) has received increasing attention in transplant recipients and may account for a significant proportion of all cases of anxiety disorder, especially during the first year posttransplant. Factors contributing to PTSD include actual events surrounding the transplant experience (e.g., critical medical events during the waiting period, the surgery itself, and the perioperative recovery period), as well as patients' experiences of hallucinated events and delusions during delirium.

One study found that PTSD experienced during the first year after heart transplantation predicted mortality during the subsequent 2-year follow-up period. However, persistent non-PTSD anxiety symptoms were unrelated to either mortality or morbidities in the cohort. The mechanism by which PTSD predicted mortality is not clear, and whether treatment would improve transplant recipients' health outcomes is unknown.

Psychotic Disorders and Mania

Psychosis, including conditions such as schizophrenia, schizoaffective disorder, mania, and psychotic depression, is rare in any population, including transplant recipients. Moreover, unlike the depressive and anxiety-related disorders discussed above, which occur in transplant recipients both with and without earlier lifetime histories of such disorders, psychosis emerging beyond the postoperative recovery period (and which can therefore not be ascribed exclusively to acute metabolic perturbations associated with surgery and marked medication changes) appears to occur almost exclusively in individuals with pretransplant histories of such illness. In fact, individuals with a history of psychosis are at high risk for posttransplant recurrences. A survey found that 37% of such patients experienced manic or psychotic episodes in the first several years posttransplant. Nevertheless, these patients can do well posttransplant if good social supports, adherence to medical and psychiatric directives, ongoing mental health treatment, and good control of psychotic symptoms are established prior to transplant. Under these circumstances, these patients do not have rates of nonadherence that differ from other transplant recipients. The potential for good management posttransplant, plus the very negative effects if psychoses are not well managed, make them important conditions to consider during the pretransplant assessment. Despite common beliefs, it is not clear that a prior psychiatric history makes a patient more susceptible to the potential of immunosuppressive medications, especially steroids, to produce or exacerbate psychotic symptoms or mania (i.e., “steroid psychosis”). The appropriate use of antipsychotic medication is usually adequate to prevent or manage these symptoms if they emerge, and care must be taken that these medications are not inadvertently discontinued in the early posttransplant phase.

Alcohol and Other Substance Use Disorders

Mental health professionals are most likely to be called on to evaluate patients for these disorders in the context of liver transplantation. ALD and HCV (most often incurred due to injected drug use) are two common indications for liver transplantation in the United States (Table 27.13–1). Alcohol use disorders, predominantly in liver transplant patients, have thus received the most attention with respect to candidacy issues, policy, and pre- and posttransplant outcomes. Although the use of other substances by the liver as well as other solid organ transplant patients is equally important, other substance use appears less prevalent and has received much less consideration.

There is no direct concordance between the alcohol amount consumed, an alcohol use disorder diagnosis, and either the development of ALD or alcohol-related cardiomyopathy. Evidence from the natural history of alcohol-related cardiomyopathy and ALD suggests that the risk for these diseases increases in a dose-dependent fashion with 10 years or more of daily consumption. Even modest exposure cumulatively over time can cause cirrhosis. For example, the daily consumption of two to four drinks increases the risk of cirrhosis two- to threefold for men and women compared to abstinent persons. Due to the focus of their interview, a mental health professional may be more likely than other clinicians to gather information on the quantity, frequency, and duration of alcohol use that can aid in establishing the ALD diagnosis. Importantly, the mental health assessor should establish the severity of the alcohol and other substance use disorders, length of abstinence, efforts at treatment, and whether treatment is needed to prepare for transplantation.

The rates of problematic alcohol and other substance use are relatively low posttransplant, perhaps due to rigorous selection criteria. Over the long term following liver transplant for ALD (including projections out to 10 years posttransplant), although 30% to 50% may drink (including those who have had only a single drink), some who drink will begin or return to addiction counseling, and only a smaller percentage, 10% to 20%, will continue to drink in a pattern potentially harmful to their health or graft functioning. However, since even a single drink can lead to a heavier or hazardous pattern of alcohol use, the identification of any use is a reason for concern and an opportunity to provide education and referral for addiction counseling. At both U.S. and European programs, posttransplant alcohol consumption that results in premature death or graft failure is uncommon, and long-term survival of patients with alcoholic liver disease is equal to those transplanted for other causes, even among those who drink most heavily. Nevertheless, among the heaviest drinkers, alcohol can contribute to medication nonadherence, which in combination with alcohol-related liver damage can lead to loss of graft function and death. Other nonalcohol substance use is uncommon. In a meta-analysis of studies reporting posttransplant substance use, the relapse to illicit drugs for transplant recipients with a pretransplant history of such drug use was 3.7 cases per 100 patients per year (PPY). The rate was significantly lower in liver transplant recipients (1.9 cases per 100 PPY) compared to other organ types (6.1 cases per 100 PPY). These favorable results suggest that the most important task for transplant teams is to continue ongoing monitoring for substance use posttransplant, not to condone or dismiss any use, and to provide resources for psychiatric assessment, intervention, and treatment as soon as any substance use is discovered.

Characteristics that predict relapse to use in the general population also appear to predict posttransplant relapse to use, including the severity of the substance use disorder, family history of addiction, failed rehabilitations, multiple substances of abuse, poor social support for abstinence, and short pretransplant length of abstinence. However, these risk factors have been largely studied in relation to alcohol relapse; risk factors for other substance use have been mostly unexplored.

A few studies have addressed outcomes in MMT transplant recipients. These studies show little evidence that MMT participants have problems with posttransplant relapse or nonadherence. Although postoperative pain management may be more complicated for these patients in the immediate posttransplant period, methadone can be used as their primary analgesic. Close coordination with and transition back to their outpatient MMT program is essential when the patient is medically and physically capable. In one study of MMT, liver recipients mostly were eventually returned to their original methadone dose, with only a few requiring a modest dose increase. There is very little literature regarding the use of buprenorphine in transplant patients.

Nicotine Dependence: Smoked and Smokeless Tobacco

There is substantial evidence of the contribution of tobacco use to posttransplant vascular complications, health problems, cancers, graft loss, and death. Transplant patients may be at higher risk of tobacco-induced cancers due to the effects of immunosuppression. For example, in one study, compared to nonimmunosuppressed individuals, ALD transplant recipients had a 4-fold increased risk for lung cancer and a 25-fold increased risk for oropharyngeal cancer, risks presumed due to tobacco use. As noted earlier, all types of solid organ transplant programs are increasingly likely to suggest or require candidates to be tobacco free for transplant listing. Nevertheless, significant rates of tobacco use resumption occur after transplantation. Tobacco use posttransplant is particularly common in patients transplanted after histories of alcohol or illicit drug use, with a rate of 10 cases per 100 PPY, compared to the rate of 3 cases per 100 PPY in the general transplant population. Thus, posttransplant clinicians should be vigilant in monitoring for tobacco use and counseling for cessation. Little is known about the effects of electronic cigarette use or any differences in risk compared to tobacco use.

Cognitive Disorders

Cognitive dysfunction is prevalent in patients with end-stage disease and transplant recipients alike. Assessing cognitive function in transplant candidates is essential to identifying deficits that could compromise patients' ability to follow medical directives and capacity to comprehend and participate in decision making before surgery. Caregivers may need to assist the patient with following medical directives or even with basic activities of daily living if deficits are severe. Pretransplant cognitive impairments can occur due to their advanced organ disease. For example, in heart failure patients, low cardiac output and hypoperfusion or reduced cerebral blood flow, impaired cerebrovascular reactivity, ischemia in the absence of cerebrovascular events, and even medications such as inotropic agents can contribute to cognitive impairments. Patients with older, pulsatile ventricular assist devices (VADs) pretransplant may have microemboli to the brain, leading to temporary or permanent cognitive impairment (see VAD section below). Newer nonpulsatile VADs are half as likely to associate with neurologic and psychiatric dysfunction. Cognitive impairment is also relatively common in lung transplant candidates due to ineffective pulmonary gas exchange. Both elevated arterial partial pressure of carbon dioxide (P_{CO_2}) and decreased partial pressure of oxygen (P_{O_2}) contribute to these deficits. Oxygen therapy may improve cognitive functioning in certain candidates, but whether hypercapnia or hypoxemia will lead to permanent brain damage is unclear.

Prior to transplantation, it is critical to differentiate between the fluctuating course of delirium that is potentially reversible posttransplantation and more persistent cognitive deficits that may represent pre-existing dementia or a static cognitive impairment. Serial cognitive examinations are crucial to properly evaluate the impact of chronic and end-stage disease as well as the impact of the transplant itself. If possible, therapies to reverse or alleviate the metabolic or physiologic effects of organ disease should be implemented and re-examination performed. Existing impairments can persist or worsen posttransplant. Even if reversible, posttransplant symptoms of cognitive impairment may take months to years to resolve, are associated with poor quality of life, and are unlikely to match healthy individuals. Older age and lower education are risk factors for ongoing poor cognitive performance.

Although organ dysfunction–related cognitive impairment may be reversed by transplantation, postoperative delirium or neuropsychiatric side effects of corticosteroids and immunosuppressants may worsen cognitive symptoms (see the section on neuropsychiatric side effects of immunosuppressants below). Delirium is the most common cognitive disorder associated with transplantation, can be caused by multiple factors (Table 27.13–5), and arises as part of pretransplant medical conditions or postoperative recovery. Patients with pre-existing delirium (e.g., hepatic or uremic encephalopathies, discussed in more detail below) often have persisting symptoms postoperatively, with added vulnerability to other deliriogenic risks. Patients with other pretransplant central nervous system (CNS) insults may similarly be more vulnerable to delirium postoperatively. In a prospective study of cognitive impairment in lung transplant patients, 57% showed neurocognitive impairment at transplant hospitalization discharge, while only 45% were impaired prior to transplant. Although delirium-associated impairment tended to improve at 3-month follow-up, 80% of those with pretransplant impairment remained impaired at follow-up with older age, and lower education associated with poorer neurocognitive performance. In a cohort of liver transplant patients, cognitive impairment characteristic of hepatic encephalopathy (HE) largely resolved after transplant. However, new-onset cognitive impairments emerged by 1-year posttransplant in 70% of recipients and were not related to a history of encephalopathy, primary liver disease, immunosuppression regimen, or prior alcohol use. Overall, transplant recipients’ cognition may be expected to improve the following transplant. However, cognitive functioning is unlikely to ever fully recover or return to the preillness baseline.

Table 27.13–5.
Perioperative Causes of Delirium

Anesthesia and the stress of surgery	Inhalational and intravenous anesthetics, sedative, analgesic and paralytic agents, intraoperative events, inflammation
Hypoxia	Anemia, hypotension, pulmonary emboli, pulmonary insufficiency, ARDS
Infection	Aspiration pneumonia, peritonitis, pneumonia, sepsis, stasis/decubitus ulcer, wound infections, urosepsis
Neurologic events	Abscess, cerebrovascular accident, central pontine myelinolysis, intracerebral hemorrhage, meningitis, posterior leukoencephalopathy syndrome, seizures
Metabolic imbalance	Acidosis, alkalosis, hypo- and hypercalcemia, hyponatremia, hypo- and hyperglycemia, hypo- and hyperkalemia, hypo- and hypermagnesemia
Nutritional deficiency	Protein-calorie, vitamin (B_{12}, folate, niacin, thiamine)
Medications	
Immunosuppressants	Cyclosporine, mycophenolate mofetil, steroids, tacrolimus
Antibiotics	Aminoglycosides, cephalosporins, fluoroquinolones, sulfonamides
Antifungals	Itraconazole, voriconazole
Antivirals	Acyclovir, famciclovir, ganciclovir, valaciclovir
Cardiovascular	ACE inhibitors, β-blockers, calcium channel blockers, clonidine
Gastrointestinal	Metoclopramide, proton pump inhibitors, ranitidine

Narcotics	Morphine sulfate, dihydromorphinone, etc.
Respiratory	Albuterol, montelukast, pentamidine
Sedatives	Benzodiazepines, diphenhydramine, zaleplon, zolpidem
Drug withdrawal	Narcotic or benzodiazepine withdrawal

ARDS, acute respiratory distress syndrome; ***ACE***, angiotensin-converting enzyme.

Hepatic Encephalopathy in Liver Disease. HE is classified by the fifth edition of the ***Diagnostic and Statistical Manual of Mental Disorders*** (DSM-5) as a delirium due to another medical condition. HE is commonly encountered in patients with significant liver disease and manifests in a constellation of signs and symptoms that can include alteration of consciousness (including stupor or coma), cognitive impairment, confusion or disorientation, affective or emotional dysregulation, psychosis, behavioral disturbances, bioregulatory disturbances, and disturbances of the motor system represented by asterixis, tremor, increased tendon reflexes, increased muscle tone, ataxic gait, bradykinesia, slurred speech, or incoordination. HE identification is important, given its effects on patient quality and quantity of life. Minimal HE predicts overt HE, and fulminant HE is associated with intracranial hypertension, cerebral edema, seizures, and death pretransplant. Persistent HE is rare and is observed predominantly in patients with extensive portacaval collateral circulation or after surgical or transjugular portosystemic stent shunting procedures.

HE reflects a continuum of symptoms, with subclinical or minimal HE lying at the milder end of the spectrum. Even minimal HE is clinically important, however, as it affects up to 60% of cirrhotic patients, impairs patient safety (such as the ability to drive a motor vehicle), job functioning, and quality of life, and is associated with persistent cognitive deficits after liver transplantation. The earliest signs of HE include sleep disturbances and subtle behavioral changes, which may not be obvious on clinical examination and should be elicited from patients and caregivers. Minimal HE is primarily associated with abnormalities in attention, memory, and response inhibition and also may not be readily discernable on typical clinical examination. Additional neuropsychological tests assessing psychomotor speed, praxis, concentration, and attention, such as Trail-

Making Tests A and B, Digit Symbol, and Block Design Tests, are useful in identifying minimal HE but are not commonly employed due to time and expertise limitations. A smartphone application utilizing the Stroop cognitive screening test has been validated against a standard neurocognitive battery to screen for HE and may represent a more feasible way to detect and track HE. Although an electroencephalogram (EEG) may also show common abnormalities, such as generalized slowing of dominant rhythm, cognitive testing is more efficient and perhaps more sensitive to minimal HE.

Although HE has a multifactorial pathogenesis, a predominant treatment strategy is to reduce the production of ammonia from protein in the intestinal tract and reduce its absorption through the utilization of daily medications. In addition, a variety of other compounds and metabolites are implicated in HE (e.g., mercaptans, false neurotransmitters, aromatic—phenylalanine, tyrosine, and tryptophan—and branched-chain amino acids, manganese, endogenous benzodiazepines, and increased concentrations of CNS γ-Aminobutyric acid (GABA)) alongside processes such as edema, inflammation, and increased permeability of the blood–brain barrier. HE can be precipitated by gastrointestinal hemorrhage, uremia, use of some psychoactive medications or diuretics, significant protein intake, dehydration, or electrolyte imbalance, and these problems should be detected and corrected first. Medications that can contribute to symptoms of HE or slow intestinal motility (anticholinergic drugs, tranquilizers, narcotics, and sedatives) should generally be avoided.

Uremic Encephalopathy in Renal Disease. Chronic renal failure frequently causes multiple pathophysiologic changes involving catabolic, metabolic, and endocrinologic processes that contribute to uremic encephalopathy (UE). In uremia, the accumulation of organic substances that act as neurotoxins (e.g., urea, uric acid, guanidine compounds, hippuric acid, various amino acids, indoleacetate, and others) are believed to be related to the encephalopathy, but no single metabolite has been identified as the sole cause. Other pathophysiologic changes also implicated in UE include hormonal elevations, electrolyte imbalances (e.g., acidosis, hyponatremia, hyperkalemia, hypocalcemia, hypermagnesemia), anemia, malnutrition, and CNS factors (increased calcium and decreased GABA and glycine activity). Reduced levels of thiamine, a water-soluble vitamin, have been noted in hemodialysis patients, and cases of Wernicke encephalopathy have been mistaken for UE.

UE symptoms typically fluctuate and can begin insidiously, with patients reporting mild symptoms of irritability, insomnia, or cognitive, attentional, and memory impairments. Patients may complain of physical symptoms of slurred speech, muscle twitches, or restless legs. Symptoms can progress slowly or rapidly to confusion, lethargy, seizures, overt delirium, psychosis, catatonia, and stupor or coma. An EEG can aid in diagnosis (typically showing a generalized slowing of the dominant rhythm) and detect seizures, including nonconvulsive status epilepticus, which can occur in uremia and could be mistaken for UE.

UE has also been associated with a cliniconeuroradiologic syndrome termed posterior reversible (leuko) encephalopathy syndrome (PRES) (see the section on PRES in the neurotoxicity of immunosuppressive medications below). Characteristic radiographic findings on CT or MRI are seen in the posterior cortical and subcortical white matter.

Removal of uremic toxins by hemodialysis, correction of electrolyte imbalances and anemia, and treatment of malnutrition can diminish UE symptoms and improve cognition. Seizures may be treated with anticonvulsants. Drugs that are sedating or require the elimination of the parent compound or active metabolite by renal clearance should be avoided or dosages appropriately adjusted for the degree of impaired renal functioning. As even patients well into their 70s are being considered for kidney transplantation, the risk of permanent cognitive deficits or dementia from chronic dialysis and comorbidities needs to be considered pretransplant. Kidney transplantation is the definitive treatment for UE, but one cross-sectional study of kidney transplant recipients who averaged 3.4 years posttransplant found that 58% met the criteria for cognitive impairment.

SPECIAL ISSUES IN TRANSPLANTATION

Life-Sustaining Technologies and Their Effects

Dialysis. Most renal transplant candidates are on maintenance dialysis prior to transplantation, although almost 20% receive kidney transplants as their first advanced renal disease therapy—the majority being living donor recipients. Without a living donor, a U.S. adult kidney candidate can expect a median wait time of 3 to 5 years. Thus, survival on dialysis until transplant remains an important issue. According to the U.S. Renal Data System statistics, dialysis survival after 1 year is 81% and only 45% after 5 years, but these rates are dependent on age, primary disease, and medical debility. Although there is no specific age limit for renal transplantation, advanced age in combination with medical comorbidity makes transplantation less of an option. Cardiovascular diseases are common reasons why patients are ineligible for transplantation and are also common causes of death on dialysis.

In addition to medical criteria, candidates should demonstrate good adherence to dialysis and follow-up with their nephrologist. Pretransplant adherence to dialysis is a useful indicator of posttransplant adherence. Patients must not only comply with dialysis sessions but with dietary and fluid restrictions as well. Attendance at dialysis and missed or shortened sessions should be evaluated. If nonadherence is identified, the reasons should be explored and addressed. Comorbid depression can affect adherence to dialysis and, if treated, can improve adherence. UE can impair functioning and quality of life during the wait period (see the section on UE above). Following transplantation, recipients often report good overall quality of life, with distinct improvements in physical, mental, and social functioning compared to pretransplantation.

VADs in Heart Transplantation. Progress in the development of VADs from extracorporeal devices to implantable devices has dramatically improved both the physical and psychological health of patients with advanced heart failure. These devices are now used both to “bridge” patients to transplantation and as long-term (“destination”) therapy for patients ineligible for transplant. The newest VADs now include implantable left ventricular or biventricular versions that have improved mobility, ease of wearability, and quality of life. Portable pneumatic drivers are compact and lightweight and can be worn on a shoulder strap or towed on a luggage-type carrier. Patients must be able to comply with postoperative care of the device and driveline site, as well as anticoagulation or other medical regimens. Following hospital discharge after implantation, patients are encouraged to be physically active (including walking, driving, dancing, and even work). Most patients can achieve New York Heart Association functional status I or II on a VAD. In addition, while supported on a VAD, patients can achieve significant gains in physical and physiologic rehabilitation and rebuild muscle mass, potentially stabilizing their cardiac condition. With the patient in a less urgent status, the transplant team can wait for an optimal donor organ. These benefits notwithstanding, the logistics of arranging for a VAD patient to live outside of the hospital require that patients and their family caregivers be extensively trained, with suitable home arrangements (e.g., availability of a reliable source of electricity). VAD program team members must be available at all times to advise patients, families, and any local emergency paramedical personnel near the patient's home if there is a sudden medical crisis. Despite these requirements, patients' ability to reside at home comes with many personal advantages for patients, including reduced financial burden, compared to prolonged hospitalization.

VADs can lead to major improvements in quality of life. They allow not only for improved mobility and function for most activities of daily living but also improvements in mental health and cognitive functioning. Adjusting to the VAD can be psychologically difficult for some patients, and incorporating the machinery into their bodies can evoke feelings of a damaged body image and sense of self, intense vulnerability, and fear of a VAD malfunction or infection. Psychotherapy to address these issues for both patients and caregivers may ease the transitions onto a VAD and help them prepare for eventual transplant. VAD recipients are also at risk for neurologic events (e.g., thromboembolism), although continued improvements in VAD design and management have reduced the risk for these events. Nevertheless, when they occur, they can lead to major cognitive impairments that may be irreversible. Thus, prevention and timely intervention for such events is critical for preserving patients' cognitive status and ability to function during VAD support and, potentially, after transplantation.

Living Organ Donor Outcomes

Living liver donation is much more surgically complex, invasive, and potentially riskier for the donor than kidney donation. In fact, since laparoscopic donor nephrectomy is most common for kidney donation, donors experience relatively limited postoperative pain, brief hospital stays (as little as 2 to 3 days), quicker recovery times than with open surgical procedures, and favorable cosmetic results. Although mortality rates are well less than 1% for both kidney and liver donors, at least one-third of liver donors have perioperative complications, with serious complications occurring in 6% to 14% of donors. Kidney donors typically require 4 to 6 weeks of recovery before returning to work. Liver donors require 8 to 12 weeks of recovery or longer, especially if the job involves heavy lifting.

Donor quality of life and emotional well-being do not appear, on average, to be adversely affected by organ donation. In fact, there is literature dating back to the mid-1960s showing that donors can accrue significant personal benefits from the donation in terms of feelings of self-esteem and self-worth. Extremely low percentages of donors regret having donated or state that they would not donate if they had it to do over again. For the majority of donors, there is little evidence that mental health, physical health, or quality of life are worsened by donation. However, adverse outcomes remain a possibility and have been observed in donors. The development of mood and anxiety disorders has been noted, and several case studies have documented completed or attempted suicide (although the extent to which the suicides were attributable to donation has usually been unclear). A growing literature shows that, on average, kidney donors' and liver donors' health-related quality of life is high and exceeds that of the general population even many years after donation. Nevertheless, up to 20% of donors may show quality of life impairments in one or more areas, including mental health and emotional well-being. A history of psychiatric difficulties increases donors' risk of experiencing mental health problems postdonation.

Treatment Issues Specific to Transplantation

Psychotherapeutic Treatment Modalities. For the pretransplant period, several randomized controlled studies have examined a variety of cognitive behavioral and education-based therapeutic interventions to address issues specific to this phase, including uncertainty of wait time or ability to survive to transplant, a decline in health and increasing dependence on others, losses of roles and functioning, and decrements in quality of life. Some studies have included caregivers as well. Results have been mixed but mostly demonstrate improvements in quality of life and reduction in psychological distress among candidates in the active treatment arms.

Following transplantation, individual psychotherapy appears beneficial in addressing three recurring psychological themes: (1) fear of organ rejection, (2) feelings of paradoxical loss posttransplant despite successful transplantation, and (3) psychological integration of the organ. Group therapy and other innovative strategies have also been employed to deal with specific issues of transplantation and to also address logistics and staffing resources. Group therapy is a frequently and successfully used technique for organ transplantation patients and family members. Educational groups are often used to provide information to transplant candidates about the transplant process, systems issues, and the responsibilities of transplant recipients and caregivers. Depending on the available resources and number of patients near to the program location, support groups may also be comprised of pre- or posttransplant patients, patients' family caregivers, and organ-specific or cross-organ type participants. Group therapy participants report decreases in negative affect, increases in positive affect and happiness, less illness intrusiveness, improved quality of life, and decreased need for contact from their transplant coordinators both in the clinic and by telephone for social support. Additional support can be found through the Transplant Recipient International Organization (www.trioweb.org), a nonprofit international organization that provides support, education, and advocacy for transplant recipients worldwide. The member chapters usually have regular meetings of transplant recipients, and they provide group support for transplant candidates, recipients, and their families.

Some transplant programs have formal and informal peer mentoring programs whereby an already transplanted recipient augments patient care by the staff. Mentors can provide information and support from a peer perspective. Mindfulness-based stress reduction and other non–transplant-specific therapeutic approaches have demonstrated benefits to patients' psychological adjustment and quality of life. Dew and colleagues developed an internet-based intervention to deal with the logistical problem of recipients living at a distance from the transplant program. The multicomponent intervention included stress and medical regimen management workshops on the website, monitored discussion groups, access to electronic communication with the transplant team, and information on transplant-related health issues. Compared to patients without access to the website, heart recipients who used the internet intervention had significant decreases in depressive and anxiety symptoms, their quality of life in the social functioning domain significantly improved, and their caregivers had decreases in anxiety and hostility symptoms. In addition, the subgroup using the website's medical regimen workshop showed significantly better adherence to multiple elements of the medical regimen.

Interventions for Adherence to the Posttransplant Regimen. As discussed earlier, lifelong adherence to the posttransplant regimen can be difficult. Many factors contribute to posttransplant nonadherence, including medication side effects, behavior problems (e.g., poor planning, poor health responsibility), and social issues (e.g., insurance coverage, copays, transportation). Eliciting patient-specific barriers will guide the selection or development of the best approach for intervention. Patient involvement in the selection of adherence intervention and allowing them, with support, to make decisions about their care is important to success. Trials designed to minimize posttransplant nonadherence show no single intervention to be superior, and a combination of interventions might be the best approach. Re-education about the regimen and its rationale can be helpful and educational interventions delivered by pharmacists have been effective. Problem-solving, cognitive behavioral therapy or other behavioral therapies may be useful. Patients traumatized by the transplant experience may have associated adherence issues that could be targeted with psychotherapy. Identification and treatment of depression and/or substance use may improve adherence.

Interventions have included a variety of approaches: case management/coaching and educational interventions, multicomponent interventions, patient-tailored interventions with a multidisciplinary approach and with the widespread use of technology, cognitive educational interventions making use of the internet, and electronic feedback (including real-time feedback). Electronic health (eHealth) interventions, including apps and other mobile health strategies, hold promise for improving patient self-monitoring and adherence to posttransplant care, ideally increasing patients' involvement in their care. Consensus is growing that short-term, time-limited interventions are inadequate; either intervention must be ongoing (e.g., via eHealth) or periodic booster or maintenance sessions may be required to maintain optimal adherence.

Psychopharmacologic Therapy. Pharmacotherapy is complicated during the pretransplant period due to the presence of advanced organ disease. Pharmacokinetic alterations caused by changes in perfusion, fluid status, protein binding, hepatic drug metabolism, and excretion are common in advanced organ disease and require careful selection of medications and dosage. In addition, pretransplant patients are more susceptible to adverse side effects (e.g., encephalopathy, arrhythmia) and potential drug–drug interactions, as it is not uncommon for them to be on a number of medications prior to transplantation. Some psychotropic agents should be avoided in advanced organ disease, particularly when advanced cirrhosis is present. In general, pharmacotherapy in pretransplant patients calls for using lower starting dosages and slower titration of drugs until they reach clinical effectiveness.

Following transplantation, pharmacokinetic alterations caused by advanced organ disease usually resolve, but considerations about potential drug–drug interactions arise. In particular, many of the immunosuppressive agents (tacrolimus [Prograf], cyclosporine [Neoral and others], sirolimus [Rapamune], mycophenolate mofetil [Cellcept, Myfortic]) are primarily metabolized by cytochrome (CYP) 450 3A4; thus, psychotropic medications that strongly inhibit CYP 450 3A4 may pose a risk. Such psychotropic drugs include, in decreasing order of inhibition: fluvoxamine, nefazodone > fluoxetine > sertraline, tricyclic antidepressants, paroxetine > venlafaxine. Although fluoxetine and tricyclic antidepressants have been used to treat depressed transplant recipients without significant changes in immunosuppressant levels, nefazodone has resulted in toxic immunosuppressant levels with serious complications. Although this suggests that the amount of CYP 450 3A4 inhibition may not create clinically meaningful drug interactions for all patients, these drugs should be used cautiously and with frequent monitoring of immunosuppressant levels. Most immunosuppressants have significant metabolic side effects (e.g., weight gain, hyperlipidemia, glucose intolerance), which must be considered with choosing psychotropics to prescribe. Table 27.13–6 summarizes primary considerations about psychotropic agents pre- and posttransplantation.

Due to its narrow therapeutic window, the use of lithium before and early after transplantation can be complicated due to fluctuations in fluid status (e.g., edema, dehydration) and renal function causing unstable lithium levels and potential toxicity. In addition, lithium can cause renal tubular damage that might be problematic when combined with the nephrotoxic effects of immunosuppressant medications. If lithium is necessary, monitoring of lithium levels, fluid status, electrolytes, renal functioning, and creatinine clearance should be closely followed and lithium discontinued if renal insufficiency develops.

The use of antipsychotic medications may be necessary posttransplant to manage delirium, marked immunosuppressant-induced mood changes, or a pre-existing psychotic disorder. Symptomatic treatment of agitation, hallucinations, or paranoia usually calls for the use of atypical antipsychotic agents or traditional neuroleptics (e.g., haloperidol). Due to the risk of QT prolongation associated with these medications, along with the likely presence of other medications (e.g., tacrolimus) or conditions that have also been linked to torsades de pointes, electrocardiogram (ECG) monitoring is recommended.

Benzodiazepines are effective for anxiolysis before and after transplant, but care is needed when selecting these agents. Oversedation increased cognitive impairment, and reactivation of a past substance use disorder are potential risks. Pretransplant, impaired hepatic oxidation can delay clearance of benzodiazepines and increase the risk for adverse side effects. Use of agents that only require conjugation (i.e., lorazepam, oxazepam, temazepam) would be preferable. Buspirone can also be beneficial for anxiety and may also have respiratory stimulating properties, especially beneficial for patients with lung disease or those attempting to wean from a ventilator.

Medications that may reduce cravings and potentially diminish the risk for relapse to alcohol (e.g., acamprosate, ondansetron, naltrexone) or opioids (naltrexone) have not been studied in transplant patients. Naltrexone has a small risk of hepatotoxicity (contained in a black box warning) and is not recommended for patients with active liver disease, hepatitis, or liver failure, making it less suitable for liver transplant patients. At doses higher than recommended (more than 300 mg per day), naltrexone can be a direct hepatotoxin. Disulfiram blocks the oxidation of alcohol at the acetaldehyde stage and can create severe nausea, vomiting, and hemodynamic instability. It requires hepatic metabolism for conversion into an active drug. In addition, a metabolite of disulfiram is an inhibitor of CYP 450 3A4, and posttransplant may interfere with immunosuppressant metabolism.

For tobacco users, in addition to smoking cessation counseling and nicotine replacement therapies (gum, patches, lozenges, and inhaled), the use of medications (bupropion or varenicline) may aid in cessation. Bupropion should not be used in those with a history of a seizure disorder or electrolyte disturbances that could contribute to a seizure. Varenicline is excreted by the kidneys, and dose adjustment is required for patients with moderate to severe renal impairment. Both medications can increase the likelihood of success and should be continued for 6 to 8 weeks after tobacco cessation to prevent relapse.

Neurotoxicity of Immunosuppressant Medications

While transplant teams tend to focus on the neurotoxic effects of corticosteroids (e.g., delirium, agitation, insomnia, depression), the impact of these agents on transplant recipients is often transient and occurs early posttransplant when higher doses are utilized. Of greater significance are calcineurin inhibitors (CNIs), which continue to serve as a mainstay of immunosuppressive medication regimens in solid organ transplantation. Tacrolimus and cyclosporine have similar neurotoxic side-effect profiles, with as many as 40% to 60% of transplant recipients experiencing mild neurologic symptoms such as tremulousness, headache, restlessness, agitation, insomnia, vivid dreams, paresthesias, and mood changes. Moderate-to-severe neurotoxic side effects (e.g., delirium, seizures, cortical blindness, catatonia, coma) occur less often but can reach 20% to 30% in the early postoperative period. Neurotoxicity is likely more common with tacrolimus but can occur with either agent even at therapeutic drug levels. Liver transplant patients appear to be more susceptible to these neurotoxic side effects than heart and lung transplant recipients.

Table 27.13–6.

Pharmacotherapy Considerations in Transplant Patients

Medication	Specific Considerations for Transplant Patients
Antidepressants	
SSRIs	
Citalopram	Lack of drug–drug interactions, possible QT prolongation
Escitalopram	Preferred, lack of drug–drug interactions
Fluoxetine	Reduce dosage or lengthen dosing interval in HI; risk of interaction with immunosuppressants
Fluvoxamine	Consider lower initial dose and slow titration in HI; risk of interaction with immunosuppressants
Paroxetine	Reduce initial dose and slow titration in moderate-to-severe RI and HI. Anticholinergic side effects
Sertraline	May consider lower or less frequent dosing in HI
SNRIs	
Venlafaxine	Reduce total daily dose in moderate-to-severe RI/hemodialysis/HI; risk of elevated blood pressure
Desvenlafaxine	Dose should not exceed 50 mg/day in moderate RI; no more than 50 mg every other day in severe RI
Duloxetine	Avoid use in HI and in severe RI.
Vortioxetine	Not recommended in severe HI/cirrhosis
Others	
Bupropion	Limit dose to 150 mg of XL formulation every other day in severe HI; 100 mg/day of SR formulation; reduce initial dose in RI
TCAs	Cardiovascular side effects are undesirable in cardiac patients. Anticholinergic side effects
MAOIs	Significant risk of drug–drug interactions make this undesirable in transplant patients
Mirtazapine	Clearance reduced in moderate-to-severe RI/HI.

	Consider lower initial dose and slower titration, sedative, and weight gain side effects
Vilazodone	No dose adjustment necessary in RI/HI, Not recommended in severe RI
Nefazodone	Avoid use due to risk of drug–drug interaction with immunosuppressants
Brexanolone	Avoid use in ESRD (eGFR <15 mL/min/1.73 m^2, no dose adjustment necessary in mild–severe RI)
Esketamine	Not recommended in severe HI, monitor longer for adverse reactions and moderate HI
Anxiolytics	
Benzodiazepines	Lorazepam, oxazepam, temazepam do not require hepatic oxidation. Safer choices in presence of HI; risk of respiratory depression
Buspirone	Not recommended in severe RI or HI
	All pose risk of QT prolongation; atypical agents also increase risk of metabolic syndrome
Antipsychotics	
Atypical	
Aripiprazole	No dose adjustment needed in RI/HI
Asenapine	Use contraindicated in severe HI
Clozapine	Risk of seizures, agranulocytosis, orthostatic hypotension complicate use in transplant
Lurasidone	Moderate HI and moderate-to-severe RI: starting dose 20 mg/day, total dose not to exceed 80 mg/day; severe HI: total dose not to exceed 40 mg/day
Iloperidone	Use not recommended in severe HI
Olanzapine	Poses greater risk of metabolic syndrome.
Paliperidone	Clearance reduced in RI; for moderate-to-severe RI: starting dose 1.5 mg/day, total dose maximum 3 mg/day

Risperidone	Severe RI: start dose at 0.5 mg bid. Slow titration is recommended
Quetiapine	No dosage adjustment needed in RI/HI
Ziprasidone	No dosage adjustment needed in RI/HI; greater risk of QT prolongation
Brexiprazole	For moderate–severe RI/HI, maximum dose 2 mg/day for major depression, 3 mg/day for schizophrenia
Cariprazine	Not recommended in severe RI/HI
Lumateperone	Not recommended in moderate–severe HI
Typical	
e.g., haloperidol	No dosage adjustment needed in RI/HI
Mood Stabilizers/Anticonvulsants	
Lithium	Moderate RI: use 50–75% of usual dosage hemodialysis-supplemental dose of 300 mg once after each dialysis session May be difficult to use due to shifting fluid status, unstable levels, risk of toxicity; cyclosporine reduces Li clearance
Carbamazepine (CBZ)	CYP 3A4 inducer effects of CBZ lower levels of immunosuppressants, risk of organ rejection
Oxcarbazepine	Avoid extended-release formulation in severe HI; severe RI: reduce initial dose to 300 mg/day, titrate slowly, CYP 3A4 inducer drops immunosuppressant levels
Lamotrigine	Severe HI with ascites: reduce initial, escalation, and maintenance dose by 50%; without ascites: reduce by 25%
Topiramate	Moderate-to-severe RI: reduce dose by 50%; cognitive side effects may be problematic for transplant patients
Valproic acid	Avoid use in HI if possible due to reduced clearance; reduced serum albumin can raise

	free VPA levels
Gabapentin	Moderate RI: 600 mg/day; severe RI: 150–300 mg/day; hemodialysis: may require both maintenance and supplemental dose after dialysis session
Psychostimulants	
Atomoxetine	Moderate HI: reduce initial and total dose by 50%; severe HI: reduce initial and total dose by 75%
Methylphenidate	No dosage adjustment needed in RI/HI; use in cardiac patients limited by cardiovascular side effects
Amphetamine	No dosage adjustment needed in RI/HI; use in cardiac patients limited by cardiovascular side effects
Modafinil	Severe HI: reduce dose by 50%.
Armodafinil	Severe HI: reduce dose due to reduced clearance; severe RI: lack of dosing info prevents safe use
Anticraving Medications	
Buprenorphine	Moderate-to-severe HI: use with caution and at lower dose; may have longer exposure and activity
Naltrexone	Use in severe RI/HI has not been studied—thus, safety is unknown
Methadone	RI/HI: lower initial dose and slow titration advised; can prolong QT interval; use with caution in cardiac patients
Varenicline	Severe RI: initial dose of 0.5 mg/day, may increase to 0.5 mg bid; risk of neuropsychiatric side effects
Acamprosate	Moderate RI: initial dose of 333 mg tid; severe RI: use is contraindicated
Disulfiram	Can cause dangerous hemodynamic instability in medically vulnerable patients if they drink, HI: use with caution; may cause hepatotoxicity even

	without alcohol use; a metabolite is a 3A4 inhibitor so could raise immunosuppressant levels

HI, hepatic insufficiency; ***RI***, renal insufficiency; ***moderate RI***, creatinine clearance 30–50 mL/min; ***severe RI***, creatinine clearance <30 mL/min; ***moderate HI***, Childs–Pugh class B; ***severe HI***, Childs–Pugh class C.

The actual mechanism of CNI neurotoxicity is unclear but possibly multifactorial. Investigations into the etiology of more serious neurotoxic side effects suggest associations with biochemical and electrolyte derangements, including higher plasma levels, intravenous administration, hypocholesterolemia, and hypomagnesemia. However, these features are not present in every case. Neurotoxicity of CNIs may involve drug-induced vascular changes or direct action at the CNS neuronal level. Studies have revealed that CNIs increase sympathetic outflow and nerve activity, alter mitochondrial function, may be selectively toxic to white matter, and are capable of disrupting the integrity of the blood–brain barrier. In some cases, correcting the metabolic disturbance or reducing the CNI drug level will lead to the resolution of neurotoxic side effects, though severe symptoms may require discontinuation of the CNI.

While CNIs are highly effective immunosuppressants, their toxic side effects, particularly nephrotoxicity, have led to the search for alternate immunosuppressant agents that might serve as substitutes or allow reductions in CNI dosage. Within the past several years, a number of newer medications have been developed, including mycophenolate mofetil, sirolimus, everolimus, belatacept, as well as monoclonal antibodies such as basiliximab and daclizumab. Monoclonal antibodies are used for the treatment of acute rejection and as induction therapy during the first days or weeks after transplantation, but not for maintenance therapy. These offer reduced nephrotoxicity and also generally demonstrate less neurotoxic side effects (Table 27.13–7). However, there are infrequent complications that can arise with either CNIs or some of the newer immunosuppressants—namely, PRES or, rarely, progressive multifocal leukoencephalopathy (PML). PRES usually arises in the first month posttransplantation, though it can arise years later, and presents with a range of symptoms including seizure, headache, visual changes, confusion, ataxia, and other neurologic symptoms depending on the location of the lesion. Diagnosis via MRI usually reveals

T_2-weighted and fluid-attenuated inversion recovery (FLAIR) MRI images showing hyperintense lesions typically in the subcortical and cortical regions of the bilateral parieto-occipital lobes, but lesions can occur in other areas and in the brainstem. Discontinuation of CNI or sirolimus usually leads to the resolution of clinical and radiologic abnormalities. PML is a rare and severe demyelinating disease caused by the reactivation of the JC virus that can develop in immunocompromised patient populations.

Table 27.13–7.
Neurotoxicity of Immunosuppressant Medications

Medication	Side Effect(s)
Calcineurin inhibitors	Headache, tremor, paresthesia, visual disturbances, insomnia, delirium, seizures, apraxia, parkinsonism, cortical blindness, anxiety, mood changes, cognitive impairment, agitation, PRES, PML
Corticosteroids	Euphoria, depression, anxiety, agitation, insomnia, delirium, irritability, hallucinations, psychosis, cognitive impairment
Sirolimus	Headache, dizziness, tremor, depression, insomnia, PRES, PML
Everolimus	Tremor, somnolence, dizziness, paresthesia, hypoesthesia, fatigue, weakness, headache
Mycophenolate mofetil	Headache, insomnia, anxiety, dizziness, weakness, fatigue, PRES, PML
Belatacept	Headache, Guillain–Barré syndrome, PML
Azathioprine	PML
Monoclonal antibodies	
Basiliximab	Headache, tremor, insomnia, anxiety, depression, fatigue
Daclizumab	Insomnia, fatigue
Rituximab	Anxiety, depression, delirium, hallucinations
Alemtuzumab	Headache, dizziness, light-headedness, weakness, fatigue

PRES, posterior reversible encephalopathy syndrome; ***PML***, progressive multifocal leukoencephalopathy.

FURTHER READINGS

Asrani SK, Trotter J, Lake J, et al. Meeting report: the Dallas consensus conference on liver transplantation for alcohol associated hepatitis. ***Liver Transpl***. 2020;26(1):127–140.

Bajaj JS, Thacker LR, Heuman DM, et al. The Stroop smartphone application is a short and valid method to screen for minimal hepatic encephalopathy. ***Hepatology***. 2013;58:1122–1132.

Coffman KL, Crone C. Rational guidelines for transplantation in patients with psychotic disorders. ***Curr Opin Organ Transplant***. 2002;7:385.

Corbett C, Armstrong MJ, Parker R, Webb K, Neuberger JM. Mental health disorders and solid-organ transplant recipients. ***Transplantation***. 2013;96(7):593–600.

De Bleser L, Matteson M, Dobbels F, Russell C, De Geest S. Interventions to improve medication-adherence after transplantation: a systematic review. ***Transpl Int***. 2009;22(8):780–797.

*De Geest S, Burkhalter H, Bogert L, Berben L, Glass TR, Denhaerynck K; Psychosocial Interest Group, Swiss Transplant Cohort Study. Describing the evolution of medication nonadherence from pretransplant until 3 years post-transplant and determining pretransplant medication nonadherence as risk factor for post-transplant nonadherence to immunosuppressives: the Swiss Transplant Cohort Study. ***Transpl Int***. 2014;27:657–666.

Denhaerynck K, Dobbels F, Cleemput I, et al. Prevalence, consequences, and determinants of nonadherence in adult renal transplant patients: a literature review. ***Transpl Int***. 2005;18:1121–1133.

Dew MA. The LVAD patient bridging to wellness – Psychological assessment and rehabilitation. In: Barac YD, Silvestry SC, Goldstein DJ, Daneshmand MA, eds. ***Textbook of Transplantation and Mechanical Support for End-Stage Heart and Lung Disease***. Wiley-Blackwell; 2023.

Dew MA, DiMartini AF, Dobbels F, et al. The 2018 ISHLT/APM/AST/ICCAC/STSW recommendations for the psychosocial evaluation of adult cardiothoracic transplant candidates and candidates for long-term mechanical circulatory support. ***J Heart Lung Transplant***. 2018;37(7):803–823.

*Dew MA, DiMartini AF, De Vito Dabbs A, et al. Rates and risk factors for nonadherence to the medical regimen after adult solid organ transplantation. ***Transplantation***. 2007;83(7):858.

Dew MA, DiMartini AF, Steel JL, Jowsey-Gregoire JG. Psychosocial evaluation, care and quality of life. In: Lentine K, Concepcion B, Lerma E, eds. ***Living Kidney Donation***. Springer Publishing; 2021:253–282.

Dew MA, DiMartini AF, Steel J, et al. Meta-analysis of risk for relapse to substance use after transplantation of the liver or other solid organs. ***Liver Transpl***. 2008;14(2):159.

*Dew MA, Fox KR, DiMartini AF. Organ transplantation: Psychological and behavioral aspects. In: Gellman M, ed. ***Encyclopedia of Behavioral Medicine***. 2nd ed. Springer Nature; 2020, e-book.

Dew MA, Goycoolea JM, Harris RC, et al. An internet-based intervention to improve psychosocial outcomes in heart transplant recipients and family caregivers: development and evaluation. ***J Heart Lung Transplant***. 2004;23(6):745–758.

Dew MA, Jacobs C, Jowsey SG, Hanto R, Miller C, Delmonico FL; United Network for Organ Sharing (UNOS), American Society of Transplant Surgeons, American Society of Transplantation. Guidelines for the psychosocial evaluation of living unrelated kidney donors in the United States. ***Am J Transplant***. 2007;7(5):1047.

Dew MA, Kormos RL, Winowich S, et al. Quality of life outcomes after heart transplantation in individuals bridged to transplant with ventricular assist devices. ***J Heart Lung Transplant***. 2001;20(11):1199.

*Dew MA, Rosenberger EM, Myaskovsky L, et al. Depression and anxiety as risk factors for morbidity and mortality after organ transplantation: a systematic review and meta-analysis. ***Transplantation***. 2015;100(5):988–1003.

DiMartini A, Crone C, Dew MA. Alcohol and substance use in liver transplant patients. ***Clin Liver Dis***. 2011;15(4):727–751.

*DiMartini A, Crone C, Fireman M, Dew MA. Psychiatric aspects of organ transplantation in critical care. ***Crit Care Clin***. 2008;24(4):949.

Dobbels F, Verleden G, Dupont L, Vanhaecke J, De Geest S. To transplant or not? The importance of psychosocial and behavioural factors before lung transplantation. ***Chronic Respir Dis***. 2006;3(1):39.

Evans LD, Stock EM, Zeber JE, et al. Posttransplantation outcomes in veterans with serious mental illness. ***Transplantation***. 2015;99(8):e57–e65.

Gupta A, Mahnken JD, Johnson DK, et al. Prevalence and correlates of cognitive impairment in kidney transplant recipients. ***BMC Nephrol***. 2017;18(1):158.

Jowsey S, Schneekloth T. Psychosocial assessment of adult liver transplant recipients. In: Busuttil RW, Klintmalm GK, eds. ***Transplantation of the Liver***. 2nd ed. Elsevier Saunders; 2005:395–404.

Kirklin JK, Naftel DC, Kormos RL, et al. Fifth INTERMACS annual report: risk factor analysis from more than 6000 mechanical circulatory support patients. ***J Heart Lung Transplant***. 2013;32:141–156.

Liu L, Schiano T, Lau N, et al. Survival and risk of recidivism in methadone-dependent patients undergoing liver transplantation. ***Am J Transplant***. 2003;3:1273.

Maldonado JR, Dubois HC, David EE, et al. The Stanford Integrated Psychosocial Assessment for Transplantation (SIPAT): a new tool for the psychosocial evaluation of pre-transplant candidates. ***Psychosomatics***. 2012;53(2):123–132.

Myaskovsky L, Jesse MT, Kuntz K, et al. Report from the American Society of Transplantation Psychosocial Community of Practice Adherence Task Force: real-world options for promoting adherence in adult recipients. ***Clin Transplant***. 2018;32(9):e13353.

*Olbrisch ME, Benedict SM, Ashe K, Levenson J. Psychological assessment and care of organ transplant patients. ***J Consult Clin Psychol***. 2002;70(3):771–783.

Parekh PI, Blumenthal JA, Babyak MA, et al; INSPIRE Investigators. Gas exchange and exercise capacity affect neurocognitive performance in patients with lung disease. ***Psychosom Med***. 2005;67(3):425–432.

Popoola J, Greene H, Kyegombe M, MacPhee IA. Patient involvement in selection of immunosuppressive regimen following transplantation. ***Patient Prefer Adherence***. 2014;8:1705–1712.

Sher Y, Maldonado J. ***The psychosocial care of end-stage disease and transplant patients***. 1st ed. Springer Publishing; 2019.

Skotzko CE, Strouse TB. Solid organ transplantation. In: Wise MG, Rundell JR, eds. ***The American Psychiatric Publishing Textbook of Consultation-Liaison Psychiatry: Psychiatry in the Medically Ill***. American Psychiatric Press; 2002:623.

Smith PJ, Rivelli S, Waters A, et al. Neurocognitive changes after lung transplantation. ***Ann Am Thorac Soc***. 2014;11(10):1520–1527.

Tryc AB, Pflugrad H, Goldbecker A, et al. New-onset cognitive dysfunction impairs the quality of life in patients after liver transplantation. ***Liver Transpl***. 2014;20(7):807–814.

27.14 Psychiatric Care of the Burned Patient

MARTHA C. GAMBOA, M.D., EJINE OKOROAFOR, M.D., F.A.P.A., REENA BAHARANI, M.D., YVETTE SMOLIN, M.D., M.ED., AND MICHAEL BLUMENFIELD, M.D.

Burn injuries represent a major global public health burden and constitute the fourth most common type of trauma worldwide, following traffic accidents, falls, and interpersonal violence. Across the globe, fire, heat, and hot substances incidents resulted in nearly 9 million injuries and more than 120,000 deaths worldwide in 2017. The mortality and morbidity related to fire, heat, and hot substances–related injuries generally decreased between 1990 and 2017 across income groups. However, approximately 96% of fatal fire-related burns continue to occur in low- and middle-income countries.

In 2019, in the United States, local fire departments responded to an estimated 1.3 million fires. These fires caused roughly 3,700 civilian fire-related deaths and 16,600 reported civilian fire injuries.

Major medical and surgical advances have greatly increased the burned patient's survival. The 2017 National Burn Repository (NBR) reviewed the combined data set of more than 212,820 acute burn admissions between 2008 and 2017, data provided by 101 hospitals from 37 states and the District of Columbia. The two most common etiologies were fire or flame burns and scalds, accounting for 76% of cases reported. Scald injuries were most prevalent in children under age 5, while fire/flame injuries dominated the remaining age categories. More than 67% of the reported total burn sizes were less than 10% of total body surface area (TBSA). These cases had a mortality rate of 0.6%. The mortality rate for all cases was 3.1% and 5.6% for fire or flame injuries. The mortality rate declined from less than 4% to less than 3% for both men and women. The length of hospitalization was slightly greater than 1 day per percentage TBSA burned. The length of hospitalization for women declined from 9.4 days to 7.3 days, and for men, it declined from 9.5 to 8.5 days.

Increased survival rates translate to a larger number of burn patients that require physical, psychological, and social rehabilitation. Burn patients often suffer severe pain, major psychological distress, and very complicated psychosocial disruption to their lives. Psychiatrists and other mental health professionals working in the medical and surgical settings have a major role to play in the care of these patients.

RISK AND ASSOCIATED FACTORS FOR DEVELOPING BURN INJURIES

Burns result from the destruction of one or more cell layers of the skin or other organ tissues in contact with hot liquids (scalds), solids (contact burns), or flames (flame burns). Injuries due to ultraviolet radiation, radioactivity, electricity, or chemicals, as well as respiratory damage resulting from smoke inhalation, are also considered to be burns.

According to the World Health Organization (WHO), some of the major risk factors for sustaining burn injuries around the world are alcohol and smoking, particularly in combination. Contributory factors include local cultural practices, such as floor-level stoves, bedside fires, loose flammable attires, and fire walking; socioeconomic status (for both developed and developing countries); overcrowded living conditions; lack of proper safety measures; and insufficient parental supervision. There is notably an increased risk of burn injuries in vulnerable populations, including children, women, and the older adults.

Table 27.14–1.
Factors that Contribute to the Risks of Childhood Burn Injury

Physical Environment

- Structural materials of residence (substandard housing)
- Heating mechanisms (space heaters), objects, or flame (candles, stoves, bathtub)
- Presence of smoke/carbon monoxide detectors/alarms, fire extinguishers (half of all home fire deaths occurred in homes without alarm)
- Awareness of fire safety precautions and services, including adequate means of egress
- Rural areas
- Fall and winter months

Family, Cultural, and Social Structure

- Level of supervision of children
- Overcrowding
- Low-income status
- Traditional practices (e.g., open flame ovens)

Developmental/Gender (Chronologic, Emotional, and Intellectual Stage)

- Incredibly young children, especially under 4 years old
- Boys, greater than girls (boys are with more risk-taking behaviors) and use of flammable materials
- Increased levels of risk-taking behavior and exploration
- Epilepsy
- Learning disabilities and mental retardation

Psychological/Behavioral Disturbances of Child or Caregiver

- Alcohol/substance abuse and smoking
- Childhood/adult depression
- Violence/abuse
- Oppositional defiant disorder, attention deficit hyperactivity disorder, bipolar disorder

The 2017 NBR reported that in the United States, the two most common etiologies for burn injuries across all age groups are flame burns and scalds, accounting for almost 8 out of 10 reported burn cases. Scalds are more prevalent in the younger population, especially among those younger than 5 years of age, while fire/flame burns prevail in the remaining age categories. The vast majority of burns seen at burn centers were considered accidental, with only 13% of these related to work and only 3% specified as nonaccidental. The home continues to be the most common place of burn injury occurrence (73%). According to the 2017 NBR, males predominate in almost all age categories in comparison to females, except in individuals older than 80. Children between 1 and 15 years of age accounted for 23.5%. Patients between the ages of 20 and 59 years old represented 55% of the burn cases. Patients older than 60 represented 15% of the burn cases.

Table 27.14–2.
Typical Signs of Physical Abuse in Burned Children

Bilateral burns to feet, buttock, or back by dipping in hot water bath
Unburned lower extremity flexion creases, as child reflexively tries to avoid pain
Cigarette-sized round lesions
Posterior deep burns, with sharp margins from sustained or forced contact with hot object (child prevented from withdrawing)
Illogical, inconsistent explanations of the burn event provided by caregiver
Failure to thrive
Physical numbness to pain
Depression, unresponsiveness, intense fear reaction
Multiple injuries at different stages of healing, including fractures

The risk of burns in children usually results from the interplay between environmental, developmental, behavioral, and familial factors. Some of the environmental risk factors include lack of supervision, proximity to burning substances, clothing made of flammable materials, congested housing, low socioeconomic status, and poor social support. Behavioral factors include risk-taking tendencies, fire setting, as well as abuse of alcohol or illicit drugs. Familial factors including parental illiteracy, parental psychopathology, and parental stress or neglect also contribute to burn injuries in children. Child neglect or abuse causes about 6% to 20% of pediatric burns, with scalds as the prevalent etiologic factor.

The average rate of burn injuries decreases from age 5 until adolescence. However, after age 15, the incidence rate of burn injuries begins to increase again, presumably because of greater exposure to hazards, experimentation, and risk taking, as well as employment. Psychiatric diagnoses usually associated with increased risk of burn injuries include childhood depression, learning disorders, intellectual disabilities, substance use disorders, oppositional defiant disorder, conduct disorder, attention deficit hyperactivity disorder, and bipolar disorder.

The risk factors for burn injuries in adults include chronic mental illness, neurocognitive disorders, alcohol and substance use disorders, character disorders, history of suicide attempts, as well as a history of abuse or homicide. Factors like prior physical or sexual abuse, substance use disorders, affective disorders, psychotic disorders, and personality disorders significantly contribute to the risk of self-inflicted burns. Conduct disorder, personality disorders, bipolar disorder, major depression, schizophrenia, factitious disorder, and substance use disorders are some of the psychiatric diagnoses associated with increased risk for self-inflicted burns. Female gender, Black ethnicity, unemployed, single, and homeless status, and Medicaid funding are common characteristics found in the majority of patients with assault-related burn injuries.

Some medical and neuropsychiatric conditions interfere with intact sensory, motor, or cognitive functions, and in turn, decrease the patient’s ability to properly recognize and respond to situations that can represent a risk for burn injury. In the setting of insufficient or inappropriate caregiver supervision, early-onset neurologic impairments including cerebral palsy, spina bifida, attention deficit hyperactivity disorder, intellectual developmental disorder, and autism increase the risk of sustaining scalds, flame burns, and contact burns during activities related to cooking, eating, and bathing. The lack of temperature control settings in the shower for this group of patients is an additional risk factor.

In addition, natural disasters can create special circumstances that increase the risks of burns. A retrospective study reviewing the burn patient records at Staten Island University Hospital Burn Center following Super Storm Hurricane Sandy estimated that the widespread power outages and losses of hot water and heat that followed the storm led to a majority of scald injuries because of the damage to hot water pipes and hot water heaters and the use of portable heating sources to cook food and boil water.

A 62-year-old widower with a history of diabetes mellitus for 20 years, complicated by peripheral neuropathy and no psychiatric history, suffered 18% scald burns to her lower extremities while taking a hot shower. She only felt mild discomfort on her feet and hours later noticed erythema and blister formation. She then decided to go to urgent care and was referred to the burn unit. During a psychiatric assessment, she reported being distracted thinking about the upcoming death anniversary of her late husband and forgetting to call her heating company for her water heater tune-up, which was overdue. Upon further questioning, she endorsed mild depressive symptoms, including low mood, decreased energy, decreased appetite, and decreased concentration, as well as anticipatory anxiety related to dressing changes. She was started on sertraline and individual supportive psychotherapy. Prior to discharge to the physical rehabilitation unit, she reported improved mood and appetite. She was referred to a private psychiatry group that took her insurance.

ALCOHOL, SUBSTANCE ABUSE, AND BURN INJURY

The association between burn injuries and alcohol is widely studied and accepted. It has both a direct and an indirect causal effect. Approximately 50% of burn patient fatalities have a history of alcohol use. Alcohol use is also known to increase the morbidity and mortality of hospitalized burn-injured patients. This is due in part to delay in treatment prior to coming to the hospital, increased ventilatory days and respiratory infections, sepsis, blood loss, and a decreased immunologic response. Older individuals are at greater risk of dying after their burns. Burn patients admitted with either alcohol intoxication or a history of alcohol use tend to have an increased percentage of TBSA burned, higher rate of inhalation injury and mortality as well as longer length of hospital stay and, consequently, increased medical costs in comparison to burn patients who are nonalcoholic users. Alcohol exposure is frequently associated with deaths from cigarette-ignited fires. An association exists between alcohol and/or substance use with suicide by burning and other intentional burn injuries, including assault and self-mutilation. Those who have used both drugs and alcohol are 10 times more likely to die from their burn injuries compared to those who have not used either.

It is sometimes difficult to distinguish the specific drug or if alcohol is involved in a particular burn injury. Illicit drug use is associated with a greater proportion of patients with flame burn and a larger average burn size (except for marijuana in most cases). In addition, these patients have longer hospital and intensive care unit (ICU) stays and rates of intubation but not a higher risk of mortality. In some studies, self-inflicted injuries were more likely to be seen if the patient had used drugs. Illegal drug manufacturing and the use of peculiar drug paraphernalia pose added and somewhat novel risks for various types of thermal injury. For example, freebasing cocaine may cause burns to the face and hands. Smoking crack has led to accidental inhalation of crack pipe “brillo” screens resulting in burns to the trachea, mouth, and esophagus. Methamphetamine laboratory explosions have been shown to inflict blast injury, and smoking in bed while intoxicated can cause flame burn and inhalation injury. Ignited alcoholic drinks are vastly becoming a popular fad, whereby mixed high-proof alcohol drinks are ignited before serving and pose significant risks for facial, chest, esophageal, and other burn injuries.

Receiving an agitated patient to a burn unit who is most likely delirious, in pain, and unable to provide a history makes it difficult to provide accurate, comprehensive care. An astute clinician rules out alcohol or other substance intoxication. Blood alcohol level and urine toxicology screens should be standard tests in the evaluation of burn patients to provide them with optimal care. Timely intervention and initiation of treatment for withdrawal symptoms undoubtedly help to save lives, time, and money. Uncontrolled withdrawal poses significant problems for both patient and staff and utilizes a staggering degree of hospital resources as patients may require any combination of services, including one-to-one observation due to multiple periods of agitation or attempts to leave against medical advice, frequent (sometimes hourly) assessment of vital signs, frequent administration of intravenous (IV) medications and close observation for effects of the central nervous system (CNS) depression secondary to benzodiazepine treatments, and intubation if withdrawal progresses to alcohol withdrawal delirium and is accompanied by seizures. Fluctuating fluid levels, multiple pain medications, anoxic injury, and CNS anesthetics greatly complicate the use of any standard alcohol withdrawal protocol. These patients must be assessed frequently for withdrawal and their treatment adjusted as needed. This treatment may be required for days to weeks, depending on the severity and duration of the withdrawal, which is affected by the age of the patient, the number of drinking years, the level of sustained drinking, the existence and severity of comorbidities, and a history of withdrawal symptoms. Drug or alcohol abuse or dependence can profoundly impact all phases of burn recovery, including the rehabilitation phase. Chronic alcohol use can (1) depress the immune function, leaving compromised patients less able to battle infection, (2) cause malnutrition, when adequate nutrition is one of the most important factors in healing from burn injury and graft surgery, and (3) damage multiple organs such as liver, heart, and brain.

This patient was a male in his 50s, an unemployed musician who lived alone, never married nor with kids. He had relationships though none lasted for more than a few years. He never smoked cigarettes, snorted cocaine and heroin, never IV, and drank alcohol. His alcohol use was a few times a week but had increased during the covid pandemic. He worked from home, had lost his gigs due to the pandemic, and saw very few people. He did have some friends but saw them less and less, especially since he spent all of his time alone these days.

As his life options became fewer and fewer, he thought on more than one occasion as to whether life was worth living. Though he had these thoughts for years, they seemed to come more frequently, especially when he was drinking alcohol. His past psychiatric history was significant for dysthymia, sometimes burgeoning into major depression, though he was never really in therapy and tried an antidepressant two times but stopped both times impulsively.

On the night of the admission, he was feeling particularly low due to his poor financial picture. He drank a six pack of beer followed by a pint of gin, finishing off the night by snorting the remaining heroin he had in his apartment. One of the last things he remembered was heating soup on the stove since he had forgotten to eat all day. In taking the soup off the stove, his shirt caught on fire. Neighbors heard his yells and smelled smoke and called the ambulance.

When arriving at the hospital, it was determined he had 35% TBSA around his face, hands, and chest. Ambulance workers reported that they had seen empty alcohol bottles in his apartment, and he was started on a withdrawal protocol of benzodiazepines. When he was able to be psychiatrically evaluated, he said this had not been a suicidal attempt but admitted to having passive suicidal thoughts. The patient did not think he had a significant substance abuse problem but acknowledged that he had been depressed for some time and consented to a trial of an antidepressant. He was unable to have visitors in the burn unit due to covid precautions but allowed his brother to be called for collateral information. He reiterated some of what the patient had told the team. It was decided that after discharge, he should go to a physical rehabilitation program. From there, he was going to move in with his brother for a while and attend an outpatient clinic to deal with his depression, still refusing help for his substance abuse.

DELIRIUM

Delirium is a highly prevalent neuropsychiatric syndrome among hospitalized burn patients. It is generally caused by the dysregulation of the baseline neuronal activity secondary to systemic disturbances, leading to alteration in the level of attention, awareness, and other areas of cognitive functioning. The onset of delirium is acute, usually developing over a short period of a few hours to a few days, and tends to fluctuate in severity. There are three major types, including the most frequent hyperactive delirium characterized by psychomotor agitation and restlessness.

The hypoactive delirium is less frequent, challenging to diagnose, and presents with decreased psychomotor activity and level of arousal. The third type is mixed, which combines hypoactive and hyperactive features.

The most identified causes of delirium in burn patients are cerebral anoxia, skin infections, fluid or electrolyte imbalance, and withdrawal from opiates and benzodiazepines. Some burn patients with an underlying psychiatric diagnosis can intermittently have the manifestation of delirium erroneously mistaken for a flare-up of their or other mental illness. If unidentified or left untreated, delirium leads to poorer patient outcomes, prolonged hospital stays, increased medical costs, and higher mortality.

The evaluation and diagnosis of delirium in burn patients, as in any patient, include full bedside assessment and measurement of vital signs, including pulse oximetry, ECG, and chest radiography. Comprehensive medical history, including the history of current illness, past psychiatric, medical, surgical, and substance abuse histories, are required. Collateral input is required in most cases because of the confused state of these patients. The confusion assessment method (CAM) remains one of the commonly used assessments for screening. Appropriate laboratory tests, including complete blood count, comprehensive metabolic profile, ammonia and uric acid levels, Vitamin B12, thyroid-stimulating hormone, urinalysis, and drug levels, are assessed. Prolactin and troponin levels are checked if clinically indicated. Radiologic and other investigations, including CT scan, MRI, EEG, cerebrospinal fluid examination, Doppler ultrasound scans, arterial blood gas, and lung perfusion scans, are intermittently indicated to ascertain pertinent underlying medical causes.

Treatment of Delirium in the Burn Patient

While the identification and correction of underlying medical issues remain paramount in the treatment of delirium, ensuring patient safety is also a priority. A delirious patient can present as confused, agitated, or disinhibited. They are also at risk of falls, wandering, hurting themselves or others, pulling out of venous accesses, dressings, other tubing, or equipment. Safety measures, including placement of side rails on the beds or soft restraints, suffice for some patients. Others will require constant one-to-one observation for their safety.

The treatment of delirium in burn patients does not differ greatly from the treatment of delirium in patients with other medical conditions. However, the evidence for definite pharmacologic treatment remains scarce, and guidelines are controversial. For hyperactive delirium associated with agitation, antipsychotics are usually the first line. QT interval corrected for heart rate (QTc) must be checked prior to their initiation and to reduce risks of torsade de pointes and ventricular tachycardia. If the delirium and agitation are perceived to derive from benzodiazepine and alcohol withdrawal, then benzodiazepine constitutes the main treatment than antipsychotic. Other factors affecting treatment choice include underlying cardiac, hepatic, renal, and endocrine illness, dystonic reactions, neuroleptic malignant syndrome, and seizures, among others. The use of IV haloperidol is not approved by the FDA but does not restrict the use in burn units and other critical care settings. It is administered frequently under close cardiac monitoring for effective and rapid control of hyperactive delirium in patients without oral access. The starting doses of IV haloperidol in burn adults and older children are 0.5 to 2 mg. Incremental and frequent doses may be necessary while the patient is closely observed on a cardiac-monitored intensive burn unit bed. Benzodiazepines are recommended if anxiety or extreme agitation dominates the clinical picture or in alcohol withdrawal delirium, but caution is necessary since the addition of benzodiazepines can increase the risk of delirium and should be tapered off quickly.

In children, it is especially important to differentiate insomnia, agitation without delirium, and anxiety from delirium as the former two are effectively treated with benzodiazepines. Some studies had found an association in the use of IV haloperidol in the treatment of acute pediatric burn patients with dystonia, hypotension, hyperpyrexia, and respiratory and renal failure, leading to death in one child. Despite these observations, haloperidol remains frequently used parenterally to control agitation in older burn children for 2 to 3 days with rapid relief. It is essential that such treatment be very carefully monitored. Second-generation antipsychotics (SGA) are more frequently used in the treatment of agitated delirium in adolescents and adults with oral access.

The alpha-2 agonist dexmedetomidine is very useful in the treatment of delirium by shortening its duration, the duration of mechanical ventilation, and the length of stay in the ICU. Another alpha-2 agonist clonidine may also be used. Relief of delirium can be achieved within minutes or hours, but it can persist for days or weeks, especially if the etiology is not correctly identified and addressed.

Alongside psychopharmacology, multiple-component nonpharmacologic interventions are implemented in the prevention and treatment of delirium. These include facilitating and promoting orientation to the environment, for example, allowing the patient who wears glasses and hearing aids to have them available and encouraging them to wear them, placing a visible calendar in the patient's room, and giving the patient straightforward statements regarding the procedures that will be done and the things that are expected from them. Maintaining consistency in the room setting and nursing staff who care for the patient may be helpful measures to help orient the delirious patient to an unusual and possibly perceived threatening environment. Early mobilization, adequate hydration and nutrition, adequate pain management, and maintenance of sleep–wake cycles are some of the other nonpharmacologic interventions used.

PAIN MANAGEMENT

A patient in significant pain has little tolerance for a meaningful conversation; therefore, it is not possible to properly assess the psychological issues such as the reaction to the injury, depression, guilt, body image, or cognitive deficits in such a state. Any effort to talk to the patient without addressing pain status will likely be experienced as insensitive and irrelevant. The proper treatment of burn injuries usually requires debridement, daily wound care, skin grafting, and retraction release surgeries, followed sometimes by weeks or months of physical therapy. These treatments can be extremely painful. Patients have different thresholds for pain, different coping abilities, and different physiologic responses to treatment. For these reasons, it is important to create a multimodal individualized pain control plan for each burn patient as well as to inquire about the adequacy of pain control during each contact with the patient and address it accordingly as is recommended by the International Society for Burn Injury Practice Guidelines.

Background pain refers to the underlying pain from the initial injury. It is ongoing and present even in the absence of activity or procedures. Management of this type of pain is generally accomplished through the use of long-acting analgesic agents.

Breakthrough pain is more intense, episodic, and associated with activities of daily living and other minor activities that require movement or manipulation of injured areas. It is usually managed with short-acting agents.

Procedural pain is associated with invasive procedures and ongoing daily burn care, such as wound cleansing, dressing changes, and physical and occupational therapy. A structured approach to pain management should aim to control all three types of pain. Acute pain disrupts normal physiologic functions and increases stress.

Poor pain control causes sleep disturbances and increased anxiety, which further exacerbate pain. It has been shown that uncontrolled pain has physiologic effects that impede wound healing. In addition to the pain that arises from the burned tissues and donor sites, burned patients may experience additional pain from wound infections, concomitant injuries suffered at the time of the burn, pre-existing conditions, or abdominal pathology related to sepsis or due to heterotopic calcifications.

Psychological addiction is infrequently caused in the context of burn care, but some patients treated with high opiate doses for the long term can develop physiologic dependence, and acute tolerance has been documented in children during the initial phases of intensive care. There is no evidence that opioid addiction occurs more often in burn patients compared to other patients requiring opiates for acute pain. Patients with active substance use disorders who are victims of burn injuries require higher doses of opiates than their counterparts to achieve adequate analgesia. Frequent assessment of the patient’s pain and general condition by experienced burn unit staff and the collaboration of the pain service consultants help to elucidate the appropriate time for changes in the medications’ dosages and forms of administration in an effort to effectively control the pain and at the same time avoid unnecessary exposure to high opiate dosing. Despite the goal of adequate pain control, patients are frequently undermedicated for pain. This can lead to a full range of psychiatric symptoms, which include anxiety, depression, and post-traumatic stress disorder (PTSD).

Although quantifying pain can be difficult, a burn injury must be considered the most painful trauma that a person can sustain. The assessment of pain should be conducted at different phases of care (at rest, during physical therapy sessions, prior, during, and after procedures, whenever feasible) and documented in a standardized manner. Methods to measuring pain include subjective reports, as determined by rating scales, questionnaires, and more objective measures such as physiologic parameters. The visual analog scale, the numerical pain scale, and the Wong-Baker FACES Pain Rating Scale appear among the most used. The Initial Pain Assessment Tool and the Brief Pain Inventory are two useful questionnaires that provide options for the initial evaluation of acute or chronic pain by a comprehensive set of questions on the nature and effects of pain. It has been accepted that on a 0 to 10 scale (where 10 equals the maximal pain), the pain is well controlled if it is patient self-rated between 3 to 5 or less at rest and between 5 to 7 or less during procedures. An unmarked 100-mm scale can also be used, where the score is determined by measuring the distance between the two poles of “no pain” and “the most pain you can imagine.” This visual analog scale translates the subjective experience of pain to a specific number from 1 to 100. An electronic device can also be used to do this. Documenting the pain in a standard manner will allow the team to communicate in a consistent language and to address the pain needs more reliably.

Objective indicators of pain such as sweating, increased pulse, elevated blood pressure, agitation, and increased respiratory frequency could be due to other causes such as hypoxia, hypoglycemia, and alcohol withdrawal, to name a few examples. The Critical Care Pain Observation Tool (CPOT) can be used when a patient is not able to interact or communicate their individual assessment of pain.

Pharmacology of Pain Management

Opioid administration is the foundation of effective burn pain management. The choice of the opioid agent should be based on physiology, pharmacology, and physician experience. Opioids should be tailored to each patient's needs and continuously adjusted throughout their care. Attempts should be made to use the lowest opiate dosing needed to achieve the desired level of pain control. Whenever appropriate, patient-controlled analgesia (PCA) should be utilized. PCA prevents delays in medication administration by nursing staff, leads to more consistent pain relief, and decreases the patient's level of anxiety by providing a sense of control.

Opiates and benzodiazepines are often given together for the management of pain and pain-associated anxiety control in burn patients. It is often observed that burn patients require exceptional doses of opiates and benzodiazepines to relieve pain and anxiety, compared to the standards in nonburned patients, to compensate for their hypermetabolic state. The preferred routes of opiate administration are IV and oral, taking into account that a normal intestinal function is needed to ensure the effectiveness of the oral route. Very close monitoring for respiratory depression is of paramount importance in patients receiving such medications.

Fentanyl (Actiq), a synthetic opioid around 100 times more potent than morphine, is available in injectable, transmucosal, and sustained transdermal presentations. The oral transmucosal form appears promising in addressing anxiety and pain in burned children.

Methadone (Dolophine) is a long-acting opioid with properties similar to morphine. Its *N*-methyl-D-aspartate (NMDA) antagonist properties are postulated to moderate opioid-induced tolerance, hyperalgesia, and neuropathic pain.

Buprenorphine has been increasingly prescribed over the last 10 years in response to the opioid epidemic as a maintenance treatment for opioid dependence. It is available in a sublingual form as Subutex and, to deter misuse, it is also available in a formulation with naloxone as Suboxone. Buprenorphine is a partial mu-agonist with a strong affinity to the mu-opioid receptor.

As a result, it only partially activates the mu-receptor and also makes it more difficult for mu-agonists to fully activate the receptor causing attenuation of other opioids' analgesic effects. In patients on buprenorphine therapy for opioid use disorder, the management of acute pain from burn injuries can become challenging. Recommended approaches include consideration of several factors, including the maintenance dose, the severity of the injury, the expected severity of the pain associated with the procedures the patient will need to endure, and the estimated time of analgesic requirements. If the maintenance buprenorphine dose is low, this can be continued while starting other opioids. Patients maintained on buprenorphine will require much higher doses of a full opioid agonist to obtain analgesia. Patients maintained at higher doses may benefit more from decreasing the buprenorphine dose while starting full mu-agonists. Stopping buprenorphine in order to manage intense pain with other opioids is acceptable as well. However, it poses the issue of opioid withdrawal symptoms during the reinduction of the partial agonist.

Meperidine hydrochloride (Demerol) is not used due to its short half-life and because normeperidine, its major metabolite, can cause agitation, hallucinations, and seizures and may accumulate to dangerous levels with repeated dosing.

Anesthetic agents such as ketamine (Ketalar), propofol (Diprivan), and nitrous oxide are being used more often to facilitate the performance of burn care. Low-dose ketamine can be used as an adjunct to opioid therapy in patients who could benefit from reduced opioid consumption, particularly in the postoperative period. It is important to keep in mind that propofol is a sedative that lacks analgesic properties.

Dexmedetomidine (Precedex) is an alpha-2 adrenergic receptor agonist that can be used as a first-line sedative in the intubated burn patient. Dexmedetomidine and clonidine are recommended as pain management adjuncts in the setting of signs of withdrawal or prominent anxiety symptoms.

Acetaminophen can be used in burn patients while monitoring the maximal daily dose. Nonsteroidal anti-inflammatory drugs (NSAIDs) can be considered in burn patients taking into account individual comorbidities, particularly renal function. Agents for the treatment of neuropathic pain, such as gabapentin (Neurontin) or pregabalin

(Lyrica), can be used as an adjunct to opioids in patients with neuropathic pain or who are refractory to standard therapy.

Antidepressants such as TCAs and SNRIs block the uptake of serotonin and norepinephrine, increasing the activity of descending inhibitory pathways. The effect on pain typically antecedes the effect on anxiety and depression and requires lower doses than those needed for effect in mood.

Psychotherapeutic Strategies for Pain

Providing the burned patient with education regarding the analgesic strategy and its ongoing re-evaluation typically provides reassurance and strengthens the therapeutic alliance. Psychological approaches to pain management can increase the patient's sense of control over the painful burn care procedures. Any psychotherapeutic approach that successfully addresses the patient's anxiety, depression, guilt, acute stress syndrome, or other psychological issues stands a good chance of reducing the patient's pain experience. Some randomized prospective small-sample studies using adjunctive psychological techniques in the burned patient utilizing pain and anxiety self-report, as well as observer, rated measures have been conducted.

Ideally, burned patients should be offered a nonpharmacologic pain control technique. Cognitive-behavioral therapy, hypnosis, and virtual reality have the strongest evidence.

In order to use cognitive interventions for the adjunctive treatment of pain, it is important for the clinician to recognize the patient's coping skills. Patients who cope with avoidance will be most likely to benefit from distraction during short procedures and deep relaxation and distraction imagery during longer procedures. Patients who focus on the procedure may benefit from reappraisal techniques. They should be encouraged to differentiate sensory from affective components of pain and evaluate its meaning. They usually benefit from being told that pain sensation is a positive sign of wound healing. Distraction therapy using pleasant videos along with administration of analgesics during burn dressing changes has demonstrated significantly reduced pain and anxiety. There is also some supporting evidence that the use of virtual reality provides analgesia with minimal side effects and little impact on the physical hospital environment. Virtual reality takes distraction one step further by using a visual stimulus instead of the power of suggestion. Visual stimulus is provided through a headset that covers the eyes. An auditory stimulus can be provided through associated earphones. A controlled study using virtual reality showed a reduction in pain, nausea, and anxiety during physical therapy. The management of burn pain is multidisciplinary and dynamic as the burn wound pain is essentially multifaceted. Several psychological approaches have been proven effective as adjuvants of the pharmacologic treatment, but larger studies are still needed. Pain management needs to be individualized to each patient’s needs and continuously monitored for optimum control.

Hypnosis and Pain

Hypnosis can help to control pain during burn wound debridement and is a viable adjunct treatment for burn pain. Hypnosis can be used in the short term. It has been reported that in less severe burns, hypnosis may help prevent the postburn inflammatory response and reduce burn severity. Hypnosis is also helpful as an adjuvant in the treatment of pain and pain-related anxiety. During burn dressing changes, anxiety appears to be better controlled by hypnosis than by stress-reducing strategies. Hypnosis is particularly valuable when there are medical contraindications to adequate narcotic control of pain.

Relaxation Techniques

It is widely accepted that relaxation enhances the effectiveness of analgesics by reducing muscle tension and increasing blood flow. Breathing exercises, guided imagery, and deep muscle relaxation are all examples of practical bedside interventions to alleviate tension and enhance the patient’s sense of control and well-being.

BURN INJURY AND PSYCHOPATHOLOGY

Through the stages of acute treatment to rehabilitation and reintegration, burn patients contend with multiple stressors, including severe pain, disfiguring scars, loss of functionality, and other distress. Some of these factors lead to enduring psychosocial and psychological difficulties. Greater psychosocial distress during acute hospitalization for burn injury has been shown to result in a more difficult rehabilitation and recovery. As recent major medical and surgical advances have greatly increased the survival rates of burn patients, there has also been increasing interest in their psychiatric and functional sequelae.

Depression and PTSD, prevalent in 13% to 23% and 13% to 45% of cases, respectively, have been the most common areas of research in burn patients. In the case of depression, risk factors have been found to include premorbid depression and female gender in combination with facial disfigurement. For PTSD, premorbid depression also as well as type and severity of baseline symptoms, anxiety related to pain, and visibility of burn injury are all contributing factors. It is also noted that the emergence of acute stress disorder (ASD) and/or PTSD symptoms in the first few weeks of hospitalization can be a predictor of the subsequent development of chronic PTSD. However, PTSD symptoms can singularly occur months after a burn injury has occurred in a significant minority of patients who had never exhibited symptoms of ASD or PTSD during acute hospitalization. Both neuroticism and avoidant coping styles are important personality dimensions linked to the development of PTSD as well.

Great levels of acute pain have been associated with negative long-term psychological effects such as ASD, depression, suicidal ideation, and PTSD for as long as 2 years after the initial burn injury. Subjective reports of feelings of “alienation and anxiety” experienced during hospitalization are also linked with distress for up to 2 years in burn patients.

Burn patients can experience difficulties in sexual life and social interactions. Even burn patients who never develop chronic psychological sequelae following burn injury will frequently experience, especially early in recovery, episodic expressions of stress, depression, dysthymia, acute stress disorder or PTSD, other anxiety disorders, or psychosis. The quality of life of burn patients initially seems to be lower in comparison to the general population; however, over a period of many years, quality of life normalizes or becomes rather good.

The treatment of psychopathologies associated with burn injuries does not differ greatly from the treatment of the general population with similar ailments. The psychiatrist on initial assessment must screen for underlying risk factors including premorbid psychiatric disorders, personality characteristics, peri- and post-traumatic factors, as well as proneness to shame, body image problems, and self-esteem issues in case of patients with disfigurement. Psychotherapy (cognitive, behavioral) and pharmacologic treatments (SSRI, SNRI, etc.) are valuable in the treatment of depression. Exposure therapy and eye movement reprocessing and desensitization have been successful in the treatment of PTSD as well as available pharmacologic treatments. Additional measures including social skills training, community interventions, sexual health promotion, and counseling have all been found helpful in enhancing the quality of life of these patients.

PSYCHOTHERAPEUTIC INTERVENTIONS

There has been a wide variety of psychotherapeutic modalities used in treating patients who have suffered a burn injury. They include cognitive behavioral, supportive, brief and short-term psychodynamic, and exposure therapies, debriefing, eye movement desensitization, and psychoeducational interventions, all of which have reportedly been used with varying levels of success. Authors reviewing general trauma treatment were rather unimpressed by any of the myriad modalities, including medication therapy. However, exposure therapy garnered substantial research results as a relatively effective treatment for ameliorating some PTSD symptoms. Pain control, as previously discussed, may be one of the most crucial treatment modalities for adults and children because, if effective, it may reduce the potential for developing post-traumatic symptoms. The choice of therapeutic intervention is often dependent on the stage of recovery from burn injury.

Acute Phase

During the acute phase, the burned patient faces multiple immediate stressors, including fear of death, acute pain, being in an acute care environment such as the emergency room and ICU (burn unit), and concerns for other victims of the same burn event. The patient may experience delirium manifested by fluctuations in the level of consciousness, confusion, agitation, and disorientation. The delirium could be related to the acute physiologic effects of the burn itself, alcohol withdrawal, metabolic complications, infection, medication, or substances of abuse. At this stage, psychological interventions focus on immediate concerns such as pain control, anxiety, management of delirium, and sleep. The priority at this stage is survival; therefore, in-depth psychological interventions are not indicated at this point.

The initial psychiatric assessment offers patients an opportunity to recount the burn event if they are psychologically prepared to do so. As patients reveal their understanding of their injuries, the consulting psychiatrist, in concert with the treating surgical team, orients patients and families to burn unit protocols and prepares them for the journey that lies ahead. A crucial part of providing psychiatric care to a burned patient is building a therapeutic alliance with their family or support system. If the patient is unable to provide information, the initial history can be provided by collaterals, which can serve as an opportunity to facilitate the family's understanding of the injury and its treatment course. The presence of distressed family members may enhance the patient's anxiety. Efforts to moderate caregiver stress may also serve to lower a patient's anxiety and stress reaction.

Chemical neurologic embedding of traumatic memories is thought to involve sympathetic nervous system release of norepinephrine. Adequate treatment of acute pain and anxiety symptoms after burn injury may prevent some psychological sequelae by moderating the serum levels of norepinephrine and possibly other neurotransmitters and endogenous hormones associated with the stress response. Historically, treatments in the acute phase have also involved pharmacologic methods such as oral and IV administration of opiates and benzodiazepines for pain and anxiety, selective serotonin reuptake inhibitors (SSRIs), tricyclic, and monoamine oxidase inhibitor antidepressants, beta-blockers, and D-cycloserine in conjunction with exposure therapy. The antihypertensive prazosin (an alpha-adrenergic blocker) is being used to treat traumatic nightmares related to various types of trauma, including combat events. It is reported to improve sleep onset and, if taken during the day, has been reported to reduce irritability and other hyperarousal symptoms in some. In one literature review, prazosin doses varied between 1 and 16 mg, though higher doses have been needed to achieve remission of nightmares and have been used with positive effects. Researchers continue to examine the role of neuroactive steroids in the development and treatment of PTSD and the biochemical or genetic individual differences that may underlie variable responses to pharmacotherapy. Benzodiazepines have not been found effective in treating symptoms of PTSD, although they are widely prescribed for anxiety and agitation. Some evidence is cited in a recent literature review to support the use of pharmacologic intervention in reducing the risk for adults of developing PTSD after trauma, specifically hydrocortisone administered immediately after trauma exposure. In the same review, researchers found no supporting data for propranolol, escitalopram, temazepam, or gabapentin in the prevention of PTSD. Critical stress debriefing used in the early treatment of traumatic stress has been shown to have neutral or harmful effects.

A 55-year-old male patient with no visible wounds sustained an electrical injury from contact with an underground power line. A psychiatric consult was ordered on day 1 of this patient’s hospitalization for repeated signs and symptoms of anxiety and panic, including increased blood pressures, diaphoresis, subjective report of nightmares, and awakening in a cold sweat, “feeling in a daze.” The patient associated the onset of symptoms with the sounds of the arrival and departure of the hospital medical transport helicopter. He had been transported to the hospital by helicopter after suffering electrocution. The patient was relocated within the burn unit, and lorazepam was ordered to be given as needed for anxiety. Interestingly, once the source of his reaction was identified, he reported a reduction in anxiety while awake, even without medication. By subjective report, benzodiazepines helped this patient achieve restful sleep with a decrease in startled nighttime awakenings and nightmares.

A 34-year-old man came to the burn unit following a chemical burn on his hands and legs that occurred at work. He complained bitterly about his care and the various policies of the burn unit to the embarrassment of his wife. The staff frequently defended the care that he received and argued on behalf of the doctors. Negative feelings about the patient were discussed at psychosocial rounds with the mental health team. This patient believed that the statements that he was making were true, and he would increase his attacks if the caregivers attempted to argue with him. The approach with this patient was to acknowledge his beliefs but not to offer an opinion about them. Occasionally, asking the patient for his patience with difficult situations steered him away from being defensive and put him in a more cooperative mode. As soon as possible, he was taught to do his own dressing changes and allowed to take showers to help him feel more in control and reduce his reliance on staff members, whom he did not trust. The family was counseled in an effort to help them understand that the patient's complaints may have stemmed from his fear and sense of helplessness, more so than with the quality of care.

Intermediate Phase

Once acute medical issues have been stabilized, increasing the likelihood of survival, the patient enters the intermediate phase. The focus of treatment shifts toward restorative care. Most of the burn wounds have been grafted and are in various stages of healing, and ventilator support, sedation, and severe infections become less prevalent. However, patients continue to undergo painful procedures. They become more aware of the physical and functional impact of their injuries and, at the same time, are more psychologically ready to participate in a psychiatric diagnostic assessment. Anxiety and depression-related symptoms often begin to appear in this phase. Sleep disturbances can be related to anxiety, depression, pain, and the noises and routines of the intensive care setting. Symptoms and behaviors related to premorbid psychopathologies such as personality disorders and substance abuse may surface in this stage. A dysfunctional pattern of coping of the individual can have an adverse impact on outcomes, including increased length of hospital stay and more serious psychopathologies after the injury. During the intermediate phase, the grieving process is more evident. Burn patients may have lost family members, friends, pets, their home, or personal property, as well as aspects of their former life: work, appearance, and physical abilities.

Optimum pain control facilitates the patient’s ability to participate in treatment protocols and allows for the development of a sense of hope for recovery. As discussed above, psychological interventions are helpful adjuvants to appropriate psychopharmacologic treatment in the management of pain and anxiety. Cognitive-behavioral therapy and hypnosis have proven to be effective in treating procedural pain. Virtual reality utilizing special computerized programs is used as a distraction technique during painful procedures. A recent pilot study using a manualized cognitive-behavioral treatment protocol developed specifically for treating PTSD and psychosocial sequelae following burn injury resulted in improvements in PTSD and depressive symptoms, self-image, and aspects of reintegration.

Brief supportive psychotherapy to treat depression and anxiety in combination with appropriate psychopharmacologic interventions is also helpful in this phase in managing depression and anxiety. Informing patients that nightmares are common and usually subside over time can help reduce stress and anxiety, increasing patients’ understanding of the psychological healing process. Other forms of psychotherapy utilize cognitively oriented psychoeducation. Large burn centers may offer a variety of rehabilitation support groups for adults, children, adolescents, and parents, using various modalities, including expressive and cognitive-behavioral therapies. The focus of interventions during the intermediate phase is the assessment of emotional and cognitive adaptation to the injury, treatment and surgical procedures, and readying patients for transfer to rehabilitation care or discharge home and return to school or work. The importance of psychoeducation and preparation for future care and other postdischarge issues must be highlighted. Knowledge of and engagement in the planning of upcoming surgeries, for example, gives patients a sense of control as they anticipate and prepare for the experience. Supportive treatments and normalizing the emergence of psychiatric symptoms related to the trauma can be very helpful during this and all phases. Appropriate mental health care referrals should be made at discharge for surveillance and treatment of current or evolving psychological concerns.

Recovery Phase

The recovery phase usually begins at the time of discharge from the hospital, when the patient begins to reintegrate into society and is marked by the promise of survival. Patients with severe burns may still require long-term outpatient physical rehabilitation, including painful procedures such as dressing changes and surgeries. Ongoing sufficient nutritional intake is still imperative for healing skin and grafts. During the recovery stage, patients have significantly more control over their treatment than while acutely hospitalized. The timing and extent of plastic and reconstructive surgeries are collaborative decisions made by the patient and surgeon. The duration of this phase depends on the extent of impairment and the need for corrective surgeries. A patient may require multiple surgeries over a span of several years to release scars and improve function or restore appearance. Forty-six percent of patients presenting for these reconstructive surgeries were found to suffer from mild to moderate depression. Body image dissatisfaction was predictive of increased depression scores.

Psychological challenges in this stage include regaining a sense of competence while adjusting to the practical limitations of the injury. Psychological considerations during this phase include body image changes, self-esteem issues, adjustment to functional impairments, and depressive and anxiety disorders, particularly PTSD and phobias. Ongoing issues of grief, guilt, and premorbid psychiatric conditions continue to require psychotherapeutic support.

Social stressors in this stage include vocational rehabilitation, readjustment to work, family strains, and lack of social support, among others. Appropriate interventions of this phase involve psychiatric outpatient follow-up, social skills training, support groups, peer counseling, vocational counseling, and substance abuse relapse prevention.

Functional and cosmetic outcomes following burn injury have dramatically improved over the past few decades. Evolving techniques of cell culturing have resulted in improved skin grafts, and advances in plastic and reconstructive surgery have reduced permanent disfigurement. With the advent of specialized burn and trauma units, climbing survival rates of even massively burned individuals has placed significant emphasis on the recovery phase.

Factors Influencing Recovery. Patients influence the rate of progress during rehabilitation by their degree of participation and cooperation. If they are or become unmotivated, anxious, or depressed, outcomes will be less than optimal. One study related larger burns and more psychological distress (during acute hospitalization) to more physical impairment and a slower recovery. In addition, a greater subjective psychological burden was associated with slower psychosocial recovery and more postburn psychological impairment. Patient participation in recovery and rehabilitation programs can be maximized by adequate attention to pain and anxiety control. Treatment plans are more successful if premorbid physical, developmental, and psychiatric status are assessed in order to set realistic goals for recovery.

Three important factors affecting burn recovery are pain, brain injury, and overall psychological adjustment. Uncontrolled pain has been linked to increased anxiety and can limit motivation, affecting optimal recovery. Brain injury potentially limits full participation in rehabilitation efforts and can involve periods of impaired impulse control and agitated delirium. Psychological distress also affects motivation and participation. In this regard, whenever appropriate, there should be some effort to explore the question that nearly all patients ask themselves: “Why did this happen to me?” Burn injuries, like many traumatic experiences, can be a turning point for some people, demanding major internal and external shifts in priorities, self-image, relationships, and life goals. Useful psychological information can be obtained by asking the patient, “Do you have any thoughts or fantasies about why this might have happened?” The answer to the question may present a starting point in the therapeutic process.

Although many burn victims sustain only small physical scars, due either to limited injury or to the skills of plastic and reconstructive surgeons, some studies have found that the size of the burn is not directly related to the degree of periburn emotional trauma or later development of psychopathology. For example, small burns have been shown to result in disproportionately severe psychological trauma, especially if the injury involved hands (decreasing functional capacity), face (altering self-image and increased visibility of scarring or disfigurement), or perineum (complicates elimination and sexual activity/identity).

A 20-year-old Hispanic female with prior history of PTSD and substance abuse suffered 65% total body flame burns primarily involving her upper body, including face and hands. Drug dealers had set the house on fire as she slept inside. Acute hospitalization lasted for approximately 1 year, and she required a chemically induced coma for over a month while fragile skin grafts healed, and plastic and reconstructive surgeons attempted to repair her ears and nose. Several years after discharge, she continued to undergo reconstructive and release surgeries to remove keloid scars. During this recovery phase, the plastic surgery department, physical therapy office, mental health clinic, and Alcoholics Anonymous (AA) became alternate homes for her, where acceptance, hope, and structure were provided. At a point in her recovery, she gave meaning to this traumatic experience by using it in a constructive way. She began speaking to groups of adolescents about the realities and dangers of substance abuse. This case exemplifies much of what has been discussed in this section, including the relationship of burns to substance abuse and premorbid psychiatric disorders, advancements in trauma and surgical treatments and survival of massive burn injury, the importance of psychiatric involvement at all phases of burn injury, psychological adjustment to new body image, and finding meaning in a life-altering traumatic event.

THERAPEUTIC APPROACH TO GRIEF, REGRESSION, AND ANGER IN THE BURN PATIENT

Grief

Grief work is necessary as any patient who suffered a burn has sustained some kind of loss: appearance, function, role, independence, sense of safety in the world, possessions, or even loved ones. In facilitating grief work, it is useful to have in mind certain principles:

1. *Acknowledging the loss:* It is useful to prompt the patient to look at the losses by bringing up the topic that family and others often avoid. Most patients are willing to talk about their losses, and providing the patient with an opportunity to do so may promote an outpouring of emotions that the listener should be prepared to hear.
2. *Allow the patient to be with pain:* The pain of the loss is not eliminated with words. It is preferable to avoid rationalizations about “the good that will come of the loss,” which often makes the caregiver feel better but not the patient. More helpful interventions are focused on providing the patient with space and time to convey the full power of their emotions and to feel heard by an empathic listener. After the patient has amply expressed their feelings, it is useful to provide reassurance about the gradual subsiding of the emotional pain with time.
3. *Affirm the individual’s experience as real and part of normal grief:* Somatization, preoccupation with the lost object, guilt, hostility, and changes in normal patterns of conduct are anticipated components of a normal grief reaction.
4. *Provide assurance by exploring past coping patterns:* Burned patients may perceive their loss as unbearable and may be helped to use their coping skills by being encouraged to think of how they dealt with other losses in the past, if any. This process requires careful probing as the cumulative pain of prior losses not fully experienced may be rekindled.
5. *Monitoring danger signals:* Persistent thoughts of self-harm, neurovegetative symptoms of depression, failure to make progression in the grieving process, and exacerbation of previous mental illness should prompt pertinent interventions. These include increasing the level of monitoring, adjusting psychotherapeutic and psychopharmacologic approaches, and, if necessary, admitting the patient to an inpatient psychiatric treatment.

Regression

A common behavior noted among burn patients is regression. The patient, feeling overwhelmed by the physical and psychological distress, may attempt to return to the comfort of earlier stages in life. Staff and family members may have negative reactions to this phenomenon. In managing regression, it is useful to:

1. Recognize the regressed behavior as a coping mechanism.
2. Examine the caregivers' responses and provide them with necessary feedback.
3. Determine if the regressed behavior interferes with the immediate patient care goals and decide the proper intervention accordingly. Behavioral modification techniques with rewards for achieved goals are effective in some patients.
4. Develop a consistent plan to handle regressed behavior. Such a plan needs to be clearly explained to the patient and their family, allowing frequent communication.

Anger

When anger begins surfacing during interactions with medical or mental health staff, it is useful to acknowledge this expressed emotion with a supportive approach instead of attempting to talk the patient out of any feelings they have. It is useful to pay attention to the interactions with family members, who usually feel uncomfortable with expressed negative emotions, and try to redirect the patient to avoid talking about them. Some family members may express their own anger, fueling the patient's emotions.

Although nursing staff can handle and provide opportunities for the patient's expression and abreaction of angry feelings, it is sometimes advisable that a psychotherapist talks with the patient about their feelings and meets with the patient regularly, even for a short period. On the surface, patients' anger is directed outward toward causes of the accident, missing work, being in pain, being hospitalized, and perceived inconsiderate treatment by hospital staff. Patients often avoid expressing anger toward the hospital staff as they depend on them. The ultimate object of the anger may be “self.” Patients blame themselves for being in the situation in which they were burned whether or not there is validity to this idea. This is not easy to accept, and the patient may displace their anger toward themselves or toward family members. They may push away their family and even their health care team by utilizing acting out behavior, projection, and projective identification. A differential diagnosis in terms of potential causes of mood changes or mood disorders needs to be considered and pharmacologic interventions implemented when needed. Patients are really angry at themselves for either their role in the burn injury (real or imaginary), their lack of ability to avoid the situation, losses they “caused,” for not recovering fast enough, for needing pain medication, for not tolerating the hospitalization, and so forth. As these feelings are unconscious and unacceptable to the patient, they are displaced onto external objects, sometimes in excess.

In this setting, the most successful interpretation is that allowing patients to realize that they are really angry at themselves while blaming others. However, it is very important to be aware that taking away the defense against inwardly directed hostility can elicit profound depression and even suicidal ideation. Precautionary measures need to be in place. Supportive techniques may be used here to express respect for the patient and for their acknowledgment of true feelings. Another technique is that of an inexact interpretation that can be used to diffuse the displacement of anger toward staff. Using objects directly or indirectly related to the burn injury such as “the malfunctioning furnace,” “the drunk driver,” and so forth, rather than focusing on the patient’s guilt and self-blame, is more likely to be accepted by the patient.

Expressive and abreactive techniques, time-limited psychotherapy, and psychodynamic life narrative treatment approaches may be successfully applied in the treatment of burn patients to allow verbalization of emotions, examination of separation, and acceptance of the meaning of the injury in the context of the patient’s life trajectory.

PLASTIC SURGERY AND FACE GRAFTS

The overall incidence of burn injuries has declined during the recent decades in the United States, but the proportion of survivors of large burn injuries has increased, carrying with it more severe aesthetic and functional sequelae. Facial and body disfigurement as a result of burns are known causes of profound shame, stigmatization, and isolation. Deep burns frequently result in loss of function and damage of cosmetically important body areas. Many of these burn victims suffer long-term morbidity as a result of functional impairment and severe facial defects.

Deep burns can involve not only the skin but also adjacent tendons, cartilage, and bone structures, contributing to complex deformities. The defects that result from burn injuries include hypertrophic scars, contractures, disorders of pigment, and loss of anatomic parts. Burn reconstructive surgery attempts to help the burn survivor to recover, as much as possible, normal appearance and functioning. There is still a need to identify and develop novel strategies to treat burn deformities in order to obtain improved outcomes. Early closure of the wounds translates to less severe scarring, lower mortality from infection, lower mortality from other complications, decreased stress, decreased hypermetabolic response, and decreased bacterial load.

Burn patients require multiple procedures arranged over months or years to address the deformities. Empathy from staff with the experience of being burned can assist the patient in adapting to a new appearance. Burns may profoundly affect body image, self-esteem, and interpersonal relations, as well as ego functioning. Skin allografts transplanted between genetically nonidentical individuals of the same species mostly come from cadaveric sources but can also come from close family members. Close relative intermingled skin allografts along with autografts have been satisfactorily used in a small number of child and adult burn victims. Among other advantages, this technique allows the active participation of family members in the treatment of the burn victim, creating other interesting dynamics and ethical issues. Synthetic acellular skin substitutes have been used in the treatment of burns, which have the advantage of being readily available for immediate coverage of the wound.

Face Reconstructive Surgery

The human face is a functional, dynamic, and esthetic organ with critical roles in individual development since birth. There is evidence supporting that newborns have an innate ability to recognize individual human faces shortly after birth. The face is ingrained in the concept of one’s identity and comprises a very individualized set of tools that allows the human being to effectively communicate and relate to others. In a patient with facial disfigurement, effective communication is impaired due to self-perception but also as a result of verbal and nonverbal responses of the nondisfigured. In addition, the face includes basic bodily functions, such as breathing and eating. The optimal result of facial soft tissue reconstruction should achieve aesthetic balance and symmetry, distinct facial planes, skin texture appropriate for camouflage, and facial animation, as well as functionality. The deformity caused by massive facial burns is an indication for the consideration of facial allograft transplantation. Partial face transplantation became a reality in recent years, with the first two partial facial transplants performed in France (2005) and China (2006). The first total facial transplant took place in Spain (2010). To date, approximately 40 face transplants have been completed across the world.

The goal of facial transplantation is to enhance aesthetic, functional, and psychological quality of life instead of improving survival. Facial transplantation requires the presence of a multidisciplinary team composed of plastic and transplant surgeons, psychiatrists, psychologists, ethicists, social workers, radiologists, infectious diseases specialists, and nephrologists. From the pretransplant phase to long-term post-transplant support, psychiatry has a very important role as a part of the facial transplant multidisciplinary team. The goals of the pretransplant psychiatric evaluation include the selection of motivated patients, discussing options besides face transplant, including risks and benefits of transplantation, educating the patient about the expected success rate and rescue procedures, immunosuppression, and diagnosing substance use and psychiatric disorders requiring treatment. During the evaluation, the psychiatrist explores the patient's motivations and expectations regarding transplant, their psychological strengths, coping style, self-esteem, body image, and identity issues, the functionality of the support network, and in detail assesses the capacity to consent for the surgery. Following the surgery, the psychiatrist provides ongoing care, which includes psychopharmacologic and psychotherapeutic intervention as the patient goes through the process of integrating the donated face into their self-schema.

Ethical Considerations of Face Transplantation. Despite the significant advances in technology and surgical technique, facial transplants remain controversial due to multiple ethical concerns surrounding the procedure. These include the rationale behind the exposure of individuals to the sequelae of lifelong high-dose multidrug immunosuppression, the development of waiting lists for non–vital organ transplantation, the risk of potential immunologic rejection of the facial graft with its potentially devastating consequences, the consent of the donor's family, the cost to society of such highly technical procedure and the potential damage caused to other organs of the donor during the face harvesting period. Results, however, show that from a clinical, technical, and immunologic standpoint, facial transplantation has achieved functional, aesthetic, and social rehabilitation in severely facial disfigured patients. Early involvement of a hospital bioethics committee is advocated to ensure compliance with ethical standards in this procedure that requires research in order to establish the risk of experimental interventions.

FAMILY SUPPORT

Family members and loved ones of critically injured burn patients face great strain in all phases of treatment. Worst still if they witnessed the accident and ensuing painful medical treatments. The health and mental well-being of family members are frequently impacted as well. In the course of acute or initial phase treatment of the burned patient, it is important to regularly provide their family with updates on the patient's condition and evolving treatment plans as information becomes available. Family members may be anxious and distressed while observing the patient undergo treatment which fosters the same response in the patient. It is important to help family members understand this effect and help them to convey a sense of calmness and hope to the patient. If possible, burn unit staff should describe to patients' loved ones any changes in the patient's appearance resulting from burn injury or treatment. This will help them prepare and process the information related to the extent of the burn injury and the impact of the injury on the patient's functioning in the future. Patients with an extensive burn injury can display disconcerting behaviors related to states of pain or delirium, and they may be agitated, anxious, or confused during family visits. In response, loved ones may make unrealistic demands on burn unit staff. The restrained patient often elicits great concern from family members. Taking the time to listen to their concerns and to describe the reasons for restraints (or 1:1 supervision), the expected course of treatment and recovery, as well as the likely etiologies of the patient's current states may help to ease their distress.

Once the family understands the nature of the injury and its treatment and begins to feel supported and reassured that staff is doing all they can, the family can often help calm the patient. Denial, and other defensive mechanisms, can be employed by patients and loved ones in the setting of acute and critical injury and should be respected as a method of psychological protection and a phase of healing. It should also be noted that the long-term consequences and disability after burns place a considerable strain on burn survivors and their families.

The presence of mental health staff can be helpful to family members in all phases of recovery. Psychiatric consultants should be present and available and introduced to patients and family members as part of the treatment team. This approach may help to destigmatize mental health involvement in the burn unit.

SELF-INFLICTED BURNS

Self-inflicted burns are rare and horrific forms of burn injuries. In the United States, these types of burns account for 1% of burns. In developing countries, they account for 25% to 30% of burn unit admissions. In the United States and other developed countries, these individuals usually have an increased TBSA percentage resulting from flame and inhalation injuries. They have a past psychiatric history, worse physical and psychological outcomes, and are more likely to be men. Additional risk factors are alcohol and substance abuse, stress, impulsivity, previous suicide attempts, low economic status/being homeless, single marital status, and chronic medical issues.

The prevalence of self-inflicted burns increases in some developing countries and within certain cultures, though the incidence can vary within continents, countries, and cultures. In developing countries, the incidence of self-inflicted burns is 10 times greater than in high-income countries. These burns are usually perpetrated by women who are uneducated, unemployed, and from rural areas. The women are often victims of domestic violence or financial stress, living in a male-dominated society. Other contributing factors are financial hardship, history of suicide attempts, marital/family conflicts, and feelings of guilt and shame.

Self-immolation is the act of setting oneself on fire with a flammable liquid in pursuit of a cause or statement. This dramatic action can be seen throughout history. Though this can be seen in Greco-Roman mythology, there is a deeper history associated with religion in Asian cultures such as Buddhism and Hinduism. More recently, self-immolation has been a sign of protesting political movements and wars such as the Vietnam War and Arab Spring. The mode of self-immolation has changed over the years from funeral pyres to gasoline.

Compared to survivors of unintentional burns, those with self-inflicted burns suffer more severe injuries with higher rates of morbidity and mortality and require longer periods of acute hospitalization. Injuries usually affect the upper body, including the face, chest, arms, hands, and ventral thighs. Self-inflicted patients are younger, unmarried, male, unemployed, burned by fire/flame, and substance abusers. They also have minimal social support, low socioeconomic status, poor coping skills, and previous self-harm experiences. In a prospective study looking into long-term physical and mental health morbidity and mortality of self-inflicted versus non–self-inflicted burn injuries, a surprising finding was that the patients with self-inflicted burns had increased all-cause mortality during the 24-month period after injury. The severity of these burns may be due to the deliberate use of liquid accelerants on clothing, which keeps the skin in contact with the fire. In addition, the person who intends to die by fire is less likely to extinguish the flame, compared with victims who suffer burns by assault or accident.

There is a paucity of written/described literature on nonsuicidal burn patients, those who burn themselves without the clear intent to kill themselves. In this population, the majority of which are women, injuries are due to scalding, chemical, or contact means of burning. The distribution of burns is the forearms and lower extremities. The TBSA percentage is less than suicidal-inflicted burn patients, and their psychiatric history is more likely due to personality disorders.

Patients who attempt suicide, especially by such unequivocal and aggressive acts, can evoke extreme emotional responses from others, including from medical providers and loved ones, the individuals responsible for the direct care and support of the patient. Some caregivers may even feel it is the right of the patient to die, given the extremes to which they have gone. These emotional responses, among others such as anger and anxiety, may affect the care and survival of the patient. Treatment of the psychiatric and emotional needs of the patient, family, and, at times, staff, is imperative.

The process of discharging patients with a recent suicide attempt and severe burns challenges all hospital services involved in arranging placement and aftercare. Inpatient psychiatric units vary in their ability to absorb severely burned patients, given the complex nature of the injury and the need for medical and surgical follow-up care. The creativity and adaptability of the burn unit staff, including psychiatry, may make the difference between a timely discharge and the patient needlessly remaining on the burn unit for months. Inpatient psychiatric units are often reluctant to manage the medical follow-up care of severely burned patients. Rehabilitation facilities are reluctant to admit patients with complex psychiatric needs and a recent suicide attempt. Some suggest that the best approach would be to establish secure psychiatric units within rehabilitation centers.

Self-inflicted burn injury continues to present a complex medical and psychiatric challenge to all involved and requires a coordinated and multidisciplinary approach. Recent studies acknowledge the paucity of quality research data related to self-inflicted burns and suicide attempts by self-immolation. Larger, controlled, prospective studies are needed to further elucidate factors that might lead to prevention and treatment modalities. There is evidence that, with appropriate treatment, patients with self-inflicted burns can go on to live productive lives.

CHILDREN AND BURN INJURY

Burns are a major source of morbidity and mortality in children. Almost one-quarter (24%) of all burn injuries occur in children under the age of 15. According to WHO, globally, more than 95,000 minors die of burns each year or approximately 262 children per day. Table 27.14–1 lists some of the risk factors for burns in children. They are usually an interplay of all or some of the listed factors.

The 2019 NBR reported that in the United States, the two most common etiologies for burn injuries across all age groups are flame burns at 41% and scalds second at 31%. Scalds are more prevalent in the younger population, especially among those younger than 5 years of age, while fire/flame burns prevail in the remaining age categories. Most burn injuries are accidental and are the fifth leading cause of injury deaths for children from ages 1 to 4 in the United

States, and third for those aged 5 to 9. However, child abuse or neglect also causes about 6% to 20% of pediatric burns, with scalds as the prevalent etiologic factor. Child abuse or neglect must be ruled out by pre-evaluation and treatment of the burned child. The nature of the burn wounds is intermittently indicative of abuse, as listed in Table 27.14–2. Other indicative factors of abuse include large tap water burns, delay in seeking care as well as if the history and description do not appear to match the injury pattern. Parent–patient interaction might also raise suspicion of abuse or maltreatment.

Underreporting of abuse is common. Another adult, besides the perpetrator of the abuse, often presents the child for treatment. Child victims of abuse are reluctant to present for treatment for fear of legal, physical, or emotional repercussions. Children generally want to remain with their primary caregivers at all costs. It is critical to recognize the signs of abusive burn injuries because they are shown to have increased mortality, require longer hospitalizations, and are often more severe than accidental burns. The average rate of burn injuries decreases from age 4 until adolescence, and after age 15, the incidence rate of burns increases again, presumably due to greater exposure to hazards, experimentation, risk-taking behaviors, and employment. The two major determinants of survival in children are burn size and the presence of inhalation Injury.

Psychological Sequelae of Burn Injury in Children

Psychological distresses are frequent and debilitating complications of the postburn injury. Recent research into the prevalence of post-trauma reactions in pediatric patients after sustaining a burn injury identified that 35% were diagnosed with at least one psychological disorder. Sociodemographic factors and clinical characteristics of the burn injury are some of the major factors influencing psychological symptoms. Among the sociodemographic factors, high living expenses, low income, and female sex appear to be good predictors of psychological symptoms after burn injury. Some studies claim that compared to sociodemographic variables, clinical characteristics such as the degree of the burn and duration of the burn have a greater effect on the development of psychiatric disorders for burn patients. Greater levels of acute pain have been found in studies to be associated with negative long-term psychological effects, including ASD, depression, suicidal ideation, and PTSD for as long as 2 years after the initial burn injury. For children, particularly, inadequate pain control and lengthy periods of hospitalization, coupled with long periods of separation from primary caregivers, can exacerbate the developing, often fragile, psychological and ego stability of children. The psychological development of burned children can be affected by a decrease in their ability to form adaptive behaviors. Adaptability is a necessary emotional tool for adjusting to life's challenges at any age and can ultimately influence maturation, which may, in turn, affect behavior.

Common assumptions exist about the negative effects of disfiguring burns on a child's general attitudes. One study evaluating attitudes of adolescents burned as children contradicted these beliefs and found them to have positive attitudes about their appearance and about how others viewed them, as well as good quality-of-life assessments.

Acute Stress and PTSD Symptoms in Children

The burn injury threatens the child's health and bodily integrity, which results in psychological trauma in the child. Furthermore, burn injury may result in permanent scarring, limited functionality, and intensive and long-standing physical treatments, all of which may place the affected children at the risk of psychiatric disorders. A significant number of burned children present with ASD during hospitalization and typically occur within the first month of injury. It is the core feature of a child's reaction to a traumatic event. The risk factors for ASD include burn size, high pain level, increased pulse rate of the child or the parent, and poor body image of the child. PTSD is another common ailment seen in burned children and can occur even in toddlers aged 1 to 2 years. They occur after a month and up to 2 years or more after the actual injury. The development of PTSD is frequently predicted by ASD. PTSD is associated with long-lasting neurobiologic abnormalities, such as reduced hippocampal size due to autonomic arousal and acquisition of conditioned fear with chronic re-experiencing of traumatic events. Two biologically separate routes have been proposed in the development. One involves separation anxiety as it relates to increased pain levels, and the other connects the expression of dissociation with burn size. The importance of these recent findings lies in the possibility of preventing the development of PTSD by detecting and treating early signs and symptoms of acute stress.

Medical and Psychiatric Treatment of Burned Children

Recent reports indicate up to 50% decline in the numbers of burn-related deaths and hospital admissions in the United States. This is attributed to more effective preventative measures and advances in therapeutic strategies, leading to a decrease in the number and severity of burns and improved clinical outcomes. Great advances in the spheres of resuscitation, wound coverage, support of the hypermetabolic response to injury, infection control, treatment of inhalation injury, as well as recognition and treatment of the psychological sequela in the burn patient accounts for the improvements in clinical outcomes.

Pain management in the burned child remains an arduous task, despite the aforementioned improvements, specialized pain management departments, an arsenal of pain medications, and alternative methods of treatment, including hypnosis and guided imagery. Too often, the pain remains difficult to treat or is undertreated. Adequate pain control is critical in the management of the burned children, especially as unresolved pain causes negative short- and long-term psychological outcomes. It may also trigger anticipatory anxiety for future medical procedures. Increased pain and anxiety are found to decrease wound re-epithelialization, which can also lead to long-term consequences for growth and mobility. Increased pain can also influence the possibility of patients and families displaying signs of PTSD.

The combination of the constant presence of caregivers on burn and trauma units, child life specialists, and administration of medications, such as benzodiazepines, SSRIs, and tricyclic antidepressants, have significantly improved the treatment of anxiety and other psychological sequelae of burn injury.

The parental level of coping with the stress of caring for a burned child influences the child's recovery. Treatment of the caregiver may be essential for optimum recovery of the child. In addition, educating parents about the psychological effects of burn injury on their child will help them to be more attuned to the child's needs and stages of healing. A holistic and family-centered approach consisting of both pharmacologic and nonpharmacologic techniques helps provide optimal care for the burned child.

Lisa, an 11-year-old girl, suffered a 70% TBSA burn. Following a protracted, critical hospital course and treatment in the acute inpatient setting for over a year, she was finally transferred to a specialized long-term rehabilitation center. Toward the later parts of her acute inpatient stay and prior to transfer, Lisa was gradually coming into her own again besides developing meaningful friendship ties with the staff. She had also started exhibiting typical adolescent behaviors of boredom, rebellion, and stubbornness while re-establishing connections with a few of her old friends via the internet. She, however, remained reluctant to confront her new appearance. A few months later, at the rehab center, coupled with continued progress in other spheres and building up trust in new staff at the rehab center, she was finally comfortable enough to request a full-length mirror. Huge shock and devastation were evident at her first glimpse of the full extent of her scarring and disfigurement. However, her family and staff were immensely supportive and reassuring. She eventually came to terms with her new appearance and started entertaining visits from friends.

Developmental Considerations

The prevalence of premorbid dysfunction in burned children is high and estimated at over 50%. Pre-existing developmental issues, including learning difficulties, emotional and behavioral disturbances, and developmental delay, are significant risk factors for a burn injury in children. The complexities of burn treatment and often lengthy recovery times can adversely affect the physical and emotional development of the burn child. Preburn physical, intellectual, and emotional levels will help determine realistic rehabilitation goals as much as will the extent and impact of the burn injury. The burden of postburn injury morbidity, including long-lasting physical and psychological effects, may, unfortunately, be experienced by children in the remainder of their lifetime. A child's level of resilience is likely due to some combination of preinjury emotional stability, level of support, family construction, socioeconomic status, temperament or genetics, and certain aspects of after-burn care and rehabilitation.

Health care providers involved in a child's or adolescent's transition to rehabilitation must take premorbid and postinjury developmental status into consideration when assessing the goals of the child, family, and caregivers during discharge planning. The survival of the burn patient, children in tow, had greatly improved over the years, even in those with large total body burns. Decreasing caregiver stress can have a direct correlation with the well-being of the patient, as shown in several studies focused on the effect of stress levels in parents of injured or ill children.

STAFF SUPPORT

Hospital burn unit staff are repeatedly exposed to severely traumatized bodies and minds. Patients and their families depend on the staff for support, clinical updates on the patient's condition, and glimmers of hope that "everything will be okay." When patient outcomes are positive, these responsibilities can provide staff with great satisfaction and the sense that they have been able to adequately minister to the patient. As in any critical care unit, morbidity and mortality rates can be high. Factors that contribute to burn unit staff burnout and increased levels of staff stress include (1) high acuity, and the labile nature, of burn-injured patients' physiologic and psychological states; (2) understaffing of the unit or lack of organizational support and guidelines to reduce work stress; (3) physical isolation of the burn unit from other areas of the hospital and avoidance by other services; (4) "secondary traumatization" via repeated visual exposure to the disfigurement and severely burned body parts of patients and by repeated descriptions in rounds of the actual burn event; (5) reactions and demands of traumatized patients and their families; (6) pressure to make life and death ethical decisions; (7) managing patients' ongoing pain; (8) lengthy hospital stays; and (9) substance abuse issues.

The pandemic has increased psychological distress, contributing to stress/burnout factors for providers on burn units. The risk of viral transmission to staff and patients has been of the greatest concern. In the current (as of this writing) world of SARS-CoV-2, concerns of social distancing, PPE procurement and use, testing, quarantining, and unit staffing have taken over ordinary routines. The daily race is to take all of this into account and learn as you go to keep ahead of the virus and prevent deaths. Keeping up with new protocols is important and helpful to feel in command of all the information and yet stressful at the same time since keeping up with the new information is time-consuming and not always clear.

Many burn units are tucked away in relatively remote areas of the hospital. The isolation of the physical plant may mirror and enhance the psychological isolation felt by the staff when handling critically ill or multiple acute patients. Caring for a burned patient may require repeated visual impressions of maimed, grafted, or amputated body parts or of the patient wrapped in gauze with only eyes and mouth exposed. Many burned patients are protected from infection by isolation precautions and live behind a glass window. One must gown and glove with each visit. This is exponentially intensified due to the covid precautions hospital staff are now taking. The patient may have suffered or required amputation(s), causing disfigurement and altering their life in many ways. The heavy weight of these burdens can create feelings of helplessness and despair in staff and loved ones as well as the patient. To avoid these painful emotions, staff and family may struggle with urges to avoid the patient at any stage of treatment or recovery. In addition, consultants may try to avoid coming to the burn unit, another factor that contributes to feelings of isolation among burn unit staff.

Studies have shown an increase in mistakes made by caregivers with increased levels of stress. Acute and chronic stress in the workplace leads to staff with lowered morale, impaired performance, increases in turnover, and absenteeism. At a particularly high risk of increased stress is the burn unit staff in charge of critical decision making, as they attempt to balance the needs of the patient and the unit and implement national guidelines and local standards of care with the resources available to the unit, hospital, and community. Training and supervision of ICU nursing staff have been shown to reduce absenteeism and turnover and improve patient care. There is a need to psychologically support health care workers. There are programs created for health care to educate, bolster support, and promote resilience through in-person as well as internet programs.

There are great challenges in taking care of burn patients. Adding substance abuse, psychiatric issues, and the stress of working in difficult conditions to the clinical picture increases those challenges, all of which can contribute to a myriad of feelings that can get played out and contribute to negative countertransference and staff burnout. The psychiatric consultant can play an integral role in reducing staff stress by making daily rounds on the burn unit, making recommendations for the treatment of patients' psychological, neurologic, pain, as well as safety issues. Liaison work with families and staff conveys a sense of connection, support, and understanding. The psychiatric consultant can explore these feelings that the staff may be experiencing, which can assist them in working with patients and their families.

END OF LIFE AND ETHICS

Consultation liaison psychiatrists are regularly called upon to help evaluate multiple ethical queries arising from the management of burn patients. Patients with massive burns are now more likely to survive than in the past. These patients can fall into protracted shock or unresponsiveness, rendering them incapable of making their own medical decisions. Thus, burn injuries can bring up a myriad of complex ethical challenges, including capacity for medical decision making, the appointment of a health care proxy, continuation of treatment, do not resuscitate or intubate orders, medical futility, quality of life concerns, and initiation of comfort measures as opposed to standard treatment.

While dealing with such complex issues, the principles of autonomy, beneficence, nonmaleficence, and justice must be considered. While there is no simple algorithm to help solve the ethical challenges that present in the management of the severely burned patient, there are frameworks that may help clinicians work through these ethical dilemmas.

One such approach, known as the "Four Quadrant Approach," described by Jonsen, Siegler, and Winslade, uses the above principles to analyze ethical problems in the context of four areas: medical indications, patient preferences, quality of life, and contextual features (i.e., social, economic, legal and administrative).

- *Medical indications:* In the first quadrant, the medical facts of the case are described. Diagnosis, prognosis, proposed treatment, and expected outcomes fall under this quadrant. In the context of burn care, this may include the type and severity of the burn injury, planned interventions, and the expected course and outcome for the patient. This stage is important to highlight the purpose and goals of the planned interventions, risks, and benefits of the proposed interventions. Having the facts, an accurate diagnosis, prognosis, and severity of the burn can help guide any ethical dilemmas that may arise surrounding goals of care and how management decisions might benefit or harm the patient. While this may occur in any clinical discussion, it becomes especially important when ethical questions arise.
- *Patient preferences:* Patient's exercise of autonomy is usually central to the resolution of ethical dilemmas often posed in their care. If the patient has decision-making capacity, their

preferences should be respected and guide medical care. If the patient cannot participate in decision making, the patient’s wishes or best interests should be carried out by a health care proxy or a surrogate decision maker.

Determining decision-making capacity in an acute burn patient poses unique challenges. Patients usually have a lucid period at the start of initial resuscitation in which patients demonstrate an ability to participate in their medical decision making. However, obtaining informed consent during this time period can be controversial because patients may be cognitively and emotionally incapable of making life-changing decisions so soon after the injury. Family members may also find themselves highly distressed and may not be able to process all the information given to them. Patients and families should be educated about their medical choices with guidance and support from the medical team.

Ideally, patients should be asked the type of care that they would wish or do not wish to receive at the start of the initial resuscitation, a period in which they are usually lucid.

- *Quality of life:* Burn injuries can negatively affect a patient’s quality of life. Patients with severe burns may lose functionality, have significant scarring, and deal with life-altering impairments. These may lead to employment and financial issues, isolation, and poorer social supports. While treatment generally focuses on preserving and restoring quality of life, clinicians should aim to determine whether therapeutic interventions are likely to positively or negatively affect the quality of life.

An often-arising intricate ethical query is withholding or withdrawal of care in severely burned patients. Even in cases when this is the patient’s wish or when the staff feels death is inevitable, making this decision can still cause high anxiety in both staff and family members. A “consensus conference” can be used to help the burn team discern the judgment of the group. All involved staff members discuss and resolve their differences of opinion and unify their position prior to presenting to the patient or their family. This group decision helps remove the burden of responsibility on individual staff members as well as on family members, who may otherwise bear the burden of making a decision that may ultimately end with the patient’s death.

Pain control and management are of paramount importance in all burn patients as this might be an underlying element behind a burn patient’s refusal of further treatment, dressing changes, and other related procedures, as well as in terminally ill burn patients where adequate comfort and sensible care should be ensured. A consulting psychiatrist must evaluate for pain control and advocate for optimal pain control in these patients. Comforting words may be of help even in seemingly unresponsive patients. All involved staff should also help patients and their families with anticipatory grief and with concrete plans in order to help alleviate their distress.

- *Contextual features:* Family dynamics, financial resources and socioeconomic issues, religious/cultural identity, and personal bias of anyone involved in the care of the patient should also be considered.

In conclusion, the consulting psychiatrist must evaluate the medical indications for treatment, including risks, benefits, and potential burdens; the patient preferences based on patient’s autonomy; quality of life issues; and contextual features, including socioeconomic, family, and legal issues. Patients’ surrogates may aid in decision making and must be adequately informed regarding the extent of the burn injury, treatment options, likely outcomes, and prognosis. The hospital’s bioethics committee is sometimes involved in the resolution of very complex ethical situations.

Several years ago, at a conference, one of the authors (MB) was on a panel discussing ethical issues in the burn patient. Another guest panelist was a nonphysician professional who had suffered a major burn injury that had severely scarred his face and body, causing blindness. He subsequently went on to graduate school, got married, had children, and reported that he enjoys his family very much. However, his point to the panel is that during his acute hospitalization, he asked numerous times that his treatment be stopped and that he be allowed to die because he found the treatment intolerable. He believed that it was wrong that his wishes were not respected, and he was advocating that this should be done in the future with other patients in this situation. He was asked by this author whether he had adequate pain control during his treatment and replied definitely not. He agreed that he might feel different if there had been better pain relief of his severe pain.

FURTHER READINGS

Abdelrahman I, Vieweg R, Irschik S, Steinvall I, Sjöberg F, Elmasry M. Development of delirium: association with old age, severe burns, and intensive care. *Burns*. 2020;46(4):797–803.

Ahrens M, Evarts B. Fire loss in the United States during 2019. *National Fire Protection Association Research*. 2020. Accessed March 2021. https://www.nfpa.org/News-and-Research/Data-research-and-tools/US-Fire-Problem/Fire-loss-in-the-United-States

American Burn Association. *National Burn Repository 2019 Update*. Accessed February 2020. http://www.ameriburn.org

Attoe C, Pounds-Cornish E. Psychosocial adjustment following burns: an integrative literature review. *Burns*. 2015;41(7):1375–1384.

Aziz S, Arabi YM, Alhazzani W, et al. Managing ICU surge during the COVID-19 crisis: rapid guidelines. *Intensive Care Med*. 2020;46(7):1303–1325.

Coffman KL. Psychological and psychosocial aspects of face transplantation. In: Sher Y, Maldonado JR, eds. *Psychosocial Care of End-Stage Organ Disease and Transplant Patients*. Springer; 2019:353–363.

Davé DR, Nagarjan N, Canner JK, Kushner AL, Stewart BT; SOSAS4 Research Group. Rethinking burns for low & middle-income countries: differing patterns of burn epidemiology, care seeking behavior, and outcomes across four countries. *Burns*. 2018;44(5):1228–1234.

ElHawary H, Salimi A, Gilardino MS. Ethics of facial transplantation: the effect of psychological trauma associated with facial disfigurement on risk acceptance and decision making. *Ann Surg*. 2020;275(5):1013–1017.

Fidel-Kinori SG, Eiroa-Orosa FJ, Giannoni-Pastor A, Tasqué-Cebrián R, Arguello JM, Casas M. The Fenix II study: a longitudinal study of psychopathology among burn patients. *Burns*. 2016;42(6):1201–1211.

Govender R, Hornsby N, Kimemia D, Van Niekerk A. The role of concomitant alcohol and drug use in increased risk for burn mortality outcomes. *Burns*. 2020;46(1):58–64.

Gueler JR, McMullen K, Kowalske K, et al. Exploratory analysis of long-term physical and mental health morbidity and mortality: a comparison of individuals with self-inflicted versus non-self-inflicted burn injuries. *Burns*. 2020;46(3):531–538.

Hodgman EI, Subramanian M, Wolf SE, et al. The effect of illicit drug use on outcomes following burn injury. *J Burn Care Res*. 2017;38(1):e89–e94.

Hudson A, Al Youha S, Samargandi OA, Paletz J. Pre-existing psychiatric disorder in the burn patient is associated with worse outcomes. *Burns*. 2017;43(5):973–982.

James DL, Jowza M. Principles of burn pain management. *Clin Plast Surg*. 2017;44(4):737–747.

James SL, Lucchesi LR, Bisignano C, et al. Epidemiology of injuries from fire, heat and hot substances: global, regional and national morbidity and mortality estimates from the Global Burden of Disease 2017 study. *Inj Prev*. 2020;26(Supp 1):i36–i45.

Kalina M, Malyutin G, Cooper ML. A comparison of burn related injuries following the natural disaster Super Storm Hurricane Sandy to the National Burn Repository of the American Burn Association. *Burns Trauma*. 2016;4:1.

Karaçetin G, Demir T, Baghaki S, Cetinkale O, Yüksel ME. Psychiatric disorders and their association with burn-related factors in children with burn injury. *Ulus Travma Acil Cerrahi Derg*. 2014;20(3):176–180.

Klifto KM, Shetty PN, Slavin BR, et al. Impact of nicotine/smoking, alcohol, and illicit substance use on outcomes and complications of burn patients requiring hospital admission: systemic review and meta-analysis. *Burns*. 2020;46(7):1498–1524.

Kukolja J, Kuhn J. SOP: treatment of delirium. *Neurol.Res Pract*. 2021;3(1):12.

Lembke A, Ottestad E, Schmiesing C. Patients maintained on buprenorphine for opioid use disorder should continue buprenorphine through the perioperative period. *Pain Med*. 2019;20(3):425–428.

Lodha P, Shah B, Karia S, De Sousa A. Post-traumatic stress disorder (ptsd) following burn injuries: a comprehensive clinical review. *Ann Burns Fire Disasters*. 2020;33(4):276–287.

Mushin OP, Esquenazi MD, Ayazi S, Craig C, Bell DE. Self-inflicted burn injuries: etiologies, risk factors and impact on institutional resources. *Burns*. 2019;45(1):213–219.

Najafi A, Nejad HZ, Nikvarz N. Evaluation of the analgesic effects of duloxetine in burn patients: an open-label randomized controlled trial. *Burns*. 2019;45(3):598–609.

Nisavic M, Nejad SH, Beach SR. Intentional self-inflicted burn injuries: review of the literature. *Psychosomatics*. 2017;58(6):581–591.

Rosenberg L, Rosenberg M, Rimmer RB, Fauerbach JA. Psychosocial recovery and reintegration of patients with burn injuries. In: Herndon DN, ed. *Total Burn Care*. 5th ed. Elsevier Inc; 2017:709–724.

Scheffler M, Koranyi S, Meissner W, Strauß B, Rosendahl J. Efficacy of nonpharmacological interventions for procedural pain relief in adults undergoing burn wound care: a systematic review and meta-analysis of randomized controlled trials. *Burns*. 2018;44(7):1709–1720.

Shadman A, Arabi YM, Alhazzani W, et al. Managing ICU surge during the COVID-19 crisis: rapid guidelines. *Intensive Care Med*. 2020;46(7):1303–1325.

Spronk I, Legemate CM, Dokter J, van Loey NEE, van Baar ME, Polinder S. Predictors of health-related quality of life after burn injuries: a systematic review. *Crit Care*. 2018;22(1):160.

Storey K, Kimble RM, Holbert MD. The management of burn pain in a Pediatric Burns-Specialist Hospital. *Paediatr Drugs*. 2021;23(1):1–10.

Teven CM, Gottlieb LJ. The Four-Quadrant Approach to ethical issues in burn care. *AMA J Ethics*. 2018;20(1):595–601.

Tuohig GM, Saffle JR, Sullivan JJ, Morris S, Lehto S. Self-inflicted patient burns: suicide versus mutilation. *J Burn Care Rehibil*. 1995;16(4):429–436.

World Health Organization. Burns. Fact sheet. Updated March 6, 2018. Accessed March 2021. https://www.who.int/news-room/fact-sheets/detail/burns

Zheng H, Wu K, Zhou Y, Fu L, Zhang D, Liu Z. Prevalence and associated factors of post-traumatic stress disorder in burned patients and their family members. *Burns*. 2021;47(5):1102–1109.

27.15 Management of Chronic Pain

J. GREGORY HOBELMANN, M.D., M.P.H., AND MICHAEL R. CLARK, M.D., M.P.H., M.B.A.

DEFINITION AND COMPARATIVE NOSOLOGY

Pain is a complex experience integrating cognitive, behavioral, and affective factors serving as a warning of impending or actual injury. The International Association for the Study of Pain defines pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage." This definition recognizes that pain may be experienced in the absence of tissue damage and is, therefore, a subjective experience. As such, the patient's report of pain should be considered valid in the absence of evidence to the contrary.

There are many terms used to describe different types of painful experiences (Table 27.15–1). Pain can be either acute or chronic. Acute pain is typically the result of actual tissue damage secondary to injury, surgery, or the exacerbation of chronic disease. It has survival value in that it promotes behaviors that minimize repeat injury. If pain persists on a daily basis for a month beyond what would be considered necessary to heal the underlying injury, it is generally considered chronic. Chronic pain seems to serve no purpose and is much more complex and difficult to manage than acute pain. The development of chronic pain is multifactorial and involves physical, cognitive, and emotional processes but is not fully understood. The intensity of pain can be classified as mild, moderate, or severe and is usually measured on a 10-point numerical rating scale.

Pain is further classified as either nociceptive or neuropathic (Table 27.15–2). Nociceptive pain results from injury or disease and can be divided into somatic or visceral pain. Somatic pain is usually well-localized and results from injury or disease of the skin, muscle, bone, or related structures. Examples of somatic pain conditions are arthritis, fibromyalgia, and myofascial pain. Visceral pain arises from internal organ dysfunction as the result of inflammation, ischemia, or organ distention and is usually poorly localized and difficult to characterize. Examples of visceral pain conditions are endometriosis, appendicitis, and interstitial cystitis. The quality of nociceptive pain is often described as throbbing, aching, or pressure-like pain.

Table 27.15–1.
Types of Painful Experiences

Allodynia	A painful response to a normally non-noxious stimulus
Deafferentation pain	Pain resulting from loss of sensory input into the central nervous system
Dysesthesia	An abnormal pain that can be spontaneous or evoked
Hyperalgesia	An exaggerated painful response to a normally noxious stimulus
Hyperesthesia	Increased sensitivity to stimulation that excludes the special senses
Hyperpathia	An exaggerated painful response evoked by a noxious or non-noxious stimulus
Hypoesthesia	Diminished sensitivity to stimulation that excludes the special senses
Nociception	Detection of tissue damage by transducers in skin and deeper structures and the central propagation of this information via A delta and C fibers in the peripheral nerves
Paresthesia	An abnormal nonpainful sensation that is either spontaneous or evoked
Sensitization	Lowered threshold and prolonged/enhanced response to stimulation

Neuropathic pain is pain arising from a lesion or disease of the somatosensory system and is often accompanied by maladaptive changes in the nervous system. The distinction is made between peripheral, as in peripheral neuropathy and postherpetic neuralgia (PHN), or central, as in multiple sclerosis and central poststroke pain. The quality of neuropathic pain is usually described as shooting, stabbing, or electric-like pain. Conditions such as headache and cancer pain can involve both nociceptive and neuropathic pain. Over time, all types of chronic pain produce more difficult-to-reverse changes in the functioning of the CNS as well as the structure of its components termed central sensitization.

EPIDEMIOLOGY

Pain is the number one reason that patients present to their physician, and the societal burden of pain is considerable. It is the most important current and future cause of morbidity and disability globally and is a feature in a wide range of medical conditions. In 2008, about 100 million adults in the United States reported a chronic pain condition. Chronic back pain is the most common condition, followed by severe headaches, arthralgias, and neck pain. Patients with chronic pain utilize about twice as many health care resources as the general population. People with moderate to severe pain lose 8 days of work every 6 months on average. The total health care costs secondary to pain ranged from $560 billion to $635 billion in 2010, eclipsing the cost of heart disease, diabetes, or cancer.

Many factors have been linked to the development of chronic pain. According to the available literature, the effect of age is inconsistent, with some studies reporting an increase in the prevalence of chronic pain with age and others reporting a decrease. Most studies report that women present with more severe, frequent, and longer-lasting pain than men. There appears to be an inverse relationship between socioeconomic status and the prevalence of chronic pain, with a higher prevalence in patients with lower income, lower levels of education, and who are unemployed. Several factors, including smoking, alcohol use, lack of physical activity, and poor nutrition, are associated with the development of chronic pain. The burden of chronic pain is worse in developing countries and appears to be increasing over time, despite improvements in medical care. The most reliable factors for the development of chronic pain are a history of a previous episode of pain and a more severe intensity of acute pain.

Table 27.15–2.
Neuropathic and Nociceptive Pain Conditions

NEUROPATHIC	
Peripheral	**Central**
Peripheral neuropathy	Spinal cord injury
Postherpetic neuralgia	Central poststroke pain
Trigeminal neuralgia	Parkinson disease
Chronic postsurgical pain	Multiple sclerosis
Phantom limb pain	Seizure disorder
Complex regional pain syndrome	
Radiculopathy	
Nerve entrapment	
NOCICEPTIVE	
Somatic	**Visceral**
Arthritis	Irritable bowel syndrome
Myofascial pain	Endometriosis
Fibromyalgia	Interstitial cystitis
Connective tissue disorders	Ulcers
Surgical/trauma-related pain	Appendicitis
Burn pain	Cholecystitis

ETIOLOGY AND PATHOPHYSIOLOGY

The neurobiology of pain is complex and involves peripheral, dorsal horn, ascending tract, and descending inhibition mechanisms. Pain results from the activation of peripheral receptors, or nociceptors, by a stimulus. Ongoing stimulation causes changes in peripheral nerves, spinal cord structures, and supraspinal structures, which can result in the development of chronic pain. Our understanding of potential mechanisms that result in chronic pain remains relatively unclear but involves dimensions such as stress, age, genetics, environment, emotional response to injury, and immune responsivity, among others.

Peripheral mechanisms begin with primary afferent nociceptors responding to a mechanical, chemical, or thermal stimulus. Myelinated A-delta fibers transmit mechanothermal information with a sharp, prickling quality. Unmyelinated C-fibers are polymodal and transmit pain with a burning, aching quality. Usually, stimulation activates only high-threshold nociceptors. In conditions of nerve injury or inflammation, peptides such as substance P and neurokinin A are released, causing neurogenic inflammation. This results in nerve fibers becoming excitable, vascular structures dilating with extravasation of plasma proteins, and inflammatory cells releasing a variety of chemical mediators. These chemicals can alter the response of the high-threshold nociceptors by reducing the response threshold called peripheral sensitization. In peripheral sensitization, there is a zone of allodynia around the site of injury in which low-intensity stimulation can produce pain.

The primary afferent nociceptors terminate in the dorsal horn of the spinal cord. The dorsal horn is the site of action of many neurotransmitters and neuromodulators that act at several receptors, including *n*-methyl-D-aspartate (NMDA), kainite, glutamate, opioid, and GABA, among others. The second-order neurons can be classified into nociceptive- specific/high-threshold or wide dynamic range neurons. Allodynia occurs when wide dynamic range neurons become sensitized and fire at an increased frequency. Central sensitization is a state of increased sensitivity to central pain signaling neurons and also occurs in the dorsal horn. It refers to several modifications resulting from prolonged activation of nociceptive afferent fibers. These modifications include windup (progressive increase in neuronal activity during stimulation), facilitation (magnification and prolongation of the neuronal response), action potential threshold reduction, receptive field expansion, oncogenic induction, and long-term potentiation (strengthening of the synaptic transmission after stimulation). NMDA receptor activation by glutamate is thought to be an important mechanism for the production of central sensitization.

Second-order neurons project to supraspinal structures in the ascending tracts of the spinal cord. These ascending tracts are located in the contralateral anterolateral spinal cord (spinothalamic, spinoreticular, spinomesencephalic) and latent ipsilateral pathways. The ventroposterior nuclei of the thalamus represent temporal and spatial aspects of pain, while the medial nuclei are involved in the affective-motivational features of pain. Other subcortical structures, such as the basal ganglia, hypothalamus, amygdala, and cerebellum, are thought to function in the transmission of nociception and perception of pain. The role of cortical structures in pain perception is not well understood, but it is postulated that the parietal lobes and somatosensory cortex contribute to the sensory-discriminative component while the cingulate cortex is involved with the affect component.

Pain can be decreased by descending inhibition. This is a complex mechanism involving many structures and neurotransmitters throughout the central nervous system. Treatment modalities involving electrical stimulation attempt to activate descending inhibition to reduce the activation of nociceptive mechanisms.

DIAGNOSIS AND CLINICAL FEATURES

Pain is a subjective experience and is difficult to assess, especially in people with cognitive limitations or other chronic degenerative diseases. Complaints of pain should always be taken seriously. There are no diagnostic tests that can ascertain the existence of pain or measure the severity of pain. The evaluation of pain should always be individualized and begins with a thorough history and physical examination. A comprehensive history should include the quality and location of the pain, associated signs and symptoms, exacerbating and alleviating factors, temporal aspects, interference with daily functioning, and responses to prior treatments. Acute pain is usually the result of an inciting insult such as trauma or surgery, but the etiology of chronic pain is usually more insidious, and often the etiology of chronic pain cannot be determined. As a result, chronic pain conditions result in more psychological sequelae than acute pain conditions.

The severity of pain can be assessed using a variety of rating scales, including verbal, numerical scales from 0 to 10 and the visual analog scale (VAS). The VAS consists of a 10-cm line with “no pain” on the left and “worst possible pain” on the right. For consistency, the same scale should be used over time in individual patients. For patients that have difficulty using these scales, such as young children or cognitively limited patients, substitutes such as the facial expression scale have been validated. In addition to severity, other important aspects of pain should be assessed, including functional capacity, psychological and emotional function, satisfaction ratings, and adverse treatment effects. Distinguishing between neuropathic and nociceptive pain is also important as treatment approaches to each vary considerably.

A psychological history must be included in the evaluation as well. A focused exam should include prior psychiatric diagnoses, work history, legal history, substance abuse history, and prior or current suicidality. Up to two-thirds of chronic pain patients have comorbid psychopathology. Depression is the most common comorbidity, followed by anxiety disorders, somatoform disorders, and substance use disorders (SUDs). It is important to elicit these disorders and treat them concurrently, as they have been associated with poorer treatment prognoses. In addition, people seeking care for chronic pain may also be more likely to have personality vulnerabilities which can act as a barrier to effective treatment.

Physical examination of the patient should include all body systems because pain is often a manifestation of systemic disease.

However, a focal neurologic and musculoskeletal examination may be appropriate in patients with limited complaints. It is important to evaluate for sensory changes, motor changes, passive and active range of motion, and postural and gait abnormalities. The examination findings are rarely pathognomonic but rather serve to select patients for further imaging studies or invasive diagnostic procedures to focus on a specific etiology.

Pathology and Imaging Studies

Various imagining techniques have become commonly utilized in the search for the cause of chronic pain. However, they have several drawbacks and are limited in their utility. The correlation between MRI findings and the severity of pain is poor. Greater than 50% of asymptomatic patients will have abnormalities on lumbar, thoracic, or cervical films, and early imaging for back pain does not appear to improve outcomes or affect decision-making. MRIs are indicated in patients with serious or progressive neurologic deficits, bowel or bladder dysfunction, suspected metastatic disease, or infection.

For nonspinal conditions, MRI is best for inflammation and soft tissue disease, and CT scanning is better for detecting bone disease and bleeding. Plain films can be useful to detect gross bone abnormalities and joint degeneration. Ultrasound is infrequently used but has the advantages of being inexpensive and does not expose patients to radiation. In most instances, imaging serves to alert the practitioner to seek further invasive diagnostic procedures.

Electromyography and nerve conduction studies (NCS) can be helpful in detecting lesions of large nerves. They do not detect impairment of smaller nerves, however, and are associated with significant false-negative and false-positive rates. For peripheral neuropathies, skin biopsies are sensitive tests that demonstrate decreased density of epidermal nerve fibers.

Recent studies have shown that the brains of people with chronic pain are different from those of people without pain complaints. There are morphologic alterations such as decreases in the gray matter in the areas involved in the perception of pain. This appears to be a result of the pain and can be reversed with successful treatment. Functional MR-based brain imaging methods are developing rapidly and will likely be utilized more in the diagnosis and treatment of pain in the future.

DIFFERENTIAL DIAGNOSIS

Acute pain is usually the result of an injury or trauma and determining the underlying cause is straightforward. Chronic pain, however, often has no identifiable etiology, and management must be focused on the suppression of symptoms. The more common chronic pain syndromes (Table 27.15–3) are discussed in the next section.

Low Back Pain

Low back pain (LBP) is one of the most common physical complaints and the most expensive condition when health care costs and loss of productivity are considered. Approximately 25% of adults in the United States have experienced at least 24 hours of LBP in the last three months, and lifetime prevalence ranges from 65% to 80%. Psychological factors such as depressed mood and somatization are highly correlated with LBP and predict the transition from acute to chronic pain. The most powerful predictor of chronicity is poor functional status 1 month after seeking treatment. Both economic and social rewards have been associated with higher levels of disability and depression in patients with chronic LBP.

The presence of a depressive disorder increases the risk of developing chronic LBP and musculoskeletal pain in general. Depression has been associated with a nearly fourfold increase in the likelihood of seeking treatment for new-onset LBP. On the other hand, chronic disabling LBP can result in significant psychopathology that affects outcome and treatment response.

Chronic LBP is a condition that exemplifies the complexity of chronic pain treatment. Patients with chronic LBP present with symptoms representing numerous diagnoses that require multimodal treatment approaches. Treatments can often reduce symptoms, but there is conflicting evidence about their ability to improve functional status and return patients to work. Almost all treatment modalities, including conservative interventions, have produced inconsistent results. A thorough evaluation and individualized treatment plan for each patient presenting with LBP are essential.

Table 27.15–3.
Common Chronic Pain Conditions

HEADACHE

- Migraine
- Tension-type headache
- Cluster headache
- Headache associated with trauma
- Chronic daily headache
- Cervicogenic headache
- Postdural puncture headache

OROFACIAL PAIN

- Trigeminal neuralgia
- Geniculate neuralgia (Ramsay Hunt Syndrome)
- Glossopharyngeal neuralgia
- Superior laryngeal neuralgia
- Tolosa–Hunt syndrome

LOW BACK PAIN

- Spondylosis
- Spondylolisthesis
- Spondylolysis
- Disc herniation
- Foraminal stenosis
- Facet arthropathy
- Spinal stenosis
- Fracture
- Musculoligamentous
- Discogenic pain
- Congenital disease

BUTTOCK PAIN

- Sacroiliac dysfunction
- Piriformis syndrome

PERIPHERAL NEUROPATHIES

- Painful peripheral neuropathies (e.g., diabetes)
- Focal entrapment neuropathies
- Postsurgical syndromes
- Traumatic brachial plexus avulsion
- Ischemic cerebrovascular injury

- Syringomyelia
- Arachnoiditis
- Complex regional pain syndrome
- Meningoradiculopathies
- Epidural spinal cord compression
- Herpes zoster and postherpetic neuralgia

PELVIC PAIN

- Endometriosis
- Infection
- Uterine pain
- Pelvic congestion
- Pelvic adhesions
- Ovarian remnant syndrome
- Cancer pain

CENTRAL PAIN STATES

- Spinal cord injury pain
- Poststroke pain
- Parkinson disease
- Multiple sclerosis

FIBROMYALGIA

MYOFASCIAL PAIN SYNDROME

PHANTOM PAIN

SICKLE CELL ANEMIA

Peripheral Neuropathic Pain

Peripheral neuropathy caused by a lesion or disease of the somatosensory nervous system is common and can have a major impact on quality of life. It often causes pain, numbness, and weakness and usually occurs in the hands and feet but can occur in other areas of the body. There are numerous etiologies of peripheral neuropathy, including diabetes mellitus and excessive alcohol consumption. Approximately 25% of patients with diabetes will experience painful diabetic neuropathy, and a longer duration of disease and poor glycemic control increases the risk. The pain from peripheral neuropathy is often described as burning and lancinating and can be constant or episodic as well as spontaneous or evoked. The pain is primarily the result of axonal degeneration, segmental demyelination, sites of ectopic impulse generation, and activation of immune cells and can occur anywhere along the peripheral nerve up to the dorsal root ganglion. Voltage-dependent sodium channels contribute to hyperexcitability, and central sensitization amplifies and sustains pain by a variety of mechanisms. Pharmacologic treatment can help to reduce symptoms, but a reversal of the pathology is rarely attainable.

Postherpetic Neuralgia

PHN is responsible for about 15% of referrals to pain clinics. It is defined as pain that persists or recurs at the site of shingles at least 3 months after the onset of an acute varicella-zoster rash. PHN occurs in about 10% of patients with acute herpes zoster and in over 50% of patients over the age of 65. Risk factors for the development of PHN include advanced age, cancer, diabetes mellitus, immunosuppression, longer duration of prodromal symptoms, greater pain and severity of the rash, sensory impairment, and psychological stress. Although it generally improves over time, 25% of patients with PHN continue to complain of pain for longer than 1 year.

Degeneration and destruction of motor and sensory fibers of the dorsal root ganglion characterize acute herpes zoster. In PHN, further neurologic damage, including inflammation of the spinal cord, myelin disruption, axonal damage, and a decrease in the number of nerve endings in the affected area, have been demonstrated. Studies have suggested the role of both peripheral and central sensitization in PHN. Early detection and treatment of herpes zoster with antiviral medication and analgesics may have the potential for the prevention of PHN. In addition, a live attenuated virus vaccination for the prevention of herpes zoster is now available and has reduced the burden of illness in adults older than 60 years.

Migraine and Chronic Daily Headache

The lifetime prevalence of migraine is 18% in women and 6% in men. The peak incidence occurs between the third and sixth decades of life. Common migraine is unilateral and pulsatile and may be accompanied by nausea, vomiting, photophobia, and phonophobia. Prodromal symptoms, such as scintillating scotomata, are typically present in the classic form of migraine. Complicated migraine includes focal neurologic symptoms such as cranial nerve palsies or other neurologic deficits and is often described by the deficit, such as hemiplegic or basilar migraine. The definitive pathogenesis has not been described, but identification of new migraine-associated genes, the visualization of involved brain regions, a greater appreciation of the potential role of cervical nerves, and the recognition of the role of neuropeptides are among the advances that have led to novel targets for treatment. Management of migraine starts with the identification and control of comorbidities and risk factors that predispose individuals to migraines. Beta-blockers, flunarizine, valproate, topiramate, and CBT appear to be the best prophylactic agents, and several studies have demonstrated the efficacy of nonsteroidal anti-inflammatory medications and triptans in the treatment of acute attacks.

The most common reason for lost productivity in the US workforce is headaches, and about 5% of the population suffers from chronic daily headaches. Patients with chronic daily headaches are more likely to have psychiatric comorbidities such as depression and anxiety and report functional disability. They also tend to overuse analgesic medications leading to rebound headaches. Patients with chronic daily headaches are a difficult group to treat, and studies suggest that treatment with medications and CBT together is better than either treatment alone.

Central Poststroke Pain and Spinal Cord Injury

Lesions of the central nervous system, such as stroke or spinal cord injury (SCI), are commonly associated with pain. Symptoms of central pain are generally poorly localized and vary over time. Patients may experience allodynia, hyperalgesia, muscle and visceral pain, and lancinating pain. The pain can be described as burning, aching, lacerating, or pricking but tends to vary widely among patients. The pain may not occur until 1 month or longer after the initial insult. There are likely many processes involved in the generation of the pain, and excitatory amino acids have been implicated in the generation of central sensitization after a stroke or SCI.

Central poststroke pain and SCI pain can be severe and disabling but are commonly underreported and often overshadowed by neurologic complications such as weakness, spasticity, and aphasia. It is associated with clinically diverse features, which has limited the ability to consolidate a definition and clinical characteristics. Studies indicate that treatment with analgesics, adjuvant medications, and in some cases, electrical stimulation can significantly reduce pain in these conditions. The studies have been small, however, and more research is needed to determine which treatments are most efficacious.

Fibromyalgia

Fibromyalgia is characterized by widespread musculoskeletal pain in all four extremities as well as the trunk, exaggerated tenderness, and stiffness. It is also commonly associated with fatigue, sleep disturbances, and other cognitive and somatic symptoms. It is diagnosed in 2% to 8% of the population, but this may be an overestimation as criteria for the condition have been debated. Musculoskeletal pain is the most prominent feature of fibromyalgia, but it should be suspected in any patient who has multifocal pain that is not explained by a definable etiology. Because pain is amplified and can occur anywhere in the body, sensory hyperresponsiveness, chronic headaches, somatic pain, and visceral pain are quite common. Studies suggest that muscle relaxants, membrane stabilizers, selective serotonin/norepinephrine reuptake inhibitors, and tramadol are effective in reducing pain associated with fibromyalgia.

Complex Regional Pain Syndrome

Complex regional pain syndrome (CRPS) was formally called reflex sympathetic dystrophy (RSD) and causalgia. It is described as an array of painful conditions characterized by continued pain that is precipitated by trauma or immobilization that is usually associated with allodynia or hyperalgesia. The pain is regional but is not limited to a specific dermatome or peripheral nerve, and there is no confirmatory test or definitive treatment. In acute stages, inflammatory symptoms prevail. In chronic stages, symptoms can include hyperalgesia, sensory loss, motor changes, sudomotor dysfunction, edema, blood flow abnormalities, perception disturbances, and complications of disuse and physical deactivation. These changes are due to central neuroplasticity and likely involve neurogenic inflammation and neuropeptides such as calcitonin gene-related peptide, cytokines, nerve growth factor, and substance P. Sympathetically maintained pain is often present and may result from the coupling of the sympathetic and sensory neurons. Patients with sympathetically maintained pain can receive temporary relief from sympathetic blockade. Mood, anxiety, and substance abuse disorders are fairly common in patients with CRPS. These disorders are generally considered a consequence rather than a cause of the condition.

Pharmacotherapy for CRPS has limited efficacy, but there is evidence supporting the use of anti-inflammatory drugs, gabapentin, pregabalin, carbamazepine, tricyclic antidepressants, and opioids. Physical therapy, graded motor imagery, and graded pain exposure can help to overcome central reorganization. If psychiatric comorbidity is present, the patient should be referred for specialized treatment. Invasive treatment, such as electrical stimulation and surgical sympathectomy, should only be offered after psychosomatic assessment and careful patient selection. The prognosis for recovery from CRPS is now thought to be better than was historically assumed if treatments facilitate the rehabilitation of the affected limb with activated movement and use.

Phantom Limb Pain

Phantom limb pain is a type of neuropathic pain that originates in the distal territory of an amputated limb. It can also occur after the surgical removal of other body parts, such as the breast, rectum, penis, eye, or tongue. It occurs in 50% to 80% of amputees and is generally described as stabbing, throbbing, aching, burning, or cramping. While studied extensively and easy to diagnose, phantom pain can be challenging to treat. Pharmacologic treatments and neural blockade seem to have limited efficacy, but morphine, calcitonin, and mexiletine have been shown to reduce pain lasting more than 6 months, suggesting a role for the usual pharmacologic treatments of neuropathic pain and central sensitization. The limited success with pharmacologic treatments has spawned interest in psychological interventions, such as CBT and biofeedback, for the treatment of phantom limb pain. The limited available evidence demonstrating the effectiveness of these treatments has shown various degrees of success.

Orofacial Pain

Orofacial pain is defined as pain occurring on more than 15 days per month that lasts for more than 4 hours daily and has persisted for greater than three months. There are a variety of conditions that are categorized as orofacial pain, and we will review a few of the more common syndromes.

Trigeminal neuralgia, or tic douloureux, is a chronic pain condition with severe, paroxysmal, recurrent, lancinating pain in the distribution of the facial nerve (cranial nerve V). It is usually unilateral and involves the mandibular or maxillary branch of the nerve. Episodes of pain are often spontaneous but can be evoked by nonpainful stimuli or environmental conditions. Motor and sensory deficits are not typically present. Uncontrolled pain with frequent or severe attacks increases the risk of developing anxiety, depression, insomnia, weight loss, and social withdrawal.

In most cases of trigeminal neuralgia, there is evidence of trigeminal nerve root compression and demyelination by blood vessels, mass lesions, or other diseases such as multiple sclerosis or herpes zoster. Compression and demyelination cause hyperactivity of the trigeminal nucleus. Carbamazepine is the first-line treatment, and there is evidence to support the use of oxcarbazepine and lamotrigine. If pharmacologic approaches fail, there are several invasive procedures such as microvascular decompression, percutaneous gangliolysis, and stereotactic radiosurgery that can be considered.

Temporomandibular disorder (TMD) refers to conditions that involve pain from the temporomandibular joint, muscles of mastication, or other orofacial musculoskeletal structures. Most commonly, pain arises from the muscles of mastication and is precipitated by the movement of the jaw when chewing or talking. The pain is generally sharp, sudden, intense, and localized to the preauricular area, and joint sounds such as “popping” or “clicking” are common. Other symptoms include feelings of muscle fatigue, weakness, tightness, and changes in bite or ability to open or close the jaw. TMD is often associated with psychological stress, which is responsive to behavioral treatments. Treatment of patients with TMD should initially be based on the use of conservative and evidence-based therapeutic modalities, with more invasive interventions, such as surgery, being reserved for refractory pain.

Burning mouth syndrome (BMS) is an array of conditions characterized by pain in the oral and pharyngeal cavities that are often associated with dryness and taste alterations. Middle-aged and postmenopausal women are most frequently affected. Although most cases are idiopathic, BMS can be associated with other conditions such as bruxism, poorly fitting dentures, oral candidiasis, hypothyroidism, diabetes mellitus, and neoplasm. Psychiatric comorbidities are fairly common and include anxiety, depression, traumatic life events, and somatization. Treatment begins with identifying potential underlying etiologies such as anxiety and depression, nutritional deficiencies (iron, folate, and B-vitamins), maladaptive oral habits, and iatrogenic causes. Symptomatic treatments include tricyclic antidepressants, anticonvulsants, soft desensitizing oral appliances, vitamins, hormonal supplements, and habit awareness counseling.

There are a number of less common orofacial pain conditions. These include cluster headache, short-lasting, unilateral, neuralgiaform pain with conjunctival injection and tearing (SUNCT), Tolosa–Hunt syndrome, and atypical facial pain.

COURSE AND PROGNOSIS

Pain is a complex perceptual experience that is not only affected by the activation of nociceptors but also emotions, cognition, expectations, and social and cultural context. In acute pain, the course and prognosis are fairly predictable with the resolution of pain when the initial insult has resolved. In chronic pain, however, the course and prognosis vary widely and are difficult to predict. Certainly, psychiatric comorbidities and their treatments influence the course and prognosis greatly. Below are some of the psychiatric comorbidities that are frequently associated with chronic pain conditions.

Somatic Symptom Disorder

When a somatic cause for the pain cannot be identified, practitioners often begin to look for psychological causes. Pain caused by emotional factors was first introduced in the DSM-II under psychophysiological disorders and has gone through several modifications in subsequent editions. The authors of the DSM-5 created a new category name, somatic symptom and related disorders, and its prototype is the somatic symptom disorder (SSD). The criteria for SSD have been modified substantially from its predecessor, somatoform pain disorder (SPD). The most radical change is the abolition of the distinction between medically explained and medically unexplained somatic complaints. This will likely increase the number of patients with SSD considerably compared to SPD.

The main feature of SSD is a patient's concern with physical symptoms for which no biologic cause can be found. A psychiatric evaluation is essential in the work-up to exclude or identify comorbid psychiatric disease. Failure to recognize SSD may lead to diagnostic or therapeutic procedures that could result in iatrogenic complications. It may also result in an increased financial burden on the healthcare system. In general, patients with SSD have a poorer prognosis with continuing use of medical services. Proven treatments include cognitive behavioral therapy, mindfulness therapy, and antidepressants. A stepped care approach with close cooperation of a primary care provider, a somatic specialist, and a mental health care professional is recommended.

Substance Use Disorder

The prevalence of severe SUD in patients with a chronic pain condition is estimated to be 7% to 20%. Although a relatively large proportion of patients with SUD report chronic pain, a much smaller but significant proportion of patients with chronic pain report SUD. The essential criteria to note for SUD in patients with chronic pain are continued use of a substance despite negative consequences, loss of control of substance use, and excessive preoccupation with using the substance. Americans account for only 5% of the world's population but consume 80% of the global opioid supply.

Aberrant medication-taking behaviors are often assumed to be abuse or addiction. However, chronic, persistent pain can lead to an increased focus on taking opioid analgesics. Understandably, patients fear the reemergence of pain or withdrawal if they run out of medication and often take measures to assure that they have an adequate supply. About 20% of patients on long-term opioid therapy for pain may have what has been termed "complex persistent dependence (CPD)," which is qualitatively more severe than simple physical dependence. A hallmark of CPD is that patients are unwilling or unable to taper opioids despite functional or medical deterioration, and classic opioid use disorder-like behaviors often emerge with a reduction of the opioid dose. There are several patterns of misuse of or nonadherence to prescribed medications that can occur when patients are concerned about medication. In general, patients take more medication than prescribed. This occurs when patients are concerned about withdrawal, tolerance, addiction, scrutiny of medication by others, or a greater perceived need for medication. Having these concerns is a risk factor for developing a SUD. Patients are at greatest increased risk for SUDs during the first 5 years of chronic pain treatment, and the risk is higher if they have a history of substance abuse or psychiatric comorbidity.

Aberrant medication-taking behaviors are present in about 50% of chronic pain patients who are prescribed opioid analgesics. However, studies suggest that only about 10% of patients with aberrant medication-taking behaviors do so as the result of a SUD.

Patients with a history of SUD are at increased risk for relapse when prescribed opioid analgesics for pain, but the process is not well understood and is likely multifactorial. The cycle of pain followed by relief after taking medication is reinforcing and certainly plays a role. In patients who develop a new SUD, the prescribed medication is commonly involved. Risk prediction tools such as the opioid risk tool (ORT), screener and opioid assessment for patients with pain (SOAPP), and current opioid misuse measure (COMM) are available to offer guidance, but they are self-assessment tools that have limitations.

Conversely, patients with SUDs are at increased risk of having a chronic pain condition. This can lead to stigmatization and undertreatment of their pain. This is potentially a difficult population to treat because opioid-dependent patients with chronic pain have relatively high rates of SUDs. Most chronic pain patients who enter a treatment facility for a SUD reported that they received a prescription for opioids from a clinician. They also frequently reported obtaining opioids illegally and altering the route of administration, which suggests that they had progressed to a severe SUD.

Frequent follow-up care is necessary for patients with chronic pain who are prescribed opioids. Prior to the initiation of opioids, the risks and benefits of the medication should be discussed with the patient, and the clinician should have the patient sign an opioid consent form. Close monitoring with intermittent, random toxicology testing should also be performed. Integrating care for chronic pain with innovative stepped-care models of substance abuse could improve outcomes by tailoring the treatments and their intensity to the needs of the individual. However, there is a lack of high-quality evidence on the prevalence, risk factors, and optimal treatment approaches related to concurrent chronic pain and substance misuse. There is an urgent need for high-quality research to guide evidence-based medicine and health policy considering the escalating epidemic of harms related to opioid misuse.

Depression and Affective Illness

There is an intimate and bidirectional relationship between pain and depression. Patients with major depression often have physical symptoms, and about 60% of depressed patients complain of pain. It has been estimated that patients with depression have about a fourfold increase in the prevalence of chronic pain conditions. Baseline symptoms of depression have been found to predict future pain and disability better than initial ratings of the pain itself.

There are higher rates of lifetime major depression in patients with chronic physical complaints. Of patients that present to pain clinics, up to 50% meet the criteria for major depressive disorder. In patients with chronic LBP, the rate of depression is greater than three times the rate in the general population. In patients with chronic pain, depression is associated with greater severity and persistence of pain, greater interference from pain, more pain behaviors observed by others, and early retirement from working. These patients are also more likely to use medication, including analgesics and sedative-hypnotics, daily. Studies indicate that comorbid pain and depression almost double the health care costs of managing the conditions.

Patients with chronic pain syndromes such as migraine, chronic abdominal pain, and orthopedic pain conditions also report greater suicidality. Rates of suicidal ideation, suicidal attempts, and suicide completions are two to three times greater in these patients. Risk factors for suicide in this population include depressive symptoms, anger problems, harmful habits, childhood adversities, family history of suicide, and occupational disability. There tends to be a decrease in self-efficacy in chronic pain patients that contributes to feelings of hopelessness. Depression is the most consistent and strongest predictor of suicide in patients with chronic pain, although other psychosocial variables likely play a role.

Comorbid pain can result in depression that is more resistant to treatment. Depression should not be viewed simply as an expected outcome in patients with pain and should be treated aggressively. Successful treatment of depressive symptoms often results in the reduction of pain and improvement in functional status and quality of life. This effect is independent of the direct treatment of pain and can likely be explained by the fact that the neurobiology of pain and depression overlap.

Anxiety Disorders, Fear, and Anger

Anxiety symptoms are present in almost half of patients with chronic pain conditions. Up to 30% of these patients meet the criteria for an anxiety disorder such as generalized anxiety disorder, panic disorder, agoraphobia, and post-traumatic stress disorder (PTSD). PTSD is increasingly recognized as a comorbid condition in patients with medical illnesses, especially chronic pain. It is associated with greater levels of pain, emotional distress, interference in function, and disability. On the other hand, anxiety is associated with high levels of somatic preoccupation and physical symptoms. Pain complaints are present in up to two-thirds of patients with panic disorder, and pain intensity is greater in patients with symptoms of anxiety. Pain is associated with higher levels of anxiety and more frequent panic symptoms.

Fear of pain, movement, and reinjury can result in the avoidance of activities that can promote the transition to and maintenance of chronic pain. Restriction of activities can result in physiologic changes, such as weight gain and muscle atrophy, and functional deterioration. This process is reinforced by feelings of low self-efficacy, catastrophic interpretations, and expectations of failure in rehabilitation. The fear-avoidance model of pain incorporates components such as fear of and attention to pain, catastrophizing, escape or avoidance behavior, disability, disuse, and vulnerability to help explain the transition from acute to chronic pain. Fear-avoidance beliefs have been shown to be one of the most significant predictors of failure to return to work. They can reinforce disability through operant conditioning if the avoidance provides short-term benefits, such as relieving anxiety or unwanted responsibilities. Decreasing work-specific fears can be more important than addressing the general fears of activity when trying to improve the physical capability to work.

Catastrophic thinking can be described as the amplification of threatening information. It involves elements of cognitive rumination, symptom amplification, expectations of pessimism, and feelings of helplessness. Catastrophizing interferes with the patient’s ability to remain involved in productive activities and intensifies the experience of pain, emotional distress, and perceived disability. Cognitive distortions such as catastrophizing and fear-avoidance beliefs predict poor coping and adjustment to chronic pain better than objective factors such as disease status and physical impairment. Catastrophizing is also a predictor of suicidality, independent of pain severity and depressive symptoms.

TREATMENT

The primary aims in the treatment of pain are elucidating a cause, alleviating suffering, and restoring function. Biologic, psychological, and social factors all play a role in the perception of pain, and each should be addressed. Pain reduction, functional improvement, minimizing psychological comorbidities, and improving social interactions are all important goals. The traditional biomedically oriented treatment with a focus on pharmacotherapy and passive interventions has not proven to be an effective treatment strategy for patients with chronic, disabling pain. A strong consensus exists that an interdisciplinary rehabilitation approach to the treatment of pain is beneficial.

Pharmacologic Treatment

There are numerous medications used to treat chronic pain. These medications target the mechanisms of peripheral and central nervous system sensitization such as sodium and calcium channel upregulation, descending modulation, loss of inhibitory interneurons, spinal hyperexcitability, and aberrant sympathetic-somatic nervous system interactions. Antidepressants, anticonvulsants, and antipyretic analgesics are the best-studied medications and are recommended as first-line agents. Opioids are also well studied and frequently used in the treatment of pain but are not without significant drawbacks, and long-term effectiveness has not been demonstrated in rigorous research. There is no simple medication algorithm that provides a straightforward approach to the complexities encountered in the treatment of chronic pain.

Antipyretic Analgesics

Aspirin, acetaminophen, and nonsteroidal anti-inflammatory drugs (NSAIDs) are the backbone of pharmacologic management of acute pain. Aspirin is the most widely used analgesic in the world. These drugs inhibit cyclooxygenase, which is the rate-limiting enzyme in the production of prostaglandins. Prostaglandins regulate inflammation and sensitize nociceptors. Acetaminophen works primarily in the central nervous system, lacks anti-inflammatory effects, and is a weaker analgesic than NSAIDs. NSAIDs work both centrally and peripherally, making them effective orally and topically. Antipyretic analgesics are more effective in the treatment of nociceptive pain than they are for neuropathic pain. They are appropriate treatments for mild to moderate pain, but they have a ceiling effect and usually cannot be used as standalone agents in the treatment of severe chronic pain. They work synergistically with opioids and have proven opioid-sparing effects.

Antipyretic analgesics, in particular NSAIDs, have significant side effects, including bleeding, gastrointestinal ulceration, renal toxicity, and increased risk of cardiovascular events. The use of cyclooxygenase-2 selective inhibitors, such as celecoxib, and acid-reducing agents, such as proton pump inhibitory, may reduce the risk of gastrointestinal complications, but nothing can attenuate the renal and cardiovascular risks. As such, they should be prescribed at the lowest effective dose for the shortest duration possible. When appropriate, topical NSAIDs can be used and have comparable efficacy for regional pain conditions.

Opioid Analgesics

Opioid analgesics interact with opioid receptors in both the peripheral and central nervous systems to reduce the sensory and affective components of pain. They are the cornerstone of treatment for cancer pain, but controversy surrounds their long-term use for chronic nonmalignant pain. Several reviews have indicated that they provide prolonged relief and improvement in function in a few individuals, and their superiority over other analgesics is inconclusive.

There are several available guidelines for the use of opioids in chronic nonmalignant pain, and careful attention to balancing the beneficial effects with undesirable adverse effects is necessary. Rational use requires an understanding of the clinical pharmacology of opioids, including potency, duration of action, bioavailability, pharmacokinetics, and pharmacodynamics. The side effects of opioids are numerous and include sedation, nausea, constipation, respiratory depression, pruritus, urinary retention, and endocrine dysfunction. Important steps in the managing of chronic pain patients on opioid therapy are a thorough evaluation, detailed documentation, a function-based treatment plan with an exit strategy for discontinuing opioids, patient education, informed consent, frequent follow-up, and close monitoring. Patients who are appropriate for opioid analgesia include those with moderate-to-severe pain lasting for more than 3 months that adversely affect function or quality of life. Prior to starting opioid therapy, factors such as the patient's particular pain condition, response to therapy, and potential for aberrant drug-related behaviors should be considered. Treatment outcomes, such as degree of analgesia, activities of daily living, and adverse events, should be routinely assessed.

Long-term opioid use is associated with tolerance and dependence. Cross-tolerance among opioids is incomplete and alternating opioids may result in up to a 50% reduction in equianalgesic dose, so opioid rotation is a strategy that may be employed when tolerance is suspected. Reported addiction rates in chronic pain patients who are prescribed opioids vary widely and range from 5% to 50%. A synthesis of evidence suggests that about 50% of patients will exhibit aberrant drug-related behaviors, 20% will abuse the medications, and 10% will develop an opioid use disorder.

Clinically available opioids include naturally occurring compounds, such as morphine and codeine; synthetic opioids, such as fentanyl, methadone, and tramadol; and semisynthetic derivatives, such as oxycodone, hydrocodone, hydromorphone, and oxymorphone (Table 27.15–4). Morphine has poor oral bioavailability due to its hydrophilicity and delayed central nervous system absorption, and the onset of action. This delay prolongs the analgesic effects of morphine, which decreases the potential for accumulation and toxicity with repeated dosing. Oxycodone is a popular analgesic used to treat chronic pain patients. It has higher bioavailability, faster onset of action, and more predictable plasma levels compared to morphine. These qualities may increase the risk of abuse and addiction. It has similar analgesic efficacy to morphine but causes less histamine release and fewer hallucinations. Hydrocodone is similar to oxycodone with regards to its rapid oral absorption and onset of action but is a less potent analgesic. It is metabolized to hydromorphone, which has properties similar to those of morphine but has lower rates of side effects. Fentanyl is highly hydrophilic with an affinity for neuronal tissues. It is frequently used intravenously in the treatment of acute surgical pain but can also be delivered via transdermal and transmucosal routes for the treatment of chronic pain.

Table 27.15–4.
Clinically Available Opioids Used in Chronic Pain Conditions

Drug	Equianalgesic Dose	Duration of Action
Pure Agonists		
Morphine	30 mg	3–6 hours[a]
Oxycodone	20 mg	3–6 hours[a]
Hydrocodone	30–60 mg	3–6 hours[a]
Hydromorphone	3–6 mg	3–6 hours[a]
Oxymorphone	10 mg	3–6 hours[a]
Fentanyl	12.5 mcg/h (TD)	72 hours (TD)
	800–1,000 mcg (TM)	1.5–3 hours (TM)
Codeine	200 mg	3–6 hours[a]
Methadone	2–20 mg	6–12 hours for pain
Agonist/Antagonists, Partial Agonists		
Buprenorphine	0.3 mg (SL)	6–8 hours (SL)
	5–70 mcg/h (TD)	7 days (TD)
Butorphanol	1 mg (IN)	3–4 hours
Pentazocine	90 mg	8 hours
Weak, Dual-Action Agonists		
Tramadol	150–300 mg	4–6 hours[a]
Tapentadol	75–110 mg	4–6 hours[a]

[a] Long-acting formulations available.

TD, transdermal; TM, transmucosal; SL, sublingual.

Methadone for use in the treatment of chronic pain warrants special consideration because of its stigma, low cost, high bioavailability, rapid onset of action, long half-life, multiple receptor affinities (NMDA), lack of neurotoxic metabolites, and incomplete cross-tolerance with other opioids. It has a significantly greater risk of overdose because of the longer time needed for adaptation and its long and variable half-life (15 to 120 hours). Methadone is also associated with QTc prolongation and torsades de pointes at high doses (>100 mg per day), as well as increased rates of all-cause mortality at all doses. Although once-daily dosing is adequate to prevent withdrawal symptoms due to its extensive tissue distribution, its analgesic effects only last about 6 hours due to biphasic elimination. Repeated dosing can increase the analgesic period to 8 to 12 hours because it accumulates in tissues, but this same phenomenon can also increase the risk of delayed respiratory depression, overdose, and death.

Tramadol and tapentadol are a newer class of analgesic drugs with weaker analgesic effects and a dual mechanism of action. They are weak opioid agonists that also inhibit the reuptake of serotonin and norepinephrine. Adverse interactions can occur when administered with coumadin, selective serotonin reuptake inhibitors, and serotonin-norepinephrine reuptake inhibitors, and they are associated with tolerance, dependence, and abuse.

Butorphanol, nalbuphine, and pentazocine are opioid agonist-antagonist drugs. Buprenorphine is a partial agonist at the mu-receptor and antagonist at the other opioid receptors. These drugs should be used with caution in patients taking pure agonist drugs because they can precipitate withdrawal. Buprenorphine (in combination with naloxone) is available orally for the treatment of opioid use disorder and as a transdermal patch for the treatment of chronic pain.

Antidepressants

The analgesic effects of antidepressants are thought to be independent of their antidepressant effects and are mediated primarily by the reuptake inhibition of serotonin and norepinephrine. Increased levels of these neurotransmitters, particularly norepinephrine, enhance the activation of the descending inhibitory neurons in the dorsal horn of the spinal cord. They may also produce antinociceptive effects through a variety of other mechanisms as well including modulation of monoamines, interaction with opioid systems, inhibition of ion channel activity, and NMDA, histamine, and cholinergic receptors. This suggests that all antidepressants would be effective in the treatment of chronic pain, but the analgesic properties of these medications remain underappreciated.

Tricyclic Antidepressants

Studies have confirmed that TCAs are the most effective agents for the treatment of neuropathic pain. They are also effective in the treatment of headache syndromes, central pain syndromes, PHN and other painful neuropathies, and postmastectomy pain syndrome. They have not been shown to be effective in treating SCI pain, phantom limb pain, or painful HIV neuropathy. All TCAs have equal analgesic effects, but the secondary amines, such as nortriptyline, are better tolerated than the tertiary amines, such as amitriptyline. TCAs may produce analgesia at lower doses than needed for the treatment of depression, but treatment should be guided by titration to efficacy. Advantages of TCAs include cost-effectiveness and that optimal titration can be monitored with serum levels to ensure adequate trials and the avoidance of toxicity.

Serotonin-Norepinephrine Reuptake Inhibitors

SNRIs include duloxetine, venlafaxine, desvenlafaxine, and milnacipran, and they inhibit the presynaptic reuptake of serotonin, norepinephrine, and, to a lesser extent, dopamine. They are associated with fewer side effects than TCAs and can be titrated quicker, but titration cannot be monitored by serum levels. Duloxetine more potently blocks serotonin and norepinephrine than venlafaxine. It has been shown to be effective in the treatment of fibromyalgia and painful diabetic neuropathy. Venlafaxine has been shown to reduce neuropathic pain breast cancer treatment, decrease hyperalgesia and allodynia in neuropathic pain and atypical facial pain, and prevent migraines. Milnacipran reduces hyperalgesia mediated by monoamine and opioid systems and is effective in the treatment of fibromyalgia.

Selective Serotonin Reuptake Inhibitors

The efficacy of selective serotonin reuptake inhibitors in the treatment of chronic pain conditions has been inconsistent in clinical trials. For some conditions, such as migraine, SSRIs do not appear to be more efficacious than placebo. For other conditions, such as fibromyalgia, abdominal pain associated with irritable bowel syndrome, and painful diabetic neuropathy, there is evidence that their use can reduce pain, independent of changes in affective disorders. Because of these inconsistencies, SSRIs are not recommended as first-line agents. Instead, they should be used as alternatives to TCAs or SNRIs when they are ineffective or not well tolerated.

Novel Antidepressants

The neurobiology of pain suggests that all antidepressants would be effective for the treatment of pain, but there is little evidence to support the use of novel antidepressants such as mirtazapine, bupropion, trazodone, vortioxetine, and vilazodone. Mirtazapine has been shown to decrease the duration and intensity of chronic tension-type headaches, and bupropion has been effective in the treatment of neuropathic pain by decreasing pain intensity and interference with pain with quality of life. Several reports have suggested that trazodone is effective for treating chronic pain, but side effects can hinder titration to effective doses. Vortioxetine and vilazodone are fairly new antidepressants that selectively inhibit serotonin reuptake and also have action at 5-HT3 and HT1A receptors. They have been shown to be effective antidepressants, but their efficacy in treating pain has not yet been established.

Anticonvulsants (Antiepileptics)

Anticonvulsants reduce pain by inhibiting excessive neuronal activity. They block voltage-gated sodium channels, modulate calcium channels, inhibit excitatory amino acid neurotransmitters, and enhance amino acid-mediated inhibitory neurotransmitters. Although top-tier evidence is lacking, studies have demonstrated efficacy for the treatment of diabetic neuropathy, trigeminal neuralgia, PHN, and migraine recurrence.

First-generation anticonvulsants have been used in the treatment of pain since the 1940s, when phenytoin was first reported to be successful in the treatment of trigeminal neuralgia. Carbamazepine is the most widely studied anticonvulsant in the treatment of neuropathic pain and is still the first-line treatment for trigeminal neuralgia. Valproic acid is commonly used in the prophylaxis of migraine but has also been shown to be effective for other neuropathic pain conditions as well.

Second-generation anticonvulsants include gabapentin, pregabalin, lamotrigine and topiramate. Gabapentin and pregabalin decrease the influx of calcium, modulate the release of substance P and glutamate and activate the descending inhibitory of nociception. They are effective in the treatment of painful diabetic neuropathy, PHN, fibromyalgia, phantom limb pain, and SCI-associated pain. Using flexible titration schedules has been associated with fewer discontinuations, higher final doses, and slightly better pain relief compared to fixed schedules. Lamotrigine is effective in treating HIV-related neuropathy and central poststroke pain but has been disappointing in treating other neuropathic pain conditions. Topiramate has been effective in migraine prophylaxis, chronic LBP, lumbar radiculopathy, and painful diabetic neuropathy. It has the advantages of minimal hepatic metabolism, few drug interactions, and a long half-life; it also has the unusual side effect of weight loss.

Oxcarbazepine is a carbamazepine derivative with an improved safety profile that is considered a next-generation anticonvulsant. Studies have demonstrated its efficacy in painful diabetic neuropathy and refractory PHN. Other next-generation anticonvulsants include tiagabine, vigabatrin, retigabine, levetiracetam, and zonisamide. They have a wide spectrum of pharmacologic actions and antinociceptive effects in animals, but studies in humans are sparse. They are also associated with significant cognitive impairment and are not recommended as first-line treatments for chronic pain conditions. Combinations of anticonvulsants with complementary mechanisms of action should always be considered as they may increase effectiveness and decrease adverse effects of high-dose monotherapy.

Other Adjunctive Drugs

Antipsychotics have been used in diabetic neuropathy, PHN, headache, facial pain, musculoskeletal pain, and pain associated with AIDS and cancer, and there is increasing evidence to support their effectiveness. Compared to typical antipsychotics, atypical antipsychotics offer a broader therapeutic spectrum, have lower rates of extrapyramidal side effects, and have greater potential for augmentation in the treatment of depression. However, they are associated with metabolic disruptions, including weight gain, new-onset diabetes, hyperlipidemias, and cardiac arrhythmias, which offset their benefits.

Oral and topical local anesthetics have long been used in the treatment of pain. Topical lidocaine is approved for the treatment of PHN but has been shown to be ineffective in the treatment of HIV neuropathic pain. It does not produce significant serum levels. Oral mexiletine is effective in the treatment of diabetic neuropathy, peripheral nerve injury, and alcoholic neuropathy. There are mixed reports for its use in phantom limb pain, and it appears to be ineffective in treating cancer-related pain. The analgesic effects of mexiletine did not correlate with serum levels.

Calcium channel blockers have also been used in the treatment of chronic pain syndromes. Verapamil is the most commonly prescribed of the calcium channel blockers and is effective in the treatment of migraine and cluster headaches. Ziconotide is a neuron-specific calcium channel blocker that is derived from the toxin of the cone snail, Conus magus. It has been approved for intrathecal use in the treatment of refractory cancer or AIDS pain and has potent analgesic effects that can be synergistic with morphine. However, it produces common side effects, including dizziness, nausea, confusion, and headache, that often make it difficult to tolerate.

Several potential targets have been identified for the development of new analgesic medications, including nerve growth factor (NGF), 1.7-selective voltage-gated sodium channels, N-type voltage-gated calcium channels, and angiotensin II (Ang II) AT_2 receptors. In addition, tumor necrosis factor alpha (TNF-alpha) has been implicated to play a pivotal role at both the peripheral and central levels of pain sensitization, and anti-TNF-alpha therapy has potential use for the treatment of neuropathic pain. Despite being counterintuitive, the use of low-dose naltrexone, an opioid antagonist, has been demonstrated to reduce symptom severity in a variety of conditions such as fibromyalgia, Crohn disease, multiple sclerosis, and CRPS, but its superiority to placebo has yet to be verified.

Benzodiazepines are commonly prescribed for chronic pain patients to treat insomnia, anxiety, and spasticity, but practitioners must use them with caution in this population. Their use in treating target symptoms is not supported in the literature. In fact, benzodiazepine use has been associated with decreased activity levels, higher rates of health care visits, depression, increased domestic instability, and more disability days. In combination with opioids, their use can be counterproductive and potentially dangerous. In studies of opioid-related mortality, concomitant benzodiazepine use was frequently implicated.

Psychological Treatments

Cognitive-Behavioral Models. Many psychological treatments have been used in the treatment of pain and its associated distress and disability. Wilbert Fordyce pioneered the psychological treatment for chronic pain using an operant conditioning behavioral model. The behavioral approach focuses on understanding pain occurring in a social context. The behavior of chronic pain patients is not only reinforced by others but also reinforces the behavior of others, including physicians. Pain behaviors, such as grimacing, guarding, and taking pain medications, are indicators of perceived pain severity and functional disability. It has been shown that these pain behaviors alone are predictors of whether or not opioid analgesics were prescribed to patients. The behavioral model assumes that pain and disability will persist if these behaviors persist. In CBT, healthy behaviors are targeted for reinforcement to replace extinguished pain behaviors.

The cognitive-behavioral model of chronic pain is based on the theory that individual beliefs, attitudes, and expectations affect emotional and behavioral reactions to life experiences. Pain and the resulting pain behaviors are influenced by biologic, psychological, and socio-environmental variables. Patients will experience more negative affective responses, increased pain, and decreased functioning if the patient believes that pain is inevitable and uncontrollable. This cycle of disability can be interrupted by CBT approaches, including relaxation, biofeedback, guided imagery, meditation, motivational interviewing, external reinforcement, cognitive restructuring, and the development of coping strategies. Patients are taught to be active participants in the management of their pain by using these methods to minimize distressing thoughts and feelings. The goal of CBT is to increase activity, independence, and resourcefulness regardless of the particular technique used. CBT has been proven to be effective in reducing pain severity, pain behaviors, distress, and depression and improving coping mechanisms. These benefits can persist for longer than 6 months after the completion of treatment.

A heterogeneous group of behavioral treatments has been termed “third-wave psychotherapies.” These include acceptance and commitment treatment, behavioral activation, cognitive-behavioral analysis system psychotherapy, dialectical behavioral therapy, mindfulness-based cognitive therapy, and metacognitive therapy. These therapies incorporate a variety of psychological and social theories that emphasize the context and function of psychological phenomena rather than just form. The goal is to create strategies for creating change in context and experience to develop broad, flexible, and effective repertoires to improve quality of life. These therapies have added to the spectrum of empirically supported treatments for mental disorders and could certainly be useful in the treatment of pain disorders.

Beliefs

Beliefs can be conceptualized as the thoughts that patients have about their personal pain problems. Psychosocial dysfunction has been correlated with believing emotions are related to pain, receiving overly solicitous responses from others, and attributing dysfunction to pain. Patients have also endorsed the belief that pain signifies injury and that, therefore, activity should be avoided. A change in perceived control over pain has been shown to be a predictor of the beneficial effects of CBT in the treatment of chronic pain.

Self-efficacy, outcome expectations, and locus of control are cognitive variables derived from the social learning theory. Self-efficacy is a belief about one's ability to perform a specific behavior, and outcome expectation is a belief about the consequences of performing the behavior. Patients are more likely to engage in coping efforts that they believe are within their capabilities and will result in a positive outcome. Patients who score higher on self-efficacy measures or have an internal locus of control have reported lower levels of pain, higher pain thresholds, increased exercise performance, and more positive coping efforts.

Acceptance of pain is associated with multiple domains of the experience of pain, including lower pain intensity, less pain-related anxiety, and avoidance, less depression, less disability (physical and psychological), more daily uptime, and better work status. It predicts a patient's adjustment to the illness and is independent of catastrophizing, coping skills, and pain-related beliefs and cognitions.

Coping

Coping has been defined as an individual's cognitive and behavioral efforts to solve personal or interpersonal problems that are appraised as taxing or exceeding the person's resources. Coping strategies support the cognitive-behavioral model of chronic pain, whether they are active, passive, adaptive, or maladaptive. However, patients who remain passive or continue to use maladaptive coping strategies such as catastrophizing, ignoring or reinterpreting pain sensations, or diverting attention from pain have a higher disability.

The effectiveness of particular coping strategies is dependent on many aspects of the patient’s experience with pain. For example, reinterpreting pain as not being a sign of ongoing injury had been found to be useful in reducing experimentally induced pain. However, focusing on pain and disability for greater amounts of time, as is needed in CBT to help patients reinterpret pain sensations, can be detrimental. It can divert patients from engaging in social activities and healthy behaviors. When patients perceive an increase in the effectiveness of a new coping skill, pain and disability can be decreased, and their use of maladaptive coping strategies can be reduced.

Physical Therapy and Exercise

As pain persists, activity is often reduced, resulting in deconditioning of the individual. The deconditioning and worsening of pain, in turn, results in less activity. As such, resuming normal activity should be an immediate goal. Delaying the resumption of normal activity may lead to greater disability. Physical exercise has been established as a safe and effective way to help manage chronic, debilitating pain. Treatment with physical therapy aims to reduce disability and suffering by reducing pain and increasing tolerance to movement. Physical therapy, aqua therapy, and exercise to improve range of motion and strength can improve function, pain, and quality of life in patients with chronic pain. Yoga, aerobic training, and supervised strength training have also been shown to reduce pain. Mood and levels of anxiety may also improve. The importance of physical therapy and exercise must be stressed as an essential component of a pain recovery program.

Placebo Response

Placebo effects are complex phenomena that are similar to those of active treatments and can be biologically measured. The literature supporting the placebo response has been criticized as flawed and overrated. Clinically, it is difficult to distinguish improvements from treatment and placebo responses. In the past, placebo interventions were a part of the paternalistic practitioner's treatment armamentarium. Currently, however, controversy exists over the evidence of clinically important effects that justify the use of placebo treatments. A positive placebo response does not prove that the patient's pain is psychogenic or that the patient would not benefit from active treatment, and the use of a placebo to determine if the patient's pain is "real" is dishonest and counterproductive. Such an intervention can result in the loss of trust in the practitioner and render future treatments less effective.

Multidisciplinary Rehabilitation

Chronic pain patients have greater reductions in physical, psychological, and social well-being and lower-rated quality of life than those with any other medical condition. The goal of treating chronic pain is to ease suffering and end disability that prevents people from returning to productive activities. Evidence suggests that early failures in single-modality treatment for chronic pain can have devastating consequences for future treatment effects. Furthermore, analysis of multidisciplinary programs that use comprehensive assessments, step-care treatments, and rehabilitation goals result in significant reductions in pain along with functional and quality-of-life improvements. Evidence-based practice guidelines emphasize interdisciplinary rehabilitation, integrated treatment, and patient selection criteria. Furthermore, interdisciplinary pain rehabilitation programs (IPRPs) are efficacious in returning patients to work and ultimately can reduce the economic burden of chronic pain.

Invasive Treatments

Nerve Blocks. In the treatment of chronic pain, injections can be performed for diagnostic and/or therapeutic purposes. Injections with local anesthetics may work by enhancing blood flow, releasing entrapped nerves, and interrupting the processes involved in central sensitization. Adding corticosteroids to local anesthetics may provide the added benefits of blocking the inflammatory cascade and inhibiting the synthesis of prostaglandins which can sensitize nociceptors.

Although nerve blocks are rarely curative, they may provide pain relief for a time and facilitate rehabilitative therapy. Peripheral nerve blocks that can provide relief in properly selected patients include trigger point injections for myofascial pain, intra-articular injections for osteoarthritis, and direct nerve blocks for entrapment syndromes. Spinal injections include epidural steroid injections for radicular pain and facet injections, sometimes with radiofrequency denervation, for facet joint pain. Other neuroablative procedures using alcohol or glycerol, such as celiac plexus neurolysis for pancreatic cancer and hypogastric plexus neurolysis for pelvic pain, can provide pain relief for several months. However, they can be associated with severe rebound pain and should be avoided in patients with nonmalignant pain conditions.

Electrical Stimulation

Spinal cord stimulation (SCS) is a reversible treatment that uses low-voltage electrical pulses to manage chronic, intractable pain conditions. The procedure includes the implantation of a pulse generator, extension cable, and leads, which are placed in the dorsal spinal column near the nerve root that is thought to be associated with the pain. An electrical current induces paresthesia or a tingling sensation that is meant to mask the pain. SCS was developed based on the gate-control theory, which postulates that activation of peripheral sensory A-fibers can attenuate pain signaling by C-fibers. SCS is typically used in pain conditions that have been refractory to other treatments, including failed back surgery syndrome and CRPS.

Surgery

Surgical interventions should be reserved for chronic pain patients who have failed to respond to more conservative measures. It can be challenging to decide when operative procedures are appropriate because of the difficulty involved in establishing a causative relationship between the targeted pathologic process and pain. In many individuals, the targeted disease may not be the primary cause of pain. Rarely does removing a painful body part result in long-term pain relief. Furthermore, the sequelae of the surgery itself can result in worsening pain, such as when scar tissue or adhesions develop. Careful patient selection is a necessity in deciding who should receive surgery for relief of a chronic pain condition.

CONCLUSION

Chronic pain is a significant public health problem and negatively impacts those who are affected by it. The traditional view that pain is either caused by tissue pathology (nociception) or psychological states (suffering) is too simplistic and outdated. It is now widely recognized that the relationship between pain and nociception is complex and must be understood in terms of the situation of the whole individual. There are high rates of psychiatric comorbidity in patients with chronic pain, and a psychiatric evaluation should be part of any comprehensive pain evaluation. Psychiatric diagnosis and treatment can add an essential and often neglected component to the conceptualization and treatment of chronic pain disorders. Care for patients with chronic pain should occur in a medical treatment setting, as this is the most effective way to reassure patients that the somatic elements of their pain are not neglected, and it allows effective integration of somatic and psychological treatments. A shift from traditional, single-provider care to a multispecialty, integrative approach results in improved outcomes and reduced economic burden.

FURTHER READINGS

Blair MJ, Krebs EE. Fibromyalgia. *Ann Intern Med*. 2020;172(5):ITC33–ITC48.

Borsook D, Youssef AM, Simons L, Elman I, Eccleston C. When pain gets stuck: the evolution of pain chronification and treatment resistance. *Pain*. 2018;159(12):2421–2436.

Busse JW, Wang Li, Kamaledin M. Opioids for chronic noncancer pain: a systematic review and meta-analysis. *JAMA*. 2018;320(23):2448–2460.

Clark MR. Pain. In: Coffey CE, Cummings JL, eds. *Textbook of Geriatric Neuropsychiatry*. 3rd ed. American Psychiatric Press; 2011.

Clark MR. Pain management. In: Levenson JL, ed. *Textbook of Psychosomatic Medicine*. 2nd ed. American Psychiatric Publishing, Inc.; 2011.

Clark MR, Owen JA. Pain management. In: Ferrando S, ed. *Clinical Handbook of Psychopharmacology in the Medically Ill*. American Psychiatric Publishing, Inc.; 2010:613–656.

Dharmshaktu P, Tayal V, Kalra BS. Efficacy of antidepressants as analgesics: a review. *J Clin Pharmacol*. 2012;52(1):6–17.

Dunphy L, Penna M, El-Kafsi J. Somatic symptom disorder: a diagnostic dilemma. *BMJ Case Rep*. 2019;12(11):e231550.

Finnerup NB, Kuner R, Jensen TS. Neuropathic pain: from mechanisms to treatment. *Physiol Rev*. 2021;101(1):259–301.

Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. *Cochrane Database Syst Rev*. 2017;4(4):CD011279.

Haibach JP, Beehler GP, Dollar KM, Finnell DS. Moving toward integrated behavioral intervention for treating multimorbidity among chronic pain, depression, and substance-use disorders in primary care. *Med Care*. 2014;52(4):322–327.

Henschke N, Kamper SJ, Maher CG. The epidemiology and economic consequences of pain. *Mayo Clin Proc*. 2015;90(1):139–147.

Hruschak V, Cochran G. Psychosocial predictors in the transition from acute to chronic pain: a systematic review. *Psychol Health Med*. 2018;10:1151–1167.

Incayawar M, Maldonado-Bouchard S. *Overlapping Pain and Psychiatric Syndromes*. Oxford University Press; 2020.

International Association for the Study of Pain. IASP taxonomy. http://www.iasp-pain.org.

IsHak WW, Wen RY, Naghdechi L, et al. Pain and depression: a systematic review. *Harv Rev Psychiatry*. 2018;26(6):352–363.

Kahla KG, Wintera L, Schweiger U. The third wave of cognitive behavioural therapies: what is new and what is effective? *Curr Opin Psychiatry*. 2012;25:522–528.

Kamper SJ, Apeldoorn AT, Chiarotto A, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. *BMJ*. 2015;350:h444.

Kessler A, Yoo M, Calisoff R. Complex regional pain syndrome: an updated comprehensive review. *NeuroRehabilitation*. 2020;47(3):253–264.

Knoerl R, Smith EML, Weisberg J. Chronic pain and cognitive behavioral therapy: an integrative review. *West J Nurs Res*. 2016;38(5):596–628.

Manchikanti L, Kaye AM, Knezevic NN, et al. Responsible, safe, and effective prescription of opioids for chronic non-cancer pain: American Society of Interventional Pain Physicians (ASIPP) Guidelines. *Pain Physician*. 2017;20(25):S3–S92.

Manchikanti L, Singh V, Kaye AD, Hirsch JA. Lessons for better pain management in the future: learning from the past. *Pain Ther*. 2020;9(2):373–391.

Manharpa A, Becker WC. Pain and addiction: an integrative therapeutic approach. *Med Clin North Am*. 2018;102(4):745–763.

Mills SEE, Nicolson KP, Smith BH. Chronic pain: a review of its epidemiology and associated factors in population-based studies. *Br J Anaesth*. 2019;123(2):e273–e283.

Obata H. Analgesic mechanisms of antidepressants for neuropathic pain. *Int J Mol Sci*. 2017;18(11):2483.

Racine M. Chronic pain and suicide risk: a comprehensive review. *Prog Neuropsy Biol Psychiatry*. 2018;87(Pt B):269–280.

Rice ASC, Smith BH, Blyth FM. Pain and the global burden of disease. *Pain*. 2016;157(4):791–796.

Seidel S, Aigner M, Ossege M, Pernicka E, Wildner B, Sycha T. Antipsychotics for acute and chronic pain in adults. *Cochrane Database Syst Rev*. 2013;(8):CD004844.

Urits I, Burshtein A, Sharma M, et al. Low back pain, a comprehensive review: pathophysiology, diagnosis, and treatment. *Curr Pain Headache Rep*. 2019;23(3)23.

Williams AC, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. *Cochrane Database Syst Rev*. 2020;8(8):CD007407.

28

Additional Conditions That May Be a Focus of Clinical Attention

28.1 Malingering

FRANK JOHN NINIVAGGI, M.D.

The condition termed *malingering* was formally introduced into psychiatry in 1980 in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III). Psychiatrists deal with health and examine disorders to discern whether they are real, imagined, or created. DSM-5 (2013) denotes *malingering* as the intentional production of false or grossly exaggerated physical or psychological symptoms motivated by external incentives and states aspects of it may represent adaptive behavior. The symptoms of malingering are under voluntary control. This behavior reflects intentionally contrived statements and actions with both conscious and nonconscious components, all of which intend to mislead the evaluator, usually in medical and legal contexts.

Emphasis is given to the conscious awareness the malingerer has of the malingering presentation as a behavioral strategy. Malingering is the production of, for example, false or exaggerated symptoms using a deliberately invented presentation. The malingered presentation is fluid, can change over time, and shift across changing contexts, circumstances, and situational demands. Malingering, as a form of illness behavior, suggests dysfunction but not real psychopathology. This section discusses the psychiatric condition and clinical indicators suggestive of malingering, primarily as they present in the psychiatric assessment. Neuropsychological investigations are often used as one part of comprehensive evaluations. Legal considerations and determinations are not part of the psychiatrist’s role. Empirically reliable evaluations include multimodal elements: clinical interviews over time, mental health histories, legal and criminal histories, objective psychological testing, family, occupational and medical, and other relevant collateral information. Informant reliability presumes that information is genuine, accurate, and not biased. An evaluation *team* is best to gather the elements mentioned above.

Misconceptions regarding malingering abound. Some consider it rare; some think it is common. Some equate all medical deception with malingering and correlate this feigning with a wholesale antisocial personality disorder. However, studies show antisocial personality disorder is not more likely to malinger or be more successful than the general population. Mild exaggeration of symptoms and efforts at "impression management" or reasonable attempts to put oneself in a good light does not constitute malingering. When malingering is present, it is thought to be a static response with a stable base rate; this is not true. Some think of it as a diagnosis with clear-cut criteria; instead, it is a condition with broad parameters. Malingering does not preclude genuine disorders, such as post-traumatic stress disorder (PTSD), depression, traumatic brain injury (TBI), and others discussed later. While DSM-5 gives malingering criteria, some have suggested these require further precision to ensure greater validity.

The 2001 edition of the *New Oxford American Dictionary* denotes a *malingerer* as one who exaggerates or feigns illness to escape duty or work. It dates the origin of the term in English to the very early 19th century, describing it as "back-formation from *malingere*, apparently from French *malingre*, perhaps formed as *mal*- 'wrongly, improperly' + *haingre* 'weak,' probably of Germanic origin." The term *malingerer* is said to have first appeared in 1785 in a French dictionary. It denoted a soldier who feigns sickness or induces or protracts an illness to avoid his duty. This 200-year history suggests only relatively recent attention to an individual problem with critical social implications. Malingering is socially problematic behavior.

The relative newness of this DSM entity has brought with it a plethora of conceptual difficulties. In the last 25 years, the psychiatric, psychological, and sociologic literature on this topic has increased. Contributions from these varied sources have been replete with various terms attempting to capture core meanings and attendant nuances. The following definitions will hone the concept and establish a tentative yet uniform lexicon for this section.

1. *Response style* denotes how an individual who is being evaluated communicates self-disclosures (e.g., in revealing or deceiving ways). Replies include personal goals, desired consequences, and their predicted utility in a specific instance and a particular context with a situational demand. Response styles include all linguistic and paralinguistic communications.
2. *Dissimulation* is the broadest term encompassing all forms of deception, often regarding psychological symptomatology. *Dissimulate* denotes deliberate misrepresentations, concealing, hiding, distorting, or disguising perceived truth under cover of deliberate, voluntarily controlled pretense. The dissimulated

appearance is fake, not genuine, counterfeit, and a sham. In the vernacular, this concept resembles a *hoax*, a falsehood intended to deceive or produce a facade of deliberate trickery. If intended to gain an external advantage, it suggests malingering, primarily if symptoms are used.

3. *Dissimulation* concepts encompass three principal subcategories:
    a. Malingering and factitious disorder
    b. “Defensiveness”
    c. Markedly inconsistent, evasive responses

Dissimulation in psychiatry correlates with malingering, *faking good*, *faking bad*, and factitious disorder.

Malingering is medical deception, asserting one has significantly distressing symptomatology in its actual absence. However, malingering spans a spectrum from complete misrepresentation to partial shifts in emphasis that skew the presentation to favor the patient’s receiving some desired external goal. Its range of shaded meanings points to a core denotation of *deliberate illness misrepresentation to obtain an external reward that is unmerited*. In this sense, the *gross* exaggeration of feigned symptoms for external gain in ordinary circumstances is typically not merely a plea for help; it suggests malingering. Nuanced terminology about such situations includes feigning, fabricating, disguising, deceiving, lying, misleading, masquerading, tricking, and faking. The terms “ruse,” “subterfuge,” and “wile” also capture the intention to deceive.

Factitious disorder is the presentation of physical or psychological symptoms (sometimes “signs”) that are self-induced, thus deceptively created with the inferred goal of achieving the sick role to receive medical help. Malingering and factitious disorder are a deliberate misrepresentation of symptoms. They differ because the malingerer seeks an actual external reward, whereas the patient with factitious disorder appears to look for sympathy and care, short of genuine material gains.

Some researchers restrict the denotation of *feigning* to mean deliberate fabrication or gross exaggeration of symptoms *with no assumption about its goals*. In feigning, the underlying motivation is not or cannot be determined. This use is a precise denotation used when psychological tests cannot validate and assess an individual’s motivations—intended purposes or reasons for actions—a key feature of malingering. Psychological tests can establish feigning. Often, feigning means, simply, intentional deception, and this is the way it is used in this section unless qualified. However, the malingerer may be considered deliberately to take on an impostor’s role, one whose *conscious motivation* seeks to achieve an intentionally misleading and deceptive appearance to get an external reward. In this section, the word “motivation” is used as found in the malingering literature. In other nontechnical contexts, motivation is used to mean reasons or the drive behind one’s actions in general; this has a positive connotation. By contrast, the word "motive" connotes ulterior and suspicious reasons behind specific actions; this use has a negative sense in those everyday frameworks.

The term *defensiveness* has been used in the malingering literature, at times, loosely. Typically, it denotes the masking (e.g., deliberate denial or gross minimization) of symptoms, difficulties, or impairments, in the malingering studies. It would be better not to use this term or to qualify it unambiguously. For example, defensiveness could signify *consciously intended denial or minimization (e.g., faking good) or gross distortion (e.g., faking bad) of symptomatology*. The malingering literature emphasizes "defensiveness" when medical patients minimize or deny their real symptoms, that is, *faking good*. Using *faking bad* in medical contexts suggests malingering since some external incentive or gain is presumed. These clarifications distinguish it from the concept "defensiveness" in psychiatry, which means ego defense mechanisms elicited to modulate and contain the anxiety caused by unrefined impulsive drives and emotions seeking to overwhelm and impair the ego's functioning. Such defensiveness is nonconsciously motivated, whose outcome is minimization and modulation of unbearable anxiety.

"Defensiveness" as *faking good* is a dissimulation by the omission of negative material, difficulties, or impairments. It denotes consciously minimizing or denying present and perceived symptoms having negative consequences, hence their intentional suppression. It is deceptive impression management. *Faking good* is a guarded attempt actively to portray oneself in a better light than is actually the case. For example, pedophiles seeking to avoid criminal culpability might use *faking good* strategies, referred to as a "dance of denial," to minimize incriminating behaviors when being examined. In child custody cases, a parent may also intentionally *fake good* to minimize perceived negative behaviors that might jeopardize that parent's chances of gaining or maintaining custody. Deception and desired external goal point to malingering. Faking good, like faking bad, is an intentionally motivated form of response bias: actively reporting material thought to be the "right" answer to achieve a goal. *Faking good* is deliberately emphasizing good or positive aspects of oneself, that is, positive impression management that is deceptive. This response style spans the spectrum from being prosocial (positive impression management; here, not malingering) in getting the desired job to antisocial (deceptive impression management; here, suggests malingering) to deny criminal or illegitimate behaviors. Impression management in the workplace shows in various ways: disclosure, concealment, or hinting at any material that may bias.

Malingering has also been termed *faking bad*, which denotes consciously creating false, untrue, yet apparently distressing symptoms. *Faking bad* is creating a bad impression, including intentionally over-reporting, embellishing, amplifying, or elaborating present symptoms, but not as problematic as the individual tries to convey in such grossly exaggerated fashion. Symptom education through Internet-derived information is commonplace today. When a clear external goal is sought, malingering needs consideration.

The term *simulation* has been used as a synonym for malingering but will not be used in this way here. *Simulation* more precisely denotes a consciously contrived pretended situation, an imitation, or a representation (not a deceptive misrepresentation) of one process or situation using a different approach or condition in an illustrative and instructive context. The simulated appearance is usually a “make-believe” condition used for illustration or learning. For example, simulation or analog research protocols examining dissimulation or contexts in which faking is studied using a quasi-experimental design. Participants randomly assigned to an experimental group (e.g., participants asked to dissimulate and deceive) or to a control group (e.g., participants asked to be honest) are the groupings.

In the last decade, the condition of malingering within the field of psychiatry has been a focus of concern in assessments such as medico-legal contexts, for example, insurance claims, disability requests, physical injury litigation, harassment suits, post-trauma cases, neurocognitive dysfunction, chronic pain, and criminal culpability situations. The malingerer seeks *unmerited recompense*. Not only attorneys and courts but also hospital inpatient services request such assessments and consultations. The forensic evaluator (e.g., psychiatrist, psychologist, or attorney)—especially in criminal contexts—acts primarily as a factfinder, discovering facts based on credible evidence, often on a short-term basis. This consultant role differs from the therapeutic relationship that the longer-term treater establishes in a clinically oriented helping situation. More traditional, forensically related general psychiatry assessments were in the realms of suicide, professional boundary violations, informed consent, capacity, and competency, medical errors, civil commitment, confidentiality, domestic violence, and abuse and neglect.

The psychiatrist’s primary goal in evaluations is first to establish the real presence of actual disorders and then discern whether and in which areas genuine symptoms are present and whether the condition of malingering might exist. Malingering is the presence of consciously produced symptoms that are not genuine and motivated by external incentives. Psychodynamic awareness can address the role of *projective identification* (e.g., implicit bias) because input from patients, legal systems, and the psychiatric community may unwittingly and implicitly act in the coauthoring of symptoms influencing the psychiatrist’s decision-making processes striving for objectivity. Making sense of a presentation with strong emotional components requires a working knowledge of emotional intelligence. An intelligent understanding of emotions and complex feelings deepens psychiatric perspectives and interpretive options. The wide array of individual differences in response styles may have crucial emotional determinants needing deliberation in assessments. One strength distinguishing the psychiatric evaluation of complex, often ambiguous, information is an explicit recognition of unconscious psychodynamics, individual differences, and subtle unique distinctions. Considerations of the presence of what malingering connotes, especially inconsistencies and discrepancies, in any patient's presentation (e.g., self-reports, complex response styles, direct observation, and especially from collateral data) will guide the astute clinician in proper assessments. Included in treatment planning is the rational prescription of medications.

DEFINITION AND COMPARATIVE NOSOLOGY

DSM-5 describes the condition of *malingering* as follows:

> The essential feature of malingering is the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs. Under some circumstances, malingering may represent adaptive behavior—for example, feigning illness while a captive of the enemy during wartime. Malingering should be strongly suspected if any combination of the following is noted:

1. Medicolegal context of presentation (e.g., the individual is referred by an attorney to the clinician for examination, or the individual self-refers while litigation or criminal charges are pending).
2. Marked discrepancy between the individual's claimed stress or disability and the objective findings and observations.
3. Lack of cooperation during the diagnostic evaluation and in complying with the prescribed treatment regimen.
4. The presence of antisocial personality disorder.
   (Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, p. 835 [Copyright © 2022]. American Psychiatric Association. All Rights Reserved.)

Malingering differs from factitious disorder in that the motivation for the symptom production in malingering is an external incentive, whereas in factitious disorder external incentives are absent. Malingering is differentiated from functional neurological symptom disorder (conversion disorder) and other somatic symptom-related mental disorders by the intentional production of symptoms and by the obvious external incentives associated with it. Definite evidence of feigning (such as clear evidence that loss of function is present during the examination but not at home) would suggest a diagnosis of factitious disorder if the individual's apparent aim is to assume the sick role, or malingering if it is to obtain an incentive, such as money.

Thus, the malingerer through misrepresentation seeks two aims: (1) to gain a desired external, material advantages such as financial, another material benefit, probation, better sentencing or accommodations in the justice system, or illicit substances; and (2) to avoid something undesired, for example, fulfilling an obligation, a jail sentence, and military or other service obligations.

Malingering is coded in DSM-5 as Z76.5 as a condition (not a mental disorder) that may be a focus of clinical attention. DSM-5 differs from DSM-IV-TR because it adds under #1 above: “or the individual self-refers while litigation or criminal charges are pending.” This addition emphasizes the prevalence of suspected malingering in this population. DSM-5 also includes a new discussion of malingering, factitious disorder, and somatic symptom-related disorders and their distinctions. Thus, in factitious disorder, the voluntariness of pursuing the sick role as a psychological conclusion is emphasized in contrast to malingering. The voluntariness of symptom production in the quest for financial or other external goals yielding a behavioral conclusion is malingering.

The 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) lists malingering in Chapter XXI of this text as code Z76.5, page 244. It codes the following:

- Malingering (conscious simulation)
- Person feigning illness (with obvious motivation)

The ICD-10 chapter containing malingering states these categories are provided for occasions when circumstances other than a disease, injury, or external cause classifiable to categories A00–Y89 (more clear-cut medical and psychiatric disorders) are recorded as “diagnoses” or “problems.” This situation can arise when a person who may or may not be sick encounters the health services for some specific purpose or when some circumstance or problem influences the person’s health status but is not in itself a current illness or injury.

International Classification of Diseases for Mortality and Morbidity Statistics, 11th Revision, (ICD-11), v2019-04 (2021) codes and describes Malingering as code QC30:

> Malingering is the feigning, intentional production or significant exaggeration of physical or psychological symptoms, or intentional misattribution of genuine symptoms to an unrelated event or series of events when this is specifically motivated by external incentives or rewards such as escaping duty or work; mitigating punishment; obtaining medications or drugs, or receiving unmerited recompenses such as disability compensation or personal injury damages award. (Reprinted from World Health Organization. ICD-11 for Mortality and Morbidity Statistics, 2023. https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/1136473465.)

The DSM-5 guidelines used to indicate the condition of malingering are minimal and relatively broad. However, the *magnitude* of the dissimulation, that is, the creation and gross exaggeration of multiple false symptoms, must be present. There are no suggestions regarding chronologic age range or clear-cut demographic or epidemiologic findings because malingering has not been studied in the rigorous, scientific manner in which the mental disorders have been researched. Also, because malingering denotes patient deception and unreliable information, it is difficult, if not impossible, to organize findings in a consistently coherent fashion. Clinical judgment, especially about inferred motivation and conscious intention to avoid perceived distress and achieve desired goals, must be sensitive to ambiguous, indeterminate presentations. Therefore, using careful and well-chosen terminology in reports suggests doubt or suspected lack of credibility (e.g., feigning, fabrication, noncredible, and dissimulation short of malingering). When present, it is essential to document this in diagnostic evaluations.

HISTORY

Examples of malingering can be found in literature, from Ulysses feigning insanity to escape the Trojan War to King David's biblical story. The image of street beggars and vagabonds who feigned disease to gain alms is familiar in early Western and Eastern cultures. It was not until about 1880 in Germany and about 1910 in the United Kingdom that malingering moved from the political, military, and social spheres to the medical profession's aegis. World War I reinforced the phenomenon of what already was known as *compensation neurosis*. At the time, medically prescribed removal from duty was added to address war fatigue, shell shock, and war trauma, particularly physical distress. At this point, physicians took or were given a leading role in the management of malingering because it presented as a symptomatic picture in the absence of typically associated material causative features. Because many symptom profiles were strongly suggestive of "mental" and emotional distress, psychiatrists and psychoanalysts then took the lead in trying to sort out genuine from feigned illness.

From its introduction in the DSM-III in 1980 to about 2005, there was an almost threefold increase in malingering scientific literature. Historical accounts have been written, and clinical and empirical case reports have appeared, but evidence-based research is still in its infancy. Hence, as a scientifically justifiable construct, malingering remains more in the context of discovery. The research phase in which evidence is brought to bear on objectively testable hypotheses remains in statu nascendi.

EPIDEMIOLOGY

The actual incidence, prevalence, and distribution features associated with malingering are imprecise. This uncertainty is because malingering has not achieved status as a formal psychiatric diagnosis with attendant studies examining validity and reliability in the context of clinically defined populations. Speculations suggest base rates can vary from 7.5% to 33%. In ordinary clinical practice, some have suggested a 1% prevalence; in military populations, a 5%

occurrence, and in litigious, criminal populations, up to 20%. A 2002 study of 131 clinicians of the American Board of Clinical Neuropsychology gave the following breakdown of base rates for different contexts of malingering: 29% for personal injury, 30% for disability or worker's compensation, 19% in criminal cases, and 8% in medical or psychiatric cases.

ETIOLOGY

The etiologic factors contributing to the spectrum of malingering presentations are broad. They are related to much larger questions concerning the structure of and motivations within human nature. A crucial element is "why" the illness behavior was deliberately manufactured. Cognitive development and refinement, introspection, insight, accurate self-analysis, ego defense mechanisms, adaptation and survival, moral development, openness to self-disclosure, honesty, and the capacity deliberately to lie all play into an assessment of the mechanisms contributing to malingering. Basic knowledge of child and adolescent development, especially the ability to lie, is useful. For example, lying is enabled for about 2 years and expands until 10. Lying is detectable by age 4. Malingering may be construed to carry a pejorative judgment, and so it poses significant professional and legal implications for the doctor if it is challenged. Hence, diligent, parsimonious, perspicacious, and paused approaches to a patient's motivations as intentions are essential ingredients. They should lead to fair and reasonable conclusions.

A thought-provoking 1994 study by Richard Rogers, Ph.D., a recognized leader in the malingering field, proposed the consideration of three explanatory models of malingering. He used a prototypical analysis and subjects who were trained psychologists with extensive experience with malingerers. Etiologic determinants showed adaptational attributes such as making rational decisions based on expected rewards, attempting to cope with challenging circumstances, weighing alternatives, trying to make the best of a bad situation, and facing an unsympathetic or adversarial system were first-rank factors explaining the behavior of response styles. Antisocial and criminologic factors were in the second rank. A psychopathological model was favored by only 3.4% of the forensic experts. This study emphasized that the choice of one's explanatory model may have important implications for treatment recommendations. For example, the adaptational and pathogenic models carried with them the sense of more positive motivation for recommending treatment. The fourth edition of Rogers' book (2020) again restates his explanatory, predicted-utility model: an adaptational and contextual position that flexible response styles over time are based on options available and desired outcomes. This perspective, rich and laudatory, differs from the current DSM-5 emphasis.

The biopsychosocial model in psychiatry, both theoretically and applied to individual cases, may be one of the best ways to consider possible etiologic factors until further research reveals more objectively verifiable causes. This approach includes the rational integration of data from clinical interviews and collateral information: biologic factors such as temperament and medical condition, psychological factors such as defense mechanisms, coping, adaptive styles, and social factors such as current stressors, interpersonal, occupational, and financial considerations, medical and legal matters, military obligations, and situationally specific challenges offering the patient pragmatic options.

DIAGNOSIS AND CLINICAL FEATURES

For the psychiatrist, the clinical diagnoses derive from several interviews with the patient during which an interpersonal and social field is established. Presenting problems, past history, and mental status are fundamental. Preexisting symptoms need clarification to discern if they are resolved, residual, or now being brought to the surface. Since the psychiatrist as a clinician might have only brief contact with one or several sessions, a highly detailed present problem survey with observable signs, reported symptoms, and the patient's detailed elaboration of these is crucial. Continuity versus gross contradictions over time must be noted. In this context and the specific situational demands, all aspects of the patient's demeanor are noted. Self-reports and response styles, complex and often nuanced, are carefully observed and explored. Concerning signals suggesting malingering include vague responses, little or no detail, and inconsistencies with reports and observed behaviors (e.g., having schizophrenia but showing no negative signs such as flat affect, a paucity of thought content and production, and anhedonia). Cognition, memory, social, and emotional responsivity, and behavior may be tested. Issues around the manner (e.g., passive and helpless versus more aggressive and grandiose) in which the patient views and may express a variety of emotions (e.g., envy and greed—with their implicit spoiling and destructive characteristics—have clinically relevant value). Initially, the patient's material may be considered raw data that only gradually undergo an evaluative and interpretive analysis and subsequent synthesis. All the patient's productions are initially taken at face value; some include sets of partial intangibles.

Conclusions can be drawn only after diligent consideration and further collation using collateral information when available (e.g., neuropsychological testing, medical reports, laboratory investigations, past medical history, past psychiatric history, and psychosocial and legal history). *Short of the only genuine gold standard for establishing malingering, which is the patient's full disclosure of deliberate intention to malinger, coding for malingering remains a matter of clinical judgment and professional opinion*. A judicious, prudent, and economical approach lists the evidence suggesting malingering but does not reach a definitive conclusion unless the evidence is overwhelming and supported by clear-cut proof.

Projective Identification

Projective identification is a fundamental communicational and implicit information processing mechanism. Contemporary psychology recognizes it as implicit bias, which refers to the attitude or stereotypes that unconsciously affect our understanding, actions, and decisions—we don't know we're doing it.

Melanie Klein in 1946 and Wilfred Bion in the 1960s and 1970s in London first introduced the construct of *projective identification*. Their seminal framework exposed a modern information technology highlighting communicational dynamics. For example, "the medium is the message" includes explicit linguistic and implicit paralinguistic data, for example, facial microexpressions of emotions the patient communicates. The sway of outside influences (e.g., suggestions in collateral information and even theoretical training) cannot be underestimated. The evaluator must recognize the pressure from the adversarial divide and the referral source. Highly trained and sensitively attentive clinicians identify these visceral elements piecing them together to formulate such complex malingering presentations.

This theoretical construct is an informational and communicational processing transaction denoting the mobilization of partial aspects of an individual's implicitly conflicted emotions, thoughts, and wishes. These data become "split off" and "intrusively projected" (i.e., information transfers) into an unwitting receiver. Material communicational indicators, that is, response styles—language, gestures, demeanor, facial expression, even silence—are the medium of these data. The goal is a subtle yet aggressive manipulation persuading the receiver/examiner to believe the sender/examinee. The core aim is to control rather than to discover. The astute clinician, sensitive to this resonance, can parse this important *additional* factor as a subtle source of subliminal persuasive data in assessing all components of malingering presentations when formulating rational clinical judgments. Further aspects of *projective identification* are discussed in the treatment and management subsection.

Truth and Malingering

Malingering depends on the clinician's conception and assessment of truth—the truthfulness and veracity of the patient's clinical presentation, especially verbal descriptions of symptoms. This calculation rests, in part, on the patient's openness to accurate self-disclosure. The assessment of malingering challenges the trust inherent in the doctor–patient relationship. Comparisons make the detection of valid from feigned symptoms and disorders to accepted DSM standards thought-provoking. The normal development of a child's sense of reality and capacity to intentionally misdirect (e.g., “lying”) is found in many child development texts. A lie is a misrepresentation of what is thought to be factual—without letting the other know that you are deliberately misrepresenting. The capacity to lie or deceive may be apparent at age 4 when the ability for perspective-taking and “theory of mind” (ToM) develops. Lies in this developmental era usually are prompted by self-preservation and the wish not to be punished.

Consistency in meeting standards, taken together with understandable atypical variations, reflects any clinical assessment’s fidelity. Marked discrepancies, mismatches, deceptive shifts of the diagnostic frame, and blatant excursions from established norms suggest deception. A high-index of suspicion is raised when the patient demands medication, divulges symptoms too eagerly or dramatically, and makes blatant conditional threats of self-harm, violence, or litigation.

Research has suggested that the following are representative detection indicators for feigned mental disorders: rare symptoms (a significant indicator), improbable combinations, symptoms that rarely occur together, spurious patterns, erroneous stereotypes, endorsing a very high-proportion of symptoms (an important detection indicator), endorsing a wide array of minor psychological problems, reporting symptoms of grossly extreme severity, and evidencing more significant impairment on reported symptoms versus observed behavior. Gross exaggeration is a detection indicator when its magnitude is strikingly amplified. When consciously deliberate intention and tangible gain can be assigned to illness deception, a malingering diagnosis can be entertained.

Table 28.1–1.
Contexts in Which Malingering May Present

1. Avoiding responsibility: Military service, child, or adolescent school avoidance
2. Criminal forensic contexts: Avoiding punishment, criminal liability, or seeking transfer from prison to a hospital (the base rate incidence in criminal cases is about 19%)
3. Medicolegal contexts: To obtain financial gain, compensation cases (personal injury, 29%; disability, 30%)
4. Emergency department or primary care outpatient clinic visits to gain admission to hospital as may be found in homeless persons seeking shelter
5. Outpatient settings: Medical requests to obtain drugs such as controlled substances, opioids, minor tranquilizers, or psychostimulants (e.g., ADHD medications: methylphenidate, dextroamphetamine, etc.)

Thus, signs of possible deception are inconsistencies and contradictions between reported history and actual facts, a clear record of misrepresentation, previous acts that are the same or similar to the current presentation, a rapid improvement when external incentives change, and absurd responses, contain unlikely combinations, and blatant contradictions. Vague responses with little detail are suspect.

Diagnosing the condition of malingering is based on careful considerations of the best evidence available and astute clinical judgment. When reaching clinical conclusions, particular attention to configurations of factors associated with malingering is useful. When psychological testing is used as collateral evidence, neuropsychological profiles are constructed using multiple tests, each tapping a different area. There are no established unambiguous criteria. From the patient's side, good faith, openness to self-disclosure, and well-intentioned honesty are presumed present but examined for possible discrepancies. The extent to which such discrepant information is conveyed in what is suspected to be a deliberately deceitful manner lends itself to a high-index of malingering suspicion. Table 28.1–1 summarizes contexts in which malingering may present itself.

CONTEXTS IN WHICH MALINGERING MAY PRESENT

Avoiding Responsibility

Avoiding responsibility has developmental and social determinants. The current presence of international unrest and war makes the context of military service a continuing impetus for some to use malingering to avoid a personal threat. Avoiding responsibility may also appear in child and adolescent contexts in the frequently encountered condition of school avoidance. Although an anxiety disorder may be the underlying cause of such behaviors, many youngsters also intentionally feign illness to avoid going to school or leaving their homes. Malingering may not be an essential feature. Truancy may be associated with more salient disorders such as conduct disorder, learning disorders, or understandable environmentally based problems such as bullying or an impaired family situation in which neglect, abuse, or both exist.

Avoiding responsibility has shown itself in the workplace. Employees may seek leaves of absence or reduced work schedules because of claimed symptoms or illnesses. If these claims are not genuine and include a wish for payment, malingering is a consideration. However, in the current era of COVID-19, a new phenomenon has emerged in more explicit ways than before. Some have speculated about an unproductive and dangerous behavior loosely termed "reverse malingering." For example, genuine malingering occurs when a person believes they are well but feigns illness to obtain a hoped-for compensation for a leave of absence from work. Malingerers, well in general, use the deception of having distressing symptoms to avoid working. By contrast, health professionals say that people who know they are ill are told not to work if sick yet continue to work. This untruthfulness may be "reverse malingering." They reverse the direction of their health status from claiming illness to behaving as if they are well.

Unlike in malingering, they wish to receive “merited” recompense. Symptomatic people put themselves and their coworkers at risk if they have symptoms or are sick, do not mention it, and go to work, although advised to stay home. They want to continue working to maintain getting paid. This passive deception is used because they fear potential unemployment and income loss. Hiding actual illness is a medical deception to achieve the gain from working. This feigning may be a variant of *faking good.* Deception and questionable judgment are key. Such situations pose dilemmas in discerning the nuances of the dissimulation spectrum.

Criminal and Forensic Contexts

Criminal and forensic contexts in which adults and adolescents find themselves in situations in which malingering offers an option to avoid unwanted legal prosecution or confinement. Feigning psychiatric illness may offer a person convicted of a crime the opportunity to go to a psychiatric hospital rather than to jail. Malingerers might feign schizophrenic illness to avoid jail terms or, if in prison, to be transferred to better quarters. The issues surrounding criminal behaviors and psychiatric illness are complex. Appropriate assessment and recommendations require skilled forensic evaluators. Malingering has been seen in undocumented immigrants caught for illegal entry into a country and threatened with deportation back to their country of origin. Again, such situations must elicit a pause so that fair and judicious determinations are made.

The court referred a 15-year-old boy to a juvenile justice psychiatric evaluation clinic. He was being held at a detention center for multiple charges: possessing cannabis and crack cocaine, setting fire to his family's mobile home, expressing violence toward a parent, and resisting arrest. The arresting officers described him as uncooperative, at times belligerent, and communicatively unclear. Several years before this episode, they had known him since he and his mother had involved authorities because of an allegation of physical abuse by one of his mother's live-in boyfriends. He was brought to a local emergency department at the point of arrest. He was examined and found to be in good physical health; he became quiet and did not appear to exhibit any clear-cut psychiatric symptomatology. Shortly after, he was brought to the clinic daily for a series of psychiatric and psychosocial examinations and psychological testing. The patient was reticently compliant and demonstrated minimal expressive communication. He did admit to illicit substance use for several years, a history of school avoidance and failure, and a chronic family conflict history. When asked about the fire setting, he did not reply. Family interviews revealed a chaotic household; the biologic father was incarcerated. The mother had multiple male companions and a history of depression with suicidal attempts, and there was an overall lack of structure and home supervision. Psychological testing showed a full-scale intelligence quotient (IQ) of 75. Some disorganization and looseness of thinking, with pervasive deficits in adaptive and academic functioning, and projective tests suggested possible psychosis but no depression. He and his mother wanted to avoid incarceration and emphasized a history of abuse. The patient then claimed to be depressed; his mother vociferously supported this. The patient then tried to elaborate on what he thought were symptoms of depression: "I can't eat or sleep or play sports; I see monsters in my dreams." The patient, however, appeared robust, muscular, and engaged in soccer at school. Recent interviews, psychological testing, diagnoses of psychotic disorder unspecified, substance abuse disorders, borderline intellectual functioning, and PTSD were made from all the available past history. Conduct disorder was to be ruled out. Although both mother and son intentionally tried to establish a primary diagnosis of clinical depression alone to avoid prosecution, the evaluator did not concur. Moreover, adding malingering to the diagnostic profile was thought to be unhelpful in the context of this already complex group of psychiatric disorders. Recommendations were made for therapeutic services: long-term therapeutic residential placement, consideration of psychosocial therapy and psychotropic medication for PTSD, substance abuse, and psychotic symptoms.

Medicolegal Contexts

Medicolegal contexts can suggest malingering when issues of unmerited financial compensation are at stake. This situation may involve personal injury cases and disability claims for psychiatric or developmental problems. These complex cases have salient medical and neurocognitive components. Assessment requires medical specialists to weigh the merits of neurologic and musculoskeletal features and psychiatric and psychological evaluations.

Police brought a 16-year-old boy to the emergency department. He was wildly out of control, moving in a disorganized, hyperkinetic fashion, and speaking incoherently. Thinking showed loose associations. Initial physical examination was unremarkable, but laboratory investigations showed elevated liver function tests; drug screening was negative. The patient's mother said he had no prior medical or psychiatric history, and she was unaware of any indication of substance abuse or cigarette smoking. The patient was admitted to the psychiatric unit, where he was moderately agitated, demonstrated a catatonic-like excitement, and reported visual hallucinations. Atypical neuroleptics were begun; after several days, his motor behavior began to normalize, and he was more coherent. His cognitive functioning cleared but was slow. He denied using illicit substances but said he had recently started a summer job in a small paint factory; this appeared to coincide with his symptoms. A tentative diagnosis of inhalant-induced psychotic disorder was made. The patient was discharged on no psychotropic medication and seen 1 week later. He appeared to be stable and in no acute distress. His thinking was clear, with no reports of hallucinations. Sleep, appetite, and energy level were normal. Blood work was within normal limits at this time. He returned to school in the fall and was reported to be doing as well as before his decompensation. His mother then requested the hospital physicians support her contention that her son was permanently disabled, thereby certifying him to be eligible for long-term financial compensation. After further interviews, no evidence of changed or limited functioning from prehospital baselines could be found to support disability claims. The mother's attorneys brought the matter to a hearing. It was agreed that the paint factory would make a small settlement. This case demonstrates how patients may shift symptom history of a real disorder and its resolution toward exaggerated claims of continued impairment.

Emergency Facilities

Emergency departments and primary care facilities may be the only resource to which socially and occupationally impaired individuals have relatively easy access. Such persons may have complex chronic medical and even psychiatric illnesses, which do not warrant immediate hospitalization. Careful diagnostic assessment of emergent, urgent, and longer-term needs is critical. Social services, particularly if an immediate shelter is needed, can be an essential resource. Whether malingering, in addition to other problems, is believed to be present, clinical judgment must dictate these cases’ appropriate disposition.

Outpatient Settings

A common situation in which malingering may present itself can be seen in typical outpatient practice for the psychiatrist and all physicians. Patients may malinger about pain symptoms that are either untrue or deliberately exaggerated. The request for controlled substances such as opioids in excess of what the physician deems reasonable always needs careful attention. Patients may be habituated to or dependent on benzodiazepines and so demand ongoing prescriptions or excessively high daily doses. The physician’s clinical judgment again must weigh reasonable risks and benefits, especially if malingering is suspected. A not infrequent request by adolescents and some adults is for psychostimulant medications such as methylphenidate and its derivatives and dextroamphetamine and its derivatives to treat alleged attention-deficit/hyperactivity disorder (ADHD) or attention-deficit disorder (ADD). Because these disorders are frequently diagnosed, careful clinical assessment is needed to discern legitimate cases from those persons wishing to get controlled substances for purposes such as weight loss or nonmedically justified stimulation, diversion, and abuse. It has been almost commonplace for adolescent girls and boys to be overly concerned about body weight and to request psychostimulant medications under the ruse of suffering from ADHD when, in fact, the underlying intent is to use the psychostimulant drug for effective weight loss. Adolescents and young adults may abuse stimulants to get through exams or merely to experience a temporary high. Of course, this represents only a subset of patients. Child and adolescent psychiatrists, and pediatricians, face this issue on an almost daily basis. Also, careful histories and assessments of the family context, particularly members with illicit drug histories or criminal backgrounds, are essential to rule out abusive use or diversion into the street for the illicit drug trade.

A 35-year-old attorney requested a psychiatric evaluation to treat what he claimed was his adult ADD. This meeting was the third such consultation he had requested in the last 6 months from different psychiatrists. The patient was a middle-class male, married with two children, and a practicing attorney for the previous 6 years. He had a reported past history of intermittent, moderate depressions since age 20 years, for which he received periodic brief psychotherapy and antidepressant medication. He stated he had not felt depressed for the last 5 years, although he described himself in the words he said his wife used: “a man who fears success.” He explained this was related to multiple law firms’ attempts at jobs that quickly ended with his sudden resignation when his intermittent bouts of worthlessness and anhedonia came to the surface. He now felt troubled with difficulties concentrating, staying focused, and completing tasks requiring attention, concentration, and perseverance. He said he was able to diagnose himself and believed he had ADD. He wanted to work and so sought a prescription for stimulant medication.

After several careful history sessions, diagnoses of anxiety and mild mood instability with possible depression were suggested, based on current and past history. After agreeing with the psychiatrist’s “skillful insights,” as he described them, he continued to assert the assessment was essentially wrong and missed understanding the core of his dilemma. He said that, although partly right, the judgment was “perfectly flawed.” However, he insisted he would come for further sessions to “show the doctor” how wrong he was and that the putative ADD required a prescription that “only the doctor had the power and authority to order. Whether or not I meet the criteria for ADD, I feel I can benefit from the drug.” Personality disorders were considered, but such diagnoses were deferred.

During about six sessions, he behaved similarly, and his uncooperative behavior increased. When it became clear that no psychostimulant medication would be prescribed, he said, “You just won’t open up and listen to me. It’s really very clear. I know myself better than you do. Even though I didn’t go to some fancy medical school, I’m competent enough to see things as they are.” He then added, “I guess we’re on different wavelengths.” Any sort of reasonable attempt to clarify his presentation, however tactful, was thwarted. The patient declared an “impasse” had been reached, and he would not return.

CASE DISCUSSION

The nature of the patient’s motivation was complex. Characterologic issues were principal. Envy of the therapist was a noticeable feature. He certainly intended to persuade the psychiatrist both of his ADD and his need for drugs. The extent to which he feigned or believed these to be true is unclear. His almost arrogant insistence on his independently arrived at self-diagnosis, inability to consider another opinion, and constant comparisons between himself and the psychiatrist, evidenced rigidity and recalcitrance, hinted at a narcissistic pathologic, unconscious envy. He partially acknowledged the therapist's skill. Yet his insistence on the assessment being "perfectly flawed" pointed to his attempt at envious spoiling of what he sensed was skillful and insightful—that is, ideal. Such uncooperativeness thwarts constructive dialogue. Envy as implicit bias is powerful and can be an aggressive effort to influence the psychiatrist. Making this unconscious and implicit display of defensive omnipotence, that is, *projective identification,* conscious of oneself as an examiner, is critical to counter-transference self-observation. It is unnecessary to interpret the patient his wish to gain control to ward off and defend against feelings of inferred inadequacy and helplessness—in a short-term setting. His choice to leave the therapeutic field due to what he called an "impasse" further attested to the usual blocking of forwarding progress resulting from envy and narcissism in the therapeutic impasse. The boundary between malingering and factitious disorder, as seen in this case, is thin.

CLINICAL PRESENTATIONS OF MALINGERING

Physical Symptomatology

Pain and Medical Presentations. Malingered symptomatology may take various forms. Table 28.1–2 summarizes clinical presentations of malingered symptomatology. They may be divided into physical symptomatology (e.g., pain, medical with cognitive incapacity) and psychological symptomatology. Some think malingered pain and malingered amnesia are common in malingering contexts. Whereas factitious disorders often present with physical symptomatology and signs, malingering appears in psychiatric settings with psychological symptomatology suggestive of psychiatric disorders. Perhaps the presentation of pain and distress in suspected malingering is most often encountered in contexts in which financial or disability compensation is at stake, for example, lower back complaints in the work situation. This issue can be related to physical injury or degenerative disease states. Forensic and litigation disputes may be venues for this. If pain is malingered, it may typically take the form of an embellishment of the severity of pain and, subsequently, what appears to be significantly impaired functional capacity. Ordinarily, some degree of real pain is present, but the excessive degree to which it is intentionally exaggerated may suggest malingering. Because this is an extremely complex area and one in which malingering has been shown to be less prevalent than once thought, such presentations need careful consideration. When physical symptoms are in question, medical specialists can help to rule out neurologic, infectious, musculoskeletal, orthopedic, and other primarily somatically based complaints. Also, one must carefully discern the presence or absence of pain disorder or the condition of pain (e.g., neuralgia, neuropathy, arthritis) associated with a general medical condition, the code for which is specifically selected based on the associated medical condition or the anatomical site.

Table 28.1–2.
Clinical Presentations in Which Malingered Symptomatology May Present

1. **Physical:**
    a. Physical pain and incapacity due to injury or chronic medical condition
    b. Psychogenic nonepileptic seizures
    c. Traumatic brain injuries
    d. Amnesia
    Dementias: *Alzheimer's, Frontotemporal, Lewy body, Vascular, Traumatic brain injury, HIV, Prion, Parkinson's, Huntington's, or due to another medical condition*
    e. Delirium
2. **Psychological:**
    a. Post-traumatic stress disorder
    b. Depression
    c. Amnesia
    d. Psychosis
    e.
    Neurodevelopmental: (unlikely)
        i. Intellectual disabilities
        ii. Borderline intellectual functioning
    f. Miscellaneous disorders: Attention-deficit/hyperactivity disorder; Ganser syndrome

Psychogenic Nonepileptic Seizures. Psychogenic nonepileptic seizures (aka PNES, pseudoseizures, psychoseizures, psychogenic nonepileptic spells, hystero-epilepsy) resemble epileptic seizures. Prevalence is suggested to be about 5% of presentations in emergency departments. Patients exhibiting PNES are unusually suggestible. They may also have real seizure disorders or a family history of epilepsy. Although they resemble tonic–clonic convulsions, no tonic phase is present; there is an asynchronous thrashing of the limbs typically lasting more than a few minutes, unlike in actual seizures. Consciousness may be normal or lost, and typical postictal phase or loss of continence, tongue bite, and headache are variable. Laboratory studies show no electrocerebral changes. Serum prolactin, ordinarily increased between 15 and 30 minutes after a real tonic–clonic seizure, remains unchanged from baseline. Antiepileptic drug treatment has little, typically no, therapeutic effect.

In some cases, hysterical conversion reactions associated with acute emotional stress or chronic physical and sexual abuse produce PNES. PNES also have been seen in malingering. When there is clear evidence of incompatibility with neurologic disease and when malingering has been ruled out, DSM-5 terms PNES medically unexplained and therefore to be a functional neurologic symptom disorder (i.e., DSM-5, F44. a conversion disorder).

Neurocognitive Presentations

Neurocognitive functions include consciousness and information processing abilities–attention, concentration, memory, reasoning, problem-solving, executive functions, and language. This group of presentations is complex because it may entail the intricacies of multiple levels of cognition having medical, injury, and deterioration over time. Assessment demands specialized clinicians skilled in such cases.

Cognitive functions are closely linked to neural pathways and cortical networks. Impairments in these—DSM-5 (p. 591) “Neurocognitive Disorders (NCD)”—represent a decline, typically acquired, from a previously achieved level of functioning. The NCDs are syndromes whose etiology frequently can be potentially determined.

NCDs were referred to in DSM-IV-TR as dementia, delirium, amnestic, and other cognitive disorders. They may result from central nervous system injury, moderate-to-severe TBI with significant declines in previous performance levels, concussion whose traumatic effects may be transient, mild traumatic brain injury (mTBI), toxicity, disease, or deterioration. Clinical presentations may be classified as major NCD (i.e., a substantial decline) or minor cognitive (i.e., a mild to modest decline) disorders. They may present themselves as *delirium*, *dementias such as Alzheimer’s, Frontotemporal, Lewy Body, Vascular, TBI, HIV, Prion, Parkinson’s, Huntington’s, or due to another medical condition (e.g., DSM-IV “amnestic disorder”)*.

Neuropsychiatric symptoms may begin with social withdrawal and anxiety. Once a definitive diagnosis of dementia is made, significant agitation, aggression, hallucinations, and delusions of psychosis may emerge.

They may also appear as various intellectual and memory impairments, such as occurring after motor vehicle accidents or war trauma following concussion and TBI. Overlap with psychological symptoms is common. Trauma and injury events—head injuries—typically present themselves in the ruling out of possible malingering. Examples of typical venues are personal injury litigation, workmen’s compensation, and disability claims. Neuropsychological testing and the assessment of symptom validity are essential in these cases.

Cognitive assessment relies on the comparative measurement of responses: incorrect, partially correct, and correct answers. Estimations of reasonable effort, which may be difficult to assess, help in interpretation. Unlikely presentations, excessive impairment, and the magnitude of test error are useful detection strategies when malingering is suspected. Pain, depression, and PTSD represent the most common differential diagnoses when considering mild TBI or modest decline in the person's previous performance level. Various forms of the Mini-Mental State Examination can screen for dementia. However, several factors may render results unreliable: imprecise testing of executive functions and performance variations based on the patient's educational level and motivation or effortful willingness. Apathy in dementia, typically frontotemporal and Alzheimer's occurs and should be differentiated from depression, both of which influence clinical presentations. An experienced forensic neuropsychologist's expertise is essential in determinations. Findings may include data from standardized scales assessing levels of cognitive response and diagnostic category.

Psychological Symptomatology: Clinical Presentations

Psychological symptomatology and the suggestion of mental disorders are presentations the psychiatrist encounters, often by referral or consultation, in ruling out suspected malingering. A high-index of suspicion is raised when symptoms are clearly atypical, grossly unusual, nontraditionally bizarre, extraordinarily exaggerated, embellished, extravagantly elaborated, or alleged to occur in the future if some incentive or personal gain in the present is not met. Laypersons, for example, may decide to exhibit dramatic symptoms or to portray some florid clinical picture based on what has been seen in the popular media. Professionals who malinger might consult psychiatric texts and attempt to feign a bona fide mental disorder using every feature and criterion listed and their precisely given time frames. Such direct textbook presentations are rarely present in day-to-day clinical contexts. Detection strategies include close attention to the patient's description of when and how symptoms arose and their sequence. Clinical incongruities in symptom presentation—for example, an individual describing an anxiety disorder in a calm and composed manner at every interview—suggest feigning. Finally, the markedly atypical juxtaposition of symptoms and the presence of an external incentive may suggest malingering. Some examples of psychological symptomatology are discussed in the following.

Post-Traumatic Stress Disorder. PTSD is a common presentation in military and civilian populations. It first appeared in 1980 in the third edition of the DSM (DSM-III) and was associated with the Vietnam War's legacy. Earlier related terms such as "soldier's heart," "shell shock," and "war neurosis" were used in the World

War eras. For example, “war neurosis” was equivalent to the *nevrose de guerre* and *Kriegsneurose* of the French and German medical literature.

Genuine PTSD has clear-cut, very distressing signs and symptoms persisting for a month or more: exposure to actual or threatened severe harm, intrusive, recurrent traumatic thoughts, avoidance of reminders of the trauma, negative emotional states, alterations in arousal such as hypervigilance, and sleep disturbances.

Cases involving personal injury, automobile accidents, and those exposed to terrorism and traumas associated with military service and combat may present with features of PTSD. Combat-related and civilian PTSD needs to be differentiated. Although those in the military are prepared for the unexpected, they may experience genuine feelings of helplessness, vulnerability, and surprise. Suicide patterns in the military have increased over the last decades. Military suicides continue to be at high-rates. Veterans complete approximately 20% of national suicides. Megaorganizational complexity, time pressures, and feeling overwhelmed to act efficiently and reliably may contribute to stress factors. More than half the suicides occur in those never deployed. Those who have been exposed to combat may experience a variety of postcombat stress-related sequelae. Being in the combat zone, physical assault, and heavy fighting may cause 18% to 30% of PTSD. PTSD is also associated with the breakup of relationships and subjective feelings of isolation. “Moral injury,” which differs from PTSD, has been described in military personnel who experience feelings of guilt and shame that might be related to killing others or having friends killed in combat and causing the harm or death of civilians, knowingly but without alternatives, or accidentally. Other circumstances include failing to provide medical aid to an injured civilian or service member and following orders believed illegal, immoral, or against the Geneva Conventions’ rules of engagement.

Genuine PTSD has been correlated with the patient’s minimizing combat as causative, blaming oneself versus others, experiencing dreams with themes of helplessness, general reluctance to actively speak about combat memories, feeling survivor guilt and shame, and anger. Feigned PTSD diverts care from veterans and military personnel who genuinely have suffered mental and physical trauma and legitimately require services. Currently, genuine PTSD is common.

In civilian populations, with the onset of COVID-19 in the last few years, healthcare professionals are experiencing rates of PTSD higher than previously. Whereas *burnout syndrome* was recognized about a decade ago as a cause of physical and emotional exhaustion, depersonalization, and a loss of personal efficacy, burnout is now added to actual PTSD associated with the virus. High-stress emergencies over time raise cortisol levels that do not decline to baseline as in more typical situations. Doctors feel they cannot “decompress” and reestablish perspectives. In healthcare and other fields, most cases in frontline workers are genuine and need identification and multiple interventions.

COVID-19 has become a global health emergency, both physical and psychological. For example, healthcare workers are exposed to the virus, supply shortages, and patients suffering from the virus. Also, patients admitted to the hospital with COVID-19 are ill, fear for their survival, and experience social isolation. Fear of being intubated and put on a ventilator is real. 35% of ICU survivors and those with “postintubation syndrome” can go on to experience PTSD. After discharge, lack of immediate social support because of the need to self-quarantine is real. Thus, the rise in PTSD in healthcare and frontline workers and COVID sufferers is real and must be taken seriously.

One motivation for malingering PTSD *in the general population* is thought to be financial gain—unmerited recompense. Symptom presentations in feigned cases, by contrast, often appear dramatic and highly embellished. Dramatic positive symptoms, such as flashbacks and nightmares, are emphasized. For example, nightmares might be described as relentlessly persistent, whereas traumatic nightmares often dissipate over time. Feigned experiences are reported in ways inconsistent with DSM definitions; they may reflect unrealistic caricatures seen in the media. Subtle features and generally incapacitating anxiety are often absent. For example, to avoid occupational duties, psychological and physical incapacity (e.g., disability) may be claimed. Often, on further examination, this incapacity is found not to extend to other, more pleasant activities, such as social events, recreation, travel, vacation, or sports. A thorough history may reveal past occupational difficulties or frequent job changes. The general demeanor of uncooperativeness, withholding information, belligerence, or personality disorders (e.g., antisocial, borderline, or histrionic) may also suggest—though not confirm—malingering. A bland refusal to undergo treatment and to consider future re-employment is also suggestive. Using all available collateral information is essential regarding substantiating or disconfirming self-reports and detecting discrepancies.

Real predictors of PTSD include a “lifetime trauma load” of past trauma and a history of adverse childhood events, including physical, sexual, emotional, abuse and neglect, and witnessing violence in the family. Detailed collateral data and the examinee’s assessment using psychological testing make possible a more apparent distinction between genuine and malingered PTSD.

Depression. Depression may be malingered. Because depressive disorder and depressive variations (e.g., melancholic, atypical, and bipolar) have been so extensively researched, the psychiatrist is usually perceptive in detecting faking in this area. For example, the malingerer may be extremely fluent in descriptions of “feeling depressed” and “not sleeping” but may not report the typical details of impairment (e.g., diminished specific hobbies), anhedonia, worthlessness, inappropriate guilt, ambivalence, and the destructive and distressing trends common to these, For example, disruptive interpersonal relationships, limitations in activities of daily living, and altered vegetative functions. The genuinely depressed person typically reports these symptoms. Appetite and weight may remain constant. The very loquaciousness often attending such presentations is inconsistent with depression. Bipolar depression must be ruled out. Also, malingered depressions may not permeate all aspects of a person's life and functioning. For example, the malingerer may continue to engage in going out to dinner, hobbies, and regular sports activities. Fatigue, loss of energy, degrees of anhedonia, worthlessness, and inappropriate guilt—all part of genuine depression—are absent.

Amnesia. Dissociative amnesia (F 44.0) is part of the DSM-5 category "Dissociative Disorders." This amnesia is the inability to recall important autobiographical information, usually of a traumatic or stressful nature inconsistent with ordinary forgetting. Autobiographical memory is a memory system comprising episodes recollected from an individual's life, based on a combination of two memory systems: (1) episodic, that is, personal experiences and specific objects, people, and events experienced at a particular time and place and (2) semantic, that is, general knowledge and facts about the world.

The disturbance is not attributable to any illicit substance or a neurologic or other medical condition Amnesia, with its cognitive and emotional dimensions, may be a form of malingering seen in forensic settings and is usually related to alleged crimes. Some suggest that up to one-third of homicide perpetrators claim amnesia. Because this is a grave situation, specialists in forensic psychiatry collaborating with other specialists are best suited to discern such presentations. Head injury after accidents also presents with degrees of amnesia. Such cases were historically referred to in the literature as *accident neurosis* or *compensation neurosis* and seen in a negative light until recent research suggested that genuine cognitive deficits might be present in large numbers of these cases. In assessing amnesia, highly selective, localized memory deficits, especially in the areas where the patient's motivation and liability are in question, may suggest feigning. Typically, genuine amnesic patients demonstrate impaired new learning and short-term memory, but procedural memory and premorbid semantic memory are preserved. If deficits are produced in immediate recall, digit span, autobiographical and self-memory, and new learning, this may suggest feigning if the broader context is corroborative.

Also, in genuine disorders (e.g., DSM-5 major or minor NCDs) with pathophysiologic concomitants, disorientation may be to place and to time. There may be a lack of insight into circumscribed memory deficits, and the patient may explicitly deny the presence of severe memory impairment. Apathy and emotional blandness, differing from depression, are common. Chronic substance abuse and alcoholism need to be ruled out. Dissociative states and histrionic personality disorder tend to lessen the probability of malingering. In dissociative amnesic disorder, the patient does demonstrate an inability to recall important personal information, usually of a traumatic or stressful nature. Patients with genuine dissociative amnesia typically score high on standard measures of hypnotizability, unlike the case with malingerers.

DSM-5 states individuals with factitious and malingering have been noted to continue their deception even during hypnotic or barbiturate-facilitated interviews. An antisocial personality disorder is a feature of malingering according to the DSM-5 guidelines and needs to be considered in forensic contexts. Of course, neurologic examination and relevant studies (e.g., electroencephalogram [EEG] and magnetic imaging resonance [MRI]) should be considered.

A 25-year-old female was seen for consultation in a forensic neurologic inpatient unit. The question of whether malingering was present was raised by her neurologist. She had set fire to her boyfriend’s trailer after a physical altercation, during which he struck her on the head with a blunt object several days ago. She was admitted for observation and holding. Neighbors called the police and firefighters. The trailer was demolished, and the boyfriend sustained severe burns and was unable to speak with authorities. The patient never lost consciousness; she sustained minor injuries. Her neurologic examination, EEG, CT, MRI, and blood work were noncontributory. The patient claimed amnesia for the event and asked for an attorney. After extensive interviews and the absence of a history of psychosis or mental illness, probable malingering was seriously entertained. When her boyfriend recovered after several weeks, he described her setting the fire. When she was told, she “broke down” in tears and admitted to the act.

Psychosis. Psychosis can be malingered. Examples include criminal defendants attempting to avoid trial or alter prison sentences, the homeless who present to the emergency department to gain entry into the hospital for temporary shelter and food, or those whose motivation is to avoid some clear-cut accountability they find undesirable, such as imprisonment for crimes. This behavior may be a more common malingered presentation in less sophisticated persons. It may not be challenging for the experienced clinician to differentiate true psychosis, as, for example, in schizophrenia, or bipolar mania, from deceptive illness. Persons with schizophrenia are often withdrawn and minimally verbal, not open to fluent descriptions of their symptoms. Individuals feigning psychosis often produce reports of very dramatic, perhaps melodramatic, hallucinations with bizarre content that is more characteristic of science fiction than ordinarily reported by those with genuine psychosis. Hallucinations may be uncharacteristically not accompanied by associated delusions. Feigned hallucinations are constant, not intermittent; irresistibly commanding to action with no chastising content, always terrifying rather than manageable; and, if visual, seen in black and white or conversely in dramatic colors in cinematic fashion. Malingerers may speak of delusions in a forthright, open manner, unlike the guardedness often accompanying descriptions of genuine delusions. These persons may describe their delusions as having all-or-nothing features rather than the subtler and continuous tone characteristic of truly psychotic delusions. Also, malingerers feign only the contents that accompany psychotic states. They often demonstrate no irregularities in the form (e.g., derailment, looseness of associations, and disorganization) of their thinking processes, another essential characteristic of psychosis. Also, negative symptomatology (e.g., affect blunting, avolition, alogia, and asociality) may be absent. Genuine chronic thought disorder is typically distressing and confusing.

A 35-year-old male pharmacist was referred to an outpatient clinic by his internist for medication evaluation of his long-standing bipolar disorder. He was in apparently good health and worked over 60 hours per week but experienced anxiety, dysphoria, and increasing disorganization. He reported that the last year had been turbulent in his relationship with a girlfriend of 3 years. Sleep, appetite, and job performance were beginning to suffer. The patient had been on lithium and an atypical neuroleptic for 4 years with reasonable control of symptoms. When his symptoms had recently escalated, the patient suggested to his internist that his ADHD and bipolar disorder were not adequately addressed. A once-per-day, long-acting psychostimulant was then added by the internist. Although the patient stated he felt he could function more efficiently on the stimulant, his observable behaviors dramatically deteriorated, speech became pressured, and thinking became progressively more disorganized. His prescribing physician made the referral when he demanded benzodiazepines for anxiety, escalating irritability, and insomnia. At the psychiatric interview, the patient was thin and agitated; he appeared highly intelligent and articulate. At first, he asked for increases in his lithium and neuroleptics but then demanded increases in his stimulant and the addition of a benzodiazepine. Then, it was apparent he was hypomanic and escalating, and his judgment was progressively becoming impaired. After several sessions of attempting to target the worsening bipolar presentation and get blood work (which the patient failed to procure), a recommendation for inpatient hospitalization was made. He protested and reiterated his ADHD was his primary problem and stimulant medication was the only answer. A second psychiatric opinion was obtained, and it confirmed bipolar disorder with little or no evidence of ADHD. Again, hospitalization was strongly recommended, at which point the patient abruptly left the clinic. Attempts to contact him failed. After several weeks, it was discovered he had left the state, and his whereabouts were unknown. Six months later, a family member contacted the clinic to request records, stating that the patient had been found dead in suspicious circumstances, and an investigation was in progress. This regrettable case presentation shows complexity of actual psychosis and a malingered plea for obtaining inappropriate medications in the context of diagnostic difficulties and medication complications.

Intellectual Disabilities. Serious cognitive limitations, now referred to as “intellectual disabilities,” must be considered and identified. Intellectual disabilities connote substantially reduced cognitive capacities: a significantly limited age-appropriate capacity and use of thinking skills, memory, reasoning, abstraction, judgment, and decision-making. Such intellectual limitations have traditionally been referred to as ranging from the extreme of what had previously been termed “mental retardation” to well below average or borderline intellectual functioning. These are not losses from a previously higher baseline level but relatively stable levels of cognitive functioning. It is also not very common for an adult to feign such conditions as an intellectual disability. Past history, especially academic performance, vocational functioning, daily living activities, and social adaptation, ordinarily appear in quite distinct ways as substantially impaired over extended periods in genuine cases of such disability. These impairments permeate all personality sectors and often require obvious caregiving, whether in childhood, adolescence, or adulthood.

The clinician may be presented with a malingering case by proxy, which suggests that the real malingerer (usually a parent) is alleging that their child has substantial intellectual disabilities that demand financial and other compensation. In these malingering cases imposed upon another, careful assessment of both child and caregiver’s past history is essential. Insofar as differential diagnoses to be considered, dementia in an adult (not an “intellectual disability”) represents a significant intellectual decline from a previously higher cognitive level. In neurocognitive dementias, there is a decrease in expressive language and marked perseveration; this is not a malingered condition. A person with mild dementia (mild NCD) may erroneously hold a false belief in a nonexistent illness and claim monetary compensation; this is not considered malingering.

Borderline Intellectual Functioning. Borderline intellectual functioning is a condition denoting an IQ between about 70 and 85, accompanied by significantly impaired adaptive functioning. The overall functioning is *well below average*. Average IQ ranges from about 85 to 110, with no need for adaptive supports in everyday living. Such persons often present as *slow learners* who may appear to be feigning their cognitive limitations to achieve an external gain, which is legitimately needed help from others. The often genuine needs of these persons should be differentiated from the deceptions of malingering. Also, persons of any intelligence level may innocently attribute an unlikely causative factor (e.g., allergies, job stress, or Lyme disease) to what they genuinely believe is a disorder. Such innocent misattributions should be distinguished from consciously motivated malingering. Cognitive decline, however, after a head injury, toxin exposure, or chronic substance abuse suggests a clear traumatic precipitant. Past history, medical examination, and laboratory screening for illicit substances and abnormal liver functions can help rule out legitimately understandable causes of genuine illness even if the patient misattributes cause and effect.

The single mother of an 18-year-old boy just graduating from a residential school for adolescents with academic, behavioral, and emotional problems requested documentation verifying her son was disabled on the grounds of “learning disabilities” to obtain financial assistance for college entrance. The mother immigrated from Europe and lived in the United States for over 10 years; the boy’s father died in an automobile accident 5 years previously. There were no other children, and the mother could not work due to disability secondary to depression. However, she stated she took care of her son in many activities of his daily living. The boy had been enrolled in the school about 3 years previously because of public school failure. Psychological testing showed his intelligence to be below average (Full-scale IQ 80), and there was no evidence of primary attentional problems or specific learning disorders. His academic failure was deemed due to the primary situational conflicts between him and his mother, in addition to probable borderline intellectual functioning. For the last 3 years, his school performance and behavior approached normal parameters, but his relational situation did not improve. After comprehensive retesting and assessment of psychological and social factors, it was judged he was not disabled. While the mother’s request may have been well-intentioned, it exaggerated her son’s level of impairment with the goal of procuring financial resources based on partially plausible data. Whether or not these hints of exaggeration suggest malingering is debatable. Attendance at a supportive college that would accommodate his learning needs away from home was suggested. Appropriate resource help and access pathways were discussed and made available.

Ganser Syndrome. Ganser syndrome is classified as a DSM-5 other specified dissociative disorder. It is denoted as giving approximate answers (*Vorbereiden*) to questions when not associated with dissociative amnesia or dissociative fugue. Dissociative disorders connote disruptions in integrating consciousness, memory, identity, or perception of the environment. Ganser syndrome was initially described in prisoners who presented as confused. Although the utility of this diagnostic classification has been questioned, such presentations require ruling out genuine intellectual disabilities, borderline intellectual functioning, and forms of delirium and malingering. Considerations of both inferred motivation and assessing the prospect of external gain need to be considered.

Attention-Deficit/Hyperactivity Disorder. ADHD is a commonly made diagnosis. It is characterized by significant inattention, hyperactivity, impulsivity, several symptoms before age 12 years, and functional impairment in two or more settings such as school and home. Task avoidance and low motivation (e.g., as poor effortful willingness), which may be primary in such presentations, must always be ruled out in assessing attentional difficulties. Because one aspect of the comprehensive educational, psychosocial, and psychopharmacological treatment of ADHD requires Class II controlled psychostimulant medication, adolescents, and adults have been known to feign symptomatology to gain these drugs for their perceived pleasurable and excitatory effects, which is drug abuse, or for diversion and financial gain. In *malingering by proxy* (e.g., malingering imposed on another), a caregiver might attempt to persuade the physician that the child or adolescent has ADHD symptomatology in the known absence of such to procure these controlled substances for personal use or illicit diversion. Drug use and abuse have strong associations with deceptive behaviors. Both DSM-5 and ICD-11 explicitly state that obtaining medications or drugs to gain a material advantage via deception suggests malingering.

An 8-year-old boy was brought to the clinic by his mother for medication management of his 2-year history of ADHD, which had been medicated by the family pediatrician until then. The referring physician reported he was unwilling to continue prescribing controlled substances because the mother had been unwilling to involve the child or herself in psychosocial interventions that included psychotherapy, behavior modification, parent management skills, family therapy, and adjunctive school interventions. Mother and child were interviewed both together and alone on several occasions. The mother reluctantly agreed to the release of school records. Previous psychological and academic testing showed the child to have low IQ scores but within normal limits. There was an adequate performance in all subjects except a clear-cut problem with reading. School personnel regarded this as probable dyslexia (i.e., a specific learning disorder of reading), not ADHD, and appropriate remediation was in place. The child was significantly below average in height and weight. He had multiple facial grimaces and tics and a history of insomnia and difficulty awakening in the morning. His medication regimen included dextroamphetamine in high doses and clonidine at night for sleep. His mother also requested the clinic verify the child's need for disability compensation based on his ADHD. The multidisciplinary clinic team assessment included social work home visits and an intensive evaluation of the mother's history. It was found she was on probation for a history of financial scams, writing bad checks, and involvement with the illicit drug trade. She gave informed consent for the release of her probation records. It was also discovered that state social services were in the process of investigating alleged child neglect. Before the psychiatric evaluation was completed, the clinic was informed that the mother had been arrested for possession of and dealing with illicit substances, some of which included stimulant medications. The child was put in custody of his grandparents. Recommendations were made for comprehensive psychosocial services for the child and his grandparents. They agreed that all medications should be tapered to get a new baseline of overall functioning and a new reassessment.

Child and Adolescent Illness Falsification

The importance of taking into account the developmental characteristics of the child or adolescent when considering whether a genuine illness is present cannot be overemphasized. Because individuals younger than 18 years fail to meet some guidelines used to code for malingering (e.g., antisocial personality disorder), the phrase *illness falsification* is best used to denote situations in which youth appear to feign a medical or psychiatric illness having some ulterior personal gain as a goal. Often, such *illness falsification* is associated with symptomatic lying in this age group. Also, lying in youth may have temporary adaptive value, for example, under the duress of threatened abuse and trauma.

Does *Illness Falsification* Present Differently at Different Developmental Stages? In childhood, up to about 12 years of age, separation anxiety, school avoidance, and PTSD may be the leading disorders associated with a child's feigning illness (e.g., stomachache, headache, forgetfulness, intentional suppression of information of abuse, and pseudoseizures) to obtain a desired goal such as staying home from school, avoiding consequences for negative behaviors, or wishing to receive gifts or increased attention for being sick. Children who are bullied in school might also use *illness falsification* to avoid exposure to a context of trauma. Symptom coaching, sometimes a consideration when a child falsely attributes illness to a parent, may occur in the context of an acrimonious divorce and child custody dispute. One parent may coach the child to describe the other parent in an unfairly negative light. This dissimulation is an example of an *illness falsification* by proxy.

In adolescence, about age 13 to 18 years of age, the emotions of anxiety, fear, and guilt may prompt the teenager toward *illness falsification*. Anxiety and fear of attending classes may be related to anxiety disorders, social phobia, autism spectrum disorders, and bullying. School avoidance may result. As is discussed in the section on attentional problems, it is not uncommon for a subset of adolescents to claim to ADD and ADHD as an excuse to rationalize both low motivation and intentional task avoidance and so excuse excessively poor school performance. Juvenile offenders often carry a diagnosis of conduct disorder. This disorder strongly connotes the presence of lying and deception in possibly up to 50%. These youth have been noted to use *illness falsification* to avoid sentencing to jail, lighten sentencing severity, or divert their adjudication from jail to the mental health system for rehabilitation and treatment.

Antisocial Personality Disorder and Malingering

As described by Richard Rogers in 1990, 1994, 2008, and 2020, the criminologic model proposes that the DSM-articulated framework is used, perhaps inaptly, in Rogers view, to explain the primary motivations (i.e., reasons for acting) underlying malingering. It includes the diagnostic variables of antisocial personality disorder, the contextual variable of a medicolegal evaluation, and two presentational variables—uncooperativeness and discrepancies with objective findings. The strong association between antisocial personality and deceptiveness is emphasized in this implied DSM risk model. Although there are obvious correlations between these, some studies have questioned their validity. The prejudicial role of confirmatory bias must be considered in all assessments.

The Nonmalingerer's Dilemma. A significant consideration in any malingering assessment is an awareness of the patient whose symptomatic presentation is real, genuine, and not feigned, yet suspect. Ordinarily, such patients have genuine disease processes, bona fide diagnoses, obvious impairments, limitations, and restricted functioning. Such functional incapacity typically affects all areas of their life, including home, occupational, and social spheres. Their biopsychosocial functional capacity is significantly diminished, and actual disability leaves them handicapped. The context in which this appears may be work-related, injury-related, associated with a degenerative disease, or a mental disorder.

In the process of assessing a patient's genuine functional incapacity and current needs, the extensive and rigorous interviewing, at times perceived as an interrogation, and the implicit level of accusation that questions the patient's credibility and forthrightness are often sources that strain the patient's already compromised emotional resilience. Some refer to this as "litigation distress." While "stress" and extended interviewing can be a tool in malingering interview techniques, its use must be judicious and individually based on each case. The integrity of the working relationship between examiner, caregivers, and patients may become unduly encumbered. Iatrogenic factors may appear in the patient's presentation and confound correct evaluation and management in such cases. Here, the healthy and adaptive components of the individual's resources may be mobilized. Considerations of these factors are useful guidelines to help temper any tendencies to conduct evaluations in an inappropriately aggressive manner that, in itself, may be experienced as a trauma. Therefore, all assessments should be performed with respect, open-mindedness, and clinical grace.

CODING MALINGERING

When malingering is considered present, it is useful to indicate its inferred type since the response style lies on a continuum suggesting its significance.

- *Full malingering* denotes complete fabrication in the genuine absence of symptoms.
- *Pure malingering* denotes the feigning of disease when it does not exist.
- *Positive malingering* denotes the feigning of symptoms that do not exist.
- *Partial malingering* denotes the exaggeration of existing symptoms or the statement that extinct symptoms are still present; this level of malingering may be common.
- *False imputation* is a more nebulous type of malingering. The malingerer deceptively shifts the diagnostic focus frame by ascribing actual symptoms to a cause consciously recognized to have no direct causative relationship to those symptoms. Such obfuscation darkens and obscures diagnostic clarity yet may be often overlooked.

It may also be helpful to place malingering in a context that connotes its degree of certainty:

- Indeterminate
- Strongly suspected
- Provisional
- Probable
- Deemed likely present

The American Medical Association has stated that confirmation of malingering is challenging. Table 28.1–3 summarizes a range of degrees of certainty in coding for malingering. Finally, the inadvertent misattribution of linking symptoms with unrelated causes, as, for example, found in those with cognitive limitations, should not be considered malingering.

Table 28.1–3.
Degrees of Certainty for Coding Malingering

1. *Indeterminate*: Findings suggest the possible presence of malingering according to DSM-5 guidelines but are too ambiguous, insufficient, and unsupported by collateral evidence.
2. *Strongly suspected*: Findings are consistent with basic DSM-5 guidelines but lack typically associated features and contexts; collateral support is equivocal.
3. *Provisional*: Findings are compatible with primary DSM-5 guidelines, although they are partial or too brief, yet a strong presumption exists that further indicators will be brought to light over time.
4. *Probable*: Findings are consistent with basic DSM-5 guidelines; associated features and contexts are present, but there is a lack of strong confirmation from collateral sources, especially psychological testing.
5. *Present*: Findings are consistent with basic DSM-5 guidelines, associated features, and contexts and are supported by a preponderance of evidence from most collateral sources. If expressed, acknowledgment by the subject is definitive.

Table 28.1–4.
Clinical Decision Model for Coding Malingering

1. Intentional production of false or grossly exaggerated physical or psychological symptoms
2. Motivation by external incentives:
   - Avoiding military duty
   - Avoiding work
   - Obtaining unmerited financial compensation
   - Evading criminal prosecution
   - Obtaining drugs (e.g., for recreational use, abuse, financial gain, and diversion)
   - Adaptive behavior (e.g., feigning illness while a captive during wartime)
3. Medicolegal context
4. Marked discrepancy between a person's claimed stress or disability and the objective findings
5. Lack of cooperation during the diagnostic evaluation or in complying with treatment
6. The presence of antisocial personality disorder

Table 28.1–4 summarizes DSM guidelines indicative of malingering and provides a clinical decision-making model for coding the condition. Although these entire threshold-screening indicators are not essential, factors 1 and 2 are fundamental and should be accompanied by some other typical features. Although DSM-5 does list the presence of antisocial personality disorder in these coding guidelines, some have questioned the strength of this association. Others have called attention to the connection between the clinical findings of deliberate misrepresentation and lying, for instance, commonly seen in persons with an antisocial personality disorder and in persons with a borderline personality disorder, and similar findings of intentional deception in malingering.

Conclusions and case formulation should derive from five considerations:

1. Is the clinical evaluator acting as a short-term consultant in a setting that implies a medicolegal or, perhaps, an adversarial context?
2. Is the patient's presentation reasonably consistent with the presentation of genuine illness?
3. Is there a medical or psychiatric disorder that plausibly explains the signs and symptoms presented?
4. Are the final diagnostic conclusions based on evidence of consistency or marked inconsistency?
5. What is the best professional opinion about the patient's intentions and goals?

Central issues in evaluations should include:

- the specific situational demands challenging the patient.
- the clinical relevance (related to the above five considerations) that the dissimulation has or may have in the examinee's proposal of impairing problems.

Genuine medical and psychiatric illnesses have several features in common. Reported symptoms are typically not in excess of signs and measurable findings. Real illnesses demonstrate a regular consistency across an adequate database, including findings from different examiners, varied contexts (e.g., work, home, and recreation), and unmistakable presence when the patient is distracted. Disability specialists, for example, weigh these heavily in impairment determinations. Understandably, both forensic examiners and clinicians may be reluctant to give an opinion of malingering for various reasons, including litigation fear, concerns about errors, and the pejorative nature of the term. Partial malingering adds to the condition being far from clear-cut. Careful documentation of objective findings and facts in a nonpejorative manner is essential. A malingering diagnosis requires firm evidence.

COLLATERAL DIAGNOSTIC DATA

The finding of malingering may also be supported by getting information from collateral sources. These include data from interviews with people who know the patient, relevant records, psychological testing, and other care professionals. Informed consent issues need to be considered and meticulously documented. All authorizations need to be in writing and in the record. Interviews with family members and significant others, employers, past therapists, jail or prison personnel, and probation or parole officers may provide important information. Past medical records, psychiatric records, occupational and educational reports, court evaluations, and prior disability claims are also useful. There are no physical examination techniques that have been demonstrated objectively to prove the presence of malingering. However, internists and family practice physicians stress that warning flags include persistent noncompliance, striking inconsistency between physical findings and stated symptoms, and a referral from an attorney or an insurance company. Case formulations must include a culturally informed assessment. Although the psychiatrist has an essential clinical role in an overall assessment, the collaborative efforts of a team of medical, psychological (e.g., using neuropsychological testing), and legal specialists may be needed to provide corroborative evidence and hence a consensus to support psychiatric findings.

Psychological testing involves using standardized psychometric instruments by a trained and experienced clinical psychologist. These detection strategies can establish feigning but not a person's underlying motivations, which are necessary to support malingering. The following summarizes a sampling of tests. The application of these sophisticated symptom validity tests (SVTs) and performance validity tests (PVTs) is best managed by those highly trained in their use.

Such tests are often used to assess competency to stand trial, child custody evaluations, criminal responsibility, violence risk assessments, and personal injury litigation. Tests must show validity and reliability to be admissible in legal contexts and adequate sensitivity to identify true positives and sufficient specificity to identify true negatives. These federally recognized criteria are known as the *Daubert* standard and are derived from the 1993 U.S. Supreme Court case *Daubert v. Merrell Dow Pharmaceuticals, Inc*. Tests with these features used to detect feigning will be reviewed. It is acknowledged that no one gold standard is considered definitive—using multiple test measures and other collateral data aids in providing converging supportive evidence suggesting malingering.

The astute clinician always considers the test taker's appearance of task avoidance, low motivation, poor effort, even time of day, and hunger status in assessing the results of psychometric tests. For that matter, any test result may be wrongly skewed in the presence of actual pain, headache, lack of sleep, fatigue, and so forth. Without a subaverage cognitive ability and substantial neurocognitive impairment, the malingerer may produce deceitfully feigned reports using dishonest, invalid, and intentionally motivated low effort to supply data.

The *Minnesota Multiphasic Personality Inventory-2 (MMPI-2*) is one of the most commonly employed instruments used to support mental illness malingering. Some of its subscales are particularly useful. The *F scale* detects symptom exaggeration or severe psychological disturbance such as deliberately exaggerated psychopathology. It is considered the best available discriminator of honest versus malingering feigned symptom profiles. It is not sensitive, however, to detecting neurocognitive dysfunction. The *K scale* is deemed good at detecting *faking good* or minimizing what the respondent perceives as incriminatingly negative symptomatology. This response shows mildly negative symptomatology an ordinary person would openly admit to experiencing. The *F-K scale* indicates a tendency to exaggerate symptoms. The F-psychopathology scale detects truly rare symptoms, and the D scale and FBS scale expose erroneous stereotypes of illness. MMPI-2-RF (2008/2011) is the restructured version currently used.

The *Wechsler Intelligence Scales* are also used to determine strengths and weaknesses in intellectual functioning areas from verbal comprehension, perceptual reasoning, working memory, and processing speed perspectives. Gross abnormalities in intellectual levels, such as intellectual disability forms and striking discrepancies in subtest areas, can be detected. Wechsler Memory Scales are also frequently used. Results are supplementary to the other tests of malingering described in this section.

The *Structured Inventory of Reported Symptoms (SIRS*), a prevalent instrument in use, is used to test for feigned mental disorders and truly rare symptoms. Some consider it the best measure or gold standard of whether someone is faking psychotic symptoms. The two underlying dimensions in this test are the Spurious Presentation used to assess symptom constellations unlikely to be true under any circumstances and the Plausible Presentation used to evaluate some symptoms that genuine patients are likely to endorse. *SIRS-2* (2010) is currently used.

The *Personality Assessment Test (PAI)* is also used to test for truly rare symptoms, unlikely psychopathology patterns, and screening out honest responders. It can be completed in less than 1 hour. The *Validity Indicator Profile (VIP*) is a psychometric instrument using verbal and nonverbal items to test cognitive feigning; it takes less than 1 hour. The VIP contains very easy and difficult items. Feigned profiles show easy items missed and difficult items responded to at chance levels by malingerers. Nonmalingerers, for example, typically pass these items. It should be noted that such tests are not suitable for those with significant intellectual disabilities because such cognitively limited persons may not try as hard or may give up easily and produce what are recognized to be careless and invalid profiles.

The *Test of Memory Malingering (TOMM*) is a more sophisticated malingering test. Also widely used, it assesses exaggerated memory complaints and the dissimulation of cognitive impairment, and it detects subtle forms of malingering. Its classification is based on the “floor effect,” a testing strategy relying on the nonmalingering individual’s easily passing most test items. Studies of its validation indicate excellent specificity and modest sensitivity. It has a very high accuracy rate, and its validity has been supported in a forensic psychiatric population. The *Millon Clinical Multiaxial Inventory (MCMI-III*) is a malingering test detecting tendencies to overstate emotional and personal problems. The *Fifteen Item Test (FIT*) is a test of recall; it evaluates feigned neurocognitive deficits.

The *Miller Forensic Assessment of Symptoms Test (M-FAST)* uses a structured interview and is considered most efficient at distinguishing feigned from bona fide schizophrenia. This tool is a screening device with highly validated studies confirming it has sound psychometric properties. The majority of its items represent psychotic or highly unusual or improbable symptoms.

Some experts state the two primary stand-alone measures for feigned mental disorders are the Miller Forensic Assessment of Symptoms (M-FAST) and the Structured Interview of Reported Symptoms (SIRS-2).

Using a polygraph or lie detector, drug-assisted (e.g., amobarbital) interviews, or hypnosis are not considered standardized in an adequately scientific fashion and are not ordinarily used to detect malingering. Ultimately, the only acknowledged gold standard undisputed in confirming malingering is full disclosure by the subject’s stating a deliberate intention to malinger.

DIFFERENTIAL DIAGNOSES

The malingering condition may coexist with actual mental disorders, such as major depressive disorder, anxiety disorders, bipolar disorders, and personality disorders. Careful assessment must discriminate genuine mental disorders and personality disorders from the pretense of malingering. Presentations are usually complicated and typically not straightforward. Often, more than one diagnosis and condition may exist together. Also, one must screen for intact cognitive status and rule out intellectual and developmental disabilities and dementias. These may inadvertently cause the patient to incorrectly describe what seems to be symptoms. These misidentifications are not deliberate deception—*feigning*. Such mistaken beliefs do not carry the motivation of intent to deceive and must be distinguished from malingering.

Malingering (not a mental disorder but a condition that may be a focus of clinical attention) has been contrasted with a group of mental disorders: “somatic symptom and related mental disorders,” notably somatic symptom disorder, conversion disorder (i.e., DSM-5, F44. functional neurologic symptom disorder), illness anxiety disorder, and factitious disorder.

DSM-5 has a new category, *somatic symptom, and related disorders*, that includes these and other related disorders, and states their initial presentation occurs mainly in medical rather than psychiatric settings. Table 28.1–5 summarizes some aspects of the differentiation of malingering from factitious and somatic symptom disorders.

Factitious disorders imposed on self are characterized by physical, psychological, or mixed symptoms plus signs of illness intentionally produced or aggravated. Symptoms are not genuine but manufactured. The patient willingly agrees to any testing suggested. This disorder may or may not be associated with an unconscious psychological need to assume the sick role; yet, it is an apparent wish to receive medical care. Some have used the phrase: “the patient with factitious disorder looks for sympathy.”

Emphasis on this diagnosis is on the objective identification of falsification rather than inference about intent, reasons, or possible underlying motivation. This diagnosis always implies psychopathology. Its prevalence is uncertain but said to be about 1%. Individuals with factitious disorder dramatically present their history but tend to be vague regarding details. Pathologic lying (i.e., “pseudologia fantastica”) may be present. If frustrated by caregivers, these individuals may become uncooperative or abusive. Intentionally produced symptoms may include depression, amnesia, PTSD caricatures, infections, pain, fevers, or neurologic symptoms. Pain and requests for analgesic medications are widespread. The most severe and chronic form of factitious disorder, primarily with physical signs and symptoms, was referred to as *Munchausen syndrome*, which includes recurrent hospitalizations, peregrination (i.e., traveling), and pseudologia fantastica. “Factitious disorder imposed on another” (previously termed *Munchausen by proxy*) refers to a form of maltreatment in which an individual in the parental or caregiving role fabricates or produces physical, psychological, or combined illness in another to assume the vicarious sick role. Other actual mental disorders and borderline personality disorder may also accompany this.

Table 28.1–5.
Differentiating Malingering from Factitious and Somatic Symptom Disorders

	Malingering	**Factitious**	**Somatic Symptom Disorders**
Presentation	Psychological/physical symptoms	Psychological/physical signs and symptoms	Physical symptoms
Deliberate feigning	Yes	Yes	No
Inferred motivation	Conscious desire for unmerited tangible gain	Sick role; to receive medical care	Multiple determinations

Malingering differs from factitious disorder because the malingerer’s intentionally manufactured symptom presentation is secondary to a clear-cut conscious motivation to attain a recognizable external reward, payment, or prize. The malingerer seeks a tangible external incentive, which is usually material, not an abstract satisfaction. The patient with factitious disorder, unlike the malingerer, can be shown to create an identifiable falsification of being ill, impaired, or injured in the absence of obvious external rewards. Dangerous, self-injurious, and self-handicapping signs often accompany symptoms. Malingering and factitious disorders typically are not to be diagnosed together. In psychodynamic terms, the inference is that the intentionally created signs and symptoms of any problem produce a *secondary gain*, a recognizable advantage, for example, being sick and needing some form of help. This idea has been challenged because it is not precise and its meaning confusing; therefore, “gain” concepts are no longer commonly used.

Somatic symptom disorders category (e.g., conversion disorder, somatic symptom disorder, and illness anxiety disorder) are characterized by real physical symptoms not intentionally produced and are not under voluntary control. These disorders are typically first seen in medical settings such as primary care and neurology rather than psychiatry.

In *conversion disorder*, also termed *functional neurologic symptom disorder*, somatic symptoms are present as the *absence* of a function. These real symptoms are clearly incompatible with neurologic pathophysiology. In contrast to somatic symptom disorder, the patient does not express excessive thoughts, feelings, and behaviors about the disorder; the focus is typically on presenting a symptom of the loss of a function. In conversion disorder (e.g., blindness and paralysis), patients are cooperative, welcome evaluations, search for answers to their symptoms, yet present with historical gaps and vagaries. Some observers have commented on their “la belle indifference” attitude of almost bland indifference. Medically unexplained symptoms remain a key feature in conversion disorder. Neurologic symptoms that are found after appropriate neurologic assessment are incompatible with neurologic pathophysiology.

In *somatic symptom disorders*, complaints suggest a medical condition. The patient genuinely experiences the symptoms. Several real somatic or physical symptoms exist that are *positive*, for example, pain. Thus, positive symptoms, not the absence of a function such as sight or motor ability, are presented. Still, they cannot be fully explained by a general medical condition, by the direct effects of a drug substance, or by another mental disorder. One or more genuine somatic symptoms are significantly moderate to severe, distressing, and disabling. Thoughts, feelings, and behaviors about the seriousness of these symptoms are disproportionate and disrupt daily functioning. Most of the somatic symptoms in somatic symptom disorders cannot be demonstrated to be clearly incompatible with pathophysiology (e.g., pain and fatigue). The patient’s preoccupation is distress at having a serious disease based on the misinterpretation of bodily symptoms. About 75% of these patients fall into the category of “hypochondriasis.” The patient *focuses primarily on the distress about having the symptoms and not on the symptoms, usually pain; has a false belief they have a serious disease that has not been detected; this preoccupation impairs their ability to function in important areas of life.*

In contrast to the relatively more euthymic patients with conversion disorder, patients with somatic symptom disorder present as chronic worriers. Some have speculated that psychological factors, inferred to be dynamically unconscious, have an essential role in these physical symptoms. In this mental disorder, the patient believes their symptoms are real, and distress is based on having a serious disease. Approximately 75% of individuals previously diagnosed with hypochondriasis are subsumed under the diagnosis of somatic symptom disorder.

About 20% to 25% of those previously holding the DSM-IV term “hypochondriasis” may now be subsumed under *Illness Anxiety Disorder* (F45.21) in DSM-5. This disorder is seen in medical rather than mental health care settings. Characteristics are a *preoccupation with having or acquiring an as yet undiagnosed serious illness, that is, concerns about getting sick, not symptoms*; *somatic symptoms are not present* or, if present, few, mild, and not severe. There is a *high level of health anxiety*. With no or mild symptoms present, fear about having or acquiring a serious disease dominates thinking for at least 6 months.

Behaviors of avoidance or repeated body checks occur. A range from care-seeking to care-avoidance follows. All these patients misinterpret their bodily sensations. Another mental health disorder does not better explain illness preoccupation. Excessive health-related behaviors exist. The complaints are more than what one would expect from history, physical examination, or laboratory findings. There may be an absence of objective conclusions fully to explain the clinical presentation. Other mental disorders and genuine medical disorders such as depressive and anxiety disorders may coexist with somatic symptom disorders. Illness anxiety disorder (i.e., excessive worry about developing an illness) and somatic symptom disorder (i.e., excessive worry about the seriousness of one's physical symptoms) lie on a quantitative continuum, with the latter being more severe.

Since COVID-19 is now a universal part of our shared experience, issues related to it are plentiful. There are tests available detecting the presence or absence of COVID-19, even though patients may need repeat testing. Although respiratory symptoms may be present, numerous other symptoms and signs (e.g., fever, muscle pain, GI distress) may be found. The precautions, quarantine of those exposed, and isolation of those infected bring bouts of loneliness to the forefront. As with other concerns over illnesses, if the patient can show reasonable insight into their anxiety, this suggests a genuine concern. Insight into one's anxiety and refraining from extreme and unnecessary behaviors may obviate adding a tentative diagnosis of Illness Anxiety Disorder. Thus, differentiating real illness from its absence requires contextualizing individual patients and their realistic or overly extreme, preoccupying worries.

Malingering differs from somatic symptom disorders because the malingerer's intentionally contrived presentation of physical and psychological symptoms is under voluntary control and has an external incentive. In malingering, the patient does not genuinely experience the symptoms presented. The malingerer avoids agreeing to uncomfortable testing. If physical symptoms are present in somatic symptom disorders, they are not intentionally produced and are not under the patient's voluntary control. Conversion disorder within the somatic symptom disorder category is susceptible to symptom relief by suggestion and hypnosis, unlike malingering. Somatic symptom disorder may also exist with conversion disorder. If factitious disorder (does not seek obvious external reward) is diagnosed, malingering (does seek obvious external reward) ordinarily is not added as an additional condition. However, the presence of some factitious or malingered symptoms, mixed with other symptoms that are clearly nonintentional, may be found. When this mixed situation appears present, both somatic symptom disorder (referring to the nonintentional symptoms) and factitious disorder (intentional symptoms) or malingering (intentional symptoms) may be acknowledged.

COURSE AND PROGNOSIS

Malingering, when deemed present, needs to be assessed in the overall biopsychosocial context of the person’s life. The presence of diagnosable mental disorders, their histories, and their response to psychotherapy and medications must be addressed. Any acute or chronic medical or surgical problems and their effect on the patient’s overall functioning need to be considered. In forensic, hospital-related settings or military cases, *malingering may be context-specific and change its presentation or even disappear if circumstances and situational challenges change*. Because malingerers do not usually comply with treatment recommendations, the malingering status often remains unaffected. Research in this area is limited but shows that if reinforced, malingering may continue; if malingered goals are met, symptoms sometimes abate; if the malingerer becomes frustrated with current care, other providers are sought.

TREATMENT AND MANAGEMENT

Since malingering is regarded as a condition that may focus on clinical attention, no routine or standardized interventions have been devised and recommended. Collaboration among evaluators is useful and significant. A range of suitable management approaches having clarification as a primary goal includes the following:

1. A clear diagnostic delineation of any medical or psychiatric symptoms and disorders, if present, in addition to malingering.
2. A mindful consideration of legal and ethical issues, especially those reflected in the medical record, because malingering is unlikely to be proved conclusively.
3. If the psychiatrist is the treater or leading evaluator, a nonthreatening approach should include concerned neutrality and the avoidance of any abrupt confrontation or accusations of lying. Safety considerations should be present because angry reactions might escalate to dangerous behaviors. Explicit recognition of issues of implicit countertransference reactions (e.g., an unreasonable or exaggerated sense of annoyance at the patient) and an awareness of the possible subtle effects of *projective identification,* mentioned earlier, and any implicit biases on the part of all parties can minimize inordinate distortions and inadvertent psychological boundary violations.
4. The clinician might discuss that all findings indicate no medical basis for the patient’s symptoms can be found. Hence, no medical or psychiatric diagnoses can be established. Suggesting that the patient’s history appears incomplete or something significant has been left out may offer the patient the last opportunity to reveal the illness deception or devise an explanation now to save face. With tactful clarification, alternative strategies that are healthier

might emerge. Parsimonious, sensitive feedback might encourage the patient to shed further light on their intent or motivations. However, in some cases, the patient may become angry, respond defensively, and decide to leave, perhaps to go elsewhere. Rarely, a patient will openly admit malingering.

5. If the psychiatrist is a consultant, suggested recommendations and management strategies can be provided directly to the referring party for implementation.

FUTURE DIRECTIONS

The introduction of malingering–deliberate symptom falsification to obtain an external incentive or reward–into the DSM-III in 1980 caused many nonmedical specialists' vociferous reactions. This vigorous effort questioned the validity of this attempt to medicalize deceptive illness behavior. What had seemingly been an orphan condition was and is continuing to be the subject of intense debate. The medical model contrasts with an alternate view emphasizing the person having agency, free will, choice, and the ability to make consciously planned decisions. From this latter perspective, malingering as illness deception is said to be better regarded within the context of individual responsibility, accountability, conscious choice, and social action. According to such a framework, the emphasis on malingering should eschew an uncritical adoption of the medical model and instead consider legal issues and questions having to do with problems of intentional social deviance. Those who advocate this position state that patterns of misconduct that are, in fact, "bad behavior" are incorrectly being redefined as bona fide illnesses.

The psychiatrist views malingering as not a mental disorder but as a behavioral strategy that healthy or sick people might use in certain situations to attain an unmerited goal—using deliberate pretense. Persons with this condition may deem the behavior morally right or perhaps wrong but justified under their current circumstances. In contrast, factitious disorder is maladaptive because the induction of injury or disease is a central feature, making it a DSM-5 disorder rather than a condition. These considerations beg the questions:

- "What is injury, and who is harmed"?
- "Can we, as physicians, accurately document deception and motivation as reasons for the choice, let alone the degree"?
- "In what way is the suspected deception consequential"?

Many have debated the extent to which malingering has clear-cut legal implications, such as by equating it with criminal fraud. However, it can be argued that a medical expert's function is not principally to make a judgment on an individual's honesty. The medical expert's role is to assess a clinical presentation, evaluate reported symptoms, render diagnoses, and points of clinical attention based on the best evidence possible. A treatment or management plan follows. This approach to malingering clearly gives the physician a medical rather than a legal or moral role.

The effort to introduce more objectivity into detecting malingering has produced sophisticated test measures from the medical field. Detecting deception has shown some success by mapping brain correlates of lying. Functional magnetic resonance imaging (fMRI) has demonstrated specific neurocircuitry in brain regions under conditions of simulated, mock crime such as stealing. For example, truthful, as opposed to deceptive responses, were correctly identified in 90% of the model testing group. Similar brain imaging studies have shown different cerebral activation processes are involved when a symptom is subjectively experienced as real compared to when it is feigned. Such objective data are intriguing but far from conclusive.

The entire area of malingering and assessing symptom validity with all it implies is complex. Partial malingering may be common and can add parameters to indeterminate conclusions. Obvious psychological, psychiatric, monetary, social, legal, moral, and philosophical issues, to mention a few, are at play. Although it is laudatory to understand malingering in an objectively accurate manner, as many in specialized fields are trying to do, the explosion of research, conjecture, simulation studies, and psychological testing has raised serious ethical questions.

At this time, no generally agreed-on consensus exists. With its relatively straightforward and presumed nonjudgmental guidelines, the psychiatric approach values a dedicated capacity for examining multiple sources of information, often ambiguous and indeterminate. Therefore, psychiatry striving for a nosologic reserve participates as a rational, even essential, player as one team member in the field.

FURTHER READINGS

Carone DA, Bush SS. *Mild Traumatic Brain Injury: System Validity Assessment and Malingering*. Springer; 2013.

Crossman AM, Lewis M. Adult's ability to detect children's lying. *Behav Sci Law*. 2006;24:703–715.

Daubert V. Merrell Dow Pharmaceuticals, Inc. 113 S.Ct. 2786;1993.

Hall HV, Poirier JG. *Detecting Malingering and Deception*. CRC Press; 2020.

Halligan PW, Bass C, Oakley DA, eds. *Malingering and Illness Deception*. Oxford University Press; 2003.

Iverson GL. Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Appl Neuropsychol*. 2006;13(2):77–90.

Kaseda ET, Levine AJ. Post-traumatic stress disorder: a differential diagnostic consideration for COVID-19 survivors. *Clin Neuropsychol*. 2020;34(7–8):1498–1514.

Klein M. Envy and gratitude [1957]. In: Klein M, ed. *Envy and Gratitude and Other Works*. Hogarth Press; 1975:176–235.

Kozel A, Johnson KA, Mu Q, et al. Detecting deception using functional magnetic resonance imaging. *Biol Psychiatry*. 2005;58(8):605–613.

Kucharski TL, Duncan S, Egan SS, Falkenbach DM. Psychopathy and malingering of psychiatric disorder in criminal defendants. *Behav Sci Law*. 2006;24:633–644.

Libow JA. Child and adolescent illness falsification. *Pediatrics.* 2000;105(2):336–342.

Malingering YG. *Feigning, and Response Bias in Psychiatric/Psychological Injury: Implications for Practice and Court*. Springer; 2014.

McCann JT. *The Significance of Malingering and Deception in Adolescents: Assessing Credibility in Clinical and Forensic Settings*. American Psychological Association; 1998.

McDermott BE. Psychological testing and the assessment of malingering. *Psychiatr Clin North Am*. 2012;35(4):855–876.

Merten T, Merckelback H. Factitious disorder, and malingering. In: Geddes JR, Andreasen NC, Goodwin G, eds. *New Oxford Textbook of Psychiatry*. Oxford University Press; 2020:1342–1350.

Mittenberg W, Aguila-Puentes G, Patton C, et al. Neuropsychological profiling of symptom exaggeration and malingering. *J Forens Neuropsychol*. 2002;3(1–2):227–240.

Mittenberg W, Patton C, Canyock EM, Condit DC. Base rates of malingering and symptom exaggeration. *J Clin Exp Neuropsychol*. 2002;24(8):1094–1102.

Ninivaggi FJ. *Biomental Child Developmental: Perspectives on Psychology and Parenting*. Rowman & Littlefield Press; 2013.

Ninivaggi FJ. *Making Sense of Emotion: Innovating Emotional Intelligence*. Rowman & Littlefield Press; 2017.

Nock MK, Stein MB, Heeringa SG, et al. Prevalence and correlates of suicidal behavior among soldiers: results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*. 2014;71(5):514–522.

Norfolk PA, Floyd RG. Detecting parental deception using a behavior using scale during assessment of attention-deficit/hyperactivity disorder: an experimental study. *Psychol Sch*. 2016;53(2):156–172.

Rios J, Morey LC. Detecting feigned ADHD in later adolescence: an examination of three PAI-A negative distortion indicators. *J Pers Assess*. 2013;95(6):594–599.

Rogers R, Bender SD. *Clinical Assessment of Malingering and Deception*. Guilford Press; 2020.

Rogers R, Granacher RP. Conceptualization and assessment of malingering. In: Drogin EY, Dattilio FM, Sadoff RL, Gutheil TG, eds. *Handbook of Forensic Assessment: Psychological and Psychiatric Perspectives*. Wiley; 2011:659–678.

Slick DJ, Sherman EMS. Differential diagnosis of malingering. In: Carone DA, Bush SS, eds. *Mild Traumatic Brain Injury: System Validity Assessment and Malingering*. Springer; 2013:57–72.

Slick DJ, Sherman EMS, Iverson GL. Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *Clin Neuropsychol*. 1999;13:545–561.

Sollman MJ, Ranseen JD, Berr DT. Detection of feigned ADHD in college students. *Psychol Ass*. 2010;22(2):325–335.

Stern TA. *Learning About Factitious Illness and Malingering: A Programmed Text*. MGH Psychiatry Academy; 2020.

Stout RG, Farooque RS. Claims of amnesia for criminal offenses: psychopathology, substance abuse, and malingering. *J Forensic Sci*. 2008;53(5):1218–1222.

Suhr J, Hammers D, Dobbins-Buckland K, Zimak E, Hughes C. The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. *Arch Clin Neuropsychol*. 2008;23(5):521–530.

Young S, Jacobson R, Einzig S, Gray K, Gudjohnson HG. Can we recognize malingerers? The association between malingering, personality traits, and clinical impression among complainants in civil compensation cases. *Pers Individ Differ*. 2016;98:235–238.

28.2 Adult Antisocial Behavior, Criminality, and Violence

DOROTHY OTNOW LEWIS, M.D., F.A.C.P., HARRELL WOODSON, PH.D., AND ROBERT J. BOLAND, M.D.

The acts subsumed under the terms “adult antisocial behavior, criminality, and violence” range from stealing candy from a baby to abducting, raping, and murdering a child. They cover a multitude of sins and a multitude of neuropsychiatric signs, symptoms, and behaviors. Despite the diverse aberrant behaviors included within these categories, official diagnostic manuals consider “adult antisocial behavior” to be a condition “not due to a mental disorder.” For decades, criminals, violent and nonviolent, were deemed “bad.” Those committing particularly heinous crimes were “evil.” They posed a religious, not a medical question. There was nothing wrong with their brains. Professional thieves, racketeers, dealers in illegal substances, thugs, and even serial killers were regarded simply as mentally healthy malefactors. They were immoral, and their acts presumably represented their conscious choices of alternative lifestyles. Such assumptions presupposed the existence of supportive scientific evidence. Alas, there was none.

People tend to divide crime or antisocial behavior into two general categories: street crime and white-collar crime. But what about organized crime?

ORGANIZED CRIME

In the United States, organized crime, as defined by the Organized Crime Control Act of 1970, is “the unlawful activities of . . . a highly organized disciplined association.” We call this structured criminal activity “racketeering.””

Until 1957, J. Edgar Hoover denied or ignored this violent if lucrative entity. Instead, Hoover focused his attention and workforce on uncovering alleged Communists at all levels of government, while the Mob (otherwise known as the Mafia or Cosa Nostra) flourished right under his nose. Prior to 1957, Hoover dismissed The Mob as hoodlums to be dealt with by local police. However, after a series of turf wars and murders, mob bosses from across the United States gathered for a meeting. This event was in the fall of 1957 at the palatial 58-acre Apalachin estate of Joe Barbara. A group of snazzy cars parked around the Barbara mansion caught the eye of the local police, who gave chase. The police caught several mobsters as they tried to run away through cornfields. Headlines the next day shouted, “Police Nab 67 Mafia Chiefs in Backwoods.” However, the courts convicted only 20 on minor charges, and the appeals court reversed all convictions. Such was the power of this theoretically nonexistent entity, The Mafia.

Although there are many books about "The Mob," its structure, rules, and alleged rituals, there is little, if any, formal research on the psychology of mobsters. In 2004, two Italian psychologists attempted a clinical description of the workings of the Mafia mind. Unfortunately, the vague, undocumented nature of their database made it impossible to generalize from their publication. Nonetheless, their description of the Mafia family structure was consistent with that of other Mafia observers. There is a consensus that at its height, many young men who joined The Mafia (or Cosa Nostra) were not joining in desperation (like their impoverished immigrant peers) but with admiration. Kingpins commanded respect. The toughest, strongest, most self-assured male children were, reportedly, groomed to take over positions of authority previously held by their fathers, uncles, or other powerful male relatives.

Absent vigorous data, Mafia psychology, is open to speculation. Some early criminologists asserted that these criminals suffered from a lack of self-control and sought only to satisfy their immediate needs or wishes. Others hypothesized the existence of a criminal personality characterized by risk-taking, unreliability, irresponsibility, and inability to tell the truth.

One systematic study compared Mafia-linked incarcerated criminals with incarcerated non-Mafiosi. The researchers found, to their surprise, that none of the Mafia-linked criminals scored in an abnormal range on a psychopathy scale, whereas "ordinary criminals" scored much higher on average. These results suggest that being a run-of-the-mill criminal is quite different from being a "made man." A fascinating finding was that Mafiosi's scores in interpersonal domains suggested that they were more caring and empathic than their everyday criminal jail mates. Many had stable marriages as well as mistresses. One expert on the Mafia concluded that the personality traits that predict leadership success (i.e., extroversion, controlled impulsiveness, a sense of adventure, megalomania, and narcissistic personality disorder) are equally suitable traits for a successful career in either business or organized crime. If the scales and methods were valid, we might conclude that the portrayal of mobsters in such shows as *The Sopranos* is surprisingly accurate.

WHITE-COLLAR CRIME

The line between legitimate business practices and racketeering often blurs and at times totally disappears. We merely have to scan through current news to marvel at the nature and scope of skullduggery perpetrated by heretofore upstanding businessmen who seem to be in their right minds. There are, it appears, hidden cracks in many pillars of society. It cannot be said with conviction that, as a group, racketeers, embezzlers, and money launderers are free of psychopathology any more than we can say the same for doctors, lawyers, stockbrokers, accountants, senators, and even presidents. We know much less about the neuropsychiatric characteristics and brain function of so-called white-collar criminals than we do about violent criminals.

Edwin Sutherland coined the term *white-collar crime* at a 1939 meeting of the American Sociological Society. Sutherland took issue with his colleagues, who theorized that crime primarily resulted from poverty and socioeconomic inequality. Sutherland called attention to the improper business practices of individuals in positions of power, practices he felt should be regarded as criminal offenses rather than simply as civil wrongs.

The Federal Bureau of Investigation (FBI) has defined *white-collar crime* as illegal acts characterized by deceit, concealment, or violation of trust that are not dependent on the application or threat of physical force or violence. These crimes thus are treated as conscious, rational, and economic decisions.

Estimates of white-collar crime’s cost range in the hundreds of billions of dollars per year, whereas street crime costs about 15 billion dollars per year. The human consequences of embezzlement and fraud are immeasurable and include losing one’s home, pension, tuition, or physical and mental health.

To understand the increasing incidence and prevalence of financial crime, one must study the enormous changes in the technology used to manipulate economic transactions since the 1990s. For example, with very little oversight, corporate officers in publicly held companies with the stroke of a key can artificially inflate the price of a stock. New technologies have facilitated such practices at the expense of customers and employees. The more automated the victims (such as the digits of their retirement funds), the easier it is to defraud them. The farther away one is from the target, the easier it is to destroy it. It is easier to kill with drones than with knives. Also, the less human the target, the less troubled the conscience. Wall Street moguls, who would never consider holding up their clients at gunpoint, think nothing of tapping a key and transferring their clients’ life savings into questionable, if not downright illegal, investment schemes.

Such moral compromises are not limited to the financial world. Most troubling is the degree to which the bottom-line economy has corrupted the medical industry.

An especially egregious example of corporate and medical malpractice was the marketing and dissemination of the Dalkon Shield, an intrauterine contraceptive device. Despite early warnings regarding the risky design of the product (it had a multifilament string with a tendency to wick and carry bacteria-laden fluid from the vagina into the otherwise sterile uterus), corporate officials went forward with promotion, production, and sales.

Shortly after that, the company began to receive complaints regarding cases of pelvic inflammatory disease and septic abortions. Nevertheless, it continued to market the device for over 3 years. By the time the corporation filed for bankruptcy, there were more than 300,000 lawsuits.

According to the National Health Care Anti-Fraud Association (NHCAA), a small percentage of health care providers commit most healthcare fraud. However, the NHCAA estimates that healthcare fraud's financial losses are in the tens of billions of dollars.

Epidemiology

As violent crime rates diminished during the first decade of the 21st century, government and private lawsuits for fraud and other financial violations soared. Scholars and law enforcement personnel have speculated that the surge in high-stake white-collar crime is a result, at least in part, of technologic advances such as the Internet, automated tellers, and plastic that facilitate certain kinds of offenses (e.g., hacking, identity theft, and embezzlement) and provide opportunities for new creative kinds of white-collar crime. The scope of white-collar crime is ever-widening. It has come to include criminal acts in violation of environmental and health laws. The Department of Justice has finally begun to focus on financial institution fraud, defense procurement fraud, health care fraud, computer crimes, telemarketing fraud, and international fraud.

Healthcare fraud is soaring as large healthcare programs, public and private, steal trillions of dollars. According to the Government Accounting Office (GAO), in 2011, Medicare and Medicaid paid a whopping $65 billion in improper payments. According to the FBI, one of the most significant trends in healthcare fraud is the participation of health care professionals—in other words, doctors—who in so doing risk their patients' lives, not to mention their patients' pocketbooks. The FBI has reported an upsurge in unnecessary treatments, including surgeries, the prescribing of unnecessary, potentially dangerous drugs, and overall substandard care.

Of particular concern to psychiatry are cases in which psychiatric/psychological professionals have engaged in fraudulent practices. For example, individual practitioners and hospital chains have pled guilty to providing unnecessary treatment to thousands of patients to collect insurance money.

When insurance reimbursements depend in large part on procedures performed, psychiatrists are at a disadvantage. Talk is cheap. There are a few procedures. At the same time, thanks to technologic advances, new procedural treatment modalities are now available (e.g., transcranial magnetic stimulation and deep brain stimulation). Electroconvulsive therapy, once rare, has experienced a rebirth.

One of the most devastating instances of white-collar crime in the United States involving the medical profession is the case involving Purdue Pharma, producer of OxyContin, an opioid touted by Purdue as being less addictive than other pain medications because of its time-release formula. Before marketing, internal documents circulated within the company indicating the product's potential for abuse. Nonetheless, sales representatives, reportedly under management's instructions, falsely assured doctors of the safety of their product. The U.S. Food and Drug Administration (FDA) approved the drug in 1995, and by the year 2000, parts of the United States, mainly rural areas, experienced soaring rates of addiction and crimes related to addiction. Although the pill as marketed was long-acting, the addict, by simply crushing the pill and swallowing the powder or diluting and injecting it, could experience an immediate, intense "high." At a news conference in 2007, a United States attorney stated that OxyContin's aggressive, untruthful marketing had resulted in rising crime rates, teenage drug addiction, and even deaths. This particular fraud was not limited to the pharmaceutical industry. Doctors reportedly were paid handsomely to promote the drug to colleagues.

White-Collar Violence

The more closely we examine corporate crime and criminals, the more we see that white-collar criminals cause tremendous physical damage. A graphic example is the production and sale of the Ford Pinto long after the corporation knew of the flaw in its design that caused it to explode in collision. The "bean-counters" at Ford concluded that the cost of lawsuits from injuries and deaths would be less than the cost of a large-scale recall, so Ford stalled the recall.

Another clear example of corporate violence, of obliviousness to human life, is the coal mining industry. In April 2010, an explosion occurred at the Upper Big Branch Mine in West Virginia, killing 29 of the 31 miners working there that day. The mine owners had previously received dozens of notifications of safety violations and had not rectified them. In its investigation, the West Virginia Office of Miners Health, Safety and Training (WVOMHS and T) issued 253 violations of safety standards.

According to the U.S. Bureau of Labor Statistics (BLS), there were 5,333 fatal work injuries in 2019, a 2% increase from 2018. That amounted to 3.5 fatal injuries per 100,000 full-time workers. This percentage is only moderately lower than the 5.3 deliberate homicides per 100,000 population in 2017. Thus, corporate malfeasance has a mortality rate similar to that of deliberate homicide.

Etiology

We do not know what causes white-collar crime. The field is wide open to speculation. Sutherland believes these are learned behaviors resulting from an association with those who approve of them and isolation from disapprovers. Others have argued that white-collar crime, like street crime, results from a tendency of individuals to seek immediate, short-term gratification without considering the possible long-term consequences of their acts. Such individuals, they believe, are impulsive, aggressive, and motivated by the pursuit of pleasure and avoidance of pain. Comparing street criminals with other kinds of white-collar offenders reveals marked demographic differences (poor vs. rich). One magnetic resonance imaging (MRI) study reported hypertrophy of parts of the frontal cortex in white-collar criminals and suggested a relationship to “abstract monetary rewards.” That is a leap! Although some have asserted that individual tendencies lead to all crime, others have argued that perceived incentives and disincentives like those described earlier are potent influences. Others have pointed to ethical deficits. For example, experiments on college students suggest that anyone will cheat when given a chance.

Why is white-collar crime so rampant? In his examination of this phenomenon, Callahan called attention to the trend in the United States toward an increasingly free-market, laissez-faire society with a focus on money. He posited that the bottom-line ethic (if it can be called that) had infiltrated areas of culture previously insulated from market pressures, such as medicine, education, journalism, and sports. In addition, he hypothesized that the greater inequality in America between the haves and the have-nots had weakened the social fabric—undermining the notion that everyone is in it together and bound by the same rules. Finally, Callahan suggests four critical reasons for the increased levels of cheating:

1. Increasing pressures to succeed in a society where we cannot count on job security.
2. Weakened safeguards against cheating. The FBI, Securities and Exchange Commission (SEC), and Justice Department currently have insufficient resources to stop white-collar crime.
3. Larger rewards for those who cheat, from the CEO who inflates company earnings to the baseball player who bulks up on steroids to maintain his million-dollar-plus annual income.
4. “Trickle-down corruption” in a society where the ordinary middle-class person perceives that the cards are stacked, and the rules are unfair.

Another factor may be the ever-increasing distance between perpetrators and victims. It is easier to harm someone on the internet than it is your neighbor.

MORALITY AND THE BRAIN

For millennia morality has been assumed to be an acquired capacity, developed with maturation and upbringing. Most religions include specific moral concepts, beliefs, and standards inculcated during childhood and encouraged throughout adolescence and adulthood. Pillars of the church and pillars of the community are considered fabricated of the same ethical stuff. Morality is, in reality, among the most sophisticated features of human judgment, behavior, social cognition, decision-making, and conflict resolution. It includes a theory of mind, meaning our ability to imagine another person's complex emotions and mental states. Scholars debate the relative biologic inheritance of this capacity versus environmental influences on it. It may be that infants are born with the innate capacity for implicit morality, whereas explicit morality requires ongoing instruction. Infants will cry when they hear other babies crying, but they cannot tell you why. This is the beginning of empathy.

In the 1990s and early 21st century, Damasio established an association between frontal lobe damage and immoral dyscontrolled behavior. Many studies have attempted to localize the areas in the brain and the pathways among them that facilitate moral thought and behavior. Greene and colleagues were the first to use functional magnetic resonance imaging (fMRI) to study moral judgments. In 2001 they reported a study that presented several different scenarios involving moral and neutral choices while subjects underwent fMRI. One example scenario was a variant of the trolley problem, developed by the philosopher Judith Jarvis Thomson in 1985 and familiar to ethicists and moral psychologists. In it, a person imagines seeing a runaway train and either doing nothing as it runs over five track workers or pushing a large man in front of the train, stopping it but killing that man. The dilemma, in part, is a challenge to utilitarianism: whether the end justifies the means.

Investigators found that certain brain areas were especially active when contemplating moral dilemmas such as the one above. These areas include the medial frontal gyrus, the posterior cingulate gyrus, and the angular gyrus. Some of these areas are more active in emotional processing than cognitive reasoning, and they were more active when subjects contemplated the "up close and personal" choices than when considering impersonal moral and nonmoral choices. Decety and Porges modified Greene's study by showing subjects scenarios in which a person was harming or helping another person. They instructed the subjects to identify themselves with the active (harming or helping) protagonist. Helping activity was associated with increased activity in the ventral striatum, while harmful activity resulted in a "strong signal decrease" in the ventromedial prefrontal cortex (VPFC) and activation in the amygdala. Thus, Decety and Porges demonstrated that the subject's perspective profoundly affects the recruitment of different brain areas during moral considerations. Investigators have found that damage to the prefrontal cortex (PFC) interferes with the aversive response to harmful actions. They have argued that what some would call psychopathic behaviors result from damage to the PFC and an inability to develop an aversion to immoral behaviors.

Does it matter when in life damage to the PFC occurs? It does. Taber-Thomas and colleagues compared the moral functioning of individuals who experienced VPFO damage in adulthood with those who had similar injuries in childhood. They found that when they confronted adults with early childhood damage to the VPFO with self-interest decisions (e.g., whether or not to kill your boss for money), the subjects were less likely to condemn the act than were adults who had received similar results of damage to their brains during adulthood. The researchers suggest that the subjects with late-onset injury had, over time, developed affective responses that prohibited self-interested behaviors. What have we learned? Learning morality starts early. Absent early learning of morality, we may never be able to catch up.

Morality involves numerous other areas of the brain interconnected and connected to the PFC. These areas include the cingulate gyrus, the amygdala, the dorsolateral PFC, the temporal lobes, and even the cerebellum. Studies of the timing of reactivity from one area to the next exist. There appear to be different aspects of morality. This brief overview should tantalize the reader into exploring the findings, agreements, and disagreements, which swirl around the fascinating exploration of the neurologic foundations of morality.

VIOLENCE

What kind of animals are humans? Was Hobbes right? Is the natural state one of "continual fear, and danger of violent death"? Without a strong government or set of rules, "is the life of man, solitary, poor, nasty, brutish, and short"? Was Samuel Butler wise to caution, "Then spare the Rod and spill (sic) the Child"? Bad advice, at least according to current scientific data. Corporal punishment correlates with subsequent violence. No matter. Butler based his belief on the Biblical "He that spareth his rod hateth his son: but he that loveth him chasteneth him betimes." This instance is not the only time in this 21st century that science and God have found themselves at odds.

What about Wordsworth? Was he on to something when in *Intimations of Immortality,* he proclaimed that the human entrance into this world is one of purity and innocence? Humans come "trailing clouds of glory" only to be corrupted by the pressures of society, of living alongside other people. Rousseau agreed.

Locke disagreed. According to Locke, each person enters this world a blank slate to be scrawled upon by their environments and experiences.

Are all men (and women) created equal, or are some innately more aggressive than others? For millennia these kinds of questions were the provenance of religion, philosophy, literature, and politics. Now, these eternal questions have fallen into the lap of science. Think of current technologies, statistically validated neuropsychological test materials, positron emission tomography (PET) scans, fMRIs, the newly devised ability to chop up yards and yards of DNA, and discover what humans are made of—will these developing technologies lead to scientific answers based more on fact and less on faith? We have made a start. We already know that each person enters the world with certain innate but not inevitable potentials for behavior. What happens to each person makes the difference. No single factor or known constellation of factors accounts for violence. Instead, as will become apparent, various intrinsic and environmental influences come together, interact, and create matrices for violence.

Epidemiology

Just over a decade ago, Nelson Mandela stated, “The Twentieth Century will be remembered as a century marked by violence. It burdens us with its legacy of mass destruction, of violence inflicted on a scale never seen and never possible before in human history.” He blamed this extraordinary violence on “new technologies in the service of ideologies of hate.” Had he lived longer, he would be dismayed, although likely not surprised at the start we have made in the 21st century.

As of 2021, there are many ongoing conflicts in the world: 4 major wars (10,000 or more combat-related deaths in the past year), 19 other wars (1,000 to 9,999 combat deaths in the past year), and at least 25 other “minor conflicts” (between 100 and 1,000 deaths in that past year). Among the major wars, the Afghanistan conflict has been the most costly, killing as many as 2 million people from 1978 to 2020. The Mexican drug war has killed as many as half a million through direct (combat) deaths or related (organized crime) homicides since 1989. The Yemeni crisis has killed around 233,000 people since 2014, and the Tigray war has killed as many as 50,000 people since 2020. In addition to murders, wars leave countless refugees and abused and neglected children. Some of these children are recruited as child soldiers and taught to kill.

PEACE AND PEOPLE

In contrast to the straightforward documentation of worldwide conflict and homicidal violence, the documentation of peaceful societies is more complicated. One must almost literally beat the bushes to find data on peaceful societies. Those studying these kinds of societies describe tiny communities that avoid contact with neighboring communities. They reported that adult members of such groups raised their children to be hesitant and fearful about the intentions of others. All physical violence was eschewed, including competitive games. One of the so-called "peaceful societies" is the Piaroa, a jungle people who live along the tributaries to the Orinoco River in Venezuela. The Piaroas society teaches children the evils of vanity, jealousy, arrogance, dishonesty, and ferocity. They are taught to protect themselves from unseen cannibalistic beings. Another reportedly peaceful tribe, the Semai of Malaysia, "see themselves as helpless in a hostile world that is beyond their control." In truth, there is little evidence that idyllic, nonaggressive societies ever existed. Margaret Mead based her account of the peace-loving Samoan people on approximately 9 months of research on the island, most devoted to studying adolescent girls. Numerous accounts of Samoan life, written before and after Mead's book, indicate that Mead's generalizations about the peacefulness of Samoan life were simply wrong. It is surprising to learn that even now, the proportion of men killed by warfare in many primitive societies reportedly dwarfs the proportion of men in the United States and Europe who die in wars.

What about Buddhist societies? Tibet is a country long regarded as a model of nonviolence. The Dalai Lama is a symbol of peace. In 1912, Tibet, formerly a Chinese principality, declared its independence from China and ruled autonomously from 1912 to 1950. Then Mao's army took over China. In 1958, following a series of uprisings in opposition to Chinese domination, a Tibetan army, made up predominantly of laypeople, was formed to fight the Chinese. Over the next few years, resentment of the Chinese grew among Tibetans, and in 1959 a full-blown revolt broke out, causing the Dalai Lama to flee his country. Despite their religious vows of nonviolence, Tibetan monks were active in protest movements. However, the extent to which even Buddhist monks participated in warfare remains unclear.

For years, the degree to which Tibetans, laypeople, and monks took part in training for battle and fighting remained secret. It is easy to understand why. Both monks and laypeople found themselves in an ethical dilemma. War violated the central principle of Buddhism, that is, to create no suffering for any sentient being. Even the Dalai Lama confessed his ambivalence, his personal and political quandaries. As he explained in his autobiography, "Nonviolence is not an intrinsic Tibetan trait. It is instead a philosophical component of Buddhism and a political component of state struggle." Thus, one sees humanity's difficulty in adhering strictly to nonviolence in its internal struggles. The source of this all too human dilemma will become more apparent when considering the complexities of brain structure and function.

Based on numerous studies of primitive and advanced societies, the consensus seems to be that violent conflicts motivated by jealousy, revenge, dominance, and sexuality (i.e., rape) are universal human behaviors and are especially characteristic of men. These findings do not mean that violence alone characterizes the human species. Counterbalancing these aggressive drives are gentler instincts and a brain that, under optimal conditions, is often capable of moral reasoning and peaceful behavior. Then again, as will become apparent, moral reasoning does not always govern our behaviors.

HOMICIDE

Epidemiology

According to the United Nations' Global Study on Homicide (2013), the major forces contributing to homicide include economic crises, food insecurity, low levels of human development, and high levels of income inequality.

A study by the Geneva Declaration on Armed Violence and Development (GDAVD) estimated that in 2004 the global homicide rate was 7.6 intentional homicides per 100,000 inhabitants. In 2012, the United Nations Office on Drugs and Crime (UNODC) did a similar study and found a global average homicide rate of 6.2 per 100,000. UNODC's most recent data was from 2017, where the homicide was essentially unchanged at 6.1 victims per 100,000. Of these, men committed 90% of homicides, and 80% of the victims were also men.

Most homicides are in the Americas (17.2 per 100,000) and Africa (13 per 100,000). There are much lower rates in Asia (7.3 per 100,000), Europe (3 per 100,000), and Oceania (2.8 per 100,000). The countries with the highest rates are El Salvador (62.1 per 100,000), Venezuela (56.8 per 100,000), Honduras (41.7 per 100,000), and Brazil (30.5 per 100,000).

In the United States, the homicide rate increased by 14% from 2010 to 2017. This increase occurred after several decades of decline. The highest rates were in the central and southern United States, whereas the northeast United States had decreased murder rates. In the United States, most homicides involved firearms (77% in 2020), whereas, in countries where firearms are primarily banned, such as the United Kingdom, firearm-related deaths are infrequent.

Who gets murdered? In the United States, about 30% of victims were killed by someone they knew, and strangers killed only about 14%. Spouses and family members make up about 15% of all victims. Felony circumstances such as rape, robbery, and burglary accounted for only 22.5% of murders.

Murder rates in death penalty states were consistently higher over the years than in nondeath penalty states. Thus, it is unlikely that murderers consider the possibility of a death sentence when they commit a crime, and forensic evaluations seem to justify this assumption.

What about wars? Unfortunately, after a war ends, homicide rates often remain high. These increases seem to occur independently of the scale of the conflict or whether the nation was victorious in the war. Homicide rates increased with the greatest consistency in nations that had suffered large numbers of combat deaths. Most of the wars studied occurred before developing the concept of post-traumatic stress disorder (PTSD), a mental illness familiar to psychiatrists practicing after the Persian Gulf, Iraq, and Afghanistan wars when mental health workers recognized numerous examples of homicides and suicides by traumatized soldiers. The extent to which these mental disorders contribute to postwar homicides is not well established.

Mass Shootings

In contrast with declining homicide rates by juveniles is the startling increase in mass or rampage shootings in the past several years. Between 1982 and 2005, most years had only one mass shooting. Several years had none. This number increased after 2005, the worst years being in 2017 to 2019 with 31 and 40 active shooter incidents respectively. About one-third of all mass shootings in the world occur in the United States. They have more mass shootings than any other country. However, the United States is also larger than most countries–it is more populous than most countries. Most school shootings are in the United States. The exact ranking of the United States among countries with mass shootings is a controversial subject, with different definitions, different ways to distinguish them from military conflicts, and different statistical approaches. Different approaches put the United States ranking in a range from 1st to 11th. However, it is notable that the United States is the most consistent in having mass shootings each year. Most other countries do not experience mass shootings.

It is hard to determine the comprehensive psychological profiles of mass shooters, as many die by suicide or are shot by first responders. Some have suggested typologies of mass shooters (e.g., psychopathic, psychotic, and traumatized). Some have reported that adult shooters tend to have experienced setbacks in social, family, or vocational domains as well as “autogenic components” (e.g., dysthymic reactions). Others have described shooters as having been bullied, being obsessional, grandiose, and harboring persecutory beliefs. We know that in most cases, the shooters are white, middle-aged men who act alone. Many mass shooters do suffer from identifiable mental illnesses. However, the relationship between mass shootings and mental illness is controversial.

The lead author of this chapter evaluated two teenaged mass shooters. Each of these young shooters (age 16 years) had long-standing histories of maladaptive school behaviors.

One of these youngsters had a history of auditory hallucinations dating back to the age of 12 years. He also, though highly intelligent, had specific learning disabilities for which he received special educational assistance. Of note, for years, he had been obsessed with war and guns. Periodically he came to school dressed in camouflage or all in black. The year before the shooting, he saw a psychiatrist (a gun collector himself) who diagnosed a depressive disorder and treated him with antidepressant medication. While in treatment and while taking his medication, he did well socially and academically. However, several weeks before the shooting, his parents decided he was doing well enough and should stop therapy and medication. Following cessation of treatment, his functioning deteriorated. His essays and presentations in class centered on war and shooting. At one point, he jumped up in class, clutched his head, and shouted, "Goddamn these voices in my head." His teacher thereupon punished him, making him write out 100 times, "I will not swear in class," completely overlooking his psychotic behavior. During the days prior to his shooting spree, he told his friends of his fantasies/plans of shooting up the school. No one did or said anything. One afternoon, following a bitter argument with his father (a high achiever who could not understand why his son had such trouble at school), the boy returned home and, using a firearm his parents had bought him, shot his father through the back of the head. When his mother came home, he shot her in the back ("so she wouldn't see what I had done"). The following day he came to school with guns and ammunition, killed two classmates, and wounded several others. In addition to the prior author, a second psychiatrist evaluated him for the defense and diagnosed schizophrenia. However, as is often the case, the psychiatrist and psychologist who evaluated him for the state found no psychosis. On the advice of his attorney, the boy pled guilty to two counts of first-degree murder and was sentenced to life in prison.

The second shooter, a 16-year-old girl, also shot up a schoolyard of children, using a gun and ammunition gifted to her by the male relative with whom she lived. Shortly before the day of the shooting, at the request of her school counselor, she was evaluated by a psychiatrist who recommended hospitalization for severe depression. The male relative with whom she had been living refused to allow this. At that time, it was suspected that she was the victim of incest (she kept knives in plain sight around her room to prevent anyone from entering). When asked about being bothered sexually, she was evasive. She did not openly reveal that she had been sexually abused by the male relative until a parole hearing 10 years later. The parole board ignored the relevance of this disclosure to mitigation.

During the psychiatric evaluation, the author learned that she had suffered a severe bicycle accident and had injured the left temporoparietal area of her skull. She also had a history of episodes of impaired memory for her actions and several symptoms consistent with possible complex partial seizures. In addition, her memory of the shooting itself was vague and distorted, and she lacked all memory for things she had said during an interview with police right after the shooting. Unfortunately, prior to her trial, she never had a neurologic or neuropsychological evaluation, although they had been recommended.

The psychiatrist for the prosecution found no evidence of mental illness. On the advice of her lawyer, she pled guilty to two counts of first-degree murder and, to this day, remains in prison. Several months after the shooting, the male relative with whom she had been living, married one of her girlfriends who was the same age as her. The author never saw this youngster after the initial interviews. However, years later, according to press releases, she was being treated in prison for what the press called a rare form of epilepsy.

Prior to the school shooting, she often talked about shooting and killing. Furthermore, approximately 3 days before her rampage, she warned her schoolmates that she would “do something big” the following week. Other shooters have made similar declarations that unfortunately are dismissed simply as ploys for attention.

In both cases evaluated by the prior author, sharing their thoughts of murder and mayhem appeared to be efforts to be stopped. In this author’s clinical experience, many murderers, young and old, indicate to others their homicidal thoughts days to weeks before they commit murder. Nobody gets the message.

Evolution and Violence

We cannot ignore the fact that, from an evolutionary perspective, *Homo sapiens* are apes. Because humans have uniquely large frontal lobes, humans seem to be the only creatures with the capacity to grow their foods, find shelter from extremes of climate, create new sources of energy, gain an understanding of their physiology, and invent specific cures for their ills. Humans are unique in our ability to try to understand the universe, mortality, and what we must do to protect ourselves from extinction. Our unique cognitive capacity allows us to distance ourselves and examine our behaviors as if we were studying another species. We can recognize our capacities for jealousy, hatred, vengeance, lust, and desire to dominate. We only need to consider the wars in our lifetimes to appreciate the advances we have made in our ability to annihilate one another.

For centuries, if not millennia, scholars assumed that humans were the only animals capable of forming groups to make war on each other, but humans are not alone. Jane Goodall documented raiding parties of male chimpanzees that perpetrate deliberate violence on other, more vulnerable groups of chimpanzees. Primatologists debate whether chimpanzee violence is innate or caused by human interference (e.g., feeding the animals and changing their habitats).

In a five-decade study of 18 chimpanzee communities, researchers reported that of 152 killings by chimpanzees in 16 communities, 92% of the attackers were male (as were 72% of the victims). Most killings involved intercommunity attacks in which the killers greatly outnumbered their victims. The investigators in this project asserted that variations in killing rates were unrelated to the impact on human beings. The researchers interpreted their findings as supporting the explanation that lethal violence among chimpanzees results from evolutionary adaptive strategies to increase access to food and mates. The primates did not catch it from humans. Other primatologists have questioned these conclusions.

There is substantial proof of human biologic closeness to the “other” great apes (chimpanzees, gorillas, orangutans, and bonobos). Advances in analyzing the structure and composition of deoxyribonucleic acid (DNA) have revealed that humans are genetically more closely related to chimpanzees than any other known species. From the perspective of chimpanzees, humans are their closest relatives. Their DNA is more similar to human DNA than it is to the DNA of other great apes. The question remains of whether the sizeable human cerebrum, which enables humans to devise weapons of mass destruction while simultaneously enabling humans to conceptualize moral codes by which to live, will, in the end, have survival value. How will people use their heads?

The recognition of our genetic similarity to chimpanzees has led to changes in the ways great apes are used for science. In the past, chimpanzees were used for experiments with toxins and infectious diseases. They were used by the Air Force and were subjected to extreme acceleration and sudden changes in atmospheric pressure. Early experiments tested the effects on primates of separation from their mothers and peers. These practices revealed the devastating effects of isolation. Ironically, though such maltreatment of chimpanzees is prohibited in the United States, immigration policies have separated human infants from their mothers.

Hormones, Gender, and Violence

Are all humans equally aggressive? Clearly not. Males of the human species, like chimpanzee males and males of most other species, are more violent than females. Whether measuring rough play during childhood or adult violent crime, the human male behaves demonstrably more aggressively than the female. These differences are not peculiar to a particular society but have been observed cross-culturally in societies as different as Ethiopia and Switzerland. In the United

States, men are about eight times as likely to commit murder, nine times as likely to commit armed robbery, and four to five times as likely to commit aggravated assault as women. Of note, men are also many times more likely than women to commit nonviolent crimes such as burglary, car theft, and drunken driving. Women are catching up to men in less physical forms of criminality such as fraud and embezzlement.

We can also witness the greater inherent aggressiveness of human males through the patriarchal nature of almost all human societies. For example, a 1971 survey of 93 societies around the world revealed that men retained all important political positions in 88% of them and that in the majority of societies, men also tended to hold all significant leadership roles within kinship groups. Most surprising is the finding that maltreatment of women by men (i.e., wife-beating) is more the rule than the exception in societies the world over. In addition, human males are more likely to commit rape than females. The human male is, in short, the predominant maker and breaker of the rules in human societies.

Much as many may wish to attribute most, if not all, of the differences between human male and female aggressiveness to cultural influences, overwhelming evidence exists that, long before birth, the die is cast; male and female brains develop differently in utero. The architectural and functional differences between male and female brains are, to this day, areas of controversy. While some investigators have noted marked differences in the “sexually dimorphic” areas of the hypothalamus and corpus callosum, others have highlighted the similarities across genders. However, human males have about 8% to 10% larger brains than females, although as we know from neurodevelopment, a bigger brain does not necessarily mean a smarter one as brain volume decreases with synaptic pruning while efficiency increases (and intelligence testing does not show a significant difference between the sexes). The most significant size discrepancies are in the cerebrum, that alleged repository of reason and impulse control.

Gonadal hormones play essential roles in the development of aggressive behaviors. Gonadal hormones are especially critical during fetal and neonatal development. In addition, these hormones affect sexually dimorphic areas of the brain, and their early presence is vital to the normal development of aggressive behaviors in such diverse species as fish, lizards, birds, and chimpanzees.

Androgens such as testosterone are metabolized in the brain to form estrogens. We now recognize that estrogens influence the sexual differentiation of the male and female brains and thus influence behavior. Estrogens influence dendrite formation at synapses. Thus, they have permanent organizing effects on the structure and function of the brain during critical prenatal periods. The transformation of testosterone to the estradiol essential to brain differentiation requires activating the cytochrome P450 aromatase enzyme. In vitro studies of neurons indicate that neurons from the male fetal hypothalamus have higher aromatase activity than neurons from the female fetal hypothalamus. We do not entirely understand the relationship of the masculinization of the fetal brain to the production of neurotransmitters and receptors in the brain. However, we have documented the differences in the amount, turnover, and localization of catecholamines in male and female rat brains. The behavioral effects of gender differences in the central nervous system (CNS) are incredibly complex. The steroid derivatives progesterone, testosterone, and glucocorticoids, referred to as neuroactive steroids or neurosteroids, can also alter behavioral responses and are essential in sex-related differences. The behavioral effects of neurosteroids are numerous and include sedative, anesthetic, and anxiety-reducing properties.

Several MRI and fMRI studies have reported structural and functional differences between the brains of male and female homo sapiens. For example, investigators have reported sex differences in different parts of the orbitofrontal gray matter. This area is of particular interest because men in conflict with the law have been reported to have less functional orbitofrontal gray matter than their law-abiding peers. An exception: Raine has reported areas of hypertrophy in parts of the frontal cortices in white-collar criminals.

As technology has advanced, brain researchers have been able to identify other, smaller and smaller brain areas in which men and women differ. Investigators are zooming in on numerous brain areas affected by fetal testosterone (FT), including the right temporoparietal junction/posterior superior temporal sulcus, the planum temporale/parietal operculum (PT/PO), and posterior-lateral orbitofrontal cortex. Subregions of both the amygdala and hypothalamus are also reportedly sexually dimorphic.

Some of the literature on dimorphism reads like an exercise in political correctness. Papers on the subject of gender tend to soft-pedal any possible intellectual differences between men and women, especially if the differences seem to favor the male brain. One publication from the Proceedings of the National Academy of Sciences got directly to the point: “Sex differences in human behavior show adaptive complementarity. Males have better motor and spatial abilities whereas females have superior memory and social cognition skills.”

Males have one X chromosome and one Y chromosome, whereas females carry two X chromosomes. The Y chromosome contains a particular gene (SRY) that, during gestation, triggers the outpouring of androgens. Not only does this flood of testosterone affect the density of neurons in different parts of male and female brains, but it also results in sexually dimorphic *circuits* among different brain areas.

The study of brain formation and function has made a giant leap ahead with the new techniques of optogenetics and pharmacogenetics and innovations in researchers’ ability to track and record in vivo brain activity during complex social behavior. At this point, most of these new technologies study animal brain function, particularly the brain of the laboratory rodent. The assumption is that the underlying neural mechanisms that drive aggressive behavior are likely to be conserved across species.

Using the new techniques of optogenetics and pharmacogenetics and newer in vivo recording methods, investigators have identified areas in the ventromedial hypothalamus and ventrolateral area (VMHvl) as a critical area for both the critical area intermale aggression and male-on-female sexual activity. David J. Anderson and colleagues identified a ventrolateral subdivision of the mouse VMHv1 marked by estrogen receptor 1 (Esr1+) that, when optogenetically stimulated, elicits attack behavior. However, to the researchers' surprise, weaker optogenetic activation of these same neurons promoted mounting (i.e., sexual) behavior. The finding that the same hypothalamic pathways associated with aggression are also associated with sexual acts at lower stimulation may shed light on the pairing of sex and aggression in human rape. Rape is not strictly a violent behavior. It is aggressive *and* sexual, and this research may help explain why that is. Research is also focusing more and more on the interactions of steroidal hormones and neurochemical systems. Many studies have looked at the relationship of hormones to serotonin (5-HT) function in males and the relationship of dehydroepiandrosterone (DHEA) to the modulation of gamma-aminobutyric acid (GABA) receptor function in females. A growing body of data suggests that serotonin, via its 5-HT 1A and 5-HT 1B receptors, significantly influences the regulation of aggression in species from crustaceans to humans. Increased 5-HT has repeatedly (though not consistently) been demonstrated to decrease aggression, whereas a lack of it is associated with increased aggression.

If particular hormones already give men a head start on violence, one can only imagine the possible consequences of their ingesting or injecting stimulants or steroids in the interest of bodybuilding and enhancing stamina. Anabolic-androgenic steroids and some of their metabolites and stimulants are associated with psychosis, disinhibition, aggression, and acts of violence.

A heretofore upstanding law enforcement officer, who stopped a young lady for an alleged traffic violation, then dragged her into a field of high grasses and raped her. Oblivious to her terror, he returned the victim to her car, gave her his card, and advised her to call him the next time she wanted to get together. The officer—a man with an impeccable record—then drove on to the department store where his recently estranged wife worked. He kept his gun in his holster when he entered the store. He just wanted to talk with his wife and win her back. While attempting to clear the store of customers, the officer brandished his revolver, which went off, killing a sales clerk. During the hours that followed, the officer kept several women hostage. He forced them to commit sexual acts with each other and with him. As time wore on, the captives got hungry, at which point the officer raided the till, took the change, and bought candies from a machine. When the police finally captured him, they charged him with felony murder (i.e., murder in the course of a robbery).

A psychiatric evaluation revealed that the defendant had periodic episodes of deep depression, which contrasted with his usual hypomanic state of mind. During the psychiatric examination, he was expansive and delusional about his powers to influence the police and the court to set him free. A detailed history obtained from close relatives and psychiatric records revealed a strong family history of bipolar mood disorder. Like many sexually violent offenders, the defendant also had a history of early, ongoing childhood sexual abuse by a neighbor.

The clinical team was nonplussed. Why had this upstanding officer of the law suddenly “lost it” and raped and murdered? After re-examining hundreds of pages of educational, social service, and medical records, the clinical team found a previously overlooked single allusion to the defendant’s recent use of “bodybuilding vitamins.” Before then, the clinicians had only inquired about alcohol and illegal substance use. When it finally occurred to the psychiatrist to ask about his bodybuilding “vitamins,” he readily admitted that over the past several months, he had been taking an over-the-counter potion called Ripped Fuel and an herbal concoction that at the time contained ephedrine and caffeine. He also had been consuming large amounts of a potentially mind-altering drug, chromium picolinate. In addition, to stay awake and work double shifts, he had been drinking dozens of cans of caffeinated sodas each day. As a result, he was wired and sleep-deprived. His wild sexual and aggressive explosion was a drug-induced, manic, psychotic episode in a man primed by early, ongoing sexually abusive treatment. The defendant was unaware of his predisposition to mania or that the “vitamins” could precipitate psychosis.

The man never ingested illegal substances. Thus, he had a solid defense of diminished capacity. An argument might even have been made for insanity because, at the time of his offenses, he could not have appreciated the nature of his acts or that they were wrong (a legitimate M'Naughton defense). Would a man who thought he had done wrong have given his calling card to one of his victims?

Unfortunately, this man's manic braggadocio and his sexual advances toward one of his female defense attorneys alienated his entire defense team. This situation is not unusual, and many grandiose, manic murderers manage to alienate, particularly their male defense attorneys. They lock horns. They also compete with their psychiatrists, threatening their authority as "professionals." Their manic grandiosity often leads psychiatrists to dismiss them as having "antisocial personality disorder."

After the evaluation, the legal defense team learned of the defendant's mania and the effects of drug ingestion on his behaviors (offenses). The defense team, unable to overcome their antipathy, advised the defendant to plead guilty and throw himself at the mercy of the court. He was thereupon found guilty and sentenced to death.

Approximately 10 years later, a newly appointed, experienced public defender whose ego was strong enough to deal with the defendant's ego came upon the clinical team's neuropsychiatric evaluation and contacted the writer of this vignette. With the previous clinical findings, the new attorney was able to get a new sentencing hearing at which two psychiatrists, a psychopharmacologist, and a neuropsychologist, testified to the defendant's compromised state of mind at the time of the offense. At the time of this writing, the case is not decided.

Genes, Neurotransmitters, and Violence

Criminality runs in families. However, it is hard to tease out why this statement is true. A 1950s study of more than 1,000 Glasgow boys who left school at age 14 revealed that the greater the number of convicted family members, the more likely a boy would be convicted of a crime.

In David Farrington's Cambridge Study of Delinquent Development, he and his colleagues found that having a convicted father, mother, brother, or sister was associated with a child's convictions. In addition, same-sex relationships were more strongly associated with criminality than opposite-sex relationships, and having an older criminal sibling was more closely associated with a child's criminality than having a younger one.

Twin studies and studies of adopted away children of criminal biologic parents support a genetic component in the genesis of violence. A recent study of a large sample of twin pairs, using parent-reported data, found an association between early temperament and later aggression. Emotionality and high activity, especially in boys, were associated with subsequent aggressiveness.

Early studies of adopted away children of criminal biologic fathers and noncriminal biologic fathers revealed that the children of criminal fathers were more likely to become antisocial than noncriminal fathers' offspring. However, those children with both criminal biologic fathers and criminal adoptive fathers were at the most significant risk of becoming antisocial, suggesting that intrinsic and environmental factors contribute to antisocial behavior. Of particular importance is that the antisocial behaviors in these samples were reportedly nonviolent.

Others have compared identical twins (concordant for all genes) and fraternal twins and have found that identical twins are more similar in terms of aggressive behaviors than fraternal twins. Even identical twins separated at birth are more likely than fraternal twins or full siblings to become involved in criminal activities. These kinds of studies seem to support a genetic predisposition to antisocial behavior. The critical two words: "seem to."

Here is the hooker. What researchers find is limited to what they look for or ask. Studies are often large and impersonal. By large it meant studies of documented characteristics of hundreds, even thousands of subjects. Such studies are too large to allow for lengthy neuropsychiatric assessments of the subjects, much less of their families. Interviews in very large studies tend to be formulaic and conducted by trained interviewers rather than experienced clinicians. They gain in uniformity what they lack in the flexibility necessary to pursue areas of particular interest when they come to light. Subjects fill out checklists, as do interviewers and observers (e.g., parents and teachers). Investigators also rely on juvenile justice records or adult criminal records. Such studies allow for sophisticated statistical analyses. More often than not, they confirm that identical twins or fathers and sons are more likely than nonbiologically related relatives to get into hot water. The question is, why?

The answers to that kind of question depend not only on the records one locates but also on the time one spends interviewing the subject and relatives and the detail of the questions one asks. If one asks only whether a father, sibling, or another relative ever had trouble with alcohol or with the law, or even with mental illness, or if one looks only at arrest records, one will often find the names not only of parents but also of the defendant's "sisters and his cousins and his aunts" (Gilbert and Sullivan). However, one cannot simply ask about mental illness. The answer to that is usually "no." One must take the time to inquire about relatives who were "going and going and going," "too loud," "too talkative"; "too suspicious"; "too controlling"; relatives who were "performers," "tycoons," "big spenders," or "gamblers."

Having identified such relatives, one must then ask in greater detail about that person's behaviors. In that way, the researcher or clinician is more likely to come upon a family history of manic or hypomanic, or even psychotic signs and symptoms in relatives. After gathering such information, the next question must be “who else in the family . . . and who else?” Suppose one takes the time to ask about people in the family with “moodiness,” “depression,” “suicide attempts,” “periods of isolation and withdrawal,” “over-suspiciousness,” “episodes of ragefulness,” as well as “alcohol” or “drug” problems. In that case, one is more likely to uncover a hitherto unknown family history of severe mental disorders that manifested themselves as aggressive antisocial behaviors and landed them in jail. Many of the relatives of the murderers the author has evaluated had documented family histories of severe psychiatric illness when an interviewer took the time to look for these data. Many had been imprisoned. Similarly, when taking a perpetrator’s medical history, one cannot settle for “Any serious accidents”? If the answer is “none,” one must pursue the topic (e.g., “How about bicycle accidents”? “Car accidents”? and “Falls”?). Any positive responses must be followed by “When else”? And “When else”? Ask. Look for scars on the face, head, or arms and the back, legs, the soles of the feet, the buttocks, and even on genitalia.

The secret is time. It takes much interviewing and record reviewing to uncover the extensive psychiatric and familial characteristics of perpetrators and their first, second, and even third-degree relatives. It takes more time to dig up the institutional records or find other family, friends, and relatives who can supplement and support indications of severe psychopathology and abuse in several generations of family members. Many of these psychiatrically impaired family members have come into conflict with the law at one time or another. That does not mean that the family carries a “warrior gene” for violence or criminality. It is tempting to settle for what one can get quickly and en masse (e.g., arrest and birth records). Extensive record review-based studies are a bit like buying in bulk at Walmart. Lots and lots of stuff, but the quality of the merchandise is pretty uniform. Now Tiffany is a different story. You pay a lot for a unique bauble, but boy, oh boy do you take your time studying the facets of those tiny jewels. Are there flaws? You will need a magnifying glass.

Some researchers who deal primarily with large cohorts or esoteric genetic concepts try to bring their studies to life with a colorful clinical example. Such was the case of Jeffrey Landrigan. Landrigan had twice killed and was sentenced to death in Arizona. After lengthy appeals regarding the failure of his lawyer to develop mitigating evidence, Landrigan was put to death. His case was used recently in scientific publications to illustrate the genetics of violence. He was said to have been adopted at birth into a loving middle-class professional family. The researchers stated that he was nevertheless a particularly troublesome child from the beginning. They said that even within this lovely family, his behavior progressed from temper tantrums at 2 years of age, abusing alcohol at 10 years of age, being arrested for burglary at 11 years of age, abusing drugs as a teenager, to killing his first victim at age 20 years. After escaping from prison, he again committed murder and was sentenced to death.

By chance, another inmate noticed that Landrigan bore a striking resemblance to another career criminal who also abused drugs and killed twice. This man turned out to be Landrigan's biologic father. With a bit more inquiry, the researchers learned that Landrigan's grandfather had been an "institutionalized criminal," but the nature of the institution remained a mystery. Alas, upon discharge from the "institution," grandpa was "shot by the police." According to the researchers, Landrigan's great-grandfather was "a notorious bootlegger." Now from bootlegging to murder is a long stretch (literally and figuratively). Nevertheless, the original researchers stated that Landrigan's case documented not just the intergenerational transmission of violence but also how "the adoption design separates the genetic influences of the biologic parents from the environmental influences of the rearing home." It did not.

At the beginning of Landrigan's trial, his lawyer had done little if any work investigating either his birth history, his medical history, or even his family history. In court, Landrigan was obstreperous, repeatedly interrupting his lawyer and boastfully telling the judge to go ahead and sentence him to death, which the judge obligingly did. Several accounts of Landrigan's behavior in court suggest that he was in an excited state and could not keep his mouth shut except to interrupt his lawyer and challenge the judge.

After the trial, Landrigan calmed down enough to realize the consequences of his court behavior and gathered mitigating evidence with the help of public defenders. His initial appeal for postconviction relief, based on his first lawyer's failure to present any mitigating evidence at all, was "summarily denied" by the Arizona Supreme Court. Eventually, in 2005, the United States Court of Appeals, Ninth Circuit, agreed to hear the petition of ineffective assistance. They stated, "We conclude Landrigan has raised a colorable claim that his counsel's performance fell below the objective standard of reasonableness and that he was prejudiced by their errors." They deemed Landrigan's initial trial attorney ineffective because "Counsel is not deficient for failing to find mitigating evidence if, after a reasonable investigation, nothing has put the counsel on notice of the existence of that evidence . . . In this case, however, the initial investigation did reveal potential mitigating evidence . . . in addition to the evidence described above . . . Counsel did little to prepare for the sentencing aspect of the case."

A psychologist, who had interviewed Landrigan pretrial, recommended additional testing, but Landrigan's attorney failed to authorize any follow-up tests. The court also cited potential mitigators that the original defense attorney had evidence that Landrigan's birth mother was alcoholic and drug-addicted during her pregnancy with him and failed to inform the court. Furthermore, she did not give him up at birth but rather attempted to raise him until Landrigan was 6 months old, a crucial developmental period. She then abandoned him at daycare. One day she just did not bother to pick him up. Thus, Landrigan's adoption was not at birth. The state took custody of him and, at 8 months of age, he was adopted. However, further investigation brought to light the mythology that he was "adopted at birth into a loving middle-class professional family and was nevertheless a particularly troublesome child from the beginning." The court continued, "Landrigan began drinking at an early age because his adoptive mother was an alcoholic (at times consumed a fifth of vodka) and would walk around nude in front of him . . ." According to court records, she would frequently slap him and once even hit him with a frying pan. There was strong evidence that Landrigan suffered from severe psychiatric illness (i.e., Landrigan overdosed in class in eighth or ninth grade). After his trial, according to court records, Landrigan did undergo a thorough neuropsychological evaluation (which concluded) the concurrence of the factors mentioned above—Landrigan's genetic makeup, in utero exposure to teratogenic substances, early maternal rejection, and his troubled interactions with his alcoholic, sexually inappropriate, abusive adoptive family—resulted in disordered behavior that was beyond the control of Mr. Landrigan . . . and left him unable to function in a society that expects individuals to operate in an organized and adaptive manner. Thus, spoke the court. After his trial, Landrigan's appeals attorney arranged for a neurologic examination by Pamela Blake MD, at the time, an assistant professor of neurology at Georgetown. According to her report, Landrigan's mother "would come home from work and start drinking; she would drink until she would pass out on the couch." Dr. Blake stated, "The neurologic examination indicates several abnormalities consistent with impairment of the frontal lobes of the brain. The frontal lobes control higher level ("executive") function, impulses, regulation of behavior, and actions. . . . Numerous studies of brain function have demonstrated a correlation of this type of brain disorder with repeated aggression." Dr. Blake documented "paratonia in all four extremities . . ." "inability to perform three-step hand movements (The Luria)," and "abnormal antisaccade functions." All of the above are signs of significant frontal lobe dysfunction.

Eventually, Landrigan's case reached the Supreme Court of the United States, where his death sentence was confirmed by a five to four majority. However, Justice Stevens wrote an eloquent dissent. Stevens said, "Significant mitigating evidence that may well have explained respondent's criminal conduct and unruly behavior at his capital sentencing hearing was unknown at the time of sentencing . . . correspondent's counsel failed to conduct a constitutionally adequate investigation."

On Tuesday, October 26, 2010, Jeffrey Landrigan was put to death. During the week preceding his execution, the sentencing judge from his initial trial told a clemency board that she would have given him a lesser sentence had she been aware of the brain damage caused by his mother's heavy drinking while pregnant and had she known of his turbulent childhood. She said, "The death penalty, in this case, is not appropriate and never has been." The clemency board reportedly split on whether or not to advise Arizona's Governor, Jan Brewer, to commute Landrigan's sentence to life without parole. She did not.

The author of this chapter chose to include this case study for two reasons: First, it illustrates the ease of dismissing violent defendants as inherently evil. Second, it calls into question the existence of a "warrior gene" that trumps the effects of a nurturing upbringing. The author's research team has evaluated hundreds of extremely violent individuals, most of whom came from chaotic, violent households. More critical, careful examination of first- and second-degree relatives usually revealed a family history of severe mental illness (e.g., bipolar disorder, psychoses, intrafamily violence, and physical or sexual abuse). Unfortunately, like the Landrigan case, many of the patients, particularly the men abused by fathers or mothers, withheld these data. The author is currently working on just such a case. She has seen murderers who go to their executions rather than reveal their mothers' and fathers' brutal, often sexually abusive behaviors. In two cases, families threatened to cease visitation if the person revealed the sexual abuse. One family cut off all communication with their son when he revealed his extraordinary physical and sexual abuse. Their son's death sentence was commuted, but he has not seen his parents in years. Because of confidentiality, the author has been obliged to remain silent in most of these cases.

This case was also chosen because, more often than not, what looks like a genetic transmission of violence is, instead, a predisposition to severe mental illness or a history of repeated CNS trauma that on clinical evaluation indicates frontal or temporal lobe/amygdaloid abnormalities. Landrigan had frontal lobe abnormalities. Violence, child abuse, and mental illness are handed down from generation to generation, and thus, on the surface, violence appears to be genetic. Identifying these constellations is a mystery to be solved. Combinations of intrinsic and extrinsic interactions perpetuate what looks to outsiders like a simple genetic trait. The authors' team has encountered family after family to whom they can trace intergenerational mental illness, violence, child abuse, and institutionalizations. State prisons and county jails appear in the histories of violent families much as Harvard and Yale appear in the histories of intellectual/upper-middle-class families. Certainly, genetics plays some role in both pedigrees, but DNA is rarely the whole story. (Of note, one of the protective factors from violence is intelligence quotient [IQ].)

In the middle of the 20th century, at least partly in reaction to Lombroso's earlier theories that criminals represented a degenerate biologic phenomenon, investigators looked almost exclusively at society for explanations of criminality. The past half-century has seen an almost 180-degree about-face in the study of violence and criminality, as investigators have turned back to biology to explain deviance, especially violence. Advances in the so-called "hard" sciences (e.g., genetics, biochemistry, cellular biology, neuroanatomy, and neurophysiology) have increased the ability to identify, measure, and even visualize biologic similarities and differences among people thus encouraging this direction of research. Studies of the relationship of such neurochemicals as serotonin, norepinephrine, dopamine, vasopressin, GABA, glutamates, and nitric oxide to aggressive behavior in animals have encouraged similar studies in humans. Initially, the finding of a relationship between low cerebrospinal fluid (CSF) 5-HIAA (a metabolite of serotonin) and impulsiveness and aggressiveness in humans seems to shed light on why some individuals are more aggressive than others. Studies of men depleted of tryptophan, the amino acid precursor of serotonin, have indicated that they become more aggressive in a laboratory setting than men not so deprived. Of note, those men with a tendency toward hostility as measured on an aggression questionnaire reportedly responded more aggressively to tryptophan depletion than did nonhostile men. Chimpanzees with high blood levels of serotonin are reportedly nicer to their peers and more likely to groom them.

However, the relationship between serotonin levels and degrees of aggression is not simple. For example, a study of young (7 to 11 years old) aggressive boys with attention-deficit/hyperactivity disorder (ADHD) revealed a *greater* prolactin response to fenfluramine (i.e., greater serotonergic responsivity) in aggressive than in nonaggressive boys. In addition, a study of blood serotonin levels in adolescents diagnosed with conduct disorder reported that those with early-onset maladaptive behaviors had *higher* blood serotonin levels than later offenders and that blood 5-HT correlated positively with ratings of aggressiveness.

Researchers included environmental factors in a study of the relationship of brain 5-HIAA levels in infants to subsequent behaviors. Clarke and colleagues followed a sample of children for their first 30 months of life and found only a modest correlation between 5-HIAA and aggression. In contrast, they found that environmental factors such as poverty and poor family cohesiveness were strongly associated with behavior problems at 30 months of age. Another study documented a relationship between low socioeconomic status, low prolactin response, and increased aggression. Finally, a meta-analysis of 39 studies linking 5-HIAA levels to violence in humans found that psychiatric patients as a group, whether violent or nonviolent, had similar 5-HIAA levels. The mean 5-HIAA level in both violent and nonviolent patient groups was lower than the mean level in groups of normal, healthy volunteers. Thus, serotonin levels may prove to be indicators of overall psychopathology that only occasionally manifest as aggressiveness.

One of the most critical and influential genetic studies was Brunner’s 1993 report of a family in the Netherlands in which five generations had behavioral disorders, including arson, rape, exhibitionism, and attempted suicide. In addition, many individuals had been aggressive. Each affected family member had defective gene coding for monoamine oxidase A (MAOA), causing increased serotonin, norepinephrine, and dopamine. As these affect cognition and behavior, some refer to this gene as the “warrior gene.” Only the male children of mothers carrying the abnormal gene showed behavioral abnormalities.

Piton and colleagues reported on a French family with a similar genetic abnormality and several family members suffering from various behavioral abnormalities. Piton reported that several affected males had cognitive difficulties, a 7-year-old boy was autistic, impulsive, and aggressive, and two maternal uncles had severely delayed psychomotor development and required placement in a special needs school. They, too, had autistic features, could not read or write, and were given aggressive outbursts. Of note, both of these family members were maltreated and sexually abused as children.

Since discovering the MAOA genetic abnormality, some clinicians/researchers have coined the term “Brunner Syndrome,” designating the MAOA abnormalities as a recessive X-linked disorder characterized by impulsive, aggressive behaviors, and mild retardation. However, others have taken issue with this phenotypic concept because of the wide variety of signs, symptoms, and behaviors reported to exist among individuals carrying this MAOA abnormality.

Another gene, the serotonin transporter gene, has been associated with behavioral problems and with several different psychiatric disorders. Studies have focused on the short and long types of alleles of the serotonin transporter gene. People carrying the short allele appear to have lower-than-normal serotonin levels in their blood. In addition, some studies have reported a greater predisposition to aggressiveness in individuals with low cerebrospinal levels of serotonin.

Dopamine, another catecholamine neurotransmitter, influences many mental functions, including cognition and emotion. The dopamine receptors belong to the family of G protein-coupled receptors (GPCRs). DRD1-type receptors are associated with stimulatory function, and DRD2-type receptors are generally associated with inhibitory functions. Investigators have identified more than 200 polymorphisms in the sequencing of this gene. Variants of the DRD2 gene are associated with alcoholism and addictive disorders, mood disorders, and psychotic disorders. In addition to genes, many other biopsychosocial influences and constellations of influences contribute to a violent lineage. Notably, although each cell carries the same load of genes, when, where, and how a gene will be “turned on” (i.e., activated) depends in great measure on environmental stimuli.

One of the most complex genetic studies illustrating the above is that of Hexuan and colleagues. They examined the possible relationships between youth delinquency and violence and 403 variants of 39 genes against what they termed high/moderate levels of social control in contrast to low levels of social control (i.e., high/moderate parental attachment vs. low parental attachment). Their theoretical underpinning was that changes in environmental conditions could influence the expression of genes related to different phenotypes in both animal and human studies.

They found that estimates of the total variance in serious delinquency attributable to the 403 SNPs were, in and of themselves, nonsignificant at the 0.05 level. The researchers then examined the collective genetic contribution of the 403 SNPs under low social control (LSC) compared with high/moderate social control (HMSC). They found that genetic risk for delinquency and violence was significantly greater for subjects weakly attached to parents and school and loosely disciplined by them. In addition, living in lower-income neighborhoods also increased the genetic risk for delinquency and violence. In short, their analysis showed “significant interactions of aggression-related genetic variants and socio-environmental variables . . . favorable social conditions are associated with smaller collective genetic contributions whereas adverse social conditions are associated with greater genetic contributions to adolescent delinquency and violence.” This association amplifies when we consider the nature and interactions of child abuse with neuropsychiatric variables.

PSYCHOPATHY

To this author’s mind, the concept and subsequent reification of the diagnosis “psychopathy” hamper the understanding of criminality and violence. Robert Hare spearheaded the term when, in 1970, he created a checklist of behaviors (based in part on Hervey Cleckley's book, *The Mask of Sanity*), combinations of which Hare believed characterized a psychopath. He subsequently revised and updated the list. The list includes such items as "grandiose sense of self-worth," "need for stimulation," "promiscuous sexual behavior," "impulsivity," and "many short-term marital relations" as well as "early behavior problems," "shallow affect," "juvenile delinquency," "revocation of custodial release," "lack of realistic long-term goals," and "criminal versatility." Then come a series of clinically subjective criteria (e.g., "callous lack of empathy," "cunning manipulation," "failure to accept responsibility for one's own actions," "lack of remorse or guilt," and "irresponsibility") some of which come under the rubric of "pleading not guilty." Unfortunately, there is little guidance on how to ascertain cunning or callousness. What follows is a group of characteristics typical of many offenders with frontal lobe dysfunction, such as "poor behavioral controls," "impulsivity," "irresponsibility," and "shallow affect."

In reality, the items above, and their combinations, cover a multiplicity of different disorders. According to Hare, in many cases, one need not even meet the patient. Just rummage through his records to determine what items seemed to fit. Nonsense.

To this writer's mind, psychopathy and its synonyms (e.g., sociopathy and antisocial personality) are lazy diagnoses. As a result, the authors' team has seen scores of offenders dismissed as psychopaths. However, detailed, comprehensive psychiatric, neurologic, and neuropsychological evaluations have uncovered many signs, symptoms, and behaviors indicative of such disorders as a bipolar mood disorder, schizophrenia spectrum disorders, complex partial seizures, dissociative identity disorder, parasomnia, and, of course, brain damage/dysfunction.

Almost 20 years ago, this author evaluated a jittery, jumpy prisoner who had killed a policeman and whose psychiatric, neuropsychological, and neurologic examinations were consistent with frontal lobe dysfunction. Nonetheless, he was advised to plead guilty and was sentenced to death. He was dismissed as a psychopath. Almost 20 years later, he was lucky enough to have a very senior, experienced appeals attorney assigned to him, a man who read our old reports and wondered why they had not been introduced at trial, at sentencing, or at least on previous appeals. We re-evaluated him, and the signs of CNS dysfunction were even more severe than when we first saw him. However, this time, we obtained an fMRI. Figure 28.2–1 shows that this man had diffuse cortical atrophy, the most serious of which was orbitofrontal thinning (see Figs. 28.2–1 and 28.2–2).

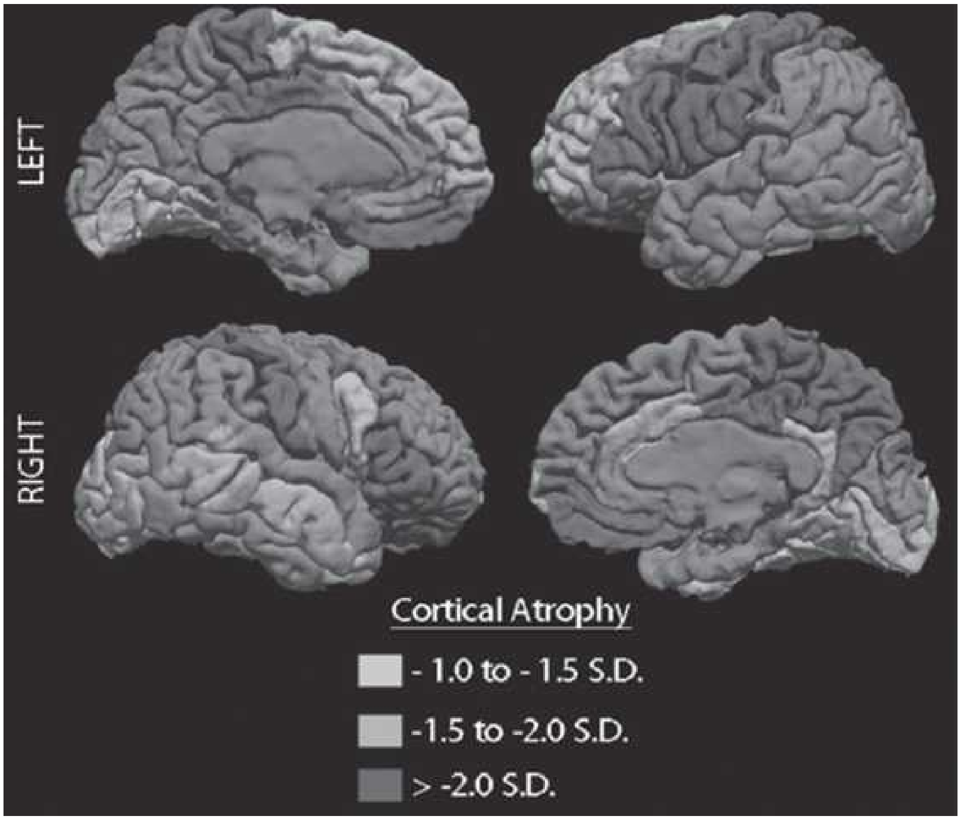

FIGURE 28.2–1. Analyzed by Ruben Gur, Ph.D. (Courtesy of Ruben C. Gur, University of Pennsylvania, Philadelphia, PA.)

The mish-mash of different characteristics listed in the psychopath checklist has encouraged otherwise impeccable researchers, using sophisticated technologies (e.g., MRI, PET, and fMRI) to lump together criminals under the label “psychopaths” and compare them to noncriminals. Some have reported at least two different types of brain function or dysfunction in their psychopathic samples. The concept of psychopathy as described by Cleckley and “measured” by Hare just does not work. Nevertheless, in the experience of the authors’ research team, just about every violent criminal they have ever examined has been diagnosed psychopathic or antisocial at one time or another. One of the criteria for diagnosing psychopathy is having been designated “conduct disordered” as a youngster. Many children labeled as having a conduct disorder have had many of the same psychopathologies as adult criminals. Thus, “conduct disorder” diagnosed in children smoothes the way to diagnosing psychopathy in adulthood. Unfortunately, clinicians use these terms indiscriminately to designate individuals young and old whose behaviors are unpleasant or dangerous and whom examiners would prefer to spend as little time as possible interviewing.

ENVIRONMENTAL INFLUENCES ON AGGRESSION

Intrauterine and Perinatal Factors

The position in utero of animals within a litter may influence behavior. For example, female mice that develop in utero between two males are more aggressive after birth than those positioned between other females. Thus, animals from the same litter who have shared the same uterus at the same time have not had identical prenatal environments. Maternal stress during pregnancy can affect the behavior of offspring long after delivery.

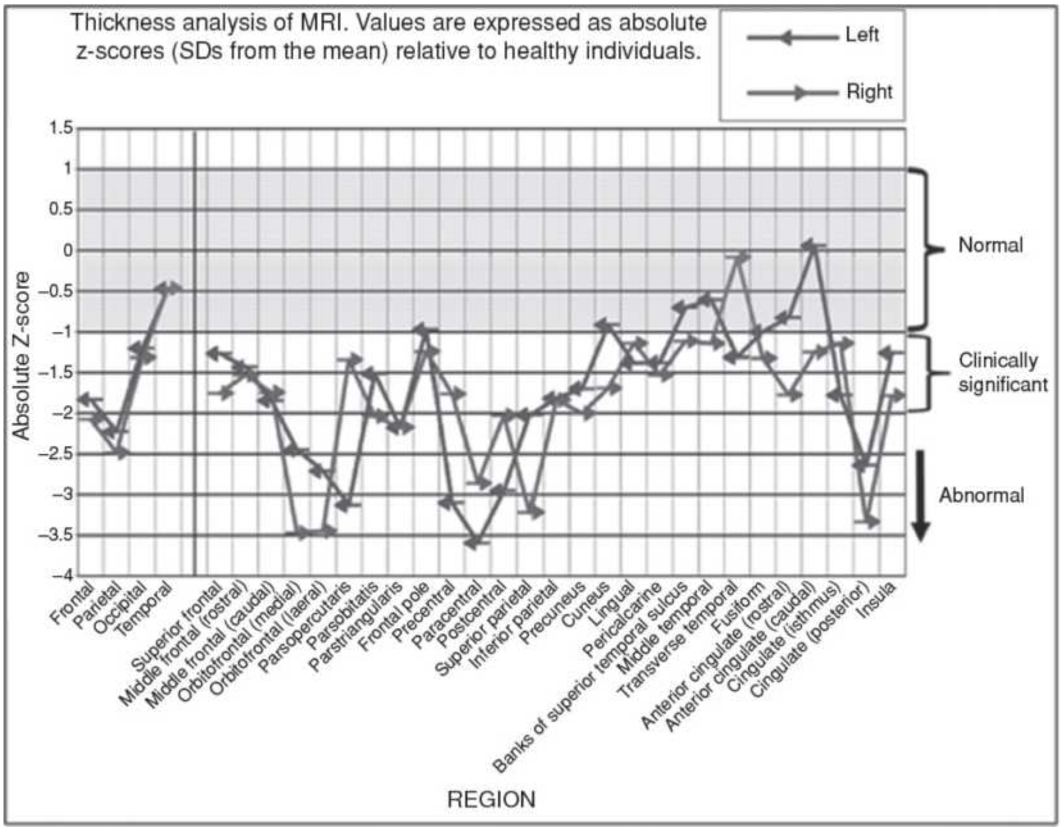

FIGURE 28.2–2. Analyzed by Ruben Gur, Ph.D. (Courtesy of Ruben C. Gur, University of Pennsylvania, Philadelphia, PA.)

In humans, there is strong evidence that different kinds of noxious prenatal influences ranging from minor viral infections to maternal anxiety and psychological stress correlate with childhood maladaptation. There is strong evidence regarding the adverse effects of maternal malnutrition, alcoholism, and other noxious substances on fetal development and subsequent postnatal social and intellectual functioning. There is a higher prevalence rate of histories of perinatal problems in the delinquent versus nondelinquent populations. Given the vicissitudes of life in utero and at birth, one should not conclude that even temperamental styles such as “emotionality” and early activity levels necessarily reflect genetic endowment. The pregnancies of mothers who release their children for adoption cannot have been uncomplicated. There is evidence that even heightened maternal cortisol due to stress affects the fetus and increases aggressiveness postnatally. Thus, the adopted away children of criminal parents, in all likelihood, had rocky in utero experiences.

Moreover, the genes such children inherited may well have predisposed them to various mental disorders other than psychopathy or antisocial personality disorder. In other words, behaviors that at first appear to be genetic may not be.

Nurturing

Nurturing affects temperament. Animals bred to have a gentle nature, if cross-fostered by adult females of a violent strain, will become more aggressive than usual. Mice reared in their fathers' and mothers' company become more aggressive than those raised only by mothers. Isolation during critical postnatal developmental periods can engender aggression in otherwise gentle animals. Thus, there is evidence from animal studies that innate temperament can be modified and that the quality of parenting and environment influence adult behaviors.

Early experiences have potent and enduring effects on physiologic and psychological development and, therefore, on behavior. Social adaptation depends mainly on adequate parenting during the first days and weeks of life in animals and humans. Although many reptiles can defend themselves shortly after birth, most mammals require parental protection for more extended periods. And humans? Even longer.

In humans, the first months and years of life are periods of extreme vulnerability, emotionally and biologically. Although the auspicious neuropsychiatric and emotional cards nature deals with at birth may seem to predict a successful game, experiences can reverse what should have been a winning hand under other circumstances. For example, research on attachment has shown that how mothers and infants interact during the earliest weeks and months of life affects subsequent aggression in the child. Remember Landrigan. Adequate mothers sense their children's needs and respond appropriately. However, mothers who lack tenderness and harbor suppressed anger toward their babies seem to engender insecure, avoidant attachments. Avoidant attachment correlates with subsequent aggressive behavior in boys.

Poor mothering takes its toll intellectually as well as behaviorally. Infants with disorganized attachments lag in their mental development irrespective of maternal IQ. Lyons-Ruth and colleagues found that the combination of early disorganized attachment and lags in cognitive development were associated with aggressive behaviors at 7 years of age. Thus, it would appear that poor mothering affects brain function itself.

ABUSE

Copious literature attests to the association of maltreatment in childhood and aggression in adolescence and adulthood. The extent to which the category of “abused” children overlaps that of children who have suffered attachment disorders and the independent influences of each are difficult if not impossible to measure. Whereas complete emotional deprivation of infants leads to severe depression and even death, lesser degrees of neglect correlate with poor peer relationships and aggressive behaviors. Some research suggests that neglect can be as powerful an influence on the development of aggression as physical abuse. Research has repeatedly shown that severe ongoing physical or sexual abuse is associated with subsequent extremely violent behaviors.

Early repeated sexual abuse seems to have especially devastating consequences. Grotesque sexual crimes have been committed by men who, as children, experienced extraordinary sexual abuse. Many have been so traumatized that they do not remember the torture.

How does severe abuse and neglect engender aggression? The answer is not straightforward, and we must glean it from various studies and observations. What is clear is, for humans, maltreatment affects every aspect of functioning—cognitive, emotional, and physiologic.

A typical result of abuse or neglect is direct injury to the CNS. Almost any CNS trauma may be associated with hyperactivity, emotional lability, diminished intelligence, poor judgment, and impulsiveness. Children with these kinds of deficits and vulnerabilities have trouble thinking ahead or understanding the consequences of their behavior. In the author’s studies of violent juveniles and adults, medical histories were so replete with evidence of CNS insults that it was usually impossible to ascertain the exact cause of brain dysfunction. We know that injuries to the CNS, even apparently mild concussions, can have a cumulative effect neurologically, cognitively, and behaviorally. Extremely violent juveniles have often experienced multiple accidents and injuries, many of them inflicted by those entrusted with their care. Of note, a recent study of perpetrators of domestic violence reported neuropsychological impairments of their executive functions, memory, attention, IQ, and empathy. In the author’s studies, they could not distinguish between children exposed to domestic violence and physically abused children since they went hand in hand in most cases. Examining the heads, bodies, limbs, and even genitalia will often reveal traumas that the subject/patient cannot recall or will not reveal.

Psychological and Cognitive Effects of Abuse

Maltreatment has psychological consequences. Abuse changes how we perceive the environment. Studies of animals and children have shown that one of the consequences of maltreatment is the development of hypervigilance. Defeat increases defensiveness in laboratory mice, and conditioned defensiveness is generalized to other situations and other opponents. Likewise, abused children seem to become hypervigilant, repeatedly misinterpreting their surroundings and perceiving ambiguous stimuli as threatening. Once a response has been established or conditioned (e.g., hormonal or neurophysiologic response to fear-inducing stimuli), it is easily re-evoked by exposure to similar stressors. Such recurrent, maladaptive physiologic responses seem to play essential roles in the hypervigilance and retaliatory aggression of many severely abused, violent children and adults.

Child abuse also contributes to violence by diminishing a child's expressive skills. Severely abused children manifest specific kinds of deficits in verbalization. For example, abused toddlers are less able to use words that reflect their inner states or feelings than normal children. In addition, the tendency of abused children to repress or deny their painful feelings as well as the pain of others, their inability to articulate emotions, their emotional lability and impulsiveness, their tendency to distort reality and perceive threats where they do not exist, and their resultant tendency to put angry feelings into actions rather than words can become a relatively enduring adaptational style.

Modeling of Abusive Behaviors

We also learn aggression. There is sound experimental evidence that modeling plays a vital role in the development of aggressive behaviors in animals and children. We learn aggressive behaviors (as we do more adaptive behaviors) through reinforcement. One of the most important contributions to the current understanding of the role of reinforcement in the genesis of aggression is Patterson's observation that when children's aggressive behaviors are punished severely by parents, they tend to continue. Farrington also found severe physical punishment to be a significant antecedent of aggressive delinquency. Murray Strauss has documented the aggressive behaviors induced by corporal punishment. Probably the most powerful generator of aggression in animals and possibly in humans is the repeated infliction of pain. It is, therefore, no surprise that children who have been painfully physically abused tend to behave in more aggressive ways than their nonabused peers.

To summarize, child abuse contributes to violence in several ways:

1. It provides a model of behavior
2. It teaches aggression through reinforcement
3. It inflicts pain, a known powerful stimulus to violence
4. It creates neurophysiologic responses in the brain that both damage the brain and affect its functioning
5. It damages parts of the brain related to impulses and their control
6. It increases paranoia
7. It impairs cognition and the ability to make judgments and foresee consequences

8. It diminishes the ability to verbalize feelings
9. It interferes with the child's ability to empathize with the feelings of others
10. In the extreme, it leads to dissociation

In short, early ongoing child abuse is the most potent known generator of child, adolescent, and adult violence.

SOCIAL FACTORS AND VIOLENCE

Courtwright, in his trenchant analysis of the sociocultural contributors to violence in the United States, observes, "other things being equal, the total amount of violence and disorder in society is negatively related to the percentage of males in intact families of origin or procreation ." He goes on to say, "Families rather than governments are the first best defense against violent and disordered behavior." He points out that the particular mixture of demographic, cultural, and social characteristics of the American frontier, its surplus of young, unparented, unmarried men, its code of honor, its racist attitudes, its heavy alcohol consumption, its lack of religious convictions or practices, its devotion to guns, and its lack of an effective system of law enforcement formed a matrix for violence. Moreover, the closing of the frontier did not eliminate the "national self-image" in that, to this day, Americans continue "to think a manly man is someone with a gun and an attitude." Courtwright states, "Indeed, it is possible to think of the urban ghettos as artificial, unusually violent frontier societies—vice-ridden combat zones in which groups of armed, unparented, and reputation-conscious young bachelors, high on alcohol, cocaine, and other drugs, menace one another and the local citizenry, undeterred." At this time, the United States is far more violent than most Western European and Asian countries.

In addition to the availability of guns, what else promotes violence? Material poverty and extreme economic inequality are associated with homicide rates; this is true if one compares violence in states, cities, or neighborhoods within cities. Whether the significant contributing factor is socioeconomic inequality or racial subordination is academic. Violence, poverty, and racial subordination cluster together in the same populations. Being somewhat low on the socioeconomic totem pole is not so much associated with violence as living in *grinding* poverty. However, poverty alone does not seem to account for violence. Sociologists believe that a strong influence on criminality springs from "social disorganization," the lack of extended families to support youngsters whose families lack the stability to nurture them properly.

Numerous studies from various parts of the world have documented an association between low or low-normal IQ and violence. It would be necessary to stress that neither impulsivity nor low IQ is necessarily inherited. The wealth of data cited above attest to the potentially adverse effects on intelligence, temperament, and behavior of impaired attachment, child abuse, neglect, and neurologic insults. In summary, the kinds of external stressors and intrinsic vulnerabilities that play roles in individual violence (e.g., physical discomfort, lack of nurturance and stability, CNS dysfunction, increased rates of mental illness, and child abuse) are more pervasive in disadvantaged communities. They contribute significantly to the disproportional rate of violence in the most impoverished sectors of society.

BRAIN DYSFUNCTION AND VIOLENCE

The part of the human brain thought to be involved in sexual and aggressive drives lies deep in the areas often referred to as the *limbic system*. The areas of this region thought to be most closely involved in aggression are the hypothalamus, septal area, and amygdala. Lesions, electrical stimulation, and infusion of neuroactive substances into the amygdala have shown that those nuclei are necessary for animals to experience particular emotions, especially fear. In species as diverse as crocodiles and humans, the amygdala processes fear, stimulation of which elicits defensive, aggressive responses.

Humans can usually perceive a threat, assess it intellectually, and choose to act or not to act in response to the stimulus. The PFC allows humans to exercise some control over their instinctive responses. The reflexive, violent response of some repeatedly aggressive individuals to innocuous stimuli is often the result of paranoid misperceptions and immediate defensive responses. In the late 1960s, Vernon Mark and Frank Ervin in *Violence and the Brain* explored the association between electrical activity in the limbic system and aggressive behaviors in humans. Using electrodes inserted deep in the brains of recurrently aggressive subjects, they found abnormal electrical activity corresponding to episodic rages. Whether directed aggression can occur during a complex partial seizure generated within the limbic area remains a subject of debate among neurologists. However, investigators have reported aggressive behavioral changes in individuals with a damaged medial temporal lobe or brain tumors in the limbic system. Violence can also occur interictally in patients with epilepsy. The author’s studies have indicated a high prevalence of various psychomotor symptoms and paranoid ideation in violent delinquents.

The serial killer, Arthur Shawcross, demonstrates the importance of obtaining a complete medical history and account of the offense(s). Shawcross had many dissociative symptoms and a history of early severe ongoing sexual abuse that could easily have influenced the rapes and murders he committed. However, a history of several severe blows to his head, evidence of which could be documented by the scars and indentations on his head, suggested the possibility that brain dysfunction might have contributed to his episodic violence. A detailed history of the offenses (i.e., what happened before, during, and after the offenses) revealed Shawcross's confusion and impaired memory of the acts. As Shawcross recalled, before the acts, he would be having a pleasant conversation with his victim. Then, out of the blue, something she did or said provoked a reaction. In response, he would start to sweat, and everything would turn bright white. A struggle would ensue and, shortly after that, his hands would close around his victim's neck. Shawcross would subsequently fall into a deep sleep. Hours later, he would awaken in his car, only to find a corpse in the seat beside him. Realizing what he must have done, Shawcross would drive around, seeking a place to hide the body. At least once, on a return visit to the body, he cut the victim's genitalia and ate them. Individuals with temporal lobe lesions have committed similar acts of extraordinary and bizarre viciousness (The psychiatrist for the prosecution opined that Shawcross's cannibalism was simply an effort to destroy the DNA. Surely, doctor, there must have been an easier way!)

The above account—an aura (bright lights and sweating), a misremembered act, followed by a deep sleep—had the hallmarks of a psychomotor or complex partial seizure. A review of prison records showed that Shawcross had been found unconscious in his cell on numerous occasions. An MRI of his brain revealed a cyst in the amygdalar area on his right temporal lobe and scars in the white matter of both frontal lobes. A single, routine electroencephalogram (EEG) was normal. The physicians advised his attorneys to obtain a more sophisticated neurologic assessment, including a quantitative EEG with hyperventilation and photic stimulation, but they refused. They chose instead to pursue a defense of dissociation. Such a defense is a big mistake when clear evidence of brain damage exists. Juries are more likely to believe what they can see (e.g., MRI and EEG printout). They can neither see nor hear dissociative states. For this reason, we try to videotape interviews in all cases in which dissociation plays a role. Predictably, Shawcross was tried, found guilty, and sentenced to life in prison for 10 of the 11 murders he had committed.

Before going to trial for the one remaining murder, Shawcross's second attorney decided to obtain a quantitative EEG. It revealed spiking, not only in one temporal lobe but also in both frontal lobes. Despite these findings, this new confirmatory evidence of complex partial seizures and their relationship to the crime in question, Shawcross's attorney advised him to plead guilty to the last charge, which he did. Given the offenses, Shawcross was destined to be institutionalized either in prison or in a hospital. The tragedy was the failure of the defense to educate the judicial system and the public about how damage to specific parts of the brain and early, ongoing sexual abuse can engender extraordinary, repetitive sexual violence.

A recently published computed tomograpbhy (CT) and MRI study of 287 male prison inmates, comparing 162 violent offenders with 125 nonviolent offenders (none previously diagnosed mentally ill), and 52 noncriminal controls reported that offenders displayed a significantly higher rate of morphologic brain abnormalities, with violent offenders having significantly more abnormalities than nonviolent offenders. Differences were statistically different for frontal/parietal cortex, medial temporal structures, third ventricle, and left but not right lateral ventricles. The authors concluded that their findings raised questions regarding the culpability of violent offenders with brain abnormalities.

PET and studies of regional glucose metabolism have revealed hypoactivation in lateral and medial zones of the PFC in individuals given to impulsive aggressive acts. Findings from neuropsychological testing of patients with known prefrontal lesions also support the significance of frontal lobe deficits evident on PET scans.

In short, studies from various fields suggest that many impulsive, aggressive individuals have abnormalities in the structure or function of the brain and in circuits that interpret outside stimuli and regulate emotional responses.

We now understand what we once called schizophrenia to be a multiplicity of different psychoses with different genetic constellations and biopsychosocial correlations. Similarly, we have enough data on the biopsychosocial differences between aggressive and nonaggressive offenders to separate them into specific diagnostic categories that go beyond mere "psychopathy." It is vital to cease cramming them all into the same cubbyhole of psychopathy, or antisocial personality disorder. A review of the behaviors of the patients Cleckley described in *The Mask of Sanity* reveals that most were manic or hypomanic, and today we would treat them with mood stabilizers. Here is yet another distinct group of patients heretofore also dubbed psychopaths.

SEVERE PSYCHIATRIC DISORDERS AND VIOLENCE

Whether mentally ill individuals are more violent than people in the general population has become a politicized question. Thus, perfectly reputable scientists have sometimes turned statistical and intellectual somersaults to find ways to prove that severely psychiatrically disturbed individuals are no more a threat to society than the little old lady living next door. Unfortunately, while well-intentioned, these politically correct efforts can have dire consequences for psychotic, brain-injured, impulsive, and aggressive patients and their neighbors. They get discharged from mental hospitals only to find themselves frightened, alone, homeless, and confused. When made uneasy (often paranoid) by their surroundings, they often lash out to defend themselves against real or imagined threats and wind up in jail more often than in the therapeutic settings they need. Moreover, their aggressive, disordered behaviors in prison and hospitals often land them in solitary confinement, a setting known to induce psychosis in normal people and exacerbate it in the mentally ill.

There is no question in this author's mind that the thrust of the 1998 MacArthur Foundation's Violent Risk Assessment, namely that with few exceptions, discharged patients from mental institutions are no more dangerous than other persons in the general population, is simply wrong. Indeed, most individuals with mental illness are not violent. Unfortunately, however, studies worldwide from more developed and less developed nations indicate that violence is more common among seriously mentally ill individuals than among mentally healthy individuals. These findings should not come as a surprise. After all, severe mental illness often distorts perceptions, increases suspiciousness, is accompanied by extremes of emotion, and impairs impulse control. When the brain is not functioning correctly, humans often take umbrage quickly and feel an urgency to protect themselves at all costs. It is called paranoia. Moreover, paranoia is part of almost every serious mental illness.

Studies have found that the prevalence of violence is approximately five times higher in individuals who meet the diagnostic and statistical manuals' criteria for an Axis I diagnosis than in people not meeting criteria for serious mental illness. The prevalence of violence in individuals with schizophrenia, major depression, or mania/bipolar disorders was similarly high (i.e., approximately 11% to 13%). Individuals with alcoholism or other substance abuse reportedly commit violent offenses (i.e., 25% and 35%, respectively) more often than patients with schizophrenia or major depression alone. Those with the highest prevalence of violence are young lower-class men with substance abuse disorders or major mental disorders. Almost all the differences in prevalence between violent and nonviolent patients are due to psychosis. Patients reportedly are no more violent than anyone else when they are not actively psychotic.

In the February 2008 issue of *Psychiatric Services*, E. Fuller Torrey and Jonathan Stanley reviewed the critical literature published since the MacArthur report and challenged their significant conclusions in a dialogue with two of the report's authors. The MacArthur study followed 951 psychiatric patients ages 18 to 40 after discharge from state and university psychiatric inpatient units.

Torrey and Stanley reported that the prevalence of violent, sometimes murderous behaviors among discharged patients was high. Among the 951 discharged patients, 262 (27.5%) committed at least one violent act over a 41-week follow-up. This group was responsible for 608 violent acts; 1% of the violent acts were homicidal, and there were 6 victims of these murders. The homicide rates in the United States during the MacArthur study were at a peak, ranging from 8.2 per 100,000 to 9.5 per 100,000. In the most impoverished, minority populations, among young adults, the rates approached 200 per 100,000. These numbers pale compared to the homicide rate in the discharged patient sample which was over 600 per 100,000 population. Thus, the contention that discharged patients are no more violent than others from disadvantaged communities is not correct.

There is a debate over which Axis I diagnoses are most closely associated with violence. In a study of a Danish birth cohort, hospitalized patients diagnosed with schizophrenia had higher arrest rates for violent crime than did nonpatients. A recent paper on the neurobiology of aggression in schizophrenia suggests a possible role of the catechol-*O*-methyltransferase gene that is involved in the metabolism of dopamine. Others have reported that patients with affective disorders have even higher rates of violence after hospital discharge than those diagnosed with schizophrenia. Some have found the onset of antisocial behaviors to coincide with the onset of affective disorders in many patients. Many so-called antisocial youths have unrecognized bipolar disorders. Their wild, antisocial behaviors are often dismissed as part of normal adolescent development. A glance at Hare criteria for the diagnosis of psychopathy reveals a plethora of manic symptoms (e.g., glibness, superficial charm, grandiose sense of self-worth, and promiscuous sexual behavior). Thus, use of the PCL-R criteria can lead to serious misdiagnosis.

Other psychiatric disorders associated with episodic violence, although usually not considered in the differential diagnosis, include PTSD and dissociative identity disorder. Violent behaviors may occur in reaction to flashbacks in cases of PTSD and in cases of dissociative identity disorder. Most patients with dissociative identity disorder develop aggressive alternate personality states early in life. In these states, they experience the actual pain inflicted and develop stoic but enraged and aggressive coping styles or personality states to deal with the pain. Subsequent stressors or threats, real or perceived, may evoke a switch into these aggressive personality states. At such times, these individuals may lash out, physically or sexually, and even become homicidal. Patients with dissociative identity disorder are often partially or completely amnesic for their behaviors and may vigorously deny responsibility for violent acts. Their stories may change frequently during interrogations, depending on their personality states and on what each state remembers of an event. These kinds of patients tend to be dismissed as pathologic liars. Because forensic psychiatrists and psychologists rarely entertain dissociative disorders as part of their differential diagnosis, dissociative individuals are routinely misdiagnosed as episodically explosive, simply antisocial, or as “pathologic liars.” Because they do hear voices in their heads, they are sometimes diagnosed with schizophrenia. Alcoholic offenders, who, because of brain damage, do not recall parts of their behaviors, fill in the missing spaces with varying stories. They, too, tend to be dismissed as liars.

Again, it is important to restate that most mentally ill individuals are *not* violent. However, they are more likely than mentally healthy individuals to act aggressively when stressed, especially if they have endured early, severe, ongoing maltreatment.

Differential Diagnosis

The differential diagnosis of violent behaviors is as broad as the fields of psychiatry, psychology, and neurology. The diagnoses of oppositional defiant disorder, conduct disorder, and antisocial personality disorder are hodgepodges that, as stated, cover a multitude of sins and a multitude of different kinds of signs, symptoms, behaviors, and combinations thereof. These categories if used at all should be thought of as interim designations on their way to proper diagnoses. Because violence toward others is the final common pathway of so many different kinds of neurophysiologic vulnerabilities and environmental stressors, the clinician faced with the task of diagnosing and treating a dangerously violent person must work hard. Every aspect of functioning—behavioral, cognitive, emotional, social, psychiatric, neurologic, neuropsychological, and environmental—must be investigated thoroughly.

There is no single treatment or intervention for violence because it has a myriad of causes. What is more, because so much violence is grotesque and abhorrent, the perpetrator often is dismissed as simply evil or bad. One need not be a clinician to make that diagnosis. Evil is not a scientific or medical term.

OFF-COLOR CRIME? VIOLENCE BY THE MEDICAL PROFESSION

How does one pick a color for psychiatrists and psychologists who participate in torture, even death of prisoners? It is known that at the time of the Second World War, American and Canadian doctors worked with the government on methods of getting prisoners to reveal what they knew. The use of mind-altering drugs, the effects of sensory deprivation, and other methods were not only suggested but also implemented by respected individuals in the medical profession. Psychiatrists and psychologists at other times have consulted with the military on how to train soldiers to kill. (It does not come naturally to healthy people.)

The current wars in the Middle East seem to have further degraded the morals and standards of some in the psychiatric/psychological professions. The role of psychologists and physicians in the Central Intelligence Agency’s (CIA)torture program has been likened to the four main qualities of behavior of the Nazi doctors. The report of the 2014 U.S. Senate Select Committee on Intelligence Report on the CIA's Detention and Interrogation Program revealed that medical professionals (1) experimented on detainees; (2) perverted medical procedures, turning them into rape and torture (e.g., giving their detainees enemas of fluids and pureed food substances in the guise of feeding and hydrating them); (3) induced disease in their detainees; and (4) provided enough medical care to put their subjects into condition to be tortured again. Only two so-called professionals were named, James Mitchell and Bruce Jessen; however, they were assisted by anonymous MDs. The use of torture was encouraged by the head of the United States' C.I.A. (George Tenet).

Although most of this chapter is devoted to the study of violence and criminality perpetrated by physically aggressive criminals and by greedy, flamboyant white-collar criminals and Mafiosi, next to nothing is known about what enables trained professionals to indulge in such brutality. Did they experience some sort of high? Some sort of sexual excitement while torturing? What were their incentives? According to government and newspaper reports, one incentive, money, may have been the draw. The sum of $ 81,000,000 has been bandied about. Was wealth the lure? Does that mean that for the right price, anyone could be turned into a sadist or murderer?

The people who indulge in such atrocities without blinking an eye are the ones who should be studied, if only to learn how to keep everyone else from falling into that inhuman (or all too human) trap.

ETHICAL IMPLICATIONS OF WHAT IS KNOWN ABOUT VIOLENCE

The diagnostic evaluation of the violent and/or criminal person is a weighty responsibility. The diagnostic evaluation for purposes of determining guilt, innocence, or mitigating factors in murder cases is weightier still. It can mean the difference between life and death. Remember Landrigan.

Extreme violence, even in war, does not come naturally. Soldiers must be trained to kill.

Studies of delinquents and criminals have shown that they can often conceptualize moral principles but are unable to behave in accordance with them. Bioethicists and philosophers have maintained that true autonomy, that is, the ability to make reasoned choices, requires the capacity not only to identify and reflect on basic, instinctual desires and impulses, but also to control them, and thereby to do the right thing (i.e., act virtuously). Part and parcel of that precept, of course, is the capacity not to do the wrong thing. Unfortunately, not all brains have the same capacity to judge consequences and monitor and control behavior. Some people by virtue of birth trauma, emotional neglect, brain injury, genetic predispositions to mental illness, and exposure to abuse and violence—by virtue of what the bioethicist Rawls has called “the natural lottery”—are more vulnerable to emotional lability, distortions of reality, and loss of impulse control than are those who, by chance, have drawn a better hand.

This chapter asks the question posed by bioethicists Beauchamp and Childress: Do recurrently violent people, at the time of their violent acts, “have the capacity to rationally accept, identify with, or repudiate a lower order of desire or preference in a manner that [was] independent of the manipulation of desire?” Virtue, as conceptualized by Aristotle in *Nicomachean Ethics*, may spring from an innate capacity to do the right thing. However, according to Aristotle, in order for a person to make use of this potential for moral action, this capacity must be cultivated actively. It must be exercised with the same perseverance, intensity, and training necessary to perform well athletically. Aristotle recognized that virtue also required mentors, and models for controlled behavior. The ethical question posed in this chapter is whether defendants at the time of their offenses could have acted virtuously, or whether the combination of psychopathology, frontal lobe dysfunction, and the ongoing exposure to violent models of behavior prevented them from thinking once, much less twice, about what they were doing? According to one of the authors’ subjects, “Ain’t no thinkin’ anymore . . . it’s a reaction.”

These data call for a re-examination of the outdated 19th-century definition of insanity (i.e., “. . . if the party accused was labouring (sic) under such a defect of reason, from a disease of the mind, as not to know the nature and quality of the act . . . or, if he did know it, that he did not know . . . (it) was wrong”). Queen v M’Naughton (1843). This rule limits the defense to cognitive insanity, although it is known that cognition plays a limited role in violence. In this author’s experience, some of her colleagues have taught defendants how to parrot the M’Naughton answers and thus have enabled them to go to trial and on to execution. Numerous well-trained psychiatrists have declared chronic schizophrenic and unalterably brain-damaged people to have been “sane” at the time of their offenses. Did their psychoses temporarily remit? Did their damaged frontal lobes function briefly during their frenzied act? Given the biopsychosocial impairments that are now known to underlie many criminal, violent acts, has the time not come for the neuropsychiatric professions to assist the judiciary in redefining the criteria for mitigation and insanity?

FURTHER READINGS

Alvaro-Gonzalez LC. Neuroethics: moral pathways in normal brain [Review] [Spanish]. *Rev Neurol*. 2014;58(5):225–233.

Arizona v Landrigan 176 Ariz 1, 8, 859 p2d 111 (1993) Arizona Supreme Court.

Arnedo J, Svrakic DM, Del Val C, et al. Uncovering the hidden risk architecture of the schizophrenias: confirmation in three independent genome wide association studies. *Am J Psychiatry*. 2015;172(2):139–153.

Baez S, Kanske P, Matallana D, et al. Integration of intention and outcome for moral judgment in frontotemporal dementia: brain structural signatures. *Neurodegener Dis*. 2016;16(3–4):206–217.

Bondu R, Beier S. Two of a kind? Differences and similarities of attacks in schools and in institutes of higher education. *J Interpers Violence*. 2015;30:253–271.

Brunner HG, Nelen M, Breakfield XO, Ropers HH, van Oost BA. Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. *Science*. 1993;262:578–580.

Callahan D. *The Cheating Culture: Why More Americans are Doing Wrong to Get Ahead*. Harcourt; 2004.

Courtwright D. *The Violent Land: Single Men and Social Disorder from the Frontier to the Inner City*. Harvard University Press; 1996.

Cristofori I, Zhong W, Mandoske V, et al. Brain regions influencing implicit violent attitudes: a lesion-mapping study. *J Neurosci*. 2016;36(9):2757–2768.

Cushman F. The neural bases of morality: not just where but when. *Brain*. 2014;137(4):974–975.

Decety J, Porges EC. Imagining being the agent of actions that carry different moral consequences: an fMRI study. *Neuropsychologia*. 2011;49(11):2994–3001.

Edelstein A. Dissociative identity disorder theory explaining serial murder and murderers. *J Trauma Treat*. 2015;S4:019.

Falkner AL, Lin D. Recent advances in understanding the role of the hypothalamic circuit during aggression. *Front Syst Neurosci*. 2014;8:168.

Fumagalli M, Priori A. Functional and clinical neuroanatomy of morality. *Brain*. 2012;135:2006–2021.

Gibney A. Crazy, not insane. *HBO*. 2020.

Glenn AL. The other allele: exploring the long allele of the serotonin transporter gene as a potential risk factor for psychopathy: a review of the parallels in findings. *Neurosci Biobehav Rev*. 2011;35:612–620.

Hare RD. *Psychopathy: Theory and Research*. Wiley; 1970.

Hart H, Rubia K. Neuroimaging of child abuse: a critical review. *Front Hum Neurosci*. 2012;6:52.

Hexuan L, Yi L, Guo Guang. Gene by social environmental interaction for youth delinquency and violence: thirty-nine aggression-related. *Soc Forces*. 2015;93;(3):881–903.

Hortis CA. *The Mob and The City: The Hidden History of How the Mafia Captured New York*. Prometheus Books; 2014.

Hutchings B, Mednick SA. In: Mednick SA, Schulsinger F, Higgins J, eds. *Registered Criminality in the Adoptive and Biological Parents of Registered Male Criminal Adoptees.* Elsevier/North Holland; 1974.

Ingalhalikar M, Smith A, Parker D. Sex differences in structural connectivity of the human brain. *Proc Natl Acad Sci USA*. 2014;111(2):823–828.

Lee H, Kim DW, Remedios R. Scalable control of mounting and attack by Esr1+ neurons in the ventromedial hypothalamus. *Nature*. 2014;509:627–632.

Lewis DO. From abuse to violence: psychophysiological consequences of maltreatment. *J Am Acad Child Adolesc Psychiatry.* 1992;31(3):383–391.

Lewis DO. *Guilty by Reason of Insanity*. Ballantine Publishing Group; 1998.

Lewis DO, Shanok SS, Pincus JH, Glaser GH. Violent juvenile delinquents: psychiatric, neurological, psychological and abuse factors. *J Am Acad Child Adol Psychiatry.* 1979;18:307–319.

Lewis DO, Yeager CA, Swica Y, Pincus JH, Lewis M. Objective documentation of child abuse and dissociation in 12 murderers with dissociative identity disorder. *Am J Psychiatry.* 1997;154:1703–1710.

Mark V, Ervin F. *Violence and the Brain*. Harper and Row; 1970.

Miczek KA, Meyer-Lindenberg A. *Neuroscience of Aggression*. Springer; 2014.

Moskowitz A. Dissociation and violence: a review of the literature. *Trauma Violence Abuse*. 2004;5:21–46.

Piton A, Poquet H, Redin C, et al. 20 ans après: a second mutation in MAOA identified by targeted high-throughput sequencing in a family with altered behavior and cognition. *Eur J Hum Genet*. 2014;22:776–783.

Raine A. *The Anatomy of Violence: The Biological Roots of Crime*. Pantheon Books; 2013.

Rawls JA. *Theory of Justice*. Revised Edition. Harvard University Press; 1999.

Schiltz K, Witzel JG, Bausch-Holterhoff J, Bogerts B. High prevalence of brain pathology in violent prisoners: a qualitative CT and MRI scan study. *Eur Arch Psychiatry Clin Neurosci*. 2013;263:607–616.

Stuber GD, Mason AO. Integrating optogenetic and pharmacological approaches to study neural circuit function: current applications and future directions. *Pharmacol Rev*. 2013;65:156–170.

Teicher MH. Wounds that won't heal: the neurobiology of child abuse. *Cerebrum*. 2000;2:4, Fall.

Wilson ML, Boesch C, Fruth B, et al. Lethal aggression in Pan is better explained by adaptive strategies than human impact. *Nature*. 2014;513:414–417.

Yoder KJ, Decety J. The good the bad and the just: justice sensitivity predicts neural response during moral evaluation of actions performed by others. *J Neurosci*. 2014;34(12):4161–4166.

28.3 Borderline Intellectual Functioning and Academic or Educational Problem

Frank John Ninivaggi, M.D.

This section introduces one of the most *clinically oriented* and relevant portions of this textbook: recognizing and treating borderline intellectual and adaptive functioning (BIAF). BIAF proposes that IQ and adaptive functioning are both *well below average*. This section's psychiatric focus rests on variable patient presentations—clinical expressions. Clinical medicine is the broader perspective encompassed by psychiatry to improve quality of life.

Mental abilities and adaptive performance skills being *well below average* is a hidden risk factor. When the threshold below typical and into BIAF is crossed, significant slowness, risk, and potential disability burden emerge. These important facts are known. Researchers across the world continue striving to expand knowledge and services for this population.

Borderline functioning, however, is neither currently viewed as a formal disability nor as representing clear-cut deficits in intellectual and adaptive domains. Thus, "borderline intellectual functioning" (BIF) is not a formal fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) disorder; it is a condition. DSM-5 only vaguely defines it. Yet, *the condition may indicate the probability of risk with potential harm and loss in the future. Marginal overall functioning is problematic. The chances of risk are present but can be prevented or lessened if identified and managed*.

Mindful awareness of BIF enhances the higher reliability of both remedial anticipation and building resilience. Such prevention improves cognitive reserve and disengages the trajectory for cumulative impairments to solidify along the developmental line. Screening and assessment for risk factors, especially early adversity, predispositions to vulnerability, and stressful adverse life outcomes thwart developing disorders. Identification and good clinical care reduce often-underestimated residual burden. Considerations about risk, risk reduction, and intelligent performance (not identical to measured IQ scores) are discussed since risk and risk interventions have a protective effect and open the way for support providing growth and development. Comprehensive treatments—a detailed part of this section—can take the borderline person over the marginal edge so that life functioning can improve. This approach maintains patient-centered values at the heart of quality medicine.

Protection averts loss and builds a fence around resiliency. However, the following question remains: While duly respecting the currently limited state of evidence-based literature regarding BIF, does identifying clinically significant risk and addressing it merit inclusion of the now diagnostic gray zone, BIF, in official DSM-5 diagnoses? Such inclusion would recognize the evolving knowledge base of neuroscience and psychiatry and the limitations of current diagnostic criteria.

I suggest vanguard trends in primary prevention—eliminating or reducing risk—merit broadening risk assessment by expanding diagnostic thresholds. This refinement would identify cases with incipient clinical significance. Therefore, this inclusion's broadened scope may accommodate the natural diversity inherent in a population's heterogeneity. It would not artificially exclude groupings of heterogeneous syndromes with inferred different pathophysiologic mechanisms. Recognizing that *designated cutoffs may be better understood as flexibly moving targets* than as tipping points may be more realistic. This expanded vision would reflect the enhanced developmental focus of DSM-5. The magnitude of the net benefit of this spectrum flexibility remains to be seen. It may include authorizing coverage and making possible needed clinical services. Identifying such needs and providing resources is the essence of a prevention model that may thwart, if not obviate, additional complications, including costs. Such harm reduction, therefore, is a clinically salient goal of excellent medicine. More significant strides toward precision medicine remain an ultimate vision.

The condition now termed BIF should be understood to denote BIAF. This phrase should be its proper designation. This inclusiveness reflects a realistic developmental, clinical, and functional understanding and diagnostic notation of the whole child and person. It is consistent with conceptions and definitions of intellectual disabilities (IDs) held by the American Association on Intellectual and Developmental Disabilities (AAIDD) and put forth in the World Health Organization's ICD-10 International Classification of Functioning, Disability, and Health.

Over the last decade, BIF has continued to be underrecognized. Lack of formal diagnosis makes reimbursement, qualification, and easy access to services sorely lacking. Thus, the trend of connecting BIF (has no qualification) and mild ID (has full qualification) in one grouping is now becoming more common, notably in Europe. This tendency is now prevalent because of the imprecise ability to distinguish between the two and their inescapable "overlap." Both may be developmental variants exhibiting substantial pleiotropy across intelligence boundaries. Many phenotypes—having one source with multiple outcomes—in intelligent functioning are to be expected and viewed as "the nuanced norm."

Epigenetics, molecular mechanisms regulating gene activity, is the complex interplay between genes and their products establishing the divergence of phenotypes. Studies show the tremendous diversity in the epigenome that allows for shifts in developmental trajectories. Epigenetic mechanisms are the basis of adaptation to environmental challenges. An optimistic interpretation would be that such pleiotropy could point to developmental risk pathways amenable to common therapeutic interventions. Perhaps, pleiotropy represents factors that reduce cognitive reserve or nonspecifically disrupt brain development. In either case, how phenotypic specificity occurs remains an open question.

A "borderline intellectual disability" diagnosis as part of "intellectual developmental disorders" (IDDs) would address the gaps in the DSM-5 BIF descriptions. Also, the qualification for needed services and reimbursement would follow. Untapped abilities and newly developed skills can further social participation. This is more than a human right; *it is the right to be human to the fullest extent.*

The international community has taken a striking lead in research and has proposed bold initiatives about pursuing serious investigations concerning BIF's significance. A recent summary paper from Spain in 2013 by a 19-member panel documents the Catalan association's Nabiu (ACNabiu) CONFIL project begun in 2007. This comprehensive study funded by the Department of Health of the Catalan government draws from CONFIL 1, 2, 3, and 2010 conclusions. Using evidenced-informed clinical data, it advances BIF as a priority condition demanding an international consensus. This accord comprises construct, operational criteria, screening, coding, incidence, prevalence, effective interventions, and an integrative care model centered on the individual while also providing for the family and extended social, educational, economic, occupational, and environmental considerations. This wealth of research trajectories demands further expansion.

This movement is actively growing in Europe and internationally. For example, in the United Kingdom, Angela Hassiotis and colleagues continue to work on BIF, documenting significant future risks: a range of adverse health consequences and mental health disorders. In Spain, R. Martinez-Leal, and in the Netherlands, J. Weiland, P. Nouwens, I. Woittiez, and their colleagues continue unearthing quality-of-life difficulties for persons with BIF and mild ID. This prolific work has highlighted the heretofore underestimated intersection, overlap, and shared needs of persons with BIF and mild ID. Recognition and policy inclusion have immense practical value in establishing eligibility for needed services. For example, adverse childhood events, coupled with lower cognitive profiles, surround the newly recognized prevalence of substance abuse and offending behaviors in these populations. The toll on individuals and families, plus the economic burden on society, is sobering. Therefore, risk management includes harm reduction and an untapped plethora of treatments facilitating progressive development.

INTELLECTUAL FUNCTIONING

Whether intelligence is best regarded as a single general capacity (i.e., “g”) proposed by Charles Spearman in 1904 and 1927 or as a range of separate, modular capabilities as first proposed by Thurstone in 1938 remains an open question. No consensus prevails. Contemporary theorists view intelligent cognition as an array of abilities, strengths, and weaknesses across multiple information processing and adaptive domains. The Cattell–Horn–Carroll (CHC) theory of intelligence, one of the dominant models in modern psychology, reflects this view.

The very nature of intelligence and its multiple definitions lingers in an unsettlingly intriguing flux—the ongoing subject of intensive research. The “IQ” test was developed by Alfred Binet around 1900. As far back as 1916, American psychologist Lewis M. Terman recognized the spectrum of intellectual capacities and formalized the notion of “borderline” intelligence to indicate those who not only functioned *well below average* but were also not intellectually disabled. The emergence of rapidly developing disciplines such as cognitive science and its subspecialties is devoted to exploring the meaning and pragmatics of all aspects of human cognition (e.g., conceptual, emotional, social, practical, and motivational) bears out the central importance of these undeniable epistemologic realities. The role of powerful emotions such as envy in eliciting comparisons of superiority and inferiority cannot be overlooked. Envy triggers cognitive discrepancies, distinctions, and emotional dysregulation that attest to this. Here, envy is the bitter realization that someone else has more of something good (i.e., an ideal possession or capacity) than oneself who has less and wants. The sense of its being unattainable leads to hopelessness and an intolerable bitterness needing to devalue or destroy what is perceived as unreachable.

Intelligence—acquiring and applying knowledge—is broader than cognition—knowing—considered in isolation or merely by IQ scores. Intelligence and IQ scores, though related, are different conceptions. Intelligence first denotes the capacity for awareness and discrimination of relevant and novel environmental and subjective stimuli. This information is then used in an integrated fashion in several ways:

1. To deductively make sense of complexity through analysis, synthesis, and reproducing meaning, primarily through perceptual reasoning.
2. To maintain mental and physiologic infrastructure.
3. To promote regulatory expansions in adaptive survival that includes flexible and effective performance interactively with the environment.

Language and effective communication facilitate executive functions—focused attention, selection, planning, organization, implementation, and refinement of both thinking and behavioral responses. Words become used as “signifiers” that represent something. These must be combined in the brain with the “signified” to form a meaning-imbued “sign.”

This unique human symbolic process correlates with the level of an individual’s capacity for abstract thought, which facilitates critical thinking and, in turn, executive functioning. Emotions and specific and nonspecific motivational urges give dynamic meaning to how these functions are integrated, experienced, and refined. Moreover, an individual’s intelligence does not exist in a vacuum; it is a function of interpersonal engagements throughout the life cycle. Real-life learning from experience contributes to eliciting strengths and accommodating to challenges. Respect for the mind’s life in the whole person underlies this quality-of-life consideration.

Intelligence or intellectual functioning is a complex construct (i.e., a measurable representation of a hypothesized abstraction) that does not necessarily manifest itself in a unitary fashion. It includes the following mental abilities: understanding relationships, adeptly and efficiently processing information, spontaneously applying problem-solving strategies, acting purposefully in novel, ambiguous, and complex situations, and learning from experience to refine adaptive and problem-solving skills. Intelligence is cognitive competence, the mental ability to function and perform successfully. *Intelligence is both* *intelligent cognition and intelligent adaptive performance.* A range of latent inner resources and potential capacities can be developed by learning, experience, and a sensitively facilitating environment. Subsequently, real-time capabilities to perform a demanded task emerges. Pragmatic and efficient skill performance denotes adaptive functioning.

From a scientific and academic perspective, measurements of intelligence reflecting an individual's position or standing relative to others traditionally have resided on two quantitative indices:

1. Measurement by psychometric (i.e., "measuring the mind") instruments yielding profile arrays with numeric ranges of component scores of cognitional abilities, including a composite score—full-scale IQ.
2. A statement about the effectiveness of adaptive behavior as expressed by arrays of varied performance skills.

For both criteria, cutoff scores establishing normal ranges are approximately two standard deviations (SDs) below the respective assessment instrument's mean. These scores vary by test, subgroup, and age group while considering the standard error of measurement for the specific tool used and its strengths and limitations. Current testing uses complex neuropsychological examinations that yield arrays of profiles that eschew previously held reductionist assessments and single scores in isolation. Such results give full weight to the nuances of multifaceted cognitive and adaptive processes. Intelligence is neither a static given nor a single number. It is a dynamic and relatively changing confluence of phenomena affected by chronologic, maturational, developmental, neurologic, psychological, social, and contextual factors. In BIF, real-world reasoning and judgment are *well below average*. These atypical features are apparent in discernibly poor academic performance—an essential activity in childhood.

Taking the role of intelligence into serious consideration is vital in assessing health and well-being. This understanding derives, in part, from research demonstrating that intelligence plays a crucial role, along with socioeconomic influences, not only in predicting academic success but also in promoting a variety of positive life outcomes. The pivotal role may have been underestimated or not relevant a century ago when the term "borderline intellectual" was first introduced. Today, it is self-evident. Along with the central importance of denoting intelligence as *mental ability*, including emotional intelligence as *emotional knowing* and motivational factors as real-life adaptation is essential in all considerations of intellectual functioning. The expression of intelligence and its limitations is multidimensional. It emerges from the dynamic and reciprocal engagement among intellectual abilities, adaptive behavior, health status, social participation, cultural context, individualized opportunities, and support availability. These considerations must temper, if not humble, any endeavors at reductionist measurements of this basic human functioning.

While this section's primary thrust is on the cognitive and performance role central to intact intellectual functioning, emotional conflict undoubtedly affects daily life's thinking and activities. When psychic conflict escalates, defense mechanisms are elicited. From a psychodynamic perspective, *intellectualizations* may block and constrain conflictual feelings, temporarily subduing anxiety. Avoiding and denying unruly emotional turmoil by using lofty, often specious, cognitive explanations normatively arise in adolescence as a way of managing newly emerging psychosexual impulses. If such strategies bring relief, they are carried forward and may become embedded as enduring character traits. However, those with BIF rarely become proponents of such abstruse, pedantic legalisms or present themselves as voluble, creamy intellectuals whose excessive verbosity mesmerizes the listener. Intellectualizations are cognitive obfuscations, rarely if ever, used by those with BIF. I raise these contrasts since they may sharpen attention to BIF's typical presentations, more literal and concrete, as they tend to be.

Furthermore, the increasingly recognized protective effect that intelligence affords merits its appreciation as a powerful therapeutic focus. Protective factors increase the chances of resilience, healthy coping, effective action-oriented behaviors, and positive life outcomes. Such knowledge has clinical utility by providing useful information needed to convey the patient's material, relevant to the clinical course, and comprehensive treatment planning. *Attention to IQ scores and adaptive strengths and weaknesses has provided helpers with a means of identifying persons who have developmental and other special needs. Only with this knowledge can services be offered to optimize social inclusion, equity, and personal empowerment. Moreover, these models influence research, policy, funding, access, and treatment reimbursement.* These considerations also comprise a vision aimed toward raising an individual's biopsychosocial quality of life with meaning and dignity. Cognitive and emotional capabilities pragmatically expressed as satisfying life skills in the real world are thereby enhanced.

BORDERLINE INTELLECTUAL AND ADAPTIVE FUNCTIONING AS A CLINICAL ENTITY OF SIGNIFICANCE

DSM-5, as have its previous renditions, recognizes and codes BIF as a condition or focus of attention and not as a disorder.

DSM-5 describes a mental disorder in the following way:

> "A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning. Mental disorders are usually associated with significant distress or disability in social, occupational, or other important activities." (Adapted from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, p. 14 [Copyright © 2022]. American Psychiatric Association. All Rights Reserved.)

Clinical utility is emphasized: helping to determine prognosis, treatment plans, and potential treatment outcomes, and not a need for treatment. The need for treatment is a complex clinical decision that considers symptom severity, salience (e.g., suicidal ideation), distress or mental pain, disability, and risks and benefits of available treatments. Clinical significance emphasizes that the diagnostic threshold is vitally influenced by the disturbance's degree, causing clinically significant distress or impairment in social, occupational, or other important areas of functioning at crucial developmental eras, for example, education in childhood.

By contrast, a *condition* is a focus of clinical attention that may affect the diagnosis, course, prognosis, or treatment of a patient's mental disorder. A condition draws attention to the scope of additional issues that may be encountered and affect or *impact* the individual's treatment or progress. Describing these conditions as "problems" or interferences strongly suggests some degree of distress, significant dilemma, trouble, or obstruction. Some examples of "conditions" are abuse, neglect, sexual violence, and adult antisocial behavior.

Practical utility in clinical practice is thus acknowledged, especially by definition of a condition, as it may impact clinical care and treatment progress. As far back as DSM-III (1980), the distinction between normality and psychopathology was seen as a gray area whose inexactness remains subject to continuous scrutiny and reexamination. The threshold for "caseness" and the "cut off area" concepts used to denote a diagnostic label of "mental disorder" is not yet deemed sufficient for BIF. While the real-world population abounds with persons having no apparent difficulties, individuals with BIF typically have noteworthy problems. BIAF, moreover, may exacerbate comorbid mental disorders. Yet, the psychological, educational, ID, and psychiatric literature meagerly yet valiantly strives to document this challenge convincingly. DSM-5 considers BIF worthy of inclusion since conditions such as these draw attention to a broad scope of problems encountered in clinical practice. BIF is coded both for its clinical importance and how it affects the diagnosis, course, prognosis, or treatment of the patient's co-occurring mental disorder(s). It is useful information that may affect patient care.

A consistent operational definition of BIF covering all or most situations, symptom presentation, structural pathology, pathophysiology, and etiology, along with valid and reliable studies, has been absent. This lack of data correlates directly with the absence of clear-cut measurement and defining criteria. To what extent BIF resembles a disorder, a disability, a deficit, or aspects of each remains indeterminate. How BIF shows features such as distress, dyscontrol, dysfunction, malfunction, syndromal patterning, pervasive inflexibility, gross irrationality, severity level, disadvantage, social or occupational impairment, risk vulnerability, heterogeneity, and meaningful statistical limitations and deviations are questioned, debated, and controversial. From the platform of typical or average development, does BIF lie on the continuum of developmental atypicality, for example, average skills not emerging as expected for age? *These matters remain unsettled, as does labeling all IDs as disorders rather than conditions with significant demand for support levels*. For example, could disability be construed not as resulting from a deficit in isolation but from neurocircuitry in conflict that competes to create chaos resulting in marginally less disorganized or "well below average" or "borderline" outcomes? The Easterseals community has defined disability in 2020 contemporary terms: "Disability is defined as any visible, invisible, emotional, social and educational challenges that are a part of normal, everyday life. Disability can refer to a medical diagnosis, a barrier, and/or a person's identity."

BIF is, in fact, a boundary or transitional phenomenon—at least by previously accepted criteria (DSM-IV-TR), and now by rather broad, if not somewhat ambiguous, parsimonious guidelines ("careful assessment of intellectual and adaptive functions and their discrepancies," DSM-5, p. 727). Since BIF had been viewed as the transition from the ID spectrum to the average intellectual range, and since psychometric and adaptive testing shows some excursion from statistical norms in one or more essential areas, a "disability model" is suggested. One could also argue that the individual must have some internal "disorder" that is elicited or brought forward when interacting with an environment (animate and inanimate) not quite capable of meeting and supporting atypical needs, impairing, and thus causing distress. The last consideration that shall not be elaborated here is regarding the term "mental." Mental disorders assume the implicit split in an individual between the brain and mind. This dualism is not merely traceable to 17th-century Western Cartesian dualism but is recorded as far back as the extant Sanskrit literature of India and some Eastern philosophical systems. Contemporary Western science now understands individuals from more unified, less anachronistic perspectives.

In psychology, psychometrics measures differences between individuals or an individual to an appropriate reference group. Measurement is key to scientific understanding because it puts the amounts measured into numbers for reference and comparison. For example, the range of normal IQs has been 90 to 109, following the Terman–Wechsler tradition. Sometimes it is defined as +/−1 SD from the mean of 100 (i.e., 85 to 115). Others consider average to be within 1 SD of the mean because the number of SDs from the mean or average norm has become an increasingly popular way of communicating test results. Of course, the sample of people used in these calculations should represent the testing standardization population.

BIF is a developmental limitation, not a psychiatric disorder. Not all persons with the condition of BIF or an IQ score of about 71 to 84 are impaired or dysfunctional. A *lower-than-average* IQ score is below the range of 85 to 115, with a mean of 100. This intellectual ability encompasses cognitive (i.e., analytical–conceptual–critical reasoning, verbal, and spatial) capacity. It is noteworthy to compare a DSM-5’s criterion for Learning *Disorders* (LDs) (p. 69) that specifies substantial and quantifiable low academic achievement of academic skills “*well below average”* to be of at least 1.5 SDs below the population mean for age. This score is 78 or less. Having this *well below average* score qualifies the student for the formal LD diagnosis and many services.

The condition of BIF adds that well below average IQ must be present concurrently with significant *limitations in adaptive behaviors* and functionally restricted practical performance skills across several domains of everyday living. Clinical experience is consistent with BIF occurring in the developmental period before age 18. Integrating assessment of IQ scores with an individual’s day-to-day functioning supports ecologic validity in interpretations.

Adaptive behavior domains include the following collection of conceptual, social, and practical dimensions of performance behaviors:

- Communication
- Self-care
- Social and interpersonal skills
- Home living
- Use of community resources
- Self-direction
- Health and safety
- Functional academics
- Leisure
- Work

Evaluation and treatment planning warrant serious consideration of adaptive performance criteria with particular attention to an individual's physical assets, strengths, limitations, preferences, and productive habits (e.g., exercise, sports, and other physical activities). Adaptive behaviors include a range of self-care and *instrumental activities of daily living*, notably skills needed to be reasonably successful at home and in the community. For example, personal hygiene such as bathing to minimize body odor, brushing the teeth to limit mouth odor and tooth decay, and trimming the nails create a positive appearance and minimize social signs that elicit perceptions of oddness and interpersonal distancing. Personal care includes a developing sense of wellness and mindful attention to good health and daily living activities. Uneven performance in adaptive skills, also in cognitional abilities, is common. Many distinguished contemporary psychologists whose work on intelligence testing is held in high esteem repeatedly emphasize the difficulties in adequately describing the ability levels of individuals whose IQ scores reside in the borderline category.

Therefore, BIF is a heterogeneous condition that, when dysfunctional, denotes the co-occurrence of clinically, significantly limited intellectual and adaptive functioning. Together, these deficits result in uneven and often marginal scholastic, practical, and social competence. In such situations, BIF acts as a functional neurodevelopmental disability, placing the individual at risk for failure and subsequent lowered self-esteem, dysphoria, and distress. As these children grow, the gap between their capabilities and demands from school, the workforce, and society increases. Tasks of daily living become more challenging to fulfill successfully; hence failure and blame ensue. The impact that BIAF plays is typically overlooked. Missing BIF has perplexing consequences for all involved.

Such academically fragile persons have been referred to as “slow learners” and “hard-to-teach” children, especially since they typically have difficulty understanding and *learning* average schoolwork amounts. Dysfunctional BIF denotes a general life skill’s coping disadvantage and a heightened risk for maladaptive outcomes. Yet, it should not be regarded as a static condition. While it may be true that IQs stabilize after the childhood period, adaptive skills do not. They respond to environmental enrichment and supports. Overreliance on IQ scores alone incorrectly de-emphasizes the critical importance of viewing an individual’s functioning as the complex interaction of strengths, limitations, opportunities, and appropriate supportive interventions. The entire conception of supports—a contemporary and novel construct—includes resources, strategies, and accommodations that enable people to live successfully in community settings.

The essence of BIAF, therefore, descriptively indicates *ranges of abilities that are well below average*. Clinically, those with BIF are very slow to understand and problem solve across areas. Critical thinking and executive functioning reflected in planning, anticipating, performing, and self-correcting is well below average. This may manifest in the face of apparently adequate language skills. BIF, in effect, can become a significant developmental limitation in learning and real-life adaptation. Therefore, it can contribute to *marginal abilities* or inability to succeed in major life functions such as self-care, academic achievement, and social–occupational tasks. Limited ability to perform may be due to unavailable inner resources or inability to access and efficiently use what is already psychologically present but undeveloped. Sensitive environments can respond accordingly and enhance unsteady performance.

Last, as discussed in this section, the condition of academic or educational problem is coded, usually in addition to other diagnoses, when scholastic performance is significantly impaired and becomes an outstanding feature of a patient's overall clinical presentation.

CULTURAL SENSITIVITY: A FRESH PERSPECTIVE AND AN ORIGINAL VOICE

The approach advocated here to recognize, screen, identify, and therapeutically work with persons with BIAF highlights direct clinical focus. This acuity is an astute individualized line of engagement. In part, it rests on the premise that the array and range of factors (i.e., conceptual, social, and practical performance skills, IQ scores, and details of an individual's culture, social context, and environment) are multidimensional and always varied.

Effective clinical work requires a center of gravity in clinical details with a periodic reflection on theoretical generalizations—sound clinical judgment. Acute clinical awareness and a high index of suspicion are needed to detect borderline functioning, indicating clusters of additional risk factors predisposing to limited and even impaired performance.

Recognition of cultural sensitivity in its broadest dimensions (i.e., language; styles of communication, movement, gesture, behavior; and the community environment typical of the individual's peers and cultural milieu) has been axiomatic of good clinical care. DSM-5 has updated its "Outline for Cultural Formulation Interviews" with field tests for diagnostic usefulness in many areas to obviate difficulties in communication that may bias diagnosis and treatment. In this spirit, clinicians who encounter persons with BIF and adaptive performance need to develop a higher level of sophisticated sensitivity and amplified empathy. This more intensive recognition will support a cultural formulation producing the following:

- Increased patient engagement
- A contoured comprehensive treatment plan
- Higher powered effectiveness
- Improvements in the patient's quality of life

Toward this vision, an acute, forwardly aware recognition and understanding of these cultural dimensions in the clinician and how they impact those served, including their families, schools, neighborhoods, and allied agencies, is essential. This partially self-reflective technique helps bring BIF's expression into sharper focus because BIF is typically "obscure." This instrumental perspective is the clinician's adding self-observing awareness as a primary tool to all assessments' repertoire. Evaluators and treaters undoubtedly experience some differences in perception, conception, and emotion —not themselves having BIF's personal experience.

Sharper *mindful self-awareness* provides the opportunity to recognize the transitional areas in the functioning and adaptive skills that approach theoretical ideals or "norms." This appreciation helps clinicians capture subtle impairments that may be severe enough to produce functional yet underrecognized disabilities. Effective clinical understanding may become more pragmatic because it traverses this transitional boundary between helper–helpee divides and typical versus slower or uneven mental abilities. The clinician's more intimate personal self-reflection and recognition of the individual with BIF and their distinctive profile of mental abilities, adaptive behaviors, and personality traits may be heard by the patient as an "original voice," fresh and empowering. The patient will experience the clinician not as an outside stranger but more as an inside ally, mentor, guide, and counselor. Such a motivated engagement magnifies care experience and may mitigate iatrogenic biases, enhance adherence, reinforce support, and bolster greater success chances.

BORDERLINE INTELLECTUAL FUNCTIONING

Definition

The DSM-5-TR describes BIF in Section II, "Other Conditions that may be a Focus of Clinical Attention," as follows [DSM-5 uses the older diagnostic terminology "mild intellectual disability (intellectual developmental disorder)"]:

> This category may be used when an individual’s borderline intellectual functioning is the focus of clinical attention or has an impact on the individual’s treatment or progress. Differentiating borderline intellectual functioning and mild intellectual developmental disorder (intellectual disability) requires careful assessment of intellectual and adaptive functions and their discrepancies, particularly in the presence of co-occurring disorders that may affect patient compliance with standardized testing procedures (e.g., schizophrenia or attention-deficit/hyperactivity disorder, with severe impulsivity). (Adapted from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, p. 836 [Copyright © 2022]. American Psychiatric Association. All Rights Reserved.)

DSM-5 (ICD-10) coding: R 41.83 has significantly revised BIF’s definition from that given in DSM-IV-TR in 2000. DSM-5 begins with acknowledging that BIF may be a focus of clinical attention or have an *impact* on treatment or progress. No numerical IQ scores (i.e., IQ 71 to 84) are now explicitly given as definitional criteria, as they had in DSM-IV-TR. Metrics for identification are suggested by indicating careful assessment of intellectual and adaptive functions and their discrepancies. Adaptive functions are now mentioned in the DSM-5 entry. DSM-IV-TR had contrasted BIF with mental retardation (MR)/ID. DSM-5 directly contrasts BIF with mild ID only. Co-occurring mental disorders now include attention-deficit hyperactivity disorder (ADHD) with severe impulsivity and mentions schizophrenia that may affect patient compliance with standardized testing procedures. DSM-5 does not give highly detailed coding instructions.

Since DSM-5 “requires careful assessment” of BIF (*well below average*) from mild ID (*significantly subaverage or* two or more SDs below the mean for the test), a brief discussion of this follows. Using traditional IQ scores based on Bell curves, a test norm or average is 100 as the mean; this norm extends to one SD above 100 and below 100. When one to two SDs occur, and scores range from about 70 to 85, this is considered *well below average or discernably subaverage*. At the 70 to the 55 mark or three SDs below the norm, this range is considered “significantly subaverage” or “severely, markedly, and seriously limited.” At the 55 level and below, this range is deemed substantially impaired and extremely limited. Terminology that describes levels of support needed includes (1) mild ID, intermittent supports;(2) moderate ID, limited supports; (3) severe ID, extreme supports; and (4) profound ID, pervasive supports.

Table 28.3–1
Clinical Features Associated with Borderline Intellectual Functioning

1. Mild, impairing global (not specific) learning problems
2. Well below average, limited ability
3. Uneven academic achievement and adaptive performance
4. Concrete mindedness
5. Well below average comprehension of functional relationships
6. Slowness in understanding, broad comprehension, and response
7. Poor attention, concentration, and organizational skills
8. Impaired ability to generalize and transfer learning
9. Low frustration tolerance
10. Intermittent dysphoria from perceived inadequacy
11. Low self-esteem
12. Variable social and adaptive behavior impairment
13. Task performance overcompensation (e.g., excels in physical tasks and sports)
14. Compensatory verbal fluency
15. Naiveté
16. Need for more-than-average direction

DSM-5 and others have shifted the emphasis in their current perspectives on IDs. Levels of severity are *determined by assessing adaptive functions and the levels of supports* required rather than by IQ scores. *Of great significance in the current Intellectual Disability literature is the serious consideration that mild ID characteristics may at times characterize BIF in more muted and less intense forms.* DSM-5's section on ID criteria (Table 28.3–1) discusses severity levels in the three adaptive domains: conceptual, social, and practical. The reader is referred to the DSM-5 discussion, mainly Severity Level Mild, to elaborate on such an individual's support needs, intensity, and support systems. Supports enhance functioning and avert the risk burden.

Detailed elaborations of these comprehensive systems also can be found in AAIDD texts. AAIDD has made and continues to make contributions of inestimable importance, both theoretical and practical. The enormous significance of supports is reflected in the AAIDD User Guide statement: “The implication is that whereas the life functioning of the individual with ID can improve with personalized and sustained supports, a valid diagnosis will not change.” (2012, p. 2).

History and Comparative Nosology

BIF definitions’ history and development directly reflect evolving conceptualizations of MR, now termed ID. Significantly subaverage capacities characterize ID in intellectual functioning (e.g., reasoning, learning, and problem solving), and adaptive functioning (e.g., conceptual, social, and practical skills) emerges before the age of 18 years. Efforts toward reducing stigmatization and unfair segregation resulted in a federal statute in the United States (Public law 111–256, Rosa’s Law, 2010) replacing the term “mental retardation” with “intellectual disability.” Whereas BIF is a borderline cognitive and adaptive limitation, ID represents an array of mild, marked, severe, and substantial developmental disabilities below those of peers that include both intellectual and adaptive functioning deficits in conceptual, social, and practical domains. The DSM-5 (2013) definition correlates with that of the AAIDD 2010 definition except that the DSM-5 categorical definition includes the terms “disorder” and “deficit.” AAIDD instead uses the phrase “significant limitations” to denote and describe features of intellectual “disability.”

Of note is that ICD-11 (to be released in 2021) proposes to use the phrase “intellectual developmental disorder” to indicate ID as a disorder that involves impaired brain functioning early in life. The tentative name and coding are ICD-11: 6A00.Z, “Disorders of intellectual development, unspecified, Intellectual disability.” IDD is the term that characterizes persons with significant impairment of cognitive functions associated with learning limitations, adaptive behaviors, and skills.

Since about 1973, in addition to an IQ score of between 70 and 75 up to 85 for BIF and onset before 18 years of age, classification schemes have placed significantly greater emphasis on functional adaptive skills, the elimination of the notion of borderline MR, and including the construct of BIF.

The following is a brief historical review of the coding and terminologic vicissitudes of BIF in the DSM. DSM-I (1952–1967), which correlated with ICD-6 of 1948, coding in Section V reads: "Code no. 60.0, mental deficiency (familial or hereditary), mild, standard nomenclature: 000-x901, international list: 325.3." DSM-II (1968), which correlated with ICD-8, 1965, used the code and phrase "310. borderline mental retardation." The name "borderline mental retardation" was eliminated in 1973 (official reclassification by the American Association on Mental Deficiency, now AAIDD). It was changed to "borderline intellectual functioning" at that time. DSM-III (1980), which followed ICD-9, 1975–1978 (ICD-9 clinical modification to increase the resolution of diagnoses and procedures, an adaption created by the U.S. National Center for Health Statistics, became official in January 1979 as "ICD-9-CM"). DSM-III first introduced the five-axis system and identified BIF on Axis I coded as "V62.89." DSM-III-R (1987), still following ICD-9-CM, now shifted BIF from Axis I to Axis II and coded it as V40.00. DSM-IV (1994), working closely with ICD-10, 1992, maintained BIF on Axis II and coded it as V62.89. DSM-IV-TR (2000) maintained previous definitions, classifications, and coding. DSM-5 (2013), now intimately aligned with ICD-10, revised its description of BIF, eliminated all axes systems, and codes BIF as "R 41.83."

Rapidly changing paradigms within the ID field have adopted an interaction-based ecologic perspective with an evolving vocabulary. For example, with the AAIDD's 2010 publication of its 11th edition, *Intellectual Disability*, the nature and amount of support (i.e., frequency, daily support time, and type of support) that a person needs to participate fully in life is defined. As of October 1, 2014, DSM-5 codes ID based on both intellectual functioning deficits and adaptive competencies, but specifiers will be based on levels of severity defined by adaptive functioning parameters. Although DSM-5 criteria remain broad, they state that identifying BIF requires “careful assessment of intellectual and adaptive functions and their discrepancies.”

Etiology and Epidemiology

As with other conditions that reflect intellectual and adaptive limitations, genetic, hereditary, congenital, health, environmental, interpersonal, sociocultural, and educational opportunities have been implicated as etiologic factors. For BIF, however, no factor in isolation has been shown to be definitively causative. Prenatal, perinatal, and postnatal medical conditions must be identified since they may provide more explicit explanatory and remedial information. With reference to a normal distribution or bell curve of psychometric intelligence on the general population, IQ scores in the borderline range (about 70 to 85) occur from between one and two SDs below the mean or average of 100, approximately 13.6% of the time. BIF’s clinical condition is strongly suggested only when noteworthy limitations or impairments in adaptive functioning accompany this finding. These considerations must be viewed in the cultural, educational, environmental, and economic context of a person’s life. Such a multimodal perspective adds needed clinical flexibility to assessments and determinations having significant consequences.

Identification and Clinical Features

Contemporary views stress the presence of *well below average* measured intelligence, limited adaptive functioning, and support requirement needs as the broadest dimensions within which dysfunction and needs arise. *Limitations and well below average performance may not be global and pervasive; areas of strength may be interspersed with marginal functioning*. Identification of BIF integrates clinical, academic, and common sense factors. There has been a shift from reliance on composite or full-scale IQ scores to an analysis of multiple subtests of cognitive functions together with adaptive performance. Psychological and neuropsychological test assessments provide cognitive functioning scores, IQ profiles, and comments on emotional and behavioral functioning. These indicators offer objective measurements, helping to position the student in areas of strengths and weaknesses.

Specific cognitive abilities predict academic achievement and can be one part of evidence-based data for prescriptive, comprehensive interventions and individualized remediation. Because information regarding academic performance and testing, including psychometric testing, is typically available for persons younger than 18 years of age, BIF may be recognized more easily in the school-aged population. Initially, assessing adaptive functioning is done through clinical screening and thorough history taking. If needed, formal specialized testing may be performed. Intelligence testing and the conceptual, social, and practical life skills aspects of adaptive testing should be performed in a nonbiased manner with due attention to cultural contexts.

Fairness in adequately identifying and not misconstruing the presence of BIF remains essential in decision making. Attention to socioeconomic status (SES) and access to reasonable educational and school programs must be noted. This inclusion in case formulations underscores a culturally informed assessment. Motivational factors, effort, and fatigue affect test taking and affect real-time performance in life skills. The effects of psychotropic medications on performance must also be considered. Information from testing is correlated with all aspects of the individual’s life functioning for conclusions to be rendered reasonably and reflect diligent clinical judgment. *For example, those with BIF are recognized implicitly as requiring more coaching early on in grade school. When this is successful, verbal skills are raised; hence, they may test at near average verbal levels, both empirically and on test scores. In such cases, verbal comprehension may eclipse well below average perceptual reasoning capacities, working memory, and processing speed deficits. A primary consideration of the presence of BIF thus may be missed.* A veneer of language may mask practical performance problem-solving abilities.

Common sense skills commensurate with chronologic age and educational level may be lacking.

Table 28.3–1 summarizes some essential clinical features associated with coding the condition of BIF. Familiarity with the basics of child and life cycle development, quickly gleaned from readily available texts on child development, offers inestimable pragmatic, clinical assistance in orienting the evaluator concerning presentations reflective of typical versus *well below average performance*. Clinical assessments and the psychiatric interview are based on firsthand observation. Seeing not only what a person can do but also *how they perform* is crucial for adaptive performance. Direct and detailed observation—*the patient's clinical expression*—is indispensable to the identification. The attentive clinician experiences that the individual's cognitional process—receptive and expressive—is slow-moving and lumbering.

These variable manifestations may show some of the following: mild but impairing global (i.e., not specific or isolated) learning problems; learning how to learn—the brain's "search engine"—appears limited; understanding is slow-moving; slow working memory; difficulty grasping functional relationships; once learned, materially performing an activity or task is intact, and may even be highly skilled; uneven academic and other skills performance; tendency to concrete-mindedness; perceptual reasoning (e.g., problem solving) may be impaired; some slowness in response time; difficulty with simultaneous management of multiple tasks; notable difficulties in understanding and learning, which may be displayed as task avoidance, insufficient attention, concentration, and organizational skills; difficulties integrating new knowledge with previously learned information and applying it to novel problem solving; impaired capacity to generalize learned information and skills; low frustration tolerance; intermittent irritability, dysphoria, and distress, mostly from poor achievement and perceived personal inadequacies; low self-esteem and low motivation; social awkwardness; mild-to-moderate social skills, although social interaction is sought (i.e., not apathetic); neither hypokinetic nor hyperkinetic motor activity; poor common sense; naivete; potential for being bullied and for victimization; and a need for more-than-average direction and supervision in general, given chronologic and developmental status. Procedural memory and habitual task performance from past practice and repetition are typically intact. Uneven performance within each and between intellectual and adaptive domains may occur for several reasons, one of which is the individual's own compensatory processes. Verbal skills testing as average or better may compensate for *well below average* (i.e., *one to two SDs below the mean for the test*) perceptual reasoning and problem-solving abilities. Clinically, this may mask the real presence of BIF.

When a number of the aforementioned clinical features are moderately pervasive, dysfunction is inevitable. All these factors then contribute to a less-than-average rate of learning and understanding, frustration in an academic environment that expects average or greater-than-average performance and compliance, educational and social failure, some difficulties with activities of daily living, and a greater chance of emotional distress and behavioral dysfunction.

Other sources that support this diagnosis come from standardized intelligence tests (e.g., Wechsler Intelligence Scale for Children, Stanford–Binet test, and Kaufman Assessment Battery for Children), adaptive behavior scales, and school psychological and academic testing and achievement reports. Working memory seems to be significantly impaired. When these clinical features, integrated psychometric profiles of multiple cognitive functions and capacity measures, and adaptive and support needs are associated with an individual's biopsychosocial and academic dysfunction, BIF may be coded. If it is clinically deemed the principal or a significant cause of the individual's problems or visit, the terms "principal" or "reason for visit" may be indicated in parentheses. Although a strong presumption is present, such uncertainty can be shown by recording "(provisional)" after the code if not enough information is available.

Differential Diagnosis

The diagnosis of BIAF is an orphan entity. Rarely do clinicians consider it when faced with its presentation. Thus, a high index of clinical suspicion needs BIF's signs and symptoms in mind. One must first rule out substantially significant forms of IDs, which can easily be detected by standardized intelligence tests, "adaptive skills" profiles, history of onset in the developmental period, and overall performance in various contexts. Table 28.3–2 summarizes the differential diagnosis of the clinical condition of BIF. A pivotal diagnostic question with significant treatment implications is whether the typically diagnosed disorders that often mask or are mistaken for BIF, such as ADHD, attention-deficit disorder (ADD), and LDs, are, in fact, present or absent. It is also essential to designate whether BIF is primary and comorbid or comorbid and secondary to the individual's presenting and fundamental borderline intellectual problems if comorbidity exists. Careful and diligent history from family and collateral sources such as school reports, including psychological testing, are indispensable.

BIF is typically underreported. The most frequently diagnosed disorders that mask, overshadow, or may be mistaken for BIF are *ADHD, ADD*, and *LDs*. Common features of all these are focusing attention and sustaining concentration that prevents adequate learning from occurring. However, there may be diagnostic overlap and definitional variations according to variable criteria used. ADHD, ADD, and LDs are often, though not necessarily, found in those with average or higher IQ scores (i.e., 85 and above).

Table 28.3–2
Differential Diagnosis of the Clinical Condition of Borderline Intellectual Functioning

1. Attention-deficit hyperactivity disorder; attention-deficit disorder
2. Learning disorders
3. Autistic spectrum disorder
4. Nonverbal learning disabilities
5. Conduct disorder
6. Antisocial personality disorder
7. Intellectual disability (previously termed "mental retardation")

In *ADHD and ADD*, impulsivity and behavioral inhibition impairments are primary, along with executive functioning impairments. Many psychologists emphasize that difficulties reside in *performing* academic skills, *not in learning* them. In ADHD, impulsive behaviors typically pose significantly disruptive problems and interferences with adaptation at home, school, and other contexts. In ADD and ADHD, academic functioning is notably affected by an impulsive need to rapidly get through the material, which is a deficit intimately linked to poor behavioral inhibition. Impulsivity leading to immediate action also prevents sufficient encoding into working memory. In BIF, academic learning is primarily limited since overall cognitive processing is well below average and concrete oriented, emphasizing physicality, not abstractions. Adaptive performance may match these limitations, including limited communication, social participation, or occupational functioning. In ADHD, organizing and skillfully executing a task is a clear-cut-weakness; in BIF, once an activity is learned, it is typically performed soundly; execution may be slow yet persistent and organized. Some studies indicate that on the WISC-IV Factor Index, Wechsler Perceptual Reasoning (Gv-Gf), those with ADHD attain average or better scores, unlike those with BIF. They score lower on solving novel nonverbal problems and fluid reasoning tasks.

Specific learning disorders (SLD) (present in up to 15% of the U.S. population) may manifest, not global, but in a specific fashion (e.g., reading, writing, mathematics, or combinations of these) and in those school-aged children whose learning and using academic skills are compromised by more modular, primarily cognitive, information processing irregularities. Samuel A. Kirk, an American psychologist, was the first to identify the construct and coin the term “learning disabilities” in 1963. DSM-5 implies IQ scores to be greater than about 70, thereby ruling out ID to better account for these LDs. These LDs may also occur in “gifted” individuals, those with IQ scores roughly above 130. DSM-5 also terms the persistent difficulties defined by LD as “disorders” rather than “disabilities.” They are localized to academic skills that are substantially and quantifiably below those expected for age. They cause significant interference with educational or occupational performance or with activities of daily living. Details in memory function, mostly encoding and retrieval, play a significant role.

The entire construct of LDs has undergone radical revisions in the last several years, and its determination is dependent on multiple complex factors. A comprehensive process that includes developmental history, medical history, psychological and achievement testing, school reports, language and communication assessments, and especially the response to intervention (RTI) data are used to determine eligibility. LDs (commonly referred to as "learning disabilities" in educational contexts) qualify an individual for special education services and an individual education plan (IEP). Also, whereas ADHD and ADD may require psychotropic medication as one part of a comprehensive management plan that includes psychosocial therapies, behavior modification, specialized academic programming, and parenting strategies, LDs and BIF do not include a medication component as part of their comprehensive treatment plan. In specific LDs, *specific* academic skills performance is well below average for age (i.e., low achievement scores ranging from 1.0 to 2.5 SD below the population mean for age). Working memory and processing speed may be specifically impaired. In BIF, well below intellectual and adaptive functioning impacts functioning *as a whole*, not merely on specific or isolated academic skills, and so affects interventions and overall progress.

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by persistent deficits in social communication and social interaction across multiple contexts together with restricted, repetitive patterns of behavior, interests, or activities. Reciprocity is severely constrained; pleasurable preoccupations rather than distressing obsessions are typical. DSM-5 codes three levels of severity. In higher functioning ASD (Level 1 severity), IQ scores are average or higher, and visuospatial competence may be high, unlike in BIF. Academic performance usually is impaired because of severe social skills deficits' interference with learning, although social approach—awkward as it is—is attempted. Although eye contact is typically impaired, there is no social or emotional indifference.

Emotions can be powerful. Anxiety and dysphoria may emerge from social rejection and isolation. Those with ASD may appear dysmorphic and “quirky” because of difficulties modulating affect in response to social cues and irregularities in prosody and one-sided, egocentric, long-winded conversational styles. ASD Level 2 is less impaired but requires substantial support, while Level 3 requires very substantial support. In contrast, children with BIF appear typical, often quiet and exhibit modulated ranges of affect, unlike the more constricted facial expression reflecting impaired socioemotional reciprocity and perspective-taking exhibited by those with ASD. Those with BIF often show socially meaningful pragmatic language and developing empathetic perspectives, unlike those with ASD.

Nonverbal learning disabilities (NVLDs), a neuropsychological and academic classification and not a DSM disorder, are similar in presentation to higher functioning ASD, and both ASD and NVLD overlap in their clinical features. NVLD has also been called “right hemisphere deficit” and “developmental socioemotional processing disorder.” Prefrontal and frontal lobe executive functioning impairments manifest as cognitive, behavioral, and emotional dysregulation. Individuals with NVLD are characterized by less severe social skills deficits (i.e., “dyssemia”) than those found in ASD. Academic problems, for example, in mathematics and reading comprehension, take center stage. NVLD’s prominent features include overattention to details and missing main themes; intact linear cognitive processing but impaired gestalt comprehension; visuospatial perceptual deficits, especially left-side inattention and neglect; impaired pragmatic language use; and speech-related rhythm, volume, and prosody irregularities. Attentional style is desultory and wandering, in contrast to distractingly impulsive, as found in ADHD and ADD. Neuropsychological testing confirms clinical impressions and may show performance IQ scores, especially low scores on coding and block design subtests (i.e., poor visuospatial functioning), to be below verbal IQ scores, which are typically average. The capacity for abstract thinking is dysfunctional, regardless of average or higher IQ scores. Children with NVLD appear more motorically passive and not overactive. Anxiety and depression, rather than disruptive behaviors, are more typical. Individuals with NVLD typically qualify for special education services more easily than those with BIF.

Disruptive, impulse-control, and conduct disorders (CDs) typically draw robust clinical attention and require treatment. CD is characterized by aggression and bullying to people and animals, destruction of property, deceitfulness or theft, and serious violation of rules. These behaviors in children and adolescents are often associated with school failure and an apparent impairment of comprehension of academic and social standards. BIF may be comorbid and contribute to the impaired ability to understand and comply with expected norms in youth with CD. Antisocial personality disorder (APD) is a pervasive pattern of disregard for and violation of others' rights that begins in childhood or early adolescence and continues into adulthood. As in CD in children, BIF can be associated with APD and impaired cognitive skills in adults, and its comorbidity, if present, needs to be identified. Any underlying psychotic disorder, whether endogenous or exogenously produced (e.g., illicit substance use), must be ruled out.

Also, it must be recognized that BIF, like other conditions found in the juvenile population, may be, and usually is, comorbid with (although usually not mistaken for) several different psychiatric disorders. Post-traumatic stress disorder (PTSD) associated with neglect, maltreatment, trauma, or abuse is now commonly encountered in children whose poor academic performance is primarily affected by anxiety, agitation, and preoccupation with traumas. Primary mood instability and cycling along with a driven, distracting, and hyperkinetic motor component (cognitive, speech, and physical activity) characterize mood disorder or bipolar disorder in children; "disruptive mood dysregulation disorder" with severe and recurrent verbal or behavioral temper outbursts and persistently (i.e., not episodic) irritable, peevish, and habitual angry mood may be seen in 7 to 18 year olds. Substance-related disorders such as cannabis use and other drugs that impair learning and memory must also be considered. Clear differentiation and targeted treatments are needed. Also, BIF has no correlation with borderline personality disorder.

Sluggish cognitive tempo (SCT) is a theoretical construct being investigated as a presentation of attentional cognitive impairment. It is not a formal diagnostic entity and has no standardized criteria. Most researchers, however, identify its most salient symptoms as including the following: daydreaming, trouble staying awake and alert, mentally foggy and easily confused, staring a lot, spacey with the mind seeming elsewhere, lethargic, underactive, slow-moving and sluggish, inability to process questions or explanations accurately, drowsy and sleepy appearance, apathetic and withdrawn, lost in thoughts, slow to complete tasks, and lacks initiative with fading effort. Although correlated with ADD and ADHD, it is thought to both be comorbid and possibly a different condition. While ADHD is considered a disorder of executive functions and externalizing correlates (i.e., disruptive behaviors), SCT is not. SCT is correlated with internalizing features such as anxiety, depression, and social withdrawal. Some have suggested renaming the condition "concentration deficit disorder," "developmental concentration disorder," or "focused attention disorder." Although problems in adaptive performance are evident, there are no data available regarding IQ scores, and, therefore, it would be premature to make correlations between SCT and BIF. The clinician must also carefully differentiate those with temperamental styles that are avoidant, reticent, and retiring (e.g., typically discernable by 12 months in infancy and onward) and schizoid personality styles from presentations having cognitive sluggishness as an isolated or primary feature.

Course and Prognosis

Specific studies showing the longitudinal course and prognosis of those with BIF continue as an evolving process. However, they detail the quality of life, risk, and potential impairments. Allied literature describing those with mild cognitive limitations reveals consistent negative trends. Studies show that the overall quality of life may be negatively affected when the condition is unrecognized, and no supports are given. *Quality of life* is reflected in the following domains of human functioning: social relationships, family relations, employment, autonomy, recreation and leisure, personal competence, independent living, community integration, social status, health, finances, future parenting, and subjective well-being. With appropriate habilitation, remediation, guidance, and support, risk reduction may be achieved, and good quality of life is realized. Early identification and treatment are essential. A naturally facilitating environment that elicits strengths is found in a good family and reinforced with a mentor, a supportive school, and social community. These platforms support positive motivation and a sound work ethic—all of which continue to encourage success.

A 13-year-old boy was referred to a partial hospital program because of anger outbursts, temper tantrums, and moodiness in the home. He was in a regular classroom setting, and his grades were marginal. He appeared passive, but his overall demeanor brightened, and his performance improved when given adult attention and guidance in school and at home. Along with the psychiatric assessment, evaluation included a pediatric medical examination, blood work, and urine screening for illicit drugs. School psychological testing was reviewed. Results were unremarkable. There was no evidence to support diagnoses of ADD, ADHD, or LDs. There were no medical problems. However, verbal IQ was 80, performance IQ was 75, and adaptive skills proficiency was several years below that expected for age. A primary diagnosis of BIF was made. Treatment included intensive emphasis on special help with resource room and accommodations in school, psychoeducation for the family, and brief individual, group, and behavior modification for the patient. Mood and behavior at home significantly improved, self-esteem was enhanced, and school performance was strengthened academically and socially.

An 11-year-old boy was admitted to a residential facility to evaluate further, treat, and stabilize a long history of ADHD and intermittent explosive disorder. He had been on dextroamphetamine (Adderall), 20 mg in the morning and 20 mg at noon, and lithium (Eskalith, Lithobid) 300 mg three times daily. His clinical presentation showed a significantly small-for-age, underweight boy with a moderate degree of facial tics and grimaces, stereotyped behaviors (e.g., picking at the skin of the hands, arms, and legs), agitation, and constricted affect. Medical examination and blood work were normal, although tics were noted. A comprehensive assessment revealed a 3-year history of current behaviors, with no previous special education resources. However, grades had been consistently poor to failing, and chronic family strife because of the patient's intermittent irritability, oppositionality, and noncompliance at home, homework, bedtime, and difficulty getting up in the morning. Evaluation and treatment included a therapeutic trial of the gradual tapering to zero of all psychotropic medications in the context of a structured therapeutic milieu. After 3 months, the patient gained 10 lb, had considerably fewer facial tics, and had few to no explosive episodes. Sleep patterns began to normalize.

Stereotyped behaviors and moderate impulsivity remained, although school performance (behavior and academics) improved. A trial of 0.5 mg of risperidone (Risperdal) twice daily was begun to target residual agitation, impulsivity, stereotypies, and tics. Because of weight gain, risperidone was gradually cross-tapered to a small amount of aripiprazole and monitored carefully. After 2 months, these signs and symptoms resolved, leaving only mild impulsivity, which was easily redirected in the structured environment. During this stabilization period, psychological testing was done and showed a verbal IQ of 78, a performance IQ of 76, and an adaptive skills profile well below chronologic age expectations. A primary diagnosis of BIF was made. Because no specific learning disabilities were noted, his global school learning problems were addressed in a classroom with general, special education accommodations. Furthermore, psychoeducation for the student and his family, along with family therapy, continued to help resolve much of the chronic strain between parents and the student, resulting from years of mutual frustration and apparent failure.

Treatment

Overview. BIF is a clinical condition showing significant risk vulnerability. Since few to no treatment guidelines exist, this section details a comprehensive summary. Risk management addresses the *whole person* who has the condition. Careful case formulations with culturally informed assessments and data from IQ and adaptive functions optimize quality treatment. Robust clinical and educational interventions follow from these. A whole-person focus sees the individual's inner resources—the terrain or field of resilience—as the platform upon which interventions depend. Although standardized techniques targeting those with borderline cognitive and adaptive functioning are not yet established, a *multimodal management approach* is best. Treatment is risk intervention and aims for harm reduction in BIF. Improvements beyond the marginal edge reinforce resilience. This comprehensive strategy centers on the individual, the family, and the school environment. As one part of an effective treatment plan, psychosocial therapies foster aptitudes eliciting innate resourcefulness—potentials for intelligent performance—and improved emotional resiliency are essential.

Psychotherapy and counseling, suitable to the developmental level and individual's receptivity, often help long-standing poor self-image. Many of those with BIAF feel they "do not fit in" with others. Many persons with borderline functioning are implicitly demoralized because of their typically undiagnosed, *well below average* condition. The emotional toll on those with BIF correlates with low self-esteem, low self-confidence, and a puzzling quality of alienation. Helping these children, adolescents, and adults identify their given strengths and amplify their inner resources is essential. Parents also need to come to terms with their child's BIF and *well below average* functioning. Last, when the adult children of those with BIF finally realize their parent or caregiver was limited, insight and renewed compassion rather than frustration and resentment for unexplained less-than-optimal overall care can bring resolution and acceptance. Family work is essential. Psychoeducation and psychosocial management strategies to help parents understand, process, problem solve, and willingly reach out to professionals, notably educators and social service systems, round out multimodal care. Transition and long-range planning must be a part of all care models. Teaching life skills can be understood and internalized as early as age 12.

Initial Assessments. Since children, adolescents, and adults present with multiple problems, often with multiple psychotropic medications, careful medication assessments are crucial. Psychotropic medications have indications for psychiatric disorders or the symptomatic features of disorders such as anxiety, agitation, irritability, impulsivity, and mood instability. There is no explicit indication for BIF. Well-chosen, low doses of an appropriate medication may target distressing signs and symptoms and can support less dysphoric attitudes for educational, psychosocial, and behavioral interventions.

Clinical presentations may be confounded with iatrogenic medication side effects such as fatigue, sluggishness, cognitive dulling, lethargy, irritability, mood lability, anxiety, poor appetite, insomnia, tics, skin picking, and trichotillomania. These potential culprits pose distractions to evaluators that, when eliminated, make more transparent the child’s actual profile of signs, symptoms, and cognitive baseline. Testing results are less adversely affected. The judicious use of psychotropic medications in the juvenile population is part of good clinical practice.

Therapeutic diversification—thoughtfully constructed, monitored, and refined over time—nuanced to identify and address individual needs is outstanding patient care. *A restructured individualized academic program is the framework of all interventions. The role of the school, whatever its context, is pivotal.* BIF’s condition rarely qualifies an individual for special education services or an IEP—a current shortcoming in official policies. Such plans outline specific and measurable educational and behavioral goals with quarterly and annually addressed objectives. Instead, a strong, compelling, and reasonable argument for an appropriate level of individualized academic intervention such as Section 504 remedial assistance can be requested to school boards. Acceptance may be successful.

After careful delineation of cognitive, emotional, social, academic, and adaptive strengths and limitations, comprehensive strategies should address the individual's specific needs and the educational supports required. Problems and interferences with adaptive activities of living are significant. The entire treatment plan should be user-friendly and not aversive in any way to the child, the family, or the academic environment. The treatment-determining context may be school based, within the family, or in a residential, independent living, or group home setting. Particular attention to significant developmental and other life transitions should be highlighted. This attention to lifelong needs includes early considerations of pragmatic vocational planning. Students graduating from high school, for example, might enter a community college transitional program specifically devoted to addressing the adaptive skills needed in adult life. Vocational training for adulthood is part of transition planning. Although interventions have emphasized attention to the intellectual domain, using the adaptive domains—including the physical body's role in kinesthetic activities such as sports and manual labor (e.g., prevocational exposure)—should not be underestimated. Procedural memory and forms of implicit learning typically remain potential strengths. Interpersonal communication skills, time management, organizational skills, career exploration, and workplace readiness are taught in such programs. For older adolescents and younger adults, referrals to bureaus of rehabilitation and vocational services, which may be associated with local and state Departments of Social Services, are appropriate vocational considerations. Treatment planning of resources and strategies must be reviewed regularly because needs and circumstances will change over time. The family and caregivers' role pervade school-based remediation, psychosocial therapies, behavior modification, and psychoeducation.

School-based remediation efforts in the developmental period up to age 18 years old are fundamental. These core interventions identify specific learning abilities, weaknesses, capabilities, adaptive skills, strengths, and limitations. Remediation first addresses increased overall academic and adaptive (i.e., conceptual, social, and practical) skills. Areas to strengthen include verbal comprehension and language use, basic problem-solving, and organizational strategies. The adaptive features of conceptual intelligence include reading, writing, mathematics, time awareness, and issues around using money. Adaptive social skills include teaching cues, communication, reciprocity, and situational judgment. Adaptive practical intelligence covers personal care, activities of daily living, and the instrumental activities of daily living—all of which are listed at the beginning of this section under Adaptive Behavior Domains. ID specifiers are based on the level of supports needed for adaptive functioning in three domains. The *conceptual* level in mild ID shows impairment in abstract thinking, executive functions, and the functional use of academic skills such as reading and money management. A concrete approach to problems and solutions compared to age-mates is evident. The *social* domain shows more concrete or immature than expected for age communication and social interactions with a limited understanding of risk in social situations. In the *practical* domain, personal care may be sufficient or need help. However, complex living tasks—the instrumental activities of daily living—grocery shopping, transportation, food preparation, and money management, require support and supervision. BIF and mild ID support needs overlap.

Moreover, *accommodations*, which are adjustments to the school program, do not change the curriculum level or performance criteria. They permit the student to use it more efficiently. Various forms of assistive technology have merit. *Modifications*, by contrast, change curriculum goals and performance criteria. Teaching adaptations and innovative workarounds to tasks must be individualized to afford the student reasonable participation and chances of success, especially since individuals perceive, interpret, and respond differently to new information. For example, public school programs in charter school settings have been launched. They include modified curriculums considering the clinical features associated with BIF, customized educational course plans, daily assessments with ongoing plan revisions, and strong parental involvement. A treatment program with appropriate educational, behavioral, psychotherapeutic, counseling, and social interventions is then constructed. Since novelty and ambiguity may be experienced as stressors, learning is significantly enhanced when teaching styles change from relying on the individual’s gathering, inferring, and generalizing information to styles that are more hands-on. Students benefit significantly from having a dedicated “go-to” person who can act as a personal mentor, guidance counselor, and mediator. All these acts as “stress inoculation,” which is preparation by the exposure eliciting stretching to go above comfort zone reactions, learned to be controllable. Also, if psychotropic medication is deemed useful to boost attentional focus and reduce hyperarousability, FDA-approved guanfacine, extended release (ER) in from 1 to 3 mg by mouth once in the morning is helpful. Besides monitoring for baseline blood pressure and periodic checks, guanfacine ER has no appreciable side effects that the group of stimulant medications poses.

Psychosocial therapies focusing on the individual are essential. Various psychotherapy forms in the broadest sense, including counseling, guidance, mentoring, coaching, and support, may enhance greater social competence by strengthening social skills, especially communication, comprehension, and self-advocacy. Whatever the therapy style, feedback in the therapeutic conversation helps the therapist see where the patient is at the moment. This feedback helps the therapist structure questions and responses that “fit” the patient’s ability to understand. These communicational guidelines help strengthen executive functions, which halo out and modulate the tendencies for impulsive choice—that is, decisions resulting in small, immediate rewards despite long-term negative consequences. A large part of this approach focuses on *risk reduction* because studies show that BIF is associated with *loneliness and poor social functioning, increased rates of psychiatric diagnoses, and drug use.* Recognizing boredom and redirecting energies are skills taught through self-management techniques that target motivation, initial encoding competence, and sequencing. This approach also imparts strategies modeling learning in a step-by-step fashion and initially learned information in as many ways as possible to enhance generalization aptitudes.

The decisive role of helping children make sense of emotions and innovate their emotional intelligence has been underestimated in the past. We now know it can be instrumental in advancing their global performance skills. Precision medicine and effective interventions rest on developing emotional intelligence. Attention to social competence's vital role as allied with cognition and enhanced symbol formation in effective overall life functioning should not be underestimated. Its roots, in part, can be seen in the pioneering work of such distinguished psychoanalysts as Melanie Klein, who, as early as the first half of the 20th century, advanced important theoretical directions that underscored the significant interplay of cognition, symbol formation, social adaptation, and the reciprocal emotional influences within the parent–child relationship in clinical applications. Contemporary psychotherapies are pragmatic and solution focused. Therapists use various theoretical models and tailor techniques to real-time needs, interests, and patients' and families' responsiveness. As with other conditions that affect learning and successful interpersonal and social adaptation, effective contemporary methods built on precedents from the past must now also use multidisciplinary team approaches.

Behavior modification, especially useful in group and academic settings, has an essential place in comprehensive interventions. Whether the psychiatrist decides to work with the patient directly or in a team with other clinicians, the psychotherapeutic and behavioral components can be a powerful adjunct in promoting change. The child's quality of life in the extended environment expands.

An essential part of team treatment is the *family*: identifying its specific needs and offering pragmatic supports. The importance of enlisting the family as an active collaborator in psychotherapies, social skills, school programs, and allied interventions cannot be overstated. Recent meta-analyses show that family behavioral therapy and parent management training have significant positive outcomes. Also, the child or adolescent, as a family member, should be offered an explicit collaborating role. A specified member of the treatment team, for example, should construct setting up benchmarks for future success in a coordinated fashion with the child.

Psychoeducation, which has a protective effect on many levels, denotes explaining the meaning and significance of the diagnosis and treatments of BIAF to families, therapists, and educators. BIF is typically underreported and unfamiliar to most professionals. Thus, knowledgeable evaluators must explain in a tactful and detailed way to caregivers of children, adolescents, and older adults its meaning and the practical limitations. Identifying appropriate resources and regularly applying them makes the potential for improvement significant.

Therefore, stressing to parents that their direct, *empathetic hands-on attentive care and guidance* are fundamental to their child's needs over time cannot be overemphasized. Clearly communicating this explanatory, attributional model is an essential task for the psychiatrist. Studies have shown parenting skills emphasizing genuine warmth and responsiveness augment adaptive outcomes. Because children with BIF are usually emotionally sensitive and responsive, especially to nonverbal communications, positive perceptions of genuine helpfulness from all caregivers effectively facilitate cooperation and change. Highlighting and reinforcing all indications of effort bolsters motivation and self-esteem. Motivation may be an underestimated explanatory factor for the question of why individuals "grow" at different rates. Motivation, more than just intention, is the energetic desire or *effortful willingness* to take action toward or away from a goal. Its emotional side is arousal. Its cognitive side entails persistence toward achieving despite challenges. Motivated performance goals are well defined and made pleasurable. The emphasis on motivation is willingness. The focus in the effort is an exertion to succeed. Both complement one another.

In working with those with BIF, *motivational messages* are communications sensitive to the developmental level, needs, "questions," requirements for guidance through transitions, self-regulation, and containment. Whether implicit or explicit, such messaging becomes sufficiently powered when delivered with enthusiasm, hopefulness, and tempered optimism. This feedback to the student provides the scaffolding to support stepwise improvement rather than unrealistic perfection. A fundamental strategy for fostering the sustainability of desired behaviors is to call attention to them as they occur—“contingent positive reinforcement.” Pragmatic strategies emphasize “build first, refine after.”

All comorbid conditions must also be addressed in the comprehensive treatment plan. Behavior modification strategies (coordinated with techniques used in school) are essential in-home settings. Meta-analyses show that family behavior therapy and management skills training, and cognitive behavioral therapy for school-aged children are both clinically and cost-effective.

Mindfulness exposure is often part of interventions having both relaxation and an attention-focusing component. *Practicing deeper breathing*, a mindfulness platform, is sometimes called “watching the breath” or diaphragmatic breathing. The following guidelines are useful:

1. Sitting in a straight-back chair or lying down on a firm surface.
2. Closed eyes may be preferred.
3. Focus on breathing with attention to the abdomen/belly with a hand on it.
4. Relax the abdominal area and breathe slowly.
5. Slowly inhale through the nose for about a count of approximately 4 to 6.
6. Notice the abdomen rise; a hand on the stomach helps your awareness of this.
7. Try to hold the breath for about a count of 6 to 7.
8. Slowly exhale through the mouth or nose for a count slightly longer than your inhalation, for example, a count of 8.

To restate the breathing technique, in starting the breathing exercise, allot about 2 to 5 minutes to establish a calm breathing baseline that may or may not be maintained. While seated and comfortable, erect posture with the spine vertically aligned, take a deep breath through the nose sensing, and noticing its course to expand the chest via the rise in the upper abdomen and diaphragm. A count of 4 to 6 is right. Putting a hand on the upper waist/belly helps center attention to this upward rise. Hold the breath in for a count of approximately 6 to 7, then gently exhale through the mouth or the nose, whichever is more comfortable in the beginning. The precise timing of in-breaths, holding, and out-breaths varies. In-breaths are shortest, out-breaths are most extended, and time holding the breath is in between these. Doing this breathing practice for 5 minutes once or more times per day is helpful. Physiologically, arterial baroreceptors are stimulated and modulate blood pressure. Also, the vagus nerve afferent stimulation to the frontal cortex increases parasympathetic tone—greater tranquility.

Given the fundamental platform of breathing, a summary of the essence of this *Learned Mindfulness* model founded on emotional intelligence, mindful awareness, and self-observation follows. Starting with a basic familiarity of how one’s sensations emerge, one can be taught to label them as feelings. Grasping, releasing, contracting, and expanding attend these. Breathing affords a model of such fundamental dynamics. For example, releasing the breath follows the pause between inhalation and exhalation. This act establishes a sense of muting and not impulsively reacting that transfers to grasping, holding, and releasing objects arising in the mind. The emerging realization of gracefully letting go can progress further in the refinement of emotional processing: perception, conception, and performance. This emotional intelligence practice correlates the mind’s “objects,” for example, sensations, feelings, and thoughts, with their mindful management.

Special mention of the current widespread use of interventions derived from complementary, alternative, and integrative medicine may have a place in overall treatment because motivated families often first initiate these aids with enthusiasm not only to promote but also to enhance well-being. Healthy food choices, exercise, stress management, and even prenatal health measures (i.e., primary prevention) can be important components of a successful treatment plan. When used adjunctively, rationally, and collaboratively, they may help sustain a positively engaged therapeutic trajectory.

Proper recognition, together with remediation, enhances both cognitive and adaptive reserve. This long-term perspective optimizes potentials that extend benefits beyond the identified individual's immediate needs with BIF. Risk reduction may reach across generations. The young children of such individuals are thereby given a significantly increased chance of success in all quality-of-life measures. The adult children of persons with borderline intelligence can be given a rational explanation of insights into parents' long-standing intellectual limitations. This epiphany may afford them a capacity for a forgiving attitude toward what might have erroneously been perceived as intentionally motivated, less-than-optimal parenting.

A 10-year-old boy was referred for consultation because of failing grades since first grade and disruptive classroom behaviors, including expressing aggression toward peers, breaking things, and having temper tantrums. He would purposefully break any toys or gifts he was given; he would tear his clothes at times. He was the only child of a single mother who reported normal gestation, delivery, and no medical problems in the child. The boy was mildly overweight in the interview and had some facial dysmorphic features but appeared otherwise healthy. His mood was sullen. In the past, his mother had been given a variety of diagnoses for the child by school and pediatrician: ADHD, bipolar disorder, ASD, and CD. No psychotropic medications had been helpful. A comprehensive assessment included multiple interviews with the child and the mother, school consultation with intelligence, adaptive and scholastic testing, and routine blood work. Verbal and performance IQ scores were in the low 70s, adaptive skills were marginal, and grades were 1 to 2 years behind. Laboratory values were within normal limits. A primary diagnosis of BIF with a mood disorder, unspecified, was made.

Intensive treatment was three pronged: child, mother, and school. School consultation provided psychoeducation about diagnoses and the need for special education, which was approved. Family therapy also included psychoeducation about borderline intelligence, the child's strengths and limitations, and his need for a structured, straightforward, and routine environment of expectations and behavioral performance. Individual therapy for the mother also began. His mother was eager to use a range of health-oriented dietary and exercise guidelines. This therapeutic alliance with the mother was crucial to holding and integrating all prongs in an effective and collaborative way. Intensive individual psychotherapy with the child included aspects of cognitive, behavioral, and age-appropriate psychodynamic work. The child's sense of self, self-esteem, peer relations, and the effects of his behaviors on others were reviewed. During play therapy, several conflictual problems were uncovered. The boy revealed his envy of other boys on multiple levels. He envied their having a father, their perceived sense of being “smarter and not retarded,” their physical competence in sports, and their acceptance and facile participation in peer-group activities. His envy created a hatred of, and a wish to destroy, both them and, moreover, aspects of himself, as illustrated in either spoiling his clothing or breaking gifts he received. He often stated: “I’m angry because I just don’t fit in with everybody else.” Even though he was cognitively limited, he showed an interest in and some aptitude for understanding the complex envy dynamic of excessive idealization, spoiling, destructiveness, and resentment—when put into simple language and delivered flexibly at his level of readiness over time.

Play therapy and interpretive discourse were sensitive to engage him on the specific and changing levels of cognition and affect that he presented at each session. Changes were slow but significant. This developing level of emotional insight, basic emotional and cognitive understanding, appeared to enhance his everyday cognition because it diminished a great deal of anxiety and preoccupying resentment. The acquisition of “envy management skills” over time helped him understand cause-and-effect sequences and constrain destructive behaviors. Once-weekly psychotherapy was supplemented using low doses of quetiapine (Seroquel) to target irritability, anxiety, and mood. After approximately 6 months, his behavior stabilized, his aggressive outbursts were significantly eliminated, and his academic performance began to approach chronologic and developmental expectations. His mother remarked that his demeanor appeared more relaxed and that his affect brightened.

ACADEMIC OR EDUCATIONAL PROBLEM

Definition

DSM-5 describes another condition (Z55.9) *academic or educational problem* as follows:

> This category should be used when an academic or educational problem is the focus of clinical attention or impacts the individual’s diagnosis, treatment, or prognosis. Problems to be considered include illiteracy or low-level illiteracy; lack of access to schooling owing to unavailability or unattainability; problems with academic performance (e.g., failing school examinations, receiving failing marks or grades) or underachievement (below what would be expected given the individuals intellectual capacity); discord with teachers, school staff, or other students; and any other problems related to education and literacy. (Adapted from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, p. 830 [Copyright © 2022]. American Psychiatric Association. All Rights Reserved.)

The title of this entry now includes the added phrase “…or educational problem.” Unlike the DSM-IV-TR description, no mention is made of this condition not being “…due to a mental disorder or, if due to a mental disorder, is sufficiently severe to warrant independent clinical attention.” About seven additional problems are now mentioned: illiteracy, low-level illiteracy, lack of access to schooling, discord with teachers, school staff, students, and other issues related to education or literacy besides the previous two in DSM-IV-TR: failing grades and significant underachievement. No mention is made of requiring the absence of a specific learning or communication disorder or any other mental disorder that would account for the problem, as DSM-IV-TR had required. Therefore, the term “illiteracy” is not defined in a specific fashion; some impairment in the ability to read and write is implied. Some of the associated factors causing academic or educational problems are remediable extrinsic agents (e.g., access, discord, implied bullying, and victimization) that, if addressed, would assuage or eliminate the designated academic problem. The way DSM-5 defines this condition suggests that it can be a focus of attention or a secondary problem affecting an already present diagnosis, its treatment, or progress. DSM-5 coding is Z55.9.

DSM-5-TR (2022) describes several specifically coded conditions under the single category, Educational Problems (Z55.0 -Z55.9) with the subcategory Z55.9, now called “Other Problems Related to Education and Literacy,” that DSM-5, Z55.9, previously termed “Academic or Educational Problem” as described above.

DSM-5-TR now describes the entire range of Educational Problems as follows:

> These categories may be used when an academic or educational problem is the focus of clinical attention or has an impact on the individual's diagnosis, treatment, or prognosis. Problems to be considered include illiteracy or low-level literacy; lack of access to schooling owing to unavailability or unattainability; problems with academic performance (e.g., failing school examinations, receiving failing marks or grades) or underachievement (below what would be expected given the individual's intellectual capacity); discord with teachers, school staff, or other students; problems related to inadequate teaching; and any other problems related to education and/or literacy. (p. 830)

Etiology

The presentation of significant academic or educational problems or those impacting diagnosis, treatment, or progress denotes this condition. Risk factors include genetic, constitutional, premature birth, congenital atypicalities, and various early and later environmental stressors, for example, poor nutrition, less than optimal care, and maltreatment. English as a second language may also be considered. DSM-5 also clearly lists several causatives and clear-cut associative factors, leading to academic and educational problems. Some of these may have intrinsic causes, such as illiteracy associated with dyslexia or LDs. Most related causative agents are apparent environmental difficulties such as lack of access to schooling (e.g., transportation, attendance in-school, online school, tele school access, tutoring, etc.), interpersonal conflict, and perhaps behavioral dissonance, as delineated in the DSM-5 description. Bullying and victimization must be primary considerations.

Identification and Clinical Features

Scholastic problems, acute or chronic, associated with the particular factors clearly mentioned in this code's description best characterize academic or educational difficulties. Specific medical conditions, disabilities, or primarily overshadowing psychiatric disorders such as psychosis, schizophrenia, bipolar, depression, ASD, substance-related disorders, or PTSD may be comorbid but not sufficient to account for this coding. Clinical judgment is required.

Differential Diagnosis

Adequate psychological and intellectual testing and academic performance profiles derived from standardized scholastic testing need to rule out forms of ID, BIF, specific LDs, and communications disorders. All medical conditions (including, e.g., vision, hearing, motor) and disabilities that might contribute to illiteracy, poor school performance, or access must also be explored. Major psychiatric disorders, along with severe oppositional defiant disorder, must be considered. Substance abuse or dependence and CD must be ruled out.

One must also rule out clinical and subclinical states of anxiety and depression, as well as environmental stressors, either hidden or explicit, which significantly affect the child in the family (e.g., marital discord, divorce, abuse, and mental illness in the parent), the socioeconomic environment (e.g., poverty, malnutrition), and the suitability of the school setting (e.g., accessibility, presence of bullying, and so forth). Significant medical conditions (e.g., anemia, seizures, hypothyroidism, chronic infections, heavy metal loads, motor and ambulation impairments, and genetic disorders) must be considered. Adolescent developmental concerns, such as preoccupation with burgeoning sexuality, identity issues, and peer relations, should not be underestimated as factors that contribute to academic problems. Typically, such preoccupations are highly significant and may contribute to academic problems. Also, motivational issues and associated family and socially influenced contextual values may directly affect scholastic performance. Socialization and developing social skills and exposure to community values take place in the school environment. With the onset of remote learning, many of these venues have been disrupted, thus posing additional risk exposure.

With the advent of COVID-19, remote academic learning is prevalent. Home learning with technology and electronics and access to these are used. Hybrid combinations, including at-home remote, partial attendance, and full attendance, are possibilities. Absenteeism, for various reasons, including illness in the child or family, needs consideration. Inappropriate or excessive use of individually chosen screen time, including the lack of needed technologies such as electronics and lack of internet access, adds other emerging layers of concern related to academic or educational problems.

A 15-year-old boy was admitted to an intensive outpatient program to evaluate insidiously declining grades at school, isolative behavior, and incipient anhedonia occurring for the first time over the last semester in his second year in high school. The student told his family that he was being bullied in school. The patient had no previous psychiatric history and was in good health. School reports showed a full-scale IQ score to be 100 and no previous behavior or academic problems. Interviews with family and patient revealed that the boy had an intact family with no apparent or unusual conflicts noted. His mother found evidence of drug paraphernalia—cigarette papers, small pipe, matches, and a suspicious-looking dried substance—in the patient's bedroom. When confronted, the patient revealed a 6-month history of regular use of marijuana associated with his statement of being bullied in school. A primary diagnosis of cannabis abuse, along with an academic problem, was made. A secondary depression was to be ruled out. Referral to a therapeutic program with a drug and alcohol rehabilitation component was immediately implemented. A comprehensive school meeting was called to investigate and address the allegation of bullying.

A 10-year-old girl was admitted to an intensive outpatient program because of various disruptive, impulsive, hyperkinetic behaviors and intermittent rage reactions only in the home environment. The patient was well behaved in school and considered bright. Her grades, for the first time over the last year, were failing. Onset began about 2 years previously when her parents separated and her mother's new male companion moved in. The child's pediatrician made a diagnosis of ADHD and started methylphenidate (Ritalin). The child's condition worsened and was marked by mood instability and periodic temper outbursts, now also in school. The medical assessment showed the patient to be in good health, and she was referred for further examination. A psychiatric interview showed a guarded, hypervigilant child with a constricted affect. No other data were forthcoming. The patient's mother was reluctant to take part in treatment. The initial diagnostic impression was of disruptive behavior disorder, along with an academic problem. After several weeks and months, the patient revealed that she was sexually abused.

Further investigation and involvement with the Department of Child and Family Services uncovered that the home's adult male was the perpetrator. He was removed. Family and individual work began and continued for several months. All psychostimulant medication was discontinued. A primary diagnosis of PTSD was made, along with abuse and secondary academic problem. Over time, the child became less irritable, her rages stopped, and her grades in school improved.

Treatment

Identifying and addressing environmental—family, school, and socioeconomic—stressors are essential. Substance abuse and traumatic exposure must be ruled out. Any concurrent major psychiatric disorders require diagnosis and treatment. Since referred children and adolescents are typically prescribed multiple psychotropic medications, careful medication assessment is crucial.

Side effects, adverse drug effects, behavioral toxicity, and iatrogenic effects (e.g., hypothyroidism from lithium, or overactivation, irritability, mood instability, underweight, tics, stereotypies, and insomnia from psychostimulants) must be ruled out. When clinically indicated, taper and discontinuation or change of medications may be beneficial. With informed consent, school consultation can provide teachers with insight into the student's difficulties and suggest enhanced teaching strategies, especially conflict resolution techniques. Matters concerning bullying are now common and require careful consideration and treatment team coordinated efforts.

Child, family, school, and professional collaboration are essential. Illiteracy, underachievement, and failing grades need to be evaluated and appropriately addressed. If academic problems are severe enough, accommodations for individual educational programming, psychotherapeutic interventions such as counseling, guidance, family therapy, or residential schooling should be recommended even though this condition may not meet current stringent criteria for formal special education eligibility.

FUTURE DIRECTIONS

Intellectual *abilities*, social and emotional cognition, motivation, and successful adaptive performance skills coalesce through integrated thinking and behavior. Innovating emotional intelligence magnifies what I term *emotion performance utilization* (EPU). EPU is not a phase but the target of emotional intelligence. Viewing children and adolescents in this integral fashion calls for assessments and interventions that are comprehensive, multimodal, and flexible through the biopsychosocial paradigm.

In the introductory comments to this section, ongoing international efforts promoting evidence-based research and implementing recognition and intervention into BIAF are active. Perhaps, a case can be made for translating such patient-centered clinical practice into universally applicable research across the world. A global initiative is imperative to speed up BIF coordinated research, development, policy, and dissemination of pragmatic interventions.

In the United States, the AAIDD (2010) has proposed a rationale for formalizing a continuum of Intellectual and Developmental Disabilities. Included is a support framework spanning IQ limitations on either side of the definitional dividing line, typically full-scale IQ 70 to 75. AAIDD urges that individuals with BIF or persons with mild limitations falling marginally above this "cutoff" should become eligible for services already in place for those with mild ID. This bold yet accurate proposal attests to BIF being significant, let alone a vital area needing clinical intervention. BIF can, and usually does, interfere with activities of daily living and instrumental activities of living in the real world. This admonition for a broader breadth of the intelligence and adaptive continuum is consonant with this section's call to integrate BIF into the mainstream of official psychiatric nosology. This clinical improvement reflects the art guiding scientific psychiatry. Dimensional measures of severity between diagnostic entities transcend the limits of categories. The *Journal of the American Medical Association* (August 2014) advances the view that investing in prevention should be a strategic national priority to improve the United States' lagging population health compared with peer countries.

A new paradigm, the *biomental* perspective, has recently been introduced into psychiatric theory and has been applied to child development and parenting. The integral confluence of brain, mind, and behavior is proposed regarding neuroscience, developmental psychology, and pragmatic parenting strategies. Adaptive intelligence, a hypothetical construct, is presumed in this model. It denotes the central plexus subtending the wide range of conceptual, emotional, and motivational abilities underlying ease of effective response to change. This hub includes a tacit reserve of potential capabilities comprising an individual's innate resilience. We can make this intelligence potential active by therapeutic techniques that foster "instinctive resourcefulness," the repository of silent compensatory adaptive aptitudes. Here, "instinctive" is not impulsive but a visceral resilience made practical by a facilitating environment. These are learning-sensitive mediators between mental and emotional resources and their timely, spontaneous, and successful implementation as practical problem-solving performance skills. This innate capacity of resourcefulness—possibly distinct from abilities currently measured by and solely dependent on IQ scores—is considered a fundamental life tool. Although its functioning may be significantly limited in those with borderline cognition and adaptive skills, therapeutic initiatives may activate its enhanced operation as described in the treatment section.

By integrating insights from genetics, neuroscience, and neuropsychiatry, behavioral and clinical studies aim to construct valid and reliable phenotypes. They span the heterogeneous range of typical to atypical and include “borderline” with its inevitable problems and interferences in daily life. The aim is not to medicalize or pathologize but to enhance BIF. Having a rational probabilistic mindset for risk provides for formal and resource-rich interventions. Circumventing quality-of-life hazards for these individual’s underdeveloped potentials expands possibilities. Incorporating such contemporary perspectives eschews considering only single categorical models. It pays tribute to heterogeneity as a fact in complex psychiatric presentations.

Studies show that emotional awareness is diminished in those with BIF, that is, alexithymia—the inability to meaningfully label emotions sensed and felt. Inclusive treatments for BIF may also embrace guidelines that recognize *emotional management skills*. These assets promote understanding, reduce cognitive and emotional dissonance, and foster greater cognitive clarity and emotional stability. This upgrade optimizes both intellectual and adaptive functioning and proficiency. Emotions are a common experience. Language offers an organizing principle to share our common experiences with one another. Helping those with BIF expand their understanding, and use of mindful language may be an underestimated, high-impact emotional management skill. These management skills and various mindfulness techniques are now commonly available. They may be adapted as part of BIF’s individualized support resources.

A paradox of COVID-19 is the need for physical distance and attendant social spacing to minimize the disease spread. This distancing has brought children apart and intensified their already overinvolvement with electronic devices (e.g., the current need to use remote learning such as Zoom classrooms, etc.). In the COVID-19 era, the academic environment has undergone dramatic changes. The prevalence of distance, hybrid and at-home learning attests to this. Besides its inherent difficulties, the lack of regular socialization is paramount. Socialization includes material contact, interaction, and face-to-face feedback for correction and improvement. This interface comprises social learning with adults and a range of peers. Social learning fosters personal growth, maturation, and social skills development. Teamwork is enhanced. Being in a school setting also affords children access to both free and structured playtime. Play is essential to healthy development in childhood and those with borderline intellectual and adaptive skills needing improvement. COVID-19 has brought a new layer of multifaceted problems as an additional toll on marginally functioning persons.

Remote learning and home-based education are established in current educational approaches. Computer hardware, access, and having internet connectivity are potential problems. Improving accessibility and usability is evolving. Home-based education’s complexity involves technology and teacher training to optimize access and the user-friendliness of educational materials. Addressing the social issues around remote learning remains and will continue to be a significant challenge.

The environmental embeddedness—*eco-corporeality*—of each individual is assumed in these approaches. Our entire corpus—human being—is embedded interactively in the fabric making up the world. The integrated use of inclusive interactional paradigms enhances a sustainable, successful, and satisfactory quality of life. This significant advance denotes a life imbued with meaning and dignity—disabilities becoming newfound abilities. This integration crosses multiple eco-contexts, including personal, interpersonal, social, academic, familial, occupational, cultural, and spiritual. Identification and active interventions cross the adaptive intelligence threshold; integrating those with BIF into this *nuanced norm* is necessary. More than a human right, this piece of social participation is the right to be human to the fullest extent.

FURTHER READINGS

Champagne FA. Epigenetics. In: Bornstein MH, ed. *The Sage Encyclopedia of Lifespan Human Development*. Thousand Oaks, CA; 2018:760–762.

Edgerton R. *The Cloak of Competence: Stigma in the Lives of the Mentally Retarded*. Rev. ed. University of California Press; 1992.

Fenning RM, Baker JK, Baker BL, Crnic KA. Parent-child interaction over time in families of young children with borderline intellectual functioning. *J Fam Psychol*. 2014;28(3):326–335.

Fernell E, Ek U. Borderline intellectual functioning in children and adolescents—insufficiently recognized difficulties. *Acta Paediatr*. 2010;99(5):748–753.

Ferrari M. Borderline intellectual functioning and the intellectual disability construct. *Intellect Dev Disabil*. 2009;47(5):386.

Fletcher R, Loschen E, Stavrakaki C, et al. *Diagnostic Manual—Intellectual Disability (DM-ID): A Textbook of Diagnosis of Mental Disorders in Persons with Intellectual Disability*. NADD Press; 2007.

Goh PK, Martel MM, Barkley RA. Clarifying ADHD and sluggish cognitive tempo item relations with impairment: a network analysis. *J Abnorm Child Psychol*. 2020;48:1047–1061. https://doi.org/10.1007/s10802-020-00655-2

Hassiotis A. Borderline intellectual functioning and neurodevelopmental disorders: prevalence, comorbidities and treatment approaches. *Advances in Mental Health and Intellectual Disabilities*. 2015;9(5):275–283.

Kaufman AS. *IQ Testing 101*. Springer Press; 2009.

Klein M. Our adult world and its roots in infancy [1959]. In: Klein M. Hogarth Press; 1975:247–263.

Martinez-Leal R, Folch A, Munir K, Novell R, Salvador-Carulla L; Borderline Intellectual Functioning Consensus Group. The Girona declaration on borderline intellectual functioning. *Lancet Psychiatry*. 2020;7(13):E8.

Mascolo JT, Flanagan DP, Alfonso VC. *Essentials of Planning, Selecting, and Tailoring Interventions for Unique Learners (Essentials of Psychological Assessment)*. Wiley; 2014.

Miller DC. *Essentials of School Neuropsychological Assessment (Essentials of Psychological Assessment)*. 2nd ed. Wiley; 2013.

Ninivaggi FJ. *Learned Mindfulness: Physician Engagement and M.D. Wellness*. Elsevier/Academic Press; 2020.

Ninivaggi FJ. Borderline intellectual functioning in children and adolescents: reexamining an under-recognized yet prevalent clinical comorbidity. *Conn Med*. 2001;65:7–11.

Ninivaggi FJ. DM-ID, Diagnostic manual-intellectual disability: a textbook of diagnosis of mental disorders in persons with intellectual disability, (book review). *J Autism Dev Disord*. 2008;38:204.

Ninivaggi FJ. *Envy Theory: Perspectives on the Psychology of Envy*. Rowman & Littlefield Press; 2010.

Ninivaggi FJ. *Biomental Child Developmental: Perspectives on Psychology and Parenting*. Rowman & Littlefield Press; 2013.

Ninivaggi FJ. *Making Sense of Emotion: Innovating Emotional Intelligence*. Rowman & Littlefield Press; 2017.

Nouwens PJG, Lucas R, Embregts PJCM, van Nieuwenhuizen C. In plain sight but still invisible: a structured case analysis of people with mild intellectual disability or borderline intellectual functioning. *J Intellect Dev Disabil*. 2016;1–9. Published online: 04 May 2016. http://dx.doi.org/10.3109/13668250.2016.1178220

Salvador-Carulla L, Garcia-Gutierrez JC, Gutierrez-Colosia MR, et al. Funcionamiento intellectual limite: guia de consenso y buenas practices. *Rev Psiquiatr Salud Ment (Bar.)*. 2013;6:109–120.

Schalock RL. *Adaptive Behavior and Its Measurement: Implications for the Field of Mental Retardation*. American Association on Mental Retardation; 1999.

Shalock RL, Borthwick-Duffy SA, Bradley VJ, et al. *Intellectual Disability: Definition, Classification, and Systems of Supports*. 11th ed. American Association on Intellectual and Developmental Disabilities; 2010.

Shalock RL, Luckasson R, Bradley V, et al. *User's Guide: Intellectual Disability: Definition, Classification, and Systems of Support*. American Association on Intellectual and Developmental Disabilities; 2012.

Stein DJ, Phillips KA, Bolton D, Fulford KW, Sadler JZ, Kendler KS, et al. What is a mental/psychiatric disorder? From DSM-IV to DSM-V. *Psychol Med*. 2010;40:1759–1765.

Thompson JR, Bryant BR, Campbell EM, Craig EM, Hughes CM. *Supports Intensity Scale: User's Manual*. American Association on Mental Retardation; 2004.

Thompson JR, Bradley VJ, Buntinx WHE, et al. Conceptualizing supports and the support needs of people with intellectual disability. *Intellect Dev Disabil*. 2009;47(2):135–146.

Tymchuk AJ, Lakin KC, Luckasson R. *The Forgotten Generation: The Status and Challenges of Adults with Mild Cognitive Limitations*. Paul Brooks; 2001.

Wieland J, Zitman FG. It is time to bring borderline intellectual functioning back into the main fold of classification systems. *BJPsych Bull*. 2016;40(4):204–206.

Woittiez I, Eggink E, Putman L, et al. An international comparison of care for people with intellectual disabilities. *The Netherlands Institute for Social Research.* The Hague, Netherlands: July 2018. © The Netherlands Institute for Social Research | scp, The Hague, 2018 ISBN 978 90 377 0871 4.

28.4 Relational Problems

WILLIAM BERNET, M.D.

As social creatures, individuals cannot be completely understood without considering their interpersonal relationships. A person's identity is determined, in part, by how others see and value them. Likewise, clinicians cannot understand a person's symptoms out of the context of the individual's social relationships. These relationships can be promotive and protective of good mental and physical health —or neutral—or detrimental to health outcomes. There is a long history of competition—and perhaps some conflict—between the proponents of biologic psychiatry and the practitioners and researchers who emphasize the psychosocial and relational underpinnings of mental illness. No doubt, both these perspectives are essential in understanding and intervening with the mental disorders that trouble the individuals and the families with whom we work.

DEFINITIONS

Relational problems are maladaptive patterns of feelings, behavior, and perceptions involving two or more members of a relational unit. They can be distinguished from a broad range of interpersonal difficulties by four features: (1) repeated, fixed patterns of painful and destructive patterns of feelings, behavior, and perceptions can be clearly recognized; (2) the patterns are long-standing and are not a response to a recent stressful event; (3) the maladaptive patterns are nonresponsive to supportive features that may occur naturally in the social environment; (4) there is clear evidence of a major impact of these patterns on psychological functioning, physical health, social adaptation, and/or occupational activities in one or more persons in the relationship. A relational problem may come to clinical attention as the primary reason that the individual consults with a physician or therapist, or it may be a problem that affects the course, prognosis, or treatment of the individual's mental or medical disorder.

Relational problems are included in a chapter that occurs toward the end of the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) and its text revision (DSM-5-TR), "Other Conditions That May Be a Focus of Clinical Attention." With regard to terminology, the topics discussed in this chapter are called *problems* or *conditions;* in contrast, most of the topics discussed in other parts of DSM-5 are called *disorders,* such as delusional disorder and panic disorder. A common explanation for the distinction is that mental disorders are “located” inside the affected individual, while relational problems are “located” between the affected individual and another person.

Previous editions of the DSM employed a multiaxial system that included the documentation of mental disorders (Axes I and II), other medical disorders (Axis III), psychosocial and contextual factors (Axis IV), and level of disability (Axis V). Thus, relational problems were listed under Axis IV of the multiaxial system. As DSM-5 was developed, the Task Force decided to abandon the use of the multiaxial system. Mental disorders, other medical disorders, and relational problems are now included in one list, typically in the order of importance in understanding and conveying why the individual client is undergoing assessment and/or treatment. Some writers have had the misunderstanding that relational problems and other conditions (identified as V codes) may not be diagnosed because they are not mental disorders; that is incorrect since mental disorders, problems, and conditions may all be listed as diagnoses.

COMPARATIVE NOSOLOGY

Relational problems were not considered or alluded to in the first DSM (1952). In DSM-II (1968), however, a new diagnostic category was introduced called “social maladjustments without manifest psychiatric disorder.” It was noted that “this category is for recording the conditions of individuals who are psychiatrically normal but who nevertheless have severe enough problems to warrant examination by a psychiatrist.” One of the conditions included in that category was *marital maladjustment,* which was intended “for individuals who are psychiatrically normal but who have significant conflicts or maladjustments in marriage.” The authors of DSM-II were prescient in saying, “The far-reaching implications of this set of terms for forensic psychiatry and for insurance compensation will become clearer with time.”

DSM-III (1980) featured a new chapter called “V Codes for Conditions Not Attributable to a Mental Disorder That are a Focus of Attention or Treatment.” That chapter included several diagnoses that were precursors to conditions that ultimately were classified as relational problems: *uncomplicated bereavement (UB), marital problem, parent–child problem,* and *other specified family circumstances.* The last mental condition included sibling rivalry. DSM-III-R (1987) provided the same terms and definitions for those mental conditions.

In DSM-IV (1994), the title of the chapter changed to “Other Conditions That May Be a Focus of Clinical Attention.” That chapter included a separate section called “Relational Problems,” which included the following: *relational problem related to a mental disorder or general medical condition, parent–child relational problem (PCRP), partner relational problem, sibling relational problem (SRP),* and *relational problem not otherwise specified.* DSM-IV was criticized because the descriptions of these relational problems were overly vague. However, DSM-IV also featured the Global Assessment of Relational Functioning (GARF) Scale, which was included in the chapter, “Criteria Sets and Axes Provided for Further Study.” The GARF was intended to indicate an overall judgment of a family relationship on a hypothetical continuum ranging from optimal relational functioning to a disrupted, dysfunctional relationship. It was analogous to the Global Assessment of Functioning (GAF) Scale that constituted Axis V of DSM-IV. The GARF and GAF Scales were continued in DSM-IV-TR (2000) but not in DSM-5 (2010).

Even before the publication of the DSM-IV, the Group for the Advancement of Psychiatry and the Family Psychology Division of the American Psychological Association was joined by 13 other professional societies interested in relational problems; they formed the Coalition of Family Diagnosis. The Coalition compiled further information about family dysfunction and relational problems and published the classic, *Handbook of Relational Diagnosis and Dysfunctional Family Patterns* (1996). The first sentence of the

Handbook stated, “Relational diagnosis is a concept whose time has come.” The contributors to that book were disappointed in the minimalistic presentation of relational problems in DSM-IV and hoped to influence the discussion of those topics in DSM-5. For example, they criticized the traditional principle in DSM that a mental disorder is something that occurs *in a person;* that approach seemed to exclude problems that emerge from disturbed relational patterns and dysfunctional family systems. Another consideration has been financial: “Other conditions” and V codes are typically not acceptable to third-party payers.

When the DSM-5 Task Force got underway, there was not a separate workgroup dedicated to relational problems. However, liaisons from the field of family therapy were included in some of the other workgroups. The *Research Agenda for DSM-5* (2002) included a chapter that emphasized both personality disorders and relational problems. It was suggested that relational problems should have the following features, which are ambitious goals not only for relational problems but also for traditional mental disorders:

- The diagnostic description must be of the relationship, not just of the individual.
- Relational problems should have distinctive features.
- Relational problems should cause severe emotional, social, and occupational impairment.
- Relational problems should have a recognizable clinical course.
- Relational problems should show recognized patterns of comorbidity.
- Relational problems are expected to appear in patterns of family aggregation.
- Relational problems should involve both biologic and psychosocial factors.
- Relational problems should show a response to specific treatments.
- Relational problems should be preventable.

A group of family researchers was organized to collect the scientific evidence and to create the conceptual frameworks necessary to clarify the place of interpersonal relationships in clinical practice. That team of researchers, the Relational Processes Working Group, advised both the DSM-5 Task Force and the component of the World Health Organization (WHO) that was revising the *International Classification of Diseases* (ICD). The Working Group published a book, *Relational Processes and DSM-V: Neuroscience, Assessment, Prevention, and Intervention,* which DSM-5 personnel utilized in revising the chapter on “Other Conditions.” The Working Group published a second book, *Family Problems and Family Violence,* which was directed more closely at ICD-11. In general, the importance and prominence of relational problems have gradually increased in both DSM-5 and ICD-11.

Eight relational problems—which all relate to psychosocial issues within families or couples—are included in DSM-5: (1) *parent-child relational problem (PCRP),* (2) *sibling relational problem (SRP),* (3) *upbringing away from parents,* (4) *child affected by parental relationship distress (CAPRD),* (5) *relationship distress with spouse or intimate partner (RDSIP),* (6) *disruption of family by separation or divorce,* (7) *high expressed emotion (EE) level within family,* and (8) *UB.* This is a rather disparate potpourri of mental conditions, which have a variety of manifestations, circumstances, and implications. The descriptions and definitions of relational problems in DSM-5-TR are more detailed and robust than those in DSM-IV, but there remains room for improvement in developing specific criteria for these mental conditions. These eight conditions are included in their own section, “Relational Problems,” which distinguish them from “Abuse and Neglect,” and the various additional topics that are included in the chapter, “Other Conditions That May Be a Focus of Clinical Attention.”

Table 28.4–1.
Comparable Terminology for Relational Problems in DSM-5 and ICD-11

DSM-5-TR	ICD-11
Parent–child relational problem, V61.20	Caregiver–child relationship problem, QE52.0
Sibling relational problem, V61.8	Relationship with parents, in-laws, or other family members, QE50.4
Upbringing away from parents, V61.8	Removal from home in childhood Institutional upbringing, QE93
Child affected by parental relationship distress, V61.29	Problems related to primary support group, including family circumstances, unspecified, QE70.Z
Relationship distress with spouse or intimate partner, V61.10	Relationship distress with spouse or partner, QE51.0
Disruption of family by separation or divorce, V61.03	Disruption of family by separation or divorce, QE70.1
High expressed emotion level within family, V61.8	Problems related to primary support group, including family circumstances, unspecified, QE70.Z
Uncomplicated bereavement, V62.82	Uncomplicated bereavement, QE62

There was considerable discussion among the writers of DSM-5 and ICD-11 when those manuals were being developed, with the hope of coordinating definitions as much as possible. See Table 28.4–1 for comparisons of terminology regarding relational problems in DSM-5-TR and ICD-11. Relational problems are included in a much broader concept called social determinants of health

(SDOH), which the WHO has introduced. SDOH are the “conditions in the environments in which people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks.”

PARENT–CHILD RELATIONAL PROBLEM

Definition

DSM-5-TR specifies that PCRP (V61.20) should be used when the main focus of clinical attention is to address the quality of the parent–child relationship or when the quality of the parent–child relationship is affecting the course, prognosis, or treatment of a mental or other medical disorder. The DSM-5-TR further explains that PCRP is associated with impaired functioning in behavioral, cognitive, or affective domains. Examples of behavioral problems include inadequate parent control, supervision, and involvement with the child; parental overprotection; excessive parental pressure; arguments that escalate to threats of physical violence; and avoidance without resolution of problems. Cognitive problems may include negative attributions of the other’s intentions, hostility toward or scapegoating of the other, and unwarranted feelings of estrangement. Affective problems may include feelings of sadness, apathy, or anger about the other individual in the relationship.

Etiology

There are multiple risk factors for PCRP. They include the child’s genetic predisposition, the child’s having a difficult temperament, parental psychopathology, parent–child attachment problems, inadequate supervision, parental relationship distress, family violence, negative peer interactions, poverty, neighborhood problems, and stressful life events. Maladaptive methods of parenting are frequently the cause of PCRPs. A common system of classifying parenting styles uses the two dimensions of parental *demandingness* (behavioral control, supervision, maturity demands) and *responsiveness* (warmth, acceptance, involvement). This model results in subclassification into four parenting styles: (1) *authoritative,* with high demandingness and responsiveness, (2) *authoritarian,* with high demandingness and low responsiveness, (3) *permissive,* with low demandingness and high responsiveness, and (4) *neglectful,* with low demandingness and low responsiveness (Table 28.4–2). One can see that there is a broad range of adequate, good enough parenting, and problematic parenting can range from too much of a good thing (being overly protective) to too little (being neglectful). Most practitioners encourage parents to strive for an authoritative approach with their children. Parents who are overly authoritarian, permissive, or neglectful are more likely to create PCRP between themselves and their children.

Table 28.4–2.
Parenting Styles Diagram

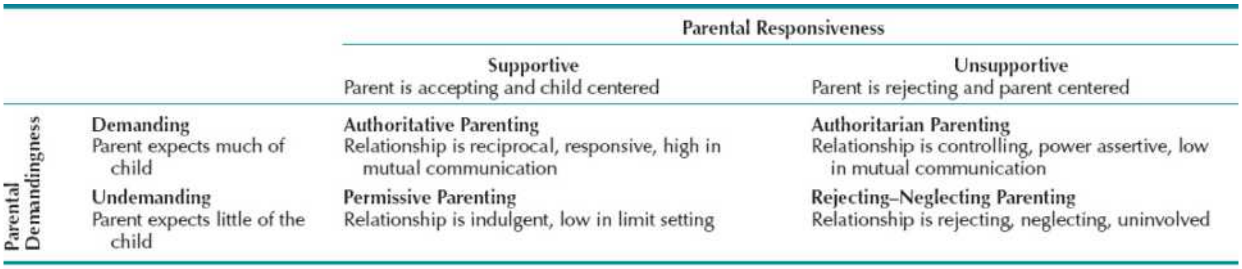

		Parental Responsiveness	
		Supportive Parent is accepting and child centered	Unsupportive Parent is rejecting and parent centered
Parental Demandingness	Demanding Parent expects much of child	Authoritative Parenting Relationship is reciprocal, responsive, high in mutual communication	Authoritarian Parenting Relationship is controlling, power assertive, low in mutual communication
	Undemanding Parent expects little of the child	Permissive Parenting Relationship is indulgent, low in limit setting	Rejecting–Neglecting Parenting Relationship is rejecting, neglecting, uninvolved

Data from Baumrind D. The influence of parenting style on adolescent competence and substance use. *J Early Adolesc*. 1991;11(1):56–95.

Diagnosis and Clinical Features

PCRP pertains to the behaviors and attitudes of both the child and the parent; one or the other or both individuals may contribute to their dysfunctional relationship. Of course, this diagnosis may also apply to the relationship between the child and other caregivers who are actively involved in raising the child, such as extended family or even close family friends. The Semistructured Interview for Caregiver–Child Relationship Problem has been introduced, which may be useful in both research and clinical settings.

In cases of PCRP, it is common for one or both parties to experience personal mental health issues—such as depression in the parent or separation anxiety in the child—which drive the PCRP that occurs between them. In order to fully express what is happening in the family, it is necessary to identify the mental disorders (if any) of both the child and the parent, as well as the dysfunction (if any) that occurs in their relationship.

The parent–child relationship is particularly important in infancy, and early childhood, when the development of the brain is essential and thus affects the child's mental maturation. A healthy parent–child relationship buffers and modulates psychopathologic risks arising from the child, the parent, or elsewhere. Impaired parent–child relationships can exacerbate and even create psychopathology. The organization, Zero to Three has focused on the assessment, diagnosis, treatment, and prevention of mental health problems in infants and toddlers. Their book, *DC:0–5,* is a diagnostic manual for young children that complements DSM-5.

Differential Diagnosis and Comorbidity

PCRP is intended to capture dysfunctional parenting, but not child maltreatment. If physical, sexual, or psychological abuse is occurring, those conditions should be the diagnosis, not PCRP. The parent participant in PCRP may have a mental disorder that is contributing to the problem, such as a substance use disorder, a mood disorder, or an anxiety disorder. The child participant in PCRP may also have a mental disorder that is aggravating their family relationships, such as oppositional defiant disorder or conduct disorder. Parents who are in conflict with their children may also be in conflict with each other, so one should consider an additional diagnosis such as relationship distress with a spouse or intimate partner or spouse or partner violence (SPV). Also, one should consider the possibility of parental alienation, where a child is manipulated by one parent to reject the other parent; that condition could explain symptoms that may occur in PCRP, such as negative attributions to the intentions of the rejected parent, extreme hostility toward that parent, and unwarranted feelings of estrangement.

SIBLING RELATIONAL PROBLEM

Definition

DSM-5-TR states that SRP (V61.8) should be used when the focus of clinical attention is a pattern of interaction among siblings that is associated with significant impairment in individual or family functioning or with development of symptoms in one or more of the siblings, or when an SRP is affecting the course, prognosis, or treatment of a sibling's mental or other medical disorder.

Etiology

Retrospective research indicates that the majority of adults have experienced sibling rivalry in some form. The birth of a new sibling usually results in an older sibling becoming competitive with the infant for the attention and affection of parents. Such rivalry becomes disordered when it is associated with extremely negative affects, overt hostility, or attempts to injure the new sibling or any other sibling.

SRP is exacerbated by marital conflict and parental displays of favoritism. When a child is idealized or is made a parental scapegoat, rivalry may become violent. Certain cultures may sanction favoritism, for example, favoring the oldest male, without increasing the risk for pathologic sibling rivalry.

A poignant form of an SRP may occur when a child experiences serious illness or death, which may be a life-altering event for that child's siblings and other family members. Such an event changes family dynamics, disrupts the family routines, and captures the parents' time and attention. The affected child's siblings may be marginalized just when they most need to feel like important members of the family.

Diagnosis and Clinical Features

Sibling rivalry refers to competition among siblings for the attention, affection, and esteem of their parents or for other recognition or reward. It typically consists of jealousy, competition, and fighting between brothers and sisters. The child may feel that they are receiving unequal amounts of the parent's attention, discipline, or responsiveness. *Sibling rivalry* is a more general term that refers to common interactions among siblings; in contrast, *SRP* is a less common and more serious mental condition that—like all mental disorders and conditions—is associated with clinically significant impairment in functioning.

In cases of SRP, the maladaptive sibling relationship adversely affects family function, is the cause of significant impairment, or contributes to problematic behaviors in one or more siblings. Behaviors that occur in SRP include recurring arguments and conflicts that distract from the activities of daily life, avoiding interaction with the sibling who is causing the stress, threats of a child's running away, a pattern of anger toward a sibling that does not resolve because of a lack of conflict resolution strategies, and excessive jealousy between siblings. SRPs may occur between full, half, step, foster, or adopted siblings.

Although verbal and physical conflicts among siblings are common, major violence and abuse present a more serious challenge. There are no bright lines to distinguish them, but it seems obvious that extreme SRP may lead to sibling violence that becomes sibling abuse. The latter features injurious behaviors such as hitting hard enough to leave a mark, cutting, burning, and putting a child in dangerous circumstances; also, the victim becomes fearful of the perpetrator. However, parents may downplay fighting between siblings and regard it as child's play. Even the children are confused about what is abusive.

Differential Diagnosis and Comorbidity

In assessing for possible SRP, there needs to be a determination as to whether there is a partner relationship problem in the family that may have destabilized the family system, resulting in an intensification of sibling needs, parental needs, the showing of favoritism, or other more pathologic means of differentiating between siblings. SPV must be ruled out as well. The presence of serious psychiatric disorders or substance use disorders in the parents must be determined.

The siblings need to be evaluated for a variety of psychiatric diagnoses and psychosocial stressors. Being teased or bullied at school may exacerbate sibling violence at home. Having been sexually abused either within or outside of the family might predispose a child to be sexually abusive of a less powerful or younger sibling. Violent siblings need to be evaluated for past or current history of physical abuse, the witnessing of intimate partner violence (IPV), and specific psychiatric disorders. Specific disorders that might predispose a child to be abusive to a sibling include attention deficit hyperactivity disorder, conduct disorder, and mood disorders. Psychotic disorders, impulse disorders, and any neurologic conditions that might result in increased impulsivity or decreased inhibition and judgment should be considered.

UPBRINGING AWAY FROM PARENTS

Definition

DSM-5-TR states that *upbringing away from parents* (V61.8) should be used when the main focus of clinical attention pertains to issues regarding a child’s being raised away from the parents or when this separate upbringing affects the course, prognosis, or treatment of a mental or other medical disorder. The child could be one who is under state custody and placed in kin care or foster care. The child could also be one who is living in a nonparental relative’s home or with friends but whose out-of-home placement is not mandated or sanctioned by the courts. Problems related to a child living in a group home or orphanage are also included.

Etiology

Some children require out-of-home placement, most often due to parental neglect and inability to provide adequate care, less often because of physical abuse, sexual abuse, parental incarceration, and abandonment. Foster care—including shelters, orphanages, nonrelative foster parents, and relative foster parents—constitute a temporary form of out-of-home care; adoption is intended to be a permanent assumption of parental responsibilities by individuals other than the biologic parents. Government data indicate that more than 420,000 children are in formal foster care—arranged and monitored by a social service agency—in the United States each year. The number of children placed through informal fostering arrangements outside of the public system, for example, with grandparents, is several times that.

Diagnosis and Clinical Features

Most children enter foster care with pre-existing handicaps and vulnerabilities, such as chronic medical and mental health problems. Separation from their families of origin and also the loss of peers, schools, and neighborhoods constitute enormous stressors. Although many children show resilience in the face of these stressors, the risk for continuing physical and mental health problems in children in foster care is higher than that of the general population. It is sadly common for children to experience multiple disruptions in placements over time, which add to their vulnerability for serious and prolonged mental health problems. Frequently the goal for the youngster is reunification with their biologic parents, which occurs in more than one-half of cases. However, many children who are reunified with their families of origin are subsequently placed again into foster care.

Various initiatives have been introduced with the goal of reducing the morbidity associated with foster care. For example, social service agencies strive to avoid unnecessary separation of children from their families in the first place by providing services for the prevention of removal. Permanency planning is a goal, with reunification, adoption, or long-term foster placement as possible options. Social service agencies also aim to free a child for adoption—by the termination of parental rights—sooner rather than later. All of these initiatives are intended to reduce the unstable and continuous cycle of children being in multiple foster homes over extended periods of time.

Kinship care—most often with the child’s grandparents—is a popular option in lieu of reunification with the child’s biologic parents. Typically, the grandparents hope that the child’s parents will resolve the problem that led to the family crisis, such as substance abuse, homelessness, etc. The advantages of placement with grandparents include the child’s pre-existing attachment to the grandparents; preservation of a connection to the extended family of origin; greater stability and permanence; and maintenance of religious, racial, and ethnic values. There are also concerns regarding placement with grandparents: These are older individuals with health concerns and physical limitations; they may be unemployed and have limited financial resources, and difficulty maintaining restrictions on parental visitation. Fostering by grandparents and other relatives is often arranged informally rather than through a social service agency. In order for the family member to access medical care, educational services, etc., it may be necessary for the caregiver to secure guardianship or legal custody.

Practitioners will face other challenges associated with the relational problem, upbringing away from parents. (1) For example, children come to the United States from all over the world, fleeing desperate circumstances, becoming separated from family, and now experiencing a new culture and language. There is a surge of unaccompanied minors from the southern border of the United States, having been exposed to violence and severe deprivation. (2) A concern for child and adolescent psychiatrists is that children in foster care are prescribed a disproportionate amount of psychotropic medication. Prescribers and consultants try to put more emphasis on psychosocial interventions. (3) Finally, both practitioners and caregivers are concerned when foster children “age out” of the system. When that occurs—typically at age 18 or 21—everyone involved in their program should create an adequate transition plan, which relates to social supports, housing, education, employment, and medical insurance.

CHILD AFFECTED BY PARENTAL RELATIONSHIP DISTRESS

Definition

This diagnosis, CAPRD (V61.29), was new to DSM-5, which provided a very brief account of its meaning. DSM-5 simply states that CAPRD should be used when the focus of clinical attention is the negative effects of parental relationship discord (e.g., high levels of conflict, distress, or disparagement) on a child in the family, including effects on the child's mental or other medical disorders.

Etiology

About 55% of American marriages end in divorce. The divorce rate in second marriages is even high than for first marriages. Also, a large percentage of unmarried parents separate when they have young children. Approximately one-half of the children in the United States will experience parental separation or divorce.

A conflict between parents is common even in harmonious and stable relationships; most children, even in high-conflict homes, do not develop emotional or behavioral problems. Nevertheless, negative child outcomes are more strongly associated with parental conflict than with the parents' divorce. In fact, children from divorced families appear to have higher levels of well-being than children from intact, high-conflict families. Of course, continued conflict between parents after divorce exacerbates the children's poor adjustment.

Diagnosis and Clinical Features

Parental relationship distress may refer to persistent disparagement of one or both parents by the other parent; high levels of conflict; intimate partner distress (IPD: dissatisfaction with the relationship as well as difficulty resolving conflicts, lack of positive exchanges, coercive exchanges, or persistently perceiving negative intentions in the partner); and IPV (physical force such as hitting, slapping, and biting; extreme psychological manipulation; and/or coercive sexual acts). A CAPRD may display impaired functioning in behavioral, cognitive, affective, and/or physical domains. Examples of behavioral problems include oppositionality and the child's reluctance or refusal to have a relationship with a parent (contact refusal). Cognitive problems may include cognitive dissonance, attempting to maintain affection for both parents simultaneously (loyalty conflict), and/or adopting the false belief that the rejected parent is evil or dangerous (parental alienation.). Affective problems may include anger, anxiety, depressed mood, and post-traumatic symptoms. Physical symptoms may include stomachaches, headaches, and exacerbation of general medical conditions.

Differential Diagnosis and Comorbidity

The scope of CAPRD includes the following four scenarios:

RDSIP, also called intimate partner distress, is the most common cause of acute emotional distress in treatment-seeking couples. RDSIP features relationship issues such as conflict resolution difficulty, withdrawal versus overinvolvement, chronic negative attributions of the other's intentions, dismissal of the partner's positive behavior, and chronic sadness, apathy, and/or anger about the other partner. Of course, the hurtful effects of marital conflict on children are well documented. Specifically, the level of marital conflict predicts children's internalizing and externalizing disorders.

SPV is also called IPV. SPV is always more serious than RDSIP. RDSIP is more amenable to treatment; SPV is more likely to be associated with antisocial personality disorder. The latter is a pattern of behavior in which one partner uses physical violence, coercion, threats, intimidation, isolation, or emotional, sexual, or economic abuse to control the other partner. There has been considerable research regarding the effects on children who have been exposed to IPV. It is traumatic for a child to witness persistent or substantial conflict between parents. When high levels of violence occur, the child may develop post-traumatic stress disorder. When low levels of violence occur over a period of time, the child may develop anxiety (such as separation anxiety disorder or generalized anxiety disorder) or depression. The clinician may use both terms together—for example, CAPRD with post-traumatic stress disorder—to convey clearly the child's condition.

A **Loyalty Conflict** occurs in a child when they try to maintain affection and good feelings toward each of their parents, even though the parents are angry and hostile toward each other. A child may experience a significant degree of divided loyalty if parental conflicts are obvious and persistent. Also, a more serious loyalty conflict may develop if one or both parents pressure the child to support that parent's side. Children frequently develop physical and psychological symptoms—such as headaches and stomachaches—when they experience intense loyalty conflict stress. These physical symptoms typically occur when the time approaches for the child to transition from the household of one parent to that of the other parent.

Parental alienation is almost always more serious than a loyalty conflict. Those mental conditions involve different psychic mechanisms for coping with parental conflict. A child with a loyalty conflict puts mental energy into maintaining a good relationship with both parents; that effort evokes cognitive dissonance, which causes discomfort and anxiety. The child may resolve those uncomfortable emotional states by aligning with one parent against the other, especially if the favored parent was able to manipulate the child into believing that the rejected parent was abusive or neglectful. In cases of parental alienation, the child typically adopts false beliefs that the rejected parent was abusive or neglectful, which drive strongly expressed contact refusal and hostility. It is important for both clinicians and forensic practitioners to distinguish parental alienation (rejection of a parent without a good reason) from realistic parental estrangement (rejection of a parent for a good reason, such as a history of abuse or neglect by that parent).

Heather was a 6-year-old girl whose parents recently divorced, and her mother was the primary residential parent. Before the divorce, Heather had a close relationship with her father. Their relationship slowly deteriorated after the divorce to the point that Heather refused to visit her father and clung to her mother during the scheduled exchanges. Beginning when the parents separated, Heather's mother cried at the transition times and repetitively told Heather, "I love you. I miss you. I can't wait for you to come back." During her time with her father, the mother telephoned Heather repeatedly, asking if she was all right, and then said, "I miss you so much. I can't wait for you to come home." When Heather returned to her mother, her mother ran to her, became tearful, and said, "I'm so happy you're back. I've been worried sick about you. I don't know what to do when you're gone." The father noticed that Heather gravitated away from him. Just before Heather began to refuse visitation with her father, she told him, "Stop being so mean to Mommy. I hate you." This is an example of a child affected by parental relationship distress, specifically parental alienation. (Adapted from Lorandos D, Bernet W. *Parental Alienation: Science and Law*. Charles C Thomas, Publisher, Ltd; 2020.)

RELATIONSHIP DISTRESS WITH SPOUSE OR INTIMATE PARTNER

Definition

DSM-5-TR states that *RDSIP* (V61.10) should be used when the major focus of the clinical contact is to address the quality of the intimate (spouse or partner) relationship or when the quality of that relationship is affecting the course, prognosis, or treatment of a mental or other medical disorder. Examples of behavioral problems include conflict resolution difficulty, withdrawal, and overinvolvement. Cognitive problems can manifest as chronic negative attributions of the other's intentions or dismissals of the partner's positive behaviors.

Affective problems may include chronic sadness, apathy, and/or anger about the other partner.

Etiology

A satisfying marriage is one of the most important goals in life. Marital happiness is more strongly associated with overall life satisfaction than is satisfaction in areas involving health, work, and children. A survey in the United States found that the most frequently cited causes of acute emotional distress were problems in intimate relationships, including separation, divorce, and other problems with intimate partners.

General etiologic factors of RDSIP include genetic and neurobiologic contributions to temperament and style of interaction. Several studies have identified genetic factors related to partner relationship problems as modestly to moderately heritable, including divorce risk, marital satisfaction, conflict, and warmth. Most personality studies have focused on temperament, which is highly heritable. The most commonly used rating scale, the Five-Factor Model (FFM) of Temperament, categorizes temperament along five dimensions: neuroticism, extraversion, openness, agreeableness, and conscientiousness. High neuroticism and low agreeableness are the most consistent predictors of marital dissatisfaction, conflict, divorce, and abuse.

Childhood attachment patterns are remarkably stable through adulthood. Adult attachment patterns are described along two domains: seeking or avoiding attachment and fear of intimacy. Securely attached adults usually have a positive view of themselves and others; they feel loved, expect responses to self-disclosures, minimize negative affects, and engage in effective conflict resolution. Developmental factors, psychosocial factors, and sociocultural expectations within and outside the family, all contribute to stability in the partner relationship or lack thereof. A history of parental divorce increases the risk for partner separation, especially if both spouses' parents are divorced.

There is a circular pattern of causation between mental illness and stressful experiences such as relationship distress. For example, individuals with symptoms of depression are less supportive toward their partner and have lower expectations regarding support from the partner. These negative behaviors and expectations lead to greater intimate partner distress. Then, the level of marital or partner stress predicts subsequent depressive symptoms. This feedback loop serves to maintain symptoms and reduce the effectiveness of interventions.

Diagnosis and Clinical Features

There are psychological tests that highlight possible marital or intimate partner relational problems. The Relationship Conflict Inventory consists of 114 items that ask how many times particular forms of conflict occurred and the distress level of each occurrence for the participant. The Marital Satisfaction Inventory-Revised (MSI-R)—which consists of 150 true/false questions—is designed to identify the nature and intensity of distress in distinct areas of partner interaction. A short form of the MSI-R can be used as a screening tool. It consists of only 10 questions, such as, “Does your partner often fail to understand your point of view on things?” Of course, if the screening test reveals a possible relational problem, a more complete assessment of RDSIP would include clinical interviews and perhaps more elaborate questionnaires.

Dysfunctional relationships are characterized by an increase in negative affects, cognitive distortions, communication problems, and difficulties in conflict resolution. Negative affect is the most significant sign of RDSIP, particularly when accompanied by negative reciprocity. Negative reciprocity exists when negative affect is countered with more negative affect, resulting in affect amplification. Dysfunctional couples tend to attribute hurtful encounters to events outside of their control, for example, their partner’s “unchangeable” personality traits, and view negative behaviors as undeserved, intentional, or malevolent. Attributions may also be affected by transference, with characteristics of a person’s partner being associated with earlier attachment figures, evoking re-enactments of earlier feelings and experiences. Relational stability is negatively affected by unrealistic expectations and conflicting personal and relational goals, for example, the desire to have children, parenting strategies, money, sex, religion, and so on.

Strategies for problem solving are critical to relationship stability. The maladaptive characteristics of conflict resolution that are highly predictive of divorce in dysfunctional couples are harsh start-ups or arguments, frequent use of criticism and contempt, defensive reactions, withdrawal and stonewalling, and failed repair mechanisms. Abrupt criticism at the beginning of a conflict results in affective and autonomic flooding. Men are more prone to feeling flooded due to their higher physiologic reactivity; however, both genders may experience autonomic flooding. Cognitive cascading is another characteristic of dysfunctional relationships, with criticism turning into contempt and contempt triggering a loss of feelings of admiration and memories about positive aspects of the relationship, producing a loss of motivation for preserving the relationship. With escalating intensity, stonewalling and emotional withdrawal follow. Intervention is required if healthy communication patterns are to be restored.

Differential Diagnosis and Comorbidity

Having a successful intimate relationship is a major developmental task of adulthood. Consequently, the failure of a relationship affects a person’s sense of self and self-esteem, resulting in sadness, despondency, and possibly suicidal risk. Because risk is high when there is a comorbid psychiatric illness, such as an affective disorder or substance abuse, a thorough psychiatric assessment is essential because untreated disorders may exacerbate the relational problem. Conversely, a relational problem may exacerbate a pre-existing psychiatric disorder.

Past history of violent behavior and the presence of a gun in the house pose significant risks. Personality characteristics associated with violence include pathologic jealousy, verbally abusive behavior, and progressive isolation of one’s partner from family and friends. Risk is particularly high when separation is imminent, and the violent partner indicates, “If I can’t have you, no one will.” If there are children in the family, it becomes important to determine whether they are being maltreated. It is likely that when there is an RDSIP, there will also be present relational problems such as CAPRD or PCRP, as children struggle to manage in the midst of their parent's difficulties.

DISRUPTION OF FAMILY BY SEPARATION OR DIVORCE

Definition

DSM-5-TR states that *disruption of family by separation or divorce* (V61.03) should be used when partners in an intimate adult couple are living apart due to relationship problems or are in the process of divorce.

Etiology

Marital distress predicts divorce, which is strongly associated with symptoms of depression for the parents and also child mental health outcomes. Separation and divorce are common: in the United States, 41% of first marriages, 60% of second marriages, and 73% of third marriages end in divorce. The median duration of first and second marriages is 7 to 8 years. The U.S. divorce rate varies by the occupation of the parties: dancers, bartenders, and massage therapists have the highest divorce rates; dentists, farmers, podiatrists, clergy, and optometrists have the lowest divorce rates. Between 2000 and 2019, both the marriage rate and the divorce rate in the United States have declined: the annual marriage rate fell from 8.2 per 1,000 population to 6.1; the annual divorce rate fell from 4.0 per 1,000 population to 2.7. This low rate continued during the pandemic, with 2.3 divorces per 1,000 population in 2020, and 2.5 in 2021.

Diagnosis and Clinical Features

Parental divorce places children at risk for a range of negative sequelae in childhood, adolescence, and adulthood. In childhood, parental divorce is associated with internalizing and externalizing behavior problems, academic problems, social difficulties, and poor physical health. When parental divorce occurs during adolescence, there are continued academic, mental health, and physical health problems, as well as risk for substance use and pregnancy. Adults who experienced parental divorce during childhood have an increased risk for major depression, dysthymia, anxiety disorders, personality disorder, and alcohol and drug dependence. Parental divorce also predicts lower educational attainment, occupational problems, physical health problems, and poorer relationships with spouses, parents, and peers in adulthood.

HIGH EXPRESSED EMOTION LEVEL WITHIN FAMILY

Definition

DSM-5-TR states that *high EE level within family* (V61.8) should be used when a family's high level of EE is the focus of clinical attention or is affecting the course, prognosis, or treatment of a family member's mental or other medical disorder. *EE* is a construct used as a qualitative measure of the "amount" of emotion—in particular, hostility, emotional overinvolvement, and criticism directed toward a family member who is an identified patient—displayed in the family environment.

Etiology

It should be noted that contemporary perspectives on EE are distinct from early attempts at developing models of the influence of family processes on severe mental illness. While historical notions—such as "poor parenting," "double-bind messages," and "schizophrenogenic" mothering—were overly parent blaming, current terminology avoids censuring the family for a person's mental illness. For example, modern theories do not regard the family as the primary causal agent of the patient's illness. Instead, the emphasis is on how the family reacts to a psychiatrically impaired relative and how these reactions protect against or contribute to the risk of recurrence of illness.

Diagnosis and Clinical Features

An established assessment device for EE is a semistructured interview known as the Camberwell Family Interview (CFI), which is conducted with the parents or spouse without the patient being present. This instrument provides ratings on the dimensions of criticism, hostility, emotional overinvolvement, warmth, and positive remarks, which allow family members to be classified as high or low on EE. The CFI is a lengthy, complex process and is practical in research settings but not in clinical practice. The Five Minute Speech Sample simply requires the family to talk about the patient for 5 uninterrupted minutes. The speech is recorded and later coded for the overall level of EE, criticism, and emotional overinvolvement. The Level of Expressed Emotion (LEE) scale is a 60-item self-report, which is completed by either the patient's important relationships or by the patient themself. The Family Attitude Scale is a 30-item self-report measure of EE; either relatives or patients may complete it. Finally, the test of Perceived Criticism (PC) consists of only two questions: "How critical do you think [your relative] is of you?" and "When [your relative] criticizes you, how upset do you get?"

High levels of EE may be benign in many family contexts, and in other families may be elicited by the stress associated with physical and mental health problems. For example, parents or family members of severely mentally ill individuals may find themselves confronted with difficult situations for which they are poorly prepared, resulting in excessive EE by the family member toward the mentally ill individual. Thus, some serious mental disorders appear to be induced—in part—by high EE in the family. For instance, genetic load plus high EE in a patient's parents appear to predict the onset of schizophrenia. That phenomenon illustrates the biosocial hypothesis, that is, that major psychotic disorders are the result of the continual interaction of specific biologic disorders of the brain with specific psychosocial and other environmental factors.

It has been observed that relatives of individuals with a major psychiatric disorder tend to respond with negative and increasingly intense emotions to subtle symptoms and negative behavior during the prodromal phase of the illness. As symptoms and functioning worsen, parents manifest more rejection and less warmth. Also, EE levels typically decline as patients are treated, and their condition improves. There are implications for interventions with these families, that is, that onset of illness might be delayed and relapses prevented if family members remain low in criticism and rejection of the patient. The family's therapist can help family members view the patient's behavioral symptoms as signs of their illness, not indications of an evil disposition.

During the first family session, JJ's father began with a list of all of the things JJ had done in the past several months to "destroy the family atmosphere." These included inconsistent school performance, staying up too late playing video games, being irritable, especially when asked to interrupt his game playing; being grumpy at the dinner table; getting into altercations with kids at school in situations where he felt insulted or left out; and more generally, "poisoning the atmosphere wherever he goes." JJ responded by intermittently crying and cursing at his father. In the next session, conducted just with JJ and his mother, JJ cursed loudly and repeatedly jumped out of his chair when his mother spoke, threatening to physically hurt her with pieces of cardboard, pens, or other implements he found in the therapy room. Despite this behavior, JJ showed some insight into his problems. He complained bitterly that his father "is always after me to do better… he never tells me I'm doing good enough." This is an example of a high expressed emotion level within the family. (Reprinted with permission of Hooley JM, Miklowitz DJ. Perceived criticism in the treatment of a high-risk adolescent. *J Clin Psychol.* 2017;73(5): 570–578. doi:10.1002/jclp.22454.)

UNCOMPLICATED BEREAVEMENT

Definition

DSM-5-TR states that *UB* (V62.82) may be used when the focus of clinical attention is a normal reaction to the death of a loved one. As part of their reaction to such a loss, some grieving individuals present with symptoms such as feelings of sadness and associated symptoms such as insomnia, poor appetite, and weight loss. The bereaved individual typically regards the depressed mood as "normal," although the individual may seek professional help for relief of associated symptoms such as insomnia or anorexia.

Diagnosis and Clinical Features

Questionnaires have been used for decades to screen for symptoms of grief. The Texas Revised Inventory of Grief (TRIG) assesses typical symptoms of UB, such as acceptance of loss, crying, and intrusive thoughts. The Inventory of Complicated Grief (ICG) assesses indicators of complicated bereavement, such as anger, disbelief, and hallucinations. Bereavement is also called "grief" and "mourning." After the loss of a loved one, the bereaved individual may experience sadness, loneliness, and perhaps anger. In some circumstances—for example, involving long-term caregivers—the process of bereavement may begin even before death occurs, which is called anticipatory grief. Cultural expectations may influence how a person experiences grief. In some cultures, a death in the family is manifested in a quiet and private manner; in other cultures, there may be loud, intense displays of emotion.

In contrast to the relational problem of UB, complicated or complex grief makes it harder for the bereaved person to adapt to the loss of the loved one. These individuals may ruminate over the timing or manner of the death. They may put excessive energy into avoiding reminders of the loss. They may question whether their own experience of grief is too much or too little. In complicated grief, these symptoms persist longer than 6 months following the death of the loved one.

Differential Diagnosis and Comorbidity

The DSM-5-TR discusses how to distinguish bereavement or grief from a major depressive episode (MDE). In UB, the predominant affect is feelings of emptiness and loss, while in an MDE, it is persistent depressed mood and the inability to anticipate happiness or pleasure. The dysphoria in UB is likely to decrease in intensity over days to weeks and occurs in waves, which tend to be associated with thoughts or reminders of the deceased. The depressed mood of an MDE is more persistent and not tied to specific thoughts or preoccupations. The pain of UB may be accompanied by positive emotions and humor that are uncharacteristic of the pervasive unhappiness and misery characteristic of an MDE. If a bereaved individual thinks about death and dying, such thoughts are generally focused on the deceased and possibly about “joining” the deceased, whereas, in an MDE, such thoughts are focused on ending one’s own life because of feeling worthless, undeserving of life, or unable to cope with the pain of depression.

CONCLUSIONS AND FUTURE RESEARCH

The relational problems are concepts that clinicians and research personnel will find useful once they become familiar with their meaning, scope, and implications. In clinical practice, it adds important information to say that a child with a generalized anxiety disorder is also experiencing the relational problem, high EE level within the family. Or to say that an adult with persistent depressive disorder is also experiencing the relational problem, disruption of the family by separation or divorce. Thus, traditional DSM-5-TR diagnoses focus solely on symptoms exhibited by individual patients, while relational problems identify the context and often the precipitating cause of the individual’s symptoms. Identification of the contextual component may lead to a more comprehensive treatment plan. The relevance of relational problems goes well beyond psychiatric disorders since they pertain to many disorders across all branches of medical practice. For example, the strategies for managing chronic medical illnesses may include an assessment of family factors and interventions with families, which creates a role for the relational problems described in DSM-5-TR.

Researchers can achieve greater specificity by using both traditional diagnoses and relational problems in describing or classifying their study participants. For example, treatment studies for children with specific disorders—for example, anxiety, depression, or disruptive behavior disorders—can assess outcomes using the occurrence of IPD or IPV in parents as covariates to determine whether the presence of CAPRD affects treatment outcomes. Or, screening for parental relationship distress may be accomplished during well-baby appointments. If there are relational problems involving the parents, they could be randomized to couples therapy or treatment as usual, and tracking participant outcomes could test whether that intervention prevents the onset or progression of child mental health problems.

It seems counterproductive for clinicians and researchers to “choose sides” between biologic and psychosocial methods for conceptualizing mental illness. These are simply different ways of understanding the same reality, along with acknowledging that biologic and mental factors mutually influence each other. For example, genetic risk for major depression puts a person at risk for marital and family problems, which in turn are risk factors for depression.

Of course, there are opportunities for further research and improvements in the classification of relational problems, which may play out in future iterations of the DSM. Many family therapists would like to see a broader range of mental and interpersonal conditions included in this section. They would prefer more detailed descriptions of each condition, and they probably would like to include well-known terms and concepts such as triangulation and cross-generational coalition. Also, it is apparent that relational problems must be described in terms that are clearly operationalized. Studies can be designed to see if clinicians who observe the same interpersonal processes make the same relational diagnosis. It is important to distinguish relational problems from similar, perhaps overlapping, conditions: for example, what criteria distinguish PCRP from

CAPRD? Finally, researchers may be able to sort out how specific relational problems lead to specific mental disorders.

FURTHER READINGS

Bakula DM, Sharkey CM, Perez MN, et al. Featured article: the relationship between parent and child distress in pediatric cancer: a meta-analysis. *J. Pediatr Psychol.* 2019;44(10):1121–1136.

Baumrind D. The influence of parenting style on adolescent competence and substance use. *J Early Adolesc*. 1991;11(1):56–95.

Beach SRH, Wamboldt MZ, Kaslow NJ, et al., eds. *Relational Processes and DSM-V: Neuroscience, Assessment, Prevention, and Treatment*. American Psychiatric Association; 2006.

Beach SRH, Whisman MA. Relationship distress: impact on mental illness, physical health, children, and family economics. In: Foran HM, Beach SRH, Slep AMS, Heyman RE, Wamboldt MZ, eds. *Family Problems and Family Violence: Reliable Assessment and the ICD-11*. Springer Publishing; 2013:91–100.

Bernet W, Greenhill LL. The Five-Factor Model for the diagnosis of parental alienation. *J Am Acad Child Adolesc Psychiatry*. 2022;61:591–594.

Bernet W, Wamboldt MZ, Narrow WE. Child affected by parental relationship distress. *J Am Acad Child Adolesc Psychiatry*. 2016;55:571–579.

Bodin AM. Relationship conflict—verbal and physical: conceptualizing an inventory for assessing process and content. In: Kaslow FW, ed. *Handbook of Relational Diagnosis and Dysfunctional Family Patterns*. John Wiley; 1996:371–393.

Cherry MG, Taylor PJ, Brown SL, Sellwood W. Attachment, mentalisation and expressed emotion in carers of people with long-term mental health difficulties. *BMC Psychiatry* 2018;18:257.

Cummings EM, Davies PT. *Marital Conflict and Children: An Emotional Security Perspective*. Guilford Press; 2010.

Finkelhor D, Turner HA, Ormrod R. Kid's stuff: the nature and impact of peer and sibling violence on younger and older children. *Child Abuse Negl*. 2006;30:1401–1421.

First MB, Bell CC, Cuthbert B, et al. Personality disorders and relational disorders: a research agenda for addressing crucial gaps in DSM. In: Kupfer DJ, First MB, Regier DA, eds. *A Research Agenda for DSM-V*. American Psychiatric Press; 2002:123–199.

Foran HM, Beach SRH, Slep AMS, Heyman RE, Wamboldt MZ. *Family Problems and Family Violence: Reliable Assessment and the ICD-11*. Springer Publishing; 2013.

Foran HM, Whisman MA, Beach SRH. Intimate partner relationship distress in the DSM-5. *Fam Process*. 2015;54:48–63.

Gillath O, Karantzas GC, Fraley RC. *Adult Attachment: A Concise Introduction to Theory and Research*. Academic Press; 2016.

Gottman JS, Gottman JM. *10 Principles for Doing Effective Couples Therapy*. WW Norton; 2015.

Hahlweg K. Parent-child relational problems: the psychological, physical, and economic consequences. In: Foran HM, Beach SRH, Slep AMS, Heyman RE, Wamboldt MZ, eds. *Family Problems and Family Violence: Reliable Assessment and the ICD-11*. Springer Publishing; 2013:203–216.

Hart SA, Little C, van Bergen E. Nurture might be nature: cautionary tales and proposed solutions. *NPJ Sci Learn*. 2021;6(1):2. doi: 10.1038/s41539-020-00079-z

Hooley JM, Miklowitz DJ. Perceived criticism in the treatment of a high-risk adolescent. *J Clin Psychol*. 2017;73(5):570–578.

Hooley JM, Miklowitz DJ, Beach SRH. Expressed emotion and DSM-V. In: Beach SRH, Wamboldt MZ, Kaslow NJ, et al., eds. *Relational Processes and DSM-V: Neuroscience, Assessment, Prevention, and Treatment*. American Psychiatric Association; 2006:175–191.

Kaslow FW, ed. *Handbook of Relational Diagnosis and Dysfunctional Family Patterns*. Wiley; 1996.

Linscott JG, Leve LD. Parent-child and sibling relational problems. In: Goldstein S, DeVries M, eds. *Handbook of DSM-5 Disorders in Children and Adolescents*. Springer International; 2017:565–581.

McFarlane WR. Family expressed emotion prior to onset of psychosis. In: Beach SRH, Wamboldt MZ, Kaslow NJ, et al., eds. *Relational Processes and DSM-V: Neuroscience, Assessment, Prevention, and Intervention*. American Psychiatric Publishing; 2006:77–87.

Miklowtiz DJ, Beach SRH, Reiss D, et al. Recommendations for research on relational disorders and processes. In: Beach SRH, Wamboldt MZ, Kaslow NJ, et al., eds. *Relational Processes and DSM-V: Neuroscience, Assessment, Prevention, and Intervention*. American Psychiatric Publishing; 2006:241–258.

Plamondon A, Bouchard G, Lachance-Grzela M. Family dynamics and young adults' well-being: the mediating role of sibling bullying. *J Interpers Violence*. 2021;36(9–10):NP5362–NP5384. doi: 10.1177/0886260518800313

Radetzki PA, Deleurme KA, Rogers SM. The implications of high-conflict divorce on adult–children: five factors related to well-being. *J Family Studies*. 2021;2021. http://dx.doi.org.proxy.library.vanderbilt.edu/10.1080/13229400.2020.1754277

Shear MK, Bloom CG. Complicated grief treatment: an evidence-based approach to grief therapy. *J Rat-Emo Cognitive-Behav Ther*. 2017;35:6–25.

Slep AMS, Tamminen T. Caregiver–child relational problems: Definitions and implications for diagnosis. In: Foran HM, Beach SRH, Slep AMS, Heyman RE, Wamboldt MZ, eds. *Family Problems and Family Violence: Reliable Assessment and the ICD-11*. Springer Publishing; 2013:185–196.

Wamboldt MZ, Cordaro AR Jr. Practical tools for assessing caregiver-child relationship problems. In: Foran HM, Beach SRH, Slep AMS, Heyman RE, Wamboldt MZ, eds. *Family Problems and Family Violence: Reliable Assessment and the ICD-11*. Springer Publishing; 2013:217–228.

Wamboldt MZ, Cordaro AR Jr, Clarke D. Parent-child relational problem: field trial results, changes in DSM-5, and proposed changes for ICD-11. *Fam Process.* 2015;54:33–47.

Whisman MA, Beach SRH. Relationship distress: Assessment, definition, and implications for mental health diagnosis. In: Foran HM, Beach SRH, Slep AMS, Heyman RE, Wamboldt MZ, eds. *Family Problems and Family Violence: Reliable Assessment and the ICD-11*. Springer Publishing; 2013:71–78.

Zero to Three. *DC:0–5: Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood*. Zero to Three; 2016.

29

Cultural Concepts of Distress and Assessment

29.1 Cultural Concepts of Distress

Roberto Lewis-Fernández, M.D., M.T.S., Laurence J. Kirmayer, M.D., and Brandon A. Kohrt, M.D., Ph.D.

Cultural psychiatry has documented wide variations in the experience and expression of mental distress around the world. Historically, these variations were often characterized in terms of “culture-bound syndromes”—distinctive patterns of co-occurring symptoms and signs that were assumed to be rooted in or “bound” to local cultural or contextual features. With advances in research, this view has been challenged, and a more refined set of terms have been proposed to capture the ways in which symptoms, explanations, and expressions of distress vary across cultural contexts. This chapter reviews these newer concepts and approaches, which are included in DSM-5-TR, and illustrates how they can be used to better understand the causes, course, and consequences of mental health problems in the community and the clinic.

Contemporary cultural psychiatry and psychology recognize that all classification schemes are rooted in particular cultural and historical frames. In turn, the diagnostic language of psychiatry influences popular culture and everyday understandings of mental health problems, ensuring that professionally defined categories help pattern social experience. Hence, Western psychiatric nosologies need to be subjected to the same kind of social, historical, and contextual analysis applied to diverse local nosologies.

Despite local variations, however, similar forms of distress have been documented in many regions of the globe, resulting in clusters of symptoms or syndromes, such as depression or schizophrenia, that can be described at a level of abstraction that allows broad cross-cultural application. Local cultural variations in symptom interpretation and experience will nonetheless influence the recognition, response, and outcome of these disorders. Moreover, some local variations may depend on specific culturally based attributions, behavioral responses, and patterns of interpersonal interaction so that they only occur where those modes of explanation and interaction are present. These indigenous illness forms may occur in Western as well as non-Western contexts. The study of cultural variations in the presentation of illness and distress yields a valuable critique of Western classification schemes by highlighting the assumptions of existing nosology and identifying alternate, indigenous categories that may better fit the experience of patients from diverse backgrounds.

Terminology and Conceptualization

Cross-cultural mental health researchers have introduced several terms to refer to and describe culturally specific forms of expressing and diagnosing emotional distress, including cultural syndromes, popular or folk illness categories, idioms of distress, and illness explanations or attributions. The need for multiple terms reflects the complexity of the ways in which people experience, understand, and communicate distress.

The term “culture-bound syndrome” was used in the past to describe patterned behaviors of distress and/or illness that appeared to be distinct from existing psychiatric categories and were considered unique to particular cultural settings. The clear implication was that Western psychiatric categories were not culture-bound but rather universal and that proper characterization would disclose a simple mapping of non-Western syndromes onto the categories of international nosology.

The term “syndrome” generally refers to (1) a collection of co-occurring symptoms and signs with (2) a distinctive temporal course, which may include an immediate precipitating cause or etiologic context and response to treatment. Most of the entities described in the DSM as disorders are syndromes in that their identification depends on symptoms and signs since their underlying causes and mechanisms remain uncertain, and there are no reliable biomarkers. The categories are clinically useful largely because they have implications for treatment choice and prognosis. As underlying mechanisms become clear, new distinctions can be made, and specific symptoms, signs, and behaviors may be distinguished and given more diagnostic weight. However, the symptom cluster itself may remain important as a central feature of patients’ experience.

The notion of a syndrome being bound by culture is more problematic. The attribution of “boundedness” emerged early in the history of cultural psychiatry in colonialist contexts in which observers had limited understanding of local cultural contexts. Hence, there was a strong tendency to exaggerate differences, generate stereotypes, and “exoticize” the other. Reports that focused on a particular cultural group in a community noted patterns of problematic behavior, which appeared distinctive or unique to the (usually) Western observers, whether they were missionaries, anthropologists, or clinicians. In some cases, the behavior pattern had a specific popular label and treatment within an indigenous medical system. The foreign observers often believed that the communities in question had been socially isolated, and therefore, saw the patterns of distress as unique and locality bound. Considering the massive migrations and cultural diffusions that have occurred throughout human history, this assumption of boundedness has rarely been tenable. For example, in the late 1950s, Raymond Prince described a Nigerian culture-bound syndrome of “brain fag” characterized by feelings of mental fatigue, difficulty concentrating, heat in the head, and myriad other somatic symptoms. Brain fag was found among students who were the first in their families or communities to pursue higher education and was linked to local ethnophysiologic ideas about the importance of central heat and the head in African psychology. However, recent historical evidence shows that “brain fag” was commonly described in England in the late 1800s, and presumably, traveled to Nigeria as part of the colonial system to be adopted as a local mode of expressing distress. Hence, while there may indeed be an important fit between the syndrome of brain fag and local cultural concepts, it must also be understood in relation to these colonial cultural influences.

It would be more accurate, therefore, to term these conditions *cultural syndromes*, de-emphasizing their boundedness because, while they originate in particular cultural areas, they spread with migration and communication and are incorporated into local concepts of illness and expressions of distress. In the contemporary global market of intensified cultural exchanges, exportable syndromes readily become detached from geographical areas. Whatever boundedness they have inheres not in specific locales but in their adoption by ethnic or linguistic groups and links to the sociocultural patterns of which they form part, whether in their countries of origin or in new homes.

The term “syndrome” implies a relatively fixed pattern of symptoms. Only a subset of culture-specific expressions of distress around the world shows this level of organization. Many are typified instead by a more diffuse set of complaints, symptoms, and predicaments that vary across situations and settings, age cohorts, and social subgroups. Rather than representing separate illnesses, they constitute a general category of distress, part of an indigenous system of classifying suffering, sometimes encompassing not only illness but also other forms of misfortune or adversity. Reports of *nervios* (nerves) in Latin America, for example, comprise a continuum of conditions, labeled with different but related terms (e.g., "being nervous since childhood," "suffering from nerves," "being ill with nerves," and "attacked by my nerves"), all denoting different gradations of vulnerability and of the effects of adverse events on the nervous system, including symptoms, signs, and causal attributions. As another example, Haitians may distinguish broadly between "*maladi dyab*" and "*maladi bon Dieu*"—the first referring to a set of illnesses ("maladi") triggered in association with evil forces ("*dyab*," literally "devil"), usually through spiritual means, and the latter referring to diseases of the natural order, sent by God ("*bon Dieu*"). A great diversity of human afflictions fall under these general headings. Anthropologists use the term *folk illness* to refer to local illness categories like these that are based on particular systems of medicine or broader cultural ontologies, independently of whether they are organized as discrete syndromes or not. A related term is *popular illness category*, which simply reflects the fact that these are labels or constructs commonly used by lay people in the general population.

Table 29.1–1.
Key Elements of the Constructs Used to Describe the Cultural Concepts of Distress

Folk Illness Categories	Cultural Syndromes	Idioms of Distress	Causal Explanations
▶ Encompass cultural syndromes, idioms of distress, and causal explanations ▶ May include afflictions and predicaments not considered illnesses in the local cultural context	▶ Organized as a collection of co-occurring symptoms and signs ▶ Have distinctive temporal courses, precipitants, causal attributions, and treatment responses	▶ Constitute a linguistic or nonverbal mode of communication ▶ Represent a local “language” for expressing distress ▶ Do not presuppose pathology	▶ Reference etiologies rooted in local systems of knowledge ▶ Are not necessarily linked to particular signs and symptoms ▶ Guide help seeking

One drawback to the term “folk illness” is that the descriptor “folk” tends to connote less affluent and formally educated sectors of society and ignores the fact that these categories are used in complex medical systems embraced by large segments of the population in modern, urban, industrialized societies. Thus, traditional Chinese medicine or Indian Ayurveda has many categories of illness defined in terms of the imbalance of energies or vital essences, respectively. Another disadvantage of the term is that many local illness labels are used to describe afflictions and expressions of distress that are on a continuum with folk illnesses but are not strictly illnesses themselves (much as the term “depression” refers to a psychiatric disorder as well as to lay descriptions of much milder reactions to adversity). DSM-5-TR uses the more general term *cultural concepts of distress* to encompass folk illnesses, cultural syndromes, and two other types of cultural expressions of distress described below.

One of these is the notion of a *cultural idiom of distress*, a key concept in recent medical anthropology. Introduced by anthropologist Mark Nichter in 1981, the term “idiom” emphasizes the idea of a local linguistic or nonverbal mode of communication. It highlights the communicative function of the cultural concepts of distress, the notion that suffering is expressed in particular local “languages of distress,” which are shaped by cultural understandings of what are appropriate and meaningful ways of denoting affliction. “Idiom” also does not presuppose pathology. Many idioms—like “thinking too much,” “stressed out,” or lay notions of “depression”—can be used to express everyday concerns or levels of distress that may be subclinical. However, idioms of distress also provide a way to express more severe suffering that leads to help seeking. Hence, the clinician must be able to decode the significance of such expressions to identify dimensions of the problem that require intervention.

A final cultural category encompassed under the term “cultural concepts of distress” involves the many expressions of distress pointing to *causal explanations* that are rooted in local systems of knowledge. These, in turn, may reference particular ontologies involving ethnophysiologic ideas about humors, energies, or other processes, ethnopsychological theories about the mind, sociomoral explanations in terms of problems in the social order, or spiritual or other religious and cosmologic systems. As etiologies, these causes are not necessarily linked to particular signs and symptoms but rather denote explanations for the source of the problem that usually guide help seeking. Table 29.1–1 summarizes the key elements of the constructs used in this chapter to describe the cultural concepts of distress.

Cultural syndromes, folk or popular illness categories, idioms of distress, and explanations commonly coexist, and the same local term, or closely related terms, may refer to more than one of these types of concepts. For example, *nervios* terms in Latin America can be used to describe a syndrome (e.g., as the subcategory *ataque de nervios*, a stereotyped sequence of precipitants, behaviors, and sequelae), a folk illness category (e.g., when denoting an illness "affecting the nervous system," as in "she is ill with nerves"), an idiom (e.g., when communicating the nature of the affliction, such as, "my nerves are killing me"), or an explanatory cause (e.g., "he was born with emotional problems because his mother suffered from nerves during her pregnancy"). Moreover, usage may change over time, reflecting cultural changes and the influence of the health care system itself. Despite this complexity, the concepts provide a window into the ways that patients and their families understand their problems and are essential for accurate epidemiologic research, clinical diagnosis, and assessment, as well as designing appropriate and effective interventions. The following sections further trace the origins and use of these terms in psychiatric theory, research, and practice.

HISTORICAL BACKGROUND

Pioneering Work

In 1962, Pow Meng Yap, an early exponent of cultural psychiatry, introduced the term *atypical, culture-bound psychogenic psychoses* to denote what he later called "forms of psychopathology produced by certain systems of implicit values, social structure and obviously shared beliefs indigenous to certain areas." In response to changing nosologic paradigms, this term evolved over time into "atypical, culture-bound reactive syndromes," from which the label of "culture-bound syndromes" emerged. Yap's selection of the term "culture-bound" was based on his view that aspects of culture (i.e., implicit and explicit values and notions, social structure, and folk practices) pattern the expression of psychopathology to such a degree that particular forms of illness may be confined to the geographical boundaries of the cultural group in question. He expected that once analyzed further, these apparently unique forms of mental illness would prove to be atypical variations in generally distributed psychogenic disorders. Basic to Yap's formulation was an acceptance of the universality of psychoanalytic constructs, on which culture could exert pathoplastic effects. This implied it should be possible to locate all of these "exotic" syndromes within the classificatory systems of clinical psychiatry, and Yap proposed a formal ordering of these syndromes within a universal nosologic schema for clinical, didactic, forensic, and administrative purposes. Yap also suggested that attempts should be made to distinguish "essential" from "accessory" symptoms as well as to establish the "natural history" of each disorder in its "elementary psychogenic form" in order to accurately parcel the syndromes among psychiatric and somatic illness categories. His expectation was that such studies would lead first to a consistent definition of each disorder in the form of modal symptoms and then to universally valid mechanisms of psychopathology, providing a firm basis for cross-cultural clinical research.

In the 1980s, Raymond Prince and coworkers expanded on Yap's writings. They proposed a definition of these phenomena as a "collection of signs and symptoms (excluding notions of cause), which is restricted to a limited number of cultures primarily by reason of certain of their psychosocial features." This definition attempted to narrow Yap's "culture-bound syndromes" exclusively to the medical denotation of "syndrome" as a collection of descriptively defined symptoms that are "a reasonably complete descriptive picture of nature, remain fairly constant over time and are verifiable by all investigators." This kind of nosologic entity was distinguished from "labels" and "causal notions" superimposed idiosyncratically on universal psychopathologic entities. In this view, these terms conveyed much less scientific accuracy and were prone to nonverifiable misconceptions. Prince attributed some of the controversy surrounding the "culture-bound syndromes" to the lack of focus on descriptive categories defined by recurrent symptom patterns. "Culture-bound syndromes" were portrayed as useful to medical anthropologists and cross-cultural psychiatrists as well as to epidemiologists. The former can use them to explain the relationships between symptom patterns and cultural processes, while the latter need to take them into account to accurately and completely measure the full range of psychopathology in a given population.

Following on the methods of Yap and Prince, other authors continued to categorize “culture-bound syndromes” according to contemporary professional paradigms, usually adhering to Prince’s formula of stressing their descriptive aspects. For instance, Ronald Simons emphasized the phenomenologic similarity of groupings of these conditions across historically unrelated and culturally dissimilar times and places, despite differences in the meaning assigned to the experiences and behaviors. He contended that it is feasible to insert the “culture-bound syndromes” into contemporary diagnostic systems, particularly once they are grouped as a cluster of related syndromes. The term *taxon*, borrowed from biologic classification, was employed to denote a grouping of similar items, but without specifying the level of abstraction of the similarity or suggesting that the grouping as presently elaborated is in any way exhaustive or final. In their 1985 text, *The Culture-Bound Syndromes*, Simons and his colleague Charles Hughes used the taxonomic principle to group the cultural concepts based on their phenomenologic similarity across diverse cultural settings. For example, under a “Startle Matching Taxon,” they listed *latah* (from Malayo-Indonesia) and *imu* (from Japan), defined in purely behavioral terms, roughly as follows: someone is startled, after which they say things that are normally socially taboo, talks in a manner locally considered nonsensical and amusing, matches the movements of others, and obeys or defies commands in an exaggerated and stereotyped way. Like Yap and Prince, this kind of classification method focused on a verifiable cluster of behaviors and experiences as the defining elements of the cultural concepts of distress and de-emphasized other aspects of the illness experience, such as causal attributions or social meanings that are dependent on the context in which the cultural concepts arise.

Subsequent Work in Cultural Psychiatry

Ruth Levine and Albert Gaw distinguished several kinds of cultural concepts of distress. “True” “culture-bound syndromes” were organized as recurrent collections of symptoms, which should be classified according to their “predominant symptom or symptoms” and listed as culturally patterned versions of the most relevant DSM category (e.g., *amok* [from Malayo-Indonesia] under the dissociation phenomena and *boufée delirante* [from Western Africa and other settings] under the psychoses). By contrast, “illnesses of attribution” were organized around indigenous etiologic notions (e.g., *susto* [from Mexico and other Latin American regions] in which sudden fright will cause the soul to leave the body, making the individual vulnerable to a variety of ills) rather than discrete symptom clusters. And “idioms of distress” are more general terms that do not have specific correspondences to psychiatric categories (e.g., *nervios* [from Latin America], a general term used as an idiom for stress-related illnesses that can apply to disparate psychiatric disorders, from adjustment problems to psychosis). The authors themselves acknowledged some of the arbitrariness of this approach, however, as well as the technical difficulties involved in parceling out indigenous expressions of psychopathology into their three categories. For instance, in *hwa-byung*, suppression of anger is seen as a causative factor in the development of a somatic symptom cluster (indigestion, dyspnea, fatigue, and the sensation of a mass in the stomach) associated with emotional symptoms (anxiety and depression). The fact that this folk condition is organized around a Korean etiologic notion that ascribes to anger (*hwa*) the capacity to produce illness (*byung*) led these authors to categorize *hwa-byung* as an illness of attribution, whereas it has been described as a cultural syndrome by others due to its usual presentation as a recurrent symptom cluster. In fact, most cultural syndromes contain embedded notions of attribution as organizing elements for the symptom clusters.

Other authors have taken a more mixed approach, combining universalistic views with the notion that each social group creates unique forms of psychopathology that are fundamentally inseparable from the cultural preoccupations that gave rise to them. Notable among these is John Carr, who hypothesized that all forms of psychopathology, including the folk illnesses, represent *culturally authorized final common behavioral pathways*. These are behavioral responses, part of a limited number of behavioral repertoires that individuals use in response to stressful conditions that are common to humanity. Although significantly shaped by cultural contexts, these pathways are mediated by universal mechanisms of learning and cognition. To elucidate the relationship between *amok* and depression, for example, Carr and his colleague Peter Vitaliano contended that these alternate expressions of distress are culturally determined variants of response to comparable life stress conditions. The critical question, in their view, is why some people respond with aggression, some with depression, and still, others cope adaptively to similar aversive events. Culture-specific and idiosyncratic responses to stress result from complex interactions between physiologic processes, predisposing personality styles, cognitive processes, and learned problem-solving skills.

Perspectives from Clinically Applied Medical Anthropology

Ethnographic studies of folk illnesses have approached them as indigenous expressions of distress found within specific cultural contexts. From the anthropologic perspective, these expressions represent local patterns of distress that reference common preoccupations or concerns of particular cultural groups. According to this view, all expressions of distress, including those incorporated into the DSM and ICD systems as diagnostic entities like major depression or schizophrenia, are culturally shaped and are thus equally “culture-bound.” This specificity does not negate the role of biology or personal biography in the formation of any of these concepts, much as universal biologic parameters constrain the features of all languages, yet cultural groups produce multiple variants of human speech.

Medical anthropologists and sociologists, such as Byron Good, Mary-Jo Delvecchio Good, and Margaret Lock, have noted that illness expressions may be organized culturally into nosologies not only on the basis of descriptive, symptom-based approaches but also according to diverse cultural logics that are rooted in distinct ontologies (e.g., the reality of spirit-mediated illness). Many cultural groups use symbolic methods of classification, including the literate medical traditions of the ancient civilizations, in which illnesses are often categorized by analogy to various aspects of the cosmos, social relations, objects in nature, and other contextual factors. The Goods noted that current Western nosologic thinking is predicated on an empiricist theory of language, which posits the existence of a one-to-one correlation between *words* and *things in the world*, with the notion that the latter will remain fixed and unchanging despite evolving social situations. Other cultural groups and periods have considered instead that things in the world are mutable and depend for their meaning on their associations with the social, environmental, and even supernatural context and that human categorizations should reflect this fundamental embeddedness.

Ethnographic work reveals that cultural concepts of distress identify socially constructed patterns of interpretation of suffering that serve pragmatic communicative functions. Clarifying these functions requires what Good and Delvecchio Good called a “meaning-centered medical anthropology” that investigates illness forms from the perspective of *semantic networks*, the clusters of experiences, words, and interpretations that give rise to particular illness forms as modes of experience and expression. These meaning networks are called “semantic” because they reveal the role of interpretation and communication in shaping reality and may include “personal trauma, life stresses, fears and expectations about the illness . . . the metaphors associated with a disease, the ethnomedical theories, the basic values, and conceptual forms, and the care patterns that shape the experience of the illness and the social reactions to the sufferer in a given society.” By taking this contextual approach, the basic organizational principle underlying a particular illness form can be established to assess the degree to which it is descriptive, etiologic, or expressive of particular concerns.

The anthropologist and psychiatrist Arthur Kleinman approached the meaning-centered approach through the concept of an *explanatory model*. This concept refers to the more-or-less systematic explanations of an illness among different actors in any given illness episode (e.g., patient, relative, and caregiver). Kleinman showed how the explanatory model approach could be incorporated into clinical assessment through the use of practical questions about basic aspects of illness experience, addressing perceived etiology, time and mode of onset of symptoms, pathophysiology, course of sickness, and treatment that can be used by clinicians and investigators to clarify the interaction of cultural, physiologic, and psychological elements in an illness episode. The Explanatory Model approach was operationalized by Mitchell G. Weiss in the form of a semistructured instrument for clinical ethnographic research, called the Explanatory Model Interview Catalogue which can be adapted for research on any illness form.

In later work, Kleinman deepened this approach to explore the ways in which the intersubjective processes of interpersonal agency —communication, engagement, transaction, and negotiation—create the flow of *social experience* that links the body–self with collective meanings and social institutions. Individuals transact their lives, including basic processes of health and illness, within relational networks of social experience, which provides “the medium through which moral commitments and negotiations join the corporeal with the collective.” For Kleinman, all of life is lived in “local worlds” that constitute the moral contexts of the illness experience. The importance of local categories of suffering then lies, not primarily in identifying discrete syndromes, but in understanding what is “at stake” for individual patients, families, and communities in terms of power, prestige, material resources, ethnic identity, social order, or survival. This intersection of vital interests and values was well illustrated in Kleinman’s studies of neurasthenia in China, which revealed the importance of relational networks at the levels of supernatural and natural worlds, human interactions, bodily processes, and politics. In traditional Chinese thought, each level affects the functioning of all the other levels. This fundamental interdependence can be seen in the existence of illnesses whose pathology is at the network level, such as the concept of *shenjing shuairuo,* “weakness of the nervous system” in Mandarin. In this cultural syndrome, the internal nervous system (*shenjing*) has become devitalized (weak or run down, *shuairuo*) in correspondence with a demoralized network of interpersonal and societal relations. The symptoms of *shenjing shuairuo* can thus be understood as embodied signs of a devitalized world, both internally and externally, which are linked by contextually located emotions.

Ethnographic fieldwork by Allan Young, Laurence Kirmayer, and their colleagues has shown the complexity of types of knowledge and modes of narration embedded in everyday explanatory accounts of illness. Individuals may link symptoms to explanations based on *chain complexes* because they were closely connected in time and space (contiguity) or as a result of logical shortcuts to denote complex experiences (e.g., metonymy), such as using the term “nerves” to refer to the anatomical and functional aspects of the nervous system, including the feelings and sensations it mediates, much as journalists use the term “Washington” to refer to the entire U.S. government. People also commonly use *illness prototypes*—illness episodes, experienced previously by the person or by others, that provide salient examples used to reason analogically about one’s own experience. In effect, patients’ illness narratives may combine models common to a cultural group, idiosyncratic assemblies of possibly disconnected events, and assumptions based on past experiences. These approaches were operationalized by Danielle Groleau in collaboration with Young and Kirmayer to develop the McGill Illness Narrative Interview, a semistructured interview designed to obtain illness narratives based on (a) contiguity of events, (b) cultural prototypes of the illness for the patient, and (c) explanatory models of illness.

Atwood Gaines’s concept of *sickness history* was another important contribution to the anthropologic understanding of concepts of distress. Gaines’s constructivist argument held that folk and professional notions of sickness evolve historically, in association with other aspects of culture, and that individuals’ understanding and experience of illness are patterned on this accumulated social knowledge. In this way, cultural concepts of distress may reflect distinct cultural histories of societies. Gaines presented the examples of *spasmophilie* (spasmophilia), *fatigué* (fatigued), and *triste/fatigué tout le temps* (sad/fatigued all the time) as sicknesses that are considered physical diseases in France but are unknown in the United States and would actually be considered nonpathologic forms of dysphoria if they came to the attention of American nosologists. Gaines showed how these concepts emerged from a common sickness history that evolved from medieval understandings of dejection as a sin (acedia) to a positive valuation of existential world-weariness as a sign of maturity and a possible and accepted etiology of a physical disease characterized by various states of fatigue, bodily aches, and irritability, among other symptoms.

Much recent work has pursued similar constructivist and historical approaches to both local constructs and the diagnostic entities of the DSM. Combining anthropologic and clinical methods, Roland Littlewood and Maurice Lipsedge traced the historical contexts of constructs like hysteria, medication overdosing, agoraphobia, anorexia nervosa, and shoplifting and linked their phenomenologies and demographic distribution to culturally patterned gender and class constraints prevalent in the United States and United Kingdom in the 19th and 20th centuries. They described these syndromes or idioms as deeply embodied expressions of societal conflicts about which the participants were frequently only marginally aware. The field of *critical medical anthropology* emphasizes the contribution of societal structures of oppression (e.g., racism, class hierarchies, restrictive immigration laws) to the prevalence, phenomenology, and course of mental health conditions, including their comorbidity with physical illnesses. The work of Nancy Scheper-Hughes, Merrill Singer, Philippe Bourgois, and others reveal the impact on the expression of illness of adverse social determinants of health that often go unnoticed or unaddressed in routine clinical care.

A constructivist approach to the ethnopsychologies underlying culturally specific expressions of illness and distress can be elaborated in terms of pragmatic, linguistic, communicative, and interactional processes, including (a) the handling of the gap between experience and expression; (b) the labeling of deviant behavior as voluntary or accidental; and (c) the interpretation of symptoms as symbols or meaningless events. This conceptual grid helps explain why in North America, it is possible to privilege "psychologized" understandings of illness since these arise out of local commitments to individual autonomy, expressiveness, and personal goals. Commitments to other values in other cultural settings may produce different ethnopsychologies.

RELATIONSHIP TO PSYCHIATRIC CLASSIFICATION

At present, most investigators consider cultural concepts of distress to be *orthogonal* to psychiatric diagnoses; that is, the cultural concepts comprise alternate systems of classification that describe in distinct ways the same realm of suffering as professional nosologies. This means that, at the level of global categories, the cultural concepts cannot be easily subsumed into existing psychiatric disorders. Instead, each cultural concept exhibits a *degree of overlap* with a group of Western diagnoses, which can be represented by percentages or by a series of correlations, as with any set of comorbid relationships within professional categories.

Rather than one-to-one, the correspondence between diagnoses and cultural concepts tends to be one-to-many in either direction. Symptoms or behaviors that might be sorted by nosologies like ICD or DSM into several disorders—including across distinct diagnostic clusters—may be included in a single cultural concept, and diverse presentations that might be classified by Western nosologies as variants of a single disorder may be sorted into several distinct concepts by an indigenous diagnostic system. Therefore, cases identified by a single indigenous label often cut across diagnostic clusters. For example, the symptoms of *amok* in Malayo-Indonesia exhibit diverse combinations of anxiety, depressive, somatic, dissociative, psychotic, and characterologic symptoms. The obverse is also true; homogeneous psychiatric cohorts will appear locally diverse. In Latin America, for example, different cases of major depression may be labeled as distinct conditions according to locally salient features of phenomenology, causation, context, or the healing ideology of the practitioner. These include cases of *padecer de los nervios* (suffering from nerves), *susto, fallo mental* (mental failure), *cólera* (rage), *facultades* (faculties, for *espiritistas*), and demonic influence (e.g., for Pentecostals), among other categories. Cross-cultural variation in the assembling of the “metastructure” of psychiatric disorders—the fact that different nosologies distribute individual illness categories into diagnostic groupings in distinct ways—as well as the changes over time in this classification process (e.g., the reassigning of post-traumatic stress disorder [PTSD] away from the anxiety disorders and into a new category of trauma and stressor-related disorders) reveals the cultural underpinnings and social evolution of any classification system.

The lack of one-to-one correspondence between diagnoses and cultural concepts is intensified by the substantial phenomenologic variation that exists even across individuals who endorse the same cultural concept of distress. A given instance of *ataque de nervios*, for example, may present as an acute dissociative episode, with seizure-like symptoms in a state akin to trance, for which amnesia is later reported. Other episodes may resemble cases of panic disorder, albeit with more suicidal and aggressive features, such as breaking things and verbally or physically attacking those around them. Some episodes may present as uncontrollable fits of nervous laughter, which evolve into bouts of crying and screaming and subsequent depression. Others may constitute episodes of acute grief at the funeral of a loved one without any pathologic sequelae. What links all four instances in the popular classification system in the Latino Caribbean is their presentation as acute episodes of uncontrolled emotionality, precipitated by interpersonal stressors that overwhelm the person’s coping ability, usually in a context of insufficient social and familial support. Yet from the psychiatric perspective, these presentations would be considered phenomenologically distinct. Clinical translation between the cultural concepts and psychiatric diagnoses must therefore be conducted at the level of the *individual*, describing for that person the overlap between both systems of classification.

Two other differences ensure that cultural concepts of distress and contemporary psychiatric nosologies overlap only partially as global systems. First, Western nosologies are based at present nearly exclusively on formal, descriptive definitions of psychopathology. Most indigenous nosologies, however, distinguish pathology from normality at least as much based on contextual characteristics and causal theories as on descriptive ones. These include assessments regarding the appropriateness of the symptom in the specific setting and time in question, the relative sufficiency of the precipitating stressors, and the nature and quality of the human, and sometimes spiritual, relationships of the sufferer.

Second, while psychiatric nosologies contain only pathologic conditions, cultural concepts of distress may apply to a wide range of severity, including presentations that do not meet DSM criteria for any mental disorder. As described in DSM-5-TR, “an individual with acute grief or a social predicament may use the same idiom of distress or display the same cultural syndrome as another individual with more severe psychopathology.” The indigenous classification systems thus tend to describe broader realms of behavior and adaptation than the narrower purview of psychiatric classification.

Despite the differences between local and professional nosologies, information on cultural concepts of distress is vital to clinical care for at least five reasons, which are outlined in DSM-5-TR. First, the use of cultural concepts in screening instruments or in reviews of systems may facilitate the identification of patients’ concerns and improve the detection of psychopathology, as individuals may be more familiar with these local concepts than with professional terminology. For example, including questions about “thinking too much” can elicit illness accounts that may not be evoked by typical symptom queries. Second, clinician awareness of cultural concepts may prevent misdiagnosis, as cultural variation in symptoms and explanatory models associated with these concepts may lead clinicians to misjudge the severity or a problem or assign the wrong diagnosis (e.g., socially warranted suspicion may be misunderstood as paranoia). Third, cultural variations in symptoms and attributions related to the cultural concepts can be associated with particular features of risk, resilience, and outcome. Clinical exploration of cultural concepts can elicit information on the role that specific contexts play in symptom development, course, and coping strategies. For example, the interpretation of trauma-related symptoms as spiritual afflictions may prompt the use of effective forms of healing and self-coping, such as church-based social support or meditation. Conversely, the presence of a cultural syndrome may be associated with specific types of risk and morbidity —for instance, endorsement of *ataque de nervios* is related to higher suicidal ideation after adjusting for psychiatric diagnosis, demographic characteristics, and trauma exposure. Fourth, understanding the messages being communicated by the patient in the idiom of the cultural concept (e.g., social isolation, being overwhelmed) can improve clinical rapport and engagement and help patients and clinicians select a mutually acceptable treatment plan. Some patients come to care expecting to be treated for the cultural concept of distress and may become disengaged if their understanding is not negotiated effectively with that of the clinician. Fifth, cultural concepts can illustrate key psychological mechanisms which need to be understood and addressed to improve clinical efficacy. For example, catastrophic cognitions can contribute to symptom escalation into panic attacks in culturally specific ways, such as through the cultural salience of certain symptom–outcome pairings. The popular linkage made in the United States between chest pain and heart attacks can precipitate panic episodes when chest pain occurs as part of an anxiety reaction but is interpreted as a harbinger of myocardial infarction. In Cambodia, by contrast, the association is more commonly between neck soreness and stroke due to the interpretation of panic attacks as signs of disordered *khyâl* circulation, a wind-like substance important in local ethnophysiology. Being aware of these linkages can help clinicians identify how cultural interpretations can constitute the actual “cogs” of the disorder.

Attending to these aspects of the relationship between cultural concepts and psychiatric diagnoses can help clinicians meet patients' treatment expectations. Moreover, the inclusion of cultural concepts in research can reveal patterns of comorbidity—between the diagnoses linked by the idiom or cultural syndrome—and clarify the epidemiologic distribution of cultural features of the illness. Information on the cultural concept of distress can be incorporated into the process of diagnostic assessment and treatment planning by eliciting an account of the person's definitions of pathology, idioms of distress, unique symptoms, contextual embeddedness, and other associated factors that are contained in the experience of the cultural concept. A systematic approach for eliciting and documenting this information is the Cultural Formulation Interview, which is included in DSM-5-TR and will be described later in this chapter.

EXAMPLES OF CULTURAL CONCEPTS OF DISTRESS

Only a few of the thousands of cultural concepts of distress have received sustained research attention integrating cultural and psychiatric research methods. This chapter focuses on three concepts from diverse cultural regions that have been the object of intensive research, including studies of their relationship with psychiatric diagnoses: *ataques de nervios*, possession syndrome, and *shenjing shuairuo*. Other examples of cultural concepts of distress are included in the DSM-5-TR chapter on culture and psychiatric diagnosis. These three examples and the few cultural concepts listed in the DSM-5-TR are meant to be illustrative of the wide gamut of folk illness categories relevant to particular populations or contexts.

Ataque de Nervios

Ataque de nervios ("attack of nerves" in Spanish) is a syndrome indigenous to various Latin American cultural contexts, notably those of the Latino Caribbean (Puerto Rico, Cuba, and the Dominican

Republic). It has received considerable attention in the psychiatric and anthropologic literature since the mid-1950s, mostly in Puerto Rican communities in the Island and in the United States. Early reports of presentations in male recruits to the U.S. armed forces by military psychiatrists in Puerto Rico mislabeled *ataque* as “the Puerto Rican Syndrome.” In keeping with psychoanalytic theory at the time, these investigators attributed the *ataque* to specific indigenous child-rearing practices allegedly predisposing to the hysterical discharge of aggressive drives. The power relationships embodied in the military, and the colonial situation represented by the presence of the U.S. armed forces in Puerto Rico, were usually left unanalyzed in these discussions.

Subsequent research focused on the role of *ataque* as a culturally recognized and sanctioned expression of emotion, mostly by women, that should be understood as a form of communication about family relationships in contexts of social and economic powerlessness. These investigations located the psychosocial etiology of *ataques* in experiences of acute family discord and individual suffering—such as grief, fear, and anger—in the context of collective distress—such as migration, poverty, and discrimination. This research emphasizes the role of *ataque* as an idiom of distress in Latino communities and illustrates how the same folk term may represent different kinds of cultural concepts—that is, a folk category such as *ataque* is a cultural syndrome that functions as an idiom of distress.

Contemporary clinical–anthropologic studies have further clarified the phenomenology, cultural meanings, epidemiology, and relationship of *ataque* to psychiatric diagnoses. The work of Peter Guarnaccia and Roberto Lewis-Fernández has identified loss of control of emotions, behaviors, and a social world out of control as the core experiences of *ataques de nervios*. Their work suggests a popular nosology based on various categories of *nervios* (nerves), several of which require more systematic research.

Phenomenology. Guarnaccia, Glorisa Canino, and colleagues surveyed a representative community sample of Puerto Rico residents (*N* = 912) for the presence of *ataque de nervios*. Qualitative coding of general descriptions of the syndrome among a subsample of 145 respondents in this larger study—of whom 77 were *ataque* sufferers—yielded four domains of *ataque* experience. These were emotional expressions, bodily sensations, action dimensions, and alterations in consciousness. Based on these data, a prototypical *ataque de nervios* can be described as composed of the elements described in Table 29.1–2.

Epidemiology. A 1987 survey by Guarnaccia and Canino in a probability sample of the population of Puerto Rico (*N* = 912), using the Diagnostic Interview Schedule (DIS), found a weighted lifetime prevalence of *ataque* of 13.8%, making it one of the most frequently reported syndromes in the Island. Twelve percent of respondents had *ataques* that met DIS criteria for a positive psychological syndrome—that is, the *ataque* met severity criteria (consulting a physician or other professional, taking medications, or reporting some functional impairment) and could not be explained as resulting from physical illness or substance use. The Latino cohort of the 2002 to 2003 National Latino and Asian American Survey, led by Margarita Alegría, also obtained lifetime *ataque* prevalence in a nationally representative sample of U.S. Latinos using a diagnostic module that followed the format of the Composite International Diagnostic Interview. *Ataque* prevalence was 10.9% in Puerto Ricans, 6.2% in Cubans, 6.0% in Mexicans, and 5.4% in other Latinos, suggesting both that *ataque* is common across Latino subgroups and that it is more prevalent among Puerto Ricans.

Table 29.1–2.
Phenomenology of *Ataque de Nervios* (Attack of Nerves)

a. Exposure to a frequently sudden, stressful stimulus, typically eliciting feelings of fear, grief, or anger, and involving a person close to the subject, such as a spouse, child, family member, or friend. Severity of the trigger ranges from mild to moderate (i.e., marital argument and disclosure of migration plans) to extreme (i.e., physical or sexual abuse and acute bereavement).
b. Initiation of the episode immediately upon exposure to the stimulus or after a period of brooding or emotional “shock.”
c. Once the acute attack begins, rapid evolution of an intense affective storm characterized by a primary affect usually congruent with the stimulus (such as anger, fear, grief) and a sense of loss of control (*emotional expressions*).
d. These are accompanied by all or some of the following:
    i. *Bodily sensations*: trembling, chest tightness, headache, difficulty breathing, heart palpitations, heat in the chest, paresthesias of diverse location, difficulty moving limbs, faintness, blurred vision, or dizziness (*mareos*).
    ii. Behaviors (*action dimension*): shouting, crying, swearing, moaning, breaking objects, striking out at others or at self, attempting to harm self with nearest implement, falling to the ground, shaking with convulsive movements, or lying “as if dead.”
e. Cessation may be abrupt or gradual, but it is usually rapid, and often results from the ministration of others, involving expressions of concern, prayers, or use of rubbing alcohol (*alcoholado*). There is return of ordinary consciousness and reported exhaustion.
f. The attack is frequently followed by partial or total amnesia for the events of the episode, and descriptions of the following for the acute attack: loss of consciousness, depersonalization, derealization, mind going blank, and/or general unawareness of surroundings (*alterations in consciousness*). However, some *ataques* appear to exhibit no alterations in consciousness.

Guarnaccia and Canino found a prevalence of *ataques de nervios* of 8.9% in a representative sample of youth in Puerto Rico aged 4 to 17 years (*N* = 1,897). A second, community-based probabilistic sample of Puerto Rican children aged 5 to 13 years in San Juan (*N* = 1,353) and the South Bronx (*N* = 1,138), led by Héctor Bird and Canino, found a lifetime prevalence of *ataque* of 5.4% and 4.3%, respectively.

Risk factors for *ataque* span a range of social and demographic characteristics. Adults reporting an *ataque de nervios* in Puerto Rico are more likely to be female, over the age of 45, with less than a high school education, formerly married (i.e., divorced, widowed, or separated), and out of the labor force. However, 10% of the men in the sample reported an *ataque de nervios*, which indicates that some men express their distress through this cultural syndrome. Among Latinos in the United States—and like other psychiatric disorders—*ataque* is positively correlated with higher acculturation (e.g., U.S. birth and English proficiency), suggesting higher psychiatric vulnerability in this postmigration population. The level of neighborhood violence is also a risk factor for *ataque*. Risk factors among Puerto Rican youth differ somewhat by the site (New York vs. Puerto Rico) and by study; they include older age (especially adolescence in girls), single-parent household status, exposure to violence, stressful life events in the past year, and higher perceived poverty. In one of the community-based studies in Puerto Rico, multiple regression analysis also revealed an independent positive association between *ataque* status in the youth and a family history of the syndrome, suggesting the possible influence of intergenerational behavioral modeling.

Lewis-Fernández and colleagues estimated the lifetime prevalence of *ataque de nervios* in two clinical settings serving adult Puerto Rican psychiatric outpatients. The prevalence rate of *ataque* was 55.1% among immigrant Puerto Ricans (*N* = 89) in an outpatient Latino mental health clinic in New England and 51.5% among outpatient referrals to psychiatry (*N* = 97) in three rural Puerto Rico primary care clinics. In both sites, the gender distribution of *ataque* sufferers was nearly equivalent: 56.3% female in New England and

54.8% female in Puerto Rico. However, women were twice as likely to seek mental health care in both settings. Finally, in these rural clinics for low-income residents in Puerto Rico, *ataque* rates also did not differ by age cohort or level of education. All social groupings among this disenfranchised sector of the population appeared equally at risk for *ataque* episodes. Two large opportunity samples of Puerto Rican or Latino primary care patients in the Northeastern United States separately found a lifetime *ataque* prevalence of 26% to 28%, confirming the elevated prevalence of the syndrome in clinical populations. An elevated clinical prevalence was also observed among youth: lifetime prevalence of *ataque* was 25.5% in a representative sample of youth utilizing the mental health system in Puerto Rico (*N* = 767).

Precipitants/Etiology. Prototypically, *ataque de nervios* was linked by sufferers to an acute precipitating event or to the summation of many life episodes of suffering brought to a head by a trigger that overwhelmed the person's coping ability. In the Puerto Rico community sample studied by Guarnaccia and Canino, the first experience of an *ataque* was closely tied to a precipitating event. In 92% of cases, the *ataque* was directly provoked by a distressing situation, and 73% of the time, it began within minutes or hours of the event. A majority of first *ataques* (81%) occurred in the presence of others, as opposed to when the sufferer was alone and led to the person receiving help (67%). Finally, unlike the typical experience of persons with panic disorder, most respondents reported feeling better (71%) or feeling relieved (81%) after their first *ataque*. These findings suggest that first episodes of *ataque de nervios* are closely tied to the interpersonal world of the sufferer and that they result in an unburdening (*desahogarse*) of one's life problems, at least temporarily. One function of the *ataque* may be to rally the person's social network, including reversing the current interpersonal arrangement in favor of the sufferer. In one primary care study of 508 Latino adults in the Northeastern United States, partner support was associated with significantly lower odds of *ataque de nervios*, while emotional strain from one’s partner, family, or friends was each associated with higher lifetime prevalence.

Specific precipitants were varied but typically consisted of acutely stressful experiences resulting in feelings of loss, grief, anger, or fear. In general, the most frequent triggers identified by the Puerto Rico epidemiologic study included arguments with a spouse or other close relatives, the drunkenness of a family member, an accident that threatened the life of a person close to the subject, death of a close family member, natural disaster, or the passing of a fright (*susto*) through a person. Certain precipitants were obviously traumatic and were correlated with globally elevated prevalence rates in exposed subpopulations. For example, this survey showed a higher prevalence of *ataques* among those severely affected by a recent natural disaster (19%) than among nonexposed individuals (14%), and intermediate rates among subjects only moderately affected (16%). However, *ataques* have been reported after apparently minor stressors, such as family squabbles, adolescent romantic disappointments, or peer razzing. In these cases, the presence of underlying interpersonal or sociocultural conflicts of longer duration—such as a marital argument or abuse, migration, poverty, discrimination, precarious employment/source of income, or family discord—has been invoked by investigators to explain the enhanced *ataque*-producing power of superficially mild triggering experiences.

Studies in psychiatric outpatient samples of the relationship between childhood trauma exposure and the presence of *ataques* have yielded conflicting results. Daniel Schechter and colleagues found a positive association, whereas Lewis-Fernández and colleagues did not. In this latter study, childhood traumatic exposure was high across Puerto Rican psychiatric outpatients regardless of *ataque* status, suggesting that, in a clinical sample, factors in addition to childhood trauma accounted for this condition. In a different psychiatric outpatient sample of diverse Latino subethnicities with a history of traumatic exposure ($N = 202$), greater trauma severity was not associated with a higher frequency of *ataques*. More research is needed in samples with and without traumatic exposure to clarify the association between this risk factor and *ataque*.

Relationship to Psychiatric Diagnoses. Psychiatric evaluations of *ataque* sufferers have usually revealed the presence of some form of psychopathology. In the Puerto Rico adult epidemiologic survey, 63% of respondents who reported an *ataque de nervios* met diagnostic criteria for one or more psychiatric disorders measured in the DIS. This suggests that rather than being an expression of nonpathologic distress, most *ataques* are associated with a psychiatric disorder. Overall, *ataque* sufferers had 4.35 times the odds of meeting criteria for a psychiatric diagnosis than those who did not report the cultural syndrome. At the same time, 37% of *ataque* sufferers did not fulfill DIS criteria. This suggests that a substantial minority of respondents with this syndrome are describing milder levels of distress using the idiom of the cultural syndrome; alternatively, they may be reporting psychiatric disorders not assessed by the study instrument.

It is important to note that *ataques* do not correlate exclusively with one psychiatric diagnosis and cannot be treated as simply a culturally shaped version of a specific psychiatric disorder. The combined research findings support a strong relationship between *ataques de nervios* and mood and anxiety disorders in youth and adults, as well as disruptive disorders in youth and dissociative disorders in adults. Of particular concern is the strong relationship between *ataque* and suicidal ideation (odds ratio [OR] = 6.2) and attempts (OR = 8.1). In a U.S. nationally representative community sample of diverse Latino subethnicities (N = 2,554), lifetime history of *ataque de nervios* was independently associated with suicidal ideation (OR = 2.4) after adjusting for multiple covariates, suggesting that the cultural syndrome conveys additional risk beyond recognized risk factors such as psychiatric diagnosis and traumatic exposure. During *ataque* episodes, sufferers may make goal-directed or ill-formed attempts at self-harm, increasing the morbidity (and occasional mortality) of the syndrome.

While early speculation by psychiatrists focused on whether an *ataque* constitutes a culturally shaped version of a panic attack, Lewis-Fernández and colleagues systematically demonstrated that only 36% of best-remembered *ataques* met DSM-IV-TR criteria for panic attacks and 17% met criteria for panic disorder. What mainly distinguished *ataques* that did not fulfill panic disorder criteria from those that did include the cuing of episodes by precipitants, symptom crescendo longer than 10 minutes, and feeling relieved rather than frightened after an episode. However, features that were common to both *ataque* and panic disorder included recurrence of attacks, the experience of fear or nervousness during an episode, and anticipatory or behavioral sequelae afterward. Patients who described feeling relieved after their *ataque* also reported significantly fewer panic symptoms than those without this feeling.

Specific characteristics of *ataque* profiles may be associated with distinct psychiatric diagnoses. A study by Ester Salmán and Michael Liebowitz of a sample of Puerto Ricans and Dominicans (N = 156) attending a Manhattan research clinic that was enriched with individuals suffering from *ataque de nervios* (N = 109) found a link between *ataque* features and specific psychiatric disorders. Among those participants in the study whose *ataques de nervios* were characterized by intense fearfulness and feelings of asphyxia and chest tightness, diagnoses of panic disorder were more common. Those whose *ataques* were more dominated by the emotion of anger and aggressive behavior, such as breaking things, were more likely to meet the criteria for co-occurring mood disorders.

Ataque is strongly associated with dissociative experiences. The severity of *ataques*, assessed as lifetime number of episodes, is positively and independently related to self-reported dissociative symptoms, measured with the Dissociative Experiences Scale (DES), and to clinician-diagnosed dissociative disorders, assessed with the Structured Clinical Interview for DSM-IV. Further evidence of the relationship between *ataque de nervios* and dissociation comes from the significant correlation between *ataque* and other Latino cultural concepts that are also related to higher dissociative capacity on the DES. One of these idioms is characterized by "pseudohallucinations" in several sensory modalities—hearing voices or noises, seeing shadows (*celajes*), and feeling presences when alone—that are popularly attributed to spiritual influences. In an outpatient U.S. Latino sample with high traumatic exposure (N =

202), *ataque de nervios*, pseudohallucinations, and a third Latino folk illness, “being currently ill with nerves,” were all significantly associated with each other and with elevated DES scores, forming a nexus of “nerves”-related cultural concepts linked to dissociative capacity.

In youth, *ataques de nervios* are associated with research diagnoses of mood, anxiety, and disruptive disorders. In a Puerto Rico community-based survey, youth aged 4 to 17 years with *ataques* had ORs of 7.6 for any mood disorder, 5.0 for any anxiety disorder, and 3.1 for any disruptive disorder, compared to youth who did not report an *ataque*. In the clinical sample of the same study, the equivalent ORs were 5.8, 3.5, and 2.0, respectively. Youth with *ataques de nervios* had 4.3 times the odds of meeting criteria for any psychiatric diagnosis in the community and 2.3 times the odds of doing so in the clinical setting. Moreover, impairment was likewise elevated among youth with *ataques* compared to those without, with an OR = 5.6 in the community and OR = 2.1 in the clinical sample. In another study, after adjusting for demographic characteristics and the level of psychosocial stressors, children aged 5 to 13 years with *ataques* in San Juan and the South Bronx had nearly 3 times higher odds of meeting criteria for an anxiety or a depressive disorder, nearly 2 times the odds of a disruptive disorder, and 1.4 times higher odds of global impairment.

Additional Clinical Features. Assessment of the phenomenologic features of *ataques de nervios* experienced by the 77 respondents in the Puerto Rico epidemiologic study with at least one lifetime *ataque* who were reinterviewed by Guarnaccia revealed that the largest number (28%) had suffered only one such episode over the course of their life. Thirteen percent had two *ataques*, 8% experienced three, 9% had four, and 4% had five. The second-largest number (23%) comprised those who had more than five lifetime *ataques*, while 15% were unsure about the total. The bimodal distribution of *ataque* frequency (one vs. six or more) suggests that persons with frequent episodes constitute a minority of *ataque* sufferers. This raises the question of whether having an isolated *ataque* in response to a markedly stressful event—as opposed to recurrent *ataques*—represents a relatively normal expression of distress in Latino communities. However, the lifetime presence of one or more *ataques* is associated with disability and outpatient mental health service use among U.S. Latinos, even after adjusting for demographic characteristics, traumatic exposure, psychiatric diagnoses, and chronic health conditions. This suggests that experiencing an overwhelming reaction such as *ataque* may not be without clinical consequences.

The association between *ataque* and a sense of loss of control and of being overwhelmed highlights the importance of the relationship between the cultural syndrome and other behaviors associated with acute emotional dysregulation. Most concerning among these is the strong association noted earlier between *ataques* and suicidal ideation and attempts. Other behaviors related to this acute dysregulation include loss of aggression control, expressed as attacks on people or property, and dissociative experiences, both of which are related to the acute *ataque* experience.

Specific Cultural Factors. The complex relationship between *ataque de nervios* and psychiatric diagnosis may be further clarified by placing *ataque* within a broader Latin American popular nosology of *nervios* (nerves) of which it is part. Experiences of adversity are linked in this nosology to ensuing “alterations” of the nervous system, which result in its impaired functioning, including the peripheral nerves. This quasi-anatomical damage is evidenced in emotional symptoms, including interpersonal susceptibility, anxiety, and irritability, as well as in physically mediated symptoms, such as trembling, palpitations, and decreased concentration. Certain people are considered at greater risk for *nervios* when confronted with adversity, such as those with an inherited predisposition or with stressful perinatal experiences (e.g., mother’s alcoholism). But in all cases, the possibility that sufficient adversity will result in an alteration of the nervous system leads to the popular view that persons faced with experiences of suffering should “control” their reactions to avoid nervous system damage. This usually takes the form of dampening the emotional impact of the experience so that the nerves will not be substantially “altered.”

Clinical and popular discussions about *nervios* and *ataques* in Latino communities thus typically reference cultural views about the importance of remaining in control and of facing adversity with *tranquilidad* (equanimity) regardless of the severity of a stressor. *Ataque de nervios* can therefore be understood as a behavioral syndrome stemming from general notions and practices about how to behave, rather than a peculiar cultural expression of a particular diagnosis. As a marker of being overwhelmed, *ataques* are associated with various forms of psychopathology, but the *ataque* itself cannot be reduced to just the psychiatric disorder; rather, it is a marker of the severity of the stressor, of the person’s reaction style, of their level of support, and of the resulting distress and pathology.

Aspects of Caribbean Latino cultural contexts may promote the observed association between *ataque de nervios* and dissociation since the appearance of dissociation during the *ataque* may be enhanced by the cultural perception that a mature social actor needs to be always in control. Dissociation may be particularly promoted by the popular view that when out of control, the person is no longer “themself” (“*ése no era yo* [that was not me]”). The uncontrolled behavior is disavowed as not being part of the self. This may facilitate the emergence of dissociative splits that allows the person to distance themself from the moral consequences of the *ataque* episode. Similarly, prevalent cultural concerns about the psychological and social dangerousness of strong emotions may reinforce the role of anxiety sensitivity in *ataque*. Individuals with high anxiety sensitivity appear to be at elevated risk for *ataque* prevalence and severity, possibly due to positive (adverse) feedback effects on *ataque* onset or persistence.

Community rates of *ataque* in adults are significantly higher in women than in men. The female-to-male ratio is 2:1 in Puerto Rico and 3:2 among U.S. Latinos. The gender difference may appear during adolescence since it is absent in both South Bronx- and San Juan-based representative samples of Puerto Rican children aged 5 to 13 years but is found in a Puerto Rico-wide representative sample of youth aged 4 to 17 years. The gendered nature of *ataque* in adults has been attributed in part to traditional gender roles pressuring Latina women to restrict some emotions in social situations, especially around sexual, aggressive, or assertive needs. The gender-specific cultural directive to remain “in control” may predispose women to express suppressed feelings in situations of interpersonal conflict as acute paroxysms of emotionality and uncontrolled behavior, which can later be disavowed as unwilled or even ego-alien.

Treatment. No treatment studies of *ataque de nervios* have been conducted. Typical treatment in community settings involves, first, ensuring the safety of the person and those around them, given the association between *ataque*, suicidality, and uncontrolled aggressivity. “Talking the person down” is usually helpful, accompanied by expressions of support from relatives and other loved ones; the use of rubbing alcohol (*alcoholado*) to help calm the person is a culturally prescribed way of expressing this support.

“Telling the story” of what led to the *ataque* usually constitutes the principal therapeutic approach in subsequent stages of treatment. Since one of the main functions of the attack is to communicate a feeling of being overwhelmed, indicating receipt of the message and the desire to offer support are usually perceived as therapeutic. The person should be allowed to set the pace of disclosure and to give enough details and circumstances to feel “unburdened” (*desahogado[a]*).

In the case of single or occasional *ataques* in the absence of a psychiatric diagnosis, brief follow-up is usually sufficient. This may be discussed with the patient and the family as a way of ensuring a full return to the previous healthy state. For recurrent *ataques*, treatment depends on the associated psychopathology, the nature of the precipitants—including traumatic exposure—the degree of family conflict or support, the social context, the previous treatment experiences, and the patients’ and family’s expectations, among other factors.

Psychotherapy is typically the mainstay of treatment, usually with a family focus, given the usual source of the overwhelmed behavior in the interpersonal milieu. A social activism stance by the therapist that acknowledges the origins of adversity among low-income Latinos in socioeconomic disenfranchisement and ethnoracial discrimination is helpful in addressing the sociocultural roots of *ataque de nervios.* Pharmacotherapy may also be useful in the treatment of *ataque*-related psychopathology; the primary emphasis should be placed on treating the underlying disorder. Given the slow crescendo of many *ataques*, judicious use of short-acting benzodiazepines also has a role in helping abort an impending episode. However, this should not be the main form of treatment for recurrent *ataques* since it only forestalls the principal function of the syndrome as a mode of communication.

Possession Syndrome

Involuntary possession trance states are very common presentations of emotional distress around the world. Cognate experiences have been reported in very diverse cultural settings, including India, Sri Lanka, China, Japan, Malaysia, Niger, Uganda, Southern Africa, Turkey, the United States, Canada, Haiti, Puerto Rico, and Brazil among others. *Possession syndrome* is an umbrella English-language term used in the clinical and anthropologic literature on South Asian presentations of involuntary possession trance that encompasses conditions with many different names in regional languages and dialects of India and Sri Lanka. These presentations are seen as a severe affliction by the person's cultural group because they are involuntary, cause distress, and do not occur as a normal part of a broadly accepted collective cultural or religious ritual or performance.

DSM-IV-TR did not list possession syndrome or any other pathologic possession trance condition as a "culture-bound syndrome." Instead, this class of clinical presentations was included in Appendix B of DSM-IV and DSM-IV-TR under the proposed category of dissociative trance disorder (DTD), one of the "criteria sets and axes provided for further study.” Categories in this section of the manual fell short of full inclusion in DSM-IV as specified diagnoses but were included in the appendix to stimulate research for possible incorporation in DSM-5. Beyond its global prevalence, another reason why involuntary possession trance was given a special status in DSM-IV was its inclusion in ICD-10 as a subtype of trance and possession disorders (F44.3), a specified diagnosis under the dissociative (conversion) disorders in the section on neurotic, stress-related and somatoform disorders. The goal of making the ICD and DSM nosologies more compatible helped to promote consideration of involuntary possession trance in DSM-IV and DSM-5. In ICD-11, the main characteristics of possession trance disorder include the following:

A disorder characterized by a trance state in which there is a marked alteration in the individual’s state of consciousness resulting in significant distress or impairment. During the trance episode, the individual’s normal sense of personal identity is replaced by another identity attributed to an external “possessing” spirit, power, deity, or another spiritual entity, and the individual engages in behaviors or movements that are experienced as being controlled by the external agent. The trance episodes must be recurrent or, if single, last for at least several days and be involuntary, unwanted, and not accepted as a part of a collective cultural or religious practice. The episodes must also not be due to the effects of a substance or medication or be better explained by another disorder of the nervous system and must result in clinically significant distress or impairment.

In DSM-5 and DSM-5-TR, pathologic possession trance was included under dissociative identity disorder, leading to its redefinition as two variants: possession form and nonpossession form. This incorporation was due largely to the phenomenologic similarity between pathologic possession trance and dissociative identity disorder. Both are disorders of identity alteration that occur during a dissociative state of consciousness and are characterized by involuntary stereotyped and culturally determined behaviors or movements that are experienced as being under the control of an alternate identity. Among 119 possessed patients who visited traditional healers in Uganda, for example, Marjolein Van Duijl and colleagues observed involuntary shaking of the head or parts of the body (attributed to spirits), talking in a different voice—which others recognized as the voice of a specific spirit—and feeling influenced by unidentified forces that cause unusual behavior. Nearly half of the possessed patients mentioned something holding the body, so they were unable to move or speak. Among traditional healers and local health workers in these villages, case vignettes of dissociative identity disorder were always attributed to spirit possession. In many healing traditions, patients or practitioners learn how to become possessed to serve as a vehicle for a healing deity or spirit, which then is able to diagnose and treat a wide range of health conditions and life predicaments. Such deliberately sought and socially regulated "executive possession" should not be seen as pathologic.

In both possession and nonpossession types of pathologic identity alteration, sufferers can report full or partial amnesia for the events of the dissociative state. Moreover, a history of exposure to traumatic experiences is common in both conditions. In the Uganda study, for example, individuals with pathologic possession scored higher on the Harvard Trauma Questionnaire and the Traumatic Experiences Checklist than 71 randomly selected healthy inhabitants of the same villages, and scores on both trauma scales were strongly correlated with symptoms of dissociation ($r = 0.64 - 0.69$) and pathologic possession ($r = 0.51 - 0.64$). To rule out the misdiagnosis of nonpathologic possession states, dissociative identity disorder requires the presence of amnesia, as well as the fact that the identity alteration states (a) be associated with clinically significant distress or impairment (criterion C) and (b) not be a normal part of a broadly accepted cultural or religious practice (criterion D). In addition, the text states that the pathologic possession episodes "present recurrently [and] are unwanted and involuntary."

Possession trance can occur in epidemic form. For example, in recent years, there have been many episodes of groups of school children in Nepal experiencing possession episodes characterized by involuntary movements and speech. Individuals who are affected in these epidemics may be high on measures of hypnotizability and more likely to have a past history of peritraumatic dissociation.

The possession-form and nonpossession-form variants of dissociative identity disorder are not mutually exclusive: individuals in one cultural setting may experience either form, and a single individual may experience both forms. Typically, in nonpossession-form dissociative identity disorder, the alternate identities are experienced as *internal* aspects of the person, as other identity states. In possession-form dissociative identity disorder, by contrast, the alternate identities are typically experienced as *external* possessing agents, usually of supernatural or spiritual origin (e.g., demons and spirits). The main contrast between these two forms of dissociative identity disorder appears to be shaped by the patient's sociocultural milieu. That is, possession-form dissociative identity disorder arises in contexts where possessing agents are considered ontologically "real" (e.g., some fundamentalist religions in the United States or traditional South Asian cultures), as opposed to nonpossession-form dissociative identity disorder, which draws its alternate identities from individualistically focused experiences, such as stages of life (e.g., childhood) or biographical roles in the patient's life (e.g., protector and perpetrator).

The decision to subsume pathologic possession trance states under dissociative identity disorder is not without controversy, however. It is an open empirical question, for example, what proportion of pathologic possession states meets dissociative identity disorder criteria, which include substantial amnestic gaps in addition to identity alteration. The nonpossession form of the disorder is typically a chronic condition arising in childhood in the wake of prolonged and severe trauma. Some pathologic possession trance states follow this pattern, but others are triggered by nontraumatic stressors, such as romantic disappointments and work or religious conflicts, and present as single or recurrent episodes rather than persistent conditions. However, the absence of systematic assessment of childhood trauma precludes excluding this exposure from the genesis of dissociative disorders in these populations. With accumulating evidence, the placement of pathologic possession in psychiatric nosology will be further clarified. Dissociative identity disorder may prove to be a heterogeneous condition, encompassing a gamut of identity alteration symptoms, courses, and precipitants across cultural contexts. Alternatively, the evidence may favor the inclusion of pathologic possession trance as a separate diagnostic entity in DSM, as in ICD-11.

Pathologic possession trance states thus constitute local presentations that are farther along the process of incorporation into Western psychiatric classification systems than the usual cultural concepts of distress. Their truly global prevalence makes them candidates for "universalistic" disorders, even though they are considered uncommon in modern Euro-American and European societies. Their status is still unresolved, however, given their inclusion as a variant of dissociative identity disorder in DSM-5 and as a distinct disorder in ICD-11. Moreover, since local presentations of pathologic possession trance vary substantially across cultural settings, possession syndrome is discussed here as a cultural concept of distress, focusing on its specific characteristics in South Asia.

Phenomenology. It is important to distinguish at the outset between possession syndrome, as an instance of possession trance, and the broader category of possession. The latter term refers to a general ideology describing the full range of direct spirit influences on human affairs, including effects on physical, psychological, spiritual, social, and ecologic realms. By contrast, as a subset of general possession experience, "possession trance" refers to specific alterations in consciousness, memory, behavior, and identity attributed to direct spirit influence. In addition to pathologic possession trance states, South Asian cultures authorize multiple examples of normal possession and possession trance in religious, healing, and other ritual contexts. When voluntary and normative, these states are typically seen as instances of religious devotion, mystical ecstasy, social commentary, asceticism, interpersonal relations, existential reflection, and the study of consciousness. This chapter discusses possession syndrome as a pathologic entity with an established phenomenology, that is, as a special case in the general continuum of etiologic ideas regarding possession illnesses. A prototypical episode is composed of the elements in Table 29.1–3.

Data on the epidemiology, precipitants, and associated psychopathology of subjects with possession syndrome in South Asia is limited by methodologic considerations. These include a lack of representative community samples and nonsystematic definitions of the syndrome, which shows considerable regional variation.

Epidemiology. Period-prevalence rates among psychiatric cohorts in India conducted mostly via chart review have been variously reported as 0.2% over 1 year, 0.06% over 7 years, and 0.93% over 12 years. Patients with dissociative disorder accounted for 0.15% to 1.5% of outpatients and 0.15% to 1.2% of inpatients at a major psychiatric facility in India between 1999 and 2008; of these, 11.5% of outpatients and 5.4% of inpatients met ICD-10 criteria for trance and possession disorder.

Community-based surveys have yielded higher rates. In 1972, Morris Carstairs and Ravi Kapur found a 6-month period prevalence of 0.97% for involuntary possession syndrome in a South Indian village, using a symptom checklist in a 50% sample of all village adults (*N* = 1,233). They also reported a 6-month period prevalence of an additional 1.8% for voluntary possession states. V. Venkataramaiah and colleagues surveyed the total population (*N* = 1,158) of a school catchment area in another South Indian village during an epidemic of possession syndrome in 1979. They found a 1-year period prevalence of 3.5% for involuntary attacks, as well as a prevalence of 0.2% for voluntary possession trance episodes.

Anthropologic accounts of widespread occurrence and cultural barriers to discussing the supernatural features of the syndrome have suggested that the actual prevalence is likely higher than that reported in public surveys. The clients of indigenous healers currently provide the main source of comparative data. For example, 17% of those patients seen by one traditional healer during a 4-month period who met research criteria for psychiatric ailments (*N* = 30) received an indigenous label of possession syndrome. Finally, the periodic occurrence of syndrome epidemics complicates the epidemiologic picture.

Possession syndrome is more common in women, with a female-to-male ratio of 2 to 3:1 in most community and psychiatric cohorts. The 1999 to 2008 cohort showed much higher female-to-male ratios among outpatients (4.7:1) than inpatients (1.3:1), suggesting that women with trance and possession disorder were significantly less likely to be hospitalized. Sex distribution may also vary historically, according to evolving cultural gender attributions, and show geographical or possibly idiosyncratic variation. Various studies conducted in India suggest that women experience considerable social disadvantage related to the typically patriarchal social system, including early marriage, childbearing as the primary social function, family and spousal violence, lack of social integration, poor social support from family members, and limited autonomy. These factors may contribute to the higher prevalence of possession syndrome among women. The age of onset is usually between 15 and 35, but many cases reportedly begin in childhood. Attacks may persist well into middle age, and geriatric cases have also been reported. The syndrome has been identified in all Hindu castes, as well as in Muslims, Christians, and tribal peoples; however, it is more commonly described among Hindus.

Table 29.1–3.
Phenomenology of Possession Syndrome

a. Onset occurs typically due to subacute conflict or stress and shows considerable variation. It may be gradual and nonspecific—e.g., diverse somatic complaints, such as dizziness, headaches, abdominal discomfort, hot–cold flashes, listlessness, or difficulty breathing—or sudden and specific, in the form of an abrupt transition to an altered state of consciousness. Onset may also vary by geographical region, presenting in certain areas of the country (e.g., parts of Uttar Pradesh) as a mounting series of rage attacks. If presentation is gradual and nonspecific, transition to the altered state may occur spontaneously or after contact with an indigenous healer specializing in the treatment of the syndrome.
b. Behavior during the altered state consists of some or all of the following:
   i. Dramatic, semipurposeful movements—such as head bobbing, bodily shaking, thrashing, gyrating, or falling to the ground—accompanied by guttural, incoherent verbalizations such as mumbling, moaning, or shrieking.
   ii. Aggressive or violent actions directed at self or at others, including spitting, striking, and impulsive suicidal or homicidal gestures. Verbalizations may be coherent, and consist of derogatory comments or threats of violence directed against significant others or against the subject (in the third person) and typically considered by observers to be uncharacteristic of the subject's usual behavior.
   iii. Specific gestures, comments, or requests denoting the appearance of a known possessing personality by (1) reference to standard attributes of culturally recognizable figures or (2) the name and degree of relation of deceased family members or acquaintances.
c. The altered state is marked by the emergence of one or several secondary identities distinct from that of the subject. Their specific identities, which may remain undisclosed for some time, typically adhere to cultural norms regulating permissible agents of possession, which vary by religion, region, and caste.

Acceptable agents include spirits of deceased family members, in-law relations, or known village acquaintances who died under specific conditions of distress, and minor supernatural figures of the Hindu pantheon (rarely major deities) and the Islamic spiritual world. They do not include the identities of alternate, living persons, nor apparently alternation of co-conscious multiple identities, as in nonpossession-form dissociative identity disorder, though consecutive possession by two or three agents is not rare, and one case of 21 successive possessions has been reported. On occasion, a single agent may possess several members of a family over time, or in succeeding generations.

d. Possession by the secondary identity(ies) is episodic, resulting in alternation between the usual personality of the subject and the altered state. The subject in their usual identity appears in a daze, exhausted, distressed, or confused about the situation and may report visual or auditory perceptual disturbances regarding the possessing agent, as well as “unconsciousness” and partial or total amnesia for the altered state. States of consciousness may alternate continuously within one episode over hours to weeks, with each state lasting minutes to hours. This pattern may persist intermittently over months or years.
e. Frequently, the specific identities of possessing agents remain undisclosed for some time, requiring the active ministrations of family members and the intervention of specialized indigenous practitioners. The process of disclosure is conceived as a struggle between the family members and the beneficent agents possessing the healer on the one side, and the troublesome possessing identities on the other. It is characterized by remarkable reactivity on the part of the subject to environmental cues, including direct questioning, strategic neglect, and aggressive manipulation. Often, this kind of ritualized intervention elicits the withheld disclosure, and leads to a process of negotiation between the family, the healer, and the possessing agent regarding the propitiatory treatment necessary for the resolution of the episode. Examples of this treatment typically consist of ritual offerings, admission to a possession

cult, resolution of family conflicts, cessation of abuse, and the postponement or precipitation of marriage.

f. Outcome is variable. Total recovery is often reported at the cessation of a single acute episode, which may be of several weeks' duration. Alternatively, prolonged morbidity may result, or even, rarely, death. The subject may seek exorcism treatment with multiple healers, or undertake a pilgrimage to a regional healing temple. At times, they are prescribed entrance into a ritual devotion cult offering programmed periodic possessions.

Caste distribution may fluctuate historically and respond to sociocultural determinants of the appropriateness and desirability of possession syndrome as an idiom of distress. It has been reported more frequently among disadvantaged castes, of low income and education, and rural residence. Nevertheless, some case studies revealed the prevalence of the syndrome among more advantaged groups, such as Brahmins, despite the cultural proscriptions against it. Case reports suggest higher rates among widowed persons, but a systematic study did not confirm this association. The presence of possession syndrome in other family members (or in ancestors) may be a risk factor, but this relationship has not been investigated systematically.

Precipitants/Etiology. Precipitants of possession syndrome are varied but typically include social or family conflicts and stressful life transitions that elicit strong feelings of vulnerability in persons without firm emotional support. Examples encountered in the literature include marital conflict, abuse, and neglect, at times associated with alcoholism; the arrival of a new bride to the home of her husband's family; delay in arranging a marriage, or in consummating it; forced marriage; widowhood; postpartum status; loss of family social standing; death of a family member; difficulty finding employment and financial difficulties; alienation from family support; and subordination to other family members and in-laws. These precipitants were identified in 67% and 73% of separate mental health cohorts.

Some older clinical studies stressed the pathogenic role of hysterical and histrionic ego defenses in possession syndrome. However, this ignores the fact that pathologic possession commonly occurs in settings where nonpathologic possession is a positive, socially mandated experience. What may distinguish pathologic episodes is the importance of traumatic precipitants—deaths of children, murder of husband, sexual abuse of a close friend, measles epidemic, physical abuse, the revelation of husband's incest with daughter. Social scientists such as Richard Castillo have emphasized the underlying context of violence against women in India—including selective abortion, infanticide, murder due to discovered premarital relations, rape, sale into prostitution, "dowry death" (the new bride is killed if dowry is considered insufficient), marital abuse, witch killings, and *sati* (widow burning). These data suggest that many instances of possession syndrome may arise not from intrapsychic vulnerability but from widespread undisclosed early trauma, an exposure that has been insufficiently investigated. Some researchers agreed on the etiologic role of childhood stress as a risk factor for possession syndrome but differed on the specific contributions of physical and sexual abuse as opposed to emotional or physical neglect. Alternatively, anthropologic investigators have argued that individual biographical data was insufficient to explain the specific association of spirit possession with marginal persons and transitional life phases, such as birth, menarche, marriage, pregnancy, widowhood, and death. They located the etiology of attacks in caste-specific cultural expectations linking social and spiritual vulnerability. Additional collectively defined precipitants of possession syndrome include certain acts and settings, such as crossing physical boundaries (e.g., the village limit, an empty field at night), which invite the attention of possessing spirits.

Relationship to Psychiatric Diagnoses. Psychoanalytically based evaluations of subjects with possession syndrome often reported evidence of hysterical traits predisposing to dissociative psychopathology in otherwise normal subjects. Others distinguished the intrapsychic origin of the acute behavioral syndrome—which they attributed to hysterical processes—from the underlying personality formation and psychic vulnerability. These authors argued that attacks could arise in subjects with diverse personality structures and biologic propensities. Some research reported a substantial proportion of attacks among patients with a primary diagnosis of schizophrenia (40% to 59%) and manic depressive illness (11% to 13%), in whom the possession symptoms took longer to resolve, compared to patients without these disorders.

As noted earlier, DSM-5-TR includes most cases of possession syndrome under dissociative identity disorder. However, cases that do not fulfill the amnesia criterion should receive other DSM diagnoses, including other specified dissociative disorder. As with all cultural concepts of distress, there is no one-to-one overlap between the cultural label and any DSM disorder; psychiatric evaluation must proceed at the level of the individual. In ICD-11, possession syndrome cases would fulfill criteria for trance and possession disorders. Prior to the inclusion of possession phenomenology in dissociative identity disorder, possession syndrome did not have a diagnostic home in DSM: a study conducted in 1987 in a New Delhi outpatient psychiatric clinic (*N* = 1,517) found that 10% of cases with dissociative pathology received an ICD-10 diagnosis of trance and possession disorders, whereas none met DSM-III-R criteria for multiple personality disorder.

Additional Clinical Features. Rates of relapse are unknown, although anecdotal reports suggest it is not uncommon. Relapse within 15 years was not infrequent and has been recorded after two to three decades. Four of 15 subjects in one study were found to have suffered previous episodes, and 1 subject in another series reported three consecutive postpartum attacks. Entry into a ritual possession cult may reduce or eliminate relapse.

Specific Cultural Factors. Possession syndrome constitutes a folk illness category throughout India and Sri Lanka. It may present initially in a variety of forms, linked by the attribution of spirit etiology. When it presents in a nonspecific fashion, indigenous diagnosis is confirmed by the appearance of the altered state during the therapeutic ritual. It is considered an affliction by its painful, involuntary nature and attributed to the intervention of specific spiritual agencies acting independently or at the behest of a witch. Certain castes and persons in transitional states (e.g., puerperium) are considered most vulnerable to spirit attacks, especially when deprived of emotional and material support. The degree of pathology ascribed to the sufferer, the therapy chosen, and the outcome attained depend on the interaction of several factors: psychological (e.g., traumatized ego), biologic (e.g., psychosis), sociocultural (e.g., caste status), and situational (e.g., family dynamics). For instance, when it presents in some Moger women of South India, possession syndrome may be seen as an affliction opening the door to a sacred practice in a possession cult. In some Havik Brahmin women, however, it may constitute the last step in a series of symptomatic expressions of acute distress instead, leading directly to exorcism and incomplete conflict resolution.

Treatment. Specialized indigenous practitioners and ritual therapies are generally available and widely utilized, but psychiatric treatment is typically avoided. Indigenous treatments include neutralization of the conflict or stress via the communal rituals involved in exorcism, as well as the reformulation of the suffering into beneficent individual and communal practice via initiation into a spirit devotion cult—such as the Siri cult of South India—or education into the roles of oracle (diviner), exorcist, or, rarely, avatar (divine incarnation).

Shenjing Shuairuo

Shenjing shuairuo (“weakness of the nervous system” in Mandarin Chinese) is a translation and cultural adaptation of the term “neurasthenia” (from Greek roots for “lack of nerve strength”), which was introduced to China from Russia in the early 1900s. As formulated by the American neurologist George Beard in 1868, neurasthenia denoted a heterogeneous syndrome of lassitude, pain, poor concentration, headache, irritability, dizziness, insomnia, and over 50 other symptoms. At first, neurasthenia was considered an "American disease" resulting from the "pressures" of a rapidly modernizing society but was later adopted by European diagnosticians. Its pathophysiology was thought to derive from a lowering of nervous system function through a physical process due to excessive use, especially among the educated and wealthier classes. In Soviet psychiatry, buttressed by Pavlovian research, neurasthenia was a central component of mental health nosology, which became especially influential in Chinese psychiatry after the communist revolution of 1949.

While neurasthenia declined in importance in Western psychiatric nosology and practice during the 20th century, *shenjing shuairuo* underwent marked popular and professional development in mainland China, Taiwan, Hong Kong, in Chinese migrant communities, and in Japan, where a similar syndrome was labeled *shinkei suijaku*. From a peak in about 1980, when it constituted up to 80% of all "neurotic" diagnoses in Chinese societies, *shenjing shuairuo* has become less common. The diagnosis was included in the *Chinese Classification of Mental Disorders, Second Edition, Revised* (CCMD-2-R; 1993), under the section on "Other Neuroses." The CCMD-2-R diagnosis required three symptoms out of five nonhierarchical symptom clusters, organized as weakness, emotional, excitement, and nervous symptoms, as well as a fifth category of sleep disturbances. Like other neurotic disorders in the Chinese manual, the condition had to last at least 3 months and (a) lower the efficiency of work, study, or social function; (b) cause mental distress; or (c) lead to treatment seeking. Depressive and anxiety symptoms, which used to be central in previous professional definitions of *shenjing shuairuo*, became exclusionary in CCMD-2-R, probably due to the growing influence of Western nosologies. Physical weakness and mental fatigue, which are prominent elements of some neurasthenic presentations in other parts of the world, are not prioritized in the Chinese classification. The third edition of the CCMD (2001) retained *shenjing shuairuo* in the section on the Neuroses as a diagnosis of exclusion. This decision was based on multicenter field trials in China on the discriminant validity and clinical utility of the neurosis subtypes.

The general category of “neurasthenia” included in ICD-10 (F48.0) emphasized the components of weakness and fatigue and relegated other possible symptoms to secondary characteristics. Although the evolving CCMD nosology has not emphasized these symptoms in its definitions of *shenjing shuairuo*, some controversy remains over the proper status of fatigue and weakness in the Chinese classification. Some investigators considered the centrality of fatigue in the ICD-10 definition of neurasthenia a misleading alignment of the international classification with the Western diagnoses of somatoform disorders and chronic fatigue syndrome. For example, Sing Lee estimated that the ICD-10 definition excluded over 80% of Chinese patients with *shenjing shuairuo* who reported prominent symptoms other than fatigue. ICD-11 no longer includes an entry for neurasthenia, apparently due to decreasing official support internationally for the category. The DSM-5-TR chapter on culture and psychiatric diagnosis includes an entry for *shenjing shuairuo*, as an example of a cultural syndrome, which follows the CCMD-3 definition and does not prioritize the fatigue component.

Phenomenology. Given the evolution of diagnostic practice regarding *shenjing shuairuo* in Chinese societies over the last decades, the phenomenologic description in this chapter is based instead on clinical histories of self-identified sufferers—or the “illness” aspect of the syndrome. The elements in Table 29.1–4 are prototypical.

Epidemiology. The changing professional definitions of ICD neurasthenia and CCMD *shenjing shuairuo* complicate the interpretation of the epidemiologic reports, which are in any case limited in European languages. Most studies have used ICD definitions, which in turn have evolved over time, or else refer to rates of “neurasthenia” without distinguishing between ICD and CCMD diagnoses. Therefore, the relationship between these reports of neurasthenia prevalence and the folk presentation of *shenjing shuairuo* remains unclear.

Table 29.1–4.
Phenomenology of *Shenjing Shuairuo* (Weakness of the Nervous System)

a. Onset is usually gradual, sometimes spanning several years, and typically emerges out of a conflictive, frustrating, or worrying situation that involves work, family, and other social settings, or their combination. A sense of powerlessness in changing the precipitating situation appears central to most illness accounts of the syndrome.
b. Symptoms show substantial individual variation, but usually involve at least some of the following spontaneous complaints: insomnia, affective dysphoria, headache, bodily pains and distortions (e.g., “swelling” of the head), dizziness, difficulty concentrating, tension and anxiety, worry, fatigue, weakness, gastrointestinal problems, and “troubled vexation” (*fan nao*). This last emotion has been described as a form of irritability mixed with worry and distress over “conflicting thoughts and unfulfilled desires,” that may be partially concealed in Chinese communities for the sake of preserving social harmony.
c. The sufferer frequently seeks the sick role, attributing his/her difficulties in meeting work, school, or other social expectations to the syndrome. Sources of treatment vary substantially across Chinese communities, depending on the availability of formal and traditional service sectors. Typically, physical health services, such as general health or neurology clinics, or traditional Chinese medicine providers are preferred to psychiatric services, though this may be less pronounced among school-related presentations, especially in more Westernized settings such as Hong Kong.
d. Course is variable, and may respond closely to changing interpersonal and social circumstances. Amelioration of the precipitating stressor(s) typically brings about substantial improvement, although residual symptoms appear common.
e. Response to treatment may be strongly mediated by the illness role and its relationship to the intractability of precipitating stressors.

Several community studies have been conducted, but most are available only in Chinese. One community-based study conducted in 1959 in Chengdu City (*N* = 10,830) found a point prevalence of 5.9% for unspecified “neurasthenia” of greater than 3 months’ duration. Prevalence rates from other community surveys prior to 1980 have been reported as between 6% and 10%. Several investigators noted a rise in presentations of *shenjing shuairuo* to medical services in the aftermath of the Cultural Revolution (1966 to 1976) when social space was opened for a public accounting of the traumas associated with this period. However, epidemiologic documentation of this likely upswing in community and clinical prevalence is limited. One nationwide community-based study from 1986 found a prevalence rate of 1.3% in the general population. Prevalence rates among students have typically been high, given the association of school-based stressors with *shenjing shuairuo*. In 1993, 5.1% of Chinese college students in Anhui, China, fulfilled the CCMD-2 criteria for the syndrome. Lee and colleagues reported a 6.3% prevalence of moderate or severe *shenjing shuairuo* among Hong Kong undergraduates in 1995. A 2014 household survey in Guangzhou by Brian Hall and colleagues, using the Patient Health Questionnaire (PHQ-9) and the neurasthenia criteria from ICD-10, found a prevalence of depression of 5.3% and of neurasthenia of 15.4%. The prevalence of neurasthenia alone or with depression did not differ significantly by any demographic characteristic studied, including age, gender, education, and income.

Clinical epidemiology studies have varied markedly over time. By numerous anecdotal reports, between about 1950 and 1980, neurasthenia, either in the ICD or the CCMD formulation, constituted probably the most common psychiatric diagnosis in China. The narrowing of the psychiatric definition of neurasthenia since 1980—such as by the exclusion of primary affective presentations—and its reduced application in diagnostic practice likely decreased prevalence rates in clinical epidemiology. Yet the prevalence of *shenjing shuairuo* in medical clinics remained high. A 1983 clinical epidemiologic study in a county general hospital in Sichuan province (*N* = 1,401) found a prevalence of 3.9% for ICD-9–defined neurasthenia. From 1997 to 1998, 5.3% of adult outpatients using internal medicine, neurology, and traditional Chinese medicine clinics in a referral hospital in Hunan province met CCMD-2-R criteria for *shenjing shuairuo*. Rates of up to 6.3% have been reported in other general clinical settings. Psychiatric inpatients appear to have much lower rates, probably reflecting the outpatient nature of most neurasthenia care. In 1990, the prevalence of neurasthenia among 2,275 inpatients discharged from the Shanghai Institute of Mental Health was only 0.26. Due to the growing official acceptance of the ICD classification in China, the removal of neurasthenia from ICD-11 may lead to less use of the diagnosis of *shenjing shuairuo* in official clinical documentation.

The 12-month prevalence of ICD-10 neurasthenia among Chinese Americans in Los Angeles County was studied by Yan-Ping Zheng and colleagues. The prevalence of all neurasthenia diagnoses, regardless of diagnostic exclusions, was 6.4%. The rate of "pure" neurasthenia—those without any lifetime comorbid DSM-III-R diagnoses—was 3.6%, higher than any other DSM-III-R disorder. These "pure" neurasthenia sufferers did not differ on most sociodemographic characteristics from respondents without a psychiatric diagnosis, with "pure" depression, or with "pure" anxiety disorder. Further analysis of the study subsample (95%) who were foreign-born revealed that those with higher neurasthenic symptom counts were older, had less education, and were more likely to be formerly married (separated, widowed, or divorced). Lower acculturation increased the risk of neurasthenia, but only in the presence of financial stress (as an interaction term). Participants with lower acculturation were more likely than those with higher acculturation to express the syndrome when confronted with financial stress. The authors concluded that "neurasthenia is a unique form of expressing distress among low acculturated groups of Chinese Americans."

Precipitants/Etiology. An empirical assessment of the precipitants of *shenjing shuairuo* has found high rates of work-related stressors, which were made more intractable by the centrally directed nature of mainland Chinese society. These included unwelcome work assignments, job postings that caused family separations, harsh criticism at work, excessive workloads, monotonous tasks, and feelings of inadequacy or incompatibility of skills and responsibilities. Students usually described somewhat less severe study-related precipitants, particularly school failure or anxiety over the mismatch between personal or family aspirations and performance. Other interpersonal and family-related stressors included romantic disappointments, marital conflict, and the death of a spouse or other relative. The work of Arthur Kleinman has emphasized the political nature of many precipitants of *shenjing shuairuo*, especially during the 1980s and 1990s as a result of the vicissitudes of the Cultural Revolution. This societal upheaval involved traumatic changes in social circumstances, family separations, acute humiliations, social ostracism, demoralization, and exposure to violence and deprivation.

A common theme to most precipitants of *shenjing shuairuo* involved frustration of a keenly held aspiration, glossed as “troubled vexation,” to which social circumstances allowed no or insufficient solution. Loss of face (*mianzi, lianzi*), a key social value in Chinese communities, was also frequently noted as a trigger of *shenjing shuairuo*.

Chinese etiologic understandings of the syndrome commonly invert the Western view of “psychosomatic” presentations, whereby social–interpersonal precipitants cause psychological distress that is displaced onto bodily experience. In what Kleinman described as “somatopsychic” interpretations, many sufferers of *shenjing shuairuo* located the pathologic impact of external stressors directly onto physical functioning, producing somatic symptoms that only secondarily caused psychological distress. This schema permits sufferers to acknowledge a psychological contribution to their illness while yet locating the nexus of pathology on physical processes.

Relationship to Psychiatric Diagnoses. Several studies in Chinese communities found a strong relationship between *shenjing shuairuo* and mood, anxiety, and somatoform disorders. Kleinman found considerable overlap between *shenjing shuairuo* and these disorder clusters in his studies in both mainland China and Taiwan. In the People's Republic, his study of 100 Chinese patients diagnosed with *shenjing shuairuo* found that 93 also had a depression diagnosis, with 87 meeting the criteria for major depressive disorder. Sixty-nine patients were also diagnosed with an anxiety disorder and 25 with a somatoform disorder. In Taiwan, 22 of 51 patients with *shenjing shuairuo* also met the criteria for major depressive disorder, which was the most common psychiatric diagnosis in this group of patients.

Subsequent investigations confirmed these associations, with rates of depressive disorders ranging from 30% to 70% in cases of *shenjing shuairuo*. The association with somatoform and anxiety disorders was also found repeatedly. The exact correlations with these disorders depended on whether the subjects came from community or clinical settings; in clinical studies, the correlations varied between general medicine and psychiatric clinics. Outpatient primary care settings tended to yield associations between *shenjing shuairuo* and the somatoform disorders. One study by Sing Lee and colleagues with 100 primary care patients in Hong Kong with unexplained fatigue of at least 6 months' duration that excluded patients already in treatment for a depressive disorder found that 72% endorsed a self-label of *shenjing shuairuo*. In the full sample, the most common diagnoses on the Structured Clinical Interview for DSM-III-R were somatoform pain disorder (47 percent) and undifferentiated somatoform disorder (35 percent). Fourteen patients met the criteria for depressive disorders, with the other diagnoses mostly in the anxiety spectrum.

A study conducted from 1997 to 1998 by Doris Chang and colleagues in three general medicine clinics in Hunan also found similar correlations. The study included 139 patients, mostly of rural origin, who presented with unexplained somatic complaints. Only 55 percent received any diagnosis on the Structured Clinical Interview for DSM-IV, all in the somatoform disorder spectrum. Undifferentiated somatoform disorder was the closest counterpart to *shenjing shuairuo* in DSM-IV, with a diagnostic concordance of kappa = 0.24 ($p < .001$). The fact that 45% of subjects with the CCMD-2 syndrome did not meet the criteria for any DSM-IV disorder, however, suggested to the authors that *shenjing shuairuo* is a distinct clinical condition with only partial comorbidity with DSM diagnoses.

A similar conclusion was reached by Hall and colleagues in their 2014 community-based epidemiologic study in Guangzhou (*N* = 751). This study did not use the CCMD definition of *shenjing shuairuo*, however, relying instead on the ICD-10 criteria for neurasthenia, which had only a partial overlap with CCMD-3. The comorbidity between neurasthenia and PHQ-9-based depression was only 2.5%, but people with depression had nearly six times higher odds of having neurasthenia. Sleep latency was highest among individuals with neurasthenia and specific to this disorder, possibly linked to higher levels of worry. Zheng and colleagues, in their 1997 community-based epidemiologic study among Chinese Americans in Los Angeles County (*N* = 1,747), likewise used the ICD-10 criteria for neurasthenia. They identified 112 (6.4%) subjects with ICD-10 neurasthenia in the community. Over half (*N* = 63; 56%) did not meet the criteria for a DSM-III-R diagnosis. Compared with those with depression and anxiety diagnoses, these respondents reported fewer psychological symptoms on the SCL-90-R but similar levels of somatization symptoms.

The overlap with ICD definitions of neurasthenia has been assessed in a few studies. Chang and colleagues found an overlap of 53% between the CCMD-2 *shenjing shuairuo* category and ICD-10 neurasthenia. An even more limited overlap with the international classification was found in a 1989 study conducted at the Shanghai Psychiatric Hospital. Clinical diagnoses of *shenjing shuairuo* obtained by two senior Chinese psychiatrists yielded only a 22.5% overlap with the ICD-9 neurasthenia category.

Additional Clinical Features. The clinical course of the syndrome may depend on the associated psychiatric comorbidity and on the degree of persistence of the precipitating stressors. One longitudinal study found complete resolution of *shenjing shuairuo* symptoms and good social adjustment 20 years after the index diagnosis in 83 out of 89 cases. Only one case was receiving continued treatment, and no subjects reported the onset of depressive disorder subsequent to the *shenjing shuairuo* diagnosis.

Chinese psychiatrists have carried out numerous studies of neurophysiologic and cognitive function in *shenjing shuairuo* patients since the 1950s. Most have reported abnormalities compared to normal controls, including in tests of polysomnography, EEG, psychogalvanic reflex, gastric function, and memory function. These findings need to be replicated with well-controlled samples using contemporary diagnostic instruments.

Specific Cultural Factors. The evolving definitions of *shenjing shuairuo* have emerged out of a tradition of syncretism in Chinese medicine between indigenous illness understandings and international contributions. Nineteenth-century Western notions of a weakened nervous system due to overuse (neurasthenia) found an ancient cognate expression in Chinese concepts of bodily meridians or channels (*jing*) binding vital organs in balanced networks along which forces (e.g., *qi*, vital energy, in *yin* and *yang* forms) could be disrupted from their normal, harmonious flow. This gave rise to *shenjing shuairuo*, an illness whereby the *jing* that carries *shen*—spirit or vitality, the capacity of the mind to form ideas, and the desire of the personality to live life—have become *shuai* (degenerate) and *ruo* (weak) following undue nervous excitement.

As noted by Kleinman, this Chinese conception is much more relational and contextual than the typical Western medical view. Bodies are seen as acting in the midst of specific contexts and social relationships rather than abstractly or individually. Physical networks are thus metaphorical representations of interpersonal and societal connections, with emotions acting as mediating channels. These social networks, just like bodies, “can be vitalized or devitalized, demoralized or remoralized. Indeed, bodies are networked, networks embodied.”

The result is an illness form, *shenjing shuairuo*, that tightly links social and interpersonal situations with bodily afflictions, such as the metaphorical linkage described earlier between dizziness and social dislocation due to revolutionary upheavals. Other contributors to the persistence of this illness form include elevated stigma against mental illness in Chinese communities, which is forestalled by the primarily physical nature of *shenjing shuairuo*; the contributions of Pavlovian psychiatry and Japanese research to Chinese nosology, which emphasized neurasthenia beyond its role in Western psychiatry; the prominent use of the syndrome by the communist authorities as a health-based, treatable cause of reduced productivity during the 1950s and 1960s; and its popular acceptability as a veiled complaint against social deprivation and oppression at a time when direct confrontation would be harshly suppressed.

Treatment. Kleinman noted a discrepancy in treatment patterns between Taiwanese and Chinese *shenjing shuairuo* patients, with those in Taiwan more frequently resorting initially to self-help remedies, which included special foods, tonics, and dietary changes. He attributed this discrepancy to the availability of inexpensive care in the People’s Republic.

When accessing formal sectors of care, most patients used both Western-trained physicians and practitioners of traditional Chinese medicine. Nonpsychiatric medical settings were preferred, including neurology and general medicine clinics, in concert with cultural understandings of the somatopsychic etiology of *shenjing shuairuo*, which emphasize its physical mediation. The modality of treatment was usually traditional Chinese medicines, which were prescribed by both Western-trained and traditional Chinese medicine practitioners. This conformed to the balanced status ascribed to both types of training among Chinese physicians. Polypharmacy was common, combining sedatives, traditional herbs, antianxiety agents, vitamins, and other tonics. In spite of active suppression of religious healing in China, almost a quarter of patients were also engaged in such treatment in the early 1980s.

ONGOING WORK ON THE RELATIONSHIP BETWEEN CULTURAL CONCEPTS OF DISTRESS AND PSYCHIATRIC NOSOLOGY

Research

In 1969, Yap challenged the field to learn more about “culture-bound syndromes” so that their relationship to psychiatric disorder could be resolved, yet much research still needs to be done. Without a more systematic and wide-ranging program of research on a broader range of cultural concepts of distress, current approaches that only focus on classificatory exercises are unlikely to further the current understanding. The strategy of trying to find the right classificatory scheme by basing it on the similarity between one or two symptoms of a cultural syndrome and of the psychiatric disorder is limited. It is particularly problematic when researchers privilege the DSM categories as the main organizing structure of relevance to the cultural concepts of distress. This approach is unlikely to produce new insights into these concepts on their own terms or in relation to psychiatric diagnosis. It ignores, for example, that many cultural concepts are not organized syndromically so that classifying them on the basis of prototypical symptoms is incompatible with their organizational structure. Comparing etiologic categories with symptom-based diagnoses on the nature of their symptoms is not only useless but counterproductive.

A simple, symptom-based comparison further ignores the function of cultural concepts as idioms or modes of expressing distress. Comparison of cultural concepts and psychiatric disorders as idioms should focus on the mutual compatibility of their intended communications (do they deliver the same message?) and the cultural fit between the idiom—whether indigenous or professional—and the situation in which it arises (is the message equally appropriate?). An idiom of distress intended as a sign of anger that is misinterpreted by the diagnostician as a sign of severe mental illness (e.g., psychosis) is not being accurately translated from the idiomatic perspective. Similarly, an idiom that is considered culturally appropriate to the situation (e.g., anger as a comprehensible reaction) would be well matched by a diagnosis that idiomatically conveys a similar sense of appropriateness.

In order to resolve the cognitive dissonance created by the complex relationship between the cultural concepts and psychiatric disorders, cross-cultural mental health researchers need to carry out an intensive research that illuminates these concepts rather than simply attempting to subsume them into psychiatric categories and obliterating them. Peter Guarnaccia and his colleague Lloyd Rogler proposed a research program on cultural concepts of distress to bring their study into the mainstream of mental health research. This framework is faithful to the holistic nature of the cultural concepts while at the same time applying current research approaches from several fields. It comprises a series of key questions that need to be answered in order to understand these concepts from the cultural as well as the psychiatric perspectives.

Guarnaccia Rogler's questions are organized into the following five broad dimensions:

1. **Nature of the phenomenon:** How is the folk category characterized within its cultural context? What are the defining features of the phenomenon? Is the folk illness organized syndromically? Does it refer primarily to an etiologic construct? What is the message it attempts to convey as an idiom of distress? How salient is the concept, defined as the rapidity and extent of its recognition within a cultural context?
2. **Location in the social context:** Who are the people who report the folk category, and what is their location in the social structure? What are the situational factors that provoke these conditions?
3. **Relationship to psychiatric disorder:** How is the folk illness empirically related to psychiatric disorder(s)? Does it cut across diagnostic groupings? What is the range of its overlap with psychiatric categories?
4. **Social/psychiatric history of the cultural concept:** When the onset and offset of the cultural concept and the associated psychiatric disorder(s) do not coincide, what is the sequence of onset? How does the life history of the sufferer, particularly the experience of traumatic events, affect the sequence? How many

episodes of the folk illness has the person experienced? How severe have they been?
5. **Treatments and outcomes:** Holistically, how should be person's suffering be addressed? If the folk illness is effectively treated, does the sufferer also experience relief of the associated psychiatric disorder(s)? And vice-versa, if the psychiatric disorder is resolved with psychotherapy or pharmacotherapy, does the folk illness resolve as well?

Clinical Application

Medical anthropology and cultural psychiatry, and psychology have developed tools and perspectives that are useful in assessing cultural concepts of distress in clinical work and in negotiating between the concepts and psychiatric diagnoses. One such model is the Explanatory Model Questions developed by Kleinman. Another model is the Outline for Cultural Formulation, included in DSM-5-TR along with its operationalization as a set of semistructured questionnaires known collectively as the Cultural Formulation Interview. These frameworks are designed to help clinicians inquire about and better understand the cultural dimensions of their patients' presentations of their mental health problems.

The key area of the Outline for Cultural Formulation is the section on cultural explanations of the individual's illness. The major elements of inquiry include the predominant idioms of distress through which symptoms are communicated; the meaning and perceived severity of the individual's symptoms in relation to the norms of the cultural reference group; local illness categories used by the individual's family and community to identify the condition; the perceived causes or explanatory models that the individual and reference group use to explain the illness; and past and current experiences with popular and professional sources of care. These areas can serve as a support for clinicians who seek to better understand their patients' perspectives on the symptoms, expressions, and difficulties they present.

In DSM-5, the Outline was rendered more operational for routine clinical practice by standardizing its implementation as the Cultural

Formulation Interview. This is composed of three types of semistructured interviews that offer clinicians concrete ways to carry out the cultural assessment of patients and their entourages. The three components of the Cultural Formulation Interview are a core 16-item questionnaire, an informant version of the core interview to obtain information from caregivers, and 12 supplementary modules that expand on these basic assessments. If used routinely to frame the intake evaluation, the Cultural Formulation Interview can help establish a foundation of person-centered information on the patient's context and perspective, and a working alliance, on which to build the rest of the diagnostic interview and treatment negotiation process.

There are several resources that clinicians can turn to for further guidance and examples using this formulation technique. The cultural formulation has been elaborated by the Cultural Committee of the Group for the Advancement of Psychiatry in their 2002 publication Cultural Assessment in Clinical Psychiatry. Illustrative cases using the cultural formulation and discussing the implications of cultural understanding of mental health problems have been available in select issues of *Culture, Medicine, and Psychiatry* starting in 1996. Key writings on the cultural formulation approach were assembled in *Cultural Formulation: A Reader for Psychiatric Diagnosis* (2008). The McGill Cultural Consultation Service published *Cultural Consultation: Encountering the Other in Mental Health Care* in 2014, describing the cultural assessment of over 400 cases with an expanded cultural formulation approach. Finally, the *DSM-5 Handbook on the Cultural Formulation Interview* (2016) describes in detail the components of the Cultural Formulation Interview, presents multiple clinical vignettes to illustrate its use, and suggests strategies for incorporating the interview into clinical training and routine practice.

Once the information is elicited, the next key step is for clinicians to lead a negotiation process whereby they and their patients come to understand each other's perspectives on the mental health problem and come to an agreement on how to proceed. Kleinman described this approach in *The Illness Narratives* (1988). The process that Kleinman developed is for the provider to first elicit and understand the patient's explanatory model and then to make clear the provider's explanatory model. The process then focuses on developing a treatment plan based on negotiation between the patient and provider models. It may be useful to include the patient's cultural terms and explanations in case formulations and treatment plans, as they may help clarify symptoms and etiologic attributions that could otherwise be confusing. This approach has been further elaborated in a variety of cultural competence training programs for physicians to aid them in this negotiation process. The impact on the quality of mental health care from these approaches for assessing cultural understandings of cultural concepts of distress and for clinician–patient negotiation are pressing areas for research.

USING CULTURAL CONCEPTS OF DISTRESS TO IMPROVE MENTAL HEALTH SERVICE DESIGN AND IMPLEMENTATION

In the previous section, we discussed the added value of incorporating cultural concepts of distress into clinical assessment using the Cultural Formulation Interview. In this section, we discuss how attention to cultural concepts of distress can play an important role in culturally appropriate and effective mental health care more broadly. As we have emphasized, cultural concepts of distress are not simply placeholders for psychiatric categories—they have their own meanings and both personal and social consequences that may interact with many different clinical conditions and social predicaments. In many instances, cultural concepts of distress may be more meaningful to patients and families than psychiatric diagnostic terms because they represent shared experiences and expectations.

Many of the cultural concepts of distress previously described in this chapter have been used to improve mental health services. In this section, we will illustrate the clinical uses of cultural concepts of distress by focusing on the example of "thinking too much"—a cultural concept of distress characterized by both worry about future events and rumination about past events that has been observed in various forms across many cultures and contexts. The qualification "too much" refers to the intrusive quality of these thoughts and difficulty controlling them once they have begun. "Thinking too much" may also be associated with sleep problems and negative perceptions of self. In some cultural contexts, experiencing "thinking too much" for a prolonged period is associated with a risk of more severe psychopathology such as delusions or hallucinations or a lack of self-control (e.g., an inability to abide by social norms). Coping with "thinking too much" varies cross-culturally, but it often includes strategies to control the intrusive thoughts such as by sharing them with others, engaging in prayer or other religious practices, using behavioral distractions, and seeking treatment from a variety of practitioners of traditional medicine as well as biomedical health workers. Local examples of "thinking too much" include *kufungisisa*, a Shona term in Zimbabwe; *reflechi twòp*, a Kreyòl term in Haiti; *pensando mucho*, a Spanish term in Latin America; *man maa kura khelne*, a Nepali term in South Asia; and *kut careen*, a Khmer term in Cambodia.

Cultural Concepts of Distress and Engagement with Mental Health Services

Challenges in mental health services are often categorized as *supply side* or *demand side*. "Supply side" refers to the availability of mental health services for a particular population and setting. "Demand side" refers to help-seeking behaviors to engage the available services. For many minority populations in high-income countries and for most residents of low- and middle-income countries, there are both supply-side and demand-side barriers. Culturally and contextually sensitive care is often lacking, and when services are available, there is a hesitance to engage such services because of lack of awareness about care, difficulties trusting providers due to prior experiences with poor-quality care, lack of recognition of a psychiatric illness, and stigma associated with using mental health services. The utilization of cultural concepts of distress can help to reduce demand-side barriers. For example, in Zimbabwe, although the label of depression carries mental illness stigma, framing psychological treatment as care for *kufungisisa* makes it more acceptable and appealing. Similarly, in Haiti, describing psychosocial and psychological support as an approach to help with *reflechi twòp* helped lay community counselors encourage persons with psychological distress to participate in care. In Peru and other Latin American countries, cognitive behavioral therapy has been a treatment for *pensando mucho* in programs for women with perinatal depression.

Cultural concepts of distress can also be added to mental health screening tools or used to develop new screening instruments to identify persons who would benefit from services. Standard biomedical screening tools such as the PHQ-9 often lead to high numbers of false positives when adapted for use in a new cultural context. Adding a question about a cultural concept of distress can help identify individuals with forms of distress for which it may be appropriate to seek help. In Nepal, for example, adding a question about problems in the “heart–mind” (*man*) as a required symptom reduced false positives on the PHQ-9. Moreover, patients referred based on both heart–mind problems, and high PHQ-9 scores had the shared understanding that they all experienced a similar type of distress.

Another approach to using cultural concepts of distress to promote engagement with different types of mental health services is through proactive case detection. This approach uses consultation with community members to identify persons in the community who may need mental health services. For example, teachers can be consulted to refer students with distress. Health workers can refer persons they encounter when delivering vaccinations or reproductive health services. Religious leaders and traditional healers can refer members of their congregations and communities. Cultural concepts of distress can be an effective way to explain who would benefit from referrals when communicating with these community members who are not mental health professionals. For example, in Haiti, Vodou priests (*Hougan*) have been trained to refer persons suffering from ongoing *reflechi twòp* to be evaluated for mental health services. Similarly, in the Democratic Republic of Congo, community members referred mothers who had *maladi ya souci* (a sickness of worry) to participate in a culturally adapted version of cognitive behavioral therapy. Recently, the World Health Organization has adopted the Community Informant Detection Tool (CIDT), which was developed by Mark Jordans and colleagues as a way of using cultural concepts of distress in a prototype narrative to train community members to identify persons with distress and link them with mental health services.

Once people are engaged through acknowledging their cultural concepts of distress, it is important to consider how these concepts may align with the available types of service. Traditional medicine systems or religious healers may recognize the same terms, and the cultural concept then may help direct the individual toward appropriate support. For biomedical services, it is important to further evaluate how the individual's cultural concepts of distress may map onto, modify, or obscure diagnostic criteria for specific psychiatric conditions. However, the increasing availability of transdiagnostic treatments means that individuals with specific symptoms or behavioral problems included in cultural concepts of distress may benefit from psychological interventions that target these symptoms or problems. Cultural concepts can also be incorporated into stepped-care services, as has been done in global mental health programs in low- and middle-income countries. In stepped-care services, patients who present with distress framed in terms of a cultural concept associated with limited functional impairment may be treated with a brief psychological intervention, such as training in stress management or problem-solving techniques. However, the presence of more severe impairment or urgent symptoms such as suicidal thoughts or behavior indicates the need for a higher tier of services, or potentially, emergency care. Similarly, endorsement of a cultural concept of distress among patients who also met criteria for a specific psychiatric diagnosis may require a disorder-specific pharmacologic and/or psychological intervention.

Cultural Concepts of Distress and Promotion of Collaborative Care

The efficacy of indigenous interventions may stem in part from their fit with local cultural explanations and expectations. Including elements of local concepts of distress in both the explanation of the treatment plan and in specific interventions may then increase effectiveness. For example, Tibetan, Mongolian, Chinese, and other related medical traditions have developed treatments for worry and rumination associated with "wind"-based and *qi*-based conditions, such as *rlung, yadargaa*, and *shenjing shuairou*. By understanding the healing theories associated with these cultural concepts of distress, biomedical mental health practitioners can better understand who may benefit from these healing practices. This runs counter to assumptions that traditional healing practices will have universal benefits for all types of patients and any type of depressive or anxiety disorder. Rather, the greatest benefit may be among those individuals who experience specific aspects of depression or anxiety but also endorse relevant aspects of the cultural concept of distress. Moreover, cultural concepts of distress can serve as a communication bridge among different types of care providers and support referrals when needed from one type of care system to another. In Haiti, persons suffering from *reflechi twòp* with suicidality report greater benefits when receiving care from a *Hougan* (Vodou priest). This suggests that care for suicidality may be optimal when both mental health specialists and *Hougan* are involved if that is the preference of the patient and family. While such cultural information can guide the development of collaborative care systems and services, in each case, it is crucial to move beyond cultural stereotypes and generalizations to determine individual patients' experiences, expectations, and preferences.

Cultural Concepts of Distress and Monitoring and Evaluation of Mental Health Services

Another role that cultural concepts of distress can play is in monitoring the progress and outcomes of treatment. When biomedical interventions are evaluated, DSM and ICD symptoms are typically used to determine if treatment is progressing and deemed to be successful. However, this may miss aspects of distress that are most salient to the person receiving care. Monitoring progress with measures that track culturally salient symptoms and outcomes can provide a more accurate assessment. For example, if a major concern of a patient is “thinking too much,” then this can be tracked during care to see if the individual feels that the distress is improving by monitoring the relevant local cultural concept. As described above in relation to the Cultural Formulation Interview, when clinicians prioritize the concerns of the patient, this will help promote patient engagement and guide decisions about what treatment will be most beneficial. At a systems level, an intervention or a whole system of care can be evaluated by monitoring changes in cultural concepts of distress. For example, in Haiti, *reflechi twòp* has been included in the Kreyòl Distress Inventory as an outcome measure. Similarly, in Zimbabwe, the Shona Symptom Questionnaire includes *kufungisisa* as one of the symptoms to evaluate the outcome of mental health interventions. For treatments of Cambodian refugees, the Cambodian Somatic Symptom and Syndrome Inventory (CSSI) was developed and used to determine the effectiveness of both medications and culturally adapted psychotherapy.

When evaluating interventions using cultural concepts of distress, it is important that results be transparently communicated. For example, in a study in East Africa of a group psychological treatment, the depression symptom measure was validated against a cultural concept for traumatic grief. However, when the study was published, it was reported as establishing evidence for treating depression. It would have been more accurate to describe the intervention as having local evidence for resolution of traumatic grief, as it was described in the local context because that construct is closer to the cultural concept of distress used to locally adapt the symptom–outcome measure. Again, the broad issue is that local cultural concepts should not simply be assimilated to standard psychiatric concepts and categories because they provide additional information about the personal and social meaning and consequences of distress.

Cultural Concepts of Distress and Adaptation and the Development of Mental Health Interventions

Finally, cultural concepts of distress can play an important role in the development and adaptation of clinical interventions, especially psychological and interpersonal or social interventions. One of the best predictors of effective cultural adaptation is when the adaptation addresses the explanatory models of participants. Eva Heim and Brandon Kohrt developed a cultural adaptation framework for psychological interventions centered around cultural concepts of distress as the core of the adaptation process. They suggest focusing the beginning of the intervention on cultural concepts of distress to present the intervention in ways that are readily understood by patients. The pathways by which the intervention alleviates psychological distress should consider the potential causes identified in cultural concepts of distress to the extent possible.

The Friendship Bench study in Zimbabwe provides a clear example of how a local cultural concept of distress for "thinking too much," *kufungisisa,* was central in developing, implementing, and evaluating a psychological intervention. The Zimbabwean psychiatrist Dixon Chibanda developed the Friendship Bench psychological intervention based on principles of problem-solving therapy, but the intervention is framed according to the concept of how to manage "thinking too much" (*kufungisisa*). In the intervention, problem solving is explained as an activity to reduce *kufungisisa*. Moreover, *kufungisisa* is not simply used as a substitute for depression but is linked to other aspects of Shona ethnopsychology in Zimbabwe. In particular, *kufungisisa* is ameliorated through the practices of *kuvhura pfungwa* ("opening of the mind"), *kusimudzira* ("uplifting"), and *kusimbisa* ("strengthening"). The intervention also makes use of cultural notions of social space and interpersonal interaction. The intervention is delivered on a bench outside the hospital. The person delivering the intervention is thought of as a "helpful grandmother-type" figure who listens closely to one's needs and provides sage advice. Discussing excessive thoughts with a trusted family member or other confidant is what one would typically do when experiencing *kufungisisa*. Parts of the Friendship Bench intervention are delivered in groups. In this context, it is easy to explain the purpose of the group to all participants because they typically have a shared experience and understanding of *kufungisisa*. This focus is more comprehensible and less stigmatizing than framing the group as depression treatment. In a randomized controlled trial comparing the Friendship Bench with a waitlist control, participants in the Friendship Bench showed a greater reduction on the Shona Symptom Questionnaire, which includes *kufungisisa*. In addition, they showed improvement in depression symptoms. Six months after receiving the Friendship Bench psychological intervention, only 13% of patients were depressed, compared with 50% of persons in a basic treatment comparison group, which included a nurse-led evaluation, brief support counseling, option for medication, education, and support for common mental disorders, and referral to a psychiatric facility if appropriate. The Friendship Bench thus illustrates the power of framing a psychological intervention according to local concepts and the benefit this has both for reducing the cultural concept of distress as well as symptoms of a psychiatric disorder. There are similar examples of adapting interventions based on *pensando mucho* among Latin American populations, *kut careen* and other Khmer idioms for Cambodian refugees, and concepts of worries in the heart–mind and tension in South Asia.

CONCLUSION

Integrating cultural concepts of distress into mental health services is essential to culturally informed care. It is a deliberate process that uses information about the relationship between professional classification systems and local traditions of distress to accurately translate the person’s experience into clinically useful formulations and diagnoses. Integrating this local information into clinical practice can also help identify individuals with psychological distress; enhance their engagement in mental health services; provide effective psychoeducation; facilitate collaboration among biomedical mental health specialists and traditional healers and helpers; monitor individual patient-level and service-level outcomes; and guide the cultural adaptation and implementation of mental health interventions. This chapter has outlined key areas for this translation and implementation process.

FURTHER READINGS

Alarcón RD, ed. *Cultural Psychiatry*. Karger; 2013.

Bass JK, Ryder RW, Lammers MC, Mukaba TN, Bolton PA. Post-partum depression in Kinshasa, Democratic Republic of Congo: validation of a concept using a mixed-methods cross-cultural approach. *Trop Med Int Health*. 2008;13(12):1534–1542.

Benish SG, Quintana S, Wampold BE. Culturally adapted psychotherapy and the legitimacy of myth: a direct-comparison meta-analysis. *J Couns Psychol*. 2011;58(3):279–289.

Bhugra D, Bhui K. *Textbook of Cultural Psychiatry.* 2nd ed. Cambridge University Press; 2018.

Chang DF, Myers HF, Yeung A, Zhang Y, Zhao J, Yu S. Shenjing Shuairuo and the DSM-IV: diagnosis, distress, and disability in a Chinese primary care setting. *Transcult Psychiatry*. 2005;42(2):204.

Cork C, Kaiser BN, White RG. The integration of idioms of distress into mental health assessments and interventions: a systematic review. *Global Ment Health (Camb)*. 2019;6:e7.

Groleau D, Young A, Kirmayer LJ. The McGill Illness Narrative Interview (MINI): an interview schedule to elicit meanings and modes of reasoning related to illness experience. *Transcult Psychiatry*. 2006;43(4):671.

Guarnaccia PJ, Lewis-Fernández R, Pincay IM, et al. Ataques de nervios as a marker of social and psychiatric vulnerability: results from the NLAAS. *Int J Soc Psychiatry*. 2010;56(3):289.

Guarnaccia PJ, Rivera M, Franco F, Neighbors C. The experiences of ataques de nervios: towards an anthropology of emotions in Puerto Rico. *Cult Med Psychiatry*. 1996;20(3):343.

Gureje O, Nortje G, Makanjuola V, Oladeji BD, Seedat S, Jenkins R. The role of global traditional and complementary systems of medicine in the treatment of mental health disorders. *Lancet Psychiatry*. 2015;2(2):168–177.

Hall BJ, Chang K, Chen W, Sou KL, Latkin C, Yeung A. Exploring the association between depression and *shenjing shuairuo* in a population representative epidemiological study of Chinese adults in Guangzhou, China. *Transcult Psychiatry*. 2018;55(6):733–753.

Heim E, Kohrt BA. Cultural adaptation of scalable psychological interventions: a new conceptual framework. *Clin Psychol Eur*. 2019;1(4):e37679.

Hinton DE, Kredlow MA, Pich V, Bui E, Hofmann SG. The relationship of PTSD to key somatic complaints and cultural syndromes among Cambodian refugees attending a psychiatric clinic: the Cambodian Somatic Symptom and Syndrome Inventory (CSSI). *Transcult Psychiatry*. 2013;50(3):347–370.

Hinton DE, Barlow DH, Reis R, de Jong J. A transcultural model of the centrality of “Thinking a lot” in psychopathologies across the globe and the process of localization: a Cambodian refugee example. *Cult Med Psychiatry*. 2016;40(4):570–619.

Jordans MJD, Luitel NP, Lund C, Kohrt BA. Evaluation of proactive community case detection to increase help seeking for mental health care: a pragmatic randomized controlled trial. *Psychiatr Serv*. 2020;71(8):810–815.

Kaiser BN, Haroz EE, Kohrt BA, Bolton PA, Bass JK, Hinton DE. “Thinking too much”: a systematic review of a common idiom of distress. *Soc Sci Med*. 2015;147:170–183.

Kirmayer LJ, Guzder J, Rousseau C, eds. *Cultural Consultation: Encountering the Other in Mental Health Care*. Springer; 2014.

Kirmayer LJ, Lemelson R, Cummings CA, eds. *Re-Visioning Psychiatry: Cultural Phenomenology, Critical Neuroscience, and Global Mental Health*. Cambridge University Press; 2015.

Kleinman A. *Social Origins of Distress and Disease: Depression, Neurasthenia, and Pain in Modern China*. Yale University Press; 1986.

Kohrt BA, Rasmussen A, Kaiser BN, et al. Cultural concepts of distress and psychiatric disorders: literature review and research recommendations for global mental health epidemiology. *Int J Epidemiol*. 2014;43(2):365–406.

Lee S, Kleinman A. Are somatoform disorders changing with time? The case of neurasthenia in China. *Psychosom Med*. 2007;69(9):846–849.

Lewis-Fernández R, Aggarwal NK, Hinton L, Hinton DE, Kirmayer LJ, eds. *The DSM-5 Handbook on the Cultural Formulation Interview*. American Psychiatric Publishing; 2016.

Lewis-Fernández R, Guarnaccia PJ, Martínez IE, Salmán E, Schmidt A, Liebowitz M. Comparative phenomenology of ataques de nervios, panic attacks, and panic disorder. *Cult Med Psychiatry*. 2002;26(2):199–223.

Lewis-Fernández R, Kirmayer LJ. Cultural concepts of distress and psychiatric disorders: understanding symptom experience and expression in context. *Transcult Psychiatry*. 2019;56(4):786–803.

Lim RF, ed. *Clinical Manual of Cultural Psychiatry*. 2nd ed. American Psychiatric Publishing; 2015.

López SR, Guarnaccia PJ. Cultural dimensions of psychopathology: the social world’s impact on mental illness. In: Maddux JE, Winstead BA, eds. *Psychopathology: Foundations for a Contemporary Understanding*. Routledge; 2012:45.

Nichter M. Idioms of distress revisited. *Cult Med Psychiatry*. 2010;34(2):401–416.

Sangraula M, Kohrt BA, Ghimire R, et al. Development of the mental health cultural adaptation and contextualization for implementation (mhCACI) procedure: a systematic framework to prepare evidence-based psychological interventions for scaling. *Global Ment Health (Camb)*. 2021;8:e6.

Sapkota RP, Brunet A, Kirmayer LJ. Characteristics of adolescents affected by mass psychogenic illness outbreaks in schools in Nepal: a case-control study. *Front Psychiatry*. 2020;11:493094.

Thomas JJ, Eddy KT, Ruscio J, et al. Do recognizable lifetime eating disorder phenotypes naturally occur in a culturally Asian population? A combined latent profile and taxometric approach. *Eur Eat Disord Rev*. 2015;23(3):199–209.

29.2 Cultural Assessments in Psychiatry

NEIL KRISHAN AGGARWAL, M.D., M.B.A., M.A., AND ROBERTO LEWIS-FERNÁNDEZ, M.D., M.T.S.

Cultural psychiatry has been the medical subspecialty most involved in developing clinical, research, and educational methods for patient cultural assessments. Cultural psychiatry—also known as cross-cultural, transcultural, or anthropologic psychiatry—has typically focused on three areas: (1) the extent to which manifestations of psychopathology are similar or different across societies, (2) the provision of mental health services for underserved ethnoracial groups in multicultural societies, and (3) the cultural bases of psychiatric knowledge and practice to reveal how mental health systems reflect underlying societal concerns and values. Cultural assessments play a fundamental role in each area, clarifying the relative biologic and environmental contributions of symptom manifestations; patient preferences on how, where, when, and with whom to receive care; and cultural assumptions within the patient–clinician relationship such as communication and relational styles. The inclusion of the Outline for Cultural Formulation (OCF) in *DSM-IV* and the Cultural Formulation Interview (CFI) in *DSM-5* indicated growing acceptance of cultural assessments within psychiatry, not just for underserved ethnoracial groups, but for *all* people, including those whose identities are defined through gender, sexual orientation, migration, and veteran status.

HISTORICAL OVERVIEW

Cultural assessments in psychiatry have evolved alongside changing definitions of culture, especially within disciplines such as anthropology and psychology. In 1904, the renowned psychiatrist Emil Kraepelin—one of the first to differentiate *dementia praecox* (schizophrenia) from manic-depressive illness—sailed to Java, Indonesia, and investigated symptom manifestations in 100 European, 100 indigenous, and 25 Chinese patients at the Buitenzorg mental hospital. Kraepelin found that Javanese patients with psychosis experienced fewer hallucinations and delusions than European patients. He concluded that “the Javanese mind” was more primitive than “the European mind.” Kraepelin carefully recorded patient symptoms to examine similarities and differences between both groups, assuming that the definition and expression of psychotic illness would be identical in both populations. Although his assumptions and interpretations were based on racist intellectual trends from 19th-century physical anthropology that posited hierarchical relationships among races based on supposed biologic differences, Kraepelin inaugurated the field of comparative psychiatry and was among the first psychiatrists to use interpreters.

Another landmark movement emerged in the 1920s and 1930s with the Culture and Personality School. Social anthropologists such as Gregory Bateson, Ruth Benedict, Margaret Mead, and Ralph Linton collaborated with psychoanalysts such as Abram Kardiner and drew from Sigmund Freud’s theories on personality structure and organization. Many researchers tried to apply psychoanalytical theories to ethnographic data. This school contended that culture molds personality types, especially during child development, and that different cultural contexts produce different personality types based on socialization practices. For this reason, its theorists held that studies of individual personalities could lead to insights into general characteristics of prototypical national cultures. The school gained prominence during World War II with the American government using “National Character Studies” to understand the character and strategies of the Axis Powers, as exemplified by Benedict’s classic work on the Japanese, *The Chrysanthemum and the Sword*. Postwar, the school lost influence for suggesting that all members in a society act similarly through a prototypical personality, for assuming that culture is static rather than dynamically created and debated by society’s members, and for insufficiently demonstrating robust links between child development practices and adult personality traits.

These examples demonstrate how the process of conducting cultural assessments has changed based on contemporary understandings of culture.

Current trends in cultural assessments can be traced to the extensive collaboration among cultural mental health experts and interpretive medical anthropologists since the 1970s. A substantial scholarship has grown from this collaboration, which is based on the following principles: (1) formal psychiatric diagnoses may not be translatable across all societies, and empirical data must demonstrate their applicability in each society to avoid a “category fallacy,” the assumption that constructs (such as those related to mental health and illness) developed in one cultural setting are equally valid in other settings; (2) the clinician’s interview bears a resemblance to the anthropologist’s ethnography in attempting to reconcile personal lived experience with scientific theory; and (3) a clinical interview can approximate ethnography by understanding the patient’s explanatory model (EM) of mental illness, such as its causes, onset, mechanism, treatments, course, and severity. This orientation acknowledges that individuals belong and contribute to their cultural groups without assuming the homogeneity of cultural experience. In addition, this orientation acknowledges that professional biomedical theories and practices are also cultural products and that successful treatment requires negotiation between patient and clinician EMs. These principles have laid the foundation for four main ethnographic interviews used in patient cultural assessments: psychiatrist-anthropologist Arthur Kleinman's EM framework (1980), the Explanatory Model Interview Catalogue (EMIC) (1992), the Short Explanatory Model Interview (SEMI) (1998), and the McGill Illness Narrative Interview (MINI) (2006). These cultural assessments identify key domains that can be used in evaluating all patients irrespective of their cultural background.

Kleinman's Explanatory Model Framework

Arthur Kleinman's work in medical anthropology and cultural psychiatry during the 1970s centered on EMs, the culturally based understandings of sickness and treatment that patients and practitioners use. Kleinman contended that the study of interactions among patient and practitioner EMs could introduce a method to understand cultural influences in clinical communication. He identified five themes in illness EMs: (1) etiology; (2) time and mode of onset of symptoms; (3) pathophysiology; (4) course of sickness, including the degree of severity; and (5) treatment. Differences in patient–practitioner EMs could be analyzed as the differences between illness and disease, the former referring to the patient's lived experience of sickness and the latter referring to the practitioner's professional knowledge. Kleinman believed that analyzing patient EMs could illuminate the personal meanings of sickness and help-seeking preferences. His framework was influential in cultural psychiatry, primary care, and medical anthropology as a form of readily applied ethnography in clinical settings throughout the 1980s and 1990s. Subsequent patient cultural assessment interviews have variously engaged with Kleinman's EM framework to expand in other directions.

The Explanatory Model Interview Catalogue

Psychiatrist-epidemiologist Mitchell Weiss and his colleagues developed the EMIC as a method for cross-cultural research in psychiatric epidemiology based on the EM framework. The abbreviation EMIC invokes the anthropologic distinction between *emic* cultural knowledge rooted in local communities and *etic* theoretical knowledge of outside professionals, as in the illness/disease distinction mentioned earlier. Kleinman warned against a mechanical use of EMs that could reduce a patient's culture to decontextualized clinical information such as a lab result or physical examination finding; this may have prevented him from elaborating a research methodology beyond unstructured in-depth interviews. In contrast, Weiss designed the EMIC to operationalize the EM framework with a clear methodology, enable systematic comparisons across sites, and enhance diagnostic epidemiology with cultural epidemiology to analyze the relationships of illness understandings and practices with clinical outcomes. Based on research with patients having mental illness and leprosy in India, Weiss also created the EMIC to quantitatively rate narrative differences among patients; for example, noting whether a patient answered about a perceived illness cause in direct response to a question or in passing during a response to another question. EMIC instruments follow a tailored question format and set of domains rather than reproducing a single, unvarying interview; therefore, an EMIC instrument is customized for every research question. Starting from general introductory questions for each domain—which may be similar across EMICs—each iteration of the interview progresses to specific questions targeting a range of local illness experiences and interpretations. EMICs include assessment domains such as (1) patterns of distress, (2) perceived illness causes, (3) help seeking and treatment, (4) general illness beliefs, and (5) disease-specific queries. In addition to operationalizing Kleinman's EM framework, the EMIC introduces this final domain to assess the patient's general understanding of a specific disease. This domain is intended to ascertain the extent to which the patient's individual illness explanations resemble or differ from other explanations in the patient's cultural background to which the individual is exposed. This domain addresses a criticism of Kleinman's EM framework that it relies too heavily on individual explanations of illness without considering their degree of representativeness within a particular cultural grouping.

The Short Explanatory Model Interview

Criticisms emerged that the EM framework was underspecified and that the EMIC was too detailed and time-consuming. Psychiatric researchers in England developed the SEMI as a standardized, valid tool to assess common mental disorders in large-scale fieldwork and primary care settings. The SEMI covers five sections: (1) the patient's individual and social background; (2) the nature of the presenting problem; (3) help-seeking behavior, including contact with nonmedical healers; (4) interactions with the clinician/healer to clarify roles and expectations for promoting satisfaction; and (5) illness understandings, covered through a presentation of three vignettes for common mental disorders—depression, phobia, and somatization—followed by open-ended questions to capture the respondent's attitudes to the vignettes. The interview was tested in four primary care sites in England and Zimbabwe. In addition to assessing patient EMs and general views on mental illness, the SEMI expands cultural assessment to inquire about the patient's background and expectations of care to inform current treatment.

The McGill Illness Narrative Interview

The final ethnographic interview comes from researchers at McGill University in Canada, who raised concerns that the EM framework may be too unstructured and incomplete for systematic comparison and that the EMIC and SEMI may be too structured for narrative spontaneity. The MINI incorporates medical anthropologist Allan Young's critique that the EM framework assumes logical, coherent accounts of an illness experience. Young's work demonstrated that individuals use multiple representations and modes of reasoning to produce narratives of illness that can be complex, internally inconsistent, and even self-contradictory. Many patients may not be able to stipulate the causes of an illness, preferring explanations based on illness prototypes common to people in their communities (e.g., learned through popular myths or media accounts) or based on chain complexes in which people reason by association, such as through a chain of events (e.g., that a symptom and a potential cause are related because they happened close together). The elicitation of illness prototypes recognizes broad social influences on an individual's meanings of illness, whereas chain complexes acknowledge reasoning based on categories salient to the patient rather than predetermined EM domains. To capture such variation, the MINI consists of open- and closed-ended interview components.

The *DSM-IV* Outline for Cultural Formulation

These four assessment interviews generated significant interest among mental health experts conducting cultural assessments in clinical practice. Anthropologists developed a related but independent scholarship demonstrating that the health care system can be analyzed as a cultural system and that psychiatric diagnosis is a form of clinical interpretation influenced by social networks and institutions. Throughout the 1980s, a consensus emerged among mental health clinicians and anthropologists that patients and providers—independently and together—routinely create and debate meanings of health, illness, and treatment. Psychiatrists and anthropologists have agreed that such acts of meaning making merit attention to close gaps between patient and provider understandings during any sickness. Mental health clinicians have found patient cultural assessments helpful to clarify mismatches between symptom manifestations and official diagnostic criteria or to engage patients in treatment based on their preferences for care; anthropologists have concentrated on the extent to which illness narratives reflect local patient worlds coming into contact with the biomedical system.

For the *DSM-IV* revision process, the U.S. National Institute of Mental Health (NIMH) convened a conference in April 1991 at the University of Pittsburgh among cultural experts, the *DSM-IV* task force, and workgroup representatives to enhance the new Manual's cultural validity. Cultural experts included cultural psychiatrists and psychologists, medical anthropologists, and members of underserved ethnic and racialized groups within the American Psychiatric Association (APA) leadership. Following this conference, the NIMH Group on Culture and Diagnosis developed recommendations for

DSM-IV, such as an introduction to the upcoming manual on the importance of culture, a section on cultural considerations for each diagnosis, a glossary of “cultural syndromes and idioms of distress,” a cultural commentary on *DSM-IV’s* multiaxial evaluation format, and guidelines for a cultural formulation. The NIMH group based the elements of its OCF on a literature review, demonstrating four domains that were common across patient cultural assessments: (1) cultural identity of the individual, (2) cultural explanations of the individual’s illness, (3) cultural factors related to the psychosocial environment and levels of functioning, and (4) cultural elements of the patient–clinician relationship. A final domain was included to record how cultural information could influence diagnosis and care. Clinicians could use the OCF to supplement the multiaxial diagnostic assessment and to describe the patient’s cultural and social reference groups in clinically relevant ways. Table 29.2–1 lists OCF domains for assessment.

Table 29.2–1.
The DSM-IV Outline for Cultural Formulation

Cultural Formulation Domain	Subtopics for Assessment
Cultural identity of the individual	▶ Individual's demographic and other socially and culturally defined characteristics ▶ Aspects of identity that are prioritized by the individual and how they interact (intersectionality) ▶ Degree of involvement with the cultural contexts of origin and the new cultural contexts (for immigrants and ethnic minorities) ▶ Degree of interaction and identification with own group and other segments of society (for those who identify with racialized and ethnic groups) ▶ Language abilities, use, and preference (including multilingualism) and their impact on access to and processes of care
Cultural concepts of distress	▶ Cultural constructs (idioms of distress, explanations, and syndromes) that influence how symptoms or problems are communicated ▶ Meaning and perceived severity of individual's distressing experience in relation to cultural norms

	► Priority symptoms, perceived seriousness of the illness, level of associated stigma, and anticipated outcomes ► Current preferences for and past experiences with professional and popular sources of care
Psychosocial stressors and cultural features of vulnerability and resilience	► Culturally relevant interpretations of social stressors, challenges, supports, and levels of functioning and disability ► Social determinants of mental health, including access to resources and opportunities, exposure to racism, discrimination, systemic institutional stigmatization, and social marginalization or exclusion (structural violence) ► Role of religion and interpersonal networks in causing stress or providing support
Cultural features of the relationship between the individual and the clinician, treatment team, and institution	► Individual differences in cultural background and identity between the individual and the clinician, treatment team, or institution ► Problems that these differences may cause in diagnosis and treatment (e.g., difficulties in eliciting symptoms and understanding their cultural significance, in determining whether a behavior is normal or pathologic, or in establishing rapport)

Overall cultural assessment	▶ Discussion of how cultural considerations specifically influence comprehensive diagnosis and care

The OCF’s inclusion in *DSM-IV* provided an institutional mechanism for mental health experts and anthropologists around the world to appraise the current state of knowledge, harmonize efforts at patient cultural assessments, and suggest future directions. In contrast to the multiaxial evaluation with its standardized codes, the OCF was designed for clinicians to narratively describe the patient’s account of suffering from an *emic* perspective rich with contextual details. The use of the word “formulation” was an attempt to stimulate the type of narrative analysis already familiar to psychiatrists trained in psychodynamic formulations who assess details of a patient’s psychological conflicts in the context of life experiences. The formulation was also designed to standardize a methodology for cultural assessment and produce outcome studies in mental health services research. Field trials with various American ethnicities validated its utility and offered tangible clinical examples.

However, the OCF faced challenges right from its inclusion. Although the NIMH group recommended its placement at the beginning of the Manual, the OCF was placed in the ninth appendix. It was also combined with the *Glossary of Cultural Syndromes and Idioms of Distress* and renamed by the *DSM-IV* leadership as the *Glossary of Culture-Bound Syndromes*, a decision that tended to exoticize the OCF as if it should be used only with unusual cultural presentations rather than with every patient. Despite these obstacles, the OCF has been used worldwide since its publication in 1994. While much of this work has found the OCF a useful clinical tool, problems have been identified. Its format as a description of domains to be assessed, lack of implementation guidelines, and brevity—which required that clinicians already be familiar with cultural topics or social science theories—hindered widespread clinical use. Nonetheless, the OCF remains the most widely taught method of cultural assessment in U.S. psychiatric residency programs. In addition, psychiatrists and psychologists in Canada, Denmark, England, the Netherlands, Sweden, and the United States have independently operationalized the OCF's implementation. These efforts range from brief semistructured interviews for individual clinicians to use with every patient to cultural consultation services that receive referrals from other health professionals and utilize cultural and linguistic brokers. A major contribution of the OCF has been to organize subsequent research along with its four domains of cultural assessment, facilitating its operationalization into the CFI in *DSM-5*.

KEY DOMAINS IN PATIENT CULTURAL ASSESSMENTS

Cultural Identity of the Individual

Cultural psychiatrists assume that understanding a patient's identity is indispensable to a thorough clinical evaluation. Early OCF guidelines suggested that cultural identity assessments be conducted through an "interpersonal grid" to record potentially relevant variables such as ethnicity, race, geographical origin, language, acculturation, gender, age, sexual orientation, religious or spiritual beliefs, socioeconomic class, and education. This suggestion can be seen as a precursor to contemporary understandings of intersectionality, the idea that individuals belong to multiple groups through which identities are ascribed and self-selected. However, the recommendation for an interpersonal grid faced criticism. The group raised concerns about stereotyping patients or minimizing sociocultural phenomena based on a clinician's determination of the patient's identity. Others have advocated a broader approach that accounts for changes in self-reported identity based on context (i.e., interviewer, setting), and that includes the patient's perception of the clinician's identity. Assessments of cultural identity can help clinicians discern whether forms of social belonging, stigma, and language fluency impact symptom experience, expression, and interpretation in certain cultural contexts and settings.

Studies show that many psychiatric trainees may feel uncomfortable and invasive in asking patients about their identities without prior training. Nearly all OCF-inspired interviews include common elements to assess identity, such as the cultural groups with which a patient identifies; the levels of importance and meaning ascribed by the patient to these identifications; evaluations of language fluency and preferences; and assessments of patients' involvement in their cultural contexts of origin and the new cultural contexts in the case of immigrants, especially, including positive and negative experiences with involvement in both contexts. Aspects of identity in some OCF-based interviews included evaluations of whether patients experience multiple identities simultaneously or over time and how cultural identity impacts the clinical problem. Several OCF-based interviews also included detailed migration histories. All OCF-based interviews have used open-ended questions to encourage interaction and discussion, rather than closed-ended checklists. All authors have also warned against an oversimplification of cultural factors in which social processes such as adverse experiences associated with migration and discrimination are mistakenly understood as inherent and representative of the cultural group itself rather than as time-limited and context-driven reactions to social predicaments.

Studies published since the OCF have highlighted unique challenges for certain populations. Some refugees may be suspicious of questions about identity, and many migrants may not know how to discuss their backgrounds in new societies. Children and adolescents may have different cultural identities than their parents, depending on mixed parentage, adoption, and socialization or acculturation. The cultural identity of older patients may be shaped by cultural meanings of old age based on their social roles and status. For all patients, cultural identity should be explored with reference to social context, developmental phase, and the possibility of change over time. Clinicians should especially endeavor to clarify which aspects of identity patients consider to be ascribed by others and chosen for themselves.

Cultural Concepts of Distress

EMs of illness have underpinned most studies on patient cultural assessments in mental health and anthropology. Compared to the original assumption of EMs as cognitively linear, later research has shown that EMs can exhibit metaphorical thinking, no clear causes, sequences of events linked by patients based on contiguity or metonymy in ways that may be unclear to others, and images or examples of illness used analogically to understand the current illness episode. EMs can change based on context, including the language, interviewer, and setting in which they are elicited; interactions with close associates, the community, or the health care system; acculturation status for immigrants and refugees; and experiences of discrimination within the health care system. Attempts to code illness narratives from patients with medically unexplained symptoms revealed three types of EMs: *nominal*, in which a condition was named without elaborating on causal attribution or mechanism; *causal attributions*, in which a symptom or illness was linked to a cause; and *process or mechanism explanations*, in which mechanistic or metaphorical devices linked cause and symptom.

Culture and context influence EMs. Clinical narratives reflect broader cultural norms affecting communications with mental health care providers, family, and friends. Access to the illness representations of crucial members in the person’s extended social network (either directly or as reported by the patient) provides a range of possible views influencing the person’s experience. The organization of health care services in a society may also determine patients’ attitudes toward disease, diagnosis, and treatment based on familiarity with the health care system. For this reason, mental health clinicians should evaluate culture and context when working with *all* patients since the act of diagnosis still relies on linguistic communication within psychiatry.

Eliciting past treatment experiences and current treatment expectations are also important components of EMs. Cultural scripts associate a patient's illness representations with patterns of self-coping and help seeking, including perceived relevance, the efficacy of mental health treatments, and their potential dangerousness. Cultural adaptations of evidence-based therapies that take account of cultural scripts may improve patient engagement and response, highlighting the value of identifying cultural interpretations prior to treatment planning. Nonmedical forms of self-coping and help seeking are common, diverse, and merit clinical attention because they may complement or conflict with biomedical care. Barriers to utilizing mental health services are also determined by cultural priorities for care, family values and interactions, and local health system and socioeconomic factors affecting access to mental health services. A sound cultural assessment strives to situate EMs against the relevant social and health contexts of the patient.

Early theoretical guidelines on the OCF included the *culture-bound syndromes* in this section. *DSM-5-TR* substitutes the term "culture-bound syndrome" with three other terms: *cultural idioms of distress, cultural explanations,* and *cultural syndromes*. These terms denote the ways that suffering, behavioral problems, and troubling thoughts and emotions are experienced, understood, and communicated, either as symptom clusters, ways of expressing distress that may not involve specific symptoms, and/or as explanations or attributions. Key aspects of patients' clinical presentations may be patterned by such cultural concepts.

Nearly all OCF-based interviews include questions on illness-related idioms, the meaning and severity of symptoms, causes, expected course of illness, experiences with care, and treatment expectations. Nearly all interviews ask the patient to describe the views of close associates such as family and friends to situate the individual's perspective in the community context. This line of questioning attempts to operationalize the understanding of culture as shared and interpersonal. OCF-based instruments vary in assessing how the patient describes the clinical problem to close associates, the most troubling aspects of the clinical problem, the contribution of illness prototypes within a given cultural context to the patient's understandings, the most useful treatments received, treatment recommendations of close associates, and the patient's use of traditional healers and alternative services.

Psychosocial Stressors and Cultural Features of Vulnerability and Resilience

This domain encourages clinicians to clarify the context of a patient's illness—especially cultural interpretations of the stressors and supports associated with onset and course—and fears and hopes affecting levels of functioning. OCF-based interviews have used open-ended questions, with some interviews asking questions directly of the patient and others instructing clinicians to identify these data based on patient responses. In all cases, the goal is to obtain a description of the life context in which illness emerges and to elicit responses from the patient, the family, and the community. The subjective, lived experience of the patient's and social network's response is crucial because culture shapes the perceived centrality, magnitude, and treatability of the problem as well as forms of stress and support. This approach synthesizes the person's subjective experience of risk, resilience, and protective factors and the clinician's analysis of the patient's social and cultural context. For example, the subjective–objective approach might rely on a functional assessment of key activities of daily living that the clinician conducts alongside the patient's and social network's cultural interpretations of normative functioning.

The OCF was partly developed to address the adverse experiences of disadvantaged ethnoracial groups. Since its publication, studies have elaborated on additional social stressors. A recurrent finding has been the effect of migration status on the mental health of refugees as a social determinant of health independent of premigratory trauma. The factors most associated with risk for psychopathology include low socioeconomic status, poor housing, unemployment or underemployment, lack of official residency status, linguistic barriers, limited social networks, racial or ethnic discrimination, role strain, family conflict, status loss, acculturative stress, nostalgia, and bicultural tension. Acculturation issues may affect disempowered or marginalized ethnoracial groups, not just immigrants. Discrimination-related stress affects members of other social groups, such as religious minorities, lesbian–gay–bisexual–transgender–queer persons, those with lower socioeconomic status, and individuals who are differentially abled.

In clinical assessment, family, friends, and other close associates can provide valuable information about cultural context. Collateral information from relevant members of the patient's social network is critical for children and adolescents, older adults, and those with conditions affecting cognitive functioning such as dementia or psychosis that could interfere with the clinical interview. OCF-based interviews have varied in the extent to which they seek collateral forms of information, with some protocols for children and adolescents identifying multiple sources of information and perspectives. Although patient cultural assessments such as the OCF and its related interviews can accommodate the use of medical interpreters and cultural brokers, there is little guidance on how to operationalize this process in practice.

Finally, psychiatrists have reported that patients frequently turn to religion, spirituality, and secular moral traditions in understanding and responding to mental illness. A "religiosity gap" may exist among clinicians trained in empirical/positivistic scientific methods and patients searching for holistic treatments, especially within complementary and alternative systems. Cultural assessments should attend to the meanings of religion and other moral systems in crafting illness assumptions and beliefs, especially when a religious community provides emotional, social, and material support. Clinicians may be able to mobilize such information with patients to support recovery through groups and organizations aligned with these traditions.

OCF-based interviews have assessed stressors, supports, and levels of functioning differently. Variations include focusing on the impact of stressors, the views of the social network, the role of the person's partner, forms of religious/spiritual or moral support, barriers to receiving help, and the views of the social network on subsequent levels of functioning.

Cultural Features of the Relationship between the Individual and the Clinician, Treatment Team, and Institution

Many OCF case studies have illustrated the challenges confronting clinicians who provide cross-cultural care. These include tensions between biomedical forms of treatment and other forms of healing, the failure of health systems to provide interpreters, and difficulties in cross-cultural communication. Clinicians have discussed their reactions to consider the patient as an individual and a member of a social group. Sincere clinicians may encounter obstacles in addressing discrimination, racism, and subtle forms of cross-cultural bias during the medical interview. This domain of the OCF provides an opportunity to consider how such experiences affect patients' perceptions of care and promote clinician self-reflection on the challenges of delivering culturally sensitive health care.

Existing OCF-based interviews address this domain differently. Some provide queries to the clinician in the form of introspection, such as questions on mutual perceptions of power and the social positions of the patient and clinician, including the history of relationships between their respective cultural groups. This work operationalizes the concepts of "cultural transference" and "cultural countertransference" to analyze patient–physician relationships. Other instruments suggest questions that can be directly posed to the patient, such as the quality of communication during the interview, the need for interpreters, important topics missed by the interview, the perceived impact of ethnic or cultural matching with the clinician or lack thereof, and the role of the patient during the interview. Although this is a sensitive topic, early, open inquiry into past negative experiences of care and potential miscommunications with clinicians may facilitate rapport, engagement, and a commitment to overcome treatment barriers.

Future Directions for the OCF

DSM-5 included the OCF as it was published in *DSM-IV* with only minor edits. Since then, mixed-methods (qualitative–quantitative) analyses of patient cultural assessments and debriefing interviews have suggested revisions to the OCF that would (1) combine the current third and fourth domains into a single domain on key social relationships and (2) create a new domain to cover the social, cultural, political, and economic contexts of care. This suggestion comes from one clinical site, particularly in response to growing calls throughout academic medicine for greater attention to the structural contexts of care, such as the social determinants of health and systemic racism. Further studies are needed to validate these findings with data from other settings. This work indicates the ongoing relevance of the OCF to contemporary research and practice in cultural psychiatry.

The Cultural Formulation Interview in *DSM-5* and *DSM-5-TR*

Nearly 20 years of experience with the OCF and its related interviews since 1994 led to the *DSM-5* revision. The *DSM-5* Cross-Cultural Issues Subgroup (DCCIS) developed corresponding semistructured interviews to operationalize the OCF for clinicians. This set of interviews is known collectively as the CFI and consists of a core questionnaire of 16 items, an informant version of the core CFI to obtain collateral material from caregivers, and 12 supplementary modules that expand the core cultural assessment based on topics of interest (i.e., explanatory models) or populations with specific needs (i.e., children and adolescents, immigrants and refugees). The APA has made these instruments available for free at http://www.psych.org/practice/dsm/dsm5/online-assessment-measures.

The DCCIS was part of the larger *DSM-5* Gender and Cross-Cultural Issues Study Group and included experts in culture and mental health from diverse countries and cultural groups. One goal of the DCCIS was to improve cultural assessment within psychiatric diagnoses. The subgroup was charged by the *DSM-5* task force with making recommendations on ethnoracial, cultural, and contextual issues to the *DSM-5* workgroups, such as differences in risk factors, precipitants, symptom presentations, prevalence, symptom severity, and course of illness for each diagnosis. These recommendations have appeared within the text of *DSM-5* and were further expanded in *DSM-5-TR*.

In addition, the DCCIS undertook a critical review of the *DSM-IV* OCF. DCCIS participants prepared literature reviews on the OCF as a whole, each of its four domains, and key implementation questions raised by the use of the OCF in clinical care. Such topics included recommended guidelines for use, special populations in need of attention such as children, the older adults, or immigrants and refugees, and the format and content of previously developed interviews, protocols, and questionnaires operationalizing the OCF. DCCIS members provided recommendations on revising the OCF, including how to facilitate implementation through the CFI.

Based on these literature reviews and previous experiences of DCCIS members developing and using OCF-based cultural assessments, the OCF was revised slightly for *DSM-5*, mostly by making the *DSM-IV* text more explicit in places. Examples of this clarification process include the incorporation of elements of cultural identity not mentioned in *DSM-IV,* such as religious affiliation and sexual orientation, among others; clinician instructions to identify stressors and supports in the patient’s social environment; and mention of the impact of racism and discrimination on the clinician–patient relationship. These clarifications were expanded in *DSM-5-TR:* for example, by referring under cultural identity to intersectionality, the interaction of various aspects of the person’s identity, and by guiding clinicians to elicit the level of interaction and identification by individuals who identify with racialized and ethnic groups both with their own group and with other segments of society.

Other changes made the OCF more operational. An initial 14-item interview was developed largely from previous question lists, interview protocols, and interviews that operationalized the OCF; most of the authors of the OCF-based approaches were members of the DCCIS. The content of the questions was formulated, discussed, and debated in 2010 and 2011 during biweekly conference calls, supported by the *DSM-5* development process. The resulting 14-item version of the CFI was tested in a field trial from November 2011 to November 2012 at 12 clinical sites in the United States, Canada, Peru, the Netherlands, Kenya, and India. This trial utilized a standardized training format (http://www.ucl.ac.uk/ccs/specialist-services/#cfi) and research design to evaluate the CFI's acceptability, feasibility, and clinical utility among patients and clinicians. Field trial results were used to finalize the 16-item version in *DSM-5*. An informant version of the CFI was prepared to elicit collateral information from parents, caregivers, and other close associates. Supplementary modules were also developed for a more comprehensive assessment. Some modules expand various sections of the core CFI, as mentioned above.

Definition of Culture

To facilitate the application of the interview, a definition of *culture* was included in *DSM-5* and updated in *DSM-5-TR*. *Culture* is defined as (1) the processes through which people assign meaning to experience, drawing from the values, orientations, knowledge, and practices of the diverse social groups (e.g., ethnic, faith, occupational, and veterans groups) in which they participate; (2) aspects of an individual's background, developmental experiences, and current social contexts and position that may affect their perspective, such as age, gender, social class, geographic origin, migration, language, religion, sexual orientation, disability, or ethnic or racialized background; (3) the influence of family, friends, and other community members (the individual's social network) on the individual's illness experience; and (4) the cultural background of the health care providers, and the values and assumptions embedded in the organization and practices of health care systems and institutions, that affect the clinical interaction.

This definition appears in the text of *DSM-5-TR* and emphasizes the all-encompassing effects of culture on individual subjective experience, as well as the fundamental role of family and close associates in shaping conceptions of illness and help-seeking expectations. In the United States, culture is frequently conflated with ethnic and racialized categories, which may unintentionally stereotype patients. Other components of culture—for example, age cohort, occupation, and sexual orientation—are often not part of a cultural assessment. By contrast, the CFI encourages clinicians to elicit the contextual frame of each patient's experience—seeking to use information about the collective to understand an individual's perspective and to clarify how local environments influence the person's situation, including how sociocultural contingencies help pattern a set of events. In addition, the CFI focuses the cultural formulation on the patient's self-identified group rather than on visible group markers of identity that may inadvertently lead to stereotyping. Finally, the updated definition focuses attention on the cultural background of the clinician and the cultural assumptions that organize health care delivery as important cultural contributors to the clinical environment.

Table 29.2–2.
Cultural Formulation Interview (CFI)

Supplementary modules used to expand each CFI subtopic are noted in parentheses.

GUIDE TO INTERVIEWER	INSTRUCTIONS TO THE INTERVIEWER ARE ***ITALICIZED.***
The following questions aim to clarify key aspects of the presenting clinical problem from the point of view of the individual and other members of the individual's social network (i.e., family, friends, or others involved in current problem). This includes the problem's meaning, potential sources of help, and expectations for services.	*INTRODUCTION FOR THE INDIVIDUAL:* I would like to understand the problems that bring you here so that I can help you more effectively. I want to know about ***your*** experience and ideas. I will ask some questions about what is going on and how you are dealing with it. Please remember there are no right or wrong answers.

CULTURAL DEFINITION OF THE PROBLEM

CULTURAL DEFINITION OF THE PROBLEM

(Explanatory Model, Level of Functioning)

Elicit the individual's view of core problems and key concerns. *Focus on the individual's own way of understanding the problem.* *Use the term, expression, or brief description elicited in question 1 to identify the problem in subsequent questions (e.g., "your conflict with your son").*	1. What brings you here today? *IF INDIVIDUAL GIVES FEW DETAILS OR ONLY MENTIONS SYMPTOMS OR A MEDICAL DIAGNOSIS, PROBE:* People often understand their problems in their own way, which may be similar to or different from how doctors describe the problem. How would *you* describe your problem?

Ask how individual frames the problem for members of the social network.	2. Sometimes people have different ways of describing their problem to their family, friends, or others in their community. How would you describe your problem to them?
Focus on the aspects of the problem that matter most to the individual.	3. What troubles you most about your problem?

CULTURAL PERCEPTIONS OF CAUSE, CONTEXT, AND SUPPORT

CAUSES

(Explanatory Model, Social Network, Older Adults)

This question indicates the meaning of the condition for the individual, which may be relevant for clinical care.	4. Why do you think this is happening to you? What do you think are the causes of your [PROBLEM]?
Note that individuals may identify multiple causes, depending on the facet of the problem they are considering.	*PROMPT FURTHER IF REQUIRED:* Some people may explain their problem as the result of bad things that happen in their life, problems with others, a physical illness, a spiritual reason, or many other causes.
Focus on the views of members of the individual's social network. These may be diverse and vary from the individual's.	5. What do others in your family, your friends, or others in your community think is causing your [PROBLEM]?

STRESSORS AND SUPPORTS

(Social Network, Caregivers, Psychosocial Stressors, Religion and Spirituality, Immigrants and Refugees, Cultural Identity, Older Adults, Coping and Help Seeking)

Elicit information on the individual's life context, focusing on resources, social supports, and resilience. May also probe other supports (e.g., from co-workers, from participation in religion or spirituality).	6. Are there any kinds of support that make your [PROBLEM] better, such as support from family, friends, or others?
Focus on stressful aspects of the individual's environment. Can also probe, e.g., relationship problems, difficulties at work or school, or discrimination.	7. Are there any kinds of stresses that make your [PROBLEM] worse, such as difficulties with money, or family problems?

ROLE OF CULTURAL IDENTITY

(Cultural Identity, Psychosocial Stressors, Religion and Spirituality, Immigrants and Refugees, Older Adults, Children and Adolescents)

	Sometimes, aspects of people's background or identity can make their [PROBLEM] better or worse. By ***background*** or ***identity,*** I mean, for example, the communities you belong to, the languages you speak, where you or your family are from, your race or ethnic background, your gender or sexual orientation, or your faith or religion. 8. For you, what are the most

Ask the individual to reflect on the most salient elements of his or her cultural identity. Use this information to tailor questions 9–10 as needed.	important aspects of your background or identity?
Elicit aspects of identity that make the problem better or worse. *Probe as needed (e.g., clinical worsening as a result of discrimination due to migration status, race/ethnicity, or sexual orientation).*	9. Are there any aspects of your background or identity that make a difference to your [PROBLEM]?
Probe as needed (e.g., migration-related problems; conflict across generations or due to gender roles).	10. Are there any aspects of your background or identity that are causing other concerns or difficulties for you?
CULTURAL FACTORS AFFECTING SELF-COPING AND PAST HELP SEEKING	
SELF-COPING	
(Coping and Help Seeking, Religion and Spirituality, Older Adults, Caregivers, Psychosocial Stressors)	
Clarify self-coping for the problem.	11. Sometimes people have various ways of dealing with problems like [PROBLEM]. What have you done on your own to cope with your [PROBLEM]?
PAST HELP SEEKING	
(Coping and Help Seeking, Religion and Spirituality, Older Adults, Caregivers, Psychosocial Stressors, Immigrants and Refugees, Social Network, Clinician-Patient Relationship)	

Elicit various sources of help (e.g., medical care, mental health treatment, support groups, work-based counseling, folk healing, religious or spiritual counseling, other forms of traditional or alternative healing). *Probe as needed (e.g., “What other sources of help have you used?”).* *Clarify the individual’s experience and regard for previous help.*	12. Often, people look for help from many different sources, including different kinds of doctors, helpers, or healers. In the past, what kinds of treatment, help, advice, or healing have you sought for your [PROBLEM]? *PROBE IF DOES NOT DESCRIBE USEFULNESS OF HELP RECEIVED:* What types of help or treatment were most useful? Not useful?

BARRIERS

(Coping and Help Seeking, Religion and Spirituality, Older Adults, Psychosocial Stressors, Immigrants and Refugees, Social Network, Clinician-Patient Relationship)

Clarify the role of social barriers to help seeking, access to care, and problems engaging in previous treatment. *Probe details as needed (e.g., “What got in the way?”).*	13. Has anything prevented you from getting the help you need? *PROBE AS NEEDED:* For example, money, work or family commitments, stigma or discrimination, or lack of services that understand your language or background?

CULTURAL FACTORS AFFECTING CURRENT HELP SEEKING

PREFERENCES

(Social Network, Caregivers, Religion and Spirituality, Older Adults, Coping and Help Seeking)

Clarify individual’s current	Now let’s talk some more

perceived needs and expectations of help, broadly defined. *Probe if individual lists only one source of help (e.g., "What other kinds of help would be useful to you at this time?").*	about the help you need. 14. What kinds of help do you think would be most useful to you at this time for your [PROBLEM]?
Focus on the views of the social network regarding help seeking.	15. Are there other kinds of help that your family, friends, or other people have suggested would be helpful for you now?
CLINICIAN-PATIENT RELATIONSHIP (Clinician-Patient Relationship, Older Adults)	
Elicit possible concerns about the clinic or the clinician-patient relationship, including perceived racism, language barriers, or cultural differences that may undermine goodwill, communication, or care delivery. *Probe details as needed (e.g., "In what way?").* *Address possible barriers to care or concerns about the clinic and the clinician-patient relationship raised previously.*	Sometimes doctors and patients misunderstand each other because they come from different backgrounds or have different expectations. 16. Have you been concerned about this and is there anything that we can do to provide you with the care you need?

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

The Core CFI

Table 29.2–2 presents the core CFI interview. The interview has been designed like other semistructured interviews, such as the Structured

Clinical Interview for DSM Disorders (SCID), in which instructions to clinicians appear next to questions for patients. As emphasized in *DSM-5-TR*, the CFI follows a person-centered approach to cultural assessment by asking the patient about their views and those of close associates to avoid stereotyping by ethnoracial background or clinical expectations for a given illness. Clinicians are encouraged to voice the questions in their own style and ask probes as necessary to understand patient answers completely. The CFI is designed for use in the initial assessment of patients in all clinical settings, regardless of the cultural backgrounds of the patient or the clinician. Even patients and clinicians who seem to share the same cultural background may differ in ways that are critical to care, such as by educational level, socioeconomic status, or role expectations. The CFI may be used in its entirety or incorporated piecemeal; however, its developers recommend obtaining the person's narrative as an unbroken flow before proceeding to the rest of the clinical interview. The CFI may be especially helpful when there is difficulty in diagnostic assessment due to differences in the cultural, religious, or socioeconomic backgrounds of the clinician and the patient; when there is uncertainty about the fit between patient symptoms and diagnostic criteria; when it is difficult to judge illness severity or impairment; when patients hold views of symptoms or expectations of care that differ from clinicians' based on their experience with other cultural systems of healing and health care; when patients and clinicians disagree on the course of care; when there is mistrust of mainstream services and institutions by individuals with collective histories of trauma and oppression; or in cases of suboptimal treatment engagement and adherence. In response to criticisms about the absence of guidelines for the OCF, the DCCIS specifies instructions to clinicians to facilitate cultural assessments.

The core CFI is divided into four main domains. The first domain is the *Cultural Definition of the Problem* and consists of the first three questions. Question no. 1 is a broad, open-ended inquiry about the patient's presenting concerns. The prompt invites patients to describe their problems and emphasizes an understanding of the patient's illness, even if these descriptions are "similar to or different from how doctors describe the problem." Question no. 2 clarifies the patient's description of the problem as would be discussed with close associates, in recognition that culture influences information sharing according to the audience and social context. At this point in the interview, patients may be reticent to describe the problem in nonbiomedical ways that the clinician could find unusual and may therefore censor such descriptions. By framing the question in relation to the person's social network, the goal is to facilitate the patient's report of descriptions that are salient in the social environment. Question no. 3 on what troubles the patient most about the problem asks about what is most at stake for the individual during the illness.

Questions no. 4 to 10 belong to the second domain of the core CFI, *Cultural Perceptions of Cause, Context, and Support*. Questions no. 4, 6, and 7 focus the interview on the patient's EMs for the causes, social stressors, and social supports of the illness. Question no. 4 includes the placeholder "[PROBLEM]" to solicit the patient's preferred terms and phrases for how a problem is framed, known as *idioms of distress*. Once obtained, the patient's vocabulary can be substituted for the placeholder throughout the interview to build rapport and bridge patient–clinician linguistic differences in understanding the illness. In some cases, the patient may use biomedical vocabulary such as "I might have depression," but in other cases, terms referencing social determinants of health may be evident, such as "I worry about my rent, electricity, and phone bill." Terms communicating the presenting problem may vary greatly, and the CFI can reveal how patients frame issues so that clinicians can tailor treatment strategies.

Question no. 5 asks the patient to consider what close associates may understand as the cause of illness in response to the *DSM-5-TR* understanding of culture as the meanings and viewpoints transmitted within social groups. In this instance, the meanings being transmitted are the understandings of illness, and the social group is the patient's close associates. As with question no. 2 on the definition of the problem, framing the inquiry through members of the social network encourages a fuller range of responses that may influence the individual's EMs.

Questions no. 8, 9, and 10 also belong to this second domain but orient the interview toward the patient's cultural identity. Question no. 8 begins with a brief introduction that clinicians can use to clarify the meaning of *cultural background or identity* as relevant to health care. The introduction provides examples of identities that may be important for the patient, such as language, ethnicity, religion, and sexual orientation. These examples are meant to be representative in order to stimulate thought, not a fixed set of identities from which the patient must choose. Question no. 8 then asks the patient directly about the most important aspects of background or identity. This method of allowing the patient to name the preferred cultural identity departs from previous models of cultural competence that have made assumptions about patient identities based on a group affiliation such as ethnoracial background without accounting for their importance to the individual. For example, one patient in the *DSM-5* field trial initially named his identity as "Cuban" but then explained that identifying as a Christian was more helpful in understanding his suffering existentially. A model of cultural competence that assumed ethnicity was the most important group affiliation—without accounting for the patient's self-ascribed identity—could have overlooked the role of religion in providing meaning to the illness experience. Question no. 9 asks the patient how this identity may impact the current illness experience or patient-centered definition of the clinical problem. Question no. 10 asks the patient to consider how identity may cause other problems in life that may not be as obviously related to the illness but may be nonetheless important, such as problems with migration, gender roles, or intergenerational conflict.

Questions no. 11, 12, and 13 form the third domain known as *Cultural Factors Affecting Self-Coping and Past Help Seeking*. This domain encourages patients to share past forms of self-coping and help seeking in recognition that most forms of help are sought outside of the biomedical health care system. Question no. 11 addresses the patient's coping practices, and question no. 12 considers help seeking broadly to include help inside and outside of the biomedical system (i.e., religion-based support and support groups). An additional prompt for question no. 12 clarifies the types of help that have been most and least helpful. This information may aid the clinician in developing a treatment plan for the current illness episode. Question no. 13 asks about past barriers to treatment to help devise the current treatment plan around available resources.

The final three questions constitute the last domain of the CFI, *Cultural Factors Affecting Current Help Seeking*. Question no. 14 asks the patient about current treatment preferences, and question no. 15 explores treatment preferences recommended by close associates. As with questions on patient terms and causes for the presenting problem, this question is examined at the individual and social levels to explore how patients fashion cultural meanings within social groups. Question no. 16 begins with an open-ended statement to help the patient anticipate how perceived differences with the clinician can adversely affect care. By asking the patient about this potential barrier, the clinician can work to resolve such differences. Even if the question is not answered during the initial visit, the clinician's openness may empower the patient to raise concerns in future visits. The core CFI adopts an ecologic approach to culture by asking the patient about the illness, moving outward to views of the social network and the role of social supports and stressors, and then finishing with the illness in relation to the health care system.

The Informant CFI

The informant CFI (I-CFI) guides clinicians in conducting a cultural assessment of the patient's clinical problem from the perspective of a key informant such as a spouse, relative, or friend who may be present during the clinical encounter. Patients are often accompanied by relatives or other potential informants during clinic visits, providing clinicians with the opportunity to include them during diagnosis and treatment planning. Social relationships can have profound effects—positive and negative—on pathways to and through mental health treatment, including decisions about when or whether to seek care, participation in clinic visits, and implementation of treatment plans in the home, such as adherence to medication. The roles of close associates and their perspectives on illness and treatment are informed by cultural meanings and practices.

Clinical care often requires consultation with family members, other kin, or friends. Certain aspects of culture may be shared among patients and others within the social network, but there may be differences within families. For example, an older immigrant man who is depressed and most comfortable speaking his native language may be accompanied by an adult child who speaks English and has different cultural values. In such cases, the patient and the informant may have different EMs for the patient's depression or view of treatment. In such situations, the I-CFI can assist clinicians with eliciting the perspectives of key informants through the lens of culture, enabling a deeper understanding of the patient's illness and the interpersonal dimensions of care. This can create opportunities for clinical negotiation and the development of shared understandings of illness and goals for treatment.

The I-CFI can play an especially important role in several situations when an informant is available. In some instances, the informant may be the primary or sole source of information to assess the role of cultural factors in diagnosis and treatment. This may occur when a patient is unable to give a meaningful or coherent history because of conditions such as severe cognitive impairment (i.e., from Alzheimer disease or a head injury), catatonia, severe paranoia, agitation, or substance intoxication. Alternatively, the patient may be a young child, and a cultural assessment of the parent or other caregiver's perspective may be helpful. In other cases, the clinician may seek clarification from the informant when the patient's responses to the CFI are ambiguous or partial (as when asking the patient about how others in their cultural group would view the symptoms). The clinician may also administer the I-CFI if there is a family member or friend who is influential in the patient's clinical care and decision making. Finally, the I-CFI may be useful when treatment hits an impasse, and an assessment of the social context may identify underlying causes.

In choosing whether to use the I-CFI, the clinician should consider several issues. First, the clinician should discuss the involvement of the informant with the patient while respecting all desires for autonomy and privacy. Patients may not want a relative involved due to concerns about burdening family members, shame about certain aspects of the problem, or discomfort in discussing conflicts with the informant. In these situations, the clinician may decide not to involve the informant or opt to negotiate family involvement over time after the patient and clinician have established rapport. In other situations, as when the patient insists on having relatives present during the clinical encounter and defers to them, the clinician may need to negotiate time during the visit to meet individually with the patient to allow discussion of sensitive issues that may involve family members (i.e., cases of conflict or abuse). After the informant interview, the clinician may decide whether to interview the relative (or another informant) in the presence of the patient and vice versa. Clinicians should balance patient preferences with medicolegal requirements related to patient confidentiality that are operational in their practice settings.

A related issue is that the clinician may find important differences in the views of patients and informants (i.e., in their illness EMs or the extent to which behavior is viewed as normative in their cultural context). One approach to such situations is to present the different perspectives in a nonjudgmental way, together with the biomedical view of illness, to try and negotiate a partially shared understanding of illness and treatment. When multiple perspectives cannot be resolved in this way, documenting differences may assist the clinician in seeking additional information to clarify diagnostic issues or in anticipating obstacles to implementing the treatment plan.

The clinician has flexibility in the timing and administration of the I-CFI. For example, the I-CFI could be used at the initial visit or later. The clinician may use all of the I-CFI or only the parts most apparently relevant to clinical care. With multiple family members or caregivers, pragmatic decisions must be made about whom to interview based on patient preferences and the informant’s involvement in the patient’s life. Finally, the supplementary modules may be useful with the I-CFI, particularly the caregiver module, which further explores the role of cultural factors in caregiving. The I-CFI can be accessed in *DSM-5-TR* and at the website noted earlier.

The CFI Supplementary Modules

The final set of CFI assessments is the supplementary modules, listed in Table 29.2–3. The core CFI operationalizes a basic cultural assessment, but the need for supplementary modules may arise when certain cultural domains are not addressed or assessed without requisite depth during the initial clinical encounter. In other circumstances, the core CFI may not be needed as much as a specific supplementary module. Supplementary modules that are based on special populations such as children and adolescents, immigrants, and older adults include questions tailored for their cultural needs. Alternatively, a clinician can administer the EM module to learn more about the patient's understanding of the problem or the spirituality, religion, and moral traditions module to better grasp the meanings of suffering. These examples suggest how 12 supplementary modules can be utilized with the core CFI and the informant version or as part of a stand-alone interview.

All supplementary modules possess a common format. They begin by noting common questions with the core CFI. Instructions for administering the module are provided, followed by an introduction for the clinician to the patient. Most modules are organized by a specific subdomain, further assessed through several questions. For example, in the cultural identity module, the subdomains include language ability and migration history, assessed by 6 and 9 questions, respectively. Like the core CFI, the supplementary modules are mostly problem focused, attempting to elicit how the module topic relates to the presenting problem. Some, like the cultural identity module, focus more on aspects of the patient's identity than the presenting problem. Other modules are focused on the presenting problem, such as the EM and coping and help-seeking modules. In some cases, questions in one module are assessed in another based on topical interrelationships. Investigators continue to develop new supplementary modules, such as for the cultural assessment of grief and bereavement.

Table 29.2–3.
The CFI Supplementary Modules

Number	Module
1	Explanatory model
2	Level of functioning
3	Social network
4	Psychosocial stressors
5	Spirituality, religion, and moral traditions
6	Cultural identity
7	Coping and help seeking
8	Patient–clinician relationship
9	School-aged children and adolescents
10	Older adults
11	Immigrants and refugees
12	Caregivers

Research on the CFI Since *DSM-5's* Publication

By 2020, a substantial evidence base emerged on the CFI. One systematic review article identified 25 studies on the CFI, with sample sizes ranging from 1 to 460 participants. Studies were divided into two types: those that used the pilot CFI from the *DSM-5* field trial, which collectively enrolled 960 subjects, and those that used the final version of the CFI in *DSM-5,* which enrolled 739 subjects. The studies covered three broad themes: how the CFI (1) affects a clinician's practice; (2) relates to existing theories in medical communication, implementation science, and family psychiatry; and (3) is used within cultural competence training programs. Most studies reported that patients and clinicians viewed the CFI as improving clinical rapport, though no study has examined longer-term clinical outcomes such as improvements in symptoms of functioning.

FUTURE DIRECTIONS

This chapter has presented a historical overview of patient cultural assessments, from Kraepelin’s initial attempts at cross-cultural comparative psychiatry over a century ago to the CFI in *DSM-5-TR*. Common themes in patient cultural assessments since the 1970s have included a focus on the patient’s cultural background and identity, explanations of illness, levels of psychosocial support and functioning, and experiences with help seeking, including with the biomedical system. The CFI represents the most international attempt at standardizing a pertinent set of questions for patients and clinicians in all clinical encounters with the assumption that culture constantly frames understandings and practices related to sickness and care.

Members of the DCCIS have continued to expand their work on cultural assessments beyond *DSM-5-TR*. The CFI presents a brief list of questions designed to elicit clinically relevant cultural information within practice constraints. Research is currently underway on the comparative effectiveness of the CFI with other forms of resource-intense cultural assessments such as cultural consultation services in changing diagnosis and recommended treatment modalities. The CFI also reflects an attempt to mobilize contemporary notions of culture from anthropology and the interpretive social sciences, but further work is needed to clarify if these notions apply in all clinical settings and how best to elicit this information through simple and succinct probes. Ongoing partnership with collaborators from anthropology, psychology, and sociology may illustrate new domains necessary for thorough cultural assessments that can be incorporated in future revisions. For example, ongoing work focuses on the best ways to assess the impact on a patient’s illness experience and access to care of social determinants of health, including societal–structural disparities by ethnoracial background and social position. Studies on CFI implementation in educational and clinical settings are needed to understand how trainees learn the CFI, which questions obtain the most pertinent cultural information applicable to diagnostic and treatment planning, and how patients perceive differences with nonculturally informed care. Outcome studies measuring the CFI's effects on patient symptoms, quality of life, and functional assessment are needed to examine potential mechanisms of action. Finally, research must continue to identify the most successful strategies for implementing the CFI during routine practice, including in team-based care.

FURTHER READINGS

Aggarwal NK. Hybridity and intersubjectivity in the clinical encounter: impact on the cultural formulation. *Transcult Psychiatry*. 2012;49:121.

Aggarwal NK, Jarvis GE, Gómez-Carrillo A, Kirmayer LJ, Lewis-Fernández R. The cultural formulation interview since DSM-5: prospects for training, research, and clinical practice. *Transcult Psychiatry*. 2020;57:496.

Aggarwal NK, Nicasio AV, Desilva R, Boiler M, Lewis-Fernández R. Does the Cultural Formulation Interview (CFI) for the revision of the DSM-5 affect medical communication?: a qualitative exploratory study from the New York site. *Ethn Health*. 2015;20:1.

Aggarwal NK, Pieh MC, Dixon L, Guarnaccia P, Alegria M, Lewis-Fernández R. Clinician descriptions of communication strategies to improve treatment engagement by racial/ethnic minorities in mental health services: a systematic review. *Patient Educ Couns*. 2016;99:198.

Alarcón RD. Culture, cultural factors and psychiatric diagnosis: review and projections. *World Psychiatry*. 2009;8:131.

Eisenberg L. Disease and illness. Distinctions between professional and popular ideas of sickness. *Cult Med Psychiatry*. 1977;1:9.

Good BJ. *Medicine, Rationality and Experience: An Anthropological Perspective*. Cambridge University Press; 1994.

Groleau D, Young A, Kirmayer LJ. The McGill Illness Narrative Interview (MINI): an interview schedule to elicit meanings and modes of reasoning related to illness experience. *Transcult Psychiatry*. 2006;43:671.

Jarvis GE, Kirmayer LJ, Gómez-Carrillo A, Aggarwal NK, Lewis-Fernández R. Update on the cultural formulation interview. *Focus (Am Psychiatr Publ)*. 2020;18(1):40–46.

Kirmayer L. Cultural psychiatry in historical perspective. In: Bhugra D, Bhui K, eds. *Textbook of Cultural Psychiatry*. 2nd ed. Cambridge University Press; 2018:1–17.

Kirmayer LJ, Guzder J, Rousseau C. *Cultural Consultation: Encountering the Other in Mental Health Care*. Springer; 2014.

Kleinman AM. Depression, somatization and the “new cross-cultural psychiatry.” *Soc Sci Med*. 1977;11:3.

Kleinman A. *Patients and Healers in the Context of Culture: An Exploration of the Borderland Between Anthropology, Medicine, and Psychiatry*. University of California Press; 1980.

Kleinman A, Benson P. Anthropology in the clinic: the problem of cultural competency and how to fix it. *PLoS Med*. 2006;3:e294.

Lewis-Fernández R, Aggarwal NK. Culture and psychiatric diagnosis. *Adv Psychosom Med*. 2013;33:15.

Lewis-Fernández R, Aggarwal NK, Bäärnhielm S, et al. Culture and psychiatric evaluation: operationalizing cultural formulation for DSM-5. *Psychiatry*. 2014;77:130.

Lewis-Fernández R, Aggarwal NK, Hinton L, Hinton DE, Kirmayer LJ. *DSM-5 Handbook on the Cultural Formulation Interview*. American Psychiatric Publishing, Inc.; 2015.

Lewis-Fernández R, Díaz N. The cultural formulation: a method for assessing cultural factors affecting the clinical encounter. *Psychiatr Q*. 2002;73:271.

Lewis-Fernández R, Aggarwal NK, Lam PC, et al. Feasibility, acceptability, and clinical utility of the Cultural Formulation Interview: mixed-methods results from the DSM-5 international field trial. *Br J Psychiatry*. 2017;210:290.

Littlewood R, Dein S. *Cultural Psychiatry and Medical Anthropology: An Introduction and Reader*. Athlone Press; 2000.

Lloyd KR, Jacob KS, Patel V, ST Louis L, Bhugra D, Mann AH. The development of the Short Explanatory Model Interview (SEMI) and its use among primary-care attenders with common mental disorders. *Psychol Med*. 1998;28:1231.

Mezzich JE, Caracci G. *Cultural Formulation: A Reader for Psychiatric Diagnosis*. Jason Aronson; 2008.

Mezzich JE, Caracci G, Fabrega H Jr, Kirmayer LJ. Cultural formulation guidelines. *Transcult Psychiatry*. 2009;46:383.

Paralikar VP, Deshmukh A, Weiss MG. Qualitative analysis of Cultural Formulation Interview: findings and implications for revising the outline for cultural formulation. *Transcult Psychiatry*. 2020;57:525.

Paralikar VP, Sarmukaddam SB, Patil KV, Nulkar AD, Weiss MG. Clinical value of the cultural formulation interview in Pune, India. *Indian J Psychiatry*. 2015;57:59.

Weiss MG, Doongaji DR, Siddhartha S, et al. Explanatory Model Interview Catalogue (EMIC): contribution to cross-cultural research methods from a study of leprosy and mental health. *Br J Psychiatry*. 1992;160:819.

Weiss MG, Aggarwal NK, Gómez-Carrillo A, et al. Culture and social structure in comprehensive case formulation. *J Nerv Ment Dis*. 2021;209(7):465–466.

Young A. When rational men fall sick: an inquiry into some assumptions made by medical anthropologists. *Cult Med Psychiatry*. 1981;5:317.

30

Special Areas of Interest

30.1 Psychiatry and Reproductive Medicine

HELEN BRADSHAW, M.D., MARIKA TOSCANO, M.D., AND LAUREN M. OSBORNE, M.D.

Reproductive events and processes have both physiologic and psychological consequences—and psychological states affect reproductive physiology and modulate reproductive events. This section examines these bidirectional relationships, intending to introduce fundamental concepts related to the classic reproductive events, such as menarche, the menstrual cycle, pregnancy, delivery, postpartum, and the menopausal transition. Both psychiatry and reproductive medicine have defined distinct diagnoses that include the functioning of both systems. For instance, premenstrual dysphoric disorder (PMDD)—mood, cognitive, and behavioral changes that can occur in association with the menstrual cycle—exemplifies a somatopsychic disorder, in which biologic changes occurring in the soma trigger changes in psychological state. In contrast, functional forms of hypothalamic anovulation represent psychosomatic illness that originates in the brain but alters somatic functioning. Medicine has traditionally separated the treatment of reproductive events from that of psychological functioning. However, increasing awareness of a false dichotomy between mind and body is beginning to shape new systems and approaches to care that place greater attention on addressing mental health needs in medical settings. The field of reproductive psychiatry has blossomed in recent years, with distinct residency training tracks and postdoctoral fellowship training for both psychiatrists and obstetricians—making this a crucial and distinctive subspecialty field of psychiatry.

Since its inception, most of this field's research, policy, and clinical activities have applied exclusively to women. Men's reproductive health, and the perinatal mental health of fathers, have only recently been recognized as worthy of study and attention. Similarly, the unique reproductive mental health needs of people who have female reproductive organs but do not identify as women are now being recognized—but evidence to support these unique needs at this time is limited. Therefore, we need to note that we speak to the interplay of reproductive and psychiatric systems on the health of all people with ovaries. Not all individuals that possess ovaries identify as women, and not all women have ovaries. An individual's genetic/chromosomal sex, either assigned female at birth (AFAB) or assigned male at birth (AMAB), does not necessarily correspond with that individual's experienced gender throughout life. Here, as we open the chapter, we remind the reader of this distinction between genetic/chromosomal sex and expressed gender. Emerging research may in the future allow us to make distinctions about the reproductive mental health of people of different genders, but as most research currently has been done on women, we will continue to use that term in this edition of this textbook.

REPRODUCTIVE DEVELOPMENT

Fetal and Neonatal Development

A fetus's genetic/chromosomal sex is defined by sex chromosome composition (XX vs. XY), which subsequently directs gonadal differentiation into ovaries or testes. This, in turn, directs the hormonal patterns that determine phenotypic sex, defined as the appearance of external genitalia and secondary sex characteristics that arise during puberty. The genetic/chromosomal sex of an embryo is determined during the fertilization process, with an oocyte (egg) contributing an X chromosome and a sperm contributing either an X or a Y chromosome. Subsequent gonadal differentiation does not occur until approximately 6 to 8 weeks of gestation, and prior to this developmental phase, the embryo contains both male and female undifferentiated gonads (both the Wolffian duct and Müllerian duct systems). Phenotypic sex is bidirectional until about 8 to 10 weeks of gestation, with undifferentiated external genitalia consisting of the genital tubercle and genital folds and swellings. The default developmental pathway occurs in the absence of a Y chromosome and testosterone when the Müllerian duct system develops into fallopian tubes, the uterus, and the upper vagina; the genital tubercle becomes the clitoris, and the genital folds and swellings become the labia. In the presence of a Y chromosome containing the SRY gene region, which codes for testis-determining factor, the undifferentiated gonads are converted to testes. The fetal testes begin to secrete testosterone and anti-Müllerian hormone (AMH), leading to regression of the Müllerian duct system and development of the Wolffian duct system to produce male internal genitalia. Testosterone's enzymatic conversion to dihydrotestosterone (DHT) by 5-alpha–reductase results in masculinization of the external fetal genitalia. The urogenital tubercle becomes the glans, the urogenital folds become the penis and male urethra, and the labioscrotal swellings in the midline become the scrotum. Male differentiation occurs between 8 and 10 weeks of gestation, with external genitalia recognizable by 14 to 16 weeks. If testosterone synthesis is absent or compromised, if enzymatic conversion of testosterone to DHT does not occur, or if the fetal tissues contain androgen receptor defects, rendering them insensitive, a spectrum of incomplete masculinization ensues, including microphallus, hypospadias, cryptorchidism, or partial vagina. In female fetuses, excess androgen exposure from 8 to 12 weeks of development from maternal or fetal sources, such as occurs in fetal congenital adrenal hyperplasia or maternal virilizing tumors, can partially masculinize a female fetus, leading to findings such as clitoromegaly or labial fusion.

Sexual differentiation of the central nervous system (CNS) is mainly dependent on the effect of testosterone on target areas within the developing brain. In response to placental human chorionic gonadotropin (hCG), the fetal testes secrete testosterone in the mid-first trimester. Fetal testicular function peaks between 15 and 18 weeks, then declines, with the maintenance of in-utero testosterone production assumed by the fetal pituitary gonadotropins. In contrast, the fetal ovary does not develop functional follicles until the third trimester. The estrogens produced by the fetal gonads and placenta do not cross the blood–brain barrier because of circulating binding proteins, and thus circulating estrogens do not have access to and cannot imprint the developing CNS in utero. In late pregnancy, gonadotropin secretion is restrained in male and female fetuses due to the negative feedback effects of increasing placental estrogen and progesterone production. At the time of birth, when maternal estrogen and hCG circulating in the neonate decrease sharply compared to fetal life, the newborn's hypothalamic–pituitary–gonadal axis escapes from this suppressive effect and gonadotropin secretion rises dramatically in both sexes.

Similarly, follicle-stimulating hormone (FSH) rises to a greater extent in girls, and luteinizing hormone (LH) rises to a greater extent in boys. In boys, the gonadotropin rise is followed by an elevation of testosterone concentrations to adult levels, with a peak at 3 to 6 months postnatally, followed by a steady decline to prepubertal levels by 9 to 12 months of age. Girls have a more marked and sustained postnatal rise in gonadotropins than boys, with FSH reaching levels greater than during a normal adult menstrual cycle and LH levels equivalent to adult levels. In girls, this gonadotropin rise peaks at 12 to 18 months and falls to prepubertal levels by 24 to 36 months. Thus, between birth and 3 years of age, the brains and bodies of girls and boys have been exposed to dramatically different patterns, durations, and doses of sex-steroid secretion.

The organizational effects of testosterone on the developing CNS are thought to depend primarily on in situ aromatization (the conversion of androgens to estrogens by the enzyme aromatase) of testosterone to estradiol. In addition, testosterone may bind directly (without conversion to estradiol) to androgen receptors in the CNS. Therefore, sexual dimorphism of brain structure may reflect androgen exposure and central dimorphism in crucial brain areas that are responding to high aromatase activity (which, thus, can convert testosterone to estradiol) and the presence of androgen receptors. In the nonhuman primate brain, androgen receptors are in the hypothalamus, amygdala, and prefrontal visual and somatosensory cortices, whereas estrogen receptors (ERs) are more highly concentrated in the female hippocampus than in the male. The degree to which gender-related behavioral asymmetries are accounted for by differences in fetal/infant-steroid hormone exposure is open to debate, but clearly, a mechanism for inducing differences exists.

Morphometric studies of the developing and adult brain support the sex-steroid theory of CNS sexual dimorphism. Sex differences cluster in brain regions with developmentally rich sex–steroid-receptor distribution. Sexual dimorphism is prominent in the cortex, which may result from rapid postnatal decreases in sex–steroid-receptor concentrations. By the age of 6 years, brain volume is 95% of that of an adult; however, boys and girls follow different trajectories in brain tissue changes between the age of 6 years and adulthood. By adulthood, men have 10% larger brain volumes relative to women, larger corpus callosum genu and overall white matter volume, and increased ventricular size. By adulthood, women have larger cortical volume relative to men, higher cortical gray matter volume, and larger volumes in Broca’s area, the superior temporal cortex, hippocampus, caudate, thalamus, anterior cingulate gyrus, dorsolateral prefrontal cortex, and interhemispheric white matter connections.

Ongoing and future research will clarify how sex-chromosome genes, independent of gonadal hormones, contribute to the sexual differentiation of brain cells. This concept has already been substantiated from research on embryonic tissues studied before gonadal development and through animals engineered to be discordant for genotypic and phenotypic sex. For example, sex-chromosome determination of the midbrain dopamine system has been described. Magnetic resonance imaging (MRI) studies have further clarified the contributions of sex chromosomes and gonadal hormones to the sexual differentiation of the brain. More recent research has identified early sex differences in microglia, the immune system of the CNS, as a possible explanation for discrepancies in the incidence of neurologic diseases and response to treatment across the sexes. Many questions remain that will inform future directions in research on the sexual differentiation of the brain, such as the permanence versus transience of the effects and the role of animal models in understanding environmental factors.

Puberty

Gonadotropin-releasing hormone (GnRH) neurons in the hypothalamus are endogenously pulsatile and active during fetal and neonatal life, making the hypothalamic–pituitary–gonadal axis fully functional before birth. However, between infancy and puberty, the hypothalamic–pituitary–gonadal axis enters a dormant period, with a mechanism proposed to be related to the effects of the inhibitory neurotransmitter gamma-aminobutyric acid (GABA). In humans, this postnatal gonadal hiatus involves desynchronizing and suppressing the hypothalamic GnRH pulse generator by undetermined central mechanisms that are largely gonad independent. Although pulsatile GnRH activity remains during this time of dormancy, the frequency and amplitude are low and irregular. Puberty—the resumption of gonadal activity—is also a centrally initiated process that depends on synchronizing the pulsatile release of hypothalamic GnRH. The character of gonadotropin secretion pulses changes around puberty, with a 10-fold increase in LH levels compared to a 2-fold increase in FSH levels and, to a lesser extent, an increase in pulse frequency. LH levels further increase to 20 to 40 times greater than prepubertal levels, particularly during nocturnal pulses. These changes are likely triggered at least in part by expression of the *KiSS1* gene (a gene that encodes a family of neuropeptides called kisspeptins), the excitatory neurotransmitter glutamate, and possible contributions from neuropeptide Y. Although prenatal exposure to sex steroids has organizational effects on the developing brain, the reinitiation of gonadal activity during puberty exerts primarily activational influences on the brain and behavior The stimulus for the onset of puberty remains unknown. However, genetics, environment, social factors, and overall health all influence the transition.

Gender

The phenotypic sex of an individual depends on chromosomal sex and exposure and responsiveness of target tissues to endogenous and exogenous sex steroids. *Gender* differs from biologic sex in that it refers to a person's self-image and sex-role identity. Although the development of gender identity is not entirely understood, it is likely influenced not only by exposure and responsiveness to sex steroids but also by expectations and behavioral patterns learned in early childhood from parental, familial, and societal models. Notions of gender direct mannerisms, marital, career, and other behavioral choices that may depart from biologic options. Consider the following example: a child born with a penis and testes identifies as female from a young age and undergoes therapy to prevent masculinization at the time of puberty, followed by feminization with hormones, and later pursues breast construction and vaginoplasty as an adult, with life-long persistence of her female identity. Consider a second example: a female-identifying individual with *Müllerian agenesis*, a condition with ovaries present but uterus and vagina absent, pursues a heterosexual relationship and motherhood via surgical and technologic advances, despite the biologic constraints of her disorder.

While the two examples in the paragraph above offer distinct situations with few similarities, the first describing the experience of a male-to-female transgender individual and the second an example of a person with a disorder of sexual development (DSD), much of our understanding of the development of gender identity is drawn from the literature on DSD. The understanding of gender as two separate categories, either male or female, often referred to as the gender binary, has predominated throughout history and has heavily influenced concepts of mental experience. However, recent literature and social activism have challenged the accuracy of understanding individuals as either male or female based upon phenotypic sex. While transgender individuals identify with their nonphenotypic sex, many identify outside of the gender binary and describe themselves as queer, agender, gender fluid, demi- or pan-gender. Our understanding of the biologic and environmental factors that contribute to an individual's development of gender identity remains limited. Available research suggests that the development of gender identity is not simply a psychological process and implicates an array of biologic mechanisms, including genetics, neuroanatomical gender differences, and prenatal and postnatal androgen exposure, in the development of gender identity.

Early Sexuality

Children's behavior can be understood by acknowledging the sexual motivations that, until Sigmund Freud's time, had been assumed to be quiescent until adolescence. The fact that incest, sexual abuse, and other sexual taboos that may occur in childhood have powerful aftereffects, even in adult life, bespeaks the importance of childhood sexuality in establishing healthy relationships.

In puberty, the overriding influence on sexuality is the dramatic change in the hypothalamic–pituitary–gonadal axis that occurs in both sexes. Major hormonal changes, combined with the issue of identity diffusion, contribute to the characteristic tension in relationships, particularly between male and female peers and between adolescents and their parents.

In adulthood, one of the major maturity indicators is establishing healthy sexual relationships. When it is in harmony with other components of the self and with others, sexuality is an integral component of the adult character.

Adolescence

Carol Gilligan observed that preadolescent girls exhibit a sense of self-confidence and a belief that they can influence the course of events in their lives, whereas, during adolescence, that self-esteem and belief in their efficacy appear to diminish. Before puberty, boys are more vulnerable than girls are to depressive illness. Girls display a precipitous rise in depression rates throughout adolescence that outpaces adolescent boys. Compared to boys, by age 15 years, girls are twice as likely to have experienced a lifetime episode of major depression. This gender difference has been linked to the onset of puberty rather than to chronologic age and persists for the next 35 to 40 years, spanning the female reproductive years. The predominance of depression in women and the adolescent-onset of this gender difference have been replicated cross-nationally and represent one of the most robust findings in the field of psychiatric epidemiology. In addition, LGBTQ+ youth have higher rates of depression than their non-LGBTQ+ peers.

Negative life events and chronic psychosocial stressors place individuals at risk for major depressive episodes. Several theories about the sex difference in depression point out that women are more likely than men to experience certain traumatic life events. These include physical or sexual abuse and chronic life stressors, such as single parenthood, gender discrimination in the workplace, or financial adversity. Growing evidence indicates that puberty may sensitize girls to the depressogenic effects of life stress. Because women tend to develop a strong affiliative style in their social relationships and a sense of self in connection with others, social stressors, such as conflicts, breaches, or losses within interpersonal relationships, contribute to their depression risk. Men, in contrast, tend to focus on issues of individuation and assertiveness within social relationships and develop an agenic and autonomous sense of self.

Early menarche is associated with an increased risk of depressive symptoms. However, additional research is needed to clarify the causality and biologic mechanisms responsible for this relationship. Results of the 2005 to 2016 National Health and Nutrition Examination Survey (NHANES) demonstrated that women with early menarche had 1.25 (95% CI of 1.16 to 1.61) times the odds of depressive symptoms at the time of the survey compared to women who experienced normal or late menarche. More than 15,000 reproductive-age women completed the survey, and comparisons were made after controlling for the likely confounders of age, race, educational attainment, income, and marital status. While the association between early menarche and depressive symptoms has been well-replicated, few studies have addressed the causality of this relationship. A longitudinal study assessing the relationship between a genetic risk score for early menarche and depressive symptoms at four time points during adolescence demonstrated a causal relationship until age 14. Hypothesis surrounding the possible mechanisms linking early menarche to depressive symptoms include dysregulation of the HPA axis and an incongruity between biologic and cognitive maturation.

Recent theories have implicated the hypothalamic neurohormone oxytocin in facilitating such affiliative behaviors as maternal caregiving and adult pair-bond formation in female mammals. Shelley Taylor and coworkers proposed the existence of a uniquely female response to stress, which they termed the *tend-and-befriend response*, based on the oxytocin-mediated caregiving system, which, for women, serves to decrease sympathetic and hypothalamic–pituitary–adrenal (HPA) responses to stress while simultaneously promoting affiliative responses. This stands in contrast to the prototypic model fight-or-flight human stress response. Instead of fighting or fleeing, quieting and tending to offspring’s needs when under stress and affiliating with other adults to promote the protection of oneself and one’s offspring may represent a more adaptive female response to environmental threats.

Although gender role socialization may begin well before puberty, social pressures to conform to stereotypically feminine versus masculine gender roles likely intensify during the adolescent transition, as the development of secondary sex characteristics make adolescents look more like adults, and they begin to engage in greater levels of cross-sex interaction and dating. Of importance, hormonal changes in the reproductive neuroendocrine axis at puberty may potentiate and interact with this socialization process to promote an adolescent intensification of gender-based behaviors. Some have speculated that the increase in testosterone in boys and the attendant aggressiveness, often directed at girls, contribute to a change in girls' behavior during adolescence.

MENSTRUAL CYCLES

Menstrual Physiology

Menstrual cyclicity results directly from ovarian cyclicity (Fig. 30.1–1). Each ovarian cycle starts with recruiting a group of follicles in response to FSH, with subsequent follicular development and maturation, and finally "selection" of one, which becomes the dominant follicle (also called a mature oocyte) that proceeds toward ovulation. As the follicle follows this progression, hormonally active theca and granulosa cells proliferate and synthesize steroid hormones, estrogen more than androgen or progesterone.

Follicular development is initiated by the hypothalamic release of GnRH at a pulse frequency of approximately one pulse every 90 minutes. GnRH stimulates the release of the pituitary gonadotropins LH and FSH. In turn, LH stimulates ovarian theca cells to synthesize and secrete androgens; FSH induces granulosa cell development, including the enzyme aromatase, which converts the thecally produced androgens to estrogens. In the presence of a constant GnRH pulse frequency of one pulse every 90 minutes, the secretion of LH and FSH in the follicular phase is regulated primarily by estradiol feedback at the pituitary level. Rising estradiol concentrations suppress FSH, limiting the number of follicles that become mature oocytes capable of ovulating. When estradiol concentrations rise exponentially to exceed a critical threshold and remain elevated for at least 36 hours, which is the pattern one fully mature follicle produces, an LH surge is triggered, and ovulation (release of the ovum from the follicle sac) ensues approximately 36 hours later. Thereafter, granulosa cells transform into progesterone-secreting luteal cells, and the ovulated follicle is then referred to as the *corpus luteum,* which secretes progesterone.

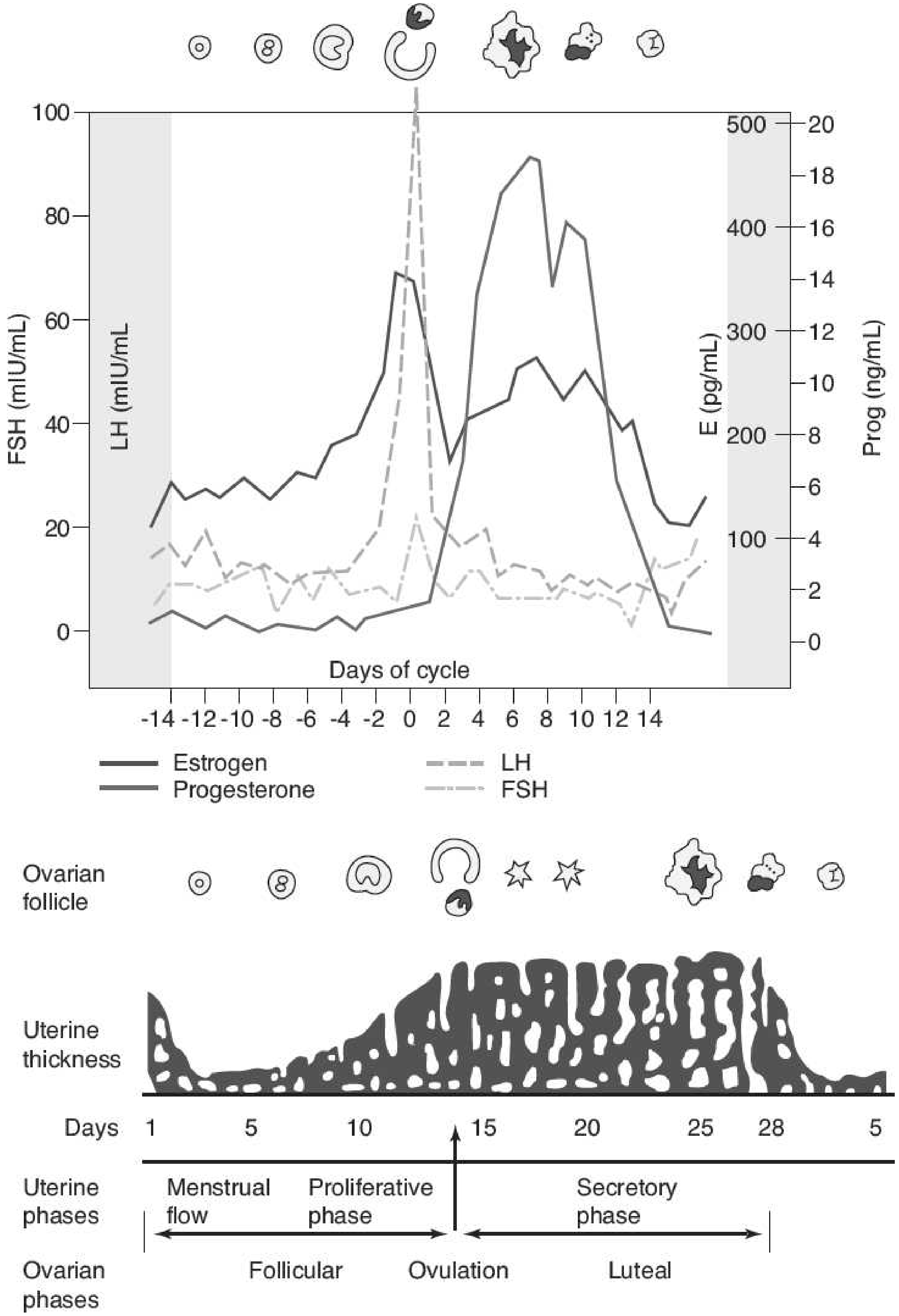
FSH (mIU/mL)
LH (mIU/mL
E (pg/mL)
Prog (ng/mL)
100
80
60
40
20
0
500
400
300
200
100
20
18
16
14
12
10
8
6
4
2
0
Days of cycle
-14 -12 -10 -8 -6 -4 -2 0 2 4 6 8 10 12 14
Estrogen
Progesterone
LH
FSH
Ovarian follicle
Uterine thickness
Days
1
5
10
15
20
25
28
5
Uterine phases
Menstrual flow
Proliferative phase
Secretory phase
Ovarian phases
Follicular
Ovulation
Luteal

FIGURE 30.1–1. Changes during reproductive and menstrual cycles **(Top)**. Chart illustrating plasma hormone concentrations in the normal female reproductive cycle **(Bottom)**. FSH, follicle-stimulating hormone; LH, luteinizing hormone. (See eBook for color images.) (Illustration depicting the ovarian events and uterine changes during the menstrual cycle. LifeART image copyright © 2024 Lippincott Williams & Wilkins. All rights reserved.)

The target tissues for ovarian steroids include the endometrium and the hypothalamic GnRH pulse generator, whose frequency is slowed dramatically by the combination of estrogen and progesterone secreted during the postovulatory or luteal phase of the menstrual cycle. This inhibition of GnRH is followed by decreased secretion of LH and FSH so that new follicular development is prevented until the corpus luteum regresses. As progesterone concentrations decline, GnRH pulsatility increases, and secretion of gonadotropin, especially FSH, rises. The menstrual cycle phases can be termed *follicular* and *luteal* in reference to ovarian events or *proliferative* and *secretory* in reference to endometrial events.

The typical menstrual cycle is 21 to 35 days in adults but may range from 21 to 45 days due to differences in luteal phase length. It is generally assumed that regular menstrual cyclicity signals ovarian cyclicity and ovulation, but that assumption is erroneous. Menses and menstrual cycle interval cannot be used to determine whether a woman is or is not ovulatory or if the luteal phase was normal. The clinician who links a psychological state to a reproductive event must know how to determine the presence or absence of gonadal activity. Ovarian activity can be estimated by using several different methods. At-home urine test kits to measure LH levels (over-the-counter "ovulation predictor kits" at drug stores) can demonstrate LH surge and presence of ovulation, which usually occurs on postmenstrual days 13 to 15 for women with 28-day cycles. Urine testing with these kits should be timed to begin on postmenstrual day 11 and continued until ovulation is detected.

Similarly, blood tests may be used to detect ovulation. A blood test to measure serum progesterone can be performed about one week before the anticipated onset of the next menstrual period, and an elevated level is usually associated with ovulation. Estrogen and LH levels may also be measured in blood just prior to ovulation. Basal body temperature monitoring with an increase of 0.5 degrees Fahrenheit for several days can indicate that ovulation has occurred. Similar changes in cervical mucous (increased, thin, slippery, and stretchy) can provide at-home clues to determine the presence of ovulation. Women may also be queried regarding the presence or absence of molimina, a group of characteristic symptoms which occur after ovulation, including breast discomfort, increased mucoid vaginal discharge, or premenstrual cramping and bloating. Menstrual cycles can be expected to become more regular with advancing gynecologic age (time since first menses), and, in the absence of pathology, they generally remain regular until 5 to 10 years before the menopausal transition. Ovarian function is characterized by higher estradiol and lower progesterone secretion during the perimenopausal years. Ovarian function does not gradually decline but rather becomes erratic and unpredictable, potentially exposing the brain and soma to large fluctuations in hormone concentrations.

Ovulatory Dysfunction

Ovulatory dysfunction (i.e., oligo-ovulation, anovulation) is identified in 3.4% to 18.6% of menstruating women. Oligo- or anovulation refers to the failure of the ovary to release an ovum during the typical menstrual cycle period of 21 to 45 days. This can sometimes be confused with oligomenorrhea (bleeding that occurs less frequently than every 35 days) or amenorrhea (absence of bleeding over three typical cycles), which are menstrual cycle changes that fall under the term “abnormal uterine bleeding” (previously termed dysfunctional uterine bleeding). While ovulatory dysfunction can sometimes be the etiology of oligo- or amenorrhea, some women with ovulatory dysfunction demonstrate normal menstrual cycle bleeding patterns despite a lack of ovulation. Additionally, many women with oligo- or amenorrhea will have alternative etiologies causing changes in their bleeding patterns unrelated to ovulation.

The physiologic causes of ovulatory dysfunction include adolescence (due to immaturity of the HPA axis), perimenopause, lactation, and pregnancy. Pathologic ovulatory dysfunction can arise from hyperandrogenic states (such as polycystic ovarian syndrome, congenital adrenal hyperplasia, or androgen-secreting tumors), endocrinopathies (hyperprolactinemia, thyroid disease), premature ovarian failure, iatrogenic causes (radiation, chemotherapy, or medications including antipsychotics and antidepressants), health conditions (obesity, diabetes), eating disorders/excessive exercise, or hypothalamic dysfunction, which is discussed further below.

Functional Hypothalamic Amenorrhea

Functional hypothalamic amenorrhea (i.e., hypogonadotropic hypogonadism), defined as menstrual cycle interval persistently exceeding 45 days and/or amenorrhea for 3 months or more for which there is no organic or pathologic cause, is a consequence of a nonorganic reduction in GnRH secretion that results in reduced pituitary secretion of gonadotropins and subsequent chronic anovulation. Previously referred to as *idiopathic* or *psychogenic amenorrhea* in the psychiatric literature, functional hypothalamic anovulation is a diagnosis of exclusion. The term “functional” indicates that correcting the underlying causal behavioral factors can reverse ovulatory dysfunction. This implies that the ovaries can respond to appropriate gonadotropin input of either endogenous or exogenous origin; that there are no structural abnormalities of the thyroid, adrenal, pituitary, or brain; that the use of drugs, including antipsychotic medications, does not account for the suppression of GnRH; and that the patient is not pregnant. Neuroimaging may be needed to establish no significant anatomical lesions of the brain or pituitary. Two major questions concerning the pathogenesis of functional hypothalamic amenorrhea remain unresolved. First, the peripheral and central signals that disrupt GnRH pulsatility are poorly understood. Second, how do behavioral, cognitive, and personality variables activate the neural systems to disrupt GnRH secretion? Physicians and patients alike wonder what lifestyle variables provoke or contribute to this type of ovulatory dysfunction and what pharmacologic and nonpharmacologic treatment interventions should be considered.

The idea that psychogenic stress can induce reproductive dysfunction in women was formally introduced in 1946. The best biochemical evidence supporting the concept that stress impairs GnRH release in women with functional hypothalamic amenorrhea is the consistent demonstration that the activity of the HPA axis is increased. For instance, a prospective study found that young U.S. women who developed transient amenorrhea while studying in Israel had higher urinary cortisol concentrations on arrival than did those whose menses remained regular. There appears to be a dose–response relationship between the type, severity, and number of stresses on one hand and the proportion of women who develop anovulation on the other. Biological and psychological predispositions may confer resistance or sensitivity to various stressors. Exercise, low weight, weight loss, affective and eating disorders, various personality characteristics such as perfectionism and unrealistic expectations, drug use, and a variety of external and intrapersonal stresses have been linked to the development of anovulation. Most women with functional hypothalamic anovulation, when carefully evaluated, display more than one of these traits or behaviors. Recent evidence suggests synergism between metabolic stressors that suppress the hypothalamic–pituitary–thyroidal axis, such as excessive exercise and nutritional restriction, and psychosocial challenges that activate the HPA axis. Thus, multiple chronic stressors are more likely to induce functional hypothalamic anovulation than is a single, larger-magnitude stressor

While further research is necessary to elucidate the mechanisms involved in the tight regulation of ovulation, eating disorder literature demonstrates that nutritional deficits or significant caloric fluctuations (through either restriction or overexercise) commonly result in decreased LH pulsatility, leading to hypogonadotropic hypogonadism and resulting amenorrhea. Amenorrhea occurs in 66% to 84% of patients with anorexia nervosa. Excessive exercising resulting in the female athlete triad (i.e., relative energy deficiency) can also disrupt the hypothalamic–pituitary–gonadal axis. It is thought that endorphin release, central stress (from elevated cortisol), and energy deprivation in these athletes leads to slowing of the GnRH pulsatility amplitude and frequency, resulting in hypogonadotropic hypogonadism and amenorrhea. Women with functional hypothalamic anovulation are more likely to require fertility treatments than women in the general population. Furthermore, untreated and protracted illness, especially in women with low BMI, increases the risk of miscarriage and preterm labor.

Recovery is possible if response patterns to ongoing psychosocial demands are developed that are less likely to activate the central and metabolic processes that disrupt pulsatile GnRH release. The current standard of practice, other than observation, is to offer pharmacologic interventions, such as oral contraceptives or hormonal replacement if fertility is not desired, and pharmacologic ovulation induction if fertility is sought. However, pharmacologic intervention alone does not lead to spontaneous recovery and cannot be expected to ameliorate stress-induced alterations in central neurotransmission and hypothalamic function or to reverse ongoing metabolic derangements secondary to exercise or weight loss. For instance, bone accretion does not proceed apace in the face of metabolic compromise, even if exogenous hormone replacement is given in supraphysiologic doses. Although women with functional hypothalamic anovulation frequently do not meet DSM-V criteria for eating disorders or depressive disorders, all these states are characterized by increased cortisol secretion, which alone can alter thyroid hormone secretion and action and alter metabolism and attendant central signals. Thus, it is not surprising that women with amenorrhea due to decreased GnRH drive, regardless of cause, have lower bone mineral density.

Furthermore, pharmacologic intervention alone may mask the recognition of psychological dysfunction and forestall the development of more effective response patterns. In addition, ovulation induction may place low-weight women with functional hypothalamic amenorrhea who conceive at risk for premature labor and intrauterine growth retardation. Fetal neuropsychological development may be compromised by the concomitant presence of hypothalamic hypothyroidism (euthyroid sick syndrome) or hypercortisolemia if ovulation is pharmacologically induced. If women’s parenting skills with functional hypothalamic amenorrhea are impaired by ongoing stress, their children may be at risk for poor psychosocial development. Clearly, treatment strategies need to consider that stress and mild psychological dysfunction can play an important role in the genesis of this form of anovulation. If an eating disorder is recognized, specialized psychiatric treatment is indicated.

Psychobiological characteristics can predispose a woman to chronic activation of central neural processes in response to commonplace events. Global hypothalamic dysfunction, including inhibition of GnRH pulsatility, can result. Although identifying the neuromodulators that mediate the development and maintenance of the disruption of GnRH is of academic interest, clinicians must make sure that all other possible causes of anovulation and amenorrhea have been excluded before recommending psychological, psychiatric, or behavioral interventions.

REPRODUCTION

Pregnancy, parturition (childbirth or delivery), and the postpartum period involve a series of tightly orchestrated hormonal events that create the potential for unique psychological states. As pregnancy advances, estrogen and progesterone concentrations rise, and a host of placentally elaborated polypeptide hormones induce metabolic adaptations. Parturition is an intense biologic event that eventuates in the abrupt withdrawal of estrogens, progesterone, and other placental hormones. Parturition is also accompanied by a large-magnitude release of oxytocin and the initiation of the neuroendocrine cascade that establishes lactation. The metabolic demands of pregnancy, parturition, and lactation, the need to deliver infant care across the usual circadian rest-activity cycle, and the hypoestrogenism accompanying lactational amenorrhea present unique physiologic challenges. Menses (and ovulation) generally resume in 6 to 12 weeks if the mother declines to nurse or when lactational frequency declines (usually after 4 to 6 months) if she does nurse. Breastfeeding, and eventual weaning, have implications for maternal mental health, as discussed later in the text in the Breastfeeding and Perinatal Depression sections, respectively.

SEXUALITY AND SEXUAL DYSFUNCTION

It has been said that the most important sexual organ is the brain. This truism emphasizes the essential role of desire and comfort in mediating sexual awareness and expression. Brain centers mediate libido in a general way. Other brain centers influence the selection of sexual partners and the circumstances under which sexual longings are expressed. Hormones made by the ovaries and testes, such as estrogens and androgens, bind in certain brain centers and increase sexual interest. Whether and how this sexual interest gets expressed as behavior depends on several factors, including general health and well-being, the availability of a partner, a sense of what is appropriate, and prioritization of this interest above other interests and obligations. One of the major aims of social and cultural institutions is to channel or direct sexual interest to minimize negative outcomes, such as sexually transmitted diseases (STDs), while ensuring that children will be born and thrive. Sexual interest can be inhibited by any number of negative emotions, including anger, fear, worry, and dislike, and by a marked decline in health or hormones. Like other personality factors, sexual interest, or libido, has a spectrum, with some people having more and others less. As discussed earlier in the Gender section, a person's expression of sexuality and sexual behavior is not necessarily dictated by expressed gender or natal sex. The sexual response cycle and the dysfunctions are covered in a separate section in this textbook on Human Sexuality.

FERTILITY AND INFERTILITY

Gametogenesis

Gametogenesis, the production of oocytes and sperm, occurs because of puberty. Gametogenesis is the minimum physiologic requirement for fertility. In women, the ovary must have oocytes that respond appropriately to gonadotropin input with the maturation of an ovulated oocyte. In men, the testes must manufacture normal sperm capable of penetrating the layer of granulosa cells surrounding the oocyte and then fusing with the outer oocyte cell wall, the zona pellucida. The fallopian tube must be capable of oocyte capture, and the density of sperm in the fallopian tube must be such that there is a juxtaposition of sperm and oocyte. The sperm must negotiate the layers of granulosa cells surrounding the oocyte (cumulus) and fuse with the outer cell membrane of the oocyte (zona pellucida) to release the paternal DNA into the oocyte cytoplasm. Thereafter, the tube must be capable of timely transfer of the fertilized oocyte or zygote to the primed endometrium of the uterus. Implantation must occur, and normal fetal development must ensue for the pregnancy to continue. In the absence of medical intervention, the customary way ovulated oocytes are fertilized is via coitus or sexual intercourse.

Infertility

Infertility is defined as failure to achieve pregnancy within 12 months of unprotected intercourse or therapeutic donor insemination in women younger than 35 years or within 6 months in women older than 35 years. Infertility is estimated to affect 15% of all couples and may arise from female factors, male factors, or a combination of male and female factors, or may be unexplained. Common causes of female factor infertility include diminished ovarian reserve, ovulatory dysfunction, tubal factors (endometriosis, history of pelvic inflammatory disease [PID], or other scarring of the fallopian tubes), uterine factors (Müllerian anomalies, endometrial polyps, or intracavitary fibroids), or abnormalities of cervical mucus. Common causes of male factor infertility include abnormal semen quality or sperm functional parameters; anatomical, endocrine, genetic, functional, or immunologic abnormalities of the male reproductive system (including chronic illness); or sexual conditions incompatible with the ability to deposit the sperm in the female genital tract. Unexplained infertility is a term used when no apparent cause of infertility can be determined after testing excludes the common male and female factors listed above. The basic infertility evaluation can involve a semen analysis, hysterosalpingogram or sonohysterogram, hormonal studies, and tests of ovarian reserve to define possible contributing or causal factors. Ovulation and ovulatory adequacy can be monitored by obtaining a luteal phase serum progesterone level (with repeated levels <3 ng/mL indicative of anovulation). Ovarian reserve testing can involve measurement of serum antimüllerian hormone level (<1 ng/dL indicates diminished reserve), antral follicle count by transvaginal ultrasound (<5 to 7 indicates diminished reserve), or postmenstrual days 2 to 5 estradiol and FSH levels (FSH > 10 IU/L with estradiol >60 to 80 pg/mL indicates diminished reserve). Semen analysis involves quantitative microscopic evaluation of sperm parameters such as semen volume, sperm concentration, morphology, and motility. However, spermatogenesis is difficult to assess because the process takes 70 to 90 days to complete, and the morphologic features of sperm that a standard semen analysis can assess do not always correlate well with function (i.e., ability to fertilize a human oocyte). Diagnostic uncertainty and a seemingly endless array of treatment approaches can easily provoke tension, indecision, and disagreement in the couple with infertility.

The sense that the window of opportunity is narrow can engender panic; a woman's age correlates well with fecundity, and treatment success declines sharply after age 35 years, even when ovulation occurs regularly. Although fertility declines with age in men, the decline is gradual until age 50, and fertility may persist beyond age 60. The offspring of older mothers have an increased risk of chromosomal aneuploidies, particularly trisomy 21. An increase in autosomal dominant conditions has been reported in the offspring of older fathers. Of the identifiable causes of infertility, approximately 20% are thought to be due solely to male factors, 40% to female factors, and 20% to both. Anovulation is the most common cause of infertility in women. Approximately 20% of men with suboptimal semen parameters have microdeletions of the long arm of the Y chromosome in a region thought to play a critical role in spermatogenesis. Approximately 30% of infertile couples have unexplained infertility, which means no identifiable cause.

Table 30.1–1.
Lifestyle Factors That Affect Infertility

Factor	Impact on Fertility	Study
Obesity (BMI > 35)	Time to conception increased 2-fold	Clark et al.
Underweight (BMI < 19)	Time to conception increased 4-fold	Clark et al.
Smoking	RR of infertility increased 60%	Augood et al.
Alcohol (>2 drinks/day)	RR of infertility increased 60%	Eggert et al.
Caffeine (>250 mg/day)	Fecundability decreased 45%	Wilcox et al.
Substance abuse	RR of infertility increased 70%	Mueller et al.
Toxins/solvents	RR of infertility increased 40%	Hruska et al.

RR, relative risk.

Data from Practice Committee of the American Society for Reproductive Medicine and the Practice Committee of the Society for Reproductive Endocrinology and Infertility. Optimizing natural fertility: a committee opinion. *Fertil Steril*. 2022;117(1):53–63. doi:10.1016/j.fertnstert.2021.10.007.

The degree to which psychogenic or lifestyle factors play a role in infertility is controversial. Because lifestyle variables, such as tobacco exposure, alcohol consumption, caffeine exposure, drug use, excessive exercise, weight loss or gain, exposure to environmental pollutants or toxicants, and psychogenic stress, can compromise reproduction, it is worth reviewing lifestyle factors with the couple and informing them of the potential adverse consequences of these factors in a supportive manner. However, the literature surrounding the contributions of many modifiable lifestyle factors demonstrates inconsistent effects on fertility (see Table 30.1–1). Psychogenic variables can interfere with coital frequency, reduce the likelihood of conception, suppress the central hypothalamic-pituitary drive to the gonad, and compromise the quality and quantity of gametes produced. In addition, a diagnosis of infertility, and the therapeutic interventions available to support conception, can result in significant psychological distress for both partners, and discussions should be approached with sensitivity. Given limited evidence surrounding the role of lifestyle factors, coupled with the psychological impact of infertility, we caution providers against ascribing causality between any given lifestyle factor and a couple's fertility challenges.

Standard interventions for family building in infertile couples include a variety of assisted reproductive techniques (ART), including ovulation stimulation (via selective ER modulator, aromatase inhibitor), ovulation induction (via injectable exogenous gonadotropins), intrauterine insemination, in vitro fertilization with fresh or frozen embryo transfer (IVF-ET), and intracytoplasmic sperm injection (ICSI), as well as assisted cycle monitoring/optimization of natural fertility, gestational surrogacy, or adoption. In in vitro fertilization techniques, the ovaries are generally hyperstimulated using injectable exogenous gonadotropins to produce at least three or four mature follicles from which the oocytes can be retrieved using a transvaginal procedure. Fertilization can then occur in two ways: (1) Sperm and oocytes are mixed in a laboratory and supported in vitro, or (2) a single sperm can be injected into each egg (ICSI) to promote fertilization. If fertilization occurs, the fertilized embryo or blastocyst is transferred to the recipient's uterus after the endometrial lining has been optimized for implantation. ART carries a substantially increased risk of multifetal gestation. Because of the increased maternal perinatal and neonatal morbidity/mortality of twin or higher-order multiples, professional societies recommend the transfer of a single embryo regardless of patient age. However, the transfer of more than one embryo may be appropriate in select patients over 37. More than 1.5% of all live-born infants in the United States are conceived via ART. However, approximately 25% of implantations (as evidenced by a transiently elevated serum β-hCG concentration) are not sustained. The likelihood of a successful full-term pregnancy increases dramatically if fetal cardiac activity is observed approximately 2 weeks after missed menses (i.e., 4 weeks after the gamete or embryo transfer). Gamete intrafallopian transfer (GIFT) is a form of assisted reproduction with similar rates of success, in which three or four oocytes are mixed with sperm and placed into the fallopian tube during a laparoscopy. Fertilization occurs in the tube and, thus, is not observed. IVF-ET has largely supplanted this procedure because GIFT requires a laparoscopy and general anesthesia and because pregnancy results are not better than with IVF-ET.

When it is not possible to correct underlying causes, substitutions are possible. Not only are donor sperm available, so too are donor oocytes, gestational surrogates, and uterine transplants. Embryos can be cryopreserved indefinitely and potentially transferred to a biologically unrelated recipient or the woman herself years after the menopausal transition. Oocyte retrieval and cryopreservation are now available outside of investigational protocols for fertility preservation in patients undergoing gonadotoxic therapy or gonadectomy (often in the setting of malignancy) or in women with a genetic predisposition for ovarian failure (such as fragile X-associated primary ovarian insufficiency). Although not yet endorsed by professional societies, this technology is being marketed for fertility preservation in healthy women who choose to delay childbearing and considered in people seeking biomedical gender transition. The most recent advances in ART focus on optimizing IVF live birth rates, fertility preservation, preimplantation screening for aneuploidy, and mitochondrial replacement to prevent certain diseases.

Infertility treatments are expensive and consume much time and energy. A sense of psychological and physical invasion is common. Infertile men and women may feel angry, damaged, or guilty. Studies suggest an association between women with infertility and increased risks of invasive ovarian, endometrial, and breast cancers, further increasing anxiety and psychological turmoil. The potential for husbands and wives to respond differently to the experience of infertility and its treatments, particularly any of the ARTs, can overwhelm customary adaptation patterns and lead to psychological decompensation. Even successful infertility treatment poses challenges. Singleton pregnancies achieved through ART have higher odds of perinatal mortality, preterm delivery, low birth weight or very low birth weight, Cesarean delivery, placental abruption, pre-eclampsia, placenta previa, vasa previa, and stillbirth. ART is also associated with increased risks of multifetal gestation. In 2011, the proportion of twin pregnancies arising from ART was 36% (17% attributed to IVF and 19% attributed to ovulation induction or stimulation) and 77% of all triplet pregnancies (32% attributed to IVF and 45% attributed to ovulation induction or stimulation). In higher-order multiple gestations, the couple must grapple with the option of selective reduction to optimize maternal and fetal outcomes of the pregnancy. ART is associated with an increased chance of monozygotic twinning among twin pregnancies, bringing the additional risk of intrauterine growth abnormalities, congenital anomalies, and twin-to-twin transfusion syndrome leading to perinatal complications. A multiple pregnancy also increases overall preterm labor and delivery risks, with associated neonatal sequelae of prematurity, pre-eclampsia, gestational diabetes, and Cesarean delivery.

Furthermore, multiple gestation complicates postpartum adjustment and carries a lifelong parenting challenge. There is some evidence that children born to previously infertile parents are viewed by their parents as special. These parents may have trouble setting limits or may be overly protective.

Whether or not it is treated, infertility can produce a variety of psychological reactions. Responses depend on personality attributes, including adaptability, cultural expectations, support systems, knowledge about reproductive processes, and the attitudes of the involved clinicians. Infertility can damage self-image and impair a sense of health and integrity. Some individuals feel that they are being made to suffer for past wrongs; others blame the medical profession for being inadequate or inattentive; some regret decisions surrounding delayed childbearing. A sense of panic, helplessness, and loss of control may compromise self-esteem. Women and men alike are generally overwhelmed by the complexity, cost, invasiveness, and uncertainty associated with medical intervention or adoption. Almost all patients who undergo infertility treatment (Table 30.1–2) experience a loss of privacy as they cope with physicians, nurses, insurance carriers, psychologists, hospitals, and laboratory personnel. The sense of intimacy that reinforces marriage and mutuality may be challenged when sex and procreation are separated by technology (Table 30.1–3). Ultimately, infertility presents the specter of psychological death. Thus, it is not surprising that patients with infertility report as much anguish and distress as patients newly diagnosed with acquired immunodeficiency syndrome (AIDS).

Same-sex and transgender couples often seek these same medicalized assisted reproduction techniques to build families and may also experience the challenges of infertility superimposed on the need for assisted reproduction. Lesbian and transgender couples often seek fertility-industry–assisted intrauterine insemination via donor sperm or pursue oocyte retrieval and in vitro fertilization from one partner followed by embryo transfer and pregnancy carried by the other partner. Same-sex male or transgender couples have increasingly sought in vitro fertilization and gestational surrogacy. Compounding the psychological stresses and potential complications of seeking these assisted forms of reproduction for family building, LGBTQ+ parents must also navigate complex legal systems with differing state-by-state regulations surrounding assisted reproduction that restrict their parenthood and parental rights.

Ideally, psychological support services should be available for all individuals and couples undergoing active evaluation and biomedical fertility interventions. Unlike the traditional medical model, in which diagnostic testing precedes and directs the course of treatment, empirical therapy may precede more-invasive diagnostic testing in infertility. Cause and effect can be difficult to identify, and magical or superstitious thought processes are common in patients and practitioners. Fertility treatments can threaten interpersonal relationships and romantic relationship adjustment, particularly if the partners react differently to the challenges. Although fertility is an attribute of a couple, blame may be prematurely or incorrectly assigned to one person in the dyad. In addition to individual or couples counseling, groups, such as Resolve, can offer support and information. When donor sperm, oocytes, embryos, or gestational surrogacy are being considered, a thorough psychological evaluation is recommended before treatment to ensure that all parties agree to the treatment plan and understand the potential implications and ramifications.

Dysfunctional attitudes and disharmony constantly threaten persons seeking fertility treatments. Prophylactic evaluation is preferred in all treatment settings, but specialized services may not always be immediately available. Stress can compromise gametogenesis and libido, and fertility treatment alone can be stressful enough to activate the central mechanisms that compromise reproductive function. Psychological support is intended to lessen the likelihood of this effect and reverse it if it is already occurring.

Pharmacologic agents used to treat ovulation disorders or hyperstimulate the ovaries in preparation for oocyte retrieval may alter mood and cognition. These agents include clomiphene citrate (Clomid, a selective estrogen receptor modulator [SERM]), Letrozole (an aromatase inhibitor), GnRH analogs, human gonadotropins, hCG, estrogen and progesterone, and bromocriptine (Parlodel). Danazol (Danocrine, a synthetic androgen) or leuprolide (a gonadotropin hormone-releasing agonist) may be used to arrest the growth of endometriosis (greater than stage I/II) as an adjunct to later fertility therapy. Most of the agents alter estradiol and progesterone levels, and the resultant hormonal fluctuations may trigger mood lability and depression in sensitive individuals. Agents such as clomiphene, danazol, and bromocriptine may also directly affect the brain. The current understanding of the effects of ovarian hormones on cognitive function is largely drawn from studies of postmenopausal women receiving hormone replacement therapy (HRT). While some observational studies have suggested that estrogen exposure may affect cognition, randomized control trials have not demonstrated cognitive benefit from HRT initiated after natural menopause. The Women’s Health Initiative Memory Study (WHIMS) and The Women’s Health Initiative Study of Cognitive Aging (WHISCA) are randomized, placebo-controlled clinical trials of 7,429 and 2,302 women, respectively. Neither found an association between HRT and cognitive improvements, nor between HRT and a decreased risk of dementia. However, cognitive decline has been associated with surgical menopause in women undergoing elective oophorectomy at the time of hysterectomy; therefore, data remain inconclusive.

Table 30.1–2.
Tests in the Infertility Workup

Possible Cause	Test	Comments
Ovulatory dysfunction	Basal body temperature chart	Patient must do each morning
	Endometrial biopsy	Infrequently used office procedure in the late luteal phase used to determine if the endometrium has been adequately stimulated with progesterone
	Serum hormone testing	Serum progesterone can be measured 1 week prior to expected menses; LH and estrogen can be measured just prior to expected ovulation
	Urinary ovulation detection kit	Home test kits that show changes in the level of LH in a woman's urine preceding ovulation
	Transvaginal ultrasound	Visualizes ovarian follicles and their size (ovulation generally occurs when the follicle measures about 1.8–2.5 cm)
Anatomic disorder	Hysterosalpingography	X-ray done in the proliferative phase, in which a radio-opaque dye is introduced through the cervix into the uterus to define

		intrauterine contours and tubal patency
	Sonohysterography	Transvaginal ultrasound examination with instillation of saline into the uterine cavity to define contours
	Diagnostic laparoscopy	Infrequently used surgical procedure to view external surfaces of internal female reproductive structures for evidence of endometriosis or scarring
	Hysteroscopy	Visualizes endometrial cavity directly
Diminished ovarian reserve	Antral follicle count	A transvaginal ultrasound done in the early part of the menstrual cycle to count the number of small (2–10 mm) follicles in the ovary
	Serum hormone levels	High FSH and estradiol on postmenstrual day 3 or low antimüllerian hormone levels may indicate diminished ovarian reserve
Abnormal spermatogenesis	Semen analysis	Normal value >20 million/mL, 2 mL volume, 60% motility
	Postcoital test	Midcycle timing to look

		at sperm–cervical mucus interaction
Immunologic disorder	Antisperm antibodies	Male semen test
Azoospermia	Testicular biopsy	Determines eligibility for intracytoplasmic sperm injection
	Semen analysis for fructose	Determines if the vas deferens are patent

Adapted with permission from Beckmann CRB, Ling FW, Barzansky BM, Bates G, Herbert W. *Obstetrics and Gynecology for Medical Students*. Williams & Wilkins; 1992:353.

Table 30.1–3.
Assisted Reproduction Techniques

Method	Comments
Ovulation induction or augmentation (multiple agents, particularly recombinant or highly purified gonadotropins)	Stimulates multifollicular development and ovulation; may produce multiple births; used in anovulation, luteal phase deficiency, unexplained infertility, and assisted reproduction
Induction of spermatogenesis	Used in men with idiopathic hypothalamic hypogonadism of functional or organic nature
Intrauterine insemination	Donor sperm is injected into the uterine cavity or the fallopian tubes; the sperm of the male partner may be used if healthy
Gamete intrafallopian transfer	Transfer of collected oocytes and sperm into the fallopian tubes; zygote may also be transferred; used for infertility from endometriosis and unexplained infertility
In vitro fertilization and embryo transfer	Transfer of developing embryos into the uterus after extracorporeal incubation of collected sperm with oocytes retrieved by laparoscopic surgery or by ultrasound-guided transvaginal aspiration; used with occlusion of the fallopian tubes or significant sperm dysfunction; permits preimplantation genetic diagnosis
Intracytoplasmic sperm injection	Injection in vitro of sperm head or sperm DNA to cause fertilization and production of embryos for transfer to receptive endometrium;

	may be used even if sperm are barely viable; genetic causes of male infertility may be transmitted to offspring
Gamete donors	Donation of sperm or oocytes to another couple; can include oocytes only, oocyte cytoplasm only to restore reproductive competence to aged oocyte, donor sperm alone, or any combination; cytoplasmic transfer considered unethical and is illegal in some states
Gestational surrogacy	An arrangement where a woman carries and delivers a child for another couple or person. Laws vary for each state in the United States and contracts are involved.
Embryo donor	Donation of a fertilized embryo in order to help an intended parent(s) have a child. No genetic link between embryo and intended parent(s), however the intended parents are able to experience pregnancy and birth

Data from Susman V. *Pregnancy in Behavioral Sciences for Medical Students*. Williams & Wilkins; 1993.

The first systematic review on the psychopathological effects of infertility treatments was recently published. Nine clinical trials were included in the review. One study demonstrated fewer mood symptoms associated with gonadotropin-releasing hormone (GnRH) antagonists than GnRH agonists. Two studies compared GnRH agonists/antagonists to natural cycle protocols and found fewer symptoms associated with the protocols that avoid gonadotropin stimulation. The National Institute on Aging (NIA) is currently conducting a study on the effects of hormonal infertility treatment on cognition in premenopausal women. This and future work may better characterize fertility treatments' potential affective and cognitive implications.

Treatment of infertility itself is anxiety-provoking because of the uncertainty and sense of expectation. A further concern of infertile couples is whether ARTs increase the risk of neurodevelopmental or other abnormalities, including childhood cancers. The uncertainty about whether there is a link between infertility treatments and poor health outcomes in the mother or child only exacerbates anxiety and guilt. Infertility is a far more devastating disorder than most laypersons and physicians realize.

Early Pregnancy Loss and Stillbirth

Unlike infertility, which can continue indefinitely, pregnancy loss is often a defined event. *Early pregnancy loss* (i.e., miscarriage, spontaneous abortion) is defined as a nonviable, intrauterine pregnancy with either an empty gestational sac or a gestational sac containing an embryo or fetus without fetal heart activity within the first 12 6/7 weeks of gestation. Early pregnancy loss is a common occurrence. Eighty percent of pregnancy losses occur in the first trimester, and an estimated 10% of all clinically recognized pregnancies result in early pregnancy loss. This rate is an under-estimation, given that many women who experience early miscarriages are unaware that they were pregnant. Because the most common reason for early pregnancy loss is a cytogenetic abnormality, it follows that the frequency of clinically recognized early pregnancy loss is highly dependent on age, with rates 9% to 17% in women ages 20 to 30 years, 20% at age 35 years, 40% at age 40 years, and 80% at age 45 years. A single miscarriage does not predict poor future reproductive performance. *Recurrent pregnancy loss* is inconsistently described but is most commonly defined as two or more clinical pregnancy losses. Fewer than 5% of women will experience two consecutive miscarriages, and only 1% will experience three or more. An etiology for recurrent pregnancy loss is made in about 50% of patients and includes factors related to genetics, age, antiphospholipid syndrome, uterine anomalies, thrombophilias, hormonal or metabolic disorders, infection, autoimmunity, sperm quality, and lifestyle issues. Therefore, recurrent pregnancy loss can more closely resemble infertility in its chronicity and sense of lost potential than a defined event. The term *stillbirth* (i.e., fetal death or perinatal loss) refers to fetal death after 20 weeks gestation and is estimated to be around 6 per 1,000 live births, though it is less frequent at later gestational ages (about 3 per 1,000 live births after 28 weeks).

Acute reactions to an isolated miscarriage or stillbirth, such as shock, blame, guilt, numbness or anger, and denial, can gradually decline and lead to acceptance and resolution. However, a minority of parents continue to experience more persistent and impairing symptoms, diagnosed as complicated grief. Parents are at an increased risk for complicated grief developing when the death occurs suddenly or traumatically. Though there is no strict timeline for the grief process, diagnosing a parent with complicated grief 6 months following a fetal loss would be inappropriate. Each parent may experience grief differently, and the event can have a lasting impact on the relationship. When a patient experiences stillbirth, bereavement care should be individualized to personal, cultural, or religious needs. It should include recognition of parenthood, acknowledgment of partners' and families' grief, an acknowledgment that grief is individual, awareness of burial cremation, and funeral rituals. Although it is uncommon to identify an exact etiology of stillbirth, good communication of test results related to a fetal autopsy, placental examination, or genetic testing and the diagnoses excluded may relieve feelings of parental guilt or anger. Individual bereavement counseling and peer support groups can also benefit symptoms of anxiety or depression after a stillbirth. Although it might be assumed that grief and emotional turmoil would be related to gestational age, observations do not support this assumption. Women who miscarry have more than twice the relative risk of major depression than community women. More than half of women with a history of major depression experience a recurrence after miscarriage. Because of the significant risk of depression, acute intervention is recommended for all women and couples who experience a miscarriage or stillbirth; services should be extended beyond the acute event when indicated.

Pseudocyesis

Pseudocyesis is the development of the classic symptoms of pregnancy—amenorrhea, nausea, breast enlargement and pigmentation, abdominal distention, and labor pains—in a nonpregnant woman. Pseudocyesis demonstrates the ability of the psyche to dominate the soma, probably via central input at the level of the hypothalamus. Predisposing psychological processes are thought to include a pathologic wish for and fear of pregnancy; ambivalence or conflict regarding gender, sexuality, or childbearing; and a grief reaction to lost potential after a miscarriage, tubal ligation, or hysterectomy. The patient may have a true somatic delusion that is not subject to reality testing, but a negative pregnancy test result or pelvic ultrasound scan often leads to resolution. Psychotherapy is recommended during or after a presentation of pseudocyesis to evaluate and treat the underlying psychological dysfunction. A related event, couvade, in which the child's father undergoes simulated labor as if he were giving birth, occurs in some cultures.

Fertility Control

Voluntary Termination of Pregnancy. Voluntary termination of pregnancy (VTOP) is distinct from spontaneous abortion (i.e., miscarriage) (Table 30.1–4). VTOP is the planned termination of a pregnancy. In 2017, approximately 862,000 abortions were performed in the United States—13.5 abortions per 1,000 reproductive-aged women (15 to 44 years). This correlates with 18.4% of all U.S. pregnancies in 2017 ending in VTOP. Abortion rates in the United States have been steadily declining and reached the lowest rates since the passage of Roe V. Wade in most recent statistics, perhaps related to increasing abortion restrictions such as counseling and ultrasound requirements, waiting periods, and clinic closures. As in the past, in the United States in 2014, a disproportionately higher number of abortions were obtained by heterosexual women (94%), white women (39%), multiparous women (59%), women using contraception the month they became pregnant (51%), poor or low-income women (75%), and women younger than age 30 years (61%).

Table 30.1–4.
Types of Pregnancy Loss and Termination

Spontaneous abortion	Spontaneous expulsion of the products of conception before 20 weeks of gestation
Recurrent spontaneous abortion	Two or more spontaneous abortions
Missed abortion	A pregnancy in which there is a fetal demise (usually for a number of weeks) but no uterine activity to expel the products of conception
Threatened abortion	An ongoing pregnancy complicated by bleeding (+/– cramping) before 20 weeks of gestation
Incomplete abortion	Some, but not all, of the products of conception have been passed; retained products may be part of the fetus, placenta, or membranes
Septic abortion	A spontaneous abortion that is complicated by intrauterine infection
Inevitable abortion	The cervix has dilated, but the products of conception have not been expelled
Complete abortion	All products of conception have been passed without the need for surgical or medical intervention
Induced abortion (voluntary termination of pregnancy)	Using medication or a surgical procedure to end a pregnancy as a means of fertility control, or when pregnancy poses a high risk of morbidity/mortality to the mother, or in cases of life-limiting fetal anomalies; medication abortion can be performed up to 70 days

gestation and includes use of mifepristone (RU-486, an anti-progestin) with or without misoprostol (prostaglandin analog) or methotrexate. Surgical procedures include manual vacuum aspiration or suction curettage in the first trimester, and dilation and evaluation (D&E) in the second trimester (up to ~24 weeks). Labor induction termination may also be used in the second trimester. “Partial birth abortion” is not a medical term but refers to a ban on intact D&E wherein a fetal demise does not occur before surgery.

In the United States in 2016, 65.4% of legally induced abortions were performed at or before 8 weeks of gestation, 14.7% between 9 and 10 weeks, 8.2% between 11 and 12 weeks, 6.3% between 13 and 15 weeks, 4.1% between 16 and 20 weeks, and 1.3% over 21 weeks. Medication abortion (used up to 10 weeks of gestation) accounted for 39% of all VTOP. Of concern are statistics reporting that up to 18% of abortion clinics in 2017 reported seeing patients who had attempted self-induced abortion. Table 30.1–5 summarizes the most common abortion techniques.

In contrast to prior discourse, there is no high-quality evidence showing significant untoward psychological effects of VTOP on women’s mental health. In fact, there is high-quality evidence demonstrating that rates of mental health problems for women with an unplanned/undesired pregnancy are the same whether they continue to live birth or proceed with a first-trimester abortion. In one study comparing the outcomes of women who sought treatment in a clinic because they thought they were pregnant, women who were pregnant and chose to abort were reported later to be doing better emotionally and in pursuing personal, educational, and work endeavors than were either those who were not pregnant or those who chose to continue the pregnancy. Another study suggested untoward consequences on women's mental health and well-being who were denied the option to abort and, thus, were required unwillingly to carry the pregnancy to term and then raise an undesired child. This study suggested untoward consequences on the mental health and well-being of the child as well. However, termination of a wanted pregnancy because of an abnormal karyotype or fetal anomalies can be traumatic, and supportive intervention is recommended.

Techniques for inducing abortion (Table 30.1–5) have been used for at least 5,000 years. Given the considerable risks associated with childbearing and the physical, emotional, and financial demands of child-rearing, VTOP was viewed historically as a means of modulating the maternal and societal risks of unrestrained reproduction. Historically, the concept of therapeutic abortion before *quickening,* which is generally defined as the time at which the mother perceives fetal movements, was well accepted. In the United States, legislation regulating VTOP was introduced in the early 1800s, with the intent to protect women from sepsis and other complications associated with the procedures. However, VTOP, whether by surgical or chemical means, now entails minimal medical risk and is safer than carrying a pregnancy to term. Furthermore, there is no high-quality evidence to support the suggestion that VTOP increases risks of breast cancer, adverse future pregnancy outcomes, or infertility, as some misinformation exists on this.

Medication abortion, which can be used in selected patients under 70 days gestation, is commonly achieved using two medications: mifepristone (previously called RU-486) and misoprostol. In cases where regulations prohibit the use of mifepristone, misoprostol-only regimens are also validated. Mifepristone is a selective progesterone receptor modulator that binds to the progesterone receptor with a higher affinity than progesterone. However, it blocks its action, leading to decidual necrosis and softening of the cervix, sensitization to prostaglandins, and subsequent increased uterine contractility. Misoprostol is a prostaglandin E1 analog that stimulates uterine contractions and softens the cervix. Uterine aspiration is a surgical procedure used to induce abortions in the first trimester, and dilation with evacuation (or D&E) is used for second-trimester abortion. For pregnancies with lethal fetal anomalies or maternal medical indications for termination of pregnancy beyond 24 weeks, induction of labor and vaginal delivery (termed “induction termination”) is recommended.

The medical risks of both surgical and medical VTOP are few, but all procedures require physician input. Medically induced VTOP, however, requires less physician input than do surgical techniques, although medical VTOP requires careful monitoring to ensure that the termination occurs in a timely fashion. Rarely hemorrhage can occur and require surgical intervention or a transfusion. Studies indicate that women who choose to terminate their pregnancies do at least as well as women who continue their pregnancies to term or were never pregnant concerning physical and mental health and educational, occupational, and marital functioning. Diana Greene Foster, a demographer with an area of focus on family planning, directed a 10-year prospective longitudinal study, entitled the Turnaway Study, that assessed the psychological, physical, and financial impact of electing a VTOP versus continuing an unwanted pregnancy to term. The study, which resulted in more than 50 peer-reviewed papers, found no evidence that pregnancy termination increases a woman’s risk of psychiatric illness, neither in the short- nor long-term. In fact, the study demonstrated that being denied a VTOP was associated with a greater risk of psychiatric illness than undergoing a VTOP in the short term.

Table 30.1–5.
Abortion Techniques

Timing	Type	Benefits	Risks
First trimester	Manual vacuum aspiration or suction curettage	Manual vacuum aspiration may be performed in an office setting (avoiding anesthesia), may be performed after 70 days gestation, immediate examination of tissue aspirate can confirm complete procedure, slightly increased efficacy compared to medication abortion	Rare complications of hemorrhage, cervical laceration, retained products of conception, infection, or uterine perforation
	Medication abortion (use of medications such as mifepristone, prostaglandin analogs, and methotrexate to induce an abortion up to 70 days gestation)	Avoids risks and costs of procedural intervention, perception of privacy	Takes longer to complete than uterine aspiration, requires active patient participation, incomplete medication abortion or ongoing pregnancy, may be contraindicated

			in some patients
Second trimester	Dilation and evacuation (cervical preparation with osmotic dilators and/or medications [mifepristone, prostaglandin analogs] followed by surgical dilation and evacuation of uterine contents using suction and grasping forceps)	Fewer complications than medication abortion, less pain reported, may be considered less emotionally challenging by patients	Rare complications of hemorrhage, cervical laceration, retained products of conception, infection, or uterine perforation
	Induction of labor termination (use of pharmacologic uterotonic medications such as mifepristone, prostaglandin analogs, oxytocin, with or without osmotic	May be preferred by some patients/providers for termination of pregnancies complicated by fetal anomalies or genetic disorders (when preservation of fetal anatomy is desired), or maternal health issues; may be	Compared with D&E, less cost-effective, greater risk of complications (such retained products of conception requiring surgical procedure), and may be prolonged

cervical dilators or transcervical Foley catheter. May sometimes be combined with an intraamniotic procedure to stop the fetal heartbeat prior to induction of labor).	necessary if a D&E trained physician is not available	
Abdominal surgery (hysterotomy or hysterectomy)	Should only be performed in rare instances when D&E or medication abortion have failed or are contraindicated	Much higher risks of complications than D&E or medication abortion

Data from ACOG Practice Bulletin #135. June 2013. https://www.acog.org/clinical/clinical-guidance/practice-bulletin/articles/2013/06/second-trimester-abortion#:~:text=The%20proportion%20of%20abortions%20performed,the%20past%20two%20decades%204; ACOG Practice Bulletin #225, October 2020, https://www.acog.org/clinical/clinical-guidance/practice-bulletin/articles/2020/10/medication-abortion-up-to-70-days-of-gestation.

Current legislative attention has shifted from safeguarding maternal health to the ethics of abortion and the right of women to exercise control over their bodily functions. Traditionally, ethical deliberations regarding VTOP were reserved for religions. Given the diversity of religious perspectives on the role of women and procreation, a consensus is neither likely nor necessarily desirable. Nonetheless, the current legislative debate centers on who has the authority to regulate induced abortion and to what degree. The 1992 Supreme Court ruling in *Planned Parenthood of Southeastern Pennsylvania v. Casey* upheld the basic right of women to elect pregnancy termination before fetal viability. However, it also upheld the right of the state to regulate and restrict abortion and define viability. Furthermore, in some states (e.g., Pennsylvania), pregnant women do not have the right to select whether they wish to receive advanced life support in life-threatening situations. The Freedom of Choice Act put before the U.S. Congress recognized the right of states to regulate the strictly medical aspects of abortion to safeguard maternal health. It restricts a woman's freedom to terminate a pregnancy after, but not before, fetal viability. This legislative effort represented a compromise between those who argued against any medical mediation of the process and those who would grant women complete autonomy. Application of this law would hinge on the interpretation of the phrase *fetal viability,* a concept that is not easily defined. Making the option of induced abortion contingent on fetal viability reveals continuing ambivalence about allowing women the autonomy to make decisions about physiologic processes that directly shape their lives. A recent example is the Texas Heartbeat Act which was passed in May 2021 and took effect in September 2021. The new legislation bans abortions following the detection of cardiac activity, which typically occurs as early as six weeks gestation when many women are not yet aware of the pregnancy. The act restricts abortion in the country's second-largest state to an extent not seen in fifty years, without exceptions for rape, sexual abuse, or incest. Various reproductive rights groups requested an injunction from the U.S. Supreme Court to prevent the law from being enacted. The Court denied the preliminary injunction and later remanded the case back to the Fifth Circuit. In December 2021, a state district judge ruled that portions of the statute's enforcement mechanism violated the Texas constitution but did not enjoin the law's enforcement. In the summer of 2022, the Supreme Court ruled in Dobbs v. Jackson that the constitution does not protect the right to abortion, and 27 states now have at least some restrictions.

Contraception. In 2018, there were 72.7 million reproductive-age women, of which 46 million were sexually active and not seeking to become pregnant. Unfortunately, 49% of pregnancies in the United States are unintended, higher than rates in most developed countries. Lack of knowledge, misperceptions of contraceptive methods as abortifacients, exaggerated safety concerns about contraceptives, and emphasis on abstinence-only sexual education all contribute as barriers to contraceptive use, and thus to family planning and pregnancy intendedness. Despite this, 99% of sexually active American women have used a contraceptive method other than natural family planning during their lifetime, and in 2018, 65% of women ages 15 to 49 were currently using a contraceptive method. Methods of contraception fall into five main categories: barrier methods (such as condoms or diaphragms), hormonal agents (including oral contraceptive pills/patch/vaginal ring, depot injection, and emergency contraception), long-acting reversible contraceptives (LARCs; such as intrauterine devices or implant), surgical sterilization procedures, and fertility-awareness methods (natural family-planning) that require abstinence around the time of ovulation (Table 30.1–6). Most women utilizing contraception use a combined hormonal (estrogen/progesterone) contraceptive (CHC), such as a pill, patch, or ring. However, LARCs are gaining popularity due to their higher efficacy. In fact, from 2002 to 2012, LARC use increased from 2% to 12% in the United States. The perfect contraceptive would be completely reversible, free of adverse effects, and completely efficacious with typical use. Because no perfect contraceptive is available, people must select a method that best meets their needs, and this involves assessing personal preferences, reproductive goals, method efficacy, acceptability, and availability. Decisions regarding contraceptive methods must also consider underlying medical conditions, which may change the balance of potential risks of certain contraceptive methods while also acknowledging the potential risks of unintended pregnancy. Frequently, the responsibility for choosing and using a contraceptive falls primarily on the woman within a heterosexual relationship. This can lead to resentment, lack of compliance, and marital disharmony. For those in a stable relationship, contraceptive counseling can include both partners.

Given the high prevalence of both mental illnesses and contraceptive use, providers should remain cognizant of potential drug-drug interactions when caring for women of reproductive age. CHC may alter the efficacy of certain psychotropics, and vice versa, through pharmacokinetic and/or pharmacodynamic interactions. Carbamazepine and St. John’s Wort may decrease CHC efficacy by increasing the metabolism of hormones, thereby effectively reducing the dose of the CHC and, in turn, the effectiveness of the contraceptive. Concurrent use of CHC and lamotrigine may reduce the serum concentration of lamotrigine, possibly requiring an increase in lamotrigine dose to maintain a therapeutic effect. However, CHC can have the opposite effect on clozapine dose: The estrogen in CHC may impair clozapine metabolism, resulting in an increase in serum clozapine level and possibly requiring a halving of the clozapine dose. Regarding anxiolytic use, CHC will increase the effect of benzodiazepines, while barbiturates reduce the efficacy of CHC. CHC use does not modify the effectiveness of fluoxetine, and conversely, selective serotonin reuptake inhibitors (SSRIs)/serotonin-norepinephrine reuptake inhibitors (SNRIs) do not appear to interact with CHC metabolism.

The assessment of the risks and benefits of a given contraceptive strategy should not be limited to considering the biologic risks alone but must consider the mindset of each party, particularly regarding gender issues and cultural perspectives. If significant discord exists within a couple, counseling may be indicated. Counseling can also be considered for the person or couple who feels guilty about the idea of interfering with natural bodily functions for this purpose.

Clinicians must be mindful of the difference between actual and perceived risk. Concerns of current and later bodily harm from the use of contraception are common, and the risks of childbearing and child-rearing frequently are underestimated. The mortality associated with childbearing is estimated to be 10 per 100,000 live births between the ages of 20 and 24 years, 44 per 100,000 between the ages of 35 and 39 years, and 71 per 100,000 after age 40 years. However, comparable mortality in nonsmoking women on oral contraceptives is 1.2 per 100,000 live births, 3.9 per 100,000 live births, and 6.6 per 100,000 live births, respectively. The potential risks of child-rearing are difficult to quantify, including role strain and stress, interrupted education or career advancement, depression, loss of independence, and compromise of the psychosexual relationship with one's partner. This misperception of risk is particularly prevalent among adolescents, who may be uninformed and have unrealistic ideas about the rewards of pregnancy and the risks of STDs. In addition, a teenager's aptitude for listening and making decisions may be limited by impulsivity or ambivalence.

In short, despite the availability and relative affordability of contraceptive means, their use is constrained by the discrepancy between actual and imagined risk.

A few recent studies report a correlation between oral contraceptives and a diagnosis of depression among adolescents. However, possible confounders and the causality of the relationship remain an active area of study with ongoing debate. A prospective study conducted in Denmark demonstrated an association between oral contraceptive use and suicide, with 2 months following the initiation of oral contraceptive identified as the time of highest risk for an attempt. A prospective cohort study followed 800,000 Swedish women over the year following initiation of a hormonal contraceptive. Adolescents taking hormonal contraceptives were more likely to be prescribed a psychotropic medication during the year of follow-up than were women naive to hormonal contraceptives. This relationship was not demonstrated among adult women. The GABA pathway, which has demonstrated sensitivity to neurosteroid changes, is possibly implicated in the relationship between hormonal contraceptives and the subsequent development of depression, though mechanisms have not yet been elucidated. However, the causality of the relationship remains in question. A confounder may be responsible for the association between hormonal contraception and depression; hormonal contraception use might serve as a marker for another characteristic directly responsible for the increased risk of depression among adolescents. A more recent study analyzing data from Swedish national registers found considerable heterogeneity in the association between hormonal contraceptive use and depressive illness, identifying the relationship to be strongest among adolescents, immigrants, and low-income earners. The association remains an active topic of study and continues to be debated. Current professional guidelines state that women with depressive disorders can safely use all hormonal contraception methods.

Because coitus is not always anticipated, there is a need for postcoital or the so-called emergency contraceptives. No clinical examination or pregnancy testing is required prior to using emergency contraception, and treatment should be initiated as soon as possible after unprotected intercourse to improve efficacy. Available methods include high doses of levonorgestrel, high doses of combined hormonal contraceptive (CHC) pills, antiprogestins, and intrauterine device (IUD) insertion. Insertion of the Copper IUD (and most recently the Levonorgestrel IUD) within 120 hours of unprotected intercourse remains the most effective method of emergency contraception but can be logistically challenging for patients to access. The use of levonorgestrel 1.5 mg pill (Plan B) within 72 hours of intercourse has greater efficacy and is associated with less severe adverse effects than a high dose of combined oral contraceptives (the Yuzpe method). In the United States, Plan B is available without a prescription to women 18 years or older. Selective progesterone receptor modulators (ulipristal acetate) require a prescription. However, they can be effective up to 120 hours after unprotected intercourse, and in a recent meta-analysis, ulipristal acetate demonstrated greater efficacy than levonorgestrel in preventing pregnancy (pregnancy rates were 1.4% in users of ulipristal, compared to 2.2% in users of levonorgestrel). Emergency contraception should not be confused with medical abortion. In general, pharmacologic emergency contraception inhibits or delays ovulation but does not prevent the implantation of a fertilized oocyte. The IUD prevents fertilization by affecting sperm viability and function. None of these methods confers risk to an established pregnancy or teratogenicity to a developing embryo.

Table 30.1–6.
Current Methods of Contraception

Type	Efficacy (as % of women who will become pregnant in a typical year of use)	Advantages	Disadvantages	Risks
Barrier (chemical or mechanical)				
Spermicidal agents	28%	Readily available and easy to use	Messy; loss of spontaneity; requires forethought	Allergic reactions
Diaphragm, cervical cap	12% (diaphragm), 17–23% (cervical cap)	Inexpensive; does not interfere with menstrual cycle	User familiarity required; prescription and fitting required; may interfere with spontaneity; requires forethought	Recurrent urinary tract infections with diaphragm; allergic reactions to latex or spermicide
Male condom	18%	Readily available and easy to use; protects against sexually transmitted infections	May interfere with spontaneity; requires forethought	Allergic reactions to latex or spermicide
Female condom	21%	Inconvenient to use; protects against sexually transmitted infections	May interfere with spontaneity; requires forethought	Allergic reactions to latex or spermicide
Hormonal (suppresses ovulation and/or impairs endometrial development)				
Combined hormonal contraceptives	9%	Includes pills, patches, and intravaginal rings. Can reduce menstrual cramps, make periods lighter or shorter, can improve acne or reduced unwanted hair growth, can decrease risks of cancer of the uterus, ovary, or colon	Side effects including headache, nausea, breast tenderness, breakthrough bleeding; require daily or weekly use, regardless of frequency of sex; must be prescribed and monitored by health care professional	Associated with a small increased risk of deep vein thrombosis, heart attack, and stroke. Should not be used by women with certain risk factors or who are within the postpartum period. Can be taken continuously to achieve ovarian suppression and amenorrhea and to treat medical conditions such as endometriosis or dysmenorrhea
Postcoital steroids (emergency contraception)	2% (pregnancy rate after use)	Can be used after unprotected intercourse or contraceptive failure to prevent a pregnancy. Methods include oral estrogen-progestin, progestin only, selective progesterone receptor modulators, or insertion of a copper IUD	Must be initiated as soon as possible (within 72–150 hours after unprotected intercourse) to maximize efficacy. Side effects can include nausea, headache, irregular bleeding	Not effective for long-term contraception, body weight can influence effectiveness of oral emergency contraceptive methods
Contraceptive implants (rods)	0.05%	Long-acting and reversible contraceptive implant (approved for 3 years, effective up to 5 years); once in place, no forethought required, few medical contraindications to use, one of the most effective forms of contraception available	Irregular bleeding and spotting due to endometrial effects and suppression of ovarian function; requires office procedure for placement and removal. Side effects such as unpredictable vaginal bleeding, headaches, breast pain, weight gain, acne	Problems with insertion or removal of the implant, slightly increased risk of ectopic pregnancy if pregnancy occurs while implant is in place
Progestin-only contraceptives	6% for injectable; 9% for oral pill	Includes progestin-only pills and depot injection. Can make periods lighter or stop periods altogether. Can be used safely in the postpartum period or while breastfeeding, and in women who have contraindications to using estrogen in combined hormonal contraception. Depot formulation can be administered every 13 weeks which improves user efficacy and it has been shown to reduce risks of uterine cancer and reduce pelvic pain caused by endometriosis	May lead to unpredictable spotting or bleeding in some women. Other side effects include headaches, nausea, breast tenderness. Depot formulation can cause weight gain in some women (about 5 lb), may not be reversed if pregnancy is desired prior to 3-month window, and can be also be associated with slow return of ovarian function (on average, 10 months to become pregnant after stopping the injection)	After prolonged use, depot formulation can be associated with bone loss and increased fracture risk later in life (though this is partially or completely reversible). Women with cardiovascular disease risk factors may have increased risk of developing cardiovascular disease while using the depot formulation

Intrauterine device	0.1–0.8%	Long-acting and reversible. Copper IUD is effective for 10 years, hormonal IUD is effective for 5 years. Once in place, no forethought required, few medical contraindications to use, one of the most effective forms of contraception available. Hormonal IUD can be used to decrease menstrual pain and heavy menstrual bleeding. There is immediate return to fertility once IUD is removed.	Requires office procedure for insertion/removal. Copper IUD can be associated with increases in menstrual bleeding and painful periods during the first months of use. Hormonal IUDs can cause frequent spotting, more days of bleeding, or heavier bleeding during first months of use, however over time length of period and amount of menstrual bleeding usually decreases or may stop completely.	In a small number of women, the IUD may spontaneously expel. During placement, the IUD could cause uterine perforation. Rarely, pregnancy can occur with IUD in place and if this occurs there is a higher chance of an ectopic pregnancy. Rarely, pelvic inflammatory disease can occur after IUD insertion.
Sterilization				
Male sterilization (vasectomy)	0.15–0.5%	Failure very rare; 20-minute office procedure; does not affect male sexual function. Safer, more effective, and less expensive than tubal occlusion procedures.	Can take 3–6 months to reach azoospermia	Can be reversed in only 80% of cases; Procedural complications include infection, bleeding, hematoma formation, granuloma formation, epididymitis
Female sterilization	0.15–0.5%	No impairment of sexual function or pleasure; total salpingectomy for tubal sterilization can reduce risks of ovarian cancer. Can be used safely by women with medical contraindications to hormonal contraception, can be performed in the immediate postpartum period, no alterations to menstrual cycles, not user dependent for efficacy and very low failure rates.	More complex procedure than vasectomy; reversal of tubal ligation is costly and very rarely successful.	Surgical morbidity; risk of post-tubal sterilization regret, increased risks of ectopic pregnancy (though lower than noncontraceptive users)
Other				
Fertility awareness-based methods	24%	Readily available	User must monitor cervical mucus and body temperature closely; imposed coital timing	None
Lactational amenorrhea method	8%	Temporary method of contraception can be used for the first ~6 months after giving birth by women who exclusively breastfeed	Can have breakthrough ovulation prior to first menses resulting in unintended pregnancy	None
Withdrawal, coitus interruptus	22%	Readily available	Difficult to implement	None

Data based on https://www.acog.org/womens-health/faqs/birth-control; https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD001329/full.

The mental and physical health risks of contraceptive use are significantly lower than are those associated with pregnancy, parturition, and parenthood. However, contraception often gives an unwarranted illusion of control over bodily processes. No method is perfect, and unintended pregnancy can occur despite responsible use. In addition, if a couple or person intends to have children, childbearing should not be postponed indefinitely hoping that the perfect time will arrive. It can be difficult to conceive on demand.

Sterilization. In women, sterilization can be accomplished by several methods of tubal occlusion, ligation, or excision as a primary surgical procedure, or may be effected through hysterectomy (with or without oophorectomy) for other medical indications. Hysterectomy for sterilization alone is associated with more morbidity and mortality than any other method of female sterilization and is therefore not primarily performed for this purpose. In men, vasectomy can be performed as an outpatient procedure with local anesthesia and is safer than any of the present female sterilization methods, with the lowest failure rate of any method of permanent sterilization (0.15% in the first year). Approximately 600,000 tubal occlusions and 200,000 vasectomies are performed each year in the United States. Postpartum partial salpingectomy has the lowest failure rate of any female sterilization method (7.5 per 1,000 procedures over 10 years), followed by silicone band method (17.7 per 1,000 procedures/10 years), bipolar coagulation (24.8 per 1,000 procedures/10 years), and titanium spring clip (36.5 per 1,000 procedures/10 years). Although uncommon, another risk of female sterilization procedures is ectopic pregnancy (7.3 per 1,000 procedures/10 years); overall, those who have undergone sterilization procedures have a much lower ectopic pregnancy risk than non-users of contraceptives. Sterilization procedures have not been shown to affect female or male sexual function or the menstrual cycle, and multiple observational studies report an association of tubal procedures with decreased incidence of ovarian cancer, particularly in women with BRCA1 or 2 mutations.

Comprehensive preoperative counseling is an important step in determining candidacy for sterilization. The permanence of the procedure, risks of failure, regret, alternatives (i.e., LARC and vasectomy), and prohibitive expenses of reversal procedures should be discussed with the patient who is deciding on permanent sterilization. In a well-informed and counseled patient, age and parity should not be barriers to accessing permanent sterilization.

It is difficult to determine how often a reversal of sterilization is sought or performed, and it is even more difficult to estimate the frequency of regret that the sterilization was performed. No formula can predict who will seek reversal, but factors such as remarriage, age <30 at time of sterilization, death of a child, and sterilization performance during the postpartum period seem to play roles. Regret is high when sterilization is undertaken to stabilize a marriage. The success of reversal depends on various factors, including the method of sterilization, the ages of the partners, and the elapsed time since the sterilization procedure. Clinicians need to make couples and individuals aware of these considerations but should not use them as reasons to deny the planned procedure. Again, when a couple is involved, the mutuality of the decision should be discussed to avoid later resentment. Voluntary sterilization, especially vasectomy, has become the most popular birth control in couples married for more than 10 years.

SEXUALLY TRANSMITTED INFECTIONS

An STI is a contagious disease acquired because of physical, sexual interaction. While STIs have coexisted throughout human history, treatments became more available in the 1950s for many previously life-threatening STIs. AIDS, caused by infection with human immunodeficiency virus (HIV), remains incurable, life-threatening, and transmissible from mother to fetus.

Trichomoniasis is the most-prevalent nonviral STI (2.1% among females, 0.5% among males in the United States). Most trichomonas infections have minimal or no symptoms, leading to untreated infection and passage between sex partners. Reproductive morbidity related to trichomoniasis is reported to be a 1.4 times greater likelihood of preterm birth and small for gestational age infants and a twofold increased risk for cervical cancer. Reproductive sequelae of STIs such as *Neisseria gonorrhoeae* and *Chlamydia trachomatis* are infertility, chronic pelvic pain (CPP), or ectopic pregnancy because of tubal scarring. Because asymptomatic infection is common among both men and women, many women diagnosed with these bacterial cervical infections already have a subclinical or clinical PID. PID is a spectrum of inflammation and infection of the upper female genital tract, including any combination of endometritis, salpingitis, tubo-ovarian abscess, and pelvic peritonitis. PID is attributed to *N. gonorrhoeae* and *C. trachomatis* in about 50% of cases, but it can also arise from other nonsexually transmitted vaginal microorganisms. *C. trachomatis* infection has the additional reproductive morbidity leading to congenital chlamydial pneumonia or ophthalmia neonatorum during pregnancy and delivery. *N. gonorrhoeae* during pregnancy and delivery can similarly lead to ophthalmia neonatorum, as well as disseminated gonococcal infection among neonates. In men, *N. gonorrhoeae* and *C. trachomatis* can also obstruct the vas deferens and chronic prostatitis in men and subsequent male infertility.

Another STI that can have serious consequences is venereal warts or human papillomavirus (HPV) infection. Genital infections with certain subtypes of HPV can lead to premalignant changes of the penis, vulva, vagina, and cervix and are thought to cause cervical cancer. An HPV vaccine is recommended for young men and women to be administered prior to the onset of sexual activity. Risk assessment with education and counseling, consistent use of primary prevention methods (such as condoms), pre- and postexposure prophylaxis or vaccination for preventable STIs, screening for asymptomatically infected persons, expedited partner therapy, effective diagnosis/treatment/and follow-up of infected persons, mutual sexual monogamy, and abstinence are advocated as public health measures for STI prevention. However, abstinence-only education results in more unintended pregnancies than holistic approaches to sexual education, including education on contraception. Therefore, we encourage clinicians to engage in open and supportive discussions with patients surrounding individual sexual practices, risks, and preferences in family planning.

PELVIC PAIN

CPP is an emerging diagnosis within women’s health care, with evolving treatment approaches. CPP is pain lasting for more than 6 months and is perceived in the pelvis or lower abdomen. It is associated with central sensitization symptoms such as multiple pain sites, sleep disturbance, anxiety, depression, rumination, catastrophizing, hyperalgesia, allodynia, or failure to respond to treatment. It is two times more common in women than men and affects approximately 15% in the United States. In women with CPP, etiologies are thought to be related to lower urinary tract, sexual, bowel, myofascial, or gynecologic dysfunction, and diagnoses include interstitial cystitis, irritable bowel syndrome, pelvic congestion syndrome, pelvic floor dysfunction, endometriosis, pudendal neuropathy, and vulvodynia.

The pathophysiologic mechanisms for CPP are complex. Theories include viscero-viscero cross-sensitization (in which pathology or injury in one organ, such as the uterus, can hypersensitize another adjacent organ, such as the bowel or bladder) and viscerosomatic convergence (in which persistent visceral nociceptive stimuli lead to noxious somatic stimulation via shared neural pathways). Some common triggers for the onset of visceral nociceptive stimuli include malfunctioning pelvic floor muscles, injury, or surgery.

CPP is often associated with negative cognitive, behavioral, sexual, and emotional consequences, and multiple pelvic pain syndromes often coexist in the same patients. These patients often have compounding nonpelvic coexisting pain syndromes, including fibromyalgia, chronic low back pain, temporomandibular joint disorder, and chronic fatigue. There is a complex relationship among trauma, abuse, mental health history, social stressors, and CPP, with a controversy of cause and effect. Catastrophizing, feelings of helplessness, and posttraumatic stress disorder (PTSD) increase the intensity of CPP and affect the quality of life. Historically, treatment of pelvic pain focused on uterine or ovarian pathology and surgical intervention; this approach, however, does not provide meaningful pain relief in CPP patients. In fact, data demonstrate that 50% to 90% of the women evaluated in CPP specialty centers have pain originating from musculoskeletal structures and not pelvic organs themselves. Treatment of CPP can include pharmacotherapy with anticonvulsants or antidepressants (particularly SSRIs/SNRIs or tricyclic antidepressants [TCAs]) and gabapentin/pregabalin.

In some cases, when pain is cyclic, hormonal therapies (HTs) such as progestins or gonadotropin-releasing hormone agonists or antagonists may be used to suppress the menstrual cycle. Recommended lifestyle modifications include pain education, elimination of dietary triggers, stress management, mindfulness and meditation, physical therapy, and exercise. Cognitive behavioral therapy (CBT) can help patients develop pain coping strategies and should be integrated into treatment plans. Adjuvant therapies addressing psychosocial modulators of pain, anxiety, depression, sleep, fatigue, cognitive impairment, and sexual dysfunction using a multidisciplinary approach are important treatment components.

PREMENSTRUAL MOOD SYNDROMES

Women commonly experience both physical and psychological symptoms in the days or weeks prior to the onset of menses. The label to give those symptoms depends on their pattern and severity, particularly their impact on functioning. While some women's symptoms are mild, transient, and nonimpairing, others may experience more serious symptoms that qualify for one of three possible syndromes.

Premenstrual syndrome, affecting up to 80% of women, is the least severe of these three syndromes and is defined by the American College of Obstetrics and Gynecology (ACOG) as the following:

- AFFECTIVE SYMPTOMS: depression, angry outbursts, irritability, anxiety, confusion, social withdrawal
- SOMATIC SYMPTOMS: breast tenderness, abdominal bloating, headache, swelling of extremities
- At least one affective and somatic symptom in the three prior menstrual cycles during 5 days before menses.
- Symptoms resolve within 4 days of onset of menses do not recur until day 12.
- Symptoms present in two cycles by prospective recording.
- Symptoms adversely affect social or work-related activities.

By definition, women with PMS do not have symptoms at other times of the menstrual cycle. Women with Premenstrual Exacerbation of an Underlying Mood Disorder (PME), by contrast, experience premenstrual symptoms but also have psychiatric symptoms at other times. The premenstrual symptoms typically represent a worsening of the symptoms present at other times rather than a qualitative change.

Among women with psychiatric illness, more than half experience a worsening of psychiatric symptoms during the luteal phase of their menstrual cycle. Individuals with personality vulnerabilities, such as borderline personality disorder, are also likely to be at an increased risk of worsening symptoms during the luteal phase. While women can sometimes have PMDD alongside an underlying mood disorder, it is likely that most women with mood disorders and luteal phase symptoms, in fact, have PME.

Neither PMS nor PME is a recognized psychiatric disorder. PME in particular needs considerably more research, as it affects such a large proportion of women with psychiatric illness and details of both etiology and optimal treatment are unknown. For example, some research indicates that women with PME do not respond to treatment with CHC, while women with PMDD do (see below).

Premenstrual Dysphoric Disorder

PMDD was recognized as a diagnosis in 2013, with the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), and is categorized as a depressive disorder. PMDD affects 2% to 8% of women during the reproductive years and is associated with emotional and physical symptoms that occur in the week prior to menses, resolve within days following onset of menses, and result in marked functional impairment. To meet the criteria for PMDD, a woman must experience at least five symptoms, including a mood symptom, that begin during the week prior to menses and resolve following the onset of menstruation. In women with another mood disorder, PMDD should be diagnosed only if the luteal phase symptoms differ in quality (not just quantity) from follicular phase symptoms. Before a diagnosis can be made, a physical examination and thyroid studies must be completed. Given the distinct pattern of symptoms characterized by PMDD, careful prospective daily symptom charting throughout two menstrual cycles is also necessary for diagnosis. PMDD occurs much less frequently than PME or PMS. Risk factors for PMDD include a high potassium diet, metabolic syndromes, increased BMI, nicotine and alcohol use, childhood trauma, low socioeconomic status, white race as compared to Black, and family history of PMDD.

Research into PMDD comorbid with other disorders is often difficult to parse, as many studies have included both women with PMS and those with PMDD and have often not considered the concept of PME. Some research suggests that bipolar disorder comorbid with PMDD is associated with worse prognosis, including earlier age of onset of bipolar disorder, more rapid cycling, increased frequency of affective episodes, and additional psychiatric comorbidities including anxiety, PTSD, bulimia nervosa, adult attention-deficit hyperactivity disorder (ADHD), and substance use disorders.

While the pathogenesis of PMDD is not entirely understood, evidence supports both hormonal and serotonergic mediators. Absolute concentrations of hormones do not predict PMDD. Rather, the disorder appears related to increased sensitivity to hormonal fluctuations. Recent literature supports allopregnanolone’s role in the pathogenesis of PMDD. Allopregnanolone acts as a positive allosteric modulator of the $GABA_A$ receptor. In rodent and human studies, impairment in the interaction between allopregnanolone and $GABA_A$ results in increased sensitivity to stress during the luteal phase of the menstrual cycle, leading to a heightened susceptibility to affective symptoms. Paradoxically, the administration of allopregnanolone has demonstrated efficacy in managing PMDD. Neuroimaging research has provided empirical evidence of hormonal effects on emotion processing–related activity in orbitofrontal, amygdala-hippocampal, anterior cingulate cortices, and brain reward circuits in monetary reward paradigms.

First-line treatments for PMDD include exercise, adoption of a complex carbohydrate diet during the luteal phase of the menstrual cycle, cognitive-behavioral therapy (CBT), SRIs, and CHCs. Second-line treatments include an increased dose of SRI and hormonal supplementation, including estradiol patches and oral or vaginal progesterone. The third-line treatment utilizes GnRH analogs to prevent menstrual cycling with concurrent continuous hormonal replacement. A recent study comparing ulipristal acetate, a selective progesterone receptor modulator, to placebo in women with PMDD found a significant increase in remission of symptoms for the drug. Symptom relief was especially noted in the mood (as opposed to physical) symptoms, holding out the promise of a potential hormonal treatment for PMDD if these results can be replicated.

PERINATAL PSYCHIATRIC DISORDERS

Baby Blues

The baby blues is the most common constellation of mood symptoms experienced by women (occurring in up to 80% of women) in the immediate postpartum period. The baby blues include transient symptoms and rapid mood shifts, including tearfulness, irritability, anxiety, insomnia, lack of energy, loss of appetite, and the general experience of feeling overwhelmed—particularly about newborn caregiving tasks. By definition, the baby blues are transient. Onset typically occurs after the third postpartum day, after the mother has left the hospital after delivery. Symptoms typically peak by day 5 and spontaneously resolve by day 10 postpartum. While the presence of the blues is unrelated to psychiatric history, it has been estimated that 20% to 25% of women who experience symptoms of baby blues will experience baby major depression. Treatment of the baby blues should include psychoeducation, validation of the mother's experience, and careful monitoring for a worsening or prolongation of symptoms.

Perinatal Depression

As perinatal mental health has gained increasing public health and scientific attention, there is a greater appreciation that many affective disorders (not just postpartum depression [PPD]) emerge both during pregnancy and postpartum. A change in the conceptualization of PPD occurred with the revision of the time specifier in DSM-5, published 2013, in which "postpartum onset" was changed to "perinatal onset." "Perinatal onset" recognizes new symptom onset anytime in pregnancy through 1-month postpartum, based on data that many postpartum affective episodes either have their onset in or predate pregnancy. However, diagnosing women who experience *either* prenatal or postpartum symptoms with a single “perinatal depression” may muddle key distinctions in the development and response to treatment of affective episodes during the perinatal period. Pregnancy and postpartum are biologically distinct states, and symptom onset during these two states may represent biologically distinct entities. For women without a history of depression, the onset of symptoms is most common during the first month postpartum and increases the risk for an eventual diagnosis of bipolar disorder. Depressive episodes that occur only postpartum are more likely to be associated with obsessive-compulsive and psychotic symptoms. Onset during pregnancy is more likely in women with a history of depression. It remains unclear whether depressive symptoms with prenatal versus postpartum onset demonstrate distinct responses to treatment. It will serve physicians and investigators to take note of the timing of symptom onset to capture potentially important distinctions between affective illnesses that frequently occur during the perinatal period.

Although pregnancy was once considered a time of mental wellness, minor or major depression affects up to 25% of pregnant women, similar to nonpregnant women. Postpartum depressive syndromes affect 15% of mothers, with rates that triple among inner-city and adolescent mothers. Completed suicide is a leading cause of death in new mothers. Infanticide, more commonly associated with postpartum psychosis (PPP) than PPD, is another devastating outcome that has brought the field of postpartum mental illness frequent media attention. Several studies have examined the course of unipolar depression in pregnancy. In a specialized perinatal psychiatry clinic, the rate of depression relapse was 68% in women who discontinued antidepressants prior to pregnancy, compared to 26% in women who maintained their medication during pregnancy.

In contrast, in a community sample with less severe illness, only 16% of the cohort stopping their medication in early pregnancy experienced a recurrence. In both studies, 50% to 75% of relapses occurred in the first trimester, which, in addition to considerable literature supporting the compatibility of antidepressants with pregnancy, further argues against delaying medication until the second trimester. Also, both studies reported that the duration of the depressive episode, a history of more recurrences, and a major depressive episode in the 6 months prior to pregnancy were associated with an increased likelihood of pregnancy relapse. A recent meta-analysis found that women with a history of severe and/or recurrent depression demonstrated more than a twofold risk of depression relapse during pregnancy with discontinuation of their psychotropic, compared to women who continued medications throughout pregnancy. Medication reduction without discontinuation resulted in a 35% risk of relapse, supporting a dose-response. A randomized control trial found that Preventative Cognitive Therapy (PCT) was ineffective in preventing relapse of depression among women who gradually tapered pharmacologic treatment during pregnancy.

While the psychosocial adjustment to motherhood lasts longer than 1 month, the DSM-5 time specifier ends at 1-month postpartum to denote a syndrome with the greatest biologic link to the hormonal shifts of pregnancy and delivery. Heightened biologic sensitivity for PPD at 1 month comes from familial studies, wherein PPD onset within 30 days was concordant among parous sisters, and postpartum mood symptoms within 30 days was associated with strong genome-wide linkage signals. From linkage studies of large obstetric and psychiatric databases based upon hospitalization and outpatient contacts, there was a two- to threefold increased risk of PPD in primiparous women evenly distributed across the first 5 months postpartum, with risk remaining elevated out to 2 years postpartum in one study.

Postpartum episodes are sometimes characterized by obsessive-compulsive, particularly intrusive violent thoughts, psychotic symptoms, and anxiety. See additional details described in the OCD section below. Risk factors for PPD include a personal or family history of depression and a personal history of previous depressions related to reproductive events (such as PMDD or previous PPDs). As with non-PPDs, women facing multiple or severe psychosocial problems, trauma, or chronic interpersonal difficulties are at increased risk of experiencing a major depressive episode during the postpartum period. One study revealed a 50% rate of childhood sexual abuse history in women hospitalized for PPD. Current theories surrounding the biologic etiology of PPD are addressed in the Mechanisms section later in the text.

A subgroup of women experiences late-onset PPD when weaning from breastfeeding. A recent review of current literature supports the hypothesis that fluctuation and dysregulation of allopregnanolone, an agent known to act on inhibitory GABA neurons, likely plays a role in late-onset PPD. Further research may serve to identify markers to identify a subgroup of women susceptible to the onset of a mood episode at the time of weaning.

Evidence-Based Treatments for Postpartum Depression. A few placebo-controlled trials for the treatment of PPD have been conducted. The SSRI fluoxetine was compared with CBT, combination treatment, and placebo. Fluoxetine and CBT were similarly effective, and fluoxetine was significantly more effective than placebo. Sertraline failed to significantly differ from placebo as an add-on treatment to brief dynamic psychotherapy. In a pragmatic trial of SSRI versus supportive care listening visits, the SSRI proved superior by 4 weeks of treatment. Paroxetine was associated with greater remission rates than placebo in an 8-week trial. For PPD with comorbid anxiety, antidepressants and CBT were efficacious; however, there was no added benefit of combining modalities. In small, open trials, sertraline, venlafaxine, escitalopram, paroxetine, and bupropion SR were effective for the treatment of PPD. In studies of the prevention of PPD, the TCA nortriptyline did not confer protection compared with placebo, but a randomized pilot trial of sertraline given immediately after birth for 17 weeks had greater benefit than placebo (93% vs. 50% remained healthy, respectively) in conferring preventive efficacy.

PPD treatment or prophylaxis with the reproductive hormones estrogen and progesterone has been examined as a means of compensating for the abrupt hormonal withdrawal that accompanies delivery. Numerous studies have attempted to parse out the contributing roles of estrogen and progesterone in the development of

PPD. In an RCT of transdermal 17β-estradiol 200 mg per day versus placebo, the estradiol-treated group had a significant reduction in depression scores beginning 4 weeks after treatment initiation that was maintained at 3 months. An open trial of sublingual estradiol (mean dose of 4.8 mg/day) in women with severe depression resulted in rapid response within 1 week and an 83% remission rate by 2 weeks of treatment. However, a more recent double-blind, placebo-controlled trial including 12 women with PPD demonstrated no significant improvement with transdermal estradiol administration.

For new-onset depression, an SSRI may be tried initially because such agents are well tolerated and are associated with a low risk of toxic effects in patients taking an overdose. However, if the patient previously had a positive response to a specific drug from any class of antidepressants, resumption of the prior agent should be strongly considered. Women may be more likely to have a response to serotonergic agents, such as the SSRIs and venlafaxine, than to non-serotonergic TCAs, yet safety and efficacy data for TCAs such as nortriptyline and desipramine are generally good, given that they have been in use for a relatively long time.

Slow increases in the dose help manage side effects. Women who have recently given birth are often sensitive to medication side effects, possibly related to elevated anxiety, which is common early in the postpartum course. Treatment should be initiated at half of the recommended starting dose (e.g., 25 mg of sertraline per day or 10 mg of paroxetine per day) for 4 days, and small increments should increase doses (e.g., 25 mg of sertraline per week or 10 mg of paroxetine per week) as tolerated until full remission is achieved. If the patient responds to an initial trial of medication lasting 6 to 8 weeks, the same dose should be continued for a minimum of 6 months after a full remission has been achieved to prevent a relapse. If there is no improvement after 6 weeks of drug therapy or if the patient has a response but then has a relapse, a medication change (within-class or different class) is recommended. Drug doses may need to be increased in the third trimester during the period of increased hepatic metabolism and expansion of blood volume. Antidepressant trials are often initiated in primary care and obstetrician-gynecology offices, and consultation with psychiatrists can facilitate optimal dose titration to achieve remission, address treatment resistance, and treat psychiatric comorbidities. Long-term maintenance treatment to prevent recurrence should be considered for women who have had three or more episodes of severe depression.

BREXANOLONE AND RELATED COMPOUNDS. The utility of compounds derived from progesterone in the treatment of PPD has gained momentum in recent years, culminating in FDA approval of a novel agent for the treatment of PPD, brexanolone, in March 2019. Brexanolone is structurally identical to allopregnanolone, an endogenous neuroactive steroid and a progesterone metabolite. Three randomized, double-blind, placebo-controlled studies demonstrated significant improvement in depression following a 60-hour continuous infusion of brexanolone. While there is significant ongoing enthusiasm for brexanolone, and it is now available at sites across the country, it remains expensive and time prohibitive for many, costing $34,000 per infusion and requiring at least a 2.5-day inpatient admission, and many have noted the strong placebo response of the clinical trials. The current understanding of brexanolone's mechanism of action is discussed in the Mechanisms section later in the text. A related oral compound, zuranolone, was granted FDA approval for the treatment of PPD in August 2023. A double-blind RCT including 151 women meeting criteria for PPD demonstrated significant improvement in depressive symptoms at day 3, with persistence to day 45, among those receiving zuranolone compared to the placebo group.

NONPHARMACOLOGIC TREATMENTS FOR POSTPARTUM DEPRESSION. Complementary and alternative medicine therapies suggest potential benefits as monotherapy or in conjunction with pharmacologic treatments. Psychosocial interventions for individuals and groups, delivered in-home, clinic, or by phone, compared to usual care, are associated with symptom reduction, reduced psychosocial stress, increased mental wellness, and prevention of PPD in an increasing number of studies. Promising modalities include brief interpersonal psychotherapy, CBT, mindfulness, computerized resources, in-home nurse visits, and peer-based telephone support. Light therapy and supplements may also provide relief. Five-week morning 7,000 lx bright light therapy was associated with a significant improvement in depressive symptoms in an RCT compared to red, dim light. A meta-analysis of 18 RCTs examining the use of omega-3 polyunsaturated fatty acids (PUFAs) in treating perinatal depression found a small beneficial effect. The result was of statistical significance, but the response appeared to be so small and inconsistent that the authors advised against prescribing omega-3 PUFAs to prevent or treat mood episodes during pregnancy. However, it might be beneficial as an adjunct therapy in treating PPD. There is strong evidence for the effectiveness of cognitive-behavioral and interpersonal therapy in managing PPD. Research is too limited to draw conclusions surrounding the effectiveness of exercise, yoga, massage, acupuncture, light therapy, and supplements.

Postpartum Psychosis

The most severe perinatal mood disorder, PPP, is rare, affecting approximately 0.1% to 0.2% (1 to 2 in 1,000) of postpartum women. The disorder was recognized as early as 400 BC by Hippocrates. PPP typically presents within the first few weeks following delivery. Depersonalization, delirium, agitation and insomnia, and delusions are characteristic of the disorder. Affective symptoms—including depression, irritability, anxiety, mixed episodes, and, most often, mania—are also present, and are an important reminder that the syndrome is an affective psychosis. (A woman with an underlying primary psychotic illness who experiences an exacerbation in the postpartum is NOT the same as a woman with PPP, see below for nonaffective psychoses.) A clinical cohort study assessing the phenotype of PPP found abnormal thought content among 75% of cases, disorganization among 25%, disturbed consciousness in 10%, and catatonia in 5%. It is important to recognize that the clinical presentation of PPP can include waxing and waning of consciousness, similar to delirium; therefore, it is imperative to obtain collateral information about symptoms as a presentation at any moment may not reveal the full extent of symptoms.

Women with pre-existing bipolar disorder are at very high risk of PPP. In addition, women with new-onset PPP are eventually diagnosed with bipolar disorder in 50% to 80% of cases. PPP is a medical emergency because of the potential risk for infanticide or suicide, and therefore imminent referral to a psychiatrist, hospitalization, and initiation of lithium is typically indicated. PPP has been associated with a 70-fold increased risk of suicide in the first 12-month postpartum, with peak rates within a few days to a few weeks after discharge. Although being a mother is protective against suicide, this narrow time interval postpartum for very ill women constitutes an exception and warrants greater vigilance. Infanticide risk is also greatly increased in this patient population. In a classification of maternal filicides, 24% occurred within the context of acute maternal psychosis and another 56% due to altruistic beliefs about alleviating suffering in the infant, which was often accompanied by maternal suicide. If recognized early and treated aggressively, PPP can have a good prognosis. Lithium is the gold standard for managing the disorder, with 98% of patients in one longitudinal study responding to lithium treatment. Benzodiazepines and antipsychotics can be useful for symptomatic management but should not take the place of lithium.

Bipolar Disorder

Although an important proportion of PPP cases will be attributable to bipolar disorder, it is critically important to recognize that the mothers with bipolar illness also have a higher risk of experiencing a nonpsychotic, nonmanic episode in the perinatal period. A birth register study found a relative risk of 37.2 for relapse in the postpartum period for women with bipolar disorder (approximately 10-fold greater than the risk for women with other psychiatric disorders). While some of these women will have PPP, most will have depressive episodes. Episodes were notable for short postpartum latency and increased postpartum duration relative to pregnancy incidence. Mothers with bipolar II disorder demonstrated more widely distributed timing of mood episodes, both during pregnancy and at 1-month postpartum. Clinicians should maintain a high suspicion for bipolarity, given some indication that this diagnosis might be missed in as many as 50% of cases presenting with unipolar PPD.

Pharmacologic treatments are the first-line therapy for the treatment of bipolar disorder; however, close clinical monitoring and efforts to enlist the aid of social supports to prevent sleep disruption and stabilize social rhythms in the new mother with a history of bipolar disorder may be particularly important. Illness recurrence occurs in most women who discontinue psychotropic management, and women are particularly vulnerable to illness during the postpartum period. In a retrospective cohort study including 101 women with bipolar disorder (some followed through pregnancy and postpartum, others nonperinatal), the risk of recurrence after discontinuing mood stabilizer therapy in pregnant women was similar to that in nonperinatal women. In contrast, postpartum recurrences were 2.9 times more frequent than those in nonperinatal women. Among the perinatal women, 85% of women who discontinued mood stabilizer therapy relapsed, while only 33% of those who continued mood stabilizer treatment relapsed (all those relapses occurred postpartum).

Lithium prophylaxis in late pregnancy or 24 hours postbirth has been associated with a nearly 10-fold decrease in the rate of postpartum illness recurrence. Sustained remission out to 9 months postpartum was optimized with lithium maintenance therapy. For unresponsive illnesses, benzodiazepines, antipsychotics, and ECT also have an important role. A low-dose, mid-potency antipsychotic, such as perphenazine, can be added as a PRN agent for prodromal or breakthrough symptoms.

Nonaffective Psychosis

This category includes diagnoses of schizophrenia, delusional disorder, and brief and unspecified psychosis. While first-generation antipsychotics (FGAs) are associated with diminished fertility rate via hyperprolactinemia and anovulation, second-generation antipsychotics (SGAs) are associated with normal fertility rates, and there is a high rate of unwanted pregnancies and repeated VTOPs in this population. These are often high-risk pregnancies due to low rates of prenatal care, high rates of poverty, homelessness, poor social support, poor nutrition, and high use of prescribed medications, alcohol, drugs, and tobacco. These women also have increased rates of sexual abuse prepregnancy and physical abuse in pregnancy. Often these mothers are less aware of signs of labor and delivery and are at greater risk of obstetric complications. Women with nonaffective psychosis represent the minority (15%) of hospitalizations of perinatal women due to psychosis, which are largely accounted for by affective psychosis. In women with schizophrenia, the relative risk of hospitalization in the first 30 days postpartum compared to 6 months postpartum is 5.7. Common delusional themes include ideas that the birth did not occur or that the baby is dead or defective; there can also be command hallucinations to harm the baby. Fifty percent of these mothers are separated from infants at discharge from the obstetric hospital due to incapacity to provide for the child's minimal needs for physical sustenance. There is a 75% puerperal recurrence rate in schizophrenia and a 100% readmission rate 10-year post index postpartum admission due to schizophrenia relapse.

This is a diagnostic group for whom severe and persistent mental illness is convincing for medication needs. Psychotic illness confers the greatest increase in the risk of poor fetal outcomes. Evidence from standard guidelines is currently used to guide treatment with no knowledge that treatment efficacy differs between postpartum and nonpostpartum populations. Although there is a tendency for physicians to discontinue or avoid initiating long-acting injectable antipsychotics (LAIs) during pregnancy and postpartum, given concern for the possible risk in exposing a fetus to such agents, the utility of these medications is supported by the literature. There are no clear contraindications to pregnancy identified, and there is a role for the judicious use of LAIs in pregnant women, especially for women at risk of relapse or where there is a concern for medication adherence. Women with psychiatric illness who discontinue antipsychotics during the first trimester of pregnancy are twice as likely to experience a return of their illness compared to those who continue on medication.

Infants of women with untreated psychosis are more likely to have congenital malformations than healthy women. As it is not feasible to conduct RCTs assessing the risks of LAI exposure to the fetus during pregnancy, we must rely on limited information, largely drawn from case studies. Based upon case reports, LAIs appear to be well tolerated by women and offspring during pregnancy, and rates of premature birth, spontaneous abortion, and congenital anomalies do not appear to differ from the general population. Theoretically, LAIs likely carry risks similar to orally administered antipsychotics, for which data are also limited. Prescribers should carefully weigh the known and unknown risks of the medication against the substantial risks associated with psychiatric decompensation for both the woman and baby. Adjunct psychosocial treatments include individual cognitive treatments, family support, and rehabilitation programs to facilitate mother-infant caregiving and bonding. Close collaboration among health professionals is necessary during the perinatal period in these women.

Perinatal Anxiety

Anxiety disorders are the most prevalent of psychiatric disorders, and women, compared to men, have a greater overall lifetime prevalence of anxiety disorders and increased peak prevalence during the childbearing years. Thus, both gynecologists and primary care practitioners need to have a basic familiarity with anxiety disorders to promote early detection and treatment. A recent systematic review and meta-analysis found prevalence rates for self-reported anxiety symptoms ranging from 18.2% to 24.6% during pregnancy and 15% postpartum. Rates of diagnosed disorders are lower (15.2% and 9.9%, respectively). Rates of anxiety comorbidity in women with perinatal depression reach 50%. Medical diagnoses such as hypo/hyperthyroidism, anemia, and hypertension have a similar phenotype and should be ruled out.

Prenatal pregnancy-specific anxiety (along with the less clearly defined stress) has been associated with increased glucocorticoid concentrations in the mother–fetus dyad, pregnancy complications, earlier delivery timing, prematurity, and low birth weight. Risk factors in pregnancy include young adulthood, single motherhood, pregnancy complications, stressful life events, poor physical health, and past year victimization. Therefore, assertive screening in pregnancy is warranted, and affective symptoms shouldn’t be assumed to be normative responses to pregnancy. Both antenatal anxiety and depression are highly predictive of PPD, which is important in framing the decision-making process for treatment choices in pregnancy.

Perinatal Obsessive-Compulsive Disorder (OCD)

Subsyndromal obsessive-compulsive disorder (OCD) symptoms may reach a 25% prevalence in perinatal women. A recent longitudinal study followed 763 women from the third trimester of pregnancy to 9 months postpartum and found 7.8% of women to meet criteria for OCD during pregnancy and 16.9% of women to meet criteria during the postpartum period. Prevalence slowly increased throughout pregnancy and into the postpartum period, peaking at 2 months postpartum. OCD has a prevalence of 1% to 1.5% in the general population. It has been hypothesized that the symptoms of postpartum OCD may be an exaggeration of adaptive maternal behavior. Postpartum OCD is more likely than OCD occurring outside the postpartum period to be characterized by obsessions involving contamination, aggression, and compulsions around cleaning and checking. More specifically, women with postpartum OCD commonly experience unwanted, recurrent, and intrusive thoughts and images of unintentionally or intentionally harming the infant. Ego-dystonic aggressive obsessions toward the infant, usually regarded as highly distressing by new mothers, must be distinguished from ego-syntonic aggressive ideation, which may accompany psychosis. However, these distressing symptoms are frequently missed by practitioners at rates as high as 70%, and it is therefore imperative that clinicians ask specifically about intrusive violent thoughts and images. Subsyndromal aggressive obsessions are significantly more common in postpartum compared with nonpostpartum women. OCD and depression are highly comorbid in postpartum women, with 40% of women with OCD also meeting the criteria for MDD. Risk factors for postpartum OCD include primiparity, history of psychiatric illness, obstetric complications, and sleep deprivation. SRIs are the first-line in the treatment of OCD. CBT, in addition to pharmacologic management, is helpful for many women. As doubt of self and guilt surrounding symptoms are common features of perinatal OCD, a strong therapeutic alliance with the treating clinician is important.

Evidenced-Based Treatment for Anxiety and Related Disorders. Standard guidelines for treating GAD, PTSD, and panic disorder should inform the treatment of postpartum women with these disorders, using specific psychotherapeutic interventions when available. Medication management described in standard guidelines includes SSRIs as a first-line treatment. There is also considerable evidence for the benefit of TCAs, such as clomipramine for OCD and imipramine for panic disorder. Nightmares of PTSD can be effectively treated with the alpha-1 adrenergic receptor antagonist prazosin and could benefit this population. The use of benzodiazepines to promote sleep and/or treat residual anxiety symptoms not addressed by the primary behavioral or antidepressant medication is a common clinical practice.

Nonpharmacologic Treatments for Postpartum Anxiety and Related Disorders. An array of nonpharmacologic treatments, including yoga, tai chi, self-guided workbook with weekly telephone support, relaxation therapies, acupuncture, exercise, and massage, have demonstrated promise in reducing antenatal anxiety in several studies with variable degrees of rigor. Prevention of postpartum anxiety has been successful in prenatal educational formats that deliver CBT for postpartum OCD symptom reduction and parenting communication for reducing postpartum generalized anxiety and improving overall postpartum adjustment and marital satisfaction. However, results remain mixed, as there was no benefit seen with 12-session group CBT when added to paroxetine in women with comorbid depression and anxiety. For women who meet full criteria for OCD, CBT with Exposure and Response Prevention (ERP) is the gold standard, as it is for nonperinatal women.

Theorized Mechanisms in the Pathogenesis of Perinatal Psychiatric Illness

Scientific investigation is beginning to clarify the pathogenesis of PPD and, to a limited extent, postpartum psychosis. Little research has been conducted on the etiology of other perinatal psychiatric disorders. Therefore, this section presents the evidence pertinent to our current understanding of the pathophysiology of PPD and offers a brief review of the literature investigating the etiology of postpartum psychosis. Future research will be necessary to inform discussions on the pathophysiology of other psychiatric perinatal disorders.

What constitutes the “hormonal trigger” prompting postpartum depressive symptoms has been a topic of significant scientific curiosity. Well appreciated are the hormonal excursions of the perinatal period, whereby late gestational increased concentrations of CRH (1,000-fold), cortisol (3-fold), estradiol (100-fold), and progesterone (10-fold) are followed by different rates of decline after delivery. Postpartum withdrawal of estradiol and progesterone, which occurs rapidly within the first 2 to 7 days postpartum, when studied under simulated conditions, induced depressive relapse in women with a history of PPD but not in women without such history. Other experimental simulations of this kind indicate that reproductive hormonal fluctuations are associated with disruptions in CNS corticolimbic connectivity that promotes healthy emotion processing. While studies have found no relationship between estradiol and PPD, numerous recent studies have implicated the dysregulation of allopregnanolone, a progesterone metabolite, in the development of PPD. Lower serum levels of allopregnanolone and reduced sensitivity of allopregnanolone to stress during pregnancy have been associated with depressive episodes during pregnancy and postpartum.

Other hormones may additionally contribute to the pathogenesis of PPD. The hormones CRH, cortisol, oxytocin, and prolactin, which fluctuate across the perinatal period and demonstrate interindividual differences, may confer differential risk for mood disorders and are implicated in stress response, emotion processing, and social cognition functions. Although there are no peripheral hormone biomarkers for perinatal mood disorders, gestationally elevated CRH and reduced oxytocin, as well as prolonged blunting of the postpartum HPA axis, have been associated with perinatal depression in as yet unreplicated studies. Subclinical hyperthyroidism poses a risk for antenatal depression, whereas thyroid hypofunction or increased antibodies was associated with PPD.

The immune system is another likely contributor to the development of PPD. The literature supports an association between immune dysfunction and MDD; additional research is necessary to elucidate the mechanisms and population implicated in the association. Given that pregnancy and postpartum are periods of significant immune dysregulation and that autoimmune disorders are associated with MDD in the general population, it would be reasonable to consider a role for dysregulation of the immune system in the pathogenesis of PPD. Early studies attempting to clarify the role of the immune system in PPD focused on the activity of proinflammatory cytokines, with mixed results, finding both elevated and reduced levels of proinflammatory cytokines to be associated with depressive episodes. The immune system's role in the development of depressive illness is an ongoing and active field of research; numerous groups are working to examine various potential regulators, including acute phase reactants, B- and T-cells, microglia, toll-like receptors, immune cell populations, and extracellular vesicles.

Functional MRI of healthy mothers in the early puerperium as they attend to visual and auditory cues of their baby reveal activation of brain regions involved in reward (striatum, midbrain, orbitofrontal cortex), empathy (superior temporal sulcus), emotional appraisal (insula and amygdala), and emotion–cognition integration centers (anterior cingulate gyrus) with transitions to the maternal role. Depressed mothers activate this circuitry to a lower extent. Genetic mechanisms research assumes a general vulnerability for mood disorders and a specific hormonal sensitivity. PPD risk increases in mothers with a personal and/or family history of depression. Within 30 days of delivery, postpartum mood disorder was concordant among parous sisters and was associated with strong genome-wide linkage signals. Several monoamine and steroid hormone gene polymorphisms and hormone-sensitive epigenetic signals have been described, all of which require further investigation.

Psychological and sleep-related mechanisms have a critical place in understanding postpartum disorders. Maternal sleep deprivation is a common, if not universal, consequence of the infant's lack of a circadian rest-activity cycle and a significant risk factor for developing PPD. Therefore, therapies supporting sleep are critical to psychiatric care in the postpartum period. The average time an infant develops a recognizable nocturnal sleep pattern is 12 weeks. Currently, infants display a clear nychthemeral pattern of melatonin secretion, indicating maturation of the master clock, the suprachiasmatic nucleus. Becoming a mother represents an unparalleled identity transition and a disruption of life's routine, including relationships, work, sleep, exercise, nutrition, and finances. High social stress, marital discord, minority status, adolescence, and neuroticism are all associated with increased perinatal mood disorders. Women with a history of attachment insecurity or trauma might have particular struggles when faced with an infant to protect from abuse or neglect.

The current understanding of the pathogenesis of PPP is limited. Like the suspected etiology of PPD, increased sensitivity to hormonal fluctuations around the time of delivery likely plays a role in developing PPP. Disruption in the physiologic reconstitution of the immune system following parturition is another plausible explanation. Immune dysfunction is more common among women with PPP; women with PPP are more likely to experience co-occurring autoimmune thyroiditis, encephalitis, and pre-eclampsia than those without PPP. In fact, a study including 23 women with PPP and 117 unaffected controls matched by primiparity demonstrated significantly elevated rates of both autoimmune thyroid disease and clinical thyroid failure. Women with PPP experienced thyroid disease at a rate of 19%, while only 5% of women in the control group demonstrated disease.

Furthermore, women with PPP progressed to thyroid failure at a rate of 67%, as compared to 20% of controls. Therefore, it is recommended that clinicians test for serum thyroperoxidase antibodies among women with PPP. The above observations surrounding the elevated rates of an array of autoimmune disorders among women with PPP have prompted an early investigation into possible mechanisms underlying immune dysfunction leading to PPP. Differences in inflammatory cells between women with and without PPP have also been observed. Elevated levels of monocytes, lower naïve CD4 and CD8 count but higher activated CD8 and memory T cells and decreased natural killer cells have been observed in women with PPP, compared to healthy controls. A recently postulated theory implicates dysfunction in immune-mediated myelination in the pathogenesis of PPP. More specifically, impairment in regulatory T cells and CCN proteins involved in CNS remyelination has been proposed. Additional investigation is necessary to explain the mechanisms involved in the pathogenesis of PPP and to develop more targeted treatments for this rare but highly morbid disorder.

Breastfeeding. Breastfeeding is the continuation of the reproductive cycle, both through providing neonatal and early childhood nourishment and through the normalization of maternal pregnancy physiology. In the United States in 2014, 83% of infants were breastfed at any point, 55.3% were breastfeeding at 6 months, and 33.7% were breastfeeding at 1 year. Breastfeeding is associated with maternal benefits such as decreased risks of breast and ovarian cancer, diabetes, and hypertensive heart disease, and infant benefits such as timely neuronal development, passive immunity, and reduction of food allergies. There are additional psychological benefits for both mother and infant. Optimally, breastfeeding should be exclusive for 6 months, and continue as complementary feeding for 1 to 2 years.

Lactation occurs because of a complex psychoneuroendocrine cascade triggered by the abrupt decline in estrogen and progesterone concentrations at parturition. Lactation becomes established in response to the neurologic stimulation transmitted by suckling. The composition and amount of breast milk change as the infant grows. In mother–infant dyads with breastfeeding difficulties, steps should be taken toward managing challenges to preserve or improve maternal mental health. When milk production or latch are suboptimal, feeding plans should be simplified to promote bonding. If demands of nighttime breastfeeding and sleep disturbance trigger mood changes, another caregiver should provide infant feedings to allow for 5 to 6 hours of uninterrupted sleep. A mother's perception of insufficient or inadequate ability to breastfeed can undermine confidence in her ability to meet her infant's needs and leave her vulnerable to symptoms of PPD. Cluster-feeding (several short feeds close together), constantly feeding for several hours, or a fussy infant awake all night should be evaluated for feeding difficulties to prevent extreme fatigue in the mother and allow for optimal infant nutrition and maternal rest. If for some reason, however, the mother cannot breastfeed, reassurance that the child can be well cared for and adequately fed with current formulas should be provided to minimize the attendant guilt that may occur. PPD and breastfeeding difficulties can often co-present. There is a theoretical argument that depression and lactation difficulties have a common neuroendocrinologic basis, perhaps mediated by disrupted oxytocin, prolactin, or stress neuroendocrine circuitries. Studies evaluating the protective or aversive effects of breastfeeding on mental illness have demonstrated inconsistent results. A recent cross-sectional study found breastfeeding to be protective against the development of PPD, with a dose–response relationship. The study attempted to control for many likely confounders (including demographics, depression severity, desire for pregnancy, reproductive and abuse history, birth complications, and substance use). Mothers with anxiety or depression history should be encouraged to obtain lactational education and support early to optimize success. The fostering of skin-to-skin contact between mother and infant during lactation has important emotion-regulatory influences for both members of the dyad and facilitates mother-infant socioemotional bonding.

Screening

Given that numerous risk factors are predictive of mental illness during the postpartum period, assertive screening in pregnancy is warranted, and affective symptoms should not be assumed to be normative responses to pregnancy. Western European countries and

Australia have been international leaders in establishing nationwide surveillance programs to detect postpartum mood disorders. Legislation at the national and state-level in the United States over the past decade reflects increased recognition of the seriousness of maternal mood disorders, with the 2010 Melanie Blocker Stokes MOTHERS Act and the recent Bringing Postpartum Depression Out of the Shadows Act, which was passed as part of the 21st Century Cures Act. The postpartum bill authorized the Secretary of Health and Human Services to provide federal grants to states for screening, assessing, and treating PPD. An increasing number of U.S. states mandate health care providers and facilities to screen newly delivered mothers for PPD. The U.S. Preventive Services Task Force (USPSTF) made a formal recommendation that all women be screened for depression during pregnancy and the postpartum period in 2019. The American Psychiatric Association (APA) also published a position statement in 2018, declaring “all pregnant and postpartum women should be assessed for both the presence of and risks for a psychiatric disorder” and continues “all perinatal patients should be evaluated for depressive, anxiety, and psychotic disorders throughout the pregnancy and postpartum period.” The statement specifies that screening be conducted with “a validated screening tool twice during pregnancy, once in early pregnancy for pre-existing psychiatric disorders and once later in the pregnancy,” and again postpartum at pediatric visits, as recommended by the American Academy of Pediatrics (AAP). The American College of Obstetricians and Gynecologists (ACOG) additionally supports screening women during pregnancy and postpartum.

As in other medical conditions, screening for perinatal psychiatric disorders has caveats and must be responsibly implemented so that cases identified can be referred to treatment promptly. Careful planning for the choice of instrument, the setting, clinical cut-off scores, and the timing of administration must be appropriate to individual settings. Screening offers little hope for symptom improvement if referrals for treatment cannot be made. The assessment of perinatal disorders can be confounded by somatic complaints common to the postpartum period, such as sleep and eating disturbances, decreased sexual interest, weight gain, and fatigue. As a result, the Edinburgh Postnatal Depression Scale (EPDS), which de-emphasizes such symptoms, has become a common instrument utilized by obstetricians, family doctors, psychiatrists, and pediatricians. The ten-question scale is available in 70 languages, has been validated cross-culturally, and is easily completed by the patient or care provider in a few minutes. Sensitivity and specificity range from 70% to 88% and studies evaluating the accuracy of the EPDS demonstrate that it identifies ill individuals more effectively than a clinician's assessment. EPDS includes a question about suicidal thoughts, which is especially important in this high-risk population. Moreover, a three-question mini-anxiety screener is embedded within the EPDS, making it possible to screen for both depressive and anxiety disorders at once. Antenatal presence of anxiety and depression is highly predictive of PPD and, therefore, important to consider in the decision-making process surrounding treatment during pregnancy.

Psychiatric Treatment Decision Making During Pregnancy and Postpartum

Structured Clinical Decision Making. Because of the temporal coincidence of mood disorders during women's reproductive years, and because pregnancy is unplanned in 50% of cases, psychiatrists must routinely ask about contraception and family planning during pharmacologic management visits. The family planning period is ideal for a structured preconceptual counseling discussion regarding perinatal treatment with prospective parents. Decision making in the treatment of pregnant women should be structured to include patient education regarding the natural history of her psychiatric disorder, the identification and review of likely risks and benefits associated both with the disorder and available somatic treatments, and frank discussion of available treatment options. This decision-making process must be tailored to address the severity of the patient's current illness and psychiatric history and the experiences, values, and concerns of the patient. Clinical decision making must balance the known and unknown risks of medication treatment with the known and unknown risks of untreated psychiatric illness.

Providers must engage patients in a comprehensive discussion of risks associated with treatment versus the considerable risks associated with an untreated mood disorder. Discussions supporting open communication with patients in a comprehensible format, respecting the patient's concerns and value structure, are critical to optimizing treatment decisions and obtaining fully informed consent for treatment. Such conversations iteratively take place. Some women will desire a withdrawal of psychotropic medications under close supervision prior to pregnancy, with iterative decisions about medication resumption over time in the event of symptom recurrence or if there is a recommendation for postpartum prophylaxis. When pregnancy is discovered during a pharmacologic treatment course, an analysis of the risks associated with medication continuation and the risks associated with illness relapse must be carefully evaluated.

Clinical decision making is complicated for pregnant and breastfeeding women because of the exponential growth in reports about psychotropic use in pregnancy that can be difficult to interpret and are often conflicting. Furthermore, clinicians in psychiatry and obstetrics may lack experience or comfort in this domain, and thus the perinatal patient may lack access to a useful decision-making process. Perinatal psychiatrists are skilled in this area and should be consulted when needed. Updated data about the risks of psychotropic use during pregnancy and lactation can be found at a variety of free websites, including the U.S. National Library of Medicine http://www.toxnet.nlm.nih.gov/Developmental and Reproductive Toxicology Database or Drugs and Lactation Database, the Organization of Teratology Information Specialists http://www.otispregnancy.org/, Thomas Hale's website on lactation http://www.infantrisk.com/, and the Massachusetts General website http://www.womensmentalhealth.org/. Subscribers to Micromedex Solutions can also access Reprotox, TERIS, and Shepard's databases. The National Curriculum in Reproductive Psychiatry (NCRP), available at https://ncrptraining.org/, is a free, online, interactive curriculum that offers up-to-date learning and teaching materials on the pathophysiology, epidemiology, diagnosis, and treatment of psychiatric disorders that occur throughout the reproductive lifespan. Notably, the U.S. Food and Drug Administration published a new Pregnancy and Lactation Labeling Rule that went into effect June 30, 2015, to improve access to a summary of all pertinent studies that bear upon perinatal use of individual medications.

This update replaces the older system of pregnancy letter categories A, B, C, D, and X, in use since 1992. Instead it "requires that labeling include a summary of the risks of using a drug during pregnancy and lactation, a discussion of the data supporting that summary, and relevant information to help health care providers make prescribing decisions and counsel women about the use of drugs during pregnancy and lactation (https://www.fda.gov/drugs/labeling-information-drug-products/pregnancy-and-lactation-labeling-drugs-final-rule)." The prior letter categorization system was oversimplified, confusing, and poorly conveyed the information available surrounding the use of medications during pregnancy. For example, sertraline, regularly used in pregnant and breastfeeding women, belongs to category C as "Animal reproduction studies have shown an adverse effect on the fetus, or there are no animal reproduction studies and no well-controlled studies in humans." When very little information is available on a drug, it gains category B classification – meaning that there is less information available than that for category C drugs. Clinicians, however, often misinterpreted this system as a simple gradation from A to and thought category B was necessarily safer than category C. These categories did little to guide clinicians and patients in navigating risk-risk discussions. Since 2018, all drugs, regardless of the date submitted to FDA, are required to use the new format.

In the absence of specialized perinatal treatment guidelines, the APA guidelines apply generally. The Centers for Disease Control "Treating for Two" initiative draws attention to the fact that, in some cases, avoiding or stopping medication use during pregnancy may be more harmful than taking a medication. The use of medication is increasingly common for women with moderate-to-severe psychiatric disorders, for whom removal of treatment would increase the likelihood of neurovegetative symptoms, incapacitating anxiety, functional impairment, drug and alcohol abuse, suicidality, or psychosis.

Risks of Psychiatric Illness During Pregnancy

IMPACT ON MOTHER. Untreated illness can lead to clinical deterioration, with concomitant psychological distress, social and occupational dysfunction, financial hardship, an inability to plan for and successfully cope with the impending life transition of motherhood, and possible suicidal ideation and self-harm. Recently acquired data from nearly 600,000 American women of reproductive age demonstrates a dramatic risk in suicidal ideation and self-harm behaviors over the last 12 years. Suicidal ideation increased from 0.1% per 100 women in 2006 to 0.5% per 100 women in 2017, and self-harm behaviors were reported by 0.1% per 100 women in 2006, subsequently doubling to 0.2% per 100 women in 2017. With many psychiatric disorders, functional impairment results in sleep disturbance, poor nutrition, lower use of prenatal vitamins, weight loss, and increased risk of cigarette, alcohol, and drug use. Poor rates of prenatal care can fail to detect gestational anemia, diabetes, hypertension, and hypothyroidism, which can harm both mother and fetus. Close attention must be paid to the host of behaviors that accompany psychiatric relapse in framing the risk-benefit discussion. As an example, cigarette smoking is associated with increased rates of miscarriage, intrauterine and perinatal death, preterm birth, and intrauterine growth restriction, outcomes that may be worse than any risk of psychotropic medications.

IMPACT ON OFFSPRING. Untreated mental illness carries significant independent risks for pregnancy outcomes. Increasingly appreciated is the impact of the prenatal milieu on fetal and infant health. The maternal HPA axis can have fetal programming effects that increase offspring risk for subsequent adult-onset medical and psychiatric disorders. Severe prenatal maternal stress during the first trimester of pregnancy has been associated with congenital anomalies. Furthermore, prenatal anxiety, pregnancy-specific anxiety, depression, and maternal stress hormones have all been associated with preterm birth, increased rates of obstetrical interventions during labor and delivery, small-for-gestational-age, low birth weight, impaired neonatal stress regulation, infant fearful temperament, adverse cognitive outcomes, and reduced gray matter brain volumes. Novel tests of fetal cardiovascular reactivity during maternally stressful tasks are associated with altered nervous system function in utero in fetuses of depressed and anxious mothers.

Fetal neurodevelopment is an active field of research. More recent neuroimaging studies demonstrate altered structure and function in infants of mothers with depression. Changes in connectivity between the amygdala and the prefrontal cortex, decreased hippocampal growth, less cortical thickening, and increased amygdala volume have been identified in women with prenatal depression.

The potential negative impact of psychiatric illness on maternal caregiving functions is well demonstrated. Mentally ill mothers have an increased likelihood of having intrusive or withdrawn interactional styles with their infants and may lack sensitivity and attunement to infant cues. These maternal behaviors have been associated with offspring vulnerability for poor self-regulation, low self-esteem, insecure attachments, behavioral problems, and motor and cognitive development delays. Ultimately, the complex multifactorial nature of the fetal environment, postbirth environment, and shared parental–infant genetic factors make unitary explanations of infant development overly simplistic. Yet, awareness of these distinct levels of infant developmental origins empowers newly pregnant women to optimize conditions for their offspring.

Nonsomatic Treatments for Perinatal Mood Disorders. Treatment options should be presented according to a hierarchy from least to most somatically invasive. Nonpharmacologic depression treatments, such as outpatient psychotherapy or mindfulness-based stress reduction, can be advantageous for mild to moderately depressed or anxious women by providing emotional support and instrumental skill development needed to facilitate coping through this life transition. Maintenance of ongoing clinic visits and psychiatric monitoring and optimizing general health and social behaviors are all important strategies that promote primary prevention or early intervention as needed, thus conferring the best prognosis and preventing the untoward consequences and refractory course that otherwise might develop.

Chronobiologic treatments were developed because depression results from a desynchronization of circadian rhythms. Treatment implications are to resynchronize the biologic rhythms of sleep and activity with those of other underlying circadian rhythms, such as melatonin, cortisol, thyroid-stimulating hormone (TSH), or prolactin. This end may be achieved by administering bright light (greater than 2,500 to 10,000 lx) or wake therapy (sleep deprivation) at critical times of the day. Although still in its experimental stages, the use of bright light or wake therapy in women with pregnancy or PPD has had promising results in pilot studies. Further studies in this area would be worthwhile, given the relatively short onset of action of these treatments and the avoidance of risks associated with pharmacologic interventions.

The benefits of social and material support that is not delivered in traditional clinics and promotes the organic routine of a mother caring for her children in her home cannot be understated for new mothers. Models for treatment delivery continue to change over the years. In the United Kingdom, Australia, and several European countries, home health care providers have had a major effect on improving recognition and management and reducing the adverse consequences of these disorders. In the United States and elsewhere internet-based intervention, health coaches by phone and support groups are increasing. For women with severe mental illness, mother-baby inpatient units (in Europe, Australia, and India) enable mothers to undergo the necessary treatment while breastfeeding and developing a relationship with their baby. In the United States, there are several mother–baby day hospitals and three Inpatient Perinatal Psychiatry Units, located at the University of North Carolina at Chapel Hill, Zucker Hillside Hospital in Queens, New York, and El Camino

Health in Mountain View, California. However, none of these options allow the baby to stay with the mother.

Safety of Psychopharmacology During Pregnancy and Lactation

MAKING SENSE OF THE LITERATURE. Although ethical considerations limit performing randomized, prospective medication studies with pregnant women, improved methods for naturalistic studies (large samples, propensity score matching) and ongoing collection of observational data in registries have increased the information available to women facing perinatal treatment decisions. Increasingly sophisticated study designs and statistical methods have been important for reducing sources of bias, such as confounding by indication, which refers to the host of demographic, behavioral, and health-related characteristics associated with the illness itself (i.e., increased smoking, poor prenatal care) or use of medication (i.e., potentially more severe illness). Both medication exposure and maternal mental illness contribute to the outcomes of interest, and thus the most informative studies can measure and control for each. A true teratogen is associated with a consistent birth defect across studies; therefore, inconsistent findings of disparate defects are harder to interpret, especially in view that the psychiatric disorder itself is not always accounted for. Additionally, well-powered studies that properly control for confounding are needed to detect rare infant malformations. The multifactorial nature of infant development suggests that large studies are needed to control for numerous contributory variables. Greater outcomes standardization and data pooling across research centers will facilitate advances in this domain.

Pharmacotherapy is an important modality for women with illnesses unresponsive to alternative treatments and for those with moderate-to-severe symptoms. The psychiatrist plays an important role in supporting the patient to make an informed decision and providing education to and fostering communication with the obstetrician and pediatrician. A guiding principle in treating pregnant women is minimizing fetal exposures both to the untreated illness and to psychotropics. The medication to which a mother has responded in the past is often the best choice, even absent substantial data, given the known risk of severe, untreated illness. Most drugs cross the placenta, albeit to a varied extent. Psychotropic transfer to the baby via lactation is much lower than placental transfer due to incomplete psychotropic absorption into breast milk and first-pass metabolism effects in the infant. If fetal exposure to a particular psychotropic has already occurred, this agent would be reasonable to continue during lactation rather than switching to one with theoretically lower exposure. If the drug was started postpartum, observing the infant's behavior before the mother is treated permits clinicians to avoid misinterpreting typical behavior as potentially drug-related. For each psychotropic class below, the available evidence is summarized in order of reproductive domain: congenital teratogenicity, late gestation effects, neonatal withdrawal/toxicity, developmental (neurobehavioral) teratogenicity, and lactation.

It is also crucial to consider the timing of making decisions about medication use in pregnancy. Ideally, pregnancy should be considered when prescribing medications for any reproductive-aged woman, regardless of her plans (as nearly 50% of pregnancies are unplanned). Ideally, a thorough conversation of the risks of untreated illness versus the risks of medication use should be conducted with each woman well before a planned pregnancy to allow time to observe the effects of any suggested modifications. Because limiting exposure to both medication and illness is the goal, women taking small doses of numerous medications might be able to pare down to fewer medications at larger doses if planning a pregnancy. Similarly, women taking a newer drug with less evidence for compatibility with pregnancy, and a history that does not include lack of response or tolerance to older drugs, might trial a drug with more evidence if planning a pregnancy. But if the same woman were to present already pregnant, the calculation is different; switching to another drug is another exposure for the fetus, and unknown efficacy for the mother may mean even more exposures down the line. Timing within the pregnancy itself also matters. While few drugs are associated with congenital teratogenicity, for those that are, it is important to remember when the period of organ formation occurs; if a pregnancy has already advanced beyond that point at presentation, there is no utility to stopping the drug (and possibly considerable potential for harm).

Finally, it is crucial to take into account pregnancy changes to metabolism and volume of distribution and their effects on serum concentration of drugs. Many clinicians are tempted to reduce medication doses in pregnancy (on the theory that this will pose less risk to the fetus). In fact, there is no evidence to support this practice, and we know that serum concentrations of many drugs will be 40% to 50% lower at the end of pregnancy—meaning that many women will require higher rather than lower doses of drugs to maintain their mental health. There are unfortunately as yet no guidelines to inform prophylactic dose changes in pregnancy, but clinicians would be wise to monitor symptoms carefully and react to relapses quickly. For those drugs that have commercially available tests for blood level (such as lithium, lamotrigine, and the TCAs), incorporating frequent levels across pregnancy can inform decisions about dose changes.

SEROTONIN REUPTAKE INHIBITORS (SRIS = SSRIS AND SNRIS). All SRIs are generally considered compatible with pregnancy. Some studies, largely those with inappropriate control groups, have demonstrated a statistically significant, though arguably not clinically relevant, relationship between SRI use and congenital teratogenicity. There was significant concern about the association of paroxetine with cardiac septal defects, resulting in the FDA revising its label to category D in 2005. However, subsequent meta-analyses failed to reproduce the relationship, challenging the conclusion that paroxetine is a cardiac teratogen. More recently, several large, methodologically rigorous studies have demonstrated no significant relationship between cardiac malformations and SRI use in the first trimester. The possible link between late pregnancy (after 20 weeks of gestation) exposure to SRIs and persistent pulmonary hypertension of the newborn (PPHN) also created public alarm and FDA label changes warning of risk in 2006, which was then rescinded in 2011. An increasing number of studies and meta-analyses draw contrasting conclusions and do not adequately control for confounders such as maternal depression severity, preterm birth, and mode of delivery. Current information, therefore, does not support SRI discontinuation due to PPHN risk.

Spontaneous abortion rates were initially estimated to be higher among women taking antidepressants. Significant limitations of this early research include lack of control for psychiatric status; lack of control for a history of prior miscarriage; and the possibility that therapeutic abortions were misrepresented as miscarriages due to stigma. Several analyses in a large cohort found no increased risk attributable to SRIs when controlling for depression status. Additionally, SRI use was not associated with an increased risk of stillbirth, neonatal death, or postneonatal death in a second well-controlled, large study.

Antidepressant exposure in late pregnancy has been significantly associated with lower gestational age and lower APGAR scores at 1 and 5 minutes, even with proper controls for psychiatric status. SRI exposure was associated with lower infant birth weight only when maternal psychiatric status was unaccounted for. Importantly, these effects are quite low and of questionable clinical significance. A 1.2- to 2-fold increased risk of preterm delivery, defined as birth prior to 37 weeks of gestation, was associated with in utero SRI exposure in a meta-analysis that controlled for illness exposure; however other characteristics of women who maintain meds through pregnancy could confer residual confounding. Depression has also been associated with preterm birth, but additional research is needed.

A neonatal adaptation syndrome affects approximately 30% of late pregnancy, SRI-exposed newborns, and substantially more preterm infants are affected. Features of poor neonatal adaptation include respiratory difficulty, hypoglycemia, hypertonia, jitteriness, vomiting, seizures, and sleep disturbances. The syndrome appears to be time-limited, with the greatest incidence in the first 48 hours postbirth but potentially extending to 2 weeks postbirth. While one study provided evidence for similar rates of neonatal signs in the absence of SRI exposure 2 weeks prior to delivery, thus implicating maternal or other neonatal factors in this syndrome, another study found increased adaptation syndrome risk even after control for maternal characteristics. Long-term sequelae of this syndrome have not been adequately studied; one study found continued differences in response to visual and auditory stimuli in SRI-exposed babies compared to unexposed at 30 days postdelivery.

Several prospective studies have found no impact of SRIs or TCAs on temperament, mood, arousability, activity level, distractibility, global intelligence quotient (IQ), or language development in children followed to 7 years of age. The concern that in utero SRI exposure is associated with a delay in motor skills and milestone achievement development has been lessened by evidence that development was still in the normative range in these studies and that no effects of in utero exposure were noted at evaluations beyond 1.5 years of age. Similarly, studies that suggest that diagnoses of autism spectrum disorders (ASDs) and ADHD as related to in utero SRI exposure inconsistently control for confounding factors such as underlying maternal diagnosis/illness severity, other exposures such as tobacco and alcohol, as well as family history of these behavioral disorders. With better control for these variables, more recent studies find no association between SRI exposure and ASD or ADHD.

Sertraline and paroxetine are the SRIs with the largest evidence base of use in breastfeeding and negligible concentrations detected in infants. Relative infant doses from breast milk are often less than 2% of adjusted maternal dose or have milk/plasma ratios less than one. Citalopram, fluoxetine, and venlafaxine are more commonly associated with infant concentrations greater than 10% of the adjusted maternal dose. In general, there are low rates of adverse effects, which are short-lived, in breast-fed infants whose mothers were treated with SRIs. With fluoxetine, in case reports, breast-fed infants of fluoxetine-treated mothers gained less weight after birth, although adverse behavioral effects were not observed. Colic was reported in three infants who were breastfed by mothers taking fluoxetine.

It is important to remember that both breastfeeding and infant metabolism change over time. The full-term neonatal cytochrome

P450 activity is approximately one-half of that found in adults. The collective data on serum levels suggest that infants older than 10 weeks are at low risk for adverse effects from antidepressants. Premature infants, and those with immature liver function, may have an increased risk of adverse effects initially. As the baby matures, those risks lessen. As the baby grows, the quantity of milk ingested will increase—but so will the intake of other food sources, leaving breast milk as a smaller proportion of nutrient intake.

In summary, there is no evidence to support stopping or failing to start SRI treatment in pregnant or lactating mothers. One caveat applies to the newer SRIs, including vortioxetine, vilazodone, and levominalcipran. Because these drugs' mechanisms of action are like other SRIs, there is no theoretical reason for risks to be different—but they have not been studied. It is, therefore, wiser to choose an older drug when possible.

TRICYCLIC ANTIDEPRESSANTS. The data show no evidence of increased incidence of congenital malformations in fetuses exposed to TCAs, although evidence is less extensive than that for the SRIs, and most reported exposures have been to clomipramine. Increased likelihood of preterm birth and low birth weight at an OR of ~1.5 mirrors the findings of SRI-exposed pregnancies, and like SRIs, there have also been reports of PPHN and cardiac septal defects related to TCA exposure. Secondary TCAs (i.e., nortriptyline, desipramine) may be preferred in pregnancy based upon fewer hypotensive and anticholinergic effects. A poor neonatal adaptation syndrome has been described with SRI exposure, and some evidence suggests it could be more severe and include anticholinergic symptoms of urinary retention and bowel obstruction. TCAs are not typically found in measurable amounts in nursing infants. Respiratory depression and sedation, the only adverse outcomes reported with any TCA, occurred in an infant whose mother took doxepin. Of the drugs in this class, nortriptyline has been studied the most as a treatment for breastfeeding women. Children exposed to TCAs both in utero and through breast milk have been followed through preschool and compared with children who were not exposed to these drugs, and no developmental problems were found. Due to increasing preferences for SRI treatment, fewer long-term studies are available following in utero TCA exposure.

MONOAMINE OXIDASE INHIBITORS. The use of monoamine oxidase inhibitors (MAOIs) is less optimal than many other agents in pregnancy for several reasons. First, studies have found growth restriction in animals receiving doses more than the maximum human recommended dose. Second, pregnancy-induced hypertension is a common complication of pregnancy that often has an insidious onset; further exacerbation of that condition using MAOIs may lead to placental hypoperfusion and have serious fetal consequences. Third, if premature labor occurs, tocolysis with β-mimetics may be obstetrically indicated but not possible because of potential interaction with the MAOI. Fourth, anesthetic management in labor may be complicated by the relative contraindication of opioids in patients taking MAOIs. If pregnancy occurs while the patient takes MAOIs, a gradual cross-taper to a more evidence-based drug is the preferred strategy. Guidelines have been suggested for appropriate obstetrical anesthesia and analgesia for women treated with MAOIs.

BUPROPION. Bupropion's dopaminergic mechanism of action and role in smoking cessation gives it an important place in the treatment armamentarium, and it is preferred by many reproductive-aged women concerned about SRI-associated weight gain. Data on bupropion are growing but still scant compared to those available for SRIs. Some small studies raised concerns about an increased risk of ventricular septal defects and left outflow tract heart defects among exposed infants, but numerous recent studies have not supported these concerns. A recent systematic review and meta-analysis including more than 5 million pregnancies reported an odds ratio of 1.23 (95% CI 1.01 to 1.49) for the use of bupropion during the first trimester and occurrence of cardiac malformations in the offspring, a statistically significant but likely not clinically significant association. The challenge of adequately powering studies to look at rare heart defects remains, and the studies to date have not consistently controlled for important confounding effects of maternal illness. A recent review on the management of ADHD during pregnancy recommends bupropion as the preferred therapy and notes that the drug appears to be compatible with pregnancy. No associations have been found between bupropion use and late gestation effects, neonatal withdrawal/toxicity, or developmental (neurobehavioral) teratogenicity, but data are limited at this point. A few small studies have assessed the degree to which bupropion passes to breastmilk. One study, conducted on two breastfed infants of mothers currently taking bupropion, was unable to detect bupropion, and its active metabolite, hydroxybupropion, in the blood of either infant. Infant exposure has been measured at 0.14% of the standard adult dose of bupropion and is therefore very small.

MIRTAZAPINE. Mirtazapine is a novel antidepressant with noradrenergic and selective serotonin effects. It can be extremely useful for women with sleep issues or nausea and vomiting of pregnancy, but there is far less evidence about the use of mirtazapine in pregnancy than is available for SRIs. The case literature showed effectiveness in rapid alleviation of hyperemesis gravidarum, based on serotonin 3 receptor blockade. Treatment benefits were obtained at doses of 7.5 to 45 mg/day, and in some cases, mirtazapine was administered initially through intravenous fluids, and the treatment course was concluded after 10 days. In a study of first-trimester mirtazapine exposure, there was no increased rate of malformations and similar rates of spontaneous abortion and preterm birth compared to pregnancies with other antidepressant exposures. The estimated infant dose through breast milk is about 1.8% to 4.4% of the maternal dose.

FIRST-GENERATION ANTIPSYCHOTICS. Although FGAs have been utilized in pregnancy for decades and are historically offered to manage hyperemesis gravidarum, our understanding remains limited to observational studies and case reports. As with any study on a gravid population, our research questions remain limited by the safety of subjects and human research ethics. The available data are reassuring, supporting that FGAs are unlikely to be major teratogens. However, an association between haloperidol exposure during the first trimester and limb abnormalities has been inconsistently demonstrated. Some studies have associated FGA use with major malformations, but the study designs failed to control for confounders. For example, in a meta-analysis of nearly 75,000 births among women treated with a low-potency antipsychotic (chlorpromazine), there was a 0.4% increased risk of malformations compared to the general population. Another systematic review found that the high-potency antipsychotic haloperidol was not associated with increased risks of malformation. Furthermore, in a meta-analysis of studies of parents with schizophrenia, the rate of malformations and perinatal death did not differ from the general population. Preterm birth and small-for-gestational-age births may be more common with FGA than SGA exposure.

Use of these agents in the second and third trimester is associated with slightly increased rates of perinatal complications, low birth weight, and transient perinatal syndromes (withdrawal symptoms, extrapyramidal symptoms, difficulty with feeding, and possible neonatal jaundice and functional bowel obstruction) that generally resolve within days but may last for up to 10 months after birth. Again, these studies failed to control for likely confounders, such as severe mental illness. Lactation studies of FGAs consistently report infant to maternal plasma concentrations at less than 10%. There may be an increased risk of gross motor skills delay among exposed infants at 6 months. However, this appears to be transient, with no difference between exposed and unexposed infants at 1 to 2 years. Continued treatment is recommended, given the high risk to the mother and offspring during a recurrence of psychosis and the minimal risks of treatment. The use of such agents as augmentation or adjunctive agents (in combination with antidepressants or mood stabilizers) in treating mood and anxiety disorders must be weighed against the severity of illness. Lower-potency FGAs have a greater risk of hypotension, while higher-potency agents confer risk for extrapyramidal symptoms. Intermediate-potency agents such as perphenazine constitute a useful compromise in this drug category.

SECOND-GENERATION ANTIPSYCHOTICS. Our understanding of the safety profile of SGA use in pregnancy has made significant strides in recent years. A recent report on data belonging to the National Pregnancy Registry for Atypical Antipsychotics confirms that SGAs are unlikely to have a major teratogenic effect, with an odds ratio comparing exposed to unexposed infants of 1.48 (95% CI, 0.625 to 3.517). A review including data on 14,000 women utilizing SGAs during pregnancy found no association between congenital malformations and the use of either aripiprazole, olanzapine, or quetiapine. However, there was a statistically significant, though of questionable clinical significance, increased risk for malformations among users of risperidone and paliperidone. Authors could not comment on the safety of ziprasidone, clozapine, amisulpride, asenapine, lurasidone, and sertindole use during pregnancy due to limited data. Another study of over 1 million Finnish women, with data collected from 1996 to 2016, demonstrated an increased risk of pregnancy complications related to glucose metabolism, including gestational diabetes, Cesarean section, large-for-gestational-age infant, and preterm birth, among women taking either an FGA or SGA. However, controls were pregnant women without any history of antipsychotic use during pregnancy. Of note, maternal mental illness was not assessed nor controlled for and therefore a likely confounder in the study. A systematic review published in 2017 supports continuing women on the minimum effective dose of SGA throughout pregnancy, recognizing that discontinuation during pregnancy results in a high risk of relapse of mental illness, with weighty repercussions for both mother and offspring.

SGAs confer less risk for extrapyramidal side effects in the mother than FGAs, and the same is likely true for neonates exposed in utero. In a study of 142 pregnancies with exposure to SGAs, the antipsychotic dose was inversely related to gestational age at birth and positively related to the likelihood of NICU admission. Higher rates of respiratory distress were reported when SGAs were combined with mood stabilizers. Intake of SGA through milk leads to 0.1% to 10% maternal dose, and most are considered compatible with lactation. (There are theoretical concerns about the effect of aripiprazole on milk production, however, and this should be considered for women who are having difficulty establishing breastfeeding.) The National Pregnancy Registry for Atypical Antipsychotics continues to accept eligible subjects to further our understanding of the benefits and risks of SGA use during pregnancy (http://womensmentalhealth.org/clinical-and-research-programs/pregnancyregistry/).

While the above demonstrates that most FGAs and SGAs are compatible with both pregnancy and breastfeeding, there is one that should be used only with caution. Clozapine has the rare but serious possible side effect of inducing fatal agranulocytosis in adults, and therefore particular caution should be demonstrated with use in pregnancy and lactation.

LITHIUM. Decisions surrounding lithium use in pregnancy and lactation should be taken with care. Women being treated with lithium are likely to have a severe illness, with potentially devastating consequences for both mother and child associated with discontinuation. This considerable known risk must be weighed against the known risks of treatment, which are less substantial than previously believed. The scientific community continues to assess the risk of Ebstein cardiac anomaly following first-trimester lithium exposure. While the risk was originally cited as 400-fold, this risk was later revised to 10- to 20-fold (0.5–1 per 1,000). Some argue the risk may be even lower due to methodologic limitations in extant studies. A recent study analyzing data from more than 1 million pregnancies found a slightly increased and dose-dependent risk of cardiac anomalies, on the order of one additional case per one hundred live births. Despite this dramatic reduction in perceived risk, fetal echocardiography is recommended to rule out the presence of cardiac malformations as a precaution. There may be other risks associated with the use of lithium in pregnancy, but thus far, there is a lack of consensus regarding gestational size increases and preterm delivery. While many clinicians are daunted by the risks posed by lithium use, it should be remembered that approximately 85% of women with bipolar disorder who discontinue mood stabilizers for pregnancy will relapse. Continuation of lithium throughout pregnancy will therefore be the choice that best minimizes risk for many patients with bipolar disorder.

Neonatal signs of lithium toxicity have been described to include muscle flaccidity, inhibition of normal neonatal reflexes, lethargy, cyanosis, and cardiovascular effects, including atrial flutter, tricuspid regurgitation, and congestive heart failure. Careful monitoring of maternal lithium concentration in pregnancy is required because of the dramatic shifts in fluid volume during pregnancy. A higher glomerular filtration rate, coupled with the increased plasma volume, often requires higher doses in the gravid woman to achieve comparable serum lithium concentrations. Lithium crosses the placenta freely, and maternal and fetal plasma concentrations are similar. Maternal serum lithium should be closely monitored as delivery approaches. Although lithium is commonly held for 24 to 48 hours prior to delivery to reduce infant exposure, there is no scientific evidence to demonstrate that this practice reduces the risk of adverse outcomes.

Furthermore, missed doses of lithium place the mother at significant risk of relapse following delivery. Because of possible dehydration during labor, rapid loss of fluid volume after delivery, and potentially elevated gestational lithium doses, reinstatement of lithium therapy in the recently delivered mother should be gradual to approximate pregravid doses. Early postpartum measurement of lithium concentrations and maternal side effects is required.

Reports of successful breastfeeding and healthy infants of well-informed and clinically stable mothers treated with lithium have shifted perceptions surrounding the use of lithium in lactating women. The mean concentration of lithium in breast milk is approximately half that of the maternal serum concentration, ranging from 20% to 70%. Lithium concentrations in the serum of breastfeeding infants were, on average, one-third of the level found in maternal serum, but with considerable variability, with infant-to-maternal serum concentration ratios ranging from 5% to 67%. However, given the large range in concentrations observed, laboratory monitoring of infant serum lithium, TSH, blood urea nitrogen, and creatinine immediately postpartum and up to 6 weeks of age is recommended, with ongoing monitoring every 2 to 3 months thereafter for the duration of breastfeeding. Long-term neurobehavioral assessments have found no difference in exposure relative to nonexposed siblings in the two studies. Recent literature suggests that lithium presents less risk to the breastfed infant than previously thought. Women who are will be well served by clinicians that engage them in personalized discussions surrounding the risks and benefits of breastfeeding while taking lithium.

ANTIEPILEPTIC DRUGS (AEDS). Our understanding of the safety and risks associated with the use of antiepileptic drugs (AEDs) during pregnancy and breastfeeding is largely drawn from neurology literature. Convergent studies from numerous AED registries associate first-trimester use of carbamazepine and valproate with neural tube defects in 0.5% to 5% and major malformation in up to 16% of exposed offspring, with valproate conferring a greater risk than carbamazepine. Valproate was also associated with greater fetal death rates than other AEDs. This finding led to the recommendation that women taking AEDs take an increased dose of daily folate (4 mg/day), thereby mitigating some of the risk of neural tube defects. Other than neural tube defects, reported malformations include heart defects, urogenital defects, and oral clefts. Oxcarbazepine, while related to and like carbamazepine, has been little studied. Experimental animal studies indicate that risks are similar to those of carbamazepine, but human data are lacking.

Lamotrigine has gained attention as a therapy for bipolar depression in pregnant women, as registry data have demonstrated no difference in rates of neural tube and cardiac malformations compared to the general population. Initial studies of comparably small population size found a small increase in the incidence of oral clefts among infants of mothers taking lamotrigine. More recent, larger studies have found no increased risk of facial malformations. A case-control study in 2016 further supported the favorable profile of lamotrigine, showing no difference in orofacial cleft or cleft palate rates among infants exposed and unexposed during pregnancy. Of note, lamotrigine carries FDA approval for the treatment of bipolar depression. However, it has not demonstrated efficacy in managing acute mania. It should also be remembered that estrogen has a major effect on glucuronidation, thus substantially increasing clearance of lamotrigine across pregnancy and likely necessitating higher doses. While there are, regrettably, no clear guidelines for psychiatry on prophylactic dose changes of lamotrigine in pregnancy, clinicians can be guided by the guidelines for neurology.

Lamotrigine, carbamazepine, and valproate all demonstrate compatibility with breastfeeding. The dose excreted in breastmilk, divided by the maternal dose, ranges from 9% to 18% for lamotrigine, 1% to 6% for valproate, and 4% to 6% for carbamazepine. Rare cases of lamotrigine-associated adverse events have been reported, including thrombocytosis and rash. Maternal levels should be monitored, and infant levels can be monitored when there is a concern for toxicity in the infant, possibly suggested by difficulty breathing, rash, poor feeding, or drowsiness. There have been no reports of adverse events clearly attributable to valproate; however, thrombocytopenia was reported in one case of a breastfed child whose mother was taking valproate, and hepatotoxicity remains a theoretical risk. While carbamazepine is generally well tolerated by the breastfed infant, some infants have demonstrated transient mild hepatic toxicity, poor feeding, and sedation associated with carbamazepine exposure. Infants should be monitored for jaundice, poor weight gain, and drowsiness. If there is a concern for elevated levels of carbamazepine, infant serum liver enzymes, blood counts, and a carbamazepine level can be assessed. When AED exposure is limited to breast milk, there is no evidence for adverse impact on IQ, verbal and nonverbal memory, or executive function by age 6.

In summary, while all antiepileptics are considered compatible with lactation, there are significant concerns about the use of carbamazepine and especially valproate in pregnancy, and they should be discontinued if possible. If used, folic acid supplementation (4 mg daily) is essential. A wise course of action is to prescribe these drugs in reproductive-aged women only when all other possible treatments have failed, thus answering in advance the question of whether it will be possible to discontinue. More information on these drugs is needed, and pregnancy registries have been established both in the United States (North American Antiepileptic Drug (AED) Pregnancy Registry: 1-888-AED-AED4 (1-888-233-2334) http://www.massgeneral.org/aed/) and in Europe (EURAP Registry (International Registry of Antiepileptic Drugs and Pregnancy): http://www.eurapinternational.org/registry/index.php).

REPRODUCTIVE SENESCENCE

The Menopausal Transition

Menopause—the cessation of ovulation and absence of menses for 1 year—is due to the depletion of responsive ovarian oocytes and generally occurs between 47 and 53 years, with an average age of 51 in the United States. Perimenopause describes the period of hormone fluctuation leading up to amenorrhea and can begin as early as ten years prior to the menopausal transition. Women frequently experience somatic and psychological symptoms during this period, and ovulation and menses typically become more erratic. On average, women experience symptoms associated with the menopausal transition for 4 to 8 years.

An effort to operationalize the stages of reproductive aging resulted in The Stages of Reproductive Aging Workshop (STRAW), a seven-stage system anchored by the final menstrual period. Menstrual cyclicity and FSH measurement are indices used to determine one's reproductive stage. However, psychiatric symptoms do not correlate with hormonal measurements, so treatment is determined based on clinical assessment alone. While the staging system may offer a helpful index for gynecologists tracking a woman's progress through perimenopause and following the menopausal transition, psychiatrists do not utilize it.

Surgical menopause occurs when the ovaries are removed. Surgical menopause differs from natural menopause in the abrupt decrease in estrogen levels and abrupt rise in FSH levels, compared to variable fluctuations in hormone levels occurring over the natural menopausal transition. In natural menopause, the ovaries remain intact and secrete androgens, including testosterone and androstenedione, which can be converted in many tissues to estrone, a weak estrogen. Elective bilateral oophorectomy at the time of hysterectomy (surgical removal of the uterus) was historically performed in women in their 40s or older undergoing hysterectomy for benign disease, with the rationale that removal reduced risks of ovarian cancer, and avoided the potential for re-operation for additional ovarian surgery. Because women in this age range were approaching natural menopause, the advantages of ovarian preservation were thought to be outweighed by the risks. This paradigm has drastically shifted in recent years, with ovarian preservation now recommended for women undergoing hysterectomy for benign disease until age 51 or even older in selected cases. This shift arose with accumulating evidence that surgical oophorectomy is associated with serious long-term health consequences, particularly in women of younger ages. In particular, elective oophorectomy was associated with increased risks of all-cause mortality, cardiovascular disease, chronic kidney disease, cognitive impairment/dementia/parkinsonism, depression, anxiety, glaucoma, sexual dysfunction, and osteoporosis. However, some of these effects were mitigated by hormone replacement with estrogen during surgical oophorectomy.

Both men and women age and experience an age-related decline in reproductive capacity, but only women experience complete gonadal cessation. Loss of reproductive capacity may present a psychological challenge to those not reconciled to the loss of fertility. However, even with gonadal failure, the availability of donor oocytes and sperm means that pregnancy can be initiated in a menopausal woman with an intact uterus who elects to pursue that option. Some women find these multiple options overwhelming. The menopausal transition is not only a marker of life stage, but it also presents biologic and psychological challenges unique to women.

Physical Symptoms of Menopause. Vasomotor symptoms, or hot flushes, are the hallmark distressing sequela of the menopausal transition and consist of sudden episodes of perspiration, flushing, chills, clamminess, anxiety, and a sensation of extreme heat in the upper body (particularly in the face, neck, and chest). Symptoms last, on average, 1 to 5 minutes and occur daily in over 85% of women, and up to 10 times daily in up to one in three women. Lifestyle alterations such as layering clothing, consuming cold beverages, lowering the surrounding environmental temperature, and avoiding consumption of alcohol and caffeine can improve these symptoms in some women. However, pharmacologic interventions are often necessary to manage symptoms and maintain quality of life. HT with estrogen alone or with progesterone is the mainstay of treating vasomotor symptoms, with various formulations including oral pills and transdermal patches, gels, or sprays. Testosterone in combination with HT has also been investigated as a treatment for vasomotor symptoms. However, symptom improvement was not demonstrated, and there were negative consequences on lipid parameters, clitoromegaly, hirsutism, and acne, and testosterone is not currently recommended for the treatment of vasomotor symptoms. Tibolone, a synthetic steroid with tissue-specific estrogen and progesterone action, is another treatment under investigation to treat vasomotor symptoms, but it is not considered first-line therapy. When there are contraindications or intolerances to HT, SSRIs, and SNRIs (one particularly well studied is paroxetine), clonidine, and gabapentin are all effective alternatives to HT for the treatment of vasomotor symptoms. Phytoestrogens (plant-derived substances with estrogenic biologic activity) found in soy and soy-related products have no demonstrated benefit in reducing vasomotor symptoms in numerous placebo-controlled trials. There is also insufficient evidence for the use of herbal remedies such as black cohosh, ginseng, St. John wort, and ginkgo biloba. Similarly, estrogen agonists and antagonists (tamoxifen, raloxifene) are ineffective in treating vasomotor symptoms.

Vaginal atrophy, which results from a hypoestrogenic state, can also lead to distressing symptoms of vaginal dryness, discharge, pruritis, and pain with intercourse. Thinning of vaginal tissue can cause bleeding and fissuring, and loss of elasticity can result in narrowing and shortening the vagina. Changes in vaginal pH and vaginal discharge ensue, with increased risks of urogenital infections and sexual dysfunction. Topical administration of estrogen in cream, ring, or tablet formulations is beneficial in reducing these symptoms and has decreased or nonclinically relevant systemic absorption. Ospemifene, an estrogen agonist/antagonist, has evidence of benefit for vaginal atrophy symptoms and is FDA approved for treating moderate-to-severe dyspareunia from menopausal vaginal atrophy. Nonestrogen water-based or silicone-based vaginal lubricants and moisturizers may also be used to treat women who wish to avoid HT.

Menopausal Status and Depression. Depression occurring around the menopausal transition was previously attributable to “empty-nest syndrome,” or what was termed “involutional melancholia.” Recent acknowledgment that women’s roles may encompass more than child-rearing has helped to change this view, and in fact, many women report an enhanced sense of well-being as they can pursue postponed goals. However, there is no question that the menopausal transition does confer a risk for depression. The perimenopausal transition is associated with a two- to fourfold increased risk of depressive symptoms and a twofold increased risk of a depressive disorder in medically and reproductively healthy women *without* prior history of major depressive disorder. These increased risks remained after correction for numerous factors, including hot flashes, poor sleep, and adverse life events, which have challenged the interpretation of earlier studies. Although there is no association between mean hormone levels and mood, one large study found that higher variability of between-visit estradiol concentrations in women relative to their nondepressed, premenopausal baseline was the strongest risk factor for the onset of a depressive disorder. This strongly argues for a mechanistic role of estradiol fluctuations in the onset of perimenopausal depression. Hot flashes were 1.3- to 2.2-fold more likely in women with perimenopausal-onset depressive symptoms and depressive disorder, a factor that argues for a shared neurophysiologic mechanism of perimenopausal depression and vasomotor symptoms.

Other risk factors for perimenopausal depression include a personal history of mood disorder, previous reproductive-related mood disturbance (such as postpartum and/or premenstrual mood disorder), and psychosocial stress.

Perimenopausal Cognitive Changes

Women commonly report cognitive changes during perimenopause; however, the symptoms are typically mild and develop gradually and are therefore difficult to distinguish from the experience of normal aging. Literature suggests that women may struggle with verbal recall and processing speed during perimenopause, though measurable changes were typically subtle and transient. A 14-year longitudinal study assessing cognitive changes among 400 women demonstrated a significant decline in immediate and delayed recall when comparing pre- and postmenopausal performance. Menopause-related hormonal changes are also correlated with cognitive performance. Limited evidence supports a role for stimulant medications in treating these symptoms.

Treatment of Menopausal Symptoms

Hormone Replacement Therapy. Several health outcomes associated with the use of HT have been studied. Generalizability is made challenging due to differences in study design in the age of initiation, duration of use, dosage, and estrogen or progesterone formulations used in these studies. The most important variable in gauging hormone effect on long-term health is likely the woman's age when hormones are initiated; women who are close to the menopausal transition may benefit from extended exposure to estrogen, whereas initiation many years remote from the menopausal transition is more deleterious than beneficial. The question of whether to initiate or continue this therapy poses a medical, financial, and philosophical challenge. The results of the Women's Health Initiative (WHI) and its secondary analysis suggested that cardiovascular risks and benefits conferred by a standard combination of an oral estrogen plus progestin (conjugated equine estrogens plus medroxyprogesterone) or a standard dose of equine estrogen preparation are dictated by the timing of their administration. Oral estrogen or estrogen-progestin protected women from cardiovascular disease if a woman was within 10 years of the menopausal transition but increased risk in women 20 or more years postmenopause. Stroke risk was increased 1.3-fold in women randomized to HRT relative to placebo and did not vary based on age at HRT use. This may be because oral estrogen promotes venous thromboembolism regardless of age. Benefits conferred by HRT in the WHI included protection from bone fracture and reduced colorectal cancer risk

However, because not all estrogens have the same molecular properties, the WHI results are not generalizable to other hormone products. In particular, evidence suggests that more physiologic preparations, such as transdermal estradiol patches, creams, and suppositories, may minimize risks while retaining benefits. A negative effect of synthetic progestins compared to progesterone has also been advanced as an explanation for some of the WHI findings and related epidemiologic studies. Although concern exists that combined oral hormone use after menopause increases the risk of breast cancer and affords minimal, if any, cardioprotection, there is also evidence of benefits. Hormone use retards bone loss and reduces osteoporosis and fracture risk, maintains urogenital integrity, stabilizes mood and other critical brain outputs, including motor abilities, and maintains body habitus and composition. Whether HT protects against dementia also remains controversial. Some experts have suggested that the increased risk of venous thromboembolism and stroke with orally delivered estrogens may have counteracted the trophic effects of estrogens on neurons and glia. It is thought that nonoral routes of delivery of estradiol may prove neuroprotective because there is no increased risk of venous thromboembolism with nonoral formulations. Although women at midlife tend to gain weight, studies have demonstrated that women who use HRT gain less weight and that the weight gained is preferentially deposited in a gynecoid (hips and thighs) rather than android (abdominal) pattern.

Of note, recent, prospectively collected population data showed that rates of dementia increase in both men and women after age 75 years but that the increase is much greater in women than in men (relative risk, 2.1). However, rates of dementia in women were reduced in proportion to years of hormone use after the menopausal transition. Ongoing estrogen use for longer than 10 years, especially when begun contemporaneously with the onset of the menopausal transition, reduced the risk of dementia in women compared with men. It is not widely appreciated that the brain converts testosterone to estradiol and that men typically have, always in their lives, more estradiol bathing their neurons and glia than women do except when women are pregnant. Because men do not experience gonadal cessation at midlife as women do, their brains continue to be protected by estradiol indefinitely.

Increasing basic research substantiates a role for estrogens in neurotrophic and neuroprotective functions. These are alternatively mediated via classical genomic functions of ERs, which, when ligand-activated, travel to the cell nucleus, where they act as transcription factors. In addition, cell membrane ERs are associated with activation of second messenger cascades and highly amplified gene transcription. SERMs target the same receptors as estrogen—ER subtypes ERα and ERβ—but differ in transcriptional ability based on unique ligand-induced conformational change of the ERs to facilitate the interaction of ERs with coactivators and consequent agonism or antagonism of ER-mediated DNA transcription.

The risks and benefits of using existing SERMS by postmenopausal women are being defined. Antagonism of ERβ may pose a risk for heightened depression and anxiety, given the demonstration of ERβ facilitatory effects on serotonergic activity, social interaction, and positive behaviors in rodent models of depression. Both ER subtypes are involved in neuroprotective functions. The relative density of each receptor subtype in a particular brain region likely underlies the ultimate action of any ER ligand. Although it was postulated that the mixed agonist/antagonist raloxifene would have adverse effects on mood and memory because it is associated with hot flushes, there are no long-term studies that have addressed this in a group of women older than age 70 years when the risk of dementia is sufficient to be able to discern effect.

Most studies find no increased risk of depressive symptoms in SERM-treated women, and some find SERM-associated amelioration of depressive symptoms. Likewise, raloxifene in doses of 60 or 120 mg was not associated with memory impairment and was recently reported to reduce the risk of mild cognitive impairment in postmenopausal women.

In contrast, tamoxifen was associated with memory impairment when added to standard chemotherapy in breast cancer survivors. Aromatase inhibitors have been shown to have the most deleterious effect on cognitive functioning, which argues for an important role for estrogen in safeguarding neural health. The effect of long-term specific SERM use on cognitive function is an area in need of further study. There is even less information about the long-term risks and benefits of partial estrogen agonists or SERMs than about more conventional estrogen preparations. The same holds for phytoestrogens derived from soy and flaxseed. Direct-to-consumer marketing of hormone preparations means that patients have more questions, concerns, and ambivalence. Although women want more information and guidance, physicians are poorly reimbursed for this type of counseling and often lack sufficient information.

Current evidence does not reveal a strong association between estrogen use and breast cancer. In particular, even in older women, oral contraceptive use is not associated with an increased risk of breast cancer. In terms of quality of life, hypoestrogenism can impair sexual enjoyment and libido and cause hot flushes, night sweats, and profound fatigue. These sequelae can generally be ameliorated or reversed with estrogen use. The trend is to use the smallest dose that relieves these symptoms.

Progestin use is generally recommended for women with a uterus to guard against the development of endometrial hyperplasia and carcinoma; whether or to what degree progestin use compromises the beneficial effects of estrogen use requires clarification. The availability of a progestin-containing IUD may minimize or obviate altogether the risks of systemic progestin exposure. The adverse effects that most limit progestin use are depression and bloating. In animals, including monkeys, progestins increase irritability and diminish sexual receptivity, and progesterone has fewer negative side effects than synthetic progestins such as medroxyprogesterone acetate (Provera). Furthermore, concern has been raised that the use of progestins other than progesterone will reverse cardioprotection by promoting vasoconstriction. Progestins may also increase the risk of breast cancer, particularly when used continuously. Women who previously experienced depression or PMS are most prone to mood disturbances with progestin use. To avoid or minimize the effects of progestin exposure, it is prudent to prescribe the smallest amount of progestin necessary to protect the endometrium from hyperplasia or use a progestin-containing IUD. Recent data support the notion that antidepressant and antidementia agents are more efficacious in women when they are estrogen replete. To achieve the best risk–benefit profile, it is worth titrating and individualizing route, dose, type, and regimen to maximize benefits and minimize adverse effects.

The literature has inconsistently supported a role for estradiol in treating depression among postmenopausal women. Women with endocrinologically confirmed perimenopause and a diagnosis of depression in a double-blind placebo-controlled trial experienced a 70% to 80% remission rate during treatment with 50 to 100 μg per day of 17β-estradiol, compared to a 22% rate of response in placebo-treated subjects. However, estradiol was equivalent to a placebo in a postmenopausal trial. Other studies have failed to demonstrate that estradiol has a significant therapeutic effect on mood symptoms during perimenopause. A longitudinal study including 291,709 female Veterans demonstrated a correlation between suicide and estradiol therapy. Women receiving estradiol showed a twofold increased risk of death by suicide compared to women not receiving estradiol. While efforts were made to control for age, race, and medical comorbidities, additional research is necessary to examine potential confounders and clarify causality.

The potential role of progesterone and its derivatives in the modulation of cognitive symptoms associated with the menopausal transition is also an active area of research. However, research has largely been limited to animal models. Inconsistent results surrounding the effects of estrogen and progesterone on affective and cognitive symptoms prompted a recent study using a rat model with induced ovarian failure. Rats were administered either estradiol, progesterone, levonorgestrel, estradiol, and progesterone, or estradiol and levonorgestrel and underwent assessments of spatial memory, anxiety-like behaviors, and depressive-like behaviors. Combined estradiol and progesterone treatments offered the most favorable outcomes in the spatial memory task, anxiety-like behaviors, and depression-like behaviors. Another study on ovariectomized rats demonstrated that rats administered exogenous progesterone performed worse on a working memory task than rats that received a control solution. The potential therapeutic applications of progesterone and its derivatives in managing depressive symptoms not exclusive to the menopausal transition have received considerable attention in recent years, as described in the Mechanisms section above.

Alternatives to HRT for Perimenopausal Symptoms. Antidepressant drugs are increasingly prescribed for hot flushes based on clinical trials that show benefit with low-dose SRIs/SNRIs (escitalopram, venlafaxine, and duloxetine, but not sertraline). It is proposed that antidepressants may improve distress tolerance of hot flushes via symptomatic improvement of coexisting mild insomnia, anxiety, dysphoria, and/or irritability. Sleep disturbance affects nearly 50% of perimenopausal women and significantly contributes to reduced quality of life. Perimenopausal insomnia is variably attributed to vasomotor symptoms, night sweats, and mood disturbance but can also be independent of those symptoms. Targeting the sleep symptoms with sleep hygiene interventions may result in sleep and quality-of-life improvement of similar magnitude to that provided by HRT.

Because the evidence does not consistently support the use of HRT for perimenopausal depression, mood symptoms are best targeted with antidepressant treatment. Clinical guidelines for antidepressant treatment in the general population apply. For women who have vasomotor symptoms and/or sleep disturbance, in addition to mood symptoms, an antidepressant that can target hot flushes, as well as mood, may be beneficial.

The utility of HRT in supporting cognition also remains questionable, but other treatments show promise. A double-blind, placebo-controlled crossover study of 16 perimenopausal women with cognitive complaints demonstrated significant improvement with atomoxetine. Participants with a history of attention-deficit/hyperactivity disorder (ADHD) were excluded from the study. Another double-blind, placebo-controlled, cross-over study of 32 perimenopausal women demonstrated improvement in delayed verbal recall and subjective memory with the administration of the stimulant lisdexamfetamine. While lisdexamfetamine was largely well tolerated in this study group of healthy women, blood pressure and heart rate increased significantly. Stimulant therapy might be appropriate for some women with impairments in executive function. However, the potential benefits of therapy must be weighed against the significant cardiac risks. An antidepressant may be helpful for cognitive symptoms in women with comorbid depression. Other medications have not shown a clear benefit in addressing cognitive symptoms associated with the menopausal transition.

Aging

Further investigations are required to distinguish the psychological and biologic effects of aging from perimenopause. Aging is associated with diminished resilience and circadian rhythmicity in men and women. Desynchronized or reduced biologic rhythms may compromise sleep, cognition, and mood. While the loss of ovarian estrogen and progesterone associated with the menopausal transition is exclusive to women, men and women experience a gradual decline in adrenal and gonadal androgen production. Androgens help to maintain libido, muscle mass, strength, and energy.

Psychological factors play a role in age-related disability. Disease and disability may present psychological challenges and engender the need for changes in lifestyle and employment. Those who successfully transition to later life can adapt to the medical, biologic, and psychosocial challenges frequently accompany aging. Stress may accelerate certain aging processes. In a longitudinal population study of the older adults, women whose cortisol levels remained elevated across the observation interval of 2.5 years displayed progressive cognitive impairment, whereas cognition improved in those whose cortisol levels returned to the normal range. The most common stress was social isolation.

Older adults are at an increased risk for depression compared to younger and middle-aged adults. Most older adults, 80%, experience an episode of depression, and 50% have two or more episodes. Yet, depression remains underrecognized and undertreated among older adults, in part because the presenting signs and symptoms may be confused with dementia or other concomitants of aging. Depression is not a normal aspect of aging. Therefore, clinicians of older adults should maintain vigilance for the signs and symptoms of depression and offer treatment for this common yet highly morbid disorder.

FUTURE DIRECTIONS

Molecular medicine is poised to exponentially expand the understanding of the mechanisms mediating sexual differentiation and reproductive physiology. The therapeutic promise will be greatly expanded as it becomes ever more possible to translate understanding into clinical therapeutics. Concomitantly, given the importance of reproduction to the sense of mortality and immortality, it is likely that even greater anxiety will be aroused by the development of ever-more-clever ways to tinker with reproductive options. The idea that sex steroids imprint the brain in fundamental ways that determine sexual identities, longevity, and aptitudes will make it difficult to think in a polarized manner about gender, families, and aging. Both individual and collective psyche will continue to be even more troubled by the implications for society and self.

Reproductive behavior is extremely personal, but governments and religions often feel compelled to interfere with personal decision making and regulate what is and is not acceptable. A sense of invasion is unavoidable. Offspring may or may not applaud their biologic and social parents’ choices and actions. Indeed, concepts of belonging and identity will likely be ever more challenged and challenging as more and more reproductive choices are offered to individuals and couples. As the choices burgeon, individuals may feel ever more compelled to seek guidance from authorities and those in the helping fields. This means that the knowledge and scientific expertise needed for those who write laws, interpret supposedly normative teachings, and provide guidance will increase. The medical insurance system is ill-equipped to deal with the reproductive options currently being offered to prospective patients. Insurance companies will likely fail to appreciate the nuances of what constitutes medical necessity or coverage eligibility when reproductive options are perceived as departures from convention. Controlled reproduction has already refined the understanding of reproductive wellness and reproductive rights, but access has not kept pace with the options. It will be interesting to see how various cultures grapple with ideas about the role of women in reproduction and family building in the face of population expansion in some parts of the world and contracture in others. It will also be interesting to see how the expansion of societal concepts of gender affects our understanding of reproductive issues and our research priorities. It will take true societal wisdom to harness the promise of molecular medicine for reproductive and women's health in a manner that protects the vulnerable and satisfies critics and regulatory authorities. Science literacy will continue to be critically necessary, especially for policy development and application.

FURTHER READINGS

ACOG Policy and Position Statements. https://www.acog.org/clinical-information/policy-and-position-statements

APA Official Action. Position Statement on Screening and Treatment of Mood and Anxiety Disorders During Pregnancy and Postpartum. Approved by the Board of Trustees, December 2018. Approved by the Assembly, November 2018.

Bloch M, Schmidt PJ, Danaceau M, Murphy J, Nieman L, Rubinow DR. Effects of gonadal steroids in women with a history of postpartum depression. *Am J Psychiatry*. 2000;157(6):924–930.

Fairbrother N, Collardeau F, Albert AYK, et al. High prevalence and incidence of obsessive-compulsive disorder among women across pregnancy and the postpartum. *J Clin Psychiatry*. 2021;82(2):20m13398.

Fruzetti F, Fidecicchi T. Hormonal contraception and depression: updated evidence and implications in clinical practice. *Clin Drug Investig*. 2020;40(12):1097–1106.

González-Rodrıguez A, Cobo J, Soria V, et al. Women undergoing hormonal treatments for infertility: a systematic review on psychopathology and newly diagnosed mood and psychotic disorders. *Front. Psychiatry*. 2020;11:479.

Guttmacher Institute. Contraceptive use in the United States. https://www.guttmacher.org/fact-sheet/contraceptive-use-united-states

Hutner L, Catapano L, Nagle-Yang S, Williams K, Osborne LM, eds. *The APA Publishing Textbook of Women's Reproductive Mental Health*. APA Publishing; Washington, DC, 2022.

McCloskey LR, Wisner KL, Cattan MK, Betcher HK, Stika CS, Kiley JW. Contraception for women with psychiatric disorders. *Am J Psychiatry*. 2021;178:247–255.

McEvoy KM, Rayapati D, Washington Cole KO, Erdleyl C, Payne JL, Osborne LM. Poor postpartum sleep quality predicts subsequent postpartum depressive symptoms in a high-risk sample. *J Clin Sleep Med*. 2019;15(9):1303–1310.

Nillni YI, Nehralizade A, Mayer L, Milanovic S. Treatment of depression, anxiety, and trauma-related disorders during the perinatal period: a systematic review. *Clinical Psychology Review*. 2018;66:136–148.

Osborne LM. Recognizing and managing postpartum psychosis: a clinical guide for obstetric providers. *Obstetrics and Gynecology Clinics of North America*. 2018;45(3):455–468.

Osborne LM, Payne JL. Clinical updates in Women's Health Care Summary: mood and anxiety disorders: primary and preventive care review. *Obstet Gynecol*. 2017;130(3):674.

Poretsky L, Hembree WC, eds. *Transgender Medicine: A Multidisciplinary Approach*. Humana Press, Springer Nature; 2019.

Practice Committee Documents, the American Society for Reproductive Medicine. https://www.asrm.org/news-and-publications/practice-committee-documents/

Reinstein SA, Cosgrove J, Malekshahi T, Deligiannidis KM. Long-acting injectable antipsychotic use during pregnancy: a brief review and concise guide for clinicians. *Clin Psychiatry*. 2020;81(6):20ac13597.

Shen Y, Varma DS, Zheng Y, Boc J, Hu H. Age at menarche and depression: results from the NHANES 2005–2016. *Peer J*. 2019;7:e7150.

Sui G, Pan B, Liu G, Liu G, Want L. The long-term effects of maternal postnatal depression on a child's intelligence quotient: a meta-analysis of prospective cohort studies based on 974 cases. *J Clin Psychiatry*. 2016;77(11):e1474–e1482.

Viguera AC, Whitfield T, Baldessarini RJ, et al. Risk of recurrence in women with bipolar disorder during pregnancy: prospective study of mood stabilizer discontinuation. *Am J Psychiatry*. 2007;164(12):1817–1824; quiz 1923.

30.2 Genetic Counseling for Psychiatric Conditions

Jehannine Austin, Ph.D., M.S., F.C.A.H.S., C.G.C., and Holly L. Peay, Ph.D., M.S.

Historically, healthcare professionals have only infrequently had discussions with patients about the etiology of psychiatric illness and/or familial risk for recurrence. When these topics are addressed with patients, it is often brief and at a high level and/or performed in a group-based context (e.g., psychoeducation). However, discussing the etiology of psychiatric disorders with patients and their families in a deeper manner can be profoundly impactful. These discussions tap into fundamental, existential questions from patients and family members–questions such as, “Why me”?, “Did I do something that caused this”? and “Could I have prevented it”? Some people with psychiatric disorders explicitly ask about the origins of their condition(s), whether genetic or genomic information can provide useful information and the chances of recurrence in the family. But other times, patients and their family members find it difficult to ask these questions because such questions are a challenge to articulate or because they are anxious about the answers. Having an explanation regarding why one developed an illness is critical to the process of adapting to the condition. In the absence of being provided with a comprehensible explanation for the cause of illness, affected individuals will develop their own explanatory model. An explanatory model is important in determining how one behaviorally responds to that illness, including one’s treatment adherence. Studies show that people are more likely to actively participate in treatment if the treatment is seen to directly address what the individual perceives to be the cause of their illness. (See Inglis et al., 2014.)

Genetic counseling about psychiatric conditions has been provided by professional genetic counselors for several decades to address these fundamental questions about why psychiatric disorders develop. Interest in the service is widening, perhaps because of strong positive data about patient outcomes of psychiatric genetic counseling, improved understanding of genetic and environmental origins of these conditions, and enhanced focus on risk modification and management. Increasingly, requests for genetic counseling may be made directly to psychiatrists. Many psychiatrists are well placed to initiate discussions about and respond to patient requests for genetics education and to provide counseling around these issues. They may have knowledge of their patients' needs and family histories within ongoing therapeutic relationships. An ideal model, however, involves providing psychiatric genetic counseling in the context of a multidisciplinary team, with collaboration between genetics and mental health professionals. Such collaborative consultation is particularly relevant when there is a need for advanced risk assessment, collection and construction of complicated family histories, and counseling about the appropriateness and clinical utility of genetic and genomic testing.

Advances in knowledge and growth in the clinical use of genomic technologies may also increase inquiries to psychiatric professionals. For example, there is discussion ongoing in the literature regarding whether genetic testing for copy number variants (CNVs) should be performed routinely for all individuals with schizophrenia. Further, people are already able to upload their raw genomic data from direct-to-consumer genetic testing sites like 23andme or ancestry.com into third-party tools that generate polygenic risk scores (a summary variable, which combines many genetic variants of small effect to estimate a person's genetic risk for a given trait or condition) to learn about genetic predisposition to a range of outcomes, including both physical and mental health.

AR is a 33-year-old client with bipolar disorder. She discusses how much she would like to be a mother, but she is fearful about her mental health during and after pregnancy. She also is haunted by the idea that her child would develop bipolar disorder. She asks for expert advice about the advisability of having children.

BN is a 57-year-old client with recurrent major depression. He has two teenage children and is concerned about their mental health. He asks for advice about how to distinguish between normal teenage moods and a developing mood disorder. He wonders what symptoms to look for and what the appropriate response would be to any emerging symptoms. BN expresses his belief that at least one of his two children is bound to become ill because the illness is so common in his family.

KC is a 20-year-old who requests a consultation. She has no history of psychiatric symptoms. She indicates that she received concerning results about her future risk for a mood disorder from a direct-to-consumer genetic testing company. She would like your advice about how to interpret the results and what she should do to prevent becoming affected.

A couple in their early 30s requests a consultation. One partner has a history of mild social anxiety and has a parent with schizophrenia. The other partner has no contributory history. The couple is planning to become pregnant and asks your advice about embryo selection using polygenic risk scores to ensure a baby does not develop schizophrenia.

Although the situations and queries appear quite different, the underlying issues are similar. Their client’s concerns are based on the etiology of psychiatric illness, symptom identification, and risk for occurrence or recurrence in family members. Mental health professionals likely have clients who express such concerns regularly. These questions can lead to discussions about the cause of psychiatric illness, symptom effects during and after pregnancy, teratogen concerns with psychiatric medications, recurrence risks for asymptomatic family members, the utility of genetic and genomic testing, and early intervention and identification efforts—all topics that fall under the umbrella of psychiatric genetic counseling.

WHAT IS GENETIC COUNSELING?

Genetic counseling is the "process of helping people to understand and adapt to the medical, psychological, and familial implications of genetic contributions to disease." The National Society of Genetic Counseling describes the process as integrating the following factors:

- Interpretation of family and medical histories to assess the chance of disease occurrence or recurrence.
- Education about inheritance, testing, management, strategies for mitigating risk, resources, and research.
- Counseling to promote informed choices and adaptation to the risk or condition.

The process aims to minimize distress and facilitate adaptation, increase one's feeling of personal control, and facilitate informed decision-making and life planning.

Although the term "genetic" counseling suggests a primary focus on the genetic contributions to a condition, in fact–in the context of complex conditions in particular–both genetic and environmental components of the presenting condition are discussed together in a holistic manner. The complex interplay between genes and environment in the etiology of these conditions requires a comprehensive approach to care.

GENETIC COUNSELING FOR PSYCHIATRIC DISORDERS

Many healthcare professionals consider referring to genetic counseling when genetic testing may be an option for a patient. Yet genetic counseling has a long history of meaningful application for a variant of conditions, even *prior to* the development of genetic tests for those conditions. For psychiatric disorders, genetic counseling can be conducted with or without the aid of results from genetic or genomic testing.

Outcomes of psychiatric genetic counseling include:

- a better understanding of etiology and risk in the family,
- psychological support related to familial risk,

- a discussion of the limited role of parenting in causing or preventing psychiatric disorders (reflecting current scientific knowledge),
- increased empowerment and self-efficacy in terms of managing one’s condition,
- early symptom identification and early intervention, and
- facilitated meaning-making and coping.

In our experience, interest in genetic counseling is generally driven by the client’s appraisal of psychiatric illness within the family. Several factors that play into a client’s evaluation of the illness in the family include the perception of the condition, beliefs about what caused the condition within the family, existing perceptions of risk to family members, and the ability to control risk. With regard to these issues, in the context of psychiatric disorders:

- Affected individuals and family members often describe the disorders as caused by genetic/biologic and environmental factors.
- When considering the risk to family members, many close relatives and individuals with bipolar disorder and schizophrenia overestimate the risk of family recurrence.
- Individuals affected with bipolar disorder and schizophrenia often perceive moderate to high illness burden, and relatives often perceive high burden.
- There is some evidence that a subset of individuals with mood and psychotic disorders, and a subset of their relatives, are less willing or unwilling to have children because of fear of illness in the children.

During a psychiatric genetic counseling consultation, it is helpful to ask the client about these factors because they will likely have a strong effect on how they perceive the information provided and the resulting decisions and life plans that they make.

Cassie and her partner, John, sought genetic counseling as a couple to understand more about the causes of mental illness. John said that he wanted to better understand—as he put it: “Why is my sister-in-law’s obsessive-compulsive disorder (OCD) so much more severe than Cassie’s”? In discussion, it emerged that he doubted that Cassie should have been diagnosed with OCD. In exploring the couple’s existing explanations for the cause of illness, it emerged that John perceived environmental factors to be more important in the etiology of OCD. Cassie’s sister had a traumatic event before her symptoms became noticeable to the family, while Cassie had experienced nothing similar. After John repeated several times that he felt that Cassie’s OCD was not as severe as her sister’s, Cassie broke down in tears, saying that she felt John minimized her OCD and that she felt her symptoms were just as severe though not as overt as her sister’s. The clinician was able to help John identify that he was terrified that genetics might be involved because of what this might mean for children that he and Cassie planned to have together. This fear caused him to deny Cassie’s experience. By the end of the session, John and Cassie had come to a better understanding and were developing a plan to work on protective factors together.

After Anna (12) received a diagnosis of an eating disorder, she and her mother requested a session to discuss the causes of eating disorders. Both Anna and her mom were very distressed about the diagnosis, and both seemed to attribute the disorder entirely to genetic factors. Anna said: “It’s genetic, right? I’m just different…and weird.” Anna’s mom was feeling helpless, watching Anna as she withdrew from her friends. It emerged that Anna’s mom had experienced anorexia as an adolescent and felt profoundly guilty for having “passed on” a condition that she experienced as life-threatening at a similar age. As the clinician discussed the roles of genes and environment in the etiology of eating disorders and relating this back to Anna’s life experiences, both mom and daughter expressed feelings of relief. For the mom, the relief was about feeling that perhaps it was not “all my fault,” and for Anna, it was about feeling less fundamentally different (flawed). As the conversation turned toward protective factors, Anna’s mom began sharing strategies with her daughter that she had found to be helpful in her own recovery. This provided her with a feeling that she was not entirely helpless and seemed to bring the pair closer together.

THE ROLES OF THE GENOME AND ENVIRONMENT IN MENTAL HEALTH

Knowledge about the etiology of psychiatric disorders is improving steadily. These improvements allow more accurate and specific answers about why a disorder has occurred in a particular individual and family. Likewise, it increases our ability to offer more accurate assessments of risk and counseling about environmental modifications.

It is widely accepted that psychiatric illnesses tend to run in families. Disorders can recur in families for several reasons, including shared genomic variation and/or shared environmental exposures. Thus, as described previously, genetic counseling for etiologically complex disorders entails discussions around *both* genetic and environmental risk factors. For example, consider a case in which a father with bipolar disorder has concerns about symptoms developing in his children. Genetic counseling may focus on the following:

- The family history is indicative of both genetic and shared environmental risk in the family.
- Risk factors for the children, stemming from those genetic and environmental factors.
- Early symptom identification and early intervention (aiming to modify the environment and thus affect the genetically mediated progression of the condition).
- A discussion of the benefits and limitations of environmental modifications aimed at reducing the risk or minimizing emerging symptoms.

The concept of complex disorders accounts for the interplay of the genome and the environment and also includes the broader biologic and physiologic mechanisms of evolution and development that operate within societies and cultures. The onset of most cases of psychiatric conditions, like the onset of other common disorders, is thought to arise when an individual has numerous risk-increasing genomic variants and has experienced one or more environmental or experiential influences that contribute to risk.

The complicated nature of the etiology of psychiatric disorders can be understood by considering an illustrative family history, such as shown in Figure 30.2–1. This pedigree illustrates many different aspects of complex conditions, including the following:

1. Although the variants predisposing to disorders A and B segregate through the family, the disorders in question aggregate in the family.
   a. Individuals may have the necessary genomic susceptibility and never experience any illness because they lack the necessary environmental exposures.
   b. Individuals may have the necessary genomic and environmental susceptibility and never experience any illness due to protective factors (genomic and/or environmental).
2. A different number or complement of susceptibility variants and/or different environmental exposures leads to the existence of two etiologically related disorders in the family, A and B.
3. The typical ages at onset of disorders A and B range from early adolescence through middle adulthood, making it difficult to determine who in the family will remain unaffected throughout their lifespan.

4. Genomic variants that increase risk are common in a general (unaffected) population.

Genetic Contributions to Psychiatric Disorders

The first evidence that genetic factors were operating in psychiatric disorders came from analysis of families with multiple affected members, twin studies, and adoption studies. Family studies clearly indicated familial etiology, although the relative contributions of genes and environmental factors were unclear. Twin and adoption studies provided insights into the relative contributions of etiologic risk factors, and for all major psychiatric disorders, consistently indicated the importance of genes and environment. Family, twin, and adoption studies also provided insight into the genetic heterogeneity and phenotypic variability of mental disorders. Table 30.2–1 provides a summary of family, twin, and adoption data for schizophrenia.

Thousands of candidate gene, linkage, and genome-wide association studies have provided insight into genetic variants that increase susceptibility to psychiatric illness. Different types of genetic variation can influence vulnerability to psychiatric illness, including single nucleotide and CNVs. These variants do not generally confer vulnerability to a single discrete psychiatric diagnosis. Rather, they confer vulnerability to a range of psychiatric disorders. It is important to remember that many individuals with genomic susceptibility will never develop a psychiatric illness.

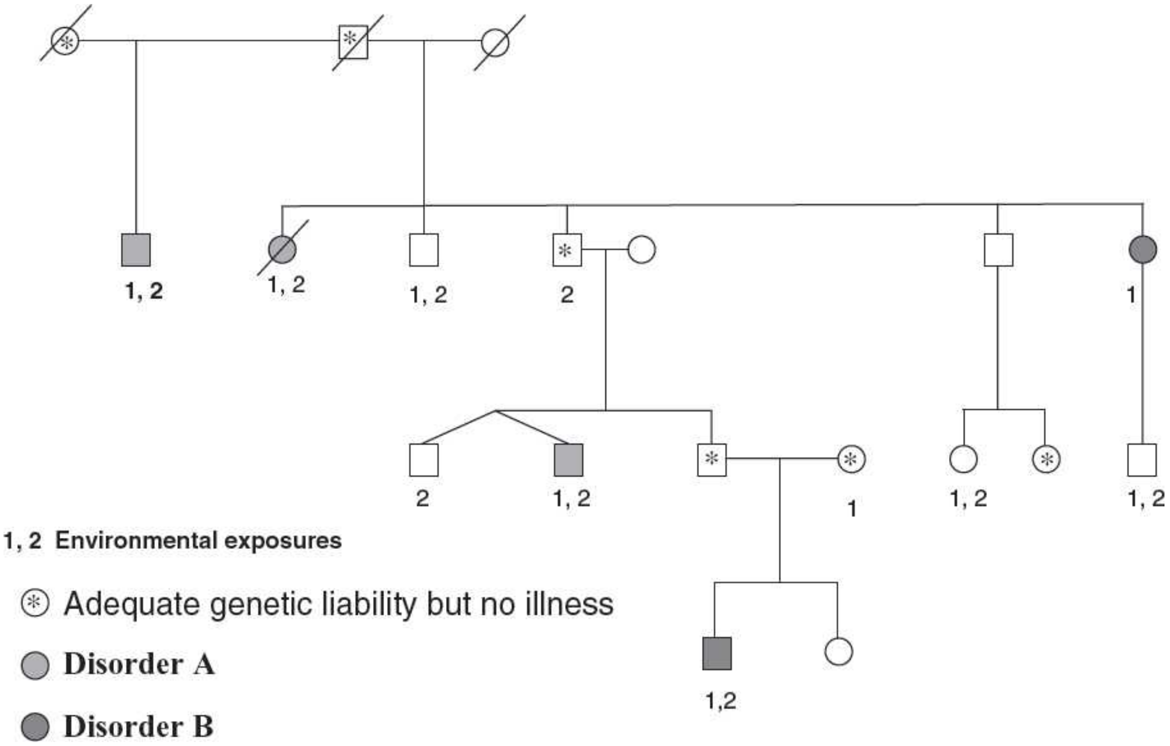
1, 2
1, 2
1, 2
2
1
2
1, 2
1
1, 2
1, 2
1,2
1, 2 Environmental exposures
⊛ Adequate genetic liability but no illness
● Disorder A
● Disorder B

FIGURE 30.2–1. Simulated pedigree displaying the aspects of complex inheritance.

Table 30.2–1.
Data from Family, Twin, and Adoption Studies for Schizophrenia

Family Studies	Twin Studies	Adoption Studies
Empirical risk for first-degree relatives: ~5–15% Empirical risk for second-degree relatives: ~2–8% Also increased risk for spectrum disorders (schizoaffective disorder, schizotypal personality disorder, and paranoid personality disorder)	Monozygotic twin concordance: 40–60% Dizygotic twin concordance: 10–16% Heritability: ~80%	Support high heritability; unique environmental exposures play a significant role, whereas the influence of shared environmental exposures is more limited

Single Nucleotide Variants. Many genetic risk variants for psychiatric disorders are of small effect and very common in the general population. For example, the hundreds of individual single nucleotide variants that have been associated with bipolar disorder and schizophrenia in the context of genome-wide association studies typically have odds ratios in the range of 1.01 to 1.3 and are so common in the population that a large percentage of the population will have some of them. While individually, these variations cannot provide much information about the risk of developing a psychiatric disorder, there has been great interest in combining these variants together to generate polygenic risk scores (PRS, as described briefly above). It is important to know that PRSs will not be able to determine who will and will not develop a psychiatric disorder. At present, the best PRS for schizophrenia explains only 11% of the overall liability for developing the condition, and it has been estimated that the upper limit of the liability for schizophrenia that could be explained by a PRS is ~25%. Therefore, the clinical utility of these tests remains in question.

Copy Number Variants. Structural variations like copy number variants (CNVs–missing or extra segments of chromosomes) also contribute to the etiology of psychiatric disorders. Though less common in the population than single nucleotide variants, when present, they have a much larger influence on the risk for psychiatric illness. The best-known example of this phenomenon is 22q11 deletion syndrome, where individuals who have this deletion have a greater chance to develop a variety of psychiatric disorders, including psychotic disorders, with an associated risk of ~25% to 30%.

Epigenetics. New research approaches and a growing understanding of the function of the human genome have opened promising studies of epigenetic phenomena (heritable changes that do not involve alteration to the DNA sequence). A detailed understanding of the causal pathway from any of the identified genomic contributions to disorder susceptibility is not yet clear. It is apparent, however, that changes in the genome do not unequivocally result in major psychiatric disorders. Rather, current knowledge suggests that particular variants may increase the risk of adverse response to an environmental stimulus or may more directly affect a biologic process in a way that facilitates the onset of particular symptoms. As described in the benchmark findings of the Schizophrenia

Working Group of the Psychiatric Genetics Consortium (2014), most genomic variants exert an effect through altering gene expression rather than through altering protein structures.

Environmental Contributions to Psychiatric Disorders

For disorders like psychiatric disorders that have a complex etiology, environmental factors are typically necessary for the onset of the condition. In this context, “environment” is a catch-all phrase meaning anything outside of the genome that can affect the expression of a gene, including the in-utero environment during development, shared and nonshared family environment, exposures to harmful substances, and medical history.

Adoption studies have indicated that the strongest environmental risk factors are unique or nonshared factors (meaning factors that make siblings less alike). Many misinterpret this to mean that the home environment is not important in the development of psychiatric illness because a shared home environment should make siblings more alike. But home environment and parenting style are not equal for all siblings and may not be stable over time. Outside effects on the home environment, sibling order, sibling relationships, and differences in sibling personality traits (themselves genetically mediated) have effects on the home environment. Even if the home environment is identical, siblings likely perceive differences in their shared environment based on their individual developmental history and personality traits. In some cases, the perception of differences may be as important as actual differences. Thus, the identification of unique factors as the most important environmental contributors does not exclude the relevance of the home environment in the etiology of psychiatric illness. It is important for parents to know, however, that in many cases, the home environment is unlikely to be sufficient to either cause or prevent psychiatric illness in children.

Unique environmental risk factors are unequivocally important in the development of psychiatric illness. Factors that are likely to play a role in schizophrenia and bipolar disorder include the following:

- Obstetrical complications
- Substance use, specifically cannabis and methamphetamine
- Childhood trauma
- Being reared in an urban environment

- Advanced paternal age
- Immigration
- Maternal prenatal infections
- Season of birth
- Head injury
- Stressful life events
- Viral infection

Unfortunately, many of these factors are pervasive and largely outside of an individual’s control.

The interplay between genomic and environmental risk factors is complicated; for example, some cases of depression may be caused through an interaction of genomics and environment, such that susceptibility variants are expressed in an adverse environment but not a supportive one.

PRIORITIES OF PSYCHIATRIC GENETIC COUNSELING

Genetic counseling integrates exploration of psychological and social issues that result from living with a psychiatric condition (personally and/or in the family) with a personal and family medical history (FMH). Increasingly genetic counseling also includes genomic risk information. Genetic counseling takes a person-centered approach that focuses on supporting the client in making autonomous informed health decisions.

Table 30.2–2.
Steps and Process of Genetic Counseling

Solicit and clarify the client's presenting questions and goals for genetic counseling.
Explore the client's existing explanation for the cause of illness.
Collect and review medical and family medical history.
Identify support systems.
Verify diagnoses in the proband and other affected family members when possible.
Address issues and concerns identified through the genetic counseling process.
Assess the client's emotional and intellectual capacity before proceeding to determine the approach to the provision of education and counseling.
Provide information at cognitively appropriate levels.
Note that the processing of emotional reactions is intertwined with the provision of information.
Assess the personal meaning of the information and the client's willingness to negotiate various risks and burdens.
When applicable, assist the client in arriving at a decision by discussing available options; discuss benefits and limitations of each alternative.
Assist the client with adapting to the risk status in the family.
Assist in formulating a plan that the client is able to carry out.
Provide follow-up counseling and support. Continue to assess the client's and family's understanding of the information and the effect of the information, risks, or decisions.

The basic components of genetic counseling are outlined in Table 30.2–2.

The optimal approach is a two-way exchange of information regarding the client's expectations for the session, understanding of the problem/diagnosis, and future plans or options.

A couple in their mid-30s with a 10-year history of infertility had been trying to adopt a child. Recently, the adoption agency told them of a baby being placed for adoption because the biologic mother was affected with bipolar disorder and did not feel she could adequately care for the baby. The FMH collected on the newborn baby did not identify others in his family with psychiatric disorders. The recurrence risk for bipolar disorder in the newborn was, therefore, estimated to be between 5% and 20%, with additional risks for other psychiatric disorders (Table 30.2–3). The couple individually reacted disparately to the estimated risks. When the professional explored factors contributing to their feelings regarding the risks, the husband shared his experience with a childhood neighbor who had "some kind of mental illness" and detailed the "torment and agony" that the child brought to the family. Retorting, the woman shared that her coworker had bipolar disorder and did "just fine" with the help of medication. She, therefore, did not feel that the risks for psychiatric disorders were of concern. The professional facilitated a discussion of the spectrum and meaning of mental illness, along with recurrence risks. Although the couple did not come to an agreement at that meeting over the potential for adopting the child, they did feel that the information and sharing of experiences and perspectives were beneficial. They agreed to return in one week to reach a decision regarding the adoption.

Providing Educating and Counseling about Etiology

By its very nature, genetic information is complex and often technical, with a vocabulary that is not familiar to clients. Traditionally, the professional selects the initial core of information presented to the client. Significant effort should be made to identify and include the information considered most relevant by the client and to identify and address the emotional and psychological issues that discussing this information can raise. The topics that are most often included are a description of the disorder and its natural history, explanation of the possible modes of inheritance and possible implications for family risk, discussion of environmental risk factors, options for early intervention, medical management, prevention or treatment, and (in the cases in which it is applicable) the availability of genomic testing and the associated risks, benefits, and limitations. Owing to the complexity of the information and discrepancies in client learning styles, providers may use a variety of means to communicate the information, including written material; graphics that visually present genetic terms, concepts, and modes of inheritance; and a detailed letter summarizing the clinical encounter; as well as other available resources (e.g., support groups and web-based resources).

Discussions about etiology can raise emotional issues related to guilt, shame, and stigma. Thus, the provision of understandable and meaningful information should be coupled with empathy, compassion, and sensitivity to provide a multifaceted genetic counseling encounter. Individuals, couples, and families who experience greater feelings of stigma, higher perceived burden, and higher perceived risk for recurrence in offspring or siblings are likely to require more intensive counseling around family vulnerability. Whereas issues of the stigma associated with mental illness are commonly encountered by the client and addressed by the practitioner within the psychiatric setting, additional efforts to assess stigma pertaining to a genomic etiology seem prudent. Likewise, some clients may experience a higher perceived burden due to the recognition of a genomic contribution and feel that they have less control over preventing or treating their condition.

Periodic follow-up provides opportunities to address emerging issues as they present to the client. The extended nature of such contact provides an optimal opportunity to assist the client or family member in receiving information when it is most relevant, as well as facilitating their use of the information and adaptation to its implications.

In some cases, genomic testing may inform risk, especially in cases with unusual presentation (i.e., highly loaded families; psychiatric symptoms manifesting along with dysmorphic features, birth defects, and/or learning difficulties that are not associated with the psychiatric phenotype). Consideration of genomic testing should include a careful review of potential benefits, limitations, and risks. A genetic counselor with expertise in psychiatric conditions should be consulted to confirm the appropriateness of genomic testing.

Communication of information to interested relatives is appropriate and encouraged within the process of genetic counseling. Invitations to family members to receive information related to the condition in question, the etiologic underpinnings, recurrence risks, and counseling to address the related psychological or emotional issues are encouraged. Summary letters describing the consultation can be particularly useful for sharing information with others in the family.

Counseling about Risk Reduction in Children

Clients with children or those considering having children often benefit from counseling specifically focused on the management of uncertainty and the potential for risk reduction. Clients are often keenly interested in information about environmental risk because it may facilitate the placement of blame away from parents while at the same time providing potential interventions that may reduce the risk for their children. Given the nature of risk factors, this counseling often focuses on their limited ability to reduce the risk of the onset of illness in at-risk individuals. There are, however, potential targets for intervention. At-risk individuals have been found to be more vulnerable to certain street drugs (most notably cannabis and methamphetamine) and should be educated about the possibility that such substances may trigger episodes of psychiatric illness. Thinking about the environment more broadly, it may be useful to suggest cognitive-behavioral therapy and the development of strategies for effectively coping with and managing stress. For affected parents with young children, parenting programs, social support, and daycare may be helpful. Providers should discuss the importance of early detection of psychiatric illness because early, and appropriate interventions improve prognosis.

Individuals with psychiatric conditions are often vigilant about monitoring for symptoms in children, though they may have difficulty determining normal behavior from abnormal. Psychiatrists should assist such families in achieving a balance between appropriate vigilance and inappropriate boundary setting or anxiety. Although it may be worrisome to have risk in the family, a family with an affected individual(s) is well placed to recognize early manifestations of symptoms. Encouraging and educating clients about their abilities in this capacity can increase their sense of control.

Risk Assessment

Different clients seek genetic counseling for different reasons, and not all will be interested in discussing chances for themselves or others in the family to be affected. Whether the client is interested in risk assessment should be explored at the beginning of the session, and–if provided–this should be done within the contextual framing of what is known about the etiology of these conditions. Therefore, it can be helpful to contract around risk communication twice, once at the beginning of the session and again once this contextual framing discussion has been completed. For clients who are interested in counseling about risk, some may prefer a qualitative discussion of the risk magnitude. These individuals may not welcome or benefit from receiving numerical risk information. For those who prefer a detailed numerical estimation of risk for recurrence, risks based on the family and medical history should be presented as a range to reflect the ambiguity of the estimates. In some cases, based on the family and medical history, it may be possible to suggest that one end of the range is more appropriate for a specific client.

Perceptions of high illness burden and limited control over onset are likely to result in less tolerance for familial risk. Clients often have preconceived ideas about their (or other family members') risk or affected status. These may be influenced by perceptions such as similarities to affected relatives or family and cultural beliefs. As such, risks, as provided, may differ from what is expected, and this should be explored during the consultation.

Individuals also vary in their capability to understand numeric risk. Many people perceive risks from a dichotomous, all-or-nothing perspective (i.e., it is or is not going to happen), and the presentation of risks in a numerical fashion (i.e., 1 person out of 10, or 10%) may be difficult for many clients to process. The baseline for risk provision should be absolute risks. Relative risks may be interpreted incorrectly, especially when the disorder is rare (such as autism). Once clients understand the absolute risk, in some cases, it may be useful to clarify understanding by discussing relative risks.

The majority of families have only one affected family member, and thus empirical risks often are in the range of 10% to 15% or lower. For individuals who do not wish to be given a risk number but instead prefer a qualitative assessment of risk, qualitative communication might be posed as follows: “You are right to think that your children have an increased chance of having (the disorder[s] in question). However, your children have a better chance of *not* having (the disorder[s] in question) than having it.” The meaning of descriptive risks is highly subjective, and patients should be consulted about their understanding.

Finally, it is important to remember that it is the client's perception of risks (perceived or presented) rather than the actual risk that is more likely to affect future decisions and outcomes. A client's perceived estimate of the likelihood that an event will occur is affected by their ability to recall or visualize examples of the occurrence and by their perceptions of the nature of the event (i.e., severity).

The FMH is used to assess the likely etiology of the disorder and estimate risks for recurrence in the family. Confirmation or clarification of the diagnosis or diagnoses in the family is essential to the provision of valid information. Topics that are specific to psychiatric disorders and are important to include in the FMH are listed below.

- Any psychiatric diagnosis, whether childhood, adult, postpartum, or geriatric
    - Age at symptom onset and at diagnosis
    - Subjective assessment of illness severity and treatment response
    - Potentially important environmental exposures (e.g., obstetric complications, cannabis use, head injury)
- Symptomatic, undiagnosed individuals and/or undiagnosed individuals treated with a psychiatric medication
- Developmental history (e.g., has the individual achieved normal milestones for age, such as independent living and employment for adults?)
- Social history
- Substance abuse
- Suicide and self-harm
- Birth defects, intellectual disability or learning disabilities, and unusual medical conditions
- The age and sex of at-risk family members

As mental health providers will be aware, challenges to collecting psychiatric FMH may include guilt and stigma that limit knowledge of diagnoses among family members; the frequency of undiagnosed psychiatric illness; and the dynamic nature of emerging psychiatric symptoms that may result in diagnostic change over time.

Risk Assessment Based on Family and Medical History

In most cases, risk assessment is based on FMH in the absence of meaningful genomic test results. FMH data are used together with empiric risks (see Table 30.2–3). It is vital to understand that empiric risks are not specific to any one individual or family but rather are risks obtained from studies that aggregate recurrence risk among large numbers of families. Factors specific to the proband, the family, and at-risk individuals allow for the tailoring of empirical risks. At its most basic, such tailoring requires considering the following series of questions.

Questions about the affected individual(s) in the family:

1. Who will be considered to be affected in the family?
   1. Many individuals who meet the criteria for a major psychiatric disorder are not diagnosed or treated, many are incorrectly diagnosed, or the diagnosis may not be known by the client.
   2. The typical late-teenage/young-adult onset of many major psychiatric disorders confounds pedigree assessment because it is possible that individuals who are genetically susceptible and will become symptomatic have not yet experienced the onset of illness.

Table 30.2–3. Empirical Risks for Selected Mental Disorders

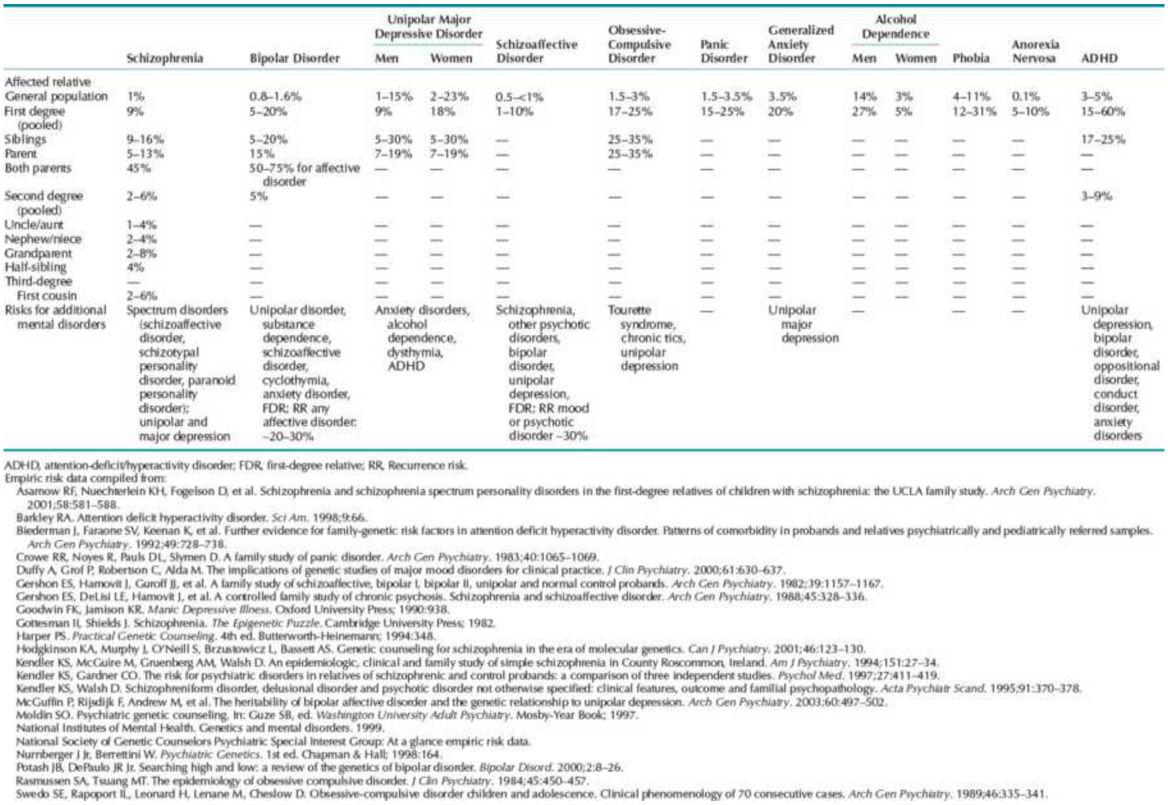

	Schizophrenia	Bipolar Disorder	Unipolar Major Depressive Disorder		Schizoaffective Disorder	Obsessive-Compulsive Disorder	Panic Disorder	Generalized Anxiety Disorder	Alcohol Dependence		Phobia	Anorexia Nervosa	ADHD
			Men	Women					Men	Women			
Affected relative													
General population	1%	0.8–1.6%	1–15%	2–23%	0.5–<1%	1.5–3%	1.5–3.5%	3.5%	14%	3%	4–11%	0.1%	3–5%
First degree (pooled)	9%	5–20%	9%	18%	1–10%	17–25%	15–25%	20%	27%	5%	12–31%	5–10%	15–60%
Siblings	9–16%	5–20%	5–30%	5–30%	—	25–35%	—	—	—	—	—	—	17–25%
Parent	5–13%	15%	7–19%	7–19%	—	25–35%	—	—	—	—	—	—	—
Both parents	45%	50–75% for affective disorder	—	—	—	—	—	—	—	—	—	—	—
Second degree (pooled)	2–6%	5%	—	—	—	—	—	—	—	—	—	—	3–9%
Uncle/aunt	1–4%	—	—	—	—	—	—	—	—	—	—	—	—
Nephew/niece	2–4%	—	—	—	—	—	—	—	—	—	—	—	—
Grandparent	2–8%	—	—	—	—	—	—	—	—	—	—	—	—
Half-sibling	4%	—	—	—	—	—	—	—	—	—	—	—	—
Third-degree	—	—	—	—	—	—	—	—	—	—	—	—	—
First cousin	2–6%	—	—	—	—	—	—	—	—	—	—	—	—
Risks for additional mental disorders	Spectrum disorders (schizoaffective disorder, schizotypal personality disorder, paranoid personality disorder); unipolar and major depression	Unipolar disorder, substance dependence, schizoaffective disorder, cyclothymia, anxiety disorder, FDR: RR any affective disorder: ~20–30%	Anxiety disorders, alcohol dependence, dysthymia, ADHD		Schizophrenia, other psychotic disorders, bipolar disorder, unipolar depression, FDR: RR mood or psychotic disorder ~30%	Tourette syndrome, chronic tics, unipolar depression	—	Unipolar major depression	—		—	—	Unipolar depression, bipolar disorder, oppositional disorder, conduct disorder, anxiety disorders

ADHD, attention-deficit/hyperactivity disorder; FDR, first-degree relative; RR, Recurrence risk.

Empiric risk data compiled from:

Asarnow RF, Nuechterlein KH, Fogelson D, et al. Schizophrenia and schizophrenia spectrum personality disorders in the first-degree relatives of children with schizophrenia: the UCLA family study. *Arch Gen Psychiatry.* 2001;58:581–588.

Barkley RA. Attention deficit hyperactivity disorder. *Sci Am.* 1998;9:66.

Biederman J, Faraone SV, Keenan K, et al. Further evidence for family-genetic risk factors in attention deficit hyperactivity disorder. Patterns of comorbidity in probands and relatives psychiatrically and pediatrically referred samples. *Arch Gen Psychiatry.* 1992;49:728–738.

Crowe RR, Noyes R, Pauls DL, Slymen D. A family study of panic disorder. *Arch Gen Psychiatry.* 1983;40:1065–1069.

Duffy A, Grof P, Robertson C, Alda M. The implications of genetic studies of major mood disorders for clinical practice. *J Clin Psychiatry.* 2000;61:630–637.

Gershon ES, Hamovit J, Guroff JJ, et al. A family study of schizoaffective, bipolar I, bipolar II, unipolar and normal control probands. *Arch Gen Psychiatry.* 1982;39:1157–1167.

Gershon ES, DeLisi LE, Hamovit J, et al. A controlled family study of chronic psychosis. Schizophrenia and schizoaffective disorder. *Arch Gen Psychiatry.* 1988;45:328–336.

Goodwin FK, Jamison KR. *Manic Depressive Illness.* Oxford University Press; 1990:938.

Gottesman II, Shields J. Schizophrenia. *The Epigenetic Puzzle.* Cambridge University Press; 1982.

Harper PS. *Practical Genetic Counseling.* 4th ed. Butterworth-Heinemann; 1994:348.

Hodgkinson KA, Murphy J, O'Neill S, Brzustowicz L, Bassett AS. Genetic counseling for schizophrenia in the era of molecular genetics. *Can J Psychiatry.* 2001;46:123–130.

Kendler KS, McGuire M, Gruenberg AM, Walsh D. An epidemiologic, clinical and family study of simple schizophrenia in County Roscommon, Ireland. *Am J Psychiatry.* 1994;151:27–34.

Kendler KS, Gardner CO. The risk for psychiatric disorders in relatives of schizophrenic and control probands: a comparison of three independent studies. *Psychol Med.* 1997;27:411–419.

Kendler KS, Walsh D. Schizophreniform disorder, delusional disorder and psychotic disorder not otherwise specified: clinical features, outcome and familial psychopathology. *Acta Psychiatr Scand.* 1995;91:370–378.

McGuffin P, Rijsdijk F, Andrew M, et al. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry.* 2003;60:497–502.

Moldin SO. Psychiatric genetic counseling. In: Guze SB, ed. *Washington University Adult Psychiatry.* Mosby-Year Book; 1997.

National Institutes of Mental Health. Genetics and mental disorders. 1999.

National Society of Genetic Counselors Psychiatric Special Interest Group: At a glance empiric risk data.

Nurnberger J Jr, Berrettini W. *Psychiatric Genetics.* 1st ed. Chapman & Hall; 1998:164.

Potash JB, DePaulo JR Jr. Searching high and low: a review of the genetics of bipolar disorder. *Bipolar Disord.* 2000;2:8–26.

Rasmussen SA, Tsuang MT. The epidemiology of obsessive compulsive disorder. *J Clin Psychiatry.* 1984;45:450–457.

Swedo SE, Rapoport IL, Leonard H, Lenane M, Cheslow D. Obsessive-compulsive disorder children and adolescence. Clinical phenomenology of 70 consecutive cases. *Arch Gen Psychiatry.* 1989;46:335–341.

2. Are illnesses in the family etiologically related?

   1. Heterogeneity within diagnostic groups and the occurrence of comorbidity make it more difficult to evaluate which disorders might be etiologically related in a family. The high prevalence of psychiatric disorders confounds the issue because some illnesses might be unrelated to other illnesses in the family.
   2. Generally, disorders that share a spectrum and disorders that are frequently comorbid are considered to be etiologically related. It is difficult to evaluate the relationship between other groups of disorders, although family studies have indicated that, on a research population level, no psychiatric diagnosis occurs in a vacuum; that is, if a relative has one psychiatric illness, family members are empirically increased risk for a range of other disorders as well.
   3. Assortative mating, whether between a couple with concordant or discordant diagnoses, complicates the assessment of the etiologic relationship.
3. What may factors in the family history raise or decrease risk?
   1. Families in which there are multiple affected individuals over several generations, or an unusual clustering of illness in one sibship, likely have the highest chance for recurrence.
4. What factors in the affected individual’s psychiatric history may raise or decrease risk?
   1. Age at onset: If age at onset is younger than expected, the risk to relatives may be increased. Relatives likely have an increased risk of experiencing early onset as well.
   2. Severity: More severe illness, and illness that is treatment-resistant, may be associated with increased risk to relatives.
   3. Sex: Depending on the psychiatric disorder in question, the sex of the affected individual may increase or reduce the risk to relatives, although the data are mixed. For example, siblings of females with schizophrenia have been shown in some studies to have a higher risk than siblings of males with schizophrenia.
   4. Environmental exposures: Relatives of affected individuals with known significant environmental exposures may have a reduced risk of the condition if they do not share the same environmental exposure. For example, relatives of an individual with psychosis who has a history of crystal methamphetamine use may have a lower risk of illness (assuming that they are not exposed to the drug). However, consideration of environmental factors is difficult to apply when providing quantitative risk assessment.

    5. Birth defects or a developmental/medical history suggesting an underlying genetic syndrome potentially increases the risk to family members.
5. What etiology best explains the illness in the family? What are alternate explanations?
    1. Although research suggests that the vast majority of psychiatric disorders are complex in etiology, it is not possible definitively to rule out Mendelian or nontraditional modes of inheritance in any particular family. When family patterns suggest the possible involvement of a genetic variant associated with a major effect, risks often exceed the empirical risk range. In many cases, such “loaded” families do not actually reflect the actions of a genetic variant of major effect, but regardless of cause, such a pattern in the family is indicative of increased risk. Families with multiple affected individuals may benefit from a referral to a genetics professional.

Questions about the at-risk individuals in the family:

1. For what disorder(s) might the individual(s) be at risk?
    1. In the case of psychiatric spectrum disorders, it may be difficult to evaluate *for what specific disorder* family members are at risk. Providers should familiarize themselves with appropriate spectrum disorders based on the at-risk individual’s family history. Family members may be at increased risk for comorbid disorders as well.
2. What factors about the individual may raise or decrease risk?
    1. Current age: At-risk individuals who are young have all of their years of greatest risk for the condition ahead of them, whereas older individuals have “lived through” a portion of their risk. Empirical risks generally reflect lifetime risk and can be qualitatively modified to reflect the at-risk individual’s age.
    2. Past psychiatric and developmental history: Individuals with existing psychiatric symptoms, another psychiatric disorder, or evidence of emerging psychiatric symptoms are at greater risk. Individuals with usual developmental histories (e.g., unusual social relationships; prolonged lack of gainful employment) may have a greater risk.
    3. Environmental exposures: Exposure to known environmental triggers may increase risk, although such exposures are currently difficult to identify and evaluate.

When evaluating any family history, providers should be aware that the presence of any of the following “red flags” may significantly increase risk. The identification of any of these factors should call into question the use of a standard empirical risk range.

- Learning disabilities or intellectual disability (preceding/unrelated to cognitive deficits due to psychiatric phenotype)
- Birth defects in the client or family members, most notably heart defects or oral clefts
- Dysmorphic facial features
- Unexplained major medical problems
- Neurologic impairment is not explained by the psychiatric diagnosis or medication side effects
- Very early onset of psychiatric symptoms (e.g., schizophrenia during childhood)
- Highly loaded family history (e.g., multiple individuals in several generations affected with psychiatric disorders; diagnoses may vary)

Figure 30.2–2 provides an overview of risk assessment for a family at moderate risk.

The pregnant couple depicted in Figure 30.2–2 presented for risk assessment related to the current pregnancy. There was no relevant family history on the maternal side of the family. Ages at onset and illness severity were typical in the family, and no obvious environmental risk factors were reported. The empirical risk for bipolar disorder for a child born to this couple (reflecting one affected parent) is between 5% and 20%. However, this risk does not incorporate the distantly related paternal cousin who also carries the diagnosis. Therefore, risks at the lower end of the range provided may underestimate the risk for the offspring of this couple. The empirical risk for this child to develop any major affective disorder is between 20% and 30% (again, the family history suggests that the lower end of the range may underestimate risk). The history of depression in the paternal great-aunt is difficult to evaluate because depression may be part of the bipolar disorder spectrum but is very common in females without family liability to bipolar disorder.

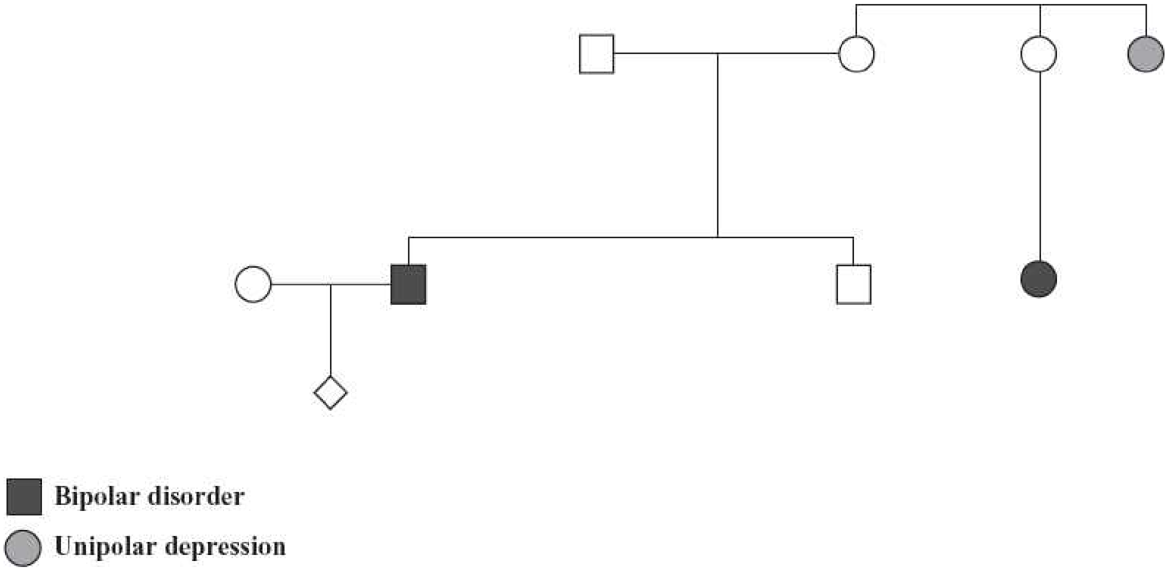

FIGURE 30.2–2. Pedigree of a family at moderate risk for affective disorder.

Families at High Risk (Exceptions). Matters are further complicated by the presence of families in which a combination of etiologic factors for psychiatric illness are skewed in a fashion that mimics a Mendelian pattern of inheritance (autosomal dominant or recessive), with multiple affected individuals in one or more generations. When evaluating such “loaded” families, it is important for the provider to consider the potential causes. Besides presenting unique psychological needs, highly loaded families have higher risks for recurrence, and genomic testing may be considered. Psychiatrists who are not highly experienced in genetics should refer such individuals and families to genetics professionals for evaluation and counseling.

The skewed presentation of psychiatric illness in selected families may be caused by a number of colliding factors that mimic single-gene inheritance. Some examples follow.

Complex Etiology with Unusual Accumulation of Risk Factors

It is possible that families with numerous affected individuals might have an unusual number of susceptibility variants segregating within the family, such that many family members have inherited multiple susceptibility-causing variants. One mechanism for this is “assortative mating,” in which individuals with psychiatric illness have children with other individuals with psychiatric illness. Such families may also experience recurrent shared environmental exposures that contribute to the onset of the disorder. In these cases, the risk to other family members might be high, although not currently reliably quantifiable.

Genetic Variants of Large Effect. Although overall, they are rare, there are a number of genetic variants that have significant effects on the risk for psychiatric illness and often can involve other features. Examples include Huntington disease, 22q11 deletion syndrome (22qDS), Wilson disease, Wolfram syndrome, late juvenile- or adult-onset metachromatic leukodystrophy, homocystinuria, acute intermittent porphyria, coproporphyria, frontotemporal dementias, and early-onset familial Alzheimer disease. These conditions are examples in which the presence of a genetic variation responsible for the syndrome increases the risk for the development of psychiatric disorders compared to the general population.

22qDS is an example of an inherited disorder with significant variability and increased risks for psychiatric disorders. It is an autosomal dominant microdeletion syndrome localized to the long arm of chromosome 22 (22q11). The clinical spectrum is variable and may include clefting of the soft palate, heart defects (most commonly tetralogy of Fallot, ventricular septal defect or right aortic arch), velopharyngeal incompetence, immune dysfunction, mild intellectual impairment, and psychiatric disorders, although rarely does an individual have all of the symptoms. Approximately 30% of individuals with 22q deletion experience psychotic disorders; risk is also increased for mood disorders. The onset of psychotic symptoms may be earlier than typical. Attention-deficit/hyperactivity disorder (ADHD) is also relatively common.

A FMH from the following case study illustrating the variability of expression of 22q deletion syndrome is shown in Figure 30.2–3.

A 29-year-old woman contacted a local psychiatrist for genetic counseling regarding her family history of schizophrenia and cognitive impairment. She was concerned that her future children might experience problems similar to those of her sister, nephew, and niece. The psychiatrist collected the FMH and constructed the pedigree. To confirm the diagnoses and to gain further information, the psychiatrist discussed the potential benefits and risks of obtaining medical records. The client decided to pursue this option and obtained permission from her affected sister. The psychiatrist received records on the affected sister and her 13-year-old son, who had mild intellectual disability. The records on the boy indicated that he had undergone genetics evaluation and testing and had been identified as having a microdeletion on the long arm of chromosome 22, consistent with 22q11 deletion syndrome. The records indicated that the client's sister also carried the deletion, although this information had not yet been shared with the family.

The identification of 22qDS in this family offered the option of genetic testing for others in the family. The 22q microdeletion is estimated to arise spontaneously in 80% to 90% of cases. However, this pedigree was suggestive of a hereditary form. Counseling in such situations is difficult owing to the variability of the disorder and the inability to predict the phenotype (e.g., the client may have the deletion and have a very subtle presentation). Genetic testing is warranted to clarify recurrence risks; however, issues regarding confidentiality of the acquired information require attention.

CHALLENGES POSED BY GENOMIC TESTING FOR PSYCHIATRIC DISORDERS

For the vast majority of individuals with psychiatric disorders, available genetic and genomic testing does not achieve clinical utility (A potential caveat is in regard to pharmacogenetic testing). As genomic technologies continue to improve and knowledge about variants that can increase vulnerability for psychiatric illness grows, identification of individuals who are at higher risk for developing psychiatric illness based on (for example) the number of illness-associated single nucleotide variants they carry will become more common. Appropriate applications of such developments clearly need careful consideration from both economic and ethical perspectives but may have important clinical uses. For example, at present, it can be very difficult to distinguish between emerging (prodromal) psychosis and normal adolescent behavior in youth who are at higher risk for psychosis (e.g., as a result of close family members who have experienced psychosis). However, some studies of prophylactic strategies to reduce risk among high-risk youth show promise, and for those in whom psychosis does emerge, early intervention is critical to allow for the best possible outcomes. Therefore, genomic technologies could be useful in determining which individuals might benefit most from prophylactic, therapeutic approaches.

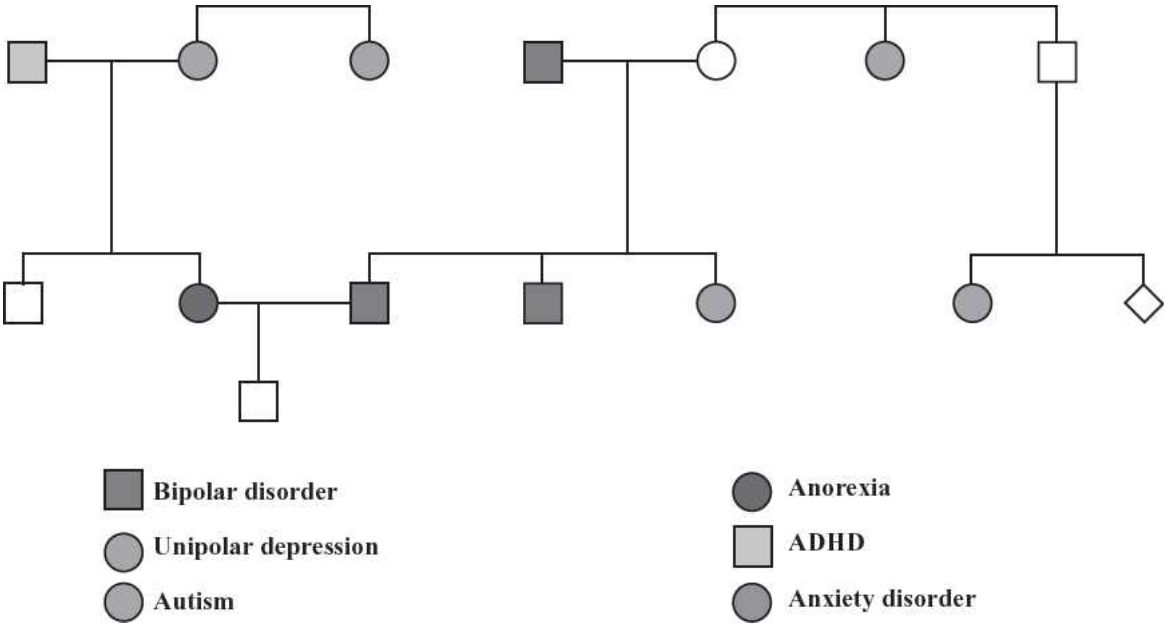

FIGURE 30.2–3. High-risk family. ADHD, attention-deficit/hyperactivity disorder.

The clinical implementation of genomic testing for psychiatric illness is and will continue to be the subject of intense debate. Decisions about the initiation of clinical testing will likely hinge on: (1) the demonstration of adequate proof that the identified variants unequivocally increase the risk for psychiatric illness; and (2) affirmation that the information gained from the testing has clinical utility (e.g., informs risk management strategies) and/or demonstration of personal utility, such as testing that aids in coping and adaptation. Conflicting opinions notwithstanding, genomic testing for susceptibility to psychiatric illness is becoming more commonplace. Psychiatrists will be on the front line for receiving requests for genetic counseling and testing because of their established relationship with clients and families with psychiatric disorders.

The clinical identification of genomic risks will most likely occur before most individuals have access to preventative options. The option of knowing risks without having access to preventative options raises concerns regarding the effect of such knowledge on the individual's mood, anxiety, distress, self-image, reproductive decisions, and decisions regarding career, family relationships, insurability, employment, and, potentially, other areas.

Across many indications, the majority of persons undergoing genomic testing and receiving their identified risk status have not been shown to experience long-term adverse effects. However, it should be anticipated that those receiving results indicating increased risk for a condition may experience elevated levels of condition-related anxiety and depression soon after receiving results and a small number have prolonged difficulty in adjusting to their results. Supportive counseling should be available to facilitate the adjustment.

The pursuit of genomic counseling and testing by an individual has clear and profound implications for other family members. Consideration should be given to helping the client understand the potential reactions and effects that their decision to pursue genomic testing could have on close family members. The health professional can play a role in assisting the client in considering ways to share the information with selected significant others. Similarly, consideration should be given to the effect of such information on the spouse or significant other.

Individuals who have psychiatric susceptibility are arguably more likely to enter genetic counseling with higher baseline levels of mood disturbance and may be more likely to respond negatively to predictive testing information. The important role of stress in the onset and recurrence of psychiatric symptoms must also be considered; by providing an at-risk individual with stressful, anxiety-provoking risk information, might providers inadvertently hasten the onset of psychiatric symptomatology? More research in this area is vital to ensuring that clinicians do not take too paternalistic a stance, refusing wanted information to individuals who might benefit psychologically and medically (as well, refusing to provide wanted information may be a stressful, anxiety-provoking experience for a client), while on the other hand considering the inherent vulnerability in a population at risk for psychiatric illness. It is likely that the manner in which these delicate conversations unfold is important to client outcomes and that a focus on counseling around these issues is key.

Given the complexity and uncertainty that comes with genomic testing, the pretest counseling, ordering, interpretation, and communication should be made by experienced researchers and clinicians. When considering genomic testing, we recommend consultation with a genetics professional.

ADDITIONAL ETHICAL, LEGAL, AND SOCIAL CONSIDERATIONS

Key questions for the clinical applications of genetic testing in psychiatry concern uncertainty around when our knowledge is “ready” for clinical implementation, when, how, and to whom to offer to test, and critical issues regarding equity. Specifically, given the populations (predominantly of European ancestry) from which datasets for PRS are generated, these tests will be more informative for some groups than others. Although there are many benefits of genomics knowledge, scientists and the lay public have worried, from the beginning of genetic research, about the potential for harmful consequences of knowing one’s genomic makeup. How will family members, neighbors, communities, employers, health care providers, and insurance companies react to such information? Might healthcare providers request that persons at increased genomic risk take psychiatric medications prophylactically? Might employers limit job opportunities for those with increased risk for psychiatric illness, keeping them in positions with lower levels of stress? Will persons at increased risk be pressured not to have children? Might providing such risk information induce significant enough anxiety/stress to trigger the onset of illness? And finally, how will we educate families and the public about the availability of susceptibility testing for psychiatric illness? Public messages about susceptibility testing should be crafted in a way that does not increase the already heavy burden of stigma attached to individuals with psychiatric disorders and does not add to “courtesy stigma” attached to at-risk relatives. Increased public awareness of the hereditary component of psychiatric illness may heighten stigmatization. Conversely, having an identified, biologic basis may supplant current public perceptions that mental illness is somehow a personal or family failure in moral, spiritual, or attitudinal perspectives.

The ability to provide susceptibility testing is likely to make psychiatric practice more complicated and will challenge the process of education, counseling, and informed consent. Psychiatrists are accustomed to communicating and receiving information that requires special consideration owing to its sensitive nature; handling genomic information from clients will be yet another sensitive topic to be considered from social and ethical perspectives. The diagnosis of a genomically influenced disorder may create risks not only for the patient but also for others within the family. Some clients may not choose to share genomic information with relatives, even though the information could be relevant to reproductive decisions, relationships, or even medical care. Coercion of affected relatives to undergo testing may become an issue, especially in the case in which at-risk relatives are pursuing genetic counseling to help guide their decisions. Parents may desire susceptibility testing of children even in the absence of preventive interventions and may risk stigmatizing the children and forcing information on them without informed consent. These situations call for careful attention to the confidential nature of the information, as well as the need to encourage family communication to discuss the potential for the associated personal, family, and social risks.

Historically, some attribution of psychiatric illness has been focused on child-rearing and the home environment. The current focus on the biologic nature of the psychiatric illness has been important for reducing guilt and feelings of blame in many families. It is unclear how a more nuanced understanding of environmental risk factors, potentially including home environment and parenting style, might affect perceptions of guilt. It is important that messages about risk reduction be crafted in such a way as to not lend credence to outdated notions of parental causation.

It can be anticipated that many individuals with clear-cut psychiatric diagnoses will not be found to have identified susceptibility variants. Test results with no identified variants are likely to cause some individuals to question the validity of their psychiatric diagnoses and may have a negative effect on treatment compliance. In addition, the absence of an identified genomic susceptibility might increase parental guilt and “force” responsibility for their child’s illness on their parenting efforts. A key aspect of pretest genetic counseling will be discussions about the meaning of a “negative” result.

Conversely, based on the very nature of susceptibility testing, many individuals identified as having susceptibility variants will never develop a psychiatric disorder. Counseling around the notion of susceptibility is complex, as is the necessary support for individuals who are at risk. Although the potential harm for such individuals is apparent, many may perceive benefits (perhaps related to perceived control and coping with vulnerability) that outweigh risks. If benign environmental modifications that reduce risk are identified, the concern around individuals with “positive” test results will decrease.

Finally, it is not yet clear how the public will react to genomic tests that indicate levels of risk that are not definitively predictive (like PRS), and indeed only modify risk estimations based on family history by a marginal amount. People with no psychiatric illness may perceive the importance of genomic results differently than those who have a psychiatric condition, while those who have a history of mild symptoms may re-evaluate past symptoms based on the results. It may be that individuals will interpret the resulting risk as higher than appropriate or that the predictive limitations will lead them to discount the results. Such appraisals are likely to vary greatly among individuals based on their history, personality types, and coping styles.

As individuals within society become more aware of the identification of genomic factors within the complex etiology of psychiatric disorders, it is anticipated that requests for psychiatric genetic counseling will increase. Although genomic testing may provide important psychological benefits for many individuals, the uncertainties inherent in the current level of knowledge pose challenges in the education and counseling of clients. Psychiatrists are appropriately positioned to provide genetic counseling (or referrals for genetic counseling) and ongoing therapy to families experiencing psychiatric disorders. The greatest opportunity for a benefit to clients will be provided through a multidisciplinary team comprised of mental health and genetics professionals who adhere to the basic principles of genetic counseling and maintain an awareness of the latest research findings relevant to psychiatric genetics. Genetic counselors within the geographical area of the psychiatrist can be identified through the National Society of Genetic Counselors (http://www.nsgc.org).

ACKNOWLEDGMENTS

Donald W. Hadley, MS and Barbara Biesecker Ph.D. MS contributed substantially to prior versions of this chapter.

FURTHER READINGS

Austin JC. Evidence based genetic counseling for psychiatric disorders: a road map. *Cold Spring Harb Perspect Med.* 2020;10:a036608.

Austin JC, Palmer CGS, Rosen-Sheidley B, Veach PM, Gettig B, Peay HL. Psychiatric disorders in clinical genetics II: individualizing recurrence risks. *J Genet Couns*. 2007;17:18–29.

Austin JC, Peay HL. Applications and limitations of empiric data in provision of recurrence risks for schizophrenia: a practical review for healthcare professionals providing clinical psychiatric genetics consultations. *Clin Genet*. 2006;70:177–187.

Baker DL, Schuette JL, Uhlmann WR. *A Guide to Genetic Counseling*. Wiley-Liss; 1998.

Baker K, Costain G, Fung WLA, Bassett AS. Chromosomal microarray analysis—a routine clinical genetic test for patients with schizophrenia. *Lancet Psychiatry*. 2014;1:329–331.

Biesecker BB, Peay HL. Ethical issues in psychiatric genetics research: points to consider. *Psychopharmacology (Berl)*. 2003;171:27–35.

Borle K, Morris E, Inglis A, Austin J. Risk communication in genetic counselling: exploring uptake and perception of recurrence numbers, and their impact on patient outcomes. *Clin Genet*. 2018;94(2):239–245.

Clatworthy J, Bowskill R, Rank T, Parham R, Horne R. Adherence to medication in bipolar disorder: a qualitative study exploring the role of patients' beliefs about the condition and its treatment. *Bipolar Disord*. 2007;9(6):656–664.

Duffy A, Grof P, Robertson C, Alda M. The implications of genetics studies of major mood disorders for clinical practice. *J Clin Psychiatry*. 2000;61:630–637.

Gamm JL, Nussbaum RL, Biesecker BB. Genetics and alcoholism among at-risk relatives I: perceptions of cause, risk, and control. *Am J Med Genet A*. 2004;128:144–150.

Hippman C, Lohn Z, Ringrose A, Inglis A, Cheek J, Austin JC. "Nothing is absolute in life": understanding uncertainty in the context of psychiatric genetic counseling from the perspective of those with serious mental illness. *J Genet Couns*. 2013;22(5):625–632.

Inglis A, Koehn D, McGillivray B, Stewart SE, Austin J. Evaluating a unique, specialist psychiatric genetic counseling clinic: uptake and impact. *Clin Genet*. 2015;87(3):218–224.

Iyegbe C, Campbell D, Butler A, Ajnakina O, Sham P. The emerging molecular architecture of schizophrenia, polygenic risk scores and the clinical implications for GxE research. *Soc Psychiatry Psychiatr Epidemiol*. 2014;49(2):169–182.

Martin CL, Wain KE, Oetjens MT, et al. Identification of neuropsychiatric copy number variants in a health care system population. *JAMA Psychiatry*. 2020;77(12):1276–1285.

Meiser B, Mitchell PB, Kasparian N, et al. Attitudes towards childbearing, causal attributions for bipolar disorder and psychological distress: a study of families with multiple cases of bipolar disorder. *Psychol Med*. 2007;37:1601–1611.

Meiser B, Mitchell PB, McGirr H, Van Herten M, Schofield PR. Implications of genetic risk information in families with a high density of bipolar disorder: an exploratory study. *Soc Sci Med*. 2005;60:109–118.

Miller SM, McDaniel SH, Rolland JS, Feetham SL. *Individuals, Families, and the New Era of Genetics: Biopsychosocial Perspectives*. WW Norton; 2006.

Murray GK, Lin T, Austin J, et al. Could polygenic risk scores be useful in psychiatry?: A review. *JAMA Psychiatry*. 2021;78(2):210–219.

Peay HL, Austin J. *How to Talk to Families about Psychiatric Illness*. WW Norton; 2011.

Peay HL, McCarthy-Veach P, Gettig B, Palmer CGS, Rosen-Sheidley B, Austin JC. Psychiatric disorders in clinical genetics I: addressing family histories of psychiatric illness. *J Genet Couns*. 2008;17(1):6–17.

Peay HL, Rosenstein DL, Biesecker BB. Adaptation to bipolar disorder and perceived risk to children: a survey of parents with bipolar disorder. *BMC Psychiatry*. 2013;13:327.

Peay HL, Rosenstein DL, Biesecker BB. Parenting with bipolar disorder: coping with risk of mood disorders to children. *Soc Sci Med*. 2014;104:194–200.

Quaid KA, Aschen SR, Smiley CL, Nurnberger JI. Perceived genetic risks for bipolar disorder in a patient population: an exploratory study. *J Genet Couns*. 2001;10:11.

Rutter M, Moffitt TE, Caspi A. Gene-environment interplay and psychopathology: multiple varieties but real effects. *J Child Psychol Psychiatry*. 2006;47:226–261.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014;511(7510):421–427.

Semaka A, Austin J. Patient perspectives on the process and outcome of psychiatric genetic counseling: an “empowering encounter.” *J Genet Couns*. 2019;28(4):856–858.

Smoller JW, Tsuang MT, Rosen Sheidley B, eds. *Psychiatric Genetics: Applications in Clinical Practice*. American Psychiatric Publishing; 2008.

Trippitelli CL, Jamison KR, Folstein MF, Bartko JJ, DePaulo JR. Pilot study on patients’ and spouses’ attitudes toward potential genetic testing for bipolar disorder. *Am J Psychiatry*. 1998;155:899–904.

Wray NR, Lin T, Austin J, et al. From basic science to clinical application of polygenic risk scores: a primer. *JAMA Psychiatry*. 2021;78(1):101–109.

30.3 Physical Violence on Adults

SHANNON F. STROMBERG, M.D., KYLE OWEN, D.O., AND NANCY MARTIN, M.D.

Violence remains a significant public health concern that can include abuse, assault, domestic violence, torture, and homicide. While there are multiple types of violence, there is no common denominator for them all. Victims span all races and genders; however, fixed risk factors are identified for specific groups. For example, the leading cause of death for black males continues to be homicide. Women continue to be at the highest risk for lifetime sexual assault. Additionally, the type of work in which one engages may predict violence. For example, sex workers or law enforcement officers may face threats daily. Statistics indicate that most Americans will become victims of violent crime during their lifetime.

The costs of violence may lie well beyond mortality. The morbidity for an individual exposed to violence results in decreased productivity, medical disability, and psychiatric sequelae. Morbidity can have far-reaching societal implications, with violence or perception of violence impacting where people decide to live, raise and school their children, how late they stay out, or even where they walk or drive.

During the past 2 decades, there has been increased emphasis on education and prevention. Statistics indicate an overall decrease in the total number of cases and rates of violence during that time. Education and resources available for victims and those at risk continue to improve. Treatment modalities available to victims have significantly expanded and have led to decreased morbidity in survivors; however, there is a need to continue preventing violence in all forms.

TYPES OF ASSAULT, PREVALENCE, AND RISK FACTORS

Assault can be viewed in the context of two variables: who is being assaulted and where it occurs. With these parameters, assault can be classified into several categories. Violent crime, rape, domestic violence, workplace violence, and torture are the most commonly used identifiers to classify assault. These different types of assaults are most often reported via two different data collection systems: the Federal Bureau of Investigation’s Uniform Crime Report (UCR), which collects reported crime information from local law enforcement agencies, and the Bureau of Justice Statistics’ National Crime Victimization Survey (NCVS). This survey generates estimates of the likelihood of victimization from different types of assaults. For each type of assault, there have also been attempts to develop risk factor assessments. Age, gender identity, sexuality, race, history of a prior assault, past psychiatric history, substance abuse, and income have been found to impact assault risk for different categories of assault.

VIOLENT CRIME

The UCR defines violent crime as murder and nonnegligent manslaughter, forcible rape, robbery, and aggravated assault. These categories exclude simple assault, which in the UCR is defined as an assault or attempted assault in which no weapon was used, or no serious injury occurred to the victim. Simple assault includes stalking, intimidating, coercing, and hazing.

Prevalence

In 2019, there were an estimated 5,813,410 violent crimes in the United States. Aggravated assaults accounted for an estimated 1,019,490 of this total, simple assaults were estimated at 3,800,190, and robberies were estimated to be 534,420. From 1993 to 2019, the rate of violent victimization (excluding simple assault) declined by 75%. The rates of all types of violent crimes in 2019 were lower than the rates observed in 2013.

Risk Factors

Males are more likely to experience violent crime than females. LGBTQ individuals are nearly four times more likely to be victims of violent crime, and they have higher rates of victimization in all types of violent crimes except robbery. Age plays a larger role, with those younger than 25 years at the highest risk. Racial and ethnic groups also appear to have different violent crime victimization rates. The groups most at risk for experiencing violent crime are persons of two or more races, American Indians and Alaska Natives. Black and Hispanic groups have higher rates than whites, while Asians have the lowest rate of any group. Lower-income households are more likely to be burglarized, and homelessness has also been shown to be a factor in increased physical assaults. Finally, multiple studies have shown substance abuse to increase the risk of victimization.

Perpetrators

Both the UCR and NCVS have limitations regarding the report of offender characteristics. It excludes any data from assaults and robbery not resulting in arrest. Another problem is that data are primarily generated based on the victim’s incident report. Although these incident data can be used to define the victim’s characteristics, they cannot be used accurately to define the offender. For example, the NCVS reports on victimization as reported by the victims. However, a tool as developed as the NCVS does not exist for offending as reported by the offender. Furthermore, incident data for victims cannot control for a situation in which one offender has assaulted multiple victims.

RAPE

The legal definition of rape varies from state to state. Some states strictly define rape, whereas others describe any sex crime as varying degrees of sexual misconduct or sexual assault. The UCR defines rape as follows: “penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim.” In this section, rape refers to forced oral, vaginal, or anal penetration, and sexual assault refers to any forced sexual contact. The stereotype of the perpetrator being a stranger has been proven inaccurate by research, with nearly two-thirds of all sexual assaults being committed by someone known to the victim.

Prevalence

Unfortunately, obtaining accurate statistics is difficult because of underreporting and unacknowledged assault. The Rape, Abuse, and Incest National Network (RAINN) and NCVS estimate that nearly 70% of sexual assaults go unreported. Over 5 years, there were 463,634 victims ages 12 or older of rape and sexual assault each year in the United States. An estimated 139,815 rapes (UCR definition) were reported to law enforcement in 2019. This estimate was 2.7% lower than the 2018 estimate and 10.8% higher than the 2015 estimate. Native Americans remain at the most significant risk of sexual violence, experiencing 5,900 sexual assaults per year. RAINN estimates that 1 in 6 U.S. women and 1 in 33 U.S. men are victims of sexual assault. An estimated 21% of TGQN (transgender, genderqueer, or nonconforming) college students have been sexually assaulted, compared to 18% of non-TGQN females and 4% of non-TGQN males. In addition to being underreported, there are many rapes that the victim does not acknowledge. They often refer to the assault in more benign terms, such as a misunderstanding, even though it meets the legal definition of rape. Research has reported that this percentage is substantial; more than 50% of sexual assaults may go unacknowledged and, thus, unreported. Victims who fail to acknowledge having been raped usually believe that rape involves two strangers and greater force, as opposed to the views of those who acknowledge having been sexually assaulted.

Risk Factors

Although females are usually victims of rape and sexual assault, greater than 10% of victims are estimated to be male. Furthermore, most experts believe that males underreport more than females. Nevertheless, females 16 to 19 years of age have four times the risk of any other group of becoming a victim of sexual assault. Eighty percent of all rape victims are younger than 30 years, and the Bureau of Justice Statistics indicates that women 16 to 24 years of age are at the most significant risk for rape in the United States. Having been a victim of childhood abuse or previous assault increases the likelihood of further assault of all types. Many studies have also shown substance use and abuse to be a risk factor for sexual assault.

Perpetrators

Rape is considered a crime of power and aggression. However, controversy remains about whether the sexual predator’s motivation is anger, power, sexual gratification, or control. Multiple research studies indicate that victims know the majority of perpetrators. Nonstrangers commit approximately 73% of all sexual assaults and 66% of rapes.

The Centers for Disease Control and Prevention (CDC) noted several factors for perpetration, including individual and relationship components. Individual risk factors included prior sexual victimization or perpetration, substance use, difficulty with empathy, aggressiveness and acceptance of violence, different sexual issues (early initiation, coercive fantasies, and impersonal sex), and hostility toward females. Relationship risk factors for perpetration included family environments characterized by physical violence, emotionally unsupportive families, associating with sexually aggressive peers, and poor parent–child relations.

DOMESTIC VIOLENCE AND INTIMATE PARTNER VIOLENCE

Domestic violence is defined as a pattern of abusive behavior in any relationship with the intent to gain power or control. It may be physical, sexual, emotional, economic, or psychological. This abuse can be directed toward children, elders, or partners. Intimate partner violence (IPV) is domestic violence by a current or former partner or spouse. Domestic violence is often broken down into two categories: “high-severity abuse,” which includes being threatened or hurt with a weapon, burned, choked, hit or kicked, and resulting in broken bones and head or internal injuries; and “low-severity abuse,” which includes being slapped, hit, or kicked without injury, but could also include bruising, minor cuts, and sprains. The CDC has defined four subtypes of IPV: physical violence, sexual violence, stalking, and psychological aggression.

Elder abuse is a variant of assault either when the abuser is an unrelated caretaker or due to domestic violence. Elder abuse can include physical, sexual, emotional, or financial abuse.

Prevalence

Prevalence estimates of domestic violence in the United States vary widely, and many studies include psychological or emotional abuse. However, one consistent statistic is that nearly 30% of women and 10% of men in the United States have experienced some form of physical or sexual abuse from an intimate partner. The National Intimate Partner and Sexual Violence Survey (NISVS) from 2011 found that 27.3% of women and 11.5% of men experienced sexual violence, physical violence, or stalking by an intimate partner during their lifetimes. The NCVS reported more than 521,870 domestic violence incidents in 2019 and 310,320 cases of IPV. The prevalence rates for both are unchanged over the previous 15 years.

Risk Factors

Studies have found that certain individual, relationship, community, and societal risk factors are associated with domestic violence and IPV. Individual risk factors may include younger age, pregnancy, low income, substance use, exposure to violence and abuse as a child, low level of education, and having traits of borderline or antisocial personality disorder. Relationship risk factors include marital discord and instability, economic stressors, the male having multiple partners, and male dominance in the family. Multiple studies have shown that community and societal risk factors such as poverty, overcrowding, acceptance of IPV or lack of sanctions against it, and traditional gender roles, or norms play roles in increased domestic violence and IPV rates.

Perpetrators

The perpetrators of domestic violence are represented across all races, cultures, ethnicities, genders, religions, educational backgrounds, and socioeconomic levels. Many studies have shown that having been a victim of abuse or witnessing abuse in the home increases the risk of someone becoming an abuser. Substance abuse is often a risk factor for episodes of perpetration. Various typologies have been used to describe perpetrators of domestic violence, grouped into two types of offenders. The first is described as being “emotionally dependent,” with a high level of jealousy, interpersonal dependency, anger, and low self-esteem. These abusers are prone to interpret their partner’s behavior as betrayal, and their anger can explode into violence. The second type is the “antisocial/narcissistic” person, who tends to have low empathy and hostile attitudes toward women. These abusers become less physiologically aroused as violence escalates and have the highest rates of alcohol dependence. Several traits can be associated with domestic violence and IPV perpetrators, including exhibiting verbal abuse, projecting blame, objectifying the victims, and isolating them from social support. Finally, abusers are noted to be charming and have manipulative personalities in public while being cruel and hostile to their intimates.

WORKPLACE VIOLENCE

Acts of violence at work include simple assault, aggravated assault, robbery, rape/sexual assault, and homicide.

Prevalence

Between 2007 and 2015, an estimated 10.3 million acts of workplace violence occurred in the United States. During that period, nonfatal workplace violence accounted for about 22% of nonfatal violent crimes. From 2014 through 2019, the number of firearm workplace homicides increased steadily after a 2-decade decline prior to that. Roughly 80% of these workplace homicides involved firearms. With some variation, total violent crime in the workplace did decline 35% between the years 2002 and 2019. This reduction correlates with a similar 37% decrease in all violent crimes (workplace and nonworkplace) during the same period.

Risk Factors

Unlike other violent crimes, workplace crime rates are higher among whites than other races. An important component of workplace violence is the relationship between the risk of violence and the type of occupation. The NCVS found that people working in law enforcement are at high risk for victimization (48 per 1,000) and constituted 19% of all workplace assaults. This included law enforcement officers, security guards, and corrections officers. Other high-risk occupations include bartenders (79.9 per 1,000), the mental health fields (20.5 per 1,000), transportation fields (12.2 per 1,000), and high school teachers (13.5 per 1,000). Preschool and college/university teachers are at the lowest risk for victimization (0.9 and 1.9 per 1,000, respectively).

Perpetrators

There are several distinct characteristics among perpetrators–of which gender is the most pronounced. As reported by the victims, four-fifths of all violent workplace crimes are committed by men, regardless of the victim's gender. The offender's race is most often white (55%), followed by black (30%). Substance use in the offender is reported more often by victims in the law enforcement, medical, and retail sales occupations as opposed to the teaching fields. Unlike domestic violence, workplace violence is more often perpetrated by strangers (53%) than by acquaintances (12%) or known intimates (1%). Mental health and education are the only fields in which attacks occur more often at the hands of a known perpetrator.

SEQUELAE OF VIOLENCE AND ASSAULT

Survivors of violence have varied reactions, but they are similar to those exposed to other types of trauma. Furthermore, the severity of the sequelae varies with the individual. However, multiple studies have shown that many people who experience violence have diminished physical and mental health, resulting in higher utilization of health care services.

The most common reported sequelae after sexual assault in women are post-traumatic stress disorder (PTSD), affective disorders, substance misuse, disordered eating, and sexual disorders. Perceived life threats and physical injury are also linked to a diagnosis of PTSD. One of the most protective factors in ameliorating the development of trauma-related symptoms is social support. Furthermore, the lack of social support or the perception of being treated differently can be quite detrimental to the survivor, causing increased trauma-related symptoms.

Domestic violence has been associated with depression, anxiety, low self-esteem, substance abuse, sexual dysfunction, functional gastrointestinal disorders, headaches, chronic pain, and multiple somatic symptoms. Research has also shown a dose-effect of escalating violence and increasing physical complaints.

Physical signs and symptoms associated with current abuse:

Loss of appetite
Frequent or serious bruises
Nightmares
Vaginal discharge
Eating binges or self-induced vomiting
Diarrhea
Fractures
Sprains or serious cuts
Pain in the pelvic or genital area
Fainting
Abdominal pain
Breast pain
Frequent or severe headaches
Difficulty in passing urine
Chest pain
Problems with sleeping
Shortness of breath
Constipation

Many studies have shown that survivors of domestic violence attempt suicide more frequently than those who have not experienced violence. Substance abuse appears to be a significant factor for both the survivor and the perpetrator. It is both a risk factor for abuse and a consequence of it.

The relationship between childhood abuse and adult abuse is complex. Childhood abuse is a risk factor for abuse as an adult, and it increases both physical and mental health sequelae of adult abuse. Survivors of sexual assault or domestic violence who are older and without a history of childhood abuse have fewer symptoms.

TREATMENT ISSUES

In evaluating assault victims, multiple treatment issues need to be addressed. Assessing safety, treating acute psychiatric and medical issues, and evaluating the need for hospitalization are primary foci of the initial assault workup. Another critical issue is the mandatory reporting requirements for abuse victims. After these treatment issues have been addressed, the clinician should assess whether the patient may benefit from psychotherapeutic or pharmacologic intervention. Evaluating assault victims requires creating a safe environment and establishing rapport reflecting the delicate nature of the circumstances. Efforts should be made to assess and respect the patient’s wishes and requests for a safe therapeutic alliance at the patient’s pace.

Initial Evaluation

Many victims of assault initially present to a medical provider to treat their injuries. In emergency departments, patients may often be treated for their injuries without the assault being recognized or addressed. Consequently, it is vital to consider any injuries as being potential stigmata of an assault. The patient may be avoidant when asked about the cause of injuries that appear related to an assault. This may be further complicated by the need to complete specific tasks (rape kit, photo documentation, reporting of legal concerns) that require the patient to describe and relive the recent assault.

Consequently, it is paramount to establish rapport while completing a thorough evaluation. The process of evaluation should be explained to the patient before initiating. Allow the patient to direct the pace and content of the interview and direct the process. The crime may represent a time when they lacked control, and similar feelings in the initial interview can trigger traumatic or anxious responses. Finally, the physician should remain attuned to nonverbal patient responses that signify discomfort and conduct the interview accordingly. This early therapeutic contact with assault victims may help influence their future prognosis. Furthermore, when a patient discloses domestic violence, this should be documented in the chart for future follow-up and possible future legal documentation.

Safety

After an assault, a safety assessment should be addressed with the patient, including the potential for suicidal and homicidal ideation pertaining to the recent assault. The patient's safety from further assault also needs to be addressed, especially in a setting where the perpetrator was an intimate partner or known acquaintance. Screening for severe psychological symptoms that would cause difficulty in self-care should be identified, including acute mood deterioration or affective instability, self-destructive behaviors, dissociative symptoms, or psychosis. If the evaluation indicates these areas are inadequate to ensure patient safety, then a plan needs to be developed and should provide for the patient's physical safety with attention to a safe space to sleep, eat, and recuperate.

Hospitalization

If the patient's safety cannot be guaranteed, hospitalization may be necessary. Common indications for hospitalization include (1) severe medical injuries, (2) suicidality or homicidality, (3) dissociative or psychotic symptoms, (4) mood instability or affective dysregulation, (5) self-destructive behaviors, and (6) continued serious threat to the patient's life or well-being. An individualized multidisciplinary treatment plan should be developed if the victim is admitted to the hospital. Safety, milieu therapy, mood stabilization, and medication evaluation are the primary treatment modalities provided in the hospital and reflect the short-stay nature of most psychiatric facilities. A coordinated inpatient-to-outpatient transition will provide ongoing treatment in the community, including psychiatric, medical, and social work issues being addressed prior to discharge.

Legal Issues

According to state law, a physician evaluating an assault victim must adhere to mandatory reporting requirements. Mandated reporting of child, elder, and developmentally disabled abuse exists in all 50 states. In cases of assault secondary to domestic violence, states do not usually require mandatory reporting; however, many require physicians to report serious injuries due to acts of criminal violence. Thus, in the setting of injury secondary to domestic violence, this can sometimes be considered a de facto mandatory reporting situation.

Continuing the Assessment

Victims of assault may report various psychiatric symptoms and psychosocial issues, including feelings of depression, anxiety, aggression, guilt, shame, and low self-esteem. Interpersonal relationships may suffer, and substance abuse may occur. These symptoms and issues should be addressed early because of the risk to the patient of future psychological sequelae. Victims of assault are at increased risk for PTSD, major depression, agoraphobia, OCD, social phobia, and simple phobia. The likelihood of these illnesses developing varies according to the type of assault and risk factors related to being assaulted. Therefore, a treatment plan must be developed that (1) assists the patient in dealing with psychiatric symptoms that they are currently experiencing because of the assault and (2) assists the patient with psychiatric symptoms that may arise later.

Postassault Psychotherapy Treatment

After the patient's safety has been ensured and the initial evaluation is complete, a variety of psychological interventions can be initiated. Cognitive-behavioral approaches are the most researched techniques demonstrating efficacy. Exposure therapy is a variant of CBT that has been shown to help victims emotionally process the assault by decreasing their fear of memories or cues of the event. Recent literature has shown that brief trauma-focused CBT may accelerate recovery in victims manifesting acute PTSD and is superior to supportive therapy. Eye movement desensitization and reprocessing (EMDR) has been shown to help patients process upsetting memories, thoughts, and feelings. These individual psychotherapies may be augmented with group psychotherapy, art therapy, dance and movement therapy, music therapy, and body-oriented approaches if these might prove beneficial to the patient.

Postassault Psychopharmacologic Treatment

Although medications are not recommended in the acute treatment of all assault victims, they may be beneficial in certain circumstances and target specific symptomatology. The clinician may decide to medicate patients with incapacitating anxiety, extreme aggression toward themselves or others, or dissociation or psychoses immediately after an assault. The patient's safety and the safety of those around the patient will help decide the need for pharmacologic intervention. Most medication treatment will be initiated much later after the assault as the patient develops symptoms of PTSD, depression, anxiety, OCD, or psychosis. Although medications may be helpful in symptom management, they should not be viewed as a replacement for psychotherapy aimed at trauma symptom resolution.

FURTHER READINGS

*Acierno R, Resnick HS, Kilpatrick DG. Health impact of interpersonal violence. 1: prevalence rates, case identification and risk factors for sexual assault, physical assault and domestic violence in men and women. *Behav Med*. 1997;23(2):53–64.

*Breiding MJ, Smith SG, Basile KC, Walters ML, Chen J, Merrick MT. Prevalence and characteristics of sexual violence, stalking, and intimate partner violence victimization—National Intimate Partner and Sexual Violence Survey, United States, 2011. *MMWR Surveill Summ*. 2014;63(8):1–18.

Bureau of Justice Statistics. *Criminal Victimization, 2019*. U.S. Department of Justice; 2020.

Bureau of Justice Statistics. *National Crime Victimization Survey*. U.S. Department of Justice; 2019.

Bureau of Justice Statistics. *National Crime Victimization Survey, Violence in the Workplace, 1993–2009*. U.S. Department of Justice; 2011.

Cantor D, Fisher B, Chibnall SH, et al. Report on the AAU Campus Climate Survey on Sexual Assault and Sexual Misconduct, 2020.

Carlson EB, Dutton MA. Assessing experiences and responses of crime victims. *J Trauma Stress*. 2003;16(2):133–148.

Doucette ML, Bulzacchelli MT, Frattaroli S, Crifasi CK. Workplace homicides committed by firearm: recent trends and narrative text analysis. *Inj Epidemiol*. 2019;6(5):1–9.

Faravelli C, Giugni H, Salvatori S, Ricca V. Psychopathology after rape. *Am J Psychiatry*. 2004;161(8):1483–1485.

Federal Bureau of Investigation. *Criminal Justice Information Services Division, Reporting Rape in 2013*. U.S. Department of Justice; 2014.

Federal Bureau of Investigation. *Uniform Crime Reports, Crime in the United States, 2013*. U.S. Department of Justice; 2014.

Flores AR, Langton L, Meyer IH, Romero AP. Victimization rates and traits of sexual and gender minorities in the United States: results from the National Crime Victimization Survey, 2017. *Sci Adv*. 2020;6(40):eaba6910.

Kilpatrick DG, Acierno R. Mental health needs of crime victims: epidemiology and outcomes. *J Trauma Stress*. 2003;16(2):119–132.
Kluft RP, Bloom SL, Kinzie JD. Treating traumatized patients and victims of violence. *New Dir Ment Health Serv*. 2000;(86):79–102.
Leserman J. Sexual abuse history: prevalence, health effects, mediators, and psychological treatment. *Psychosom Med*. 2005;67(6):906–915.
McCauley J, Kern DE, Kolodner K, Derogatis LR, Bass EB. Relation of low-severity violence to women's health. *J Gen Intern Med*. 1998;13:687–691.
McCauley J, Kern DE, Kolodner K, et al. The "battering syndrome": prevalence and clinical characteristics of domestic violence in primary care internal medicine practices. *Ann Intern Med*. 1995;123(10):737–746.
Palmero GB. The mind of the sexual predator. *Curr Opin Psychiatry*. 2007;20:497–500.
*Resnick PJ, Scott CL. Legal issues in treating perpetrators and victims of violence. *Psychiatr Clin North Am*. 1997;20(2):473–487.
*Roberts NP, Kitchiner NJ, Kenardy J, Bisson JI. Early psychological interventions to treat acute traumatic stress symptoms. *Cochrane Database Syst Rev*. 2010;(3):CD007944.
Siegel M, Johnson CY, Lawson CC, Ridenour M, Hartley D. Nonfatal violent workplace crime characteristics and rates by occupation—United States, 2007–2015. *Morb Mortal Wkly Rep*. 2020;69:324–328.
Sijbrandij M, Olff M, Reitsma JB, Carlier IVE, de Vries MH, Gersons BPR. Treatment of acute posttraumatic stress disorder with brief cognitive behavioral therapy: a randomized controlled trial. *Am J Psychiatry*. 2007;164(1):82–90.
*Tharp AT, DeGue S, Valle LA, et al. A systematic qualitative review of risk and protective factors for sexual violence perpetration. *Trauma Violence Abuse*. 2012;14(2):133–167.
Ullman SE, Filipas HH. Predictions of PTSD symptom severity and social reactions in sexual assault victims. *J Trauma Stress*. 2001;14(2):369–389.

30.4 Complementary, Alternative, and Integrative Approaches in Mental Health Care

JAMES H. LAKE, M.D.

DEFINITIONS AND HISTORICAL BACKGROUND

The National Center for Complementary and Integrative Health (NCCIH) of the National Institutes of Health defines complementary and alternative medical (CAM) practices as "those healthcare and medical practices that are not currently an integral part of conventional medicine.” In the United States and other Western countries, where biomedicine is the dominant healing tradition, *alternative* treatments are approaches for which a mechanism of action is not supported by *current* biomedical theory. Acupuncture is regarded as an alternative form of treatment in Western countries because Chinese medicine postulates the existence of an energetic principle called “*qi*” that is not congruent with the assumptions of current science. In contrast, complementary therapies are based on mechanisms that are empirically verifiable by current science and consistent with biomedical theory but remain controversial or are not endorsed by biomedicine for ideologic or other reasons. St. John wort (SJW) is presently regarded as a complementary treatment in the United States. Although the postulated mechanism of action of the herbal is consistent with biomedical theory and well-designed studies have demonstrated beneficial effects in depressed mood, there is an ongoing debate in biomedical psychiatry over the appropriate uses of SJW in the treatment of depressed mood. Integrative medicine is a rapidly evolving subspecialty that encompasses both biomedicine and CAM and is personalized to the unique history, needs, and values of each patient.

In North America, Europe, and other industrialized world regions, an increasing percentage of the population is using complementary, alternative, and integrative approaches to treat or *self-treat* medical and mental health problems. Large patient surveys have reported that females, college graduates, and persons who believe they receive poor health care from conventionally trained providers are more likely to use CAM and that roughly one-half of individuals who use CAM notify their health care providers. The majority of CAM users are well educated, committed to personal growth, satisfied with conventional care, and use both medications and CAM for the same problem. An increasing number of medical schools, nursing schools, and psychology graduate programs offer courses on CAM. Symposia on CAM modalities are included in the annual meetings of the American Medical Association (AMA), the American Psychiatric Association (APA), and other professional medical associations. Approximately half of U.S. physicians—and the majority of European physicians—believe that acupuncture, chiropractic, and homeopathy are valid therapeutic modalities and refer patients to practitioners of these therapies. Increasing numbers of primary care physicians are becoming certified to practice massage, acupuncture, herbal medicine, homeopathy, and other nonallopathic modalities.

Persons diagnosed with a major psychiatric disorder are significantly more likely to use CAM treatments than the general population, and the majority of people who use CAM to self-treat a mental health problem take one or more psychotropic drugs concurrently. Almost one-half of persons diagnosed with major depressive disorder or panic disorder use at least one CAM treatment, compared to less than one-third of the general adult population. Roughly two-thirds of severely depressed or anxious persons who use CAM consult with a mental health professional, and 90% of the time, this is a psychiatrist. The majority of persons who use CAM therapies for a mental health problem do not disclose this to their mental health care provider. More than one-half of persons who self-treat severe depression or anxiety using CAM take one or more psychotropic medications concurrently, resulting in potentially serious safety problems.

LEGAL AND ETHICAL ISSUES

CAM use in mental health care raises important legal and ethical issues for psychiatrists and allied mental health professionals. The increasing use of CAM in mental health care is resulting in competing demands of patient expectations, clinical judgment, and liability considerations. Clinicians who recommend or prescribe CAM often do so in the absence of institutional oversight or professional guidance. This is complicated by the absence of treatment guidelines, policies, and procedures to help clinicians deal with liability issues.

Scope of practice refers to legally authorized boundaries of care for a particular conventional medical specialty or CAM practice. There is no clearly established legal–ethical framework that defines the scope of practice for physicians or CAM practitioners who treat patients using CAM modalities. Furthermore, substantial differences may exist between state jurisdictions for the kinds of medical practices that are legally sanctioned for a particular CAM specialty. For example, chiropractors are authorized to give nutritional advice in some states, but not others and massage therapists are prohibited from engaging in mental health counseling in most states but not all. All health care providers are ethically obligated to provide competent professional care. This duty includes the demonstration of professional competence when treating patients and the exercise of sound judgment when referring a patient to other conventionally trained medical specialists or CAM practitioners. Medical practitioners who use interventions that are not recognized as falling within the scope of their professional competence may invite disciplinary action by the state medical board or other regulatory agencies if patients or professional peers make formal complaints to those bodies. Disciplinary action may result in probation, suspension, or revocation of a professional license to practice.

It is prudent for health care providers who recommend or prescribe CAM approaches to first learn about all applicable restrictions imposed by their state medical boards on the scope of medical practices in their respective states. When a patient's medical or mental health problem is outside of the scope of the provider's expertise and experience, the provider is ethically obligated to refer the patient to an appropriate and competent provider. A conventionally trained provider who refers a patient to a CAM practitioner assumes liability for negative outcomes resulting from the referral, including harmful effects that may result from treatment. Referrals should be made only after confirming that a particular CAM practitioner is qualified to practice in their specialty. Malpractice refers to negligence, which is defined as failure to use due care in treating a patient, resulting in injury to the patient. Courts make rulings on malpractice using different standards of care for physicians and CAM practitioners. The issue of malpractice insurance coverage pertaining to the practice of CAM is poorly defined. States vary significantly in their requirements for malpractice coverage, and in a given state, different insurers often vary in their coverage of the same CAM therapy. Statutes pertaining to fraud in health care are intended to prevent intentional deception of patients by physicians and other providers. Submitting a reimbursement claim for a treatment that is medically unnecessary may lead to charges of fraud.

SAFETY

Evaluating the safety of CAM and integrative approaches poses complex challenges. Many herbal treatments are widely regarded as safe because of their long-standing historical use in traditional healing practices. There are limited safety data on the majority of natural supplements used to treat medical or mental health conditions. The safety of imported herbal medicines and other natural supplements used for medicinal purposes from Asia is generally more difficult to assess than Western botanicals because of the absence of strong import controls.

Herbals and other natural supplements may cause adverse effects and potentially interact with other supplements or drugs with harmful consequences. In contrast, few safety risks are associated with somatic therapies, mind–body practices, and the so-called “energy” healing. Most safety information on CAM comes from case reports of limited reliability, and published reviews of adverse effects often cite conflicting findings. SJW has the greatest number of reported cases of adverse effects because it is one of the most widely used CAM treatments of depressed mood. In contrast, relatively little is known about adverse effects and toxicities associated with the majority of natural supplements. Safety information on natural supplements is especially incomplete with regard to young children, pregnant and nursing women, and patients with serious medical illnesses. As already noted, many individuals who take medications self-treat using natural supplements without disclosing this information to their prescribing physician, resulting in unknown risks of toxic interactions and underreporting of potentially serious interactions between medications and natural supplements.

Patients should be advised about adverse effects or toxic interactions before taking *any* natural supplement alone or in combination with *any* medication. Table 30.4–1 shows adverse effects, potential interactions, and contraindications of select natural supplements widely used in North America, Western Europe, and Australia to treat or self-treat psychiatric disorders.

Rare cases of injury or adverse effects have been reported with CAM therapies that are not based on discrete biologic mechanisms of action, for example, therapeutic massage, acupuncture, yoga, *qigong,* and Reiki. Where applicable, these safety concerns are reviewed in the text.

OVERVIEW OF CAM ASSESSMENT APPROACHES IN MENTAL HEALTH CARE

Recent advances in neuroscience suggest that the causes of mental illness are more *subtle* and more *complex* than the prevailing biopsychosocial model is capable of elucidating. Models of consciousness based on complex systems theory, quantum mechanics, and other emerging theories may be consistent with the assumptions of “energetic” assessment used in Chinese medicine and other nonallopathic systems of medicine. CAM assessment approaches may eventually permit a more *complete* “picture” of underlying psychological, biologic, and *energetic* causes of symptoms. CAM approaches used to assess the causes of mental illness include assays of biologic structure or function, measurements of scientifically validated forms of energy or information, and measurements of postulated *subtle* forms of energy or information that cannot be explained by current science. Quantitative electroencephalography (QEEG) and heart rate variability (HRV) analysis are examples of scientifically validated assessment methods that provide useful clinical information about functional dysregulations of electromagnetic activity in the brain and heart related to symptoms of mental illness. Assessment approaches that purport to detect *subtle* energetic phenomena associated with both physical and mental illness include pulse diagnosis in Chinese, Tibetan, and Ayurvedic medicine and analysis of electrodermal potentials following acupuncture or medical *qigong.*

Table 30.4–1.
Safety Issues Associated With Botanicals and Other Natural Products

Natural Product	Adverse Effects	Potentially Toxic Interactions and Contraindications
Ginkgo biloba	May increase bleeding risk if used before surgery or labor/delivery Gastrointestinal discomfort, skin allergy, dizziness, headache	**Contraindicated** in combination with warfarin due to interference with platelet activation factor (PAF)
Kava (*Piper methysticum*)	Gastrointestinal discomfort (generally mild), allergic skin rashes, mild headache Case reports of severe hepatoxicity (always recommend routine liver function tests for all patients taking kava products)	Warn against combined use with benzodiazepines Case reports of lethargy and disorientation when used with alprazolam, other benzodiazepines
St. John wort (*Hypericum perforatum*)	Uncommon adverse effects include mild stomach discomfort, allergic skin rashes, tiredness, restlessness Rare case reports of mania Rare case reports of mild serotonin syndrome	**Contraindicated** in combination with cyclosporine, other immunosuppressive agents, oral contraceptives, warfarin, protease inhibitors, and other antiretroviral drugs (may induce cytochrome P450 hepatic enzymes, causing increased metabolism and decreased serum levels of these drugs)

Valerian (*Valeriana officinalis*)	Uncommon reports of morning grogginess; headache Caution against driving for up to a few hours after using	Warn against combined use with benzodiazepines or barbiturates (may potentiate central nervous system sedating effects)
L-Tryptophan	Uncommon adverse effects include drowsiness, dry mouth, and blurred vision Fatal cases of eosinophilia-myalgia syndrome (EMS) attributed to a single contaminated batch of L-tryptophan in the late 1980s L-Tryptophan remains in widespread use and no subsequent cases of EMS have been reported	Rare case reports of possible serotonin syndrome when combined with SSRIs and other antidepressants
5-Hydroxy-tryptophan (5-HTP)	Case reports of mild nausea, dry mouth, stomach irritation, and drowsiness at commonly used doses (50–100 mg thrice daily)	No case reports of serotonin syndrome when used in combination with SSRIs, other antidepressants
S-Adenosyl-methionine (SAMe)	Well tolerated Dose-related gastrointestinal distress is most common side effect (can range from mild	Case reports of serotonin syndrome when used in combination with SSRIs, other antidepressants

	nausea to vomiting and diarrhea) Case reports of mild nervousness and insomnia	Warning: Risk of mania induction in bipolar patients; should be avoided in this population
Omega-3 fatty acids	Well tolerated at low doses, but there are reports of nausea and loose stools at higher doses; gastrointestinal side effects are reduced by taking with meals	Mild blood-thinning activity but no bleeding problems caused when used alone or with aspirin Warning: Increased risk of hypervitaminosis A and D with chronic high doses of cod liver oil because of high concentrations of vitamins A and D Caution: Pharmaceutical-grade products are recommended to minimize the risk of contamination with heavy metals, polychlorinated biphenyls, organochlorines

SSRI, selective serotonin reuptake inhibitor.

Functional medicine is an interdisciplinary field that uses established quantitative analysis methods to assess relationships between nutritional status, neurotransmitters, endocrine and immune function, and psychological symptoms. Functional tests used to evaluate mental health problems include urinary assays of neurotransmitters and their metabolites and serum assays of cholesterol, triglycerides, vitamins, minerals, amino acids, and their metabolites, hormones, fatty acids, proinflammatory cytokines (e.g., interleukin-6 [IL-6], IL-8, and IL-1b), and immunologic factors. Discrete biomarkers of specific psychiatric disorders have not been identified, and functional assessment findings generally do not provide insights about specific treatments that improve outcomes. For example, although low serum folate has been causatively linked to depressed mood, research findings are inconsistent, and this *apparent* relationship may be due to chance, bias, confounding, or reverse causality. Preliminary findings suggest that total cholesterol and triglyceride levels may be inversely correlated with the severity of depressed mood. Depressed patients who attempt suicide frequently have abnormally low serum total cholesterol and triglyceride levels, which may provide useful clinical markers for suicide risk. Nonsuicidal depressives tend to have cholesterol levels in the vicinity of 180 mg/dL, while severely depressed suicidal patients tend to have significantly lower serum total cholesterol levels in the vicinity of 150 mg/dL. On the basis of these findings, it is prudent to check cholesterol levels of severely depressed patients, make appropriate dietary recommendations, and address issues related to the use of cholesterol-lowering medications. In contrast to abnormal low cholesterol in severely depressed patients, *elevated* serum cholesterol may be a sensitive marker for anxiety disorders, including generalized anxiety disorder (GAD), panic disorder, and possibly OCD. Amelioration of anxiety in response to anxiolytics is often associated with a reduction in serum total cholesterol to normal levels. Assessment of vitamin deficiency states is important when working with patients who do not take supplements or who are at risk of one or more vitamin deficiency syndromes because of chronic malnutrition or a medical problem that interferes with normal absorption or metabolism.

Low serum zinc levels result in impaired zinc transport to the brain and have been cited as a possible factor in the formation of amyloid fibrils. Consistently reduced red blood cell (RBC) membrane levels of two fatty acids—arachidonic acid (AA) and DHA—have been reported in nonmedicated schizophrenics compared to medicated schizophrenics and healthy controls. Low brain levels of fatty acids and cholesterol may indirectly affect the number and activity of neurotransmitter receptors by modifying their three-dimensional structure in the lipid bilayer of nerve cell membranes. Assessing flushing response following niacin challenge may provide useful information when evaluating individuals with symptoms of psychosis. Research findings suggest that normal flushing response to niacin is attenuated in schizophrenia and other psychotic disorders. Approximately 80% of patients with schizophrenia do not flush when challenged with niacin compared to healthy controls. In one series, 43% of patients with schizophrenia experienced abnormally low vasodilation in response to a 200-mg challenge dose of niacin, compared to only 6% of individuals diagnosed with bipolar disorder. The mechanism of action probably involves dysregulation of phospholipid-dependent signal transduction and increased phospholipase A2 activity.

There is preliminary evidence for the clinical utility of HRV monitoring and QEEG in psychiatric assessment; however, both approaches are still mainly used for research. Abnormal EEG findings occur in up to 40% of depressed patients, and *small sharp spikes* are often seen in severely depressed suicidal patients. QEEG analysis of unipolar depressed patients typically reveals increased alpha or theta power and decreased interhemispheric coherence. In contrast, bipolar depressed patients often have reduced alpha activity and increased beta power. Reduced alpha activity and increased beta power are more likely in bipolar disorder than in depression. EEG changes frequently associated with anxiety include decreased alpha activity in generalized anxiety, increased theta activity in OCD, and paroxysmal activity in patients who experience panic attacks. QEEG findings may predict differential medication response rates in patients with OCD. In one series, approximately 80% of OCD patients who exhibited increased alpha power *responded* to SSRIs, compared to 80% of OCD patients with increased theta activity who *failed* to respond to SSRIs. QEEG findings associated with other anxiety disorders are highly inconsistent. Neurometric brain mapping is a specialized QEEG approach that compares EEG characteristics of the individual being evaluated with normative databases for the same age group. Neurometric mapping helps clarify functional brain correlates of cognitive impairment and yields useful information when planning EEG biofeedback protocols addressing different symptoms (see discussion of EEG biofeedback). Neurometric brain mapping is being increasingly used to differentiate cognitive impairments due to head injuries, medical disorders, progressive dementia, alcohol or substance abuse, depressed mood, and learning disorders.

Cordance is a measure of localized electrical brain activity relative to averaged brain EEG patterns. More than one-half of patients with severely depressed mood who subsequently had the highest response rates to SSRIs showed significant decreases in prefrontal cordance during the first 48 hours of therapy, suggesting that improvement was due to normalization of abnormal low CNS serotonin levels. In contrast, nonresponders to SSRIs or placebo did not show negative prefrontal cordance following treatment, suggesting that low serotonin was not the primary cause of depressed mood in these cases. It has been suggested that differences in prefrontal cordance may also predict response differences of depressed patients to homeopathic remedies.

Assessment approaches used by energy healers are based on the belief that *imbalances* in postulated subtle energies manifest as psychological symptoms. Analysis of the pulse is used to assess energetic imbalances associated with both physical and mental symptoms in Chinese medicine, Ayurveda, and Tibetan medicine. A treatment plan consisting of psychological, physical, and energy “balancing” therapies are constructed with the goal of normalizing energetic imbalances and improving health. Anecdotal reports suggest that subtle characteristics of the pulse described in Chinese, Ayurvedic, or Tibetan medical theory correspond to specific energetic *imbalances* that manifest as physical or psychiatric symptoms. For example, mild to moderate symptoms of anxiety and depressed mood are often associated with a “tight” or rapid rate when the pulse is palpated at the pericardium position or as a subtle sensation of “smooth vibration” over the entire pulse at all depths. More severe symptoms of mental illness are often associated with a “rough vibration” over the entire pulse or a sensation of “slipperiness” in some aspects of the pulse. Severe psychiatric symptoms such as psychosis and acute mania are reportedly associated with the most abnormal energetic pattern of the pulse.

Electrodermal skin testing (EDST), the measurement of electrodermal potentials at specific acupoints, may provide useful information when assessing clinical responses to acupuncture or *qigong.* EDST may also help determine whether conventional or subtle energy treatments of anxiety are likely to be effective for a particular patient. Healing Touch (HT) practitioners conducted six trials comparing anxiolytic effects of applied relaxation with HT and concluded that EDST reliably measured changes in “bioenergy” in meridians following both conventional and subtle energy treatments of anxiety.

COMPLEMENTARY, ALTERNATIVE, AND INTEGRATIVE APPROACHES USED TO TREAT

MENTAL ILLNESS

Advances in the understanding of the pathogenesis of mental illness have important implications for therapeutic uses of natural supplements for treating depressed mood, anxiety disorders, bipolar disorder, schizophrenia, and other disorders. Select natural supplements and other CAM modalities such as regular exercise, bright light therapy, and biofeedback are based on well-understood mechanisms of action and are substantiated by positive findings from large placebo- or sham-controlled studies. However, many CAM modalities reviewed in this chapter presently lack compelling evidence. Studies on "energy" therapies, including Reiki, *qigong,* HT, and energy psychology, pose complex challenges to contemporary Western-style research methods. A comprehensive understanding of clinical benefits associated with light, electricity, and sound, as well as *"energy"* therapies, may require conceptual advances in the foundations of science and greater openness to emerging models of physics, including complex systems theory, quantum mechanics, and other nonclassical theories.

Approaches used as treatments in both biomedicine and nonallopathic systems of medicine can be assigned to five general categories corresponding to differences in known or postulated mechanisms of action (Table 30.4–2):

- *Biologic treatments* operate through one or more discrete biologic mechanisms of action at the molecular or cellular level (examples include conventional medications, vitamins, minerals, medicinal herbs, amino acids, and omega-3 fatty acids).
- *Somatic therapies* improve general well-being and affect the body as a whole but do not operate through a discrete (i.e., molecular or cellular) biologic mechanism of action (examples include exercise, massage, and dance therapy).
- *Mind–body and mindfulness practices* use a particular mental technique or combine a mental technique with physical training with the goal of optimizing mental or mind–body functioning and alleviating symptoms (examples include yoga and mindfulness-based stress reduction [MBSR]).
- *Scientifically validated forms of energy or information* have been investigated and verified by research studies. Examples include electromagnetic fields, bright light, dim light, microcurrent electrical

stimulation, high-density negative ions, music, binaural sound, and biofeedback. Specific therapeutic applications use different durations, frequencies, and power characteristics of these established forms of energy to achieve beneficial effects.

- *Postulated forms of energy or information* have not been validated by (current) science but are theorized to be the mechanisms underlying “energy” healing approaches, including Reiki, *qigong,* prayer, HT, Therapeutic Touch (TT), and energy psychology.

Table 30.4–2.
Nonconventional Modalities Reviewed in This Section

Category	Modality
Biologic treatments	Dietary modification Vitamins Minerals Amino acids Essential fatty acids Nutrient formulas Botanicals Essential oils Hormones and prohormones
Somatic therapies	Exercise Expressive movement therapy Animal-facilitated psychotherapy Snoezelen Massage
Mindfulness training and mind–body practices	Applied relaxation Yoga Meditation
Energy–information modalities validated by Western science	Music and sound Light High-density negative ions Electroencephalography (EEG) biofeedback Heart rate variability (HRV) biofeedback Virtual reality graded exposure therapy Microcurrent electrical stimulation
Energy–information modalities not validated by Western science	Acupuncture Homeopathy Healing touch and therapeutic touch *Qigong* Reiki Energy psychology Craniosacral therapy and somatoemotional release

Identifying Appropriate CAM Modalities

Deciding on which CAM or integrative approach to recommend to a patient poses many complex issues. When formulating clinical treatment recommendations, five basic issues should be kept in mind:

- The symptom or disorder that is the focus of clinical attention.
- The patient's history of response to previous treatments for the same mental health problem.
- The relative strength of evidence for various biologic, somatic, mind–body, and energy therapies.
- Practical and personal considerations, including safety, cost, availability of treatment, and cultural and personal preferences that will lead to realistic and acceptable treatment recommendations.
- Appropriate methods for assessing treatment response.

REVIEWS OF COMPLEMENTARY, ALTERNATIVE, AND INTEGRATIVE APPROACHES

The balance of this section is devoted to a review of the evidence for specific CAM and integrative modalities in current use to treat or *self-treat* common mental health problems. Because of space constraints, only modalities that have been validated by research findings are discussed in detail, whereas CAM and integrative approaches supported by limited evidence are briefly reviewed.

Pharmacologic and Related Treatments

Dietary Modification. Dietary patterns that result in a suboptimal intake of specific macro- and micronutrients such as omega-3 fatty acids, magnesium, methylcobalamin (B12) and zinc, negatively impact the microbiome and gut–brain axis and are associated with increased risk of depressed mood. Reducing or eliminating refined sugar and caffeine can significantly improve mood in some depressed patients. Caffeine consumption increases serum epinephrine, norepinephrine, and cortisol levels, resulting in general feelings of "nervousness" and increased risk of panic attacks in predisposed individuals. Foods rich in folate, pyridoxine

(B_6), and methyl-cobalamin (B_{12}) such as legumes, leafy greens, and beets may have mood-enhancing benefits as these vitamins function as essential cofactors for the synthesis of serotonin, dopamine, and norepinephrine. Important dietary sources of omega-3 fatty acids include salmon, halibut, other deep-sea fish, and flaxseed oil. Relatively greater intake of omega-3 fatty acids (i.e., compared to omega-6 fatty acids) is correlated with lower C-reactive protein levels and reduced risk of inflammatory-mediated disorders, including possibly depression and Alzheimer disease. Prevalence rates of severe depression and suicide attempts are lower in countries where fish is an important part of the average diet. In Japan, where fish consumption is very high, the annual incidence of depressed mood is only 0.12%, compared to 6% in New Zealand, where fish consumption is relatively low. A large epidemiologic study concluded that fish consumption two to three times weekly significantly reduced the risk of cognitive decline in the older adults. High fish consumption was inversely correlated with cognitive impairment. In contrast, individuals who preferred foods rich in linoleic acid (an omega-6 fatty acid) exhibited significantly higher rates of cognitive decline. Lifestyle factors—including exercise and alcohol consumption—and educational level are associated with dietary preferences, and thus a relatively lower risk of developing Alzheimer disease cannot be ascribed to fish consumption alone.

Excess caloric intake and high-fat intake promote the formation of free radicals that cause diffuse neuropathologic changes in the brain. High-fat, high-calorie diets correlate with a significantly greater risk of developing Alzheimer disease compared to diets with lower fat and fewer calories. A meta-analysis of findings from 18 community-wide studies concluded that the risk of Alzheimer disease increased linearly at a rate of 0.3% with every 100-calorie increase in daily intake. In the same meta-analysis, the average daily fat consumption was strongly correlated with an increased risk of developing dementia. Epidemiologic studies suggest that moderate but not heavy consumption of wine (two to four glasses per day) is associated with a reduced risk of Alzheimer disease. In a large 5-year cohort study, 19,000 women aged 70 to 80 years who drank 14 g of alcohol daily (the equivalent of one and a half glasses of wine or one 12-oz beer) were 20% less likely to experience significant cognitive impairment compared to nondrinkers. Ketogenic diets intended to shift the metabolism from primarily glycolysis to ketosis are sometimes used in individuals with treatment-refractory seizure disorders. Limited findings, mainly from animal studies, suggest that the ketogenic diet may have both antidepressant and anxiolytic effects.

Chronic alcohol abuse is often associated with malnutrition due to malabsorption of essential nutrients through the mucosa of the stomach and small intestines resulting in significantly reduced blood levels of thiamin, folate, B_6, and important trace minerals. Hypoglycemia can result from toxic effects of alcohol on the liver, manifesting as confusion, anxiety, and impaired cognitive functioning. Alcoholics who improve their overall nutrition may have a better chance of maintaining sobriety than those who do not. In one study, 80% of alcoholics who stayed on a caffeine-free, low-refined sugar diet high in wheat germ and fruit remained abstinent at 6 months.

Probiotics. The putative health and mental health benefits of probiotics have been the focus of intensive research in recent years. Emerging findings suggest that the pathogenesis of many neuropsychiatric disorders, including autism spectrum disorders, depressed mood, anxiety disorders, and schizophrenia involves dysregulation of the microbiota–gut–brain (MGB) axis. It has been suggested that reported antidepressant effects of probiotics are mediated by immune modulation, anti-inflammatory effects, and modulation of neurotransmitter pathways, including serotonin, glutamate, dopamine, GABA, norepinephrine, and acetylcholine. Inter-individual variation in gut microbiota compositions may call for a highly individualized approach to selecting a probiotic formula that is most likely to yield therapeutic benefits for depressed mood or other mental health conditions.

Preliminary findings suggest that taking probiotics may have therapeutic effects on mild to moderate depressive symptoms but not more severe forms of depressed mood. In a small 8-week RCT ($N = 71$), adults with mild to moderate depressed mood randomized to a patented probiotic supplement experienced a significantly greater reduction in symptom severity compared to placebo. The probiotic formula did not significantly alter gut microbiota; however, a significant correlation was found between a specific microorganism (*Ruminococcus gnavus*) and depressed mood. The significance of findings was limited by the small study size and a high attrition rate. A meta-analysis of placebo-controlled studies on probiotic supplementation in depressed adults (total $N = 1{,}349$) found nonsignificant overall differences in mood in subjects taking probiotics versus placebo. A comprehensive review of probiotic supplementation in prophylaxis and treatment of depressive and anxiety disorders found inconsistent evidence for reductions in both depressed mood and anxiety. A 1-year placebo-controlled RCT (*N* = 423) found that supplementation with a probiotic formula containing *Lactobacillus rhamnosus* HN001 in pregnant and postpartum women significantly reduces the severity of depressed mood and anxiety and lowers the risk of postpartum depression. Findings of studies on probiotics in neuropsychiatric disorders are limited by small study size, methodologic problems, and the absence of clinical populations in most studies. Different outcomes in response to probiotics are poorly understood and may reflect differences in probiotic formulas, study populations, and study design. Further research is needed to establish optimal formulations and dosages of probiotics for the treatment of depressed mood. Probiotics are generally well tolerated in healthy individuals. Probiotics are contraindicated in individuals who are immune-compromised or have serious illnesses as they may result in severe medical complications, including pneumonia, endocarditis, and sepsis.

Vitamins

B-FAMILY VITAMINS. Folate is required for the synthesis of *S*-adenosylmethionine (SAMe) from homocysteine. Folate deficiency results in increased serum homocysteine levels and an increased risk of depressed mood. Folate in the form of L-methylfolinic acid does not require methylenetetrahydrofolate reductase (MTHFR) to become activated, more readily crosses the blood–brain barrier, and is a more effective adjuvant than dietary folate. In one study, clinical improvement in depressed patients treated with an SSRI and L-methylfolate (0.5 to 1 mg) was 30% greater than that in matched patients treated with an SSRI only. A systematic review of controlled studies (total *N* = 247) concluded that folate augmentation (1 to 15 mg per day) enhanced the efficacy of antidepressants. The findings of a double-blind study (*N* = 75) suggest that folic acid at 200 to 400 μg per day might enhance the therapeutic benefits of lithium carbonate in acutely manic patients.

Widespread folate deficiency in patients with schizophrenia is probably caused by chronic malnutrition and the effects of antipsychotic medications on absorption. Abnormal low brain folate levels in patients with schizophrenia may result from a heritable deficiency in glutamate carboxypeptidase II (GCPII), the enzyme required for folate absorption through the gut. Patients with schizophrenia who take daily folic acid in the form of methylfolate 15 mg adjunctive to their antipsychotic medication may have fewer positive and negative psychotic symptoms and may respond more rapidly than patients taking antipsychotics alone. Low serum levels of folic acid, niacin, and thiamin are associated with cognitive impairment in general. Individuals with dementia frequently have normal RBC folate levels and fail to improve with supplementation; however, the evidence for vitamin B supplementation as a treatment of cognitive impairment, including dementia, is inconclusive. In a prospective study ($N = 370$), a deficiency of folate or B_{12} doubled the risk of developing Alzheimer disease. However, a systematic review of four studies found weak evidence supporting the use of folic acid with or without B_{12} as a treatment of dementia or other forms of severe cognitive impairment. Because folate supplementation can mask the macrocytic anemia caused by B_{12} deficiency, patients should be checked for B_{12} deficiency before starting folate. However, supplementation with L-methylfolinic acid does not mask B_{12} deficiency and can be started without first checking B_{12} levels.

Vitamin B_{12} is often recommended for older adult nondemented individuals who complain of impaired cognition; however, research findings are inconsistent. In a pilot study ($N = 22$), 18 older adult patients with abnormally low serum B_{12} levels and cognitive impairment were treated with B_{12} (1 mg/day intramuscularly) for 1 week, followed by weekly 1-mg injections for 1 month, then monthly 1-mg injections for 6 months. Patients who had been cognitively impaired for less than 1-year experienced significant improvements in cognitive function, whereas patients who had been impaired for longer than 1 year became worse, suggesting that supplementation may ameliorate symptoms related to B_{12} deficiency only in the early stages of cognitive impairment. Chronic alcoholics are often deficient in thiamin, folate, B_6, and B_{12} because of the toxic effects of alcohol on the mucosal lining of the stomach and small intestine and may benefit from supplementation with a broad-spectrum vitamin B supplement. Thiamin is essential for the metabolism of acetylcholine and its release from presynaptic neurons. Depressed patients report improved mood and energy with daily thiamin supplementation of 50 mg. Thiamin (vitamin B_1) 500 mg TID adjunctive to an antipsychotic may lessen symptom severity in patients diagnosed with schizophrenia. Thiamin-dependent enzymes may be abnormal in Alzheimer disease; however, research findings on high-dose thiamin (3 to 8 g per day) supplementation in Alzheimer disease are inconsistent. Animal studies suggest that low serum thiamin levels are associated with increased alcohol craving. Niacin in the form of 1.25 g per day of nicotinamide taken with a meal before drinking may protect the liver against the acute toxic effects of alcohol in individuals who have relapsed. Early double-blind RCTs reported that large doses of niacin (up to 3 to 8 g per day) ameliorated symptoms of schizophrenia. A combined regimen of high-dose niacin and a conventional antipsychotic medication reportedly resulted in differentially greater improvement in positive psychotic symptoms compared to schizophrenic patients treated with an antipsychotic alone. A subsequent review of 53 trials on niacin and other vitamins in the treatment of schizophrenia and other chronic psychiatric disorders identified methodologic flaws in early studies and concluded that findings reached only marginal statistical significance. Nevertheless, high-dose niacin has become standard therapy for schizophrenia in orthomolecular psychiatry, and case reports of dramatic outcomes continue to be published. A retrospective analysis of negative findings of controlled trials reported that nonmedicated patients with schizophrenia in the early acute phase of illness may experience significant and sustained benefits from high doses of niacin, in contrast to medicated patients with schizophrenia or individuals in the chronic phase of the illness.

Inositol is regarded as a member of the B family of vitamins and is the precursor of phosphatidyl inositol, a second messenger required for normal neuronal function. A systematic review and meta-analysis of seven RCTs ($N = 242$) on inositol in mood and anxiety disorders found nonsignificant benefits for inositol over placebo for depressed mood, OCD, PTSD, and generalized anxiety, and a trend toward superior response over placebo for premenstrual dysphoric disorder. In a 4-week, double-blind, crossover study ($N = 21$), inositol at 12 g per day and imipramine were equally effective in reducing the frequency and severity of panic attacks in individuals diagnosed with panic disorder with and without agoraphobia.

VITAMIN D-3 (CHOLECALCIFEROL). Vitamin D_3 has anti-inflammatory, immunomodulatory, and neuroprotective effects and increases levels of the rate-limiting enzyme needed for the synthesis of serotonin, dopamine, and norepinephrine. Vitamin D deficiency is common in Western populations—especially among the older adults—and may be associated with an increased risk of depression in susceptible populations. Preliminary findings suggest that daily supplementation with vitamin D_3 at doses of 400 to 800 IU may reduce the severity of depression in patients diagnosed with seasonal affective disorder when serum D_3 levels are raised to 35 ng/mL or greater. Findings of RCTs suggest that vitamin D supplementation is more effective against major depressive disorder than milder forms of depressed mood. A meta-analysis of RCTs on vitamin D supplementation in MDD identified 4 trials (total $N = 948$) and concluded that vitamin D at doses of 1,500 to 7,000 IU/day resulted in moderate improvement in depressed mood. Parenteral vitamin D was no more effective than oral vitamin D. Daily supplementation with high doses of vitamin D up to 100,000 IU per day may be more effective than bright light therapy in the treatment of seasonal depressed mood.

Minerals. Magnesium is essential for maintaining neuronal cell membrane fluidity and neuroplasticity and plays a central role in the brain’s antioxidant protective mechanisms. Magnesium deficiency is an established risk factor for overall poor mental health, depressed mood, and cognitive impairment. Zinc has general neuroprotective effects, is involved in hippocampal neurogenesis, and modulates the activity of the hypothalamic-pituitary axis. Increased risk of depressed mood associated with abnormal low zinc serum levels may be mediated by dysregulation of the immune system. Dietary sources of zinc include seafood and red meats; hence strict vegetarian diets may result in zinc deficiencies.

Research findings on the mood-enhancing effects of magnesium and zinc are inconsistent. A systematic review and meta-analysis that included 11 RCTs on zinc depressed mood found that zinc supplementation 25 mg/day resulted in modest improvements in symptom severity. A different systematic review found no evidence for antidepressant effects of magnesium supplementation at doses between 225 and 4,000 mg/day. A narrative review of preclinical and human studies investigating putative antidepressant effects of supplementation with magnesium, zinc, or selenium found modest evidence for zinc and inconsistent findings for magnesium and selenium. Preliminary evidence from randomized controlled studies supports the antidepressant efficacy of zinc supplementation used as an adjuvant to prescription antidepressants. Animal studies and preliminary human findings suggest that abnormally high copper levels may be a causative factor in the pathogenesis of Alzheimer disease by damaging the molecule required to eliminate amyloid-beta plaques soon after they are formed.

The trace element selenium is believed to reduce the risk of cancer, cardiovascular disease, and infectious diseases, including HIV. It is estimated that a significant percentage of the world's population is selenium deficient because of geographic differences in selenium in the soil. Selenoproteins play important roles in healthy brain function. The findings of animal studies suggest that severe selenium deficiency may cause degeneration of the GABA neurotransmitter system, increasing the risk of Alzheimer disease, Parkinson disease, and other neuropsychiatric disorders. Limited findings suggest that selenium supplementation may reduce symptoms of anxiety and depressed mood in some populations. In a 5-week double-blind RCT ($N = 50$), healthy adults randomized to selenium 100 mcg versus placebo reported improvements in mood and decreased anxiety. In a 12-month RCT ($N = 63$), HIV^+ adults were randomized to selenium 200 mcg/day reported generally increased vigor, reduced levels of anxiety but no improvements in measures of depressed mood.

Amino Acids

5-HYDROXYTRYPTOPHAN, L-TRYPTOPHAN. 5-hydroxytryptophan and L-tryptophan have been extensively investigated for their putative antidepressant and anxiolytic effects. Therapeutic dosages for depressed mood range between 1,500 and 5,000 mg/day. L-Tryptophan is moderately sedating and is typically used at night. Early studies reported antidepressant effects at doses of 2 g/day; however, a systematic review of placebo-controlled RCTs found equivocal evidence of efficacy. 5-HTP is generally preferred over L-tryptophan because it crosses the blood–brain barrier more readily, is converted into serotonin more efficiently than L-tryptophan, and has more marked antidepressant effects. Findings of placebo-controlled studies support that 5-HTP starts to have antidepressant effects at doses between 100 and 300 mg/day. In a small

8-week double-blind study (*N* = 70), depressed patients randomized to 5-HTP versus fluoxetine reported equivalent improvement in mood. Patients were started on 5-HTP in three divided doses starting at 150 mg for the first 2 weeks, increased to 300 mg in three divided doses starting week 3, and further increased to 400 mg after the fourth week. Both groups showed significant and equivalent reductions in depressed mood scores starting at week 2 and continuing to study end. 5-HTP is moderately sedating, and doses greater than 100 mg should be taken at bedtime. RCTs have reported consistent anxiolytic effects of 5-HTP for generalized anxiety and panic disorder. 5-HTP dosed at 50 to 100 mg thrice daily is well tolerated alone or in combination with antianxiety agents. In a double-blind study, 58% of generally anxious patients (*N* = 79) randomized to 3 g per day of L-tryptophan reported significantly greater reductions in baseline anxiety compared to placebo. L-Tryptophan, 1 g, at bedtime reduces time to sleep onset in mild situational insomnia, and doses up to 15 g may be necessary for severe insomnia. Case reports suggest that nighttime use of 5-HTP at 300 to 600 mg may improve mild to moderate insomnia and lessen sleep disturbances related to obstructive sleep apnea and narcolepsy.

Recent findings suggest that treatments employing the amino acid precursors of serotonin (i.e., 5-HTP) and dopamine (i.e., L-tyrosine) must be properly balanced in order to be efficacious over the long term. These findings come mainly from studies on monoamine transporter optimization (MTO). According to MTO theory, competitive inhibition resulting from taking exogenous 5-HTP or L-tyrosine alone leads to the unbalanced synthesis of the “opposite” neurotransmitter resulting in its relative deficiency compared to the dominant monoamine neurotransmitter. It has been suggested that imbalances in the monoamine neurotransmitters resulting from chronic antidepressant use leads to gradual loss of efficacy, increased incidence of adverse effects, and (over time) complete loss of efficacy when the nondominant neurotransmitter has diminished below a critical threshold level in the CNS. Limited research evidence suggests that the optimal balance between serotonin and dopamine is achieved in response to a daily regimen that includes 5-HTP, L-tyrosine, and L-dopa.

Uncommon, usually mild side effects of both L-tryptophan and 5-HTP include dry mouth, gastrointestinal distress, and fatigue. Both amino acids are contraindicated in combination with MAOIs as the combination may result in a serious serotonin syndrome. Both amino acids are also contraindicated in combination with lithium because they can elevate serum lithium levels to toxic levels.

ACETYL-L-CARNITINE. Acetyl-L-carnitine (ALC) has been investigated as a novel antidepressant because of its beneficial effects on neuroplasticity, neuronal cell membranes, and neurotransmitter modulation. Emerging findings suggest that ALC may slow the rate of cognitive decline in normal aging, dementia, and traumatic brain injury. Most studies on ALC in older adult depressed patients report sustained antidepressant effects after 1 month of treatment at doses of 2 to 3 g/day divided into three to four doses. ALC may accelerate response when taken as a monotherapy or an adjunct to antidepressants. A narrative review found consistent positive evidence for the superiority of ALC over placebo. **A 2018** systematic review and meta-analysis of placebo-controlled RCTs on ALC as a monotherapy or adjunctive to antidepressants found comparable efficacy of ALC and antidepressants. Relatively greater antidepressant effects were observed in older subjects, and side effects were significantly lower in subjects taking ALC. Findings of RCTs suggest that ALC 1.5 to 3 g/day improves age-related symptoms of cognitive impairment in healthy older adult nondemented individuals and depressed older adult patients. In a 2-month RCT, depressed patients with dementia randomized to ALC 3 g/day experienced significantly greater improvements in mood and global cognitive functioning compared to the placebo group. RCTs show that ALC 1,500 to 3,000 mg/day improves reaction time, memory, and cognitive performance in patients with dementia and may slow the rate of progression of cognitive impairment. A systematic review of 11 RCTs of ALC in dementia confirmed significant positive effects at weeks 12 and 24. However, improvements were not sustained at 1 year with continued treatment. Younger individuals experiencing presenile dementia may benefit more than older individuals, and for unclear reasons, 62 years may be the “optimal age” beyond which ALC loses much of its cognitive-enhancing effects. ALC is generally well tolerated and may be safely used in combination with antidepressants and cholinesterase inhibitors.

L-THEANINE. L-Theanine (*N*-ethyl-L-glutamine) is found in green tea and is widely used in Asian countries to treat anxiety and depressed mood. Animal studies confirm that L-theanine increases brain serotonin, dopamine, and γ-aminobutyric acid (GABA) levels and may have general neuroprotective effects. Anxiety-reducing effects of L-theanine are believed to be mediated by enhanced EEG alpha activity in the occipital and parietal regions that are dose-dependent and reportedly similar to EEG changes observed during meditation. Calming effects may last 8 to 10 hours and are usually experienced within 30 minutes following ingestion of L-theanine at doses between 50 and 200 mg. Moderately severe anxiety often responds to L-theanine 200 mg twice daily, and more severe anxiety may require doses up to 600 to 800 mg per day divided into 200-mg increments every 3 to 4 hours. In one RCT (*N* = 16), healthy adult volunteers randomized to alprazolam (1 mg) versus L-theanine (200 mg) or placebo experienced equivalent and nonsignificant anxiolytic effects during an experimentally induced anxiety state. In another RCT (*N* = 12), individuals taking L-theanine experienced a relatively greater reduction in acute stress response as measured by changes in HRV and salivary immunoglobulins (s-IgA) compared to the placebo group. These findings suggest that stress-reducing effects of L-theanine may be mediated by inhibition of cortical neuron excitation. In a crossover study, healthy adults were randomized to L-theanine (250 mg) alone or in combination with caffeine (150 mg). Compared to the L-theanine–only group, the combination group experienced improved visual information processing, reduced mental fatigue, faster reaction time and faster working memory, generally increased alertness, and fewer headaches. A 2020 systematic review of RCTs on L-theanine for reducing stress-response and anxiety symptoms in healthy individuals found consistent positive evidence for efficacy at doses of 200 to 400 mg/day. The significance of findings on L-theanine in anxiety is limited by small study sizes and the use of nonclinical populations in most studies. L-Theanine is generally well tolerated, and there are no published reports of serious adverse side effects or interactions with other natural products or conventional medications.

***N*-ACETYL-CYSTEINE.** The findings of animal and human studies suggest that NAC decreases neuronal glutamate release. Limited findings suggest that NAC 2 to 3 g/day may be a safe and effective adjunctive treatment of moderately severe OCD and trichotillomania (TTM) in adults but not severe OCD or TTM in children and adolescents. In a 16-week double-blind placebo-controlled study (*N* = 40) adults with treatment-refractory

OCD randomized to NAC 3,000 mg/day versus placebo while continuing on their medications experienced equivalent non-significant reductions in OCD symptom severity by study end.

In a 12-week double-blind placebo-controlled trial (*N* = 50) adults with TTM were randomized to NAC (1,200 to 2,400 mg/day) versus placebo while continuing their current medications. Individuals in the NAC group experienced significantly greater reductions in hair-pulling behavior compared to the placebo group, and no adverse effects were reported. In contrast to adult TTM, adjunctive NAC is probably not an effective adjunctive treatment of pediatric TTM. In a placebo-controlled trial (*N* = 39), children and adolescents randomized to NAC while continuing treatment as usual reported nonsignificant differences in hairpulling compared to the placebo group. Findings on NAC in bipolar disorder are inconsistent. In a 2-month open-label phase of a randomized placebo-controlled study, 149 moderately depressed bipolar patients treated with NAC 1 g BID experienced significantly decreased indices of functioning and improved life quality compared to the placebo group. In a subsequent double-blind placebo-controlled study, 149 stable bipolar patients randomized to NAC 2 g/day versus placebo while continuing on treatment as usual did not experience significant differences in recurrence rates or quality of life.

GLYCINE. Glycine functions as an inhibitory neurotransmitter. Glycine and glutamate are required co-agonists for the activation of NMDA receptors. Case reports and RCTs suggest that glycine and other naturally occurring NMDA agonists help correct NMDA receptor dysfunction underlying negative symptoms and cognitive impairments in schizophrenia. However, only a few small clinical trials have been done, and findings are inconsistent. A small open trial found that negative psychotic symptoms improve with glycine at doses up to 60 g per day. In a 6-week crossover RCT (*N* = 22), patients with treatment-resistant schizophrenia randomized to glycine at 0.8 g/kg in addition to their antipsychotic versus placebo plus antipsychotic experienced a 20% to 30% greater improvement in negative symptoms compared to drug plus placebo. In a multicenter 16-week RCT (*N* = 157), inpatients diagnosed with schizophrenia or schizoaffective disorder were randomized to glycine versus D-cycloserine or placebo. No significant differences in outcomes were observed in negative psychotic symptoms or cognitive impairment. In a 10-week open-label prospective study, stable patients diagnosed with schizophrenia (*N* = 29) randomized to glycine (up to 60 g per day) while continuing on antipsychotic medications experienced significant improvements in depressed mood and reductions in extrapyramidal symptoms. Glycine is well tolerated.

LYSINE AND ARGININE. Limited findings suggest that lysine and arginine may enhance stress-response and reduce general anxiety when used alone or in combination. In a 10-day placebo-controlled RCT (*N* = 29), healthy male volunteers with high trait anxiety were randomized to receive a formula containing L-lysine and L-arginine (10 g/day) or a placebo before being exposed to a psychosocial stressor (Jezova et al., 2005). The group receiving the amino acid formula experienced enhanced ability to cope with a psychosocial stress (i.e., being asked to give a public speech) that was mediated by increased cortisol. In a 1-week placebo-controlled RCT (*N* = 108), healthy adults were randomized to receive a formula containing L-lysine (2.64 g/day) and L-arginine (2.64 g/day) versus placebo. The group receiving the amino acid formula was observed to have significant reductions in both short-term stress response and state anxiety. L-Lysine and L-arginine are generally well tolerated, and serious adverse effects have not been reported.

***S*-ADENOSYL METHIONINE.** SAMe is a methylation product of the amino acid methionine and adenosine triphosphate. The molecule is essential for many important biosynthetic pathways, including transmethylation reactions required for the synthesis of serotonin, dopamine, and norepinephrine. After donating a methyl group, SAMe is converted to *S*-adenosyl-homocysteine, which has been implicated in the pathogenesis of Alzheimer disease, atherosclerosis, and other inflammatory diseases. SAMe is a widely used treatment of depression, arthritis, and fibromyalgia. The compound is available by prescription in Europe in oral, injectable, or intravenous forms, but in the United States, it can be obtained only as a nonstandardized, over-the-counter oral supplement. Folic acid and B_{12} are required cofactors for re-methylation of homocysteine back to SAMe and are routinely recommended to patients taking SAMe.

Although SAMe has been widely used for many years to treat depressed mood, large placebo-controlled studies have been conducted only in the last few decades. Recent well-powered studies and systematic reviews have reported non-significant or inconsistent findings. A systematic review published in 2016 included eight trials (total *N* = 934 adults) comparing both oral and parenteral SAMe with placebo, imipramine, desipramine, or escitalopram found non-significant response differences between SAMe and placebo, and non-significant outcome differences when SAMe and placebo were used adjunctively to an SSRI. A 2020 systematic review that included 8 placebo-controlled randomized trials (total *N* = 1,011) evaluating SAMe at doses between 200 and 3,200 mg/day as a monotherapy or adjunctive therapy reported inconsistent and nonsignificant findings.

Although the evidence supporting SAMe is uneven, some patients report significant sustained improvements in depressed mood. Advantages of SAMe over antidepressants include relatively rapid onset of action, usually within 1 week of starting treatment, the absence of clinically significant interactions with drugs, and relatively few side effects compared with antidepressants. The standard maintenance regimen for depressed mood is between 800 and 1,600 mg in two to four divided doses. Absorption is improved when SAMe is taken before meals.

The butane–disulfonate form of SAMe has significantly greater bioavailability than the tosylate form and is available as an enteric-coated tablet for longer shelf life. SAMe is generally well tolerated. However, transient mild anxiety, insomnia, gastrointestinal side effects, tachycardia, dry mouth, and dizziness have been reported. Rare cases of mania or hypomania have been reported in patients with bipolar disorder.

Omega-3 Fatty Acids and Phosphatidyl Serine

OMEGA-3 FATTY ACIDS. Omega-3 fatty acids (n3 PUFA) are polyunsaturated fatty acids found in oily fish, algae, nuts, and seeds. Two omega-3 fatty acids, EPA and DHA, have been extensively investigated as treatments of depressed mood, ADHD, schizophrenia, and other psychiatric disorders. Several mechanisms of action may help explain observed beneficial effects of omega-3 fatty acids, including increased CNS serotonin activity, anti-inflammatory effects, and suppression of phosphatidyl–inositol second messenger activity. There is strong evidence that omega-3 fatty acids (N-3 PUFA)—especially formulas containing high doses of purified EPA—are effective adjunctive treatments of depressed mood but are not effective when used as a monotherapy. Effective adjunctive doses are 1 to 2 g/day. Formulations containing more EPA than

DHA may have greater antidepressant efficacy. Consistent findings of high-quality RCTs support that formulas containing higher EPA dosages are associated with better outcomes than formulas with higher amounts of DHA. Omega-3 may have the greatest antidepressant efficacy in obese individuals and others with abnormally high levels of inflammation. The finding that omega-3 fatty acids may reduce the incidence of coronary artery disease by influencing the production of pro-inflammatory cytokines in the heart may help to explain the observed correlation between heart disease and major depressive disorder. Preliminary evidence suggests that omega-3 supplementation may reduce anxiety associated with substance abuse or withdrawal, test-taking anxiety, and in premenstrual, pregnant, or menopausal women. Omega-3 fatty acids are generally well tolerated. Mild adverse effects include gastrointestinal upset, nausea, bloating, and diarrhea. There is a small risk of increased bleeding time, and concomitant use with anticoagulants should be avoided.

A meta-analysis of RCTs on omega-3 fatty acids in bipolar disorder identified only one study in which omega-3 fatty acids used adjunctively with a mood stabilizer showed a differential beneficial effect on depressive but not manic symptoms. EPA augmentation may be an effective preventive strategy in individuals at high risk for developing schizophrenia or in the early prodromal phase of psychotic illness but not as a treatment of established cases of schizophrenia. The neuroprotective role of omega-3 fatty acids during the early prodromal phase of schizophrenia may be related to general effects on the brain's antioxidative intracellular defense mechanisms or direct interactions between EPA and the glutamate receptor system. EPA augmentation may be especially appropriate for younger or antipsychotic naïve individuals in the early phase of illness or cases in which metabolic or sexual adverse effects may result in medication nonadherence and relapse. Preliminary findings suggest that a combination of EPA and DHA may reduce the rate of transitioning to full-blown schizophrenia in high-risk populations, including individuals who chronically abuse substances or already have impairments in neuropsychological functioning. A meta-analysis of placebo-controlled studies on EPA augmentation in patients with chronic schizophrenia or other psychotic disorders found no evidence for a greater benefit of EPA augmentation over antipsychotic medications alone. All studies included in the meta-analysis lasted 12 weeks and used EPA doses between 2 and 3 g per day. Possible explanations of reported negative findings include small study size, absence of clinical efficacy of EPA supplementation in chronic schizophrenia, and a “ceiling effect” on response to EPA augmentation associated with concurrent treatment with antipsychotics.

A systematic review and meta-analysis of studies on the effects of omega-3 supplementation on cognitive development, function, and decline throughout the life span concluded that omega-3 supplementation —especially DHA—may significantly improve cognitive development in infants but not in children, adolescents, adults, or the older adults. omega-3 fatty acids supplementation was not associated with a reduced rate of cognitive decline in healthy older adults or with a slowing in the rate of progression of Alzheimer disease. A systematic review and meta-analysis of placebo-controlled trials on omega-3 fatty acids in the treatment of ADHD in children and adolescents reported inconclusive findings due to methodologic problems including small sample sizes, variability of selection criteria, and type and dosage of supplements and short-term follow-ups. Studies done in Japan following the 2011 tsunami investigated the effectiveness of pretreatment with omega-3 fatty acids in preventing the development of PTSD following exposure to trauma in first medical responders. In a small, 8-week controlled trial (N = 30), women diagnosed with moderately severe borderline personality disorder randomized to ethyl-EPA (1 g/day) versus placebo reported less severe symptoms of aggression and depressed mood compared to the placebo group.

Omega-3 fatty acids are generally well tolerated and pose few safety issues. Transient mild gastrointestinal distress is sometimes reported by patients who take omega-3 fatty acids. There is one case report of apparent hypomania induced by omega-3 fatty acids. Rare cases of increased bleeding times have been reported in patients who take aspirin or anticoagulants concurrently.

PHOSPHATIDYL SERINE AND PHOSPHATIDYL CHOLINE. Phosphatidyl serine is one of the most important phospholipids in the brain and is an essential component of nerve cell membranes. Findings of RCTs suggest that phosphatidyl serine (300 mg per day) may improve global functioning and memory in the early stages of Alzheimer disease. However, the findings are inconsistent. Controlled trials on older adults suggest that phosphatidyl serine (300 mg per day) may result in small but consistent improvements in age-related cognitive decline. Open trials and one small controlled trial reported that supplementation with phosphatidylcholine

(PC) 15 to 30 g/day reduced the severity of both mania and depressed mood in bipolar patients and that symptoms worsened when PC was discontinued. Cytidine 5-diphosphocholine (CDP-choline) is a precursor of PC and other important structural phospholipids in neuronal cell membranes. CDP-choline increases mitochondrial energy production, increases CNS noradrenaline and dopamine levels and has general neuroprotective effects following hypoxic or ischemic brain injury. Preliminary findings suggest that CDP-choline (1,000 mg per day) may improve cognitive performance in the early stages of Alzheimer disease. Two systematic reviews concluded that CDP-choline has consistent positive effects on recovery rate following stroke and mitigates the severity of cognitive impairment related to cerebrovascular disease in older adult individuals.

Multinutrient Formulas. A systematic review of RCTs on micronutrient formulas for the treatment of diverse psychiatric disorders found evidence of efficacy for a high-dose vitamin B complex formula and three different micronutrient formulas for acute stress, general anxiety, and PTSD. In general, high-dose formulas had differentially greater effects compared to low-dose formulas. The significance of findings was limited by small study size, nonblinding of many studies, heterogeneity in designs, and the absence of clinically diagnosed subjects in most studies. A systematic review of RCTs on single supplements and supplement formulas including essential fatty acids (EFA), B vitamins, vitamin C, Mg, and/or Zinc for female stress and anxiety identified 14 studies that met inclusion criteria. Significant findings included stress-reducing effects of EFAs during pregnancy and in premenopausal and menopausal women; reduced premenstrual anxiety in response to combined Mg and B6 (but no anxiolytic effects when either supplement was used alone); reduced anxiety in older women in response to B6; high-dose vitamin C reduced anxiety and blood pressure in response to stress. Micronutrient formulas are generally well tolerated at doses commonly recommended for treating psychiatric disorders. Some micronutrient formulas contain very high amounts of vitamin A, other vitamins, trace minerals, and amino acids, potentially resulting in toxicity. There have been case reports of hypervitaminosis A when taking micronutrient formulas that contain very high amounts of vitamin A.

Findings from several case series suggest that a proprietary nutrient formula (EMP+) containing 36 separate constituents, including chelated minerals, vitamins, essential fatty acids, amino acids, phytonutrients, and trace elements, may significantly reduce symptoms of mania, depressed mood, and psychosis in patients with bipolar disorder when taken alone or together with mood stabilizers. The putative mechanism of action involves the correction of inborn metabolic errors that manifest as cyclic mood changes when the diet is deficient in certain micronutrients. In a 10-week double-blind placebo-controlled RCT (*N* = 93), children with ADHD randomized to EMP+ experienced significant improvements in inattentive symptoms, emotional regulation, and aggressive behavior, but no improvements in hyperactive or impulsive symptoms. Preliminary findings suggest that taking EMP+ before exposure to trauma may increase emotional resilience and reduce the severity of PTSD symptoms following exposure. The significance of findings on EMP+ is limited by the small study size and the absence of a placebo group, blinding, and randomized protocols. Potentially toxic micronutrient–medication interactions have been reported in patients taking EMP+ with mood stabilizers, and individuals taking the formula should be closely monitored by a knowledgeable physician.

Botanicals

ASHWAGANDHA (*WITHANIA SOMNIFERA*). Ashwagandha is used in Ayurvedic medicine to enhance general wellbeing, improve stamina, and treat a range of medical and mental health conditions. The bioactive constituents (withanolides) have anti-inflammatory and antioxidant properties, influence the GABA neurotransmitter system, and affect the HPA axis resulting in a reduction in cortisol levels. Findings of RCTs support that Ashwagandha is an effective monotherapy of anxiety symptoms related to chronic stress and may be an effective adjunctive treatment of GAD. A systematic review of RCTs on Ashwagandha for anxiety and stress found evidence for consistent positive effects at doses between 125 and 1,200 mg/day. Limited findings suggest that Ashwagandha may be an effective adjunctive treatment of GAD. In a 6-week placebo-controlled RCT (*N* = 40), adults diagnosed with GAD were randomized to a standardized Ashwagandha formula (1 g/day) versus placebo while continuing on an SSRI. The group receiving adjunctive Ashwagandha experienced significantly greater reductions in anxiety symptoms starting at the second week of treatment. Adjunctive treatment in combination with SSRIs was well tolerated. The significance of findings on Ashwagandha in GAD is limited by publication bias, small study size, heterogeneity of study designs, and variability in dosing strategies. Ashwagandha is generally well tolerated. Uncommon mild adverse effects include gastrointestinal distress and diarrhea. Ashwagandha may cause miscarriages and should be avoided during pregnancy.

CANNABIDIOL (*CANNABIS SATIVA*). Cannabidiol (CBD), a constituent of *Cannabis sativa,* has been extensively investigated as a potential treatment of diverse neuropsychiatric disorders. Findings of animal studies and early clinical trials suggest that CBD may be an effective treatment of GAD, panic disorder, social anxiety disorder (SAD), OCD, and PTSD. Anxiolytic effects are believed to involve modulation of CB1, CB2, and 5-HT1A receptors in limbic structures such as the amygdala, hippocampus, hypothalamus, cingulate cortex, as well as the prefrontal cortex. A retrospective chart review of a large case series ($N = 103$) found evidence for consistent anxiety-reducing effects of CBD. A review of animal and human studies investigating CBD as a potential treatment of panic disorder found promising early evidence; however, all clinical trials done to date have been in healthy volunteers. A systematic review of RCTs on the safety and efficacy of CBD as a treatment of diverse psychiatric disorders identified 27 studies, including 2 RCTs on CBD in SAD. In both studies, never-treated individuals with SAD randomized to receive CBD at doses of 400 or 600 mg experienced significantly greater reductions in social anxiety compared to the placebo group. CBD is well tolerated at doses up to 1,500 mg/day. Uncommon generally mild adverse effects include dry mouth, low blood pressure, lightheadedness, and drowsiness. There are no described contraindications to CBD use.

GOLDEN BOUQUET (*GALPHIMIA GLAUCA*). Galphimia is used as a sedative and anxiolytic in traditional Mexican medicine. The findings of animal studies suggest that the mechanism of action is mediated by the modulatory effects of Galphimine-B on dopamine, NMDA, and serotonin. Findings of RCTs support that a standardized extract of the herbal dosed at 310 mg BID has anxiolytic effects comparable to or greater than benzodiazepines in individuals diagnosed with GAD. In a 15-week study

(*N* = 191), subjects with GAD randomized to Galphimia experienced significantly greater reductions in anxiety than subjects treated with lorazepam (1 to 2 mg/day). No adverse effects were reported by subjects taking Galphimia, and one-fifth of subjects taking lorazepam reported excessive sedation. In a 10-week placebo-controlled RCT (*N* = 167), individuals with moderate or severe anxiety were randomized to a standardized Galphimia formula versus alprazolam experienced significant, equivalent reductions in anxiety. No adverse effects were reported in the Galphimia group, and over 20% of subjects taking alprazolam reported daytime sleepiness. Galphimia is generally well tolerated. Mild adverse effects include headaches, gastrointestinal distress, and sedation.

GINKGO BILOBA. Ginkgo (*Ginkgo biloba*) extract is used in Europe and North America to treat dementia and other neurodegenerative diseases. Most commercially available preparations are standardized to two bioactive constituents: 24% flavone glycosides and 6% terpenoids. The flavonoid constituent functions as a strong antioxidant and is believed to have a general neuroprotective effect. The terpenoid fraction antagonizes platelet-activating factor (PAF), facilitating poststroke recovery through vascular reperfusion by decreasing thrombosis and nerve cell death associated with cerebral ischemia. Constituents of ginkgo may inhibit neurotoxicity and nerve cell death caused by nitric oxide. Two RCTs found that a compound herbal product containing both *G. biloba* (160 or 320 mg) and *Panax ginseng* (400 mg) resulted in transient improvement in recall performance in healthy adults. Enhanced cognitive functioning appears to reach its peak approximately 6 hours after the herbal preparation is taken. Preliminary findings suggest that combining ginkgo with another traditional Chinese herbal, *dangshen* (*Codonopsis pilosula*), may improve cognitive performance in healthy adults more than either herbal alone. In a 24-week, double-blind RCT, patients with mild to moderate dementia were randomized to a standardized ginkgo extract (160 mg per day) versus the cholinesterase inhibitor donepezil (5 mg per day) or placebo. At 24 weeks, patients taking ginkgo and donepezil experienced equivalent improvements in cognitive performance and global functioning. A meta-analysis of 11 clinical trials on standardized ginkgo extracts in older adults cognitively impaired individuals who did not meet full criteria for dementia suggested consistent cognitive-enhancing effects. Early systematic reviews of RCTs on ginkgo in dementia concluded that ginkgo preparations in doses between 120 and 600 mg per day taken over several weeks to 1 year result in modest improvements in memory, general cognitive functioning, and activities of daily living in mild to moderate Alzheimer dementia and multi-infarct dementia. However, a systematic review of all RCTs published up to 2006 on ginkgo in the treatment of dementia, and other types of acquired cognitive impairment found inconsistent evidence, noting that early trials were generally small or methodologically flawed, including the absence of standardized preparations and the use of different dementia rating scales across studies. Findings of a recent systematic review and meta-analysis support that ginkgo has beneficial effects on behavioral and psychological symptoms associated with dementia, including aggressive behavior, apathy, and anxiety. Ginkgo is generally well tolerated. Mild transient adverse effects include upset stomach, dizziness, and headaches. Cases of spontaneous bleeding have been reported. Ginkgo should be avoided in patients taking aspirin, warfarin, or heparin. *G. biloba* preparations should be discontinued at least 2 weeks prior to surgery. *Ginkgo. biloba* may cause elevation of hepatic enzymes, and there are case reports of possible serotonin syndrome when taken with SSRIs.

GINSENG (*PANAX GINSENG*). Individuals diagnosed with schizophrenia frequently show marked deficits in working memory, which can limit functional improvements with conventional drug treatment. In a 4-week placebo-controlled trial, 64 stable individuals diagnosed with schizophrenia were randomized to a proprietary North American ginseng extract versus placebo while remaining on treatment as usual. Visual working memory was significantly improved in patients receiving the ginseng preparation but not in the placebo group. Of note, individuals taking the ginseng preparation reported significantly fewer neurologic adverse effects after 4 weeks than the placebo group. Further studies are needed to confirm these findings and determine the optimal dosing of ginseng as an adjuvant to antipsychotics.

KAVA (*PIPER METHYSTICUM*). In traditional Polynesian cultures, Kava (Fig. 30.4–1) is used for ceremonial purposes and as an inebriant. Preclinical studies suggest that kavalactones in the roots interfere with serotonin neurotransmission in the amygdala, act as norepinephrine reuptake inhibitors that have a binding affinity with both GABA and NMDA receptors. Kava may also reduce anxiety by directly modulating vagal heart tone.

Most studies on Kava use preparations standardized to 70% kava lactones at doses between 70 and 240 mg per day for the treatment of “stress” and moderate anxiety in healthy adults. Early RCTs and systematic reviews reported consistent positive anxiolytic effects of Kava; however, most studies were small, of short duration, or did not enroll subjects with diagnosed anxiety disorders. More recent studies have reported inconsistent findings. In a 4-week double-blind RCT ($N = 69$), older adult individuals with diagnosed DSM-5 anxiety disorders were randomized to a high-dose Kava formula (containing 200 mg/day of kavalactones) versus a low-dose Kava formula (20 mg/day of kavalactones). At the study end, the high-dose Kava group reported statistically significantly greater reductions in anxiety compared to the low-dose group. In contrast, a large 16-week multicenter study found equivocal evidence for Kava 120 mg/day in individuals diagnosed with GAD. Findings of the most recent systematic review of placebo-controlled RCTs on Kava in individuals with diagnosed GAD identified only three studies that met inclusion criteria (total $N = 130$) and found modest anxiolytic effects of Kava that were not superior to placebo.

Kava is generally well tolerated at therapeutic doses. However, some individuals report nausea, rash, headaches, or dizziness. Kava use should be avoided during pregnancy. Kava may potentiate the sedating effects of alcohol and benzodiazepines. Rare case reports of hepatitis and fulminant liver failure have led to restrictions in the sale of kava products in many European countries and a warning by the FDA. A thorough review concluded that most reported cases of liver failure were associated with a processing mistake resulting in potentially toxic levels of alkaloids in a single batch of kava.

FIGURE 30.4–1. Kava (*Piper methysticum*).

LAVENDER (*LAVANDULA OFFICINALIS*). The essential oil of lavender has been widely used in traditional Persian medicine for millennia to treat medical disorders and mental health problems, including anxiety, insomnia, and depressed mood. The findings of animal and human studies suggest that the essential oil of lavender has significant anxiolytic properties mediated by the effects of linalool and linalyl on serotonin and NMDA receptors. A systematic review of studies on different lavender preparations found evidence for modest anxiolytic effects of lavender essential oil but nonsignificant evidence for aromatherapy. A systematic review and meta-analysis of RCTs on lavender aromatherapy in populations with or without clinical anxiety that included 22 trials found evidence for modest anxiety-reducing effects. Placebo-controlled RCTs and systematic reviews of studies on Silexan, a standardized oral preparation of lavender oil support comparable efficacy to SSRIs and benzodiazepines in individuals with GAD or anxiety-related restlessness and disturbed sleep. When dosed between 80 and 160 mg/day, Silexan has comparable or superior efficacy to paroxetine and lorazepam. One placebo-controlled study found superior efficacy of Silexan 160 mg/day compared to lorazepam 0.5 mg for symptoms of generalized anxiety.

ROSEROOT (*RHODIOLA ROSEA*). *R. rosea* is used in traditional Russian medicine to enhance stamina, reduce fatigue, improve sleep, and enhance mood. Findings of in vitro and animal studies suggest that bioactive constituents of *R. rosea* affect the HPA axis reducing serum cortisol levels reducing fatigue, and decreasing stress-related anxiety. Recent placebo-controlled RCTs support that standardized extracts of *R. rosea* at doses of 340 to 680 mg/day are superior to placebo for mild to moderate depressed mood and may be an effective adjuvant of antidepressants. In a 12-week double-blind placebo-controlled RCT by Gao and colleagues (*N* = 100), patients with mild to moderate depressed mood were randomized to sertraline plus *R. rosea* 0.6 g/day, sertraline plus *R. rosea* 0.3 g/day, or sertraline plus placebo. At the study end, both *R. rosea* groups reported significantly greater reductions in depressive mood symptoms compared to the placebo group. Subjects receiving the higher dose of *R. rosea* responded better than subjects receiving the lower dose. Limited findings of small RCTs suggest that standardized preparations of *R. rosea* dosed at 200 mg BID have significant stress-reducing effects. To date, only one small open-label study has investigated *R. rosea* in individuals with GAD with positive findings. Findings of recent open studies suggest that 500 mg of a standardized *R. rosea* extract has antidepressant effects, improves overall mental performance and stamina, and may accelerate the recovery of normal cognitive functioning following traumatic brain injury.

R. rosea is generally well tolerated at doses used to treat stress-related fatigue. Uncommon adverse events include dizziness, dry mouth, and excessive salivation. *R. rosea* can be safely combined with antidepressants and may augment antidepressant efficacy. There are no cases of serious adverse effects or drug–drug interactions; however, there have been reports of possible mania induction. Patients diagnosed with bipolar disorder should be advised against using this herb.

SAFFRON (*CROCUS SATIVUS*). Saffron, obtained from the stigmas of the flower of *Crocus sativa*, is widely used as a spice and medicinal agent in traditional Persian medicine. Antidepressant effects may involve anti-inflammatory and immunomodulatory effects and reuptake inhibition of serotonin, norepinephrine, and dopamine. Findings of meta-analyses and systematic reviews suggest that saffron 20 to 30 mg/day has comparable efficacy to antidepressants for mild to moderately severe depressed mood; however, recent double-blind RCTs have reported inconsistent findings. So far, only one study has investigated saffron as a potential adjunctive treatment of MDD. In an 8-week double-blind RCT ($N = 160$), depressed individuals were randomized to a standardized saffron extract (14 mg BID) versus placebo while continuing on their antidepressant medication. Clinicians rated the saffron group as showing greater improvements in the depressed mood; however, subjects in the saffron group did not report subjective improvements in mood. Preliminary findings suggest that saffron may have comparable efficacy to cholinesterase inhibitors in Alzheimer's patients. Saffron is generally well tolerated as a monotherapy and in combination with antidepressants. In a 6-week RCT, saffron supplementation improved sexual adverse effects associated with fluoxetine. Uncommon mild adverse effects include nervousness, nausea, headache, and increased appetite.

ST. JOHN WORT (*HYPERICUM PERFORATUM*). SJW (Fig. 30.4–2) is widely used to treat depressed mood. Different constituents function as serotonin reuptake inhibitors, decrease binding to benzodiazepine receptors, weakly inhibit monoamine oxidase (MAO), and possibly also bind to *N*-methyl-D-aspartate (NMDA) receptors. Hyperforin probably contributes more to the antidepressant efficacy of SJW than hypericin. However, many commercial preparations are standardized to 0.3% hypericin. Therapeutic doses of standardized SJW preparations for moderate depression typically range from 300 mg twice daily to 900 mg twice daily. More than 40 controlled trials on SJW in patients with depressed mood have been conducted with mixed methodologic quality and inconsistent results. A systematic review of 27 trials (total $N = 2{,}200$), 17 of which were placebo-controlled, found that SJW is significantly more effective than placebo for the treatment of mild to moderate depressed mood. However, a more recent meta-analysis using stringent inclusion criteria controlling for publication bias and small study effects found significantly lower antidepressant efficacy than reported in previous meta-analyses.

SJW is generally well tolerated. Uncommon adverse include nausea, restlessness, and mild sedation. There are no published reports of serotonin syndrome or other serious side effects when SJW is used as monotherapy at recommended doses, but there are several case reports of possible serotonin syndrome when SJW and a conventional antidepressant were used together. Pregnant women should not take

SJW. There are no case reports of serious interactions when SJW is taken with MAOI antidepressants. However, their combined use should be avoided to minimize risk. SJW should not be taken with protease inhibitors, immunosuppressive agents, Coumadin, digoxin, and oral contraceptives.

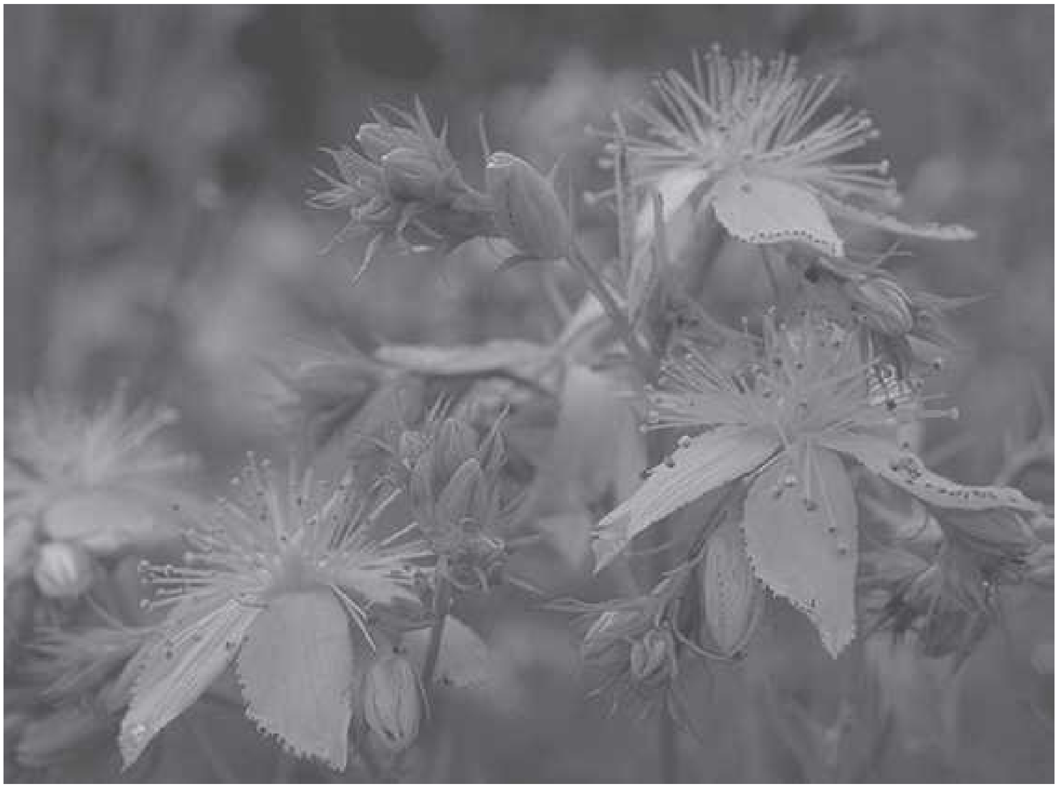

FIGURE 30.4–2. St. John wort (*Hypericum perforatum*).

TOOTHED CLUBMOSS (*HUPERZIA SERRATA*). Huperzine A, an alkaloid derivative of the herb *Huperzia serrata* is a central ingredient of many compound herbal formulas used in Chinese medicine to treat cognitive impairment related to normal aging. Huperzine A reversibly inhibits acetylcholinesterase, the enzyme that breaks down the neurotransmitter acetylcholine, which plays a critical role in memory. Huperzine A may also slow the production of nitric oxide in the brain, reducing age-related neurotoxicity. Findings from animal studies suggest that huperzine A may be a more potent and more specific inhibitor of acetylcholinesterase than some cholinesterase inhibitors. Early controlled trials reported consistent beneficial effects in both age-related memory loss (i.e., benign senescent forgetfulness) and Alzheimer disease at doses between 0.2 and 0.8 mg/day; however, more recent studies report inconsistent findings. Two systematic reviews of placebo-controlled randomized trials on Huperzine A in patients diagnosed with AD found inconclusive evidence for improvements in memory loss and other cognitive problems. Because most studies on Huperzine A are of short duration, 8 to 12 weeks, it has been suggested that observed beneficial effects might reflect temporary symptomatic improvement. Huperzine A is generally well tolerated. Infrequent adverse effects include transient dizziness, nausea, and diarrhea.

TURMERIC (*CURCUMA LONGA*). Curcumin is the principal bioactive constituent of Turmeric. Animal studies have established that curcumin has strong antioxidant and anti-inflammatory activity, suggesting that it may have beneficial effects on a range of medical and neuropsychiatric disorders associated with chronic inflammation. The findings of animal studies suggest that antidepressant effects of curcumin may be related to the modulatory activity of the monoamine neurotransmitters serotonin, norepinephrine, and dopamine.

Recent findings support that curcumin at doses between 500 mg and 1,000 mg/day has antidepressant efficacy as a monotherapy and is a safe and effective adjuvant to antidepressants. A systematic review of RCTs on curcumin for neuropsychiatric disorders found strong evidence for antidepressant and antianxiety effects and provisional findings for beneficial effects of curcumin on PTSD, bipolar disorder, OCD, psychotic disorders, and autism. A meta-analysis of 6 placebo-controlled RCTs (total *N* = 377) on curcumin in MDD found evidence of superior antidepressant and antianxiety efficacy over placebo. Curcumin is generally well tolerated, and severe adverse effects have not been reported.

VALERIAN (*VALERIANA OFFICINALIS*). Valerian (*Valeriana officinalis*) has been widely used in many traditional systems of medicine for more than a millennium. The mechanism of action may involve both inhibition of GABA reuptake and binding to serotonin receptors by valepotriates. Findings on Valerian extract in insomnia are inconsistent. An early systematic review of RCTs on valerian for insomnia found that a standardized extract of 600 to 900 mg taken at bedtime improves the subjective quality of sleep with few side effects. However, a more recent systematic review of 29 RCTs that met rigorous criteria for study design and methodologic quality concluded that the most rigorously designed studies failed to demonstrate significant outcomes differences between valerian and placebo. Another systematic review found a statistically significant benefit of valerian extract in insomnia but noted significant methodologic flaws in most studies examined and evidence of publication bias. In a large, multicenter, 6-week study, chronic insomniacs randomized to a standardized valerian extract at 600 mg versus oxazepam at 10 mg reported comparable improvements in sleep quality and duration. There is no risk of addiction with valerian preparations, and there are no reports of daytime psychomotor slowing. Because of its potentiating effect on GABA, valerian use should be avoided with benzodiazepines or alcohol. There are rare reports of elevated liver enzymes. Valepotriates may be mutagenic, and this herb should be avoided in pregnant or nursing women.

OTHER BOTANICALS. Select herbals and other natural products can be safely used adjunctively to depressed mood and other psychiatric disorders or to mitigate adverse effects caused by psychopharmaceuticals. For example, *G. biloba* can be safely combined with antidepressants to mitigate adverse sexual effects. Limited findings suggest that Passion flower (*Passiflora incarnata*) extract may have equivalent efficacy to benzodiazepines for the management of moderate anxiety symptoms. Preparations of two botanicals used in Ayurvedic medicine, *Bacopa monnieri* and *Centella asiatica,* traditionally used in Persian medicine to treat generalized anxiety, are well tolerated and diminish symptoms of moderate anxiety. Limited findings from human clinical trials suggest that an aqueous extract of Catmint (*N. menthoides*) 400 mg BID may have anxiolytic effects. Preclinical animal studies have reported promising early evidence for many other herbals in which anxiolytic effects are believed to be mediated by GABA or cannabinoid receptors. A comprehensive narrative review of studies on plant-derived anxiolytics found evidence of the efficacy of several herbals in the treatment of generalized anxiety, including *Matricaria recutita, Crataegus* spp., *G. biloba, Scutellaria lateriflora, Silybum marianum, P. incarnata, Withania somniferum, Galphimia glauca, C. asiatica, R. rosea, Echinacea* spp., *Melissa officinalis,* and *Echium amoenum.* Acute-onset anxiolytic activity was found for *C. asiatica, Salvia* spp., *M. officinalis, P. incarnata,* and *Citrus aurantium.*

Compound Herbal Formulas. Meta-analyses and systematic reviews of Chinese herbal formulas used to treat depressed mood have identified several promising formulas, including Chaihu-Shugan-San, Xiao Yao San, Ganmai Dazao, Wuling, Shuganjieyu, Wendan Decoction, and Jiu Wei Zhen Xin. Many systematic reviews have reported comparable or superior efficacy to antidepressants; however, the significance of findings is limited by methodologic flaws and possible publication bias. Significantly, most studies report very low rates of adverse effects when Chinese herbal formulas are used as monotherapies or adjunctively to antidepressants. Some studies reported that Chinese herbal formulas lessened antidepressant adverse effects.

Jiu Wei Zhen Xin Granula (JWZXG) is a Chinese herbal medicine formula that can be used alone or in combination with Western anxiolytic medications for GAD and other anxiety disorders. The formula is standardized to extracts of nine herbals with established tranquilizing or anxiolytic properties: *Panax Ginseng* (ginseng), Spina Date Seed (seed of wild jujube), *Schisandra chinensis* (the fruit of Chinese magnolia-vine), *Poria cocos* (hoelen), Radix Polygalae (the root of Polygala tenuifolia Willd), Rhizoma Corydalis (corydalis tuber), *Radix Asparagi* (Cochinchinese Asparagus Root), *Rehmannia glutinosa* (prepared Rehmannia root), and Cinnamon (cassia bark). A meta-analysis of 14 trials (total *N* = 1,358) on JWZXG in individuals with diagnosed GAD found superior efficacy of SSRIs but comparable efficacy of JWZXG and the azapirones. JWZXG was better tolerated than SSRIs. Although the formula is often used in combination with buspirone and other azapirones, potentially unsafe herb–drug interactions have not been evaluated. The significance of findings was limited by small study sizes, short study duration (most studies were 4 to 8 weeks long), the absence of placebo groups in many studies, inadequate blinding or nonblinding, and the small number of studies and possible publication bias. Adverse effects of JWZXG include dry mouth, constipation, dizziness, and nausea.

Kami-untan-to (KUT) is a compound herbal formula consisting of 13 different herbs used in Japanese traditional Kampo medicine to treat symptoms of cognitive impairment and dementia. Animal studies show that KUT increases CNS levels of nerve growth factor (NGF) and choline acetyltransferase (ChAT), the enzyme required to synthesize acetylcholine. KUT protects against memory impairment induced by thiamin deficiency in mice, suggesting a possible beneficial role in the management of delirium tremens and other syndromes of cognitive impairment related to thiamin deficiency. In a small 12-month open trial (*N* = 20), moderately demented patients with Alzheimer disease were treated with KUT alone versus a combined regimen of vitamin E, estrogen, and a nonsteroidal anti-inflammatory drug or placebo. Both treatment groups deteriorated at a significantly slower rate compared to the placebo group. Large, controlled trials are needed to confirm the potential therapeutic effects of this promising herbal formula in dementia.

Brahmyadiyoga is a compound herbal formula in Ayurvedic medicine used to treat syndromes resembling schizophrenia. The formula contains six herbs, including *Rauwolfia serpentina,* from which reserpine was isolated in the early 20th century and shown to have significant antipsychotic efficacy. Reserpine is seldom used to treat schizophrenia in Western countries today because of concerns over reports of severe depressed mood. In an early 2-month RCT, 108 chronically psychotic patients were randomized to 8 to 12 g per day of brahmyadiyoga, 8 to 12 g per day of valerian, 200 to 300 mg per day of chlorpromazine, and placebo. Patients treated with brahmyadiyoga and chlorpromazine reported significant comparable improvements in symptom severity. A subsequent RCT confirmed the antipsychotic efficacy of brahmyadiyoga. Findings of a systematic review of three small RCTs (total *N* = 250) comparing brahmyadiyoga with an antipsychotic or placebo were inconclusive because of small study sizes, high attrition rates, methodologic differences, and the use of nonstandardized assessment approaches. Cases of nausea have been reported with brahmyadiyoga.

Essential Oils. Essential oil extracts of certain botanicals can be inhaled as aromatherapy or applied topically; however, research findings are inconsistent. Some studies report general calming of agitated behavior in demented persons when the essential oils of lemon balm and lavender are applied to the face and arms. Other studies report that exposure to the same essential oils is no more effective than placebo in agitated patients with dementia. It has been suggested that essential oils are effective in this population through systemic absorption or the combined effects of massage and olfaction but *not* through olfaction alone. Infrequent adverse effects of essential oils include skin allergies, phototoxic reactions, and potentiation of sedative-hypnotics when used with lavender or other oils known to have sedating effects. Pregnant women should avoid topical use of essential oil therapies because of possible toxic effects on the fetus or the uterus caused by systemic absorption.

Hormones

DEHYDROEPIANDROSTERONE. Dehydroepiandrosterone (DHEA) is a precursor of testosterone and other hormones. The sulfated form of DHEA—DHEA-S—is the most abundant steroid in the body. DHEA is an important neuroactive steroid and modulates neuronal excitability by acting as an antagonist at the GABA receptor complex. Preclinical studies suggest that the putative mood-enhancing effects of DHEA are mediated through androgen receptors, estrogen receptors, serotonin, GABA, NMDA, and possibly norepinephrine. DHEA's antipsychotic effects may involve increased dopamine release in the frontal cortex and increased activity of NMDA and sigma receptors. Physiologic replacement doses of DHEA 30 to 90 mg/day, corresponding to DHEA serum levels at age 40, may improve mood in middle-aged or older adult depressed patients. Higher serum levels of DHEA and its sulfated form DHEA(S) are associated with healthy aging. A recent study found that healthy older adult individuals with relatively higher DHEA(S) levels are at reduced risk of developing depressed mood.

Findings of placebo-controlled RCTs support that DHEA supplementation of 90 to 450 mg/day reduces the severity of depressed mood; however, many studies are underpowered or lack methodologic rigor. In one 6-week RCT (*N* = 46), moderately depressed adults were randomized to DHEA 90 mg/day for 3 weeks followed by DHEA 450 mg/day (150 mg TID) for three weeks versus placebo (Schmidt, 2005). A 50% or greater reduction in depressive symptoms was observed in the majority of patients in the DHEA group, most of whom reported improvements in baseline sexual functioning, and most responders remained asymptomatic at 12 months follow-up. The findings of a recent study showing that higher circulating DHEA(S) levels correlate to increased remission rate in depressed individuals taking SSRIs suggests that DHEA may be a safe and effective adjunctive treatment of depressed mood.

Pilot studies suggest that DHEA (25 to 50 mg per day) improves memory and enhances general cognitive functioning in healthy adults.

However, findings are inconsistent, and negative findings have been reported at doses less than 90 mg per day. A systematic review of RCTs on DHEA or DHEA-S in healthy older adults found no evidence of improved memory or cognitive functioning. DHEA may differentially improve memory in older adult patients who have low DHEA serum levels more than in healthy adults with normal serum levels. However, no correlations were found between serum DHEA-S levels and severity of cognitive impairment or cumulative mortality in a large cohort of individuals newly diagnosed with Alzheimer disease. Preliminary findings suggest that 200 mg per day of DHEA may improve symptoms of cognitive impairment in multi-infarct dementia; however, no placebo-controlled trials have been done. Large prospective studies of at least 1 year or longer are needed to further investigate claims of long-term cognitive benefits of DHEA.

In a 6-week RCT (N = 30), inpatients with schizophrenia randomized to DHEA at 100 mg per day in addition to their regular antipsychotic medications experienced significant improvements in *negative* psychotic symptoms, including reduced apathy and social withdrawal. However, there were no significant changes in *positive* psychotic symptoms, including auditory hallucinations and delusions. Findings of another small RCT (N = 30) suggested that augmentation of antipsychotics with DHEA 100 mg/day for 6 weeks significantly reduces negative symptoms and may be especially effective in women. However, a systematic review that included three small RCTs (N = 126) on DHEA as an adjunctive to antipsychotics found equivocal evidence for an augmentation effect on measures of positive and negative symptoms, global functioning, and quality of life.

Increased severity of PTSD is correlated with reduced DHEA blood levels, and limited findings suggest that DHEA may lessen symptoms of PTSD. In an open-label pilot study (N = 5), women with treatment-refractory PTSD related to early childhood abuse who had not responded to pharmacologic therapy reported decreases in numbing, re-experiencing, hyperarousal, and other PTSD core symptoms, as well as improved sleep and improved libido with DHEA at doses between 25 and 100 mg/day.

DHEA is generally well tolerated at therapeutic doses, but cases of acne, hirsutism, and rare cases of mania have been reported with chronic high doses. DHEA supplementation may promote cancer in women who have a history of estrogen receptor-positive breast cancer and may increase the risk of prostate cancer in men with early or undetected prostate cancer. DHEA should be avoided in men at risk of prostate cancer and women who have a history of estrogen receptor-positive breast cancer. Women should be cautioned about an increased risk of heart disease and breast cancer.

MELATONIN. Melatonin is a hormone manufactured in the pineal gland and plays an important role in regulating the sleep–wake cycle. Synthetic over-the-counter forms of melatonin are widely used to treat insomnia. A meta-analysis of 17 RCTs (total *N* = 284) found that exogenous melatonin at doses between 0.3 and 3 mg decreases the time needed to fall asleep and increases total sleep duration. Doses that replace normal physiologic levels (0.3 mg) in insomnia patients with low serum melatonin levels improve restful sleep in the middle part of the sleep cycle. Higher pharmacologic doses (3.0 mg) also improve sleep quality but may cause hypothermia, elevated daytime melatonin levels, and residual grogginess. A systematic review of 10 RCTs found that melatonin in jet lag was effective at doses of 0.5 to 5.0 mg when taken close to bedtime in the country of destination. Overall efficacy was relatively less for westward flights and relatively greater when more time zones were crossed. Melatonin has been used with moderate success to treat insomnia in patients with schizophrenia and agitated patients with mania. Uncommon mild adverse effects include headaches and itching.

Psychedelics. Psychedelics including D-lysergic acid diethylamide (LSD), ketamine, psilocybin, and ayahuasca are extensively investigated as potential treatments of depressed mood, anxiety disorders, PTSD, and other psychiatric disorders. Antidepressant effects of LSD and psilocybin are believed to be related to modulatory activity on serotonergic, dopaminergic, and glutamatergic neurotransmitter systems. A comprehensive review of studies on psychedelics for all psychiatric disorders published between 2007 and 2019 identified a total of 1,603 articles, including 14 articles reporting well-designed clinical trials investigating LSD, MDMA, psilocybin, and ayahuasca for mood and anxiety disorders, PTSD, substance use disorders and end-of-life care. The review found preliminary but promising evidence for MDMA in the treatment of PTSD, and for psilocybin in the treatment of depressed mood and cancer-related anxiety. A 2019 systematic review of RCTs on psychedelics in mood and anxiety disorders (7 studies, total *N* = 130 subjects) found that low doses of ayahuasca, psilocybin, and LSD consistently resulted in rapid onset significant antidepressant and anxiolytic effects that often persisted for months following treatment. A systematic review of RCTs on psychedelics in unipolar depressed mood that included 19 studies (total *N* = 423) found that almost 80% of subjects experienced improvements in depressive mood symptoms. A different systematic review found that psychedelics also have beneficial effects on group therapy in diverse populations and clinical indications. A 2018 systematic review of systematic reviews on the efficacy, safety, and tolerability of psychedelic agents found evidence of beneficial effects of psilocybin and LSD on anxiety and depressed mood, alcohol and tobacco dependence, and of ayahuasca on depressed mood, including cases of treatment-refractory depression. The significance of findings on psychedelics as treatments of unipolar depressed mood and other psychiatric disorders is limited by the small size of many studies, absence of blinding in many studies, and short study duration.

Psychedelics are generally well tolerated when taken at doses established to have antidepressant effects and do not carry a risk of dependence. Psychedelics can cause intense negative psychological effects, including paranoia and psychosis, and it is prudent to take them in medically controlled environments. Concomitant use of psychedelics and SSRIs may result in serotonin syndrome and should be avoided. Commonly reported generally mild adverse effects include transient anxiety, nausea, headaches, and mild transient increases in heart rate and blood pressure.

Somatic Therapies

Exercise. Infrequent physical activity is an established risk factor for developing depressed mood. Physical exercise increases the levels of brain-derived neurotrophic factors and may enhance neural plasticity and new synapse formation. Regular exercise is associated with increases in the relative size of the frontotemporal and parietal lobes, which are important centers for learning, memory, and executive functioning. Brain levels of endorphins, dopamine, norepinephrine, and serotonin are increased following sustained exercise. Two meta-analyses of controlled trials confirmed consistent positive effects of regular exercise on depressed mood. Aerobic conditioning and strengthening exercises are equally effective against depressed mood. The optimum duration and frequency of exercise for depression have not been determined but are probably related to age and fitness level. In a 16-week study (*N* = 156), depressed patients older than 50 years randomized to aerobic exercise three times a week versus sertraline (up to 200 mg per day) versus exercise plus sertraline experienced equivalent improvements in standardized measures of mood, self-esteem, and negative thoughts. Patients in the antidepressant-only group initially improved faster. However, patients in the exercise-only group had a lower 6-month relapse rate. The antidepressant effects of running or fast walking may be equivalent to those of CBT and conventional antidepressants for moderate depressed mood. Depressed individuals who exercise in a brightly lit (2,500 to 4,000 lx) indoor environment reported greater improvements in mood and vitality than depressed individuals who exercised indoors in ordinary room light (400 to 600 lx).

Routine physical activity is associated with a reduced risk of all categories of dementia. More than 2,000 physically nonimpaired men aged 71 to 93 years were followed with routine neurologic assessments at 2-year intervals starting in 1991. At the end of the study period, men who walked less than 1/4 mile daily had an almost twofold greater probability of being diagnosed with any category of dementia than men who walked at least 2 miles daily. Findings of the Nurses' Health Study based on biannual mailed surveys over 10 years showed that women aged 70 to 81 years who engaged in regular vigorous physical activity were significantly less likely to have been diagnosed with dementia than women with more sedentary lifestyles. A small RCT found that regular daily exercise in individuals with moderate dementia receiving in-home care reduced depressed mood but did not improve cognitive functioning, suggesting that regular exercise does not slow the rate of cognitive decline once dementia has begun.

A period of 20 to 30 minutes of regular daily exercise significantly reduces symptoms of generalized anxiety. A prospective 10-week study on exercise in individuals diagnosed with panic disorder found that regular walking or jogging (4 miles three times a week) reduced the severity and frequency of panic attacks. In a small open study, alcoholics who regularly exercised while hospitalized for acute detoxification reported significant improvements in mood and general well-being. Abstinent alcoholics enrolled in outpatient recovery programs reported improved mood with regular strength training or aerobic exercise. A systematic review of studies on the relationship between exercise and sleep in the older adults concluded that exercise probably enhanced sleep quality and improved overall quality of life.

Expressive Movement Therapies. Dance therapy and other expressive movement therapies are widely used in both outpatient and inpatient settings as treatments for depression, anxiety disorders, and schizophrenia. Most studies on expressive movement therapies are qualitative in design and difficult to analyze because of methodologic differences, small study sizes, and the absence of blinding. However, evidence suggests that the regular practice of dance therapy and other expressive movement therapies is associated with significant reductions in the severity of anxiety and depressed mood.

Animal-Facilitated Psychotherapy. Animal-facilitated therapy provides a sense of social connection and renewed purpose and has beneficial effects on mood, anxiety, and global cognitive functioning in both outpatient and inpatient settings. Animal-facilitated psychotherapy has not been rigorously evaluated in controlled trials; however, observational studies suggest that regular contact with therapy dogs and other animals may reduce anxiety, improve mood, and increase attention span. Preliminary findings suggest that the use of animal companions in the context of psychotherapy probably improves general emotional well-being in depressed and anxious persons.

Massage. The anxiety-reducing effects of massage are believed to be mediated by decreased cortisol and increased parasympathetic tone. However, the mental health benefits of massage are difficult to quantify. Findings of open trials suggest that regular massage therapy reduces the severity of chronic moderate state anxiety related to test-taking, problem-solving, or work stress. A review of massage techniques for depressed mood identified several open studies with positive outcomes.

Snoezelen. Snoezelen is a multimodal treatment of dementia that employs music, light patterns, tactile surfaces, and essential oils to stimulate the senses of sight, hearing, touch, taste, and smell. Snoezelen sessions are usually administered two to three times weekly to patients with dementia in skilled nursing homes. A systematic review of RCTs of snoezelen in dementia concluded that apathy, speech skills, and psychomotor agitation improved in demented patients after four to eight sessions.

Mindfulness Training and Mind–Body Practices. Mindfulness training includes many styles of meditation and guided imagery. Mind–body practices include *taijijuan, qigong,* yoga, and others. Improved capacity for focused attention and reflection has been proposed as an important nonspecific psychological benefit of mind–body practices. MBSR and mindfulness-based cognitive therapy (MBCT) are secular approaches widely used to treat mental health problems. Research findings support that MBCT is a beneficial adjunctive therapy for unipolar depression and both MBSR and MBCT are beneficial adjunctive interventions for anxiety disorders. The regular practice of a mind–body skill may be as effective as CBT or antidepressants in patients with moderately depressed mood. Mindfulness training combined with cognitive therapy may enhance outcomes compared to cognitive therapy alone. Findings of a small open pilot study suggest that regular use of MBCT may significantly improve executive functioning, memory, and ability to complete tasks in stable patients diagnosed with bipolar disorder. Improvements in executive functioning persisted up to 3 months after treatment ended. Twelve individuals diagnosed with bipolar disorder who participated in 12 weekly group sessions of MBCT exhibited increased mindfulness, lower residual depressive mood symptoms, increased capacity for emotional regulation abilities, and generally improved psychological well-being and psychosocial functioning. Large prospective trials are needed to confirm these findings and determine the optimal type and frequency of MBCT for cognitive symptoms of bipolar disorder.

Relaxation Techniques. Relaxation techniques include sustained deep breathing, progressive muscle relaxation, guided imagery, and systematic desensitization. The mechanism of action may involve reducing muscle tension using behaviors or cognitions that diminish tension and autonomic arousal. The regular practice of progressive muscle relaxation or sustained deep breathing at bedtime reduces the time needed to fall asleep in chronic insomniacs. Individuals who have difficulty falling asleep because of somatic complaints respond better to progressive muscle relaxation, whereas those who stay awake because of chronic worrying are more likely to sleep better using guided imagery. A meta-analysis of 66 studies (total *N* = 2,000) on somatic and mind–body treatments of insomnia concluded that all interventions resulted in “reliable and durable benefits” that were superior to placebo, including improved quality of sleep and reduced sleep latency. A separate meta-analysis concluded that nonpharmacologic treatments of chronic insomnia are initially more expensive and require more time than conventional drug therapy but are reliable, yield sustained benefits, and are thus more cost-effective in the long run.

In a 5-month prospective study, individuals diagnosed with GAD randomized to relaxation only versus antidepressants plus relaxation experienced equivalent and significant improvements in state anxiety levels by the end of the trial. In a small RCT (*N* = 36), anxious adult outpatients randomized to 12 weekly sessions of applied relaxation versus conventional cognitive therapy experienced significant and comparable reductions in anxiety. Depressed patients who had previously been diagnosed with somatoform disorders experienced greater improvements in mood when guided imagery was combined with antidepressants than with antidepressants alone. In an open trial, 60 women reporting anxiety and postpartum depressed mood experienced significant reductions in both anxiety and depressed mood using a combined relaxation-guided imagery protocol during the first four weeks after childbirth. Guided imagery has beneficial effects on the immune system, physiologic stress responses, and general mental-emotional functioning.

Yoga. More studies have been done on yoga than any other mind–body discipline. The calming effects of yogic breathing are probably related to the modulation of parasympathetic–sympathetic tone. The regular practice of yoga postures (asanas) probably results in beneficial changes in autonomic activity, resulting in improved cardiorespiratory performance and increased feelings of general well-being. A review of recent studies on yoga in depressed persons reported significant positive findings, but most studies were limited by methodologic flaws. Combining a mindfulness practice with yoga may have greater mood-enhancing benefits than forms of yoga that do not include mindfulness. Moderately depressed patients improved significantly by the end of a 5-week class in Sudarshan kriya (SK) yoga. Improvements associated with SK breathing practices in severely depressed hospitalized patients were comparable to responses achieved with antidepressants and only slightly less robust than patients receiving electroconvulsive therapy. Preliminary findings support that the regular practice of forms of yoga that do not include specialized breathing exercises also improve depressed mood, including hatha yoga, Omkar meditation, and Iyengar yoga. Regular SK yoga practice may reduce the need for anxiolytics in individuals diagnosed with GAD. Findings of a small RCT suggested that the regular practice of a specific kundalini yoga left-nostril breathing technique may significantly reduce symptom severity in patients with OCD. A review of studies on yoga and other mind–body practices used to treat PTSD found that many approaches in current use reduce some PTSD symptoms, including intrusive memories, avoidance, and emotional arousal.

Meditation. Meditation has been extensively investigated as a treatment of anxiety. Beneficial physiologic effects of meditation include decreased oxygen consumption, respiratory rate, and blood pressure, and decreased autonomic arousal. Findings of RCTs show that the consistent practice of mindfulness meditation significantly reduces generalized anxiety. MBSR is an eclectic approach that incorporates elements of different Eastern meditation practices and Western psychology. In one series, the majority of anxious individuals (*N* = 322) who successfully completed a 10-week MBSR program reported significantly decreased physical and emotional distress, improved quality of life, enhanced general well-being, greater optimism, and increased feelings of control. Patients diagnosed with irritable bowel syndrome (IBS), a frequent concomitant of generalized anxiety, experienced significantly fewer symptoms of IBS and anxiety when practicing two brief (15 minutes) daily sessions of mindfulness meditation.

Studies have evaluated mindfulness training, mantra reciting, and compassionate meditation (i.e., “Vipassana”) for their potential beneficial effects in PTSD. A review of meditation practices aimed at preventing PTSD found more evidence supporting mindfulness meditation than mantra reciting or compassionate meditation. Preliminary findings suggest that symptoms of reexperiencing and psychic numbing may be less responsive to meditation than other PTSD symptoms. In a 12-week pilot study (*N* = 5), veterans from the middle east conflict diagnosed with PTSD who trained in transcendental meditation experienced significant improvements in overall quality of life and reductions in core PTSD symptoms.

Therapies Based on Scientifically Validated Forms of Energy and Information

CAM approaches based on scientifically validated forms of energy or information include biofeedback, sound, full-spectrum bright light exposure, dim blue light, high-density negative ions, virtual reality graded exposure therapy (VRGET), brain–computer interface (BCI) technologies, and microcurrent electrical stimulation. Research findings point to both direct and indirect effects of well-understood forms of energy and information on the psychological and neurophysiologic causes of mental illness.

Music and Sound. In a small RCT (N = 40), anxious adult patients were randomly assigned to conventional cognitive therapy versus music-assisted reframing. Patients in the music group experienced greater reductions in overall anxiety. When headphones are used to route sounds of slightly different frequencies to the right and left hemispheres, the brain stem “constructs” binaural beats based on the frequency difference between sounds processed in each hemisphere. Preliminary findings support that certain binaural beats (16 to 24 Hz) may improve depressed mood, enhance attention, and improve short-term and immediate recall in healthy adults, whereas binaural beats at lower frequencies may consistently induce a calm, relaxed state. Although antidepressant effects of music have not been established, frequent listening to music probably has beneficial effects on endorphins and other neurotransmitters that enhance mood.

Findings of RCTs suggest that music alone or in combination with guided imagery improves mood in depressed patients with cancer. Patients with cancer who participated in weekly sessions of combined music and guided imagery reported greater improvements in depressed mood and overall quality of life than patients in a wait-list group. Depressed outpatients reported significant improvements in mood and beneficial changes in heart rate and blood pressure soon after listening to tranquil music. A meta-analysis of studies on music therapy in individuals with dementia concluded that regular singing, engaging in dance therapy, listening to music, and participating in musical games improve cognitive and behavioral functioning in individuals with severe dementia, resulting in reduced agitation, reduced wandering, enhanced social interaction, improved mood, reduced irritability or fear, and increased cooperative behavior. Calming background music significantly reduces irritable behavior, anxiety, and depressed mood in nursing home patients with dementia. Regular music therapy reduces irritability and improves expressive language in demented individuals.

Light. Light of different intensities and colors is used in both biomedicine and many nonallopathic systems of medicine to treat diverse medical and psychiatric disorders. Entrainment of sleep–wake cycles by external bright light cues and the associated suppression of melatonin production by the pineal gland is the established mechanism of action underlying the therapeutic benefits of light exposure on sleep and daytime fatigue. Exposure to natural sunlight, especially in the early morning, has significant antidepressant effects and may reduce the length of hospital stays in severely depressed inpatients. A systematic review of controlled studies confirmed antidepressant effects of bright light (10,000 lx) exposure therapy in seasonal depressed mood but provided only limited evidence supporting bright light as a treatment of nonseasonal unipolar depressed mood. A more recent meta-analysis of controlled studies concluded that bright light exposure or dawn simulation for *seasonal* depressed mood and bright light exposure (but not dawn simulation) for *non*seasonal depression had comparable efficacy to antidepressants. Findings of a small RCT suggested that bright light therapy may be an effective alternative to antidepressants in pregnant depressed women. Recent studies suggest that regular exposure to dim red or blue light might be as efficacious as bright light, especially in seasonal depressed mood. In a 4-week single-blind study (N = 57), patients diagnosed with seasonal affective disorder were randomized to daily bright light versus dim red light exposure. Both groups experienced an equivalent and significant (40%) reduction in symptoms. Preliminary findings suggest that antidepressant effects of early-morning exposure to narrow-spectrum blue or green light may be equivalent to benefits obtained from full-spectrum bright light exposure.

Exposing depressed patients to low-intensity artificial light approximately 2 hours before they would naturally be exposed to early-morning daylight increased the speed of response to antidepressants.

Thirty depressed inpatients treated with citalopram 40 mg/day were randomized to early-morning dim green light, 400 lx, versus a sham, non–light-emitting device during the first 2 weeks of drug treatment. Patients in the combined citalopram–light exposure group reported significantly greater and more rapid improvements in mood compared to the citalopram-only and placebo groups.

An expert consensus report on light treatment for sleep disorders concluded that there is compelling evidence for the efficacy of bright light exposure in the management of circadian rhythm sleep problems but not for other types of insomnia. Most protocols recommend 30 to 40 minutes of bright light exposure daily for beneficial effects in shifting circadian rhythms and changing sleep–wake cycles. Beneficial effects of bright light exposure therapy are usually noticeable within a few days, and an appropriate treatment plan often resynchronizes the patient's sleep–wake cycle with their time zone in 2 to 3 weeks. Correct timing of light exposure is essential for successful treatment of circadian sleep problems, and bright light exposure between the hours of the late morning and late afternoon probably has no effect on circadian rhythm phase. Preliminary findings suggest that melatonin used in conjunction with bright light exposure may achieve more rapid normalization of the sleep–wake cycle.

Some patients exposed to bright morning light at 10,000 lx on a regular basis reported transient side effects, including mild jitteriness or headaches (10%) and mild nausea (16%). Sporadic cases of hypomania have been reported, especially in winter depressives or bipolar patients exposed to early-morning bright light. Almost two-thirds of patients who use bright light exposure therapy in the evening report insomnia. Patients interested in bright light exposure therapy should be cautioned to avoid using sun lamps (which emit harmful ultraviolet rays) and instructed to keep the eyes open while viewing the light fixture at an oblique (approximately 45 degrees) angle to avoid excessive eye strain. No adverse effects have been reported with regular exposure to dim light.

High-Density Negative Ions. Regular daily exposure to high-density negative ions may be an effective treatment of seasonal depressed mood and may have comparable efficacy to bright light exposure for this condition. Twenty-five depressed patients with seasonal depressed mood were randomized to high-density negative ions (2.7×10^6 ions/cm^3) versus low-density negative ions (1×10^4 ions/cm^3) using in-home ion generators 30 minutes daily for 3 weeks. Fifty-eight percent of patients exposed to high-density negative ions experienced significant improvements in mood on standardized rating scales, compared to 15% of patients exposed to low-density negative ions. In a 2-week RCT (N = 158), patients with seasonal depressed mood were randomly assigned to bright light exposure (10,000 lx) versus high-density or low-density negative ions. Patients exposed to high-density negative ions or bright light experienced significant and equivalent improvements in mood. In an RCT (N = 128), bright light exposure, a pleasing auditory stimulus, and high-density negative ions resulted in equivalent and rapid improvements in mood and alertness in mildly depressed and nondepressed adults. A review of studies on the effects of air ionization on mental health found evidence for a consistent antidepressant effect of high-density negative ions (but not low-density ions) on unipolar depressed mood and seasonal depressed mood and no consistent effects on anxiety, mood, relaxation, sleep, and measures of well-being. Adverse effects have not been reported with regular daily exposure to negative ions following the protocols used in published studies.

Biofeedback Using EEG and HRV. Electroencephalographic (EEG) biofeedback and HRV biofeedback employ information extracted from the body's intrinsic electromagnetic rhythms as sources of feedback. EEG biofeedback (also called *neurofeedback*) translates electrical brain activity into a visual or auditory signal with the goal of modulating the *general arousal level* of the brain or correcting *dysrhythmias* in specific brain regions or frequency domains to achieve therapeutic changes in psychological and psychosomatic symptoms associated with depressed mood, bipolar disorder, generalized anxiety, PTSD, social anxiety disorder, attention-deficit disorder, and other psychiatric disorders. HRV biofeedback is based on the theory that positive emotional states result in increased *coherence* of several physiologic aspects of heart rhythm and that feed-forward and feedback loops between the heart, brain stem, and neocortex result in complex relationships among specific emotional or mental states and specific patterns of HRV.

Double-blind studies have evaluated the efficacy of EEG biofeedback (also called *neurofeedback*) as a treatment of anxiety, depressed mood, attention-deficit disorder, and PTSD. Sustained “shifts” in the dynamic equilibrium of electrical brain activity have been proposed as a general mechanism underlying the therapeutic effects of EEG biofeedback. Case reports and controlled studies show reductions in the severity of depressed mood using photic stimulation or alpha-asymmetry EEG biofeedback protocols. Beta-frequency training in EEG biofeedback may result in improved performance on standardized intelligence tests in nonimpaired individuals. EEG biofeedback has been successfully used to treat depressed mood in chronic alcoholics and to reduce relapse rates. Regular EEG biofeedback training and conventional antianxiety medications may have comparable efficacy for the management of generalized anxiety.

Recent research findings support that EEG biofeedback involving very low frequencies, between 0.02 and 0.2 Hz, results in rapid, significant reductions in the severity of PTSD symptoms and improvements in global cognitive functioning. It has been suggested that *infra-low* frequencies (ILF) may induce therapeutic shifts in the functional connectivity of the brain's resting-state networks resulting in reduced overall arousal, enhanced cognitive functioning, and increased emotional resilience. At a large military base, more than 500 active-duty combatants diagnosed with PTSD were trained in ILF EEG biofeedback. Training was done one to three times per week and was administered by licensed psychotherapists certified in EEG biofeedback therapy. Symptom severity was evaluated weekly using the PCL-5—the military version of the PTSD Checklist. A cohort analysis of 300 Marines in the original cohort found that 75% of persons with moderate to severe PTSD symptoms experienced significant clinical improvement, including reductions in psychological, cognitive, psychophysiological, and physiologic symptoms. Twenty-five percent of subjects in the cohort reported that all PTSD symptoms had completely resolved with fewer than 20 EEG biofeedback sessions; another 50% experienced significant reductions in symptom severity after 40 sessions. The remaining subjects took much longer to respond to treatment, continued to report clinically significant symptoms, discontinued training prematurely, or were nonresponsive to the neurofeedback training protocol.

In a review of nonpharmacologic treatments of insomnia, EEG biofeedback was found to be more effective than progressive muscle relaxation. The American Academy of Sleep Medicine recommends EEG biofeedback as a treatment of chronic insomnia. Several EEG biofeedback protocols are beneficial in insomnia, but a protocol that modulates alpha and theta activity along the primary sensory and motor regions in the frontal and prefrontal areas of the cortex may yield the most consistently beneficial outcomes in improving the overall quality of sleep and reducing the frequency of middle awakenings.

Biofeedback using information obtained from measures of HRV may have therapeutic benefits in anxiety, depressed mood, and PTSD. Abnormal low HRV is associated with deficits in attention and short-term memory in combat veterans diagnosed with PTSD. Case reports support that HRV biofeedback may improve depressed mood and anxiety resulting from chronic stress and significantly reduces arousal, and improves general emotional well-being in individuals who are subjected to acute job-related stress. Beneficial changes in baseline anxiety achieved with HRV biofeedback are associated with decreased serum cortisol levels and increased serum DHEA levels. In a 4-month RCT, the majority of 29 police officers trained in HRV biofeedback reported significant improvements in baseline anxiety compared to 36 officers assigned to a wait-list group who did not improve. Individuals diagnosed with GAD reported significant improvements in general emotional well-being and reduced baseline anxiety with regular HRV biofeedback training. Preliminary findings support that increased cardiac coherence may lessen the severity of PTSD; however, results are inconsistent. In a small 3-week open exploratory study, a group of active-duty service members diagnosed with PTSD or depressed mood who received HRV biofeedback plus treatment as usual did not report greater reductions in symptom severity compared to a group receiving treatment as usual only. Correlations between electrocardiogram (EKG) and EEG activity provide evidence for complex feed-forward and feedback loops between the heart and brain that regulate emotion. This integrative biofeedback approach is called “energy cardiology” and is only beginning to be examined in controlled trials.

Virtual Reality Graded Exposure Therapy. VRGET is a rapidly evolving therapy with important implications for the management of PTSD and severe phobias that are difficult to treat using more conventional therapies. VRGET combines advanced computer graphics, three-dimensional visual displays, and body-tracking technologies to create realistic virtual environments with the goal of simulating feared situations or objects. Virtual environments have been designed to provide visual, auditory, tactile, vibratory, vestibular, and olfactory stimuli to patients in highly controlled settings. During a virtual exposure session, the therapist closely tracks the patient's state of arousal by monitoring physiologic indicators of stress, including heart rate and respiration. Patients are ready to practice in vivo exposure to the feared object or situation after they have been desensitized to a virtual environment.

VRGET is more effective than conventional imaginal exposure therapy and has comparable efficacy to in vivo exposure therapy for the treatment of specific phobias, agoraphobia, panic disorder, and PTSD. Like in vivo and imaginal exposure therapy, VRGET desensitizes the patient to a situation or object that would normally cause anxiety or panic. In one RCT, VRGET and conventional CBT were equally effective in panic disorder with agoraphobia. Patients who underwent VRGET required 33% fewer sessions to achieve similar results. RCTs confirm that VRGET is an effective treatment of fear of flying, fear of heights, fear of small animals, fear of driving, and other specific phobias. VRGET is as effective as and more cost-effective than in vivo exposure therapy for fear of flying because patients and therapists avoid significant time commitments and the need to use airplanes.

A virtual environment simulating the devastation of the September 11, 2001, attacks on the World Trade Towers has been successfully used to treat individuals diagnosed with severe PTSD following the attacks. Findings of a study on combined multisensory exposure and VRGET reported significant reductions in the severity of PTSD symptoms in active-duty combatants who had failed to respond to other forms of exposure therapy. Several subjects reported significant improvement following only five VRGET sessions; however, there was considerable variability in the number of VRGET sessions needed to reduce symptom severity to the same level. Combining VRGET with D-cycloserine, a partial NMDA agonist may yield differentially greater improvement in symptoms of acrophobia compared to VRGET alone.

Considerable research is ongoing to develop virtual reality tools for assessing and preventing combat-related PTSD. STRIVE (stress resilience in virtual environments) is a highly integrative "stress resilience training" program aimed at enhancing emotional coping strategies prior to active deployment. STRIVE employs an immersive VR environment to simulate combat situations that includes a "virtual mentor" who guides the combatant through a virtual experience while coaching them in relaxation and emotional self-regulation skills. The intensity of the virtual stimulus used is determined by the individual's habituation based on HRV and other measures of autonomic arousal. Physiologic biomarkers of the stress response are measured before and after VRGET sessions. The STRIVE system permits users to be immersed in stressful combat scenarios and interact with virtual characters for training in a variety of coping strategies that may enhance resilience in the face of extreme stress. The STRIVE protocol may provide a useful tool for predicting the risk of developing PTSD or other psychiatric disorders in new recruits prior to actual combat exposure. It has been suggested that recruits who display high resilience and are presumably at relatively lower risk of developing PTSD might be more suitable for direct combat roles, while individuals who display low resilience might preferentially be assigned to noncombat roles.

Future integrative approaches to phobias, panic attacks, and other severe anxiety disorders will combine VRGET with biofeedback in outpatient settings or in the patient's home via broadband Internet connections, with CBT, relaxation, mind–body practices, and appropriate medications. Combining VR environments with real-time feedback based on neurophysiologic responses to stress may permit each unique patient to optimize the level and type of VR exposure to enhance resiliency training and speed the rate of recovery from PTSD. Human–computer interface (HCI) systems based on CBT and biofeedback are being developed for resilience training in individuals at risk of developing PTSD following exposure to trauma. Larger studies on patient populations diagnosed with PTSD using head-mounted displays and other technologies that create more immersive virtual environments are needed to determine whether combining VRET and EEG biofeedback is practical in clinical settings and yields superior outcomes compared to either approach alone.

Infrequent cases of disorientation, nausea, dizziness, headache, and blurred vision have been reported following VRGET. Intense sensory stimulation during VRGET can trigger migraine headaches, seizures, or gait abnormalities in individuals who have these disorders. Patients with schizophrenia should not use VRGET because immersion in a virtual environment can exacerbate delusions.

Cranioelectrotherapy Stimulation. Cranioelectrotherapy stimulation (CES), also called microcurrent electrical stimulation, is based on the application of extremely weak electrical current to the head and neck to treat anxiety, insomnia, and substance use disorders. The mechanism of action may involve stimulation of endorphin release and induction of EEG frequency changes that reduce the severity of anxiety. A meta-analysis of controlled trials comparing CES with a sham treatment found that generalized anxiety improved in seven of eight studies, and the magnitude of improvement reached statistical significance in four studies. A review of 34 sham-controlled trials concluded that regular CES treatments resulted in short-term symptomatic relief of generalized anxiety through direct effects on autonomic brain centers. Patients diagnosed with one or more phobias reported significant reductions in state anxiety when exposure to the phobic stimulus was followed by 30 minutes of CES treatment. Daily application of weak electrical current to the neck or the head may stimulate global brain activity, resulting in beneficial modulatory changes in neurotransmitters implicated in dementia. A meta-analysis of three studies on CES in patients with dementia found evidence of significant but transient improvements in word recall, face recognition, and motivation immediately following treatment that was not sustained at 6-week follow-up.

CES is also used to lessen symptoms of alcohol and drug withdrawal. The postulated mechanism of action involves stimulation of release of endorphins, enkephalins, and other endogenous opioid peptides and may prove to be similar to the mechanism underlying electroacupuncture. In a 4-week RCT (*N* = 20), depressed alcoholics were randomized to receive 20 CES treatments at 70 to 80 Hz, 4 to 7 mA, versus sham treatments. Patients who received CES treatments experienced significantly reduced anxiety by the end of the study. Findings of a sham-controlled study (*N* = 60) on inpatient alcohol or polysubstance abusers suggest that daily 30-minute CES treatments significantly improves cognitive functioning and reduces stress and anxiety during the acute phases of withdrawal. In a 7-year prospective study of CES in the treatment of alcohol, drug, and nicotine addiction, acute and chronic withdrawal symptoms were diminished, normal sleep patterns were restored more rapidly, and more patients remained addiction-free following regular CES treatments compared to those receiving psychopharmacological management. The discrepancy between largely positive outcomes from studies on CES and frequent negative findings on electroacupuncture (see the section on acupuncture) in detoxification may be partly attributable to suboptimal current or frequency settings in electroacupuncture protocols investigated for this clinical application.

CES is widely used to treat insomnia; however, the findings of sham-controlled studies are inconsistent. Whereas some studies found no beneficial effects of CES in insomnia, others have reported sustained improvements in the timing of sleep onset and total sleep duration. Disparate findings on the therapeutic benefits of CES in insomnia may be related to differences in equipment, protocols, and duration, and timing of treatment. Serious adverse effects of microcurrent electrical brain stimulation have not been reported.

Therapies Based on Postulated Forms of Energy or Information Not Validated by Current Science

In contrast to scientifically *validated* modalities, treatments based on postulated *subtle* forms of energy or information are not verified by *current* science. CAM approaches based on hypothesized *subtle* forms of energy include acupuncture, homeopathic remedies, HT, *qigong,* Reiki, and energy psychology. Beneficial effects of these modalities rest on the postulated role of poorly characterized energetic *fields* or directed human intention that are difficult to examine using contemporary research methods and may, in fact, not be *explainable* by current science. Chinese medicine, Ayurveda, and Tibetan medicine posit the existence of a subtle “energy” body and a gross physical body. All Asian healing traditions use physical, psychological, and spiritual approaches to treat postulated *energetic* imbalances and restore optimal health. In Chinese medicine, “*qi*” is regarded as a fundamental kind of energy that cannot be explained by contemporary science but may have properties consistent with quantum mechanics. Ayurveda posits an analogous energetic principle called “*prana.*” In both systems of medicine, a central goal of treatment is to restore “balance” in the energetic and physical body, thus enhancing the body’s capacity for self-healing.

Like Asian healing traditions, homeopathy posits that physical and emotional symptoms result from imbalances in a fundamental *vital force.* Current science cannot explain the postulated curative mechanism of homeopathic remedies, which are typically *diluted beyond the limit of possible biologic action*. Reiki holds that a fundamental energetic principle, "*ki,*" promotes healing when guided by a skilled practitioner. Physical contact between the practitioner and the patient is not believed to be necessary for beneficial effects to take place. HT practitioners claim that "energetic" contact between the practitioner and the patient results in the alleviation of symptoms. In HT, the practitioner positions the hands over specific areas of the body but does not physically *touch* the patient's body. TT is similar to HT but uses gentle physical contact to promote healing. In contrast to HT and TT, the Reiki practitioner does not consciously direct healing intention to specific regions of the body but serves as a "conduit" for *ki* to *flow into* the patient to restore healthy energetic functioning. Somatoemotional release is an energy therapy in which the practitioner "releases" pathologic energetic states that accumulate in the body following physical injury. The technique involves gentle touch while guiding the patient to assume positions that stimulate body memories of past trauma.

Energy psychology is an eclectic combination of Western psychological theory and Chinese medical theory that posits that energetic imbalances in the meridians (i.e., pathways through which *qi* flows) are associated with different physical, emotional, and mental symptoms. "Thought field therapy" (TFT) and "emotional freedom technique" (EFT) are specific energy psychology approaches used to treat anxiety and depression. In TFT, the patient is asked to invoke a "thought field" associated with depressed mood, a traumatic memory, or other principal symptom pattern. The TFT practitioner then *reattunes* energetic imbalances associated with the target symptom by tapping on specific acupuncture points in a specified manner in an effort to alleviate symptoms. EFT is a simplified version of TFT that involves only one routine for stimulating acupuncture points. Twelve of 18 RCTs included in a review of noncontact *biofield* therapies including Reiki, HT, Johrei, and TT reported statistically significant positive outcomes; however, findings were limited by the small sample size of most studies and methodologic problems.

Acupuncture. Evaluating the clinical efficacy of acupuncture in depressed mood and other mental health problems poses complex methodologic challenges because of differences in the severity and comorbidity of mental, emotional, and physical symptoms in study populations; concurrent use of other conventional or CAM treatments; conceptual differences between Chinese medicine and biomedical diagnoses; and the use of individualized acupuncture treatment protocols reflecting each patient's unique "energetic" balance. Sham-controlled studies suggest that conventional needle acupuncture, electroacupuncture, and computer-controlled electroacupuncture (CCEA) have consistent beneficial effects on depressed mood. The findings of a double-blind, sham-controlled study support that traditional manual acupuncture (i.e., needles in the absence of electrical current) is an effective treatment of severe depressed mood. By the end of the 8-week study, 68% of 33 female outpatients being treated with a specialized acupuncture protocol directed at depressed mood had achieved full remission. However, matched depressed women in a wait-list group showed equivalent improvement. In a large 6-week, multicenter study (N = 241), depressed inpatients randomized to receive electroacupuncture plus placebo versus electroacupuncture plus amitriptyline experienced equivalent improvements in mood. Significantly, electroacupuncture was found to be superior to amitriptyline in patients who did not report comorbid anxiety. Patients treated with electroacupuncture had significantly elevated plasma norepinephrine concentrations following a 6-week course of treatment, suggesting that the mechanism of action of electroacupuncture involves stimulation of norepinephrine release. CCEA uses computer-guided modulation of the frequency and waveform of electrical current delivered through acupuncture needles. Preliminary findings suggest that high frequencies (up to 1,000 Hz) may yield responses in depressed patients superior to both conventional acupuncture and electroacupuncture. A narrative review of controlled studies, outcomes studies, and published case reports on acupuncture as a treatment of anxiety and depressed mood supported that sham-controlled studies are associated with consistent improvements in anxiety using both conventional acupuncture and electroacupuncture.

A review of prospective trials on acupuncture as a treatment of PTSD identified four quality sham-controlled studies and two quality uncontrolled trials that met inclusion criteria. One high-quality trial included in the review showed statistically significant differences between the acupuncture and wait-list group but nonsignificant differences between the acupuncture and CBT groups. Patients receiving acupuncture or CBT continued to report clinical improvements in PTSD symptoms 3 months after the study endpoint.

Findings on acupuncture for relapse prevention in alcohol abuse are inconsistent, reflecting differences in the selection of acupuncture points, the treatment protocol used (i.e., conventional vs. electroacupuncture), frequency of treatments, the duration of total treatment, and the relative skill or specialized training of practitioners. While some sham-controlled trials have reported significant reductions in alcohol craving and relapse rate in recovering alcoholics, other studies have found nonsignificant differences in craving or relapse rates between subjects receiving an acupuncture protocol traditionally used to treat addiction, sham transdermal stimulation on random points, and a wait-listed group.

Findings of most controlled trials on acupuncture for smoking cessation have been negative or equivocal. Three meta-analyses of sham-controlled studies on the efficacy of acupuncture for smoking cessation (total *N* >2,000) concluded that therapeutic acupuncture protocols and sham acupuncture have equivalent efficacy, suggesting a significant placebo effect. Two systematic reviews of sham-controlled trials found that both conventional acupuncture and electroacupuncture are equally *ineffective* in reducing symptoms of nicotine withdrawal and controlling cocaine addiction. A large study examining three auricular acupuncture protocols for relapse prevention in narcotics abusers concluded that all three protocols were associated with reductions in drug use over time. An 8-week RCT (*N* = 32) compared acupuncture with two drugs versus placebo in cocaine addicts on methadone maintenance therapy. Half of the subjects dropped out; however, almost 90% of those who completed the study achieved abstinence by study end. Subjects who achieved abstinence reported diminished craving and significantly improved mood.

A systematic review of all studies on acupuncture for insomnia published in the English language through 2002 identified few studies for analysis because most studies had not employed randomized, double-blind procedures. Acupuncture may be beneficial for insomnia associated with generalized anxiety and schizophrenia. In a large-case series (*N* = 500), schizophrenics reported significant improvements in sleep following acupuncture treatments. In a small RCT (*N* = 40), patients complaining of insomnia were diagnosed using Chinese medical assessment methods and randomized to an acupuncture protocol addressing the specific energetic “imbalance” underlying their sleep disturbance versus a sham protocol. Significant and sustained improvements in the subjective quality of sleep and changes on EEG polysomnography reflecting normalization of sleep were found in the treatment group but not in the sham group. The absence of standardized acupuncture treatment protocols for insomnia has interfered with efforts to design rigorously controlled double-blind studies. The significance of findings of studies on acupuncture in psychiatric disorders is limited by small study size, the absence of standardized symptom rating scales in most studies, limited follow-up, and poorly defined differences between protocols used in different studies.

Uncommon transient adverse effects associated with acupuncture include bruising, fatigue, and nausea. Infrequent cases of infection with human immunodeficiency virus, hepatitis B, and hepatitis C have been reported when nonsterilized needles are used. Rare cases of pneumothorax and cardiac tamponade have been reported as a result of accidental puncturing of the lungs or the pericardium.

Homeopathy. Preliminary findings from open studies suggest that certain homeopathic remedies may improve depressed mood. Short-term and long-term constitutional remedies are used, depending on symptom severity, personality factors, and the “energetic pattern” associated with depressed mood. Homeopathic preparations are widely used to self-treat generalized anxiety and other anxiety symptoms; however, research findings are inconsistent and based largely on case reports. Case reports of symptomatic improvement following the administration of specific homeopathic remedies have been published for PTSD, social phobia, panic disorder, and OCD. However, the findings of two placebo-controlled studies were negative or equivocal. Students with test-taking anxiety (N = 62) randomized to a specific homeopathic preparation (*Argentum nitricum* 12X) versus placebo twice daily for 4 days reported equivalent modest reductions in anxiety. In a 10-week prospective study (N = 44), patients diagnosed with GAD were randomized to an energetically appropriate homeopathic treatment versus placebo. All subjects were medication-free at least 1 month prior to the start of the study. All subjects were individually diagnosed by a classically trained homeopathic physician and treated with a homeopathic remedy addressing identified “energetic imbalances.” Significant and *equivalent* improvements in baseline anxiety were found in the placebo and treatment groups at week 5 and at the end of the study. Homeopathic remedies are generally too dilute to potentially cause adverse effects. However, rare cases of toxicity have been reported when low dilutions of arsenic, cadmium, or other heavy metals are used.

Concerns about the possible carcinogenic effects of remedies containing *Aristolochia* have been raised, and there are rare reports of allergic reactions.

Healing Touch and Therapeutic Touch. HT practitioners claim to be able to intuitively assess energetic *imbalances* manifesting as physical or psychological symptoms and "*direct*" healing energy to specific bodily regions to correct purported imbalances, resulting in improved functioning or enhanced well-being. In a 3-week open study (*N* = 30), moderately to severely depressed patients were randomized to receive HT versus a waitlist. Individuals in the treatment group received two HT sessions weekly. Control patients received no treatment, including no exposure to a sham HT protocol. By the end of the study, patients in the treatment group experienced significantly improved mood that was sustained at 1-month follow-up. Open studies, case reports, and one double-blind trial suggest that HT and TT may reduce the severity of agitation in dementia. In one small open study, measures of agitation improved significantly in 14 individuals with dementia in a residential setting who received three HT treatments weekly over 4 weeks. Three individuals in the study needed fewer sedating medications during the active treatment phase, and two residents required dose increases in the first 2 weeks after HT treatments were stopped. In a small 5-week, sham-controlled study, two weekly 10- to 20-minute HT treatments were administered to patients with Alzheimer disease. Patients who received regular HT treatments displayed consistent reductions in disruptive behaviors, globally improved emotional and cognitive functioning, enhanced socialization, a more regular sleep schedule, improved compliance with nursing home routines, greater emotional stability, and improved communication with staff. Individuals who received regular HT treatments complained of physical distress or discomfort significantly less often than the control group. Measures of agitation improved significantly in the HT group but did not change in the control group. In a double-blind study (*N* = 57) that included mock-TT in the control arm, agitated patients with dementia who received two brief TT treatments daily for 3 days exhibited significantly fewer behavioral symptoms of dementia, including reduced restlessness and fewer disruptive vocalizations, compared to patients who received mock-TT. In a 3-week RCT (*N* = 123), active-duty military diagnosed with PTSD randomized to semi-weekly sessions of HT plus guided imagery experienced greater improvement in PTSD symptom severity and quality of life compared to those receiving treatment as usual. The significance of research findings on TT and HT is limited by study design flaws, including lack of blinding, the confounding influence of frequent human contact, small numbers of enrolled subjects, the lack of standardized methods, and the absence of control groups receiving sham HT or TT treatments in most studies.

Qigong. *Qigong* is an ancient energy healing method used in Chinese medicine that has its origins before written language. Most published studies have examined the effectiveness of regular *qigong* practice on general indicators of physical or emotional health. Individuals can engage in a regular *qigong* practice or receive "emission *qigong*" treatments for a particular medical or psychological problem from a skilled *qigong* master. Findings of open trials on *qigong* as a treatment of "neuroticism" suggest that anxiety, depressed mood, and quality of sleep were significantly better in subjects who had practiced *qigong* consistently for at least 5 years than in those who had never practiced *qigong.* An open study reported that depressed individuals who practiced *qigong* consistently for at least 2 years experienced improved mood and general improvement in their baseline emotional state. In a 2-month open pilot study ($N = 39$), patients with a diagnosis of major depressive disorder, dysthymic disorder, or bipolar disorder based on criteria of the text revision of the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR) received *qi* "emission" treatments administered by *qigong* masters. Half of the subjects engaged in the regular daily practice of *qigong.* By the end of the study, standardized ratings of depressed mood had improved for all subjects, and there were no significant differences in response between individuals who practiced *qigong* while receiving *qi* "emission" therapy and subjects who received *qi* "emission" therapy only. A systematic review and meta-analysis of studies on qigong "exercise" for stress and anxiety in adults identified seven RCTs that met the inclusion criteria. Four RCTs reported positive findings of qigong for generalized anxiety, and three studies reported beneficial effects on stress; however, the significance of findings was limited by methodologic flaws and small study size. A systematic review and meta-analysis of 15 studies on qigong for anxiety, depressed mood, and psychological well-being found no conclusive evidence for qigong as a treatment of a specific mental health condition; however, the findings suggested that regular qigong practice may have positive effects on general psychological well-being in patients with chronic illnesses.

The *unskillful* or excessive practice of *qigong* can result in serious psychological or psychosomatic symptoms. Psychotic symptoms that have been reported following *erroneous qigong* are described as a culture-bound syndrome, “*qigong* psychotic reaction,” in the DSM. Schizophrenics and individuals diagnosed with severe personality disorders should be cautioned to practice *qigong* moderately and only under the guidance of a skilled *qigong* instructor.

Reiki. Findings of a 6-week, single-blind, sham-controlled trial of Reiki on depressed mood (*N* = 73) suggested that regular Reiki treatments may increase feelings of general well-being and reduce stress. The study population consisted of volunteers who complained of depressed mood or “stress” but did not meet diagnostic criteria for a major psychiatric disorder. Subjects were randomized to weekly 1½-hour, hands-on Reiki versus distant Reiki or distant sham Reiki treatments. Patients who received hands-on or distant Reiki treatments experienced significant and equivalent improvements in depressed mood and stress, which were sustained 1 year after treatment was terminated. No one in the sham Reiki group improved during or following the study. In a small sham-controlled study (*N* = 12), chronic pain patients were randomized to receive Reiki versus sham Reiki, progressive muscle relaxation, or no treatment. Improvements in pain and state anxiety were significantly greater among patients receiving Reiki than sham-Reiki and relaxation training. Preliminary findings are suggestive of the therapeutic effects of Reiki. However, large sham-controlled studies in clinical populations are needed to evaluate the putative mood-enhancing and stress-reducing effects of hands-on and distant Reiki.

Energy Psychology. EFT, TFT, and body psychotherapy are “energy psychology” techniques used to treat mood disorders, phobias, and other anxiety disorders. Few controlled studies have been done, most studies are methodologically flawed, and findings are highly inconsistent. EFT treatments can be self-administered using a manual following an initial training session, and no technical expertise is required. In a double-blind, prospective trial (*N* = 35), patients who met DSM-IV-TR criteria for a specific phobia were randomized to receive one EFT treatment versus one session of diaphragmatic breathing (DB) while exposed to a feared small animal. EFT-treated patients experienced greater improvement than DB-treated patients in four of five measures of anxiety, and improvements were sustained at 3 months in the absence of further EFT or DB treatments. The above findings were replicated in a follow-up study. A study comparing EFT and sham-EFT in anxious individuals who had not been formally diagnosed found no differences in anxiety reduction between a defined EFT protocol and a sham protocol. These findings suggest that the therapeutic benefits of EFT may be attributable to desensitization and distraction. A sham-controlled study on EFT in depression found equivalent nonsignificant changes in mood following EFT and sham treatment.

Somatoemotional Release and Craniosacral Therapy. Practitioners of craniosacral therapy gently guide the spinal and cranial bones with the goal of bringing the CNS into greater “harmony.” The movement of the cerebrospinal fluid (CSF) is believed to be restricted by bodily trauma or chronic stress. Gentle manipulation of the spine, the skull, and fascia is believed to ease restrictions on the openings through which nerves pass while optimizing the flow of CSF through the spinal cord. Vietnam veterans (*N* = 22) diagnosed with PTSD completed a 2-week intensive program in which they were treated with both somatoemotional release and craniosacral therapy. The majority of veterans reported significant reductions in pain and physical distress and improvements in depressed mood, anxiety, suspiciousness, guardedness, and behavioral isolation. The significance of findings on somatoemotional release and craniosacral therapy is limited by the small study size, the absence of a sham treatment arm, lack of follow-up, and the use of nonstandardized symptom rating scales.

Shamanic Healing. A shaman is a person who is regarded by their tribe or sociocultural group as having the power to heal the sick, predict the future, and communicate with the supernatural world, including primitive natural forces and spirits of the dead. The term *shaman* has ancient roots in Siberia. However, similar traditions of spiritual healing can be found in all world regions. Shamanic healing ceremonies often focus on retrieving the patient’s *lost soul*, which is believed to be a root cause of many physical and psychological illnesses. The therapeutic goals of ritual acts, stories and chants used in shamanic healing vary widely depending on the parent culture. Research on shamanic healing is generally descriptive, and there are few published studies using clinical research methodology. One study reported on shamanistic contributions to the treatment of schizophrenia and other mental illnesses in Japan. Another study examined the outcomes and satisfaction of 188 patients treated by shamans in Taiwan with a matched sample of patients treated by physicians. Five distinctive sickness types were identified, including mental health problems. Patient and experimenter reports suggested that three-fourths of patients improved. However, shamanic treatment was more likely to be described as difficult or uncomfortable than conventional medical treatment. Findings of a more recent study suggested that shamanic healing approaches contribute to the psychiatric stability and general emotional well-being of individuals diagnosed with bipolar disorder.

Outcomes research on shamanic healing is at a very preliminary stage. Applying standard biomedical research designs to indigenous healing practices poses many complex challenges, resulting in studies that are primarily naturalistic and observational in design. However, increasing interest in indigenous healing practices in the context of promising early findings on shamanic healing and the importance of fostering cultural belief systems in community mental health care will probably lead to the growing acceptance of shamanic healing in North America, Western Europe, and other developed world regions.

FURTHER READINGS

Ashford JW, Mahoney L, Burkett T. A role for complementary and integrative medicine in Alzheimer’s disease prevention. *J Alzheimers Dis*. 2015;48(1):13–14.

Aung SK, Fay H, Hobbs RF III. Traditional Chinese medicine as a basis for treating psychiatric disorders: a review of theory with illustrative cases. *Med Acupunct*. 2013;25(6):398–406.

Bonaccorso S, Ricciardi A, Zangani C, Chiappini S, Schifano F. Cannabidiol (CBD) use in psychiatric disorders: A systematic review. *Neurotoxicology*. 2019;74:282–298.

Borre YE, Moloney RD, Clarke G, Dinan TG, Cryan JF. The impact of microbiota on brain and behavior: mechanisms & therapeutic potential. *Adv Exp Med Biol*. 2014;817:373–403.

De Oliveira AM, Radanovic M, De Mello PC, et al. Nonpharmacological interventions to reduce behavioral and psychological symptoms of dementia: a systematic review. *Biomed Res Int*. 2015;2015:218980.

Dos Santos RG, Bouso JC, Alcázar-Córcoles MÁ, Hallak JEC. Efficacy, tolerability, and safety of serotonergic psychedelics for the management of mood, anxiety, and substance-use disorders: a systematic review of systematic reviews. *Expert Rev Clin Pharmacol*. 2018;11(9):889–902.

Fava M, Mischoulon D. Folate in depression: efficacy, safety, differences in formulations, and clinical issues. *J Clin Psychiatry*. 2009;70(Suppl 5):12–17.

Freeman MP, Fava M, Lake J, Trivedi MH, Wisner KL, Mischoulon D. Complementary and alternative medicine in major depressive disorder: the American psychiatric association task force report. *J Clin Psychiatry*. 2010;71(6):669–681.

Gao L, Wu C, Liao Y, Wang J. Antidepressants effects of Rhodiola capsule combined with sertraline for major depressive disorder: A randomized double-blind placebo-controlled clinical trial. *J Affect Disord*. 2020;265:99–103.

Guu TW, Mischoulon D, Sarris J, et al. International society for nutritional psychiatry research practice guidelines for omega-3 fatty Acids in the treatment of major depressive disorder. *Psychother Psychosom*. 2019;88(5):263–273.

Hammerschlag R, Marx BL, Aickin M. Nontouch biofield therapy: a systematic review of human randomized controlled trials reporting use of only nonphysical contact treatment. *J Altern Complement Med*. 2014;20(12):881–892.

Jezova D, Makatsori A, Smriga M, Morinaga Y, Duncko R. Subchronic treatment with amino acid mixture of L-lysine and L-arginine modifies neuroendocrine activation during psychosocial stress in subjects with high trait anxiety. *Nutr Neurosci*. 2005;8(3):155–160.

Jiao J, Li Q, Chu J, Zeng W, Yang M, Zhu S. Effect of n-3 PUFA supplementation on cognitive function throughout the life span from infancy to old age: a systematic review and meta-analysis of randomized controlled trials. *Am J Clin Nutr*. 2014;100(6):1422–1436.

Kasper S. Phytopharmaceutical treatment of anxiety, depression, and dementia in the elderly: evidence from randomized, controlled clinical trials. *Wien Med Wochenschr*. 2015;165(11–12):217–228.

Kim YK, Shin C. The microbiota-gut-brain axis in neuropsychiatric disorders: Pathophysiological mechanisms and novel treatments. *Curr Neuropharmacol*. 2018;16(5):559–573.

Lopresti AL, Drummond PD. Saffron (Crocus sativus) for depression: a systematic review of clinical studies and examination of underlying antidepressant mechanisms of action. *Hum Psychopharmacol*. 2014;29(6):517–527.

Lopresti AL. Curcumin for neuropsychiatric disorders: a review of in vitro, animal and human studies. *J Psychopharmacol*. 2017;31(3):287–302.

Marchand WR. Mindfulness-based stress reduction, mindfulness-based cognitive therapy, and Zen meditation for depression, anxiety, pain, and psychological distress. *J Psychiatr Pract*. 2012;18(4):233–252.

Marx W, Moseley G, Berk M, Jacka F. Nutritional psychiatry: the present state of the evidence. *Proc Nutr Soc*. 2017;76(4):427–436.

Muttoni S, Ardissino M, John C. Classical psychedelics for the treatment of depression and anxiety: A systematic review. *J Affect Disord*. 2019;258:11–24.

Opie RS, O'Neil A, Itsiopoulos C, Jacka FN. The impact of whole-of-diet interventions on depression and anxiety: a systematic review of randomised controlled trials. *Public Health Nutr*. 2015;18(11):2074–2093.

Perica MM, Delas I. Essential fatty acids and psychiatric disorders. *Nutr Clin Pract*. 2011;26(4):409–425.

Ravindran AV, Balneaves LG, Faulkner G, et al. Canadian network for mood and anxiety treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: Section 5. complementary and alternative medicine treatments. *Can J Psychiatry*. 2016;61(9):576–587.

Rucklidge JJ, Kaplan BJ. Broad-spectrum micronutrient formulas for the treatment of psychiatric symptoms: a systematic review. *Expert Rev Neurother*. 2013;13(1):49–73.

Sarris J, Ravindran A, Yatham LN, et al. Clinician guidelines for the treatment of psychiatric disorders with nutraceuticals and phytoceuticals: The world federation of societies of biological psychiatry (WFSBP) and canadian network for mood and anxiety treatments (CANMAT) taskforce. *World J Biol Psychiatry*. 2022;23(6):424–455.

Sarris J, Kavanagh DJ, Byrne G. Adjuvant use of nutritional and herbal medicines with antidepressants, mood stabilizers and benzodiazepines. *J Psychiatr Res*. 2010;44:32–41.

Sarris J, Marx W, Ashton MM, et al. Plant-based medicines (Phytoceuticals) in the treatment of psychiatric disorders: A Meta-review of Meta-analyses of randomized controlled trials: Les médicaments à base de plantes (phytoceutiques) dans le traitement des troubles psychiatriques: une méta-revue des méta-analyses d'essais randomisés contrôlés. *Can J Psychiatry*. 2021;66(10):849–862.

Sarris J. Herbal medicines in the treatment of psychiatric disorders: 10-year updated review. *Phytother Res*. 2018;32(7):1147–1162.

Sarris J, Murphy J, Mischoulon D, et al. Adjunctive nutraceuticals for depression: a systematic review and meta-analyses. *Am J Psychiatry*. 2016;173(6):575–587.

Schefft C, Kilarski LL, Bschor T, Köhler S. Efficacy of adding nutritional supplements in unipolar depression: A systematic review and meta-analysis. *Eur Neuropsychopharmacol*. 2017;27(11):1090–1109.

Yong SJ, Tong T, Chew J, Lim WL. Antidepressive mechanisms of probiotics and their therapeutic potential. *Front Neurosci*. 2020;13:1361.

Zagórska A, Marcinkowska M, Jamrozik M, Wiśniowska B, Paśko P. From probiotics to psychobiotics—the gut-brain axis in psychiatric disorders. *Benef Microbes*. 2020;11(8):717–732.

30.5 Micronutrient Monitoring and Supplementation in Psychiatric Illness

JESSICA L. PRINCIPE, M.D., M.S., AND NORMAN K. POLLOCK, PH.D.

There is growing interest in using micronutrient supplements as an adjunct to standard psychopharmacology or monotherapy for common psychiatric illnesses. Indeed, inverse relationships between micronutrient intake, nutrition biomarkers, and psychiatric illness are widely reported; however, these associations cannot be used to guide treatment decisions as the studies on which they are based do not support causal relationships. Placebo-controlled randomized controlled trials (RCTs) of nutrition interventions are often unavailable, vary in quality, and conflicting results can limit practice decisions. Meta-analyses of RCT, considered the gold standard for practice guidance, are often sparse; however, given the added cost for patients and potential for adverse effects, recommendations for nutrition supplementation should be based on the highest level of available evidence. Therefore, this chapter will focus on available evidence from published meta-analyses to supplement four nutrients of historical interest due to purported neuropsychiatric benefits: vitamin D, vitamin B12, folate (vitamin B9), and omega-3 fatty acids. In intervention trials, an effect size is calculated to quantify the magnitude of difference between two groups (and therefore the magnitude of benefit, if any) of treatment compared with placebo.

Effect sizes are expressed as Cohen *d* or standardized mean difference (*SMD*), whereas

SMD = **0.2** is a small effect size
SMD = **0.5** is a moderate effect size
SMD > **0.8** is a large effect size
SMD ≥ **1.3** is a very large effect size.

For reference, in depression, effect sizes for antidepressant medications are often small to moderate (*SMD* = 0.19 to 0.62), with potentially larger effect sizes seen with transcranial magnetic stimulation (TMS) (*SMD* = 1.33) and the largest with electroconvulsive therapy (ECT) (*SMD* = 2.83). Hopefully, keeping these in mind will allow the reader to put into context the magnitude of benefit (or lack thereof) reported for individual micronutrients discussed in this chapter.

For each nutrient, potential mechanisms, epidemiologic associations, and considerations regarding biomarker measurement will be addressed, while emphasis will be on evidence for the efficacy of supplementation, guided by biomarker monitoring where applicable, in common psychiatric, neurodevelopmental, and neurocognitive disorders. Please note that this chapter's scope does not include a review of the evidence for micronutrient supplementation in peripartum, trauma, obsessive-compulsive, or substance use disorders (SUDs). Furthermore, concentration is solely on benefit for individuals with psychiatric illness or clinically significant symptoms and *not on prevention* of illness or symptoms, as this is an entirely different topic.

VITAMIN D

Vitamin D is a fat-soluble steroid hormone, created endogenously with sufficient exposure to UV light and present in a few foods. Vitamin D is a unique nutrient because its requirement can be met endogenously from sunlight and dietary sources, complicating determining nutritional requirements. Serum 25-hydroxyvitamin D [25(OH)D] is considered the optimal biomarker of vitamin D status due to its relatively long half-life of 15 days and its reflection of endogenous production as well as vitamin D obtained from food and supplements. The most commonly used cut-points for serum 25(OH)D levels are: >75 nmol/L (>30 ng/mL) for vitamin D sufficiency, 50 to 70 nmol/L (20 to 29 ng/mL) for insufficiency, and <50 nmol/L (<20 ng/mL) for deficiency. Due to modern lifestyles and the importance of limiting sunlight exposure to reduce skin cancer risk, vitamin D deficiency and insufficiency are not uncommon. According to data from the National Health and Nutrition Examination Survey (NHANES) 2001–2010, one in three U.S. adults was considered vitamin D deficient, and an additional 40% met the criteria for insufficiency. Risk factors include dark pigmented skin, older age, obesity, physical inactivity, smoking, limited milk intake, limited sunlight exposure, fat malabsorption, primary hyperparathyroidism, chronic granulomatous disease, and medications (anticonvulsants, glucocorticoids, anti-HIV). Various mechanisms for a beneficial role of vitamin D in psychiatric illness have been suggested, including support for neurodevelopment, neurogenesis, neurotransmitter synthesis, immune modulation, anti-inflammatory actions, and HPA axis regulation. Inverse relationships between biochemical vitamin D status and depression, psychosis, attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and neurocognitive disorders have been reported; however, evidence for vitamin D supplementation as an adjunctive or stand-alone treatment option in most psychiatric disorders is limited.

Vitamin D and Depression

A meta-analysis of four RCTs ($N = 948$) in adults with unipolar depression showed that vitamin D supplementation improves depressive symptom scores (measured by Hamilton Depression Rating Scale or Beck Depression Inventory-II) with a moderate effect size ($SMD = 0.58$, $P < .05$). Two trials were selected for patients with depression and vitamin D deficiency at baseline, and in the two studies that did not select for vitamin D deficiency at baseline, average baseline vitamin D levels were in the insufficient range. However, subgroup analysis was not performed to determine whether vitamin D deficiency or insufficiency at baseline predicts response. Supplementation strategies varied from oral 1,500 IU cholecalciferol (vitamin D_3) daily, oral vitamin D 50,000 IU weekly, or one-time single-dose intramuscular injection of vitamin D 150,000 IU or 300,000 IU. It is relevant to consider that supplementation strategies of 50,000 IU cholecalciferol weekly, and 1,500 to 2,000 IU daily are supported by current Endocrine Society Guidelines for managing vitamin D deficiency in normal-weight adults (Table 30.5–1). Adjunctive antidepressant therapy was either excluded or not reported, except in one study, where 1,500 IU cholecalciferol daily was tested as adjunctive to fluoxetine therapy with positive results. Study durations ranged from 8 to 52 weeks, and no significant adverse events were reported. Overall, these results suggest that oral or parenteral vitamin D supplementation may be a clinically effective adjunctive or stand-alone intervention for improving depressive symptoms in adults with depression. However, it is important to note that vitamin D is a fat-soluble vitamin, excess supplementation poses the risk of toxicity, and many participants had vitamin D deficiency or insufficiency at baseline. Therefore, checking serum 25(OH)D levels in individuals with depression prior to initiating vitamin D supplementation and following evidence-based guidelines (Table 30.5–1) for repletion and maintenance dosing is the most responsible practice.

Vitamin D and Attention-Deficit/Hyperactivity Disorder

A recent meta-analysis of four RCTs ($N = 256$) in children and adolescents aged 2 to 18 years with ADHD demonstrated that vitamin D supplementation as adjunctive therapy to methylphenidate significantly improved ADHD symptoms. Small effect sizes were seen in ADHD total scores ($SMD = 0.39$), while medium effect sizes were seen in inattention ($SMD = 0.67$), hyperactivity ($SMD = 0.60$), and behavior ($SMD = 0.54$) (all $P < .01$). Supplement doses ranged from oral vitamin D 1,000 to 3,000 IU daily or 50,000 IU weekly and study durations ranged from 6 to 12 weeks. Three of the four RCTs in this analysis reported deficient average vitamin D levels at baseline; however, only one RCT performed subgroup analysis to determine whether vitamin D deficiency at baseline predicted improvement. In this study, subgroup analysis revealed that children with deficient or insufficient vitamin D levels at baseline improved with vitamin D supplementation while those with adequate baseline levels did not. Only one study reported adverse events, and no differences were found between vitamin D and placebo groups. Given the limited available evidence, vitamin D supplementation as adjunctive therapy to methylphenidate appears to improve inattention, hyperactivity, and behavior in children and adolescents with ADHD without serious adverse events associated with improved vitamin D status. As mentioned previously, due to the risk of toxicity with over-supplementation, it is most prudent to check vitamin D levels before supplementing per published guidelines in children or adolescents (Table 30.5–1) with ADHD and evidence of vitamin D deficiency or insufficiency.

VITAMIN B12

Vitamin B12 is a water-soluble vitamin naturally found in animal foods and necessary for normal neuronal function. Vitamin B12 status is measured in serum or plasma with 200 to 250 pg/mL as lower limits for most laboratories; however, some suggest that levels up to 400 pg/mL should be screened further for deficiency. In these borderline cases, methylmalonic acid (MMA) is recommended as a more sensitive measure of vitamin B12 status, with elevated MMA level (>0.271 µmol/L) as a reliable cutoff indicating deficiency. Estimates of Vitamin B12 deficiency in older adults range from 1.5% to 15%, and low vitamin B12 status has been linked to depression and cognitive dysfunction. Deficiency symptoms include depression, irritability, insomnia, cognitive slowing, and psychosis. Risk factors for deficiency include older age, vegan or vegetarian diet, pernicious anemia, malabsorption syndromes (gastric bypass surgery, inflammatory bowel disease, celiac disease, small intestinal bacterial overgrowth, ileal resection), and chronic use of metformin, proton-pump inhibitors, and histamine-2 receptor antagonists. Vitamin B12 deficiency may be increasing in younger adults. Data from the Framingham Offspring Study found similar deficiency rates in age groups 26 to 49 years, 50 to 64 years, and 65 years and older.

Unfortunately, few RCTs to date have tested isolated vitamin B12 supplementation as an intervention in patients with psychiatric illness. Vitamin B12 intervention trials in psychiatry almost exclusively employ cosupplementation with folate or vitamin B6. For this reason, it is not possible to make specific, evidence-based recommendations regarding vitamin B12 supplementation in psychiatric illnesses at this time. However, due to this compound’s water-soluble nature, excess quantities are not stored to a significant degree, and supplementation poses *little risk of toxicity*. Additionally, given that many, albeit nonspecific, neuropsychiatric symptoms can be associated with low B12, it is not unreasonable to check vitamin B12 levels in patients presenting with symptoms consistent with deficiency, as above, and consider supplementing per published guidelines if deficient or borderline levels are found. If a practitioner desires a more formal diagnosis of deficiency prior to supplementation, checking serum MMA is a reasonable confirmatory step to guide decision-making regarding the initiation of vitamin B12 supplementation. A 2018 *Cochrane* review by Wang and colleagues established that low-quality evidence exists to indicate that daily oral supplementation with 1,000 to 2,000 mcg of vitamin B12 is as efficacious, safe, and potentially more cost-effective than IM supplementation in vitamin B12 deficiency.

Table 30.5–1.
Endocrine Society Guidelines for Treatment of Vitamin D Deficiency

Patient Characteristics	Supplementation[a]	Duration[b]	Maintenance[c]
Children/adolescents ages 1–18 years	2,000 IU/day or 50,000 IU/week	≥6 weeks	600–1,000 IU/day
Adults ages >18 years	50,000 IU/week or 6,000 IU/day	8 weeks	1,500–2,000 IU/day or 50,000 IU q2 weeks
Obese adults (Also: malabsorption syndromes, medications affecting vitamin D metabolism[d])	6,000–10,000 IU/day	Monitor level and adjust dose to achieve 30 ng/mL	3,000–6,000 IU/day

[a] Supplementation with cholecalciferol (vitamin D3) or ergocalciferol (vitamin D2) is acceptable.

[b] Durations estimated to achieve sufficient level of ≥30 ng/mL (>75 nmol/L).

[c] Maintenance doses estimated to keep steady levels ≥30 ng/mL (>75 nmol/L) without inducing toxicity.

[d] Anticonvulsant medications, glucocorticoids, other medications that enhance the activation of the steroid xenobiotic receptor, which results in destruction of 25(OH)vitamin D.

Note: Vitamin D deficiency is defined here as ≤20 ng/mL (<50 nmol/L) per 2011 Endocrine Society Guidelines. While no formal guidelines exist for the treatment of vitamin D levels in the insufficient range (21–29 ng/ml; 50–70 nmol/L), it is reasonable to follow dosing guidelines for deficiency with shorter duration before initiating maintenance dosing, or alternatively, initiate maintenance dosing and re-check level in 3–6 months to ensure sufficiency ≥30 ng/mL (>75 nmol/L).

FOLATE (VITAMIN B9)

Folate is a water-soluble vitamin known for its role as an enzymatic cofactor involved in many essential neuropsychiatric processes, namely: neurotransmitter synthesis, central nervous system (CNS) phospholipid synthesis, and homocysteine metabolism. Folate deficiency (defined as serum folate <3 ng/mL) is rare in the United States, owing to the 1996 Food and Drug Administration mandate requiring folate supplementation of enriched grain products, completed in 1998. Indeed, a 2014 report of outpatient serum folate testing in 84,187 individuals in Boston, MA, found folate deficiency in only 0.06% and low to normal levels in only 0.2% of individuals tested. However, recent data from NHANES 1999 to 2016 reported that folate insufficiency rates are ~20% in the general population. This rate is significant considering folate insufficiency (serum folate approximately <7 ng/mL) in women of childbearing age is a risk factor for neural tube defects in offspring. Risk factors for folate deficiency include inadequate diet, excessive alcohol intake, women of childbearing age, non-Hispanic black women, pregnancy, dialysis, celiac disease, inflammatory bowel disease, and methylenetetrahydrofolate reductase (MTHFR) gene mutation. Symptoms of folate deficiency include cognitive impairment and depression, and low folate levels have been associated with major depressive disorder (MDD), dementia, and schizophrenia.

Folate and Depression

A meta-analysis of nine RCTs ($N = 1{,}061$) found folate supplementation as adjunctive to antidepressant therapy improved depressive symptoms in adults with MDD with a small effect size ($SMD = -0.38$, $P = .01$). Folate supplementation was also associated with a higher remission rate ($NNT = 8$, $RR = 1.46$, $P = .008$), with remission induced in 48% of patients in the folate group compared to 27% in patients receiving placebo. Supplement strategies included folic acid 0.5 to 10 mg/day and L-methylfolate 15 mg/day, with a mean study duration of 15.2 weeks. The mean age of the participants was 42.1 years, and 35.9% were male. Subgroup analysis revealed significant effects in trials lasting less than 15 weeks and in participants younger than 42.1 years, with a weakening of the effect size in predominantly female studies. Gender differences in effects of folate supplementation for depression are controversial with mixed results to date, with some studies suggesting a need for increased folate dose in men and limited effect of supplemental folate in older than middle-aged women, but these findings are not confirmed. Age effects on folate benefits in depression are also unclear, but it has been suggested that an age-dependent decrease in folate absorption plays a role. In this study, subgroup analysis was not performed to evaluate the effects of baseline folate level, folate dose, or formulation (i.e., folic acid vs. L-methylfolate, the most active form, which does not require MTHFR activity for metabolism, and which crosses the blood–brain barrier.) Rates of discontinuation due to any reason were not significantly different between groups. In summary, folate supplementation as an adjunct to antidepressant therapy appears effective for reducing depressive symptoms and inducing remission in adults with MDD. However, the role of monitoring folate levels to guide treatment decisions and optimal folate dose and formulation remains unclear. It is important to note that, in contrast to vitamin B12, the Food and Nutrition Board at the National Academies of Sciences, Engineering, and Medicine has set an upper limit for supplemental folate of 1 g/day for adults >19 years. This limit is due to concerns of exacerbating anemia and cognitive concerns associated with vitamin B12 deficiency, acceleration of preneoplastic lesions, and lower cognitive scores in children of mothers who consumed greater than 0.4 to 0.999 mg/day folic acid (400 to 999 mcg/day) during pregnancy. Additionally, with the use of L-methylfolate 15 mg/day, a few cases of triggered irritability and agitation have been reported (albeit one of these cases was in a patient with bipolar disorder type 1). For these reasons, and considering that no additional benefit of folate supplementation >15 weeks was found in the meta-analysis above, in select patients with inadequate response to standard treatment, it may be reasonable to trial folic acid doses up to 1 g/day or L-methylfolate 15 mg/day for no longer than 15 weeks, with close monitoring for irritability and agitation with L-methylfolate. Future studies may suggest additional biomarkers that will predict response to L-methylfolate supplementation.

Folate and Schizophrenia

A small meta-analysis of four RCTs (*N* = 210) in adults with schizophrenia (mean age of 38 years) demonstrated that folate supplementation as an adjunct to standard treatment was not superior to placebo in terms of total psychopathology, positive, negative, or general symptom scores. The mean study duration was 13 weeks, and 62% of participants were male. Doses ranged from 2 mg/day of folate to 15 mg/day of L-methylfolate. Rates of discontinuation due to any reason and adverse reactions were similar between folate and placebo groups. Subgroup analysis of the type of folate supplementation or baseline level was not performed; however, in one small RCT (*n* = 17), patients with schizophrenia and low or borderline folate levels who received 15 mg/day of methylfolate showed improved Clinical Rating Scale scores after 6 months (*P* < .01). A second RCT utilizing 15 mg/day of L-methylfolate (*n* = 55) for 12 weeks showed improvement in PANSS total, negative, and general psychopathology scores with moderate effect sizes (Cohen *d* = 0.63, 0.68, and 0.84, respectively) along with beneficial changes in magnetic resonance imaging measures (working memory-related activation, resting connectivity, and cortical thickness) (all *P* < .05).

Interestingly in this study, total and general PANSS scores appeared to depend on genotype, specifically a variant of *MTR,* notably the same variant that influenced response to L-methylfolate in depression, while improvement in PANSS negative scores did not. These individual RCT results suggest a differential effect of L-methylfolate and the possible importance of baseline level in this population, but further evidence is needed. In patients with schizophrenia who are compliant with standard treatment, who have low or borderline folate levels, or for whom nutrient supplementation is appealing, L-methylfolate at 15 mg/day may be a safe and effective intervention for reducing symptom burden with the caveat that evidence is limited, and standard therapy should be prioritized.

OMEGA-3 FATTY ACIDS

Omega-3 fatty acids are considered essential fatty acids, which means they are not produced in adequate amounts endogenously and must be consumed. To date, there is insufficient evidence on which to base a recommended daily allowance (RDA) or estimated average intake (EAR), so an adequate intake (AI) recommendation is utilized instead. Current AIs for total omega-3 fatty acid intake range from 0.7 to 0.9 g/day in ages 1 to 8, 1.0 to 1.6 g/day in ages 9 to 19, and 1.1 to 1.6 g/day in adulthood. Various biomarkers may be utilized in a research setting to determine biochemical status; however, none are used clinically at this time, and no cutoffs have been established to indicate deficiency. The two most prominent omega-3 fatty acids in brain tissue are docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA), with DHA accounting for 40% of total brain membrane phospholipids. DHA is thought to play important roles in neurologic development, maintenance of membrane integrity and fluidity, receptor binding, neurotransmission, and signal transduction, while EPA is thought to have potent anti-inflammatory and antioxidant effects. Omega-3 fatty acids are perhaps the most widely studied nutrients in neuropsychiatry to date. Inverse associations have been found between omega-3 intake or biochemical status and depression, anxiety, ADHD, ASD, bipolar disorder, schizophrenia, mild cognitive impairment, and Alzheimer dementia.

Omega-3 Fatty Acids and Depression

A recent meta-analysis of 26 RCTs (*N* = 2,160) demonstrated that supplementation with omega-3 fatty acids resulted in a small but statistically significant improvement in depressive symptom scores in adults with MDD or depressive symptoms. Depression symptom rating instruments included: Hamilton Depression Rating Scale (16 studies), Montgomery-Asberg Depression Rating Scale (4 studies), Beck Depression Inventory (4 studies), and Geriatric Depression Scale (2 studies). The supplemented doses of EPA ranged from 180 to 4,000 mg/day, whereas DHA ranged from 120 to 2,000 mg/day. Trials ranged in duration from 4 weeks to 6 months. The effect size for all types of omega-3 supplementation was small (*SMD* = −0.28, *P* = .004); however, subgroup analysis revealed increased therapeutic effects based on EPA formulation. EPA-pure formulations (containing 100% EPA) and EPA-major formulations (containing ≥60% EPA), each delivering a total EPA dose of up to 1 g/day of EPA, demonstrated significantly greater benefits when compared with other omega-3 formulations. The greatest effect size was seen in EPA-major formulations (containing ≥60% EPA, delivering total EPA up to 1 g/day) with a large effect (*SMD* = −1.03, *P* = .03). The effect size for EPA-pure supplements (containing 100% EPA, delivering up to 1 g/day of EPA) was moderate (*SMD* = −0.50, *P* = .003). Studies included treatment with antidepressants (10 studies), without antidepressants (8 studies), and or a mixed sample (8 studies). Subgroup analysis was not performed to determine differences in mono- versus adjunctive therapy. Studies included patients with medical comorbidities (8 studies), without medical comorbidities (11 studies), and mixed samples (7 studies). This meta-analysis did not report adverse events or time to discontinuation in treatment versus placebo groups. Based on available data, the use of daily omega-3 fatty acid supplementation containing

≥60% EPA, delivering up to 1 g/day of EPA, appears to be an effective intervention in adults with MDD or clinically significant depressive symptoms.

Omega-3 Fatty Acids and Anxiety

In a recent meta-analysis of 19 RCTs (*N* = 2,240) in adults with clinically significant anxiety symptoms, supplementation with omega-3 fatty acids resulted in a small but significant improvement in anxiety symptoms (*SMD* = 0.37, *P* = .01). Reflecting the range of anxiety disorders and symptoms studied, a large number of rating instruments were used to determine clinically significant anxiety symptoms as follows: Yale-Brown Obsessive-Compulsive Scale, Profile of Mood States, State-Trait Anxiety Inventory, Hamilton Anxiety Rating Scale, Generalized Anxiety Disorder Questionnaire, Depression, Anxiety, and Stress Scales, Clinician-Administered Post-traumatic Stress Disorder Scale, Beck Anxiety Inventory, Visual Analog Scale of Anxiety, Impact of Event Scale-Revised, Conners Score Anxiety Subscale, Neuropsychiatric Inventory, Test Anxiety Severity, Hospital Anxiety, and Depression Scale Anxiety Subscale, and Child Behavior Checklist anxiety subscale. Psychiatric and medical conditions in participants also varied widely, including no specific clinical conditions, acute myocardial infarction, borderline personality disorder, mild-to-severe depression, obsessive-compulsive disorder, severe accidental injury, traumatized by disaster, substance abuse disorder, premenstrual syndrome, Alzheimer disease, Parkinson disease, generally healthy undergraduate college students with test anxiety, children with ADHD, and individuals with Tourette syndrome. The total omega-3 fatty acid supplementation doses ranged from 225 to 4,074 mg/day, and the treatment duration ranged from 3 to 24 weeks. In subgroup analysis, patients with a clinical diagnosis of an anxiety disorder experienced a moderate benefit of omega-3 supplementation compared to those without a clinical anxiety diagnosis (*SMD* = 0.51, *P* = .01). Further subgroup analysis demonstrated that omega-3 fatty acid doses ≥2 g/day resulted in a small but significantly increased benefit compared to doses <2 g/day (*SMD* = 0.21, *P* = .02). Subgroup analysis of trials utilizing EPA supplements containing less than 60% versus 60% or more were opposite to the depression literature, with supplements containing less than 60% EPA conferring benefit versus placebo, whereas supplements containing 60% EPA or more did not. Adverse effects were not reported, and subgroup analysis was not performed to determine differences in monotherapy versus adjunctive therapy. Modest sample sizes and significant heterogeneity in the included studies should be considered; however, this study lends support to the idea that omega-3 fatty acids, particularly in doses >2 g/day and perhaps with <60% EPA, may be an effective intervention to reduce anxiety symptoms in individuals with clinical anxiety disorders. Further study in more narrowly defined illness groups would be helpful to clarify the optimal dose.

Omega-3 Fatty Acids and ADHD

In a meta-analysis of eight RCTs ($N = 628$) in children with ADHD aged 7 to 14 years, omega-3 fatty acid supplementation demonstrated small but significant improvement in parental reports of inattention ($SMD = 0.42$), hyperactivity ($SMD = 0.48$), and total ADHD symptom scores ($SMD = 0.38$) (all $P < .05$). Subgroup analysis revealed that inattention and total ADHD symptom scores improved with EPA doses <500 mg/day and ≥500 mg/day, with small effect sizes (*SMD*) ranging from 0.36 to 0.48, while EPA doses ≥500 mg/day showed a large effect on hyperactivity scores ($SMD = 0.81$, $P = .02$). Mixed (EPA and DHA) omega-3 doses used in the trials ranged from 80 to 650 mg/day of EPA and from 3 to 640 mg/day of DHA. Trials ranged in duration from 12 weeks to 6 months. Data on adverse effects and concurrent medication use were not reported. Omega-3 fatty acid supplementation in modest dosages appears to improve ADHD symptoms in children with greater benefit in hyperactivity seen with mixed supplements containing EPA ≥500 mg/day.

Omega-3 Fatty Acids and ASD

In a small meta-analysis of five RCTs ($N = 183$) in children with ASD aged 2 to 17 years, there were no differences between the omega-3 fatty acid and placebo groups in Aberrant Behavior Checklist (ABC) scores of irritability, stereotypy, hyperactivity, or inappropriate speech assessed by teachers/clinicians or parents. However, omega-3 supplementation demonstrated significant improvement in parents' ratings of ABC lethargy ($SMD = 1.98$, $P = .02$), but not when assessed by teachers/clinicians ($P = .58$). In studies using the Behavioral Assessment System for Children scoring tool, parent ratings reflected significant externalizing behavior and social skills worsening. One study used the Vineland Adaptive Behavior Scale and reported significant improvement in the daily-living component with omega-3 supplementation. Omega-3 doses used in the trials ranged from 375 to 840 mg/day of EPA and from 200 to 1,125 mg/day of DHA. Trials ranged in duration from 6 weeks to 6 months. No significant differences in adverse effects were observed between the omega-3 and placebo groups. Concurrent medication administration use was not reported in the trials, but this should be understood with the context that the mainstay of treatment for symptoms in ASD is nonpharmacologic. Given the paucity of data and mixed results, no recommendations can be made regarding the benefit of omega-3 supplementation for improving symptoms in ASD at this time.

Omega-3 Fatty Acids and Mild Cognitive Impairment

In a recent meta-analysis of 15 RCTs (*N* = 2,524) in adults with MCI, supplementation with mixed omega-3 fatty acids (EPA + DHA) revealed small but statistically significant improvements in memory (*SMD* = 0.31) and executive function (*SMD* = 0.19) (both $P < .01$). Trials ranged in duration ranged from 2 months to 26 weeks. Mixed omega-3 doses ranged from 60 to 2,187 mg/day of EPA and 180 to 1,550 mg/day of DHA. Subgroup analysis showed that the effect size for improvement in either domain was not related to the dose of EPA or DHA. Data on adverse effects were not reported. The results of this review suggest that mixed EPA + DHA omega-3 supplementation may confer benefit in the cognitive domains of memory and executive function in adults with MCI, albeit with small effect sizes, and further study is needed to outline more specific dose ranges.

Omega-3 Fatty Acids and Alzheimer Disease

A recent meta-analysis of six RCTs (*N* = 758) in patients with Alzheimer disease demonstrated that omega-3 fatty acid supplementation did not improve cognitive deterioration, cognitive dysfunction, or neuropsychiatric symptoms measured by the Mini-Mental State Examination, Alzheimer Disease Assessment Scale-Cognitive subscale, and the Neuropsychiatric Inventory, respectively. Variability in DHA or DHA/EPA doses ranged from 240 mg/day to 2.3 g/day, and treatment periods ranged from 90 days to 1.5 years. The authors suggested that variability in DHA or DHA/EPA dose, the ratio of DHA/EPA, concurrent medications, and lack of information on participant diets may have limited conclusions. Adverse effects were reported in two of the studies, and neither found significant group differences in the incidence of events. The current evidence does not suggest that omega-3 fatty acid supplementation improves cognition or neuropsychiatric symptoms in adults with Alzheimer disease.

SUMMARY AND FUTURE DIRECTIONS

Given that many patients will choose to utilize dietary supplements, the psychiatrist needs to offer evidence-based recommendations when feasible. Additionally, given the often inadequate responses to standard treatment approaches, it is worthwhile for mental health providers to be aware of nutrition-related evidence-based interventions to augment standard therapy or offer alternative approaches when appropriate. Considering that the first vitamin was isolated just under 100 years ago, it is heartening to see well-designed studies emerging to help elucidate relationships between specific micronutrients and mental illnesses, particularly since trials of this kind did not emerge to a large degree until the 1990s. More well-designed RCTs are still needed to determine the role of supplemented micronutrients outlined in this chapter (vitamin D, vitamin B12, folate, and omega-3 fatty acids) in many psychiatric disorders.

Our review suggests a benefit for

- Vitamin D supplementation (depression, childhood ADHD)
- Folate (depression)
- Omega-3 fatty acids (depression, anxiety, ADHD, and mild cognitive impairment)

Supplementing with nutrients should be undertaken with care, guided by biomarker monitoring where applicable, and dosed appropriately (in amount and duration) for specific conditions based on the available evidence. As interest in this area appears to be ongoing, providers should hopefully expect additional data to emerge to guide practice recommendations for using these and other nutrients in psychiatric illness in the future.

FURTHER READINGS

Alex A, Abbott KA, McEvoy M, Schofield PW, Garg ML. Long-chain omega-3 polyunsaturated fatty acids and cognitive decline in non-demented adults: a systematic review and meta-analysis. *Nutr Rev*. 2020;78(7):563–578.

Araya-Quintanilla F, Gutiérrez-Espinoza H, Sánchez-Montoya U, et al. Effectiveness of omega-3 fatty acid supplementation in patients with Alzheimer disease: a systematic review and meta-analysis. *Neurologia (Barcelona, Spain)*. 2020;35(2):105–114.

Chang JP-C, Su K-P, Mondelli V, Pariante CM. Omega-3 polyunsaturated fatty acids in youths with attention deficit hyperactivity disorder: a systematic review and meta-analysis of clinical trials and biological studies. *Neuropsychopharmacology*. 2018;43(3):534–545.

Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet (London, England)*. 2018;391(10128):1357–1366.

David N, Osser MD. Sorting Out Comorbidities. 2020. https://www.psychiatrictimes.com/view/l-methylfolate-augmenting-agent-may-contribute-agitation-and-mania

de Meer K, Smulders YM, Dainty JR, et al. [6S]5-methyltetrahydrofolate or folic acid supplementation and absorption and initial elimination of folate in young and middle-aged adults. *Eur J Clin Nutr*. 2005;59(12):1409–1416.

Gan J, Galer P, Ma D, Chen C, Xiong T. The effect of vitamin D supplementation on attention-deficit/hyperactivity disorder: a systematic review and meta-analysis of randomized controlled trials. *J Child Adolesc Psychopharmacol*. 2019;29(9):670–687.

Godfrey PS, Toone BK, Bottiglien T, et al. Enhancement of recovery from psychiatric illness by methylfolate. *Lancet North Am Ed*. 1990;336(8712):392–395.

Holick MF, Binkley NC, Bischoff-Ferrari HA, et al. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2011;96(7):1911–1930.

Horvath A, Łukasik J, Szajewska H. ω-3 fatty acid supplementation does not affect autism spectrum disorder in children: a systematic review and meta-analysis. *J Nutr*. 2017;147(3):367–376.

Institute of Medicine (US) Committee to Review Dietary Reference Intakes for Vitamin D and Calcium. *Dietary Reference Intakes for Calcium and Vitamin D*. National Academies Press; 2011.

Kantor ED, Rehm CD, Du M, White E, Giovannucci EL. Trends in dietary supplement use among US adults from 1999–2012. *JAMA*. 2016;316(14):1464–1474.

Liao Y, Xie B, Zhang H, et al. Efficacy of omega-3 PUFAs in depression: a meta-analysis. *Translational Psychiatry*. 2019;9:190.

Lindenbaum J, Rosenberg IH, Wilson PW, Stabler SP, Allen RH. Prevalence of cobalamin deficiency in the Framingham elderly population. *Am J Clin Nutr*. 1994;60(1):2–11.

L-Methylfolate: Augmenting Agent May Contribute to Agitation and Mania. Psychiatric Times. Accessed December 19, 2021. https://www.psychiatrictimes.com/view/l-methylfolate-augmenting-agent-may-contribute-agitation-and-mania.

Micallef-Trigona B. Comparing the effects of repetitive transcranial magnetic stimulation and electroconvulsive therapy in the treatment of depression: a systematic review and meta-analysis. *Depression Research and Treatment*. 2014;2014:135049.

Office of Dietary Supplements—Folate. Accessed December 19, 2021. https://ods.od.nih.gov/factsheets/Folate-HealthProfessional/

Office of Dietary Supplements—Omega-3 Fatty Acids. Accessed December 15, 2021. https://ods.od.nih.gov/factsheets/Omega3FattyAcids-HealthProfessional/

Office of Dietary Supplements—Vitamin B12. Accessed December 13, 2021. https://ods.od.nih.gov/factsheets/VitaminB12-HealthProfessional/

Papakostas GI, Shelton RC, Zajecka JM, et al. Effect of adjunctive L-methylfolate 15 mg among inadequate responders to SSRIs in depressed patients who were stratified by biomarker levels and genotype: results from a randomized clinical trial. *J Clin Psychiatry*. 2014;75(8):855–863.

Pfeiffer CM, Sternberg MR, Zhang M, et al. Folate status in the US population 20 y after the introduction of folic acid fortification. *Am J Clin Nutr*. 2019;110(5):1088–1097.

Roffman JL, Petruzzi LJ, Tanner AS, et al. Biochemical, physiologic, and clinical effects of L-methylfolate in schizophrenia: a randomized controlled trial. *Mol Psychiatry.* 2018;23(2):316–322.

Su K-P, Tseng P-T, Lin P-Y, et al. Association of use of Omega-3 polyunsaturated fatty acids with changes in severity of anxiety symptoms. *JAMA Network Open*. 2018;1(5):e182327.

Theisen-Toupal J, Horowitz G, Breu A. Low yield of outpatient serum folate testing: eleven years of experience. *JAMA Intern Med*. 2014;174(10):1696–1697.

Tucker KL, Rich S, Rosenberg I, et al. Plasma vitamin B-12 concentrations relate to intake source in the Framingham Offspring study. *Am J Clin Nutr.* 2000;71(2):514–522.

Vellekkatt F, Menon V. Efficacy of vitamin D supplementation in major depression: a meta-analysis of randomized controlled trials. *J Postgrad Med*. 2019;65(2):74–80.

Wang H, Li L, Qin LL, et al. Oral vitamin B12 versus intramuscular vitamin B12 for vitamin B12 deficiency. *Cochrane Database Syst Rev*. 2018;3:CD004655.

Zheng W, Li W, Qi H, et al. Adjunctive folate for major mental disorders: a systematic review. *J Affect Disord*. 2020;267:123–130.

30.6 Disaster Psychiatry: Disasters, Terrorism, and War

David M. Benedek, M.D., Joshua C. Morganstein, M.D., James C. West, M.D., and Robert J. Ursano, M.D.

Disasters impact hundreds of communities and affect the lives of millions of people around the world every year. They may result from the destructive forces of nature and climate-related events (e.g., hurricanes, earthquakes, tsunamis, or infectious disease outbreaks, such as COVID-19), accidents (e.g., plane crash, industrial fire), negligence (e.g., forest fires sparked by carelessness), or deliberate acts (e.g., large-scale terrorist attacks or military combat operations). Regardless of origin, the frequent chaos of disaster can be pervasive and enduring over time. Psychological injury often occurs in conjunction with physical injury, exposure to the injury and death of others, loss of social connections, disruption of the physical environment, and accompanying feelings of terror and helplessness. Many disasters have both climate and environmental causes that are often exacerbated by human systems and behaviors. Hurricane Katrina began as a storm that exposed weaknesses in levees; The triple-disaster of Fukushima began as an earthquake that created a tsunami that damaged vulnerable nuclear reactors; COVID-19 appears to have been initially transferred by animals, but then rapid, mass global travel quickly transmitted the virus around the world.

Disaster psychiatry is an area of psychiatry that includes public health planning, clinical care, and research. Core to this area is anticipating the scope of behavioral, emotional, and distress-related responses of individuals and communities, planning mental health disaster response, and, ultimately, providing care for those exposed. War and terrorism are subtypes of human-caused disasters and typically cause more significant adverse psychological impacts than natural and climate-related disasters. During war, military personnel may die or be injured in combat, but noncombatant civilians are also intentionally or unintentionally injured and killed during military operations. As a result, war is a particularly unique form of disaster, exposing various groups and populations to distressing and grotesque imagery and physical harm, creating risk for adverse psychological and behavioral outcomes.

Because war, military peacekeeping, and humanitarian relief operations place military health care providers in direct and prolonged contact with disaster-affected populations, the knowledge base, skills, and professional attitudes required for psychiatric practice in military operations are closely aligned with those necessary for other types of disasters. A psychological injury must be addressed promptly in any disaster, including war and terrorism. Mental health care must be provided in hospitals or clinics and near sites of chaos and destruction, where traditional medical resources may be scarce or overwhelmed. Care providers use public health models of health surveillance, outreach, and education to overcome barriers to care, including those safety and survival interests coexisting with mental health needs, limited physical and psychiatric care resources, and stigma associated with requesting or receiving mental health assistance. Military and disaster psychiatry disciplines address these care demands in nontraditional environments and mass casualty situations in which resources are overwhelmed by emphasizing disease prevention through a range of interventions by individuals, teams, and leaders that foster resilience.

Disaster psychiatry highlights the importance of integrating mental health and public health. Care in these environments relies on not only psychiatrists but also other physicians, social scientists, epidemiologists, nurses, and emergency responders such as police and firefighters. The coordination of these diverse care providers is necessary to ensure that needs are met and that no subgroup of victims is overlooked. This approach to population health care delivery and management, including primary (prevention), secondary (early treatment), and tertiary (treating established disease and rehabilitation) care, is a critical aspect of disaster psychiatry.

In the United States, the terrorist attacks on the Pentagon and the World Trade Center and the distribution of anthrax-tainted mail in the following weeks resulted in widespread fear and uncertainty in the American public. Subsequently, around the globe, the issues of severe acute respiratory syndrome (SARS), Sarin gas attacks in Tokyo, train bombings in Spain, and the earthquake in Fukushima drew attention to disaster care for individuals and groups across explosive, radiologic, and biologic agents. Most recently, nations worldwide have grappled with the global COVID-19 pandemic and resultant economic crises, political upheaval, and social unrest. These events also resulted in neuropsychiatric syndromes in many survivors and their relatives. Response to these events involved numerous government, public health, law enforcement agencies, civilian disaster relief organizations, military medical, mental health, and law enforcement personnel. The response to the COVID-19–related crises in the United States, as well as military participation in humanitarian relief efforts after the Southeast Asian tsunami of December 2004, Hurricanes Katrina and Rita of 2005, Haiti earthquakes of 2010, the Philippines’ Typhoon Haiyan of 2013, West Africa Ebola outbreak of 2014, and COVID-19 pandemic, illustrate the complex response required for disasters of all types.

ENVIRONMENTS OF CARE IN DISASTER AND MILITARY OPERATIONS

The term *disaster* has several definitions or connotations. The word is derived from the Latin words *dis* (meaning *against*) and *astrum* (meaning *stars*)—*the stars are against us.* Ancient societies may have conceived of disasters, such as floods or earthquakes, as acts of the gods. However, the distinction between *natural* disasters, such as earthquakes, and *human-caused* or technologic disasters, such as explosions or plane crashes, is increasingly challenging to make. Much of the death and destruction from an earthquake may result from damage to poorly constructed housing in urban areas—thus, there is often a human-caused aspect even to natural disasters. Certainly, human-caused disasters can result from accidents (e.g., factory explosions), but human-caused disasters also result from intentional acts in the cases of war or terrorism. The degree to which a disaster may be anticipated and the characteristics of the affected population (soldiers or civilians) may be more critical distinctions than whether the disaster was natural or human-caused.

War is an all-too-common human-caused disaster. At the time of writing, there are more than 57 ongoing armed conflicts involving 45 countries occurring around the globe. War has been defined as a political act (generally involving violence) to achieve national objectives or protect natural interests. In these actions, militaries create their own and civilian casualties in armed conflicts of all kinds. During recent years, militaries worldwide have also been prominently engaged in humanitarian disaster relief, as for hurricanes and pandemic response in the United States, earthquakes in Turkey, and the recent tsunami in Southeast Asia, to name only a few. Indeed, military forces typically represent the most significant portion of disaster-response personnel resources for most countries. In addition, militaries are involved in government stabilization and peacekeeping efforts (military operations other than war). Terrorism, like war, involves violence to achieve political objectives and may occur between members of different countries or spawn from internal, domestic threats. Here, however, the objective is to create fear or terror to undermine society’s confidence in its government and leaders. As such, fear, anxiety, and a sense of chaos are the goals of terrorism rather than the (albeit predictable) aftereffects observed in wars and other human-caused or natural disasters. Therefore, mitigating the emotional and psychological effects of terrorism is crucial to sustaining and protecting a nation from the impact of terrorist attacks.

The causes of emotional symptoms in response to disasters, including war, have been historically attributed to sources ranging from the gods, the concussive force of a passing cannonball, and various natural or supernatural sources of contagion. Emotional consequences are alluded to frequently in the descriptions of the terror induced by the hero's attack in Homer's *Iliad.* Ancient Greeks attributed epidemic illness to Apollo's wrath after the desecration of his temple. The French military surgeon Dominique-Jean Larrey commented clearly on the ill effects of war on the health of Napoleon's soldiers. Combat-related behavior, similar to the symptoms of posttraumatic stress disorder (PTSD), defined in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), is described in Civil War medical treatises. The science of neurology entered military medicine with Weir Mitchell's work in this era. During the remainder of the 19th and 20th centuries, concepts of neurasthenia, dissociation, hysteria, and psychological suggestion were developed to define psychoneurologic states for which no demonstrable neuroanatomical pathology could be identified. Military physicians in the Russian–Japanese War made similar observations. Recognition that nontraumatic injuries followed train derailments occurred even as World War I and II physicians initiated treatment models for conditions such as *shell shock, gas neurosis,* and *conversion paralysis.* Civilian physicians have also long recognized the concept that war results in physical and emotional human suffering. In 1859, Henri Dumont arranged for civilian medical services for the injured after observing soldiers die from lack of medical attention during the Battle of Solferino. His efforts led to the establishment of the International Red Cross and international guidelines for the humane care of the sick and wounded during wartime.

Beginning in the last quarter of the 20th century, the Red Cross, and other international medical and relief agencies, such as Doctors without Borders and Partners in Health, have increasingly provided mental health-related consultation and direct care for victims and responders in the aftermath of war and natural and human-caused disasters. Observations during these military and civilian medical efforts formed the basis for today's understanding that aspects of the environment and the unwanted environmental changes brought about by war or disaster result in emotional trauma.

Regardless of the source, potential stressors in disaster and military environments include exposure to the dead and grotesque, immediate and pending threat to life, loss of loved ones, loss of personal property, and physical injury with associated disability and pain. Such stressors may occur transiently but must be endured for months or years more often. Although the military brings supplies and logistical infrastructure, including food, shelter, and specialized equipment and clothing to protect soldiers at war, civilians exposed to combat in modern times may be subject to large-scale destruction within their communities. They may become refugees and experience shortages of food or shelter that threaten life. These noncombatant victims often do not receive treatment for psychiatric symptoms that may emerge because of exposure to bombings, battles, rape, torture, and genocide. Although an earthquake may last only seconds or minutes, the traumatic experiences resulting from the destruction of home, geographic displacement from familiar surroundings, a loss of employment, or because of injury or loss of loved ones may continue for weeks, months, or, possibly, years. Combat environments, ongoing terrorist campaigns, and cases of natural disasters, such as earthquake exposure, over time may include vaguely anticipated or entirely unexpected life-threatening experiences, followed by daily life in an austere or disrupted environment.

TRAUMATIC STRESS RESPONSE IN POPULATIONS AND INDIVIDUALS

Psychological and behavioral responses after a disaster change over time. Understanding the phases of psychological response to disaster can inform planning and caring for the mental health needs of individuals and communities. A *predisaster phase* occurs when an event is anticipated or an advanced warning is given, such as a tropical cyclone or wildfire. Feelings of vulnerability and worry about safety are common in this phase and, for some, will promote protective behaviors such as evacuation. For others, these feelings are ignored, not present, or may lead to paralysis with indecision or feeling emotionally overwhelmed.

In the immediate aftermath of an event, during the *impact phase*, survivors experience disbelief, numbness, fear, and confusion as they witness devastation and suffering. These are the “normal,” that is, expected, emotional responses to an extraordinary event. A *heroic phase*, lasting days to weeks, typically follows. Rescue and survival activities are the central focus. Disorientation may persist for some and render them unable to participate in response and recovery. Injury or separation from family can produce increased anxiety and worry and detract from immediate problem-solving. The *heroic phase* includes the initial appearance of outside assistance and marks the beginning of adaptation to the postdisaster environment.

The *honeymoon phase* follows as outside assistance becomes fully available and communities band together in response to shared adversity, giving and receiving assistance. Survivors are hopeful that the help they receive will make them whole again. During this phase, disaster behavioral health workers are often perceived as helpful and readily accepted by community members, developing a foundation to assist in the future.

As the challenges of prolonged recovery set in, optimism gives way to *disillusionment*. Assistance agencies and volunteer groups begin to pull out of the community, and resources for recovery are stretched thin, leading to feelings of disappointment and resentment as hopes for a restoration of the predisaster environment are not met. The sense of community, so strong in the *honeymoon phase*, diminishes as individuals focus on personal needs or express frustration over unmet needs. Ill will and resentment may surface as survivors receive unequal monetary compensation for what they perceive to be equal or similar damage. Survivors may be physically and emotionally exhausted in the face of enormous multiple demands: financial pressures, relocation or living in a damaged home, family discord, bureaucratic challenges, and a reduced time for recreation or self-care. New health problems and exacerbation of preexisting conditions emerge because of ongoing stress and fatigue. *Anniversary reactions* should be anticipated and represent an opportunity for leaders to support the psychological well-being of disaster victims through memorializing, helping make meaning of the events, and focusing on “building back better.” Failure to effectively address a disaster anniversary experience can demoralize survivors and exacerbate underlying psychological distress and feelings of disillusionment.

As needs are met, and communities stabilize, the *reconstruction phase* emerges and may last for years. Survivors face deciding whether to return to old jobs or find new work, rebuild homes or relocate, reestablish social ties and emotional support systems, or develop new ones. Individuals may find meaning and experience posttraumatic growth, emerging from the disaster with an increased sense of personal strength.

Individuals and communities progress through the phases of disaster at variable rates. Planning and delivering care to disaster victims must recognize this. Depending on the intensity and duration of trauma exposure, resources available, and individual coping skills, individuals may develop persistent symptoms requiring prolonged treatment. Failure to account for these factors in medical and psychological response plans can lead to anger at caregivers and community leaders.

Symptoms versus Functioning

Disaster psychiatry addresses the clinical concerns of identified patients and supports interventions to prevent potentially incapacitating morbidity in entire populations. Distress-related symptoms are universally experienced during an evolving disaster, as they are during combat. Symptoms occurring in persons who are not functionally impaired are, by necessity, a secondary concern when resources are overwhelmed.

The label given to clusters of symptoms has political, economic, and research-related significance. However, in the context of ongoing battle or relief efforts, the perception that one is ill can lead to impaired functioning, which may compromise these efforts and increase long-term disability. Indeed, transient and nearly universal symptoms that represent normal responses can become medicalized if health care personnel reinforce a view that these symptoms constitute a disease. Therefore, initial mental health interventions must focus on mobilizing effective functioning to allow rescue and relief efforts or battle in the context of war-related disasters to continue.

Psychological and Behavioral Syndromes in the Aftermath of Disasters, Terrorism, and War

A considerable body of literature exists on the effects of traumatic events on individuals and communities in response to disasters and war. Each of these informs the other. Many individuals emerge with limited or no adverse effects, effectively resuming their social and occupational roles (resilience). Some may even experience an increased sense of competence, self-efficacy, and belief in their ability to manage future stressors (sometimes termed “posttraumatic growth”). However, a sizable minority experiences a range of adverse psychological and behavioral effects, including distress reactions, health-risk behaviors, and psychiatric disorders (Fig. 30.6–1).

In the immediate aftermath of a disaster, distress reactions predominate. Individuals feel a sense of vulnerability and often blame, scapegoat, and express anger at organizations and leaders perceived as responsible. Demoralization and a loss of faith may also occur. Many individuals experience insomnia, irritability, and feelings of distractibility. Some present to health care settings with physical symptoms as a manifestation of psychological distress. Symptoms such as headache, dizziness, nausea, fatigue, and weakness are common in the wake of a disaster, even when an identifiable physical disorder cannot be found. In addition, alterations in health-related behavior are not uncommon after traumatic events. Increased use of alcohol, tobacco, and illicit or prescription drugs may occur as individuals seek to manage feelings of distress.

Disasters are time-consuming and lead to work-life imbalance leading to erratic sleep, deterioration in eating habits, and reduced exercise. Family conflict and violence may also occur. These are common reactions to extraordinary events. Planning for these distress reactions and risky health behaviors requires adequate resources to respond to individuals with distress symptoms in a timely and supportive manner and triage at emergency and other health care settings to enable the management of other physical and mental disorders.

Much of disaster psychiatry focuses on the neurobehavioral responses to stressful, emotional events. Well-defined psychiatric illnesses certainly result from such exposures. The DSM-5 definition of PTSD necessitates direct or indirect exposure to death, threats to the physical integrity of self or others, or aversive details about an event. PTSD includes (1) intrusion symptoms, such as involuntary memories, traumatic nightmares, or dissociative reactions; (2) avoidance manifested by persistent efforts to avoid trauma-related thoughts, feelings, or external reminders such as people, places, or activities; (3) negative alternations in cognition and mood, including the inability to recall key features of the traumatic event, persistent negative beliefs about oneself and the world, distorted beliefs, diminished interest in activities and constricted affect; and (4) alterations in arousal and reactivity demonstrated by irritability or aggression, self-destructive behavior, hypervigilance, concentration problems, and sleep disturbance. When such symptoms result in functional impairment and persist for more than 1 month, PTSD is diagnosed. Transient symptoms occurring during the first month and resulting in dysfunction are diagnosed as *Trauma and Stressor-Related Disorder Not Otherwise Specified or Adjustment Disorder.* War, disaster, and terrorist attacks often result in exposure to stressors meeting the criterion for PTSD. PTSD is conceptualized as a response to catastrophic trauma or stressor. It has been as carefully defined as many other psychiatric disorders and targets symptoms that military and disaster psychiatric preventive and evidence-based treatments seek to mitigate.

In addition to PTSD, other disabling responses to trauma may occur (Table 30.6–1). Adjustment disorders (characterized by anxiety, depression, and changes in behavior) can occur in response to the stressors associated with war, terrorism, or disaster. Conversion disorders are characterized by unexplained deficits in sensory or motor function (e.g., sudden paralysis of an extremity) without demonstrable neuroanatomical injury. Bereavement may occur in response to the death of a loved one as a result of war, disaster, or another catastrophe. Other observed distress-related responses include anxiety and depressive syndromes and antisocial behavior (including criminal behavior, military misconduct, and war-related atrocities).

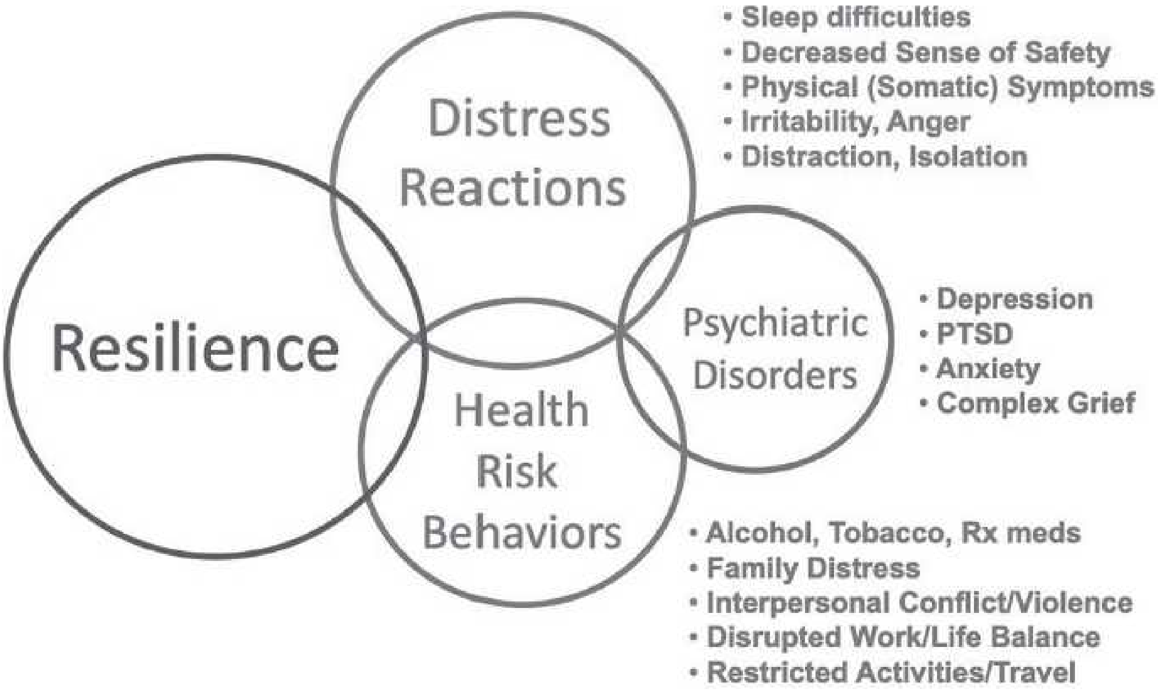
• Sleep difficulties
• Decreased Sense of Safety
• Physical (Somatic) Symptoms
• Irritability, Anger
• Distraction, Isolation
Distress Reactions
Resilience
Psychiatric Disorders
• Depression
• PTSD
• Anxiety
• Complex Grief
Health Risk Behaviors
• Alcohol, Tobacco, Rx meds
• Family Distress
• Interpersonal Conflict/Violence
• Disrupted Work/Life Balance
• Restricted Activities/Travel

FIGURE 30.6–1. Psychological and behavioral responses to disasters, terrorism, and war include distress reactions, health risk behaviors, psychiatric disorders, and resilience.

The constellation of symptoms, severity, and refractoriness of symptoms to conservative measures distinguishes between subsyndromal responses and psychiatric disorders.

Combat and Operational Stress and Medically Unexplained Physical Symptoms

Besides DSM-5 conditions, in the practice of military and disaster psychiatry, *combat and operational stress* (previously termed “battle fatigue”) are also significant. These terms encompass a variety of somatic and emotional responses to disaster and military operations that do not fully match the specific constellations of symptoms seen in major depressive disorder, adjustment disorder, PTSD, or conversion disorder. Symptoms such as gastrointestinal (GI) distress, tremulousness, and perceptual disturbance, including depersonalization and derealization, occur transiently in response to the traumatic exposure, sleep deprivation, loss of social support, or any combination of these stressful situations factors associated with military operations. Minor injury, parasitic infection, starvation, heat exhaustion, or cold injury may also contribute to a decreased ability to cope and, therefore, to operational stress symptoms.

Table 30.6–1.
Psychological Responses to Disaster and War: Associated Mental Disorders and Symptoms

Mental Disorders
- Acute stress disorder
- Posttraumatic stress disorder
- Generalized anxiety disorder
- Panic disorder
- Major depressive disorder
- Brief psychotic episode
- Adjustment disorders
- Alcohol abuse
- Substance abuse

Symptoms
- Anger and irritability
- Fear
- Restlessness
- Concentration and attention difficulties
- Sadness
- Insomnia (with or without nightmares)
- Somatic complaints
- Headaches, gastrointestinal distress, musculoskeletal pain
- Exacerbation of existing medical illness
- Increased alcohol and tobacco use
- Diminished interest in activities

Finally, war historians and military and civilian physicians have observed that unexplained physical syndromes are a common sequela of combat. Syndromes such as *soldier’s heart* and various physical symptoms (e.g., arthralgias, rashes, and dysesthesias) have been attributed variously to undetected neurologic or chemical attacks, Agent Orange, depleted uranium, or other toxic factors on the battlefield. Although debates among scientists, clinicians, veterans, and journalists persist, the consistent appearance of these syndromes—affecting various persons in various ways—after wars or other military operations points to the likelihood that psychosocial factors contribute to etiology.

A 28-year-old Army sergeant was involved in U.S. military operations in Afghanistan. His commander escorted him to the troop medical clinic because “he just isn’t acting right.” On interview, he reported that he had been involved in frequent combat, fighting from house to house for over 5 days. His temperature was 100°F, and his pulse was 90 beats per minute; his blood pressure and respiratory rate were normal. His physical examination was unremarkable. His mental status examination was notable for poor grooming; he was unshaven and smelled of perspiration. He was oriented to month and year but unsure of the day and date. He displayed poor eye contact and psychomotor retardation. His thought processes were logical, but he asked for several questions to be repeated. There was no evidence of psychosis, but he seemed focused on his perceived need to “get back to my men.” He noted that he and others had been under machine-gun fire for “a few days, I guess,” and that members of his unit were wounded, and two were killed because of a grenade blast. He reported these events with little display of emotion. He denied any history of psychiatric illness or contact with mental health professionals and denied any such history in his family.

DISCUSSION

This service member’s history suggests an adjustment disorder or acute stress disorder. However, the presentation was also consistent with a combat stress reaction. His mental status examination indicates the possibility of depression. His vital signs suggest that dehydration may be contributing to his presentation. Infectious diseases endemic to the region may also contribute. His report of proximity to a hand-grenade explosion suggests the possibility of blast-related traumatic brain injury (TBI) as well. Appropriate initial interventions include rest, rehydration, and respite from the ongoing battle. In addition, applying psychological first aid (PFA) steps, such as establishing safety and calming, will aid recovery and help differentiate a brief stress reaction from persisting illness or impairment. Additional history from members of his unit might clarify the extent of his involvement in previous combat and his degree of resiliency to these exposures. A negative evaluation for infectious disease, head injury, or other medical conditions and a rapid return to baseline function with respite and nourishment suggests a combat stress reaction and would allow him to return to duty. The possibility of future emotional sequelae from this event points to the need for continued surveillance.

ETIOLOGY AND EPIDEMIOLOGY

The mental health effects of a disaster have been historically well recognized. Epidemiologic data derived from methodologically sound studies in the aftermath of disaster have increased in recent years but are limited understandably, given the research challenges imposed by the chaotic environment of disasters. Studies conducted in Manhattan and Washington, D.C., in the months after the September 11, 2001, terrorist attacks demonstrated a substantial burden of PTSD and depression, as well as increased use of alcohol, tobacco, and marijuana. Rates of PTSD ranged from 7% to more than 17%, and rates of depression ranged from 9% to 15%. In one survey, nearly 29% of persons reported increased alcohol, cigarettes, or marijuana use. Studies of the terrorist bombing of the Oklahoma City Federal Building demonstrated that nearly 35% of the directly exposed population developed PTSD within 6 months of the event. Of importance, nearly 40% of those with PTSD had no history of a psychiatric disorder. The National Comorbidity Replication Study Survey established a prehurricane prevalence rate for mental disorders of 15% in the region affected by Hurricane Katrina.

In contrast, a posthurricane survey of the same population effectively demonstrated a doubling of overall rates of mental disorders to 30% 6 months after the hurricane. A systematic review of suicidality following a wide range of international and U.S. natural disaster events revealed a trend toward decreased suicidality in the first few months following a disaster, during the “Honeymoon” phase. However, numerous studies observed increased rates of suicidal thinking and behavior, which commonly began during the 6- to 12-month period following a disaster event, coinciding with the “Disillusionment” phase. Consistent with these findings, when compared to predisaster data, surveys given to greater than 1,000 affected residents 9 to 18 months following Hurricane Katrina found an increase in suicidal ideation (6.4% vs. 2.8%) and suicide plans (2.5% vs. 1.0%).

Considering exposure to war, historically, the suicide rate in the U.S. military has been below that in the civilian population, but it climbed steadily after the beginning of the Iraq and Afghanistan conflicts to the point where suicide is presently the second leading cause of death behind combat deaths and has exceeded demographically matched civilian rates since 2008. The rise in U.S. Army suicides has persisted despite substantial efforts to publicize and encourage the use of mental health services. The Department of the Army responded to these trends by collaborating with the National Institute of Mental Health (NIMH) to fund the Army Study to Assess Risk and Resilience in Service members (Army STARRS; http://www.armystarrs.org), a multicomponent epidemiologic and neurobiologic study of risk and resilience factors for suicidality and its psychopathologic correlates among Army personnel. Army STARRS is the most extensive study ever to address risk and protective factors for suicide in the U.S. Military, which has leveraged big data analytics to more precisely determine where risk is concentrated and develop actionable interventions to protect health and enhance military readiness.

A 2002 study of 2,583 veterans aged 18 to 54 years who received standardized psychiatric interviews in the National Comorbidity Survey demonstrated that combat exposure resulted in high rates of psychiatric diagnoses and psychosocial difficulties: 28% had PTSD; 21% engaged in spousal or partner abuse; 12% experienced a loss of employment (9% unemployed at the time of the survey); 8% reported substance abuse problems within 1 year after combat, and 7% reported symptoms consistent with major depressive disorder. In its oft-cited 2008 study, the RAND Corporation, Center for Military Health Policy Research, examined the prevalence of PTSD among previously deployed Operation Enduring Freedom and Operation Iraqi Freedom (Afghanistan and Iraq) Service members. PTSD was assessed using the PCL, as in the Gulf War Veterans study. Among the 1,938 participants, the prevalence of current PTSD was 13.8%. The study also showed that a substantial proportion of Iraq War veterans accessed mental health services in the year after returning home and suggested that concerns over stigma and potential career implications prevented others from doing so. Traumatic brain injury, resulting from blast and other injuries in disaster and other settings, has received increased attention, including as a comorbidity with PTSD when occurring in the context of exposure to psychologically traumatic events. The presence of TBI is associated with more complex and chronic dysfunction and impairment, particularly when comorbid with trauma-related psychological disorders. As the incidence and prevalence of intermediate and long-term effects of combat have become more apparent, so too have the neurobiologic and psychosocial explanations for observed psychopathology.

Predisposing Factors

PTSD, other anxiety and depressive disorders, and physical syndromes in response to a stressor are more frequently diagnosed among women than men. Explanations for this involve neurobiologic and psychosocial factors, including the greater rate at which women seek treatment for stress-related symptoms and the fact that the duration of illness may be longer for women and, therefore, may be more likely to reach clinical attention. In the aftermath of a disaster or during war, women are at higher risk for victimization through intimate partner violence and physical and sexual assault outside the home. Women may also be at increased risk due to having greater responsibility for managing family emotions and distress. Men are at greater risk for posttraumatic problems with substance abuse and antisocial and violent behavior. Gender-specific cultural factors and as-yet-unspecified neurophysiologic factors are likewise implicated in these differences.

The level of functional impairment in the aftermath of combat or disaster is also related to pretrauma functioning. Individuals who function marginally in social or occupational roles before combat or disaster exposure are at increased risk of poor function after trauma exposure. A history of childhood abuse or significant traumatic exposure generally appears to be a risk factor for developing PTSD after exposure in adulthood. If persons have successfully negotiated disaster or combat situations in the past, they may be less adversely affected by subsequent exposures (even if the circumstances of the trauma change from combat to disaster or vice versa). On the contrary, if past traumatic experiences produced PTSD or other psychiatric syndromes, further traumatic exposures may trigger exacerbations or the emergence of new psychiatric disorders. Specific subpopulations, such as children, may be particularly vulnerable with the well-being of children intimately tied to their caregiver's level of distress and function. While a history of mental health diagnosis does not necessarily increase risk, those with serious and persistent mental illness may be more vulnerable to the emotional consequences of traumatic experience due to a dependence on systems of care (e.g., case manager, injectable medication, day treatment programs), which may be disrupted or inaccessible during a disaster event. Older adults with cognitively impaired or limited mobility are less able to respond to the requirements of a disaster environment, including movement, evacuation, and seeking medical care, and are generally at increased risk for psychological injury. However, age and experience are also protective during disasters. For instance, older adults across various nations throughout the COVID-19 pandemic frequently reported lower levels of distress and health risk behavior than younger counterparts; a reminder that disasters create unique risk, and it is essential to be aware of assumptions risk (e.g., “being older makes someone vulnerable”). The loss of social support after disasters and in war-torn societies also results in the need for assistance in others who may have been previously able to cope (Figs. 30.6–2 and 30.6–3). The loss of access to medical care (i.e., medications) and health care providers is an indirect but substantial cause of casualties after disasters and terrorism.

Precipitating Factors

Precipitating factors are the proximate circumstances—the traumatic event itself—that initiate the various sequelae of trauma. For disaster responders and military populations, disaster and combat deployments and peacekeeping missions create additional stressors at the onset, including disrupting families. The high intensity and long duration of disaster or combat exposure increase the likelihood of psychiatric casualties. Specific experiences, such as physical injury or witnessing grotesque deaths, torture, or other atrocities, place individuals at increased risk for adverse mental health consequences. Rape, harassment, or assault can precipitate significant distress or illness in those victimized. Sexual assault is a powerful precipitant of adverse neurobehavioral changes.

Mitigating or Perpetuating Factors

Ongoing factors, including the security and safety of recovery environments, the extent of secondary traumatization, and—in first responder and military populations—rotation schedules, the degree of recognition or compensation for efforts, and one’s belief in the mission all affect the trajectory of the stress or illness/recovery response and the rate and severity of psychiatric illness and distress symptoms. Personal attributes or behaviors, such as over-dedication to the task at hand, may further contribute to dysfunction development. Symptoms in victims of disaster and war may be mitigated or exacerbated by perceptions of community leadership's preparedness for disaster, response to a crisis, recognition of heroes, and the distribution of resources such as the provision of medical, financial, or emotional assistance immediately after the crisis and over time.

Nongovernmental organizations (NGOs), such as the American Red Cross and the Salvation Army, help minimize the stress after a disaster. By attending to basic human needs, such as food, clothing, and shelter, they reduce the psychological and physiologic effects of the event. In recent years, the American Red Cross and other NGOs have developed training for volunteer health care workers to recognize, minimize, and treat stress responses in disaster workers and victims of disaster based on PFA principles.

Protective Factors

In groups such as military units, police, firefighters, and paramedic rescue units, protective factors may diminish the potential sequelae of trauma. Loyalty and interpersonal cohesion among members of such organizations may provide immediate access to social supports during the acute experience of a disaster and its aftermath. When such organizations possess strong leaders, these persons create and foster an environment of strong loyalty, mutual trust, and respect. Effective leaders communicate, manage their own stress, model self-care, and address issues of grief that are universal in disasters. Such an environment may reduce the extent to which one member of the group becomes excessively critical of another's performance during an emergency, may contribute to voluntary participation in high-risk rescue-and-recovery or combat missions, and may facilitate the synchronization of individual efforts that is necessary for success in these types of operations.

A critical factor in developing such trust and synchronization is the number of demanding, specific, and repetitive overtraining exercises in which elite police, rescue-and-recovery, and combat units participate. Disaster responders (firefighters, police, physicians, and civic leaders) assemble to train for responding to terrorist attacks, as well as natural disasters. Government emergency preparedness agencies, such as the U.S. Federal Emergency Management Agency, are increasingly coordinating such training. If the level of training is high, individuals in such units rely on ingrained principles aimed at mutual support in a quest for mission success. In addition, the National Center for Disaster Medicine and Public Health and Center for the Study of Traumatic Stress published disaster behavioral health curriculum recommendations that provide a comprehensive overarching framework for the development of training in disaster behavioral health preparation and response.

FIGURE 30.6–2. Children are particularly vulnerable to the emotional consequences of disaster.

FIGURE 30.6–3. Separation from social support networks contributes to stress reactions in the context of military operations.

The nature and extent of the fit between persons, equipment—such as chemical protective suits and masks, remote-controlled bomb detonators, or mine detectors—and the environmental extremes that may be experienced during a disaster (e.g., heat, cold, and sleep deprivation) modify stress responses in individuals and communities. The fit between the pilot, aircraft, and other aircrew members may also be improved by specific training. The extent to which subtler factors within the environment may modify stress response has been noted in those forced to exist in close quarters with limited contact with the outside world, such as those aboard submarines or ships. Although these factors cannot be eliminated from the practice of military operations or disaster response efforts, the extent to which these additional stressors exacerbate responses to other primary stressors may be reduced through training.

MANAGEMENT AND CARE DELIVERY

General Principles

Disasters, including terrorist attacks and military conflicts, shatter the expectation of a just and safe world within populations in which such notions of basic safety and justice are cultural norms. Because a lack of emotional preparedness makes chaos and disorder more likely after such events, primary prevention must begin with a recognition of the need for rapid dissemination of information regarding the threat (e.g., early warnings for tornados and hurricanes or reporting of anticipated terrorist attacks gleaned from reliable intelligence sources). The development of a planned response should follow this effort, including the practice of that response and the provision of funding and facilities supporting the plan. These steps can mitigate a community's sense of helplessness before such an event occurs and increase a community's capacity to provide social support to victims.

Secondary prevention (i.e., early intervention) begins with the initial triage and treatment of victims. First responders, medical care personnel, and mental health care providers must be prepared to provide psychoeducation and symptomatic relief for individuals with acute (and frequently transient) emotional or behavioral disorganization or other symptoms. It is vital to create a *holding environment*—locations in which such symptoms can be observed and monitored. This area should be sufficiently removed not to disrupt the ongoing triage and stabilization of life-threatening physical injuries but should be close enough to allow re-evaluation and further medical intervention should symptoms worsen. A holding environment may also prove to be a location in which affected individuals attend to such basic needs as food, shelter, and personal hygiene, which are critical to restoring a sense of normalcy after a disaster or traumatic event (Table 30.6–2). For military personnel and other responders, maintaining individuals near other units or team members maintains a sense of social connection and identity, which can aid in symptom resolution and return to normal function.

Initially, establishing a safe environment and managing life-threatening physical injury and the myriad disease possibilities that result from lack of food, potable water, or infection are the most critical psychiatric interventions. Subsequently, identifying high-risk populations, such as disaster workers, persons at impact zones, children, and the chronically ill, can focus further interventions. Outreach and public education strategies are critical because disaster and trauma victims rarely specifically seek mental health care. It is important to educate medical and community groups about the normal and predictable responses to traumatic events of disaster and war and when to refer for additional mental health treatment. Informing community and governmental leaders about the behavioral and emotional sequelae associated with disasters ensures the allocation of funds and resources for these phenomena in the disaster recovery plan.

Table 30.6–2.
Disaster Mental Health Response: Management Principles

Primary Interventions
- Early warning systems
- Development, rehearsal, and resource allocation for event response, including identification and roles of responders, methodology for information dissemination, potential locations for triage, initial treatment, and holding environments

Secondary Interventions
- Initial triage and identification of at-risk individuals or subpopulations
- Creation of holding environment, ensuring that needs for food, shelter, and hygiene are met, and allowing for close observation and medication management of acutely agitated individuals
- Brief support and psychoeducation, application/implementation of the principles of psychological first aid

Tertiary Interventions
- Medication management
- Cognitive-behavioral therapy
- Population surveillance for further development of symptoms or disorders

The responsibility for preventive measures and the initial recognition and treatment of the psychological consequences of disaster cannot be limited to available psychiatrists. General physicians, psychologists, and other social scientists must use their diverse skills to care for the victims of disasters and wars. It may be necessary for well-trained allied professionals to make preliminary diagnoses and provide initial psychoeducation and consultation to medical and surgical colleagues and other first responders. PFA has been developed as a concept to organize evidence-based and informed interventions in the acute phase of trauma and disaster exposure. A manual for PFA was developed by the National Center for PTSD and the National Child Traumatic Stress Network for mental health professionals.

Mobile apps and online resources for mental health are increasingly used during disaster events. PFA Mobile is an app-based adaptation of the PFA manual for use in the field, and online training is also available. Other organizations, such as the Substance Abuse and Mental Health Services Administration, have developed apps and other resources that use principles of PFA to support a variety of responders and emergency workers, including school crisis teams, clergy, emergency medical personnel, and military rapid deployment teams. Online training in PFA for leaders and supervisors is also available from the National Association of County and City Healthcare Organizations. During the global COVID-19 pandemic, the Covid Coach app has provided helping resources and support for lay persons. The Heroes Health app offers regular self-health assessments and local resources to understand better the health and well-being of personnel to tailor interventions.

Table 30.6–3.
Core Actions of Psychological First Aid

Action	**Goal**	**Examples**
Contact and engagement	Initiate or respond to contacts by survivors	"Hi. My name is ______. I'm just checking in to see how you are doing."
Safety and comfort	Create a physically and emotionally safe environment	Get survivor to a protected location Offer a blanket or water
Stabilization	Promote calming	Engage survivor in relaxation or focusing exercises Enlist the help of caregivers
Information gathering	Identify immediate needs	Ask about injuries and concerns. Ask about family or friends the survivor may be concerned about Ask about losses or other concerns
Practical assistance	Address immediate needs Develop an action plan (promote efficacy)	"It sounds like your most important concern right now is finding out where your sister may be. Let's think through some ways you might find out."
Connection with social supports	Facilitate connectedness with family, friends, community	Assist with phone contact with family Help identify other persons nearby that might offer support
Information on coping	Provide information about stress reactions and encourage adaptive coping	Explain intrusive reactions, avoidance, and hyperarousal as expectable reactions to trauma

		Discuss health risk behaviors and healthy coping
Link with collaborative services	Connect survivors to needed services	Assist with referral to assistance organizations as appropriate

PFA is an evidence-informed intervention framework that operationalizes social support, psychoeducation, stress management, and cognitive restructuring across the domain of short-term and intermediate mental health response supported by empirical evidence. The goals of PFA are to establish and enhance individuals’ actual and experienced safety; increase feelings of and ability to calm; foster and provide skills in self-efficacy; enhance social connectedness, and instill hope. These “essential elements” are associated with improved psychosocial function after disasters. Actions to facilitate these essential elements are listed in Table 30.6–3.

Therapeutic Interventions

Other interventions are needed for those with significant disease/illness or developing symptoms. Symptoms ranging from lethargy and depression to disorientation, dissociation, hallucinations, paranoia, and cognitive dysfunction may be observed in the immediate aftermath of a disaster. The diversity of response is complicated by the fact that symptoms may result purely as an emotional response because of a neurologic injury from a blast or other blunt trauma or, in the case of terrorist attack, from the neurotoxic effects associated with specific chemical or biologic agents (Tables 30.6–4 and 30.6–5). Rapidly identifying the underlying infectious, chemical, or physical insult to the brain directs the treatment strategy. It may be difficult to differentiate transient distress reactions from lasting illness in the acute setting. Acutely agitated patients may endanger themselves or other disaster workers and place exceptional demands on treatment staff that would otherwise be unavailable to deliver group-related interventions and education.

Pharmacologic interventions in the immediate aftermath of a disaster tend to be problem-focused and straightforward. The importance of keeping treatment as simple as possible under these confusing circumstances cannot be overemphasized. When a safe environment exists, brief supportive interventions or observation may alleviate the need for psychotropic medication. Sleep disruption will be a common complaint, and short-acting hypnotics may be used to restore standard sleep patterns. Antipsychotic and anxiolytic medications may be helpful to manage acute agitation from a range of causes. However, their metabolism by way of the hepatic cytochrome P450 system may alter levels of antimicrobial agents and acetylcholinesterase inhibitors used to treat biologic or chemical terrorist attack victims. The potential for drug–drug interactions bears on selecting antibiotics, chemical antidotes, and antipsychotic agents in specific cases. Persisting symptoms of depression, anxiety, and PTSD will respond to treatment with antidepressant medication, and treating persisting symptoms may enable survivors to engage in recovery activities more effectively.

Table 30.6–4.
Neuropsychiatric Syndromes or Symptoms in Selected Biologic Agents

Biological Agent	Syndrome or Symptom(s)	Comment
Anthrax	Meningitis	May be rapidly progressive
Brucellosis	Depression, irritability, headaches	Fatalities associated with central nervous system involvement
Q fever	Malaise, fatigue	In one-third of patients
	Encephalitis, hallucinations	In advanced cases
Botulinum toxin	Depression	Due to lengthy recovery
Viral encephalitides	Depression, cognitive impairment	Other mood changes are also possible
All biologic agents	Delirium	Acutely impaired attention and memory and perceptual disturbances

Beyond pharmacologic intervention, several psychosocial interventions have been described as helping prevent PTSD or other neurobehavioral responses to disaster, often, however, without evidence-based support. Community meetings or informal debriefings in which survivors can share experiences and offer mutual support can be helpful and should be encouraged. This practice should not be confused with originally proposed Critical Incident Stress Debriefing, which was for all people regardless of symptoms or exposure and can aggravate or create symptoms in those with none or few symptoms.

Cognitive-behavioral psychotherapeutic approaches (CBT) can target anxiety, social withdrawal, depressive symptomatology, and hyperarousal. CBT may reduce the immediate symptoms of depression, anxiety, and PTSD and may diminish the development of long-term morbidity. Trauma-focused psychotherapies have some evidence in the acute aftermath of trauma, mainly the benefit of brief exposure and cognitive restructuring. Trauma-focused psychotherapies are indicated and effective for those with persisting symptoms of PTSD. Because medical resources may be quickly overwhelmed in a large-scale disaster or combat situation, appropriate training of nonphysicians in delivering these therapies may allow for more effective delivery of care.

Table 30.6–5. Neuropsychiatric Effects of Chemical Agents

Organophosphates (nerve agents such as sarin, soman, VX, and tabun)	Depression, sleep disturbance, and impaired cognition; poor judgment, attention, and short-term memory; poor concentration, confusion, drowsiness, delirium, irritability, insomnia with nightmares, excessive dreaming, and psychosis
Atropine (treatment for organophosphates)	Blurred vision, tachycardia, dry mouth, sweat suppression, urinary retention, cognitive impairment, psychosis, and delirium (disorientation, hallucinations, and delusions)
Cyanide	Anxiety, confusion, giddiness, and hyperventilation
Blistering agents, mustard gas, and phosgene	No direct acute effects; inactivation of charcoal may result in sensations of suffocation or anxiety; psychic distress over disfigurement
Mycotoxins (fungal toxins)	Psychosis and somatic complaints, involuntary movements, anxiety, and agitation

Three weeks after a train derailment, a 42-year-old budget analyst presented to the mental health clinic. He noted that he was embarrassed to seek care, as he was previously a firefighter, but he felt that he needed “some reassurance that what I’m experiencing is normal.” He reported that, since the wreck, he had been feeling nervous and on edge. He experienced some difficulty focusing his attention at work, and he had occasional intrusive recollections of “the way the ground just shook; the tremendous ‘bang’ and then the screaming when the train rolled over.” He noted that he had spoken with five business colleagues on the train, and three acknowledged similar symptoms. However, they said that they were improving. He was more concerned about the frequency of tearful episodes, sometimes brought on by hearing the name of a severely injured friend, but, at other times, occurring “for no particular reason.” In addition, he noted that, when he evacuated the train, rescue workers gave him explicit directions about where to report, and, although he complied, he now felt extremely guilty about not returning to the train to assist in the rescue of others. He reported a modest decrease in appetite and denied weight loss but noted that he had stopped jogging during his lunch break. He had difficulty initiating sleep, so he had begun consuming a “glass or two” of wine before bed to help with this. He did not feel rested on awakening. He denied suicidal ideation or any psychotic symptoms. His sister took an antidepressant several years ago, but he did not desire medication. He feared that side effects could further diminish his ability to function at the workplace and could cause him to gain weight. Although this individual reported initially experiencing intrusions and altered arousal symptoms, these symptoms did not necessarily follow an expected improving trajectory in the weeks after the trauma, suggesting acute stress disorder. His family history of depression suggests that major depressive disorder may also be possible. Given the patient’s reticence to take medication, brief supportive and cognitive-behavioral therapeutic approaches are appropriate. In addition to allowing time for trust and rapport to develop, early response to simple interventions may suggest a more transient course and negate the need for medication.

SPECIAL TOPICS IN DISASTER PSYCHIATRY: RESPONDING TO NATURAL, HUMAN-MADE, AND WAR EVENTS

Military Mental Health Care in Peacetime

Military health care is part of disaster response around the globe for nonwar events and war. When not directly involved in combat operations, peacekeeping, or humanitarian relief, military psychiatric practice is, in many ways, like civilian practice, except for the fact that military psychiatrists go to work in military uniforms. The U.S. Department of Defense has tertiary care hospitals throughout the United States, Europe, and Asia. These institutions all have outpatient psychiatry or interdisciplinary behavioral health clinics or inpatient wards for severely mentally ill military personnel and family members. Some military psychiatrists receive their undergraduate medical and residency training at civilian institutions. However, the U.S. Congress in 1975 authorized the Uniformed Services University of the Health Sciences (USUHS) to provide medical education and produce physicians for military service. USUHS provides a 4-year medical degree program and several graduate degree programs in the basic and clinical sciences. The USUHS Center for the Study of Traumatic Stress conducts research and consults communities and federal and international agencies on matters surrounding individual and community responses to trauma, disaster, and war. Japan, Great Britain, and Russia are among nations with institutions that teach military-specific curricula to military medical care providers. As in other nations, these countries also call to national duty physicians not explicitly trained in military institutions during war or crisis.

The U.S. Department of Defense has residency programs in psychiatry at tertiary care Army, Air Force, and Navy Medical Centers within the United States. Congress mandates these institutions to provide military-unique curricula to ensure that the physician in training will provide care in combat or disaster environments. This type of training is not necessarily provided in civilian programs. Military training programs also introduce physicians to aspects of the military lifestyle (e.g., frequent moves and separations from family, living and working overseas, special occupational duties, and requirements for maintaining security over classified information) that affect military members and families, even when service members are not deployed. In addition, the USUHS offers the nation’s only Disaster Psychiatry Fellowship in collaboration with the Center for the Study of Traumatic Stress.

The U.S. military has attempted to decrease the incidence and severity of combat and operationally induced psychiatric disorders through several mechanisms. Mental health teams are now routinely assigned to U.S. forces in combat and are deployed in operations other than war. Each branch of the U.S. military services has specialized rapid intervention teams to provide consultative assistance and acute treatment as necessary to units that have experienced traumatic events. These teams provide prompt needs assessments, educate leaders on likely behavioral responses to stressful events, and recommend leadership interventions to reduce negative stress responses. Embedded behavioral health teams have served within military units throughout the wars in Iraq and Afghanistan in the U.S. armed forces and many allied nations.

Military units frequently are involved in disaster and other humanitarian response missions. Military units can quickly provide services and supplies to affected areas with the ability to rapidly deploy resources and an existing command and control structure. The U.S. military has been called upon to respond to disasters ranging from the Great Japan Earthquake to establishing field hospitals in the United States during COVID-19. Combat and operational stress control (COSC) has been a central part of these missions. In preparation for deployment, COSC entails educating personnel on hazardous exposures they can expect to encounter, expectable psychological reactions to these exposures, and techniques of self-care and peer support to manage stress reactions. During a response, COSC involves monitoring exposures of personnel through a needs assessment or simply through personnel presence walking around and offering immediate emotional support. Following humanitarian or disaster response, COSC is focused on educating responders on typical recovery trajectories, facilitating healthy self-care and peer support, screening for more severe symptoms, and treating as appropriate. This model was highly successful in the military response to COVID-19 in New York City and led to the dissemination of COSC principles throughout health care systems in the area.

In addition, embedded mental health teams have been established to improve access to care for those who might not otherwise request assistance and destigmatization of those who seek mental health care.

Military leadership has strongly supported the increased use of these units to establish immediate access to care that reduces distress and improves functioning. There are ongoing efforts to study the efficacy of such interventions in reducing the incidence of PTSD and somatic symptom syndromes that may be tied to psychological responses to combat. These teams have also provided consultation and intervention in military humanitarian relief efforts after hurricanes, floods, infectious disease outbreaks, and other natural disasters within the United States and abroad. In addition to direct patient care, military psychiatrists provide consultation to community leaders and civilian physicians not accustomed to responding to large-scale physical and emotional traumas. During the COVID-19 pandemic, at the request of city officials in hotspot outbreak locations, such as New York City, the military deployed embedded mental health teams in civilian health care facilities to train personnel and leaders in the principles of COSC.

Historically, different mission types, patterns of deployment, and medical support systems among the various branches of the U.S. military posed challenges to developing a unified tri-service approach to the management of operational stress. For instance, Armies typically deploy large units for extended periods and allocate large amounts of medical assets to support these units. This medical support includes specialty services. The Navy and Marine Corps deploy smaller units at sea and ashore. General medical officers and nonphysician providers furnish medical support, and specialty care is not routinely available in the operational theater. The Air Force has short- and long-range missions and, increasingly, deployment in place through remote drone warfare. However, the past decade has seen an increased emphasis on the use of fundamental principles of COSC that are predicated on the essential elements of PFA. The Department of Psychiatry at the Uniformed Services University provides a tri-service curriculum that educates uniformed medical students on these principles and provides support, including efforts to enhance standardization, to military psychiatry residency programs. While each uniformed service will likely continue to have different mission tasks resulting in exposure to some unique stressors, trauma responses to these events are similar and standardized interventions based on evidence-based principles will enhance operational sustainment and military readiness.

In the United States, definitive treatment of psychiatric illness is often provided in the military's system of hospitals. Medical care is provided to active duty personnel and their families. Other mental health specialists, nurses, social workers, and psychologists augment this care. Military members who develop psychiatric disorders while on active duty are eligible for medical retirement disability pay and continued treatment through a system of Veterans Administration hospitals. Under present regulations, some individuals may be separated from service because of personality disorders that existed before service entry without disability payment or ongoing medical care from the military.

Other nations with recent wartime or military operational experience, such as Israel, Croatia, and Canada, have developed programs to evaluate and treat soldiers and civilians exposed to combat or other operational stressors. Their experiences are somewhat different from experiences in the United States because they rely much more heavily on reserve forces. These nations and many of our North Atlantic Treaty Organization (NATO) allies have a more inclusive social, medical infrastructure; therefore, treatment programs are less reliant on the military medical system. Nonetheless, coordinating civil and military medical treatment efforts (e.g., in response to the COVID-19 pandemic) across member nations and partners has become a priority of military medical planners. Other nations are increasingly confronted with operational stress management in peacekeeping and humanitarian missions. Asian nations that have recently experienced natural disasters and terrorist events are also studying approaches to evaluating and treating individuals exposed to trauma and are incorporating lessons learned into their pandemic response.

FUTURE CHALLENGES AND EVOLVING ISSUES

As political, social, scientific, and technologic factors evolve, societies will change their responses to the consequences of disasters and wars. The psychiatric practice associated with wars and disasters has changed with the evolution of the scientific understanding of illness. In the future, the resources to deal with the consequences of disaster or war and the relative importance assigned to dealing with the resultant injuries and disabilities are likely to be influenced by political and sociocultural values and advances in knowledge of disease processes, treatment paradigms, and technology. Disaster and military psychiatry must anticipate these changes to ensure that current trainees can deliver care in a changing environment.

Emerging Threats

The COVID-19 pandemic has raised international awareness of the disruption and calamity resulting from global infectious disease outbreaks, and increased focus on availability and access to protective equipment, medications for prevention and acute treatment, the likelihood of widespread morbidity, death, chaos, and distress, as well as the convergence of multiple disasters, including concomitant economic crisis, political upheaval, and widespread social unrest. Implementing necessary physical distancing measures (e.g., quarantine, isolation, the closing of schools and workplaces) required to control infection spread created significant obstacles to the establishment of a holding environment and the fostering of social connectedness critical to public health response in a disaster. Ongoing uncertainty and shifting public health requirements taxed societies worldwide and required health risk and crisis communication. The unprecedented surge in health care demand altered the delivery of health and the frequency with which people accessed it. The COVID-19 pandemic refocused awareness of society on the role of preparedness and the critical importance of all-hazards planning, which considers the full breadth of threats and mitigating strategies.

Advanced technology has implications for those responding to human-caused disasters, such as terrorist attacks involving chemical, biologic, radiologic, and nuclear agents (CBRN). In the United States, the terrorist attacks on the Pentagon and the World Trade Center and the fatal distribution of anthrax-tainted mail in the fall of 2001 precipitated an ongoing “war” on terrorism. These events intensified efforts to increase coordination among civilian, nonmilitary, governmental, and military agencies for response to future terrorist events within the United States and abroad, as the scope of terrorist activity has become more globalized and technologically advanced. Though post-9/11 concerns about escalating CBRN terrorist attacks went largely unrealized, the use of coordinated multisite attacks has become an increased terror threat. Recent such events worldwide have resulted in unprecedented levels of community fear and disruption. Clarifying the roles of military and civilian responders and integrating mental health and public health activities in terms of triage, treatment, consultation, and education in any joint responses to crisis are challenges for military and disaster psychiatrists.

New technologies in combat continue to require modification to the means of sorting and treating persons with medical and psychiatric injuries for those exposed to war. Militaries in technologically advanced nations are increasingly smaller, move more rapidly across the battlefield, use advanced sensors, and direct intense fire across a considerable distance. These capabilities, coupled with the use of weapons of mass destruction, will likely make the battlefield more chaotic and inhospitable to human life. The increasing use of uncrewed aerial vehicles in combat operations and disaster reconnaissance has reduced exposure to physical injury but, paradoxically, created a new avenue for psychological and emotional distress as those who remotely operate this technology have a close-up and personal experience of human suffering and death, possibly at their own hands. Individuals who gather intelligence information from unmanned aircraft engaged in combat operations are effectively “deployed in place,” creating several potential challenges that result from being persistently exposed to the stimulus of war without the geographic or psychological distance afforded to traditional combat units. For instance, the disruption to unit cohesion and mission focus, coupled with a need to balance and manage personal life stressors by night while participating in warfare during the day, has created unique personnel sustainment challenges. For special operators working within small, remote units, emergency care and evacuation of those with disease and injury may become increasingly difficult. The inability to maintain contact with rapidly moving units may preclude returning individuals to their original units. Future military casualties may increasingly rely on unit buddies, medics, and front-line leaders rather than specialized medical units or specialists at hospitals in the rear. Underdeveloped nations may have limited access to advanced technologies, so more traditional ways of organizing medical and psychiatric practice may remain relevant in certain circumstances.

Additional changes in how war is waged have altered risk and care delivery. The evolution of highly mobile units on widely dispersed battlefields will decrease the opportunity for exchanging rested troops from the rear area for those exhausted by front-line combat. Provision of brief respite for exhausted troops—a hallmark of battle fatigue management—may become impossible because each individual performs a critical specialized task. Small medical units operating within the combat area are likely to be eliminated from this technology-intensive battlefield. Although treatment may, by necessity, move to the battlefront, medical specialists at the rear may render triage decisions and diagnoses using telemedicine communication technology, the use, and familiarity of which has rapidly accelerated due to the shift to telepsychiatry during the COVID-19 pandemic. Experience has shown that front-line mental health providers take a pragmatic view of acute psychiatric symptoms and tend not to make hasty formal diagnoses on overstressed troops. Rear-echelon providers, by contrast, tend to assign formal psychiatric labels that may be inaccurate and that may stigmatize troops without contributing to treatment. Rear-echelon mental health specialists in future battles must address the challenge of providing practical therapeutic advice from afar while avoiding unnecessary diagnostic stigmatization.

Cyber terrorism represents a rapidly increasing global threat. Cyber-attacks have become increasingly complex and malicious in recent years, with governments and large corporations becoming frequent targets of hacking with the intent to extort financial payments. For instance, the corporation that manages petroleum supply for the eastern half of the United States was attacked in 2021, disrupting gas access for individuals and corporations that precipitated fear-based excess buying and widespread shortages. This event added to the ongoing uncertainty and distress of the COVID-19 pandemic, enhancing a sense of fear and concern within the communities most affected. Infiltration of electrical grids, water supplies, and fuel reserves and the systems that deliver these to citizens represent potential mechanisms of inflicting catastrophic harm and inducing a sense of terror within a society, with developed nations potentially most vulnerable. Beyond financial gain, the use of cyber-attacks by those intended to disrupt and harm represents real and present threats requiring coordination and collaboration by intelligence, military, homeland security, and private sectors to protect national resources upon which societies rely.

Cultural Issues

Social scientists note that trauma responses may be considered normal or pathologic, depending on the cultural context. Many have expressed fears that mental health practitioners, motivated by profit, will try to convince individuals experiencing normal uncomfortable responses that they need treatment. Because of the persistent stigma associated with mental health care, a future challenge to military efforts at aiding other nations in crisis may be overcoming the belief that accepting assistance signals weakness.

Most individuals exposed to traumatic events deserve to be reassured that they will recover with a return to work, community, and family. However, some individuals (and, perhaps, some cultures) will experience more significant psychopathologic responses and more prolonged symptoms after trauma. The nature of the behaviors and symptoms associated with trauma response across cultures is uncertain. Psychiatrists providing care during combat and humanitarian environments in a cross-cultural environment must do so with an awareness of the unique aspects of how culture influences the experience of mental illness. The extent to which social supports, biologic–genetic predisposition, concurrent illnesses, and other political and economic parameters contribute to variability across cultures is unknown. Although it appears clear that the severity of the trauma is important, reliable severity measures remain to be determined. Trauma may, under some conditions, create the opportunity for personal growth as well, and further understanding of this potential must also be exploited to reduce morbidity.

Ethical Challenges

Providing care in the mass casualty situation raises ethical questions about the equitable distribution of resources and the application of moral values. The hyper-suggestibility of individuals recently exposed to trauma has, at times, led to the exercise of political influence and the manipulation of loyalties. Governments in trouble have withheld treatment to minority racial or political groups—clearly, an ethical breach. Because governmental terrorism is a common form of terrorism, care providers and leaders must be sensitive to the possibility that disasters can afford unscrupulous leaders an opportunity to manipulate citizens for their purposes. As seen in previous public health emergencies, the COVID-19 pandemic highlighted these concerns as political strife and elections worldwide focused primarily on varying messages and agendas about the appropriate pandemic and societal responses, often leveraging the frustrations of highly distressed communities for political purposes.

Military and combat operations raise challenging ethical dilemmas. To facilitate command assessment of troop health status, militaries have denied members' confidentiality in medical communication. Mental health care providers must strike a balance between a promise of privacy that encourages persons to seek care and responsible reporting to higher command regarding situations that pose a danger to larger groups. Thus, dual allegiance to individuals and the larger community presents an ethical challenge that the military care provider must negotiate. Persons in extreme circumstances may behave in ways that they later view as shameful. This self-view may contribute to symptom development and, at the same time, discourage care-seeking.

An innate aspect of war involves the collection of military intelligence to enhance the effectiveness of combat operations. Prisoners of war or other detainees are often a source of intelligence information. Commanders may seek the involvement of mental health care personnel in managing detainees. Providers will be asked to deliver mental health care to alleviate detainee suffering. They may also be sought out to guide behavioral interventions to facilitate intelligence information gathering. Physicians must be aware of ethical prohibitions against participating in cruel, inhumane, or coercive practices with prisoners of war or other detainees and of the challenges that emerge when their involvement is intended for purposes other than the direct provision of care. In addition to personal ethical dilemmas that arise, medical and allied professional organizations have provided guidance regarding the appropriateness of these roles.

Disaster triage is frequently carried out in large open areas that allow everyone present to hear what patients say to caregivers. Given the social stigma assigned to the manifestations of psychiatric illness, it is easy to understand patients' reluctance to communicate and doctors' reluctance to inquire. Research on disaster- and war-exposed populations also raises significant ethical issues surrounding balancing requirements to develop knowledge and to help others with the potential exploitation of those in need. Perhaps re-educating the population can reduce ethical and therapeutic problems associated with stigma. However, altering deeply ingrained cultural expectations is just as challenging as providing privacy in chaotic triage environments.

CONCLUSION

Natural and human-caused disasters, including terrorism and war, disrupt societal function. Wars and acts of terrorism, with their attendant large-scale death, injury, and destruction, affect populations much like massive natural disasters. However, these events' intentional, human-caused nature increases their mental health consequences. Disasters and war leave in their wake populations experiencing psychological disturbances that have been described by social scientists, civil leaders, physicians, and other care providers throughout the ages.

The consequences of exposure to disaster and war may take the form of distress reactions, health risk behaviors, or psychological disorders, such as PTSD, or may manifest as various (and sometimes more subtle) forms of behavioral change, anxiety, or depression. Symptoms may present at different times during and after traumatic exposure. Many factors complicate the evaluation and treatment of neuropsychological syndromes in the aftermath of war or disaster. Resources are overwhelmed, life-threatening illnesses require immediate treatment, and psychological casualties are often reluctant to seek assistance.

Progress has been made in identifying the nature of trauma-related psychological responses. Predisposing, exacerbating, and mitigating factors have been identified, as have traumatic, stress-related, neuronal, and intercellular events. The value of multidisciplinary preparation and training for disaster management and the need for outreach programs have also been demonstrated. The U.S. experience in the aftermath of September 11, 2001, the ongoing war on terrorism, and increased frequency of large-scale natural disasters, including the recent global COVID-19 pandemic, have clarified the need for greater integration among military, government, and civilian organizations charged with the responsibility of mental health response to these events. However, further study must clarify how psychotropic medications and various forms of psychosocial support and psychotherapy in the treatment of war- and disaster-related morbidity can be delivered by cooperating agencies in times of national need. With technologic advances and global economic shifts, the nature of war and other human-caused disasters has changed. Disaster and military mental health care delivery must anticipate future changes to develop improved prevention, evaluation, and care methods for individuals and groups devastated by war or disaster.

ACKNOWLEDGMENT

This chapter represents an update and revision of previous chapters in this series, most recently by Harry C. Holloway, M.D. from the last edition. We are indebted to Harry C. Holloway, M.D. for their work and retain much of their prose in this current edition.

DISCLAIMER

The views expressed are those of the authors and do not necessarily reflect the views of the Department of Defense, the Uniformed Services University, the Department of Health and Human Services, or the United States Public Health Service.

FURTHER READINGS

Beaglehole B, Mulder RT, Frampton CM, Boden JM, Newton-Howes G, Bell CJ. Psychological distress and psychiatric disorder after natural disasters: systematic review and meta-analysis. *Br J Psychiatry.* 2018;213:716–722.

*Benedek DM, Fullerton C, Ursano RJ. First responders: mental health consequences of natural and human-made disasters for public health and public safety workers. *Annu Rev Public Health.* 2007;28:55–68.

Brooks S, Amlot R, Rubin GJ, Greenberg N. Psychological resilience and post-traumatic growth in disaster-exposed organizations: overview of the literature. *BMJ Mil Health.* 2020;166:52–56.

Brymer M, Jacobs A, Layne C, et al. *Psychological First Aid: Field Operations Guide*. 2nd ed. National Child Traumatic Stress Network and National Center for PTSD; 2006.

Canyon DV, Ryan BJ, Burkle FM. Rationale for military involvement in humanitarian assistance and disaster relief. *Prehosp Disaster Med.* 2020;35:92–97.

Davidson TM, Price M, McCauley JL, Ruggiero KJ. Disaster impact across cultural groups: comparison of Whites, African Americans, and Latinos. *Am J Community Psychol.* 2013;52:97–105.

*Dodgen D, Donato D, Kelly N, et al. Mental Health and Well-Being. In: *The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment*. U.S. Global Change Research Program; 2016:217–246.

Fernandez CA, Choi KW, Marshall BDL, et al. Assessing the relationship between psychosocial stressors and psychiatric resilience among Chilean disaster survivors. *Br J Psychiatry.* 2020;217:630–637.

Flynn BW, Vance MC, Morganstein JC. Curriculum recommendations for disaster health professionals: disaster behavioral health. 2nd ed. 2020. https://www.cstsonline.org/assets/media/documents/CSTS_Curriculum_Recommendations_2nd_ed.pdf

Flynn BW, Speier AH. Disaster behavioral health: legal and ethical considerations in a rapidly changing field. *Curr Psychiatry Rep.* 2014;16:457.

Fullerton CS, Mash HBH, Benevides KN, Morganstein JC, Ursano RJ. Distress of routine activities and perceived safety associated with post-traumatic stress, depression, and alcohol use: 2002 Washington, DC, sniper attacks. *Disaster Med Public Health Prep.* 2015;9:509–515.

Goenjian AK, Steinberg AM, Walling D, Bishop S, Karayan I, Pynoos R. 25-year follow-up of treated and not-treated adolescents after the Spitak earthquake: course and predictors of PTSD and depression-ERRATUM. *Psychol Med.* 2021;51:989–990.

*Goldmann E, Galea S. Mental health consequences of disasters. *Annu Rev Public Health*. 2014;35:169–183.

Hobfoll SE, Watson P, Bell CC, et al. Five essential elements of immediate and mid-term mass trauma intervention: empirical evidence. *Psychiatry.* 2007;70:283–315.

Kessler RC, Galea S, Gruber MJ, Sampson NA, Ursano RJ, Wessely S. Trends in mental illness and suicidality after Hurricane Katina. *Mol Psychiatry.* 2008;13:374–384.

Kino S, Aida J, Kondo K, Kawachi I. Long-term trends in mental health disorders after the 2011 Great East Japan Earthquake and Tsunami. *JAMA Netw Open.* 2020;3:e2013437.

Morganstein JC, West JC, Huff LA, et al. Psychosocial responses to disaster and exposures: distress reactions, health risk behavior, and mental disorders. In: Shigemura J, Chhem RK, eds. *Mental Health and Social Issues Following a Nuclear Accident*. Springer Japan; 2015:99–117.

*Morganstein JC, Flynn BW. Enhancing psychological sustainment & promoting resilience in healthcare workers during COVID-19 & beyond: adapting crisis interventions from high-risk occupations. *J Occup Environ Med.* 2021;63:482–489.

Motreff Y, Baubet T, Pirard P, et al. Factors associated with PTSD and partial PTSD among first responders following the Paris terror attacks in November 2015. *J Psychiatr Res.* 2020;121:143–150.

National Academies of Sciences, Engineering, and Medicine. Exploring lessons learned from a century of outbreaks: readiness for 2030: proceedings of a workshop. 2019.

Norris FH, Friedman MJ, Watson PJ, et al. 60,000 disaster victims speak: part I. An empirical review of the empirical literature, 1981–2001. *Psychiatry*. 2002;65:207–239.

North CS, Pfefferbaum B. Mental health response to community disasters: a systematic review. *JAMA.* 2013;310:507–518.

North CS, Ringwalt CL, Downs D, Derzon J, Galvin D. Postdisaster course of alcohol use disorders in systematically studied survivors of 10 disasters. *Arch Gen Psychiatry*. 2011;68:173–180.

Orengo-Aguayo R, Stewart RW, de Arellano MA, Suárez-Kindy JL, Young J. Disaster exposure and mental health among Puerto Rican youths after Hurricane Maria. *JAMA Netw Open*. 2019;2:e192619.

*Pfefferbaum B, Flynn BW, Schonfeld D, et al. The integration of mental and behavioral health into disaster preparedness, response, and recovery. *Disaster Med Public Health Prep*. 2012;6:60–66.

Tanielian T, Jaycox L, eds. *Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery*. RAND Corporation; 2008.

Tucker P, Pfefferbaum B, Nitiéma P, Wendling TL, Brown S. Intensely exposed Oklahoma City terrorism survivors: long-term mental health and health needs and posttraumatic growth. *J Nerv Ment Dis*. 2016;204:203–209.

Ursano RJ, Fullerton CS, Weisaeth L, Raphael B, eds. *Textbook of Disaster Psychiatry*. 2nd ed. Cambridge University Press; 2017.

*Ursano RJ, Colpe LJ, Heeringa SG, et al. The Army study to assess riskand resilience in servicemembers (Army STARRS). *Psychiatry*. 2014;77:107–119.

West JC, Morganstein JC. Psychological effects on military personnel assigned to humanitarian assistance and disaster response missions. In: Christodoulou GN, Mezzich JE, Christodoulou NG, Lecic-Tosevski D, eds. *Disasters Mental Health Context and Responses*. Cambridge Scholars; 2016:69–82.

30.7 Psychiatry and Spirituality

SHWETA KAPOOR, M.D., PH.D.

Historians have traced the slow but progressive decline of religion since the 16th century, particularly in the West, when the Protestant Reformation led to the fragmentation of Christianity. Among the key events contributing to secularization are the scientific revolution started by Copernicus, the rise of mathematics and empirical investigation, the Age of Enlightenment in which intellectuals opposed religion as an institution, the French Revolution that expunged religious influence from public life, and the 19th-century theories of thought that dispensed with God by stressing the importance of economics (Karl Marx), biologic evolution (Charles Darwin), and sexual urges (Sigmund Freud). Secularization in the West is epitomized by increasingly empty European churches: 72% of the population of Britain and Germany identify as Christians, but less than 2% attend church regularly. Most indicators of traditional religiosity are either stable or declining; for example, there is lessening confidence in religious leaders and a decline in those choosing religious vocations. In the current times, the transforming sociocultural environment and reformative norms have facilitated engagement in life without necessitating the integration of religion, such as attending school, marrying, making social connections, and entertainment.

Spirituality has emerged as an important construct parallel to the waning prominence of religion. Spirituality embraces a diverse set of definitions, theories, and explanations. Earlier on, it was described as “a fuzzy concept that embraces obscurity with passion.” More recently, with the accelerating empirical interest in this construct, more distinct descriptions have emerged. Social scientists have variously described it as a “sacred connection with God” and “a subjective experience of the sacred.” The word itself has biblical origins and derives from the Latin translation of the Hebrew word variously meaning breeze, wind, breath, and by extension, the breath of life. Traditionally, it has been linked with religion, but in polls, many people, particularly those born between 1980 and 2000 (Millennials or Generation Y), self-identified as “spiritual but not religious.” They seem to be put off by religious dogmatism and intolerance. While they still may profess a belief in God, they are more likely to encounter God while walking in the woods at sunset rather than at a place of worship. Consistent with this trend, those born after 2000, commonly referred to as Generation Z, are the least likely to profess a religious identity. In fact, they are two times more likely to identify as atheists than the general United States adult population. To many of them, the problems and suffering they see in the world appear to be a significant source of ambivalence in believing in a just and loving God and finding value in more traditional forms of religion.

George Vaillant describes spirituality as an amalgam of positive emotions—compassion, forgiveness, love, hope, joy, faith/trust, awe, and gratitude. The spirituality that arises from these emotions “binds humans to one another to experience ‘God’ as we may understand Him/Her/Them.” The positive emotions have an evolutionary basis and architecture. In this regard, our limbic system plays an important role and underlies spirituality. The human neocortex has evolved over 200 million years to generate science, culture, ideas, behaviors, and religion and has brought the limbic positive emotions into consciousness. Cultural evolution, working over a shorter time frame, was facilitated by the development of language that allowed for mastery of past and future tenses. A transformative cultural evolution occurred during the Axial Age, variously described as ranging from about 800 BCE to 100 CE, when in many areas of the world major religions were formed along with the development of moral conscience. As rationalism in political and ethical behavior became increasingly important, the concepts of self-consciousness and a transcendent element in life gained prominence. The third evolution in the development of spirituality is that which occurs when the positive emotions are learned and grow during the maturation process from infancy to old age. Love, for example, is Vaillant’s shortest definition of spirituality and “lives in the anterior cingulate gyrus.” It is catalyzed by genes and by neurochemistry; for example, oxytocin in limbic dopaminergic centers is linked with lasting human attachment. It is learned both through cultural examples—poetry, songs, stories—and through identification with persons who exemplify it by their behaviors.

Religiosity and spirituality are often used interchangeably. However, now more than ever, scholars in this area are delineating the two constructs, and increasingly these two are very different from each other. Religion arises solely through cultural evolution with its sacred texts and dogmas, where it has played an uneven role. It has produced intolerance and rigid beliefs and behaviors while being an opening through which the positive emotions are brought to conscious attention and reflection. Religion often involves a rigid set of beliefs, it binds believers to their own community, and it is imposed from without. On the other hand, all the positive emotions that relate to spirituality are grounded in a threefold evolutionary heritage. They allow for transcendent moments that appear to represent "religion without the side effects." Spirituality primarily involves emotions and trust, accepts the experiences of others as valuable, is democratic, and is imposed from within. The ultimate promise of many of the world's religions is a happy eternal afterlife in some sort of heaven, but the promise of spirituality is living an ethical, moral, and compassionate life guided by the positive emotions and sometimes punctuated by transcendent moments.

A biologic basis for religious experiences has long been evident in cases of persons with temporal lobe epilepsy and religious delusions and obsessions in the context of Schizophrenia and Obsessive-Compulsive Disorder. Over the past few decades, there has been mounting evidence that spirituality positively influences a person's health and well-being. Neurologic studies have generally revealed a frontal parietal circuit in which increased frontal-lobe activity is associated with advanced meditative/prayer states and with an enhanced propensity for the representation of religious concepts. A more robust finding is that decreased activity in the right parietal lobe (RPL) is associated with a decreased focus on the self and is the primary neuropsychological foundation for spiritual transcendence. More recently, studies have suggested a possible common brain circuit for both spirituality and religiosity in the periaqueductal gray. The latter is located in the brainstem and is related to pain and altruism.

Brick Johnstone has been a pioneer in developing a functional geographic model based in part on the studies of persons with traumatic brain injury, where it is possible to relate specific cerebral structures and neurophysiologic abilities to specific spiritual traits and abilities. After administering comprehensive neuropsychologic evaluations and various measures of spirituality and religion, he determined that decreased activation of the RPL, due to brain injury, surgery, transcranial magnetic stimulation, or meditation/prayer practices, provides the foundation for transcendent experiences. The decreased activation results in the experience of selflessness and a diminished personal reference to space, time, or relationships. This selflessness "allows individuals to focus on and emotionally connect with things beyond the self, i.e., the very definition of self-transcendence." The reason that this selflessness is experienced as spiritual transcendence rather than typical symptoms of RPL dysfunction such as anosognosia or left-sided neglect "appears to be partially related to the context in which the selflessness is experienced, e.g., in an artistic/musical context, personal relationships, religious/spiritual setting, etc."

Johnstone notes that neurologically based spiritual transcendence can be perceived as a numinous connection with God or a high power or as a mystical connection with the universe based on the environmental context and a person's religious and cultural upbringing, for example, numinous if reared in a monotheistic/polytheistic culture or mystical if reared in Buddhist culture. Differences in neurologic processing may reflect different cultural values, such as the Western independent self in contrast with the East Asian collective self. With its connection to the limbic system, the left temporal lobe may serve as a brain center for the universal experience of common religious archetypes (God, demonic figures, prophets, etc.) that can be a basis for differing theologic, ethical, and moral constructs. Increased frontal lobe activity is associated with increased frequency of prayer attendance, but its association with transcendence is unclear. It may have a role in focusing attention on spiritual thoughts, beliefs, and practices.

In addition to rational understandings of spirituality, there is also an esoteric tradition exemplified by the Greek Eleusinian and Northern European Odinic Mysteries. New Age spirituality emerged in the 1950s. It emphasizes irrational personal experiences and notions such as spirituality-laden force fields, extraterrestrial and earthly vibrations, and topsy–turvy interpretations of quantum mechanics. It is amazing that so many people worldwide hold such irrational, albeit usually harmless, beliefs.

FORMS AND METHODS OF ACHIEVING SPIRITUALITY

Because of the exceedingly broad definitions that people apply to the notion of spirituality, there are myriad ways people employ to achieve a spiritual moment and lead a spiritual life. Being a vegetarian, for example, may be considered by some people to be spiritual because it aims at sparing animals from being slaughtered, thus preserving the sanctity of living creatures. This section describes some of the many ways that people induce or enhance spirituality. It is extremely broad in scope, though, and to capture it all is a challenge. For instance, some individuals utilize altered states of consciousness while others do not. Some individuals rely on reducing sensory stimuli, whereas others rely on increasing stimuli. Some persons find spirituality deep within themselves, whereas others find it in the world around them.

Consciousness

Consciousness is difficult to define or explain except to state that it is the context in which one experiences the physical world and their mental world, including a sense of self. How it emerges is unknown. In addition to awareness, it has an unconscious component—dream states and the memories and feelings of past experiences that are not readily available to conscious attention. It is scientifically possible to parse the various contents of consciousness but attempts to explain any findings and to understand why they are experienced lead people in circles because any explanations and understandings arise from the totality of consciousness itself.

Consciousness can be altered and the sense of self changed, but the concept of self is multilayered. I can cut myself, for example, but I cannot cut my Self. Be that as it may, one perspective on spirituality is that it is a meaningfully interpreted alteration of consciousness, especially the sense of self.

People spend much of their lives in an altered state of conscious sleep, during which they have dreams. For almost a millennium, dreams have been associated with spirituality. In ancient Greece, asclepeions were healing temples sacred to the god Asclepius. Sick pilgrims slept overnight in the temple and reported their dreams to a priest, who then prescribed a cure. If barren women dreamed of snakes, they believed this to portend pregnancy. Throughout the Bible, God interacts with people through dreams. Many religious visions may have actually been remembered dreams.

When people are awake, altered states of consciousness are brief (from a moment to an hour, except in cases of pathologic dissociation, drug intoxications, or sicknesses such as coma or delirium), typically pleasurable, unusual experiences that cause a break with their normal sense of reality. This detachment is a form of dissociation and a gateway to the world of spirituality. Depending on the social context, the altered state of consciousness may be interpreted as having either a religious or a naturalistic meaning. The psychologist Abraham Maslow, for example, depicted a hierarchical pyramid ranging from basic physiologic needs upward to safety, love, esteem, and self-actualization needs. At the top of the triangle is the “peak experience,” which consists of a sudden moment of well-being, wonder, awe, and at times an awareness of a higher truth and the unity of everything in the world. Triggers for these peak experiences can come from exposure to beautiful scenes of nature or to great art and music, intense feelings of love, and a confrontation with severe illness or death. Maslow believed these moments were therapeutic because they increased empathy, creativity, self-determination, and personal growth. He claimed that everyone can experience peak moments and that they should be cultivated.

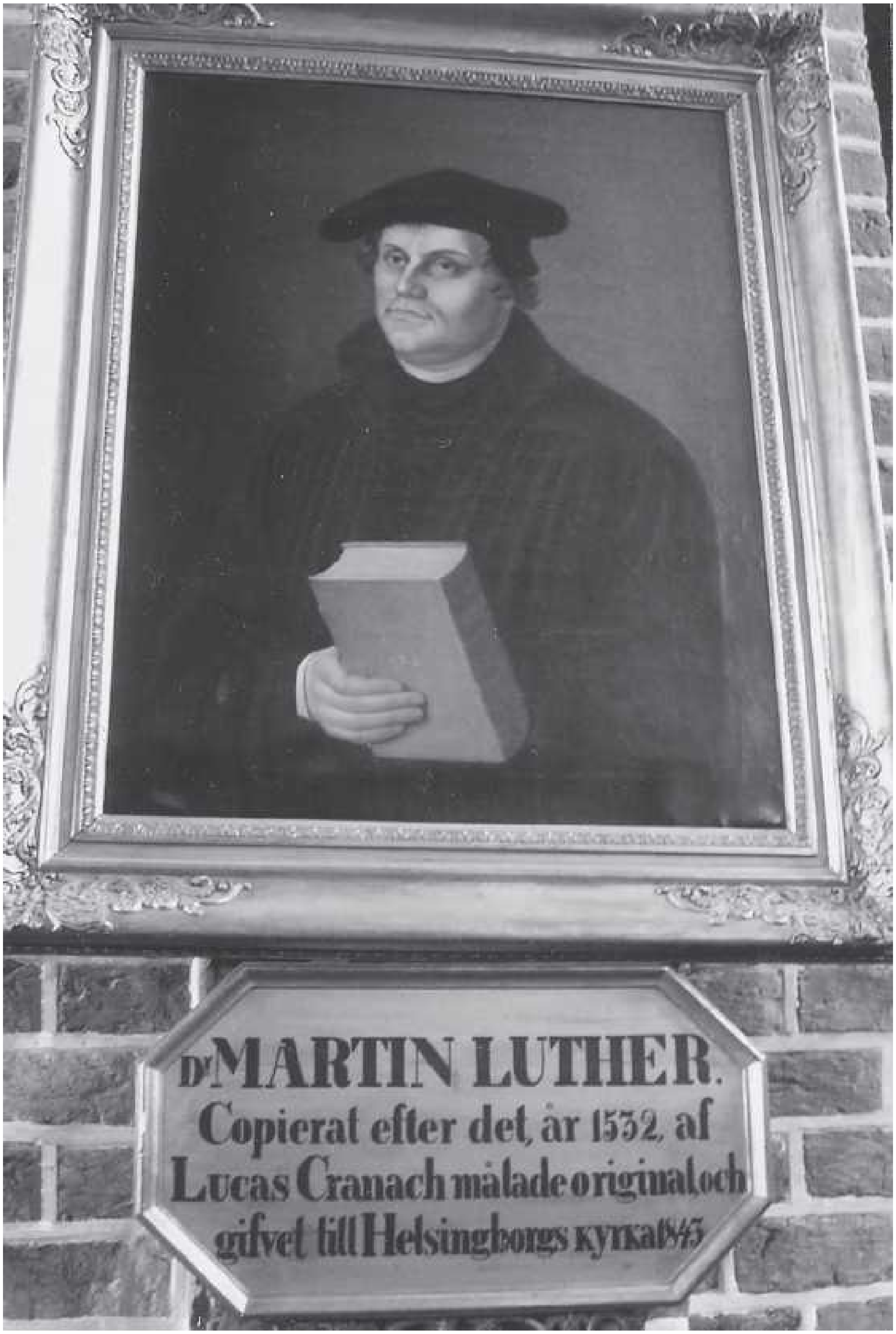
Dr MARTIN LUTHER.
Copierat efter det, år 1532, af
Lucas Cranach målade original och
gifvet till Helsingborgs kyrka 1845

FIGURE 30.7–1. Martin Luther (1483–1547) led the Protestant Reformation and spurred the movement to translate the Bible into local languages. He believed that salvation could not be achieved by good deeds but rather by

faith alone. (Frederick Kemmelmeyer. Corcoran Collection [Museum Purchase], National Gallery of Art.)

It is not uncommon for people to experience inexplicable fleeting insights that are less than peak moments. The latter may fall short of an actual altered state of consciousness but still serve as a spiritual dimension and may incorporate into a person's spirituality. Martin Luther (Fig. 30.7–1), for example, wrote that when he heard a fellow monk recite the words from the Apostle's Creed, "I believe in the forgiveness of sins," he saw Holy Scripture in an entirely new light, felt that he was born anew, and had found the door of paradise thrown wide open. A single word or phrase, a line in a poem or a song, particular sounds from a symphony or an opera, even the play of sunlight on rocks or water can strike a chord of a seemingly important moment.

Meditation

Meditation is an activity usually practiced in a tranquil setting that attempts to self-regulate the mind by freeing it of wandering thoughts and emotions, especially negative ones, by promoting psychophysiologic relaxation. Routine practice is shown to help with emotion regulation and promote peace within oneself. There are many types of meditation practices that are best understood by relating them to the specific sociocultural traditions from which they arose. Generally, Western meditation practices focus on increasing a deeper love for God and a sustained focus on religious symbols and the revelations found in sacred texts. Spirituality is achieved when the mind is filled with thoughts of God and devotional images. Transcendent states may be welcomed but play a minor role. Eastern meditation practices, especially Buddhist ones, focus on the primacy of the mind without reference to religion or faith. Spiritualism is achieved when the foundations of the self are transcended by comprehending that the self is an illusion and by eliminating all thoughts. In both traditions, when meditation practices are mastered—no easy task—they can cause decreased activation of the RPL that may result in transcendent moments.

As the world is in the throes of a deadly pandemic, COVID-19, meditation may provide a way to alleviate a part of the intense levels of stress being experienced by the population in general but particularly frontline health care workers. Meditation has the potential to provide some form of relief. Meditation is known to provide multifold benefits such as decreased stress, anxiety, and mood disturbances. Neurologic investigations have indicated that regular meditation practice brings about brain changes and provides a window to the mechanism of neuroplasticity. It can be tailored by the patients to maximize positive impacts, and in the midst of the pandemic with all its uncertainties, meditation is ideal to introduce to patients. It also provides access to a more indigent population as it is an inexpensive method of addressing the myriad of stressors.

Mysticism

The quintessential example of pure spirituality is the intense altered state of consciousness known as a mystic experience, which has traditionally been described within a religious context. There are several characteristics of these experiences. Individuals who have these experiences claim that they are so uniquely personal that their quality and worth cannot be fully communicated to others. They are a totally absorbing combination of feelings and revelatory insights felt to be of such enormous significance that they transcend logic and intellectual comprehension. During the experience, mystics lose their sense of will and feel totally absorbed by the awe of the unity of the universe and their immersion in the infinite ocean of God or of nature or of nothingness. They feel an elation and joyfulness unlike any other experience. The ultimate phase of the intense mystic experiences, at least in the Christian tradition but probably in others as well, is often preceded by what St. John of the Cross called “the dark night of the soul.” It is a period of helplessness that may be brief or last for decades before the Self disintegrates, surrenders itself to the divine will, and eventually merges in rapture with the only true reality God. As normal consciousness returns, mystics who have had this experience return to normal life but have been transformed into intercessors between God and humankind. Although today such claims are met with skepticism, it was not uncommon in the past for mystics to claim ecstatic raptures in which they dueled with devils and interacted with Jesus and the angels. One of the best-known mystics is St. Teresa of Avila (1515–1582). In one of her most famous raptures, she described an angel who pierced her heart into her entrails with a long dart with a fiery gold tip. As he withdrew it, she burned with a great love of God: “So great was the pain that it made me give those moans; and so utter the sweetness that this sharpest of pains gave me, that there was no wanting to stop, nor is there any contenting of the soul with less than God." This orgasmic rapture is immortalized in a statue in the Roman Church of Santa Maria della Vittoria by Gian Lorenzo Bernini.

Dhyana. *Dhyana* is a form of concentrated meditation taught by Buddha and Hinduism. It is a practice of controlling your mind to achieve an altered state of consciousness during which all thinking is abolished and truth is realized. Some practitioners of *dhyana* own only one appropriate garment, eat simple foods in small quantities, live in a quiet, humble abode, and are disentangled from worldly affairs. They must censor the desires that arise from the five senses, depriving themselves of gratifications and their ensuing cravings. They must clear their minds of thoughts that may result in greed, anger, laziness, sleepiness, recklessness, remorse, and doubt. They must sit quietly in a certain position, cross-legged with the spine erect, and breathe silently for 1 to 6 hours. Their minds must be tranquil and in harmony with their bodies and their breathing, and they should be totally aware of the emptiness and deceptive nature of the world with all its fraud and suffering. They may have visions of swollen, dead bodies with pus oozing out and maggots fattening on them so as to lose all pride and confidence in the ego-self and in the selfhood of others and thus develop a dissolution of all attachments, as well as compassion for all sentient beings. They then realize that neither the phenomena of the world nor the ego-self are real. As they focus on the merits and purity of all the Buddhas, they experience a pervasive sense of bliss and safety. They are desireless and develop insight into both the emptiness of all phenomena and the potentiality within their emptiness, an insight that allows them to follow a "middle way" to attain supreme, perfect wisdom and Nirvana, which ends the cycle of reincarnation.

Samadhi. In Hinduism, *Samadhi* is a process of meditation that also leads to awareness of one's existence without the distractions of normal consciousness. It often involves an initial trance state that evokes a joyous sense of well-being and then, with the help of spiritual masters known as gurus or yogis, grows into detachment from one's body and eventually from one's ego. The meditator may attain the deepest moment of self-actualization during which the heart supposedly stops beating, breathing ceases, and there is dissolution into the "divine ocean of love" and a realization of the unity of the universe.

Zen. The core practice of Zen Buddhism is *zazen,* the Japanese word for seated meditation in which a spiritual master instructs his students, some of whom will go on to become masters themselves, thus establishing a lineage of authority. Zen meditation uses *koans,* paradoxical or seemingly meaningless dialogues or questions, to test students' progress. "What is the sound of one hand clapping?" is an example of a *koan* and is akin to the Christian theologic question, "How many angels can dance on the head of a pin?" The goal of *zazen* is an epiphany (*satori* in Japanese) or lasting enlightenment about the true nature of creation.

Meditation in the Christian tradition is primarily an intellectual, sustained focus on spiritual themes such as Christ's suffering on the Holy Cross and his resurrection. Meditation on verses from the Bible is a common practice and derives from the book of Joshua 1:8, which states, "Do not let this Book of the Law depart from your mouth; meditate on it day and night, so that you may be careful to do everything written in it, then you will be prosperous and successful." Adoration of the Eucharist, which demands focused concentration, sometimes for hours at a time, is a particularly Catholic practice.

Transcendental Meditation. Transcendental meditation, derived from Hindu meditative practices again, is the most popular form of New Age meditation. It was trademarked by its founder, Maharishi Mahesh Yogi, in 1958. Although it is "officially" nonreligious, its goals are to attain "God Consciousness" and "Unity Consciousness," which are described as pure states of "being" in which a person sees the Self as integrated into all creation. This form of meditation uses physical exercises with softly spoken suggestions that result in a mild trance facilitated by a group setting with teachers. Repetitive suggestions of what is and what will happen lead students to believe what they are told. Maharishi refers to Transcendental Meditation as a pathway to God. Some clergy see spiritual value in it, whereas some critics see it as bogus, as in the false claim that people actually levitate during Transcendental Meditation, whereas all that happens is that meditators sit cross-legged on rubber mats and hop-forward on their derrieres while being told and believing that they are levitating.

Mindfulness Meditation. The concept of mindfulness meditation and its practices derives from a Buddhist tradition that focuses on the true nature of reality. In 1979, Jon Kabat–Zinn developed a secular Mindfulness-Based Stress Reduction (MBSR) program for the treatment of chronically ill patients. It is now also used to promote better mental and physical health for persons in the community and in schools, businesses, government agencies, prisons, and veterans' centers. The book *Waking Up: A Guide to Spirituality Without Religion* by Sam Harris is a thoughtful presentation of mindfulness.

A simplistic view is that it is the nonjudgmental, undistracted acceptance of all person's thoughts, emotions, and sensations in the present moment. For Harris, a person's experience "is always now" and what is subjectively real in every moment. In advanced mindfulness meditation, the uncluttered mind aims not at being lost in thought or thinking more clearly about experiences but rather at experiencing consciousness itself more clearly. Thinking is indispensable to life, but a person's identification with thoughts and failure to recognize thoughts as experiences in consciousness is a primary source of human suffering and of the illusion that a separate sense of self lives inside one's head. The self is an illusion, even though people feel that it exists as their body and brain. Mindfulness breaks the spell of thought and of an illusory self. It allows a person to know what pure consciousness is between one thought and the arising of the next one and reveals that consciousness is intrinsically selfless. "The conventional self is a transitory appearance, and it vanishes when it is looked for." Enlightenment, the direct experience of pure consciousness, reveals a void filled with everything and is the basis of a spiritual life. Selflessness gives rise to spiritual/mystical experiences and gives one's life a moral dimension. Despite Harris's belief that it is conducive both to physical and mental well-being, specific health claims have probably been exaggerated.

Prayer

For most of humankind, prayer is a primary source of entry into the world of spirituality. Prayer is a method of communicating with what is believed to be a supreme spiritual being, other spiritual beings (angels, saints, the Virgin Mary, totemic animals, etc.), or ancestors. The communication may be verbal or silent, private or public; sung, danced, or acted out behaviorally; spontaneous or planned as part of a religious service, spinning prayer wheels, reciting the rosary, praying at certain times of the day, assuming a particular physical position, or facing a specific direction.

Prayer may be used to worship or to placate spiritual beings; to confess sins; to petition for help in a variety of cases such as illness or poverty; to seek guidance, justice, or mercy; to enhance spiritual growth; to curse enemies; and to exorcise demons (Fig. 30.7–2).

FIGURE 30.7–2. The monastery of Grigoriou on Mount Athos, Greece, the major center of institutions devoted to Christian Greek orthodox prayer. (Courtesy of Scott Cairns.)

People who pray assume that their prayers are received by an appreciative spiritual being who may or may not acknowledge or act on them. Prayer, as defined traditionally, posits a spiritual universe that can become operational in the phenomenal world—that is, prayer activates a spiritual energy that becomes manifest in everyday reality. Belief in a deity who has a personal interest in the life of all persons heightens the intensity of the prayer experience, but the experience may also be salient for those who believe in an impersonal deity that is not a being but rather is *being* itself, the god above god, the force of ultimate reality from which grace flows and on whom everything is dependent (Fig. 30.7–3).

Spirit Possession

Belief that a supernatural entity may enter into and even control a person's body and mind is an ancient and still-current spiritual belief found in most cultures. It is a belief that underlies such comments as “What’s gotten into you?” or “What possessed you to do that?” and serves as the original prototype for Axis I mental disorders that people may have but from which they may be dispossessed with treatment. The possessing spirits may be malevolent or benevolent, and the possessed person may be perceived to be disturbed and sick or to be a healer or a prophet. A worldview that allows for spirit possession can be understood fully only on its own terms and not scientifically. Within a given culture, spirit possession may be quite logical, and the pantheon of consensually validated spirits is unlike a projected paranoid pseudocommunity.

In Egypt, Sudan, Ethiopia, and the Arabian peninsula, *zar* spirits, somewhat akin to *jinn,* are thought to possess women and cause headaches, anxiety, nightmares, and similar symptoms. The diagnosis of *zar* possession is made during a ceremony. A possessed woman will twitch and dance when a song that becomes special to her is played by musicians. She then goes into a trance, and her possessing *zar* will demand gifts such as jewelry or clothing as an appeasement price for it to depart. The woman’s husband and family give her the gifts, which she can keep for a while before returning them as her symptoms disappear. Thus, the *zar* interaction results in the woman’s gaining the attention of her husband and asserting her position as a wife whose wishes should be respected. Many communities have weekly *zar* ceremonies during which women dance when their special song is played.

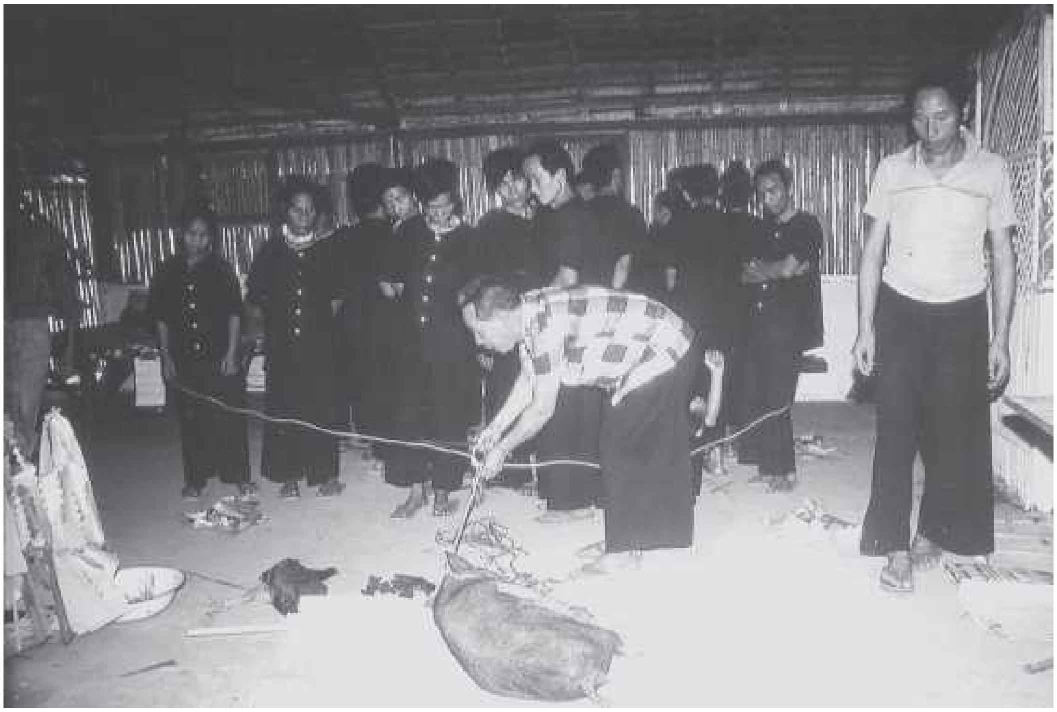

FIGURE 30.7–3. A Lao Hmong shaman who is a specialist in curing opium addiction prays as he ties a group of addicts to a sacrificial pig that will travel as a messenger to the opium divinities with "spirit money." The therapeutic ritual is meant to free the addicts' souls from dependence on opium. (Courtesy of Wolfgang Jilek.)

The utility of spirit possession is demonstrated in many shamanic rituals. A shaman is a generic term for a person who follows a divine call to healing, has received instructions directly from spiritual contacts, and may be a medium who communicates those spiritual instructions. Different types of shamans can be found in almost all cultural groups, such as Haitian *hungans,* West Indians *obeah* men, Puerto Rican spiritists, and Christian faith healers. Psychosocial profiles of shamans show that some are imposters, some are mentally disturbed, and some are mentally healthy, mature persons with a special cognitive style. The last may be considered true shamans who possess an ability to deal with and organize unconscious needs and primary process thinking. Often, through the mechanism of trance and spirit possession, the shaman acts as an agent of a powerful spirit and is thus able to perform psychological cures that may be temporary or permanent; for example, the spirit may declare that a patient's lost soul has been retrieved, as in cases of *susto* (spirit loss) in Hispanic cultures (Fig. 30.7–4).

Shamanic rituals are usually public affairs, and the persons in attendance include not only identified patients or troubled persons but also family, friends, community members, and cult devotees. This openness helps dissipate the stigma of illness, clearly expresses societal approval for attempts at remediation, and demonstrates recognition of the connection between the private health of individuals and the public health of society. The healing experience often involves altered states of consciousness in both the shaman and the patient. Persons who enter a trance state as part of a healing ritual are relieved from personal responsibility for what they say or do because it is the possessing spirit who is "really" talking or acting. Such a mechanism allows for the expression of thoughts or emotions that otherwise might remain suppressed or repressed. It can bring out into the open any interpersonal and social conflicts that are troubling individuals, families, or the entire social group. Such conflicts have a better chance of resolution when they are brought out in the context of a healing ritual rather than in other forms or private confrontations.

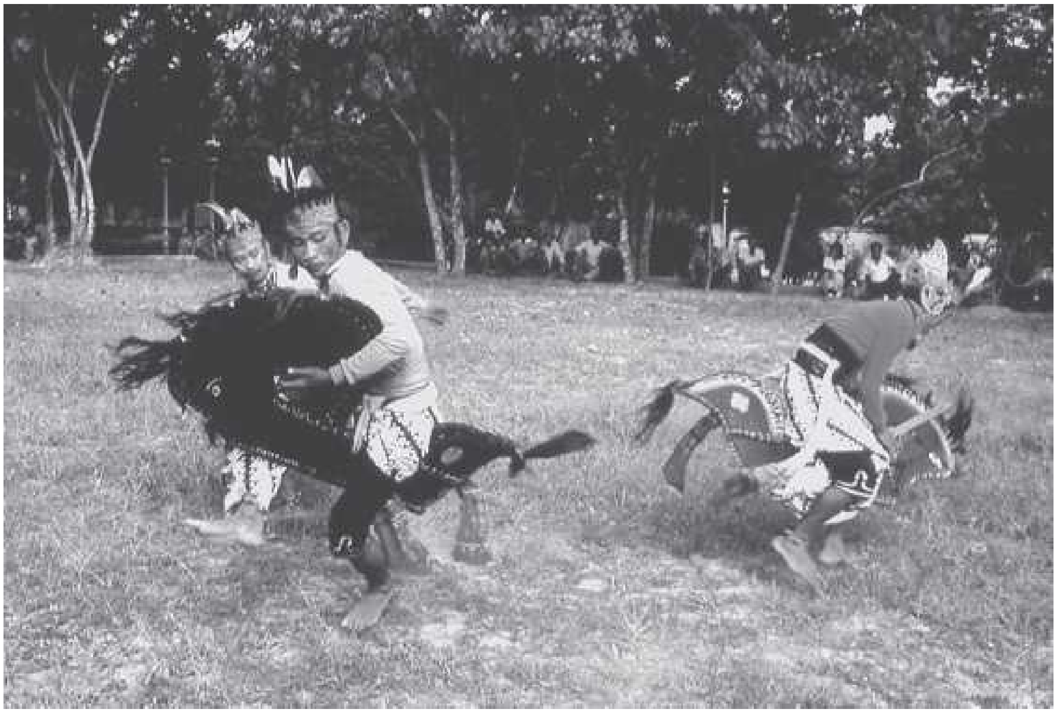

FIGURE 30.7–4. Horse possession dance (*Kuda lumping*) in Java, a ceremony performed for spiritual and health benefits. Participants are possessed by the spirits of horses of historical heroes. In trance states, they act like horses and dance-jump to exhaustion. (Courtesy of Wolfgang Jilek.)

In Puerto Rican *espiritismo,* there are several procedures for removing pathologic spiritual influences. In *despojo* (exorcism), the mediums may fumigate patients with cigar smoke and stroke their necks. In addition to attracting benign spirits to the patients, the mediums may also attempt to draw off the evil spirits into themselves. *Trabajando la causa* (working the cause) is a technique used when special pathogenic spirits have been identified; it is the treatment of choice for sorcery. Possessed mediums interrogate the patients' spirits and try to convince them to depart. *Dando la luz el spirito* (giving light to the spirit) is a technique in which a patient follows a set of rituals to remove a spirit's harmful influences, such as purchasing flowers or candles, performing specific therapeutic tasks, and attending meetings at the medium's *centro. Desarollo* (spiritual development), which is indicated for persons with dissociative symptoms, is a process by which patients are taught to display their symptoms at culturally and ritually appropriate times and places.

Many religions allow for possession by both benign and evil spirits. In Christianity, for example, possession by the Holy Spirit is a highly esteemed experience, especially among Pentecostals, in whom it may become manifest in the gifts of speaking in tongues and prophecy. The exorcism of harmful possessing demons is prominent in the New Testament. Nearly 20% of the Gospels are devoted to Jesus' healings, with the commonest ailment being illness caused by demon possession. According to Christian teaching, he passed on this ability to all baptized Christian believers: "And these signs will follow all who believe; in my name they will cast out demons; they will speak in tongues; they will take up serpents; and if they drink anything deadly, it will not hurt them; they will lay hands on the sick, and they will recover" (Mark 16:16–18).

Almost all religions have rituals for exorcising demons. Exorcisms were common in the early centuries of Christianity and then fell to a low level until a renewed interest in healing emerged at the turn of the 20th century, especially in the U.S. Midwest. The Catholic Church retains its formal, infrequently used rite of exorcism, although one or more simple exorcisms are routinely pronounced in the sacrament of baptism. Fundamentalist ministers perform exorcisms when they perceive evidence of often-dramatic demon possession. Televangelists have become more sophisticated in their approach; they are more apt nowadays to speak of demonic *oppression,* especially in persons with impulse control problems such as alcohol and substance abuse, self-injury, gambling, bulimia, and promiscuity.

Psychedelics

Psychedelics have been used throughout the world for over a millennium as a portal into the world of spirituality. Drugs such as LSD (lysergic acid diethylamide), DMT (*N,N*-dimethyltryptamine), peyote, ayahuasca, and psilocybin are called hallucinogens, although they rarely cause hallucinations. In the lay press, they are more properly called psychedelics ("mind-manifesting"). The new name for these drugs, especially among those who use them, is entheogens (derived from the Greek word for "God within").

The psychedelic experience can range from mild to very intense, depending on the dose of the drug and, perhaps more importantly, on the psychological expectations of the user and the social setting of the drug use. The experience, or "trip," involves visual, auditory, and time distortions as well as synesthesia. At its peak, the drug users have a loss of ego with feelings of merging with space and objects or the universe. They may feel a sense of reaching to the beginning or end of space and time. The psychiatrist Daniel X. Freedman, a pioneer in psychedelic research, described the portentousness of the LSD experience: "The capacity of the mind to see more than it can tell, to experience more than it can explicate, to believe in and to be impressed with more than it can rationally justify, to experience boundlessness and boundary-less events, from the banal to the profound." The use of psychedelics may alter peoples' concept of reality, change their views on life or death, and challenge or promote their most passionately held beliefs. Like spiritual transcendence or ecstatic religious rapture, the peak LSD experience cannot be fully described verbally. However, when "trips" go sour, the experience can be devastating, nihilistic, fearful, and filled with demonic visions.

Psychedelics have come a long way from outright rejection to skepticism to being embraced by a diverse set of multidisciplinary researchers. In the 1970s, the FDA classified psychedelics as controlled schedule I drugs, accessibility was severely limited, and research stalled for almost 20 years. As more results came out suggesting extremely positive and clinically meaningful outcomes in those treated with psychedelics for a variety of psychiatric disorders (anxiety, depression, OCD, etc.), funding became more accessible. As a result, the research on psychedelics has exploded in recent years. They are being studied as novel psychotherapeutics. Notably, neuroscientific studies have indicated that psychedelics operate through pathways in the brain that influence immunity and neurotransmitter systems. In a blinded study, 36 volunteers with spiritual or religious learning who had never taken a psychedelic drug were administered alternate doses of psilocybin and methylphenidate for an 8-hour session in a comfortable room. Based on responses to a Mystical Experience Questionnaire, 60% had a mystical experience, and 70% rated the experience among the top five in their lives, on a par with the birth of a child or the death of a parent. The effects persisted for 16 months and were confirmed by persons who knew the subjects intimately. Unpleasant sessions manifested by short-term anxiety and troublesome personal struggles were reported in 22% of subjects, some of whom nonetheless found the experience meaningful. More recently, a considerable interest in psychedelics for the treatment of Posttraumatic Stress Disorder (PTSD) has gathered momentum, and psilocybin is being envisioned as one of the most efficacious modes to alleviate stress-related symptoms. These have shown a positive and significant treatment effect in treatment-resistant disorders.

Psychedelic-assisted psychotherapy is another modality linked with this class of substances and is gaining prominence. It entails providing psychotherapy sessions in conjunction while patients are receiving psychedelics. It is postulated that, during this time, the patient is more open to getting in touch with their innermost self and open to authentically experience and be much more engaged in therapy. Different psychotherapeutic modalities may be employed, and a supportive form of therapy is popular. Individuals often describe their experience as a life-changing, effective, and transcendental connection with Self.

Drugs

Cannabis enjoys an important place in several religions for a variety of reasons and is considered an entheogen, a substance used to attain spiritual bliss. It has been used to achieve spirituality since ancient times and is still used in more ritualized fashion during religious ceremonies. It is consumed as a drink called *bhang* by many Hindus during the festival of Holi, and it is smoked by devotees of Lord Shiva. Although the Qur’an forbids wine, most Muslims interpret this to include all drugs that affect consciousness. Sufis are known to have used cannabis for spiritual reasons. The Rastafari religion, which emerged in Jamaica in the 1930s, equates cannabis with the biblical Tree of Life and encourages its use sacramentally as a means of discovering the truth and getting close to JAH (God). Rastafari claim that the drug must be used regularly in order to achieve cosmic consciousness and to rid their hearts of corruption. Its use is often associated with Bible study and with truth-seeking group discussions. The Ethiopian Zion Coptic Church, a small religious movement incorporated in Florida in 1975, equates cannabis with the Christian Eucharist (consecrated communion wafer). In the 1960s, marijuana became almost synonymous with the "Hippie movement," and it was described as a way to achieve spiritual peace and attain a transcendental state of bliss. With the legalization of cannabis in several states in the United States, it is more freely available. It will be some time before we know if the patterns and epidemiology of cannabis use have changed.

Alcohol has long had spiritual connotations. Many Greeks worshipped the god Dionysus, who supposedly invented wine. At sacred rites, better known by their Roman name, the bacchanalia, his religion was characterized by the drinking of wine, which then released almost bestial emotions. Women were especially attracted to the chaste orgies, where they were free to forget their domesticity, dance ecstatically in the darkness of the mountains, and transcend human cares and fears. With alcohol-driven superhuman agility and strength, they ripped apart wild animals and devoured them sacramentally, believing that each bloody morsel contained a bit of Dionysus himself. The wine cult spread rapidly to areas as far apart as India and Britain. It also took hold in Palestine and Asia Minor, where, according to some scholars, the Dionysian orgy evolved into the Canaanite Feast of Tabernacles.

The Bible refers to wine as "the blood of the grape," and in its most revered transformation, it signified the blood of Christ. At the Last Supper, Christ instituted the Eucharistic sacrifice, in which he offered his disciples bread and wine that he declared to be his body and blood. For many Christians, this is of prime importance and serves as a reminder of His death and resurrection. By remembering Jesus Christ's sacrifice, people urge Him to lead them and provide direction and safety in their lives.

The psychologist William James wrote of alcohol's intoxicating power to stimulate human nature's mystical qualities and provide a core religious feeling of transcendence and a sense of meaning in life for alcoholics. Alcohol's physiologic power to transform consciousness can, depending on the context, be interpreted as a transcendence of everyday life, an entry into the cosmic world of spirituality. Alcohol fosters an experience of trust, relatedness, and unity through an illusory closeness to others. In some ways, alcoholism is religion gone awry, a metabolic error of the religious impulse. Over time, as alcohol dominates the lives of alcoholics, it replaces a higher spiritual power with itself, and alcoholics turn to it for redemption and the alleviation of guilt and insecurity.

Body Modification and Mortification

Nonmedical alteration and mortification of body tissue, such as tattooing, flagellation, and piercing for spiritual reasons, have a lengthy history. Pope Benedict XIV (1675–1758) wrote: "With the exception of the martyrs, the Church venerates and gives the sanction of her authority to the sanctity of those only whom she finds to have been zealous in the modification of the flesh and senses." One such person was a mystic Heinrich Suso (1300–1366) who, naked, beat himself with a spiked scourge until blood poured off his arms, legs, and torso, and bits of his flesh were ripped from his body. Gazing at his beaten body, he was reminded of the beatings endured by Christ. Another was St. Catharine of Siena (1347–1380), who ate very sparingly, whipped herself with chains for 90 minutes three times a day, and deliberately scalded her body. After drinking the pus that oozed from the breasts of a cancerous woman, she said that she was visited by Christ, who invited her to drink the blood from his wounds.

Flagellation was a common Christian practice. In the 14th century, for example, itinerant bands of self-flagellators were received as heroes by a public who prayed that the bloody penitents would protect them from the plague. Mild flagellation is today practiced during prayer by some strict monastic orders and, along with brief crucifixions, by lay penitents in the Philippines and Latin America. In Islam, the death of the martyr Husain ibn Ali, Muhammed's grandson, is memorialized yearly by Shiites. On this day, some mourners beat themselves, often severely, with chains and spiked clubs while recalling his words, "Trial, afflictions, and pains, the thicker they fall on men, the better do they prepare him for his journey to heaven."

Within the Eastern Orthodox tradition, the sect known as the Skoptsi (eunuchs) included castration as a central religious practice. Its founder believed a medieval legend that Adam and Eve were born sexless. After they ate the forbidden fruit in the Garden of Eden, God supposedly grafted the halves of the fruit onto Adam as testicles and onto Eve as breasts. Only through castration and removal of the nipple or breasts in women could believers return to the pristine state of purity that existed before original sin. Despite severe repressive measures by the Russian government, the sect persisted for several hundred years until the turn of the 20th century.

Among Hindus, the yearly festival of Thaipusam commemorates both the birthday of the deity Lord Murugan and the day he was given a *vel* (lance) to conquer an evil demon. The highlight of the festival is a brief pilgrimage to a holy site before which devotees, after much preparation, enter a trance state and then pierce their skin, tongue, and cheeks with a *vel.* Because Lord Murugan's conveyance is a peacock, some persons carry a *kavadi* (somewhat of a portable altar) that may weigh 100 lb on their shoulders. The *kavadi* is festooned with peacock feathers and is attached to the carriers by hooks in their backs and chests. The festival, a joyous event held in several sites, is growing in popularity. Those who pierce themselves not only give thanks to the deity but also accrue spiritual merits. In Islam, some Sufis practice potentially dangerous forms of body piercing to achieve a spiritual state.

Fakir Musafar, the adopted name of a Caucasian South Dakotan, is a primary founder of the Modern Primitive movement. He has devoted his life to understanding the spirituality of body modification in rituals throughout the world and putting this knowledge into practice. He is a shaman of sorts who has established a licensed school for body piercing and modification in California in which he indoctrinates students into the spiritual underpinnings of his craft.

Confession, Penance, and Reconciliation

Sins are offenses against the deities, ancestral spirits, and certain taboos in many religions. Sins, especially grave ones such as blasphemy and murder, separate people from God and destroy the charity and love of God in their hearts. Only by confession and reconciliation with God can sinners avoid divine punishment, sometimes in this life or certain eternal damnation in the afterlife, restore a right relationship with God, and regain spiritual purity. In many religions, atonement is achieved by participation in holy rituals. For example, Christians believe that atonement comes from the sacrificial crucifixion of Jesus Christ, which washed away the sins of the world and allowed humankind to be reconciled with God.

Visiting Spiritual Sites

Every religion and culture has designated sites considered sacred and whose spiritual forces may affect the lives of visitors. The sites may be churches, temples, or natural landmarks such as mountains, rivers, or rock formations. The Temple Mount in Jerusalem is one of the holiest places on earth. The Church of the Holy Sepulcher has been a pilgrimage destination for Christians since the fourth century, who believe that it contains the sites of both Jesus' crucifixion and burial. Also on the Mount is the Islamic shrine, the Dome of the Rock, which Muslims believe is the place from which Muhammad made his Night Journey to visit Allah in heaven. This site was built over a Roman temple devoted to Jupiter and, before that, the Jewish Temple. The rock or Foundation Stone in the center of the Dome is thought by Jews to be the place where the patriarch Abraham was tested by God to see whether he was willing to sacrifice his son, Isaac.

Charity and Altruism

In all the major religions, charity is considered the greatest or among the greatest of all spiritual values. Its fullest explication in Christianity is found in Paul's First Letter to the Corinthians. Paul used the Greek word *agape,* which denotes the purest and most absolute level of unconditional love. The word also refers to the theologic virtue by which people love God above all things for his own sake and love themselves and their neighbors for the love of God.

The love of one's neighbors, a highly spiritual virtue, can be manifested physically by engaging in acts of kindness and providing both comfort and material things to the less fortunate. The Salvation Army, for example, aims to save people's souls but provides shelter, food, and clothing to the poor. Giving alms is one of the Pillars of Islam. Hindus give food to sadhus and Buddhists to monks. Such giving enhances the spirituality of the giver. Buddhist monks may demonstrate against unjust political rulers by walking with their begging bowls face down, thus denying their oppressors the opportunity to give them food.

For some individuals, irrespective of their religious affiliation, altruism serves as a means to finding a deeper meaning and purpose in life. Altruism is a secular term referring to a selfless obligation of benevolence, respect, and caring for humankind, all living creatures, and the environment. Philanthropists and those who selflessly give food and clothes to those in need, preserve and improve the environment, shelter the homeless, care for the sick, protect the defenseless, assist the weak, and enlighten the uneducated achieve a sense of peace and the belief that they have accrued spiritual merits for their good deeds.

New Age Spirituality

The term New Age refers to a cultural movement that began in the late 20th century with multiple roots, beliefs, and concerns that vary widely and include aspects of the following: mainstream religious traditions, Sufism, mesmerism, shamanism, Kabbalah, occultism, neopaganism, theosophy, magic, karma, unidentified flying objects (UFOs), meditation, psi energy, astrology, synchronicity, numerology, fortune-telling, parapsychology, ecologic responsibility, guardian angels, cosmic forces, yoga, fasting, contact with aliens, near-death experiences, levitation, dream interpretation, and Wicca. This list is hardly inclusive but does give a sense that New Age spirituality has no sacred texts, central organization, membership, clergy, central headquarters, dogma, or creed.

New Age believers are generally optimistic and eager to enhance human potential by tapping into so-called spiritual vibrations and powers that emanate from the cosmos, crystals, pyramidal objects, and geographic locations such as Sedona, Arizona, where the Center for the New Age employs psychic readers and energy healers specializing in spiritual quests and inner transformation.

A major result of the New Age movement is the development of a thriving spiritual tourism industry. Using a computer in the comfort of one’s home, it is possible to book spiritual pilgrim tours anywhere in the world. Thus, one can seek spirituality among the Kung Bushmen of the Kalahari Desert, Australian aborigines, Amazonian shamans, or at exclusive spas in Bali or ashrams in India. Such tours often may dramatically affect the lives of participants.

Spirituality and Yoga

Yoga is a form of exercise consisting of performing a variety of bodily postures while being aware of your breath, your thoughts, and being in the moment. It is often described as a practice promoting a deeper mind–body connection leading to a profound sense of peace and well-being. Yoga is seen as a way to bring harmony and closeness between an individual's emotional and physical self. There are a variety of forms and routines rooted in the healing process and promoting mindfulness about one's own self in relation to the world outside. Yoga is increasingly being studied for its significant beneficial outcomes in chronic physical and psychiatric illnesses such as depression, PTSD, chronic pain, and pervasive anxiety. Yoga has universal appeal, regardless of religion, attributed to its association with spiritualism and meaningfulness.

SPIRITUALITY AND RELIGION IN CLINICAL PRACTICE

Spirituality can be a positive force in peoples' lives. In 1984, the California Psychological Association's Task Force on Spirituality and Psychotherapy provided a new definition of spirituality as "The courage to look within and to trust what is seen and what is trusted appears to be a deep sense of belonging, of wholeness, of connectedness, and openness to the infinite." The psychologist David Elkins outlined the values of spiritual persons. These include beliefs in "a greater self" or a personal God; a sense of purpose in life and a quest for meaning; acceptance of the sacredness of nature and all human experience; knowledge that ultimate fulfillment is found in spirituality and not in material things; altruism; idealism; awareness of suffering and death; and leading a life that has a positive effect on people, nature, and their relationship with whatever they consider to be the ultimate and transcendent reality. This is an impressive list, but it is a challenging task, for example, to reconcile spirituality in the sacredness of nature and the occurrence of devastating tsunamis, tornados, volcanic eruptions, earthquakes, congenital malformations, and stillbirths unless they are considered as part of God's great plan for humankind, which includes all the wonderful and horrible events of the past and present or as tough-love warnings by Mother Nature to tread lightly on the Earth.

The list of spiritual values would also seem to apply to religious values, perhaps the closest being those of Buddhism. However, in the case of

Christianity, for example, the list would need to contain specific beliefs about Jesus Christ as the savior of humankind, personal salvation as the ultimate goal in life, participation in church rituals, and acceptance of the Bible as the only true sacred text of the only true religion. Consideration would then have to be given to sectarian beliefs such as obedience in spiritual matters to a church hierarchy, reservation of the priesthood for men, a wife's duty to be obedient to her husband, and denial of homosexuality as an acceptable form of human personhood.

Religion, Religiosity, and Spirituality

Religion, religiosity, and spirituality may have some overlapping dimensions, but the concepts are not interchangeable. Spiritual persons may neither participate in religion nor possess religiosity, or they may participate in religion but lack religiosity, or they may participate and have religiosity. Other permutations are possible; for example, people who attend church may be spiritual but lack religiosity, or people may possess religiosity but not attend church, and so on. Great confusion results when these distinctions are not made. Thus, in many studies, "religion" is measured by participation in religious services at a church, mosque, synagogue, or temple. However, some people attend only a few special services each year, whereas others may call themselves a Christian, Jew, and so on, and consider themselves to be a member of a religion even though they have not attended services for years.

Religiosity refers to peoples' devotion to their religion; Allport and Ross's Religious Orientation Scale, for example, contains the statement, "My whole approach to life is based on my religion." An affirmative answer ideally implies a sincere commitment to one's religious beliefs, which are internalized and serve as the guiding motivation of behavior. In reality, religiosity is not an all-or-none phenomenon. More than 20 variables have been identified as components of religiosity, and people may score high or low on different variables. Religiosity usually does conform to many spiritual values, but it may be spiritually destructive, for example, in the case of cults that prey on children and vulnerable persons who are disillusioned with life, frustrated by failed relationships, lonely, or experiencing a sense of disorientation that may be apparent, for example, in newly arrived college students, foreigners, and runaways. Destructive cults use deceitful recruiting practices and masterful indoctrination programs that, often in the course of several days, can cause persons to abandon their autonomy and conform to the groups' behavioral norms in exchange for a false sense of well-being and exclusivity. Once indoctrinated, cult members lose their cognitive flexibility and adaptability; demonstrate a narrowing, blunting, or artificiality of affect; and develop a childish dependence on cult leaders to make decisions for them. In the worst scenarios, cults may use perverse religiosity to justify their sexual abuse of children, as was the case with the Branch Davidian cult in Waco, Texas, whose compound was destroyed and many of its members (including 21 children) killed during a raid by federal agents in 1993.

In an effort, perhaps to counterbalance the overwhelming focus of medicine on biology and biotechnology, it is not surprising that a host of the major health care, medical, graduate medical education, nursing, and social work organizations have jumped onto the spirituality bandwagon. Most medical schools have courses on spirituality and health, and residents in psychiatry, family medicine, and internal medicine are required to learn about spirituality. The movement has gained traction through numerous correlational and cross-sectional studies that make unwarranted conclusions about the effects of religion and prayers on medical and psychiatric illnesses.

The lay press has publicized widely exaggerated claims about the "God spot," an area in the brain that supposedly demonstrates the biologic basis of spirituality. Neurologists have long known that patients with temporal lobe epilepsy may develop a newfound religiosity (often obsessive in nature), have transcendent experiences, and experience religious auditory and visual hallucinations. Some have speculated that St. Paul's "thorn in the flesh" was temporal lobe epilepsy, which facilitated his conversion to Christianity on the road to Damascus. In fact, before Hippocrates, epilepsy was known as "the sacred disease." The controversial psychologist and countercultural icon of the 1960s Timothy Leary championed the use of LSD for its therapeutic and spiritual benefits; he wrote a book titled *Your Brain Is God.* Recent functional magnetic imaging studies have shown that a large number of neural circuits involving the frontal, prefrontal, temporal, and cingulate cortices; the caudate nucleus; and the parietal lobes "flickered into life" when 15 Carmelite nuns reported feeling the presence of God. The psychiatrist Herbert Benson stated, "To put it simply, we're wired to God." Albert Ellis, the psychologist who developed rational emotive therapy, retorted, "The whole field is off its rocker." Benson reasoned that because faith in a medical treatment can be therapeutic in up to 90% of the most common medical problems, then “faith in an invincible and infallible source carries even more healing power.” In other words, belief in the healing power of God is the greatest placebo of all. Benson is probably correct in that this belief engenders hope, which is no small thing for people who feel hopeless.

At a conference sponsored by the Association of American Medical Colleges in 1999, a consensus of clinicians, medical educators, and chaplains agreed that spirituality involves a person’s search for ultimate meaning in life through participation in religion, as well as through relationships with God or family, and is expressed in nature, rationalism, humanism, and the arts. This declaration is somewhat idealistic in that very few persons lead an examined life, much less search for ultimate meaning. Sages from the West, wise men from the East, prophets, philosophers, poets, priests, psychiatrists, and Ph.D. candidates have all struggled to find the ultimate meaning of life. Multiple answers have been proposed—to have offspring and occupy and subdue the Earth; to act justly, love mercy, and walk humbly with God; to glorify and worship God; to love one’s neighbor; to defeat one’s enemies; to experience oneness with the universe; to be content with cultivating one’s garden; to experience pleasures; to seek truth and be wise; and so on—but no single answer suffices for everybody.

From a global perspective, most people lead simple, unquestioning lives that go from paycheck to paycheck. Sports, sex, movies, children, and television shows are more apt to be topics of concern and conversation than transcendence or the meaning of life. The consensus definition of spirituality is somewhat high-minded, although it states that spirituality “may be expressed” by listening to music or painting, for example. Does listening to rap music qualify, or must it be mostly Mozart? Does painting-by-the-numbers count? What about relationships with pets?

Because the word “spirituality” is used in many different contexts, a distinction should be made between *Spirituality* with a capital S and *spirituality.* The former refers to a unique and very special experience as defined in the following discussion, whereas lower-case spirituality is a more general term that refers to meritorious, estimable, and honorable behaviors and feelings that benefit both individuals and the common good.

Based on this distinction, Spirituality refers to the following:

Spirituality is a feeling of a special connection with a person, a group, any living thing, nature, a cause, sacred or culturally significant objects or activities, the universe, or a supernatural being such as a deity. This feeling is special because it penetrates to the perceived inner core of a person's being and may evoke deep passion, intense love, joyful bewilderment, visionary experiences, or a sense of profound understanding that cannot be expressed adequately in words. The settings where this feeling occurs are highly variable, as are their degrees of intensity. Traditionally, the most exalted and intense feelings are those of mystical transcendence and understanding the true nature of reality and the ultimate meaning of life. However, equally valid and more commonly experienced are the mundane feelings of connections such as wonderment in the beauty of nature and reverence for sacred landmarks; sexual merging with a loved one; caring for children, grandchildren, or a faithful pet; working for charitable causes; cheering for a favorite sports team; and maintaining good relationships with friends, family, and neighbors in good times and bad. A person may seek after and prepare for this feeling, but it almost always comes spontaneously and lasts for a moment or, in exceptional cases of mystical transcendence, for up to 30 minutes. Although this feeling is transitory, it may have enduring effects. It can be a positive force in coping with the travails of daily life and a sustaining force, particularly in times of physical and emotional crisis and in old age, especially when the specter of death is on the horizon.

The adjective "spiritual" in everyday usage usually refers to spirituality rather than the specific feeling of Spirituality. Many persons may consider themselves to be spiritual or to lead spiritual lives because they follow the Ten Commandments and practice the behaviors specified by Jesus in his Sermon on the Mount; attend religious services regularly; pray often; fulfill the requirement of the Five Pillars of Islam; follow the Noble Path; live in a cloistered environment; evangelize for their faith; make pilgrimages; participate in charitable, patriotic, and ecologic causes; remain faithful to their marriage vows; provide for the well-being of their children, parents, and pets; accept people for the content of their character; forgive those who have harmed them; appreciate the beauty of nature; serve as positive role models; and, in certain contexts, attempt to influence otherworldly beings on behalf of humankind. This list is not exhaustive but provides a sense of the attributes of spiritual persons and spiritual lives. Spiritual persons who lead spiritual lives may never experience Spirituality as defined here but may approximate and even share some portion of it. If the Virgin Mary's love for her child Jesus was spiritual, then there is some measure of Spirituality in every mother's love for her children. If the Native American spiritual belief that all things are connected is true, then there is a potential spirituality in all the thoughts, emotions, and deeds of people's lives. Evangelical, Fundamentalist, and Pentecostal Christians claim to have a personal conversion encounter with God in which they are born again. The Biblical book of 2 Corinthians states, "Therefore, if anyone is in Christ, he is a new creation; old things have passed away; behold, all things have become new." When this conversion moment is intense and occurs suddenly, the experience may be one of true Spirituality or, at the least, a close approximation.

Unfortunately, some people who consider themselves spiritual can be misguided and create mischief and even havoc when they try to force their beliefs and values on others in the name of a cause or a religion. Much suffering can be caused by those who believe in their own righteousness. The austere Puritans, for example, escaped religious persecution by fleeing to America from England in the mid-16th century. Historians have characterized them as persons who were distressed by the thought that some people, somewhere, were enjoying themselves.

Psychiatrists' Perceptions about Religion and Spirituality

Surveys of physicians in general and psychiatrists, in particular, have always reported that they have both a lower level of belief in God and participation in religious activities than the general population. Although involving only 100 psychiatrists, a recent survey also showed a lower level of positive attitudes toward Spirituality, which was defined in terms of religion and feelings of connectedness with an entity greater than oneself that gives meaning to life. This is certainly not anything new: a medieval aphorism was, "Three doctors, two atheists." These attitudes have probably resulted from physicians' experiences in watching so many good people and innocent children die. It takes great effort to reconcile these experiences with the will of a just and merciful deity or to find in them any transcendence. Sigmund Freud contended that religion was an infantilizing, neurosis-producing, tyrannical illusion, and most psychoanalysts followed his critique, with the prominent exception of Carl Jung, who, for example, believed that every person's mind contained unconscious prehistoric memories of the real Father–God who is perceived as one’s biologic father, contrary to Freud’s belief that it was the other way around. When psychoanalysts had something positive to say about religion, they often managed to find a rankling counterbalance. Ernest Jones, for example, wrote about “the enormous civilizing influence of Christianity” and “sublimated homosexuality” in the same sentence. Even Henry Maudsley, the most influential British psychiatrist of his time and certainly not a Freudian, wrote in 1918 that “the corporeal or the material is the fundamental fact—the mental or spiritual only its effect.” In 1975, the outspoken psychologist Albert Ellis declared that religion itself is a self-deprecating, dehumanizing mental illness. These attitudes have probably attracted some people to gravitate toward psychiatry as a career.

The change among psychiatrists to a more positive view of religion began in the 1960s and 1970s. The psychiatrists William Meissner and Ana-Marie Rizzuto each proposed a nontheologic and easier-to-accept understanding of the human experience of God through the life cycle. A child’s initial image of God is based primarily on the image of their parents. The child’s God may be loving and protective, as well as fearsome, cruel, and not always available to gratify the child’s wishes. A positive mother–child experience at this stage results in a basic sense of trust that, in later years, may lead to a trusting faith in God, whereas a negative, insecure experience may distort the experience of God. As the child begins the process of separation and becoming an individual with some independence, it hears about something called God but lacks the capacity to understand what a spirit or transcendent force is supposed to be. The child hears that God has made the world and, thus, likens God to the most powerful person that they know, namely, both parents, but especially the father, who tends to be the more forceful or aggressive parent. Over the next few years, as the child comes to recognize that their parents neither know everything nor are all-powerful, a romantic view of a perfect, heavenly father may emerge, as well as a distorted view of the family. God and parents may be idealized as totally protecting, or there may be a split into a good God and mean parents or into good parents and a mean God. Between 6 and 11 years of age, the child begins to appreciate symbols. The concept of God shifts from that of a flesh-and-blood person to a still imperfectly understood spirit whose power must be revered and feared because God exacts punishment for every misdeed. In this stage, an immature obsessional religiosity may emerge in conjunction with ritualistic behaviors.

During adolescence, God becomes more personalized as “my Savior” or “my Father” and more connected to emotions and subjective attitudes, such as love, obedience, trust, and fear. God is experienced as the ideal leader and advisor who patiently and kindly listens to one’s innermost problems and desires. When not tempered by some maturity of judgment, this realization of God may lead to fanatical tendencies. As adolescents become moralists, the burden of sinfulness emerges, and they may engage in the extremes of loose or puritanical behavior. Rebellion against parental authority may carry over to the parent’s religion, as well as to the God whom the parents symbolically represent.

As adolescents step into adulthood, they must decide whether they will follow their parents’ religion, seek a different religion, or avoid religion altogether. A study undertaken by Michaelson et al. (2016) examined the evolution of spiritual health in adolescents in six countries, including Canada and the United Kingdom. Results indicated that as adolescents advanced toward young adulthood, they continued to highly rate the importance of connection with self while their perceived importance of connection with others decreased. The authors attributed this to the development of a more independent style of thinking, challenging the traditional beliefs they grew up with, and an increased capability for abstract thought. For some, the inner representation of God remains solid; for others, it may change; and for still others, it may be rejected or replaced. Thoughts about the memory of God usually return in old age, especially when death nears. A person may decide to embrace God gracefully or grudgingly, for reasons that include mature acceptance of the cycle of life, the hope of entering a heavenly afterlife, avoidance of eternal damnation, and neurotic fear, or to spurn God’s ultimate encroachment; a person’s last breath may be an inspiration or an expiration.

Freud held that belief in God was an illusion that softens fears about the dangers of life; that the moral world order that God promises instills hope that justice will be served; and that belief in an afterlife creates an ultimate setting for the fulfillment of wishes. For Freud, this illusion was noxious because it distorted reality and promoted a state of psychic infantilism. However, in some modern psychiatric thought, God is a benevolent illusion necessary for the growth and nourishment of mental and spiritual health. God is like the baby blanket or teddy bear that is special to the infant who endows it with a unique quality, cuddles it lovingly, and sometimes hates it. As the child matures, these transitional objects are neither mourned nor forgotten but rather simply lose meaning. They are the stuff of illusion and exist mentally somewhere between subjective and objective reality. In this formulation, God is a special transitional object that does not lose meaning totally and is always available through religion. Rizzuto concluded that a person's humanity depends on illusion; the type of illusion selected, be it science, religion, or something else, "reveals our personal history and the transitional space each of us has created between his objects and himself to find a resting place to live in."

Another factor that began to emerge in the 1960s and 1970s was that psychoanalysis was giving way to the community mental health movement, which decreased the traditional role of psychiatrists as therapists. Nowadays, the majority of psychiatrist–patient interactions involve 15-minute medication checks. In such an environment, psychiatrists usually have little trouble calling on the clergy and religious institutions to help with patient care, for example, by providing relief from social isolation through attending religious activities, advising about life issues as well as spiritual ones, and giving emotional and material support. Some clergy and religious institutions are quite amenable to including persons with mental illness and mild mental retardation in their congregations, whereas others may be less inclined to do so. Psychiatrists need to be aware that the clergy, a group with diminishing members, has little to no training in providing mental health services. Their major priorities are saving souls and turning sinners toward holiness. Certainly, there are pastoral counselors, but their numbers are few, and they are most likely to be called on in hospital settings in times of impending death or major medical crises.

Undoubtedly, some psychiatrists harbor conscious and unconscious negativity toward religion and religious patients. This disdain may stem from multiple sources, such as contact with or reports of pederastic and philandering clergy; exploitative cults; greedy televangelists; anger toward God over the premature death of a loved one; personal guilt (the Semitic religions are quite effective in stirring this emotion); the belief that religion is an irrational and coercive force in peoples' lives; negative experiences while attending a religious grade school; and so on. There are also undoubtedly psychiatrists who overvalue the role of religion and religiosity in the treatment of the mentally ill. Such under- or overvaluation distorts the clinical encounter and compromises the care of religious patients, who may detect this bias and skew the information they provide, which may be interpreted as denial, resistance, paranoia, or characterologic psychopathology. It is imperative that psychiatrists be aware of their feelings and beliefs about religion and religiosity to provide appropriate clinical care. The same caveat holds true for encounters with patients whose spirituality is manifested in secular rather than religious beliefs and practices, although it would be rare for psychiatrists to harbor disdain in dealing with patients who love their family and their pets, engage in charitable and altruistic activities, and find peace in natural beauty. Psychiatrists may consider some of the practices of New Age spirituality to be odd, but, with the exception of protracted use of psychedelic drugs, these practices do not carry the same baggage as religious ones.

Spiritual Assessment

For a thorough clinical assessment of a patient, the biopsychosocial paradigm is insufficient, and the adjective “cultural” should be added to include consideration of such factors as ethnicity and spirituality. Several spirituality assessment tools have been published and share many commonalities, such as the (British) Royal Free Interview for Religious and Spiritual Beliefs and the University of Minnesota Module on Taking a Spiritual History.

The initial core question is, “Do you consider yourself to be a spiritual person?” If the answer is “No” or “What do you mean by spiritual?” some appropriate responses include “Are there any things that you feel passionate about or that give a special meaning to your life?” and “What are the things in your life that give you the strength to go on living?” Sometimes, a more dramatic question may be posed, such as “What is it that keeps you from ending your life?” The most common answers, in no particular order, are linked to pets, grandchildren, and God and/or religion.

It should be no surprise that pets are ranked so high; animals are significant in religious history and imagery. Among the ancient Israelites, the Leah tribes were linked to antelopes, the Rachel tribes to lambs, the Manasseh tribe to asses, and so on. Australian aborigines have long revered the totemic kangaroo. The Sacred Jaguar was central to the Aztec and Olmec religions and remains important to Amazonian tribes. Native Americans have animal spirit guides. Ancient Egyptian religion was polytheistic and often zoomorphic; for example, the jackal god Anubis watched over the dead; the falcon-headed Horus was the god of the sky; Menhit was the lion goddess of war, and Wadjet, the cobra goddess, was the protector of Egypt, and her image always appeared on the crown of the pharaohs. Cows are sacred to many Hindus. In Christianity, Jesus is often depicted as the Lamb of God and the Holy Spirit as a dove.

In Western cultures, dogs and cats are the most common pets, and not infrequently, patients state that they could not commit suicide because of the disastrous effect it would have on their pets. Popular fictional animals such as Snoopy the dog, Garfield the cat, and Bambi the deer demonstrate peoples' willingness to accept, be amused by, and cherish animals. Dogs, perhaps because of their fidelity, transparent emotionality, and the seemingly unlimited amount of unconditional "love" that they offer, often stir the passions of their owners: "I would kill anyone who tried to hurt my dog"; "No one understands me like my dog." Pets certainly have a calming influence, and taking care of them does provide some meaning to many peoples' lives. Asking about the status of a pet, even in a brief clinical encounter, often helps with treatment compliance because patients feel that the psychiatrist is interested in more than their symptoms. Encouraging depressed patients to take walks with their dogs exposes them to the beneficial effects of exercise and sunlight.

The birth of a grandchild often changes the relationships between grandparents and their children, especially when these relationships are strained. Grandparents may experience both delight and a renewed meaning in their lives in caring for grandchildren, passing on their wisdom, and providing some respite for weary parents. Intergenerational bonding and continuity of the family partake of the spirituality that exists in daily life.

When patients say, for example, that God helps them persevere and gives meaning to their lives, it is important for psychiatrists to gather more information by responding, for example, as follows: "Tell me about your relationship to God." "Do you believe that God somehow plays a role in your illness?" "Has God ever made his presence felt in your life?" "Has God ever disappointed you?" "Have you ever disappointed God?" "Do you believe that God will be a factor in recovering from your illness?" "How do you stand with God now?" "How do you think that God will judge you?"

It is quite rare for patients to describe themselves to psychiatrists as sinners, a trend that has grown over the last century. Karl Menninger's popular 1973 book *Whatever Became of Sin?* noted that sin no longer is the topic of daily discourse but rather has been transmuted into mental illness and criminality. In addition, in recent decades, the concepts of free will and personal responsibility, necessary components of sin, have been greatly diminished by explanations of behavior that stress genetics, childrearing practices, childhood experiences, chemical unbalances in the brain, defects in neural circuits, exposure to traumatic events, peer pressure, and the influence of advertising. Only in extraordinary circumstances is it appropriate for psychiatrists to ask patients about sin. Patients who regard sin as one factor in their mental illness should be referred to a pastoral counselor to deal with this issue but still may need to receive treatment by a psychiatrist.

When patients state that religion provides meaning to their lives, this needs to be explored. Some appropriate questions are as follows:

What religion do you belong to?
Has religion been important to you throughout your life?
Do you attend church (synagogue, mosque, temple) services? How often? When was the last time that you attended?
Are you satisfied with your church?
Do you feel that the congregation welcomes you?
Tell me about the congregation.
Were you exposed to religion as a child?
Is your current religion the same one you grew up with?
Do other family members attend religious services with you?
Do you discuss religion at home?
What is it about your religion that you find most gratifying?
Does your religion or your religious practices help you cope with stress and difficult times in your life?
Have you ever had any negative religious experiences?
Are there any aspects of your religion that you disagree with?
Do you read Holy Scripture or religious materials? Does this reading influence how you feel or how you lead your life?
If you were to die right now, would you go to heaven or hell?

Most patients tend to elaborate their responses, but if they do not, then more detailed questioning is necessary. It should not be surprising that not all congregations or church leaders are alike, even though they are members of the same religion. Each church, synagogue, mosque, and temple has its unique dynamics, personalities, and ordering of priorities. In some Sikh temples, for example, major confrontations may occur over the practice of eating at a table or on the floor. In highly hierarchical religions, such as the Catholic Church, individual members or a majority of a congregation may choose to disregard pronouncements by the pope or local cardinals or bishops over such issues as birth control and not allowing divorced persons or those who are tolerant of abortion to receive communion. Mosques attended by both Shia and Sunni Muslims may be troubled by internal politics. Congregations may be politically liberal or conservative. Some may be somber, and others joyful in their religious practices. Some may be more tolerant of the mentally ill than others. A thorough spiritual assessment should consider the characteristics of the specific church, synagogue, mosque, or temple that a patient attends.

Spiritual and Religious Problems in Clinical Practice

Asking all of the questions involved with a spiritual assessment in a respectful, nonbiased manner may set a positive tone for the doctor–patient relationship, especially with patients for whom spirituality and religion are important. However, even patients who profess no interest in religion may be intrigued by the idea that spirituality does not automatically imply religion and that even atheists can lead spiritual lives.

If a thorough biopsychosociocultural assessment demonstrates that a patient is not mentally ill but rather has a spiritual or religious problem such as issues about a loss of faith or conversion to a different religion, *Diagnostic and Statistical Manual of Mental Disorders* code V62.89 may be the appropriate diagnosis. Such problems may arise during treatment or may be the chief complaint of new patients who may think that they are mentally ill or who are seeking reassurance that they are not mentally ill. Once a diagnosis is made, a referral to a pastoral counselor should be made, even if the patient remains in treatment for an Axis I or II disorder.

Psychosis, Depression, and Anxiety. As defined in this section, some of the feelings associated with Spirituality are similar to those experienced during a psychotic episode—for example, visions, fleeting profound insights, closeness to God. There is an eeriness to a psychotic episode, however, that may be manifested by turmoil in the patient’s life, by disorders of thought (loose associations, thought blocking, etc.), and often by emotion (mania, depression, etc.) and general appearance (odd clothing, mannerisms, and comportment). When one is unsure whether a patient is psychotic because of religious claims, the patient’s peers—such as church members—should be consulted. They commonly will be able to identify the patient’s claims as either abnormal or within the group’s concept of normalcy. Psychotic persons demonstrating high religiosity and concrete thinking may interpret scriptural passages literally and are in peril of harming themselves (eye enucleation and self-castration are mentioned in the Gospels) and others (killing children during a postpartum depression to “save” them). Claims made by a person who has gone through a rapid conversion from a traditional religion to a cultic or somewhat “exotic” religion, such as Hare Krishna or Druidism, are highly suspect because such conversions are usually associated with emotional tumult and sometimes with psychotic episodes.

However, it is possible for persons with severe, chronic mental illness who are not currently psychotic to experience Spirituality or to lead spiritual lives. In a Swiss study of 100 outpatients with severe, chronic mental illnesses, some subjects reported that religion instilled hope, purpose, and meaning in their lives (71%), although it induced spiritual despair in others (14%). Subjects also reported that religion lessened (54%) or increased (10%) psychotic symptoms, increased social integration (28%) or social isolation (3%), reduced (33%) or increased (10%) the risk of suicide attempts, reduced (14%) or increased (3%) substance abuse, and fostered adherence (16%) or opposition (15%) to psychiatric treatment. Religious delusions worsen the prognosis of patients and are associated with higher severity of hallucinations, bizarre beliefs, lower levels of functioning, longer durations of illness, and the need for higher amounts of antipsychotic medication.

The relationships between depression, religion, and spiritual life are complex. In almost all cultures, severe major depression replete with suicidality, deep regression, marked agitation or immobilization, and total hopelessness has been recognized as an illness. However, the concepts of moderate depression and dysthymia have been considered pathologic conditions only in relatively recent times and have spread from Western cultures to the rest of the world. The number of adjectives that have referents to depression but are used to describe aspects of the human experience is truly vast: For example, dejected, sad, downtrodden, disheartened, downcast, despondent, disconsolate, gloomy, pessimistic, melancholic, despairing, grave, lugubrious, funeral, lamentable, sullen, grim, and so on.

All persons experience both sadness and happiness in their lives, but the former tends to be more persistent and is usually interpreted as more meaningful; thus, Oscar Wilde's epigram, "Where there is sorrow, there is holy ground." The most iconic image of Western civilization is the image of Jesus on the cross, bloody and beaten. The New Testament promises Christians suffering and persecution and urges them to follow Christ's example by sharing in his afflictions. Unlike gladiators, who glistened with oil and fought for their lives, martyrs aggressively sought a passive death —the bloodier, the better. In an impassioned tract written during the worst days of Roman persecution, the theologian Tertullian boasted,

> Some Christians have been tested by the fire, some by the sword, some by the beasts; some, lashed and torn with hooks have just tasted martyrdom, and lie hungering for it in prison…. Nothing whatever is achieved by each more exquisite cruelty you invent; on the contrary, it wins men for our cause. We are made more often as you mow us down. The blood of Christians is seed (semen est sanguis Christianorum). (From Glover TR. *The Conflict of Religions in the Early Roman Empire.* Vol. 94. 1909:324.)

Over the centuries, many Christians willingly accepted suffering for Christ's sake. Perhaps the greatest devotional manual of Christianity, *The Imitation of Christ,* written in the 15th century by Thomas à Kempis, is based on the principle of self-renunciation. Many of the great mystics have recounted long periods of despair, dryness of spirit, terrible darkness, and questioning of their faith in their quest for holiness.

The ancient notion that suffering is a just punishment for sins is found in the Hebrew scriptures, but some of its writers were aware that this is not invariably the case. Thus, Jeremiah (12:1) pleaded with God to answer, "Why is it that the wicked live so prosperously? Why do scoundrels enjoy peace?" The answer that emerged was described by the theologian John Bowker as "perhaps the supreme contribution of Israel to a human response to suffering, that suffering can be made redemptive, that it can become the foundation of better things, collectively, if not individually."

In Hinduism, the deities often represent opposite and complementary qualities. Siva personifies the forces of both creation and destruction in the universe and must be propitiated on the basis of the relationships of the devourer, devoured, and devouring. Kali, the Dark Goddess, is a bloodthirsty aspect of Siva. Her teeth are fangs, and she wears a necklace of skulls and a belt of amputated limbs. Her stomach is a void that can never get enough blood from beheaded offerings in her temples. Human sacrifices to Kali were abundant until forbidden by law in 1835.

However, because Siva also represents the force of creation, suffering must be put into perspective as the connection between the two conflicting universal principles, the urge to life and the urge to death. Suffering is inherent to the tensions of the cosmos and is a significant part of the human experience. It is the process of being devoured, the better to be free from worldly attachments in one's search for transcendent spirituality. Sadhus are so revered in the towns and villages of India because they allow themselves to be devoured by living a life of asceticism and of twisting their wasted bodies into odd positions, sometimes for years at a time. It is this suffering that creates true liberation from the bondage of the flesh.

Buddhism is the belief system that focuses most directly on suffering. Indeed, the First Sermon of the Buddha at Benares contained the essence of his wisdom. In this sermon, Buddha disavowed both devotion to bodily pleasures and self-mortification. Instead, he preached a middle path that gives knowledge and leads to calmness, insight, enlightenment, and Nirvana. To reach this path, a person must comprehend the Noble Truth of suffering: Aging is suffering; sickness is suffering; death is suffering; sorrow, lamentation, pain, grief, and despair are suffering; association with the unpleasant is suffering; disassociation from the pleasant is suffering; not to get what one wants is suffering; all types of attachment are suffering. The origins of suffering are cravings for bodily pleasure, for existence and nonexistence, for becoming, and for annihilation. Only renunciation and detachment from these cravings will allow a person to travel the Noble Path leading to the cessation of suffering: Right view; right thought; right speech; right action; right livelihood; right effort; right mindfulness; and right concentration. In areas where these Buddhist beliefs are prevalent, the thoughts, acts, and emotions associated with the Western concepts of moderate depression and dysthymia have been perceived traditionally as within the normal range of human experiences. It has taken a large public health campaign and much advertising by pharmaceutical companies, for example, to convince people in Japan that they are truly depressed.

Patients' declarations that they are sinners, evil persons, and, to a lesser extent, guilt-ridden are encountered less frequently by psychiatrists than in the past. Some religious patients may try to extract meaning from their depression or psychosis or feel that God is testing them. A simple and truthful approach should be used. All of the patient's symptoms should be enumerated, and the diagnosis explained. They should be told that in addition to psychosocial factors contributing to their illness, modern science has shown that disturbances in brain chemistry and neural circuitry play a significant role. In addition, it is often ill-advised to seek meaning during the course of an illness because cognitive and emotional distortions compromise their ability to reason clearly. A quest for meaning may be best after the illness has either passed or come under good control. This approach does not discount the use of spiritual or religious resources during the course of treatment.

Anxiety is an experience of everyday life, and everyone, from the born-again Christian to nasty sociopath, has anxious moments—for example, when the telephone rings in the middle of the night, when one is crossing a busy street with six lanes of traffic, when a laboratory report indicates a possibly fatal illness. Persistent anxiety can be terribly corrosive both by itself and in panic, social phobia, posttraumatic stress, and OCDs.

The literature on anxiety, religion, and spirituality is sparse and consists mostly of cross-sectional studies. The only clear findings are that the presence of religious obsession and compulsions offer a poor prognosis and that the collapse or weakening of faith in PTSD patients often results in unresponsiveness to therapy. An Internet survey of 1,402 adults examined the association between mental disorders (using a subscale of the Symptom Assessment-45 Questionnaire) and religious characteristics. The study found that frequency of prayer was significantly related to higher phobic scores, whereas belief in an afterlife was related significantly to lower phobic scores. No association was found between phobia and religious fundamentalism or religious attendance. A 2002 study of 37,000 representative Canadians is the largest on the interactions among religiousness (defined as how often a person worships in an organized religious setting), spirituality (defined by how important a search for a meaning for life was for a person), and mental illness. The study found that religiousness and spirituality were linked positively with older age, higher income, and greater social support. Separated, divorced, and widowed persons were the most spiritual. With regard to depression, manic episodes, panic disorder, and social phobia, spirituality was significantly associated with higher lifetime odds, whereas religiousness had the opposite relation. The study was unable to determine whether spirituality helped set the stage for the genesis of the disorders, was a response to the disorders, or was simply a correlative finding.

Existential anxiety was a topic of psychiatric concern, especially in the mid-20th century, when it was heralded as the root cause of all mental illness. This anxiety supposedly stemmed from the experiences of being thrown from the womb into a meaningless world and then spending one's life confronting the fearful freedom that comes with meaninglessness and eventual death, the trauma of nonbeing. People were said to cope with this anxiety by rationalizations and leaps of faith that attempted to impose some order in the world; for example, religions redefine death as the end of physical existence while allowing for the continued existence of the soul, which endures eternally. Although existentialism has a rich and provoking literary heritage, it is rarely even mentioned in modern psychiatric textbooks.

Anomalous Experiences. Psychiatrists may encounter nonpsychotic patients who report anomalous experiences that defy rational explanations, such as so-called near-death experiences, alien abductions, being touched by an angel, telepathy, and clairvoyance. Although these phenomena are odd, persons who have experienced them and have been studied do not appear to be mentally ill but rather go on to develop increased empathy with others and involvement with family members, enhanced interest in spirituality, appreciation of the purpose of life and mundane occurrences, and a stronger belief in an afterlife.

Of interest, the literature on this topic includes mystical experiences—defined in this section as true Spirituality—as anomalous events. In various surveys, up to 40% of subjects reported a mystical experience at least once in their lives. Undoubtedly, some of these represent Spirituality, but it is probable that many are out-of-the-ordinary emotional New Age experiences and may even include dissociative or brief psychotic episodes.

NEAR-DEATH EXPERIENCES. Near-death experiences or something like them have a lengthy and widespread tradition. In *The Republic,* the Greek philosopher Plato described the hero Er, who was slain in battle and carried home for burial. Twelve days later, his body, surprisingly unaffected by decay, was placed on a funeral pyre. Er then came to life and reported that his soul had left his body and gone on a journey with a great company to a mysterious place where the souls of the dead resided. He saw a rainbow-colored column of light extending through heaven and earth and scenes of judgment that were both terrifying and joyful. Descriptions have been found in Buddhist and Islamic texts and in Chinese, Siberian, Finnish, and North and South American accounts of out-of-body experiences in which a person’s spiritual body experiences a bright light accompanied by joyful peace and meets with ancestors and departed friends after crossing the border between the living and the dead.

The psychologist Raymond Moody’s book *Life After Death* contained 150 interviews with persons who reported near-death experiences and brought attention to this phenomenon in 1975. These experiences occur when persons are in a state of what appears to be death (they may even be pronounced dead by a clinician) but then are revived. During the period of not-quite-life and not-quite-death, many report that they were out of their bodies, looking down at themselves, and felt themselves moving through a dark tunnel peacefully and quietly. Their lives were reviewed as in a panorama. They saw a glowing light with a human shape, cities of light, or a border from which there was no return. They meet deceased relatives and friends. They heard doctors or spectators who pronounced them dead, but they were rescued from death by a spirit.

Remarkable similarities exist between the reports of near-death survivors and users of hallucinogenic and other drugs. Patients recovering from anesthesia, for example, may recall words spoken by the operating team during surgery. Bright lights and tunnel imagery can occur when drugs that mimic the effects of light on the retina stimulate the central nervous system or when the threshold for light perception is lowered, so that bright lights are seen surrounded by darkness. Published statements from hallucinogenic drug users describe:

> …bright colors not like here on earth but just indescribable…. I could see a city, a city of light…. There are tall structures all around me in all colors…. It seems like I’m getting closer and closer to the sun, and there’s a geometric network or lattice in the distance.

Even the panoramic, rapid review of a person’s life has been described by users of drugs that excite the central nervous system. Such excitation may also be induced by fever, exhaustion, and psychological processes associated with dying; decreased sensory input from failing body organs; a turning inward of attention and blocking out of the environment;

reminiscence; and a fear of death that may trigger both comforting fantasies and dissociative mental states. A fanciful theory is that the near-death experience is so traumatic that it kindles the memory of birth trauma when the infant travels through the vaginal tunnel toward the light. Yet another theory is that all memories are stored, but most are suppressed by new experiences that create more memories. In situations such as the near-death experience, new input is markedly decreased, thus allowing the old memories to be released and to be psychodynamically experienced as hallucinations, dreams, or fantasies. When faced with imminent death, a person’s thoughts often turn to God, heaven, and culturally accepted religious images.

Many scientists believe that the imagery and sensations of the near-death experience are caused by biologic processes and mediated by cultural expectations. Another example of a biologic process that results in a strange experience occurs when a person whose painful limb had been amputated feels pain as if the leg was still present. This phantom limb pain starts at the limb stump and persists in the sensory centers of the brain. A different type of experience, the autoscopic phenomenon, is associated with epilepsy and migraine headaches. It consists of the sudden, brief perception of a person’s face or body as a colorless, transparent phantom that imitates the person’s actions. The German writer Johann Wolfgang von Goethe described a likely autoscopic experience when, while horseback riding, he saw an approaching horseman who was his double and who then rapidly disappeared. The biologic process of sleep paralysis is associated throughout much of the world with culturally elaborated beliefs and visions of a ghost, the spirit of a dead person, or an “old hag” who presses on the sleeping person’s chest. The Hmong of Cambodia claim that they see a tiny figure, or demon, which crushes their chests. In fact, so many young Hmong men die in their sleep that a sudden unexpected nocturnal death syndrome is now medically recognized. It is a genetic disorder that causes fatal ventricular arrhythmia. Survivors describe intense fear, paralysis, and pressure on their chests caused by an alien being.

Most persons endow their near-death experience with a favorable spiritual explanation, such as having been “touched by the light,” which supposedly has some association with heaven and with God. Many religious traditions associate darkness with hell and the forces of evil, whereas light epitomizes heaven, goodness, and God. In *The Divine*

Comedy, one of the greatest literary evocations of God in Western literature, the poet Dante saw in Paradise both a cross that brilliantly “flames forth Christ” and God, “the eternal light.” Unfortunately, some near-death experiences are not so splendid. In various studies, up to 22% have been described as hellish and terrifying, with the person having a sense of being condemned to a barren void for eternity or falling into a demon-filled pit.

ALIEN ABDUCTIONS: A PSYCHOSPIRITUAL PHENOMENA? A few individuals believe that they were abducted by extraterrestrial aliens. In many reports, these abductions occur at night in a rural setting while the victim is driving a car and spots a UFO. The victim may stop to investigate, or the car will stop because of mechanical failure presumably induced by the UFO. The driver and passengers, if any, get out of the car, become amnesic for several hours, and then regain awareness. Usually, they are mystified by the lost time, but in subsequent dreams, hypnotic sessions, or regression therapy to recover memories, they remember being paralyzed, transported in a beam of light, and abducted by gray-skinned humanoids with large heads and very large, dark eyes. During this terrifying experience, the victims report being surgically probed, having devices implanted in their bodies, and being sexually abused or impregnated by the aliens. A few victims may have some sort of mystical experience, but this is soon followed by medical and psychological problems.

This scenario seems to be a North American phenomenon that first received public attention in the 1960s. Well-respected psychiatrists and psychologists who have interviewed persons claiming to be abductees have found them to be rather ordinary, with a level of mental health reflecting that of the general population. Many have created support groups and a subculture devoted to explaining the abductions, including the belief that they represent alien attempts to better understand human beings or to repopulate the Earth with human-alien hybrids, secret government mind control experiments, and so on. Similar reports of abductions from around the world tend to be more benign, with the abductors being demons or spiritual entities.

The physical reality of the abduction phenomenon is scientifically absurd, but there is no doubt that it might be a culturally elaborated psychospiritual reality. As reports of abduction rely so often on the recovery of lost memories, one theory is that the basic repressed memory is of childbirth: The fetus is forcibly plucked from the safety of the womb and placed in an alien, cold environment; bright lights are shined on it, and it sees the eyes of masked strangers who probe its body. Because newborns cannot distinguish colors, the strangers appear to be gray, which accounts for the perception of abductors as "grays."

The historical, religious, and anthropologic literature of the world contains accounts of many visionary encounters with deities, demons, and ancestors. The Hebrew Bible, for example, reports encounters with God and his surrogates. In 1149, an Irish monk wrote *The Vision of Tundal,* a highly successful book that was translated into most of the languages of Europe. Tundal saw Satan in hell viciously torturing sinners. At that time, no one doubted the veracity of Tundal's accounts or those of many other writers. The slow demise of visionary encounters began with the publication of Dante's *The Divine Comedy* in the 14th century. Dante provided a compelling and profound first-person account of his voyage to hell, purgatory, and heaven but admitted it was a poem derived from his imagination. In Western culture, skepticism about visions continued through the 17th and 18th centuries up to modern times, in which visions have usually been associated with psychosis, although in folklore and shamanic and some tribal cultures, visions are accepted as reality.

In today's world, only a few persons believe in the physical reality of the abduction phenomenon, a surprisingly larger number would endorse its possibility, and the vast majority reject the idea. John Mack, the most serious psychiatric investigator of the phenomenon, concluded that it is a modern type of visionary experience but overoptimistically felt that it might portend a dramatic change in human consciousness and a blow to the world's collective egoism in the tradition of Copernicus, Darwin, and Freud. Although it may be futile to try to convince abductees that their experiences were delusional and not physically real, Mack determined that many eventually come to regard the phenomenon more positively in that it resulted in a broadening of their consciousness.

TOUCHED BY ANGELS. Belief in guardian angels is a staple of both New Age spirituality and most persons who consider themselves religious.

Computer search engines typically show more than 2 million "guardian angel" sites ranging from offering guardian angel pins, wallpaper, and poems to tattoos, newsletters, and dog rescue teams specific to certain breeds. Large bookstores have sections devoted to angels, and in the film, *It's a Wonderful Life,* one of the most-watched movies of all time, a guardian angel saves the day. There have been literally millions of testimonials throughout history, especially in the last half-century, attesting to the power of a person's guardian angel miraculously to heal relationships and illness, provide expert guidance in life decisions, expand consciousness and spiritual evolution, and render material support in hard times. In his 1975 best-seller *Angels: God's Secret Agents,* the evangelist Billy Graham's language was probably influenced by the recent U.S. experience in Vietnam. He wrote:

> "All of the major religious traditions profess a belief in angels or similar spiritual beings, who usually are considered to be God's messengers and bringers of good news."

Treatment Issues. There is much confusion in both the literature and in most people's minds, including psychiatrists and patients, because of the conflation of the terms Spirituality, spirituality, spiritual, religion, and religiosity (or religiousness). The word Spirituality should be used to refer to the extraordinary moment or brief period of time during which a person feels a special connectedness to living beings, other forms of life, nature, causes, sacred or culturally significant artifacts, the universe, supernatural entities, or even ancestors. This feeling is special because it strikes deep into the perceived inner core of a person as a revelation or epiphany that cannot be fully described in words. Spirituality appears suddenly, and it is possible for some persons to cultivate its reappearance. Some religious mystics, St. Teresa of Avila, for example, described multiple episodes of a transcendent connectedness with God. In a secular context, the polymath S. Giora Shoham described "experiences of blissful temporary dissolution of the self's cognitive partitions" while watching the sun set behind some ruins of the Mediterranean shore while the waves, the darkening sky, and his self fused into an opaque unison. He developed a capacity to recapture the experience; for example, "By concentrating on the yellow petals of a wild chrysanthemum while the petals extended and widened, my whole self was engulfed into a vision of a single omnipresence." Some patients report having had such a moment of Spirituality (not to be confused with moments of therapeutic clarity or insight) that influenced them to lead spiritual lives. The experience of Spirituality is a highly prized phenomenon but cannot be induced by psychiatric expertise, which can only enhance the probability that it might occur.

The Spiritual Life. Patients seek psychiatric treatment because they feel ill and beset by stress and suffering. Increasingly, medication has become the most common treatment modality, but supportive psychotherapy should be an integral component of every clinical encounter, even brief ones. As opposed to more traditional psychotherapy that occurs over an extended period of time and symptom-focused, shorter-term cognitive and behavioral forms of therapy, supportive therapy is the most difficult. It presupposes that the psychiatrist understands the basic nature and course of an illness, the individual psychological and social dynamics of the case, and the patient's cultural beliefs and practices. The psychiatrist then uses this information to ensure treatment compliance through advice, exhortations, admonitions, and even threats, such as commitment to a hospital. The psychiatrist's duty should be not only to reduce or eliminate symptoms but also to enhance a patient's quality of life.

Biologic interventions rarely cure mental disorders, but they do make it possible for patients to deal with the issues that cause them stress and suffering. Through advice, exhortations, and so on, the psychiatrist's role is to point patients in the right direction on life's path. Some advice is immediately practical—for example, get a payee to help with finances or move to a different apartment to avoid the drug dealers next door. Much of the advice, however, is geared toward having patients lead spiritual lives, even if the adjective *spiritual* is not articulated by the psychiatrist or the patient.

Some of the things that constitute a good spiritual life are based on mundane experiences such as doing charitable deeds; heeding one's marriage vows; appreciating the beauty of nature; accepting the responsibilities of parenthood; caring for pets; protecting the environment; visiting the ill; forgiving those who have harmed one; maintaining good relationships with parents, relatives, and neighbors; caring for the integrity of one's body; cheering for a favorite sports team; supporting the troops, in whatever fashion one envisages; engaging in civic affairs; developing hobbies; maximizing one's talents and special abilities; refraining from illegal activities; not drinking to excess; and so on. However, no one is perfect, and a person does not have to be a saint to lead a mostly spiritual life. Very few people can follow all these precepts all of the time. There will be occasions when laziness, envy, affairs, divorces, drunkenness, and other problems will appear on the scene. When this occurs, it is the psychiatrist’s role to minimize their deleterious effects and provide advice to help a patient get back on track.

THE RELIGIOUS LIFE. For some patients, a spiritual life includes religion, a facet of life that provides for both much that is wonderful and much that is horrible in human experience. In his 1902 masterpiece *The Varieties of Religious Experience,* the psychologist William James went so far as to declare that religion is an essential organ of a person’s life and that it performs a function that no other portion of human nature can fulfill. He wrote that during a special state of mind known only to religious persons:

> What we most dreaded has become the habitation of our safety, and the hour of our mortal death has turned into our spiritual birthday. The time for tension in our soul is over, and that of happy relaxation, of calm, deep breathing, of an eternal present, with no discordant future, has arrived. Fear…is positively expunged and wasted away.

A contrary view is articulated in Sam Harris’s 2005 book *The End of Faith: Religion, Terror, and the Age of Reason,* in which he provides voluminous accounts of the historical and current episodes of bloodshed and terror spawned by religion, such as the killing of heretics and the suicide bombing of innocent civilians.

When psychiatrists are sure that religion is not a factor in patients’ psychopathology and that their participation in religious activities fosters a spiritual life, patients should be encouraged to follow their religious proclivities. Many patients find much solace simply in reading holy scriptures, even though most sacred books can be difficult to understand, contain conflicting messages, and may even lead to invidious and harmful acts if interpreted concretely or wrongly.

There is much to be said for the value of participating in organized religious activities and holding religious beliefs. Certainly, each religion imposes its particular order on life and provides rules for living that, if followed, usually result in the belief that an externally felicitous afterlife will be achieved. It gives hope to all impoverished, unfortunate, disfigured, badly treated, sick, lonely, and dying people that they might be happy and blessed in the next world, be it heaven, Nirvana, or the happy hunting grounds. Questions about the meaning and purpose of life arise, particularly in old age and during palliative treatment. In such situations, believers want to hear religious explanations rather than medical clichés. Questions about faith are also common in traumatized patients who feel abandoned by God and experience survivor guilt. Wounded soldiers may feel that they deserve to be tormented as a punishment for their wartime activities. A traumatic event may become so central to victims' lives that, especially in times of crisis or severe stress, they re-enact it in their minds or, more tragically, in their behaviors as a pathologic form of worship. A religious approach to their therapy is often needed to provide a more benign object of worship.

By belonging to a caring religious community, persons can count on emotional and material support in times of crisis, but the largest study on religion and coping with crises found that "On the face of it, people appear to be able to cope just as effectively without religion as with it." Furthermore, religious rituals in times of crisis were helpful in 40% of cases and harmful in 23%. It would seem that even for true believers, the faith, hope, and compassion offered by religion might be necessary but not sufficient for coping with crises or dealing with illness.

There is a growing literature demonstrating some health advantages for persons who attend religious services. One explanation is that better physical health leads to an increased ability to attend services. It may reflect a placebo effect or simply a correlation. There is little information comparing the relative health advantages among religions or branches of religions: Do Muslims have greater health advantages than Christians, Jews, or Hindus? How do Baptists compare to Catholics or Pentecostals? At any rate, depending on the results of a spiritual assessment, psychiatrists should encourage church-going patients to continue their religious practices. Lonely, shy, or morally lost persons who have stopped attending religious services for any number of reasons (panic attacks, getting caught up in other activities, etc.) but for whom religion was once important should be gently encouraged to at least consider searching for a church and a congregation where they feel comfortable.

Prayer is the most common form of religious worship, and there is no reason for psychiatrists to dissuade patients from praying because it does no harm but rather offers hope and consolation. Hands-on prayer often provides comfort and temporary relief from some illnesses and can be explained by psychological mechanisms. Studies of prayer-at-a-distance in which prayed-for sick persons are unaware they are being prayed for implicitly test a supposed supernatural mechanism. Such studies have not demonstrated any advantages to such subjects. A major study of 748 patients with coronary artery disease who were undergoing certain cardiac procedures showed that neither prayer-at-a-distance nor bedside

MIT (music, imagery, and touch) therapy significantly affected the primary clinical outcome measure. A confounding variable in these studies is that nonprayed-for control group subjects may be offering their own prayers or know that relatives, friends, or even hospital staff are praying for them.

FIGURE 30.7–5. St. John writing his Gospel, as portrayed in a painting in St. Isaac’s Cathedral in Saint Petersburg, Russia. The eagle was his representative animal and indicates the spreading of John’s Gospel throughout the world. (Domenichino. National Gallery, London.)

For Christians, the New Testament contains seemingly difficult-to-understand comments on prayer. In three instances, the Gospels promise that believers who pray to God will have their prayers answered. When prayers are not answered, however, the explanation offered is that even Jesus was unable to convince God the Father to spare him from his crucifixion, recognizing that God’s will must take precedence. In other words, God only answers prayers that conform to his will, and there is no way to convince him otherwise. Whatever happens, *is* God’s will. Nonetheless, Christians continue to offer intercessory prayers for themselves and others. Pastoral and Christian counselors often pray with patients, as do some psychiatrists. As a general rule, however, when patients ask psychiatrists to pray with them, the most judicious response is for psychiatrists to say that they will witness the patients' prayers in respectful silence (Fig. 30.7–5).

Domesticity, particularly in Christian families, may be threatened when a family member is discovered to be a homosexual. Either the homosexual person, who may be depressed, or his angry family members may seek psychiatric treatment because, in their tradition, homosexuality is reviled as an unnatural, repulsive condition. For Christians, the basis for this attitude is biblical. The book of Leviticus contains laws that God presented to Moses; verse 20:13 reads, "If a man lies with a man as he lies with a woman, both of them have committed an abomination. They shall surely be put to death. Their blood shall be upon them." Death sentences are also proscribed for adulterers, men who sleep with their father's wife or their daughter-in-law, people who mate with animals, and fortune-tellers. The hundreds of Levitical laws condemn many behaviors, such as trimming one's beard; sitting on a seat that a menstruating woman has sat on; letting defective people such as those who are blind, dwarfs, have marred faces, very long limbs, or eczema defile the Temple sanctuary by their presence; touching a mouse; eating shellfish, rabbits, or pork; and wearing garments made of a mixture of linen and wool.

Psychiatrists should at least be aware that the concept of homosexuality, as we know it currently, did not begin to take shape until the 11th century when St. Peter Damian coined the word "sodomy." Before then, persons who were intimate with members of their own sex were simply regarded as persons with special sexual preferences. Other biblical references to "homosexuality" are ambiguous and can be interpreted in differing ways. For example, for at least two millennia, the destruction of the city of Sodom was attributed to the lack of civility and hospitality to God's angels rather than to homosexual activity. John Boswell's *Christianity, Social Tolerance and Homosexuality,* and Armando Favazza's *PsychoBible* are books that might be consulted by psychiatrists and patients for more details on this issue.

Psychiatrists should also be aware that, according to Christian theologians, Jesus overturned the Levitical laws and that Christians are neither expected nor obligated to follow them. Some theologians, however, hold that only the ceremonial laws were overturned. Even if this were the case, Christians have no problem with men being in close presence with menstruating women, men trimming their beards, and people eating pork, and they certainly do not condemn adulterers or fortune-tellers to death. The persistent Christian fixation on homosexual behavior as unnatural and homosexuality as a deliberately chosen preference is irrational from the perspective of psychiatry, which dropped homosexuality from its list of mental disorders in 1973. Fortunately, the Western culture is showing more and more acceptance of homosexuality, along with the realization that both homosexual and heterosexual persons may lead spiritual lives.

Psychiatrists are often called on to treat abused women and should be alert to the possible misuse of religion in such situations. The Semitic religions originated in patriarchal cultures, and many of their beliefs and practices reflect a belief in the superiority of men over women. The book of Ecclesiasticus, found in Catholic and Greek Orthodox Bibles, states, “From a woman (Eve) sin had its beginning, and because of her we all die.” In the book of Numbers, women are classified as chattel property. In the New Testament, Paul notes that “the head of every woman is man,” and this view is repeated in various writings: For example, “Wives, submit to your own husbands, as to the Lord” (Ephesians 5:22), and “Wives, be submissive to your own husbands” (1 Peter 2:3). In the nonbiblical Gospel of Thomas, Peter the disciple says to Jesus, “Let Mary [Magdalene] leave us, for women are not worthy of life.” Jesus replies, “I myself shall lead her in order to make her male, so that she also may become a living spirit like you males.” St. Augustine, four centuries later, settled things by declaring that man, either alone or when joined by his wife, is the image of God, whereas women alone are not. However, the Old Testament also states in Genesis 1:27, “So God created man in his own image, in the image of God he created him; male and female he created them.”

Some Christian, Jewish, or Muslim men may misuse religion by citing sacred texts to rationalize the humiliation and abuse of their wives. By using passionately pious appeals, abusers may gain the support of their in-laws and congregations. Psychiatrists who treat the victims of such persons may be portrayed as evildoers who subvert the will of God unless they endorse the status quo. However, if therapeutic efforts to ameliorate the abusive situation fail, divorce must be considered, even though the patient may find herself estranged from her family, her congregation, and maybe even her religion. Escaping from an abusive relationship is a spiritually wise decision.

Specific Spiritual Therapies. Alcoholics Anonymous (AA) exemplifies the clinical efficacy of treating a specific illness by leading a spiritual life. The movement started in 1931 when a wealthy American alcoholic was unsuccessfully treated for a year by Carl Jung. Jung advised him to place his hope in a spiritual experience, which he eventually found in the Oxford Group, an interdenominational organization devoted to re-experiencing the dynamism of Christianity's early days through inspirational talks, group discussion, and mutual support. His religious conversion and the process through which he achieved it was a founding moment of AA.

One of his alcoholic friends also experienced a conversion and, in 1934, tried to convince an alcoholic stockbroker named Bill Wilson that religion was the answer. Wilson responded by going on a drinking binge, was hospitalized by a psychiatrist, and cried out for God to show himself. He then experienced a moment of Spirituality: "Suddenly the room lit up with a great white light. I was caught up in an ecstasy where there are no words to describe." He pictured himself on a mountain in a Spiritual windstorm. "And then it burst on me that I am a free man." As the ecstasy subsided, he experienced,

> A new world of consciousness. All about me and through me there was a wonderful feeling of Presence and I thought to myself, 'So this is the God of the preacher.' A great peace came over me and I thought that no matter how wrong things seem to be, they are all right.

AA is not a religion, although it comes close, and some critics have called it a religion in denial. It is not a Christian organization either, but its roots are in Christian teachings. Each of its famous 12 steps contains a spiritual message, such as the necessity for alcoholics to turn their will and lives over to the care of God as they understand him; to make a searching and fearless personal inventory; to make direct amends to persons that they have harmed; and to carry the AA message to other alcoholics and to practice these principles in their lives.

According to AA, alcoholics' acceptance of self as human is founded on their rejection of any claim to be more than human. AA members must accept the limitation of their being: An alcoholic cannot drink alcohol safely. Formal religions take personal limitations as a problem that can be overcome by participation and belief in the mechanisms of salvation offered by a particular church. Religion aspires to perfection and contains what it considers to be absolute truths. However, organized religion promised too much for the founders of AA. When asked why AA separated from the Oxford Group, Bill Wilson replied, “The Oxford Group wanted to save the world, and I only wanted to save the drunks.”

Another example of a psychiatric therapy that utilizes guidelines for achieving a spiritual life is found in a series of DVD modules, *The Happy Life—Voyages to Well Being,* developed by the psychiatrist C. Robert Cloninger. The modules attempt to enhance character traits associated with frequent positive emotions and to enhance a sense of hope and mastery for self-directedness, kindness, forgiveness, cooperativeness, mindfulness, and finding spiritual meaning. The therapy begins with a focus on what makes patients happy or unhappy, the experience of well-being, and awakening their physical and spiritual senses. Patients then observe and strive to elevate their thoughts and human relationships, chart their maturity, and contemplate their being. The therapy ends with modules on stress reduction, quieting fears, observing the power-seekers in their life, contemplation of mysteries, and constant awareness. Although a happy life for patients is supposed to be a result of the therapy, there is a branch of psychology devoted to searching for what actually constitutes happiness. The theologian–physician Albert Schweitzer, who became depressed after concluding that Jesus had not really been resurrected, noted that “Happiness is nothing more than good health and a bad memory.” This new form of therapy is based on some solid research findings; for example, low self-directedness scores on the Temperament and Character Inventory indicate a vulnerability to major depression, whereas high scores predict a rapid and stable response to both antidepressant medication and cognitive-behavioral therapy. Whether this therapy will be widely used or proven effective is still unknown.

Evidence from psychological studies of personality points to a biologic consideration for a partial understanding of spirituality. A widely accepted model of fairly stable core personality traits rooted in biology identifies five dimensions: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Each dimension has associated descriptors or facets such as degree of emotional stability, compassion, self-discipline, and so forth. Characteristics of the five dimensions arise because of environmental forces and result in the development of beliefs, attitudes, goals, and a sense of self. Adaptations come into play during a person’s lifetime experiences and are translated into behaviors and emotional reactions. In research findings, Openness to Experience, Extraversion, and Agreeableness correlate highest with spirituality. A recent study among persons from differing faith traditions suggests that spirituality may be a sixth core personality dimension. The core personality traits impacted health outcomes in this study, but spirituality and religious subscales did not. Only the forgiveness subscale predicted better mental health outcomes. Forgiveness is one of Vaillant's positive emotions that comprise spirituality, but it is possible to regard forgiveness as a coping strategy that is related to but distinct from spirituality.

The "hard-wired" for God/spirituality concept is popular, especially in the lay press, and has given rise to the pseudoscientific field of neurotheology with notions of God genes, neurons, and neurotransmitters. Monoamine neurotransmitters have been identified in a popular book as the molecules partly responsible for self-transcendence and spirituality. A variation in the vascular monoamine transporter 2 gene in some persons supposedly enhances a tendency for spirituality. However, the book ends with a strange disclaimer: Spirituality may be partly genetic, but genes are like verbal cooking recipes, and what you end up with depends upon how you interpret and execute the directions.

Most people believe that almost any type of meditation does no harm and probably does some good just by reducing stress. But only some types that are readily accessible, such as mindfulness, hold the promise of achieving spirituality. In most instances, mindfulness is practiced in group workshops, although it may be utilized in conjunction with cognitive-behavioral therapy. The attainment of mindfulness spirituality is a gradual process and may take years of hard work. In 2014, the U.S. Agency for Healthcare Research and Quality issued a review of 41 meditation trials with matched control groups. The review revealed that mantra meditation had very little if any effect on psychological stress and well-being outcomes, while mindfulness meditation had a moderately positive effect on anxiety, depression, and pain. The conclusion was that meditation programs were not more effective than exercise, progressive muscle relaxation, cognitive-behavioral therapy, or drugs in changing health outcomes. The review did not address the issue of spirituality.

FURTHER READINGS

Cardena E, Lynn SJ, Krippner S. *Varieties of Anomalous Experience*. 2nd ed. American Psychological Association; 2014.

Cook C, Powell A, Sims A, eds. *Psychiatry and Spirituality.* Royal College Psychiatry Publications; 2009.

*Crescenti C, Di Bucchianico M, Fabbro F, et al. Excitatory stimulation of the right parietal cortex lessens implicit religiousness/spirituality. *Neuropsychologia*. 2015;70:71–79.

de Souza M, Francis CJ, O'Higgins-Norman J, Scott D. *International Handbook of Education for Spirituality, Care and Wellbeing*. Springer; 2009.

*Favazza A. *PsychoBible: Behavior, Religion, and the Holy Bible*. Pitchstone; 2004.

Galanter M. *Spirituality and the Healthy Mind*. Oxford University Press; 2005.

*Harris S. *Waking Up: A Guide to Spirituality Without Religion*. Norton; 2014.

Inserra A, De Gregorio D, Gobbi G. Psychedelics in psychiatry: neuroplastic, immunomodulatory, and neurotransmitter mechanisms. *Pharmacological Reviews*. 2021;73(1):202–277.

Jain AR. *Peace love yoga: the politics of global spirituality*. Oxford University Press; 2020.

*Johnstone B, Bodling A, Cohen D, Christ S, Wegrzyn A. Right parietal lobe-related "selflessness" as the neurophysiological basis of spiritual transcendence. *Int J Psychol Religion*. 2012;22(4):267–284.

Johnstone B, Gloss B. Support for a neuropsychological model of spirituality in persons with traumatic brain injury. *Zygon*. 2008;43(4):861–874.

*Johnstone B, Cohen D, Konopaki K, Ghan C. Selflessness as a foundation of spiritual transcendence. *Int J Psychol Religion*. 2016;26:287–303.

Kabat-Zinn J. *Coming to Our Senses*. Hyperion; 2006.

Krediet E, Bostoen T, Breeksema J, et al. Reviewing the potential of psychedelics for the treatment of PTSD. *Int J Neuropsychopharmacol*. 2020:23(6):385-400.

Koenig H, Huguelet P. *Religion and Spirituality in Psychiatry*. Cambridge University Press; 2009.

Michaelson V, Brooks F, Jirásek I, et al. Developmental patterns of adolescent spiritual health in six countries. *SSM Popul Health*. 2016;2:294–303.

Oviedo L. Religious attitudes and prosocial behavior. *Religion Brain Behav*. 2016;6:169–184.

*Pargament K. *Handbook on Psychology, Religion and Spirituality*. American Psychological Association; 2013.

Peteet JR, Lu F, Narrow W. *Religious and Spiritual Issues in Psychiatric Diagnosis*. American Psychiatric Publishing; 2009.

Purzycki BG, Willard AK. MCI theory: a critical discussion. *Religion Brain Behav*. 2016;6:207–274.

*Vaillant G. *Spiritual Evolution*. Broadway Books; 2008.

Van Cappellen P. Effects of oxytocin on spirituality and emotional responses to medication. *Soc Cog Affective Neuroscience*. 2016;10:1579–1587.

30.8 Physician and Medical Student Mental Health

NEETA JAIN, M.D., AND CYNTHIA M. STONNINGTON, M.D.

The World Health Organization (WHO) defines mental health as “a state of well-being in which the individual realizes his or her own abilities can cope with the normal stresses of life, can work productively and is able to make a contribution to his or her community.”

CURRENT FORCES IN HEALTH CARE

The U.S. health care delivery systems have undergone enormous shifts over the last two decades. Consolidations and mergers greatly transformed practice structures, with many physicians now serving as employees within a larger network. Health care regulations and policies, such as the Affordable Care Act and the narrowing of commercial insurance networks and noncommercial insurances (Medicare, Medicaid), resulted in declining reimbursements. Widespread implementation of electronic medical records (EMRs) was associated with increased clerical burden due to requirements of new quality metrics, charting completion for billing purposes, and answering inbox messages, all of which lengthened workdays well into evenings and weekends (an additional 28 hours per month on nights and weekends). These “efficiencies” created a greater expectation of productivity, usually increasing workloads with the same amount of time/resources, that is, efforts to improve efficiency came at the expense of physician autonomy. Furthermore, competent and acceptable practice required physicians to stay abreast of an expanding knowledge base, implement the latest updates on regulatory requirements (meaningful use, medication reconciliation, electronic prescribing), complete burdensome maintenance of certification requirements, and face new levels of scrutiny with quality metrics and patient satisfaction surveys. These factors have taken a personal toll on health care professionals. Physicians, trainees, and learners are therefore particularly vulnerable to experiencing burnout and mental health difficulties by the nature of their workplace demands.

The impact of burnout was not only recognized as taking a toll on the individual clinician, but mounting evidence pointed to a link between burnout and quality of care, medical errors, patient satisfaction, declining productivity, and turnover, all burdening health care organizations with further financial costs. Institutions began expanding the triple aim of reducing costs, improving quality, and increasing patient satisfaction to a quadruple aim that further includes health care professionals' wellness. Medical and resident education has begun appreciating the impact of the learning environment on well-being and the need for adequate resources to support mental health in recent years. Hippocrates wrote the infamous oath in the fifth century obliging physicians to "first do no harm" to patients but disregarded to include harm to oneself. Many medical schools use the modern version of the Hippocratic Oath to include a pledge "an awareness of my own frailty."

COVID-19 Impact

Not surprisingly, health care workers can experience significant psychological distress during an epidemic or pandemic. In January of 2020, the first reported coronavirus disease (COVID-19) case appeared in the United States. In a relatively short period, the spread of the virus led to a worldwide pandemic. As was previously observed with the SARS and Ebola epidemics, the sudden onset of life-threatening illness led to exceptional levels of pressure on health care professionals. In this instance, health care workers and health care systems were faced with limited access to personal protective equipment (PPE), testing, and treatment. Early in the pandemic, 20% of the workforce caring for those with COVID-19 became infected themselves, and many physicians witnessed colleagues becoming seriously unwell or dying due to their work. Many states implemented a lockdown with many physicians abruptly required to work remotely from home, limiting direct contact with not only patients but coworkers and colleagues, increasing social isolation, and affecting family systems, with schools switching to online education and parents now straddling a new work–life schedule entirely at home. As a society, this pandemic robbed the natural inclination to seek human comfort from loved ones such as family and friends, delivering a more severe impact than other disasters.

Frontline workers directly responsible in caring for COVID-19 patients not only had to face shortage of PPE, but were subjected to evolving and changing treatment guidelines, and felt inadequately supported with weighing responsibilities. Unfortunately, health care professionals entered the COVID-19 outbreak with an “empty tank,” as evidenced by multiple studies reflecting a high prevalence of stress, anxiety, and burnout among health care professionals, trainees, and students in the preceding decade. This pandemic strained the health care systems nationally and globally, reinforcing the concern for health care workers, including physician, resident, fellow, and learner burnout. The mental health impact of COVID-19 on health care workers became a major concern. The studies showed an increase in acute self-reported symptoms of depression and anxiety, as well as high rates of insomnia, substance use, and distress. A variety of studies from different countries and settings surveying health care workers during the COVID-19 pandemic revealed substantially increased prevalence rates of moderate to severe anxiety, depression, insomnia, and traumatic distress, compared to 2019. Long-term follow-up studies of health care workers who have worked in previous pandemics, for instance, the 2003 outbreak of severe respiratory syndrome, suggest that as many as 1 in 10 with early distress symptoms will have progressed to long-lasting mental health conditions for some individuals, such as post-traumatic stress disorder (PTSD). According to the Institute of Medicine, during the SARS outbreak of 2002–2003, four ethical issues arose: the roles and responsibilities of medical workers, the impact of infection on the global economy, equitable care, and the challenge of balancing public welfare and individual rights.

In the training manual for mental health and human service workers in major disasters, published in January 2000, Dr. Deborah DeWolfe outlines the psychosocial and emotional phases of disaster, including warning/threat, impact, rescue or heroic, honeymoon, inventory, disillusionment, and reconstruction or recovery. At the outset of a disaster, providing daily communication, adequate PPE, and developing safety protocols are top of mind. In the immediate aftermath, or “heroic” and “honeymoon” phases, the community comes together, appreciates the “heroes,” and may accept mental health interventions if offered. Psychological first aid is an approach used to help people in the immediate wake of a disaster decrease their acute level of distress and encourage short- and long-term adaptive functioning. During the disillusionment phase, the earlier feelings of optimism turn to discouragement and physical and emotional exhaustion and feelings of abandonment. During this time, we often see increased demand for psychiatric services. The disillusionment phase can last months and even years and is often extended by one or more trigger events, usually including the anniversary of the disaster. Similarly, the reconstruction phase may last for years, involving working through grief and loss and post-traumatic growth.

Beyond psychological first aid, a variety of services and programs can help health care professionals process the losses and stressful clinical events that they encounter. “Schwartz rounds,” for example, involve bringing staff of all levels together to have conversations about the emotional impact of their work. Peer support training teaches active listening skills, empathy, and validation so that peers can be mobilized to reach out to colleagues with offers of support. They can also normalize help-seeking when their colleague may need more professional help. Employee assistance programs and counseling services for learners are critical resources for those needing additional support. “Psychological PPE” and a variety of community and peer support offerings may gather steam during disasters but ideally would robustly always remain in place, given the prevalence of burnout, stress, and mental health conditions that physicians face even when there is not a global pandemic.

BURNOUT

The field of medicine is a highly competitive and demanding profession that attracts high-achieving and ambitious individuals. By the nature of the training, physicians have been able to endure long hours, irregular and/or inflexible schedules, increasing workloads, emotionally taxing situations, and stress that surpasses most other work environments. A growing body of literature has demonstrated a greater prevalence of occupational distress among physicians compared to the general workforce. The research has been rapidly expanding, with concerns for “burnout and moral injury” rising to a national level of public health concern.

Burnout was first recognized in many workplaces in the United States in the 1970s. This can affect workers, particularly those whose work involves an intense interaction with people (i.e., teachers, social workers, police officers, health care workers). Burnout syndrome was first described in

1974 by psychologist and psychoanalyst Herbert Freudenberger, Ph.D., who often volunteered at a free clinic for substance abusers in East Village of New York City. He observed that over time, volunteer clinical staff were emotionally depleted with accompanying psychosomatic symptoms. He borrowed the term “burnout” from drug-addict slang, with the burning out of substance abusers on the street with those mental health providers who tried to help these substance abusers. Independently, Christina Maslach, Ph.D., became interested in related work issues about the same time. She developed a method for assessing burnout as a multidimensional construct by creating the Maslach Burnout Inventory (MBI) with three subsections for a total of 22 questions in 1981. Burnout has been commonly characterized by a clinical constellation of symptoms of emotional exhaustion, depersonalization (DP), and decreased feeling of personal accomplishment (PA) that result from the chronic stress of medical practice. Emotional exhaustion is when an individual feels emotionally drained and has no more to give as a caregiver. This feeling of being unable to replenish differs from depression and compassion fatigue, with the caregiver beginning to distance themselves as a protective mechanism. This can lead to DP with the arising of cynicism and detachment toward the patient. A lower personal sense of accomplishment is the result, with less efficiency at work and a lower sense of pride and job satisfaction.

Burnout is not an expected reaction to hard work, though deeply satisfying work can involve tremendous personal sacrifice. It is not the same as depression, yet severe burnout can evolve into it. For burnout to be present, interpersonal stress in the workplace is present to some degree. Self-valuation appears to be particularly low for physicians and an important contributor to burnout, for example. The interactions between individuals and organizational factors likely contribute to a high burden of responsibility, neglecting self-care and needs, low perceived control, discordance between individual and organizational values, unsupportive work environments, isolation, and loss of meaning. The components of emotional exhaustion and DP are most commonly affected. Initially, burnout might coexist with empathy, satisfaction, and caring, but unrelieved work stress for months diminishes the ability to call forth and sustain those attributes. Burnout is not an acute self-limited illness. Numerous publications have confirmed that burnout is not confined to specific practice settings (academic, government, or private practice) or career stage. For an individual to experience burnout, the presence of only one of these symptoms needs to be present. The highest rates of burnout are found among physicians in emergency medicine and the lowest rates among preventive medicine/occupational medicine. In 2019, WHO revised the definition of burnout, which aligns with but is different from the Maslach paradigm: burnout is a syndrome conceptualized as resulting from chronic workplace stress that has not been successfully managed. It is characterized by three dimensions: (1) feelings of energy depletion or exhaustion (EE); (2) increased mental distance from one's job, or feelings of negativism or cynicism related to one's job, that is, DP; and (3) reduced professional efficacy or PA. Burnout refers specifically to phenomena in the occupational context and should not be applied to describe experiences in other areas of life.

Determining the scope of physician burnout is challenging. There is substantial heterogeneity in the research method and definitions employed by investigators studying burnout, including variability in the measures used to evaluate burnout symptoms. Studies have varied in sample size, practice settings, clinician types, and participation rates. The first large, national study on burnout among U.S. physicians across all specialties was published in 2012 by Tait Shanafelt and colleagues. There were 7,288 participating physicians, with approximately 45% experiencing at least one symptom of burnout. Furthermore, this study revealed that burnout was more common in U.S. physicians than in general U.S. workers. In the past decade, several studies of U.S. physicians found prevalence rates of burnout ranging from 40% to 54%, although with variable methodologic definitions and scales.

Personal Consequences of Physician Emotional Distress

Physicians of all ages experience a milieu of emotions from guilt, despair, and isolation after a complication or judgment call with devastating consequences. Physicians suffer, generally silently, when dealing with bad outcomes, and peer reviews or mortality and morbidity rounds offer minimal comfort. Numerous studies have shown positive associations between physician emotional distress and a range of deleterious physical and psychological conditions, including increased risk for cardiovascular disease, high cholesterol levels, type 2 diabetes, musculoskeletal pain, fatigue, headaches, gastrointestinal issues, respiratory problems, and severe injuries. Resident physicians with burnout have shown higher rates of needle sticks, bodily fluid exposures, and motor vehicle accidents. Clinicians experiencing burnout may also have higher rates of alcohol misuse, which may be related to self-medicating to cope with job-related stress.

Factors Driving Burnout

In the mid-1990s, Linzer et al. identified four factors associated with burnout in primary care practices: (1) lack of control over work conditions and decision making; (2) time pressure with emphasis on productivity; (3) a chaotic and inefficient work environment that inappropriately burdens physicians to do clerical and other mundane tasks; and (4) lack of alignment among physicians and executives regarding values, mission, purpose, and compensation. Factors that can increase the risk of burnout in health care environments include job demands (work frequency, duration, and intensity), higher workloads and time pressure (working more than 52 hours/week, night calls, and taking work home), computerized physician order entry, 24/7 remote access to patient records, tests results, and administrative burdens. This is often combined with a low level of autonomy or flexibility with work schedules. Some data suggest that patient portals may increase patient phone calls and overall workload, creating additional work for health care professionals. Remotely accessing electronic health records on personal time is stolen time from families and personal activities. Clinical documentation and online patient messages are not the only professional tasks clinicians might perform on personal time. Continuing Medical Education (CME) and maintenance of certification activities are frequently performed on nights, weekends, and even vacation time, since most organizations do not provide dedicated time for their completion. Clinicians must complete a number of required training modules each year, typically mandated by regulators, covering topics such as patient safety, infection control, the Health Insurance Portability and Accountability Act (HIPAA), and human subject protections. Because the pace of medical knowledge is accelerated and rapidly expanding, the time required to stay current has also increased, which is another demand on personal time.

For academic physicians, time pressures can be particularly acute. Historically, their time has been divided between clinical work, research, and training learners in education. The performance of these physicians is typically measured using an academic yardstick, requiring them to engage in educational activities (such as lectures and presentations), publications, and grants within defined time limits as a requirement for retaining their positions. To increase revenue, many academic centers have reduced the proportion of time allocated to scholarly pursuits and increased the time devoted to clinical care without adjusting any of these academic performance expectations. Most academic centers still have similar criteria for publications, presentations, grants, and educational responsibilities, which have increasingly shifted to personal time on nights and weekends. These variables have led to the encroachment of professional tasks on interpersonal time for physicians to a much greater degree than in many other fields. Often, this work is not clearly compensated. These professional activities create problems with work–life integration as well as adequate sleep, which all fuel burnout. Of note, physicians who were able to spend >20% of their professional time on what they valued most personally were less likely to experience burnout. Finally, a culture of prioritizing patient care and being “tough” over personal needs and help-seeking has created work–life conflicts, inadequate support or space to process stressful clinical events, and increased risk of workplace attrition.

In a consensus report released in 2019, the National Academy of Medicine developed a systems-based framework that conceptualizes clinical burnout as a consequence of frontline care delivery, health care organization, and external environmental factors. The report highlighted the need for efforts to both prevent and address burnout among clinicians through the engagement of health care organizations, electronic health record developers, private payers, and other stakeholders. Accordingly, the American Medical Association (AMA)’s STEPS Forward initiative was developed as a national tool kit for disseminating strategies to ameliorate physician burnout and improve patient care and cost containment.

Consequences of Burnout

The phenomenon of burnout affects all medical disciplines. It has been associated with increased medical errors, lower patient satisfaction, longer postdischarge recovery times, and decreased professional work effort. In addition, a direct negative impact is seen on fatigue, stress, anxiety, depression, substance abuse, suicides, poor quality care, early retirements, and unexpected resignations. Furthermore, higher rates of absenteeism, chronic work disability, hospital admissions due to mental and physical disorders, higher turnover, and poorer job performance can occur. Collectively, these factors make it more difficult for health care organizations to maintain an adequate and competent professional workforce. A 2019 study estimated that the economic cost of turnover and reduced productivity due to burnout among U.S. physicians was approximately $4 billion each year.

Moral Injury/Distress

The dynamic challenges of medical practice give way to moral injury, a syndrome that has been noted in military personnel for years and is now increasingly recognized as a source of physician distress. Moral injury is the harm experienced from a violation of moral boundaries. This distress is experienced when clinicians perceive they have participated in an undesirable situation or have been unable to act in accord with their professional and ethical values under conditions of constraint or duress. Pressure to act contrary to ethical standards can arise from patients or their families, clinical teams, or external environments. Pushback on “resilience training” as an answer to burnout grew out of frustration with the current structure of the health care delivery system; targeting the intervention largely on individual physicians, who already have substantial personal resilience, fueled skepticism associated with burnout. Health care professionals described that their ability to deliver necessary care was compromised by the system at large (i.e., insurance, reimbursements, electronic health record) seen in hospitals, clinics, and medical practices. The lack of control over the work environment, lack of balance between effort and reward, lack of transparency and fairness by decision makers, fragmentation of community among providers, and incongruence between one’s personal values and work demands have eroded wellness among physicians. The most common sources of moral distress among physicians were related to a lack of resources that compromised patient care and lack of administrative action.

Personality Traits

It comes as no surprise that many medical students, trainees, and physicians endorse certain personality traits such as perfectionism, competitiveness, intellectualization, caregiver, rigidity, and strong work ethic that make this profession desirable. Focus on these qualities during admission into medical school, residency, and fellowship training programs, and physicians seeking to hire a new associate, may self-select individuals vulnerable to mental illness in the future. Ironically, some characteristics such as idealism, perfectionism, and a strong work ethic that allow physicians to excel can make them more vulnerable to burnout. Traditionally, physicians are rewarded for being the first to arrive, not attending to their physical and emotional needs, and being the last to leave. What is it about the culture of medicine? The longstanding sentiment that if you are “not strong enough” to handle the demands, then consider seeking another profession has diminished over time but not entirely disappeared. Some senior faculty and administrators express dismay about the “softening” of the culture and students. There are ingrained aspects of medical culture that may have slowed the process of change. Traditionally held beliefs are that more pressure, more hours, and more demands would lead to better educational outcomes, and with less rigorous training, standards are being lowered. Such perceived beliefs can act as barriers to seeking and receiving the help needed.

Tracy and Robin’s Model of Self-Conscious Emotions provides a well-studied theory about self-conscious emotions in medical education. Congruence between current and ideal self-representation can elicit positive emotions of pride. If there is incongruence between self-representation and ideal self, negative emotions of shame or guilt can be elicited. Medical learners are apt to hold themselves to unrealistic standards of performance. Achieving perfection is a common, if not a pervasive, goal. It is a distinguishing character trait among learners and practicing physicians. Maladaptive perfectionism has been shown to correlate with psychological distress in medical learning. This is often additive with social, personal, and family expectations maintaining a constant level of pressure. Shame and guilt are not inherently tied to this theory. How the learner perceives a negative outcome distinguishes feelings of shame (inner locus of self) compared to a specific action (guilt). If a perfectionist learner frequently falls short but attributes the shortcomings to correctable actions, they are likely to learn and grow. However, highlighting imperfection deficiencies within themselves is likely to elicit a shame reaction, which can lead to withdrawal and disengagement from learning.

While many of these attributes of being fastidious to detail and setting high performance standards for oneself may be beneficial, there is a risk for more harm than good if this mindset is characterized by excessive self-criticism leading to psychological distress called maladaptive perfectionism. Maladaptive perfectionism is related to negative feelings and psychological distress, as seen in several studies on health care professional students. It is also strongly correlated with and predictive of future burnout among medical students.

There are several studies on the psychological construct of imposter syndrome. This phenomenon is characterized by a pervasive feeling of inadequacy and irrational fear of being discovered as an intellectual fraud despite one's capabilities and accomplishments. Estimates of prevalence among medical students range from 15% to 50%, with female medical students at higher risk for experiencing significant frequency and intensity of these feelings.

Women in Medicine

Building upon years of progress, women now comprise the majority of enrolled U.S. medical students, according to the 2019 data released by the American Association of Medical Colleges (AAMC). In 2017, women represented the majority of the first-year medical class. Even so, the gender gap with women in leadership positions persists. According to the AAMC 2018 data, women accounted for 16% of medical school deans, 18% of all department chairs, and 25% of full professors. The six academic departments with 50% or more full-time women faculty, such as obstetrics, pediatrics, public health and preventive medicine, psychiatry, dermatology, and family practice, nonetheless lagged in leadership roles and academic promotions for women, with most holding midlevel or associate dean positions. Women aspire to leadership roles equally to men. There is good data that women in leadership roles improve operational performance, advance innovation, promote group performance, and increase recruiting ability. Obstacles to equity that women face remain subtle, with overt biases based on gender stereotypes and work–life conflicts.

Women across the world are affected by sexual harassment in their daily lives. The medical field is no exception. This topic rose in prominence in late 2016 with the rise of the #MeToo movement. A survey in 2016 by Jagsi and colleagues found that 30% of female clinician-researchers reported experiences of sexual harassment compared with 4% of their male counterparts. In 2018, the National Academy of Science, Engineering, and Medicine reported that in a large university system, female medical students were “220 percent more likely than non-STEM (science, technology, engineering, and mathematics) majors to experience sexual harassment by faculty and staff.”

In 1993, the Family Medical Leave Act (FMLA) was passed in the United States with the intent of helping employees balance the demands of the workplace with the needs of the family, minimizing gender discrimination, and promoting equality for men and women. FMLA allows for 12 weeks of unpaid job protection. However, the United States remains the only country that does not allocate maternity or paternity leave as a national policy. While these are well-intended policies to promote breastfeeding and provide financial support to women of childbearing age, this reinforces traditional gender roles of women as primary caregivers and stigmatize women.

Professional isolation among women can arise from a relative lack of respect, collaboration, mentorship, sponsorship, and role models and can occur at any career stage. Gender-based isolation may stem from a history of a paucity of women in academic medicine, though this doesn’t fully explain the gaps that remain today. Generational constructs persist in salary, professional opportunities for promotion, and the ability to work flexible hours with adequate work–life integration. Parenthood can pose additional hurdles for early-career mothers and fathers. Finally, compared to men, studies show that women physicians are more likely to be unmarried or be in a dual professional marriage, do more of the domestic duties, spend more time with patients, be assigned patients with more complex psychosocial concerns, have a harder time setting limits with their time, write longer notes, adhere to patient-centered communication and clinical guidelines, and provide preventive care. All the above factors may contribute to the relatively higher burnout and depression rates among women physicians.

Underrepresented Minorities

According to the Perelman School of Medicine at the University of Pennsylvania, African Americans, Hispanics, and Native American students remain underrepresented in medical schools matriculants compared to the U.S. population despite the increasing effort to create a more diverse physician workforce. As of 2019, this underrepresentation did not change significantly since the institution of Liaison Committee of Medical Education diversity accreditation guidelines of 2009. Medical schools, graduate education programs, and health systems are increasingly aware of the additional stressors faced by minority learners and physicians and the resultant disparate impact on well-being. Efforts such as new curricula to help raise awareness of implicit bias, structural racism, how to respond constructively to microaggressions, cultural competency modules, allyship groups, mentorship and sponsorship programs, and revamping educational materials and content to better represent and include minorities, promise to reduce marginalization and isolation. However, cultural change is slow, and those experiences may also be internalized, further adding to one's emotional burden and suffering.

MENTAL HEALTH CONCERNS

Illness doesn't belong to us…It belongs to them, the patients. Doctors need to be taught to be ill. We need permission to be ill and to acknowledge that we are not superhuman.

(interviewee quote from McKevitt and Morgan's 1997 qualitative study of doctors who were ill)

The debate continues about the association between burnout and depression. There is symptomatic and epidemiologic overlap; however, burnout arises predominately from workplace problems, whereas depression carries several risk factors with impairment in numerous spheres. Burnout and depression have distinct symptoms and require different interventions. In a large study designed to distinguish associations with burnout and depression, depression was independently associated with suicidal ideation, whereas burnout was directly associated with self-reported medical errors.

Physicians are not immune to a full spectrum of mental health conditions. Mental health conditions may go underrecognized and/or inadequately treated among physicians for various reasons, including reluctance to seek treatment secondary to stigma associated with mental health problems, attempting self-diagnosis and treatment, and receiving "VIP" treatment from other health care providers. There is a general reluctance to admit any emotional "weakness." Like the general population, the most common mental disorders reported among physicians are depressive and anxiety disorders.

A 2010 literature review found that depression was observed in all stages of medical training. Depression tends to begin in medical school (up to 28% of medical students by the fourth year compared to 8% in the general population). Persistence of depression into residency and clinical practice is likely. Lifetime prevalence rates are 13% in male physicians and 20% in female physicians. There is some evidence that rates of mental health symptoms have gradually increased among physicians in the past three decades, with a meta-analysis suggesting that the rates of depression have increased on average by 0.5% per year. Most alarming are the rates of suicide among physicians, which are double that of the general population. Fewer studies have examined anxiety disorders among physicians. A small number of cross-sectional studies suggest rates of generalized anxiety disorder as high as 24% and rates of PTSD ranging between 4% and 16%.

Suicide rates among physicians are 1.4 times the general population for men and twice the rates for women, resulting in 300 to 400 physician deaths by suicide in the United States yearly. Physicians later in their careers have higher completed suicide rates compared to the general population, with male physicians at 40% increased risk and female physicians at 130% increased risk. Physicians who screened positive for depression were more like to have alcohol use disorder.

There is an emerging link between specialty area and the risk of suicide, with a 2019 review suggesting that anesthesiologists, psychiatrists, general practitioners, and general surgeons might have higher rates of suicide than other specialties. The apparent increased risk of suicide among anesthesiologists appears to be related to a combination of higher rates of psychological distress in addition to access to lethal means. Michael Myers and Carla Fine wrote on the aftermath of physician suicide in “Touched by Suicide: Hope and Healing after Loss.” Fine’s husband, a urologist at NYU School of Medicine, killed himself at the early peak of what promised to be a bright career. In particular, Fine wrote about the isolation and ostracization she faced from her husband’s medical colleagues after his death. Other accounts of physician suicide, both attempts and completions, anchor this life-altering event in the complexity of its biologic, personal, professional, social, and family webs. More recently, advocacy to reduce mental health stigma in the medical community grew out of the tragic suicide of Dr. Lorna Breen, a vibrant and respected emergency department director at New York Presbyterian

Hospital, who died by suicide during the first wave of the COVID-19 pandemic. Dr. Breen’s sister and brother-in-law started a foundation and campaign to raise awareness of the conditions within the culture of medicine. They advocated for the Dr. Lorna Breen Health Care Provider Protection Act (S. 610 and HR 1667), which unanimously passed the U.S. Senate on August 6, 2021, and aims to reduce and prevent suicide, burnout, and mental and behavioral health conditions among health care professionals. These stories drive home the magnitude of such loss and the urgency of preventative measures, education, and destigmatization of mental illness within the house of medicine. They also underscore the value of outstanding and empathic psychiatric care, not only from the perspective of the patient in treatment but also from that of the wounded healer evolving as a more sophisticated physician.

The limited published studies have estimated that the rate of physicians with mental health conditions seeking help is only between 13% and 36%, suggesting a need to address further stigma and other barriers to seeking treatment and develop a stronger culture of support. It is a shame that with many good, evidence-based treatment options available for mood and anxiety disorders, ranging from psychotherapy (individual, family, or group) and medication management to neuromodulation, medical students, trainees, and practicing physicians often fail to seek treatment. As physicians tell their stories of having a mental illness and getting treated, as barriers to treatment are reduced, and as organizations prioritize the need for mental health care, fewer physicians will have to shorten their careers or life because of mental health concerns.

Substance Misuse

Substance use is an ongoing public health concern. Just as physicians are not immune to mental health conditions, they are not exempt from the use of illicit substances. It is not uncommon to rationalize alcohol or substance use to relieve work-related distress, cope with relationship issues, or lessen physical pain. Denial is a normal defense mechanism to cope with adversity. Some may not see their misuse as problematic. Some may recognize problematic substance use but not think medical intervention is necessary, and others may acknowledge problematic use but feel unable to take further action. Most research investigating alcohol and substance misuse among physicians have been done largely through self-report surveys. Research suggests that between 5% and 20% of doctors across high-income ranges consume alcohol problematically. The data suggest that a substantial proportion of physicians have experimented with illicit substances at some point in their lifetime. The use of prescription drugs (benzodiazepines and minor opioids) occurred higher than in the general population, likely due to the ease of accessibility either through self-prescribing or through colleagues. Substance use disorder (SUD) is the most frequent cause of impairment among physicians. A national study of SUDs in a large sample of U.S. physicians was conducted in June 2011, representing specialties across all disciplines. Physicians misused alcohol at rates comparable to the general population, but opioids and benzodiazepines were misused at rates five times that of the general population. Approximately 13% of male physicians and 21% of female physicians met diagnostic criteria for alcohol use disorder, and younger and partnered physicians were more likely to be affected. There was a strong association with alcohol use disorder and a positive screen for depression, suicidal ideation, decreased quality of life, and decreased career satisfaction. Female surgeons and physicians suffered more than their male counterparts from work–life conflicts, which appeared to be associated with depression, burnout, and alcohol use disorder. On average, women progress faster to SUD after first use than men (so-called “telescoping”), and women are more likely to have a comorbid psychiatric disorder with SUD. The highest prevalence of hazardous alcohol consumption was seen among the specialty groups of dermatology, orthopedic surgery, internal medicine, and emergency medicine. Over the past 30 years, physicians consistently consumed alcohol at a similar or perhaps higher rate than their general population counterparts (15.4% vs. 12.6%). Published in 1973, one of the first major publications reporting on physician drug use and resultant impairment was “The Sick Physician.” This seminal paper called for the need to treat impaired physicians and led to the development of physician health programs (PHPs). PHPs aim to help physicians avoid the potential for serious professional and/or legal consequences of their SUD. Currently, many physicians are motivated to enter SUD treatment to avoid such consequences.

The Joint Commission (TJC) requires that medical centers have a standing committee for physician health and that physician health matters be dealt with separately from disciplinary matters. States differ in terms of the criteria of evidence required for taking action against doctors and the type of intervention. Any successful program that diverts these complaints to treatment committees also carries consequences, depending on the severity of the offending behavior. For example, a physician may need to have a chaperone present for all examinations, perform a procedure with a fellow specialist present, or be subject to random drug testing or office visits to assess record keeping. Infractions of these treatment programs can lead to a medical license suspension if the behavior continues to place patients at risk or be otherwise unamenable to treatment. The focus remains on prevention to facilitate safe, regular, and reliable participation in the program and successfully model the best ethical medical care. Each state's society provides the particular details of its program. Reassuringly, more physicians who participate in a PHP typically achieve remission and return to a fulfilling career as compared to those in the general population or without a PHP model of care. Seventy-eight percent of physicians never had a single positive test for drugs or alcohol during the first 5 years of monitoring. After successfully completing a 5-year SUD monitoring contract, 96% of participants were licensed to practice (none reported lack of licensure was related to substance use), and 91% were currently practicing medicine. Thirty-eight percent of participants voluntarily extended their monitoring at some point, and 20% reported being currently voluntarily monitored by a PHP.

Barriers to Seeking Treatment

Physicians with mental health problems might be more reluctant than nonphysicians to seek help for a variety of reasons. Major barriers include fear regarding confidentiality, the potential consequences for their career, medical board and licensure, and a belief that they can manage any symptoms on their own. Qualitative studies of physicians who have mental health problems have shown high prevalence of stigma, often driven by views that doctors should be invincible and fear of discrimination. These factors act as major barriers to seeking help and recovery and contribute to the cultural stigma in the medical profession. The most cited reasons why physicians are unwilling to seek help for depression or anxiety were lack of time (81%), fear of lack of confidentiality or privacy (41%), embarrassment (39%), and concerns of medical licensure.

To reduce fears associated with help-seeking, state medical boards and hospital credentialing processes are increasingly changing the wording of their license applications and renewals to focus on impairment rather than the mere existence of a mental health or SUD. Those physicians who lived in states where the initial or renewal application asked about mental health conditions irrespective of impairment were much less likely to seek help for a mental health condition than in states that only asked about impairment. In an effort to remove barriers to seeking help, The Federation of State Medical Boards' recommended language does not include any specific reference to a mental health condition: "Are you currently suffering from any condition for which you are not being appropriately treated that impairs your judgment or that would otherwise adversely affect your ability to practice medicine in a competent, ethical and professional manner? (Yes/No)" (https://www.fsmb.org/siteassets/advocacy/policies/policy-on-wellness-and-burnout.pdf). The hope is that physicians will seek treatment in time to prevent later impairment.

More recently, in the United States, formal avenues to address the issue of the impaired physician burgeoned beginning in 1969 (Florida) and 1971 (Texas) with the passage of "sick doctor statutes," which allowed for doctors with probable cause of distress to be referred and receive care without the actual legal allegation by a patient of harm. The number of programs providing treatment to physicians misusing substances and the accompanying statewide committees on physician health has mushroomed, and these eventually became part of every state's laws. Initially focused on SUD, physician mental health programs have grown in scope to acknowledge the impact of untreated mood and bipolar disorders, more subtle disorders of personality that may drive the sexual exploitation of patients or the abuse of colleagues, and impairment from comorbid medical conditions. Finally, the field now fully recognizes the impact of wellness and has campaigns to prevent medical illness, psychiatric disorder, or practice burnout. It recognizes the value of education of the profession at all levels of training, beginning in medical school and continuing throughout, with a constant provision of confidential access to care.

Risk Factors Unique to the Profession of Medicine

> The expectation that we can be immersed in suffering and loss daily and not be touched by it is as unrealistic as expecting to be able to walk through water without getting wet.
>
> Dr. Rachel Naomi Remen.

The demands of clinical care of patients can prove exceptionally challenging and worthy of the respect bestowed on the healing professions by society. The regular grief involved with exposure to death and dying of longstanding patients; feelings of helplessness at not knowing which path to take with a patient and of limitation in the face of human nature; exhaustion from working long or emotionally taxing hours with the accompanying deprivation of sleep; and invitations to allow responsibility for clinical care to assume the weight of guilt when patients' conditions decline precipitously all contribute to the occupational hazards of the profession. Beyond the immediate clinical level, complex communications with perfection-seeking colleagues, demanding patients, or cultural and language boundaries unfamiliar to the physician add to the background stress of health care delivery. A professional culture unwelcoming of medical error, high levels of work–home interference, limited control over call schedules, and an increasing educational debt burden further magnify the impact of any given stress. Just as each of these factors creates stress, each also illuminates directions for better physician self-care and for change within the culture of medicine.

Physician Impairment, Medical Boards, and Licensure

It is not uncommon for many physicians to carry disability insurance to protect themselves and their families. Many physicians have been protected from disability-based discrimination since 1973 under the Rehabilitation Act, which has comparable provisions but less comprehensive coverage than the Americans with Disabilities Act (ADA) of 1990. The term "impaired physician" is often used to describe a disability related to psychiatric illness and alcohol use disorder or SUD. Whereas the ADA's legal definition tends to be more inclusive of physical, mental, and behavioral limitations, impairment is present when a physician is unable to perform in a manner that conforms to acceptable standards of practice, exhibits serious flaws in judgment, and/or provides incompetent care.

SUDs account for the most frequent cause of impairment among physicians. It is estimated that as many as 15% of practicing physicians suffer from SUDs at one point in their career, almost equivalent to the rates for the general population. An important distinction is that the drugs chosen by physicians are more often alcohol or prescription medications and less commonly illicit drugs.

Mental illnesses such as personality or affective disorders and interpersonal issues within or outside the workplace could disrupt patient care. Numerous programs have been established across the country to help evaluate and treat disruptive physicians. Remediation programs can help identify and offer education.

Each state’s medical practice act explicitly states that each physician must remain physically, mentally, and emotionally competent to practice medicine and to self-report if any of these are compromised. An affirmative "duty to report" is required of a physician health care provider if public safety is imperiled. The National Practitioner Data Bank must also be notified of disciplinary action taken against an impaired physician.

In a national survey of physicians, only 45% of respondents indicated that they had notified the state licensing board of a colleague who was incompetent or felt to be impaired. Yet almost all indicated that these individuals should be reported. An ethical requirement to report must be balanced against the obligation to maintain confidentiality.

Countertransference Issues

Physicians opens the door to a particular set of countertransference issues that can impede the provision of a high standard of care. For instance, the treating physician might fear that the physician-patient is smarter—which may well be the case—thus minimizing a contribution to the careful management of the patient, which is often more emotional in nature than an issue of “book-smarts.” The treating physician (and their peers) can feel compelled to join the physician-patient in thinking about the clinical struggle as purely an intellectual one, thus depriving the patient of crucial emotional clinical data. Most significantly, the treating physician can identify with their health care professional colleagues and normalize their distress as secondary to the trials of medical training, thus stretching the realm of “normal” to miss significant avoidance of psychic pain in need of mending. The treating physician might forget to think of suicide as a central concern, given the mutually shared value of life in the profession; wish not to see a colleague as impaired; mistake the patient’s high intellect as an absence of serious disturbance; or wish to avoid destructive feelings as the provider. All of these factors can lead the treating physician to buy into the medical student or physician-patient’s wish to avoid treatment and the wealth of uncomfortable feelings likely experienced, even if with a sensitive, compassionate listener. Buying into a medical student or physician’s wish to avoid treatment when indicated will only defer the struggle until a later time, when it may be much more difficult to handle, if only because many more people are then affected. It also postpones the very relief that is longed for physician- or medical student-patient musters the courage to pick up the phone and seek consultation for issue.

LIFE PHASES IN MEDICINE

Medical Students

Students enter medical school with similar mental health status to same-age peers. Their overall mental health declines in medical school. This may suggest that medical education itself contributes to student distress. A significant stressor from college to the first year of medical school is “not being the smartest” in their class. While some stress may enhance academic performance, high stress levels are associated with burnout, depression, and somatic distress. Major stressors for medical students are academic workload and volume of information, studying and time management, work–life conflicts, peer relationships, poor student guidance/support, evaluative pressures, finances, lack of time for oneself, responsibility, and the need to succeed.

Research on burnout prevalence rates among medical students ranges from 28% to 55% for high EE, 26% to 38% for high DP, and 27% to 84% for low PA. A 2009 national survey of U.S. medical students reports burnout strongly correlated in the first and second year with perceived level of dissatisfaction with support from staff and medical school faculty. During the third and fourth years, burnout was more closely associated with poor clerkship organization and exposure to cynical trainees. One episode of mistreatment by faculty or resident has been reported by 83%, with recurrent mistreatment as another contributing factor to burnout. Most commonly, the reported mistreatment is in the form of public humiliation from other staff, such as nurses, peers, and clinical faculty. Long hours studying, financial debt, interpersonal struggles, witnessing stressful clinical events, and lack of adequate sleep contribute to burnout. There are multiple potential ramifications of burnout for medical students, including poor academic performance, loss of empathy, negativity toward patients, impaired academic integrity, SUDs, and suicide. There is evidence that burnout is strongly associated with thoughts of leaving medical school. Medical students with burnout often don’t seek care for fear of discrimination, stigma, and loss of confidentiality. Lotte Dyrbye and colleagues suggested that the reluctance to seek help is related to the misperception of being weak. Unfortunately, this may lead to maladaptive coping with excessive alcohol use. In a large national study of 2,682 medical students, students with burnout were more likely to report dishonest behavior and have less altruistic views toward society. Decreased empathy and professionalism associated with burnout continued in the transition from medical school to residency.

Rotenstein and colleagues did a systematic review and meta-analysis of 195 studies in 2016 based on medical students from 47 countries. Based on assessment modes, they estimated prevalence rates of depression of 27.2%, ranging from 9.3% to 55.9%. From 24 cross-sectional studies of medical students, they found prevalence rates of 11.1% expressing suicidal ideation. The Center for Disease Control and Prevention report in 2019 ranked suicide in the top three leading causes of death among teens and young adults, noting an increase of 36% among persons aged 20 to 24 years from 2000 to 2017. A 2019 report from the AMA’s Council on Medical Education indicated that “medical students are three times more likely to commit suicide than the rest of the general population in their age range in other educational settings.” Associated risk factors for suicidal ideation include depressive symptoms, burnout, psychological distress (anxiety, irritability, loss of control), living singly, having poor health and/or economic status, and family history of mental illness and suicide. The prevalence of depressive symptoms and suicidal ideation is significantly higher through medical school than in other postgraduate training. Excessive consumption of alcohol is common among U.S. medical students and strongly related to the distress experienced by learners. Younger age and being single are known risk factors for problematic alcohol use among students, similar to the general population. Educational debt was noted to increase alcohol consumption in a mass survey sent out in 2012 to all students listed in AMA Physician Masterfile; from 1995 to 2014, the average cost of private medical schools increased by 209% compared to an increase of 286% cost to attend public medical schools.

In their systematic review of 28 interventions designed to promote medical student wellness, Wasson and colleagues found that interventions involving the learning environment were associated with improved wellness among medical students. These include converting to a pass/fail grading system, available and easily accessible mental health programs, mind–body skills training programs, curriculum restructure with balancing between clinical and nonclinical learning environments, and implementation of formal faculty advisor/mentor programs. These and other studies suggest that addressing the environment of medical education and the culture of medicine is critical if we hope to improve the well-being of rising physicians. The AAMC's recommendations regarding health services for medical students remain relevant from 1992 for mental health and SUDs. Recommendations suggest providing all students with access to confidential counseling by mental health professionals, confidential policies protecting students' mental health records, and clear administrative mandates for evaluation and treatment. However, a lack of resources often prevents schools from implementing these guidelines. Access to community providers, long wait times, insurance and self-pay issues, concerns about future employability and licensure, and the scramble to orchestrate evaluations for medical leave and clearance for returns remain a challenge. Another barrier includes self-stigma, which describes the phenomenon whereby individuals have a loss of self-esteem and self-efficacy as they adopt and internalize the external social stigma of needing help or appearing weak. It has been long suggested that physicians stigmatize mental illness more than the general population. Furthermore, some learners may be on parents' health insurance plans with concerns about the privacy of receiving mental health treatment. Finances, time off, and co-pays pose another challenge for learners. Some researchers suggest incorporation of alcohol reduction programs, student debt reduction, wellness curricula, mindfulness programs, teaching cognitive reappraisal, and small group facilitated discussions. Ideally, well-being programs would be integrated into existing curricula with buy-in from students and faculty to maximize engagement and minimize further resentment toward overloaded schedules.

During the unprecedented era of the COVID-19 pandemic that emerged in February 2020, lockdown and quarantine were universally implemented by March 2020 to contain the spread of infection. Universities closed classrooms, and medical education shifted across the country to digital learning lectures for social distancing and safety, barring a student from in-person clinical work. Numerous studies have established the negative impact of social isolation on mental health. Before the COVID-19 outbreak, several studies suggested a high prevalence of stress, anxiety, and burnout among medical students. The impact of the relatively rapid adaptations to revise course structures, new online requirements, social isolation, and reduced student–patient and student–faculty interactions created even more challenges for preclinical students. Third- and fourth-year students experienced disruptions in their clinical clerkships while navigating applications for residency training and virtual interviews.

Crises, whether a pandemic such as COVID-19, natural disasters such as Hurricane Katrina, school shootings, or a campus suicide, demand immediate response to provide safety to the community and restore normalcy (even if a “new normal”) as soon as possible. In the event of a suicide, it is critical to have a plan of action and a postvention team already in place. The AAMC has created a helpful toolkit for schools called “After a Suicide: A Toolkit for Medical Schools.” Assembling a crisis response team and creating a crisis response communication plan are critical first steps. The toolkit offers guidance on how to gather the facts, inform the emergency contact person/family, share the news, help medical students cope, and memorialize the deceased. In the case of a suicide of a learner, it is of utmost importance to identify whether there are other at-risk students or trainees and to ensure their safety. In addition, it is necessary to reach out to individuals who are known to be close to the deceased person, such as roommates, significant others, close friends, and academic mentors. It is important to provide psychoeducation about typical reactions to trauma, including disturbances in sleep, appetite, energy, and interest, and to emphasize the importance of self-care in returning to precrisis levels of well-being and available resources. Professional counselors and chaplains may offer group debriefings and a safe place to process the inevitable mix of emotions arising after such tragic events. Leadership typically allows and encourages their learners to take the time needed for emotional processing without penalty.

Mr. A was a 24-year-old Korean American who had just started his third year of medical school. He was a popular and well-liked student who had been active in the Christian fellowship. He was an excellent athlete and often played pickup basketball games with his peers. His classmates saw him as affable, kind, and funny. However, behind closed doors, Mr. A struggled with severe self-doubt and insecurities. Despite his high grades, he thought that his peers knew more than he did, and he questioned whether he could get into a competitive plastic surgery residency. Two weeks before the first step of his national board examinations, his girlfriend of 1 year broke up with him and started dating one of his close friends. Mr. A was devastated. The breakup added to his feeling that he was inferior to his peers. Mr. A had come from a high-achieving family, and he was expected to do the same. Despite evidence of significant mental illness in his family, it was never acknowledged or discussed. The day after Mr. A got his board scores back, he was found dead in his dorm by his roommate. He had hanged himself from his door frame. His roommate notified campus security, leading to the activation of the postvention response plan. The family was contacted immediately, and they expressly indicated that they did not want the cause of his death to be disclosed, which was honored. After the family was notified, the public affairs office of the medical center sent out an email to the entire medical community, alerting them to Mr. A’s untimely death. In addition, the dean of student affairs emailed the medical student community, including medical school faculty, confirming that a medical student had died and the availability of staff to speak with any student wishing to do so. The same day, the postvention team set up shifts at the medical student dorm from 9 in the morning until 9 at night to provide psychological first aid for any students or staff. The counseling center contacted medical students who were known to have had suicidal thoughts in the past or who were being treated for severe depression to offer additional support. In addition, the counseling center posted flyers in the dorm announcing a support group for those students who were especially close to or distressed by Mr. A’s death. The leader of the Christian fellowship also held a special session to help its members talk about the recent loss of Mr. A. The dean of students also met with medical school faculty who were involved in teaching Mr. A to provide support and review common reactions to a crisis so that they could be on the lookout for themselves and students exhibiting any symptoms. In addition, the dean stressed to the faculty that some students might need extensions on examinations or other accommodations. Meanwhile, the postvention team met daily to discuss the effectiveness of their efforts, determine whether any adjustments were necessary, and provide the necessary support to each other. Over the subsequent weeks and months, the medical school community resumed its usual functioning. However, the postvention team continued to meet to discuss improvements for future events.

Promoting Wellness

Well-being and wellness are often terms used interchangeably. Well-being is a state of being comfortable, healthy, or happy. Wellness includes a spectrum of domains that include social, physical, emotional, intellectual, spiritual, financial, and environmental. Resilience refers to the ability to rebound or grow from a stressful experience, often by connecting with others or with one's humanity or beauty rather than staying stuck in the stress response; resilience is believed to contribute positively to well-being and wellness. Over the past decade, there has been increasing awareness and recognition of mental health challenges faced by medical students, residents, and physicians and the importance of prevention. The AAMC, Institute of Medicine, and the Liaison Committee on Medical Education unanimously support and/or require policies and programs to promote well-being in medical school. The AAMC longitudinally tracks issues critical to medical education using student surveys. These voluntary surveys include the post–Medical College Admission Test survey, the Matriculating Student Questionnaire, the Medical School Year 2 Questionnaire, and the Medical School Graduation Questionnaire. Previous approaches to medical student mental health were largely reactive. Early initiatives relating to medical student mental health primarily focused on improving access to care, educating students on mental health issues, and addressing the stigma associated with treatment. In recent years, some medical schools have taken a top-down approach, incorporating into the curriculum basic education about personal mental health assessment, mindfulness strategies to build resilience, teaching devoted to managing and coping with errors in practice, gender- and racial-sensitive perspectives on the experience of being a learner and practitioner, and providing more resources and tutoring. Many medical schools have designed and implemented programs that promote optimization of learner well-being. For example, the University of Pittsburg School of Medicine addresses the challenges of access, privacy, and stigma with a dedicated mental health team. The medical school financially supports this team to provide care for their 560 medical students and 360 graduate students. A faculty psychiatrist receives 30% salary support, and a full-time master's level psychologist receives 100% salary support. This mental health team is further supported by the Student Health Advocacy Resource Program (SHARP), a confidential peer-counseling referral and advocacy service for medical students. Ten peer counselors provide guidance and assistance with life transitions, family stress, relationship difficulties, depression, anxiety, academic concerns, leaves of absence, and substance misuse. SHARP is organized by medical students and supervised by faculty from the psychiatry and medicine departments. There are other schoolwide initiatives for mental health and stress reduction for students at the University of Pittsburg. The Faculty and Students Together program helps students transition to medical school, be assigned faculty advisors, and observe how faculty balance work and life choices.

The Saint Louis University School of Medicine began an initiative in 2006. They developed a simple well-being model with three components. The first component was to mitigate unnecessary stressors and enhance the learning environment. The second was to teach students stress management skills and encourage the use of the available psychological and emotional support resources. The third component was to create more opportunities for students to find meaning in their work. This was followed 2 years later with interventions and changes in the learning curriculum and preclerkship environment. The grading system was changed to pass/fail, and a 10% decrease in hours for all curriculum courses and longitudinal electives and theme-based learning communities were instituted. Over the next decade, they found a noticeable reduction in depression and anxiety rates in the first-year class.

Well-being programs in medical education have been seen throughout the country. Components include mind–body skills, a longitudinal curriculum focusing on personal development with mentoring wellness activities, weekly 30-minute lectures on the neuroscience of yoga, relaxation and breathing exercises, resiliency mindfulness programs, and appointing faculty "wellness advisors." The University of San Diego implemented the Healer Education Assessment and Referral (HEAR) program to educate medical and pharmacy students, house staff, faculty, and trainees about burnout, depression, and suicide by providing confidential assessments. Other resources available to medical schools include free materials developed by the American Foundation for Suicide Prevention (afsp.org) to prevent burnout, depression, and suicide among health care learners.

Figure 30.8–1 depicts the various barriers, interventions, and outcomes at the medical school level, though the same infrastructure should be applied during postgraduate training. Wellness and well-being efforts cannot start and end at the medical school level. Training directors and employers must prioritize well-being at each phase of education and practice by measuring it, creating an infrastructure to address the barriers to well-being with adequate resources, and implementing strategies and tactics appropriate for the particular developmental level and specialty. Faculties who supervise and educate are instrumental in modeling self-care and a culture of support.

Physicians in Training

Residency and fellowship have long been thought to be stressful times in a physician's career. The literature is mixed regarding whether resident physicians begin residency with burnout. Most studies find that the rates of burnout significantly increase during the first year of residency. Interns with passive coping styles and struggles with perceived stress were more likely to have higher levels of emotional exhaustion and DP. Some contributing factors to burnout are long hours, financial hardship, sleep deprivation, and emotionally difficult situations with patients, families, and colleagues. Burnout symptoms contribute to more medical errors and decreased professionalism. In July 2003, citing a concern for patient safety, the American Council of Graduate Medical Education (ACGME) implemented a resident duty hour restriction of an 80-hour workweek for patient care and didactics. In 2011, after studies showed no improvement in burnout or self-reported medical errors, the ACGME instituted limited single-shift hours to less than 28 hours and a required 1 day off per week. The attention to burnout among residents is of significant interest in how professional identity and healthy habits are shaped during this critical time. The ACGME has made resident wellness a top priority, not only with work hour restrictions but also by alerting program directors of their role in assessing and promoting the wellness of their residents.

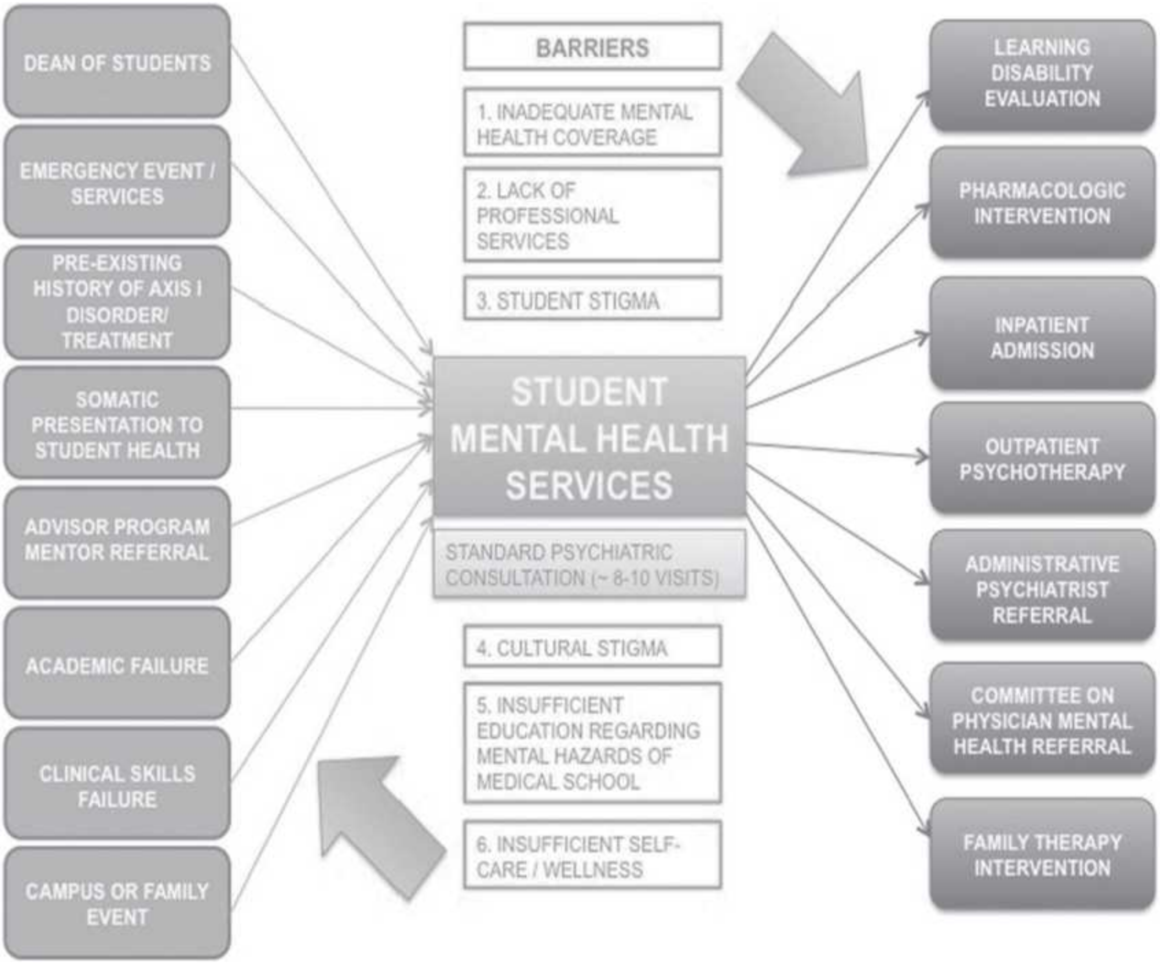
DEAN OF STUDENTS
EMERGENCY EVENT / SERVICES
PRE-EXISTING HISTORY OF AXIS I DISORDER/ TREATMENT
SOMATIC PRESENTATION TO STUDENT HEALTH
ADVISOR PROGRAM MENTOR REFERRAL
ACADEMIC FAILURE
CLINICAL SKILLS FAILURE
CAMPUS OR FAMILY EVENT
BARRIERS
1. INADEQUATE MENTAL HEALTH COVERAGE
2. LACK OF PROFESSIONAL SERVICES
3. STUDENT STIGMA
STUDENT MENTAL HEALTH SERVICES
STANDARD PSYCHIATRIC CONSULTATION (~ 8-10 VISITS)
4. CULTURAL STIGMA
5. INSUFFICIENT EDUCATION REGARDING MENTAL HAZARDS OF MEDICAL SCHOOL
6. INSUFFICIENT SELF-CARE / WELLNESS
LEARNING DISABILITY EVALUATION
PHARMACOLOGIC INTERVENTION
INPATIENT ADMISSION
OUTPATIENT PSYCHOTHERAPY
ADMINISTRATIVE PSYCHIATRIST REFERRAL
COMMITTEE ON PHYSICIAN MENTAL HEALTH REFERRAL
FAMILY THERAPY INTERVENTION

FIGURE 30.8–1. Flow diagram of various intervention avenues and outcomes.

Similar to entering medical students, studies show that medical residents are at higher risk for depressive disorders, depressed mood, and suicidal ideation than their peers in the general population. A meta-analysis showed prevalence rates of depressive symptoms among residents to be 28.8%, ranging from 20.9% to 43.2%, depending on screening instruments. This raises concern for developing greater long-term morbidity and risk for suicide even after training in these learners. The spotlight on suicide was elevated when, in 2014, two internal medicine interns committed suicide. Among residents in ACGME-accredited programs from 2000 to 2014, the leading cause of death was neoplastic disease, which raises the concern that trainees sacrifice medical care that might have discovered problems when still treatable. The second most prevalent cause of resident death was suicide. Suicide is the single most common reason for male physician mortality during training and the second most common reason for female physicians in training. A majority of suicides occurred during years 1 and 2 of training. Barriers to seeking help include time constraints and inconvenience, documentation and confidentiality concerns, cost, stigma, and future employability.

Strategies to prevent and respond to the emotional burdens associated with residency and fellowship training start with having an infrastructure dedicated to well-being, support, and mental health care. Deans of well-being and teams dedicated to measuring well-being and distress take the pulse of the community, encourage efforts that address drivers of distress and burnout, and build community and support. Examples include free and accessible mental health, academic success, and disability services; teams dedicated to the assessment and remediation of struggling trainees; peer support training and infrastructure to deploy peer counselors who are aware of resources; engagement of program directors and chiefs in promoting peer support; support groups; allyship groups; childcare, paternal leave, and lactation facility options; and having a crisis response plan and team in place.

Dr. E presented for consultation in the third year of his psychiatry residency for complaints of depression and burnout. He lamented that his passion for learning and zest for life had each atrophied and dried up since the beginning of his PGY-II year. He could not say for sure why he thought this descent had occurred, although he readily identified a couple of immediate possibilities: failure in his love life relative to his high expectations and experiencing himself as a cog within the greater wheel called profit-driven medical care. These factors left little protected time for his formerly passionate thirst for knowledge. He also owed more money than he could calculate to credit cards and student loan agencies—when would a day come that he could ever pay this debt or live in the expensive city of his training on such a thinly stretched salary? While he enjoyed basic invitations to know more about the human condition in sickness and health and the ubiquity of human suffering, he found himself directionless and listless in the bigger picture of his life. His prior experience in psychotherapy had soured, which left him in a hypocritical stance toward his patients and his work: how could he recommend and/or perform a procedure that had so wounded him?

Consultation over several sessions focused on several core concepts: the establishment of a supportive, mending therapeutic alliance; the restoration of hope; the normalization of Dr. E’s plight relative to his stage of training and deep capacity for empathy; and the exploration of pharmacologic options to manage his mood and anxiety. Numerous treatment options were considered, including participation in a group therapy program for physicians. Over the course of the year, Dr. E made steady gains in his life and his mood. He broke up with his on-again, off-again lover with whom he had spent years in a passionate bad fit and worked on some searing emotional pain that stemmed from his family of origin, identifying his mother’s history of alcoholism and his father’s covert history of PTSD from Vietnam as significant drivers of his wish to become a psychiatrist. He reconnected with his brother in new and rewarding ways, fell in love with a fellow resident from the emergency medicine program, and refined a professional path focused on intensive inpatient hospital work with physicians in crisis.

Early Career

The early career phase spans from medical school and postgraduate training to the first 5 years out of training. The transition from training to practice may be a particularly challenging time in trying to establish a professional identity and adjust to a new structure of work environment, uncompromising schedules, and learning the business aspects of medicine. Newly trained physicians may be reluctant to ask for help or demonstrate uncertainty, invoking tendencies to compare themselves to others, lack of belonging, and the experience of imposter syndrome. This phase coincides with the desire for romantic partnerships, community building, and starting a family. For single women, the reproductive complexities with their biologic phase in life pose alternative routes of in-vitro fertilization, surrogates, and adoption. These personal challenges are often superimposed on workplace expectations for productivity. Systematic factors of flexible hours, job share, mentorship, role models, and support networks are often lacking yet necessary for academic advancement. The AAMC has identified a “good work–life policy,” which includes paid leave offered to full- and part-time faculty, no required length of prior service to qualify for leave, 12 weeks of paid leave, and a 6-month total (paid and unpaid) leave. The Family and Medical Leave Act requires employers to provide 12 weeks of unpaid family leave once the employee has worked 12 months. However, the United States has no federal paid parental leave policy and only a minority of states offer paid parental leave. Consequently, parental leave time policies are inconsistent and often come with constraints under the discretion of the institutional and departmental leadership. Return to work following parental leave can be a difficult transition period with childcare availability, feeling pressure to return to work while on leave, makeup time with extra calls or clinic time, and feeling discriminated against. Many women lack access to lactation facilities and inadequate time for pumping breast milk.

Strategies to support physicians during their early career include identifying and responding to bias (especially women and minorities), mentorship, sponsorship, career and life coaching, implementing early and effective onboarding, and providing faculty development workshops on time management, self-compassion, and setting limits. Key initiatives include (1) support groups to normalize the stressors; (2) trained peer supporters: (3) training leaders to be authentically open to input, focused on the need for flexible schedules/telecommuting and advocating for staff’s well-being and work–life integration; (4) developing enough cushion to allow for sick leave or parental leave; (5) making childcare leave and stop-the-clock policies opt-out versus opt-in; (6) having programs that reduce on-the-job stress (e.g., emergency sick child care programs, on-site child care, elder care, daycare hours in line with physician schedules); (7) having adequate lactation facilities and allowing time that is counted toward relative value units (RVUs) for breast pumping; (8) ensuring that promotion and tenure are the same for all faculty regardless of leave use; and (9) recognition and rewards for contributions to the team or department.

Dr. A was a junior faculty member of a surgery department at an academic institution who presented to the psychiatry outpatient clinic after her husband urged her to seek treatment for possible postpartum depression. In the initial consultative appointment, 5 weeks after giving birth, she exhibited insomnia, tearfulness, loss of energy, low mood, anhedonia, and difficulty attaching to her twin babies. The latter made her feel shame and guilt, as she had desperately wanted to be pregnant yet now did not feel able or worthy of caring for her precious newborns, though she denied suicidal ideation. She explained how she dedicated her life to becoming a surgeon and delayed having children because there was never a right time. She and her physician spouse had been trying to conceive, but it was difficult. It felt strange not to be able to achieve this goal more easily when she had already achieved so much in her career. She assumed that her colleagues, mostly men whose wives carried out most of the domestic responsibilities, wouldn’t relate to the level of stress she felt with this, and it would not be something she would feel comfortable discussing anyway. She prided herself on being tough and not complaining. Ultimately, with the second round of in vitro fertilization, she became pregnant. The pregnancy went okay after she figured out strategies to manage first-trimester nausea and vomiting (a particular challenge during her longer surgical cases), though she increasingly had symptoms of fatigue, insomnia, and free-floating anxiety. She had hoped to complete a book chapter and research manuscript before the end of her pregnancy but was struggling to manage her clinical responsibilities by the time she went into labor. Past history included symptoms of depression and anxiety in college during a time of interpersonal stress. She had responded to a combination of sertraline and cognitive behavioral therapy, but she attributed most of the benefit to the therapy. Subsequently, she tapered off the sertraline and found that a strategy of daily exercise, meditation, healthy eating, and strong organizational skills kept her reasonably stable throughout medical school and training. She was highly driven and very action oriented. Although medical school and surgery training were stressful, she excelled and loved being a surgeon.

After diagnosing postpartum depression, the psychiatrist recommended that she initiate sertraline, as she had done well with it previously, and it was known to be relatively safe with breastfeeding.

She had partial improvement in depressive symptoms a few weeks after initiating sertraline, and over time, she felt more attached to the babies. Psychotherapy this time focused on allowing, accepting, and tolerating her emotions and imperfections with self-compassion, as well as the need to widen her support network. Her husband attended a few of these sessions, during which she began to acknowledge and communicate her needs, and they negotiated the sharing of domestic responsibilities. She was connected to a local peer mentor program and joined an online support group through Postpartum Support International. She was encouraged to extend the maternity leave to 4 months. When she returned to work, she talked with her Chair about options to remain on the tenure track via a “stop the clock” program at her institution, an idea shared by her peer mentor. The institution had lactation facilities; she was able to continue breastfeeding between surgical cases and, on clinic days, was given a reduced schedule to accommodate the breastfeeding. Over time, she found ways to integrate her personal needs and career goals rather than suffering in silence.

Midcareer

This has been defined by some as 11 to 20 years out of training with a general age group between the ages of 40 and 55 years. On a personal level, raising school-aged children and teenagers with social activities has its own set of developmental issues. Focus on child-rearing may overshadow attention to one’s partner, leading to separation or divorce. Many can get caught between raising children and caregiving to aging parents’ health issues. Balancing dual-career households with family can be difficult. Personal health issues may arise. The career trajectory for women may slow down compared to male counterparts.

Many of the solutions to supporting midcareer physicians are like early-career physicians. In addition, practices should find ways to lessen the burden of the EMR by providing adequate allied health staff support, optimizing the “right person for the right job,” and workflows. The AMA Steps Forward has many helpful modules directed at saving time and reducing the less rewarding aspects of clinical responsibilities. Departments and leaders should have zero tolerance for abusive behaviors at work. Institutions should facilitate and fund organizational science. Providing formal space for physician peer interactions, for example, physician lounges, and encouraging peer support groups and programs help diminish the isolation and exhaustion that can be felt during this very busy time in life. It is essential to encourage “rest and recharge” breaks and self-care and normalize help-seeking and mental health treatment. Midcareer is a time when physicians are often pulled in many directions, and burnout, depression, anxiety, and SUDs are common.

Dr. L was a 50-year-old married sports medicine physician in a small, private group practice in a midsized town. In addition to a very active practice, he was the physician for a national sports team and engaged in his children's activities, such as helping to coach soccer and attending their games. A few years prior, his father, who lived nearby, was diagnosed with dementia, so he also began helping navigate his father's care. Just before the COVID-19 pandemic, the office manager for his private practice resigned, his father's dementia progressed, and his mother's health also started failing to the point that he needed to either look for an assisted living facility or have his parents move in with him and his wife. Knowing that his parents would resist moving to an assisted living facility, he and his wife began to work on a remodel to accommodate his parents. To further add to the stress level, his teenage daughter began showing signs of an eating disorder with excessive exercise and food restriction. Over time, the one glass of wine with dinner became two and was supplemented with a cocktail before dinner and sometimes before bed if he couldn't sleep. He became more irritable with his wife and kids. His office partner, Dr. E, observed him being impatient and less compassionate than usual with staff and patients and so asked him how he was doing, commiserating that she knew it was a stressful time with their practice and his life circumstances. Dr. L said he was fine but was just having a little difficulty sleeping. He said he would be ok if he just got a better night's sleep and asked Dr. E if she would prescribe some zolpidem because he didn't have time to see his primary care physician (in fact, he hadn't gone to his doctor in years). Dr. E sympathized with his dilemma but worried about the consequences of doing so in terms of her licensure, especially if there was a bad outcome with one of Dr. L's patients. She convinced Dr. L to make the time to be evaluated by his doctor. She wondered if she was obligated to report him but decided not to do so.

At the appointment with the internist, Dr. L was noted to have high blood pressure. The internist expressed concern about the alcohol use. She frankly discussed the duty to report if there were any concerns about impairment. However, they both agreed that the impact of his maladaptive drinking had not impaired Dr. L's functioning as a physician yet. She offered to connect him with another physician who was open about his experience with a PHP and mentored other physicians. She further proposed a “warm handoff” for Dr. L to meet briefly with the licensed social worker in their integrated behavioral health team. After meeting with the social worker, he began to acknowledge the realities of his stress level and maladaptive responses. He didn’t want to threaten everything he had built, especially his involvement with the national sports team. He agreed to regular counseling sessions for accountability and to eliminate alcohol; he committed to seeking further treatment with a PHP program should he have difficulty changing these behaviors or managing his understandable stress level.

Late Career

Late-career physicians can be defined as those over 55 years of age. Work–life integration, a constantly changing health care environment with the requirement to master new technology, and providing quality patient care with increased regulation and decreased autonomy all impact career satisfaction. Possible cognitive and physical decline in one’s own personal and spousal health may provide an impetus for early retirement. Physicians prefer gradual retirement with flexible and fewer hours, part-time employment, reduced on-call duties, and job share. The key to successful retirement is planning. In her chapter entitled “Late Career Solutions. Transitional Zone–Retire or Retread,” Dr. Suzanne M. Conolly discussed the five Fs: fitness, family, friends, fun, and financial planning. These components cultivate a sense of well-being with the importance of being active with meaningful activities, that is, volunteer work, mentoring, teaching, writing, and community networking. Barriers to retirement planning include poor financial management, health insurance, and lack of interest or activities outside of medicine.

Dr. B was a 60-year-old hospitalist and associate professor of medicine at an academic institution. A beloved teacher, she had mentored many students and trainees over the years. Patients appreciated her thoughtful approach to their care, and she remained dedicated to maintaining and expanding her skills and knowledge to keep abreast of new developments. She had served on many committees, often in leadership roles, including Division Chair. Together with her same-sex partner, she had raised three children who were now out of the house and pursuing their own careers. Other than recurrent, mild depression, stable on venlafaxine 75 mg daily, she remained relatively healthy for much of her career. She had seemingly managed to do it all. Yet, not long after her 60th birthday, she began noticing that she was progressively more emotionally exhausted, physically tired, lacking efficiency in her work, and overwhelmed. Feeling irritable and not enjoying her work, she felt increasingly annoyed with the requirements of the EMR and dealing with angry or opioid-seeking patients. She was frustrated with the realities of a much busier service, including greater numbers of agitated and violent patients, with a lack of support to go along with the realistic clinical demands. She thought about cutting back or retiring almost every day, though due to the cost of her children's undergraduate and graduate education and neglecting to attend to a financial plan, she didn't think she could afford to retire, which added a sense of being trapped. In the midst of this, she became ill with a flu-like illness, and in the course of the workup, lung cancer was incidentally discovered. Luckily it was found early, enabling a surgical cure. The medically required 6-week pause for surgery triggered some unexpected outcomes. First, her colleagues and friends showered her with kind words and offers of help and support. At the urging of her partner, who was aware of both her history of depression and recent symptoms of burnout, she agreed to meet with a psychiatrist. The psychiatrist offered an increase in the venlafaxine dose. However, Dr. A preferred to wait on making any medication changes now that she was away from work, which the psychiatrist agreed was reasonable. They continued to meet biweekly, during which time Dr. A began to reflect on how little time she devoted to past hobbies, such as playing piano and gardening, and the importance of taking time to be with friends or just having fun. She contemplated her mortality, values, and priorities. She appreciated that she valued the teaching and mentoring, though she had lost a sense of autonomy with the current demands at work. After meeting with a financial planner, she realized that she could afford to reduce her work hours and still manage financially. They worked out a plan of gradual reduction of work hours over the next 5 years. The prior symptoms of irritability and anxiety dissipated without the need for an increase in her antidepressant. Upon return to work, she maintained monthly visits with the psychiatrist to monitor the need to adjust her antidepressant medication. She talked with her department leadership about the clinical gaps, and although there were still problems, she also noted constructive efforts from the institution. With intention, she managed to reorganize enough to build her outside interests, reappraise how she viewed and approached her clinical work to allow meaning and purpose to flourish, and build in time for rest and recharge. She felt more in control of her schedule and choices and could see a path to a fulfilling retirement.

FUTURE DIRECTIONS

Much as the military has had to confront an epidemic of suicides since September 11, 2001, the medical profession is at its own tipping point. A window of opportunity exists now to expand upon the work of so many individuals, medical student mental health services, families of medical students or physicians who have died by suicide and are still grieving in isolation, and the pioneering work of the American Foundation for Suicide Prevention. If much of the mental duress that occurs during training stems from the emphasis on science over intuition, rational thought over feelings, and taking care of one's patients over taking care of one's self, then it makes sense that much of the risk of training comes from impossible expectations of omniscience, suppression of personal emotion in the name of patient care, and extinguishing of one's capacity to tend to physical, mental, and spiritual health.

As outlined in Tait Shanafelt's reflection on where physician well-being was and where it is going, health care professionals must cultivate human qualities (over deity or hero qualities) of vulnerability and growth mindset (over perfectionism), work–life integration (over no limits on work), self-compassion (in addition to self-care), community (over isolation), and meaning and purpose (over performance or frustration). Organizations must take action (over simple unawareness or simply being aware), focus on the needs of people (over just institutional or patient needs), be flexible (over rigidity), prevent distress and cultivate professional fulfillment (over ignoring distress or just responding after the events), emphasize alignment of organizational and professional values (over unfettered autonomy or carrots and sticks), create a shared responsibility (over blame or neglect), place value on investment (vs. return on investment or ignorance of economic impact), encourage physician and administrator collaboration (vs. functioning independently or with an adversarial relationship), and redesign the system to have sustainable workloads, provide coverage when physicians are ill, and incorporate appropriate breaks and rest. Commitment to developing leaders who value well-being, listening to input from health care professionals, optimizing team-based care, and developing and preserving a trusting environment will be essential to a healthy health care organization. Success will require an infrastructure with enough resources to support these organizational changes.

ACKNOWLEDGMENTS

Khleber Chapman Attwell, M.D. and Graciana Lapetina, M.D. were the authors of the previous edition of this chapter, and portions were retained for this revision.

FURTHER READINGS

Aaronson AL, Backes K, Agarwal G, Goldstein JL, Anzia J. Mental health during residency training: assessing the barriers to seeking care. *Acad Psychiatry*. 2018;42:469–472.

Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med*. 2014;12:573–576.

Brady KJS, Trockel MT, Khan CT, et al. What do we mean by physician wellness? A systematic review of its definition and measurement. *Acad Psychiatry*. 2018;42:94–108.

Brennan-Wydra E, Chung HW, Angoff N, et al. Maladaptive perfectionism, impostor phenomenon, and suicidal ideation among medical students. *Acad Psychiatry*. 2021;45(6):708–715.

Brower KJ, Riba MB. *Physician Mental Health and Well-Being: Research and Practice*. Springer International Publishing AG; 2017.

Bynum WE 4th, Artino AR Jr. Who am I, and who do I strive to be? Applying a theory of self-conscious emotions to medical education. *Acad Med*. 2018;93:874–880.

DeWolfe DJ. *Training Manual for Mental Health and Human Service Workers in Major Disasters*. DHHS Publication No. ADM 90–538;2000.

DuPont RL, Merlo LJ. Physician health programs: a model for treating substance use disorders. *Judges J*. 2018;57.

Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. *Med Educ*. 2016;50:132–149.

Dyrbye LN, West CP, Sinsky CA, Goeders LE, Satele DV, Shanafelt TD. Medical licensure questions and physician reluctance to seek care for mental health conditions. *Mayo Clin Proc*. 2017;92:1486–1493.

Garcia LC, Shanafelt TD, West CP, et al. Burnout, depression, career satisfaction, and work-life integration by physician race/ethnicity. *JAMA Netw Open*. 2020;3:e2012762.

Han S, Shanafelt TD, Sinsky CA, et al. Estimating the attributable cost of physician burnout in the United States. *Ann Intern Med*. 2019;170:784–790.

Harvey SB, Epstein RM, Glozier N, et al. Mental illness and suicide among physicians. *Lancet*. 2021;398:920–930.

Hopkins L, Morgan H, Buery-Joyner SD, et al; Undergraduate Medical Education Committee, Association of Professors of Gynecology and Obstetrics. To the point: a prescription for well-being in medical education. *Am J Obstet Gynecol*. 2019;221:542–548.

Jagsi R, Griffith KA, Jones R, Perumalswami CR, Ubel P, Stewart A. Sexual harassment and discrimination experiences of academic medical faculty. *JAMA*. 2016;315:2120–2121.

LoboPrabhu S, Summers RF, Moffic HS. *Combating Physician Burnout: A Guide for Psychiatrists*. American Psychiatric Association Publishing; 2020.

Menon NK, Shanafelt TD, Sinsky CA, et al. Association of physician burnout with suicidal ideation and medical errors. *JAMA Netw Open*. 2020;3:e2028780.

Myers MF, Fine C. *Touched by Suicide: Hope and Healing After Loss*. Gotham/Peguin Books; 2006.

Myers MF, Gabbard GO. *The Physician as Patient: A Clinical Handbook for Mental Health Professionals*. American Psychiatric Publishing; 2008.

Oreskovich MR, Shanafelt T, Dyrbye LN, et al. The prevalence of substance use disorders in American physicians. *Am J Addict*. 2015;24:30–38.

Rotenstein LS, Ramos MA, Torre M, et al. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. *JAMA*. 2016;316:2214–2236.

Shah K, Kamrai D, Mekala H, Mann B, Desai K, Patel RS. Focus on mental health during the coronavirus (COVID-19) pandemic: applying learnings from the past outbreaks. *Cureus*. 2020;12:e7405.

Shanafelt TD, Noseworthy JH. Executive leadership and physician well-being: nine organizational strategies to promote engagement and reduce burnout. *Mayo Clin Proc*. 2017;92:129–146.

Shanafelt TD, Boone S, Tan L, et al. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. *Arch Intern Med*. 2012;172:1377–1385.

Shanafelt TD. Physician well-being 2.0: where are we and where are we going? *Mayo Clin Proc*. 2021;96:2682–2693.

Slavin SJ. Medical student mental health: culture, environment, and the need for change. *JAMA*. 2016:316:2195–2196.

Stonnington CM, Files JA. *Burnout in Women Physicians: Prevention, Treatment, and Management*. Springer Nature Switzerland AG; 2020.

Tawfik DS, Scheid A, Profit J, et al. Evidence relating health care provider burnout and quality of care: a systematic review and meta-analysis. *Ann Intern Med*. 2019:171:555–567.

Trockel M, Sinsky C, West CP, et al. Self-valuation challenges in the culture and practice of medicine and physician well-being. *Mayo Clin Proc*. 2021;96:2123–2132.

Vayr F, Herin F, Jullian B, Soulat JM, Franchitto N. Barriers to seeking help for physicians with substance use disorder: a review. *Drug Alcohol Depend*. 2019;199:116–121.

30.9 Computers in Psychiatry

JOHN LUO, M.D.

Computers in medicine were primarily used in large hospital systems that could afford them in the late 1950s. At Massachusetts General Hospital, Dr. Octo Barnett and his team developed a language known as Massachusetts General Hospital Utility Multi-Programming System, known as MUMPs or M. This language supported the lab system at the time and was the basis of the U.S. Veterans Administration VistA (Veterans Health Administration System and Technology Architecture). Early use of computers in mental health included the Eliza program written in the 1960s at the Massachusetts Institute of Technology Artificial Intelligence Laboratory by Joseph Weizenbaum. This program implemented the Turing test, a hypothetical test to determine whether a computer can think like a human. Eliza emulated a Rogerian therapist, asking you how you felt and then to elaborate on those feelings. These early uses of computing in medicine, much less in psychiatry, kept computer use in medicine on a large system level.

In the 1970s, computing became available more individually. These computers were primarily used by hobbyists and did not reach mainstream use. The International Business Machine (IBM) Personal Computer with MS-DOS (Microsoft Disk Operating System) earmarked the beginning use of computers in the workplace. Software packages handled more of the administrative aspects of health care, such as billing and accounts receivable. Data were saved on floppy disks, which evolved from the floppy 5.25-inch version with paper covers to the 3.5-inch version encased in hard plastic with a metal slider to reveal the media. Data also could be transmitted over dial-up modems, which sent basic billing files to clearing houses. Some providers began to use word processing documents to generate patient notes, but most practices were still on paper. Much of the slow adoption of computer use in medicine was related to the challenges in learning to use DOS commands.

The advent of the Microsoft Windows system in the mid-1980s began to change that paradigm. Initially, Windows was a graphical user interface built to work with DOS, making it easier for users to identify where their programs and files were stored and navigate the operating system. Windows software was much more intuitive to use, and personal computers began to proliferate in offices, along with the development of improved printing with laser printers and storage with hard drives with increasing storage capacity. Personal computers using the Apple Mac operating system, such as the Apple II+ and Macintosh, were mostly used at home for their graphics capabilities. There was a significant power disparity between personal computers used at home and the office compared to mainframe computers that large organizations such as hospitals used to manage data such as laboratory tests. Moore’s Law is an observation by Gordon Moore in 1965 that the number of integrated circuits on microchips would double every 2 years. Increasingly faster and more capable computers and software development meant that computer technology began to play an integral role in daily life.

Online communities began to develop in the early 1980s with bulletin board systems, where users visited and shared messages that stayed on the host site, such as America Online, which provided online access and a destination to visit. In a parallel development, Tim Berners-Lee developed the World-Wide-Web in 1989 while at CERN, the European Particle Physics Laboratory. He wrote the first web client and server and set the foundation for the specifications of HTML (hypertext markup language), HTTP (hypertext transfer protocol), and URLs (uniform resource locator). Information exchange using web browsers such as Mosaic, developed by the National Center for Supercomputing Applications, which enabled the display of graphic images, helped spread the popularity of websites. Early organizational sites for information on the web, such as Yahoo.com, developed by Jerry Yang and David Filo at Stanford University, helped users find information in a growing online sea of information. Sergey Brin and Larry Page, also at Stanford University, developed a more efficient and effective search engine, Google.com, which dramatically facilitated Internet use. Information, including medical information, was becoming more accessible online and in hand.

EARLY INTERNET AND COMPUTERS

The internet as we know it today is a far cry from the early days. The Internet Archive Wayback Machine is an archive of old websites that preserves the early days of the Internet as far as the 1990s. Visiting some early websites illustrates that many were just sources of information that required manual updating. For example, this author wrote the American Psychiatric Association/Center for Mental Health Services Minority Fellowship webpage in HTML 1.0 using just Notepad, a basic Microsoft text editor. Early website pages linked to other pages both on the same site and external, hence creating the “web” of information.

Over time, the computing power of server computers increased along with higher-speed connections, and the Internet network increased its speed. Dial-up modems with initial speeds of 1,200 baud (one bit of information per second) were slowly supplanted by ISDN (integrated services digital network), capable of 64 kbits/second or 64,000 baud. In essence, ISDN lines were digital versions of the analog dial-up modems that ran over a hundred times faster. DSL (Digital Subscriber Line) delivered even higher speed internet over existing copper phone lines using a different frequency. Faster network speeds meant the Internet could be used to exchange more information than just text and pictures, including video and audio.

Along with increasing computing power, computing became much more portable. Early portable desktop computers such as the Commodore SX-64 eventually gave way to laptop computing. While not light, early laptop computers were significantly more portable than their early portable computer predecessors. The next level of computing was palmtop computing, such as the Apple Newton MessagePad and the Palm Pilot. These devices brought computing power into the palm of your hand, and when coupled with cellular capabilities, the smartphone was born.

The Palm Pilot was largely successful as a portable device due to its ease of use and well-designed user interface. Navigation involved using the stylus to tap on icons and scroll up and down. Input on the device was using letters via “Grafitti,” a single-stroke shorthand handwriting recognition system. A keyboard accessory made typing much easier and faster than using Graffiti. An infrared port could exchange information such as an address or note between devices. Software written for the Palm Pilot created many opportunities for medical applications. Many health providers took courses at the American Psychiatric Association annual meeting to learn how to use this device to access medical information.

Many consider drug reference software as the “holy grail” or “killer application” of the Palm Pilot and other mobile devices in health care. Gone was the need to find a Physicians Desk Reference or a textbook on medications. It was extremely convenient to find available dosages, indications, common side effects, pregnancy and black box warnings, and other information. However, the major strength was the ability to select medications and check the drug–drug interaction between them. Likewise, reference texts such as the DSM-IV and Washington Manual of Medical Therapeutics could be purchased and installed for reading using specialized applications. At the time, Internet access on these devices was unavailable, and there were not many health care–related websites online, so all health care content had to be installed or entered.

Physicians found other ways to use this portable device beyond storing and retrieving information. Electronic sign-out via the Palm Pilot was one way that two providers shared management of an inpatient service in a county hospital. Notes and tasks for each patient kept the care plan consistent and ensured laboratory results were followed up. Using a handheld database system, the consult service at UC Davis medical center likewise kept the continuity of care between the weekday team and the weekend on-call providers by managing patients using this database. Residents and attendings each carried their own Palm Pilot with the patients they were following and beamed via infrared updates to the ‘consult Palm Pilot’ for use by residents on call. A portable infrared printer also printed consult notes for the paper chart for full portability since desktop computers were not readily available. While this process nowadays does not sound so remarkable, at the time, charting was still done on paper, and orders were handwritten and entered manually by a unit secretary.

Telemedicine was just beginning to take root with the use of dual ISDN lines and expensive camera systems. Hilty and colleagues wrote about patient satisfaction with telepsychiatry, even with the rather limited bandwidth. Concerns were abundant regarding the use of the technology, including worry about what to do in the event of a clinical emergency, technology failure, or the inability to properly assess the patient for movement disorders. However, patients were quite satisfied with telemedicine visits because, for many, it meant much easier access to care. For example, Lopez and colleagues described the use of telepsychiatry for deaf patients using American Sign Language to communicate over video. It was remarkably faster and more convenient for psychiatrists to see patients using telemedicine instead of driving from patient to patient for in-person visits. Patients in rural areas could see specialist providers unavailable in their local area. Primary care providers could also access psychiatric consultants to help manage their patients. There were many reasons why telemedicine did not take off at the time in the early 2000s, mostly related to the availability of technology such as expensive videoconference devices as well as access to the ISDN lines since dial-up Internet access from home and businesses provided speeds insufficient for good quality video referred to as the “last mile problem” in telecommunications.

IMPROVING TECHNOLOGIES

Over time, the speed and availability of the Internet improved. Dial-up modem access gave way to broadband access via DSL or cable modem with speeds measured in megabits per second (Mbps or 1,000,000 bits/second) compared to dial-up speeds of (kbps or 1,000 bits/second). Increased Internet access speeds helped develop better HTML standards to support additional website features, such as incorporating video, improved navigation elements, and database integrations. In turn, these improved features brought about website developments enabling the delivery of more than just health care information.

The relative privacy of searching for health information online made searching for health topics one of the more popular search terms. Online assessment tools for mental health began to proliferate on the Internet. For example, the PHQ-9 screening tool for depression is widely available on numerous websites such as Mental Health America. Online support groups for patients, such as DailyStrength.org, provided a community where patients with anxiety, depression, bipolar disorder, and schizophrenia could share their experiences. Patients felt less alone with their illness and received peer support. However, these forums were not moderated by health professionals, which meant that if someone posted thoughts of suicide, there may be delays in that person getting urgent assistance via a welfare check. The relative anonymity as well of an online profile also meant that sending a crisis evaluation team would be nearly impossible. TogetherAll.org has peer-to-peer support and registered mental health practitioners to monitor the safety and anonymity of members. These mental health practitioners can also provide online one-on-one sessions if needed. TogetherAll has a process of escalation if there are safety concerns, and their policy states that they will pass information to emergency responders. TogetherAll is freely available to all citizens in the United Kingdom, but in the United States, access is limited to universities and colleges that have arranged for services to augment student mental health services. PatientsLikeMe.com takes sharing to another level. On this website, patients share data such as symptoms and treatments used to manage those symptoms, including medications. Patients share their experiences with different medications, psychotherapy, lifestyle modifications, and treatment programs, as well as track their progress with those treatments.

More than a thousand health information and support websites are available to patients on the Internet, so it is unrealistic for any health care provider to have knowledge or access to all of them. When an online website influences their decision-making on whether to start a medication or undergo a procedure such as electroconvulsive therapy, it is helpful to ask to see from what site or article they are getting their information. Reviewing this information together will help them understand their concerns and provide direction as to what further education may be necessary. Larger health organizations often contract with patient education platforms such as X-Plain to integrate these educational materials into the electronic health record. Providers can add relevant information to the printed after-visit summary or send a link for viewing in the patient portal. Ultimately, it is a process of listening, validating their concerns, and using motivational interviewing skills to facilitate decision-making.

One of the earliest forms of electronic communication has been electronic mail or email. In many ways, it is easier to use than just voice mail, which often is difficult to transcribe and not always accessible outside of the office. The American Medical Informatics Association Internet working group developed guidelines in 1989 for email communications with patients. Their recommendations included informing patients that email is not constantly monitored and is not to be used for urgent communication. They recommended that patients sign an affidavit recognizing these appropriate parameters for communication. Many of their recommendations still hold today. Joel Yager, MD, used email as an adjunct for communication with his patients with eating disorders in New Mexico. Many of his patients lived in rural communities and traveled long distances to see him; therefore, email was a convenient way to follow up on their progress and provide documentation for their medical records.

The Health Insurance Portability and Accountability Act (HIPAA), enacted in 1996, marked a major milestone regarding health information privacy in the United States. This law provided provisions to improve the effectiveness and efficiency of the health care system with national standards for electronic health care transactions, code sets, unique health identifiers, and, most significantly, security. The privacy rule sets standards for protecting individuals’ medical records and other personally identifiable health information. It requires appropriate safeguards to protect the privacy of protected health information and limits appropriate disclosures of such information without the patient’s authorization. Other elements include rights to copy the records, direct a covered entity to transmit this information to a third party, and correct this information. The security rule requires physical, administrative, and technical safeguards to ensure the confidentiality of health information. The enforcement rule highlights civil monetary penalties for breach of confidentiality and provisions for compliance and investigations. Any adoption of health information technology needs to ensure that the guidelines laid out in HIPAA are followed.

Communication with patients using email in a HIPAA-compliant manner was not a simple matter in the early days. End-to-end encryption has been considered the gold standard in ensuring that communication between provider and patient remains secure. The use of encryption software such as Pretty Good Privacy, developed by Philip Zimmerman, was initially cumbersome and required several steps. The first was to exchange public “keys,” which are the encryption rubric, so only the key owner could decode the message. Then the message text had to be copied into a separate security software application for encryption using the key. The now encrypted text could be copied into the email for sending to the recipient. The recipient would then use their key and password to decrypt and read the message. Nowadays, such encryption is managed in a variety of ways. Enterprise-level communication via email has built-in encryption invoked by a security code word in the subject of the email, thereby reducing the administrative burden on the sender and recipient with no exchange of keys nor use of a separate application. Companies such as ProtonMail and Tutanota provide secure email for individual practices via their Internet or web-based platform.

In this period, electronic health records also began to proliferate and develop. These are software applications designed to capture health data in a structured manner for ease of retrieval and organization. Early psychiatry adopters of computer use began with word processing or database software to manage health information. The advantages of electronic health records over paper include legibility, access by many simultaneously, and ease of analysis for population health. The Health Information Technology for Economic and Clinical Health (HITECH) Act in the United States in 2009 provided incentives for using electronic health records in clinical practice, known as “meaningful use.” Hospitals and individual providers could earn extra compensation from Medicare and Medicaid by meeting objectives in the electronic health record incentive program. Stage 1 involved data capture and sharing. Stage 2 involved advanced clinical processes such as drug–drug interaction and alerts regarding best practices. Stage 3 required improved clinical outcomes such as decreased falls or adverse events such as inappropriate dosing. Nowadays, without meeting these requirements, there are penalties via a reduction in reimbursement from Medicare and Medicaid.

Determining which health record system to use is a complex decision, requiring consideration of many factors. Documentation tools such as macros add standard text to help increase the speed of documentation. Not all providers are fast typists, so using voice recognition software or integrating medical transcription services may be a deciding factor. Electronic prescribing nowadays is a must since the reconciliation of medications is part of the meaningful use criteria and saves time over phone calls and sending facsimiles. Integration with billing, scheduling, and practice management software are key features. Likewise, integration of lab orders and results decreases time spent tracking down facsimiles or logging in separately into a web-based portal. Connections to a regional health information exchange may appropriately and securely share medical information electronically to improve health care quality, speed, safety, and cost. The ability to create a continuity of care document for ease of information exchange with other health care providers is another important feature of electronic health records. Other significant details include how the data or system is backed up and whether there is secured access remotely or online. An oft-overlooked feature is how the health record system manages users in the same medical record and addresses the need for users to enter and edit information without data loss and appropriate version control. A patient portal facilitates secured communication to providers and staff and is a must-have to decrease dependency on voicemail, email, and fax. Management of user access is often overlooked as a key feature of electronic health record systems, and it is part of the HIPAA requirement to ensure security and monitor access logs. To decide which of the above features are the highest priorities is certainly a challenge, and most vendors should have a training environment to try out the system for a test drive.

Like deciding which car to drive, an electronic health record system must be fairly intuitive. The user interface may be the most important feature of all. Navigation should be easy with both mouse and keyboard, and graphic visuals help orient the user to what function is used. Some developers will welcome requests for new or custom features, which they may build for minimal cost if that improves their system. Training demands and the ability to customize features are significant factors in selecting an electronic health record. Many health record systems offer mobile applications for the tablet or smartphone in additional to traditional desktop access. Information should be accessible and actionable no matter the provider’s platform. The company’s longevity may be the ultimate deciding factor because updates and maintenance become onerous tasks when the company becomes defunct.

HEALTH 2.0

A significant shift on the Internet has been the concept of Web 2.0, where new applications and sites began to harness collective intelligence and opinion. Wikipedia, an online encyclopedia where any web user can add and edit entries, is a good example of this paradigm. Real Simple Syndication (RSS) is a technology that allows people to not only link to a page but to subscribe, such that they are notified when there are changes. YouTube is another example where the site is a platform for users to upload any video content of their creation. In Web 1.0, there are hyperlinks to different sites, but with Web 2.0, the added element is interaction, engagement, and shared experiences. Health 2.0 is merely the application of these newfound principles to websites and services linked to health care. Prime examples of Health 2.0 sites were discussed earlier with patient support sites such as DailyStrength and PatientsLikeMe. These features of user-generated content to be shared online via various tools led the way for social media to become one of the more predominant features of online use.

At first, the general public's adoption of social media use was slow. Facebook was initially only for college students before opening to anyone over 13. Although Facebook is now considered the largest social media site, it was not always the most popular. MySpace was the first general social media site in popularity, but with the rising popularity of games such as Farmville on Facebook, the ability to post classifieds to sell products, chat features, and the addition of businesses on Facebook, it began to quickly become the main social media destination. The health care industry as a whole was slow to adopt the use of social media for professional purposes. The American Association for Directors of Psychiatric Residency Training created vignettes for discussion and exploration of resident physician views regarding appropriate professional versus personal use of social media in its Virtual Training Office.

The general public and health care providers' attitudes regarding social media use began to change. The initial recommendation to have separate professional versus personal accounts began to fade as people began to accept that both personal and professional online information became aggregated. This blurring of professional and personal use is best exemplified by how LinkedIn, a professional social media site, began to incorporate more features similar to Facebook with news feeds, hobbies, shared videos, likes, and messaging, while Facebook began to offer news and a marketplace for businesses. It also heralded the beginnings of how the convenience of using one login, such as Facebook, Google, or AppleID, to log into multiple websites began to override privacy concerns regarding information shared across these platforms. Linked accounts indicated one's interests, preferences, and purchasing tendencies thereby creating a digital information profile that commercial companies use to market goods and services. The value of convenience developed alongside the burgeoning "gig economy" where freelancers provide food delivery or transportation services.

Hospital systems began to develop a presence on social media primarily for marketing, extending the brand beyond the expected website to provide information. YouTube, X, formerly known as Twitter, and Facebook are popular platforms to tout the services available and message the public regarding expertise and quality recognized by U.S. News and World Report, Leapfrog group, and many others. Community outreach with online seminars linked to national health campaigns such as mental health awareness month provided hospitals and large groups a presence online, just as valuable as television and radio marketing. Health fairs, traditionally held in person, are shifting toward video conferencing to engage an audience beyond the local arena. Social media presence for health care agencies, health systems, providers, and companies is no longer novel or innovative but a must to engage the public or target audience.

Individual providers initially used professional social media to market their practice as an extension of their practice website; however, many are now using social media platforms such as Facebook and Yelp as primary ways to hang their virtual shingle. Links on these sites to curriculum vitae, blogs, and other social media platforms may lead to scheduling an appointment via a telephone call, email, or self-scheduling online. Voicing opinions on new research via X or retweeting other posts creates a professional presence and identity that serves many purposes. The ever-present #hashtag, popularized by X and now used universally across many social media platforms, helps the audience find providers with expertise in their area of interest, whether for patient care, research, or networking for future collaborations. Doximity, a professional network for providers, offers the ability to connect for ease of referrals and useful tools such as web-based electronic faxing and secured dialing that allow providers to keep mobile phone numbers private. In addition, Doximity provides individual providers the ability to rate hospital systems and residency programs in conjunction with U.S. News and World Report, as well as continuing medical education opportunities with articles and recorded webinars. Social media marketing of medical practices follows the natural trend of consumers who eschew the traditional white pages and guides such as Zagat for search engines and socially driven review sites such as Yelp.

Patients have benefitted tremendously from social media as a medium to share stories of life experiences beyond vacation photos, birthday parties, or gatherings. It has become a multimedia interconnected platform of photos and videos with linked comments and emoji reactions, where people share everything, including political views, favorite restaurants, and anything that comes to mind. Social media has bridged the communication gap that email, phone calls, and text messages could not with its broader community reach and engagement. Many social media users have shared their feelings with family and friends, which decreases isolation and loneliness. Searching on Facebook for “depression” has Facebook curated resources on how to take care of your emotional health or find health for a friend. Articles on depression from various websites are just a click away, and Facebook pages about depression are a collection of posts with comments from people regarding their experiences. People have found social media sites helpful in extending and improving relationships with those they already know and making connections to new people and groups. It is this extended yet separate sense of community that facilitates people to make announcements on Facebook that they are coming out or having suicidal thoughts. Many patients have been brought to the emergency room because they have written concerning posts about suicide, which has led Facebook to use Artificial Intelligence (AI) to review phrases in posts and comments for worrisome statements such as “had anyone heard from them?” In some ways, social media has created an extended community that oversees the well-being of its users besides their direct friends and family.

Just like any device or tool, social media has its detractions. The anonymous nature of creating a profile without the need for validation of the owner has fostered an opportunity for cyberbullying as well as “trolling,” which is a quite different concept than using the word when fishing. Internet trolling is when someone comments or responds to a post, usually in a confrontational way to evoke a strong emotional reaction, often for amusement. Trolling can be emotionally harmful, especially for young developing egos. Internet pranks, such as the

#cutforbieber, can propagate through social media platforms, causing physical harm in addition to its emotional impact. Overuse of social media has led to problematic use that impacts psychosocial function, and well-being has led to social media addiction. A Healthy Minds Monthly poll by the American Psychiatric Association in January 2022 of 2,200 adults revealed that a third of Americans say social media does more harm than good for their mental health. It may be that asking about social media or video game use should be part of the standard regular new patient intake inquiry.

Social media and other online content also has another consequence. Together, they build an informational “picture” of a person online with the mosaic of social media posts, comments, likes, pictures, articles, videos, and anything else tagged about that person on the Internet. All this information creates a veritable Internet Rorschach test for others to decide if this person is a professional with whom they may visit in the future. Physician rating sites such as Healthgrades and Vitals summarize the prospective patient details such as medical school education, residency training, board certifications, office locations, and hospital privileges. They also rate the provider and office based on friendliness, wait times, cleanliness, and how the interaction went. More significantly, there is an opportunity for patients to post comments about the provider. Although a referral from a colleague or friend is still the most common way patients find their providers, the Internet and physician rating sites may soon become the norm. Issues arise when the reviews are not positive, as that stain on reputation online may deter future patients from considering that provider. Most sites provide little recourse for the provider regarding negative reviews as they say it threatens the integrity of their platform to be an open space. If the review uses profanity or appears to be bullying, sites will consider deleting negative comments. Some physician rating sites will offer a premium account, which offers the opportunity to have a higher profile spotlighted or marketed on the site, pushing down the negative reviews lower, akin to the last page of an Internet search. It may be tempting to have patients post positive reviews as a rebuttal; however, the ask itself is problematic as a boundary-crossing. There is a report of a New York City OB/GYN who sued a patient for false and defamatory reviews in 2018, which was successful in a court order to remove all online postings. This measure is quite costly regarding attorney fees, given that online physician rating sites are not yet the main source for patient referrals. Firms such as ReputationDefender offer to fix search results, remove personal data, and manage business reviews. These services also come at a significant cost that must be weighed to decide the service's value. Self-help tips include monitoring your name searches with alerts from Google, validating your profile on physician rating sites to fix errors, and creating your own website with links to desired sites to promote their search engine optimization.

In addition to the online reputation issue, social media and other online content also create a privacy risk to the provider. Search engine websites have indexed countless search terms, including people's names, so that they can save a search results list for future reference. People search engines such as TruePeopleSearch and Intellius take the Internet search further by linking other information such as court records, property records, and other publicly available databases. These search results identify significant details such as a home address, relatives, and age. While these details perhaps are not as alarming as social security numbers, credit cards, bank accounts, and other details for sale on the dark web, this readily available personal information poses a significant privacy risk. Scott McNealy, Sun Microsystems cofounder, is often quoted from a Fox Business interview as saying, "You have no privacy; get over it." Despite the truth of his statement, there are many ways to limit information exposed online. People Search Sites have opt-out options due to the privacy laws of states and the federal government. The Reputation Defender website also offers a self-help guide to removing personal information from the internet, such as providing the bare minimum of information in your profile and locking down privacy settings by limiting access to posts. Jumbo, an app available on both Apple and Android mobile operating systems, also ensures privacy by archiving old Facebook, Instagram, and Twitter posts; deleting your Google, Gmail, and YouTube search history; as well as notifying you if data such as credit cards, email passwords have been compromised. DeleteMe is another vendor that will remove private information from over 500 data brokers. Privacy is a relative commodity that requires active monitoring and a decision to actively prune the weeds of exposure or hire a gardener.

APPS AND ARTIFICIAL INTELLIGENCE

The most recent wave of technology use in mental health has been the explosion of “apps” or applications for the smartphone. Social media use encouraged smartphone adoption as many posts are created on the fly and in real-time using these devices. The almost ubiquitous cellular signal availability and open wi-fi networks have made interacting with the Internet and each other with these devices an every day and all-day phenomenon. Indeed, leaving home without a smartphone is worse than leaving the house without keys. Apps such as maps are used daily for travel by foot or car as well as email, notes, and text messages. The COVID-19 pandemic has accelerated the use of telemedicine and apps for mental health due to the heightened need for services and increased access to mental health services that they provide. With so many apps available, the question of which ones to recommend to patients becomes a daunting task.

Many types of mental health apps are available in the Apple App Store and Google Play. Some are conduits to mental health providers, providing secure video conferencing, secured texting, or asynchronous communication. Other apps provide direct services using cognitive behavioral therapy approaches or have rating scales to provide the psychiatrist with status reports in-between visits. Reading reviews on the App Store or Google Play from users is not optimal as the prior user criterion for assessment is quite variable. The American Psychiatric Association has created an “App Evaluation Model,” a helpful framework for assessing whether an app is worthy of recommendation. The model implements several steps: access & background, privacy & safety, clinical foundation, usability, and therapeutic goal. The detailed questions asked are appropriate but may easily overwhelm those who intend to recommend them. The Argosy Foundation has funded Mindapps.org, an online database of apps reviewed using the APA framework. This site facilitates selecting apps for consideration with various filters such as cost, supported conditions, privacy, and features and is updated with new reviews by trained app raters. Ultimately, your patient will vote with their fingers as many apps are downloaded, but few are used regularly.

Artificial Intelligence in apps like chatbots is the next wave of behavioral health computing. Woebot and Wysa are examples of chatbots in production, using CBT, DBT, meditation, and motivational interviewing to help users build mental resilience and feel better. While most of the interaction is with the AI coach, there are links to human therapists when needed. The chatbots are not designed for crisis intervention or severe and persistent mental illness but are best used by the patient focused on wellness. They ask questions in a non-threatening tone and may include probing questions about local news events such as the COVID-19 outbreak and its impact on the user. Keywords such as “suicide” or “homicide” trigger safety protocols, including a human intervention or a call to 911 services. Chatbots are not poised to resolve the problem of not enough therapists, but certainly are a tool that may provide a bridge until a therapist is available.

A related product are digital therapeutics. Many of these are applications available on the smartphone and online via a web browser that treat serious conditions such as attention deficit hyperactivity disorder, depression, insomnia, and substance use disorders (SUD). Some of these are prescription based and therefore have been evaluated by the FDA and are reimbursed as products via pharmacy and medical benefits. They increase patient access to effective therapies and often extend the clinician’s ability to care for patients. Payors are involved as these solutions are scalable and cost-effective. Digital therapeutics are still in their infancy, but with the growing demand for mental health services, they will certainly grow and play a role soon.

CONCLUSION

The use of technology in psychiatric practice has grown slowly at first, but now adoption is at a more rapid pace. Traditional barriers used to be computing power, Internet access, and availability. Over time, desktop computers have given way to laptop computers and now smartphones and tablets with the continual development of smaller and more powerful central processing units, flash memory, and faster wireless Internet every year. Cloud computing and software as a service have also created an infrastructure where computing runs on the device and the network. Improved user interfaces and better multimedia graphics, along with a plethora of digital content online—websites, games, social media, and the like, have created a Matrix-like world where everything and everyone is connected. At first, the digital immigrants, who did not grow up with technology, were slow to embrace the digital world that digital natives, who grew up with technology, find incredibly necessary to their daily lives. It remains to be seen how growing up with technology changes the developing brain and how the preponderance of digital communication via text, social media, and email changes the value of direct human interaction. Will future generations have greater difficulty with nonverbal cues, or will they be better since AI-based avatars will help them determine what next steps to take? More significantly, will health care providers need to retrain much as factory workers and taxi drivers have done because technology has changed the need for their services with robots and automation? Health care has traditionally lagged behind in technology implementation due to privacy concerns and hefty regulations, but in the new future, it may be a different picture.

FURTHER READINGS

American Psychiatric Association. One in Three Americans Worry About Social Media's Impact on Mental Health Nearly Half Say it has Hurt Society at Large. Accessed February 26, 2022. https://www.psychiatry.org/news-room/news-releases/one-in-three-americans-worry-about-social-medias

American Psychiatric Association App Evaluation Model. Accessed April 16, 2022. https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model

Cheng C, Lau YC, Chan L, Luk JW. Prevalence of social media addiction across 32 nations: Meta-analysis with subgroup analysis of classification schemes and cultural values. *Addictive Behaviors*. 2021;117:106845.

Commodore 64. Accessed August 7, 2021. https://history-computer.com/commodore-64-history-of-the-commodore-64-computer/

DailyStrength.org. Accessed November 13, 2021. https://www.dailystrength.org

Facebook Suicide Prevention. Accessed February 5, 2022. https://www.facebook.com/safety/wellbeing/suicideprevention

Fearnow B. NYC Gynecologist Files $1 Million Lawsuit Over Negative Yelp Reviews. Accessed February 26, 2022. https://www.newsweek.com/joon-song-bad-review-yelp-lawsuit-gynecologist-michelle-levine-defamation-new-950538

Hilty DM, Luo JS, Morache C, Marcelo DA, Nesbitt TS. Telepsychiatry: an overview for psychiatrists. *CNS Drugs*. 2002;16(8):527–548.

HIPAA for Professionals. Accessed January 8, 2022. https://www.hhs.gov/hipaa/for-professionals/index.html

HITECH Act. Accessed January 8, 2022. https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html

Hoffberger C. 4chan Sparks sick #cutforbieber Trend, Reaches New Low in Trolling. Accessed February 5, 2022. https://www.dailydot.com/unclick/4chan-cut-for-bieber-hashtag-trend-hoax/

IBM PC. Accessed August 7, 2021. https://www.ibm.com/ibm/history/exhibits/pc25/pc25_birth.html

Internet Archive. Accessed September 1, 2021. http://archive.org

Kane B, Sands DZ. Guidelines for the clinical use of electronic mail with patients. The AMIA internet working group, task force on guidelines for the use of clinic-patient electronic mail. *J Am Med Inform Assoc*. 1998;5(1):104–111.

Lopez AM, Cruz M, Lazarus S, et al. Use of American Sign Language in telepsychiatry consultation. *Telemed J E Health*. 2004;10(3):389–391.

Luo J. A guide for the 21st century psychiatrist to managing your online reputation, your privacy, and professional use of social media. *Psychiatr Clin North Am*. 2019;42(4):649–658.

Luo J, Hales R, Hilty D, Brennan C. Clinical computing: electronic sign-out using a personal digital assistant. *Psychiatr Serv*. 2001;52(2):173–174.

McNealy S. You have no privacy, get over it. Accessed March 12, 2022. https://video.foxbusiness.com/v/5797171997001#sp=show-clips

Meaningful Use. Accessed January 8, 2022. https://www.healthit.gov/faq/what-meaningful-use

Mental Health America Depression Test. Accessed November 13. Depression Test—Free mental health tests from Mental Health America (mhanational.org); 2021.

Moore's Law. Accessed September 1, 2021. https://www.intel.com/content/www/us/en/history/virtual-vault/articles/moores-law.html?wapkw=moore%27s%20law

Moreau E. Internet Trolling: How Do You Spot a Real Troll? Accessed February 5, 2022. https://www.lifewire.com/what-is-internet-trolling-3485891

Norma J. Steve Wozniak and Steve Jobs release the Apple I. Accessed August 7, 2021. https://www.historyofinformation.com/detail.php?id=3019

O'Kane K. Mumps Programming Language. Accessed August 6, 2021. https://www.cs.uni.edu/~okane/

Oppy G, Dowe D. The Turing Test. Accessed August 6, 2021. https://plato.stanford.edu/entries/turing-test/

PatientsLikeMe.com. Accessed November 13, 2021. https://www.patientslikeme.com

Philip Zimmermann. Accessed January 8, 2022. https://philzimmermann.com/EN/background/index.html

Reputation Defender. The Complete Guide TO Removing Personal Information from the Internet. Accessed March 12, 2022. https://www.reputationdefender.com/remove-personal-information

Ronkowitz K. Eliza: A Very Basic Rogerian Psychotherapist Chatbot. Accessed August 6, 2021. https://web.njit.edu/~ronkowit/eliza.html

Tim Berners-Lee. Accessed September 1, 2021. https://www.w3.org/People/Berners-Lee/

Timex Sinclair 1000 Personal Computer. Accessed August 7, 2021. https://americanhistory.si.edu/collections/search/object/nmah_334354

TogetherAll.org. Accessed November 13, 2021. https://togetherall.com/en-us/

What is Web 2.0. Accessed February 5, 2022. https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html

X-Plain. Accessed December 4, 2021. https://www.patient-education.com

Yager J. Clinical computing: monitoring patients with eating disorders by using email as an adjunct to clinical activities. *Psychiatr Serv*. 2003;54(12):1586–1588.

31

Psychiatric Emergencies

31.1 Suicide: Overview of Epidemiology, Risk and Protective Factors, Treatment, and Prevention

DANUTA WASSERMAN, M.D., PH.D., AND VLADIMIR CARLI, M.D., PH.D.

The problem of suicide cannot be limited to just one discipline, and the complexity of self-destruction transcends the psychiatric field. Suicide can be viewed from many perspectives like religious, philosophical, anthropologic, sociologic, ethical, psychological, and psychiatric/medical. A new body of literature stemming from research across numerous research disciplines informs a multidimensional understanding of suicide risk, which can be summed up by understanding suicide as a complex health outcome.

A stress-vulnerability model has been proposed to consider the complex interaction of several risks and protective factors in the development of the suicidal process. According to this model, a constitutional predisposition ("trait" or "diathesis") to suicidal behavior is developed through the influence of genetic factors and/or early prenatal, perinatal, and postnatal experiences, and determines the individual’s suicidal threshold. Internal and external stressors interact with a diathesis and determine suicidal risk. Throughout the course of life, the resilience to stress and the suicidal threshold may be enhanced or diminished by the influence and interaction of different risk and protective factors, such as psychiatric or somatic illnesses, alcohol or drug use, traumas, coping, and problem-solving strategies, and social support. Furthermore, acute stress factors, such as relapse or exacerbation of illnesses, separation or losses, bullying, and economic problems, may act as triggers for suicidal behavior.

The development of suicidal behavior is a dynamic process. It may evolve from death wishes to suicidal ideation and then into suicide attempts and suicide. However, most suicidal processes do not end in death by suicide since death wishes and suicidal ideation fade away thanks to others’ protective responses to suicidal communication, as well as treatment interventions. Nevertheless, if new life stressors occur, the suicidal process may be reactivated with fewer positive outcomes.

Given this complexity, suicide prevention requires a comprehensive and multidimensional approach targeting both public health and health care sectors. The health care strategy aims at identifying risk groups, improving diagnostics, treatment, and offering better rehabilitation for suicidal patients. Health care interventions can be selective and target subgroups displaying risk factors for suicide, or indicated, that is, targeting individuals with identified psychiatric disorders, who have severe suicidal ideation or have attempted suicide.

The public health approach utilizes universal and selective strategies that aim to increase awareness about the role of protective factors for mental health, improve support and attitudes toward individuals with mental illness or suicide risk, and promote measures to eliminate fears and misunderstandings that surround suicide. It also comprises strategies for promoting mental health, social welfare, and education policies. Both clinical treatment and community-based interventions have growing evidence for actions that can reduce suicide risk.

Protective factors for suicide include cognitive flexibility, active coping strategies in difficult life situations, healthy lifestyles, active social networks, confidence, and a sense of personal value, as well as the ability to seek advice from others and help from the health care system.

This chapter focuses on the broad spectrum of psychopathology in suicidality and on treatment and preventive interventions used in the health care and public health sectors.

Universal public health interventions like restricting lethal means to suicide and school-based interventions are also shortly described.

EPIDEMIOLOGY

Suicide

Suicide occurs in all regions of the world and throughout the entire life span. According to the World Health Organization (WHO) 2019 estimates, with more than 700,000 deaths every year, suicide is among the leading causes of death worldwide. Indeed, suicide accounts for approximately 1.5% of all deaths, more than those due to breast cancer, war, and homicide. Suicide rates are likely to be even higher due to under-reporting, as not all countries report suicide mortality to the WHO. Furthermore, only 60 out of 194 Member States have good quality mortality registration data, and even in countries with good mortality statistics, suicides can be misclassified as accidents, homicides, or unknown causes of death.

In 2019, the global age-standardized suicide rate was around 9 per 100,000 population, even if large variations between different countries and regions are reported. The Eastern Mediterranean region showed the lowest suicide rate (6.01 per 100,000), while the highest rates were found in European (12.08 per 100,000), South East Asian (11.19 per 100,000) and African (10.28 per 100,000) regions. Within the same region, suicide rates can vary from one district to another. Therefore, it is important for tailored suicide prevention to survey local data from both population and clinical settings, as risk groups and risk factors can vary considerably at the local level. Moreover, in the same year, the global suicide rate for males (13.25 per 100,000) was more than double that of females (5.74 per 100,000).

Globally, suicides account for 57% of all violent deaths in men and 71% of all violent deaths in women. In high-income countries, 85% of violent deaths in both males and females are caused by suicide.

Among young people 15 to 29 years of age, suicide is the fourth leading cause of death globally, following road injury, tuberculosis, and interpersonal violence, accounting for more than 150,000 deaths per year.

Table 31.1–1.
Age-Standardized Suicide Rate by Sex in WHO Regions[a]

Year 2019	Age-Standardized Suicide Rate, Both Sexes (per 100,000)		
	Both Sexes	**Males**	**Females**
Africa	10.28	16.57	4.66
Americas	8.51	13.75	3.65
Eastern Mediterranean	6.01	8.51	3.29
Europe	12.08	19.94	4.69
South-East Asia	11.19	13.45	9.01
Western Pacific	7.65	10.33	15.13
Global	9.39	13.25	5.74

[a] Data from World Health Organization. Global Health Observatory. 2019. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/age-standardized-suicide-rates-(per-100-000-population)?bookmarkId=5be1bd45-0345-4d38-b75f-3602374891ea.

Tables 31.1–1 and 31.1–2 show the number and suicide rates by sex and age group in all WHO regions.

Suicide rates for males are highest in the European Region, followed by the Regions of Africa and the Americas. Suicide rates for females are highest in Southeast Asia and Western Pacific Regions.

In Table 31.1–3, the absolute numbers, crude, and age-adjusted suicide rates per 100,000 persons in each state in the United States are shown. The age-adjusted suicide rate in the United States is 14.1 per 100,000 population. The highest age-adjusted suicide rates are in Wyoming (32.3), Montana (32.0), Alaska (30.8), New Mexico (25.0), and South Dakota (23.1), while the lowest are observed in the District of Columbia (6.2), New Jersey (7.1), New York (7.9), and Massachusetts (8.0).

Table 31.1–2.
Crude Suicide Rates by Age and Sex in WHO Regions[a]

	Crude Suicide Rates (per 100,000)					
Year 2019	**5–14**	**15–19**	**20–24**	**25–49**	**50–69**	**70+**
Africa	0.57	3.52	6.08	9.43	20.78	44.69
Males	0.85	5.21	10.03	15.67	32.31	72.21
Females	0.29	1.88	2.36	3.59	10.16	21.98
Americas	0.69	7.77	11.16	11.45	13.62	14.67
Males	0.79	10.89	17.95	18.28	21.98	28.10
Females	0.59	4.54	4.27	4.86	5.97	4.32
Eastern Mediterranean	0.50	5.93	9.63	8.77	7.47	8.92
Males	0.70	7.91	13.97	12.40	10.73	12.59
Females	0.29	3.81	5.04	4.77	3.88	17.47
Europe	0.57	7.11	12.60	18.44	18.62	22.33
Males	0.76	9.90	19.98	30.69	30.58	39.11
Females	0.37	4.16	4.78	6.08	7.90	11.24
South-East Asia	0.92	11.30	18.34	15.92	13.33	19.47
Males	0.85	8.35	18.02	20.21	17.42	25.65
Females	0.99	14.43	18.67	11.50	9.33	14.16
Western Pacific	0.38	3.65	6.93	8.03	12.94	32.82
Males	0.45	4.54	9.60	11.19	16.59	42.74
Females	0.30	2.67	4.05	4.75	9.33	24.87
Global	0.64	7.02	11.61	12.05	14.25	24.53
Males	0.75	7.45	14.75	17.45	20.40	36.23
Females	0.53	6.57	8.38	6.57	8.33	15.38

[a] Data from Institute for Health Metrics and Evaluation. Global Burden of Disease Study 2019 (GBD 2019). GBD Results Tool. http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/a7cb270a3838b12d7e28ca2e59fdd698.

Socio-Demographic Factors

Although 77% of suicides occur in low- and middle-income countries, the highest age-standardized suicide rate is found among high-income countries (10.9 per 100,000). In particular, suicide rates are highest among males in high-income (16.4 per 100,000) and low-income (15.2 per 100,000) countries. In contrast, the highest suicide rate among females is reported in lower-middle-income countries (7.11 per 100,000). Differences between genders in coping with stress and conflicts, patterns of alcohol consumption and help-seeking behaviors for mental disorders, as well as access to lethal methods of suicide, can explain regional and gender differences.

As shown in Table 31.1–2, suicide rates are highest both for males and females in persons aged 70 years and older in nearly all regions of the world. An exception is represented by South-East Asian females, where the suicide rates seem to peak in young adulthood.

Status of being unmarried, divorced, widowed, older adults, unemployed, immigrant, and having different sexual orientations are risk factors increasing the propensity for suicide. Moreover, individuals in the armed forces, veterans, prisoners, and indigenous people have an increased risk of suicide. Since Durkheim, religion, marriage, and parental responsibilities are known key factors that play a crucial role in controlling suicide impulses. Moreover, the impact of values, social-political trends, societal patterns of drug and alcohol use, and violence is also of great relevance.

Methods of Suicide

National-level data on suicide methods are limited, as many countries do not collect this information, especially in low-income areas of the world. In Europe, hanging (54.3% of male suicides and 35.6% of female suicides) is the most common suicide method. Among males, it is followed by the use of firearms (9.7%) and self-poisoning (8.6%), while, among females, other common suicide means are self-poisoning (24.7%) and jumping from high places (14.5%). Hanging and self-poisoning also appear to be the preferred suicide methods in Africa and Asia. Suicides by suffocation (59.2%), self-poisoning (17.7%), and firearms (13.8%) are predominant in Latin America and the Caribbean. In 2021, 55% of suicides in the US involved firearms. Suffocation (26%) and poisoning (12%) were other common means.

Table 31.1–3.
Number, Crude, and Age-Adjusted Suicide Rates per 100,000 of Population in Each State of the United States by Sex, 2021[a]

State	Number of Deaths			Crude Rate			Age-Adjusted Rate		
	Both Sexes	Males	Females	Both Sexes	Males	Females	Both Sexes	Males	Females
United States	*48,152*	*38,330*	*9,822*	*14.5*	*23.3*	*5.9*	*14.1*	*22.8*	*5.7*
Alabama	827	680	147	16.4	27.8	5.7	15.8	27.3	5.4
Alaska	220	170	50	30.0	44.2	14.3	30.8	45.0	15.0
Arizona	1,475	1,183	292	20.3	32.6	8.0	19.5	31.2	7.9
Arkansas	618	491	127	20.4	32.9	8.3	20.6	33.0	8.9
California	4,143	3,253	890	10.6	16.6	4.5	10.1	16.1	4.4
Colorado	1,382	1,077	305	23.8	36.6	10.6	22.8	35.4	10.2
Connecticut	401	287	114	11.1	16.2	6.2	10.0	14.9	5.4
Delaware	137	110	27	13.7	22.6	5.2	13.6	22.8	5.2
District of Columbia	46	39	Suppressed[b]	6.9	12.2	Suppressed[b]	6.2	11.4	Suppressed[b]
Florida	3,350	2,554	796	15.4	23.8	7.2	14.0	21.7	6.6
Georgia	1,676	1,330	346	15.5	25.2	6.3	15.3	25.2	6.1
Hawaii	202	144	58	14.0	19.9	8.1	13.7	19.6	7.7
Idaho	387	324	63	20.4	33.8	6.7	20.5	34.0	6.9
Illinois	1,451	1,130	321	11.5	18.0	5.0	11.1	17.6	4.9
Indiana	1,129	926	203	16.6	27.4	5.9	16.4	27.4	6.0
Iowa	549	454	95	17.2	28.4	6.0	17.5	28.7	6.3
Kansas	560	470	90	19.1	32.0	6.1	19.4	32.4	6.5
Kentucky	816	695	121	18.1	31.1	5.3	17.9	31.0	5.3
Louisiana	689	551	138	14.9	24.3	5.9	14.8	24.7	5.7
Maine	277	227	50	20.2	33.6	7.2	19.5	32.0	7.4
Maryland	620	505	115	10.1	16.8	3.6	9.7	16.5	3.5
Massachusetts	603	462	141	8.6	13.5	3.9	8.0	12.7	3.5
Michigan	1,484	1,198	286	14.8	24.1	5.6	14.2	23.3	5.5
Minnesota	807	648	159	14.1	22.7	5.6	13.9	22.2	5.7
Mississippi	479	389	90	16.2	27.1	5.9	16.1	27.7	5.5
Missouri	1,177	983	194	19.1	32.3	6.2	18.7	31.8	6.2
Montana	350	284	66	31.7	50.8	12.1	32.0	51.2	12.3
Nebraska	288	223	65	14.7	22.6	6.7	15.0	22.7	7.3
Nevada	691	525	166	22.0	33.1	10.6	21.5	32.9	10.3
New Hampshire	223	174	49	16.1	25.1	7.0	15.1	23.8	6.6
New Jersey	688	539	149	7.4	11.8	3.2	7.1	11.4	3.0
New Mexico	532	437	95	25.1	41.5	8.9	25.0	40.9	9.2
New York	1,658	1,288	370	8.4	13.3	3.6	7.9	12.7	3.5
North Carolina	1,446	1,169	277	13.7	22.7	5.1	13.2	22.2	4.9
North Dakota	156	125	31	20.1	31.4	8.2	20.8	31.6	9.2
Ohio	1,763	1,436	327	15.0	24.7	5.5	14.6	24.3	5.4
Oklahoma	877	702	175	22.0	35.4	8.7	22.1	35.9	8.9
Oregon	888	686	202	20.9	32.4	9.5	19.5	30.7	8.7
Pennsylvania	1,885	1,526	359	14.5	23.8	5.5	13.9	23.0	5.3
Rhode Island	117	86	31	10.7	16.0	5.5	10.3	15.7	5.2
South Carolina	802	645	157	15.5	25.6	5.9	15.2	25.2	5.9
South Dakota	202	153	49	22.6	33.6	11.1	23.1	33.8	12.1
Tennessee	1,221	983	238	17.5	28.7	6.7	17.0	28.1	6.6
Texas	4,191	3,368	823	14.2	22.9	5.6	14.2	23.2	5.5
Utah	643	509	134	19.3	30.1	8.1	20.1	31.8	8.4
Vermont	142	114	28	22.0	35.5	8.6	20.3	33.0	8.2
Virginia	1,186	950	236	13.7	22.2	5.4	13.2	21.6	5.2
Washington	1,228	962	266	15.9	24.7	6.9	15.3	24.0	6.7
West Virginia	375	307	68	21.0	34.5	7.6	20.6	33.7	7.9
Wisconsin	905	705	200	15.3	23.9	6.8	15.1	23.5	6.7
Wyoming	190	154	36	32.8	52.0	12.7	32.3	51.8	12.6

[a] CDC Wonder, Centers for Disease Control and Prevention. https://wonder.cdc.gov/

[b] Data suppressed due to confidentiality constraints (fewer than 10 persons per year).

It is estimated that self-poisoning with pesticides is responsible for approximately 20% of global suicides, and its prevalence is as high as 48.3% in low- and middle-income countries in the Western Pacific region. In some countries in Asia, jumping from high buildings is a common method of suicide. Since 1998, the use of barbeque charcoal, which produces highly toxic carbon monoxide gas, as a means for suicide has spread rapidly in many Asian countries.

Suicide Trends

Between 2000 and 2019, the global age-standardized suicide rate has fallen by 36%, with larger decreases in the European (47%) and Western Pacific (49%) regions. On the contrary, a 17% increase was reported for the Region of the Americas. According to CDC (Centers for Disease Control and Prevention), suicide rates in US increased by 37% between 2000–2018 and, after a 5% decrease between 2018–2020, they returned to their peak in 2021. The reasons for the observed decreasing suicide rates can be many. One likely explanation is the improvement of global health and the use of new treatment methods.

Attempted Suicide

There is a lack of systematic registration of attempted suicides across national levels. Data on suicide attempts typically derive from self-reported surveys in the general population or from hospital-based data of medically treated suicide attempts. Suicidal behavior tends to be rare during childhood and increases during adolescence and young adulthood. Both among adolescents and adults, suicide attempts are more frequent among females than among males. However, there is a wide variation in the attempt-to-suicide death ratio with respect to age, sex, region, and methods used.

Suicide attempts are known to be the single most important risk factor for completed suicide. In spite of this, there are very few countries that contain registries of medically treated suicide attempters.

In clinical work, the presumed continuum from suicidal ideation through suicide attempts to suicide completions can be a helpful model in understanding the suicidal process, risk detection, and treatment choices, even if this model cannot be applied to all suicides.

RISK AND PROTECTIVE FACTORS FOR SUICIDE

According to the stress-vulnerability model, genetics and neurobiologic factors play an important role in the diathesis of suicide. Family and genetic studies show that the risk factors for suicide are both individual and familial. Suicidal behaviors aggregate in families, and a family history of suicidal behaviors is an independent risk factor for suicide attempts and death by suicide. The hereditability of suicidal behavior results from the contribution of multiple genes and genetic variants with small effect size, and it is at least partially independent of the hereditability of psychiatric disorders. To date, a large part of research on the genetics and neurobiology of suicide is focused on the study of the serotoninergic system, the hypothalamic–pituitary–adrenal (HPA) axis, and genetic and neurobiologic correlates of neuroplasticity. Other emerging areas of research involve the GABAergic neurotransmitter systems and inflammation parameters.

In the context of suicide, there is a growing body of evidence showing that exposure to childhood adversities can affect genes involved in crucial neural processes in the regulation of behavior through methylation and histone modification, which are supposed to induce behavioral deviations during the early development, and possibly later in life. This mechanism is called epigenetics. Childhood abuse and other detrimental environmental factors seem to target the epigenetic regulation of genes involved in the synthesis of neurotrophic factors and neurotransmission. On the contrary, family and school connectedness and support may act as a protective factor against suicide. Indeed, social relationships (e.g., social isolation, loneliness, sense of belonging, and social support) may play a major risk or protective role in the suicidal process, especially among adolescents and older people.

Impulsive and aggressive personality traits are largely recognized as suicide risk factors. Introverted, negativistic, perfectionistic, avoidant, dependent, and neurotic personalities are also associated with increased suicide risk. Other psychological characteristics, such as extroversion, curiosity, resilience, and self-esteem, seem to negatively correlate with suicidal thoughts and behaviors.

Increased suicidality is found to be associated with a wide range of health-risk behaviors, such as smoking, alcohol use, and sedentary behavior. On the contrary, healthy lifestyles such as being physically active and having a healthy diet may play a protective role against suicidality.

Finally, socioeconomic inequalities, such as high rates of unemployment, indebtedness, and job insecurity, as well as racism and discrimination toward ethnic and sexual minorities and people with mental health problems, are considered risk-conferring factors at community and societal levels. Conversely, country levels of social acceptance of people with mental health problems and availability of mental health treatments and services showed to negatively correlate with suicide rates in some countries.

PSYCHIATRIC DISORDERS AND SUICIDE

The presence of current and lifetime psychiatric diagnoses represents an important risk factor for suicide, even if the extent and the causal nature of this relationship are debated. Indeed, even if the majority of psychiatric patients do not die by suicide, an eightfold increased risk for suicide in psychiatric patients has been estimated using record linkage studies combining data on mental health service use and suicide deaths. Patients with psychotic disorders (risk ratio [RR] = 13.2) showed the highest risk, followed by mood disorders (RR = 12.3) and personality disorders (RR = 8.1). An increased risk for suicide was also found among patients with substance use disorders (RR = 4.4) and anxiety disorders (RR = 4.1). Considering these data, up to 21% of suicides may be prevented through the prevention and treatment of psychiatric disorders. Comorbidity between different psychiatric disorders, especially alcohol and substance use, largely increases the suicide risk. A suicide rate of 147 per 100,000 psychiatric inpatient years has been calculated, with a history of suicide attempts, depressive symptoms, suicidal ideation, and a family history of suicide being among the factors increasing the risk of inpatient suicide.

Depressive Disorders

Major depressive disorder (MDD) is strongly linked with suicide, especially when recurrent and long-term conditions are triggered by acute negative life events.

The clinical picture of MDD is characterized by symptoms like weight loss or gain, sleep disturbances, fatigue, concentration difficulties, changes in psychomotor capabilities, feelings of worthlessness, guilt, and recurrent thoughts of severe suicidal ideation with suicidal plans.

Patients diagnosed with persistent depressive disorder (dysthymia) complain of irrational patterns of negative thinking and chronic dysphoria.

The risk of suicide varies between the different subtypes of depression. Depressed suicidal patients have similar risk factors for suicide as patients with other psychiatric diagnoses, namely, living alone, being unmarried, unemployed, and especially having a history of previous suicide attempts. Furthermore, an increased risk of suicide was found among depressed patients with alcohol and/or substance use, anxiety symptoms, and personality disorders. Physical illnesses also increase the risk of suicide and attempted suicide in depressed patients, especially in the older adult population.

Affective temperaments like cyclothymic disorder and anxiousness are associated with both suicide attempts and suicide. Such temperaments, along with irritability and rapid mood switches, are important contributors to triggering suicidal acts. To prevent suicidality, treatment of depression with a combination of pharmacologic and psychological methods is recommended.

Bipolar Disorders

Bipolar disorders are separated from depressive disorders in the DSM-5 classification and are placed between the diagnostic class of depressive disorders and schizophrenia/other psychotic disorders in order to denote a bridge between those two diagnostic classes with respect to symptoms, family history of psychiatric disorders, and molecular genetic findings. Bipolar disorders constitute a high-risk group for suicide and attempted suicide. It is estimated that suicide accounts for up to 15% to 20% of deaths in people with bipolar disorders and that one-third to one-half of bipolar patients attempt suicide. The majority of deaths by suicide in this population occur when they are in a major depressive episode or in episodes with mixed features. Suicide during the manic phase is rare. The prevalence of suicide attempts is similar in both type I and type II bipolar disorders. However, younger age at illness onset and depressive polarity of both first illness episode and current or most recent episode significantly increase the suicidal risk. Other common risk factors are a family history of suicide and comorbidity with borderline personality disorder (BPD), anxiety disorder, and substance use disorders, especially alcohol use.

Effective pharmacologic treatment supported by psychological techniques is the primary strategy to prevent completed suicide.

Alcohol and Substance Use Disorders

Excessive alcohol and substance use, leading to significant impairment and distress, are well-recognized conditions for an increased risk of suicidal behaviors. Comorbidity with personality disorders magnifies the risk of suicide. Suicide mortality is highest among drug users, especially when opioid substances are involved, and lower, but still high, among persons with alcohol use disorders, as well as more prevalent among males compared to females.

Excessive alcohol and substance use increase sadness and feelings of despair, as well as aggression and impulsivity. Furthermore, it causes deterioration in cognitive capacity and flexibility to find constructive coping strategies. Finally, it might reduce the level of social integration. Suicidal alcohol users with a fairly good psychosocial coping abilities could make their suicide appear more astounding to their social networks. Good psychosocial functioning could potentially explain why they did not seek treatment or receive attention from significant others to motivate them for treatment.

Triggers for suicidal behaviors in persons with alcohol and substance use disorders are losses of important relationships, work, economic security, and self-esteem. The more dependence the substance users experience in their relationships, the greater the risk that separation may push them into self-destructive acts.

Guidelines for the treatment of alcohol and substance use are published by the American Psychiatric Association (APA) and by the National Institute for Health and Care Excellence (NICE). The duration of treatment may vary, but it is important to use both pharmacologic and psychological treatments.

Decreasing the accessibility to alcohol both on a societal and individual level, as well as molding attitudes toward alcohol intake, show significant results in diminishing suicide, as described during the Perestroika period in the former USSR.

Anxiety Disorders

Anxiety as a risk factor for suicide has been neglected for a long time. There is a high comorbidity between mood and anxiety disorders, and severe anxiety and agitation strongly contribute to suicidal behaviors. There are many ways to assess the severity of anxiety, and it is important that suicidal psychiatric patients, especially those with mood disorders, substance use disorders, and psychotic states, are assessed for the presence of severe anxiety and treated for it.

Patients who suffer from severe anxiety may sometimes deny suicidal thoughts or suicidal intent and refuse hospitalization. It is recommended to document in the medical records the patient's refusal and level of anxiety and secure that outpatient treatment includes frequent follow-ups involving the family or social network whenever possible.

Schizophrenia and Other Psychotic Disorders

In the schizophrenic population, estimates show that approximately 5% to 10% die by suicide, and the risk is higher at illness onset. Already Bleuler drew clinicians' attention to the fact that the most serious of the schizophrenic symptoms is the suicidal drive.

Moderate-to-severe depression is one of the most frequent features of schizophrenic patients who die by suicide.

Schizophrenics who die by suicide usually have poor treatment compliance, not seldom due to the side effects of antipsychotic medication like akathisia. Suicides often occur after abrupt discontinuation of medication due to negative attitudes toward medication and treatment. Poor treatment compliance, social isolation, and increased expectation of good performance from others and from patients themselves are risk factors for suicide in schizophrenics.

For a long time, there has been controversy surrounding the question of whether suicide in schizophrenia occurs during intense and frightening psychotic activity or during periods of remission. Studies found that command hallucinations are rare among completed suicides.

On the contrary, a good premorbid functioning, a higher level of education, or a higher intelligence quotient may predispose to suicide in younger schizophrenic patients, as they experience more disruption of performance and may have more difficulty accepting chronic illness and prospects of mental deterioration than older schizophrenic patients. Therefore, when giving information about the diagnosis, course of illness, and treatment, one should be aware that it is a risk situation for suicide.

Eating Disorders

Patients with eating disorders have an increased suicide risk. Thus suicide risk assessment should be included in a comprehensive clinical evaluation. High comorbidity of mental illnesses like MDD, bipolar disorder, anxiety disorders, and BPD among persons who are underweight as well as overweight is to a great extent responsible for the high suicide risk in all eating disorders. Treatment is often complicated and requires behavioral therapy combined with medication and supportive measures.

Low body mass index (BMI) and low serum cholesterol have been shown to be associated with a higher risk of attempted and completed suicide.

Personality Disorders and Suicide

Personality disorders have been previously understudied in clinical and general population studies on suicidal behaviors. Nevertheless, it is estimated that more than half of individuals who died by suicide had at least one personality disorder, and a similar prevalence of personality disorders is found among psychiatric patients who died by suicide.

Persons diagnosed with Cluster B comprising borderline, antisocial, histrionic, and narcissistic personality disorders are characterized by poor impulse control, aggressivity, unstable identity, poor self-image, affective instability with proneness to experience, and strongly react to real or imagined environmental stresses and losses. Difficulties with interpersonal relationships, problem-solving, and the inability to plan and think about the future in positive terms increase the propensity for triggering suicidal behaviors. The highest suicide rate is found among people with BPD (10%). Furthermore, a history of a previous suicide attempt is reported in 50% to 90% of these patients.

Cluster C is characterized by a high level of anxiety and includes avoidant, dependent, and obsessive-compulsive subtypes. Few studies have investigated the suicide risk in these disorders. However, they may increase the suicidal risk when depression or other psychiatric disorders are present.

The study of suicide risk among Cluster A personality disorders has received little attention, even if some evidence exists for increased suicide risk for individuals with schizotypal personality.

Personality disorders contribute even more to the risk of suicidality when comorbid with other psychiatric diagnoses and in the presence of negative life events, especially those involving interpersonal distress and loss. Concurrence of psychiatric disorders with personality disorders is now recognized as a major risk factor in suicide and attempted suicide.

Somatic Disorders and Suicide

Studies performed in the United States, Australia, and Europe show an association between suicidal behaviors and severe or chronic somatic disorders.

The comorbidity of somatic disorders with psychiatric disorders, especially with MDD and personality disorders, substantially increases the risk of suicide. However, the increased risk persists even when adjusting for the presence of psychiatric disorders. Indeed, somatic disorders involve physical, psychological, and social repercussions. They imply stress, pain, sometimes handicaps, limited social performance, a decrease in the capability to work, and the increased need for help from others. Age and separation from loved ones, loneliness, hopelessness, helplessness, and social isolation are parameters of importance for suicide risk.

The suicidal situation and the suicidal propensity vary during the course of the somatic disorder and depend on the treatment outcomes and pain control, as well as on psychosocial comfort.

Cancer, HIV infection and AIDS, stroke, diabetes mellitus, epilepsy, Parkinson disease, trauma with subsequent brain damage, spinal cord injury, multiple sclerosis, Huntington disease (HD), and amyotrophic lateral sclerosis are associated with an elevated risk of suicide.

Older adult patients with chronic or incurable diseases need to have adequate somatic and psychiatric treatment, as well as good psychosocial care in times of shrinking economical resources. Different opinions on the distribution of economic resources can sometimes be associated with advocacy concerning euthanasia and assisted suicide.

Organizing home visits, case management, and regular telephone contacts with somatically ill and vulnerable older adults are effective preventive methods, as it diminishes isolation and provides the opportunity to detect risk factors and risk situations for suicide early.

SPECIAL POPULATIONS

Children and Adolescents

Poor Mental Health. Mental ill-health is among the leading causes of disability in young persons aged 5 to 14 years, ranking in second place in Europe and the Americas. The prevalence of mental health disorders among children aged 3 to 17 years in the United States is estimated to be 7.1% for anxiety, 7.4% for behavioral/conduct problems, and 3.2% for depression.

The results of the Saving and Empowering Young Lives in Europe (SEYLE) study showed that in a usual school class of teenagers, approximately 10% of young people display high-risk behaviors like excessive alcohol and illegal drug use, heavy smoking, truancy, etc. Thirty percent display unhealthy lifestyles like poor sleep and diet, physical inactivity, and excessive Internet/media use. Both of those groups have high levels of depression, anxiety, and suicidal behaviors. Approximately 12.5% of adolescents required qualified psychiatric and psychological help. However, experiences from the SEYLE study in Europe and the Columbia University TeenScreen program in the United States show that there is a fear of stigma when screening for mental health problems in schools, and both parents and adolescents show little trust in the mental health care systems. Therefore, the process of destigmatization of screening for mental disorders and increasing help-seeking behaviors is an important public health issue. SEYLE results show that younger adolescents are more likely to adhere to rules, procedures, and recommendations with higher attendance rates in the health care system for girls and for pupils victimized by peers. Suicidal behaviors also appear to be an important predictor of help-seeking and a predictor of referral to mental health care.

Risk for Suicide. A meta-analysis calculated a pooled suicide rate of 6.04 per 100,000 in 15 to 19 years adolescents, which is far higher than the rate found in 10- to 14-year-olds (0.93 per 100,000). Suicidal thoughts and behaviors at this age are also common, with a lifetime prevalence of about 30% for suicidal ideation and 4% to 10% for suicide attempts. For most adults, it is difficult to acknowledge the child’s despair and suicidality. Caretakers, as well as parents, can deny serious childhood diagnoses due to frustrations and feelings of guilt when having difficulties handling a suicidal, depressed child or a suicidal child with serious conduct difficulties. Suicidal behavior in children and adolescents often occurs in the context of a poor family environment characterized by low perceived support, parent-child conflicts, and a lack of family cohesion. Parental suicide and psychopathology, particularly affective disorders and alcohol use, also increase the suicide risk in offspring. Suicidal children and adolescents have poor self-esteem and poor personal identity, are often truant, not seldom bullied, and have poor school grades, which leads to a sense of inadequacy. Impulsive-aggressive traits and impaired decision-making have also been described. Issues of gender identity are well recognized as risk factors for adolescent suicides. The presence of a mental health condition is associated with a 10 times higher risk of suicide death in young people, especially in cases of affective disorder, disruptive/conduct disorders, and substance use disorders. Attempted suicide is an important risk factor for future suicide.

Treatment. The treatment of suicidal children and adolescents should include the individual child and the entire family. Besides developing a safety plan, the aim is to improve coping styles and communication between the parents and the child. Psychosocial rehabilitation measures, including school consultation and academic remediation, are important.

The treatment of underlying psychiatric disorders requires psychotherapeutic and pharmacologic treatment when deemed necessary. Medications can play an important role in managing severe depression and reducing suicide risk when managed judiciously. When medication is used in the management of suicidal young persons, the family can get a false sense of security, as well as the fantasy that the problem has been solved, which may lower motivation to work through the problems in family and school situations.

Major advances have been made in effective treatment and care for youth at risk for suicide. Dialectical behavior therapy, cognitive-behavioral therapy (CBT), as well as a number of family-based therapies show evidence for reducing suicidal behavior in high-risk youth (e.g., attachment-based family therapy, family intervention for suicide prevention, Youth Nominated Support Team for Suicidal Adolescents). The common themes of these treatment approaches include the involvement of family to enhance positive communication and relationship dynamics, parent education and support, and treatment that goes beyond addressing the primary psychiatric condition to approaching suicide risk as to its own related but separate clinical target.

If the child or adolescent manifests psychosis, then hospitalization is required. Hospitalization may also be necessary in the case of extreme agitation, violence, or alcohol and drug use in the context of a disruptive home situation. If a young person perceives rejection in the family, a replacement home can be a solution during the period of treatment and rehabilitation.

Parent education using specialized resources—such as for parents of LGBTQ youth (Family Acceptance Program resources) or parents of bipolar or borderline personality disordered youth (Family Connections Program)—can be extremely helpful as well.

Older Adults

Risk Factors. In many regions of the world, older adults have the highest suicide rate compared to any other age group. The ratio of suicide attempts to suicide is very low among older adults, and therefore suicidal thoughts and plans in older persons should be taken seriously, given the increased fatality rate for attempts among older age populations. Decreased economical resources after retirement and placement into long-term care or residential care can increase the risk of suicide. Older adult suicides are characterized by high lethality methods, social isolation, and the presence of physical illnesses.

Diagnoses of MDD are the most common in older adult suicides, followed by a diagnosis of alcohol use and organic brain syndrome. Changes in sleep pattern, appetite changes, as well as somatic complaints or exacerbation of chronic concerns which stop an older adult from going out, can be a sign of depression. Symptoms of guilt, sadness, or anhedonia can be masked by somatic expressions. It is not seldom that cognitive functions are severely impaired, and depression can sometimes be misdiagnosed as dementia. Unconscious negative attitudes toward older adults may create a risk of not taking a suicidal risk in older persons seriously.

Evidence shows that dementia is not associated with an increased risk for suicide. However, the suicide risk can be higher in the period immediately after the diagnosis, especially in those of younger age, highly educated, and with preserved insight.

Treatment. The treatment of suicide risk in older adults benefits from focusing on aggressive identification and treatment of depression and enhancing social support. In cases of severe depression, when medication does not help, electroconvulsive therapy (ECT) can be recommended. In the treatment of older adults, drug-to-drug interactions are important to observe, as older adults often have multiple medications.

In talk therapy, it can be of value to focus on what the person has accomplished in life rather than talk about what may or may not lie ahead. Discussing suicidal ideation can be part of a process that puts one's life into perspective, and by acknowledging suicidal thoughts, the clinician can help the patient move away from the wish to die. Inclusion of family and other resources in the community should be encouraged in order to diminish loneliness and social isolation.

SUICIDE RISK ASSESSMENT

Even if suicide is regarded as a psychiatric domain, decisions regarding the level of suicide risk are made every day in all clinical specialties. Indeed, it is estimated that almost half of suicide victims had contact with primary care providers and one in five had contact with mental health services in the previous month. There are many obstacles when performing the suicide risk assessment due to the uniqueness of each individual possessing a unique set of risk factors. However, by engaging the patient in identifying drivers of suicide risk and reasons for living, it is possible to engage in a prevention-oriented suicide risk formulation. Foreseeable changes in the patient's life and environmental influences, which are likely to contribute to the outcome, are another key areas to engage the patient in safety planning for these foreseeable changes.

Clinical Interview

Suicide risk assessment is the most important instrument in clinical inquiry. This inquiry has an impact on the suicidal patients' communication, and to minimize the possibility of denial, it should be clear, empathetic, free from prior criticism, and focus on facts and patients' emotional and communication styles.

In a systematic clinical assessment of suicide risk, the presence of severe anxiety, MDD, use of alcohol or illicit substances, and previous suicide attempts should be covered. Personality type should be scrutinized. Intensive affective states characterized by severe anxiety, helplessness, hopelessness, rage, guilt, sleep problems, and impulsive behaviors are strong indicators of acute risk for suicide. Negative life events, such as bereavement, financial difficulties, and relationship problems, may act as precipitating factors. Indeed, disruption of social and occupational functioning is frequently reported among suicidal patients. Previous psychiatric disorders, suicide, or suicidal behaviors in the family should also be examined.

The recognition of direct and indirect suicidal communications is particularly important, and the clinician should explore if there is a suicide plan, its potential lethality, and the possible access to lethal means.

Repeated assessments are recommended to understand the mechanisms that generate the suicidal urge or impulse. Interview with the family or significant others is an important source of information. However, permission from the patient must be given, and it can sometimes take time to motivate the suicidal person to involve the family.

The evaluation of the social network and quality of family support should always be included when suicide risk assessment is done before temporary or final discharge from the hospital.

Psychometric Scales

A set of scales can be used to complement clinical inquiry. The Ask Suicide-Screening Questions (ASQ), Suicide Intent Scale (SIS), Scale for Suicidal Ideation (SSI-C), the Beck Hopelessness Scale, and the Columbia Suicide Severity Rating Scale (C-SSRS) are used for screening and assessing the risk of suicide.

Some biologic and genetic markers have been described, but none of them has practical application for clinical suicide risk assessment at the present moment.

The final decision in the suicide risk assessment is always based on an intuitive judgment, and therefore the clinician should be aware of feelings toward a suicidal patient, which can contribute to a denial of suicide risk, a rigid approach to the patient, or even negative emotional states like ambivalence or hostility. Those feelings and behaviors are called countertransference.

MANAGEMENT OF THE SUICIDAL PATIENT

Hospitalization

Following the assessment of suicide risk, the first important decision to be taken is the choice between admission of the patient to a psychiatric ward or treatment in outpatient care. The decision does not depend only on the status of the patient but also on the quality of the available family support, availability of other community resources such as mobile crisis teams and respite centers, and the accessibility to qualified outpatient treatment.

Hospitalization is reserved for concerns about imminent safety risks, and most regions have a mechanism for involuntary treatment since it was traditionally felt that, in severe cases, hospitalization provides security if safety measures at the psychiatric ward are followed. In the management of the suicidal patient in the hospital, precautions are important in restricting access to means of suicide in the ward and establishing different degrees of freedom while maintaining an empathetic encounter. However, current advocacy and research support the use of least restrictive measures to support individuals through suicidal crises, and more evidence supports brief interventions like Safety Planning and other treatments such as CBT, DBT, CAMS, and other forms of therapy which engage the patient to not only stabilize but engage more deeply with effective treatment and reasons for living.

Suicide in the Ward

When a suicide or suicide attempt takes place in the psychiatric ward, it is important to give support to all patients. A regular suicide risk assessment should be carried out for patients who are at risk. In addition to the usual treatment, intensive psychosocial support should be given to work through the emotions and the possible identifications with the patient who died by suicide among patients in the same or neighboring unit. Also, the personnel at the ward should be given psychological support.

Treatment of Acute Suicidal Crisis

In the treatment of an acute suicidal crisis, it is important to focus on reducing anxiety, insomnia, depression, and psychotic symptoms. Collaboration with the family and significant others must be secured. Safety measures should also be considered in the home environment, and family members should be alerted of the increased risk of suicide after discharge from the hospital. The plan for follow-up treatment and establishing collaboration with rehabilitation units and community services for chronically suicidal patients, especially those with the diagnosis of substance use, schizophrenia, or other psychotic disorders, is recommended.

Finally, it is important to have proper documentation about communication between different departments, families, and social services and have a clear follow-up plan.

The patient and the family should also be informed that they need to seek psychiatric emergency services as soon as a new suicidal crisis arises. Brief intervention and follow-up contact are considered successful strategies for reducing future suicidal behaviors. It is aimed at identifying warning signs of suicide (SOS), promoting coping skills, increasing family and social support, and ultimately, engaging the individual in mental health care through follow-up contacts (e.g., telephone calls, postcards, e-mails, and other digital means).

PHARMACOTHERAPY

Mood-Stabilizing Treatments

Lithium. Meta-analyses, reviews, and original studies show that long-term lithium treatment is associated with a substantial reduction in the risk of suicide and attempted suicide in patients with bipolar spectrum disorders. The analyses by the CDC show that mortality from overdoses of lithium is similar to the mortality from overdoses of new antidepressants and antipsychotic agents.

Ecologic studies also show that lower rates of suicide, as well as other violent behavior, are found in populations exposed to drinking water with a relatively high concentration of lithium salts. There is no plausible explanation for this finding, as the daily intake of lithium in drinking water is much lower than the dosages required for clinically effective treatment of mood disorders.

Other Mood Stabilizers. In bipolar depression, other mood stabilizers, such as anticonvulsants or second-generation antipsychotics, have beneficial effects on suicidal behavior. However, lithium is noted to be significantly superior in reducing suicidal behaviors and suicide death.

Antidepressants

Selective serotonin reuptake inhibitors (SSRIs) are nowadays widely used in the treatment of suicidal patients with MDD and related conditions. When choosing an antidepressant for a suicidal depressive patient, compounds with a sedative profile are preferable. Some experts have suggested that medications that increase drive may increase agitation, insomnia, and the risk of suicide. However, that risk is primarily theoretical.

Adverse outcomes in some patients during treatment with SSRI antidepressants like agitation, restlessness, irritability, dysphoria, anger, and insomnia can worsen the suicide risk in these patients. Due to those responses to serotonergic antidepressants, especially in some children and young adults below the age of 25 years, the Food and Drug Administration (FDA) and the European Medicine Agency (EMA) recommend continuous and competent monitoring of side effects. It should be noted that the data leading to these recommendations was on patients who reported suicidal ideation, not on actual deaths from suicide. Many experts, however, question the reality of this risk, particularly when compared to the very real risk of not treating a mood disorder.

Infusion of ketamine has shown promising results in reducing acute suicidal ideation. Nevertheless, further research is needed to assess its long-term safety and effectiveness in reducing suicide risk.

Antipsychotics

Suicidal symptoms in schizophrenic patients require, in addition to the standard treatment of schizophrenia with antipsychotics, an additional medication to control anxiety or agitation. Sedating antipsychotics are mostly used in these cases. When high dosages are required, attention must be paid to the risk of acute hypotension, particularly orthostatic. In some cases, to better control anxiety, benzodiazepines can be administered for a short period of time. Clozapine, although less used than many antipsychotics, has been shown to reduce suicide risk and is the only FDA-approved medication with that indication.

In the case of post-psychotic depression, additional treatment with antidepressants is recommended. If the suicidal symptoms are part of a depressive syndrome due to the side effect of neuroleptic treatment, the dosage should be reduced, if possible.

Anxiety Control

In controlling anxiety as a comorbid symptom of depression and in anxiety disorders, antidepressants are the preferred choice for acute and maintenance treatment. In cases where treatment response is not sufficient, adding short-term administration of benzodiazepines may be necessary.

Electroconvulsive Therapy (ECT)

Cases of depression with suicidality, which are difficult to treat by other means, can be treated by ECT, which has a rapid onset of action and relief of symptoms.

PSYCHOTHERAPY AND OTHER PSYCHOLOGICAL INTERVENTIONS

Treatment in the prevention of suicide requires a complex approach of psychosocial, psychotherapeutic, and psychopharmacologic interventions. Suicide cuts across the whole spectrum of psychiatric illnesses, and therefore therapists should be familiar with several therapeutic schools and strategies as they give valuable contributions to the treatment of suicidal persons. The choice of treatment depends on the condition of the patient. The combination of psychopharmacologic treatment with psychotherapy should always be taken into consideration as part of a complex treatment strategy.

Suicidal Statements

Suicidal communication, both verbal and nonverbal, must be taken seriously, independent of whether the clinician considers that the patient has a genuine wish to kill him/herself or if the purpose is to get attention. Suicidal communication always conveys an underlying sense of helplessness and hopelessness and a cry for help.

Risk of Suicide during Treatment

Treatment of suicidal patients is challenging, as the possibility of death by suicide is ever-present, independent of whether the treatment is pharmacologic or psychotherapeutic. Establishing an empathetic relationship with the patient when the threat that the patient might commit suicide is present can be difficult. This can be a discouraging condition to work under.

Treatment of a suicidal person requires self-involvement and confrontation with one’s own attitudes toward life and death.

In treating suicidal patients, one must allow the patient to feel suicidal and not minimize, reject, or deny the seriousness of their suicidal intent. Suicidal patients must be allowed to feel that their pain can be shared with another person. The challenge is to follow the patient and to understand, as well as to utilize psychotherapeutic methods without denying suicidality.

Transference and Countertransference

Suicidal patients have many different feelings toward their therapists and doctors, which in the psychotherapeutic literature are called transference. Therapists and doctors also have feelings toward their patients, which is called countertransference. Countertransference feelings toward suicidal persons can be empathetic but also ambivalent, aggressive, and even hostile. It implies an understanding of one’s own attitudes and values, as well as the origins of suicidal patients’ positive and negative feelings toward the clinician. A therapeutic alliance, which is important to establish, comprises elements of a real relationship between the therapist/doctor and the patient, as well as aspects derived from the patient’s experiences of earlier relationships, which are transferred to the present treatment situation.

Long-Term Psychological Treatments

Two long-term–based treatments, CBT and dialectical behavioral therapy (DBT), show the strongest evidence of preventing suicidal behaviors.

There are also a sufficient number of studies of good quality showing promising results in the treatment of suicidal patients with family psychotherapy, developmental group therapy, psychodynamic therapy, and in-home interpersonal psychotherapy (IPT). IPT and psychodynamic approaches like transference-focused psychotherapy, mentalization-based treatment, and sequential Brief–Adlerian psychodynamic psychotherapy are supported by randomized controlled trials in which the psychodynamic approach was found to be effective in reducing suicidal behavior among patients with BPD. Other therapies like Collaborative Assessment and Management of Suicidality (CAMS) use a combination of therapeutic techniques from multiple previously studied interventions.

Brief Psychological Treatments

Brief psychosocial suicide preventive interventions with some empirical support are CBT-based interventions, contact interventions, a combination of those two approaches, and manual assisted cognitive behavioral therapy (MACT).

Psychosocial support and involvement of family, friends, and significant others after obtaining consent from the suicidal person is also recommended.

Supportive Telephone Calls and Letters

Treatments consisting of supportive telephone calls, SMS, or written contacts after discharge of the suicidal patient from the emergency department or hospital showed preventive effects. The use of new digital technology is promising, but more research is needed to evaluate its effectiveness. Involving community resources with the goal that the social intervention should give support to the patient for combatting loneliness and finding social networks, and in this way, enhancing a sense of meaning in their lives is essential.

LOSS SURVIVORS

The term suicide survivors usually refer to those who have lost a significant other to suicide. Since stigma has been diminishing, people with a history of suicide attempts also identify with the term “suicide survivor.” Therefore, it has become necessary to use more specific terminology, that is, suicide loss survivors and suicide attempt survivors. When a suicide occurs, the family and the surrounding people experience severe trauma and a range of reactions from shock, sense of unreality, sadness and grief reactions, anger, and feelings of abandonment and rejection. Suicide loss survivors experience a devastating event and desperately seek an explanation of what has happened. Often they make contact with the family doctor or doctor who treated the patient and who also is a survivor of the patient's suicide. It can be wise to ask an experienced colleague to help with advice on how to manage this situation. The doctor needs not only to work with their grief but also to assess the potential suicide risk in loss survivors, who have a higher risk for suicide.

The wide range of interventions targeting those bereaved by or exposed to suicide is called postvention. Support groups, grief counseling, outreach, and internet-based interventions have shown some effectiveness in reducing depression, anxiety, and distress, as well as increasing help-seeking behaviors.

The development of support groups can be considered an important step forward in working through the aftermath of suicide in a compassionate and supportive way, educating about grief, demystifying suicide, as well as eliminating excessive guilt, shame, and blaming. The WHO issued guidelines on how to form support groups, and national organizations of suicide survivors also offer resources and programs for those who have experienced suicide loss.

PUBLIC HEALTH PERSPECTIVE AND PREVENTION

Controlling the Environment

The fatality of suicide methods largely varies from that reported for medication overdose (1.5% to 4%) up to 90% reported for firearms. An analysis of state gun regulations demonstrated that increased state gun restrictions and lower gun ownership rates were associated with fewer firearm related suicides in both adults and the youth. The rationale behind the WHO recommendations to diminish access to suicide methods is to curb a sudden and strong impulse to end one's life. Controlling the access to suicide means is an effective universal suicide prevention strategy as shown in several studies on the restriction of access to firearms, pesticides, and charcoal, as well as on the detoxification of domestic gas and vehicle exhaust, placing barriers and safety nets at jumping sites, reducing access to medications, and controlling access to railroads and subways. Furthermore, it should be incorporated not only at the public health level but also at the clinical level.

Alcohol policies, such as the restriction on alcohol availability, increased alcohol pricing, and more severe drunk driving laws are associated with reduced suicides, especially in males.

One of history's most effective suicide prevention programs was observed in the second half of the 1980s in the former USSR during ***Gorbachev's Perestroika*** (restructuring), characterized not only by openness and freedom but also by the strict limitations on access to alcohol by cuts in the production and sale, primarily with the aim to improve the health of the population. Suicide for men decreased by 40% from the years 1984 to 1986 in comparison with 3% in 22 European countries during the same period. This decline occurred in all 15 republics of the former USSR. The sharp decrease in suicide applied to women as well. The largest decreases were observed among men in the workforce aged 25 to 54 years. No corresponding decline for this age group was noted in any country during the 20th century.

The Role of the Media in Suicide Prevention

Sensational and irresponsible reporting by different types of media may precipitate or induce suicidal acts through imitation or identification mechanisms in suicidal persons and even trigger suicide clusters. The WHO issued guidelines on media coverage, describing how the press and broadcasting media should report on suicide in order to avoid copycat effects. It is essential to avoid the description of suicide as courageous or desirable.

Furthermore, even if the impact of fictional portrayals of suicide is still to be ascertained, resources for filmmakers, stage, or screen artists were also released by the WHO.

Conversely, media may play a protective role against suicidality, the so-called Papageno effect, when providing positive coping examples or when providing educational materials.

School-Based Suicide Prevention

Children and young people spend many hours each day at school, which is a good arena for comprehensive suicide preventive work involving a continuum of activities from mental health promotion, screening and prevention to intervention (treatment), and postvention when a suicide occurs.

Overall school-based suicide prevention interventions are promising, and results from randomized controlled trials of classroom-based intervention programs in the United States, like SOS and Good Behavior Game (GBG), along with the SEYLE study in Europe, show that suicidality can be prevented.

The Youth Aware of Mental Health (YAM) program used in SEYLE aims at educating youth on mental health promotion and suicide prevention and promoting coping and help-seeking strategies through interactive discussions and role-plays, as well as through educational materials. Compared to the control group, the YAM program has reduced the incidence of suicide attempts and severe suicidal ideation with suicide plans by 50%. This effect is higher than effects noted in other successful public health interventions like bullying and bullying victimization (7% to 23%) and school-based interventions addressing smoking cessation (14%). Besides being perceived as stimulating by young people and adult instructors, YAM may be particularly effective due to the promotion of adaptive coping strategies, increased socialization and help-seeking, and decreasing violence/aggressive behaviors.

To secure both preventive aspects and treatment aspects for young people in schools, stakeholders must continuously build models for collaboration across the education and health care systems interface.

FURTHER READINGS

American Psychiatric Association Steering Committee on Practice Guidelines. ***Treatment of Patients with Substance use Disorders***. 2nd ed. 2010.

Asarnow JR, Baraff LJ, Berk M, et al. An emergency department intervention for linking pediatric suicidal patients to follow-up mental health treatment. ***Psychiatr***

Serv. 2011;62(11):1303–1309.

Asarnow JR, Hughes JL, Babeva KN, Sugar CA. Cognitive-behavioral family treatment for suicide attempt prevention: a randomized controlled trial. ***J Am Acad Child Adolesc Psychiatry***. 2017;56(6):506–514.

Boritz TZ, Sheikhan NY, Hawke LD, McMain SF, Henderson J. Evaluating the effectiveness of the Family Connections program for caregivers of youth with mental health challenges, part I: a quantitative analysis. ***Health Expect***. 2021;24(2):578–588.

Brent DA. Antidepressants and suicidality. ***Psychiatr Clin North Am***. 2016;39(3):503–512.

Carli V, Hoven CW, Wasserman C, et al. A newly identified group of adolescents at “invisible” risk for psychopathology and suicidal behavior: findings from the SEYLE study. ***World Psychiatry***. 2014;13(1):78–86.

Diamond GS, Wintersteen MB, Brown GK, et al. Attachment-based family therapy for adolescents with suicidal ideation: a randomized controlled trial. ***J Amer Acad Child Adolesc Psychiatry***. 2010;49(2):122–131.

Doupnik SK, Rudd B, Schmutte T, et al. Association of suicide prevention interventions with subsequent suicide attempts, linkage to follow-up care, and depression symptoms for acute care settings: a systematic review and meta-analysis. ***JAMA Psychiatry***. 2020;77(10):1021–1030.

Ghandour RM, Sherman LJ, Vladutiu CJ, et al. Prevalence and treatment of depression, anxiety, and conduct problems in US children. ***J Pediatr***. 2019;206:256–267.

Glenn CR, Kleiman EM, Kellerman J, et al. Annual Research Review: a meta-analytic review of worldwide suicide rates in adolescents. ***J Child Psychol Psychiatry***. 2020;61(3):294–308.

King CA, Arango A, Kramer A, et al. Association of the youth-nominated support team intervention for suicidal adolescents with 11- to 14-year mortality outcomes: secondary analysis of a randomized clinical trial. ***JAMA Psychiatry***. 2019;76(5):492–498.

Lindow JC, Hughes JL, South C, et al. The youth aware of mental health intervention: impact on help seeking, mental health knowledge, and stigma in U.S. adolescents. ***J Adolesc Health***. 2020;67(1):101–107.

Mann JJ, Michel CA, Auerbach RP. Improving suicide prevention through evidence-based strategies: a systematic review. ***Am J Psychiatry***. 2021;178(7):611–624.

Niederkrotenthaler T, Voracek M, Herberth A, et al. Role of media reports in completed and prevented suicide: Werther v. Papageno effects. ***Br J Psychiatry***. 2010;197(3):234–243.

Paul ME, Coakley BA. State gun regulations and reduced gun ownership are associated with fewer firearm-related suicides among both juveniles and adults in the USA. ***J Pediatr Surg.*** 2023;58(9):1796–1802.

Preventing Suicide: A Resource for Pesticide Registrars and Regulators. World Health Organization and Food and Agriculture Organization of the United

Nations; 2019.
Ryan C, Russell ST, Huebner D, Diaz R, Sanchez J. Family acceptance in adolescence and the health of LGBT young adults. ***J Child Adolesc Psychiatr Nurs***. 2010;23(4):205–213.
Turecki G, Brent DA, Gunnell D, et al. Suicide and suicide risk. ***Nat Rev Dis Primers***. 2019;5(1):74.
Wasserman D. ***Oxford Textbook of Suicidology and Suicide Prevention***. Oxford University Press; 2021.
Wasserman D. ***Suicide: An Unnecessary Death.*** 2nd ed. Oxford University Press; 2016.
Wasserman D, Carli V, Iosue M, Javed A, Herrman H. Suicide prevention in childhood and adolescence: a narrative review of current knowledge on risk and protective factors and effectiveness of interventions. ***Asia-Pac Psychiatry***. 2021;13(3):e12452.
Wasserman D, Carli V, Iosue M, Javed A, Herrman H. Suicide prevention in psychiatric patients. ***Asia Pac Psychiatry***. 2021;13(3):e12450.
Wasserman D, Hoven CW, Wasserman C, et al. School-based suicide prevention programmes: the SEYLE cluster-randomised, controlled trial. ***Lancet***. 2015;385(9977):1536–1544.
Wasserman C, Postuvan V, Herta D, Iosue M, Värnik P, Carli V. Interactions between youth and mental health professionals: the Youth Aware of Mental health (YAM) program experience. ***PLoS One***. 2018;13(2):e0191843.
Wasserman D, Rihmer Z, Rujescu D, et al. The European Psychiatric Association (EPA) guidance on suicide treatment and prevention. ***Eur Psychiatry***. 2012;27(2):129–141.
Wasserman D, Sokolowski M. Stress-vulnerability model of suicidal behaviours. In: ***Suicide: An Unnecessary Death***. Oxford University Press; 2016:27–37.
Wasserman D, Värnik A. Suicide-preventive effects of perestroika in the former USSR: the role of alcohol restriction. ***Acta Psychiatr Scand Suppl***. 1998;394:1–4.
Wilcox HC, Kellam SG, Brown CH, et al. The impact of two universal randomized first- and second-grade classroom interventions on young adult suicide ideation and attempts. ***Drug Alcohol Depend***. 2008;95(Suppl 1):S60–S73.
Wilkinson ST, Ballard ED, Bloch MH, et al. The effect of a single dose of intravenous ketamine on suicidal ideation: a systematic review and individual participant data meta-analysis. ***Am J Psychiatry***. 2018;175(2):150–158.
World Health Organization. Preventing suicide: a resource series.
World Health Organization. ***Suicide Worldwide in 2019: Global Health Estimates***. World Health Organization; 2021.

31.2 Suicide Treatment

KATRINA A. RUFINO, PH.D.

This chapter focuses on ways to identify individuals and groups at high risk for suicide and how to prevent death. Although the aspirational goal as clinicians may be to prevent ***all*** suicide attempts and deaths, in reality, we are not there yet—however, reducing attempts and deaths is both possible and attainable. Prediction is not equivalent to prevention, and as such, prediction is not necessary for prevention to be feasible. Research shows that almost half of suicide decedents had a visit to their primary care doctor in the month before their death. As such, the National Action Alliance for Suicide Prevention created a recommended standard of care for treating suicide in an effort to standardize care. This standard of care includes identification and assessment, safety planning, means reduction, and caring contacts. Ideally, this effort will also increase the number of patients receiving follow-up treatment once their suicide risk has been identified. However, suicide treatment encompasses a broad spectrum of patients. Whom are we treating? The attempter, the individual considering it, the family of the attempter or decedent, or even the public when the victim is prominent? A discussion of working with each of these groups follows.

SUICIDE-RISK ASSESSMENT

It cannot be stressed enough that every psychiatric evaluation must include a documented suicide-risk assessment. If there is no indication of prior suicide ideation, an attempt, or depression, this may be brief. Be certain you are not inadvertently signaling that you're looking for a negative answer, as in "You're not suicidal, are you"? Better to start with something like, "With all you've been through lately, have you started to wonder if you could go on"? followed by "Are you having any thoughts of ending your life"? and at some point in the conversation asking, "Through all of this, what has you kept going"? The inquiry proceeds stepwise until the physician feels reassured that the patient is in no need of more restrictive treatment. If the patient is psychotic and/or depressed, one needs to expand the assessment. “How depressed are you”? or “Does it ever get so bad that you wished you were dead”? or “That you ever thought about suicide”? “Have you ever felt suicidal at another time in your life”? “Have you ever attempted suicide”?

Any complete assessment must include “Any family history of suicide attempts or deaths”? If a patient responds affirmatively to “Have you now or ever felt you wished you were dead”? it is often helpful to ask “Are you concerned that you might harm yourself”? If the patient says “no” you then need to wonder if they are only saying “no,” because they think that’s what you want to hear. Asking other informants, such as the patient’s spouse or other family members accompanying the patient to your office or calling the patient’s family should you be suspicious that the patient may not have been forthcoming with you, is also helpful (assuming that the patient agrees).

However, proper interviewing techniques can help increase comfort of the patient, as well as the information they are willing to disclose. According to Shawn Shea (2009 a,b), the real suicidal intent of any patient is the combination of their stated intent, or what they directly tell you, their reflected intent, or the amount of suicidal thinking, planning, and actions that may reflect the intensity of their actual intent, as well as their withheld intent, or the suicidal intent withheld from the clinician either intentionally or subconsciously. Shea’s Chronological Assessment of Suicide Events (CASE) approach to interviewing utilizes four techniques including the behavioral incident, gentle assumption, symptom amplification, and denial of the specific to elicit truthful and comprehensive information about suicide, a historically sensitive and stigmatized topic.

When making a suicide-risk assessment, one must weigh risk factors and warning signs as well as protective factors, after gathering all the pertinent information. Risk factors are generally considered more distal indicators of suicide, while warning signs are more proximal. It is important to remember that suicide risk is dynamic, and can rapidly shift. Thus, it is important to include intervention strategies in any thorough risk assessment.

Unfortunately, a recent meta-analysis of the last 50 years of research showed that no single risk factor predicted suicidal thoughts or behaviors much better than chance (Franklin et al., 2017), instead suicide risk is likely the confluence of many factors interacting simultaneously. However, there are several suicide measures available to help clinicians gather pertinent information regarding suicide history as well as scales that aid in risk determination. For example, the Self-Injurious Thoughts and Behaviors Interview (SITBI; Nock et al., 2007) is a comprehensive interview with strong psychometric properties that can be administered in both clinical and research settings. The Columbia Suicide Severity Rating Scale (C-SSRS, Posner et al., 2011), is also a clinician-administered interview with demonstrated psychometric properties. However, unlike the SITBI, the C-SSRS allows for the creation of an ideation intensity score and has triage and risk classifications. Additional suicide risk measures include the Beck Scale for Suicide Ideation (BSS; Beck & Steer, 1991), the Ask Suicide-Screening Questions (ASQ; Horowitz et al., 2008), and the Suicide Behavior Questionnaire—Revised (SBQ-R; Osman et al., 2001). The BSS has been widely used and has demonstrated predictive validity for suicide attempts and deaths by suicide (Brown, Jeglic, Henriques, & Beck, 2006), and the SBQ-R provides clinical cutoffs for use with both general and psychiatric inpatient populations (Osman et al., 2001). The ASQ was originally developed for use with pediatric populations (Horowitz et al., 2008) but has recently been expanded for use with adults (Horowitz et al., 2020).

Once you have completed a thorough history and evaluation, your assessment regarding the acuity of the situation determines what you do next. If you determine that suicide risk is greater than low, it is imperative a safety plan is completed (Stanley & Brown, 2012). A safety plan will identify warning signs and internal coping strategies for a patient, as well as various levels of external support including people and locations to provide distraction, people who will provide support, and professional support for times of crisis. For each person identified, telephone and contact information will be provided, so the patient has all the information in one place should a crisis arise. In addition, the final step of a safety plan will discuss means restriction, an integral step of any risk assessment. It is imperative the provider ask and document the patient's access to firearms, medications, and other lethal means, as well as discuss actions to mitigate danger when safety is a concern. Research shows that Safety Planning decreases suicidal behavior and increases treatment engagement for patients discharged from the ED (Stanley et al., 2018). For more information on Counseling on Access to Lethal Means (CALM), resources are available from the Suicide Prevention Resource Center at https://www.sprc.org/resources-programs/calm-counseling-access-lethal-means.

When determining a treatment plan for a patient, ideally the process should be collaborative, but it should always be patient-centered. More specifically, a treatment plan should be written based on the individual needs, resources, supports, and risk level of a patient. If a one-size fits all approach is used, patients might not have the ability to follow through with treatment, even with the best intentions, or worse, the plan sets the patient up for failure before they have even begun. In addition, a patient-centered treatment plan provides for the least restrictive environment possible for a patients' determined level of risk, such as an intensive outpatient program or partial hospitalization, if traditional outpatient would not be enough. If concerned about liability, the clinician can carefully document their rationale, as well as steps taken to mitigate potential risk (e.g., Safety Plan, means restriction, increased visits, communication with family, crisis hotlines, consultation with other clinicians, suicide-specific treatment, etc.). Clinicians can also increase communication with suicidal patients until the risk subsides. Research shows that “Caring Contact” or follow-up calls, emails, postcards, etc. reduced the rate of suicide behaviors after people were discharged from inpatient psychiatry or emergency department treatment (Luxton et al., 2013). Finally, inpatient hospitalization remains an option, but should be reserved for when there is a serious concern for safety, and your patient is at an imminent risk of harm.

THE TREATMENT OF SUICIDE

The Biologic Treatment of Suicide

There are several biologic treatments with research support for the reduction of suicide symptoms. For example, there is a large body of research support from meta-analyses and other studies showing that lithium reduces the risk of both suicide death and attempt in patients diagnosed with bipolar disorder. In addition, the antipsychotic drug clozapine has also been shown to have a protective effect on suicide death in patients diagnosed with schizophrenia. Other biologic treatments include electroconvulsive therapy (ECT) and transcranial magnetic stimulation (TMS). While ECT has not been studied as extensively as a treatment for suicide, research does support ECT as a treatment for depression. A meta-analysis of RCTs showed ECT as a valid treatment of both severe and treatment-resistant depression. Other neurostimulation techniques, including TMS are gaining popularity because they are less invasive, have fewer side effects, and do not require anesthesia, but there is less available research into their efficacy with suicidal patients.

Psychotherapy

Ellis and Goldston, guest editors for a special series, “Working with a Suicidal Client—Not Business as Usual,” noted that “until recently, suicide literature focused on risk factors and assessment, and treatment focused on the “underlying DSM-5 disease” in which suicide is seen merely as a symptom of another disease such as depression, psychosis, etc., and on prevention.” However, there is evidence of success for psychotherapies that specifically target suicide risk. An example of this is dialectical behavior therapy (DBT) for borderline personality disorder (BPD). Although DBT has come to be regarded as the gold standard for treating BPD, there are studies showing it may be equally valuable in reducing suicidal behaviors for patients with other psychiatric disorders as well. In addition to DBT, there are other therapeutic approaches that appear to be helpful for suicidal patients. One of them is mindfulness. As Luoma and Villatte (2012) point out, experiential avoidance, the tendency to avoid or escape from unwanted psychological experiences, may produce harmful consequences like suicidal behaviors. Mindfulness and psychological acceptance may serve as a means to target experiential avoidance, thereby decreasing the risk of suicide. This approach directs treatment of suicide not solely as a symptom of disease but as the primary target of therapeutic efforts. Similarly, mindfulness is a key feature that overlaps the four psychological processes in acceptance and commitment therapy (ACT).

CBT is another highly valued treatment for suicidal behaviors. A 2012 article in ***The Behavior Therapist,*** indicates that suicidal behavior should be viewed as the primary problem rather than merely viewing this behavior as a symptom of a psychiatric disorder. The authors target specific cognitions and behaviors which fuel suicidal behavior. They detail ***Cognitive Therapy for Suicidal Patients*** (CBT-SP; Wenzel, Brown, & Beck, 2009), an approximately 10-session program divided into early, intermediate, and late stages of treatment. Results from Brown et al.'s (2005) RCT showed that participants in the cognitive therapy group had a 50% reduction in reattempts when compared to participants receiving usual care.

Another suicide-specific treatment is David Jobes' Collaborative Assessment and Management of Suicidality (CAMS; Jobes, 2016). CAMS is a therapeutic framework for working with suicidal clients, regardless of the theoretical orientation of the therapist. It is a patient-centered approach guided by the principle that the patient is the expert in their own suicidal experience. Utilization of a Suicide Status Form (SSF) keeps suicide a focus of treatment. A distinctive aspect of CAMS is that the first session of treatment incorporates a thorough assessment of history and risk, which becomes part of the case formulation. Because of this patient-centered approach, a CAMS clinician is uniquely situated to be able to simultaneously build rapport with their client while also gathering pertinent risk assessment information. Several RCTs have shown strong research support for CAMS specifically for reducing suicidal ideation, but not necessarily attempts or death as of yet.

Role of the Hospital

When the patient reports intense ideation, the therapist is unsure of intent, or the patient reaches a severity that the therapist feels hospitalization is necessary to protect them, then hospitalization is required as a last resort. Although this sounds simple and straightforward, as discussed earlier, this is a complex situation, even when the patient and family agree. Since the therapist generally is not the decider of who gets admitted, the therapist will need to provide a convincing reason for “Why now in”? Assuming that hospital staff agree to keep the patient, they may or may not want your input in helping decide what level of care is required: closed or open unit, single or double room, frequency of observation, etc.?

Presumably, you (the therapist) and the hospitalist will work together on this. Before your patient can be safely discharged, however, virtually the same question must be asked by the attending, namely “Why now out”? What has changed to indicate that the patient is now safe? The mere fact that they have been in the hospital for a few days in itself is insufficient evidence. How has his improved outlook been documented? Just because they are taking medicine and attending groups is not sufficient evidence by itself. What do staff and family report? Such evidence must inform the discharge plan.

For the treatment of suicide while hospitalized, postadmission cognitive therapy (PACT), combined with enhanced treatment as usual has shown clinically significant reductions in depression, hopelessness, suicide ideation, and post-traumatic stress symptomology when compared to enhanced treatment as usual (Ghahramanlou-Holloway et al., 2020). PACT is six, 60 to 90 minutes sessions, adapted from Brown et al.’s Cognitive Therapy for the Prevention of Suicide Attempts (2005) for use in an inpatient setting. When longer-term inpatient hospitalization is an option, research shows that patients treated with an inpatient adaptation of CAMS showed greater improvement in measures of suicide ideation and suicide cognition (Ellis et al., 2015). In addition, patients treated with CAMS showed greater improvement on measures of suicide ideation, depression, well-being, and disability than patients receiving treatment as usual (TAU), even after controlling for severity of suicide ideation and previous attempts at admission. Furthermore, these gains were durable at 6 months post-treatment (Ellis, Rufino, & Allen, 2017).

The Patient Who Survives an Attempt or "Person with Lived Experience"

Suicide remains a leading cause of death, with psychiatric patients at an elevated risk. A patient who indicates suicidal ideation or who has just attempted suicide represents a psychiatric emergency; individuals who have attempted and those planning an attempt comprise the majority of the psychiatric emergencies managed by clinicians, aside from pharmacologic and general medical emergency situations. Traditionally, having serious thoughts of suicide was seen as a "symptom" of a disease, not a disease per se, and conventional wisdom has dictated that it is the disease that needs to be treated, not one of its symptoms. Over time, there has been a shift in the conceptualization of suicide from a model where it was considered a symptom that would decrease if the overarching disorder was treated, to the current conceptualization where it is the target of treatment.

DBT, a treatment specifically designed for patients with BPD also appears useful in some other impulsive psychiatric conditions such as nonsuicidal self-injurious behavior as well. Similarly, lithium, a treatment for stabilizing bipolar patients, appears to be useful in other conditions. Antipsychotics are commonly used for nonpsychotic agitation, and so forth. The most common condition associated with suicide is depression, and the psychotherapies and pharmacotherapies which address this are all part of the arsenal employed to decrease the risk of a suicide attempt or death. Unfortunately, "suicide attempter" has become almost a pejorative label, and such individuals too often are treated dismissively and peremptorily by emergency personnel, as though the attempt is merely a way to control or manipulate others. Certainly, most emergency rooms (ERs) are understaffed with stressed-out personnel who too often view attempters as unnecessary burdens which may stem from staff's own discomfort, anxiety, and lack of knowledge. However, a recent follow-up study revealed that from a total of 65,784 patients seen in ERs in Canada, during a median follow-up period of

5.3 years, 4,176 had died, including 976 (23.4%) by suicide (Finkelstein et al., 2015). This is hardly a low-risk subpopulation. In fact, immediately following discharge from the hospital is the highest risk period for death by suicide. A meta-analysis of 100 studies examining the postdischarge suicide rate from mental health facilities revealed a pooled rate of 484 per 100,000 person-years, significantly higher than the 2016 global suicide rate of 10.5 per 100,000. This rate increased dramatically when the samples were limited to three months from the time of discharge (1,132 per 100,000 person-years) or patients who were admitted to the hospital for suicidality (2,078 per 100,000 person-years) (Chung et al., 2017). However, research continues to show evidence of more effective treatments and clinical tools as discussed in this chapter.

Helping Significant Others When a Suicide Occurs

Foremost among “others” is the decedent’s family. Assuming that they desire help, they may be referred to a local “SOS” (Survivors of Suicide Loss) group near them. Both the American Foundation for Suicide Prevention (AFSP—headquartered in New York) and the American Association of Suicidology (AAS—headquartered in Washington, DC) can provide a geographic directory for such groups. SOS groups are ***not*** designed for those with lived experienced, since SOS groups exist to support individuals bereaved by suicide. Some SOS groups are reluctant to have family members attend meetings immediately after their loved one’s death, as it can be too overwhelming. Meetings are generally conducted by other survivors, rather than mental health professionals, on the principle that they “know” and empathize with what the families are suffering better than the professionals. This certainly does not mean that the treating professional’s role ceases following a patient’s death. Attending the funeral, meeting with the family to offer condolences, and discussing with them whatever is possible within the bounds of confidentiality is generally perceived by grieving families as helpful. Following a patient’s suicide, many mental health professionals are reluctant to have any contact with the decedent’s family—often because they feel so guilty or are so frightened of litigation that they wish to terminate all contact. However, such contact may actually diminish such an outcome and be comforting to the family.

The mental health provider is a survivor, as well. Approximately 20% of psychologists and 50% of psychiatrists report having lost a patient to suicide (Ruskin et al., 2004). Because having one's patient die by suicide is so traumatic for therapists, they often seek help for themselves. Institutional review boards which routinely conduct suicide reviews, postmortems, or "root-cause" reviews are generally designed with the institution's needs in mind, not necessarily the clinician's, so that therapy and support for the involved provider can be extremely helpful. Ellis, Schwartz, and Rufino (2018) discuss the importance of candid communication with supervisees and trainees when discussing loss or struggle. Specifically, having struggle normalized by respected mentors, and that seeking consultation or help does not equate to weakness or incompetence can be an invaluable lesson for colleagues and trainees alike. Additional resources for clinician survivors are also available from AAS.

"Significant others" also include classmates, coworkers, and friends of the victim, and mental health professionals may become involved in assisting and supporting them, as well. There are a number of postvention programs to assist schools in handling some of the emotional distress produced by a schoolmate's suicide. The AFSP has a postvention toolkit (2018) available to schools at no cost. These suggestions include the recommendation that schools should avoid permanent memorials to decedents, for instance, and other similar steps to avoid "contagion risk" resulting in additional suicides. Similarly, AFSP has toolkits customized for medical schools, residency programs, and veterinary medicine schools and clinics. In addition, there are published guidelines for the media to help them present suicide information in ways that do not engender "copycat" suicides.

Suicide and Malpractice. What standard of care applies to patient suicide? Conventional wisdom holds that "the psychiatrist isn't expected to predict all suicides," only to properly appraise the "risk" of it and to determine how "foreseeable" the risk is. Although correct in theory, it is more accurate to conclude, “it all depends on the judge and the jury.” Most psychiatrists will have at least one suicide in their careers. The great majority of these do not proceed to litigation. Psychiatry residents, unfortunately, not uncommonly have patients who die by suicide. Then, if litigation ensues, it is generally against the institution and the attending physician as well as the resident. The resident is often called as a witness but not held financially accountable. This is of small comfort to the resident because they have less capacity for perspective to fall back on and less confidence—hence, the critical need for support programs, lawsuit or not.

How does the prudent psychiatrist keep risk at a minimum? The medical record will generally be decisive in court proceedings; so it behooves clinicians to absolutely ensure that the record accurately reflects the evaluation, diagnosis, the suicide-risk assessment, treatment plan, course, results, etc. Any contact with the patients’ family to corroborate the diagnosis and explain the treatment should be documented. The inquiry about available means for suicide (e.g., guns in the home, potentially lethal medications, etc.) should be detailed. If there is reason to believe that the patient might be suicidal, were they asked about it at every visit and was this documented in the chart? Did the therapist ask about past history and family history of suicide? Did they ensure that the doses and types of medicine prescribed were reasonably safe, effective, etc.? Were the number of pills prescribed and their doses excessive or insufficient for the situation? A good relationship between the provider and the patient’s family may be the most critical factor in terms of avoiding lawsuits after a death.

Special Populations

Military. For both active duty and, to a lesser extent retired, military personnel have increased rates of attempts and completions. A recent study by Rudd et al. (2015), compared brief CBT in patients with previous attempts or ideation with intent to act. Half the participants received CBT while the other half received TAU. Results revealed the CBT subjects were 60% less likely to attempt than the subjects who received TAU. Suicide in military and veteran populations is researched extensively through the Department of Defense Military Suicide Research Consortium (MSRC), an effort led by Dr. Thomas Joiner at Florida State University and Dr. Peter Gutierrez at the Rocky Mountain Mental Illness Research, Education, and Clinical Center at the Denver VAMC. More information about their work can be found at msrc.fsu.edu.

Adolescents. Depression in adolescents is less likely to manifest the same symptoms as in adults. Verbal or even physical aggression is just as apt to be seen as sadness, tearfulness, etc. Giving prized possessions away or changes in behavior (becoming unruly, uncooperative, abruptly starting to use drugs or alcohol) may be all that the parents note as changed. Similarly, a drop in grades in a student who has always done well may be all that is observed by the parents. More “moodiness” or more sleep irregularity are common symptoms but not as common in adolescence as to be poor markers for depression. Suicide is the second leading cause of death in children and adolescents, so a thorough assessment of suicide risk should be completed when indicated, regardless of age. Similar to adults, this can include interviews and screening instruments such as the ASQ. If suicide is indicated, suicide-specific treatment should be implemented. CAMS-4Kids and CAMS-4Teens are age-appropriate adaptations of CAMS (see Jobes et al., 2019). This treatment utilizes an adapted SSF but also incorporates parent/caregiver involvement to optimize success.

The Older Adults. The geriatric population has more deaths by suicide than any other age group. In 2019, older adults comprised 16.5% of the population, but accounted for 19.3% of the suicide deaths (Drapeau & McIntosh, 2020). Risk factors prominent in the geriatric population include mood disorders, hopelessness, cognitive rigidity, physical illness and pain, isolation through loss of friends and family, particularly a spouse, work problems coping with retirement, and use of lethal means (Maris, 2019). Research on suicide prevention in the geriatric population has proposed four main drivers to decrease suicide in this demographic: (1) early detection, (2) general health promotion, (3) evidence-based mental healthcare, and (4) increased social connectedness (Conwell, 2014).

The Publicly Shamed. Shame, a risk factor for suicide, is further exacerbated when the shame is made public. Defrocked clergy, disbarred attorneys, delicensed physicians, and other prominent and respected individuals, who have become publicly shamed due to fair or unfair reasons (e.g., online shaming; criminal charges; financial "mischief," etc.) are anecdotally reported to have suicide rates higher than any other subgroup within the year following their exposure.

Victim-Precipitated Suicide. Also known as suicide-by-cop, victim-precipitated suicide is a method of killing oneself that engenders pain not only to the decedent's friends and family but also to an innocent person. The unfortunate police officer who is confronted by a stranger who appears to be homicidal and who shoots and kills that stranger thinking their own life or some third party is in danger also ends up becoming a victim. Jumping to one's death in front of a passing train or car leaves a similar painful legacy for innocent second parties.

ACKNOWLEDGMENT

The author wishes to express her appreciation to Drs. Tom Ellis and Christine Moutier for comments on an earlier version of this chapter.

Further Readings

Beck AT, Steer RA. ***Manual for the Beck Scale for Suicide Ideation***. Psychological Corporation; 1991.

Brown GK, Jeglic E, Henriques GR, Beck AT. Cognitive therapy, cognition, and suicidal behavior. In Ellis TE, ed. ***Cognition and Suicide: Theory, Research, and Therapy***. American Psychological Association; 2006: 53–74l.

Brown GK, Ten Have T, Henriques GR, Xie SX, Hollander JE, Beck AT. Cognitive therapy for the prevention of suicide attempts, a randomized controlled trial. ***JAMA.*** 2005;294(5):563–570.

Chung DT, Ryan CJ, Hadzi-Pavlovic D, Singh SP, Stanton C, Large MM. Suicide rates after discharge from psychiatric facilities: a systematic review and meta-analysis. ***JAMA Psychiatry.*** 2017;74(7):694–702.

Conwell Y. Suicide later in life: challenges and priorities for prevention. ***Am J Prev Med.*** 2014;47(3 Supple 2):S244–S250.

Ellis TE, Goldston DB. Working with suicidal clients: not business as usual. ***Cogn Behav Pract***. 2012;19(2):205–208.

Ellis TE, Rufino KA, Allen JG. A controlled comparison trial of the Collaborative Assessment and Management of Suicidality(CAMS) in an inpatient setting: outcomes at discharge and six-month follow-up. ***Psychiatr Res.*** 2017;249:252–260.

Ellis TE, Rufino KA, Allen JG, Fowler JC, Jobes DA. Impact of a Suicide-Specific Intervention within Inpatient Psychiatric Care: the Collaborative Assessment and Management of Suicidality (CAMS). ***Suicide Life Threat Behav***. 2015;45(5):556–566.

Ellis TE, Schwartz JAJ, Rufino KA. Negative reactions of therapists working with suicidal patients: a CBT/mindfulness perspective on “Countertransference.” ***Int J Cogn Ther.*** 2018;11:80–99.

Finkelstein Y, Macdonald EM, Hollands S, et al. Risk of suicide following deliberate self-poisoning. ***JAMA Psychiatry.*** 2015;72(6):570–575.

Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. ***Psychol Bull***. 2017;143(2):187–232.

Ghahramanlou-Holloway G, LaCroix JM, Perera KU, et al. Inpatient psychiatric care following a suicide-related hospitalization: a pilot trial of Post-Admission Cognitive Therapy in a military medical center. ***Gen Hosp Psychiatry***. 2020;63:46–53.

Hayes SC, Strosahl KD, Wilson KG. ***Acceptance and Commitment Therapy: The Process and Practice of Mindful Change.*** 2nd ed. The Guilford Press; 2012.

Horowitz LM, Bridge JA, Teach SJ, et al. Ask Suicide-Screening Questions (ASQ): a brief instrument for the pediatric emergency department. ***Arch Pediatr Adolesc Med.*** 2012;166(12):1170–1176.

Horowitz LM, Snyder DJ, Boudreaux ED, et al. Validation of the Ask Suicide-Screening Questions for adult medical inpatients: a brief tool for all ages. ***Psychosomatics***. 2020;61(6):713–722.

Jobes DA. ***Managing Suicidal Risk: A Collaborative Approach.*** 2nd ed. Guilford; 2016.

Jobes DA, Comtois KA, Gutierrez PM, et al. A randomized controlled trial of the Collaborative Assessment and Management of Suicidality versus enhanced care as usual with suicidal soldiers. ***Psychiatry***. 2017;80(4):339–356.

Jobes DA, Vergara GA, Lanzillo EC, Ridge-Anderson A. The potential use of CAMS for suicidal youth: building on epidemiology and clinical interventions. ***Children’s Healthcare***. 2019;48(4):444–468.

Linehan MM, Comtois KA, Murray AM, et al. Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. ***Arch Gen Psychiatry***. 2006;63(7):757–766.

Luoma JB, Villatte JL. Mindfulness in the treatment of suicidal individuals. ***Cogn Behav Pract.*** 2012;19(2):265–276.

Luxton DD, June JD, Comtois KA. Can postdischarge follow-up contacts prevent suicide and suicidal behavior? A review of the evidence. ***Crisis.*** 2013;34(1):32–41.

Maris RW. ***Suicidology: A Comprehensive Biopsychosocial Perspective***. Guilford; 2019.

Moutier CY. Innovative and timely approaches to suicide prevention in medical education. ***Acad Psychiatry.*** 2021;45(3):252–256.

Nock MK, Holmberg EB, Photos VI, Michel BD. Self-injurious thoughts and behaviors interview: development, reliability, and validity in an adolescent sample. ***Psychol Assess.*** 2007;19(3):309–317.

Osman A, Bagge CL, Guitierrez PM, Konick LC, Kopper BA, Barrios FX. The Suicidal Behaviors Questionnaire Revised (SBQ-R): validation with clinical and nonclinical samples. ***Assessment.*** 2001;8(4):443–454.

Pisani AR, Murrie DC, Silverman MM. Reformulating suicide risk formulation: from prediction to prevention. ***Acad Psychiatry.*** 2016;40(4):623–629.

Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. ***Am J Psychiatry.*** 2011;168(12):1266–1277.

Rudd MD, Bryan CJ, Wertenberger EG, et al. Brief cognitive-behavioral therapy effects on post-treatment suicide attempts in a military sample: results of a randomized clinical trial with 2-year follow-up. ***Am J Psychiatry.*** 2015;172(5):441–449.

Ruskin R, Sakinofsky I, Bagby RM, Dickens S, Sousa G. Impact of patient suicide on psychiatrists and psychiatric trainees. ***Acad Psychiatry.*** 2004;28(2):104–110.

Shea SC. Suicide Assessment: part 1: uncovering suicidal intent-a sophisticated art. ***Psychiatric Times***. 2009a;26(12):17–19.

Shea SC. Suicide assessment: part 2: uncovering suicidal intent using the chronological assessment of suicide events (CASE Approach). ***Psychiatric Times***. 2009b. Accessed May 12, 2021. https://www.psychiatrictimes.com/view/suicide-assessment-part-2-uncovering-suicidal-intent-using-case-approach

Stanley B, Brown GK. Safety planning intervention: a brief intervention to mitigate suicide risk. ***Cogn Behav Pract***. 2012;19(2):256–264.

Stanley B, Brown GK, Brenner LA, et al. Comparison of the safety planning intervention with follow-up vs usual care of suicidal patients treated in the emergency department. ***JAMA Psychiatry.*** 2018;75(9):894–900.

Swift JK, Trusty WT, Penix EA. The effectiveness of the Collaborative Assessment and Management of Suicidality (CAMS) compared to alternative treatment conditions: a meta-analysis. ***Suicide Life Threat Behav***. 2021;51(5):1–15.

Wenzel A, Brown GK, Beck AT. ***Cognitive Therapy for Suicidal Patients: Scientific and Clinical Applications***. APA Books; 2009.

31.3 Other Psychiatric Emergencies

ROBERT T. COBB, M.D., AND KATHERINE CAMFIELD, M.D., M.P.H.

The Community Mental Health Centers Act of 1963 established Emergency Psychiatry as one of the five essential services mandated for all federally funded community mental health centers nationwide. This legislation ultimately helped establish Emergency Psychiatry as a subspecialty through the increased recognition of its importance to the health care delivery system. Recently, legislative impacts have continued to expand psychiatric emergency services (PES) and led to a rise in advanced training programs. Along with providing further historical background of the field, this chapter will discuss frequently encountered PES diagnoses and how to prioritize care. Recent advances in management, both pharmacologic and nonpharmacologic strategies, also are highlighted. This edition of the chapter includes updated references and new subject material, in addition to the most relevant information from the previous version.

HISTORY

Emergency psychiatry originated in the Russo-Japanese War (1904 to 1905). Psychiatrists in the Imperial Russian Army championed treating psychiatric emergencies at the front lines. Healthy components of personality were strengthened, esprit de corps was encouraged, and soldiers returned to action as soon as they were stabilized and reconstituted. This countered the effects of stigmatization, secondary gain, and avoidable iatrogenic effects of longer-term military therapies that entail the removal of soldiers from the front lines. With this approach, the Italians found that 60% of what would be defined as post-traumatic stress disorder remitted. American military psychiatrists did not at first adopt this technique; only 18% of U.S. psychiatric casualties returned to action in 1943. A year later, when crisis-oriented therapy was introduced using abreaction, ventilation, and social support, as many as 85% returned to fighting. It is apparent that brief crisis-oriented approaches at times of emotional crisis can not only restore functioning but also reinforce and help develop healthy coping styles that protect against future stresses. Crisis intervention approaches have been incorporated into psychiatric practice and extended to the civilian populations over the years and now serve as an important arm of mental health care services available to patient populations.

Several major forces converged in the 1950s and 1960s that gave rise to the field of emergency psychiatry as it is known today. With the beginning of deinstitutionalization in the mid-1950s, and the emergence of emergency medicine as a medical specialty, hospital emergency departments began to see an increase in the number of chronically ill psychiatric patients who were being admitted for assessment and treatment. These patients were generally seen in the medical emergency department by a psychiatrist who acted as a consultant and with recommendations for either referral to an outpatient clinic or a psychiatric hospitalization. With the passage of the Community Mental Health Centers Act of 1963, which mandated 24-hour emergency services, the PES was created as a specialized program with its own staff and designated clinical area. The PES evaluated and treated patients suffering from acute psychopathology with the goal of providing treatment at the least restrictive level of care.

The American Association for Emergency Psychiatry was founded in the 1980s for psychiatrists with a special interest in emergency psychiatry. Other programs for providing emergency psychiatric care have since evolved, including crisis hospitalization programs, crisis residences, and mobile crisis teams.

The passage of the Emergency Medical Treatment and Labor Act (EMTALA) in 1986 made it illegal for hospitals to refuse emergency medical care to patients, which included psychiatric emergencies. Although this law was enacted to help prevent refusal of care for patients who are unable to pay and prevent “patient dumping,” it also developed a system where patients would remain in emergency departments for extended periods of time awaiting placement. Thus, it became imperative for psychiatrists to be present in the emergency department for both initial evaluations of psychiatric emergencies and re-evaluation assessments to decompress emergency department patient volumes.

The most recent major policy shift affecting PES came in 2010 with the passage of the Affordable Care Act (ACA). With millions of more Americans receiving health insurance and the expansion of Medicaid, the hope was that there would be decreased utilization of emergency department resources. The overall impact of the ACA on emergency psychiatric care is still unclear, but the need for these services definitely has not waned with the continued growth of PES as a specialty with advanced training programs.

A psychiatrist working in a PES must have an in-depth knowledge of psychiatry, substance-related illnesses, medicine, neurology, and the available community resources and advanced training programs aim toward the development of these specialty skills and knowledge.

In emergency psychiatric care, as with emergency medicine, prioritization of specific patient needs is essential to safe and effective health care delivery.

The priorities and goals of an emergency psychiatric assessment are as follows:

- To control aggressive behavior in order to protect other patients, the PES staff, and the patient from him- or herself.
- To rule out any etiology for a patient’s behavior that might be life-threatening or may increase medical morbidity.

- To facilitate a thorough clinical evaluation and an appropriate longitudinal plan.

First Priority

The first priority is to ensure that the PES environment is safe for clinical evaluation. Clinical personnel should be trained in the management of aggressive behavior and have a working understanding of their clinical responsibilities and institutional policies and procedures. Various formal training programs have been designed to ensure that providers are trained accordingly in management strategies. These programs serve to mitigate the risk of injury to patients and clinical personnel and are essential to providing a secure and safe environment for patient care.

Certain environmental variables should be modified to decrease the potential for violence. These include, but are not limited to: shortening waiting times, offering the patient drink or food, and decreasing environmental stimuli. Providing the patient with a comfortable chair in a quiet area or the opportunity to lie down is also effective in mitigating the risk of worsening agitation. Early identification of potentially dangerous pathology and early intervention can markedly decrease negative outcomes due to behavioral escalation, and a skilled practitioner should be able to intervene safely and appropriately.

It is important to understand that the physical structure of the PES is also crucial to facilitating a safe environment for patient care. Having shatterproof windows in the doors of interview rooms provides privacy but allows an interview to be observed. In addition, security cameras can allow for observation of clinical providers and ensure another layer of safety. Cameras can also provide the patient with a sense that their behavior is being monitored. Posting rules visible to patients within the PES can be useful in making it clear that violence will not be tolerated and has consequences is another useful strategy.

Safety and security can be further enhanced by:

1. weapons screening prior to entry of the secured area;
2. rooms in which the examiner cannot easily be trapped;

3. a choice of open or enclosed interviewing areas;
4. an alarm system to call for help, preferably a panic button in every room; and
5. adequate personnel to respond if help is needed, including trained security personnel.

If a person is brought in detained by the police with a history of a very recent assault, then it may be necessary to examine the patient in handcuffs or possibly in physical restraints, if the agitation or threats of violence persist. If the patient is acutely agitated and threatening, then it may be necessary to administer medication to help them regain control before beginning a complete interview. Interventions for calming the agitated and aggressive patient are discussed later in this chapter.

Second Priority

The second priority in managing emergent psychiatric pathology is to rule out a medical etiology for the patient's behavior. The incidence of patients presenting with psychiatric illness who have a medical etiology for their symptoms varies quite dramatically based on age and complexity of known medical history. Medical examinations of psychiatric patients in the emergency department are often limited in scope, and psychiatrists and emergency department physicians may have different viewpoints on the utility of laboratory screening. Therefore, psychiatrists must maintain a high index of suspicion for underlying medical problems and may need to initiate further studies than those that are recommended by emergency medicine guidelines.

Clinical history, signs, and symptoms that are suggestive of a medical etiology include:

- patients older than 40 years or younger than 12 years with no previous psychiatric history
- acute onset (hours to weeks)
- fluctuating course
- altered state of consciousness

- visual or olfactory hallucinations
- abnormal vital signs
- disorientation
- known medical illness or neurologic symptoms
- memory impairment
- complex medication regimen
- a history of alcohol or drug use
- atypical or unique psychiatric symptomatology

When medical conditions are suspected as the cause of the patient's symptoms, a physical examination and an appropriate medical workup should be completed, if possible. Initial diagnostic studies should include a complete blood count with differential, a complete metabolic panel (including electrolytes, blood urea nitrogen, creatinine, liver function studies, glucose), thyroid hormone level, an electrocardiogram, and urine drug screen, urinalysis, blood alcohol level, and neuroimaging when indicated. Medication blood levels should be obtained if applicable, for example, valproic acid, lithium, or carbamazepine. An ammonia level should be obtained for patients taking valproic acid. If the initial diagnostic studies are negative, but the clinician still suspects an underlying medical etiology, then additional tests can be completed once the patient is admitted to the hospital. Various medical illnesses that present with psychiatric signs and symptoms are listed in Tables 31.3–1 through 31.3–4.

Third Priority

Facilitating appropriate treatment and disposition through a comprehensive psychiatric evaluation is the third priority. The patient interview should always be regarded as a therapeutic opportunity. Patients in the PES are frequently in distress and should be given the opportunity to express their concerns and ventilate their affect. An open-ended question will usually elicit the primary reason for the visit and potentially aid in rapport development. Even if the staff has already obtained an initial history, it is still important to let the patient talk for a few minutes without interruption. Closed-ended questions should be reserved for the end of the interview or for those patients who are unable to adequately provide a meaningful and coherent history. Reassurance should be given frequently, and some clinicians may prefer to avoid long silences during the interview to ensure efficiency in the provision of care.

Table 31.3–1.
Differential Diagnosis of Anxiety

Medical
AIDS/HIV
Cardiac arrhythmia
Cerebral arteriosclerosis
Encephalitis
Epilepsy
Essential hypertension
Hyperthyroidism
Hyperventilation syndrome
Hypocalcemia
Hypoglycemia
Hypokalemia
Myocardial infarction
Internal hemorrhage
Mitral valve prolapse
Pheochromocytoma
Postconcussion syndrome
Pulmonary embolism
Subacute bacterial endocarditis
Temporal lobe disorder

Substance-Induced
Alcohol delirium and withdrawal
Cannabinoid intoxication
Caffeine intoxication
Stimulant Intoxication
Stimulant withdrawal disorders
Sedative, hypnotic, or anxiolytic withdrawal and withdrawal delirium

Psychiatric
Anxiety disorders
Bipolar disorders
Borderline personality disorder
Schizophrenia
Major depressive disorder

Generalized anxiety disorder
Panic disorder
Phobias
Post-traumatic stress disorder
Sexual disorders

The emergency evaluation varies in length and may, on occasion, exceed several hours. Ideally, the evaluation should include:

- identifying information: age, gender identity, preferred pronoun, ethnicity, marital status, and source of referral
- chief complaint
- history of present illness
- past psychiatric history
- substance use history
- medical history
- social history
- educational, occupational, and military history
- legal history
- family history
- review of systems, if indicated
- physical examination, if indicated (breast, genital, pelvic, or rectal examinations should not be done)
- mental status examination, including a cognitive screen

It is essential that suicidal and homicidal ideations are assessed thoroughly. A structured suicide and violence risk assessment should be performed and documented during the course of the examination. The most common liability in the PES is failure to hospitalize a patient who is at significant risk of violence to themself and/or others. If a decision is made to allow for discharge from the PES and not hospitalize, a thorough record of the clinical determination should be documented. This includes protective factors, risk factors, and the overall risk stratification of the patient for the potential for harm to self or others. While there is not a “gold standard” for documenting the clinical decision, it is important to have documentation reflect one’s decision should an unfortunate adverse event occur. Additionally, with a recent focus on decreasing the risk of suicide by the Joint Commission, the Behavioral Health Care and Human Services National Patient Safety Goals created minimum standards of care to help ensure patient safety due to suicide. As policies and guidelines are heterogeneous among institutions, it is important for clinicians to understand and practice within institutional policies.

Table 31.3–2.
Differential Diagnosis of Depression

Medical

AIDS/HIV
Antihypertensive medication use
Cerebral neoplasia
Cirrhosis of the liver
Dementia of the Alzheimer type
Hepatitis
Hypokalemia
Hyperthyroidism
Hypothyroidism
Infectious mononucleosis
Occult Malignancy
Pancreatic carcinoma
Postviral infection syndrome
Steroid psychosis
Vascular dementia

Substance-Induced

Stimulant withdrawal

Psychiatric

Adjustment disorder with depressed mood
Bipolar disorder
Borderline personality disorder
Brief psychotic disorder
Cyclothymic disorder
Persistent depressive disorder
Major depressive disorder
Schizophrenia
Schizoaffective disorder
Schizotypal personality disorder

Table 31.3–3.
Differential Diagnosis of Mania

Medical

AIDS/HIV
Antidepressant-induced mania
Amphetamine-induced mania
Bronchodilator-induced mania
Decongestant-induced mania
Delirium
Epilepsy
Hyperthyroidism
L-Dopa-induced mania
Postencephalitic syndrome
Steroid-induced mania

Substance-Induced

Alcohol intoxication
Amphetamine-induced mania
Cocaine-induced mania
Phencyclidine-induced mania
Synthetic cannabinoid-induced mania

Psychiatric

Atypical psychosis
Bipolar disorder
Schizoaffective disorder

Table 31.3–4.
The Differential of Thought Disorder

Medical
AIDS/HIV
Dementia due to Pick disease
Dementia of the Alzheimer type
Endocrinologic disease
Frontal lobe neoplasm
Migraine
Pernicious anemia
Steroid psychosis
Syphilis
Temporal lobe epilepsy
Vascular dementia

Substance-Induced
Alcoholic hallucinosis
Alcohol intoxication
Cannabinoid intoxication
Stimulant Intoxication

Psychiatric
Adjustment disorders
Atypical psychosis
Bipolar disorder
Brief reactive psychosis
Delusional disorder
Dissociative disorders
Major depression
Schizophrenia
Schizoaffective disorder
Schizophreniform disorder

Obtaining a comprehensive history in a PES can be challenging as patients who are psychotic or agitated may provide unreliable or inconsistent histories. This is especially true if the patient is mute, cognitively impaired, or agitated. In these situations, it becomes even more important to utilize collateral sources of information, including significant others, family, therapists, neighbors, police, or paramedics. Managed care companies may have significant clinical information and may be able to provide useful patient data and history. Previous hospital records, if available, should be reviewed, and important information should be summarized in the final assessment.

PRINCIPLES OF TREATMENT

Overview

Treatment in the PES usually focuses on behavioral and symptom management related to psychiatric pathology once other medical pathology has been ruled out and/or treated appropriately. Patients often have multiple confounding clinical problems. These include psychosis, depression, substance intoxication or withdrawal, extreme psychosocial stresses, and overlapping character pathology. Furthermore, psychosis and depression have an extensive differential (Tables 31.3–2 through 31.3–4). Given the complexity of patients presenting for care, making a definitive diagnosis may be difficult in the limited amount of time available to the PES clinician. Beginning diagnosis-specific treatment may be problematic unless continuity of care can be coordinated and assured with the next provider. The focus of the PES intervention should be acute stabilization and safety with subsequent longitudinal planning based on the outcomes of initial interventions.

The Initial Evaluation

Treatment should ideally begin at the time of the initial evaluation. The initial interview should serve to elicit important diagnostic information but is also an important therapeutic tool, if approached correctly. While the goal of the interview is to gather information about the present illness, past psychiatric and medical history, family history of mental illness, social history, and drug or alcohol use, it can serve as an important vehicle in establishing rapport and ensuring a patient is aware of their specific treatment plan.

Ideally, the interview should take place in a quiet, comfortable setting with both the patient and the clinician seated. However, often the intervention may take place in a less than ideal setting. In these instances, at minimum, the clinician should have an unimpeded exit in case of patient behavioral escalation. If the patient is seen after the medical evaluation, then it is important to remove unnecessary intravenous needles and ask the patient to sit up on the stretcher. This helps restore the patient's autonomy and improve the physician-patient alliance. The physician should always introduce themselves and address the patient by utilizing the patient's preferred pronouns formally to ensure a patient's sense of dignity.

The interview should begin with nonspecific, less intrusive questions. Only after the patient begins to show some comfort with the interviewer should the physician start questioning them about the specific details of the present illness. Questions should be open-ended. The physician should be flexible in their approach to the interview as each patient presents a unique set of circumstances, behavioral triggers, stressors, and life experiences that will influence the diagnostic evaluation.

When interviewing an acutely disturbed patient, it is frequently necessary to structure the interview by asking straightforward questions. If the patient begins to ramble, then the interviewer should indicate that they understand what the patient is saying while helping them to be redirected in providing important content. When patients have disordered perceptions, it is important not to attempt to correct the misperceptions but to clarify how the patient experiences them. The clinician should avoid the temptation to use logic to convince the patient that they are wrong. Such an approach makes the patient more defensive and can lead to escalation with the potential for dangerousness. The interview should be geared toward fact-gathering and establishing a baseline understanding of psychopathologic-related impairments.

Establishing emotional contact with the patient to help identify the patient's feelings will help build the treatment alliance. Even a psychotic patient's feelings are not unique, and the interviewer should be capable of identifying and empathizing with them. A patient's need to protect their sense of integrity and self-control may initially make them reluctant to engage in the interview. If patients are resistant to the idea of talking, then it is useful to elicit their feelings about the situation and try to understand the patient's discomfort with being in a PES and not knowing what may happen to them. Sometimes it may be better to drop a particular subject and come back to it later in the interview when the patient is more at ease. It is of great importance to understand that the patient is experiencing a significant loss of autonomy in requiring an assessment in the psychiatric emergency room setting. Individuals react differently to their circumstances depending on the level of impairment and previous life experiences. Impulse control is often compromised and is often a reflection of the normative human response to stress. The innate "fight or flight" response based on sympathetic nervous system drive can be elevated in PES patients, especially those who do not want assistance with their underlying psychopathology. The goal of the interview should be to develop rapport with the patient and aid in their understanding of the unique set of circumstances that have led to the clinical intervention. It is also important to understand that the clinician may be in a position of perceived power by the patient. As a result, the goal should be to have the patient understand that you are an advocate and there to help aid in the recovery process.

When psychotic patients do not respond to questioning, the interviewer should use whatever data are available to make a meaningful connection with the patient. This may include the patient's words, expressions, appearance, or behavior, as well as a physician's feelings. Comments should be as specific as possible in order to provide some structure to an interview that has little to no momentum —"I don't know what Jesus says to us" or "I see you are in a bathrobe. I gather you were brought to the hospital unexpectedly." By attempting to develop rapport in understanding the patient's experiences, a clinician can have a significant impact on developing a therapeutic alliance and improving patient outcomes.

Of all patients, the most difficult to evaluate is the hostile, paranoid patient. If a patient is frightening, then the physician should not ignore their own fear. Even the most impaired patients can quickly sense a physician’s discomfort and may become frightened, resulting in reactive behavior. This may, in turn, lead to a violent episode. At times, however, the physician may have to acknowledge their fears to the patient, “The way you are looking at me is scary like you are on the verge of striking out. I won’t be able to give you my full attention if I am afraid.” If the physician remains professionally confident and does not avert their eyes, then the patient is reassured. The patient’s anxiety can be further ameliorated if the interview is conducted with the door open, thus allowing for a sense of safety for the patient. If the patient engenders a level of anxiety in the clinician such that it interferes with the evaluation, then the clinician should request another clinician to interview the patient or interview the patient with another staff member present in the interview room. It is often difficult for clinicians to understand their limitations in engaging with patients, but ensuring your safety and the safety of others is paramount in the PES. Termination of a clinical assessment may be necessary and is not usually a function of poor training.

It is important to be extremely tactful with paranoid patients because they are vulnerable to being easily humiliated and made to feel guilty for their actions and thoughts. The clinician should never become involved in how the patient’s beliefs are distorted. If the physician tries to determine what is real and what is not, then they will be experienced as a prosecutor, and the patient may become more argumentative and defensive. By identifying a patient’s thought distortions in their delusional belief system, especially those that are paranoid in nature, a clinician may become incorporated into the delusional construct and disrupt rapport development.

Frequently, paranoid patients who are angry begin with a tirade of accusations about being mistreated. The physician may have to interrupt these tirades and ask the patient, “How can I help you?” in an attempt to develop a supportive therapeutic alliance. With the hostile, paranoid patient, the physician should always keep a professional distance by not becoming too friendly, avoiding jokes, and avoiding parental reassurance as patients can quickly escalate on the basis of external stimuli or a misperception of circumstances. Occasionally, highly paranoid patients can make the interviewer feel defensive and foolish by twisting the meaning of the interviewer's words, making it impossible to sustain any direction in the interview. Under these circumstances, it is advisable to alter the course of the interview and attempt to process with the patient their conduct. If meaningful contact cannot be made, then the interview should be terminated, reassuring the patient of concern for their safety and ensuring that they are aware of the clinician's role as an advocate and care provider.

When patients are unable or unwilling to provide clinical information, significant others should be contacted (e.g., spouse, relatives, immediate family, or previous treating physicians) to obtain collateral information. Whenever possible, consent to contact others should be obtained. Written consent is preferable to verbal consent. In potentially life-threatening situations such as suicidal ideation or gestures, homicidal ideation or threats, or an inability to communicate, the emergency takes precedence over confidentiality, and significant others should be contacted even without consent. In these cases, the clinician should carefully document in the medical record the reasons for breaking confidentiality.

An important aspect of the interview is its termination. Although there is considerable pressure on the clinician to work rapidly, the interview should be structured to make the patient feel validated and that they has been given sufficient time to present their story. Allowing a patient an opportunity to present information independent of the structured interview can often deescalate a tense interaction, and the clinician should validate the experience of the patient. The clinician should give a summary of their understanding of the patient's problem, followed by recommendations for further care.

Difficult Patients

Difficult patients encompass a wide category of diagnoses. A common theme with these patients is their ability to induce extremely negative reactions in PES staff and clinicians. Most often, patients with personality disorder-trait specified (PD-TS) or formal personality disorder (particularly borderline personality disorder), those who repeatedly present for care, and those with addiction-related behaviors which might be demanding, manipulative, and entitled may fit this category. Difficult patient populations often demand immediate attention, have a poor frustration tolerance, and relate to others in a negativistic manner. As a result, clinicians may feel anxiety, ambivalent anger and rage, and even fear. It is essential to remain objective and ensure appropriate professional responses to the patient. The clinician's responses to challenging patient populations can have a significant impact on treatment and outcome.

Because of the wide diagnostic variability with difficult patients, similarly, diverse clinical presentations may be seen in the PES. Patients may come voluntarily or may be brought in by police for disruptive or violent behavior. Those with borderline traits or personalities may present with suicidal ideation, suicidal gestures, suicide attempts, homicidal ideation or gestures (and rarely attempts), drug abuse, brief psychotic episodes, and impulsive behavior, often precipitated by interpersonal conflicts. Patients with antisocial personality disorders often use the emergency service for the purpose of secondary gain, such as obtaining medication, avoiding legal issues, or using the prospect of hospitalization to escape from some psychosocial predicament.

The management of these challenging populations is similar, beginning with a clear understanding of acceptable behavior and firm but supportive limit setting by the clinician. An effective strategy to accomplish this is clearly stating that certain behavior is unacceptable and the reasons why it is unacceptable. The patient should be offered two therapeutic alternatives. For example, ***you cannot keep punching the walls because your behavior is scaring other patients, and you could injure yourself. Therefore, you can take some medication or lie down in a quiet room until you feel more in control.*** It is important to understand, especially in dealing with patients who are involuntarily being observed, that a loss of autonomy can precipitate untoward behaviors. In providing a patient with alternatives to maladaptive behaviors and allowing them to make decisions about their own care, a clinician can potentially deescalate a tense situation. Threatening the patient with punitive consequences will exacerbate the negative behavior and may lead to further acting out behaviors. Medication such as benzodiazepines or antipsychotics may be useful in reducing the patient's agitation, which is discussed further in this chapter. Effective communication between staff members, clinicians, and the patient is essential, as these patients may be manipulative or try to divide the PES staff. The strategies discussed earlier in dealing with agitated and aggressive patients are effective in dealing with difficult patients. These patients may require more time for an evaluation in order to discharge their affect, regain control of their behavior, and realistically discuss their options. Effective interpersonal intervention may avert hospitalization. A primary diagnosis of adjustment disorder, mood disorder, or brief psychosis will show that the decision to involuntarily hospitalize a patient is based on an illness rather than a pattern of maladaptive behavior. However, involuntary commitment may be necessary for patients with unremitting homicidal or suicidal ideation to ensure safety.

Pharmacologic Management of Psychiatric Emergencies

In the PES, a high priority is given to the pharmacologic treatment of agitated patients to reduce the incidence of patient and staff injuries and to reduce the patient's psychological discomfort. Management of agitation and aggression is complex because these nonspecific symptoms can occur in a wide variety of clinical conditions. These may include delirium, dementia, alcohol and drug intoxication or withdrawal, personality disorders, and psychosis secondary to primary psychotic illnesses. At times, patients are so agitated that they are unable to cooperate with a psychiatric or medical evaluation and are incapable of providing any relevant information.

Pharmacologic management may initially serve as the primary therapy or as an adjunct to a verbal intervention. Clinically significant agitation may be defined as abnormal and excessive verbal or physical aggression, purposeless motor behaviors, heightened arousal, and significant disruption of the patient's functioning.

Typical behaviors of clinically significant agitation include:

- explosive and/or unpredictable anger
- intimidating behavior
- restlessness, pacing, or excessive movement
- physical and/or verbal self-abusiveness
- demeaning or hostile verbal behavior
- uncooperative or demanding behavior or resistance to care
- impulsive or impatient behavior
- low tolerance for pain or frustration

The goal of pharmacologic intervention is to calm the patient without sedation so that they can participate in the evaluation and treatment plan. Whenever possible, the patient should be given the option of the route of administration as this can facilitate their sense of having some measure of control. The use of oral liquid or dissolving tablets is the least invasive pharmacologic intervention and allows the patient to have a feeling of control and participation in treatment.

The most frequently used medication strategies consist of benzodiazepines, second-generation antipsychotic medication alone or in combination with a benzodiazepine, and haloperidol alone or in combination with a benzodiazepine. The most commonly used benzodiazepine is lorazepam.

Studies of the pharmacologic treatment of agitation in a busy PES are methodologically challenging. A treatment algorithm for the management of acute agitation or aggression based on the extant literature is as follows:

- Agitation due to a primary psychiatric illness
  - Orally disintegrating or liquid risperidone (2 mg) or orally disintegrating olanzapine (5 to 10 mg) is a first-line intervention for agitation.
  - An oral benzodiazepine can be added if deemed necessary and should be done so before adding additional antipsychotic doses.
  - For severe agitation secondary to psychosis when oral medications are not possible, IM ziprasidone (20 mg) that can

    be supplemented with IM lorazepam (2 mg) is first-line to limit the potential for side effects.
  - Second-line recommendations for severe agitation secondary to psychosis are IM haloperidol (5 mg) and IM lorazepam (2 mg), or IM olanzapine (5 to 10 mg). ***Lorazepam should not be used in combination with olanzapine because of the risk of cardiorespiratory depression.***
- Agitation due to stimulant intoxication OR alcohol withdrawal
  - Benzodiazepines are first line, such as Lorazepam 1 to 2 mg or diazepam 5 to 10 mg.
  - Oral medications should be attempted prior to IM or IV medications when possible.
- Agitation due to alcohol or other CNS depressant intoxication
  - Haloperidol (2 to 10 mg) has been shown to be safe and has low concern for additional respiratory compromise, therefore is a first-line treatment.
  - Second-generation antipsychotics (either PO or IM) are the best alternatives, if needed.
  - Benzodiazepines should be avoided due to the risk of respiratory depression.

Medication may be given every hour, up to three or four doses every 24 hours, though most patients will respond to a single dose of medication. The most common side effects of antipsychotic medications are dystonic reactions that can be effectively treated with IM benztropine (2 mg). Akathisia may develop with the use of highly potent agents and should be monitored. Management is with PO beta-blockers such as propranolol, and patient management can be optimized with the concomitant use of benzodiazepines. The most common side effects of benzodiazepines are sedation and ataxia.

Pharmacologic management of agitation in children and adolescents in the PES setting is a complex and difficult task for the psychiatric provider. Although the number of atypical antipsychotics FDA approved for use in children has increased, the use of these medications is complicated by the indications varying based on age and diagnosis. However, in emergency situations, a consensus of

PES medical directors responded that risperidone or lorazepam are their preferred medications.

Unfortunately, there are also no outcome studies or formal guidelines for treating agitated pregnant patients. In general, clinicians should use the minimal amount of medication necessary to reduce agitation and aggression in these two groups of patients and work collaboratively with specialty services to ensure consensus care as a treatment team. It is important to weigh the risks of acute agitation with that of a one-time dose of an antipsychotic or benzodiazepines. Mood stabilizers are not typically helpful in the rapid resolution of agitation symptoms.

Management of Agitation and Aggression

Agitation is a nonspecific symptom that occurs across all medical and psychiatric diagnoses. Untreated agitation can escalate to violence. Managing agitation in the emergency setting is complicated by the need to make clinical decisions on the basis of relatively incomplete psychiatric and medical assessments and histories. Severe agitation may make it impossible to obtain an initial medical history, complete a medical examination, or collect screening laboratory studies. Once agitation subsides and the patient is more cooperative, a complete medical history and evaluation should be performed as medical illness can precipitate violent and aggressive behavior. The differential diagnosis for violence is summarized in Table 31.3–5.

Table 31.3–5.
Differential Diagnosis of Violent Behavior

Medical
Cerebral infection
Cerebral neoplasm
Electrolyte imbalance
Hepatic disease
Hypoglycemia
Hypoxia
Infection
Renal disease
Temporal lobe epilepsy
Vitamin deficiency

Substance-Induced
Alcohol intoxication
Alcohol withdrawal
Stimulant intoxication
Delirium tremens
Inhalant intoxication
Phencyclidine (PCP) intoxication
Sedative/hypnotic withdrawal
Synthetic cannabinoid intoxication

Psychiatric
Antisocial personality disorder
Bipolar disorder
Borderline personality disorder
Schizophrenia
Decompensated obsessive-compulsive disorder
Delusional disorder
Dissociative disorder
Impulse control disorder
Post-traumatic stress disorder
Schizophrenia
Schizoaffective disorder

Interventions for agitation should be conceptualized as hierarchal, beginning with the least restrictive and least intrusive interventions and only using the most restrictive interventions when all other treatment approaches have been exhausted. The core clinical strategy for managing agitation is the use of interpersonal strategies that emphasize verbal deescalation interventional techniques. It is important for the clinician to reduce the patient’s anxiety and fear by maintaining a humane and respectful disposition. A patient who is treated with honesty, dignity, and respect is likely to respond with reciprocity if there is retained reality testing. Affect management is central to any effective aggression management technique. It involves acknowledging the patient’s affect, validating it when appropriate, and encouraging the patient to talk about their feelings. Attempting to use logic and rational thought with a patient who is effectually dysregulated may lead to a further increase in agitation. An emotionally distraught patient requires an active response from the clinician. The use of active listening techniques, such as paraphrasing to the patient a brief encapsulated form of their statements, helps convey that the clinician is attempting to understand what the patient is experiencing. It is important to be honest and unambiguous. Verbal deescalation techniques require time and are often sufficient to reduce agitation and eliminate the risk of violence. Offering food or drink may also help facilitate a therapeutic alliance and reduce a patient’s agitation. Management of aggressive patients may be compromised when the clinician inadvertently evokes feelings of fear, anger, or rejection. This can generate thoughts of retaliation aimed toward the clinician. If identified, the clinician should stop the interview and ask another staff member to continue with the therapeutic intervention.

Patients with a history of documented or witnessed assaultive behaviors should be provided a significant amount of space by all clinical providers. A general rule of thumb is to keep at minimum two quick steps or at least two-legs lengths distance from the patient. Personal space can be visualized as an oval zone extending 4 to 6 ft all around. If the patient is standing, then the clinician should stand. If the patient is sitting, then the clinician should also sit down, preferably at the same level as the patient. The clinician should never stand over the patient during the interview as this may further escalate patients and places the clinician in a compromised position for potential harm. If the patient is pacing, then the clinician can walk with the patient but at a much slower pace. The clinician should take a posture that makes them appear small and thus less threatening. This can be done by holding the hands at waist level with palms up and open. This posture ensures the patient, especially if paranoid, does not misperceive the intentions of the clinician and ensures that there are no syringes or other devices of perceived harm in the practitioner’s hands. Assuming the “Thinker” stance (one forearm crosses the chest, and the opposite elbow rests on it with the index finger on the cheek or chin) is also an effective technique. In this position, the patient’s content can be validated by affect but also ensure the practitioner is ready for any unexpected behaviors by the patient. Some patients may not respond or only partially respond to verbal intervention and will require psychotropic medication to aid in behavioral management.

When both interpersonal and pharmacologic interventions fail to reduce agitation, and the patient remains uncooperative with treatment efforts, a show of force with five or more staff and/or security guards may encourage the patient to cooperate with treatment. The presence of additional staff supporting the clinician conveys to the patient that their aggressive impulses are not appropriate in the given circumstances and will need to be controlled. A show of force should not be implemented in a way that threatens or humiliates the patient but serves as a vehicle for ensuring the safety of the patient, patient peers, and staff.

When less restrictive interventions fail to reduce a patient’s agitation, physical restraints may be necessary to prevent imminent harm to the patient, patient peers, or staff. Prevention of serious disruption of the treatment setting, prevention of bodily harm, and/or prevention of significant damage to property is the goal, and restraints should be considered a last resort in the PES setting. Recent evidence suggests that the acute symptoms and presentation characteristics in a PES clinical encounter are more likely to be associated with restraint/seclusion than the patient’s underlying disease process or demographics. A careful approach to every patient is warranted, especially if there is a history of impaired impulse control or assaultive behaviors. PES providers should always be prepared for behavioral escalation. Once the decision is made to restrain a patient, the restraint process should be implemented immediately and without negotiation. Rigorous attention to the patient’s safety is paramount as restraints are contraindicated in the patient with an unstable medical condition such as infection, cardiac illness, body temperature instability, metabolic illness, pregnancy, or orthopedic problems. In addition, patients with delirium or dementia may experience worsening symptoms secondary to the sensory isolation induced by restraints and seclusion. In all restraint episodes, documentation should clearly outline the behavior requiring restraints, the interventions that were made to reduce the patient’s agitation prior to restraints, and all interventional efforts to remove the patient from restraints. While a patient is in restraints, continuous monitoring of the patient should occur to prevent injury with 15-minute checks of extremities to insure adequate circulation, adequate hydration, and limb exercise. All clinical efforts should focus on removing the patient from restraints as quickly as is clinically possible. All clinicians should review and thoroughly understand the restraint guidelines, policies, and procedures of their respective institutions.

A PES should have written policies and procedures that are understood and practiced by all clinical staff. Patients should be evaluated and treated in areas that have adequate space, appropriate lighting and alarm systems, and the ready availability of security officers. Interviewing rooms should have multiple exits and a minimum of materials and furniture that could be used as weapons. As a clinician, it is important to understand your physical and psychological limitations in dealing with the unpredictability of a psychiatric emergency setting. Ensuring the safety of patients and staff is the first priority and will mitigate the risk of a potential secondary emergency. Be aware of your surroundings, communicate effectively, and utilize resources available on site.

PSYCHIATRIC EMERGENCIES RESULTING FROM SUBSTANCE INTOXICATION

Intoxicated patients tend to evoke negative responses from hospital staff, which in turn may further agitate an intoxicated patient. As with all difficult patients, personal feelings should not be allowed to influence professional judgment and clinical care. Utilization of all available resources is the standard of care to ensure safety, as intoxicated patients can be unpredictable. Security forces, though seldom needed, should be alerted before interviewing an intoxicated patient. If possible, the interview should be conducted in a large comfortable room. The clinician should assume a tolerant, nonthreatening manner and be prepared for potential verbal insults and physical threats. Offering food, coffee, or juice may also be helpful, as well as other strategies for intervention with aggressive patients.

Consideration of patient needs longitudinally will require an understanding of each patient's level of intoxication or potential for withdrawal, medical comorbidities, emotional and psychiatric comorbidities, acceptance or rejection of treatment, relapse potential, and availability of a supportive environment. Due to the heterogeneity of institutional resources, it is important for each provider to become familiar with available resources, both within the care setting and within the community. Resources such as The Patient Placement Criteria developed by the American Society of Addiction Medicine can be very helpful for determining the level of aftercare for patients with substance-related pathology.

Stimulant Intoxication

Sympathomimetics substances, also known as stimulants, which include cocaine and amphetamine salts, also are relatively short-acting substances with rapid effects when snorted, smoked, or used intravenously (IV). Symptoms of acute stimulant intoxication include a heightened sense of well-being, increased sense of alertness, decreased anxiety, reduced social inhibitions, increased talkativeness, and poor judgment manifested by loss of money, sexual indiscretion, and potentially illegal activities. Patient symptoms usually resolve within 48 hours. Stimulant intoxication may also present with grandiosity, psychomotor agitation, paranoid ideation, formication (sensation of insects crawling on the skin), bizarre behavior (such as sorting objects into various groups), stereotyped movements of the mouth and the tongue (such as teeth grinding, lip biting, and skin picking), tactile or visual hallucinations, and distorted perceptions. Delusional states, especially paranoia, can also occur.

Physical findings include tachycardia or bradycardia, hypertension or hypotension, hyperthermia, dilated pupils, perspiration or chills, hyperreflexia, nausea, and vomiting. Patients may have evidence of weight loss, muscular weakness, respiratory depression, chest pain, or cardiac arrhythmias. Intoxication can lead to confusion, seizures, or coma. Patients who abuse stimulants IV are at risk for numerous medical conditions, including endocarditis, tetanus, hepatitis, abscesses (MRSA), and AIDS. Urine and serum drug screens provide corroboration of the diagnosis and should be part of the PES intervention, if suspected or if there is a need for determination of diagnosis based on readily available information.

Treatment should include placing the patient in a quiet room away from the general traffic of the PES as external stimuli may exacerbate the physiologic effects of sympathomimetics. Explaining the treatment in simple language coupled with reassurance to the patient that the effects of the stimulant will subside may aid in providing comfort. Because the patient should not be left alone, family members or friends can be included in the treatment to provide further reassurance and support, if safe to do so. Strategies discussed earlier in this section for managing aggressive patients should be utilized. If interpersonal interventions fail to relieve the patient’s agitation, a benzodiazepine can be administered to calm it. If the patient does not respond or is delusional, extremely agitated, or combative, then an antipsychotic medication with or without a benzodiazepine can be given as per the medication algorithm discussed earlier. The patient should be discharged to the care of family or friends with a follow-up outpatient appointment for the next day in a drug treatment program, if possible.

Hospitalization should be considered if:

- the psychosis does not remit and affects the ability to negotiate their needs.
- the patient has a polysubstance abuse problem and requires detoxification.
- the patient has no family or other support system, and appropriate housing or shelter cannot be arranged.
- the patient has protracted cognitive impairment.
- the patient has serious associated medical problems.

Alcohol Intoxication

Patients who are intoxicated with alcohol in the PES may be self-referred or brought in by police, friends, or relatives because of self-destructive, disruptive, or violent behavior. It has been estimated that more than one-half of murderers and their victims are intoxicated at the time of the act. In addition, approximately one-quarter of deaths by suicide in the United States involves alcohol.

Blood alcohol levels usually correlate with the clinical presentation of the alcohol-intoxicated patient:

- 30 mg/dL: attention difficulties and/or mild euphoria
- 50 mg/dL: coordination problems and/or unsteadiness
- 100 mg/dL: impaired memory and/or ataxia
- 200 mg/dL: confusion and/or stupor
- >400 mg/dL: seizure, coma, respiratory arrest, and/or death

Chronic alcohol use disorder patients will have a higher tolerance to the effects of alcohol, so blood levels of alcohol will not always correlate with the clinical presentation. Blood alcohol levels should always be obtained in the assumed intoxicated individual. Alcohol is metabolized at the rate of about 1 oz per hour or 10 to 30 mg per hour in patients with healthy hepatic functioning.

Symptoms of alcohol intoxication include aggressive or inappropriate behavior, mood lability, impaired judgment, slow or slurred speech, incoordination, unsteady gait, and nystagmus. Objective physical findings of intoxication may include autonomic instability; tremulousness; nausea and vomiting; transient visual, tactile, or auditory hallucinations or illusions; psychomotor agitation; and anxiety. Progression to withdrawal may be rapid in chronic and heavy users of alcohol, and symptoms should be monitored closely in patients with complicated use histories. At a minimum, vital signs and neurologic function checks should occur every 4 hours for patients at high risk for withdrawal.

Medical complications are common in chronic use of alcohol. These include but are not limited to: gastrointestinal bleeding, gastritis, stomach or duodenal ulcers, cirrhosis, pancreatitis, vitamin deficiencies, cardiomyopathy, peripheral neuropathy, and malnutrition. In an acute presentation, patients may have sunburn exposure and subdural hematoma from head trauma, thus making an exploration of patients' recollection of events prior to arrival imperative. Duration of use, quantity used, and timeline of the most recent exposure must also be explored. Other drugs may also have been ingested, thus further complicating the clinical picture. Death may occur from aspiration of vomitus or respiratory depression, if not clinically identified.

The first treatment priority is to protect the acutely intoxicated patient from harming him- or herself or others and to identify medical problems that require immediate attention. A patient may be hyper- or hypoglycemic, bleeding internally, hypertensive, have end-stage cirrhosis, a cardiac arrhythmia, pneumonitis, hematoma(s), or one of the alcohol-related syndromes such as alcohol-induced persisting amnestic disorder or Wernicke encephalopathy. A person may appear obtunded by alcohol, but clouding of consciousness from hyper/hypoglycemia, blood loss, or a subdural hematoma may be the cause. Clouded consciousness in the clinical presentation of a negligible serum alcohol level is concerning for medical pathology warranting intervention. The patient's vital signs should be obtained every 15 minutes if possible, and depending on the patient's medical history and physical examination, appropriate laboratory studies should be obtained. If control of agitation is required, then the use of antipsychotics should be used, as described earlier in this section. Benzodiazepines should be avoided, if possible due to the increased risk of respiratory depression. Patients may require a higher level of care than may be offered in the PES setting, and emergency medical personnel should be involved in care as acute intoxication, especially in the context of the potential for withdrawal, can be a medical emergency.

Disposition should be to home with family, if available, or other community placement if the patient presents no immediate danger and has no acute medical problems. The patient should be given a follow-up appointment in an outpatient program and referral to alcoholics anonymous (AA). Many facilities offer on-site substance use counselors, which may be beneficial in aiding in longitudinal planning as well.

Benzodiazepine, Sedative, and Hypnotic Intoxication

Benzodiazepine, sedative, or hypnotic intoxication is marked by impaired judgment, loquacity, mood lability, dreamlike feelings, and a disinhibiting of sexual and aggressive impulses. Slurred speech, incoordination, unsteady gait, lateral nystagmus, areflexia, muscle hypotonicity, hypotension, and hypothermia are common physical findings.

Management of benzodiazepine, sedative or hypnotic intoxication is a complex medical intervention best left to clinicians who have experience in treating such cases. These patients should be treated in the medical emergency department with continuous monitoring of vital signs and immediate availability of airway protection interventions. Barbiturate overdose, in particular, is a medical emergency. Most patients require admission to a medical intensive care unit for continued evaluation and treatment. Patients may require management for a comorbid psychiatric illness, but intervention should be deferred until medical stability is ensured.

Opioid Intoxication

Symptoms of opioid intoxication include drowsiness, slurred speech, and potentially coma. Physical findings include decreased respiration, blue lips, blue or pale skin, pinpoint pupils (if brain damage has occurred, usually secondary to anoxia, pupils may be dilated, or meperidine can cause mydriasis), hyperemic nasal mucosa, recent needle marks, pulmonary edema, cardiac arrhythmias, or seizure.

Death is secondary to pulmonary or cerebral edema.

If a patient is unstable due to the direct physiologic effects of opioids, then the signs and symptoms are rapidly reversed by the administration of a narcotic antagonist, for example, naloxone. Formulations of naloxone include injectable, auto-injectable, and nasal spray. Symptoms of opioid intoxication may be complicated by varying levels of intoxication or withdrawal from other drugs. Therefore, patients should have urine and serum drug screens. Because of the life-threatening risk or its potential medical complications, opioid intoxication or overdose should be treated in the medical emergency department with continuous monitoring of vital signs and immediate availability of airway protection interventions. Patients may require management of a comorbid psychiatric illness, but intervention should be deferred until medical stability is ensured.

Cannabis Intoxication

Cannabinoids include substances such as hashish, bhang, marijuana, and purified Δ9-tetrahydrocannabinol (THC). Abusers commonly use cannabinoids with other substances, particularly alcohol and cocaine. The most common methods of use are smoking and oral consumption. Drug effects develop within 20 to 30 minutes. There has been an increasing prevalence of the use of cannabinoids due to recent legalization in various jurisdictions. Providers should always inquire about use independent of legality as cannabinoids may directly influence the psychiatric phenotype of the patient in the PES.

Clinical findings of acute intoxication include:

- Impaired motor coordination
- Euphoria
- Anxiety
- The sensation of slowed time
- Impaired judgment

- Social withdrawal

Within 2 hours of cannabis use, the patient may develop the following:

- Conjunctival injection
- Increased appetite
- Dry mouth
- Tachycardia

Marijuana intoxication usually will not be the primary reason for evaluation in the PES unless an adverse reaction to the drug occurs. Treatment involves reassurance and a calm environment. Symptoms will usually clear within 4 to 8 hours. If the patient's anxiety or panic becomes severe, then a benzodiazepine can be given. Paranoid delusions are rare but may require antipsychotic medication. If a psychotic reaction lasts more than a few days, then the patient should be evaluated for a comorbid psychiatric illness.

It is important to note that patients may have exposure to synthetic cannabinoid products in presenting to the PES. Patients may endorse the use of "marijuana" use only, but synthetic cannabinoids have seen recent increases in incidence rates in the emergency setting. Patients may present with symptoms similar to naturally occurring cannabinoids but can develop severe tachycardia, hypertension, marked restlessness, agitation, anxiousness, nausea with emesis, mydriasis, somnolence, hallucinations, the elevation of blood glucose, hypokalemia, and even seizures. Patients may require supportive medical interventions to ensure stability, and physicians may need to treat symptomatically underlying psychopathology, especially when agitation and psychosis are salient features. If the presentation of a patient is inconsistent with the use of cannabinoids alone, clinicians should ensure thorough screening for other substances, including synthetic cannabinoids. Toxicology testing, whether urine or serum, may have limitations in its utility as synthetic compounds are not regularly part of commercially produced screening tests. Patients should be monitored closely, and if a psychotic reaction lasts more than a few days, the patient should be evaluated for a major psychiatric illness.

Hallucinogen Intoxication

Hallucinogens include the following:

- Lysergic acid diethylamide (LSD)
- Psilocybin (“magic mushrooms”)
- Dimethyltryptamine (DMT)
- Peyote
- Dimethoxy methylamphetamine (DOM or “STP”)
- Mescaline
- Phencyclidine (PCP)
- Ketamine
- Methylenedioxymethamphetamine (MDMA, “Ecstasy,” or “X”)

Hallucinogens cause no known withdrawal symptoms. Unfortunately, drugs are often mixed, adulterated, and sold under a different name and, therefore, clinicians can never be certain what hallucinogen was ingested on the basis of a patient’s provided history. Therefore, serum and urine screens must be obtained. Pre-existing psychiatric or medical illness or premorbid personality disorders may intensify hallucinogenic symptoms as the effects of substances are not homogenous among users. The symptom complex present in the clinical setting is often multifactorial and the result of drug–drug interactions and the dose ingested. The duration of action and rate of onset of symptoms is also heterogeneous. Additionally, a user’s premorbid personality, expectations, and social setting can directly affect the behavioral phenotype and described effects. Hallucinosis typically lasts 16 hours to 2 or 3 days.

Symptoms include perceptual, emotional, and cognitive changes. Patients may see colorful mobile displays of geometric designs or complex human or animal forms. Colors appear more intense, and fixed objects undulate and flow. The patient has a greater sensitivity to touch and an altered sense of taste and smell. Synesthesia may occur (an overflow of one sensory modality into another; for instance, the patient “hears” colors or “sees” music). Other perceptual changes include depersonalization, derealization, and body image distortions in which body parts seem to grow, shrink, or disappear.

Emotional presentations range from euphoria, elation, and bliss to tension and anxiety and may culminate in a panic attack. The patient may laugh or cry for prolonged periods of time. Cognitive changes include loosening of associations with unusual content, illogical and fantasy-laden thought sequences, paranoid grandiosity, and, less commonly, persecutory delusions. Physical findings include hypertension, hypothermia, dilated pupils, tachycardia, sweating, palpitations, tremors, blurred vision, incoordination, and hyperreflexia.

Hallucinogen intoxication usually will not present to the PES unless an adverse reaction to the drug occurs or a patient places themselves in imminent danger due to impaired judgment and is brought in by a concerned family member or agency. The primary intervention involves reassuring the patient that the drug’s effects will subside within several hours. It is helpful to remind the patient that the hallucinogen has produced their extraordinary ideas and feelings and that the drug effects will soon dissipate. When possible, friends and family members can help alleviate the patient’s anxiety by maintaining a continuous dialogue in a supportive, nonthreatening environment. Because closing the eyes can intensify the hallucinogenic state, the patient should sit up or walk. Clinicians should try to avoid pharmacologic intervention. If the patient’s anxiety becomes overwhelming, then a benzodiazepine can be given. Severely agitated patients may require an antipsychotic agent.

Phencyclidine Intoxication

Originally marketed as an anesthetic agent, phencyclidine (PCP) was taken off the market because of the psychotic reaction experienced by patients postoperatively. PCP, commonly called “angel dust,” is an illegal drug often sold to consumers under many names and may be sold as “crack” or “LSD” (it is neither). The drug may be taken orally, inhaled, snorted, or injected IV. PCP presents a confusing clinical picture with diverse psychological, physiologic, and behavioral effects and may mimic schizophrenia. PCP highlights the importance and utility of urine and serum drug screening in complicated patient presentations.

Patients can become extremely violent when intoxicated with this illicit substance. It can create a significant paranoid psychosis resulting in harm to others and, when taken in sufficient amounts, can produce death from respiratory failure, cardiac arrhythmias, and adrenergic crisis. Clinical findings are diverse and can include calmness to the point of sleep, agitation, hostility, disorientation, emotional lability, a sense of slowed time, out-of-body feelings, unpredictability, impaired judgment, hallucinations, assaultive behavior, and paranoid ideation. Because of the amnestic quality of PCP, patients will often not remember how they behaved while intoxicated.

Clinical signs and symptoms tend to correlate with PCP blood levels and include vertical and horizontal nystagmus, ataxia, loss of pain response, drooling, nausea, and mild pulse and blood pressure elevation at low doses. Higher blood levels include catatonia, rigidity, myoclonus, fever, elevated pulse and blood pressure, hyperreflexia, and elevated creatinine phosphokinase secondary to muscle trauma. PCP intoxication can proceed to death as the dose increases.

Patients with mild to moderate intoxication should improve with supportive care. The patient will require supervision in a nonstimulating environment that should include protection from self-injury. If symptoms remit and there is no cognitive impairment after 12 hours, then the patient can be discharged from the PES. Moderate to severe intoxication will require supportive care in an emergency department or medical inpatient unit with special attention to respiratory and cardiac function. It is often important to obtain serum creatinine kinase levels, especially in the context of significant abnormalities in vital signs. Patients may require IV fluids to ensure that renal damage does not ensure when creatinine kinase levels are markedly elevated. Vital signs need to be monitored for laryngeal stridor and/or respiratory depression. Restraints should be avoided to avoid the increased risk of rhabdomyolysis and resultant acute renal failure.

The preferred method of physical restraint for combative patients is total body immobilization by rolling the patient in a sheet. Sedation with a benzodiazepine may be required for anxiety, agitation, or harmful behavior. An antipsychotic may be given for psychosis.

Inhalant Use

Inhalants can cause irreversible damage. They can cause death by suffocation, aspiration vomitus, or myocardial infarction. Patients who abuse inhalants sniff aerosols, cleaning solutions, nail polish remover, lighter fluids, paint thinner and other petroleum products, or glue containing toluene. Intoxication occurs within 5 minutes and ceases in 1 to 2 hours.

Clinical findings include belligerent or assaultive behavior, apathy, impaired judgment, and impaired social and occupational functioning. Physical findings include dizziness, nystagmus, slurred speech, unsteady gait, depressed reflexes, tremors, blurred vision, diplopia, stupor, coma, rash around the nose and mouth, breath odors, residue on the face, hands, and clothing, redness and swelling of the eyes, and irritation of the throat, lungs, and nose. Major medical problems associated with inhalant abuse include cardiac arrhythmias, hypoxia, hepatitis, renal failure, bone marrow suppression, impaired pulmonary functioning, and skeletal muscle weakness. Death can result from central nervous system depression, cardiac arrhythmias, bronchospasm, and ventricular fibrillation.

Care is mainly supportive. Patients should be kept in an environment with minimal stimulation. Clothes smelling of solvents should be removed, and solvents on skin or hair should be washed away with mild soap and water. Marked agitation should be controlled with antipsychotic medication as benzodiazepines may potentiate the effects of inhalants. Hospitalization may be warranted. However, chronic inhalant users often have polysubstance involvement and clinicians should carefully screen for other substances and treat accordingly.

PSYCHIATRIC EMERGENCIES RESULTING FROM SUBSTANCE WITHDRAWAL

Withdrawal from Alcohol

Alcohol withdrawal is a medical emergency. If left untreated, it may lead to death. Clinical signs of alcohol withdrawal include disorientation, fluctuating attentiveness, and intermittent agitation, commonly violent in nature. Physical findings include tachycardia, sweating, hypertension, nausea, vomiting, tremulousness, hyperreflexia, orthostatic hypotension, and, occasionally, generalized seizures. Patients in withdrawal are usually depressed, irritable, and anxious and may experience illusions or transient visual, tactile, olfactory, and auditory hallucinations. Hallucinations are usually frightening—the patient sees and feels mice or lice crawling on the skin or sees animals, especially snakes. Alcohol withdrawal can be complicated by seizures or delirium tremens. Delirium tremens is the most severe form of ethanol withdrawal. It can manifest clinically with an altered level of consciousness and/or global confusion coupled with autonomic hyperactivity (specifically the sympathetic nervous system), which can lead to cardiovascular damage or failure.

Death occurs in 4% to 20% of cases from complications such as hyperthermia, aspiration, or vascular collapse. The onset of complicated withdrawal is usually 2 to 10 days after cessation or decreases in alcohol use and may last hours to days. Symptoms of uncomplicated alcohol withdrawal peak 24 to 48 hours after the last drink and subside within 5 to 7 days, even without treatment. Irritability and insomnia may last for 10 days or more. Withdrawal is more severe in patients with previous episodes of withdrawal or medical illnesses.

Alcohol withdrawal symptoms are generally effectively treated with benzodiazepines. Chlordiazepoxide (25 to 100 mg every 4 to 6 hours and as needed) provides for smoother detoxification given its long half-life. Lorazepam (1 to 2 mg orally or intramuscularly every 6 hours and as needed) is preferred when IM medication is necessary, or patients have markedly impaired liver function. Lorazepam is safer for patients with hepatic disease or brain damage due to renal excretion and a shorter half-life. An antipsychotic, as recommended earlier in this section, could be used for patients with hallucinations. Pulse, blood pressure, and temperature must be monitored every 30 minutes in the PES. One hundred milligrams of thiamine is given IM to address expected vitamin deficiency. Delirium tremens or a seizure is a medical emergency and must be treated in the medical emergency department in a monitored setting. If a patient has a documented or verbalized history of alcohol withdrawal-related pathology, it is considered a medical emergency, and patients will require monitoring and potential hospital admission, independent of any comorbid primary psychiatric pathology.

Withdrawal from Benzodiazepines, Sedatives, and Hypnotics

Withdrawal from benzodiazepines, sedatives, and hypnotics is similar to alcohol withdrawal. Physical findings include tremors, increased pulse and respiration, hyperthermia, loss of appetite, nausea and vomiting, postural hypotension, hypertension, seizures, and delirium. Patients who abuse sedatives are at risk for infections and hepatitis.

Withdrawal from short-acting agents will have an onset 12 to 24 hours after the last dose, peaking between 24 and 72 hours. Withdrawal symptoms with longer-acting agents peak 5 to 8 days after the last dose. The older adults or patients with hepatic disease may develop symptoms more slowly because of decreased drug metabolism.

There are several pharmacologic options for withdrawing a patient from benzodiazepines, sedative, and hypnotic drugs.

- One method is to gradually withdraw the agent of dependency.
- Substitute the dependent agent with phenobarbital or a longer-acting benzodiazepine and taper. For short half-life benzodiazepines, treatment would be with a longer-acting drug such as clonazepam or chlordiazepoxide.
- In cases of barbiturate or other mixed sedative/hypnotics, a phenobarbital taper can be used, as this will allow for safer detoxification because it produces few changes in blood levels

between doses, it has observable signs of toxicity, and intoxication does not produce euphoria. A phenobarbital taper is done as follows:

1. Administer a 200-mg test dose of pentobarbital orally and assess neurologic status after 1 hour. Pentobarbital can be converted to phenobarbital, which has a longer half-life with fewer variations in blood level (100 mg of pentobarbital is equivalent to 30 mg of phenobarbital).
2. If the patient is asleep but arousable, then no treatment is required. If the patient is drowsy with slurred speech, coarse nystagmus, or ataxia, then the patient's estimated pentobarbital requirement is 400 to 600 mg per 24 hours.
3. If the patient is comfortable with only fine lateral nystagmus, then the 24-hour pentobarbital dose is 800 mg.
4. If the pentobarbital has no effect, then the 24-hour pentobarbital dose is estimated at 1,200 mg or more.

Patients who are withdrawn from sedative-hypnotics should be admitted to the hospital for the withdrawal protocol because this is a life-threatening condition. Clinicians should also familiarize themselves with the dose equivalents of benzodiazepines and other sedative drugs.

Some common benzodiazepine dose equivalents are:

- Alprazolam (Xanax) 0.5 mg
- Chlordiazepoxide (Librium) 10 mg
- Clonazepam (Klonopin) 0.25 mg
- Diazepam (Valium) 5 mg
- Lorazepam (Ativan) 1 mg
- Oxazepam (Serax) 15 mg

Withdrawal from Opiates

Patients' opioid use disorders may present to the PES requesting detoxification or, more commonly, present in withdrawal requesting relief. In addition, patients may have significant mood symptoms that may potentiate the untoward effects of withdrawal. Cessation of opiates, while uncomfortable, is not life-threatening. Opioid withdrawal is precipitated by the abrupt cessation of opioid administration. When heroin, morphine, or short-acting prescription opioid use is discontinued, withdrawal symptoms can appear within 6 to 8 hours, peak within 36 to 72 hours, and last for 7 to 10 days. Meperidine withdrawal may not begin until 1 to 3 days after the last dose, with symptoms dissipating after 10 to 14 days. Fentanyl, one of the most potent opioids, has both faster onset and resolution of withdrawal symptoms, with onset from 2 to 5 hours, peak within 8 to 12 hours, and resolution within 4 to 5 days.

Withdrawal symptoms from opioid use include nausea, vomiting, diarrhea, dilated pupils, muscle aches and pains, insomnia, lacrimation (tearing eyes), flu-like weakness, rhinorrhea (running nose), piloerection, sweating, flushing, abdominal cramps, increased temperature, yawning, and semen ejaculation. These symptoms can be managed with nonopioid adjunct medications. For example, ondansetron can be given for nausea, dicyclomine for gastrointestinal cramping, and ibuprofen for body aches. Although opioid withdrawal is not considered life-threatening, patients who are dehydrated or debilitated could potentially have life-threatening complications. Seizures and delirium tremens do not occur.

Recently it has become more common for the management of potential opiate withdrawal actively in the PES setting. Buprenorphine (with or without naloxone) is a partial ***u-***opioid agonist that has been shown to alleviate acute opioid withdrawal symptoms and be an effective long-term treatment for opioid use disorder. Although sedation can occur, it does not have the same risk for respiratory depression as other opioids. It is considered safe to initiate in the emergency department and can help bridge patients to outpatient addiction treatment. One caveat is that patients must already be experiencing withdrawal symptoms prior to initiating buprenorphine, as it may precipitate withdrawal. Induction dosing of buprenorphine is variable depending on the severity of the withdrawal, typically with a maximum dose of 32 mg per day.

Patients can also be referred to methadone maintenance or drug rehabilitation program for further evaluation and treatment, but clinicians should not initiate methadone management in the emergency setting. However, if a patient is in established treatment for opioids with methadone, efforts should be made to continue care. This includes contacting current providers once consent is obtained to prevent a lapse in care, if possible.

Withdrawal from Cocaine and Amphetamines

Withdrawal from amphetamines and cocaine produces dysphoria, hypersomnia or insomnia, increased appetite, apathy and lack of energy, muscle aches, chills, depressed mood, anxiety, irritability, and vivid and unpleasant dreams. Patients will have a heterogeneous presentation in sympathomimetic withdrawal and may be agitated or have psychomotor retardation. Drug cravings may be intense. Suicidal thoughts can occur, and in some cases, patients may experience paranoia or hallucinations. The severity of the symptoms will depend on the potency of the stimulant of use. With amphetamines, these symptoms commonly occur 3 or 4 days after a decrease or cessation of heavy amphetamine use. The depression can last for months, although symptoms may last only 8 to 48 hours, with milder symptoms persisting for several weeks.

There is no standardized detoxification regimen for stimulants. Usually, after several nights of sleep, the dysphoria and suicidal ideation resolve. A calming environment can decrease agitation, and clinicians should avoid a confrontational approach. Quiet time, food, and sleep are the best interventions. However, while in PES care, the patient has to be carefully evaluated for suicide risk and may require inpatient hospitalization. Many of these patients do not regularly use the health care system, and clinicians should always maintain a high index of suspicion for undiagnosed and/or untreated medical diseases. Because these patients are at high risk for relapse, an inpatient rehabilitation program should be considered. Given that antidepressants have a delayed onset of action, their use is of no value in substance-induced mood disorders in the PES. If the patient is discharged home, then it is preferable to have family or another supportive individual accompany the patient home from the PES.

Withdrawal from Inhalants

Chronic, heavy use of inhalants has been reported to precipitate withdrawal symptoms within a few hours to several days after cessation of use. These symptoms include tremulousness, anxiety, inattention, diaphoresis, diarrhea, restlessness, and insomnia. Withdrawal symptoms can be treated with benzodiazepines. Whether there is an actual withdrawal state from inhalants is controversial

OTHER PES-RELATED ISSUES

Domestic Abuse

Domestic abuse refers to a behavior pattern in which a person repeatedly inflicts physical injury, pain, fear, or mental anguish on another member of the household. The abuse may be imposed through physical, psychological, sexual, or economic means. Neglect occurs when a care provider fails to provide the basic necessities for living to a dependent individual. Given the marked prevalence of the problem and the difficulty of identifying domestic abuse and neglect, clinicians should maintain a high index of suspicion. Knowing the signs that suggest domestic violence can help the clinician to determine the necessary intervention. Although men are less likely to have a history of abuse than women, they are more likely to be attacked by their partners with knives, guns, or household objects and suffer serious injury or death.

Treating a patient’s physical injuries is the first priority. It must then be determined whether the patient’s significant other (or another household member) caused the injuries. The clinician should obtain a history of the current relationship, as well as a history of the previous types of abuse. Victims of abuse often will give inconsistent explanations for their injury. Frequently, the abused adult patient may minimize the pain suffered from the attack and impede investigating the abuse or claim that the abuser did not really mean any harm. Issues such as whether the patient should report the abuse and whether the patient should seek placement other than their home are complicated and require a great deal of patience and skill on behalf of the clinician. If a patient has a pattern of injuries sustained by an abuser, it is advised to separate them from the abuser if possible and consider long-term treatment to aid in recovery. Whether or not to seek further social services and legal aid also factors into the intervention. Mandatory reporting guidelines can vary between jurisdictions; therefore, it is very important to be familiar with local laws and reporting guidelines. Treatment options may include psychotherapy, special abuse programs, and self-help and advocacy groups. Knowledge of available community resources is important for PES staff to ensure appropriate care.

Elder Abuse

Older adult patients can be victims of various types of abuse: passive neglect, physical abuse, and emotional or sexual abuse. The most common form of abuse is neglect, although physical abuse and financial exploitations are also common. The clinician's high index of suspicion is a critical component in making the diagnosis. The first priority is always to treat the physical consequences of the abuse or neglect. An older adult patient may minimize the problem, appear afraid of the caregiver, and express fears about returning home. Patients should be asked about being hit, being physically harmed in the past, or being deprived of care or necessities such as food and clothing, if any suspicions arise in clinical care. Signs of neglect in an elder person may include malnourishment, dehydration, poor hygiene, or pressure sores. The misuse of drugs, including overdosing, underdosing, or inability to obtain prescribed medications, may also reflect abuse. Sometimes additional evidence of neglect can be obtained from the person(s) bringing the patient to the emergency department. They may have first-hand knowledge of the patient's domicile and can comment on the level of cleanliness, the functioning of utilities, and the adequacy of medical appliances such as walkers and commodes.

Risk factors for those that perpetrate elder or dependent abuse include substance abuse (especially alcohol), psychiatric illness, psychosocial stressors, emotional and financial dependency, history of abuse themselves, social isolation, inexperience in caregiving, and misinformation as to how to provide care. In addition, there are risk factors specific to institutional care settings. Staffing shortages and provider burnout, lack of oversight, and lack of sympathy for residents have all been shown to increase the potential for elder abuse.

Interventions for those providing care to elders may include reducing the stress on the family by expressing empathy and providing support to the caregiver. Identifying interpersonal conflicts that underlie the abuse may also help alleviate tension in the family. Treatment for a patient, the abuser, and another family member will occur outside the hospital, but the clinician must inform the family of appropriate community resources and make appropriate referrals. If elder abuse is suspected, mandated reporting laws require a concerned clinician to report findings to the appropriate authorities. Clinicians should be familiar with the local statutes concerning this issue and the available mechanisms for reporting suspected abuse.

Child Abuse

The majority of child abuse cases reported annually are related to neglect, with smaller proportions related to physical abuse, sexual abuse, and emotional mistreatment. Because child abuse is underreported, the clinician should always maintain a high index of suspicion. Risk factors associated with child abuse include a single-parent household, especially if the father is the sole caretaker, children between 3 and 12 years of age, behavioral problems, large family size, and poverty. Of all of the risk factors, poverty is most strongly associated with abuse and neglect. Caretaker factors that have been associated with or have contributed to child abuse and neglect include younger age, poverty, personal history of abuse, large family size, unplanned pregnancies, illicit drug abuse and alcohol use, inadequate parenting skills, low–birth-weight babies, and social isolation.

The interview of a child who may be a victim of abuse requires a compassionate and sensitive approach, which assures the child that they are in no way at fault. The most successful interviews occur when the child is interviewed alone. The clinician should sit at the child’s eye level, explain the purpose of the interview, use open-ended questions when possible, and not display shock or horror at the child's descriptions. Depending on the outcome of the interview, the clinician has numerous options for treatment, including hospitalization, outpatient treatment and family therapy, contacting the state or city social welfare agencies, and, depending on the state legal statutes, contacting legal authorities. Medication may be considered to treat underlying psychopathology but not to treat reactive emotional states. If underlying mental illness is discovered, the use of child and adolescent psychiatrists is optimal to ensure appropriate longitudinal care by specialized providers.

Rape

Rape is a major psychiatric emergency. Due to social stigma, taboo, or fear of retaliation by the offender, the victim will often not immediately report rape or wish to report it at all. However, a victim who is not managed properly may develop enduring patterns of psychological and sexual dysfunction. Thus, the treatment of a patient for known or suspected rape is complicated. The accompanying fear of future violence and death at the hands of a known perpetrator and of contracting sexually transmitted diseases are also difficult acute stressors in the context of rape. For women, there is the additional fear of becoming pregnant. The management of the rape victim involves medical, mental health, and forensic interventions that may be initiated in the PES. However, the primary goal of the assessment for a victim of rape is to restore psychological safety by providing information, correcting misattributions, restoring and supporting effective coping, and ensuring social support. A respectful, compassionate, and sensitive approach is necessary to forge a therapeutic relationship with victims.

Helping the patient recognize that they are in the hospital and are safe can restore psychological safety. Because patients may be sensitive to power and control issues, a common respectful approach will help develop this sense of safety. Patients who react to the trauma with dissociative symptoms may appear calm and unaffected by their ordeal yet have a high risk of developing post-traumatic stress disorder. Once a sense of safety is achieved, it is essential to provide information to the rape victim about their current medical status. Addressing fears of future health problems, including pregnancy, sexually transmitted diseases, and HIV infection, is paramount and will aid in reducing anxiety. Information must also be given about the legal system, the responsibility of the police, and the need for evidence collection. Clinicians should acknowledge the patient's fears of being touched and undergoing vaginal and rectal examinations and discuss the potential for these examinations to trigger flashbacks or intrusive thoughts of the rape or sexual assault. The presence of a trusted family member or friend may help the patient to cope with the examination. Victims of rape or sexual assault commonly blame themselves for what happened and may have persistent fears of being out in public and perceiving being out alone as dangerous. Helping the patient to recognize the potentially impairing nature of sexual trauma may decrease anxiety, guilt, and anger. Education about common reactions to rape, such as irritability, hypervigilance, and problems with sexual intimacy, is important to share with the patient once a therapeutic relationship has been established. Discussing sleep problems, intrusive thoughts, nightmares, avoidance, and numbing may also aid in restoring psychological stability and allay fears that they are crazy or have lost control. Ongoing counseling and psychotherapy should be recommended for all sexual abuse victims. Facility-specific services may be available, and it is important for the PES provider to be familiar with all potential resources to help patients longitudinally with sexual trauma.

PSYCHOSOCIAL ISSUES IN THE PES

Homelessness

Homelessness has become an increasingly difficult social stressor for many in the United States. Additionally, many urban homeless are chronically mentally ill. PES and social services personnel are frequently called on to provide aid to homeless persons after medical care has been rendered in the emergency department. Although the typical “street person” is difficult to characterize, these patient populations may suffer from one or more psychiatric illnesses, including schizophrenia, substance use disorders, neurocognitive disorders, and personality disorders. Because of the extensive biopsychosocial pathology seen in the homeless, numerous resources must be devoted to their evaluation and treatment. Recent novel care programs in various jurisdictions have aided in providing longitudinal community-based care to mitigate the potential use of scarce hospital-based resources. Awareness of these services is crucial for a PES physician.

It is often difficult to engage homeless populations. Homeless individuals, especially those with severe mental illness, may not trust or actively seek to avoid contact with the medical or social service systems. If the clinician cannot develop a rapport with the patient, it is not generally an effect of poor training. In these instances, other PES staff members may be utilized to try to establish a relationship with the patient in order to assess their needs. The clinician’s initial task is to determine what the patient requires (shelter, medical attention, or antipsychotic medication), especially when self-presenting. However, police or other agencies may bring in homeless patients for care, and the chief complaint may be obscure or often nonexistent.

Helping the patient to clean up and offering food and drink may be the best ways to engage the patient and learn about their psychiatric and social needs. As a general rule, all homeless persons should be at minimum screened for any medical illness. Some people may have unidentified illnesses due to head trauma or nutritional deficiencies. If possible, these patients should be hospitalized for further evaluation.

When mental health care is rendered, and medication is utilized, it is less likely the patient will continue to take medication following discharge from the PES unless a more structured living situation can be arranged. The mentally ill homeless person is commonly resistant to a meaningful change in lifestyle unless the patient is actively and accurately describing their needs. A caring, nonthreatening clinician may be able to place the patient into the mental health care system to meet needs. Linking a patient to a drop-in center or day hospital may be worthwhile, if possible, and many urban facilities may offer specialized programs in providing assistance. Working with the patient and targeting specific identified goals can potentially develop a positive therapeutic alliance and allow a homeless patient to receive necessary services longitudinally.

High Utilizers of Service

Patients may utilize psychiatric emergency care services for inappropriate secondary gain. Beginning with symptoms, patients may complain of urgent needs that are ripe with unrealistic expectations and wishes for immediate relief. Patients may relate in a generally negativistic manner and are often regarded by staff as demanding and/or manipulative. These patients' desires and demands are usually quite different from their clinical needs on assessment and can create intense frustration on the part of both the patient and the treatment team. High utilizers of resources have a paucity of supportive interpersonal networks as they tend to interact within their social environments in much the same way they do in the PES, leaving available resources exhausted and depleted.

Managing these patients requires a flexible approach. The patient's autonomy should be supported, but each emergency visit should be viewed as an opportunity to encourage both the patient's engagement and participation in treatment. In addition, every effort should be made to enhance the patient's understanding of what the PES provides and the limitations of service. Some of these patients may ultimately develop a positive transference to the PES that will help them maintain a relatively consistent level of functioning. While the optimal goal would be to have them engaged in an outpatient program, their relationship with the PES may keep some patients under the protective mantle of a single health care facility and ensure appropriate and safe longitudinal care. A partial hospital or intensive outpatient program is an effective, relatively stabilizing disposition and should be encouraged by PES staff. However, the availability of such resources may be limited, and the care in many instances rests on PES providers.

Perhaps the most difficult aspect in managing any patient whose expectations are unmet is the clinician's own discomfort. Because these patients may be manipulative, uncooperative, and ambivalently help-rejecting, the clinician must be aware that the patient’s style is likely a reflection of interpersonal deficits. The deficit is usually the result of early-life developmental exposures having little to do with the clinician. Such an understanding should help the provider resist being drawn into the transference struggle and instead help them to respond empathically despite provocation. In addition, a supportive approach combined with operant shaping strategies that set limits on demanding or inappropriate behavior can be effective when interacting with this population.

LEGAL ISSUES IN EMERGENCY PSYCHIATRY

Relevant legal issues in the PES include confidentiality and release of information, duty to warn and/or protect, informed consent and right to refuse treatment, competency, and involuntary commitment. An overarching concern for all of these issues is the need for adequate documentation in the medical record. The chart is a legal document; therefore, a court can subpoena a medical record. This is the most important piece of evidence in most malpractice cases. Charting should include an accurate assessment of the patient and all procedures, interventions, treatments, and plans for continued care. All chart entries should be accurately timed, dated, written legibly, and signed. Fraudulent documentation or “doctoring” the chart is a very serious legal offense.

Confidentiality

Since the time of Hippocrates, information shared between a doctor and a patient has been considered confidential. Accordingly, the American Psychiatric Association has established ethical codes addressing confidentiality. A physician is ethically and legally obligated to maintain the privacy of all information shared by the patient in a clinical doctor-patient setting. This “privileged” information belongs to the patient and can be shared with others only if the patient gives their permission to release the information or if the court system demands it. There are, however, several clinical situations in which confidentiality is waived. These include situations of child/dependent/elder abuse or neglect, competency hearings, court-ordered psychiatric assessments, danger to self, danger to others, expressed intent to commit a crime, and communication with other treatment providers involved in the care of the patient.

Duty to Warn

For the PES physician, the duty to warn is of particular significance. If a physician determines that there is a substantial risk of harm to another person if no intervention is made, then they are legally obligated to warn the person at risk. It is imperative for the psychiatrist to determine whether the information provided is a credible threat. Vague statements about thoughts of harm should not be taken at face value as a legitimate threat. However, after further questioning and clarification, if a threat with intent to harm others is elucidated, it is the responsibility of the clinician to make an effort to warn the threatened individual. Obtaining information from family, friends, and other treatment providers will often help clarify the credibility of a threat as well. When available, medical records of prior psychiatric treatments or the patient's managed care company's clinical database can also provide pertinent clinical information related to historical threats. The medical record should accurately detail the patient's statements and all follow-up information obtained, including the physician's opinion on the credibility, or lack thereof, of any threat. The record should justify the clinical action taken by the doctor and should include any specific plans expressed by the patient, capacity to carry out the plan, prior history of violent behavior or aggression, access to weapons, current mental status examination, and a current psychiatric differential diagnosis. The clinician should also document any consultation with colleagues, supervisors, administrators, or other authorities. If the threat is deemed credible or the clinician is unsure, then the patient should be admitted to a psychiatric unit for further evaluation as imminent danger to others is a viable justification for inpatient care.

Evaluation of Capacity and Competency

Clinical capacity is broadly defined as the ability of a patient to make informed decisions and consent to receive medical treatment as assessed by any physician. Similarly, competency is a patient's legal capacity to engage in the decision-making process regarding their medical care. A patient's incompetence, however, may be determined only by a judge who is often supplemented with a physician's capacity evaluation. By law, all adults are presumed competent unless otherwise determined by a court of law. In a psychiatric emergency, waiting for a judgment from a court is rarely practical.

In assessing a patient's decision-making capacity, a patient must be able to:

1. understand the basic information necessary to make a decision;
2. understand alternative treatment strategies available;
3. appreciate the consequences of treatment and of no treatment; and
4. integrate all of the relevant information to make a consistent decision on their own behalf.

Aside from these four criteria, the standard for capacity may also vary on the risks to benefit ratio of the proposed treatment. A high benefit with a low-risk procedure may carry a higher standard of capacity to refuse and a lower standard of capacity to accept the treatment.

Competency is situation-specific, but a patient may also be assessed for global incompetence. A patient may be competent to consent to treatment and still suffer from major mental illness. A patient who is delirious, floridly psychotic, demented, or who is acutely intoxicated or withdrawing from drugs is likely not competent to make decisions concerning their treatment. In these cases, a surrogate decision-maker, usually a family member acting in the best interest of the patient, may decide on the patient's behalf. Additionally, facility-specific risk management services may be available for consultation to ensure practice within institutional guidelines.

Competency is a medicolegal determination and should not be used to coerce a patient into treatment, even if the physician feels it is in the best interest of the patient. If found competent for decision-making capacity, a patient has the legal right to determine their own treatment independent of the physician's recommendations for care. Accurate and timely documentation, including a mental status examination in the medical record, is mandatory in mental health cases regarding patient competency, especially when the patient refuses the recommended treatment.

Informed Consent

Physicians often equate informed consent with a person giving permission to participate in a clinical research trial. The concept of informed consent, however, is relevant to all clinical settings, especially the PES. The doctrine of informed consent mandates that a patient must voluntarily give permission to receive or deny medical or psychiatric treatment. However, in order to provide consent, the patient must be deemed competent to make medical decisions on their own behalf (see "Evaluation of Capacity and Competency" previously) and be free of any coercion. A key component of informed consent is that adequate information must be provided to the patient or surrogate decision-maker. In some states, the amount of information needed for an average individual to make a decision on a proposed treatment is required (materiality standard). However, in other states, the amount of information required depends on the needs of the specific individual. Typically, the information for informed consent includes the patient's diagnosis, nature of the treatment proposed, treatment risks, benefits, and alternatives, if available, likely outcome with or without the proposed treatment, and the unpredictability of some treatment outcomes.

The medical record should document informed consent granted for treatment given or why treatment was necessary despite a patient's refusal. In some instances with patients who have chronic psychiatric conditions, a court-appointed delegate might be designated to provide consent on the patient's behalf. This is a legal determination and not one made by a psychiatrist. In children, parents generally have the right to treatment decisions except in the case of emancipated minors. The age and nature of treatment where parental authorization is required vary from state to state.

There are four circumstances in which a psychiatrist may treat a patient without first obtaining informed consent from the patient.

1. The patient is incompetent (a legal determination).
2. Emergency care is necessary to prevent serious risks to the patient or others.
3. There is a therapeutic waiver, where a competent patient designates the physician to make treatment choices for them.
4. Therapeutic privilege allows the physician to withhold information from the patient if it poses a threat to the patient or would impair their ability to make a competent decision.

Whenever informed consent is not obtained, precise and thorough documentation of the clinical circumstances should be placed in the medical record in the form of a legible, dated, and timed progress note.

Involuntary Commitment

Psychiatrists are in a unique position of having an ethical and legal responsibility to care for patients against their will if they meet specific criteria. Depriving an individual of their civil liberties is often a difficult decision, as patients may not have the insight into their actions and repercussions of behaviors due to an underlying mental illness. Although each state has its own involuntary commitment procedures, virtually all include criteria related to dangerousness to self, dangerousness to others, or an inability to care for oneself as a result of a mental illness. A patient may demonstrate symptoms consistent with severe mental illness, but if they are not deemed at imminent risk of harm to self or others and able to utilize resources for self-care, then they generally cannot be placed into involuntary psychiatric care. Family members, friends, police, or physicians can petition for the patient to be evaluated and allow a certified mental health professional to determine the safety of the patient’s behavior in most jurisdictions. If deemed dangerous, the patient may then be hospitalized for a few days to a week, depending on the state mental health laws. Within a designated time period, the patient must have their specific case reviewed by a mental health court to determine whether a release is appropriate or detainment for longer treatment is warranted. It is important to emphasize that mental illness is necessary but not the only criteria sufficient for involuntary psychiatric hospitalization. Behaviors consistent with concern for safety or inability to care for self must be provided to ensure appropriate detainment. Homelessness, for example, does not justify forced hospitalization.

CONCLUSION

This section has reviewed some of the unique challenges that present to clinicians working within PES. Although emergency psychiatry is challenging for clinicians, it is also very rewarding and a critical component of the mental health delivery system. As services expand, so do programs for advanced training in the specialty of emergency psychiatry. As is the case in all emergency department settings, the need for a focused, thoughtful initial evaluation is the key to clinical success. Unlike an outpatient service, time pressures are essential to effective and efficient health care delivery. The PES clinician must be able to complete a biopsychosocial assessment, formulate a differential diagnosis, and make treatment or placement decisions, often in a few hours or less. The primary focus is on patient and staff safety with concomitant initiation of appropriate treatment. The importance of concise, accurate charting which accurately reflects clinical evidence is critical. It is unlikely that the PES clinician will be following the patient long-term, and the medical record will be the enduring material supporting placement decisions and must contain the supporting clinical information. Accurate dating and timing of all notes are particularly important in the PES, as this may impact legal issues (such as involuntary commitment) and clinical issues (such as medication dosing). Maintaining a high index of suspicion for medical illness will minimize the potential mortality and morbidity of PES patients. Managing violence is of particular importance to ensure the safety of patients and staff. Thus, all patients presenting to the PES should be considered potentially violent. This “universal precaution” mentality will help mitigate dangerous situations from developing. PES staff must be familiar with the array of social resources available in the community, including shelters, drug and alcohol treatment services, and community mental health services. These are vital to running an effective PES. PES staff must also be familiar with the legal aspects of providing the standard of care in a psychiatric emergency setting.

FURTHER READINGS

Allen MH, Currier GW, Carpenter D, Ross RW, Docherty JP. Expert Consensus Panel for Behavioral Emergencies 2005. The expert consensus guideline series. Treatment of behavioral emergencies 2005. ***J Psychiatr Pract***. 2005;11(Suppl 1):5–108.

American Psychiatric Association. ***Diagnostic and Statistical Manual of Mental Disorders***. 5th ed. American Psychiatric Association; 2013.

Buckley PF, Noffsinger SG, Smith DA, Hrouda DR, Knoll JL IV. Treatment of the psychotic patient who is violent. ***Psychiatr Clin North Am***. 2003;26:231–272.

Cameron CL. Hospital based psychiatric emergency services: a guide to service delivery and clinical best practices. ***Int J Emerg Ment Health***. 2006;8:23–33.

Crowley TJ, Sakai JT. Inhalants. In: Galanter M, Kleber HD, eds. ***Textbook of Substance Abuse Treatment***. 3rd ed. American Psychiatric Publishing; 2004.

Currier GW. Impact of the patient protection and affordable care act on emergency psychiatry. ***J Nerv Ment Dis***. 2015;203:901–905.

Currier GW, Allen MH. Emergency psychiatry: physical and chemical restraint in the psychiatric emergency service. ***Psychiatr Serv***. 2000;51:717–719.

Currier GW, Medori R. Orally versus intramuscularly administered antipsychotic drugs in psychiatric emergencies. ***J Psychiatr Pract***. 2006;12:30–40.

D’Orio BM, Purselle D, Stevens D, Garlow SJ. Reduction of episodes of seclusion and restraint in a psychiatric emergency service. ***Psychiatr Serv***. 2004;55:581–583.

Dubin WR, Weiss KJ. ***Handbook of Psychiatric Emergencies***. Homestead Schools; 1997.

Duke AA, Smith KMZ, Oberleitner LMS, Westphal A, McKee SA. Alcohol, drugs, and violence: a meta-meta-analysis. ***Psychol Violence***. 2018;8(2):238–249.

Hermanns-Clausen M, Kneisel S, Szazbo B, Auwarter V. Acute toxicity due to the confirmed consumption of synthetic cannabinoids: clinical and laboratory findings. ***Addiction***. 2013;108(3):534–544.

Herring AA, Perrone J, Nelson LS. Managing opioid withdrawal in the emergency department with buprenorphine. ***Ann Emerg Med***. 2019;73(5):481–487.

Hillard R, Zitek B. ***Emergency Psychiatry***. McGraw-Hill Companies; 2004.

Kahn MW. Emergency psychiatry: tools of engagement: avoiding pitfalls in collaborating with patients. ***Psychiatr Serv***. 2001;52:1571–1572.

Kosten TR, O'Connor PG. Management of drug and alcohol withdrawal. ***N Engl J Med***. 2003;348:1786–1795.

Logan JE, Haileyesus T, Ertl A, Rostad WL, Herbst JH. Nonfatal assaults and homicides among adults aged ≥60 years – United States, 2002–2016. external icon. ***MMWR Morb Mortal Wkly Rep***. 2019;68(13):297–302.

Osterman JE, Barbiaz J, Johnson P. Emergency psychiatry: emergency interventions for rape victims. ***Psychiatr Serv***. 2001;52:733–734.

Petit JR. ***Handbook of Emergency Psychiatry***. Lippincott Williams & Wilkins; 2004.

Petit JR. Management of the acutely violent patient. ***Psychiatr Clin North Am***. 2005;28:701–711.

Rodriguez-Cabezas L, Clark C. Psychiatric emergencies in pregnancy and postpartum. ***Clin Obstet Gynecol***. 2018;61(3):615–627.

Rudolph MN, Hughes DH. Emergency psychiatry: emergency assessments of domestic violence, sexual dangerousness, and elder and child abuse. ***Psychiatr Serv***. 2001;52:281–282.

Simpson SA, Joesch JM, West II, Pasic J. Risk for physical restraint or seclusion in the psychiatric emergency service (PES). ***Gen Hosp Psychiatry***. 2014;36(1):113–118.

Slaby AE. Definitions and conceptual framework of psychiatric emergencies. ***Emerg Health Serv Rev***. 1985;3:9–24.

Stone AL, Becker LG, Huber AM, Catalano RF. Review of risk and protective factors of substance use and problem use in emerging adulthood. ***Addict Behav***. 2012;37(7):747–775.

Theriault KM, Rosenheck RA, Rhee TG. Increasing emergency department visits for mental health conditions in the United States. ***J Clin Psychiatry***. 2020;81(5):20m13241.

U.S. Department of Health and Human Services. Agency for Healthcare Research and Quality. Effective Health Care Program (2012, August). First- and Second-Generation Antipsychotics for Children: Comparative Effectiveness.

Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. ***Lancet***. 2013;382(9904):1575–1586.

Wilson MP, Pepper D, Currier GW, Holloman GH Jr, Feifel D. The psychopharmacology of agitation: consensus statement of the American association for emergency psychiatry project BETA Psychopharmacology workgroup. ***West J Emerg Med***. 2012;13(1):26–34.

Work Group on Psychiatric Evaluation, American Psychiatric Association Steering Committee on Practice Guidelines. Practice guideline for the psychiatric evaluation of adults. Second edition. American Psychiatric Association. ***Am J Psychiatry***. 2006;163(Suppl 6):3.

32

Psychotherapies

32.1 Psychotherapy Integration

Adam M. Brenner, M.D., and Amy S. Brenner, M.S.W., L.C.S.W.

The term eclecticism has long been used to identify the practice of blending ideas or techniques from various schools of psychological thought in order to provide a more multifaceted, targeted treatment for patients. Although the field of clinical mental health is vast, and there exists many schools of thought on practice, technique, and aims, the one commonality all schools share is the desire to help patients improve their state. The ideas of eclecticism, currently known as integration, sought to further this idea by increasing the ability to be effective treaters, using the manner, time frame, and logistics already in place for the patient.

As the idea of treating the brain through talk therapy evolved, so did the idea that one's own particular talk therapy was superior to the next. Beginning with Freud, psychoanalytic thought held sway till it began to be challenged in the 1950s by behavioral therapy. From there, an increasing number of theoretical orientations proliferated, and partisan allegiance began to further separate clinicians into an array of "name brand" therapies. Trainees were indoctrinated by teachers expert in a particular school and graduated believing, "*The way I practice is better than the way you practice, and my patients get better faster, or change more deeply, or recover more completely.*"

While this rivalry among schools led to a proliferation of theories and therapies, there was little evidence that any one of them worked better than another. The factors that most therapies have in common, such as alliance, positive regard, empathy, and emotional relief, are the factors that researchers find contribute the most to positive outcomes for patients. Only 10% to 15% of outcome differences in the research literature are attributable to varying techniques. Further, in the promotion of their new brands of therapy, experts were prone to overstate what was unique, and many commonalities were given new jargon. As one of our mentors used to remark, many "advances" in psychotherapy were simply "old wine, new bottle." (K. Altshuler, personal communication, multiple.) As for the actual added value of the specific techniques, the sum of psychotherapy research and clinical experience made clear that no one theoretical school could help every patient.

In treating something as complex as the human brain and mind, psychotherapy practitioners found both limitations to their orientations and overstatements of the effectiveness of the same. Frustrated with what appeared to be clinical failures, some clinicians reexamined their professional identities and began to reach out to other therapeutic schools to see what was there. Consequently, some practitioners began to carefully choose techniques outside of the approaches they were originally trained in, blending these with their original school of thought. As one eclectic proponent, Arnold Lazarus, put it, while remaining grounded in one's original training, one should "borrow, purchase, pilfer, and import" other techniques from other schools of thought. A particularly significant milestone in the development of these efforts came when like-minded practitioners launched the Society for the Exploration of Psychotherapy Integration in 1983. In 1991, the society began publishing the Journal of Psychotherapy Integration.

Unfortunately, the monotheoretical communities often responded as if affronted by these “eclectics.” (One of the ironies to keep in mind about the current “pure forms” of therapy is that they were often created as a blend of theories and techniques from the major thinkers of the time.) It was suggested that deviation from the pure form of an established school of psychotherapy was evidence that the therapist was either poorly trained, indecisive, or disorganized.

When the authors trained in the 1990s, this had progressed to the point that the term “eclectic” became almost derogatory. As a result, the term “integration” slowly replaced it. And yet, eclecticism/integration is not syncretism. Syncretism implies a hodgepodge of quickly gathered methods without attention to a grounding in good practice or the evidence base. This can indeed imply poor training and focus. The goal of the integration of therapeutic approaches does not imply a blind acceptance that all theories and all techniques work. Theories and techniques must be grounded in an evidence base and operate within the established guidelines of responsible clinical practice.

Norcross has identified several barriers to therapy integration beyond partisan bias. The obstacles begin with training—most clinicians are schooled in one idea about psychotherapy, sometimes two. Being immersed in these ideas would, of course, lead new practitioners to a partiality toward that school of thought. In addition, psychotherapy as a field lacks a common language with which to describe, teach, and supervise integrative ideas—by contrast, the jargon of individual schools can be impressive. The use of such jargon can, of course, also serve competitive and narcissistic aims. Finally, there is the question of whether psychotherapies that are grounded in different theories of human nature may be ultimately philosophically incompatible.

Nonetheless, many clinicians have gone about integrating therapeutic techniques and ideas over the course of their careers, whether they choose to disclose this to their peers or not. Below we discuss the four major ideas regarding psychotherapy integration: the common factors approach, technical eclecticism, assimilative integration, and theoretical integration. Additionally, we will examine more recent proposals, including pluralism and differential psychotherapeutics. Each has a distinct way of viewing integration, all with the idea that flexibility and thought outside a monotheoretical view can lead to better patient outcomes.

Mrs. E is a 47-year-old woman. She is coming to therapy because she has been feeling "overwhelmed" for the past few months. She works as an accountant in her own business from home. She is married, and her husband also works full-time as an EMT. They have 3 children, 2 boys, 16 and 11, and a 13-year-old daughter. The boys have chronic medical conditions that lead to significant expenses.

Mrs. E reports that she initially started having trouble sleeping, with early waking around 3 AM and trouble returning to sleep. She has been taking over-the-counter Benadryl, but this has only led to feeling groggy upon early waking, but not helping with prolonging sleep. She then feels "draggy" during the day, has bouts of rumination about her inability to get enough done, and has some anxiety around the chaos she feels her household is in and also regarding various accounts she is dealing with at work. She reports this negativity is making her cry almost every day and feel "less resilient" than she has felt in the past. She is alarmed that she can't enjoy anything in her life and ashamed that she has had thoughts of wishing her life was over "so I can be at peace." She has gotten some response to antidepressant medication, and her psychiatrist suggests the addition of psychotherapy as the next step.

Mrs. E's business has taken off recently, which helps with the financial strain of healthcare expenses but causes a new set of problems. She was always able to be a very hands-on mother to her children but now has found she has less time available for all her tasks. She says, "I feel disgusted with my inefficiency and sloppiness. Taking care of the kids isn't rocket science, but I have kept dropping the ball and letting things fall through the cracks."

Childrearing has been very rewarding, and she has some sadness that as the children have aged, they have needed her less emotionally. She reports they are doing well and feels much of her concerns are her own feelings about the change in her role, less about any specific complaints from the children.

Mrs. E reports no substance use other than an occasional glass of wine with meals when she and her husband go out to eat. She is neatly dressed, with a normal range of affect that is most often sad and anxious. She is well-spoken and has some insight into her ongoing issues.

Her husband works as an EMT and has rotational shifts throughout the week. So, though he is very helpful, this help is not regular or reliable. He is troubled by his work, with increased pressure and secondary trauma from the scenes he is called to. His shift changes make it difficult to be together, but they both work hard to make that happen. His work has provided financial stability for them as she has built her business, and he has been looking forward to cutting back. Mrs. E feels that their relationship has suffered as her work has become greater, and she has wanted more help at home. She feels he is disappointed in her, but she has avoided discussing this. She is angry that he seems more distant and withdrawn.

She reports being very religious and engaged in her church. She has been on a number of mission trips to Mexico and Central America over the years, and these have been rewarding. However, in the last few years, she has experienced a crisis in her faith and the teachings of her church, especially as her daughter has matured. The culture of the church and its teachings have not aligned with how she wants her daughter to perceive herself, and this has been difficult.

She was raised in a very religious household, the eldest of three sisters. Their father was very exacting, expecting “perfection,” as she reports. He was never satisfied with their best, always feeling more could be done to be better, and his tone was often angry, sarcastic, and demeaning. Mrs. E’s mother never chastised her particularly but never spoke up to the father either. She wonders if her mother was intimidated or depressed herself. One of her younger sisters is completely estranged from the family, and the other is living with the parents following years of ups and downs in relationships and an ongoing alcohol use problem. Mrs. E is “a dutiful daughter” but tries to keep her distance from her parents and limits interactions with them. She is especially careful to keep their influence on her children to a minimum.

She reports feeling similarly in college. She was overwhelmed with doing well and getting perfect grades. At times she found the work too much and would resort to “disordered eating” to make herself feel better. When asked about this, she reports that she would often forgo eating to “have some control” over her life–deciding when she had done enough work to be “rewarded” with food. One of her roommates reported this to a college resident assistant, and she was seen by the college counseling service for help with this. She reports that the counseling was helpful in some aspects but finds that this type of thinking and behavior about food has returned. One of the drivers for appearing in treatment is to minimize such behavior in the future.

COMMON FACTORS APPROACH TO INTEGRATION

One of the most well-established findings in psychotherapy research is the importance of the “common factors.” This refers to those elements which the whole variety of therapies share. Yet, we believe they are still underemphasized in the scholarly literature and in training programs.

Why this relative lack of enthusiasm? Perhaps the reason lies in the matter of prior investment on the part of the therapist. Learning the theory, formulations, and interventions of any specific therapy requires hundreds of hours of study, practice, and supervision. After all that work, therapists can be forgiven for wanting to believe that those specific techniques are at the center of the treatment’s benefit.

Brown et al. argue that therapists should pay much more attention to common factors and that this is an ethical requirement. They begin with the idea that ethics guidelines call for the use of evidence-based therapy, but if the evidence argues that it is common factors that are most important, perhaps we should be talking about evidence-supported therapies instead. Instead of assessing whether there is an evidence base for the specific therapy applied to the patient's specific problem, the ethical practitioner can fulfill their duty by practicing any reasonable approach with a strong emphasis on maximizing the common factors.

What exactly are the common factors? There is no universally accepted answer to this question. Different authors have proposed different lists, and no one can claim that theirs is definitive.

For example, Frank's "Persuasion and Healing" did much to draw the field's attention to common factors. Frank's list included: an emotionally charged confiding relationship; hope; placebo effects; healing setting; mutually accepted belief system; therapeutic ritual; warm and inspiring therapist; catharsis and practice of new behaviors; therapeutic suggestions; and interpersonal healing. More recently, Feinstein et al. argue for the following: (1) Patient characteristics; (2) Hawthorne effect; (3) Hope; (4) The therapeutic alliance; (5) Therapist characteristics and behaviors; (6) Impact of external life events.

However, rather than attempt to exhaustively review the many proposed collations, we will instead focus on those common factors which, in our reading, find the greatest consensus in the literature. These are (1) Empathy; (2) Expectations or hope; and (3) Alliance.

Empathy is a complex construct, and much happens when a patient experiences true empathy from the therapist. There are cognitive, affective, and interpersonal aspects. The therapist actively listens and imagines what the patient is feeling, based on all that the therapist has learned about the human experience from her own personal history, from psychological theory, from narrative art and entertainment, and from prior patients. Along with this cognitive work, the therapist allows herself to resonate affectively with the patient's feelings. She may be moved directly by the patient's expression of emotion or may have associated it with her own emotional experiences. She creates space within herself to feel with the patient. The final piece of the picture is the therapist's attitude of compassion and acceptance of the patient's feelings and the effort to communicate this to the patient with tact and kindness. As the reader can see, calling something a “common factor” does not imply that it is automatic or lacks effort and skill.

Empathy is so powerful because we are social animals with a powerful biologic system of pair-bond attachment. Human beings are not designed to be alone with their distress. A secure attachment is an absolutely crucial element in child development. And loneliness in adults is a toxic stress that contributes to much psychiatric morbidity.

Experiencing empathy creates at least two very powerful therapeutic effects. First, the patient feels that they matter. This is the root of the Hawthorne effect—the gift of sustained attention. Second, many patients come to therapy believing that their experience is unacceptable. They often feel shame in addition to whatever the primary problem is. Empathy is a powerful antidote to shame. Knowing that the listener understands my experience and doesn’t recoil from me is innately therapeutic. This is all-encompassed in what Wampold et al. refer to as the “real relationship.” Their review of the literature suggests a Cohen-d effect size of 0.63 for empathy, in the medium range.

Once an empathic connection has been made, therapists of all stripes convey to the patient that they have some manner of making sense of the problem and a treatment approach that they believe will help. The impact of this step is not limited to psychotherapeutic interventions, but extends across all of medicine. It is the root of the placebo effect and is both widespread and powerful. Having brought their distress to an expert, the expert’s expression of confidence can feel transformative. (Contemporary anthropologists argue that the development of a penchant for identifying and learning from experts was a pivotal step in the move from biologic to cultural evolution.) What had been a source of despair now pivots to an expectation of healing and a feeling of hope.

This leads us directly to the therapeutic alliance, perhaps the best studied of the common factors. The therapeutic alliance is the condition created by a patient and a therapist who agree on goals, pursue tasks together toward these ends, and work in the context of a warm interpersonal bond. The presence of a strong therapeutic alliance is one of the best predictors of success in therapy.

From the perspective of any one psychotherapeutic approach, this seems obvious. To a psychoanalyst, of course, the therapy will depend on the patient and therapist pursuing the goals and tasks of an analytic therapy. Perhaps less intuitive is that it doesn't seem to matter what the specific goals and tasks are. It may be to explore the transference, practice mindfulness in the face of painful feeling, correct maladaptive cognitions, or any number of other possibilities. What matters is the experience of working together.

We have all become more aware of the problem of burnout in our helping professions. While the degree of stress is an important element, the literature also demonstrates the power of teamwork and alignment of the mission. A physician who feels they are part of a team rowing in tandem, aimed at a shared and worthy goal, is less likely to become burned out by stress. This is similar to the effect of the therapeutic alliance. Feeling that one is contributing to a joint project with members you trust is a powerful antidote to feelings of demoralization and worthlessness.

There is an irony in how the positive expectations and productive alliance rest on the pair's confidence in the specific therapy they are commencing together. We are unlikely to generate the full effect of the common factors if we were to meet with a patient without the scaffolding of a specific theory and its accompanying technical maneuvers.

What, then, does a common factors approach to integration look like when put into practice? Perhaps like supportive psychotherapy. We might suggest that much of what is called supportive therapy is actually psychotherapy integration with an emphasis on common factors. Welton and Crocker, in their recent overview of supportive therapy, note that the "common elements become the primary target of therapy" and that supportive therapy then overlaps and borrows specific techniques from other schools.

Feedback informed treatment is another approach that attempts to put an emphasis on common factors into concrete practice. Its advocates argue that regular formal measure and review of the therapeutic alliance should be built into the treatment. The aim is to create a "culture of feedback" with a particular emphasis on the patient's view of the alliance, which is generally a better predictor of success than the therapist's.

The therapist prioritizes making sure that Mrs. E feels empathically listened to. The therapist comments, "I can hear how overwhelmed you feel. You work so hard but you're still not sure you're doing enough or doing it well enough. And then you feel ashamed at how depressed you've become." As a result, Mrs. E feels understood, accepted, and no longer alone with her problem. The therapist says, "I believe therapy can help you," which is both an evidence-based prognosis and also meant to foster hope. The therapist then focuses on what Mrs. E's goals are and what to work on first. Together they decide to focus on maintaining healthy eating habits, identifying activities that relieve stress, and talking about what kinds of support she most needs. After just a month in therapy, Mrs. E reports improvement in all domains.

TECHNICAL ECLECTICISM

Technical eclecticism is the form of integration that is probably the most used, and least discussed. It flows from the techniques, not the theories. Technical eclecticism uses data, both from the literature and from clinical practice, to choose the best therapeutic options for a patient, options that have been shown to work in the past. Practitioners don't necessarily champion the particular theories behind each technique, but rather are able to see beyond the theory and value the particular techniques from various theoretical sources. Practitioners also tend to use techniques that already exist, as opposed to creating grand new theories and techniques. This practice is grounded in what has worked in the past and translating that into what hopefully will work now.

Wakefield points out that becoming well-versed in many theoretical frameworks is not realistic. The difficulty in cost and time of training argue against such a quest. Nor is referring a patient to multiple practitioners a realistic way of practice. Psychotherapies occur within the limitations of available providers who are affordable or covered by insurance, not to mention the difficulty in forming multiple new alliances for a patient-changing practitioners.

As a result, many practitioners have used some form of technical eclecticism in their practice. These are often clinicians who have run up against the limits of what a monotheoretical approach has to offer. Interestingly, they tend not to be found in academic settings, where there is often relatively little support for ideas regarding integration. Each eclectic clinician has found on his/her own that there are great commonalities in practice techniques across the different schools, and these interventions work the same, no matter what vocabulary one uses. Unfortunately, eclectic clinicians often move under the radar, lest they offend the assumptions of theoretical purity. One of the authors recalls being asked about her therapeutic orientation at a job interview. When she declared an eclectic mindset, she experienced a disdainful silence and awkward end to the meeting.

There have been several noteworthy attempts to systematize what eclectic therapists do spontaneously. The work of Lazarus et al. on multimodal therapy is a prominent example. They argue that using empirically supported techniques will be more useful in treating patients than trying to use theories of psychopathology to lead the treatment. Multimodal therapy begins with an assessment of patients on a wide spectrum of function—behavior, affect, sensory, imagery, cognition, interpersonal, and drug/biologic. This is called a modality profile and represented by the acronym BASIC ID. A clinician working in multimodal therapy appreciates the common factors, such as the therapy relationship, as a valuable driver of change. Consistent with the general flexibility of the approach, the therapist’s way of relating to the patient should be based on what the patient responds best to—the therapist is “trying to be an authentic chameleon.” Regarding therapeutic technique, the MMT clinician will be free to draw from any method that has demonstrated efficacy and is appropriate to the problem at hand. However, proponents of the approach do not generally include psychodynamic interventions related to transference within their scope. While multimodal therapy draws on techniques that have demonstrated efficacy independently, there is little rigorous controlled research to support the overall approach.

Systematic treatment selection (STS) also de-emphasizes theories of etiology and tries to provide an alternative approach to structuring eclectic intervention. STS is built around a set of principles to help guide the choice of technique. For example, careful attention should be paid to the alliance, and the need to repair ruptures with empathy and flexibility. Additionally, the characteristics of the patient should guide technique. They reasonably suggest that greater impairment of the patient should be met with greater intensity of treatment. Patients who are highly resistant to therapy should be met with less directive approaches, while those who are less resistant may find comfort in being directed. Similarly, focusing on behavioral change and symptom reduction is recommended for patients who externalize, while insight and relational learning may be preferred for those who internalize. A small but intriguing research base suggests that matching characteristics of the patient to the technique may enhance outcomes.

Mrs. E's therapist prioritizes forming an alliance that engages the patient in shared work. The therapist observes that Mrs. E tends to internalize her difficulties and is very open to exploring new ways to think about her experience and relationships. This exploration makes clear that Mrs. E is grieving several losses. She conveys her sadness about her withdrawal from the church that she was raised in. She expresses her bitterness about the need to keep her children from their grandparents, lest her father's toxicity harms them as well. Her therapist moves from an exploratory to a more problem-solving stance. They discuss how to set effective limits with her father when the family is together, and practice this in role play. Appreciating Mrs. E's longing for spiritual support, her therapist helped her examine the anxieties that prevented her from trying different churches and denominations. As these were connected to a fear that she would be rejected by the pastor and congregation, Mrs. E and her therapist began to use an approach of gradual exposure and guided relaxation to overcome these inhibitions. Mrs. E was appreciative that her therapist was willing to try different approaches to her different problems, and this deepening trust allowed them to return to an exploratory approach to the long-standing issues around food and body image.

ASSIMILATIVE APPROACH TO INTEGRATION

In this mode, the clinician works from a theoretical and technical foundation of one distinct and established school of therapy and adds elements from other therapies onto that framework. It should not be surprising that this would be a common mode of integration. Generally, clinicians are trained in programs that teach distinct schools of therapy. The practice of integration does not usually begin with a therapist who is an empty receptacle acquiring a variety of eclectic techniques. Nor is it likely that most practicing integrationist therapists begin with construction of an integrated theoretical model.

In a study of experts in psychotherapy integration, three-quarters reported that though they drew on multiple orientations or schools of therapy, they considered one to be primary. As Wachtel points out, “All proponents of integration start from somewhere, and how they look at the end is in good part shaped by where they started.” This is consistent with an assimilative approach.

Therapists often become open to integration when they find themselves up against the limits of their chosen therapeutic style. That experience of “limits” can take a number of different forms.

First, there are those times when a new problem surfaces that the current form of therapy is not as well suited for. For example, a patient may be deep in a psychodynamic therapy to explore difficulties with intimate relationships when they reveal that they are also suffering from an obsessive-compulsive disorder. The therapist feels the dynamic therapy has been valuable but believes that CBT is a better, more evidence-based approach to severe OCD symptoms. Similarly, the therapist may learn of a substance use disorder that might be best managed with motivational interviewing. Rather than refer the patient to another therapist, an assimilative approach would be to add some CBT-based management of compulsions or motivational interviewing for addiction into the current treatment.

Another kind of limit, however, is the experience of simply being stuck. Consider again a clinician working from a psychodynamic foundation. The patient and the therapist understand their shared work to be an exploration of the patient’s mind and an exploration of the therapist–patient relationship, in the service of developing insight into unconscious conflicts, and implicit relational patterns. There are several ways that exploratory work that seemed to be progressing can become stuck. One way is when the exploratory process seems to still be moving along well, but deepening insight is not leading to symptomatic or behavioral change in the patient’s life. Another way is when the exploration itself has become stalled. Perhaps the transference is too volatile to productively discuss, or perhaps the patient’s symptoms (such as anxiety) become too painful to tolerate. The assimilative therapist may decide that these are times when a more directive or symptom-focused approach is necessary to break the impasse.

On the flip side, let's imagine a CBT therapist who finds that the patient is unable to adhere to the planned homework because any such expectation from the therapist evokes intense feelings of resentment and defiance. The therapist may suggest a brief exploration of the relational origin of this paralyzing reaction as a means to clear the space for a return to a working alliance focused on exposure-based tasks.

This kind of work is clearly not easy to do—it requires all the usual skills of the dynamic or CBT therapist and also the judgment of when to switch tracks. This is the critical issue. For example, our psychodynamic clinician could err in either direction. They might incorporate a more directive technique unnecessarily and enable the patient's avoidance of a crucial, but painful, emotional experience. Or they might continue to explore the symptom's meaning while the patient requests a more direct approach and simply repeat and reinforce an experience of a rigid authority more interested in "following the rules" than being of actual help. A psychodynamic purist may suggest that in the latter case, the solution is more exploration of this experience, but when the patient drops out of treatment in frustration, exploration generally diminishes markedly.

In their writing, Stricker and Gold have attempted to define an assimilative psychodynamic psychotherapy (APP) and ground it in a three-tiered model of mental function ([1] Behaviors/relationships; [2] Conscious cognitions and emotions; [3] Unconscious conflicts and self/object representations). They retain the core dynamic emphasis on exploration and insight but incorporate techniques of exposure and cognitive learning as indicated. The therapy may begin with either exploratory or more directive methods, depending on what the patient prefers, which "tier" is most prominent in the presentation, and what kind of interventions the patient can tolerate. Their default mode will remain exploratory but they note that sometimes people "need to learn new skills and to unlearn maladaptive skills in order to change." They argue that more active work on symptoms can be synergistic with exploration; by providing relief from acute distress new areas of potential exploration become accessible.

Castonguay and colleagues have described an approach that uses CBT as its foundation but acknowledges its limitations. In their view, CBT remains too focused on the individual’s cognition, and doesn’t sufficiently emphasize interpersonal factors and the value of deepening emotional experience (contrasted with controlling it). As a result, they created a program where therapists did an hour of CBT followed by an hour of interpersonal emotional processing (IEP) therapy. It is worth noting that the IEP hour is a fairly rigid and directive version of dynamic work. The authors report preliminary data that didn’t show particular gains compared with CBT alone except in secondary analysis for patients with dismissive attachment style. It is hard to draw much inference from this work to a more organically integrated approach where the clinician utilizes the interpersonal/emotional focus within the framework of a CBT hour.

Proponents of assimilation can argue that they are providing patients with the “best of both worlds” the benefits of the core mode of therapy while addressing the gaps in its utility or the limits of its reach. Critics might suggest that the result will actually be a watered-down version of the foundational theory and that switching approaches when faced with an impasse prevents the clinician from learning how to deepen their work and turn such challenging periods into opportunities for therapeutic growth.

Mrs. E begins treatment with a psychodynamic psychotherapist. The focus is on the roots of her perfectionism and self-criticism about her role as a mother. They make connections to her father's harsh standards and also to her own sadness that her children are growing more emotionally independent. Together they wonder if this sadness is related to her determination to never have her children experience her as she did her own mother—distant and emotionally removed. Some of these themes of judgment and closeness come up in the therapeutic relationship and are used to deepen their understanding. After several months, Mrs. E feels much less depressed and overall is functioning very well.

However, her eating symptoms have not gotten any better. Insight into their meaning is not leading to change. Mrs. E's therapist suggests that they continue dynamic therapy, which has proven so productive, but that they devote some time each session to using a CBT approach to her eating. They map out how thoughts of inadequacy trigger ruminations about her weight. Mrs. E begins recording all her food-related thoughts, emotions, and behaviors. Overevaluation of appearance and weight is identified as a core problem and strategies for minimizing her dysfunctional eating behaviors are developed.

Mrs. E's eating symptoms improve, but she and her therapist find that directly addressing these issues also opens up new areas of exploration in her therapy. She begins to report a recurring dream about her body and her associations lead to anxieties about physical intimacy that have never surfaced before. The psychodynamic work deepens.

THEORETICAL INTEGRATION

Instead of focusing on what therapies have in common, or how to modify existing therapeutic techniques toward integration, some authors have proposed that we begin our efforts at integration with an underlying theory of psychotherapy. They have asked, can we construct a theoretical model that embraces the entire range of psychotherapy strategies? There is an obvious appeal to this approach that makes it something of a “holy grail” of therapy integration.

Integration that begins at the theoretical underpinnings—what is our theory of psychopathology and of healing/change—would have the potential to streamline education and training, focus research efforts, and provide consistent standards of practice. In this section, we will review some of the more prominent efforts.

Wachtel’s “Integrative Relational Therapy” in some ways straddles the categories of assimilative and theoretical integration. It is assimilative in that Wachtel was trained in psychoanalytic psychotherapy and builds upon the long traditions of psychodynamic formulation and technique—ego psychology, self-psychology, and relational psychoanalysis. Yet it is also a theoretical integration, because there is not just addition of techniques from outside the dynamic tradition, but also a reworking of the dynamic model termed “cyclical psychodynamics.”

As in traditional psychodynamics, the patient is seen as having internal conflict—wishes/longings that pull in opposing directions or provoke the internal experience of threat and anxiety. However, Wachtel’s emphasis is equally on behavior—how does the patient act in the world, or fail to act, that provokes responses from others that actually confirm the anxiety? Insight into internal conflict is not enough to undo these “vicious circles.” The therapist and patient must focus on learning new behaviors that carry significant anxiety. This means working with the patient on a process of gradual exposure and desensitization.

One of Wachtel’s most important contributions is the idea that “exposure” is not a technique limited to behavior and that it is equally applicable to disturbing thoughts. He argues that it is a core concept that helps unify behavioral and psychoanalytic approaches. The psychodynamic technique of clarification and interpretation also results in the patient learning that something feared will not be catastrophic. “Exposure” to disturbing ideas about the self and important relationships is uncomfortable, but if done gradually and with tact, not intolerable.

Eubanks and Goldfried similarly argue that there are principles of change that cut across different therapies. These extend beyond what are generally studied as “common factors,” as they begin with fostering hope and strengthening the alliance, but also include increasing insight, corrective emotional experience, and improvements in reality testing. As Wachtel reframes exposure for internal experience, the authors reframe insight to encompass action, arguing that patients are often unconscious of what is incompetent about their behavior and must move to “conscious incompetence” in order to change. The principle of corrective emotional experience is intriguing. It is the place where learning theory and relational psychodynamics can come together. It is also problematic, as the question of whether the therapist allows corrective emotional experience to occur spontaneously versus intentionally providing a prescribed experience is one that depends on the theory.

Prochaska and DiClemente’s Transtheoretical Model (TTM) is an attempt to create a comprehensive structure that could contain all therapy approaches. They propose to describe the processes of psychotherapy using four dimensions: processes of change, stages of change, pros and cons of change, and levels of change. Levels of change attempt to locate and define the problem by specifying whether the target is a symptom, a cognition, an intrapersonal conflict, an interpersonal conflict, or a family/systems conflict. Next, we can assess the readiness to change. The TTM outlines a continuum that progresses through precontemplation, contemplation, preparation, action, and maintenance. Describing the pros and cons of change captures the decision from the patient’s unique individual perspective. Finally, the processes of change define the range of techniques and interventions and intend to pair these to specific levels of readiness. These include consciousness raising, self-liberation, social liberation, counterconditioning, stimulus control, self-reevaluation, environmental reevaluation, contingency management, helping relationships, and dramatic relief/emotional arousal.

The model's creators have attempted to define where specific psychotherapies are most relevant and useful given the intersections of levels and stages of readiness, but this is not a strength of their work. The model's strength lies in its tested efficacy for specific maladaptive behaviors—most commonly involving substance use disorders and general health promotion (e.g., smoking cessation). In these circumstances, there is substantial evidence that matching change processes (interventions) to the stages of change (readiness) increases the effectiveness of the therapy.

Unfortunately, the paradigm was designed for the kinds of difficulties that are binary in nature—either the patient will quit smoking or they will not. It is much harder to apply the model to the typical complexity of mental health disorders, and not surprisingly, there is only very preliminary evidence of value.

Prochaska and DiClemente's levels of change are echoed in the "levels of meaning" that Wakefield and colleagues propose. They argue that psychotherapy integration is crucial not only because every therapy doesn't work for every individual but because different treatments target very different levels of meaning. The levels of meaning they propose are biology, neuroscience, behaviorism, cognitive, psychodynamic, family systems, and culture.

The first two levels are largely targeted by pharmacologic and neurostimulation interventions. Behaviorism focuses on difficulties in learning and reinforcement. Cognitive therapy aims at conscious beliefs that motivate action, while psychodynamics addresses meanings outside of consciousness that can only be activated associatively. Family systems therapies focus on the implicit rules and roles in our intimate relationships, and culture brings into focus the ways that social context shape identity and the presentation of symptoms. Embracing the idea that there are different levels of meaning seems to be an essential first step in creating a theory that can encompass all therapies. The authors, unfortunately, but understandably, are not able to provide similar clarity regarding the implications for technique—should we continue the pursuit of integrated technique or use different therapies in a "stepped care" model?

We do not think that any of these authors have fully succeeded in their quest for theoretical integration—the "grail" is not yet in our grasp. This does not mean that it is impossible, nor does it mean that the efforts have not been valuable. The journey of a quest can be transformative regardless of whether the final object is yet attained.

During their initial consultation, Mrs. E's therapist notes that she is very motivated to address her depression. They decide to begin by understanding the ways that Mrs. E's depression is currently being sustained and focus on her marital relationship. She describes that she has not directly discussed her need for more help at home because she is afraid of how her husband will react. Unfortunately, in the absence of communication, Mrs. E has nonetheless experienced her husband as disappointed. She wonders if her hurt and anger in response have contributed to the growing distance between them. Her therapist calls this a "vicious circle," which Mrs. E finds very helpful. Together they decide to continue to explore Mrs. E's feelings and their origins in her family while also taking small steps in tolerating the anxiety of talking directly with her husband about her own needs and disappointments. She notes that it would be nice if he felt the same responsibility for the household that she did, for example. Mrs. E and her therapist find that behavioral steps work synergistically with allowing herself more "exposure" to her own mixed feelings in her marriage.

OTHER RECENT APPROACHES TO INTEGRATION: PLURALISM, STEPPED CARE, AND DIFFERENTIAL PSYCHOTHERAPEUTICS

Cooper and McLeod have proposed an approach they call "pluralism," grounded in the observation that many very different kinds of therapy have proven helpful to different patients or the same patient at different times. They note that these different therapies can consist of very different models to explain the cause of psychological distress and the mechanisms of change.

Pluralism means accepting that these different models may all hold some truth, some piece of the complexity of human life, but that attempts to resolve them into one comprehensive theoretical model of pathology/cure have been unsuccessful. But further, Cooper and McLeod believe such a resolution is unnecessary. Instead, they advise us to move away from theory and into three “domains” that are more pragmatic: goals, tasks, and methods.

“Goals” refers to the big picture of what the patient and therapist are trying to accomplish through therapy. Goals can include “to gain more insight into myself ” or “to cope more successfully with stress at work.” Tasks are more discrete and focused. They are specific action items in the larger project of therapy, in the same way, that “sorting through the closets” and “taking out the trash” are steps toward the larger goal of making one’s home feel livable. A partial list of therapy tasks would include: talking openly; expressing/letting go of emotion; finding and analyzing information; exploring meaning; problem-solving; changing behavior; enhancing self-care.

Cooper and McLeod do note that these are conjectural, and there is a need to develop an empirically grounded “taxonomy of therapeutic tasks” through qualitative research.

Once the task is defined, an appropriate method is applied. The available methods are all the techniques of the various schools of therapy that the therapist is able to offer. Pluralism pragmatically accepts that no therapist can be expected to offer every possible method to pursue each task. Thus far, it may be very hard to distinguish pluralism from technical eclecticism. One especially strong emphasis of pluralism is that respect for the patient’s agenda and preferences are prioritized. As a result, the choice of each of the three domains is negotiated in collaboration with the patient. (This is called “metatherapeutic dialogue.”) This should occur at the beginning of therapy and again as tasks are completed, new ones are decided on, and methods are chosen.

We should note here that the term “pluralism” means something quite different in contemporary psychoanalytic thinking. While some essentials of psychoanalytic technique have been preserved—free association, interpretation of the unconscious, working within the transference—there are great differences among the proliferating psychoanalytic schools in their theories of psychopathology and cure. Theoretical models include foci on drives, defenses, internalized relationships (object relations), formation of the self, or intersubjective relationships. Instead of the atheoretical emphasis on tasks and methods of Cooper and McLeod, psychoanalytic pluralists tend to focus on which theoretical frame of reference to apply to best explain the clinical moment.

A very different approach to the different needs of patients at different clinical moments is taken by the advocates of “stepped care.” Markowitz and Milrod, for example, argue against eclectic treatment. They suggest it is confusing to patients and that therapist adherence to specific therapies has been shown to be useful. Instead, they would have us develop an evidence base for conducting sequential algorithms for psychotherapies, similar to those we have for psychopharmacology. They suspect that changes in psychotherapy type should occur more often and be driven by having time-limited treatments with outcome measures/monitoring.

Cabaniss and Holoshitz propose a process called Differential Psychotherapeutics to guide changes in therapy type. They begin from the perspective that no one kind of therapy fits all patients and that the strongest evidence available is for discrete, not integrated, therapies. The challenge is then to have a systematic approach to “prescribing” when and what to change. There are four steps to the Differential Psychotherapeutics cycle: learning, thinking, matching, and discussing. “Learning” covers all the usual components of an initial psychotherapy assessment—establishing the current problem, obtaining the patient’s life story, and identifying internal and external resources. “Thinking” means a preliminary jargon-free formulation—what is wrong, why is it happening, and what needs to change. Cabaniss and Holoshitz acknowledge that we cannot think about “why” without some theoretical framework, but no one theory can explain everything. “Matching” means generating the possible fits between the formulation and the potential therapies. Although no therapist will be able to offer all treatments, we are expected to become knowledgeable about the frameworks so we can think with patients about the options.

The choice of “match” focuses on using the evidence base when possible and always keeping the patient’s safety and preferences as central concerns. The final step is the discussion of one’s thinking and recommendations with the patient. The differential psychotherapeutics cycle is to be used at the beginning of the treatment but at other key points as well—if there is an impasse, the patient’s goals change, or the patient’s circumstances, symptoms, level of function, etc., change.

Mrs. E presents initially to a therapist who believes that patients often need more than one kind of therapy and that the choice of modalities should be openly discussed and decided on with the patient. The therapist suggests that both the depressive and eating symptoms are rooted in distorted cognitions—some about self-worth and perfection, others about food and body image. She suggests beginning with a trial of CBT, and the patient enthusiastically agrees. Mrs. E's eating symptoms respond very well and are soon in remission, but her depressive thinking proves more complicated. Her harsh self-criticism proves resistant to improvement. In fact, Mrs. E often hears the therapist's encouragement to work on symptoms as actually a veiled and sarcastic criticism of her lack of progress. Whenever the therapist notices this, Mrs. E is reminded of some painful childhood experiences of her father. More and more of their time is spent on this kind of exchange. The therapist tells Mrs. E that psychodynamic therapy would be indicated for this kind of problem and notes that the patient seems spontaneously drawn to this approach. The therapist primarily does CBT and has much less experience in a psychodynamic approach. She explains this to the patient and offers to do this work with her or to refer her to a colleague who specializes in this approach. Mrs. E decides that her growing trust and connection with her current therapist outweigh any disadvantage in relative expertise in the modality. They begin a new course of dynamic therapy together.

FUTURE DIRECTIONS

Even today, very few training programs teach an "integrated" approach to psychotherapy—one might be lucky to be exposed to multiple approaches as monotheories. The Accreditation Council of Graduate Medical Education (ACGME) only requires psychiatry residents to become "competent" in supportive, psychodynamic, and cognitive-behavioral psychotherapies. Each of these components tends to be taught by practitioners of that school and often remains siloed clinically and educationally. One of the reasons for this is the relative lack of academics who focus on integration. More integrationists tend to be found in clinical practice, where clinicians often see a wide range of patients and feel the need to expand the range of their techniques to assist them.

For integration to grow further, training programs need to be more proactive in including the topic in their curricula. While it is, of course, important to learn the basics first, being exposed to ideas from other schools of thought, as well as opportunities to observe the fluidity of integrationist thinking and practice, would go a long way toward normalizing the practice.

Having supervisors who can discuss a more integrated approach for specific patients would also be of great benefit.

Graduating psychiatrists who are interested in, or at least curious about, the integration will also lead to an increase in practitioners who identify as integrationists. The ability to be fully oneself with one’s practice will boost not only individual identity but also the identity of the integration practice ideas as a whole.

The more accepted the practice is, the more funding will become available for research of techniques and the creation of a more robust evidence base. Most research in psychotherapy, in general, is still woefully underfunded and confounding, as distinct approaches are pitted against each other in trials that rarely demonstrate a significant difference in efficacy. Research in integration instead offers the opportunity to focus on interventions and change processes that cut across schools of therapy.

The fluidity of thinking in integration may foster a more patient-centered way of thinking about care, as opposed to a theory-oriented focus. The focus can be less reliant on rigid DSM-V diagnoses and focus more on the various concerns, goals, and needs of a patient. Indeed, many of the most challenging patients we see, termed treatment-resistant, do not present with only one DSM-V diagnosis—they arrive at treatment with distress that represents the “integration” of many strains of psychopathology. The complexity of each person —when viewed through the various lenses of personality/temperament, child development, and the social determinants of mental health—is very difficult to capture within any one theory of psychotherapy.

This patient-centered focus may allow for more multicultural sensitivity. In viewing a patient's cultural context, race, sexuality, biology, beliefs, biases, and abilities as part of the whole, various techniques drawn from different schools of therapy may be valuable in order to reach each facet of a patient's needs and their interaction within society. Wakefield argues that the "moral imperative" we as practitioners have in helping the most patients possible is only possible with an integrated, multitheoretical approach. This makes for a more complete recovery and a more resilient sense of healing.

At the core of all treatments is the therapist's desire to help. The ability to offer aid to the diverse problems a patient brings to treatment can have a positive effect on the challenge of retaining our mental health workforce. The more help one can be, the more rewarding the work is, and the greater the desire and capacity to remain in the profession. Though burnout is a complex issue, the ability to create a robust and flexible clinical identity can go a long way toward retention and career satisfaction in the mental health field.

The flexibility of integrative treatment, which evolves as the patient's illness and recovery evolves, not only encourages us to remain hopeful about the most difficult mental health issues. Integrative approaches also provide a path toward rethinking the partisanship that has limited the help each clinician can offer their patient.

FURTHER READINGS

Accreditation Council for Graduate Medical Education. Program requirements for graduate medical education in psychiatry. https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/400_Psychiatry_2021.pdf?ver=2021-06-15-161510-003

Bloch-Elkouby S, Knopf L, Gold J. Assimilative psychodynamic psychotherapy: a critical case study reconsideration. *J Psychotherapy Integration*. 2017;27(2):254–264.

Boswell JF, Newman MG, McGinn LK. Outcome research on psychotherapy integration. In: Norcross JC, Goldfried MR, eds. *Handbook of Psychotherapy Integration*. Oxford University Press; 2019:405–431.

Brown J. Specific techniques vs. common factors? Psychotherapy integration and its role in ethical practice. *Am J Psychotherapy*. 2015;69:301–316.

Cabaniss DL, Holoshitz Y. *Different Patients, Different Therapies: Optimizing Treatment Using Differential Psychotherapeutics*. W.W.Norton and Company; 2019.

Castonguay LG, Newman MG, Grosse Holtforth M. Cognitive-behavioral assimilative integration. In: Norcross JC, Goldfried MR, eds. *Handbook of Psychotherapy Integration*. Oxford University Press; 2019:228–256.

Cooper M, McLeod J. A pluralistic framework for counseling and psychotherapy: implications for research. *Couns Psychother Res*. 2007;7:135–143.

Eubanks CF, Goldfried MR. A principle based approach to psychotherapy integration. In: Norcross JC, Goldfried MR, eds. *Handbook of Psychotherapy Integration*. Oxford University Press; 2019:88–104.

Feinstein R, Heiman N, Yager J. Common factors affecting psychotherapy outcomes: some implications for teaching psychotherapy. *J Psychiatric Practice*. 2015;21:180–189.

Frank JD, Frank JB. *Persuasion and Healing*. Johns Hopkins University Press; 1993.

Lazarus AA. Is there still a need for psychotherapy integration? *Current Psychology: Developmental, Learning, Personality, Social*. 2005;24:149–152.

Lazarus CN, Lazarus AA. Multimodal therapy. In: Norcross JC, Goldfried MR, eds. *Handbook of Psychotherapy Integration*. Oxford University Press; 2019:125–140.

Maeschalck CL, Prescott DS, Miller SD. Feedback informed treatment. In: Norcross JC, Goldfried MR, eds. *Handbook of Psychotherapy Integration*. Oxford University Press; 2019:105–124.

Markowitz JC, Milrod BL. What to do when psychotherapy fails? *Lancet Psychiatry*. 2015;2:186–190.

Norcross JC, Alexander EF. A primer of psychotherapy integration. In: Norcross JC, Goldfried MR, eds. *Handbook of Psychotherapy Integration*. Oxford University Press; 2019:3–27.

Norcross JC, Goldfried MR. The future of psychotherapy integration: a roundtable. *J Psychotherapy Integration*. 2005;15(4):392–471.

Prochaska JO, DiClemente CC. The transtheoretical approach. In: Norcross JC, Goldfried MR, eds. *Handbook of Psychotherapy Integration*. Oxford University Press; 2019:161–183.

Stricker G, Gold J. Assimilative psychodynamic psychotherapy. In: Norcross JC, Goldfried MR, eds. *Handbook of Psychotherapy Integration*. Oxford University Press; 2019:207–226.

Wachtel PL. Pathways to progress for integrative psychotherapy: perspectives on practice and research. *J Psychotherapy Integration*. 2018;28:202–212.

Wakefield JC, Baker JC, Conrad JA. Levels of meaning and the need for psychotherapy integration. *Clinical Social Work J*. 2020;48:236–256.

Wampold BE. How important are the common factors in psychotherapy? An update. *World Psychiatry*. 2015;14:270–277.

Wampold BE, Ulveus PG. Integration of common factors and specific ingredients. In: Norcross JC, Goldfried MR, eds. *Handbook of Psychotherapy Integration*. Oxford University Press; 2019:69–87.

Welton RS, Crocker EM. Supportive psychotherapy. In: Brenner AM, Howe-Martin LS, eds. *Psychotherapy: A Practical Introduction*. Wolters Kluwer; 2021:105–128.

32.2 Supportive Psychotherapy

JOHN BATTAGLIA, M.D.

The science of psychotherapy, the “talking cure,” was developed by Freud based on his topographical and structural theories of the mind in the early 1900s. Freudian psychoanalysis became the *sine qua non* for treating neuroses and spread rapidly throughout the industrialized world. Fidelity to the psychoanalytic technique was seen as essential for treatment success, and therapy practice that strayed from classical psychoanalysis was disavowed. Despite this worldwide popularity, practicing therapists began to realize that many neuroses, as well as more serious mental illnesses, were not helped by psychoanalysis. Also, therapists noticed some patients improved even when psychoanalytic practice guidelines were not strictly followed. In one of the earliest papers to address improvement in psychoanalysis from something other than insight, Edward Glover proposed that inexact interpretations sometimes led to symptom improvement by repression and displacement, thereby reducing anxiety.

Franz Alexander was trained in the classical Freudian technique. However, he recognized that many patients were unable to tolerate the stress of relatively unstructured therapy sessions and needed more guidance from the therapist. He outlined the need for the therapist to be flexible and use shifting tactics to accommodate the patient. He advised modifications from the standard psychoanalytic technique, including changing the frequency of sessions, using directives in the interview, helping the patient address real-life problems, and modulating the transference to make the sessions more tolerable for the patient. These modifications of the psychoanalytic technique comprise the fundamentals for what we now define as supportive psychotherapy. The therapeutic modulation of the transference, specifically maintaining a positive transference, is now considered the cornerstone of supportive psychotherapy. These supportive techniques were not developed from a theoretical perspective. Rather, they sprung from the keen observations of Alexander in response to what patients needed from the therapist in order for healing to occur.

Alexander realized that people had transformative, life-changing experiences from relationships outside of therapy and questioned the degree to which insight was necessary for change. While classical psychoanalysis emphasized insight as the curative force in therapy, he realized that the process of obtaining comfort from the therapist also provided a curative effect. This process–that of the patient overcoming painful experience in therapy (either from addressing their difficult emotions or experiencing turbulence in the transference) in the context of an accepting, nurturant therapist–was labeled the corrective emotional experience by Alexander. The concept of the correctional emotional experience as a fundamental healing component in supportive psychotherapy remains cogent today.

Over the next several decades, a number of therapists began recognizing that supportive elements in psychotherapy were not only acceptable but necessary in many cases, especially with more seriously ill patients. Also, therapists were naturally “drifting” into using supportive techniques even with traditional psychoanalysis. These changes did not occur without controversy, and there was a considerable professional argument that supportive techniques were causing the infidelity of psychoanalytic therapy in a damaging way. In fact, many psychoanalysts did not consider supportive types of therapy as valid psychotherapy, epitomized by this quip from Crown, “if it is supportive, it cannot be psychotherapy; if it is psychotherapy, it cannot be supportive.”

At the heart of the controversy with supportive psychotherapy was the notion that patients might improve and display symptom improvement. However, had they really “changed”? Many psychoanalysts argued that supportive psychotherapy helped relieve symptoms but did not change the person in any fundamental way. The concept of *structural* change, that is, change of personality in a permanent way versus *behavioral* change (symptom resolution), was put to the test in perhaps what was the most ambitious psychotherapy study ever conducted. An extensive longitudinal study that spanned 23 years was developed by the Menninger Institute Psychotherapy Project in the 1950s to compare classical psychoanalysis (several days a week on the couch) to psychoanalytic psychotherapy (weekly “upright” psychoanalytic sessions face to face). A supportive psychotherapy arm was added as a control condition. The results were quite unexpected in that there were no significant differences among the three types of psychotherapy. The researchers noted that supportive elements were found in both psychoanalysis and psychoanalytic psychotherapy and that these elements accounted for many of the observed changes. Also, they concluded that change did not occur in proportion to insight, and thinking of change as “structural” as opposed to “behavioral” was not a useful concept.

Although supportive psychotherapy lacked an underlying theoretical model, the practice techniques and strategies became standard among therapists of many different persuasions. The widespread effectiveness of supportive techniques in therapy could not be denied, and eventually, the term “supportive psychotherapy” became accepted as a separate, bona fide therapy. The first textbook on supportive psychotherapy was published by Werman in 1984, and at least a dozen others have followed since. Supportive psychotherapy skills are now listed as core requirements for psychiatry residency training in the United States.

RESEARCH

The Menninger Psychotherapy Research Project epitomized what would become a default standard practice for research on supportive psychotherapy, and that is, supportive psychotherapy has often been studied as a “treatment as usual” or “comparison” condition when other types of therapy were being examined. Despite this lack of concerted investigational focus on supportive psychotherapy, it often performs well in psychotherapy research even when compared to treatments that are the areas of interest for the academic institutions studying them. For example, Klein and associates investigated behavior therapy versus imipramine treatment for specific phobias and added supportive psychotherapy as a control arm of the study. Therapists trained in the behavior therapy techniques followed a highly structured, manualized protocol that included relaxation training, systemic desensitization in imagination, in vivo desensitization homework assignments, and assertiveness training (including modeling, role-playing, behavioral rehearsal, and in vivo homework). The supportive therapists were instructed to be empathic, nondirective, and nonjudgmental, encourage the ventilation of feelings, and allow discussion of life problems. The results were surprising and showed that both behavior therapy and supportive psychotherapy were highly effective, with no significant differences between them. The authors pointed out that supportive psychotherapy performed beyond their expectations.

McIntosh and associates conducted a randomized, controlled study looking at cognitive behavioral therapy versus interpersonal therapy for anorexia nervosa, and nonspecific supportive clinical management was added as a third arm to mimic treatment as usual in the community. The cognitive-behavioral and interpersonal therapy arms of the study included highly structured treatment manuals, which were followed by therapists trained in the techniques. The cognitive behavioral protocol included psychoeducation, motivational assessment, cognitive behavioral skills training (including thought restructuring and homework), relapse prevention, and recovery strategies. In the interpersonal therapy arm, the patient’s life events, interpersonal relationships, and eating problems were assessed, and links between them were examined. Problem areas in domains of grief, role transitions, interpersonal disputes, and deficits were identified, and their links with depressive symptoms were examined. In the nonspecific supportive clinical practice, patients were provided education and support about their eating disorder, and therapists provided praise, reassurance, and advice. The researchers fully expected the two specialized psychotherapies would be superior. However, the patients who received the nonspecific supportive clinical management had outcomes as good or better on most measures. The authors of the study were surprised by these results and noted that supportive management techniques likely best provided the important nonspecific factors of psychotherapy, such as the therapeutic alliance, empathy, and positive regard.

Rutledge and associates looked at cognitive behavioral versus supportive psychotherapy in nurse-delivered telemedicine treatment for chronic back pain. The cognitive behavioral arm was highly structured and followed a manualized protocol that included the treatment rationale, pain education, patient activity log monitoring, relaxation training, monitoring of sleep hygiene, sexual activity and exercise, homework assignments, and examination of pain cognitions, cognitive distortions, and pain behaviors. In the supportive psychotherapy condition, the patients were provided a self-help book on back pain, and the nurses were instructed to employ active listening techniques following Rogerian principles. Both cognitive behavioral and supportive psychotherapy was moderately effective in reducing pain, with no significant differences between them.

These three studies spanning nearly four decades, epitomize the research frequently conducted on supportive psychotherapy. It is often used as a control/comparison condition for more rigorous and specialized treatments under study. It is usually equally efficacious to other psychotherapies. Finally, it requires little training or expertise. Supportive psychotherapy techniques are easily adaptable for practitioners outside of psychiatry, including nurses and family practitioners. Supportive psychotherapy research has shown effectiveness for a variety of psychiatric and medical conditions, including schizophrenia, bipolar disorder, depression, anxiety disorders, personality disorders, substance use disorders, eating disorders, perinatal stress, breast cancer, ovarian cancer, diabetes, leukemia, heart disease, chronic bronchitis, emphysema, inflammatory bowel disease, back pain, and for hemodialysis patients.

CLINICAL ISSUES, STRATEGIES, AND TECHNIQUES

A key component for successful, supportive psychotherapy is to maintain a positive therapeutic alliance. Negative transference is handled by the therapist actively providing comfort, reassurance, and "repairing" the disruption in the therapeutic alliance as opposed to interpretation. A number of studies have supported the view that a positive therapeutic alliance is most strongly correlated with therapy treatment success, regardless of the type of therapy utilized (cognitive behavioral, interpersonal, psychodynamic, other), patient characteristics, or country. A recent study by Leibovich and associates showed that supportive techniques acted as mediating factors for therapy success in the treatment of depression by strengthening the therapeutic alliance and subsequently decreasing symptoms.

In supportive psychotherapy, the therapist not only engages in maintaining a positive therapeutic alliance, but also in modulating the pacing of the sessions, the anxiety of the patient, and the degree to which the patient should "open up" versus "cover-up." Contrary to more exploratory or provocative therapies, it is the job of the supportive therapist to proactively guide the patient throughout the session so they can adequately cope with their psychic pain. First and foremost, the supportive therapist helps the patient develop coping skills and methods for soothing themselves when feeling bad. This will lay the groundwork for calming in later sessions when the patient addresses difficult feelings and lower the degree to which the patient may “act out” as a result of a painful therapy session (“acting out” is unconsciously acting on the painful emotion outside the psychotherapeutic session, often in a physical way, and often with adverse consequences). Time management of the sessions is also an important ingredient for the therapist to actively monitor. Enough time must be conserved at the end of each session in order to help the patient cope with the material that was worked on, as well as to prepare for possible ripple effects of emotions that might arise later after the session has ended.

The style of the supportive psychotherapist can be described as conversational. That is, the therapist is able to use emotion in their voice much the same as when talking with friends. They do not mimic a “blank slate,” stare, or subject the patient to long, anxiety-provoking silences. This conversational style, when used in a comforting manner, helps the patient feel more at ease with the therapist, less like an alienated “subject,” and helps form a positive therapeutic alliance. Many, if not most, supportive therapy strategies are aimed at helping the patient feel less anxiety and safe in the therapeutic environment.

There are many supportive psychotherapy techniques, limited only by the creativity of the therapist. Providing *encouragement, hope, reassurance*, and *praise* are common supportive techniques that help the patient’s self-esteem. Over time the patient is sometimes able to incorporate some of the comfort and understanding of the therapist into their own self-acceptance. In psychoanalytic terms, the therapist is “lending ego,” and the patient develops a more positive self-image by introjection of the therapist. This process falls within the theoretical perspective of the correctional emotional experience as described by Franz Alexander so many years ago.

Psychoeducation (teaching the patient about their symptoms), *coping skills training,* and helping the patient find ways to soothe themselves are also mainstays of supportive therapy. When teaching *self-soothing* techniques, it is useful for the therapist to overtly label the process "self-soothing," as this concept is foreign to many patients and increases their awareness that they can be a curative force for their own suffering rather than rely on external supports. This also reduces their tendency to act out when experiencing difficult emotions. These techniques are all most effective when practiced while maintaining a positive therapeutic alliance.

Explaining behavior can greatly benefit the patient to feel less alienated and also contribute to a positive therapeutic alliance. By explaining behavior, the therapist helps the patient understand that they are acting in ways that make sense to their particular situation, even if seemingly "crazy" or unhealthy. This *normalizes* their behavior, thereby reducing stigma and helping improve self-esteem.

The use of *storytelling* and *metaphor* are powerful techniques that often can have a lasting impact when used in ways most cogent to the patient's situation. Storytelling and metaphor are sometimes more palatable for the patient because these techniques seem less directed toward them and, therefore, less confrontative. This allows the patient to listen less defensively, and the images from the story or metaphor then have more enduring effects. For example, a patient might become anxious or defensively reactive if directly told they were self-sabotaging. However, they would hear it more easily if told a story about a person who would "shoot themselves in the foot" with their behavior.

The technique of *confrontation* is therapeutically holding a mirror to the patient's behavior, so they may see how their own patterns contribute to their suffering. Contrary to lay usage, confrontation by the therapist does not involve any kind of anger or stern demeanor and is most effective with employed in a nurturant, accepting manner of the therapist. Confrontation is a powerful and useful supportive therapy technique and can only be used with a strong therapeutic alliance. The stronger the therapeutic alliance, the better the patient is able to tolerate confrontation.

There is a wide range for the number of sessions needed to perform supportive psychotherapy, dependent upon both the nature of the stressors present and the degree of psychopathology in the patient. Supportive psychotherapy is sometimes conducted over just a few sessions to help patients with adjustment disorders, while at other times can be utilized over the lifetime of the patient. Patients with primitive personality disorders or major mental illnesses are seen as having significant mental “defects” causing chronic symptoms as opposed to having neurotic “defenses.” With this perspective, some patients will require ongoing supportive psychotherapy on a chronic basis, much as patients with brittle diabetes will need continual, ongoing endocrinologist care to maintain their medical health. Thus, a specific process of termination in supportive psychotherapy is sometimes not necessary. When the termination process is indicated, the therapist provides the patient ample opportunity to express their feelings about it as well as review what insights or learning from the therapy may be taken forward with them.

Supportive psychotherapy can be an intense, intimate process, and some patients will internalize their positive experience with the therapist and utilize the interpersonal gains well beyond termination. For patients with significant losses in their lives, successful termination with a nurturant, supportive psychotherapist can be a corrective emotional experience.

FURTHER READINGS

Accreditation Council for Graduate Medical Education. *ACGME Program Requirements for Graduate Medical Education in Psychiatry*.

Alexander F, French TM. *Psychoanalytic Therapy: Principles and Applications*. Ronald Press; 1946.

Battaglia J. 5 keys to good results with supportive psychotherapy. *Current Psychiatry*. 2007;6(6):27–34.

Battaglia J. *Doing Supportive Psychotherapy*. American Psychiatric Association Press; 2019.

Conte HR, Plutchick R. Controlled research in supportive psychotherapy. *Psychiatric Annals*. 1986;16(9):530–533.

Crown S. Supportive psychotherapy: a contradiction in terms? *Br J Psychiatry*. 1988;152:266–269.

Fluckiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: a meta-analytic synthesis. *Psychotherapy*. 2018;55(4):316–340.

Glover E. The therapeutic effect of inexact interpretations: a contribution to the theory of suggestion. *Int J Psychoanalysis*. 1931;12:397–411.

Klein DF, Zitrin CM, Woerner MG, Ross DC. Treatment of phobias, II: behavior therapy and supportive psychotherapy. Are there any specific ingredients? *Arch Gen Psychiatry*. 1983;40:139–145.

Kleinman K, Wenzel A. Principles of supportive psychotherapy for perinatal distress. *J Obstet Gynecol Neonatal Nurs*. 2017;46(6):895–903.

Leibovich L, McCarthy KS, Zilcha-Mano S, Front O. How do supportive techniques bring about therapeutic change: the role of the therapeutic alliance as a potential mediator. *Psychotherapy*. 2020;57(2):151–159.

Manne SL, Rubin S, Edelson M, et al. Coping and communication-enhancing intervention versus supportive counseling for women diagnosed with gynecological cancers. *J Consult Clin Psychol*. 2007;75(4):615–678.

McIntosh VW, Jordan J, Carter FA, et al. Three psychotherapies for anorexia nervosa: a randomized, controlled trial. *Am J Psychiatry*. 2005;162(4):741–747.

Rockland LH. A review of supportive psychotherapy. *Hosp Community Psychiatry*. 1993;44(11):1053–1060.

Rogers CR. *Client-Centered Therapy; Its Current Practice, Implications, and Theory*. Houghton Mifflin; 1951.

Rutledge T, Hampton Atkinson J, Holloway R, et al. Randomized controlled trial of nurse-delivered cognitive behavioral therapy versus supportive psychotherapy telehealth interventions for chronic back pain. *J Pain*. 2018;19(9):1033–1039.

Szigethy E, Bujoreanu SI, Youk AO, et al. Randomized efficacy trial of two psychotherapies for depression in youth with inflammatory bowel disease. *J Am Acad Child Adolesc Psychiatry*. 2014;53(7):726–735.

Wallerstein RS. The psychotherapy research project of the Menninger foundation: an overview. *J Consult Clin Psychol*. 1989;57(2):195–205.

Werman DS. *The Practice of Supportive Psychotherapy*. Psychology Press; 1984.

Williamson PS. Psychotherapy by family physicians. *Primary Care*. 1987;14(4):803–816.

32.3 Psychoanalysis and Psychoanalytic Psychotherapy

ADAM CRITCHFIELD, M.D., CHRISTINA KITT GARZA, M.D., AND PHILIP R. MUSKIN, M.D., M.A., D.L.F.A.P.A., L.F.A.C.L.P.

We live in an era in which psychoanalysis is no longer the dominant paradigm of psychiatric treatment. Analytic training institutes are not the prevailing institution that they were before the revolution of psychopharmacology beginning in the 1960s. There has been a differentiation and proliferation of various short-term symptom-targeted treatments, many of which are derivatives of psychoanalytic theories, ideas, processes, and practices. Nevertheless, psychoanalysis continues to evolve and thrive as a robust contemporary discipline. It is an effective treatment for many of the psychic ills that bring patients into consulting rooms, clinics, and hospitals.

Unfortunately, because of past problems, including the rigidity with which it was once practiced, there continues to be an erroneous idea that psychoanalysis still represents a “one size fits all” treatment. This is a misconception as psychoanalytic treatment is deeply individualized to the patient, to the person’s specific life history and circumstances, and to their unique subjectivity, modes of thinking, fantasies, affects, and actions. Despite differences among analysts, there are core values to which most would subscribe and that this chapter will describe.

This chapter will detail what constitutes psychoanalysis as a treatment and explore its derivative psychotherapies, including various forms of psychodynamic or psychoanalytic psychotherapy. Our goal is to convey something about the psychoanalytic sensibility and the attention to interpersonal and psychological processes that are applicable not just to these individual outpatient treatments but also can be applied to many domains of the psychiatrist’s work across diverse patient populations and treatment settings. As Nancy McWilliams writes, “Most fundamentally, psychoanalytic practitioners take seriously the evidence that the sources of much of our behavior, feelings, and thoughts are not conscious. There is a dynamic organization to how we unconsciously register experience, an organization that prompts analysts to refer to *the unconscious.”*

What in nonconscious sources of thought, feeling, and behavior might lead someone to seek psychoanalytic treatment? Consider these three cases:

A 30-year-old man who gets engaged, appearing enthusiastic as he tells family and friends. Almost immediately, he begins to have panic attacks and nightmares about death. Seeking treatment for the panic attacks, he gives a history that his father died soon after he was born. In his early childhood, his mother endlessly repeated stories of their lives as newlyweds before his unexpected and untimely death. It becomes clear that there is a set of *affects* that the patient is not immediately aware of, involving an emotional recollection and re-experiencing of fears in his early childhood mind that linked the marriage of his parents with his father's sudden unexpected demise. The man makes no conscious connection between his current symptoms and his childhood experience.

A woman in her mid-20s who repeatedly goes to various physicians is convinced that she has an undiagnosed illness that would explain her chronic pain. Following a comprehensive evaluation and physical exam, she is told, "there is nothing wrong with you." The woman feels angry and dismissed, believing herself to be rejected by yet another purported caretaker. This replicates the neglect she experienced from her alcoholic mother as a child during her mother's binges. There is a particular *object relationship* that is not yet conscious. This is how she experiences herself in relation to the internal representation of an important other that profoundly shapes how she views these emotionally charged encounters with physicians.

A financially successful screenwriter who has worked mostly by himself accepts a leadership position for a theatre organization, only to become angry and depressed at his inability to feel appreciated at work. He goes to great lengths to avoid chance encounters with the very people who selected him, afraid that if an interaction does not go well, he will feel ignored and disrespected, leading to an angry outburst that might ruin his increasingly public reputation. This man is not consciously aware of the impact of his unmet childhood need for mirroring and validation that plays out over and over again in his adult life.

All of these examples reflect unconscious aspects of an individual’s psychological functioning that can lead to suffering, inhibitions, and restrictions in personal growth. This chapter will explain how it is that psychoanalysis and psychodynamic psychotherapy affect change, in part, by identifying such unconscious processes. In contemporary psychoanalytic thinking, there has been a shift away from the noun of Freud’s early writing: “*the* unconscious” as if it is a concrete place in the mind or a “second mind.” There is now an emphasis on the unconscious as an adjective, describing an important aspect of various mental processes that can render them problematic.

We acknowledge that this is an enormous subject. It includes the rich history and the extensive set of practices that constitute a treatment for psychopathology, a set of theories and working models of the mind, and the various derivative psychotherapies based on those theories and practices.

The field has come a long way since Freud. Multiple schools now exist that focus on different models of the mind and different lenses through which to examine particular problems in living as well as types of psychopathology. Some of these models overlap. There are also points of tension that are constantly being renegotiated in an effort to clarify, reconsider, or integrate contested viewpoints. This has led to a modern psychoanalytic landscape known for ambiguity and plurality of views that can be deeply enriching as well as confusing, disorienting, and frustrating for the beginner and even seasoned practitioners.

Psychoanalytic training takes years, following an already lengthy postgraduate education, which can limit its appeal to practitioners and availability to patients. Psychoanalytic treatments are focused on individual *meaning* and facilitate in the patient a gradual elaboration and expansion of their inner worlds. This fosters a greater capacity of awareness and ability to observe, reflect on, and understand the multidetermined associations and sources of thoughts, feelings, and behaviors. The enormous richness and complexity of the field cannot be conveyed in one chapter; however, we hope to provide a solid foothold for the reader to venture a climb through the fundamental concepts of the psychoanalytic models of mind as well as the clinical process.

The emergence of psychoanalysis, as created by Freud in the late 1800s, was a direct outgrowth of the treatment of neurotic patients at that time by physicians in general practice—the use of hypnosis for hysteria. Freud's friend and older colleague, the Viennese internist Josef Breuer, set the stage in 1881 with his serendipitous discovery that a patient's symptoms disappeared as a consequence of their verbal expression under a hypnotic trance. This was the famous case of Fraulein Anna O., who resides in the annals of psychoanalysis for having given it its impetus as well as its lasting name, the "talking cure" or, as she jokingly referred to it, "chimney sweeping." When Freud heard of the success of the technique in 1882, he was just completing his medical training with an assignment at Vienna's Allgemeine Krankenhaus, under the direction of Theodor Meynert. Struck by Breuer's compelling account, he spent the year from 1885 to 1886 studying hypnosis at the Salpêtrière in Paris with the French neurologist Jean-Martin Charcot. The latter's direct demonstration that with hypnosis, he could both elicit and abolish hysterical symptoms so captured Freud's imagination that it became a major turning point in his thinking. In his practice in the late 1880s, Freud used hypnosis to enable patients to suppress or deny their symptoms while under its influence. More important to Freud's subsequent thinking, however, was the use of hypnosis in retracing the history of symptoms.

When Freud returned to Vienna and subsequently established his medical practice, he had no success in treating hysterical patients with the more typical methods of the day, such as hydrotherapy, electrotherapy, and rest cures. He then decided to draw on the profound influences of Breuer and Charcot. The hypnotic method at that time was used strictly for psychic catharsis, attempting to gain access to patients' reminiscences that they could not remember under any other conditions. The particular method was a form of cathartic therapy borrowed from Breuer.

In their collaborative work on hysteria in the mid-1890s (Studies in Hysteria), Breuer and Freud discussed their theory that the symptoms represented repressed ideas that had not reached consciousness and were a transformation or “conversion.” They suggested that hypnotic treatment was effective because of the discharge of affect previously attached to the repressed mental acts, called “abreaction.” A major requisite for the cathartic method, which was contingent on that theoretical position, was that the patient could be hypnotized and thus be enabled to revert to the psychological state in which their morbid symptoms had appeared for the first time. Their clinical experience and observations suggested to them that the patient’s memories, thoughts, and impulses that were no longer conscious could enter the mind under hypnosis and that when recounted to the physician, especially under circumstances of intense emotion, the symptoms would disappear automatically. It was consistent with the theory that hysteria was the result of traumatic experiences, usually sexual in nature, whereby large quantities of affect that had been barred from consciousness appeared in disguised form as symptoms. Freud poignantly noted that “hysterics suffer from reminiscences.”

Following Breuer’s precedent with Anna O., Freud used the cathartic method (in conjunction with hypnosis) in 1889 to locate the traumatic origin of the hysterical symptomatology of Frau Emmy von N. In describing the case, Freud anticipated his later thesis, which speculated that inhibited sexuality played a role in the etiology of her illness. Freud grew dissatisfied with hypnosis because he was not able to hypnotize all his patients who suffered from hysteria. He was also frustrated by the transitory effects of hypnosis that seemed to last only as long as the patient’s contact with the physician. He began to suspect that the therapeutic effects depended more on the relationship between patient and doctor than on the technique per se. His suspicions were prompted by a striking incident in which one of his female patients threw her arms around him on waking from a hypnotic trance. (Breuer had earlier reported a similar incident.) As a result, Freud turned his thinking in the direction of the physician–patient relationship itself and, with that, the origins of his eventual discovery of transference. It was not until his use of hypnosis failed in the case of Fräulein von R. in 1892 that Freud felt obliged to experiment with modifications of the original hypnotic–cathartic method. The physician, Hippolyte Bernheim of the Nancy School, an adversary of Charcot, influenced the technical direction that Freud took. Freud's decision to give up the hypnotic–cathartic method was largely based on Bernheim's observation that the events recalled during hypnosis (in a hypnoid state) could be re-evoked by patients in states of consciousness when prompted by leading questions and urgings of the physician. Basing his concentration method on that premise, Freud developed a strategy in which he requested the patient to lie on a couch with eyes closed, then concentrate on a particular symptom and attempt to recall past memories; Freud would place his hands on the patient's forehead as a physical way of facilitating the memory process.

Through his work with the same recalcitrant patient Fräulein von R., Freud discovered the method that is identified with the true creation of psychoanalysis—free association. Freud soon realized that his interruptions of the patient and his constant prodding for additional details hindered the spontaneous flow of the patient's thoughts. A simple remark by Fräulein von R., namely that she did not know what to say because she was not sure what Freud wanted to hear, sparked his realization of the undesirable effects of censorship and direction by the analyst. He began to instruct his patients to express even the most irrelevant, unpleasant, or trivial material. The method of free association is thus the gradual consequence of experimentation with technique—from hypnosis and the cathartic method to a concentration method, to an approach almost antithetical to the original suggestive methods. As a new technique, free association was (1) unlike hypnosis, which required patient suggestibility; (2) unlike catharsis, which focused on the release of affects only;

and (3) unlike the concentration method, which prodded, directed, and interrupted the patient.

In altering his therapeutic technique, Freud realized that a “different though not contradictory conception of the therapeutic process” was warranted. With the new emphasis on the removal of amnesia and the recovery of repressed memories, he observed that the same instinctive forces that brought about repression to obliterate pathogenic material from consciousness continued to exert force—resistance—against full disclosure through free association. Moreover, unexpected feelings toward the analyst—transference—provided additional resistance to treatment. The persistence of repression with free association brought to the fore those two concepts, the unconscious and transference, which remain the critical constructs of modern psychoanalysis.

The first section of this chapter will describe core concepts of psychoanalytic models of the mind, including unconscious process, motivation, conflict, defense, transference, countertransference, and fantasy. The second section will focus on core concepts of clinical process, including the psychoanalytic frame of treatment, mechanisms of therapeutic action, and resistance. Through the description of these core concepts of mind and treatment process, the various “major tradition” or psychoanalytic schools will be reflected in the evolution of these bedrock concepts.

For example, our discussion of unconscious process and motivation will reflect Freud’s initial discoveries and formulations of the topographical model of the mind as well as his emphasis on the way that the mind grapples with physiologic needs. The concepts of defense and internal conflict reflect developments in ego psychology which became paradigmatic of psychoanalysis during its heyday in American psychiatry from the 1940s to 1960s. We will discuss what continues to be useful as well as what is potentially problematic about these early models. We will consider findings from modern neuroscience that may shed light on their scientific validity as well as the limits of such an enterprise.

Transference and unconscious fantasy, while Freud's discoveries, were greatly elaborated and developed by object relations theorists, including Melanie Klein, Wilfred Bion, Ronald Fairbairn, and D.W. Winnicott. Countertransference became a focus of contemporary relational psychoanalysts and intersubjectivists. This chapter will then explore various derivatives of psychoanalysis, including psychodynamic psychotherapy and several evidence-based, manualized treatments that have emerged under the umbrella of psychodynamic psychotherapy. Some of these manualized treatments were developed by psychoanalysts looking to integrate aspects of the aforementioned schools. For example, transference-focused psychotherapy for personality disorders features a useful integration of object relations theory and ego psychology as developed by Otto Kernberg and others. Lastly, we will look at how specific psychodynamic therapies and psychoanalytic concepts can be variously applied to major DSM-5-TR syndromes.

CORE CONCEPTS OF PSYCHOANALYTIC MODELS OF MIND

Motivation and Unconscious Process

Why do human beings do the things that they do? When a person does something that causes them to suffer, it may or may not be pathologic. The behavior may lead them into treatment with the hope of understanding how it is that they do something repeatedly that causes pain to them or to people with whom they are in relationships. As noted by George Vaillant, "Over the short haul, neurotic defenses make the user suffer; immature defenses make the observer suffer." Part of the clinician's work is to understand the "why" or what is *motivating* the behavior or set of choices that the individual makes. We can take this deeper and ask such questions about not just behavior but the feelings themselves. That is, we can question the motivation(s) underlying a particular experience of an affect, or even an experience of the self.

Freud started theorizing a metapsychology of motivation using the concept of drive. Contemporary psychoanalysis uses this theoretical concept less, but it is worthwhile to consider why this was a starting point for Freud and for psychoanalytic psychology. For Freud, the drive is “the psychical representative of the stimuli originating from within the organism and reaching the mind, as a measure of the demand made upon the mind for work in consequence of its connection with the body.” A drive is a *frontier concept* between body and mind, an unconscious representative of physiologic needs in the mind, and therefore a way of thinking about the pressure that the body places on the mind for the mind to “do work.”

We can distinguish the concept from instinct, an unconditioned behavior pattern prompting reproductive success, in its emphasis on psychological representation. People can feel driven toward certain feelings and behaviors on the basis of biologic imperatives. This was an idea that, while it seems obvious to us now, was threatening to the culturally sanctioned ways of viewing human thought and behavior in Freud’s time. Humans are motivated to satiate biologic needs like hunger, thirst, and sexual expression. We now understand this is also true for attachment needs.

At times those needs come into conflict with what is allowable or available in the external or cultural environment. In circumstances when the expression of a physiologic need is restricted, the mind must negotiate with the mental representation of the need to find the maximum allowable expression of the need in order to obtain gratification. This “drive satisfaction” is the maximum possible reduction in physiologic tension and all within the limits of what is acceptable to the person’s (more or less conscious) sense of self and the internalized limits imposed on the self from parents or wider society.

Freud also thought that this was a key function of dreams, allowing “wishful hallucination” to keep the dreamer asleep; take a common example of dreaming, one is drinking water only to wake up parched with thirst. The role of dreams in analytic therapy is now known to be more complex and nuanced than simple wish-fulfillment, as will be explored later in this chapter, but we can see how Freud was grappling with ways that the mind is forced to reckon with the desire-filled demands of the body. It was in that grappling that the original theoretical ideas, as well as the nature of the clinical practice of psychoanalysis, was born.

Dreams reflect both the inner struggles of the mind and the world at large.

A patient described the change from a typical “anxiety dream,” that is, being back in college and discovering he had not attended a class all semester and it was the day of the final. During the height of the COVID pandemic, he noted the dream had changed to being on campus with no one wearing a mask, crowds of people walking towards him, and him not having a mask. “Did COVID change how we dream” (NYT Brooke Jarvis, 11/3/2021).

A person’s motive for their behavior may not be immediately apparent, observable, or understood. Much of the clinical work in psychotherapy, irrespective of the specific psychopathologic syndrome, is to identify and elaborate causal links between thoughts, affects, and behavior, as well as their situational contexts and historical antecedents. It is when motivation is not conscious to the patient that achieving change is most difficult. They may feel at a loss for how to modify their behavior or feel that a behavior is purely “automatic” that they “can’t help it,” or have no motivation to modify that behavior. A person can feel “lived by” one’s impulses and desires rather than feel themselves to be the experiencing agent who desires the object of the impulse. This can obscure the person’s awareness of the reasons for such impulses.

When a person comes to see the function that a particular choice serves in the maintenance of their psychological equilibrium or identifies the “hidden gratification,” the behavior becomes mutable as the true cost/benefit analysis of the action becomes clearer. As explored in the following section on mechanisms of therapeutic change, making such “hidden gratifications” or motives conscious is one of the ways that analysis may lead to change over time. This is rarely an immediate change.

A patient with dependent personality traits has somatization, which worsens in the setting of relationship conflict. Over the course of analytic treatment, this patient becomes aware of the physical symptom as a “safe” expression of aggression towards, and control of, their partner, upon whom they are dependent. The dependency is unconsciously anxiety-producing with fears of abandonment. The physical symptoms keep the partner involved but at the cost of suffering to both individuals. The awareness of this mental mechanism allows the person to reflect on the physical symptoms rather than succumb to them.

The evolution of the history of psychoanalytic thought involves the tendency of different psychoanalytic theories or schools to posit one or two supposedly supraordinate motivational systems to which other motivations are considered subordinate or derivative expressions. Freud initially posited a dual instinct of sexuality (libido) and a “self-preservative” instinct. He would ultimately revise his thinking, and in his second dual-instinct model, he assumed a primary sexual and a primary aggressive drive. The various psychoanalytic schools would, in turn, shift the emphasis, partially due to reactions against and desires to correct earlier models when it was felt certain core motivations had been de-emphasized.

For example, contemporary psychoanalysts, informed by infant observational research, consider attachment needs to be a primary motivational system. This is reflected in the perspective of object relations theorists and relational psychoanalysts. This perspective emphasizes *relational needs* (such as the need for a sense of self-coherence found in an understanding other) as a primary motivation as opposed to drive discharge or expression.

One of the strengths of the psychoanalytic approach is that it explores and includes, but is not limited by, the conscious subjectivity of the patient. The analytic therapist listens intently on multiple levels, not just to the conscious verbal output of the patient but also to what may lie behind that content. This has also been called listening *from surface to depth*. Psychoanalysts contend not just with the words the patient is saying but what is latent in the manner in which a narrative is communicated, as well as what may be left out. Analysts listen for the "affective music."

Defense and Conflict

The motivation for using a defense is the avoidance of unpleasant affect. Different psychological defenses may be adaptive or maladaptive at this task in the short term (Table 32.3–1). Defenses, which may have conscious elements but are largely unconscious, may exact a long-term price for that short-term gain of comfort. A person's typical relied-upon defenses become part of their character, or what Wilhelm Reich first described as "character armor." One person may be very adept at dealing with painful situations by finding the humor in them without disregarding the difficulty. This helps the person stay engaged with the unfolding events and those who are involved along with them in a shared experience. We would describe this as a "mature" defense since it improves the ability to cope or adapt to difficulty without avoiding or distorting reality. This is in contrast to the less mature defense of denial.

> A middle-aged patient diagnosed with severe coronary artery disease reacts with, "This can't be right. I'm as healthy as a horse." Instead of pursuing the recommended evaluation and treatment recommendation, the person does nothing, leading to a bad outcome.

The immature defenses are more reality-distorting, meaning they get rid of a negative affect at the cost of the person's ability to adequately negotiate with reality. Thus, immature defenses are usually, therefore, more costly over the long term. The kinds of negative feelings that are defended against include negative feelings about the self, loss of love or approval from significant others, or painful realities like mortality or loss of a loved one, as well as the inability to achieve desired goals.

Psychoanalytic theory about defense was heavily elaborated by the ego psychologists who made particular use of Freud's *structural model of the mind* with its emphasis on wish, defense, and internal compromise. In the original metapsychology, the defenses were initiated "against the threat of the drives." Defense analysis became the predominant focus of ego psychology that dominated American psychiatry from the 1940s to 1970s.

Table 32.3–1. Defense Mechanisms

Mature Defenses

- Suppression: consciously putting a disturbing experience out of mind
- Altruism: vicarious but instinctively gratifying service to others
- Humor: overt expression of normally unacceptable feelings without unpleasant effect
- Sublimation: attenuated expression of drives in alternative fields without adverse consequences
- Anticipation: realistic planning for inevitable discomfort

Neurotic Defenses

- Repression: involuntary forgetting of a painful feeling or experience
- Controlling: manipulation of external events to avoid unconscious anxiety
- Displacement: transferring an experienced feeling from one person to another or something else
- Reaction formation: expression of unacceptable impulses as directly opposite attitudes and behaviors
- Intellectualization: replacing of feelings with facts and details
- Rationalization: inventing a convincing, but usually false, reason why one is *not* bothered
- Isolation of affect: separating a painful idea or event from feelings associated with it
- Undoing: ritualistically “removing” an offensive act, sometimes by atoning for it, or in thought by immediately “taking back” an assertion

Immature Defenses

- Splitting: experiencing oneself and others as all good or all bad
- Idealization: seeing oneself or others as all-powerful, ideal, or God-like
- Devaluation: depreciating others
- Projection: attributing unacceptable impulses or ideas to others
- Projective identification: causing others to experience one’s unacceptable feelings. One then fears or tries to control the unacceptable behavior in the other person

- Acting out: direct expression of an unconscious wish or impulse in order to avoid being conscious of the affect and thoughts that accompany it
- Omnipotent control: a fantasy of total control of another perceived to be hostile, usually as a result of projective identification
- Passive aggression: expressing anger indirectly and passively
- Intermediate denial: refusal to acknowledge painful realities

Psychotic Defenses

- Psychotic denial: the obliteration of external reality
- Delusional projection: inner conflicts are externalized and given tangible reality—minimal reality testing
- Schizoid fantasy: withdrawal from conflict into social isolation and fantasizing

Adapted from Muskin PR, Haase EK. Personality disorders. In: Noble J, Greene HL. *Textbook of Primary Care Medicine*. 3rd ed. Mosby, Inc.; 2001:458–464.

A useful aphorism of Freud is that the “child is the father of the man.” Conceptually this is how the developmentally young mind of the child copes with intense bodily wishes and feelings of frustration when its wishes are denied. Eventually, a child learns to delay gratification. This occurs as the child internalizes parental prohibitions and expectations as well as positive early feelings of mastery, accomplishment, and the synergy of good feelings about oneself alongside parental pride. The intensity of the internal “No!” that the child takes on can have a harsh, punitive quality. This impacts how the child’s mind will treat itself even years later as it comes to measure itself against that internal standard. That standard exists in a psychological structure that Freud, in his structural model, called the *superego*.

This was differentiated from the *ego* and the *id*. It is worth noting that these terms entered popular culture long ago. The concepts are nearly universally well-known and equated with the Freudian model of the mind; however, these labels were not quite the actual terms used by Freud. This results somewhat from the translation by James

Strachey from the original German in an effort to give these ideas the scientific sheen of latinate terms. The words Freud used were the “Ich,” the “Es,” and the “Uber-Es,” which can more directly be translated to the “I” (or *ego*), the “It” (the *id*), and the “Over-I” (*superego*). The plain language that he used is evocative and inclusive of ordinary meanings and invites one to see these different “agencies” of the mind from the subjective conscious perspective of the “I.”

From this perspective, it is one’s “I” that feels itself to be looking out at the world (through sense perception) and goes about mediating the interactions between the first-person, agentic self, and external reality. The structure that would come to be called *the ego* in ego psychology, and is said to be the mediator between wishes and desires (often unconscious, sometimes repressed, and therefore said to be located in the “it” or *the id*) and external reality. With the term “Over-I,” one can imagine the perspective of the small child’s self looking up at the towering authority that sets the rules and metes out punishment, but that also represents the wished-for ideal standard of a person that one hopes to grow into.

These three entities are not locations in the brain nor are they expected to correspond to any one neuroanatomical structure. Instead, they describe aspects of developmentally achieved psychological structures that are reflected in clinical phenomena, especially when it comes to internal conflict.

What does this look like in practice?

A premedical student is eager to please and help others. He holds to rigid standards that are also projected in the form of judgment upon others for not being adequately selfless. The student feels guilt for this judgment upon others as well as guilt for not living up to an idealized standard of morality informed by a religious upbringing that provided a sense of belonging and early models of moral authority. Seeking treatment for anxiety, the student is aware of feeling overly deferent and uncomfortable around authority figures. Aware that he feels compelled to be deferent, he has an inkling of an “irrational” motivation. He explores his feelings in relationships through an analytic treatment. This allows him to reflect on a growing awareness that, while his family instilled a sense of hope for him to achieve, they also found many opportunities to communicate an imperative not to achieve too much, lest he surpass them and no longer be someone to whom they could relate. These mixed messages around the dangers of achievement were a constant during his adolescence. Eventually, this patient takes on the role of monitoring himself lest he goes too far, restricting his interests and doubting his capacities. He develops a chronic sense of dysphoria and frustration, accompanied by guilt whenever he spends his time exploring personal interests.

We can say that this patient is experiencing a conflict between his nascent adult sense of self (separate from his family with his own interests and capacities) and harsh superego restrictions upon his achievement. This can lead to less-than-fulfilling life choices. He chooses not to be a doctor, despite having outstanding grades and accomplishments, and instead chooses to work in an ancillary capacity supporting other healthcare workers. He avoids situations that might lead him to feel angry at authority figures. His choice of a career that is more supportive and from the sidelines is an example of a *compromise formation* that partially, but only incompletely, satisfies both sides of an inner conflict. Through treatment, he develops a growing awareness of what each side of his conflict feels like and what are their historical determinants. Ultimately, he develops a greater sense of agency in making choices that are informed by a growing awareness of what he wants for himself, as an autonomous adult. This may include making new and different, but hopefully more satisfying, compromises.

Transference and Fantasy

Transference is one of the concepts most associated with psychoanalysis, which is logical as it is an important and ubiquitous psychological phenomenon. “Transference is of central importance in psychoanalysis because it demonstrates vividly that the past lives within the present and exerts a powerful force upon it. While transference is a universal feature of mental life, its role within the analytic situation provides a unique experience for the patient, one that provides the greatest leverage for therapeutic gain.” Alessandra Lemma defines transference as reflective of “the human predisposition to see current interpersonal experience through the largely unconscious lens of selected forms of past experience” but importantly adds, “it is not a photocopy of the patient’s childhood.”

When the psychoanalytic patient develops an attachment relationship to their analyst, the quality and nature of that attachment will contain elements of past relationships. These elements are then revealed and explored in the treatment. This process occurs through feelings, thoughts, and fantasies about the analyst, or what is known as *the transference*.

A patient arrives late for a session and talks about the fear they have disappointed their analyst with their lateness and "wasted your time." Associating to the concern, they remember the manner in which their depressive mother never seemed to care when they were around as a child. This patient arrives at an understanding that the worry conceals the wish to be important enough that the analyst would be disappointed when they didn't show up to a session on time.

It can be useful to understand transference through the lens of object relations theory. Freud coined the term transference, recognizing the impact of past relationships on the lived experience with the analyst. He focused on the idea of transference being a misreading or a distortion of current experience because of expectations formed in past relationships. Transference has come to be a more capacious concept and is seen as a working ground for interpersonal growth, not merely a distortion. The classical analytic literature indicates that an analysis is used to promote a "transference regression" or to establish a "transference neurosis" that is based on infantile wishes and conflicts that can then be worked through.

More modern psychoanalytic theories such as object relations, self-psychology, and intersubjectivity offer broader approaches to understanding the nature of transference in a psychoanalytic treatment. The elaboration of the transference in treatment allows a patient to understand the meanings inherent in their expectations of others' minds, including expectations about how another person will respond to the wishes and feelings of the patient. When responses from the analyst do not correspond exactly to what the individual expected in their early relationships, the potential for new forms of relationships begins.

Inasmuch as transference relates to early childhood relationships, common types of transferences may be characterized as a *paternal* or *maternal*, or even *sibling transference*. A series of transferences develop according to the stage of treatment. The transferences reflect particular relationship conflicts occurring in the person’s outside life and/or situations arising with the analyst that recall dynamics from the person’s early family life.

We can say that relationship schemas, or expectations, about the other’s mind and their responses are represented in *fantasies*. The fantasies may be initially unconscious and come to consciousness via the treatment. In a fantasy, we are relating in our mind to another person or feel that we are being treated by another person in a highly specific way. Fantasies are *affective narratives* and usually take the form of *who is doing what to whom and why?* This contributes to the affective intensity of the transference experience, that is, in a *positive transference*, the individual generally feels cared for, attended to, loved, and helped. They feel that the analyst is interested and curious and is doing something for their benefit. This experience contains derivates (fantasies laden with specific memory traces, feelings, and imaginings) related to earlier helpful, caring figures in the person’s life.

If the person experienced deprivation in their nuclear family but was aided and taken care of by trusted others (teachers, guidance counselors, or other relatives), this would play out in a variety of transferences. The analyst might expect the patient to alternately experience a *positive transference* associated with a precarious longing as well as a *negative transference* as if confirming the feared reality of neglect from parents. The greater the overall impairment in ego function and personality organization, which is usually a result of serious deprivation and outright trauma in childhood, the more prone a person is to predominantly negative transferences, which may be experienced as concrete reality. That is to say, they lose the “as-if” quality most advantageous to an ability to reflect on transference and to integrate insights about it into the personality over time. Such negative transferences are often hostile and paranoid in nature, accompanying a sense of being exploited or harmed. For more examples of such negative transferences as seen in borderline personality organization (BPO), refer to the transference-focused psychotherapy section of this chapter.

Countertransference and Intersubjectivity

Countertransference is one of the core concepts that has been most fruitfully elaborated after Freud's lifetime and has benefited particularly from theoretical developments in self-psychology and intersubjectivity. As psychoanalysis has developed over time, the analyst's participation in, and contribution to, analytic process and outcome has also evolved. The earlier view of them as being an objective, outside observer and interpreter of the patient's mind has evolved to greater recognition and inclusion of the analyst's own emotional life and responses as fundamental aspects of psychoanalytic understanding. The so-called *relational turn in psychoanalysis* is also regarded as a shift from the predominantly "one-person psychology" of the classical ego psychological model to the "two-person psychology" of intersubjectivity.

The original understanding of countertransference was that of the analyst's potentially idiosyncratic or personal response to a patient's transference. This was generally seen as a possible obstacle to treatment and something for the analyst to be aware of and work through in their own analysis. When Freud recognized that patients developed loving feelings towards their analysts as attachment figures that recapitulated feelings towards parents or other love objects, he saw the potential vulnerabilities this could lead to in the analyst. The analyst might feel gratified by these expressions of strong feelings in patients that appear to be directed toward the person of the analyst rather than the transference figure. This experience could set the stage for boundary crossings, as happened with early analysts in Freud's inner circle. For example, Carl Jung began an intimate, boundary-violating relationship with patient Sabina Spielrein. One of Josef Breuer's patients developed a delusional pregnancy in conjunction with an erotic transference prompting Breuer, frightened by this, to immediately terminate the treatment. This might be understood as Breuer's *acting out of the countertransference*. While those are dramatic examples, we now understand how these and the more subtle and productive manifestations of countertransference are a critical part of psychoanalytic work.

As a result of the frequency and concomitant affective intensity of analytic treatment, analysts come to be immersed in their patients' minds in a way that no other treatment can replicate. As analysts become attuned to their own reveries with patients, they hone an ability to use their associations, pattern-recognitions, and feelings to better understand patient material and behavior. The felt resonances of bodily affect and thought content often tell an analyst a great deal about the inner workings of their patients' minds. The epistemic status of countertransference data as "knowledge" about the patient is a complex issue. Rich theorizing about countertransference comes together in several strands under the header of relational psychoanalysis or intersubjectivity.

This includes the theory of self-psychology of Heinz Kohut and his understanding of narcissism. The works of Bion and Ogden are often cited in understanding how an analyst's own reverie in sessions allows for a more nuanced understanding of the unique analyst–analysand dyad. Robert Stolorow and Steven Mitchell each, in their own ways, integrate the infant observational research of Daniel Stern. Stern described "interaffectivity," or the mutual sharing of affect states as seen in the mother–infant pair, as an important potential forerunner of the analytic intersubjective process. It is the mother's responsiveness to the infant's recurrent affect states and internalization of those patterns that play a major role in the infant's developing affect- and self-regulatory capacities. Stolorow explains, "the child becomes vulnerable to self-fragmentation because his affect states have not been met with the requisite responsiveness from the caregiving surround and cannot become integrated into the organization of his self-experience." Defenses against affect become necessary to preserve a brittle self-structure from the experience of *fragmentation* that lies at the root of narcissistic problems, including narcissistic rage. As Stolorow describes, Kohut's "discovery of the developmental importance of phase-appropriate mirroring of grandiose-exhibitionistic experiences points to the critical role of attuned responsiveness in the integration of affect states involving pride, expansiveness, efficacy, and pleasurable excitement.”

Intersubjectivists tend to see transference not as distortion, projection, or displacement but rather as an organizing activity of the self. In turn, this has implications for countertransference to be understood not as a pathologic response to an individual’s “distorted” experience of the analyst but as a necessary part of being involved, responsive, and engaging with a patient’s subjectivity.

Stephen Mitchell outlines a useful framework for describing and situating the intersubjective process in clinical work in a way that both incorporates other aspects of psychoanalytic theory and helps us see what is uniquely intersubjective about this way of working. He describes an “interactional hierarchy” of four modes of interaction between the analyst and the patient that are of increasing organizational sophistication. These modes are (1) nonreflective behavior; (2) affective permeability; (3) self-other configurations; and (4) intersubjectivity. These are different “levels” of relationality as well as different lenses that organize a clinical moment. In any given interaction, one may be in the foreground while others operate in the background.

Nonreflective Behavior. This refers to the level of action involving two people or *what* is being done to *whom*. This behavioral dimension was seen as central to John Bowlby’s attachment theory and later infant observation that examined the behavioral microinteractions between child and mother, activities such as smiling, clinging, crying, and turning to or turning away from the other. This kind of careful tracking of mutual but “presymbolic” or prereflective behavior is a surface-level “way in” to an understanding of complex processes of mutual regulation.

Affective Permeability. This refers to the properties of affect that enable it to resonate from person to person, particularly with regard to intense affects like anxiety, euphoria, or rage. Humans are built for sharing affect, and Mitchell notes that at this level of affective permeability, questions like “who started it and who did what to whom? tend to be meaningless when intense affective connections are involved, as in strong sexual attraction, terror, rage, or joyous exhilaration." This level of analysis is often helpful in understanding the interpenetrability of transference–countertransference experiences and how the analyst's own feelings can be used as a window into the patient's affects, even dissociated ones.

Self-Other Configurations. This describes the level of object relations, where the complexities of *symbolic representations* of self and other can be understood. These are the internal object relations or object dyads of W.R.D. Fairbairn and Otto Kernberg's work, where a specific representation of self is linked to a specific representation of another by an affect state. For example, at any given moment, the patient may feel themselves a child in relation to a parent with the attendant affects, fantasies, and associative memories, which may then get played out in the therapeutic dyad. It is at this level that we can understand the contribution of Heinrich Racker's distinction between *complementary* and *concordant countertransference*.

If the patient is experiencing themselves as a child seeking understanding and advice from a caring parent, and the analyst feels themselves pulled into just such a parental attitude towards the patient, this would be considered a *complementary countertransference*. In other words, the analyst's own experience resonates with the object representation from the patient's internal object world. The analyst may also note the ways in which they are *not* resonating with the patient's experience in the transference.

A *concordant countertransference* is an identification with the affective experience of the patient—feeling what the patient is feeling—which may indicate empathic immersion in the experience of the patient but may also alert the observant analyst to a potential blind spot. For example, if the analyst is having personal problems with their spouse, and a patient is complaining of feeling disappointed and criticized by their spouse, this analyst would want to be aware of how their own personal feelings might be stimulated in such a way as to lead them to be more inclined to believe they knew exactly what the patient was experiencing. This can help the analyst maintain a neutral position with respect to the patient's conflicts.

Intersubjectivity. This recognizes that both persons in the dyad are complex, self-reflective subjectivities, or subjects "in their own right." Each will have ways of mutually regulating with others through behavior and affective resonances. Each has an internal object world that will be at play in the therapeutic encounter. Each can strive to understand the complex intentionality of mind at play, with the caveat that the psychoanalytic frame requires a degree of asymmetry between patient and analyst that is unavoidable in order for it to be clear that it is the patient who is there to be helped.

An early example in the psychoanalytic history of asymmetry in the therapeutic encounter being disregarded was in the work of Sandor Ferenczi. He experimented with "mutual analysis," in which the analyst also free-associated to the patient and discussed his own feelings and conflicts. It is easy to imagine how this could create a sense of impingement on the patient, who may feel pulled to "take care of" the clinician, with pathologic consequences. While an extreme example of a disregarded practice, it highlights the importance of asymmetry in the roles of analyst and patient, even in this contemporary era of intersubjective exploration. This asymmetry is constituted of the relative anonymity of the analyst, clear role expectations for the patient to freely associate, and the need for the attention of both parties in the room to be focused on the problems, feelings, and conflicts of the individual seeking help.

Intersubjectivity, especially as developed in the work of feminist psychoanalysts, was a necessary corrective to the dangers of a view of the analyst as a silent, all-knowing authority who has unique possession of knowledge of the patient's unconscious. However important a corrective provided by the postmodernist deconstruction of the analyst's authority, it is precisely the asymmetry of the clinical encounter that creates conditions for a safe playground of the mind in which two subjectivities work together for the benefit of the patient.

CORE CONCEPTS OF CLINICAL PROCESS

The Psychoanalytic Frame

The "frame of psychoanalytic treatment" or *psychoanalytic frame* refers to the contractual conditions that comprise the psychoanalytic situation, including but not limited to meeting times, location of sessions, payment, and use of the couch. Psychoanalytic sessions are typically 45 to 50 minutes in duration and take place three to five times per week. During the sessions, the patient is encouraged to *free associate* or to speak as freely as possible about their thoughts as they come into mind, regardless of how illogical or disconnected the thoughts may seem. Freud referred to this as the *fundamental rule* of psychoanalysis. He believed that through free association, patients could bring repressed material into conscious awareness. The psychoanalyst responds to the patient's free associations by asking questions to facilitate the deepening of the material and through the use of interpretations. In the course of free association, patients will, often out of their conscious awareness, show signs of avoiding discussion of certain topics, expression of affects, or to the analytic procedure itself. Freud termed this phenomenon *resistance*.

Psychoanalysts of all orientations tend to agree that a critical part of the analytic process should be devoted to making the patient aware of their resistances and to help the patient view these resistances as obstacles to the analysis that must be understood and overcome. Eventually, the patient will produce material that contains overt or covert references to their thoughts and feelings about the psychoanalyst. This material is derived from a combination of the patient's current perceptions of the analyst as well as past wishes, experiences, and relationships (particularly with significant figures such as parents) that the patient has experienced, which is known as *transference*. Transference analysis is at the heart of psychoanalytic technique and what distinguishes psychoanalysis from all other forms of psychotherapy.

Aside from the fundamental rule and the expectation that communication can be in any form except through physical contact or violence, analysands are otherwise subjected to few rules. Most of the rules governing the analytic situation apply to the psychoanalyst, who generally maintains a stance of anonymity in keeping with the “rule of abstinence.” According to Laplanche and Pontalis, the “[the analyst] should refuse to satisfy the patient’s demands and to fulfill the roles which the patient tends to impose upon him” in order to ensure that “the patient finds as few substitutive satisfactions for his symptoms as possible.” In addition to abstinence, the analyst attempts to maintain an analytic space in which the patient feels sufficiently safe and free to share their intimate thoughts and to fully engage in the analytic process. These aspects of the treatment frame are essential for preserving the analytic space from both internal and external influence, as described by Casement in 2013:

> “The consulting room space gives privacy and protection from intrusion by others; the set times, kept specifically for each patient, provide reliability and continuity; the professional ethic offers a guaranteed confidentiality and nonexploitation of the patient; and the relative anonymity of the analyst aims to keep to a minimum the impact upon the patient of knowing personal details about him/her… The analytic space also offers a freedom from the intrusive pressures (influence, reassurance, advice, or moral judgment) that could arise from any personal or theoretical predisposition of the analyst.” (Excerpted from Casement P. *Learning from the Patient.* Guilford Press; 1985:340.)

The analytic situation has been referred to as a “transitional space” in which the patient, facilitated by their analyst, can “regress in the service of progression” (as described by Balint and Sandler in 2018), test out new ideas, and challenge old ones. Winnicott referred to the analytic space as a “holding environment” and emphasized the analyst’s role in ensuring the patient can feel safe and “contained” even when in the midst of a severe regression. Schafer believed that the analytic attitude of “gentleness, undemandingness, open-mindedness, flexibility, patience, tentativeness, spontaneity, and individuality, and willingness to go along” was also crucial to maintaining an atmosphere of safety in the analytic situation.

Traditionally, patients in psychoanalytic treatment are invited to lie recumbent on a psychoanalytic couch facing away from the psychoanalyst. This technique was believed to allow analysands greater access to their unconscious minds by freeing them from inhibitions related to censoring or judging thoughts that they may experience when they see their analysts face-to-face. Over the years, at times out of necessity, such as during the COVID-19 pandemic, analysts have made modifications to the standard analytic technique, including but not limited to allowing patients to sit up or meeting with patients via telephone or video chat to accommodate for special circumstances. Such changes are known as “parameters of technique” and can be employed either temporarily or on a longer-term basis. When working in an analytic treatment, changes to the frame or shifts in parameters are always to be explored with the patient in order to understand the meanings that such changes generate.

Mechanisms of Therapeutic Action

Each major school of psychoanalysis has a theory of *therapeutic action* or idea of how psychoanalysis works to produce a therapeutic benefit that is based upon its own unique model of the mind and the associated theory of technique. Contemporary theories of therapeutic action tend to emphasize the role of interpretation and of the function of the therapeutic relationship in producing change. Though there are limited studies to support certain techniques being associated with specific outcomes, clinical data seem to suggest that there are multiple modes of therapeutic action.

Prior to the establishment of psychoanalysis proper, Freud’s and Breuer’s rudimentary theory of therapeutic action was that patients suffering from hysteria could be cured via *catharsis*—that the recovery of traumatic memories and discharge of their associated affects could remove symptoms of hysteria.

Freud went on to propose the topographic model, which emphasized: “making the unconscious conscious” as the primary mode of therapeutic action. According to this model, neurosis was attributable to repressed childhood wishes and the analyst’s role of interpreting the patient’s resistance to the analytic process. In addition, the patient’s transference(s) toward the analyst was critical to lifting the repression, providing the patient with insight, and thus effecting the cure. With his shift to the structural model, Freud amended his psychoanalytic technique to center around interpreting the patient's conflicts within and between the id, ego, superego, and reality but continued to focus on the patient's acquisition of insight as the primary mode of therapeutic action. Freud believed that the key to this process was through the patient's improving their capacity for regulating drives and strengthening ego capacities.

Ego psychologists continued to emphasize the role of interpretation and insight in their work. Some of those theorists deviated from the original Freudian theory to postulate the superego as being the site of therapeutic action or transference interpretation as being the primary mechanism for therapeutic change. Ideas about the role of the therapeutic relationship as the driver of therapeutic action also gained traction. Franz Alexander introduced the concept of patients experiencing a "corrective emotional experience" in the treatment. This was controversial at the time but paved the way for ego psychologists who focused on the "therapeutic alliance." They included both transferential and nontransferential aspects of the therapeutic relationship as a force for mutative change. Hans Loewald emphasized the role of the analyst as a "new object" for the patient, not dissimilar to a parent, who assists the patient in integrating their ego, self, and relationships with others.

These ideas informed Object Relations theorists who hypothesized that object relations are activated in the patient–therapist relationship. For them, the therapeutic action derives from using the analytic situation to better understand internalized relationship schemas, including in the "here and now" of the clinical moment, with the analyst functioning as a new object. Melanie Klein and her followers in the Object Relations school emphasized the role of the patient's transference and the analyst's countertransference in this endeavor.

The school of Self-Psychology, derived from Heinz Kohut's work with patients suffering from disorders of the self, views psychoanalysis through a developmental lens focusing especially on the consolidation of the sense of self and problems when this aspect of development goes awry. In this model of the mind, patients develop disturbances in their sense of self-cohesion, self-esteem, and capacity for investment in ideals and goals when their primary childhood caretakers did not adequately provide the child with certain crucial functions, known as *self-object functions*.

Examples of these functions include *mirroring*, which consists of validating the child’s need for exhibitionism and accomplishments, and the *idealized parental imago*, in which an idealized caregiver provides adequate strength and calm needed for the child to develop affect regulation. According to Self Psychologists, previously arrested development is reactivated in the analytic setting, with the analyst serving important *self-object functions* that the patient felt were missing in childhood so that the weakened self can be repaired. Kohut believed that central to therapeutic action is the examination of *empathic failures* in the treatment, that is, moments of disruption in the analyst–patient relationship, and an understanding of what is essential for their repair. Empathy is thought to play an important role in facilitating therapeutic change.

Modern-day psychoanalysts have also looked to neuroscience research in an attempt to elucidate potential mechanisms for therapeutic action. An fMRI study looked at the neurologic effect of verbal labeling of affective stimuli. There was increased activity in the right ventrolateral prefrontal cortex and diminished amygdala and limbic responses to negative emotional images. This indirectly suggests that there is therapeutic value in making the unconscious conscious. Identifying an unconscious affect and labeling it, making it conscious, appears to cause an appreciable change in neurologic functioning.

Similarly, psychotherapy research suggests that the strength of the therapeutic alliance, regardless of the therapeutic approach, is correlated with treatment outcome. Contemporary psychoanalysts Glen Gabbard and Drew Westen suggest that there are multiple modes of therapeutic action. These include, but are not limited to, the role of interpretation, the role of the analyst–patient relationship, and various other strategies that work together to, per Auchincloss and Samberg, “change unconscious associative neural networks that represent problematic emotions, defenses, and interpersonal patterns, as well as conscious patterns of thought, affect, and motivation.” More empirical research is needed to help characterize the mechanisms for therapeutic action.

DERIVATIVES OF PSYCHOANALYSIS

Psychodynamic Psychotherapy

Psychoanalytic theory and treatments were the primary approaches to care taught and used in mental health training for many years. By the mid-1980s, psychoanalytic approaches were increasingly being marginalized as a treatment. This is in large part due to limited research on psychoanalytic treatments, given how difficult these treatments are to manualize and systematically study. Training practitioners of psychoanalysis is a lengthy and difficult process, and criticism grew that psychoanalysis was not tailored to specific DSM disorders. In the late 1980s, Barbara Milrod et al. developed a psychodynamic formulation for panic disorder and used this to create a symptom-focused manual of psychodynamic psychotherapy that could be used in both training and research. This treatment model has since been adapted for the treatment of depression and post-traumatic stress disorder. It provided one of the first systematically studied manualized dynamic treatments that demonstrated efficacy.

In many ways, this created a foundation for other manuals to be developed using key principles of psychoanalytic treatment. These treatments could then be tailored to treat a broader range of issues since patients do not typically present with a single disorder. Luborsky developed a psychodynamic treatment manual that could be used in research based on data from the psychotherapy research project of the Menninger Foundation. A set of core conflictual relationship themes (CCRT) was developed, underlying key psychopathologic syndromes and presenting problems such as anxiety and depression.

These approaches can collectively be referred to as problem-focused psychodynamic psychotherapies (PrFPP). In PrFPP, the therapist is more active and focused and provides more psychoeducation than in traditional psychoanalytic treatments. The treatment approach aims to address core dynamics (core conflictual themes), trauma, and adverse developmental events as they contribute to various difficulties. The usual psychodynamic approaches to transference and countertransference and interpretation of conflicts and feelings in response to separation from the therapist are employed.

Psychodynamic therapy, especially as it has been operationalized in order to study and teach, can be seen as consisting of essential building blocks. Summers and Barber outline five building blocks: therapeutic alliance, core psychodynamic problem, psychodynamic formulation, strategies for facilitating change, and termination.

The *therapeutic alliance* has been the subject of much examination, as an essential common factor in therapeutic efficacy. It refers to a collaborative, here-and-now working alliance between patient and clinician. Each person has an understanding of tasks that must be performed, a sense of an attachment bond, and a notion of working towards shared goals. The therapeutic alliance is sometimes thought of as the "unobjectionable positive transference" and does not, therefore, require interpretation. Rather, it provides a fulcrum upon which productive therapeutic work takes place, which may include the elaboration of transference feelings and fantasies which are related to but can be distinguished from the so-called "real relationship" with the therapist.

The *core psychodynamic problem* refers to one of several frequently presenting problems with common underlying patterns that are typically well-suited to psychodynamic treatment. Summers and Barber refer to six core psychodynamic problems, including depression, obsessionality, fear of abandonment, low self-esteem, panic anxiety, and trauma. These are similar to the CCRT, which recognize these common patterns as the basis for the empirically supported psychodynamic therapy developed by Lester Luborsky. In the early (diagnostic and induction) phases of treatment, the forging of a working alliance and the identification of core problems will lend themselves to developing a psychodynamic formulation.

A *psychodynamic formulation* is a synthesis of data gathered from the patient encounter. It includes working hypotheses of factors leading to the development of core problems within the context of individual character vulnerability and strengths, historical events, nondynamic predisposing factors, and life stressors that help explain the patient's central conflicts. In other words, it is an integrated understanding of the patient's inner mental life, with a focus on unconscious aspects that may explain how patients came to feel and behave as they do and how their symptoms developed. A formulation will offer a degree of predictability about the expected course of treatment, including prognosis, strategies for facilitating change, probable transference manifestations, and resistance to change. The heart of the treatment, or the so-called mid-phase, is where the *working through* of these conflicts takes place through the use of psychodynamic psychotherapy techniques, including empathic listening, facilitating emotional exploration, and the sequence of clarification, confrontation, and interpretation.

The eventual *termination* of treatment is itself seen as a critical phase allowing the patient to work through feelings related to separation and autonomy as well as consolidation of gains and reflection on unresolved conflicts and disappointments.

Evidence Base for Psychodynamic Psychotherapies. Randomized controlled trials (RCTs) are considered gold-standard evidence of effectiveness in medicine. To the extent that DSM psychiatric diagnoses are able to be reliable diagnoses, despite limitations regarding heterogeneity and comorbidity, RCTs have enabled treatments to be evaluated for effectiveness in reducing symptoms as expressed descriptively in DSM. Over the last 20 years, research into the effectiveness of psychodynamic treatments has caught up to other modalities.

Leichsenring et al. conducted a systemic review finding 39 RCTs demonstrating the efficaciousness of psychodynamic therapies for major depressive disorder, social anxiety disorder, panic disorder, borderline and other personality disorders, somatoform pain disorder, and anorexia nervosa, with "possible efficaciousness" (because of only one RCT) in dysthymia, generalized anxiety disorder, and complicated grief. A recent Cochrane Report by Abbas and colleagues found psychodynamic treatment to be efficacious in common mental disorders. Effects were also found to be stable or even increased in follow-up assessments. This may point to specific strengths of psychodynamic approaches, which attempt to address psychological structure, enable personal growth, and thereby are particularly positioned to “keep working” over time.

There are important controversies, caveats, and limitations when discussing effectiveness research into psychoanalysis and psychodynamic treatments. The settings in which psychotherapy RCTs take place are generally quite different from real clinical situations, which limits their generalizability. Some argue that psychoanalytic interventions are competing in research trials where the outcome measures are not necessarily the same as the overarching goals of psychoanalysis and are more suited to those goals targeted by medical intervention. That is, we must question the real-life significance of changes on the symptom questionnaires typically used in RCTs to evaluate the outcome. A common criticism of psychotherapy research in general (inclusive of CBT, IPT, and others) is that we may find manualized treatments targeting the scales of measurement rather than the underlying pathologic process.

We can also think more broadly about what it means to invite “science” into the debate about the validity of psychoanalytic interventions. Some psychoanalysts argue that psychoanalysis should not be considered a science and it is, therefore, meaningless to evaluate it by those standards. This argument sidesteps the critical concern: if any therapy is to be a treatment for psychological problems, we have a responsibility to ensure that it works. As Alessandra Lemma puts it, “if psychoanalysis only claimed to be a philosophy, experimental validation would not be an issue.” In the chapter sections below on specific problem-focused treatments, the relevant clinical trials supportive of their efficacy will be discussed.

Psychodynamic Treatment of Panic. Panic-focused psychodynamic psychotherapy (PFPP) was developed by Milrod and colleagues as one of the first rigorously tested short-term symptom-focused psychodynamic treatments. The treatment emphasizes core psychodynamic conflicts and meanings of symptoms involved in panic disorder and has since been extended to address a range of DSM-5 anxiety disorders. It has demonstrated efficacy in RCTs as a 12-week, 24-session psychotherapy and has been studied in comparison to CBT and applied relaxation training. The psychodynamic view includes the necessity of uncovering previously unconscious meanings of symptoms in contrast with the “atheoretical” and acontextual DSM definition of “out of the blue” panic disorder phenomena. Core psychodynamic conflicts addressed in the treatment include difficulties with separation and autonomy, particularly in light of early attachment difficulties and fearful dependency on caregivers, with recognition and management of anger and perceived dangers of sexual excitement.

Defense mechanisms such as denial, reaction formation, and undoing are employed in an unconscious effort to reduce the threat from such feelings, fantasies, and conflictual wishes related to separation, anger, and the perceived threat to the sense of safety in core attachment relationships. Milrod and colleagues note that other DSM-5 anxiety disorders have psychodynamic constellations and core conflicts that overlap in varying degrees with those typically found in patients with panic disorder. They note that patients with anxiety disorders are prone to viewing themselves as inadequate and ineffectual, which reinforces dependency fears and the need to squelch feelings that they perceive would be disruptive to those essential safety relationships.

In generalized anxiety disorder, patients may maintain a state of internal hypervigilance in an effort to prevent the conscious emergence of “prohibited” fantasies that could arise if they were to relax. Worries about certain life stressors or somatic concerns can also function as defenses against unacceptable feelings and fantasies. The time-limited nature of the psychotherapy paradigm also permits panic patients to re-experience their difficulties with separation and autonomy directly with the therapist. This enables them to better understand underlying fantasies and renders them less “magical” and frightening. Patient reaction to termination is addressed for at least the final eight sessions of the treatment.

Psychodynamic Psychotherapy for Major Depressive Disorder. Dynamic Interpersonal Therapy (DIT) (not to be confused with Interpersonal Therapy or IPT developed by Klerman and Weissman) is another form of short-term psychodynamic psychotherapy. DIT is a 16-session dynamically oriented intervention informed by attachment and mentalization theory which owes much to conceptualizations of dynamic therapy advanced by Luborsky and Crits-Christoph, and Otto Kernberg. DIT focuses on how dysfunctional thoughts and beliefs in relationships cause depression. DIT aims to help patients improve their ability to cope with current attachment-related interpersonal challenges by better understanding their subjective reactions to them as threats. It makes implicit anxieties and concerns explicit and improves patients’ ability to reflect on their own and others’ thoughts and feelings. DIT helps patients understand the connection between their presenting symptoms and their relationships by identifying a core, nonconscious, repetitive pattern of relating. This pattern, termed the interpersonal affective focus, becomes the therapy’s central focus, and the therapist works on the conscious cost of such repetitive patterns. DIT primarily targets the assumed dynamic origins of depressive symptoms. Secondarily it fosters enhanced interpersonal functioning through an improved capacity to reflect on the individual’s own states of mind as well as those of others.

There are RCTs and meta-analyses examining the efficacy of a variety of psychodynamic and short-term psychodynamic therapies on efficacy for MDD. Leichsenring conducted a meta-analysis of seven studies comparing short-term dynamic therapy to CBT and found no significant differences between them in outcome scores or response rates. A 2013 meta-analysis by Barber and colleagues suggested that dynamic treatments were as therapeutically effective as alternative treatments at the end of active treatment and at short-

and long-term follow-up and were more effective than control conditions.

Transference-Focused Psychotherapy for Borderline Personality Disorder

Transference-Focused Psychotherapy (TFP) is an evidence-based treatment for borderline and narcissistic personality disorders. TFP is a modification of a psychodynamic approach to treating personality pathology with a strong emphasis on contracting, setting goals, and the employment of limit-setting. It is a particular application of psychodynamic therapy for BPO, not just the DSM-5-TR diagnosis, as will be further explained shortly. It was developed and tested as a twice-weekly psychotherapy, typically of at least a one- to two-year duration. TFP relies on an ego psychological and object relations framework initially derived from the work of Melanie Klein and most recently elaborated by Otto Kernberg. This came about in part from the challenges of treating patients that did not improve using a classical psychoanalytic frame. Their treatment required modifications based on the borderline patients' characteristically distorted senses of self and others, which would inevitably get played out in the relationship with the therapist or analyst.

In the TFP model, the diagnostic assessment focuses less on DSM descriptive criteria of personality disorders and more on the assessment of an axis called *level of personality organization*. This assessment reflects the severity of the personality pathology and aligns more with the alternate DSM-5 model for personality disorders that can be found in the DSM-5 appendix section III, written by Oldham and colleagues. One of the limitations of the current DSM diagnosis is the high degree of comorbidity among categorical diagnoses of personality disorders. There is an advantage to focusing on core features of personality pathology that cut across categorical diagnoses.

BPO is a broader construct than borderline personality disorder (BPD) proper. It reflects deficits in six intimately related domains of personality that are present in BPD but also in many of the major personality disorder diagnoses. The domains of assessment in the structural interview include reality testing, aggression, defenses, identity diffusion, object relations, and superego pathology. One can refer to the STIPO-R, a validated semistructured interview that assesses each of these domains and has been used extensively in personality disorder research.

The core therapeutic target of long-term TFP is considered to be *identity diffusion*, though the six domains are deeply interrelated. In keeping with the emphasis on identity diffusion as a core feature of personality disorder, achieving a more consolidated or integrated identity (or cohesive self-structure) is a major goal of the treatment. Identity diffusion is a lack of integration of a coherent sense of self, especially in relation to others. It is evidenced by nonreflective, contradictory, and incoherently oscillating experiences of self. Identity pathology may be reflected in the inability to pursue consistent goals in career or in relationships. Tastes, values, and opinions may be highly inconsistent and shift easily according to the social environment.

Identity diffusion is also characterized by a predominance of subjective feelings of emptiness, free-floating anxiety, and chronic dysphoria. In contrast, normal identity formation is understood as an expression of an adaptive, flexible, and stable organization of component object relations. This is a process called identity consolidation, whereas pathology of formation is understood in terms of rigid and distorted organization of component internal object relations. Patients with neurotic personality organization have a consolidated identity, while those with borderline organization do not. As discussed earlier in the section “defense and conflict,” psychological defenses can be thought of as generally unconscious strategies for managing difficult affects.

Key in the TFP model is that borderline patients use a predominance of splitting-based defense mechanisms, which means that they separate (split) idealized from persecutory components of their experience. Any individual may use a mix of characteristic defenses, some more mature, such as intellectualization, rationalization, humor, and sublimation; some less mature such as denial, projection, and projective identification. The more an individual patient relies on splitting-based mechanisms, the more severe the personality pathology. This is because splitting-based defense mechanisms may function to remove painful feelings from awareness but at the expense of a clear sense of reality, especially in the interpersonal domain.

The domain of assessment of an individual's *object relations* involves an understanding of the person's functioning in their intimate relationships. The most straightforward way of describing an *internal object relation* is as a mental representation of self-connected to a mental representation of another person linked by an affect. Representations are mental constructs that are vulnerable to alteration by the individual and thus may not reflect reality. Object relations are an internal representational system that emerges in childhood in the context of attachment relationships and is fundamental to how personality is structured. Object relations inform the quality of our relationships in adulthood.

This concept is closely linked with unconscious fantasy, as these are deeply embedded, stereotypically reproduced relationship schemas that contain aspects of an "affective narrative" of the self's relation to the other at a given moment. *Object dyads are* a single specific internal representation of self, connected to the representation of someone else (the other) linked by an affect. A negative object dyad might occur in an individual with a personality disorder whose self-representation is of being weak, vulnerable, and sick. This self-representation is connected to a representation of the other as exploitative but needed, linked by a negative affect of suspicion, rage, and helplessness. A positive object dyad would be a comfortable sense of oneself as longing and dependent but perceiving the other as caring and idealized, linked by an affect of love and sense of being cared for. These dyads are activated repeatedly in interpersonal situations and may fluctuate rapidly. This is seen in the classic "flip" in the borderline patient who may idealize their significant other or their therapist one minute and then devalue or hate them the next.

The last three domains are reality testing, aggression, and superego pathology. In regards to *reality testing*, the greater the use of splitting-based defenses, and the more distorted and stereotyped the patient's internal object relations, the more impaired the capacity for social reality testing. The hallmark distinguishing this experience from psychosis is that it depends on affect activation in an interpersonal context, rather than being a persistent state in the person. Along with affect management, reality testing is important to specifically assess how the patient manages aggression. The more severe, or low-level, borderline patients are likely to have more prominent paranoid and antisocial personality traits; their object relations are more highly suffused with aggression. The more severe the personality pathology, the more disturbed the domain of moral structure or superego formation. Low-level BPO patients, such as the severely narcissistic and antisocial, will use others for their own gain without a sense of guilt or remorse.

Much of the work in the twice-weekly treatment proceeds according to principles of dynamic therapy technique outlined throughout the chapter, including the clarification of internal conflict and the confrontation and linking of defenses to the object dyads that emerge in treatment. Patients with borderline character structure often have trouble recognizing internal conflict, instead perceiving a conflict as problem between themselves and another person, that is, they project one side of their own conflict while enacting the other. Dissociative defenses (splitting) are empathically addressed and object relationships are interpreted. Over time this leads to greater degrees of integration in the personality, improvement in consolidation of identity, and improvements in reflective functioning.

Mentalization-Based Treatment of Borderline Personality Disorder

Bateman and Fonagy have developed a manualized treatment for patients with BPD with a focus on increasing mentalization. Their approach was initially studied in a partial hospitalization setting with patients randomized to a Mentalization-Based Treatment (MBT) versus a treatment-as-usual group. Patients in the MBT group showed a statistically significant reduction in depressive symptoms, suicidal and self-injurious behaviors, number of inpatient hospitalization days, and better social and interpersonal functioning. The improvement began after 6 months of treatment and was maintained following termination of treatment at 18 months. Follow-up evaluations suggested that the patients who received the MBT intervention continued to experience significant improvement relative to the control group after the treatment had ended. MBT has since been adapted to the outpatient setting consisting of one group and one individual psychotherapy session per week and can be administered by any mental health practitioner trained in MBT and supervised by experts.

According to MBT theory, mentalization is a capacity that is typically gradually acquired in the first few years of life in the setting of secure child and caregiver attachments. It entails “making sense of the actions of oneself and others on the basis of intentional mental states, such as desires, feelings, and beliefs.” Prior to the development of capacity for mentalization, children are thought to experience the world in terms of psychic equivalence. In this experience, the internal and external worlds are equivalent. In pretend mode, the mental state exists independent of external reality but separate from the rest of the child’s mental state. A playful, attentive, and sensitive caregiver who respects the thoughts, feelings, and wishes of the child as significant, while also not being on the same order as physical reality helps the child modify the psychic equivalent and pretend modes of functioning. This style of parenting eventually allows the child to develop the capacity for organizing and recognizing their internal states as internal states.

When this process goes awry, as in the case of early neglect or physical maltreatment, the child’s internal experience remains unlabeled, chaotic, and uncontained. A child who does not receive adequate mirroring of their current state is likely to internalize the caregiver’s actual state as their own self-structure. When a caregiver exhibits primarily negative affects of rage, hatred, or fear, the child takes in the image of themself as frightening or unmanageable. The child then must externalize this painful, persecutory image in order to achieve a coherent self-representation. This sets the stage for the affect dysregulation and impaired interpersonal functioning that characterizes conditions such as BPD.

Important features of this treatment include the “mentalizing stance” of the therapist. The treatment focuses on the here and now rather than on the past. The relatively safe, secure base of the attachment relationship to the therapist provides a relational context in which it is safe to explore the mind of the other and to find one’s own mind within it.

CONCLUSION

Psychoanalysis has evolved from its early days, historically rooted in hypnosis and the treatment of conversion phenomena, to a robust and flexible discipline of depth psychological psychotherapies. Different “schools” of psychoanalysis emphasize different supraordinate motivational systems or models of mind. Ego Psychology focuses on internal conflict and compromise formation. Object Relations schools emphasize relationship schemas and paved the way for attachment research. Self-Psychology focuses on the development of self-esteem and self-cohesion and the ways that these are often disrupted in narcissistic disorders. Relational Psychoanalysis and Intersubjectivity emphasize the “two-person psychology” of the patient–therapist dyad as a corrective to outdated models, which tended to ignore the analyst’s own mind in working with their patients.

Readers of this chapter should now be familiar with the distinctive features of psychoanalysis and psychoanalytic therapy that underpin their effectiveness. These distinctive features include a focus on affect; examination of characteristic avoidances (defenses); identification of recurrent themes in thought, feeling, and behavior (enactments and repetitions); use of a developmental lens; a focus on interpersonal relationships (attachment and object relations); a focus on the therapeutic relationship in particular (transference and countertransference); and an exploration of fantasy life. This last feature, exploration of fantasy life, refers to a privileging of developmentally determined personal meanings that emerge through a sustained focus on the patient's inner world. By making such affects, avoidances, repetitions, relational themes, and fantasies conscious, their connection to symptoms can be worked through and clarified in a process that leads to enduring psychological structural change.

ACKNOWLEDGMENT

This chapter represents a major update and revision of previous chapters in this series, most recently by Dr. Silvia Karasu from the last edition. We have included a brief section of the last chapter that outlines Freud's early work.

FURTHER READINGS

Abbass AA, Kisely SR, Town JM, et al. Short-term psychodynamic psychotherapies for common mental disorders. *Cochrane Database Syst Rev*. 2014;1(7):CD004687.

Auchincloss EL, Samberg E. *Psychoanalytic Terms and Concepts*. 4th ed. Yale University Press; 2012.

Auchincloss EL. *The Psychoanalytic Model of the Mind*. American Psychiatric Association Publishing; 2015.

Barber JP, Muran JC, McCarthy KS, Keefe RJ. Research on psychodynamic therapies. In Lambert MJ, ed. *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. 6th ed. John Wiley & Sons, Inc.; 2013:443–494.

Barsness RE, ed. *Core Competencies of Relational Psychoanalysis: A Guide to Practice, Study and Research*. Routledge; 2017.

Bateman A, Fonagy P. Effectiveness of partial hospitalization in the treatment of borderline personality disorder: a randomized controlled trial. *Am J Psychiatry*. 1999;156(10):1563–1569.

Bateman AW, Fonagy P. Mentalization-based treatment of BPD. *J Pers Disord*. 2004; 18(1):36–51.

Busch FN, Milrod BL. Panic-focused psychodynamic psychotherapy–extended range. *Psychoanal Inq*. 2013;33(6):584–594.

Caligor E, Kernberg OF, Clarkin JF, Yeomans FE. *Psychodynamic Therapy for Personality Pathology: Treating Self and Interpersonal Functioning*. American Psychiatric Pub;2018.

Carlat DJ. *The Psychiatric Interview: A Practical Guide*. Lippincott Williams & Wilkins; 1999

Casement P. *Further Learning from The Patient: The Analytic Space and Process*. Routledge; 2013:339–345.

Eagle MN. *Core Concepts in Classical Psychoanalysis*. Routledge; 2018.

Freud S. *The Interpretation of Dreams*. Vol. 4. Standard Edition. Hogarth Press; 1900.

Freud S. *Instincts and their Vicissitudes*. Vol. 14. Standard Edition. Hogarth Press; 1915:117–140.

Gabbard GO, Litowitz BE, Williams P, eds. *Textbook of Psychoanalysis*. American Psychiatric Pub; 2012.

Leichsenring F. Comparative effects of short-term psychodynamic psychotherapy and cognitive-behavioral therapy in depression: a meta-analytic approach. *Clin Psychol Rev*. 2001;21(3):401–419.

Lieberman MD, Eisenberger NI, Crockett MJ, Tom SM, Pfeifer JH, Way BM. Putting feelings into words: affect labeling disrupts amygdala activity in response to affective stimuli. *Psychol Sci*. 2007;18(5):421–428.

Leichsenring F, Klein S, Salzer S. The efficacy of psychodynamic psychotherapy in specific mental disorders: a 2013 update of empirical evidence. *Contemp Psychoanal*. 2014;50(1–2):89–130.

Leichsenring F, Luyten P, Hilsenroth MJ, et al. Psychodynamic therapy meets evidence-based medicine: a systematic review using updated criteria. *Lancet Psychiatry*. 2015;2(7):648–660.

Lemma A. *Introduction to the Practice of Psychoanalytic Psychotherapy*. John Wiley & Sons; 2015.

Lemma A, Target M, Fonagy P. The development of a brief psychodynamic protocol for depression: dynamic interpersonal therapy (DIT). *Psychoanal Psychol*. 2010;24(4):329–346.

Lemma A, Target M, Fonagy P. The development of a brief psychodynamic intervention (dynamic interpersonal therapy) and its application to depression: a pilot study. *Psychiatry*. 2011;74(1):41–48

Luborsky L, Crits-Christoph P. *Understanding Transference: The Core Conflictual Relationship Theme Method*. American Psychological Association; 1998.

Luborsky L, DeRubeis RJ. The use of psychotherapy treatment manuals: a small revolution in psychotherapy research style. *Clin Psychol Rev*. 1984;4(1):5–14.

Luyten P, Blatt SJ. Psychodynamic treatment of depression. *Psychiatr Clin North Am*. 2012;35(1):111–129.

Milrod B, Leon AC, Busch F, et al. A randomized controlled clinical trial of psychoanalytic psychotherapy for panic disorder. *Am J Psychiatry*. 2007;164:265–272.

Milrod B, Chambless DL, Gallop R, et al. Psychotherapies for panic disorder: A tale of two sites. *J Clin Psychiatry*. 2016;77(7):927–935.

Mitchell SA. *Relationality: From Attachment to Intersubjectivity*. Routledge; 2000.

Muskin PR, Haase EK. Personality disorders. In: Noble J, Greene HL, eds. *Textbook of Primary Care Medicine*. 3rd ed. Mosby, Inc.; 2001:458–464.
Oldham JM. The alternative DSM-5 model for personality disorders. *World Psychiatry*. 2015;14(2):234–236.
Perry S, Cooper AM, Michels R. The psychodynamic formulation: its purpose, structure, and clinical application. *Am J Psychiatry*. 1987;144(5):543–550.
Racker H. The meanings and uses of countertransference. *Psychoanal Q*. 1957;26(3):303–357.
Sandler J, Dare C, Holder A. *The Patient and the Analyst*. Routledge; 2018:16.
Sandler J, Dare C, Holder A, Dreher AY. *The Patient and the Analyst: The Basis of Psychoanalytic Process*. International Universities Press; 1992:13–22.
Schafer R. *The Atmosphere of Safety: Freud's "Papers on Technique" (1911–1915)*. Routledge; 1983:14–25.
Schafer R. *The Analytic Attitude*. Routledge; 2018:28.
Shedler J. The efficacy of psychodynamic psychotherapy. *Am Psychol*. 2010;65(2):98–109.
Stern DN. *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*. Basic Books; 1985:138–161.
Stolorow RD, Brandchaft B, Atwood GE. *Psychoanalytic Treatment: An Intersubjective Approach*. Routledge; 1988.
Summers RF, Barber JP. *Psychodynamic Therapy: A Guide to Evidence-Based Practice*. The Guilford Press; 2013.
Vaillant GE. *The Wisdom of the Ego*. Harvard University Press; 1993.
Wallerstein RS. The psychotherapy research project of the menninger foundation: An overview. *J Consult Clin Psychol*. 1989;57(2):195–205.
Wolson P. Some pros and cons of psychoanalytic teletherapy. *Psychoanal Psychol*. 2021;38(2):109–110.

32.4 Psychoanalytic Treatment of Anxiety, Obsessive-Compulsive and Trauma-Related Disorders

ERIC M. PLAKUN, M.D.

Since the history of psychoanalysis extends back over 100 years, its conceptualization of psychiatric disorders has evolved over time, extending and correcting but remaining consistent with Sigmund Freud’s earliest formulations of the importance of the unconscious, mental representations, transference, countertransference, resistance, and defenses. The reader is referred elsewhere in this volume for a description of the history and theory of psychoanalysis. This clinically focused section endeavors to make accessible a psychoanalytic way of understanding and treating patients with anxiety and several related disorders, and to put theory and clinical practice into ordinary language.

Although psychoanalysis is a theory of the mind and of human development, as well as a specific form of treatment developed by Freud for treating “psychoneuroses,” the adjective “psychoanalytic” refers to more than these. Today its more usual form in psychiatry is as the word “psychodynamic.” This term refers both to a less intensive form of individual therapy based on psychoanalytic theory and to an entire way of looking at things that happen in the minds of (and between) patients and clinicians, but also within (and between) families, groups, and systems, and to a way of understanding, formulating, integrating, and responding to them. Psychodynamic psychiatry is the interface between psychoanalysis and psychiatry. A psychodynamic approach to anxiety and related disorders includes both a specific form of psychotherapy used to treat patients and a way of integrating the overall treatment plan of patients with these disorders. For example, and as illustrated later, this might include a specific way of conducting the psychopharmacologic treatment of anxiety and other disorders that has been called “psychodynamic psychopharmacology.”

THEORETICAL ISSUES

Two Types of Anxiety

When Freud initially studied the affective state known as anxiety in the 19th century based on his observations of himself and patients and the relevant neuroscience of his time, he concluded that anxiety was a physiologic phenomenon resulting from blocked discharge of a form of psychic energy that he referred to as libido. Thirty years later, in “Inhibitions, Symptoms and Anxiety,” Freud modified this view, as he would others, to differentiate between *primary anxiety* and *signal anxiety.*

Primary anxiety is associated with an overwhelming, inchoate experience of total disintegration and is thought to have origins in the earliest threats to the cohesiveness of an individual at or shortly after birth. It implies no capacity to think about or respond to whatever the overwhelming stimulus may be.

Signal anxiety is a warning of potential danger or of a potential threat to equilibrium. It warns that primary anxiety may not be far off and leads an individual toward thought and action. In psychodynamic thinking, signal anxiety is seen as a symptom of a deeper problem rather than necessarily as a specific focus of clinical attention, just as the dashboard “check engine” warning light in a motor vehicle is a warning to deal with some hidden but discoverable danger, not a warning to turn off the light. In both cases, there is an underlying problem that has led to the signal, and the solution lies in the direction of dealing with the underlying problem.

In psychoanalytic structural theory, structures like the ego, id, and superego are used to understand internal mental processes. Viewed through the lens of structural theory, signal anxiety is the ego’s response to sexual and aggressive drives originating in the id that are in conflict with the moral standards of the superego. Humans are never far from conflict between their wish to gratify id impulses immediately and the superego’s demands for adherence to standards of behavior that require delay in acting on impulses. Signal anxiety is ubiquitous and part of ordinary human functioning. Indeed, anxiety is the price humans pay for civilization. However, some individuals suffer from marked signal anxiety in ways distressing enough to come to clinical attention.

Psychoanalytic theory includes recognition of a predictable progression of developmental issues that emerge in the course of human life. Signal anxiety may occur in response to conflicts that emerge as the developmental sequence unfolds, including but not limited to conflicts about loss of the self, separation from an attachment figure who is loved and needed (separation anxiety), loss of the love of an attachment figure who is loved and needed, fear of massive retaliation by a powerful rival for the love of an attachment figure who is claimed by the powerful rival (castration anxiety within the Oedipus complex), loss of control, dependency, aggression, fears of sexual impulses, and others. Signal (or even primary) anxiety may also emerge from severe adverse experiences of loss, neglect, deprivation, trauma, or abuse.

Nosologic Issues

DSM-5-TR anxiety disorders are listed according to age of onset, and include separation anxiety disorder, selective mutism, specific phobia, social anxiety disorder, panic disorder, agoraphobia, generalized anxiety disorder (GAD), substance/medication-induced anxiety disorder, anxiety disorder due to another medical condition, other specific anxiety disorder, and unspecified anxiety disorder. This chapter will focus on the psychoanalytic psychotherapy of three of these: specific phobia, panic disorder, and GAD. In the transition from DSM-IV to DSM-5, two anxiety disorders were relocated to other chapters—obsessive-compulsive and related disorders like body dysmorphic disorder, and trauma and stressor-related disorders. This chapter on psychoanalytic treatment of DSM-5-TR anxiety disorders will also address obsessive-compulsive disorder (OCD) and posttraumatic stress disorder (PTSD).

Although disorders will be discussed as separate entities, there is often comorbidity among anxiety disorders and between anxiety and other disorders. Except for research purposes, psychoanalytic therapists do not generally join what is arguably a false assumption of psychiatry that most or even many patients will present with a single disorder that will respond to a specific evidence-based treatment of that disorder. Studies like the STAR*D for depression have shown that the vast majority of patients present with multiple comorbid conditions despite the “one disease-one treatment” models of psychopathology that predominate in the field.

Indeed, increasing recognition of the ubiquity of comorbidity set the stage for replicated factor analyses of patients' symptoms, leading to the elucidation of a single psychopathology or "p" factor that underlies most mental disorders. Concurrently, some cognitive behavioral therapy researchers have demonstrated the efficacy of transdiagnostic "unified protocol" therapies for anxiety disorders that target neuroticism and other underlying factors rather than disorder-specific anxiety symptoms. Psychoanalysis and psychodynamic therapy are arguably the original transdiagnostic therapies and are often indicated in the presence of complex comorbid conditions.

While psychiatry groups a series of disorders together as anxiety disorders, anxiety is often part of these disorders in different ways that are meaningful to a psychodynamic psychiatrist. Although some direct experience of a form of anxiety is part of GAD and panic disorder, other anxiety disorders may come to clinical attention because of efforts to avoid anxiety, as in agoraphobia to avoid the risk of exposure to sources of anxiety, other phobias in which anxiety is limited by limiting the risk of exposure to the phobic stimulus to a degree that may compromise functioning, or when obsessions and compulsions in OCD may serve to ward off the experience of anxiety, or because of symptoms beyond anxiety (as in avoidance, re-experiencing, negative alterations in cognition, or hyperarousal phenomena in PTSD).

These distinctions are meaningful to a psychodynamic psychiatrist. Signal anxiety is an affect, whereas phobic avoidance, obsessions and compulsions, and re-experiencing phenomena are symptoms that may be associated with anxiety (e.g., a flashback in PTSD) or with its relief (e.g., counting rituals that ward off anxiety in OCD). Psychodynamic psychiatrists are generally less interested in eliminating a symptom or affect than in helping a patient to discover and then take charge of what lies hidden beneath. What is the conflict hidden beneath the affect of generalized (signal) anxiety? What is the meaning of the symptom of phobic avoidance or of a counting ritual? Psychodynamic treatment understands symptoms and affects like signal anxiety as the visible phenomenologic tip of an iceberg that really goes much deeper than a surface view suggests

—the "check engine" light suggesting the need to look beyond the dashboard to discover the problem.

The Unconscious and Psychotherapeutic Technique

Elsewhere in this volume, the evolution of psychoanalysis is addressed, including its move from a topographical model to a structural model, to an object relations–ego psychological model, and more recently to an intersubjective-relational model. For the current chapter, the model used will be based on the notion that psychoanalytic or psychodynamic psychotherapy of any disorder presupposes that there is a relationship between the disorder and the functioning of the dynamic unconscious. This point of view is based on the notion that much of the behavior of any individual is governed by forces outside conscious awareness. Unconscious processes are active and make significant contributions to psychiatric disorders, including anxiety disorders. Although a patient may be aware of anxiety or of behavioral symptoms that cause distress, the conflicts or other meanings that underlie these may well remain out of awareness, although they are potentially discoverable —and discovery can lead to mastery. These conflicts and the issues associated with them are repressed or kept out of conscious awareness by defense mechanisms.

From this perspective, the task of psychodynamic therapy is to bring into conscious awareness and under conscious control these unconscious conflicts and other meanings that are signaled by anxiety or encoded as nonverbal communications in behavioral symptoms. Making the unconscious conscious is a significant challenge, but psychodynamic therapists understand that there is a range of ways to do this.

Psychodynamic therapists conceptualize three parts of the therapeutic relationship. The *real relationship* always exists and is based on the reality elements of two individuals involved in the therapy situation, including age, gender, ethnicity, and personality features, among others. The *transference relationship* also always exists and is based on life experiences from the patient's past that are projected onto the therapist (and the therapist's transference to the patient or countertransference based on the therapist's own life history). The third part of the therapeutic relationship, the *therapeutic alliance,* exists only if it is negotiated between therapist and patient, and this negotiation is an essential early step in treatment.

Psychodynamic therapists negotiate an agreement to explore the patient's mental life, encouraging patients to *free associate,* that is, verbalize their thoughts and feelings freely as they arise in sessions, including thoughts and feelings about the therapist. Psychodynamic therapists recognize that a patient's fantasies, dreams, fleeting thoughts, slips of speech, and habitual or unusual behaviors offer opportunities for therapist and patient to discover clues about unconscious motives, fantasies, and conflicts that can illuminate and make sense of the patient's disorder. These discoveries make therapeutic work possible in order for the patient to acknowledge, bear, and put into perspective these issues. From a psychodynamic perspective, the therapeutic alliance includes an agreement for both parties to be committed to the task of therapy, especially when this is most in question, as in periods of regression or amplification of the transference or countertransference.

Psychodynamic therapists understand that, in addition to the wish to change that brings them to treatment, patients, like other people, are fundamentally ambivalent, and thus also fear change and prefer the familiar. This is partly because patients' symptoms are not only problems but also represent the best solutions patients have been able to find for challenges they face in life. Patients may pay a high price for their solutions, but their symptoms are nevertheless solutions.

Hence, it is understood that patients will inevitably bring to the treatment a reluctance to change which is known as *resistance.* Given the request that patients free associate, resistance may take the form of conscious censorship or unconscious repression of thoughts, fantasies, or feelings. It is important to emphasize that not all parts of resistance are conscious or entirely under the patient's control. While resistance is motivated and powerful, attention to the patient's (and therapist's) associations allows it to be discovered and brought into conscious awareness. In psychodynamic therapy, bringing the patient’s resistances into conscious awareness is generally a necessary step before change can occur.

Psychodynamic therapists work with a focus on transference and countertransference as an inevitable therapeutic phenomenon. Transference and countertransference are important opportunities to learn more about what a patient is struggling with, to engage relevant issues in the here and now of the therapy relationship, to deepen the work, and to help a patient to move toward change.

In psychodynamic psychotherapy, the therapist maintains a stance of *technical neutrality* rather than being directive or self-disclosing. Technical neutrality should not be misconstrued as being cold and distant. It is in fact a warm, empathic, and nonjudgmental listening stance with evenly hovering attention and responsiveness. Patients who meet with a therapist operating from a stance of technical neutrality will react to the therapist in ways shaped substantially by past relationships that have influenced their experience of others. The psychodynamic therapist strives to accept whatever the patient brings to the sessions, including distorted perceptions of the therapist based on past relationships, that is, the transference, while striving to renounce ambition and desire about how the patient’s life should be or what choices the patient should make. The psychodynamic therapist expects that central experiences from the patient’s life history will inevitably be recreated in the transference to the therapist.

For example, it is likely a patient with a history of early sexual trauma perpetrated by a relative will come to experience the therapist at different times as like the abuser or like another relative who failed to stop the abuse or even as the victim of abuse. Psychodynamic therapists try to accept the transferences that are offered, including unpleasant negative transferences, without refusing them or trying to talk the patient out of them. This is because they recognize that this is the way a patient’s problems come to life in the therapeutic relationship. Transference is the vehicle for “exposure and response prevention” in psychodynamic therapy.

Because therapists are also human beings, they will have a range of reactions to a patient. Some of these will be the reactions anyone would have to a patient acting in that particular way, whereas others may be more uniquely related to the therapist's character as shaped by their life history.

For example, a therapist who was unable to protect their own younger sibling from abuse may respond to the aforementioned patient's transference to the therapist as an abuser with guilt and a wish to talk the patient out of the transference through reassurance about how professional and kind he is. This kind of situation, in which a patient's transference engages the therapist's character and life history in a way that creates a reciprocal countertransference that the therapist cannot endure, leading to action that refuses to accept the transference, is called an *enactment.*

Enactments are ubiquitous therapeutic phenomena. In fact, the difference between beginning and experienced therapists is better characterized by the experienced therapist's ability to detect, analyze, and utilize enactments to deepen the work, rather than entirely eliminate them. To be sure, enactments are a slippery slope, but, as in skiing on a slippery slope, the task is not to avoid sliding down at all, but rather to develop skill in maneuvering and maintaining one's balance and control while sliding down the slippery slope. It is partly because of the inevitability of countertransference enactments that it is so valuable for psychodynamic therapists to undergo psychoanalysis or psychodynamic therapy themselves so that they become aware of their blind spots, character styles, and countertransference vulnerabilities.

In this chapter, case examples are deeply disguised and/or composites of several patients. The case examples are offered to illustrate theory and technique. They are not offered as single-case proofs of the efficacy of psychodynamic psychotherapy. Although a review of the evidence base for psychodynamic psychotherapy is beyond the scope of the current chapter, there are over 300 studies demonstrating the efficacy of psychodynamic therapy, including its efficacy in treatment of specific disorders, complex comorbid disorders, and including studies that demonstrate the “noninferiority” or “equivalency” of psychodynamic therapy to other forms of therapy.

SPECIFIC PHOBIA

In phobic disorders, the affect of signal anxiety is evoked in response to a particular phobic stimulus. Avoiding the stimulus often eliminates anxiety, but the price paid to be free of anxiety may be high, seriously compromising social or role functioning. Psychodynamic therapists approach phobias by exploring the unconscious meaning of the phobic stimulus to the individual as revealed by associated feelings, memories, dreams, and fantasies. As these emerge in the transference relationship, there is an opportunity to engage them and to work through the resistances and the associated issues and conflicts. The following case example illustrates treatment of a woman with a specific phobia.

D, a woman in her 40s, had a long history of a specific phobia about driving over bridges. Over the course of years, she had been treated for this phobia with benzodiazepines and then, with good effect, exposure and response prevention. After the sudden death of her father, D’s phobia returned worse than ever, making it nearly impossible to commute to work. This time behavioral treatment did not seem to help. D was referred to a psychodynamic therapist, who worked with her to explore and elaborate the details of her phobia, which was most anxiety-producing when it involved bridges with open sides. Associated with her anxiety was a sense of dread that she would be swept off the side of the bridge and fall to her death. She could sometimes force herself to cross them, but there was a horrible swirling knot in the pit of her stomach the entire way across. Having someone else in the car helped, but only if they agreed to hold her hand while they drove across the bridge.

The therapist explored D’s relationship with her father, given that his death seemed to have made the phobia worse. D had been very attached to her father, who seemed more loving and tender toward her than her aloof mother, but, despite his warmth, her father was often absent. She had not really grieved his sudden death because of how much her widowed mother needed D to be her solid support.

At one point in a session, when the therapist asked if she had ever previously felt the kind of feeling in the pit of her stomach that was associated with driving over bridges, D recalled a memory from a family vacation when she was 13 years old. With her father, she had climbed a steep and exposed fire tower. Going up had been exciting and the view was exhilarating, but when she tried to climb down D became terrified by the steep pitch and small steps and felt the same horrible, swirling knot in the pit of her stomach that was now associated with crossing bridges. D froze and asked her father to hold her hand during the descent, but he declined, saying she had to make it down on her own, then turned away from her and descended. He did not stop when she called him. Alone at the top, D was terrified, but she had no choice but to make her way down alone. She sat on the steps and slid down one at a time, all the while terrified she would slip off a step and fall to the bottom, where her family was gathered waiting and joking about her timidity. When she got down her father laughingly commented that he knew she could get down on her own one way or another.

In the next phase of her therapy, driving over bridges became somewhat easier while therapeutic work was done around the earlier experience at the fire tower that caused the same knot in her stomach, the same fear of being swept into space and falling, and the same sense of shame and humiliation for being laughed at for not being able to do what others could do easily. These feelings were also associated with a painful sense of abandonment and ridicule from the father she needed and loved. She had never spoken to him again about the incident. A great deal emerged in the therapy about her longing for her father and the way his absence in her life and tendency to ridicule and tease her had hurt her. She felt that there was something that made her less interesting to him than her older brother, with whom he spent more time. As these feelings emerged, with them came intense grief over her father’s death. Now she would never have him with her again, and the hope of understanding why he had acted so uncaringly on the tower, or that he would apologize for it, was gone forever. As she grieved in and outside the sessions, D’s bridge phobia diminished further.

After about 6 months of weekly therapy, D felt ready to think about ending therapy, so she and her therapist set a termination date a couple of months in the future. In the termination phase of their work, D began to have more trouble commuting over bridges again. This led to an opportunity to address the ways in which the loss of her therapist was presenting D with another abandonment. Termination was another occasion in which she would feel all alone and in danger of being swept into the void. During the termination phase, abandonment and dependency issues were explored in the transference. D’s symptoms subsided again by the time of termination. At a 2-year follow-up she was free of the bridge phobia.

D's phobia about bridges was revealed to be a translation into a symptom of a humiliating memory of an experience of abandonment by her beloved but belittling father, whose death could not be fully grieved until she gained access to issues in their relationship that had been frozen in time in her symptom. In this instance, the "anxiety disorder," that is, D's phobia, and the signal anxiety associated with driving over bridges, concealed issues of unresolved grief and loss. The therapy provided an opportunity to discover this and to grieve the relevant losses.

PANIC DISORDER

Psychodynamic therapists view panic attacks as a form of signal anxiety triggered by unconscious, unresolved past conflicts often involving aggression, anger, separation, and dependency. By following affects, by listening to the patient's wishes, dreams, and fantasies, and by carefully exploring the patient's life history for therapeutic stories that illuminate the meaning of the panic attacks, as well as by attending carefully to the triggering context of panic attacks, a psychodynamic therapist joins the patient in a search for unconscious past conflicts related to panic. These conflicts are brought into awareness and under the patient's conscious control by engaging these issues as they arise in the transference relationship.

A specific form of psychodynamic therapy has been developed and studied for panic disorder. A randomized, controlled trial compared a manualized, twice-weekly, 24-session therapy called panic-focused psychoanalytic psychotherapy (PFPP) to twice-weekly, manualized, applied relaxation training. PFPP was found to be efficacious in patients with panic disorder.

PFPP is conducted in three phases. Phase 1 focuses on acute panic. The therapist works with the patient to uncover unconscious meanings of panic to achieve relief from panic attacks. PFPP therapists explore the circumstances and feelings surrounding the onset of panic, the personal meaning the patient assigns to panic symptoms, and the feelings and mental content of panic attacks. Conflicts involving separation, autonomy, and suppressed anger are frequently discovered. Within phase 1 some panic relief is typical, together with a reduction in any associated agoraphobia.

Phase 2 of PFPP focuses on lessening vulnerability to panic through the understanding and alteration of core unconscious conflicts. These are addressed as they arise in the transference and are shown to be ubiquitous in the patient's life. Responses expected in phase 2 include improved relationships, less conflicted and anxious experiences of separation, anger, and sexuality, and reduced recurrences of panic.

Phase 3 of PFPP occurs during the last third of the sessions and focuses on termination. Conflicts—for example, those about separation or anger about losing the relationship with the therapist—are re-experienced directly within the transference. Although there may be a temporary recrudescence of symptoms during termination, as if the patient were saying, "If you let the treatment end and abandon me, then all my symptoms will return," this allows consolidation of gains in the termination phase through transference work on the central separation, anger, and autonomy issues associated with panic. Patients generally emerge with a new ability to manage separation and anger and to function autonomously.

Patients with panic disorder who responded to 12 to 14 weeks of PFPP maintained their gains 12 months later at the same rate as those who responded to 12 to 14 sessions of cognitive-behavioral therapy or applied relaxation training.

GENERALIZED ANXIETY DISORDER

For a psychodynamic therapist, even GAD, in which (signal) anxiety seems to be ubiquitous and without specific content, would be approached with curiosity about and a search for whatever pieces of context may emerge in the course of exploration of a patient's fantasies, fleeting thoughts, dreams, and other clues from the unconscious that may emerge as associations to the affect of anxiety are followed and explored. Treatment is oriented toward discovering the unconscious fantasies that illuminate the conflicts associated with anxiety, some of which may involve fears of separation from attachment figures, fears of loss, fears of castration or other loss of agency, or fears of sexual impulses. The therapist approaches the patient with no preconceived notion about what these conflicts are, although with knowledge of potentially conflictual issues in human development. The following case example illustrates psychodynamic therapy with a patient with GAD.

B, a 28-year-old man with a history of GAD, had alcohol use disorder now in remission and regularly attended Alcoholics Anonymous (AA) meetings. Because of their sexual side effects, B was unwilling to take selective serotonin reuptake inhibitors (SSRI), while serotonin norepinephrine reuptake inhibitors and buspirone had been ineffective, and gabapentin was too sedating. Clonazepam was effective, but B's peers in AA urged him to give up benzodiazepines. Partly because of these pressures, B sought weekly psychodynamic therapy with a psychiatrist, who suggested that they begin tapering clonazepam in light of B's history of alcohol use disorder. B balked at this idea, worried that he would become more anxious. The therapist suggested that it might be useful to bring his anxiety to sessions if their task really was going to be to learn more about it. B agreed to the plan after some reflection.

B's anxiety increased as the dose of clonazepam decreased. He complained that his male therapist was unempathic, forcing B to suffer with anxiety while the therapist did nothing. As the treatment unfolded, the therapist learned B had been especially close to his mother, who, with B, had been the target of criticism from his often absent, short-tempered, mean-spirited, alcoholic father. B's mother had surgery and chemotherapy for breast cancer when B was 14 years old. It was shortly after this that B's anxiety symptoms began.

When clonazepam was discontinued, B experienced an increase in anxiety and had an outburst of anger at the therapist for making B suffer so much. The therapist quietly accepted B's anger at him, empathically noting that he had asked B to endure anxiety while leaving him alone and on his own most of the week. When he pointed out that he and B might want to be curious about the way he was finding in the therapist someone like his absent and sadistic father, B thought this made sense and began to trust the therapist more. Over time B stated his appreciation that the therapist could endure and understand his anger without retaliation. B also realized that the therapist was sticking to a treatment plan they had agreed to at the outset. As the alliance deepened, B struggled to put words to his experience of anxiety. B spoke more of his attachment to his mother and to the way he would cling to her to support her, pressing himself against her ample bosom, while his father would rage at them both while drunk, sometimes suggesting that B's clinging to her was unnatural and inspired by lust.

B reported a dream in one session in which he watched passively, frozen with fear and guilt and unable to move, as a man murdered and dismembered a naked woman. B's associations to the dream led to painful memories of his mother's disfiguring cancer surgery and to his guilt about his inability to stop his father from angrily criticizing her before or after the surgery. B then added there was another part of the dream he had left out because of shame; B had been sexually aroused during the dream. B suddenly reported an intrusive thought that upset him—a thought that the breast cancer had come because he had been unable to protect his mother—and because he had been aroused by her breasts. B wept for the first time in the therapy. Over time therapist and patient explored the dream and his intrusive thoughts, learning that B guiltily felt as if he were the cause of his mother's illness and disfiguring surgery. It felt as if she had become ill as retribution for his failure to protect her from his father's rages, but also as retribution for his attraction to his mother's breasts. He spoke of the way his father's drunken accusation of lust toward his mother had actually been partly right. B feared, too, that he would be disfigured because of a disease or accident, perhaps by castration, for what he had done to his mother. It was not easy for B to explore these feelings, but as he did his anxiety diminished.

By the end of B's treatment, it was evident that what had appeared initially to be GAD without context had actually been the surface presentation of unconscious psychological conflicts around attraction to his mother, guilt about her disfiguring illness, and fear of his father.

These conflicts were revealed in the course of B’s therapy, with resolution of his symptoms.

POSTTRAUMATIC STRESS DISORDER

In PTSD, a patient has experienced something horrifying that is out of the range of ordinary human experience. The experience of anxiety during the traumatic event may approach that of primary anxiety. The clinical presentation of PTSD includes more than anxiety, with a range of symptoms that carry discoverable meaning.

Freud’s early theoretical formulations dealt with the horror of actual but repressed sexual trauma as the cause of neurosis as he came across patients whose unfolding unconscious conflicts led to descriptions of childhood sexual abuse by adults. Freud later moved away from theoretical formulations based on the notion of actual seduction of children by adults to formulations based on the notion that such seductions were wished for and feared fantasies of patients. There is no simple resolution of such matters. Some patients have experienced early sexual trauma, whereas others may struggle with such “seduction” within the realm of fantasy.

Patients with PTSD have experienced actual sexual or other kinds of trauma. Psychodynamic therapists recognize that patients with PTSD are likely to struggle with difficulty trusting authority figures, given that past abusers may well have been untrustworthy authority figures. Psychodynamic therapists negotiate an alliance within which patients are encouraged to explore the details of their experience of abuse and the associated affects as they are able to bear them.

Psychodynamic therapists also anticipate that patients with PTSD will likely mobilize in their transference relationships with therapists aspects of the past abusive relationship. Thus, psychodynamic therapists understand they may be experienced within the transference as the abuser, as the helpless victim of abuse in response to the patient’s sadism, or as the passive witness or bystander, like the other parent in an abusive family who seems to the child like someone who ought to know about and put a stop to the abuse—unless they are sanctioning it. These are often particularly difficult transferences for clinicians to endure because therapists are motivated primarily by the wish to be healers and helpers. Such transferences may evoke powerful countertransference reactions in the therapist—and enactments in therapy.

The following case illustrates psychodynamic therapy with a patient with PTSD.

M, a 33-year-old woman, was employed by a domineering male boss who singled her out at work as the target of his anger, publicly humiliating her over the slightest failures. Her efforts to report him to superiors were ignored as his behavior was minimized. After a frightening sexual approach by him that she successfully rebuffed, she became symptomatic, with re-experiencing phenomena, negatively altered cognitions, hyperarousal, and intrusive memories of the abuse by him. She quit her job, felt unable to follow advice from friends to take legal action against her boss, and sought outpatient treatment. M was initially treated by a psychiatrist who prescribed medications for her symptoms and urged her to file a complaint against her boss. Although M felt unable to file a complaint because of her unhelpful past experience with her boss' supervisor, her psychiatrist urged her to do so because it might protect other employees, but she continued to refuse. When she experienced significant side effects, M stopped the SSRI he had prescribed. Her psychiatrist pressured her to try another medication, but this one, too, was associated with intolerable side effects. Unhappy, M ended her treatment, but, when symptoms persisted, began work with a new psychiatrist, although with considerable trepidation that she would have another bad treatment experience.

M began psychodynamic therapy with the new psychiatrist. She quickly felt comfortable with her new psychiatrist when he helped her see that her previous treatment experience had unwittingly recreated the experience of abuse by a domineering authority figure as the well-meaning first psychiatrist had badgered her about making a formal complaint, she did not feel ready to file and pressured her to take pills associated with intolerable side effects. The alliance was off to a good start, with a strong positive transference attachment.

Soon, though, M started to miss or come late to sessions. Although still symptomatic, she came to a session with a decision to end the therapy. Taken aback, the therapist pressured her to continue the therapy, and M reacted with anxiety, anger, and a flashback to the experience of abuse by her boss. The therapist realized he was in an enactment, noticing to himself that the abusive relationship had once again been recreated, this time between them. Now it was the new psychiatrist who was experienced as domineering and pressuring her to do something she did not want to do but had trouble resisting. Aware of the enactment, the therapist noted to the patient that they had unwittingly recreated the abusive scenario. The psychiatrist suggested that they slow down to take stock of what was happening between them that would lead to another repetition of a now-familiar pattern. The therapist noted that a choice to end the work was hers to make, but he thought that they were in the midst of important work and thought there was more to understand about her wish to flee, if she were up for the learning that might unfold. M agreed to continue the work.

With difficulty, M revealed that she had been troubled by somatic sensations she had avoided revealing. For some weeks, about as long as the lateness and missed sessions had been occurring, she had been having strange, confusing, and sometimes arousing genital sensations before and during the sessions. She realized with fear and disgust that she was feeling attracted to him. In further work it became apparent that becoming sexual with the therapist was not just something she feared, but also something she guiltily desired. This led to the patient’s revelation of an experience of date rape by an 18-year-old man when she was 14 years old. Although she had never really forgotten this incident, it was an “unthought known” in her mind, something she tried not to think about or tried to make unimportant on the rare occasions when she did acknowledge it to herself.

As the experience of date rape was explored, it emerged that it was not her first terrifying experience of being forced to do something that she did not fully want to do by a domineering man. Her father, too, had been frightening when angry, although usually loving and never sexually inappropriate. As the exploration continued, M reported discomfort and a feeling that bringing the memory of the date rape to the therapy felt like being invited to reveal more than she wanted to. When she complained that the therapist was pressing her to reveal things that made her uncomfortable, he realized that this was true and that he was again in an enactment. He apologized for pressuring her, noting he could see he was pressing her to keep going beyond the point that felt safe in a way that replicated her traumatic experiences with men. After the apology, M guiltily revealed that the experience of date rape, as terrifying as it genuinely was, was also associated with sexual pleasure. M wondered if she were to blame if that was the case. Over the remainder of the therapy, work was done exploring her conflicted and guilty sexuality as her PTSD symptoms subsided. When she felt strong enough, M decided to file a complaint against her former boss.

M presented with symptoms of PTSD in response to a recent experience at work that had important connections with an earlier instance of sexual abuse and with her relationship with her father. The first abuse experience was recreated in her work life and then, repeatedly, in enactments in therapeutic relationships, until the underlying fears and wishes were brought to consciousness, explored in the transference, and worked through over the course of several repetitions. Both therapist and patient recognized their own roles in recreating the problem.

OBSESSIVE-COMPULSIVE DISORDER

OCD is a disorder in which signal anxiety may be present (as when intrusive obsessions are associated with anxiety), although other patients with OCD ward off anxiety through compulsive rituals, expending huge amounts of energy through their symptoms to control and contain anxiety about unacceptable thoughts, risks, or impulses.

From a psychodynamic perspective, OCD is thought to have its origins in an individual’s earliest struggles to achieve a capacity for autonomous functioning in the period of toilet training. During this period, a child is encouraged, lovingly and otherwise, to give up a form of dependence on parents and to take charge of their own excretory functions. This creates the potential for a complicated struggle about control in the form of complying with or defying parental expectations. Defiance is a way of refusing to grow up, whereas complying and accepting change, though a kind of surrender, is ultimately in the service of achieving greater autonomy. A psychodynamic therapist approaches a patient with OCD by searching for hidden meaning encoded in the obsessions and compulsions, with anxiety often emerging as the meaning of the symptoms becomes apparent.

P, a passive, emotionally vacant, exceedingly polite, apologetic, and quiet man in his 30s, sought psychodynamic psychotherapy for OCD because he was having difficulty functioning at work and in relationships. P reported counting rituals and compulsive behaviors that filled hours. Among his bedtime and cleanliness rituals were a need to keep shoes shined and properly hung on shoetrees or aligned in closets, and behaviors to verify that knives were safely locked up. In sessions, he often droned on endlessly about emotionally empty details of his work life. The therapist became drowsy at one point as he struggled to remain attentive to P, who noticed this and, with uncharacteristic affect in his voice, asked, “Doctor, excuse me, but are you listening?” To this the therapist replied, “No, I guess not. Are you?” P seemed taken aback, and then characteristically apologized for having been boring.

This incident led to a direct discussion between them about the way P’s obsessive, circumstantial, and emotionally empty recounting of details was a form of resistance and one in which his therapist had joined him by becoming drowsy in the session. This was a sign their work was in trouble. Were they going to do the work of therapy together or not?

In subsequent sessions, P made efforts to speak more about the origin of his symptoms, which began with a ritual of kissing each of his parents goodnight nine times lest he be unable to sleep. On one occasion, while explaining this, P made a slip of speech about having to kiss his father nine times, instead substituting the word “kick” for “kiss.” The therapist asked P if he had noticed the slip. P insisted he could not have made such a mistake, escalated his protests about this for a minute or two as he became sadder, then began crying. While weeping, P accused the therapist of pretending there had been a slip of speech to make him look bad, and then, with an outburst of anger, recalled the hurt and injury he had felt when the therapist had become drowsy in the session.

Surprised at the intensity of his feelings, P recalled when his need to kiss his parents goodnight in a ritualized way had begun. It was after he had gotten and then lost a new puppy. His controlling and intrusive father had kicked P's beloved new puppy after the latter had a series of toileting accidents in the house. P was filled with authentic sadness as he recalled his revulsion and rage at his father and the way he had later comforted his dog when alone. P recalled having repeatedly unfolded and refolded the blade of his jack-knife to show his dog his weapon and swore to his dog that he would use it on his father if the latter ever tried to hurt him again. Soon after the kicking incident, P's father decided he had had enough of this messy puppy and sent the dog away while P was at school. P was bereft for a while but settled into an affectless, timid, apologetic, and passive way of being.

P responded to what he had learned about his compulsions by actively trying to suppress them and becoming more anxious. As his anxiety was explored, memories emerged of P's earlier struggle with his parents around toileting before the puppy came into his life, when he had received regular enemas from his father to control the frequency of bowel movements. Despite loving his father, P was also enraged at him for taking his dog away without even a chance to say goodbye and for the intrusive and terrifying experience of the enemas, while he was also furious at his mother for not stopping his father. P also felt humiliated that he had experienced this kind of intrusion into his body and that he had let his dog be given away by his father after swearing that he would protect the puppy from his father.

The meaning of P's specific rituals about knives and shoes became apparent in the course of his therapy as a result of his reaction to his therapist's entry into an enactment in becoming drowsy and as a result of a slip of speech that revealed aggression hidden beneath passivity and compliance. P's compulsive need to be sure that no knife blades were exposed related to the way such blades signified the threat of a terrifying assault on his father or an equally terrifying failure to protect the puppy that he loved but lost. Similarly, P's compulsive need to put shoes properly and safely on shoetrees was linked to an effort to put away the memory of and prevent any recurrence of his beloved lost puppy being kicked and injured by a shod foot. After these were clarified and P tried to control his rituals, signal anxiety emerged, and with it the recovery of memories of an earlier struggle with his parents about intrusive control of his toileting behavior.

PSYCHODYNAMIC PSYCHOPHARMACOLOGY IN THE TREATMENT OF ANXIETY DISORDERS

In addition to psychodynamic therapy as a treatment of anxiety disorders, a psychodynamic perspective informs psychopharmacologic treatment of anxiety as well as other disorders. If anxiety is a signal, then a psychodynamic psychiatrist will wonder what the advantage is of medicating it away rather than bringing the relevant conflicts into consciousness. The psychodynamic psychopharmacologist will hold this point of view in mind while formulating a treatment plan and selecting drugs and dosages. Using medications to achieve optimal capacity to do the work of therapy is a goal of psychodynamic psychopharmacology.

Psychodynamic psychiatrists attend carefully to the meaning of medications when they treat patients with anxiety and other disorders. They recognize that the way medications are prescribed and their intended and unintended effects will have important meaning for patients, whether or not there is concurrent therapy, and whether or not any such therapy is provided by the psychodynamic psychiatrist or another clinician. Psychiatrists practicing psychodynamic psychopharmacology pay attention not only to *what* to prescribe but also to *how* to prescribe. An example of this was part of the case of B, who presented with a GAD. His therapist's stance about reducing benzodiazepines so that B could bring his anxiety to the sessions led to the emergence of feelings that opened up an understanding of hidden issues in B’s life.

The meaning of medications may not be the same to patients as to therapists. Psychiatrists who practice psychodynamic psychopharmacology recognize that their wish to help patients with medications may not be perceived simply as help by patients. The prescription of an SSRI to reduce anxiety may be perceived by a patient who develops sexual side effects as an intentional interference with sexual functioning. Prescription of medications, then, becomes an area in which psychiatrists may step into countertransference enactments. This was illustrated in the case of M, the woman with PTSD, whose first psychiatrist’s well-intended efforts to medicate her PTSD symptoms was experienced as a replication of abuse by an intrusive and controlling boss. Her new psychiatrist’s eventual recognition of this recurring between them helped strengthen their alliance and created a precedent for examining and learning from other enactments as her treatment unfolded.

There is evidence of the power of meaning effects of medication. One double-blind placebo-controlled study suggests that a patient’s readiness to change is more predictive of response of panic disorder to a benzodiazepine than whether the patient is receiving an active drug or placebo. Similarly, there is evidence the quality of the therapeutic alliance may be more important for psychopharmacologic treatment outcome than medication effects. A large, multicenter, National Institute of Mental Health–supported study found that, in psychopharmacologic treatment, patients had the greatest reduction in symptoms when a strong alliance was paired with an active drug. Patients receiving placebo who had strong therapeutic alliances with their doctors had greater symptom reduction than patients with poor therapeutic alliances receiving an active drug. These studies support a basic premise of psychodynamic psychopharmacology that symbolic aspects of medications, that is, their meanings, may be as potent as the biologically active ingredients and should not be neglected in the care of patients.

FURTHER READINGS

Barlow DH, Farchione TJ, Bullis JR, et al. The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: a randomized clinical trial. *JAMA Psychiatry*. 2017;74(9):875–884.

Beitman BD, Beck NC, Deuser WE, Carter CS, Davidson JR, Maddock RJ. Patient stage of change predicts outcome in a panic disorder medication trial. *Anxiety*. 1994;1:64–69.

Busch FN, Milrod BL, Singer MB, Aronson AA. *Manual of Panic-Focused Psychodynamic Psychotherapy—Extended Range*. Routledge; 2012.

Caspi A, Moffitt TE. All for one and one for all: mental disorders in one dimension. *Am J Psychiatry*. 2018;175(9):831–844.

*Freud S. On the grounds for detaching a particular syndrome from neurasthenia under the description “anxiety neurosis.” In: *The Standard Edition of the Complete Works of Sigmund Freud*. Vol. 3. Hogarth Press; 1996.

*Freud S. Inhibitions, symptoms and anxiety. In: *The Standard Edition of the Complete Works of Sigmund Freud*. Vol. 20. Hogarth Press; 1996;3:87.

Kayatekin MS, Plakun EM. A view from Riggs: treatment resistance and patient authority, Paper X: from acting out to enactment in treatment resistant disorders. *J Am Acad Psychoanal Dyn Psychiatry*. 2009;37(2):365–381.

Krupnick JL, Sotsky SM, Simmens S, et al. The role of therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the national institute of mental health treatment of depression collaborative research program. *J Consult Clin Psychol*. 1996;64:532–539.

Leichsenring F, Rabung S. Effectiveness of long-term psychodynamic psychotherapy: a meta-analysis. *JAMA*. 2008;300(13):1551–1565.

McCarthy KS, Chambless DL, Solomonov N, Milrod B, Barber JP. Twelve-month outcomes following successful panic-focused psychodynamic psychotherapy, cognitive-behavioral therapy, or applied relaxation training for panic disorder. *J Clin Psychiatry*. 2018;79(5):17m11807.

Meissner WW. The concept of the therapeutic alliance. *J Am Psychoanal Assoc*. 1992;40:1059–1087.

Meissner WW. Notes on the therapeutic role of the alliance. *Psychoanal Rev*. 1999;86:1–33.

*Milrod B, Leon AC, Busch F, et al. A randomized controlled trial of psychoanalytic psychotherapy for panic disorder. *Am J Psychiatry*. 2007;164:265–272.

*Mintz D. *Psychodynamic Psychopharmacology: Caring for the Treatment-Resistant Patient*. American Psychiatric Publishing; 2022.

Plakun EM. Making the alliance and taking the transference in work with suicidal patients. *J Psychother Pract Res*. 2001;10:269–276.

Plakun EM, ed. *Treatment Resistance and Patient Authority: The Austen Riggs Reader*. Norton Professional Books; 2011.

*Plakun E. Treatment resistance and psychodynamic psychiatry: concepts psychiatry needs from psychoanalysis. *Psychodyn Psychiatry*. 2012;40(2):183–209.

Wisniewski SR, Rush AJ, Nierenberg AA, et al. Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR*D report. *Am J Psychiatry*. 2009;166:599–607.

32.5 Intensive Short-Term Psychodynamic Psychotherapies

MANUEL TRUJILLO, M.D.

The terms Short-Term Dynamic Psychotherapies (STDPs) and Intensive Short-Term Dynamic Psychotherapies (ISTDPs) denote a robust corpus of theories, clinical practices, and therapeutic techniques deriving from psychoanalysis, as discovered, developed, and formalized by Sigmund Freud at the end of the 19th century and the first third of the 20th century. Those short-term therapy systems that include the label psychodynamic in their self-definition usually wish to signal a strong adherence to the classical psychoanalytic metapsychology and a very extensive inclusion in their therapeutic armamentarium, of essential psychoanalytic techniques, such as the utilization of transference concepts and transference interpretations, and the use of the unconscious and of unconscious mental functions to understand the causal underpinnings of psychic conflict and to frame the development of the psychopathology of common psychiatric disorders. Finally, most of the currently active systems of STDPs use the interpretation of impulses, sexual and aggressive feelings, loss and abandonment, and shame and guilt as key therapeutic interventions.

The growth of psychotherapy in general and of dynamic psychotherapies derived from the psychoanalytic framework, in particular, represents a landmark achievement in the history of psychiatry. Most impressive is the capacity of the STDP field to meet the challenging standards of contemporary biomedical research paradigms by producing high-quality treatment research studies evaluating the therapy outcomes obtained in well-defined patient populations suffering from various psychiatric disorders such as anxiety, depressive disorders, personality disorders, and others that, together, represent a vast majority of the large and growing suffering and disability burden imposed by psychiatric disorders on most societies.

Probably because of these advances, and despite growing and spirited competition from many other biomedical and psychosocial interventions, more people are treated with psychodynamic approaches than ever before in history. Those who are treated achieve clinical results at least comparable to those treated with alternative psychological/psychosocial or biomedical approaches. In the cases complicated by the presence of significant personality pathology or substantive psychosocial problems, traumatic experiences, and psychosocial crises, the results achieved by psychodynamic psychotherapies are often superior to the clinical outcomes delivered through standard clinical management, including the use of psychotropic medications.

Since the methodologically sound meta-analysis conducted in the 1970s by Mary Lee Smith, Gene V. Glass, and Thomas I. Miller, which concluded that 75% of patients treated with psychotherapy were better off at termination and follow-up than the untreated controls, many rigorous research studies have demonstrated the efficacy and effectiveness of various models of the STDPs when treating such common disorders as anxiety and depression, selected personality disorders, certain posttraumatic stress syndromes, and selected eating and substance abuse disorders.

In 2013, the American Psychological Association’s Council, after an extensive and comprehensive review of psychotherapy research studies, concluded that the general or average effects of psychotherapy are widely accepted to be significant and large and are quite constant across most diagnostic conditions, with variations being more influenced by general severity than by particular diagnoses.

The council members asserted that the results of psychotherapy tend to last longer, and be less likely to require additional treatment courses, than psychopharmacological treatments. For example, in the treatment of depression and anxiety disorders, psychotherapy clients/patients acquire a variety of life skills that are used after the treatment termination, and generally may continue to improve after the termination of treatment. Within this picture of broad effectiveness for therapies of long duration, STDPs hold a position of distinction. Systematic review and meta-analyses conducted since 2010 of the effect of STDPs on patients suffering from mood disorders, anxiety disorders, somatic conditions, and personality disorders, concluded, after reviewing studies meeting high standards of quality, that the therapies studied were clinically effective, provided respectable effect sizes, and were cost-effective compared with the control conditions. The fact that these robust outcomes were also obtained working with patients suffering from personality disorders, who are not considered very responsive to other forms of psychotherapy, highlights the potential therapeutic outcome, and the value, of these ISTDPs, designed by Habib Davanloo starting in the 1970s.

A question stemming from a public mental health perspective is, can these interventions be made readily available to the large number of patients who may benefit from them?

In parallel to the growth of research in these short-term techniques, the use of planned, psychodynamically oriented, short-term modes of treatment in clinical settings has also grown. Hanna Levenson and colleagues, in a comprehensive national survey of psychiatrists, psychologists, and clinical social workers, found that psychodynamically oriented clinicians provided a sizable proportion of all brief therapy conducted nationally. Most such clinicians acknowledged treating many patients in brief therapy and would welcome supervisory and learning opportunities to enhance their technical expertise in STDP and ISTDP. The challenge to the field is to provide such learning opportunities to sufficient number of therapists to ensure the availability of such therapies to the large number of patients who, suffering from gravely disabling conditions, remain untreated, or undertreated in the community.

The current state of the field of STDPs can thus be described as healthy. The field has developed well-documented models of treatment and has proposed specific techniques that are often codified in research-based treatment manuals. Practitioners can count on numerous well-conducted research studies to support efficacy and effectiveness claims, and large numbers of practitioners offer short-term dynamic approaches nationwide to the growing number of patients who consult them for their psychiatric disorders and life crises.

HISTORY

The Early 1900s

The history of ISTDP traces back to Sigmund Freud, in that there is reason to believe that many of Freud's early treatments lasted from a few hours to a few months. Many such early treatments were in fact training analyses conducted with psychiatrists and other clinicians traveling from locations throughout Europe or the United States with stringent travel time limits. Freud's treatment of the conductor Bruno Walter that took place in 1906 lasted a few sessions, and the composer Gustav Mahler was treated in 1908 during a 4-hour walk through the woods of Vienna for an episode of psychogenic impotence.

As the time went on, various patient factors and technical developments may have contributed to the lengthening of the course of psychoanalysis. The increasing reliance on free association as a key technical ingredient diminished, by necessity, the analyst activity and may have diffused the therapeutic focus of the clinical situation. The use of the couch and the increasing commitment to allow for the slow development of a transference neurosis may have further contributed to the lengthening of the psychotherapeutic process.

The consolidation of the early psychoanalytic technical triad of reduced therapist activity, encouragement of free association, and the development of the transference neurosis has yielded substantial heuristic fruits and has undoubtedly contributed to the construction of the vast edifice of psychoanalytic metapsychology and theory of psychopathology. The price, however, may have been the reduction of therapeutic potency for those patients for whom the nature and extent of their resistances (particularly what Freud termed superego resistances) proved impervious to the technique of free association —thus Freud acknowledged growing pessimism regarding the power of free association to overcome such powerful resistances.

The Middle 1900s

Among early psychoanalytic practitioners, Sandor Ferenczi and Otto Rank deserve credit for attempting to create alternative techniques to enhance the therapeutic effect of psychoanalysis. In their book *The Development of Psychoanalysis,* they delineated a therapeutic position that included enhanced activity of the therapist, the active development of a therapeutic focus, and the setting of time limits—a triad that was to become the core of the technical offerings of later innovators such as Peter Sifneos, David Malan, and Habib Davanloo. These investigators were also aware that when patients acquired just intellectual knowledge, without integrated affects, such knowledge may serve as resistance, and they recommended an active maintenance of the focus on affect through the course of therapy to overcome such as intellectualized resistance. In the late 1930s, Franz Alexander, by then a prominent psychoanalytic leader in Chicago, focused his efforts on a pioneering research project aimed at making psychoanalytic treatment more effective, to specify the recovery techniques (process variables) necessary to achieve such a goal, and to shorten the duration of the average course of treatment. More than 600 patients were included in his studies, the findings of which were reported in 1946 in the book *Psychoanalytic Therapy* coauthored with Thomas French.

Consonant with the work of Ferenczi and Rank, Alexander also emphasized the importance of the therapist's increased activity and the maintenance of the therapeutic focus on the patient's emotional experience—thus his emphasis on what he called the "corrective emotional experience," an experience that he considered essential to any successful therapeutic venture. Such corrective emotional experience affords the patient an opportunity to relive, in the safe and supportive presence of the therapist, the original childhood traumatic experiences that contributed to the development of the patient's neurotic symptoms and maladaptive defenses. In his 1946 book, Alexander reported therapeutic results comparable to those of long-term psychotherapy and emphasized the fact that, besides symptomatic improvement, short-term therapies can also promote significant character changes. Of note, both his belief in the therapeutic power of bringing affect into the focus of treatment and in facilitating emotional experiences in the here and now of the therapy session (inevitably related to the transference) have been substantiated again and again in high-quality research studies.

The Contemporary Situation

Starting in the 1960s, the field of STDP underwent tremendous growth. Clinicians and researchers across the United States and the United Kingdom proposed diverse short-term models anchored in the psychoanalytic tradition. Authors such as David Malan in London, Peter Sifneos in Boston, Habib Davanloo in Montreal, and others generated models of psychoanalytically based psychodynamic psychotherapies applicable to increasingly broader populations of patients and—with significant boldness—provided transcripts, audio recordings, or videotapes of treatment sessions. These authors defined in very specific terms their treatment techniques and initiated basic research strategies to examine the outcomes of their treatments. Additional models were developed by clinicians and investigators such as Mann, Horowitz, Luborsky, Strupp, and the New York Beth Israel group (Winston, Trujillo, Pollack, Been, McCullough, Laikin, and others). The Beth Israel group, and other investigators, such as A. Abbass, have contributed seminal research studies demonstrating the efficacy and effectiveness of STDP as well as its broad applicability to patients suffering from complex psychiatric and personality disorders.

CLINICAL SPECTRUM OF STDPS

This chapter will describe in more detail the models proposed by Sifneos, Malan, and Davanloo as their models (which have much in common) can be ordered in a spectrum of complexity of techniques as well as of clinical severity of the pathology of patients who can benefit from them. Sifneos's short-term anxiety-provoking psychotherapy (STAPP) initiates the spectrum and occupies its relatively less complex position as it is mainly applicable to patients suffering from psychodynamically uncomplicated conditions, which can be successfully treated using basic technical interventions (development of a treatment focus, increased therapist activity, minimal to moderate resistance, and transference work). Increasing technical complexity, and patient applicability are provided by Malan's short-term model, which can be situated in the middle range of the theoretical spectrum. Davanloo's ISTDP provides more complex interventions to reach more complicated psychopathology.

Short-Term Anxiety-Provoking Psychotherapy—Peter Sifneos

Theoretical Issues. Developed by Peter Sifneos as a result of his 7-week treatment of a patient who presented to his clinic at the Massachusetts General Hospital complaining of severe anxiety, but needed to move from Boston after a few months, STAPP has found its place in the contemporary therapeutic armamentarium as a brief psychotherapy approach for patients who, having an oedipal conflict as the basis of their symptomatology, display high motivation for change, and suffer from few characterologic complications.

Typically, the oedipal dynamics lying at the core of the patient's clinical problems have to do with the conflictual aftermath of unresolved competition with the parent of the same sex for the love and affection of the parent of the opposite sex (or their surrogate). The presence of significant anxiety and the defenses deployed by the patient to counter such anxiety make it difficult for the patient to resolve such conflicts through "living," leading, as time goes on, to the development of overt clinical symptoms, subjective suffering, and/or interpersonal difficulties. Common chief presenting complaints of STAPP patients include various experiences of anxiety, alone or in conjunction with other symptoms, grief reactions, mild-to-moderate depression, somatic symptoms, and/or interpersonal difficulties unrelated to enduring character disorders.

Techniques. As Sifneos indicates, “the technical therapeutic considerations for this kind of psychotherapy are specific and clear-cut. The therapist must encourage the development of a therapeutic alliance and must create the atmosphere where learning takes place.”

Taking advantage of the prevailing positive transference feeling, staying within the specified focus, and using anxiety-provoking confrontation questions, clarifications, and interpretations, the therapist must strive to clarify with the patient the conflictual impulses and feelings underlying their psychopathology, while avoiding getting entangled in pregenital characterologic issues such as passivity, dependence, sadomasochistic or narcissistic features, and acting-out tendencies, which are often used defensively by the patient and tend to prolong treatment.

Sifneos’ prescription synthesizes three core technical dimensions: the use of positive transference, the active maintenance of the focus, and the avoidance of regression. Beyond this general prescription, Sifneos provides detailed descriptions of techniques covering the process of therapy and including the following domains: selection criteria, initial assessment, early and middle phases of therapy, and termination.

SELECTION CRITERIA. Sifneos summarizes his selection criteria for STAPP as follows:

1. Presence of a circumscribed chief complaint.
2. Evidence of a give-and-take or meaningful relationship in childhood.
3. Capacity to relate flexibly to the evaluator and to express feelings.
4. Above-average intelligence and psychological mindedness.
5. Motivation for change.

INITIAL ASSESSMENT. Beyond the completion of a classical psychiatric history and mental status examination necessary for the development of a criteria-based clinical diagnosis following the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5-TR), the initial assessment of patients selected for STAPP treatment must provide evidence of the fulfillment of the selection criteria and must lead to the joint development of a sufficiently explanatory psychodynamic formulation, a therapeutic focus, and a therapeutic contract. Such developments are usually the result of an active and dynamic interaction between the patient and the therapist, an interaction that optimally ends in the mutually agreed delineation of therapeutic foci that clearly and succinctly crystallize the specific psychological conflicts underlying the patient's psychological problems.

THE OPENING PHASE. The goal of the opening phase of treatment is to deepen the patient's commitment to work on the mutually agreed upon therapeutic contract as delineated in the initial interviews. To accomplish that goal, Sifneos recommended the use of techniques such as high therapeutic activity, the maintenance of the therapeutic focus, the consolidation and extension of the therapeutic alliance, the use of positive transference, and the avoidance of the defensive (and regressive) use of pregenital characterologic issues.

THE MIDDLE PHASE. From the technical point of view, the middle phase of treatment has as its goal a significant working through the patient's psychological conflicts as delineated early in the treatment in the dynamic formulation and therapeutic focus. Here the facilitation of transference-based "corrective emotional" experiences is central. As depicted by Alexander, "the key to shorter therapy lies in the ability to experience and control the transference." To accomplish this task, the STAPP therapist continues to use anxiety-provoking questions, confrontations, and clarifications as they link the emerging transference feelings with similar experiences in the patient's past and/or in other areas of the patient's present life. In this phase of therapy these interpretative links should lead to new learning and healthy problem solving. To facilitate the consolidation of such learning, the therapist needs to support the emerging evidence of such growth and continue to use supportive, clarifying, educational, and insight-promoting techniques.

THE TERMINATION PHASE. Because STAPP is designed as short-term therapy with open-ended termination, the therapist must rely on evidence of significant clinical improvement before approaching the termination phase. To the persisting controversy in academic psychotherapy as to who should initiate the terminations of the therapy, Sifneos provides clarifying answers: The therapist, being more able to discern objectively the state of resolution of the patient's original conflicts, should initiate termination discussions. In practice, however, more than half of Sifneos' patients spontaneously mentioned the possibility of ending the therapy as they experienced better moods and enhanced self-esteem and enjoyed their new problem-solving capacities.

Regardless of who initiates the discussion, termination is "in the air" when evidence is provided that points toward substantial differences in the attitudes and behavioral patterns present at the initiation of therapy, and new problem-solving skills are applied by the patient to the solution of their own conflicts. Abilities such as progressive independence in the pursuit of personal goals, increasing self-assurance, and improved self-esteem signal the emergence of new adaptive capacities liberated from the shadows of the unconscious conflicts that dominated the patient's life prior to the initiation of therapy.

Clinical Issues

INDICATIONS. Clinical experience with Sifneos' techniques suggests its application to the patient who fits the following DSM-5-TR diagnostic categories:

Adjustment disorders with depressive mood, with anxiety, and with mixed anxiety and depressed mood, generalized anxiety disorder, panic disorder, selective somatic symptom disorders, illness anxiety disorders, psychological *factors affecting other medical conditions, acute stress disorder, persistent depressive disorder (DSM-IV's dysthymic disorder), and major depressive disorder.*

As taught by Sifneos, beyond those diagnostic parameters, these patients should fulfill the following selection criteria: relatively circumscribed chief complaint, a history of at least one personal relationship characterized by mutuality ("give and take"), ability to interact flexibly with the therapist/evaluator, average intelligence with responsiveness to insight, and motivation for change.

LIMITATIONS. As reported by Sifneos, patients with significant character pathology, excessive dependency, multiple foci of loss, or active substance abuse may not be suitable candidates for STAPP.

COMPLICATIONS. Beyond premature abandonment of treatment, the most commonly reported complications include the development of a transference neurosis that may need additional time for resolution. Fear of termination may produce clinical regressions that can, and needs to, be handled by active interpretation within the framework of the technique. This complication is more likely to appear in those cases in which one or more early losses have not appeared in the initial psychodynamic evaluation and thus have not been incorporated in the therapeutic focus, promoting its expression through distinct behaviors in the transference, such as excessive dependency and clinging, as well as through other signs of clinical regression.

CONTRAINDICATIONS. Active dependence on alcohol or other drugs needs to be addressed independently, and a period of abstinence (30 to 90 days) should be documented prior to the initiation of STAPP. Active suicidal ideation and overwhelming interpersonal, work, or financial adversity should prompt the use of crisis intervention techniques. Once the crisis situation is resolved, the use of STAPP could be considered after a careful evaluation of the stability of the patient's psychosocial situation and the restoration of sufficient ego strength.

Chris, a 31-year-old single man, sought help for a moderate depressive episode precipitated by the loss of his relationship with his girlfriend, Joanna. She had broken off the relationship after approximately 1 year, tired of Chris' erratic work ethic and emotional instability, and discouraged by his fear of commitment to the future of their relationship. This cycle of infatuation, increasing fear of commitment, and relationship loss had become a pattern in Chris' interpersonal life. His work life was plagued with similar problems. Chris lost several jobs because of serious conflicts, and threatening confrontations with his superiors. As conflicts arose at both work and home, Chris typically suffered increasing anxiety and episodic panic attacks. In the aftermath of each loss, Chris usually was stunned, sad and often endured prolonged depressive episodes at times accompanied by suicidal ideation.

DYNAMIC FORMULATION. The development and delineation of an accurate dynamic formulation is an essential clinical activity in STDP. The challenge is to construct, based on the findings of the clinical situation, a clinical synthesis that links the patient's subjective symptoms with the underlying original childhood foci that determine or drive such symptoms. These genetic foci may be related to repressed unconscious impulses (aggressive or sexual) and feelings (loss, grief, self-doubts) that may have first occurred at key developmental points, and remain suppressed and unresolved. The end result of the dynamic formulation can be represented by a dynamic holographic (Fig. 32.5–1) representation of the pathogenic sequence that links deep genetic impulses or feelings with broad dynamic forces (repetitive experiences such as irritability or ready to explode anger or habitual maladaptive behaviors such as automatic opposition to authority figure). The final step of the hologram, links those dynamic forces with their surface, conscious conflicts or behavioral manifestations such as sullen withdrawal and passivity or self-destructive angry acting out.

For Chris, in the foregoing case, the dynamic hologram can be represented as follows:

Life problems: Recurrent episodes of anxiety and depression; work problems; unstable interpersonal relationships; conflict with authority figures; antagonism toward, and emotional distance from, his father, brother, and other male friends; and fears of heterosexual intimacy and of commitment.

Dynamic forces: Ongoing hostility and envy toward males, authority figures, and successful people, and compulsive and possessive craving and seeking female love objects while simultaneously having great difficulties considering fulfilling, or even recognizing their independent needs.

Genetic pathogenic foci: Unconscious loss of maternal objects precipitated by birth of brother at age 2 years; uncontrolled grief for that loss with compulsive drive to experience child-like possession of love objects; and compulsive hostility toward others perceived as rivals.

PROCESS AND GOALS OF TREATMENT. The typical clinical process in STAPP can be summarized as follows:

Objective: Resolution of Oedipal conflicts

Focus/Foci: Triangular/Oedipal

Duration: From 10 to 20 sessions, lasting from 50 to 60 minutes, occurring once a week

Termination: Open

After the evaluation phase, the typical therapy process proceeds on the strength of the therapeutic alliance optimally established during the early sessions. The efforts of the therapist in the initial phase aim to enrich the dynamic formulation through an active inquiry around the foci established in the initial formulation. This active inquiry, in the case of Chris, yielded additional confirmation of the persistence of repressed sexual feelings toward his mother and the presence of hostile feelings toward all rivals for his mother's affection. A memory suffused with very visceral feelings emerged in this phase as a result of the therapist's active inquiry. In this memory, Chris saw himself in his mother's arms in a dark room. He remembered vividly the intense pleasure of the contact with the warm skin of his mother, the texture of her clothes, and the smell of her perfume. While narrating this memory to the therapist, Chris was so absorbed in the experience that he blushed intensely while describing it to the therapist. He first giggled with delight and amazement, a look of disbelief on his face, seeking the complicity of the therapist in this timeless experience of imaginary possession, and enjoyment of his mother's body. He next described the painful termination of this moment of pleasure by his father's sudden, angry forceful, and disruptive opening of the closet door, and intense pain of the flood of light that disturbed his pleasurable absorption. After this sequence, Chris felt sad and open himself to the experience of grief at this loss of the intense and exclusive bond with his mother; a loss that in his childhood he had also after his brother's birth.

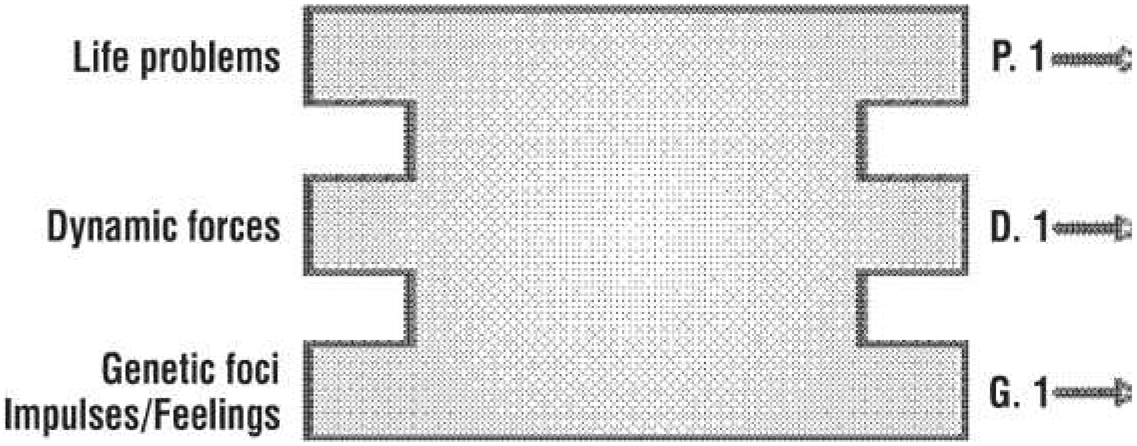

FIGURE 32.5–1. Dynamic hologram.

Feeling safe in the therapeutic situation, he was able to experience again and again, the sadness, loneliness, sense of abandonment, and rage that had been all too familiar through his childhood and that recurred as well in his present life when his romantic attachments were threatened or actually lost. The affective link between this childhood experience and his intimacy problems in the present became viscerally obvious to Chris, and the acceptance of this link enhanced his capacity to work through this essential component of his pathology. A parallel conflict appeared in the transference as the patient resented the “intrusion” of the inquiring therapist into the zealously guarded privacy of this primal fantasy of maternal possession.

OUTCOMES. In addition to his role as a clinical pioneer, Sifneos deserves additional credit for his commitment to make his work visible through the pioneering utilization of the rigorous outcome evaluation of his work. In two studies, Sifneos reported respectable clinical outcomes, which included moderate symptomatic improvement, significant enhancement of self-esteem, and positive resolution of the crisis situation that initially brought the patients into treatment.

Beyond this early evaluation, Sifneos’ approach has been studied in a modified form by many clinical researchers. The work of Arnold Winston and colleagues represents an excellent investigational effort. In these studies, two forms of brief therapy based on procedures specified in therapy manuals were used. One of them, labeled STDP was based on principles developed by Davanloo. The second one—brief adaptive psychotherapy (BAP)—used a composite of brief therapy techniques and was heavily influenced by the work of Sifneos. The results of both of Winston’s studies indicate that both treatment conditions were equally effective and obtained clinical outcomes that were significantly better than those achieved by the patients who remained on the waiting-list control for the duration of the study. Of note, the Sifneos-derived BAP technique achieved better outcomes than the STDP in measures of anxiety and phobic anxiety, whereas STDP showed significant advantage over BAP on depression subscales.

Short-Term Dynamic Psychotherapy—David Malan

Theoretical Issues. The work of David Malan flows from the interest of Michael Balint and his group at the Tavistock Clinic in

London, in expanding the boundaries of psychoanalysis by training general practitioners in basic psychotherapy techniques, and in developing brief psychotherapy techniques to enrich the clinical armamentarium of experienced psychoanalytically oriented psychotherapists.

Anchored in classic psychoanalytic metapsychology and theory of technique, Malan's brilliant contributions to STDP involve both technical aspects and research insights. His contributions to technique include the focusing of the therapist's interventions on the triad of impulses, defenses, and anxiety (triangle of conflict) as it plays out in relation to the therapist and to key figures in the patient's current life and the past. In this regard, Malan advocates for a comprehensive exploration of all components of the triangle of conflict and recommends a thorough working of the impulse component of the past conflict.

His contributions to research are many. By keeping the research enterprise very close to the clinical ground and to the way in which psychodynamic clinicians actually think and behave, he has fostered the meaningful generalizability and reliability of his findings. His attempts to make outcome predictions from baseline clinical data, place him as a pioneer in research techniques that later grew exponentially in the 1970s and 1980s. A key finding of his studies is that the higher the number of interpretations linking the current behavior of the patient with the past and the transference (TCP), the better is the outcome. This finding has stood the test of time and has been replicated in numerous other psychodynamic psychotherapy studies.

Techniques. The treatment process in Malan's STDP can be summarized in the following manner:

OBJECTIVE. To work through the patient's dynamic problems by linking the patient's impulses and feelings, defenses, and anxiety in relation to each relevant psychodynamic focus, while promoting "corrective emotional experiences" with the therapist in the transference. The typical outcome of such working through is increased tolerance of interpersonal conflict (ego strength), enhanced assertion, and substantial lessening of dependence. Negative feelings and pessimism are also reduced.

FOCI. Unresolved interpersonal conflicts and suppressed impulses and feelings, including feelings of abandonment and loss, links between traumatic experiences in the past, congruent situations in the present, and lived perceptions and emotions in the transference. A patient with unresolved abandonment and separation fears may come to understand the link between this fear (based on childhood abandonment experiences) and her present day compulsive neediness and clinginess. Supported by her positive transference experience, such patient may feel confident enough to endure fears of loss and initiate assertive independence in key areas.

SELECTION CRITERIA. Motivation for psychological change, response to early tentative interpretations (trial interpretations), and ability to tolerate and work with intense feelings.

EXCLUSION CRITERIA. Inability to create emotional contact between therapist and patient, extremely rigid defenses, lack of motivation for dynamic change, and the presence of risk factors for the development of intense dependency.

Therapists working within Malan’s framework used all psychoanalytic technical interventions, organized around two conceptual triangles labeled the triangle of conflict and the triangle of the person (Figs. 32.5–2 and 32.5–3). The objective of treatment is the repeated, and affect-laden, clarification of defense in the triangle of conflict, which permits access to the experience of underlying impulses and feelings when anxiety is appropriately modulated. Increased awareness of formerly repressed impulses and feelings permits their elaboration and working through as they appear in the patient’s current life, the past, and the transference. As noted previously, Malan observed that the higher the number of links between experiences in the current life of the patient, the past, and the transference, the better is the outcome.

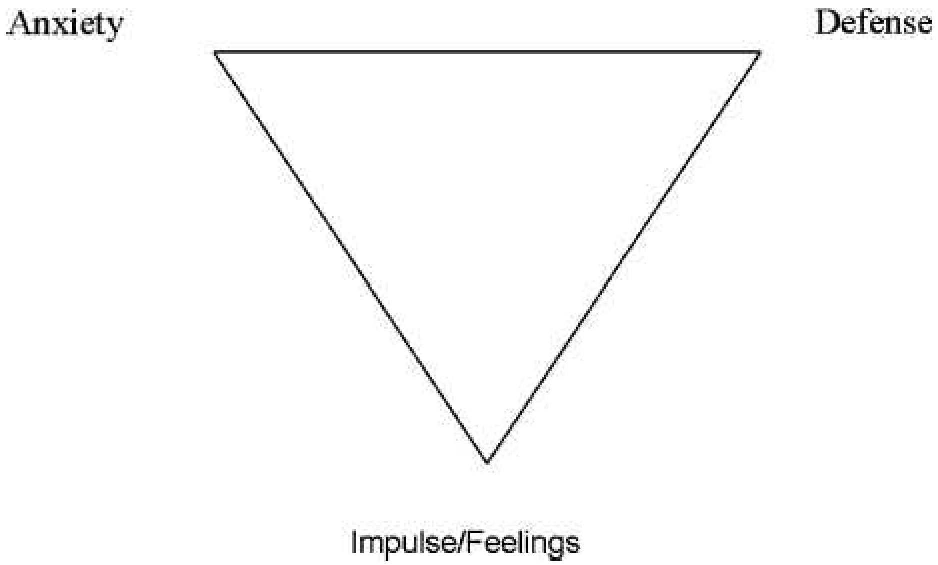

FIGURE 32.5–2. Triangle of conflict.

The techniques used by Malan are the ones used in standard analytically oriented psychotherapy. In addition to the clinical material spontaneously communicated by the patient or developed through the therapeutic interaction, dreams, fantasies, slips, and associations are incorporated into the clinical work.

The clinical material is kept manageable by the efforts of the therapist to keep the therapy in focus, an objective that is achieved by following Malan's recommendation to use "selective interpretation, selective attention, and selective neglect." The therapy is terminated when there is evidence of sufficient symptomatic improvement. In most cases reported by Malan, patients also achieved substantial improvements in long-standing characteriologic problems.

Clinical Issues

INDICATIONS. Clinical experience with Malan's techniques suggests its application to the patient who fits the following DSM-5-TR diagnostic categories: *Adjustment disorders with depressive mood, with anxiety, and with mixed anxiety and depressed mood, generalized anxiety disorder, panic disorder, selective somatic symptom disorders, illness anxiety disorders, psychological factors* *affecting other medical conditions, acute stress disorder. Selective separation anxiety disorders. Persistent depressive disorder, dysthymic disorder and major depressive disorder. Selected personality disorders such as avoidant and dependent personality disorders*.

LIMITATIONS. The range of techniques proposed by Malan can provide treatment to patients suffering from oedipal psychopathology complicated by loss and abandonment, as long as the patients have not developed too intense, character-based interpersonal defenses or express regressive trends such as diffuse pervasive anxiety, paralyzing helplessness, extensive somatization, and the like. In his series of studies, Malan and his group predicted poor outcome in patients who failed to make contact with the evaluator, remained poorly motivated after the initial interviews, or developed intense transference-based dependency.

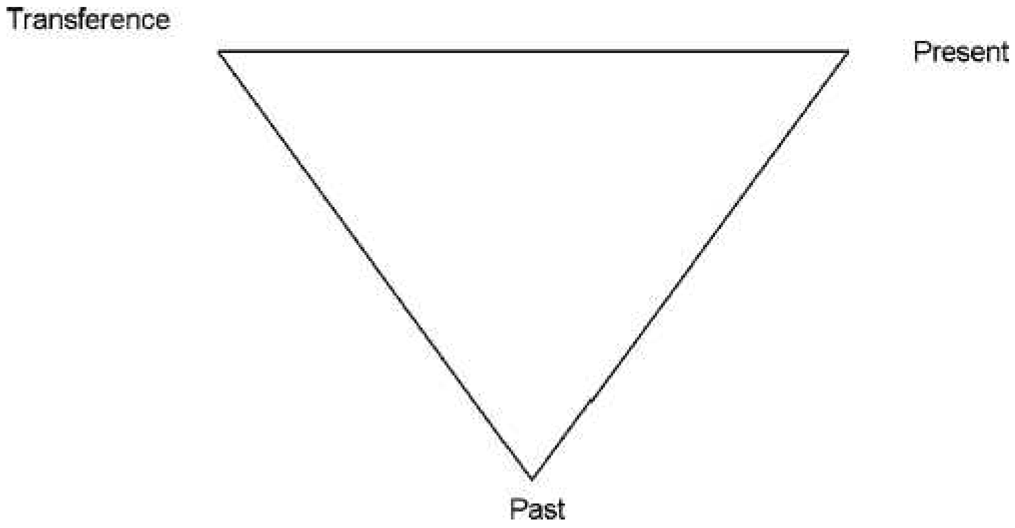

FIGURE 32.5–3. Triangle of the person.

COMPLICATIONS. The most common complication reported in Malan's studies relates to the inability to terminate the therapy within the 40-session upper limit. In Malan's study most of these patients went on to medium-length and long-term psychoanalytic psychotherapies but ultimately achieved delayed, but respectable, clinical outcomes. Of the 30 patients treated in Malan's second study, only 1 was reported to be clinically worse at follow-up.

CONTRAINDICATIONS. As reported in Malan's studies, there were certain absolute contraindications that eliminated patients from being considered a candidate for this treatment. They included a recent history of suicide attempts, active substance abuse, history of more than one course of electroconvulsive therapy (ECT), chronic incapacitating obsessional symptoms, and gross destructive or self-destructive acting out. Whether selected samples of these originally excluded patients would be good candidates for therapy should they respond positively to trial therapy needs to be tested in additional studies.

Ana, a divorced 60-year-old woman, sought psychiatric help following a severe depressive episode lasting several months. This episode, which was one of many in her life, was especially severe in terms of loss of energy, interest, and motivation, as well as in terms of the intensity of her sadness and her wish to die. Only her profound religious convictions protected her from acting on these wishes. Ana had lost a lot of weight, had trouble sleeping, experienced many nightmares, and had difficulty with concentration. She was plagued by pervasive feelings of hatred for her mother, who was by now very old, ill, and dependent on Ana, who was unable to forgive her for having abandoned her in an orphanage when she was 5 or 6 years of age.

After an extensive assessment, the dynamic formulation of Ana's problem was represented as follows:

Life problems: Recurrent depressive episodes plagued by feelings of guilt and self-reproach; problems with men involving choosing partners who are commonly cold, distant, or otherwise unavailable; involuntary and painful emotional distance from her children, friends, and other close relationships; and unproductive and unrewarding work-life, despite considerable intellectual gifts.

Dynamics: Ambivalent relationship with her mother, whom she blames for most of the tragedies of her life; guilt and need for punishment in relation to her unrelenting hatred for her mother; and pathologic grief reaction for the loss of an idealized and more optimal relationship with her mother, the one she remembers she had prior to her orphanage placement. From this focus there flows a melancholic conviction of the inevitable failure of human relationships.

Pathogenic foci: Grief and inability to mourn the loss of her mother after she was placed in the orphanage, with attendant rage and guilt; a second pathogenic focus results from her pathologic grief for the loss of her father, who, because of severe alcoholism, had abandoned the family; a move that caused the mother to place her children in an orphanage in order to be able to work and ultimately recover their care. Unconsciously, Ana had always blamed her mother for the family catastrophe, thus “protecting” an idealized view of her father, to whom she was profoundly attached.

For Ana, the initial phase of treatment focused on facilitating, through quite intense resistance, the actual affective experience of her destructive anger toward her mother, the work through of which, made possible the appearance of a modicum of empathy with her mother’s painful life situation around the time she placed Ana and her sisters in the orphanage. Next, the therapy focused on Ana’s father. Deep feelings of idealization, disappointment, anger, and grief were experienced with increasing clarity and intensity, frequently via displaced feelings in the transference and almost always after overcoming considerable resistance. The last phase of treatment permitted the gradual development of realistic feelings of empathy and appreciation for her mother, experienced now without the cloud of anger or the numbed emotional distancing which had accompanied such feelings in the past. The reawakening within Ana of feelings of joy and hope, as well as professional ambition also needed support and encouragement, as such “luxuries” had been blocked in her development by her feelings of guilt.

OUTCOMES. Although Malan’s carefully designed and elegant research studies would not conform to today’s requirements for research in STDP, they contributed rich insights to the field. Malan reported on two different naturalistic, noncontrolled samples. The results of the initial study of 21 patients were reported in Malan’s 1963 book, *A Study of Brief Psychotherapy.* A replication study involving 39 additional patients was reported in 1976. Both studies generated a wealth of therapeutic insights, which have guided the field since their publication. His findings confirmed that neither severity of baselines symptoms nor their relative chronicity predicted clinical outcome at termination, whereas the presence of motivation for insight was highly predictive of success. In terms of therapeutic process, use of transference and transference–past-present links was also found to be highly predictive of positive outcome.

Malan’s proposals, both in direct or modified forms, were later thoroughly tested in efficacy studies. Leichsenring’s meta-analyses included several studies using Malan’s STDP as one of the study groups. Probably the most rigorous test of Malan-derived STDP principles and techniques is represented by a study by Svartberg and colleagues, which showed statistically significant improvements to patients with Cluster C personality disorder suffering from anxiety and depressive disorders. This study compared STDP and cognitive therapy and found changes in symptom distress after active treatment termination which were statistically significant only for the STDP group.

Intensive Short-Term Dynamic Psychotherapy—Habib Davanloo

Theoretical Issues. Davanloo developed his technique of ISTDP in a series of clinical research projects at the Department of Psychiatry of the Montreal General Hospital of McGill University. The initial studies (conducted between 1963 and 1974) established that excellent results could be achieved working largely through interpretations on a population of highly responsive and motivated patients suffering from relatively simple psychodynamic problems. A major naturalistic research project was then designed to broaden the base of eligible patients by testing the effect of techniques designed to deal with increasing resistance. This study established for Davanloo that the combination of transference focus and special handling of resistances could be used to achieve major breakthroughs into the patient’s unconscious, leading all the way to major unlocking of the unconscious, a process found to have great therapeutic impact.

These discoveries led to the development and consolidation of his technique of ISTDP and to a new theoretical understanding of the function of the human unconscious and of the role of transference and the therapeutic alliance in the therapy of highly resistant patients. For Davanloo the key therapeutic force involves the mobilization of the therapeutic alliance through a highly specialized form of interviewing designed to deal with the patient's overwhelming resistance. The second major area of scientific concern is Davanloo's views of the centrality of the process of attachment for human development and of the importance of its disruptions for psychopathology. For Davanloo, it is the severe traumatization of core attachment bonds during development what causes psychic traumas, the pain of trauma, the mobilization of primitive rage and murderous impulses, and the guilt, grief, and character distortions that one reliably finds at the core of the neurotic processes of complicated psychoneurotic patients.

Techniques. The core of Davanloo's techniques flows from his research on the twin dynamics of transference and resistance. The central aim of Davanloo's system of ISTDP is the rapid reduction or removal of resistance through the careful application of a system of planned interventions that end—when successful—in direct access to the unconscious as a result of the breakdown of the major resistance and defense systems with the attendant mobilization of the unconscious therapeutic alliance.

This planned sequence of steps to deal with resistance is what Davanloo called the central dynamic sequence. The eight technical steps of this sequence are as follows:

1. Phase of inquiry into the patient's clinical and interpersonal difficulties.
2. Pressure over the avoided or repressed feelings.
3. Mild to intense challenge to the patient's resistance: Davanloo's "head-on-collision."
4. Detection and challenge to the transference resistance.
5. Direct access to the unconscious; usually in the form of viscerally intense, sudden eruption of vividly experienced primal and primitive rage, sadness, grief, and the dominant ascent of a deep therapeutic alliance.

6. Systematic analysis of the transference.
7. Dynamic exploration of the core unconscious configuration of conflict.
8. Consolidation and preparation for working through of such unconscious configuration.

The end product of this process, when successful, involves increasing visceral experience and awareness of intense murderous rage and other deep feelings, which usually appear first in the transference and which then are focused on key figures in the patient’s development and childhood trauma. As these feelings burst into open, they typically bring forth deep waves of sadness and guilt, both of which are experienced deeply and viscerally in a manner that promotes the resolution of the pathogenic foci. Following the completion of the tasks of the initial interview, the therapy enters the working-through phase. Key to this phase is the ongoing affective experience of personal reactions and conflicts. The tasks of the working-through phase include the continuation of the process of resolution of previously repressed, extensive, murderous impulses and feelings.

The transformation of pathologic grief reactions into acute grief processes facilitates their resolution and contributes greatly to clinical improvement.

The process of termination is initiated when there is evidence of symptomatic resolution and of character and functional changes. In complicated cases, mourning the termination of the therapeutic relationship becomes a significant focus and the key to the prevention of repressive dependency.

The working-through phase aims at the systematic resolution of the multiple foci of conflictual painful feeling accessed in the initial interview. The following principles apply to the working-through process:

The therapist should place emphasis on the patient’s actual experience of their feelings. Feelings such as intense grief reactions, sadness, and rage followed by guilt and remorse repeatedly appear in the psychodynamic focus, facilitating their working through and resolution and promoting healthy adaptive growth.

Often, as a result of the breakdown of resistance systems, the patient experiences vivid reactivation of memories and dreams, which enrich the working-through process, as long as the therapist holds the focus on the direct affective experience of the recovered memories and dreams.

Clinical Issues

INDICATIONS. *Adjustment disorders with depressive mood, with anxiety, and with mixed anxiety and depressed mood, generalized anxiety disorder, panic disorder, selective somatic symptom disorders, illness anxiety disorders, psychological factors affecting other medical conditions, acute stress disorder. Selective separation anxiety disorders.*

Persistent depressive disorder, dysthymic disorder, and major depressive disorder. Treatment-resistant anxiety and depressive disorders and chronic pain syndromes.

Major dissociative disorder (depersonalization); social anxiety disorder; obsessive-compulsive disorder; somatic symptom disorder, conversion disorder and hypochondria.

Selected personality disorders such as avoidant and dependent, obsessive compulsive and histrionic personality disorders.

Davanloo, further characterized the core psychodynamic problems presented by patients eligible for treatment through these techniques, on the basis of a descriptive sample of close to 1,000 patients. The large majority of these patients suffered from diffuse and pervasive symptoms such as anxiety, depression, and widespread somatic complaints. They displayed severe maladaptive personality and character problems, leading to significant interpersonal and work problems and to intensely self-sabotaging and self-destructive behaviors. They had clinical histories of overt (and covert) severely traumatic experiences and presented to treatment with high degrees of major resistance. The core psychopathology of these patients was dominated by the presence of highly primitive unconscious rage often directed toward multiple figures of their past who participated in their highly traumatic development or otherwise induced repeated experiences of abandonment and loss. The widespread experience of unconscious rage often led to profound feelings of guilt and grief and commonly promoted masochistic self-destructive lifestyles.

LIMITATIONS. These techniques can provide effective treatment to a large majority of patients suffering from what had been classically called severe psychoneurotic disorder and character disorders. These patients are diagnosed today with what was formerly called Axis I disorders, such as anxiety (including panic) and major depressive disorders, and with personality (predominantly Cluster C) disorders. They represent large cohorts from among the patients who seek treatment in outpatient settings and in private practice and present a serious public mental health burden because of the treatment-resistant nature of their problems and their high utilization of health and mental health resources.

As these techniques are designed to deal with impulse (rage)-based dynamics, they may reach their limits in the growing population of patients suffering from primary self-disorders such as narcissistic and borderline syndromes (self-disorders).

The treatment of self-disordered patients needs to go beyond the resolution of their unconscious rage and secondary guilt and grief. Treatment needs to reach two key components of their disorder: their severely devalued self-images and self-representations, and, more importantly, the healthy and vital self-representations that often lie dormant and repressed under the weight of more consciously experienced negative views of self. Such healthy self-representations will make possible the healthy ego development and personal growth that can be induced by treatment.

Techniques for the treatment of these disorders have been proposed, and they are reviewed later in this section.

COMPLICATIONS. The most common complication of ISTDP consists of the development of a protracted or stagnant therapeutic process as a function of the therapist’s inability to break through the highly elaborate structure of resistances presented by the patient at baseline. Transference-based negative therapeutic reactions might ensue and result in a temporary worsening of the presenting symptomatic picture.

CONTRAINDICATIONS. These techniques are not indicated for the treatment of psychiatric disorders such as schizophrenia, substance abuse disorders, antisocial personality disorders, or life-threatening psychosomatic conditions.

The application of ISTDP for psychosomatic and borderline disorders requires modifications of the technique. Selected problems of patients with bipolar disorder (certain maladaptive behaviors, interpersonal problems, and pathologic grief reactions) can be treated with these techniques with positive effects on the patients’ overall affective functioning.

Davanloo provided numerous detailed case reports that give step-by-step descriptions of the different phases of his therapeutic approach.

The following case was treated following Davanloo’s system of ISTDP.

At the time of the first interviews, John was a 35-year-old professional seeking help because of a worsening of lifelong feelings of depression, hopelessness, and impotence. He had an almost delusional conviction that he was destined to catastrophic failure and that he was helpless to change the course of such a fate. A very brilliant man, he struggled to hold back his professional ambition for fear of being perceived as arrogant. He suffered from pervasive feelings of physical and intellectual inferiority and experienced severe loneliness, with the conviction that he would never be able to marry. He believed that no worthy woman would ever show an interest in him.

John was also plagued by episodes of severe anxiety, which became intensified in conflict situations. He feared any interpersonal conflict that could lead to angry words or deeds. In his struggle to suppress any remnant of anger, he had adopted an ego-syntonic, submissive, and self-effacing character structure.

The second of two siblings raised in a middle-class family, John remembered a lifetime of submission to the capricious and sadistic acts of his older brother, who tyrannized him daily through angry words and physical violence. Neither his mother (perceived as distant, sad, and helpless) nor his authoritarian and emotionally removed father ever intervened effectively to curb his brother's relentless aggression. As early as he could remember, John's subjective experience was suffused with terror. Usually, he was afraid of triggering his brother's brutal temper. Occasionally, he feared losing control of his intense destructive impulses toward his brother. Such impulses were deeply repressed and denied.

The dynamic formulation of John's problems can be represented as follows:

Life problems: Ongoing feelings of depression, sadness, hopelessness, and loss of motivation and interest; deep loss of self-esteem, with feelings of physical and intellectual inferiority;

a resignation to a life of limited relational and intellectual ambitions; fear of rejection by women and fear and avoidance of falling in love due to the subjective certainty of rejection; and subjective sense of helplessness, worthlessness, and impotence.

Dynamic forces: Struggle to suppress even minimal experience of anger in any context; suppression of assertive wishes and of ambitious hopes and aspirations, and intense guilt and self-reproaches when hoping for assertion and success.

Pathogenic foci: Powerful deep unconscious destructive feelings (including death wishes and murderous impulses) toward his older brother; and parallel destructive impulses toward John's father for his lack of emotional involvement and toward his mother for her failure to prevent the abuse perpetrated by John's older brother. These foci were sharply portrayed via a particularly violent dream that occurred in the early phase of treatment. In this dream "a serial killer, in a murderous rampage stabbed and killed several members of the same family."

GOALS OF TREATMENT. Davanloo established ambitious goals for his system of ISTDP. He asserted that he aimed at a total resolution of the patient's core psychopathology, leading to a clinical situation in which the patient was symptom free and showed demonstrable structural, character, and behavior changes.

There should be visible and stable changes in the way in which the patient is related to others and used their capacity to love and work. A key component of such progress was the patient's often new ability to be in touch with and appropriately express deep feelings.

OUTCOMES AND RESEARCH. Very few studies have been conducted rigorously to assess the efficacy and effectiveness of Davanloo's STDP. In fact, in several meta-analyses of brief dynamic therapy studies conducted between 1978 and 2004, only the studies presented by Winston and colleagues purported to test the efficacy of Davanloo's STDP when compared with an alternative dynamic therapy and a wait-list control. These studies, however, have been widely disseminated, are cited by all selective meta-analytic reviews, and are often praised for their careful scientific design and clinical implementation.

In Winston's 1991 study, 32 patients with diagnoses of personality disorder were randomly assigned to two treatment conditions, and an additional 17 patients were assigned to a wait-list control group. Both therapy conditions were brief dynamic therapies, one being an STDP based on the principles of Davanloo and a second (BAP) drawing heavily on the work Sifneos. At the end of the study, the two therapy groups obtained similar outcomes (significantly better than the wait-list controls) but showed differences in outcomes in symptomatic subscales. Davanloo-based STDP obtained better outcomes on ratings of depression, whereas the Sifneos-based BAP offered better results in ratings of anxiety. Both therapy conditions demonstrated robust and clinically meaningful effect sizes (0.70 to 1.35). An expansion of the earlier study to include 81 patients reported in 1994 obtained similar results.

In a recent report by Allan Abbass and colleagues, 27 patients diagnosed with DSM-IV personality disorders (borderline, obsessive-compulsive, and avoidant personality disorders constituted 65% of the sample) were randomized for treatment with ISTDP or to a minimal contact, delayed treatment, control condition.

The ISTDP treatment was provided as outlined by Davanloo, and treatment adherence was carefully controlled. After an average of 27.7 (range 2 to 64) sessions, the ISTDP patients did significantly better than controls in symptomatic and functional measures, and these gains persisted in long-term follow-up. Of note, the treatment sample reached the normal range on the individual symptomatic scales used (Brief Symptom Inventory and Inventory of Interpersonal Problems), had an 83% reduction in personality disorder diagnosis, and achieved large reductions in the use of medications. Just as important, these patients achieved a significant reduction in functional disability.

EMERGING SHORT-TERM DYNAMIC PSYCHOTHERAPIES

Accelerated Experiential Dynamic Psychotherapy—Diana Fosha

Theoretical Issues. Building on the work of Sifneos, Malan, Davanloo, and others, Diana Fosha proposed a new model of therapy, which she called accelerated experiential dynamic psychotherapy (AEDP). This approach integrates the technical contributions of its predecessors under the aegis of an affect-centered model of relational psychodynamics; a dynamic which Fosha has later anchored in the neurobiology of attachment and affect.

Whereas the reduction of resistance by breaking through the defense of patients constituted a central tenet of Davanloo's ISTDP, Fosha proposed an alternative route to deal with such defenses: fostering the experience of deep affects and the development of an emotional bond between patient and therapist. Such bonds, by promoting feelings of safety and connectedness in the here and now, are hypothesized ultimately to reduce the need for resistance and defensiveness and to facilitate "corrective emotional" experiences of intimacy and closeness within the patient/therapist relationship.

Fosha's AEDP places the experience of affect, in the context of an emotional relationship, at the center of its understanding of normal psychic development and of the psychotherapeutic process.

Following the teachings of Donald Winnicott and Heinz Kohut, Fosha understands psychic activity as motivated by the person wishing to maximize the integrity and vitality of self-experience and the strength of meaningful attachments. In this framework, affects are, in the context of normality, deep sources of meaning and vitality and, in turn, affect frame experiences of integrity and authenticity. However, they may also become sources of suffering and anxiety when caretakers are actively destructive (through rejection, excessive criticism, or shaming) or passively destructive (by being detached, uninvolved, or absent). Affects experienced under such painful conditions may eventually overwhelm the mastery capacity of the developing ego and may make it necessary, to ensure survival, the use of increasingly complex defense mechanisms. The basic therapeutic stance of the AEDP therapist—supportive, affirming, and emotionally engaged from the initiation of therapy—is intended to create a therapeutic environment in which defenses can be momentarily withdrawn and deep feelings can become available for therapeutic reconstruction.

Techniques. The central goal of AEDP is the mobilization (and working through) of core affects and genuine experiences. According to Fosha, the key mutative agent in AEDP is the state transformation leading to the visceral experience of core affective phenomena within an emotionally engaged dyad.

AEDP aims to facilitate affective experiences such as (1) previously unbearable emotional experiences (including grief, anger, sadness, fear, and joy), (2) self-experiences (positive or negative self-images), and (3) relational experiences (feelings of closeness or distance from others), and (4) affects associated with healing and with breakthrough feelings emerging as liberation from emotional traps. To achieve these aims, the AEDP therapist can use what Fosha calls (1) relational techniques (animated by empathy), (2) restructuring interventions, using the classical triangles of conflict and triangle of person to facilitate cognitive understanding, and (3) experimental/affective interventions (facilitating deep, genuine affective experiencing).

Clinical Issues

INDICATIONS. AEDP has been used widely with a variety of outpatients suffering from the previous DSM-IV-TR Axis I disorders (anxiety disorders, dysthymia) and Axis II disorders (avoidant, dependent, and histrionic personality disorders). For patients with more severe psychopathology, such as somatoform, dissociative, and/or borderline disorders, a positive response to a “trial therapy” dose is necessary prior to the conduct of the full therapy course.

Success with narcissistic personality disorders has also been reported in selected cases.

LIMITATIONS. Exclusion criteria include all psychotic disorders, bipolar disorders, major depression, impulse disorders, marked acting-out behaviors, and substance abuse disorders.

COMPLICATIONS. No complications have been reported in the AEDP literature.

Psychoanalytic Psychotherapy for Panic Disorders—Barbara Milrod and the Cornell University Group

Theoretical Issues. The psychoanalytic psychotherapy for panic disorder proposed by Barbara Milrod and the Cornell University Group shares common theoretical underpinnings and metapsychology with classical psychoanalysis but differs in its application. For these researchers, panic symptoms carry psychological meanings, and the aim of PFPP (panic-focused psychoanalytic psychotherapy) is to uncover such unconscious meanings to achieve relief. In addition, their conceptual framework assesses continuous psychodynamic conflicts which tend to be quite specific to patients suffering from panic disorders: conflicts related to separation and the development of individuality. Anxiety is often aroused in these patients when confronting abandonment or separation threats. Conscious or unconscious anger is also a common trigger of anxiety reactions, including panic.

The theoretical underpinnings of their method (PFPP) incorporate key generally psychoanalytic precepts such as the following:

1. The role of unconscious fantasies. Unacknowledged (unconscious) rage is considered by Milrod's group to be an important part of the mental life of these patients, which—in their cases—relates to the onset of panic.
2. The common presence of characteristic defenses such as reaction formation, undoing, and denial, which are extensively

used by panic patients to minimize and deny anger, and protect their bonds with others.

3. The central role of transference in organizing the patient's relationship to others, including the therapist. As children, these patients often lived in fearful dependence, plagued by concerns about rejection and/or abandonment. Such concerns fueled repeated experiences of anger, which in turn exacerbated the fear of separation and loss.
4. The sudden failure of the overburdened system of defenses may lead to the onset of acute panic.

Techniques. Emerging from classic psychoanalytic metapsychology and from the common ground of psychoanalytic psychotherapy, panic-focused psychotherapy is an active, focused, interpretative technique that—at least in research projects—can be carried out in 24 sessions conducted once or twice a week. There are three phases to the therapeutic process: (1) an initial evaluation and early treatment phase normally focused on addressing the acute panic, and its surrounding dynamics, (2) a middle phase focused on the underlying core conflicts, and (3) a termination phase usually focused on the patient's substantive conflicts with separation and independence.

Techniques of work vary with the phase of treatment, but in a complete course of therapy, most of the classic psychoanalytic techniques are used. Clarifications, interpretations of impulses and feelings, and transference interpretations are covered. Affects (related to anger, interpersonal dependence, and fear of autonomy) are also mobilized and worked through.

Clinical Issues

INDICATIONS. The central indication of PFPP is the treatment of panic disorder, and this is the venue in which panic-focused psychotherapy has established its efficacy in randomized, controlled clinical trials. Many of the patients included in the reported trials suffered comorbid depression and reported improvement of their depressive symptoms as well, which may suggest the ready applicability of this technique to the treatment of depression with whatever modifications are called for as a function of different psychodynamic configurations.

LIMITATIONS. The most common current limitation of this technique relates to its specific focus or panic disorder. This limitation, however, needs to be balanced by the awareness of the public health effect of the successful treatment of panic disorder, whose patients carry a high suicide risk, suffer high rates of medical and psychiatric comorbidities, and incur very high overall medical costs.

COMPLICATIONS. The reported dropout rate of PFPP is lower than that reported for psychodynamic psychotherapies in general.

Deterioration of clinical status, worsening of panic symptoms, and the development of comorbid depression are also possible complications.

CONTRAINDICATIONS. The presence of active substance abuse, high active suicide risk, bipolar disorder, and psychosis contraindicate the use of this technique. Cognitive impairment of enough magnitude to interfere with learning represents another contraindication.

GOALS OF TREATMENT. Building on the psychodynamic formulation of panic, which understands the disorder as mediated by conflicts related to separation and autonomy and with difficulties with the recognition modulation and expression of anger, the goals of treatment are designed to facilitate the resolution of those conflictual dynamics. Thus, the initial phase of treatment brings into focus detailed and specific exploration of the feelings surrounding the onset of panic and the personal meanings evoked as associated with the symptoms of the panic episode. Goals of the initial phase (phase 1) include providing relief from panic and help to overcome the functional limitations that the disorder may impose on the patient as a function of the use of avoidance mechanisms.

The goals of the middle phase (phase 2) of treatment include highlighting and working through the core dynamics of the disorder: fear of unconscious impulses (anger and sexuality) and feelings of abandonment. The goal here is to desensitize the patient to the anxiogenic effect of angry affects and to facilitate the move from dependency to autonomy while modulating fear of loss and abandonment.

Phase 3 (termination) permits the re-experiencing of conflict about reparation and dependency directly in the therapeutic situation and in relation to the therapist, thus potentially consolidating the therapeutic work of phases 1 and 2. When the treatment is successful, patients report enhanced autonomy, less fear of anger and interpersonal conflict, and acceptance and enjoyment of independence.

OUTCOMES. PFPP has been tested in an open trial and a randomized, controlled clinical trial, both of which showed a significant reduction in panic symptoms. In the second trial, twice as many experimental psychotherapy patients met response criteria at termination than patients in the control conditions, which consisted of a cognitively enriched form of relaxation therapy. In addition, PFPP showed a significantly superior reduction in the severity of a broad range of panic symptoms as measured by the Panic Disorder Severity Scale: 73% of PFPP patients versus 39% of the controls met the response criteria for panic (40% reduction in scale score). PFPP patients also achieved a significant reduction in functional impairment.

Milrod and her group presented a manualized form of psychodynamic psychotherapy for panic disorder, which they tested in a randomized clinical trial. The treatment, consisting of an open-ended number of sessions in clinical practice and of 24 sessions in the clinical trial, is conceptualized as consisting of three phases:

Phase 1 involves the treatment of acute panic, phase 2 addresses the core dynamic conflicts that support the vulnerability for panic, and phase 3 is the termination phase. The goal of phase 1 is the elucidation of the conscious and unconscious meaning of panic symptoms with the goal of obtaining a map of the personal architecture of unconscious conflicts surrounding the panic. Common themes highlight the role of separation and dependence and problems in the modulation of sexual and aggressive impulses. Key techniques in this phase of treatment include keeping the focus on the symptoms of panic, engaging in a detailed and specific exploration of the circumstances surrounding the panic experience, and providing a firm, supportive and reassuring stance that the problem can be treated and the symptoms ameliorated.

The goal of phase 2 is the reduction of vulnerability to panic by engaging and working through (often in the transference) additional unconscious conflicts related to dependence and independence and building tolerance for instinctual experiences related to anger and sexuality.

The phase of termination, by bringing to the focus the key vulnerability of these patients (separation experienced as abandonment), provides an opportunity for consolidation of the therapeutic work. Milrod and her group addressed this problem by keeping an active focus on termination for the last one third of the duration of the therapy and by helping the patient work through such fears in the context of the therapeutic relationship.

STDP of Narcissistic and Other Self-Disorders—Manuel Trujillo

Theoretical Issues. The term narcissism has pervaded popular culture and the sociologic and psychiatric clinical literatures for many decades, depicting the increasing isolation and self-oriented individualism of affluent citizens in the United States and other developed countries.

Many features of contemporary culture mirror the narcissistic traits of individuals: the pursuit of materialistic cravings and the instrumental nature of many relationships; the seeking of success for success’s sake; and the cultivation of hedonic pleasure rather than the realization of ideals. In such contexts, narcissistic personality traits may be valued, adaptive, and, therefore, preferentially fostered.

Witness the success of the “Master of the Universe” meme in the late 20th century.

Freud first referred to narcissism to describe a libidinal position in which cathexes were invested in the subject’s ego and not in objects. Under those conditions, true transferences could not develop, making these patients ineligible for psychoanalysis. The clinical expression of nonobject libidinal cathexes included psychotic processes.

Dynamically, significant advances have been made over the last three decades in the clarification of the metapsychological dimensions of narcissistic disorders. Two authors, Otto Kernberg and Heinz Kohut, have advanced well-developed metapsychologies for these disorders and have proposed comprehensive techniques for their treatment through appropriately modified psychoanalytically based interventions.

There are, however, fewer attempts to apply techniques of STDP to patients suffering from self-disorders, despite the fact that such techniques have demonstrated respectable efficacy in the treatment of patients suffering from complex psychoneurotic disorders, selected personality disorders, and clinical conditions such as anxiety and depressive disorders. Patients meeting criteria for these disorders have been shown to make rapid and impressive gains, as long as their dynamic problems are related to difficulties in the management of sexual and/or destructive impulses, problems in the regulation of object loss, or complex combinations of both types of pathology. Patients suffering from extensive self-pathology are typically reported to do less well, largely because the structural underpinnings of self-disorders have not been as well understood as the constructs underlying Oedipal or attachment disorders.

A significant shift in the deep psychological understanding of narcissistic disorders is represented by the work of Kohut and summarized in his book *The Analysis of Self.* Kohut advanced a complete metapsychology of self-disorders, including (1) the superordinate position of the self as a whole system organizer to which other mental agencies (e.g., drives, ego, and superego) and mental capacities (e.g., motivation, vitality, goals, and values) are, to different degrees, subordinate; (2) and the relative independence—but rich interdependence—of the development of the self in relation to the development of the other components of the mental apparatus, for example, drives, id, ego, and superego.

Kohut described the relatively independent developmental line of the self and the various disordered responses to failures of environmental support, including parental mirroring. The subjective impact of these failures is registered through deficits and distortions of the perception of self (affecting self-esteem), the relative lack of integration of ambitions and capacities (creating grandiosity, self-effacement, and/or underachievement), or the incomplete development of appropriate, socially accepted values and ideals. For Kohut, repairing these deficits requires facilitating the development of a reparative transference of one of three types: (1) mirror transference proper, involving the use of the therapist as an admiring self-object and/or the resolution of perceived therapeutic failures via empathic interpretations; (2) idealizing transference, involving the acceptance by the therapist of the patient's unconscious needs to see them as a quasi-omnipotent object, capable of soothing and consoling and (transferentially speaking) worthy of emulation; and (3) twinship transference, which allows the patient to feel like others and develop a sense of connectedness under conditions of existential equality.

In the area of the self, and thus in narcissistic disorders, the key pathogenic phenomenon is lack of integration; that is, the inability of the mind to maintain a sense of wholeness. The signal of failure to maintain self-homeostasis is the affect of shame. Its presence indicates the inability of the self to maintain sufficient wholeness and fullness. In cases of more severe pathology, fragmentation anxiety denotes the organism's fear for its own basic integrity.

A very important theoretical and practical construct in self-psychology is the developmental process that, when successful, promotes the maintenance of a coherent, whole, optimally altruistic persona. When a "good enough" environmental fit is present, this process, fraught with complexities and opportunities for derailment, facilitates vital human functions such as (1) the slow maturation of a primal boundless, grandiose self into realistic, enduring, and resilient self-esteem, personal identity, and grounded ambitions and (2) the gradual transformation of undifferentiated, global, idealizing needs into values, ideals, and a moral compass that color and mediate each person's investment in, and involvement with, external social systems of beliefs, meaning, and action.

Narcissistic psychopathology occurs via two interrelated mechanisms: (1) the failure to complete key developmental steps along the pathway just delineated. Such failure may lead to a lack of maturation, integration, and realization of large segments of grandiose self (or idealized object image) and may create split-off (or dissociated) self-object representations, which entail potential problems for identity formation for the maintenance of self-esteem, and for the development of an appropriate and satisfying system of values, ideals, and accomplishments; (2) pathologic interrelatedness of self-development with drive-developmental processes. In this regard, the presence of significant unresolved self-developmental pathology may color drive-developmental and interpersonal processes such as separation-individuation, the ability to cope with object losses, and the resolution (or lack thereof) of triangular Oedipal pathology. This clinical situation often leads to hypersexuality (sex defending from abandonment), the sexualization of violence, and to other compulsive behaviors. Kohut, in fact, stated unequivocally that clinically significant Oedipus complexes occurred because the developing child's normal drives and conflicts were disturbed and intensified by unempathic parental responses.

Once consolidated, narcissistic psychopathology is expressed via (1) painful and pervasive subjective experience of shame, emptiness, and/or fragmentation anxiety; (2) the presence of extreme intrapsychic defenses (e.g., splitting and cognitive-affective dissociation) and interpersonal defenses (e.g., distancing, feeling extreme superiority or inferiority, intense neediness, or equally unyielding indifference); and (3) symbolic or manipulative behaviors designed to restore the integrity of self by providing soothing of vicarious experiences of wholeness via symbiotic fusion. At times, such soothing is obtained through the use of addictive drugs and or compulsive sexuality.

Techniques. Therapy can be conceptualized as a process that allows the patient—via the development, experience, verbalization, and resolution of intense transferences—to reactivate primitive, dissociated, anxiety-provoking views of self and others. Once activated, a more normal developmental pathway can be resumed in which primitive self-perceptions mature and become more integrated and healthier.

INTERVENTIONS. The paradigmatic intervention for self-problems is empathic understanding coupled with active interpretations of impulses, feelings, and self-perceptions; all in a context of pervasive use of transference. In this realm empathy involves an accurate grasp, both cognitive and affective, of what others experience and its presence creates a supportive milieu. In it, the patient can either bask in the use of the therapist as an optimum self-object or, bitterly but safely, complain about the therapist's failure while benefiting from the healing effects of the possibility of resolving the damage caused by that failure.

To accomplish this task in a relatively short-term program (50 sessions or less) requires the following:

1. **Empathy.** Accurate, empathically mediated grasp of the patient's core narcissistic vulnerabilities. For many patients, these problems are hidden behind a well-cemented wall of character defenses. Thus, these narcissistic vulnerabilities are apparent only after significant breakthroughs into the patient's unconscious rage and/or grief. For this part of the process, the ISTDP techniques developed by Sifneos, Malan, and Davanloo are crucial.
2. **Transference focus.** Active and early attention to the development of various types of self-object transferences.
3. **Activation of primitive self-states.** Early detection and reactivation (largely through active inquiry) of primitive,

dissociated, grand, or grandiose self-images and the containment of defensive affects, such as dissociative anxiety, fragmentation anxiety, and pervasive sense of embarrassment and shame. A simple inquiry such as “and how do you see yourself as a person now?” can bring up a mélange of primitive, childlike, exaggerated, or minimized view of self with a poorly differentiated mixture of affects such as shame, fear, pride, and others. Once all those feelings are “in the open,” the patient can do the necessary work of pruning and shaping such personal (at times idiosyncratic) views of self in order to carve a more grounded and satisfying personal identity.

4. **Imaginary reconstruction**. A special form of corrective emotional experience called “imaginary reconstruction.” This intervention permits the soothing and healing of narcissistic wounds via the imaginary reconstruction following a painful shameful, often catastrophic experience or failure of self or of a valued other. At these moments, the therapist inquiry of an imagined optimum (not perfect, but not wounding) version of the “failure narrative” often provides balsam to the narcissistic wounds. A few such interventions are provided in the case description below.

Just as the course of an episode of ISTDP treatment can be depicted by the use of the two triangles discussed earlier, the course of an episode of treatment of a (pure or mixed) self-disorder problem can be depicted as shown in Figure 32.5–4.

The typical therapy process focuses first—and repeatedly—on the overt manifestations of defensive activities, such as character defenses buttressed by splitting and cognitive–affective dissociation. Subsequent and repeated therapeutic challenges to the defenses may prompt the emergence of anxiety or paralyzing shame, and finally lead to the reactivation of formerly repressed and distorted self-representation. Distorted negative self-images are then deactivated via this exposure because the patients can see themselves more realistically and can experience and resolve the emerging grief. Positive self-images are allowed again to flow through the barriers of dissociation and are reactivated, owned (experienced by patients as rightfully belonging to them) and appropriately linked to hopes, dreams, skills, and values. In this way, profoundly personal projects of much meaning to the patient can re-emerge as valued blueprints for purposeful and meaningful life plans and actions.

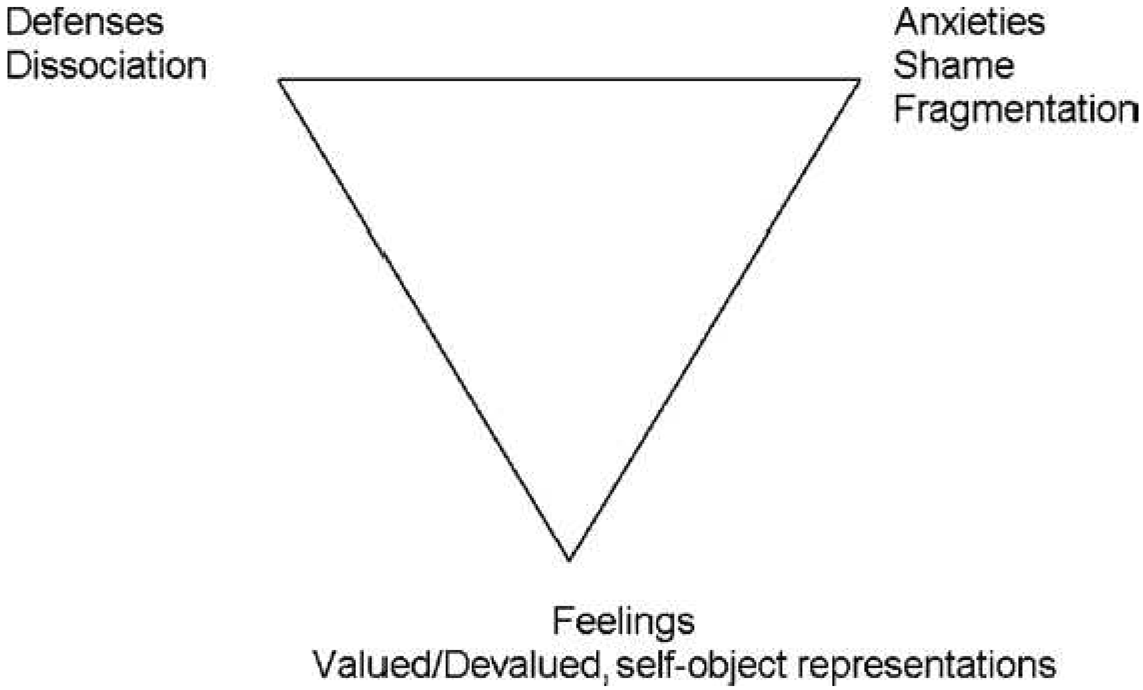

FIGURE 32.5–4. Triangle of the self.

WORKING THROUGH OF RAGE. The technique of working through narcissistic rage requires a sequence of interventions that achieve (1) keeping the focus on the rage, for example, a specific memory of rage in a patient's interaction with the patient's father, (2) reactivation of the physiologic–emotional–cognitive continuum underlying the rage response to allow the patient a simultaneous visceral and cognitive experience of the rage, (3) pressure and challenges to the defenses against rage to ensure a fuller experience, and (4) pressure and challenges to the defenses against interpersonal closeness and, especially, closeness with the therapist. As the therapist acquires, via the transference, a certain standing in the patient's emotional life, the attachment to the therapist heightens the patient's ambivalent attachment to the primary object and intensifies the patient's unconscious conflict over emotional loyalty. This is yet another source of considerable resistance and active therapeutic work. Rage toward the therapist may reach murderous proportions.

REACTIVATION OF SELF

Reactivation and Integration of Self. As patients gain mastery and knowledge of their experience of rage and the pathologic grief is lessened, they are often ready to reactivate grand and grandiose self-images, and to reinitiate and complete the reintegration of such self-images both with reality and with the person's unique set of skills and personal capacities. An enhanced sense of serenity and self-regard, a capacity for confident performance of work and play activities, and increased interpersonal empathy and closeness are usual results of such processes. The work of reactivation of primal, dissociated self-images also requires clarification and correction of multiple self-distortions.

For example, for GR—the patient to be presented next case—a rather large, athletic, and physically powerful man, the self-image of utter fragility as a premature, inhumanly small baby in danger of being trampled underfoot at any moment contaminated the way he saw himself and acted physically. His hunched-over posture made him appear considerably smaller than he actually was. As the therapy advanced, he started to derive considerable pride in his body, his physical prowess, and his dress style, for which he often sought the therapist's admiration and approval. Increasingly, his dreams and fantasy life are filled with images of grandness ("I feel like the captain of my own ship"), pride, and pleasure at his own physical, mental, and emotional characteristics.

The resolution and integration of this sequence of work required a process akin to a recapitulation of the developmental work of adolescence. Much as in late adolescence, through newly developed capacities for introspection, one takes some ownership and responsibility for one's self and destiny, patients often experience a mini-recapitulation of this developmentally appropriate sequence.

There is often a lessening of the earlier dependence on external mirroring and a sense of pride and pleasure at discovering, defining, and taking responsibility for one's unique personal assets and aspects; assets that can now be used to create a life of personal purpose.

Clinical Issues

INDICATIONS. *Adjustment disorders with depressive mood, with anxiety, and with mixed anxiety and depressed mood; generalized anxiety disorder; panic disorder with agoraphobia and without agoraphobia; dysthymic disorder; major depressive disorder; depressive disorder (recurrent).*

Table 32.5–1.
Diagnostic Criteria for Narcissistic Personality Disorder 301.81

A pervasive pattern of grandiosity (in fantasy or behavior), need for admiration, and lack of empathy, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following:

- Has a grandiose sense of self-importance (e.g., exaggerates achievements and talents, expects to be recognized as superior without commensurate achievements).
- Is preoccupied with fantasies of unlimited success, power, brilliance, beauty, or ideal love.
- Believes that they are “special” and unique and can only be understood by, or should associate with, other special or high-status people (or institution).
- Requires excessive admiration.
- Has a sense of entitlement, i.e., unreasonable expectations of especially favorable treatment or automatic compliance with their expectations.
- Is interpersonally exploitative, i.e., takes advantage of others to achieve their ends.
- Lacks empathy; is unwilling to recognize or identify with the feelings and needs of others.
- Is often envious of others or believes that others are envious of them.
- Shows arrogant, haughty behaviors or attitudes.

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, p. 760 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Narcissistic personality disorder and selected borderline personality disorder as well as the self-disorder core of many other clinical conditions such as eating disorders, certain addictions, and others.

Experienced self-psychologists recognize the following additional syndromic presentations:

Primary Self-Disorders

1. Primary, overt, self-disorder: the narcissistic personality disorder, as defined by the DSM-5-TR (Table 32.5–1).
2. Covert disorder (the hidden “mirror” image of the overt narcissist): while the overt narcissist has a grandiose sense of self-importance (e.g., exaggerates achievements and talents, expects to be recognized as superior without commensurate achievements), the covert narcissist may have many of the same traits but are not easily apparent; their narcissism appears as an exaggerated sensitivity to being or feeling criticized, ridiculed, devalued, or ignored.

   While the overt narcissist is preoccupied with fantasies of unlimited success, power, brilliance, beauty, or ideal love, and they are loud and aggressive about such claims, the covert narcissist cover such traits under a façade of self-effacement and even modesty while the potent narcissistic claims remain hidden. Often their narcissistic traits are expressed through perfectionism: nothing is good enough and they can also be self-deprecating: an indirect way to claim the appreciation of others. Covert narcissist can be as angry as the overt type when feeling unjustly slighted, but instead of exploding in expressive overt rage, the covert narcissist expresses it often in passive-aggressive ways or may generate extensive revenge fantasies and actions.

   Just like the overt type, the covert narcissist believes that they are “special” and unique and can only be understood by, or should associate with, other special or high-status people, but, often unsure of the truth of such conviction of goodness, constantly searches and critiques their own self-image or behavior in search of perfection. Such self-absorption limits their capacity for empathy.

3. Mixed self-disorders: disorders that reflect a mixture of narcissistic traits and conflict/drive problems. Narcissistic vulnerabilities and traits may fuel and compound many classic clinical conditions disorders, such as anxiety disorders, depressive disorders, and eating disorders.
4. Disorders mixed with other personality disorders. Pronounced self-disorder traits may color the clinical presentation and outcome of other personality disorders, such as avoidant and dependent personality disorders.

Secondary Self-Disorders

1. Disorders in which self-esteem and self-representation may be impaired through the effect of functional biologically based illnesses such as schizophrenia and bipolar disorders as they interact with the normal development of self.
2. Disorders in which self-esteem and self-representation may be impaired through the effect of structural, biologically based disorders such as attention-deficit/hyperactivity disorder in the normal process of development of the self.

The importance of secondary self-disorders lies in the fact that, if unattended, their presence may complicate the treatment and outcome of the primary disorder and force the patient to live with unresolved unconscious, defective, and pathogenic self-images. Each clinical presentation requires different interventions, priorities, and techniques.

LIMITATIONS. In addition to narcissistic personality disorders, many other psychiatric conditions affect significant aspects of the structure and/or development of the self. Patients suffering from such disorders as schizophrenia and bipolar disorders do indeed suffer impaired self-experiences and self-representations and object representations. In these conditions, such views of self and others, develop through neurobiologic alterations that may impede the structural integrity of the self. Other patients, such as those with antisocial or borderline personality disorders, may suffer from lack of ego integration or from a structural inability to develop the minimum of empathic relatedness necessary for therapy.

COMPLICATIONS. The most common complication is the development of a relatively stable transference neurosis, which requires additional time for its resolution. Negative therapeutic reactions in response to actual or perceived narcissistic injuries have been described.

GR was a 44-year-old unmarried professional man who first consulted because of unbearable anxiety, depression, and loneliness. Although periodically present throughout his life, these feelings appeared in a clinically explosive form at the end of a troubled 6-month relationship with Ms. B, a divorced professional woman who was the mother of a 10-year-old girl. Ms. B ended the relationship after a few months of sustained—if unsuccessful—efforts to please GR and cater to his many needs and demands. His escalating and often frantic demands reflected a deep and potent craving for perfectly attuned attention and responsiveness to the endless flow of his physical and emotional needs. Despite all of Ms. B's efforts (now painfully recognized by the patient), he usually treated her with a mixture of cruel devaluation and condescension, forcing her to admit, in practically every encounter, her incredible luck at being with a man with his "unique" sensitivity and passion and her pitiable incompetence at "adequately" responding to him.

Dynamically, his conscious life was dominated by the need to enact scenarios of uniqueness and specialness, in a never ending quest for positive confirmation from whomever was chosen to function as a confirming self-object. Failures in this regard met with a cascade of subjective effects such as unbearable anxiety, pervasive rage, and fury, which often led to prolonged, sadistic devaluing and attacks on the hapless offender. When such a person left GR's presence, his affect shifted to intense loneliness and profound depression, often requiring immediate medical attention.

The dynamic picture for patients like this can be represented as follows:

1. Ongoing, somewhat unstable conscious experience of entitled grandness or frank grandiosity, with large segments of the patient's adaptive behavior devoted to seeking opportunities for confirmation of the patient's grand(iose) view

of self and to searching for, and holding onto, complementing, confirming self-objects.
2. Storms of primitive anxiety and despair often accompanying the experience of narcissistic traumas and self-object rejections. This may include painful episodes of fragmentation anxiety, depression, and self-devaluation and prolonged experiences of emptiness. When these subjective states coincide with serious professional, health, or financial setbacks the risk of suicide is high.
3. An unconscious life dominated by negative self-images and expectations of rejection and devaluation. Such experiences color patients' conscious fears and frequently populate their dreams.

GOALS OF TREATMENT. At the end of the evaluation phase, GR's dynamic hologram appeared to have the following configuration:

Problems

- Inability to achieve reasonable success despite evident intellectual assets.
- History of stormy failed relationships with multiple women despite a craving to achieve stability, be married, and have a family.
- Chaotic life dominated by a craving for emotionally intense professional and emotional relationships (self-object-like).

Dynamics

- Ongoing and intense need for outside emotional supplies. The need to search for the perfect (i.e., all-admiring, all-giving, and nurturing) partner dominated his life to the practical exclusion of more mature relatedness.
- Primitive rage and vindictiveness triggered by perceived or actual failure of his partners to meet his unexpressed needs.
- Ongoing professional and personal self-sabotage.

Core conflicts

- Profound repressed primitive, ambivalent feelings toward the maternal object, resulting in an insoluble tension between libidinal investment and sadistic-destructive derivatives in his actual relationships with women. Thus, his relationships with women typically started on a high of idealization and, just as often, crashed in a low of violent recriminations.
- Impulse/Feeling: Ambivalence toward paternal object complicated by unresolved pathologic grief, with the consequent disruption of the development of dynamically equivalent actual adult relationships.
- Self-representations: To the degree that adaptive and realistic maturation of primal grand self-representations was delayed, the patient showed incongruent behavior: The grandiose self-images were acted out in consciousness in the form of perfectionistic demands on others while the helpless self-fed enormous self-doubts and neediness.

DYNAMIC HOLOGRAM OR ROADMAP. This dynamic hologram was obtained in the initial evaluation at a time when GR was often flooded with anger, grief, and other painful feelings. The hologram represents a roadmap for the unfolding of the therapeutic experience. For this patient, therapy will proceed through repeated runs around the triangle of impulses, person, and self. The two key pathogenic foci for him are (1) the focus of self and (2) the focus of unresolved grief for the death of his father. The pathogenic capacity of each focus reinforces the pathogenic capacity of the other. The images of helplessness and utter fragility derived from his repressed, disturbed negative self-images are generated first by identification with the failed aspects of his father's life and function as a way to maintain some bond with the father. Such a bond is also reinforced by his need to expiate guilt for his unconscious murderous impulses toward his father. The trigger for such destructive feelings relates to his father's failure to act as an adequate narcissistic self-object (optimum mirror), as his father, probably immersed in his own pervasive sense of failure, was unable to recognize GR's person, accomplishments, and inclinations.

The working through of these problems will require revisits to the experience of anger and destruction in relationship to the people in his current life (C) or toward the therapist (T) and subsequently toward his father (P). Multiple TCP interpretations were made before a sense of peace and emotional understanding and a modicum of closure and closeness obtained for the patient in relationship to memories of his father. Simultaneously, expressed and covert demands for closeness, support, feedback, extra time, and extra resources showed up in the therapeutic situation. The slightest disappointment in the therapist's responsiveness to such needs, evoked feelings of desolation and devastation in the session.

The previously described technique of imaginary reconstruction allowed the patient to evoke emotionally satisfying moments of perceived optimum mirroring from the therapist, often following the expression of intense anger toward the therapist as a function of real or perceived failures on the therapist's part. After one such episode, after the patient bitterly berated the therapist for not responding in the way he "needed" during a telephone interchange, the following dialogue ensued:

CLINICAL VIGNETTE

Therapist: You were obviously very disappointed with my response. What would you have wished me to say?

Patient: You know—something warm, something supportive, like "Of course, G[. . .], she should have agreed with you."

Therapist: And how would you have felt if I said that?

Patient: Relaxed, accepted, or like I have value.

Therapist: And how do you imagine I feel toward you?

Patient: Never thought of that—I think you want me to succeed, I think [with emotional conviction] you love me!

Therapist: And how do you feel about that?

Patient: Warm—Great it makes me think of my father. I think that if he would have been able to, he definitely would have loved me, how sad!

Therapist: What would he have loved especially about you?
Patient: That I am sensitive, that I am creative, that I am funny.
Therapist: Because—How do you see yourself as a person now?
Patient: Just like that: Funny, creative

The ultimate goal of the therapy is the rapid reactivation of formerly dissociated positive self-images and their fuller reintegration with such dimensions of the self as hopes, dreams, and aspirations for the future and such means of self-realization as actual skills and capacities. This aim may be succinctly expressed by paraphrasing Freud’s well-known dictum “Where id was, there ego shall be” as “where emptiness is, there self will be.”

Mutually agreed termination of therapy with patients suffering from narcissistic personality disorder or substantive narcissistic traits requires the fulfillment of the following criteria: Significant reduction of initial clinical symptoms and substantive lessening of the interpersonal and work dysfunctions presented at the initial evaluation. From the dynamic viewpoint, a sense of agency and self-possession often appears de novo, partially replacing the overwhelming sensitivity of these patients to the negative behaviors of others.

OUTCOMES. The outcome of the treatment of these disorders has not been tested through appropriately designed research studies. Clinicians treating many of these patients report substantial gains in symptoms, interpersonal function, personal integrity, and object-relatedness. Characteristics of the treatment process, such as short duration, enhanced activity of the therapist, development of treatment foci, use of transference, and mobilization of intense affects, have been satisfactorily researched in many studies of the efficacy and effectiveness of STDP in the treatment of severe psychoneurotic and personality disorders. These studies represent a good background for the design of future studies of the treatment of self-disorders.

By the 1950s, the randomized clinical trial became the gold standard of studies purporting to demonstrate the efficacy and effectiveness of medical interventions. Since then, these trials, which have guided the development of many such medical interventions, have slowly been introduced for verbal therapies as well. By the 1970s sufficient numbers of well-conducted studies permitted the meta-analysis by Smith, Glass, and Miller that demonstrated the general effectiveness of psychotherapy when applied to the treatment a broad range of conditions. More recently, the efficacy of psychotherapy has been demonstrated for various anxiety disorders, depression, certain personality disorders, including borderline personality disorder, and selected eating and addictive disorders.

Smith, Glass, and Miller studied the outcome of psychotherapy in 475 controlled studies that fulfilled preset criteria in terms of diagnostic validity, specification of the interventions provided in both the experimental and control conditions, and the systematic measurement of outcome. The main finding established that typical patients treated in the active experimental arm of the clinical trials were significantly better than 75% of untreated controls. The therapeutic gains reported at the conclusion of the therapy were maintained at 2-year follow-up. Both behavior and psychodynamic therapies appeared to be superior to other therapies.

No separate analysis was conducted in this review of the efficacy of STDPs. To assess the efficacy of STDP in treating specific psychiatric disorders, Falk Leichsenring and colleagues conducted a more recent meta-analysis. The review included studies conducted between 1970 and 2004 and were selected for meeting stringent criteria, such as randomized, controlled trials, therapies specified in treatment manuals, verification of treatment integrity, training of therapists to preset proficiency standards, and the use of reliable and valid diagnostic and outcome measures. Core features of the STDP selected included increased activity of the therapist, focus on the core themes and conflicts defined in the initial evaluation, mobilization of affects, and active use of transference with particular emphasis on the here and now.

In this meta-analysis, STDP yielded large and significant effect sizes for target problems (1.39), psychiatric symptoms such as anxiety and depression (0.90), and social functioning (0.80), effects that significantly exceeded those attained by controls consisting of wait-list and treatment as usual. Of note, patients in the active treatment arm of the studies continued to achieve clinical improvement between termination and 1- to 7-year follow-up, as reflected in Table 32.5–2.

Table 32.5–2.
Summary of Effect of Short-Term Dynamic Psychotherapy (STDP)

Mean Effect Site	STDP Term/Follow-Up	Usual Treatment Term/Follow-Up	Wait-List Term/Follow-Up
Target problems	1.39–1.57	0.55–0.84	0.27–NA
Symptoms	0.90–0.95	0.22–0.24	0.12–NA
Social function	0.80–1.19	0.38–0.95	0.21–NA

Outcome of STDP in Specific Psychiatric Disorders

Anxiety Disorders. Although there are far fewer studies of STDP applied to the treatment of anxiety disorders than trials of CBT, those studies provide some evidence that STDP is effective in the treatment of generalized anxiety disorder, social phobia, panic disorder, and posttraumatic stress disorder. Some of these studies included comparisons between psychodynamic psychotherapy and CBT, and in most of these the two methods of treatment were reported as being equally effective.

Generalized Anxiety Disorder. Treatments developed for GAD are represented by cognitive therapy and behavioral therapies focused on anxiety management (biofeedback, relaxation). There are few studies that include well-constructed STDP therapies in the comparison arm. The results of these few studies tend to favor CBT over dynamic interventions. One well-conducted study that included

99 patients randomly assigned to CBT or analytic psychotherapy showed the effect of CBT to be superior at termination but not at 1-year follow-up.

Panic Disorder. There are few studies of brief dynamic psychotherapy for panic. Milrod's group conducted several trials of manualized PFPP. In an open study of psychoanalytic psychotherapy that included 21 patients seen twice a week for 12 weeks and followed up at 6 months, they reported remission rates of 76% for the total sample and 93% for those who completed treatment.

The Helsinki Psychotherapy Study, a well-designed randomized study of the psychotherapeutic treatment of mood disorders, included a sizable number of patients suffering from panic disorder and generalized anxiety disorder. This study compared STDP (modeled after Malan's approach) with problem-solving, solution-focused therapy (SFT). In this study, 56% of patients with anxiety disorder did not meet criteria for any anxiety disorder after 7 months of STDP, compared with 42% of those in the SFT group. At 12 months, STDP patients had experienced a 34% reduction in the Hamilton Anxiety Rating Scale, compared with a 28% decrease achieved by SFT patients.

In a second study, a randomized, controlled trial of psychoanalytic psychotherapy was compared with applied relaxation training. Subjects treated with psychotherapy had significantly greater reduction in severity of panic symptoms and manifested improved psychosocial functioning. The effect size for symptomatic improvement was a robust 0.95 and for disability improvement was 0.74. Significant reductions in the score for depression, based on the Hamilton Scale for Depression, were also achieved, with a respectable 0.53 effect size.

DEPRESSION. A sizable number of reviews of psychodynamic psychotherapy trials for the treatment of depression is available. A review by Leichsenring in 2001 comparing STDP and CBT concluded that the clinical results achieved were not substantially different between the techniques. A. A. Abbass and collaborators, in a 2014 study sponsored by the Cochrane Library, reviewed all published studies in which dynamic psychotherapy lasting less than 40 hours were included.

The authors included 33 studies involving 2,173 participants and measured symptom reductions across somatic, general, anxiety and depression domains, as well as social function and interpersonal problems. Despite the studies diversity, and except for the somatic dimension, all other measures showed substantial superiority of the STDP arm of the studies over the control conditions in the short and the long term. Effect sizes increased over the long term.

PERSONALITY DISORDERS. Personality disorders represent a major challenge for clinical practice and for research due to their heterogeneity and the high level of comorbidities they share with Axis I and other Axis II disorders.

A specific meta-analysis of psychodynamic psychotherapy applied to personality disorders conducted by Leichsenring and Leibung showed robust gains in outcome domains such as target complaints, broad-based symptoms, and social functioning but did not demonstrate large outcome differences between the STDP and alternative CBT therapies. The effect size for psychodynamic psychotherapy was 1.31 versus 0.95 for CBT.

A study by Martin Svartberg and colleagues comparing STDP to CBT for Cluster C personality disorders with anxiety and depression comorbidities reported significant improvement at termination and 2-year follow-up. At the 2-year follow-up, 56% of patients treated in the STDP group and 42% of CBT patients had returned to community standards of normal functioning.

A survey by Abbass and colleagues evaluated the efficacy and long-term effectiveness of ISTDP in the treatment of patients with DSM-IV-TR personality disorders. The study compared Davanloo’s ISTDP to a minimal-contact controlled condition. The most common personality disorders included in the treatment sample (borderline, obsessive-compulsive, and avoidant) corresponded to reported prevalence rates in clinical samples drawn from broadly representative outpatient facilities. The sample also exhibited high levels of interpersonal problems, work disability, and use of psychotropic medications, thus qualifying as a severe sample. The ISTDP-treated group showed superior outcomes relative to controls in measures related to symptoms, social functioning, return to work from disability, and reduced use of psychotropic medications. Most participants (86%) no longer met criteria for any personality disorder at the completion of treatment or at the follow-up assessment.

ISTDP in Other Psychiatric Disorders. The development of manualized therapies based on the psychoanalytic psychodynamic framework has fostered their application in therapeutic trials designed to assess ISTDP's efficacy when applied to a broad range of psychiatric disorders. A study of STDP in anorexia nervosa compared it with dietary advice and found no difference in body weight at termination and 1-year follow-up, although ISTDP patients reported better social and sexual adjustment. A larger trial compared ISTDP (based on Malan's approach) with specialized family therapy, cognitive therapy, and standard psychiatric care. Within a context of modest gains, all therapy conditions performed better than the control.

Although there have been a modest number of trials of ISTDP for alcohol dependence and abuse, there is insufficient evidence to support psychodynamic psychotherapy as a robust first-line therapy for these patients. Trials with cocaine-dependent patients were also inconclusive.

A study by Woody and colleagues compared a dynamic (supportive-expressive) psychotherapy condition with CBT and with drug counseling in opiate-dependent individuals maintained on methadone. Patients in all three groups showed improvements in lessened drug use, reduced criminal behavior, and improved psychological functioning. STDP showed better results than CBT in reduction in drug use, improvement in work status, and lessening of psychological problems, whereas CBT improved the patients' ability to deal with their legal problems.

In a multicenter prospective trial, 157 breast cancer patients with comorbid depression were randomized to either individual STPP (intervention group, *N* = 78) or “treatment as usual” (control group, TAU, *N* = 79). As the primary outcome measure, the authors hypothesized a higher rate of remission defined as no diagnosis of depression (Structured Clinical Interview for DSM-IV) and reduction in depression score by at least 2 points (Hospital Anxiety and Depression Scale, HADS-D) in STPP versus TAU at treatment termination. Secondary outcomes mainly refer to quality of life (QoL).

In the intention to treat (ITT) analysis, 44% of the STPP group achieved highly significantly more remission than TAU (23%). STPP treatment (OR = 7.64; $p < 0.001$) was the strongest predictor for remission post-treatment; time was also significant (OR = 0.96; $p < 0.05$). A high effect favoring STPP ($d = 0.82$) was observed for the HADS-D score post-treatment (secondary outcome). Regarding further secondary outcomes (QoL), analyses of covariance yielded main effects for group (favoring STPP with an effect size of at least $d = 0.5$) for global QoL, role, emotional and social functioning, pain, treatment side effects, breast symptoms, and upset by hair loss. STPP is an effective treatment of a broad range of depressive conditions in breast cancer patients improving depression and functional QoL. The interpretations of these findings are limited by the drop-out rate (∼1/3) and by delayed post-treatment assessments. Future trials may consider stepped-care approaches, tailored to patients’ needs and requirements in the acute treatment phase.

COMBINED ISTDP AND MEDICATIONS. Many trials combining specific psychopharmacological agents and various forms of brief dynamic psychotherapy and CBT have supported the superiority of the combined condition over each intervention provided alone. In the treatment of panic disorder, Wiborg and Dahl studied the effect of adding psychodynamic psychotherapy to treatment with the tricyclic antidepressant clomipramine. The addition of the therapy (a manualized blend of Davanloo, Malan, and Strupp and Binder’s work) significantly strengthened key therapeutic outcomes. At termination, all patients in the combined therapy conditions were free of panic. Patients in the combined therapy arm suffered far fewer relapses at 18-month follow-up (20% vs. 75% for clomipramine alone).

Regarding depression, many studies comparing psychodynamic therapy with medication have established comparable efficacy for psychotherapy and medication and the superiority of combined therapies over the medication-alone condition.

PROCESS-OUTCOME CORRELATIONS—MECHANISM OF CHANGE. Beyond established efficacy and effectiveness for specific psychiatric disorders, therapies classifiable under the umbrella term ISTDP need to refine their knowledge regarding what specific ingredients of the proposed therapy are responsible for what aspects of the process of change that determines the outcome of the intervention. Among the ingredients mentioned in the spectrum of therapies described in this section, ingredients such as the enhanced activity of the therapist, the maintenance of a relevant focus, the mobilization of unconscious affects, and the use of the transference are most often mentioned as relevant to the therapeutic outcomes obtained.

The therapeutic efficacy of those factors has been supported by well-conducted process—outcome correlation studies and selective meta-analyses. Diener and colleagues systematically examined the relationship of emotional experience and expression and the therapeutic outcome of dynamic psychotherapy through a meta-analysis of 10 independent studies. By calculating an overall effect size of the relationship between therapist-directed focus on affect and the outcome of psychotherapy, they concluded that in the sample analyzed, such focus on affect was associated with considerable patient improvement. The overall average weighted effect size across all outcomes measured was reported as significant ($r = 0.30$), suggesting a 30% difference in success rate for patients who received an affective therapeutic focus.

Regarding transference, transference interpretations are generally assumed to be a key treatment ingredient for psychoanalytically oriented short-term psychotherapy. Yet, despite their theoretical importance, few empirical studies have attempted to measure the long-term effects of transference interpretations in brief dynamic psychotherapy.

Glund and coworkers conducted such a study by randomly assigning 100 patients to brief dynamic psychotherapy with or without transference interpretations. Although both treatments demonstrated significant improvement after treatment termination, patients with a lifelong pattern of poor object relations benefited more from the 1-year transference-based condition than from therapy without transference interpretations, an effect that was sustained through the 4-year follow-up period. The findings of this study suggest that transference interpretations may be especially suited to improving the interpersonal functioning of patients with prior histories of poor object relationships.

Review of the growing literature on the application of brief dynamic psychotherapy to many specific psychiatric disorders has broadened the original findings of the meta-analysis of Smith, Glass, and Miller. Beyond their conclusion that the typical psychotherapy patient is better off than the untreated control at termination of the treatment, it is now possible to assert the specific efficacy of dynamic approaches when treating anxiety disorders, depression, a broad range of personality disorders and other selected eating disorders, and substance abuse conditions. Efficacy has also been established for the combination of psychodynamic therapies with medication. The challenges for future outcome research include identifying precise and specific mechanisms of change, generating additional innovative treatments for resistant conditions, and refining the technical specificity of existing psychodynamic techniques.

Psychotherapy and Neurobiology

It is to Freud’s everlasting credit that psychoanalysis transformed the current understanding of mental life by contributing, for the first time in history, a scientifically grounded theory of mind. Unfortunately, his deeper ambition—to see the grand edifice of psychoanalysis anchored in the solid foundation of neurobiology—has not yet been realized. Freud was forced to abandon such ambitions, expressed in his “Project for a Scientific Psychology,” because of the immaturity of the brain science of his time, among many other reasons. In “Beyond the Pleasure Principle” Freud asserted “The deficiencies in our description would probably vanish if we were already in a position to replace the psychological terms with physiological or chemical ones. We may expect physiology and chemistry to give the most surprising information.”

He also asserted in his paper “On Narcissism: An Introduction” that “we must recollect that all of our provisional ideas in psychology will presumably one day be based on an organic substructure.”

The absence of a sound neurobiologic base with which to link his emerging discoveries about the structure of mental life forced Freud to abandon his early ambition and to concentrate his efforts on the development of a psychoanalytic metapsychology that has stood, independent of its biologic underpinnings, for more than a century and has produced abundant clinical and metapsychological fruits.

The growth of neural science in general and of cognitive neuroscience and neuropsychology in particular affords contemporary psychoanalysis a second chance to anchor classical metapsychology in a newly evolving neural science.

The following findings are highly suggestive of potential avenues for the construction of a neurobiologically anchored psychological science. Such pathways include the capacity for verbal therapies to modulate basic neurophysiologic brain functions and alter serotonergic function, modify synaptic plasticity, and changing the activation of key brain circuits in areas such as the anterior cingulated cortex, various limbic structures, the prefrontal cortex, and other brain centers that play key roles in the processing of key emotions and views of self and others.

Psychotherapy and the Modulation of Brain Function. The work of Kathleen Shear in the early 1990s demonstrated that the application of CBT to patients suffering from panic disorder appeared to reverse their pretreatment vulnerability to lactate-induced panic, raising the possibility that CBT, at a minimum, modulated the biologic cascade of events that trigger a panic episode in vulnerable patients following the infusion of sodium lactate and thus providing evidence of the biologic effect of talk therapies.

EFFECTS OF VERBAL THERAPIES ON KEY NEUROTRANSMITTERS: SEROTONIN. More evidence for the robust biologic effects of verbal therapies was subsequently produced by L. R. Baxter Jr. and his collaborators. In a study published in the Archives of General Psychiatry in 1992, Baxter investigated local cerebral metabolic rates for glucose in patients with OCDs before and after treatment with either fluoxetine or behavior therapy. After treatment, the investigators observed significant reductions in glucose metabolic rates in the head of the right caudate nucleus, a reduction not observed in nonresponding patients or in normal controls. In considering how behavior therapy could produce brain function changes similar to those caused by neurochemically specific drugs, the authors invoked the findings of Eric Kandel, who demonstrated, in the sea slug Aplysia the changes at synapses that use serotonin to mediate learned changes in stimulus-response behavior.

It is thus possible that both medication and behavior therapy achieve their therapeutic effect via the modulation of the availability of serotonin to the synapse. Although medication is reported to act by blocking the reuptake of serotonin, behavior therapy—after all, a specialized form of learning—may achieve its therapeutic effect by enhancing the functional availability of serotonin at the synapse level.

Also in the late 1990s, Heimo Viinamäki demonstrated, via single-proton emission computed tomography, the normalization of serotonin transporter densities in the medial prefrontal cortex, midbrain, thalamus, and striatum in a 25-year-old patient suffering from depression and personality disorder who received 1 year of once-a-week psychodynamic psychotherapy.

Psychotherapy and Synaptic Plasticity. A study by Arthur L. Brody and colleagues demonstrated that subjects with major depressive disorder had regional brain metabolic abnormalities at baseline that seemed to normalize equally with the application of either of two forms of treatment: the antidepressant paroxetine and interpersonal psychotherapy (ITP). In testing the hypothesis that in subjects with major depressive disorder, metabolism would (with treatment) increase in the dorsolateral prefrontal cortex (DLPFC), and decrease in the ventrolateral prefrontal cortex, the authors found that both paroxetine and ITP normalized the metabolic abnormalities defined at baseline. Specifically, both groups showed decreases in prefrontal cortex and left anterior cingulate gyrus metabolism, and increases in left temporal lobe metabolism. In discussing the potential mechanisms of action of ITP the authors invoked changes in synaptic plasticity through learning-induced retraining of implicit memory systems.

Because of ITP's special focus on improved socialization, areas of the brain associated with that function may undergo an attenuation of neuronal connectivity during such treatment. Just as signs of distress in socially isolated animals correlate with increased metabolic activity in the cingulate cortex, improvements in the socialization of depressed patients achieved through ITP may lead to reductions in the activity of the anterior cingulate cortex (ACC) and a corresponding amelioration of depression-induced negative evaluation of self and others.

A more recent study by S. Saxena and collaborators reported findings that establish the intriguing possibility that psychotherapy (CBT in their experiment) may share with medication some of their effects in brain physiology while producing their own specific effects in addition. In seeking to elucidate the nature of the brain response to brief intensive CBT for OCD, Saxena obtained positron emission tomography (PET) brain scans of patients suffering from OCD before and after a course of 4 weeks of daily intensive CBT supplemented by disorder-specific homework assignments, and compared the brain scans of patients undergoing therapy, with the brain scans of untreated controls. Treated patients, as expected, showed significant decreases in thalamic metabolic activity, and, unexpectedly, they also showed a significant increase in right ACC activity, a response that the authors considered necessary to explain the patients' response to CBT, as well as the relative rapidity and robustness of such responses. In contrast to the effect of pharmacotherapy on brain function, augmentation of dorsal ACC activity maybe a primary mechanism of action of CBT for OCD.

It is known that subregions within the ACC have different functions. A critical such function, which may mediate the effect described here, involves the conscious regulation of emotion via re-evaluation of negative emotions, suppression of excessive arousal, and suppression of amygdala-mediated negative emotionality. These effects, specific for the CBT arm of the study, tended to occur rather rapidly and were correlated with treatment response at 4 weeks.

Table 32.5–3.

Functional Neuroanatomy of Neurobiologic Effects of Verbal Therapies

Disorder	**Treatment Technique**	**Post Treatment Findings**
Obsessive-compulsive disorder–social phobia (PET—Baxter, 1992; SPECT—Nakatani, 2003) Pediatric OCD (fMRI—Huyser, 2010)	Cognitive-behavior therapy/behavior therapy CBT CBT	↓ Metabolism in right caudate nucleus Uncoupling of cortico-striato-thalamic activity ↓ Limbic metabolism ↓ Thalamic metabolism Increased dorsolateral prefrontal and parietal (from baseline low)
Major depression	Cognitive-behavior therapy	Decreased ↓activation in dorsolateral prefrontal cortex Decreased ↓activation in para hippocampal gyrus ↑ Activation in limbic regions
Major depression (PET—Brody 01, SPECT—Martin 01) fMRI—Ritchey, 2011	Interpersonal psychotherapy CBT	↑ Activation in right basal ganglia ↑Activation in right posterior cingulate ↓Metabolism in prefrontal cortex, which correlates with symptom improvement ↑Ventromedial, ↑Amygdala, caudate, hippocampus
Disorder	Treatment technique	Posttreatment Findings
Panic Disorder	Psychodynamic	↓Amygdala and

(fMRI—Beutel, 2010)	Thx	hippocampus ↑Prefrontal activity

The biologic relationship between negative affects and their modulation is central to disorders such as anxiety and depression, and it is a core paradigm that governs the psychotherapeutic treatment of these disorders. A central task of psychotherapy involves the repeated re-examination of painful memories, the experience and expression of the negative emotions associated with those memories, and the acquisition of a modulated soothing perspective. Such a sequence has also been studied in normal subjects, on whom the process of memory reappraisal has been found to be associated with improved mood, increased activity in prefrontal cortex subregions, and decreased activity in the amygdala and orbitofrontal cortex.

The results of studies examining the functional neuroanatomy underlying the neurobiologic effects of verbal therapies are summarized in Table 32.5–3.

Brain Structures Key for Psychotherapy

The Amygdala. Although every region of the brain may participate in any mental process, selected regions may have special relevance for the health (or disorder) of key mental functions and for the treatment of such problems through psychotherapy. The amygdala is probably the brain structure most commonly understood by clinicians as essential to the experience of emotions, emotional dysregulation and, conversely, emotional modulation through psychotherapy. It tends to show elevated activity in depression and posttraumatic stress; its dysregulation worsens in the presence of reduced prefrontal cortex activity (which supports key ego-functions) and improves through various means including verbal therapies. This dyad: amygdala, frontal cortex drives much of what is adaptive in humans and is capable of creating just as much emotional pain and dysfunction when it overwhelms human prudence and judgment.

Prefrontal Cortex. Within the frontal lobes, in addition to the already mentioned ACC, the dorsomedial prefrontal cortex (DMPFC) is central to the exercise of top-down cognitive modulation of fear and other emotions, and to paying attention (self-awareness) to one's own (and other's) emotional responses facilitating mentalization and empathy. It executes many of the functions traditionally ascribed to the ego and to the normative superego. It also supports object relationships. It is hypoactive in depression. As long as it remains hypoactive, it limits the potential success of psychotherapy. Even without the availability of imaging technologies, the wise psychoanalytic clinicians of yore treating such patients, often insisted on affording them an initial period of therapy designed to "build up the ego," largely through being supportive, providing empathy, and supporting the patient's own efforts for tender loving self-care. Exercise however minimal in the early stages, rest, proper nutrition (protein, minerals, etc.) help as well. If necessary, treatment with stimulating antidepressants rapidly refills the empty "battery."

In the authors' own practice, they have been able to detect through neuroimaging (SPECT) significant under activity in the DMPFC in a sizable number of patients who became "stuck" in short-term therapy after an initial period of rapid progress, and who could resume their expected progress after the SPECT guided treatment of detected neurobiologic dysfunctions.

The Dorsolateral Prefrontal Cortex. A close neighbor of the DMPFC reviewed above, the DLPFC provides, under normal conditions, the basic infrastructure of the person's executive functions.

Attention, concentration, control of effort, relational processes, working memory, emotional memory, and many other functions under its control, are absolutely essential for successful adaptation to function and to people. It can also be hypoactive in depression and may contribute to the experience of loneliness, social isolation, loss of interest, and low standards of performance. In geriatric depression, the poor concentration, inability to focus, and no desire and no capacity to get much done raises the question of the differential diagnosis of dementia versus depressive pseudo dementia. Imaging studies showing hypoactive DLPFC without the widespread frontotemporal or temporoparietal underactivity usually present in true dementia, points toward a diagnosis of depression.

Neuroimaging and Basic Psychodynamic Principles

Imaging of Core Psychoanalytic Constructs. Of the many challenges that lie ahead as psychiatry and psychoanalysis continue their march toward becoming true science-based, biopsychosocial disciplines, two are really within the reach of currently emerging science.

The first challenge consists of the need to delineate the brain systems which underlie normal psychological functions, and which may signal disorder when functioning outside of defined ranges. Validated research findings are emerging, for example, which enable one to know how the temporal lobes and thalamus act in concert to modulate mood and to restrain wild fluctuations of mood in response to every day moderately challenging stimulus. Key research findings provide information as well, how the orbitofrontal cortex facilitates (or inhibits) impulsive activity, depending upon its functional state, and what conditions may lead to poor impulse control in a particular single event or situation, in response to a particular class of stimulus (sexual stimulation, risk taking in gambling) or as an expression of enduring personality characteristics.

The second fascinating task is represented by the need to re-evaluate, and if necessary reformulate, core psychoanalytic constructs, which were formulated by Freud while he waited for a robust brain physiology that would help trace neural activity as it was converted into basic mental function. As he humbly said: “The deficiencies in our description would probably vanish if we were already in a position to replace the psychological terms with physiological or chemical ones. We may expect physiology and chemistry to give the most surprising information.”

As summarized by Bradley S. Peterson and other investigators, imaging technologies can be used to develop experimental paradigms that may permit the understanding of the neural basis underpinning central theoretical constructs of psychoanalysis.

Drives and Instincts. The classical Freudian tenet that humans, as well as all living organisms, seek to maximize pleasure and minimize displeasure has been expanded by modern neuroscience into a complex theory of affects and rewards that describes numerous emotional systems served by different neural substrates. Such neural substrates, under the label pleasure, can discern and delineate and quantify various reward-systems and can help the organism to select which behaviors it enacts (or avoids) when a particular neural circuit is activated or inhibited. Thus, such circuits—and their attendant neurotransmitter(s)—may direct or even compel the organism to seek extremely risky behaviors, if driven by a severe enough "dopamine deficiency" state for example.

The nucleus accumbens and related limbic circuitry contribute to the "final reward equation" by assigning values of pleasure or displeasure to affective experiences, and the alerting/arousal system recruits the reticular formation, thalamus, amygdala, and cortex to the task of assigning personal meaning and salience to those affective experiences. Different emotional states are, in this paradigm, the result of complex and distinct interactions between the reward/pleasure systems (which promote appetitive drives) and the activation of arousal systems that assign priority to the pursuit of certain objects or activities over alternative behavioral choices. Imaging studies have shown increased activation of the nucleus accumbens and the orbitofrontal cortex while experimental subjects experience the euphorigenic effects of certain drugs, whereas the intense dysphoria experienced during abstinence correlated with metabolic underactivity of the same circuits.

As these investigations go forward, they are likely to yield a more refined theory of affects, drives, and emotions on which to create a modern psychoanalytic metapsychology and a theory of techniques built on hypotheses that can be tested in controlled studies. The role of the ego and superego as mediating agencies in the modulation of drives is as central to psychoanalysis as Freud's theory of drives.

In this area, modern neuroscience in general and neuroimaging in particular are generating seminal insights.

The relevance of frontostriatal systems is key to understanding how individuals plan, execute, and monitor goal-directed behaviors by selecting from the innumerable choices presented to them by the environment, as well as by their own perceptions, affective experiences, memories, and cognitions. Neuroimaging studies indicate that the ACC and prefrontal cortex are involved in the top-down regulation of the capacity for self-regulation through their influence on broader brain regions such as temporal and parietal cortices and the basal ganglia. Neuroimaging studies of disorders of self-regulation, such as attention-deficit disorder and bipolar disorder, have established the existence of functional disturbances of cortico-striato-thalamo-cortical activity.

These discoveries have direct clinical application as well; witness, for example the very positive clinical impact of adding anticonvulsant mood regulators to the treatment mix of patients suffering from anxiety or depressive disorders or even attention-deficit disorder, when neuroimaging determines the presence of temporal lobe over or underactivity. Under such conditions, pure antianxiety, antidepressant, or attention focusing treatments are not sufficiently effective because of the severe mood disruption precipitated by temporal disorganization.

Memory is also a core player in psychoanalytic discourse. Neuroscience research has described numerous memory systems with diverse distributions in the central nervous system. A key distinction for psychoanalysis is that between declarative and procedural memory systems. Declarative (explicit) memory is mediated by the medial temporal lobe and the hippocampus and stores conscious memory for people, places, and things. Procedural memories, on the other hand, are completely unconscious and are evident only in behavioral performance rather than conscious recall. Procedural memory involves the striatum (mediating the acquisition of skill and habits), the neocortex, the amygdala (facilitating conditioned emotional response), and the cerebellum.

Many investigators have advanced the idea that a part of the unconscious (the procedural unconscious, which maps onto procedural memory) is central to key processes of psychoanalytic therapy. Early attachment relationships, for example, are internalized and encoded through procedural memory and unfold in the therapeutic process in the form of transference. Psychotherapy, in this context, represents the potential for the development of a new primal attachment relationship that may create the conditions necessary to change attachment-related, implicit procedural memory as new interactions with the therapists are internalized by the patient in corrective self-object representations. Such maybe the deepest meaning and therapeutic value of the concept of "corrective emotional experience" advanced by Franz Alexander in the 1940s and of the "moment of meaning" developed by Sanders and Stern to describe moments in the interaction between patient and therapist that break the impasse of a transference-based pattern of relatedness and advance the therapeutic relationship to a new level.

Neuroimaging and the Prediction of Outcome

The most persuasive outcome predictions have developed out of neuroimaging studies of depression. In one such study, increased activity in the rostral ACC predicted a positive response to treatment, a finding that has been replicated in other studies.

The anterior cingulate performs many functions related to conflict resolution and response to basic emotions. The central region of the cingulate (rostral subregion) may be recruited to resolve conflict between emotional stimuli or between conflictive cognitive mental contents. Thus, patients with higher pretreatment levels of activity in this area of the cingulate may be in a better position to resolve their pathogenic conflicts.

In studies related to the prediction of outcome of treatment in OCD, the orbitofrontal cortex is a key area. Neuroimaging studies have shown that lower pretreatment levels of activity predicted a good response to pharmacotherapy, whereas higher pretreatment levels or activity predicted a better response to psychotherapy.

FURTHER READINGS

Abbass A, Sheldon A, Gyra J, Kalpin A. Intensive short-term dynamic psychotherapy for DSM-IV personality disorders. *J Nerv Ment Dis*. 2008;196(3):211–216.

*Abbass AA, Kisely SR, Town JM, et al. Short-term psychodynamic psychotherapies for common mental disorders. *Cochrane Database Syst Rev*. 2014;(7):CD004687.

Alexander F, French TM. *Psychoanalytic Therapy: Principles and Application*. Ronald Press; 1946.

Baxter LR, Schwartz JM, Bergman KS, et al. Caudate glucose metabolic rate changes with both drug and behavior therapy for obsessive-compulsive disorder. *Arch Gen Psychiatry*. 1992;49(9):681–689.

Beutel ME, Stark R, Pan H, Silbersweig D, Dietrich S. Changes of brain activation pre- post- short-term psychodynamic inpatient psychotherapy: an fMRI study of panic disorder patients. *Psychiatry Res*. 2010;184(2):96–104.

Busch FN, Milrod BL, Sandberg LS. A study demonstrating efficacy of a psychoanalytic psychotherapy for panic disorder: implications for psychoanalytic research, theory, and practice. *J Am Psychoanal Assoc*. 2009;57(1):131–148.

*Davanloo H. *Intensive Short-Term Dynamic Psychotherapy*. Wiley; 2000.

Davanloo H. Intensive short-term dynamic psychotherapy. In: Sadock BJ, Sadock VA, eds. *Comprehensive Textbook of Psychiatry*. 8th ed. Lippincott Williams & Wilkins; 2004.

Driessen E, Cuijpers P, de Maat SC, et al. The efficacy of short-term psychodynamic psychotherapy for depression: a meta-analysis. *Clin Psychol Rev*. 2010;30(1):25–36.

*Etkin A, Pittenger C, Polan HJ, Kandel ER. Toward a neurobiology of psychotherapy: basic science and clinical applications. *J Neuropsychiatry Clin Neurosci*. 2005;17(2):145–158.

Forbes EE, Olino TM, Ryan ND, et al. Reward-related brain function as a predictor of treatment response in adolescents with major depressive disorder. *Cogn Affect Behav Neurosci*. 2010;10(1):107–118.

Fosha D. Emotion and recognition at work: energy, vitality, pleasure, truth, desire and the emergent phenomenology of transformational experience. In: Fosha D, Siegel DJ, Solomon MR, eds. *The Healing Power of Emotion: Affective Neuroscience, Development, Clinical Practice*. Norton; 2009.

Fosha D, Paivio SC, Gleiser K, Ford J. Experiential and emotion-focused therapy. In: Courtois C, Ford JD, eds. *Complex Traumatic Stress Disorders: An Evidence-Based Clinician's Guide*. Chapter 14. Guilford Press; 2009:286–311.

Freud S. The resistances to psycho-analysis. In: Strachey J, ed. *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XIX (1923–1925): The Ego and the Id and Other works*. Hogarth; 1925.

Gerber AJ, Peterson BS. Measuring transference phenomena with fMRI [abstract]. *J Am Psychoanal Assoc*. 2006;54(4):1319–1325.

Hilsenroth MJ. A programmatic study of short-term psychodynamic psychotherapy: Assessment, process, outcome, and training. *Psychother Res*. 2007;7(1):31–45.

Hilsenroth MJ, Defife JA, Blake MM, Cromer TD. The effects of borderline pathology on short-term psychodynamic psychotherapy for depression. *Psychother Res*. 2007;17(2):172–184.

Huyser C, Veltman DJ, Wolters LH, et al. Functional magnetic resonance imaging during planning before and after cognitive-behavioral therapy in pediatric obsessive-compulsive disorder. *J Am Acad Child Adole Psychiatry*. 2010;49(12):1238–1248.

*Kohut H. *The Analysis of the Self*. International University Press; 1971.

Leichsenring F, Rabung S, Leibing E. The efficacy of short-term psychodynamic psychotherapy in specific psychiatric disorders: a meta-analysis. *Arch Gen Psychiatry*. 2004;61(12):1208–1216.

Leichsenring F, Schauenburg H. Empirically supported methods of short-term psychodynamic therapy in depression—towards an evidence based unified protocol. *J Affect Disord*. 2014;169:128–143.

Malan D. *The Frontier of Brief Psychotherapy: An Example of the Convergence of Research and Clinical Practice*. Plenum Medical Book Co; 1976.

Malan D. *Towards the Validation of Dynamic Psychotherapy*. Plenum Press; 1976.

Milrod B, Leon AC, Busch F, Rudden M, Schwalberg M. A randomized controlled clinical trial of psychoanalytic psychotherapy for panic disorder. *Am J Psychiatry*. 2007;164(2):265–272.

Nakatani E, Nakgawa A, Ohara Y, et al. Effects of behavior therapy on regional cerebral blood flow in obsessive–compulsive disorder. *Psychiatry Res Neuroimaging*. 2003;124(2):113–120.

*Peterson BS. Clinical neuroscience and imaging studies of core psychoanalytic constructs. *Clin Neurosci Res*. 2005;4:349–365.

Reneses B, Galian M, Serrano R, et al. A new time limited psychotherapy for BPD: preliminary results of a randomized and controlled trial. *Actas Esp Psiquiatria*. 2013;41(3):139–148.

Ritchey M, Dolcos F, Eddington KM, Strauman TJ, Cabeza R. Neural correlates of emotional processing in depression: changes with cognitive behavioral therapy and predictors of treatment response. *J Psychiatr Res*. 2011;45(5):577–587.

Roffman JL, Marci CD, Glick DM, Dougherty DD, Rauch SL. Neuroimaging and the functional neuroanatomy of psychotherapy. *Psychol Med*. 2005;35(10):1385–1398.

Sifneos P. *Short-Term Anxiety Provoking Psychotherapy: A Treatment Manual*. Basic Books; 1992.

Smith ML, Glass GV, Miller TI. *The Benefits of Psychotherapy*. John Hopkins University Press; 1980.

Trujillo M. Treating pathological narcissism with short-term dynamic psychotherapy. In: Ogrodniczuk JS, ed. *Understanding and Treating Pathological Narcissism*. American Psychological Association; 2013.

Trujillo M. Short-term dynamic psychotherapy of narcissistic personality disorders. In: Kaslow FW, ed. *Comprehensive Handbook of Psychotherapy*. Vol. I. Wiley; 2002.

Winston A, Pollack J, McCullough L, et al. Brief psychotherapy of personality disorders. *J Nerv Ment Dis*. 1991;179(4):188–193.

32.6 Individual Psychodynamic Psychotherapy and Cognitive-Behavioral Therapy for Psychosis

MICHAEL D. GARRETT, M.D.

This chapter reviews the individual psychodynamic psychotherapy (PDTp) and cognitive-behavioral therapy (CBTp) of psychosis. It is organized historically, first describing clinically based psychoanalytic theories developed in the last century, then covering the newer, more research-based CBTp approaches that have evolved in the past 25 years, and finally suggesting the practical integration of both bodies of knowledge.

There is much overlap in these approaches, and there are some significant differences. Both place central importance on a close, collaborative therapeutic relationship, active listening, and a technique guided by psychological theory. PDTp, framed in a theory that assumes an open-ended receptive technique, places greater emphasis on psychological conflict, dissociated affect, fantasy, and the patient's past, while CBTp, employing a more structured technique, places more emphasis on cognition and behavior in the patient's present life. Each treatment pivots on a different foundation aiming to arrive at a similar place. PDTp seeks to bring insight and reduce suffering in the present by showing patients how psychotic symptoms are disguised expressions of emotionally charged experiences from the past. CBTp seeks to reduce suffering by marshalling the patient's capacity to examine evidence to change delusional beliefs. PDTp therapists focus on the emotional truth of psychotic symptoms; CBTp therapists focus on their literal falsity. These approaches meet and can be integrated where the heart meets the mind. Both may be of value to the same patient; at different times in treatment, patients with different needs may benefit from either approach separately or in combination.

We begin with two clinical vignettes, one of a man with psychosis treated with PDTp and one of a woman treated with CBTp. Both had a similar core psychological schema: abused as children, they were left with a psychological legacy of fear, need, rage, and expectations of mistreatment and abandonment that made it very difficult to establish stable, satisfying adult relationships.

First the patient treated with PDTp.

THE MAN WHO CHOKED PEOPLE

Bertram Karon was invited to provide psychotherapy for a man with chronic psychosis, who had been hospitalized for 10 years, a man with a history of unprovoked violence against other people in the hospital. The patient was grossly incoherent, and when he became coherent, he stuttered very badly. Review of his chart revealed that he came from a poor family, his father was an alcoholic, he had developed the stutter as an adolescent, and he had reported to sick call in the Army with a venereal infection of his mouth just before his first assault on a stranger. He would creep up behind other patients and choke them, but he had never killed anyone, as he would drop his victims when he had rendered them unconscious. After establishing rapport with the patient, in the first session Karon discovered that when he was a little boy, his mother, for minor offenses like not eating, would put a cloth around his neck and choke him. After the patient had revealed this piece of history to the therapist, he stopped choking other patients.

The patient began a subsequent therapy hour by yelling, “Why did you do it to me, Dad?” Exploring the psychotic transference, Karon learned that the patient’s alcoholic father had come home drunk one day and anally raped him. The patient’s terrible stutter was revealed to have an extraordinary cause. Latin words were discernable in the middle of his stutter. When asked if he had been an altar boy, he said, “You swallow a snake, and then you stutter. You mustn’t let anyone know.” He was extremely ashamed and guilty at having been compelled to performed fellatio on a priest. The therapist offered direct reassurance that what he had done was all right, and said, “Anyone as hungry (for love) as you were would have done the same thing.” At that point the stuttering stopped. When he started to stutter in later sessions, it was only necessary to repeat this interpretation, and the stuttering immediately ceased. As a boy the patient had turned from his mother who choked him to a father who raped him to a priest who sexually abused him. This succession of horrors drove him out of his mind. Yet his hospital record contained not one word of this history, and 10 years of confinement and treatment as usual had done nothing to allay his primary symptoms. This case illustrates how seemingly bizarre symptoms can be understood to have meaning, and how psychotherapy can ameliorate them. There are many such reports of successful psychotherapy scattered through the literature.

Now the patient treated with CBTp by the author.

THE WOMAN WHOSE CAT SOUGHT TO MURDER HER

The patient was a 40-year-old woman who had made several serious suicide attempts. She reported feeling depressed, hopeless, and terrified when she came to believe that her cat was planning to murder her in her sleep. In the weeks prior to the emergence of this delusion she had been feeling unusually needy and had begun telephoning her boyfriend at work many times a day, until he began to set limits on the phone calls. She interpreted this as meaning he no longer loved her. She blamed her cat for the loss of his love.

At night when she and her boyfriend went to bed, the cat would snuggle up between them, which the patient experienced as the cat "coming between" her and her boyfriend. The boyfriend showed the cat affection, which indicated to the patient that the cat's plan to steal her boyfriend's love was working. She believed the cat planned to murder her in order to have her partner all to itself. The psychodynamic meaning of the delusion was readily apparent. Having been abused as a child and hungry for love, she felt a compelling need to be in sole possession of her boyfriend's affection.

The therapist inquired about the details of how the patient imagined the cat would kill her. The patient responded as though this should be perfectly self-evident. "You see it all the time on TV! Lions and tigers bite their prey in the neck. They bite the jugular!" The patient expected the cat to bite her jugular when she was asleep. While respecting the patient's conviction that the attack was inevitable, the therapist inquired in an off-hand manner if the patient knew where her jugular was located. She said she knew it was somewhere in the neck but did not know exactly where. The therapist commented, again in an off-hand manner, that the cat would have to have a much more precise knowledge of human anatomy than she did. While speculating where in the neck the jugular was the patient acknowledged that her being overweight and quite thick in the neck would make it especially difficult for her cat to locate her jugular. The therapist opined that the cat would certainly have one and only one chance to attack her precisely at the jugular, because as soon as she felt the cat's teeth on her neck she would surely awaken immediately and throw the cat off, compress the bite on her neck, and call an ambulance. The patient returned the next session less anxious and less depressed. When the therapist explored her change in mood, she said, “I still believe my cat plans to kill me, but I no longer believe she has the means to do it.” This positive shift in mood and decrease in depression, while not resolving the delusion entirely, was a significant gain obtained in a single session. The psychotherapy then continued to explore her relationship with her boyfriend.

Chronic psychosis is the cause of enormous suffering and individual disability. In contrast to significant advances in the treatment of heart disease, cancer, and stroke that have dramatically improved quality of life and life expectancy, most chronically psychotic patients remain functionally disabled their entire lives and live 15 to 20 years less than the general population. While medication is undoubtedly helpful to some patients at some times in some ways, compared with other branches of medicine, therapeutic outcomes in the treatment of chronic psychosis leave room for improvement. Enter and add psychotherapy.

Psychotherapy (broadly defined) is a disciplined professional interaction between a trained psychotherapist and a patient that aims to improve the patient’s quality of life by exploring the patient’s thoughts, feelings, and behaviors, with the goal of facilitating new possibilities for emotional growth and more adaptive thinking and functioning. This chapter will assume that the reader has at least some passing familiarity with such essential concepts in PDT as the existence of unconscious mental processes, unconscious fantasy, transference, and the therapist’s interpretation of psychological meaning, and CBT concepts including the existence of cognitive biases in thinking and the A-B-C cognitive sequence in which an activating event is followed by a belief about the meaning of that event, which in turn leads to an emotional and/or behavioral consequence.

Although psychotherapy is widely practiced in the treatment of non-psychotic patients, psychotherapy has for decades taken a back seat to pharmacologic treatments for psychosis. The reason psychotherapy for psychosis is so little practiced is not that psychotherapy has no value. Research to be cited in a later section of this chapter demonstrates its efficacy. Why has psychotherapy lagged so far behind medication as a frontline treatment? In the author’s view, the following issues have contributed to marginalizing psychotherapy for psychosis.

1. The biggest obstacle to providing psychotherapy to patients with psychosis is the widespread belief that psychotherapy is of no value. The stunning initial successes in the work of pioneers like Harry Stack Sullivan, Frieda Fromm-Reichmann, and Silvano Arieti and the solid evidence base for CBTp should have put this mistaken idea to rest, but few people have been trained in these approaches. It is not surprising that approaches that are helpful to other patients are insufficient with patients suffering from psychosis and have left clinicians with feelings of inadequacy, failure, and pessimism about therapy for psychosis.
2. The same symptom picture might be diagnosed as “brief reactive psychosis” or “schizophrenia” depending solely on length of illness. This convention implies that schizophrenia is, by definition, a bear of a disease that never gets better. Such fatalism embedded in diagnostic practice allows clinicians to be complacent about poor outcomes rather than to press hard for better treatments.
3. Understanding the psychology of psychosis requires the clinician to empathize with levels of anguish not often encountered in everyday life. Clinicians may take for granted that psychotherapy can help patients with non-psychotic depression and anxiety because they know from personal experience that talking things through with the right person can help, but when they see little connection between ordinary mental life and psychotic symptoms, the utility of talk therapy for psychosis may not be intuitively obvious.
4. The introduction of the neuroleptic drug chlorpromazine (Thorazine) in 1954 led to the dramatic amelioration of acute psychotic symptoms in many patients. The observation that effective neuroleptics diminish dopamine transmission led to the dopamine theory of schizophrenia, which added credibility to the idea that schizophrenia is simply a brain disease, and as such is more likely to respond to medication than to psychotherapy.
5. The obvious efficacy of neuroleptics in relieving some psychotic symptoms fueled a backlash against the unfortunate hypothesis that the “schizophrenogenic mother” was the cause of psychotic illness, but this needed course correction went too far in the other direction, lending support to the idea that psychological factors are of little importance.
6. Freud did not believe that psychotic patients formed a transference that could be analyzed in a conventional psychoanalytic treatment, and so had no faith in the efficacy of psychoanalysis, at least as he

practiced it, for psychosis. With some notable exceptions, the psychoanalytic community mirrored Freud’s attitude and largely abandoned the care of the severely mentally ill.

7. PDTp failed to establish itself not only because it was crowded from the scene by biologic treatments, but because a classical psychodynamic approach is rarely effective without a modification of technique. Psychodynamic therapists have paid too much attention too early to the unconscious psychological meanings of psychotic symptoms, and too little attention to the patient’s conscious experience of the psychotic symptoms. Pioneers of PDTp have often been mythologized as exotically talented clinicians who could do what the average well-trained clinician could not. In fact, doing effective PDTp and CBTp is well within the range of most clinicians in public clinics when they are provided with specialized training and supervision.

The time has come to give renewed consideration to the place of individual psychotherapy in the treatment of psychotic individuals. Not only is individual psychotherapy often effective, but there is also mounting evidence of the limitations of neuroscience and psychopharmacology. A recent review by a group of prominent neuroscientists (Stephan et al., 2016) concludes “… few, if any, breakthroughs in basic scientific research have led to substantive improvements in psychiatric clinical practice.” What is psychiatry missing? The CATIE study found that newer medications are not decisively more effective than older drugs and cause serious side effects of their own, including metabolic syndrome and diabetes. Two recent studies suggest that structural changes in the brain correlate not only with total period of untreated psychosis but also with total exposure to neuroleptics, that is, both psychosis and the drugs used to treat psychosis may damage the brain. A recent review of the traumagenic neurodevelopmental hypothesis suggests that the brain changes typically attributed to a primary brain disease may reflect changes in the hypothalamic-pituitary-adrenal axis and dopamine system in response to psychological trauma. These studies suggest that psychiatrists should aim for the lowest dose of medication possible for the shortest period of time. Psychotherapy may help to achieve this goal.

In the author’s experience, effective psychotherapy can reduce the necessity for dosage increase, poly-pharmacy, and inpatient admission. Consider an example that illustrates the possibilities for combining medication and psychotherapy.

THE WOMAN WHO HAS NEVER BEEN RE-HOSPITALIZED

A 30-year-old woman with a chronic paranoid psychosis with a 6-year history of treatment as usual, with 3 to 4 hospitalizations a year and intermittent homelessness, was referred to the author for psychotherapy. She had failed multiple neuroleptics including clozapine (Clozaril) and was being prescribed a high dose of aripiprazol (Abilify) at the time of initiating psychotherapy. In the course of a 5-month, once-a-week CBTp treatment, the patient understood for the first time that her belief that she was being pursued by thugs, her conviction that she was being monitored through a computer chip in her body, and her experience of “voices” telling her she was a child abuser were all manifestations of an illness (the convention “voices” will refer to verbal auditory hallucinations). Aripiprazol was titrated downward to a lower dose but could not be stopped without a flare-up of psychotic symptoms. She attends session every other week and has never been readmitted. During her 18 years in treatment she suffered two significant recurrences of paranoid ideas and ideas of reference, both of which were worked through psychotherapeutically without recourse to increased medication or hospitalization.

HISTORY

In the interest of brevity, one can divide the history of PDTp after its beginnings in the early 1900s into two schools, a European Kleinian School heavily influenced by the ideas of Melanie Klein that was active in the first half of the 20th century, represented by Hannah Segal, Wilfred Bion, and Herbert Rosenfeld, and an American Interpersonal School based on the ideas of Harry Stack Sullivan that has been influential from the mid-20th century on, represented by Harry Stack Sullivan, Frieda Fromm-Reichmann, Silvano Arieti, and Harold Searles. See the review by Michael Stone for a more comprehensive listing of contributors. The history of CBTp, which began in the early 1990s, will be covered in a later section, “Cognitive-Behavioral Therapy for Psychosis.”

Psychodynamic thinking about schizophrenia began at the Burghölzli Mental Hospital in Zurich, Switzerland when Eugen Bleuler, M.D. became Director in 1898. He emphasized the importance of psychological "complexes" in the formation of psychotic symptoms. Carl Jung worked there during Bleuler's tenure. Jung recalled that most psychiatrists at the time paid attention to diagnosis but ignored the meaning of psychotic symptoms.

Jung described a psychotic woman who had spent many years in the hospital and made stereotyped, repetitive, seemingly meaningless gestures. He read her chart and found an entry that was years prior describing these gestures as a "cobbler's motion," that is, like the movement a cobbler would make drawing a thread through leather in a shoe he was holding between his knees. The patient died shortly after Jung met her, and her elder brother came for the funeral. Jung asked him, "Why did your sister lose her sanity?" The brother told Jung that she had been in love with a shoemaker who had not wanted to marry her, and when he rejected her, she had "gone off the deep end." Jung concluded that the shoemaker's movements represented the heartbroken patient's identification with her beloved cobbler, a way of keeping her sweetheart with her psychologically despite his having rejected her in reality. In short, he broke her heart, she lost her mind, and she remained "married" to him in gesture until death did her part.

Paul Federn was one of the few analysts in Freud's circle with significant experience treating psychotic patients. He emphasized the concept of the "ego boundary," a fundamental idea in the psychoanalytic theory of psychosis. Federn believed that the fundamental problem in schizophrenia was that the ego boundary that ordinarily separates thoughts and feelings from perceptions is faulty, allowing the content of the unconscious to spill into consciousness, where it is perceived as reality.

THEORETICAL ISSUES

In both psychodynamic and CBT literature on borderline personality disorders (e.g., work by Kernberg, Fonagy, Meares, Linehan, and Young), one finds research-based efforts emphasizing specific deviations from ordinary psychoanalytic or cognitive-behavioral treatment required for that population. Although there is a small, clinically based psychoanalytic literature on “supportive therapy” and treatments of schizophrenia, comparable refinements of theory and technique have yet to be fully realized and widely practiced in the treatment of individuals with psychosis.

As clinical technique should derive from research-supported theory, let us turn to the psychodynamic theory of psychosis. Despite disagreements about the relative importance of biologic and psychological factors in the etiology of psychosis, there is substantial agreement among psychodynamic theoreticians about its fundamental psychology: *Psychotic states represent the intrusion into adult mental life of primitive psychological processes that are ordinary in childhood; these early processes normally recede from consciousness by the time a person reaches later childhood (the “latency” period) but they re-emerge in psychosis*.

These intrusions of early mental life are of two sorts. First, individuals in psychotic states utilize *primitive cognitive processes* similar to those operative in the first years of life. Instead of thinking a thought and expressing it in verbal symbols, or feeling a feeling that is subjectively located within the psychological boundary of the self, psychotically disturbed individuals experience a body sensation or experience a distorted perception of the outside world that expresses feelings and thoughts that a non-psychotic person would express in words. For example, instead of thinking the thought, “I am a bad person” a psychotic individual might *perceive* a “voice” saying, “You are a bad person.” Federn called this failure to distinguish between subjective experience and perceptions of the outside world a *breakdown of the ego boundary*. This confusion of thought, feeling, and perception plays a significant role in the formation of ideas of reference, hallucinations, and delusional beliefs.

Second, *primitive object-related phantasies* (more on this language later) that ordinarily operate in the first years of life re-enter consciousness in psychosis and dominate the person’s experience of relationship with other people and the world, shaping the content of delusions and hallucinations and the narratives that patients construct to explain their lives. For example, while normal children might play at being king or queen, treating their parents as vassal subjects while resisting efforts to spoil the fun and expose the game as play-acting, people with schizophrenia might make a literal claim to being a king or queen, insisting that family members and clinical staff bow to their will, and resisting efforts to characterize their belief as illusory.

What remains unconscious in psychosis and in non-psychotic conditions, respectively, is different. The psychotic person, while closer to primitive childhood mental life, is not consciously aware of using archaic mental processes to keep painful thoughts and feelings about real life outside conscious awareness.

Although there is widespread psychoanalytic agreement about the intrusion of primitive psychological processes in psychosis, historically there has been disagreement about how to deal with these intrusions psychotherapeutically (see Table 32.6–1). Because Federn viewed the psychotic process as a defect in the brain's ability to invest internal mental events with sufficient "ego feeling" to mark them as belonging to the self, he advised clinicians not to encourage the emergence of unconscious material and to harbor no ambition to interpret unconscious processes. Instead, they should attempt to compensate for the patient's defective ego boundary by supportive reality testing. His approach is echoed in current-day psychosocial programs such as day hospitals, supportive housing, and employment arrangements that center treatment in a clinical team rather than a single therapist. Federn's concern that exploratory psychodynamic treatment might undermine the reality-testing of psychotic individuals is mirrored in the empirically unproven belief of some in the 1980s that psychotherapy for patients with psychosis is ineffective or even harmful.

Followers of Melanie Klein, in contrast, advocated a direct interpretation of primitive thinking with no modification of classical psychoanalytic technique. They regarded the "psychotic transferences" that then emerged in treatment as important opportunities to make interpretations revealing the unconscious ideas dominating patients' experiences.

Also, in contradiction to Federn's ideas, analysts in the American Interpersonal school of Harry Stack Sullivan sought, above all, to establish trusting relationships with patients that would facilitate their ability to open a dialogue. They did not encourage the development of a psychotic transference, however, nor regard it as the essential focus of therapeutic change. Rather, they would interpret transference distortions when they interfered with the therapeutic alliance. Sullivanians made interpretations at the interface between patients' subjective experiences of painful situations in childhood, latency, and adolescence and the meaningful expression of these experiences in the form of symptoms such as delusions and hallucinations.

Table 32.6–1.

Supportive, Kleinian, and Sullivanian Approaches to Psychotherapy for Psychosis

	Focus of Theory	Focus of Interpretations	Handling of the Transference
Federn and the current clinical zeitgeist	A defective ego filter that fails to bar primitive unconscious content from consciousness	Do not encourage the emergence of unconscious content with the aim of interpreting it. Develop a supportive relationship that eventually allows the therapist to encourage reality testing.	Try to avoid a psychotic transference by avoiding free association and including other people as therapy aides to diffuse the intensity of the relationship with the therapist.
European Kleinian School Klein, Segal, Bion, Rosenfeld	Primitive unconscious object-related phantasies	Interpret the primitive unconscious phantasy that operates through mechanisms of projective identification.	Welcome the psychotic transference with the intent to interpret it as an essential opportunity for therapeutic change.
American Sullivanian School Sullivan, Fromm-Reichman, Arieti, Searles	Failure to achieve independence and stable social competency within one's core family group	Show the patient how psychotic symptoms are meaningful reflections of painful early interpersonal failures.	Diminish the blank screen of classical analytic technique that encourages a psychotic transference by making a special effort to establish a real

trusting
relationship
with the
patient.

THE PSYCHOLOGY OF PSYCHOSIS

These pioneers, despite their differences, all tried to reduce the apparent distance between psychotic processes and ordinary human psychology. If one can't understand psychotic states psychologically, how can one hope to treat them psychologically? How can one relate the seemingly bizarre symptoms of psychosis to ordinary human psychology, and in so doing forward a psychological theory of psychosis and a rationale for psychological treatment? Addressing these questions begins in an understanding of the psychology of young children that can be inferred from direct child observation and clinical experience.

The experience of newborns and infants under 4 months old, as best we can determine, consists of isolated fragments of somatic sensations associated with perceptual images that have not yet been integrated within the boundary of an organized self. The newborn does not know yet that the hand in its immediate visual field of vision is *its* hand. Children do not recognize themselves in a mirror until roughly 18 months. Very young children do not know that sensations are part of their own personal subjective experience. As development proceeds, consciousness moves beyond somatic perceptions to increasingly complex "mentalized" forms of experience, allowing for symbolic representations, verbal thought, and conscious and unconscious phantasy.

Klein came to believe that early in life, experiences are organized and represented mentally in "phantasy" as *internal object relationships* between psychological "objects," where the self and an "other" (the *object* of the child's interest) are connected by an affectively charged link: the self is doing something to the object, or the object is doing something to the self. This chapter adopts her use of "phantasy" to refer to primitive unconscious phantasies, and "fantasy" to refer to conscious daydreaming. Given that human beings are fundamentally interdependent social animals, it is not surprising that human phantasy is organized in an object-related form. In object relations theory the primitive self defends against painful mental experience by splitting into two primal psychological objects, a pristine self that attempts to preserve its freedom from pain by identifying itself with a natural state of being phantasied to be pain-free, and a persecutory object that is causing the pain by intruding upon what would otherwise be the satisfactions of the primitive self. In this construction, human suffering does not arise from internal conflict, but is rather caused by persecutory attacks of elements experienced as outside the self.

This basic tendency to split off the pristine self from a persecutory object is apparent in childhood stories and their adult derivatives (Superman is all good; he does not cheat on his taxes. Lex Luther is all bad; he doesn't spend weekends volunteering at soup kitchens for the homeless). This psychology leaves many residuals in adult mental life. For example, during a headache we may say, "My head is killing me." We are imagining two psychological objects, a "me" that in its natural state would be headache-free were it not for the murderous activity of the head, and a "head," a persecutory object intruding pain into the otherwise pain-free "me." In psychotic illness, this split between the self and a persecutory object becomes the basic psychological structure of a paranoid delusion. The fundamental tendency to split the mind into contrasting mental representations of self and persecutory object is the source of the ubiquitous human tendency to blame others for our failings and sorrows; for example, the train is responsible for my lateness (not my failure to leave early enough); a particular ethnic group is responsible for society's social ills, and so on.

Child observation suggests that phantasy begins pre-verbally. Susan Isaacs described a 20-month-old child with delayed speech development who saw her mother wearing a shoe with a loose sole that was flapping about. The child screamed in terror, and for a week would shrink away if she saw her mother wearing any shoes at all. At 2 years 11 months, when she was able to express herself verbally, she suddenly asked in a frightened voice, "Where are Mummy's broken shoes?" Her mother reassured her that she had sent the frightening shoes away, to which the child replied "They might have eaten me right up." The child at 20 months had seen the shoe as a threatening mouth but could not express this phantasy verbally. Her conception of the shoe as a dangerous mouth can have no other origin than her psychology. Her oral aggressive impulses were projected outside the self and experienced as though residing in the shoe. In this phantasy, the internal danger posed by the child's aggression is transformed into a danger threatening from outside the self; that is, an intra-psychic conflict has become an interpersonal problem between two internal objects, with the persecutory object experienced as located in the outside world. A similar process happens in psychosis, when a man in a state of great internal turmoil conceives his difficulty as an external threat from the Mafia or CIA.

In the first 3 years of life, the child's conception of the world is dominated by projected elements of its own mental life; for example, the shoe is a mouth. This developmental period can be characterized as a two-person psychology in which the child feels inherently related to particular persons. Young children live in a psychic world in which what they feel and think is experienced as real. The girl who was frightened of the shoe was not telling her mother "I am having a phantasy that your shoe is a scary mouth, but I know it is just a shoe." Rather, she *saw* the shoe as a threatening mouth. The child's state of feeling *is* the child's reality during this early phase, and its perception of the world is conditioned by its emotions. This is also the case in psychosis.

Peter Fonagy and Mary Target distinguish two modes of cognitive functioning in young children, the stage of "psychic equivalence" in which whatever is psychically experienced is felt as real, and the "pretend mode," in which ideas are representational but their correspondence with reality is not examined, allowing children to play at pretend games. Consistent with the preoccupations of young children, the primitive object-related phantasies that predominate in the first 3 years of life often have oral and anal themes, such as shoes that bite. The child's experience of eating becomes the body-based metaphor for the psychological processes of *introjection* and *identification*, via which aspects of the environment are taken into the self; defecating becomes the body metaphor for *projection*, in which aspects of the self are extruded and deposited in mental representations of psychological objects outside the self.

In effect, in the service of defense, the mind appropriates the body processes of taking in and pushing out content and transforms these processes into psychological defenses. At this maturational stage, the mind is like an open international border where heavy trucks rumble back and forth day and night, exporting projected psychic content and importing it back again after it has been modified by the external environment. Residuals of these early, body-grounded mental processes remain in adult metaphors such as “That’s food for thought!”, where the comprehension of an idea is conceived as the ingestion of a satisfying food, a sort of mental “good breast” experience. When we say, “I had to swallow my pride” we mean “I didn’t abandon my pride. Rather, I temporarily hid it in an interior body space from which it can safely re-emerge when the coast is clear.”

If young children are to achieve a firm sense of being real in a real world, they must grow out of their experience of the world as projected fragments of the self. How does this happen? Donald Winnicott, a practicing pediatrician and psychoanalyst, described how a secure attachment to an empathic mothering figure becomes the developmental base for a child’s experience of reality. Winnicott’s “good-enough mother” responds empathically to the child’s authentic strivings while not imposing an alien agenda on the child’s gradually emerging unique identity. The child’s earliest experience of the external world is its encounter with the mother’s body, especially the breast. The empathic mother, attuned to the baby, learns to read the infant’s communications so as to feed the baby when it is hungry, and change the diaper when it is wet, and not the other way around.

As described by Winnicott, we can fancifully imagine the baby’s experience of hunger and feeding as, “I was feeling uncomfortable just now. I was starting to imagine something that would eliminate the discomfort when suddenly, as if by magic, the breast appeared. Then I knew I was trying to imagine something like a breast all along. I am of this world. I am in sync with this world. The world sees my existence.” The empathic mother delivers the real world to the child at an opportune moment of need and thereby builds a bridge between the child’s psychic reality and the real external world.

Children who find sufficiently resonant correspondences between their mental states and the outside world turn toward reality and begin to feel real in a real world. A child whose authentic stirrings are not met empathically may turn inward and begin living as a “false self” that merely reacts to the impingements of others rather than expressing its authentic nature. If the world returns no resonant ping to a child’s sonar probe, the child adapts to living in a dead world that appears to not know the child’s true self is alive. In some family systems, the child is recruited to fulfill the psychological needs of others in the family, in which case the family environment impinges upon and distorts the development of the child’s authentic individuality. This impingement beggars the self, leaving the child little base from which to build a sustaining sense of interpersonal competence and independence, a state of vulnerability that can predispose to psychosis.

Some children suffer not only impingements of their authentic identity but overt abuse. Research in the last two decades has shown that adverse life experiences in childhood or young adult life, including sexual and physical abuse, physical and emotional neglect, and being bullied, are associated with a two- to threefold increase in the incidence of psychosis in adulthood. The fact that the risk for delusions has been seen to increase more than 10-fold in people who have been raped and who also manifest symptoms of PTSD suggests that what is typically diagnosed as schizophrenia may, for many people, be a form of complex PTSD in individuals biologically predisposed to the neuropsychological aberrances observed in psychosis. Auditory hallucinations appear to have a particular association with sexual abuse. Several research studies that have looked for links between trauma and the content of psychotic symptoms have all found that the content of psychotic symptoms often reflects the historical details and themes of abuse particular to the individual. For example, Corstens and Longden found that in 94% of 100 cases of auditory hallucinations, it was possible to clearly formulate the underlying emotional conflicts resulting from trauma that was embodied in the voices. Such links between the content of psychotic symptoms and trauma have important implications for the psychotherapeutic treatment of psychosis because they demonstrate that psychotic symptoms should not be regarded as meaningless products of a broken brain but rather displaced and distorted memories of traumatic experiences that have reshaped the psychotic person’s experience of their social reality.

In secure, empathic environments, the primitive unconscious phantasies of early childhood fade over time, but when the social environment impinges upon a child’s authentic self when these phantasies are active, they do not fade. Cumulative disappointments give rise to painful conscious fantasies and unconscious phantasies that are reinforced. These will influence how later life experience is perceived and given meaning, and may leave people with such experiences feeling they have no space for tolerable living. What others have described as the pre-psychotic child’s difficulty with separation and individuation from a symbiotic relationship with the mother can be viewed as a massive impingement of caregivers on the authentic independence of the child.

Recall the little girl who was frightened that her mother’s shoe would eat her. Oral experiences are important to young children, who often phantasize about biting and eating, and being bitten and eaten. Frightening oral themes of this sort appear in fairy tales and also in psychotic delusions. In “Hansel and Gretel,” a famine descends upon the land, and a terrible mother plans to leave her two children to die in order to conserve the limited food supplies for herself. The children narrowly escape being eaten by a witch. Little Red Riding Hood and her grandmother are both eaten by a wolf who has assumed the grandmother’s clothing.

Compare the internal world of a psychotic man from a chaotic family with a drug-addicted mother and absent father, who suffered a delusion composed of primitive oral phantasies. He reported that he had developed a cancer when at birth his mother touched his penis (the mother has put a “persecutory bad object” into him in the form of the cancer), and that this cancer now coated all his internal organs (the “bad object” is about to engulf him). Because he feared that medication would disturb a delicate balance between his organs and the cancer, giving the cancer the upper hand, he refused all pharmacologic treatment. In phantasy, taking medication would destabilize his resistance to the persecutory object. This man was frozen in a state of terror in which he feared his true self would disappear inside a bad object (the cancer) and be annihilated. In effect, his delusion is a horrifying version of “Hansel and Gretel” and “Little Red Riding Hood” with no happy ending, a personal fairy tale with a readership of one.

Similarly, a 30-year-old man with two prior psychotic episodes reported in a first psychiatric consultation an intense anxiety he associated with his belief that a terrible drought was about to occur, which had led him to fill his apartment with many gallons of stockpiled water. At the end of the consultation, the patient said, “You have asked me a lot of questions. Can I ask you some questions?” The therapist said he would be happy to answer any questions the patient had relating to his treatment. The man then asked how old the therapist was, was he married, did he have children, and where did he live? Before offering any answer the therapist stated, “I am wondering why it is important for you to know the answer to these particular questions.” The patient responded in a matter-of-fact tone as though the reason were perfectly obvious. “If I am to model myself after you, I must know everything about you! I must know details.”

A primitive phantasy of being deprived of essential oral supplies, akin to the famine of fairy tales, emerged in his delusion of an impending drought. He imagined his relationship with the therapist as equivalent to swallowing him whole; that is, indiscriminately absorbing attributes of the therapist’s identity into himself. Before swallowing the therapist and taking on his identity (just as the wolf swallowed Little Red’s grandmother and assumed her identity) the patient sought to determine in detail what the therapist was made of, so as not to ingest a toxic persecutor inadvertently. Indeed, the devil is in the details. In psychosis there is regression to Fonagy’s “psychic equivalence” mode, in which what is felt to be real is experienced as being real.

Segal called this cognitive process “symbolic equivalence”; other writers have referred to “concrete thinking” or the experience of projected parts of the personality as “things.” Segal describes a women who, when non-psychotic, wrote a story depicting her mother and the analyst as being *like* a Lancashire witch (a fantasy), but when psychotic, believed the analyst *was* a Lancashire witch (a concrete metaphor that constitutes a delusion). Franz Kafka’s famous story *The Metamorphosis*, in which Gregor Samsa comes to experience himself as a monstrous insect, can be viewed as a psychotic sequence beginning with Samsa’s feeling of being a nobody, an anonymous functionary in an industrial society, thus being *like* an insect, then experienced as the concrete metaphor that he *is* an insect. In the "psychic equivalence" mode, if two people or things share a common attribute, they are experienced not as being similar, but as being the same.

This primitive mode of psychic functioning can be observed in young children trying to figure out what categories different objects belong to. One imaginative 3-year-old who was delighted to receive a paper origami bird requested "More ohwogamee!," but burst into tears when offered an origami flower. "That not ohwogamee!" Are all instances of "ohwogamee" avian? Does a folded paper bird belong to the category "paper" or "bird"? In psychosis, instead of figurative metaphorical thoughts of the "X is like Y" variety that form a basis for consensual symbols, individuals aggregate disparate objects and people that share some quality of emotional significance to themselves in a single identity, like a child sorting experience according to categories different from those of most adults.

Such "X = Y" statements, when reported to clinicians, are seen as delusions. For example, a man diagnosed with schizophrenia reported that a nurse's aide on the inpatient unit, who had a last name with four letters in common with his mother's name, was in fact his mother in disguise. This primitive type of cognition forms a basis for a variety of delusional beliefs and ideas of reference in psychosis.

In his description of what others have called "concrete thinking," Arieti describes how in psychosis, thinking regresses to "paleologic" or "predicate logic" (according to the Von Domarus principle), where parts or attributes of a person or thing become equated with the whole. This regression in psychosis has a psychodynamic purpose. When people in psychotic states try to think about their lives and feelings, clear thinking that maintains discernable symbolic emotionally charged connections between thoughts and events can be unbearably painful. Concrete thinking damps down the symbolic emotional resonance of thought, thereby blurring its clarity and emotional significance, making it harder for the patient to think clearly to a painful endpoint.

Concrete thinking also turns attention away from intolerable internal experience, outward to the perception of concrete objects and external entities. But the suffering that the patient wishes to avoid nevertheless persists in disguised form, hidden in plain sight in the suffering associated with psychotic symptoms. A psychotic person continues to suffer mightily, but achieves a modicum of relief by displacing the suffering away from the core interpersonal relationships that gave rise to the pain and projecting it outward into mental representations of the real world.

For example, a psychotic woman who was conflicted about her wish for a life insurance payout following her husband's death developed a delusion that her father-in-law intended to kill her to steal the money. Hospital staff were unable to confirm the existence of any insurance policy, and there was no evidence the father-in-law wished her harm. She reported seeing a red car drive by her home and concluded that the driver was her father-in-law (who had a red car), circling in the neighborhood with plans to kill her. Her paleologic equation of any driver of any red car with her father-in-law degraded her ability to think clearly in her bereavement. Instead of feeling her pain over her husband's failure to provide a comfortable inheritance for her, she dealt with her guilt over her wish to profit from her husband's death by becoming preoccupied with an external danger. Greed and murderous feeling have gone missing from her mind and reappeared in the imagined mind of her father-in-law.

In the 4th and 5th years of life, the "psychic equivalence" and "pretend" modes evolve into a more fully developed "theory of mind," in which the child can distinguish between dream images, thoughts, and real things, and appreciate that others may have thoughts and feelings that different from those of the child. The stage is set for the three-person psychology of Freud's oedipal period, where the child registers the parents' relations to each other and also begins relating to a wider network of people. Table 32.6–2 depicts the normal course of development of children: mental life before age 3 is dominated by primitive phantasies and metaphors for body processes, especially taking in external objects and extruding internal ones. As the child grows up, the fantastic, primitive two-person psychology of early childhood is gradually replaced by more realistic mental representations of others.

Examples of the disappearance of fantastic conceptions of the world in favor of more realistic representations include the disappearance, normally around age 8, of beliefs in Santa Claus, the Easter Bunny, and the Tooth Fairy. Just as the Wizard of Oz is revealed as an ordinary man operating a hand crank, not a phenomenal being wielding a magic wand, such entities become understood as parental inventions. This change is shown in Table 32.6–2 as the gradual increase of more realistic mental representations of others, progressing through latency and adolescence into young adulthood.

Table 32.6–2.
Stages of Development and Internal Object Relations

Stages of Development	States of Internal Object Relations
AGE 18 MONTHS Psychology of the primal self Primary anxiety: annihilation of self Primary defense: splitting, projection, denial Fonagy—psychic equivalence mode Klein—paranoid schizoid position Mahler—symbiosis Erikson—basic trust	Primitive internal object relations consisting of split-off fragments of the self that are projected into mental representations of others.
AGE 3 YEARS Two-person psychology of attachment Primary anxiety: loss of mother; shame Primary defense: projective identification Bowlby—attachment Fonagy—pretend mode Klein—depressive position Mahler—rapprochement and individuation Erikson—autonomy	Diminishing influence of primitive internal object-related phantasies. The ability to pretend and play with reality. *“Mother’s shoes might have eaten me.”*
AGE 6 YEARS Three-person (Oedipal) psychology Primary anxiety: body injury; guilt Primary defense: repression Fonagy—“mentalization” Klein—depressive position	More realistic multidimensional images of other people as separate individuals. The ability to “mentalize” the feelings and perspective of other people. *Hansel & Gretel/Little Red Riding Hood*
YOUNG ADULTHOOD	**PSYCHOSIS**: Primitive object-related

phantasies re-emerge in young adulthood.
The psychotic man who feared a drought:
“If I am going to model myself after you, I must know everything about you.”

THE IMPACT OF LESS THAN OPTIMAL FAMILY RELATIONS IN SOLIDIFYING PRIMITIVE UNCONSCIOUS PHANTASY

Tolstoy’s observation that “Happy families are all alike; every unhappy family is unhappy in its own way” might be rephrased here as “Happy children are all alike, but unhappy children who are vulnerable to developing a psychotic illness are each unhappy in their own way.” Because the failures in development that make people vulnerable to psychosis can take many forms, the roads to psychosis are as varied as there are different patterns of troubled development. Just as no single gene has been identified to account for all instances of psychosis, no single psychological dynamic has been implicated in all cases of psychosis. There are many fault lines along which the mind can shatter. In some cases there is a massive failure in the early caretaking environment that ushers in the psychosis, or an evident trauma such as physical or sexual abuse, rape, or persistent bullying at school, but in many cases no such massive trauma is apparent.

In ordinary mental life a person might remark in response to a severe upset, “I almost lost my mind!” implying they felt like they were approaching the abyss of madness without falling in, but close calls of this sort fail to convey the profound alteration of a psychotic person’s relationship to reality. Individuals who have suffered too grievous hurts in life culminating in a succession of social failures may not be able to step back from the abyss. Some autobiographical accounts of psychosis describe a near voluntary letting go of reality, a surrender to the enveloping tide of psychosis in the face of the seeming impossibility of meeting the ongoing demands of everyday life. To grasp the enormity of this transformation the average person might ask how much pain they would need to be in to wake up one morning and decide to not go to work, to renounce the prior premises of family life with one's partner, children, friends, and extended family, to live instead in a private world where one's urgent concerns cannot be shared with anyone for fear of being labelled crazy. The high incidence of psychosis in abused children suggests that mistreatment by caregivers results in a legacy of pain sufficiently intense to drive a person to abandon a consensual reality in which they are unable to live. In her book *I Never Promised You a Rose Garden*, Joanne Greenberg describes in detail such a flight from and return to reality over the course of her psychotherapy.

Arieti observed the so-called "schizophrenogenic mother" in less than 20% of families, and found that most parents of children who became psychotic were loving and concerned. Karon makes the important point that psychological processes in parents that may predispose to psychosis in children often operate outside the parents' conscious awareness. The caregivers fail to give the child what is needed but do not harm the child intentionally; in such cases both child and parent may be victims of the parent's unconscious psychology. In most cases, the erosion of good-enough development occurs gradually, one disappointment at a time, one failed interpersonal exchange after another, until the young person's psyche can no longer hold on. After having failed to achieve a positive standing in the peer group despite repeated efforts to do so, the about-to-be psychotic person one day surrenders and says simply, without much fanfare, "I'm done. I give up."

In her account of her psychosis and psychotherapy with Fromm-Reichmann in her book *I Never Promised You a Rose Garden*, Joanne Greenberg describes her mother and father as loving parents desperate to help her, but not knowing how. In her case, her psychosis issued from the accumulation of many factors—her younger sister's birth, traumatizing urologic surgery at age 5, the impinging presence of her charismatic, demanding maternal grandfather (who to a great extent defined her parents identities and accordingly hers), a move away from a familiar neighborhood where she felt safe, and the anti-semitic bullying she encountered in summer camp.

Greenberg describes her retreat from the real world, where she found it impossible to live, into a psychotic world she called Yr, inhabited by a host of magical delusional characters with their own language, personas as beautiful, compelling, and terrifying as in any fairy tale. She hears the “voice”of a hallucinated being named Anterrabae, who says to her “Bird-one, you are not of them.” Which confirms her conclusion that she is impossibly misfit to the real world. As she descends into psychosis, her sense of self disintegrates, and her mind drifts away from verbal thought that would require her to think clearly about her pain, toward a succession of images and sensations characteristic of the pre-verbal mental processes of early childhood.

> A black wind came up. The walls dissolved and the world became a combination of shadows. Seeking for the shadow of firm ground on which to stand, she was only deceived again when it warped away like a heat mirage; she looked toward a landfall and the wind blew it away. All direction became a lie. The laws of physics and solid matter were repealed and the experience of a lifetime of tactile sensation, motion, form, gravity, and light were invalidated. She did not know whether she was standing or sitting down, which way was upright, and from where the light, which was a stab as it touched her, was coming. She lost track of the parts of her body; where her arms were and how to move them. As sight went spinning erratically away and back, she tried to clutch at thoughts only to find that she had lost all memory of the English language and that even Yri was only gibberish. Memory went entirely, and then mind, and then there was only the faster and faster succession of sensations, unidentifiable without words or thoughts by which to hold them. These suggested something secret and horrible, but she could not catch what it was because there was at last no longer a responding self. The terror, now, could have no boundary.

(Excerpt from I NEVER PROMISED YOU A ROSE GARDEN by Joanne Greenberg, Copyright © 1964 by Hannah Green, Copyright renewed 1992 by Joanne Greenberg. Used by permission of Penguin Books, an imprint of Penguin Publishing Group, a division of Penguin Random House LLC. All rights reserved.)
(From I Never Promised You a Rose Garden by Joanne Greenberg published by Penguin Classics. Copyright © Joanne Greenberg 2022. Reprinted by permission of Penguin Books Limited.)

How can we conceive of the ways adverse experiences trigger psychosis? To be psychologically significant, an event must, by definition, have personal meaning for an individual, whether that meaning is good or bad. Because the unconscious phantasies laid down in childhood become the foundation of the adult’s system of meanings, inferences about significant life events are inevitably linked to, and elaborated by, unconscious phantasy. We pass a test, and the world appears to welcome us; we fail a test, and the world turns away, and so on, in the ongoing, never-ending cascade of meaning that is the human psyche. Unconscious phantasy always operates in the background of conscious experience, inevitably shaping our thoughts, feelings, and actions over the course of an ordinary day. Most such phantasy is not experienced as an ego-dystonic intrusion, but rather as a source of value, meaning, and goal-directedness.

When we have a sense that we are *overreacting* to an event, we know we are experiencing the intrusion of an unconscious phantasy that is emotionally coloring our view of the event. We know that what has just happened is somehow not as bad as we have felt it to be. With cognitive effort, and in a brief time, we can usually temper our overreaction. The unconscious phantasy that lent its overcharged emotions to the experience fades, and we go on with our day. But when a traumatic event or succession of events kindles powerful emotions that resonate with primitive unconscious phantasies that have been reinforced by early failures of development, we may become convinced that a scene expressive of our primitive unconscious phantasies is actually happening now, in reality. Any feeling person can empathize with the reactions of helplessness, rage, fear, and betrayal of someone raped by a parent or sexually abused by a priest—or both, as was true for Karon’s patient.

But in psychosis, the enormity of what has happened goes even deeper. A child who desperately wants a parent’s love and needs the parent’s protection with the intensity that only a vulnerable child can feel, who is instead used like a condom and thrown away, may not be able to bear such events within an integrated experience of self. As the trauma resonates with the deep object-related phantasies of early childhood, the mind begins to vibrate with pain, then shake, then fall apart into pieces. At the deepest levels of the psyche, where the most joyous phantasies of loving others should ordinarily reside and give hope at times of adversity, a sort of wholesale genocide of good psychological objects occurs. The psychotic person’s mind then becomes a fragmented collage of fantastic figures that tend to be all good or all bad.

This experience of one’s worst unconscious anxieties coming true is accompanied by a pervasive terror that has been termed *annihilation anxiety*. On the brink of psychosis, the self sees itself beginning to fall apart and disintegrate. As individuals descend into the “psychic equivalence” mode, where what they feel is experienced as real, primitive unconscious phantasies surface in consciousness as delusions, hallucinations, or other psychotic symptoms, whereby the psychotic person’s thoughts and feelings are expressed in a psychotic “fairy tale,” in idiosyncratic concrete metaphors, and in altered perceptions of other people and the outside world.

The author once supervised the psychotherapy of a young woman, Samantha, who routinely sat down to what on the surface was expected to be a warm family dinner where one might be expected to say such things as “How was school today?” or “Can you pass the potatoes?” knowing that within three hours her father would likely be raping her. When she was unable to bind this experience within an integrated view of herself and the world, her mind fell apart, first into two pieces in addition to the self, and later into four. She began to hear the ‘voice’ of a “good angel” on her right shoulder, and a “bad angel” on her left. The bad angel urged her to kill herself, the good angel to stay alive. As she descended further into psychosis, she came to hear three distinct “voices,” that of “Darryl,” who combined traits of her violent father and a fantasied older brother who might have defended her; “Baby,” who cried and was weak, who represented the needy child who had been betrayed; and “Tamika,” who was brash, irreverent, street-wise, so tough she could not be hurt. She described how the Tamika “voice,” like a professional wrestling match where one wrestler tags off to another, would command her to take certain actions when she felt at a loss as to what to say, feel, or do. Her psychotherapist established a trusting rapport, which over time allowed her to put Humpty Dumpty back together again. Samantha came to understand that the states of mind composed of her cast of hallucinated characters were aspects of her own feelings. Darryl’s violence and Tamika’s indifference evolved into a confident self-assertiveness that allowed Samantha to stand up for herself without need for recourse to auxiliary internal objects.

Here are examples of how different family constellations may predispose to psychosis by impinging upon the good-enough development of children, leaving them unhappy in their own way. In good-enough development, parents support independence when the growing child shows this developmental striving but allow dependence when needs for reassurance and security are paramount, as is particularly apparent in the practicing and rapprochement phases of development described by Mahler. If we keep this empathically alternating relationship between parent and child in mind as a central arc of development that leads from the nuclear family out eventually into the adult world, the disturbances that predispose to psychosis can be viewed, broadly defined, as a disturbance in the child's separation and individuation from the parent; i.e., a failure in the child's becoming an authentic, self-inhabiting person.

In the severe cases of childhood onset psychosis, florid acute psychosis in adults, and severe chronic psychosis, individuals may literally not know where other people end and they begin. More often, however, the disturbance is less dramatic but nevertheless profoundly important. Some parents need their child or imagine their child to be someone other than who the child authentically is. They impose their own mental representation of an internal object on the identity of the child. Rather than discovering who the child is as the child's identity unfolds, they decide who the child is and require the child to inhabit that identity as a condition of relating to the parent.

The father of a son might have wanted a daughter to avoid the competitive feelings he might experience with a son. Or a mother with an inadequate relationship with her husband may want an enmeshed relationship with a child who will remain dependent upon her, allaying the mother's loneliness and separation anxiety. Or a father may have wanted an athlete, but the son prefers to draw, or a mother who has lost a child in infancy may subtly communicate to another child in the family the parent's wish that it had been the other child who had died, or the next child born may be viewed as a replacement for the dead child, and so on. The possibilities for mismatches between the child's authentic identity and the imposition of a parent's need are many. These sorts of stories routinely emerge in the individual psychotherapy of psychosis.

Theodore Lidz described two basic patterns in which the psychology of one or both parents impinges upon and dominates the development of the child, patterns he termed the “skewed” and the “schismatic” family, respectively. The former serves the needs of one of the parents, most commonly the mother. The other parent and the children accept the dominance of one parent's needs, knowing that servicing them is necessary to the parent's psychological survival and therefore the survival of the family. In the latter, both parents attempt to impose their own differing and distorted view of the world on the family, and no consistent consensus is achieved. When two very different parental agendas are imposed upon the identity of the child, the child has difficulty forming a stable identity.

Arieti described a psychotic patient who grew up in a schismatic family environment. His father was a war hero killed in World War II. The father's loss so crippled the patient's mother that she was unable to invest emotionally in her son. The patient translated this seeming lack of interest into a feeling of his being a bad person. His maternal grandmother, however, encouraged him to believe he was an exceptional person who would grow up to be a hero like his father. In the face of these two contradictory imposed identities, the patient grew up with no core identity and no realistic self-esteem, a condition that left him vulnerable to psychosis.

Another patient, Jason, grew up in an atmosphere of constant bickering and fighting between his parents, who frequently urged him to take sides. The parents, who divorced when he was 16, forced him to decide which parent he wished to live with, making no effort to arrange amicable joint custody. He chose his father, who promised Jason a motorcycle when he was 18. The father had little interest in his son except to demonstrate the son's preference for him over his wife. Jason soon became disenchanted and one evening returned to his mother's house, hat in hand. At first, she barred him from her home, haranguing him about how he had betrayed her by choosing his father, saying he must live with his choice. She eventually relented and allowed him in, but she said, amidst angry declarations of what a monster the father was, that nothing would please her more than to read in the newspaper that Jason's father had been murdered. Jason got the message: The price of his return was to be his father's life. He was psychotic by daybreak. Both parents viewed Jason as belonging either to themselves or to the other parent, and if the other parent, their mental representation of Jason was infiltrated by their hatred of their ex-spouse.

Karon describes a mother–child dyad in which an insidious mismatch arose around feeding. A psychotic man who went on hunger strikes and believed he was being poisoned reported, “The Athenian girls are laughing at me. They say their breasts are poisoned.” There was, at first pass, no apparent trauma in the patient’s history. His mother, who had breast fed him for a year and weaned him gradually, asked her son, “Did I give you enough milk?” He replied cryptically, “The cow gave her calf milk and then kicked it. She shouldn’t do that. It’s something that happened hundreds of times in the history of the world.”

Karen inferred that the cow represented his mother, the calf represented the patient, and the history of the world represented his life history. Thus, the patient was complaining that every time his mother fed him, she got angry or “kicked” him. This pattern was apparent in their relationship as adults. If the mother cooked a meal and he ate it, they quarreled afterward. If she cooked and he was on a hunger strike, or if someone else cooked, there was no quarrel. When the patient was a child, his need to be fed produced anxiety in the mother, which made her angry. The same sequence occurred in their adult relationship, in which she would find pretexts for a squabble after feeding her son.

The essential aim of PDTp can be stated rather simply in relationship to the theory outlined above. The therapist tries to help patients tell their stories, to say what has happened to them in their lives and how this has made them feel, and what thoughts and feelings are currently being expressed in the psychotic symptoms. In this sense each patient’s psychotic symptoms are a Rosetta Stone by which one person’s emotional life can be translated into simple human terms.

TECHNIQUE

As noted, most psychoanalytic theorists would agree that psychotic mental states represent the intrusion of primitive psychological processes into adult waking life, and that those processes were originally most clearly explicated by Melanie Klein and her followers. Although Kleinian analysts made brilliant contributions to the psychoanalytic understanding of psychosis, their way of working using classical psychoanalytic technique has no direct application in the public mental health sector. But if one regards psychoanalysis as a body of ideas applicable in a variety of settings, not as a specific clinical technique, Kleinian ideas are a compass in psychotic chaos, invaluable tools for understanding the psychotic process and guiding care wherever it occurs with whatever technique is applied. We begin with a short review of Kleinian technical ideas, as they illuminate some general principles of therapy for psychosis, followed by the recommendations of therapists of the American Sullivanian School.

The European Kleinian School

Segal describes the treatment of a chronically psychotic man with multiple delusions and hallucinations who believed that he was a prisoner, not a mental patient, and that the world had been changed and was being destroyed. His two core delusions, that he was in prison and that the world was changing, were connected to primitive internal object-related phantasies. He complained that he was being given plenty to eat in the hospital, but not enough to do or think about, which led him to believe his head, arms, and legs were shrunken while his stomach had become huge. Instead of saying he was bored and that he felt like he was being force fed, he expressed these sentiments in a somatic delusion (a concrete metaphor). Because in the hospital he thought about little else than eating, he believed the world was obsessed with eating, to the point that people had neglected to plant new crops, which would lead to famine and the destruction of the world. After Segal consistently interpreted his primitive object-related phantasies and analyzed the psychotic transference, within a year all his delusions disappeared.

Wilfred Bion extended the work of Freud, Klein, and Segal in novel ways with his concept of *attacks on linking*. In psychosis, individuals turn away from painful realities by internally undermining (attacking) the connecting links between memories and verbal thoughts that give experience meaning. Instead of thinking a verbalizable thought, psychotically disturbed individuals project the part of their cognitive apparatus that would ordinarily maintain a conscious awareness of reality outside the self into what Bion called a “bizarre object.” The familiar computer chip in the brain, a common delusion, is a “bizarre object.” The person's cognitive capacity to observe reality has gone missing from the person's mind and has been projected into a mental representation of the chip. The chip operates within the person's body but is separate from the person's ability to think. The capacity most often projected into the chip is that of monitoring thoughts and transmitting them to an external agency such as the FBI, in which case the chip has taken over one's capacity to know one's own thoughts and evaluate them.

Herbert Rosenfeld believed in the importance of analyzing the psychotic transference. In his experience, primitive phantasies of intruding upon others or being intruded upon to the extent of annihilating the self are common in the psychotic transference. He describes a patient who feared that one day she would begin to speak in a strange voice that belonged not to her but to Rosenfeld. In keeping with Klein's emphasis on primitive body-related phantasies in which children imagine taking something out of or putting something into the mother's body or having the same done to themselves, Rosenfeld interpreted the negative psychotic transference to this woman, telling her she feared that he was going to force himself into her body and her mind and take control of her thoughts so as to cause her to lose her identity entirely.

The American Interpersonal School

Harry Stack Sullivan's ideas about the etiology and treatment of psychosis are important not only because they are persuasive but because of their simplicity. Sullivan believed that disturbances in the early mother–child relationship predispose to psychosis. According to Sullivan, all human beings have basic needs for *security* and *satisfaction*. His concept of security roughly parallels current concepts of secure attachment, positive self-esteem, and associated expectation of esteem from others. His concept of satisfaction roughly corresponds to Freud's conception of drive satisfaction, including the gratification of sexual needs. In order to achieve security and satisfaction, children must cultivate sufficient interpersonal competence to ensure that these basic needs are met in childhood, latency, adolescence, and adulthood.

According to Sullivan, when an infant's need for maternal care makes the mother anxious, this anxiety is communicated to the child, undermining the child's capacity to feel secure and to develop the interpersonal competence necessary for independent living. This failure of relatedness arises in interactions with the small group of people that compose the child's core care-giving environment. He considered the developmental milestone of having a "chum" during latency to be an important marker of mental health, as it signals the growing person's ability to establish a stable and gratifying relationship outside the nuclear family.

Sullivan died at 57 before ever writing a comprehensive manual for technique. His recommendations are scattered throughout his writings.

Much of his initial work with psychotic patients was done on a special 6-bed inpatient unit that he designed at Sheppard Enoch Pratt Hospital, where without medication, he reportedly achieved an average 85% recovery rate for acute psychosis. Consistent with his theory, he placed a significant emphasis on the quality of interpersonal contacts in the ward milieu. He took care to alert ward staff to the exquisite sensitivity of psychotic patients, encouraging them to spend considerable time fostering a relationship in which the patient could feel secure. Believing that his patients felt interpersonally incompetent in their family settings, he placed them in a humane environment in which all patients could feel like valued members of the community. He resisted objectifying diagnoses while teaching that, however strange their symptoms, people with psychosis should be considered "more simply human than otherwise." (In a similar vein, CBTp recommends the psychotherapy technique of *normalizing* to emphasize the continuity between psychosis and ordinary mental life.)

A central aim in Sullivan's approach was to reconstruct the chronology of the onset of the psychosis, so as to impress upon patients that whatever psychotic reaction had befallen them was the result of their profound disappointments in life experiences with a small number of people. Sullivan cautioned that it is extremely important to use simple, clear language because people in psychotic states have difficulty communicating in consensually familiar verbal symbols. Talking in too complex or fancy a way would be, in his words, "asinine behavior." With hebephrenic patients who had retreated into an autistic self-involvement, or paranoid patients with fixed delusions, he sought to revive a state of confusion present in an earlier phase of the psychosis. This goal is similar to the CBTp aim of encouraging a state of uncertainty in patients about their delusional beliefs. The therapist must spend sufficient time with the patient to let the treatment unfold. When the patient is not communicating clearly, the therapist should welcome the attempt to communicate but not pretend to understand what is still confusing.

Sullivan offers occasional examples of his work with patients. A psychotic man who had been severely disappointed in love said of the woman who had rejected him, she is “of the past” and “out of my mind” and someone “of no importance.” Sullivan took advantage of this moment of denial to say something along the lines of, “Nonsense! You are happy with her.” Having startled the patient into alertness, Sullivan continued, “and there is no reason on earth why the pleasure you had in her company should be thrown away just because the relationship didn’t last forever.”

If all goes well enough on the inpatient service and some degree of human contact can be re-established, allowing some degree of social recovery, the patient can be treated as an outpatient. Sullivan believed the training, experience, and personality traits of the therapist were crucial to a positive outcome. (Training and experience do matter, a fact often lost in RTCs of PDTp.)

Frieda Fromm-Reichmann worked closely with Sullivan. She distinguished *psychoanalytically oriented psychotherapy* from classical psychoanalytic technique and believed that even the most psychotic patients were capable of some degree of interpersonal relatedness. Her success as a clinician stemmed from her warmth, humility, and exquisite psychological intuition. She stressed that when the therapist inevitably disappoints the patient in some way, it will likely be experienced as a repetition of early traumas, and will provoke anger and intense hostility that can be understood and worked through. She emphasized how lonely psychotic people are, noting that in spite of their retreat from the real world, people in psychotic states nevertheless retain some dim notion of the unreality of their substitute delusional world. A longing for human contact and understanding remains, yet patients are afraid to admit this to themselves or to the therapist for fear of further disappointments.

Fromm-Reichmenn emphasized that to establish trust, the therapist must treat patients with kindness and consideration, without infantilizing them. She would explain to them that symptoms were ways of remodeling life experiences in consequence of, or in accordance with, their thwarted past or present relationships. She recommended telling hallucinating and delusional patients that the therapist does not see, hear, and believe the same things the patient does, but that the two parties can investigate together the reasons for the differences in their experiences. She wanted patients to become aware of and to grieve the losses they had sustained in early life without distorting them symbolically as psychotic symptoms. This approach is reflected in the title of Greenberg's book *I Never Promised You a Rose Garden*. Interestingly, at one point in Greenberg's psychotherapy Fromm-Reichmann used CBPp like technique to array evidence against her patient's belief she had tried to kill her sister.

Silvano Arieti won the National Book Award in Science in 1975 for his seminal work *Interpretation of Schizophrenia*. This 756-page book is arguably the most comprehensive and coherently organized text currently available on PDTp, with Karon and VandenBos's *Psychotherapy of Schizophrenia* a close second. Sullivan, Arieti, Fromm-Reichman, and Karon share much in their clinical style, an approach which might be called an *interpersonal interpretive technique*. What follows is a description of this technique following Arieti's outline of three phases of treatment, written as an amalgam of his approach and others.

Interpersonal Interpretive Technique

Arieti conceptualized PDTp as proceeding in three stages. First, the *engagement* phase, in which the therapist attempts to connect with a person who has largely withdrawn from social contact and consensual reality. Next, a phase that focuses on showing patients how they express themselves in a variety of *psychotic mechanisms*, such as hallucinations, delusions, and ideas of reference. Specific techniques are required to bring the thoughts and feelings experienced as perceptions back within the boundary of the self, where they can be experienced as mental events and examined in more traditional psychodynamic fashion. CBTp techniques that can be a stand-alone practice can be extremely useful in this second phase, about which there is more in the CBTp portion of this chapter.

Third, when patients have learned to identify the psychotic mechanisms that disguise their thoughts and feelings, they can begin to reclaim their inner life as the therapist interprets the meaning of symptoms in relationship to disowned emotions, memories, distorted thinking, and current life events. The therapist attempts to help the patient see his psychotic symptoms as meaningful reactions to problems in living, without accepting the psychotic distortions or promising the patient compensation or an ideal world. These three phases of treatment will now be described in more detail.

Engaging the Patient and Establishing a Therapeutic Alliance. Establishing a trusting emotional relationship with a patient suffering psychotically is the first and most important step in the psychotherapeutic process. The patient is caught in a devilish version of Schopenhauer's porcupine dilemma, in which too much closeness with other human beings brings the prick of the quill, and too much distance the chill of a cold that withers the spirit. Having found that attempts to relate to others have in the past occasioned intense anxiety and repeated interpersonal failures, the psychotic person withdraws from consensual reality and ordinary social contact into a solitary delusional world. Patients who can be engaged in psychotherapy must find a way to set the value of a relationship with the therapist against the unaltered continuation of their psychotic existence.

All clinicians, whether in psychiatry, internal medicine, or surgery, will see in their patients a balance of health-promoting factors and pathology. The clinician will not know all the variables involved, or their ultimate strength in the clinical equation defining outcome. The same is true in the initial engagement phase of therapy with patients suffering psychosis. The therapist cannot always know beforehand how the need for human contact will play out against the patient's aversion to the real world. Patients who lack basic trust will approach all relationships, including a relationship with the therapist, with trepidation. When a person's development has been significantly impinged upon by others, the patient may experience being influenced by the therapist in any way as an impingement.

For example, a psychotic woman revealed after months of psychotherapy that she did not think about what the therapist said during the sessions; instead, she tried to memorize the therapist's words so that she could digest them later, alone, when she felt she could exert more control and avoid the therapist's ideas taking over her mind. Patients who have had a symbiotic relationship with their primary caregiver may adopt an opposite course and rush to fuse with the therapist's identity in an attempt to reserve no mind of their own with which to feel their own pain.

The minimum conditions for psychotherapy require the patient to be able to attend sessions, whether as an inpatient or outpatient, and to curtail violent behavior directed toward themselves, the therapist and others. The therapist cannot allow socially intolerable behaviors such as stealing from the therapist, inappropriate touching, or significant disruptions of the therapist's life. A minority of patients, even when properly approached, are too damaged or too terrified to engage in treatment, at least within the parameters of the current mental health system. The therapist needs to strike a flexible balance between authoritarian controls over the patient's behavior and sitting passively by while the patient contemplates some ruinous action.

Confidentiality should be discussed early in treatment. With respect to parents, significant others, and public authorities, the therapist should explain that the treatment will remain confidential except when the therapist believes that the patient presents a clear danger to self or others, in which case the therapist may need to contact family, friends, or the authorities. If a patient cannot allow the therapist to call in help if need be, the patient may be best treated in a more structured clinical setting. The therapist can share a draft of insurance forms, psychosocial summaries, and disability applications prior to releasing confidential information.

Patients with psychosis don't necessarily expect to have rewarding conversations with mental health professionals. They may find that the delusions and hallucinations that are the most compelling concerns in their lives may be brushed aside as "imaginary" by others.

After establishing rapport with his therapist, a psychotic man acknowledged that the most important relationship in his life was his ongoing dialogue with a female "voice" he believed to be the voice of God. He said, "I never tell the doctors much at all about God. I know what they are going to think! They will just think I am crazy. They don't want to talk. They are just Haldog (*sic*) terrorists. So I take the medication and then after a while they let me go, and I stop taking the medication after I leave. I trust you. Maybe I shouldn't, but I do. There is something about your way of speaking, something about your manner, that makes me trust you."

Establishing rapport is not a mechanical matter of technique, but fundamentally *it is a matter of the heart*. The therapist's relationship with the patient will reflect what the therapist genuinely thinks and feels about the psychotic person. Just as children readily discern when meeting an adult for the first time whether it feels safe to proceed, psychotic individuals will know by the therapist's words, gestures, facial expressions, and tone of voice, whether the therapist's greeting is a well-intentioned but generic professional posture, or whether the therapist is motivated by an affectionate, abiding respect for the patient's individual personhood. Patients must know that they are truly seen by the therapist as discrete individuals. Unlike non-psychotic patients, who can explore their fantasies about who the therapist is, the psychotic patient cannot wait to know, for fear of the worst. Adrift in a sea of madness, the psychotic patient must find something real in the therapist to hold on to that is sufficiently reassuring to allow the patient to risk connecting.

Greenberg illustrates the sensitivity of psychotic patients to the therapist's true feelings in her description of the reaction of inpatients to a conscientious objector who was unhappily assigned to the hospital as a therapy aide in lieu of military service. The patients sensed that he was frightened of them and that he defensively held them in contempt. "His rigid fundamentalist beliefs made him see insanity is a just desert for its victims, as God's vengeance, or as the Devil's work, and sometimes as all three at once. As the days passed, his fear waned in time and his righteous wrath was at hand. … Against his loathing, the sick fought in their sick way. The literate rewrote the Bible or ridiculed its passages to make him horrified. Constantia made flagrant sexual advances to him. And Debra made a few pointed observations about the similarity between psychotics and religious fanatics."

The seeming failure of people around them to see the world as they do leads psychotic patients to believe that at best the therapist will not understand, and at worst the therapist is directly involved in the patient's delusions. The therapist's attempt to communicate effectively is paramount. Initially the therapist should not ask the patient any complicated questions, because a person who is desperately fighting for psychological survival will be unable to attend to the therapist's queries. A patient's wish to talk to the therapist is a good sign, no matter the subject. When the patient is unable to talk directly about the psychosis, the therapist should encourage the patient to talk about something else.

When the therapist does not understand what the patient has said, the therapist should not pretend to understand, but rather say the therapist is trying to understand but is not yet certain what the patient is trying to say. When the patient cannot or will not speak, the therapist must be able to tolerate periods of silence, but the therapist should not allow protracted silence to develop. Based on whatever the therapist knows about the patient, the therapist can keep talking, musing, meandering, speculating, allowing the patient to feel the therapist's effort to communicate. The therapist should be emotionally engaged, firm, honest, and familiar in his manner as opposed to emotionally distant.

The therapist need not withhold direct reassurance. Arieti describes his approach with a patients with post-partum psychosis. "In many cases, I have told the patient that she is in great distress because she rejects the baby, she feels guilty about it and believes that the baby is going to suffer. I have reassured the patient by telling her that she must accept the fact that she rejects the baby and that the baby will not suffer. Adequate provisions are being made for the baby. We shall eventually discover why she feels this way and her feelings will change. In the meantime, she cannot feel guilty because she can't help rejecting the baby. I have found that even very disturbed women understand these explanations, partially or totally, and are relieved by them." Karon routinely reassured terrified patients saying, "I will not let anyone kill you."

The therapist should ally with the non-psychotic part of the psychotic patient's personality that remains capable of standing outside the psychosis and reflecting on it, thin ledge along a deep chasm though this foothold may be. The therapist should not collude with the patient's delusional beliefs, nor should the therapist side-step the issue of the patient's being ill, though the matter of what to call the patient's condition must be handled with finesse. The patient's condition may have been called "schizophrenia," "mental illness," "bipolar disorder," or "a chemical imbalance." It is important to understand the labels that have been applied, the meaning of these descriptors to the patient, and what words the patient might use to describe what is going on. How the label is worded shapes patients' self-image and the way they believe the therapist will see them. Patients will sometimes talk comfortably, for example, about an "emotional breakdown" while being allergic to the word "schizophrenia."

A woman with a delusion about a man at her church was willing to consider that she might be mistaken about his intentions toward her, but she was absolutely certain that the label “schizophrenia” did not apply to her. She had looked up “schizophrenia” on the Internet and had taken away the idea that “schizophrenics imagine things that are not real.” This could not possibly apply to her, she thought, because she had no doubt that the man at church had in fact said certain things to her, as indeed he had. Although she granted that she may have misinterpreted what he said, she had certainly not *imagined* him saying what he said. She had understood the diagnostic label “schizophrenia” to apply to people who imagined things that never happened rather than people who misinterpreted things that did happen. The fact that her doctors had given her a diagnostic label that she believed clearly did not apply to her led her to conclude that they must have some malicious purpose. The therapeutic alliance foundered until this issue was identified and worked through.

Although it would not be accurate to characterize psychosis as a freely chosen, alternative way of living, people who have recovered from psychosis and written autobiographical accounts of their illness describe needing to give up on an impossibly painful real world and to surrender, subjectively voluntarily, to psychosis. Greenberg describes her escape from the real world into her delusional Yr, where in her mind she lived apart from the pain of the real world for years. Fromm-Reichmann stood patiently at the boundary between Yr and the real world, encouraging Greenberg to step across the boundary back into reality when she could, accepting Greenberg’s need to step back into Yr when she needed to.

People with psychotic illnesses may feel they have less to lose than non-psychotic individuals because they have already lost so much. Having already given up on life, such patients may feel freer to tell the truth about what they think has been going on in their core relationships, including with the therapist. When a patient with a psychotic condition comments accurately on something the therapist has said or done, the therapist must acknowledge the truth of the perception. As Arieti put it, “… on the one hand [the psychotic person] is very sensitive and is capable of seeing through a situation to perceiving the truth even more so than the normal person; on the other hand, what he does with what he sees is so distorted that it will increase, rather than decrease, his difficulties.”

If attempts to engage the patient are successful, the clinician may recognize a “breakthrough” from the profound isolation of psychosis.

Arieti quotes a patient who describes such a breakthrough moment. “There was a reading room on the ward, with a piano and magazines. I went there and I looked at the cover of a magazine. The nurse stepped in. I had never seen her before. She started a conversation with me as if I were a normal person. I told her with tremendous emotion, “You’re the first person who has broken through to me.” She was not on guard. I was not on guard. She was an ordinary girl. She made me feel communication with people was worthwhile. Before that nothing was worthwhile. People were hateful.”

Identifying Psychotic Mechanisms. Arieti distinguished between *psychotic mechanisms*, meaning the form the symptom takes, and the content and psychodynamic meaning of the psychotic symptom. For example, auditory hallucinations may be similar in form from patient to patient but different in content and personal meaning. Arieti helped patients recognize when these psychotic mechanisms were operating in their minds. CBTp-like techniques were thus essential to Arieti’s method. Through a process that CBT therapists would now call “guided discovery” of a cognitive bias, Arieti helped patients understand that they heard “voices” when they *expected* to hear them: “A patient goes home after work and expects the neighbors to talk about him. As soon as he expects to hear them, he hears them. In other words, he puts himself in what I have called *‘the listening attitude.’*” The patient comes to see that there is an internal mental contingency to what would otherwise be experienced as an external event. A similar realization happens when patients in CBTp understand that their “voices” follow stress or lack of sleep.

Even after recognizing that their expectation of abuse leads them to imagine an abusive attitude in strangers, or to trigger critical “voices,” patients may remain puzzled as to why specific charges have seemingly been brought against them. They have no conscious awareness of ever having had thoughts about themselves of the sort given voice in the attack. Klein’s ideas about primitive splitting and projection are extremely useful in understanding and working with this state of mind. In psychosis, repudiated thoughts and feelings literally go missing from the patient's mind and reappear as infiltrations of the patient's perception of reality. The patient who reports being unaware of ever having any thoughts similar to the persecutory charges is telling the truth. Arieti offers the example of a woman who heard "voices" telling her she was a prostitute, who was puzzled because she had no conscious awareness of thinking such thoughts about herself, until in PDTp she became aware of a variety of sources of poor self-esteem that had condensed in her mind into the "prostitute" charge.

Arieti approached ideas of reference and delusions with a technique similar to his approach to auditory hallucinations, by helping the patient identify what he called a "referential attitude." A man enters a public park feeling a vague, abstract sense of foreboding. Once in the park he notices children running in a direction opposite from him, and he jumps to the conclusion that he is being perceived as a child molester and the children are running from him. What Arieti called a "referential attitude," CBT therapists would call a jumping-to-conclusions bias and a self-referential bias.

Arieti taught patients to recognize the vague antecedent referential attitude that preceded and became concretized in ideas of reference. PDTp attempts to reverse this sequence, to move from the perception of concrete exterior events to the mental interior of more specific thoughts and feelings, and finally to the developmental origins of these feelings.

For example, Arieti describes a patient who believed she was getting Morse code messages she could feel when she palpated pulsations in her wrist that gave her directions about what to do and not to do to remain safe. Arieti interpreted that that she was actually giving herself advice, but she needed to experience the directives as originating in some august authority outside herself. He related this symptom to her parents never having provided the confident guidance she needed to grow into an independent person.

In some cases, the patient's delusions reflect a sensitive awareness of an actual social reality. Arieti calls this a "punctiform insight."

He gives the example of a compliant woman who was in fact being overworked and taken advantage of by her employer, who received a present of yellow roses from her boss for her birthday. She associated the color yellow with jealousy and in a turn of “paleologic,” she decided that yellow roses had been selected to show her that her office-mates knew she was jealous of the boss’s wife. Arieti interpreted that she was in a state of anxiety at work relating to her self-doubts and poor self-esteem, and that she had used a special type of thinking to express her feelings of unworthiness, in which she attributed her thoughts about herself to others. The boss did in fact appear to be hostile and demanding, and it did in fact appear she was being exploited, but her punctiform insight into this social reality had been expressed in a widening circle of paleologic ideas of reference in which the entire social surround was seen as being hostile.

Interpreting the Relationship Between Psychotic Symptoms and the Patient’s Current and the Past Life. Assuming that the therapist has been able to engage the patient and has made progress in the second phase of treatment in helping the patient to identify psychotic mechanisms and not regard their content reflexively as reality, the therapy can turn gradually toward a more traditional psychodynamic approach. Trauma and impingements upon the good-enough development of an authentic independent self have left the psychotic person with a socially incompetent personality incapable of meeting the person’s adult needs for security and satisfaction. Painful, repressed elements of the person’s mental life have so distorted the psychotic person’s mental representations of other people and reality that it has become impossible for the person to form gratifying relationships with others. Psychotherapy offers patients the opportunity to re-think these distortions so as to do better with other people and have a second chance at finding their lost authentic self.

The therapist tries to create a psychological holding environment in which, session by session, the patient’s authentic thoughts and feelings can emerge in treatment. Week after week, the patient speaks to the therapist of thoughts, feelings, and actions, and the therapist helps the patient by interpreting the unconscious meaning of the patient’s productions, including especially the patient’s psychotic symptoms. The patient may say things that are fragmented and initially incomprehensible. The therapist must gather up these mental shards over time to make coherent sense of what the patient is saying. The therapist's genuine concern and ability to bear empathic witness to the patient's emotional life provides a corrective emotional experience that offsets a childhood environment that failed to bear witness to the child's true self.

The third phase of PDTp, while similar in broad outline to work with non-psychotic patients, poses unique challenges. The non-psychotic person in psychotherapy has the challenge of trying to find links between an affectively charged experience in the present and an affectively charged experience in the past; such patients come to see that the way they are feeling in the present is an expression of a childhood state of mind. Once non-psychotic persons can see that the possibilities of the present have been constrained by the past, they can work themselves free of the past and return to the full possibilities that life offers in the present. The psychotic person's task in PDTp involves an additional intermediate step that the therapist must help the patient accomplish. In psychosis, affectively charged experiences from the past are experienced as altered perceptions of the outside world that are elaborated in delusional accounts of reality. The psychotic person in psychotherapy must first translate a current perception, or a belief regarded as a real perception of the objective world, into an affectively charged thought and/or feeling about people and the world and then link that affectively charged experience with an affectively charged experience from the past.

Consider, for example, a typical sort of session that might be occurring in any number of public clinics across the country.

THE MAN WITH PEOPLE INSIDE HIM

The author supervised a social worker who had been seeing a 22-year-old man who had become psychotic at the age of 17 for supportive sessions once a month. He had been tried on adequate doses of several neuroleptics with no seeming impact on his core delusion that people he did not know were having sex inside his body, an experience he found very distressing. The patient reported vague somatic sensations that he took as indications of the movement of other people's bodies inside his body. The therapist was encouraged to see the patient more often and to take a careful developmental history, which he did.

Lo and behold, the therapist discovered that the patient had been seduced into an ongoing sexual relationship by his older sister when he was 8 years old, a fact not known previously, which appeared nowhere in the patient's chart. In supervision, the author suggested that the therapist might consider the patient's perceptions of movements inside his body as a kind of sexual daydream in which the patient was experiencing his thoughts and feelings as perceptions. Following up on the history he had obtained, the therapist remarked to the patient in an offhand manner that most young men his age have fantasies about sex from time to time and asked if it was possible that his sense of movement inside his body, which he thought was sexual activity, might be a form of sexual daydreaming?

The patient took readily to this suggestion. As if to say "Now that you mention it," he shared with the therapist for the first time some doubts he had about his delusion. He was at a loss to understand how the people having sex entered and left his body without leaving scars. He acknowledged that when he imagined what the woman inside him might be like, he imagined her as having some of his sister's attributes. After exploring his memories, thoughts, and feelings about his relationship with his sister in four subsequent sessions, the body sensations faded, his delusion receded, and his interest turned to how he might get a job.

In psychoanalytic terms, his sister's early sexual overstimulation was an impingement upon what should have been a slower, more natural development of his sexuality. At age 8, he was not emotionally prepared to reject his sister's advances, nor was he able to absorb them emotionally. Unable to integrate these highly charged experiences within his latency age identity, he found them persisting in the form of unabsorbed introjects that he experienced as moving around inside his body. Consistent with primitive object-related phantasy that imagines the movement of internal objects in and out of the body, the patient believed that two people were moving in and out of his physical self. This is an example of the meat-and-potatoes sort of psychotherapy work that can occur in public clinics where earnest, well-trained clinicians are supervised by a clinician with more experience doing PDTp or CBTp.

The therapist will want to accumulate a chronology of significant life events and interpersonal relationships with the aim of relating this history to the patient's psychotic symptoms. It will likely not be possible to take an orderly history in a single sitting. In modern treatment-as-usual, the life histories of psychotic patients are sometimes reduced to a clinical shorthand that may be adequate for pharmacologic treatment but fall short of what is needed for PDTp. For example, one might read in a chart, "This is the fifth psychiatric hospitalization for this 26 y.o. Male with a diagnosis of chronic paranoid schizophrenia who presents with auditory hallucinations telling him to kill himself that recurred 10 days after he stopped taking his neuroleptic." This sentence contains enough information to sketch out a bare-bones pharmacologic plan: re-start his medication because it worked before and discharge him when the "voices" abate. A good clinical service will address the questions "Why five admissions?" and "Why did he stop his medication?" although important issues like these are difficult to work through in a 10-day admission, and this focus is often entirely lost in the transition to outpatient services. Rarely will the origins of the patient's self-hatred, so intense as to be expressed in a "voice" telling him to kill himself, be a central concern. Rarely will his self-hatred be the focus of a treatment plan that includes PDTp or CBTp as the treatment of choice to explore this issue.

While helping the patient translate altered perceptions and concrete metaphors into more ordinary human thoughts and feelings, the therapist must also help the patient deal with the strong affects that in the first place drove the patient's thoughts out of the patient's mind and into the patient's perceptions of the outside world. Repressed affects that commonly emerge in PDTp include intense terror, rage, shame, guilt, longing, loneliness, despair, and self-hatred. Andrew Lotterman advocates the therapist’s occasionally disclosing the therapist’s feelings to the patient as a way of modeling how a person can contain and verbally express distressing affects. For example, if the patient were to arrive quite late for a session, the therapist might say simply, “I was worried when you weren’t at our appointment on time.” The therapist’s modest, measured disclosure of the therapist’s feelings helps the patient to keep in mind the person the therapist really is.

In psychosis, affect may cease to operate as a signal-triggering defense but may course through the patient’s mind like a runaway train, becoming traumatic in its own right. The patient can be taught that no emotion, no matter how strong or seemingly out of place, is entirely irrational. If angry, it is likely the patient has felt hurt, though the hurt may not be conscious, may relate to something in the past, or be symbolic in nature. All felt emotion is real in the sense that it is really experienced, and all emotion is commensurate and in proportion to the mental events that aroused the emotion. The therapist can aim to say, “Given your experience of those events, it is not surprising you felt as angry [frightened, needed, guilty, etc.] as you did.” In keeping with the playground adage, “Sticks and stones can break my bones, but words can never hurt me,” the patient can be encouraged to understand that emotion alone does not harm.

Psychotic individuals inevitably suffer from poor self-esteem, and more often from self-hatred. People who has not felt their existence empathically recognized, who enter adolescence and young adulthood lacking the social competence to meet age expectable goals in social interaction will judge themselves to be a failure and a lost cause. Many delusions directly reveal patients’ poor self-esteem, as when they think they have a vile odor, or skin eruptions, or are harangued by “voices” telling them that they are worthless and should kill themselves to spare the world their presence.

For example, a young man who regarded it as his duty to lead his family out of poverty and set a positive example for his siblings believed he was a total failure after this younger brother was killed in a drug-related slaying. The patient came to believe that dogs in the neighborhood were looking through his clothing, and with their eyes were mocking his puny body. This man's self-hatred had gone missing from his own mind and reappeared in the mocking eyes of the dog. The therapist's devotion of considerable time and effort on the patient's behalf can go a long way to shore up self-esteem. Unfortunately, it is difficult to spend much time with patients in current care delivery models.

As the patient comes to doubt that his problems are caused by a delusional persecutor, the patient may mount an anti-parent campaign with a similar accusatory theme. Descriptions of parents that first emerge may be malignantly distorted images of the parents, infiltrated by primitive projections quite at variance with the therapist's impression of the parents upon meeting them. Because the patient has come to give some credence to what the therapist says, the therapist can point out distortions in current events related to the parents. For example, a psychotic man, who had come to understand that his perceptions of what his students said in the classroom were distorted, accused his parents of tormenting him by the off-hand mention at family dinners of the names of his prior homosexual lovers. In the course of therapy, he acknowledged that his parents could never have been aware of the existence of the lovers in question and so could not have known their names. He was able to see that events similar to the “voice”-hearing intrusions in the classroom were happening around the family dinner table.

When over time, the therapist has come to a more measured view of the patient's development, the therapist should acknowledge the very real limitations of the parents and the hurts they inflicted. In some cases, the patient may be told that despite the parent's actions being hurtful, there is no clear evidence the parents intended to hurt the patient in the way they ultimately did. In keeping with Klein's conception of the *depressive position* as including complex, integrated images of other people that tolerate ambivalence, the therapist endeavors to help patients away from the *paranoid-schizoid position* to the more mature *depressive position*, where they can see their parents as having both good and bad traits.

Patients can also be shown that their behavior may reinforce negative interactions with parents that foreclose the possibility of making whatever connection might be possible. As is true in the psychotherapy of non-psychotic individuals, patients with psychosis must learn that they will not get what they want by putting their parents on trial with the expectation they will now make good on what was missing in childhood. Patients must face the necessity of going out into the world and finding their own adult satisfactions. When Greenberg ("Deborah" in her book) complained to her therapist about how hard it was to live in the real world, Fromm-Reichmann responded, "I never promised you a rose garden," from which the book took its title.

Over the many months of her treatment, Greenberg worked to understand the meaning of her psychotic symptoms in terms of her life experience. Well into treatment, she revealed to her therapist that she believed herself to be a Japanese soldier. She knew about the "Japanese menace" in World War II and had heard the phrase on the radio, "If you aren't with us, you are against us." Socially estranged from her peer group, she heard her "voices" repeating, "You are not of them," which she took to mean that people thought of her as they thought about the hated Japanese soldiers. Instead of feeling *like* a Japanese soldier (a figurative metaphor), in her delusion, she believed she *was* a Japanese soldier (a concrete metaphor). As she imagined herself into the identity of the hated Japanese soldier, her urologic surgery at age 5 became a soldier's war wound, and the soldier's wish to escape was her wish to escape from a painfully overwhelming real world. At a point late in her treatment, when Greenberg had recovered her will to live, she made an additional discovery regarding the Japanese soldier. Greenberg writes:

"It's a description of … why … why, it's grandfather!" Deborah [Greenberg] cried, having unearthed the familiar tyrannical Latvian to whom she had given such an unrecognizable mask. It was a description of him and it fitted him better than height or weight or number of teeth. Dr. Fried [Fromm-Reichmann] then said, "Coming to see this … does it hurt so much?" Deborah replied, "A good hurt." Then Dr. Fried continued, "The symptoms and the sickness and the secrets have many reasons for being. The parts and facets sustain one another, locking in and strengthening one another. If it were not so, we could give you a nice shot of this or that drug or a quick hypnosis and say, Craziness, begone!' and it would be an easy job. But these symptoms are built of many needs and serve many purposes, and that is why getting them away makes so much suffering."
(Excerpt from I NEVER PROMISED YOU A ROSE GARDEN by Joanne Greenberg, Copyright© 1964 by Hannah Green, copyright renewed 1992 by Joanne Greenberg. Used by permission of Penguin Books, an imprint of Penguin Publishing Group, a division of Penguin Random House LLC. All rights reserved.)

(From I Never Promised You a Rose Garden by Joanne Greenberg published by Penguin Classics. Copyright © Joanne Greenberg 2022. Reprinted by permission of Penguin Books Limited.)

In addition to disclosing emotions, Lotterman advocates structured modifications of free association that he calls “naming” and “enlargement” that can help the patient get in touch with repressed emotions and put them into words, as well as role playing, in which the therapist might say “Imagine you are a poet trying to put your feelings into worlds.” For example, when a patient offers a vague description of a feeling state such as “I felt bad,” the therapist might encourage the patient to linger over the statement to name more specifically the emotions involved. The therapist may note a particular phrase that appears to have a significant emotional valence, as when a patient said her feelings had been “interred. “When the therapist drew her attention to this word, she at first dismissed it as just an expression, but when encouraged to stay on point and enlarge her associations to the word, she began to talk about how cut off from people she felt and how hopeless she felt about anyone ever discovering her; that is, about her ever being disinterred.

At the beginning of treatment, the relationship with the therapist as a supportive real presence is paramount. As awareness of psychotic mechanisms increases and examination of the patient’s family relationships proceeds, treatment may pay more attention to transference distortions. For example, a patient who believed she had an offensive odor recalled that the therapist had opened the window to his office when she arrived—to air the stink out of the room, she assumed. She believed that the therapist and most strangers on the street thought she had a bad smell. The therapist was able to point out that he had actually *closed* the window to shut out noise from the air well, but she had distorted her memory of his actions in keeping with her negative expectations. The patient was able to work through her conviction that the therapist thought she had a bad smell (but withheld telling her because he didn’t want to hurt her feelings).

At the beginning of treatment, the therapist strives to be a trusted supportive parental figure who offers a sincere commitment to provide whatever support is needed for the patient to survive. Later in treatment, the therapist may move toward a more demanding stance that conveys the expectation the patient will be able to live in the real world. As is true in the back-and-forth movement between dependence and independence in childhood, the patient may resent and resist such encouragement. The therapist must go slowly, showing the patient that the expectation of independence is an expression of confidence.

Aside from the fear of psychosis that may inhibit a therapist's close work with psychotic people, and aside from feelings in the countertransference that can provide clues to the patient's pre-verbal states, several other forms of countertransference reactions may arise during the course of treatment. Therapists may feel so deeply gratified by the intimately meaningful reward of the work that they do not recognize that a patient has made significant progress toward termination. Or they may become so habituated to a patient's peculiarities that they fail to recognize the pervasiveness of a patient's psychosis. Or they may experience a diminished interest in a patient as the person becomes less psychotic. There is little idle chit-chat in work with people who suffer from psychosis. Therapists may find the courage and raw vitality in psychotic patients an appealing environment in which to work, especially when compared with the deeply defended "normalcy" one often encounters working with personality disorders.

As in any psychotherapy, the patient changes not only from insight but from internalizing aspects of the therapist. The patient may identify with the therapist's accepting attitude such that the patient's excessively harsh superego is softened. The patient may take the relationship with the therapist as a model of how a human relationship might be. When therapy succeeds in reducing a patient's distortions of reality, expressed in psychotic symptoms, and when the relationship with the therapist provides a sufficient basis for self-esteem, the patient can get on with the sometimes difficult business of adult living. After all, the therapist never promised the patient a rose garden—just a better chance at living.

Cognitive-Behavioral Therapy for Psychosis (CBTp)

Cognitive-behavior therapy, as first described by Aaron Beck in 1979, was not applied to the treatment of psychosis in earnest until the early 1990s. The first comprehensive textbook of CBTp appeared in 1996, and many have appeared since. Interventions shifted from the primarily behavioral focus in token economies to cognition in psychosis, and a vibrant several decades of clinical research and refinement of technique ensued and continues. If behavioral therapy evolved into cognitive-behavioral therapy, and, as it appears to headed, CBTp evolves into “affective cognitive-behavioral theory,” CBTp and PDTp will lie in closer proximity.

A basic tenet of cognitive theory is that a person’s beliefs about events mediate between those events and the emotional and behavioral consequences that follow them. This paradigm is summarized in the A-B-C sequence familiar to all cognitive therapists, in which “A” is an activating event, “B” is a belief about the meaning of that event to the individual, and “C” is the emotional and behavioral consequence of that belief. Assuming that emotional distress and maladaptive behavior are a consequence of delusional beliefs, CBTp researchers set out to determine how delusional beliefs and beliefs about “voices” arise, hoping to find ways to help patients change these beliefs and reduce suffering. The subsequently evolved “CBT triangle,” which depicts reciprocal links between cognition, affect, and behavior reinstates the importance of emotion in symptom formation in CBT theory, a view that has always been central in PDTp.

Their research revealed that a number of cognitive biases found in ordinary people were exaggerated in people with psychosis, including attributional bias, self-referential and attentional bias, a tendency to jump to conclusions, dichotomous thinking, belief inflexibility, and theory of mind deficits. A cognitive model of psychosis emerged, in which prior expectations were seen to shape the cognitive processing of an external life event or an internal event like hearing a “voice.” Informational events that have the qualities of being expected, external, clear, common, neutral in affective significance, voluntary, public, and irrelevant to the person, may be ignored. But when, as is often the case with psychotic experiences, an event is unexpected, internal, ambiguous, unusual, affectively loaded, involuntary, private, and personally relevant, the event proceeds to further cognitive processing according to the individual’s information-processing style under conditions of high emotional arousal. Delusional people tend to have a cognitive style that focuses on current stimuli as opposed to past regularities of experience, and they tend to make rapid, overconfident judgments that are lacking in nuanced cognitive complexity.

Descriptions of the cognitive biases operating in psychosis allowed clinicians to tailor CBTp interventions to research findings. Through a process of *guided discovery*, patients can be shown that their belief is not the only possible explanation for events. Patients were taught to identify the cognitive biases operating in their own thoughts and to question the conclusions that followed from these biases. Many of the cognitive biases revealed by empirical research are known in the psychoanalytic clinical literature by different names. For example, Klein's idea of a primal split, in which the self is identified with good elements of the psyche and bad elements are projected into a persecutory object, is a description of what CBTp therapists would call an "attributional bias." What Freud described in 1926 in *Inhibitions, Symptoms, and Anxiety* as a constriction of the ego in the service of defense, CBTp would call a "safety behavior." There are many such crosswalks between psychoanalytic and cognitive theory. It may be reassuring to experienced clinicians to find that the psychoanalytic and cognitive-behavioral traditions have found their way to many of the same ideas and that CBTp has demonstrated the value of those ideas in empirical research.

Clinicians and researchers continue to develop new modes of treatment. In a so-called "third wave" of clinical interventions, new types of therapy emerged after CBTp gained acceptance. Over time, early attempts to mitigate suffering by changing beliefs has made room for more meta-cognitive approaches; that is, helping patients change the way they think about their own thinking. These innovative techniques include, among others, mindfulness-based approaches that draw upon the wisdom of Eastern traditions to foster a more accepting attitude toward suffering, with an emphasis on kindness and compassion. In contrast to traditional CBTp, which might encourage patients who are hearing distressing "voices" to turn their attention away from the "voice" and to change their beliefs about the "voice," a mindfulness approach would encourage them to adopt a more tolerant, more meditative, less emotionally reactive response to the presence of the "voice."

Acceptance and commitment therapy (described elsewhere in this textbook) does not have the primary aim of changing beliefs, but rather hopes to encourage patients to identify and keep clearly in mind their personal values, and to act on these values rather than acting on the prompt of a psychotic symptom. Compassionate mind training aims to help patients increase self-esteem by resisting self-critical elements in their personality more effectively. For example, the "two chair technique" allows the patient to imagine sitting in one chair while conducting a dialogue with an inner "bully." This technique echoes the voice dialogue approach developed in the late 1980s, and Gestalt approaches promulgated earlier, and is in keeping with a fundamental goal of most PDTp: diminishing the impact of an overly punitive superego.

CBTp offers a variety of effective interventions that can be used in their own right that can also be integrated with PDTp to help patients recognize the operation of distorting psychotic mechanisms. This sets the stage for more effective PDTp work if the patient needs wants and can use this treatment approach. Like psychodynamic approaches, CBTp therapies are too richly diverse to summarize. There follows a description of how a more traditional CBTp might be conducted in the following sequence: (1) engaging the patient; (2) eliciting the patient's story; (3) introducing CBTp with an emphasis on coping mechanisms; (4) reviewing the three treatment models underlying CBTp; (5) applying these models to the patient's experience; (6) challenging beliefs and re-formulating the patient's personal life narrative; (7) maintaining contact and employing "booster sessions" as needed.

Engaging the Patient in CBTp. The therapist should pay attention to how the patient came to be referred. Was it the patient's idea or someone else's? What is the patient's understanding of the purpose of the meeting? Does the patient have any specific anxieties about the meeting? Patients may wonder why the therapist is offering to see them more frequently than they expected. One patient joked, "Do you have a new supervisor?" The therapist might say, "If it is okay with you, we can meet a few times and talk about what has been happening in your life and see if we can figure out a way for things to be less distressing." Does the patient have an immediate, overriding concern that must be addressed before the patient can focus on the therapy? A patient who is worried about being evicted from home or who doesn't have enough to eat will have difficulty paying attention until these needs are addressed.

The therapist will want to know about prior interactions with the mental health system, particularly those that have been notably problematic for the patient. Allowing the patient a virtual "panic button" that the patient can activate if the patient gets overly anxious in the session may be helpful. The therapist can ask occasionally, "How are we doing so far?" The therapist must be sensitive to the fact that the patients may believe they are putting their life at risk by even talking with a therapist. If the patient is reluctant to talk about what the therapist sees as psychotic symptoms, the therapist can find something else the patient is willing to discuss. In summary, as was emphasized by the American Sullivanians, contemporary CBTp therapists establish a relaxed conversational manner that levels the power relationship between patient and therapist.

Elicit the Patient's Story. It is vital to ask what the patient thinks is happening and why. What evidence does the patient have for a problematic idea? The patient will often point either to one seminal event that fostered the delusional belief, or to a series of events given meaning by ideas of reference, in which case the patient may have pondered, "If this happened once or twice, it might be a coincidence, but it has happened over and over again, so there must be something going on." It is extremely important to elicit any aspects of the patient's experience about which the patient feels confused, as these doubts can often serve as a useful base to explore alternative explanations of the patient's experience.

For example, a man who had no doubt that he was an angel reported he had had an experience when he was being taken to the hospital that made no sense to him. The ambulance had passed children playing in school yard. He saw a good aura around the heads of some of the children, and a bad aura around others, but he knew there must have been something wrong with his perception because he believed all children were innocent and good. To encourage a therapeutic alliance based on evaluating data, the therapist might ask such a patient to rate his degree of preoccupation and conviction regarding a delusion on a 1% to 100% scale. If the patient reports being 99% certain of the delusion, the therapist might quip, "So we can work with the 1%."

Because the wish to reduce distress drives treatment, the therapist must determine what the patient finds most distressing about the current situation. While a diagnostically oriented clinician might aim at eliminating symptoms like "voices," the patient may be initially unconcerned about the "voices" and yet quite distressed that the patient's family regards them as imaginary. The therapist will want to establish rapport by empathizing with the patient's beliefs and feelings from the patient's point of view. The therapist might say, "Having heard what you went through, I can see why you came to the conclusions you did. When you first noticed strangers in your neighborhood looking at you, you concluded that you were under surveillance. I can see how that made sense to you at the time." The therapist may go on to ask what other people think about the patient's belief. What does the patient's family say? Does anyone share the patient's point of view?

The therapist can record the patient's life history on a *timeline* sketched out along a horizontal line on paper and add to the timeline as history accumulates. Putting the patient's history down on paper will frame the therapeutic alliance around a shared data set. In many cases, the timeline will be the first occasion the patient has ever constructed any organized, linear sequence of personal history. Life events that psychotic individuals have trouble connecting in their minds can sometimes be connected on paper by noting their physical proximity on the page. For example, the therapist might say, "We marked your parent's getting divorced right here when you were away at school, and I see the 'voices' started up a short time after that."

Temporal sequences, and therefore causal links, can often be better established on paper than in the flow of conversation. The use of paper also gets the patient used to working with diaries and documents in homework assignments. At a later time in treatment, it will be extremely useful for the therapist to draw two columns representing alternative beliefs so that evidence for and against each belief can be arrayed in each column. It is important to inquire about a history of physical or sexual abuse or bullying, but it is wise to establish some rapport with the patient before venturing into this territory. Even when the patient has been abused, the patient may not tell the therapist early in treatment. Substance abuse is usually more easily acknowledged.

If the patient hears "voices," the therapist should inquire about the identity, purpose, and power relationship to the patient, the consequences of non-compliance with the "voices," and whether the patient regards them as omnipotent or omniscient. As noted earlier, the therapist should inquire about what diagnosis the patient has been given, by whom, and what it means to the patient. The therapist will want to know what the patient's daily life is like and what relationships are important to the patient. The therapist should consult the chart and collaterals for additional information and, of course, screen for danger to self or others.

Introduce the Principles Underlying CBTp With an Initial Emphasis on Coping Mechanisms. After the therapist has constructed a comprehensive timeline and identified the primary sources of the patient's distress, goals for treatment with the patient can be addressed. The fundamental aim is to reduce distress rather than necessarily eliminate symptoms. Goals might include an increased understanding of what triggers psychotic symptoms, improved coping with residual symptoms, reduction in stress related to delusions and hallucinations, and maintaining gains and preventing relapse. The therapist should inquire as to what coping mechanisms the patient currently uses.

If a patient has abandoned hope of coping, this may be a grave sign. The therapist may ask patients if they would be interested in learning how other patients cope with distressing experiences. Most will say yes. This conveys the implicit message that what is happening to the patient has happened to other people: the patient is not alone. Most coping techniques involve the re-direction of attention away from distressing symptoms or toward more rewarding interests. Patients may distract themselves by listening to music, or practice focusing away from symptoms and on to pleasant imagery. The therapist can go through a list of possible coping techniques with the patient and choose several to try as a homework assignment. Coping cards that the patient can carry are useful. By highlighting the patient's coping capacity, the therapist underscores that symptoms are not entirely outside the person's control.

The therapist can explain the three basic models underlying the CBTp approach: (1) the stress/vulnerability model that emphasizes the importance of stressful life events in the genesis of distressing experiences; (2) the continuum model that finds analogies to psychotic symptoms in ordinary mental life, and 3) the A-B-C model noted earlier. The therapist should be fluent in each of these models and be able to explain them clearly to the patient. The stress/vulnerability model is a good place to start. The therapist should avoid appearing to be an expert lecturing the disempowered student. One might say, "Everyone experiences stress at some point in life. Stress can result in worries and a lack of sleep. I have certainly felt the effects of stress myself."

Discuss the Normalizing and A-B-C Models in More Detail. *Normalizing* psychotic symptoms equalizes the power relationship between the patient and therapist, which encourages a good-faith exploration of the patient’s symptoms without the serious risk that the patient will feel diagnostically stigmatized. The patient will hear the therapist saying, “We all have a bit of what you have, though you have more than most.” Normalizing encourages patients to place their symptoms in a psychological realm rather than regard them as problematic realities. The therapist should be prepared with a variety of personal examples of psychosis-related mental events that can be shared. One might normalize hearing “voices” by describing hallucinations of everyday life, such as hearing one’s name on the street when no one is speaking, or by describing how “voice”-hearing is a variant of the self-talk and internal dialogues everyone engages in from time to time, as when we say to ourselves sub-vocally, “You idiot!”, or when, on a diet, we say to ourselves “Who are you kidding! You don’t need dessert!”

Although affect and cognition interact in a dialectic manner, it is useful to teach patients the A-B-C model that emphasizes beliefs. This can be taught first in abstract outline, then from the therapist’s personal examples, then from mundane examples from the patient’s experience, and eventually in relationship to the patient’s core delusional beliefs. The A-B-C model can be used to analyze a relatively neutral event in the patient’s life. This gives the patient and therapist practice with the concept. The patient is invited to examine a distressing “A,” ideally an event about which the patient already expresses some doubt.

As the patient and therapist begin to use the A-B-C model to examine the patient’s core psychotic experiences, the therapist can return to normalizing to emphasize the range of beliefs and perceptions that occur in ordinary mental life. The author sometimes tells patients the story of how in dim light, he ‘hallucinated’ the image of a black dog curled up on a bed that turned out to be a black sweatshirt thrown on the bed in such a way that the arms of the sweatshirt looked like the folded legs of a curled up dog. This example illustrates how we sometimes “see” what we expect to see.

Apply These Models to the Patient’s Experience and Begin Investigating Core Psychotic Beliefs. The therapist can with advantage adopt a “Columbo-style,” off-hand way of talking. “Columbo” was a detective on television, played by Peter Falk, who spoke in a casual manner that belied what he knew. The therapist fosters a therapeutic alliance not based on the therapist’s claim to know what is in the patient’s mind, but on the joint investigation of data consciously available to both patient and therapist. The treatment team consists of Watson and Crick sitting side by side, rather than Freud sitting behind Dora (one of his first patients).

The A-B-C model can be used to examine core psychotic experiences. The therapist can put two columns down on a piece of paper to examine alternative theories. First, the therapist and patient clarify the “A”: What precisely did the patient experience? What precisely did the patient hear? The patient may have actually experienced only a muffled sound but later concluded it was a threat. An “A” may be a mental event, like a “voice,” or an external event, like the glance of a stranger. Patients often merge the “A” and the “B” such that the event and the patient’s belief about an event are simultaneous equated experiences. CBTp aims to separate the “A” from the “B” so that the patient’s beliefs can be examined. The therapist might say, “If your explanation for ‘A’ is correct, your reaction is understandable. If there is any other way of understanding what is happening to you that would be less distressing, it would be important for us to discover it. Let’s look at the evidence and see what we can figure out.”

A rich toolbox of CBTp techniques has been developed to examine beliefs; for example, the practice of rating the patient’s degree of conviction on a scale from 1% to 100%. If patients can be made aware of their convictions changing, the therapist can point out that while facts don’t change, beliefs do, thereby categorizing the delusion as a belief rather than a fact. Other techniques include rating the likelihood of a belief, rating the value of the evidence, agreeing to disagree about final conclusions pending further assessment, peripheral questioning, informational handouts that increase real-world knowledge, googling topics during the session, inference chaining (“If this turned out to be true, what would it mean to you? Which explanation would you prefer to be true?”), homework assignments and diaries, reality-testing experiments, and self-esteem-building exercises.

Challenge Delusional Beliefs and Re-Formulate the Patient’s Personal Life Narrative. On the basis of their joint investigation, the therapist can offer a tentative formulation based on the stress/vulnerability, cognitive, or biopsychosocial model that leads to a more adaptive alternative belief. A formulation using the stress/vulnerability model might be: “Your mother’s death was very stressful for you. You weren’t sleeping well and were using drugs to ease the pain. In that stressed-out state, you had some disturbing experiences, like thinking your neighbor was trying to poison you. We can now see that those beliefs were the result of the upset state of mind you were in at the time.” Or, a formulation based on identifying cognitive biases: “You have a cognitive bias that leads you sometimes to jump to the conclusion that people know you and want to hurt you. You tend to personalize events, believing people are referring to you when they are not.” Or, a fuller biopsychosocial formulation: “You felt that if you had set a better example for your brother, he might not have died so young. You felt guilty, and when you couldn’t get your life together after the Navy, you felt ashamed. You criticized yourself. Over the years, those feelings of guilt and shame built up, and you came to feel that shame everywhere you went, including in the eyes of the dogs in your neighborhood.” The therapist should determine whether the patient engages in safety behaviors that constrict life experience, precluding positive experiences that would contradict delusional beliefs.

As described in the earlier part of the section on technique, the CBTp work outlined above can lead into a final phase of therapy that seeks to find psychological explanations for the formation of the patient’s psychotic symptoms. CBTp is a preferred method for showing the patient what delusional beliefs are *not* the case. PDTp is well suited for showing the patient what *is* the case; that is, showing how the patient came to form delusional beliefs and what emotional truths the psychotic experience expresses.

Table 32.6–3.

Resistance to the Use of Logic in Cognitive-Behavioral Therapy for Psychosis

Logic evasion	Evading the implications of logic by claiming to be an exception to whom logic does not apply
Logic monopolizing	Claiming to be a master logician who offers definitive logical proof of a delusion not open to any alternative line of logical reasoning
Logic blinding	Acknowledging a logical conclusion while remaining blind to the emotional implications of a logical inference
External expectancy	Relying on external agents and contingencies as instruments of change rather than one's own hard work in therapy
Logic partitioning	Conceding that logic suggests that some beliefs are likely untrue without changing core delusions or recognizing the psychological processes that continually generate new delusional beliefs
Equivocation	Offering vague, nonspecific noncommittal responses that provide little traction for logical thinking
Reactive reassertion	Expressing a reactive, strident reassertion of a delusion when the patient senses a conflict between the delusion and logic
Feeling-percept fusion	Ignoring the considerations of logic because of a strong "feeling" that the delusion is true, despite logical evidence to the contrary
Mind-guarding	Resisting alterations of beliefs because the change threatens to erode a vital psychological structure upon which the patient's identity, vitality, self-esteem, or emotional regulation depends
Peripheral preoccupation	Preoccupation with an underlying emotional

	concern that is not embedded in a psychotic symptom per se that has not been recognized as a central focus of therapy

CBTp and PDTp both access logic, but compared with PDTp, CBTp places more emphasis on the use of logic to examine evidence to consider alternative beliefs that are more adaptive than delusions. However, even in a well-conducted CBTp treatment, patients may show a variety of resistances to the use of logic which may hinder a therapeutic approach that relies on logical cognition (see Table 32.6–3) as opposed to the interpretation of unconscious meaning (PDTp). CBTp therapists may reduce the impact of resistances to logic by identifying them and addressing them in the treatment.

Maintain Contact and Employ "Booster Sessions" as Needed. Therapists should develop with each patient a description of possible stressors and early signs of relapse to look out for. They need to anticipate what the patient should do if these signs appear. Working collaboratively, they should encourage the patient to practice the ability to identify signs of relapse and to implement the recommended course of action. Continuing to see the patient on a regular basis is critical for supporting positive changes.

ETHICAL ISSUES

The ethical constraints, such as maintaining professional boundaries that govern the practice of psychotherapy with non-psychotic patients, are well known and apply equally to the treatment of psychotic patients and so will not be addressed here. Because most individuals with chronic psychosis are treated in the public psychiatry sector or in hospitals or clinics funded by third-party government insurance, their well-being is directly affected by how resources are allocated and what kinds of care are reimbursed, a circumstance that poses ethical concerns.

In the study of ethics, there is a time-honored tradition of using a hypothetical situation to think through an issue. Consider the following hypothetical: Assume that chronic psychosis is a complex biopsychosocial disorder that has no single cause. Assume that no mental health system has developed a way of treating psychosis that avoids the prescription of neuroleptic medications in all patient cases, despite evidence that these drugs may cause structural damage to the brain. Also assume that individual psychotherapy is an effective treatment for psychosis that in some cases may resolve psychotic symptoms unresponsive to medication, improve quality of life, and reduce or even on occasion eliminate the need for medication. If we make these assumptions in our thought experiment, and assume that physicians are ethically bound to provide patients access to quality care, what ethical implications follow from the above assumptions?

The author suggests that psychiatry has an ethical obligation to consider the following:

1. At present, there are woefully insufficient numbers of therapists trained in psychotherapy for psychosis to meet patient needs, and there are not enough supervisors who can train additional therapists. Federal and state agencies should fund pilot training programs with a research component to develop supervisors and therapists with the needed skills. Psychiatric residency programs should find ways to teach psychotherapy for psychosis.
2. In understaffed public clinics, where too many patients are assigned to too few clinicians, administrators should pilot psychotherapy initiatives coupled with research that allow clinicians to see at least a few patients frequently enough to provide PDTp and/or CBTp.
3. To foster a therapeutic alliance that creates a bridge between inpatient acute care and outpatient PDTp and CBTp, suitable patients nearing discharge on inpatient units should begin working with an outpatient therapist while the patient is still in the hospital.
4. Granting agencies should encourage research that explores whether PDTp and CBTp can help reduce the total neuroleptic exposure patients receive.

Despite the enormous amount of information accumulated about the functioning of the brain, most psychiatrists regard neuroscience as still in its infancy. Psychiatry should also regard the psychotherapy of psychosis as still in its infancy, but as a childhood with much promise. Improvements in outcome seen in the last two decades with CBTp suggest there is much progress to be made in the psychotherapy of psychosis when a critical mass of researchers and clinicians put their minds to the challenge.

Since Bleuler's time, there has always been a small chorus of voices in the clinical community who were convinced from personal experience that psychotherapy can be an effective mode of treatment for psychosis. If it is true (as personal experience and research evidence suggest) that psychotherapy can be an effective element of treatment, psychiatry has an ethical obligation to understand why patients with psychotic disorders have little or no access to psychotherapy and to devise the psychiatric educational, political, economic, and clinical vehicles that would allow psychotherapy to be delivered. As Karon put it a decade ago, "The real tragedy of schizophrenia is not the severity of the symptoms and the suffering that results for patients and for their families, but that we know psychotherapies that work and we are not using them."

RESEARCH AND EVALUATION

There is solid evidence for the efficacy of CBTp, supported by multiple randomized-controlled trials (RCTs) and meta-analyses. An early meta-analysis of 34 CBTp clinical trials found an effect size of .4, in the range of effect sizes for some neuroleptics (typically in the .4 to .5 range, with the exception of clozapine as high as .8). A more recent meta-analysis of CBTp RCTs showed a somewhat smaller average pooled effect size of .33. CBTp is included in the British National Institute for Health and Care Excellence (NICE) Guidelines. A recent review of six different psychological interventions for psychosis found that CBTp was significantly more efficacious than other interventions in reducing positive symptoms, while social skills training showed an advantage in improving negative symptoms. Thus it is clear that CBTp works. There are no head-to-head RTCs comparing psychopharmacology alone against CBTp alone.

The many published case histories, autobiographical accounts, and the testimony of many therapists and patients bear witness to the efficacy of PDTp, and the positive experiences of the Soteria Project and the Open Dialogues program lend support to treating psychosis without neuroleptics, but if one considers RCTs to be the gold standard of evidence, the RCT base for PDTp is small, methodologically flawed, and outdated. The Michigan State Psychotherapy Project, in which hospitalized patients with psychosis were randomly assigned to either individual psychotherapy without medication, therapy with medication, or medication without therapy, reported that psychotherapy produced significantly greater improvement than medication alone and was particularly effective in changing thought disorder. The investigators also concluded that more experienced therapists produced more improvement. A more recent study compared expressive insight-oriented therapy (EIO) with reality adaptive supportive therapy (RAS) and found that both were effective. RAS had a clearly greater impact on recidivism and social functioning, while EIO had a modestly more positive effect on ego functioning and cognition. The conclusion of a 2001 Cochran Database Review of randomized individual PDTp and psychoanalysis for schizophrenia, however, did not support the efficacy of those approaches. The review concludes by saying there is an urgent need for research in this area. It is apparent that while RCT support for CBTp is strong, research support for PDTp is not.

In addition to PDTp and CBTp, additional modes of psychotherapy for psychosis with demonstrable efficacy have emerged, including *mentalization* therapy (MBTp) and *metacognitive* therapy (MERIT). In MBTp the therapist aims to help the patient toward a more accurate understanding of the thoughts, feelings, and wishes that transpire in the patient’s mind and the minds of other people in the context of social relations. The MBTp therapist provides psychoeducation about mentalization and fosters an interpersonal attachment that builds a positive therapeutic alliance that minimizes the distortions of affect arousal. MERIT encourages the patient’s capacity to think about their thinking (metacognition) to strengthen the coherence of a personal narrative constructed from episodic memories that strengthens the patient’s sense of personal agency and self-directed recovery. In MERIT sessions, the therapist is mindful of eight elements, including (1) encouraging personal agency; (2) focus on episodic memory (often of recent events); (3) identification of a plausible psychological problem; (4) setting reasonable goals; (5) reflection on the therapeutic relationship; (6) assessment of progress in therapy; (7) interactions matched to the individual’s capacities for metacognition; and, (8) developing strategies to master psychological problems. *Avatar* therapy is another promising new psychotherapy for auditory hallucinations in which the patient creates an avatar in a computer program that represents the entity the patient experiences as persecuting them. The therapist helps the patient to engage in a dialogue with the avatar where the patient’s control of the avatar fosters a change in the patient’s relationship to the avatar from abusive to more friendly and supportive, a shift that is mirrored in the patient’s voice-hearing experience.

FUTURE DIRECTIONS

CBTp is structured and brief compared with PDTp, which is less structured and more open-ended. This makes PDTp more difficult to study than CBTp. Developing new ideas and conducting research requires a critical mass of clinicians trained in the techniques to be studied. At present, there exists more than a critical mass of CBT-trained clinicians working and doing research in many centers. The same cannot be said for PDTp, where more teaching and practice opportunities need to be developed.

At this point, what do we know about the comparative efficacy of psychopharmacology and psychotherapy? We know that patients treated in a care system with a predominantly pharmacologic paradigm have a life expectancy of 15 to 20 years less than the general population. We know that all neuroleptics have potentially significant side effects and that psychotherapy for psychosis has never been proven to have any. We know that roughly a quarter of patients initially treated with medication after years of pharmacologic treatment can do well off medication, but we do not know how to identify these people or how to wean them off medication even if we could identify them.

We don’t know how to teach psychotherapy and integrate it into current care systems. We don’t know whether psychotherapy combined with medication can reduce dosage creep or reduce total neuroleptic exposure for patients on maintenance medication. We don’t know if the current psychotherapeutic techniques are the best that can ever be devised. We don’t know if psychotherapy is more effective for some symptoms and if medication is more effective for others. Many important questions remain to be answered, but until there is a more vibrant clinical community doing psychotherapy with psychotic patients that can collaborate doing research with psychopharmacologists, these questions will likely go unanswered.

CASE STUDY

The author will now describe the psychotherapy of a psychotic woman who, prior to her beginning an integrated CBTp/PDPp treatment, had been an inpatient in a state hospital for 15 years after she stabbed both her parents to death while in an acute psychotic state. At the time she began psychotherapy, she had been taking 600 mg of clozapine for the prior 8 years, with no change in her core delusions and, accordingly, no hope of ever leaving the hospital. After a successful course of psychotherapy, she was discharged to a community residence.

THE WOMAN WHO MURDERED HER PARENTS

Loretta was an only child who returned home after college and worked a number of low-demand jobs. She joined an Evangelical church and became preoccupied with church activities. Soon afterward, she suffered her first psychotic episode and psychiatric admission. After discharge, she was followed in a community clinic on psychotropic medication. Five years later, she again became preoccupied with the church. One evening after an altercation with her father, she murdered her parents. She recalled that prior to the homicides, “one of the Evangelicals had looked at me with a piercing stare” that she believed had placed her in a hypnotic trance that led her to murder her parents. Staff believed that she might seek revenge if she were discharged. In 15 years in the hospital, she had never been involved in a violent incident, though she feared assault by other patients and believed that nursing staff were intentionally provoking her to violence.

The therapist followed a CBTp approach for the first twenty sessions. At one point during this period, the therapist focused in detail on the subjective events that had occurred during the psychotic episode when Loretta killed her parents, framing these as “A”s to be investigated in A-B-C format. Loretta acknowledged that people might stare for a variety of reasons. The therapist drew two columns of the paper, one for the “Evangelical” explanation and one for the “mental health” explanation. He then asked which Loretta hoped was true. He had been imagining that the patient might welcome the mental illness explanation, but much to his surprise, she said the opposite. “If the Evangelicals hypnotized me, I was just stupid. But if I killed my parents because of mental illness, I am a piece of garbage, and I should rot in hell!”

Moving into the PDTp phase of treatment with more focus on life history and affect, the therapist conceived a role-playing exercise to explore the patient’s dissociated guilt. The therapist asked if she were familiar with the case of the woman who claimed to be David Letterman’s wife, who kept breaking into his house, and she said she was. The therapist then pretended to be an attorney who presented three different scenarios to Judge Loretta, asking whether she would recommend treatment or punishment. In the first scenario, the woman resists the police, but no one is hurt. She recommended treatment. In the second scenario, a policeman receives a minor physical injury; she again recommended treatment. In the third scenario, the woman strikes a policeman in the head with a lamp and kills him. Judge Loretta paused. She said she needed more information. Was the attack premeditated? If there was premeditation, Judge Loretta would recommend punishment. If not, she would rule for treatment. The therapist then asked Loretta if she had premeditated the murder of her parents. As if taken aback, she replied “No” with confidence and calm. In 15 years of inpatient treatment-as-usual, no one had ever given her an opportunity to talk about her terrible guilt over murdering her parents.

The next week as the therapist turned the key and stepped through the door into Loretta’s ward, she rushed to greet him, a broad smile on her face, a brand new shawl around her shoulders. She announced she had recovered the ability to pray, and that she felt once again able to communicate with God! The therapist was taken aback. He feared that the previous session had undone her, and that she was now hallucinating. But this was not the case. She explained that after she murdered her parents, she had a feeling that God had turned his back and would no longer listen to her prayers. In the previous week, she had not heard the “voice” of God, but rather she had recovered a feeling that God was once again listening to her! The therapist’s eyes welled with tears. Her psychotherapy had begun to mitigate the agonizing guilt that she had carried for 15 years in the Evangelical delusion. Her obsession with the Evangelicals faded, staff grew increasingly confident in her recovery, and she was discharged to a community residence within a year. The therapist met Loretta at Dunkin’Donuts after her discharge. She was doing well. Both ordered coffee and a jelly-filled donut, and both agreed the donuts were delicious.

Loretta’s story and the stories of many patients like her who have benefited from PDTp and CBPp constitutes evidence, even if those stories have not been studied via RCTs, of the importance and efficacy of psychotherapy in the treatment of individuals with psychotic suffering.

FURTHER READINGS

Andreasen NC, Liu D, Ziebell S, Vora A, Ho BC. Relapse duration, treatment intensity, and brain tissue loss in schizophrenia: a prospective longitudinal MRI study. *Am J Psychiatry*. 2013;170(6):609–615.

*Arieti S. *Interpretation of Schizophrenia*. 2nd ed. Basic Books, Inc; 1974.

*Beck A T, Rector N, Stolar N, Grant PM. *Schizophrenia: Cognitive Theory, Research and Therapy*. Guilford Press; 2009.

Corstens D, Longden E. The origins of voices: links between life history and voice hearing in a survey of 100 cases. *Psychosis*. 2013;5(3):270–285.

Federn P. *Ego Psychology and the Psychoses*. Weiss E, editor. Basic Books; 1953.

Fonagy P, Bateman A. Adversity, attachment, and mentalizing. *Compr Psychiatry*. 2016;64:59–66.

Garety P A, Kuipers E, Fowler D, Freeman D, Bebbington PE. A cognitive model of the positive symptoms of psychosis. *Psychol Med*. 2001;31(2):189–195.

*Garrett M. *Psychotherapy for Psychosis: Integrating Cognitive-Behavioral and Psychodynamic Treatment*. Guilford Press; 2019.

Garrett M, Ahmed A, Athineos C, et al. Identifying psychological resistances to using logic in cognitive-behavioral therapy for psychosis (CBTp) that limit successful outcomes for patients. *Psychosis*. 2019;11(4):287–297.

*Greenberg J. *I Never Promised You a Rose Garden*. St Martins Press; 1964.

Gunderson JG, Frank AF, Katz HM, Vannicelli ML, Frosch JP, Knapp PH. Effects of psychotherapy in schizophrenia: II. Comparative outcome of two forms of treatment. *Schizophr Bull*. 1984;10(4):564–598.

Isaacs S. The nature and function of phantasy. *Int J Psychoanal*. 1948;29:73–97.

**The International Society for Psychological and Social Approaches to Psychosis Book Series*. Routledge.

*Karon BP, VandenBos GR. *Psychotherapy of Schizophrenia. The Treatment of Choice*. Rowman & Littlefield; 2004.

Leff J, Williams G, Huckvale M, Arbuthnot M, Leff AP. Avatar therapy for persecutory auditory hallucinations: What is it and how does it work? *Psychosis*. 2014;6(2):166–176.

Lidz T. *The Origin and Treatment of Schizophrenic Disorders*. Basic Books; 1973.

Lotterman A. *Specific Techniques for the Psychotherapy of Schizophrenic Patients*. Madison: International Universities Press; 1996.

Lysaker P, Keane J, Culleton S, Lundin N. Schizophrenia, recovery and the self: An introduction to the special issue on metacognition. *Schizophr Res Cogn*. 2020;19:100167: 1–5.

Malmberg L, Fenton M. Individual psychodynamic psychotheraty and psychoanalysis for schizophrenia and severe mental illness. *Cochrane Database Syst Rev*. 2011;(3):CD001360.

Marcus ER. *Psychosis and Near Psychosis*. Revised Third Addition. Madison: International Universities Press; 2017.

Morrison AP, Renton JC, French P, Bentall RP. Think you're crazy? *Think Again: A Resource Book for Cognitive Therapy for Psychosis*. Routledge; 2008.

Read J, Fosse R, Moskowitz A, Perry BD. The traumagenic neurodevelopmental model of psychosis revisited. *Neuropsychiatry*. 2014;4(1):656–679.

Segal H. Notes on symbol formation. *Int J Psychoanal*. 1957;38:391–397.

Stephan K, Bach D, Fletcher P, et al. Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. *Lancet Psychiatry*. 2016;3:77–83.

Stone MH. The history of the psychoanalytic treatment of schizophrenia. *J Am Acad Psychoanal*. 1999;27(4):583–601.

Sullivan HS. *Schizophrenia as a Human Process*. Norton; 1962.

Summers A, Rosenbaum B. Psychodynamic psychotherapy for psychosis. Empirical evidence. In: *Models of Madness:Psychological, Social and Biological Approaches to Psychosis*. Routledge; 2013.

Tai S, Turkington D. The evolution of cognitive behavior therapy for schizophrenia: Current practice and recent developments. *Schizopr Bull*. 2009;35(5):865–873.

Turner DT, van der Gaag M, Karyotaki E, Cuijpers P. Psychological interventions for psychosis: a meta-analysis of comparative outcome studies. *Am J Psychiatry*. 2014; 171(5):523–538.

Wykes T, Steel C, Everitt B, Tarrier N. Cognitive behavior therapy for schizophrenia: effect sizes, clinical models, and methodological rigor. *Schizophr Bull*. 2008;34(3):523–537.

32.7 Treatment of Borderline Personality Disorder

LOIS CHOI-KAIN, M.D., M.ED., EVAN ILIAKIS, B.A., M.D./PH.D., GRACE MURRAY, B.A., AND ANA RODRIGUEZ-VILLA, M.D., M.B.A.

All health care professionals clinically manage patients with borderline personality disorder (BPD) in the routine delivery of medical, surgical, and mental health services. The emotional reactivity, self-destructive impulsivity, relational dysfunction, and stress-induced departures from realistic thinking that define BPD can overwhelm, destabilize, or frustrate the efforts of clinicians to provide standard-of-care interventions. Understanding that these are symptoms of a common, disabling, but treatable disorder can enhance health care providers' likelihood of adhering to best practices in their domain. However, adequate education and training for the recognition and management of BPD are not widely available, even for mental health clinicians.

Empirically validated manualized psychotherapies for BPD have revolutionized attitudes toward this highly stigmatized, once thought to be untreatable, illness associated with severe morbidity, mortality, and costs to society. These *specialized* therapies for BPD, most notably dialectical behavioral therapy (DBT), mentalization-based treatment (MBT), and transference-focused psychotherapy (TFP), yield comparable effects. No pharmacotherapy has shown significant, consistent superiority over a placebo; therefore, there are no Food and Drug Administration (FDA)-approved medications for BPD treatment. There is a significant disparity between the short supply of therapists trained in these specialized therapy approaches and the demand created by the high prevalence of patients with BPD seeking treatment, especially in acute mental and general health care environments. The intensive, complex, and lengthy formats of these treatments limit broad implementation outside specialist programs.

Clinical professionals of all disciplines, as well as students and trainees, need pragmatic tools that distill up-to-date knowledge about what works for the treatment of BPD and how that knowledge can be applied in cross-sectional, brief, or acute care situations. To meet this need, this chapter will review common factors from prevailing specialized psychotherapies that clarify BPD's features in constructive terms. Since most clinicians, even in mental health, will not deliver these specialist therapies, we will also describe good psychiatric management (GPM), a generalist approach that medicalizes BPD as a treatable diagnosis to create objective ways of discussing its symptoms and emphasizes tasks such as diagnostic disclosure, psychoeducation, goal setting, management of safety and comorbidities, and conservative prescribing that most clinicians use for the treatment of other disorders. To further relate GPM's approach to common clinical scenarios that most medical trainees and practitioners will encounter, GPM's applications to inpatient psychiatric units, emergency departments (EDs), and general health care settings will be discussed.

WHAT IS BPD: A REVIEW OF SYMPTOMS

In order to best understand the problems of any personality disorder, it is essential to first understand how personality usually functions. Everyone has a personality, and it defines who we are, both in terms of our self-concept and how others see us. Just as we all have moods and can be prone to mood disorder, under specific conditions of stress and maladaptation, we are also all prone to personality dysfunction. Personality functions as a mental envelope for our psychologies. On the one hand, it encompasses self-management (i.e., coping styles, self-direction, self-esteem, and coherent identity). On the other hand, personality functioning also involves relational management, which relies on capacities for understanding and collaborating with others. Personality functioning is optimal when it balances reliable containment of our thoughts, emotions, and impulses with adaptive flexibility as we navigate through our journeys to developmental milestones. Stressors can challenge the integrity of that envelope. Resilient functioning occurs when individuals can learn new means of self and relationship management in changing contexts, while also maintaining consistency and coherence. When personality is too porous or rigid to contain and adapt to life stressors or transitions, our personalities may show disorder in a variety of ways that fulfill criteria of different personality disorders. With instabilities in emotion, cognition, behavior, social relationships, and sense of self, BPD is thought to be a prototype and severity indicator for all personality disorders.

The DSM-5's diagnostic definition of BPD requires five or more of the following nine criteria. The most discriminating features that distinguish it from other disorders are its interpersonal criteria which include (1) frantic efforts to avoid abandonment and (2) unstable interpersonal relationships that vacillate between devaluation and idealization. Self-destructive action often brings people with BPD to clinical attention in the form of (3) impulsivity in at least two potentially damaging areas (e.g., substance use, reckless spending, binge eating), and (4) recurring suicidal threats, gestures, or self-harm. Emotional instability in the form of (5) affective instability due to reactivity of mood, (6) chronic feelings of emptiness, and (7) intense, uncontrollable anger also occurs. Affective instability is the feature that has the most negative predictive value, meaning in its absence, the patient is less likely to have BPD. Last of the BPD criteria are symptoms of cognitive disorganization including, (8) identity disturbance and/or unstable self-image, and (9) stress-related paranoid thoughts or extreme dissociative symptoms. These symptoms destabilize major domains of functioning for individuals with BPD, interrupting their development, increasing risk for both not finishing school or unemployment as well as poor physical and psychiatric health overall. Even one symptom of BPD increases risk for inpatient psychiatric hospitalization, suicide attempts, other psychiatric disorders, and greater psychiatric severity and dysfunction.

WHAT YOU NEED TO KNOW ABOUT BPD: PREVALENCE, COURSE, AND COSTS TO PATIENTS AND THEIR COMMUNITIES

BPD is common, affecting 1 to 3 individuals per 100 in the general adolescent and adult populations, and up to 6 individuals per 100 in terms of lifetime prevalence. This matches or exceeds the prevalence of other well-recognized disorders such as schizophrenia, obsessive-compulsive disorder, or bipolar disorder. Patients with BPD actively seek care and represent an even larger proportion of patients in clinical settings, so there is ample opportunity to provide them with care. In acute psychiatric settings, individuals with BPD are overrepresented, with studies demonstrating that patients with BPD account for 9% of all ED visits and 50% of recurrently suicidal patients in emergency rooms, and 20% to 25% of psychiatric inpatients meet criteria for BPD. They are also over-represented in primary care settings and psychiatric outpatient care practices. However, despite its high frequency, the BPD diagnosis is often missed or not disclosed.

Suicidality and deliberate self-harm are the most urgent features of BPD and often precipitate contact with health care. Four of five patients with BPD will have a history of engaging in some form of suicidal behavior, with an average number of attempts nearing three. Statistics show that 1 suicide per 23 attempts is completed, suggesting that in general, most attempts provide opportunity for intervention. Studies published early in the literature estimate that 10% of patients with BPD complete suicide, but the most recent studies, conducted since the introduction of multiple effective treatments became available, report rates of 3% to 5%. This figure still reflects a 54-fold increase in risk for completed suicide over the risk for the general population, based on data from published trials. Suicide attempts are associated with identity disturbance, frantic efforts to avoid abandonment, and emptiness for those with BPD, and these features are not routinely the focus of assessment and treatment planning. After suicidal crises temporize, the roots of these difficulties are often not identified or addressed, contributing to their chronicity and recurrence.

Despite the serious risks and severity of BPD, research on its longitudinal course reports that over 80% of individuals who at baseline were diagnosed with BPD achieved symptomatic remission over periods of 10 years or more with less than 25% experiencing relapse. These studies challenge the assumption that BPD is enduring and inflexible. Prospective studies of youths in the community describe fluctuations in BPD symptoms starting in childhood, peaking in adolescence, and normatively declining in early adulthood, highlighting the importance of maturational processes. These studies also show that youths who begin with more symptoms relative to peers maintain a higher level of BPD severity throughout time. These youths are more likely to meet criteria for BPD in adulthood, and also more likely to experience significant social and functional impairment with lower academic and occupational achievement, higher use of medical and psychiatric services, and more unstable social support even when BPD symptoms decline. Early identification of those at risk for BPD can curb the trajectory away from more chronic and severe forms.

While high rates of remission and low rates of relapse occur in adults with BPD, people with this disorder continue to struggle with significant functional impairment, medical comorbidity, extensive treatment utilization, suicide risk, and negative prognostic impact on major depression, anxiety disorders, and substance use problems, especially when BPD symptoms persist. The costs to society associated with BPD exceed that of most psychiatric conditions, including depression and anxiety disorders, and are comparable to severe psychiatric disorders such as schizophrenia. In addition, BPD symptoms increase risk of unemployment 22-fold and disability 15-fold, contributing to high indirect societal costs of the disorder. Moreover, the reduction of life expectancy is estimated to range between 13 and 25 years, when compared to the general population. People with BPD are more likely to encounter adversity. Treating BPD predicts better psychosocial functioning, improvement in relations with friends and family, greater resilience in the face of adverse life events, and fewer adverse life events overall.

These findings present a rationale for greater attention to BPD as a priority in clinical care rather than a diagnosis to be deferred as a last resort after treatment for comorbidities fails. Considering there are now numerous treatments that work for BPD and that remission can be facilitated in most patients, curbing some of the detriments of the disorder, there are few arguments against integrating attention to BPD early on in a patient’s treatment.

WHAT WORKS FOR BPD: COMMON FACTORS FROM EVIDENCE-BASED TREATMENTS COMBINED IN A GENERALIST APPROACH TO CARE

Various psychotherapeutic interventions can be used for BPD's treatment, spanning the gamut from cognitive behavioral to psychodynamic approaches. Common factors among these treatments likely lend a substantial degree of influence on outcomes. Human connection is a common factor in the effectiveness of psychotherapies. The fact that psychotherapies are the treatment of choice for BPD highlights the healing power of humanism. Human beings are highly social and rely on social networks for problem solving and healing. All of us naturally make sense of the world using folk psychology, using common wisdom of how people and the world work. For individuals with a personality disorder, their folk psychology, or their understanding of how the world works, is maladaptive. Their difficulties in understanding themselves and others in realistic and flexible ways limit their abilities to navigate the social world and use human connection in any healing practice, including traditional medical care. The development of a shared constructive therapeutic narrative between the patient and the provider can be used to enhance the social collaboration and human connection within any treatment relationship. The corrective experience in which patients have a more benevolent, empowering, and trust-building interaction with care providers can be the basis on which the patient can apply these healing myths in their activities outside of treatment.

A common maladaptive folk psychology for someone with BPD might be "I am bad" and "bad things happen to me" and/or "people aren't trustworthy, they always hurt me." The patient and clinician can work to replace that maladaptive belief with one that is more flexible and objective, such as "I am sensitive and when stressful things happen, I can't manage." This replacement narrative or healing myth serves as a foundation for creating expectations for change, for example, "stressful things happen so I need to find ways of coping reliably when they do." This replacement narrative reduces pessimism and fear while promoting a focus for a collaborative treatment process. The most transformative and integrative phase of treatment occurs when the patient can apply their healing narrative outside of treatment, starting to cope independently and building a life worth living despite ongoing vulnerabilities. This process of shared narrative building between the patient and their treater highlights how psychotherapy is a healing form of social influence. Figure 32.7–1 illustrates the therapeutic process of correcting maladaptive folk psychology.

Effective treatments for BPD share common ingredients. Fundamentally, they all propose various healing paradigms that define core factors that explain BPD's symptoms and also specify pathways to help patients optimize growth to remit and recover from the disorder. The ingredients of these healing paradigms are (1) a model of how BPD develops, emphasizing the transactional relationship between individual's particular innate characteristics and responsiveness (or lack thereof) in their immediate environment; (2) an explanation of how these BPD symptoms in turn continue to interrupt the development of social and emotional skills that are necessary to successfully navigate developmental milestones; and (3) recognition of how these difficulties leave those with BPD unable to develop an identity that is coherent and affirming.

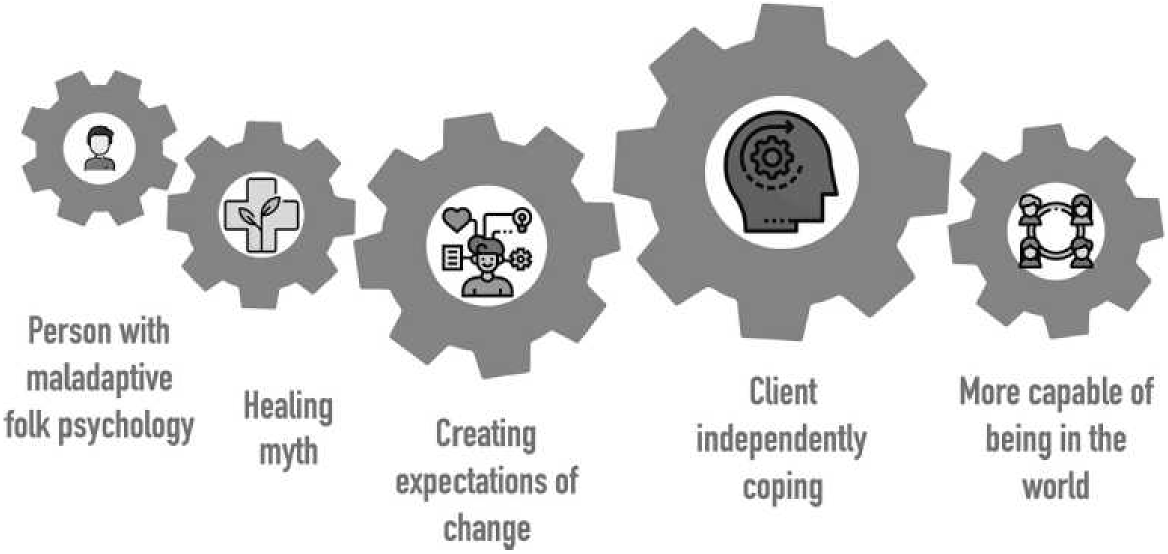
Person with maladaptive folk psychology
Healing myth
Creating expectations of change
Client independently coping
More capable of being in the world

FIGURE 32.7–1. The healing myth as a common factor in psychotherapy.

Leading evidence-based psychotherapies have transformed BPD from a diagnosis associated with hopelessness into a treatable entity by providing constructive healing myths that explain BPD, its symptoms, and possible remedies to reduce its detriments. They all employ healing paradigm that objectively describes BPD, including its origins, symptoms, and effects in a depersonalized and nonpejorative way. DBT proposes

BPD is at its core a problem of emotional dysregulation and points to skills to enhance abilities to tolerate distress, regulate attention and emotion, and be more interpersonally effective. MBT highlights attachment insecurity and instabilities in being able to understand oneself and others accurately, flexibly, and coherently (i.e., mentalize) as the core engine of BPD's symptoms. MBT, therefore, points to the stabilization of the capacity to mentalize or better understand oneself and others in social exchanges as a way to interrupt the process that gives rise to BPD symptoms. These formulations of why patients with BPD struggle with this specific profile of symptoms provide a hopeful and productive focus for clinicians and patients to organize their interactions and efforts more reliably toward a shared goal of recovery. To date, there is no proof that one of these treatments is more effective than the other. This finding may suggest that the key formulation of how BPD works, rather than any specific concept or technique, is what works in these treatments.

Generalist treatments have more recently emerged as a realistic alternative to more intensive and specialized treatments such as DBT and MBT when such options are either unavailable or not feasible. The empirical foundation for generalist clinical management for BPD was established with the publication of the two largest outpatient BPD psychotherapy studies comparing DBT and MBT with less intensive but well-informed treatment for BPD. Based on John Gunderson's essential clinical guide, general (later manualized as "good") psychiatric management (GPM) led to similar reductions in suicidality and self-harm, BPD symptoms, and depression as did DBT, with no differences in improvement in social functioning or quality of life. These outcomes were maintained 24 months post-treatment. A more recent meta-analysis confirmed comparable efficacy of both specialized and generalist protocolized treatments for BPD in adults.

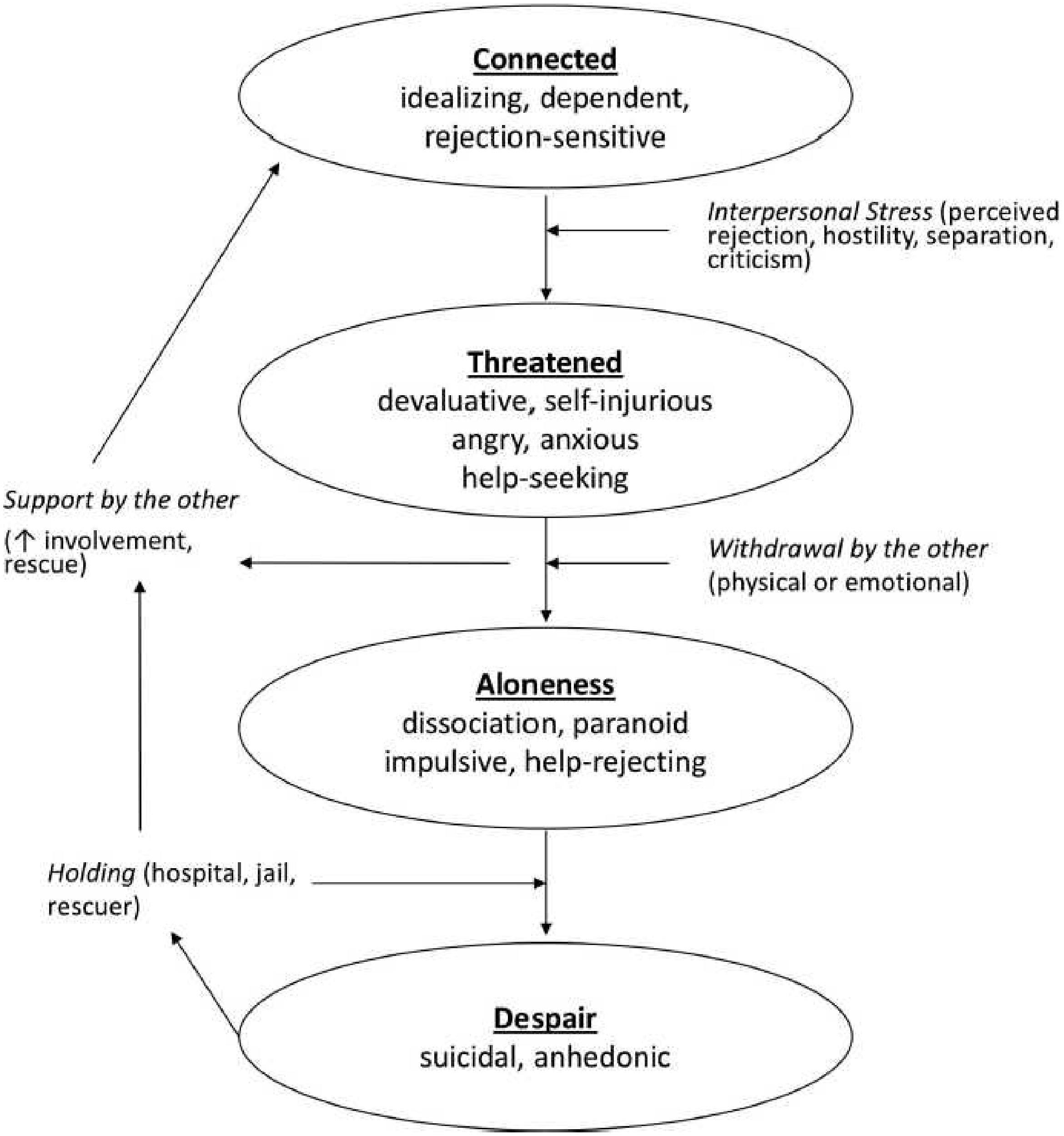
Connected
idealizing, dependent,
rejection-sensitive
Interpersonal Stress (perceived rejection, hostility, separation, criticism)
Threatened
devaluative, self-injurious
angry, anxious
help-seeking
Support by the other
(↑ involvement, rescue)
Withdrawal by the other
(physical or emotional)
Aloneness
dissociation, paranoid
impulsive, help-rejecting
Holding (hospital, jail, rescuer)
Despair
suicidal, anhedonic

FIGURE 32.7–2. GPM’s Interpersonal Coherence Model. (Reprinted from Gunderson JG, Links PS. *Handbook of Good Psychiatric Management for Borderline Personality Disorder*. American Psychiatric Publishing; 2014. Used with permission.)

Good Psychiatric Management

GPM is a principle-driven generalist approach to BPD grounded in the understanding of BPD as a problem of interpersonal hypersensitivity. GPM recognizes that most mental health care providers will not be trained intensively in psychotherapy, and few will have the specialized training and supervision that most evidence-based treatments for BPD require. Instead, it relies on the use of common factors in psychotherapy that most clinicians already know and use: diagnostic disclosure, psychoeducation, goal setting, multimodal treatment, crisis planning, a structured approach to psychopharmacology, managing comorbidities, and most importantly, functional rehabilitation. It is rooted in principles of the treatment of BPD shared by many other specialist and generalist treatments, substantiated by evidence from clinical trials and studies.

GPM's interpersonal coherence model (Fig. 32.7–2) provides a helpful roadmap to navigating the fluctuating states of an individual with BPD. It is another variant of a healing paradigm that allows the patient and clinician to discuss symptomatic events. Due to their interpersonal hypersensitivity, individuals with BPD are most stable and collaborative when they feel connected to someone who provides care in a way that makes them feel worthy. Their sense of self-worth is often rooted in the status of their exclusive relationships, leaving them intensely vulnerable and fearful of abandonment. In order to avoid abandonment, patients with BPD present as compliant and idealizing in a way that optimizes their existing capacities for collaboration. When interpersonal stressors invariably arise, patients enter a threatened state demonstrating anger amidst abrupt transitions between idealization and devaluation. In threatened states, people with BPD act aggressively toward caregivers and other important figures and frequently express this distress and anger toward themselves in the form of self-harm. Although threatened, patients are still desperately attempting to reinstate connection with important others.

Anger, cycling between idealization and devaluation, and self-harming behaviors predictably evoke fear, confusion, and anger in others, and often precipitate the very abandonment that people with BPD fear. As others withdraw, the patient with BPD enters a state of aloneness. In alone states, people with BPD lack the interpersonal anchor they rely on and become disoriented both cognitively and behaviorally. Here, patients with BPD are paranoid, disconnected from reality, recklessly impulsive and disregarding of consequences, and descend into the most unstable state of despair. Without the holding or grounding influences of others, people with BPD lose reasons to live and become more dangerously suicidal. At this point, concerned others may intervene unilaterally, putting patients in the hospital when suicidal, or prison when recklessly impulsive, in order to create a holding environment that reconstitutes either the connected or threatened state once again. The revival of interpersonal contact stabilizes the patient with BPD temporarily and while it may be necessary for short-term safety, without attention to the repeated patterns of interpersonal hypersensitivity underlying symptomatic states, this pattern will continue. Clinicians can expect that an individual with BPD will transition through the different states of this model. They can use the GPM model to understand how BPD symptoms manifest themselves in each state as well as how external factors, including the provider's own attitude and actions influence the patient and their reactions to care. Over time, this model can become a shared and explicit tool for both provider and patient.

While working toward greater understanding of the patients' symptoms through this model, GPM also promotes attention to patients' functioning with the imperative to get a life outside treatment. Informed by the outcomes of longitudinal studies showing that functional recovery remains challenged even with symptomatic remission, GPM focuses on promoting corrective experiences for people with BPD in work, school, or community-based activities where opportunities for contributing in specified roles and developing mastery are emphasized. By moving away from preoccupation with exclusive relationships, which are most prone to produce idealized expectations and dependency, people with BPD can benefit from a socializing environment that has more transparent social norms and expectations, as well as a larger social network for a more stable sense of connections. By understanding that people with BPD will have greater interpersonal hypersensitivity, tendencies to read others as untrustworthy, and difficulties with collaboration in behavioral exchange, clinicians and patients can work together to anticipate difficulties that will be encountered, aim to increase effectiveness of patients' social transactions in the service of patient's stated goals, and enhance social learning and self-esteem.

Table 32.7–1.
Principles of Good Psychiatric Management (GPM)

Principle	Description
1—Be active, not reactive	A discussion on the BPD diagnosis generally helps the patient commit to treatment. GPM recommends that the diagnosis is openly shared and discussed with the patient. Its complex etiology (including genetic and environmental aspects), clinical course, and treatability are explained. Tailored psychoeducation is necessary for patients, as well as for their loved ones and relatives.
2—Provide support	The patient may find it difficult to explore and face their core problems, but doing so is important, and the patient is actively supported by the clinician in the session.
3—Focus on life outside treatment	Life challenges are assessed in therapy. These difficulties are experienced by the patient as obstacles to having a meaningful life outside treatment, from which the patient can develop an identity and be connected to others. Therefore, it is recommended to focus on life outside of treatment.
4—Therapeutic relationship is real and professional	The GPM clinician pays special attention to relational issues in the patient's life. Relational issues between the patient and therapist are not GPM's primary focus, but are expected. Instead, links between relational events, internal feelings, and behaviors, both in the therapeutic relationship and in outside life, are understood. The therapist provides both a professional relationship and a

	personal one marked by authenticity when needed.
5—Change is expected	Change is expected in GPM. Symptomatic amelioration is expected when in treatment. A short-term treatment should already have some effects on certain aspects of BPD symptomatology since we assume that change starts with session one.
6—Promote responsibility	Psychoeducation and the clinician's attitude during the sessions promote responsibility. The patient should know that their active participation has an effect on their real-life difficulties and therapeutic outcomes.

BPD, borderline personality disorder; GPM, good psychiatric management.

Reprinted from Gunderson JG, Links PS. *Handbook of Good Psychiatric Management for Borderline Personality Disorder*. American Psychiatric Publishing; 2014. Used with permission.

Principles of Good Psychiatric Management of BPD

The following section elaborates on the core principles of GPM with tangible takeaways for clinical settings. Table 32.7–1 summarizes these principles.

Diagnostic Disclosure. Many clinicians fear that disclosing the diagnosis of BPD will cause harm to the patient. Secrecy around the diagnosis can increase the risk of iatrogenic harm (e.g., polypharmacy) and block the patient from receiving appropriate treatment. Disclosing the diagnosis informs the direction of care, allows accurate communication with colleagues, and is a good risk management strategy since it indicates a rationale for clinical decision making guided by evidence-based treatment. Diagnostic disclosure challenges the patient's maladaptive folk psychology through a clinical destigmatization process that localizes the patient's problems to a discrete disorder, rather than a pervasive "badness."

Diagnostic disclosure can be done in collaboration with the patient. The clinician can use the DSM-5TR criteria list for BPD as a concrete reference point to return to throughout the discussion. For each item on the list, the clinician and patient review the ways in which the criterion plays a role in the patient's life. This discussion naturally leads to the diagnosis and helps the patient understand their own profile of symptoms. The process should be approached collaboratively, emphasizing that BPD is common so that they are not alone and that their problems are understandable. Oftentimes, BPD patients have been misdiagnosed and treated with heavy-handed and ineffective interventions. Reframing the clinical focus of treatment on BPD can therefore restore hope that more effective treatment may exist.

Psychoeducation. Psychoeducation around BPD can begin with diagnostic disclosure, followed by a discussion about the favorable prognosis of the disorder, common comorbidities, self-harm and suicide risk, and available treatments. Psychoeducation around BPD should be realistic and emphasize that (1) evidence-based treatment likely facilitates remission; (2) the risk of suicide needs to be considered, given data about significant rates of suicide completion in the BPD population, and (3) many patients with BPD will continue to have challenges engaging in meaningful work and relationships even when they technically no longer meet criteria for the disorder.

Growing evidence suggests psychoeducation about BPD symptoms, comorbidities, etiology, course, and available treatments is a useful form of pretreatment in and of itself, delivered in person or remotely using technology, reducing BPD symptoms at 12-week and 1-year follow-up. Besides review of diagnostic criteria, essential topics of GPM psychoeducation include etiologic factors such as heritability, frontolimbic dysfunction, and neurohormonal disturbance that combine with life adversity, in addition to longitudinal course and core formulation of BPD as a disorder of interpersonal hypersensitivity. The basic message conveyed by the heritability and biologic basis of BPD is “it is not your fault” to both patients and family members, while also communicating, “It is your responsibility to optimize your health and wellbeing.”

Goal Setting/Functional Rehabilitation. Setting goals is a central task in the treatment of BPD that promotes agency, self-improvement, self-reliance, and productive collaboration rather than dependency on others.

GPM encourages patients to refocus their energies to build a meaningful and functional life outside of treatment, rather than remain mired in an identity as a patient. Getting a job is an important task within GPM that provides structure and interaction with others to counteract the sensitivity BPD patients have to loneliness. It also provides a means of building self-reliance and self-esteem that is not staked in an exclusive relationship. GPM advises that this strategy helps people become more sustainable romantic partners if one's goal is to be in a stable life partnership.

Case-management strategies can help guide the patient toward necessary life skills that can be achieved in the short term, such as making a budget, improving physical health, or maintaining hygiene. While GPM integrates concepts from psychodynamic psychotherapy with more practical life management strategies, the focus on taking concrete steps to build a meaningful life should be prioritized over exploration of the patient's developmental history. Employment status has been identified as one of the best predictors of remission from BPD, and unemployment has been associated with rapid response to treatment followed by relapse.

Multimodality. The use of different treatment interventions either in sequence or together is a key aspect of GPM's flexibility. The GPM therapist is open to including whatever might be of use to the patient, including group treatment options (DBT skills groups, 12-step programs, family coaching, and psychoeducation, along with adjunctive pharmacotherapy). As noted, GPM itself borrows from various approaches to treatment. Rather than considering concurrent treatment interventions as a liability or distraction, the GPM therapist would remain open to whatever might be beneficial during treatment. In general, connecting patients who are sensitive to being alone to a broader network of reliable and well-defined community settings can allow the patient access to social support without an active crisis.

GPM also integrates guidelines for psychopharmacology, based on the American Psychiatric Association (APA) Practice Guideline for the disorder (see psychopharmacology section for more information).

Crisis Planning and Management. Suicidal and self-injurious symptoms are to be expected and prepared for in the treatment of BPD. These symptoms are expressions of distress that require response with expressions of concern and careful risk assessment (e.g., the Columbia Suicide Severity Rating Scale), with the understanding that the baseline risk for suicide is chronically elevated in this population. The generalist clinician responds with the knowledge that acute exacerbations of risk for any person relates to comorbid mood disorders, substance use, withdrawal of support (e.g., recent hospital discharge), and negative interpersonal events.

Crisis planning should involve the patient as an active participant in creating a crisis plan at the beginning of treatment rather than as a crisis is unfolding. Safety plans for suicidal crises include identifying warning signs for a suicidal crisis, reduction of access to lethal means, coping strategies that the patient can use independently, ideas for social interaction to reduce isolation, and contact information for both supportive interpersonal relationships and professionals. In GPM, the patient and clinician collaboratively develop alternative ways to cope with unbearable emotions, with their intolerance of aloneness and interpersonal hypersensitivity in mind. The patient is thus encouraged to recruit additional social support to be available when they are feeling distressed or alone and also discouraged from engaging in relationships that are likely to be rejecting during crises. Actively involving the patient in the safety planning process, rather than asking them to passively agree to a “safety contract,” promotes a sense of agency and autonomy.

The suicidal or self-injurious patient predictably creates high anxiety in the clinician. Clinicians are therefore strongly encouraged to seek support, supervision, and consultation from colleagues. It is also important for a clinician treating BPD to be open and clear about their limits. No clinician is omniscient, omnipresent, or omnipotent, and it is therefore unreasonable for a patient’s safety to depend on the clinician’s availability. Having a plan that allows for different means of coping and gaining social support is most empowering to the patient, and their ability to rely on themselves to manage their safety over time is a constructive goal for treatment.

Psychopharmacology. Medications are adjunctive to psychosocial interventions for BPD. Medications can be helpful when they are targeted to specific symptoms (e.g., antipsychotics/mood stabilizers for anger, mood instability) or comorbidity (e.g., selective serotonin reuptake inhibitors for depression), but no medication is uniformly or consistently helpful, nor has been approved by the U.S. FDA for the treatment of BPD. Polypharmacy may increase the risk for side effects, promote dependency, and send the message that the patient can expect to improve by taking medications alone, as opposed to mobilizing to build a more satisfactory life. Due to the tendency to accumulate polypharmacy, especially with repeated hospitalizations, “collaborative deprescribing” or judicious and deliberate shared efforts with patients to taper medications when they are not working or reducing doses when possible, can become the focus of a treatment plan while inpatient or in primary care and outpatient psychiatric settings. Benzodiazepines should be avoided as they are habit forming and reinforce the belief of individuals with BPD that their emotions are unbearable and unmanageable.

Managing Co-occurring Disorders. Personality pathology should be prioritized in cases where the comorbid disorder is unlikely to remit otherwise or does not preclude learning. These disorders include major depressive disorder (MDD), non-manic bipolar disorder, panic disorder, bulimia, and post-traumatic stress disorder related to trauma that occurs in adult life. Recent findings suggest the presence of a substance use disorder does not impact the efficacy of brief GPM. However, comorbidities that interfere with active learning, including psychosis, manic states, or anorexia nervosa, should be prioritized before use of BPD treatment.

Duration and Intensity. GPM is distinguished by its relative flexibility about the length of treatment recommended and the intensity of the treatment as reflected in frequency of visits. In this regard, GPM differs from the other empirically validated treatments that require a set number of visits weekly and/or an expected duration of treatment. Greater flexibility allows for treatment of patients with BPD symptoms with a range of severity of illness and schedules. This means that relatively stable patients who are not in crisis, for example, may be seen sporadically, while those patients with prominent active difficulties can be seen more frequently as determined by the therapist. In situations where treatment is offered to patients who may not have a set schedule, such as college students who are away for vacations or summer break, this flexibility can allow for good enough treatment.

Treatment should continue only if the patient and therapist feel that progress toward stated goals is being made. The patient and provider should actively evaluate the benefit of the intervention. The idea that any medical intervention should only be continued if it is working frames the treatment as useful if improvement occurs, rather than solely required due to crises or desperate need. Patients vary in their readiness for and responsiveness to treatments, so if this occurs with generalist treatment, a referral for a different, more intensive treatment where available should be considered.

Applications of Good Psychiatric Management Principles to Different Care Settings. The minority of providers in inpatient, emergency, or general health care settings are trained in specialist treatments for BPD. What follows are principles of GPM of BPD in these settings so that any provider in these settings can become “good enough” in the treatment of this population. Table 32.7–2 summarizes the key principles of each application.

Inpatient Management of BPD. Patients with BPD constitute a large proportion of all psychiatric admissions (20% to 25%). These patients tend to spend more time in inpatient settings than many other disorders (14.3 weeks of hospitalization on average, compared to 2.4 weeks for patients with MDD) and have nearly five times as many separate hospitalizations. Given these facts, it is important that inpatient units are prepared to treat this patient population, but they are often not provided guidance on how to do so since most evidence-based treatments for BPD are not designed for implementation at this level of care. Hospitalization prevents unsafe action but often does not address the relationship between the patient’s BPD and their presenting problem. Hospitalization can become addictive when patients view it as necessary to their basic survival, and it serves a secondary function of separating the patient from interpersonal stressors and unwanted responsibilities in the patient’s life outside the hospital. To mitigate the harmful cycle of recurrent hospitalizations, inpatient units can implement the following basic principles: diagnostic disclosure and psychoeducation, structure, focusing on the patient’s life stressors, and aftercare planning. See Figure 32.7–3 for recommended phases of inpatient treatment.

Table 32.7–2.

Key Principles Guiding the Application of GPM in Health Care Settings

Application	Key Principles
Inpatient Unit	1. **Psychoeducation.** Focus on BPD as the primary diagnosis rather than comorbidities such as mood or anxiety disorders. Educate the patient about the limited benefits of medication. 2. **Structure.** Set goals and expectations. Clearly communicate mandatory and prohibited activities. Daily schedules should include multiple hours of groups and homework. Unstructured milieus can be actively harmful. 3. **Focus on life stressors.** Staff should engage patients in discussing the stressors that contributed to their admission and ways to minimize these upon discharge. 4. **Aftercare planning.** Schedule the first week after discharge, including a safety plan. Involve the patient in a group (e.g., AA, spiritual community, structured outpatient therapy).
Emergency Department	1. **Disclosure of a possible BPD diagnosis** in the ED can provide patients with reasonable expectations and information to discuss with outpatient providers. 2. **Building rapport and managing interpersonal hypersensitivity.** A detached interpersonal style could be interpreted as dismissive or rejecting by patients with BPD. Remain active and curious to form a rapport. 3. **Self-harm** is not a good indicator of acute suicide risk. Focus on acute risk factors and warning signs (e.g., stressful interpersonal events,

	concurrent major depression, substance abuse) to avoid unnecessary hospitalizations. 4. **Psychopharmacology.** Benzodiazepines and antipsychotics are the most commonly prescribed medications for BPD in the ED. These medications are not associated with significant benefits to the core symptoms of BPD and are associated with high rates of adverse effects.
General Health Care Settings	1. **Alliance Building**. Effective alliance building includes displaying concerned attention, active listening, validation of painful experiences, and demonstrating helpfulness through effective and practical assistance. GPM actively discourages overemphasis on rescuing a patient and splitting. 2. **Real Relationship.** Therapeutic relationships are two sided. A treater can own their mistakes, set limits about realistic expectations, and disclose their real emotional reaction to the patient's behavior where appropriate. 3. **Medication Management.** There are no FDA-approved medications for BPD treatment. Prescribe medication cautiously when needed, and where appropriate, explain its role as an adjunctive, short-term tool rather than a first-line defense. 4. **Diagnostic disclosure** of BPD can mitigate the iatrogenic effects of polypharmacy, derailing hospitalizations and aimless psychotherapies, as well as facilitating the patient's access to appropriate treatment.

AA, Alcoholics Anonymous; BPD, borderline personality disorder; FDA, Food and Drug Administration; GPM, good psychiatric management.

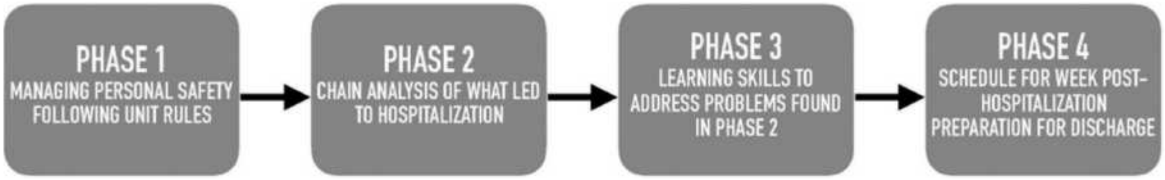
PHASE 1
MANAGING PERSONAL SAFETY
FOLLOWING UNIT RULES
PHASE 2
CHAIN ANALYSIS OF WHAT LED TO HOSPITALIZATION
PHASE 3
LEARNING SKILLS TO ADDRESS PROBLEMS FOUND IN PHASE 2
PHASE 4
SCHEDULE FOR WEEK POST-HOSPITALIZATION
PREPARATION FOR DISCHARGE

FIGURE 32.7–3. Recommended phases of care for BPD in inpatient settings.

DIAGNOSTIC DISCLOSURE AND PSYCHOEDUCATION. Mood or anxiety disorders are common primary diagnoses in inpatient settings. Listing BPD as the primary diagnosis is an important first step for inpatients with BPD to receive the care they need. Furthermore, in the United States, BPD falls under DRG 883 for Medicare patients, which is reimbursed at a higher rate than DRG 885, which includes MDD. Following diagnosis, clinicians can use standard GPM principles to proceed with psychoeducation.

STRUCTURE. A structured inpatient milieu benefits all patients, but it is especially important for patients with BPD. Unstructured time can lead to impulsive action and may reinforce a perception that the hospitalization is an escape from outside stressors. Useful structures in the milieu include active schedules and clear expectations. It is helpful for staff to clearly communicate both expected and prohibited activities, as well as resulting consequences and rewards that recognize their capacities for self-management (e.g., a privilege system). It is also beneficial for patients to set short-term goals that are achievable while holding a long-term goal in mind as motivation.

FOCUS ON LIFE STRESSORS. Since psychological growth is often not feasible in a short-term hospitalization, it may be more productive to focus on the real-life situations that need attention in discharge. Inpatient admissions are often preceded by self-destructive behaviors such as suicide attempts, which are usually triggered in BPD by real or perceived rejection or loss of support. It is important that staff engage patients in thinking critically about the stressors that contributed to their inpatient admission and support them as they develop a plan to mitigate the stressors after discharge. In some cases, this focus on life stressors can even lead to symptom remission (e.g., through disclosure of shameful stressors that once shared might bring support, leaving an abusive partner, or finding less stressful work). Group sessions in the milieu that focus on both identifying and mitigating stressors, as well as planning for educational or vocational activities postdischarge are especially useful to inpatients with BPD.

AFTERCARE PLANNING. A sound aftercare plan helps prevent relapse and readmission, as well as manage risk for suicide, which is 200 times the global rate for patients in the 3 months after discharge. Three key components of creating an aftercare plan are (1) identify in the *intake interview* the major life stressors, what additional supports are needed, and how school/vocational re-entry can be facilitated; (2) create a detailed schedule for the first week after discharge; and (3) develop a safety plan to manage suicidal crises that occur after discharge. With regards to the patient’s network of support after discharge, it is helpful for friends and families to know and prepare for the fact that patients with BPD tend to become more symptomatic in times of transition, and in some cases, inpatient clinicians may wish to work with the patient to re-evaluate the helpfulness of their prior outpatient treater. It is also important to note that the patient may not be ready to fully re-engage in school or employment; taking a small number of classes or volunteering may be the first step. Another valuable way to introduce structure and support in the patient’s life after discharge is to involve them in a group such as Alcoholics Anonymous, a spiritual community, or a structured outpatient group therapy. For more information on forming a safety plan, see the standard guidance outlined in the *Crisis Planning and Management* principle of GPM.

BPD in the Emergency Department (ED). Patients with BPD should be considered a special population warranting specific treatment guidelines in the ED due to their recurrent visits, increased disruptive behavior compared to other diagnoses, and chronic suicide risk. While an ED visit serves as a basic triage function for many patients with BPD, it can be an integral part of treatment for others. The following guidelines can help emergency room doctors make the visit a therapeutic experience for the patient with BPD.

DIAGNOSTIC DISCLOSURE AND PSYCHOEDUCATION. It’s common for ED clinicians to hesitate in making a definitive diagnosis because patients in crisis may present with features that are not typical for them. However, GPM maintains that diagnostic disclosure and psychoeducation are beneficial, particularly in cases where collateral information is available that supports the diagnosis or for patients who have made recurrent visits to the ED. In all cases, even where less supporting information is available, suggesting the possibility of the diagnosis and providing more information gives the patient something to discuss with their outpatient care providers and can give families hope.

BUILDING RAPPORT. The fast-paced, often chaotic nature of the ED necessitates efficiency and expeditious disposition where possible. It is challenging to connect with an individual with BPD who is in crisis but demonstrating genuine concern and adopting an active interview style pave the way. The interpersonal hypersensitivity common in BPD is an important consideration when building a rapport. A detached interpersonal style could be interpreted as dismissive or rejecting by patients with BPD, and a passive style inhibits alliance development. A good strategy for fostering connection includes active curiosity about the patient's experience and validating the high stress situation and the patient's emotional intensity in order to demonstrate caring and potentially inspire a more open, honest conversation. Providers can emphasize that the patient's honest disclosure and active participation are important aspects of providing the right care, while also validating the difficulty of sharing intimate experiences.

MANAGING SUICIDALITY AND SELF-HARM. Clinicians in the ED should use a model of "acute on chronic" risk for patients with BPD. This population has high rates of chronic suicidal thoughts and behaviors and self-injury, and it is important for ED clinicians to understand whether the patient wants and intends to die or if they are expressing feelings of loneliness, depression, or rejection. An unnecessary hospitalization will likely not be beneficial and may be iatrogenic. A step-by-step analysis of the events leading to the ED visit is a useful activity to undertake with the patient in order for both parties to better understand what led to the patient's crisis and where earlier intervention could have occurred. Ultimately, the decision to hospitalize a patient with BPD should be made on the basis of acute risk factors, which may emerge in the case of stressful interpersonal events, concurrent major depression, or ongoing substance use. Acute changes in symptoms or deviations from the patient's baseline are also potential indicators of risk, and family members or the patient’s outpatient team can often provide collateral information to assist in risk assessment. In the case that acute risk is apparent, short-term inpatient hospitalization should be considered until the crisis passes or sufficiently diminishes. If the patient is to be discharged to home from the ED, safety planning should include how to recognize warning signs, how to make their environment safer by means of reduction, and identifying who to contact when in crisis. ED staff can further mitigate risk in the days following discharge by placing follow-up phone calls to BPD patients. Also note that in the case of recurrent visits to the ED, clinicians may work with the patient and their outpatient therapist to evaluate the helpfulness of the current outpatient treatment.

PSYCHOPHARMACOLOGY. Patients with BPD tend to receive more medications in the ED than patients with other psychiatric diagnoses, but the medications commonly used for BPD in the ED (benzodiazepines, antipsychotics) do not significantly benefit the core BPD symptoms and often have adverse effects. In general, it is good practice to treat medication as a last resort, for use when acute agitation and/or anxiety endangers the patient or others, or when the patient’s symptoms render a thorough interview impossible. It is also important to note that benzodiazepines are habit forming, and patients with BPD may be especially prone to this due to their reactivity, intense reactions to stress, and affective instability. As with all patients, in the absence of imminent safety risks, verbal de-escalation is an appropriate first-line strategy to managing agitation and anxiety. If medication is deemed necessary, the clinician should clearly communicate that the purpose for the medication is to move forward with evaluation and not to provide a long-term treatment.

Changing a medication regimen in the ED is typically not recommended, given the usual lack of historical information regarding the patient’s baseline. However, if a medication change is deemed necessary, it is helpful to contact the outpatient provider for collaboration.

FAMILY INTERVENTIONS. The patient’s family and loved ones can be especially helpful when a patient with BPD presents to the ED because they can provide collateral information about the patient’s level of risk and baseline symptoms. BPD symptoms can be diminished or exacerbated by families, and an ED visit is also an intervention point where the clinician can provide psychoeducation and guidance to the family. Furthermore, some patients with undiagnosed and likely BPD present to the ED with several diagnoses and a history of ineffective treatments. In these cases, families may feel desperate or lost, and psychoeducation about the diagnosis of BPD itself, its appropriate treatment, and how the family may be able to help can be exceptionally useful.

Treatment of BPD in General Health Care Settings. Patients with BPD disproportionately experience medical complications compared with matched cohorts and have high rates of health care and surgical care utilization. Medical complaints that are more common among people with BPD include poor sleep, regular nightmares, obesity, metabolic syndrome, functionally incapacitating headaches, and polycystic ovaries. Health care providers are encouraged to keep their stated roles, goals, and standards of care in mind, using an understanding of BPD and its basic treatment strategies to minimize destabilization or departures from the routine provision of care. Providers who are not mental health professionals are not expected to provide definitive treatment for the patient's BPD-related problems. The following principles may be useful for treaters in nonpsychiatric, nonemergent health care settings to optimize their care for other conditions in patients who have BPD.

ALLIANCE. Effective alliance building often includes (1) displaying concerned attention, (2) active listening, (3) validation of painful experiences, and (4) demonstrating one's helpfulness through providing effective and practical assistance. GPM principles discourage overemphasis on rescuing a patient and splitting (both where the patient views part of the team as good and the others as bad, as well as where part of the team views the patient as innocent while others view them as problematic). According to GPM, a strong alliance is "real" (two sided), which includes (1) owning one's mistakes, (2) limit setting around unrealistic expectations, and (3) using self-disclosure of one's emotional reactions to the patient's behavior when appropriate. Grounding the terms and expectations of the roles of the patient and clinician by focusing on the diagnoses, treatment plans, and immediate goals at hand can provide focus or organization to the interaction. Making recommendations with a collaborative effort to understand risks and benefits, even if repeated like a broken record, is what clinicians can do in their role. The patient has the choice to follow them or not. The use of case management services to reduce obstacles may assist the patient in reducing barriers to follow through.

MEDICATION MANAGEMENT. There are a number of potential pitfalls with medication management that most medical and surgical professionals may not have been trained to know. First, it is important for both the care team and the patient to know that there are no FDA-approved medications for BPD, and many medications come with more risks than benefits. Low-dose antipsychotics can be helpful on a short-term basis for psychotic-like symptoms (paranoia, extreme agitation, hallucinations). In order to avoid reinforcing the idea that medications are first-line rescue interventions for distressing symptoms, anxiolytic medications can be prescribed as standing orders rather than as needed. Similarly, while non–habit-forming medications that promote sleep may reduce vulnerability factors, behavioral interventions are sufficient for some patients. If medication is necessary, the clinician can explain that medication for sleep promotion is a short-term intervention rather than a long-term treatment.

Antidepressants and mood stabilizers may be useful in the case of untreated comorbidity (e.g., mania) but are not the intervention of choice for BPD symptoms. If short-term medications have been used during hospitalization or medical treatment, GPM recommends that they be discontinued at discharge to avoid polypharmacy. Patients with BPD are at high risk of misusing analgesic medications, and referrals to structured pain management programs or expedited appointments with the primary care physician should be made as needed. The standard as-needed medications for mild symptoms such as constipation, nausea, indigestion, and headaches can be given if these symptoms are related to the underlying illness that requires hospitalization but minimizing their use where possible is encouraged to mitigate overreliance on medication.

DIAGNOSTIC DISCUSSION. GPM maintains the recommendation of discussing the diagnosis of BPD, even in nonpsychiatric settings. For those patients with a prior diagnosis of BPD, reviewing how it may influence care in other settings can be helpful when broached prior to conflict or crisis. Diagnostic discussion of BPD can mitigate the iatrogenic effects of polypharmacy, derailing hospitalizations and aimless psychotherapies, as well as facilitating the patient’s access to appropriate treatment. Bringing up the question of the BPD diagnosis to the patient in helpful terms, rather than reactively and punitively, is what most health care providers can do without a definitive diagnosis if a referral or further information is needed. While many clinicians worry about a negative reaction after diagnostic disclosure, patients with BPD who are told about their diagnosis after an initial evaluation are as satisfied with the evaluation as patients without BPD.

CONCLUSIONS

In this chapter, we present tools for the clinical management of BPD, a common, disabling, and painful disorder that increases the risk for negative medical, psychiatric, and functional outcomes, including completed suicide. All health care professionals can benefit from knowing basic facts about BPD and applying constructive frameworks for understanding the instability of patients with this disorder as symptoms that require treatment and management rather than a reflection of a bad patient or poor care. The more consistently BPD is addressed in the patient’s overall care across different clinical environments by different members of their care team, the more stable, organized, and predictable treatment can be. As the patient improves in their capacity to function, both in their roles and responsibilities as a patient and, more importantly, in life outside of treatment, they can attain a more stable sense of self and more positive expectations of others.

Further Readings

Bateman A, Fonagy P. Randomized controlled trial of outpatient mentalization-based treatment versus structured clinical management for borderline personality disorder. *Am J Psychiatry*. 2009;166(12):1355–1364.

Bender DS, Dolan RT, Skodol AE, et al. Treatment utilization by patients with personality disorders. *Am J Psychiatry*. 2001;158(2):295–302.

Black DW, Blum N, Pfohl B, Hale N. Suicidal behavior in borderline personality disorder: prevalence, risk factors, prediction, and prevention. *J Pers Disord*. 2004;18(3):226–239.

Choi-Kain LW, Reich DB, Masland SR, Iliakis EA, Ilagan GS. Longitudinal course of borderline personality disorder: what every clinician needs to know. *Curr Treat Options Psychiatry*. 2020;7:429–445.

Cristea IA, Gentili C, Cotet CD, Palomba D, Barbui C, Cuijpers P. Efficacy of psychotherapies for borderline personality disorder: a systematic review and meta-analysis. *JAMA Psychiatry*. 2017;74(4):319–328.

Fineberg SK, Gupta S, Leavitt J. Collaborative deprescribing in borderline personality disorder: a narrative review. *Harv Rev Psychiatry*. 2019;27(2):75–86.

Gunderson JG, Links PS. *Handbook of Good Psychiatric Management for Borderline Personality Disorder*. American Psychiatric Pub; 2014.

Gunderson JG, Stout RL, Shea MT, et al. Interactions of borderline personality disorder and mood disorders over 10 years. *J Clin Psychiatry*. 2014;75(8):829–834.

Hastrup LH, Jennum P, Ibsen R, Kjellberg J, Simonsen E. Societal costs of borderline personality disorders: a matched-controlled nationwide study of patients and spouses. *Acta Psychiatr Scand*. 2019;140(5):458–467.

Iliakis EA, Sonley AKI, Ilagan GS, Choi-Kain LW. Treatment of borderline personality disorder: is supply adequate to meet public health needs? *Psychiatr Serv*. 2019;70(9):772–781.

Linehan MM, Armstrong HE, Suarez A, Allmon D, Heard HL. Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Arch Gen Psychiatry*. 1991;48(12):1060–1064.

McMain SF, Fitzpatrick S, Boritz T, Barnhart R, Links P, Streiner DL. Outcome trajectories and prognostic factors for suicide and self-harm behaviors in patients with borderline personality disorder following one year of outpatient psychotherapy. *J Pers Disord*. 2018;32(4):497–512.

McMain SF, Guimond T, Streiner DL, Cardish RJ, Links PS. Dialectical behavior therapy compared with general psychiatric management for borderline personality disorder: clinical outcomes and functioning over a 2-year follow-up. *Am J Psychiatry*. 2012;169(6):650–661.

McMain SF, Links PS, Gnam WH, et al. A randomized trial of dialectical behavior therapy versus general psychiatric management for borderline personality disorder. *Am J Psychiatry*. 2009;166(12):1365–1374.

Oldham JM, Gabbard GO, Goin MK, et al. *Practice Guideline for the Treatment of Patients with Borderline Personality Disorder*. American Psychiatric Association; 2001.

Penzenstadler L, Kolly S, Rothen S, Khazaal Y, Kramer U. Effects of substance use disorder on treatment process and outcome in a ten-session psychiatric treatment for borderline personality disorder. *Subst Abuse Treat Prev Policy*. 2018;13(1):10.

Ridolfi ME, Rossi R, Occhialini G, Gunderson JG. A clinical trial of a psychoeducation group intervention for patients with borderline personality disorder. *J Clin Psychiatry*. 2019;81(1):19m12753.

Sharp C, Wright AGC, Fowler JC, et al. The structure of personality pathology: both general (‘g’) and specific (‘s’) factors? *J Abnorm Psychol*. 2015;124(2):387.

Soeteman DI, Verheul R, Busschbach JJV. The burden of disease in personality disorders: diagnosis-specific quality of life. *J Pers Disord*. 2008;22(3):259–268.

Stanley B, Brown GK. Safety planning intervention: a brief intervention to mitigate suicide risk. *Cogn Behav Pract*. 2012;19(2):256–264.

Stoffers J, Völlm BA, Rücker G, Timmer A, Huband N, Lieb K. Pharmacological interventions for borderline personality disorder. *Cochrane Database Syst Rev*. 2010;(6):CD005653.

Temes CM, Frankenburg FR, Fitzmaurice GM, Zanarini MC. Deaths by suicide and other causes among patients with borderline personality disorder and personality-disordered comparison subjects over 24 years of prospective follow-up. *J Clin Psychiatry*. 2019;80(1):18m12436.

Trull TJ, Jahng S, Tomko RL, Wood PK, Sher KJ. Revised NESARC personality disorder diagnoses: gender, prevalence, and comorbidity with substance dependence disorders. *J Pers Disord*. 2010;24(4):412–426.

Wampold BE. Humanism as a common factor in psychotherapy. *Psychotherapy (Chic)*. 2012;49(4):445.

Yen S, Pagano ME, Shea MT, et al. Recent life events preceding suicide attempts in a personality disorder sample: findings from the collaborative longitudinal personality disorders study. *J Consult Clin Psychol*. 2005;73(1):99.

Yen S, Peters JR, Nishar S, et al. Association of borderline personality disorder criteria with suicide attempts: findings from the Collaborative Longitudinal Study of Personality Disorders over 10 years of follow-up. *JAMA Psychiatry*. 2021;78(2):187–194.

Zanarini MC, Conkey LC, Temes CM, Fitzmaurice GM. Randomized controlled trial of web-based psychoeducation for women with borderline personality disorder. *J Clin Psychiatry*. 2018;79(3):16m11153.

Zanarini MC, Frankenburg FR, Reich DB, Fitzmaurice G. Attainment and stability of sustained symptomatic remission and recovery among patients with borderline personality disorder and axis II comparison subjects: a 16-year prospective follow-up study. *Am J Psychiatry*. 2012;169(5):476–483.

Zanarini MC, Frankenburg FR, Reich DB, Harned AL, Fitzmaurice GM. Rates of psychotropic medication use reported by borderline patients and axis II comparison subjects over 16 years of prospective follow-up. *J Clin Psychopharmacol*. 2015;35(1):63.

Zimmerman M, McGonigal P, Moon SS, et al. Does diagnosing a patient with borderline personality disorder negatively impact patient satisfaction with the initial diagnostic evaluation? *Ann Clin Psychiatry*. 2018;30(3):215–219.

32.8 Behavior Therapy

Jenessa S. Price, Ph.D.

HISTORY OF BEHAVIOR THERAPY

Behavioral treatment of psychiatric symptoms has existed since at least the first century, when Pliny the Elder used a type of aversion therapy to modify drinking behavior by putting spiders at the bottom of a drinker's glass. In the early 19th century, Alexander Maconochie set up a token economy wherein prisoners were awarded points for positive behaviors in a penal colony at Norfolk Island, Australia. The formal application of learning principles and the movement toward a behavioral approach to the treatment of abnormal behavior emerged in the 1900s.

The first reference to the concept of behavior modification occurred in 1911 by Edward Thorndike (*Provisional Laws of Acquired Behavior or Learning*). In 1920, the classic case of "Little Albert" was published, demonstrating the application of classical conditioning principles to the acquisition or learning of fear. In a variant of the famous Pavlovian conditioning paradigm, Little Albert was conditioned to fear a white rat when steel bars were banged together over his head while a white rat was presented to him. These principles were similarly applied to the deconditioning or unlearning of fear in another child, Peter, who was afraid of rats, rabbits, and other furry objects. To decondition the fear, a caged rabbit was gradually moved closer to Peter while he was eating a favorite food. Repeated and gradual pairings of the rabbit and the pleasant food led to a reduction in fear. Fourteen years later, another set of investigators utilized classical condition principles with the use of a *bell-and-pad* technique for treating enuresis or bed wetting. A bell (buzzer), embedded in a pad on the child's bed, sounded when the pad got wet, conditioning the child to awaken at the onset of urination and contract the sphincter. Ultimately, the child learned to awaken at the sensation of a full bladder without needing the bell/buzzer.

Behavior therapy evolved rapidly in the mid-1950s in three distinct areas of the world: the United States (B. F. Skinner), South Africa (Joseph Wolpe), and the United Kingdom (Hans Eysenck, Stanley Rachman). A driving factor was the pursuit of scientifically based paradigms and the need for effective short-term therapies for anxiety and depression in World War II veterans returning with significant emotional problems. The characteristic feature of this “first-wave” behavior therapy movement was a focus on classical conditioning and operant learning.

The “second wave” of behavior therapy occurred in the 1970s, which was characterized by increased focus on information processing, rather than observable behaviors alone, and paved the way for a more integrated cognitive-behavior therapy (CBT). Much of this work derived from Albert Bandura’s *social learning theory,* which posited that people learn not only through the process of classical and operant conditioning but also by observing the actions of others in their environments. The process of *observational learning* emphasized the roles of thoughts, images, and expectations in the development of psychological or psychiatric disorders and set the stage for treatments that target these internal components of behavior more directly. The first two major cognitive-behavioral treatments, Aaron Beck’s CBT for depression and Albert Ellis’ rational emotive therapy (RET), had major impact on the field.

Although “second-wave” CBT is at present the dominant contemporary system of psychotherapy worldwide, the “third-wave” movement largely emerged from a component analysis study directed by the late Neil Jacobson, wherein a behavioral activation (BA) intervention was as effective as the more comprehensive CBT treatment of individuals with major depression. This finding was further highlighted in a 2006 study where Sona Dimidjian and colleagues demonstrated that BA was as effective as antidepressant medication and more effective than cognitive therapy for individuals with severe depression. This “third wave” of behavior therapies is characterized by themes new to behavior interventions: metacognition, cognitive fusion, emotions, acceptance, mindfulness, dialectics, values, spirituality, and the therapeutic relationship. Today, behavior therapy (including CBT) has become one of the most well-accepted and well-researched approaches for treating many psychiatric symptoms and disorders.

PRINCIPLES OF BEHAVIOR THERAPY

Here, “behavior therapy” refers to the many interventions and programs combining behavioral (and cognitive) techniques demonstrated as efficacious for a range of psychiatric disorders. Despite the breadth of these techniques and treatments, however, all variations in behavioral treatments share certain basic assumptions or principles.

- *Maladaptive behaviors are acquired through learning, according to the same principles that govern the learning of adaptive behaviors.*
- *It is not necessary to identify an underlying cause or motive for maladaptive behaviors.*
- *Learning principles can be used to modify maladaptive behavior.*
- *The focus of treatment is on operationally defined factors that maintain current behavior rather than on historical issues.*
- *Psychologists and therapists should be knowledgeable about scientific literature relevant to the patients they treat.*
- *It is important to set specific, measurable treatment goals and measure outcomes.*

TECHNIQUES OF BEHAVIOR THERAPY

Assessment and Conceptualization

The primary goals of an initial behavioral assessment are to identify and define target behaviors and the variables that maintain them and to conduct a *functional assessment*. This would include historical information about the onset and course of symptoms, relationships with environment and/or others, and other related factors (e.g., substance use, medical issues). A functional assessment should lead directly to a treatment plan as it establishes causal (or functional) links between problematic behaviors and the internal and external antecedents and consequences that serve to maintain them. When these links are identified, learning principles and treatment techniques from empirically supported interventions can be applied to modify antecedents, behaviors, and consequences such that problematic symptoms are reduced. Assessment continues as treatment is implemented to ensure that appropriate changes are occurring.

Methods of Behavioral Assessment

As with many other theoretical approaches, the individual conducting behavior therapy often begins the assessment with clinical interviews. Additional strategies help establish links between target behaviors and maintaining factors, evaluate symptom severity against normative data, and set the stage for measuring outcomes over the course of treatment.

Unstructured and Structured Interviews. The clinical interview has tremendous variability, ranging from a primarily unstructured and completely flexible approach to structured methods that are more restrictive and goal directed. Several positive correlates are associated with unstructured methods, including increased rapport (and thus engagement in psychotherapy), observation of the patient's response style, and opportunity to gather unique details of a patient's history. Most contemporary practitioners allow for some degree of flexibility, although most also make use of some type of intake form or checklist relevant to their particular practice to facilitate the assessment process. The goals of initial interviews involve gathering data on target behaviors and maintaining factors, toward the goal of *operationalizing* all components specifically for the individual patient. In addition, the interview includes psychoeducation about the behavioral philosophy, including expectations about day-to-day approaches taken outside of the session. Behavioral approaches typically emphasize individualized assessment given that target behaviors and maintaining factors may vary greatly across patients with common presenting complaints.

Nevertheless, those conducting behavior therapy typically assign DSM-5-TR based diagnoses because they help guide selection of appropriate evidence-based techniques (while also providing evidence to payors of the utility of treatment). As a means to optimize reliability and validity across providers (and in the context of research), the most commonly used structured (clinician-rated) interviews are the Structured Clinical Interview for DSM-IV (SCID), for which there are both Axis I and Axis II versions; the Mini-International Neuropsychiatric Interview (MINI), a brief interviews are the Structured Clinical Interview for DSM-5 (SCID-5); the Mini-International Neuropsychiatric Interview (MINI), available for both adults and children/adolescents; and the Primary Care Evaluation of Mental Disorders (PRIME-MD), which also is brief and was designed for use by primary care clinicians. Other diagnostic interviews such as the Anxiety Disorders Interview Schedule for DSM-5 (ADIS-5), the Schedule for Affective Disorders and Schizophrenia (SADS), and the Addiction Severity Index (ASI; typically used alongside a substance use history/frequency interview) are available for collecting more detailed information on a limited set of diagnostic categories and functioning.

Questionnaires and Clinician Rating Tools. Self-report questionnaires allow for normative comparison of symptom severity. They can screen and augment assessment data, monitor progress across sessions or at longer follow-up intervals, and assess change associated with various psychological and pharmacologic interventions. The majority of these instruments have adequate excellent psychometric properties and assess a wide range of symptoms. The Patient Health Questionnaire (PHQ-2; PHQ-9) is brief and commonly used to screen for depression in primary care or other health settings. The clinician-rated Hamilton Rating Scales for Depression (HRSD) and Anxiety are useful toward assessing the severity of depressive and anxiety symptoms and measuring patient progress and treatment outcome. The 24-item *Brief Psychiatric Rating Scale (BPRS)* is one of the most commonly used clinician-rated scales of depression, anxiety, psychotic symptoms, and other unusual behaviors. More comprehensive personality tests such as the *Minnesota Multiphasic Personality Inventory (MMPI) and Personality Assessment inventory (PAI)* may provide useful assessment data on various psychological symptoms such as anxiety, depression, hypochondriasis, psychosis, and symptoms of personality disorders; these questionnaires require significant time for the patient to complete and comprehensive training to ensure accurate interpretation.

Self-Monitoring. Self-monitoring techniques are in contrast to retrospective assessments, wherein the patient observes, measures, and records behavior as it occurs. Self-monitoring can be used to evaluate the frequency and duration of symptoms as well as to identify sequences of events, behaviors, thoughts, and feelings that may provide information about antecedent and consequent factors that maintain problematic behaviors. Techniques can be simple, such as noting each time they smoke a cigarette, have a panic attack, binge eat, or perform a ritual. Daily diary logs can augment frequency data to explore the difficulty, enjoyment, and reward value of activities and behaviors. More complicated monitoring procedures can be used to evaluate behavioral sequences and environmental or internal antecedents and consequences;

for example, one can record precipitating situations, associated thoughts and feelings, physical symptoms, and consequent behaviors with an anxiety experience. Self-monitoring can be used to assess treatment progress. Data plots of symptom severity over time can be a highly useful visual indicator of treatment efficacy for both the patient and psychologist or therapist. Self-monitoring is also a behavioral component of the treatment program, not just of the assessment process. Recording one's own behavior is a *reactive* process, meaning that awareness of behavioral antecedents and the behavior being recorded often change as a result of monitoring.

Behavioral Observation. Like self-monitoring, behavioral observation involves someone other than the patient measuring the operationalized target behavior as it occurs, with attention to frequency, duration, and associated contextual variables. Observations of this type might involve *event recording,* where each episode of the target behavior is monitored, or *interval recording,* where the frequency of the behavior is monitored as it occurs during a restricted interval of time. In the case of children with behavioral problems, parents and teachers may track retrospectively through use of measures such as the *Child Behavior Checklist (CBCL)* or *Behavior Assessment Scale for Children (BASC-3)* or they may conduct direct observation in the classroom environment using a structured *Student Observation System.* The patient may be observed in the natural environment or engage in *analog* or *simulated* situations in session (e.g., having a conversation with a patient with social anxiety disorder). *Role plays* are a common method of simulated observation and provide opportunity to observe interpersonal skills. Patients in these natural or simulated environments also can be asked to rate the severity of their anxiety or depression using a standardized rating scale or Subjective Units of Distress Scale (SUDS), in which patients rate distress from 0 to 10 (or 100).

Case Conceptualization. A *case conceptualization* integrates assessment data and information about historical context and formulates a behaviorally focused treatment plan to address multiple disorders and their presenting features impacting functioning. The core difficulties then are examined in terms of predisposing, precipitating, and maintaining factors. *Predisposing factors* reflect both biologic components (e.g., genetic influences) and those elements of patients' learning histories that put them at high risk of developing significant symptoms. For example, a patient may be predisposed to develop anxiety or depression due to a strong family history of mental health problems, being raised by authoritarian parents who provided minimal nurturance and were overly controlling and manipulative, or experiences with parents that were perceived as cold, rejecting, and invalidating. *Precipitating factors* are those events that occur at the start of significant psychiatric difficulties. For instance, this same patient may have recently ended an unpleasant interpersonal relationship in which she frequently tried to express her thoughts and feelings that largely were ignored, minimized, and invalidated by her partner. She also might have simultaneously lost her job, resulting in decreased income and less frequent contact with her social support system. *Maintaining factors* are those antecedents and consequences that reinforce and maintain problem behaviors once they have started. The patient who is experiencing chronic worry may be receiving feedback from parents or significant others that inadvertently reinforces the patient’s fears and worry behaviors. Careful analysis of these factors identifies targets for empirically supported behavioral interventions and consideration of each patient’s unique learning history.

“First-Wave” Behavior Therapy Treatment Strategies

Here, we discuss a number of behavioral interventions, following the course of the three “waves” (or movements) discussed in the introduction.

Progressive Muscle Relaxation. Relaxation often is used as a component of systematic desensitization therapy (presented below), as a primary intervention for stress and some medical disorders (headache, pain, postsurgical distress), and as part of a comprehensive intervention strategy for reduction in anxiety. Relaxation therapy increases parasympathetic nervous system activity, thereby decreasing the opposing sympathetic nervous system activity and decreasing arousal. Several different approaches can be used to produce a relaxation response. The most popular is progressive deep muscle relaxation (PDMR), in which patients are taught to tense and relax muscle groups in their bodies while focusing on the differing sensations of tension and relaxation. PDMR begins with diaphragmatic breathing, followed by tensing and relaxing each of the identified muscle groups (e.g., feet, legs, hands, arms, shoulders, face). The act of tensing before relaxation serves to help patients identify tension, including tension they may have not been aware of, and facilitate relaxation. A few negative side effects have been associated with relaxation. These include muscle soreness or strain, depersonalization (when practiced continuously for several hours per day, or trauma re-experiencing in the context of general relaxation or visualization techniques), and relaxation-induced panic attacks (occasionally seen among patients with panic disorder who are attuned to physical sensations).

Exposure Therapies. Exposure therapies are based on the premise that fears are acquired through associative learning (classical or operative conditioning). Interventions to eliminate fear require exposure (contact) with the feared object, event, or situation. Exposure therapy is based on two models: The *counterconditioning model* pairs a feared object with an incompatible response. The *extinction model* is based on habituation of the feared response.

Systematic desensitization consists of relaxation therapy (usually PDMR) and the presentation of fear-producing stimuli arranged within a hierarchy representing successive approximations to the feared object, situation, or event. SUDS are rated for each item on the list. The intervention focuses on achieving a relaxed state and pairing with the first item on the list, with interruptions to imagining if anxiety returns. When the patient reports the imagery no longer elicits anxiety, the psychologist or therapist proceeds to the next item, and over sessions, the cycle is repeated until the patient is able to imagine the items at the top of the hierarchy without reporting any anxiety symptoms. In *in vivo desensitization,* partial relaxation, deep breathing, or no competing response is used as the patient is actively moving. Real-life hierarchical steps are developed, but each step represents only a small incremental increase in arousal in comparison with the preceding step.

An exposure procedure that uses the extinction process to reduce anxiety is known as *flooding.* The term *extinction* describes the process of decreasing fear and maladaptive anxiety. In flooding, the patient is exposed immediately to the fear-producing stimuli at its full strength and without relaxation. In most cases, flooding (exposure) is accompanied by response prevention (RP) or discontinuing all avoidance and other anxiety-reducing behaviors, so as to eliminate the emotional response over time (used in obsessive-compulsive disorder).

When a feared stimulus is first presented, anxiety typically rises significantly. Over time, with continued exposure, indicators of anxiety (SUDS levels, physiologic indices like heart rate, skin temperature) should decrease. Reduction in anxiety is often referred to as *within-session extinction.* Across a number of sessions, the key elements for successful treatment are that the peak anxiety level is reduced and the time to return to minimal anxiety is shortened. These responses to treatment are called *between-session extinction,* when the fear-producing stimulus produces reduced amounts of anxiety and faster within-session extinction over sessions.

Two relatively new strategies may hold promise. In *virtual reality exposure,* the patient wears a head-mounted display that produces computer-generated images of feared situations and "interacts" with what feels like a real-life situation. This has been used to treat phobias, posttraumatic stress disorder (PTSD), social anxiety disorder, and panic disorder. Second, eHealth utilizes technology-based self-help (i.e., computer or smartphone applications) techniques.

Operant Strategies. *Operant* is a term used to describe a mechanism whereby a behavior operates on the environment. In this learning model, behavior is not evoked or elicited (as in classical conditioning) but independently occurs, and environmental responses influence whether the behavior is continued or stopped. Changes in behavior occur as a result of consequences that reinforce or punish emitted responses. For a consequence to alter a certain behavior, the consequence must be *contingent* on the behavior, meaning it is delivered only after the target behavior is performed, and it is otherwise not available. *Reinforcement* is a contingent event that strengthens the response that precedes it. See Table 32.8–1, which outlines reinforcement of target behaviors, schedules, and outcomes associated with each subtype.

Several behavioral interventions exist to assist patients in the acquisition of new behaviors. *Shaping* is a process of reinforcing (rewarding) successive approximations that come increasingly close to a desired behavioral goal. Intermediary behaviors may be reinforced because they resemble the desired behavior or include a component of the desired behavior. BA is a good example; longer-term goals like running a 5k, can include reinforcement for successive approximations along a trajectory, such as researching training schedules, signing up for the race, walking increasing distances, and eventually running increasing distances consistent with the chosen schedule. *Chaining* refers to the series of specific, small behaviors or steps that when performed in sequence, form a chain. One response produces the conditions for the next response, and so forth, and at the end of the chain, the final behavior is reinforced. *Forward chaining* consists of teaching the sequence of behaviors beginning with the initial step in the chain and reinforcing along the series. In *backward chaining,* the final response is taught first. Behavior chain analysis is particularly useful for exploring maladaptive behaviors (e.g., self-harm) as well as behaviors not met (e.g., assigned homework for a therapy session), as it incorporates processing the connections between series of behaviors as well as negative reinforcement for target behaviors.

Table 32.8–1.
Operant Theory Principles

BEHAVIOR

Action	Increases	Decreases
Given	*Positive reinforcement*	*Positive punishment*
Taken away	*Negative reinforcement*	*Negative punishment*

SCHEDULES

Type	Application	Utility
Continuous reinforcement	Applied with each target behavior	Establishing new behavior; increasing low-level response
Intermittent reinforcement	Applied with a subset of target behaviors	Maintaining behaviors over time; reducing extinction
Fixed interval	Applied at consistent time points or number of behaviors	
Variable interval	Applied randomly	

Contingency Management. Contingency management is defined as the general application of operant principles in the process of behavior change and relies on consistent administration of reinforcement to strengthen target behavior (and ultimately reduce or extinguish the undesired response). This has been efficacious in the treatment of substance use disorders (SUDs) and smoking cessation among teens and adults, including those with comorbid mental illness. Vouchers or gift cards are received for meeting behavioral expectations (e.g., negative drug screens, treatment attendance) at fixed or variable intervals/quantities; this is also implemented in the context of larger treatment programs including CBT or other relapse prevention skills. It is challenging to implement in practice, as stigma surrounding substance use and limited interpretations of “reward” breed controversy, and it is difficult to identify funding for vouchers.

Token economy programs are a formalized method of contingency management often used in adult and child inpatient settings, day treatment programs, and classrooms. Token economy systems clearly spell out a series of behaviors that are expected and the contingencies for compliance/noncompliance with these behavioral goals. Achievement of behavioral goals leads to the acquisition of tangible rewards (snacks, poker chips that can be redeemed for prizes) or privileges (time off the ward, home visit privileges). Noncompliance does not result in privileges or rewards. Some programs include a provision called *response cost* (a form of punishment), which is the removal or withdrawal of a reward following inappropriate or problem behavior (e.g., fining a child for swearing). The combination of token reinforcement and response cost may be more effective than either alone. Token economies also are most effective when administered in a patient’s natural environment and when reinforcers eventually are shifted from tangible rewards (e.g., food and points) to social and more natural reinforcers (e.g., praise, feeling good about oneself). Shaping procedures can be used to facilitate this transition.

A number of behavioral procedures incorporate contingency management strategies as part of an overall intervention. For example, *parent–child interaction therapy* (PCIT) is a structured program that teaches parents child-directed interaction (CDI; e.g., praising, reflecting, imitating, describing, being enthusiastic, and providing contingent attention) and Parent-Directed Interaction (PDI; e.g., giving good commands, praising compliance, using time-out in a chair for noncompliance, and establishing structured house rules). PCIT is a short-term intervention that has been used successfully with a range of behavioral problems, including oppositional behavior, aggression, eating behaviors, and homework completion.

Aversion-Based Approaches. Like exposure therapies, aversion-based approaches consist of interventions based on operant principles or counterconditioning strategies. Here, punishment is defined functionally as the reduction in the future probability of a response as a result of the immediate delivery of a noxious stimulus for that response. Punishing stimuli include loud noise, aversive tastes, mild electrical shock, or chemical agents, used as a last resort when other behavioral methods such as time-out or response cost are ineffective. Punishment is also used when the negative behavior is very dangerous to the person (e.g., self-mutilative behaviors). To be effective, punishment must be of high intensity (only enough to disrupt the behavior), immediate, and continuous (at least initially). Less aversive stimuli, such as a drop of lemon juice applied to the tongue, are tested prior to escalating intensity. Punishment may produce undesirable effects such as negative emotional reactions, avoidance, and aggression, in addition to the increased response of problem behavior when intensity is insufficient. Numerous ethical issues are also considered, and consultation and review are necessary to ensure overall patient well-being, particularly for children and those with developmental disabilities.

Other aversive procedures, such as *aversion therapy,* are based on a classical conditioning paradigm. Used to treat sexual deviations and SUD, an aversive physical and/or emotional state is paired with cues that elicit abusive behavior or sexually deviant behavior, resulting in an aversion to the sensory aspects associated with the behavior and ideally, suppression of the behavior. *Chemical aversion* may be administered via the olfactory or gustatory systems. *Olfactory aversion* is the pairing of extremely noxious but harmless odors (such as ammonia) with either sexual fantasies or sexual behaviors, typically in the setting of the deviant sexual behavior. For SUD, *emetic aversion* (a chemical that induces vomiting; such as disulfiram or Antabuse) has also been used, as implementing immediate consequences may help to promote sobriety. Similarly, electrical stimulation to the forearm (*faradic aversion*) and *verbal aversion* (covert sensitization: presentation of an imaginal scene involving deviant behavior immediately followed by a scene that produces nausea) have been used. Recent evidence supports the superior effectiveness of disulfiram (relative to oral naltrexone or acamprosate) in treating alcohol use disorder, such that it can serve as adjunct to a comprehensive treatment program rather than as aversive therapy alone.

Skills Training. Skills training approaches have a long history of use in behavioral therapy. From a behavioral perspective, lack of skill may play a role in the onset and maintenance of various types of psychopathology. For example, people with social phobia or schizophrenia may have limited functionality related to social isolation, whereas those with personality disorders have difficulty maintaining relationships with other people. In these and other instances, *social skills training (SST)* may need to be an integral part of the treatment program. SST is a broad-based program designed to teach a wide variety of social skills applicable to individual needs. Social Environment Awareness involves teaching the nuances of when, where, and why to initiate and terminate interpersonal interactions. Interpersonal Skill Enhancement utilizes education about the mechanics of successful social encounters (e.g., initiating, maintaining, and ending conversations; establishing and maintaining friendships; and assertiveness). Techniques used for SST include instruction in the skill, modeling the skill, behavioral rehearsal, corrective feedback, and positive reinforcement. Role-playing assesses baseline skill and facilitates behavioral rehearsal of new skills which research identifies as the critical element in social skill acquisition.

“Second-Wave” Behavior Therapy Treatment Strategies

In the late 1960s, neo-behaviorists began to question operant and respondent principles alone and turn to potential mediators introduced in cognitive theory. The “cognitive revolution” emerged in the context of social psychology, where pioneers such as Fritz Heider, Harold Kelley, and Bernard Weiner were beginning to research and understand the complex ways in which individuals interact and think about one another’s behaviors.

Cognitive Therapy. A founding co-father of cognitive therapy is Aaron Beck, and the fundamental premise of cognitive therapy is that abnormal behavior is the result of distorted cognitive processes. Beck frequently referred to the *cognitive triad* in depression: a negative view of the self, the world, and the future. These negative beliefs are often called *cognitive distortions,* or patterns in thinking that affect individuals' mood and behavior. The focus of intervention is on increasing awareness of such *automatic thoughts* and engaging in the process of *cognitive restructuring* or reframing, which examines the validity and/or utility of automatic thoughts and modifying content when in the face of inconsistent evidence from their life and daily experiences. The other founding co-father of cognitive therapy was Albert Ellis, who developed rational emotive behavior therapy (REBT), previously called rational emotive therapy and RET. One of the fundamental premises of REBT is that the major cause of human suffering is that people are largely distressed by events and adversities because of how they construct their views of reality through their language, evaluative beliefs, meanings, and philosophies about the world, themselves, and others. In REBT, the patient begins to understand how beliefs about adversity and consequences maintain distress and instead engages in adaptive perspectives.

Several other cognitive therapies subsequently have been developed, including Patricia Resick's *cognitive processing therapy* (1993) and Jeffrey Young's *schema therapy* (1990), interventions demonstrated to be effective for problems such as depression, suicidal ideation, anxiety, trauma and PTSD, SUD, eating disorders, personality disorders, and marital problems. Although beyond the scope of the chapter to describe in detail, the important point is that all cognitive therapies, to a greater or lesser degree, include behavior-based assessment and intervention methods.

Problem-Solving Therapy. Psychological difficulties can result from ineffective methods by which to approach life's problems. Problem-solving therapy (PST) teaches the necessary skills to adapt to challenges and discover effective solutions and is effective for psychiatric disorders and stress associated with chronic medical disorders. The specific goals of PST are to increase understanding of the connection between current mood symptoms and everyday problems, increase the ability to clearly and accurately define current problems, teach patients a specific problem-solving method to allow for a more structured skill set to address life problems, and develop an increased sense of self-efficacy. During the course of PST, patients increase agency in modifying their problem list and tackling solutions.

Self-Control Approaches. Common self-control strategies include physically restraining oneself, chaining the stimuli conditions, depriving or satiating oneself, using mild aversive events, using drugs or alcohol, administering self-reinforcement or self-punishment, and doing something else to avoid aversive consequences. Self-control therapies are efficacious for changing maladaptive and excessive behaviors that contribute to an unhealthy lifestyle, such as overeating and smoking.

Self-control therapies generally involve three basic processes. The first stage is self-monitoring of behaviors (described earlier, e.g., keeping a food diary), self-evaluation (wherein data is aggregated and compared to a standard or normative information, such as number of calories per day), and self-reinforcement (setting contingencies based on performance). Self-control programs are most likely to work when reinforcement is directed at a specific behavior (e.g., reward for engaging in exercise) and not effects of behavior (e.g., weight loss). Another type of self-control strategy is called *stimulus control.* Adjusting stimulus presence and cueing can reduce problematic behavior, as seen in sleep hygiene techniques for insomnia.

Habit Reversal Training. Habit reversal training (HRT) is a behaviorally based treatment approach designed to reduce repetitive behaviors (habits) such as hair pulling (trichotillomania), nail-biting, and tics. From a behavioral perspective, these habits emerge through classical conditioning and operant reinforcement to reduce stress. HRT, as developed by Nathan Azrin and his colleagues, was designed to break the conditioned habit cycles using a combination of procedures. In *awareness training*, self-monitoring increases patient awareness of internal and external precipitants of the habit, considering a pattern of behaviors preceding the motoric responses, internal states, and consequences. *Habit control motivation* explores challenges and motivations associated with the behavior. *Competing response training* is the core of HRT. This process involves identifying a motoric behavior that is incompatible with the habit and engaging alternatively engaging in this in the context of associated precipitating behaviors or internal cues. For example, fist clenching (or clenching a book, or firmly grasping a steering wheel) is considered incompatible with body-focused repetitive behaviors.

Relaxation training discussed earlier, is also used to help break the conditioned habit cycles, particularly when stress or another negative mood is a precipitant for the target behavior. Finally, reinforcement from others or through contingency management or self-control mechanisms can help to facilitate and maintain treatment gains.

“Third-Wave” Behavior Therapy Treatment Strategies

According to Hayes, “Grounded in an empirical, principle-focused approach, the third wave of behavioral and cognitive therapy is particularly sensitive to the context and functions of psychological phenomena, not just their form, and thus tends to emphasize contextual and experiential change strategies in addition to more direct and didactic ones. These treatments tend to seek the construction of broad, flexible, and effective repertoires over an eliminative approach to narrowly defined problems and to emphasize the relevance of the issues they examine for clinicians as well as clients. The third wave reformulates and synthesizes previous generations of behavioral and cognitive therapy and carries them forward into questions, issues, and domains previously addressed by other traditions, in hopes of improving both understanding and outcomes.”

Behavioral Activation. Based on the premise of behavioral theory, individuals with depression get less positive reinforcement via anhedonia (finding limited or no enjoyment in activities) and behavioral withdrawal (limited engagement in enjoyable activities). Positive or negative reinforcement from others or exposure to aversive or unpleasant life experiences may also serve to maintain depression.

BA emphasizes structured attempts to increase overt behaviors that bring patients into contact with reinforcing environmental contingencies and corresponding improvements in thoughts, mood, and quality of life. Within the BA model, the process of increasing response-contingent reinforcement follows the basic behavioral principles of extinction, shaping, fading, and in vivo exposure. Initial sessions involve assessing the function of depressed behavior, establishing patient rapport, motivational exercises focusing on the pros and cons of behavioral change, depression psychoeducation, and introduction of the treatment rationale. Once efforts are made to reduce reinforcement for depressed behavior, a systematic approach for increasing healthy behavior is initiated by increasing the value of reinforcers for such behavior and devaluing reinforcers for depressed behavior.

BA has been effectively used with patients in community mental health centers, inpatient psychiatric facilities, college psychology clinics, and primary care settings, even when administered by previously untrained mental health nurses. In addition to major depression experienced by younger and older adults, BA has been shown to be effective for individuals with grief and bereavement, SUD, anxiety disorders such as PTSD and GAD, schizophrenia, and depressed patients with cancer, obesity, and HIV/AIDS. Data suggest BA leads to greater improvement than cognitive therapy for patients with severe depression, with equivalent outcomes to antidepressant medication.

Acceptance and Commitment Therapy. Originally referred to as comprehensive distancing, acceptance and commitment therapy (ACT) was developed in the 1990s by Steven Hayes and colleagues. ACT encompasses acceptance and mindfulness strategies together with behavior change strategies to increase psychological flexibility toward the goal of adapting cognitively and behaviorally to stressors and one's reactions to these. Tenets include contacting the present moment fully as a conscious human being and, based on what the situation affords, changing or persisting in behavior in the service of chosen values. The core premise of ACT is that much of human (psychological) suffering is a consequence of *experiential avoidance,* or the unwillingness to experience negative feelings, thoughts, and sensations. Through metaphor, paradox, and experiential exercises, clients learn how to make healthy contact with thoughts, feelings, memories, and physical sensations that have been feared and avoided. Clients gain the skills to recontextualize and accept these private events, develop greater clarity about personal values, and commit to needed behavior change. ACT views the core of many problems to be due to the concepts represented in the acronym, *FEAR*: *F*usion with your thoughts, *E*valuation of experience, *A*voidance of your experience, and *R*eason-giving for your behavior. According to ACT, the healthy alternative is to *ACT*: *A*ccept your reactions and be present focused, *C*hoose a valued direction, and *T*ake action.

Mindfulness-Based Therapy. Mindfulness-based treatments are a relatively new variation in behavioral (and cognitive) interventions that integrate various techniques based on Eastern philosophies of contemplation and meditation. Mindfulness-based therapies (MBTs), such as mindfulness-based stress reduction (MBSR) and mindfulness-based cognitive therapy (MBCT), show promise as effective interventions for anxiety, depression, and other psychological problems. The fundamental principles of MBT involve attention regulation, openness to present experience, curiosity, acceptance of the "here-and-now," and nonjudgmental awareness of thoughts, emotions, sensations, and the environment. The efficacy of MBT in treating emotional disorders is believed to be a function of an increased ability to manage life stressors, enhanced concentration and mindfulness, and reduced rumination. In MBSR, the primary strategies include learning mindful eating (e.g., the raisin exercise), mindfulness meditation, body (scanning) awareness, mindful walking, urge surfing, and yoga. Originally developed in the mid-1990s as a treatment to prevent relapse into depression, MBCT integrates fundamental theories and applications of MBSR with those of more traditional CBT. Unlike CBT, however, MBCT does not emphasize changing thoughts themselves; rather, the goal is to develop an awareness of and relationship to thoughts. Aspects of CBT included in MBCT are primarily those designed to facilitate "decentered" views, such as "thoughts are not facts" and "I am not my thoughts." Although interventions such as MBSR and MBCT are predominantly administered over 8 (3-hour) weekly sessions followed by a daylong retreat, some data support abbreviated formats of these protocols as well as the use of these strategies as part of a comprehensive treatment approach for specific presentations (e.g., anxiety, adjustment difficulty, or pain).

Dialectical Behavior Therapy. Developed by Linehan in the early 1990s, dialectical behavior therapy (DBT) was originally developed as an intervention for chronically suicidal individuals. DBT blends behavioral and crisis intervention theories with an emphasis on acceptance and tolerance based on Western contemplative and Eastern meditation practices. The theoretical framework emphasizes exchange and negotiation in the therapeutic dyad, between the rational and the emotional, and between acceptance and change (hence the term "dialectical"). Treatment targets are agreed upon, with self-harm taking priority. The learning of new skills is a core component, including mindfulness (e.g., nonjudgmental awareness of one's own feelings and actions), interpersonal effectiveness (e.g., assertiveness and social skills), coping adaptively with distress and crises, and identifying and regulating emotional reactions. DBT can be implemented in individual and/or group formats; in particular, DBT skills groups for adolescents and adults can provide psychoeducation, skills training, and feedback normalizing the intensity of emotional experiences.

Cognitive-Behavioral Analysis System of Psychotherapy. Developed by James McCullough in the late 1990s, cognitive-behavioral analysis system of psychotherapy (CBASP) is an integrative therapy for chronically depressed adults that combines components of cognitive, behavioral, interpersonal, and psychodynamic therapies. According to this model, those with chronic depression experience disconnectedness from their environment and therefore have decreased access to important feedback on problematic interpersonal patterns and relationships. The therapeutic relationship is actively used to help patients generate empathic behavior, identify and change interpersonal patterns related to depression, and heal interpersonal trauma. CBASP consists of three techniques: (1) *situational analysis,* a problem-solving technique designed to help the patient realize the consequences of their behavior on others and modify it; (2) *interpersonal discrimination exercises,* an examination of past traumatic experiences with others and differentiation of those from healthier relationships; and (3) *behavioral skill training/rehearsal,* such as assertiveness training, to further help depressed individuals modify maladaptive behavior. McCullough also strongly recommends that all patients beginning CBASP begin a regime of antidepressant medication.

Functional Analytic Psychotherapy. Functional analytic psychotherapy (FAP) is a treatment based on a contextual behavior-analytic understanding of the therapeutic relationship as an agent of change. Although much of behavior analysis is centered on environmental manipulation, FAP presumes that significant change occurs in the context of the therapeutic relationship during the traditional weekly psychotherapy session, such that the focus is on clinically relevant behaviors (CRBs) that occur in the therapeutic relationship rather than processing life problems. In FAP, the psychologist or therapist seeks to reduce (functionally punish or extinguish) the client’s dysfunctional behavior occurring in session (CRB1) while increasing the probability of (functionally reinforce) more appropriate alternative behavior occurring in session (CRB2). This is done by using naturalistic feedback and reactions in session. The psychologist or therapist also becomes an important mediator of interpersonal reinforcement. Awareness of CRBs and contingent responding to clinical problems and improvements that occur in therapy support change that can be maintained outside the therapy setting. Kohlenberg, Tsai, and colleagues recommend the following: create trust and safety, move the conversation from content to in-the-moment process, evoke and naturally reinforce client target behaviors related to authentic expression, instill hope, be aware of clinical impact, and promote generalization of in-session gains to daily life.

CLINICAL ISSUES

Developmental Considerations

Effective behavioral treatments have been developed for use across the full developmental spectrum, including children. These approaches are particularly valuable for patients whose verbal skills are not appropriate for more traditional therapeutic approaches. However, it is important to select and implement treatment techniques with careful attention to cognitive, social, and other developmental considerations. In children, assessment and treatment may utilize more visual representations (e.g., rating scales with faces or other pictures to describe mood instead of words). Further, children have not yet developed the cognitive skills to participate effectively in interventions that have a large cognitive component. For example, prior to midchildhood, few children have metacognitive skills (i.e., the ability to think about thinking). Therefore, interventions with children often de-emphasize cognitive strategies in favor of more behavioral interventions. Finally, treating children also involves consistent feedback with a broader treatment team, including parents, teachers, and other caretakers, and may occur exclusively in schools to increase access to services. A broad assessment of ability in younger populations may also yield individualized treatment recommendations, as seen in applied behavior analysis (ABA) for children with autism spectrum disorders (ASDs). While positive reinforcement and patterns of antecedents, behaviors, and consequences are core components, treatment is tailored according to individual factors and skill deficiencies.

Among older adults, consideration of sensory limitations (e.g., written documents prepared with larger fonts) and any medical issues are relevant. In addition, there are expected cognitive shifts in normal aging, and neuropsychological testing can be especially useful in identifying cognitive impairment that may impact treatment choice and engagement (e.g., in dementia, cognitive approaches are not appropriate). In such cases, exploration of relative strengths may help to elucidate potential avenues for therapy, whereas noted weaknesses may indicate needed resources (e.g., the inclusion of collaterals or caregivers who can assist with the assessment and treatment process). Indeed, case studies have also demonstrated that older adults with dementia and anxiety are able to learn coping skills such as relaxation, self-talk, and BA with the help of a psychologist or therapist and a collateral or caregiver. For these patients, different learning strategies are used to teach skills, including increased repetition and practice, the use of more cueing procedures to facilitate memory, and spaced retrieval, a procedure based on classical conditioning that repeatedly pairs questions (e.g., What can you do when you feel nervous?) and answers (e.g., I can breathe deeply) over increased intervals to facilitate memory. Reminiscence can also serve to optimize mood, particularly when utilized in a BA context.

Cultural Considerations

Cultural competence is an integral part of ethical, patient-centered care and is included as one of the American Psychological Association (APA)'s core competencies for psychologists. Yet attention toward this tenet has historically been limited with respect to empirical review according to race/ethnicity and thus generalizability of efficacious findings to marginalized groups. In general, data supports improved effectiveness when empirically supported therapies, including CBT, are culturally adapted to better represent language and nuanced meaning in treatment protocols and subsequent delivery. We must do better in the inclusion of underrepresented groups and analysis according to individual factors like race. A recent review noted less than half of RCTs of behavioral treatment for weight loss reported outcomes by race or considered race in design, and overall, Black patients lost less weight relative to White patients. Indeed, while differences according to race and ethnicity may reflect access to and request for mental health services, the paucity of individualized intervention approaches is a fault of our community that may maintain these disparities. An ideal approach to address this clinical need is FAP, which allows a safe, nonjudgmental environment to explore interpersonal components of minority stress (e.g., gender and sexual minority patients) and contingencies associated with vulnerability with others.

Implementation Variables

Behavioral treatments typically are of relatively brief duration, although the total length of treatment depends on the nature of the target problem(s) and the maintaining variables. The use of homework assignments in behavioral treatments, in effect, extends the actual duration of treatment. Sometimes, sequential phases of treatment are required; for example, the ideal approach for SUD treatment is longitudinal and multidisciplinary to account for the chronically relapsing nature of the disease as well as the varying presentations according to severity. The initial intensive phase of treatment may require physician- or psychiatrist-supervised detoxification to safely manage withdrawal symptoms while coincidentally also limiting exposure to environmental cues. Step-down treatment may then incorporate relapse-prevention strategies to maintain sobriety and engagement in alternative coping behaviors. A variety of self-help treatments not involving traditional therapy have also derived from behavioral approaches (e.g., bibliotherapy and internet- or phone-based treatments).

Strengths of a Behavioral Approach

The strong foundation in empiricism and the continued reliance on efficacy to drive treatment selection is a strength of behavior therapy and minimizes the potential for harm. The scientific base for behavioral treatments is larger than for any other psychosocial approach, and behavioral treatments are the treatment of choice for many psychiatric disorders, including depression and anxiety. Specifying targets for change and evaluating outcomes in an objective fashion helps ensure that treatment is appropriately administered (i.e., with regard to content and duration) as well as suitable for the problem. As such, accountability is high; it should be relatively clear to the clinician and the patient (and insurance companies) when treatment has the desired effects. In addition, the collaborative nature of the interventions allows for active patient engagement and flexibility in approach to best meet the patient's needs across the treatment continuum. Behavior therapy is also efficient and cost-effective. Many treatment techniques described above have an immediate effect (e.g., when a sufficiently powerful reinforcement is used to change behavior, when intensive exposure is used to treat anxiety), and the course of behavioral treatment can be a few sessions with much of the work of treatment occurring outside of therapy. Finally, behavioral approaches often match well with the requirements of managed care, which require ongoing documentation of identified problems and outcomes.

Limitations, Complications, and Contraindications

Behavior therapy is an action-oriented approach that requires a patient's readiness to change. Behavioral treatment, therefore, may not be appropriate for patients who perseverate on the historical origins of their problems without wanting to take action. Treatment approaches such as *motivational interviewing* have been designed to help move patients toward readiness to change such that behavioral treatments can be effective. However, reinforcement for not making a change may be more powerful than reinforcement for making a change (particularly in substance use). If personality factors underlie limited engagement in behavior therapy sessions, DBT approaches, centered on the balance of validation of distress and the expectation to do more, may be more appropriate. However, patients who are consistently uninterested in making real changes are not optimal candidates for behavior therapy, or for that matter, virtually any form of therapy.

At times, although behavioral treatments generally match well with managed care reimbursement practices, some forms of behavior therapy may be at odds with insurance-based practice and reimbursement availability. Intensive exposure treatments, longer and more frequent than the traditional 50-minute weekly session or home visits, may not be fully reimbursed. Finally, in many places, psychologists and therapists with appropriate expertise are simply not available. With the swift implementation of telehealth practices since the onset of the COVID-19 pandemic, virtual sessions have increased access to care at the expense of factors that may be central to behavior therapy (e.g., interpersonal interaction and environmental reinforcement).

ETHICAL ISSUES

Ethical considerations are inherent in this work, particularly in aversion and exposure treatments. Aversion-based interventions are used only as a last resort when the patient has not responded to other interventions and is at risk of harming themselves or someone else. These treatments involve informed consent and are reviewed by a full team of providers, including physicians evaluating medical risk. The provider should have sufficient training in a particular treatment and seek supervision or consultation as needed. In many cases, ethical and legal review boards also have input. And perhaps most importantly, these procedures should be used in conjunction with alternative techniques that promote other, more adaptive behaviors. With exposure, there is a risk of increased anxiety and distress in the context of inconsistent or poor treatment. The choice to use exposure should be based on the scientific literature, a thorough case conceptualization, and a collaborative decision between the psychologist or therapist and patient following thorough discussion and evaluation of the cost–benefit ratio. In addition, ensuring the patient does not have a medical condition that contraindicates increased arousal and monitoring that reductions in arousal are occurring as expected is important.

In summary, although less than 100 years old, behavioral interventions are now considered established, standard-of-care interventions. They have been applied to a range of psychiatric disorders, functional deficits, medical disease states, and a combination of these presentations across the developmental spectrum, including children, adolescents, adults, and mature adults. They are translatable across settings, including traditional inpatient, outpatient, and integrated medical settings. Finally, unlike some other schools of therapy, the success of behavior therapy is not limited by intellectual or disability status, having been successfully used for those with various developmental disorders and sensory limitations. Clearly, behavioral interventions are broad based, rooted in science, and applicable across contexts.

FURTHER READINGS

Azrin NH, Nunn RG. *Habit Control in a Day*. Simon & Schuster; 1977.

Baer RA. Mindfulness training as a clinical intervention: a conceptual and empirical review. *Clin Psychol Sci Pract*. 2003;10:125–143.

Bandura A. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall; 1986.

Barlow DH. *Clinical Handbook of Psychological Disorders*. 5th ed. Guilford; 2014.

Bernstein D, Borkovec TO. *Progressive Relaxation Training*. Research Press; 1979.

Chambless DL, Ollendick TH. Empirically supported psychological interventions: controversies and evidence. *Ann Rev Psychol*. 2001;52:685–716.

Dimidjian S, Hollon SD, Dobson KS, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol*. 2006;74:658–670.

Eysenck HJ. The effects of psychotherapy: an evaluation. *J Consult Psychol*. 1952;16:319–324.

Farha LC, Olfson E, Nasir M, et al. Pharmacological and behavioral treatment for trichotillomania: an updated systematic review with meta-analysis. *Depress Anxiety*. 2020;37:715–727.

Hayes SC. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies–Republished article. *Behav Ther*. 2016;47:869–885.

Hayes SC, Strosahl KD, Wilson KG. *Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change*. Guilford; 1999.

Hofmann SG, Sawyer AT, Fang A. The empirical status of the “new wave” of cognitive behavioral therapy. *Psychiatr Clin North Am*. 2010;33(3):701–710.

Hollon SD, Ponniah K. A review of empirically supported psychological therapies for mood disorders in adults. *Depress Anxiety*. 2010;27(10):891–932.

Hopko DR, Lejuez CW, Ruggiero KJ, Eifert GH. Contemporary behavioral activation treatments for depression: procedures, principles, progress. *Clin Psychol Rev*. 2003;23:699–717.

Jones MC. The elimination of children’s fears. *J Exp Psychol*. 1924;7:383–390.

Kabat-Zinn J. An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: theoretical considerations and preliminary results. *Gen Hosp Psychiatry*. 1982;4:33–47.

Kazdin AE. *Behavior Modification in Applied Settings*. 6th ed. Wadsworth/Thomson Learning; 2001.

Kohlenberg RJ, Tsai M. *Functional Analytic Psychotherapy*. Springer US; 1991:169–188.

Lewinsohn PM. A behavioral approach to depression. In: Friedman RJ, Katz MM, eds. *The Psychology of Depression: Contemporary Theory and Research*. Wiley; 1974:157–178.

Linehan MM. *Cognitive–Behavioral Treatment of Borderline Personality Disorder*. Guilford Press; 1993.

McCullough JP. *Treatment for Chronic Depression: Cognitive Behavioral Analysis System of Psychotherapy (CBASP)*. Guilford; 2000.

Nezu AM, Nezu CM, Perri MG. *Problem-Solving Therapy for Depression: Theory, Research, and Clinical Guidelines*. Wiley; 1989.

Persons JB. Empiricism, mechanism, and the practice of cognitive-behavior therapy. *Behav Therapy*. 2005;36:107–118.

Sattler JM, Hoge RD. *Assessment of Children: Behavioral, Social, and Clinical Foundations*. Jerome Sattler Publications; 2006.

Spiegler MD, Guevremont DC. *Contemporary Behavior Therapy*. 4th ed. Wadsworth/Thomson Learning; 2003.

Summerfeldt LJ, Antony MM. Structured and semistructured diagnostic interviews. In: Antony MM, Barlow DH, eds. *Handbook of Assessment and Treatment Planning for*

Psychological Disorders. The Guilford Press; 2002.

Teasdale JD, Segal Z, Williams JM. How does cognitive therapy prevent depressive relapse and why should attentional control (mindfulness) training help? *Behav Res Therapy*. 1995;33:25–39.

Tsai M, Yoo D, Hardebeck EJ, Loudon MP, Kohlenberg RJ. Creating safe, evocative, attuned, and mutually vulnerable therapeutic beginnings: strategies from functional analytic psychotherapy. *Psychotherapy (Chic)*. 2019;56:55–61.

Watson JB, Rayner R. Conditioned emotional reactions. *J Exp Psychol*. 1920;3:1–14.

Wolpe J. *Psychotherapy by Reciprocal Inhibition*. Stanford University Press; 1958.

32.9 Cognitive Behavior Therapy

JUDITH S. BECK, PH.D., AND ROBERT HINDMAN, PH.D.

Cognitive behavior therapy (CBT) was developed by Aaron T. Beck, M.D., in the 1960s and 1970s. He and colleagues worldwide have continued to expand and refine the underlying theory and treatment ever since and Beck's work, emphasizing the central role of cognition in psychopathology, has led to a paradigm shift in the field. It is a comprehensive biopsychosocial system of psychotherapy, with an operationalized treatment based on an elaborated and empirically supported theory of psychopathology. CBT has become the most heavily researched and widely practiced form of psychotherapy in the world. It has been demonstrated in approximately 2,000 outcome trials to be effective for a wide range of psychiatric disorders. The benefits of CBT persist after the discontinuation of treatment, resulting in a lower relapse rate in comparison to those taking psychotropic medication. CBT has also been demonstrated to be effective for a range of psychological problems (such as stress, anger, procrastination, and relationship distress) and medical conditions with psychological components (such as tinnitus, chronic pain, and hypertension).

CBT has been adapted and research has demonstrated its effectiveness in a variety of settings, including outpatient clinics, psychiatric hospitals, primary care offices, schools, forensic settings, and community mental health centers. Its efficacy has been demonstrated for populations from young children through older adults, and for individuals who vary in gender, sexual orientation, socioeconomic status, and racial or ethnic background, and who come from a variety of cultures around the world. Studies also support its efficacy in individual, couples, family, and group treatment. A substantial number of studies now demonstrate its efficacy in computer-based programs and telehealth formats.

CBT is based on the cognitive model, which has been widely studied and has received substantial empirical support. The cognitive model posits that it is not situations that directly impact one's reaction (emotional, behavioral, and physiological). Rather it is one's *perception* of the situation, expressed in automatic thoughts, that is more closely associated with the reaction. One important component of CBT involves teaching patients to identify their thoughts that lead to emotional distress or dysfunctional behavior, to evaluate the accuracy and/or utility of their thinking, to effectively respond to these cognitions, and to reduce negative emotion and/or maladaptive actions. Figure 32.9–1 depicts the cognitive model.

Beck and colleagues developed a manual for CBT for depression to facilitate treatment fidelity in research studies (Beck et al., 1979). Manuals, however, often fail to allow clinicians to vary therapy as appropriate for the individual patient, and most clinicians find their treatment is significantly more effective when it is based on a cognitive formulation of the patient's disorder(s), an individualized cognitive conceptualization of the patient, and the patient's personal values and aspirations. A strong therapeutic alliance is essential for effective treatment and the relationship is collaborative. Both patients and clinicians are active in solving problems, evaluating cognitions, and modifying behavior. CBT is educative; clinicians teach patients how to use the techniques at home that they have employed in session. The clinician's stated goal is to help patients become their own therapists. They emphasize that the way patients get better is to apply what they have learned in session to make small changes in their thinking and behavior every day.

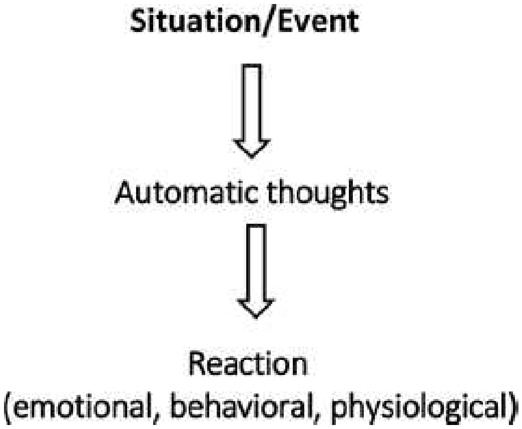

FIGURE 32.9–1. Cognitive model.

Originally designed as a short-term treatment for depression, CBT has evolved over the years. Patients with straightforward cases of anxiety and depression, who were relatively psychologically healthy before the onset of their disorder, often benefit from a short course of treatment, perhaps 6 to 12 sessions. At the other end of the spectrum are patients whose disorders are chronic or complex, or those for whom standard treatment needs to be varied due to comorbidities such as substance use disorders (SUDs) or personality disorders. They may need many more sessions over a significantly longer period of time. Those with recurrent episodes or with serious mental illnesses may require intermittent treatment throughout their lifetime. This latter group of patients often benefit from a modified version of CBT: Recovery-Oriented Cognitive Therapy, which emphasizes empowerment, purpose, and resilience, and focuses on helping patients operate in an adaptive mode where they are able to live in alignment with their values and aspirations instead of focusing on problems and symptoms.

HISTORY

Aaron T. Beck, who was Emeritus professor of psychiatry at the University of Pennsylvania, had trained and practiced as a psychoanalyst. A scientist at heart, he conducted a series of studies starting in the late 1950s to test psychoanalytic theories of depression. Much to his surprise, his research demonstrated that depression was not a result of retroflected hostility and did not support a need to suffer. Instead of viewing depression as a result of unconscious drives, wishes, and defenses, Beck proposed that depression was associated with highly negative, distorted, and dysfunctional beliefs and thoughts, the result of biased information processing. When he began to tailor treatment toward helping his patients identify and evaluate their biased thinking and change their behavior, they began to rapidly improve.

In his landmark monograph, Beck (1967) detailed the cognitive triad of depression: negatively biased cognitions about oneself, one's world (including other people), and the future. He described the role of schemas, hypothesized mental structures that organize information, and he outlined a treatment approach that emphasized testing the accuracy of patients' cognitions. Ten years later, the first randomized controlled trial of CBT for depression was published (Rush et al., 1977). Patients in the CBT group responded slightly better to treatment than patients in the psychopharmacological group and had fewer relapses after treatment. Later studies confirmed these findings and showed that CBT was effective for severe depression.

Throughout the decades, Beck and colleagues studied other disorders. They used clinical material to identify and specify key cognitions and behaviors characteristic of a number of psychiatric problems. They used existing rating scales or constructed and validated new instruments to measure relevant clinical variables, adapted the general cognitive model and treatment interventions for each disorder, and created treatment manuals for randomized controlled trials.

In developing CBT, Beck was influenced by Greek Stoic philosophers and theorists contemporary to the time, such as Adler, Alexander, Sullivan, and Horney, as well as Kelly, Lazarus, and Arnold. Behaviorists and cognitive behaviorists such as Ellis, Meichenbaum, Mahoney, Lewinsohn, and Bandura were also influential in the development of his theories and treatment.

THEORETICAL ISSUES

Most automatic thoughts are verbal in nature, but some are in the form of mental images.

As Kevin was thinking about looking for job opportunities on the internet, he pictured telling a potential interviewer how long he had been out of work. He imagined the interviewer looking scornfully at him, shaking his head.

Core Beliefs

Individuals display characteristic, recurrent themes in their thoughts which are associated with a deeper level of cognition, their core beliefs, representing the content of cognitive schemas. Core beliefs are one's most central ideas about the self, other people, and the world. These beliefs are generally identified by asking patients for the *meaning* of their thoughts. Core beliefs develop during childhood. The interaction between one's genetic predisposition (toward certain personality traits and styles of thinking) and their experiences affects the way they interpret their experiences.

People without significant psychological difficulties tend to view themselves throughout their lives as reasonably effective, able to protect themselves, likable, and worthwhile. If they develop an acute disorder such as depression, however, previously latent negative schemas may become activated and dysfunctional beliefs become manifest. Patients begin to view themselves, their worlds, and/or other people through a negative lens, interpreting their experience in a highly negative way, and may then overgeneralize the meaning of a situation. They may see themselves as ineffective, vulnerable, inferior, unlikeable, and/or worthless. They may view others as potentially hurtful, rejecting, or superior. They may see their world as unpredictable or unsafe. During times of psychological distress, these individuals tend to regard their negative core beliefs as fixed, absolute truths, which make these deeper-level ideas less amenable to change than automatic thoughts.

Throughout his school years, Kevin performed at a level slightly below average but more or less at the level of most of his friends. He did not value getting good grades and was more interested in developing his above-average athletic skills. He felt competent enough academically and saw himself as a success athletically. His parents, however, had unrealistic academic expectations for him and criticized him when he did not perform as well as they desired (which was much of the time). Accordingly, from time to time, especially when his parents were harsh, he viewed himself as a failure. Fortunately, these episodes were generally fleeting, and his view of himself as basically adequate and competent predominated. It was not until Kevin was fired from his job and could not quickly obtain another one that he began to view himself as a failure. The dominance of this negative core belief contributed to the onset and maintenance of his depression.

Coping Strategies

The activation of a negative core belief typically leads patients to engage in maladaptive coping behaviors as an attempt to decrease or avoid distress.

As Kevin began to view himself as a failure, he started to avoid tasks that he assessed were difficult, at which he could fail. He gave up job hunting after a few weeks and spent a considerable amount of time sleeping, watching television, or playing video games. He began to withdraw from his wife, family, neighbors, and friends, believing that they all saw him as a failure, too.

Intermediate Beliefs

Intermediate beliefs are ideas at a deeper level than automatic thoughts but are more malleable than core beliefs. They consist of generalized rules, attitudes, and assumptions, and can be a link between patients' deeper core beliefs and their coping strategies. Patients generally believe that if they engage in certain coping strategies, their negative core beliefs may not be evident; however, if they do not, they believe these beliefs will likely be shown to be true.

Once Kevin became depressed, several key assumptions became activated. “If I take on challenges, I’ll fail, but if I avoid difficult tasks, I’ll be okay (at least for the moment).” He also held the attitudes that “It’s terrible to fail” and “Failing means I’m a failure as a person.” His related rules, which often took the form of “should” statements, included “I should do great at everything I do.”

Schemas and Modes

Schemas are responsible for processing, storing, and retrieving different types of information, including perceptions, goals, expectations, memories, fantasies, and prior learning. They provide meaning and, when activated, affect the way individuals process their experiences. Their function is to aid the individual in selecting the most relevant information from the practically infinite amount of information in the environment. Schemas may be cognitive, behavioral, motivational, or affective. They range from adaptive to dysfunctional and vary in complexity, strength, flexibility, permeability, level of abstraction, and valence.

When Kevin is depressed and his maladaptive depressive schemas are activated, he is likely to process information in a distorted manner that confirms his core beliefs. For example, he views his difficulty getting out of bed in the morning as a sign that he is a failure, even after receiving psychoeducation about the effect depression has on patients’ thinking, motivation, and behavior. He fails to register the tasks he completes at home or the ways he is helpful to his wife even in the face of severe depression. He discounts his wife’s positive comments, believing that she is “just trying to be nice.”

The diathesis-stress model (Beck, 1967) posits that schemas are triggered when individuals experience events relevant to their underlying vulnerability.

Kevin had had negative experiences from time to time throughout adulthood, such as the death of his best friend, arguments with coworkers, and conflict with a neighbor. These incidents, however, did not activate his depressive schemas as they did not match his underlying vulnerability toward interpreting a deficit in performance as signifying failure. His depressive mode was activated only after he was fired from his job for “negligence and incompetence” and was unable to secure a new job. These events matched stimuli from his “failure” schema, which led to its activation.

Modes are a complex organization of multiple schemas related to individuals’ self-evaluations, rules, memories, and expectancies. Primal modes are related to survival; constructive modes allow the individual to increase resources. Modes help explain the complex symptoms that patients experience and their global, systematic bias. Each psychiatric disorder is characterized by a particular mode. One constellation of schemas, for example, characterizes a depressive mode, while a different constellation is involved in an anxious mode.

The Generic Cognitive Model

In recent years, Beck has described an expansion of the cognitive model. This “generic cognitive model” explains how cognitive schemas contribute to the development and maintenance of psychopathology. Neurophysiologic hyperactivity is genetically influenced and may, in turn, negatively bias individuals’ attention and memories, resulting in dysfunctional cognitive schemas. Beck also describes the nature of individuals’ functioning along a continuum from adaptive to maladaptive, the process by which schemas are activated, the nature of dual processing (through the automatic and then reflective processing systems), and protoschemas (which quickly but crudely detect and interpret critical stimuli), along with an expanded theory of modes and common processes (such as rumination, worry, attentional fixation) across psychopathologic states.

Strength-Based Cognitive Conceptualization

Increasingly, the field has recognized the importance of adding a strength-based conceptualization and focus on understanding and treating patients. CBT clinicians identify patients' historic *adaptive* cognitions and coping strategies, often by questioning them about the best period in their lives and focusing on their achievements, strengths, positive qualities, and resources. During these best periods, individuals often had core beliefs such as "I'm reasonably competent/safe/likeable/worthwhile." They usually had relatively benign and realistic beliefs about people: "Many people are trustworthy or neutral and will not intentionally harm me" and their worlds: "The world is reasonably safe and predictable." Their intermediate beliefs tend to be functional ("If I try, I can succeed, at least to some degree"; "If I interact with most people, they'll accept me.") When they face challenges, they tend to persevere because of their adaptive automatic thoughts. ("I can probably solve this problem, but if I can't, I can ask for help.")

In contrast to a more traditional, problem-based treatment, therapists ask clients for their goals for each session. They help them draw realistically positive conclusions about the positive experiences they had in the previous week. They use standard techniques to resolve obstacles that could interfere with their ability to take steps in alignment with their values and aspirations.

Applications Across Specific Disorders

According to the cognitive model, psychiatric disorders are characterized by an interaction among triggers, cognitions, behaviors, emotions, and common psychological processes. Each psychiatric disorder is characterized by a specific set of beliefs and behavioral strategies. The following is an overview of the cognitive theories for several disorders.

Depression. The maladaptive cognitions specific to depression are referred to as the cognitive triad: negative beliefs about oneself, one's life experience (and the world in general including other people), and one's future (Beck, 1967). Depressed patients often predict that they will inevitably face only unending problems and pain; these predictions affect both their mood and their behavior. They tend to be excessively self-focused. They are likely to disproportionately allocate attention to their negative thoughts, feelings of sadness, and physiologic symptoms. Depressive rumination typically results in an exacerbation of symptoms and continued inactivity. As the depressive mode continues to be active, individuals' sense of hopelessness about the future may increase and lead to increased suicidal ideation when they believe suicide is the only solution to their problems.

Bipolar Disorder. The schemas of bipolar patients are activated when biologic and psychological processes interact with external and internal triggers, leading to depression, mania, or mixed-mood episodes. In a depressed state, these patients display the same kinds of cognitive biases and dysfunctional beliefs as patients with unipolar depression, especially negative cognitions about the self, the world and other people, and the future. In a manic state, patients with bipolar disorder tend to view themselves in an inflated, expansive way. Hyperpositive core beliefs ("I am powerful"; "I am invulnerable"; "I am amazing") predominate. Their behavior changes are consonant with the changes in their thinking. They become impulsive, minimize potential threats, and take risks. Typical conditional assumptions include: "If something seems like a fantastic idea, it must be"; "I should follow my impulses and take action immediately"; "If I'm feeling great, I don't have to take meds or be careful about my daily habits." In a euthymic mood, patients with bipolar disorder may think, "I have to drive myself hard to make up for how little I accomplished when I was depressed." This belief may lead them to overextend and overstimulate themselves, particularly in reference to goal-oriented behavior; they may then sleep poorly and ultimately become hypomanic.

Anxiety Disorders. When patients have anxiety disorders, they tend to significantly overestimate the risk of a situation and greatly underestimate their ability to cope. Worry about a variety of events or activities is common, especially among individuals with generalized anxiety disorder (GAD), as are futile attempts to control the worry process.

Many clients with anxiety disorders have maladaptive core beliefs such as "I'm incapable and vulnerable," and "The world is a dangerous place." Individuals with social anxiety disorder (SAD) tend to see themselves as defective and unlikable and predict that others will reject them; they become hyperalert for signs of anxiety which they try to hide from others; paradoxically, they therefore sometimes behave in a way that others find odd.

Cognitive models for obsessive-compulsive disorder (OCD) propose that intrusive thoughts, images, or urges develop into obsessions when they are given excessive importance. Patients with OCD believe that their intrusive thoughts and images are uncontrollable, that they will be responsible if bad things happen, and that they cannot overcome the urges to engage in compulsive behaviors that they hope will neutralize the threat.

The automatic thoughts of patients with anxiety tend to center around potential future catastrophes. Their attention becomes fixed on signs of threat, danger, and/or uncertainty. When clients have panic disorder, for example, they predict that a bodily or mental sensation (or a small set of sensations) will precipitously increase and lead to a catastrophe.

Behaviorally, clients with anxiety disorders tend to engage in maladaptive "safety behaviors," coping strategies to ward off anxiety, such as avoidance of situations in which they feel vulnerable, reassurance seeking (from other people and the internet), and excessive checking. These strategies prevent individuals from being exposed to situations that would have demonstrated to them that anxiety, while uncomfortable, is not dangerous.

Anxiety and worry are often triggered by an intolerance of uncertainty, or a view of ambiguous or uncertain situations as upsetting and stressful. When worrying about uncertainties, individuals experience a negative problem orientation. They do not trust their ability to solve the problems that are the focus of worry, and they perceive problems as threats. They continue to predict unfortunate or even catastrophic outcomes and they continue to worry, hoping to find just the right solutions. Worry has also been proposed to be a form of cognitive avoidance.

Eating Disorders. Anorexia nervosa and bulimia nervosa have been extensively studied and a cognitive model has been developed for each disorder. However, Fairburn and colleagues (2003) proposed that the disorders share the same underlying psychopathology. Their transdiagnostic model postulates that individuals diagnosed with eating disorders evaluate their self-worth disproportionately both in terms of their eating habits, shape, or weight and in their ability to control these factors. Their preoccupation with how they see themselves may have been influenced by early life factors, such as parents or peers who placed too much emphasis on or were overly critical of their weight or shape.

The difference between the two disorders is hypothesized to be the result of the compensatory behavior used to cope with negative self-evaluations or other triggers related to eating, body image, or sense of control. Patients with anorexia predominantly cope by restricting food intake and/or engaging in excessive exercise, while individuals diagnosed with bulimia use a compensatory strategy such as purging when they perceive they have overeaten or engaged in binges. In contrast, patients with binge-eating disorder do not use compensatory strategies after a perceived binge.

Many patients with binge eating disorder or bulimia have difficulty with emotion regulation and turn to food as a means of eliminating negative affect. Although their mood temporarily improves while they are eating, negative affect returns more intensely afterward in the form of disgust, embarrassment, guilt, or depression, further strengthening negative beliefs about the self.

Substance Use Disorders. Individuals may begin to use substances to obtain pleasure; feeling “high” may improve their experience of activities. Or they may begin to use substances to decrease or eliminate unwanted experiences such as negative emotion or physical pain. Once substance use has started, several factors may contribute to becoming dependent on the substance, including beliefs such as, “If I use, I can endure the stress and frustration of life;” “I can escape or avoid distress.” Individuals for whom substance use is a problem also tend to have negative core beliefs. The activation of these beliefs leads to dysphoria; the individual focuses on the negative emotion and then has a permission-giving cognition which leads to the compensatory behavior of substance use to alleviate or cope with the distress. They may also use because of their belief, “I won’t enjoy myself if I don’t use.” As substance use becomes more frequent, external, and internal triggers become associated with use. These triggers also become associated with the experience of cravings and urges; they may believe “I have no choice; I have to give in to urges,” “Cravings are uncontrollable,” and “This craving will never go away unless I use.”

As substance use increases, individuals are likely to encounter financial, social, and/or medical problems about which they have negative cognitions, leading to negative mood states and subsequent substance use. Their focus of attention is excessively inward, as they monitor themselves for dysphoria. They generally minimize potential consequences or develop additional permission-giving beliefs. The cycle of substance use is maintained by the dysfunctional beliefs about the self, substance use, cravings and urges, permission to use, and negative mood states.

Schizophrenia. The negative symptoms of schizophrenia are mediated by dysfunctional beliefs: “I’m powerless”; “I’m defective”; “People will reject me”; “I’ll fail at anything I try.” Delusions are believed to develop, in part, due to information processing biases, cognitive errors, and impaired reality-testing. Individuals experiencing delusions tend to perceive themselves to be the center and central focus of events, while also interpreting their subjective experiences as having an external causation. A cognitive model of hallucinations describes factors that are precursors to hallucinations, lead individuals to fixate on them, and then maintain their occurrence. One predisposing factor is a low threshold for internal auditory and visual imagining that seems real or almost real. This low threshold interacts with activated cognitive schemas that produce cognitions related to the schema and not to reality. For instance, an individual with a “failure” schema may have an intrusive, automatic thought that is perceived as an auditory hallucination saying, “You can’t do anything right.” The hallucination is perceived as coming from an external source due to externalization bias, cognitive errors, and impaired reality testing. Hallucinations are maintained by dysfunctional beliefs such as “These voices are real.” Safety behaviors, such as avoidance, are also used to minimize the presence or impact of hallucinations.

Beliefs about positive symptoms, in addition to the symptoms themselves, influence the development and maintenance of negative symptoms of schizophrenia. Moreover, as positive symptoms are often triggered in social situations, individuals may develop an aversion to socialize and choose to reduce social contact. Overall activity level may also decrease as individuals become excessively focused on delusions and/or hallucinations and less focused on engaging in valued activities. Additional factors that influence the development and maintenance of negative symptoms are negative beliefs about one’s performance, low expectations for obtaining pleasure and success, and the perception of having limited resources for engaging in activities, for example, too little energy.

Personality Disorders. As with the previously discussed disorders, personality disorders develop from the interplay of genetic and epigenetic factors and the meaning individuals put to their early adverse or traumatic experiences. The schemas present in individuals with personality disorders resemble those present in other syndromes, but they tend to be more rigid and remain activated on a more or less continuous basis.

Each personality disorder is characterized by a specific set of beliefs, assumptions, and coping strategies. An individual with dependent personality disorder, for example, may have a belief of helplessness and cope by attaching to and depending on others. An individual with avoidant personality disorder usually has core beliefs of emotional vulnerability and unlovability and avoids not only social interactions but also thinking about situations that could lead to emotional distress. When individuals have comorbid personality disorders, they often hold multiple sets of dysfunctional beliefs that cross diagnostic lines.

TECHNIQUES

The goal of CBT is to help clients reach collaboratively set treatment goals, live a valued life, facilitate remission of their presenting problems and/or psychiatric disorders, and prevent or reduce relapse by teaching patients skills they can use for their lifetime, within a highly supportive, collaborative relationship. Both patients and clinicians are active participants, jointly resolving current issues, especially those that interfere with goal attainment, through a variety of strategies, including teaching patients to evaluate their cognitions and working toward needed behavioral change. As a result, patients' symptoms are reduced, their mood and day-to-day functioning improve, they are able to live in greater accordance with their values and aspirations, they are more resilient, and they have an increased sense of purpose, hope, and well-being. Many patients achieve treatment goals and remission without exploration of childhood issues. When longstanding highly dysfunctional beliefs acquired in childhood interfere with "standard" treatment focused on present-moment experiences, as may be the case when working with patients with personality disorders, clinicians may help patients restructure the meaning of early experiences and apply what they have learned to help them overcome current difficulties.

In this section, we describe how clinicians use CBT techniques in a standard 50-minute appointment with an individual outpatient. Many of the techniques, however, can be used in brief sessions.

Clinicians may adapt techniques from almost any psychotherapeutic modality, using them in the context of a cognitive conceptualization of the patient. In recent years, cognitive therapists have drawn on techniques from the range of evidence-based therapies, including dialectical behavior therapy, exposure therapy, cognitive processing therapy, acceptance and commitment therapy, and compassion-focused therapy. Techniques from positive psychology, psychodynamic psychotherapy, mindfulness-based therapies, interpersonal therapy, and others are adapted and used with some patients.

As part of the initial assessment, clinicians collect data to diagnose patients' difficulties. Treatment is guided by both a problem-focused and a strength-focused formulation. Clinicians begin to formulate a cognitive conceptualization of individual patients and their strengths and difficulties; they refine this conceptualization at each session, as they receive additional data. They hold their hypotheses as conditional, subject to confirmation by the patient.

Clinicians use this ongoing conceptualization to plan treatment across sessions and within sessions. An accurate conceptualization allows the clinician to appropriately select and adapt techniques. Clinicians also take into consideration the patient's goals, stage of treatment, degree of distress, personal characteristics and preferences, and values and aspirations. They consider the strength of the therapeutic alliance, the nature of the problem or goal under discussion, the amount of time remaining in the session, and other factors. When suggesting an intervention, they provide a rationale and elicit the patient's agreement. They often check on the intensity of patients' emotions before and after an intervention, to assess whether additional strategies might be needed. Two questions also help guide the selection of techniques: How can I help the patient feel better by the end of the session? How can I help the patient have a better week?

The strategies described below are somewhat artificially grouped, as many techniques fall into more than one category. For example, most behavioral techniques contain both cognitive and behavioral components and vice versa.

Therapeutic Relationship Techniques

Clinicians use good Rogerian counseling skills: warmth, support, accurate understanding and reflection, genuineness, empathy, realistic optimism, and compassion. They use positive reinforcement to praise and encourage change. They show genuine interest in their patients and may deliberately be more conversational at times, for example, when inquiring about positive experiences patients had in the previous week. They also make judicious use of self-disclosure. Clinicians are highly collaborative and seek patients' feedback. During sessions, they are attuned to patients' affect, and when they notice a negative shift, they ask patients how they are feeling and what they are thinking. When patients provide negative feedback, clinicians positively reinforce them and mutually develop a plan to address patients' concerns. Clinicians also routinely ask for feedback at the end of sessions: "What did you think of the session?… Was there anything that bothered you or anything you thought I didn't understand? . . . . . Is there anything you want to do differently next time?"

Some patients, especially those with personality disorders, bring dysfunctional beliefs about other people and relationships to treatment and may react to clinicians in dysfunctional ways. Clinicians work through therapeutic relationship problems, helping patients evaluate their beliefs about the clinician. They then help clients generalize what they have learned to other relationships. It is essential for clinicians to vary their interpersonal style according to the preferences and sensitivities of patients. Some patients prefer a close, warm relationship while others react poorly to direct expressions of empathy or self-disclosure and prefer a more clinical approach.

Finally, research has demonstrated that a reduction in symptoms leads to a better therapeutic alliance. Therefore, it is important to spend sufficient time developing a good working relationship with patients, but clinicians need to keep as their central focus the facilitation of cognitive, emotional, and behavioral change so patients are equipped to deal more effectively with their difficulties outside of the therapy session.

Structuring Techniques

CBT sessions tend to follow a standard structure, as it allows patients to gain the most from the appointment. The structure is implemented flexibly, though, and sometimes has to be modified, especially initially, if patients react negatively.

At the beginning of sessions, clinicians do a mood check. Research has shown that psychotherapy is more effective when including feedback on treatment progress. Clinicians ask for an update (of positive and negative events) since the previous session, and elicit patients' conclusions about their positive experiences, successes, or improvement to foster positive beliefs and build confidence. Patients' Action Plans (therapy homework) are reviewed; these self-help assignments have been demonstrated to enhance therapy outcomes. Clinicians also set an agenda for the current session ("What problems/issues do you most want my help in solving today?" or "What are your goals for today's session?"). Clinicians may contribute items to the agenda, too. Then they collaboratively prioritize agenda items. If later in the session patients bring up items that were not on this initial agenda, the clinician notes the change of topic, and they make a collaborative decision about whether to return to the initial item under discussion or to continue discussing the newly identified topic or difficulty.

In the middle of sessions, clinicians focus on the problem or goal that is most important to the patient. They ask patients to recall a specific recent instance in which they felt the most distressed about the problem or they may ask clients to identify a problem they expect to encounter in the coming week as they are pursuing an important goal. Therapists then seek additional data, summarize their perception of the patient's experience in the form of the cognitive model, and ensure that they correctly understood the patient.

The clinician might then propose an intervention aimed toward any part (or at several parts) of the cognitive model. For example, they might do problem-solving about the problematic situation, evaluate the client's automatic thoughts, elicit, and examine the patients' deeper beliefs, teach the patient emotional regulation skills, or do behavioral skills training. When they have finished discussing the problem or goal, patients are asked what they want to remember (and clinicians might suggest additional points). This summary is recorded as therapy notes that clients read daily as part of their Action Plan. At the end of sessions, clinicians ensure that patients have a good grasp of key points of the session and are highly likely to read their therapy notes and carry out their Action Plans. Then they elicit and respond to feedback.

Cognitive Techniques

The clinician elicits patients' thinking when they have experienced a significant affect shift, have engaged in significantly dysfunctional behavior, or have experienced an obstacle to progress. Clinicians generally ask, "What was just going through your mind?" or "What were you thinking?" Automatic thoughts may be in verbal and/or imaginal form.

Clinicians share their conceptualization with the patient; then they often collaboratively decide to evaluate the patient’s thoughts.

Guided discovery, primarily through Socratic questioning, is the major methods for helping patients assess their thinking. Typical techniques involve labeling distortions, reality testing the evidence for and contrary to the thoughts, seeking alternative explanations, decatastrophizing, examining the utility of the thought, and asking patients to consider how they would advise another person who had the same thoughts (Fig. 32.9–2).

If the process of evaluation results in a positive affect shift, the clinician makes sure patients’ new understandings are recorded, on cards, in a notebook, or on a smartphone so patients can reflect on these new ideas daily. When thoughts are in imaginal form, clinicians often use imaginal techniques in addition to or instead of verbal techniques.

Sometimes a realistic appraisal of patients’ thinking indicates that a given thought is largely valid. In this case, the clinician may engage them in problem-solving, identifying, and assessing the deeper meaning they have attached to their thoughts, and/or moving toward acceptance of the situation, along with focusing on ways to enrich other parts of their lives.

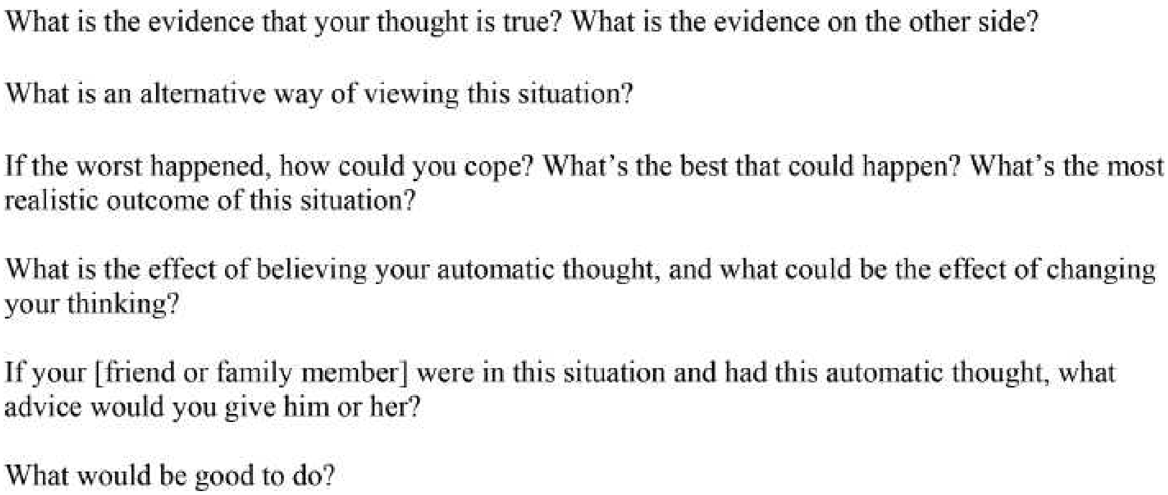

What is the evidence that your thought is true? What is the evidence on the other side?

What is an alternative way of viewing this situation?

If the worst happened, how could you cope? What’s the best that could happen? What’s the most realistic outcome of this situation?

What is the effect of believing your automatic thought, and what could be the effect of changing your thinking?

If your [friend or family member] were in this situation and had this automatic thought, what advice would you give him or her?

What would be good to do?

FIGURE 32.9–2. List of Socratic questions. (Reprinted from © Beck J. *Cognitive Behavior Therapy: Basics and Beyond*. 3rd ed. Guilford Press; 2020.)

Behavioral experiments are commonly used to help patients evaluate their predictions. Examples of techniques used in session include panic inductions (so patients can learn that no matter how intense their symptoms become, a feared catastrophe does not happen), reading a paragraph in a book (so depressed patients can learn that they have not lost the ability to concentrate), or rubbing their hands on the bottom of their shoes and refraining from washing (so patients with OCD and a fear of contamination can learn that their obsessions are inaccurate, they can withstand the anxiety, and the anxiety that arises does not increase indefinitely without resolving when they prevent themselves from engaging in a compulsion). Clinicians also help patients devise behavioral experiments to test predictions about situations that arise outside of session (such as expressing a strong preference to see if a friend becomes angry).

To identify patients' underlying beliefs (their assumptions and basic beliefs about themselves, their worlds, and others), clinicians ask patients for the meaning of their thoughts (referred to as the "downward arrow technique"), examine the recurrent themes in patients' thoughts, directly elicit a rule or attitude, review a questionnaire patients have completed, such as the Personality Belief Questionnaire (J. Beck, 2005), or provide a list of common core beliefs to see which one(s) resonate.

Clinicians use a variety of techniques to help patients assess and respond to their deeper-level cognitions. They use psychoeducation to explain how the patient's information processing system automatically processes data that seems to support their beliefs but discounts or fails to register data that contradicts it. Socratic questioning, examining advantages and disadvantages of maintaining a given belief, formulating a more adaptive, realistic belief, and historical tests of the belief are frequently used techniques. Clinicians also teach patients how to continually identify and reframe evidence that seems to support the old belief and how to identify evidence that is contrary to this belief but supports the new belief. These techniques help patients change their ideas at an intellectual level. To modify their ideas at a deeper, "gut" or "emotional" level, other techniques may be necessary. Imagery, behavioral experiments, acting "as if" the patient fully accepts the new belief as true, metaphors, and experiential techniques (such as roleplaying or psychodrama to restructure the meaning of specific current or childhood experiences) can facilitate a new understanding.

Behavioral Techniques

Problem-solving is a key technique for most patients unless they have a circumscribed problem with anxiety (e.g., specific phobia). Some patients also need to learn effective methods for solving problems and making decisions on their own. For younger patients or those with limited problem-solving skills, a more structured problem-solving approach can be used (e.g., ITCH: Identify the problem; Think about solutions; Choose one solution to put into action; “How did it work out?”).

Behavioral activation, accomplished through activity scheduling, is very important for most depressed patients and others, especially those with entrenched avoidance. Clinicians ask patients to verbally describe or track their hour-to-hour activities. They discuss which activities, usually passive and solitary, are maladaptive if overused because they maintain or worsen their mood (for example, napping, resting, watching excessive television, surfing the web, and playing video games) and which activities have the potential to bring patients a sense of mastery and pleasure or are more in line with their values. Then they formulate a plan in session to increase their active engagement, either creating a model schedule with specific activities for the full day or scheduling specific activities throughout the week. Clinicians help patients predict and respond to unhelpful thoughts just before, during, and after their activities and help them set up reminders (e.g., by setting alarm reminders or posting their schedule in a prominent place).

Some patients, especially those who are depressed, have difficulty making decisions. Creating a four-square chart of costs and benefits or advantages and disadvantages of one option versus another option can help.

Graded task assignments are important for patients who are depressed, procrastinating, or overwhelmed by what they have to do. The clinician teaches them to break a large task into smaller components. A patient whose apartment is quite messy might initially just tackle the dishes in the sink or go one room at a time with a trash bag, looking for obvious things to throw away. Tasks can also be broken down by time: spend 10 minutes separating junk mail from important mail, spend 10 to 15 minutes paying bills.

Exposure Exercises

Exposure exercises are used with anxiety disorder clients to counteract avoidance and test feared predictions. Prior to engaging in the exposure exercise, clinicians identify the patient’s predictions. They then develop a plan to engage with the stimulus, including how to determine if their feared predictions are accurate and how to cope more effectively. Clinicians emphasize an inhibitory learning approach during the exposure, which involves taking note of discrepancies between the feared predictions and what actually occurs. Patients are instructed to record the outcome immediately after the exposure, and clinicians help them draw helpful conclusions.

Various forms of exposure exercises are used based on what stimulus the patient fears and avoids. Interoceptive exposures are used to expose panic disorder clients to feared sensations (heart racing; dizziness). Imaginal exposures expose patients to feared memories or images (intrusive memories of a traumatic event). In-vivo exposures expose patients to feared and avoided external situations (driving over a bridge; answering a question during a class).

Exposure and response prevention is often used with patients who have OCD. In a graded fashion, the clinician exposes patients to feared situations that often trigger obsessions, coaching them to refrain from engaging in compulsions so they can test their predictions related to feared outcomes and increase their tolerance of anxiety.

Motivational Techniques

Patients vary in their interest and engagement in therapy. To boost motivation to engage in treatment, work collaboratively in sessions, and follow through with Action Plans between sessions, clinicians may use a variety of techniques including motivational interviewing, reminders of patients’ key values and goals, an examination of advantages and disadvantages of various behavioral options, and Socratic questioning of cognitions that decrease their willingness to change.

Emotional Regulation Techniques

An important therapeutic goal is for patients to understand and accept their negative emotions and to discover that they can tolerate negative emotion without turning to maladaptive strategies such as avoidance, self-harm, aggression, or the use of substances. Psychoeducation, Socratic questioning of beliefs about the experience of emotion, and behavioral experiments to test these beliefs are often used to facilitate this understanding.

Patients also benefit from learning specific skills to better regulate their emotional arousal. They include mindfulness techniques, relaxation exercises, refocusing on an engaging or valued activity, reading therapy notes, evaluating automatic thoughts, positive social interaction or seeking support from others, and self-soothing activities such as listening to music, physical contact with a pet, walking in nature, or focus on a sensory experience. Patients may also benefit from learning self-compassion skills if their distress is related to a high degree of self-criticism.

Environmental Interventions

Some patients continually face such adverse circumstances that an environmental change is required. Patients who are emotionally or physically abused in their living environment, for example, may need help to identify a new place to live for some period of time. Other patients might benefit from an examination of the pros and cons of staying in a hostile work environment, and, if needed, coaching on job-seeking and interviewing skills.

Interpersonal Techniques

Techniques focus on solving interpersonal problems, teaching social skills, and modifying maladaptive beliefs about themselves, others (including, at times, the clinician), and relationships. Roleplaying is often used to help patients increase their facility in initiating and sustaining casual conversation, being appropriately assertive, accurately reflecting what an intimate has said, and skillfully conversing during emotional interactions. Including a family member or intimate in session, whether or not there is friction in their relationship, can help clinicians round out their conceptualization of patients and their problems and provide an opportunity for psychoeducation and cooperation in helping patients reach their goals.

Biologic Interventions and Treatment Adherence

It is important for clinicians to look broadly at patients and, for some, to encourage healthier habits: facilitating change, for example, in eating habits, caffeine intake, use of substances, and exercise. CBT for insomnia (CBTi) has been shown to be effective (van der Zweerde et al., 2019). Medication adherence can be increased by considering advantages and disadvantages of taking medication, examining interfering cognitions, promoting tolerance of side effects, behavioral experiments (such as taking the first dose in the office), and checking on practical matters such as potential problems in affording and obtaining the medication or remembering to take it.

CLINICAL ISSUES

Indications

CBT may be used to (1) eliminate or moderate the symptoms of a psychiatric disorder or psychological distress as an independent treatment or in combination with medication and/or other biologic treatment; (2) modify underlying beliefs and schemas that contribute to the development and maintenance of disorders; (3) address psychosocial problems (e.g., marital discord, job stress, caregiver burnout) that may have preceded, coincided with, or been caused by the disorder; (4) reduce the likelihood of relapse or recurrence of a disorder after symptoms have resolved; (5) increase adherence to recommended medical treatment (e.g., medication, weight-loss, diabetes management); or moderate the impact of a medical condition on level of functioning (e.g., chronic pain, cancer); (6) live a more valued, fulfilling life.

Patients. Ideal CBT patients quickly learn to recognize and label their emotions, automatic thoughts, behaviors, and the connection between these experiences. Nonetheless, patients who initially have difficulty with this kind of self-monitoring are typically able to learn these skills through instruction and practice. When patients have learning difficulties or impaired intellectual skills, they often learn better when clinicians arrange experiences and help them draw conclusions that fortify their adaptive cognitions. Accordingly, even average intelligence is not a prerequisite for treatment. Motivation for therapy is also helpful, but not required, as patients who are hopeless, unmotivated, or lethargic may become more invested in treatment as they begin to experience improvement. Patients who are new to psychotherapy often report appreciating the CBT approach as they see it as being pragmatic and comprehensible. Some long-term patients of other psychotherapy approaches may have an initial period of adjustment due to CBT’s more structured nature. Nonetheless, most patients are able to make the transition, and structure can be employed flexibly for others.

Clinicians. The ideal cognitive therapist is attentive, empathic, and nonjudgmental, as is desired in most models of psychotherapy. To be maximally effective in an active treatment such as CBT, clinicians need to be moderately directive in a collaborative manner. They also must be able to ask for and receive feedback in a nondefensive manner. Although a skilled clinician knows how to sit back and listen when appropriate, they do not allow long silences as a matter of routine, and they are willing to answer most questions with a therapeutically direct answer. CBT clinicians aim to be transparent and share their conceptualization and rationales for interventions. Treatment frequently lasts for weeks or months, but not usually years.

As would be expected, several studies have demonstrated a high correlation between a clinician's level of competency, as measured by the Cognitive Therapy Rating Scale (Young & Beck, 1980) and successful treatment outcomes. Competency in CBT requires more than a passive knowledge of the theory and techniques. Clinicians who wish to specialize in CBT usually require a minimum of 1 year of supervised clinical experience based on the theoretical formulation of the cognitive model, though periodic training and supervision are recommended as research validates refinement in cognitive formulations and associated treatment. Ideally, supervision is performed by an expert cognitive therapist who reviews, rates, and provides feedback on recordings of therapy sessions.

Limitations

Early outcome studies on CBT indicated that depressed patients could show significant improvement in approximately 12 weeks (of biweekly and weekly sessions). Likewise, individuals with panic disorder evinced a significant reduction in panic attacks after 4 to 8 sessions of CBT. While the short-term results are impressive, they have, at times, been interpreted too broadly, creating an inaccurate perception that all types of pathology and clinical presentations will respond as rapidly.

Research on CBT for individuals diagnosed with bipolar disorder, chronic depression, schizophrenia, and personality disorders allow for a longer course of treatment (e.g., 1 year), and results have been promising. Moreover, even for uncomplicated cases, booster sessions at longer intervals (e.g., 3 or 6 months) after termination of the course of regular treatment can help with the maintenance of treatment gains and prevent relapse.

Complications

At present, there are no known complications caused by CBT when it is used properly and in the appropriate context. The model is designed to teach patients a variety of skills to manage their presenting problems, which gives them a sense of empowerment and self-efficacy. Accordingly, the CBT model does not encourage patient dependence on the clinician. Termination issues, which are relevant in any form of psychotherapy, are dealt with proactively by the clinician who explicitly educates the patient about the time-effectiveness of treatment, the importance of learning to be one's own therapist, and the use of booster sessions and self-therapy sessions. Unhelpful cognitions related to termination are addressed through standard techniques.

Clinical crises are handled with a combination of standard risk-reduction procedures (e.g., increased frequency of sessions, inclusion of significant others, emergency consultations, hospitalizations), along with a cognitive conceptualization of the crisis situation. Once an acute crisis subsides, clinicians explore the patient's vulnerability (the activation of their schemas) that was triggered by the activating event and help the patient develop alternative coping strategies for similar future situations. The goal is to help patients prevent such instances, gain confidence in their ability to cope in the future, and become generally less vulnerable to life stressors.

Contraindications

CBT has been demonstrated to be efficacious as part of a treatment package for severe and persistent mental illness, although it is generally recommended that pharmacotherapy options be assessed as part of the treatment regimen for disorders such as schizophrenia and bipolar disorder.

GOALS OF TREATMENT

CBT has three major goals. The first is to reduce patients' symptomatology and suffering as quickly as possible and help them reach remission. The second is to help clients live a more fulfilling life consistent with their values. The third is to teach them skills that they can employ long-term to increase their resilience, maintain improvement, and reduce relapse.

ETHICAL ISSUES

The ethical issues that arise in many forms of psychotherapy also may arise in CBT: eliciting informed consent, reporting child abuse, breaking confidentiality when the patient's life or the lives of other people are in imminent danger, maintaining appropriate boundaries, and avoidance of dual relationships. Additionally, clinicians are cognizant of not continuing regularly scheduled sessions when ongoing treatment is no longer deemed necessary. Once patients have made strides in accomplishing their goals, are no longer experiencing acute symptoms, and have demonstrated an ability to maintain treatment gains, clinicians may suggest tapering the frequency of sessions, followed by scheduling only periodic "booster" sessions.

RESEARCH AND EVALUATION

CBT has more research evidence in support of its efficacy and effectiveness than any other form of psychotherapy, and as a result, has been argued to be the gold standard of psychotherapy (David, Cristea, & Hofmann, 2018). Systematically controlled studies have demonstrated the efficacy of CBT for a range of psychiatric disorders, including serious mental health conditions. A review of meta-analyses concluded that the strongest support for CBT therapy exists in the treatment of anxiety disorders, somatoform disorders, bulimia, anger, and general stress. CBT has been successfully adapted across populations of different ages, ethnicities, gender and gender identities, sexual orientation, socioeconomic status, and cultural backgrounds, in a wide variety of settings, including real-world settings.

A common misconception in psychological intervention research is the notion of the Dodo Bird verdict, which concludes that all psychological treatments are equal in their effects. A meta-analysis, however, concluded that CBT was significantly more efficacious than other bona fide interventions (e.g., psychodynamic, interpersonal, and supportive therapies) for anxiety and mood disorders and is considered the first-line treatment for anxiety disorders. CBT has also demonstrated superiority to other treatments when comparisons are limited to studies with high methodologic quality (e.g., random assignment, evaluator blind to condition). Additionally, Hofmann and colleagues (2012) noted that 7 of 11 meta-analyses demonstrated higher response rates for CBT than comparison conditions, while only one identified lower response rates for CBT.

Researchers have used functional neuroimaging to study the influence of psychological interventions on brain function. Neuroimaging studies have demonstrated that individuals with psychological disorders tend to have different brain functioning than nonclinical controls. Interestingly, studies have found that following CBT, the brain functioning of individuals with certain psychological disorders resemble control subjects (who did not receive CBT). Neuroimaging research has also compared brain functioning of individuals undergoing treatment with CBT to those taking psychotropic medication. The results indicated that CBT causes similar changes in brain function to psychotropic medication in the treatment of OCD, SAD, specific phobias, and depression. Nonetheless, there are differences in how disorders present in terms of both symptoms and brain function. A more nuanced study found some differences in brain function between successful CBT and pharmacotherapy.

CBT has been adapted to be administered using telehealth, the internet, and computer-based formats. Video-delivered CBT has been demonstrated to be as efficacious as in-person treatment when compiling outcomes across studies. A meta-analysis concluded that internet- and computer-based CBT demonstrated large effect sizes when compared to a control group in the treatment of all disorders that were studied (depression, panic disorder, SAD, and GAD). Moreover, patient adherence was good, and patients reported being satisfied with treatment.

In sum, although CBT was originally developed as a treatment for unipolar depression and anxiety disorders, it has been successfully adapted to treat various clinical populations, including patient groups who were traditionally thought to be unresponsive to psychosocial treatments (e.g., schizophrenia). A growing body of research has demonstrated the efficacy and effectiveness of CBT in addressing various disorders and patient populations. While some researchers study mechanism of change, additional studies, using mediational analyses, should be conducted to continue to establish empirical support for the cognitive model and to aid in refining treatment. Additionally, as researchers become better able to identify biologic processes of psychological disorders, future research may attempt to integrate biologic processes with the cognitive model, consistent with the National Institute of Mental Health Research Domain

Criteria initiative, which was implemented to develop an alternative, more homogenous and etiologically relevant, psychological disorder classification system to the DSM. Finally, efforts to disseminate CBT to clinicians have lagged behind the growing efficacy and effectiveness research base. Future research should place a greater emphasis on determining the most effective and efficient methods to disseminate CBT to clinicians so that patients have greater access to competently delivered evidence-based treatment.

FURTHER READINGS

Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. *PLoS ONE*. 2010;5:e13196.

Beck JS. *Cognitive Therapy for Challenging Problems: What to Do When the Basics Don't Work*. Guilford Press; 2005.

Beck JS. *Cognitive Behavior Therapy: Basics and Beyond*. 3rd ed. Guilford Press; 2020.

Beck AT. *Depression: Clinical, Experimental and Theoretical Aspects*. Hoebner Medical Division, Harper and Row; 1967.

Beck AT, Davis DD, Freeman A. *Cognitive Therapy of Personality Disorders*. 3rd ed. Guilford Press; 2015.

Beck AT, Finkel M, Beck JS. The theory of modes: applications to schizophrenia and other psychological conditions. *Cognit Ther Res*. 2021;45:391–400.

Beck AT, Grant P, Inverso E, Brinen A, Perivoliotis D. *Recovery-Oriented CBT for Serious Mental Health Conditions*. Guilford Press; 2020.

Beck AT, Rector NA, Stolar N, Grant P. *Schizophrenia: Cognitive Theory, Research, and Therapy*. Guilford Press; 2009.

Beck AT, Rush AJ, Shaw BF, Emery G. *Cognitive Therapy of Depression*. Guilford Press; 1979.

Beck AT, Wright FD, Newman CF, Liese BS. *Cognitive Therapy of Substance Abuse*. Guilford Press; 1993.

Clark DA, Beck AT. *Scientific Foundations of Cognitive Theory and Therapy of Depression*. John Wiley & Sons; 1999.

Clark DA, Beck AT. *Cognitive Therapy of Anxiety Disorders: Science and Practice*. Guilford Press; 2010.

David D, Cristea I, Hofmann SG. Why cognitive behavioral therapy is the current gold standard of psychotherapy. *Front Psychiatry*. 2018;9:4.

DeRubeis RJ, Hollon SD, Amsterdam JD, et al. Cognitive therapy vs medications in the treatment of moderate to severe depression. *Arch Gen Psychiatry*. 2005;62:409–416.

Dugas MJ, Marchland A, Ladouceur R. Further validation of a cognitive-behavioral model of generalized anxiety disorder: diagnostic and symptom specificity. *J Anxiety Disord*. 2005;19;329–343.

Fairburn CG, Cooper Z, Shafran R. Cognitive behaviour therapy for eating disorders: a "transdiagnostic" theory and treatment. *Behav Res Ther*. 2003;41:509–528.

Fernandez E, Woldgabreal Y, Day A, Pham T, Gleich B, Aboujaoude E. Live psychotherapy by video versus in-person: a meta-analysis of efficacy and its relationship to types and targets of treatment. *Clin Psychol Psychother*. 2021;28;1535–1549.

Fuchs T. Neurobiology and psychotherapy: an emerging dialogue. *Current Opinion Psychiatry*. 2004;17:479–485.

Furmark T, Tillfors M, Marteinsdottir I, et al. Common changes in cerebral blood flow in patients with social phobia treated with citalopram or cognitive-behavioral therapy. *Arch Gen Psychiatry*. 2002;59:425–433.

Goldapple K, Segal Z, Garson C, et al. Modulation of cortical-limbic pathways in major depression: treatment-specific effects of cognitive behavior therapy. *Arch Gen Psychiatry*. 2004;61:34–41.

Hofmann SG, Asnaani A, Vonk IJJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. *Cognit Ther Res*. 2012;36:427–440.

Hollon SD. Is cognitive therapy enduring or antidepressant medications iatrogenic? Depression as an evolved adaptation. *Am Psychol*. 2020;75;1207–1218.

Kazantzis N, Whittington C, Zelencich L, Kyrios M, Norton PJ, Hofmann SG. Quantity and quality of homework compliance: a meta-analysis of relations with outcome in cognitive behavior therapy. *Behav Ther*. 2016;47:755–772.

Knapp P, Kieling C, Beck AT. What do psychotherapists do? A systematic review and meta-regression of surveys. *Psychother Psychosom*. 2015;84:377–378.

Kovacs M, Rush AJ, Beck AT, Hollon S. Depressed outpatients treated with cognitive therapy or pharmacotherapy. *Arch Gen Psychiatry*. 1981;38:33–39.

Miller SD, Duncan BL, Brown J, Sorrell R, Chalk MB. Using formal client feedback to improve retention and outcome: making ongoing, real-time assessment feasible. *J Brief Ther*. 2006;5:5–22.

Newman CF, Leahy RL, Beck AT, Reilly-Harrington NA, Gyulai L. *Bipolar Disorder: A Cognitive Therapy Approach*. American Psychological Association; 2002.

Rush AJ, Beck AT, Kovacs M, Hollon S. Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognit Ther Res*. 1977;1:17–37.

Strunk DR, Brotman MA, DeRubeis RJ, Hollon SD. Therapist competence in cognitive therapy for depression: predicting subsequent symptom change. *J Consult Clin Psychol*. 2010;78:429–437.

Tolin DF. Is cognitive-behavioral therapy more effective than other therapies? A meta-analytic review. *Clin Psychol Rev*. 2010;30:710–720.

van der Zweerde T, Lampros B, Kyle SD, Lancee J, van Straten A. Cognitive behavioral therapy for insomnia: a meta-analysis of long-term effects in controlled studies. *Sleep Med Rev*. 2019;48:101208.

Wells A. *Metacognitive Therapy for Anxiety and Depression*. Guilford Press; 2009.

Wenzel A, Liese BS, Beck AT, Friedman-Wheeler DG. *Group Cognitive Therapy for Addictions*. Guilford Press; 2012.

Wright JH, Sudak DM, Turkington D, Thase ME. *High-Yield Cognitive-Behavior Therapy or Brief Sessions: An Illustrated Guide*. American Psychiatric Publications; 2010.

Young J, Beck AT. *Cognitive Therapy Rating Scale*. Beck Institute for Cognitive Behavior Therapy; 1980.

32.10 Interpersonal Psychotherapy

ROBERT W. GUYNN, M.D.

It is more accurate to speak of the interpersonal psychotherap*ies* (plural). There have been long-term, insight-oriented, dynamically based, and open-ended interpersonal psychotherapies, and there are more modern, short-term, time-limited approaches that avoid both extensive theory and insight. What all these therapies hold in common is the belief that the interpersonal relationships of a patient, now and in the past, to a greater or lesser extent, have a role in the etiology of his or her psychopathology and its treatment. It is unfortunate that all forms of the interpersonal psychotherapies bear the same name. In order to minimize confusion, for the rest of the discussion the field will be split into the insight-oriented interpersonal psychotherapies (IO-IPT) and non–insight-oriented interpersonal psychotherapies (IPT in keeping with the current usage of that abbreviation).

HISTORY

To the American psychiatrist Harry Stack Sullivan (1892–1949) goes the honor of the title of the Father of Interpersonal Psychiatry. It was from the 1920s through the 1940s that Sullivan developed his ideas as an alternative formulation to classic psychoanalytical theory. Simply put, Sullivan theorized that everything people consider human comes from their interactions with others. He postulated that there are two basic drives: (physical) satisfaction (food, warmth, shelter, sex…) and (interpersonal) security. Security here is the sense that one is an adequate human being (having a healthy self-esteem, feeling confident to handle stress, being comfortable among others, being interpersonally competent, being relatively anxiety-free, etc.). Sullivan pointed out that not only do humans learn language and culture from the members of their family, but they also learn how to separate fantasy from reality, develop a sense of self and self-esteem (or lack thereof), develop strategies for coping with anxiety, etc. Sullivan postulated social stages of child development (rather than psychosexual ones), the normal outcome of which being the more-or-less enduring patterns of thinking and behaving that constitute one's personality. These patterns develop not so much because of inner conflicts, according to Sullivan, but as an attempt to deal with the human environment of one's family, school, and community. To the extent that these patterns are significantly maladaptive, one has a mental illness.

It is probably fair to say that, although Sullivan himself was a clinician, he is primarily remembered for his theories, through his lectures and largely posthumous books. It was left to others to flesh out the therapeutic techniques. Some, such as Frieda Fromm-Reichmann (1889–1957), Harold Searles (1918–2015), Karen Horney (1885–1952), and Hilda Bruch (1904–1984) blended some of Sullivan's ideas with more classic psychoanalytic thinking. They, thereby, were able to adapt the traditional theory and therapy to nontraditional populations, including schizophrenia. Others, such as Donald Kiesler (1933–2007), proposed alternative therapeutic strategies. Of note, among all of these early writers is the recognition that a therapist is not and cannot be a neutral observer.

The general tasks of an insight-oriented interpersonal psychotherapist are to identify patterns of behavior and thinking from the patient's history, support the self-esteem of the patient and re-enforce it appropriately, address the positive aspects of the patient, and his or her successes. The therapist is friendly, even an advocate, but not a friend. An important element of the therapy is the assumption that the therapist is a participant, and, therefore, cannot be entirely objective and cannot avoid being the target of the patient's stereotypical behavior. If the patient tends to interact with others by being manipulative, challenging, provocative, or suspicious, sooner or later they will interact with the therapist in the same way. The ability of the therapist to read his or her own reactions to the patient objectively and reflect them therapeutically is an essential part of the treatment.

Kiesler, to take but one specific example, built on Sullivan's ideas and developed a more formal therapeutic strategy. He very much emphasized that the therapy sessions are dyadic—that is, the therapist is a participant observer. He uses the term "impact messages" to designate all the feelings, thoughts, and behaviors of the therapist in response to the patient. Kiesler distinguishes several groups of impact messages: direct feelings, action tendencies, cognitive attributions, and fantasies. The therapist consciously and continuously monitors his or her reactions (impact messages) and assumes that the patient is provoking in the therapist what they provokes in other significant relationships. That is, the therapist, far from being aspiring to be neutral, uses his or her own reactions to the patient to understand the patient's maladaptive interpersonal style. The impact messages are not considered counter-transference, as such, but the essentially normal response that anyone might have to the repetitive interpersonal patterns that the patient has developed as a strategy for living, perhaps because it is what worked in the nuclear family. The difference is that the therapist's response and reactions will attempt to be therapeutic rather than "usual." In fact, the therapist's failure to respond to the patient's interpersonal psychopathology in the expected way begins to undermine the "usefulness" of the psychopathology.

For example, if a person treats people in his or her life as objects to be controlled, then there is no reason to think that the patient would treat the therapist any differently. In this example, the therapist might find themself becoming progressively irritated with the patient as the sessions go on. The therapist would then internally analyze the source of the irritation and, hopefully, ascribe it to the subtle, or not so subtle, attempts at control by the patient. This strategy makes the therapist a very active participant, monitoring his or her own feelings and reactions as a strategy to both understand and treat the patient. If the insight-oriented interpersonal psychotherapist were to choose to share his or her own reaction with the patient, in this case, the interpretation would not be to reflect back on a specific early relationship of the patient but point out to the patient what the therapist is sensing ("I feel that you are trying to control me") and wonder aloud if this is not an example of the patient's automatic and well-established behavioral pattern presumably identifiable from the patient's interpersonal history. The therapist might then discuss with the patient alternative and less stereotypic strategies.

Another example might be the shy and anxious person who treats interpersonal interactions as painful reminders and proofs of their inadequacy ("I never know what to say and must come across as stupid") will probably be looking for judgments from the therapist and be unusually sensitive to any reaction of the therapist that implies criticism. One might expect such a patient to provoke in the therapist feelings that the patient must not be criticized. The IO-IPT therapist might choose to comment to the patient that they are feeling that the patient is projecting a sense of fragility and vulnerability—is that really necessary?

In fact, Kiesler suggests for any given statement by the patient, the therapist can choose among any of the six categories of reply. They may simply respond to the manifest content of the statement or may ignore it and change the subject. One may also interpret the statement of the patient as an unconscious message either of the patient to himself or the patient to the therapist. Finally, the therapist may choose to share either his personal reflection or his reaction to the patient's statement.

In this example of IO-IPT, Kiesler therapy has four phases: (1) hooking, (2) uncovering, (3) resolving, and (4) termination. No set number of sessions is proposed. The last three phases would seem to be self-evident and do not need exposition here. The first phase ("hooking") might be worth addressing. It was an assumption that many individuals seek to come to therapy not so much to become better but to become better at their psychopathology. Such patients will quickly tire of therapy once whatever provoked their anxiety and drove them to seek help abates. The "hooking phase" is a conscious attempt by the therapist to intrigue the patient in the goal of health.

In other instances, the problem may at least partially reside outside the patient's own interpersonal style. An example would be the case of the depressed woman married to a perfectionist. Without a trace of insight, the patient pinpointed the onset of her depression to the date of her marriage. Being not as educated as her husband, she readily accepted him as an authority at all levels. His pattern was to "lovingly" make suggestions about how the patient might do things better, from how she cooked to how she held the leashes when walking the dogs. After a couple of years, the patient gradually lost much of her self-esteem and had become paralyzed when trying to make any decision. During the exploration of the pattern of the interpersonal relationship between the patient and her husband, he announced that he had found another. Much to the surprise of all, the patient was not devastated but actively threw him out and never looked back. The depression resolved.

The caveat to all of the above is the fact that such approaches have not been subjected to rigorous scientific study. However logical and attractive the theories and approaches may appear, they have not been put to sufficient testing.

TIME-LIMITED INTERPERSONAL PSYCHOTHERAPY

A more modern version of interpersonal psychotherapy, IPT, is a nonpsychodynamic, time-limited psychotherapy developed in the 1970s by Gerald Klerman and colleagues for the treatment of nonbipolar, nonpsychotically depressed outpatients. This version of IPT makes no assumptions about the causes of psychiatric illness, other than that the development and maintenance of some psychiatric illnesses occur in a social and interpersonal context and that the onset, responses to treatment, and outcomes are influenced by the interpersonal relations between the patient and significant others. Klerman's approach is reminiscent of cognitive-behavioral therapy in that the focus is on the here and now. There are, also, a set number of sessions, transference may be ignored or undermined, and a workbook may be used. As with all interpersonal psychotherapies, in IPT the focus is primarily on the patient's current and recent interpersonal relationships. Other differences between IO-IPT and IPT are that in the latter there is less interest in the origin of the problem and the therapist does not necessarily reflect on the relationship between the therapist and the patient. A major advantage of ITP over the other interpersonal approaches is that IPT can be researched.

Time-limited IPT moves through defined phases, each of which is associated with specific strategies and tasks for the therapist and patient. Its well-defined treatment strategies are aimed at resolving problems within four social domains: grief, interpersonal role disputes, role transitions, and interpersonal deficits. The overall goal is to reduce and/or eliminate psychiatric symptoms by improving the quality of the patient's current interpersonal relations and social functioning. Time-limited IPT has been found to be efficacious for major depression and has also been successfully adapted to treat other types of mood and nonmood disorders. IPT has also been successfully applied to a group format and is being evaluated in a number of trials. The extensive empirical background and theoretical foundation, as well as the strategies and techniques of IPT, are fully described in comprehensive books by Myrna Weissman and her colleagues.

Differences from Other Psychotherapies

Time-limited IPT is, of course, not the only psychological approach for the treatment of psychiatric disorders. The methods and techniques of many schools of psychotherapy share common ground, including attempts to help patients gain a sense of mastery and reduce social isolation. A major difference among the therapies lies in their conceptualization of the causes of the patient's problems and the typical length and foci of treatment. Although they use similar techniques, IPT also differs from cognitive and behavioral approaches in that maladaptive thoughts and behaviors are addressed only as they apply to problematic interpersonal relations, with the goal to change the relationship pattern rather than associated cognitions. In contrast to other therapies, IPT recognizes but does not necessarily focus on the patient's personality characteristics.

The focus of IPT is on identifying and changing the maladaptive interpersonal context in which the psychiatric problem has developed and is perpetuated. It is designed to help patients identify and address largely current interpersonal problems within four domains: grief, identified as the problem area when the onset of the patient's symptoms is associated with the loss of a person or a relationship, either recent or in the past; interpersonal role disputes, which are conflicts with a significant other (e.g., a partner, other family member, coworker, or close friend) that emerge from differences in expectations about the relationship; role transitions, which include difficulties associated with a change in life status (e.g., graduation, leaving a job, moving, marriage/divorce, retirement, change in health status); and interpersonal deficits, which apply to those patients who are socially isolated or are in chronically unfulfilling relationships.

TECHNIQUES

Therapeutic Stance

Similar to the other therapies, IPT places importance on establishing positive therapeutic alliance between therapist and patient. Specifically, the interpersonal psychotherapeutic stance is one of warmth, support, and empathy. The interpersonal psychotherapist is active and advocates for the patient rather than remaining neutral. Moreover, the therapeutic relationship is not viewed as a classic manifestation of transference/countertransference, nor is it treated as a friendship. As noted, IPT is similar to many other therapies at the level of psychotherapeutic techniques and therapeutic stance but is distinct at the level of strategies.

Timeline of Treatment

The typical course of IPT lasts 12 to 20 sessions over a 4- to 5-month period. The initial phase is dedicated to identifying the problem area that will be the target for treatment. The intermediate phase is devoted to working on the target problem area(s). The termination phase is focused on consolidating gains made during treatment and preparing the patients for future work on their own (Table 32.10–1).

Initial Phase. Sessions one through five typically constitute the initial phase of IPT. After assessing the patient’s current psychiatric symptoms and obtaining a history of these symptoms, the therapist explains the rationale of IPT, underscoring that therapy will focus on identifying and altering dysfunctional interpersonal patterns related to psychiatric symptomatology. To determine the precise focus of treatment the therapist will conduct an interpersonal inventory with the patient and will develop an interpersonal formulation based on this. In the interpersonal formulation, the therapist will link the patient’s psychiatric symptomatology with the therapist’s identification of the problem area(s). The patient’s concurrence with the therapist’s assessment of the problem and agreement to work on this area are essential before beginning the intermediate treatment phase.

Table 32.10–1.
Phases of Time-limited Interpersonal Psychotherapy

- Initial phase: Sessions 1–5
  - Give the syndrome a name; provide information about the prevalence and characteristics of the disorder
  - Describe the rationale and nature of the therapy
  - Conduct the interpersonal inventory to identify the current interpersonal problem area(s) associated with the onset or maintenance of the psychiatric symptoms
  - Review significant relationships, past and present
  - Identify interpersonal precipitants of episodes of psychiatric symptoms
  - Select and reach consensus about the interpersonal problem area(s) and treatment plan with patient
- Intermediate phase: Sessions 6–15
  - Implement strategies specific to the identified problem area(s)
  - Encourage and review work on goals specific to the problem area
  - Illuminate connections between symptoms and interpersonal events during the week
  - Work with the patient to identify and manage negative or painful affects associated with his or her interpersonal problem area
- Termination phase: Sessions 16–20
  - Discuss termination explicitly
  - Educate the patient about the end of treatment as a potential time of grieving; encourage the patient to identify associated emotions
  - Review progress to foster feelings of accomplishment and competence
  - Outline goals for remaining work; identify areas and warning signs of anticipated future difficulty
  - Formulate specific plans for continued work after termination of treatment

DIAGNOSIS AND ASSIGNMENT OF THE SICK ROLE. After a thorough psychiatric interview has been conducted, the therapist gives the patient a formal diagnosis from the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), and the patient is formally diagnosed and assigned the sick role. Therapist and patient then discuss the diagnosis, as well as what might be expected from treatment. Assignment of the sick role during this phase serves the dual function of granting the patient both the permission and the responsibility to recover. The purposes of assigning the sick role are both theoretical and practical. Consistent with the theoretical model, receiving a formal diagnosis reinforces the idea that the patient has a known condition that can be treated. The giving of a diagnosis also explicitly identifies the patient as being in need of help.

The sick role is not assigned to condescend to the patient but, rather, is designed temporarily to exempt the individual from the responsibilities in order to devote full attention to recovery. This is particularly important for patients who tend to set aside their needs and desires to care for and please others. If this applies in a particular case, the therapist may explicitly highlight the patient's excessive caretaking tendencies and encourage the patient to redirect this energy from others toward his or her recovery. In doing so, the therapist clarifies the rationale behind IPT—that by improving the patient's patterns of interpersonal functioning, the patient's psychiatric symptoms are expected to improve as well. Symptom relief starts with helping the patient to understand that his or her psychiatric symptoms are part of a known syndrome that responds to several treatments and has a good prognosis.

IDENTIFY CURRENT INTERPERSONAL PROBLEMS: THE INTERPERSONAL INVENTORY. The therapist conducts an interpersonal inventory, which includes a review of the patient's past and current social functioning and close relationships. A thorough interpersonal inventory is essential for adequate case formulation and development of an optimal treatment plan. The clinical importance of investing the time involved in conducting a comprehensive interpersonal inventory cannot be overemphasized; accurate identification of the patient's primary problem area can be difficult, but it is the key to success in therapy.

For each individual who is important in the patient's life, the following information is assessed: frequency of contact, activities shared, satisfactory and unsatisfactory aspects of the relationship, and ways in which the patient would like to change the relationship. The interpersonal inventory provides a structure for elucidating the social and interpersonal context of the onset and maintenance of psychiatric symptoms and delineates the focus of treatment. Of particular importance are changes in relationships (e.g., the death of a significant other, a change in jobs, an increase in marital discord, and disconnection from a friend). The therapist obtains a chronologic history of significant life events, fluctuations in mood and self-esteem, interpersonal relationships, and psychiatric symptoms. From this review, the therapist can make a connection between certain life experiences and psychiatric symptoms. As this interrelationship is delineated, patients understand more clearly the rationale of IPT.

ESTABLISH THE RELEVANT PROBLEM AREAS. On completion of the interpersonal inventory, the therapist develops an individualized interpersonal formulation, linking the patient's symptoms to one of the four interpersonal problem areas—grief, interpersonal deficits, interpersonal role disputes, or role transitions. Although the therapist explicitly assigns a problem area, the patient needs to concur with the salience of the problem area proposed and agree to work on it in treatment. Because the four problem areas are relatively inclusive, it is usually easy for the therapist to provide an individualized formulation. Although some patients may present for treatment with difficulties in several problem areas, the time-limited nature of the treatment necessitates a focused approach. The therapist, with the agreement of the patient, should assign one or, at most, two problem area(s). If more than one problem area is identified, the problem area that seems most likely to be responsive to treatment is addressed first. Tackling the most soluble problem first has the advantage that progress on one front often leads to progress on others. Not only are the patient's morale and overall sense of competence enhanced when progress is made on a problem, but also barriers to progress in other areas may be eroded or disappear.

COLLABORATIVELY DEVELOP TREATMENT GOALS. After the major interpersonal problem area(s) associated with the onset or maintenance of the disorder is identified, the therapist and patient work together to formulate a treatment plan with specified goals. The goals developed at this stage will be referenced at each future session and will guide the day-to-day work of the treatment. Goals should flow from the identified problem area and be designed to address it. After identification of a goal, the therapist and patient collaboratively identify specific steps the patient will take to improve relationships and socialization. To the extent possible, goals should be summarized in such a way as to include reference to specific individuals, events, and interpersonal themes to help ensure that they are as personally meaningful to the patient as possible. Before the end of the initial phase, patients may be given a written summary of their goals, which serves as a treatment contract to guide their work for the remainder of treatment.

Table 32.10–2.

Interpersonal Problem Areas: Description, Goals, and Strategies

Problem Area	**Description**	**Goals**	**Strategies**
Grief	Complicated bereavement after death of a loved one	Facilitate the mourning process Help the patient to establish interest in new activities and relationships to substitute for what has been lost	Reconstruct the patient's relationship with the deceased Explore associated feelings (negative and positive) Consider ways of becoming reinvolved with others
Interpersonal deficits	A history of social impoverishment, inadequate, or unsustaining interpersonal relationships	Reduce patient's social isolation Enhance quality of any existing relationships Encourage the formation of new relationships	Review past significant relationships, including negative and positive aspects Explore repetitive patterns in relationships Note problematic interpersonal patterns in the session and relate them to similar patterns in the patient's life
Interpersonal role disputes	Conflicts with a significant other	Identify the nature of the dispute	Determine the stage of the

	—a partner, other family member, coworker, or close friend	Explore options to resolve dispute Modify expectations and faulty communication to bring about a satisfactory resolution If modification is unworkable, encourage patient to reassess the expectations for the relationship and to generate options to either resolve it or dissolve it and mourn its loss	dispute: renegotiation (calm down participants to facilitate resolution); impasse (increase disharmony to reopen negotiation); dissolution (assist mourning and adaptation) Understand how nonreciprocal role expectations relate to the dispute Identify available resources to bring about change in the relationship
Role transitions	Economic or family change—the beginning or end of a relationship or career, a move, promotion, retirement, graduation, diagnosis of a medical illness	Mourn and accept the loss of the old role Recognize the positive and negative aspects of the new role and assets and liabilities of the old role	Review positive and negative aspects of old and new roles Explore feelings about what is lost Encourage development of social support system and new skills

	Restore self-esteem by developing a sense of mastery regarding the demands of the new role	called for in new role

Reprinted with permission from Wilfley D, Stein R, Welch R. Interpersonal psychotherapy. In: Treasure J, van Furth E, Schmidt U, eds. *Handbook of Eating Disorders*. 2nd ed. Wiley; 2003:258.

Intermediate Phase. The intermediate phase typically lasts from 8 to 10 sessions and continues the “work” of the therapy. Once the patient and therapist have agreed on the primary problem area and have set treatment goals, the intermediate phase of treatment begins. An essential task throughout the intermediate phase is to strengthen the connections the patient makes between the changes they are making in their interpersonal life and the changes in his or her psychiatric symptoms.

During the intermediate phase, the therapist implements the treatment strategies specific to the identified problem area as specified by Weissman and colleagues (Table 32.10–2). For an in-depth discussion of the IPT goals and strategies, the therapist is encouraged to read the comprehensive text by Weissman and colleagues.

PROBLEM AREAS

Grief. Grief is identified as the problem area when the onset of the patient’s symptoms is associated with the death of a loved one, either recent or past. Grief, however, may not be limited to the physical death of a loved one. Grief can also result from the loss of a significant relationship or the loss of an important aspect of one’s identity. The goals for treating complicated bereavement include facilitating mourning and helping the patient to find new activities and relationships to substitute for the loss. Reconstructing the relationship—both its positive and its negative aspects—is central to the assessment of not only what has been lost, but also what is needed to counter the idealization that commonly occurs. As patients become less focused on the past, they should be encouraged to consider new ways of becoming more involved with others and establishing interests.

Role Transitions. Role transition (which can overlap with grief) includes any difficulty resulting from a change in life status. Common role transitions include a career change (promotion, firing, retirement), a family change (marriage, divorce, birth of a child), the beginning or end of an important relationship, a move, graduation from school, or diagnosis of medical illness. The goals of therapy include mourning and accepting the loss of the old role, recognizing the positive and negative aspects of both the old and new roles, and restoring the patient’s self-esteem by developing a sense of mastery in the new role. Key strategies in achieving these goals include thoroughly exploring the patient’s feelings related to the role change and encouraging the patient to develop new skills and adequate social support for the new role.

Interpersonal Role Disputes. Such disputes are conflicts with a significant other (e.g., a partner, other family member, coworker, or close friend) that emerge from differences in expectations about the relationship. The goals of treatment include clearly identifying the nature of the dispute and exploring options to resolve it. It is important to determine the stage of the dispute; once the stage of the dispute becomes clear, it may be important to modify the patient’s expectations and remedy faulty communication to bring about adequate resolution. It may be particularly helpful to explore how nonreciprocal role expectations relate to the dispute. If resolution is impossible, the therapist assists the patient in dissolving the relationship and mourning its loss.

Interpersonal Deficits. Interpersonal deficits refer to patients who are socially isolated or who are in chronically unfulfilling relationships. The goal is to reduce the patient’s social isolation by helping them to enhance the quality of existing relationships and encouraging the formation of new relationships. To help these patients, it is necessary to determine why they have difficulty forming or maintaining relationships. Conducting a review of past significant relationships will be particularly useful in making this assessment. During this review, attention should be given to both the positive and negative aspects of each relationship, as well as to an investigation of potentially recurrent patterns in these relationships. It may also be appropriate to examine the nature of the patient–therapist relationship because this may be the patient's only close relationship, and it can be observed and commented on.

FACILITATE PATIENT'S WORK ON IDENTIFIED GOALS. In structuring the intermediate phase, the IPT takes a moderate position between the extremes of being highly active and merely reactive to the patient's concerns. The therapist actively guides the patient to cover material that is relevant to the treatment goals. If the patient does not bring in material, the therapist may elicit an update or more detailed information in one of the agreed-on problem areas and goals. The need for the patient to change is stressed at regular intervals. It is important to note that this constitutes a general encouragement to change rather than pressure to take a specific course of action. Meetings during the intermediate phase should not pass without reference to the interpersonal problem areas and attendant goals. Phrases such as "moving forward on your goals" and "making important changes" are used to encourage patients to be responsible for their treatment while reminding them that altering interpersonal patterns requires attention and persistence. In sessions, unfocused conversations are redirected to central themes of treatment, and abstract and vague discussions are minimized to maintain focus. Therapists avoid asking questions that elicit general or passive responses, such as general inquiries about the patient's week. Instead, sessions open with questions such as "What would you like to work on today?" and "How have things been since we last met?" These questions keep the patients focused on recent interpersonal events and psychiatric symptoms, which the therapist attempts to link. Above all, unfocused, unproductive complaining about others should be discouraged, in which case it should be pointed out that neither the patient nor the therapist is likely to change someone else in the patient's life.

A critical goal in the intermediate phase is to facilitate and help to strengthen the patient's connections between their psychiatric symptoms and the difficulties they have in their interpersonal lives. As the patient continues making these connections and develop strategies to alter the interpersonal context in which his or her symptoms occur, the patient will disrupt the cycle. The patient is encouraged to explore the problem areas and consider ways of changing; the attempts to change then become the focus of subsequent sessions. The therapist helps the patient to remain focused by ensuring that the subject matter is relevant and by providing clarification when needed. Clarification takes the form of pointing out themes and inconsistencies or highlighting points that the patient might miss. Clarification does not extend to making interpretations, in which reference is made to a theoretical view of the disorder and its treatment. Throughout, the focus remains on the present.

The behavior of the patient toward the therapist is expected sooner or later to reveal the patient's well-practiced behavioral patterns toward significant others. These may become more obvious during the intermittent phase. It may be useful for the therapist to comment on them.

By the end of the intermediate phase, patients are often acutely aware that treatment will soon be ending. The therapist begins to discuss termination explicitly and addresses any anxiety that the patient may be experiencing regarding it. The therapist also begins to prepare the patient for emotions that may arise with termination, including grief related to the ending of treatment.

Termination Phase. In the termination phase (usually four to five sessions in length), the therapist discusses termination explicitly with the patient and assists them in understanding that the end of the treatment is a potential time of grief. During this phase, patients are encouraged to describe specific changes in their psychiatric symptoms, especially as they relate to improvements in the identified problem area(s). The therapist also assists the patient in evaluating and consolidating gains, detailing plans for maintaining improvements in the identified interpersonal problem area(s), and outlining remaining work for the patient to continue on his or her own. Patients are also encouraged to identify early warning signs of symptom recurrence and identify plans of action.

DISCUSS TERMINATION EXPLICITLY. The therapist systematically raises the issue of reactions to impending termination in each of the last several sessions. Introducing the idea that termination is an explicit stage will plant the seed that it is an important topic to discuss. Therapists always ask patients how they feel about the ending of treatment—not least because this provides an opportunity to emphasize what has been achieved and to stress the patient's competence in dealing with future areas of difficulty.

DISCUSS TERMINATION AS A POTENTIAL TIME OF GRIEF. Patient reactions to the conclusion of treatments are often varied. As therapy nears completion, and sometimes well before that, patients may develop anxiety about saying goodbye and going it alone. Because termination marks the end of a connection to the therapist, it has a theme of loss, an analog of grief. It is important to state this possibility explicitly because unacknowledged sad feelings may lead to fears of relapse and an increase in symptoms.

REVIEW PROGRESS. An important aspect of the termination phase is encouraging patients to talk about the progress they have made and giving feedback about the changes that have been seen. This review helps to consolidate the work that has been done. It is not uncommon for individuals to attribute changes in treatment to the therapist rather than to their own hard work. Misplaced credit may erode the patient's confidence in the ability for continued success and improvement without treatment. Therefore, the therapist needs to emphasize how the patient has begun to successfully manage his or her outside relationships.

ENCOURAGE MAINTENANCE OF THERAPEUTIC GAINS. Patients are encouraged to maintain the interpersonal behaviors that have led to changes in their psychiatric symptom profile, and therapists work to develop a plan for continued work on identified goals. However, there are always goals that are not accomplished within the time frame of IPT, and patients need to be aware of this. Indeed, future difficulties can be expected, and the therapist needs to cultivate a discussion of how to handle them. Patients should also be encouraged to view their psychiatric disorder as vulnerability or an "Achilles' heel," in the sense that it may recur at future times of difficulty. Guiding the patient in a discussion of strategies for managing future problems will bolster feelings of competence. It is vital to assist patients in thinking about warning signs and symptoms that suggest a need for intervention.

General Therapeutic Techniques

Throughout therapy, the therapist should maintain a consistent focus on the interpersonal context of the patient's life. Symptoms should thus be analyzed in light of the interpersonal difficulties constituting the backdrop of those symptoms. Although this approach is unique to the interpersonal psychotherapies, some of the therapeutic techniques used in IPT are similar to techniques used in other therapies. These include exploratory questions, encouragement of affect, clarification, communication analysis, and the use of the therapeutic relationship.

Exploratory Questions. The use of general, open-ended questions can facilitate the free discussion of material. This is particularly true in the initial phases of a session. For example, the therapist might say, “Tell me about your relationship with your husband.” Once this has generated discussion, progressively more specific questioning will ensue.

Encouraging Affect. IPT focuses on affect evocation and exploration. Specifically, the therapist helps patients to (1) acknowledge and accept painful affects—as the feelings are expressed, it is important for the therapist to validate and help the patient to accept them; (2) use affective experiences to bring about desired interpersonal changes—the goal is to help the patient to act more constructively in interpersonal relationships, and this may involve learning when his or her needs are met by expressing affect and when they are better met by suppressing affect; and (3) experience suppressed affects—frequently patients are emotionally constricted in situations in which others would typically experience strong emotions. For example, sometimes patients will deny feeling upset when it is clear that an upsetting interaction has just occurred. In this situation, the therapist might say, “Although you said you were not upset, it appears to me that you have shut down since mentioning the situation with your husband.” By explicitly noting the discrepancy, the therapist attempts to draw out the affect that has been suppressed.

Clarification. This technique is useful in calling attention to contradictions that may have occurred in the patient’s presentation of material and increases the patient’s awareness about what they has actually communicated. For example, contradictions between the patient’s affect and speech may be noteworthy (e.g., “While you were telling me how upset you are about your father, you had a smile on your face. What do you think that’s about?”).

Communication Analysis. This technique is used to identify any communication difficulties that the patient may be experiencing and to help the patient change ineffective communication patterns. Typically, the therapist will ask the patient to recall in great detail an important recent interaction or argument with a significant other. Together, patient and therapist work to identify any communication difficulties and find more effective communication strategies.

Use of Therapeutic Relationship. The premise behind this technique is that people have characteristic patterns of interacting with others. Use of the therapeutic relationship in IPT is less routine. If appropriate, however, the patient's thoughts, feelings, expectations, and behavior in the therapeutic relationship are explored and related to the patient's characteristic way of behaving or feeling in other relationships. In IPT it is particularly relevant and useful for patients with interpersonal deficits (the patient develops a relationship with the therapist as a model for other relationships) and interpersonal role disputes (the patient receives feedback on how they come across and thereby has the opportunity to understand the nature of his or her difficulties in relating with others).

INTERPERSONAL PSYCHOTHERAPY DELIVERED IN A GROUP FORMAT

A relatively recent approach in the ongoing development of IPT has been its use in a group format. Interest in this area emerged from the successful adaptation by Denise Wilfley and colleagues of two individually-based treatments—cognitive-behavioral therapy and IPT—to a group format for the treatment of overweight individuals with binge-eating disorder. Subsequently, Wilfley and colleagues published a book that details the procedures for implementing IPT in a group format. Numerous investigators are using group IPT to treat other patient populations and psychiatric disorders.

IPT delivered in a group format has many potential benefits in comparison to individual treatment. For example, a group format in which membership is based on diagnostic similarity (e.g., depression, social phobia, eating disorders) can help to alleviate a patient's concerns that they are the only one with a particular psychiatric disorder while offering a social environment for patients who have become isolated, withdrawn, or disconnected from others. Given the number and different types of interpersonal interactions in a group setting, the interpersonal skills that are developed may be more readily transferable to the patient's outside social life than the relationship patterns that are addressed in a one-on-one setting. Moreover, a group modality has therapeutic features not present in individual psychotherapy (e.g., interpersonal learning). The group format also facilitates the identification of problems common to many patients and provides a cost-effective alternative to individual treatment.

Techniques

Timeline and Structure of Treatment. The typical course of group IPT lasts 20 sessions over a 5-month period. It is recommended that group size ranges from six to nine members, with one or two group leaders, depending on resources and training needs. In contrast to the initial phase of individual IPT, in group IPT, the assignment of the sick role, the interpersonal inventory, the development of the problem area(s), and the development of treatment goals occur in a structured, individual, 2-hour pregroup meeting. The initial phase (sessions 1 through 5) of group IPT is then devoted to the development of positive group norms as group members test out and refine their individual target problem area(s). The intermediate phase (sessions 6 through 15) of group IPT is dedicated to group members' work on their identified problem area(s). At the midpoint of the intermediate phase, a 1-hour individual meeting is held between therapist(s) and each group member. The termination phase (sessions 16 through 20) of group IPT is devoted to the consolidation of gains made by group members over the course of treatment and the preparation of the group members for termination of the group and for future work on their own. Within 1 week of the final session of the termination phase, a 1-hour post-group meeting is held between therapist(s) and each group member. The three individual meetings (pre-, mid-, and postgroup), sequenced to correspond with critical time points in the three phases of IPT, in combination with other techniques, were designed to maintain the exclusive and strategic focus on individual patients' interpersonal problem areas—the hallmark of IPT.

Therapeutic Stance. Similar to individual IPT, the therapeutic stance of the group interpersonal psychotherapist is one of warmth, support, and empathy. The therapist is active and advocates for the group members rather than remaining neutral. The open-ended group approach used in group IPT, with its emphasis on fostering direct communication between group members, is also highly compatible with the therapeutic stance, goals, and structure of individual IPT. The approach provides a framework for facilitating interaction among group members and emphasizes the establishment of a safe environment in which patients can receive feedback on how they are perceived by others and can experiment with new and risky interactional styles. Successful implementation of an open-ended group approach requires specific training for the therapist and attention to important developmental processes.

Pregroup Meeting. The pretreatment meeting is crucial for facilitating a patient's individualized work in the first phase of group IPT. In individual IPT, the first five sessions are devoted to a detailed examination of the patient's interpersonal history (i.e., an interpersonal inventory) and the formulation of problem areas and goals, which will guide the therapeutic work. In these critical sessions, the patient is also oriented in a more general way to the work of psychotherapy and is educated about the nature of his or her disorder and the means by which the work is used to bring about a recovery. Although it is possible to undertake individual preparation in the presence of other group members, such an approach would simply not be an efficient or therapeutically valuable use of group time. Thus, in group IPT, the 2-hour pretreatment meeting is used to focus on identifying interpersonal problem areas, establishing an explicit treatment contract to work on problem areas, and preparing patients for group treatment. After identifying a patient's interpersonal problem(s) (interpersonal deficits, role disputes, role transitions, or grief), the therapist works collaboratively with the patient to formulate concrete prescriptions for change, in addition to the specific steps the patient will take to improve social relationships and patterns of relating. These goals of treatment are expressed in language that is as specific and personally meaningful to the patient as possible. Before the start of the group, each group member is given a written summary of his or her goals and told that these goals will guide his or her work in the group.

Another important element of the pregroup meeting involves adequately preparing patients for group treatment. That is, patients are encouraged to think of the group as an "interpersonal laboratory" in which they can experiment with new approaches to handle challenging interpersonal situations. In this regard, patients are informed about the important interpersonal skills that are learned while participating in a group (e.g., interpersonal confrontation, honest communication, expression of feelings) and are encouraged to learn from others as they see changes occur. The therapist stresses to patients the importance of keeping their work in the group focused on changing their current interpersonal situations or intensifying important existing relationships and not using the group as a substitute social network.

Initial Phase. The first five sessions of group treatment comprise the initial phase of group IPT. During this phase, the therapist works to cultivate positive group norms and group cohesion while emphasizing the commonality of symptoms among members and how they will be addressed in the group context. During this phase, group members are encouraged to review their goals with the group and begin to make some initial changes in their respective interpersonal problem areas. As members begin to experiment with the changes outlined in their goals, the therapist works collaboratively with each group member to refine and make any alterations in the target areas before the beginning of the intermediate phase.

Intermediate Phase. During the intermediate “work” phase of group IPT (sessions 6 through 15), the therapist works to facilitate connections among members as they share the work on their goals with one another. In contrast to open-ended group approaches, the group interpersonal psychotherapist is much less likely to focus on intragroup processes and relationships unless they are specific to the work on the member’s interpersonal problem area (e.g., interpersonal deficits). However, the therapist consistently and continuously encourages group members to practice newly acquired interpersonal skills both inside and, most important, outside of the group. As is the case with individual IPT, an essential task throughout the intermediate phase is to strengthen the connections that the group members make between difficulties in their interpersonal lives and their psychiatric problems. As with individual IPT during the intermediate phase of treatment, the interpersonal psychotherapist implements treatment strategies that are specific to the four problem areas as specified by Weissman and colleagues (Table 32.10–2).

MIDTREATMENT MEETING. The midtreatment meeting is held midway through the intermediate phase (usually between sessions 10 and 11). This meeting provides an opportunity to conduct a detailed review of each group member’s progress on his or her individual problems and to refine interpersonal goals. The therapist(s) recontract with group members during this meeting as a means of outlining and emphasizing the work that remain, both inside and outside of the group, before the conclusion of treatment.

Termination Phase. In the termination phase (sessions 16 through 20), the therapist discusses termination explicitly with the group members and begins to help them to recognize that the end of treatment is a time of possible grief and loss. The therapist helps members to recognize their own progress and the progress made by other group members. During this phase, group members are encouraged to describe the specific changes in their psychiatric symptoms, especially as they relate to improvements in the identified problem area(s) and relationships. Although it is common for group members to want to keep meeting on their own or to have frequent reunions, group members are encouraged to use this phase of the group to formally say goodbye to one another and to the therapist(s). The therapist(s) also uses this time to encourage members to detail their plans for maintaining improvements in their identified interpersonal problem area(s) and to outline their remaining work. In the final session, a specific structure is established to allow each group member to give feedback to other members.

Table 32.10–3.

Linking the Phases of Interpersonal Psychotherapy to the Stages of Group Development

Phases/Tasks	Group Stages	Members’ Work	Therapist Interventions
Initial: Sessions 1–5; identify problem areas	Engagement: Sessions 1–2	Members look for structure as they grapple with the anxiety of being in a group and sharing their problems	Establish a structure that encourages appropriate self-disclosure; facilitate norms for effective communication
	Differentiations: Sessions 3–5	Members work to manage negative feelings over interpersonal differences as they emerge in the group	Help members to understand their reactions in the context of interpersonal differences in their outside social lives
Middle: Sessions 6–15; work on goals	Work: Sessions 6–15	Members work out differences and strive toward common goals	Facilitate connections among members as they share their work with each other; encourage practice of newly acquired interpersonal skills in and outside of the group
Final: Sessions 15–20; consolidate treatment	Terminations: Sessions 16–20	Members struggle with how to manage the impending loss of	Help members to consolidate their work and to plan continued

		connection with other group members	work; assist members in grieving the loss of the group

From Wilfley DE. *Interpersonal Psychotherapy for Group*. Basic Books; 2000.

POSTTREATMENT MEETING. The posttreatment meeting is scheduled within 1 week after the final group session. The therapist(s) use this final individual meeting to develop an individualized plan for each group member's continued work on his or her interpersonal problem area(s). The therapist(s) reviews the group experience and the changes the patient has made in his or her interpersonal problem area(s) and significant relationships.

Maintaining the Interpersonal Focus in Group Interpersonal Psychotherapy

Use of Group Summaries. In contrast to individual IPT, in which patients have the undivided attention of the therapist, in the early phases of group formation, patients may feel overlooked or confused about how others' interpersonal problems relate to their own. One way to address this issue is to provide weekly group summaries to group members. These four- to five-page summaries focus on the transactions that occurred during each session and their implications for each patient's recovery. Summaries include information about the group process, the common interpersonal problems of those in the group (e.g., difficulty expressing feelings, excessive caretaking of others to the detriment of meeting their own needs, conflict avoidance), and how to use the group most effectively (e.g., how to use feedback and learn from other members' work on problem areas). Wilfley and her colleagues used group summaries in their treatment study for binge-eating disorder. However, writing group summaries can be quite time intensive. Therefore, one alternative to the use of group summaries may be the use of self-help manuals. With this approach, all group members (assuming diagnostic similarity among the group membership) would receive the self-help manuals and be instructed to read them over the course of treatment. For example, Weissman developed self-help manuals for mastering depression that would be useful companions to an IPT group for depressed patients.

Use of Interventions That Are Informed by Group Stage Development Theory. As discussed earlier, maintaining the important individual focus in group IPT also requires attention to important developmental processes noted in other time-limited groups. The developmental sequence of a group merges naturally, but successful movement from one stage to another can be disrupted by an ineffectual leadership style and ill-informed intervention strategies. At the very least, a therapist unaware of group development may use interventions that lead to stagnation, and, at the worst, they may use strategies that are detrimental to an individual's progress. Therefore, to manage the group effectively while maintaining the interpersonal focus required in group IPT, it is essential that therapists understand the stages of group development. Attending to these normative transitions in the group intensifies group cohesiveness, prevents premature dropouts, and assists patients in making otherwise difficult interpersonal transitions. In addition, therapists who can predict critical transition points in the group and make stage-appropriate interventions are more likely to facilitate the work of the group and its members (Table 32.10–3).

CLINICAL ISSUES

The following case example illustrates the individual treatment of a 51-year-old woman with binge-eating disorder. This vignette highlights the therapist's presentation to the patient of the interpersonal formulation linking the onset of her eating disorder to one of the four interpersonal problem areas. As discussed earlier, the therapist's ability to discern patterns in interpersonal relationships rapidly, connect events with the onset and maintenance of the disorder, and formulate goals is crucial to the time-limited nature of IPT. In addition, this case illustrates several of the strategies and techniques used in treating the interpersonal problem area.

G, a 51-year-old woman, presented for treatment of binge-eating disorder. She was college educated, had her own business, and was a divorced mother of one adult son in his early 20s. Before treatment, G had a body mass index (BMI) of 42 and had been binge eating approximately 10 to 15 days per month for the last 8 years. Along with her current diagnosis of binge-eating disorder, G struggled with recurrent major depression.

During the initial phase, G and her therapist began to review her history and the interpersonal events that were associated with her binge eating. G shared that she began overeating and gaining weight at the age of 14 years. When she was 18 years old, she moved to another country with her parents. Soon after the move, G's father left her and her mother to return to the United States. G was enraged at her father for leaving them and would still get very tearful and angry when discussing the separation. G and her mother decided to stay abroad because G had started university and her mother was working. Both had developed strong social ties and felt comfortable in their new home. During this time, G continued to gain weight and started dieting. Shortly after graduating from university, G met and married a foreign national and, at the age of 28 years, delivered their only son. Two years later, G and her husband went through a very bitter divorce. G described this as a terrible time in her life. She maintained close ties with her friends and her mother. During this time, she began to diet and reached her lowest adult weight. At the age of 35 years, when her mother died of a heart condition, G had her first episode of major depression, which was treated and resolved with antidepressants and a brief course of psychotherapy. Although G had previous cycles of weight loss and weight regain, she did not evidence any sign of eating disturbance at this point. She continued to maintain close social ties and enjoyed her close relationship with her son. When G was in her early 40s, an economic downturn in her adopted country forced her to return to the United States. Having lost all of her savings, G struggled financially while she looked for work. During this time, she started binge eating and gaining weight. Within 1 year of this move, G's son decided to return to live with his father (who was very wealthy). G felt angry and betrayed. Yet, when her son would visit, she would assume a subservient role with him because she was afraid of losing his affection. He, in turn, became quite demanding and critical of her.

Before seeking treatment, her heightened feelings of isolation and loneliness were leading to increased binge eating, depression, and weight gain.

By session three of the initial phase, G's therapist began to consider which problem area would be the focus of the remainder of treatment. G had a history of important relationship losses and subsequent grief—the loss of her father, her husband, her mother, and, most recently, her son. However, none of these losses was associated with the development of binge-eating problems (although her dieting was clearly linked to her feelings of anger after the divorce from her husband, and her depression was intimately linked with her mother's death). G's anger at her son for returning to live with the enemy was clearly a role dispute, yet her binge eating had begun 2 years before his departure (although it clearly worsened after he left). Because neither of these problem areas was directly linked to the onset of the eating disorder, G's therapist decided that the focus of treatment would be to assist G in managing her role transition. Her move back to the United States, with the subsequent loss of her support and friendship networks, was clearly associated with the onset and continued maintenance of her binge eating. During the fourth session of the initial phase, G's therapist shared her formulation of the problem area with her:

"From what you have described, your binge eating really began after you returned to the United States. After that transition, you were more isolated and alone than you have ever been. It seems that binge eating was a way for you to manage that transition and the subsequent feelings of isolation and loneliness. Your transition has also had a negative impact on your relationship with your son. Even though you are a very social person and enjoy the company of others, you have yet to develop the kind of support that you had before you moved. Although you have struggled with some very significant issues over the course of your life—your father's leaving, the pain of the divorce, and the death of your mother—your friends and support systems sustained you. If we work together to help you find and develop more intimate and supportive relationships here, I believe you will be much less likely to turn to food and binge eating as a source of support or comfort."

G agreed with the formulation and worked with her therapist to establish some treatment goals to help her resolve the problem area. First, she was encouraged to become more aware of her feelings (especially isolation and loneliness) when she was binge eating and of how binge eating seemed to be the way she managed those feelings. A second goal was for G to take steps to increase her social contacts and develop more friendships. The third goal, which was identified as a secondary problem area, centered on helping G to resolve the role dispute with her son. Specifically, the therapist developed a goal with G to help her to establish a clearer parental role with her son. During the intermediate phase, the therapist helped G to grieve the loss of her previous role and the extensive support that she once had. G and her therapist worked to identify several sources of support and friendships of which she had not been aware. Soon after, G reported significant progress in initiating and establishing relationships with others. This change appeared to help give her confidence in her new roles. In fact, she had begun to receive a few social invitations. She was more attuned to the ways that she would rely on food, especially when she felt lonely or felt that she was not receiving enough time from others. The connection between the lack of supportive contacts and binge eating was becoming very clear to her in these intermediate sessions. During this phase, the therapist also assisted G in setting appropriate limits in her relationship with her adult son and in recognizing his adult-like responses in return.

By the termination phase, G reported that she no longer felt so lonely and isolated and that her binge eating had all but disappeared. She remarked how the quality of her relationship with her son had changed dramatically. He was more supportive and respectful, visited more frequently, and stayed with her for longer periods of time. In the final sessions, she talked about her need to let go of the past and move on with her life as it was at that point, assuming her new roles more fully. She worked closely with her therapist to develop a plan to maintain the gains that she had made in treatment and used the final session to review the important work that she had accomplished.

Indications

Indications for time-limited–oriented IPT have been defined by several efficacy trials. IPT has been shown to be effective in the treatment of mood disorders and eating disorders (bulimia nervosa and binge-eating disorder). IPT is also described by the American Psychiatric Association's (APA) *Practice Guidelines for Major Depressive Disorder in Adults* as useful for patients who face conflicts with significant others or who are having difficulty adjusting to a life transition.

Limitations

IPT has not been shown to be effective in treating substance abuse. The literature on the treatment of anxiety with IPT is evolving.

Complications

IPT is a very straightforward treatment that is highly acceptable to patients. In fact, very few, if any, complications from IPT have been noted. However, after a full course of IPT, treatment nonresponders should be evaluated for the possible prescription of an alternative, evidence-based treatment.

Contraindications

IPT was not developed to treat psychotic depression or other forms of psychosis, such as schizophrenia.

Goals of Treatment

As indicated earlier, the goal of IPT is to reduce or eliminate psychiatric symptoms by improving the quality of the patient's current interpersonal relations and social functioning.

ETHICAL ISSUES

To maintain high ethical standards, therapists wishing to conduct IPT should seek training and supervision to develop proficiency. Specific IPT training guidelines are available in the comprehensive text by Weissman and colleagues, and specific group training guidelines are available in the group text by Wilfley and colleagues. In general, interpersonal psychotherapists need to be mindful of imposing their goals and values onto their patients (e.g., pushing a patient to end a relationship). As with any psychological treatment, it is critical to thoroughly evaluate (medically and psychiatrically) all patients entering treatment to determine the best evidence-based treatment and most prudent course, including the range of treatment options.

RESEARCH AND EVALUATION

Outcome Research

IPT has been found efficacious for major depression and has also been successfully adapted to treat other types of mood and nonmood disorders.

IPT has demonstrated efficacy in the treatment of several disorders, including nonpsychotic major depression, recurrent depression, bipolar disorder, postpartum depression, dysthymic disorder, bulimia nervosa, and binge-eating disorder. In addition, a number of new research applications are under investigation for various disorders, some of which include traumatic grief, social phobia, body dysmorphic disorder, chronic somatization, borderline personality disorder, and anorexia nervosa.

IPT has been adapted for various populations, including older adults, posttraumatic stress disorder (PTSD) patients, adolescents, and patients with comorbid medical conditions (e.g., human immunodeficiency virus [HIV] infection and cancer). IPT also has been modified for long-term treatment, groups, couples, telephone intervention, primary care, and adjunctive self-help, as well as been translated into several languages. Together, these findings indicate that IPT is not only an effective and promising treatment for a range of psychiatric disorders, but also has the potential for wide dissemination.

Research Using Group Interpersonal Psychotherapy

Group IPT has recently become a topic of various research interests. It has been used to treat depression by K. Roy MacKenzie and colleagues, postpartum depression by Claudia Klier and colleagues, and binge-eating disorder by Wilfley and colleagues. Studies are underway to look at group IPT for depressed adolescents, depressed older adults, and individuals with eating disorders, PTSD, and social phobia.

FUTURE DIRECTIONS

There are a number of important areas within the study and practice of IPT that warrant further consideration. First, although a number of outcome studies in IPT document its efficacy, little is known about the mechanisms by which IPT exerts its effects. How much is specific; how much is nonspecific, held in common with other psychotherapies? A greater understanding of the mechanisms by which change occurs in IPT would assist in further refinements of the treatment and yield insights into the nature of the psychiatric syndrome under investigation. Second, increased efforts to improve the effectiveness of IPT are warranted. Timing, duration, frequency of the sessions, and identifying specific therapist behaviors and patient responses associated with better outcome could be studied. Third, refining the definitions of the target populations is advisable, which would define those for whom IPT is most indicated or the treatment of choice. Fourth, there exists a need to translate IPT efficacy data into effectiveness studies and appropriate clinical practice. The decade and more has seen the expansion of applications of IPT to a wider circle of situations, many needing replication. It has also seen continued expansion beyond the English-speaking world.

FURTHER READINGS

Agras WS, Walsh T, Fairburn CG, Wilson GT, Kraemer HC. A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. *Arch Gen Psychiatry*. 2000;57:459–466.

Anchin JC, Kiesler DJ. *Handbook of Interpersonal Psychotherapy*. Pergamon Press; 1982.

Bernecker SL, Coyne AE, Constantino MJ, Ravitz P. For whom does interpersonal psychotherapy work? A systematic review. *Clin Psychol Review*. 2017;56;82–93.

Cuijpers P, Donker T, Weissman M, Ravitz P, Criztea LA. Interpersonal psychotherapy for mental health problems: a comprehensive meta-analysis. *Am J Psychiatry*. 2016;173;680–687.

Frank E, Spanier C. Interpersonal psychotherapy for depression: overview, clinical efficacy and future directions. *Clin Psychol Sci Pract*. 1995;2:349–369.

Heisel MJ, Talbot NL, King DH, Tu XM, Duberstein PR. Adapting interpersonal psychotherapy for older adults at risk for suicide. *Am J Geriatric Psychiatry*. 2015;23(1):87–98.

Kiesler DJ. *Contemporary Interpersonal Theory of Research: Personality, Psychopathology, and Psychotherapy*. Wiley; 1996.

Kiesler DJ, Strupp H. *The Process of Psychotherapy: Empirical Foundations and Systems of Analysis*. Aldine Transaction; 2006.

Klier CM, Muzik M, Rosenblum KL, Lenz G. Interpersonal psychotherapy adapted for the group setting in the treatment of postpartum depression. *J Psychother Pract Res*.

2001;10:124–131.

Lenze EJ, Dew MA, Mazumdar S, et al. Combined pharmacotherapy and psychotherapy as maintenance treatment for late-life depression: effects on social adjustment. *Am J Psychiatry*. 2002;159:466–468.

Lillie R. Getting clients to hear: applying principles and techniques of Kiesler's interpersonal communication therapy to assessment feedback. *Psychol Psychother*. 2007;80:151–163.

Lipsitz JD, Markowitz JC. Mechanisms of change in interpersonal therapy (IPT). *Clin Psychol Rev*. 2013;33:1134–1147.

MacKenzie KR, Grabovac AD. Interpersonal psychotherapy group (interpersonal psychotherapy-G) for depression. *J Psychother Pract Res*. 2001;10:46–51.

Markowitz JC. *Interpersonal Psychotherapy for Dysthymic Disorder*. American Psychiatric Press; 1999.

Markowitz JC. Interpersonal psychotherapy for chronic depression. *J Clin Psychol*. 2003;59:847–858.

Miller MD, Frank E, Cornes C, Houck PR, Reynolds CF 3rd. The value of maintenance interpersonal psychotherapy (IPT) in older adults with different IPT foci. *Am J Geriatr Psychiatry*. 2003;11:97–102.

Miniati M, Callari A, Calugi S, et al. Interpersonal psychotherapy for postpartum depression: a systematic review. *Arch Womens Ment Health*. 2014;17:257–268.

Mufson L, Sills R. Interpersonal psychotherapy for depressed adolescents (IPT-A): an overview. *Nord J Psychiatry*. 2006;60:431–437.

O'Hara MW, Stuart S, Gorman LL, Wenzel A. Efficacy of interpersonal psychotherapy for postpartum depression. *Arch Gen Psychiatry*. 2000;57:1039–1045.

Orgeta V, Qazi A, Spector AE, Orrell M. Psychological treatments for depression and anxiety in dementia and mild cognitive impairment. *Cochrane Database Syst Rev*. 2014;(1):CD009125.

Parker G, Fletcher K. Treating depression with evidence-based psychotherapies: a critique of the evidence. *Acta Psychiatr Scand*. 2007;115:352–359.

Reynolds CF 3rd, Miller MD, Pasternak RE, et al. Treatment of bereavement-related major depressive episodes later in life: a controlled study of acute and continuation treatment with nortriptyline and interpersonal psychotherapy. *Am J Psychiatry*. 1999;156:202–208.

Scocco P, Toffol E. Group interpersonal psychiatry: a review. *Sante Ment Que*. 2008; 33(2):105–131.

Van Bronswijk SC, Lemmens LHJM, Keefe JR, Huibers MJH, DeRubies RJ, Peeters FPML. A prognostic index for long-term outcome after successful acute phase cognitive therapy and interpersonal psychotherapy for major depressive disorder. *Depress Anxiety*. 2019;252–261.

Weissman MM. Interpersonal psychotherapy: history and future. *Am J Psychother*. 2020; 73;3–7.

Weissman MM. Recent non-medication trials of interpersonal psychotherapy for depression. *Int J Neuropsychopharmacol*. 2007;10:117–122.

Whiston A, Bockting CLH, Semkovska M. Towards personalizing treatment: a systematic review and meta-analysis of face-to-face efficacy moderators of cognitive-behavioral therapy and interpersonal psychotherapy for major depression. *Psychol Med*. 2019;49;2657–2668.

Wilfley DE, MacKenzie KR, Welch RR, Ayers VE, Weissman MM. *Interpersonal Psychotherapy for Group*. Basic Books; 2000.

Wilfley DE, Welch RR, Stein RI, et al. A randomized comparison of group cognitive-behavioral therapy and group interpersonal psychotherapy for the treatment of overweight individuals with binge-eating disorder. *Arch Gen Psychiatry*. 2002;59:713–721.

Zlotnick C, Johnson SL, Miller IW, Pearlstein T, Howard M. Postpartum depression in women receiving public assistance: pilot study of an interpersonal-therapy-oriented group intervention. *Am J Psychiatry*. 2001;158:638–640.

32.11 Dialectical Behavior Therapy

M. Zachary Rosenthal, Ph.D., Jacqueline Trumbull, B.A., Ph.D., and Rachel Guetta, B.A.

DEFINITION

Dialectical behavior therapy (DBT) is a multimodal, cognitive-behavioral treatment originally developed to treat women who meet the criteria for BPD with a history of chronic self-harm and suicide attempts. BPD is a severe, costly, and difficult-to-treat psychiatric disorder characterized by problems across interpersonal, behavioral, cognitive, and emotional domains of functioning. Most individuals diagnosed with BPD report high rates of nonsuicidal self-injury (NSSI) and multiple suicide attempts. The lifetime mortality rates by suicide are estimated to be 50 times higher among BPD patients compared to the general population. Individuals with BPD require more mental health services and evidence higher rates of treatment dropout and relapse compared to patients with other psychological disorders. Moreover, BPD is associated with therapist burnout and is one of the most stigmatized psychological disorders. According to the American Psychiatric Association (APA), there is no single efficacious pharmacologic intervention for BPD. However, current practice guidelines recommend psychotherapy as the primary treatment for BPD, and, to date, DBT is the psychosocial treatment that has received the most empirical support for patients with BPD.

DBT was developed in the context of the broader rise of behavioral therapies and CBTs. Although behavioral and cognitive interventions are fundamental treatment strategies in DBT, this is a therapy based on the amalgamation of principles from diverse sources, including interpersonal systems, theories, and scientific research related to emotional functioning, and Eastern mindfulness and Western contemplative spiritual practices. A biosocial model featuring the reciprocally interactive effects of nature and nurture, along with a dialectical worldview, together influence case conceptualization and are woven into the transactional process between patients and therapists. As a result, DBT offers a comprehensive treatment approach that uses a wide range of interventions to create a compassionate and hopeful therapy for patients who have been historically considered “manipulative,” virtually untreatable, or even deliberately resistant.

Structurally, there are four primary modes of treatment in DBT: group skills training, individual therapy, telephone skills consultation, and a consultation team of DBT therapists. Preliminary data suggest that DBT skills training may hold promise as a stand-alone treatment for emotion dysregulation among a transdiagnostic sample. Still, there is inconclusive empirical evidence to determine which specific components and processes within DBT are necessary. Until further dismantling and component analysis studies have been completed, it is inappropriate to consider a treatment as DBT unless all of the four primary modes are being implemented. Accordingly, this chapter focuses on DBT as originally developed for the treatment of BPD. However, it also highlights several areas where DBT has been adapted for specific populations and mental health problems.

HISTORY

DBT was developed by Marsha Linehan at the University of Washington in the beginning of late 1970s. For years, she had used standard CBT strategies to work with suicidal and actively self-injurious patients. However, Linehan found that therapists’ unrelenting focus on trying to help patients change their problematic behaviors often led patients to feel invalidated. In turn, patients responded by expressing anger, ambivalence about engaging in treatment, and even threats of self-harm. Therapists working with patients diagnosed with BPD tended to report higher levels of burnout, in part, because of the sheer volume of severe and highly complex problems their patients presented during and between sessions and, in part, because they could not realistically address all of these problems in session and still have time to teach patients new skills.

Accordingly, Linehan and her team developed a multifaceted approach aimed at helping these high-risk patients who suffer tremendously from chronic and pervasive problems across many areas of their lives to create “a life worth living.” They adopted acceptance and validation strategies to balance the change-focused strategies of cognitive and behavioral therapies; they implemented skills training groups and phone coaching to ensure that patients learn new skills and generalize their learning into the naturalistic environment to cope with everyday stressors; and they added consultation team meetings in order to provide therapists with support, structured problem solving, and a context for ongoing learning and assessment of fidelity to the treatment model.

The first randomized controlled trial (RCT) for DBT was published in 1991. This study demonstrated the superiority of DBT over treatment as usual for women with BPD. Subsequently, Linehan published the DBT treatment manual and skills training manual in 1993. In her book, Linehan introduced a novel conceptualization of BPD, reframing it as a dysfunction of the emotion regulation system that can be addressed with the different components of DBT. When Linehan and her team published positive results for a second RCT in 1999, DBT became the first psychological treatment that could claim to be an evidence-based treatment for BPD, according to the criteria outlined by the APA Division 12 Task Force.

In order to facilitate the dissemination of DBT and the training of DBT clinicians around the world, Dr. Linehan established a nonprofit organization, Behavioral Tech LLC, in 2004. Around the same time, Linehan and her team developed a coding scale for rating individual and skills group sessions for adherence to the DBT treatment manual. In order to enhance adherence and competence, the DBT-Linehan Board of Certification has recently begun to offer DBT certification for individual therapists. To date, hundreds of clinicians in the United States have been certified through the DBT-Linehan Board of Certification. In 2014, the revised skills training manual was published, which features both a revision of many original DBT skills as well as the addition of select skills (e.g., the nightmare protocol for emotion regulation, walking the middle path skill for mindfulness, and new distress tolerance skills to be used) to be used with specific clinical problems (e.g., substance use).

As an empirically driven treatment, it is expected to evolve over time as a function of further research. To date, RCTs have been published from many investigators around the world examining applications of DBT for different problems (e.g., eating disorders, substance abuse problems, and depression with co-occurring personality disorders) and in different settings (e.g., outpatient private practices, inpatient psychiatric hospitals, and forensic settings). Indeed, in 2012, DBT was officially recognized by the Cochrane Review as the treatment of choice for BPD. In little over 20 years, DBT advanced from a novel behavioral therapy to a gold standard intervention for BPD.

THEORETICAL ISSUES

The basic elements of DBT are outlined here. For a comprehensive description of this treatment approach, interested readers are referred to the treatment manual and group skills training manual by Marsha Linehan. The following section includes an outline of the dialectical philosophy and biosocial theory that together bind the various treatment modes and strategies of DBT. Treatment techniques and a case example are then presented, followed by a discussion of ethical issues and empirical findings related to DBT.

Dialectics

For DBT, dialectics subsumes both a worldview and a process of change. The dialectical worldview refers to an ontologic perspective of interrelatedness and wholeness applied to the understanding of human suffering. Accordingly, behavior is conceptualized contextually and systemically, across private (e.g., thoughts, feelings, physiologic sensations) and publicly observable (e.g., overt motoric behavior) experiences. The patient as a whole is treated rather than a discrete disease, disorder, or symptom. Dialectical thinking provides a solution to the tension between formalistic universalistic thinking (there is absolute truth) and relativistic thinking (there are many “truths”). Dialectics adds a third way of thinking; there is both absolute truth, and truth that is contextual and always evolving. A core feature of dialectical thinking is understanding that for every point (a thesis), an opposite position (an antithesis) can be held while searching for a synthesis between these apparent contradictions.

Throughout the course of DBT, polarities are repeatedly identified in how the patient is thinking or behaving. For example, patients might express anxious thoughts in catastrophic but unrealistic terms or, at an overt behavioral level, might repeatedly try to reduce guilt by excessively apologizing. Through consideration of opposing or contradictory ways of thinking (e.g., the catastrophe will never happen) or behaving (e.g., perhaps no apology is necessary), natural tensions arise between two possible ways of thinking or behaving. The dialectical process of change promotes the resolution of apparently contradictory stances by yielding an integrated alternative, or a synthesis, that takes into account the fluidity of a constantly dynamic context. The catastrophe is highly unlikely to occur but could still happen; apologies are necessary in some situations but not all. As another example, a patient might insist that expressing anger in a certain way helps to improve mood in the moment, whereas the therapist might insist that this anger expression actually is the problem, resulting in a polarized stalemate. Dialectical thinking would help the therapist to create a synthesis with the patient by identifying how anger expression may be effective when done in some ways and in some contexts, and, that at the same time, anger expression in some contexts can exacerbate problems, not solve them. Working together, the therapist and the patient may look for ways to effectively express anger in some contexts and not express anger in others contexts to minimize the extent to which anger expression is an enduring problem. In other words, the synthesis would be to neither totally inhibit anger nor express anger, but inhibit or express, depending on the context. This sensitivity to context—seeking what is needed in any given moment and moving away from extreme or rigid ways of thinking or behaving—is a principle that lies within the core of interventions in DBT.

Not only does a dialectical approach reduce extreme ways of thinking or behaving, it also explicitly provides the therapist with a process designed to give up on always being right about every issue, without losing perspective. In addition, dialectics acknowledges that the patient influences the therapist while being influenced by the therapist. Dialectical strategies, because they can be implemented in a way that is not always perfectly predictable by the patient, also may keep the patient slightly off balance (e.g., lose the rigid conviction that their perception is the literal truth). This has advantages for patients learning new behaviors, in that the conditions under which adaptive ways of thinking or behaving in session are broad and not restricted to a narrow topographical therapist behavioral repertoire. In other words, the dialectical process is designed to promote broad repertoires of patient behavior and cognitions. For patients with BPD, this means less rigid, extreme, or otherwise problematic behavior, emotions, and patterns of thinking. Of importance, this also means more flexible ways of thinking and behaving that are sensitive to coupling the patient’s values and goals with the contingencies of the present context.

Although there are many dialectical tensions in DBT, the fundamental dialectic is that of *acceptance and change.* For the therapist, this entails balancing an acceptance of patients as they are in the moment with an explicit goal of meaningful change. This dialectic requires that patients and therapists accept the reality that painful emotions and thoughts exist (but not necessarily the literal truth of the content of thoughts) while simultaneously working to change dysfunctional and seemingly intractable problems. A perfectly balanced "middle-of-the-road" compromise between acceptance and change is not necessarily the goal. Instead, a synthesis of polarities may be more acceptance based in one moment and more change focused in another, depending on the context and what is likely to be most effective. This notion is akin to the need, when playing golf, to factor in the speed and direction of the wind when teeing off. The goal is to get the ball to land on the fairway, not to hit it in the air over the exact middle of the fairway; thus, the ball is sometimes hit to the left or to the right, factoring in the intensity and direction of the wind in the moment before the ball is struck.

Biosocial Theory of BPD

If dialectical thinking is the philosophical heart of DBT, then the biosocial theory of BPD forms the backbone. In articulating this theoretical model, Linehan suggested that BPD is fundamentally a disorder of the emotion regulation system and results from continuous reciprocal transactions among an invalidating environment, emotional vulnerability, and emotion dysregulation (Fig. 32.11–1). Prototypical examples of environmental invalidation are childhood sexual or physical abuse. Childhood sexual abuse may be a key etiologic feature of the disorder, with an estimated 40% to 70% of borderline individuals reporting a history of childhood sexual abuse. Childhood sexual abuse, however, does not appear to cause BPD. Indeed, meta-analytic data from 21 studies found only a moderate effect size (r = 0.28) in the association between BPD and childhood sexual abuse. As such, sexual abuse may be an important etiologic factor for some individuals with BPD, but the development and maintenance of BPD is complex and determined by a wide range of neurobiologically and socially mediated factors.

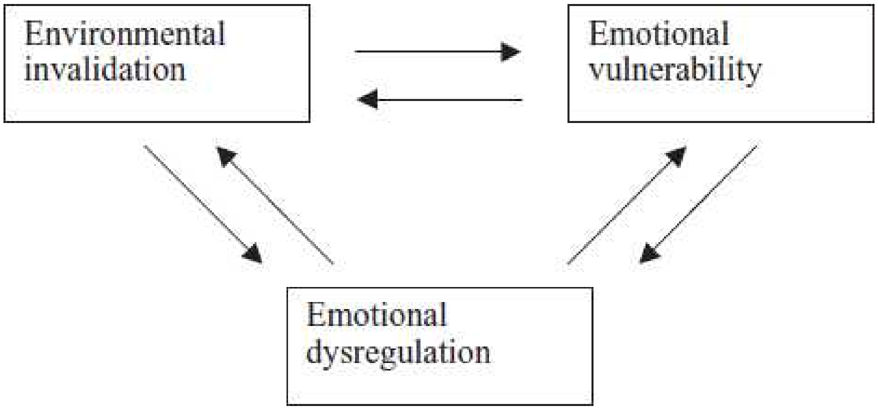

FIGURE 32.11–1. Components of the biosocial theory of borderline personality disorder.

The Invalidating Environment. According to Linehan's biosocial model, the emotion dysregulation characteristic of BPD commonly emerges among individuals with emotional vulnerabilities raised within (and often still living in) an invalidating environment. The invalidating environment is broadly characterized by pervasive criticizing, minimizing, trivializing, punishing, or erratically reinforcing communication of internal experiences (e.g., thoughts and emotions), and oversimplifying the ease of problem solving. For example, a parent may pervasively communicate "You're not that upset, you're just faking it" or "You say no, but you don't really know what you want." Verbal communication is indiscriminately rejected, and the individual may be chronically pathologized as having socially undesirable personality traits (e.g., too sensitive, paranoid, or lazy). This leads to a lack of trust in self, such that those with BPD learn that what they think, feel, and do is inappropriate, untrue, unimportant, or due to shameful personal flaws. Instead, an excessive reliance on the social environment for cues on how to think, feel, or behave develops, resulting in a hypersensitivity to interpersonal cues and often impairment with the interpretation of such cues. In addition, appropriate emotional displays are pervasively punished, and extreme emotional displays are intermittently reinforced, increasing the likelihood of emotionally escalated behavior. For example, individuals with BPD may not be taken seriously at home (e.g., dismissed, criticized, etc.) until they threaten suicide or self-injury, after which they may receive nurturance, respect, soothing, and the like, which may maintain or increase the likelihood of future self-injurious behavior in similar contexts.

Emotional Vulnerability. In addition to environmental invalidation, the biosocial model predicts that the emotion dysregulation problems characteristic of BPD are exacerbated by emotional vulnerability. Emotional vulnerability refers to a biologically mediated predisposition for affective instability involving genetic, intrauterine, and temperamental factors. As a construct, emotional vulnerability is defined as heightened sensitivity, heightened reactivity, and slow return to a baseline level of emotional arousal. That is, individuals with BPD respond quickly to a wide range of contextual emotional stimuli, respond with a high magnitude of physiologic arousal, and take a long time before arousal decreases to baseline. Once emotionally aroused, the slow return to baseline makes it increasingly likely that subsequent events will evoke extreme emotional responses. Similar to the searing physical pain a burn victim feels when the wind gusts, emotionally vulnerable individuals feel lasting emotional pain in response to seemingly ordinary daily events. Parents with emotionally vulnerable children are commonly frustrated and have difficulty managing their child's behavior. However, this may be particularly true for parents with existing psychopathology or significant deficits in parenting skills. In addition, high emotional vulnerability may make an already invalidating environment even worse, with a mismatch between child and parent emotionality contributing to additional invalidation, even if inadvertent. According to the model, in a reciprocal fashion, such increased invalidation may further exacerbate emotional vulnerability.

Table 32.11–1.
Diagnostic Criteria for Borderline Personality Disorder

- Emotion dysregulation
  - Affective instability
  - Problems with anger
- Behavioral dysregulation
  - Impulsive behavior
  - Self-injurious behavior
- Cognitive dysregulation
  - Dissociation
  - Paranoia
- Interpersonal dysregulation
  - Chaotic relationships
  - Fears of abandonment
- Self-dysregulation
  - Identity disturbances
  - Chronic feelings of emptiness

Emotion Dysregulation. In the context of environmental invalidation and emotional vulnerability, the biosocial model states that emotion dysregulation occurs across behavioral/motoric, physiologic, and cognitive/experiential emotional systems. In a transactional, back-and-forth manner, problems with emotion dysregulation can aggravate emotional vulnerability and intensify environmental invalidation, and these problems can then further exacerbate emotion dysregulation. As shown in Table 32.11–1, the biosocial model asserts that BPD criterion behaviors stemming from emotion dysregulation can be organized across domains of functioning (i.e., emotional, behavioral, cognitive, interpersonal, and self). Processes of dysregulation thought to mediate borderline symptoms include problems with up- or downregulation of physiologic arousal, inhibition of mood-dependent behavior, reliance on avoidance strategies, attentional dyscontrol, processing of emotional information, self-soothing, and self-validation. For example, an inability to readily up- or downregulate physiologic arousal may lead to the development of extreme behavioral dyscontrol, such as self-injurious, impulsive, and aggressive behavior. Extreme emotional arousal may narrow attention, making it difficult to turn attention away from emotionally evocative stimuli, thereby biasing interpretations of events in a fashion consistent with mood state. For DBT therapists, the biosocial model provides a framework for understanding how a variety of topographically distinct problems (e.g., self-injury, substance use, anger problems, dissociation) may be actively targeted for change through interventions aimed at similar mechanisms. Of importance, it also provides a nonpejorative, compassionate, and accepting approach sometimes difficult to adopt when working with BPD patients.

TECHNIQUES

DBT is a principle-driven, flexible, and comprehensive treatment for multidiagnostic patients. As a behavioral therapy, it is quite change focused and driven by behavioral principles and models of classical and operant conditioning. Yet, at the same time, DBT incorporates numerous acceptance-based strategies to implement when changing behavior may not be possible or effective.

There are four primary modes of treatment in DBT: group skills training, individual therapy, phone consultation, and consultation team. Other, ancillary treatments that are typically uncontrolled in DBT are pharmacotherapy and acute inpatient psychiatric hospitalization. To practice DBT, it is necessary to include all standard modes of treatment into a treatment program. Patients must be seen by a DBT-trained therapist, and therapists must participate in a DBT consultation team. The following section briefly reviews the function, process, and structure of DBT treatment modes.

Skills Training

Skill training occurs in the group format and is a skill class rather than a traditional interpersonal process-oriented psychotherapy group. The primary function of the group is the acquisition of new coping skills. The assumption is that these patients must learn specific behavioral, emotional, cognitive, and interpersonal skills not learned earlier in life. On the other hand, it is also assumed that group members are not fragile, and that emotional arousal during group provides natural opportunities for skills practice.

Groups occur on a weekly basis, typically last for 2 hours, and usually include between 4 and 10 patients, as well as two DBT-trained co-leaders.

The first hour is usually devoted to a brief mindfulness practice and homework review from skills learned the previous week. During homework review, group members take turns discussing their successes and challenges in implementing skills learned the previous week. Group participation is strongly encouraged, but observations about others in the group and traditional “process” therapy are actively discouraged. During the second hour of group, new skills are taught didactically from the skills training manual. This workbook provides a wide variety of handouts and worksheets designed to decrease emotion dysregulation, including separate skill modules for distress tolerance, emotion regulation, and interpersonal effectiveness, as well as the core skills of mindfulness that are taught recurrently after each skills module.

Skills Modules: Mindfulness

Mindfulness is considered the core set of skills in DBT. Derived from Eastern meditative and Christian contemplative traditions, mindfulness is the practice of paying attention in a particular way: on purpose, in the present moment, and without judgment. Patients learn that their behavior is a function of current emotions (emotion mind) or logical analysis (reasonable mind). “Wise mind” knowing is emphasized as an integration of emotion and reasonable minds. The concept of wise mind relates to a synthesis between emotional and logical thinking, such that coping decisions are both effective and remain within personal values but do not need to be free of all of emotions. For instance, this can translate into suspension of short-term behaviors that have harmful consequences (e.g., drinking) for long-term gains in self-esteem or feelings of mastery, even if negative emotions are elicited. Wise mind can also be experienced as spiritual or “knowing from the gut” that a decision or action makes sense, given the context.

Mindfulness skills also include the ability to observe, describe, and participate fully in one’s actions in a nonjudgmental, one mindful, and effective manner. Observing refers to noticing experiences without becoming attached, allowing thoughts or other internal experiences to flow freely with full awareness. Describing follows observing and involves labeling or putting words on experiences. Participating is somewhat different and involves behaving effectively without necessarily observing and describing internal experiences (e.g., an athlete at peak performance). Being nonjudgmental, including being aware of judgments and letting go of their literal truth, is a central skill repeatedly practiced by DBT patients and therapists. Being one mindful entails a sharpening of attentional focus on one thing or activity at a time. This skill involves staying in the present moment and not becoming distracted by events in the past or concerns in the future. Finally, the focus on effectiveness is a key aspect of mindfulness. Effectiveness refers to behaving in a way that is consistent with one's values and long-term goals, which often translates to not immediately reducing distress despite adverse consequences. The emphasis on effectiveness as a DBT skill illustrates how mindfulness is a behavioral, psychological, and spiritual practice—not purely a meditational pursuit.

Skills Modules: Distress Tolerance

Patients with BPD frequently experience extreme emotional distress. Many symptoms may develop as a means of reducing chronic and intense distress, albeit in maladaptive fashion. For example, self-injurious behavior may function temporarily to reduce distressing emotional states. Beyond overtly self-injurious behavior, a wide variety of topographically distinct behaviors may function abruptly and dramatically to reduce emotional distress, such as impulsivity, risk taking, anger, and dissociation. The distress tolerance module is designed to teach patients how to tolerate aversive emotional experiences without behaving maladaptively. A list of crisis management skills is taught, including strategies for effective temporary distraction, such as activities, eliciting opposite emotions, and squeezing ice or a rubber ball. Self-soothing skills are introduced; by which patients learn intentionally to soothe themselves during periods of crisis with calming visual stimuli, sounds, smells, tastes, and objects to touch. In addition, other skills such as imaginal and relaxation exercises are taught to improve the current moment in order to avert doing something that would make a crisis even worse. Other distress tolerance skills include awareness, breathing, and half-smile exercises, as well as radical acceptance of reality in the current moment. Overall, distress tolerance skills are intended to interrupt and change habitual and problematic responses to acute emotional distress, allowing the opportunity for newer, more effective behaviors to emerge.

Skills Modules: Emotion Regulation

The emotion regulation skills module is designed to help patients to better understand their emotions, reduce emotional vulnerability, and decrease emotional suffering. Specific skills taught include increasing awareness of emotions, identifying and challenging distorted ways of thinking about emotions, learning how emotions are related to problem behaviors, accurately labeling emotions, understanding the functions of emotions, reducing emotional vulnerability, increasing pleasant emotions, and *acting opposite* to behavioral urges associated with emotions. Although all of these skills are useful, the opposite-action skill is particularly helpful for most borderline patients because it can be applied in many contexts to change a variety of problematic behaviors.

The opposite-action skill provides an algorithm for knowing when to try to change an emotional response, in order to develop a new and more effective way of responding in the future. This includes first determining whether the emotion or the intensity of the emotion is justified or effective based on the current situation. Next, it is important to know the action urge of the emotion being experienced. Each emotion has its own urging component (e.g., anger—attack, fear—run, sadness—withdraw, guilt—repair, shame—hide). When the patient is experiencing an unjustified emotion that they wish to change, the skill is to go opposite to the action urge of the emotion. For example, if a patient is feeling ashamed or guilty after telling a friend that the patient feels excited about doing a new activity, opposite-action skills would teach the patient to ask first whether their behavior was egregious according to the patient’s value system. If so, then a repair (e.g., apology) would be in order and would function to lower justified guilt. However, if the shame or guilt was unjustified (i.e., the patient did no wrong), then the patient would be instructed not to repair but to repeat the behavior (i.e., expressing enthusiasm about new activities) multiple times. This leads to less guilt over time as the patient learns that giving opinions does not always result in negative outcomes.

Skills Modules: Interpersonal Effectiveness

Because chaotic interpersonal relationships are a hallmark characteristic of BPD and because interpersonal stressors are a common trigger preceding suicide attempts, the development of interpersonal skills is crucial in the treatment of BPD. This skill module teaches group members how to identify factors interfering with interpersonal effectiveness, challenge common cognitive distortions associated with interpersonal situations, and determine the appropriate level of intensity for making requests or saying no in a given situation. Patients are taught what to say and how to say what they say depending on their priority in an interpersonal situation. In addition, group members are strongly encouraged to practice these skills in a wide variety of new situations. Guidelines for being taken seriously, attending to relationships, and preserving self-respect are introduced, and patients are instructed to practice developing new interpersonal skills based on these guidelines.

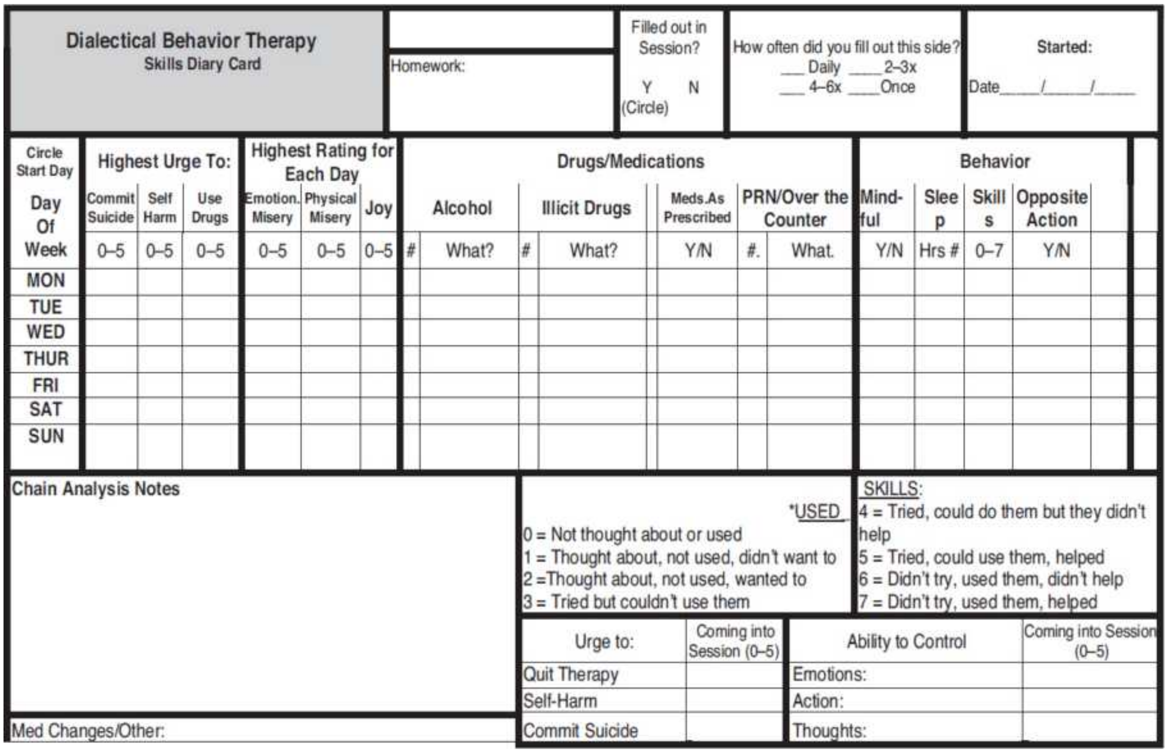

Dialectical Behavior Therapy Skills Diary Card	Homework:	Filled out in Session? Y N (Circle)	How often did you fill out this side? ___ Daily ___ 2–3x ___ 4–6x ___ Once	Started: Date___/___/___

Circle Start Day	Highest Urge To:			Highest Rating for Each Day			Drugs/Medications							Behavior				
Day Of Week	Commit Suicide	Self Harm	Use Drugs	Emotion. Misery	Physical Misery	Joy	Alcohol		Illicit Drugs		Meds.As Prescribed	PRN/Over the Counter		Mind-ful	Slee p	Skill s	Opposite Action	
	0–5	0–5	0–5	0–5	0–5	0–5	#	What?	#	What?	Y/N	#.	What.	Y/N	Hrs #	0–7	Y/N	
MON																		
TUE																		
WED																		
THUR																		
FRI																		
SAT																		
SUN																		

Chain Analysis Notes

Med Changes/Other:

*USED

0 = Not thought about or used
1 = Thought about, not used, didn't want to
2 =Thought about, not used, wanted to
3 = Tried but couldn't use them

SKILLS:

4 = Tried, could do them but they didn't help
5 = Tried, could use them, helped
6 = Didn't try, used them, didn't help
7 = Didn't try, used them, helped

Urge to:	Coming into Session (0–5)	Ability to Control	Coming into Session (0–5)
Quit Therapy		Emotions:	
Self-Harm		Action:	
Commit Suicide		Thoughts:	

FIGURE 32.11–2. A typical diary card.

Individual Therapy

Individual therapy sessions with a DBT-trained therapist are typically 50 to 60 minutes once per week, but this can vary, depending on whatever is needed for a given patient in a given week. A primary function of individual therapy is to improve and maintain patient motivation. A validating environment is created in which patients are treated with compassion and acceptance. Within this context, the therapist and patient relentlessly pursue behavioral change. Skills learned during group sessions are woven into skill plans in anticipation of upcoming events. In addition, episodes of emotion dysregulation from the previous week are discussed in light of skills that could have been used. Factors interfering with progress in treatment are also discussed in individual therapy sessions, preventing problems that might impede the development of new skills and helping patients to remain engaged in treatment despite urges to dropout.

Because it may take a long time for a patient's arousal to return to baseline, care must be taken to structure sessions in a way that facilitates learning and does not lead to a depletion of attentional resources. Once they are highly aroused in session, it is common for patients with BPD to become so emotionally dysregulated that new learning is not likely to occur during the session. A mindfulness or relaxation exercise may be used toward the end of session to facilitate return to baseline arousal before the patient leaves.

Diary Cards

To monitor a variety of targets, a daily diary card is used. For example, patients rate their mood, monitor the frequency of self-injurious behavior and urges, and track other relevant targets. Figure 32.11–2 shows a prototypical diary card. The diary card is reviewed at the beginning of each session, and the therapy session is organized around targets evident on the diary card. Given the plethora of treatment targets and the possibility that patients will not remember salient events from the week, the diary card is instrumental in directing therapy sessions toward highly relevant targets.

Behavioral Analysis

To change dysfunctional behaviors, DBT uses a number of problem-solving strategies. Behavioral analysis is frequently used to identify problem behaviors and understand the relevant context in which these behaviors occur. In a nutshell, behavioral analysis involves an active, directive effort by the therapist to identify specific antecedents and consequences associated with the problem behavior being analyzed. A thorough elaboration of events before, during, and after problem behaviors facilitates the selection of appropriate treatment interventions. Based on a functional-analytic approach to behavioral assessment, the goal of behavioral analysis is prediction and control of functional classes of problem behavior rather than traditional diagnostic assessment of disease entities. In other words, in DBT, borderline symptoms are conceptualized as problem *behaviors.* These may be external, publicly observable behaviors such as self-mutilation or impulsive aggression. Alternatively, problem behaviors often are internal, publicly unobservable experiences such as self-judgmental thoughts or urge to use substances. In sum, behavior refers to anything a person does or experiences (e.g., thoughts, emotions, sensations, and overt actions). The inclusion of internal experiences as targets is consistent with other contemporary behavioral therapies and is crucial to understanding the phenomenologic experiences of the patient associated with treatment targets.

Because behavioral analysis involves explicating the links in a chain of events, it is often referred to as a chain analysis. During a chain analysis, the topography, intensity, duration, and frequency of the target problem, all are discussed. As links in the chain are explored, the DBT therapist considers the role of classical and operant conditioning. Classically conditioned behaviors are under the control of an antecedent stimulus, and operantly conditioned behaviors are under the control of consequent events. For example, strong urges to self-injure may be classically conditioned to occur after arguing with a partner. On the other hand, self-injurious behavior may be operantly conditioned by the consequences that follow, such as less hostility or increased attention from a partner. Knowing the controlling variables (i.e., antecedent or consequent) of a problem behavior informs the choice of interventions. Strategies for changing antecedents include behavioral exposure (e.g., rehearsing arguments) and stimulus control (e.g., avoid partner). In addition, other behavioral principles used during chain analyses are positive and negative reinforcement, punishment, extinction, and shaping. Doing a chain analysis is like being a detective solving a crime. All of the events that lead to and follow the problem behavior are uncovered, revealing critical links in the chain amenable to change.

Dysfunctional links in the chain are examined and replaced with more adaptive responses during a *solution analysis.* Typically, this is guided by three questions: (1) Can the patient change the circumstance? (2) Can the patient change their emotional reaction? (3) Can the patient better tolerate the pain associated with the problem? Together, the patient and the therapist collaboratively develop strategies to replace problematic links and then commit to using new solutions the next time the problem behavior emerges.

Skill Enhancement

Skills learned during group sessions are enhanced in individual therapy. It is often fruitful to have patients rehearse behavior in session. Behavioral rehearsal may occur in the form of covert rehearsing of challenges to distorted cognitions or cheerleading statements said silently. Rehearsal also includes role-playing interpersonal scenarios with an emphasis on skill use. Therapists provide reinforcement and coaching, and rehearsal occurs with repetition. Whenever effective behavior occurs in session, it is the DBT therapist's job to reinforce such behavior. Alternatively, ineffective behavior evoked during session is often ignored or punished. In this way, the therapist is actively involved in *contingency management.* Positive regard is applied contingently. For example, when the patient is behaving inappropriately (e.g., threatening self-injury), the therapist remains cool and matter of fact, whereas skillful behavior (e.g., committing to not self-injure and using skills) is greeted with therapist's warmth.

Another method of facilitating skill use is to use behavioral exposure and response prevention in session. Patients will often become angry, ashamed, or fearful in session, and a range of behavioral problems may emerge in response to these emotions. Patients feeling angry may lash out, whereas patients feeling ashamed may look down or dissociate, and patients feeling fearful may abruptly terminate the session. Behavioral exposure and response prevention applied to these emotions in session would target paying attention to these emotions nonjudgmentally, observing urges to behave ineffectively, and blocking these urges by not lashing out, keeping eye contact, or remaining in the therapy room. In addition to changing the behavioral pattern, maladaptive cognitions are modified through standard cognitive therapy techniques aimed at identifying and challenging unhelpful ways of thinking.

Validation

All efficacious therapies require verification of what the patient does effectively and disconfirmation of what is ineffective. Indeed, verification of self-constructs may reduce physiologic arousal, and validation as an intervention may function to enhance patient learning by modulating arousal and enhancing the therapeutic relationship.

Although validation may mean many different things, in DBT validation is operationalized on several levels. These levels of validation are shown in Table 32.11–2 and are integral to doing DBT. In addition, validation may be explicitly verbal, or it may occur more implicitly and functionally, as when the therapist offers a tissue when an emotionally inhibited patient appears on the verge of tears. Validation may be used as pure acceptance, with no directed effort toward change. DBT therapists attempt to validate what is valid and usually ignore or if needed can invalidate what is invalid. For example, a therapist may say, “It makes sense that you felt angry in that situation, because your colleague lied about you, but cussing at her in the break room is not effective.” This requires the DBT therapist to discern carefully what is valid (i.e., feeling anger) and to apply validation thoughtfully. In addition to fostering a positive therapeutic relationship and providing a compassionate environment, validation by the therapist functions as a model for patients and may be taught directly as a set of skills.

Table 32.11–2.
Levels of Validation in Dialectical Behavior Therapy

1. Listening and observing
2. Accurately reflecting patient experiences
3. Helping patients to articulate unverbalized emotions, thoughts, and patterns of behavior
4. Communicating an understanding of behavior as valid, given past learning history or biologic vulnerability
5. Communicating an understanding of behavior as valid, given a current context or what is deemed normative
6. Providing radical genuineness, treating the individual as an equal and not as a sick and fragile patient

Table 32.11–3.
Dialectical Strategies Used in Dialectical Behavior Therapy

1. Providing acceptance and change-focused interventions
2. Nurturing the patient and demanding that patients help themselves
3. Being stable and persistent, as well as flexible
4. Highlighting the patient's strengths and deficits
5. Structuring the session with an agenda, and responding to in-session patient behaviors as they occur
6. Highlighting both ends of continua, and making synthesizing statements
7. Pointing out paradoxes when present (e.g., patient's behavior, therapeutic process)
8. Using metaphors
9. Playing devil's advocate
10. Extending the seriousness or implications of patient's communication
11. Helping the patient to activate "wise mind"
12. Helping to make lemonade out of lemons
13. Allowing natural changes in therapy without attempting to control all outcomes

Dialectical Strategies

As mentioned, efficacious therapy and learning requires both confirmation and disconfirmation. Dialectical strategies are fundamentally based on this notion and revolve around behavioral acceptance (e.g., validation) and change (e.g., problem solving). Dialectical reasoning is pursued with the patient, by which the therapist helps the patient to move from a polarized position of "either-or" to a dialectical synthesis of "both-and." There are a number of specific dialectical strategies used with patients (Table 32.11–3). Of importance, to be effective, these strategies must be used with therapist genuineness, not as gimmicky techniques. On the other hand, the therapist must be willing to let go of the truth or rightness of any dialectical strategy and instead must continually search for other ways to help patients change maladaptive patterns.

Telephone Consultation

To enhance the generalization of skills, patients in DBT may contact their individual therapists for telephone consultation between sessions.

Because DBT patients may experience unrelenting crises, it is crucial that therapists observe personal limits associated with telephone consultation. Telephone calls are intended to be brief and usually last less than 10 minutes. Patients are oriented during pretreatment to the purpose of skills calls and are told to call when they are unable to implement skills in necessary situations but before crises occur. To reduce inadvertent reinforcement of self-injurious behavior using warmth and validation over the telephone, DBT uses a *24-hour rule*: Patients will not receive skills help from their individual therapists until 24 hours after any self-injurious behavior. Instead, early in treatment, a contingency plan is created for occasions when patients self-injure, and DBT therapists may refer patients to this plan in the event that they call after self-injuring. In all consultation calls, DBT therapists assess for immediate danger and provide appropriate assistance when the patient is deemed to be in imminent danger of harming themself or others.

Table 32.11–4.
Consultation Team Agreements in Dialectical Behavior Therapy

1. Meet weekly for 1–2 hours
2. Discuss cases according to the treatment hierarchy (i.e., self-injurious/life-threatening behavior, treatment-interfering behavior, and quality-of-life–interfering behavior)
3. Accept a dialectical philosophy
4. Consult with the patient on how to interact with other therapists, and do not tell other therapists how to interact with the patient
5. Do not expect consistency of therapists with one another (even across the same patient)
6. Allow all therapists to observe their own limits without fear of judgmental reactions from other consultation group members
7. Search for nonpejorative, empathic interpretation of patient's behavior
8. Acknowledge that all therapists are fallible

Consultation Team

As mentioned earlier, the use of a consultation team is a necessary mode of treatment. Team members commit to weekly meetings and agree to a team structure and process (Table 32.11–4). There have been modifications and new consultation team agreements delineated, including how to develop and maintain a DBT team culture, as well as skills such as being willing to call out the “elephant in the room” when others do not. In important ways, team members treat each other with DBT by providing validation, support, and motivational enhancement strategies. This support is invaluable and can help DBT therapists with a more balanced approach toward their patients. A consultation team also provides opportunities for fresh perspectives and new solutions, helping therapists to get unstuck and become hopeful. It is not uncommon for DBT therapists periodically to become rigid in their thinking and behavior with a patient. The consultation team offers problem solving and validation for the therapist, and team members actively use a dialectical process to help find effective syntheses between polarized positions. For example, the team can help remind the therapist to continue managing contingencies in session appropriately (e.g., not being warm in response to patient hostility). If possible, it is extremely helpful to videotape therapy sessions and watch important segments during the consultation team meeting because this engenders a full appreciation for the difficulty a therapist may be having and allows the team to ensure that all members are indeed adhering to the treatment.

CLINICAL ISSUES

In DBT, problem behaviors are seen as the inevitable sequelae of dysregulated emotions or as maladaptive methods of altering emotional experiences. For example, self-injury may function to regulate dysregulated emotional responding. This theoretical stance has resulted in a treatment package that largely focuses on modifying various aspects of the patient’s emotional system and treatment formulations are apt to focus on reducing ineffective action tendencies linked with dysregulated emotions. In DBT, the problem behavior is targeted as a priority, in an effort to break cycles of reinforcement (Fig. 32.11–3). The present section begins with an outline of the assumptions and agreements of DBT, followed by a discussion of the levels and targets of treatment. A brief case example has been included to illustrate some of the strategies in DBT.

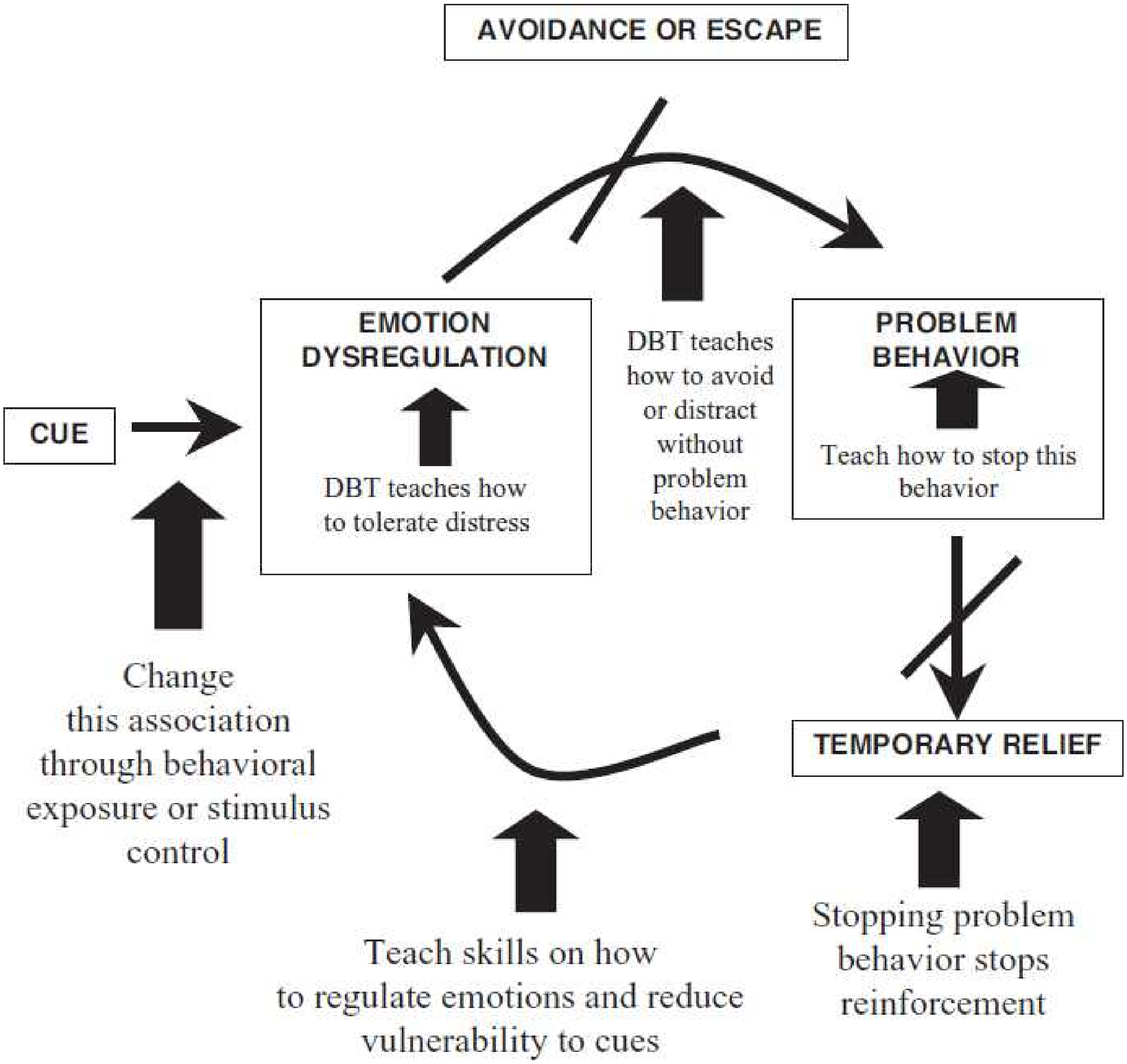
AVOIDANCE OR ESCAPE
CUE
EMOTION DYSREGULATION
DBT teaches how to tolerate distress
DBT teaches how to avoid or distract without problem behavior
PROBLEM BEHAVIOR
Teach how to stop this behavior
TEMPORARY RELIEF
Change this association through behavioral exposure or stimulus control
Teach skills on how to regulate emotions and reduce vulnerability to cues
Stopping problem behavior stops reinforcement

FIGURE 32.11–3. How dialectical behavior therapy (DBT) works.

Assumptions and Agreements

Most therapeutic approaches have underlying assumptions about what is expected from patients and from the treatment. In DBT, these assumptions are openly discussed with patients in the first few sessions, during what may be called pretreatment (Table 32.11–5). Patients are also oriented to the philosophy, theories, levels, and modes of treatment. Patient and therapist agreements for treatment are also discussed during pretreatment sessions (Table 32.11–6). In addition, it is important for the therapist to discuss the requirement that patients commit to participating in DBT. Indeed, commitment to treatment may wax and wane over time for many psychotherapy patients. Thus, the therapist providing DBT needs continually to be alert to changes in commitment by the patient and be willing to go back and work on commitment to treatment again and again if needed.

Table 32.11–5.
Patient and Therapy Assumptions in Dialectical Behavior Therapy (DBT)

Patients

1. Patients are doing the best they can
2. Patients want to improve
3. Patients need to do better, try harder, and be more motivated to change
4. Patients must solve their current problems, regardless of who caused them
5. Patients are living lives that are unbearable as they are currently being lived
6. Patients must learn new behaviors in all relevant contexts

Therapy

1. Patients cannot fail in DBT, but the therapy or therapist can fail the patient
2. Helping patients work toward their ultimate goals in life is the most caring thing a therapist can do
3. DBT therapists need support
4. The therapeutic relationship is a relationship of two equals
5. Principles of behavior are universal, affecting both patients and therapists alike

Levels of Treatment and Treatment Targets

There are four stages of DBT. During stage I, patients often present with severe behavioral dyscontrol and life-threatening problems, including self-injurious behavior such as cutting, burning, or otherwise deliberately harming oneself physically with or without the intent to die. Patients in this stage commonly meet criteria for multiple Axis I disorders as well as BPD. In addition, unrelenting crises that may vary from week to week in content, as well as a multitude of treatment-interfering behaviors, commonly occur for these individuals. To provide a framework for therapeutic targets, a hierarchy is used to guide the therapist toward specific goals in a given session: (1) reduce life-threatening and self-injurious behaviors; (2) reduce treatment-interfering behaviors, such as arriving late to session, not attending skills training group, lying, not completing homework (e.g., diary cards), arriving to therapy intoxicated, avoiding emotions during therapy, talking excessively about topics that are not germane to the problems being targeted, etc.; and (3) reduce quality-of-life–interfering behaviors, such as drug use, dysfunctional eating behavior, anxiety, depression, and physical health problems.

Table 32.11–6.
Patient and Therapist Agreements in Dialectical Behavior Therapy

Patient Agreements

1. Stay in therapy for a specified period of time, usually 1 year
2. Attend all therapy sessions
3. Therapy will be discontinued if four consecutive sessions are missed
4. Work toward terminating self-injurious behavior and other therapeutic targets
5. Participate in skills training
6. Abide by relevant research conditions of therapy
7. Pay agreed-on fees for service

Therapist Agreements

1. Maintain competence and effort
2. Provide ethical and professional treatment
3. Be available for weekly sessions and phone consultation
4. Treat patients humanely, with respect and integrity
5. Maintain confidentiality
6. Seek appropriate consultation

Within and across therapy sessions, this hierarchy largely informs the selection of treatment targets. Often, treatment- or quality-of-life–interfering problems occur in the context of life-threatening or self-injurious behavior. In these cases, it is crucial that the DBT therapist refers to the hierarchy, targeting the reduction of life-threatening/self-injurious behavior first, even if this is less desirable than talking about something else to the patient.

Once skillful behaviors learned during treatment have replaced major life-threatening/self-injurious, treatment-interfering, and quality-of-life–

interfering behaviors, patients begin stage II. During this stage of treatment, post-traumatic stress is often treated using empirically supported exposure therapy protocols while continuing to increase the patient's ability to experience aversive emotions skillfully. When post-traumatic stress is not present, longstanding difficulties with emotional experiencing and expression are targeted. For example, patients may need further to develop skills such as emotion labeling, reducing vulnerability to emotion dysregulation, emotional acceptance, or adaptive emotional change. These skills may have been learned already in stage I, but in stage II, they are generalized more directly to contexts that do not involve acute behavioral dyscontrol. Patients in this stage of DBT may revert to behavioral dyscontrol characteristic of stage I. Thus, it is crucial that stage I treatment targets (e.g., severe behavioral dyscontrol) are closely monitored while patients are in stage II.

For those who persist in DBT beyond stage II, in stage III of DBT, the therapist and patient target problems in living that are not debilitating but that interfere with experiencing *ordinary* happiness and unhappiness. This includes addressing unresolved problems with social and occupational functioning. Finally, stage IV targets a sense of completeness, spiritual growth, insight, enhanced awareness, sustained joy, and other fulfilling life endeavors. Although Linehan articulated these four stages of DBT, the first stage has been developed most thoroughly and is what has been tested and implemented for the most egregious and out-of-control behavioral problems among those with BPD.

In this transcript, the patient was a 32-year-old Black man presenting for psychotherapy with BPD and substance abuse. In this prototypical example of DBT, the therapist conducted a behavioral analysis of self-injurious behavior, identified possible skills, and began to obtain a commitment to skills-based solutions.

T: It sounds like this week you hurt yourself by punching the basement wall repeatedly and then you drank a lot of alcohol. Let's do what we do and go through a chain analysis to slow the situation way down and figure out exactly what happened. Does that sound okay?

P: I know what happened. My ex-wife told me I couldn't talk to my kid on the phone because I'd had a couple drinks with my buddy Sean. She's always using the kids to try to control me and make me feel like a monster. So I lost it.

T: I know you've been really hurt by her in the past, and this kind of thing will probably keep coming up. How do you feel about the way you reacted?

P: Pretty terrible. I think I probably drank after so I wouldn't feel so ashamed.

T: Did you punch the wall in order to hurt yourself?

P: Yeah. Sometimes it seems better to feel physical pain than what I'm feeling after I talk to her. But then I wind up feeling worse.

T: How do you feel about committing to not hurting yourself again?

P: I think that's also why I feel so terrible. I had committed and I broke it. But yeah, I'm willing to commit again.

T: I know your word is important to you, and I can see that it feels really awful to have broken a commitment. But that's exactly why we do these behavioral analyses—so that we can look at what happened together that made you hurt yourself again. That way, we can stay firm to the commitment. As you know, it's really, really important to keep a strong commitment to not hurt yourself so we can get you out of this pain you've been experiencing for so long.

P: Yeah, I hear you. Okay, let's go.

T: Okay, so tell me exactly where you were when you called your wife.

P: I was in my basement because Sean and I were having beers and tooling around on the guitars. Then he left because it was my time to call my kid, so I said "bye" to him and called my ex.

T: And how were you feeling when you called her?

P: I mean I never like calling her, but I was in a decent mood. Sean and I were having a good time.

T: So your mood was relatively high, and then you called your ex-wife and what happened?

P: I asked to speak to MaKayla and she said no, because I'd been drinking. She said she can always tell when I've had a drink.

T: And how did you feel then?

P: Really angry. I wasn't even drunk, and it felt like she was just trying to keep me away from MaKayla. It's like I can never just get any peace. There's always a roadblock, always a reminder of how I screwed everything up.

T: You actually look pretty overwhelmed and sad right now. You're moving around a lot in your chair and rubbing your face. Are these emotions coming up for you now?

P: Yeah, I guess (patient looks down at the floor and shakes his head).

T: Is this how you felt on Wednesday when you hurt yourself?

P: I felt a lot of things. Overwhelmed, angry, sad, frustrated, hating myself, hating her. All of them. You name it.

T: That's a whole lot of emotions to hold inside. I can understand why you would want to let them out.

P: Yeah. Well, they're all back.

T: I know going over this stuff can feel really, really hard. I appreciate you sticking it through with me. Can you tell me what happened when she said you couldn't talk to your daughter?

P: I kind of froze up and panicked. I asked nicely one more time and she said no, that I'd hurt MaKayla enough with my drinking and needed to stay away from her until I could learn to be a man.

T: That must have been awful to hear. What did you do next?

P: I had like 2 seconds of control left that kept me from cussing her out so I threw the phone and everything kind of went black. Like not literally black, but it's like the way I felt got so huge I couldn't feel anything anymore except this total loss of control.

T: Is that when you punched the wall?

P: Yeah, it's like I started punching the wall so I could get myself back in control. That's the messed up part. It actually works. I just punched the crap out of the wall and I came back into my body.

T: How did you feel when you came back?

P: Like a total loser. She told me to go be a man and then I punched the wall like a teenager. So, I guess I failed at being a man again. With my wife, with my kid, and now even when I'm just alone in the basement. I even tried to dunk my face in the beer cooler full of ice, so I'd at least have tried a skill to calm myself down. Even though it was too late.

T: That's actually great you thought of using a skill, Devon. Did it help bring down your emotional arousal?

P: Maybe a little bit. Sometimes these skills feel like using duct tape to keep a house upright in a hurricane.

T: I hear you. You've gotten used to using other means to calm you down, and it can be hard to try something new. That's why it's important to get really clear on what sets these big emotions off so we can intervene when the skills have the biggest chance of working. Tell me what happened after you dunked your head in the cooler.

P: I just felt so bad I couldn't stand it. Like I couldn't look at myself, so I just blacked myself out on alcohol. I finished the cooler and then found a half-bottle of bourbon and finished myself off.

T: What do you mean by "finished yourself off?"

P: Drank until I passed out and couldn't feel anything anymore.

T: Devon, you've done a great job filling in these details for me. It's really helped me to understand. I can also see that it's a similar pattern to other times you've left interactions with your wife feeling really trapped, angry, and ashamed. Is there any point in this chain where you think you could have tried a skill earlier?

P: Maybe I could have used those 2 seconds of control I had to dunk my face in the cooler.

T: That's exactly right. Are there any other times you could have tried?

P: Maybe when she first told me I couldn't see MaKayla and I started to get really upset.

T: Absolutely. What kind of skills you could you have tried?

P: Maybe I could put my face in the cooler and then do 40 push-ups. Or just take a few breaths and try to use the mindfulness skills y'all have been teaching me. Or instead of throwing my phone, I could have called you. It's just really difficult to try this stuff when I feel like I only have 2 seconds of control.

T: Yeah, it’s really hard for most people to use these skills in the moment. We can’t go back and fix the past, but, the good and bad news is that we can expect that you’ll have more confrontations with your ex, and we can plan for the next time that happens. Are you willing to commit to practicing skills in anticipation of future conflicts with her?

P: Yes. I really don’t want to punch anything again.

T: Good. Any of these skills can work. Let’s choose at least two that feel most effective for you. What do you think? Which would work best, and which would you be most able to do?

P: Probably. I feel like I need something really intense and fast.

T: Would you be willing to practice one right now?

P: Okay. I’ll do mindfulness practice with you

T: Fantastic. Before we do, let’s just remember that when these situations arise and your emotions get really big, you are at risk for wanting to hurt yourself again, because that’s calmed you down many times before. This could happen even if you use the skills. But it’s really important to stay very committed to not hurting yourself. Just know if your commitment starts to waiver, I am always here for you to call, but only before you harm yourself. Deal?

P: Deal. I’m committed.

T: Okay.

ETHICAL ISSUES

Most ethical issues faced by DBT therapists are not specific to the practice of DBT. The following section includes a discussion of ethical issues commonly faced by DBT therapists across the four modes of treatment. One such example could include a patient disclosing self-harm or suicidal ideation in the context of skills group where it is not possible to do a full risk assessment in front of other group members. These types of situations are not uncommon, which is one of the reasons why it is important to have two group leaders so that one group leader can take the patient outside to assess risk and develop a safety plan. Furthermore, as discussed previously, the consultation team can also provide assistance to the therapist in managing these situations and documenting appropriately.

The *Four-Miss Rule,* whereby a patient is dropped out of therapy after missing four consecutive sessions of individual or group therapy, presents an ethical dilemma for DBT therapists. On the one hand, if a patient is not receiving therapy, then they are, in a matter-of-fact way of thinking about it, simply not in therapy. Similarly, irrespective of reasons for not taking a dose of medicine, if one does not take the dose, the medicine is not going to take effect. On the other hand, there may be many reasons why patients miss 4 consecutive weeks of individual or group therapy visits. Still, after missing 4 consecutive weeks, in DBT, the patient is considered out of treatment. Ethically, at this point, the therapist or clinic is responsible for the patient until they have been successfully referred to a provider in the community. As such, the interim period presents an ethical challenge. Therapists should, therefore, carefully discuss and frequently review with their patients the DBT treatment agreement and its rationale. The intent of the *Four-Miss Rule,* for instance, is to prevent patients from simply drifting out of therapy. It is also important for therapists to be proactive in communicating with and motivating patients to remain engaged in therapy if they notice a patient is approaching four missed sessions. If a patient is out of therapy due to the *Four-Miss Rule,* the treatment team needs to support the therapist in helping the patient find treatment elsewhere, or develop opportunities to overcome the problems that led to them missing treatment in the first place. After some time and evidence that the problems have been resolved, the patient can petition to the therapist, who can consult with the team, about the pros and cons of the patient being permitted back into DBT.

Another common set of ethical challenges in DBT occurs when therapists take telephone coaching calls to generalize skills in everyday contexts where the skills need to be used. Telephone coaching introduces unique concerns related to confidentiality and risk assessment. For example, sensitive personal information is at risk if a patient loses their phone or if someone overhears the content of a coaching call. Although the purpose of telephone coaching is skills generalization, therapists often need to assess the risk for self-injurious behaviors. Doing an assessment over the course of a brief conversation on the phone can cause complications. Although there are many possible solutions to this problem, ultimately, the individual therapist leverages the individual sessions (e.g., by motivating patients to use skills and problem-solving skills use) to prevent self-harm–related crisis calls.

Another ethical issue occurs in the context of DBT team consultation, and is related to the dialectical tension between communicating with other treatment team members on behalf of a patient versus allowing the patient to speak to their provider directly. Ethical guidelines require consultation among treatment providers. Although it would often be easier to provide certain information about a patient to other members of the treatment team, one important aim of DBT is to help patients learn the skills necessary to get their needs met autonomously. Across each of the modes of DBT, therapists must balance apparent conflicts between what is required ethically and what is indicated clinically.

RESEARCH AND EVALUATION

DBT has garnered sufficient empirical evidence for its efficacy in treating BPD, warranting designation as well established when utilizing criteria outlined by the Division 12 Task Force. Well established, according to these criteria, means that treatment must have demonstrated efficacy in at least two rigorous RCTs with superiority over placebo control conditions or another bona fide treatment. To date, more than 20 RCTs on full-model DBT have been conducted across multiple independent research teams, examining diverse populations, including veterans, community samples, individuals with eating disorders and substance use, and varying severities of BPD. DBT efficacy studies have been conducted in various countries, including the United States, Canada, Germany, the Netherlands, New Zealand, Spain, and Sweden. In addition, 15 RCTs on skills-only DBT have been conducted across the world, as more emerging literature highlights the centrality of skills training to the effectiveness of DBT. DBT has also been implemented and evaluated in inpatient, outpatient, school, forensic, military, and community settings. A comprehensive summary of DBT studies and adaptations is beyond the scope of this chapter; however, this section will review some recent research findings.

In a recent meta-analysis of 16 studies, 8 of which were RCTs across different research teams, DBT was associated most with reductions in suicidal and self-injurious behaviors. A more recent meta-analysis of five RCTs corroborated these findings. Although DBT was designed for individuals with BPD, research on standard DBT protocols has paved the way for more studies examining adaptations for a variety of other populations. DBT has demonstrated efficacy in the treatment of substance use disorders, eating disorders, chronic depression in older adults, and depression with co-occurring personality disorders. The use of DBT with substance use disorders leads to enhanced emotion regulation. DBT in the treatment of eating disorders can lead to decreased bingeing and purging behaviors. In depression, DBT has been found to lead to faster remission compared to medication alone. The effects of DBT may last after treatment ends. For example, at a 6-month follow-up, in one recent study, DBT was associated with reduced visits to the emergency room (ER).

In the treatment of posttraumatic stress disorder (PTSD), studies have yielded promising results suggesting that DBT is effective in the treatment of women who are actively self-harming, victims of sexual assault or harassment in the military, and adults with a history of childhood sexual abuse. For instance, DBT was found to lead to reductions in depressive symptoms and hopelessness among female victims of domestic abuse. More recent evidence has substantiated the effectiveness of DBT for PTSD (DBT-PTSD) to specifically target trauma treatment–resistant individuals. Patients undergoing DBT-PTSD, as compared to cognitive processing therapy, demonstrated lower dropout, higher rates of PTSD symptom remission, overall symptom improvement, and sustained recovery.

Recent research suggests that DBT is also effective in treating patients characterized by dysfunctional eating behaviors. This includes adaptations of DBT used to treat binge eating, as well as those designed to address co-occurring BPD and anorexia or bulimia. Another adaptation of DBT, Radically Open-DBT (RO-DBT), was associated in a pilot study with improvements in eating disorder–related psychopathology symptoms among patients with anorexia nervosa–restrictive type. A more recent multiple baseline single-case experimental design study demonstrated the efficacy of RO-DBT among 13 adult female patients with anorexia nervosa, with all completers meeting full remission after treatment.

DBT also has been adapted for children and adolescents. This development is important because the evidence suggests that DBT helps to reduce suicidal ideation and suicide risk behavior, and suicide is the third leading cause of death among youth aged 10 and 15 years. The first RCT of DBT for adolescents (DBT-A) demonstrated greater reductions in self-injurious behaviors compared to enhanced care as usual. A second multisite RCT indicated a significant advantage of DBT-A for reducing repeat suicide attempts, NSSI, and self-harm following treatment. To date, 5 RCTs, 3 controlled clinical trials, and 13 prepost evaluations of DBT-A have been conducted. Some of the ways DBT-A is different from standard DBT include the following: families are included in skills training, interventions directly target the invalidating environment (in family therapy), the length of treatment is abbreviated (from 1 year to 6 months), new skills are added (e.g., the "Middle Path"), and skills examples more relevant to teens are included in the treatment manual by Rathus and Miller. In school contexts, DBT-A has demonstrated promising results (e.g., reduction in the number of referrals to the assistant principal, detention/suspensions, and visits to the ER) in an open trial study conducted at a high school in New York. More recently, DBT-A has also been successfully applied to adolescents with bipolar disorder, street-involved youth, adolescents with binge eating, purging, NSSI, suicidal behavior, and BPD. In addition to DBT-A, DBT has also been adapted to help children (aged 8 to 11 years) to reduce self-injurious behaviors. Perepletchikova et al. found that the 6-week DBT skills group intervention led to significant decreases in depressive symptoms and suicidal ideation.

Despite its strong empirical foundation and growing list of promising adaptations, a number of gaps remain in the scientific literature examining DBT. Chief among the gaps in research is the lack of research on males and other gender, sexual, racial, and ethnic minorities. Similarly, the literature must address the increasing need to adapt DBT globally. In addition, there continues to be a need for research investigating what predicts treatment response via component, dismantling, or process-analytic studies, as well as large-sample effectiveness research in community settings. Finally, as recent modifications of DBT to diagnoses other than BPD have yielded promising results, future research will likely increasingly target transdiagnostic processes, such as emotion regulation, that are common in, but not specific to, BPD.

FURTHER READINGS

American Psychiatric Association. *Practice Guidelines for the Treatment of Patients with Borderline Personality Disorder*. American Psychiatric Association; 2001.

Aviram RB, Brodsky BS, Stanley B. Borderline personality disorder, stigma, and treatment implications. *Harv Rev Psychiatry*. 2006;14:249–256.

Axelrod SR, Perepletchikova F, Holtzman K, Sinha R. Emotion regulation and substance use frequency in women with substance dependence and borderline personality disorder receiving dialectical behavior therapy. *Am J Drug Alcohol Abuse*. 2011;37:37–42.

Bohus M, Kleindienst N, Hahn C, et al. Dialectical behavior therapy for posttraumatic stress disorder (DBT-PTSD) compared with cognitive processing therapy (CPT) in complex presentations of PTSD in women survivors of childhood abuse: a randomized clinical trial. *JAMA Psychiatry*. 2020;77(12):1235–1245.

Bornovalova MA, Daughters SB. How does dialectical behavior therapy facilitate treatment retention among individuals with comorbid borderline personality disorder and substance use disorders? *Clin Psychol Rev*. 2007;27:923–943.

Buerger A, Fischer-Waldschmidt G, Hammerle F, von Auer K, Parzer P, Kaess M. Differential change of borderline personality disorder traits during dialectical behavior therapy for adolescents. *J Pers Disord*. 2019;33(1):119–134.

Carmel A, Fruzzetti AE, Rose ML. Dialectical behavior therapy training to reduce clinical burnout in a public behavioral health system. *Community Ment Health J*. 2014;50:25–30.

Fischer S, Peterson C. Dialectical behavior therapy for adolescent binge eating, purging, suicidal behavior, and nonsuicidal self-injury: a pilot study. *Psychotherapy*. 2015;52(1):78–92.

Glisenti K, Strodl E. Cognitive behavior therapy and dialectical behavior therapy for treating obese emotional eaters. *Clin Case Stud*. 2012;11(2):71–88.

Goldstein TR, Fersch-Podrat RK, Rivera M, et al. Dialectical behavior therapy for adolescents with bipolar disorder: results from a pilot randomized trial. *J Child Adolesc Psychopharmacol*. 2015;25(2):140–149.

Gunderson JG. *Borderline Personality Disorder: A Clinical Guide*. American Psychiatric Pub; 2009.

Harned MS, Jackson SC, Comtois KA, Linehan MM. Dialectical behavior therapy as a precursor to PTSD treatment for suicidal and/or self-injuring women with borderline personality disorder. *J Trauma Stress*. 2010;23:421–429.

Hörz S, Zanarini MC, Frankenburg FR, Reich DB, Fitzmaurice G. Ten-year use of mental health services by patients with borderline personality disorder and with other axis II disorders. *Psychiatr Serv*. 2010;61:612–616.

Kliem S, Kröger C, Kosfelder J. Dialectical behavior therapy for borderline personality disorder: a meta-analysis using mixed-effects modeling. *J Consult Clin Psychol*. 2010;78:936–951.

Koons CR, Robins CJ, Tweed JL, et al. Efficacy of dialectical behavior therapy in women veterans with borderline personality disorder. *Behav Ther*. 2001;32:371–390.

*Linehan MM. *Cognitive Behavioral Treatment of Borderline Personality Disorder*. Guilford Press; 1993.

*Linehan MM. *DBT Skills Training Manual*. Guilford Press; 2014.

*Linehan MM, Armstrong HE, Suarez A, Allmon D, Heard HL. Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Arch Gen Psychiatry*. 1991;48:1060–1064.

*Linehan MM, Comtois KA, Murray AM, et al. Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Arch Gen Psychiatry*. 2006;63:757–766.

Linehan MM, Heard HL, Armstrong HE. Naturalistic follow-up of a behavioral treatment for chronically parasuicidal borderline patients. *Arch Gen Psychiatry*. 1993;50:971–974.

Linehan MM, Schmidt H 3rd, Dimeff LA, Craft JC, Kanter J, Comtois KA. Dialectical behavior therapy for patients with borderline personality disorder and drug-dependence. *Am J Addict*. 1999;8:279–292.

Linehan MM, Tutek DA, Heard HL, Armstrong HE. Interpersonal outcome of cognitive behavioral treatment for chronically suicidal borderline patients. *Am J Psychiatry*. 1994;151:1771–1776.

Lynch TR, Cheavens JS, Cukrowicz KC, Thorp SR, Bronner L, Beyer J. Treatment of older adults with co-morbid personality disorder and depression: a dialectical behavior therapy approach. *Int J Geriatr Psychiatry*. 2007;22:131–143.

Lynch TR, Gray KLH, Hempel RJ, Titley M, Chen EY, O’Mahen HA. Radically open-dialectical behavior therapy for adult anorexia nervosa: feasibility and outcomes from an inpatient program. *BMC Psychiatry*. 2013;13:293.

McCay E, Carter C, Aiello A, et al. Dialectical behavior therapy as a catalyst for change in street-involved youth: a mixed methods study. *Child. Youth Serv Rev*. 2015;58:187–199.

Mehlum L, Tørmoen AJ, Ramberg M, et al. Dialectical behavior therapy for adolescents with repeated suicidal and self-harming behavior: a randomized trial. *J Am Acad Child Adolesc Psychiatry*. 2014;53:1082–1091.

Neacsiu AD, Eberle JW, Kramer R, Wiesmann T, Linehan MM. Dialectical behavior therapy skills for transdiagnostic emotion dysregulation: a pilot randomized controlled trial. *Behav Res Ther*. 2014;59:40–51.

Panos PT, Jackson JW, Hasan O, Panos A. Meta-analysis and systematic review assessing the efficacy of dialectical behavior therapy (DBT). *Res Soc Work Pract*. 2014;24:213–233.

Paris J. *Treatment of Borderline Personality Disorder: A Guide to Evidence-Based Practice*. Guilford Press; 2008.

Perepletchikova F, Axelrod SR, Kaufman J, Rounsaville BJ, Douglas-Palumberi H, Miller AL. Adapting dialectical behaviour therapy for children: towards a new research agenda for paediatric suicidal and non-suicidal self-injurious behaviours. *Child Adolesc Ment Health*. 2011;16:116–121.

Safer DL, Telch CF, Agras WS. Dialectical behavior therapy for bulimia nervosa. *Am J Psychiatry*. 2001;158:632–634.

Sayrs JH, Linehan MM. *DBT Teams: Development and Practice*. Guilford Publications; 2019.

Skodol AE, Gunderson JG, Pfohl B, Widiger TA, Livesley WJ, Siever LJ. The borderline diagnosis I: psychopathology, comorbidity, and personality structure. *Biol Psychiatry*. 2002;51:936–950.

Telch CF, Agras W, Linehan MM. Dialectical behavior therapy for binge eating disorder. *J Consult Clin Psychol*. 2001;69:1061–1065.

van den Bosch LMC, Verheul R, Schippers GM, van den Brink W. Dialectical behavior therapy of borderline patients with and without substance use problems. Implementation and long-term effects. *Addict Behav*. 2002;27:911–923.

Van Dijk S, Jeffrey J, Katz MR. A randomized, controlled, pilot study of dialectical behavior therapy skills in a psychoeducational group for individuals with bipolar disorder. *J Affect Disord*. 2013;145:386–393.

32.12 Solution-Focused Brief Therapy

Anne Bodmer Lutz, B.S.N., M.D.

Solution-focused brief therapy (SFBT) is a short-term goal-focused evidence-based therapeutic approach that helps clients change by constructing solutions rather than dwelling on problems. In the most basic sense, SFBT is a hope-friendly, positive emotion–eliciting, future-oriented vehicle for formulating, motivating, achieving, and sustaining desired behavioral change. This approach to treatment is in accordance with the positive psychology movement that emphasizes well-being and optimal functioning instead of pathology and etiology. Empirically derived rather than theoretically based both in construct and practice, the SFBT approach is a goal-directed collaborative approach to psychotherapeutic change primarily conducted through direct clinical observation of clients' responses to a series of carefully constructed questions. SFBT has continued to grow in popularity, both for its usefulness and its brevity, and is currently one of the leading schools of psychotherapy in the world. Besides psychotherapy, solution-focused (SF) thinking has become a major influence in such divergent fields of education, social policy, and business.

SFBT starts by developing a detailed understanding of the patient's best hope for their future and collaboratively looks for patient-related resources and actions that build this outlook. The focus is on detailing what a patient will be doing when their problem is solved rather than diagnosis and symptom exploration. Based on the premise that people have the necessary resources to solve their problems, SFBT amplifies these strengths and abilities by building a shared dialect that focuses on what has worked and is working in a patient's life. Perhaps, there was a time that a patient was able to manage their pain and get to work on time, even if only 1 day that week. Perhaps, there was a day that a patient was able to get up and go for a walk, even if only for 10 minutes. Perhaps there was a day they were able to manage their anxiety and depression in a tolerable way. These moments are easily missed. The SFBT is skilled at listening for these moments and amplifying with patients how they were able to have these times of success. This awareness enhances hope as patients have already had success and thus can do it again.

SFBT focuses on the patient's expertise in knowing their condition best. Asking patients (and their loved ones) what they know about their condition and what they have tried that helps attest to their competence. Many patients' problems can be relapse remitting and/or change over time which can lead to a sense of uncontrollability. SFBT helps patients to focus on how they have adapted and managed to deal with their symptoms and condition in a "tolerable/bearable" way. This increases their self-efficacy in managing their condition.

EMPIRICAL EVIDENCE

SFBT is one of the few approaches in psychotherapy that began as "evidence based" versus being "theory driven." SFBT developers Steve de Shazer and Insoo Kim Berg and their team spent countless hours carefully observing real-life therapy sessions, meticulously noting any words or behavior on the part of the therapist that resulted in progress toward the client's stated goal. Language or behavior that reliably led to positive movement in the direction of the client's goal was preserved and incorporated into the SF approach, and those that did not were discarded.

There has been a great deal of empirical research on SFBT, most of it occurring over the past 15 years. The conclusion of eight meta-analyses, a number of systematic reviews, and the overall conclusion of the most recent scholarly works on SFBT are that SFBT is an effective approach to the treatment of psychological problems, with effect sizes similar to other evidence-based approaches, such as cognitive behavior therapy and interpersonal therapy, but that these effects are found in fewer average sessions, and using an approach style that is more benign. That is, the more collegial and collaborative approach of SFBT does not involve confrontation or interpretation. Given its equivalent effectiveness, shorter duration, and more benign approach, SFBT is considered to be an excellent first-choice evidence-based psychotherapy approach for most psychological, behavioral, and relational problems.

- There have been approximately 150 randomized clinical trials (RCTs, the "gold standard" of clinical research).
- Seventy-four percent of SFBT RCT studies reported significant positive benefit.
- There are been eight meta-analyses.
- Most of the research was done in "real-world" settings, so the results are more generalizable.
- Effect sizes found in these meta-analyses are in the low to high range, suggesting that SFBT is an effective approach for the populations studied.
- SFBT's effectiveness is equivalent to other evidence-based practices, such as cognitive behavior therapy and interpersonal psychotherapy.
- While effect sizes are similar to other evidence-based approaches, these effects are attained in fewer sessions, averaging about 5 sessions and rarely extending over 8 or 10.

CORE SFBT SKILLS

Engaging in SFBT is similar to learning a new language because it requires therapists to "speak" in a way that focuses on solutions rather than their "first" language, which tends to focus on problems. The order of the SF sessions is also important. Early and throughout the interview, rigorous attention is paid to engagement through activating resources, including the patient's individual and social strengths. This is followed by goal negotiation, exploring best hopes for treatment, and utilizing scaling questions to develop small "doable" next steps to move forward. See Figure 32.12–1 for a visual representation of the key components of SFBT.

Engagement-Activating Resources From the Start

Activating resources at the onset of the conversation is critical and helps build patient engagement. *Resource activation* is realized in interventions that focus not on the patient's problems but rather on the sound and healthy parts of their life.

Example questions to promote the activation of resources:

- What do you enjoy?
- What are you good at?
- How did you learn these skills?
- How else did you learn them?

Engagement Complimenting

Patients coping with complex challenges and conditions endure many symptoms affecting the quality of their daily life. Asking patients what they know and have tried regarding their diagnosis, condition, treatments, and medications affirms their knowledge in managing their condition. Compliments can impart much-needed energy and stamina to both patients and their loved ones. Most clinicians are fluent in what is known as the direct compliment with phrases such as “Wow!” “Congratulations!” “Good job!” Clinicians are much less fluent in the powerful and highly effective indirect compliment. Indirect compliments are in the form of a question and are used early and throughout an SF conversation.

Examples of indirect compliments:

- How did you do it (e.g., manage your pain, get to work)?
- How did you decide to do it (e.g., to take your medication, to make healthy food choices, to get the rest you require)?
- How did you choose to do it (e.g., to stop smoking, to seek help)?
- How did you manage to do it (e.g., get a better night’s sleep, get to your medical appointment)?

Engagement–Building the “Yes-set”: “For You” Statements

The yes-set is a SF skill that involves creating a conversation in which both the clinician and patient say yes and agree on as many aspects of the conversation as possible. “For you” statements enhance the “yes-set” while providing validation and acknowledgment of the patient’s predicament. Patients and families coping with complex conditions often experience feelings of lack of control, exhaustion, fear, frustration, and anxiety, requiring a great deal of “for you” responses. “For you” statements are often combined with coping questions, enabling patients to recognize how they have been managing.

Yes-set “for you” responses:

- This must be very difficult “for you.”
- This must be so frustrating “for you.”
- This must be painful “for you.”
- This must be grueling “for you.”

Coping Questions

Coping questions are based on the appreciation of how painful and intense life experiences are for many patients. Coping questions are especially helpful in shifting a patient’s narrative from one of victim and self-blame to one of courage, strength, and self-efficacy. Incorporating the one-word “cope” within the question attests that patients already have managed to have some measure of success. Other verbs can also be used in coping questions, such as endure, manage, deal with, hold up, persevere, bear, and tolerate.

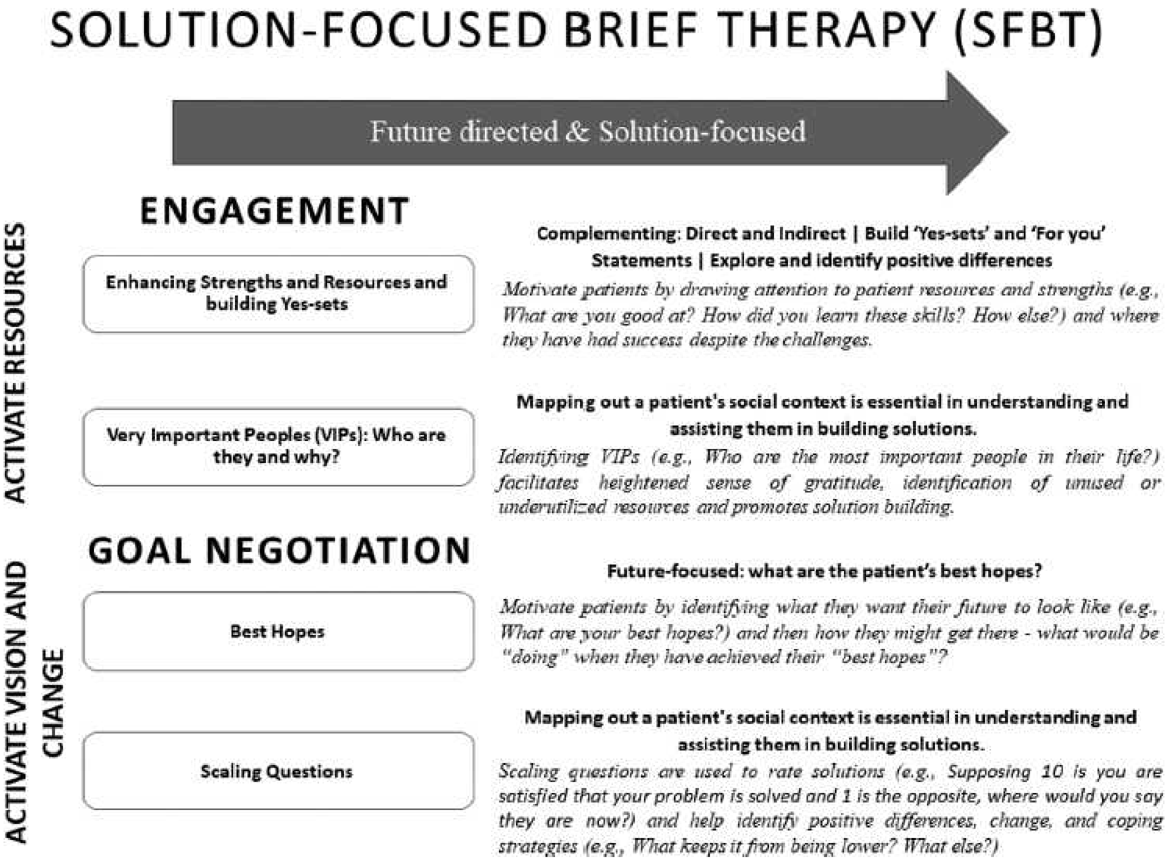
SOLUTION-FOCUSED BRIEF THERAPY (SFBT)
Future directed & Solution-focused
ACTIVATE RESOURCES
ENGAGEMENT
Enhancing Strengths and Resources and building Yes-sets
Very Important Peoples (VIPs): Who are they and why?
ACTIVATE VISION AND CHANGE
GOAL NEGOTIATION
Best Hopes
Scaling Questions

Complementing: Direct and Indirect | Build 'Yes-sets' and 'For you' Statements | Explore and identify positive differences

Motivate patients by drawing attention to patient resources and strengths (e.g., What are you good at? How did you learn these skills? How else?) and where they have had success despite the challenges.

Mapping out a patient's social context is essential in understanding and assisting them in building solutions.

Identifying VIPs (e.g., Who are the most important people in their life?) facilitates heightened sense of gratitude, identification of unused or underutilized resources and promotes solution building.

Future-focused: what are the patient's best hopes?

Motivate patients by identifying what they want their future to look like (e.g., What are your best hopes?) and then how they might get there - what would be "doing" when they have achieved their "best hopes"?

Mapping out a patient's social context is essential in understanding and assisting them in building solutions.

Scaling questions are used to rate solutions (e.g., Supposing 10 is you are satisfied that your problem is solved and 1 is the opposite, where would you say they are now?) and help identify positive differences, change, and coping strategies (e.g., What keeps it from being lower? What else?)

FIGURE 32.12–1. Key components of solution-focused brief therapy.

Coping questions:

- How have you coped with this? How else?
- How have you endured this? How else?
- How have you managed? How else?
- How have you tolerated this? How else?
- Where do you get your strength? Where else?

Effective solution building requires getting as many details as possible about prior successes. Questions that amplify details of positive coping (e.g., *How else* did you manage your symptoms? *How else* was it helpful?) leave no potential resource uncovered. It is also important to ask patients about their "good reasons" for seemingly unhelpful behaviors (e.g., medication nonadherence), as these "good reasons" can reveal how patients view these behaviors as beneficial. The question does not condone the behavior. Rather it helps patients understand their own motivation and elucidate what they could do differently. Examples include:

"You must have a good reason to … not take your medications/make it to your appointments/follow-up with treatment recommendations."

Punctuating Positive Differences

Positive differences, also known as exceptions, are those times when the expected problem could have occurred but did not or was less severe. Listening for *past successes* and discovering positive differences when patients have had or are having some measure of success is crucial. These times may include when patients were able to have periods of relief from their symptoms—even if very brief. Perhaps they were able to manage a specific symptom in a tolerable way. Perhaps they had the strength to work and manage their family responsibilities despite their symptoms. Perhaps they persevered in their efforts to care for their needs and remembered to take their medications.

Questions to identify and amplify positive differences:

- Was this different for you?
- How was it different for you?
- Was it helpful for you?
- How was it helpful for you?
- How did you do it?
- How else did you do it?

Identify Those Who Are Most Important in Their Life (VIPs)

Inviting a patient to identify their VIPs and what they most appreciate about them is especially critical when engaging patients who are coping with complex conditions. Mapping out a patient's social context is essential in understanding and assisting them in building solutions. Identifying VIPs also helps patients broaden their perspective about possible unused or underutilized resources in their situation.

Exploring VIPs:

- Who are the most important people in your life?
- Who else?
- What do you most appreciate about them?
- What else?
- Who has been most helpful for you in managing your situation?
- Who was concerned about you that they thought coming here would be helpful for you?
- Who relies on you?
- What do your VIPs know about you that you can get through this? What else?

Goal Negotiation—Exploring Best Hopes for Their Preferred Future

SFBT's translation of the chief complaint is, "What are your best hopes for this appointment so you can say it was helpful and worth your time?" This question is future directed, assists in goal negotiation, and conveys that patients *do have hopes*.

Asking for a patient's "best hopes" is only the beginning of goal negotiation. Moving from their "best hopes" to specific actions requires discovering with patients what they would be "doing" when they have achieved their "best hopes"? Many patients are more fluent in what they "don't want" and have greater difficulties describing what they "do" want. A patient may say their best hopes are to not have anxiety or pain. Asking patients what they would be *"doing instead" when they are managing* their pain and anxiety in a good enough way can assist patients in what they have control over.

Inviting patients to imagine a satisfying week is another way to explore their best hopes. Supposing when you wake up on a Sunday morning, and you look back at your week, you say to yourself, that was a satisfying week. I managed my condition in a "tolerable" "good enough" way. What would you have done? What else? What would others have noticed you did? What else? An important question is, "what else." This question compels the patient to persevere in detailing as yet undiscovered positive differences and prior successes.

Questions to explore “best hopes”:

- What are your best hopes for this appointment so you can say it was helpful and worth your time?
- How can I be most helpful for you today so that this meeting is worthwhile?
- What do you know about your condition? What else?
- What do you know about yourself that you can get through this? What else?
- What do your VIPs know about you that you can get through this? What else?
- What have you tried to help with your condition? What else?
- What will you be “doing instead” when your problem is solved? What else will you be “doing instead”?

Goal Negotiation–Scaling Questions

Scaling questions invite patients to rate their goals on a numerical scale from 1 to 10. Scaling questions are used to rate solutions, different from many other scales which rate problems. They are developed from the patient’s best hopes and can be amplified by asking patients to predict how their VIPs would rate them. Scaling questions convey empathy, are patient centered, and can assist patients and clinicians in understanding perspectives from multiple viewpoints quickly. They are extremely useful in even brief medical encounters.

It is essential to amplify scaling questions. “Working the scale” provides additional opportunities to activate resources. Asking patients what keeps the number from being lower, and what else? What’s the highest number it has been? What would be a “good enough” number? Patients often do not need the “best” but are very satisfied with “good enough.”

Scaling questions:

On a scale from 1 to 10, where 10 is the best:

- How well are you managing your condition?
- How helpful are your medications and current treatments?
- How well are you (coping/enduring/dealing) with your condition?
- How satisfied are you with how you are managing your daily tasks?
- How well are you remembering to take your medication?
- How well would your VIPs say you are coping with your condition, remembering to take your medications, and managing your daily task?
- How confident are you that you can manage your condition in a good enough way?
- How confident are your VIPs that you can manage your condition in a good enough way?

Working the scale:

- What's the highest it's been (another way to explore positive differences)?
- What's a good enough number?
- What would you be doing at that "good enough" number?
- What else?
- What keeps the number from being lower? What else?
- What would you be doing when the number goes up by one point?
- What else?
- What number would your VIPs give you?

End of Session

In SFBT, the end of a session is embedded with the compliments, positive differences, VIPs, and successes that were discovered in the conversation. Scaling questions challenge patients to consider the next small steps to take toward achieving their goals. Feedback should also confirm that there is no other information patients think would be helpful for the clinician to know in order to be most helpful for them.

Follow-Up Sessions

Follow-up sessions should begin with searching for positive differences and opportunities to amplify what is already working in a patient's life. If things are worse, offer the patient enough “for you” responses, followed by coping questions. Follow-up sessions also include exploring their best hopes for the appointment, illuminating positive differences, incorporating VIPs, and asking scaling questions.

Follow-up session questions:

- What’s been better since we last met? What else?
- What’s been even a little bit more tolerable since we last met? What else?
- What’s been happening that you want to continue to happen? What else?
- If things are worse: This must be very difficult “for you,” how have you been coping/enduring/dealing with this? How else?
- What are your best hopes for today’s appointment that would tell you it was helpful for you? What else?
- What have you tried?

Scaling questions from 1 to 10 (10 being the best):

- How well are you managing your pain/nausea/diarrhea/constipation/bloating/gas/heartburn?
- How well are you managing your anxiety/stress?
- How well are you remembering to take your medication?
- How well are they managing to get through their daily tasks?
- How well would your VIPs rate these areas from 1 to 10?

Working the scale:

- What would be a good enough number?
- What keeps it from being lower? What else?
- What would they be doing when it goes up by one point?

Pearls of wisdom:

- Translate language to highlight solutions.
- Notice opportunities to provide indirect compliments, attend and amplify positive differences, use and explore patient’s exact words, and provide “for you” statements bridged to coping questions all enhance treatment engagement and promote self-efficacy.
- Assess a patient’s VIPs to broaden “who” the patient is by detailing their social context to enhance possibilities for building solutions.
- Explore a patient’s vision and their “best hopes” to assist them in detail with what they will be doing when their problem is solved.
- Scale questions quickly to enhance goal negotiation, detail what has worked, and the next actionable steps to move forward. Questions are patient centered, collaborative, and empathic.

SUMMARY

SFBT is an evidence-based approach that promotes change by building on a patient's strengths, exploring past successes, and assisting in developing with patients a clear vision of what they hope to do moving forward when they are managing their condition in a "good enough" way. It is a pragmatic patient-centered collaborative approach based on solution building rather than problem solving. This approach has a belief in patient competence, assumes patients have the necessary resources to live a more satisfying life, and enhances hope and possibilities.

FURTHER READINGS

Fiske H. *Hope in Action: Solution-Focused Conversations about Suicide*. Routledge; 2008.

Franklin C, Trepper TS, Gingerich WJ, McCollum EE. *Solution-Focused Brief Therapy: A Handbook of Evidence-Based Practice*. Oxford University Press; 2012.

Franklin C, Zhang A, Froerer A, Johnson S. Solution focused brief therapy: a systematic review and meta-summary of process research. *J Marital Fam Ther.* 2017;43(1):16–30.

Froerer AS, Walker CR, Kim JS, Connie EE, von Cziffra-Bergs J. Language creates a new reality. In: Froerer AS, von Cziffra-Bergs J, Kim JS, Connie EE, eds. *Solution-Focused Brief Therapy with Clients Managing Trauma*. 2018:24–47.

Gingerich WJ, Peterson LT. Effectiveness of solution-focused brief therapy: a systematic qualitative review of controlled outcome studies. *Res Soc Work Pract*. 2013;23:266–283.

Kim JS. Examining the effectiveness of solution-focused brief therapy: a meta-analysis. *Res Soc Work Pract*. 2007;18(2):107–116.

Kim JS, Froerer AS. Intersection of SFBT and trauma. In: Froerer AS, von Cziffra-Bergs J, Kim JS, Connie EE, eds. *Solution-Focused Brief Therapy with Clients Managing Trauma*. 2018:10–23.

Kim J, Jordan SS, Franklin C, Froerer A. Is solution-focused brief therapy evidence-based? An update 10 years later. *Fam Soc*. 2019;100(2):127–138.

Kim J, Kelly M, Franklin C. Solution-focused brief therapy in schools. Oxford Scholarship Online, 2017.

Lutz AB. *Learning Solution-Focused Therapy: An Illustrated Guide*. American Psychiatric Publishing, a Division of American Psychiatric Association; 2014.

Shazer SD, Dolan Y, Korman H, Trepper T, McCollum E, Berg IK. *More than Miracles: The State of the Art of Solution-Focused Brief Therapy*. Routledge; 2021.

Zhang A, Franklin C, Currin-McCulloch J, Park S, Kim J. The effectiveness of strength-based, solution-focused brief therapy in medical settings: a systematic review and meta-analysis of randomized controlled trials. *J Behav Med*. 2018;41(2):139–151.

32.13 Single-Session Therapy

MICHAEL F. HOYT, PH.D.

"A journey of a thousand miles begins with a single step."

—Lao Tzu (6th century BCE)

"Remember then: there is only one time that is important – Now!
It is the most important time, because it is the only time
when we have any power."

—Leo Tolstoy (1885, *What Men Live By, and Other Tales*)

"… the fierce urgency of NOW …"

—Dr. Martin Luther King, Jr., "I Have a Dream" (28 August 1963)

The term *single-session therapy* (SST) refers to therapy in which each session is approached with the expectation that there *may* only be the one therapeutic meeting. There is the option that more sessions may follow, but it is a therapy that is understood, from the beginning, to be deliberately a single session. It is not just therapy that could have gone on (and on …) and happens to end after one visit. It is one session *by plan*, not by default. Each session is *one-at-a-time (OAAT).* The basic concept is to "view each encounter as a whole, complete in itself."

As Hymmen, Stalker, and Cait have succinctly written:

> *SST refers to a conscious approach to make the most of the first session knowing it may be the only session the [patient/]client decides to attend – not to the situation where there is an expectation that the [patient/]client will attend multiple sessions but chooses to attend just one.*

The idea that one session could be enough—and often is—may seem contrary to and disruptive of a more-is-better ethos, but there is abundant quantitative and qualitative that a single session of therapy is *de facto* the most common length of treatment, and that SST is often effective with a wide variety of clinical problems (including anxiety, depression, eating disorders, sleep disturbances, autism-spectrum issues, addictions and substance misuse, and family conflicts). SST does not claim to cure or solve all problems in one visit, of course, but even with conditions likely to be long-term or chronic, some patients may prefer to be seen on a OAAT basis, and "continuity of care" and treatment compliance may be facilitated by honoring those preferences via the availability of as-needed focused episodes. There is clear and growing support for the value of SST with children and adolescents. There are also assorted health-psychology single-session programs and protocols (e.g., stress reduction, stop smoking, improve your sleep) with empirical support of efficaciousness.

Sometimes a problem is "solved" or "resolved"—other times, things just get "unstuck" and "processed" and maybe coped with or managed better. A patient may use a single session to make a significant life change, take a step in the right direction, consider options, and/or validate their efforts, or perhaps gain support and avoid a worsening of symptoms. Each session is bespoke: one size does not fit all. Thus, Hoyt, Rosenbaum, and Talmon noted:

> *The therapist needs to be versatile, innovative, and pragmatic, asking: "What would help this patient today"? Patients may need to begin a process or complete a process; they may need to take hold or to let go; they may need reassurance or confrontation; they may need to look at something deeply or to shift perspective. […] Nothing works all the time, but what might work this time?* (Used with permission from Hoyt MF, Rosenbaum R, Talmon M. Planned single-session psychotherapy. In: Budman SH, Hoyt MF, Friedman S, eds. *The First Session in Brief Therapy.* Guilford Press; 1992:63.)

There are certain common practices—discussed below—that increase the likelihood of successful results in a single session. Without suggesting that one session is always enough, these can inform and enhance a range of clinical interventions—not just those intended to be potentially "stand-alone" single sessions. Evidence about treatment contacts clearly documents that many patients will only come once, whether we like it or not, so it is best to make the most of that session. However, even if approached as if one meeting may be enough, OAAT does not necessarily mean "only one time." SST patients may sometimes return for additional assistance regarding the same or another problem.

Although there are many positive SST reports drawn from different theories of psychotherapy, SST/OAAT may be thought of more as a format or delivery system (Young, 2018) and an affirmative and hopeful expectation and mindset rather than as a particular theoretical or technical approach. Making the possibility of a single session available has been shown to be very helpful in addressing waiting lists and in making services more available to more people.

Single-session thinking is not necessarily restricted to only one session. Rather, it can be useful for approaching and organizing each session—

establishing an alliance and determining session goals, identifying and applying resources, staying on task, planning next steps, and follow-through—even when there are subsequent meetings. Single-session thinking can also inform other types of consultations, including educational and career counseling, hospital social work, sports coaching, and postdisaster responses; and can help organize the agenda of each session of ongoing meetings.

A BRIEF HISTORY OF SST

Throughout the psychotherapy literature, there are scattered case reports of successful one-session treatments. Indeed, several of Freud's early cases involved a single session. He met one time, for example, with an emotionally upset young woman named Katharina while he and his family were on vacation in the Alps (the innkeepers were Katharina's parents); and, in another instance, he went for a long walk around Leyden, Holland with the great music composer, Gustav Mahler, who had requested a meeting while they were both there to discuss sexual and marital problems that seemed related to issues Mahler had with his mother. We don't know exactly what Katharina and Freud, and Mahler and Freud, talked about, but we do know that they knew they would only have one meeting, and we know from reports that the single sessions were helpful. A myriad of subsequent positive SST reports has come from cognitive-behavioral therapy (CBT), psychodynamics, solution-focused brief therapy (SFBT), narrative therapy, motivational interviewing, strategic interactional therapy, hypnotherapy, structural family therapy, and others.

Statistical studies have repeatedly documented that one session, usually unplanned, is the most common length of treatment. (Actually, *no* session is the most common length of treatment—people think about seeing a psychiatrist or psychologist, but never do—and "no treatment" may be the prescription of choice for those who can solve their problems without professional help.) Researchers have also repeatedly demonstrated the *medical utilization offset phenomenon*, the finding that unnecessary medical utilization is often reduced after one session of psychotherapy.

The Kaiser SST Study

In the late 1980s, Moshe Talmon, a psychologist then working at the Kaiser Permanente Medical Center (a large HMO) in Hayward, California, noticed that many clinic patients who were seen for a single outpatient visit did not return for further sessions—regardless of the staff therapist's respective theoretical orientation or the patient's diagnosis. Intrigued, Talmon reviewed charts and contacted patients for follow-up. Despite the anticipation that many seen-only-once patients would be dissatisfied, it was found that the great majority actually expressed satisfaction with their session and indicated that they had not felt a need for additional meetings. This was very different from what was expected and was termed *unplanned SST*.

Encouraged by the retrospective finding of one-session successes, Talmon then collaborated with two psychologist colleagues (Robert Rosenbaum and Michael Hoyt) to conduct the first prospective study of possible (planned) SST. A series of outpatients, ages 8 to 80 with a wide range of potential diagnoses, were told on the telephone while making their appointment and then at the beginning of the session, that an effort would be made to provide what they wanted in the one meeting and that, at the end of the session, they would be asked whether the session had been sufficient or if they wanted to return for more meetings.

Early in each session, the therapist would say something along these lines:

"Hi! I'm Dr. ______. The purpose of our meeting today is to find out what brings you here and what would be helpful. Sometimes we may be able to get everything dealt with today in one visit; if we are not able to, then we will figure out what the next steps might be. Are you interested in that, eh?" (*Comment*: Insistence produces resistance, imposition produces opposition, push produces pushback—so it is important to offer and invite, but not demand, a one-session format.)

And at the end of the session:

"Would you like to wait and see how it goes when you use some of what we've discussed—or would you prefer to make another appointment now?"

Comment: Suggesting waiting until what has been discussed is used emphasizes that the patient has an active role in making desired changes.

The basic findings of the Kaiser study were:

- 34 of 58 patients (58.6%) elected to complete their therapy in one session even when more sessions were available;

- more than 88% of the one-session patients reported significant improvement in their original “presenting complaint,” with more than 65% also reported “ripple” improvements in related areas of functioning; and
- while not experimentally assigned to one session or longer, on follow-up, there was no difference in satisfaction and outcome scores between those who chose to stop after one visit versus those who continued for more sessions.

Other Early Single Sessions

In the late 1960s, walk-in psychological (and medical) services were founded at the Los Angeles Free Clinic, at the Haight-Ashbury (San Francisco) Free Clinic, and at the Minneapolis Walk-In Counseling Clinic. These sessions often involved single visits. In the early 1990s, clinicians at the Eastside Family Centre in Calgary, Canada developed planned single-session walk-in services. Walk-in SST services were subsequently taken to Our Lake of the Lake University in San Antonio, Texas, where they have been taught to graduate students, provided to the community, and promulgated in the United States and Mexico. By 1994, The Bouverie Centre at La Trobe University in Melbourne, Australia began to offer extensive by-appointment SSTs and began training large numbers of professionals. In the mid-2000s, The Bouverie Centre, recognizing that many SST workshop participants (such as youth workers and vocational counselors) found single-session ideas very relevant and helpful but did not identify as clinical therapists, broadened the terminology and began referring to *Single Session Work* (SSW). In 2019, The Bouverie introduced the umbrella term *single-session thinking* to refer to the mindset common to all single session approaches—the characteristics of which are discussed in what follows.

So far, four international SST and Walk-In Services conferences have been held in:

1. March 2012, on Phillip Island near Melbourne, Australia
2. September 2015, in Banff, Canada
3. October 2019 in Melbourne
4. June 2023 in Rome

Reports of successful SSTs have come from the United Kingdom, Sweden, Ireland, Italy, Mexico, Haiti, France, Spain, Germany, China, Cambodia, South Korea, Chile, Canada, Israel, Turkey, New Zealand, Australia, the United States, and elsewhere. Public interest has also continued to expand; in 2018, an article featuring SST appeared in *O: The Oprah Magazine* (monthly circulation of 2.5 million copies).

SINGLE-SESSION THINKING: ONE AT A TIME

Although Talmon defined SST "as one face-to-face meeting between a therapist and patient with no previous or subsequent sessions within one year," that specification was somewhat arbitrary and was used for research purposes. It also did not emphasize the possibility of repeat single sessions. Hence, Hoyt's group proffered this more preferred general description:

> *The essence of single-session thinking is to approach the first session as if it will be the only session while creating opportunities for further work if it is requested by the client. What emerges is a collaborative, direct, and transparent approach to providing services that put the client in a very active role in determining the focus and length of the work.* (Used with permission of Hoyt MF, Young J, Rycroft, P. Single session thinking and practice going global one step at a time. In: Hoyt MF, Young J, Rycroft P. *Single Session Thinking and Practice in Global, Cultural, and Familial Contexts.* Routledge; 2021:3.)

One session may yield the positive results that patients are seeking—as discovered when Talmon followed up with many clinic patients who had only been seen once—but not recognizing that patients may have achieved what they wanted in the one session, these are sometimes (mistakenly) called *drop-outs* or *premature termination*.

Planned single sessions can occur either (1) *by appointment* or (2) *by walk-in* (which, with the advent and increasing use of teletherapy, may be via phone or the internet). In addition to private offices and mental health clinics, SSTs also often occur in primary-care medical settings, hospital and E.R. meetings, college counseling centers, inpatient psychiatric groups, second-opinion consultations, and clinical demonstration interviews. *Crisis intervention* is not generally considered an SST—unless that is what is indicated in a particular situation.

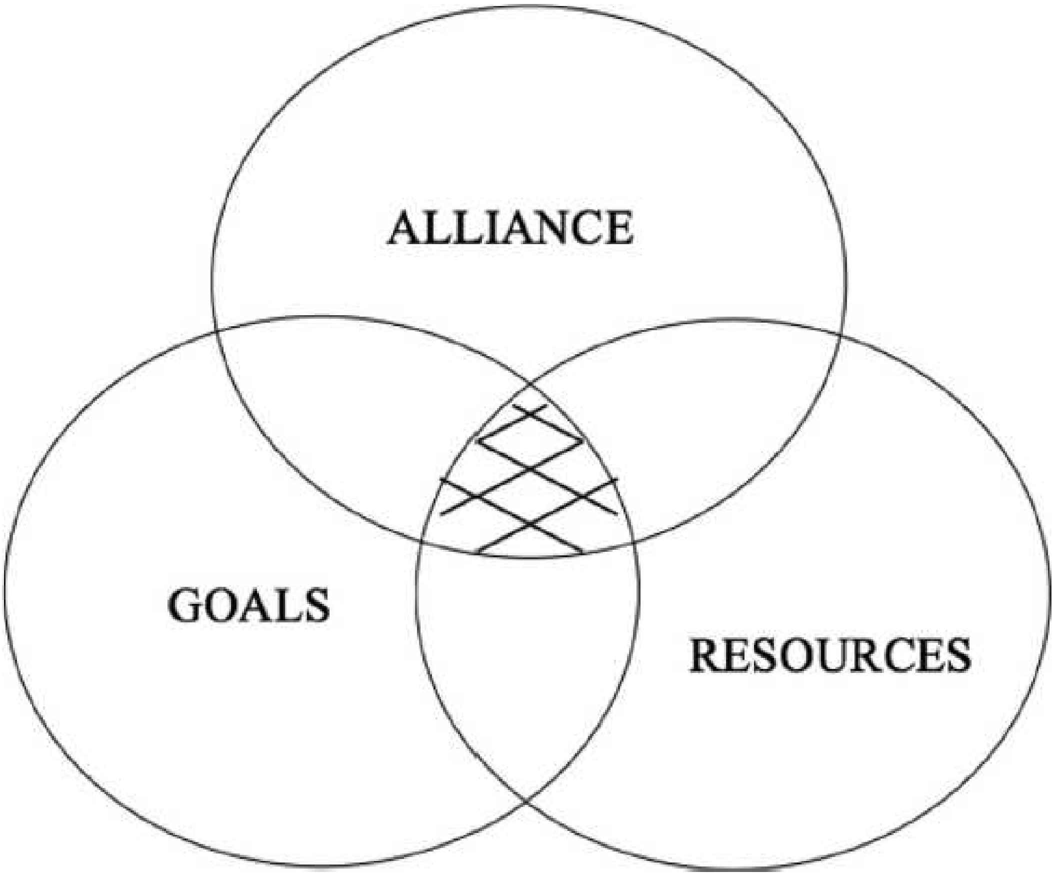

FIGURE 32.13–1. Context of competence.

Although the possibility of a single session may be available when therapy takes more than one session, it is not an “SST failure”—it just means it took more than one session.

SST/One-at-a-Time Basics

Regardless of how the patient gets there (by appointment or walk-in), once they have arrived, there are general principles that guide SST. When patients are asked what they want, the therapist and patient can work collaboratively to honor those hopes and intentions. Underlying effective single-session thinking and practice are issues of (1) *mindset*, (2) *time*, and (3) *empowerment/client activation*.

Four common themes thus cut across single-session thinking and practice:

1. *Attitude*—treating the session “as if” it might be the only one and hence making the most of every encounter, underpinned by the

paramount acceptance that one session could be (and often is!) enough.

2. *Accessibility*—responding in a timely manner without any unnecessary barriers to patients/clients receiving help when they are most ready.
3. *Acting now*—accepting that the best opportunity to address change is NOW, no matter the diagnosis, severity, or complexity of the problem.
4. *Alliance*—asking what patients/clients want to achieve by the end of the session so that the therapist and patient/client can work collaboratively, in the here and now, toward that goal.

Context of Competence. Effective therapy, whether one session or longer, occurs within a Context of Competence that we cocreate with the patient. As seen in Figure 32.13–1, therapy takes place where ALLIANCE/GOALS/RESOURCES overlap. If you identify GOALS (or PROBLEMS) and have an ALLIANCE but don't access RESOURCES, nothing happens. If you have an ALLIANCE and RESOURCES but don't identify GOALS (or PROBLEMS), there's nothing to do, no purpose. If the patient could match up their GOALS and RESOURCES on their own, bringing to bear competencies to deal with problems, they wouldn't need the therapist. Even if Shakespeare was right in *All's Well That Ends Well* (Act I, Scene1) when he wrote, "Our remedies oft in ourselves do lie," all three are often needed: the *Context of Competence* is where GOALS and RESOURCES come together through the ALLIANCE. Effective therapy, one session or otherwise, involves all three.

THE TEMPORAL STRUCTURE OF SESSIONS AND THERAPY

In SST, each session is a stand-alone, potentially complete-unto-itself "event." There is often a temporal structure to single sessions, with different issues and tasks associated with the early, middle, and late phases of a meeting (see Fig. 32.13–2).

The course of each therapy session can be conceptualized as having five phases or stages: pretreatment, early, middle, late, and follow-through. They tend to be pyramidal or epigenetic—that is, they tend to occur in sequence, and each flows from the preceding so that successful work in one preconditions the next. The patient elects therapy before forming an alliance; the alliance establishes the ground for defining goals;

goal setting leads to focusing on specific change strategies, homework, relapse prevention, and leave-taking; and points toward continued growth and a possible return to treatment.

Change begins even before meeting the patient—he/she/they have decided there is a problem and would like assistance. When the patient arrives in the office, it is important to establish a focus, finding out about any developments since the patient decided to seek therapy and emphasizing the patient's initiative. Mid-session involves various change techniques and keeping track of the patient's goals. Later in the session, termination—extracting the therapist from the successful equation—becomes central. Issues to be considered include goal attainment, homework/post-session tasks, relapse prevention, and leave-taking. Everything may not be resolved by the end of the session, of course, but the patient has learned or recalled some useful skills, has enhanced motivation, and ideas for how to move forward. Benefit and change continue after the meeting, and the patient can be invited to call or return as needed. Extensive lists of guideline SST questions associated with the different phases are provided in Hoyt (in preparation).

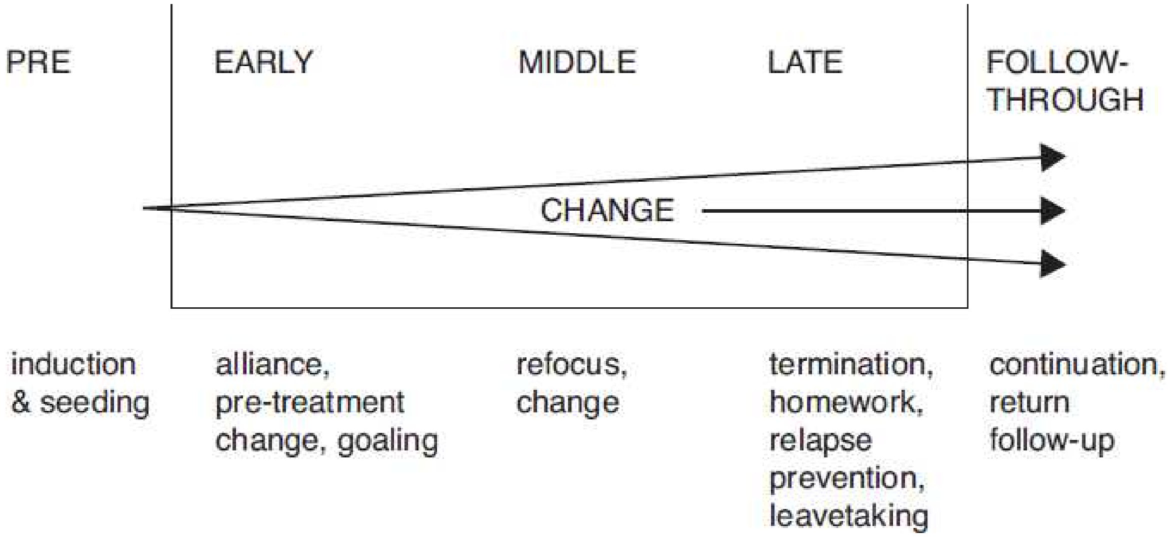
PRE
EARLY
MIDDLE
LATE
FOLLOW-THROUGH
CHANGE
induction & seeding
alliance, pre-treatment change, goaling
refocus, change
termination, homework, relapse prevention, leavetaking
continuation, return follow-up

FIGURE 32.13–2. The temporal structure of single-session therapy (SST).

KEY SINGLE-SESSION THINKING AND PRACTICE IDEAS

Over the years, a variety of useful general SST attitudes have been described. Bobele and Slive, for example, list a number of ideas that inform a OAAT mindset:

- Patients/clients know what works best for them.
- Patients/clients are far less interested in psychotherapy than therapists and usually prefer brief therapeutic encounters.
- Patients/clients frequently choose to attend only one session and overwhelmingly express satisfaction with that session.
- Research demonstrates that most change occurs early in therapy, followed by ever-decreasing improvements as sessions continue.
- Rapid change is not only possible but also common in human experience.

There are different ways to translate single-session thinking into practice. Various lists have been offered, all emphasizing the importance of clear goals and accessing resources. An early investigator, Bernard Bloom, for example, offered 15 SST transtheoretical suggestions:

1. Identify a focal problem
2. Do not underestimate patients' strengths
3. Be prudently active
4. Explore, then present interpretations tentatively
5. Encourage the expression of affect
6. Use the interview to start a problem-solving process
7. Keep track of time
8. Do not be overambitious
9. Keep factual questions to a minimum
10. Do not be overly concerned about the precipitating event
11. Avoid detours
12. Do not overestimate a patient's self-awareness (i.e., don't ignore what may seem obvious)
13. Help mobilize social supports
14. Educate when patients appear to lack information
15. Build in a follow-up plan.

Drawing from their experiences in the Kaiser SST study, Rosenbaum, Hoyt, and Talmon highlighted:

1. "Seed" change through induction and preparation

2. Develop an alliance by cocreating, with the patient/client, an attainable treatment goal
3. Allow enough time for the session to be a complete process or intervention
4. Look for ways to meet patients/clients in their worldview while, at the same time, offering a new perspective or hope about the possibility of seeing and acting differently
5. Go slowly and look for the patient’s/client’s strengths and resources
6. Focus on “pivot chords,” ambiguous situations that can be reframed in therapeutic ways
7. Practice solutions experientially, using the session to help patients/clients rehearse a solution or solutions—thus inspiring hope, readiness for change, and forward movement
8. Consider taking a time-out, break, or pause during the session to think, consult, focus, prepare, and punctuate
9. Allow time for last-minute issues to help the patient/client have the sense that the session has been complete and satisfactory
10. Give feedback, emphasizing the patient’s/client’s understanding and competency to make changes
11. Leave the door open, follow up, and let the patient/client decide if the session has been sufficiently helpful or if another session (or more) is needed.

Another list, offered more recently by Rycroft and Young (also see https://events.bouverie.org.au/sst) at The Bouverie Centre in Melbourne, contains a distillation of 10 SST core elements:

1. Negotiating a client/patient-led outcome
2. Establishing the client’s/patient’s (or clients’/patients’) priorities
3. Finding a focus and talking about the most important things (“cutting to the chase”)
4. Checking in with the client(s)/patient(s) at regular intervals
5. Interrupting respectfully when necessary (to help clients/patients get what they want)
6. Making time your friend—collaborating with the client/patient to optimize use of the time available, however short or long
7. Preparing to end well (reaching closure even if no solution or resolution)
8. Sharing your thoughts openly with clients/patients

9. Leaving the door open (an “open door” policy)
10. Listening to client/patient voices (following up, seeking feedback, and utilizing it).

SOME EXAMPLES OF SST

SST thinking can be flexibly applied in a wide variety of situations. Here are some typical clinical scenarios drawn from published reports. Each is offered as a prototype with many possible variations and applications—readers are encouraged to extrapolate and modify to suit their situations.

SST With an Individual

Mary was a 30-year-old woman who contacted the clinic because she wanted help learning to cope better with her highly intrusive biologic father (who had been divorced from her Mum since Mary was a child). She was offered a choice of either a by-appointment session or the option of walking in. She chose to make an appointment so that she could schedule childcare for her 4-year-old.

When she met with the therapist and they began to talk, he remarked: “Sometimes we may be able to get everything dealt with today in one visit, figuring out your next steps. Are you interested in that?” to which the client responded: “Yeah, that’s what I need.”

After briefly reviewing the history of her father’s misdeeds, she clarified as her session goal her wanting to handle him better in the present when he attempted to intrude on her and her husband and child. In the course of the session, she recognized how other family members had successfully dealt with the father, as well as lingering feelings of responsibility and guilt that impeded her. She then was guided, in session, to role-play, being assertive, saying “No,” and practicing ways to set what she felt would be healthier boundaries. Although quite anxious when she had first called, by the end of the one session, she felt ready with skills and motivation and did not feel she needed to make another appointment.

Patient: It makes me feel better already …. As long as I see the light.

Therapist: You can lighten the load you’ve got by taking control and not letting him run your life anymore. So you’re not trapped.

Patient: Exactly.

Therapist: Good.

Patient: Be in control and not get trapped.

Therapist: All right. Let’s speak in a month or so … If something comes up before and it is a question that can wait, wait. But if it’s something that you have to know or have to ask, do not hesitate to call me.

Patient: Like if he comes up with a new trick.

Therapist: You’ve got tricks, too. You can fight back.

Patient: Yeah. Now that I know how!

Therapist: Yes! *[they shake hands]*

Comment: On follow-up, she reported that she was doing well and that no further appointments were needed. She expressed appreciation and agreed to return on an as-needed basis.

Couple/Family Conflict

Georgie and Anne were a couple who sought an appointment to discuss their concerns regarding their 14-year-old son's angry outburst, during which he had threatened his younger brother with a knife. The parents wondered whether they had responded appropriately and/or whether their son needed psychiatric evaluation. After completing an intake questionnaire and having a brief telephone contact, they were then seen in session with the son. Careful attention was paid to connecting with all the participants, complimenting Georgie and Anne for their concern, and recognizing the son's strengths and not just his temper. To keep things focused on the search for solutions (Context of Competence), at one point, the therapist remarked: "Given that we have about 20 minutes left … I'm wondering what hasn't been said—or asked—at this point?" The son had talked things out with his brother. He acknowledged his role in the altercation, was remorseful and pledged to do better. Both Georgie and Anne felt reassured and validated, both in their concern and also in their parenting.

Comment: Further discussion revealed no indication for additional psychiatric evaluation or treatment. They agreed to a phone follow-up in 2 weeks and to recontact the therapist sooner if needed. At follow-up, the family reported that at home, they had talked more about the incident and the session. They were doing well, did not feel a need for another session, and expressed appreciation for the open-door invitation to return if/when appropriate.

Serial (or Sequential) Single Sessions

OAAT doesn't necessarily mean only one time. Many patients may come for a single session and later decide to return for either another single session or for a longer course of therapy—sometimes for more work on the original presenting issue, sometimes to discuss other issues.

One family with a disabled parent with acquired brain injury (ABI) was seen for a total of 32 SSTs! After each session, when they were asked if they wanted to make another appointment, they responded: “Not now. This is what we need for now. We’ll call when we’re ready for another meeting.” These repeated single sessions came to be called “coming in for tune-ups. Although OAAT, the family was seen, with varying combinations of attendees, over a 5-year period with an average of 6.5 times per year. Each session addressed different problems as determined by individual and family needs.

Comment: *Serial* (or *sequential*) SST is sometimes called *episodic*, *intermittent*, *repeat*, *booster*, *targeted,* or *catalytic*. SST is often useful when working with patients and families coping with new or ongoing medical-psychiatric problems (ranging from infants to adolescents and adults to seniors, including those with Alzheimer disease and other illnesses involving cognitive decline).

In the Course of Ongoing Medication Management

Mrs. Jones had been doing very well on an antidepressant for several years. One day she called Dr. Jenkins:

Patient: Hi, Doctor. I've been having a lot of anxiety about some problems with my job, and I'd like an appointment to see you to talk about what's been going on.

Doctor: Has your depression gotten worse or been interfering?

Patient: No, not really. This is more about problems with my boss at work.

Doctor: Would you like me to refer you to someone who specializes in working with occupational problems? I would be glad to see you, but as you know, I primarily do biologic psychiatry. I have a couple of colleagues, one a psychologist and the other a licensed clinical social worker, who are very good.

Patient: Thank you, doctor, but I think I'd really rather come to see you. I already know you and trust you, and you know me. I think I may only need one visit. And I'll be glad to pay for a full appointment, not just a brief medication check.

Comment: While Dr. Jenkins' main professional focus was on the use of medications to treat mood disorders, she also appreciated the opportunity to use a range of her biopsychosocial therapeutic skills. They met, and by focusing on the patient's specific session goals and the skills she could utilize to achieve those goals, Mrs. Jones got the psychotherapeutic support and direction she had sought; and she continued with her medication regimen.

Single-Session Family Consultation (SFCC)

As described by O'Hanlon and Rottem, SSFC is a model of brief family engagement and inclusion in the care of an individual client, typically involving between one and three sessions and then as needed. It includes a meeting organized by a practitioner with their client (a service user) and their family to collaboratively respond to family-identified needs and to clarify how the family will be involved in the individual's ongoing care.

Noel was a 57-year-old man who had been on disability leave from work after a recent acute episode of bipolar affective disorder in which he had become disinhibited and aggressive toward his work colleagues. He was now on medication and restabilized but having difficulty bringing himself to return to work. He felt it would be helpful for his wife and 21-year-old live-in daughter to understand more about what he had been through and how he was feeling. With the patient's consent, the mental health practitioner contacted the family, and a meeting was arranged.

Comment: An SSFC has three parts: (1) *Convening*, which refers to the key tasks and activities associated with the therapist getting the patient's consent and the therapist then setting up the session with the family; (2) *Conducting*, which refers to the actual session with the patient and family. This will typically involve welcoming, scoping the issues, finding a focus for the session, working on prioritized issues, summarizing progress, and agreeing on the next steps; and (3) *Following-up*, typically a phone call (to the patient and possibly involved family members) to check on the status of the problem, follow up on any actions or referrals decided at the patient and family meeting, and to decide upon next steps (which might include the patient's ongoing therapy, and possibly another single family session or referral to another service).

Clinical Supervision

Single-session thinking can also be used to structure each supervision meeting to have a clear beginning, middle, and end with specific goals for the particular session. Useful single-session supervision questions might include the following:

- If this could be as useful as possible _____?
- Are we talking about what we need to be focusing on, or are we getting off track?
- Of everything you have tried, what has gone well, and what has not gone so well?
- What do you know about yourself/your client(s)/patient(s) that tells me that you are ready to try something different?

- Before I tell you what I'm thinking, is there anything further I need to know, do you think, to be helpful to you?
- How far have we come in meeting what you had hoped for?
- Any last questions/concerns?
- What are you going to be taking away?

There is also sometimes a parallel process between therapy and supervision, in which single-session/OAAT phase-specific (beginning, middle, ending of a therapy session) aspects of the patient–therapist relationship get repeated in the therapist–supervisor relationship.

Many Psychiatric Medication Consults Involve One Session

A planning session with colleagues informed by SST/OAAT principles also can be helpful before consulting with a patient.

As Rosenberg and McDaniel have written:

> *Single session interventions are common in collaborative, biopsychosocial health care. Certain clinical scenarios are best suited for these kinds of interventions. These scenarios include interdisciplinary meetings in the hospital as well as in outpatient clinics, around critical moments in clinical care, and as consultations with existing care relationships or teams.*

Comment: Rosenberg and MçDaniel describe different types of planned single-session medical family therapy meetings, including hospital-based family meetings; outpatient family meetings around a new diagnosis, treatment adherence, or bad news; consultation by behavioral health or biomedical provider; and team-based preclinic "huddles."

THE ROLE OF ASSESSMENT AND RISK MANAGEMENT IN SST

As Yalom (2002, p. 4, emphasis in original) has noted, "Though diagnosis is unquestionably critical in treatment considerations for many severe conditions with a biological substrate […] diagnosis is often counterproductive in the everyday psychotherapy of less severely impaired patients."

—Irvin Yalom, The Gift of Therapy

SST involves seeking a pragmatic and parsimonious solution for the problem presented by the patient. As Talmon has described, this way of thinking, in which the search for psychohealth replaces the search for psychopathology, represents an alternative to the traditional model in psychiatry and psychotherapy. Focusing on *what works* is usually better in therapy as in life. Psychiatrist Anne Lutz and many others have written about the profound shift and "new direction" entailed in interviewing to identify strengths and resources rather than focusing on symptoms and weaknesses.

Bobele and Slive observe the frequent concern clinicians express about possibly "missing something" in a single session, including potential risk factors. They note that the need for possible further risk assessment can be identified with one item in a client's intake materials, such as "Do you now, or have you had, any fears that you (or your child, or anyone with you) is at risk of harm to themselves, others, or pets"? Along related lines, Paul Denborough, a psychiatrist in Melbourne, Australia, familiar with SST, remarked, "It's actually the perfect model to deal with risk—Single Session—because you see people quickly. You get the whole social network and family around them, and you try and fix the problem as soon as possible. You couldn't get a better way to manage risk, in my opinion."

Whatever questions are asked, it is good to remember that the word *diagnosis* comes from Greek and Latin words (*via gnosis),* meaning *the way of knowledge*. We never know everything; information should be sought that will point the way forward.

SST: TOWARD THE FUTURE

It is hard to make predictions because there are so many possible influencing factors, but these seem likely:

1. More SSTs, both by appointment and by the open-access option of "no appointment needed" walk-ins. This will call for more attention to implementation and administrative support.
2. More use of the internet—"Counselors are available now waiting for your call."
3. More clinics (including public and private healthcare organizations, university counseling centers, employee assistance programs, training institutes, and freestanding services) offering SST around the globe
4. More publications and more training.
5. More research—which groups and problems respond best to which approaches, and what actually happens—what are the "active ingredients" that make SST effective? In his Foreword to Talmon's

Single Session Therapy, psychiatrist Jerome Frank wrote about how people making changes in one session challenges a lot of the assumptions that therapists have had: the beliefs that you have to gradually form an alliance, that you have to gradually uncover the underlying schemas or neuroses, that you then have to gradually work your way through or there will be too much resistance. When things happen relatively quickly, it's an interesting question for researchers, what really happened, because it can't be a gradual working-through process; something shifted more quickly. SST, because it is circumscribed and concise, should be an excellent research ground.

6. More emphasis on patient/client strengths, resources, and competencies.
7. More attention to cultural nuances. Appreciating, respecting, and sometimes clinically utilizing the different strengths of different cultures, valuing the various belief systems people are coming from and how they go about it. It is important to work with, and especially to not work against patients' cultural orientations.

SST is responsive to patients' natural help-seeking behavior. SST was really "invented" by patients—even Freud's patients Katharina and Mahler knew in advance that there would be only the one session. Many hundreds of thousands of patients since have chosen to be seen for one session, even when more visits were available, and have found one session beneficial and sufficient. These sessions were approached and concluded as single sessions complete unto themselves (a "one-and-done," or a "one-off," as our Brit and Aussie colleagues might say)—with the possibility of return as needed. Clinicians and researchers are endeavoring to define and refine what facilitates those successful outcomes.

Patients, clinicians, managers, policymakers, and funders are all interested in the benefits of single-session thinking and practice. Offering the option of SST—by appointment and/or by walk-in—allows patients and therapists to efficiently "capture the moment" and "strike while the iron is hot." SST is responsive to the natural help-seeking behavior patterns of patients, has strong empirical support for increasing patient satisfaction, and can help conserve resources for those patients who want and need additional services. Increasing accessibility and affordability is also a social justice issue, and recognizing the potential of SST is especially important nowadays as so many more people would benefit from mental health services if they were affordable and did not carry the often daunting stigma of seemingly endless dependency. Resources are limited, and SST has a valuable ecologic function: it preserves time and money. More is not better; *better* is better—and *better* can sometimes be achieved here and now, even in one visit.

CODA

SST is a treatment that is approached one session at a time. Each meeting is conceived as an event, potentially complete until the next round. It is not intended as a replacement for other therapy modalities. It is not an abbreviated form of conventional therapy, nor even a substitute for other "brief" or "time-sensitive" therapies. SST should not be "forced" and is most appropriate for those patients who desire and choose to attend one session, with the option of more as needed. Single-session thinking can also infuse other nonclinical situations. SST is a common, effective, flexible, increasingly recognized, and important component in the landscape of mental health service delivery.

ACKNOWLEDGMENTS

Thanks to Moshe Talmon and Robert Rosenbaum, my original SST collaborators; and to Jeff Young and Pam Rycroft. Additional thanks to the APF Cummings Psyche Prize for support of my research and writing.

Further Readings

Bloom BL. Bloom's focused single-session therapy. In: Bloom BL, ed. *Planned Short-Term Psychotherapy: A Clinical Handbook*. 2nd ed. Allyn & Bacon; 1992:97–121. An earlier version, entitled "Focused single-session therapy: Initial development and evaluation, appeared in Budman SH, ed. *Forms of Brief Therapy*. Guilford Press; 1981:167–216.

Bobele M, Slive A. An open invitation to walk-in therapy: opening access to mental health care. In: Hoyt MF, Young J, Rycroft P, eds. *Single Session Thinking and Practie in Global, Cultural, and Familial Contexts: Expanding Applications*. Routledge; 2021:54–65.

Bobele M, Slive A. One session at a time: when you have a whole hour. In: Hoyt MF, Talmon M, eds. *Capturing the Moment: Single Session Therapy and Walk-in Services*. Crown House Publishing; 2014:95–120.

Cannistrà F, Piccirilli F, eds. *Terapia a seduta singola: principi e pratiche*. (In Italian). Giunti; 2018. (Published in English as *Single Session Therapy: Principles and*

Practices. Firenze: Giunti; 2021.)

Cornish P. *Stepped Care 2.0: A Paradigm Shift in Mental Health*. Springer Nature; 2020.

DeMelo J. Bull's eye! One-and-done sessions give new meaning to the phrase targeted therapy. *O: The Oprah Magazine*. 2018;19:63–64.

Dryden W. *Single-Session Therapy: 100 Key Points and Techniques*. Routledge; 2019.

Dryden W. *Single Session Therapy and its Future: What SST Leaders Think*. Routledge; 2021.

Frank JD. Foreword. In Talmon M, *Single Session Therapy*. Jossey-Bass; 1990:xi-xiii.

Hoyt MF. *Brief Therapy and Beyond: Stories, Language, Love, Hope, and Time*. Routledge; 2017.

Hoyt MF. *Brief Therapy and Managed Care*. Jossey-Bass; 1995.

Hoyt MF. Psychology and my gallbladder: an insider's account of a single-session therapy. In: Hoyt MF, Talmon M, eds. *Capturing the Moment: Single-Session Therapy and Walk-In Services*. Crown House Publishing; 2014:53–72.

Hoyt MF. *Some Stories are Better Than Others: Doing What Works in Brief Therapy and Managed Care*. Brunner/Mazel; 2000.

Hoyt MF, Bobele M, Slive A, et al., eds. *Single Session Therapy by Walk-in or Appointment: Administrative, Clinical, and Supervisory Aspects of One-at-a-Time Services*. Routledge; 2018.

Hoyt MF, Rosenbaum R, Talmon M. Planned single-session psychotherapy. In: Budman SH, Hoyt MF, Friedman S, eds. *The First Session in Brief Therapy*. Guilford Press; 1992:59–86.

Hoyt MF, Talmon M, eds. *Capturing the Moment: Single Session Therapy and Walk-in Services*. Crown House Publishing; 2014.

Hoyt MF, Talmon M. Single-session therapy in action: a case example. In: Talmon M, ed. *Single Session Therapy*. Jossey-Bass; 1990:78–96.

Hoyt MF, Young J, Rycroft P, eds. *Single Session Thinking and Practice in Global, Cultural and Familial Contexts: Expanding Applications*. Routledge; 2021.

Hymmen P, Stalker CA, Cait C. The case for single-session therapy: does the empirical evidence support the increased prevalence of this service delivery model and walk-in services? *J Ment Health*. 2013;22(1):60–71.

Lutz AB. *Learning Solution-Focused Therapy: An Illustrated Guide*. American Psychiatric Press; 2014.

McElheran N. The story of the Eastside Family Centre: 30 years of walk-in single session therapy. In: Hoyt MF, Young J, Rycroft P, eds. *Single Session Thinking and Practice in Global, Cultural and Familial Contexts: Expanding Applications*. Routledge; 2021:125–139.

O'Hanlon B, Rottem N. Single session family consultation (SSFC). In: Hoyt MF, Young J, Rycroft P, eds. *Single Session Thinking and Practice in Global, Cultural, and Familial Contexts: Expanding Applications*. Routledge; 2021:66–76.

Rosenbaum R, Hoyt MF, Talmon M. The challenge of single-session therapies: creating pivotal moments. In: Wells RA, Giannetti VJ, eds. *Handbook of the Brief Psychotherapies*. Plenum Press; 1990:165–189. Reprinted in: Hoyt MF, ed. *Brief Therapy and Managed Care*. Jossey-Bass; 1995:105–140.

Rosenberg T, McDaniel SH. Single session medical family therapy and the patient-centered medical home. In: Hoyt MF, Talmon M, eds. *Capturing the Moment: Single*

Session Therapy and Walk-in Services. Crown House Publishing; 2014:349–362.

Rycroft P. Capturing the moment in supervision. In: Hoyt MF, Bobele M, Slive A, et al., eds. *Single-Session Therapy by Walk-in or Appointment: Administrative, Clinical, and Supervisory Aspects of One-at-a-Time Services*. Routledge; 2018:347–365.

Rycroft P, Young J. SST in Australia: learning from teaching. In: Hoyt MF, Talmon M, eds. *Capturing the Moment: Single Session Therapy and Walk-in Services*. Crown House Publishing; 2014:141–156.

Rycroft P, Young J. Translating single session thinking into practice. In: Hoyt MF, Young J, Rycroft P, eds. *Single Session Thinking and Practice in Global, Cultural, and Family Contexts: Expanding Applications*. Routledge; 2021:42–53.

Slive A, Bobele M, eds. *When One Hour is All You Have: Effective Therapy for Walk-in Clients*. AZ: Zeig, Tucker, & Theisen; 2011.

Slive A, MacLaurin B, Oakander M, Amundson J. Walk-in single-sessions: a new paradigm in clinical service delivery. *J Syst Ther*. 1995;14:3–11.

Story K. “Coming in for tune-ups”: a family’s experience of episodic long-term single-session therapy at The Bouverie Centre, Melbourne. In: Hoyt MF, Bobele M., Slive A, et al., eds. *Single-Session Therapy by Walk-in or Appointment: Administrative, Clinical and Supervisory Aspects of One-at-a-Time Services*. Routledge; 2018:202–220.

Talmon M. *Single-Session Therapy: Maximizing the Effect of The First (and Often Only) Therapeutic Encounter*. Jossey-Bass; 1990.

Talmon M. *Single Session Solutions: A Guide To Practical, Effective, And Affordable Therapy*. Addison-Wesley; 1993.

Talmon M. Foreword. In: Hoyt MF, Young J, Rycroft P, eds. *Single Session Thinking and Practice in Global, Cultural, and Familial Contexts: Expanding Applications*. Routledge; 2021:xxvii–xxviii.

Yalom ID. *The Gift of Therapy: An Open Letter to a New Generation of Therapists and Their Patients*. HarperCollins; 2002.

Young J. Single-session therapy: the misunderstood gift that keeps on giving. In: Hoyt MF, Bobele M, Slive A, Young J, Talmon M, eds. *Single-Session Therapy By Walk-In Or Appointment: Administrative, Clinical, And Supervisory Aspects of One-At-A-Time Services*. Routledge; 2018:40–58.

Young K, Jebreen J. Recognizing single-session therapy as psychotherapy. *J Syst Ther*. 2019;38(4):31–44.

32.14 Mentalization-Based Treatment

ANTHONY W. BATEMAN, M.A., F.R.C.PSYCH., AND PETER FONAGY, O.B.E., F.MED.SCI., F.B.A., F.AC.S.S.

DEFINITION

Mentalization-based treatment (MBT) is an evidence-based treatment for borderline personality disorder (BPD). Randomized controlled trials (RCTs) have demonstrated their effectiveness in reducing the core symptoms of BPD and it is adapted for the treatment of other conditions, including antisocial personality disorder (ASPD) and avoidant personality disorder, substance abuse, depression, and eating disorders, and complex post-traumatic stress disorder.

This chapter will outline the theoretical basis of MBT and the core treatment model in relation to BPD.

What Is Mentalizing?

Mentalizing is the social cognitive ability to understand actions of other people and oneself in terms of mental states, including thoughts, feelings, wishes, and desires. It is a very human capability that underpins everyday interactions. In nontechnical language, it is attentiveness to thinking and feeling in oneself and others. It is beyond question that mental states influence behavior. Beliefs, wishes, feelings, and thoughts, whether within or outside people's awareness, always influence what people do. Mentalizing involves a whole spectrum of capacities: critically, this includes the ability to experience one's *own* behavior as coherently organized by mental states and to differentiate oneself psychologically from others. These capacities are reduced in individuals with a personality disorder, who lose cognitive and emotional coherence, particularly at moments of interpersonal (relational) stress, which challenge one's mentalizing capacities. In the authors' view, many symptoms characteristic of BPD emerge in association with a distortion or reduction in mentalizing.

The general reduction in mentalizing capacities leads to fear of social isolation and experience of loneliness; the loss of integrated and balanced mentalizing results in a reduction in joint intentionality and a sense of togetherness, which leaves people with BPD feeling isolated from others and unable to engage in and feel part of everyday social interaction. The key to constructive prosocial interaction is the capacity to coordinate perspectives between oneself and others and thus appreciate the distinction between the subjective (one's own view) and the objective (actual physical reality "out there"), and then intertwine these with knowledge of another individual’s mental state. Joining of minds—the so-called “we-mode”—is the crucial ingredient of social mentalizing and friendship making. The route to this, often obstructed for people with BPD, is the experience of joint attention, which defines a commonly regarded object while also acknowledging that different individuals may have different perspectives on it.

Based on this understanding of BPD, MBT is a psychotherapy that focuses specifically on the mentalizing vulnerabilities of the patient in the context of an understanding of attachment process, which is the developmental context in which mentalizing is originally acquired.

HISTORY

The word “mentalizing” has been in existence for two centuries and in the *Oxford English Dictionary* for the past century. French psychoanalysts introduced the concept into the professional psychotherapy literature in the second half of the 20th century. Mentalizing came into the English professional literature on the cusp of the final decade. It was conceived initially as a deficit in autism and as a transiently impaired process associated with profound insecurity in attachment relationships in the developmental psychopathology of BPD. This proposal was the origin of the development of MBT.

The MBT approach is rooted in attachment theory and psychoanalytic ideas, but in the interest of parsimony, it sheds many of the core assumptions of both theoretical approaches while negating neither body of work. Its historical origin is born from the authors’ wish to extend what they saw as the clear benefits of using a psychotherapeutic approach to professionals (nurses, rehabilitation or addiction counselors, activity therapists) who normally did not have opportunities to undertake extensive training in particular psychotherapeutic modalities. The authors wished to extract core elements of a therapeutic approach that could be linked to experimentally observed distortions of mental function in personality disorders.

THEORETICAL ISSUES

The Role of Mentalizing in Therapy

MBT is based on the assumption that failures of mentalizing, while common to everyone, become a dominant feature in individuals with personality disorders, leading to both serious interpersonal problems and profound psychological distress. A number of therapies that have been shown to be successful in addressing the difficulties experienced by individuals with BPD appear to strengthen the patient's capacity to mentalize; for example, transference-focused psychotherapy, conversational therapy, cognitive analytic therapy, and certain aspects of dialectical behavior therapy (DBT). MBT is unique in attempting to understand the problematic and distressing aspects of severe personality disorder in terms of a failure of mentalizing (see later) and to focus a structured therapeutic approach on addressing problems of mentalizing as they occur in a therapy session. The overarching principle of MBT is to enhance mentalizing in the context of the therapeutic relationship (in both individual and group sessions) by systematically addressing instances of nonmentalizing and using these moments of discourse as opportunities to work with the patient to achieve a fuller psychological understanding of behavior.

The significance of this process rests on the fact that mentalizing is the fuel that drives learning from life experiences in the social environment. It allows the individual to be discerning and to know from whom they can learn and who they can trust in the world. Epistemic trust encourages social learning by enabling individuals to be open to benefit from their social environment. This mechanism for opening the channel to social learning exists because it cannot be left open by default: it is adaptive for humans to adopt a position of epistemic vigilance toward another person unless they are reassured that that person's intentions in communicating are not malign. The disruption of epistemic trust—or the emergence of outright epistemic mistrust as a result of environmental adversity, genetic propensity, or an interaction between the two—can lead to a fundamental breakdown in the capacity for meaningful social communications that are necessary for change to occur in how others or the self are seen.

Learning from others and calibrating oneself in a trusting relationship improves emotional regulation, reduces impulsivity, and facilitates social function. These are the problematic areas that characterize individuals with entrenched and severe mental health difficulties—particularly those with BPD. MBT trains the individual in mentalizing in the hope of generalizing their abilities from the therapeutic situation to the wider social context. In doing so, MBT can reduce mistrust, isolation, and rigidity and enhance the capacity for trust and, with it, the capacity to benefit from social experience.

The Multidimensional Nature of Mentalizing

Mentalizing is not an entirely stable, consistent, or unidimensional capacity. Neuroscience has identified four distinct components to mentalizing, which the authors have organized into dimensions that they see as helpful for therapists to identify in the clinical practice of MBT:

1. Automatic versus controlled mentalizing
2. Mentalizing the self versus others
3. Mentalizing with regard to internal versus external features
4. Cognitive versus affective mentalizing.

To mentalize effectively requires the individual not only to be able to maintain a balance across these dimensions of social cognition but also to apply them appropriately according to context. Consistent favoring of one or other side (or *pole*) of these dimensions leads to distorted understanding of mental states, associated with profound social and emotional difficulties.

In an adult with personality disorder, consistent distortions of social cognition consequent on imbalanced mentalizing on at least one of these four dimensions would be evident. Commonly, one or more of the dimensions underperforms at one end, and consequently, the opposite pole comes to dominate social cognition. For example, excessively emotional thinking emerges in the absence of cognitive mentalizing, or the influence of others dominates if subjective experience of self-states is reduced. From this perspective, different types of psychopathology can be distinguished on the basis of different combinations of impairments along the four dimensions. In other words, personality disorders (and, to some extent, other psychiatric disorders) can be understood according to different characteristic *mentalizing profiles.*

Automatic Versus Controlled Mentalizing. The most fundamental dimension of mentalizing is the spectrum between *automatic* (or *implicit*) and *controlled* (or *explicit*) mentalizing.

Controlled mentalizing reflects a serial and relatively slow process, which is typically verbal and demands reflection, attention, awareness, intention, and effort. For example, a person might misunderstand someone, so they stop them and ask them to explain more clearly what is underlying their statements, or they focus their own mind to work out what their opinion may be. People tell stories to others about their mental states; they think back into their past and report how they felt long ago; they enjoy autobiographical coherence in their personal lives and tell others about it.

The opposite pole of this dimension, *automatic mentalizing,* involves much faster processing, tends to be reflexive, and requires little or no attention, awareness, intention, or effort. In day-to-day life and ordinary social interaction, most mentalizing tends to be automatic because most straightforward exchanges do not require more attention. People in conversation naturally take turns, adapt their tone and posture to others' emotional states, and reflexively take into account their knowledge.

Self Versus Others. This mentalizing dimension involves the capacity to become aware of one's own state—the *self* (including one's own physical experiences)—or the state of *others.* The two are closely connected, and an imbalance signals vulnerability in the understanding of both others and/or the self. Individuals with mentalizing difficulties are likely to preferentially focus on one end of the spectrum, although they may be impaired at both. Awareness of the mental states of others is in part mediated by neural structures that organize one's own actions (the *mirror neuron system*), moderated by explicit reflective processes that reinforce the self–other distinction. If explicit mentalizing is weakened, the influence of the current mental state of the other will increase.

Internal Versus External Mentalizing. Mentalizing can involve making inferences on the basis of the *external* indicators of a person's mental states (e.g., facial expressions, tone of voice, body posture) or figuring out someone's *internal* experience from what one knows about them and the situation they are in. This dimension does not simply refer to a process of focusing on the externally visible manifestations versus the internal mental state of others. It also applies to the self—it includes thinking about oneself and one's own mind state versus considering one's current (external) situation and physical (interoceptive) state. Someone who has poor access to and great uncertainty about their subjective experience, for example, as is often seen in individuals with BPD, may come to a conclusion about what they are feeling from the reactions of others as well as from observing their own behavior: for example, their legs are restless. Therefore they must be feeling anxious; they took an overdose, so they must have felt suicidal.

Cognitive Versus Affective Mentalizing. *Cognitive* mentalizing involves the ability to name, recognize, and reason about mental states (in both oneself and others), whereas *affective* mentalizing involves the ability to experience and understand associated *feelings* (again, in both oneself and others). Both are required for any genuine experience of empathy or true sense of self-coherence. Some individuals give undue weight to either cognitive or affective mentalizing. People with obsessional characteristics may be masters at explicating in detail the internal states of themselves and others, but this may be devoid of emotional content and meaning. Conversely, people with BPD are flooded with emotion and so prone to automatic process, reactivity, emotional contagion, and poor self–other differentiation. Intense emotion can disrupt the process of cognitive appraisal that normally helps to regulate it.

Dimensional Mentalizing Profile and BPD

The clinician and patient work on a personalized mentalizing profile in the assessment phase of MBT treatment, identifying which dimensions the patient tends to over- or underuse when under stress, and which modes are commonly prominent. This profile forms part of the shared formulation. Particularly when arousal increases, as is typical in the context of intense attachment relationships, individuals with BPD easily find themselves switching to automatic mentalizing. Stress and arousal, especially in an attachment context, bring automatic mentalizing to the fore and disengage the neural systems that are associated with controlled mentalizing. Under these conditions, interactions become nonquestioning precisely when they need to be more controlled and contextualized. Thinking becomes impulsive: the individual makes quick assumptions about others’ thoughts and feelings, which are not reflected upon or tested. Logic is intuitive, unreasoned, and nonverbal; it is marked by an unwarranted certainty, which betrays its unreflective origin. As a consequence, patients show severe impairments in interpersonal and intimate relationships; for example, they may be overly distrustful (paranoid) or, indeed, overly trustful (naive).

Patients with BPD may show excessive concern about their own internal state and those of others, that is, they *hypermentalize* in relation to the self and attribute motives to others without any basis in reality. At the same time, these views of the self-develop without reference to social reality, namely an awareness of how others perceive one. Failure to balance self-perception with sincere curiosity about how one is perceived by others can lead to exaggerations of the self-image, in both positive and negative directions. A balanced, adaptive form of self-mentalizing conditioned by the social context is lost.

Patients with BPD pay more attention to external indicators of mental states, and their initial ideas arising from automatic mentalizing are left unchecked by controlled/reflective mentalizing. For example, if the clinician frowns, perhaps pensively, the patient may interpret this as looking angry or disgusted with them; seeing the clinician look at the clock during a session can stimulate an internal state of overwhelming unease and an experience that the clinician wants to be rid of them when, in reality, the internal state of the clinician is concern about the time left to work on the issue being discussed. A focus on external features, in the absence of reflective mentalizing, makes an individual highly vulnerable in social contexts, as it generates interpersonal hypersensitivity.

The Reemergence of Nonmentalizing Modes and BPD

While the dimensions of mentalizing can reflect anomalies in terms of mechanisms, that is not what the clinician sees. What the patient and the mentalizing clinician experience is a product of a malfunctioning mentalizing system, driven by imbalances in the dimensions of mentalizing. The outcomes of these malfunctions can be grouped into three typical modes of subjectivity for the purpose of illuminating clinical experience. The modes are termed *psychic equivalence mode, teleologic mode,* and *pretend mode.* These modes are summarized in Tables 32.14–1 through 32.14–3, which also outline MBT interventions to address each nonmentalizing mode.

The nonmentalizing modes are important for the clinician and patient to recognize and understand, as they tend to emerge in the consulting room and reflect core aspects of the patient’s experience. It is important to address them because they cause considerable interpersonal difficulties and result in destructive behaviors.

Table 32.14–1.
Modes of Nonmentalizing: Psychic Equivalence

Clinical appearance	Certainty/suspension of doubt Absolute Reality is defined by self-experience Finality—“It just is” Internal is seen as equivalent to external
Clinician’s experience	Puzzled Wish to refute Statement appears logical but is obviously overgeneralized Not sure what to say Angry or fed up and hopeless
Intervention	Empathic validation with subjective experience Curious—“How did you reach that conclusion?” Presentation of clinician’s puzzlement (marked) The “Mentalizing Loop”: Linked topic diversion to trigger mentalizing, then revisit the area of psychic equivalence
Iatrogenic	Argue with patient Excessive focus on content Cognitive challenge

Table 32.14–2.
Modes of Nonmentalizing: Teleologic Mode

Clinical appearance	The expectation of things being “done” Outcomes in the physical world determine the understanding of the inner state—“I took an overdose; I must have been suicidal”; “You were late; you don’t want to see me” Motives of others are based on what actually happens Only actions can change mental process “What you do and not what you say”
Clinician’s experience	Uncertainty and anxiety Wish to do something—medication review, letter, telephone call, extend session
Intervention	Empathic validation of need Do (or do not do) according to the exploration of need Affect focus of dilemma of doing or not doing The “Mentalizing Loop”: Linked topic diversion to trigger mentalizing, then revisit teleologic understanding
Iatrogenic	Excessive “doing” “Prove” clinician cares in the belief it will induce positive change Elasticity (extending what clinician does, e.g., providing extra sessions, only to rebound with extra constraints) rather than flexibility

Psychic Equivalence Mode. In the *psychic equivalence mode*, thoughts and feelings become “too real” to a point where it is extremely difficult for the patient to entertain possible alternative perspectives. When mentalizing gives way to psychic equivalence, what is thought is experienced as being real and true, leading to what clinicians describe as “concreteness of thought” in their patients.

Patients with BPD who are in this mode describe an overriding sense of certainty about their beliefs, for example, “the therapist does not like me” or “I am a wicked person.” Such a state of mind can be extremely frightening, adding a powerful sense of drama and risk to life experiences. The sometimes extreme reactions of patients are justified by the seriousness and realness with which they can experience their own and others’ thoughts and feelings. The vividness and bizarreness of subjective experience can appear as quasi-psychotic symptoms and are also manifested in the physically compelling memories associated with trauma.

Table 32.14–3.
Modes of Nonmentalizing: Pretend Mode

Clinical appearance	Inconsequential talk/groundless inferences about mental states Lack of affect. Absence of pleasure Circularity without conclusion —“spinning in sand” (hypermentalizing) No change Dissociation—self-harm to avoid meaninglessness Body and mind decoupled
Clinician’s experience	Boredom Detachment The patient agrees with the clinician’s concepts and ideas Identification with clinician’s model Feels progress is made in therapy
Intervention	Probe extent of mode—is it embedded in all content? Explore immediate patient-therapist interaction Counterintuitive Challenge
Iatrogenic	Nonrecognition Joining in with acceptance as real Insight-oriented/skill acquisition intervention

More important to appreciate is that negative affect in this mode will be overwhelming and cause very deep distress to the sufferer. Psychic equivalence permits no alternative perspectives to be taken. The patient can feel locked into an extraordinarily profound sense of pain without the slightest possibility that matters could be any different from how they currently are. This state of hopelessness makes suicidality comprehensible.

Teleologic Mode. In the *teleologic mode,* states of mind are recognized and believed only if their outcomes are physically observable. Hence, the individual can recognize the existence and potential importance of states of mind, but this recognition is limited to very concrete situations. For example, affection is perceived to be true only if it is accompanied by physical contact such as a touch or caress. The teleologic mode shows itself in patients who are imbalanced toward the external pole of the internal–external mentalizing dimension—they are heavily biased toward understanding how people (and they themselves) behave and what their intentions may be in terms of what they physically do.

Impulsivity involves function in teleologic mode and a heavy emphasis on the automatic pole of the automatic–controlled dimension. There is insufficient reflection concerning the impact of one's actions on others, or on oneself. In teleologic mode, the individual cannot accept anything other than a physical action as a true expression of the other person's intentions. Friends of a patient might constantly assure the patient of their love and support, and yet none of that feels real: it does not address the "hole" that the patient falls into at certain, especially lonely, times when they feel terrible emptiness. Feeling that interpersonal affection can only be real if it is accompanied by physical behavior explains some risky sexual behavior, but also the need to create physical distractions that help with the feeling that all verbal expressions of interpersonal affection are without real meaning (see case example, later).

The teleologic mode makes the need for action and generation of "real" change overwhelming. The so-called "manipulativeness" of patients with personality disorder is little more than the experience of a pressing need to feel genuine reactions from others in terms of actions rather than words.

Pretend Mode. In the *pretend mode,* thoughts and feelings become severed from reality. Taken to an extreme, this may lead to feelings of derealization and dissociation. Patients in pretend mode can discuss experiences without contextualizing these in any kind of physical or material reality as if they were creating a pretend world. The patient may *hypermentalize* or *pseudomentalize,* a state in which they may say much about states of mind but with little true meaning or connection to reality.

Attempting psychotherapy with patients who are in this mode can lead to lengthy but inconsequential discussions of internal experience that have no link to genuine experience and will achieve no change. The limited capacity to experience a sense of an internal world gives rise to a deep and extremely distressing sense of emptiness. Dramatic action—for example, sometimes violent action against the self—may be experienced as the best way of addressing such feelings.

In summary, imbalances within the dimensions of mentalizing predictably generate the nonmentalizing modes. Psychic equivalence is inevitable if emotion (affect) dominates cognition. Teleologic mode follows from an exclusive focus on external features to the neglect of the internal. Pretend mode thinking and hypermentalizing are unavoidable if reflective, explicit, controlled mentalizing is not well established.

Attachment

It is a central tenet of the mentalization-based approach that a sense of self and the capacity to mentalize both develop in the context of attachment relationships. The child observes, mirrors, and then internalizes their attachment figures' ability to represent and reflect mental states. So the reflections need to be *contingent*—that is, related to the child's internal experience—accurate, and *marked*—that is, indicating, for example, using "motherese" (a special tone of voice), that what is being expressed is a representation of the mind of the child and not that of the caregiver. Hence, the self and others—and the capacity to reflect on the self and others—are inevitably closely intertwined.

Disorders that are characterized by severe impairments in feelings of self-identity, a central tenet of the pathology of BPD—are also characterized by severe deficits in the ability to reflect about others' mental states. However, this should not be taken to mean that an individual whose capacity to mentalize themselves is impaired will always show similar impairments in their ability to mentalize others. For example, individuals with ASPD can often be surprisingly skilled in "reading the mind" of others only then to use this skill for exploitation but typically lack any real understanding of their own inner world.

In BPD there is commonly a history of early (in particular emotional) neglect, and a disrupted early social environment in general, and this may contribute to undermining the ability of some individuals to develop full mentalizing capacities. Subsequent adversity or trauma may further disrupt mentalizing, in part as an adaptive maneuver on the part of the individual to limit exposure to a brutalizing psychosocial environment, and in part because the high level of arousal generated by attachment hyperactivation and disorganized attachment strategies serve to disrupt less well-practiced and less robustly established higher cognitive capacities. In addition, genetic influences may be expressed through the mediation of mentalizing.

In sum, the mentalizing model is not strictly an etiologic model, although it clearly prioritizes a social psychiatry perspective; it points to a final common developmental pathway that a range of biologic, family, and broader social contextual influences may take to generate the range of difficulties that are normally considered under the term “personality disorder.”

TECHNIQUES

MBT is a group and individual treatment focusing on three primary areas of personality function:

1. Mentalizing
2. Emotion processing and management
3. Attachment process

It is anticipated that, at times, the patient will experience strong affect while focusing on identified problems in treatment sessions. At these times their mentalizing will be limited or failing, and/or the patient’s understanding of the way mental states link to behavior will be inadequate.

The clinician addresses this through a structured process (the sessional intervention trajectory) of:

1. Empathy and validation of problem areas
2. Clarification, exploration of affective narratives, and, where necessary, challenge
3. Following a structured process to gently expand mentalizing when it is out of balance and encourage the patient to identify the mental states and relational processes previously outside their awareness.

The therapeutic process is primarily in the here and now of the session but increasingly, as the patient's mentalizing improves, it comes to concern core attachment relationships, including how these are activated with the clinician and key figures in the patient's life and how they influence mentalizing itself. Gradually, improvements in mentalizing serve to enable the patient to address their distorted representations of personal and social relationships.

Therapeutic Alliance and Engagement in the Model

The assessment and introductory process in MBT facilitates the alliance between patient and clinician (see Table 32.14–4) and introduces the patient to the treatment frame. An MBT-Introductory group of 10 to 12 sessions assists in the development of the formulation and facilitates the alliance. This psychoeducational intervention covers all areas of mentalizing, attachment processes, personality disorder, emotion management, and treatment itself. This preparatory work aims to ensure patients know what they are facing in trying to address their problems and are fully aware of the method and focus of treatment.

MBT is collaborative and democratic. Nothing can occur without joint discussion, taking into account the mental experiences and ideas of both patient and clinician. The process of mentalizing requires an authentic desire to understand the mental processes of oneself and others. This applies as much to the clinician as to the patient. So the MBT clinician focuses on the patient's mind and attempts to understand their experience. Similarly, the patient is asked to aim to do the same in relation to the clinician—for example, the patient's perspective "Why does my clinician want me to focus on this at the moment?" may be paired with the clinician's, "Why does my patient *not* want to focus on this at the moment?" The therapeutic process has to become a shared endeavor aimed at extending the influence of explicit, reflective, cognitive, internally focused mentalizing. Initial goals, on the road to improved mentalizing, are jointly developed and focused on. The goals cannot solely be those of the patient, although the patient's aims take priority unless they are antithetical to the whole process of treatment. The sharing of responsibility for the therapeutic process is at the core of the effectiveness of the treatment approach in the pursuit of improved mental state understanding.

Table 32.14–4.
Alliance Building in MBT

Identification of patient's mentalizing vulnerabilities presented to the patient in an understandable form
Identification of attachment strategies commonly activated in relationships
Formulation of problems—agreed between patient and clinician
Identification of patient's risk profile and crisis management strategies
Agreement on short-term and long-term goals

Assessment involves delineation of the patient's mentalizing vulnerabilities and mentalizing profile, identification of nonmentalizing cycles (see case example, later), and a shared formulation, which includes specific detail of attachment patterns and areas of vulnerability to emotional dysregulation. This has to be understood by the patient and is for both patient and clinician.

The formulation identifies common relational fears, for example, abandonment, which stimulates the patient's attachment system and results in the use of maladaptive attachment strategies in interpersonal interactions. In brief, the pattern of the patient's relationships informs an understanding of the relationship in treatment, and the relationship in treatment is used to reappraise the relationships in life outside treatment.

Finally, it is important that the patient and clinician consider establishing a goal of improving social function. This will include work, social activity, voluntary work, education, and other constructive life-affirming activity.

Clinical Principles in MBT

Clinicians follow a number of principles when treating patients with MBT (see Table 32.14–5). MBT recommends an authentic "not-knowing" stance that forms the bedrock for exploration of the patient's perspective. The not-knowing stance refers to respecting the opacity of mental states. Minds can never be "known" and it is important that the clinician recognizes that mental processes generate experiences imbued with uncertainty. The clinician's task is to take an inquisitive stance—a wish to inquire and a willingness to be surprised by the patient's response—with the aim of facilitating the patient's increased awareness of their internal states through a social process.

This may be particularly important around the point at which experiences of ideas and feelings begin to collapse into nonmentalizing modes, leading to destructive behaviors or intolerable feeling states. It is not for the clinician to compensate for the patient’s mentalizing failure with their own high-level mentalizing, “explaining” to the patient what they may have experienced. Nonmentalizing in the patient cannot be met by mentalizing in the clinician; it can be met only by “switching on” mentalizing in the patient via the range of techniques that MBT uses in the clinical situation.

Primarily, the clinician is alert to nonmentalizing in terms of the different manifest modes of nonmentalizing, but also in terms of the indications that the patient’s functioning is fixed at one pole of any of the dimensions of mentalizing. In general, as described earlier, mentalizing is optimal when the dimensions—for example, emotion and cognition, or representation of self and other—are in balance and nonmentalizing modes are inactive. The key for the clinician is to be constantly aware of imbalance and lack of flexibility in terms of the dimensions and if any dimension is in danger of becoming fixed or rooted at one pole, intervention is necessary.

Table 32.14–5.
Clinical Principles in MBT

The clinician:

- Remains alert to imbalances of the mentalizing dimensions and emergence of nonmentalizing modes
- Monitors the patient’s arousal levels to maintain optimal mentalizing
- Seeks moments of mentalizing vulnerability related to events in the patient’s life or in the session itself
- Maintains their own mentalizing
- Avoids meeting nonmentalizing in the patient with high-level mentalizing
- Avoids joining with nonmentalizing
- Matches interventions to the patient’s mentalizing capacity
- Sensitively presents their own state of mind in relation to the patient’s state of mind to generate an understanding of the reciprocity of interaction

The clinician engages in “contrary moves” to activate the dormant pole of the dimension. For example, an exclusive focus on feelings and ideas in relation to the self should suggest to the clinician that exploring the mental states of the other is called for and they direct the session accordingly. Similarly, dominance of emotional mental experience without calibration from cognition suggests the need to bring cognitions into the foreground through judicious rational inquiry. Fluidity and adaptability of the dimensions to context and content are hallmarks of good mentalizing.

Second, the clinician monitors arousal levels carefully, ensuring that anxiety is neither too low nor too high, as both interfere with mentalizing. Similarly, if attachment feelings (e.g., in relation to the therapist) become too strong, a shift to less charged relationships may help restore mentalizing. Attachment stress triggers mentalizing vulnerabilities and may overwhelm the patient’s capacities so they panic just as they are learning to balance the two effectively as part of their focus on improving relationships. Once mentalizing is restored, the relationship can be revisited for further learning.

Third, the focus of a session is maintained through the clinician always seeking moments of mentalizing vulnerability either in relation to events in the patient’s life or in the session itself. Mentalizing failure is best indicated by the clinician’s experience of being challenged to understand the patient and consequently feeling a degree of confusion about how to respond. At these times, rewinding to moments when shared understanding characterized the discourse is the best solution.

Fourth, the clinician makes sure that their own mentalizing is maintained. It is not possible to deliver effective treatment if the clinician’s mentalizing is compromised. So the MBT clinician always monitors their own capacities and may even have to explicitly own this experience, for example, by saying that their mind has become muddled and they cannot think. This type of self-disclosure of the mind state of the clinician should not be confused with sharing personal information. Sharing the effect that a patient’s actions and state of mind is having on the clinician acts in the service of asking the patient to consider another mind as well as their own and has the implicit aim of enhancing mentalizing.

Sessions are focused. They do not consist of free-associative dialogue that seeks to illuminate unconscious process, as may be the case in psychoanalytic therapy. The target area in MBT is working memory or preconsciously held experience. It is expected that a focus for a given session will have been achieved after 10 to 15 minutes of the session, and this focus will then become the pivotal point around which the clinician and patient orient themselves, returning to it whenever nonmentalizing comes to dominate the interaction.

Finally, interventions are carefully matched to the mentalizing capacities of the patient. It is no good offering complex interventions that require considerable thought and appraisal to an individual functioning in psychic equivalence mode. At best, this serves to take over their mentalizing for them, rather than facilitating its rekindling. As stated earlier, nonmentalizing in the patient cannot be recovered by mentalizing in the clinician, but only by reactivating mentalizing in the patient. This is achieved through a series of steps, which underpin the trajectory of every MBT session and may recur several times within each session.

Trajectory of MBT Session and Interventions

Empathic Validation. The initial step in a session is listening to the patient's narrative. Listening to the patient's story allows the clinician to begin by empathically validating their experience. Empathic validation and establishing a shared affective platform held between patient and clinician increases the patient's experience that they are not alone and indicates that another mind can be useful to clarify mental states and increase a sense of agency. This is the first step to advancing the development of "we-ness" ability. The intervention needs to be skillfully implemented to instill a sense of being together with a shared focus. Increasing focus on affect and interpersonal interaction during a session and over time provides the context in which to explore ever more complex states of mind within an attachment context that would normally trigger loss of mentalizing.

Empathic validation requires the clinician to find something in the story that they can empathize with. This is not the same as behaving in a sympathetic manner or saying things that repeat the patient's story. Empathic validation seeks to engender in the patient a sense that the clinician really understands the patient, their internal state, and the issue they are talking about. This is the clinical equivalent of attachment-based contingent responsiveness.

Validation is an affectively based intervention with important cognitive components; the key component is the creation of a sense of alignment with the patient's internal emotional state. This is accomplished by demonstrating an appreciation of the experience and the consequent secondary emotions triggered by a powerful emotional reaction (e.g., when the patient’s fury with a partner engenders deep distress and anxiety). In contrast, a lack of appreciation of the patient’s emotional experience and its impact on the patient’s current state (noncontingent responsiveness on the part of the clinician) is likely to trigger nonmentalizing (e.g., pseudomentalizing) or generate avoidant or other insecure and disorganized attachment strategies. Once contingent responding has increased collaboration and reduced the patient’s arousal, maintaining emotions at a manageable level, the clinician can consider sensitive but less contingent responses to try to stimulate mentalizing about the story the patient brings.

Clarification and Exploration. The second step is clarification and exploration (see case example, later). The “story” the patient brings is clarified. This is not clarification of facts about the narrative or events, although this must also take place. The clinician establishes the facts as quickly as possible. For example, if the patient speaks about an act of self-harm or a suicide attempt, a drunken brawl, or an emotional outburst, the clinician quickly elicits when it occurred, who was there, what were the circumstances, and so on. This will indicate the level of risk and provide other important information.

More than this, though, the MBT clinician wishes to contextualize events with mentalizing. Clarification establishes the reflection the patient has on the events—what was their “premorbid” state of mind, what were their hopes, what was their experience when they were waiting for their boyfriend to return home, what thoughts intruded into their mind, what feelings did they identify, and can they reflect on it differently now? This process of clarification, in the service of engaging mentalizing, links inextricably with affect identification and exploration, which is the third step.

Affects and Affect Identification. Affects and interpersonal relationships reciprocally interact and are core to the personality problems characteristic of BPD. Unmanageable emotions impinge on relationships, and relationships stimulate powerful feelings. Patients may not be able to identify their feelings accurately but experience them primarily as inchoate bodily experiences. Working with the patient to identify a range of feelings is part of the clarification and exploration component of MBT (see case example, later).

Sometimes, emotions in specific contexts have to be normalized. Too often, patients feel that their experience is wrong; in effect, they invalidate their own internal perceptions and feel ashamed. It may be that their feeling is appropriate but excessive, or at other times inexplicably absent. All this becomes clear if the clinician systematically focuses on affects. Initially, the affects associated with the events are established, then any reflection on those feelings is clarified, followed by eliciting current concerns about the events.

Clarification of current affect in the session is included as the third step if the patient and clinician retain their capacity to jointly mentalize around the focus. This is more than asking the patient how they feel at the moment or how they “feel about” an experience they reported, although this may be an initial component. The process requires the identification of *current* affect *related to talking about events to the clinician in the session.*

This expansion from identifying affects in relation to events to affect experienced while talking to the clinician about the events is named the *affect focus* of the session. The patient may initially bring intense rage into the session in relation to an experience of rejection, which is gradually clarified as having been caused by characteristic inappropriate behaviors on the part of the patient. The affect focus turns out to be a sense of humiliation the patient experiences in once again failing to manage their own actions in a more effective manner. This is a subdominant theme that is then explored to increase the complexity of personal meaning and understanding. The aim throughout this process is to build a robust platform of multilayered mentalized experience jointly recognized and shared by the patient and clinician, which incorporates the complexity of mental states and moves the patient beyond a nonmentalizing, narrow oversimplification of their uncontextualized feelings.

The affect focus, that is, identification of the interpersonal interaction in the session and the associated affect, if accurate, heightens the focus on the clinician–patient interaction in the moment of the session. This often indicates that a patient’s attachment strategies and relational patterns, or possibly those of the clinician, are being activated. Thus, it allows a move toward *mentalizing the relationship,* the final step of the intervention trajectory.

Presenting Alternative Perspectives. Before extending to incorporate the relationship with the clinician into the patient’s narrative, the clinician endeavors to enhance the patient’s mentalizing by broadening the patient’s perspectives on an event. This follows naturally from clarification and the affect focus and entails elaborating the mental states of the participants in an event. However, it gradually moves the focus from emotion to cognition. The clinician explores alternative ways of looking an event, at first playfully but sometimes in a mildly challenging manner. If the affect focus helped to reestablish more balanced mentalizing, then considering additional possibilities of what may have happened, at the level of thoughts and feelings, helps further re-engage the patient’s mentalizing processes.

Alternative perspectives may entail reconsiderations of entire scenarios from the point of view of the patient’s hypotheses about the thoughts and feelings of others, or a re-evaluation of the putative sequence of the patient’s reactions—reframing their version of events in the light of alternative thoughts and feelings which they may have also been experiencing. The aim of alternative perspectives is the gentle expansion of the patient’s mentalizing, to move from the certainty of affect toward the doubt of cognition. In general, the move is also from situational (external determinants) to exploring putative internal states. If the patient’s focus was exclusively on their own state of mind the alternative perspective may bring in the other, while if the focus was on all protagonists other than the patient, the alternative perspective can focus on the self. Overall, the clinician engenders a recognition that a pause for reflection and explicit mentalizing may be of value in addition to intuition and the certainty automatic mentalizing can bring.

Mentalizing the Relationship. The groundwork for mentalizing the relationship will have been done through the development of the formulation, which identifies the predominant attachment strategies of the patient.

Mentalizing the relationship in a session is conceptualized as a training ground for managing difficult feelings in interpersonal situations in daily life through maintaining mentalizing while within an emotional interaction. It is the process that fosters the development of a capacity for “we-ness.”

The authors have identified a number of steps for the clinician to consider, which follow closely the three steps of technique described above. First, the clinician has to empathically validate the patient’s perception of the clinician. If the patient says that they experience the clinician in a particular way, then the clinician needs to find part of that experience that they can validate. The clinician actively avoids invalidating the patient's experience.

Second, the clinician needs to work out their contribution to the patient's experience of the clinician. The clinician does this explicitly by thinking aloud about it and asking the patient to explain how they have come to that conclusion. This questioning must be authentic and genuinely curious and must not come from a perspective that implies that the patient's experience is inaccurate or a reemergence of the past distorting the present. Such an invalidating attitude will lead to therapeutic rupture because invalidation (a noncontingent response) will lead to increased arousal and a consequent reduction in mentalizing. Mentalizing the relationship can meaningfully take place only in the context of mentalizing.

Once the clinician has accepted their role in the relational process, the next step of more detailed exploration can occur. In this step, the aim is to generate a more complex understanding of the relationship, to see it from a different angle, and to see what its relevance is for the patient's life. It is not to engender insight in the sense of understanding the operation of the past in the present, although this may arise as part of the patient's broader understanding of their emotional experience.

Critically, the past is used only sparingly as an explanation of the present, in case considering the past leads to a nonmentalizing, reductionist "short-cut" that obscures rather than elaborates the patient's current experience. In MBT, the clinician is cautioned about offering historical interpretation in the context of nonmentalizing process. For the exploration of the therapeutic relationship to contribute to enhanced mentalizing, the clinician works toward increased complexity and the establishment of multiple mental models of relationships.

Mentalizing the *counter-relationship,* or the feeling in the clinician, is the counterweight to mentalizing the relationship. The feelings and mind state of the clinician are given considerable weight in MBT—not as representing the patient's projected feeling, but as a meaningful aspect of an interactive relationship, to be used to demonstrate how minds affect minds. MBT clinicians monitor their experience of the patient. Not knowing what to say to the patient may be the best indicator of the patient's inadequate mentalizing. The dominance of the pretend mode in the patient may be indicated by a sense of boredom, while teleologic mode may be indicated by a sense of confusion and anxiety. The interaction becomes the subject of useful concern, removing obstacles to mentalizing and enhancing accurate mentalizing of the relationship.

For example, if the clinician is frightened of a patient with ASPD, this is an important feeling in the clinician that interferes with treatment and dictates the form of the therapeutic relationship. The clinician finds a way of expressing their experience of the patient in a way that makes it palatable and recognizable as something worth exploring.

The authors recommend that this is done through a number of steps. First, the clinician works out exactly what their feeling is and what it relates to in the patient-clinician interaction. Second, the clinician considers the patient's likely response to an explicit statement of this current feeling state and states this before talking about the current feeling. Third, in the dialogue, the clinician identifies the experience as their own, marks it, and, finally, monitors the patient's reaction to the statement:

"It is possible that what I am going to say may make you feel I am telling you off or being critical but I assure you that is not the case (anticipating the response of the patient).

The problem is that when you sit forward like that and raise your voice, I start to feel anxious and under threat (identifying the behavioral evidence and focus of external mentalizing, presenting his own affect and the effect it has on him).

I realize that this may be me (marking the feeling) but it makes it difficult for me to concentrate on what you are talking about (additional effect on him, interfering with the relationship)."

From here the patient's reaction can be taken into account and the session can continue. But if the threatening attitude and angry presentation are something that permeates all the patient's relationships, then further exploration is essential.

In conclusion, the key to MBT is to:

1. Organize treatment around a collaboratively agreed formulation
2. Develop a focused narrative around the problems of the patient, especially interpersonal issues
3. Infuse the narrative with mentalizing process and prevent a collapse into nonmentalizing experience
4. Work with the patient to reinstate mentalizing when it is lost to prevent destructive behavior and personal distress. Mentalizing oneself with

> others is the basis of satisfactory social and personal relationships, which so often is the goal of people with psychiatric problems and yet seems to them so unreachable

CLINICAL ISSUES

Indications

MBT is effective in treatment of severe BPD. Patients in the early studies of the intervention (see Research and Evaluation, later) had made serious suicide attempts, been admitted to psychiatric hospital for risk, and/or had self-harmed. Both men and women were included in trials and patients showed high levels of comorbidity. Analysis of the data suggested that patients who showed comorbidity for a number of personality disorders, including ASPD, did preferentially better in MBT than comparison treatment. At a clinical level, patients with marked interpersonal problems who have a personality disorder rooted in mentalizing vulnerability and attachment problems may benefit from MBT.

Limitations

Patients treated with MBT show a reduction in life-threatening behaviors and distressing psychiatric symptoms at the end of treatment, require less mental health care, and demonstrate improved social and interpersonal functioning. Over 8-year follow-up of patients in an RCT, the number of patients meeting primary recovery criteria was significantly higher in patients treated with MBT compared with those receiving a well-structured control treatment. Participants treated with MBT showed better functional outcomes in terms of being more likely to be engaged in purposeful activity and reporting less use of professional support services and social care interventions than those in the control group. Nevertheless, many of the patients did not reach a level of normal social adaptation.

It should be noted that the evidence for mentalizing mediating therapeutic change in MBT is currently limited (see Research and Evaluation).

Complications

Psychotherapy can be harmful and MBT is probably no exception. However, MBT pays particular attention to ways in which patients may be harmed by treatment. On the basis that patients with BPD are uniquely sensitive to attachment process, and the stimulation of attachment reduces mentalizing, the MBT clinician focuses on levels of arousal in treatment sessions, constantly trying to balance arousal and mentalizing, ensuring that the therapeutic relationship is not a source of excessive attachment stress. Overstimulation of patients with avoidant attachment patterns is likely to trigger retreat and drop-out. Drop-out rates in MBT in clinical services in the United Kingdom are around 15% of people offered treatment, although in Scandinavia drop-out is as low as 2%. This suggests that the treatment overall is acceptable to people with BPD.

The aim of treatment is to increase the robustness of the patient’s mentalizing capacity, and yet clinicians in many psychotherapeutic modalities may often tell patients “how they feel” or “what they are really saying.” This undermines the mentalizing of the patient and so is avoided in MBT.

Contraindications

MBT is a generic treatment constructed to optimize access both by creating a low-demand treatment protocol and by facilitating access to training by a range of professionals. Individuals who have problems with mentalizing rooted in nonattachment contexts may not benefit from MBT; if they do, the mechanism of change should be assumed to be different from that in BPD. Individuals with relatively simple problems, such as phobias or uncomplicated depression, may do better with more direct approaches, such as cognitive-behavioral therapy. Even within a population of patients with BPD, patients with more complex presentations (multiple personality disorder diagnoses) are more likely to require MBT than those with a single BPD diagnosis, who may do as well in structured clinical management.

A 24-year-old patient, Sharon, was referred following a number of suicide attempts. She self-harmed, engaged in frequent polydrug misuse, and described emotionally volatile and occasionally violent relationships with men. She had a 2-year-old daughter who had recently been removed by child protection services. The clinician was able to take a detailed history, which elicited a number of vulnerability factors. Sharon was taken into care at the age of 5 and had a number of foster placements. She experienced recurrent sexual abuse from a caregiver around the age of 8. Her behavior was described in social reports as “oversexualized” around the time of puberty, and she attended an adolescent unit from the age of 13 to 15 years. She left school at 16 years having not attended for the previous year. The formulation was developed, which included these vulnerability factors but also identified the interpersonal vulnerabilities that led to mentalizing failures. The clinician explored two relational events with her boyfriend in which she had been violent and two contexts in which she had tried to kill herself. Unsurprisingly, these events were interlinked.

FORMULATION AND NONMENTALIZING INTERACTIONAL PATTERNS

In the formulation of MBT, mentalizing vulnerability points are identified and placed in the context of attachment strategies. In this patient, there was evidence that she had an insecure attachment with marked ambivalence and some disorganization. She would seek reassurance or indicators of care from her boyfriend, but when he did not respond, she became desperate and increasingly demanding and clinging, eventually attacking him. She stated that if she could not “have him,” she would “prefer to die,” or alternatively, she would “trap him” forever by getting pregnant. This nonmentalizing interactional cycle of need, demand, rejection, coercion, and the associated nonmentalizing modes were jointly explored, written down, and shared with the patient. It is anticipated that the pattern will occur in relation to the clinician in a less intense way—the patient might seek some reassurance, for example, and the clinician may fail to respond contingently, triggering feelings of rejection. Identification and reduction of the nonmentalizing interactional process was an initial goal of treatment because the interactions occurred frequently and were an important area of vulnerability leading to suicide attempts.

Sharon reported that she had been in a fight with her boyfriend. She had telephoned her boyfriend to find out when he would be back from work because she was looking forward to seeing him and telling him that she loved him. He said that he was leaving soon and would be back within an hour. A few minutes later, a friend telephoned and, in their conversation, said that she had seen Sharon's boyfriend in a bar with a blonde woman only 30 minutes ago. When the boyfriend arrived home, Sharon asked why he had said that he was at work when she knew he was in a bar with a blonde woman. He said that he had called in for a drink on the way home. They had an argument during which Sharon attacked him and then, after smashing some crockery, cut herself.

CLARIFICATION AND ELABORATION AND AFFECT IDENTIFICATION

Clarification of mental states about this event suggested that Sharon initially felt that she was looking forward to seeing her boyfriend, which was why she phoned him. His response was "contingent" on her feeling for him—he said he would be home soon. When her friend reported having seen him in a bar, this feeling was replaced with doubt about his love for her and a sense of rejection. Rapidly, she started to have thoughts that he was having an affair with the blonde woman, and in the context of feeling a loss of being loved, this was experienced in psychic equivalence—"I am unlovable"—with some elements of hypermentalizing about his motives—"he was having an affair, and I am certain of it." This was the nonmentalizing cycle and associated modes identified in her formulation.

When Sharon's boyfriend arrived home, there was no doubt in her mind about what had happened. He was unable to persuade her it was not true. In nonmentalizing states, actions are the only meaningful way of communicating, and she was therefore left with being coercive to "make" him love her. But to do this, she tried to lock him in the house so he could not go out, and she demanded sex. He resented this, refused, and the violence occurred. At this point, her demand was driven by the teleologic belief that only his actions could prove his love for her.

The task in MBT is now to rewind in time and identify how Sharon manages her initial excitement about seeing her boyfriend, the sudden collapse of these feelings, and her collapse into psychic equivalent thoughts about his infidelity. To do this, the clinician engages in the detailed delineation of her mental states while asking her to re-present them to herself to instill a sense of uncertainty: can she manage emotional turmoil between excitement and disappointment; can she question her certainty over her boyfriend's activities and motives; can she engage with him to establish a more robust way of managing the distrust in their relationship other than teleologic demand?

AFFECT SESSIONAL FOCUS AND MENTALIZING THE RELATIONSHIP

The clinician asked Sharon how she felt about her boyfriend now, and she said that she still felt that he did not care for her. She was miserable. But, in addition, she felt that she had created problems, which was "typical of me." This is the identification of the current affect in relation to the focus. But, as she was talking to the clinician, she said that she felt that the clinician was judging her and would give up on treating her as it had happened again. She was a failure. This is the identification of the current affect in the session and was quickly explored. At this point, Sharon became coercive to some degree, suggesting that the clinician could not stop seeing her, and to prevent it, she would take an overdose. It becomes possible at this point to mentalize the relationship. Sharon thinks that the clinician will not meet her needs and so automatically engages in a nonmentalizing interactional process. This needs to be discussed in the context of the clinician stating that it is not in his mind to stop seeing her in treatment (statement of clinician counter-relationship), so it is important to understand what is fueling her belief and questioning it. It is not for the clinician to interpret the repetition in the session of the interaction with the boyfriend but to define the emotional need that Sharon reacts to that makes her become coercive.

Goals of Treatment

The primary aims of MBT are to rekindle mentalizing when it is lost, to maintain it when it is present, and to increase the resilience of the individual's capacity to keep it going when it would otherwise be lost. But mentalizing is about something. So, the development and maintenance of mentalizing is initially focused on the core symptoms of BPD (or another condition), such as suicide attempts, self-harm, and other self-destructive behaviors and emotional instability. This is followed by emphasis on interpersonal problems because the key area of vulnerability to losing mentalizing in BPD is the interpersonal domain, especially when attachment processes are activated. So the clinician–patient relationship is a significant area of scrutiny. Patient and clinician increase attentiveness to mental states and interpersonal contexts in which they become disordered.

ETHICAL ISSUES

A key feature of MBT is the collaborative stance, which engages patients in a mutually agreed protocol, each phase of which focuses on achieving shared understanding between the patient and the clinician. It is not for the MBT clinician to impose cultural values and beliefs. In other words, it may be argued that MBT clinicians are less likely to encounter ethical issues than proponents of other therapeutic approaches in which collaboration is perhaps seen as a necessary condition but not the aim of the treatment.

However, MBT clinicians should be well aware that the nature of the clinical problems they are dealing with invariably and unavoidably creates ethical issues daily. Working with a person whose capacity to represent themselves in an agentive way is lacking, whose attachment system is disorganized, and who is extremely vulnerable to creating a dependent relationship, will present a consistent problem for the clinician to avoid exerting undue influence. MBT clinicians—like therapists working in other orientations—can prolong relationships beyond their useful timespan, citing the patient's need for their input as justification.

In particular, when patients pay for their treatment either themselves or through a third party, financial exploitation is a real and present risk. Beyond this, undue influence may manifest through the uncritical acceptance of the clinician's frame of reference: a nonmentalizing individual has no alternative and gratefully grabs hold of a powerful, coherent model when it is presented to them. The teleologic predisposition of patients sometimes leads clinicians to find themselves caught up in action-oriented attempts to address the patient's distress, offering more than they should, yet delivering less than they could.

RESEARCH AND EVALUATION

MBT for BPD

There have been several reviews of psychosocial interventions for BPD. DBT and MBT have the highest numbers of primary trials, with MBT having been investigated in seven RCTs. The reviews recognize the evidence base for MBT for BPD as generally inferior only to that for DBT, although not from a health-economic perspective. A recent Cochrane review concluded from a meta-analysis that MBT is more effective than treatment as usual (TAU) at reducing self-harm, suicidality, and depression. Encouragingly, a large qualitative study of patients' treatment goals established that the goals of MBT were closely allied with what patients hoped to gain from their therapies. A number of naturalistic studies have also tested the effectiveness of the MBT approach for BPD patients.

In a RCT of MBT for BPD in a partial hospital setting (MBT-PH), an 18-month program achieved significant and enduring changes in mood states and interpersonal functioning. Outcome measures included frequency of suicide attempts and acts of self-harm, number, and duration of inpatient admissions, service utilization, and self-report measures of depression, anxiety, general symptom distress, interpersonal function, and social adjustment. The benefits, relative to TAU, were large, with a number needed to treat approximately two; in addition, the benefits were observed to increase during the follow-up period of 18 months. Analysis of participants' healthcare use suggested that day hospital treatment for BPD was no more expensive than general psychiatric care and showed considerable cost savings after treatment.

A follow-up study of BPD patients 5 years after all treatment was completed and 8 years after initial entry into treatment, comparing patients treated with MBT and those receiving TAU, found that those who received MBT remained better than the TAU group. Superior levels of improvement were shown on levels of suicidality (23% in the MBT group vs. 74% in the TAU group), diagnostic status (13% vs. 87%), service use (2 years vs. 3.5 years), and other measurements such as the use of medication, global functioning, and vocational status.

These early positive findings have been replicated by independent groups comparing MBT-PH with specialist treatment as usual (S-TAU), with MBT-PH being associated with higher acceptability to BPD patients compared with S-TAU, as reflected in significantly higher early drop-out rates of 34% in S-TAU versus 9% in MBT. A further study compared MBT-PH with MBT delivered as a less intensive outpatient treatment (MBT-IOP). MBT-PH and MBT-IOP both showed significant improvements of moderate to large effect sizes on all measures, but neither treatment was superior to the other on the primary outcome measure, despite the difference in treatment intensity. However, MBT-PH showed more rapid change and a tendency to be more effective on secondary outcomes, particularly in terms of relational functioning. Economic analysis revealed that MBT-PH, the more intensive and expensive treatment, had some advantage, albeit at additional cost, in effecting greater likelihood of remission.

Well-controlled single-blind trials of outpatient MBT have been conducted with adults with BPD and with adolescents presenting to clinical services with self-harm, the vast majority of whom met BPD criteria. In a study comparing MBT-IOP for adults with BPD with structured clinical management, MBT was found to be more effective on nearly all outcome measures, and the improvements were maintained at 5-year follow-up. A study on suicidal and self-harming adolescents found MBT to be superior to TAU in reducing self-harm, including suicidality, and depression. Importantly, in the adult trial, the control group received a manualized, highly efficacious treatment, namely structured clinical management; MBT was superior to this intervention, particularly in the long term.

A posthoc analysis of moderators found that the number of personality disorder diagnoses in addition to BPD was the key indicator of severity that predicted the need for the MBT approach, as structured clinical management appeared to have little benefit on most outcome measures among these patients. Furthermore, in the trial with an adolescent sample, improvements generated by MBT appear to have been mediated by improved levels of mentalizing, reduced attachment avoidance, and amelioration of participants’ emergent BPD features; participants treated with MBT showed a recovery rate of 44%, compared with 17% of those who received TAU. Ongoing follow-ups of both these trials indicate that improvements in the MBT groups have been at least maintained, and in most cases, improvements continued after treatment termination and differences relative to the comparison group remain significant.

Further small studies on adolescents suggest that those who engage in an MBT group show improvement in borderline symptoms, depression, self-harm, peer attachment, parent attachment, mentalizing, and general psychopathology including social anxiety and emotion regulation. The mechanism of change may be through improved mentalizing. In one study enhanced trust in peers and parents in combination with improved mentalizing capacity was associated with greater decline in borderline symptoms. In another study, mentalizing emerged as a significant predictor of change over time in self-reported self-harm and hospital presentation for self-harm.

Additional independent studies provide further support for the efficacy of MBT in BPD in adults. An RCT from Denmark investigated the efficacy of MBT versus a less intensive, manualized supportive group therapy in patients diagnosed with BPD. Patients were randomly allocated to MBT ($n = 58$) or manualized supportive therapy ($n = 27$). Each intervention was delivered in combination with psychoeducation and medication. Both the combined MBT treatment and the less intensive supportive therapy brought about significant improvements in a range of psychological and interpersonal measures (e.g., general functioning, depression, and social functioning) and decreased the number of diagnostic criteria met for BPD; effect sizes were large ($d = 0.5$ to 2.1). The combined MBT was superior to the less intensive supportive group therapy on clinician-rated Global Assessment of Functioning.

An 18-month naturalistic follow-up found that treatment effects at termination were sustained at 18 months. Half of the patients in the MBT group met criteria for functional remission at follow-up, compared with less than one-fifth in the supportive therapy group, but three-quarters of both groups achieved diagnostic remission, and almost half of the patients had attained symptomatic remission. A limitation of this study is that the same clinicians delivered both interventions (and thus there was a high risk of spillover effects between the two treatments); incomplete data was a further significant limitation.

In a second study from Denmark, a cohort of patients treated with partial hospitalization followed by group MBT showed significant improvements after treatment (average length 2 years) on a range of measures, including Global Assessment of Functioning, hospitalizations, and vocational status, with further improvement at 2-year follow-up. In the New Zealand public health service, an RCT of MBT delivered by clinicians with brief training and some ongoing clinical supervision compared MBT with enhanced TAU. Both groups showed significant gains in reducing self-harm and suicide attempts, with little significant difference between groups at the end of treatment.

A quality improvement study examined the outcomes for BPD patients treated in an MBT program in a Norwegian specialist treatment unit compared with a former psychodynamic treatment program. This longitudinal comparison had a sample of 345 BPD patients, including 282 patients treated on the psychodynamic program and 64 who received MBT, who had comparable baseline severity and impairments of functioning on all measures. Outcome measures included Symptom Checklist-90 symptom distress, interpersonal problems, and global functioning assessed routinely throughout treatment, and suicidal/self-harming acts, hospital admissions, medication, and occupational status assessed at baseline and discharge.

The change in the program from traditional psychodynamic therapy to MBT led to a reduction in unplanned discharges (MBT had a low drop-out rate of 2%). Measured benefits from the change of program included greater improvements in symptom distress and interpersonal, global, and occupational functioning. Although the change was associated with the introduction of MBT, specific causal attributions are hard to establish in such a design. The same group reported that greater clinical severity was associated with poorer rates of improvement in psychodynamic psychotherapy but was not associated with significant differences in outcomes for patients in MBT. Differences in outcomes for patients in MBT and psychodynamic therapy increased significantly with greater severity of the disorder. This observation is consistent with earlier evidence that MBT is more effective than comparison interventions in patients with more severe symptoms.

One notable naturalistic study in the Netherlands investigated the effectiveness of an 18-month manualized program of MBT in 45 patients diagnosed with severe BPD. There was a high prevalence of comorbidity of DSM-IV Axis I and Axis II disorders. Results showed significant positive change in symptom distress, social and interpersonal functioning, and personality pathology and functioning; effect sizes were moderate to large (d = 0.7 to 1.7). The study also showed that the use of additional treatments and psychiatric inpatient admissions decreased significantly during and after treatment. The lack of a control group in this study limits the ability to draw conclusions about the efficacy of MBT.

In a related study, applied propensity score matching was used to determine the best matches for 29 MBT patients from within a larger (n = 175) group who received other specialized psychotherapeutic treatments. These other specialized treatments yielded improvement across domains, which was generally only moderate; in contrast, pre–posteffect sizes were consistently large for MBT, with Cohen's d for a reduction in psychiatric symptoms of −1.06 and −1.42 at 18 and 36 months, respectively, and ds ranging from 0.81 to 2.08 for improvement in domains of personality functioning. Given the nonrandomized study design and the variation in treatment dose received by participants, the between-condition difference in effects should be interpreted cautiously.

MBT has ameliorating effects on some of the neurocognitive problems of people with BPD, perhaps through enhancing social-cognitive skills by altering neural systems underpinning some aspects of neurocognition, such as executive function and attention, which support mentalizing ability. Patients with attentional and reasoning deficits showed increased sustained attention and better perceptual reasoning than controls after 6 months of treatment. Initial deficits in sustained attention among patients resolved after treatment. Patients also showed improved emotion regulation associated with better working memory capacity, and enhanced perceptual reasoning was accompanied by better interpersonal functioning.

MBT for Antisocial Personality Disorder (ASPD)

Research into treatment for ASPD up to 2009 is summarized in the United Kingdom's National Institute for Health and Clinical Excellence clinical guideline for ASPD, which confirmed that interventions for ASPD are poorly researched and that evidence on its treatment is scarce. The authors of the guideline concluded that the evidence for treatments for ASPD was extremely limited and did not support the development of any guidance on treatment recommendations. Two Cochrane reviews concluded that there was no consistent evidence for any intervention for ASPD and recommended that research to test interventions for the disorder is urgently needed.

A feasibility study of MBT for ASPD reports findings from a small sample (*n* = 9) receiving group and individual MBT. Preliminary results on the Overt Aggression Scale suggested that the participants rated the severity of their aggression toward others and themselves as decreasing over the first 6.5 months of treatment; in contrast, their rating of irritability did not change. Psychiatric symptom severity on the Brief Symptom Inventory showed a reduction in the distress participants experienced in relation to their symptoms at a 6-month follow-up, with participants reporting the greatest decreases in distress resulting from symptoms of depression, anxiety, and hostility.

A number of trials on the effectiveness of MBT have not excluded patients with co-occurring ASPD. A separate analysis of individuals with comorbid ASPD revealed that they benefited significantly from MBT. Patients showed significant reductions in anger, hostility, paranoia, and frequency of self-harm and suicide attempts compared with controls, and there were significant improvements in negative mood, general psychiatric symptoms, interpersonal problems, and social adjustment.

FURTHER READINGS

*Allen JG. *Restoring Mentalizing in Attachment Relationships: Treating Trauma With Plain Old Therapy*. American Psychiatric Association Publishing; 2012.

Asen E, Fonagy P. *Mentalization-Based Treatment With Families*. Guilford Press; 2021.

Bales DL, Timman R, Luyten P, Busschbach J, Verheul R, Hutsebaut J. Implementation of evidence-based treatments for borderline personality disorder: The impact of organizational changes on treatment outcome of mentalization-based treatment. *Personal Ment Health*. 2017;11(4):266–277.

*Bateman A, Campbell C, Luyten P, Fonagy P. A mentalization-based approach to common factors in the treatment of borderline personality disorder. *Curr Opin Psychol*. 2018;21:44–49.

Bateman A, Fonagy P. Randomized controlled trial of outpatient mentalization-based treatment versus structured clinical management for borderline personality disorder. *Am J Psychiatry*. 2009;166(12):1355–1364.

*Bateman A, Fonagy P. *Mentalization-Based Treatment for Personality Disorders: A Practical Guide*. Oxford University Press; 2016.

*Bateman A, Fonagy P, eds. *Handbook of Mentalizing in Mental Health Practice*. 2nd ed. American Psychiatric Association Publishing; 2019.

Bateman A, Holzer S, Constantinou MP, Fonagy P. Eight-year prospective follow-up of mentalization-based treatment versus structured clinical management for people with borderline personality disorder. *Personal Disord*. 2021;12(4):291–299.

Bateman A, O'Connell J, Lorenzini N, Gardner T, Fonagy P. A randomised controlled trial of mentalization-based treatment versus structured clinical management for patients with comorbid borderline personality disorder and antisocial personality disorder. *BMC Psychiatry*. 2016;16:304.

Debbané M, Salaminios G, Luyten P, et al. Attachment, neurobiology, and mentalizing along the psychosis continuum. *Front Hum Neurosci*. 2016;10:406.

Fonagy P, Allison E, Campbell C. Mentalizing, resilience, and epistemic trust. In: Bateman A, Fonagy P, eds. *Handbook of Mentalizing in Mental Health Practice*. 2nd ed. American Psychiatric Association Publishing; 2019:63–77.

Fonagy P, Campbell C, Bateman A. Mentalizing, attachment, and epistemic trust in group therapy. *Int J Group Psychother*. 2017;67:176–201.

Fonagy P, Luyten P. A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Dev Psychopathol*. 2009;21(4):1355–1381.

*Fonagy P, Luyten P, Bateman A. Treating borderline personality disorder with psychotherapy: where do we go from here? *JAMA Psychiatry*. 2017;74(4):316–317.

Fonagy P, Yakeley J, Gardner T, et al. Mentalization for offending adult males (MOAM): study protocol for a randomized controlled trial to evaluate mentalization-based treatment for antisocial personality disorder in male offenders on community probation. *Trials*. 2020;21:1001.

National Institute for Health and Clinical Excellence. *Antisocial Personality Disorder: Treatment, Management and Prevention*. National Institute for Health and Clinical Excellence; 2009.

Rossouw TI, Fonagy P. Mentalization-based treatment for self-harm in adolescents: a randomized controlled trial. *J Am Acad Child Adolesc Psychiatry*. 2012;51(12):1304–1313.e3.

*Sharp C, Vanwoerden S. Hypermentalizing in borderline personality disorder: A model and data. *J Infant Child Adolesc Psychother*. 2015;14(1):33–45.

Smits ML, Feenstra DJ, Eeren HV, et al. Day hospital versus intensive out-patient mentalisation-based treatment for borderline personality disorder: multicentre randomised clinical trial. *Br J Psychiatry*. 2020;216(2):79–84.

*Stoffers JM, Völlm BA, Rücker G, Timmer A, Huband N, Lieb K. Psychological therapies for people with borderline personality disorder. *Cochrane Database Syst Rev*. 2020;5(5):CD012955.

Weijers J, Kate CT, Viechtbauer W, Rampaart LJA, Eurelings EHM, Selten JP. Mentalization-based treatment for psychotic disorder: a rater-blinded, multi-center, randomized controlled trial. *Psychol Med*. 2021;51(16):2846–2855.

32.15 Narrative Psychiatry

BRADLEY LEWIS, M.D., PH.D.

More than anything else psychiatrists do, they listen to stories. These stories so saturate the clinical encounter that it would be impossible to imagine a clinical encounter without them. In the very first meeting between psychiatrist and patient, the psychiatrist begins with an open-ended invitation to a story: "What brings you here?" or "What seems to be the problem?" Patients respond to these questions by telling psychiatrists about their life, their troubles, when the troubles began, what seems to have caused them, how they create difficulty, and what kinds of problem solving they have tried. Such stories may be rudimentary, they may be only partially worked out, and they may even be baffled and confused. The patient may even be perplexed enough to answer, "I don't know why I came" or "I'm not really sure what's wrong, my family sent me." Nonetheless, the patient's response to the psychiatrist's initial questions always involves a story.

As the interview proceeds, going into greater detail about the person's life and troubles, the psychiatrist and patient never leave the story frame. After the initial broad questions, focused questions aimed at particular symptoms also invite stories. Even finely-grained, highly medical-sounding questions like "Do you have any medical problems?" or "Does depression run in your family?" involve stories. Moreover, the clinical evaluation is only the beginning of the story-centered encounter between the psychiatrist and the patient. Follow-up interviews and meetings go deep into a clinical treatment process, including those that only involve medication checks once every so often and are also saturated with stories. "How have you been?" "What has happened since we last spoke?" "Are you having any side effects with this medication?" These questions all elicit stories from the patient.

One might ask what is the importance of these stories in clinical life? Until recently, little attention has been paid to the question of stories within clinical settings. A patient's narrative of their troubles were seen as merely a vehicle to get across the facts. The “art” of clinical work was to abstract away the facts embedded in patient’s stories, but the stories themselves were secondary. In the last several years, however, the view of narrative as being merely functional has been thrown into question, and, increasingly, clinicians see stories as indispensable for understanding human phenomena. For these clinicians, stories are just as important as facts because it is the story that allows people to select, organize, and shape the facts that matter.

Narrative psychiatry emerges out of this increased interest in clinical stories. The two main tributaries that lead to narrative psychiatry come from the two different sides of psychiatry: “narrative medicine” on the one side and “narrative psychology and psychotherapy” on the other. Narrative psychiatrists are psychiatrists who combine the wisdom of these two domains.

Following the lead of narrative medicine, narrative psychiatrists recognize that psychiatric patients, like medical patients, come to clinics with intense stories to tell. Contemporary narrative medicine has developed from decades of work in bioethics and medical humanities devoted to humanizing the clinical encounter through a better understanding of patient stories. The term “narrative medicine” comes from Rita Charon, an internist and literary scholar, who used it to describe an approach to medicine that uses narrative approaches to augment scientific understandings of illness. Narrative medicine brings together insights from human-centered medical models, such as George Engel’s biopsychosocial model and Eric Cassel’s person-centered model, with research and insights from phenomenology, the humanities, and interpretive social sciences.

Narrative medicine uses these resources to better understand the illness experience, “to recognize, absorb, interpret, and be moved by the stories of illness.” As Charon argued, when clinician’s possess narrative competency, they can enter the clinical setting with a nuanced capacity for “attentive listening …, adopting alien perspectives, following the narrative thread of the story of another, being curious about other peoples motives and experiences, and tolerating the uncertainty of stories.” Charon further argued that doctors “*need* rigorous and disciplined training” in narrative reading and writing not just for their own sake (helping them to deal with the strains and traumas of clinical work), but also “*for the sake of their practice.*” Without such narrative competency, clinicians lack the ability to fully understand their client’s experience of illness. For Charon and others in narrative medicine, narrative study is not a mere adornment to a doctor’s medical training; it is a crucial and basic science that must be mastered for medical practice.

A major task of narrative medicine, and therefore narrative psychiatry, is to be a good listener and to connect empathically with the patient’s story. A narrative psychiatrist, like a narrative physician, seeks to understand the patient first and foremost. This understanding brings patient and clinician together into a shared experience of the patient’s world. This narrative understanding cannot be reduced to psychology, neuroscience, or sociology. Narrative understanding is much more than a causal explanation of problem A or problem B that the patient might have. It does not simply abstract from the person’s situation, a categorical label that groups problems under a well-known abstract grid. Instead, narrative understanding tunes into the uniqueness of the individual and the unrepeatability of the person’s experience and difficulties. Narrative understanding, in short, is a deep appreciation of the person as a whole—what it feels like for this person, in this particular context, going through these particular problems.

By tuning in to the person’s story, first and foremost, as a goal in and of itself, narrative psychiatrists, like their narrative medicine colleagues, create an environment in which the person can be known and their difficulties can be appreciated. This step can be one of the most healing acts a clinician can perform. By allowing the person to be known, the clinician creates a bond that reduces the patient’s isolation because the patient immediately becomes less alone with the problem. Whatever else may happen in the clinical encounter, this sense of togetherness will reduce the person’s suffering. The patient now has a comrade, a foxhole partner, someone who is in it with them. This partnership may not change the extent of the person’s troubles, but it changes immensely the experience of the situation. Although clinicians are not the only ones who can perform this role, a good clinician can sometimes be one of the few people in the person’s world who can truly understand. Clinicians have extensive experience with similar experiences of human suffering, as well as the advantage of working in clinical settings in which many of the blocks to discussing shameful or embarrassing problems often are removed.

In addition to following the lead of narrative medicine colleagues, narrative psychiatrists also follow the lead of contemporary colleagues in narrative psychology and psychotherapy. The history of narrative psychology and psychotherapy goes back to Sigmund Freud's early work at the inception of psychoanalysis. At that time, Freud lamented about how his case histories sounded more like narrative fiction than hard science. As Freud explained, "I have not always been a psychotherapist. Like other neuropathologists, I was trained to employ local diagnosis and electroprognosis, and it still strikes me as strange that the case histories I write should read like short stories and that, as one might say, they lack the serious stamp of science."

Despite this awareness of the role of story in his clinical work, however, Freud never embraced a narrative theory of psychoanalysis, and throughout his career, he remained attached to the natural science model he inherited from neurology.

Contemporary narrative psychology and psychotherapy's motivation for returning to the role of narrative comes partly from the broader turn to narrative in humanities and social science and partly from the history of psychotherapy since Freud. The last century of psychotherapy has been a century of strife, with one faction after another splitting off from psychoanalysis. Leading alternatives to psychoanalysis included behavioral, humanistic, family, cognitive, feminist, and interpersonal, just to name some. All of these splits are characterized by further splits within splits, which has fragmented the field of psychotherapy to the point that there are now more than 400 active approaches to psychotherapy. Narrative approaches emerge at this particular moment as part of an important trend away from further fragmentation and toward psychotherapy integration. Narrative approaches are invaluable for psychotherapy integration because they provide a metatheoretical orientation from which to understand and practice psychotherapy. As Lynne Angus and John McLeod pointed out in *The Handbook of Narrative and Psychotherapy,* narrative approaches allow psychotherapists to understand the common factors within psychotherapy and to appreciate the diversity of psychotherapeutic models without attempting to close these differences down. Thus, in many ways, the contemporary emergence of narrative approaches brings the history of psychotherapy back full circle. More than 100 years since Freud introduced "talking therapy," psychotherapy has returned to its early appreciation of the similarity between case histories and narrative stories.

Similar to narrative psychotherapists, narrative psychiatrists recognize that the stories people tell about themselves not only describe their lives, they also shape their lives. The stories people use to frame and organize their lives are not simply incidental to their identity; they are the primary tools people have for configuring and reconfiguring their identity. Narrative psychiatrists recognize that this shaping of identity through narrative is a major tool for recovery. When people come for help in a therapeutic situation, they not only want to be understood, they also want help making changes in their lives. Narrative psychiatrists collaboratively encourage and foster these changes by helping coauthor new stories of purpose and thriving. As SuEllen Hamkins puts it in *The Art of Narrative Psychiatry,* a narrative approach helps patients discover "untold but inspiring stories of resiliency and skill in resisting mental health challenges while dismantling aspects of presenting narratives that fuel problems. It examines what the doctor's kit of psychiatry has to offer in light of the values and preferences of the person seeking consultation, authorizing the patient as the arbiter of what are helpful and what is not." Working in this way, narrative psychiatrists understand that collaboratively reauthoring stories people tell about themselves is a powerful way to make changes toward recovery goals.

WHAT IS NARRATIVE?

Both narrative medicine and narrative approaches to psychotherapy are deeply informed by narrative theory. But how have writers and literary scholars tried to describe "narrative?" What are the important elements of narrative? How does narrative work in human life? Although these questions may sound straightforward, they are not so easy to answer. The literature on narrative theory is vast, and it has many tributaries. Historian and narrative theorist Hayden White has argued that to "raise the question of the nature of narrative is to invite reflection on the very nature of culture and, possibly, even on the very nature of humanity itself." For clinical work, therefore, it is important to select which aspects of narrative theory are most helpful.

Narrative physician and novelist Abraham Verghese provides a good place to start. Speaking to colleagues at the American Society of Internal

Medicine, Verghese explains that “As physicians, most of us become involved in the stories of our patient’s lives…we become players in these stories. Our actions change the narrative trajectory…and our patient’s stories come to depend heavily on repetition of what we say.” The inescapable implications for medicine are that:

> *Story* helps us link and make sense of events in our lives; we as physicians *create* stories as often as we record them …; and we are characters in [these] various stories, walking on and off the stage in tales that take place in our hospitals and clinics.

Verghese points to examples in which physicians reach narrative competence through years of attentive practice, but he finds a more direct route through combining medical practice with the “tools that the writer possesses,” or what Verghese called the “storytelling craft.” During Verghese’s immersion into the craft of writing, he found that the pillars of narrative invariably involved the author’s selection and organization of *metaphor, plot,* and *character.* Making selections from these narrative elements are “fundamental to good writing in the same way that internal medicine skills rest on understanding the mechanisms behind dyspnea, edema, polyuria, and other cardinal manifestations of disease.”

Following Verghese’s lead, narrative psychiatrists must, at a minimum, understand the narrative elements of metaphor, plot, character, and we can add to this list, point of view. The following sections use narrative theory, particularly the richly synthetic efforts of philosopher Paul Ricoeur, to develop these narrative elements. In addition, recent memoirs of mental illness diagnoses are explored to see how people use narrative elements to tell the story of mental health concerns.

Metaphor

The stories people tell are rich with metaphor. Consider this example from clinical life: “My boss looks at me lately like I am a piece of garbage. In his eyes, I am used up, something to discard and throw away. My whole interactions with him leave me feeling like shit. I feel utterly worthless, and my life is meaningless.” This snippet of a clinical story has at least four metaphors. The person compares herself to garbage, shit, monetary value, and a literary text. Commonsense would indicate that these kinds of metaphors have little importance. Metaphors, most people believe, are embellishments and ornamentations to the content of a story, but they do little else. From this perspective, the person behind the snippet feels bad and believes her boss does not value her. End of meaning. The literal paraphrase is all you need.

For influential literary theorists, however, such as I. A. Richards, a number of philosophers, particularly Max Black, Mary Hesse, Paul Ricoeur, and Mark Johnson, and increasingly for cognitive scientists like George Lakoff, metaphor has a much larger function than simple ornamentation. In their book *Metaphors We Live By,* Lakoff and Johnson put it most succinctly: “metaphor is not just a matter of language, that is, of mere words…On the contrary, human thought processes are largely metaphorical.” Lakoff and Johnson argue that metaphor is a pervasive aspect of human life, not just in the words that are used, but also in the very concepts: “Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature.” By shaping concepts, metaphor structures the way in which the world is perceived, what is experienced, how people relate to other people, and the choices one makes.

Metaphor performs this function by allowing one to understand and experience one thing in terms of something else. Hesse uses the example, borrowed from Black, of “man is a wolf.” Hesse explains that the metaphor works because it transfers ideas and associations from one term to the other. The metaphor selects, accentuates, and backgrounds aspects of two systems of ideas so that they come to be seen as similar: “Men are seen to be more like wolves after the wolf metaphor is used, and wolves seem to be more human.” Hesse goes on, “Nature becomes more like a machine in the mechanical philosophy, and actual, concrete machines themselves are seen as stripped down to their essential qualities of mass in motion.” In short, metaphor allows the two terms, or more precisely the two systems of terms, to interact and adapt to one another.

Understanding metaphor in this way connects to broader work in continental linguistic philosophy, and that work, as a whole, shifts standard ideas about truth and objectivity. It allows one to sidestep the usual binary traps between relativism (anything goes) and realism (there is only one correct or true way to describe the world). When the role of language and metaphor as a mediator between concepts and the world is understood, it no longer makes sense to think in these highly modernist either/or terms. Rather than using a rigid binary splitting between true and false, people can think instead in a postmodern language of semiotic realism and pluridimensional consequences. Or, as Paul Ricoeur puts it, people can let their notions of universal truth (or single objective truth) give way to ideas about “metaphorical truths.”

The multiple metaphors that shape one’s concepts and perception allow one to see and experience the world in a particular way and from particular perspectives. Ricoeur calls these particular ways “metaphorical truths” to emphasize that the metaphorical process does allow one to understand something real about the world. That “reality” is not absolute, and it is always in tension with other ways of understanding something equally real about the world. However, these other ways of understanding are not available to unless people inhabit new or alternative metaphors for perception and description. For Ricoeur, this means that the “seeing as” of metaphor generally also means a “being as” of lived experience.

Plot

The plot works like a metaphor in that it also orders one’s experience and provides a form to the narratives. Plot, or the process of emplotment, adds to metaphor two key dimensions: (1) it brings together what would otherwise be separate and heterogeneous elements, and (2) it organizes one’s understanding and experience or time, or what could be called temporal perception.

The critical function of plot for narrative is that plot creates a narrative synthesis between multiple individual events and brings them together into a single story. It allows an intelligible connection to be made between them. Remarkably, plot can create a synthesis between events and elements that are surprisingly incongruous or heterogeneous—events that do not seem to fit together. In any one story, vastly incommensurable elements can all be brought together through the process of emplotment. Consider this list of diverse elements that might come up in even a small segment of clinical work: human desires, human rationality, human purpose, human agency, unconscious motives, unintended circumstances, unplanned accidents, other peoples’ desires and rationality, their unconscious motives, material force, natural laws, divine intervention, discoveries, nightmares, historical chance, social prejudice, economic downturns, failures of communication, contemporary styles and fashions, the effects of collaboration and conflict, grassroots organizing, spiritual conversion, the weather, technologic capacity, and cell phones (just to name a few). Without a plot, there is no way to make sense of such a diversity of phenomena.

Plot also configures these multiple elements into a temporal order. This temporal order is of two sorts. First, each plot is composed of a discrete series of incidents, of theoretically infinite *nows.* Second, each plot takes these infinite nows, proceeding one after another in succession, and organizes them into a humanly manageable experience. As Ricoeur points out, without plot, clock time is only a series of nows. The aporia, or gap, between clock time and the experience of "human time" comes from a well-known paradox: "the future is not yet, the past is no longer, and the present does not remain." This paradox means that clock time cannot be directly experienced. Although humans have a sense of time, when they are asked to describe it, they are baffled.

Ricoeur finds a phenomenologic solution to the paradox of human time in St. Augustine's "thesis of the threefold present." Augustine argued that present, past, and future can all exist simultaneously in the human mind through a coming together of a plural, or threefold, present: the present moment accompanied by *memory* and *expectation.* Memory in this formula recalls the past, and expectation, in an analogous way, looks toward the future. The present, in other words, contains one's memory of the past and one's expectation of future events. For Ricoeur, the threefold present of time highlights a parallel between the experience of time and the structure of plot. Plot, like human time, is organized in a temporal order of beginning, middle, and end. Through plots, people configure their confused, unformed, and otherwise mute temporal experience.

Ricoeur sees this relation between time and plot as a two-way phenomenon: time becomes human time when it is organized by plot, and conversely, plot is meaningful because it portrays the features of temporal experience. Ricoeur admits that his conclusion is "undeniably circular," but he argues that this circularity is not in itself a problem: The circle of plot and time, he explains, is "not a vicious, but a healthy circle." Emplotment, from this perspective, can be seen as a kind of prosthetic device. It does not solve the riddle of time so much as it allows one to keep going without a solution. The riddle of human time, in other words, leaves humans with an inconclusive rumination to which the process of emplotment alone can respond: "Emplotment… replies to the speculative aporia [of time] with a poetic making of something capable, certainly, of clarifying the aporia …, but not of resolving it theoretically."

Character

In narrative theory, the concept of character connects directly to contemporary controversy surrounding the related and more basic concept of personal identity. The controversy around identity may be understood as a tension between essentialist and nonessentialist approaches. Essentialist notions of identity indicate that humans have a fixed personality, perhaps biologically stamped, that is, authentically belongs to one's self and that is at the core of one's being. This "true self" or "core self" may be distorted or covered over, but it is nonetheless there for the discovery if people apply themselves patiently and persistently to the task. Nonessentialist critiques, however, have deconstructed this ideal of identity and its notion of an integral, originary, and unified self. This critique has been comprehensively advanced by influential philosophers like Roland Barthes, Jacques Derrida, and Michel Foucault and by psychoanalytic writers working in the tradition of Jacques Lacan politically inflected through cultural studies and feminism. As a group, these writers have deconstructed the essentialist ideal of identity and its similar notions of an integral, originary, and unified self.

One of the most productive ways to navigate the tension between essentialist and nonessentialist understandings of identity is to draw a comparison between identity (in life) and character (in fiction). Paul Ricoeur calls this approach to identity "narrative identity," and the basic idea has recently been developed into a school of psychology known as "narrative psychology." As Ricoeur puts it, "fiction, in particular narrative fiction, is an irreducible dimension of self-understanding…fiction is only completed in life, and life can be understood only through the stories that we tell about it."

Self-understanding, on this account, is an interpretive event, and narrative is the privileged form for this interpretation: "a life story [is] a fictional history or, if one prefers, a historical fiction, interweaving the historiographic style of biographies with the novelistic style of imaginary autobiographies." Literary theorist Peter Brooks sums up this idea beautifully in his book *Reading for the Plot: Design and Intention in Narrative:*

> Our lives are ceaselessly intertwined with narrative, with the stories that we tell and hear told, those we dream or imagine or would like to tell, all of which are reworked in that story of our own interrupted monologue. We live immersed in the narrative, recounting and reassessing the meaning of our past actions, anticipating the outcome of

our future projects, and situating ourselves at the intersection of several stories not yet completed.

Narrative approaches to identity allow one to navigate the tension between essentialist and nonessentialist identities because narrative identity allows for a kind of continuity over time, and relative stability of self, without implying a substantial or essentialist core to this stability. People's interpretations of themselves use the cultural stories with which they are surrounded to tell a story of self that escapes the two poles of random change and absolute identity. In this way, a narrative identity is also an intersectional cultural identification. People's identifications may seem original, but they are narrated with the resources of history, language, and culture.

As cultural theorist Stuart Hall puts it, identity is a "process of becoming rather than being: Not 'who we are' or 'where we came from,' so much as what we might become, how we have been represented and how that bears on how we might represent ourselves." Culturally narrated identity stories, Hall argues, involve "not the so-called return to roots, but a coming-to-terms with our 'routes.'" They involve a deep appreciation of what routes have been taken, the effects of those routes, and what new routes might be chosen ahead. This coming to terms with routes, although it involves the use of inessential cultural narratives, in no way undermines the personal or political efficacy and value of identity. Nor does it undermine a genuine sense of belonging that can come from a well-narrated self.

First-Person Narratives

Bringing this theory to psychiatry, the importance of understanding the elements of narrative can be highlighted by looking at first-person memoirs of mental illness. As Hillary Clark, a scholar of mental illness memoirs puts it: "One cannot feel well one day, numb and oppressed the next, suicidal after a few months have elapsed…without seeking a narrative explanation—the cycle of bipolar disorder, for instance, or the Christian narrative of sin, repentance, and redemption—in order to make sense of it all, to trace a single self through all these changes." Clark finds that studying personal narratives are valuable because they "give voice to the ill, the traumatized, and the disabled," and they restore insights and wisdom they have achieved—wisdom that can be lost by medical research and clinical case studies that risk turning people into statistics and diagnostic labels.

Memoirs are also helpful because they show how models of mental illness work, like the narrative element of metaphor in the process of creating illness stories. Clinical models may be understood as containing root metaphors developed over time through a research and practice community. In ordinary writing or conversation, people generally use metaphors to provide a quick glimpse: “This hotel is a palace.” But, in a clinical model, research communities (e.g., biomedical, psychoanalytic, cognitive behavioral, family, social/political, art therapy, spiritual, or pastoral counseling) develop metaphors into systematic models that frame their work and that explain the world through a metaphorical redescription. The metaphor, “Joe is a pig,” redescribes Joe in a new way and allows for something new to be perceived about him. So too, the disease model of mental illness—“mental illness is a disease”—redescribes mental difference and allows something new about it to be perceived. Philosophical psychiatrists Bill Fulford and Tony Colombo use these insights to develop the following definition of psychiatric models: “[Psychiatric] models…are the conceptual frameworks, or sets of ideas, by which, in any given area, people structure and make sense of the world around them.”

Memoirs reveal that the root metaphors, which organize psychiatric clinical models, do not stay in the clinics; they seep into the culture and become cultural resources that join with the larger pool of mental models as well as the other narrative elements for organizing self-stories. In times of trouble, people look through the metaphorical structures of mental models to perceive, select, and plot aspects of their lives that they believe to be important. These culturally located self-stories and the priorities within those stories combine with other cultural resources to scaffold narrative identity and provide a compass for living. They tell where people have been and where they are now, and they provide a trajectory into the future. Studying memoirs of mental illness, a wealth of first-person stories can be found that help one to see how people plot the models and metaphors of mental difference into their narrative identity and their stories of recovery.

Lesley Dormen, in her memoir “Planet No,” emphasizes themes of loss and depression through a metaphor of unresolved grief common in psychoanalytic models. Dormen explains that she initially spiraled into depression by two losses in her life: “An architect I’d loved for under a year left me for a woman he met in a supermarket parking lot. And my best friend, Tessa, a woman as central to my understanding of myself as my brown eyes and small feet, disappeared into marriage and withdrew from our friendship completely.” Dormen used psychotherapy to explore these losses and came to realize that they were compounded by earlier losses and traumas from her childhood. She lost her father after her parents divorced when she was 6 years old, she was sexually abused by her stepfather, and she had a painful emotional relationship with her mother. Dormen felt so isolated and rejected by these experiences that when her therapist asked her, “if [you] were a girl in a story, how would [you] see yourself?” Dormen answered: “Like an orphan.” Building on this insight, she eventually came to understand that her many ungrieved losses amplified one another, and these unresolved feelings were repeated in the present life. As an adult, she relived her past experiences of solitary isolation and failure as an unmarried woman in New York, feeling the pressure to be coupled like “normal and successful” people.

Psychoanalytic therapy and self-reflection helped Dormen to understand that her stepfather’s sexual abuse was the central “curve in [her] depression story. His touch had caused the spine of the story to veer off into a new direction, had reconfigured the terms of my loves and longings.” Most pernicious for her ongoing development, the abuse left her confused about her desires, and this confusion left her “frozen in a girlhood made curiously old.” The dense and painful emotional conflicts that grew up around the sexual abuse knotted together shame, illicit sexual pleasure, heartbreaking loss, pity, betrayal, guilt, duplicity, and an aching distance between her mother and herself. These emotional knots were so tight that she could not begin to open them or untie the feelings, but at the same time she could not forget them. She found herself constantly and repeatedly returning to them “with the vague uneasiness that was a familiar part of [her] internal landscape.”

Slowly and often painfully, therapy helped her move past these knots in her character that seemed to rope her into depression and despair. She was able to separate herself from this history: “Something happened—something that was separate from me. There was a me that existed before the something and a me that existed after. Before and after were separate, not the same.” No longer tied together in knots, “it was possible to stand outside those events, to observe them and draw a conclusion about them. It was possible not to just be them." And, not only that, it became possible to move past the grief, to feel joy, and to make new connections: "I fell in love with New York, with the giddy green of the first spring after the worst of the depression lifted, with new friends, travel, and wonderful books. I looked around my apartment and saw that there was pleasure in things, in their colors and shapes, and that one did not actually need a marriage certificate to buy good pots." Dormen's memoir suggests that she came to understand her troubles through the metaphors of psychoanalysis, the plot of psychotherapeutic change, and the narrative identity of someone who used psychotherapy for recovery and transformation.

In contrast to Dormen's psychoanalytically inflected story, William Styron's memoir, *Darkness Visible,* is paradigmatic of the disease model as a narrative template. Styron describes himself as "laid low by the disease," a "major illness" of "horrible intensity," which came on him like a "brainstorm…a veritable howling tempest in the brain, which is indeed what clinical depression resembles like nothing else." This connection between depression and bad weather perfectly captures the disease model of depression as lying outside the frame of human goals, desires, losses, and disappointments. Like the weather, depression comes from the material world of inhuman forces, from physical and chemical interactions. For a strong disease model advocate like Styron, it makes no more sense to give human meaning to depression than to tell an atheist that thunder is caused by God's anger toward us. Styron explains how he sees it: "I shall never learn what 'caused' my depression, as no one will ever learn about their own…so complex are the intermingling factors of abnormal chemistry, behavior and genetics."

Styron's memoir was one of the first and most influential to adopt such a narrative frame. He relies on biologic metaphors for depression that perfectly match the disease model in psychiatry as well as an avalanche of pharmaceutical marketing devoted to promoting antidepressant medications. Depression, Styron writes, "results from an aberration of biochemical process. [It] is chemically induced amid the neurotransmitters of the brain…a depletion of the chemicals norepinephrine and serotonin." This "upheaval in the brain tissues" causes the mind "to feel aggrieved, stricken," and it creates "muddied thought processes" like the "distress of an organ in convulsion."

Styron is far from alone in his approach, and the disease metaphor runs through many other memoirs. For example, in a collection of Depression memoirs, *Unholy Ghost: Writers on Depression,* it shows up in several examples. Virginia Hefferman laments the “substandard physiology,” which creates her depression. Chase Twitchell moves past the phrase, “chemical imbalance,” to chime in with a more updated disease model explanation: “What happens in depression, for reasons that are still unknown is that the limbic-diencephalic system malfunctions.” And similarly, Russell Banks fears that his wife’s “malfunctioning limbic diencephalic systems” has created a depression so severe it may result in her suicide. All of these memoirs show the metaphors of biology shaping the plots of recovery and the narrative identity of the people involved.

In contrast to biopsychiatry or psychoanalysis, David Budbill’s memoir, “The Uses of Depression: The Way Around is Through,” provides a good example of alternative models and metaphors. He approaches painful periods of depression through the less common but still vital models of spirituality and creativity. Because of what he calls his “rebellious, contrary” personality, none of the mainstream approaches is suitable for him, and he decides to push past the mainstream to develop his “own way of dealing with and using depression” that he calls his “give in” method: “I discovered that the only way around my periods of depression is directly through them; in other words, the sooner I can resign myself to the Angel of Depression, the sooner she will be done with me and leave me alone.”

Drawing from the narrative frames of Zen Buddhism and creative writing, Budbill discovers that if he lets go of his critique of depression and allows himself to be with the intensity of the feelings, the depressive episodes are not all negative. Indeed, depression can be experienced as generative (or positive) as well. Budbill finds that one of its most important positive dimensions is that it slows him down. Depression forces him to enter a “negative space” where he can neither work nor produce; instead, he tunes into a very different experience—“empty, open, quiet, passive, receptive, dark.” These slothful, withdrawn periods function as a kind of dormancy where he can store up energy for times when the Angel of Depression lets him go. During such times, he can better hear his muse, and this “receptacle” period later emerges in his poetry. For Budbill, the metaphorical frames of spirituality and creative writing helped him cultivate a narrative identity of spiritual seeker and artist whose experiences show sensitivity rather than pathology. The metaphor of sensitivity can be seen to scaffold a recovery plot that involves tuning into his depression experience as a muse and a generator of wisdom rather than a clinical plot that works to remove or transform the experience.

Joshua Wolf Shenk gives a good sense in his memoir how a self-reflective awareness of the many models and metaphors of mental illness can be a tool of coping. Responding to his depressive experiences by researching the history of depression and the role of metaphor in shaping what is known about depression, Shenk concludes that “what we call ‘depression,’ like the mythical black bile, is a chimera…It is cobbled together of so many different parts, causes, experiences, and affects as to render the word ineffectual and perhaps even noxious to a full, true narrative.” Even so, the pain and suffering of depression is anything but a chimera, and Shenk understands the inevitable need to “abbreviate and simplify” in trying to understand it. He finds “no way around words like ‘depression’ and ‘melancholy’…but it is one thing to use shorthand while straining against the limits of language. It is quite another to mistake such brevities for the face of suffering.” For Shenk, simplified phrases like “biochemical malfunctions” and “biological brain diseases” favored by public health campaigns, well-intentioned clinicians, and pharmaceutical companies can mystify as much as they inform. “When we funnel a sea of human experience into the linguistic equivalent of a laboratory beaker, when we discuss suffering in simple terms of broken or fixed, mad and sane, depressed and ‘treated successfully,’ we choke the long streams of breath needed to tell of a life in whole.”

This awareness allowed Shenk to consciously seek an array of different approaches to his depression. He gave medications a try (although found none that were helpful) and sought solace and guidance in several hundred afternoons of psychotherapy. He worked through the pain of his parents’ divorce when he was seven—“the slow leakage of affection and kindness from my parent’s marriage, the grim entrance of resentment, confusion, and anger,” and he overcame the automatic tendency from his family life that “forbade expressing these emotions.” All of this work went into the larger task of creating a life story that put his depression in perspective: “to find my story by living it, following moments of emotional clarity through life’s maze. I look for help in therapy, [but also] in relationships, and faith in its broadest sense—faith in the gardener, the faith of the lover, the faith of the writer. The faith that I can experience what is real about the world, that I can hurt plainly, love ravenously, feel purely, and be strong enough to go on."

The story of the Icarus Project, a consumer-oriented mad pride group active from 2002 to 2020, provides another instructive example of how people can bring conscious awareness of the diversity of narrative options into the process of recovery. The Icarus Project was initiated by Sascha DuBrul and Ashley McNamara—two arts and humanities-educated college students who were given diagnoses of "Bipolar Disorder" in their 20s and who found themselves struggling as much with their psychiatric treatment as with their emotional intensities. DuBrul and McNamara, similar to many others in consumer activist groups, felt that the core of the problem with the psychiatric perspectives they encountered was an overemphasis on scientific models of "disease" and "treatment." DuBrul and McNamara did not respond against these models with an antipsychiatry perspective or dismiss psychiatric disease models outright, but they felt this way of telling the story of their difficulties missed something crucial. They agreed that their psychic differences could cause them tremendous difficulties similar to the way that medical illness can do, but they also found it a mistake to over-emphasize a model of pathology and story of broken brains.

Instead of accepting or rejecting the prevailing biologic narrative, DuBrul and McNamara set out to create a grassroots alternative community for healing and recovery that was open to the many stories of psychic differences. As McNamara explains, "We began The Icarus Project as a way of creating space for people to share their trajectories through this under-charted world of blackness and brilliance and the million shades of gray that the medical establishment has no idea how to describe." Throughout this process, Icarus Project organizing has been led by an insightful narrative organized around the story of the boy Icarus:

> Defining ourselves outside convention we see our condition as a *dangerous gift* to be cultivated and taken care of rather than [simply] a disease or disorder needing to be 'cured.' With this double-edged blessing we have the ability to fly to places of great vision and creativity, but like the boy Icarus, we also have the potential to fly dangerously close to the sun—into realms of delusion and psychosis—and crash in a blaze of fire and confusion. At our heights we may find ourselves capable of creating music, art, words, and inventions which touch people's souls and change the course of history.

DuBrul and McNamara go on to acknowledge the risks in using the Icarus story and the possible perils of narrating their emotional difficulties as dangerous gifts. However, in spite of these risks, they also recognize how metaphors of mental and biologic pathology and metaphors of creative, political, and spiritual sensitivity can be as close as a turn of the kaleidoscope. They recognize how stories centered on pathology can turn into stories of creativity and that this narrative flexibility can create “tools of inspiration and hope in this repressed and damaged society.” As they put it, “We understand that we are members of a group that has been misunderstood and persecuted throughout history, but has also been responsible for some of its most brilliant creations. And we are proud.”

The Icarus Project is similar to Shenk in that it consciously creates integrated narratives of psychic troubles that are open to pathologic metaphors as well as creative, spiritual, and political metaphors. The difference is that Shenk is working on his story alone, and the Icarus project uses a mutual aid community. The goals for both, as the Icarus project puts it, are to recognize the dangers and pitfalls of mental difficulties, to celebrate and tap into the “potential that lies between brilliance and madness,” to put in the necessary effort to stay balanced and grounded, and to plot recovery stories that lead to full and meaningful lives (which might include stories of psychiatric medications, psychotherapy, artistic practice, activism, or spirituality). By foregrounding the narrative process of recovery, the Icarus Project aims to separate mental health issues “from the reductionist framework offered by the current mental health establishment” and in doing so, learn “from each others’ mistakes and victories, stories and art, and create a new culture and language that resonates with our actual experiences of ‘disorder.’”

NARRATIVE PSYCHIATRY

The deep implication of narrative theory and memoirs of mental illness for psychiatric practice is that there are many ways to tell the story of mental difference and suffering—not just one right way and many other wrong ways. To be helpful for people with the tremendous diversity of values and life-style preferences seen in memoirs, psychiatrists cannot be limited to one or two storytelling options. Psychiatrists must be comfortable using a range of metaphors and models to help people plot their recovery stories and narrative identities. To see how this works in practice, it is helpful to develop a character from literary fiction as a “case example” for a narrative psychiatry approach.

Narrative psychiatrists find literary fiction an invaluable resource for "case examples" because literary examples, as a rule, provide much richer contextual data than clinical case studies. All too often clinical case studies tell a story that has been tightly organized along a particular narrative frame. Literary examples do this as well, but the framing is generally much less heavy handed and much more open to alternative frames and perspectives. That said, literary examples usually require additional imaginative work to make them suitable as clinical case histories.

To see how this works, consider Marina, a character in Claire Messud's 2007 novel *Emperor's Children.* Marina is a single woman, in her early 30s, lives with her parents, and is unemployed. She struggles throughout the novel with symptoms of depression and anxiety. To get to know Marina's story, start with a scene where she and her college friend Danielle are meeting for dinner. The dinner is at Marina's house, but Danielle arrives first, which leaves her spending several minutes alone with Marina's parents. Being alone with Marina's parents makes Danielle feel awkward and uncomfortable. Here's the passage:

> Danielle…felt like a teenager, as she used to feel in the kitchen of her parents' house in Columbus, before the divorce of course, and she was suddenly, powerfully, aware of the profound oddity of Marina's present life. A life arrested at, or at least returned to, childhood. Danielle couldn't imagine eating nightly with her parents, not only because they now lived in different states and didn't speak to each other, but because she was entering the fourth decade of life and hadn't been through the wearying rigmarole of family life for anything more than an almost supportable day since she was seventeen and had gone off to college.

In this passage, Danielle uses the metaphor of childhood to organize her understanding of Marina's life. Through the childhood metaphor, Danielle perceives Marina's adult life as "arrested, or at least returned to, childhood." If Danielle were to tell the story of Marina's life using this childhood metaphor, she might very well select events from Marina's past that led up to this moment. Perhaps she would recall a difficulty Marina had in committing to relationships, or a series of immature reactions to authority figures, or perhaps several examples of her being dreamy and irresponsible. These diverse elements would come together in a plot that leads to the present moment of Marina as trapped in childhood. The story would not only make sense of the past and the present and provide a narrative identity for Marina, but it would also provide clear direction for the future. In this narrative frame, Marina must “grow up.” She must leave her parents and home, get a job, get married, and have children (or at least make steps in these directions).

This example highlights the importance of “point of view.” Narratives tell a story from a particular perspective. Even if an author uses the omniscient point of view, that is still not a story from nowhere; it remains a story told by an author who has a particular take on the events that they select and frames. In the selection above from *Emperor’s Children,* readers learn about Marina’s life from Danielle’s point of view. This passage does not tell the reader whether Marina would see herself as stuck in childhood; it only tells the reader that Danielle sees her this way.

A little further on in the story, readers learn that Marina does feel similarly about herself. Indeed, the perspective of herself as childish and immature has left her deeply unhappy and anxious about her life and even about her worth as a person. Not only this, but as the novel proceeds, readers see Marina’s deep discontent with her “childish” self-perception leading her into a poorly thought-out romance and decision to marry an older man, a magazine editor, after they had been going out for a few weeks. In the process of this marriage, Marina gains in quick succession not only a husband but also a job (as she joins the editorial staff of her new husband’s magazine). This rapid marriage and entrance into the job market effectively disrupt the childhood metaphor (Marina has now moved into “adulthood”), but, unfortunately, Marina’s judgments turn out to have been potentially premature, as the magazine quickly fails and the relationship falters. By the end of the novel, it is not known whether Marina’s marriage will work, but Messud leaves her readers feeling uncertain and doubtful.

To use this story as a literary “case example,” the additional narrative work involves imagining Marina coming to see a psychiatrist at or near a moment of crisis. A common plot structure in narrative fiction involves what theorists call the Freitag triangle. In this plot structure, the story and its characters move from increasing tension and conflict to climactic crisis and then finally to resolution. If it is a novel or a complicated short story, the plot is often composed of multiple Freitag triangles that may or may not come together into an overarching Freitag triangle. In using literary material as case examples, it is helpful to imagine that a character in the story comes to see an imaginary clinician just at or near this point of climactic crisis. For example, imagine that Marina came to see a narrative psychiatrist for help when she was at the height of her crisis. This would be when she was in the middle of her most intense anxiety and feelings of inadequacy, and before these feelings were resolved (at least temporarily) by marrying the magazine editor and getting a job.

How would the narrative psychiatrist work with Marina in this situation? Like with narrative medicine and other person-centered approaches, the narrative psychiatrist's first and in many ways foremost task in helping Marina will be in getting to know her and developing a rich appreciation of her story. This empathic connection helps Marina become much less alone with her problems. She now has a nonjudgmental partner with whom to share them and to relate them. Psychiatrist Robert Coles captured this deeply person-centered dimension of narrative psychiatry in his book *The Mind's Fate: A Psychiatrists Looks at His Profession.* Coles sees the person-centered core of narrative psychiatry as the "*ultimate mission* of those who are doctors of the mind." To be a psychiatrist is to be an "attentive listener first" and to be able to make sense of the "inside cry for a knowing assistance that is at the heart of a person's decision to visit a physician." This kind of attentive listening remains central to psychiatry regardless of evolving science and research into mind and brain because:

> Human beings are creatures of awareness, of language, of long and articulated memory—and also, creatures whose special skills and triumphs live side by side with a special capacity for tragedy. Losses, vulnerabilities, weak spots and soft spots, one's ultimate destiny of disappearance from this planet are known to humans, but they are not to all other forms of life on this earth. Through words, whether spoken to others or silently put to one's self, humans constantly struggle to make sense of things, even as they try (if they can) to prepare themselves for the worst: hunker or huddle down, reach out, whistle in the dark, while trembling mightily within themselves. Not rarely, at such moments, people have learned in this century to seek the counsel of psychiatrists, men and women presumed to understand one's many possible predicaments, one's wild or loony sides, one's lapses into incoherence or rage, into lusts and times of darkness that threaten one's reliability, one's stability, one's very own life (or the lives' of others).

For Coles, psychiatrists must understand that to be sought out at these moments, to be asked to "attend such a cry" is an honor, a privilege, and the "ultimate bedrock" of psychiatry's professional competence.

This empathic core, however, will only be part of the interaction between Marina and the narrative psychiatrist. In addition, Marina will likely wish that the psychiatrist recommend interventions and treatment options to help her move beyond her current crisis and conflict. Narrative psychiatrists are closely guided in this process of care planning by the bioethical principles of *respect for autonomy* and *informed consent.* Bioethicists Tom Beauchamp and James Childress explain that respect for autonomy acknowledges the moral right of every individual to choose and follow their own plan of life and actions. Informed consent is the practical application of the respect for autonomy. Following Beauchamp and Childress, narrative psychiatrists agree with Johnson et al.'s description of informed consent in their *Clinical Ethics:*

> Informed consent constitutes a central feature of an encounter between physicians and patients, characterized by mutual participation, good communication, mutual respect, and shared decision making. Informed consent requires a dialogue between physician and patient leading to agreement about the course of medical care. After initial consent to treatment has occurred, and ongoing dialogue between patient and physician, concerning the patient's continuing medical needs, reinforces the original consent. A properly negotiated informed consent benefits both the physician and the patient—a therapeutic alliance is forged in which the physician's work is facilitated because the patient has realistic expectations about results of the treatment, is prepared for possible complications, and is more likely to be a willing collaborator in the treatment.

Following this approach to informed consent, the narrative psychiatrist initiates a collaborative process of selecting strategies for change by giving a background on the array of treatment options available and explaining the benefits and risks of these options. Narrative psychiatry is not unique in this way, however narrative psychiatry does share bioethicists concern that despite a vast literature in bioethics that physicians in general and psychiatrists in particular are often deficient in observing the practice and spirit of informed consent. Narrative psychiatry, by contrast, puts this form of ethical alliance building and collaborative planning at the center of care. Narrative psychiatrist SuEllen Hamkins describes this stage of treatment as exploring with the patient the "doctor's tool kit of psychiatry." As she puts it, "when we have a collaborative therapeutic stance, we can look side by side without our patients at the wealth of treatment options that might be helpful and weigh the pros and cons together."

In the case of Marina, because her problems are relatively mild, the narrative psychiatrist would suggest psychotherapy. However, which psychotherapy should Marina choose? The standards of care consist of a tremendous variety of options. The collection *Current Psychotherapies* features 14 different approaches, and *Theories of Counseling and Psychotherapy* focuses on 20 different options. Both of these texts cover psychoanalysis, cognitive therapies, and humanistic therapies, and the latter includes additional options of feminist therapies, narrative psychotherapy, expressive arts therapies, transcultural therapies, and neuropsychotherapy. But there are still many other possibilities. Neither text mentions alternative approaches to psychic troubles, such as pastoral counseling, acupuncture, yoga, peer support, and mad pride activism. Although these last options are not psychotherapy by definition, they do represent viable approaches to psychic difficulties that Marina might find helpful. Taking this tremendous variability of psychotherapies and other treatment options into account leaves the narrative psychiatrist with two problems. How to make sense of this diversity, and how to present these options to Marina in a way consistent with autonomy and informed consent?

Fortunately, one of the most helpful aspects of narrative theory for psychiatry is that it provides an overarching, or metatheoretical, rationale for understanding how these many psychotherapies work. From a narrative perspective, all therapies involve a process of storytelling and story retelling. No matter which style of psychotherapy one uses, the process of therapy involves an initial presentation of problems that the client is unable to resolve. The client and therapist work together to bring additional perspectives to these problems, allowing the client to understand them in a new way. These additional perspectives vary greatly depending on which style and therapeutic point of view are used. It matters whether the therapy is psychodynamic, cognitive, humanistic, feminist, spiritual, or expressive. But from the vantage point of narrative theory, what these different approaches all have in common is that they rework, or “re-author,” the patient’s initial story into a new story. This new story allows new degrees of flexibility for understanding the past and provides new strategies for moving into the future.

Having a metatheoretical narrative understanding of the variety of psychotherapy options helps make sense of perhaps the most important empirical psychiatric finding to date—an array of alternative approaches to psychotherapy are all *equally effective.* This finding is known lovingly in the therapeutic research community as the “Dodo effect.” The appellation comes from a line in *Alice in Wonderland:* “At last, Dodo said, everyone has won, and all must have prizes.” Although this finding is controversial and has been only partially integrated into clinical practice, empirical studies confirm with remarkable consistency that the positive effects of therapy (of which there are many) are not due to the specific interventions of the therapist. The benefits of therapy come instead from “common factors” of the therapeutic setting. These studies suggest that the process of setting up a strong therapeutic alliance with a quality therapist is much more important than the content of the specific models and theories from which the therapist works.

The Dodo effect suggests that a variety of theoretical models can be used to understand, cope with, and ameliorate painful emotional states. The Dodo effect does not mean that any old interpretation will do. Therapy does not work if the therapist says things completely out of context, such as “your sadness is because the cow jumped over the moon.” On the contrary, for psychotherapy to be effective and for the alliance to be strong, the therapist and the client must have a sense of belief and confidence in the interpretive frame being used. In the words of Saul Rosenzweig, who first coined the term Dodo effect, “Whether the therapist talks in terms of psychoanalysis or Christian Science is…relatively unimportant as compared with the formal consistency with which the doctrine employed is adhered to, for by virtue of this consistency the patient receives a schema for achieving some sort and degree of personality organization.”

When the narrative psychiatrist applies these insights to Marina’s situation, it suggests that an array of psychotherapy approaches might be helpful for Marina. Any number of different interpretations of Marina’s situation could be developed and worked through in a therapeutic relationship with a well-meaning quality practitioner. If that were to happen, there is a very good chance that an array of different insights and narrative structures could effectively help Marina understand her anxiety and feelings of inadequacy. If the insights and narrative frames were not imposed on Marina but were ones in which she participated in creating and to whose meaning she felt strongly connected, the therapy would “work” by encouraging modifications in Marina’s assumptive worlds. These modifications would transform the meanings of her painful feelings to more favorable ones and free her up for alternative forms of coping.

A narrative psychiatrist working with Marina would have all of this in the back of their mind as they began the process of making psychotherapeutic recommendations. Following the basic ethical principles of autonomy and informed consent, even in the fine-grained choices between types of psychotherapy options, the psychiatrist would give Marina background on the array of psychotherapies so that Marina could make informed choices about which therapy to choose. Most narrative psychiatrists would explain to Marina that they practice a narrative approach that collaboratively selects options and resources from a variety of treatment approaches as they seem appropriate for the therapeutic process. This means that narrative psychiatry tends to use an integrative form of psychotherapy that works through collaborate eclecticism. The strategic and collaborative frame for selecting from a range of different therapeutic approaches is guided by the narrative goal of creating preferred identities of meaning and thriving. Eclectic openness is consistent with narrative psychiatry because it appreciates that not only are a variety of therapeutic models valuable for people, but also a variety of therapeutic insights and strategies can be woven together to form a hybrid narrative of a person's past, present, and future that can be helpful. These insights may come from psychoanalysis or cognitive–behavioral, feminist, or creative arts perspectives, to name a few. Eclectic therapists are not constrained by loyalty to one point of view over another and, therefore, can borrow and mix from a variety of schools.

If this form of therapy did not sound helpful to Marina, or if she had preferences for a more specific form of treatment beyond what the narrative psychiatrist felt comfortable with, then the psychiatrist would work with her to find a suitable referral. If there was a match between what Marina would like to try and the narrative psychiatrist's collaborative, eclectic approach, then they could continue working together. It is important to point out that although there are overlaps, a narrative approach to the many different psychotherapies is different from a specific narrative psychotherapy approach. A "branded" narrative psychotherapy approach would not have particular pride of place in the therapeutic process. Narrative psychotherapy as a particular form of psychotherapy, in other words, would be one approach among others, which Marina and the narrative psychiatrist might eclectically choose from in developing her preferred recovery stories.

However, for the sake of this overview of narrative psychiatry, it is useful to consider what narrative psychotherapy approaches might look like for Marina. Narrative psychotherapy is a form of psychotherapy that is in some ways similar to psychodynamic or insight-oriented psychotherapy and in other ways similar to humanistic and existential psychotherapy. The key idea is that clients come to therapy with what narrative psychotherapists Michael White and David Epston call "problem-saturated stories.” The process of therapy deconstructs this presenting story by carefully and lovingly pulling it apart. In this example, Marina and the narrative psychiatrist would look together for gaps and breaks in her story, for inconsistencies, for left-out information, for overly emphasized events, and for metaphors and linguistic distinctions that ground the assumptions of the story. Finding these kinds of gaps and inessential authorial choices would help to establish that Marina’s presenting story is not an absolute necessity and that other stories could be told. Marina and the therapist would then work together to find a new story that would help Marina reauthor her past and present and expand her future options in desirable ways.

This process is similar to the process of other psychotherapy approaches, except that it would be less likely than most to rely on therapeutic jargon and pretold story frames—such as are more common in psychoanalytic therapy, cognitive therapy, interpersonal therapy, or feminist therapy. Rather than using therapeutic jargon, narrative psychotherapy would try to stay with the language of everyday meaning and folk psychology. In addition, narrative psychotherapy, more than most other therapy options, tends to be very conscious of cultural power dynamics and the way in which people’s problems often have to do with internalized hierarchies that they have adopted from mainstream culture.

In Marina’s case, the metaphor of “childhood” provides a good example of the pain cultural power dynamics can play. Marina lives in an elite and still deeply patriarchal culture, where “autonomy” and “independence” are prioritized over “connection” and “interdependence.” These cultural norms are loaded into the meaning of the “childhood” metaphor. If Marina lived in a different culture or if she thought through these values preferences in the context of therapy, she might feel very differently about her current life and what it means to live at home. In narrative therapy, Marina and the therapist would look for aspects of Marina’s life—what narrative therapists call “unique outcomes”—that go against the grain of the childhood metaphor. They would look at the many ways in which Marina’s life is very different from the life of a child and they would deconstruct the wisdom of a cultural insistence on norms of autonomy and independence. Are these value preferences absolute, do they necessarily lead to a contented life, and are there examples in which people have followed these goals and ended up discontented and demoralized? Indeed, as it turns out in the novel, Marina’s friend Danielle—the very one who could not imagine being in Marina’s situation—is far from happy in following a path of relentless independence and “maturity.” By working through this kind of narrative reauthoring, Marina could use narrative psychotherapy to reach a new level of peace with her current life and a new degree of flexibility as she makes plans for the future.

In Marina’s “case,” as it has been imagined thus far, the main treatment strategy has been psychotherapy. But it is important to point out that the process of using narrative as a guiding frame for treatment planning also includes more intensive psychiatric treatments such as medications, hospitalizations, and involuntary commitments. To enter into these additional aspects of narrative psychiatry, make a further imaginative adjustment to Marina’s case history. Imagine that Marina did not enter therapy in the early stages of her crisis but that she continued to suffer alone with her problems. Imagine that without a resolution for her crisis, her sense of inadequacy worsened to the point at which Marina became not only anxious but also deeply sad and tearful, with intense feelings of worthlessness and a profound sense of hopelessness that she could ever move forward with her life. Imagine that she loses interest in the things that she had previously enjoyed and starts spending most of her time in bed watching television. These feelings go on for weeks, and it gets to the point at which Marina is unable to sleep, her appetite becomes erratic, and she even begins to fantasize about suicide.

If Marina were to see the narrative psychiatrist at this moment of worsened crisis, she would now have a presenting problem that fits the *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition (DSM-5-TR), diagnosis of Major Depression, severe without psychosis. With this diagnosis, Marina would be considered by many to have a “clinical depression,” which the standard of care narrates as a kind of “disease,” not unlike diabetes or high blood pressure. How would the narrative psychiatrist respond in this situation? Would it differ significantly from the previous situation? The answer is yes and no. As before, the first task for the narrative psychiatrist would be to get to know Marina’s story in all of its unique complexity.

The main difference in this new situation would be the added complication of Marina’s suicidal ideas and the need to make an assessment of her danger to herself. If Marina met the standard of care criteria for hospitalization, then the narrative psychiatrist would collaboratively work toward that goal. If the standard of care for hospitalization was met, but the patient refused, and collaborative treatment planning broke down, the narrative psychiatrist would work toward involuntary commitment. This would, unfortunately, interrupt the collaborative process of meaning-making as the narrative psychiatrist would have to act unilaterally. But the larger narrative process of the treatment would not stop. During the hospitalization and afterward, Marina and the psychiatrist would continue to process the event. The eventual narrative meaning of the involuntary hospitalization would be open to ongoing storytelling. It may turn out that Marina comes to agree with that "danger to self" plotting of the story or it may turn out that the psychiatrist comes to restory the involuntary hospitalization as an overreaction and even a mistake.

If hospital care was not an issue, the outpatient process of assessment and treatment planning would be similar to before, except for the question of medication. After getting to know Marina, or even while this process was proceeding, the narrative psychiatrist would explain that in view of the severity of her depressive symptoms, many, if not most, psychiatrists would recommend medication treatment along with psychotherapy. The narrative psychiatrist would explain the risks and benefits of the medication options along with a basic review of the outcome studies for the medicine's effectiveness.

If Marina opted for medication and it was indicated in the standard of care, the narrative psychiatrist would be comfortable with this choice. Narrative psychiatry, in other words, does not change its basic storytelling approach when it comes to biologic explanations, disease models, or psychopharmacologic interventions for mental states. Narrative psychiatry does not resemble antipsychiatry approaches that understand disease models of mental illness as a myth. However, narrative psychiatry also does not essentialize biologic or disease models of explanation and treatment as somehow more real and substantial than other models. For the narrative psychiatrist, biologic models are not an exception to a general understanding of the ways in which metaphor and plot work to shape narrative identity because narrative psychiatry conceptualizes biologic models not as myths or mirrors of truth but as metaphors. Narrative psychiatry stands on the shoulders of work in the philosophy of science and science studies to draw out the similarities between scientific models and metaphors more broadly. It recognizes that scientific models also create a kind of "seeing as." First-person narratives of mental illness show that when people inhabit scientific frames, they come to experience "seeing as" in a deeply naturalized way, to the point at which "seeing as" becomes "being as," and people experience themselves as having some form of "broken brain." However, this process is not different from other uses of metaphor to structure experience—it is just another example of the role of metaphor and linguistic structures in the formation of narrative identifications.

Biologic models do pose a problem for informed consent. However, because of the degree, the biologic point of view is overhyped in the media and in the medical literature. Sociologist Nicolas Rose uses the term "neurochemical self," and anthropologist Emily Martin uses the phrase "pharmaceutical person" to describe the dramatic rise in neurochemically mediated self-identity. To become a neurochemical self, one does not have to become a fundamentally different kind of human. One only has to internalize the metaphorical structures of neuroscience and the practices of psychopharmacology and use these metaphors and practices to organize one's sense of self. Consider the example of Gary, a character from Jonathan Franzen's book *The Corrections:*

> Although in general Gary applauded the modern trend toward individual self-management of retirement funds and long-distance calling plans and private-schooling options, he was less than thrilled to be given responsibility for his own personal brain chemistry, especially when certain people in his life, notably his father, refused to take any such responsibility. But Gary was nothing if not conscientious. As he entered the darkroom, he estimated that his levels of Neurofactor 3 (i.e., serotonin: A very, very important factor) were posting seven-day or even thirty-day highs, that his Factor 2 and Factor 7 levels were likewise outperforming expectation, and that his Factor 1 had rebounded from an early-morning slump related to the glass of Armagnac he'd drunk at bedtime. He had a spring in his step, and agreeable awareness of his above-average height and his late summer suntan. His resentment of his wife, Caroline, was moderate and well contained. Declines led key advances in key indices of paranoia (e.g., his persistent suspicion that Caroline and his two older sons were mocking him), and his seasonally adjusted assessment of life's futility and brevity was consistent with the overall robustness of his mental economy. He was not the least clinically depressed.

Whatever else one might want to say about Gary, it is clear that the metaphors and plots of neuroscience have become a core part of his self-understanding. For another example, this time from nonfiction, consider this quote from neurophilosopher Patricia Churchland. The context is that she is describing her feelings to Paul Churchland (her neurophilosopher husband) after coming home from a particularly frustrating faculty meeting:

> Paul, don't speak to me, my serotonin levels have hit bottom, my brain is awash in glucocorticoids, my blood vessels are full of adrenaline, and if it weren't for my endogenous opiates, I'd have driven the car into a tree on the way home. My dopamine levels need lifting. Pour me a Chardonnay, and I'll be down in a minute.

Neurochemical selves, as these examples make clear, are selves like other selves. The only difference is that they have different metaphors, plots, and narrative identifications than nonneurochemical selves. Narrative psychiatrists have an ambivalent relationship with this rapidly emerging phenomenon. On one hand, narrative psychiatrists are open to this form of selfhood and they respect it as much as they would other forms of self. William Styron's memoir and the many similar ones show that people are clearly able to integrate the biologic model and neurochemical self-identity into their lives in a helpful way. On the other hand, narrative psychiatrists are also aware of the deeply problematic power dynamics of this emerging form of self and that there are many people who prefer to avoid biologic stories as much as possible.

Both Nicolas Rose and Emily Martin point to contemporary pharmaceutical practices as a major determinant of this new form of self-making. Since 1980, the pharmaceuticals have transformed themselves from research-driven companies to companies that operate more like Procter & Gamble, the maker of Tide. The pharmaceutical business plan of expanding markets through chronic, life style, conditions (such as arthritis, asthma, diabetes, mental illness) means that the industry has moved away from a focus on product per se and toward the creation of climate of opinion and desire surrounding the product.

Jan Leschley, the former CEO of Squibb pharmaceuticals, explains: "We realized we are an information company more than we are a pill company." For Leschley, the modern transnational drug company does not sell its products, the pills, as much as the information that surrounds their pills. Today's pharmaceutical industry has become an information industry, an industry of cultural shaping, more than a pill industry, and this cultural shaping very consciously includes the internal culture of psychiatric research, education, and practice. These powerful marketing strategies and tremendous promotion budgets that go with them means that pharmaceutically inspired neurochemical metaphors for self-making can often drown out other options.

All of this becomes grist for the mill, as psychoanalysts like to say, in the clinical setting. Narrative psychiatrists navigate their complicated relationship to pharmaceutical self-making by sharing with their clients their ambivalence during their discussion of these options. They let their clients know that there is considerable hype around these kinds of interventions, and they make a point to discuss the reality of side effect risks along with the ambiguity of most outcome studies for these medications. However, they also make the point that, for any one person, there is no way to know—except through a treatment trial—how much they might benefit or risk from a medication. Should the patient wish to undertake such a trial in the context of informed decision making, the narrative psychiatrist is happy to offer whatever assistance is possible. For narrative psychiatrists, these kinds of psychopharmacologic frames and strategies become part of an array of possible tools that can be woven together into an eclectic narrative of treatment, and the baggage of "big pharma" promotion dynamics does not necessarily rule out their usefulness for a particular person in a particular situation.

FUTURE DIRECTIONS

In summary, the time is right for a narrative turn in psychiatry. Recent work in narrative medicine, narrative psychology and psychotherapy, and narrative theory has opened the door for the development of narrative psychiatry. Narrative is particularly valuable for psychiatry at this moment in history because it helps self-correct from an overreliance on biologic models without falling into a morass of antipsychiatry and without throwing away the bioscience baby with the bathwater. Plus, the narrative ferment in the intellectual life surrounding psychiatry—the arts, humanities, social science, medicine, and psychotherapy—means there has never been better resources available for cultivating narrative psychiatry. Finally, narrative psychiatry is valuable for helping psychiatry tell its own story because key identity questions for psychiatry as a field and for individual psychiatrists are narrative questions. Everyone involved in psychiatry must ask themselves: where have they been, where are they now, and where will they go in the future? All of these questions are about narrative identity. For the field, these questions are about how to prioritize future research, education, and practice models. For individual psychiatrists, these questions are about determining who they want to be and how they want to practice. These kinds of narrative identity questions for the field and for individual psychiatrists are storytelling questions; they cannot be answered by science alone without the help of narrative meaning making.

None of this means the end to neuroscience, the biologic model, psychiatric medications, or emergent neuroplastic training practices such as mindfulness-based stress reduction. Just the opposite. The biologic model is now, and will remain, a critical tool for understanding and intervening in psychiatric troubles. This will only become truer as biology and neuroscience moves into a brave new world that maps out physical markers for mental states and mental processes. Taking a narrative turn in psychiatry does not mean abandoning the biologic model, or any other model. A narrative turn says all models and metaphors are valuable if they are used self-reflexively in dialogue with patients and in the service of recovery. Ultimately, it is the patient's choice which models and metaphors fit best with their goals and desires. The task of psychiatrists is to help navigate the process and the options. This kind of narrative self-reflexivity helps psychiatrists remember that their allegiance is to their patients—not their models and metaphors. Recovery is the goal, not loyalty to particular narrative templates.

Narrative theory also does not take away from empirical research. Indeed, empirical studies of neural mapping and neuroplasticity are rapidly articulating a neurologic basis for narrative theory itself. The work of neuroscientists and humanists Nancy Easterlin, Paul Armstrong, Michael Gazzaniga, Joseph LeDoux, Antonio Damasio, V.S. Ramachandran, Jeffery Gray, and Gerald Edelman support the emerging consensus in neuroscience that self-understanding takes a narrative form. Philosopher Leslie Paul Theile summarizes the biologic research that understands humans as storytelling animals this way: "Neural mapping is best understood as a narrative accounting of lived experience. It constitutes a silent, synaptic story. This narrative accounting is an "obsession" for humans in the sense that, phylogenetically and ontogenetically, it constitutes our being. From the evolutionary perspective of the species and the developmental perspective of the individual organism, brain mapping makes us who we are. Synaptic storylines inventory the individual's life, serve as its engine of development, and foster a sense of self." Or, as literary theorist Paul Armstrong puts it: "The ability to tell and follow stories requires cognitive capacities that are basic to the neurobiology of mental functioning."

In other words, at the neurobiologic level, everyone can be understood as fabricated characters. The brain constructs a model of the external world though synaptic maps and it constructs a model of the self as actor in the world. This emergent convergence of neuroscience and narrative theory can be invaluable in helping psychiatry move beyond deeply problematic mind/brain and science/humanities binaries. After a century of psychiatric swings between mind-centered and brain-centered versions of itself, narrative provides a biocultural third-space opportunity for psychiatry to be true to both its biologic dimensions and its human meaning-making dimensions.

Empirical research of recovery is also adding to an understanding of narrative processes in psychiatry. In one such study, Damien Ridge and Sue Ziebland interviewed 38 men and women who had recovered from depression to discover the strategies they used to revitalize their lives following depression. Ridge and Ziebland found that the people they interviewed did not use any single approach. Instead, they selected a variety of "narrative tools," such as talking therapy, medication, yoga, and complimentary therapies to regain mastery of their situation. Ridge explains: a "key finding…was that recovery tools considered effective by patients go hand in hand with telling a good story about recovering from depression." The very process of assembling a range of tools for recovery helped create the stories that were effective. The people interviewed found that therapists were most helpful if they could function not as experts but as "recovery allies." The most helpful health professionals were those who were empathic, good listeners, provided hope and encouragement, and who had a positive belief in the person going through the depression. A recovery ally of this type does not function in a "doctor knows best" capacity, but is someone who helps the person select the narratives and the recovery tools that feel most true to them.

Thus, when psychiatrists take a narrative turn, they do not throw out their other skills, knowledge, or research methods. The shift to narrative is, as much as anything else, an attitude shift and an opening out to additional sources of information. It starts by bringing to the foreground that the clinical encounter is a human encounter, and it follows by opening out to colleagues in the humanities, interpretive social sciences, and the arts to help to better understand this human encounter. In addition, narrative psychiatry joins with other contemporary efforts in psychiatry—such as in bioethics and the recovery movement—to make clinical encounters much more client focused and collaborative. Narrative psychiatry, at its core, recognizes that there are many ways to tell the story of one's life. The choice among these different options is a key way in which people create their identity. These choices should not be reduced to expert choices or scientific choices because they are always also personal and ethical choices. In the end, they are choices about what kind of life one wants to live.

For psychiatric practitioners, taking a narrative turn means joining with colleagues in narrative medicine and narrative psychotherapy to develop their own narrative competency. Narrative psychiatrists must understand the basics of narrative theory, and they must learn to read widely in a range of different contexts. They must come to appreciate the many stories of biopsychiatry, psychoanalysis, cognitive therapy, interpersonal therapy, family therapy, humanistic approaches, cross-cultural approaches, feminist approaches, mad pride and disability activist approaches, postmodern approaches, and spiritual approaches, just to name a few. Furthermore, they must come to understand the value of biography, memoir, and fiction for developing a repertoire of narrative frames and options. In the end, narrative competency in psychiatry means a tremendous familiarity with the many possible stories of psychic pain and psychic difference. The more stories clinicians know, the more likely they are to help their clients to find a narrative frame that works for them.

For clients and potential service users, a narrative understanding means that there is a range of possible therapists and healing solutions that might be helpful. An approach that is right for one person may not be right for another. There must be a fit between the person and the approach, and people should feel empowered to take seriously their intuitions and feelings. If the person getting help does not feel this fit, they are likely right. There may well be another approach that would work better with the person's proclivities. Like everything else, however, judgment is critical. Therapeutic experiences of all kinds can be frustrating, slow, and uncertain. How, for example, does one know when an approach misses their need and when it is something that will take time, patience, and perseverance to be helpful? From a narrative perspective, there can be no gold standard or simple answers. Only judgment, wisdom, and trial and error can decide.

FURTHER READINGS

Angell M. *The Truth about Drug Companies: How They Deceive Us and What to Do about It*. Random House; 2005.

Armstrong P. *Stories and the Brain: The Neuroscience of Narrative*. John's Hopkins University Press; 2020.

Budbill D. The uses of depression: the way around is through. In: Berlin R, ed. *Poets on Prozac: Mental Illness, Treatment and The Creative Process*. Johns Hopkins University Press; 2008.

Casey N. *Unholy Ghost: Writers on Depression*. Harper Perennial; 2002.

Cassel E. The nature of suffering and the goals of medicine. *N Engl J Med*. 1982;306(11):639–645.

Charon R, DasGupta S, Herman N, et al. *The Principles and Practice of Narrative Medicine*. Oxford University Press; 2016.

Clark H. *Narratives and Depression: Telling the Dark*. State University of New York; 2008.

Coles R. *The Mind's Fate: A Psychiatrist Looks at his Profession*. Little, Brown, and Co; 1995.

Dormen L. Planet No. In: Casey N, ed. *Unholy Ghost: Writers on Depression*. Harper Perennial; 2002:229–242.

DuBrul S. The Icarus Project: a counter narrative for psychic diversity. *J Med Humanit*. 2014;35:257–271.

Easterlin N. Introduction: Cognitive literary studies and the well-lived life. *Poetics Today*. 2019;3:377–394.

Engel G. The need for a new medical model: a challenge for biomedicine. *Science*. 1977;196(4286):129–136.

Fulford B, Colombo T. Six models of mental disorder: a study combining linguistic-analytic and empirical methods. *Philos Psychiatr Psychol*. 2004;11(2):129–144.

Hall S. Introduction: who needs identity. In: Hall S, du Gay P, eds. *Questions of Cultural Identity*. Sage; 1996.

Hamkins S. *The Art of Narrative Psychiatry*. Oxford University Press; 2013.

Hesse M. The explanatory function of metaphor. In: McErlean J, ed. *Philosophies of Science: From Foundations to Contemporary Issues*. Wadsworth; 2000:349–355.

Jones-Smith E. *Theories of Counseling Psychology and Psychotherapy: An Integrative Approach.* 2nd ed. Sage; 2016.

LeFrançois BA, Menzies R, Reaume G. *Mad Matters: A Critical Reader in Canadian Mad Studies*. Canadian Scholars Press; 2013.

Lewis B. *Narrative Psychiatry: How Stories can Shape Clinical Practice*. Johns Hopkins Press; 2011.

Lewis B. A deep ethics for mental difference and disability: the 'case' of Vincent van Gogh. *Medical Humanities*. 2017;43:172–176.

Martin E. The pharmaceutical person. *BioSocieties*. 2006;1:273–287.

Messud C. *The Emperor's Children*. Vintage Books; 2006.

Ricoeur P. *The Rule of Metaphor: Multi-Disciplinary Studies in the Creation of Meaning in Language*. University of Toronto Press; 1977.

Ricoeur P. *Time and Narrative. Vol 1*. University of Chicago Press; 1984.

Ridge D, Zeibland S. "The old me could never have done that": how people give meaning to recovery following depression. *Qual Health Res*. 2006;16(8):1038–1053.

Rose N, Abi-Rahced JM. *Neuro: The New Brain Science and the Management of the Mind*. Princeton University Press; 2013.

Rosenzweig S. Some implicit common factors in diverse methods of psychotherapy: “At last the Dodo said, ‘Everybody has won and all must have prizes.’ ” *Am J Orthopsychiatry*. 1936;6:412–415.

Sismondo S, Greene J. *The Pharmaceutical Studies Reader*. Wiley Blackwell; 2015.

Wampold BE, Imel Z. *The Great Psychotherapy Debate: The Evidence for What Makes Psychotherapy Work*. 2nd ed. Routledge; 2015.

White M, Epston D. Narrative means to therapeutic ends. WW Norton; 1990.

32.16 Positive Psychology

NANSOOK PARK, PH.D.

Positive psychology is an umbrella term describing the scientific study of what makes life most worth living. The good life, that is the happy, healthy, productive, and meaningful life, entails more than the mere absence of disease, disorder, and problems. Traditional approaches in psychology largely focus on identifying problems and reducing them. However, it is limited in addressing how to build and sustain a good life, which requires additional skills and conditions. The field of positive psychology is the science and practice that has the potential to fill this need.

Research findings from positive psychology are intended to provide a more complete and balanced scientific understanding of the human experience. The new field of positive psychology calls for as much focus on strength as on weakness, as much interest in building the best things in life as in repairing the worst, and as much concern with making the lives of normal people fulfilling as with healing pathology.

Positive psychology does not replace business-as-usual psychology, which often focuses on people’s problems and how to remedy them. Rather, positive psychology intends to complement and extend problem-focused psychology. The attention of positive psychologists is increasingly turning to deliberate interventions that promote the well-being of individuals and groups, and the same qualifications apply.

For decades, scientific psychology has neglected the study of what can go right with people and often has had little more to say about the good life than pop psychologists, inspirational speakers, and armchair gurus.

More subtly, the underlying assumptions of psychology have shifted to embrace a disease model of human nature. Human beings are seen as flawed and fragile, victims of cruel environments or casualties of bad genetics, and if not in denial, then at best, in recovery. This worldview has even crept into the common culture, and some people have become self-identified victims, trying to survive but not flourish. Positive psychology proposes that it is time to correct this imbalance.

Positive psychology studies what goes right in life, from birth to death. It is concerned with an optimal experience, people being their best, and doing their best. Everyone's life has peaks and valleys, and positive psychology does not deny the low points. Its signature premise is more nuanced: what is good about life is as genuine as what is bad and therefore deserves equal attention from psychologists. Positive psychology assumes that life entails more than avoiding or undoing problems and that explanation of the good life must do more than reverse accounts of distress and dysfunction.

HISTORY OF POSITIVE PSYCHOLOGY

The most basic assumption that positive psychology urges is that human goodness and excellence are as authentic as disease and disorder. Positive psychologists argue that these topics are neither secondary nor derivative. There are many good examples of psychological research, past and present, that illustrate the fundamental premise of positive psychology, although only recently has there been explicit emphasis on their common concerns.

Accordingly, positive psychology has a very long past but only a short history. The long past of positive psychology stretches at least to the Athenian philosophers in the West and to Confucius and Lao-tzu in the East. In the writings of these great thinkers can be found the same questions posed by contemporary positive psychologists. What is the good life? Is virtue its own reward? What does it mean to be happy? Is it possible to pursue happiness directly, or is fulfillment a by-product of other pursuits? What roles are played by other people and society as a whole?

Early Ancestors

Within psychology, the premises of positive psychology were laid out long ago. More than 100 years ago, when psychology began as a formal discipline in Germany, it had a narrow focus on sensation and perception. Psychology soon spread. When it came to the United States, it was blended with American pragmatism by William James, John Dewey, and others to result in *functionalism,* which was concerned with the consequences—functions—of psychological states, traits, and processes. Inevitably, functionalism gave rise to a host of applied psychologies and out-of-the-laboratory endeavors with practical goals.

Among the early goals of applied psychology was to understand accomplishment and satisfaction in domains such as education, religion, work, and the family. The very first applied psychologists had the explicit intent to identify what was successful in these important areas of life and to apply this knowledge to make the lives of normal people more fulfilling.

The first psychological clinic in the United States was founded in 1896 by Lightner Witmer at the University of Pennsylvania. Witmer is credited with coining the term clinical psychology, a field which he defined as “the study of individuals, by observation or experimentation, with the intention of promoting change.” Many of the clients at Witmer’s clinic were school children with learning difficulties, and the goal of the clinical work was to make them better students. Sounding like a contemporary positive psychologist, Witmer argued that “The progress of psychology, as of every other science, will be determined by the amount of its contribution to the advancement of the human race.”

Had this original mission of psychology been sustained, there would be no positive psychology today, although all of its current concerns would be well-represented. However, events occurred that changed psychology. With the rise of behaviorism and the positivism that provided its philosophical basis, a strict distinction between fact and value emerged, with the fact being seen as the province of science and value as the province of philosophy. Many psychologists stopped studying things that mattered morally.

In the early 20th century, psychologists interested in individual differences, what is now called personality, divided their subject matter into character—which could be changed—and temperament—which could not. The moral flavor of character was seen as a critical feature. Gordon Allport, the major personality trait theorist in U.S. psychology at that time, was uncomfortable with the notion of character and actually banished the term from academic discourse concerning personality. Allport argued that character was the subject matter of philosophy and not psychology. The traits he urged psychologists to study were presumably objective entities stripped of moral significance and linked to “adjustment” but not endowed with inherent value.

Another turning point was the separation between basic psychology and applied psychology. Whereas the early functionalists made no such distinction, the next generation of psychologists in the United States began to draw a line between laboratory experiments investigating “basic” processes of perception, thought, and learning and field studies investigating outcomes of interest because of their substantive content. The former studies were regarded as more important, and the elite schools stopped recruiting applied psychologists to their departments. Applied work, of course, was continued by psychologists in schools of business, education, kinesiology, and elsewhere, but what they did was not part of the acknowledged mainstream.

As important as these events were, the biggest change in psychology was the result of an external event. The realities of World War II changed psychology and spurred the growth of clinical psychology, now the numerically dominant specialization within psychology. With an unprecedented proportion of the adult male population serving in the armed forces and experiencing the stresses of modern war, the need for acute clinical services outpaced what could be provided by available psychiatrists. Psychologists, heretofore involved in clinical work mainly in devising, administering, and interpreting psychological tests, were called upon to provide treatment. What resulted was a generation of basic psychologists who through necessity became practitioners.

The immediate aftermath of World War II saw no reduction in the need for clinical services. Accordingly, the new generation of scientist–practitioner psychologists consolidated, formalized, and eventually institutionalized an identity and a profession, encouraged and supported by the U.S. federal government. The National Institute of Mental Health (NIMH) was created to support research into disorders and their treatment, and stipends for graduate training programs in clinical psychology were part of the NIMH portfolio. The Veterans Administration (VA) provided clinical psychology internships for students in these programs.

The growth of clinical psychology as a profession has been phenomenal. The downside of the growth of clinical psychology is that its view of people has come to dominate much of psychology and even the common culture. What students learn is that people are victims—of unfortunate genetics, cruel environments, and their own cognitive shortcomings. This characterization applies not just to clinical psychology but to the entire discipline. For example, the field of social psychology began in the 1940s with Kurt Lewin's stirring call for *action research,* using psychology to solve social problems, but it has since become a field that documents repeatedly what people cannot know and cannot do.

This is the context in which Martin Seligman, a mainstream academic psychologist, called for psychology to reclaim its original mission and study the rest of the human condition. The field was named in 1998 as one of the initiatives of Martin Seligman in his role as president of the American Psychological Association. A trigger for positive psychology was Seligman's realization that psychology since World War II had focused much of its efforts on human problems and how to remedy them.

The yield of this focus on pathology has been considerable, but there has been a cost. Much of scientific psychology has neglected the study of what can go right with people and often has little to say about the psychological good life. More subtly, the underlying assumptions of psychology have shifted to embrace a disease model of human nature. People are seen as flawed and fragile, casualties of cruel environments or bad genetics. This world view has even crept into the common culture, where many people have become self-identified victims.

Positive psychology is not intended to replace business-as-usual psychology but rather to complement and extend it, by drawing on past work into the psychological good life as well as doing new work from the positive psychology perspective.

In one sense, positive psychology is nothing new, and William James, John Dewey, Lightner Witmer, and other psychologists from a century ago can be considered positive psychologists. Yet, the label "positive psychology" has considerable value in the here and now because it draws together what had been scattered lines of inquiry within psychology and allied social science—for example, investigations of life and work satisfaction, utopian visions of education like those of Alexander Sutherland Neill; primary prevention programs based on notions of wellness as pioneered by George Albee and Emory Cowen; work by Albert Bandura and others on human agency and efficacy; work by Mihalyi Csikszentmihalyi on creativity, flow, and optimal functioning; George Vaillant's longitudinal studies of the benefits of mature defense mechanisms; investigations of resilience; studies of giftedness; conceptions of intelligence as multiple; and assessments of the quality of life among medical and psychiatric patients that went beyond their symptoms and diseases.

Humanistic Psychology

Chief among the immediate ancestors of today's positive psychology is *humanistic psychology* as popularized by Carl Rogers and Abraham Maslow (who actually used the phrase "positive psychology" to describe his own approach). Humanistic psychology is one of psychology's venerable perspectives, popular in the 1960s and 1970s and still with adherents today. *Humanism* is the doctrine that the needs and values of human beings take precedence over material things, and further, that people cannot be understood simply as part of the material world. Humanists argue that scientific psychologists miss what is most important about people by focusing on the supposed causes of behavior, as if people were simply billiard balls, doing poorly or well depending on what other billiard balls happen to have ricocheted into them.

Both Rogers and Maslow emphasized that people strive to make the most of their potential in a process called *self-actualization.* Self-actualization can be thwarted by various conditions, but if these conditions are changed, then the potential within each individual will necessarily unfold.

This is a very different way of thinking about human nature than that embodied in psychoanalysis or behaviorism, dominant perspectives within psychology during the 20th century. Humanistic psychology stresses the goals for which people strive, their conscious awareness of this striving, the importance of their own choices, and their rationality. The attention of psychology is thereby directed away from mechanical causes and toward fundamental questions about existence and meaning.

In a 2000 discussion of positive psychology, Martin Seligman and Mihalyi Csikszentmihalyi tersely distanced the new field from humanistic psychology. They made two arguments. First, contemporary positive psychology regards both the good and the bad about life as genuine, whereas humanists often—but not always—assume that people are inherently good. Second, positive psychology is strongly committed to the scientific method, whereas humanists often—but again, not always—are skeptical of science and its ability to shed light on what really matters.

As points of relative and occasional contrast, these arguments are reasonable, and they highlight what is novel about positive psychology. Yet, as positive psychology has evolved and as positive psychologists have more carefully examined allied perspectives, positive psychology and humanistic psychology should be seen as close relatives.

Many humanists would agree that each person has the capacity for good and bad, just as positive psychology assumes. That "the good life is simply a matter of choice" seems to go too far, given the well-documented barriers to thriving posed by external circumstances like pestilence, poverty, and prejudice, but many positive psychologists nowadays acknowledge that notions of choice and will are indispensable ones. Finally, many humanistic psychologists, from decades ago to the present are as committed to science as any positive psychologist—although one can still question the success of humanistic psychology in creating a cumulative database.

Positive Mental Health

One more important historical precursor of positive psychology needs to be mentioned. In 1958, the U.S. Joint Commission on Mental Illness and Health was authorized by a unanimous resolution of Congress. The commission published a monograph series, and the first book in the series was *Current Concepts of Positive Mental Health* by social psychologist Marie Jahoda. Foreshadowing the arguments of positive psychology decades later, Jahoda argued that mental health was not simply the absence of mental illness, and further that psychological well-being needed to be understood in its own right.

Jahoda surveyed what previous thinkers—mainly clinicians—had to say about positive mental health and integrated their views. She proposed half a dozen underlying processes that produce the states and traits usually identified as psychologically healthy:

- Acceptance of oneself
- Accurate perception of reality
- Autonomy (freedom from social pressures)
- Environmental mastery
- Growth, development, becoming
- Integration of personality

Note that she did not include happiness or satisfaction as part of her definition, believing that these were more properties of the situation than of the individual.

At the same time that Jahoda published her book, social psychologist William Scott surveyed the existing research literature on mental health, focusing on research definitions (measures) of well-being and empirically established correlates of these measures. Then as now, the majority of studies ostensibly looking at health were really studies of disorder. Scott could draw firm conclusions only about the factors that characterized the lack of pathology. Good social relationships were the most common correlate. Scott cited a personal communication from Jahoda in which she described her own attempts to develop measures of several of her criteria for mental health (i.e., accurate perception of reality and environmental mastery) as less than satisfactory, which may explain why her ideas did not stimulate further work for decades.

More recently, Carol Ryff and her colleagues repeated Jahoda’s analysis by surveying and integrating what different theorists, again mostly clinicians, said about the psychological components of being and doing well—striving and thriving, as it were. They identified six points of convergence across discussions of psychological well-being, which agree substantially with those specified decades earlier by Jahoda:

- Autonomy
- Environmental mastery
- Personal growth
- Positive relations with others
- Purpose in life
- Self-acceptance

Notable about Ryff’s work is that she and her colleagues have solved the measurement problem and have created reliable and valid self-report surveys of these components of psychological well-being. Using these measures, they have explored the links between psychological well-being and physical health, finding positive associations and implicating the left prefrontal cortex—believed to aid in the organization of goal-directed activity—in these associations.

Building on Jahoda’s earlier work, contemporary positive psychologists suggest that doing well (mental health) and doing poorly (mental illness) are not simple opposites. They argue that these are distinguishable dimensions in terms of which people can be described (Fig. 32.16–1). Complete mental health is defined by high levels of psychological well-being along with the absence of signs and symptoms of psychological disorder. Said in DSM IV terms, complete mental illness is described by one or more Axis I diagnoses and a low Axis V (global assessment of functioning) score.

Figure 32.16–1 represents how we think about intervention from a positive psychology perspective. It is based on one of the signature premises of positive psychology that well-being is not simply the absence of pathology. Both well-being and pathology are multidimensional, and their various components can and do coexist. The framework is shown in Figure 32.16–1. This framework depicts strengths and problems as in principle independent of one another.

Individuals can be placed in Figure 32.16–1 depending on their particular profiles of strengths and problems. In Quadrant 1 are strengths accompanied by problems. We like to describe individuals in Quadrant 1 as "dealing with it," meaning that they have and use strengths despite experiencing problems and challenges.

The fully functioning person is found in Quadrant 2: high strengths and low problems. Here we find health as defined by in the constitution of the World Health Organization (WHO): "A state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity."

Individuals in Quadrant 3 are the ones most frequently studied by clinical psychologists, poorly functioning individuals with psychiatric diagnoses or problems such as alienation, demoralization, and loneliness.

Quadrant 4 is both unremarkable and understudied, occupied by individuals with neither problems nor strengths, passing time quietly, complacent but not joyful. What may characterize Quadrant 4 is a history of insufficient challenge. Because life has been easy, strengths have not developed or have not needed to be displayed.

Traditional psychotherapy or consultation is usually concerned with solving problems, moving individuals from Quadrant 3 to Quadrant 4. Figure 32.16–1 suggests additional intervention goals. The ultimate goal of practitioners should be to move people and groups into Quadrant 2 and to keep them there. Depending on where individuals and groups start, different interventions are probably needed because change will entail different processes. In some cases, strengths need to be built from scratch, whereas in other cases, individuals and groups need to capitalize on existing strengths.

Even among individuals and groups already in Quadrant 2, their well-being can be further promoted. Promotion needs to be distinguished from prevention, the goal of which is to prevent problems by reducing risk factors. Prevention is of course important, but attention to risk factors alone is not sufficient because people who develop problems may lack obvious risk factors. The positive psychology perspective is that building or enhancing strengths may be an additional and broader strategy for intervention.

When first introduced by Seligman, positive psychology was described as focusing on complete mental health and its further enhancement, but the scheme suggested by positive psychologists sensibly expands positive psychology and provides a useful way to describe the range of therapeutic interventions. For example, what can be called *positive psychotherapy* attempts to move people with problems closer to the state of complete mental health. The goal of *promotion* is to increase the well-being of those without problems. *Maintenance* keeps healthy people healthy. Given that existing treatments for psychological problems rarely if ever “cure” people in the sense of precluding recurrence of problems, it is surprising how little attention has been paid to the matter of maintenance. Along these lines, positive psychology may provide new insights into the challenging task of *relapse prevention:* helping people whose problems have been resolved to stay healthy. *Traditional therapy* simply tries to mitigate someone’s psychological problems.

Incomplete mental illness and incomplete mental health are common states, and positive psychology has potential value in helping people who can be so described, as well as learning lessons from them. Incomplete mental illness might be a suitable label for people like Abraham Lincoln or Winston Churchill, who functioned well despite recurrent depression; somehow, they had learned to “deal with” their problems. Incomplete mental health might be a suitable label for those who lead lives of quiet desperation or those who earn NOS (not otherwise specified) diagnoses, should they enter the mental health system.

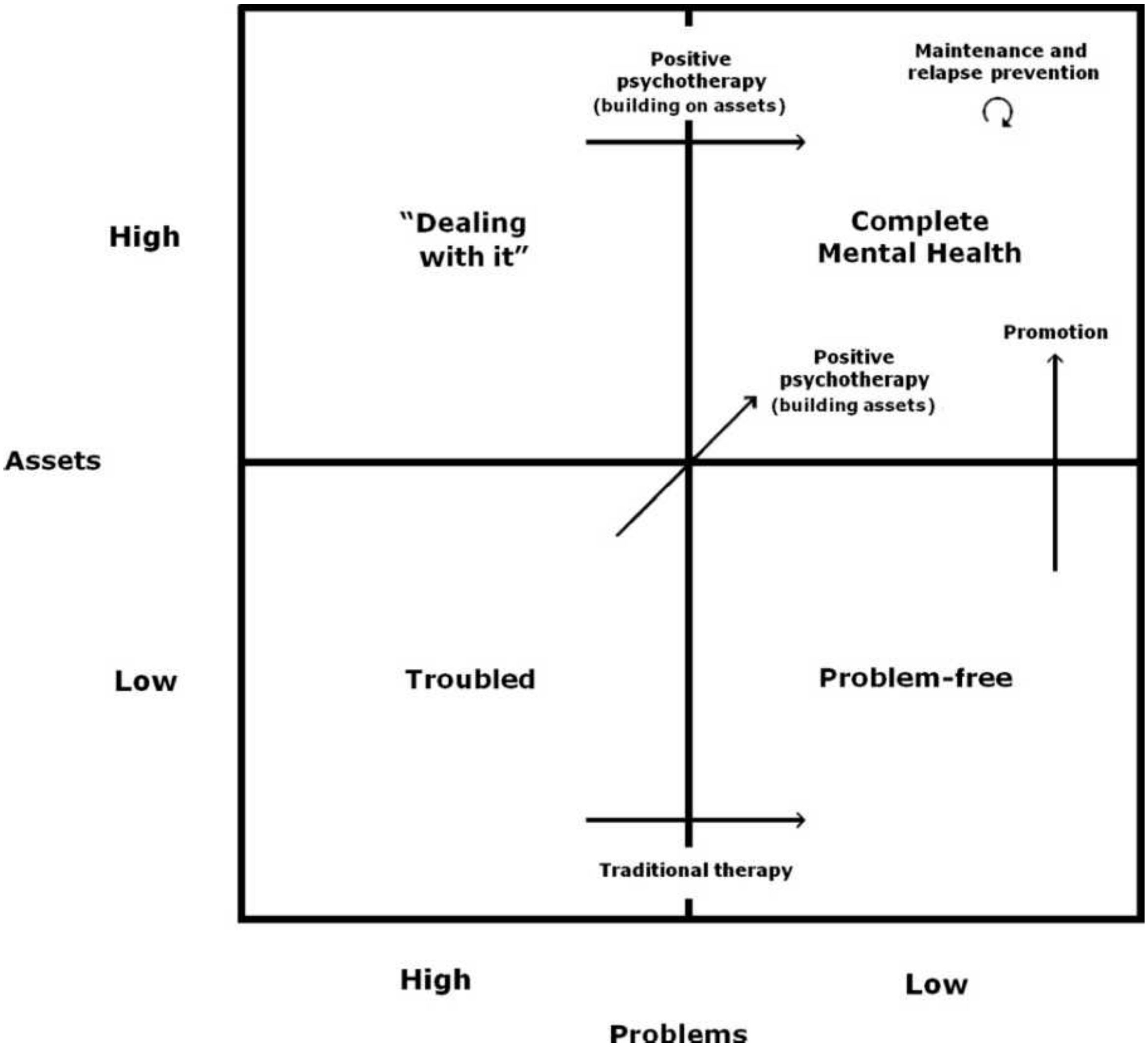
Positive psychotherapy (building on assets)
Maintenance and relapse prevention
High
"Dealing with it"
Complete Mental Health
Promotion
Positive psychotherapy (building assets)
Assets
Low
Troubled
Problem-free
Traditional therapy
High
Low
Problems

FIGURE 32.16–1. Mental health and mental illness. These are dimensions. Quadrants are shown for the purpose of illustration only.

CONCERNS OF POSITIVE PSYCHOLOGY AND CRITICISMS

The framework of positive psychology provides a comprehensive scheme for describing and understanding the good life and how it can be attained and sustained, regardless of one's starting point. The concerns of the field can be divided into four related topics:

- Positive subjective experiences (happiness, gratification, fulfillment, growth, and flow)
- Positive individual traits (strengths of character like humor, love, gratitude, forgiveness, modesty, spirituality, optimism, kindness,

creativity, courage, and wisdom; talents; interests; values; and meaning and purpose)

- Positive interpersonal relationships (friendship, marriage, and colleagueship)
- Positive institutions (families, schools, businesses, communities, and societies)

A theory is implied here: positive institutions enable the development of positive relationships, which enable the display of positive traits, which in turn enable positive subjective experiences.

The word “enable” avoids strict causal language. It is possible for people to be happy or content even in the absence of good character, and good character can operate against the interpersonal and institutional grain. However, people are at their best when institutions, relationships, traits, and experiences are in alignment. Doing well in life represents a coming together of all four domains.

As a new field, positive psychology raises questions. Despite the pervasiveness of a victim mentality, everyday people seem to know that the elimination or reduction of problems is not all that is involved in improving the human condition. In contrast, the scientific community is more skeptical. Contributing to skepticism are widespread assumptions within the social sciences about human nature as brittle and broken, notions more explicit among social scientists than the general public.

Positive psychology is also criticized in some quarters for a relentless emphasis on being positive, that is, happy and cheerful. This criticism of the field reflects a misunderstanding, because positive psychologists merely propose that what is positive about life is worth studying, in addition to what is negative. Happiness is one of many topics of interest to positive psychology, but so too are virtues like gratitude and forgiveness, generosity, wisdom, optimism, and resilience, meaning and purpose, engagement, and good relationships.

Positive psychologists do not deny the problems that people experience, and positive psychologists do not ignore the negative—stress and challenge—in their attempts to understand what it means to live well. As already emphasized, positive psychology intends, by expanding the topics of legitimate study, to complement business-as-usual psychology, not to replace it.

Indeed, what is most troubling in life can set the stage for what is most fulfilling. Consider that complex emotional experiences often blend the positive and negative, that optimism is most apparent when people confront setbacks and failures, that crisis reveals strengths of character, and that ongoing challenge is a prerequisite to experience flow in the moment and to achieve something important in a lifetime. Along these lines, many positive psychologists believe that identifying and using what one does well can be an effective way to address and resolve psychological problems.

In fairness to critics, positive psychology has become so popular that at least some people appropriate the label and present "positive psychology" as a simplified version of the power of positive thinking. All one can do here is assert that Pollyanna is not the poster child for positive psychology, that the fortunate in life—those who are happy—deserve no special congratulations, and that the unfortunate—those who are unhappy—deserve no special blame.

Positive psychology does not endorse wishful thinking, self-deception, or quick fixes. Rather, it uses the scientific method to understand and enhance well-being. The lesson of history is that patience is needed. Great thinkers for many centuries have pondered well-being, and progress in positive psychology will be slow and difficult.

EMPIRICAL FINDINGS

Positive psychology already has a canon of established findings worth considering. Here are some of the research findings that provide some insight into ways to build a happy, healthy, resilient, and thriving life.

In contrast to stereotypes that happy people are naïve, stupid, or in denial, empirical studies show that happiness and positive emotions are beneficial. The real worry should be about people who are unhappy and not satisfied with their life.

Most people are happy. When psychologists study self-reported happiness and life satisfaction, usually under the rubric of subjective well-being, they administer numerical rating scales. The consistent and perhaps surprising result is that most people in most circumstances most of the time score above the scale midpoint, whether they are multimillionaires in the United States or pavement dwellers in Calcutta. This conclusion holds across demographic characteristics like age, gender, ethnicity, and education, each of which has a surprisingly small association with avowed happiness.

Happiness is causal not epiphenomenal. People who are successful in life's venues are of course happy, but the less obvious and more interesting finding from experimental and longitudinal research is that happiness actually foreshadows success in academic, vocational, and interpersonal realms.

For example, researchers analyzed pictures from the 1958 and 1960 yearbooks of a women's college. All but three of the young women were smiling, but the smiles varied. Some displayed what is termed a Duchenne smile: a genuine, full-faced expression of happiness indexed by the degree to which the muscles surrounding one's eyes are contracted. Others showed the forced grimace known as a flight-attendant smile. The researchers chose these particular pictures for analysis because the women in them were participants in a long-term study of important life events. Specifically, the researchers knew—decades after their yearbook photos—whether the women were married and if they were satisfied with their marriage. As it turns out, ratings of the Duchenne-ness of their yearbook smiles predicted both of these outcomes. Young women who expressed genuine positive emotions (happiness) in yearbook photos, and presumably in other areas of their lives, as middle-aged women had better marriages. The skeptic might wonder if these results reflect the operation of some confound like physical attractiveness. As it turns out, rated "prettiness" was independent of the Duchenne-ness rating and moreover did not predict who had an intact or satisfying marriage.

Happiness can be increased. The possibility of deliberately pursuing happiness is a relatively recent arrival on the historical scene, immortalized by Thomas Jefferson in the U.S. Declaration of Independence. Prior to this time, happiness was relegated to the afterlife or seen as a matter of luck. Even today, some psychologists are skeptical that happiness can be increased.

Several factors contribute to this skepticism. First is the well-documented fact that people adapt to pleasure. When they repeatedly encounter the same pleasure-producing stimulus, they experience increasingly less pleasure in response. Adaptation to pleasure is so common that theorists have proposed that people live on a *hedonic treadmill,* meaning that they continually adapt to improving circumstances to the point that they always return to a point of relative neutrality. Another reason for pessimism about the possibility of increasing happiness is its demonstrable heritability, which has led theorists to posit a genetically determined *happiness set-point,* above which one cannot rise.

Research challenges these arguments. Current thinking about the hedonic treadmill and the happiness set-point holds that these are at best loose metaphors and hardly biologic realities. Heritability does not mean unchangeable. Remember the finding that most people most of the time are happy, which means that the metaphorical treadmill is not set at neutral. Furthermore, longitudinal studies show that the metaphorical set-point can be altered by life events like marriage and unemployment.

Other studies show that happiness can be increased when people simply act differently. For example, researchers asked introverted and extraverted college students to keep a careful diary for 3 weeks of their activities and mood. Both introverts and extroverts were happiest when doing extraverted things, implying that anyone who behaves in an outgoing way will be happy as a consequence, whether or not this behavior comes naturally. Happiness results not from what one is predisposed to do but instead from what one actually does. Growing evidence shows the importance of individual circumstances, cognitions, and behaviors for happiness.

Along these lines, researchers have shown that people instructed to do the sorts of things that characterize happy people indeed become happier as a result:

- Being grateful (e.g., counting blessings on a regular basis)
- Expressing sincere gratitude (e.g., writing a gratitude letter or gratitude visit)
- Creating flow (e.g., seeking out activities at work or at leisure that challenge one’s abilities)
- Giving service to others (e.g., performing volunteer work)
- Performing acts of kindness (e.g., doing small good deeds for others on a regular basis)
- Counting acts of kindness (e.g., recalling our prosocial behaviors on a regular basis)
- Practicing self-compassion (e.g., treating ourselves with compassion, not being too hard on ourselves)
- Being socially connected (e.g., building positive and deep relationships with others who matter)
- Engaging in a positive approach to life (e.g., finding the good in life, holding a positive future outlook)

- Spending time in nature (e.g., hiking, taking a walk in the park, gardening, having more interactions with nature)
- Experiencing healthy and adaptive humor (e.g., having good laughs, watching comedy films or shows)
- Savoring good things that happen (e.g., sharing experiences and memories with others, building memories, getting observed in the moment, using all senses, congratulating the self)
- Identifying and using character strengths in novel ways (e.g., seeking out different ways to display one’s humor, teamwork, or love of learning)
- Articulating one’s hopes and wishes (e.g., writing about an envisioned and desired “possible self”)

People are terrible at emotional forecasting. People of course know that a promotion at work will make them happy and that a divorce will make them sad, but research consistently shows that people cannot accurately estimate how long these emotional reactions will last. They consistently overestimate their duration and further seem not to learn from experience that they do so. Clinicians know that part of treating depression is addressing the convictions of clients that they will never feel any differently, when in fact most depressive episodes are self-limiting. Conversely, the positive psychology practitioner needs to convey the information that even the most exhilarating experiences and activities lose their magic, sometimes quickly.

Happiness leads to physical well-being. People who are happy recover more quickly from infectious disease and live longer than their peers. For example, in an experimental study, volunteers were exposed to the virus that causes the common cold. All participants became ill, but those who were happier had less severe symptoms and recovered more quickly. For another example, from the well-known longitudinal study of Catholic nuns, women who described positive emotions in essays they wrote in their early twenties were more likely to be alive decades later. Indeed, the happier nuns (those in the upper 25% of the essay writers) lived on average 10 years longer than their less happy counterparts (those in the bottom 25%). Positive psychology interventions that bolster happiness may be of considerable value in the applied fields of behavioral medicine and health psychology.

Positive emotions and negative emotions are distinct. Theorists who catalogue basic emotions usually include one or two positive ones—for example, joy and sometimes surprise or curiosity—but their focus is on the negative ones. Part of the reason for neglecting positive emotions is the problem-focused nature of psychology in general; “emotional” problems are after all a synonym for psychological disorders. However, there are important differences between positive emotions and negative emotions. They have separate bases in brain chemistry, and they can coexist. They also function differently. Negative emotions are associated with specific behavioral predispositions. Fear makes people want to run, and anger makes them want to attack. In contrast, positive emotions are not linked to such specific action tendencies. Positive emotions like joy may activate people, but in a much more vague and diffuse way.

Positive emotions broaden and build people’s psychological and behavioral repertoires. Negative emotions—fear, anxiety, or anger—alert people to danger. When one experiences a negative emotion, response options narrow, and the person acts with haste to avoid, escape, or undo whatever danger is signaled. In contrast, positive emotions signal safety and the inherent response to them is not to narrow options but to broaden and build upon them. Laboratory studies have shown that inducing positive moods among research participants (by showing them brief video clips) leads to enhanced performance on tasks requiring creativity and flexibility. Perhaps it is advantageous to experience positive emotions because they lead people to engage in activities that add to their behavioral and cognitive repertoires, building so-called psychological capital.

Positive emotions undo the physiological effects of negative emotions. For example, in one study, college students were given just 1 minute to prepare a brief speech on “why you are a good friend” that would—so they were told—be videotaped and evaluated by their peers. They did so while having their heart rate, peripheral vasoconstriction, and blood pressure (all indices of sympathetic nervous system activation) measured. Following this preparation time, research participants saw one of four different films; two produced positive emotions (joy or contentment), one produced feelings of sadness, and the last one was emotionally neutral. None of these films in and of itself had an effect on sympathetic arousal. However, following the speech preparation exercise, research participants who then saw either of the positive emotions films showed more rapid cardiovascular recovery (return to baseline) than those who saw the sad film or the neutral film.

Engagement matters. Flow is the term used to describe the psychological state that accompanies highly engaging activities. During flow, time passes quickly for the engaged individual. Attention is focused on the activity itself. The sense of the self as a social actor is lost. The aftermath of the flow experience is invigorating. Flow is not to be confused with sensual pleasure. Indeed, flow in the moment is nonemotional and arguably nonconscious. People describe flow as highly and intrinsically enjoyable, but this is an after-the-fact summary judgment, and joy is not immediately present during the activity itself. Flow can be described as the experience of working at full capacity, which is why the concept figures prominently in positive psychology.

Unlike many positive psychology concepts, flow has been extensively investigated, using an experience sampling method which asks research participants to carry a pager that goes off at random intervals. They are then asked to describe what they are doing and how they are feeling at the time. From studies of this sort, psychologists have learned that flow is most likely to occur when there is an optimal balance between skill and challenge. One need not be an expert in a domain to experience flow. All that matters is that the presented challenge meets one’s skills, and vice versa.

Flow can be experienced in all sorts of activities, at work or play, but usually among those perceived as voluntary. For example, homework for most school children strikes a balance between challenge and skill, but because homework is regarded as coerced, it rarely produces the flow state. Indeed, flow is rarely experienced by youth during any school activity. The other common activities of adolescents in the United States—watching television or “hanging out” with friends—also fail to produce flow, because they fail to meet one or another of its enabling conditions. People differ greatly in the frequency with which they experience flow, for reasons not well understood, and young people who experience more flow during adolescence show long-term desirable consequences, such as achievement in creative domains. They may even have better physical health.

“Good days” have common features. Wishing someone a good day is a common conversational gambit, and positive psychologists have been able to describe what a good day entails. For most people, it is one in which they do things that make them feel autonomous, competent, and connected to others. Encouraging people to experience pleasant events

has long been a staple of behavioral therapy for depression, and this research makes the intervention less a matter of trial-and-error.

A rosy view of matters is associated with physical, psychological, and social well-being. For example, optimism assessed early in life predicts health decades later and even whether someone is still alive. For another example, viewing one's spouse in a more positive way than do others—presumably an idealization—predicts a good and lasting marriage. Data showing that positive illusions like these are beneficial stand in sharp contrast to theoretical arguments mounted by typical clinical psychologists and psychiatrists that realism and accuracy are the hallmarks of health.

Positive expectations drive analgesic placebo effects through physiologic pathways. Specifically, dopamine—implicated in the experience of positive emotions—triggers the release of endorphins. Optimism and hope are not just in one's head but also in one's nervous system.

Most people are resilient. Stressful life events can take an enduring toll on well-being, and their consequences have been extensively studied. Usually not highlighted in research on post-traumatic stress disorder is that the majority of people exposed to life-threatening events do not experience lasting negative effects.

One of the effective strategies that contribute to resilience is the positive reframing of unfortunate situation or events. One example is the phenomenon of *downward social comparison* described by psychologist Shelley Taylor. Her research program began as an investigation of depression that followed in the wake of a diagnosis of breast cancer, a terrible and serious life event. However, Taylor had trouble identifying a sufficient number of cancer patients who were severely depressed. Instead, most were able to deal with the diagnosis by a process by thinking of someone who was faring more poorly than they were:

> I have breast cancer, but it could be worse…after all, I could be younger … the surgery required is a lumpectomy rather than a mastectomy, a single mastectomy rather than a double mastectomy … I can tolerate the side effects of chemotherapy better than most … I have a supportive family … and so on.

A previous generation of psychologists might have concluded that these women were in denial, but they were open-eyed, lucid, and sober. This study foreshadowed the premises of positive psychology by concluding that resilience and the cognitive work that sustains it were the rule rather than the exception.

In some cases, people may even show psychological growth following adversity, evidencing improved relationships with others, openness to new possibilities, greater appreciation of life, enhanced personal strength, and spiritual development. Interventions that specifically target these components of posttraumatic growth may help people recover from the effects of traumatic events.

Virtue is more than its own reward. Positive psychologists have begun to study the sorts of character strengths enumerated by philosophers and theologians over the centuries. Several findings are important. First is that notions of good character are not idiosyncratic or culturally bound. There is a strong consensus about the qualities of a “good” person. Second is that character strengths have demonstrable consequences. For example, the strengths of love, hope, curiosity, gratitude, and zest are robustly linked to greater life satisfaction and less depression. Perseverance foreshadows successful at school and at work. Hope and optimism predict good health and a long life. So too does generosity as indexed by altruistic acts.

Meaning and purpose matter. According to positive psychologists, a meaningful life is one in which people feel connected to something larger than themselves. Research consistently links the presence of meaning to well-being. Individuals with a sense of purpose report greater life satisfaction, more positive affect, higher levels of optimism, and better self-esteem. They are less likely to have psychological problems. However, the search for meaning is related to lower life satisfaction. It seems that although having meaning and purpose is beneficial, the process of finding life meaning may entail challenge and confusion. Research further shows that a life framed by meaning is more satisfying than a life centered on pleasure.

For example, researchers asked undergraduate students to keep an online diary for several weeks, noting the frequency they did pleasurable things or more meaningful things like helping others or simply listening to their problems, as well as how satisfied they were each day. The more people participated in meaningful activities, the happier they were. Pleasure, in contrast, was unrelated to satisfaction. In order to investigate the direction of these effects, the researchers looked at the sequencing of activities and satisfaction, finding that participants became happier after they did something good, suggesting that happiness does, in fact, result from meaningful activities and not vice versa.

Other people matter. Good relationships with other people are the single most important contributor to a satisfied life and may even be a necessary condition for happiness. Having a “best friend” at work is a strong predictor of work satisfaction and even productivity. A good relationship is one in which the amount of positive communication considerably outweighs the amount of negative communication.

Positive psychologists have taken a close look at the features of positive communication, describing four ways that a person can respond to someone else when he or she shares good news, including good events like a raise at work:

- Active–constructive responding—a positive and enthusiastic response: “That’s great; I am so proud of you; tell me more about it.”
- Active–destructive responding—a response that is critical or points out the potential downside: “Are they going to expect more of you now?”
- Passive–constructive responding—a muted response: “That’s nice, dear.”
- Passive–destructive—a response that conveys disinterest: “It rained all day here; what’s for dinner?”

Couples who use active-constructive responding have good relationships. The other responses, if they dominate, are associated with relationship troubles. Although this research has only been done in the context of romantic relations, it may well generalize to other relationships.

The important correlates of happiness are social in nature. In contrast to the modest demographic correlates of happiness and well-being, consider the following robust correlates:

- Number of friends
- Being married
- Being extraverted
- Being grateful
- Being religious/spiritual
- Pursuing leisure activities
- Employment (not income)

In a study that compared happy people to very happy people, there was one striking difference: good relationships with other people. Of the very happy people in the sample, all had close relationships with others. Psychology research documents very few necessary or sufficient conditions, but these data suggest that good social relationships may be a necessary condition for extreme happiness.

Kindness and generosity to others matter. Studies consistently show that doing volunteer work, helping, and giving are associated with high life satisfaction and good health. The implication of these findings is that encouraging people who have experienced difficult events to turn their attention to others might be beneficial. Doing so changes how the givers think about themselves, in positive ways that make them happier and healthier. It is also conceivable that doing service for others, as well as the feeling of contributing to others and society, gives people life meaning, that is the sense that their life matters. In particular, when people experience adversity and uncontrollable events, they feel helpless and depressed.

Perhaps encouraging people to engage in prosocial or service activities would empower them and increase their sense of self-worthiness. The activities do not need to be organized volunteer activities. Even simple acts of kindness in the community, followed by reflection on them, as through journaling, could have therapeutic effectiveness. For example, in a study with Japanese college students, merely asking them to count the kind acts they performed during the day led to increased happiness and more kind behaviors.

Money makes an ever-diminishing contribution to happiness. People who are so poor that they cannot meet their basic needs of course are unhappy, but above extreme poverty, increased income has a surprisingly small relationship with life satisfaction. In fact, money was related to experiencing less daily sadness, but no more daily happiness. The relationship between money and happiness is more complex. What people do with money matters more for happiness. That is, spending money on experiences increased happiness, while spending money on buying things did not. Also, when people spent money on other people, they became happier regardless of the amount of money they spent. Money has also been shown to have a downside. People with larger amounts of money savored less positive experiences compared to their lower-income counterparts.

Work matters. Despite the small contribution of income to well-being, whether or not someone is working is much more strongly related to happiness. People who are employed and engaged in what they do are happy, regardless of the status or compensation associated with their job. Happiness and engagement lead people to regard their work as a calling and to be more productive at whatever they do, take fewer sick days, and even postpone their retirement.

Eudaimonia trumps hedonism. According to Aristotle's notion of *eudaimonia*—being true to one's inner self (demon)—true happiness entails identifying one's virtues, cultivating them, and living in accordance with them. Contrast this notion with the equally venerable idea of *hedonism*—pursuing pleasure and avoiding pain—that is the foundation for *utilitarianism,* which in turn provides the underpinning of psychoanalysis and all but the most radical of the behaviorisms. Research shows that eudaimonia consistently trumps pleasure as a predictor of life satisfaction. Those who pursue eudaimonic goals and activities are more satisfied than those who pursue pleasure. This is not to say that hedonism is irrelevant to life satisfaction, just that all things being equal, hedonism contributes less to long-term happiness than does eudaimonia.

The "heart" matters more than the "head." As already mentioned, positive psychologists have begun to study character strengths—individual differences like curiosity, creativity, kindness, and teamwork. Measures have been devised and administered to people across the lifespan. Research consistently shows that strengths of the "heart" that connect people together—like love and gratitude—are much more strongly associated with well-being than are strengths of the "head" that are individual in nature like creativity, critical thinking, and aesthetic appreciation. Formal education of courses stresses the latter strengths, but if one goal of education is to encourage the good life, the research results suggest that the former strengths deserve attention as well.

Happiness, strengths of character, and good social relationships are buffers against the damaging effects of stressful life events. No one will go through life without problems and setbacks, but to the degree that people have more life satisfaction, greater character strengths, and better social support, they will experience fewer psychological or physical problems in the wake of difficulties and recover more quickly from them. These findings provide targets for primary prevention and may represent a unique contribution of the positive psychology perspective. Although all can agree that prevention of psychological problems is more efficient and humane than their treatment once they occur, understanding of prevention has lagged far behind what has been learned about the etiology of psychopathology. Perhaps researchers have been looking at the wrong things—risk factors rather than protective factors.

Positive institutions have common features. Over the last decade, research on positive institutions has been growing steadily. There is agreement that institutions that allow people to flourish—whether families, schools, workplaces, or even entire societies—share a core of common characteristics:

- Purpose—a shared vision of the moral goals of the institution, one reinforced by remembrances and celebrations
- Safety—protection against threats, danger, and exploitation
- Fairness—equitable rules governing reward and punishment and the means for consistently enforcing them
- Humanity—mutual care and concern
- Dignity—the treatment of all people in the institution as individuals, regardless of their position

POSITIVE PSYCHOLOGY AND CLINICAL WORK

The goal of interventions based on positive psychology was not only relieving problems but also helping people with or without problems to lead a fulfilling life. The therapeutic goals were not to move people from −5 to 0 (problem-free)—the presumed goal of conventional psychology—but rather break through the zero point.

Interventions based on positive psychology can help people to do more than help people doing well in life do even better. These interventions may also help people with psychological problems lead a more thriving life, moving them from −5 to +2 or beyond. The core of positive psychology interventions is the strength-based approach. Whatever the presenting complaints, people also bring into therapy assets and strengths that can be used to resolve their problems. A crucial task of any treatment is, therefore, to identify a client's strengths and resources and encourage their use, not just to solve problems but more generally to encourage flourishing. Such a balanced approach should build rapport and bolster client confidence. It would certainly make sense to clients.

When asked how they can tell that treatment has been effective, individuals with DSM diagnoses of depressive disorder describe their own view of "remission" in positive psychology language, spontaneously mentioning that would be more optimistic and confident, would have more energy, and would function well. They did *not* mention merely the reduction of symptoms or a change in diagnostic status. That is, depressed people, want to be more than just relieved of negative symptoms from their therapy. They also want to attain certain positive aspects of life.

Positive psychology interventions may also help psychologically healthy people stay that way—keeping them in the right-hand quadrants of Figure 32.16–1—by expanding the targets of prevention programs beyond risk factors to include those that maintain and create resilience. For those who have recovered from psychological difficulties, positive psychology interventions may prevent relapse, keeping them in the right-hand quadrants of Figure 32.16–1. Finally, these interventions may encourage and sustain physical health.

Perhaps the most general contribution of positive psychology to clinical work is to provide a vision of psychological health. Even if therapeutic concern is with traditional diagnostic categories, positive psychology provides substantive goals for any therapy that go beyond symptom reduction or removal.

Positive Psychology's Vision of Psychological Health

In its 1948 constitution, WHO defined *health* as "a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity." In more recent years, this statement has been expanded to include the ability to lead a life that is socially and economically productive. This definition is an important declaration that health entails more than the absence of illness, but it is circular, inasmuch as "well-being" is a synonym for "health." Work by positive psychologists makes this definition more concrete and thus useful as a guide for research and intervention.

If one can extrapolate from the sorts of topics that have been studied, positive psychology assumes that people are doing well when they experience more positive feelings than negative feelings, are satisfied with their lives as they have been lived, have deep positive relationships, have identified what they do well and use these talents and strengths on an ongoing basis, are highly engaged in what they do, are contributing members of a social community, and have a sense of meaning and purpose in their lives. Physical health and safety of course provide an important context for psychological well-being. It is difficult to imagine a cultural group in which these components of the good life are not valued. Respect for human diversity need not entail extreme cultural relativism.

Note that this fuller characterization of health reflects the WHO definition and is drawn from research in all of the domains of concern to contemporary positive psychology. The relevant research cautions that health so defined is not unitary. No one can have it all, at least at the same time, given tradeoffs among the psychological states and traits that reflect doing well. Psychological health therefore needs to be described with a profile of features and not a single summary score. DSM describes hundreds of psychological problems, and there may be just as many different manifestations of the good life.

The positive psychology vision of optimal functioning can also be used to make sense of the psychological problems that people do have. Problems (from a positive psychology perspective) represent shortcomings with respect to one or more of the components of the psychological good life. For example, people may experience more negative affect than positive affect; they may be dissatisfied with their lives; they may be socially estranged; they may be bored, alienated, or nihilistic; and so on.

These ways of thinking about problems of course differ from the DSM vision of psychopathology in terms of a set of symptom-defined disorders that are either present or absent. The positive psychology vision regards so-called “symptoms” as the problem and further expects them to exist in degrees. These visions need not compete. Traditional DSM diagnostic categories are not independent of the components of well-being implied by a positive psychology perspective, and a positive psychology perspective explains why given DSM symptoms and diagnoses constitute problems in living by identifying the ways in which optimal personal and social functioning are compromised by the disorder and its symptoms.

For example, it can be difficult to say why depression *per se* is considered a disorder, especially in light of discussions about its occasional utility and its basis in the evolutionary history of the human species. Yet, if depression is examined further in terms of its association with such difficulties as disinterest, indifference, helplessness, isolation and loneliness, pessimism, and alienation, then one can readily understand when and why depression is a psychological problem. At least in principle, a person who is “depressed” only in the sense of being sad and subdued does not have a problem. This person might simply be solemn and cautious, and Prozac should not be prescribed to counteract what may really be strengths of character.

Positive psychology can be criticized for being value-infused. The offsetting virtue is that the values inherent in positive psychology are clearly stated. The real question is whether these values are idiosyncratic or culture-bound. At a certain level of abstraction, the vision of psychological well-being that emerges from positive psychology seems ubiquitous if not universal.

As a perspective on topics that deserve scientific study, positive psychology has no single theory. Like much of contemporary psychology, it instead relies on midrange theories that draw on a variety of larger perspectives, from evolutionary to behavioral to cognitive to sociocultural models, in order to make sense of specific phenomena. Different topics are explained with different theories. The eventual integration of psychology may be a worthy goal, but it has not yet been achieved.

Positive Psychology Assessment

Much of psychological assessment has been tilted toward identifying weaknesses, deficiencies, and problems. The positive psychology perspective is that problem-focused assessment should be expanded to include strengths, assets, and competence. For example, low life satisfaction can occur in the absence of psychopathology, and it is nonetheless related to psychological and social problems. Conversely, high life satisfaction is linked to good functioning even in the presence of symptoms. The absence of problems is not equal to health. Problems and strengths can coexist.

Table 32.16–1.
Positive Psychology Measures

Positive Affect

- Positive and Negative Affect Schedule (PANAS)
- Authentic Happiness Inventory
- Orientations to Happiness Scale (measures endorsement of pleasure, engagement, and meaning)
- Subjective Vitality Scale
- Savoring Beliefs Inventory
- Life Satisfaction
- Satisfaction with Life Scale
- Marital satisfaction
- Work satisfaction

Positive Traits

- Values in Action Inventory of Strengths (VIA-IS) (adult and youth version)
- Ryff and Singer's Psychological Well-Being Scales
- Search Institute's Developmental Assets (for youth)
- The Life Orientation Test-Revised (measures optimism)
- Attributional style questionnaire (measures optimistic explanatory style)

Values

- Values Inventories of Rokeach, Schwartz, Scott, and others
- The Gratitude Questionnaire
- Transgression-Related Interpersonal Motivations Inventory (measures forgiveness)

Life Meaning and Purpose

- The Meaning in Life Questionnaire

Interests

- Strong–Campbell Vocational Interest Blank

Abilities

- Multiple intelligences

Social Support and Attachment

- The Multidimensional Scale of Perceived Social Support
- Adult Attachment Style Questionnaire
- Experiences in Close Relationships—Revised

Positive psychologists have already developed an impressive set of measurement instruments that allow someone doing assessment to break through the zero point of deficiency measures. For example, the healthiest that one can score on a typical measure of depression is 0, but this lumps together people who are blasé with those who are filled with zest and joy. The distinction seems well worth making, and the self-report surveys and interviews developed by positive psychologists allow it.

Space does not permit a detailed presentation of positive psychology measures, but Table 32.16–1 provides some examples. For example, the Values in Action Inventory of Strengths (VIA-IS) is a comprehensive survey that measures an individual's character strengths, 24 positive traits like curiosity, social intelligence, hope, kindness, zest, and teamwork. Positive psychology interventions are often described as strengths-based, which means that the results of this survey can provide useful information for the therapist and the client in identifying the client's strengths and devising interventions.

Many of the popular positive self-report surveys are available online at no cost at www.authentichappiness.org. Upon completion of a survey, individual feedback is provided, which means that clients can be directed to this website and asked to keep a record of their scores.

Most of the existing positive psychology measures were developed for research purposes, and they are most valid when aggregated to yield conclusions about groups of people. They can also be used ipsatively to describe the psychological characteristics of an individual and how they stay the same or change over time, but the cautious use of these descriptions is a point of discussion and departure in treatment. They are not strong diagnostic tests, and none should be treated as if they are. They should not be used alone for important decisions. Prudence is necessary for its interpretation and use, especially in the clinical setting. Such prudence is appropriate for all psychological assessments, but it is worth emphasizing in the special case of positive psychology measures.

Positive Psychology Techniques

Positive psychologists have shown that a variety of brief interventions can, in the short term, increase well-being as measured by happiness, satisfaction, and fulfillment. In some cases, there is also evidence that they can alleviate depression. These interventions have an obvious role not only in remedying distress following experience with adversity but also in building resources that blunt the negative effects of adversity in the first place.

When people experience adversity, they could be overwhelmed to appreciate the good things in life. In an experimental study by Lauren Conklin and her colleagues, depressed people showed no difficulties in learning negative information but had trouble learning positive information. Depressed people may have trouble recognizing and remembering the good things in life. This suggests that therapists should consider helping clients to identify, appreciate, and celebrate positive aspects of their everyday life. Here are some promising positive psychology intervention techniques of potential value in prevention, promotion, and/or treatment interventions.

Savoring. Savoring is one's awareness of pleasure and one's deliberate attempt to make it last. One way to cultivate positive emotion is to maximize when good things happen in one's life. Fred Bryant and Joseph Veroff examined the effects of savoring good events and found that people who do so are more satisfied. They also identified simple strategies for savoring, such as sharing good events with others, either in the moment or after-the-fact; building memories of the good events (e.g., photographs, diaries, souvenirs); congratulating the self when good things happen; sharpening perceptions during the experience of good events, and becoming fully absorbed in pleasure and not thinking about other matters. Professionals could also help their clients to use more of these strategies to maximize even simple and routine life experiences such as eating or walking. People who habitually savor are happier and more optimistic, and less depressed than those who do not savor.

Counting Blessings. Across the lifespan, people who are grateful have higher life satisfaction. Gratitude can be added to someone's repertoire by deliberate intervention, and a number of studies by different research groups have shown the effectiveness of so doing for both children and adults by increasing life satisfaction and optimism and reducing depression. Participants are usually asked to write down at the end of the day things that went well during the day and for which they are grateful. Details of this intervention may vary across the number of things listed and the frequency of listing them. This exercise can be modified depending on the client's age and situation. To avoid making this exercise burdensome to some people, especially children, it could be done a few times a week and count one or two blessings instead of three blessings. Although more evidence is needed for the effect of different frequencies and the amount of blessings on the outcome across different age groups, a study reported the negative effect when this exercise becomes excessive and, as a result, a burden.

In our own work, we have asked participants to write down why each good thing happened, encouraging them to be more mindful:

Every night for 1 week, set aside 10 minutes before you go to bed. Use that time to write down three things that went really well on that day and why they went well. You may use a journal or your computer to write about the events, but it is important that you have a physical record of what you wrote. It is not enough to do this exercise in your head. The three things you list can be relatively small in importance or relatively large in importance. Next to each positive event in your list, answer the question, “Why did this good thing happen?”

Participants, specifically those who continue to do this exercise, reported more happiness and less depression. The effect was maintained even in a 6-month follow-up. The mechanisms for these effects have not been closely studied, but there are likely several of them. This exercise could help people to pay more attention to the positive aspects of ongoing life and appreciate life more and be more optimistic and experience more positive emotions.

Strengths-Based Approach: Identifying and Using Strengths of Character in Novel Ways. Our multidimensional character strengths measures can be scored within the person (e.g., rank ordered)—to identify an individual’s “signature strengths” relative to their other strengths. Everybody has strengths regardless of where they may stand relative to others. This strengths-based approach is particularly useful for working with people with a history of disability, low achievement, or low self-esteem. They often have a hard time finding anything they are good at. However, if the strengths *within* a person are compared, those strengths that are stronger than others can be identified. Subsequently, professionals can help them to use these strengths in their lives.

After clients identify top character strengths using online questionnaires (VIA-IS), they are encouraged to use them in novel ways in their daily life:

Every day for the next 7 days, use one of your top five strengths in a way that you have not before. You might use your strength in a new setting or with a new person. It’s your choice.

Our research found that this exercise reduces depression and increases happiness and that the effect is evident even at a 6-month follow-up, so long as people continue to do the exercise.

Once individuals build their confidence by using their signature strengths, they can be taught how to use these strengths to work on weaknesses or less-developed strengths. If discussions and interventions start with the strengths of clients—things at which they are good—this can build rapport and increase motivation.

Research support is still accumulating. Enough outcome studies have been conducted to conclude that strengths-based approaches to change are more than just promising. Not known in most cases is how these expanded therapies fare in direct comparison to traditional treatments for problems and what are the mechanisms involved. Perhaps, attention to both strengths and weaknesses is critical and that no useful purpose is served by regarding these as mutually exclusive therapeutic goals.

See Table 32.16–2 for other examples.

Outcome research shows that a variety of psychotherapies are effective in alleviating problems, and usually equally so, despite the different forms they take. One interpretation of the equal effectiveness of different therapies is that nonspecific factors common to all treatments are responsible. Perhaps the sorts of strategies being studied by positive psychologists reflect these common factors and give names to them. Strategies like instilling hope and building strengths may be critical to the effectiveness of any therapy.

Table 32.16–2.
Positive Psychology Techniques and Interventions

Exercises to increase positive feelings
- Performing acts of kindness
- Savoring
- Writing a gratitude letter
- Walking in nature

Exercises to decrease negative feelings
- Turning one’s head to see the positive
- Positive reframing
- Self-compassion

Exercises to increase life satisfaction
- Counting one’s blessings

Exercises to develop talents and strengths
- Using talents and/or signature strengths of character in novel ways

Exercises to increase engagement
- Finding a challenging hobby

Exercises to increase social connectedness
- Being a good teammate
- Active–constructive responding

Exercises to increase meaning and purpose
- Performing secret good deeds
- Writing one’s own legacy
- Working for a valued institution

Some qualifications are in order if these techniques are used in the context of treatment. First, they are not therapies *per se* but simply exercises to be deployed as part of an overall treatment program. The therapist must ascertain a client’s readiness to change in the particular ways requested in the exercise, as well as the client’s capacity to make the change. Like any psychotherapeutic procedure, these techniques cannot be imposed on the unwilling or the unable.

Second, none of these techniques is akin to a crash diet or an antibiotic. To the degree that they have lasting effects, it is because clients integrate them into their regular behavioral routines. Counting blessings for a week will make a person happier for that week, but only if the person becomes habitually grateful will there be a more enduring effect. Research finds—not surprisingly—that the people who showed lasting benefits were those who continued to use the exercise.

Third, these exercises are typically presented as one-size-fits-all, but there is no reason to think that they are equally useful for all clients. Nothing is known about the match of an exercise with a client's particular presenting problems or goals or with a client's age, gender, social class, or ethnicity.

Fourth, little is known about the parameters of these interventions. For example, how many blessings should one count, and how frequently should this be done? With college students, counting blessings three times a week may be more effective in increasing happiness than counting them more frequently. Is this a general phenomenon or one specific to young adults attending college?

Fifth, all interventions run the risk of unintended harm, and while positive psychologists would like to believe that their techniques avoid iatrogenic effects, this assertion cannot be made with thorough confidence. For example, although optimism is related to mental and physical health, it would be simplistic and potentially hazardous to tell clients that positive expectations will solve all their difficulties. Along these lines, if a positive psychology intervention overemphasizes a client's choice and responsibility, considerable damage could be done in cases of abuse and victimization, where self-blame needs to be undone and certainly not encouraged. Interventions based on positive psychology should not preclude the use of existing therapeutic strategies when these are indicated.

Positive Psychotherapies

Positive psychotherapies are beginning to appear: therapeutic interventions based on the theories and findings of positive psychology. What distinguishes these emerging positive psychotherapies from conventional treatments is that their stated goal is not just symptom reduction or relief but also enhanced happiness, life satisfaction, fulfillment, productivity, and the like—one or more components of positive psychology's vision of the good life. These new therapies target people with psychological problems as well as those without (see Fig. 32.16–1). In the latter case, positive psychotherapies make contact with life coaching.

However, there is a danger in the label because it may give the misleading impression that an altogether different approach to treatment has been created. Almost all treatments labeled as positive psychotherapy are derived from well-established cognitive-behavioral approaches. The authors prefer to describe these interventions as *informed* by positive psychology and stress the continuity between them and existing approaches.

In identifying positive psychotherapies, the definition is crucial but difficult. As noted, positive psychotherapies have the goal of helping people thrive. They attempt more than the solution, eradication, or mitigation of problems. The possible field of positive psychotherapies is so broad that it needs narrowing, and somewhat arbitrarily, the focus here is on approaches characterized by an explicit *therapeutic alliance* between the positive psychologist and the client. This feature goes by many names and has been variously defined, but its recurring themes include collaboration between therapist and client, an affective bond between them, and agreement on the goals and tasks of the intervention. Asking people to write about their goals or to perform acts of kindness, despite beneficial consequences, may or may not be instances of positive psychotherapy; what matters is the relational context of the request.

This more narrow definition of positive psychotherapy is met by several approaches, most of which are recent. These are described in the next section. In most cases, there is only limited evidence for their effectiveness. This is not to say that the evidence argues against their use but simply that not enough research has been done. The good news is that most of these positive psychotherapies have been devised by research-minded psychologists, who are in the process of putting their ideas to the empirical test.

The further good news is that there are decades of available lessons from the psychotherapy outcome research literature to guide the evaluation efforts of positive therapists. The need for well-characterized samples, plausible comparison groups, random assignment, treatment manuals, objective outcome measures, effect size estimates, long-term follow-ups, and analyses of the process is recognized. All are readily applicable to the investigation of positive psychotherapy effectiveness.

In the meantime, investigation of the effectiveness of conventional therapies will continue, and it would be desirable for researchers to add positive psychology measures to their assessment batteries. These measures include a number that is brief yet well-validated. Although the inclusion of these measures would not transform conventional therapies into positive ones, they would provide important information for practitioners of all stripes. Do therapies that effectively treat anxiety or depression return clients to a typical level of life satisfaction or still leave them short? Might they, in some cases, lead to actual flourishing? Are the individual differences of concern to positive psychology moderators or mediators of traditional therapy effectiveness? One simply does not know, but these are easily answered questions given the availability of appropriate assessment techniques.

Positive Psychotherapy Examples. Here are descriptions of some of interventions that can be considered positive psychotherapies and—if the information is available—what is known about their effectiveness. As noted, positive psychology is a perspective, and no single theory underlies this perspective. When positive psychologists propose full-blown approaches to psychotherapy consistent with the premises of positive psychology, they turn to existing therapies for the overall model of treatment.

1. Decades ago, Michael Fordyce developed a *Personal Happiness Program* and showed it to be effective in boosting the long-term satisfaction of college students. Fordyce surveyed the research literature on happiness and identified predictors presumably under the short-term control of ordinary people—for example, keeping busy, socializing with others, doing meaningful work, and making the pursuit of happiness a priority. This information was conveyed to individuals along with suggestions to carry out 14 different behavioral and cognitive exercises over the course of a semester. In seven studies that included no-treatment comparison groups, Fordyce found that self-reported happiness increased and feelings of depression decreased. Follow-up was as long as 18 months. Fordyce believed that mere exposure to what he called “happiness principles” drove these effects, but one can suspect that the exercises were crucial, as well as their encouragement of positive emotions.
2. Already described were investigations of specific *positive psychology techniques* like counting one’s blessings and using signature strengths in novel ways. These studies were part of a deliberate program to develop and evaluate a positive psychotherapy. Much like Fordyce,

Seligman and his colleagues surveyed the scientific and popular literature for interventions purporting to increase happiness. More than 100 of these were identified, of which 40 were able to be manualized. Ongoing work is systematically testing the effects of these manualized interventions in randomized trials against a plausible placebo control (asking individuals to write about their early memories). The Internet is used to recruit participants, deliver the intervention, and collect data. So-called cybertherapy is obviously new and invites skepticism, but the only way to evaluate what promises to be a cost-effective strategy for helping people is to do the relevant outcome studies. Two of the interventions so far evaluated—counting one's blessings and using signature strengths in novel ways—increase happiness and decrease depression through 6 months of follow-up, with effect sizes as large as those of conventional psychotherapy. A third intervention—writing and delivering a letter of gratitude to another person—had a very large effect on happiness and depression, but only for 1 month. Two other interventions—writing an essay about "you at your best" and identifying one's signature strengths (without the instruction to use them in novel ways)—did not affect well-being. As noted, research is ongoing, and it will be just as important to learn what does not work as to determine what does.

3. The *positive psychotherapy* sketched by Martin Seligman, Tayyab Rashid, and Acacia Park packages together different positive psychology techniques and uses them sequentially in one-to-one therapy with depressed clients (see Table 32.16–3). Out-of-session homework assignments are given, and their results are the focus of discussion in treatment. Pilot data suggest that this form of treatment reduces depression and increases life satisfaction. However, more research is needed to determine the critical techniques for therapeutic outcomes, how each technique contributes to its result, and whether the order in which different techniques are given matters.

Table 32.16–3.
Idealized Sequence of Positive Psychotherapy Sessions

1. Orientation to therapy
2. Identifying strengths of character and approaches to happiness (pleasure, engagement, and meaning)
3. Using signature strengths of character in novel ways
4. Positive emotions in the past (reminiscence)
5. Undoing negative emotions in the past (forgiveness)
6. Gratitude
7. Taking stock
8. Meaning
9. Optimism and hope
10. Love and attachment
11. Family
12. Savoring
13. Giving to others
14. Integration: the full life

4. *Mindfulness-based cognitive therapy* (MBCT) combines techniques of cognitive therapy with the Buddhist practice of mindfulness meditation, which encourages people to be more aware of what they are thinking, to live in the moment, and to refrain from evaluating judgments about their own thoughts. Whereas typical cognitive therapy tries to change the content of thought, MBCT emphasizes nonjudgmental awareness. MBCT grew out of stress-reduction strategies developed by Jon Kabat-Zinn to help people with physical problems like chronic pain and hypertension. Zindel Segal, John Teasdale, and Mark Williams adapted these strategies to help individuals with recurring depression. MBCT entails weekly group sessions and homework assignments. Outcome studies show the effectiveness of MBCT in decreasing depression and anxiety. MBCT may also be more effective than conventional pharmacologic or psychotherapy treatment in reducing relapse rate. MBCT also improves the quality of life among medical patients, which makes it an instance of positive psychotherapy.
5. The *Penn Resiliency Program* (PRP) was developed by Jane Gillham and Karen Reivich to prevent depression by teaching children cognitive–behavioral skills before the development of problems. Its most-studied version is a 12-session program designed for group

administration to middle-school students. PRP contains two main components, one cognitive and the other based on social problem-solving techniques. In the cognitive component, core cognitive techniques are translated, through the use of cartoons and skits, into a language that adolescents can understand and apply to their own lives. In the social problem-solving component, students learn skills that help them better interact with this world: assertiveness, negotiation, relaxation, avoidance of procrastination, social skills, decision making, and problem solving. PRP has been successfully evaluated in school and managed care settings in both the United States and China. Results indicated that prevention participants report fewer depressive symptoms and are less likely to report symptoms in the moderate to severe range through 2 years of follow-up. Researchers are in the process of investigating the effects of PRP on life satisfaction, which means that it can be construed as a positive psychotherapy.

6. *Action and commitment therapy* (ACT), developed by Steven C. Hayes and colleagues, has not been explicitly discussed by its originators as a positive psychotherapy, although its goals are unambiguously positive: to reduce the psychological suffering of clients and to help them live a life of value. ACT is a short-term structured therapy that has been used with individuals and groups with a variety of psychological and medical problems. It provides exercises and assignments that encourage an individual to accept psychological pain as normal and indeed important but also to avoid increasing it unnecessarily. ACT takes issue with the premise of typical cognitive therapies that the direct targeting of problematic ways of thinking is helpful. Indeed, the proponents of ACT argue that too much attention to thinking is counterproductive and may even exacerbate problems. ACT distinguishes between language and experience and teaches clients mindfulness techniques that identify experience without the entanglement of language. Cognitive defusion is the term used for learning to regard thoughts as mere thoughts. ACT also helps clients to identify their core values, commit to them, and to take action to live in accordance with them. ACT has a Buddhist flavor acknowledged by its originator, but this therapy was derived from behavior therapy and an explicit psychological theory of thought and language. Of all the positive psychotherapies discussed here, ACT has been among the most frequently tested, and it is as effective in reducing psychological problems as conventional cognitive–behavioral therapies. Most

outcome research to date has used pathology measures, although some studies have also used the well-being measures of interest to positive psychology.

7. C. Rick Snyder's *hope therapy* uses cognitive–behavioral techniques to increase hope through a process of setting and monitoring explicit goals. It consists of formal and informal strategies for building and enhancing hope among children and adults. Versions of hope therapy exist for individuals and groups, and several studies demonstrate that by enhancing hope, hope therapy increases well-being and decreases symptoms of depression and anxiety.
8. *Well-being therapy* was developed by Giovanni Fava and is based on Ryff's model of psychological well-being. As described earlier, Ryff's model distinguishes several components of doing well and overlaps considerably with the positive psychology vision of psychological health. Well-being therapy is a short-term individual treatment that takes a thoroughly cognitive approach. In early sessions, clients are asked to identify and situate episodes of doing well. Then they are asked to keep a diary or journal about such episodes and to identify the thoughts and beliefs that interrupt them. Finally, they are taught to challenge and test these beliefs. Special attention is paid to the components of well-being that are frequently impaired, which makes well-being therapy more remedial than other positive psychotherapies. Well-being therapy is effective in reducing anxiety and depression, and it enhances well-being.
9. *Positive behavioral support* is a form of behavior therapy that focuses on the environment and specifically on problem contexts, defined as counterproductive and unfair situations that produce or exacerbate difficulties for people. Originally devised by Jacki Anderson, Edward Carr, Glen Dunlap, Robert Horner, Robert Koegel, Wayne Sailor, and Ann Turnbull to help people with disabilities, positive behavioral support has more general applicability. It attempts to impart to clients skills, coping strategies, and the motivation to change. It also enlists the support of family members and other key figures in the person's life. Positive behavioral support becomes an instance of positive psychotherapy when it expands its attention from what produces problems to what promotes happiness and life satisfaction. Like positive psychology per se, positive behavioral support is based on the premise that happiness is not simply the absence of conditions that produce problems.

10. *Quality-of-life therapy* was developed by Michael Frisch, who presented it as an instance of cognitive therapy. Its goal is to increase one's quality of life (happiness) by using the sorts of techniques developed by Aaron Beck and colleagues to decrease depression. Like other positive psychotherapies, quality-of-life therapy assumes that there are multiple routes to the good life, and it describes specific techniques for achieving these, from changing one's objective situation to revising one's standards for defining personal success. Quality-of-life therapy can be used with individual clients or groups. Different versions of quality-of-life therapy have been created for clinical versus nonclinical clients, and evidence is starting to accumulate that quality-of-life therapy is successful in its intended goal of increasing life satisfaction.

Taking Stock. The unique and explicit goal of emerging positive psychotherapies is to enhance well-being and promote the good life among those with obvious psychological problems as well as those without. They are also similar to more established therapies. Positive psychotherapies are short-term structured interventions for individuals or small groups. Most can be placed in the cognitive–behavioral realm, although their techniques could be integrated easily into other treatment models. Most positive psychotherapies entail out-of-session exercises and homework assignments, the results of which are discussed in sessions. A number of positive psychotherapies ask clients to keep journals, and many of these therapies rely on ongoing assessment.

Like other cognitive–behavioral interventions, positive psychotherapies take issue with assumptions of the medical model that people in treatment are ill and that their problems are best described as discrete (present-or-absent) entities *à la* DSM. According to positive psychology, people's weaknesses and strengths exist in degrees.

As emphasized, research support is still accumulating. Enough outcome studies have been conducted to conclude that positive psychotherapies are more than just promising, with effect sizes in the small to moderate range typical of psychological interventions. Not known in most cases is how positive psychotherapies fare in direct comparison to conventional treatments for anxiety or depression. As already mentioned, the boundary conditions of effective positive psychotherapy are unknown.

In the past, it was thought that positive psychology interventions were inherently light-handed, meaning that they are met with little resistance and that they are self-sustaining. Why would someone *not* want to be happier and more fulfilled, especially given the additional—presumably rewarding—benefits of so being?

Current thinking is that positive interventions can challenge both the client and the therapist. People do not always do the right or good (i.e., happy) thing, even when they know perfectly well what is to be done. Challenging one’s skills is more exhilarating than watching television. Spending time with close friends is more fulfilling than surfing the Internet. Doing volunteer work pays more dividends than reading the tabloids. Living in accord with explicit values is more satisfying than constant compromising. So why do people live as they do?

This question is among the most crucial for positive psychologists to answer, especially as they develop and deploy positive psychotherapies. The not-so-good life has considerable inertia. In the case of happiness, theorists have made sense of this inertia by positing a hedonic set-point. This set-point may reflect genetic constraints, cognitive styles like defensive pessimism, or habits sustained by given environments. Although recent research shows that the hedonic set-point can be changed by major life events, nothing suggests that change for the better is typical or easy.

CONCLUSIONS

Positive psychology goes beyond a focus on problems and their absence to reflect health and well-being. Positive psychology focuses on identifying and capitalizing on strengths and capacities. Evidence is accumulating that positive constructs such as positive emotions, good relationships, meaning and purpose, character strengths, and services for others play important roles in health and well-being, not only as broad protective factors, preventing or mitigating pathology and problems, but also as enabling conditions that promote thriving and flourishing life.

The good life can be taught and learned, but it requires hard work. Psychologists, at least in the United States, have long believed that the human condition can be improved by the intelligent application of what they have learned. Positive psychologists are no exception, and many have turned their attention to interventions that make people more happy, hopeful, virtuous, productive, and socially involved. In some cases, these applications are running in front of data that would support them, but in other cases, outcome research has been done. Even the most compelling research is not based on a follow-up that extends beyond a few years, and the research participants are usually motivated and willing volunteers. How well these interventions will generalize—across diverse people and over time—is, therefore, a research topic of high priority.

It is difficult to predict the future of positive psychotherapies except to say that their success will hinge on the relevant research. Although many therapy techniques informed by positive psychology seem promising, so far, only a few of them have been tested rigorously. More work needs to be done.

One point of view is that a thoroughly successful positive psychology will result in the fading away of the perspective, leaving a balanced psychology, one that recognizes the positive and the negative and of course their interplay. As emphasized, positive psychology does not deny the negative, and it may well be that what is most troubling in life can set the stage for what is most fulfilling. Complex emotional experiences often blend the positive and negative; optimism is most apparent when people confront setbacks and failures; crisis reveals strengths of character; ongoing challenge is a prerequisite to experience flow in the moment and to achieve something important in a lifetime; relationship success is foreshadowed not by the absence of problems but by how people resolve those that arise.

All people want to live happy and fulfilling lives. It is a fundamental human desire and right. Positive psychology provides a valuable perspective on the good life as well as strategies for assessment, treatment, prevention, maintenance, and promotion. Positive psychology expands the view of psychological health beyond the absence of symptoms and disorders and provides hope that a healthy, fulfilled, and productive life is possible for all.

FURTHER READINGS

Bryant FB, Veroff J. *Savoring: A New Model of Positive Experience*. Lawrence Erlbaum Associates Publishers; 2007.

Conklin LR, Strunk DR, Fazio RH. Attitude formation in depression: evidence for deficits in forming positive attitudes. *J Behav Ther Exp Psychiat*. 2009;40:120–126.

Gable SL, Reis HT, Impett EA, Asher ER. What do you do when things go right? The intrapersonal and interpersonal benefits of sharing positive events. *J Pers Soc Psychol*. 2004;87(2):228–245.

Hui BP, Ng JC, Berzaghi E, Cunningham-Amos LA, Kogan A. Rewards of kindness? A meta-analysis of the link between prosociality and well-being. *Psychol Bull*. 2020;146:1084–1116.

Jahoda M. *Current Concepts of Positive Mental Health*. Basic Books; 1958.

Jeste DV, Palmer BW, Rettew DC, Boardman S. Positive psychiatry: its time has come. *J Clin Psychiatry*. 2015;76:675–683.

*Joseph S, ed. *Positive Psychology in Practice: Promoting Human Flourishing in Work, Health, Education, and Everyday Life*. 2nd ed. John Wiley & Sons; 2015.

Park N, Peterson C. Methodological issues in positive psychology and the assessment of character strengths. In: Ong AD, van Dulmen M, eds. Oxford *Handbook of Methods in Positive Psychology*. Oxford University Press; 2006;292–305.

Park N, Peterson C. Positive psychology and character strengths: its application for strength-based school counseling. *J Prof Sch Couns*. 2008;12:85–92.

Park N, Peterson C, Szvarca D, Vander Molen RJ, Kim ES, Collon K. Positive psychology and physical health: Research and applications. *Am J lifestyle Med*. 2014;10:200–206.

*Peterson C. *A Primer in Positive Psychology*. Oxford University Press; 2006.

*Peterson C, Seligman MEP. *Character Strengths and Virtues: A Classification and Handbook*. American Psychological Association, Oxford University Press; 2004.

Quoidbach J, Mikolajczak M, Gross JJ. Positive interventions: an emotion regulation perspective. *Psychol Bull*. 2015;141:655–693.

Ryff CD, Heller AS, Schaefer SM, Van Reekum C, Davidson RJ. Purposeful engagement healthy aging, and the brain. *Curr Behav Neurosci Rep*. 2016;3(4):318–327.

Sansone RA, Sansone LA. Gratitude and well being: the benefits of appreciation. *Psychiatry (Edgmont)*. 2010;7(11):18–22.

Seligman ME, Csikszentmihalyi M. Positive psychology: an introduction. *Am Psychol*. 2002;55:5–14.

*Seligman ME, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. *Am Psychol*. 2005;60:410–421.

Shen M, Vu JV, Englesbe MJ. Using positive psychology to optimize patient-centered surgical care. *Ann Surg*. 2020;271:1018–1019.

*Snyder CR, Lopez SJ, Edwards LM, Marques SC. *The Oxford Handbook of Positive Psychology*. 3rd ed. Oxford University Press; 2021.

Song Y, Liu Y, Wang M, Lanaj K, Johnson RE, Shi J. A social mindfulness approach to understanding experienced customer mistreatment: a within-person field experiment. *Acad Manag J*. 2018;61:994–1020.

*Van Lange PA, Columbus S. Vitamin S: why is social contact, even with strangers, so important to well-being? *Curr Dir Psychol Sci*. 2021;30:267–273.

Werner-Seidler A, Tan L, Dalgleish T. The vicissitudes of positive autobiographical recollection as an emotion regulation strategy in depression. *Clin Psychol Sci*. 2017;5:26–36.

32.17 Acceptance and Commitment Therapy

Courtney Barry, Psy.D., M.S., Zabrina Ebert, D.O., and Joanna Petrides, M.B.S., Psy.D.

Acceptance and Commitment Therapy (ACT) is a therapeutic orientation rooted in behaviorism, functional contextualism, and relational frame theory (RFT). According to Hayes, the focus of ACT is to increase psychological flexibility through the utilization of various processes, including acceptance, cognitive defusion, and mindfulness. Within these processes, experiential exercises and metaphors are used to help one identify and engage in a healthy way with one's thoughts, feelings, and sensations. The ultimate goal is to assist in guiding one to be more in line with their values and to allow them to make value-driven behavioral changes, even when experiencing difficult thoughts or emotions. It is important to note that ACT is one of the "third wave" behavioral therapies, which focuses more on how context influences behaviors rather than eliminating behaviors completely.

FUNCTIONAL CONTEXTUALISM

Functional contextualism looks to understand one's thoughts, behaviors, emotions, and memories within a specific context to better recognize the functions they serve, the meaning they carry, and the effect they have on a person (moving toward or away from their values). This understanding ultimately allows for increased prediction and influence of behaviors toward one's values to create a more meaningful life. For example, a pen that ran out of ink may not serve its purpose in the context of writing. However, in a different context, such as using it to point to something, the pen may serve its function. In a clinical setting, a thought, emotion, or behavior may not be problematic. However, it provides insight into how thought, emotion, or behavior affects an individual in moving toward their values. ACT postulates that functional contextualism allows the clinician to see the individual as an individual and not as "damaged." This ultimately permits the clinician to change how an individual relates to their thoughts, feelings, and behaviors. When individuals become more aware of how their thought, feeling, or behavior works in the context of their life, they can make choices in relation to their values leading to more satisfying outcomes.

RELATIONAL FRAME THEORY

RFT is another approach that comprises the foundation of ACT. RFT builds on being able to identify language and derive relationships. At a young age, children learn what certain objects are and begin to label those objects with words. For instance, a child sees a picture of a dog in a book, and an adult tells the child, “That is a dog.” The child has learned the picture of a dog carries the sound D-O-G. When asked to point to a picture of a dog, the child points correctly and also makes the sound D-O-G. The child now derives relations based on what they learned. As children grow, they develop more relations based on their experiences and derive others. They connect the relations between events, as well as connect different functions to the events, which is RFT. In a clinical example, experiencing anxiety in one event and relating it to different events in one’s life evokes more anxiety across other events. It is important to understand how relationships between language and cognition have significant impacts on our lives despite simply being words.

EMPIRICAL EVIDENCE

Empirical evidence supports the effectiveness of ACT within a wide variety of contexts and conditions. ACT has been validated for various health and psychiatric disorders, including but not limited to chronic pain, depression, anxiety, stress, weight loss management, substance use, psychosis, quality of life, and smoking cessation. It has been shown to be effective in increasing psychological flexibility, thus, reducing depressive or anxious symptoms, which can be sustained for months following the intervention. In terms of patient populations, effectiveness has been supported in children, adolescents, adults, and individuals with cancer. ACT can be utilized in a wide variety of settings, such as inpatient, outpatient, group, and individual. Furthermore, ACT has been utilized with demonstrated effectiveness in various methods of delivery, including in-person as well as web-based and smartphone applications.

PROCESSES

The ultimate goal of ACT is to increase a person's ability to connect with the present moment and to persist in or change behaviors toward bringing them closer to their values. This process is known as psychological flexibility and is gained through the utilization of the six core processes discussed below. ACT uses a hexaflex (Fig. 32.17–1) to illustrate the six core processes. Each is interconnected, and together, they come together to increase an individual's psychological flexibility. The processes can be highlighted using certain skills and metaphors, as highlighted in Table 32.17–1. It must be noted that these processes are not intended to be used as methods of avoidance but rather as positive skills.

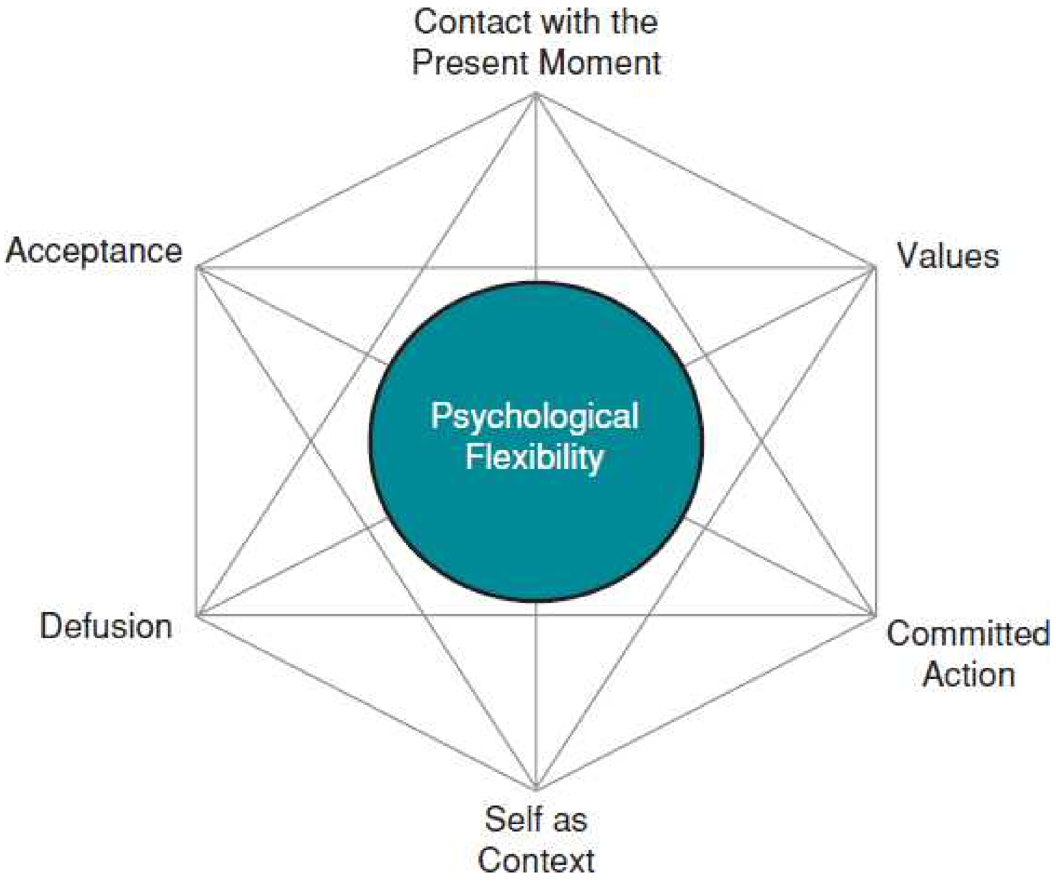
Contact with the Present Moment
Acceptance
Values
Psychological Flexibility
Defusion
Committed Action
Self as Context

FIGURE 32.17–1. ACT hexaflex. (Copyright Steven C. Hayes, Ph.D., used with permission.)

Acceptance

Acceptance is the process in which a person learns to actively embrace internal experiences without working to avoid or change them. In other words, acceptance allows a person to engage with thoughts and feelings as they are, regardless of whether they are pleasant or painful. Because it is defined as an act, a person must be willing to engage in acceptance through action, which ultimately requires a decision to embrace acceptance. It is important to emphasize that acceptance and willingness are not the same as wanting. For example, a person does not have to want their physical/emotional pain to be present, but they must work toward accepting their pain will continue to exist as they engage in actions to live life more fully and with less pain. This process ultimately allows a person to move toward their values without engaging in experiential avoidance.

Cognitive Defusion

Cognitive defusion is a process in which a person works to alter their interactions with their own thoughts. Rather than working to alter the content, function, or frequency of distressing thoughts, a person learns to observe thoughts from a more neutral position. In essence, a person learns to notice thoughts passively over getting caught up in them and believing the content. The final result is a person decreases “buying-in” to their thoughts, leading to a reduction in emotional response and, ultimately, being able to more readily move on from the thought itself. Because thoughts often drive behavior, cognitive defusion is an important step to behavioral change. As an example, a person may continually have self-deprecating thoughts such as “I am a bad person” or “I am no good” and think because they had the thought, it must be true. Through cognitive defusion, a person learns to view the thought as just a thought without reacting behaviorally to address the thought or attaching their identity to what the thought is saying. By doing so, individuals let thoughts flow through their minds with no further examination and without assigning meaning, thus, allowing their life to continue.

Table 32.17–1.
ACT Processes and Examples

Process	Brief Definition	Theoretical Clinical Examples (All case examples are fictional)
Acceptance	Actively embracing internal experiences without working to avoid or change them.	A young adult with a history of PTSD on escitalopram and prazosin presenting with continued anxiousness related to having previously experienced a traumatic event. Feels medications are helpful but unable to move past the frustration that this happened to him. *Brief Intervention:* Can focus on ideas of acceptance by addressing the idea that willingness to accept trauma in order to live more fully is not the same thing as having wanted trauma to happen.
Cognitive Defusion	Rather than working to alter the content, function, or frequency of distressing thoughts, a person learns to observe thoughts from a more neutral position, i.e., a thought is just a thought until one assigns meaning to it.	An adult patient with a history of MDD with psychotic features on fluoxetine presenting with distress secondary to having negative thoughts about others. Expresses the belief that this must mean hallucinations have returned because there is no way they could think negative thoughts about others. Does note these thoughts to be internal and in their own voice. No other psychotic or mood symptoms noted. *Brief Intervention:* Discussion surrounding the idea that a thought can be just a thought and doesn’t necessarily have to have

		a deeper meaning can increase the resolution of distress.
Being Present	Working toward living in the here and now to promote increased awareness of oneself, others, and experiences.	An adolescent with a history of social anxiety disorder presents with decreased social interaction in the context of being hyper-focused on past social interactions. *Brief Intervention:* Exploration of various mindfulness techniques can improve contact with the present moment and minimize the over-evaluation of past interactions.
Self-as-Context	Aims to have a person identify themselves as the observer; a person is the one observing what is happening, not the observation itself.	An adult patient with a history of OCD on clomipramine presented with a worsening mood secondary to ruminating thoughts. Expresses distress with an inability to silence their thoughts and feels their thoughts define them as a person. *Brief Intervention:* Have a patient listen to their thoughts and then discuss with them how they were able to observe their thoughts. They were, in fact, the outside perspective.
Values	Identification of the person one wants to be, what one wants to stand for, and what will ultimately make life both fulfilling and meaningful.	A geriatric patient with a history of depression, unspecified, on venlafaxine presents with family with concerns of anhedonia and low energy. When discussing with the patient, they express feeling useless and without purpose since they retired a few years ago. Prior to retirement, they felt

		they had been doing well on their medication regimen. *Brief Intervention:* Discussions about the different stages of life and the idea that despite this change, there is still room to work toward values to provide purpose and meaning that can improve the sense of satisfaction and fulfillment.
Committed Action	Commitment to making a change. Guided by identified values and accomplished through identification of goals.	A child patient with no past psychiatric history presents for evaluation of insomnia. Various behavioral barriers to proper sleep, including screen time before bed, varying bedtimes, etc., were noted. No acute mood or anxious symptoms were elicited. *Brief Intervention:* Create an action plan for proper sleep hygiene, including the identification of an ultimate goal (better sleep) and numerous small goals (reduce screen time, have a consistent bedtime, etc.) to reach the ultimate goal.

Being Present

Preoccupation with the past and/or future is a common pattern in many people. This preoccupation usually results in a judgmental evaluation of themselves, their environment, and their actions, as well as a disconnection from relationships and events. All too often, it is common to dwell on the past or be consumed with worries about the future. When engaging in the processes of being present, a person works toward living in the here and now to promote increased awareness of themselves, others, and experiences. This ultimately allows them to more readily utilize behaviors consistent with their values and achieve increased fulfillment from the present context. Commonly, this process is achieved through the use of mindfulness exercises ranging widely from breathing techniques to meditation and even mindful use of one’s senses.

Self-as-Context

One’s sense of self is often grounded in self-as-content, which is based in language. Self-as-content is the label one assigns to the self to define who someone is. For example, a person may define themselves based on their position in their family (parent, child, etc.), profession (businessperson, health care provider, athlete, etc.), the circles they exist within (religiously, economically), and even based on who they are not (the opposing groups). Self-as-context, on the other hand, aims to have a person remove language from the algorithm altogether and identify themselves as the observer. In other words, a person is the one observing what is happening, not the observation itself. For example, a person may experience various thoughts about themselves, including “I am successful” or “I am a daughter,” and believe this is who they are. However, in actuality, they are the ones having the thoughts and not being the thought themselves. Therefore the language becomes “I’m having the thought ‘I am successful.’”

Values

ACT embodies the importance of a person identifying who they truly want to be, what they want to stand for, and what will ultimately make their life both fulfilling and meaningful. In simpler terms, a person must identify their values. Given the nature of the definition, values are individualized and cannot be assumed based on one’s culture, religion, or other identifying groups. Values provide a person with a sense of direction and will guide decision-making and behaviors taken when clearly defined. It is important to note that values are not tangible, and they cannot be finite; they are an ongoing influence on what motivates someone to engage in the way they do. The other ACT processes (especially acceptance, cognitive defusion, being present, and self-as-context) work to increase value-driven behaviors and actions.

Committed Action

After identifying values, one must commit to the behavioral changes that align with those values. Phrased another way, values guide the behavior change, and committed action is practical steps taken toward making a change. While values and goals are different (goals can be checked off when accomplished, whereas values are ongoing), goals can help a person move toward their values. The utilization of SMART goals, in particular, can be of great use to a person as it helps them narrow down how exactly to initiate a change. The SMART mnemonic stands for Specific, Measurable, Attainable, Relevant, and Time-based. With each accomplishment of a SMART goal, another can be created, thus, always driving a person closer to their values.

The changing of behaviors is not automatic and by continuing to utilize the processes above, the individual can move toward a value-driven life.

FURTHER READINGS

Bach P, Hayes SC. The use of acceptance and commitment therapy to prevent the rehospitalization of psychotic patients: a randomized controlled trial. *J Consult Clin Psychol*. 2002;70(5):1129–1139.

Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. *JAMA Intern Med*. 2020;180:1472–1480.

Dindo L, Van Liew JR, Arch JJ. Acceptance and commitment therapy: a transdiagnostic behavioral intervention for mental health and medical conditions. *Neurotherapeutics*. 2017;14(3):546–553.

Doran GT. There’s a S.M.A.R.T. way to write management’s goals and objectives. *Manage Rev*. 1981;70(11):35–36.

Hancock KM, Swain J, Hainsworth CJ, Dixon AL, Koo S, Munro K. Acceptance and commitment therapy versus cognitive behavior therapy for children with anxiety: Outcomes of a randomized controlled trial. *J Clin Child Adolesc Psychol*. 2018;47:296–311.

Harris R. *Act Made Simple: An Easy-to-Read Primer on Acceptance and Commitment Therapy*. New Harbinger Publications; 2019.

Hayes SC. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies—republished article. *Behav Ther*. 2016;47(6):869–885.

Hayes SC, Levin ME, Plumb-Vilardaga J, Villatte JL, Pistorello J. Acceptance and commitment therapy and contextual behavioral science: examining the progress of a distinctive model of behavioral and cognitive therapy. *Behav Ther*. 2013;44(2):180–198.

Herbert MS, Afari N, Liu L, et al. Telehealth versus in-person acceptance and commitment therapy for chronic pain: a randomized noninferiority trial. *J Pain*. 2017;18:200–211.

Khoramnia S, Bavafa A, Jaberghaderi N, et al. The effectiveness of acceptance and commitment therapy for social anxiety disorder: a randomized clinical trial. *Trends Psychiatry Psychother*. 2020;42(1):30–38.

Lanza PV, García PF, Lamelas FR, González-Menéndez A. Acceptance and commitment therapy versus cognitive behavioural therapy in the treatment of substance use disorder with incarcerated women. *J Clin Psychol*. 2014;70:644–657.

Lillis J, Niemeier HM, Thomas JG, et al. A randomized trial of an acceptance-based behavioral intervention for weight loss in people with high internal disinhibition. *Obes*. 2016;24:2509–2514.

Mak YW, Leung DYP, Loke AY. Effectiveness of an individual acceptance and commitment therapy for smoking cessation, delivered face-to-face and by telephone to adults recruited in primary health care settings: a randomized controlled trial. *BMC Public Health*. 2020;20:1719.

Mohammadi Masiri F, Taher M, Moradi S, Motaghedifard M. Effectiveness of acceptance and commitment therapy on anxiety and stress in mothers of children with attention deficit hyperactivity disorder. *J Child Adolesc Ment Health*. 2020;7(2):312–321.

Mohammadi S, Soufi A. *J Health Psychol*. 2020;8(32):57–72.

Mohsenzadeh F, Yarahmadi H, Zahrakar K, Soleimani A. Effectiveness of group therapy based on acceptance and commitment therapy on quality of life in HIV patient. *J Psychol Sci*. 2019;18(80):923–932.

Viskovich S, Pakenham KI. Randomized controlled trial of a web-based acceptance and commitment therapy (ACT) program to promote mental health in university students. *J Clin Psychol*, 2020;76(6):929–951.

Zemestani M, Mozaffari S. Acceptance and commitment therapy for the treatment of depression in persons with physical disability: A randomized controlled trial. *Clin Rehabil*. 2020;34(7):938–947.

32.18 Hallucinogen-Assisted Psychotherapy

ERIKA STEINBRENNER, M.D.

DEFINITION

This section provides an overview of the known literature on the clinical use of hallucinogens as adjuncts to psychotherapy. Most research relates specifically to psychedelics or *serotonergic* hallucinogens (i.e., those hallucinogens that exert consciousness-altering effects primarily through agonist or partial agonist activity at serotonin 2A [5-HT$_{2A}$] receptors).

Ketamine has often been classified as a psychedelic, although it is chemically distinct and profoundly different. It will be addressed at the end of this section.

The past decade has seen a resurgence of interest in the clinical applications of psychedelics such as psilocybin, lysergic acid diethylamide (LSD), and 3,4-methylenedioxymethamphetamine (MDMA). Although psychedelics have been used by humans for millennia, known to science for over a century, and subjected to extensive basic research for years, the rigorous study of their clinical use as adjuncts to psychotherapy is in its infancy. Results from clinical trials have been auspicious, although additional research is needed. If the data ultimately support treatment efficacy, psychedelic-assisted psychotherapy would genuinely constitute a novel treatment paradigm in psychiatry. Many believe we are on the verge of a paradigm shift that will fundamentally change the field of mental health. A push is underway to legalize these substances, and leading researchers are gathering evidence of their safety and efficacy in various patient populations.

PSYCHOPHARMACOLOGY

Hallucinogens are a highly diverse group of substances that can alter and distort perception, thought, and mood. Hallucinogens are categorized by their chemical structure and include tryptamines, phenethylamines, dissociative anesthetics like ketamine and phencyclidine (PCP) (which interact with the glutamate system), anticholinergic hallucinogens, and cannabinoids. In addition, there are also several unclassified hallucinogens, such as kava kava, kratom, nutmeg, and Salvia. Psychedelics refer specifically to two classes: variations on the structure of tryptamine (i.e., psilocybin, LSD, *N,N*-dimethyltryptamine [DMT], ibogaine) and substituted phenethylamines (i.e., mescaline and MDMA). What they share is a common mechanism of hallucinogenic action: agonism or partial agonism at the 5-HT$_{2A}$ receptor. Psilocybin, DMT, LSD, and mescaline are considered the “classic psychedelics.” Each psychedelic has unique properties and understanding the basics of each is vital for potential therapeutic use.

Psilocybin

Psilocybin is a psychoactive substance found in over 200 species of mushrooms. Psilocybin is a prodrug and is rapidly dephosphorylated to the psychoactive metabolite psilocin in the intestinal mucosa and liver. The concentration of psilocybin and psilocin varies significantly between mushroom species. Effects first become apparent after approximately 20 to 45 minutes, peak between 60 and 90 minutes, then gradually come in progressively less intense waves for a total of about 6 hours. In terms of comparative potency, psilocybin is approximately 45 times less potent than LSD and 66 times more potent than mescaline.

LSD

LSD is a synthetic substance first synthesized in 1938 by Albert Hofmann. The onset of effects occurs around 30 minutes, with peak effects generally reached in 2 to 3 hours. Peak effects generally hold for another 2 to 3 hours and slowly dissipate for up to about 12 hours.

DMT

DMT is a psychoactive compound found naturally in a variety of plants and animals and endogenously produced in humans. When consumed orally, DMT is rapidly metabolized by the enzyme monoamine oxidase A (MAO-A) in the GI tract, thus making it inactive. Therefore, it is typically smoked or vaporized. This form of ingestion produces a potent hallucinogenic state, considered one of the most intense psychedelic experiences in existence. DMT is exceptionally fast-acting when consumed via inhalation, with peak subjective experiences occurring around 2 minutes after inhalation and completely resolving within 15 to 20 minutes. Due to its endogenous production in humans, DMT has been theorized to be associated with several nonordinary states of consciousness (i.e., dreaming, spiritual experiences, encounters with nonhuman intelligence, out-of-body experiences, near-death experiences).

Ayahuasca

Ayahuasca is a very complex brew in which plants containing DMT are combined with plants containing MAO-I’s (β-carboline alkaloids), which increase the oral bioavailability of DMT. Besides protecting DMT from degradation, the β-carboline alkaloids also have their own psychoactive properties. Ayahuasca's effects begin to be felt between 20 to 40 minutes after ingestion, peak between 1 to 2 hours, and subside gradually within about 6 hours.

Ibogaine

Ibogaine is a natural compound derived from the root bark of a West African shrub, *Tabernanthe iboga Baill.* Ibogaine is pharmacologically distinct from the classic psychedelics: it simultaneously interacts with several neurotransmitter systems, with the highest affinity for the σ-2 receptor. It is commonly used in several other countries for opioid detoxification, although the mechanism of action for its efficacy in this setting is unknown as clinical observations suggest it is not acting as a mu-opioid agonist. The ibogaine experience is extremely long, lasting anywhere from 18 to 30 hours.

Mescaline

Mescaline is a psychoactive compound found in several species of cactus, most notably peyote. Mescaline binds the 5-HT$_{2A}$ receptor at a much lower affinity than LSD, perhaps explaining the much higher doses needed for mescaline to have similar effects. Mescaline does not readily cross the blood–brain barrier, contributing to its low potency and substantial physical side effects, especially nausea and vomiting. These adverse effects are posited to account for the eclipse of mescaline research by LSD in the 1950s. Effects are generally felt within 45 to 90 minutes of consumption, peak at 3 to 6 hours, and last for 10 to 12 hours.

MDMA

3,4-Methylenedioxymethamphetamine (MDMA) is often referred to as a psychedelic, but it is quite different in many ways from classic psychedelics such as LSD or psilocybin; it produces mood elevation and an increased sense of empathy but the ego remains relatively intact. Therefore, it has been proposed that MDMA is actually an entactogen or empathogen, a distinct class of drugs. For organization and clarity, MDMA will be addressed in a separate section from the classic psychedelics.

SUBJECTIVE EFFECTS

The range of possible subjective experiences under classic psychedelics is extensive and impossible to characterize fully. Effects occur along a dose-dependent range and vary between psychedelic substances. An individual's experience is also unique from one experience to another and is highly influenced by the extra pharmacologic factors of *set* (psychological frame of mind, intention) and *setting* (environment in which use takes place). Very broadly, experiences can be generalized along several dimensions: (1) perceptual distortions (i.e., visual and auditory illusions or hallucinations, synesthesia), (2) shifts in cognition (i.e., tangential thought processes, dissociative phenomena), (3) mood fluctuations (i.e., euphoria, sadness, anger), (4) the dysphoric effects of ego dissolution (i.e., anxiety, panic, fear of losing control), and (5) mystical experiences (dissolution from self, oneness with the universe, transcendence of space and time).

PROPOSED MECHANISM OF PSYCHEDELICS

Psychedelic drugs are fundamentally different from all other psychotropic medications. They are not inherently therapeutic themselves; rather, the benefits accrued by users are dependent on the integration of enduring memories and insights of the particular states of consciousness that were experienced during the period of drug action. That is why those who encounter profound states of mind typically express little interest in repetitive ingestions in the near future.

Although neuroimaging studies have begun to characterize the acute effects of psychedelics on brain function, how these changes neurobiologically mediate therapeutic benefit remains poorly understood. One of the most consistent observations on imaging is the disintegration of normal brain network communications after ingestion of a psychedelic substance. One key network that is disrupted is the default mode network (DMN), which is involved in introspection and sense of self; in many ways it represents the seat of what Freud called "the ego." The DMN becomes more chaotic and statistically random, and thoughts, perceptions, and memories typically inaccessible to awareness have a much higher probability of activation. During a psychedelic experience, the segregation between the DMN and related cortical networks is also lost, and new between-network communication patterns are formed. Thus, psychedelics lead to an overall brain dynamic characterized by a less-constrained consciousness, an increased degree of entropy, and the opportunity to experience a broader spectrum of cognitive insights and emotional breakthroughs.

A common characteristic of a lot of mental health disorders is that they can be defined by an entrenchment into a rigid and “stuck” way of thinking and behaving (i.e., a negative cognitive bias in depression, excessive rumination about some type of threat in anxieties, excessive control, and specificity of whatever the object of relief is in addiction, repeated behaviors in obsessive-compulsive disorder (OCD), repeating thoughts and ways of looking at oneself in eating disorders). It may be that psychedelics temporarily “shake up” the system and relax these overweighted habits of mind and behavior, which then provides the opportunity for recalibration into a healthier form once things resettle.

At cellular levels, psychedelics have been found to enhance processes associated with neuroplasticity, such as the growth of dendritic spines. They increase the expression of genes associated with brain-derived neurotrophic factor (BDNF) and other mediators of synaptic plasticity. Additionally, research has discovered that psychedelics possess incredibly potent anti-inflammatory properties at very low, sub-behavioral doses, although whether this ability contributes to its mental health effects is not yet known.

HISTORY

Early Academic Research With Psychedelics

Indigenous cultures have used psychedelics for religious and healing purposes for more than 10,000 years. However, Western scientific study of hallucinogens dates to the late 1800s when Arthur Heffter isolated mescaline. Initial interest focused on the possibility that mescaline could produce insight into the basis of psychotic disorders and serve as an experimental model for schizophrenia. In 1943, Albert Hofmann discovered the psychoactive effects of LSD when he accidentally ingested traces of an ergot derivative he had synthesized. The therapeutic potential of LSD was quickly recognized; it was marketed under the name Delysid and made widely available to psychiatrists worldwide, mainly as an adjunct for psychotherapy. Between 1950 and 1965, some 40,000 patients had been prescribed one form or another of LSD-assisted therapy as a treatment for neurosis and mood disorders, alcoholism and other addictions, and psychosis (schizophrenia).

Research investigating psychedelics grew as well. By 1961, over 1,000 scientific papers on psychedelics (mostly LSD) had been published. Results strongly suggested therapeutic benefits for psychedelic-assisted psychotherapy, with data strongest for alcohol use disorder and cancer-related psychological and existential distress. Unfortunately, early investigators did not employ rigorous clinical methods such as randomized trials, specified outcome measures, and controlled treatment settings, and thus these studies did not provide definitive results. Even so, the preliminary findings seemed to warrant further research into the therapeutic benefits of psychedelic drugs.

However, clinical research on psychedelics came to a halt in the early 1970s, in large part due to sociopolitical issues. LSD, with its ability to completely reframe one's relationship to reality (at least temporarily), contributed to the emergence of a counter-culture in which one was not only permitted but expected to question every aspect of social norms, from politics to religion to morality. In one of the most famous slogans of the 1960s, Timothy Leary popularized LSD as a way to "turn on, tune in, drop out." This was occurring against the backdrop of the Vietnam War and growing opposition to the perceived military–corporate control of society in general. The backlash was fierce. The Controlled Substances Act of 1970 placed LSD, psilocybin, and related compounds in Schedule I, and human research was not allowed to continue. The United States applied political and economic pressure on other U.N. members so that a global ban was fully in effect by 1971. LSD, despite more than two decades of scientific investigation pointing to an array of likely therapeutic benefits, was stigmatized as the most dangerous drug in the world. Decades of positive work with psychedelics were effectively discounted.

Reemergence of Research Using Psychedelics

After a hiatus of two decades, human research on psychedelics resumed in the early 1990s with Rick Strassman's exploration of the subjective and physiologic effects of intravenous DMT on normal volunteers. Researchers began conducting basic science studies, and a considerable body of work has characterized the acute effects of classic psychedelics on physiology, cognition, emotion, and brain function. In turn, this led to a renewed interest in psychedelics' potential therapeutic value for mental health. Clinical trials have resumed, investigating the effects of psychedelics (primarily psilocybin) in the treatment of a number of psychiatric illnesses and addictive disorders. The studies that have been completed thus far are primarily open-label or randomized controlled trials (RCTs) with small sample sizes, but the results have been extremely promising. The FDA granted breakthrough therapy designation to psilocybin for treatment-resistant depression twice: in 2018 (COMPASS Pathways) and 2019 (Usona Institute). A breakthrough therapy designation is meant to accelerate the typically sluggish process of research and development for drugs that treat a serious condition and for which preliminary clinical evidence indicates may demonstrate substantial improvement over available therapies. At the time of writing, there are over 200 ongoing or planned research studies involving psychedelics. Suffice it to say we are just beginning to explore the potential possibilities and uses for psychedelics.

PRINCIPLES AND GOALS OF THERAPY

Two models of therapy were used in the 1950s through the early 1970s: psycholytic and psychedelic therapy. Although both combined administration of hallucinogens with psychotherapy to achieve therapeutic change, they emphasized different processes in bringing about these changes.

Psycholytic Therapy

In the psycholytic method, low-to-moderate doses of hallucinogens are administered on multiple occasions to facilitate therapy based on traditional psychoanalytic principles (helping the patient become aware of unconscious desires, emotions, attachments, self-representations, and intrapsychic conflicts).

Psychedelic Therapy

Psychedelic therapy was developed based on the observation of sudden and dramatic clinical improvements associated with high-dose hallucinogen administration. Quantum shifts in perception and cognition were noted in patients who typically experienced “peak” experiences marked by highly salient spiritual, religious, or mystical phenomena. One of the main goals of psychedelic therapy is to create an optimal treatment environment during the preparatory psychotherapy and in the dosing session to foster such “peak” experiences. Contemporary research in the 21st century has used a variant of the psychedelic approach. Although the details of the protocol used in clinical trials have varied, several core principles are shared.

Preparatory Sessions. There are typically several preparatory sessions where participants discuss their intentions, expectations, and concerns and talk about meaningful aspects of their lives. These sessions serve a twofold purpose: First, the discussion helps build trust and rapport with the facilitators and reduces the likelihood of paranoia during the trip. Secondly, related personal material may emerge under the effects of psychedelics; by knowing details about the participant, facilitators can better understand the participant’s experience and how best to provide support should difficult emotions arise.

Dosing Sessions. During a psychedelic session, participants are typically invited to lie down with eyeshades and listen to carefully selected music through headphones. The most common issue in psychedelic sessions is resistance to the emerging unconscious material and an unwillingness to “go with the experience.” For this reason, conversation between participants and facilitators is best kept to a minimum, especially during the early, most intense hours. Therapeutic activity is primarily oriented toward assisting the participant in focusing attention inward and facing and fully experiencing whatever emerges without avoiding anything. Facilitators periodically check in with the participant and offer support and reassurance if fear or distressing emotions arise, but mostly the journey is uninterrupted. Occasionally, therapeutic touch is appropriate and can provide stability and connection. Length of sessions varies depending on substance (i.e., typically 4 to 6 hours for psilocybin).

Integration Sessions. Integration sessions follow in the days or weeks after a psychedelic session and are a time for participants to process their experience and sort out new insights and discoveries. Old maladaptive defenses are discarded, and new ways of experiencing, looking at, or reacting to events are explored. These integration sessions are a crucial component for optimizing benefits from the psychedelic experience.

CLINICAL ISSUES

Indications

Psychedelic-assisted psychotherapy has historically been applied to a wide variety of psychiatric conditions. In the past decade since contemporary research has resumed, the most studied indications (and the ones with the most robust data from Phase II trials) are the use of psychedelic-assisted therapy for cancer-related distress and treatment-resistant depression. There is also encouraging research for its use in addiction and OCD, and trials are exploring its potential for several other indications (see Clinical Research section).

Limitations

There are several limitations to providing psychedelic-assisted psychotherapy. First and foremost, all psychedelics are currently Schedule I drugs and illegal in most states; thus they are largely not available for clinical use. If psychedelics are rescheduled and available as adjuncts to psychotherapy, there are several other obstacles to their broad use among mental health practitioners. The long duration of psychoactive effects (i.e., approximately 6 hours for psilocybin) necessitates daylong treatment sessions, which increases the clinical and infrastructure resources required and would thus likely result in high costs associated with treatment. Additionally, it has been shown that chronic use of SSRIs and other medications that bind the 5-HT_{2A} receptor (i.e., tricyclics, buspirone, and antipsychotics [chlorpromazine, risperidone]) reduce the subjective effects of psychedelics by 40% to 99%. Therefore, treating medication-heavy patients with psychedelics will be especially challenging, and a period of destabilization from medication discontinuation will likely be required before receiving psychedelic treatment.

Complications

Physiologic Toxicity. The safety of classic psychedelics (particularly psilocybin) in clinical research settings has been well-documented. In general, classic psychedelics possess remarkably low physiologic toxicity and are not associated with end-organ damage, carcinogenicity, teratogenicity, lasting neuropsychological deficits, or overdose fatalities.

They do produce sympathomimetic effects and can moderately increase pulse and blood pressure, although this has not been associated with cardiac, neurologic, or other organ damage. They do not result in overdose deaths after reasonable dosages. Classic psychedelics are not typically considered drugs of addiction, as they are not capable of producing sufficient reinforcing effects to produce addiction.

Psychological Toxicity. Acute, adverse psychological experiences ("bad trips") are the most significant concern in psychedelic administration and typically include anxiety, panic, dysphoria, paranoia, agitation, and fear of losing one's mind. These unpleasant experiences are more likely to occur in recreational settings involving poorly prepared individuals and those who have psychological risk factors (i.e., severe mental illness, recent trauma). Despite such adverse reactions, users usually retain insight into the fact that their symptoms are related to drug ingestion, and they typically respond to verbal reassurance. Classic hallucinogens can acutely engender frank psychosis marked by hallucinations, thought disorders, and delusions, although this is rare in individuals without underlying psychotic spectrum illness. Hallucinogen-persisting perception disorder (HPPD) is an uncommon condition that can occur sometime after the cessation of a psychedelic or other hallucinogen. It is more likely to occur in individuals who chronically use hallucinogens or who concurrently use several substances (i.e., cannabis, MDMA, alcohol).

Contraindications

Given the known link between psychedelics and their ability to initiate or exacerbate psychosis in patients with underlying psychotic spectrum illness, they are contraindicated in anyone with a significant history of psychosis, whether it be from a primary psychotic spectrum illness (i.e., schizophrenia, schizoaffective disorder), affective psychotic illnesses (i.e., Bipolar I disorder with psychotic features, major depression with psychotic features), or severe personality disorders associated with psychosis. Furthermore, they are contraindicated in patients at high risk for psychosis (i.e., strong family history of psychosis). All contemporary research trials have excluded participants with bipolar disorder, given the risk of triggering mania; however, a current study is investigating the safety and efficacy of psilocybin for Bipolar II depression.

CLINICAL RESEARCH

Use of Psychedelics for Treatment-Resistant Depression

Psilocybin. The use of psilocybin-assisted therapy for major depressive disorder (MDD) has been one of the largest areas of recent research. Several trials have been published in the last few years with encouraging results. Two consecutive psilocybin-assisted therapy sessions (spaced 1 to 3 weeks apart) were shown to significantly decrease depressive symptoms in an open-label setting or when compared to a waiting list group. Three randomized, placebo-controlled trials have investigated the effects of a single psilocybin-assisted therapy session (along with preparatory and integration sessions) in MDD. In these trials (one with inactive and two with active controls), psilocybin-assisted therapy produced rapid and significant decreases in depression symptomatology. Effects persisted to study endpoints, although notably these follow-up periods were limited (21 to 43 days). There was some variability in study design between the trials, but all contained a moderate dose of psilocybin and psychological support during the psilocybin intervention.

Psilocybin-assisted therapy has also shown noninferiority to a commonly prescribed antidepressant. In a study by Carhart-Harris, 59 patients were randomized to receive either two separate doses of psilocybin (25 mg; 3 weeks apart) + daily placebo (psilocybin group) or two separate control doses of psilocybin (1 mg; 3 weeks apart) + daily escitalopram (escitalopram group). All patients participated in a preparatory session and integration sessions following the two dose days. Results indicated that psilocybin-assisted therapy was as effective as escitalopram in reducing depressive symptoms at the 6-week endpoint. Of note, psilocybin-assisted therapy outperformed escitalopram on a number of secondary endpoints, including measures of well-being, anhedonia, work and social functioning, and suicidality. At the time of writing, there are multiple additional trials underway, including the first Phase III trial investigating psilocybin-assisted therapy for MDD.

Ayahuasca. Contemporary clinical trials with ayahuasca suggest an antidepressant effect, particularly in the subacute period. An open-label, within-subject trial investigated the effects of a single dose of ayahuasca on six subjects with varying degrees of depression. Ayahuasca was well-tolerated by all participants, and there were significant reductions in depressive symptoms at 21 days. The same team conducted another study in a larger sample (*N* = 17) and found that the results held. The first double-blind RCT for ayahuasca was published in 2018. Twenty-nine patients with treatment-resistant depression received a single dose of either ayahuasca or a placebo. Significant antidepressant effects were observed at all-time points following administration compared to baseline (1, 2, and 7 days after dosing). Notably, the effect size increased with time (from postdose day 1 to day 7).

DMT. There is much interest in whether DMT could be useful for treating mental health problems, given that it has a quick onset and short duration of action, which may provide a more cost-effective approach to treatment. There are currently no published studies investigating DMT for clinical use. Preliminary trials are underway in several academic medical centers investigating the safety and efficacy of specific doses of intravenous DMT in humans, as well as for the treatment of depression.

LSD. LSD was extensively investigated in humans in the 1950s and 1960s and demonstrated promise for attenuating depressive symptoms, but no contemporary studies have been conducted on this subject. A Phase II trial investigating the efficacy of LSD for patients with major depression is in progress.

Use of Psychedelics for Psychological and Existential Distress in Patients With Life-Threatening Illness

LSD. The first clinical study of LSD since the 1970s was conducted in 2014. This randomized, double-blind, placebo-controlled trial with a crossover design (2 months) included 12 participants with life-threatening medical disorders. Participants received two treatment sessions (2 to 3 weeks apart) with either LSD or active control in conjunction with psychotherapy. The experimental group had significant short-term (2-month follow-up) reductions in anxiety compared to the control group. A follow-up study found sustained reductions in anxiety at 12 months, and no reports of any adverse psychological or medical sequelae. A recent trial (*n* = 42) corroborated and expanded on those findings. Using a similar design, it found that two LSD-assisted treatment sessions (2 weeks apart) significantly reduced anxiety in patients both with and without life-threatening diseases for up to 16 weeks.

Psilocybin. Three contemporary research studies have been published investigating psilocybin to treat cancer-related psychiatric distress. A double-blind, randomized, placebo-controlled trial at UCLA (*N* = 12) suggested an acute and short-term sustained antidepressant and anxiolytic effect associated with a single low dose of psilocybin versus active placebo (niacin). Two other trials at the NYU Langone Medical Center (NYU) and Johns Hopkins University Medical Center used a similar treatment model, although varied on the dose of psilocybin and the comparator. Both studies had significant reductions in anxiety and depression at short-term follow-up.

A long-term follow-up of patients from the NYU study found that reductions in anxiety, depression, hopelessness, demoralization, and death anxiety were sustained at 3.2 and 4.5 years after the original psilocybin dosing date. At the second (4.5 years) follow-up, approximately 60% to 80% of participants met the criteria for clinically significant antidepressant or anxiolytic responses. Participants overwhelmingly (71% to 100%) attributed positive life changes to the psilocybin-assisted therapy experience and rated it among the most personally meaningful and spiritually significant experiences of their lives. A Phase II trial investigating psilocybin for depression in cancer patients is currently being performed.

Use of Psychedelics in the Treatment of Addiction

Alcohol Use Disorder

LSD. The use of LSD in treating alcohol addiction was studied extensively from the 1950s through the early 1970s. Uncontrolled trials had variable but generally encouraging results. A recent meta-analysis looking at the six published randomized trials of LSD for alcohol dependence (1943–2010) revealed consistent and clinically meaningful effects of LSD over control treatment. LSD was administered in a single high-dose session in all the studies, in doses ranging from about 210 to 800 mcg. At the first posttreatment follow-up, LSD-treated patients were more likely to show significant improvement in drinking outcomes. Treatment effects remained significant at 6 months. The effect of LSD treatment was homogeneous across the six studies, despite significant variability in the psychotherapeutic components of treatment and the control treatments employed. These promising findings strongly support renewed clinical investigation of LSD for the treatment of alcohol use disorder.

PSILOCYBIN. The first double-blind study investigating psilocybin-assisted therapy for alcohol use disorder was recently completed, and demonstrated positive results. Ninety-five participants engaged in 12 weeks of manualized therapy, and were randomly assigned to receive high-dose psilocybin or active placebo (diphenhydramine) during 2-day-long medication sessions at weeks 4 and 8. The therapy included motivational interviewing, cognitive behavioral therapy, and material for managing and integrating the psychoactive effects of the study medication. Results found that the psilocybin group had significantly reduced number of heavy drinking days in the 32-week follow-up period (10% vs. 24% in the placebo group). This supports findings from an earlier open-label pilot study. At the time of writing, six other trials investigating psilocybin for alcohol addiction are in progress.

IBOGAINE. There have not been any published studies on ibogaine in the treatment of alcohol use disorder. However, a Phase II trial investigating ibogaine in the treatment of alcohol addiction is in progress.

Tobacco Use Disorder. The first investigation of psilocybin for the treatment of tobacco use disorder has reported remarkably positive outcomes. In this open-label study, 15 nicotine-dependent smokers received psilocybin in moderate to high doses over 2 or 3 sessions during a 15-week course of manualized CBT therapy. At the 6-month follow-up, 80% of participants were abstinent based on self-report and verified by biologic measures (urinary cotinine and breath carbon monoxide). This success rate far exceeds those previously seen in clinical trials of any currently available pharmacotherapies for tobacco addiction. Smoking outcomes were significantly correlated with measures of mystical experience during psilocybin sessions, as well as the spiritual significance and personal meaning attributed to the sessions. Based on these very encouraging results, a comparative efficacy trial (psilocybin vs. nicotine replacement) is being performed.

Opioid Use Disorder. Ibogaine was first promoted in the West as having antiaddictive properties in 1962 by Howard Lotsof, who was a heroin addict himself. Uncontrolled data from hundreds of patients suggest that ibogaine is effective in both acute detoxification as well as treatment for opioid addiction. Clinical trials to validate its therapeutic use have begun: a Phase II trial is currently investigating ibogaine's efficacy for methadone detoxification. Phase I trials for noribogaine, as well as an ibogaine analog, are also underway.

Cocaine Use Disorder. There is much hope that psychedelics could be effective for cocaine addiction, given the considerable prevalence of this condition and the limited pharmacotherapies that are currently available. The first clinical trial investigating psilocybin for cocaine use disorder is currently being conducted at the University of Alabama School of Medicine.

Use of Psychedelics in the Treatment of OCD

Only one published study has evaluated the use of psilocybin for OCD. Nine subjects were given up to 4 different doses in a modified escalating blinded protocol. A subthreshold dose was given randomly after the first dosage as a blind. Testing days were separated by at least 1 week. Psilocybin was well-tolerated by all participants, and marked decreases in OCD symptoms of differing degrees were observed in all subjects during one or more of the experimental sessions (23% to 100% decrease in YBOCS score). Research is continuing: at the time of writing, there are three studies investigating psilocybin for OCD.

Other Potential Indications

There are currently registered trials looking at the use of psilocybin for anorexia nervosa, body dysmorphic disorder, and depression in early Alzheimer disease. Clinical trials investigating psilocybin for opioid use disorder are also beginning or planned at multiple academic medical centers. While we await more definitive results of psilocybin's potential for cocaine, tobacco, and opioid use addictions, the possibility of an approach with cross-addiction generality is looking more realistic. Psychedelics could also potentially be explored to treat other drugs of abuse (i.e., methamphetamines, tetrahydrocannabinol [THC]) or behavioral addictions (i.e., gambling, overeating, and sex).

Given the anti-inflammatory and neurotrophic properties of psychedelics, a Phase I study has been completed on a low-dose LSD formulation for the treatment of neurodegenerative diseases; a second Phase I study on the same molecule has been announced.

MDMA

Psychopharmacology

MDMA's primary mechanism of action is indirect—it stimulates the release and inhibits the reuptake of serotonin, dopamine, and norepinephrine. Its most potent effects are on serotonin, and it has an affinity for 5-HT_{2A} receptors. It also stimulates the release of hormones, including oxytocin, vasopressin, and cortisol.

Subjective Effects

MDMA facilitates states of positive mood, increased responsiveness to emotions, a heightened openness, and a sense of closeness to other people. Notably, it does not cause alterations in perception or sense of self commonly seen with classic hallucinogens.

Proposed Mechanism of MDMA-Assisted Psychotherapy

Slightly different mechanisms may be involved in the healing with MDMA than appear to be the case with classic hallucinogens. While the specific neurobiologic mechanisms underlying MDMA's efficacy for treating PTSD are poorly understood, they appear to be related to the alteration of emotional memory. MDMA can mitigate the fear response and decrease defense mechanisms without blocking access to memories or preventing a deep and authentic experience of emotion. This facilitates trauma processing, allowing participants to stay emotionally engaged while revisiting traumatic experiences without being overwhelmed by anxiety or other painful emotions. Emotional memory reconsolidation subsequently decreases or eliminates chronic hyperarousal and stress reactions to triggers. MDMA can also promote a sense of empathic rapport with the therapists, facilitating a corrective experience of secure attachment. Often a clearer perspective emerges that the trauma was a past event and a strengthened awareness of the support and safety that exist in the present. MDMA's use for other indications is thought to flow from this same principle: MDMA is a catalyst that enhances intra- and interpersonal growth without the typical interference of defensive conditioning.

History

MDMA was developed by Merck Pharmaceuticals in 1912. Researchers began investigating its potential to create controllable altered states of consciousness in the 1970s. Informal investigations found it was helpful for treating a wide range of conditions (PTSD, phobias, psychosomatic disorders, depression, suicidality, drug addiction, relationship difficulties, and the psychological distress of terminal illness), and it became the new drug of choice for therapists seeking a legal approach to psychedelic-assisted psychotherapy. However, despite the reported therapeutic potential, MDMA was officially categorized as a Schedule I drug in 1985. An expert committee of the World Health Organization (WHO) recommended the international prohibition of MDMA as an adjunct to psychotherapy. The committee, however, was not opposed to the further study of MDMA. In 1986, the Multidisciplinary Association for Psychedelic Studies (MAPS) was founded, and for more than 30 years promoted research into MDMA. Between 1994 and 2018, nine Phase II clinical trials of MDMA-assisted psychotherapy were completed at seven different sites in four different countries. Six of these trials treated individuals with PTSD and are the basis upon which the FDA ultimately granted breakthrough therapy designation for MDMA-assisted psychotherapy in 2017.

Principles and Goals of Therapy

Similar to psychedelic-assisted therapy with classic psychedelics, the design of MDMA-assisted therapy involves preparatory sessions before the dosing session and integration sessions following it. However, a core difference between classic psychedelic-assisted therapy and MDMA-assisted therapy is the therapists’ interaction with participants during the dosing experience. Whereas participants are largely left undisturbed during a dosing session with a classic psychedelic, interaction with therapists is a vital part of sessions with MDMA. The therapeutic method that has been used in clinical trials for PTSD is described in detail in the MAPS MDMA-Assisted Psychotherapy Training Manual. Key concepts include (1) a nondirective approach based on empathetic presence and listening, (2) encouragement that the participant trusts their own inner healing intelligence, and (3) intervention in the form of guidance or redirection when appropriate to help facilitate the processing of trauma. In clinical trials, therapy has been delivered by two cotherapists (usually a male and a female), and sessions lasted 8 hours. Typically, patients receive two or three dosing sessions within a 12-week period with multiple integration sessions throughout.

Clinical Issues

Indications. The majority of research has focused on the use of MDMA-assisted therapy for PTSD. However, there have also been exploratory studies for three other indications: social anxiety in autistic adults, anxiety associated with a life-threatening illness, and cognitive-behavioral conjoint therapy. These have all produced promising results calling for further studies in the future. There are also trials underway investigating several other conditions (see Clinical Research section).

Limitations. MDMA is currently still a Schedule I drug, which limits its use to research trials. Similar to psychedelics, chronic use of SSRIs blunts the effects of MDMA. Given that SSRIs are commonly prescribed for PTSD, consideration must be given to managing potential destabilization during a period of tapering off.

Complications—Physiologic Toxicity. MDMA appears to be physiologically safe when administered in a clinical setting to healthy individuals. MDMA produces sympathomimetic effects, including elevation in blood pressure and heart rate, primarily due to norepinephrine release. Elevation in systolic blood pressure above 160 mm Hg and heart rate exceeding 110 beats per minute occurred in research participants but was self-limiting. Due to adequately conducted medical screening, there have been no adverse cardiovascular events to date in any clinical trial. Hepatotoxicity has rarely been reported in people using Ecstasy in unsupervised and nonmedical conditions. No cases of liver disease or hepatotoxicity have occurred in clinical trials with MDMA.

Contraindications. Given the risk of hypertensive crisis, concurrent use of MAOIs is contraindicated. In the recent Phase III trial by MAPS, exclusion criteria included participants with primary psychotic disorder, Bipolar I disorder, eating disorders with active purging, MDD with psychotic features, personality disorders, current alcohol, and substance use disorders, pregnancy or lactation, as well as any medical condition that could make receiving a sympathomimetic drug harmful due to increased blood pressure and heart rate, including uncontrolled hypertension, history of arrhythmia, or marked baseline prolongation of QTc interval.

Clinical Research

MDMA-Assisted Therapy for PTSD. The only completed Phase III trials on a psychedelic molecule involve MDMA-assisted therapy for PTSD. The first landmark study was a randomized, placebo-controlled trial of 90 adults with severe, chronic PTSD. Notably, the study included people with common comorbidities such as dissociation, depression, a history of alcohol and substance use disorders, and childhood trauma. Participants were randomized to receive three 8-hour sessions of manualized therapy with either MDMA or placebo, combined with three preparatory and nine integrative therapy sessions. Results showed that three doses of MDMA given in conjunction with therapy over the course of 18 weeks resulted in a significant improvement in PTSD symptoms and functional impairment. After three MDMA-assisted therapy sessions, 67% of participants with PTSD no longer met the criteria for the diagnosis, and 88% experienced a significant reduction in symptoms. The results from this trial demonstrated greater efficacy for PTSD than any other psychotherapy or medication. Of importance, MDMA did not increase the occurrence of suicidality during the study. A pivotal second Phase III trial (n = 104) with similar design was recently completed and also demonstrated robust findings of reduced PTSD symptom severity and decreased functional impairment. It is anticipated that MDMA-assisted therapy for PTSD will likely be approved by the FDA in the near future. The FDA approved an expanded access program in which 50 patients will receive the treatment ahead of official approval.

MDMA-Assisted Therapy for Social Anxiety in Autistic Adults. A pilot study at UCLA demonstrated rapid and durable improvement in social anxiety symptoms in autistic adults following two 8-hour MDMA-assisted sessions of mindfulness-based therapy that were spaced 1 month apart. Results held at follow-up 6 months after the second session.

MDMA-Assisted Therapy for Anxiety Associated With Life-Threatening Illness. A recently completed pilot study randomized 18 participants with anxiety from a life-threatening illness to receive either MDMA or a placebo during two 8-hour psychotherapy sessions. One month following the second experimental session, the MDMA group showed a greater decrease in anxiety symptoms than the placebo group, with group differences trending toward significance.

MDMA-Assisted Cognitive-Behavioral Conjoint Therapy (CBCT). Before MDMA was scheduled in 1985, it was used in couples therapy. It reportedly benefited couples in this context by fostering a deeper connection, enhanced communication, and the ability to receive criticism and work through problems more easily. An open-label treatment development study was recently completed. Six couples were treated with two sessions of MDMA-assisted psychotherapy within a compressed course of CBCT lasting 6 weeks. The study was not designed to separate the effects of MDMA from the effects of CBCT, and there was no control group. However, researchers' impression was that MDMA helped the couples overcome obstacles to communication and mutual empathy, and CBCT provided valuable skills and structure for integrating the MDMA experiences and maintaining the gains over time. Controlled studies are planned based on these promising results.

Other Possible Indications. There are currently registered trials underway investigating MDMA-assisted therapy for alcohol use disorder and eating disorders. Child and adolescent psychiatry is another area that deserves to be explored, as MDMA-assisted therapy offers the potential to address trauma, abuse, and family dynamics issues decades before they lead to poverty, addiction, or chronic mental illness.

KETAMINE

Difference From Psychedelics

As mentioned at the beginning of this section, although ketamine is a hallucinogen, it is not technically a psychedelic, as it does not bind to the 5-HT$_{2A}$ receptor. This makes it profoundly different from psychedelic hallucinogens, both in terms of its effects as well as the way it can be utilized in mental health settings.

A critical difference between ketamine and psychedelics is the limited duration of ketamine's positive effects. Repeated treatments with ketamine are often necessary, if not the norm. The standard treatment model for ketamine currently in use also does not require the combination of ketamine with psychotherapy. Some ketamine clinics offer ketamine-assisted therapy (KAP), but that is the exception. As such, ketamine is intended to be used as an antidepressant drug that is continued indefinitely in contrast to psychedelic medicines that seek transformational change in a setting that includes preparation for and integration of insights from the psychedelic experience.

Ketamine is also a Schedule III drug, whereas all psychedelics are classified as Schedule I. This allows all physicians, nurse practitioners, and physician assistants who have DEA licenses to prescribe ketamine (although there is a large diversity of clinical practices). The remainder of this section will highlight other notable differences between ketamine and psychedelics.

Psychopharmacology

Ketamine acts primarily as a noncompetitive *N*-methyl-D-aspartate (NMDA) receptor antagonist. It is also active at a number of other neuroreceptors in the brain, including opioid, dopamine, and muscarinic and nicotinic acetylcholine receptors; however, the contribution of these possible targets for ketamine's myriad effects remains mysterious and controversial. Ketamine can be administered by a variety of routes, including intravenous (IV), intramuscular (IM), intranasal (IN), sublingual, and oral. The bioavailability of ketamine is approximately 90% when given IV or IM, compared with 25% to 50% intranasally, 30% sublingually, and 16% orally. The majority of clinical research and practice to date has used IV administration, although an IN formulation of one of its enantiomers (esketamine) was recently FDA-approved for treatment-resistant depression and acute suicidality.

Subjective Effects

The subjective experiences induced by ketamine have similarities to those produced by psychedelics but tend to be much gentler. Like psychedelics, ketamine can bring people into nonordinary states of consciousness that have the potential to disrupt rigid and maladaptive mental structures. However, whereas psychedelics can evoke challenging states of mind, ketamine more reliably induces a dose-dependent dissociation and gentle “lifting” of existential burden. The anxiety or fear reactions often produced by psychedelics seem to be attenuated or even absent. Moreover, while a ketamine experience can still be quite powerful, the intensity of the experience is significantly determined by the dose and rate of delivery. This allows clinicians a high degree of control over the treatment (especially when administered IV).

Proposed Mechanism

Similar to all antidepressant drugs, the mechanism by which ketamine exerts its antidepressant effect is a mystery. It is known that ketamine acts as a partial NMDA antagonist. Ketamine appears to affect glutaminergic function and initiate a number of cellular effects on neuron signaling pathways, resulting in new synaptogenesis and elevated levels of BDNF. This neural plasticity is hypothesized to be the basis for the antidepressant response.

Principles and Goals of Therapy

Over the past several years, ketamine has come into use in mental health as an IV medication intervention without a psychotherapeutic component. Ketamine’s potential to provide a legally available form of hallucinogen-assisted therapy has mainly been ignored by researchers in academic psychiatry. However, many psychiatrists are starting to use ketamine (administered IV, IM, or sublingually) in combination with psychotherapy or partnering with therapists who provide psychotherapy. There have been reports that KAP, following treatment models that combine elements of psycholytic and psychedelic methods, may enhance treatment outcomes for patients by fostering greater access to the unconscious. Nonetheless, in general, models of psychotherapy with ketamine have not been formally researched.

Clinical Issues

Indications. Esketamine (an IN form of the s-enantiomer) is FDA approved for treatment-resistant depression and acute suicidal ideation or behavior. Ketamine itself (IV, IM, sublingual, oral) is not FDA-approved for any mental health condition but is routinely used off-label for depression and mood disorders.

Limitations. As previously discussed, ketamine has a limited duration of positive effect and thus is currently utilized as a medication treatment that could continue indefinitely. Although KAP holds a lot of promise, there has been very little research on this model yet. Ketamine can also be expensive for patients: only esketamine has FDA-approved indications covered by insurance, so ketamine currently exists mainly as a self-pay treatment.

Complications—Physiologic Toxicity. In a meta-analysis of 60 studies of the standard ketamine model for the treatment of depression, the most common acute side effects were headache, dizziness, elevated blood pressure, and blurred vision, with most effects resolving shortly after the procedure. Longer-term side effects of patients using ketamine under medical supervision remain mostly unstudied. Evidence from heavy recreational users of ketamine (at doses and frequency much higher than is used to treat psychiatric illness) includes significant urologic pathology, hepatotoxicity, and cognitive deficits. Ketamine is not thought to be a physically addictive substance, as it does not produce physical withdrawal symptoms. However, there have been a few reports of iatrogenic abuse or addiction stemming from ketamine-assisted treatments, and clinicians are warned to inquire about a history of substance dependence in those seeking ketamine therapy.

Contraindications. Ketamine is contraindicated for anyone with active psychosis or mania. It is also contraindicated in patients with underlying conditions in which increased blood pressure would pose a risk of complications (i.e., aortic dissection, uncontrolled hypertension, myocardial infarction, or aneurysms). It is not currently recommended for use during obstetrics, pregnancy, or breastfeeding, as it is unknown if this medication passes into breast milk. Unlike psychedelics, since ketamine does not rely on serotonin as part of its mechanism of action, patients can remain on SSRIs while undergoing ketamine treatment or start ketamine treatment during the process of tapering off serotonergic medications.

Clinical Research

There has been little published research in the newly emerging field of office-based KAP. A recent complication of demographic and clinical outcome data of 235 patients treated with KAP in three psychiatry practices (using sublingual or IM routes of administration) showed a significant reduction in both depression and anxiety. Scores dropped an average of 11.24 points on the Beck Depression Inventory and 5.5 points on the Hamilton Anxiety Rating Scale. The most considerable improvement was seen in patients with more severe symptoms. Although this data is preliminary, it suggests that office-based KAP could be a safe and helpful treatment for a number of psychiatric disorders. Nonetheless, further studies are needed to evaluate the parameters and efficacy of KAP.

CONCLUSION

The efficacy of psychedelic-assisted therapy is self-evident. Modern medicine has already started to consider what value these molecules may have in treating certain populations and conditions (i.e., patients with significant trauma, depression, or substance use disorders), particularly given the limited efficacy of existing treatments. In addition, research is offering the prospect of a much-needed paradigm shift in psychiatry, which bridges the gap between psychopharmacology and psychotherapy. Psychedelics are starting to open our eyes to a deeper understanding of the nature of the psyche and show us how to more effectively access the potential of each individual for psychological healing and growth. A general conclusion drawn from nearly seven decades of clinical research is that psychedelic-assisted treatments activate endogenous self-healing processes.

As a result of diligent and rigorous research by a growing number of investigators, it now appears assured that both psilocybin and MDMA will achieve regulatory approval for clinical use within the next 3 to 5 years. This will pose a challenge to the current economic models of health care, which will need to develop a production and distribution infrastructure as well as a cost-effective treatment model to achieve widespread clinical usage.

Ketamine, despite having a different mechanism than psychedelics, also holds promise as an additional tool in the growing field of consciousness medicine. There are multiple unanswered questions about the potential role of KAP in psychiatry, including how to best optimize the dose, route, frequency, and role of adjunctive psychotherapy. What is clear is that treatment needs to be individualized while these best practices are investigated. Also, it should be noted that ketamine and psychedelics appear complementary and not in opposition. Due to different mechanisms of action and effects, they may be better suited for different uses or have value at different points in a patient’s recovery. While psychedelics will likely become more widely available in the near future, ketamine is likely to have a durable role due to its unique capacity for transient mind–body dissociation, established safety, and its rapid efficacy as a short-term antidepressant and antisuicidal agent.

We are at the dawn of a transformative and exciting era of development in mental health.

FURTHER READINGS

Agin-Liebes GI, Malone T, Yalch MM, et al. Long-term follow-up of psilocybin-assisted psychotherapy for psychiatric and existential distress in patients with life-threatening cancer. *J Psychopharmacol*. 2020;34(2):155–166.

Bogenschutz MP, Ross S, Bhatt S, et al. Percentage of heavy drinking days following psilocybin-assisted psychotherapy vs placebo in the treatment of adult patients with alcohol use disorder: a randomized clinical trial. *JAMA Psychiatry*. 2022;79(10):953–962.

Carhart-Harris RL, Friston KJ. REBUS and the anarchic brain: toward a unified model of the brain action of psychedelics. *Pharmacol Rev*. 2019;71(3):316–344.

Carhart-Harris R, Giribaldi B, Watts R, et al. Trial of psilocybin versus escitalopram for depression. *N Engl J Med*. 2021;384(15):1402–1411.

Danforth AL, Grob CS, Struble C, et al. Reduction in social anxiety after MDMA-assisted psychotherapy with autistic adults: a randomized, double-blind, placebo-controlled pilot study. *Psychopharmacology (Berl)*, 2018;235(11):3137–3148.

Dore J, Turnipseed B, Dwyer S, et al. Ketamine Assisted Psychotherapy (KAP): patient demographics, clinical data and outcomes in three large practices administering ketamine with psychotherapy. *J Psychoact Drugs*. 2019;51(2):189–198.

Gasser P, Holstein D, Michel Y, et al. Safety and efficacy of lysergic acid diethylamide-assisted psychotherapy for anxiety associated with life-threatening diseases. *J Nerv Ment Dis*. 2014;202(7):513–520.

Goodwin GM, Aaronson ST, Alvarez O, et al. Single-dose psilocybin for a treatment-resistant episode of major depression. *N Engl J Med*. 2022;387(18):1637–1648.

Griffiths RR, Johnson MW, Carducci MA, et al. Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: a randomized double-blind trial. *J Psychopharmacol*. 2016;30(12):1181–1197.

Grob CS, Danforth AL, Chopra GS, et al. Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. *Arch Gen Psychiatry*. 2011;68(1):71–78.

Grob CS, Grigsby J. *Handbook of Medical Hallucinogens*. The Guilford Press; 2021.

Holze F, Gasser P, Müller F, Dolder PC, Liechti ME. Lysergic acid diethylamide-assisted therapy in patients with anxiety with and without a life-threatening illness: a randomized, double-blind, placebo-controlled Phase II study. *Biol Psychiatry*. 2023;93(3):215–223.

Johnson MW, Garcia-Romeu A, Griffiths RR. Long-term follow-up of psilocybin-facilitated smoking cessation. *Am J Drug Alcohol Abuse*. 2017;43(1):55–60.

Krebs TS, Johansen PO. Lysergic acid diethylamide (LSD) for alcoholism: Meta-analysis of randomized controlled trials. *J Psychopharmacol*. 2012;26:994–1002.

Mitchell JM, Bogenschutz M, Lilienstein A, et al. MDMA-assisted therapy for Severe PTSD: A randomized, double-blind, placebo-controlled Phase 3 study. *Nat Med*. 2021;27(6):1025–1033.

Mithoefer M, Mithoefer A, Jerome L, et al. *A Manual for MDMA-Assisted Therapy in the Treatment of PTSD (version 8.1)*. The Multidisciplinary Association for Psychedelic Studies; 2017.

Monson CM, Wagner AC, Mithoefer AT, et al. MDMA-facilitated cognitive-behavioral conjoint therapy for post-traumatic stress disorder: an uncontrolled trial. *Eur J Psychotraumatol*. 2020;11(1):1840123.

Moreno FA, Wiegand CB, Taitano EK, Delgado PL. Safety, tolerability, and efficacy of psilocybin in 9 patients with obsessive-compulsive disorder. *J Clin Psychiatry*. 2006;67(11):1735–1740.

Osório F, de Sanches RF, Macedo LR, et al. Antidepressant effects of a single dose of ayahuasca in patients with recurrent depression: a preliminary report. *Braz J Psychiatry*. 2015;37(1):13–20.

Palhano-Fontes F, Barreto D, Onias H, et al. Rapid antidepressant effects of the psychedelic ayahuasca in treatment-resistant depression: a randomized placebo-controlled trial. *Psychol Med*. 2019;49(4):655–663.

Raison CL, Sanacora G, Woolley J, et al. Single-dose psilocybin treatment for major depressive disorder: a randomized clinical trial. *JAMA*. 2023;330(9):843–853.

Ross S, Bossis A, Guss J, et al. Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: a randomized controlled trial. *J Psychopharmacol*. 2016;30(12):1165–1180.

Sanches RF, de Lima Osório F, Dos Santos RG, et al. Antidepressant effects of a single dose of ayahuasca in patients with recurrent depression: a SPECT study. *J Clin Psychopharmacol*. 2016;36:77–81.

von Rotz R, Schindowski EM, Jungwirth J, et al. Single-dose psilocybin-assisted therapy in major depressive disorder: a placebo-controlled, double-blind, randomised clinical trial. *EClinicalMedicine*. 2023;56:101809.

Wolfson PE, Andries J, Feduccia AA, et al. MDMA-assisted psychotherapy for treatment of anxiety and other psychological distress related to life-threatening illness: a randomized pilot study. *Sci Rep*. 2020;10(1):20442.

32.19 Hypnosis

ALLAN DAVID AXELRAD, M.D., DANIEL P. BROWN, PH.D., AND HAROLD J. WAIN, PH.D.

Why and what is it about hypnosis that, for centuries, it has procured myths and the imagination of not only the medical community but the lay public as well? Is it a treatment, a state, or a trait? Where and how can it be used effectively? Myths and misconceptions are abundant. There is a myriad of perceptions among clinicians and the lay public. Even the term hypnosis can be misleading, coming as it does from the Greek root *hypnos* (meaning "sleep"). Hypnosis is not sleep. It is more likely a complex process that requires alert, focused, and receptive attention. Many also accept the myth that the clinician projects the hypnotic trance onto the patient or has the power to influence the patient. In reality, it is the patient who has the hypnotic gift, and the clinician's role is to assess the patient's capacity to capitalize on their assets and to help them to discover and use them effectively. Patient motivation, personality style, and biologic predisposition may contribute to the manifestation of this talent.

The lay public and many clinicians have frequently witnessed hypnosis as a form of entertainment. Though this has kept the intrigue of hypnosis alive, it has also contributed to a possible lack of appreciation or a diminished impact of its possible clinical utility. However, for the trained clinician, hypnosis can be a remarkably effective tool in helping with pain, surgery, stress habits, anxiety disorders, somatic symptom and related disorders, dissociative disorders, trauma- and stressor-related disorders, as well as many medical issues and unexplained symptoms. For experienced clinicians, hypnosis is not always a treatment of a syndrome but is a treatment of the individual patient as long as the clinician is comfortable and experienced using the intervention. With its effect on a myriad of medical and psychiatric conditions that range from PTSD to surgery, hypnosis can be facilitative for researchers to track both biologic and psychological variables. The milieu of hypnosis can also be an exceptional research milieu to explore mind–body interactions.

HISTORY

Early Chinese and Egyptian writings, as well as discussions in the Hebrew Bible and New Testament, seem to refer to hypnotic phenomena, trance states, and spontaneous dissociation being used for medical intervention. Franz Anton Mesmer was perhaps the first to describe the significance of this phenomenon. He postulated that magnetic energy or an invisible fluid could be provided by a clinician or an object to correct imbalances and restore health to an individual with an illness. Mesmer's style was dramatic and included physical passes over the body of the patient by using his own magnetic field in an attempt to restore equilibrium to the patient's "fluid." Many times his patients appeared to have experienced nonepileptiform seizures on their road to recovery.

His theories were highly controversial and, although anecdotally successful, attracted negative attention. In 1784, the French government convened a panel of experts that included Benjamin Franklin, the chemist Antoine Laurent Lavoisier, and Joseph-Ignace Guillotin to meet in Paris. They concluded that the phenomenon was nothing but "heated imagination." As a result, Mesmer's influence in France declined, although history credits him with beginning contemporary medical hypnosis.

In the 1840s, James Braid, an English physician and surgeon, described phenomena similar to those that Mesmer had noted. He reported that he could produce trance states using eye fixation and eye closure. In 1847, Braid rejected magnetic theory and created a psychological concept that he called monoideism, by which he meant mental concentration on a single dominant idea. In this state, subjects were highly suggestible and could focus their attention on specific ideas that would influence behavior.

In 1846, James Esdaile, a Scottish physician serving in India, described in his book *Mesmerism in India, and Its Practical Application in Surgery and Medicine,* the use of such techniques as the only anesthetic in more than 300 major surgical procedures. During the mid-19th century, John Elliotson, a noted professor of medicine and an editor of *Lancet,* became a strong supporter of hypnosis. Because hypnosis was still rejected by many physicians at the time, Elliotson fell out of favor with his colleagues and was forced to resign from the journal. Later many others became impressed with the effect and results of hypnosis and began studying the technique.

By 1880, the noted French neurologist Jean-Martin Charcot at the Salpêtrière in Paris considered the hypnotic state a neurophysiologic phenomenon and a sign of mental illness. Some prominent physicians, such as Pierre Janet, supported this position about hypnosis, but others, such as Hippolyte Bernheim, believed that hypnosis was a function of the normal brain. Bernheim emphasized the role of suggestibility in hypnosis over pathologic theories. Sigmund Freud studied hypnosis with Charcot, and it was central to his classic work on hysteria with Joseph Breuer. The latter facilitated a shift in hypnotic treatments by recognizing the value of abreaction under hypnosis. Of interest, Freud noted in his autobiography his setting aside the use of formal hypnosis and focusing on the concept of transference. When a patient threw her arms around his neck during a hypnotic trance, he related that “I was modest enough not to attribute the event to my own irresistible personal attractions, and I felt that I had now grasped the nature of the mysterious element that was at work behind hypnotism.” To concentrate on the elements of transference, he had to exclude extraneous variables, and it was, therefore, necessary to abandon hypnotism. Along with the rise of behaviorism, Freud’s rejection of hypnosis contributed to its decline in the early 20th century. However, later Freud spoke of the need to blend “the pure gold of analysis plentifully with the copper of direct suggestion.”

After Freud discarded hypnosis, its influence continued to wax and wane. During World War I, Ernst Simmel, a German psychoanalyst, became interested in hypnosis for the treatment of “war neurosis.” He developed a technique for accessing repressed material, which he called hypnoanalysis. Other psychiatric pioneers, such as Sandor Ferenczi, Paul Schilder, Lawrence S. Kubie, Sydney Margolin, Merton Gill, Margaret Brennan, John Watkins, and Erika Fromm, contributed to the early development of a psychodynamic approach to clinical hypnosis. During World War II, hypnosis played a prominent role in the treatment of pain, combat fatigue, and neurosis through the work of John Watkins and John Spiegel. Some of the first research on memory recall and the control of physiologic activity with hypnosis were carried out at the U.S. Army School of Military Psychiatry. Despite Freud’s earlier perceived rejection of hypnosis, many researchers and clinicians helped to shape the modern period of scientific hypnosis. For example, Ivan Pavlov focused on a biologic perspective in the understanding of hypnosis. World War I brought more interest in treating “battle fatigue” and “shell shock.” Other researchers and clinicians, such as Clark Hull, Ernest and Josephine Hilgard, Andre Weitzenhoffer, Milton Erickson, Martin Orne, and Herbert and David Spiegel, all contributed to establishing the contemporary place of hypnosis. In 1955, the British Medical Society formally recognized hypnosis and recommended that it be taught in medical schools. In 1958, the American Medical Association (AMA) and the American Psychiatric Association officially recognized it as a safe and effective treatment. Three specific national societies were soon established, including the Society for Clinical and Experimental Hypnosis, the Society of Psychological Hypnosis (American Psychological Association [APA] Division 30), and the American Society of Clinical Hypnosis.

THE TRAIT OF HYPNOTIZABILITY

When Clark Hull first studied responses to direct ideomotor suggestions in the laboratory he found that responsiveness was not evenly distributed across subjects into a more suggestible and a more resistant group. In the 1950s and 1960s, Ernest R. Hilgard and Andre Weitzenhoffer developed standardized instruments to measure hypnotic responding. The Stanford Hypnotic Susceptibility Scale (SHSS) consists of a formal hypnotic induction ceremony, followed by 12 specific suggestions to alter ideomotor response or cognitive-perceptual experience, or posthypnotic behavior. Administration of these scales to large numbers of subjects led to the view that hypnotizability is distributed, with 10% to 20% of the general population being highly or very highly hypnotizable and the remainder of the population being mildly or moderately hypnotizable. Figure 32.19–1 illustrates the distribution of hypnotizability.

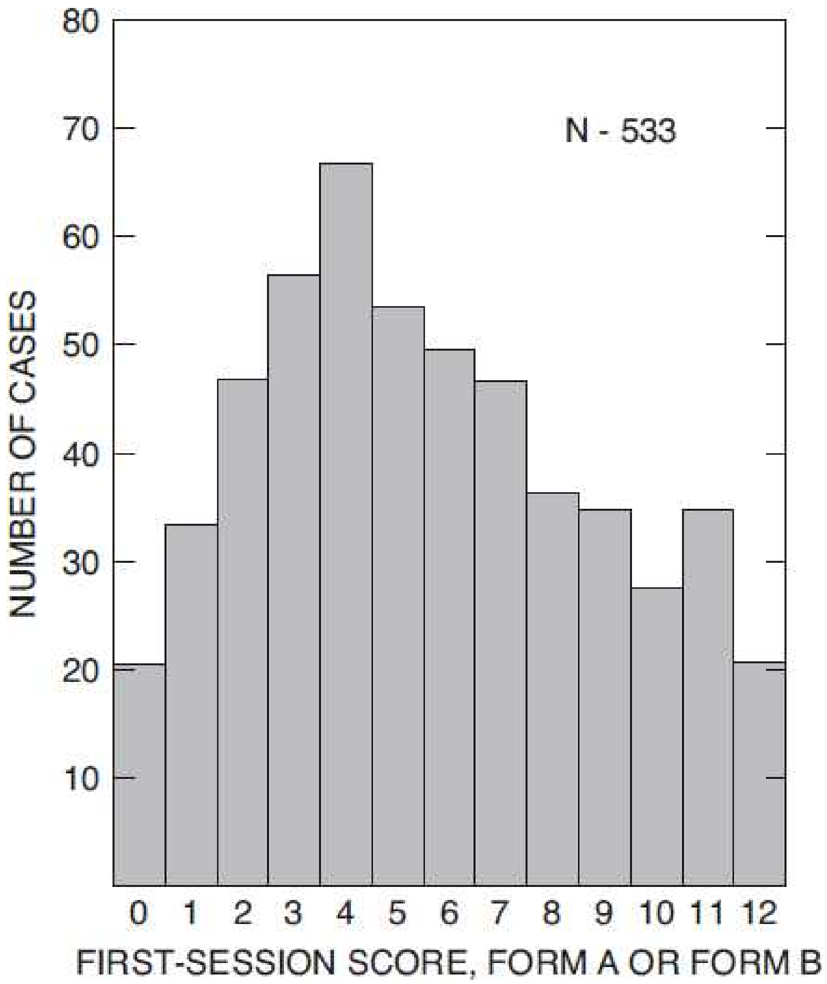

FIGURE 32.19–1. Distribution of hypnotizability resulting from administration of the Stanford Hypnotic Susceptibility Scale to a large number of subjects, reflecting that 10% to 20% of subjects are highly or very highly hypnotizable and the remainder of subjects are mildly or moderately hypnotizable.

The use of a variety of standardized instruments to assess hypnotizability over the last half century in large samples of subjects from different geographic areas representing various cultural and ethnic groups has consistently demonstrated individual differences in hypnotizability. The convention has been to divide hypnotic subjects into three groups:

1. High hypnotizables represent 8% to 10% of the general population. These are the virtuosos of hypnotic talent.
2. Moderate hypnotizables make up the bulk of hypnotic subjects—from 84% to 88%.
3. About 4% to 6% of the general population are low hypnotizable.

A comparison of results obtained using different instruments to measure hypnotizability shows that the cut-off points distinguishing high, medium, and low hypnotizability do not always exactly correspond.

Accurate assessment of hypnotizability in children depends on adapting the hypnotic technique to the level of cognitive development of the child. Preschool children are less hypnotizable than school-age children. Nonetheless, their degree of hypnotic responding increases with brief, open-eye hallucinatory pretend play suggestions matched to the preoperational thinking and short attention span of that age group. Early school-age children respond best to action-oriented fantasy suggestions of longer duration, matched to the late preoperational thinking of that age group. Older school-age children (ages 9 to 12 years) respond well to closed-eye detailed fantasy suggestions and ideomotor suggestions, as well as to a wide range of suggestions for cognitive and perceptual changes, corresponding to the development of concrete operational thinking, along with a rich internal world of fantasy and affective states and stable attentional skills characteristic of that age group. Longitudinal studies of changes in hypnotizability in the same children over the course of child development consistently have shown that irrespective of the issue of matching the style of hypnotic suggestion to the developmental age of the child, there is a progressive increase in hypnotizability from younger to older childhood, with the degree of hypnotizability reaching a peak between 9 and 12 years of age, an age roughly corresponding to the development of concrete operational thinking. Because most children from 9 to 12 years old are highly hypnotizable, this age range has been considered to have peak hypnotizability. For most individuals, the degree of hypnotizability drops off by mid-adolescence, so the majority of older adolescents and adults fall within the medium range of hypnotizability.

Test–retest correlations (0.7 to 0.8) of standardized hypnotizability scales and/or repeated experiences associated with interpersonal trust, such as psychotherapy or participation in growth groups, can lead to very modest, stable increases in hypnotizability over time, at least in some individuals. Drastic alterations in consciousness, for example, induced by hallucinogenic drugs or prolonged sensory deprivation, lead to significant but temporary changes in hypnotizability, after which the degree of hypnotizability returns to baseline for most individuals. There is little evidence that any procedure leads to significant stable changes in the degree of hypnotizability for most adults. In other words, low- and medium-hypnotizable subjects are likely to remain within that range of hypnotizability and are unlikely to become hypnotic virtuosos under any circumstances.

The trait of hypnotizability is distributed across a wide range of personalities, and as research with a wide range of personality tests has indicated, there is no simple stereotype of the type of individual who is highly hypnotizable. With respect to the correlation of hypnotizability to single personality traits, measures of attachment style, attentional skills, absorption, dissociation, fantasy-proneness, and active imagination show a mild to modest, yet significant correlation with hypnotizability.

The correlation between hypnotizability and psychiatric diagnosis has yielded more promising results. Comparing patient groups with independently validated psychiatric diagnoses on standardized measures of hypnotizability has shown that as a group

- patients with anxiety disorders (generalized anxiety, panic, and OCDs), somatoform disorders (e.g., conversion), and irritable bowel syndrome (IBS) are slightly but significantly more hypnotizable;
- patients with PTSD are moderately more hypnotizable; and
- patients with dissociative identity disorder are much more hypnotizable than the general population.

These differences in the hypnotizability of various psychiatric groups have led to the belief that hypnosis may be a preferred treatment, at least for certain psychiatric conditions. Again, this is not to suggest that it is only the hypnotic intervention that accounts for the change; it is the strategy one utilizes with the milieu of hypnosis that accounts for the positive outcome.

The relationship between degree of hypnotizability and clinical efficacy has not been clearly established. Some clinicians select only high hypnotizables for hypnotherapy. However, the majority of clinicians who use hypnotic procedures offer hypnosis most often to medium and high hypnotizables and some of the low hypnotizables under the assumption that anywhere from a low-to-moderate to higher level of hypnotizability is sufficient for most clinical work. Some clinicians believe that establishing certain expectancies, as well as a therapeutic alliance, facilitates therapeutic change even in patients with low hypnotizability. The prevailing view is that the clinical applications of hypnosis are not restricted to highly hypnotizable individuals. Some clinicians suggest that intervention strategies can be based on levels of hypnotizability.

Research consistently supports the view that hypnotizability predicts clinical outcomes for a variety of DSM-5 diagnoses, somatic symptoms, and related disorders and/or DSM-IV Axis I psychiatric conditions and not others. Hypnotizability as a personality trait, irrespective of whether the clinical procedure is conducted in a waking or hypnotic state, predicts clinical outcomes for a variety of psychiatric conditions and medical conditions, such as in the treatment of anxiety disorders, conversion disorders, posttraumatic or dissociative disorders, bulimia, pain, headache, asthma, certain skin conditions, and functional gastrointestinal disorders, but does not predict the outcome for a variety of addictive behaviors, such as smoking, weight control, and alcohol/drug abuse.

The implication of this research on the distribution of hypnotizability is that hypnotic responding, at least for high hypnotizables, is a personality trait or talent. Such a view should dispel the stereotype that hypnosis somehow exemplifies the special power of the clinician to exert undue influence. Appreciating hypnosis as a talent or gift within the patient naturally leads to a more egalitarian, less hierarchical or power-oriented approach to hypnotherapy. In permissive hypnosis, the hypnotherapist provides the relational conditions to facilitate the patient's exploration of their hypnotic abilities and inner resources relevant to therapeutic change.

HYPNOSIS AND SUGGESTIBILITY

Clark Hull systematically studied the effects of direct suggestions on motor response and believed that responsiveness to ideomotor suggestion constituted the primary factor of overall hypnotic responding. Andre Weitzenhoffer studied the effects of suggestion as a form of external social influence through verbal communication. Weitzenhoffer noted that some individuals significantly increased their suggestibility after a hypnotic induction ceremony, and this was used to induce a trance state. This led to the view that trance induction caused a *condition of heightened suggestibility* so that the individual immediately following trance induction was significantly more responsive to subsequent suggestions than to suggestions given outside of the context of trance induction. Weitzenhoffer coined the term “classic suggestion effect” to illustrate how high-hypnotizable subjects following the induction of trance accepted subsequent suggestions involuntarily and with uncritical acceptance and deep conviction, even when the suggestions violated conventional logic or normal reality testing. The *heightened suggestibility model* of hypnosis has had an important influence on hypnosis research since the last half of the 20th century, in that most of the standardized instruments to assess hypnotizability are constructed around this assumption; namely, each instrument begins with a formal hypnotic induction ceremony, followed by a series of specific suggestions given in a standard format. Overall hypnotizability in these scales is assessed as a function of how many specific suggestions are accepted by the subject following the induction ceremony.

In a broader view, heightened suggestibility is best seen as one important dimension within the overall domain of hypnotic responsiveness, with a variety of state effects occurring at least partially independent from suggestion effects in certain hypnotic subjects. The state model of hypnosis assumes that a formal induction ceremony results both in a hypnotic state and an increase in responsiveness to suggestion. Subsequent suggestions serve to shape the nature of hypnotic experiences and behaviors.

THE DOMAIN OF HYPNOTIC EXPERIENCES AND BEHAVIOR

Ernest Hilgard's research on hypnotizability defined six main areas of hypnotic responding: ideomotor response, hypnotic imagery and dreaming, effects on memory, sensory and perceptual effects, cognitive-affective changes, and posthypnotic effects.

Ideomotor responsiveness refers both to the facilitation and the inhibition of motor response through specific suggestions, for example, to facilitate eye closure, postural sway, movement of outstretched hands together as if pulled by a magnet, and hand/arm levitation and lowering, and/or suggestions, for example, to inhibit the bending or lifting of an arm, the unlocking of clasped fingers, the opening of closed eyelids, or the verbalization of certain words. Ideomotor responsiveness accounts for the largest portion of the variance of overall hypnotic responding.

Some hypnotized subjects experience a richer quality and greater quantity of imagery in hypnosis. Imagery, as compared to cognitive information processing, may play a greater role in hypnotic over waking experiences.

The effects of hypnosis on memory are complex. Hypnotic suggestions may increase the recall, which is known as hypnotic hypermnesia, but not the recognition of personally meaningful material (semantic or procedure memories). The accuracy of recalled information varies according to whether the hypnotist created realistic or unrealistic expectancies about memory retrieval, which did or did not encourage guessing and/or did or did not suggest misinformation about the recall target.

Hypnotic analgesia is a well-established phenomenon, especially among high-hypnotizable subjects. A relatively small group of high-hypnotizable individuals are able to produce positive or negative hallucinations in response to specific hypnotic suggestions. A hypnotically suggested heat hallucination not only results in an intense subjective experience of heat that is virtually indistinguishable from actual heat in certain high hypnotizables but may also result in burn blisters. Hypnotic suggestions can also inhibit the perception of actual experiences (negative hallucination), such as failing to see an actual light or color, failing to see a hand on a watch, or selectively becoming deaf to certain words or sounds.

Hypnotic suggestions can also affect changes in meaning systems, affective experience, and self-representation. For example, hypnotic suggestions can produce temporary agnosias and aphasias. Direct suggestions to experience a discrete emotional state, for example, fear, result in not only the subjective experience of fear but also an increase in catecholamine and cortisol output. Hypnotic suggestions can directly alter self-representation, such as the suggestion to imagine oneself completely free of pain or the suggestion to no longer identify oneself as a smoker. It can also be used to create somatic symptoms as a defense or bridge while looking for the metaphorical meaning of the presenting symptom.

Certain suggestions given during a hypnotic state are designed to have an effect after the formal termination of the hypnotic session. These are called posthypnotic effects and include posthypnotic amnesia and other posthypnotic behavior. Posthypnotic suggestions are either cue-induced, that is, the experience or behavior occurs in response to a specific cue given previously during trance, or are graded, that is, the effect occurs progressively, such as the posthypnotic suggestion that the meaning of a hypnotic dream will become increasingly more clear over time.

DIMENSIONS AND PHENOMENA OF THE HYPNOTIC EXPERIENCE

Clinicians suggest that although hypnosis is an altered state, the capacity to experience the phenomenon varies among individuals according to their biologic substrata and cognitive and emotional styles. Most would agree there is both a general and individual specific capacities of hypnotic responding.

Ronald Shor identified eight dimensions of hypnotic experience, each of which varies according to the hypnotic subject: Fading of the individual's generalized reality orientation, nonconscious involvement, archaic involvement, drowsiness, relaxation, the vividness of imagery, absorption, and access to the unconscious. Peter Sheehan and Kevin McConkey concluded from their research that the subjective experience in hypnosis depends on specific expectancies, cognitive style, the nature of imagery, perception of the hypnotist, and reality awareness. The data revealed "patterns of individual differences in reported hypnotic experiences," and that hypnotic experiences are different across subjects, depending on that subject's cognitive style. Three cognitive styles were identified: concentrative, independent, and constructive. Ronald Pekala attempted to quantify the patterns of consciousness in a hypnotic trance and concluded that there were specific differences between low and high hypnotizables. Low hypnotizables scored lower on voluntary control, self-awareness, internal dialogue, rationality, and availability of memory. High hypnotizables scored significantly higher on measures of hypnosis as an altered state, absorption, and altered time sense. These data suggest that a high hypnotizable's experience of an altered state of consciousness or trance is fundamentally different from the state experienced by medium or low hypnotizables.

DEFINITION OF HYPNOSIS

Until recently, there did not appear to be a consensual definition of hypnosis. Earlier hypnosis was defined by any induction procedure followed by a series of suggestions. Most experts agree that hypnosis is associated with a set of unusual subject experiences and behaviors, and most agree with a limited *state definition* of hypnosis with respect to the redistribution of attention. The central issue of ongoing debate is whether hypnosis entails a discrete altered state of consciousness. The 2003 definition of hypnosis by the American Psychological Association (APA) failed to mention either self-hypnosis or hypnosis as an altered state of consciousness. In 2015, a new APA task force was able to reach a consensual definition. Hypnosis was defined as “A state of consciousness involving focused attention and reduced peripheral awareness characterized by an enhanced capacity for response to suggestion.” A hypnotic induction was defined as “A procedure designed to induce hypnosis.”

Sakari Kallio and Antti Revonsuo proposed a solution to the state/nonstate controversy. They convincingly argued that there might be two groups of hypnotizable individuals—a smaller group of highly hypnotizable individuals who have a propensity toward extreme changes in sensation, perception, and cognition, and for whom an altered state definition of hypnosis best applies, and a much larger group of mild to moderately hypnotizable individuals, whose state changes are less extreme and whose experiences are more a product of social context. Although both the state and sociocognitive theories of hypnosis partially explain overall hypnotic responding in both groups, the altered-state and trait models of hypnosis may explain a greater portion of the variance of the high-hypnotizable group, and the sociocognitive task-motivational and response-expectancy models may explain a greater portion of the variance of the medium-hypnotizable group. (A similar argument was made by Herbert and David Spiegel.) Kallio and Revonsuo’s proposal resurrects the problem discussed at great length by Ernest Hilgard, namely that hypnotic responding appears to be distributed into two groups.

Kallio and Revonsuo have expanded the idea that medium hypnotizables may respond more to social influence, whereas high hypnotizables may respond in a way that is at least partially independent of both social influence and hypnotic context.

HYPNOSIS AS REDISTRIBUTED ATTENTIVENESS

According to Amir Raz, “attention is an integral, defining aspect of hypnosis.” Hypnosis is seen as a condition of “atypical attention.” Limited state models emphasize hypnosis as a condition of altered or heightened attentiveness and the effects of altered attentiveness on hypnotic experience and behavior. The hypnotic state is associated with changes in at least three of the normal attentional systems: alerting, orienting, and executive control.

Relaxation suggestions commonly incorporated into a hypnotic induction alter waking alertfulness but do not cause sleep. David Spiegel viewed hypnosis as alert activation without arousal or vigilance, a type of alert consciousness lacking sympathetic arousal.

Hypnotic trance is also associated with a diminished generalized orientation to external reality and inward absorption but also with enhanced orienting to specific hypnotic suggestions while avoiding distraction. Hypnotic trance is also associated with changes in executive attentional processes, which results in (a) enhanced selective focus on suggested effects along with a reduced focus on peripheral cues or on parallel competing for attentional demands; (b) alteration in the allocation of divided attention; (c) alteration in error processing, such as the suspension of critical judgment and normal reality monitoring; and (d) alteration in response monitoring, as exemplified by the facilitation or inhibition of motor response, the shift in response criteria for reporting memory recall, or the suspension of belief with respect to suggested hallucinatory experiences.

In addition to these three frontal attention systems, the parietal lobes play an important role in levels of awareness, broadening or narrowing the field of consciousness, and shifting from local to global attention, and they may also play an important role in the hypnotic experience. The magnitude of these attentional shifts in a hypnotic state is significantly greater than in, for example, a waking absorbed state while reading a book, a deeply relaxed, nonhypnotic state, a state simulating hypnosis, or in whatever state is characteristic of low hypnotizables.

TRANCE CHARACTERISTICS

Another version of the state model uses “state” to imply a *discrete altered state of consciousness* or *trance,* uniquely different from the normal waking state and characterized by a set of unique *trance characteristics,* which include a fading of the generalized reality orientation, involuntarism; a suspension of normal critical judgment and reality monitoring (sometimes called trance logic); dissociation; and the capacity for altered cognitive-perceptual information processing. Some versions of the altered state model stress the relationship between the trance induced by a hypnotic induction and naturally occurring states outside of a hypnotic context, such as self-hypnosis, daydreaming, stares of absorption, and dissociative states.

Induction of trance is associated with a relative *reduction in generalized external reality orientation* and increased inward absorption. Hypnotized individuals are either less aware of, or aware of, but less responsive to, external stimuli than in the normal waking state.

A related component of hypnosis is *absorption*—a narrowing and intensification of attention resulting in complete focus. The shift into the hypnotic state is something like a psychological zoom lens that can shift toward highly focused attention. As attention becomes more intense and focused, awareness of orientation in time and space decreases.

The experience of *involuntariness* or automatic responding is considered a benchmark of high hypnotizability. For example, although it is clear that the hypnotized subject is using active imagination to experience the hands moving together in response to the suggestion that the hands are drawn together like magnets, the subject typically reports that the hands "seemed" to move together by themselves, as if top-down, voluntary executive control over the suggested response was suspended and it occurred in an automatic, nonvolitional manner. John Kihlstrom emphasized that perceived involuntariness does not necessarily imply suspension of volitional control during trance. Sociocognitive theorists have stressed that perceived involuntariness is not really involuntary but is a function of purposeful, goal-directed cognitive activity in response to hypnotic expectancies and the nature of the hypnotic interaction. Irrespective of differences in the explanations for involuntariness, most experts concur that perceived involuntariness is a central feature of hypnotic subjective experience.

Some high-hypnotizable subjects manifest a transient suspension of *critical judgment* during trance, which allows them more easily than in the walking state to believe in the actuality of suggested positive hallucinations, to deny the actual reality in suggested negative hallucinations, accept suggestions for reduced pain experience and hypnotic analgesia in the face of actual pain experience, accept suggestions that confuse imagination and actual perception, or accept the actuality of hypnotically suggested pseudomemories. However, the extent to which a suspension of critical judgment is a defining feature of trance remains largely untested.

Few studies exist that experimentally test suggestions to improve critical judgment during trance, except for a few studies that suggest warnings about possible inaccuracies in hypnotic recall, which tend to show lower memory error rates while not eliminating memory error altogether. Trance logic is often referred to as a strong belief about suggested experiences in hypnosis that is logically incongruous from the perspective of the normal waking state, such as the presence of a hypnotically hallucinated person. Sometimes hypnosis is associated with a suspension of normal waking initiative, planning, and goal-directed behavior so that the hypnotizable subject passively and uncritically responds to suggestions. However, some experts have emphasized hypnosis as an active cognitive, problem-solving endeavor. The common factor may be hypnotically altered action planning, either in the direction of less or more action planning, depending on the hypnotic context.

The hypnotic state has been characterized as a *dissociative state.* However, there are two fundamentally distinct uses of the term *dissociation:*

1. dissociation as shifts in a state characterized by a narrowing or expansion of the field of conscious awareness and
2. dissociation as the disintegration of the normally integrated operations of waking consciousness resulting in the dissociative compartmentalization of experience.

Dissociation as state shifts and changes in the field of conscious awareness, as exemplified by absorption, consistently accounts for a significant but small portion of the variance of overall hypnotic experience. *Dissociation as compartmentalized states* is more characteristic of high than less-hypnotizable subjects, such as the dissociation of the “part of the mind” experiencing hypnotic analgesia from the “hidden observer” part of the mind still registering actual pain experience or dissociation of the emergence of conflictual mental content during trance from the normal waking defensive operations against the emergence of this material.

High hypnotizability may represent a capacity for altered *cognitive-perceptual information processing.* The fading of the generalized reality orientation results in an overall reduction in the amount of information and information processing is relatively restricted to the verbal suggestions supplied by the hypnotist. Temporal or sequential (serial) information processing is characteristic of the normal waking state.

Evidence from two sources implies that information processing in the trance state favors parallel, simultaneous over serial, temporal processing: (1) disordered recall after removal or suggestions for posthypnotic amnesia and (2) increased ability to generate random numbers in the hypnotic state as compared to the waking state, wherein the spontaneously generated numbers tend to be patterned. As a consequence of altered information processing, suggested or spontaneous time distortion is a characteristic feature of high hypnotizability.

Advocates of the altered states-of-consciousness model of hypnosis, such as Charles Tart, view the trance state as a discrete state of consciousness with a set of special characteristics. From their perspective, the hypnotic induction ceremony constitutes a set of expectancies and specific suggestions designed to induce and ratify the special characteristics of the hypnotic state as markedly different from the normal waking state. The unique characteristics of the trance state are believed to correlate with heightened suggestibility on the one hand and the clinical efficacy of hypnosis on the other. Cardena and Terhune asked hypnotic subjects to report their deepest level of hypnosis as compared to a non-hypnotic resting condition. Those in the hypnotic conditions gave more reports of altered state experiences, body images, and time distortion and as being less rational or operating out of self-agency. Expectancies were significantly correlated with altered state reports of hypnosis.

SOCIAL INFLUENCE MODELS FOR HYPNOTIC RESPONDING

According to the *task motivational theory* of hypnosis, responsiveness to hypnotic suggestions does not require the induction of a hypnotic state but depends on the tasks set by the hypnotist, the manner in which these tasks are communicated, and the way in which the hypnotist shapes the attitudes, motivation, and efficacy expectancies of the hypnotic subject. Attitude pertains to whether hypnosis is seen by the subject as a desirable or feared state. Motivation pertains to the willingness of the subject to be hypnotized. Efficacy expectancies refer to the hypnotic subject’s belief in their capacity to become hypnotized.

Theodore Barber’s research demonstrated that most of the unusual subjective experiences and behaviors characteristic of high hypnotizables could be produced without a hypnotic trance induction if the subject were motivated to the hypnotic task. Although these experiments led to a greater appreciation of the relative contributions of attitudes, motivation, and efficacy expectancies to overall hypnotic responding, they did not refute the contribution of a trance trait state to overall hypnotic responding, at least for certain high hypnotizables.

More recently, social influence models of hypnotic responding have been revitalized in the form of a *response expectancy model.* Advocates emphasize the relative omission of expectancy effects in Andre Weitzenhoffer’s suggestion model and Ernest Hilgard’s state model. Response expectancies are defined by Irving Kirsch as “anticipations of automatic reactions to particular situations,” such as automatic, nonvolitional behaviors or emotional states, and are especially relevant to the perception of internal states. According to sociocognitive models, response expectancies regarding hypnotic response are the consequence of culturally mediated a priori expectancies about hypnosis and/or specific information shaped by the interaction with the hypnotist. Response expectancies pertain to the prediction of the probability of successfully responding to a given hypnotic situation. Within a hypnotic context, hypnotic subjects learn to produce a given hypnotic response according to their cultural understanding of hypnotic role behavior and/or the specific expectancies shaped by the hypnotist. In other words, the hypnotic response is the outcome of social influence and the systematic shaping of hypnotic expectancies by the hypnotist. For example, suggestibility increased by only a small amount with a hypnotic induction when the procedure was labeled as “relaxation” but increased by a much larger amount when the procedure was labeled as “hypnosis.” Although expectancy models of hypnosis are valuable in helping in the appreciation of the contribution of expectancy effects to overall hypnotic responding, these models have several limitations. The overall contributions of expectancy effects are sometimes overstated as if expectancy were “the final common pathway to hypnotic suggestibility.” Although expectancy theory has helped in the appreciation of the relative contribution of social influence to overall hypnotic responding, it does not serve as a complete explanation for hypnotic responding, any more than placebo effects serve as a complete explanation for drug effects. Both expectancy effects and state-trait effects contribute to overall hypnotic responses.

THE STATE/NONSTATE CONTROVERSY

An ongoing debate continues between state theorists and sociocognitive theorists of hypnosis. State theorists have turned to psychophysiological and functional imaging studies of hypnotic responding to support limited- or altered-state theories of hypnosis and claim that sociocognitive theorist ignore the evidence from neuroscience. The evidence supports altered brain functioning during hypnotic responding, at least for high hypnotizables, and the hypnotic induction does activate or deactivate certain frontal attentional systems and other brain areas. A review of the neurobiologic evidence follows in the next subsection.

Sociocognitive theorists argue that trance characteristics and the altered state itself are a product of social influence and the shaping of response expectancies and that all of the experiences and behaviors characteristic of the domain of hypnosis can be produced through social influence without prior hypnotic induction or a trance state.

USING HYPNOTIC CHARACTERISTICS IN CLINICAL CONTEXT

Although the controversy over the definition and specifics of the theory remains, the clinical outcome of using hypnotic phenomena is undisputed. The effect of the clinical phenomena of hypnosis may rest on the biology of the patient, the presenting clinical problem, and the skills of the clinician. What follows are some of the hypnotic phenomena identified in the previous subsections that summarize strategies that can be used:

- *Catalepsy.* An involuntary tonicity of the muscles so that limbs remain where they are positioned with inhibition of fatigue. It is used for the assessment of responsivity to hypnosis, for trance ratification, to help with obtaining radiography, for induction of deepening and to facilitate anesthesia and analgesics.
- *Amnesia.* Inhibition or disruption of memory, which may be either spontaneous or suggested. Spontaneous amnesia suggests a deeper trance state. The duration of amnesia is unpredictable. Amnesia

assists posthypnotic suggestions to occur by reducing conscious "spectatoring."

- *Dissociation.* A detachment from the immediate environment, similar to dreaming when one sees oneself doing something from outside of one's body. It is useful in promoting increased depth, in pain relief, and in facilitating "automatic writing" for purposes of unconscious hypnotic exploration.
- *Analgesia and anesthesia.* Experiencing analgesia is like turning down the dimmer switch on a light and is characterized by some milder grimace. Estimates are that 70% to 90% of subjects may develop at least some degree of analgesia, and perhaps 20% can develop complete anesthesia (even during surgery). This phenomenon is useful for managing pain, reducing inflammation and facilitating more rapid healing, and providing a method of trance ratification.
- *Hyperesthesia.* Increased sensitivity to touch.
- *Ideosensory activities.* This refers to the capacity of the brain to develop responses on any of the sensory modalities: auditory, visual, kinesthetic, olfactory, or gustatory.
- *Somnambulism.* This is one of the deepest levels of hypnosis and refers to being in a trance but with the eyes open. A patient may go into a trance, and the untrained clinician may not be aware of its significance.
- *Hallucinations.* A negative hallucination refers to not seeing or sensing a stimulus that is present. Negative hallucination for smell (anosmia) may be produced for a pungent odor in perhaps 25% to 35% of subjects. Having a positive hallucination is perceiving something that is not there. All five senses may be hallucinated. Taste and smell are the easiest to hallucinate, vision is the most difficult (only 3% were able to do this in Hilgard's research sample), and hearing is the second most difficult to hallucinate.
- *Hypermnesia.* Refers to increased capacity for memory recall.
- *Age regression*—(a) *Revivification* refers to a complete regression or reliving of a past event; voice and writing become child-like; (b) In *partial regression* to a past time or event, the patient retains some awareness of the hypnotist and the patient's adult identity and thus has the capacity to apply an adult perspective to the past experiences. This can be an instrumental means in facilitating psychotherapy, uncovering repressed or dissociative data, and working through poor object relations and attachments.

- *Ideomotor activity and ideomotor exploration.* This refers to the involuntary capacity of muscles to respond instantaneously to thoughts, feelings, and ideas—for example, the experiencing of fingers or limbs floating and seeming to move by themselves. Ideomotor exploration is a method for rapid, unconscious exploration using involuntary finger signals. These signals may also be used to signal the therapist when internal processes have been completed (e.g., when numbness is present).
- *Age progression.* Pseudo-orientation in time to the future may seem like a fantasy or mental rehearsal about the future to some subjects, but to talented hypnotic subjects, it can feel quite real. Purposes include covert rehearsal and confidence building; the diagnostic determination of how the patient will react to a future situation; and, in some cases, the determination of what treatment “helped you recover.” In addition, for some depressive states, it also facilitates allowing the patient to foresee the possible future. This can be an invaluable approach in facilitating psychotherapy and working through poor object relations and affiliations.
- *Time distortion.* There are two possible aspects: (a) Time expansion, by which 2 minutes, say, can seem like 20 minutes, with a subjective sense of time going by very slowly. An everyday example would be waiting for a ride on a snowy day or waiting in a slow checkout line. (b) Time contraction, by which 10 minutes, say, can pass in what seems like 1 minute during a pleasurable activity or by suggestion. Purposes can range from sports activities to pain or migraine, to a past review of etiologic factors, to trance ratification to convince patients that they were in an altered state of consciousness. It can be used to expedite painful times while going through medical procedures, as well.
- *Depersonalization.* Through suggestion, an individual can temporarily forget their identity and view past traumatic experiences like an objective intelligence or may take on the identity of another person and identify with them.
- *Induced dreams.* Hypnosis may influence dream content at night, and “dreams” may be produced in a hypnotic state. Purposes include exploring and revealing causes and conflict, identifying solutions and insights, altering feelings, and recalling nocturnal dreams. In a hypnotic state, sleep ideation also may be “searched” in treating insomnia or physiologic disturbances (e.g., morning ulcer disturbance or premature labor).

NEUROBIOLOGY AND NEUROPSYCHOLOGY OF HYPNOSIS

A limitation in the physiologic studies of hypnosis is the failure to distinguish between the physiologic effects of the hypnotic state per se and the physiologic effects of specific hypnotic suggestions given within that state. Seemingly innocuous changes in the wording of specific suggestions, such as "you're becoming more and more relaxed," "focus your attention on," "your breathing will become slower and slower," or "allow a scene to come into your mind," may produce substantial physiologic changes, so that without careful control of the wording of suggestions, the overall physiologic effects of hypnotic suggestions may seem to be a wash.

Electroencephalographic (EEG) studies of hypnosis have failed to reveal a clear-cut EEG signature of the hypnotic state. The hypnotic state consistently is associated with increased theta activity in high versus low hypnotizables, yet the type and location of theta activity may be more important in hypnotic responding than theta activity per se. Posterior theta activity is more a function of relaxation than hypnotic responding per se, but it might signify hypnotic responding if the overall hypnotic suggestions include engaging in progressive relaxation and passively allowing experiences to arise, whereas anterior theta may reflect the activation of attentional systems and active cognitive processing in response to specific hypnotic suggestions. Furthermore, the frequency of theta activity may indicate the presence of imaginative as compared to realistic responses in high versus low hypnotizables and increased attentional focus. Increased alpha activity has sometimes been associated with the hypnotic state in high hypnotizables, but these findings have been disputed.

Increased high-frequency gamma activity (40 Hz) is associated with focused arousal in high as compared to low hypnotizables with and without hypnotic induction. The eye roll sign in certain high hypnotizables is associated with left frontal and bilateral occipital activation. EEG findings clearly establish that hypnosis is not sleeping.

Thus, although no definitive overall EEG signature of hypnosis exists, specific EEG patterns are associated with various components of overall hypnotic responding, such as suggestions for eye roll, hypnotic relaxation, imaginative involvement, receptive attention, and intensely focused attention. Functional imagery studies such as positron emission tomography and MRI studies have consistently documented significant differences in the brain functioning of high and low hypnotizables so that high hypnotizability may constitute a unique brain state.

Induction of a hypnotic state may represent the alteration of brain functions operative in the regulation of consciousness. Helen Crawford and her associates argued that high as compared to low hypnotizables have more efficient executive control systems and supervisory attentional systems. Low hypnotizables do not activate these same brain-attentional systems. Nevertheless, it is also clear that the activation and deactivation of localized brain areas during hypnosis are strongly associated with the specific hypnotic task suggested. There may be no single brain activation/deactivation profile for hypnosis, but instead, a number of neurocognitive patterns, depending on the nature of the hypnotic induction and the specific suggestions, are given.

Various components of the hypnotic induction may also be associated with specific patterns of brain activation/deactivation as suggestions affect the alerting, orienting, and executive attentional networks accordingly. Amire Raz provided fMRI images of the three attentional networks activated in hypnosis.

Hypnotic relaxation was associated with lower levels of activation of the mesencephalic tegmentum of the brainstem, the thalamus, and the anterior cingulate cortex (ACC), comparable to changes produced by waking decreased vigilance or sleepiness, whereas hypnotic absorbed attention was associated with higher levels of activation in the brainstem, thalamus, and rostral ACC. The overall hypnotic induction resulted in lower levels of activation of the parietal areas consistent with an overall reduction in the field of conscious awareness, temporal and spatial distortion, and disorientation of self. According to John Gruzelier, the traditional induction is not a haphazard assortment of roles that the subject is encouraged to enact but follows a sequence with neurophysiologic logic—“the engaging of anterior left-sided attentional mechanisms, which, once engaged, can be suppressed, which in turn allows selective inhibition of frontal functions permitting cardinal features of the hypnotic experience such as automaticity and involuntariness to take place.” Gruzelier added that such selective inhibition and enhancement of certain brain attentional systems also explain the dissociation component of hypnosis. Overall, the evidence strongly implicates the role of the frontal/parietal system in hypnotic responding, particularly the role of the ACC as an explanatory system for many of the phenomena of hypnosis. High as compared to low hypnotizables, have stronger attentional filtering abilities as manifested in the capacity for focused and sustained focused attention and also in the ability to ignore irrelevant stimuli; that is, high hypnotizability is a condition of both "extremely focused attention and disattention."

The left dorsolateral prefrontal cortex (DLPFC) is part of the neurocircuitry of meta-cognitive monitoring. Low-frequency repetitive transcranial magnetic stimulation of the left DLPFC, which presumably blocked the functioning of the DLPFC, increased hypnotizability. This finding is consistent with the view that suspension of critical judgment occurs in high hypnotizable subjects, and that deactivation of the left DLPFC may play an important role in hypnotizability.

The frontal systems are implicated in both the facilitation and inhibition of motor responding in hypnosis. Functional imaging studies both of patients with motor conversion paralysis and normal subjects given hypnotic suggestions for motor inhibition have shown that similar patterns of functional brain changes are involved in each case, namely right ACC activation and right orbitofrontal activation, along with deactivation of the right primary motor cortex. When high-hypnotizable subjects who respond positively to a range of ideomotor suggestions were given either hypnotic suggestions for limb paralysis or suggestions to feign paralysis, different patterns of brain activation were associated with hypnotically suggested and simulated paralysis.

Hypnotically suggested elimination of interference on the Stroop test is mediated by ACC activation. In the Stroop test, subjects name a color in which words are displayed. Response time is generally slower if the name is different from the color. Hypnotic suggestions to ignore competing responses during the test allow for response override and quicker response to the target stimulus. However, whereas both waking and hypnotic suggestions improve Stroop performance in high versus low hypnotizables, hypnotic induction is not necessary for suggestive reduction in the Stroop effect in high hypnotizables.

High hypnotizables with the ability to respond to hypnotically suggested hallucinations as compared to subjects who could not hallucinate showed activation of the right ACC when they were actually hearing an audiotaped message and also when hypnotically hallucinating the taped message but did not show ACC activation when instructed to imagine hearing the message. These data suggest that in terms of brain activation, hypnotic hallucinations are comparable to actual perceptual events and not simply the product of imagination.

Functional imaging studies have shown comparable ACC activation in persons with schizophrenia with auditory hallucinations. With respect to hypnotic visual hallucinations, highly hypnotizable subjects showed activation of the fusiform and lingual areas typically associated with color perception in response to a hypnotic suggestion to perceive a color whether actually shown a color or a gray-scale stimulus. In terms of brain activation, hypnotically hallucinated color perception and actual color perception are similar.

Hypnotic suggestions to increase or decrease pain experience cause changes in the activation of the ACC, insula, and parietal areas, and sometimes deactivation of the primary sensory cortex. Direct suggestions for pain alleviation activate cognitive (via prefrontal and orbitofrontal) and emotional (via insula, inferior parietal, lenticular nucleus and ACC circuitry). These neurobiologic findings are consistent with the view that hypnotic analgesia suggestions may prevent higher cortical processing of nociceptive input, with inhibition being mediated by the cingulate cortex. High as compared to low hypnotizables also show a decreased amplitude in the later P100 and P300 wave components of pain information processing. The difference between high and low hypnotizables in the ability to control pain may represent not only functional but also structural brain differences, in that high hypnotizables who had the ability to control pain had significantly larger corpus callosum size than low hypnotizables.

Overall, the growing consensus is that hypnotic suggestions exert a top-down influence on frontal executive control systems characterized mainly by selective inhibition of various frontal functions. In particular, the ACC plays a central role in the distribution of attention in the context of conflicting attentional demands, in the parallel distribution of attention, in distributing attention between cognitive and emotional processing, in meditating autonomic responses, in coordinating the intensity of attention, in motivational states and the initiation of goal-directed behaviors, in working memory, and in nociceptive experience. The selective inhibition of various frontal systems in general, and the alteration of ACC functioning in particular, in response to verbal hypnotic suggestions, serves as a neurobiologic explanation for such diverse hypnotic phenomena as the fading of generalized reality orientation; intense focus and distraction resistance in hypnosis; hypnotic relaxation; hypnotic analgesia effects; the selective inhibition of competing response potentials, as exemplified by response to the Stroop test in hypnosis, or illustrated by hypnotic suggestions for selected deafness or focal blindness; the production of positive hallucinations; the selective recall of personal memories in hypnosis, but also the diminished memory-error monitoring in certain subjects; the efficacy of hypnotic suggestion in evoking strong emotional states; and the seeming involuntary nature of hypnotic responding. It is well established that high hypnotizables have exceptional attentional filtering and sensory-perceptual gating abilities, so high hypnotizability may constitute a selective pattern of brain activation/deactivation.

Heidi Jiang, Matthew P. White, Matthew D. Greicius, and colleagues, did provide a study, Brain Activity and Functional Connectivity Associated with Hypnosis, and have reported a study of 545 healthy subjects with very high hypnotizability or low hypnotizability using two hypnotizability scales to address subjects with high hypnotizability and low hypnotizability to confirm greater functional connectivity between the executive control network (ECN) and the salience network (SN). They utilize functional magnetic resonance imaging to investigate the activity and functional connectivity among three networks in hypnosis. Utilizing a sophisticated study, they were able to address activity in the ECN, SN, and decreased default mode network (DMN) to address whether high hypnotizability is associated with greater functional connectivity between the ECN and SN.

The a priori hypothesis they addressed was that highly hypnotizable subjects relative to low hypnotizables, when hypnotized, would show (1) decreased activity in the dorsal anterior cingulate cortex (dACC); (2) increased connectivity between the ECN regions, such as the dorsolateral prefrontal cortex (DLPFC) and the SN attentional control regions, such as the frontoinsular cortex and the dACC; (3) decreased connectivity in brain regions immediating self-referential processing, such as the medial prefrontal cortex (MPFC) and the posterior cingulate cortex (PCC) (i.e., DMN). Based upon this sophisticated study, Jiang, White, and Greicius et al., did identify three brain regions whose activity and functional connectivity changed during hypnosis consistent with the a priori hypothesis addressed in this study. The findings in this study were evident in the two hypnosis conditions among high, but not low, hypnotizables; and they were different from the memory and rest conditions. Each of these major findings was also associated with the intensity of hypnotic experiences reported among the highs. The hypnotic state among the highs was characterized by decreased low-frequency amplitude in the dACC. The dACC is a central node in the SN and has been shown to be critical for attentional control, especially context evaluations, such as anxiety, about what we should attend to and what we can ignore. This study does confirm the findings of Helen Crawford and her associates.

In a seminal neuroscience review article, Mark P. Jensen, Graham A. Jamieson, and Antoine Lutz et al. did summarize the key advances in hypnosis research during the past two decades following a 1-day meeting among hypnosis researchers before the International Congress of Hypnosis in Paris, France. The goal of this meeting was to discuss the state of the science of hypnosis research from the purview of clinical and cognitive neuroscience. The publication from this conference did provide a summary of research activity concerning the key issues and points of agreement and disagreement among the participants concerning the recent advances in hypnosis research.

There were consensus recommendations for hypnosis research that were addressed by the conference participants. The consensus view, based upon this review article, was that despite some barriers to hypnosis research, a great deal has been learned regarding the neurophysiologic underpinnings of hypnosis in the past two decades. Highlights include the clinical research supporting the efficacy of hypnosis for managing a number of clinical problems and symptoms, research supporting the role of various divisions in the anterior cingulate and prefrontal cortices in hypnotic responding and that high hypnotic susceptibility is characterized by an atypical brain connectivity profile.

Another highlight of the neuroscience review article by Jensen, Jamieson, and Lutz et al., is the consensus view that hypnosis entails heightened attentiveness, associated with activation of the anterior cingulate cortex (ACC); enhanced metacognitive distraction-detection, associated with dorsolateral prefrontal activation (DLPFC); being less invested in the default mode, namely mind-wandering associated with persistent self-referential thinking and engaging the negative content of the mind (McGeown WJ, Mazzoni G, Venneri A, et al.); deactivation of the medial prefrontal cortex, associated with top-down (self) regulation; and being aware of moment-by-moment changes in the unfolding experience, associated with an activation of the temporoparietal juncture. Finally, high hypnotizability is likely associated with an atypical pattern of learned functional connectivity between these neurocircuits. The overall conclusion would include that as more is learned about the clinical applications of hypnosis treatments, the more individuals suffering from those conditions can be helped with hypnotic treatments—including most certainly chronic pain, IBS, and PTSD, but also likely many other conditions, as well—can be helped.

THE ASSESSMENT OF HYPNOTIZABILITY

As noted earlier, the first standardized hypnotizability scale was the SHSS, which consists of a standardized induction ceremony based on suggestions for progressive relaxation, slower breathing, and going deeper and deeper into trance. Following the induction there are 12 specific suggestions, 9 of which are ideomotor suggestions; the other 3 include suggestions for a fly hallucination (presence of a buzzing fly), posthypnotic amnesia, and a simple posthypnotic suggestion. There are two comparable forms, Form A and Form B, both of which include the same hypnotic induction ceremony, but with somewhat different sets of the 12 specific suggestions. The SHSS is read verbatim to the hypnotic subject, and there is a standard set of scoring rules. Scoring is based on objective, behavioral response to a given suggestion. A given subject can score anywhere from 0 to 12 on the SHSS. An audiotape self-scoring version of the SHSS was also developed—the Harvard Group Scale of Hypnotic Susceptibility (HGS). It contains a similar induction and many, but not all, of the same items as the SHSS. Both the SHSS and the HGS take just about 1 hour to administer, but the HGS is time saving in that its administration does not require the presence of a hypnotist or clinician.

The Stanford group also developed a graded difficulty scale, the SHSS, Form C. It contains a standard induction procedure followed by 12 specific hypnotic suggestions, each of which has been empirically shown to be somewhat more difficult than the previous item. The selection of items on Form C shows greater diversity and range than the items on the A or B scale. It contains only four ideomotor items, five hallucination suggestions, a suggested dream, a suggestion for age regression, and a suggestion for posthypnotic amnesia. There is also a self-scoring audiotape version of Form C, the Waterloo-Stanford Group Scale-Form C. London also used the Stanford scales to compose a standardized scale to assess hypnotizability in children—the Children's Hypnotic Susceptibility Scale (CHSS). The Barber Suggestibility Scale is unique in that it measures response to eight suggestions in a standardized manner but does not include a hypnotic induction ceremony. All of these scales meet acceptable standards of reliability and validity.

Some controversy exists about the value of using laboratory-derived hypnotizability scales in clinical settings. Some critics argue that these scales contain a number of challenge items and create a performance-oriented ambiance that is contrary to the holding environment of therapy. Others have argued that low performance on standardized scales may discourage patients from using hypnosis. However, one study showed that given feedback of low scores as compared to high scores, they were *more* not less inclined to request hypnotherapy. Another criticism is that scales like the SHSS A and HGS lack a range in that 9 of the 12 items are ideomotor items. The scales do not contain items representative of the range of hypnotic phenomena directly relevant to the clinical context. The SHSS C has a greater range and is relatively exempt from this criticism. A more practical criticism is that most of these scales require about 1 hour of face-to-face clinician–patient interaction to complete and, in that sense, are not time efficient. The use of audiotape self-scoring versions like the HGS and the WSGC fully addresses this problem.

Having a patient listen to a tape in another office and then returning the self-scored result to the clinician at the end of 1 hour is remarkably time efficient. Moreover, the clinician is able to chart normative data of hypnotizability with little time investment. The HGS is the assessment instrument of choice in clinics seeking normative documentation on hypnotizability with little allotment of time to accomplish this.

A number of standardized clinical scales were developed to meet the needs of clinicians. The Hypnotic Induction Profile (HIP) begins with an eye roll sign, in which the patient is instructed to look upward toward the eyebrows, then toward the top of the head, and then, while continuing to look upward, to close the eyelids slowly. The eye roll sign is alleged to be a biologic indicator of dissociative capacity. The eye roll is measured in terms of the visibility of the sclera between the cornea and the lower eyelid as the eyelids are closing. The eye roll is scored from 0 to 4. Rapid objectivity also fits well into the medical model. Observing the sclera is easily demonstrated on medical rounds and can facilitate debunking myths of the skeptical. Upon closing the eyelids, floating suggestions are used to induce a trance state. The patient is then told that the left arm will feel light and will float into the air until a cut-off signal is given. Then the arm is lowered prior to giving the cut-off signal, and hypnotic responsiveness is assessed in terms of the speed and number of prompts given before the arm floats up again. The great advantage of the HIP is that it only takes 5 to 8 minutes to administer. The experienced clinician is able to complete the administration in less than 5 minutes.

As the first standardized clinical scale developed, the HIP has enjoyed enormous popularity among clinicians. The Spiegels have collected a normative sample of HIP scores on almost 10,000 patients and nonpatients. The disadvantage of the HIP is that it lacks range. The instrument is largely constructed around a response to an arm levitation suggestion along with subjective experiences associated with that action. However, because the response to ideomotor suggestions like arm levitation is predictive of response to other dimensions of hypnosis, the choice of building a quick, standardized assessment scale around arm levitation is reasonable.

The Stanford Clinical Scale of Hypnotic Susceptibility (SCSHS) contains a standardized hypnotic induction followed by five specific suggestions: hands moving together, dream, age regression, posthypnotic amnesia, and posthypnotic suggestion. It is administered face-to-face and takes about 20 minutes. The SCSHS has a number of advantages.

It is time efficient, and it assesses mostly the range of hypnotic abilities that are directly relevant to the clinician: dream, age regression, posthypnotic amnesia, and posthypnotic suggestion. The disadvantage is that scoring allows for the ready identification of high hypnotizables but does not easily discriminate between low and moderate hypnotizables.

Because the items are reasonably difficult, a small number of patients score 4 to 5, and the majority score 1 to 2. It is difficult then to know whether a score of 1 to 2 represents low or medium hypnotizability.

Overall, no single standardized hypnotizability scale perfectly meets the needs of the clinician. Clinicians wishing a time-efficient way to assess hypnotizability must decide between the HGS and the HIP, with the HGS having better normative data and greater range and the HIP being better at detecting high hypnotizables. The SCSHS taps the best range of clinically relevant dimensions and is relatively time efficient. Although not always practical, of course, the most empirically sound approach is to administer several standardized hypnotizability scales with the same patient, although the scales result in similar but not identical classifications of high-, medium-, and low-hypnotizable subjects.

THE HYPNOTIC INDUCTION CEREMONY

Purpose and Preparation

The hypnotic induction ceremony is a metaphorical performance that suggests a change in conventional verbal discourse. In a more traditional sense, it implies that a ritual to facilitate induction into a different mental state is going to occur. Despite the great variation in hypnotic induction ceremonies, most contain certain common elements. For example, most induction ceremonies contain specific suggestions to focus attention on some target, as well as suggestions to disattend to everything else.

Most induction ceremonies strategically use language to shape the progressively developing and deepening experience: “You are becoming more and more relaxed,” “You are going deeper and deeper into this state.” Some trance inductions, especially with highly hypnotizable subjects, explicitly contain suggestions for entering an altered state of consciousness or trance state, and they may also contain explicit suggestions to enhance or ratify certain trance characteristics such as involuntarism, dissociation, altered time sense, and so on. Hypnotic discourse is often very repetitive. Because repetition and monotony are associated with the destabilization of ordinary waking consciousness, the hypnotic subject becomes more open to unique subjective experiences characteristic of trance. As part of the hypnotic induction ceremony, suggestions are given to increase the subject’s motivation to be hypnotized and find the hypnotic experience desirable and to increase the expectation that the subject will be able to respond successfully to the given suggestions.

Many different inductions are used clinically, but the most successful one will be the one with which both the patient and the practitioner feel most comfortable. Both will have preferences that will influence the success of any given technique. It is appropriate for each clinician to develop their style that facilitates the incorporation of responsiveness to the patient to allow for the progressive depth of trance to occur.

Although many textbooks and pamphlets describe the actual litany that should occur, they are merely templates of ideas that have been shown to work. In the clinical course of intervention, it is imperative that the practitioner be able to change course smoothly if it becomes clear that a particular technique or suggestion is not effective in order to allow the patient to remain in a state of relaxed attentiveness. A good guideline is that all hypnosis is self-hypnosis, and the clinician's role is to facilitate the patient's ability to make maximum use of the patient's trance capacity.

Before beginning, it is helpful to prepare the patient for the experience by discovering what their expectations are and modifying them accordingly with prehypnotic preparation. For example, popular culture's depiction of stage hypnosis performances may convey a total lack of control, and such negative expectancies can be corrected during the prehypnotic interview. Discussion of prehypnotic expectancies may also address helping the patient to become more comfortable with the idea of suggestibility and the nature of subjective experiences typically associated with hypnosis.

If the patient fails to enter what he believes is a trance state, then it is beneficial to try to reframe the failure as something positive, such as suggesting that the patient has unusually good control over their muscles and that the discovery of this fact will be very helpful to know for planning future therapy options: The patient can use that control for increasing relaxation and reaching further depths than might not otherwise be considered possible. In most cases, patient expectations play an even greater role in success than the patient's level of suggestibility. Facilitating appropriate expectancy of change will likely enhance the therapeutic alliance and set the stage for induction. Most induction ceremonies are relaxation and/or awareness based, the difference being whether the patient is concentrating on releasing tension or ignoring the body to concentrate on something, such as the rhythm of the breath or walking down stairs. The most commonly used inductions are as follows.

Progressive Relaxation

The patient is asked to relax, first concentrating on slowing the breathing and then progressively releasing tension in all parts of the body, beginning with the feet and ending with the head/eyes.

Holding On and Letting Go

For some patients who are reticent to relinquish control, it is easier to first tense the body and then release the tension. Emphasis is placed on relaxing and feeling safe and secure, and holding on to only as much tension as is required for the patient to feel in control enough to avoid anxiety about submission. The patient is asked to take deep breaths, release the breaths completely, and then concentrate on how good it feels to do this. The patient is then asked to progressively tense and release each muscle group, starting either from the feet or the head.

Staircase/Escalator

This is a version of progressive relaxation in which the patient is asked to imagine that they are walking down a staircase with 10 steps to the bottom. At the top of the steps is the patient’s current state, and at the bottom is the most relaxed, safe, and comfortable state possible. The patient is instructed to descend each stair slowly on the clinician’s invitation, and in each step, the patient relaxes further (often suggestions such as “You will find that you are twice as relaxed on the next step as you were on this last one” are helpful), slows the breath further, and becomes more at ease. As with the previously described techniques, it can be helpful sequentially to mention relaxing muscle groups such as eyes, face, neck, shoulders, chest, abdomen, pelvis, thighs, calves, and then feet on each step.

Eye Fixation/Closure

The patient is asked to stare at a target that is preferably on the wall or ceiling that is somewhere above the normal field of vision so that some eyestrain is provoked. It is then suggested that the patient’s eyes are beginning to become heavy and that the body is becoming more and more relaxed. The patient is monitored for signs that this is working, such as blinking, watering eyes, eyelids beginning to droop, and so on. These are noted verbally to the patient as assurance that they are entering the hypnotic state. If sufficient time has passed without the patient closing the eyes completely, they may be invited to do so, with the statement that this is to allow the patient to further experience and imagine the experiences that the clinician will be suggesting.

Visual Imagery Techniques

The patients are instructed to use an image to visualize a scene or scenario from the past, present, or future. Throughout the sequence, patients are instructed to relax and focus, and they are reinforced for their style.

Eye Roll Technique

During this application of this technique, patients continue typically with the screening procedure developed by Spiegel, and the technique is reinforced. Patients are asked to roll their eyes to the ceiling while taking a deep breath and holding. They are then given a suggestion for relaxation and for feeling more and more comfortable.

After the induction occurs, the specific strategy begins for most patients, such as for pain, habit control, anxiety, phobia, and so on. Some clinicians describe using the same strategy for all patients, but individual strategies based on each patient's personality traits, motivation, and so on may be more successful. All of these techniques have in common paying attention, focusing attention, and decreasing peripheral awareness.

HYPNOTHERAPY TREATMENT PLANNING AND HYPNOTHERAPY TECHNIQUES

Hypnosis in and of itself is not a treatment. Hypnosis is the *medium* of treatment. Hypnosis is a state of heightened attentiveness (and sometimes an altered state of consciousness and sometimes a condition of heightened suggestibility) within which psychotherapy or suggestions are used for intervention to achieve a goal. The patient, by virtue of going into a hypnotic state, is absorbed in a task, blocks out peripheral awareness, and has rapid assimilation of data.

It is, therefore, imperative to use the best-matched treatment protocol for a given condition. A number of comparative outcome studies have been conducted in which the same treatment protocol was offered with and without hypnotic induction. The overall treatment effect sizes are typically comparable in both the nonhypnotic and hypnotic conditions. The advantage is that a treatment protocol offered to a patient in a hypnotic state generally yields a faster effect for a number of reasons:

1. Discovering and learning new coping strategies is quicker in a nondistracted state.
2. When more focused, the undercurrent of self-talk, schemas, affects, personal memories, and self-states is more accessible than in a

nonhypnotic state, thereby giving an advantage to hypnosis in exploratory therapy.

3. Hypnosis also mobilizes and intensifies the transference, aiding analytic work.
4. Treatment planning with hypnosis entails decisions.
5. Assessing hypnotizability and deciding whether the patient has sufficient hypnotizability for hypnotherapy.
6. Deciding whether the present problem is appropriate for hypnotherapy.
7. Choosing the approach to hypnotherapy.
8. Choosing the style of hypnotherapy.
9. Using patient interests and personality style to facilitate induction and to maintain a level of trance.
10. As in all treatment, developing a therapeutic alliance with the patient is the first necessary step for successful intervention.

A review of problems most amenable to hypnotherapy is included in the next subsection. There are a number of viable approaches to hypnotherapy. The goal of time-effective *problem-solving hypnotherapy* is symptom ameliorization, for example, in the ameliorization of pain or tinnitus. A good deal of the hypnotherapeutic interventions used in behavioral medicine focus on active coping and problem solving. *Exploratory hypnotherapy* emphasizes the exploration of psychodynamic factors operative in the development and maintenance of symptoms, for example, in the treatment of a conversion disorder.

Exploratory therapy also addresses the exploration of inner resources and active coping strategies to address symptoms. *Supportive hypnotherapy* uses suggestions to enhance the patient’s sense of confidence and mastery.

John Hartland called such suggestions ego-strengthening suggestions. *Developmentally informed hypnotherapy* aims at fostering normal development along developmental lines such as self-development, object relations development, and affective development. The hypnotherapist must decide which approach best meets the needs of the given patient’s presenting problem.

There are three very different hypnotic styles. *Directive hypnotherapy* is derived from persuasion theory. The hypnotist serves as an authoritative source and gives direct suggestions for symptom amelioration. This model is best suited to cultures in which doctor–patient relationships are more hierarchical than egalitarian, as was the case in the United Kingdom, as represented in the older *British Journal of Medical Hypnosis* prior to 1970. *Permissive hypnotherapy* is based on the assumption that hypnotizability is a talent within the patient and that the patient has the inner resources to get well. In permissive hypnosis, the hypnotherapist provides just the right relational conditions in the context of an egalitarian relationship in which the patient is provided with a permissive relational context and encouraged to explore inner resources and coping strategies.

Neo-Ericksonian *indirect hypnosis* is derived from the work of Milton Erickson with difficult patients—those who were not very hypnotizable and/or who had a history of treatment failure. Indirect methods include the strategic use of metaphors and scripts, interspersal technique, and binds, double binds, and paradoxes. Although these methods are well suited for the so-called difficult patient, they are sometimes used as general hypnotherapy techniques. The therapist must decide which hypnotic style is best matched to the given patient. In contemporary Western culture, a permissive approach is best suited to the majority of hypnotherapy patients. Directive hypnotherapy works with a small percentage of patients who fit the profile of higher hypnotizability and lower intelligence. Indirect hypnotherapy is best matched to patients with intractable symptoms of a history of treatment failure.

For example, consider the introduction of hypnotic pain-coping strategies in a 12-week group of eight patients with chronic pain. One of these patients may respond quickly to direct, graded hypnotic suggestions that the pain will become less and less severe over time. The majority of the patients may discover and gradually master individualized pain coping strategies over the first two-thirds of the group using a permissive, individualized approach to group hypnotherapy. Two patients may not respond to either of these approaches. Both may be likely to complain that the pain is becoming worse through treatment. These patients would then be given indirect hypnotic suggestions. For example, such a patient is told that it is very difficult to tolerate the uncertainty, never knowing when the pain will become worse. Therefore, as a means of gaining more control and certainty, the patient is asked to try voluntarily to increase the pain. Such paradoxical interventions, when matched to the right patient, serve to ameliorate the pain.

Although such indirect methods should never be overused, they work well with a particular type of patient. There are two main approaches to hypnotherapy: (1) a *protocol-based approach* and (2) an *individualized approach.* In this era of empirically supported treatment, there is a growing trend toward using established manualized treatments that include a set of necessary treatment foci for the treatment to be effective. A well-established treatment protocol derived from outcomes research outside of the hypnosis domain can be used with a patient in a hypnotic state.

This *hypnotic induction plus manualized treatment protocol* is becoming increasingly popular in, for example, the treatment of anxiety conditions, depression, sexual dysfunctions, and behavioral medicine. An alternative approach is to use hypnoprojective techniques (described later) to identify coping resources within the individual that helps the patient to resolve the problem or condition. For example, the hypnotized patient is asked to imagine a video about a person who had a similar problem but discovered ways of overcoming it. A second video is about another person with the same problem who discovered a way to overcome the problem that was very different from that used by the first person. After a series of three to five such scenes, the patient is told to reflect on the scenes until the meaning of each becomes more and more clear and until the patient sees how such scenes inform them about the best way to overcome the problem. Next, the patient is asked to imagine themself in the scene using these coping resources effectively until they begin to develop a sense of mastery. The patient is given a posthypnotic suggestion that thoughts about these coping strategies will naturally come to mind in everyday life whenever the patient notices the very first signs of the symptom/problem occurring. Using this displacement technique, patients discover powerful inner coping resources through the metaphor of the “displaced other” and then learn to apply these coping strategies to their symptoms, first in fantasy and then in everyday life as the symptoms just begin to intensify. The power of the displacement technique rests in the fact that all of the strategies for overcoming the symptoms come from the patient, leading to a strong sense of mastery and self-efficacy. A comparable approach is the inner wisdom method, in which the hypnotized patient is asked to consult with their “inner wisdom” to discover the best way of effectively dealing with a given problem.

There is a large repertoire of hypnotherapy techniques. *Hypnoprojective techniques* entail the symbolization of inner coping resources and unconscious processes. The “theatre technique,” for example, entails having the patient visualize a theater with a play spontaneously appearing. To explore coping resources, suggestions can be given that “the play will somehow be about effective ways of coping with (name problem).” To explore psychodynamic processes, suggestions can be given that “the play somehow will be about whatever has contributed to the development of (name problem).” To explore transference, suggestions can be given that “the play will somehow be about the treatment relationship.” The hypnotherapist can structure a play with several acts so as to gradually uncover the meaning of a given problem or gradually discover a variety of new coping resources for more effectively dealing with the problem.

Other hypnoprojective techniques include the “television or video technique,” the “old house technique,” and the “library technique.” Not all hypnoprojective techniques use imagery. *Automatic writing and drawing techniques* entail giving suggestions for dissociation of the hand, followed by suggestions that the hand is able to move involuntarily and to make writing or drawing movements on its own. Then suggestions are given that the dissociated hand is able to express inner coping resources and/or unconscious processes. A more detailed discussion of hypnoprojective techniques can be found in Daniel Brown and Erika Fromm’s *Hypnotherapy and Hypnoanalysis* and John Watkins’ *Hypnotherapy Techniques.*

In Calvert Stein’s clenched-fist technique, the patient is asked to make a fist, then scan the body, then imagine all the tension from various sources in the body gradually flowing into the arm and then into the fist until it gets tighter and tighter. Once all of the tension flows into the fist, the patient is told to let it go. Hypnotic suggestions can be given for soothing (“A scene will come to mind that is somehow about a deep sense of soothing or comfort”). Additional suggestions are then used to amplify a feeling of soothing or comfort. The metaphor of an affect dial is useful in regulating the intensity of the affect state. The patient is asked to imagine themself in a video at a time when they are feeling a particular emotional state that is difficult to tolerate.

Then the patient is asked to imagine holding a remote control device with a special volume dial that can turn up or down the intensity of the feeling until just the right level of feeling is found. A wide variety of affect regulatory techniques exist within the hypnotherapy literature, and many have recently been collected in Carolyn Daitch's *Affect Regulation Toolbox.*

The field of hypnotherapy contains a number of *methods for the rapid identification and resolution of conflicts.* John Watkins' *ego state therapy* consists in calling forth in a trance each "part of the mind" associated with a given conflict, identifying the adaptive function of each "part," and then suggesting that other "parts of the mind" come forth, giving various perspectives on resolving the conflict. For example, if a patient is not ready to commit to the treatment for problem drinking, the clinician can call forth in a trance the "part of the mind" that wants to stop drinking, then call forth the "part of the mind" that does not want to stop drinking, then heighten the cognitive dissonance between these two "parts," and then suggest that other "parts" come forth that have a perspective on resolving the conflict between the "parts" wanting to stop and not wanting to stop drinking.

Ultimately the dissonance is resolved *within* the patient, and the internal state shifts away from the extreme resistance to stopping drinking. Using Leslie LeCron's *ideomotor questioning technique,* the clinician establishes communication directly with the so-called "unconsciousness mind," which is told that it can communicate directly using automatic ideomotor finger signals, with one finger signifying "yes," another finger "no," another finger "don't know," and another finger "don't want to say." Then the patient's "unconscious mind" is asked a series of questions pertaining to the meaning of a given symptom, such as whether it serves an emotional need, whether it is associated with a past conflict, whether it is associated with a limiting belief, it is a way of identifying with someone, and so on. In this manner, the hypnotherapist is able rapidly and precisely to identify the conflict central to the development of and/or maintenance of a symptom. This ideomotor method has been used extensively in identifying factors involved in resistance to treatment, exploring psychosomatic problems, exploring psychological factors involved in premenstrual symptoms and infertility problems, and accurately assessing motivation and level of risk of suicidal behavior.

In general, a good suggestion for clinicians is to recognize that the hypnotic evaluation can be made part of the overall evaluation and can offer information regarding patient processing and response. Based on patient response to the HIP, Herbert and David Spiegel described patients as ranging along a continuum of low to high that they call the Apollonian–

Odyssean–Dionysian cluster. Hypnotic and psychotherapeutic interventions are based on each patient’s hypnotic capacity.

Prior to doing any hypnotic intervention, it is recommended that a biopsychosocial evaluation be performed along with a mental status exam. Recognizing and understanding the development of the behavior, symptom, or underlying problem can help to avoid later complications. Removing symptoms without understanding their development can lead to negative consequences.

HYPNOSIS—CLINICAL APPLICATIONS

As previously noted, hypnosis has been used as a clinical intervention in medical and mental health care in a variety of settings. Following the recognition of hypnosis as a form of medical treatment by the AMA in 1958, hypnosis has been used in the treatment of a variety of medical problems and a number of medical interventions. Although there is no uniform availability of hypnotic interventions, physicians, psychologists, and other mental health professionals do provide opportunities for patients to undergo hypnotherapeutic interventions for the alleviation of medical problems and as a form of medical intervention. Harold Wain, in the paper “Hypnosis on a Consultation-Liaison Service,” described a variety of areas in which hypnosis has been used in the hospital setting.

In psychiatric settings, hypnosis has been employed for a number of psychiatric disorders, which are addressed later. Over the last 25 to 30 years, professional societies for hypnosis have developed beginning, intermediate, and advanced training programs for health care professionals for the application of clinical hypnosis to their patients. Furthermore, there have been a significant number of evidence-based studies that document the efficacy of hypnotic interventions for medical problems, medical procedures, psychiatric disorders, and psychological disorders.

Trauma and Hypnosis

Freud suggested that trauma is an overwhelming experience that creates havoc within the individual and puts into question one’s homeostasis and psychological identity. According to David Spiegel, traumatic stress manifests as a feeling of helplessness and loss of control over one’s safety.

Individuals who experience the extremes of arousal during a traumatic event are more likely to experience PTSD symptoms over time. Too high a level of anxious arousal has a disorganizing effect. Too low a level of arousal as a consequence of dissociation results in failure to integrate the traumatic experience. In either case normal processing of the traumatic event is disrupted, and PTSD symptoms are more likely to develop over time. Dissociation is used as a means of defending oneself against what might otherwise be too overwhelming to manage. The effects of trauma are generally classified into immediate aftermath effects, or acute stress disorder; chronic or delayed PTSD; and complex PTSD. Simple PTSD is generally associated with a single or circumscribed series of traumatic event(s).

Complex PTSD is generally a function of multiple comorbid psychiatric conditions, including a number of other psychiatric disorders (sometimes with a major dissociative disorder) and personality disorders. Chronic cumulative trauma, early childhood trauma, and/or trauma aggravated by early childhood attachment disorganization are associated with complex PTSD. When one is contemplating intervention it is always important to look at not only the psychiatric comorbidities but the personality styles of the patient and any medical condition that is occurring prior to intervention. Integrating a biopsychosocial approach is facilitative in gaining positive outcomes. Because the development of simple PTSD after a traumatic event is correlated with higher levels of either anxiety or dissociation during the trauma, many have theorized that hypnosis might be an effective treatment due to its ability to calm anxiety and to breach the dissociative symptoms. Specifically, when hypnotic techniques are used as a way to facilitate exposure and work through the painful memories of the trauma in a controlled way, the frequency and intensity of PTSD symptoms often appear to decrease.

The clinical use of hypnosis as a treatment of trauma was first formally described in the late 19th century by Pierre Janet, and during World War II clinical hypnotic techniques were highly successful in treating soldiers suffering from “combat neurosis.” Since that time, use of clinical hypnosis has continued, both alone and as an adjunct to other psychodynamic and CBTs, as a treatment of acute stress disorder, PTSD, and complex PTSD.

Randomized controlled studies of hypnosis and acute stress disorder, PTSD, and complex PTSD are much less prevalent in the research literature than are case studies. One clinical trial found hypnosis alone to be just as efficacious in the treatment of PTSD as psychodynamic psychotherapy and more efficacious than wait-list controls. With respect to simple PTSD, Daniel Brom and coworkers suggested that different treatments were efficacious for different PTSD symptoms: Whereas hypnosis and desensitization were more efficacious for intrusion symptoms, psychodynamic therapy was more useful for avoidance symptoms.

A meta-analysis by Jeffery Sherman, including the study by Brom and coworkers, as well as others, concluded that the beneficial effects of hypnosis for PTSD might be seen more at follow-up than immediately after treatment termination. In a more recent randomized, controlled trial, hypnosis was added to brief CBT in the treatment of acute stress disorder. The findings suggested that CBT combined with hypnosis was more efficacious in preventing re-experiencing symptoms immediately after treatment than CBT alone. However, no differences were found between the CBT alone and CBT-plus-hypnosis groups in terms of prevention of PTSD at the 2-year follow-up: The two groups had a similarly high rate of prevention of PTSD compared to supportive counseling. Manualized hypnotic ego state therapy has also been used as an alternative to CBT for simple trauma.

Because hypnosis has been demonstrated to be effective when used as an adjunct to many CBT and psychodynamic treatments for various disorders, many researchers contend that there should be more systematic research in the use of hypnotic techniques for the treatment of trauma. Reinforcing this recommendation, many case studies have demonstrated the effectiveness of hypnosis in treating PTSD-like symptoms such as dissociation, sleep and dream alterations, and pain, further justifying investigation into these techniques.

Aside from the psychiatric sequelae of a traumatic event, there is often a co-occurring physical trauma as well, and physical recovery and pain management are often an integral part of the overall recovery process. Although medical care is often essential after the trauma, the medical care received in hospital settings can sometimes be viewed by patients as an extension of the traumatic experience.

Similar to other disorders, patients experiencing pain after a trauma are not motivated and are less likely to respond to treatments for acute stress disorder and PTSD prior to the reduction in their physical discomfort. For such patients, pain control is imperative. Complicating the picture is the possibility that their pain is a defense against the trauma. Hypnosis has been demonstrated to be a clinically useful adjunctive tool in the treatment of pain, both in trauma and nontrauma patients, and can often be used as a successful analgesic for phantom limb pain that would normally be difficult to treat. Furthermore, recognizing that hypnosis can treat some of the physical symptoms of pain, patients may be more amenable to the use of hypnosis for related trauma residuals, including PTSD.

When used appropriately as a treatment of simple PTSD symptoms, there has been no evidence that hypnotic abreactions retraumatized patients or caused psychotic breaks. Similarly, not only did hypnotic analgesia cause no negative side effects, but it was also demonstrated to produce a multitude of benefits not exclusively related to pain reduction, including increased control over pain, increased sense of relaxation and well-being, decreased stress, and increased positive mood states. Thus, hypnosis might prove to be a better option for highly hypnotizable patients than more “traditional” treatments of pain and PTSD-like symptoms, especially if there is a high risk of medical complications and side effects from the traditional treatments.

Although levels of hypnotizability vary to a large extent among the general population, researchers have found that patients with acute stress disorder, as well as those developing subsequent PTSD, appear to be more hypnotizable than trauma survivors who did not develop acute stress disorder or PTSD. Several studies with standardized hypnotizability scales have suggested that individuals suffering from posttraumatic symptoms tend to be more highly hypnotizable than most other clinical and/or nonclinical groups. However, other studies have suggested that the nature and duration of the trauma play a role in the level of hypnotizability, such that hypnotizability was more highly correlated with those who experienced repeated, as opposed to isolated, traumas. These results were bolstered by similar research findings that chronic dissociation is more likely to occur in the context of repeated traumas than in the context of single traumatic events.

Military combat soldiers seem to be at particular risk of dissociation during traumatic combat situations. Charles Marmar posited that combat soldiers often protect themselves from the full psychological effect of traumatic events by engaging in “peritraumatic dissociation,” which allows them to distance themselves from an event while still staying engaged in a peripheral way. In a subsequent study, J. Douglas Bremner and Elizabeth

Brett found that patients with PTSD reported higher levels of dissociation during combat-related traumas than did patients with non-PTSD. In fact, when compared to those experiencing non-combat traumas, soldiers citing combat as their worst traumatic experience were more likely to have lifetime PTSD and unresolved PTSD symptoms and were also more likely to be unemployed, fired, divorced, and physically abusive to their spouses. There has been recent significant work at Walter Reed Army Medical Center (WRAMC) and after the move to and merger with Bethesda Naval Hospital, Walter Reed National Military Medical Center (WRNMMC) using hypnosis in the treatment of trauma-related symptomatology in medically hospitalized service members returning from Operation Iraqi Freedom (OIF) and Operation Enduring Freedom (OEF).

Harold Wain and his associates on the Psychiatry Consultation Liaison Service described many techniques and approaches in attempting to decrease psychiatric comorbidity in physically and emotionally traumatized soldiers. They stressed the need to develop a therapeutic alliance with the patient and then use hypnotic techniques initially as a way of facilitating the stabilization of the patient.

In order to decrease the stigma associated with psychiatry, Wain renamed part of the service “Preventive Medical Psychiatry.” Ego strengthening and reframing techniques are used, along with methods to enhance the patient’s assets and help to support ego defenses, using a procedure called “empathic exposure,” in which the patient is asked to recall the trauma, and then their assets are reinforced. Hypnotic recall through age regression or the affect bridge technique is used to help a patient to learn to reframe and master trauma-related symptoms. The affect bridge may be used as the patients are recovering and symptoms remain; they are asked to recall an earlier like feeling while under hypnosis. The affect and cognitions are processed and worked through. If the experience becomes too overwhelming, the patient is taught to go to a safe place or to use the split-screen technique described by Herbert and David Spiegel. The use of these techniques does not preclude other psychological or pharmacologic interventions.

Wain described two cases demonstrating the above approach:

A 36-year-old male suffered a traumatic above-the-knee amputation while in IRAQ after having been blown out of his vehicle. After being air evacuated to the states and while in the ICU, he heard a door slam, and as that occurred, he jumped and fell out of bed, dislodging his IVs. Upon evaluation, he described hearing what he thought was a blast that reminded him of what occurred in combat. Hypnotic techniques were used to help him decrease his initial startle response. We then went back to talk about the trauma and that he felt somewhat responsible and blamed himself for not changing the route of the jeep. Prior to reliving the incident, he found a safe scene to be absorbed with if his anxiety became too overwhelming for him. He then again described the incident where the jeep was blown up, and he was thrown several feet away, but he crawled back, dragging his leg behind him to save one of the other soldiers strapped into the vehicle that was on fire. He was reinforced for his heroism in saving the individual life and for the strength and determination he demonstrated. This procedure of reinforcing his assets, along with reframing techniques, was repeated for several sessions.

As he began to feel more comfortable, dreams of others getting hurt both in combat and in his civilian life were explored. An affect bridge of going back to an accident he had as a teenager where a passenger in his car was injured while he was driving became very prominent for him. As he was able to work that through, his negative dreams and nightmares ended, and his pain levels decreased.

A 42-year-old married mother of three children had been kidnapped and locked in a large packing trunk. After she freed herself and broke out, her abductors stabbed her multiple times, tied her up, put her back in the trunk, and threw her down a cliff. She eventually managed to break out and crawl to safety. Eventually, she was picked up by a passerby. She reported that others had seen her lying on the road and appeared frightened to approach her. Eventually, 911 was called, and she was transported to a hospital. Following medical stabilization, she was discharged and found herself developing nightmares, re-experiencing avoidance, and having hyperarousal symptoms. She was referred by her internist for treatment and was initially started on 25 mg of sertraline, which was increased to 50 mg 4 days later. She was evaluated on the HIP and determined to be a mid- to the high-range hypnotic subject. She was taught to go to a safe place and to use a split-screen technique. She was also given permission to describe her nightmares, re-experiences, and overwhelming anxieties and fears that she faced while being captive, as well as her feelings of abandonment while lying on the road. She was reinforced for her ingenuity in breaking out of the trunk. Her feeling of blame for her capture was reframed while she was under hypnosis. She was taught to calm herself and to reframe her negative feelings about her helplessness. Hypnotic age regression was used to help her master her experiences and facilitate their becoming like a bad movie. Initially, her startle response was used as a signal for her to go to her comfort zone. Age progression was used to help her to rehearse the future. The treatment used the milieu of hypnosis along with exposure, cognitive reframing, psychodynamic approaches, and pharmacology.

Although there is a paucity of randomized, controlled trials studying the effect of hypnotic techniques in the treatment of trauma, leading clinicians and researchers in the field have nonetheless outlined a strong rationale for its use. According to Etzel Cardena, Onno van der Hart, and David Spiegel, there are many solid reasons why hypnosis should be used as an adjunct in the treatment of simple PTSD victims. First, the highly hypnotizable nature of many patients with PTSD allows for greater effectiveness of hypnotic techniques. Similarly, because many patients with PTSD suffer from dissociation, they can recollect, reintegrate, and work through varying parts of their experiences while in a controlled setting under hypnosis.

It is known that some people who have experience a traumatic event may have distorted, fragmented memories of it, and under certain circumstances, as in cases of full dissociative amnesia, the use of hypnotic dissociation could facilitate access to previously forgotten memories or events. Finally, although hypnosis might not be a treatment in and of itself, it is a medium that can expedite treatment and can easily be added to many of the leading interventions. Research has further demonstrated that hypnosis can have a synergistic effect when added to previously established therapies.

With respect to complex PTSD, the standard of care is phase-oriented trauma treatment (POTT), reviewed extensively in Daniel Brown and coworkers' *Memory, Trauma Treatment, and the Law.* Most experts concur that the three main phases of treatment are stabilization, memory and representational integration, and postintegrative normalization.

Because processing trauma as a sole treatment focus in complex PTSD can have a disorganization effect, most experts concur that hypnosis helps in the development of a variety of trauma-specific skills and coping strategies, as well as in facilitating the developmental repair of the sense of self, attachment representations, and affect regulatory structures *prior* to trauma processing, as part of a multifocused stabilization phase. Guidelines for the assessment and treatment of complex PTSD, as well a review of outcome studies, can be found in Chris Courtois and Julian Ford's *Complex Traumatic Stress Disorders.*

Hypnosis and DSM-5-TR Somatic Symptom and Related Disorders

Hypnosis in the evaluation and treatment of somatic symptoms as defined in DSM-5-TR is a broad category that has not yet been well studied with evidence-based trials, with the exception of conversion disorders where there are many well-documented case reports.

The association between conversion disorder and hypnosis was first suggested by Pierre Janet in 1907 as a form of autohypnosis. More recent clinicians have also suggested that conversion symptoms may result from spontaneous self-hypnosis involving a dissociation of sensory-motor function in reaction to traumatic events or to other nontraumatic internal conflicts. In a sized sample of patients Wain, while working on a consult liaison service, observed that patients who exhibit conversion disorder symptoms showed a significantly higher capacity for hypnosis. Several Dutch studies demonstrated similar findings. Experts on somatic symptom disorders concur that hypnosis may therefore be used to the clinician's advantage to both diagnose and treat the disorder. Wain often uses his modification of the HIP to ascertain hypnotizability in medical patients who have physical symptoms lacking physical finding. The eye roll measurement fits well into the medical model. Interestingly, Vuilleumier based on new data emerging from functional neuroimaging reported similarities in brain functioning between conversion disorders and hypnotic suggestions but also described differences.

Charcot, Breuer, and Freud in the late 19th and early 20th centuries contributed greatly to the development of the concept of conversion disorder. In the first edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-I) conversion disorder was called a conversion reaction, in the second edition it was called hysterical neurosis, and in the third and fourth editions the conversion process was restricted to symptoms affecting the voluntary motor and sensory nervous systems. Motor conversion symptoms include nonepileptic seizures, contractures, paralysis, and stance and gait disturbances. Sensory-based conversion includes focal blindness, elective muteness, and conversion pain. Symptoms for which there was some physiologic understanding (generally involving the automatic nervous system) were subsumed under the psychophysiological disorders.

A broadened definition of conversion disorder was used in the DSM-III, which was characterized by symptoms involving any “loss or alteration in physical functioning suggesting a physical disorder” as long as the mechanism of conversion was evident, that is, the symptom was “an expression of a psychological conflict or need.” In the DSM-IV, DSM-5, and DSM-5-TR, conversion disorder was again restricted to symptoms affecting the voluntary motor and sensory systems and where there was incompatibility between symptom and neurologic or medical conditions.

All somatic symptom disorders share in common and are characterized by physical complaints lacking known physical basis or demonstrable physical findings where there is significant distress and impairment. These disorders are conditions characterized by physical symptoms suggestive of but not fully explained by a general medical condition or the direct effects of a substance. Symptoms are not intentionally produced and are not attributable to another mental disorder and cause distress or impairment in important areas of functioning. The disorders include illness anxiety disorder, conversion disorder, psychological factors affecting other medical conditions, factitious disorders, other specified somatic symptoms and related disorders, and unspecified somatic symptoms somatic symptoms and related disorders. The somatic symptoms may transiently help to alleviate the internal conflict and primary gain is established. Secondary gain may also occur, allowing the patient external reinforcement by escaping from the situation, adopting a sick role, and so on.

Once a medical condition has been carefully ruled out, hypnosis can be an effective way to reduce the intensity of the symptoms and give the patient back some control over their discomfort. In particular once the symptom is reduced and a therapeutic alliance is enhanced, the clinician can then use other hypnotic interventions to uncover the traumatic or nontraumatic conflict that precipitated the conversion symptom. Wain demonstrates this approach by teaching nonepileptiform patients to allow their fingers to have the tremors and then he begins to look to uncover the causes.

A 19-year-old male is hospitalized after having had a concussion but continued to have nonepileptiform episodes. Upon evaluation he is a mid range to high hypnotic subject who initially was able to obtain symptom substitution with his right index finger gaining tremors as he began sensing he was going to “pass out.” As he gained control we were then able to use age regression to take him back to the injury and he described the comfort he had in escaping from some of the bullying he was receiving. He was able to use self-hypnotic techniques to interrupt the episodes and then to eventually work through the problem and find other coping strategies and deal more effectively with ridicule.

Exploratory hypnotic techniques, such as age regression and affect bridge, are generally used in helping the clinician and patient to identify and understand the nature of the conflict. Procedurally, one would use trance phenomena to take a patient to a safe place prior to going to the conflictual area. This establishes a safety valve for the patient and models a way of moving away from the conflict when anxiety becomes overwhelming. Thus reassured, the patient can face the conflict but transiently find safety without having a dystonic symptom. It further can expedite the attainment of insight, facilitate mastery, and allow for a more rapid working through of the conflict. Helping the patient to find and understand the conflict that perpetuated the need for the symptoms can be expedited. The use of age progression can then also help patients to see themselves facing a similar conflict in the future with the added benefit of having a rehearsal to facilitate success and a more appropriate way of responding. While the patient is resolving the conflict, hypnotic techniques of symptom substitution can be used that are more appropriate—for example, shaking fingers versus having an entire body tremor can be extremely useful. Extinguishing the symptom without exploring the reason for its presence at times can enhance the level of anxiety and/or depression and lead to other potential acting-out behavior.

Hypnocognitive methods are also used to modify learned behavior. Remembering the symptom is beneficial for one's protection and its ablation can be destructive to one's homeostasis.

A 22-year-old male patient was brought to the emergency room with bilateral blindness. Following an evaluation by ophthalmology, it was determined that the blindness was psychogenic. After initial evaluation by psychiatry, a therapeutic alliance was developed, and hypnosis was used to take the patient to a safe place and then back to the time immediately prior to the blindness. After two sessions the patient was able to describe seeing his wife in an adulterous relationship. At that moment the patient vocalized a desire to harm his wife and her suitor. Immediately after this vocalization, he became amnesic for the event and blind. On describing this under hypnosis, he was given a suggestion that when he became alert "He would only remember what he felt comfortable remembering." Subsequent to the patient becoming alert, he had no idea what had occurred, and each day after the hypnotic intervention was initiated the patient's anger was reframed. When the patient felt comfortable he then confronted his wife. The patient became aware that the amnesia was being used to prevent him from acting out. Use of a psychodynamic, cognitive reframing approach with a hypnotic milieu helped this patient to gain control and understanding of his symptoms. The patient and his wife were then referred for marital counseling.

A 30-year-old male with a history of bipolar disorder who presented to neurology for a worsening motor tics which included rhythmic flexion of the left wrist and hand, eye blinking and grimacing, left-sided lower extremity repetitive kicking movements, and repeated jerking of the neck and shoulder left upper extremity tremor. The patient underwent a full neurologic work-up which was negative and a trial of medications was without result. After evaluation the patient was screened for hypnotic capacity and was determined to be a midrange subject. He was taught to be sensitive to his physical cues and utilize a hypnotic intervention to curtail symptoms. Following the first session the patient had an immediate reduction in the frequency and amplitude of his tremor. He then practiced the hypnotic technique and was motivated to understand the significance and etiology and gain greater control of the symptom. With a psychodynamic approach the patient recognized the tic was his defense against the anxiety of being shot. And the tremors were his way as a child of expressing his fear responses especially when he perceived his father or other adults might hurt him. Following treatment the patient was able to control his symptoms and the initiation of the hypnosis gave him a sense of confidence to explore more of the psychological conflicts that he had faced.

A 29-year-old woman was referred for evaluation and treatment of ongoing facial pain that was not responding to traditional methods of intervention. Neurologic evaluation showed no objective physical correlations. Her high mid-range performance on the HIP added support to the potential of a psychological mechanism for the pain. Initially the pain was controlled by a hypnotic intervention, but it returned 24 hours later. Her self-hypnotic technique ceased to be effective. A decision was made to explore more completely the meaning of the pain. Age regression under hypnosis was used, and the patient was regressed to a time prior to the pain.

She related that her brother had been injured by a car while he was running in the street. The patient was babysitting at the time, and her father was so angered that he hit her. Recently her friend’s dog ran away, and she felt responsible. As she began to recognize her need to punish herself because of her guilt over what had occurred, she was able to understand her feelings and reframe her thoughts in a more productive manner. An “affect bridge” was also used, and the patient was asked to go to back to a previous time when she fell guilty and was punished. She then was able to describe her feelings of being hit by her alcoholic, abusive father. She continued to gain insight and mastery over the past and was able to ablate her pain.

The foregoing vignettes demonstrate a variety of somatic disorder cases where hypnotic techniques were used to help to resolve the patients gain control over the symptom and resolve the conflicts. They further demonstrate how hypnosis can be used as adjunctive psychological therapy together with reliance on pharmacotherapy, CBT, psychodynamic therapy, and/or eye movement-desensitization-reprocessing (EMDR). Wain advocates the need for the clinician to be viewed as safe and to facilitate the development of trust and a therapeutic alliance to enhance the success of treatment. Hypnotic intervention can be introduced after these needs have been met. The clinician further needs to avoid being impatient and to recognize the consequences of immediate demands for the resolution of the conflict.

Hypnotherapy and Relational Disturbances

There are two broad categories of relational problems that patients bring to psychotherapy—problems *with* relationships, or attachment pathology, and problems *within* relationships, or focal, recurrent conflictual relational themes played out in most relationships—called core conflict relational themes (CCRTs). Hypnosis is useful for both types of relational disturbance.

Attachment disturbance is assessed in children between 12 and 20 months using the Strange Situation Paradigm, which entails direct observation of an infant and mother in an unfamiliar play room. Children are classified into four attachment types—secure, avoidant (dismissing in adults), resistant (anxious preoccupied in adults), and disorganized. Longitudinal studies have shown that these attachment types are relatively stable and predict preschool, adolescent, and adult attachment behavior. Base rate studies consistently have shown that about 65% of the general population has secure attachment, and about one-third of adults in the general population in contemporary Western culture have one or another of the three forms of insecure attachment.

Use of the Adult Attachment Inventory shows that anxious preoccupied attachment or disorganized attachment is strongly associated with borderline personality disorder. Disorganized attachment is associated with a range of psychopathology. Disorganized attachment combined with physical and/or sexual abuse in later childhood is associated with the development of dissociative disorders in adulthood. Psychodynamic psychotherapy with personality-disordered patients with attachment pathology is based on the interpretation of defenses against attachment along with the provision of a stable, consistent therapeutic attachment relationship that serves as the basis of the development of new internal working models for secure attachment. However, research with the Strange Situation Paradigm in infants shows that it is the *representation* of attachment between 12 and 20 months when symbolic or representational thinking develops, not an attachment relationship per se, that has an organizing effect on lines of development like affect regulation and self-development.

Therefore, hypnotherapeutic imagery focused on the development of positive representations of secure attachment more directly addresses the central problem of attachment pathology, namely the failure to develop positive stable internal representations of a secure base. Using an object relations model, Elgan Baker first articulated the use of hypnotic imagery to visualize the therapist as a “good object” and then to visualize interactions between the patient and therapist in which the therapist is seen a nurturing, supportive figure.

Daniel Brown’s hypnotherapeutic approach to treating attachment pathology follows John Bowlby’s model that the establishment of a secure base is naturally followed by exploration. Structured hypnotic visualizations of ideal parent figures are used to develop new, positive internal representations of secure attachment.

An outcome study is underway using this protocol with personality- and dissociative-disordered patients with the pre/post scoring of the Adult Attachment Inventory as the main measure of change in attachment status. In the initial run of the outcome study Brown et al., compared 22 adults treated for 1 to 3 years with the Three pillars attachment-based protocol to 25 control adults had taken a 2-month class on a psychoeducational approach to attachment, but no treatment, followed by 1 to 3 years of mindfulness training. A comparison of post-treatment attachment status found that all subjects in the treatment group had achieved secure attachment, whereas many in the control group had not, a treatment effect significant at the .001 level. Significant effects also occurred for an increase in overall coherence of mind and mentalization. A second group of treatment subjects are now being analyzed. Eric Spiegel classifies hypnotic treatments for attachment into three domains—those that emphasize attunement, those that foster positive internal working models for attachment through having the patient repeatedly represent it, and those that foster metacognitive skills or mentalization.

CCRT maps are more complex. They develop between the third and fourth years of life at a time that the child has developed the capacity for emotional ideas, wishes, complex schemas and beliefs, and psychological defenses. Whereas attachment representations develop from 12 to 20 months at an age prior to the development of narrative memory, CCRT maps develop concurrent with the development of narrative memory and are therefore subject to interpretation in dynamic psychotherapy.

The typical approach to modifying CCRT themes through dynamic psychotherapy takes the following steps:

1. Formulating the CCRT accurately in terms of the underlying wish of what the given patient is looking for in each relationship, the negative reactions of others expected, and the maladaptive reactions of the self in response to not having this wish fulfilled.

2. Making a test interpretation so that the patient comes to see how this pattern is operative in most relationships—past and current—and in the relationship with the therapist.
3. Interpreting the manifestation of the CCRT elements in the here-and-now transference to the therapist. By adopting an emotionally corrective stance, the therapist tries to alter the structure of the CCRT. Some methods are designed to change the reactions of others, whereas others focus on changing the reactions of the self. Outcome studies show that reactions of others and reactions of the self-elements change in roughly 30 to 50 hours of dynamic psychotherapy that consistently focuses on CCRT themes.

Hypnotherapy can play a useful role in CCRT treatment. Hypnoprojective methods can assist in formulating the CCRT ("The scene will somehow be about the kind of problems that keep coming up in relationships"). Structured hypnotic visualizations can assist in having the patient articulate the best aspects of the central wish and then imagine ideal reactions of others and healthy adaptive reactions of the self toward the goal of developing a new, positive CCRT internal map ("The scene will be about the kind of relationship you wish for, in the best sense," "The scene will be about the ways others respond, in the best sense"). There are no outcome studies specifically on the use of hypnotherapy for CCRT problems. Carolyn Daitch has developed a short-term hypnotherapeutic protocol for anxiety in the context of romantic partnerships.

Hypnosis and Anxiety Disorders

Anxiety disorders can be classified in terms of whether the anxious arousal is more or less continuous, as in GAD and OCD, or discontinuous or episodic. Disorders of discontinuous anxiety can be further classified in terms of whether the source of the anxiety is external, as in specific phobia, or internal, as in panic. Unwanted thoughts are also characteristic of continuous anxiety conditions, with excessive worry about realistic situations being characteristic of GAD and misappraisal of the importance of unwanted, senseless obsessional thoughts being characteristic of OCD. Learned progressive avoidance of unwanted, senseless thoughts makes such thoughts salient and persistent and tags these thoughts with excessive anxiety, thereby justifying the progressive use of compulsive-behavioral and mental rituals to mitigate the anxious arousal.

Learned progressive avoidance of discomforting bodily sensations and/or unfamiliar situations loosely associated with an increased probability of having panic attacks gradually transforms panic attacks in agoraphobia. Furthermore, while it is well established that exposure-based treatments are effective, each type of anxiety disorder requires a specific type of exposure treatment. For example, because excessive worry generates continuous arousal in GAD, worry exposure in the best-matched treatment. Because discomforting bodily sensations generate panic, exposure to such uncomfortable sensations is critical to successful treatment of panic, as exemplified by fear-of-fear exposure, interoceptive exposure, and panic control therapy (PCT). Because fear of external objects or situations is central to the development of specific phobias, exposure to the feared situations in fantasy and in real life through systematic desensitization is critical to the treatment of phobias.

Because compulsive rituals eradicate anxiety, behavioral exposure and response prevention (ERP) aids the patient to tolerate the anxiety and master it without resorting to the compulsion. ERP alone, however, does not directly address the obsessive thinking. Restructuring core beliefs that cause the misappraisal of senseless, unwanted obsessive thoughts is the best approach for working with the obsessive-thinking component of OCD.

Most clinicians concur that hypnosis is not a treatment; it is the medium of treatment. Hypnosis is merely a state of redistributed attention and heightened focus and, for some patients, an altered state of consciousness or trance that allows greater access to imagery, self-talk, affects, self-states, and inner coping resources than would be available in the normal waking state. However, it is still necessary to bring to hypnotherapy the protocol of treatment that is best matched to the given psychiatric condition. With respect to anxiety conditions, exposure-based cognitive-behavioral treatment protocols consistently show favorable outcomes.

Hypnotherapy can serve as a useful adjunct to the cognitive-behavioral treatment of anxiety disorders. J. Christopher Clarke and Arthur Jackson devoted an entire text to hypnotherapeutic treatment of anxiety disorders. With respect to specific phobias, hypnotherapy is a useful adjunct to desensitization treatment. Sharon B. Spiegel has provided an excellent review of the current issues in the treatment of a specific phobia and has made significant recommendations for innovative applications of hypnosis in the treatment of specific phobias. She fully describes the benefits of utilizing hypnosis in the treatment of specific phobias, and she has emphasized that hypnosis does significantly manage resistance to treatment in specific phobias. She has documented that the use of hypnosis does reduce the length of treatment, and she has demonstrated that hypnosis does provide substantial reduction in anxiety, based upon the significant advances in cognitive neuroscience. The therapist and patient construct a stimulus hierarchy from least- to most-feared situations. Exposure consists of imaginal exposure to each situation on the hierarchy, from least to most, first in fantasy and then in real life. Hypnotic desensitization has advantages over nonhypnotic behavior exposure, in that hypnotic treatment entails greater focus, less distraction, richer affective imagery, a greater capacity to control the exposure imagery, enhanced sense of learning and mastery through time distortion suggestions, and greater likelihood of facilitating effective real-life exposure through posthypnotic reinforcement.

Valentine et al., conducted a meta-analytic study of 15 studies on hypnotherapy for anxiety disorders. They concluded “the average participant receiving hypnosis reduced anxiety more than about 79% of control participants” (p. 336). Comparative outcome studies of nonhypnotic behavioral exposure and hypnotic desensitization illustrates that both approaches have comparable treatment effect sizes but that the hypnotic treatment is typically faster for the aforementioned reasons.

With respect to panic, the first proven exposure-based treatment was interoceptive exposure. David Barlow and his associates reasoned that because the escalation of panic states starts with the misappraisal of uncomfortable bodily sensations, exposure to uncomfortable bodily sensation was a core ingredient in successful treatment. In interoception exposure, the patient is given instructions to produce analogs to the kind of bodily discomfort operative in panic states, such as twirling the head around rapidly until becoming dizzy, running in place until becoming winded, or holding the breath for longer periods until becoming uncomfortable.

The rationale of interoceptive exposure is that the patient learns to sit through and master uncomfortable bodily states. Barlow more recently developed a more comprehensive treatment package for panic disorder—PCT—which includes interoceptive exposure, cognitive reframing of panic escalating cognitions, and controlled breathing. Because panic is a rapidly escalating state exacerbated by hyperventilatory breathing, learned control over breathing serves to de-escalate the panic state. Hypnotherapy can serve as a viable alternative to PCT.

Interoceptive exposure to uncomfortable bodily sensations is an indirect form of exposure because the bodily discomfort generated by head twirling or breath holding is not equivalent to the bodily discomfort naturally occurring during panic states. In contrast, hypnotherapeutic exposure using a fear-of-fear hierarchy is more direct. The clinician constructs a fear-of-fear hierarchy of least to most intensity of escalating panic experience.

In hypnosis, the therapist suggests the progressive intensification of the panic state from very mild, to moderate, to full intensity, and then instructs the patient to sit through the full discomfort until it diminishes and mastery over the discomfort is attained. The hypnotherapeutic fear-of-fear exposure more closely matches the sequence of cognitive and somatic events operative in the development of an actual panic state than does interoceptive exposure.

Another interesting disorder vocal cord dyskinesia (VCD) in the past has often been diagnosed as a conversion disorder rather than under the anxiety spectrum. This disorder is often known to be an asthma simulator and is characterized by upper airway obstruction and stridor without an anatomic cause. The disorder is highlighted by inappropriate inspiratory vocal cord adduction seen on fiber optic laryngoscopy. This can be distinguished from bilateral vocal cord paralysis by episodic, momentary abduction of the vocal folds during cough or patient distraction. Arterial blood gases values remain normal. These patients rarely exhibit hypoxia or hypercarbia. In addition, chest radiographs do not show hyperinflation or other immune functioning as is characteristic of asthmatics.

A 26-year-old female with a dx of VCD was referred from pulmonology for evaluation and treatment. She described an incident of sexual abuse at age 16 and had no previous symptoms or recurring thought intrusions or nightmares. She relates her present symptoms began after running and then beginning to gasp for air. Initially the patient is diagnosed with asthma and antiasthmatics are prescribed but all treatments she was refractory to and the symptoms persist. Following laryngoscopy a diagnosis of VCD was determined. The patient reported a successful college and vocational experiences and initially denied any psychological concerns. During the initial evaluation a HIP screening is utilized and she is a mid- to high-range subject. As the arm levitates in the posthypnotic section she then looks with amazement at the involuntarism of the arm and she is given the interpretation that her arm is responding the way other parts of her body can respond as well.

She is then able to recognize what may be occurring with her larynx. She is taught a self-hypnotic procedure to gain control over her symptoms. Whenever she feels tightness in her chest she is instructed to do her self-hypnotic procedure. In later sessions she begins to acknowledge that she is feeling choked by the stress in her present life and recalls a similar sensation when she was choked during a past abuse episode. She begins to acknowledge the feeling of having been held down and overwhelmed. As frequently occurs she begins to discuss more of her past and present trauma and is able to reframe and gain control. An affect bridge is used to help her master her feelings of the past as well as the present. As the intervention progresses the patient slowly begins to acknowledge her difficulty with her present position and her desire to terminate her job. However, she feels compelled to remain in the position because of her contract. She gains motivation to be able to protect herself and her symptoms begin to dissipate. One year after her last treatment there still were no return of symptoms and she was able to regain control. Wain and his colleagues at WRNMMC utilize this approach with many patients with VCD who are referred for treatment. After a pulmonary evaluation and a therapeutic alliance are developed, we initially teach them symptom control then help them understand the underlying conflict.

With respect to GAD, worry exposure directly addresses the excessive worry that generates sympathetic arousal. The patient first constructs a list of favorite categories of worry. The patient is instructed to intentionally worry about something common on the list, then to think about something worse, then something worse, and so on. Worry exposure is systematic worst-scenario thinking. After a while the content of worst-scenario thinking becomes so extreme that it fails to generate physiologic arousal, and the link between excessive worry and physiologic arousal is broken. There are a number of hypnotic approaches to GAD. Suggestions for relaxation, cue-induced soothing imagery, and the clenched-fist technique address the somatic arousal characteristic of GAD. Direct ego-strengthening suggestions to experience a worry-free state and/or hypnotically assisted worry exposure address the excessive worry characteristic of GAD.

With respect to the treatment of specific phobia, hypnotizability is a strong predictor of hypnotic treatment of flying phobias. Hypnosis reduces the over length of treatment, optimizes relaxation in exposure-based treatments, and builds self-regulation skills in the treatment of phobias. Pre/post neuroimaging studies of patients with specific phobia showed an MPFC top-down dampening of the fear center of the brain, the amygdala (Halsband and Wolf).

With respect to OCD, hypnosis can aid ERP. Hypnotic suggestions are given to intensify and then tolerate the anxiety associated with ritual compulsive behaviors such as checking or washing until a marked reduction in the frequency of compulsions is achieved. Modification of core beliefs that drive the misappraisals of senseless, unwanted thoughts and transform them into obsessions is best accomplished by hypnotic ego state therapy.

For example, if the patient reports a core belief of excessive responsibility, ego state therapy is used to identify the adaptive function of the overly responsible “part of the mind.” Then suggestions are given to contact the part of the mind that is most opposed to the overly responsible part. By accessing that counterpart to the overly responsible part, cognitive dissonance is created. For example, a spontaneous playful part of the mind might be identified as a counterpart to the overly responsible part.

Having heightened the dissonance by calling forth the spontaneous, playful part, the therapist uses conflict resolution strategies to reduce the dissonance between the overly responsible and spontaneous parts. Other parts are called on to offer a solution to the conflict between the overly responsible and spontaneous parts. To reduce the dissonance, the system typically shifts away from the extreme position of the overly responsible part. A similar approach is used for modifying any of the other core beliefs operative in maintaining the OCD symptoms until a marked reduction in unwanted obsessive thoughts is reported.

Hypnosis and Depression

Depressive disorders are among the most frequent referrals in psychiatry. The National Institute of Mental Health (NIMH) outcome study on depression showed that antidepressant medication and interpersonal therapy were equally effective, and CBT was modestly effective, in the treatment of major depressive disorder. Assen Alladin studied hypnotherapy combined with CBT as compared to CBT alone in treating depression in a 16-week randomized clinical trial. He found that although both groups significantly improved, the treatment effect size was significantly greater in the hypnosis group, and that hypnosis had an additive effect on traditional CBT treatment of depression. Miaozun Shih and associates conducted a meta-analysis of hypnotic treatment of depression showed a clinical significant effect of hypnotherapeutic treatment that significantly improved symptoms. In hypnosis, depressed patients can be given ego-strengthening suggestions for increased confidence and control, suggestions for positive affective states, and rehearsal of positive, hopeful future events to help the patient to come out of the hole in which they have fallen. Severe depression sometimes can interfere with concentration, thereby challenging the clinician to keep the patient focused on developing positive states and preventing a sense of failure. Hypnosis can be used as an adjunct to pharmacologic approaches with depressed patients.

Hypnosis and Personality Disorders

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, text revision (DSM-5-TR) defines a personality disorder as follows: A personality disorder is an enduring pattern of inner experience and behavior that deviates markedly from the norms and expectations of the individual’s culture, is pervasive and inflexible, has an onset in adolescence or early adulthood, is stable over time, and leads to distress or impairment.

A variety of nonhypnotic approaches to the treatment of personality disorders have evolved, ranging in diversity from behaviorally oriented DBT, cognitively oriented schema therapy, and dynamically oriented transference-focused psychotherapy (TFP) to developmentally informed psychotherapy that stresses attachment repair mentalization, self-development, and affective development. Although Marsha Linehan's DBT enjoys great popularity in the United States, outcome studies merely show that patients with personality disorders treated with DBT tend to stay in therapy and have less self-destructive tendencies over time relative to patients in conventional therapy, but no data are available to document whether DBT leads to the clinically effective result of no longer meeting a personality disorder diagnosis. Likewise, Otto Kernberg's TFP shows some promise in diminishing personality disorder criteria, but in comparative outcome studies developmentally informed treatments like Peter Fonagy's mentalization-based treatment (MBT) and cognitive-based treatments like schema therapy consistently show better outcomes than DBT or TFP in terms of numbers of patients who no longer meet sufficient criteria for a personality disorder posttreatment. More recently, outcome measures have included not only personality disorder diagnostic criteria, but also developmentally informed outcomes, such as coherence of state of mind and metacognitive capacity. Such measures are used in the attempt to demonstrate that patients with personality disorder are actually changing the organization of their state of mind in some fundamental way through treatment.

Hypnotherapy holds greatest promise as an adjunct to cognitively based approaches like Jeffrey Young's schema therapy and developmentally informed treatments. The Young Schema Questionnaire serves to identify on which of 18 maladaptive schema domains a given patient with personality disorder measures high or very high. Treatment then targets each of these areas using both traditionally cognitively focused methods, such as testing the evidence and correcting distortions, and emotion-focused methods. The main emotion-focused method entails a reparenting approach in which the therapist adopts an emotionally corrective stance. For example, if the patient characteristically expects abandonment, the patient is asked to imagine the therapist as being fully there, or if the patient expects deprivation, the patient is asked to imagine the therapist as providing the kind of care that is most needed. This emotion-focused technique is conducted in a relaxed state with structured waking imagery. A recent comparative outcome study showed that twice as many patients with personality disorder in schema therapy improved as did matched patients with personality disorder in TFP. Young believed that a main component of the success of schema therapy was the emotion-focused reparenting imagery work.

Although Young did not use hypnosis in schema therapy, there is no reason that his emotion-focused technique could not be adapted to using hypnotic imagery. There may be several advantages using hypnosis, especially with hypnotizable patients with personality disorder.

Although patients with personality disorder have a variable self-observational capacity, the hypnotic context may provide a necessary ingredient within which such patients can better resist distract, sustain focus on the emotion-focused imagery, and more quickly and efficiently alter the target maladaptive schema. Hypnotherapy can also serve as a useful adjunct to a developmentally informed approach to the treatment of personality disorders. The work of Elgan Baker and of Daniel Brown illustrates how structured hypnotic imagery can be used to develop positive, stable internal representations of secure attachment.

With respect to self-development, Janine Stevenson and Russell Meares obtained favorable results in reducing a personality disorder diagnosis using Fred Pine’s approach to fostering self-development. Hypnotherapy can also be used to foster self-development. The normal line of self-development includes the development of a cohesive self-representation (self-definition), the representation of the self as having an effect on the physical environment and on unfolding events and on eliciting desired responses from others (self-agency), and the linking of positive affective states to the self-representation (self-esteem). Patients with personality disorder with deficits along the line of self-development, such as borderline patients, show deficits in self-definition and agency, and those such as narcissistic patients show deficits in self-esteem development. Repeated structured visualizations are introduced in hypnosis to foster self-definition (“the scene will be about . . . when you are most you . . . your real self . . . the unique qualities that manifest when you are being most yourself”); self-agency (“the scene will be about . . . making an impact on . . . eliciting the kind of responses from others that you need”); and self-esteem (“the scene will be about . . . some situation in which you feel especially good about yourself . . . especially like yourself . . . feel especially confident”). Hypnotic suggestions are used to generate the target scene, amplify the positive affect associated with the scene, extend the duration of experiencing the target state of mind, and then transferring that state of mind to situations in which it would normally feel difficult to hold that sense of self. The structured visualizations are repeated in treatment sessions and outside of the sessions in self-hypnosis until they foster a stable, positive internal representation of self-definition, self-agency, and/or self-esteem.

With respect to affective development, hypnotherapy can be used to link the visceral substrate of affective states to imagery and thought in patients with personality disorder who show developmental deficits in the capacity for affect experience. Hypnotically suggested soothing imagery and/or hypnotically structured exposure (hypnotically suggested progressive intensification of affective states so that the patient learns to sit through them) can be used in patients with personality disorder who show deficits in affect tolerance. There are no outcome data on developmentally informed hypnotherapeutic approaches to the treatment of personality disorders.

HYPNOSIS AND PAIN

An excellent review of controlled studies and anecdotal reports concerning research and clinical applications of hypnosis in treating clinical pain. David R. Patterson and Mark P. Jensen’s excellent review of evidence-based studies of the use of hypnosis in assessment and treatment of pain disorders. Since the publication of that review, there have been other significant studies documenting that hypnosis has become an important method of intervention in the treatment of acute pain disorders, pain during medical procedures, surgical procedures, and chronic pain disorders.

Hypnosis in Acute Pain Disorders—Procedural Pain Interventions

Over the last 10 years, there has been increasing use of hypnotic interventions in the treatment of acute pain disorders and the alleviation of pain during medical and surgical procedures. There have been evidence-based studies document the efficacy of hypnotic interventions in these contexts, and these studies correspond to research findings regarding hypnotic analgesia in the hypnosis laboratory setting. Kendrick et al., have conducted a critical review of 29 randomized clinical trials on the use of hypnosis for procedural pain. The conclusion was that "hypnosis was most effective in minor surgical procedures" (p. 75).

Hypnosis and Burn Care. Hypnosis has been used in the analgesic management of burns. These hypnotic interventions include suggestions with progressive muscle relaxation, followed by suggestions for direct analgesia. Imagery for cooling the inflammation and suggestions for reappraisal of feared bodily harm also have been used. In the various reports in the literature concerning the use of hypnosis in the management of burns, the hypnotic interventions appear to have been efficacious for both debridement of burns and changes of dressing involving burns. These controlled studies have also documented that the hypnotic interventions had significant effect on pain perception, anticipatory anxiety, and relaxation during and after burn care.

In addition to Theodore Barber's rapid induction analgesia method of hypnosis, other hypnotic interventions have used eye fixation associated with progressive muscle relaxation, guided imagery, and direct suggestions for pain relief in the affected area. Overall, both clinical reports and controlled studies have demonstrated that hypnosis is associated with reduced use of opiate medications, decreased pain ratings, and reduced use of sedative–hypnotics and tranquilizers. Many of these studies included assessment of hypnotizability using the HIP and the Stanford Hypnotic Clinical Scale (SHCS).

Roberts et al., reviewed the literature and described how hypnotic intervention is an effective tool at reducing pain and procedural anxiety with burn patients.

Hypnosis in Obstetrics and Gynecology. In view of the fact that labor pain represents a significant acute pain and that women are certainly motivated for both reduction in pain and diminished anxiety, hypnosis has been used as an intervention in the acute pain associated with labor and delivery. Another benefit reported in both anecdotal reports and controlled studies involves the reduction in stress and anxiety in the labor and delivery process. This reduction in stress and anxiety provides a better maternal environment during both stage II and stage III of labor. The literature provides numerous anecdotal reports concerning the use of hypnosis in the labor and delivery process.

Ralph August and Simon Chaisson described using hypnosis as a sole anesthetic for cesarean births. Karl Fuchs in Israel described the successful use of hypnotic intervention with hyperemesis gravidarum.

Hypnosis and Women’s Issues

In a study by Karl Fuchs, he demonstrated that hypnotic intervention can be very successful for hyperemesis gravidarum. Out of 51 gravidas patients treated individually, 35 (68.5%) showed excellent results with no vomiting and no nausea, and 2 (3.9%) showed good response and had no vomiting but some nausea. Fourteen (25.5%) patients showed vomiting less than six times per day and had improvement in their nausea and poor response to treatment, and one patient (2.3%) could not be hypnotized in spite of repeated trial sessions. Out of 87 gravida patients treated in groups, excellent results were obtained with 61 patients (70.1%), and 24 patients (27.6%) had no vomiting but some nausea, and 2 patients decreased their vomiting to less than six times per day but had improvement in their nausea.

Gary Elkins and his associates, as well as Cassie Kendrick and associates, have used hypnosis in the treatment of hot flashes in premenopausal women. Maitri Shah and associates found that both hypnosis and medications were equally effective for treating dysmenorrhea.

Hypnosis in the Emergency Room

Past myths about the nature of hypnosis may have contributed to a reluctance among many to use hypnosis in the emergency department. Some feel that hypnosis takes too long and would not be appropriate in acute care situations. The time for formalized trance induction may range from a spontaneous occurrence to several hours. However, given the focused state of the individual in an emergency situation, hypnotic induction is frequently quicker than under other conditions. Frequently, because of the discomforts of the presenting ailment, anxiety, and confusion pertaining to the extent and significance of one’s symptoms, patients in the emergency department feel as if they are out of control. However, their preoccupation with acute symptoms may decrease their peripheral awareness and thus narrow their field of concentration, enabling them to gain some control. This highly focused state is like a hypnotic trance state.

The psychological trauma associated with physiologic malfunctioning often precipitates an altered state of consciousness for the patient. Awareness of the existence of this altered state (trance or dissociated state) can enable the staff to use it in treatment. In this altered state, the patient may be highly receptive to comments and suggestions. The staff must be very cautious with the words and statements used within the patient’s hearing. The same guidelines should be followed with patients who have not been introduced to the formal hypnotic trance. Patients, by virtue of being in a focused state, are vulnerable to staff’s comments and assimilate the data rapidly and, at times, without vigilance. This is why a mere suggestion of help or of a positive outcome can be facilitative for a patient in general.

In the emergency department the staff typically adopts a positive, calm, supportive approach to the patient. This supportive style may provide the necessary ingredients for an enhanced therapeutic alliance, which can facilitate the patient’s ability to use their trance state more effectively. Many individuals use their trance state without being cognitively aware. The natural trance is frequently used by individuals as a protective measure, and it typically allows individuals to distance themselves perceptually from their discomfort, whether it is physical or emotional. The astute clinician can recognize a patient who is doing this and capitalize on this state. For patients who are having difficulty focusing constructively, the clinician can help to find an appropriate focus for them.

Paul Delfito described the value of hypnosis in the emergency department for treating pain due to orthopedic injuries, burns, corneal abrasions, abdominal distress, headaches, menstrual discomfort, renal calculi, herpetic lesions, muscle spasms, and dental syndromes. Isserson reviewed the impressive usage of hypnosis for clinical emergency care. He suggested that the lack of training has prevented more staff from using this technique.

Wain, working on a Consultation Liaison Psychiatry Service, described similar results with a variety of patients in the emergency department.

A 32-year-old man presented to the emergency department with a severe headache. He was a chronic migraine sufferer and had been unable to control the pain on this occasion with his propranolol. The emergency department recognized that he had a high hypnotic capacity. The imagery of an icepack being placed on his forehead was suggested. Initially, some real ice was placed on his forehead to help. The patient was able to control his pain completely with this imagery. He did not require narcotics, as he had on previous visits. On follow-up several weeks later, the patient reported being able to use this strategy to control, as well as prevent, migraine attacks, and he no longer had to rely on frequent emergency department visits for pain relief.

A 55-year-old female patient suffering from coronary artery disease presented to the emergency department with severe chest pain. Her cardiac enzymes and electrocardiogram (ECG) were completely normal. The treating clinician used hypnosis, during which the patient reported that her internist had died of a myocardial infarction the previous day. When it was clear that there was no actual cardiac involvement on this occasion, the patient was allowed to reinterpret her perception of the pain. She came to the realization that she had been identifying with her internist, and the pain was her way of grieving his loss. As she began to practice the hypnotic techniques, her pain dissipated. Although pain-free, she was nevertheless observed on a monitor for 24 hours. There continued to be no evidence of myocardial damage. She was discharged pain-free. She returned as an outpatient and again was symptom-free. She found a new internist and was monitored regularly.

Hypnosis for Alleviation of Pain in Bone Marrow Aspirations. Studies have demonstrated the efficacy of hypnosis in reducing the pain of bone marrow aspiration using such techniques as storytelling, fantasy, imagery, and deep breathing. Control studies demonstrate that hypnotic interventions with children provided lower ratings of pain than a control group.

Some studies also used specific hypnotic interventions involving posthypnotic suggestions for re-entering hypnosis following cues provided by the hypnotherapist. The hypnotic interventions used eye fixation, as well as progressive muscle relaxation, and the involvement of storytelling and guided imagery as methods of hypnotic intervention. The use of hypnosis for bone marrow aspiration in children is associated with the administration of the SHCS for Children. Ernest Hilgard and Samuel LeBaron's study of children's hypnotizability predicted analgesic response, as well as a reduction in anxiety.

Hypnosis in Surgical Procedures

There have been a number of studies and reports describing the use of hypnosis in surgical procedures. Esdaile perhaps highlighted its early use by doing over 300 major procedures between 1845 and 1851. Later studies include hypnosis used in hemorrhoidectomies, anorectal surgery, cesarean sections, cervical surgery, surgical removal of pedicle attachments, neurosurgical procedures, dental surgery, breast surgery, coronary artery bypass surgery, endocrine surgery, hernia repair, and angioplasty. These studies reported efficacy and positive effect on surgery, including reduced postoperative nausea and vomiting, decreased anxiety during the procedure, fewer postoperative side effects, reduced medication use, and quicker recovery time.

Wain published a paper on the use of hypnosis as a sole anesthetic in septoplasty (Fig. 32.19–2). A patient demonstrated total anesthesia during surgery, rapid healing, a decrease in bleeding and edema, as well as analgesia without pharmacology in changing the packing postoperatively. Hypnoanesthesia is not a sole anesthetic for all patients, however—probably only for hypnotic "virtuosos." There are many more patients who can use hypnosedation as an adjunct to chemical anesthesia, thus reducing reliance on the latter and decreasing risks. In his article on hypnosis for septoplasty, Wain reviewed many surgeries in which hypnosis was used as a sole anesthetic or as an adjunct to traditional anesthesia.

FIGURE 32.19–2. Photographs of septoplasty.

A 34-year-old man presented with a history of nasal obstruction for many years. He had an injury to the nose at age 12 years and had a continuous obstruction to breathing. He was found to have a grossly deviated septum and a small hump on the nose, which required removal. His physical examination revealed a well-developed, well-nourished 34-year-old man in no acute distress with normal vital signs. Physical examination was unremarkable except for the presence of a grossly deviated nasal septum to the right with total obstruction and total dislocation of the septum to the right nostril obstructing the airway, with compensatory hypertrophy of the inferior turbinates on the left side. Screening on the (HIP) scale resulted in him being a high hypnotizable 4 to 5. A septoplasty was performed under hypnosis. Although no chemical anesthesia was used, an anesthesiologist was in the operating room to ensure efficacy for the patient. Postoperatively the patient was able to change his packing without any pain. The strategy consisted in suggesting that he was absorbed in floating on or near a beach. He had no recollection of the surgery and reported no perception of pain. He was highly hypnotizable.

Hypnosis in Chronic Pain Disorders

In contrast to acute pain and procedural pain, chronic pain is a persistent and intractable pain disorder that persists beyond the time needed for healing to occur from an injury. Although chronic pain is sometimes associated with tissue damage, in the majority of cases tissue damage has not been substantiated.

According to the modified gate control theory of Ronald Melzack and colleagues, chronic pain is mainly a signal detection dysfunction in which a variety of haptic sensory inputs are misinterpreted as pain. Daniel Brown likened the development of chronic pain to the development of panic states, both of which entail a progressive learned misappraisal and phobic avoidance of somatic distress. Extensive phobic avoidance of pain sensations maintains the signal detection dysfunction, in that the patient rarely attends to somatic distress in a way that would allow for correction of signal detection errors. Chronic pain syndrome is also frequently associated with significant cognitive distortions and illness behaviors.

According to Wain, “Nowhere in medicine does there seem to be a stronger link between the physiological, psychological, and sociological variables than there is in the understanding of the diagnosis and treatment of the patient with chronic pain. It is the trauma of the pain, suffering, and anguish that unites every physician, nurse, and health care specialist."

Wain further noted that a chronic pain experience is associated with significant problems in coping with associated social and environmental factors. In contrast to acute pain disorders, interventions such as prolonged rest and immobility are frequently contraindicated in the effective management of chronic pain syndromes.

Hypnosis frequently is also used to address the illness behaviors and suffering associated with chronic pain. In *Hypnosis and Behavioral Medicine,* Brown and Fromm described a phase-oriented approach to hypnosis and chronic pain. The first phase entails identification of specific pain coping strategies available to the given patient with chronic pain, followed by the application of these pain coping strategies in trance, along a hierarchy from less intense, peripheral to more intense, target pain locations.

Hypnotic techniques are used to correct negative self-talk and distorted pain beliefs to the extent to which they interfere with the successful application of pain-coping strategies. The second phase is skill development, in which the patient with chronic pain develops mastery over the pain experience within and outside of the treatment session and learns successfully to use a range of pain-coping strategies whenever pain occurs in everyday life. At this stage, phobic avoidance of pain is reduced. The patient learns not only to tolerate pain adequately but also to discriminate a wide range of sensory signals other than chronic pain. The third stage, the rehabilitation phase, addresses the issues of chronicity, for example, the normalization of daily life, developing a healthy, pain-free self-representation, and eliminating analgesic addiction.

Tiara Dillworth and Mark P. Jensen have provided an excellent review of controlled trials that have demonstrated hypnosis is an efficacious treatment of chronic pain. Their studies did review a number of controlled studies that have documented hypnosis is an efficacious treatment of chronic pain syndromes.

Multiple Convergent Pain Management. The Institute of Medicine published a landmark report in 1987 entitled "Pain and Disability: Clinical, Behavioral, and Public Policy Perspectives." The report, prepared by a committee of authorities, contained a literature review, an analysis of chronic pain and its treatment, and specific recommendations for future developments in medical research and public policy concerning chronic pain.

Although the Institute of Medicine study suggested the need for further research on the different modes of treatment and their efficacy, it stated that multimodality and multidisciplinary therapeutic interventions lessen the disability and dysfunction associated with chronic pain.

Ronald Melzack and Patrick Wall described multimodality and multidisciplinary therapy for chronic pain as a multiconvergent pain management program that includes physical, psychological (behavioral medicine), and pharmaceutical interventions and interventional anesthesia (Fig. 32.19–3). This type of pain management program uses a rehabilitation and collaborative model of intervention. The psychological interventions include several modes of treatment based on the principles of both operant conditioning and cognitive-behavior management. Hypnosis has a definite role in any treatment program that uses contingency management and cognitive-behavior management approaches to chronic pain. Certain features of the behavior of the patient with chronic pain demonstrate the value of hypnosis as a mode of treatment—specifically, demoralization, suffering, and low self-efficacy.

In addition to the aforementioned psychological problems associated with chronic pain, it is also associated with significant suffering based on the patient’s fear that they will be unable to control the pain. The suffering component of the chronic pain experience is also amenable to being ameliorated through hypnotic interventions.

In addition to the medical suffering associated with chronic pain, chronic pain is also associated with demoralization, which is amenable to hypnotic interventions. Because of the influence of cognitive processes on the perception and integration of pain and the significant complications regarding the psychological problems associated with chronic pain, interventions that significantly modify cognition and perception would be efficacious in the management of pain. This observation provides a firm foundation for using hypnosis in chronic pain management.

Hypnosis lends itself to both operant conditioning (contingency management) and cognitive-behavior modification therapies that are well established in multiple convergent pain management programs. In view of the fact that there is a nonspecific (placebo) response and a direct effect of the hypnotic experience, both the nonspecific and the direct effect of hypnosis can be used appropriately in the management of chronic pain syndromes.

There are significant anecdotal reports and some cases with controls that demonstrate that patients with higher hypnotizability scores on hypnotic assessment tests generally report greater analgesic effects of hypnosis and significantly improved cognitive functioning through the use of hypnosis. The use of hypnosis should involve multiple interventions during the course of the chronic pain management and self-hypnosis training. In their review “Hypnosis and Clinical Pain,” Patterson and Jensen provided significant evidence concerning the lack of control studies in chronic pain studies. They recommended that hypnosis should be used in a more organized manner in the management of chronic pain. They also suggested that hypnosis may be efficacious in the treatment of the suffering and cognitive distortions experienced by patients with chronic pain. Their review included an important meta-analysis suggesting that more than 70% of the patients in multiple convergent pain management programs could have benefited from adding hypnotherapy to the treatment program. Such interventions could be directed toward both the alleviation of the pain experience and the treatment of the suffering and cognitive distortions associated with chronic pain behaviors.

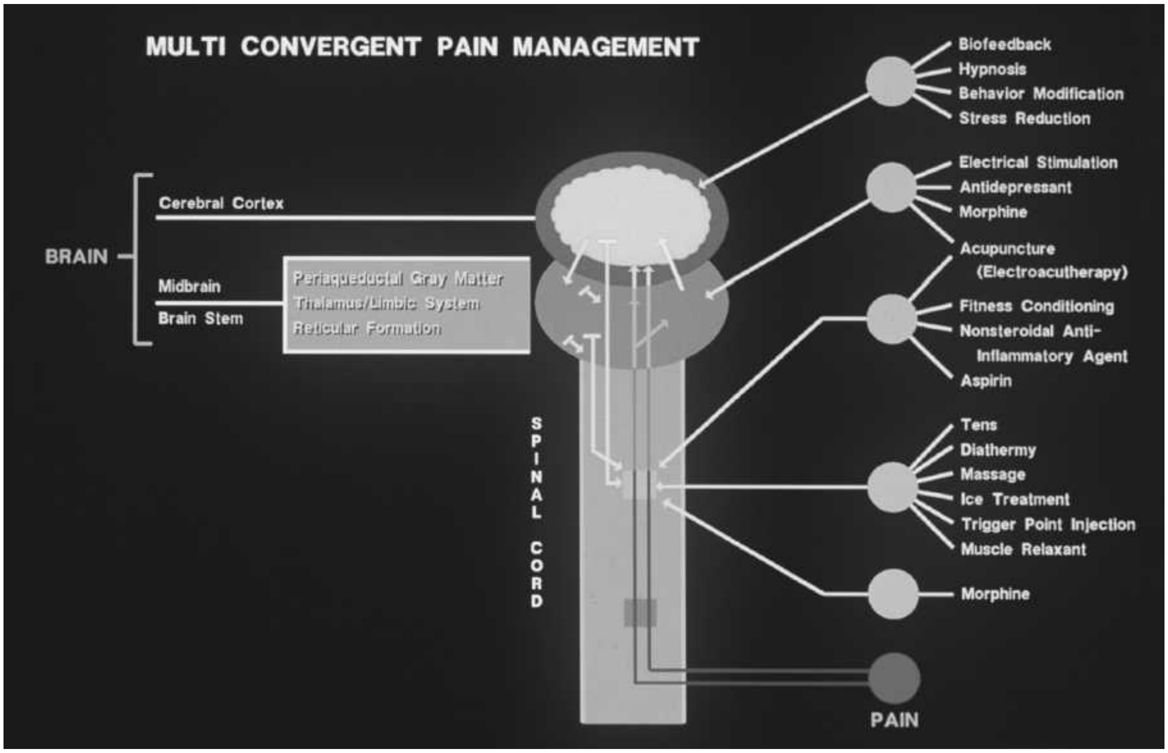
MULTI CONVERGENT PAIN MANAGEMENT
BRAIN
Cerebral Cortex
Midbrain
Brain Stem
Periaqueductal Gray Matter
Thalamus/Limbic System
Reticular Formation
SPINAL CORD
Biofeedback
Hypnosis
Behavior Modification
Stress Reduction
Electrical Stimulation
Antidepressant
Morphine
Acupuncture
(Electroacutherapy)
Fitness Conditioning
Nonsteroidal Anti-
Inflammatory Agent
Aspirin
Tens
Diathermy
Massage
Ice Treatment
Trigger Point Injection
Muscle Relaxant
Morphine
PAIN

FIGURE 32.19–3. Multiconvergent pain management involves multiple interventions that act on different parts of the central nervous system and spinal cord. These interventions substantially diminish the pain, suffering, and depression produced by the chronic pain experience. Hypnosis plays an important role in significantly reducing the pain, demoralization, and suffering of chronic pain. Hypnosis also facilitates the efficacy of biofeedback, behavior modification, stress reduction, and muscle relaxation in multiconvergent pain management. TENS, transcutaneous electrical nerve stimulation.

A more recent meta-analysis comparing hypnosis to behavioral methods for chronic pain Tononori Adachi and his associates found that hypnotic treatment of chronic (nonheadache) pain had a larger treatment effect size as compared to other psychological interventions like cognitive-behavioral treatment or guided imagery.

During the initial evaluation of such patients for hypnosis and self-hypnosis training, they should be given a full explanation of the benefits of this training for the management of their chronic pain, as well as a full understanding of the misconceptions and myths about hypnosis. They should also undergo a hypnotizability assessment, such as the SHCS for Adults or the HIP. David Axelrad developed a method of induction—open neurosensory fixation—for use in the individualized hypnotherapy and self-hypnosis training of the patient with chronic pain. This method, described in the paper “The Role of Hypnosis in Multiple Convergent

Management of Chronic Pain," includes engaging in eye fixation on a visual source, listening to ambient environmental sounds, experiencing kinesthetic and other pleasurable sensations in the skin, muscles, and joints, and engaging in progressive muscle relaxation (Fig. 32.19–4). This method can be used by patients with high, moderate, or low hypnotic capacity.

Patients with chronic pain should be provided with individualized hypnotherapy that incorporates suggestions for relaxation and progressive muscle relaxation, or some other form of relaxation intervention, that permits the patient to undergo relaxation responses from the hypnotic experience. The suggestions should also be directed toward direct pain reduction and pain relief.

In view of the significant association of depression and demoralization with chronic pain, the hypnotic interventions should also include suggestions that are directed toward alleviation of depression and demoralization.

Displacement suggestions, distraction suggestions, and transformation suggestions may also be employed during the course of the hypnotherapy. Following several individualized hypnotic interventions, the patient should also be provided with training in self-hypnosis in a group setting. The self-hypnosis training experience is an integral part of a hypnotic intervention for patients with chronic pain. During the course of a multiple convergent pain management program, patients should be encouraged to undergo self-hypnosis in a variety of situations that provoke increased pain. Such experiences reinforce the "well" behavior and help the patient to establish self-efficacy and sense of personal mastery.

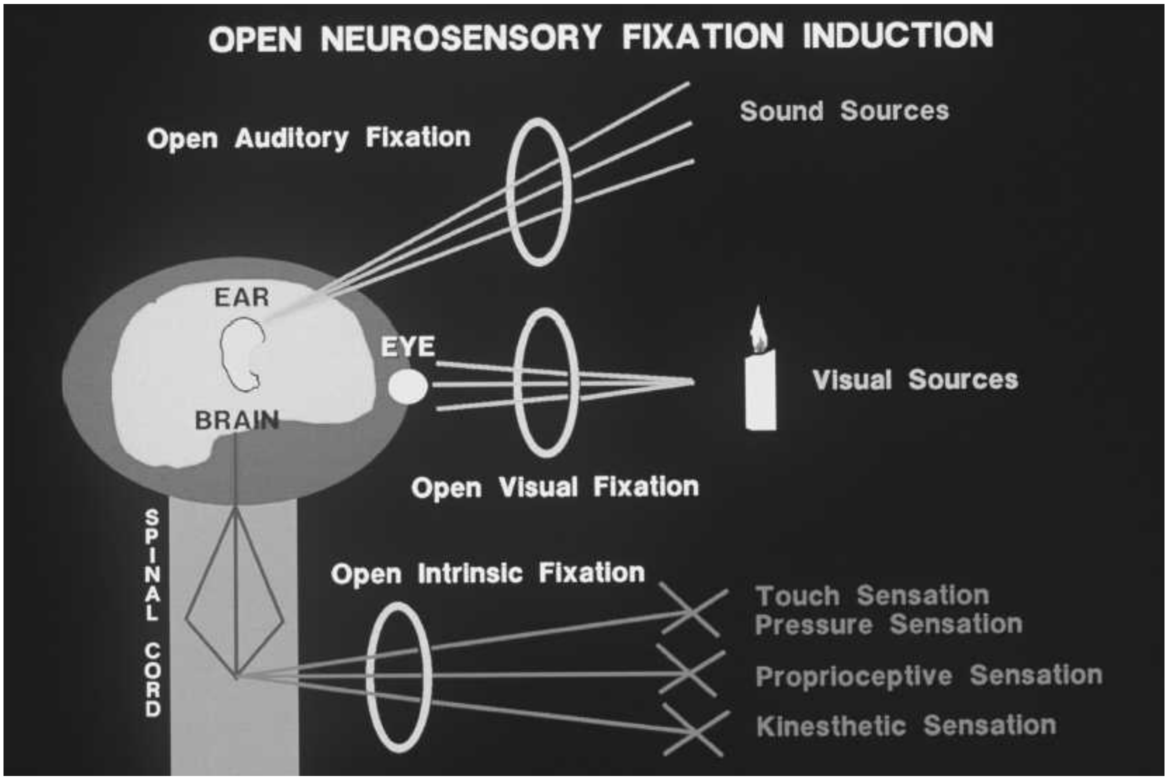
OPEN NEUROSENSORY FIXATION INDUCTION
Open Auditory Fixation
Sound Sources
EAR
EYE
BRAIN
Visual Sources
Open Visual Fixation
SPINAL CORD
Open Intrinsic Fixation
Touch Sensation
Pressure Sensation
Proprioceptive Sensation
Kinesthetic Sensation

FIGURE 32.19–4. Induction of a patient, using open neurosensory fixation induction.

Ms. D was a 63-year-old White woman referred for symptoms of chronic pain and PTSD caused by the experience of being a passenger in a motor vehicle that was struck by an 18-wheeler. Ms. D had been in the front passenger seat, restrained by a seatbelt and shoulder harness. During the accident, she experienced severe fright and the fear that she would be killed. She had sustained injuries to her left shoulder and neck, which required an emergency room assessment and evaluation. During the emergency room evaluation, a computed tomography (CT) scan of her neck and back reflected significant degenerative disease in her cervical spine, a disk protrusion in her cervical spine, and a degenerative disease in her lumbar spine. An MRI examination also documented these injuries. Subsequent studies also revealed a negative axial CT scan and a negative MRI of the brain. The patient's chronic pain was severe. Because of the patient's significant history of insulin-dependent diabetes and other problems, her orthopedic surgeon and internist recommended conservative pain management interventions. Ms. D also experienced an acute stress disorder, followed by PTSD, after her injuries. Her symptoms included recurrent nightmares and re-experiencing images of the motor vehicle accident. She also exhibited hyperarousal and a significant startle reflex. She reported marked avoidance behavior, avoiding freeway driving, and experiencing increased anxiety and panic attacks whenever a motor vehicle in which she was riding was approached by an 18-wheeler. She reported significantly increased anxiety in response to both internal and external cues reminding her of the motor vehicle accident. The patient underwent treatment for her PTSD, which included individual and group hypnotherapy and a self-hypnosis training program. She was also provided with relaxation exercises and biofeedback training. She also required multiple convergent pain medicine interventions, which included physical therapy interventions and behavioral interventions for her chronic pain syndrome. She was also provided treatment with an SSRI antidepressant. Over the course of her treatment, she was able to use both hypnosis and self-hypnosis training to significantly reduce her anxiety, abort her panic attacks, and significantly diminish chronic pain experiences involving her neck and upper back. This self-hypnosis training program also provided her with improved control over her posttraumatic headaches. Her orientation to the hypnosis experience included the administration of the SCSHS, on which she had a score of 5, demonstrating that she had marked hypnotic capacity. The focused open neurosensory induction program Ms. D included eye fixation on a visual object and environmental experiences (sound) and direction of her attention inward to neurosensory experiences, including kinesthetic sensations. She was instructed in diaphragmatic breathing, which is a critical component of most hypnotic inductions and relaxation or meditative experiences. The training program included suggestions calling for eye closure and progressive muscle relaxation. During the course of the patient’s training, she was provided with suggestions for diminishing her fear and anxiety associated with both internal and external cues reminding her of the accident, as well as direct suggestions for pain relief in her neck and back. She also incorporated these suggestions for reducing her headaches by imagining having a cold pack placed on her forehead and neck.

Over the course of Ms. D’s treatment, she experienced a significant remission of her PTSD symptoms and a significant reduction in her pain experiences, as judged by both the McGill-Melzack Pain Inventory and the Visual Analog Scale.

Mr. G was a 55-year-old African American man who experienced an occupational injury that caused a herniated disk at L4–5. He experienced intractable pain, and he failed to respond to conservative interventions. He underwent a laminectomy and fusion procedure for this injury. Unfortunately, his pain returned, and he was unable to return to full-time employment. Over the course of his medical history, he underwent four additional surgical interventions to his low back region. He was referred to a pain center because of intractable pain associated with significant tolerance to opiate medications. At the time of his admission to the pain program, he was requiring methadone at a dose of 60 mg four times daily, augmented with hydrocodone at 2 mg two to three times daily. Following his admission to the pain medicine program, he was administered the Stanford Hypnotic Susceptibility Scale, on which he had a score of 11, indicating marked hypnotic capacity. At the time of his entry into the program, he was seriously depressed and demoralized. He was provided with cognitive-behavioral management, self-hypnosis training, and physical therapy interventions, emphasizing aerobic conditioning and strengthening exercises. He was also treated with antidepressant medication to assist in the treatment of his depression and his chronic pain syndrome. Mr. G. was also provided with hypnosis and self-hypnosis training similar to that described in the previous case. He was able to use open neurosensory induction to significantly reduce the pain in his back, increase relaxation of his back muscles, and improve his self-image. The hypnosis was also directed toward enabling the patient to experience significant reduction in pain without the need for opiate medications. Based on his training, the patient was able to use self-hypnosis once or twice daily to significantly reduce his pain and discomfort. Over the course of his treatment, he was successfully withdrawn from opiate medications. He was discharged free of opiate medications was successfully rehabilitated through a state-supported rehabilitation program.

Hypnosis in Headache Management

As noted in the review by Patterson and Jensen, far more studies have focused on the use of hypnosis for headaches than for any other etiology of chronic pain. Their review included nine studies that involved some controls. The hypnotic interventions in the headache management studies lasted for significant periods of time, and the patients were provided with suggestions for relaxation, visual imagery techniques, verbal reinforcers, and direct suggestions for pain reduction. Patients were frequently provided with self-hypnosis training as a part of their headache management program. The headache management program was also combined with biofeedback training and other methods of autogenic relaxation. A controlled study demonstrated that patients provided hypnotic interventions experience significantly greater pain control than patients provided biofeedback training alone. The hypnotic interventions provided for patients in this study used eye fixation followed by suggestions for relaxation, analgesia, and numbness.

Many of the headache studies reviewed by Patterson and Jensen included hypnotic assessments and measures of suggestibility as a part of the hypnotic intervention program. The researchers, who adopted social influence models for hypnotic responding, also reported that patients provided with such interventions reported significant improvement of their headaches without the induction of a hypnotic state. In another systematic review Flynn found a significant hypnosis treatment effect specifically for migraine sufferers in a randomized trial.

Hypnosis and Irritable Bowel Syndrome

Functional abdominal pain (FAP) and IBS are highly prevalent in both childhood and adulthood. IBS entails gut motility changes and recurrent pain and is accompanied by a functional dissociation of awareness of somatic events combined with excessive worry about bodily experience. Since the pioneering work of P. J. Whorwell and his associates reported in *Lancet* there have been several dozen controlled, evidence-based studies that have demonstrated that hypnotherapy for FAP and IBS is a highly effective treatment both for children and adults. These studies show hypnotherapy to be superior to standard medical care or supportive psychotherapy for the reduction in pain and stabilization of gut motility. Some of these studies also elucidated both the physiologic and psychological mechanisms of the hypnotic effects.

One of the mechanisms of action of hypnotherapy involved changes in both colonic motility and rectal sensitivity. Hypnotherapy also reduced excessive worry and somatic dissociation in patients with IBS. In view of these studies, hypnotherapy—individual, group, and self-hypnosis training—is recommended for the treatment of IBS in both adults and children. Vivian Miller and Peter J. Whorwell developed a similar protocol in the treatment of inflammatory bowel disease (IBD), but studies in this area are only just beginning to appear.

Susanna Carolusson has provided an excellent review and significant applications of dynamic hypnosis in the treatment of IBS. She has provided a randomized controlled study of the use of hypnosis to the treatment of IBS.

HYPNOSIS AND ASTHMA

Asthma entails intermittent, variable airways obstruction due to chronic bronchospasm. It may also involve an ongoing inflammatory response. In the 1960s, the Hypnotherapy in Asthma Subcommittee of the British Tuberculosis Association conducted a series of randomized clinical trials using direct hypnotic suggestions to treat asthma. Whereas the results were promising, the findings were largely restricted to subjective outcome measures, with one important exception: Asthmatics treated with hypnosis as compared to a relaxation group showed a significant increase in forced expiratory volume over a 1-year span.

In New Zealand, Timothy C. Ewer and Donald E. Stewart conducted an important randomized, controlled hypnotherapy trial with 39 adults with mild-to-moderate asthma. The hypnosis group underwent six weekly 1/2-hour hypnosis sessions plus daily self-hypnosis. The hypnosis consisted of hypnotic induction, relaxation suggestions, ego strengthening, guided imagery, and direct suggestions for mastery over asthma symptoms. The control group intervention consisted of a nurse review of daily symptom diary entries. The hypnosis treatment group showed a significant reduction in airway obstruction as measured by peak flow, a significant reduction in airway irritability as measured by a methacholine challenge test, and a significant reduction in the use of bronchodilator medication. High hypnotizability strongly predicted a positive treatment response. Marianella Castes and coworkers reported a Venezuelan study in which 35 asthmatic children used guided imagery to change their inflammatory response. The researchers found a significant decrease in forced expiratory volume, a decrease in immunoglobulin E antibodies, and a decrease in the number of $CD23^+$ cells, which mediate the inflammatory response.

Unfortunately, neither of these studies has been replicated. Nonetheless, they suggest that hypnosis is possibly efficacious in the treatment of asthma with respect to increasing airway flow, decreasing airway hyper-responsiveness, and possibly curtailing the inflammatory response and that high hypnotizability predicts a positive response. With respect to peak flow, some critics have pointed out that these seemingly positive biologic changes are largely restricted to effects on the large air passages. Nevertheless, given the fact that the magnitude of airway change is clinically significant, especially in high hypnotizables, hypnosis should be seriously considered for this group. Wain demonstrated successful treatment of VCD, an asthma simulator that demonstrates a specific laryngoscopy profile. The symptoms have been shown to be controlled and/or ablated with hypnotic intervention.

HYPNOSIS AND IMMUNE-RELATED DISORDERS

Hypnotically assisted profound relaxation and specific hypnotic suggestions for cooling localized target tissue have been shown to reduce acute inflammation in, for example, burns and also to diminish chronic inflammation associated with autoimmune conditions like rheumatoid arthritis. Stephen Black and his associates found that direct hypnotic suggestions markedly reduced the expected weal and flare in response to a passive transfer skin test to a known allergen but did not alter cellular infiltration. Despite this clear laboratory demonstration of reduced allergic response, there have been only a few clinical studies applying hypnosis to the treatment of the gross symptoms of allergic sensitivity.

The main area in which hypnosis has been used in the treatment of immune-related disorders is psycho-oncology. Hypnosis can be used to facilitate the development of active coping strategies to deal with the crisis of the cancer diagnosis, death anxiety, life disruption following diagnosis, uncertainly of course of illness, helplessness, and noxious medical procedures.

Hypnotic intervention can be used effectively as an adjunct to controlling the nausea and vomiting effects of chemotherapy. Anticipatory vomiting, which is also a common occurrence just prior or even the week of treatment, can also be treated with hypnosis by having the patient learn self-hypnosis. Suggestions that Wain and others have used and that have been effective in this area include numbness, relaxation, floating and absorption in a task and/or even suggesting the first sign of the discomfort demonstrates that the chemo is working and having an impact on their body and by relaxing and visualizing the cancer cells decreasing they can help themselves.

Hypnosis can be used to ameliorate psychiatric symptoms associated with a cancer diagnosis, such as depression, anxiety, sleep disturbance, and pain. Montgomery et al., in a systematic review of hypnosis and cancer found that hypnosis for cancer focused on common symptoms like fatigue, sleep disturbances, pain, appetite loss, and dyspnea. Exploratory hypnosis can also be used for meaning making and benefit finding following the initial cancer diagnosis, as well as for enhancing quality of life in the face of the diagnosis.

A meta-analysis of psychosocial interventions in cancer showed that the largest treatment effects are found in altering quality of life, with more modest treatment effects associated with coping enhancement and emotional adjustment. The role of psychosocial interventions in prolonging survival rates is controversial, with studies yielding both positive and negative results. Most studies find that facilitating active coping and encouraging emotional expressivity are associated with improved immune regulation, but it is less clear whether improved immunoregulation affects the overall survival rate. James Hall found that high hypnotizability is a strong predictor of which individuals are more likely to alter their immune response following psychosocial interventions for cancer.

HYPNOSIS AND SEXUAL DYSFUNCTION

The classification of sexual dysfunctions typically corresponds to the three phases of sexual response—desire, arousal, and climax. Conjoint sex therapy is an empirically supported approach to the treatment of disorders of arousal and climax in both men (erectile dysfunction and pre- or postmature ejaculation) and women (frigidity, anorgasmia), in which couples are given a structured series of sexual assignments designed to impart new sexual skills and identify/work through conflicts that interfere with normal sexual response. Hypnosis can be used as an adjunct to conjoint sex therapy in couples. Daniel Araoz's book *The New Hypnosis in*

Sex Therapy contains a wide range of sex hypnotherapy techniques for supporting conjoint sex therapy.

The consensus among sex therapists is that disorders of desire in men and women (inhibited sexual desire; sexual compulsivity) are more complex in that long-standing unresolved traumatic or neurotic conflicts and maladaptive cognitive learning is important in their development. At the Masters and Johnson Sex Clinic, Mark Schwartz used a combination of traditional conjoint treatment along with dynamic exploratory hypnotherapy to identify the conflicts contributing to sexual desire disorders. Although no randomized clinical trials exist on hypnotherapy for sexual desire disorders, Schwartz's data constitute a large cohort study followed over time that offers promise for hypnotherapy in an area in which outcome success has generally been lacking.

HYPNOSIS AND DERMATOLOGY

One of the classic psychosomatic disorders is neurodermatitis. The link between emotions and skin lesions has been well-reported in the literature. In an early study by Surman et al., they demonstrated that more than half of their patients who had flat warts were able to eliminate them on one side with five sessions of hypnotic treatment. In particular, several clinicians have reported the successful treatment of warts with hypnotic interventions that were refractory to routine dermatologic treatment. One clinician reported the successful removal of penile condyloma acuminatum. Clinicians have also reported successful treatment of ichthyosis, pruritus, alopecia areata, eczema, boils, angular cheilosis, and psoriasis.

Areed Barabasz and his associates found that treatment of human papilloma virus (HPV)-related warts by hypnosis was more effective than medical tissue assault methods in long-term follow-up. In contrast to these clinical reports, some of the research has been less positive. Therefore, it may be appropriate to consider the use of hypnosis if other treatments fail. The more hypnotizable the patient, the more likely it is that physical alterations can occur. The use of hypnotic techniques as an adjunct to other treatments also appears warranted. It is safe to conclude that emotions have a strong effect on the skin and that hypnosis can have an effect on emotions.

HYPNOSIS AND ADDICTIVE BEHAVIORS

Hypnosis and Smoking Cessation

Hypnosis has been used for smoking cessation for many years. Many anecdotal studies have been published, as have several retrospective clinical studies that provide some encouragement for the use of hypnosis in smoking cessation. Unfortunately, many of these studies were limited because they did not include biologic markers of smoking cessation, nor was there randomization. Furthermore, the studies also varied in the amount of hypnotherapy provided during the hypnotic intervention. There are two *major* approaches to using hypnosis for treating smoking: (1) Multimodal cognitive-behavioral treatment conducted in trance. Brown and Fromm described an integrated hypnobehavioral protocol that addresses motivation for treatment; self-monitoring of urges and smoking behavior; nicotine fading; stimulus control; self-regulation of smoking behavior; smoking-related cognitive distortions; the effect of regulatory use of smoking; the social context of smoking; identification and prevention of relapse; and smoker self-image. Single-session or brief hypnosis, as exemplified in the work of Herbert and David Spiegel. The Spiegel approach uses the eye roll induction method, followed by a series of brief health-benefit suggestions designed to increase commitment for smoking cessation and the use of posthypnotic suggestions coupled with self-hypnosis training. During the course of the hypnotic intervention, positive imagery may also be used. (2) Others used a modified approach of the above where strategies are based on first developing the therapeutic alliance with the patient and structuring individual strategies based on their past experiences and what might motivate them to change. Those individual motivations are incorporated into the hypnotic suggestion given during the treatment session and in their self-hypnotic interventions.

At least one study using the longer hypnobehavioral protocol demonstrated that the intervention associated with an intensive hypnotherapy intervention program can be an effective means of achieving smoking cessation. In this one study, the rate of smoking cessation at 26-week follow-up was 40% and was confirmed by an expired carbon monoxide value of 8 ppm or less. The investigators noted that limitations in the study were the small sample size and a lack of measurement of hypnotizability.

Anecdotal reports of the single-session Spiegel approach included hypnotizability assessment using the HIP. These reports documented significant abstinence in 20% to 35% of subjects, with abstinence rates as high as 40% in subjects with higher hypnotizability scores on the HIP. The Spiegels reported that a consecutive series of 226 smokers referred for the smoking cessation program were treated with a single-session habit restructuring intervention involving self-hypnosis; they were followed for 2 years. The criterion for successful outcome was total abstinence for smoking; 52% of the study group achieved complete smoking abstinence 1 week after the intervention, and 23% maintained their abstinence for 2 years. The hypnotic interventions for smoking cessation also included self-hypnosis training in which the patient was provided with suggestions to minimize negative self-talk and to incorporate the benefits of nonsmoking and the ability to become a nonsmoker. Both approaches typically incorporate basic educational experiences regarding the deleterious physical effects of smoking and the significant benefits of overcoming the smoking addiction. The suggestions provided during the hypnosis generally include increased control in mastery and the positive experiences of healthy living. The suggestions should also be provided to counterbalance the symptoms of withdrawal.

Hypnosis in Weight Management

Hypnosis has been used in a number of weight control and weight management programs. It is generally combined with exercise and comprehensive dietary management. The purpose of a brief hypnotic intervention is to provide the patient with an opportunity to restructure their relationship to food. During the course of the hypnotic experience, patients are provided with a number of suggestions to facilitate greater control over their eating behavior. The hypnotic interventions are combined with a number of affirmations that facilitate the patient’s dietary management. Use of imagery may also assist in the patient’s weight management. Patients are provided with encouragement to use self-hypnosis on a daily basis to facilitate management of their eating behavior. Brown and Fromm described a manualized, multimodal protocol integrating hypnosis and cognitive-behavioral treatment that addresses motivation for treatment, stimulus control, self-regulation of eating behaviors, problematic eating patterns, managing urges, cognitive control over eating, relapse risk assessment and prevention, and dealing with dynamic factors associated with problematic eating. Irving Kirsch found that the addition of hypnosis to a cognitive-behavioral protocol for weight control made a significant contribution to the overall treatment effect.

Entwistle, Webb, and Abayomi have provided an excellent review of the literature concerning the use of hypnosis in weight management programs. Their review does provide an excellent understanding of the value of hypnosis in weight management programs.

Hypnosis and Eating Disorders

Beginning with Pierre Janet’s classic case studies more than 100 years ago through the modern hypnodynamic treatment approach of Moshe Torem, hypnotherapy has a long history in the treatment of anorexia and bulimia. Studies consistently have shown that patients with bulimia, and certain patients with anorexia, are more hypnotizable than normal control subjects. Johan Vander Linden in Belgium developed a detailed manualized approach to hypnotherapeutic treatment of patients with eating disorder with and without a history of trauma that combines the development of cognitive-behavioral skills in hypnosis and exploratory hypnotherapy with family therapy. The 13-step eating disorder protocol conducted over a 6-month period addresses providing psychosocial education, lowering medical risk, establishing therapeutic goals, increasing motivation, normalizing eating habits, evaluating progress, enhancing self-esteem, improving bodily experience, learning relaxation, enhancing autonomy, learning to recognize and tolerate negative and positive feelings, addressing trauma, and enhancing relapse prevention. Hypnosis is used in different ways across most of the 13 steps.

Clinical outcome data show that 55% of the patients no longer met sufficient criteria for an eating disorder diagnosis after 1 year, with 70% improvement in patients with bulimia, 52% improvement in patients with restricting anorexia, and only 18% improvement in patients with mixed anorexia and bulimia.

Hypnosis and Complex Habits

Hypnosis has also been used in the treatment of complex habits such as nail biting, tongue sucking, skin picking, and hair pulling or trichotillomania. The plucking of hair may occur for some individuals whose anxieties or internal tensions reach a crescendo, but the behavior becomes so overlearned that it serves to regulate a wide variety of affective and cognitive states in the patient. Trichotillomania is defined as the recurrent failure to resist the plucking of hair. Hair, eyebrows, and moustache are the targets. Patients are often unaware of engaging in the behavior. Hypnosis has been utilized in the treatment of complex habits in both individual and group interventions.

HYPNOSIS IN FORENSIC PSYCHIATRY

Hypnosis has been used as a forensic investigative tool to refresh the recollection of witnesses and victims, as well as sometimes suspects, of crimes. In the 1960s and 1970s, hypnotically refreshed memory was openly admitted as testimony in most courts. The forensic use of hypnosis greatly increased its popularity in those years, especially among law enforcement agents. Hundreds of police officers were given brief training in hypnosis. One popular program, for example, consisted of only 32 training hours. Given the general lack of psychological training of the police taking the courses and the cursory training given in forensic hypnotic interviewing, it was inevitable that the quality of investigative hypnosis would deteriorate. Martin Orne first articulated the misuses of forensic hypnosis and was greatly influential in cautioning the courts about abuses of forensic hypnosis and the questionable accuracy of hypnotically refreshed testimony. In the early 1980s, a series of landmark legal decisions changed the future of forensic hypnosis. Many states then adopted a per se *exclusion* standard, according to which hypnotically refreshed testimony was found inadmissible. Other states found the per se rule too extreme and adopted a *guidelines rule* in which hypnotically refreshed testimony was admitted, provided that the hypnotic interview was conducted by a competent mental health professional with specialty training in hypnotic interviewing, that no one else was in the room, and that a tape recording of each session preserved the record so that it could later be reviewed and scrutinized for possible undue influence.

In 1985, the AMA published a cautionary statement specifically on the forensic, not clinical, use of hypnosis. Unfortunately, most of the landmark court decisions limiting hypnotically refreshed testimony and also the 1985 AMA report were premature, in that the findings were written before the greater bulk of research findings on hypnosis and memory had appeared in the professional literature. The AMA report only cited a small number of the scientific studies on hypnosis in print at the time. Likewise, the only meta-analytic study on hypnosis and memory included 24 studies, only 4 of which pertained to memory recall. An expanded list of 38 studies was reviewed by Matthew Erdelyi in the same year. Erdelyi included 17 studies on the recall of meaningful material, only 3 of which were included in the aforementioned meta-analysis. Reviews and updates on scientific research on hypnosis and memory that are more representative and inclusive of the available research findings can be found in *Memory Guidelines* by the American Society of Clinical Hypnosis and *Memory, Trauma Treatment, and the Law* by Daniel Brown and coworkers. The main research findings fall into three broad areas: hypnotic hypermnesia, hypnotic pseudomemories, and confidence in hypnotically refreshed memory. *Hypnotic hypermnesia* pertains to the relative *completeness* of memory. Laboratory subjects are shown a series of target stimuli and are then asked to recall or recognize the original stimulus material under waking and hypnotic conditions. Hypnotic hypermnesia occurs if the subject remembers significantly more new information under hypnotic as compared to nonhypnotic conditions. Hypermnesia refers to the amount of information or the completeness of the recollection, not necessarily its accuracy.

Although the results across numerous studies are quite variable, the overall results suggest little or no advantage of hypnosis in recognition memory studies. Results across many studies show a modest increase in the amount of information recalled with hypnosis for more- but not for less-meaningful stimulus material. However, this conclusion is complicated in several respects. First, research designs do not always discriminate repeated recall and hypnotic recall; if the initial recall attempt occurs in the waking state and the second recall in hypnosis, there is no way of knowing whether the advantage of the second recall was due to hypnosis or simply to a second recall. Second, many studies show that hypnotized subjects asked to recall a target event produce a remarkable increase in, yet a widely variable amount, of responses.

Martin Orne hypothesized that hypnosis resulted in a lowering of response criteria, so that subjects used lax criteria when deciding whether to report a memory or other mental content like a thought or image as if it were a memory. However, subsequent research showed that when a shift to lax response criteria occurs, it is more a function of establishing unrealistic expectancies regarding memory retrieval and less a function of a hypnotic procedure per se. Although the *lax response-criteria* *hypothesis* served as the dominant explanation for hypnotic hypermnesia for several decades, a series of experiments in memory research outside of the hypnosis field by Henry L. Roediger and Matthew Erdelyi largely refuted this explanation.

The method used to control the variability in productivity in response to hypnotic hypermnesia suggestions was *forced recall,* according to which subjects were shown a fixed number of stimulus pictures, say 60, and after recalling as many as possible, say 38, were instructed to fill in the other 22 lines on the answer sheet. Unfortunately, although this method controlled for variable response criteria across subjects, it created an additional problem. The findings from many of the studies using the forced-recall procedure show an increase in both accurate and inaccurate information about the target event. This became known as the “hits and misses” view of hypnotic hypermnesia. However, this increase in the memory error rate is likely an artifact of the forced-recall method, which essentially encourages guessing. Research comparing both forced and nonforced recall failed to show any significant differences in the number of correct or incorrect responses in the hypnotic condition.

Overall, if any pattern emerges from these studies, it is that hypnosis may facilitate a modest recall of new information about a meaningful target event, at least in certain subjects, but that the effects of hypnosis per se are not always clearly distinguishable from relaxation and repeated recall effects. Although some critics assert that hypnotic hypermnesia runs the risk of inaccurate recall, this “hits and misses” hypothesis is misleading in that it does not distinguish between the effects of hypnosis per se and the effects of poor interviewing techniques and/or making unrealistic claims about memory retrieval and encouraged guessing.

The research on hypnotic *pseudomemories* tries to decipher the *accuracy or inaccuracy* of the memory obtained. There is an extensive scientific literature demonstrating that some normal (nonhypnotized) individuals are vulnerable to changing their memory reports for a target event when subjected to postevent suggested misinformation. This is known as the misinformation suggestion effect. Vulnerability of misinformation suggestions occurs both in the normal waking state and in the hypnotic context. In the hypnotic context, this is known as the hypnotic pseudomemory effect.

Kevin McConkey, Amanda Barnier, and Peter Sheehan defined hypnotic pseudomemory as “memory reported when a suggestion is given following a hypnotic induction that a person will remember specific information that the hypnotist knows (or believes) to be false." Extensive research on the hypnotic pseudomemory effect has compared the response to misinformation suggestions in both high- and low-hypnotizable subjects, in the waking state, and following a hypnotic induction. This extensive research generally shows that pseudomemories were reported both by high and low hypnotizables but at higher rates in high hypnotizables, and also that pseudomemories were reported across hypnotic, waking, and simulation-of-hypnosis conditions. There is a trend that some portion of but not all high hypnotizables who were also hypnotized have at best modestly higher rates of pseudomemories than high hypnotizables in the waking condition or than low hypnotizables in either condition. Cox and Barnier list three types of memory distortion or confabulation produced by hypnosis: (1) That produced by suggested amnesia for everything associated with a target event; (2) That produced by instructions explicitly to fill in the gaps of memory; and (3) That associated with the temporal order of the events in the target experience (memory disorganization).

High hypnotizability, not the hypnotic procedure, per se, is associated with a modest increase in pseudomemory rates. However, Graham F. Wagstaff and his associates found that when high hypnotizables as compared to low hypnotizables are given instructions that imply that hypnosis leads to more, not less accurate reporting that are less likely to produce false memories. However, the consistent finding from hypnotic pseudomemory research is that a hypnotic procedure per se poses no more risk of pseudomemory production than a walking interview procedure. The primary risk factor for inaccuracy due to suggested misinformation is an interview that includes suggested misinformation, irrespective of whether this misinformation is given in the waking state or in hypnosis. The problem of pseudomemory inaccuracy lies in suggestive interviewing and setting unrealistic expectations, not hypnosis. Some critics have argued that the effects of hypnosis and misinformation suggestions are additive and that both contribute to the overall variance to the hypnotic pseudomemory effect. However, an attempt to replicate the findings with a more sophisticated signal detection design failed to replicate the earlier findings that hypnosis had a negative effect on memory accuracy, but did replicate the rather robust finding that suggestive interviewing independent of hypnosis significantly increased memory inaccuracy. If anything has been learned from the numerous pseudomemory studies, it is that hypnosis as a procedure per se does not cause pseudomemories, and that interviews conducted with and without hypnosis must scrupulously avoid supplying misinformation, that is, avoid supplying *any* content that has not been already reported by the interview subject.

Martin Orne and Bernard Diamond cautioned against the use of hypnosis because they believed that hypnotically refreshed memory may foster undue confidence about inaccurate memories and that such undue confidence would make such individuals unusually resistant to cross-examination in the courtroom. However, the many research studies pertaining to hypothesized increased target memory confidence following a hypnotic induction have yielded equivocal results: High hypnotizability and/or a hypnotic induction were found significantly to increase confidence in hypnotically retrieved memories in some studies and yet failed to do so in other studies. The pattern of results implies that undue confidence is not a consequence of a hypnotic procedure per se, but rather is a function of directly communicating unrealistic expectancies about hypnotic memory retrieval ("You will be able to remember absolutely everything," "Your memory is like a tape recorder and you can review all the details on a television screen in hypnosis"). Other factors contributing to inflated confidence may include high hypnotizability, less meaningful or ambiguous stimulus material, shorter retention intervals, and permission to guess. There is no compelling evidence that hypnotic procedures per se are necessarily associated with inflated confidence, and in most of the studies demonstrating inflated confidence the magnitude of the effect is not large.

Some clinicians state that undue confidence in hypnotic pseudomemories (memory hardening) would make a witness impervious to cross-examination in the courts. Laboratory studies designed to test Orne's hypothesis have consistently failed to support the memory hardening view. These findings provide no support for the notion that the influence of hypnotic procedures on memory are so powerful that witnesses with hypnotically refreshed testimony cannot be effectively cross-examined or that hypnotically refreshed testimony should be found inadmissible because it is particularly resistant to cross-examination.

The overall findings from extensive research on hypnosis and memory indicate that whereas hypnosis may contribute to a modest increase in new information about a target event, at least in certain subjects, there is little evidence that hypnosis per se contributes to an increase in the memory error rate, and that the greater portion of the variance of inaccuracies associated with hypnotically refreshing memory is best attributed to nonhypnotic factors—mainly to poor interviewing techniques that include fostering unrealistic expectancies about memory retrieval, encouraged guessing, and suggesting misinformation.

As an outgrowth of the enormous research literature on memory retrieval, the misinformation effect, and the effects of interview components on the completeness and accuracy of recall, there is a better understanding of how to conduct nonhypnotic and hypnotic investigative interviews that minimize memory error rate while maximizing the retrieval of new information about a target event. These findings for forensic hypnotic interviewing can be summarized in the following guidelines:

1. *Proper informed consent.* Any forensic hypnotic interview begins with informed consent about the nature of memory, memory retrieval, and hypnosis, the risks and benefits of the procedure, and possible alternatives. In the mental health field, informed consent is not seen as a single event, namely providing information at the onset of treatment, followed by signing a consent form; it is best seen as an ongoing process of open communication and adjustment of the unfolding conceptualization of treatment problems, methods, and goals, as part of an unfolding relationship. In some jurisdictions, clients need to be informed that the use of hypnosis runs the risk that any testimony associated with hypnosis, even in a clinical and not a forensic context, may be inadmissible in court. The guidelines cited earlier from the American Society of Clinical Hypnosis include a sample informed consent form.
2. Certain *personality characteristics* are associated with memory distortion. The Gudjonsson Suggestibility Scale (GSS) is a standardized assessment instrument that assesses the degree to which a given individual is vulnerable to accepting misinformation suggestions, confabulating memory reports, and shifting memory reports in response to the emotional tone of an interview. Administering the GSS at the onset of a forensic examination allows a normative assessment of the risk of memory distortion prior to any forensic interview. The Inventory of Childhood Memories and Imaginings (ICMI) evaluates fantasy proneness. About 3% to 5% of

the general population is very high in fantasy proneness, which predicts significant internal confabulation of fantasy and memory. Empirically derived instruments like the Structured Inventory of Malingered Symptomatology (SIMS), the Malingering Probability Scale (MPS), and the Structured Interview of Reported Symptoms (SIRS) also help to identify the type of individual in a forensic setting that is likely to malinger psychiatric symptoms, and the Test of Memory Malingering (TOMM) and the SIMS help to identify individuals who specifically malinger memory symptoms. These instruments, albeit not foolproof, allow for a reasonable assessment of a variety of situational and personality factors that predict high distortion of memory reports in a forensic context.

3. *Establishing rapport,* making the patient feel at ease with the interview context, and building a relationship are important components of successful interviewing. The interview can begin with statements like, "I want to get to know you; tell me about yourself, something that is important to you." This section of the interview also includes the rules of interview, such as the importance of telling the truth and being accurate; the permissibility of saying "I don't know" or "I don't want to say"; the fact that if the interviewer repeats a question, it does not mean that the previous answer was incorrect; and the fact that the interviewer really does not know what happened and must rely on whatever the interviewee is able to say about the target event.
4. The preamble should set forth the *expectancies externally supplied by the hypnotist* regarding hypnosis, memory, and memory retrieval. The extent to which explicitly supplied explanations about hypnosis, memory, or memory retrieval overwrites the interviewee's *a priori* expectancy is unclear. Nonetheless, the forensic examiner should make explicit both the interviewee's a *priori* and explicitly suggested response expectancies.
5. The preamble should include the *establishment of realistic expectancies* about memory retrieval ("You may or may not recall more about") and the limits of accuracy ("Some of the things you may recall may be more and some less accurate") and should include explicit instructions not to guess ("Report everything that comes to mind but do not guess") and explicit instructions to foster an increase in, not a suspension, of critical judgment ("Throughout the procedure you will retain your best judgment as to which things you report are likely to be more and which are likely to be less accurate, knowing that not all details you report will be accurate").

6. Next, the interview focuses on *establishing a waking baseline memory* for a meaningful, neutral event from the same time frame (or developmental age) as the target event. In this section of the interview, free recall is encouraged, “Tell me everything you can about.”
7. The interview directly addresses the *target event* without explicit cuing. The interviewer must avoid leading the interviewee or being suggestive. Open-ended questions are desirable, such as “Tell me your understanding about what you’re being interviewed about today.” For interviewees who fail to disclose the target event to an open-ended prompt, the interviewer can state what they have been told, such as “Officer X told me that you might have witnessed an accident on (date) at (location).” Then, free recall is encouraged, with questions such as “Tell me whatever you can recall about (name target event) from the very beginning to the end.”
8. Field research on investigative interviews by child protective service workers or law enforcement agents with children suspected of being abused consistently has shown that the failure of many children to disclose abuse in an initial interview is an artifact of the interview situation. A similar finding comes from field research on adult suspects during police interrogations. In such situations, forensic interviewers typically make the mistake of asking too many closed-ended or option-limiting questions at the onset of the interview.

A series of seemingly innocuous, factually based questions like “What is your name?” “Where do you live?” “What grade are you in?” establishes expectancies that discourage free recall and shapes the interviewee’s behavior in the direction of only answering questions posed by the interviewer. Under such circumstances, the volume of free recall is significantly reduced, including much potentially valuable information about the target event. Such field research strongly shows that the greater amount of information and the greater accuracy of that information about the target event occurs in the context of *extended free recall.* There should be at least two periods of extended free recall about the target event—one prior to and one following the hypnotic induction—followed by additional periods of extended free recall in hypnosis in subsequent sessions. Free recall is encouraged by such questions as, “Tell me everything you can remember about.”

Hypnotic forensic interviewing is likely to be less controversial when the hypnotic induction conveys a *heightened focus or redistribution of* *attention* but avoids specific suggestions about trance characteristics, for example, involuntariness, the suspension of critical judgment, or dissociative compartmentalization that might bias retrieval. When strong emotional arousal, like fear, has interfered with retrieval of a memory for a target event, some evidence shows that the inclusion of *hypnotic relaxation* and instructions for focused recall increases the recall of meaningful information. After a hypnotic induction, it is best to allow the interviewee an extended duration of free recall of the target event after hypnotic induction. Frequent closed-ended questions should be avoided.

The Cognitive Interview was developed by R. Edward Geiselman and Ronald P. Fisher as an alternative procedure to hypnosis for refreshing memory. It is based on basic principles of memory retrieval, with extended free recall and context reinstatement serving as the main components of the interview. More than 50 laboratory studies exist on memory retrieval with the Cognitive Interview. These studies consistently show that the interview yields around 25% to 35% of new information about a target event, typically without a significant increase in the memory error rate. Apart from the Cognitive Interview, per se, these data strongly support the value of *context reinstatement along with free recall* as a primary memory retrieval technique. In hypnotic forensic interviews, there should be an extended period of free recall, followed by context reinstatement plus free recall. The interviewee is reoriented to the time and place of the target event, asked to place themself in the scene, and then to freely recall the event.

Throughout the hypnotic interview, the interviewer should pay careful attention to the *type of interview questions* used. Instructions that encourage extended free recall are best. After an extended period of free recall, “wh-” questions are acceptable, like “what, where, when?” Research has shown that option posing and yes/no questions are of more limited value. The best guideline is to keep interview questions as open-ended as possible and select questions that pose the greatest number of options as answers.

Introduction of follow-up *probe questions* occurs only after an extended period of free recall and context reinstatement in hypnosis fails to yield any more new information. Probe questions must avoid supplying new information, but instead must simply reflect back to the interviewee their words (“Earlier you said the assailant was ‘tall’; tell me more about ‘tall’”). The entire session should be *videotaped*—not just the hypnotic session, but all of the preliminary explanations and the preamble, from the point that the interviewee and interviewer meet. The most important principle of any forensic interviewing is completely to *avoid leading and misleading questions.* The interviewer should not suggest or alter *any* content that has not already been reported by the interviewee.

There are no hard and fast rules as to when to use a hypnotic as compared to a nonhypnotic forensic interview. Reasonably prudent forensic examiners familiarize themselves with the case law on hypnotic admissibility for the jurisdiction is question. Strong case law in a state following a per se exclusion rule would merit caution. Cases entailing the highest standards of evidence, like reasonable doubt in a murder case, also merit caution in choosing a hypnotic forensic interview over the nonhypnotic Cognitive Interview.

Most experts further concur that conducting forensic hypnotic interviews with suspects of crimes fraught with difficulties should be avoided except under special circumstances. Hypnosis is indicated in cases involving full dissociative amnesia that have not been resolved by other methods.

The most frequent, justifiable use of forensic hypnosis is in interviewing witnesses and victims of crimes, not to “solve” the crime, but to provide additional evidence William Wester and D. Corydon Hammond have reported that the most frequent use of forensic hypnosis is interviewing witnesses and victims of crimes not to solve the crime but to provide additional evidence. All evidence derived from hypnotically refreshed testimony, as well as evidence from other, nonhypnotic procedures to refresh memory, like the Cognitive Interview, should not stand alone without independent corroborative evidence. The most common use of forensic hypnosis is in seeking new evidence about a crime from witnesses and/or victims.

The most famous case in which forensic hypnosis provided valuable evidence was the Chowchilla school bus kidnapping in the 1970s. Three assailants kidnapped a school bus of mixed-age children, buried the bus and the children alive in a rock quarry, and demanded a large ransom. The bus driver and some of the older children dug their way out by hand. Later the bus driver was hypnotized and asked about the license plate and other identifying details. Under hypnosis, the bus driver recalled a number of but not all of letters in the license plate, and recall sequence was inaccurate. Nevertheless, the recall of some of the letters accurately along with recall of other unusual features of the assailants’ vehicle led to a correct computer identification of the vehicle. Witnesses at the scene of sniper incidents and crime scenes involving suspected serial killers or rapists are interviewed by a team whose practice includes hypnotizing witnesses for the purpose of generating a composite drawing of the assailant to be circulated in the news media. This team's protocol has been effective in successfully apprehending a number of high-profile assailants. Most federal jurisdictions do not follow the per se exclusion rule and follow the guidelines approach. Federal guidelines have been developed for the forensic use of hypnosis.

FUTURE DIRECTIONS

The art and science and the domain of hypnosis have evolved into a biopsychosocial model of diagnosing and treating patients with psychiatric and/or medical disorders. As we continue to explore mind–body interaction one can anticipate increased emphasis on evidence-based controlled studies. Recently, Afik Faerman et al., in a very promising study in 2024, using transmagnetic stimulation demonstrated that levels of hypnotic ability were temporarily enhanced even though hypnotic capacities were considered stable through adulthood, by short stimulation of the of the left dorsal lateral prefrontal cortex. Capacities were measured by a hypnotic screening tool, the Hypnotic Induction Profile. Patients with a diagnosis of fibromyalgia were split into two groups. Those receiving the above stimulation showed a decrease in pain when compared to the sham group who did not receive the same stimulation. This study adds greater credence in reinforcing the view that hypnosis integrates the impact of suggestion on one's neurophysiology. In conclusion, hypnosis is not just a technique; it requires listening, observing, and understanding the patient and this approach further reinforces Osler's quote, "The good physician treats the disease, but the great physician treats the patient who has the disease."

ACKNOWLEDGMENT

This chapter represents an update and revision of previous chapters in this series, most recently by Dr. Daniel P. Brown from the last edition. We are indebted to Dr. Brown for their work and retain much of their prose in this current edition.

FURTHER READINGS

Axelrad AD. The role of hypnosis in multiple convergent management of chronic pain. *Houston Med*. 1990;6:111.

Barabasz AF, Barabasz M. Single-session manualized ego state therapy (EST) for combat stress injury, PTSD, and ASD, Part 1: The theory. *Int J Clin Exp Hypn*. 2011;60:381.

*Brown D. How to treat attachment disturbances in adults: the tree pillars of comprehensive attachment repair. Presented at “Attachment and Trauma—Relationships, Consciousness, and the Developing Self,” *International Attachment Conference, London, England*, November 14–17, 2019.

*Cardena E., Terhune DB. The roles of response expectancies, baseline experiences, and hypnotizability in spontaneous hypnotic experience. *Int J Clin Exp Hypn* 2019; 67(1):1–27.

Carolusson S. Dynamic hypnosis, IBS, and the value of individualizing treatment: a clinical prospective. *Int J Clin Exp Hypn*. 2014;62(2):145.

*Casiglia E, Finatti F, Tikhonoff V, et al. Mechanisms of hypnotic analgesia explained by functional magnetic resonance (fMRI). *Int J Clin Exp Hypn*. 2020;68(1):1.

Castes M, Hagel I, Paleque M, Canelones P, Corao A. Immunological changes associated with clinical improvement of asthmatic children subjected to psychological intervention. *Behav Immunol*. 1999;13:1.

*Cox RE, Barnier AJ. A hypnotic analogue of clinical confabulation. *Int J Clin Exp Hypn.* 2015;63(3);249.

Daitch C. Hypnotherapeutic treatment of anxiety-related relational discord: a short-term hypnotherapeutic protocol. *Am J Clin Hypn*. 2014;56:325.

Dienes Z, Hutton S. Understanding hypnosis metacognitively: rTMS applied to the left DLPFC increases hypnotic suggestibility. *Cortex*. 2013;49:386.

Dillworth T, Jensen MP. The role of suggestions in hypnosis for chronic pain: a review of the literature. *Open Pain J*. 2010;3(1):39.

Entwistle PA, Webb RJ, Abayomi JC. Unconscious agendas in the etiology of refractory obesity and the role of hypnosis in their identification and resolution: a new paradigm for weight-management programs or a paradigm revisited. *Int J Clin Exp Hypn*. 2014;62(3):330.

Faerman A, Bishop JH, Stimpson KH. Stanford Hypnosis Integrated with Functional Connectivity-targeted Transcranial Stimulation (SHIFT): a preregistered randomized controlled trial. *Nature Mental Health*. 2024;2:96–103.

*Halsband U, Wolf TG. Functional changes in brain activity after hypnosis: Neurobiological mechanisms and application to patients with a specific phobia—limitations and future directions. *Int J Clin Exp Hpyn.* 2019;67(4):449.

*Jensen MP, Jamieson GA, Lutz, MG, et al. New directions in hypnosis research: strategies for advancing the cognitive and clinical neuroscience of hypnosis. *Neurosci Conscious*. 2017;2017:1–14.

*Jiang H, White MP, Greicius MD, Waelde LC, Spiegel D. Brain activity and functional connectivity associated with hypnosis. *Cereb Cortex*. 2017;27(8):4083–4093.

*Kendrick C, Sliwinsi J, Yo Y, et al. Hypnosis for acute procedural pain: a critical review. *Int J Clin Exp Hypn.* 2016;64(1):75.

*McGeown WJ, Mazzoni G, Venneri A, Kirsch I. Hypnotic induction decrease anterior default mode activity. *Conscious Cgn.* 2009;18:848.

Miller V, Whorwell PJ. Hypnotherapy for functional gastrointestinal disorders: a review. *Int J Clin Exp Hypn.* 2009;57(3):279.

*Montgomery GH, Sucala M, Baukm T, Schnur JB. Hypnosis for symptom control in cancer patients at the end-of-life: a systematic review. *Int J Clin Exp Hypn*. 2017;65(3):296–307.

Peter B, Vogel SE, Prade T, et al. Hypnotizability, personality style, and attachment: an exploratory study, Part 1—General results. *Am J Clin Hypn*. 2014;57:13.

*Rotaru TS, Rusu A. A meta-analysis for the efficacy of hypnotherapy in alleviating PTSD symptoms. *Int J Clin Exp Hypn.* 2016;64(1):116.

Shih M, Yang YH, Koo M. A meta-analysis of hypnosis in the treatment of depressive symptoms: a brief communication. *Int J Clin Exp Hypn*. 2009;57(4):431.

Spiegel D, Maruffi B, Frischholz EJ, Spiegel, H. Hypnotic responsivity and the treatment of flying phobia. *Am J Clin Hypn*. 2014;57:156.

*Spiegel EB. Attachment-focused hypnosis in psychotherapy for complex trauma: Attunement, representation, and mentalization. *Int J Clin Exp Hypn*. 2016;64(1):45.

Spiegel SB. Current issues in the treatment of specific phobia: recommendations for innovative applications of hypnosis. *Am J Clin Hypn*. 2014;56:389.

*Valentine KE, Millin LS, Clark LJ, Moriarty CL. The efficacy of hypnosis as a treatment for anxiety: A meta-analysis. *Int J Clin Exp Hypn*. 2019;67(3):336.

Wagstaff GF, Wheatcroft JM, Caddick AM, Kirb LJ, Lamont E. Enhancing witness memory with techniques derived from hypnotic investigative interviewing: focused meditation, eye-closure, and context reinstatement. *Int J Clin Exp Hypn*. 2011;59:146.

Wain HJ. Reflections on hypnotizability and its impact on successful surgical hypnosis: a sole anesthetic for septoplasty. *Int J Clin Exp Hypn*. 2004;46(4):313.

Wain HJ, Amen DG. Emergency room use of hypnosis. *Hosp Psychiatry*. 1986;8:19.

Wain H, Perito R. Care of combat injured service members. In: Cozza SJ, Goldenberg MN, Ursano RJ, eds. *Care of Military Service Members, Veterans, and Their Families*. APPI Press; 2014.

Wester WC, Hammond DC. Solving crimes with hypnosis. *Int J Clin Exp Hypn*. 2011;53(4):255.

32.20 Family and Couples Therapy

HENRY I. SPITZ, M.D., D.F.A.P.A., D.F.A.G.P.A., AND SUSAN SPITZ, A.C.S.W., C.G.P.

The family is the foundation upon which most societies are built. The study of families in different cultures has been a subject of fascination and scientific interest from viewpoints as diverse as sociology, group dynamics, anthropology, ethnicity, race, evolutionary biology, and, of course, the mental health field. The confluence of information gleaned from family studies has set the backdrop against which the contemporary practice of family therapy has evolved.

Historically, the family was not viewed in the most favorable light by early observers. Beginning with Freud, the family was seen as the single most important factor in the genesis and maintenance of psychopathology. Psychoanalysis aimed to repair the damage created by early experiences in the patient's family. Unresolved Oedipal conflicts, maternal over-protectiveness, and rivalrous and excessively competitive sibling relationships typify the focal points that illustrate early psychoanalytic perspectives on the family. In point of fact, while psychoanalytic theory applied to family life afforded a rich descriptive language for the understanding of intrafamilial dynamics, much of which is still in use today, it fell short of intervening in the solution to family problems by actually involving more than one person in the session.

Until the 1950s, psychoanalysis was the prevailing theoretical standard and the primary form of psychotherapeutic treatment available to patients with psychological problems. With the emergence of group psychotherapy and the courageous attempts to break with tradition by forward-thinking, individually oriented practitioners who began to see whole families in the therapeutic setting, a sea change in the orientation towards families and family therapy had begun. The central premise of the family as a resource rather than a villain was the critical element in liberating psychotherapists to explore the heretofore-untapped resources for change that were an inherent part of every family.

Modern family and couple therapies see the family and the couple as both part of the problem but also as an essential resource available for the resolution of the presenting problem. Family strengths, as opposed to weaknesses, are reinforced, educational elements offer families new modes of adaptation in place of rigid, well-practiced but dysfunctional interactional patterns, and the therapist is the ally of and advocate for the family and each of its members. The specific ways in which this shift in orientation has come about can be readily seen when one looks at the developments in the family therapy field in its historical context.

HISTORICAL TRENDS

Both family and group therapists often point to the work of Alfred Adler at the Vienna Child Guidance Clinic. Adler's involvement of parents in the therapy for their children, a practice that seems axiomatic today, represents a significant departure from the dyadic psychoanalytic model prevalent at the time. It is not until the 1940s that developments relevant to the evolution of contemporary family therapy practice can be seen. During this decade, the biologist Ludwig von Bertalanffy, applied concepts from mathematics and physics in an attempt to create a unifying theory for human interaction. His work emphasized the systemic nature of the human organism and the capacity for change in open human systems. He saw a system as more than the sum of its parts and capable of interacting with other systems.

Bertalanffy used the concept of equifinality to denote the capability of a system to find many routes to the same solution of a problem. A strong opponent of the dehumanized mechanistic theory prevalent at the time, he preferred a view that regarded human systems as ecologic organisms with a well-defined value system.

Midway through the 1940s, simultaneous seeds of family therapy were being sewn. First-generation family therapists, including Murray Bowen working at the Menninger Clinic, and Carl Whitaker, at Emory University, became interested in working with schizophrenic families. John Bowlby's work about infant–mother bonding and the perils of insecure or inadequate attachments in early life and its subsequent impact on critical adult relational issues such as trust and forming an enduring interpersonal bond caught the attention of interested family therapists.

Nathan Ackerman is often credited with being the founding father of family therapy. Trained as a child psychiatrist and psychoanalyst, Ackerman retained his emphasis on understanding individual personalities but also had a broader view of how individuals articulated with one another in ways that clashed or meshed. The logical extension of this point of view was to conduct sessions with the whole family and to intervene in often provocative and confrontational ways that would mobilize covert family conflict such that it was available for examination and change in the session.

In the late 1940s and early 1950s, Gregory Bateson began working in Palo Alto, the city whose name was to become synonymous with other influential family therapists, including Don Jackson, Jay Haley, and Virginia Satir (the Palo Alto group), who conducted their work at the Mental Research Institute. Bateson met Norbert Weiner, a mathematician, and expert in the field of cybernetics, and became intrigued with the application of cybernetics to human systems. Cybernetics enabled the discovery of key constructs in family therapy that included the premise that change was not linear as had been previously assumed, but rather it was circular, determined by both positive and negative feedback loops within the family interaction. The "double bind" is perhaps the best-known example of a negative feedback loop in family communication. The famous mother-to-son comment, "I hate you, you're such an angel," is a case in point.

Later on, Jackson, Haley, and Satir turned their attention to the process of communication in severely disturbed patients and their families. They were among the first to understand psychotic communications as having meaning and making a certain kind of sense in schizophrenic families. They began to label communication patterns and coined terms that are widely used today. Jackson studied couples and identified complementary and symmetrical relationships and the decoding of redundant relationship patterns or family and couple “rules.” Satir contributed clinical techniques that were aimed at reducing distance among family members. Her experiential approach and therapeutic use of self-disclosure and appropriate physical interaction with family members was later to be seen in the work of many who followed her.

Jay Haley was influenced by the work of Milton Erickson, an exponent of the use of indirect hypnotic techniques, metaphor, and suggestion. Haley’s therapeutic stance was an active and pragmatic one directed toward the rapid resolution of the presenting problem in the family.

Another central influence in the formation of family therapy came from the work of Murray Bowen. Among his many talents, Bowen was a strong theoretician as well as a clinician. His focus began when working with families with a schizophrenic member, during the course of which he made observations about how family members cooperated to maintain an interpersonal space and a degree of individual differentiation from the undifferentiated family ego mass. Bowen’s awareness of triangulation, a process whereby couples and families involve a third person (extramarital affair; mother–father–child triad) or third “thing” (work, alcohol, and drugs), all of which have the effect of attenuating the fears related to closeness and intimacy in primary relationships.

Three other major figures in the development of family therapy were Carl Whitaker, Lyman Wynne, and Theodore Lidz. Whitaker is perhaps best known for his emphasis on the family therapist’s need to interrupt and repattern disruptive, rigid, well-practiced family habits. His open attitude toward his own work being observed by others was the embryonic beginning of the use of the one-way mirror, the consulting team approach, and the use of videotaped sessions as regular fixtures in couple and family therapies. Though often flamboyant in style, Whitaker, as is the case with many noted family therapists who came on the scene later, used themselves as an agent of change employing unpredictable, spontaneous, and often confusing techniques designed to upset the homeostasis of an ineffective family system.

Lyman Wynne was interested in repetitive, stereotypical roles taken by family members and theorized that disordered or communication deviance was the conduit through which transgenerational transmission of irrationality went from one generation to the next. He coined the term “pseudomutuality” to denote the role of a false front presented by families who fear both intimacy and separateness. Although the bulk of Wynne’s work was with schizophrenic families, subsequent research and clinical studies have shown that the patterns he identified go beyond the schizophrenic spectrum and are readily observable in personality, mood, and anxiety disorders.

Theodore Lidz, also working with schizophrenic families, implicated the critical role played by fathers in these families. Prior to Lidz’s work, the main influence on the intrafamilial environment of these families was felt to be the mother. Lidz and his colleagues defined specific destructive patterns of paternal interaction with spouses and children that led to serious adverse psychological consequences. When Lidz worked with dysfunctional couples, he also delineated maladaptive patterns that led to marital discord or disruption. Marital schism, a state wherein partners compete rather than collaborate and fail to integrate their differences in a functional manner, and marital skew, where a conspicuous power imbalance exists, and active antagonism predominates, are the most widely recognized examples of Wynne’s terminology for dyadic relationships.

These family and couple therapy pioneers set the stage for the decades to follow that bore witness to the virtual explosion of family therapy research and practice and the emergence of defined schools of family therapy. The family and couples therapy literature expanded exponentially following the groundwork they laid. Corresponding acceptance of family therapy as a viable clinical entity now supported by empirical research studies gave greater momentum to family and couples work and the creation of several mainstream models of family therapy practice.

DEFINITION

Family therapy is a mosaic of techniques, all of which have the shared goal of direct alteration of maladaptive family processes. In point of fact, it is more accurate to speak of family, and couple therapy in the plural as family and couple therapies since the array of interventions with the above goal in mind is impressive. The coalescence of discreet therapeutic orientations has resulted in the creation of several distinct tracks in the actual clinical practice of family therapy. These orientations toward the treatment of family problems will be the subject of the next section.

SCHOOLS OF FAMILY THERAPY

Schools of thought in family therapy are like windows in a greenhouse, each with its own orientation when observing the family through a different pane. Each has its own lexicon, goals, and focal points but all share a common purpose in attempting to find innovative ways to provide time and cost-effective therapies for families in distress.

A representative sampling of the current state of popular family therapy techniques shows the clear linkage between first-generation family therapy described above and how it manifests itself in modern clinical practice. Gurman, in the chart that follows, has simplified the apparent overlap in the different schools by outlining the major schools of family therapy, the influential practitioners of each method, and the focus of each technique (Table 32.20–1). This schema allows one to compare and contrast each family therapy model in more depth.

Cognitive/Behavioral

The *cognitive/behavioral family approaches* have their theoretical base in social learning theory, operant conditioning, and cognitive psychology. Cognitive and behavioral therapies require active therapist participation both in structuring the framework of therapy, finding circumscribed as opposed to global goals, and having a planned sequence of interventions and instruction in mind in order to reach the solution of the presenting problem. For the most part, cognitive/behavioral family therapy is brief, symptom focused that places low value on revisiting historical material in sessions. Like most family and couples interventions, with the possible exception of the psychodynamic approach, it is extremely pragmatic, focused in the here-and-now, very structured, interactive, and directive.

The overriding premise is that problems are learned behaviors that are well-practiced and consequently almost reflexive in nature. This model helps expand the couple or family’s repertoire of functional communication and actions and, in so doing, incorporates new ways of thinking about familiar but dysfunctional family interactions. For example, the wife who feels that her husband is working too hard delivers that message from a position of caring and concern. The husband in this case “hears” the message as critical of him and feels that his partner is accusing him of being a workaholic and a neglectful father. The cognitively oriented therapist would identify this irrational belief or cognitive distortion, clarify it by supplying an alternative, more positive reading of the message, encourage the use of reality testing to keep catastrophic, or all-or-nothing thought patterns in check and reinforce all new, more adaptive and rewarding responses on the husband’s part.

An essential ingredient in adapting individually based cognitive/behavioral techniques to the couple and family milieu is the added emphasis on social exchange theory. This adds a dimension of mobilizing the reward system in the system. Families and couples are taught to replicate those behaviors that increase the number of rewarding interchanges: decrease negatively cast feedback and acquiring collaborative problem solving skills.

The therapeutic tool kit for cognitive/behavioral therapy with families and couples contains many of the same techniques used in individually based cognitive therapy. Positive reframing, cognitive restructuring, insuring a shared base of assumptions and expectations between partners and among all family members are all parts of the therapeutic equation.

In order to induce behavioral change, in-session exercises and educational elements are coupled with homework assignments to be completed between sessions. Cognitive/behaviorists do see whole families but feel more comfortable than practitioners from some other schools in seeing those subsystems in the family that are most relevant to the resolution of the identified problem. Therapists of this orientation are also comfortable with the concept of repetitive brief modules of treatment designed to address a specific issue each time over the course of time.

Jacobson and others have validated the benefit of the cognitive/behavioral approach. Advocates of this model have created impressive arrays of instruments for measuring parameters of change. Generally speaking, the cognitive and behavioral therapies are the most specific about technique (what they actually do in family sessions) and about the ways in which they assess the efficacy of their work.

Novel applications of these approaches have been used for parenting skills training, childhood conduct disorders, developmental disabilities, premarital counseling, and bipolar disorder. The integration of cognitive/behavioral family work with other therapies has been helpful with eating disorders, substance abuse, and anxiety disorders.

Table 32.20–1. Schools of Family and Couple Therapy

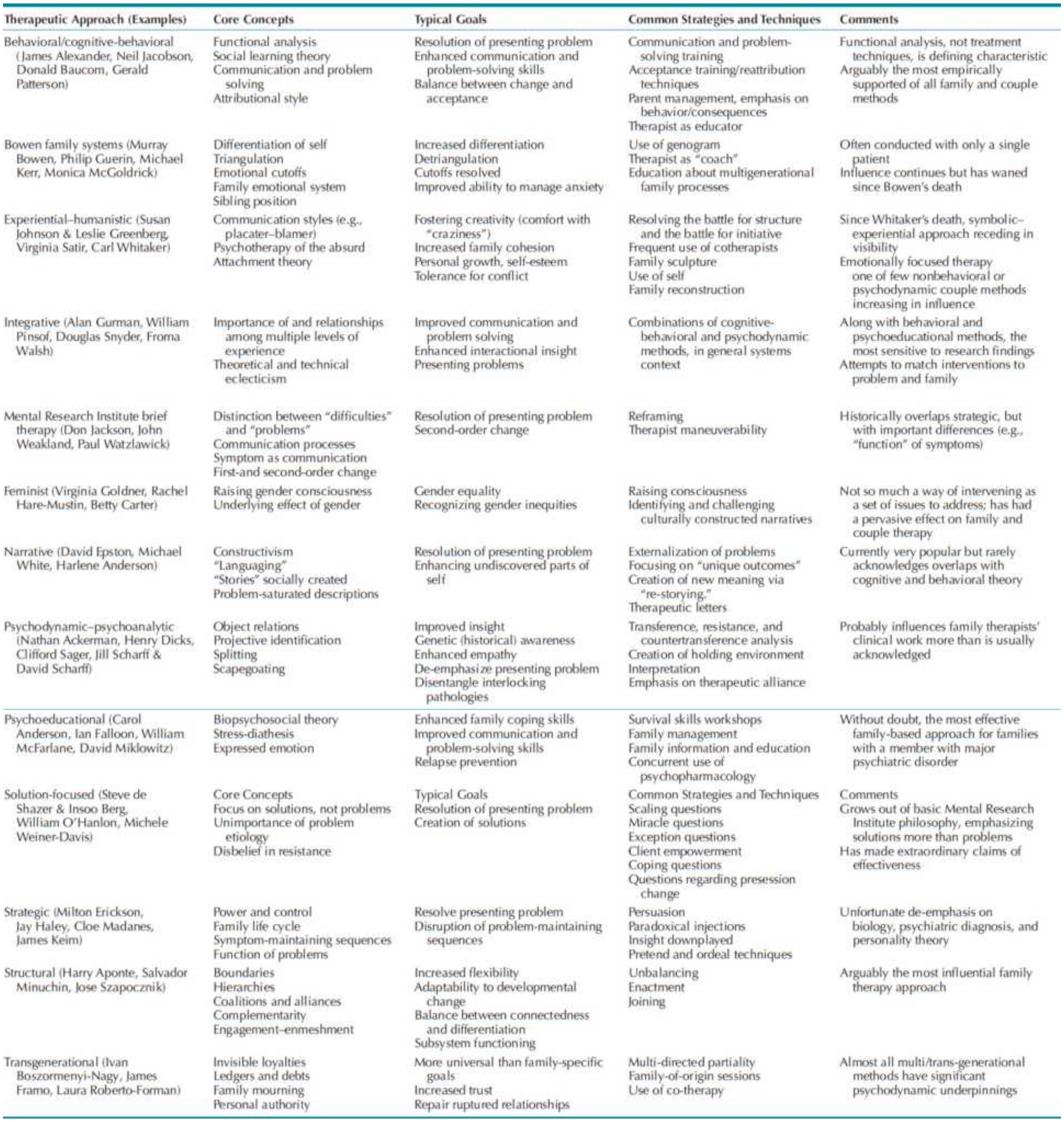

Therapeutic Approach (Examples)	Core Concepts	Typical Goals	Common Strategies and Techniques	Comments
Behavioral/cognitive-behavioral (James Alexander, Neil Jacobson, Donald Baucom, Gerald Patterson)	Functional analysis Social learning theory Communication and problem solving Attributional style	Resolution of presenting problem Enhanced communication and problem-solving skills Balance between change and acceptance	Communication and problem-solving training Acceptance training/reattribution techniques Parent management, emphasis on behavior/consequences Therapist as educator	Functional analysis, not treatment techniques, is defining characteristic Arguably the most empirically supported of all family and couple methods
Bowen family systems (Murray Bowen, Philip Guerin, Michael Kerr, Monica McGoldrick)	Differentiation of self Triangulation Emotional cutoffs Family emotional system Sibling position	Increased differentiation Detriangulation Cutoffs resolved Improved ability to manage anxiety	Use of genogram Therapist as "coach" Education about multigenerational family processes	Often conducted with only a single patient Influence continues but has waned since Bowen's death
Experiential–humanistic (Susan Johnson & Leslie Greenberg, Virginia Satir, Carl Whitaker)	Communication styles (e.g., placater–blamer) Psychotherapy of the absurd Attachment theory	Fostering creativity (comfort with "craziness") Increased family cohesion Personal growth, self-esteem Tolerance for conflict	Resolving the battle for structure and the battle for initiative Frequent use of cotherapists Family sculpture Use of self Family reconstruction	Since Whitaker's death, symbolic–experiential approach receding in visibility Emotionally focused therapy one of few nonbehavioral or psychodynamic couple methods increasing in influence
Integrative (Alan Gurman, William Pinsof, Douglas Snyder, Froma Walsh)	Importance of and relationships among multiple levels of experience Theoretical and technical eclecticism	Improved communication and problem solving Enhanced interactional insight Presenting problems	Combinations of cognitive-behavioral and psychodynamic methods, in general systems context	Along with behavioral and psychoeducational methods, the most sensitive to research findings Attempts to match interventions to problem and family
Mental Research Institute brief therapy (Don Jackson, John Weakland, Paul Watzlawick)	Distinction between "difficulties" and "problems" Communication processes Symptom as communication First-and second-order change	Resolution of presenting problem Second-order change	Reframing Therapist maneuverability	Historically overlaps strategic, but with important differences (e.g., "function" of symptoms)
Feminist (Virginia Goldner, Rachel Hare-Mustin, Betty Carter)	Raising gender consciousness Underlying effect of gender	Gender equality Recognizing gender inequities	Raising consciousness Identifying and challenging culturally constructed narratives	Not so much a way of intervening as a set of issues to address; has had a pervasive effect on family and couple therapy
Narrative (David Epston, Michael White, Harlene Anderson)	Constructivism "Languaging" "Stories" socially created Problem-saturated descriptions	Resolution of presenting problem Enhancing undiscovered parts of self	Externalization of problems Focusing on "unique outcomes" Creation of new meaning via "re-storying." Therapeutic letters	Currently very popular but rarely acknowledges overlaps with cognitive and behavioral theory
Psychodynamic–psychoanalytic (Nathan Ackerman, Henry Dicks, Clifford Sager, Jill Scharff & David Scharff)	Object relations Projective identification Splitting Scapegoating	Improved insight Genetic (historical) awareness Enhanced empathy De-emphasize presenting problem Disentangle interlocking pathologies	Transference, resistance, and countertransference analysis Creation of holding environment Interpretation Emphasis on therapeutic alliance	Probably influences family therapists' clinical work more than is usually acknowledged
Psychoeducational (Carol Anderson, Ian Falloon, William McFarlane, David Miklowitz)	Biopsychosocial theory Stress-diathesis Expressed emotion	Enhanced family coping skills Improved communication and problem-solving skills Relapse prevention	Survival skills workshops Family management Family information and education Concurrent use of psychopharmacology	Without doubt, the most effective family-based approach for families with a member with major psychiatric disorder
Solution-focused (Steve de Shazer & Insoo Berg, William O'Hanlon, Michele Weiner-Davis)	Core Concepts Focus on solutions, not problems Unimportance of problem etiology Disbelief in resistance	Typical Goals Resolution of presenting problem Creation of solutions	Common Strategies and Techniques Scaling questions Miracle questions Exception questions Client empowerment Coping questions Questions regarding presession change	Comments Grows out of basic Mental Research Institute philosophy, emphasizing solutions more than problems Has made extraordinary claims of effectiveness
Strategic (Milton Erickson, Jay Haley, Cloe Madanes, James Keim)	Power and control Family life cycle Symptom-maintaining sequences Function of problems	Resolve presenting problem Disruption of problem-maintaining sequences	Persuasion Paradoxical injections Insight downplayed Pretend and ordeal techniques	Unfortunate de-emphasis on biology, psychiatric diagnosis, and personality theory
Structural (Harry Aponte, Salvador Minuchin, Jose Szapocznik)	Boundaries Hierarchies Coalitions and alliances Complementarity Engagement–enmeshment	Increased flexibility Adaptability to developmental change Balance between connectedness and differentiation Subsystem functioning	Unbalancing Enactment Joining	Arguably the most influential family therapy approach
Transgenerational (Ivan Boszormenyi-Nagy, James Framo, Laura Roberto-Forman)	Invisible loyalties Ledgers and debts Family mourning Personal authority	More universal than family-specific goals Increased trust Repair ruptured relationships	Multi-directed partiality Family-of-origin sessions Use of co-therapy	Almost all multi/trans-generational methods have significant psychodynamic underpinnings

Family Systems

Family systems formats encompass a broad range of techniques all of which are unified by their belief that the family is a human system with unique properties that contribute to the current emotional and behavioral configuration. Much like a clock with interlocking gears where one does not move without affecting the others, without therapeutic intervention, dysfunctional families become synchronous and stuck over time.

Structural

The most frequently employed subdivisions within the systemic approach are the closely allied structural and strategic therapies. Salvador Minuchin and his colleagues at the Philadelphia Child Guidance center were the originators of *structural family therapy*. Their early work with families that had a symptomatic child formed the basis for the observations about dysfunctional patterns in families with a symptomatic member. In particular, strategic therapists look at boundary issues in families to see where the family lies along the spectrum of over-involvement (enmeshment) and detachment (disengagement). Similarly, those who think systemically look for mechanisms within the family that do not serve them well. The detouring of parental marital conflict by focusing on a child with psychological symptoms is a case in point.

Structural family therapy is a very leader-centered modality that uses directives geared to initiate change by stimulating affect, increasing the therapeutic alliance, and by employing both logical and paradoxical instructions that would induce change in the rigid family system. Once identified, the structural family therapist aims to reset the imbalances in the family system through a variety of therapeutic interventions.

Commonly, the structural therapist tries to help the family solve the current problem by defining clear and attainable goals, establishing appropriate generational boundaries, assisting parents to take leadership positions in the family, offering alternative, often provocative viewpoints that differ from the family's skewed understanding of the problem, and restoring or helping to create functional subsystems within the family.

According to Nichols, structural family therapy follows seven steps: (1) joining and accommodating—reducing initial anxieties about therapy; (2) enactment—encouraging spontaneous interaction among family members to demonstrate the properties of the system and to allow for insertion of behavioral instructions that run counter to the family's established pattern of behavior; (3) structural mapping—broadening the focus from the identified patient to a family-as-a-whole level; (4) highlighting and modifying interactions—turning the volume up or down in order to illustrate a behavioral sequence; (5) boundary making—aligning relationships between subgroups in the family; (6) unbalancing—using joining maneuvers to induce change in an unequal or destructive dyad or triad in the family; (7) and challenging unproductive assumptions–introducing notions that run counter to the established family belief system. The structure of the family with respect to concepts of boundaries, alliances, splits, cut-offs of relationships, complementarity, and positive and negative subsystems within the larger family system are the focal points for the structural approach.

Life Cycle

As an example of another focus under the umbrella of systemic/structural therapies is the *life cycle approach* favored by McGoldrick. This practice uses the in-session drawing of a genogram, similar to a genetic diagram or family tree but with much more verbal commentary by the family. Bowen created this technique to map and flesh out a family's multigenerational history over time. The assumption that underlies this method is that transition points in normal family development can become sticking points for families who now seek therapy. The use of a life stage, transgenerational orientation helps engage resistant families by depathologizing their current struggles and relabeling it as a common and understandable function of family life.

Strategic

Strategic family therapy has its historical base in the early work of Bateson and the Brief Therapy Group at the Mental Research Institute in Palo Alto. Don Jackson, Jay Haley, Virginia Satir, John Weakland, and Paul Watzlawick are famous charter members of this school. Over time, the ranks of strategic family therapists have been joined by other prominent figures in the field of family, including Mara Selvini Palazzoli and the Milan Strategic Therapy group who have incorporated their own variations on the theme of intervening strategically through the use of a team approach, circular questioning, and attempts to provide alternative understanding for dysfunctional or “stuck” family systems.

Noteworthy contributions have come from Jay Haley and Cloe Madanes, building of the hypnotic techniques of Milton Erickson who emphasized a new way of thinking about and changing psychological symptoms. Madanes defined six useful parameters that were essential to understand before formulating a plan of intervention. These include assessing whether the symptomatic behavior is voluntary or involuntary, reorienting power dynamics such that the person who acts the weakest in the family is often the most powerful or controlling, viewing symptoms as part of the communication in the system and determining whether or not maintaining the symptom serves a purpose (albeit maladaptive) for the family; evaluating the hierarchical structure and corresponding role that go with them in order to approximate fairness and equality, understanding the motivational basis (benign or malignant) for problematic behavior, and deciding whether or not the troublesome symptom is a misdirected attempt to gain acceptance love and approval.

Once an appreciation of the presenting problem is attained, a plan for intervening in the system is created. Strategic therapists, like their structural colleagues are very active in sessions and in their use of out-of-session assignments. Stemming from their origins in communication theory, strategic family therapists focus heavily on those forces that generate and reinforce resistance to change in families and work in a brief therapy format alone or with a cotherapist.

Some of the more recognizable techniques include the use of directives, circular questioning, paradoxical messages including prescribing the very symptom for which the family seeks help, assigning rituals or ordeals that disrupt the family homeostasis, and reframing behavior in a less pejorative and pathologic definition to one that emphasizes the protective nature of family or individual defenses. Strategic therapists are keenly aware of the context in which the symptom arises in the family. A child's refusal to go to school but their willingness to go out of the home to pursue pleasurable relationships with peers can be a context dependent issue worthy of exploration in family sessions.

Strategic therapists take positions in therapy that poke and test the response surface of the family system. Depending upon the reaction evoked, a strategy is designed to alter the process. Families can be compliant, noncompliant or pseudocompliant in response to directive or prescriptions put forth by the therapist. Noncompliant and responses that are actually resistances disguised as compliance call for astute identification and construction of tasks that tap the roadblocks to change in the family.

An often-cited criticism of strategic therapy is that it is manipulative and nonempathic. On the contrary, reputable strategic therapists are extraordinarily sensitive to the families they are seeing and are aware of the feelings evoked in them when using the strategic interventions. Additionally, strategic therapists are often called upon to see individuals or families who have been refractory to other forms of therapy and where a family may contain a member with a potentially life-threatening condition such as severe anorexia nervosa. When conventional therapies have been ineffective, the creativity, incisiveness, and pragmatism that typify good strategic therapy will often succeed where others have failed.

Experiential/Humanistic

Experiential/Humanistic family therapy is associated with the names Carl Whitaker, Virginia Satir, and, more recently, Susan Johnson's emotionally focused couples therapy format. The practice of family therapy has undergone dramatic change since its inception resulting in an array of technical formats that rely heavily on the here–and–now experience of the family in therapy. Satir retained her focus on communication patterns but added an emphasis on experiencing emotions as they would arise in a nurturing therapeutic climate. Satir was not afraid of showing compassion by touching family members at times of profound emotional experiences. This approach never encourages boundary violations but does advocate human responses to authentic suffering, struggle, despair, depression, and other intense affective states. A humanistic belief system and an appreciation of the vicissitudes of inadequate or damaging attachment experiences with their consequent feelings of isolation, deprivation, and emotional hunger form the foundation of most of the popular experiential family therapy experiences today.

A positive but not falsely reassuring tone is the backdrop against which therapy is conducted. The person of the therapist is one of the most powerful aspects of this technique. The therapist's positive belief in the family or couple's ability to change, the conviction that people have untapped resources within themselves that can be mobilized in the service of change, contribute to an optimistic but realistic view that resolution of longstanding problems is possible. This is not to say that experiential/humanistic family therapy is only supportive. Confrontation, in sensitive but direct ways also forms a regular component of the treatment philosophy.

Interventions designed to liberate submerged affect in present time can be accomplished through family sculpting, role playing and other action-oriented methods.

The overarching goal of experiential/humanistic family therapies is to validate aspects of the self and promote emotional well-being. Problem solving is a desirable but less urgent goal than it is in the structural, strategic, and behavioral approaches. Perhaps the renaissance of interest in humanistic family approaches has to do with a reaction to the under emphasis on the internal feeling state of the individual that received less attention in earlier cognitive/behavioral methods and the impact of the impersonal, often insensitive, attitude of managed care and third party payers. Regardless of cause, the experiential and humanistic approaches have earned a secure place in present day family therapy.

Psychoeducational

Psychoeducational family therapy has been and continues to be a mainstay in the integrated treatment of people with severe and disabling mental illness. The value of small groups for emotional support and the dissemination of accurate information about a condition shared by members have been amply demonstrated in the group psychotherapy literature. In more recent times, the scope of psychoeducational has branched out to include family problems that are child focused. These include attention deficit disorder, autism spectrum conditions, and a range of behavioral and conduct issues. Families coping with a serious medical condition are also ideal candidates for a psychoeducational experience either as a solo family or in a multifamily group format.

In principle, psychoeducation is a biopsychosocial orientation that provides a vital support function for families in crisis. Initially, psychoeducation helps stabilize a traumatized family system through the provision of a supportive and knowledgeable therapist or interdisciplinary cotherapy team. In the case of medical illness like cancer, a team is ordinarily composed of a family therapist and a coleader with a background in oncology.

Strong emphasis is placed on teaching coping skills to better deal with the condition in question. Much of the original work in psychoeducation addressed schizophrenia and its impact on family life. When applied to psychosis, psychoeducational therapy teaches families several critical skills, including how to tell when someone is showing signs of decompensation, how to arrange for rehospitalization, plans for ensuring the physical safety of all involved, how to be helpful in ensuring that the schizophrenic family member complies with taking medication, and keeping scheduled appointments for out-patient treatment. Techniques to control the level of expressed emotion (EE) to a level that is not overwhelming for the psychotic family member are also essential parts of the psychoeducational approach. Studies have clearly confirmed that lowering the levels of EE is correlated with a reduction of the relapse and rehospitalization rate in patients with schizophrenia. Similarly, psychoeducational methods, in general, have been shown to be more effective than other nonpharmacologic interventions for the prevention of decompensation leading to the need for hospital readmission

Carol Anderson and Ian Falloon are two of the best known exponents of this method and have expanded their use of the psychoeducational model to include home visits to hold family meetings, providing intensive, day-long workshops for families and coping strategy seminars for family and significant others involved in the life of a person dealing with schizophrenia. The thrust of these meetings is to help the family understand mental illness as a biologically driven entity and to avoid the temptation to blame, ridicule, or chastise the affected person. Self-care for those family members who are charged with overseeing the daily life of a psychotic family member is strongly advocated in order to avoid caregiver burnout.

The psychoeducational team's stance is very supportive and, in a nonpatronizing way, synthesizes and translates the medical and psychiatric data to the family in a form that they can convert into constructive action. McFarlane has done just this in his list of guidelines for family and friends of schizophrenics. These guidelines orient the family along several parameters: recovery is a slow process, set realistic limits, avoid harsh confrontation or arguments, comply with the psychiatrist's recommendations regarding medication and frequency of visits, avoid drugs and alcohol, and have modest goals for the patient and expect change to occur in small steps over time. This combination of distilled scientific knowledge and applied common sense are emblematic of the psychoeducational model of family therapy.

Very often, psychoeducational practitioners choose to use the multifamily group composed of as few as two or as many as 10 families as participants. The advantages of this format is that it enhances the support elements for families, destigmatizes the shame in having a psychotic family member allows for cross-family learning by seeing in other families in the group what one is unable to see in themselves, facilitates a broader exchange of information and coping strategies among peers, and counters the isolation that often evolves in families consumed by having to care for a medically or psychiatrically ill member.

Psychodynamic

Psychodynamic family therapy—The thread that runs through psychoanalysis to modern psychodynamic therapies has followed a linear path. Many early family therapists such as Ackerman, Bowen, Jackson, and Minuchin were trained psychoanalysts. Henry Dick's classic book *Marital Tensions* offers an early recognition of the differences in transference and other analytic techniques when working dyadically as contrasted to individual psychotherapy. How to directly apply, adapt, or discard time-honored elements of the analytic model transition to make the transition to family work was the first challenge these therapists faced.

Would it be relevant and appropriate to use the basic cornerstones of psychoanalysis, the analysis of transference and resistance, dream interpretation, the nondemand posture of the analyst, the quest for access to the unconscious mind, and the focus on early childhood experiences as the vehicles for treating families? The history of modern psychiatry tells us that many of the components of psychodynamic theory and practice are applicable to other therapies including group, family, and couple therapies but that significant modifications in technique and attitudinal shifts on the part of the therapist are required in order to do so.

Simply stated, Freudian and earlier analytic models have given way in favor of object relations and self-psychological perspectives. Generally speaking, exponents of the modern psychodynamic school believe that viewing interpersonal as well as individual themes is essential if dynamically oriented family therapy is to succeed. Systemic concepts have been grafted onto the exclusive focus on the inner world of the individual resulting in a broader canvas on which the family therapist works. Empathy, establishing a therapeutic alliance, functioning as a container for intense affects, supplying a secure holding environment for each individual and for the whole family, and providing a tight frame within which to conduct the therapy are all parts of the modern psychodynamic method. Interpretation is used more in this model than in any other family technique. The presenting problem is important but forays into the past life of the family are actively encouraged.

The leadership style is more active than in traditional dynamic therapies but there is ample room for the development of transference. Using transference as the prototype for how psychodynamic family and couple therapists view modify and work with traditional analytically derived concepts with families and couples, the object relations orientation as advocated by Jill and David Scharff affords an excellent example. In this model, transference in family and couples exists on three levels: contextual transference, focused transference, and shared contextual transference.

Contextual transference relates to the early phase of treatment when the role of the therapist is to identify and address, in a supportive and nurturing manner, those initial anxieties that families have about unfamiliarity with therapy, the administrative aspects of treatment (fees, scheduling, physical setting), and anticipating early resistances. The therapist strives to convey a feeling of emotional safety and serves as a positive "holding environment" or container for real or imagined fears about the therapeutic process.

Focused transference is most closely allied to the transference manifestations in conventional psychoanalytic therapy. It consists of the projections onto the therapist, distortions or misinterpretations of the therapist's words or actions, and other present-day phenomena that reflect unresolved issues stemming from early childhood experience.

Shared contextual transference is easiest to see when working with couples. It entails understanding a couple's response to a holding environment, a critical issue for dyadic relationships insofar as one of their major tasks is to create a secure holding environment for each other.

Additionally, recognition of regressive recreations of earlier family configurations, examining triadic transferences, and appreciating the transference issues that determine the process of partner selection in relationships are all illustrative of how transference, a core concept in psychoanalysis has been creatively incorporated into twenty-first century family and couple therapies.

Narrative

It is not possible to comprehensively cover all the emerging schools of family therapy other than to note that narrative and solution focused therapies are other prominent postmodern constructivist theories about family life. The premise in *narrative therapy* is that the shared narratives constructed over time create the perception of reality embraced by the family often resulting in narrow, rigid family myths and misconceptions. The therapy endeavors to broaden the family perspective by helping the family construct new, shared stories leading to more accurate narratives that, in turn, mediate a change to new behaviors.

Solution Focused

Solution-focused therapy is, as the name implies, a quest for positive resolutions to dilemmas with which the family is struggling. Through the use of a series of techniques that involve pinpointed questions geared to expand the cognitive and behavioral repertoire of the family, the therapist guides people toward new ways of thinking about old issues relying heavily on positive reinforcement, an optimistic attitude that looks toward solutions not problems, and a search for exceptions to the situations that the family feels mired down and powerless.

CLINICAL APPLICATIONS

General Clinical Principles

Family therapy and individual psychotherapy may be viewed as opposite poles of a spectrum of therapeutic interventions each requiring different modes of engagement. Individual therapy is a new relationship where interaction between therapist and between therapist and patient stimulates affects that form the basis for examination in dynamically oriented individual work. Initial goals are to engage the patient and create a relationship into which strong feelings, both past and present will develop.

In family and couple therapy, the opposite is the case. Families and couples have relationships and histories that predate the new therapy and usually have a surplus of affect that is out of control. Coming in when a crisis has arisen is a frequent entry posture of couples and families. These circumstances call for decelerating interventions designed to stabilize the system. Affective intensity needs to be reduced to a degree that the family system can gain a sense of equilibrium. The family therapist has to find a point of entry into the current family configuration. Doing what is appropriate in early individual therapy would definitely produce untoward results in the family setting. With this principle in mind, it makes clinical sense to understand the strengths and limitations of family and couple therapies and for whom they are best suited and the common clinical management issues that occur that require thoughtful treatment planning and interventions.

Indications and Contraindications

Family and couple therapies are relational therapies. The clinician is primarily focused on the unresolved interactional issues that the family or couple presents at the time of the initial interview rather than on individual diagnoses. Despite this, since family and couples therapy has been demonstrated to be useful in specific psychiatric conditions, some practitioners prefer to use DSM criteria for deciding when family therapy is indicated. In these circumstances, conditions including schizophrenia, bipolar disorder, major depressive disorder, and substance abuse are the yardsticks by which a determination is made about suitability for a family or couples approach. The majority of family therapists prefer a more systemic or group-based orientation when thinking about the indications and contraindications for family or couples therapy (Tables 32.20–2 and 32.20–3).

Table 32.20–2.
Indications for Family and Couple Therapy

1. When problems exist across generational boundaries—Grandparents feel excluded from access to grandchildren, conflicts between a father and son or daughter working together in a family business, sharing the living space with more than one generation in the family.
2. At significant times in the family life cycle—Leaving home, birth of a child, contemplating a geographical move, death in the extended family.
3. Whenever a child or adolescent is symptomatic and designated as “the problem”—Adolescent acting out with drugs, alcohol, and promiscuity; cutting; children who refuse to attend school and/or have psychosomatic symptoms.
4. When a couple or family calls or comes in requesting a family therapy.
5. When serious medical or psychiatric illness is present in a family member—cancer, schizophrenia, bipolar disorder.
6. Families stressed due to racial, cultural or religious heritage, interfaith and interracial partnerships.
7. Issues that stem from recombined or blended families—friction between step-parent and biologic child; problems with role definition as parents; children who feel themselves to be in a loyalty bind between divorced parents.
8. Families in which someone intrudes upon or sabotages another member’s individual psychotherapy—intrusive parents threatened by the independence of the adolescent or young adult in individual therapy; parental abuse of power by refusing to pay for treatment.
9. Families or couples where societal pressure and prejudice has significant emotional spill over into family life—gay parents with children; older couples contemplating adoption of a child; common-law partnerships.
10. When sibling issues are beyond the control of the parent(s)—physically assaulting a younger sibling; excessive competition for parental attention and affection.

For couples:

1. When the chief complaint is presented as a sexual issue—This may indeed be an indication for sex therapy but other unresolved relationship issues masquerading as a sexual problem should be ruled out first.
2. Following the discovery or disclosure of an affair—the threat of divorce or break up; restoring the power imbalance in the dyad.
3. When communication problems are the basis for seeking treatment—escalating anger; repetitive arguments without resolution; emotional detachment and break off of communication.
4. When partnerships are ending and children are involved—partners seek therapy not to resolve relationship issues but to minimize damage to children; decision-making that involves coparenting.

Table 32.20–3.
Contraindications for Family and Couple Therapy

1. When a family has an emotionally unstable member or members and the risk of stirring up intense affect in session might lead to decompensation or other adverse effects—schizophrenia, major depression, impulse control disorders, addictions, dementia.
2. When there is a history of violence in the family—elder abuse; domestic violence, drug and alcohol dependence.
3. When a family is collusive, rigidly unified and inflexible in considering that there may be a problem in the family/couple.
4. When therapy runs counter to the belief system in the family—sense of shame; religious or cultural beliefs discourage therapy.
5. When essential members of the family cannot or refuse to be included in the sessions—father feels that “real men” handle problems on their own; family member offers work schedule as a reason for being unable to participate.
6. When family therapy should not be the first-line treatment—detoxification of addicted family members should precede decision for family/couples therapy; psychotic family member who is not adequately medicated.

For couples:

1. When a pairing defense is so strong that both partners recoil at the suggestion of couple therapy; rare relationships where a folie a deux exists such that the couple has a shared delusional system and/or a rigid, ego-syntonic view that they do not require therapy.
2. When the level of motivation is unequal or deceptive—one partner has not disclosed his/her plan to leave the relationship.
3. When a member of the couple has an active relationship with someone other than the partner—extramarital affairs; homosexual relationship simultaneous with heterosexual primary partnership.
4. The threat of domestic violence or crisis creation in vulnerable partnerships—history of physical abuse.
5. In cases of documented child physical, sexual and/or emotional abuse, extreme caution should be exercised before exposing the

child to the abusing parent(s) even in an ostensibly therapeutic and safe climate of family therapy.

Table 32.20–2 is an integrated model that incorporates both diagnosis and presenting problem as criteria for recommending a family or couple form of therapy.

Assessment and Evaluation

Evaluation procedures take numerous forms in family therapy. It is important to note that aside from data collecting, the initial contact with the family or couple may simultaneously be the first session of their family therapy experience. The most common approach is to see the whole family for family therapy evaluation and both members of the couple for couple therapy. The closer the therapist subscribes to a strict systemic approach the more the requirement that all members be present. Alternatively, other family therapists work with whatever portion of the family system presents itself for the initial session with the thought in mind of recruiting absent members by using the existing forces within the family. The usual time allotment for a first family or couple session is 1 hour but time-extended sessions are common if, for logistical or clinical reasons, a longer meeting would allow for the most comprehensive view of the family/couple system.

Table 32.20–4.
Couples Therapy Initial Interview

1. Orient the couple to the joint format—emphasize the need for both partners to be present in order to have input and how couple therapy differs from individual psychotherapies.
2. Request that each person state his/her perception of the problem—a majority of couples experience the locus of disharmony as being in the partner rather than in themselves.
3. Try to engage both partners in the first session—balance all interventions especially praise and confrontation such that neither partner feels rejected.
4. Identify and clarify relationship patterns and structure—is the present problem ongoing since the relationship began or is it in reaction to new circumstances.
5. Formulate the problem without singling out either partner—shift the focus from the identified patient model to the identified problem model—the question is often posed in the following form, “How long has this been a problem for the two of you?”
6. State the purpose and some goals of couples therapy in nontechnical language or jargon—to resolve communication problems; to learn how to fight fairly; to address specific issues around parenting.
7. Think of ways of gaining entry into an established system—wait for a naturally occurring pause before entry or try to interrupt a pattern to see how the couple responds.
8. If couples therapy is indicated, try to define an initial treatment contract and get agreement of both partners to participate—plan for a next meeting and give some indication of what the general focus might be, i.e., communication issues.
9. Anticipate early resistances—denial defenses with alcohol and drug related issues; the temptation to stay with familiar behavioral patterns even though they may not be working but are less frightening than the prospect of change through couple therapy.
10. Clarify myths and misconceptions—joint therapy is more likely to lead to dissolution of the partnership; being in joint therapy is an indication of an unworkable relationship.

11. Leave time for the couple to ask questions about anything in the session that was unclear or worrisome—common concerns are the expected duration of therapy, cost factors.
12. Assign “homework” if indicated—instruct them to arrange to spend 5–10 minutes each day or alternate days in face to face conversation on any nonconflictual subject without outside distractions that include television, children, or other impediments to engagement with one another.
13. Summarize what you want the couple to take away from the first meeting—point out the themes that appear to be urgent ones; commend the couple for being open enough to explore the possibility of joint therapy.
14. Delineate what your role will be in the therapeutic process—make sure to emphasize how a collaborative treatment model works and that the therapist will be an active participant in the ongoing therapeutic process and will be unavailable for individual phone contacts or sessions that usually result in telling “secrets” to the therapist.
15. Address “housekeeping procedures”—appointment times; frequency and length of sessions; fees; what happens when one partner cannot attend a session; policy about individual sessions during the course of joint therapy; get permission that is HIPPA compliant for communication with other (individual) therapists; obtain informed consent from both partners.
16. Stop on time unless there is a crisis which cannot be handled in any other way than by extending the session—commonly a catharsis of intense feelings is opened up in a first meeting, allow time for both expression of emotion and putting a cognitive structure on it in order to gain closure on the session.

Seven global areas are subject to assessment in the first meeting with a family or couple: (1) the presenting problem (stated as the family defines it); (2) communication patterns; (3) styles of problem solving; (4)regulation of affect; (5) family roles—alliances, splits, coalitions; (6) family rules; (7) impulse control; (8) and the prevailing family beliefs, traditions, and narratives.

Pragmatically, the data for the assessment comes from the therapist's observation of the spontaneous interactional patterns and from the specific questions posed by the interviewer (Table 32.20–4). Many of the techniques used in family and couple therapies are interchangeable and applicable to both. The material that follows details the conduct and content of an initial evaluation of a couple with particular emphasis on the role and posture of the therapist but the points outlined apply equally when modified for working with a family format.

These are the general clinical tasks for practitioners of diverse theoretical orientations to cover in the first contact with a couple. In order to ensure that relationship-relevant themes are included in the interview, it is essential that the therapist have in mind a set of variables that will glean the information they need in order to understand the partnership issues and to begin formulating a treatment strategy. Table 32.20–5 lists a practical set of questions and areas of observation that will yield the information most relevant to creating and directing the ongoing treatment process.

When converting principles into practice, the process is rarely without potential pitfalls that call for prompt recognition and astute clinical management. In both family and couple therapies, it behooves the therapist to anticipate that there will be resistance to change. More often than not, this is fear based and can be traversed if the therapist is alert for common roadblocks to progress. To illustrate this point, a representative sampling of the types of clinical management decisions that confront the therapist working with couples and some guidelines for handling them are highlighted in Table 32.20–6.

Table 32.20–5.
Specific Areas of Inquiry and Observation in Couple Therapy

1. Why do they seek evaluation now?
2. Opening position/posture of each person in the session
3. Locus of the problem
4. Communication styles
5. Collaborative abilities
6. Problem-solving preferences—insight or behavioral preferences
7. Families of origin of both partners—three generational model starting with grandparents.
8. Overt and covert relationship "contract"
9. Methods of managing intense affect
10. Function of the presenting symptom(s) in the relationship and/or family system
11. Assessment of individual as well as joint strengths and assets
12. Power and control issues
13. Role definitions and gender-based issues
14. Mechanisms used to control interpersonal closeness and distance
15. Work function of each partner
16. Sexual issues
17. Prominent defenses—individual and joint
18. Duration of the disharmony
19. Examples of historical crisis situations and their outcome
20. Motives for selecting each other initially and for staying together currently
21. Parenting issues
22. Reaction of each family of origin to the partnership
23. External stressors on the relationship—work, time, money, illness, sleep deprivation
24. Interpersonal skills outside the relationship
25. Ability to have pleasurable and restorative experiences
26. Emotional resiliency of each partner and of the relationship
27. Ethnicity, religious and cultural, and gender issues
28. History of significant intimate relationships prior to the present one

29. Boundary issues—within the relationship and between the relationship and significant others, especially with original families
30. Involvements outside then relationship—“Triangulation”
31. Substance use and abuse
32. Joint and individual life cycle factors
33. Questions about couples therapy

Table 32.20–6.
Common Clinical Management Issues

1. The absent or resistant partner or family member—joint sessions are not held if one partner (with families, if significant family members) is not present. Doing so runs the risk of unbalancing the therapy, being perceived as participating in a collusive relationship with the partner who is present and risks being told “secrets” by the person seen individually. It also tacitly appears to be sanctioning the idea that that the presence of both partners is not necessary for successful couples therapy.
2. Distinguishing true emergencies from “pseudocrisis” situations—a common resistance to dyadic therapy is to try to create a crisis designed to have an individual relationship with the therapist. The form this very often takes is an urgent individual phone call between sessions. The therapist has to rapidly assess whether there is indeed a true emergency that requires immediate attention or if this is an effort to sabotage the joint therapy. In the latter case, the caller needs to be informed that the content of the call will be shared with the other partner in the next couple session.
3. Communication with other current or past therapists—pivotal in formation can be obtained from prior individual, family, or couple therapists referable to what has worked well and what has not. Obtain informed consent from both partners but make this a nonnegotiable condition of therapy.
4. Problems with payment for sessions—delayed payment, disputes about fee arrangements, and a host of money-related issues not only serve as points of resistance but may also be informative that money matters are an issue for the relationship itself.
5. Management of psychotropic medication issues both pharmacologically and psychodynamically—when one member takes medication it is important to insure that that person is not seen in disadvantageous or “one down” role.
6. Management of “secrets” in relationships and in original families—the therapist should make every effort to avoid situations in which one partner tells something that is not known to the other. This presents a threat to the viability of the entire therapeutic enterprise and must be handled immediately. Not permitting

individual sessions is the most common method of handling secrets. If, by chance, a person tells a secret by phone or by lingering after the other partner has left at the end of the session, the therapist needs to clearly communicate that it is the person's responsibility to share this information openly, or else the therapist will decide how best to reveal the content to the absent partner.

7. What happens when one partner plans to or has seen a lawyer—many individuals have different motives for seeking couples therapy. In instances where one has contacted a lawyer, the underlying agenda may be one of guilt reduction by leaving the partner in the hands of the therapist rather than his/her stated desire to work on the relationship. In these instances the covert actions of the dishonest partner is a priority issue to be confronted in the next session. The obvious risk is to have one person exposed and vulnerable thinking that they are working toward problem resolution, only to be abandoned by their partner who plans to end the relationship.
8. Videotaping of sessions—the purpose of taping whether for review by the therapist between session to show the couple aspects of problematic and effective communication or for purposes of training and supervision always need to be clearly explained. Obtaining informed consent from both partners is a precondition for audio or videotaping sessions.
9. Confidentiality—as is the case in all psychotherapies, confidentiality is a sine qua non of ethical family and couple therapy practice. Unique forms of confidentiality with couples can occur in a couples group where several couples are privy to intimate details in the lives of other group members. Breaches of confidentiality are grounds for dismissal from the group.
10. Whether or not (and when) to include children or extended family members in sessions—couples with children and those with living parents frequently raise issues in joint sessions that involve family members who are not present. One of the unique aspects of couple and family therapy is flexibility. Couple formats allow for the inclusion of other family basis or for conversion of couples therapy

to larger family groupings in order to address issues in the larger system.

11. How to handle intense negative affect in sessions—rage, blaming, threats to divorce, potential physical violence are all alarming phenomena in couples therapy sessions. The therapist has to have a tolerance for intense affect, know when to intervene before a dangerous situation escalates, set a clear set of ground rules for the modes of expression of anger and help the couple find more viable mechanisms for dealing with conflict.
12. How to avoid repetitive, unproductive, or destructive interchanges in sessions—a consistent property of systems is the repetitive, cyclical nature of behavioral sequences. Couples stalled in these ways tend to get demoralized and despair over the fate of their relationship. Couple therapists can supply input in the form of feedback both of a general nature and specific observations in the here and now in order to supply new information as fodder for interrupting the recycling of old arguments.
13. How to provide a different experience in the session from what the couple would be inclined to do spontaneously—asking a couple to do something in the session that they are not doing on their own is a central concept in all couple and family therapies. Asking a couple with a problem of sexual inhibition and avoidance to spend the meeting talking about sex allows them to experience that they possess the capacity to do something in the session that they feared might not be possible.
14. What constitutes appropriate timing of interventions in treatment—although couple and family therapies are brief therapies as compared to individual psychotherapy, they still go through definable early, middle, and termination phases. Therapists need to time interventions accordingly. Building a therapeutic alliance must precede the middle or working stage of therapy; otherwise, interventions will be ineffective. Similarly, termination in couples may feel abrupt to patients who have been used to the ritualized, extended termination techniques of individual approaches.

15. Cotherapy considerations—when a team treats a couple, many advantages accrue but potential pitfalls need to be monitored as well. Splitting the cotherapists is the most common defense against the team format. Opposite-sex coleaders are seen as parental figures and mobilize feelings related to the family of origin in both partners. Ideally, the advantages of cotherapy are the increased option for modeling aspects of healthy communication, respect for differences, and a working partnership between adults.
16. How to make the therapeutic experience one that is sensitive to issues of diversity—It is essential that the couple therapist make themself finely attuned to the variables that fall under the heading of issues related to diversity in the partners that they are seeing. Diversity requires modification of therapeutic understanding and technique in ways that are respectful and free of bias and potentially negative preconceptions. Race, social class, culture, disability, age, gender, financial status, ethnicity, sexual preference, and religion are noteworthy examples of factors that must be incorporated into any treatment plan in order to avert formulaic approaches that disrespect or ignore facts that are central to the identities of the partners. The therapist needs to be vigilant in self-monitoring so that their own beliefs do not insinuate themselves in ways that would hinder or destroy the therapeutic alliance.

COUPLES GROUP THERAPY

Few modalities reflect the interface between family, couples, and group psychotherapies better than a group composed of couples. Couples groups also exemplify the growing trend toward more integrative approaches in family work. In spite of their popularity and utility, couples groups are poorly described in the professional literature.

The prototypical couples group is an out-patient modality composed of three to six couples led alone or by a cotherapy team. The group typically meets once per week for a session of one and a half hours. Groups can be brief, time-limited cognitive/behavioral or skills acquisition training models or longer-term psychodynamically focused models with more ambitious individual as well as couples goals.

The benefits for patients of including couples in groups is to counter isolation, destigmatize the idea of having relationship problems, increase the options for handling issues differently by seeing how other couples manage them, enhance the possibilities for varied feedback, experiment with new communication opportunities, and counter depressive and demoralized states that invariably accompany the frustration of unsuccessful efforts to make progress.

Couples groups have distinct advantages for the therapists as well. Coleadership is very desirable in terms of sharing the workload. Working with a partner allows for role modeling and demonstrating behaviors for the couples in the group that reflect alternative modes of handling individual differences, equally distributing power and control, clarifying roles of men and women as individuals and as partners, and opening up issues that couples handle by avoidance. The presence of a coleader permits the balance between therapist support and confrontation of members by having a back-up system not present when one works alone. When it is necessary to confront denial defenses in a group of couples with substance abuse issues, cotherapy facilitates one leader's ability to take a stronger stance than they might when working solo. If necessary, the other therapist can offset any unforeseen negative fallout if it occurs. Alternating these roles between cotherapists over time demonstrates productive styles of collaboration for the group to experience.

For the psychodynamically oriented couples therapist, seeing couples in a group milieu exponentially increases the number and variety of transference possibilities in the leadership and membership vectors. The cotherapy team is often perceived by members as having a therapeutic marriage despite the group's awareness that they might be colleagues who have chosen to work with each other. It is also wise to choose coleaders who have different therapeutic styles but who show mutual respect and readily incorporate suggestions from their coleader. Therapists with an object-relations preference can see the inner world of members enacted in their transactions with other part-objects represented by the group leader(s) and fellow members in the group.

Case Vignette

During the middle phase of a couples group composed of four couples, the theme of whether or not to have children arose. One couple had just come from a visit to the gynecologist, who informed them that they were running out of time because of the wife's age. The woman in the couple did not want to have children, but her husband did. His complaint about the marriage was that his wife never was demonstrative in showing her loving feelings for him. He felt her to be detached, distant, and sexually inhibited.

The prevailing sentiment among the other couples who had children was that children only added additional stress to an already stressed relationship. One other couple, however, voiced their different view by describing how their children had enriched their lives. As the talk about going forward and getting pregnant progressed, the group leader noted the nonverbal communication between the ambivalent couple. Whenever the tone of the group leaned toward having children, the wife would reach out and grasp the hand of her husband in a tender way. This invariably had the effect of stopping him from pursuing the topic for fear of the withdrawal of the affection he hungered for. All this transpired without words. Once identified, this repetitive nonverbal pattern was available for examination in the group, and the supportive elements provided by other members and the leader encouraged a frank, direct, and open conversation between the partners, who eventually chose to go forward and attempt to have a child.

Couples groups share many commonalities with multifamily groups in terms of cross-relationship learning, psychoeducational elements, strong support elements common in homogeneous groups, and as a vehicle for reality-testing ideas that are inhibiting progress.

Therapists who are experienced with the couples group modality frequently report that they have more therapeutic leverage once an initial base of group cohesion is established. Group cohesion acts as a supportive buffer that allows the therapist greater freedom to use appropriate anxiety-evoking interventions or apply group pressure as a force for change.

A common indication for couples groups is for couples with problems of long duration (years rather than months) where it is felt that the combination of cohesion, a strong therapeutic alliance, the multiplicity of feedback from the group and the leaders, plus the option to apply pressure when needed add up to the empirical enthusiasm expressed for this format.

RESEARCH AND FUTURE DIRECTIONS

Family therapy research presents several challenges. In the broadest sense, family research to date is a robustly evolving process resulting in a much clearer delineation of the significant variables to be taken into account in order to produce clinically relevant, scientifically informed studies. Part of the problem in conducting credible family and couples research on outcomes resides with the multidimensional nature of the individual, couple, and family and therapist levels present in all family work. What and who is being measured are central concerns, as is the selection process of how cases are being assigned for particular therapeutic interventions. Therapist variables such as theoretical orientation, level of experience, how issues of diversity are managed, gender similarities and differences, and the precise form and timing of technical choice are all being studied in the current research climate in family and couples therapies.

As a result, more specificity is being applied to studies in constructing instruments for measuring the outcome of family approaches, understanding what contributes to the process of change in family work, and in randomizing families under study. In particular, greater attention is being paid to clarity of intervention technique with tendency to use those methods that can be replicated under in other settings by other practitioners. Closely allied to this is the notion of goal specificity in family and couple therapies. In cases where a family seeks therapy for help with a member with a chronic mental illness, the target goals can be reduction of caregiver burden, change in familial interaction patterns, regulating affect in the family, improved family ability for shared problem solving, or all of the above. Clear and attainable goals make outcome appraisal of treatment an easier process. This is one of the welcome emergent trends in both family and couple research.

Comparisons of different schools of family and couples therapy have not demonstrated superiority of one technique above another. Studies have shown documentable empirical support for behavioral and long term, supportive, psychoeducational family therapy for schizophrenia. The Community Reinforcement Approach is a representative example of a behaviorally based form of social network therapy for alcoholism and substance abuse, and a range of childhood behavioral problems that includes adolescent delinquent behavior and substance abuse, phobic states in preadolescent children, and aggressive behavior in children and teenagers. Problems associated with various aspects of parenting were demonstrated to show benefits following parental stress management training and behaviorally focused parental problem-solving education.

Other interventions that show promise in early studies but lack sufficient large-scale data at present include parent management training (PMT), functional family therapy (FFT), and multisystemic therapy (MST). Adjunctive family and/or couples therapy have a strong foothold as part of a more global treatment plan for conditions that encompass anorexia, alcohol and substance abuse, and bipolar disorder and schizophrenia.

Where great progress is being made in family and couples research is in aligning research findings with direct clinical application, thereby reducing the gap between family studies and family practice.

The state of couple therapy research is also showing great promise in defining those attributes that contribute to better outcome. Of the three prevailing orientations in couples therapy, psychodynamic, behavioral couples therapy (BCT), and emotionally focused therapy (EFT) only the latter two have found empirical research support for their utility. Social exchange theory informed BCT while attachment theory formed the underpinnings for EFT.

BCT has added cognitive components to the model, but research to date has not demonstrated the appreciable advantage of the combined approach. EFT, on the other hand, has been found to be valuable in several areas of couple functioning: emphasizing the attribution of meaning that partners incorrectly ascribe to their partner's verbal communication and behavior; teaching the exchanges that promote secure attachment; identifying and editing and supplying alternatives to the number and forms of negative cycles of behavior; stressing the beneficial use of intense emotion, and delineating maladaptive interactional patterns that lead to the dissolution of the relationship.

Some specific observations that have emerged from couples therapy research are in the behavioral approaches, lower levels of initial distress were correlated with a better outcome. In EFT and psychodynamic approaches, the stronger the therapeutic alliance and the more both partners of the couple are engaged in therapy, the better the outcome in terms of problem solving, communication alignment, reduction of hostility and blame, perceived partner sensitivity, and a softening of the negative affective tone present upon entry into treatment.

Future directions in family and couples therapies will include closing the gap between statistical studies and their real-world application, a quest for manual-driven therapy protocols that can be standardized for use by many practitioners, the emergence of suggested treatment guidelines for the condition that causes the family/couple to seek help; attention to ethical practice standards that deal appropriately with issues of confidentiality in multiperson therapies; treating specific conditions that include major affective disorders, personality disorders, and developing more sophisticated measures for assessing the process of change and the outcome of therapy. The wingspan of family and couples therapies is expanding and the field will continue to flourish as it accumulates greater experience and specificity with the elements most conducive to positive outcome.

The Covid Pandemic has put families and couples at great risk: separation, isolation, uninterrupted time in confined spaces, increased substance abuse, increased domestic violence—physical and emotional, financial concerns, over and under expression of feelings, increase in existing conflicts, and divorce, for example. For those seeking treatment for these issues, Tele-Mental Health has been very useful and may well persist beyond this disruptive disaster in the world.

FURTHER READINGS

Atkinson BJ. *Emotional Intelligence in Couples Therapy: Advances from Neurobiology and the Science of Intimate Relationships*. WW Norton; 2005.

Bigner JJ, Wechtler JL, eds. *Relationship Therapy with Same-Sex Couples*. Haworth Press; 2004.

*Clarkin JF, Carpenter D, Fertuck E. The state of family therapy research: a positive prognosis. In: Sholevar GP, Schwoeri LD, eds. *Textbook of Family and Couples Therapy: Clinical Applications*. American Psychiatric Publishing, Inc.; 2003.

Fields L., Morrison TL, Beels CC. Couple and family therapy. In: Hales RE, Yudofsky SC, eds. *The American Psychiatric Publishing Textbook of Psychiatry*. 4th ed. American Psychiatric Publishing, Inc.; 2003.

Glick ID, Berman EM, Rait DS, Haru A, Ascher M, eds. *Couples and Family Therapy in Clinical Practice*. 4th ed. Wiley, London & Singapore; 2016.

*Gurman AS, Lebow JL, Snyder DK, eds. *Clinical Handbook of Couple Therapy*. 5th ed. Guilford Press; 2008.

Johnson SM. *The Practice of Emotionally Focused Couple Therapy*. Brunner-Routledge; 2004.

Johnson SM. Couples therapy research: status and directions. In: Sholevar GP, Schwoeri LD, eds. *Textbook of Family and Couples Therapy: Clinical Applications*. American Psychiatric Publishing, Inc.; 2003.

*Lebow J, Chambers A, Breunlin DC, eds. *Encyclopedia of Couples and Family Therapy*. Springer Publishing; 2019.

*Kelly S, ed. *Diversity in Couple and Family Therapy: Ethnicity, Sexuality, and Socioeconomics, ABC-CLIO, LLC.* Santa Barbara, California; 2016.

McFarlane WR. *Multifamily Groups in the Treatment of Severe Psychiatric Disorders*. Guilford Press; 2002.

Nichols MP, Schwartz RC. *Family Therapy: Concepts and Methods*. 7th ed. Pearson Education Inc.; 2006.

Pinsof WM, ed. *Special issue: Marriage in the 20th century: trends, research, therapy, and perspectives*. Family Process; 2002; 41:2.

*Shackelford TK. *The Sage Handbook of Domestic Violence*. Thousand Oaks, California; 2021.

Sholevar GP, Schwoeri LW, eds. *Family and Couples Therapy: Clinical Applications*. American Psychiatric Publishing; 2003.

Sprenkle DH. *Effectiveness Research in Marital and Family Therapy*. AAMFT; 2002.
Watzlawick P, Weakland JH, Fisch R. *Change: Principles of Problem Formation and Problem Resolution*. Guilford Press, 2002.

32.21 Group Psychotherapy

HENRY I. SPITZ, M.D., D.F.A.P.A., D.F.A.G.P.A.

Group psychotherapy has always been an evolving field: sensitive to changes in the cultural climate, reactive to the lessons learned from history, cognizant of world events, aware of developments in scientific research, open to new systems of diagnosis of emotional illness, and updating theoretical orientations and clinical technique in the spirit of providing a level of care that is relevant, effective, and capable of being incorporated into new health care delivery systems.

Over the past years, a range of issues have been integrated to expand group therapy themes. Examples include the results of trauma from sources as broad as posttraumatic stress disorders in soldiers returning from war and others who have experienced childhood psychological, physical, and sexual abuse; gender issues including homosexuality, transgender, lesbian, and bisexuality; internet and computer themes including internet addictions, pornography, gambling, and hand-held phones leading to interpersonal problems; new discoveries in brain structure, genetics, disease, and the course and therapy of central nervous system disease; borderline personality disorder has been studied and dialectic behavioral group therapy has become a successful intervention.

Groups for schizophrenia and serious inpatient and outpatient psychological conditions have earned prominence with the emphasis on cognitive problems, behavioral skills, and socialization learning for its members. Relationship problems of many kinds are common as well. Heterogeneous couples have witnessed an increased divorce rate and often seek group therapy as a useful tool in exploring their marriage. New relationship forms that have emerged include same-sex couples, families that consist of a single parent, and homosexual parents.

Group therapy as a core element in the treatment of substance abuse and alcoholism continues to develop. Different group approaches and treatment opinions vary from classic 12 step or alcoholics anonymous groups to mid-range groups as in Recovery Incorporated and lastly, a harm reduction group model where members are taken into group without present sobriety being a requirement.

The list of variety of group approaches is wide ranging. Medical illness, affective disorders, bereavement and isolation, eating disorders, psychopharmacologic problems, geriatric groups, combined individual and group treatment, self-help groups, adolescent and family groups, are a brief overview to set the stage for what follows in this chapter.

FOUR PHASES OF GROUP PSYCHOTHERAPY

The field of group psychotherapy can be conceptualized as being in a new phase of its modern history. The first wave, ushered in by World War II and the enormous amount of psychological carnage that emerged from the war created a pressing need for treating the psychiatric casualties of wartime. Working in a group setting, as opposed to the conventional individual psychoanalytic milieu, facilitated seeing more people in less time and led to an exponential increase in the growth of psychotherapy groups predominantly those of a psychodynamic nature. The paradigm for phase two of group therapy’s development was the community mental health movement in the late 1960s and 1970s. The mandate to deinstitutionalize the chronically mentally ill patient from state hospitals back into the community, presented a set of circumstances similar to those following the Second World War—a surplus of those in need of psychological help and a shortage of professionally trained therapists to treat them. The practice of group therapy, particularly in the areas of homogeneous groups designed to treat members who shared a common condition expanded during this period. With greater comfort and familiarity in employing group techniques came a corresponding trend toward experimentation with the application of small and large group methods to meet the needs of a broad range of patient populations and psychiatric conditions. The 1990s was the age of health care reform and the third phase. With it again came the pressure to see more people in shorter time frames but the mandate this time had come from a different source: managed mental health care. These four emotionally charged words reflect an economically motivated model of health care delivery that places a high premium on time and cost-efficiency, often at the expense of good patient care. Despite its many drawbacks, managed care allowed creative group therapists to expand our knowledge about brief, symptom-focused treatments and the use and possible hazards of time-limited groups.

The first decade of this century has borne witness to the fourth phase of group work ushered in by the rising threat of terrorism, human and natural disasters, and, consequently, to the influx of group therapies designed to address issues of psychological trauma. The immediate and longer-term reactions following catastrophic events commonly include panic, emotional anesthesia, alienation, social isolation, a profound sense of vulnerability, and a host of symptoms, some of which result in true posttraumatic stress disorder syndromes. The group setting is ideal for help in healing those whose lives have been touched by traumatic events.

DEFINITION

Group psychotherapy is a modality that employs a professionally trained leader who selects, composes, organizes, and leads a collection of members to work together toward the maximal attainment of the goals for each individual in the group and for the group itself. Certain properties present in groups such as mutual support can be harnessed in the service of providing relief from psychological suffering and supply peer support to counter isolation experienced by many who seek psychiatric help.

Similarly, homogeneously composed small groups are ideal settings for the dissemination of accurate information about a condition shared by group members. Medical illness, substance abuse, and chronic and persistent severe psychiatric conditions including schizophrenia and major affective disorders are cases in point.

HISTORICAL ROOTS

Group therapy has a long and fascinating history spanning more than one hundred years. Early efforts from varied sources are directly traceable to elements of clinical practice today.

Most historians agree that the first recognizable group intervention was conducted in 1905 by Joseph Pratt, an internist treating tuberculosis patients in Boston. The treatment for tuberculosis at the turn of the century consisted of isolating afflicted patients from the general population and housing them in long-term care facilities. Pratt conducted “classes” on the ward designed to educate patients about what was known at the time about their disease. A serendipitous observation by Pratt was that people who attended class regularly seemed to be more compliant with treatment protocols of that era and also seemed to have improvement in their state of depression.

Parallel efforts took place in 1921 when Edward W. Lazell applied the group method to schizophrenic and manic–depressive patients at St. Elizabeth’s hospital in Washington, D.C., and observed similar results. Patients who participated in group sessions appeared to be less depressed and isolated as well as being more accurately informed about their psychological condition and the rationale for the nature of their therapy.

At the same time, Alfred Adler and Rudolph Dreikurs departed from the exclusively psychoanalytically based dyadic models of individual psychotherapy by involving significant others in the lives of the children they were treating at the Vienna Child Guidance Clinic. Parents, schoolteachers, and clinic staff were directly involved in small group sessions designed to integrate the therapy of the children and adolescents with whom they were connected.

L. Cody Marsh, a minister working at Kings Park State Psychiatric Hospital in New York, used a revivalist–inspirational style to help bridge the alienation he felt was at the root of many of the psychiatric problems of the time. Marsh's now famous quote, "By the crowd they have been broken, by the crowd they shall be healed" captures the spirit of his orientation to mental illness and his emphasis on the need for the social support provided by therapy groups in helping patients cope with their problems.

The 1920s saw a confluence of interest in group dynamics and early group therapies. Sigmund Freud addressed the properties of large groups and their potential for good or dangerous influence in "Group Psychology and the Analysis of the Ego." Jacob Moreno, the founder of psychodrama, was working in Vienna experimenting with the use of intense affect and physical activity in a group setting when he created his "Theater of Spontaneity." In addition to creating psychodrama, Moreno is also cited as the first person to use the term "group psychotherapy" in conjunction with his applications of the psychodramatic technique.

In 1925, Trigant Burrow applied the group method to psychiatric outpatients and is credited with creating the term "group analysis" to describe his technique. In addition, Burrow was interested in the dynamics of small groups and created some of the earliest group "laboratory" formats in which the task of the group was to examine its own internal interaction. Burrow's work represented an extension of Freud's individual focus to the larger social context in which an individual existed.

The 1930s were characterized by efforts to expand the interface between psychoanalysis and group therapy. Louis Wender, in the early 1930s, formed six- to eight-member groups of patients and met with the group two or three times per week. Paul Schilder, working at Bellevue Hospital in New York City was also enthusiastic about outpatient psychoanalytic groups for patients with a broad range of emotional problems.

A key figure in the evolution of group psychotherapy for children and adolescents was Samuel Slavson. Slavson is considered by many to be the founder of modern group psychotherapy. Slavson focused on both the individuals and the group itself. His innovative work with activity and play group therapies was based heavily on using the therapy group as a vehicle for observing behavior as participants interacted in group sessions. He combined an analytic superstructure to his observations and utilized the analysis of transference and resistance as mainstays of his technique. Slavson founded the American Group Psychotherapy Association and was also the organization's first president.

The marriage of psychoanalysis and group psychotherapy as practiced by contemporary psychodynamically oriented clinicians is closely associated with the work of Alexander Wolf. Wolf was intrigued by the challenge of adapting the tools of psychoanalysis to the group setting. He believed the depth and availability of transference were accentuated in groups largely because the presence of other people stimulated feelings related to the families of origin of group members. He used dream interpretation, free association, and other analytic staples in innovative ways. For example, the "go-around" or "round robin" technique used in groups to include all members was originally born out of Wolf's efforts to use free association in the group session. He began a session by asking a member to report a dream and then would ask all other group participants to share their own spontaneous associations to the elements contained in the dream.

Significant theoretical and applied group work was taking place in Europe during the late 1930s and through the 1940s following the Second World War. Wilfred Bion, Henry Ezriel, S.H. Foulkes, and E. James Anthony all made seminal contributions to the understanding of the dynamics of groups. Bion focused on the "group as a whole" and delved into the unconscious life of the group using an object relations perspective based in the work of Melanie Klein. Concepts like the "basic assumption group" emanated from Bion's studies of group behavior and have led to a popular subgroup of modern group therapists that subscribe to Bion's precepts.

Ezriel agreed with many of Bion's ideas but emphasized the importance of focusing on the individuals in the group and their unique concerns, as well as concentrating on the whole group level.

To do so, Ezriel drew attention to interactions and the member-to-member transferences in the “here and now” of the group. In the same vein, Foulkes, a British psychoanalyst, coined the term the “group matrix” which reflected his view of the group as a nexus of communication shared by all members. The matrix took into account the cultural similarities and differences among members along with their styles of communication as reflected in their group interactions. Foulke’s work was significant in integrating historical material with their manifestations in the group process itself, thereby bridging past and present. Foulkes, along with E. James Anthony and others, were instrumental in forming the Group Analytic Society in England.

In the early 1950s, another British figure, Maxwell Jones, created the “therapeutic community” concept. Jones’s ideas were embraced and adapted by clinicians working with the chronic and persistent mentally ill and those dealing with problems related to substance abuse.

The history of group therapy incorporates the contributions of social scientists like Kurt Lewin. Lewin’s field theory helped solidify the view that the group was greater than the sum of its individual parts and that groups were social systems, a concept that had great utility for clinicians in both group and family therapies. The term “group dynamics” was introduced by Lewin.

From the 1950s through the present, enthusiasm for experimentation with the application of group models led to further awareness of the uses and potential liabilities of experiences in a group setting. Among the former was the work of Nathan Ackerman who made noteworthy contributions at the interface between group and family therapy when he began seeing whole families as a unique small group in the therapeutic setting.

Simultaneously, the National Training Laboratories conducted community groups in Bethel, Maine. These groups formed the prototype for the T-group or training group, encounter and sensitivity groups that flourished through the 1960s. Important lessons about the risk factors in groups that were improperly or irresponsibly led and the psychiatric sequelae that followed directly from passive or charismatic leadership, extended time formats as in marathon group therapy came from the overzealous use of this group format.

Exponents of new “schools” of group therapy began to emerge. Eric Berne developed transactional analysis, Fritz Perls designed groups based on Gestalt therapy theory, and Robert Goulding’s redecision group therapy are but a few of the burgeoning points of view leading to the broad array of theoretical points of view represented in current group therapy practice.

The appreciation of the small and large group setting as a vehicle for the communication of information about a condition or a set of goals or problem sets shared by its members is as evident in the work of these early practitioners as it is in present-day group work.

THEORETICAL ORIENTATIONS

Many individually derived forms of psychotherapy have their group counterparts. A representative sample of major theoretical orientations toward group therapy can be broadly subsumed under the psychodynamic, interpersonal, and cognitive-behavioral schools of thought.

The *psychoanalytic spectrum* has evolved to include more than classical Freudian psychoanalytic theory. Exponents of the British school of object relations based on the work of Melanie Klein, Ronald Fairbairn, D.W. Winnicott, Harry Guntrip, and other self-psychologists utilize the unique properties of the therapy group to observe and alter the behavioral and ideational manifestations that originate in the inner world of the group member. Group members function as or represent part objects that parallel those that exist in the mind of an individual and are projected onto and enacted within the interactions in the group. In this way, the group provides an unparalleled setting for behavioral information about how group members perceive and misperceive their relationships to fellow group members and to the group leader. Practitioners with this orientation use the group to highlight unresolved issues involving themes of emotional attachment, acceptance and rejection, managing intense affect, and tolerance of frustration.

Ego psychologists also fall within the penumbra of modern psychoanalytic practices in group therapy. Their theoretical base builds on earlier work by Heinz Hartmann, Ernst Kris, and David Rapaport who viewed Freud's structural theory of mental function into id, ego, and superego components as incomplete. They postulated that a conflict-free or adaptational portion of the ego exists as well. In clinical group work, ego psychologists rely on an appreciation of the mechanisms of defense that members use to counter their anxieties, the utilization of the peer and authority transferences in the group, and on the countertransference reactions of the group therapist.

Group therapy is an *interpersonal* form of treatment. Interpersonal theorists, most notably Irvin Yalom, advocate a group approach that centers on the interaction of group participants in the "here and now" transactions in the group session. Yalom also added universal existential dilemmas to a method that calls for active group leadership in creating an early, cohesive group framework.

Psychodrama is perhaps the most physically active form of group therapy. By employing structured dramatic activities in a group setting, the "director" or group leader guides the 'protagonist" (individual with an issue to be resolved) through a dramatization that is followed by feedback from other group members.

Cognitive and behavioral approaches rely primarily on learning theory and experiential elements to reach their goals. Structured, often manual-driven protocols are utilized in a time-limited group composed of members with a shared set of problems. Cognitive approaches are the most standardized of all group interventions and are, therefore the types of groups that are most easily measured in terms of group process and outcome.

Systems-oriented group approaches have found a bona fide place in group therapy as they have in family therapy. The assumption that the therapy group is an entity larger than the sum of its parts and that each member plays a role in the success or failure of the group enterprise is a central core to the systems approach.

Lesser-known but still actively practiced orientations include *Gestalt, redecision, and transactional analysis group therapies*. The first, based on Kurt Lewin's field theory and the work of Fritz Perls and Paul Goodman enjoyed great popularity during the sensitivity and encounter group movement in the late 1960s and 1970s. A central tenet of the Gestalt approach is the emphasis on group process over group content and efforts to mobilize strong affect that stands in the way of a person actualizing their full potential. The group focuses on emotional experience in the present moment and employs a variety of active, noninterpretive exercises designed to heighten self-awareness and to facilitate change.

Closely allied to the Gestalt method is redecision group therapy, a short-term model created by Robert and Mary Goulding as an offshoot of their work with Eric Berne and Fritz Perls. The format encompasses the creation of a clear behavioral contract at the outset. The group therapist aligns with the client's "authentic self" or that part of the person that is searching for new solutions to old impasses. The model eschews the use of the psychodynamic techniques, such as the use of interpretation and the analysis of transference and resistance relying instead on an array of redecision interventions that aim at identifying maladaptive, rigidified, unconscious defenses, and interrupting their reflexive nature in order to substitute more functional responses and problem-solving behaviors.

Transactional Analysis, commonly known as TA, is a component of redecision therapy but is also a stand-alone form of group therapy. Its founder, Eric Berne, replaced Freud's structural theory of the psyche from id, ego, and superego to what he termed "ego states" of Parent ("exteropsychic"), Adult ("neopsychic"), and Child ("archaeopsychic").

Berne codified a series of dysfunctional interpersonal patterns into what he identified as "games" that people engage in repetitively without success. The goal of the therapy group was to identify transactional states among group members and to change childhood "life scripts" into more adaptive adult patterns.

Table 32.21–1 illustrates some of the similarities and differences among representative group formats.

Table 32.21–1.
Comparison of Major Group Therapy Orientations

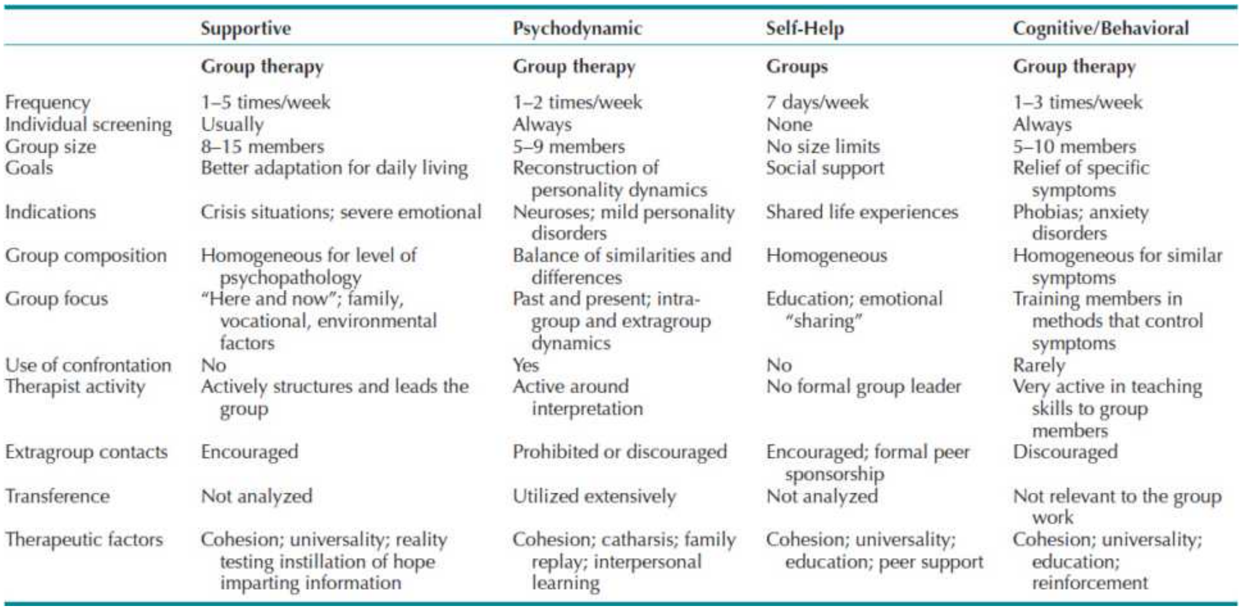

	Supportive	Psychodynamic	Self-Help	Cognitive/Behavioral
	Group therapy	Group therapy	Groups	Group therapy
Frequency	1–5 times/week	1–2 times/week	7 days/week	1–3 times/week
Individual screening	Usually	Always	None	Always
Group size	8–15 members	5–9 members	No size limits	5–10 members
Goals	Better adaptation for daily living	Reconstruction of personality dynamics	Social support	Relief of specific symptoms
Indications	Crisis situations; severe emotional	Neuroses; mild personality disorders	Shared life experiences	Phobias; anxiety disorders
Group composition	Homogeneous for level of psychopathology	Balance of similarities and differences	Homogeneous	Homogeneous for similar symptoms
Group focus	"Here and now"; family, vocational, environmental factors	Past and present; intra-group and extragroup dynamics	Education; emotional "sharing"	Training members in methods that control symptoms
Use of confrontation	No	Yes	No	Rarely
Therapist activity	Actively structures and leads the group	Active around interpretation	No formal group leader	Very active in teaching skills to group members
Extragroup contacts	Encouraged	Prohibited or discouraged	Encouraged; formal peer sponsorship	Discouraged
Transference	Not analyzed	Utilized extensively	Not analyzed	Not relevant to the group work
Therapeutic factors	Cohesion; universality; reality testing instillation of hope imparting information	Cohesion; catharsis; family replay; interpersonal learning	Cohesion; universality; education; peer support	Cohesion; universality; education; reinforcement

SELF–HELP GROUPS

Leaderless or self-help group experiences are usually homogeneously composed of participants with a common issue representing an additional use of the group setting as a vehicle for change. Although these groups are not strictly considered a form of psychotherapy they do share a number of parameters and can have therapeutic benefits for participants. Perhaps the best-known examples are the 12-step groups of alcoholics anonymous (A.A.), Al-Anon, adult children of alcoholics (A.C.O.A.), and other substance abuse–centered groups. In addition, many self-help groups are formed around health and illness, loss, or a shared life experience. Cancer support groups, groups for widows or widowers, parents without partners, and bipolar support groups are but a few well-known self-help groups.

THERAPEUTIC POTENTIAL OF GROUPS

Therapy groups contain special properties that can be utilized in the service of change. Some of these factors are unique to the group milieu, and others are amplifications of more generic components of all effective psychotherapies. For many years, researchers and practitioners have tried to identify those therapeutic or “curative factors” present in groups in order to determine how they can be used to maximum effect.

Many authors have offered systematized categorizations of the natural resources available in therapy groups over time. Popular examples of these efforts can be seen in the work of Yalom and Leczsz, MacKenzie, and Sadock. Yalom and Leczsz use a 13-point model that includes universality, altruism, instillation of hope, imparting information, corrective recapitulation of the primary family experience, development of socializing techniques, imitative behavior, cohesiveness, existential factors, catharsis, interpersonal learning (both input and output) and self-understanding.

MacKenzie divided the therapeutic factors into four groups: supportive factors, learning factors, self-revelation factors, and psychological work factors. The supportive factors include universality, acceptance, altruism, and hope as contrasted to learning factors such as modeling, vicarious learning guidance, and education. Self-disclosure and catharsis comprise the factors of self-revelation while interpersonal learning and insight round out the psychological work factors in this schema.

Sadock has delineated a more extensive description and definition of therapeutic factors listed in Table 32.21–2.

Depending upon the nature of the group and its stated goals, the group leader draws upon particular therapeutic factors in order to enhance the likelihood of group members deriving maximum benefit from group therapy. As an example, in the early stages of an adult psychotherapy group, the group leader emphasizes those factors that facilitate the creation of a rapid sense of group cohesion. In practice, this means relying on similarities among members, creating an emotionally safe group structure, using the support elements available to the group, avoiding premature confrontation by the therapist, and monitoring intermember hostility.

SELECTION OF GROUP MEMBERS

The extensive application of diverse group therapies over time has resulted in the liberalization of the inclusion and exclusion criteria for group membership. The selection process is a three-dimensional process consisting of evaluating individuals to see if they are appropriate for group therapy in general; member-to-group matching, does the suitable patient fit into the group for which they is being considered; and therapist-to-group matching, does the leader possess the clinical skills and emotional temperament necessary to understand and lead the group in the most effective way.

Traditionally, the group therapy literature has been more specific about exclusion criteria than inclusion criteria. Contemporary group therapy has been significantly modified from its historical roots in order to incorporate a broader range of difficult personality styles, severe psychopathologic states, and interpersonal deficits. Still, there remains a nucleus of patients for whom group therapy is either contraindicated or the timing of their entry into group treatment is premature.

Among these candidates for exclusion are those inpatients who are acutely psychotic and not stabilized on psychotropic medication, people who are acutely suicidal or for whom the high level of affect in the group might exacerbate their symptoms, and those whose primary language is not the language of the group. Debate continues about whether some conditions formerly considered for exclusion from group therapy can be treated in specialized groups in venues where the leader and the system are tolerant and familiar with the condition shared by members. A person with diagnosis of antisocial personality disorder was virtually certain to be excluded from group therapy. However, groups in prison settings and many substance-abuse groups are examples of the possibility of having homogeneous groups of members with a diagnosis heretofore thought to be unworkable in the group milieu.

Ideally, the person well suited for inclusion in group therapy is one who presents with an interpersonal problem and possesses sufficient motivation to work on the manifestations of these issues as they play themselves out in the ongoing arena of the group. For interactive, outpatient groups, members must have the capacity to empathize with others, possess sufficient ego strength to tolerate not being the exclusive focus of therapy, have some capacity for self-awareness or psychological-mindedness, be willing to self-disclose in the group meetings, and to benefit from the reality testing and interpersonal feedback supplied by the group therapist and group members.

In actual clinical practice, many people are referred for group therapy precisely for the very reason that they lack some or all of the aforementioned qualities. This should not be a de facto reason for exclusion from group membership. Pregroup alternatives like participation in a waiting-list group or a short-term orientation to group psychotherapy experience can give the therapist and the patient an expanded opportunity to see a prospective member in action. In so doing, it may become clear that the patient will either not be able to derive benefit from the group or will be a disruptive or undermining influence on the group, in which case, they will be deemed inappropriate for inclusion. On the other hand, pregroup participation can be invaluable in helping the group therapist design a treatment plan that is more patient-specific based on direct experience of the patient’s interpersonal interactions.

Objective measures such as the Group Therapy Questionnaire and the Group Selection Questionnaire, both sections of the comprehensive revised CORE (Clinical Outcomes in Routine Evaluation) Battery, yield important data derived from patient self-reports about variables relevant to the decision about who can/is likely to flourish in groups and who should be excluded from group therapies. Once the decision about inclusion has been made, the critical phase of pregroup preparation of the prospective group member follows next.

PREGROUP PREPARATION

All prospective group members must receive a pregroup orientation. Both the individual and group psychotherapy literature attest to the observation that the better prepared a patient is for psychotherapy, the better the outcome. In group therapy, the pregroup phase has been shown to be particularly important in reducing group drop-outs, diminishing patient anxiety about group entry, having members with more realistic expectations of group therapy, enhancing motivation for therapy, avoiding goal incompatibility among members, and building a strong therapeutic alliance.

Table 32.21–2.
Therapeutic Factors

Factor	Definition
Abreaction	A process by which repressed material, particularly a painful experience or conflict, is brought back to consciousness. During the process, the person not only recalls, but also relives the material, which is accompanied by the appropriate emotional response; insight usually results from the experience.
Acceptance	The feeling of being accepted by other members of the group; differences are tolerated, and there is an absence of censure.
Altruism	The act of one member's being of help to another; putting another person's need before one's own and learning that there is value in giving to others. The term was originated by Auguste Comte (1798–1857), and Freud believed it was a major factor in establishing group cohesion and community feeling.
Catharsis	The expression of ideas, thoughts, and suppressed material that is accompanied by an emotional response that produces a state of relief in the patient.
Cohesion	The sense that the group is working together toward a common goal; also referred to as a sense of we-ness; believed to be the most important factor related to positive therapeutic effects.
Consensual validation	Confirmation of reality by comparing one's own conceptualizations with those of other group members; interpersonal distortions are thereby corrected. The term was introduced by Harry Stack Sullivan. Trigant Burrow had

	used the phrase consensual observation to refer to the same phenomenon.
Contagion	The process in which the expression of emotion by one member stimulates the awareness of a similar emotion in another member.
Corrective familial experience	The group recreates the family of origin for some members who can work through conflicts psychologically through group interaction (e.g., sibling rivalry, anger originally toward parents).
Empathy	A capacity of a group member to put themself into the psychological frame of reference of another group member and, thereby, understand their thinking, feeling, or behavior.
Identification	An unconscious defense mechanism in which the person incorporates the characteristics and qualities of another person or object into their ego system.
Imitation	The conscious emulation or modeling of one's behavior after that of another (also called role modeling); also known as "spectator therapy," as one patient learns from another.
Insight	Conscious awareness and understanding of one's own psychodynamics and symptoms of maladaptive behavior. Most therapists distinguish two types: (1) intellectual insight—knowledge and awareness without any changes in maladaptive behavior; (2) emotional insight—awareness and understanding leading to positive changes in personality and behavior.
Inspiration	The process of imparting a sense of optimism to group members; the ability to recognize

	that one has the capacity to overcome problems; also known as installation of hope.
Interaction	The free and open exchange of ideas and feelings among group members; effective interaction is emotionally charged.
Interpretation	The process during which the group leader formulates the meaning or significance of a patient’s resistance, defenses, and symbols; the result is that the patient develops a cognitive framework within which to understand their behavior.
Learning	Patients acquire knowledge about new areas, such as social skills and sexual behavior; they receive advice, obtain guidance, and attempt to influence and are influenced by other group members.
Reality testing	Ability of the person to evaluate objectively the world outside the self; includes the capacity to perceive oneself and other group members accurately.
Transference	Projection of feelings, thoughts, and wishes onto the therapist, who has come to represent an object from the patient’s past. Such reactions, although perhaps appropriate for the condition prevailing in the patient’s earlier life, are inappropriate and anachronistic when applied to the therapist in the present. Patients in group therapy may also direct such feelings toward one another, a process called multiple transferences.
Universalization	The awareness of the patient that they are not alone in having problems; others share similar complaints or difficulties in learning; the patient is not unique.

Ventilation	The expression of suppressed feelings, ideas, or events to other group members; the sharing of personal secrets that ameliorates a sense of sin or guilt (also referred to as self-disclosure).

The use of a pregroup orientation session with each incoming group member affords the group therapist an ideal opportunity to accurately inform members about the parameters of the group they will be entering and to anticipate early resistances and behaviors that require sound clinical management in order to prevent avoidable drop-outs from the group. There are many ways to prepare new group members ranging from informal discussions with patients to structured pretraining programs, the distribution of written orientation manuals, employing videotapes of simulated group sessions, group orientations involving many new group members simultaneously, and others. Regardless of the format of the orientation to the group, the group leader is well advised to include an organized pregroup preparation component as a routine part of the pregroup phase of group therapy.

The use of a group therapy orientation checklist is one method of addressing the central elements of pregroup preparation of group therapy participants. Fourteen items comprise the pregroup orientation checklist. Depending upon the type of group and the personality of the incoming member, group leaders have the option to focus more in-depth on those areas on the list which may be most specific to the group and the patient in question and less on the topics which are less applicable to the situation.

The pregroup session not only addresses the “nuts and bolts” of the administrative and logistical aspects of groups, but it is simultaneously the start of the therapeutic process. When viewed from this perspective, the group therapist can accomplish the dual goals of preparing patients for group entry and get a preview of individual issues that will form the therapeutic agenda for a group member when they enter the group.

Group Therapy Check List

GROUP THERAPY ORIENTATION CHECKLIST

1. Goals and general purpose of the group.
2. Composition of the group.
3. The role of the group leader.
4. Observing and recording group meetings.
5. The arrangement of the therapy setting.
6. Time factors.
7. Loss and addition of group members.
8. Rules about attendance.
9. Fees and billing procedures.
10. Coordination of other therapies.
11. Extragroup socialization.
12. Modifications of the group contract.
13. Confidentiality.
14. Questions and answers

1. Goals and general purpose of the group
   The therapist clearly states the purpose and goals of the group. Goals that are vague or too ambitious are to be avoided in favor of goals that are patient and group-specific.
2. The composition of the group
   The therapist tries to convey a picture of the group the patient will be entering without disclosing information that violates the confidentiality of other group members. The goal is to provide the member with an understanding of who will be in the group, why specific members have been selected, and what the size of the group will be.
3. The role of the group leader
   A key concern for new members of any group revolves around what to expect from the group leader. Members who are used to individual psychotherapy may be anxious about sharing the therapist with other group members, puzzled about how help can come from others who themselves have significant emotional

problems, and wondering how much reliance they can place upon the group leader.

What the group leader actually tells members in the orientation dialogue is aimed at addressing common patient concerns about leadership activity versus passivity; how the relationship with the leader in the group relates to the kind of help they can expect from group therapy; how structured the group will be; will the group leader be available for individual sessions outside the group; and a host of recognizable pretherapy fears which affect new patients in groups.

4. Observing and recording group meetings
   There are settings and circumstances that may call for groups to be observed. In training programs, observation and recording of group sessions are primarily used for the benefit of teaching therapists a particular technique or method. There are also times when recording, especially videotaping of meetings, is used as part of the group therapy itself, as well as a vehicle for training, supervision, and quality control of therapeutic groups. Observation of any kind must be handled openly before the first session and with the patient's informed consent.
5. The arrangement of the therapy setting
   When working with more seriously disturbed, suspicious, paranoid patients or with group members who display excessive amounts of anticipatory anxiety, a simple statement about, and a short visit to, the room in which the group will be meeting helps allay many fears in the minds of these patients.

   During or after the visit to the group therapy room, most people will report some sense of relief, but some will describe the experience in terms that are clinically revealing. Initial reactions like, "The room feels claustrophobic to me." or "The chairs seem so close to each other; can people actually touch you in the group meeting?" are cases in point.
6. Time factors in the group
   The time frame of a group correlates directly with the purpose of the group. Time varies according to the group's goals. As a general rule, short-term groups have a fixed time limit and aim to

change maladaptive behaviors and/or to teach specific skill acquisition. Ongoing or open-ended groups are more ambitious and aspire to reach many of the same goals as their individual psychotherapeutic counterparts. Personality and attitudinal shifts as well as behavioral change are part of the long-term group model.

Considerations about the optimum time frame for each group session are also addressed in the pregroup orientation. The group parallel of the 45- or 50-minute individual session is the 90-minute group meeting. Examples of variations on the standard time frame designed to meet the unique needs and tolerance level of members include a brief cognitive/behavioral group that meets for 1-hour sessions over the course of 12 weeks and a 30- to 45-minute session group for psychiatric inpatients who may not be able to sit still or tolerate the level of affect that often accompanies a group of longer length.

7. Loss and addition of group members

Every group requires a sufficient number of members in order to reach its stated objectives. Some groups, usually those of a short-term nature, start and end with the same members. If, in the course of the time, members drop out of brief therapy groups they are not replaced. Many short-term therapy groups have a structured protocol for the conduct of each session and groups of this type do not allow for any changes in membership.

Long-term groups employ an open membership model that assumes that members will be leaving the group over the course of its existence. The two primary ways in which members leave a group are by terminating prematurely for psychological or practical reasons or through the realization of their treatment goals. In either case, gaps are left in the group.

Issues of timing are important when considering additions to the group. Long-term dynamically oriented groups do not seek immediate replacements for a departed member. Whatever the reason for a member leaving the group, it still is experienced as a loss by the group membership. The leader needs to allow time for the group to deal with the issues ushered in by the loss of a group member. Powerful emotional themes related to death, abandonment, separation, competition, family re-enactment, and loss are but a few of the feelings that accompany the exit of a member who has been in group for a long time. The dynamically inclined therapist in a long-term psychotherapy group can make extraordinarily productive use of the group members' unresolved conflicts related to loss, feelings emanating from inclusion and exclusion, and sibling relationships.

These potent emotions, most of which are historically based, get replayed in the group and are available for therapeutic exploration. Neglecting to address the issues related to loss and addition of group members invites avoidable problems in the ongoing conduct of the group.

8. Rules about attendance

All groups strive to establish regular attendance as a central group norm. It is a primary leadership responsibility to advocate and model regular group attendance. In so doing, the leader makes a position statement by their words and actions about the importance of the group and how vital each person is to the endeavor. Instructing members to call prior to a session if they are going to be absent or late is also discussed in this phase of the pregroup orientation.

Members in groups where absences and lateness are tolerated tend to devalue the group experience, lose confidence in the leader whom they regard as weak, passive, or inadequate, and feel like they are participating in a failure. All of these factors are, at the very least, demoralizing but they may actually be injurious to patients who come to the group with preexisting feelings of low self-esteem, a sense of personal failure, and a reasonable expectation that their lives will be enhanced, not made worse, by their joining a therapy group.

When working with very seriously ill psychiatric patients, rigid restrictions on attendance may be asking patients to comply with a task that is beyond their adaptive capacity. Therapists who work with schizophrenic outpatient groups can modify this portion of the group preparation module to allow for some expected irregularity in attendance. If the patient can make an attachment to the group and the system or hospital where the group meets and remain with it over time, these patients may have poor per-session attendance records but those who regard themselves as members of the group or members of the program do well over the longer lifespan of the group.

Higher functioning patients can be held to a higher standard of attendance. For purposes of the pregroup preparation of incoming members, the group leader decides in advance of the group what the policy on attendance will be and orient all members in a uniform way.

9. Fees and billing procedures

The mechanism by which therapy will be paid for is included in the orientation session. Issues of precertification for sessions, proper referral procedure, and compliance with the criteria set forth by third-party payers form the essentials of this process.

The setting of fees has significant clinical implications. Group leaders must decide whether it suits the goals of the group to have a standard fee for everyone in the group or, in cases where finances are a problem for a group member, should there be fee adjustments or a sliding fee scale. Whatever choice is made, it is important for the group leader to be clear about the fee policy for each incoming member and to anticipate the dynamic issues that graft themselves onto the issue of money.

It is common in psychodynamic groups where members pay different fees for group members to feel that the therapist prefers the people who are charged the lowest fees. This is a group replay of unresolved family issues of parental preference of one sibling over another displaced onto what appear to be practical matters involving money. The alert group leader can differentiate genuine issues about money from other psychodynamic and group process issues if he/she takes care to explain to each member the policy for group fees during the orientation meeting.

10. Coordination of other therapies

Group therapy is frequently part of a broader treatment plan for patients. The group leader has the responsibility for coordinating the treatment of each group member. In the pregroup orientation session, the therapist goes over other concurrent or potential treatment efforts and discusses how each will be handled.

This aspect of the pregroup member preparation encompasses several important considerations. If the group member is also in another form of psychotherapy, most usually individual, couples, or family therapy, the group leader has to get permission from the patient to contact the other therapist(s) involved in the patient’s care at any point the group leader deems necessary. In this way, not only can the group therapist’s database of information be expanded, but some potentially troublesome clinical issues, such as the use of “splitting” defenses by group members with borderline personality disorders, an attempt that aims to divide members of the treatment team.

The two other areas to be covered in this portion of the orientation concern medication management and the provisions for hospitalization of patients should that become necessary. When someone other than the group leader is prescribing psychotropic medication for a group member, the leader has to be knowledgeable about the drug regimen. At the same time, the clinician providing the medication can obtain valuable information referable to the patient’s response to drug therapy as demonstrated by his individual and interpersonal actions in the group.

When a group member has a history of prior psychiatric hospitalization or has a condition which becomes unstable at times, a plan for how, when, and by whom the patient will be hospitalized must be agreed upon. Patients who become overtly psychotic in group sessions may require admission to an inpatient facility directly from the group. In these instances, the understanding conveyed to the patient in the pregroup orientation is that it will be the group leader who will oversee the decision to hospitalize anyone from the group. The group therapist has to be sure that the patient understands the plan and is in accord with it in order to avoid group crisis situations for which no appropriate contingency plan has been made.

11. Extragroup socialization

    Conventional groups, especially those with a psychodynamic point of view, actively discourage or prohibit contact among members outside of group sessions. The reasons for adopting this group policy are designed to counteract the dangers which can arise when the group leader is not present. The formation of dysfunctional subgroups that exclude other group members, business or professional affiliations which spring from the group, and romantic and sexual relationships that develop between members of the same psychotherapy group are a few of the countertherapeutic results that may follow extramural socialization among group participants.

    In some selected groups, the leader may encourage contact among members for some reason that is consistent with the group goals. A “buddy system” in a group of patients who are lower functioning is an example of extragroup contact that is designed to be an extension of the therapeutic group process. Group members are paired together and come to the group and go home together. Since social isolation, loneliness, and coming to group regularly are problems for these patients, the creation of a group-generated plan that involves all members and excludes no one, is qualitatively quite different from spontaneous or planned acting out via outside contacts among members of higher-level groups. The latter is almost always an example of enactment of some issue outside the confines of the group where it would have to be openly addressed.

    In the same spirit, some substance abuse groups encourage the creation of a sobriety telephone network among members. The purpose of the networking is to give members who are at risk to relapse into drug use an alternative option whereby they can call a sober group member who will support their healthy efforts to resist using drugs.

    Whatever the group therapist decides is in the best interests of the group regarding extramural contact among group members, the pregroup orientation, not the established group, is the time and place to discuss this subject in some detail. Here again, the group leader may encounter some prospective group members who are unable to agree to adhere to this portion of the therapeutic contract. These members may have to be screened out for the group on the basis of their high-risk potential for undermining the group effort.

12. Modifications of the group contract

    This category concerns the rare but significant instance in which something in a patient's life could be at cross-purposes with their effective participation in the group. Many people in groups have scheduling constraints, however, it is a mistake for the leader to accept a member for the group when it is clear that they will have a regular pattern of absences. Individual requests to change the meeting time or date of the group based solely on any single member's personal limitations regarding scheduling are denied. The time, date, and meeting place are nonnegotiable items unless, for some reason, the therapist has to change the arrangements due to their professional commitments.

    The problems that occur when group meeting times are changed to accommodate one member, when episodic absence is accepted, and other deviations from the standard group structure are abandoned, can be of such magnitude as to threaten the viability of the entire group experience.

13. Confidentiality

    Effective psychotherapy of any kind cannot take place in the absence of the assurance of strict confidentiality. Incoming members are notified of the gravity of the confidentiality rule and are asked whether or not they feel they can adhere to it. Breaches of confidentiality are grounds for immediate dismissal of a member from the group. Any incoming member who cannot commit to strict confidentiality in the initial interview is excluded from joining the group.

    Confidentiality takes on added meaning in group therapy since members have the ability to take information about other people out of the group. In order to protect the confidentiality of all group members, the group leader must draw a clear distinction between gossip and the constructive use of information originating in the group to the member’s relationships outside of the group. This transfer of learning function encourages group members to apply insights and behaviors learned in the group to significant interpersonal relationships in their everyday lives without identifying the source of information in the group. Rules about confidentiality are restated in the initial group session and again at any point in therapy where potentially sensitive material is discussed.

14. Questions and answers

    The preparation session concludes with a portion that encourages patients to ask any questions related to the group they are joining or about any aspect of group therapy in general. It is best to leave time for questions near the end of the meeting after the other checklist items have been explained. Allowing for a question-and-answer period affords an opportunity for the group leader to get a sense of how the pregroup orientation meeting has gone. It also gives the group member and the leader time to go over any areas of confusion or misunderstanding concerning the content of the session.

    Since many people enter groups with preconceived notions concerning what actually transpires in group therapy, the pregroup interview encourages the expression of doubts, fears, and irrational beliefs about the upcoming group. The therapist can clarify myths and misconceptions and help ease many of the avoidable anxieties attendant to the process of being a new member of a therapy group.

    A sense of mistrust, fear of being attacked in the group, and concerns about humiliation, public exposure, and feeling inadequate often come up in the question–and-answer period. These can all be clinical clues to the leader about the patient’s particular difficulty with entry into the group.

    Another way of viewing the fears voiced in the interview is to see them as psychodynamic harbingers of the kinds of initial resistances that a person is likely to demonstrate in the interpersonal milieu of the actual group. The group therapist can use this opportunity to join with the patient, reassure them, and begin the process of establishing a therapeutic alliance aimed toward facilitating both group entry and retention in the group.

In a newly formed group, once all members have had the pregroup orientation, the stage is set for the first group meeting.

STAGES OF GROUP DEVELOPMENT

Well-designed therapeutic groups go through developmental processes over the course of their lifespan. The stages of group development are characterized by observable individual and group-as-a-whole phenomena. Numerous writers have identified and labeled group stages in keeping with the author's theoretical and clinical belief systems. The American Group Psychotherapy Association's Practice Guidelines for Group Psychotherapy integrated these major historical efforts at defining group stages with contemporary research findings resulting in a consensus about a five-stage paradigm of group development.

In this system, the stages that define the group's progress across its flight path are forming/preaffiliation, storming/power and control, norming/intimacy, performing/differentiation, and adjourning/separation. It is essential that the group leader be aware of the stage of the group's development and intervenes with the most stage-specific techniques available. Ill-timed or poorly sequenced interventions are not only ineffective, they can be damaging to individuals and to the whole group endeavor. Confronting a fragile member before sound group cohesion and support are established is an example of this error.

A fuller description of the elements of the five group stages will serve the twofold purpose of delineating the stages and the phenomenology of each and also shed light on the group leader's role at different points along the therapeutic spectrum. Two caveats to bear in mind are that stages of development are related to the composition, ego strength, and goals of the group and that they are not necessarily linear. Debate exists in the field regarding whether or not gains made in earlier group stages are retained even when a member of the group drifts back into an earlier phase posture. Similarly, the capacity and pacing of how far and how fast a group can progress in its development are limited by the level of psychopathology in the group members. With these two caveats in mind, the outpatient interactive psychotherapy group composed of midrange to high-functioning members is an opportune venue in which to appreciate the clinical significance of a developmental model of group therapy.

Preaffiliation

Initial concerns around inclusion and exclusion, acceptance by the leader and other members, the search for similarities, fears of interpersonal aggression, and competitive themes typify the first stage of a psychotherapy group. The group leader’s focus is on knitting the nucleus of the group into a cohesive whole.

Therapist interventions encourage identification of goal compatibility among members, modeling and eliciting supportive comments, promoting “here and now” interaction among members, discouraging excessive dependency on the leader, establishing clear group norms, and constructing a secure therapeutic frame to contain the group as it goes forward. A strong educational function also characterizes the therapist’s role in the first stage.

Power and Control

Apprehension about the emotional safety of the group, testing the group leader’s strengths and weaknesses, negotiating differences with others in the group, and experimentation with self-disclosure and self-assertiveness are hallmarks of the second or “storming” stage. Among the leadership tasks in this phase are mediating power struggles and conflicts in the group, demonstrating firm but fair limit setting, being on the alert for issues that might lead to drop-outs from the group, promoting collaboration among members, and keeping the group on task toward the attainment of its goals.

Performing/Differentiation

As the group progresses toward a more mature level, intragroup transferences become more apparent. Both vertical (member-to-leader) and horizontal (member-to-member) transferences are accessible. Levels of self-disclosure are higher and deeper interpersonal bonding takes place among members. There is a gradual transition from leader-directed group interplay to members sharing more responsibility for behaving in concert with the group norms. A cooperative group climate and greater reciprocity between members typify the intimacy stage.

The role of the group therapist during this stage can be less active than that in earlier stages however, the group therapist still plays a critical role in guiding the group process and preventing it from getting detoured into unproductive or counter-therapeutic channels. The leader aims to keep the group on track and reinforce those behaviors that promote the acquisition of insight, recognize constructive attempts at problem-solving, encourage the group to keep going forward in a collaborative spirit and to guard against the formation of subgroups.

Norming/Intimacy

Against the backdrop of the secure base of established group cohesion, group members in this stage are more willing to expose deeper feelings that might have activated real or imagined risks at an earlier point in time. This stage is an active one in which members are less afraid to show individual differences. The prevailing group climate is one of mutuality and acceptance. This atmosphere fosters greater understanding of relationships through the experiential learning that comes by virtue of being in stable relationships with others in the group over time.

Leadership focuses on letting the group be more self-sufficient and rewards spontaneous interaction among participants. Empathic responses are valued as are sincere attempts of members to be unselfishly related to others. The leader has to guard against abdicating the leadership role by becoming too passive during this phase. Since the groundwork has been laid for members to know how to express themselves in keeping with the group norms the leader may intervene less frequently but still bears overall responsibility for monitoring the group's therapeutic trajectory. Allowing and urging the group to be proactive and trusting the group are issues for the therapist during the phase of establishing intimacy.

Adjourning/Separation

The termination stage of therapy groups is noted for the triad of denial, regression, and recapitulation. In open-ended psychotherapy groups an event often sparks the realization that termination is the last stage of group therapy. Not uncommonly, this occurs when a member announces their plan to stop therapy because they feel that they have realized their treatment goals. Ambivalent feelings are evoked as members contemplate the impact of the loss of a valued member. Feelings of altruism, admiration, and having contributed to a fellow group member's success are accompanied by concurrent feelings of loss, competition, envy, anger, and concerns about each member's future. Members voice anxieties about the viability of the group when a known member leaves during this stage.

The group therapist has many leadership considerations related to an outgoing member and the remaining group. First, a clinical judgment needs to be made in order to distinguish between a genuine resolution of a member's therapeutic agenda versus a flight from the group. The latter is primarily fear-based and as such, calls for the patient to remain in the group to work through the unresolved issues that prompted the desire to terminate prematurely from the group.

Second, the termination phase is notorious for denial defenses, a re-emergence of earlier problems, and regressive behaviors all aimed at demonstrating to the leader that ending the group would be a damaging decision. The group therapist has to help members stay engaged with the plethora of strong emotions engendered by the prospect of losing a close relationship.

The leader conducts a review of the departing member's clinical course and the group's progress. In doing so the contributions made by fellow group members are acknowledged. This helps consolidate gains made for all in the group and helps buffer feelings of despair, loss, hopelessness, and sadness that regularly accompany the final stage of group work.

Finally, plans for life after the group are always part of a thorough termination stage. In instances where the whole group is ending, the experience of gaining closure on the group experience and saying good-bye to one another rounds out the tasks of the group leader as the group disbands. An appreciation of group development and the selection of carefully planned, stage-appropriate interventions by the therapist are essential ingredients of any successful therapy group.

LEADERSHIP ISSUES

The role of the group leader is the linchpin upon which the ultimate success or failure of the group enterprise rests. The leader creates, composes, and manages the group process from the selection and orientation of members through the working phases until the group reaches termination. The needs of the whole group and its individual members pose a series of potential challenges for the leader. The group therapist’s self-awareness about their strengths and limitations in carrying out these diverse leadership demands is essential for a positive group outcome.

Generic leadership functions and desirable personal and professional qualities in the person of the group leader have been a subject of study since the inception of group psychotherapy. Brabender, Fallon, and Smolar have eloquently outlined the core elements of successful group leaders. These qualities consist of an ability to look into oneself, the capacity for empathy, having a positive belief about the value of group therapy, and possessing good interpersonal skills. Additionally, executive leadership functions that encompass the ability to set and maintain the therapeutic frame of the group experience, a capacity to appreciate multiple simultaneous levels of interaction in the group session, helping members supply accurate meaning attribution to their behavior and maximizing the use of the role of authority figure in the group are other central leadership attributes.

In terms of specific leadership technique, the leader has to be proficient in tolerating as well as containing intense affect, managing their own countertransference engendered by the group, performing a teaching function for members in which the leader clarifies misperceptions and provides factual information, and uses interpretation in a timely manner. Since the leader is not immune to group pressure, care has to be exercised when using confrontational techniques and with respect to therapist self-disclosure in group meetings. As an example, when the group is in a state where anxiety in members is high and expresses itself in the form of confronting, testing, and challenging the leader, their competence, and/or the value of the group experience, the leader has to be able to separate personal feelings of rejection and inadequacy in favor of focusing on what and why this material is coming up at this point in time.

A helpful rule of thumb for judicious use of therapist self-disclosure is that the therapist should make sure that when self-disclosing information of a personal nature that it meets the criteria of being in their conscious awareness, is confined to the “here and now” of the group, and that the same goal could not be accomplished by other means. In this way, inappropriate or reflexive countertransference reactions by leaders that are harmful to the group are avoided. It is incumbent upon the leader to examine any temptation to self-disclose that has to do with the therapist’s personality as it articulates with the dynamics of the group.

Cotherapy

In many groups, the complexity of issues presented by the group suggests that sharing the leadership responsibility with a cotherapist would be advantageous. The workload on the leader who functions alone can be formidable in certain types of groups. As an example, cotherapy is often used in groups treating drug and alcohol addiction. There are many problems of living which involve the family of the substance abuser and require attention on the part of the group leader. Extragroup issues including legal problems, debt, housing, and employment are all factors which have an impact on conducting a group composed of substance abusers.

These issues coupled with the empirical observation that staff burnout is high among those who work exclusively with the problems of substance abuse, has led to a trend to use more than one leader in these groups. Team leadership is advantageous in allowing for the sharing of leadership tasks in the group. The presence of more than one leader increases the options for identification, interaction, and the exploration of intragroup relationships. As a result, the leader's therapeutic load is lightened, and the pace of the group can be accelerated by the presence of two leaders as objects of psychological investment by members.

When cotherapy is utilized, leaders have to be clear about their role and task definition in the group in order to function collaboratively. Coleaders who integrate their realistic differences concerning professional discipline, gender, and theoretical orientation, provide the group with an excellent role model for collaborative function in adult relationships. More opportunities for role modeling and for transference possibilities accrue when an opposite-sex team leads the group. This leadership combination stimulates family feelings in members by having a new therapeutic parental pair with whom members can interact.

A specialized form of cotherapy occurs in some substance abuse groups when a recovered alcoholic or drug addict is part of the therapy team. The recovered user occupies a position midway between that of a leader and a member. Their member-like qualities, including their first-hand experiences with life on drugs and alcohol, add a dimension to the group that cannot be supplied by the professional mental health staff.

In order to be considered for this “elder statesperson” role in the group, former substance abusers must have a documented history of at least 2 years of sobriety. They must also consent to urine screening during the course of the group therapy. The professional staff and recovered counselors meet immediately following each group session to discuss group events and the peer counselor's reactions to them. The postgroup discussion or “rehash” is designed to plan a course for future sessions and to ensure that the dynamics of the group do not pose a threat to the sobriety of the counselors themselves.

Variations in this specific coleadership model are also found in groups for bipolar disorder, medical illness, psychological trauma groups, and other conditions where a person has made a recovery and wants to share their experience with others trying to cope with an earlier stage of the problem.

CLINICAL APPLICATIONS

The versatility in the use of the group matrix can be seen in the following descriptions of contemporary innovative clinical applications of group therapies:

Combined Individual and Group Therapy

Many conditions require more than either group or individual therapies alone can supply. The simultaneous treatment of a patient in both one-to-one and group therapy is commonly referred to as "combined therapy." Therapy can be conducted by the same therapist or by two separate therapists. Different clinical issues are engendered by both types of combined therapy.

When the group and individual therapist are the same person, members only in group therapy often see the patient in combined therapy as being in a favored position by the therapist/group leader. This often induces group themes of competition, acceptance and rejection, and unresolved emotional residue from the families of origin of the group members.

In circumstances where one therapist is the dyadic therapist and another is the group therapist, professional colleagues must regularly communicate with each other. This ensures the coordination of a treatment plan for the patient and helps monitor whether the patient is doing an editing or splitting function by reserving some material only for individual therapy and not disclosing it in the group or vice versa.

A patient with a borderline or narcissistic personality disorder is at higher risk to play off one therapy against the other as a resistance to change. Astute use of combined therapy along with a dialectical group experience can identify this dysfunctional behavior and offer more adaptive alternatives. Specifically, the personality-disordered patient in combined therapy may use individual therapy to focus on intrapsychic issues and can apply the insights gained in therapy and experiment with new modes of behavior, thinking, and defenses in the safe and controlled interpersonal setting of the dialectic group.

At the same time, groups are concentrated on themes of appropriate interaction with peers and authority figures, the constructive expression of intense feelings, such as anger and intimacy, and the enhanced reality testing which groups provide. These attributes of the group help counter irrational beliefs and socially deviant behavior and offer more viable alternatives. Ideally, consistent communication between therapists staves off avoidable pitfalls and adds to the efficiency of the total therapeutic effort.

Bauer and McBride have created a life goals group program that combines psychoeducational, cognitive-behavioral, and self-management skills training for patients with bipolar conditions.

Groups for Mood Disorders

Affective disorders play havoc with the interpersonal relationships of those diagnosed with mood instability. The advances in diagnosis and psychopharmacologic options for treating depression have been dramatic. Group therapy has proven to be a widely accepted component of the treatment plan for patients with mood disorders. When one looks at the interpersonal sphere of the bipolar patient, their histories are riddled with interpersonal problems.

Conflicts at work, with domestic partners, original family members, and the law are common occurrences. As a corollary, depressed patients are invariably withdrawn from social contact and/or show signs of irritability as a prominent symptom.

A supportive and structured group psychotherapy milieu is ideal for addressing some of the issues that confront patients with serious mood disorders. Educating patients about their condition and the latest pharmacologic options available to them, using the group to focus on learning structured problem-solving techniques while simultaneously experiencing working in a cooperative way with fellow group members and staff, plus positive use of reframing techniques designed to counter stigmatization of mental illness are the central components of the life goals paradigm.

Groups for Personality Disorders

The model above is emblematic of many group approaches not only for affective problems but also for personality disorders where affective dysregulation is a core component of the therapeutic agenda. Exponents of group therapy know that group pressure is a powerful force for change that is unique to the group setting. When working with personality disorders, in group therapy, the therapist has more leverage available to induce change in maladaptive behavior, negative attitudinal stances, and interpersonal deficits.

The ingredients in the interpersonal group (IPG) psychotherapy protocol for patients with borderline personality disorder typify current thinking about the utility of groups to achieve significant change. The therapy group brings to light the borderline patient’s inner self-system with its feelings of emptiness, hunger for affect and gratification, and the inefficient interpersonal styles they have created in order to reach these ends. The IPG utilizes the dilution of the transference that sharing a group leader provides to turn down the emotional temperature of the group so that the atmosphere is more conducive to learning.

The interpersonal emphasis strives to help the patient form and experience more positive, gratifying identifications with others in the group. Using a theoretical premise that views problems with attachment as a common starting point for the interpersonal difficulties found in people with personality disorder, the group therapist uses a noninterpretive, empathic approach. In contrast to more traditional psychoanalytic approaches, the group leader avoids confrontation that invites strong affect, particularly in the early stages of the group, in favor of functioning as a “container” for the patient’s confusion, anger, and dysphoria. Once a therapeutic alliance is formed, the leader works with individuals in the group to counter their self-deprecating beliefs and replace them with a positive group climate of hopefulness and the possibilities for change.

These techniques rely heavily on the group leader’s subjective responses to the behavior by patients in the group and use caring feedback and suggestions of alternative courses of action as the tools to implement change. At the heart of this method is the assumption that in the IGP model, “change is more due to the experience of interactions in the group and less to the acquisition of insights about the genesis of internalized conflicts.”

Dialectical behavioral therapy (DBT) has also been adapted to work effectively in groups as well as in the individual setting. Although DBT was originally designed for working with patients diagnosed with borderline personality disorders it has been customized for use with patients with eating disorders, families with suicidal adolescents, and substance abusers. Still, more applications of IGP and DBT-based group therapies are expanding rapidly.

Groups for Anxiety Disorders

Anxiety is present in the vast majority of psychiatric conditions. For many people, it is the primary reason for which they seek professional help. While there are effective pharmacologic agents available for treating aspects of anxiety disorders, group therapies are particularly helpful as either adjunctive or primary modalities of treatment geared to managing the symptoms of anxiety and offering a vehicle for working on the interpersonal fallout from the habits formed by people who have unsuccessfully managed the stressors in their lives.

Groups, particularly those with a cognitive-behavioral orientation (CBGT), have been applied with cognitive behavioral group therapy (CBGT) to the full gamut of anxiety disorders from adjustment disorders through panic disorder up to and including posttraumatic stress disorder and phobic states.

Anxiety, fear, and panic are all characterized by avoidance behaviors. The cognitive behavior therapy (CBT) group aims to empower patients to engage with their anxieties by supplying them with a set of anxiety-reducing skills. Universal components of individual cognitive-behavioral therapies are adapted to the group setting. The advantages of a CBT group are: greater opportunities for role modeling, practicing new behaviors in vivo before testing them out in the world outside the group, getting peer feedback from others with similar anxieties, and the positive reinforcement that comes from many people, not just the therapist, which serves to reinforce new habits and thought patterns.

Examples of CBT group techniques include task assignments, including personal journals, and homework assignments, teaching progressive muscle relaxation, meditation, deep breathing, and other stress management mechanisms, and identifying the schemas or belief systems that underlie cognitive distortions. Maladaptive thought processes such as catastrophic thinking, polarized “black or white” thinking, over-personalization can be labeled in-group members, and cognitive restructuring is used to substitute more reality-based responses to stressors.

Cognitive behavioral therapy groups for anxiety-based problems are the fastest-growing form of group therapy today. They have gained clinical popularity for several reasons; the therapeutic interventions are often time-limited; the techniques are clearly described and can be standardized thereby facilitating research on the outcome by self-report, clinician-administered instruments, and/or cognitive and behavioral testing; precise treatment plans can be made into manuals that have a specific agenda and goals for each session, and the role of the group leader is very clearly defined.

Anxiety cuts a wide swath across generations, gender, ethnicity, race, culture, and physical illness. In CBGT commonality of anxiety problems is the central factor around which these groups coalesce. CBGT has the expectation that change will take place outside of the group as well as in formal group sessions. Out-of-session assignments, such as monitoring and recording intrusive thoughts that provoke anxiety are regular staples in cognitively based groups. CBT groups are prototypical of most brief or short-term group psychotherapy models insofar as they have a fixed time frame, a strong educational component, a clear focus for each group session, and very active, directive leadership that strives to keep the group on task and terminates the group at a predetermined point in time.

The awareness of traumatic conditions experienced by men and women who are returning soldiers has formed a large basis for the application of effective trauma groups.

Groups for Schizophrenia and Other Psychotic Disorders

Group therapy for patients with severe psychiatric disorders has truly come of age. Historically, inaccurate assumptions were made about the viability of therapy groups for psychotic patients. Most of these doubts came from psychoanalytic theory which proposed using techniques that were beyond the capabilities of schizophrenic patients and, in some instances, actually harmful to fragile patients.

Groups address many of the central issues involved in managing patients with severe psychiatric impairment. From crisis control to diagnostic evaluation and as a primary form of psychosocial therapy, the group milieu offers enormous flexibility by creating a foundation for a comprehensive treatment program for the chronic patient. Inpatient, transitional, and outpatient groups all constitute the group resources available to the clinician. The choice of group options is primarily related to the patient's diagnosis and to the stage of severity of the primary psychiatric problem.

Using this paradigm, groups for chronically ill psychiatric patients can be conceptualized as acute, transitional, and maintenance or aftercare in nature. Patients with acute first episodes of a condition or with an exacerbation of a longstanding problem are primarily hospital-based. Inpatient group therapies encompass quiet activity groups, art, poetry, and dance programs; occupational therapy; discussion groups, and community meetings on the inpatient service. Goals that are overly ambitious should be abandoned to suit the relatively brief time period of hospitalization. Simply stated, one of the highest priorities for inpatient groups is to get patients to be comfortable with group therapy. Having a successful group experience in the hospital increases the probability that patients will continue in aftercare groups and not get lost in the transfer from inpatient to outpatient status.

Transition groups refer to the transition from hospital-based treatment to outpatient and community-based programs. The start of this process begins with discharge planning in the hospital. Discharge planning groups, social skills training groups, family groups, job-focused behavioral groups, and medication groups are all representative of groups for people in the process of leaving the hospital.

Groups in an outpatient setting are designed to prevent relapse, which would require rehospitalization, and to help participants develop rewarding social, familial, and vocational lives. In general, these groups aim to reduce social isolation, monitor the effect of medication as measured by the patient's level of interpersonal function, aid in reality testing, provide social support, reduce the stigma associated with being mentally ill, and transmit accurate, up-to-date information about the primary condition. Leadership is very active in setting a therapeutic structure around which the group can consolidate. Psychotropic medication is invariably an integral part of group life whether or not the group is a medication group per se.

Most groups for chronic and persistently mentally ill patients are based on principles of supportive psychotherapy. Exploration of unconscious material, delusional themes, and affect related to anger and hostility are to be avoided. The emphasis is on patient strengths and the accumulation of small, incremental successes started in the group and extended to life outside the group. Problem-solving, advice-giving, and teaching new coping strategies, are all hallmarks of outpatient groups for patients with chronic and debilitating mental illness. The emphasis on improving cognitive function is also a central group goal.

Table 32.21–3 illustrates the similarities and differences between groups which have psychotic members and those which do not.

Table 32.21–3.

Group Therapy With Psychotic and Nonpsychotic Patients

	Pychosis	**Nonpsychotic**
Composition	Homogeneous for level of function	Homogeneous: Brief group Heterogeneous: Long-term group therapy
Functional capacity	Poor ego strengths Interpersonal deficits	Able to learn from group feedback Capacity for empathy and altruism
Goals	Crisis intervention Stabilization Improved communication Social skills development	Improved self-perception Improved interpersonal function Resolution of personality issues
Educational elements	Central to the group model	Elective: Yes: Brief groups No: Psychodynamic group therapy
Therapeutic factors	Reality testing Identification Social Support Bolster defenses	Cohesion Challenging defenses Use of transference and unconscious material
Cohesion	Desirable but hard to attain	Mainstray of the group model; easier to attain quickly
Leader role	Active around structuring the group Avoids aggressive and irrational	Spontaneous group interaction Leader is catalyst for group process

	themes	
Confrontation	Strongly avoided	Actively encouraged
Interpretation	Only on occasion and limited to the "here and now"	Regularly used for in-group transactions and transferential issues
Medication	Almost always part of the group	May or may not be a group issue
Group focus	Present and future	"Here and now" plus historical focus
Socialization	Strongly encouraged	Generally discouraged
Time span	Brief and long-term models Open-ended "maintenance" group therapy	Brief: Symptom focused Long-term: Personality and behavior change
Group participation	Involuntary and voluntary	Voluntary

A group focus on medication compliance, education, socialization, vocation, and psychological integration is the constellation of benefits chronically ill patients can derive from ongoing participation as members of groups.

RESEARCH ISSUES

The field of group psychotherapy has departed from an antiquated, predominantly anecdotal literature to a firm footing in scientific inquiry. Data from well-designed and well-controlled studies attest to value and utility of groups as a primary form of psychiatric treatment. Significant findings related to group cohesion show that pregroup selection and orientation; higher levels of early group structure; balancing individual and intragroup factors; active group leadership, facilitating emotional expression, relationship building, and verbal feedback all result in high levels of group cohesion which, in turn, favors positive outcome. The role of a strong therapeutic alliance is also correlated with positive outcomes in both brief and long-term groups.

A greater degree of both specificity of focus and research study design characterizes the cutting-edge efforts to support the clinical enthusiasm for group psychotherapy. Additionally, research instruments have been created to decode the mechanisms by which groups work. Empirical observation and research focus of group work now studies aspects of group process, leadership variables, and therapeutic factors that form the basis for the change resulting from group participation. The CORE Battery developed by the American Group Psychotherapy Association is the best-known example of this effort.

FUTURE DIRECTIONS

There is a clear trend toward eclectic group therapies characterized by technical integration. Greater emphasis is being placed on those group interventions that are specific, documentable, and capable of replication when employed by other practitioners. Many group therapists draw upon a psychodynamic understanding of each group member but their modes of intervention have become liberalized and less restricted to traditional analytically based practice. Cognitive elements, educational components, structured group exercises, and the assignment of tasks based on the expectation of increased patient responsibility for change are but a few of the additional eclectic modes of intervention added to the therapeutic armamentarium.

Training modules and manuals have been created for practitioners and the public covering several critical aspects of group therapy. How to train future supervisors of group therapy, explicating the issues involved in the ethical practice of group therapy and guidelines for teaching a basic course in group psychotherapy representing three examples of a trend to educate practitioners in order to assure quality control and to strive to continue to provide first-rate group services to patients.

Beginning in March of 2020 it became clear that our lives had changed dramatically due to the onset of the Coronavirus epidemic. Most of the world went into shock, panic, despair, and retreat. In-person treatment became a deadly choice. For former patients and those seeking help for the first time, a very good option has been through the use of computers and cell phones where a group of people can be heard and seen simultaneously. This mediates feelings of isolation and offers members support for new issues as well as those predating the pandemic. Tele-mental health is a new arena for group work that will undoubtedly outlast the current disaster and be useful in an ongoing way for homebound people and communities far from health services for example.

Times have changed and group psychotherapy has kept pace with them. As new developments occur, tandem reflection in the addition and modification of group technique and sensitive therapeutic styles ensure the progress of group therapy as a welcome transition to modern psychiatric clinical work.

FURTHER READINGS

Alonso A, Swiller HI, eds. *Group Therapy in Clinical Practice*. American Psychiatric Press; 1993.

*American Group Psychotherapy Association Science to Service Task Force. *Practice Guidelines for Group Psychotherapy*. American Group Psychotherapy Association; 2007. Springer Publishing Co.; 2003.

Bernard HS, Spitz HI. *Training in Group Psychotherapy Supervision*. American Group Psychotherapy Association; 2006.

Brabender VA, Fallon AE, Smolar AI. *Essentials of Group Therapy*. John Wiley & Sons, Inc.; 2004.

Brook DW, Spitz HI, eds. *The Group Therapy of Substance Abuse.* The Haworth Press; 2002.

*Buchele BJC, Spitz HI, eds. *Group Interventions for Treatment of Psychological Trauma*. American Group Psychotherapy Association; 2004.

*Burlingame GM, Strauss B, Joyce AM, et al. *American Group Psychotherapy Association's CORE Battery—Revised*. American Group Psychotherapy Association; 2006.

Feld B, ed. *Special Section:Couple Therapy, Group:The Journal of the Eastern Group Psychotherapy Society*. Vol. 31:3. Catskill; 2007.

Flores PJ. *Group Psychotherapy with Addicted Populations: An Integration of Twelve-Step and Psychodynamic Theory*. 3rd ed. The Haworth Press; 2007.

Fuhriman A, Burlingame GM. *Handbook of Group Psychotherapy: An Empirical and Clinical Synthesis*. John Wiley & Sons; 1994.

Gantt SP, Cox P. Introduction to the special issue: neurobiology and building interpersonal systems: groups, couples, and beyond. *Int J Group Psychother*. 2010;60(4):455–460.

Grundy D. The Short-Term Inpatient Group. *Int J Group Psychother*. 2014;62:2.

Heimberg RG. *Cognitive-Behavioral Group Psychotherapy for Social Phobia*. The Guilford Press; 2002.

*Kaklauskas FJ, Greene LS, *Core Principles of Group Psychotherapy: An Integrated Theory, Research, and Training Manual*. Routledge/Taylor & Francis; 2019.

Kanas N. *Integrative Group Therapy for Psychosis*. Routledge; 2021.

*Kaplan HI, Sadock BJ, eds. *Comprehensive Group Psychotherapy*. 3rd ed. Williams & Wilkins; 1993.

Klein RH, Schermer VL. Toward understanding and treating violence in America: some contributions from group dynamic and group therapy perspectives: introduction to part I. *Int J Group Psychother*. 2015;65:1–28.

MacKenzie KR. *Time Managed Group Psychotherapy*. American Psychiatric Press, Inc.; 1997.

MacNair-Semands R. *Ethics in Group Psychotherapy*. American Group Psychotherapy Association; 2005.

Marziali E, Munroe-Blum H. *Interpersonal Group Psychotherapy for Borderline Personality Disorder*. Basic Books; 1994.

Motherwell L, Shay, JJ. *Complex Dilemmas in Group Therapy*. Brunner-Routledge; 2004.

Paquin JD, Coleman MN. Introduction to the special issue on women and group work. *Int J Group Psychother*. 2021;71:1–12.

Rutan JS, Stone WN, Shay JJ, eds. *Psychodynamic Group Psychotherapy*. 4th ed. Guilford; 2007.

Schein LA, Bernard HS, Spitz HI, Muskin PR. *Psychosocial Treatment for Medical Conditions: Principles and Techniques*. Brunner-Routledge; 2003.

*Spitz HI, Spitz ST. *A Pragmatic Approach to Group Psychotherapy*. Brunner-Mazel/Taylor & Francis Group; 1999.

Spitz HI. *Group Psychotherapy and Managed Mental Health Care: A Clinical Guide for Providers*. Brunner/Mazel; 1996.

Weber RL. *Principles of Group Psychotherapy*. American Group Psychotherapy Association; 2003.

White JR, Freeman As. *Cognitive-behavioral Group Therapy for Specific Problems and Populations*. American Psychological Association; 2000.

Wolf A, Schwartz EK. *Psychoanalysis in Groups*. Grune and Stratton; 1962.

Yalom ID. *Inpatient Group Psychotherapy*. Basic Books; 1983.

*Yalom ID, Leszcz M. *The Theory and Practice of Group Psychotherapy*. 6th ed. Basic Books; 2005.

Zorzella KPM, Muller RT, Classen CC. Trauma group therapy: the role of attachment and therapeutic alliance. *Int J Group Psychother*. 2014;64:24–47.

32.22 Psychodrama, Sociodrama, Sociometry, and Sociatry

Edward J. Schreiber, T.E.P., and Scott Giacomucci, D.S.W., L.C.S.W., B.C.D., C.G.P., F.A.A.E.T.S., T.E.P.

THE PSYCHODRAMATIC SYSTEM

> Man does not live alone and does not get sick by himself. His problems develop in groups …. the mental and physical equilibrium of an individual depends to a considerable degree upon the dynamic interplay of these various individual and social forces.—J.L. Moreno

Psychiatrist Jacob Levy Moreno (1889–1974) hypothesized that human beings learn primarily through interaction with relationship systems, which he called social atoms. Moreno believed that an encounter of the self with creative processes, which he defined as spontaneity-creativity, could deepen and encourage growth for individuals and groups. Psychodrama and related action methods of sociodrama and sociometry facilitate creativity, and the discovery of strengths, based upon J.L. Moreno's understanding of human growth, from infancy to the end of life, with spontaneity-creativity as a central factor.

Psychodrama is used as therapy in various forms of education and in training therapists, doctors, lawyers, teachers, social workers, political leaders, public speakers, businesses and organizations, and more. Psychodrama can often be used in conjunction with other forms of therapy. Though at the same time, psychodrama is a comprehensive system and can stand on its own with its own philosophy, theories, and interventions.

In a psychodrama, one person called the protagonist, works with the director and a group or just individually with a director. In a group, participants may be asked to assume a role in the protagonist's exploration of an aspect of life. Those asked are not obligated to enact any role; it is a matter of choice. The director's role is complex and includes the facilitation of the protagonist, the role players, the group, the multiphased psychodrama process and techniques, and the integration and sharing phase with the protagonist and the group. Psychodrama follows a sequence of phases, which includes the *warm-up, action, and sharing*. In a training program, there is often an additional *processing* phase for students to become skilled in their application of the method.

Sociodrama is related to psychodrama and is a process for a group, organization, or community to explore relevant concerns and issues. In sociodrama, the whole group is on the stage in a role-playing theme, idea, and setting-based encounter in which participants have taken a role to play in the sociodrama. While psychodrama enacts an individual's specific story and roles, a sociodrama enacts a group's collective story using generalized roles. Sociodrama is used in educational settings, in professional development programs for a wide range of professions, and in clinical settings. Sociodrama can be employed to explore social issues, common themes, and social, organizational, and cultural problems.

Sociometry, related to psychodrama and sociodrama, although a method of its own, is used to understand and evaluate and to support, enhance, change, and develop relationship systems and subgroups in organizations. Sociometry is a tested process to address and improve communication, cohesion, functioning, and growth. Sociometry is rich with techniques to measure and examine the often unseen but felt networks which are very much alive and impact a group.

Sociatry (healing society) utilizes psychodrama, sociodrama, and sociometry to study the forces J.L. Moreno described as impacting all of humankind. J.L. Moreno understood there to be *an actual, underlying, dynamic, central structure impacting all formal and* *informal groups*. This underlying structure may be seen in groups looking through a sociometric lens, known within sociatry as the social microscope. The social microscope is an educational tool used to show the forces J.L. Moreno named *the sociodynamic effect* and *the social and organic unity of humankind*. Sociatry deepens our relationship to the creative process, our common humanity, and our interrelatedness.

HISTORY

J.L. Moreno was born in Bucharest, Romania, and received his M.D. in Vienna in 1917. In Vienna, he began The House of Encounter and formulated ideas about the nature of groups, healing, and society. Moreno describes the first psychodrama to have taken place with a large theater audience on April 1, 1921; it was focused on the task of finding a new leader after the collapse of the Austria-Hungarian Empire. In 1925 he immigrated to the United States and beginning in 1929 inaugurated Impromptu Theater in New York City. In the early 1930s, he conducted sociometric research at Sing Sing Prison, the Hudson School for Girls, and public schools in N.Y. While presenting at the 1932 American Psychiatric Association conference in Philadelphia, Moreno coined the terms “group therapy” and “group psychotherapy.” In 1936 Moreno opened the Beacon Hill Sanatorium and the psychodrama theater in Beacon, New York. In 1942, Moreno founds the first group therapy organization, the American Society of Group Psychotherapy and Psychodrama (ASGPP). Later in 1950, he actually petitioned the American Psychiatric Association to start a Group Psychotherapy section and ran for president of the organization–both attempts were unsuccessful. In 1954 he organized the First International Congress of Group Psychotherapy in Toronto, Canada and in 1973 he cofounded the International Association of Group Psychotherapy (IAGP). He is credited with publishing many books, journals, and articles, along with the organizations he began, and his worldwide teaching and training.

J.L. Moreno was a doctor, scientist, psychiatrist, and mystic. To note his mysticism causes alarm for some but is important in grasping the fullness of his ideas. It is from J.L. Moreno's integration of mysticism and science that this work has come. As a young man, J.L. Moreno was fascinated with the cosmos, fixated on the realm of the unknown. He dedicated his life and work to understanding and mapping the connections we make and have in life. J.L. Moreno created dozens of techniques, a complex method to experience and integrate this creative force–a kind of intelligence as he found it. He taught and trained thousands of practitioners in this method; how the force of creativity can be found; how to touch it, to the benefit of oneself and others. He was a social scientist as well and placed enormous effort into the investigation and cure of social problems. The goal of J.L. Moreno was ultimately to realize an organic unity within humankind and promote society's movement toward unity and healing.

Zerka Toeman Moreno began her study and work with J.L. Moreno in 1941 after having emigrated from England to the United States. Her path to finding J.L. Moreno and the Beacon, N.Y. Hospital is a remarkable story written in her memoirs *To Dream Again.* Zerka joined J.L. Moreno as his assistant, then wife and primary collaborator. J.L. Moreno's work might be just a footnote in history, were it not for the dedication, insight, and skills of Zerka Moreno. During the long period of their developing the method, Zerka emerged as codeveloper and went on to teach thousands of students throughout the world, and helped heal many wounded souls.

Zerka Moreno coauthored, edited, and helped translate most of J.L. Moreno's books and articles, and served as his primary teacher of the method. After his death in 1974, Zerka Moreno continued to teach and develop psychodrama until her death in 2016 at the age of 99 years old. She authored many articles, chapters, and books, with *To Dream Again–A Memoir* as her last, with this chapter's primary author as editor.

J.L. Moreno and Zerka Moreno were world minded. The issues of access to resources and our attention to those left out were of great importance to their work. J.L. Moreno created a therapy of experience, an experience of qualities of being human, to bring understanding and regeneration to individuals and groups.

The method created by J.L. and Zerka Moreno helps many problems. Work is being done throughout the world to bring the method and its practices to assist with the widespread issues we face as a human family. There are many sources of information regarding social application including IAGP, ASGPP, and dozens of other psychodrama associations around the world.

METHODS, TECHNIQUES, CLINICAL ISSUES

The method includes psychodrama, sociodrama, sociometry, group psychotherapy, spontaneity testing and training, role testing and training, the therapeutic stage, the empty chair, role-playing, role models, role relationships, and the exploration of subgroups and systems in organizations, groups, and communities. The method also involves sociometry, surplus reality, tele, the cultural conserve, the social atom, the protagonist, the double, auxiliary egos, the canon of creativity, the director, role theory, role reversal, and a developmental theory based on spontaneity-creativity. Furthermore, there are aspects of Moreno method that are not as widely known, such as the cultural atom, the sociodynamic effect, socio-genetic law, sociostasis, socio-gravitation, the autonomous healing center, and sociatry.

Psychodrama

Psychodrama takes place on a stage. A stage can be like the three-level-stage J.L. Moreno designed (relocated from the original hospital in Beacon, N.Y. to Boughton Place in Highland, N.Y.) or any space designated as a stage separated from the group. The stage is the location in which the psychodrama and sociodrama takes place and in most psychodrama groups today, the stage is simply designated as the center of the room where the group is held. There is flexibility in the stage, the director and protagonist may adjust and change the stage when needed.

The protagonist works with the director in a group, or just individually with a director. In a group, participants may be asked to assume a role in the protagonist's exploration of an aspect of their life. The director works with the group and group process during psychodrama, although the focus of the group is on the psychodrama and the protagonist. Psychodrama requires a combination of techniques and experiential methods to be complete. In psychodrama, a protagonist may examine aspects of life, from the past, the present, or the future. The protagonist can explore a relationship, a moment in the past or future, an interaction, a scene from life, a dream, parts of self, or internal roles.

According to Zerka Moreno, the effective practice of psychodrama is focused on the integration after catharsis (catharsis of integration) rather than solely accomplishing catharsis (catharsis of abreaction). Lasting change comes from the catharsis of integration–integration of what happened, what was needed but did not happen, what is needed now or in the future, and how this is to be healed or processed. Integration may take place during the psychodrama or days later and may include a new defining and developing of skills, relationships, and roles.

An initial task of the director is to discern the general frame and need the protagonist brings, and to determine a place (scene, setting, situation) to begin the psychodrama. There is a warm-up phase in which the protagonist emerges, and also a warm-up between the director and the protagonist. The director may suggest an initial scene, often elicited during an interview with the protagonist at the beginning of the psychodrama. A psychodrama may commence with scene setting in which the protagonist creates in action and imagination a setting, a scene, or simply wants the empty stage. There is no requirement to begin with the past. Zerka Moreno suggests that what matters is not so much where the psychodrama begins, *but that it begins.* Psychodrama develops a protagonist's capacities and strengths in their work to create, and recreate, in an authentic and adequate way.

The double technique in psychodrama is when a member of the group (who is guided by the director) offers statements and actions of support, or stabilization, clarification, amplification to the protagonist during the psychodrama. The double is role-played by someone from the group chosen by the protagonist or the director. The double can add depth, expand expression, extend or clarify statements or feelings, and help the protagonist to gain perspective, to integrate, hold and contain the experiential-based learning.

The mirror technique is when the protagonist is outside of a scene, as though watching-seeing a scene from their life presented by others. A group member may take the place of the protagonist in the scene, and members may be asked to assume other roles. The protagonist gains an observation, perspective, and emotional distance from the events.

Role reversal is a primary technique in psychodrama. It is a directed experience, and as with other techniques, it can be used as therapy, in education, and in training programs. In a role reversal, the protagonist presents what needs to be presented to another, imagined, in a setting or scene, on the stage in the psychodrama. The protagonist offers to “that other” what needs to be said. With direction and when appropriate, the protagonist “takes the role” of that other. The protagonist moves to be in the role of that other and eventually responds from the role of the other. Role reversal can continue for a series of directed exchanges, but the protagonist finishes in their own space, place, and role as themself, with a period of sharing after the psychodrama has been finished. Role reversal may take place with an imagined person, an object, an idea, a place in time, etc. A protagonist may role reverse with lessons in life or, as J.L. Moreno described, with creation.

Role reversal expands creativity, though there are contradictions. If there is difficulty with role reversal, this may indicate that other techniques, such as the double, soliloquy, or the mirror, may be more useful and appropriate. One needs to have strong and secure boundaries to return to. A contraindication and caution are for any role reversal with trauma, perpetrators, or violence of any kind–this will be addressed more fully in a later section. Training is required to facilitate role reversal or any aspect of psychodrama and related methods.

Sociodrama

In psychodrama, one person is the protagonist; in sociodrama, the whole group participates in action learning. In sociodrama, the group, organization, or community explores concerns, problems, situations, or relevant issues. Sociodrama is used in education, professional development, and in clinical settings. The group is role-playing within a theme, exploring issues in an imagined setting.

Sociodrama has been used to address global warming; and with exchanges between people of differing cultures and religions; with survivors of the Holocaust on both sides; in working with issues and differences between generations; between police and gangs, etc. Sociodrama develops a group in the capacity to double and reverse roles with others, see the world from the point of another, and find new meaning and learning.

Sociodrama is used for training doctors to better understand the experience of being a patient and parents to better know the experience of being a parent. Sociodrama increases communication and cohesion and *tele* (two-way empathy). Sociodrama includes techniques, directorial and group processes, how a sociodrama is directed, and how the concerns, themes, roles, and settings are chosen.

Psychodrama and sociodrama are experientially woven applications of the social atom, encounter, the canon of creativity, role theory, concretization, sociometry, the catharsis of integration, group psychotherapy, and other techniques, which will be explained further in this chapter. It is our relationships in life that can be concretized and explored through psychodrama, sociodrama, and sociometry.

Sociometry and Tele

Moreno hypothesized that human society has an atomic structure, which corresponds to the atomic structure of matter, and that its existence can be brought to an empirical test by social microscopy, related to sociometry. Sociometry is a method for the examination, exploration, and alteration of the connections and dynamics taking place within a group, organization, and community. The connections are often felt but often not fully recognized. Sociometry makes the connections visible, approachable to growth and change, to increase communication, creativity, cohesion, and resolution. Moreno's sociometric system was celebrated and adopted by sociologists and many consider his early sociograms and psychosocial network research to be the foundation of modern social network theories and participatory action research.

Sociometry is a method with many instruments for the measurement of connection, the web of tele "meetings" in operation. Telic connections can impact the functioning of a group, organization, and individuals within it, the dynamics of cohesion, inclusion, exclusion, subgroups, etc. Telic constellations form between two or more people, often unrealized in their capacity to harness and unleash a kind of interfeeling that impacts a group's embrace of its members, its creativity, its functioning, cohesion, and learning. Tele is a phenomenon below transference, not always seen, with multiple types of often interlocking relationships and groupings in an organization and community. While transference is a one-way distortion of reality, tele is an accurate two-way connection.

Telic constellations are made visible with sociometry. Sociometric structures and configurations can regenerate to increase communication, cohesion, creativity, and functioning. Sociometry is used in clinical settings, training programs, organizations, educational settings, and in communities.

Zerka Moreno explains tele:

> *Empathy and transference are part of a more elementary and more inclusive process of tele. The process of reciprocation does not enter into the meaning of empathy, and transference is considered the factor responsible for the dissociation and disintegration of social groups. Tele is the factor responsible for the increased mutuality of choices surpassing chance possibility and responsible for the increased rate of interaction among members of a group.*

Telic structures are central to one's personal sociometry, noticed in the feeling between oneself and others. For personal and professional development, sociometry evaluates and measures this.

With a group, sociometry may knit together positive telic connection and communication. Sociometry can open blocked communication by revealing to the group its telic structures. When the structures are more fully understood, they can be changed and developed in a productive way. Telic structures often operate below the formal organization as a matrix of cohesion or not and function at the moment in time. Sociometry shows a group, the living social atom, the dynamic telic processes commonly unseen but often felt, and the intelligence of life called spontaneity-creativity in action.

THEORIES THAT INFORM THE METHOD

The Social Atom

The social atom is used to depict the relationships of one's life and the nature of each relationship. The social atom can be an instrument used in a written format and as a psychodrama enactment that depicts a social atom by bringing it alive with others taking roles. The social atom can include one's relationship with individuals, groups, animals, communities, or other entities. The social atom may depict the past, the present, and the needed or hoped-for relationships. It is a visual representation of positive, negative, neutral, ambivalent, mutual, and nonmutual relationships.

The social atom depicts those with whom we feel emotionally impacted, connected, or related. The social atom is a relationship map; it brings awareness to relationships. The social atom looks like a bulls-eye; the nucleus represents oneself. In the first circle beyond oneself, one notes those with whom there is such a significant impact. One draws a line connecting self to that other, and there are specific line notations to show the nature of the relationship. Beyond that, in the next circle, one depicts those with impact best depicted at that level, and so on. *The social atom is a depiction of one's own truth and perception.* The spatial distance from the center to others reflects the perception of the nature and impact of those persons. A menu of symbols and lines is given that denote genders, as well as positive, negative, neutral, lost, ambivalent, mutual, and one-way relationships.

There is flexibility with this tool, which provides information and insight and points to areas of further exploration. J.L. Moreno believed we are as closely tied to our social atoms as we are to our bodies. The social atom predates the genogram and ecomap by over 40 years and may have influenced their development in the family therapy and social work fields.

Encounter

In 1914, Moreno published a seminal poem articulating his philosophy of the human encounter. This poem highlights the importance of the existential experience of seeing the world and one's self through the eyes of another. The philosophy of the encounter is central to Moreno's developmental theory and is operationalized through the psychodramatic intervention of role reversal (which also influenced Buber's 1923 I and Thou concept, Fritz Perls' development of Gestalt Therapy, and many others).

Encounter represents a theoretical and practical underpinning to psychodrama. Encounter is an action instrument, a basic idea, for an existential dialogue with another, imagined or real. With the encounter, one finds oneself in the *psychodramatic experience of me and not me.* This is an existential fiber in life for what connects us and an experience often had with psychodrama. J.L. Moreno gave a name to the mutual feeling of the essentialness of another, calling that sense of another *tele,* noted earlier in this chapter. The interweaving of tele is explored in psychodrama, sociodrama, and sociometry.

J.L. Moreno wrote about the existence of a primordial nature in life, found within us and within groups. In *To Dream Again–A Memoir,* Zerka Moreno named this primordial nature the *autonomous healing center* within. It is generated in psychodrama to craft a new, novel, and adequate response to life.

This is the case when the protagonist explores a role reversal, in taking another's role in language, behavior, and feeling. One finds a sense of what that other might feel like; and what might be learned from this. It is in the encounter of oneself with others and life that spontaneity-creativity is discovered and strengthened. Facilitating role reversal and encounter requires training; for information regarding training and certification for psychodrama, refer to the American Board of Examiners in Psychodrama, Sociometry, and Group Psychotherapy. There are similar Boards throughout the world.

Canon of Creativity: Formula for Change

The Canon of Creativity is a visual diagram to represent a theoretical underpinning to the psychodrama method. The diagram identifies stages of spontaneity-creativity for a person or group–the warming-up process, spontaneity, creativity, and the cultural conservations. It is a map for change that depicts the process of creating something new based on spontaneity-creativity theory. The Canon of Creativity is a theory actualized and brought to life by psychodrama and includes the warm-up where group members prepare for spontaneity-creativity in action. There is then a movement to create something new, which is met at first with resistance. Once resistance is addressed, then renewal and creativity can emerge. There is the movement to form–and presence–in the world and life. The Canon of Creativity can be thought of as a clinical map embodying psychodrama's theory of change which proposes spontaneity as the curative agent in the treatment process. Moreno defines spontaneity as providing the ability for adequate responses to new situations and novel responses to old, reoccurring situations. This framework can be applied to conceptualize the change process psychologically, interpersonally, or within society. The renewal of society was J.L. Moreno's full vision to bring creativity and spontaneity into the culture.

Human Development

Jacob Moreno's developmental theory is simplified into three phases which also reflect psychodrama interventions–the double, the mirror, and the role reversal. Zerka Moreno expands on the developmental stages while synthesizing them with Erickson's psychosocial stages in *The Quintessential Zerka.* She outlines periods from birth to death, each stage offering guidance for tasks to be considered developmentally. The developmental stages and tasks are templates for the realization of creative potential and have life tasks for growth.

Psychodrama helps to experience the tasks, even those not originally completed. To become more able to role reverse with others, including groups of people, is one of the developmental tasks. There is a task of completing unfinished business with one's parents and preparing to yield parents back to the cosmos as we eventually ready ourselves to leave this world. Developmental stages include taking on new roles, finishing others, and crafting one's continued path.

It is useful to consider that in each of the developmental phases, different psychodrama tools are highlighted as instruments to help with that phase. Also, in conducting psychodrama, it's wise to have as a mental construct, the notion that what protagonists are showing could be related to the tool needed for a developmental stage being completed.

Role Theory

Role theory suggests that the self develops and emerges through the process and actualization of roles throughout one's life. Humans are fundamentally role-players, and all of our relationships are composed of role relationships. Roles include internal and external aspects or functions of self-ranging from somatic roles, psychodramatic roles, and social roles. Role theory proposes a nonpathologizing and interdisciplinary theory of personality that encompasses psychiatry, sociology, theater, and culture. Role theory is a holistic and comprehensive system of conceptualizing human experience in a biopsychosocial-spiritual context.

Role training, or learning new roles in life, works with a three-stage process: (1) Role-Taking: deciding to take a role or not, addressing ambivalence and resistance, (2) Role-Playing: assuming a role, finding role models, examples of how to function with increasing skill in a role, (3) Role-Creating: emerging more fully as oneself and with increasing skill, by creating new and authentic expressions of oneself. Role theory also contains an articulation of various role phenomena and dynamics, including role reciprocity, role conflict, role fatigue, subroles, role-self congruence, and more.

The method considers our psychological and emotional growth comes through how we start, learn, play, live, change, and finish the roles and role relationships throughout our lives. A personality characterized by health has a wide role repertoire, which allows for spontaneous and skillful living. To attend to this consciously can be supported with the methods stated here. Psychodrama can be used to explore and prepare for the roles we want, or want to change, in some fashion.

Surplus Reality

Surplus reality is a wide-ranging theoretical tool for a protagonist and group to explore psychodramatically what can be imagined beyond present reality. A variety of techniques are used to work within surplus reality, which is a foundation for psychodrama. The protagonist creates a psychodrama that includes and can go beyond known reality, using imagination, perception, creativity, and psychodramatic activity. The director helps the protagonist explore the real or imagined and forge an adequate response.

Techniques include role reversal, double, mirror, soliloquy, aside, concretization, presentation of oneself in the role of another, and changing time, space, and location. Surplus reality can involve empty chairs representing people, events, aspects of oneself, stages of life, or anything else. Surplus reality can offer the protagonist an experience as it was needed but did not happen, aspects of the past, present, or future, and roles that may be required.

Surplus reality touches upon imagination and creativity. It opens to new worlds, exploring an event, redoing an event, or exploring an event that never happened. As such, psychodrama is particularly useful in the treatment of trauma in that it can provide corrective emotional experiences that would have been impossible otherwise. Auxiliary egos, as they are called, are group members asked to take part in the protagonist's exploration. Regardless of the interaction, in many ways, the protagonist explores strengths, capacities, and resources along with vulnerabilities and creativity. There is a great range of choices that surplus reality offers.

Universalia of Time, Space, Reality, the Cosmos

The psychodrama method uses time, space, reality, surplus reality, and the existential realm of the cosmos, as J.L. Moreno described it, as psychodramatic tools. The past, present, and future are reflected in the here and now, making time a flexible instrument. A protagonist may explore a scene that happened in childhood, then change scenes to one imagined in the future. A sequencing of scenes and times may be explored or markers on one's life timeline.

Psychodrama explores life as it happened or was needed or wished for. The director uses psychodrama to facilitate the protagonist's exploration and repair. The connection of human beings to life relates us to an inherent dignity. Many report psychodrama to be an experiential encounter that reveals this universal dimension.

Warm-Up, Action, Sharing—Structure of Psychodrama

Psychodrama is a three-step process of warm-up, action, sharing, with processing for those in training. The warm-up constitutes approximately one-third of the session, the action phase one-third, and sharing phase one-third. There are many approaches to conducting a warm-up, which can be quite organic to help the group meet one another, for participants to feel present, to make connections to others, and for the director to introduce themes. The warm-up is a natural tendency in human life: we have a warm-up, for example, as a preparation for going to sleep, a warm-up for going to work, or for addressing difficult issues with another. This phase is natural to how we live.

The warm-up process introduces creativity in some form to the group and individuals. A warm-up may be planned or spontaneous and may include movement, conversation, reflection, art, or music. The warm-up is preparation for the action phase. The director weaves and orients the positive telic connections, enhancing the social atom of the group.

The action phase is the psychodrama or sociodrama itself. In psychodrama, a protagonist may be chosen by the group, self-selected, selected by the director, or scheduled. Some have suggested that a sociometric/democratic selection of the psychodrama topic is the most effective approach as it ensures a protagonist that most represents the group-as-a-whole. This process helps to facilitate a group-as-a-whole experience and side-steps the potential pitfall of doing individual therapy with a group audience. The protagonist enters the stage, and there is often an initial dialogue between the protagonist and the director about goals, boundaries, time frames, and elements of the psychodrama.

The director works in collaboration with the protagonist. The group remains a group, even as some group members may be asked to enact a role in the psychodrama. The action phase often starts with a scene or some expression of present concerns. There may be a number of scenes or just one scene. There is often a peak of action and integration of experiential learning.

Sharing is the final third of psychodrama. With the psychodrama completed, group members share with the protagonist and group how the psychodrama was related to one’s life. Those who assume a role in the psychodrama generally share first, initially from the experience of the role, then as it relates to their life. Sharing is not advice-giving or analysis; the director may share as well.

Do–Undo–Redo

Do–Undo–Redo is one option for a methodologic sequence for a psychodrama as a strategy for change and structure for the action phase. The protagonist may begin by showing a condition or concern, perhaps stepping in and out of the concern, concretized in some way on the stage. The protagonist may encounter an aspect or aspects they wish to address or change. The protagonist can undo things and, following that, redo them differently or from a different vantage point. There can be a scene and then a redone scene in which something new has been created. There can be the reconstruction of the situation or elements in it. This is a flexible structure adjusted to meet the needs of the protagonist.

One quite beautiful application of this has been the “moments from one’s life” in which the protagonist (often many protagonists) may “relive” an experience in life, often a joyful and enlightening one.

PSYCHODRAMA AS A TREATMENT OF CHOICE FOR TRAUMA

Many are endorsing psychodrama and experiential therapies as a treatment of choice for trauma based on the influx of new neuroscience research illuminating how trauma impacts the brain and body at levels beyond cognition or verbal processing. Psychodrama’s capacity for undoing and redoing moments in the past is ripe with potential for corrective emotional experiences and renegotiating trauma without the need for re-enacting a trauma scene. Trauma can be approached from a place of safety through strength-based roles that promote transformation and post-traumatic growth. In a recent study on trauma-focused psychodrama, participants commented on therapy as being more tolerable and less intimidating than other trauma therapies that they had participated in–even describing it as fun and playful.

Jacob Moreno created psychodrama 100 years ago, long before post-traumatic stress disorder (PTSD) was recognized as a diagnosis in 1980. Although he did not have the same language to articulate what we now know about trauma and its impact on people, most of Moreno’s work (and much of the larger psychodrama community’s work) was with trauma survivors–including immigrants/refugees, inmates, at-risk youth, veterans, and patients with severe mental health conditions in inpatient psychiatric hospitals. The emergence of contemporary trauma theory and neuroscience findings both support the use of psychodrama and have been further integrated to refine the use of Moreno’s methods in safe and effective ways. When employed without proper training, psychodrama can be retraumatizing or harmful for clients. Its use as a treatment modality with trauma survivors requires advanced training and careful consideration.

PSYCHODRAMA RESEARCH

Psychodrama is used in both psychotherapy contexts and other spheres of practice, such as education, religion, activism, and community life. Its expansive scope, along with its foundation in spontaneity-creativity theory, has made the production of psychodrama research challenging. It has obtained status as an evidence-based practice (EBP) in other countries (especially in Europe), but its establishment within the American mental health system remains insecure. The scarcity of psychodrama’s presence in academia further limits its access to research support and funding. Useful resources for exploring psychodrama research further include the research committees of the ASGPP, IAGP, and the Federation of European Psychodrama Training Organizations (FEPTO).

Psychodrama research continues to emerge, supporting its effectiveness in the treatment of various mental health disorders, providing experiential training, and enhancing well-being or social skills. Psychodrama research is also included in the larger research bases of group psychotherapy, creative arts therapies, humanistic-experiential psychotherapies, and body- and movement-oriented interventions–all of which demonstrate levels of efficacy comparable to cognitive-behavioral therapy and other EBPs.

ETHICS IN TRAINING, EDUCATION, AND PRACTICE

A challenge in conducting psychodrama is keeping the experiential structures intact while having flexibility in the application. This places a demand on the director, educator, trainer, and practitioner to evaluate and apply the psychodrama method with a solid foundation of theory, training, education, practice, and certification. The strategic thinking for directing requires training, supervision, and professional development. The American Board of Examiners in Psychodrama, Sociometry, and Group Psychotherapy sets standards and examines candidates for two levels of national certification: Certified Practitioner (CP) and Trainer-Educator-Practitioner (TEP). The certification process is extensive and requires 780 training hours (among other requirements for entry-level certification). Similar Certification Boards exist in South America, Australia, Europe, and Asia. The American Board has adopted the ethical standards of the American Psychological Association. Psychodrama should not be practiced without Board approved training in accordance with the established certification guidelines. The necessary intensity of the psychodrama training process reflects the complexities of the model and the potential for harm when practiced outside the scope of one's competency.

Though psychodramatists officially adopt the APA code of ethics, Moreno also published a *Code of Ethics for Group Psychotherapy and Psychodrama: Relationship to the Hippocratic Oath,* which is embedded into the philosophy of his approaches. One of his novel ideas articulated in the code of ethics is the empowerment of each patient as a therapeutic agent. He suggests that because each patient has the power to promote healing for each other, patients should also be prompted to commit to aspects of the Hippocratic Oath. Specifically, the code of ethics suggests that the group leader and participants should commit to: (1) do no harm, (2) maintain confidentiality, (3) promote healing for others in the group and contribute to the overall welfare of the group, and (4) foster equality of status within the group, regardless of privilege or marginalized statuses outside of the group.

SOCIAL AND ORGANIC UNITY OF HUMANITY

J.L. Moreno contributed to the ideas of social psychiatry in many ways as he wrote extensively about groups, communities, society, and social forces impacting humankind. He is mostly known as the developer of psychodrama, but was also deeply involved in societal problems, which is depicted in his book *Who Shall Survive?*:

> *The discovery that human society has an actual, dynamic, central structure underlying and determining all its peripheral and formal groupings may one day be considered as the cornerstone of all social science. This central structure—once it has been identified—is either found or discernible in every form of human society, from the most primitive to the most civilized: It is in the genesis of every type of society. In addition, it exerts a determining influence upon every sphere in which the factor of human interrelations is an active agent in economics, biology, social pathology, politics, government, and similar spheres of social action.*

To see society's actual, dynamic, central structure underlying and determining all peripheral and formal groupings offers a perspective on what impacts our present situations. The situation calls for a willingness to reverse roles with others and other groups of people and find what is common among us to find solutions that consider all of mankind and the impacted natural world. Sociatry endeavors to understand this underlying phenomenon in human society and create education for personal and professional development about the underlying structures impacting humankind and the ecologic fabric. It is a creative path for which J.L. Moreno offers:

> *A truly therapeutic procedure cannot have less an objective than the whole of humankind. But no adequate therapy can be prescribed as long as humankind is not a unity in some fashion and as long as its organization is unknown. It helped in the beginning to think, although we had no definite proof, that humankind was a social and organic unity.*

Based on this, it may be an idea developed of necessity. As the whole of mankind is a unity in some fashion, tendencies emerge between different parts, drawing some apart, drawing some together. This may be advantageous for some at times and disadvantageous to others. If true that there is a unity among us, to humanity, biologically or otherwise, if a unity of humankind does exist, how can we find and realize it? From this, further research could be generated.

FURTHER READINGS

Blatner A. *Foundations of Psychodrama: History, Theory and Practice*. Springer; 2000.

Dayton T. *NeuroPsychodrama in the Treatment of Relational Trauma: A Strength-based, Experiential Model for Healing PTSD*. Health Communications, Inc.; 2015.

Dayton T. *The Living Stage: A Step-by-Step Guide to Psychodrama, Sociometry, and Experiential Group Therapy*. Health Communications Inc.; 2005.

Fox J. *The Essential Moreno*. Springer; 1987.

Giacomucci S. Experiential sociometry in group work: mutual aid for the group-as-a-whole. *Social Work with Groups*. 2021.

Giacomucci S. *Social Work, Sociometry, and Psychodrama: Experiential Approaches for Group Therapists, Community Leaders, and Social Workers*. Springer Nature; 2021.

Giacomucci S. The trauma survivor's inner role atom: a clinical map for post-traumatic growth. *J Group Psychother Psychodrama Sociom*. 2018;66(1):115–129.

Giacomucci S, Karner D, Nieto L, Schreiber E. Sociatry, psychodrama, and social work: moreno's mysticism and social justice tradition. *Social Work with Groups.* 2021.

Giacomucci S, Marquit J. The effectiveness of trauma-focused psychodrama in the treatment of PTSD in inpatient substance abuse treatment. *Front Psychol*. 2020;11:896.

Giacomucci S, Stone AM. Being in two places at once: renegotiating traumatic experience through the surplus reality of psychodrama. *Social Work with Groups*. 2019;42(3):184–196.

Hovatin T, Schreiber E, eds. *The Quintessential Zerka: Writings by Zerka Toeman Moreno on Psychodrama, Sociometry and Group Psychotherapy*. Routledge; 2006.

Hudgins MK, Toscani F. *Healing World Trauma with The Therapeutic Spiral Model: Stories from the Frontlines*. Jessica Kingsley Publishers; 2013.

Marineau R. *Jacob Levy Moreno 1889–1974: Father of Psychodrama, Sociometry and Group Psychotherapy*. Routledge; 1989.

Minkin R. *Sociodrama for Our Time—Sociodrama Manual*. East West Sociodrama Institute; 2011.

Moreno JD. *Impromptu Man: J.L. Moreno and the Origins of Psychodrama, Encounter Culture, and the Social Network*. Bellevue Literary Press; 2014.

Moreno JL. Code of ethics of group psychotherapists. *Group Psychother*. 1957;10:143–144.
Moreno JL. *Sociometry and the Science of Man*. Beacon House; 1956.
Moreno JL. *Who Shall Survive?* 2nd ed. Beacon House; 1953.
Moreno JL. In: Schreiber E, Kelley S, Giacomucci S, eds. *The Autobiography of a Genius*. North West Psychodrama Association; 2019.
Moreno JL, Moreno ZT. *Psychodrama*. Vol. 2. Beacon House; 1959.
Moreno JL, Moreno ZT. *Psychodrama*. Vol. 3. Beacon House; 1969.
Moreno ZT. In: Horvatin T, Schreiber E, eds. *The Quintessential Zerka*. Routledge; 2006.
Moreno ZT. *To Dream Again – A Memoir*. Mental Health Resources; 2012.
Moreno ZT, Blomkvist LD, Rutzel T. *Psychodrama, Surplus Reality and the Art of Healing*. Routledge; 2000.
Van der Kolk BA. *The Body Keeps the Score: Brain, Mind, and Body in the Healing of Trauma*. Viking Press; 2014.

32.23 Other Methods of Psychotherapy

ADAM M. BRENNER, M.D., AND AARON VAN WRIGHT, M.D.

It is well established that a majority of the effectiveness of psychotherapeutic treatments can be accounted for by the common elements that all psychotherapies share. The features shared by all include a healer–patient relationship, in which different roles carry different expectations and a differential balance of power; a nonjudgmental, supportive acceptance of the patient; and an alliance of working together on shared goals. These common elements are rooted in our universal need for attachment and connection. Our brains are designed to seek out attachment when we feel isolated or threatened and to reciprocally be moved by the pain and loneliness of those who reach out to us. Resting on this foundation, psychotherapies of many different stripes are able to provide an experience–sometimes as quickly as the initial session–of feeling comforted, finding hope, and experiencing an amelioration of the sense of being alone or marginalized by one's presenting difficulties.

But if all therapies share these comforting elements, what accounts for the existence of so many different psychotherapies? The development of a specific psychotherapeutic method begins with the creation of a psychological model. Each individual model is built upon a unique conceptualization of how psychopathology evolves and what must happen to achieve an effective and satisfying life, either through the resolution of the psychopathology or in spite of the ongoing psychopathology. Thus, psychotherapy is always a learning process, taking advantage of the brain's neuroplasticity and the related human capacity for durable change at the most fundamental levels of being. Yet, as in formal education settings, learning within psychotherapy is a unique process among individuals and requires a diverse range of approaches. One way to understand the wide array of psychotherapeutic methods, therefore, is as an attempt to accommodate the various learning styles.

In addition, what the therapist hopes the patient will learn in the course of treatment is remarkably diverse and often rooted in very different fundamental ideas about the nature of psychopathology and what defines an effective and satisfying life. As an example, the cognitive therapist believes patients need to learn about implicit maladaptive thoughts, which eventually results in changes to deeply held belief systems that contribute to recurrent emotional distress and dysfunctional interpersonal engagement. The psychoanalytic therapist, on the other hand, believes the most effective learning approach involves the exploration of unconscious wishes and conflicts, which allows for the release of long-standing, self-imposed obstacles that hold people back from living life in the fullest way possible. In both instances, the aim is to relieve suffering and promote growth and development that, hopefully, will prevent suffering in the future. How this aim is achieved, and the specific goals that occur within the treatment itself are quite different.

Through the years, new psychotherapeutic approaches have been developed, usually borne out of a desire to fill an identified need in the treatment of mental illness. This section reviews several psychotherapies, some with a particular political slant or dealing with a special population, and some older therapies that continue to have discrete followings. Currently, most of these therapies do not have a rigorous evidence base to support their efficacy, but some continue to be studied. In addition, some of the ideas in these treatments are still in use in altered formats and may have relevance in their ongoing influence on the field. Each of these therapies, if effective, has been able to generate and benefit from the comforting elements common to all therapies, but each of them has had very different ideas about what kind of learning will be of benefit to the patient and what the therapist should do to help “teach” the patient.

CLIENT-CENTERED THERAPY

History

Client-centered therapy was created in 1940 by Carl Rogers, a psychologist. Believing that diagnosis, planning, and interpretation manipulate the patient to conform to the therapist’s ideas, Rogers tried only to listen, to understand, and to reflect what he perceived the patient to feel. When Rogers did so, he found that the patient often responded with a deep feeling of personal involvement and the production of new and important material. That passive or nondirective therapy was later named *client-centered therapy* and was published as such in 1951. Rogers’ ideas found wide acceptance and a movement around the principles of the therapy gathered momentum and grew over the next three decades. Rogers himself continued to develop the method and its applications at the University of Wisconsin, the University of Chicago, and Ohio State University and in La Jolla, California, where he died in 1987.

Client-centered therapy is based on the idea that each person has the ability to solve psychological problems and that feeling understood and highly valued aids the forces within to do so. Ideas such as an instinct toward self-realization and limitless inner powers had enormous appeal to the humanistic idealism that was developing at approximately the time that the method was published.

A focus on the value of the individual and Rogers' later precept that long and particular training was not necessary for becoming a good therapist were similar rallying points for the antiestablishment, antiauthority emotions of the 1960s and 1970s.

Rogers played a major role in wresting the practice of psychotherapy from psychiatry alone and distributing it into the hands of psychology and other disciplines. He is also credited with being the first to research his results, the first to record sessions word by word and in movies and on videotape, the first to publish transcripts of his failures to encourage the study of what goes wrong, and the first to put forward the thesis that one does not need academic degrees to become a therapist. All but the last idea have become well-accepted elements in the field.

Client-centered therapy has been reported to be useful for children and adults, problems of neuroses, situational problems, speech difficulties, psychosomatic problems (e.g., allergies), and, to some extent, psychosis. It has been practiced with individuals and groups, and its principles have been applied in industry, education, child-rearing, and intensive encounter groups. Just before his death in 1987, Rogers was training policy makers and leading communication groups in South Africa and in workshops in Hungary and the former Soviet Union.

Theoretical Issues

The overarching presupposition of client-centered psychotherapy is that a basic, forward-moving force exists within—a tendency to grow, actualize, and integrate. A number of other principles flow from that idea: All people are believed to have the capacity to understand what factors cause them unhappiness and to overcome those factors; to do so, they must bring what is going on inside into awareness; and, effectuate the circumstances that make awareness possible, the therapist must establish a relationship that is warm and accepting.

Rogers believed that human beings immediately attach meaning to whatever is perceived so that perception becomes reality. A group of perceptions is organized as the self-concept—people's perceptions of their own characteristics, their relationships with others, and the value they place on those perceptions. Experiences contradicting the preferred perceptions of that organization cause anxiety and must be shut out. In therapy, the structure of the self is relaxed, so that the patient can drop the barrier and admit, “Yes, I do have those parts of myself that are unacceptable.” When this happens, the previously denied parts are accepted and then integrated, and the result is a better-functioning person, one who is less tied up in denial and conflict, more effective, and more tolerant of others than in the past.

The aim of the therapy is to create an interpersonal environment in which the self-structure can relax and in which material can come into consciousness. The therapist sets the three conditions that foster relaxation and the personal growth that follows. The three conditions are *congruence* (or *genuineness*), the ability to be present as a person, to be real in the relationship with the client; *unconditional positive regard*, which conveys to the client, “I accept you as you are, without making judgments or evaluations”; and real *empathic understanding*, a sense of standing in the client’s shoes, seeing the world from that vantage point, and then communicating what is understood from that view to the client. When these conditions are met, the client feels released and understood and is able to accept and integrate, view the world differently, and continue the process by bringing out new material.

Client-centered therapy was first practiced in a course of 5 or 6 hours, then 15 or 20 hours, and, still later, 100 or more hours. Whether the shift was caused by clients’ seeking greater levels of change, the enlistment in treatment of more difficult cases, an evolution in technique, or a fall-off in results after the initial enthusiasm is unclear.

Techniques

The techniques of client-centered psychotherapy serve the goal of having the therapist convey a sense of the worth and significance of the client and the idea that the therapist is trying to understand. Far from being laissez faire, the therapist needs to help actively and empathetically clarify what the patient feels. To be entirely nondirective, the therapist restates and reflects a close approximation of the client's feelings. When this is done correctly, clients feel as if the therapist were almost a part of them, working on their problems as they want to work on them; they feel accepted warmly, without any judgmental overtones. The acceptance and empathy are themselves therapeutic.

Rogers cited an example of an explosive 13-year-old boy who was changed and calmed by sessions in which he and the therapist only played tic-tac-toe or in which the therapist allowed silence while the patient sat at the window with his back to him, quietly counting cars. The therapist conveyed respect and belief in the child and the idea that the hour belonged to the child and that the adult would not be shocked by anything he did. The only limitation was that anger was allowed no physical expression, only symbolic and verbal expression.

The therapy aims not to interpret the client's unconscious motivations or conflicts but to reflect what the client feels. As the counselor and the client progressively clarify their feelings within the warm and accepting relationship, denied feelings come to light and can be dealt with and integrated.

Rogers described a male patient with considerable anger. As the therapist accepted the client's feelings and clarified to whom they were aimed, the client came to see which parts were rational and which were not. He was able to say, "If you are like my father, or if I only see you that way, I hate you as I hated him." When the unrealistic aspects were expressed openly, they dissolved, leaving the real relationship in place.

The theory also claims that resistance is overcome rather than interpreted. That is accomplished by the consistent acceptance and valuing of the client, the thoroughly undemanding nature of the relationship, and the empathic effort to clarify the client's feelings.

A woman with denied anger at her husband embarrassed him with dizziness and faintness at gatherings that were important to him. The therapist consistently reflected her feelings to her with such questions as “You seem to feel…,” “You’re confused by…,” and “It seems to you that….” That clarification helped her to move from affirmations of a good relationship with her husband to recollections of old resentments and the recognition of present ill will. Keeping the focus on her also aborted any tendencies for her to project or accuse the therapist of anything but interest.

Transference is allegedly avoided in a similar fashion. Rogers gave examples of an angry client:

Client: Am I supposed to tell you everything?

Therapist: It’s up to you; you can talk about anything you want to.

Client: I had a big fight with you in my mind as I came here. I thought the whole thing was a fraud—I’m being very frank—you can talk like we do down here almost anywhere if you get someone to listen.

Therapist: It seemed it was kind of a cheat, something you could get anywhere.

Client: It’s like advertisements you see where, for a few dollars, people will sit and listen to you.

Therapist: Just listen to your troubles.

Client: Of course, they don’t have your extensive background, but they do the same thing. Then, after I talked it out with myself, I realized that the reason I was mad at you is because I feel nervous and jittery coming here (pause), because I’m facing something I don’t like to, talking things out myself.

Therapist: So the feelings about someone listening to you being a fraud–you recognize as being connected with your fear and irritation at having to face things in yourself?

Client: That’s right. That’s the real logic.

In another example with chaotic angry feelings, the following dialogue took place:

Client: I hate you! I wish you never were born!
Therapist: You just hate me bitterly.
Client: I think I'll throw you in the lake. I'll cut you up! You think people like you, but they don't. I wish you were dead!
Therapist: You detest me, and you'd like to get rid of me.
Client: You think my father did bad things with me, but he didn't! You think he wasn't good, but he was! You think I want intercourse, but I don't.
Therapist: You feel I absolutely misrepresent all your thoughts.
Through this type of work, the client later confessed to a hallucination and said: "I had to get rid of this somewhere. I knew you'd understand. It's true I hate myself, but I couldn't say it, so I thought of all the ugly things I could say to you instead."

Rogers wrote that those interviews showed how feelings of transference were dealt with in the same way as other feelings and how a consistent reflection of the client's feelings essentially dissolved a transference issue by helping the client recognize their irrational emotions, after which they could deal with them and what they hid.

Indications

The general indications for therapy are when clients feel that their organized self-structures are no longer effective in meeting their needs in the reality situation or when they perceive discrepancies in themselves or see that their behavior is out of control. The description generally covers problems in the neurotic range, but a client-centered approach has been used with children who are 2 years of age and adults who are 65+ years of age, with people who have mild adjustment problems and disorders diagnosed as psychotic, and with persons of all social classes and levels of intelligence. The theory affirms that a trial of the therapy is unlikely to do harm.

Early in the development of the method, Rogers believed that diagnosis was inadequate and prejudicial. Placing the locus of evaluation in the hands of the therapist rather than the client had undesirable consequences, undermining the client's power of self-determination and raising the specter of misuse for social control. Nosologic limitations may have supported this early position, and the social circumstances of the 1960s may have lent it credence, but Rogers' followers believed in it as late as 1987.

In recent years, client-centered therapy is referred to as person-centered therapy. It has been applied in such areas as working with transgender and gender-diverse clients, and in community-based treatment of acquired brain injury (ABI). In the case of transgender and gender-diverse clients, the authors of a recent article discuss the use of a person-centered approach integrated with current scholarship to offer specific interventions for use in these clients. The authors describe using these interventions to focus on treatment of rifts within the individual that are internalized from societal rejection. In the treatment of ABI, one recent article presents a case that successfully incorporated person-centered therapy to treat challenging behaviors and avoid hospitalization.

Ethical Issues, Research, and Evaluation

Rogers' ideas sparked and rode the crest of the humanistic movement in psychology. His emphasis on the value of clients and on the need to accept and respect them in therapeutic relationships was timely and had a refreshing influence on child-rearing, education, and group relations, as well as on psychotherapy. His attempts to train nonprofessionals were less well-received, and the attempts foundered; the method became so simplified that blank verbal repetitions too easily replaced a true empathic congruence and effort to understand.

Practitioners of client-centered psychotherapy did not believe that diagnosis was particularly useful and did not follow the medical model of illness. That a warm and genuine relationship is sufficient to dissolve resistances that are dynamically motivated and tenaciously held is another tenet open to question. Reflecting on what the client feels without interpreting where it comes from or the purposes that it may serve may drive the client to action as a further defense. Rogers noted that the method is frequently unsuccessful with clients who feel incapable of managing themselves and who insist that the therapist take over; a different approach interprets the client’s insistence and resentment, rather than just reflecting them. Similarly, reflection in the face of other intense transferences seems to be less effective than a focus on the fantasies that lend force to the transference.

In its contributions to psychotherapy research, client-centered therapy has been outstanding. Rogers and his followers were the first to record and review transcripts of sessions, the first to document the process of treatment, and the first to do follow-up studies of results. They developed such innovations as the q-sort and scales to objectify the Rorschach test. They showed that, in successful therapy, negative attitudes about the self are replaced by positive attitudes, that acceptance of others occurs in parallel, that psychological tension decreases, and that plans take the place of old complaints. Although initial claims about success in treating schizophrenia patients had to be dropped, and although Rogers and his colleagues did not do controlled comparisons of their method with others, they were able to document that client-centered therapy is effective. By clearly delineating that a successful outcome is associated with genuine acceptance and an empathic congruence that facilitates the recognition and expression of emotion, they recognized two of the features that empower psychotherapy of whatever stripe. Perhaps most importantly, their emphasis on progressive clarification of the client’s feelings sounds an insistent cautionary note about the need for clear understanding, rather than facile interpretation based on the assumptions of one’s theory.

DIGNITY THERAPY

History

Beginning in the 1960s in the United Kingdom and gaining traction in the United States in the 1980s, the modern hospice movement changed the way in which end-of-life care is delivered, placing an emphasis on maintenance of functioning and physical comfort over aggressive illness-focused treatment. The subsequent establishment of palliative care as a board-certified internal medicine subspecialty in 2006 brought further credibility to the approach championed by the modern hospice movement. The traditional palliative care model uses a multidisciplinary approach to address the far-ranging issues associated with terminal illness, including pain management, assistance with the provision of home medical care, and education and support for caregivers. While the scientific evidence overwhelmingly demonstrates positive gains in quality of life for both patients and loved ones when palliative care methods are utilized, observational studies continue to demonstrate high rates of depression and anxiety in patients with terminal illness. Complicating matters, research also suggests patients with terminal illness, especially those who become sick later in life, do not respond well to traditional antidepressant and anxiolytic treatment approaches.

Beginning in the late 1990s, a group at the University of Manitoba led by Dr. Harvey Max Chochinov set out to address the both prevention of psychological distress and improving response to psychiatric treatment in end-of-life care. Dr. Chochinov and his research group designed and tested a brief, manualized psychotherapeutic approach coined “Dignity Therapy” that was shown to reduce psychological distress in both patients with terminal illness and their caregivers, leading to improved outcomes in quality of life and lower rates of depression and anxiety. Subsequent to the publication of Dr. Chochinov’s research, other groups tested the approach and found similar rates of effectiveness across an even wider range of diagnoses and conditions. This growing body of evidence has led some to call for similar approaches to be incorporated not only into routine palliative care practices but also into the care of those with debilitating but nonterminal chronic medical illness.

Theoretical Issues

It is not entirely clear why patients with terminal illness have a high prevalence of depression and anxiety, and why they may not respond as well to traditional therapies. A common formulation is that the decline in physical, cognitive, and social functioning associated with terminal illness leads to a loss of “dignity” on the part of the patient, which in turn, leads to feelings of helplessness, hopelessness, isolation, and despair. Therefore, in order to prevent the development of depression and anxiety and to treat the symptoms of these conditions if they emerge, one must find a way to maintain a person’s dignity as they move toward death.

One of the clear results from dignity research in terminal illness is the mere exploration of dignity with most dying patients leads to the desired therapeutic effect. For some patients, though, a more systematic process is required. When thinking about the development of a systematized management approach, a complex picture emerges. First, dignity is a highly subjective and variable concept, and so common elements must be identified that cut across populations within the context of end-of-life care. Then, careful analysis and categorization of these common elements are necessary so that a cohesive model for dignity can be formed. Finally, a therapeutic process that operationalizes this model must be tested to gauge the validity of the model and the effect of the therapeutic process on those parts of a person’s identity that contribute to dignity. Finally, the model of dignity must be sufficiently dynamic to account for the physical, mental, and social changes that occur with terminal illness, and the therapeutic process must be both accessible and engaging for patients whose ability to participate in traditional psychotherapeutic approaches often is compromised.

Drawing on techniques of narrative, client-centered, and existential psychotherapies and set within a framework of dignity grounded in

Eriksonian developmental ideals, Dignity therapy seeks to assist patients as they strive to resolve questions and internal conflict surrounding the terminal and often debilitating nature of their illness, their evolving role in pre-existing social constructs, and the impact these two factors have on self-identity.

Dignity therapy aims to promote a lasting sense of generativity while simultaneously engaging in the development of ego integrity. The underlying idea is that very often, terminal illness develops during the final years of one's middle adulthood. It also, then, compresses late adulthood into a much shorter timeframe. One of the keys, therefore, to alleviate psychological distress is to help patients complete the developmental tasks of these life stages. In order to accomplish this feat, the patients are charged with telling their story and describing themselves in their own words, while being prompted by the interviewer with probing questions focused on the exploration of their life's meaning and being refocused continually on those things that appear most central and engender the greatest sense of pride. The end result is a retrospective process that helps patients see their life through a lens of contentment and satisfaction, along with the production of a written portrait, with which patients are allowed to do as they please.

Techniques

Dignity therapy is a brief, manualized psychotherapeutic treatment intended to be delivered on an individual basis by a qualified mental health professional. The treatment typically requires three to four sessions, spaced over a week's time. The duration of each patient visit varies with the specific goals of the session and the volume of material discussed. Typically, though, sessions last between 30 and 60 minutes. Dignity therapy has been validated in outpatient, residential, and inpatient settings.

The basic structure of a course in Dignity therapy is as follows:

Session #1. The first session consists of four main components.

1. Administration of rating scales—These can be self-report or clinician-administered measurements. The tools used should include, at a minimum, screening questions for depression, anxiety, hopelessness, and suicidality, as well as ratings for sense of well-being and quality of life.
2. Introduction to the process and goals of therapy—Special attention should be given to the collaborative nature of the therapy and the production of the "generativity document," which is the edited written transcript provided to the patient at the conclusion of the therapy.
3. Informed consent for treatment and audio recording—This will have critical importance when considering the ethical implications of the treatment.
4. Written copy of the questions included in the Dignity therapy protocol provided for review and consideration—The same nine questions are used in every course of treatment.

Session #2. The second session is the audio-recording session and begins with the therapist asking the patient the first question on the Dignity therapy question protocol:

"Tell me a little about your life history; particularly the parts that you either remember most or think are the most important?"

The therapist uses the remaining questions included in the Dignity therapy protocol in a flexible manner, following the patient's lead and paying particular attention to the patient's affect to determine when to ask relevant follow-up questions. The goal is not to obtain a response to each question included in the protocol, but to prompt the patient to verbalize those aspects of their life that are viewed as essential to positive self-identity. The therapist utilizes a combination of reflective listening, displays of empathy, dynamic clarification, and existential probing. Additionally, the use of guided imagery and relaxation techniques can be utilized if patients struggle to put together their thoughts or experience acute psychological distress.

Following Session #2, the audio recording is transcribed and edited into a coherent narrative that conveys the tone and message desired by the patient. To this effect, an essential component of this editing process is finding a single, concluding statement that captures the patient’s desired lasting legacy.

Session #3. During the third session, the narrative produced from the edited transcript of the audio-recorded session is read aloud to the patient. This allows the patient to provide feedback on the historical accuracy of the narrative and to determine if there are any anecdotes or information that should be removed or included. In addition, reflecting upon their own self-portrait evokes a strong emotional response, and this helps to solidify ego integrity.

Session #4. Any feedback provided by the patient is used to create a final draft of the narrative, which is termed the “generativity document.” In the final session, this document is provided to the patient, and the patient’s plans for the document are considered. Additionally, the patient is invited to reflect upon the therapeutic process. Any concerns expressed by the patient are clarified and addressed. Finally, the initial rating scales readministered to gauge the therapeutic effect of the treatment and to determine if additional treatment steps are necessary.

Clinical Issues

The original research validating Dignity therapy was performed within a fairly narrow patient population, namely, adults from Canada and Australia in hospice-care settings with terminal illness (primarily cancer) and a life expectancy of less than 6 months. Further studies have included more variable demographics, different illness types, and lesser illness severity, including multiple studies done with older adult patients in long-term care facilities without a specific terminal illness.

Notably absent from most of these studies are specific inclusion or exclusion criteria for pre-existing psychiatric symptoms or psychological distress, although these were identified in a large proportion of patients as part of the research protocols. Since all of the studies demonstrated effectiveness across multiple domains (dignity, quality of life, depression, anxiety, hopelessness, suffering, etc.) for the whole of the included populations, the research findings suggest Dignity therapy is indicated not just for those persons identified as depressed or anxious, but for all persons as they approach the end-of-life. It should be noted, however, that those persons identified during study enrollment as having significant psychiatric symptoms or psychological distress experienced the most significant positive effect from a course of Dignity therapy.

While Dignity therapy may have widespread application, there are two particular patient groups for whom the effectiveness of the treatment is unknown. Dignity therapy has not been studied in pediatric or adolescent populations, or in patients with significant cognitive impairment. The current approach requires a level of patient engagement and awareness that excludes younger patients and those with cognitive deficits. It is possible Dignity therapy can be adapted to these populations, but additional study will be required.

Ethical Issues

Helping patients resolve inner psychological conflict that is influenced by their terminal condition is a central feature of Dignity therapy. Although acceptance of death is not a primary goal of the treatment, this process often leads to a greater acceptance of human mortality. It is therefore possible to imagine situations in which Dignity therapy could be misused with the purpose of influencing patients as they consider various treatment options to either prolong or foreshorten life. Another risk of Dignity therapy is the potential impact on matters of estate and other legal issues of the generativity document and the audio recording from which it is transcribed. This is why it is essential to obtain the patient’s informed consent for both treatment and audio recording and to review the generativity document with the patient before it is revealed to others.

Research and Evaluation

The research generally demonstrates beneficial effects on psychiatric symptoms and psychological distress in those who are suffering. Additionally, Dignity therapy leads to improvements in quality of life and overall sense of well-being in patients who are not struggling with the psychological impact of their illness. Finally, there is evidence Dignity therapy has a positive impact on caregivers.

Future research considerations likely include expansion of the Dignity therapy approach to nonterminal illness, the study of Dignity therapy in adolescent and young adult populations, and the adaptation of Dignity therapy for persons who are unable to participate in the process as it currently is constructed (e.g. children and those with cognitive impairments).

PSYCHOTHERAPY FOR THE DEAF

History

Psychotherapy for the deaf is not a particular type of psychotherapy but a group of therapies that are adapted to the needs of a special population—people who develop a profound hearing impairment before speech is established. The impairment is defined operationally as one that renders audition useless as a means of intercourse with the hearing world, and it generally involves a hearing loss of 80 dB or more across the speech frequencies, even with the aid of amplification. Its most obvious consequence is to interfere with the ability to learn language by simple auditory imitation and at the normal time and pace of development.

Psychiatric interest in the deaf began with the work of Franz Kallmann, John Rainer, and Kenneth Altshuler in the mid-1950s. Then the country's most prominent geneticist in psychiatry, Kallmann was approached by a branch of what is now the U.S. Department of Health and Human Services for assistance in obtaining case closure among psychiatrically ill deaf clients. Kallmann recruited Rainer and Altshuler and began a series of normative studies and studies of deaf persons with mental illness. The first effort established an outpatient psychiatric clinic, and subsequent projects developed an inpatient unit for statewide services, a vocational rehabilitation liaison and halfway house program, and consultation programs at a school for the deaf. Centers focused on psychiatry and the deaf were subsequently established in Chicago, San Francisco, Washington, and several other cities and countries around the world. At present, a small caucus of psychiatrists interested in working with the profoundly deaf meets regularly at annual meetings of the American Psychiatric Association.

Theoretical Issues

To adapt a given psychotherapy to the needs of persons with profound early deafness demands that the therapist understand what the handicap means to one's life experience and that they take into account the limitations that this experience places on the therapeutic exchange. At the same time, the therapist keeps in mind that people who are deaf also experience the same longings, fears, and conflicts as the general population.

A first guiding principle is that the simple absence of audition has no relevance to innate intelligence potentials; intelligence quotients (IQs) among the deaf follow the same bell-shaped curve as in the hearing. Among the deaf, performance scores are generally higher than verbal scores because of the lag in language development. Deaf children of deaf parents develop vocabulary at the same rate as their hearing peers, albeit in sign language, because their parents are able to communicate with them freely. Occasionally, an enlightened hearing parent learns sign language to provide the only free, two-way accessibility that the handicap allows.

Too often, however, parents are caught up in an educational controversy that has gone on since Alexander Graham Bell and the 1880s, about whether signing discourages the learning of speech. Evidence suggests that an educational approach that embraces all avenues of communication—combining oralism, speech and lipreading instruction, amplification, and sign language—gives the best results. Moreover, the combination supports the exercise of age-appropriate interests and language skills in accordance with their natural developmental timetable. In the absence of a widespread combined educational approach, studies in the 1970s showed that the average deaf adolescent left high school with approximately fourth-grade and fifth-grade English and mathematical skills.

As a result of these findings and the urging of mental health professionals, since the 1970s, there has been a major shift in educational practices from a nearly exclusive emphasis on oralism toward some variety of the combined mode of instruction. Pockets of resistance exist and the availability of cochlear implants at a relatively early age may recast expectations and heat up the issue again.

A factor contributing to the endurance of the controversy is that parents are understandably hurt and saddened when the handicap is recognized, and they wish that it would go away. Motivated partly by denial and partly by the commendable desire that their child’s life be like any other child’s life, parents resist sign language as the visible symbol of the handicap and the harbinger of a separate and sequestered existence. What is lost in the face of these pressures is the recognition that language, generally mediated through audition, brings a tremendous amount of information and experience. Without hearing or the effective alternative of sign language, uncountable hours of speech and lipreading instruction can only provide language late and often incompletely.

Sound also brings nuance and music, and it is the sole distance receptor that scans 360 degrees of the environment at all times, perhaps even in sleep. What may be associated with missing those elements of experience is uncertain.

A final concomitant of deafness, at least in the 90% of those born to hearing parents, is that the parents of deaf children feel confused, uncertain, and, often, guilty. Their own ambivalence, the conflicting advice that they may receive, and the stranger to communication in their midst often makes them overprotective, inconsistent in discipline, and erratic in their hopes and approach.

Techniques

Communication is the sine qua non of psychotherapy. The first requirement with the deaf is to understand what life experiences the handicap entails; the second requirement is to interact freely with patients in order to apprehend their personal and particular experiences and to exchange understanding about the details. By adulthood, virtually every person with early profound deafness is comfortable and fluent in sign language (barring an accompanying organic problem that interferes with language acquisition, regardless of mode). Having the therapist equally versed in manual language would be ideal but is rarely possible, despite the fact that a growing number of psychiatrists and others have learned to sign. The problem is lessened to the extent that the patient's language, speech, and lipreading skills are developed. Often, these skills are insufficient to allow the conduct of meaningful psychotherapy, even with exchanges in writing. In such instances, an interpreter is necessary. Deaf persons often rely on interpreters, and the therapist, rather than the patient, may be the one who is uncomfortable with the third party's presence. One must be alert, however, to the interpreter's tendency to interpret when the therapist wants an exact translation; for example, when the patient's thoughts are to be evaluated, the interpreter's efforts to make sense of them may be counterproductive. Supportive psychotherapy can be conducted with the aid of an interpreter, but substantial insightful work probably demands a subtlety of understanding that is only available with manual skill on the part of the therapist or adept speech and lipreading skills on the part of the patient.

Clinical Issues

Less than profound deafness and deafness after speech is well-established are problems of a different nature and require different adaptations. To the extent that language has been developed before the loss of hearing, and to the extent that the loss is remediable by amplification, the patient's life experiences are less affected, and therapy requires less in the way of modification. However, the loss itself, its cause, and the patient's changed life circumstances because of it are factors that affect the person's problems and the therapy.

A couple of additional points must be kept in mind in counseling the hearing parents of deaf children. Eager for their child to have a normal life, the parents may be impatient with modest results in attempts to remediate the handicap and its effects. Complaints of bad treatment at the hands of many institutions or agencies must be evaluated with that possibility in mind. Also, denial on the part of a family may be brought to an abrupt end with the child's impending adolescence. When parallel play and participation in simple games give way, and the horizons of social interaction loom large, the effects of the handicap become unavoidable, and parents may become depressed.

Research

The educational, developmental, and psychiatric literature on early profound deafness is considerable, but no controlled studies of psychotherapy have been reported. The psychiatric literature, mostly from the 1970s and before, suggests that major mental illness is no more prevalent among the deaf than in the hearing, that impulsive or aggressive behavior may be more common in the presentation of illness among the deaf, and that guilt and psychomotor retardation may be less frequent in cases of depressive disorders in the deaf. Anecdotal reports also indicate that various forms of psychotherapy are as useful in the deaf as in the hearing when modifications in view of the handicap are made.

EXISTENTIAL PSYCHOTHERAPIES

History

Existential psychotherapies are a diverse group of treatments that were inspired by the writings of the existential philosophers and novelists. The therapies emphasize exploring and validating the individual's subjective experience with a minimum of theoretical expectations; fostering greater authenticity in affective experience and decision-making; and confronting the unavoidable and painful limitations of human existence. Defining the essential underlying qualities of these treatments in order to create a unifying framework for practice and research has been difficult. There is no single founder of existential therapy to invoke as a defining authority; the philosophies in which existential therapies are rooted differ greatly and are open to multiple interpretations; and the existential stance tends to be suspicious of authority and of overarching systems of meaning, making it inherently inhospitable to attempts to codify the disparate approaches. Nonetheless, there are several significant elements these theoreticians and clinicians share, and which together yield a reasonably coherent body of theory and practice.

Existential philosophy arose in the late 19th and first half of the 20th centuries as a challenge to traditional religious and philosophic optimism, which held that truth could be attained through reason and that individual values could be grounded in the ever-expanding knowledge of the universe. Existentialism asserted that reason has limits, that the universe was a place of arbitrary facts, and that the irrational elements of the human psyche always limit the potential for self-knowledge and social order. Existentialists were suspicious of abstract truths and placed their trust in the reality of an individual's experience. Within this perspective, existentialists showed great diversity and included religious thinkers such as Martin Buber, Soren Kierkegaard, and Paul Tillich as well as secular and atheistic philosophers such as Martin Heidegger, Friedrich Nietzsche, and Jean Paul Sartre. Some, like Sartre and Albert Camus, expressed their ideas in fictional writing that explored the individual's confrontation with the inherent meaninglessness of life and the responsibility to nonetheless make authentic choices.

Kierkegaard and Nietzsche were aware of the profound anxiety and confusion resulting from feeling uprooted from a grounding morality and faith, while at the same time being cognizant of mankind's potential for acts of destruction. Kierkegaard advocated a religious “leap of faith,” while Nietzsche pronounced that “God is dead,” which if true would leave mankind forced to create new systems of value. As different as they were, both thinkers shared a conviction that a meaningful life requires a profound effort and commitment on the part of the individual.

The clinicians in Europe who created existential psychotherapy felt that Freud did not adequately address the human need to find meaning in an innately meaningless world. They felt that Freud's technique was profoundly helpful, but that his theory of biologic drives was reductive and overly focused on the physical world. The school of Existential Analysis, represented prominently by Ludwig Binswanger and Medard Boss, grew out of the application of Heidegger's philosophy to problems of psychotherapy. Heidegger created the concept of "being in the world" to define what is unique to human existence, specifically, that man is always both a thing in the world and the creator of meaning and knowledge.

Logotherapy is an existential psychotherapy based on the work of psychologist Viktor Frankl, whose ideas were influenced by his experience as an inmate in Nazi concentration camps. He came to believe that while people cannot control what happens to them, they can choose the attitudes with which they face and respond to their circumstances. Frankl felt this to be a fundamentally hopeful stance, and termed it "tragic optimism."

Somewhat later, Rollo May focused the attention of the psychiatric community on existential psychotherapy and generated substantial interest among American psychotherapists; his particular approach blended humanistic psychology, psychodynamic techniques, and an emphasis on the individual's alienation from their own passions and inner life. Irwin Yalom's case studies and his text *Existential Psychotherapy* were similarly influential in sustaining interest in this existential-humanistic tradition among a broad audience of psychotherapy practitioners.

Theoretical Issues

Existential psychotherapies draw on diverse theoretical roots, which include the elaboration of the phenomenologic method, as well as explorations of alienation and anxiety. Phenomenologists attempt to step outside pre-existing conceptions and ideas about the world in order to experience each new object or person directly. The phenomenologic method, therefore, inspired existential psychotherapists to seek a spontaneous empathic encounter with each new patient. Clinicians within this tradition emphasize an engagement with the patient in their unique wholeness. Therapists strive to avoid letting their preconceived ideas about patients and their pathology interfere with their understanding of the patient in the present moment. In their view, a focus on diagnosis, or on ego structure and defenses, is thought to divide the experience of the patient into pieces and miss the opportunity for an encounter with the whole person.

Existential therapists see anxiety and guilt as essential aspects of human existence, and not just responses to specific events or conditions. Anxiety and guilt can be realistic, neurotic, or existential. Realistic anxiety is the signal of actual danger, while neurotic anxiety follows from fantasized, projected, or displaced threats. Existential anxiety is seen as the experience of the dread of nonbeing—which includes death, but also a person's unrealized potential. Realistic guilt is in response to actual crimes or transgression, and neurotic guilt follows from fantasized, generally unconscious, forbidden wishes. Existential guilt means guilt derived from suppressing one's real potential in favor of inauthentic conformity.

Techniques

Although the practice of therapists varies widely, there remain some common points of technical emphasis. First, existential therapists try to find the patient's unique worldview and enter into this experience with the patient. They emphasize the importance of empathy with the totality of the patient—that is, actually "knowing" as opposed to "knowing about" the person. Second, existential therapists attempt to uncover and work through deep anxiety that stems from such universal threats such as aloneness, meaninglessness, and mortality. Third, existential therapists emphasize that all people are subject to inescapable reality-based limits and try to help their patient bear these facts. One such limit is the fact of being who one is and not someone else; another is the impossibility of fully knowing, or being known by, another person. Fourth, existential therapists consider the denial of one’s mortality and the fear of death to be critical determinants of pathology and focus their attention on these issues. Finally, the therapists support patients recognizing their freedom to choose their actions within these limits, and their freedom to create their own meanings for the events of their lives. The existential therapist pays close attention to whether patients’ choices represent their true desires and values of the patient, or whether they are better understood as distractions or narcotics.

Yalom described a patient who presented with compulsive sexual affairs. A dream of a magic bow and arrow was interpreted as representing the use of his phallus to transcend the ordinariness and banality of death and his affairs as providing him with the illusion of immortality. As the patient began to give up the compulsive behavior, it was possible to use therapy to also confront the question “What are people for?” and open up an exploration of existential isolation.

Existential analysis borrowed the technical structure of psychoanalysis (frequent sessions, use of the couch, free association, nondirective exploration), but remains squarely within the frame of the patient’s current experience. Therapists eschew interpretations of unconscious conflict, symbolism, or transference and view the essential pathology as that of being closed off to parts of one’s own experience. Thus, treatment aims to increase patients’ openness to what is disavowed. To this end, some therapists use direct confrontation, paradoxical interventions, and active reframing. The techniques of logotherapy aim to promote the identification of meanings that transcend one’s own self; this is done through “self-distancing,” which seems very similar to developing and observing ego. Such specific maneuvers are meant to be used in the context of a “Socratic dialogue” that attempts to trigger the patient’s own resources of wisdom.

Existential therapy, unlike psychoanalytic approaches, is not primarily developmental in its thinking. The basic concerns of human existence—death, freedom, isolation, and meaning—underlie anxiety for patients of any age. And May, reversing the classical psychoanalytic tendency to find present meanings in autobiographical events, suggested that people's understanding of the past is actually determined by their experience of the present and future. As with many contributions of the existential clinicians, these ideas have since been integrated into the body of mainstream psychodynamic thought.

Indications

Existential psychotherapists claim success with this modality in the same range of conditions as dynamic therapies in general—depression and dysthymia, anxiety disorders, complicated bereavement, trauma, adjustment disorders, and the like. An extended application of existential therapy suggests that much of the depressed affect seen in bipolar patients may be better conceptualized as existential despair; thus, psychiatrists should address core existential issues such as the conflict between hope versus demoralization. It has been reported also that logotherapy may be a useful adjunctive treatment for combat veterans whose difficulties include existential concerns such as a foreshortened future, survivor guilt, and a sense of nihilism. A variety of interventions grounded in existential concerns are reported in the psychosomatic psychiatry literature in recognition of the significance of such issues for patients coping with severe or chronic medical illness. Group existential psychotherapies are increasingly reported on as adjunctive treatment for cancer patients. Brief psychotherapeutic interventions for these and other medically hospitalized patients may help the patient move from a position of defeat and hopelessness to one of resilience, agency, and connection with others. Most recently during the COVID-19 pandemic, existential therapy in the form of existential-humanistic therapy was proposed as when employed as beneficial as part of a disaster response program. The pervasive widespread human suffering during the global crisis provided a unique opportunity to adapt the approaches of this therapy type to the situation.

Ethical Issues, Research, and Evaluation

There has been little systematic study of the indications or outcomes of existential psychotherapies. It may be that the emphasis on the individual's authentic and autonomous creation of meanings is especially challenging to systematize, and this has created an impediment to empirical study.

The existential therapies aim to understand patients according to their unique life situations and to foster their authentic choices without forcing them to fit into a preordained theory of mental structure or development. The inevitable influence of the therapist's subjectivity, however, has made it increasingly difficult to accept the existential assumption that the patient can be perceived uncontaminated by the therapist's theory and technique. Any attempt to restate the patient's narrative, distill central themes, or define affective currents is likely to be influenced by the therapist's theories of mind, conflict, development, and change. The existential psychotherapists have, however, provided a valuable service to the field of psychotherapy through their impassioned attention to the individual patient's experience and choices and their conviction that it is all too easy for therapists to collude in an avoidance of the existential challenges of mortality, meaning, and disconnection. As a result, the priorities, interests, and techniques of the existential therapies are today often integrated into the work of eclectic therapists of different persuasions.

PARADOXICAL AND STRATEGIC THERAPIES

History

According to the principle of paradoxical intention, fighting the symptoms of performance anxiety increases their severity, but invoking them deliberately may yield the opposite. Milton Erickson elaborated those ideas into paradoxical therapy, and the method was further refined and promulgated widely by Jay Haley.

Erickson was a hypnotherapist whose prowess in the use of paradoxical techniques was reputed to be extraordinary. Haley, while working on the double-bind theory of schizophrenia, thought that perhaps there could be therapeutic double binds, as well as pathologic ones. He attended one of Erickson's hypnosis workshops in 1953 and, with John Weakland, began a regular series of visits that endured for 20 years. Haley and his colleagues learned from Erickson and published what they had learned. Others made similar visits or were attracted by the ideas or the stories that they heard, and they added to the reputation and the literature of the growing movement. Haley later developed strategic therapy, essentially a method that applied Ericksonian paradoxical principles, along with other directive techniques, to family therapy.

Interest in Erickson's paradoxical therapy grew over the years and reached its height in the late 1970s and early 1980s. By this time, the method claimed success in treating individuals, couples, and families for a range of problems, including phobias, compulsions, insomnia, schizophrenia, Tourette syndrome, urinary retention, transvestic fetishism, anorexia nervosa, agoraphobia, childhood fecal incontinence, alcohol use disorders, depressive disorders, fire setting, asthma, constipation, truancy, and a variety of sexual performance and relationship difficulties.

Erickson died in 1980, and a foundation was formed to continue the work that he started. He was one of the first to show that brief therapy can be an effective and primary approach. He advocated that therapists should not be just passive listeners but active interventionists and guides; focus on symptoms rather than on dynamics and on action rather than on understanding; prescribe, as part of the treatment, the very resistances and symptoms the patient brought to it; and use indirection and metaphor to achieve the goals of therapy.

Theoretical Issues

The overarching idea in paradoxical therapy is to embrace and amplify the notion of paradoxical intention and thereby reduce performance anxiety or destabilize a fixed internal or social system. The paradox and the destabilization may be effective in themselves or may cause secondary reverberations, such as increased self-awareness and the evocation of new behavior or emotions that lead to helpful change.

Paradoxical therapy is relatively atheoretical about cause and is not much concerned with it. Behavioral prescription is emphasized, with the idea that altering internal and social feedback loops can alter the system. Symptoms that block change can be made vehicles for it if they are embraced and are prescribed or altered in their timing or the function that they serve. Instructing an impotent patient under no circumstances to try intercourse, for example, may dissipate anxiety and may allow him to do so; telling an insomniac to get up and polish the floors for the rest of the night may similarly invite sleep. Paradoxical therapy also recommends defining the problem clearly, establishing clear goals for the treatment, offering a plan, and enlisting the collaboration of the patient or the family in the suggested strategy.

Techniques

Some of the principles of paradoxical therapy can be directly translated into techniques: Using patients' tendencies toward compliance or defiance—for example, by overtly expecting a relapse and being surprised about improvement; reframing (i.e., relabeling behavior and emotions so that their experience is changed from negative to positive or vice versa); using metaphor, indirect directives, and cryptic paradox; and distancing patients from their symptoms or challenging patients by humor or sarcasm. These techniques and the extent of the therapist activity encouraged in paradoxical therapy are illustrated in the following cases of Erickson and his followers.

Making the Activity Deliberate

An alcoholic wife secretly hides her bottles, and the husband searches for them with increasing anger. With the relationship at a breaking point, Erickson prescribed that she deliberately hide the bottles and that he searched for them. If he found one, he could keep it. If it eluded him, she would be allowed to drink in peace. He also suggested that they buy a trailer and go fishing, an activity that they both disliked. The reported results were that the drinking abated, and the couple later happily used the trailer to go camping. The deliberate game apparently subverted the purpose of the prior secretive activity, and forcing the couple to share a disliked recreation brought them into close contact (with alcohol unavailable) and pressured them to communicate, enjoy each other, and make pleasant vacation plans.

Prescribing the Symptom

Instead of telling her husband what she felt, a woman would unconsciously start a fight with him when she wanted affection. She was instructed to start a fight on purpose and intensely the next time that she felt empty or unloved. Doing so, she noted the connection and realized that, under attack, her partner could not possibly respond with support or love. She also felt ashamed, and the experience helped her acknowledge and express her wishes more directly.

Aversive Conditioning

A woman consulted Erickson about her schizophrenic son, who did nothing but sit around the house and moan. Erickson recommended taking him for a drive; she was to drop him off, drive 3 miles farther, then sit and read while she waited for him. After doing that a few times, the son volunteered to walk without being dropped off, and later asked if he could go bowling while she read, because he liked bowling more than walking.

A second patient told Erickson that he could drive only on certain streets and that he would become nauseated, vomit, and faint if he tried to drive outside the city limits. Erickson required him to drive to the edge of town in his best clothes, to park and to lie down in a sandy ditch at the roadside, and to lie there for at least 15 minutes. He was then to drive one or two car lengths farther, repeat the procedure, and do that until he could drive from one telephone pole to the next. He was also told to stop immediately and to get into the ditch at the first sign of any symptoms. The patient reported: “The more I did it, the madder I got. So I just quit and began to enjoy driving.” Thirteen years later, Erickson reported, the former patient still had no trouble driving. Anger and defiance allegedly brought about a cure.

Destabilization

A boy who was 10 years of age engaged in compulsive, public masturbation; the compulsion had endured for years, despite many treatment efforts. Taking the behavior seriously, the therapist asked the boy to keep a list of when he masturbated and to note when he found it most pleasurable. With the idea that he should do it when he found it most pleasurable (on weekends), he was assigned to do it eight times on Sunday but not on Monday. When the boy failed and masturbated on Monday, the disappointed therapist prescribed that he do it 12 times the next Sunday. When the boy was not up to the task, the therapist suggested that, as punishment, the boy masturbate once daily in public and eight times on Sunday. Masturbation was also prescribed as a punishment for thumb-sucking and as an alternative to homework. Gradually, the boy stopped and even enjoyed his homework.

Examples reported as successful include telling a woman with headaches, which were symptomatic of family relationship problems, that the headaches were probably incurable and that therapy should aim at helping her live with them; asking the parents of a child with severe and long-standing encopresis whether they could tolerate being normal and being better parents than their mothers (the father then threatened the child with laxatives, and the problem disappeared); and combining reframing and negative emphasis or challenge by telling a family with a school-phobic child: “Maybe it’s better she stay at home. It keeps the family from being normal, and being normal may not be tolerable.” (The therapist reported that the family did not get insulted and leave but changed spontaneously to prove him wrong.)

Most recently paradoxical and strategic therapies appear in the treatment literature of anxiety-related disorder and binge eating disorders respectively in the form of *paradoxical timetable therapy* and brief strategic therapy. Paradoxical timetable therapy–also referred to as PTC (Paradox + Timetable = Cure)–pairs the previously described paradoxical prescription or disorder symptom with a timetable. The timetable dictates that the planned paradox task should be performed within the limits prescribed or not be performed at all. The components of the paradoxical task and timetable are linked and are prescribed together. A form of strategic therapy called brief strategic therapy recently compared favorably to cognitive-behavioral therapy (CBT) in a 2018 randomized clinical trial in women with binge eating disorder.

Clinical and Ethical Issues

Paradoxical therapy is clearly not a psychotherapy itself but a set of techniques potentially usable in therapy. Making explicit the assumptions that underlie the techniques reveals their range of usefulness and their pitfalls. In addition to the idea that prescribing or restraining the symptoms can reduce performance anxiety, the assumptions are that symptoms always reflect a problem in a social context or relationships, that they are somehow and ultimately volitional, and that they, therefore, respond to being emphasized or manipulated.

Such assumptions may be acceptable in some cases, in which emphasizing the symptom is similar to flooding, an accepted technique of behavioral therapy, and in relationship problems, in which prescribing the symptom or altering the context of its appearance may help the participants recognize their inappropriate emotions or behavior patterns.

Using paradoxical therapy in biologically based illnesses is far more of a problem. Expecting a patient with Tourette syndrome to respond to such techniques alone suggests a volitional element greater than what is present. The basic symptoms of schizophrenia are unlikely to respond to paradoxical therapy, and, although behaviorally conditioned symptoms may respond, the case histories show the difficulty of distinguishing what is conditioned from what is biologically based. The literature about paradoxical therapy is primarily anecdotal, and the rare research study is less enthusiastic than authors who claim widespread success.

Most importantly, many of the case histories in the literature include manipulations of the patient or the families that may cover or even invite aggressive acting out by the therapist, and some case histories suggest a simplistic conception of illness in which punishment is curative. When the operating field permits behavior that is less than open and straightforward, therapists need even more vigilance and self-awareness than when procedural safeguards protect the patient.

TRANSACTIONAL PSYCHOTHERAPY

History

Transactional analysis or psychotherapy was the creation of Eric Berne, a psychiatrist with psychoanalytic training who parted with the Freudian movement to focus on social interaction as the unit of analysis. *Transactional Analysis in Psychotherapy* was published in 1961, and *Games People Play* was published in 1964; *Games People Play* became a record bestseller, with sales of more than 2.5 million copies. Berne's thesis was that, in social interactions, people seek intimacy, comfort, or recognition, which they get through a series of mutual exchanges or *strokes*. If the ego states of the participants are adult and appropriate, the transaction is harmonious and realistic. Ulterior motives, sponsored by other than an adult ego state, cause discomfort in the transaction, and transactions habitually colored by such motives may become ritualized and predictable.

Written with wit and humor, Berne's books inspired considerable controversy. Some accused him of facile psychologizing better left to cocktail parties rather than true contributions to psychotherapeutic theory and practice. The fact that the movement has endured inspiring a journal, an accrediting body, and an occasional article in the psychiatric literature suggests that a number of its ideas have merit.

Theoretical Issues

Transactional analysis rests first on the principle that human beings have a hunger for stimulation and contact in the infantile style. When such strivings are blocked, efforts are made to regain intimate contact. Compromises and sublimations are made in that pursuit so that a hint of the original intimacy may suffice, and the stimulus hunger may be satisfied symbolically. Thus, the original physical strokes of caress and comfort are replaceable by metaphorical stroking by which the person is recognized, flattered, or supported and even by acknowledgments that may appear to bring discomfort. A stroke is a fundamental unit of interaction. An exchange of strokes is a *transaction*, the unit of social intercourse; a stimulus from one person produces a corresponding response in another person. Because social stroking satisfies the original infantile need, it follows that any social intercourse is better than none.

In addition, social interactions fill time, provide structure, and help avoid boredom. Interactional programs, all of which contain stroking opportunities, maybe *material* (dealing with reality issues); *social* (the norms and manners of a social order); or *individual* (incidents sneaked into the other programs to get extra, personalized strokes). The sequences of a person's program are developed in line with that person's *life script*, a set of overall goals and patterns, and these sequences are the individual *games*. The games range from harmless and pleasantly satisfying to dangerous and destructive. What develops in a game is assumed to be the purpose of the game, and what eventuates as its end result are the strokes that constitute the aim and the satisfaction.

To understand and diagram the interactions in a game, the theory posits three ego states, the parent, the child, and the adult. In the parental ego state, the person reacts as the parent or as the parent would have directed. In the child state, one may be adaptively compliant or whiny and withdrawing or, as the natural child, rebellious, creative, or impulsive. The adult state appraises reality and acts in tune with it; the adult also moderates between the parent and the child. All three states are present and potentially active at all times.

When the adult in one person addresses the adult in another and is responded to in adult terms, the stimulus and the response are complementary, and the transaction is smooth. The same is true if the child addresses the parent, and the parent responds appropriately—for example when a child (to parent) demands "Water!" and the parent (to child) responds "Here, baby." Such transactions are normal, two-stroke, harmonious social interactions.

However, if one partner in a couple asks, "Where are my cuff links?" and the other responds, "You're always losing things!" the lines of communication are crossed; the query may be adult to adult, but the response is parent to child, and the situation gets sticky. A comparable child-to-parent cross in response to an adult-to-adult stimulus may occur when one notes to a person driving, "You missed the turnoff," and the driver responds, "You're always criticizing me!"

Such ulterior transactions are the basis of games, which are ongoing series of complementary ulterior transactions progressing to a predictable, defined outcome that satisfies some aspect of the person's life-script strivings. On the surface, the interactions may appear plausible, but their recurrent nature, concealed snares, and predictable payoffs define them as parts of a neurotic game.

The test of a game is that a respondent's refusal to play causes increasingly intense efforts to force its continuation; continued refusal may elicit despair or bewilderment. The cure for a game is recognition of its presence, its rules of play, and its motivation. A cure is signified when the initiator can note with a laugh, "There I go again," and can desist volitionally.

Techniques

Because games generally involve at least two parties, they reveal themselves and are best treated in couples or groups. Transactional analysis has therefore had its most widespread application in couples and group therapies, although it has also been seen as useful in individual treatment. The techniques aim to illustrate the game, dissect the transactions, and frustrate players in their pursuit of strokes. In so doing, the players may learn a considerable amount about their life scripts and how they evolved, although that is not necessarily the goal of treatment.

Games are named for convenience of recognition, with the name usually connoting something about the ends sought. Thus, in “if it weren’t for you,” a woman (usually) picks a domineering man to marry, who then forbids her to do things she allegedly wants. She blames him for her unhappiness, they quarrel, their sex life deteriorates, he gives her gifts to make up, and the cycle goes on. In reality, the woman fears the activities for which she says she longs, has a sexual inhibition (which the husband may share), and is pleased by the opportunity to prove that men are tyrannical. In the game of “schlemiel,” a man (usually) messes up or breaks things to be castigated and then forgiven.

Other life games include “debtor,” “alcoholic,” and “see what you made me do.” The aim of each is to obtain one or several ulterior satisfactions without accepting conscious responsibility. In “alcoholic,” the players may include the alcoholic, a persecutor (usually a spouse), a rescuer (often, a health professional), a patsy (who gives money or sympathy), and, sometimes, a connection (who may tempt the alcoholic to drink). Each player may also be in the game for idiosyncratic reasons, games, or strokes. For the alcoholic person, the payoff is the hangover and its chance to indulge in psychological torment and suffering; for the rescuer, it is an opportunity to play “I’m only trying to help you”; for the persecutor, it is an opportunity to play “Look what you’ve done to me.” The antithesis, the goal of treatment, is to have each one recognize the role played and to frustrate the primary player (and the other persons addressed as well) by giving up the role assigned.

Groups may be randomly selected or homogeneous. In treatment sessions, the theory of transactional analysis is taught assiduously; with blackboard illustrations, the aims of treatment are outlined in operational terms; and contracts are made about individual large goals and are redrawn for small, particular steps in the course of therapy. The therapist’s role is active and confrontational but egalitarian.

Research and Ethical Issues

Interest in transactional analysis peaked long before the advent of criterion-based diagnosis and controlled clinical trials, and little research on it can stand up to today's critical scrutiny. It was claimed to be effective for persons with neuroses, psychoses, personality disorders, sexual disorders, and mental delay; for individuals and groups; and as a means of self-help for young children and teenagers. What seems sensible includes an emphasis that is typical of later cognitive-based treatments—namely, identifying elements of behavior that reflect neurotic problems so that their maladaptation can be acknowledged, their driving emotions and underlying assumptions can be recognized, and the pattern of behavior can be altered. Psychoanalytic therapies share the same aims but include the additional goal of genetic uncovering. Transactional analysis' emphases on confrontation, contracts, homework, active diagramming of conflict, and teaching about behavior are also more like cognitive therapies.

Most recently, transactional analysis interventions were studied for use in outpatient addiction patients participating in a methadone maintenance program. The results suggested that transactional analysis significantly influenced interpersonal problems in a positive way and the client's readiness for treatment. Transactional analysis was also recently used in a 2020 clinical trial to investigate its effect on imprisoned women's self-esteem. The authors reported that transactional analysis in a group-training format improved the self-esteem of the study's intervention group.

The weakness in the system and the potential it has for abuse in work with patients lie in its theoretical assumptions. *Neurotic behavior*, by definition, is not realistically aimed and is often maladaptive. To assume that all bad consequences are deliberately sought is incorrect and unfair. To the extent that such a view is held, important motivations may be missed, and therapy may be tainted by a tendency to level accusations. If this position is avoided, and complaints are not trivialized by being seen routinely as goals, the dissection of social transactions and the focus on personal responsibility offered by transactional therapy conform to what is generally considered to be of therapeutic value.

GUIDED IMAGERY THERAPY

History

The use of guided imagery or guided affective imagery as a form of therapy was systematized by Hanscarl Leuner in 1954. Leuner considered himself to be an analyst, and he was influenced by Carl Gustav Jung's interest in active imagination and Johannes Heinrich Schultze's autogenic training. Having confronted himself with disturbing emotions and images during his self-analysis, Jung thought that spontaneous, inner, dream-like images were translations of unconscious emotions and that observing and interpreting them could bring unconscious contents to consciousness. He called the technique *active imagination*, and he made it a part of what clinicians following his analytical tradition would use. Schultze framed the principles of *autogenic training* as a form of self-hypnosis and relaxation training, a mind- and organ-strengthening system based on expansive powers alleged to reside in the mind.

Leuner found that resting comfortably with the eyes closed for some time led to the appearance of images and that an exploration of the images with a guide was a pleasant and beneficial experience. He induced the images actively to shorten the interval before their appearance, added analytical insight to the guided exploration, and so framed *guided affective imagery*.

The ideas of guided affective imagery caught on in the United States and other Western countries, and the treatment was reported to be useful in chronic neuroses, personality disorders, depression, phobias, obsessive-compulsive disorders (OCDs), sexual disorders, hyperhidrosis, substance abuse, anorexia nervosa, ulcerative colitis, and hives. The niche it filled in organized psychiatry, however, was never more than a small one. The ancillary disciplines of medicine, particularly nursing, embraced it with more enthusiasm than did psychiatry, and positive claims have been made regarding its use in postsurgical pain control, in general adaptation to illness, and, particularly, in coping and dealing with incurable, progressive disease.

Theoretical Issues

Guided affective imagery attempts systematically and psychotherapeutically to use the human ability to create images. It holds that so-called catathymic images, images appearing before one's closed eyes as if on a screen and ultimately coming together and taking on movement to tell a story, represent a person's interpersonal world in autosymbolic imaginal form. The images include color, plasticity, and action, and the more one is absorbed in them, the richer, more colorful, and more subtle they become. The symbols and the stories project the person's conflicts. Psychoanalytically informed guidance as the person recounts the experience can materially enhance conflict resolution by manipulating the symbols or assisting the story toward a successful outcome. The crucial idea is that influencing the symbols and the story influences the conflict.

Techniques

The patient is placed in a comfortable chair and is taught to practice relaxation, with suggestions of warmth, heaviness, and regular breathing. The patient is then invited to imagine a harmless, pleasant image, such as a flower, and is asked to describe it in detail, to look in its center, or to touch or smell it. Involvement in the images heightens the relaxation and vice versa; the patient soon becomes able to induce a readiness for imagery quickly. The hypnoidal state is ended with a command, such as “Clench your fists, bend your arms, open your eyes, and you are back, awake and refreshed.”

When such learning is well-established, the therapist relaxes the patient once more and suggests an image to be used in the therapy—for example, a meadow. The patient is told to nod when it comes into view and then to describe the scene.

Specific images used in the therapy are classified as basic, intermediate, or advanced, according to the level of conflict that they are likely to involve and the skill and training required by the therapist. Basic images include a meadow, a brook, a mountain, a house, and a wood; intermediate images include a car ride and a rosebush (representing women’s and men’s sexual feelings toward the opposite sex); and advanced images are such things as a swamp, a volcano, and a lion.

Descriptions of a meadow or a brook (the patient may see a brook in a meadow spontaneously or may be invited to imagine one) reflect current psychological issues. The meadow may be rich and green or scorched to stubble; the brook may be cool and inviting or a raging torrent. The therapist follows and guides with questions, such as “Can you describe that?” “What do you think of it?” or “What do you see as you follow the brook?”

The therapist’s understanding is informed by knowledge and theory and determines whether the therapist interprets patients or leads them to helpful activity. The therapist may suggest, for example, that a darkening scene reflects the patient’s inner mood or that too perfect a picture shows an effort to mask distress. If patients encounter a fence in the meadow or steep rocks as they follow the brook, the therapist may suggest that the patient climb over or find a way around the obstacle.

A young male patient saw a mountain from which rose another mountain and an observation tower. The young man had problems with authority and would stutter in the presence of his father or his boss. The therapist interpreted the scene to himself to mean that the young man felt watched and that too much was expected of him. The therapist guided him to ascend the tower, which the patient did successfully but amid images of fearful wandering and threats of thunder; the symptoms resolved.

The suggested motifs represent different aspects of the patient’s psychology: The house and the edge of the wood represent one’s own person; the brook represents vitality and energy; the mountain and its panorama represent the patient’s view of the past, present, and future. The therapist looks to see whether supplies and a warm fire are in the house, what comes out of the wood, and whether the brook is cool and refreshing or an icy harbor of dangerous creatures. The management principle is to resolve the conflict symbolically.

One young student got over a block about taking an oral examination by being led in fantasy to feed a dangerous animal (the professor) coming out of a wood; those fearful of their own impulses might do well to approach such animals, to pat them, and to make friends with them. The guide conducts the exploration, using the symbols as a protective disguise, with empathy and warmth, positive feedback, praise, and, at times, direct suggestion.

Clinical and Ethical Issues

Guided affective imagery combines hypnotherapy with certain psychoanalytic ideas. Its major assumption is that manipulation of the symbols of conflict can resolve the conflict, a tenet that is open to question. The supportive relationship, the sense of involvement and effectuating something through the process, and the direct suggestions must contribute considerably to whatever effectiveness the technique has.

Interest in guided image therapy as a formal, psychoanalytically-based psychotherapy has waned. Truncated and modified, however, guided imagery has gained a following in pain control, holistic medicine, and the treatment of incurable illness. Most recently, guided imagery is described in treatment of trauma-related mental health issues. Zehetmair et al. used qualitative semistructured interviews to assess the use of guided imagery and stabilizing techniques in the treatment of traumatized male refugees. The most common disorders in this population include posttraumatic stress disorder (PTSD), anxiety disorders, and depression. The paper suggested that stabilizing techniques and guided imagery are an appropriate intervention in this population. In pain control, it has been reported to be useful as a relaxation and distraction technique, with additional suggestions of finding a second and then several seconds of decreased pain, and so extending the period of pain relief. A recent meta-analysis of randomized controlled trials studied guided imagery's efficacy as an adjunct treatment for fibromyalgia when paired with hypnosis and argued that there is efficacy and tolerability in guided imagery/hypnosis and the combination of hypnosis with cognitive-behavioral therapy in reducing key symptoms of fibromyalgia. The appeal of holistic medicine is to the limitless powers of the mind, a hopeful approach that has found favor in the nursing of patients with cancer and other terminal illnesses.

TRANSPERSONAL PSYCHOTHERAPY

History

Transpersonal psychology grew out of the humanistic movement of the 1960s. Defining itself as “the personal, plus,” it means to include all facets of personal psychology and psychiatry *plus* deeper or higher, transcendent experiences. It thus embraces a range of religious and spiritual experiences and parapsychological phenomena, such as conversion; crises of faith; ecstasy, peak, and near-death experiences; spiritual mystical experiences; and past-life experiences. Although the term *transpersonal* was first used by William James and later was expanded on by Jung, it was writers such as Abraham Maslow, Robert Assagioli, Ken Wilber, and Stanislav Grof who generated its theoretical underpinnings. Each of these authors developed theories of needs and motivations or of stages of development that encompassed and were roughly modeled on traditional developmental theory but added the extra dimensions required by the transpersonal emphasis. The *Journal of Transpersonal Psychology* started publication in 1969 and continues to exist. The first Council Grove Conference, sponsored by the Menninger Clinic that same year, focused on topics such as meditation, biofeedback, yoga, and psychedelic agents. Since then, several transpersonally oriented institutions were founded, mostly in California, and, for the past several years, approximately 15 dissertations dealing with transpersonal experiences are written yearly. The *Diagnostic and Statistical Manual of Mental Disorders*, Fifth edition Text Revision (DSM-5-TR) includes religious and spiritual problems under Other Conditions That May Be a Focus of Clinical Attention.

Theoretical Issues

In the free-for-all of theory building, several authors have contributed to transpersonal psychology and psychiatry, but no single contribution dominates. Thus, the theory is an overlapping pastiche rather than a coherent whole: Maslow studied a series of self-actualized persons and found them similar to traditional descriptions of enlightened persons. He concluded that human beings have a biologically based instinctive nature that is fulfilled in spiritual self-actualization, and he and Grof coined the term and founded the field of *transpersonal psychology*. Assagioli suggested that each person has many subpersonalities and that a spiritual psychosynthesis is needed to unleash the transforming and creative energies of a *superconscious*, which also exists. Wilber proposes prepersonal, personal, and transpersonal developmental stages; disorders of the prepersonal developmental level result in infantile psychoses or narcissistic or borderline personality disorders, disturbances at the personal level give rise to neuroses, and spiritual emergencies, kundalini crises, and dark nights of the soul rest on disturbances at the transpersonal level.

Grof, during 17 years as a scholar in residence at the Esalen Institute in California, experimented widely with psychedelics in psychotherapy and proposed that, in psychedelic and other altered states of consciousness, patients reexperience the trauma of their birth. He suggested the existence of four basic perinatal matrices (BPMs) representing experiences from the entry into the intrauterine state (BPM-I), entry to the birth experience (BPM-II), passage toward a light with intense anxiety and near-death experiences (BPM-III), and culmination in total surrender, a sense of overwhelming and unending love, light, expansion, and forgiveness (BPM-IV). After experiencing or reexperiencing BPM-IV in adulthood, further transpersonal expansions may occur in two domains: Expansion of ego boundaries in ordinary real terms and extensions beyond the usual framework. The first includes a sense of identification with the earth and all of life in the universe, and the second includes spiritual communication, out-of-body experiences, and telepathic, precognitive, and other paranormal experiences. These experiences were often seen as psychotic by traditional psychologists and therefore constitute spiritual emergencies that require special help. Grof urged mental health professionals to learn to distinguish these nonordinary transpersonal states from mental illness.

Techniques

Technique in transpersonal psychotherapy is an amalgam of the generally accepted and the esoteric, depending on the worldview of the practitioner and the issues being addressed. The range is from the psychodynamic and behavioral, through meditative approaches or the recommendation of spiritual texts, to rapid-breathing techniques, guided imagery, and the use of psychedelic drugs to promote altered states of consciousness. The teachings are variously applied to a similarly wide range of issues. *Defensive spirituality*, in which religious or spiritual beliefs are used to deny feelings or conflicts, or *offensive spirituality*, in which there is an obnoxious, self-reassuring, but intrusive stance of "having the one true path," may be amenable to dynamic techniques. Similarly, loss of faith or a conversion experience may be dealt with by a sensitive therapist to determine what factors are at work in the patient's life and whether the conversion, for example, is a legitimate search for truth and the comfort of faith or a magical denial doomed to fail. Expressive explorations of hallucinations or delusions or of a psychotic religious experience after the psychosis is controlled are recommended by some as a means of finding personal insights.

Transpersonal psychotherapists also try to distinguish genuine religious experiences from psychoses, the latter being often more terrifying, accompanied by deterioration in other areas of life, and containing delusions of being addressed by specific religious figures. The boundaries, however, are hard to define in *spiritual emergencies*, when the individual feels dazzled by truths too powerful to grasp, or in *kundalini awakenings*, states in which energies expand upward from the base of the spine and explosively alter consciousness. Supportive therapy is recommended in these cases, along with education about the experience. Meditation and the use of mantras, such as "May I be happy" and "May I be peaceful," are also reported to be useful if the patient is having anxiety over a crisis of faith. Near-death experiences may also lead to a transformed view of the world, and support is recommended along with help in "those parts of the ego that may have died."

The traditional psychotherapist and psychiatrist may have difficulty in areas of transpersonal psychiatry that aim to help individuals recognize and realize a presumed "ultimate state that transcends the usual state of functioning, a state of mystical illumination or cosmic unity." Some of these include experiences such as channeling, encounters with unidentified flying objects (UFOs), possession states, and past-life experiences. In these areas, the transpersonal literature is ambiguous, on the one hand, cautiously attesting that such experiences are meaningful to the person and embody content relevant to the individual's personality, and, on the other hand, accepting them as normative spiritual experiences, an ultimate state of mystic illumination or cosmic unity that is reflective of a reality beyond the norm.

Research and Evaluation

Research on the paranormal, although always of interest because of its supernatural allure, has failed to confirm the legitimacy of past-life experiences, precognition, psychokinesis, and similar phenomena. Also, because transpersonal psychotherapy is not a single discipline, there have been no systematic studies of its efficacy. The field emphasizes the personally meaningful and possibly transformative effects of things such as conversions or near-death experiences; the importance of religion and mysticism in the lives of many; the difficulty of teasing apart the boundaries between psychiatric illness, personal mystique, and cultural differences in expected experience; and the need for sensitivity in working with patients whose life and experiences are different from one's own.

FURTHER READINGS

Ali Besharat M, Naghipoor M. Paradox therapy for the treatment of social anxiety disorder: a case study. *J Syst Integr Neurosci*. 2019;6:1–5.

Altshuler KZ. Will the psychotherapies yield differential results: a look at assumptions in therapy trials. *Am J Psychother*. 1989;43:310–320.

Boorstein S. *Clinical Studies in Transpersonal Psychotherapy*. State University of New York Press; 1997.

Bozarth JD, Zimring FM, Tausch R. Client centered psychotherapy: the evolution of a revolution. In: Cain DJ, ed: *Humanistic Psychotherapies: Handbook of Research and Practice*. American Psychological Association; 2002.

Chochinov HM, Hack T, Hassard T, Kristjanson LJ, McClement S, Harlos M. Dignity therapy: a novel psychotherapeutic intervention for patients near the end of life. *J Clin Oncol*. 2005;20:5520–5525.

Cooper M. *Existential Therapies*. Sage Publications; 2003.

Etkin A, Pittenger C, Polan J, Kandel ER. Toward a neurobiology of psychotherapy: basic science and clinical applications. *J Neuropsychiatry Clin Neurosci*. 2005;17:145–158

Gill IJ, Fox JR. A qualitative meta-synthesis on the experience of psychotherapy for deaf and hard-of-hearing people. *Mental Health, Religion & Culture*.

2012;15:637–651.

Greyson B. The near-death experience as a transpersonal crisis. In: Battista JR, Chinen AB, Scotton BW, eds. *Textbook of Transpersonal Psychiatry and Psychology*. Basic Books; 1996.

Haley J. *The Uncommon Therapy*. 2nd ed. Norton; 1986.

Hoffman L. Existential–humanistic therapy and disaster response: lessons from the COVID-19 pandemic. *J Humanist Psychol*. 2021;61(1):33–54.

Jackson JB, Pietrabissa G, Rossi A, Manzoni GM, Castelnovuou G. Brief strategic therapy and cognitive behavioral therapy for women with binge eating disorder and comorbid obesity: a randomized clinical trial one-year follow-up. *J Consult Clin Psychol*. 2018;86(8):688–701.

Jones L, River EM. Current uses of imagery in cognitive and behavioral therapies. In: Vande Creek L, Knapp S, Jackson TL, eds. *Innovations in Clinical Practice: A Source Book*. Professional Resource Press; 1997.

Khalili E, Kheirabadi G, Khodadadi F. Effectiveness of transactional analysis intervention on treatment acceptance, primary maladaptive schema, and interpersonal problems in patients under methadone maintenance treatment. *J Subst Use*. 2021;27(3):1–5.

Knutson D, Koch JM. Person-centered therapy as applied to work with transgender and gender diverse clients. *J Humanist Psychol*. Advance online publication. 2022;61:104–122.

Lytle LR, Lewis JW. In: Glickman NS, Harvey MA, eds. *Culturally Affirmative Psychotherapy with Deaf Persons*. Lawrence Erlbaum Associates; 1996.

Markman KD, Proulx T, Lindberg MJ, eds. Psychotherapy and the restoration of meaning: existential philosophy in clinical practice. American Psychological Association; 2013.

May R. *Existence: A New Dimension in Psychiatry and Psychology*. Simon and Schuster; 1958.

Meadow KP. *Deafness and Child Development*. University of California Press; 1980.

Perez VB, Swerdlow NR, Braff DL, Naatanen R, Ligh GH. Biomarkers in schizophrenia: using biomarkers to inform diagnosis, guide treatments, and track response to interventions in psychotic illnesses. *Biomark Med*. 2014;8:9–14.

Rainer JD, Altshuler KZ, Kallman FJ, eds. *Family and Mental Health Problems in a Deaf Population*. Springer; 1963.

Ricciardi JN, Bouchard SN, Luiselli JK, Dould T. Integrated behavioral intervention and person-centered therapy within community-based treatment of an adult with acquired brain injury. *Sage Journal*. 2019;19(2):133–144.

Rogers CR. *Client-Centered Psychotherapy*. Houghton Mifflin; 1951.

Schultz JH, Luthe W. *Autogenic Training: A Psychophysiologistic Approach in Psychotherapy*. Grune and Stratton; 1959.

Shmukler D. Reflections on transactional analysis in the context of contemporary relational approaches. *Trans Anal J*. 2001;31:94–102.
Singer JL. *Imagery in Psychotherapy*. American Psychological Association; 2005.
Southwick SM, Gilmartin R, Mcdonough P, Morrissey P. Logotherapy as an adjunctive treatment for chronic combat-related PTSD: a meaning-based intervention. *Am J Psychother*. 2006;60:161–174.
Summerton O. The development of game analysis. *Trans Anal J*. 2000;30:207.
Teasell RW, Shapiro AP. Strategic-behavioral intervention in the treatment of chronic nonorganic motor disorders. *Am J Physiol Med Rehab*. 1994;73:44.
Torkaman M, Farokhzadian J, Miri S, Pouraboli B. Correction to: the effect of transactional analysis on the self-esteem of imprisoned women: a clinical trial. *BMC Psychol*. 2020;8:14.
Utay J, Miller M. Guided imagery as an effective therapeutic technique: a brief review of its history and efficacy research. *J Instruct Psychol*. 2006;33:40.
Weeks GR, L’Abate L. *Paradoxical Psychotherapy: Theory and Practice with Individuals, Couples, and Families*. Brunner/Mazel; 1984.
Yalom ID. *Existential Psychotherapy*. Basic Books; 1980.
Zech N, Hansen E, Bernardy K, Häuser W. Efficacy, acceptability and safety of guided imagery/hypnosis in fibromyalgia—a systematic review and meta-analysis of randomized controlled trials. *Eur J Pain*. 2017;21(2):217–227.
Zehetmair C, Tegeler I, Kaufmann C, et al. Stabilizing techniques and guided imagery for traumatized male refugees in a German state registration and reception center: a qualitative study on a psychotherapeutic group intervention. *J Clin Med*. 2019;8(6):894.

32.24 Evaluating Psychotherapy: Research Methods, Findings, and Future Directions

LINDSAY E. HOLLY, PH.D., MEGAN L. PETRIK, PH.D., AND STEPHEN M. SAUNDERS, PH.D.

PSYCHOTHERAPY RESEARCH

The principal goals of psychotherapy research are (1) to understand what the effective components of psychotherapy are and how they work; (2) to determine how patient, therapist, family, and social factors influence the outcome; (3) to improve the effectiveness of psychotherapeutic interventions; (4) to guide the development of new therapeutic techniques and evaluate their effectiveness; (5) and to inform public policies that increase the availability of quality mental health care. Past controversies regarding the ability to examine scientifically various psychotherapeutic techniques have largely been settled.

Valid and reliable methods for measuring therapeutic events and their effects exist including: (1) intensive analyses of patient and therapist variables; (2) in-depth assessment of therapeutic processes; (3) and implementation of outcome measures that assess general distress, symptoms related to specific disorders, and functional impairment. Controlled clinical trials comparing replicable, distinct psychotherapeutic interventions are normative in research, as are sophisticated analytic methodologies, such as growth curve analysis, time-series panel analysis, and structural equation modeling. Most significantly, various methods of meta-analysis and systematic reviews—techniques that combine results across different studies to evaluate the effectiveness of particular treatments for specific patients and problems—have been applied to psychotherapy research.

This section begins by describing research on the characteristics of psychotherapy consumers. Next, definitions of efficacy and effectiveness research are provided, and empirical evidence supporting widely used psychotherapy models is presented. The importance of the therapeutic alliance is reviewed, and research on the impact of patient and therapist variables on treatment outcomes is discussed. Finally, emerging themes and future directions in psychotherapy research are considered.

CONSUMERS OF PSYCHOTHERAPY

The effectiveness of psychotherapy must be understood within the context of who receives help and how they obtain it. Approximately 20% of the U.S. population meets the criteria for a psychiatric disorder in any given year, but most do not seek formal mental health care. A recent large-scale national survey of mental health service use in the last year found that women are more likely to engage in psychotherapy than men; Individuals identifying as non-Hispanic white are most likely to engage in psychotherapy compared to other ethnic/racial groups, and individuals aged 18 to 44 are more likely to engage in psychotherapy than older adults.

In addition, the likelihood of obtaining treatment increases with emotional distress and functional impairment. Accordingly, those with comorbid mental illnesses or with severely disabling disorders, such as schizophrenia, are the most likely to seek professional help. In contrast, a lack of insurance and concerns about how to pay for services decrease the likelihood of seeking treatment.

However, the primary barriers to mental health care are psychological, including the stigma related to having a mental illness. Negative attitudes toward admitting that one has a psychiatric problem and toward seeking care from a specialty mental health provider have actually increased over the last several decades. Research suggests that this is related to the increasing conceptualization of mental health problems as medical, which may translate into the impression that the problem is permanent. Consistent with this, most who seek care increasingly do so in the general medical sector rather than from a mental health specialist.

EFFICACY AND EFFECTIVENESS

Psychotherapy researchers distinguish between efficacy and effectiveness research. Each has a distinct purpose, methodology, and interpretative context.

Efficacy

Efficacy studies evaluate the sufficiency of a specific treatment to reduce distress, symptoms, and impairment with a group of patients having a particular psychiatric disorder. To minimize the influence of confounding or extraneous factors, efficacy studies are conducted using randomized clinical trial methodology in a controlled setting. Patients are screened to control for excessive patient variability, and they are randomly assigned to interventions that are being compared. The treatment under investigation might be compared to no treatment, a waiting list, placebo, minimal intervention (e.g., psychoeducation), or an alternative treatment. Therapists are trained to conduct the competing treatments to maximize the integrity of putative therapeutic ingredients. Most studies also perform integrity checks, such as reviews of taped sessions, to see whether the therapy under study was implemented with fidelity.

Most efficacy studies use treatment manuals, which direct (sometimes greatly, sometimes minimally) the therapist with regard to the intervention offered. Thus, treatment manuals have been essential in establishing that certain psychotherapeutic techniques are efficacious. Prominent examples of treatment manuals include Aaron Beck, John Rush, Brian Shaw, and Gary Emery's *Cognitive Therapy of Depression;* Lisa Najavit's *Seeking Safety: A Treatment Manual for PTSD and Substance Abuse;* and Myrna Weissman, John Markowitz, and Gerald Klerman's *Guide to Interpersonal Psychotherapy*. Other manuals have been published in journals and as chapters in books, such as in David Barlow's *Clinical Handbook of Psychological Disorders.*

Effectiveness

Effectiveness studies are concerned with whether psychotherapy delivered in actual clinical settings is effective in reducing the symptoms, distress, and dysfunction associated with mental illness. The experimental controls used in efficacy studies are absent in effectiveness studies, just as they are absent in community settings. Indeed, most effectiveness studies are conducted in the community. For instance, in real-world clinics, patients with comorbid disorders are commonplace, whereas they tend to be screened out of efficacy studies. Likewise, at the typical clinic, length of treatment is determined by patient need rather than treatment prescription. Thus, the goal of effectiveness research is to describe patient responses to the typical, community-based delivery of psychotherapy.

Effectiveness studies typically assess the characteristics of settings and providers, the generalizability of the patient samples, the features of treatments that are actually delivered, barriers to treatment experienced by patients, and the average costs associated with treatment. Overall, effectiveness studies have demonstrated that manualized treatments found to be efficacious in randomized clinical trials can be effective when applied in community settings.

Research also tends to find that the magnitude of treatment effects in effectiveness studies is similar to that found in efficacy studies. Effectiveness studies have demonstrated a dose–response relation between the number of sessions attended and the probability of improvement. For example, one study found that about half of patients are measurably improved by their eighth session of treatment. Other studies have found that chronically distressed patients, including those with personality disorders, improve more slowly than those who are more acutely distressed.

The results of effectiveness studies are sometimes seen as more relevant to actual practice than the results of efficacy studies, but both are important to the implementation of evidence-based practice. Although stringent experimental control reduces the likelihood of confounding variables in efficacy studies, the results can have limited external validity (generalizability) because of those very controls. On the other hand, whereas effectiveness research is more likely to be conducted in actual clinical settings, the causal implications of such research must be viewed with caution due to the high likelihood that confounding variables influenced the results.

Systematic Reviews and Meta-Analyses

Common methods to summarize and integrate efficacy and effectiveness research are systematic reviews and statistical meta-analyses. Systematic reviews apply specific strategies to assemble, critically evaluate, and integrate relevant studies that address a specific clinical question. Systematic reviews are exemplified in treatment guidelines published by the American Psychiatric Association and other organizations. Meta-analysis is a particular type of systematic review that uses statistical methods to combine and summarize the results of original studies. The individual studies are quantified and coded so that they can be compared with one another. Typically, an effect size is a measure of treatment effect that is calculated for each study by computing the size of the difference between the treated and untreated (or control) groups.

Hundreds of meta-analyses of psychotherapy have been conducted, and most have concluded that psychotherapy is an effective intervention for psychiatric illness across diverse populations and settings. Most meta-analyses have focused on specific disorders (e.g., the average effect of psychological treatment for major depressive disorder) or on specific therapeutic approaches for various disorders (e.g., comparing CBT to other therapies). In general, meta-analytic studies have shown that psychological treatments are vastly more effective than no treatment (e.g., placebo, wait-list controls, or minimal interventions), are about as effective as biomedical treatments (e.g., pharmacotherapy) for most disorders, and are about equally effective when compared to each other.

EFFICACY AND EFFECTIVENESS OF COMMON EVIDENCE-BASED TREATMENTS

Cognitive Therapy

Cognitive therapy is designed to help patients recognize, challenge, and modify irrational or dysfunctional beliefs, ideas, perceptions, and self-statements that are related to maladaptive emotional reactions to negative or stressful events. The general term for modifying such beliefs is “cognitive restructuring,” which entails identifying habitual thinking, understanding that thinking strongly influences emotions, evaluating the rationality of thinking, and modifying the thinking if it is found to be erroneous. Cognitive therapy has been investigated intensively because it has highly detailed putative causes and cures so that these components are easy to describe in treatment manuals and examine in outcome studies.

Cognitive therapy has been found to be effective in a variety of psychological disorders. A review of meta-analyses of cognitive therapy concluded that it is highly effective in alleviating the symptoms of mood and anxiety disorders in particular. The reported benefits of cognitive therapy for depression are comparable to those of pharmacotherapy. There are some indications that a combination of psychotherapy and medication is the most effective, but that cognitive therapy may be better tolerated by patients and may have more enduring effects. Specific cognitive techniques found to be helpful in treating depressive symptoms include monitoring mood, identifying cognitive distortions, and creating rational responses. These interventions are typically conducted via daily or weekly logs that allow the patient to keep track of their fluctuating emotional states, to list the related maladaptive thoughts, and to dispute the identified irrational beliefs. Maintenance sessions of cognitive therapy following positive response to the treatment of depression have been shown to reduce relapse rates.

In the treatment of anxiety disorders, cognitive therapy has been shown to be effective with generalized anxiety disorder (GAD), panic disorder, social anxiety disorder, and specific phobias. Patients with GAD benefit from identifying and challenging habitual and irrational beliefs about events and their ability to cope with worry. Similar results for panic disorder have been reported, in which clinically significant reduction in panic symptoms and panic-related cognitions has been demonstrated. Phobic disorders are also amenable, to a somewhat lesser extent, to modifications of anxiety-provoking cognitions. The effectiveness of cognitive therapy with various other disorders, including substance abuse, schizophrenia, personality disorders, and eating disorders, has been documented, if not as extensively as with mood and anxiety disorders.

Janice, a 29-year-old married mother of two school-age children, struggled with moderately severe depression on and off for several years before seeking treatment from a cognitive therapist. With the assistance of the therapist, she learned to become aware of negative self-statements and to identify irrational beliefs related to the requirements of being a “good mother.” She came to recognize that these thoughts caused her to feel incompetent and generally upset about how her children were doing. She learned to evaluate and appreciate evidence that her children were, in reality, well-adjusted and well-liked by both teachers and peers. Via cognitive restructuring, she learned to challenge and modify her thinking so that she could judge herself more accurately as a good mother. Other long-held, negative beliefs that contributed to depressed mood were also identified and challenged. Four months after terminating treatment, she returned for some “booster sessions” that helped her to get even better at recognizing and challenging excessively negative self-evaluations.

Behavior Therapy

Behavior therapy uses operant conditioning, classical conditioning, and modeling to help patients to change maladaptive behavior patterns that create and maintain psychiatric disorders. Techniques include the use of extinction, positive and negative reinforcement, stimulus control, shaping, and vicarious conditioning. One of the most elemental behavioral techniques is relaxation training, which has been found effective for various problems in hundreds of studies. Through repetition of relaxation-inducing strategies (such as deep breathing or progressive muscle relaxation), patients are trained to voluntarily produce the relaxation response, which alleviates anxiety, depression, involuntary muscle tension, pain, somatic complaints, and various other stress reactions.

Exposure-based behavioral interventions are often combined with relaxation training, and they have been shown to be highly effective in the treatment of anxiety and related disorders. The typical behavioral intervention involves confronting a feared stimulus while preventing the usual anxiety-based reaction (e.g., fleeing or engaging in a compulsion). Research has shown that patients will habituate to the feared stimulus and that the fear-based response will extinguish. For example, exposure and response prevention for obsessive-compulsive disorder (OCD) is an extremely effective approach that involves presenting anxiety-provoking stimuli and preventing the patient from performing the compulsions normally engaged in to reduce anxiety. Exposure-based treatments have likewise been shown effective for other anxiety and related disorders, including GAD, posttraumatic stress disorder (PTSD), panic disorder, and both social and specific phobias. Behavior therapy has also been shown to be effective in the treatment of pain disorders, sleep problems, behavioral disturbances (such as those associated with schizophrenia), and substance use disorders.

Behavior therapy is also commonly used to treat childhood psychopathology. Behavioral interventions might be implemented directly with the child. For example, early intensive behavioral interventions, which heavily rely on principles of positive reinforcement, have been shown to improve social skills, intelligence test scores, and school performance for children with autism spectrum disorders. Alternatively, behavioral principles have been successfully taught to parents for the alleviation of distress, symptoms, and impairment related to attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder, school phobia, separation anxiety disorder, and enuresis. Behavioral interventions have also been shown effective in reducing self-injury in children with intellectual disabilities.

Tony, a 10-year-old boy with ADHD, was being treated with a tolerable dose of stimulant medication. On the advice of their pediatric psychiatrist, the family sought adjunctive behavior therapy to deal with Tony's poor school performance. The therapist and his parents developed a behavioral modification program to change his homework-related behavior. They developed a system for rewarding Tony if he engaged in adaptive behaviors, such as writing down homework in his assignment notebook, bringing home any necessary homework-related school materials, and completing and turning in homework assignments. His parents worked with the teacher to set up similar contingencies to reinforce Tony's compliance, positive peer-related behavior, and general attentiveness in the classroom. After several months, Tony's schoolwork improved dramatically. His parents and teacher noted that he appeared to enjoy school more, to be more self-confident, and to be better-liked by classmates.

Cognitive-Behavioral Therapy

Cognitive-behavioral therapy (CBT) combines the effective elements of cognitive therapy and behavior therapy. Although initial iterations of CBT focused mostly on behavioral *or* cognitive interventions, current CBT approaches are intended to help patients develop strategies to change maladaptive thoughts *and* behaviors. For example, CBT for anxiety might include cognitive restructuring to change irrational fears, relaxation training to reduce physiologic arousal, and exposure to extinguish fear reactions.

CBT has been the most extensively studied psychological intervention, and there is ample evidence supporting CBT's efficacy and effectiveness with numerous psychiatric disorders. A review of hundreds of meta-analytic studies determined the average effect size for CBT for different disorders. Regarding depressive disorders, CBT has demonstrated clear benefits as compared to a waiting list or no treatment conditions, but meta-analyses have shown mixed results in suggesting CBT is more effective compared to other active treatments (i.e., psychodynamic or interpersonal therapies). CBT and psychopharmacology have been shown to both produce medium to large effect sizes in treating depression, although a combination of both treatments was more beneficial than either alone. Individual studies suggest that CBT is also effective with certain subgroups of patients with depression, including older adults, adolescents, patients with comorbid diagnoses, and low-income minority women.

The effect sizes for CBT for most anxiety disorders and somatoform were consistently medium to large. CBT has also been shown to have a strong effect on treating insomnia, anger, and bulimia. Meta-analytic reviews suggest that CBT has a clear benefit for special subgroups of patients, including treating internalizing disorders in children and emotional distress in older adults. CBT has been shown to have small to medium effects for substance use disorders, with more benefits for cannabis and nicotine dependence, as well as for chronic pain. Although there are small to medium effects of CBT for treating bipolar disorder and schizophrenia, there is less support for the use of CBT without the combination of psychopharmacology for these disorders.

Dan had experienced disabling panic attacks for 20 of his 45 years but only sought treatment when they started to interfere with his career aspirations. Through his engagement in CBT, Dan learned about the physiology of panic attacks, which helped to reassure him that he would not die of an attack. Based on the use of cognitive techniques, other fears and misperceptions regarding his body's reactions to stress were identified and modified (e.g., that increased heart rate was not a sign of impending heart attack but was a normal biologic response to stress). Using behavioral techniques, Dan was engaged in relaxation training to breathe normally rather than hyperventilate during a panic attack. Over the course of treatment, Dan learned to confront previously feared internal and environmental cues without experiencing a panic attack. After 3 months, Dan felt that he was managing his panic disorder and discontinued treatment.

Third Wave Cognitive-Behavioral Therapies

Over the last 40 years, CBT approaches have evolved with new iterations incorporating a greater focus on contextual, existential, and developmental issues. These approaches have been referred to as "third-wave" cognitive-behavioral treatments and have growing research support.

Acceptance and commitment therapy (ACT) proposes that psychopathology is due to cognitive fusion (unquestioning adherence to the truth of one's thoughts) and efforts to avoid painful experiences, which lead to maladaptive behavioral restriction. ACT attempts to inculcate a more skeptical perspective toward one's cognitive habits, a more accepting attitude toward the pains and foibles of life, and a commitment toward behaviors that are consistent with one's deeply held values. In ACT, there is a focus on altering how a person relates to their internal sensations (e.g., approach sensations with mindfulness and acceptance) rather than a focus on reducing or eliminating such sensations. Over 60 randomized controlled trials have found ACT efficacious for a range of psychiatric (e.g., depression, OCD, substance use disorders, PTSD) and medical conditions (e.g., tinnitus, chronic pain).

Dialectical behavior therapy (DBT) proposes that skill deficits in emotion regulation are core causal factors of various disorders. DBT was developed for patients with borderline personality disorder (BPD), but it has also been applied to substance use and eating disorders with positive outcomes. Systematic reviews have demonstrated that DBT is efficacious in reducing nonsuicidal self-injury, suicidal ideation, and other symptoms related to BPD for both adults and adolescents.

Recent iterations of CBT also utilize mindfulness techniques (e.g., mindfulness-based cognitive therapy). Mindfulness refers to a particular way of focusing attention on the present in a nonjudgmental manner. Attitudinal qualities such as curiosity, patience, equanimity, letting go, and trust are important for mindfulness practice. Increasing mindfulness is intended to reduce the pathognomonic habit of ignoring or suppressing one’s emotions and thoughts. The practice essentially entails experiencing thoughts and feelings as mental events that may or may not be related to external reality but can be experienced without judgment or fear. A systematic review of over a hundred randomized controlled trials demonstrated that mindfulness-based interventions are effective for alleviating symptoms of depression, anxiety, substance use disorders, smoking, and chronic pain.

Interpersonal Therapy

Interpersonal therapy (IPT) is a structured, time-limited intervention that is based on the notion that interpersonal stress is a causal and maintaining factor in psychological distress. IPT was originally developed to treat depression, and it targets the following four domains: interpersonal deficits, role disputes, role transitions, and complicated bereavement. Therapy involves three phases; the therapist initially links the diagnosis of depression to a specific interpersonal problem, implements strategies for resolving the problem in the middle phase, and finally supports the patient’s self-efficacy to continually use the strategies after the end of the treatment.

Research has consistently supported the use of IPT, with the majority of studies focusing on the treatment of depression. A recent meta-analysis found that the effect size for IPT was moderate to large compared to control groups in the treatment of acute-phase depression but showed nonsignificant differential effects when compared to other psychological treatments. However, other studies found IPT to have slightly better dropout and remission rates but may take longer to show a therapeutic effect when compared to CBT. When compared to pharmacotherapy, the combination of IPT and pharmacotherapy prevented relapse better than medication alone. For those with subthreshold depression, IPT prevents the onset of major depression. IPT has also been successful in the treatment of depression in particular subgroups, such as adolescents and women with postpartum depression.

Beyond depression, IPT has been shown to be efficacious in treating eating disorders, particularly bulimia nervosa and binge eating disorder, with modest outcomes. In relation to eating disorders, IPT may be most appropriate for patients with bulimia nervosa who are also experiencing interpersonal relationship difficulties. In addition, IPT is efficacious compared to control groups in the treatment of anxiety disorders, with effects comparable to CBT. There is also emerging evidence for the use of IPT for the treatment of PTSD.

Maureen sought therapy for depression and anxiety at 45 years of age, 2 years after her youngest of four children departed for college. With her therapist, she was able to recognize that her successful rearing of her children had the unfortunate consequence of their departure from home. Maureen was able to identify and pursue new activities outside of the home, to appreciate and transfer the skills that she had mastered as a successful mother outside of the home, and to reestablish her musical interest and pursuits that had waned while she raised her children. After several months of therapy, Maureen was less depressed and anxious, as she pursued a teaching certificate in music, with the encouragement of her husband, at her community college.

Psychodynamic Psychotherapy

Psychodynamic therapies derive from Sigmund Freud's psychoanalysis. They are intended to promote self-understanding and insight into recurring problems with others. Psychodynamic therapists typically promote the development of a therapeutic alliance, within which the interpersonal problems are recurring in the patient's life transfer onto the therapist. In therapy, these interpersonal problems, which are the result of defense mechanisms and intrapsychic conflicts, are identified, interpreted, and managed with the intent that such understanding and new behaviors then transfer to other relationships and alleviate problems.

Psychodynamic psychotherapies can range in length and intensity from long-term psychoanalysis to short-term psychodynamic psychotherapy. Short-term psychodynamic therapies have been manualized (e.g., Summers and Barber's *Psychodynamic Therapy: A Guide to Evidence-Based Practice*) and evaluated in controlled trials.

Research supports both the efficacy and effectiveness of this treatment, suggesting it may be as helpful as cognitive and behavioral approaches. Some studies suggest that patients in psychodynamic therapy better retain long-term gains from treatment. Researchers have documented the effectiveness of psychodynamic therapy in the treatment of various disorders, including depression, substance use, generalized anxiety, panic, and somatic disorders, as well as personality disorders.

Dario, a 34-year-old married father of three boys, sought therapy at the insistence of his wife for recurrent problems with anger and depression, both of which seemed to stem from feelings of inadequacy. His marital relationship had worsened considerably since his wife had started a new job. Toward the beginning of psychodynamic therapy, Dario expressed frustration at the situation but insisted that he did not resent his wife's successful career. Dario's underlying feelings toward his wife became evident in the course of treatment, as the feelings were transferred toward his therapist. His therapist helped him to identify and interpret his feelings of anger, fear, and sadness and to recognize that they were a manifestation of feelings of intense inadequacy and fears leftover from an abusive childhood. Dario was able to confront honestly and openly these feelings with his therapist and, ultimately, with his wife. His anger and depression lessened considerably, and his marriage improved.

TREATMENT VARIABLES RELATED TO OUTCOME

Therapeutic Alliance

Effective psychotherapy requires a therapeutic alliance between the therapist and the patient. This alliance entails the patients' perception that the therapist understands, is capable of helping, wants to help, and genuinely cares for the well-being of the patient. This emotional aspect of the alliance enables agreement between the patient and therapist on the necessary therapeutic tasks to achieve the goals of symptom reduction, alleviation of distress, and/or improving life functioning.

Carl Rogers proposed that the alliance is curative in and of itself, theorizing that the patient’s experience of a nonjudgmental, warm, genuinely caring attitude on the part of the therapist mediates positive change. Psychoanalytic and cognitive-behavioral theorists, in contrast, argue that the relationship is only a necessary precondition for the implementation of the active ingredients of therapy that are responsible for positive change, such as interpretation or cognitive restructuring.

Research suggests that both perspectives are correct. Several meta-analyses have concluded that there is a consistent relationship between the quality of the alliance and positive outcomes. A strong therapeutic alliance, whether measured from the perspective of nonparticipant observers, the therapist, or the patient, is related to therapy effectiveness. These findings cut across various interventions, and a recent task force on evidence-based therapeutic relationships concluded that the alliance contributes to outcomes as much, and likely more than, the particular method of treatment. Research shows that a positive alliance facilitates patient adherence to treatment, suggesting that it is a necessary condition for successful interventions. Research likewise shows that patients who dropped out of treatment liked the clinician less, felt less respected, experienced a weaker therapeutic alliance, viewed the psychotherapist as more passive, and regarded psychotherapy as being less potent than other interventions.

The alliance is similarly important in nonindividual interventions. Studies examining group therapy have indicated that the benefit of this treatment modality is more likely when there is a strong level of group cohesion. Likewise, studies of family and couples therapy show that the alliance with the therapist is an important predictor of positive response. Research into the causes, uses, and effects of the alliance are ongoing.

Therapist Characteristics

Psychotherapy interventions and procedures have been extensively studied, typically through the utilization of treatment manuals that specify them. Some research suggests a significant positive relationship between compliance with a treatment manual and outcome, but other research suggests that the therapist's relationship skills and perceived supportiveness may be diminished by excessive compliance with a manual. Moreover, numerous reviews and meta-analyses have concluded that not all therapists obtain equivalent results even if they implement manual-based treatments. Some studies find that therapists determine much of the positive effects of treatment, suggesting a general effect of therapist skill rather than a particular intervention. Perhaps consistent with this, studies have also found that a small portion of therapists accounts for most of the occasional detrimental effects observed in treatment.

The most commonly studied therapist variables are those proposed by Carl Rogers: Research has consistently found therapist in-session empathy, genuineness, and unconditional warmth to be predictive of treatment success. Additional therapist characteristics that predict effectiveness include skill (but not necessarily experience), attention to the therapeutic alliance, capacity to question and adjust their approach, and their perceived ability to handle professional challenges. Research suggests that therapy is most likely to be successful if therapists attend to the therapeutic alliance and make efforts to address potential problems or "ruptures" in the alliance. In contrast, there is minimal evidence to suggest that therapist gender, age, race or ethnicity, socioeconomic background, or indeed therapeutic stance (i.e., preferred treatment modality) is predictive of treatment success.

PATIENT VARIABLES RELATED TO OUTCOME

Patient Demographic Characteristics

Research indicates that most interventions work equally well for individuals of different genders across the lifespan. In fact, a meta-analysis found that there is no association between patient age or gender and treatment effectiveness. Given that older adults have been found to prefer therapy over medication, research showing that older adults benefit from various forms of psychotherapy as much as younger adults are important. Broadly, research shows that several types of psychotherapy (e.g., CBT, behavioral therapy) are effective for children and adolescents; however, interventions requiring high levels of insight and advanced or abstract cognitive abilities (e.g., ACT, psychodynamic therapy) may be more challenging for younger patients.

When it comes to ethnic and racial background, research demonstrates that members of minority groups obtain fewer benefits from conventional psychotherapies compared to their non-Hispanic white counterparts. Consistent with this finding, dropout rates are relatively higher for minority groups, especially if the provider and patient come from dissimilar racial, ethnic, and cultural backgrounds. These findings have been shown to be related to a lack of trust, a preference for self-reliance, greater skepticism about the benefits of treatment, more concerns about potential negative effects, and enhanced perceptions of stigma, all of which make both treatment initiation and treatment adherence less likely.

Research further suggests the validity of these concerns. For example, the quality of services available to ethnic and racial minority groups tends to be significantly poorer compared to the services available to non-Hispanic white individuals, including underdiagnosis and inappropriate, ineffective (e.g., nonspecialty), or culturally incompetent treatment. Perhaps not surprisingly, research shows that culturally adapted interventions are typically more effective than conventional interventions for minority individuals, and the multicultural competence of the therapist has been linked to patient outcomes.

Patient Cognitive Characteristics

Numerous cognitive variables have been found to be predictive of success in psychotherapy. Research indicates that motivation to change and willingness to engage in the therapeutic process are essential. Related to this, numerous studies have documented that enhancing a patient’s motivation to change and participate in therapeutic efforts leads to a greater likelihood of successful treatment. These are likely related to evidence showing that a patient’s expectation that therapy will help is a key predictor of effectiveness. Another cognitive characteristic of patients is attachment style, and research shows it is related to the capacity to develop a therapeutic alliance. Evidence also suggests that outcomes are related to a patient’s ego strength (i.e., the capacity to question assumptions and perceptions) and intelligence, with higher levels related to better outcomes for both characteristics.

Patient Clinical Characteristics

Investigation of the relationship between patient disturbance and outcome has yielded consistent findings: Patients with less disturbance at the beginning of treatment are relatively healthier at the end of treatment. Paradoxically, however, patient distress is the single best predictor of the likelihood of seeking professional treatment. Specific to mood disorders, patients with high levels of impairment tend to require relatively long-term, intensive treatment. Research has also delineated that the diagnosis of a personality disorder (either as the primary focus of treatment or as a comorbid condition of other disorders) is an indicator of relatively poorer treatment outcomes.

The quality of the alliance appears to be influenced by a variety of intra- and interpersonal patient characteristics. Patients with the capacity for positive social relationships tend to develop a better alliance than defensive or negativistic patients, whereas problem severity does not appear to affect the quality of the alliance. That said, research that tracks the alliance across therapy clearly demonstrates that problems in the alliance can be attended to and repaired via active intervention on the part of the therapist to the benefit of the patient.

Patient–Therapist Matching

Patient and therapist similarity across age, gender, and racial and ethnic identity appears to be related to treatment outcome. When it comes to age and gender, the extent of the patient–therapist match has been shown to have a small to modest effect on outcomes. Patient–therapist ethnic similarities have been shown to be more related to outcome, and research shows that some racial and ethnic minority patients may prefer a therapist that is ethnically similar.

However, the advantages of matching based on ethnicity might be less important than the cultural sensitivity of the therapist and adjustments to therapy based on cultural considerations. For example, one study examining the effect of matching patient and therapist ethnicity in the treatment of Asian Americans, Mexican Americans, African Americans, and White Americans found that patients who were matched with their therapists on ethnicity stayed longer in therapy, but matching was positively related to outcome only for the Mexican American group. Still, other studies have found that if the therapist tries to address the dissimilarity of patient-therapist ethnicity openly, the therapeutic relationship is strengthened.

Perhaps more important than ethnic match is whether the therapist and patient have compatible beliefs about the causes and treatment of psychological problems. Unfortunately, the explanations of mental health conditions have, for the most part, been based on European and Western notions. Explanations and treatment approaches have largely ignored religious and indigenous cultural explanations for mental health problems. Current professional guidelines and the scientific literature concur: Therapists should not only respect patient beliefs about the etiology of their condition but also should take these attitudes into account when developing and implementing intervention strategies.

Patient–Treatment Matching

There have been efforts to optimize the effects of psychotherapy by matching patient variables to treatment interventions. One example is the aptitude-by-treatment interaction (ATI) approach to psychotherapy. This research attempts to identify treatment strategies and procedures that produce optimal outcomes for individual patients. ATI research first categorizes patients and therapists on different characteristics (e.g., resistive, impulsive, extroverted, directive, supportive). Then, ATI research tries to determine which procedures are important to effective treatment in different contexts, whether certain patients are likely to respond to a certain therapist or therapy characteristics, and thus whether patients can be helped most by assignment to certain therapies or therapists.

For example, one study of major depression found that externalizing patients (those who tended to blame their environment for their problems) and patients low in defensiveness did better in cognitive therapy, which is fairly directive. Whereas internalizing patients (those who tended to blame themselves for their problems) and patients high in defensiveness did better in supportive and self-directed therapy.

Another study also found that externally focused patients respond better to more directive therapies, such as cognitive-behavioral psychotherapy. In contrast, highly anxious patients are more responsive to emotion clarification and supportive approaches. Similarly, although supportive therapy has been shown to be more effective with patients having higher levels of distress, cognitive therapy has been shown to be more effective with patients presenting with lower levels of distress. There is some evidence that panic disorder responds best and most quickly to a combination of behavioral and cognitive interventions. ATI research is fairly recent and, by definition, requires standardized therapies.

EMERGING THEMES AND FUTURE DIRECTIONS

The question of whether psychotherapy works has been definitely answered: There is a plethora of evidence from efficacy and effectiveness studies indicating that therapy is effective in alleviating emotional distress and behavioral dysfunction. Two models of psychotherapy—the medical model and the contextual model—have competed in the last several decades. The medical model suggests that psychotherapies have specific ingredients that address a specific problem within the patient and therefore account for treatment outcomes. In contrast, the contextual model holds that therapies are effective due to universal relational and psychological factors, which vary according to the therapist's skill and unique presentation of the patient.

Research suggests that both models have merit: Certain psychotherapy methods have been repeatedly demonstrated to be more effective than others in the treatment of certain disorders (e.g., exposure therapy for specific phobia), *and* the therapeutic relationship is a necessary aspect of treatment, regardless of the type of psychotherapy, with therapists' and patients' contributions playing an important role in outcomes. However, the evidence for both models is imperfect, so current psychotherapy research continues to examine the relative importance of specific (e.g., interventions) versus nonspecific (e.g., the alliance) curative factors, differential effects of treatment techniques (i.e., are some interventions more powerful for some clients or conditions?), and the transfer of knowledge gained in the research setting to actual clinical practice.

Relatedly, one of the challenges facing the field is increasing the implementation of evidence-based treatments that have been developed in the laboratory by psychotherapists in community settings. As such, there has been a large increase in the last several years in the number of studies that examine how efficacious clinical interventions can be delivered in usual care settings or target patient-centered outcomes.

In addition, efforts to evaluate the effects of dissemination strategies (i.e., attempts to promote the use of research-supported psychotherapy techniques) via translational research are also on the rise. One advancement in this domain has been in the role of technology in the provision of psychotherapy.

In particular, the recent stay-at-home orders resulting from the COVID-19 pandemic contributed to a significant increase in the use of telemedicine. A recent survey of clinical psychologists showed that 85% of clinical work was conducted via telemedicine during the pandemic, compared to just 7% prior to the pandemic. Additionally, psychologists estimated about 35% of their clinical work would remain virtually delivered after the pandemic ends. A recent literature review and meta-analysis found emerging evidence that the effectiveness of videoconference and telephone-delivered psychotherapy is comparable to in-person delivery of care. Further evaluation of the effectiveness of virtual treatment is essential for ensuring evidence-based care for patients and has exciting implications for improving access to high-quality mental health care.

Another trend over the last several years is the emphasis on patient-focused care, which uses research to inform personalized treatment decisions about the effectiveness of a specific treatment with a specific patient. Several systems of case-based decision support have been developed and applied to patient monitoring and therapist feedback. These approaches include the use of a global measure of patient functioning, repeated measurement during the course of therapy, and norms based on large patient and nonpatient samples to assess reliable change and clinically significant change.

One approach uses patient profiling techniques to predict the course of response to treatment, given the clinical characteristics of the patient. Monitored progress can be compared to the expected rate of change qualified by the patient’s clinical characteristics. Another method for personalizing treatment is shared decision-making (SDM), an interactive process in which clinicians and patients collaborate to make healthcare decisions. Initially designed for use in medical settings, SDM in psychotherapy usually involves the therapist and patient working together to define the presenting problem, followed by the clinician presenting treatment options as well as the empirical evidence supporting those options and a discussion of the pros/cons of options. The result is a treatment plan that is guided by research and incorporates patient values and preferences. Initial evidence for the utility of SDM in mental health care is encouraging.

Looking forward, the main questions remaining for psychotherapy researchers focus on maximizing the likelihood of positive effects for as many persons in distress as possible. To this end, several areas of research show promise. In addition to the research outlined above, the development of interventions that target transdiagnostic factors linked to multiple diagnoses (e.g., emotional dysregulation and cognitive appraisal issues) may lead to more efficient and effective treatments. One example is David Barlow and colleagues' *Unified Protocol for the Transdiagnostic Treatment of Emotional Disorders,* which has been shown to effectively treat a range of anxiety, depression, and other emotional disorders.

Second, advancements in neuroscience research have identified several genetic, anatomical, circuitry, and functional causes of behavior as well as measurable brain changes following a successful course of psychotherapy. Further integration of neuroscience and psychotherapy could help develop more comprehensive interventions with longer-lasting effects.

Third, culturally appropriate treatment paradigms are an essential aspect of psychotherapy research. Accounting for cultural factors involves developing new interventions and/or adapting established protocols to address the unique characteristics of patients from traditionally marginalized backgrounds (e.g., racial/ethnic minority, LGBTQIA+, disabled, neurodivergent).

Fourth, understanding treatment failure is a critically important issue, given that the failure rate for certain disorders, such as substance abuse disorders, remains high.

Finally, although there has been some progress in promoting psychological interventions as legitimate and beneficial, there remains an intense and widespread stigma associated with mental illness and its treatment.

FURTHER READINGS

American Psychological Association. Recognition of psychotherapy effectiveness. *Psychotherapy (Chic)*. 2013;50:102–109.

Barlow DH, ed. *Clinical Handbook of Psychological Disorders: A Step-by-Step Treatment Manual*. 6th ed. Guilford Press; 2021.

Batastini AB, Paprzycki P, Jones AC, MacLean N. Are videoconferenced mental and behavioral health services just as good as in-person? A meta-analysis of a fast-growing practice. *Clin Psychol Rev*. 2020;83:101944.

Bernecker SL, Coyne AE, Constantino MJ, Ravitz P. For whom does interpersonal psychotherapy work? A systematic review. *Clin Psych Rev*. 2017;56:82–93.

Butler AC, Beck JS. Cognitive therapy outcomes: a review of meta-analyses. *J Norwegian Psychol Assoc*. 2000;37:1–9.

Cabral RR, Smith TB. Racial/ethnic matching of clients and therapists in mental health services: a meta-analytic review of preferences, perceptions, and outcomes. *J Couns Psychol*. 2011;58:537–554.

Carpenter JK, Andrews LA, Witcraft SM, Powers MB, Smits JAJ, Hofmann SG. Cognitive behavioral therapy for anxiety and related disorders: a meta-analysis of randomized placebo-controlled trials. *Depress Anxiety*. 2018;35(6):502–514.

Cristea IA, Gentili C, Cotet CD, Palomba D, Barbui C, Cuijpers P. Efficacy of psychotherapies for borderline personality disorder: a systematic review and meta-analysis. *Jama Psychiatry*. 2017;74(4):319–328.

Cuijpers P, Donker T, Weissman MM, Ravitz P, Cristea IA. Interpersonal psychotherapy for mental health problems: a comprehensive meta-analysis. *Am J Psychiatry*. 2016;173(7):680–687.

Driessen E, Hegelmaier L, Abbass AA, et al. The efficacy of short-term psychodynamic psychotherapy for depression: A meta-analysis update. *Clin Psychol Rev*. 2015;42:1–15.

Gelo O, Pritz A, Rieken B, eds. *Psychotherapy Research: Foundations, Process, and Outcome*. Springer; 2015.

Goldberg SB, Tucker RP, Greene PA, et al. Mindfulness-based interventions for psychiatric disorders: a systematic review and meta-analysis. *Clin Psychol Rev*. 2018;59:52–60.

Hall GC, Ibaraki AY, Huang ER, Marti CN, Stice E. A meta-analysis of cultural adaptations of psychological interventions. *Behav Ther*. 2016;47(6):993–1014.

Heinonen E, Nissen-Lie HA. The professional and personal characteristics of effective psychotherapists: A systematic review. *Psychother Res*. 2020;30(4):417–432.

Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. *Cognit Ther Res*. 2012;36:427–440.

Jiménez JP, Botto A, Herrera L, et al. Psychotherapy and genetic neuroscience: an emerging dialog. *Front Genet*. 2018;9:257.

Lambert MJ. The efficacy and effectiveness of psychotherapy. In: Lambert MJ, ed. *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. 6th ed. Wiley; 2013.

Mackenzie CS, Erickson J, Deane FP, Wright M. Changes in attitudes toward seeking mental health services: a 40-year cross-temporal meta-analysis. *Clin Psychol Rev*. 2014;34:99–106.

McAleavey AA, Xiao H, Bernecker SL, et al. An updated list of principles of change that work. In: Castonguay LG, Constantino MJ, Beutler LE, eds. *Principles of Change: How Psychotherapists Implement Research in Practice*. Oxford University Press; 2019;13–40.

Mojtabai R, Olfson M, Sampson NA, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. *Psychol Med*. 2011;41:1751–1761.

Nathan PE, Gorman JM, eds. *A Guide to Treatments That Work*. 4th ed. Oxford University Press; 2015.

Norcross JC, Lambert MJ, eds. *Psychotherapy Relationships That Work: Evidence-Based Responsiveness*. 3rd ed. Oxford University Press; 2019.

Olfson M, Wang S, Wall M, Marcus SC, Blanco C. Trends in serious psychological distress and outpatient mental health care of US adults. *JAMA Psychiatry*. 2019;76(2):152–161.

Öst LG. The efficacy of acceptance and commitment therapy: an updated systematic review and meta-analysis. *Behav Res Ther*. 2014;61:105–121.

Soto A, Smith TB, Griner D, Domenech Rodríguez M, Bernal G. Cultural adaptations and therapist multicultural competence: two meta-analytic reviews. *J Clin Psychol*. 2018;74(11):1907–1923.

Steele RG, Elkin TD, Roberts MC. *Handbook of Evidence-Based Therapies for Children and Adolescents*. Springer International Publishing; 2020.

Substance Abuse and Mental Health Services Administration. *Key substance use and mental health indicators in the United States: Results from the 2018 National Survey on Drug Use and Health External Icon*. HHS Publication No. PEP19-5068, NSDUH Series H-54. Center for Behavioral Health Statistics and Quality; 2019.

Wampold BE, Imel ZE. *The Great Psychotherapy Debate: The Evidence for What Makes Psychotherapy Work*. 2nd ed. Routledge; 2015.

Wang PS, Demler O, Olfson M, Pincus HA, Wells KB, Kessler RC. Changing profiles of service sectors used for mental health care in the United States. *Am J Psychiatry*. 2006;163:1187–1198.

33

Biologic Therapies

33.1 Preface

WHAT'S NEW IN THE BIOLOGIC THERAPIES SECTION?

The section is structured in a pharmacologically driven order (rather than the traditional disease-based) and in line with the concepts of *Neuroscience-based Nomenclature (NbN)* (see Chapter 33.1c).

Accordingly, the chapters are arranged in 21 subsections, focusing on pharmacologic domains (i.e., neurotransmitters) and modes of action (e.g., antagonists, reuptake inhibitors). Hence, for example, risperidone is now included in the "dopamine blockers" subsection, categorized as a "dopamine-serotonin antagonist." Concordantly, formerly "antidepressants" such as clomipramine or bupropion are now classified as "monoamine enhancers." However, in instances of unique indications (e.g., sexual dysfunction) or cases where agents of different mechanisms were more intuitively discussed in one "package" (e.g., substance abuse), the chapters remained indication driven.

In order to ease the reader's orientation to the main neurobiologic properties of each drug, an "NbN ID"—stating the pharmacologic domain and mode of action—was assigned to the relevant medications throughout the section.

Finally, aside from new agents introduced in the section (e.g., esketamine, pimavanserin), the current edition expands its scope to the global angle of biologic therapies (the essential list of medicines, biologic therapies in Africa, India, and Latin America) as well as aspects concerning gender and sex differences (e.g., Sex Differences in Psychopharmacological Treatment), and novel fast-acting treatments (e.g., MDMA and psilocybin).

We, the section editors, hope that these changes will help toto present to you, the reader, the depth and richness of the neuroscience fabric of psychotropics.

Joseph Zohar
Sasson Zemach

33.1a General Principles of Psychopharmacology

OLUSOLA AJILORE, M.D., PH.D., AND MARK M. RASENICK, PH.D.

Psychopharmacology is the use of medications as a treatment for mental disorders. Previous editions of this textbook have chronicled the evolution of psychopharmacology as a medical discipline. The first edition of *Kaplan and Sadock's Comprehensive Textbook of Psychiatry,* published in 1967, appeared a decade after the appearance of modern antidepressants, antipsychotics, mood stabilizers, and anxiolytics. Of the nearly 1,700 pages in that first edition, a mere 25 pages were devoted to drugs as treatments for mental disorders. Chlorpromazine (Thorazine), approved by the U.S. Food and Drug Administration (FDA) 15 years earlier, was alone among the drugs acknowledged as being a significant therapeutic breakthrough. Other psychotropic agents, including the benzodiazepines (introduced in 1959), were characterized as playing a minor therapeutic role, serving as adjuncts to psychotherapy or as a last resort for patients failing to respond to therapy. There was no intimation that benzodiazepines would become the most widely prescribed drugs of any type within a few years. Lithium (Eskalith), which had been described as an effective treatment of mania in 1949, was mentioned in a brief paragraph. Indeed, in reading the first edition of *Kaplan and Sadock's Comprehensive Textbook of Psychiatry* and other textbooks of that time, it would have been difficult to predict how psychotropic drugs would soon transform the treatment of mental disorders. In a reversal of positions, psychoanalytic theory, which once served to define the practice of psychiatry, has been supplanted by psychopharmacology as the most widely used form of treatment of psychiatric disorders. However, there is concern by many in the field that psychopharmacology has stalled in the past decade. This crisis is due, in part, to the high cost of failure for new drug trials, particularly as it is difficult to separate out from placebo for many psychiatric drugs. This crisis is also due to insufficient government budgets for basic research. In the absence of new molecular targets, the ability of pharmaceutical companies to develop truly novel psychiatric drugs is in jeopardy. To some degree, there has been relief in recent years. After a relative drought in psychotropic drug development, 2019 alone saw three new FDA-approved medications for treatment-resistant depression, postpartum depression, and schizophrenia. Of note, these medications, esketamine, brexanolone, and lumateperone, represented novel mechanisms of action compared to current medications. Thus, highlighting that in order to advance the current knowledge in the field, it is important to understand the present state of psychopharmacology. It is within this context that this chapter describes the basic principles of both pharmacology and psychopharmacology.

In order to engender a complete understanding of the basic principles of psychopharmacology, it is first necessary to understand the principles of pharmacology—particularly those relating to receptors and cellular signaling. This chapter is organized into two sections: the first section will introduce the basic principles of pharmacology, and the second section will build upon them to discuss general issues with psychotropic drugs.

RECEPTOR PHARMACOLOGY

The majority of drugs used for palliation of psychiatric symptoms either mimic, block, or modify the actions of neurotransmitters, and most do so by binding to receptors for those transmitters. While receptors can be located either on the plasma membrane or intracellularly, the neurotransmitter receptors relevant to the drugs to be discussed in this chapter are transmembrane molecules, with outer surfaces forming binding pockets for the neurotransmitter.

The receptors discussed in this chapter are either ion channels (ionotropic) or metabotropic (G protein–coupled), and many neurotransmitters (e.g., acetylcholine, glutamate, gamma-aminobutyric acid [GABA], and serotonin) act as agonists for receptors in each class. This chapter will discuss the basic biology of many of these transmitters and receptors. It begins with a description of the characterization and quantitation of different receptors.

Receptor Analysis

There are two basic mechanisms for studying receptors. One is to measure the binding of a ligand (often radioactively labeled), and the other is to measure the physiologic consequences of agonist binding (and the effects of antagonist).

Binding Assays. In order to study receptors with radioligand binding, hormones, neurotransmitters, or their analogs have been labeled with a variety of radioactive compounds. Membranes prepared from cultured cells or brain tissue (or intact cells) can be assayed for their ability to bind these radioactive drugs. The radioactive compounds can be either agonists or antagonists and through the use of these compounds, a quantitation of both the number of receptors and the affinity of those receptors for the drug can be determined.

In order to quantify binding, ligands are added to membrane preparations at known concentrations and competed with nonradioactive ligands in order to determine the extent to which binding is specific. Membranes are trapped on filters, the filters are washed, and radioactivity is determined by scintillation spectrometry. Typical results are illustrated in Figure 33.1a–1. Specific binding can be transformed by Scatchard analysis (Fig. 33.1a–2) to reveal the existence of multiple affinity states for agonists and Hill analysis to determine whether binding is cooperative.

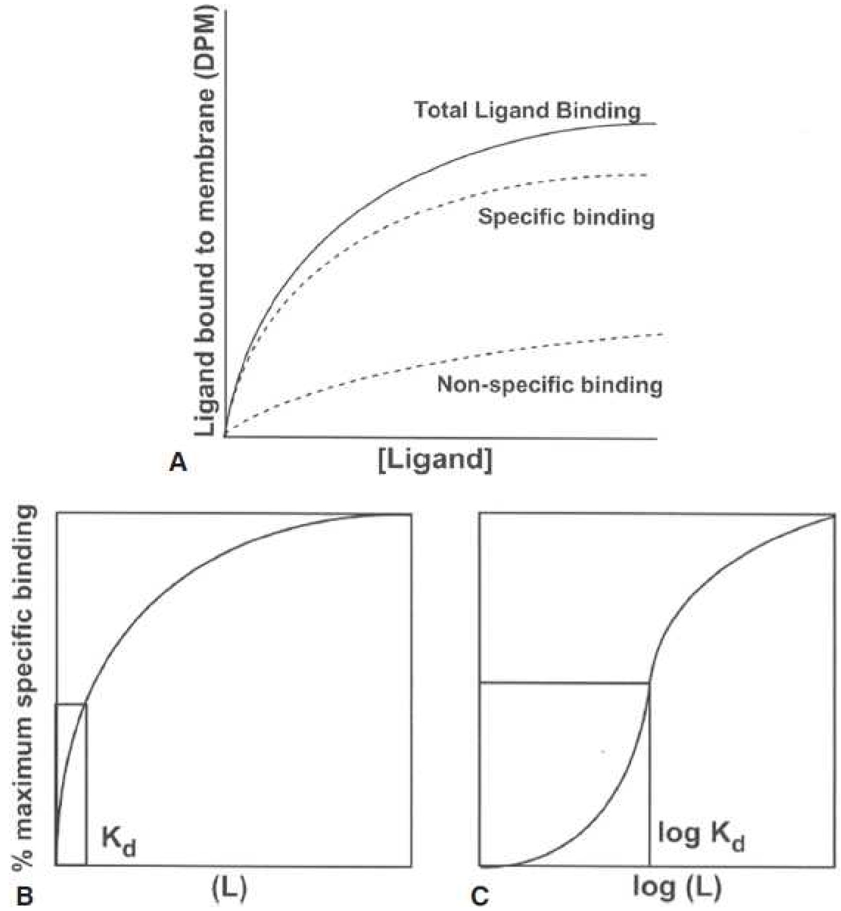

FIGURE 33.1a–1. Kinetics of ligand (agonist or antagonist) binding. **A:** Measurement of receptors using radioligand binding. Membranes (or cells) are incubated with radioactive agonist or antagonist and collected by sedimentation or filtration. Radioactivity in particulate fractions is calculated and plotted as shown above. Nonspecific binding refers to that portion of the total binding that persists in the presence of an excess of unlabeled compound. Specific binding is saturable, within the range of the experiment shown, but nonspecific binding does not saturate. **B:** Two plots of specific binding versus concentration of ligand. **C:** The approximate Kd is represented by the ligand concentration at 50% saturation of specific binding. DPM, disintegrations per minute. (Reprinted with permission from Felig P, Frohman LA. *Endocrinology and Metabolism*. 4th ed. McGraw Hill Education; 2001.)

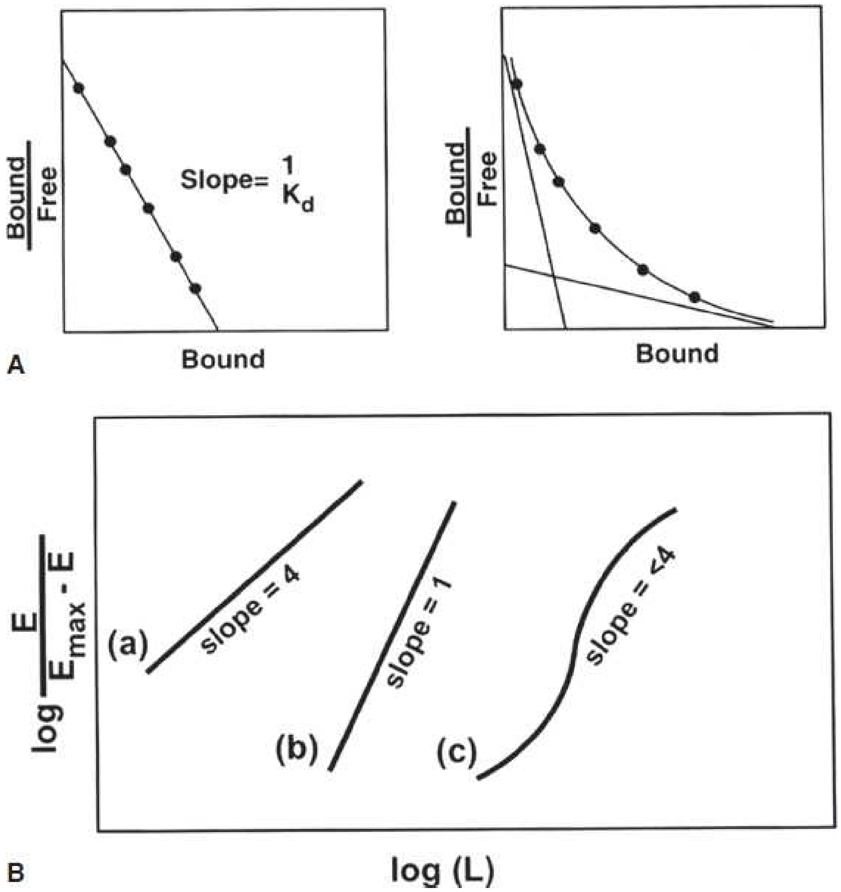

FIGURE 33.1a–2. Computational analysis of receptor binding. **A:** Scatchard plots. (Left) a receptor with a single affinity state for the binding of ligand; (right) two-site fit. **B:** Hill plots for the determination of cooperativity in ligand–receptor binding. Three Hill plots are shown here. (**a**) Four binding sites (possibly a tetramer); (**b**) is representative of a single site; (**c**) cooperative binding of fewer than four but more than one site. (Reprinted with permission from Felig P, Frohman LA. *Endocrinology and Metabolism*. 4th ed. McGraw Hill Education; 2001.)

Some receptors are oligomeric. Many ionotropic receptors have multiple subunits, and two or more of these bind agonists, often displaying cooperative binding. G protein–coupled receptors (GPCRs) have been shown to dimerize as well, and this can take the form of homodimers or heterodimers in the case of multiple receptor species. Although dimerization is demonstrably obligatory only for GABAB and sweet or umami taste receptors, many other forms of receptors have been demonstrated to dimerize.

Biologic (Efficacy) Assays. In addition to the direct determination of agonist or antagonist binding to a receptor, the biologic consequences of this binding can be assessed. One of these methods involves a biologic assay. A classical biologic assay utilizes an isolated organ strip, such as the guinea pig ileum, and measures changes in both the frequency and strength of contraction of the intestinal smooth muscle in response to the presence of an agonist. Such a system allows the investigator to coordinate the dose of a hormone with the strength of the response, but quantification is often imprecise.

More recent biologic assays rely on enzyme “effectors” activated subsequent to receptor activation. In the case of the guinea pig ileum, this might involve measuring the cyclic adenosine monophosphate (cAMP) produced by the enzyme adenylyl cyclase subsequent to the activation of the β-adrenergic receptor. Higher throughput assays designed to screen drugs that may act as agonists or antagonists for a given receptor might involve fluorescent sensors detecting an increase in Ca^{2+} subsequent to activation of a receptor that activates phospholipase c (e.g., 5HT2A), a process that ultimately increases cellular calcium. Such assays are referred to by the acronym FLIPR (fluorometric imaging plate reader) and are used routinely in the process of drug discovery and development. These high-throughput assays employ fluorescent calcium and cAMP (or other small molecules) sensors to determine GPCR activation and determine the efficacy of an agonist or potency of an antagonist at a given receptor.

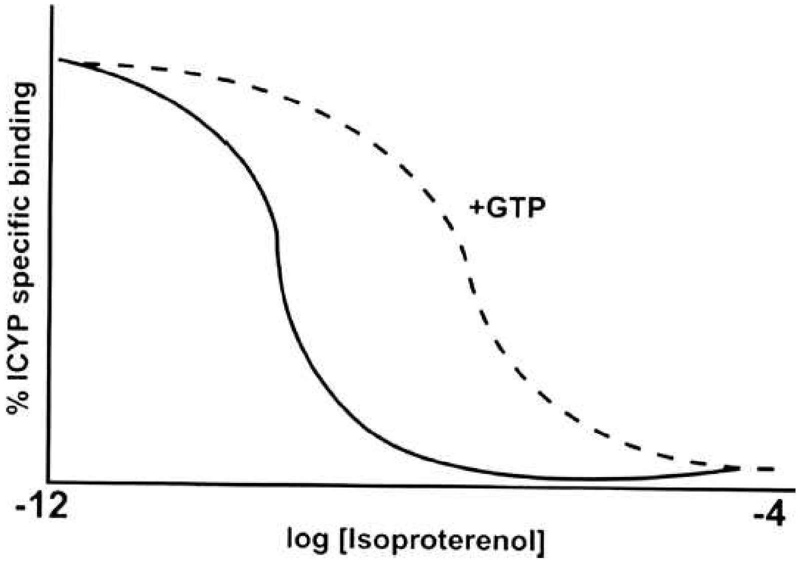

FIGURE 33.1a–3. Guanosine 5’-triphosphate (GTP) decreases agonist affinity. In the example shown, the β-adrenergic agonist, isoproterenol, is used to displace the labeled β-adrenergic antagonist, 125I-iodocyanopindolol (ICYP). In the presence of GTP, significantly more isoproterenol is required to displace the iodocyanopindolol. This “right shift” of the binding curve indicates that, in the presence of GTP, affinity for agonist by the receptor is decreased. (Reprinted with permission Felig P, Frohman LA. *Endocrinology and Metabolism*. 4th ed. McGraw Hill Education; 2001.)

Efficacy could also be plotted in a manner similar to that used in Figure 33.1a–1. The X-axis would continue to represent the amount of drug required (potency), and the degree of activation (efficacy) would be represented on the Y-axis.

G Protein Modulation of Agonist Binding. GPCRs show a significant decrease in agonist affinity following the introduction of guanosine 5′-triphosphate (GTP) into the medium. This is demonstrated in Figure 33.1a–3. Isoproterenol, a β-adrenergic agonist, displaces iodocyanopindolol (a radioactive antagonist) with less potency in the presence of GTP. It is suggested that the complex between receptor and G protein represents the high-affinity state for agonist binding. Agonist allows the active receptor to induce the G protein to bind GTP (this will be explained more completely below), which then causes the G protein to dissociate from the receptor. This “free receptor” has a lower affinity for agonists than does the receptor–G protein complex. Note that some antagonists also show the ability to undergo this “GTP shift.” As will be noted below, some antagonists also modify the conformation of certain other GPCRs.

Structure and Function of Receptors

Ionotropic receptors were first to be characterized structurally, as the nicotinic acetylcholine receptors (nAChRs) are present in large concentrations in the electroplax organs of electric fish and eel and proved rich sources of receptor that was modeled by x-ray diffraction. Many years later, outer segments of retinal rod photoreceptors proved a similarly rich source of rhodopsin, the first GPCR to be crystalized by Palczewski and colleagues. While solving the structures of GPCRs more relevant to psychiatry was daunting, Kobilka, Stevens and their colleagues devised a mechanism for modifying these receptors, and a number of pharmacologically relevant GPCRs have now been determined structurally by Roth and colleagues.

In the case of ionotropic receptors, the binding of an agonist elicits a cascade that initially creates a channel for ion conductance. Second messenger systems (elicited subsequent to GPCR activation) modify channel function considerably. In the case of GPCR, activation subsequent to agonist binding results in a change in receptor morphology that releases GDP from the bound G protein α subunit. The subsequent binding of GTP activates the G protein, allowing α subunits to engage intracellular effector enzymes such as adenylyl cyclase or phospholipase and β/γ subunits to engage voltage-dependent ion channels.

Ionotropic Receptors

VOLTAGE GATED. These receptors, which are intrinsic ion channels, are either voltage-gated or ligand-gated. The first class, consisting of Na^+, K^+, and Ca^{2+} channels, possesses sensors (arrays of charged amino acids within transmembrane regions). Many of these receptors have multiple (4 or 5) subunits that combine to form a functional receptor. Several mutations have been identified for voltage-sensitive channels, and these result in a variety of maladies (channelopathies) ranging, in the case of calcium channels, from familial migraine to dystonia (Fig. 33.1a–4).

LIGAND-GATED. These receptors, which also form ion channels that open in response to agonists, often share the quatrameric or pentameric structure of the voltage-gated receptors but have specialized regions that bind one or two molecules of agonist. nAChRs, GABAA receptors, glutamate receptors (either kainate or *N*-methyl-D-aspartate [NMDA]), and 5HT3 receptors fall into this class (Fig. 33.1a–5). While some of these receptors (e.g., nAChR) are regulated by the binding of an agonist, others have more complex regulations. GABAA has binding sites for benzodiazepines that increase agonist potency as well as binding sites within the channel region that decrease chloride conductance, hence acting as convulsants (Strychnine) (Fig. 33.1a–6). Similarly, NMDA receptors, which normally act as Na^+ channels, conduct Ca^+ subsequent to membrane depolarization (Fig. 33.1a–7) and can even participate in rapid remodeling of the synapse. Curiously, glycine, which is an agonist for a Cl^-channel receptor similar to GABAA (but found primarily in the spinal cord), also acts to promote the activity of glutamate at the excitatory NMDA receptor.

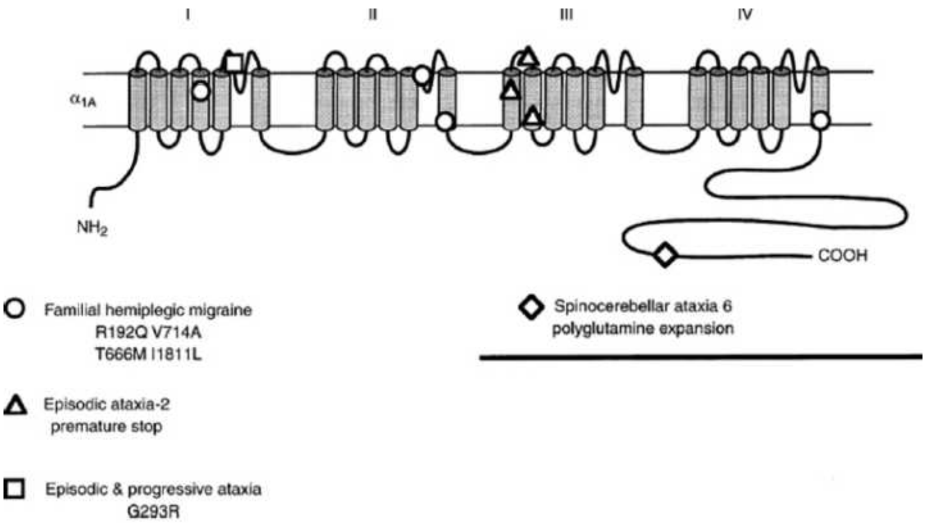
I
II
III
IV
α1A
NH2
COOH
○ Familial hemiplegic migraine
R192Q V714A
T666M I1811L
◇ Spinocerebellar ataxia 6
polyglutamine expansion
△ Episodic ataxia-2
premature stop
□ Episodic & progressive ataxia
G293R

FIGURE 33.1a–4. Mutations in Ca^{2+} channels and disease phenotypes (channelopathies). This model of a voltage-dependent Ca^{2+} channel shows clusters of mutations and the relevance of mutations in a particular region for a particular malady. FHM, Familial hemiplegic migrane; EA2, Episodic ataxia type 2; CSNB, X-linked congenital stationary night blindness. (Adapted with permission. Lorenzon NM, Beam KG. Calcium channelopathies. *Kidney Int.* 57(3):794–802. Copyright © 2000 International Society of Nephrology. doi: 10.1046/j.1523-1755.2000.00917.x.)

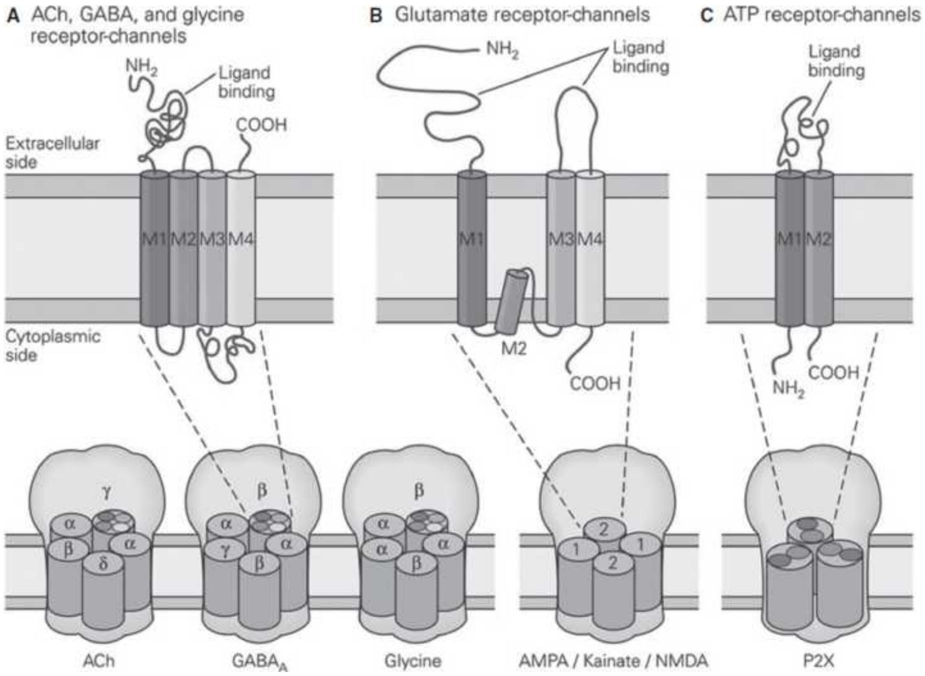
A ACh, GABA, and glycine receptor-channels
B Glutamate receptor-channels
C ATP receptor-channels
NH2
Ligand binding
COOH
Extracellular side
Cytoplasmic side
M1
M2
M3
M4
NH2
Ligand binding
M1
M3
M4
M2
COOH
Ligand binding
M1
M2
NH2
COOH
γ
α
β
δ
α
β
α
γ
β
α
β
α
α
β
α
2
1
2
1
ACh
GABAA
Glycine
AMPA / Kainate / NMDA
P2X

FIGURE 33.1a–5. Families of ligand-gated ion channels. **A:** Nicotinic AChR, GABA, and glycine receptor channels are all pentamers and are comprised of many different types of subunits. Their actions are explained in the text. **B:** Glutamate receptors are tetramers, usually comprised of two each of similar subunits. **C:** Purinergic (ATP) receptors are trimmers, each subunits with two membrane-spanning domains. (Adapted with permission from Kandel ER, Schwartz JH, Jessell TM, Siegelbaum SA, Hudspeth AJ, eds. *Principles of Neural Science*. 5th ed. Mcgraw-Hill Medical; 2013.)

Receptors for Growth Factors. Receptors for neurotrophins, such as nerve growth factor (NGF) and brain-derived neurotrophic factor (BDNF), have intrinsic tyrosine kinase activity and dimerize upon agonist binding (Fig. 33.1a–8). These receptors include trkA, trkB, and p75NTR. Regardless of the specific receptor type, most of the members of this receptor family consist of two subunits. Sometimes receptors form a homodimer and, at other times, a heterodimer. The formation of a dimer, which is a result of receptor activation, leads to autophosphorylation or transphosphorylation of the receptor within its cytoplasmic domain. Consequently, other substrates for that receptor phosphate are found and phosphorylate. Signal propagation often is a result of an activation of a series of kinase cascades resulting in cell growth or differentiation.

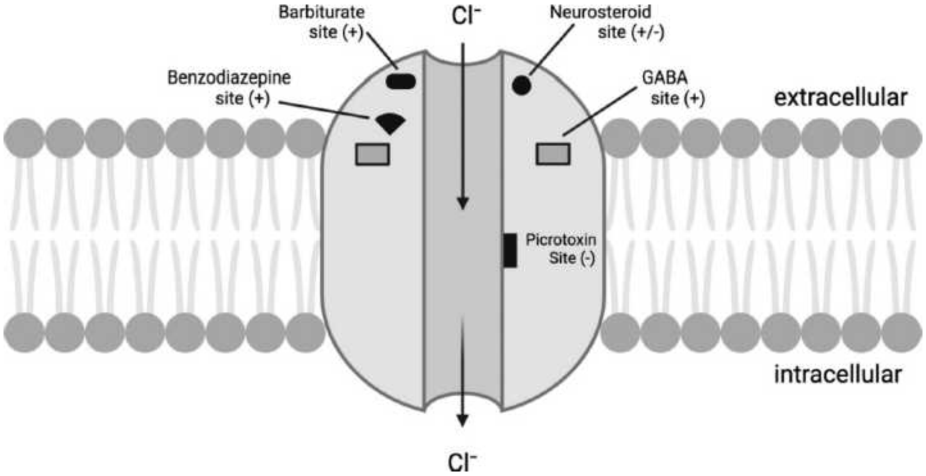
Barbiturate site (+)
Cl⁻
Neurosteroid site (+/-)
Benzodiazepine site (+)
GABA site (+)
extracellular
Picrotoxin Site (-)
intracellular
Cl⁻

FIGURE 33.1a–6. Structure and function of the [GABAA/GABA-A/GABAA] receptor. A model for the receptor shows binding sites for GABA as well as compounds that affect GABA signaling. Activation of the receptor increases chloride conductance. Compounds not binding directly to the GABA site modify (+/-) GABA-[activated/actuated] channel opening. (Created with Biorender.com)

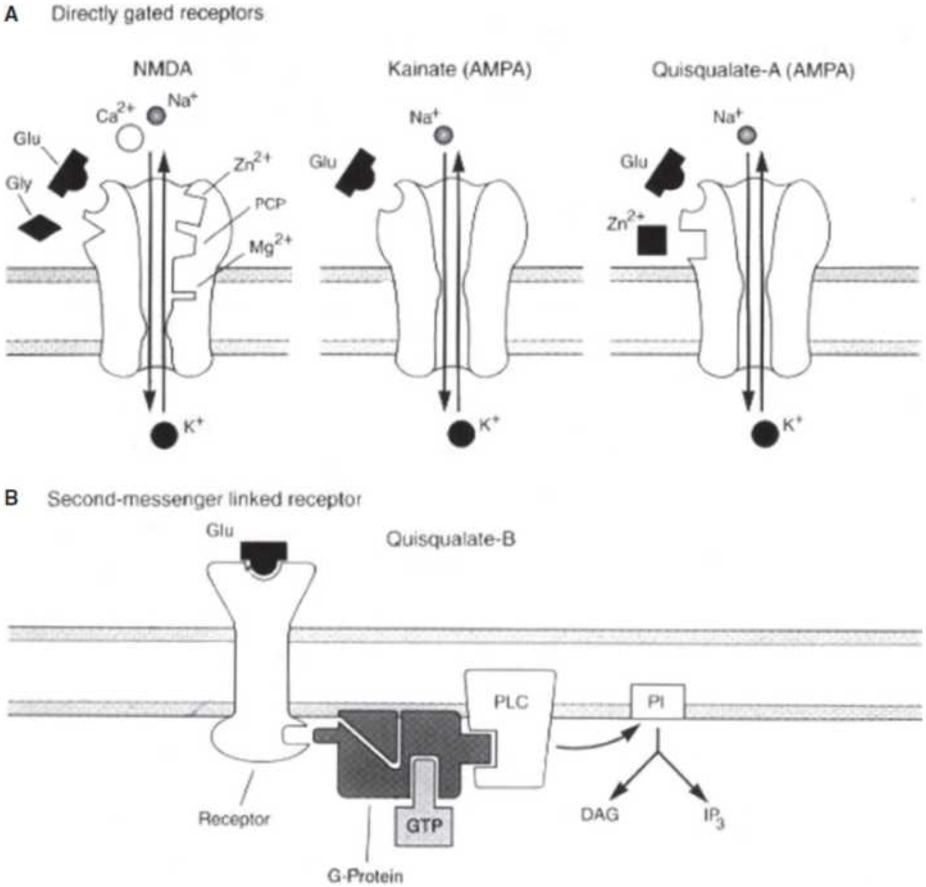
A
Directly gated receptors
NMDA
Kainate (AMPA)
Quisqualate-A (AMPA)
Ca2+
Na+
Glu
Gly
Zn2+
PCP
Mg2+
Na+
Glu
Na+
Glu
Zn2+
K+
K+
K+
B
Second-messenger linked receptor
Glu
Quisqualate-B
PLC
PI
Receptor
GTP
G-Protein
DAG
IP3

FIGURE 33.1a–7. Glutamate receptors. Three classes of glutamate receptor regulate excitatory synaptic actions in neurons in the spinal cord and brain. Two types of ionotropic glutamate receptor gate ion channels directly. Two subtypes of non-*N*-methyl-d-aspartate (NMDA) receptor bind the glutamate agonists kainate or alpha-amino 3-hydroxy-5-methylisoxazole-4-proprionate (AMPA) and regulate a channel permeable to Na^+ and K^+. The NMDA receptor (left) regulates a channel permeable to Ca^{2+} and Na^+ and has binding sites for glycine, Zn^{2+}, phencyclidine (PCP or "angel dust"), MK801, and Mg^{2+}, which modulate the functioning of this channel in the direction indicated.

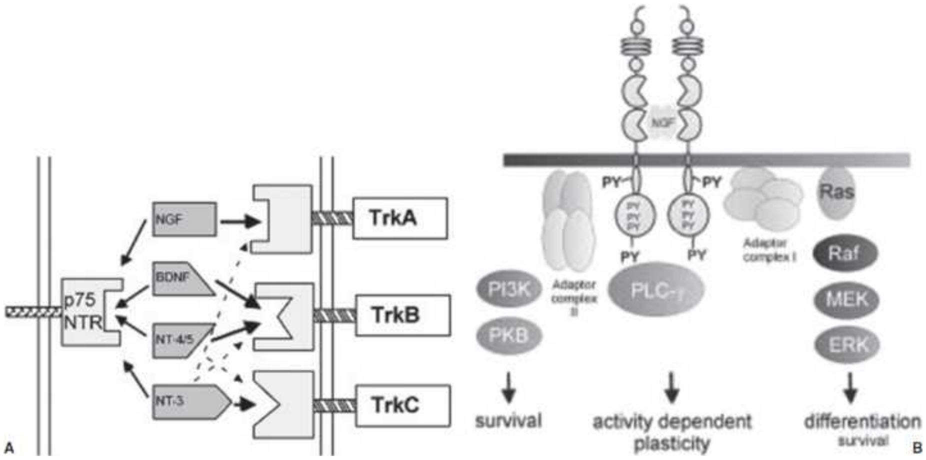
NGF
BDNF
NT-4/5
NT-3
p75 NTR
TrkA
TrkB
TrkC
A
NGF
PY
PY
PY
PY
Ras
Adaptor complex I
Raf
Adaptor complex II
PI3K
PLC-γ
MEK
PKB
ERK
survival
activity dependent plasticity
differentiation
survival
B

FIGURE 33.1a–8. Neurotrophin receptors. **A:** All neurotrophins bind p75NTR, a receptor lacking tyrosine kinase activity. The receptors with tyrosine kinase activity are selective for the molecules indicated in the figure. **B:** Upon activation, the three Trk receptors dimerize, autophosphorylate, and phosphorylate intracellular targets. A variety of survival or differentiation pathways can be activated.

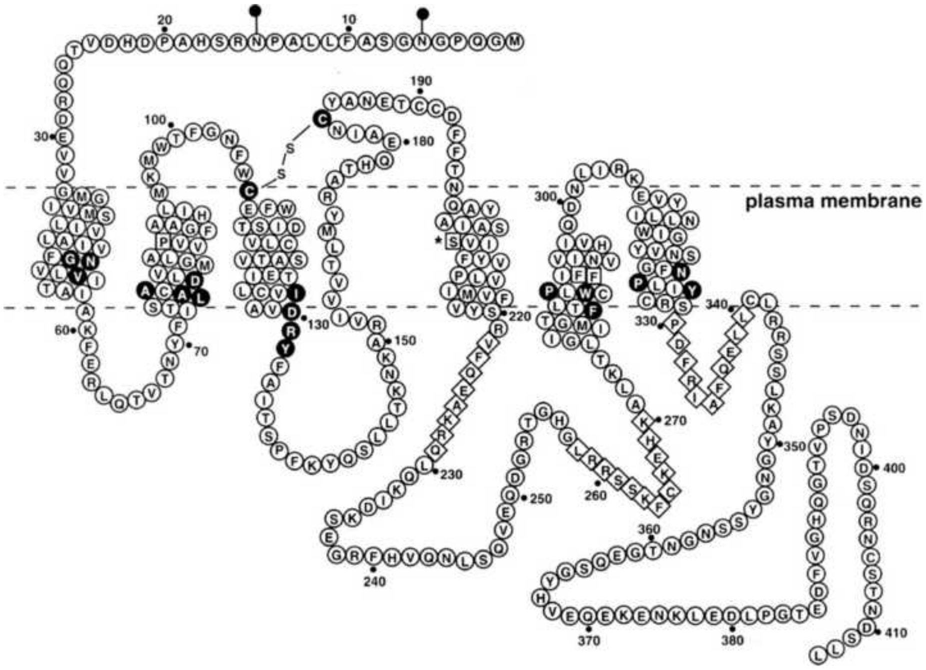

FIGURE 33.1a–9. Model of a “classic” G protein–coupled receptor. The β1-adrenergic receptor is shown. Highly conserved residues (among all G protein–coupled receptors) are shown in black. Boxed residues are conserved in catecholamine receptors and the serine marked with * is conserved in all G protein–coupled serotonin receptors. Diamond residues are involved in G protein coupling. (Adapted with permission of Probst WC, Snyder LA, Schuster DI, Brosius J, Sealfon SC. Sequence alignment of the G-protein coupled receptor superfamily. *DNA Cell Biol.* 1992;11(1):1–20. doi: 10.1089/dna.1992.11.1.)

Metabotropic (G Protein–Coupled) Receptors. GPCRs are a large family of receptors for light, odorants, small molecule neurotransmitters as well as a variety of peptide agonists. It is estimated that somewhere between 30% and 70% of current drugs relate to GPCRs, elevating their importance for psychopharmacology. GPCRs regulate a multitude of physiologic functions through activation of heterotrimeric G protein α and βγ subunits that bind to target proteins to initiate cellular signal transduction cascades, a process that is critically important for neurons and neurotransmitter action. Nearly 1,000 GPCRs have been identified in the human genome and include both chemosensory receptors and receptors for hormones and neurotransmitters. Approximately 100 of these GPCRs have not been linked to an endogenous ligand. These are known as orphan receptors. All GPCRs share a similar structure, which includes seven transmembrane domains, an extracellular amino-terminal, an intracellular carboxyl-terminal, three extracellular loops, and three cytosolic loops (Fig. 33.1a–9). While each of the seven transmembrane domains comprises somewhere between 20 and 27 amino acids, the end-terminal segments, as well as either the extracellular or cytoplasmic loops, can vary greatly among the different receptor types.

Changes in Receptor Conformation upon Agonist Activation. Mutational analysis suggested several regions of GPCRs were involved in coupling to and activating G proteins. Curiously, the elucidation of the crystal structure of the β_2 receptor, suggested that there were distinct patterns among GPCRs and their cognate G proteins.

Receptors with constitutive activity, which are active in the absence of an agonist, can be generated by mutation. The importance of the C-terminal portion of the third intracellular loop was demonstrated in studies on the adrenergic receptor that is linked to norepinephrine-induced activation of phosphatidylinositol metabolism.

Another region relevant for G protein coupling is located at the junction between the third transmembrane segment and the second intracellular loop of many GPCRs. This region is delineated by a motif known as E/DRY (Glu/Asp-Arg-Tyr), which is highly conserved among GPCRs, and mutations result in constitutive activation of many different GPCRs.

Some naturally occurring mutations result in constitutive receptor activity. Shenker and colleagues (1993) identified a mutation (Asp578Gly/Tyr) in the sixth transmembrane domain of the luteinizing hormone receptor, which confers constitutive activation. Males carrying this mutation display precocious puberty. Some types of GPCRs display constitutive activity in their normal state. While D1a dopamine receptors require the presence of an agonist to activate G protein, D1b receptors show high-constitutive activity, almost completely activating G protein in the absence of an agonist.

INVERSE AGONISM. The demonstration of naturally occurring constitutive activity for GPCRs raises the possibility that inhibition of such constitutive activity should have a functional effect. Thus, for a number of receptors, compounds exist that simply inhibit the binding of agonists (neutral antagonists), and other compounds exist that inhibit the activity of a constitutively active receptor (inverse agonists). This was demonstrated for the δ opiate receptor in studies that revealed that the “inverse agonist” ICI 174864 lowered the basal activation of Gi by these receptors. Considering the model of receptor–G protein complexes presented above, it would be predicted that inverse agonists would show preferential binding to receptors that were uncoupled to G proteins. Inverse agonism has been demonstrated at a number of receptor types.

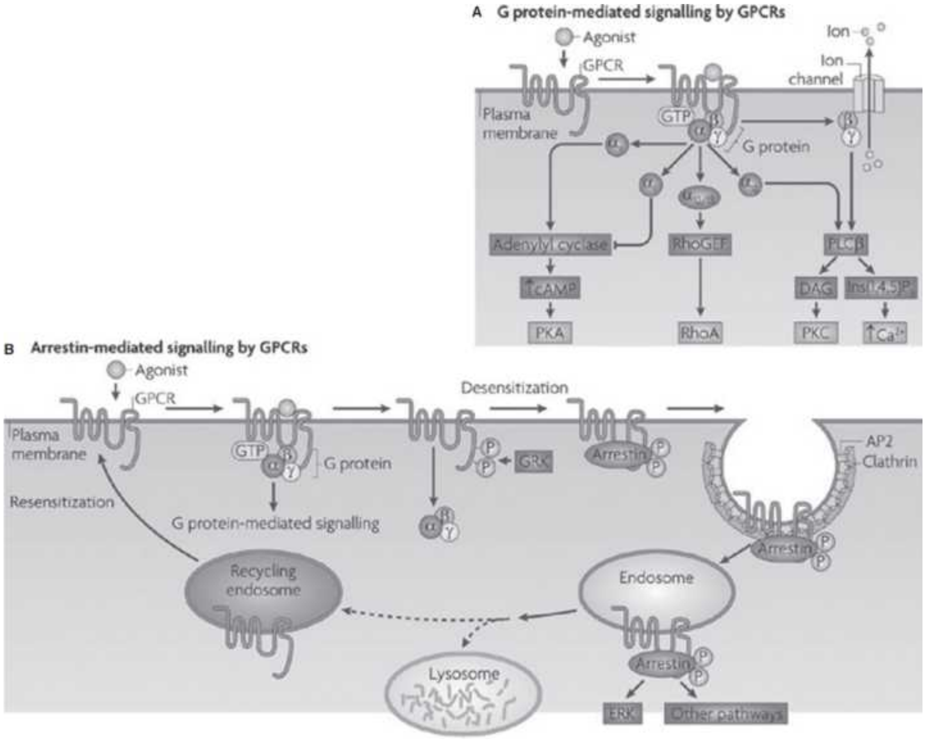
A G protein-mediated signalling by GPCRs
Agonist
GPCR
Plasma membrane
GTP
α
β
γ
G protein
Ion
Ion channel
Adenylyl cyclase
RhoGEF
PLCβ
↑cAMP
DAG
Ins(1,4,5)P3
PKA
RhoA
PKC
↑Ca2+
B Arrestin-mediated signalling by GPCRs
Agonist
GPCR
Plasma membrane
Desensitization
GTP
G protein
GRK
P
Arrestin
AP2
Clathrin
Resensitization
G protein-mediated signalling
Recycling endosome
Endosome
Lysosome
ERK
Other pathways

FIGURE 33.1a–10. G protein–coupled receptor (GPCR) internalization and biased agonism. In the long-established model of GPCR signaling, an agonist binds to extracellular and/or transmembrane regions of the receptor, leading to its activation of G protein that results from exchange of GDP for GTP on Gα subunit, creating active Gα and Gβγ subunits. Termination of receptor activation can occur through GPCR phosphorylation by GPCR kinases (GRKs) and concomitant GPCR association with arrestins, which interact with clathrin to internalize GPCRs in endosomes. Recruitment of arrestins to activated GPCRs can also lead to the initiation of distinct arrestin-mediated signaling pathways, including activation of the mitogen activated protein kinase extracellular signal-regulated kinase (ERK) pathway. This can be done absent involvement of G protein, and agonists engaging either G protein or arrestin pathways are referred to as biased. Subsequent to internalization, GPCRs can be trafficked to lysosomes, where they are ultimately degraded, or to recycling endosomes for reinsertion to the plasma membrane. (From Ritter SL, Hall RA. Fine-tuning of GPCR activity by receptor-interacting proteins. *Nat Rev Mol Cell Biol.* 2009;10(12):819–830, with permission from Springer Nature. doi: 10.1038/nrm2803.)

BIASED AGONISM. Endogenous agonists for most GPCRs) activate the receptor so that it engages both a G protein and β-arrestin. While the G protein pathway often generates cellular second messengers, the β-arrestin pathway evokes receptor internalization and a series of β-arrestin-mediated intracellular events. Over the past few years, compounds have been developed that select (are biased for) either the G protein pathway or the β-arrestin pathway. For psychiatry, of clear interest would be the development of compounds that ameliorate pain yet pose low potential for addiction (Fig. 33.1a–10).

ALLOSTERY. The discussion of GPCR agonists and antagonists presented above concerns molecules to that bind to the orthosteric ligand binding site on those receptors. Orthosteric agonists (and some antagonists) have been demonstrated to evoke specific structural changes in the receptor that engages G proteins and initiates the signal transduction cascade. Studies over the last several years have revealed the presence of allosteric sites on receptors. Molecules that bind to these sites, known as positive or negative allosteric modulators (PAM or NAM), might regulate the activity of a GPCR by either sustaining (PAM) or restricting (NAM) the active conformation of a receptor elicited by the orthosteric agonist.

Table 33.1a–1.
Mammalian Heterotrimeric G Protein α Subunits

Family	Subtype	Effectors
Gs(S)α	Gsα	Adenylyl cyclase, GTPase of tubulin
	Golfα	Adenylyl cyclase (olfactory)
Gi/oα	Go1α	Adenylyl cyclase inhibition (Gi, o, za) Go2
	Gi1-i3α	K^+ channels; Ca^{2+} channels inhibition (Gβγ)
	Ggust α	K^+ channels; Ca^{2+} channels inhibition (Gβγ)
	Gtα1/2	(transducin) cGMP-PDE in retina
Gq/11α	Gqα	Phospholipase C
	G11α	p63-RhoGEF
	G14α	
	G16α	
G12/13a	G12α	p115RhoGEF and related proteins
	G13α	

RECEPTOR DESENSITIZATION. Receptors are also capable of becoming desensitized. This process involves the attenuation of receptor response in the continued presence of an agonist. The process is quite different for ion channels and ionotropic receptors versus GPCRs. In the case of voltage-gated ion channels, there are processes that either desensitize or actually inactivate channels, and some of these events result from second messengers created subsequent to the activation of GPCRs. Similarly, ionotropic receptors are often modified, either by altering affinity for agonists or by altering channel characteristics. Again, this is due to second messengers, often through phosphorylation events.

GPCRs, however, have specific accessory proteins that mediate their desensitization. During the process, G protein receptor kinases (GRKs), an enzyme family, phosphorylate receptors in the activated (agonist-bound) conformation and these receptors become a target for β-arrestins, which herd them into clathrin-coated pits from whence they internalize. Once internalized, they either recycle back to the surface or are degraded. Figure 33.1a–10 illustrates the process. Of note is the evidence that some GPCRs might either continue to signal while internalized or become activated while localized within the cell.

Heterotrimeric G Proteins. Table 33.1a–1 reveals the current list of G protein α subunits. Currently, there have been identified 17α, 5β, and 13γ genes in humans. Alternative splicing increases the number of available proteins. Each of these G proteins interacts with specific effector molecules. These can be enzymes like adenylyl cyclase, which makes cAMP; phospholipase C, which generates lipid messengers and increases intracellular calcium; or voltage-gated ion channels (particularly K^+ or Ca^{2+}), which are modified by Gβγ subsequent to activation of Gα.

THE ACTIVATION CYCLE. Prior to binding agonists, GPCRs exist in a complex with heterotrimeric G proteins (αβγ subunits). The latter two exist as a “fused complex,” and the former releases GDP and binds GTP upon receptor activation. This shifts the Gα to an active conformation and alters the association between Gα and Gβγ in such a manner that both engage effector molecules. Hydrolysis of GTP by a GTPase intrinsic to the α subunit terminates the activation cycle (Fig. 33.1a–11).

RGS Proteins. For many Gα proteins, the GTPase is relatively slow, and regulators of G protein signaling (RGS) proteins facilitate the process. RGS proteins are a family of GTPase-activating proteins with conserved RGS domains (also called the RGS box) of about 120 amino-acid residues in length. RGS proteins interact directly with a subunit of heterotrimeric G proteins to catalyze GTP hydrolysis by enhancing the intrinsic GTPase activities of Gα, thus converting G proteins from a GTP-bound active state to a GDP-bound inactive state and terminating G protein-mediated signaling. While most G protein α subunits have been demonstrated to have their GTPase activated by a member of this large family of proteins, Gsα may be excluded. There are two sorting nexins that associate with Gsα (SNX13 and SNS14) that bind specifically with Gsα and are also known as RGSPX1 and RGSPX2, respectively. It is not clear that these proteins activate GTPase, and RGSPX2 clearly lacks this effect. Curiously, RGSPX2, a brain-specific protein, helps mediate the trafficking of Gsα from the cell membrane and also specifically modulates the cell-surface expression of the Gs-associated 5HT6 receptor.

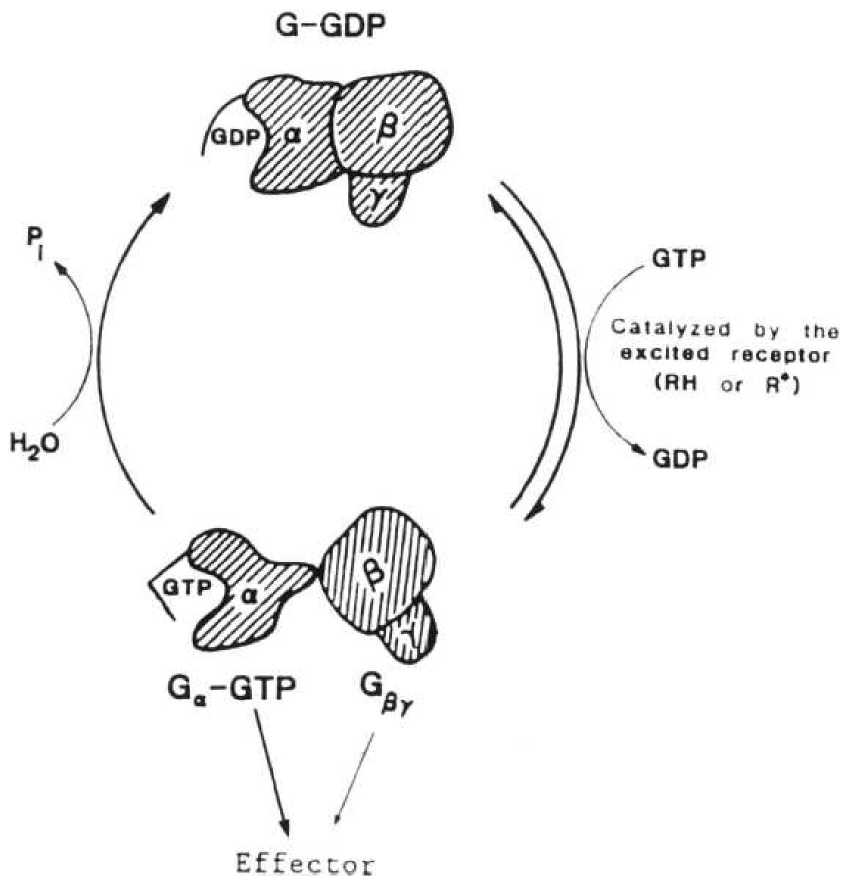

FIGURE 33.1a–11. Activation cycle for G protein heterotrimers. Activation of the GPCR by agonist (described in Fig. 33.1a–10) evokes release of GDP by G protein α subunit, allowing the binding of GTP. The GTP bound conformation of Gα activates Effector molecules such as adenylyl cyclase of Phospholipase C. The activation of Gα is terminated by the hydrolysis of GTP for GDP. While the GTPase is intrinsic, there are molecules, such as RGS proteins, that facilitate the process.

Bacterial Toxins and G Proteins. Two bacterial toxins are known to modify G proteins by covalently ADP-ribosylating select amino acids within those G proteins. Cholera toxin inactivates the intrinsic GTPase of Gsα which results in sustained (permanent) activation of Gs and subsequent hyperactivation of adenylyl cyclase. Pertussis toxin primarily acts on Gi and Go proteins and uncouples these proteins from the receptors that activate them. Since a prominent effect of Gi/o is inhibition of adenylyl cyclase, pertussis toxin will attenuate this inhibition with the net effect of increasing cAMP.

Microdomains in the Plasma Membrane and GPCR Signaling

The plasma membrane, while comprised of lipids and proteins, enjoys a more complex organization than would be predicted from proteins floating in a lipid bilayer. Membranes are highly compartmentalized and are rife with domains of highly ordered and less ordered membrane lipids. The more highly ordered regions tend to be cholesterol-rich, and these domains are normally associated with cytoskeletal proteins as well. The term used to describe these areas is “lipid rafts,” and they are thought to range in diameter between 20 and 100 nM. Some of these structures contain the protein caveolin and form invaginated structures known as caveolae (little caves). At this point, these structures are conceptual since they are below the level of resolution of the light microscope. Regardless, they can be manipulated experimentally, and it now appears that neurotransmitter action is, depending on the ligand and second-messenger pathway, facilitated or inhibited in caveolae and lipid rafts. Figure 33.1a–12 shows the postulated organization of signaling molecules into lipid rafts and caveolae. These membrane microdomains have been suggested to be important sites of action for antidepressants and may be instrumental in evoking the disrupted signaling seen in depression.

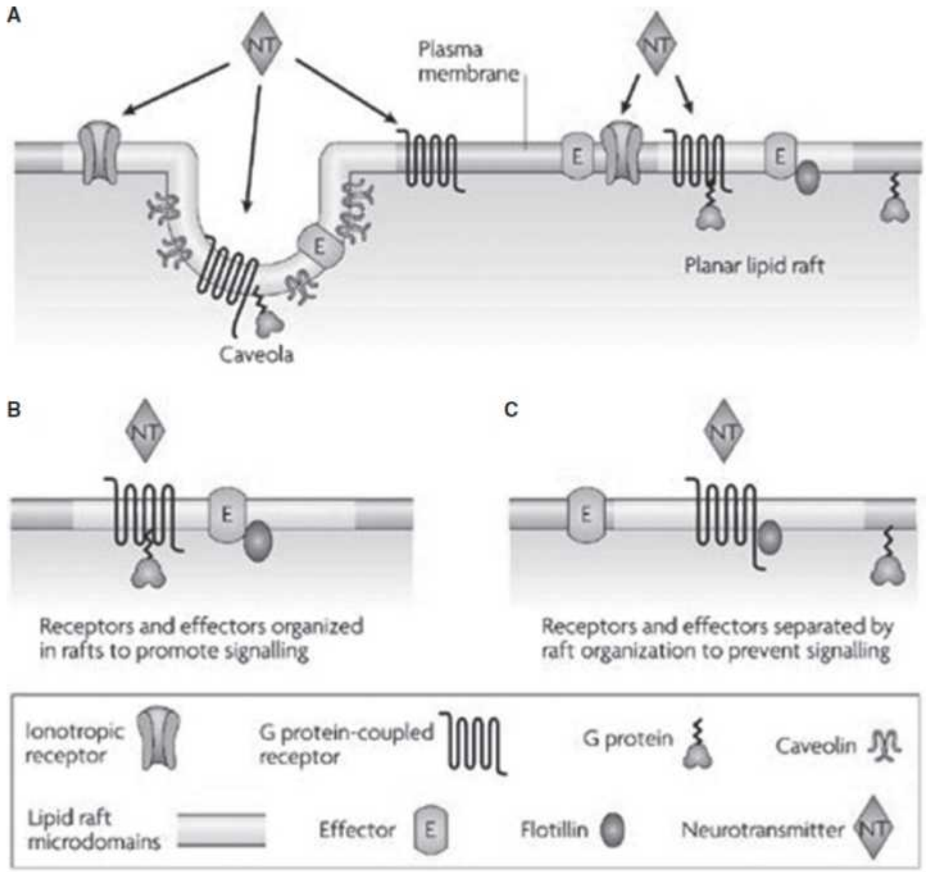
A
NT
Plasma membrane
NT
E
E
E
Planar lipid raft
Caveola
B
NT
E
Receptors and effectors organized in rafts to promote signalling
C
NT
E
Receptors and effectors separated by raft organization to prevent signalling
Ionotropic receptor
G protein-coupled receptor
G protein
Caveolin
Lipid raft microdomains
Effector
E
Flotillin
Neurotransmitter
NT

FIGURE 33.1a–12. Lipid raft microdomains and membrane organization of neurotransmitter signaling molecules. **A:** There are two common raft domains in mammalian cells: planar lipid rafts and caveolae. Both possess a similar lipid composition. Planar rafts are essentially continuous with the plane of the plasma membrane and lack distinguishing morphologic features. By contrast, caveolae are small, flask-shaped membrane invaginations of the plasma membrane that contain caveolins. Caveolins and flotillin can recruit signaling molecules into lipid rafts. Many neurotransmitter receptors (both ionotropic and G protein–coupled), G proteins, and signaling effectors such as second-messenger-generating enzymes are found in lipid rafts. Neurotransmitters might activate receptors that are located both within and outside lipid rafts. **B:** The lipid raft signaling hypothesis proposes that these microdomains organize, spatially, signaling molecules at the membrane, perhaps in complexes, to promote interactions that are necessary for signal transduction. **C:** Alternately, lipid rafts might inhibit interactions by separating signaling molecules, disrupting complexes, and damping neurotransmitter response. (From Allen JA, Halverson-Tamboli RA, Rasenick MM. Lipid raft microdomains and neurotransmitter signalling. *Nat Rev Neurosci*. 2007;8:128–140, with permission from Springer Nature. doi: 10.1038/nrn2059.)

G Protein Effectors

A vast array of cellular processes is mediated by G proteins. These molecules can be either activated or inhibited by G proteins, and either α or βγ subunits have been demonstrated to function as activators. A list of some of these G protein-mediated events is given in Table 33.1a–1, and a scheme is shown in Figure 33.1a–13.

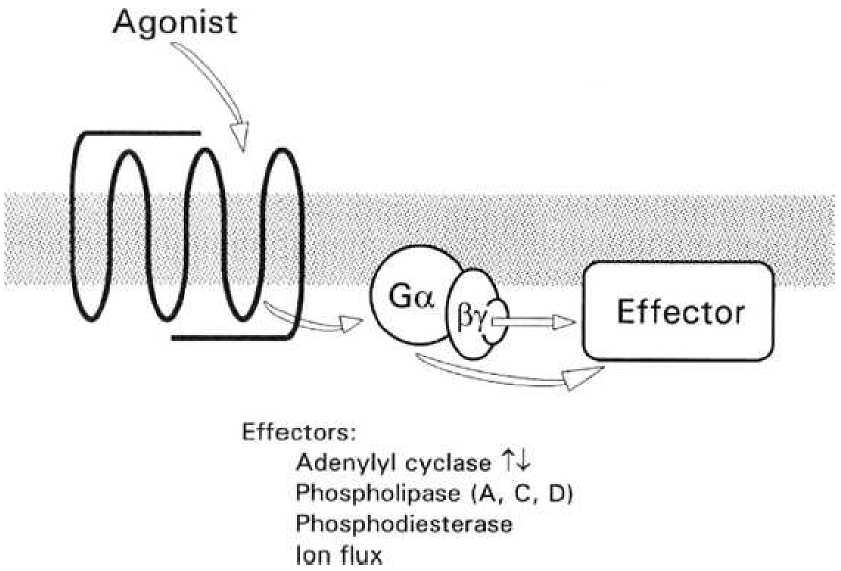

FIGURE 33.1a–13. Coupling of G proteins to receptors and effectors. The activation of a GPCR by agonist allows-activation of the G protein (the α subunits exchange guanosine diphosphate [GDP] for guanosine 5’-triphosphate [GTP]) and subsequentactivation of an effector molecule by α or βγ subunits. Several types of effector are listed. (Reprinted with permission from Felig P, Frohman LA. *Endocrinology and Metabolism*. 4th ed. McGraw Hill Education; 2001.)

Table 33.1a–2.
Adenylyl Cyclase Isoforms and Their Regulation

Regulator	Effect	Cyclase Isoform
Gsα	Stimulatory	All
Giα	Inhibitory	I, V, VI
GOα	Inhibitory	I
Gzα	Inhibitory	I, V
Gβγ	Inhibitory	I
Gβy	Stimulatory[a]	II, IV, VII
Forskolin	Stimulatory[b]	I–VIII
Ca^{2+}/calmodulin	Stimulatory	I, III, VIII
Ca^{2+}	Inhibitory	V, VI, IX[c]
PKC	Stimulatory	I, II, III, V, VII
PKC	Inhibitory	IV
Calmodulin Kinase	Inhibitory	III

[a] Enzyme must first be stimulated by Gsα.

[b] Very weak stimulation for type IX.

[c] Requires calcineurin.

Adenylyl Cyclase

This is the effector system that was being studied by Sutherland and Rall in their pioneering work on the activation of glycogen phosphorylase—work that led to the identification of cAMP and the concept of intracellular messenger molecules. Adenylyl cyclase, which generates cAMP from adenosine triphosphate (ATP), is a family of nine membrane-bound enzymes (plus a soluble version), each with a different mode of regulation. A common feature of all the adenylyl cyclase enzymes is activation by Gsα and (to varying degrees) by the diterpene, forskolin (Table 33.1a–2). While Giα (three genes encode different forms of protein) was called “i” because it was inhibitory to adenylyl cyclase, only a few forms of adenylyl cyclase are inhibited directly by Giα. Calcium can act to stimulate some forms of the enzyme (in conjunction with calmodulin), and other forms are inhibited by calcium. While all forms of adenylyl cyclase are found in the brain, these calcium-regulated types may be particularly relevant for psychopharmacology, as neurotransmitters that increase calcium (e.g., glutamate or acetylcholine) can also modify cAMP indirectly.

Phospholipases

Phospholipases responsive to G proteins fall into three families: phospholipase A, phospholipase C, and phospholipase D (Fig. 33.1a–14). The products of phospholipase A are metabolites of arachidonic acid and are known as eicosanoids. The eicosanoids fall into the subgroups of prostaglandins and leukotrienes, and these molecules function as extracellular signals. The next group of phospholipids includes those that result from the breakdown of diacylglycerol and inositol phosphates, and these are the result of the actions of the phospholipase C enzymes. Gq and G11 activate phospholipase C, which has the net effect of increasing intracellular cellular calcium, both through the release of Ca^{2+} from intracellular stores and increasing capacitive calcium entry. 5HT2A receptors activate phospholipase C through a Gq intermediary. Finally, phospholipase D, which uses phosphatidylcholine as a substrate, can generate diacylglycerol and phosphatidic acid as well as lysophosphatidic acid. This latter molecule also has extracellular signaling properties.

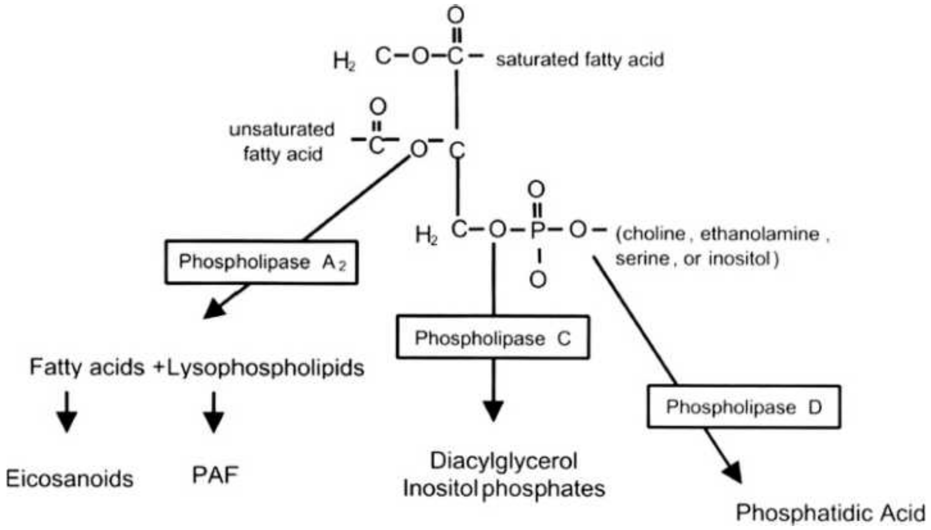

FIGURE 33.1a–14. Substrates for, and products of, G protein–coupled phospholipases. PAF, platelet activating factor. (Reprinted with permission from Felig P, Frohman LA. *Endocrinology and Metabolism*. 4th ed. McGraw Hill Education; 2001.)

G Protein–Activated Ion Channels

Direct activation of potassium channels by G proteins is important for the agonist-induced, rapid inhibition of membrane excitability in nervous tissue, and is essential for modifying both excitation and neurotransmitter release. As noted above, activation of GPCRs induces activation of G proteins, and both Gα and Gβγ can activate effector molecules, in this case, channel proteins. These channels (G protein–gated inward rectifier Kþ channels, GIRK) have been shown to bind directly and become activated by Gβγ. They are strictly voltage-independent, and channel gating requires association with intracellular proteins. Just as prominently (and also resulting in an inhibition of a voltage-activated process), Gβγ subunits inhibit voltage-gated Ca^{2+} channels, attenuating neurotransmitter release. In these processes, the G protein itself is the second messenger (Fig. 33.1a–15). Thus, the activation/inhibition of ion channels is somewhat more rapid than the generation of a second messenger and subsequent modification of channel function by phosphorylation.

INTRACELLULAR SIGNALING PATHWAYS

cAMP is synthesized by adenylyl cyclase and degraded by cAMP phosphodiesterase (PDE), which causes cAMP to be hydrolyzed to 5′AMP. Initial hormonal signals are more likely to regulate cAMP by the activation or inhibition of adenylyl cyclase than they are by altering PDE, but this is primarily because the receptors are coupled to G proteins that then directly stimulate or inhibit adenylyl cyclase. However, activation of a Gq-coupled receptor, an NMDA receptor, or a nAChR will increase cellular Ca^{2+}, which will activate calcium-sensitive variants of PDE, decreasing cAMP.

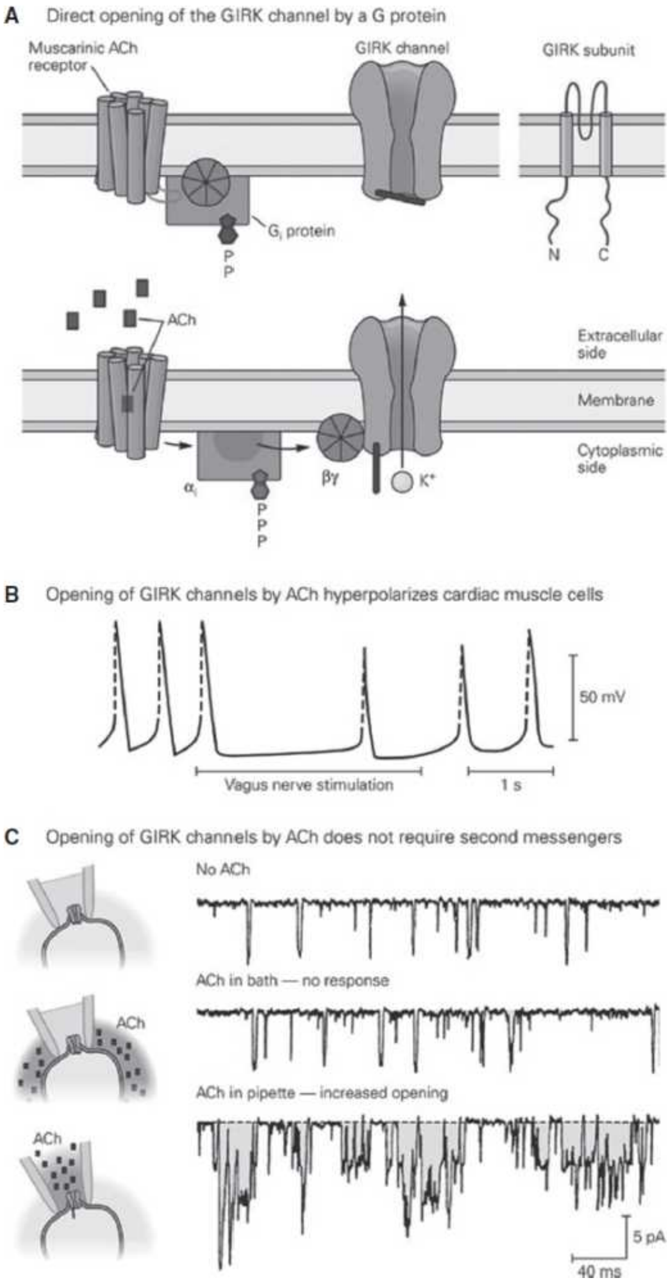
A
Direct opening of the GIRK channel by a G protein
Muscarinic ACh receptor
GIRK channel
GIRK subunit
G$_i$ protein
P
P
N
C
ACh
Extracellular side
Membrane
Cytoplasmic side
α$_i$
βγ
K$^+$
P
P
P
B
Opening of GIRK channels by ACh hyperpolarizes cardiac muscle cells
50 mV
Vagus nerve stimulation
1 s
C
Opening of GIRK channels by ACh does not require second messengers
No ACh
ACh in bath — no response
ACh
ACh in pipette — increased opening
ACh
5 pA
40 ms

FIGURE 33.1a–15. Direct opening of an ion channel by a G protein. In this instance, activation of an mAChR by acetylcholine allows Gβγ to bind to and open a K$^+$ channel known as a G protein–coupled inwardly rectifying K$^+$ channel (GIRK). Similarly, Gβγ can bind to and inhibit voltage-gated Ca^{2+} channels.

Regulation of cAMP and Ca^{2+} signaling begins with the activation of specific protein kinases sensitive to those messengers. For cAMP, the primary kinase is protein kinase A (PKA) (Fig. 33.1a–16), while Ca^{2+} activates either calcium/calmodulin-activated kinases (CaM kinase) or protein kinase C, which is activated by Ca^{2+} and diacylglycerol. Each of these has a series of targets that engage a series of intracellular pathways, including modification of ion channels, modification of cytoskeletal elements (and fine cellular structure), and modification of neurotransmitter synthesis, transport, release, or catabolism, as well as modification of gene expression via cyclic AMP response element (CREB) phosphorylation (Fig. 33.1a–17).

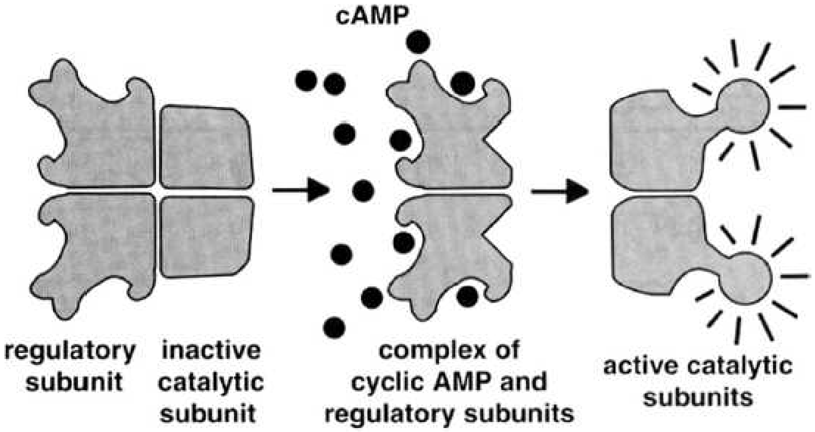
cAMP
regulatory subunit
inactive catalytic subunit
complex of cyclic AMP and regulatory subunits
active catalytic subunits

FIGURE 33.1a–16. Activation of cyclic adenosine monophosphate (cAMP)-dependent protein kinase (PKA). Binding of cAMP to the regulatory subunits induces a conformational change, causing these subunits to dissociate from the complex, thereby activating the catalytic subunits. Each regulatory subunit has two cAMP-binding sites, and the release of the catalytic subunits requires the binding of more than two cAMP molecules to the tetramer. This greatly sharpens the response of the kinase to changes in cAMP concentration. There are at least two types of A kinase in most mammalian cells: type I is mainly in the cytosol, whereas type II is bound via its regulatory subunit to the plasma membrane, nuclear membrane, and microtubules. In both cases, however, once the catalytic subunits are freed and active, they can migrate into the nucleus (where they phosphorylate gene-regulatory proteins such as CREB; see Fig. 33.1a–17). (Reprinted with permission from Felig P, Frohman LA. *Endocrinology and Metabolism*. 4th ed. McGraw Hill Education; 2001.)

As a rule, intracellular messages require a specific termination switch, and phosphoprotein phosphatases served this purpose for the phosphorylations described above. Some of these enzymes are specific for PKA (protein phosphatase IIb) or CaM kinase (calcineurin).

Regulation of adenylyl cyclase has been discussed above. The regulation of PDEs is quite complex, and much of it is due to the exquisite and unique subcellular localization of a myriad of PDE species. Further, the localization of cAMP-activated protein kinases (PKA) and a variety of PDEs is regulated by a family of proteins known as A kinase anchoring proteins (AKAPs). These proteins, often referred to as cellular scaffolds, provide a spatial dimension to second messenger signaling as well as a temporal discretion provided by the close physical relationship of the relevant proteins.

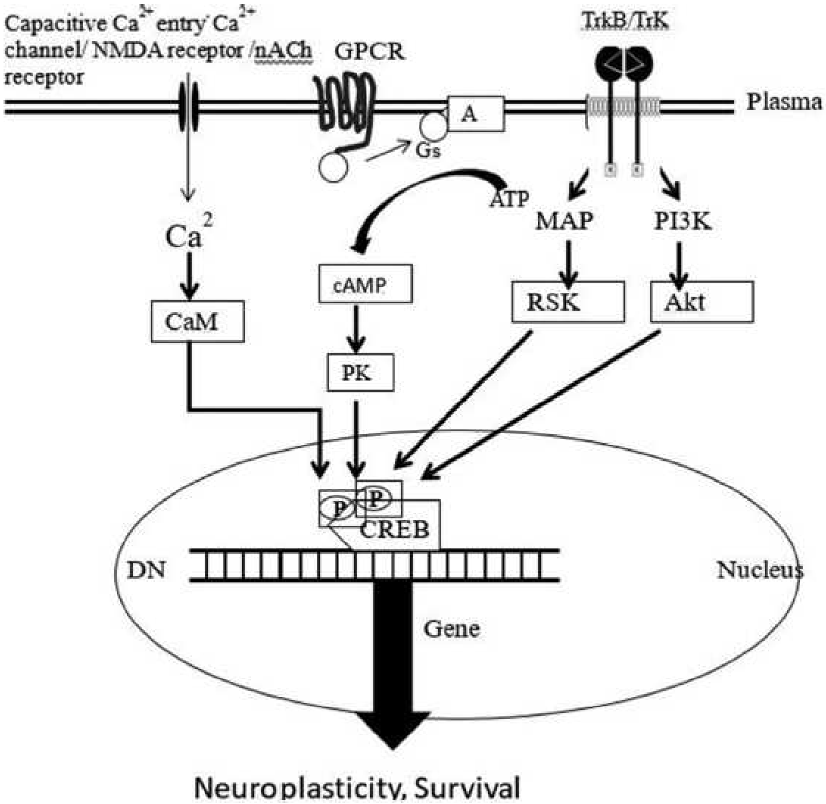
Capacitive Ca2+ entry Ca2+ channel/ NMDA receptor /nACh receptor
GPCR
TrkB/TrK
Plasma
A
Gs
ATP
MAP
PI3K
Ca2
CaM
cAMP
PK
RSK
Akt
P
P
CREB
DN
Nucleus
Gene
Neuroplasticity, Survival

FIGURE 33.1a–17. cAMP and Calcium signaling orchestrate gene expression. Mechanisms for cAMP generation, calcium entry, and regulation of nuclear and intracellular calcium-responsive proteins are demonstrated. This scheme also suggests an interface for the action of growth factor receptors, GPCRs, and Ca^{2+}.

Epac, Another cAMP Target

The default cAMP pathway is PKA, but it is noteworthy that other cellular proteins are regulated by cAMP-independent of phosphorylation. One of these cAMP targets is a family of proteins known as Epac (exchange proteins activated by cAMP). These elicit guanine nucleotide exchange and subsequent activation of a number of small G proteins involved in cell growth and synaptogenesis. Furthermore, there are ion channels (particularly important in olfaction) that are directly opened subsequent to the binding of cAMP.

INTEGRATION OF NEUROTRANSMITTER SIGNALS

Most diagrams of synaptic signaling, including those in this chapter, depict a single neurotransmitter engaging a single receptor, which then evokes a programmed series of responses through a set pathway. While as a didactic principle this is useful, neurotransmitter signaling, both at extracellular pathways and within cells, enjoys a far greater complexity. First, hundreds (even thousands) of axons converge upon a single dendrite (or even on a single dendritic spine). Second, a single transmitter, acting through a myriad of receptor types, could evoke multiple intracellular pathways. This is illustrated for serotonin in Table 33.1a–3. Finally, each of these pathways is interdependent, as illustrated by Ca^{2+} and cAMP signaling pathways in Figure 33.1a–18.

GENERAL ISSUES IN CLINICAL PSYCHOPHARMACOLOGY

Reflecting the expanded role of psychopharmacology, as well as the variety and the sheer number of psychotropic drugs available, the sections on psychotropic drugs that follow represent a virtual textbook within a textbook. These sections are also complemented by discussions in other chapters that focus on the treatment of specific disorders and special patient populations. This section is intended as a review of important principles underlying the selection and use of psychotropic medications. It provides a review of common issues that arise in the clinical setting, as well as considerations for more effective use of available agents. Every effort has been made to provide a comprehensive discussion of the properties of the drugs, as well as guidance on their most appropriate use. Nevertheless, it needs to be emphasized that psychopharmacology is a rapidly evolving field and that many patients experience idiosyncratic or paradoxical responses to treatment.

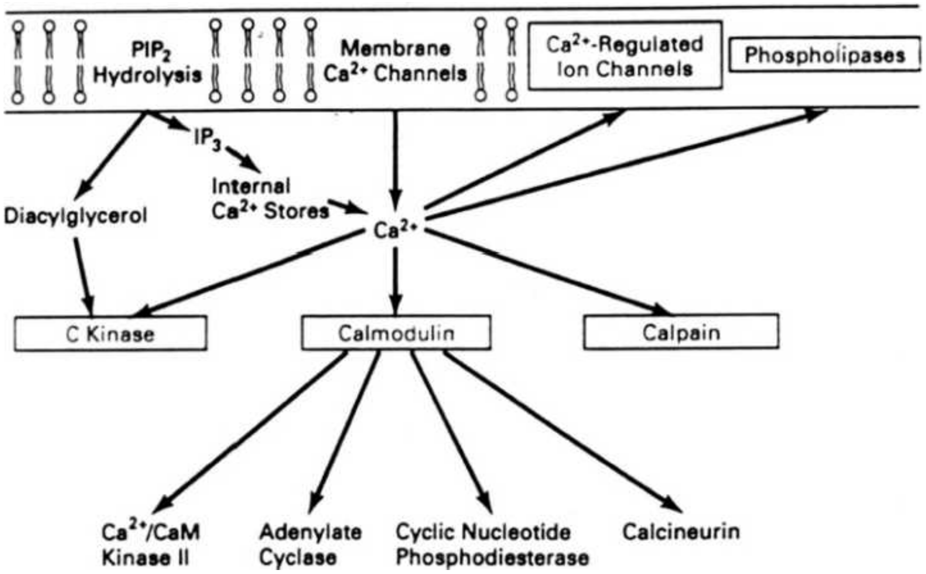
PIP₂ Hydrolysis
Membrane Ca²⁺ Channels
Ca²⁺-Regulated Ion Channels
Phospholipases
IP₃
Internal Ca²⁺ Stores
Diacylglycerol
Ca²⁺
C Kinase
Calmodulin
Calpain
Ca²⁺/CaM Kinase II
Adenylate Cyclase
Cyclic Nucleotide Phosphodiesterase
Calcineurin

FIGURE 33.1a–18. Interface among second messenger systems: many roles for calcium. Several conduits for Ca^{2+} signaling are shown and the interactions among them and cAMP signaling pathways are described in the text.

Table 33.1a–3.
Psychotropic Drugs with Postmarketing Black Box Warnings

Drug	Psychiatric Application	Basis for Warning
Clozapine (Clozaril)	Refractory schizophrenia	Myocarditis
Droperidol (Inapsine)	Sedation of agitated patients	Torsades de pointes or QT and QTc prolongation, or both, and other cardiovascular events
Thioridazine (Mellaril)	Psychosis	Torsades de pointes or QT and QTc prolongation, or both, and other cardiovascular events
Divalproex (Depakote)	Acute mania	Hemorrhagic pancreatitis and hepatic failure
Nefazodone (Serzone)	Depression	Acute hepatic failure
Pemoline (Cylert)	Attention-deficit/hyperactivity disorder	Acute liver failure
Antidepressants (in patients <24 years old)	Depression	Increased risk of suicidality
Antipsychotics (for dementia-related psychoses)	Psychosis	Increased risk of mortality

History

For centuries, people have used plant extracts and naturally occurring substances to influence mood and anxiety, modulate sleep and arousal, produce euphoria, and alter consciousness. Notable examples include alcohol, coffee, tea, opium preparations, bromides, and hallucinogens. The use of cocaine-containing soothing syrups and elixirs was common. In the nineteenth and early twentieth centuries, several types of *sedatives* were developed. Some of these compounds, chloral hydrate (Aquachloral Supprettes), and barbiturates, are still in use. From the 1930s to the 1950s, the most effective biologic treatment of a mental disorder was electroconvulsive therapy (ECT). While ECT remains an effective treatment, the image of the so-called “shock therapy” is unpalatable to many patients. In 1931, two psychiatrists in India, Ganpath Sen and Kartik Chandra Bose described the ability of *Rauwolfia serpentina* to reduce symptoms of psychosis, but this report was overlooked by Western medicine for many decades. When rediscovered, reserpine (Serpalan) became one of the first agents to offer a benefit to patients with schizophrenia. Understanding of reserpine’s effects on monoamine neurotransmitters spurred further research into the involvement of norepinephrine, serotonin, and dopamine in the treatment of psychiatric disorders and the synthesis of drugs that targeted these agents. Similarly, published findings on the antimanic properties of lithium salts appeared in 1949 but were neglected by most psychiatrists. It was more than two decades before lithium became accepted as the standard treatment for bipolar disorder.

The field of psychopharmacology evolved rapidly during the second half of the 20th century. It was the nearly simultaneous discoveries of several classes of medications that began the transformation of psychiatric treatment from a psychoanalytic to a biologic orientation. The first breakthrough occurred between 1950 and 1952 with the observation by Jean Delay and Pierre Deniker that chlorpromazine—a presurgical antihistamine—brought about dramatic control of psychotic symptoms associated with mania and schizophrenia. The effect of the drug was described as "miraculous" and "astonishing." Chlorpromazine was released for clinical use in 1954 and, within 8 months, was administered to more than 2 million patients. Patients with long histories of hospitalization owing to disorganized thinking and bizarre behavior demonstrated sufficient improvement to be discharged from their overcrowded institutions. For a time, there were expectations that chlorpromazine and its analogs, then referred to as *tranquilizers,* could eradicate or permanently cure schizophrenia. Although this turned out to be a false hope, newer versions of these drugs improve the quality of life of many patients, enabling them to be productive and live in the community.

In 1952, iproniazid (Marsilid), a drug being tested as a treatment for tuberculosis, was observed to cause a dramatic elevation in mood. Over the next few years, psychiatrists came to recognize this antidepressant action of iproniazid and similar isonicotinic hydrazide analogs. Cases of unexpected hypertensive episodes associated with monoamine oxidase inhibitors (MAOIs) were reported soon after the introduction of these drugs. Some patients died of strokes. An explanation for this so-called cheese reaction did not emerge until 1963 in an article by Barry Blackwell.

Shortly after the discovery of MAOI antidepressant activity, imipramine (Tofranil), a derivative of chlorpromazine, was observed to alleviate depressive symptoms. Imipramine, the first tricyclic antidepressant (TCA), and iproniazid, the first MAOI, became available for clinical use in 1957. Within 1 year of its release, more than 400,000 patients received imipramine for depression.

Despite the progress that has been made in more than a half-century, no antidepressants have surpassed the TCAs and MAOIs in efficacy. Nevertheless, these drugs have never lived up to their full potential because of troublesome side effects and concerns about safety. The dietary restrictions reduced the acceptability of MAOI treatment. The anticholinergic side effects of TCAs and their toxicity in overdose also had a restraining effect on their use.

The availability of meprobamate (Equanil) in 1955 marked the first time that treatment of anxiety had a relatively wide margin of safety. Compared to barbiturates, meprobamate caused less dose escalation and a less severe withdrawal syndrome. By the end of that decade, an even better tolerated and safer alternative emerged. In 1957, the benzodiazepine chlordiazepoxide (Librium) was found to produce sedative, muscle relaxant, and anticonvulsant properties but without significant risk of serious toxicity in overdose or during withdrawal. By 1960, chlordiazepoxide was marketed, soon followed by diazepam (Valium). It is worth noting that chlordiazepoxide was introduced into clinical practice just 2.5 years after its pharmacologic studies were started.

In 1958, Paul Janssen synthesized another important chemical family of compounds used to treat psychoses—the butyrophenones. The prototype of this class—haloperidol (Haldol)—remains one of the most popular drugs used for the management of psychosis. The first typical antipsychotic was synthesized in the late 1950s, but it was another of Paul Janssen's compounds, risperidone (Risperdal), that became the first widely prescribed *atypical* neuroleptic or serotonin-dopamine antagonist.

Attitudes toward psychotropics were transformed in the early 1980s after reports were published that alprazolam (Xanax) could suppress panic attacks. Until that time, only TCAs and MAOIs had been shown to control panic symptoms, and benzodiazepines were not thought to be effective in treating this severe form of anxiety.

Based on the newly gained understanding that TCAs block not only norepinephrine reuptake but also serotonin, Arvid Carlsson and his colleagues set out to develop a selective inhibitor of serotonin reuptake. Zimelidine, the first of these selective serotonin reuptake inhibitors (SSRIs), was developed in the early 1970s and was found to be a potent antidepressant. It was marketed in Europe but was quickly withdrawn from clinical use because of a rare neurologic side effect. It was not until the introduction of fluoxetine (Prozac) late in 1987 that an SSRI became available in the United States. Early clinical experience with fluoxetine was followed by the surprising realization by patients and physicians that the drug sometimes produced dramatic responses without the same type of tolerability problems associated with the TCAs. Quickly, first-line treatment of depression quickly shifted from the TCAs to fluoxetine and other SSRIs. Even around the start of the 21st century, the majority of new prescriptions for antidepressant treatments were for SSRIs.

The first psychotropic drugs were thus discovered through observations of behavioral effects in animals or humans. As the mechanisms underlying the actions of these drugs have been determined, it has facilitated rational drug development of new compounds, such as the SSRIs, with more selective action. It also has led to the identification of disorders that are preferentially responsive to specific agents. Patients now have many more treatment options; the drugs used are better understood. One of the more remarkable transformations in the field of psychopharmacology has been the degree to which psychotropic drugs are now routinely prescribed by nonpsychiatrists.

Two large studies have helped to provide a perspective on how difficult it is to differentiate the growing number of psychotropic agents in terms of their overall effectiveness. The findings of these studies, the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) and the Sequential Treatment Alternatives to Relieve Depression (STAR*D) trial, point out that patient responses to medications tend to be unpredictable, that no interventions are effective in all patients with a specific disorder, and that tolerability and safety often become the determinant of treatment success or failure.

The STAR*D trial was a large, federally funded study that investigated treatment effectiveness in a “real-world” sample of depressed outpatients treated in psychiatric and primary care practices. There were 3,671 participants in the trial. It found that remission was more common among patients in their first or second treatment trial (37% and 31%, respectively) than among those requiring three or four trials (14% and 13%, respectively). Patients needing more treatment steps tended to have more chronic depression and more coexisting psychiatric and general medical disorders. Overall, 67% of patients who completed up to four treatment steps reached remission. Results for STAR*D revealed higher relapse rates for those who required more initial treatment steps. Patients were allowed to choose among groups of subsequent treatments, which included cognitive therapy alone or combined with medication, as well as several different antidepressants used alone or in combination. The study failed to show any advantage in terms of switching or combining drugs for one strategy over another. The study found higher relapse rates among patients who failed to reach full remission and among those who went through a higher number of treatment steps to reach remission.

The CATIE study, funded by the National Institutes of Health’s National Institute of Mental Health (NIMH), analyzed the effectiveness of an older (first-generation) antipsychotic medication, perphenazine, versus the newer (second-generation) medications during the initial treatment. It found that the older medication might be more cost-effective for some patients with schizophrenia. Perphenazine was less expensive and no less effective than the newer drugs used in the trial. In phase 1 of the trial, patients were randomly assigned to receive olanzapine, perphenazine, quetiapine, risperidone, or ziprasidone for up to 18 months. Clozapine was included in phase 2 of the study. The study found that olanzapine had the longest time to discontinuation in phase 1 but caused the most weight gain and metabolic syndrome. Perphenazine did not significantly differ in overall effectiveness from quetiapine, risperidone, and ziprasidone. There were no differences in neurocognitive or psychosocial functioning in response to medications.

Classification

Traditionally, the classification and categorization of psychotropic medications have been based on the primary indication for these medications. With each subsequent edition of this textbook, the classification and categorization of psychotropic drugs have changed. This is not merely a consequence of the development of new compounds. It is also the result of advances in the understanding of how these compounds produce their clinical effects. It was not long ago that psychotropic drugs were broadly defined as either stimulants or sedatives. The term *tranquilizer* was applied to antianxiety/sedative drugs (minor tranquilizers) and antipsychotics (major tranquilizers). At first, the only antidepressants were TCAs or MAOIs. However, when newer antidepressant drugs emerged in the 1970s and 1980s, they were labeled as second- or third-generation antidepressants.

Defining drugs by their major indications is problematic. For example, the standard categories of drugs are (1) antipsychotic drugs or neuroleptics used to treat psychosis, (2) antidepressant drugs used to treat depression, (3) mood stabilizers used to treat bipolar disorder, (4) antianxiety or anxiolytic drugs used to treat anxious states, (5) hypnotic agents to promote sleep, and (6) psychostimulants to treat attention-deficit disorder. As a rule, clinicians think of drugs primarily in terms of their major clinical application. However, few psychotropic drugs are diagnosis-specific, and most of the newer drugs assume multiple therapeutic roles over time. Benzodiazepine “anxiolytics,” for example, are also very effective as hypnotics. Antidepressant drugs have a wide spectrum of action. Tricyclics like amitriptyline, which have significant histamine H1 antagonism, have been used in urticarial and other dermatologic disorders. Agents such as the SSRIs and venlafaxine (Effexor), all originally used to treat depression, have gained FDA approval as treatments for disorders as diverse as panic disorder, generalized anxiety disorder, social anxiety disorder, obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), and premenstrual dysphoric disorder and chronic pain. Similarly, all atypical antipsychotic agents have become commonly used as mood stabilizers and as augmentation agents to treat depression and OCD. Such drugs as clonidine (Catapres), propranolol (Inderal), verapamil (Isoptin), and pregabalin (Lyrica), all not officially recognized as psychotropic agents, are widely used to manage a variety of psychiatric disorders. As a result, there is a growing movement to reclassify psychotropic drugs according to the mechanism of action rather than by indication. The major neuropsychopharmacology societies around the world have come together to create a pharmacologically driven nomenclature for psychiatric drugs. While this system, dubbed the Neuroscience-based Nomenclature (NbN), has yet to be widely accepted and adopted, it suggests a more rational way of conceptualizing these medications.

ANTIDEPRESSANT PROPERTIES OF PSYCHEDELICS

In his 2018 book, “How to Change Your Mind,” Michael Pollan presents a lively history of psychedelics from both a recreational and therapeutic vantage point. While that book covers a broad array of psychedelics, this chapter focuses on two compounds, psilocybin (and its active metabolite, psilocin) and lysergic acid diethylamide (LSD). For the former, psilocybin has been shown to confer long-lasting antidepressant efficacy following a single administration. In 2019, the FDA designated psilocybin as a “breakthrough therapy” to facilitate its investigation as a rapid-acting antidepressant. This designation accelerates the time scale for full approval, and Oregon voters have authorized therapeutic use. Psilocin, the active psilocybin metabolite, reaches high-levels in the central nervous system (CNS) and is responsible for perceived effects following psilocybin administration. Psilocybin and psilocin are potent agonists of 5-HT2A and 5-HT2C receptors but have some binding activity at all serotonin GPCRs. The Gαq coupled 5HT2A and 5HT 2C receptors have been studied extensively in an effort to understand psilocybin’s hallucinogenic action. Preclinical data from studies designed to understand the antidepressant and anxiolytic properties of psilocybin/psilocin suggests that 5-HT2A/C receptor activation is insufficient to explain psilocybin’s rapid and sustained antidepressant efficacy. Furthermore, clinical trials indicate some correlation between therapeutic effects and hallucinogenic action, but not enough to account for therapeutic success. So, while hallucinogenic effects may depend on 5-HT2 receptor activation, antidepressant effects may not. If the cellular mechanisms underlying psilocin’s antidepressant effects can be identified, it may be possible to develop novel rapid-acting antidepressants that share psilocin’s antidepressant action while avoiding its (not always unwelcome) hallucinogenic side effects. Psilocin, and related psychedelic compounds, evoke 5-HT2A/C receptor desensitization and suppress downstream G protein signaling. Leading to a long-lasting (a week or more) tolerance to psilocybin’s psychoactive effects. This suggests that a desensitized G protein signaling may actually contribute to prolonged antidepressant efficacy, a phenomenon observed recently for monoamine antidepressants. Although less intensively studied than psilocybin, LSD appears to share many of the characteristics, both on the cellular (involving increased neural plasticity) and clinical levels. While the exact mechanisms underlying psilocybin’s long-lasting effects on the brain have not been determined, understanding them may bode well for the development of future antidepressant therapies.

Ketamine

Ketamine will be covered in a separate chapter in this text, but we will add some comments on the properties of this drug and its confluence and contrast with other antidepressants. Ketamine was developed initially as an anesthetic and is still used widely in that capacity. About two decades ago, ketamine began its life as an antidepressant, and intravenous ketamine enjoyed success as a treatment for intransigent depression. The recent approval of nasal esketamine has simplified ketamine administration somewhat, but patients still require monitoring. A single administration of ketamine acts rapidly and provides alleviation of depression for as much as 2 weeks, often for those patients who have failed to respond to other antidepressants. Ketamine has NMDA receptor antagonist properties, but other NMDA receptor antagonists, including those with a binding site on the receptor similar to that of ketamine, lack antidepressant properties. It is suggested that NMDAR antagonism by ketamine increases the synthesis of BDNF, and increased synthesis of BDNF is a nearly universal finding in animal and cellular ketamine studies. However, in addition to the lack of effect of nonketamine NMDA antagonists, hydroxy nor ketamine analogs, which lack NMDA antagonist properties, show antidepressant effects. Furthermore, knockout of NMDA receptor does not attenuate antidepressant effects of ketamine in a cellular model, wherein ketamine mobilizes the G protein Gsa from lipid rafts, similar to most other antidepressants but much more rapidly, resulting in augmented cAMP signaling and increased BNDF production. Recent studies suggest that ketamine interacts directly with lipid rafts and receptors for BDNF. For the current edition of this textbook, ketamine will be considered an interesting antidepressant with rapid effects on intransigent depression that acts through a yet-to-be-determined mechanism.

REGULATORY ISSUES

In the United States, the FDA determines the safety and efficacy of medication. The FDA has the authority to approve a drug for clinical use if it is found to be “safe for use” and “effective in use.” The FDA is also charged with ensuring that product labeling is truthful and contains all information pertinent to the safe and effective use of that drug. This includes monitoring the postmarketing experience with a drug, proposing changes in the product information, and overseeing the way in which drugs are manufactured. Antidepressants, antipsychotics, hypnotics, mood stabilizers, and psychostimulants have been among the drugs with additional warnings due to postmarketing monitoring. In addition to the FDA, another level of government regulation is directed by the Drug Enforcement Agency (DEA), which classifies drugs according to their abuse potential. Clinicians are advised to exercise increased caution when prescribing controlled substances (control levels II through IV).

Prescribing Information

FDA-approved product information for marketed drugs appears as a package insert. It is also referred to as *product labeling.* The *Physicians’ Desk Reference* (PDR) is an annual collection of product information for prescription drugs. The product information for most drugs—not just psychotropic agents—contains an extensive list of potential side effects, drug interactions, the need for special monitoring, and restrictions for use. This information reflects the results of critical evaluations submitted to the FDA about the safety and efficacy of the drug for the approved indications. Postmarketing experience is added to the product information over time. The limitations of the prescribing information are evident in the many adverse events that are under-reported during clinical trials and the frequent use of doses of medications in clinical practice that exceed the recommended dosing limits. Product information is also widely available through online resources such as drugs.com and Epocrates.

Black Box Warning

The prescribing information for a drug sometimes contains a *black box warning* to alert physicians about potentially important safety information. The term describes a black box at the beginning of the package insert that encloses the warning. It is considered the strongest warning, being more serious than just bolded text. A black box warning may be included in the product information when a drug is approved by the FDA, or it can be added later. The presence of a black box warning deters some clinicians from using a drug as a first-choice agent. In some cases, manufacturers have decided against bringing a drug to market because of the competitive disadvantage that a warning label confers. A controversial example is the black box warning for the increased risk of suicidality in young patients taking antidepressants issued in 2004. The biostatistician Robert Gibbons and his colleagues noted that prescribing rates decreased and suicide rates increased since the publication of the warning. Studies such as this led to revised wording of the warning by the FDA so that clinicians would not be discouraged from using antidepressants when clinically indicated. In 2005, the FDA issued a black box warning about the increased risk of mortality with the use of atypical antipsychotics for dementia-related psychosis. This was expanded to include conventional antipsychotics in 2008.

The longer that a drug has been in clinical use, the better the understanding of its side effect profile. Even though new agents undergo extensive scrutiny for safety and tolerability, clinical trials do not typically capture the full range and severity of all side effects. Most clinical trials are of too short duration and lack sufficient power to detect the full extent of side effect risk. Approximately 10% of all drugs approved in the United States in recent years have acquired a black box warning after marketing or have been withdrawn. In terms of timing, one-half of black box additions occurred within the first 7 years after a drug was marketed. Psychotropic drugs with added black box warnings (Table 33.1a–3) include nefazodone (Serzone), pemoline (Cylert), droperidol (Inapsine), lamotrigine, clozapine (Clozaril), divalproex (Depakote), and thioridazine (Mellaril). In the last-named case, the warning was added after the drug had been in use for many decades. In the context of safer alternatives, the benefits of thioridazine no longer outweighed the risk of cardiac conduction abnormalities. A warning about the risk of stroke in older adult patients with dementia has been added to the prescribing information for risperidone.

Approximately one-half of drug withdrawals due to safety issues occur within the first 2 years after a drug is marketed. Between 1993 and 2002, there were seven drug recalls that followed reports of death or severe health complications. None of these involved psychotropic drugs. This is a remarkable testimonial to an often-overlooked fact: Psychotropic drugs are among the safest drugs available. Since nomifensine (Merital) was withdrawn in 1986, within 1 year of its marketing, no other psychotropic agent has been withdrawn because of serious side effects.

PHARMACOLOGIC ACTIONS

Heterogeneity of response characterizes all psychotropic drugs. Among other considerations, genetic and environmental factors influence individual response to and tolerability of psychotropic agents. Thus, a drug that may not prove effective in large numbers of patients with a disorder can dramatically improve symptoms in some. In these cases, the identification of characteristics that might predict potential candidates for that drug becomes important but often remains elusive. Attention is being focused on gaining a better understanding of the thousands of genes that regulate the expression of drug-metabolizing enzymes and of the complex functions of the brain.

Even drugs within the same class are distinguished from one another by often subtle differences in molecular structure, types of interactions with neurotransmitter systems, differences in pharmacokinetics, the presence or absence of active metabolites, and protein binding. These differences, combined with the biochemistry of the patient, account for the profile of efficacy, tolerability, and safety and the risk-benefit ratio for the individual. These multiple variables, some poorly understood, make it difficult to predict a drug’s effect with certainty. Nevertheless, knowledge of the nature of each property increases the likelihood of successful treatment. The clinical effects of drugs are best understood in terms of pharmacokinetics, which describes what the body does to a drug, and pharmacodynamics, which describes what the drug does to the body.

Pharmacokinetics and pharmacodynamics need to be seen in the context of the underlying variability among patients with respect to how drug effects are expressed clinically. Patients differ in their therapeutic response to a drug and the experience of side effects. It is increasingly clear that these differences have a strong genetic basis. Pharmacogenetics research attempts to identify the role of genetics in drug response.

DRUG FACTORS

Pharmacodynamics

The time course and intensity of a drug's effects are referred to as its *pharmacodynamics.* Major pharmacodynamic considerations include receptor mechanisms, the dose-response curve, the therapeutic index, and the development of tolerance, dependence, and withdrawal phenomena. Drug mechanism of action is subsumed under pharmacodynamics. The clinical response to a drug, including adverse reactions, results from an interaction between that drug and a patient's susceptibility to those actions. Pharmacogenetic studies are beginning to identify genetic polymorphisms linked to individual differences in treatment response and sensitivity to side effects.

Mechanisms

The mechanisms through which most psychotropic drugs produce their therapeutic effects remain poorly understood. Standard explanations focus on ways that drugs alter synaptic concentrations of dopamine, serotonin, norepinephrine, histamine, GABA, or acetylcholine. These changes are said to result from receptor antagonists or agonists, interference with neurotransmitter reuptake, enhancement of neurotransmitter release, or inhibition of enzymes. Specific drugs are associated with permutations or combinations of these actions. Thus, a drug can be an agonist for a receptor, thus stimulating the specific biologic activity of the receptor or an antagonist, thus inhibiting the biologic activity. Some drugs are partial agonists in that they are not capable of fully activating a specific receptor. Partial agonists may also act as competitive antagonists in the presence of full agonists. Some psychotropic drugs also produce clinical effects through mechanisms other than receptor interactions. For example, lithium may act by directly inhibiting the enzyme inositol-1-phosphatase. Some effects are closely linked to a specific synaptic effect. For example, most medications that treat psychosis share the ability to block the dopamine type 2 (D_2) receptor. Similarly, benzodiazepine agonists bind a receptor complex that contains benzodiazepine and GABA receptors.

It is important to note that action mechanisms for most psychotropic drugs are poorly understood and that even the current classification of drugs (e.g., SSRI and 5HT2A antagonist) is problematic. There are two reasons for this. Studies of the binding properties of many drugs have revealed complex patterns of receptor agonism and antagonism. For example, olanzapine has 5HT2 antagonist properties, but it also is an antagonist for 5HT 1B, 1D, 2A, 2C; α1 adrenergic; H1 histamine and D2 dopamine receptors while acting as a partial agonist for dopamine D1 and muscarinic AChR m1 receptors. This polyspecificity has been extensively by Roth and colleagues. The second issue is the assignation of a mechanism based on a single target. For example, while SSRIs do, indeed, inhibit 5HT transport, they do so within hours after being administered, while the clinical effects of these compounds require weeks to develop. Furthermore, at least some biochemical properties have been attributed to multiple classes of antidepressant drugs. SSRIs can elicit these in cells lacking 5HT transporters.

Thus, accounts of so-called mechanisms of action should nevertheless be kept in perspective. Explanations of how psychotropic drugs actually work that focus on synaptic elements represent an oversimplification of a complex series of events. If merely raising or lowering levels of neurotransmitter activity is associated with the clinical effects of a drug, then all drugs that cause these changes should produce equivalent benefits. This is not the case. Multiple obscure actions, several steps removed from events at neuronal receptor sites, may well be responsible for the therapeutic effects of psychotropic drugs. These *downstream* elements are postulated to represent the actual reasons that these drugs produce clinical improvement.

Pharmacokinetics

Pharmacokinetic drug interactions are the effects of drugs on the plasma concentrations of each other, and pharmacodynamic drug interactions are the effects of drugs on the biologic activities of each other. Pharmacokinetic concepts are used to describe and predict the time course of drug concentrations in different parts of the body, such as plasma, adipose tissue, and the CNS. From a clinical perspective, pharmacokinetic methods help explain or predict the onset and duration of drug activity and interactions between drugs that alter their absorption, distribution, metabolism, or excretion.

Pharmacogenetic research focuses on finding variant alleles that alter drug pharmacokinetics and pharmacodynamics. Researchers are attempting to identify genetic differences in how enzymes metabolize psychotropics, as well as CNS proteins directly involved in drug action. It is likely that the identification of patient genotypes will facilitate the prediction of clinical responses to different types of drugs.

Drugs differ in their propensity to interact with other drugs. Awareness of potential drug–drug interactions has grown dramatically over the last two decades. Most clinicians need to consult charts or online resources to determine when potential interactions may occur and, if so, how clinically relevant they may be. Whenever possible, it is preferable to use a medication that produces a minimal risk of drug interactions. In addition, it is recommended that prescribers know the interaction profiles of the drugs that they most commonly prescribed.

Examples of pharmacokinetic interactions include one drug increasing or decreasing the concentrations of a coadministered compound. These types of interactions can also lead to altered concentrations of metabolites. In some cases, there also may be interference with the conversion of a drug to its active metabolite. There is enormous variability among patients with respect to pharmacokinetic parameters, such as drug absorption and metabolism. Another type of interaction is represented by interactions involving the kidney. Commonly used medications, such as angiotensin-converting enzyme (ACE) inhibitors, nonsteroidal anti-inflammatory drugs (NSAIDs), and thiazides decrease the renal clearance of lithium, increasing the likelihood of its severe elevation. Drug interactions may be pharmacokinetic or pharmacodynamic.

Pharmacogenetics is being used to study why patients differ in the way that they metabolize drugs. In patients who are ultrarapid or extensive metabolizers, the concentrations of a drug may be lower than expected. There are now commercially available assays that determine the potential of individual subjects for metabolizing various psychiatric drugs in an attempt to create a profile of agents to be used and those to be avoided. Unfortunately, the actual therapeutic efficacy of a compound may not be dictated simply by the ability of the patient to process it since an individual’s profile of receptors, transporters, and intracellular signaling cascades (described above) is likely essential to the action of a given agent.

PATIENT-RELATED FACTORS

Response to medication, as well as sensitivity to side effects, is influenced by factors-related to the patient. This is why there is no one-size-fits-all approach to pharmacologic treatment. This is reflected in the finding that, although drugs used to treat a specific disorder demonstrate comparable efficacy in head-to-head studies, these drugs are not equally effective in all patients. Patient-related variables include diagnosis, genetic factors, lifestyle, overall medical status, concurrent disorders, and past history of drug response. A patient's attitude toward medication in general, aversion to certain types of side effects, and preference for a specific agent also need to be considered.

Diagnosis

Accurate diagnosis is an integral component of the treatment process. Failure to correctly diagnose a disorder diminishes the likelihood of optimal drug selection. Misdiagnosis not only can result in a missed opportunity, but at times it also can produce a worsening of symptoms. Inadvertently diagnosing and treating a patient in the depressed phase of bipolar disorder as having unipolar depression can result in the induction of mania or rapid cycling. Treatment failure or exacerbation of symptoms should prompt a reassessment of the working diagnosis.

Past Treatment Response

A specific drug should be selected according to the patient's history of drug response (compliance, therapeutic response, and adverse effects), the patient's family history of drug response, the profile of adverse effects for that drug with regard to the particular patient, and the prescribing clinician's usual practice. If a drug has previously been effective in treating a patient or a family member, the same drug should be used again. While this makes intuitive clinical sense, the evidence base for this has not been well-established, particularly for family history of treatment response. One study by Fava and colleagues demonstrated the response to a second trial of fluoxetine after placebo was similar to an initial trial with fluoxetine, suggesting that past performance predicts future results. Current studies like the PReDICT trial, examining predictors of depression remission, will be key in determining relevant patient-related factors that predict treatment response. A history of severe adverse effects from a specific drug is a strong indicator that the patient would not be compliant with that particular drug.

It is helpful if patients can recall the details of past psychotropic drug treatment: The drugs prescribed, in what doses, for how long, and in what combinations. Often while patients may have tried a medication, they may not have had an adequate trial, either by dose or duration. Unfortunately, because of their mental disorders, many patients are poor historians. If possible, patients' medical records should be obtained to confirm their reports. Family members are a good source of collateral information.

Concurrent Medical or Psychiatric Disorders

Initial assessment should elicit information about coexisting medical disorders. In some cases, there may be a medical disorder that is responsible for the symptoms. Patients with thyroid disease who are not adequately treated may appear depressed. Sleep apnea produces depression and cognitive impairment. Rare conditions, such as Kleine–Levin syndrome, can mimic bipolar disorder. A drug should be selected that minimally exacerbates any preexisting medical problems that a particular patient may have.

Recreational drug use, excessive consumption of alcohol, and frequent ingestion of caffeine-containing beverages can complicate and even undermine psychotropic drug treatment. These compounds possess significant psychoactive properties and, in some cases, may represent the source of the patient's symptoms. It is reasonable to ask patients to abstain from the use of these substances at least until the benefits of psychotropic drug treatment have been unequivocally established. Gradual reintroduction of moderate amounts of alcohol, tea, and coffee can then take place. Patients can then observe whether there are any untoward effects on their clinical status.

Drug Preference

Some prescribers have obvious favorites among the available psychotropic agents. This bias or *allegiance effect* regarding treatment can stem from a combination of factors. A dramatically successful patient response during an early experience with a new drug can produce a *conversion experience,* in which the prescribing physician forms a highly positive opinion of that drug. Patients being treated by that physician are more likely to receive that medication than alternatives. Previous studies have cited physician experience and prior treatment history as among the most highly ranked factors in treatment selection.

INFORMED CONSENT AND PATIENT EDUCATION

Establishing trust and providing motivation to comply with the medication regimen are essential components of successful treatment. Patients should be informed about treatment options and the probable side effects and unique benefits of each treatment. Patient preference should be respected unless there is a compelling advantage involving efficacy, tolerability, or safety with an alternative agent. If a particular medication is being recommended, the reasons for this recommendation should be explained. Patients are more likely to continue taking their medication if they fully understand the reasons why it is being prescribed.

Physicians no longer have a monopoly over the knowledge regarding psychotropic drugs. Many patients now routinely search the Internet for information about available treatments. What physicians do have is an understanding of the clinical circumstances that would lead to the choice of one drug among the many that are available.

Medication treatment benefits from a strong therapeutic alliance between a clinician and a patient. Given the unpredictability of medication response, the frequent occurrence of side effects, and underlying ambivalence about or fear of taking medication, a positive, trusting relationship serves to improve patient compliance. Repeated failed trials may be needed before a response is seen. A patient’s confidence in the physician’s knowledge and judgment enables medication trials and more complex regimens, such as the use of multiple medications.

Discussions about drug selection should be documented in notes. In particular, patients who are informed of potential adverse effects report a higher incidence of side effects but do not have higher rates of premature discontinuation.

How the patient and family are engaged in the treatment plan can determine the success of treatment. With the patient’s consent, relatives and other clinicians should be instructed on the reasons for the drug treatment, as well as the expected benefits and potential risks. In addition, the clinician may also find it useful to explain the theoretical basis for pharmacotherapy to the patient and other significant parties.

DOSING, DURATION, AND MONITORING

Dosing

The clinically effective dose for treatment depends on the characteristics of the drug and patient factors, such as inherited sensitivity to and ability to metabolize a drug, concurrent medical disorders, use of concurrent medication treatments, and history of exposure to previous medications.

Doses used in clinical practice often differ from those used in pivotal studies, which, in turn, form the basis for dosing recommendations that eventually appear in the product information. The reason for this divergence is that patients in pivotal trials are highly selected and are not representative of the people who are ultimately treated. Once a drug achieves widespread use in real-world patients, it might be discovered that the doses needed for optimal efficacy need to be higher than suggested by clinical trial findings. Similarly, the doses needed for tolerability might be lower than thought. Even when a disparity in dosing between actual practice and a clinical trial is established, the package insert is rarely, if ever, changed. This would involve the conduct of costly and time-consuming studies and the submission of an application for change. At times, it may be clinically beneficial to exceed the upper dose limit for a drug—so-called high-dose therapy. However, it is best not to exceed the recommended upper dose threshold without a clear understanding of the potential benefits and risks involved. This caution is particularly important when a drug is associated with a severe concentration-dependent side effect.

Plasma concentrations of many psychotropics can vary up to 10-fold. Thus, to some extent, the optimal dose for an individual is ultimately determined by trial and error, guided by the empirical evidence of the usual dose range for that drug. Some drugs demonstrate a clear relationship between increases in the dose and clinical response. This dose-response curve plots the drug concentration against the effects of the drug.

The *potency* of a drug refers to the relative dose required to achieve certain effects, not to its efficacy. Haloperidol, for example, is more potent than chlorpromazine because approximately 5 mg of haloperidol is required to achieve the same therapeutic effect as 100 mg of chlorpromazine. However, these drugs are equal in their clinical efficacy—that is, the maximum clinical response achievable by administration of a drug.

Drugs must be used in effective doses for sufficient periods. Although drug tolerability and safety are always a consideration, subtherapeutic doses, and incomplete therapeutic trials should be avoided. The use of inadequate doses merely exposes the patient to the risk of side effects without providing the probability of therapeutic benefit. In view of the wide margin of safety associated with most currently prescribed medications, there is more risk of underdosing than overshooting the recommended dose range.

The time of dosing is usually based on the plasma half-life of a drug and its side effect profile. Sedating drugs are given all at night or in disproportionate daily doses at night. The opposite is true with activating drugs. The frequency of dosing is less clear-cut. Most dosing regimens of psychotropic drugs, such as once-a-day versus divided doses, are based on measurements of plasma concentrations rather than receptor occupancy in the brain. Evidence suggests that there is a significant dissociation between brain and plasma kinetics. Reliance on plasma kinetics as the basis for dosing regimens leads to a misunderstanding of necessary schedules.

Duration of Treatment

A common question that a patient typically asks is, “How long do I need to take the medication”? The answer, of course, depends on multiple variables, including the nature of the disorder, the duration of symptoms, the family history, and the extent to which the patient tolerates and benefits from the medication. Patients can be given a reasonable explanation of the probabilities but should be told that it is first best to see whether the medication works for them and whether the side effects are acceptable. Any more definitive discussion of the duration of treatment can be held once the degree of success is clear. Even patients with a philosophical aversion to the use of psychotropic drugs may elect to stay on a medication indefinitely if the magnitude of improvement is great. Most psychiatric disorders have high rates of chronicity and relapse. Because of this, long-term treatment is often needed to prevent a recurrence. Nevertheless, the fact remains that psychiatric treatments are not said to cure the disorders they treat but rather to help control the symptoms.

Treatment is conceptually broken down into three phases: the initial therapeutic trial, the continuation, and the maintenance phase. The initial period of treatment should last at least several weeks because of the delay in therapeutic effects that characterize most classes of psychotropic drugs. The required duration of a *therapeutic trial* of a drug should be discussed at the outset of treatment so that the patient does not have unrealistic expectations of an immediate improvement in symptoms. Patients are more likely to experience side effects early in the course of pharmacotherapy than any relief from their disorder. In some cases, medication may even exacerbate some symptoms. Patients should be counseled that a poor initial reaction to medication is not an indicator of the ultimate outcome of treatment. For instance, many patients with panic disorder develop jitteriness or an increase in panic attacks after starting tricyclic or SSRI treatment. Benzodiazepine agonists are an exception to the rule that there is a delay in clinical onset. In most cases, their hypnotic and antianxiety effects are evident immediately.

Ongoing use of medication does not, however, provide absolute protection against relapse. Continuation therapy provides clinically and statistically significant protective effects against relapse. The optimal duration of continuation or maintenance therapy is variable and depends on the clinical history of the patient. Early-onset chronic major depression, for example, has a more severe course and greater comorbidity than late-onset chronic major depression. In addition to early onset, a history of multiple past episodes, the severity of the current episodes, and the length of the current episode would make longer, even indefinite, treatment appropriate.

TREATMENT OUTCOMES

The goal of psychotropic treatment is to eliminate all manifestations of the disorder, thus enabling the patient to regain the ability to function and enjoy life as fully as before they became ill. This degree of improvement below the syndromal threshold is defined as *remission.*

Response and Remission

Remission is the preferred outcome of treatment not only because of the immediate effect on functioning and state of mind but also because emerging evidence suggests that remitted patients are less likely to experience relapse and recurrence of their disorder.

Patients who improve but do not experience a full resolution are considered to be responders. They may exhibit significant improvement but continue to experience symptoms. In depression studies, the *response* is usually defined as a 50% or greater decrease from baseline on a standard rating scale, such as the Hamilton Depression (HAMD) Scale or the Montgomery-Asberg Depression Rating Scale (MADRS). *Remission* is defined as an absolute score of 7 or less on the HAMD Scale or 10 or less on the MADRS. Expectations about the likely degree of improvement should be based on what is known about the responsiveness of specific disorders to medication therapy. OCD and schizophrenia, for example, are more likely to be associated with residual manifestations of illness than major depression or panic disorder. The probability of full remission from OCD with SSRI treatment alone over a 3-year period is 38%, and the probability of partial remission is approximately 65%.

Treatment Failure

The initial treatment plan should anticipate the possibility that the medication may be ineffective. A next-step strategy should be in place at the initiation of treatment. Repeated drug failures should prompt a reassessment of the patient. First, was the original diagnosis correct? In answering this question, the clinician should include the possibility of an undiagnosed medical condition or recreational drug use as the cause of the psychiatric symptoms. Second, are the observed symptoms related to the original disorder, or are they actually adverse effects of the drug treatment? Some antipsychotic drugs, for example, can produce akinesia, which resembles psychotic withdrawal, or akathisia and neuroleptic malignant syndrome, which resemble increased psychotic agitation. Long-term use of SSRIs can produce emotional blunting, which can mimic depression.

Intolerance of side effects may be the most common reason for treatment failure. Third, was the drug administered at an appropriate dose for a sufficient length of time? Because patients can vary greatly in their absorption and metabolism of drugs, the clinician may need to measure the plasma levels of a drug to ensure a sufficient dose of the drug. Fourth, did a pharmacokinetic or pharmacodynamic interaction with another drug that the patient was taking reduce the efficacy of the newly prescribed drug? Fifth, did the patient take the drug as directed? Drug noncompliance is a common clinical problem that arises as a result of complicated drug regimens (more than one drug in more than one daily dose), adverse effects (especially if unnoticed by the clinician), and poor patient education about the drug treatment plan. Patients may discontinue medication when they recover, thinking that they are cured and no longer benefiting from the medication.

Treatment Resistance

Some patients fail to respond to repeated trials of medication. No single factor can explain the ineffectiveness of the various interventions in these cases. Strategies in these cases include the use of drug combinations, high-dose therapy, and the use of unconventional drugs. There is limited evidence on the comparative success rates associated with any given strategy. This is one area that illustrates the critical need to develop biomarkers to determine trait, state, and therapeutic response. Often patients are given extended (2 months) drug trials before a decision is made to revise therapy. Thus, it may take 8 months to declare a patient “treatment-resistant.” As indicated above, it is not known whether a given patient is inherently treatment-resistant or whether treatment resistance evolves as a result of repeated failures.

Tolerance

The development of tolerance is marked by a need, over time, to use increased doses of a drug to maintain a clinical effect. This decreased responsiveness to a drug occurs after repeated doses. Tolerance also describes decreased sensitivity to adverse effects of the drug, such as nausea. This phenomenon is used as the basis for starting some drugs at subtherapeutic doses, with the plan to adjust the schedule once the patient can tolerate higher doses. Clinical tolerance appears to represent changes in the CNS, such as altered receptor configuration or density. Drugs with similar pharmacologic actions often exhibit cross-tolerance.

Sensitization

Clinically manifested as the reverse of tolerance, sensitization is said to occur when there is increased sensitivity to a drug effect over time. In these cases, the same dose typically produces more-pronounced effects as treatment progresses. Conversely, if the therapeutic index is low, as it is for lithium, careful monitoring of serum levels is essential. For most psychotropic drugs, the therapeutic index is high.

Withdrawal

The development of physiologic adaptation to a drug, with a subsequent risk of withdrawal symptoms, has been reported for many classes of psychotropic drugs. Technically, withdrawal should be considered a side effect. The probability and severity of these reactions are remote with most drugs and more common with others. As a general rule, the more abruptly a drug is stopped and the shorter its elimination half-life, the more likely it is that clinically significant withdrawal symptoms will occur. With some short-acting drugs, withdrawal reactions can result from missed doses and during daily intervals between doses. Gradual tapering of medications after prolonged use is recommended when possible. Although this reduces the risk of withdrawal reactions, it does not ensure that they will not occur. So-called sedative hypnotics and opiates are the agents most often associated with mentally and physically distressing discontinuation reactions. In some cases, such as the barbiturates, withdrawal can be fatal.

There are marked differences among agents, even within a given class, with respect to the probability and severity of discontinuation effects. For example, among the benzodiazepines, alprazolam commonly produces more immediate and intense withdrawal symptoms than other compounds. Among the SSRIs, there is a well-described withdrawal syndrome that appears to be more frequent and severe with paroxetine (Paxil). It can, however, occur with any SSRI. Even fluoxetine may be associated with discontinuation symptoms, but the symptoms may be delayed and attenuated due to the long elimination half-life of its active metabolite. These manifestations are subtle and are delayed for weeks after the last dose. Venlafaxine also produces a severe SSRI-like withdrawal syndrome.

In addition to half-life, many variables can influence the likelihood and degree of discontinuation symptoms. Changes in the rate of drug metabolism, for example, can play a role. Paroxetine is primarily metabolized by the cytochrome P450 (CYP) 2D6 isozyme. However, paroxetine is also a potent inhibitor of CYP 2D6. This results in *autoinhibition,* a dose-dependent inhibition of its own metabolism, with a subsequent increase in plasma concentrations of paroxetine. If the dose of paroxetine is decreased or the drug is stopped, the decline in its plasma concentrations can be steep, causing withdrawal to occur. Withdrawal can occur in rare cases in which the dose of a drug is not decreased, but a second agent that had been inhibiting its metabolism is stopped. For example, paroxetine is metabolized via the CYP 2D6 enzyme system. Bupropion inhibits that enzyme. If a patient taking both agents for several weeks discontinues the bupropion, it could result in a rapid increase in the rate of paroxetine metabolism and a consequent drop in plasma concentrations.

The development of sustained-release versions of drugs such as alprazolam, paroxetine, and venlafaxine has not reduced the severity of their withdrawal reactions. The prolonged half-life of those agents results from delayed absorption rather than prolongation of the elimination phase. The frequency of drug dosing is reduced but not the rate of the fall-off in plasma concentrations.

Poor bioavailability with a generic agent may account for an unexpected loss of clinical effect in the emergence of withdrawal symptoms. The occurrence of these events soon after refilling a prescription should prompt examination of the new medication. It should be confirmed whether the dispensed medication and dose are both correct. It is difficult to ascertain whether generic medications are truly equivalent, so the possibility exists that differences in potency may underlie adverse changes in clinical status.

Withdrawal symptoms invariably occur hours or days after dose reduction or discontinuation. Symptoms resolve within a few weeks, so the persistence of symptoms argues against withdrawal. Although depletion studies have been shown to provoke a rapid return of symptoms, in clinical practice, psychotic and mood symptoms do not usually reappear abruptly after long-term treatment.

COMBINATION OF DRUGS

According to the American Psychiatric Association Practice Guidelines for the Treatment of Psychiatric Disorders, “the use of multiple agents should be avoided if possible” in the treatment of psychiatric disorders. Although *monotherapy* represents the ideal, *polypharmacy* —the simultaneous use of psychotropic medications—has been commonplace since chlorpromazine was combined with reserpine in the early 1950s. The practice of combining drugs is commonplace, and the merits of various *augmentation* or *combination* strategies are routinely discussed in the literature and at scientific meetings. The mean number of simultaneously prescribed medications has increased in recent decades. In response to some of these trends, The Joint Commission has recently instituted quality measures to reduce the inappropriate use of antipsychotic polypharmacy.

Sometimes distinctions are made between augmentation and combination therapy. When two psychotropics with the same approved indications are used concurrently, this is termed *combination therapy.* Adding a drug with another indication is termed *augmentation.* Augmentation often entails the use of a drug that is not primarily considered a psychotropic. For example, in treating depression, it is common to add thyroid hormone to an approved antidepressant. Some strategies are more popular or are preferred but are not necessarily more effective than any other. Augmentation was built into the treatment algorithm for the STAR*D trial, where there was the option to add either lithium or thyroid hormone. In a STAR*D study that examined these augmentation options, the authors found that lithium and thyroid for equally effective in improving response rates in patients that failed monotherapy but that thyroid hormone was better tolerated.

The widespread reliance on multiple-drug therapy, in the face of official discouragement and the absence of objective evidence supporting this practice, is easy to explain: Many patients do not become well with monotherapy with existing drugs. In a typical scenario, a patient has little or no response to a medication, so the physician adds a second agent to induce a better response. This need not be a second drug, as some studies suggest improved efficacy for SSRIs with the addition of omega-3 polyunsaturated fatty acids (fish oil).

In some cases, the use of multiple medications is the rule. Almost all patients with bipolar disorder are taking more than one psychotropic agent. Combination treatment with drugs that treat depression and a dopamine-receptor antagonist or serotonin-dopamine antagonist has long been held as preferable in patients with psychotic depression. Similarly, SSRIs typically produce partial improvement in patients with OCD, so the addition of a serotonin-dopamine antagonist may be helpful.

Medications may also be combined to counteract side effects, treat specific symptoms, and as a temporary measure to transition from one drug to another. It is common practice to add a new medication without the discontinuation of a prior drug, particularly when the first drug has provided partial benefit. This can be done as part of a plan to transition from an agent that is not producing a satisfactory response or as an attempt to maintain the patient on combined therapy.

One limitation is that there is increased noncompliance and adverse effects and that the clinician may not be able to determine whether it was the second drug alone or the combination of drugs that resulted in a therapeutic success or a particular adverse effect. Combining drugs can create a broad-spectrum effect and also changes the ratio of metabolites.

The merits of going to a single drug with a different pharmacologic profile include lower risk of drug–drug interactions, simplicity, and lower cost. It is less burdensome to be taking one medication than two or three and is less likely to meet resistance by the patient. Many patients are ambivalent about taking even one medication, let alone two. This practice is less justifiable in newly treated patients than in those who have failed repeated trials.

SPECIAL POPULATIONS

Although every patient brings a unique combination of demographic and clinical variables to the clinical setting, certain patient populations require special consideration. When treating the young, the older adults, and those with medical disorders, awareness of risks associated with medication assumes increased importance. Data derived from clinical trials are of limited value in guiding many decisions because populations in these studies consisted of healthy young adults. Studies of children and adolescents have become more common, so the understanding of treatment effects in this population has grown.

Children

Understanding the safety and efficacy of most psychotropic drugs, when used to treat children, is based more on clinical experience than on evidence from large clinical trial data. Other than attention-deficit/hyperactivity disorder (ADHD) and OCD, commonly used psychotropic drugs have no labeling for pediatric use, so results from adult studies are extrapolated to children. This is not necessarily appropriate because of developmental differences in pharmacokinetics and pharmacodynamics. Dosing is another special consideration in drug use with children. Although the small volume of distribution suggests the use of lower doses than those used in adults, a child’s higher rate of metabolism suggests that a higher ratio of milligrams of the drug to kilograms of body weight should be used. In practice, it is best to begin with a small dose and increase it until clinical effects are observed. However, the clinician should not hesitate to use adult doses in children if these doses are effective and the adverse effects are acceptable.

The paucity of research data is a legacy of many years in which manufacturers avoided conducting trials in children because of liability concerns, small market share, and, hence, limited profit potential represented by this population. To correct this problem, the FDA Modernization Act (FDAMA) of 1997 provided special encouragement and incentives to study drugs for pediatric use. Since then, there have been expanded indications for FDA-approved medications to treat psychiatric conditions in children, such as risperidone for the treatment of schizophrenia in adolescents and manic/mixed episodes of bipolar I disorder in children and adolescents ages 10 to 17.

Some drugs used to treat psychiatric disorders carry a higher risk of serious side effects. Lamotrigine, for example, is associated with a substantially higher incidence of serious rash among children than among adults.

Older Adult Patients

Older adult persons may be more susceptible to adverse effects (particularly cardiac effects) and may metabolize and excrete drugs more slowly, thus requiring lower doses of medication. In practice, clinicians should begin treating geriatric patients with a small dose—usually approximately one-half of the usual starting dose. The dose should be raised in small increments, more slowly than for middle-aged adults, until a clinical benefit is achieved or unacceptable adverse effects appear. Although most geriatric patients require a small dose of medication, there are others who require a full therapeutic dose. Thus, sound clinical judgment and close monitoring are necessary.

A major concern about psychotropic drugs in the older adults is the well-established causal relationship between the use of these agents and falls. It has been shown that older adult users of psychotropic medications have a 47% increased risk of falls. This risk of falls and hip fractures may be minimized by selecting compounds that produce less unwanted sedation, dizziness, parkinsonism, and postural hypotension.

Historically, antipsychotics have been used to treat nonpsychotic, agitated patients. This practice has been called into question primarily as a result of the CATIE-AD trial that showed a limited benefit of atypical antipsychotics in the treatment of agitation in dementia. Although they may reduce paranoia, there was no significant benefit in improving function or quality of life, and there is evidence of worsening cognitive outcomes with these medications. There is an FDA warning regarding the use of atypical antipsychotics in the treatment of behavioral disorders in older adult patients with dementia. Studies have shown a higher death rate associated with their use. Accordingly, antipsychotic drugs should be reserved for psychosis.

Response to treatment is less predictable among older adults than younger adults. Age-related changes in renal clearance and hepatic metabolism make it more important to be conservative with the starting doses of medication, as well as with the rate of dose titration. The frequent presence of medical disorders and the use of medications to treat them make it necessary to recognize how psychotropic drugs interact with nonpsychiatric disorders and medications. Within any class of agents, those with fewer drug interactions or side effects with potentially serious consequences, such as hypotension, cardiac conduction abnormalities, anticholinergic activity, and respiratory depression, are not suitable choices. Drugs that cause cognitive impairment, such as benzodiazepines and anticholinergics, may mimic or exacerbate symptoms of dementia. Similarly, dopamine-receptor antagonists can worsen or induce Parkinson disease, another age-related disorder. Some side effects, such as SSRI-associated syndrome of inappropriate secretion of antidiuretic hormone (SIADH) and oxcarbazepine-associated hyponatremia, occur more commonly in older patients. A common ethical dilemma with the medically ill older adults or those with dementia is the question of their capacity to give informed consent before treatment with psychotropic drugs or ECT.

Medically Ill Patients

When selecting and monitoring psychiatric drugs, the patient’s medical disorders and their treatment need to be considered. Most importantly, the medical disorder or disorders and their treatments should be ruled out as a cause of the observed psychiatric symptoms. For example, patients with neurologic or endocrine disorders or human immunodeficiency virus (HIV)–infected patients may experience disturbances in mood and cognition. Common medications, such as corticosteroids and L-dopa, are associated with the induction of mania.

Although obvious, when various options are available, a psychotropic should be chosen that does not exacerbate or counteract the treatment of the existing medical disorder. For example, a psychotic patient with diabetes mellitus is better treated with an agent without the risk of weight gain or glucose dysregulation. Depending on the diagnosis, drugs that might treat the primary psychiatric disorder and also cause weight loss, drugs such as bupropion and topiramate (Topamax), can be prescribed for these patients. Patients with obstructive pulmonary disease should not be given sedating drugs, which raise the arousal threshold and suppress respiration. Patients with medical disorders are also taking other medications, which can result in pharmacodynamic and pharmacokinetic interactions. Combined treatment with an inducer of multiple CYP enzymes and a drug that is a substrate for those enzymes could result in subtherapeutic levels, leading to inadequate symptom control. The use of the tuberculosis treatment rifampicin (Rifadin) with carbamazepine is an example. The use of drugs that inhibit CYP 2D6—agents such as paroxetine and fluoxetine—can prevent the conversion of hydrocodone (Robidone) and other opiates into an active analgesic form.

NSAIDs are also a rare cause of perceptual disturbances and psychotic symptoms. Other issues include a potentially increased sensitivity to adverse effects, including increased or decreased metabolism and excretion of the drug and interactions with other medications. Drug interactions are an obvious concern when drugs with a narrow therapeutic range are being used. Any change in the rate of metabolism or interference with the formation and elimination of metabolites can profoundly influence the activity of that drug. Similarly, interactions that interfere with drug metabolism can produce an increase in side effects and toxicity.

As with children and geriatric patients, the most reasonable clinical practice is to begin with a small dose, increase it slowly, and watch for clinical benefits and adverse effects. Determining the plasma drug concentrations may be helpful for such patients, but therapeutic blood concentrations for most psychotropic drugs are neither necessary nor routinely available. Special caution is needed with potential drug–disease interactions. In patients with diabetes mellitus, drugs such as valproic acid or olanzapine, which are associated with increased body weight, and insulin resistance, should be avoided. Patients with a seizure disorder should not receive bupropion, or clomipramine (Anafranil), which lowers the seizure threshold.

Substance Abuse

Many patients who seek or need treatment for a psychiatric disorder engage in chronic use of illicit substances or drink excessive amounts of alcohol. Marijuana is the most commonly used illicit drug in the United States.

Discontinuation of chronic drug or alcohol use can result not only in craving but also in clinically significant psychiatric and physiologic withdrawal symptoms. For many patients, successful treatment of their underlying psychiatric disorder may not be possible in the presence of ongoing marijuana, cocaine, and alcohol use. If several trials of medications fail, hospitalization for detoxification may be necessary. There is little research and no consensus about how to use psychotropic agents in patients who are regular users of cocaine, marijuana, or other recreational drugs. If these agents are used as a method of coping—so-called avoidance coping—it appears to not be effective.

FUTURE DIRECTIONS

After decades of progress in bringing new treatments to the clinical setting, there has been a marked decline in the development or discovery of truly innovative drugs. Although researchers continue to search for medications with new modes of action that are both more effective and better tolerated than existing treatments, there needs to be a better understanding of the underlying neurochemical abnormalities that cause psychiatric disorders. Researchers may be close to significant breakthroughs by which research findings can be translated into better treatments. It is also possible that it will be years before new options are available to clinicians. There is a need to be able to predict more reliably which individuals will respond to specific agents because it is clear that there is no "one-size-fits-all" approach to psychopharmacology. There also needs to be more systematic postmarketing surveillance of drugs for potentially serious side effects. In the meantime, it is imperative that clinicians remain current with the clinical literature and remain thoughtful about the drug choices that they make.

FURTHER READINGS

Berwian IM, Walter H, Seifritz E, Huys QJM. Predicting relapse after antidepressant withdrawal—a systematic review. *Psychol Med*. 2017;47:426–437.

Carhart-Harris RL, Bolstridge M, Rucker J, et al. Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *Lancet Psychiatry*. 2016;3:619−627.

Casarotto PC, Girych M, Fred SM, et al. Antidepressant drugs act by directly binding to TRKB neurotrophin receptors. *Cell*. 2021;184:1299–1313.

Castrén E, Monteggia LM. Brain-derived neurotrophic factor signaling in depression and antidepressant action. *Biol Psychiatry*. 2021;90:128–136.

Catapano F, Perris F, Masella M, et al. Obsessive-compulsive disorder: a 3-year prospective follow-up study of patients treated with serotonin reuptake inhibitors OCD follow-up study. *J Psychiatr Res*.2006;40(6):502–510.

Cocchi M, Tonello L, Rasenick MM. Human depression: a new approach in quantitative psychiatry. *Ann Gen Psychiatry*. 2010;9:25.

Czysz AH, Schappi JM, Rasenick MM. Lateral diffusion of gαs in the plasma membrane is decreased after chronic but not acute antidepressant treatment: role of lipid raft and non-raft membrane microdomains. *Neuropsychopharmacology*. 2015;40(3):766–773.

Davis AK, Barrett FS, May DG, et al. Effects of psilocybin-assisted therapy on major depressive disorder: a randomized clinical trial. *JAMA Psychiatry*. 2020;78(5):481–489.

Domino EF, Chodoff P, Corssen G. Pharmacologic Effects of CI-581, A new dissociative anesthetic, in man. *Clin Pharmacol Ther*. 1965;6:279–291.

Donati RJ, Dwivedi Y, Roberts RC, Conley RR, Pandey GN, Rasenick MM. Postmortem brain tissue of depressed suicides reveals increased Gs alpha localization in lipid raft domains where it is less likely to activate adenylyl cyclase. *J Neurosci*. 2008;28(12):3042–3050.

Eichel K, von Zastrow. Subcellular organization of GPCR signaling. *Trends Pharmacol Sci.* 2018;39:200–208.

Gardner DM, MacKinnon N, Langille DB, Andreou P. A comparison of factors used by physicians and patients in the selection of antidepressant agents. *Psychiatr Serv*. 2007;58(1):34–40.

Gibbons RD, Brown CH, Hur K, et al. Early evidence on the effects of regulators' suicidality warnings on SSRI prescriptions and suicide in children and adolescents. *Am J Psychiatry*. 2007;164(9):1356–1363.

Goodwin FK, Ghaemi SN. The impact of the discovery of lithium on psychiatric thought and practice in the USA and Europe. *Aust N Z J Psychiatry*. 1999;33(Suppl):S54–S64.

Gurevich VV, Gurevich EV. How and why do GPCRs dimerize? *Trends Pharmacol Sci*. 2008;29(5):234–240.

Ha CM, Park D, Kim Y, et al. SNX14 is a bifunctional negative regulator for neuronal 5-HT6 receptor signaling. *J Cell Sci*. 2015;128(9):1848–1861.

Hibicke M, Landry AN, Kramer HM, Talman ZK, Nichols CD. Psychedelics, but not ketamine, produce persistent antidepressant-like effects in a rodent experimental system for the study of depression. *ACS Chem Neurosci*. 2020;11:864–871.

Hurst JH, Hooks SB. Regulator of G-protein signaling (RGS) proteins in cancer biology. *Biochem Pharmacol*. 2009;78(10):1289–1297.

Johnell K, Jonasdottir Bergman G, Fastbom J, Danielsson B, Borg N, Salmi P. Psychotropic drugs and the risk of fall injuries, hospitalisations and mortality among older adults. *Int J Geriatr Psychiatry*. 2016;32(4):414–420.

Kang J, Bowman FD, Mayberg H, Liu H. A depression network of functionally connected regions discovered via multi-attribute canonical correlation graphs. *Neuroimage*. 2016;14:431–441.

Khan A, Khan SR, Leventhal RM, Krishnan KRR, Gorman JMM. An application of the revised consort standards to FDA summary reports of recently approved antidepressants and antipsychotics. *Biol Psychiatry*. 2002;52(1):62–67.

Kim K, Che T, Panova O, et al. Structure of a hallucinogen-activated Gq-coupled 5-HT2A serotonin receptor. *Cell*. 2020;182(6):1574–1588.

Kobilka BK. Structural insights into adrenergic receptor function and pharmacology. *Trends Pharmacol Sci*. 2011;32(4):213–218.

Kroeze WK, Roth BL. Screening the receptorome. *J Psychopharmacol*. 2006;20(4 Suppl):41–46.

Langeberg LK, Scott JD. Signalling scaffolds and local organization of cellular behaviour. *Nat Rev Mol Cell Biol*. 2015;16(4):232–244.

Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med*. 2005;353(12):1209–1223.

Lumsden EW, Troppoli TA, Myers SJ, et al. Antidepressant-relevant concentrations of the ketamine metabolite (2*R*,6*R*)-hydroxynorketamine do not block NMDA receptor function. *Proc Natl Acad Sci USA*. 2019;116:5160–5169.

Meltzer HY. The rise of psychopharmacology and the story of CINP. *J Clin Psychiat*. 2000;61:65–66.

Murray AJ, Tucker SJ, Shewan DA. cAMP-dependent axon guidance is distinctly regulated by Epac and protein kinase A. *J Neurosci*. 2009;29(49):15434–15444.

Pollan M. *How to Change Your Mind: What The New Science of Psychedelics Teaches Us About Consciousness, Dying, Addiction, Depression, And Transcendence*. Penguin Press; 2018.

Rasmussen SGF, DeVree BT, Zou Y, et al. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature*. 2011;477(7366):549–555.

Reiter E, Ahn S, Shukla AK, Lefkowitz RJ. Molecular mechanism of β-arrestin-biased agonism at seven-transmembrane receptors. *Annu Rev Pharmacol Toxicol*. 2012;52:179–197.

Sarris J, Murphy J, Mischoulon D, et al. Adjunctive nutraceuticals for depression: a systematic review and meta-analyses. *Am J Psychiatry*. 2016;173(6):575–587.

Senese NB, Rasenick MM. Antidepressants produce persistent $G\alpha_S$-associated signaling changes in lipid rafts after drug withdrawal. *Mol Pharmacol*. 2021;100:66–81.

Sultzer DL, Davis SM, Tariot PN, et al. Clinical symptom responses to atypical antipsychotic medications in Alzheimer's disease: phase 1 outcomes from the CATIE-AD effectiveness trial. *Am J Psychiatry*. 2008;165(7):844–854.

Swartz MS, Stroup TS, McEvoy JP, et al. What CATIE found: results from the schizophrenia trial. *Psychiat Serv*. 2008;59(5):500–506.

Thal DM, Glukhova A, Sexton PM, Christopoulos A. Structural insights into G-protein-coupled receptor allostery. *Nature*. 2018;559: 45–53.

Vargas AS, Luís Â, Barroso M, Gallardo E, Pereira L. Psilocybin as a new approach to treat depression and anxiety in the context of life-threatening diseases-a systematic review and meta-analysis of clinical trials. *Biomedicines*. 2020;8(9):331.

Vigen CLP, Mack WJ, Keefe RSE, et al. Cognitive effects of atypical antipsychotic medications in patients with Alzheimer's disease: outcomes from CATIE-AD. *Am J Psychiatry*. 2011;168(8):831–839.

Viktorin A, Lichtenstein P, Thase ME, et al. The risk of switch to mania in patients with bipolar disorder during treatment with an antidepressant alone and in combination with a mood stabilizer. *Am J Psychiatry*. 2014;171(10):1067–1073.

Wray NH, Schappi JM, Singh H, Senese NB, Rasenick MM. NMDAR-independent, cAMP-dependent antidepressant actions of ketamine. *Mol Psychiatry*. 2019;24:1833–1843.

Zohar J, Nutt DJ, Kupfer DJ, et al. A proposal for an updated neuropsychopharmacological nomenclature. *Eur Neuropsychopharmacol*. 2014;24(7):1005–1014.

33.1b Drug Development and Approval Process in the United States

AMIR KALALI, M.D.

SCIENTIFIC CHALLENGES IN CNS DRUG DEVELOPMENT

Anyone following the news will be aware of the challenges faced by the pharmaceutical industry. The financial challenges are driven by many factors, including the "patent cliff," where the expiration of the patents on many recent blockbuster treatments has led to severely reduced revenues for many companies resulting in cutbacks in research and development. Increased scrutiny by payors and governments as to the value and novelty of new treatments, as well as price controls in many countries, has added to these challenges.

The latest figures estimate that the cost to develop a new drug is an average of $2.6 billion. Drugs that treat CNS diseases take more than a year longer to develop and are less than half as likely to obtain marketing approval as other drugs, according to a newly completed study by the Tufts Center for the Study of Drug Development.

In addition to the financial challenges, companies must leverage basic science to develop new and innovative treatments that are more effective or safer than current treatments available as generics. This has proved particularly challenging in psychiatric disorders for several reasons, including an incomplete understanding of the basic science and lack of a diagnostic system based on etiology, leading to a heterogeneous population in clinical studies.

Add to this the methodologic challenges in conducting clinical trials in psychiatry, including subjective outcome measures, lack of medication adherence, and increasing placebo response; it is not a surprise that this area is seen as high risk and has seen several companies abandon the area. These factors will be further elaborated below.

The good news is that due to the clear unmet patient need, many new companies are entering this area of drug development.

The Basic Science Challenge

In spite of research conducted in both academic and industry settings, including under the auspices of the Decade of the Brain at the end of the 20th century and subsequent large initiatives involving academia, government, and corporate pharmaceutical entities, the pathophysiologic underpinnings of most conditions affecting the CNS are virtually unknown. This lack of understanding of etiology complicates attempts to construct in vitro or in vivo nonclinical models with which to identify or test therapeutic molecules before taking them into the clinical phases of drug development. In addition, when potential compounds are tested in accepted nonclinical models, the benefits conferred in these models do not predict similar benefits in man. An example would be Alzheimer disease. For more than 40 years, a deficit of brain acetyl cholinergic function has been hypothesized to play a role in the condition, and therefore it was rational to test compounds that diminished the metabolism of acetylcholine: acetylcholinesterase inhibitors. Resulting from a global drug development process that tested many compounds with this pharmacologic mechanism of action, four acetylcholinesterase inhibitors have been approved by regulatory authorities to treat dementia of the Alzheimer type. This might be viewed as a triumph of disease understanding, except that these agents offer symptomatic treatment of lesser magnitude and shorter duration than was predicted on the basis of the known acetyl cholinergic deficit.

The Biomarker Conundrum

A more cautionary tale is the meager success of very expensive and long-duration clinical trials that have tested monoclonal antibodies designed to clear beta-amyloid from the brains of patients with the more advanced stages of Alzheimer disease. Beta-amyloid has been considered by some to be a surrogate biomarker for the Alzheimer disease process because the progressive brain deposition of beta-amyloid plaques has been a well-known pathophysiologic hallmark of the disease since its description by Alois Alzheimer more than 100 years ago. Indeed, this lack of translation from bench to bedside is the rule in the field of Alzheimer disease, where more than 99% of clinical trials have failed. Clearly, more predictive nonclinical models and a better choice of surrogate biomarkers are needed.

THE CLINICAL SCIENCE CHALLENGE: FACTORS IMPACTING SIGNAL DETECTION IN CLINICAL TRIALS

There is no one reason some CNS clinical trials fail. There are multiple factors that occur to one extent or another in any clinical trial. The following section reviews some of the most important of these.

Study Design

Clearly, the design of any clinical trial is a key factor in its success or failure. In general, there are some basic trial principles, such as blinding and randomization, that are followed in most drug development clinical trials. However, sometimes small design changes can have unintended consequences, which emphasizes the importance of thoughtful clinical trial design.

Effect Size

In general, the effect sizes of compounds developed so far for psychiatric disorders are not large, which makes signal detection more challenging. One may argue that if treatments had larger effect sizes, other methodologic issues would not matter.

Dose

Animal models for psychiatric disorders are not good at predicting dosing in humans, and there is a history in CNS drug development of the optimal clinical dose of a compound not being established during the development stage. Several compounds have had negative programs only to be later shown to be effective at different doses.

Diagnosis and Appropriate Patient Selection

Translation from bench to bedside is only one aspect of the issues confronting CNS drug development. Even if clinical research investigators knew a priori that they were about to test the right therapeutic molecules, enrolling the right patients can be just as much of a challenge to successful trial outcomes. The *Diagnostic and Statistical Manual of Mental Disorders* (DSM) provides a standard language by which clinicians, researchers, and public health officials in the United States communicate about mental health disorders. The DSM Fifth Edition (DSM-5), published in 2013, marked the first significant revision since the DSM-IV in 1994.

Although useful for administrative and clinical reasons, as DSM is not based on a biologic etiology of psychiatric disorders, the subjects selected for a trial are clearly a heterogeneous group, and it is not surprising that this dilutes the drug effects shown. There have been recent attempts to categorize patients with new models of diagnosis, such as the NIMH Research Domain Criteria Project (RDoC).

Because there is little understanding of the pathophysiology of CNS disorders, it is not known whether the phenotypically similar features of clinically diagnosed patients may, in fact, result from a multiplicity of underlying genotypes. Compounds with discrete mechanisms of action may offer benefit to patients with only one or a few, but likely not all these underlying genetic etiologies. Given this etiologic heterogeneity of the population with a clinical disease phenotype, how does the clinical trial investigator choose the right patient who might benefit from a promising new compound? Greater knowledge of the heterogeneous pathophysiologic causes and the contributing roles of genetic variants or environmental factors on CNS conditions are necessary.

Another question to ask is whether one treatment can be realistically expected to treat all aspects of a disorder such as depression, schizophrenia, or dementia? However, currently, in psychiatry, the labels granted by regulatory agencies map to the DSM.

Other fields, such as oncology, have faced these issues earlier, and it has been new knowledge about the basic science and etiology that has led to breakthrough success.

Expectation Bias

This effect describes a situation where either the staff conducting the trial or subjects in the trial or both groups expect that the test drug will confer a benefit or that it is unlikely to do so. In the case of a positive bias, this expectation creates a cognitive bias that increases the effect that is being measured. Suppose a large clinical trial of a new antidepressant molecule (Drug X) is being conducted. Going into the trial, it is already known that Drug X has a mechanism of action that is similar to other drugs which produce true antidepressant effects. It is also known that in a prior, smaller clinical trial, Drug X produced a measurable effect when compared to a placebo. These facts are known by the staff conducting the new and larger trial, and they are also explained to each subject enrolling in the trial as a part of informed consent. These facts may create an expectation of drug benefit in the staff who use rating scales to measure the symptoms produced by depression, and so, the staff may thereby have an unconscious bias toward rating more beneficial effects than are actually present. What might be the outcome of this larger trial? Drug X produces the same benefit as it did in the earlier trial, but a placebo produces a greater effect in this trial, shrinking or completely eliminating any measurable difference in the effects produced by Drug X and placebo.

Subjective Outcome Measures

Unlike objective measures used in other therapeutic areas such as blood pressure measurements, death, or the presence of a microorganism, psychiatry trials currently rely on outcome measures based on questions that subjects answer on a rating scale. These outcome measures are prone to a greater degree of uncertainty (and often a greater variance), as well as issues such as recall bias and inter-rater reliability in a multicenter clinical trial.

Regression to the Mean

As already discussed, variance is a way to capture the random change in repeated measurements of the same variable. In a group of people in whom repeated measurements are performed, some people will have values that are far from the mean on one measurement and values that are quite close to the mean on another measurement—all due to the randomness with which repeated measurements may vary. The presence of an extreme value (one far from the mean) on one measurement, and a value much closer to the mean at another measurement, is what is termed regression to the mean. Extreme values tend to move closer (regress) to the mean on different measurements.

As many trials enroll patients during an exacerbation of their illness, the phenomenon would tend to mean that future measurements would be less extreme even if the subject is on a placebo. This phenomenon will therefore act to diminish any drug placebo difference that might be present.

Increased Interaction with Health Care Professionals

During the conduct of a clinical trial, subjects are seen much more frequently than in clinical practice, and the visits can last several hours to conduct all the assessments required by the protocol. This increased interaction is well known to increase placebo response.

Medication Nonadherence in Clinical Trials

Another important reason for failed clinical trials is nonadherence to clinical trial medication. Reasons for this are complex and multifactorial in nature. Whether intentional or unintentional, nonadherence to the medication has major consequences in clinical trials leading to a lack of good-quality data and poor signal detection, which can oftentimes lead to an incorrect assumption that a drug is ineffective when it just has not been taken correctly or taken at all.

New digital and mobile technologies are emerging where adherence can be monitored remotely, which allows for active intervention by Investigators with nonadhering patients. While these technologies have great potential, it is important to remember that these are only the instruments that will aid investigators in oversight of treatment adherence and enhanced patient engagement.

Statistical Considerations

The proper use and interpretation of statistical methods are important in interpreting any clinical trial result.

FEW USEFUL DEFINITIONS

Positive, Negative, and Failed Trials

How does the scientist or clinician detect whether a therapeutic molecule (we will call it Drug X) offers a benefit (i.e., greater efficacy) in a clinical trial? One method is to perform a test to determine if the effect of Drug X on a variable of interest (like blood pressure) has a greater benefit when compared to the effect of either an inert agent that cannot be visually distinguished from the molecule (i.e., a placebo) or a greater benefit when compared to some other compound (Drug Y) that is already known to produce a beneficial effect. But then how does one detect whether there is a beneficial difference between Drug X and placebo or between Drug X and Drug Y?

Before this question is answered, one needs to look a little more closely at what happens when a variable is measured over and over again. Again, take blood pressure as an example. Successive measurements of blood pressure in the same person, even when performed just moments apart, will yield different values with each measurement—tiny differences, perhaps, but real ones. A portion of this difference may be explained by physiologic changes in the blood pressure from moment to moment, but another portion of the difference arises simply from small errors in the measurement itself. If this is the case, then how does one decide if Drug X produces the same benefit to drop high blood pressure as Drug Y if measurements of either drug's effect on blood pressure are subject to error? If one does detect a difference in their measured effects on blood pressure, is that difference real, or could it be due completely to the variation in repeated measurements? To circumvent this inherent difficulty in comparing two measurements, statisticians have defined methods to measure the extent to which multiple measurements will result in a spread of their values around the true average (mean) of that measurement. Not surprisingly, one of these methods to detect the random differences between repeated measurements is termed variance. Signal detection methods take the variance of two sets of measurements (say, from two groups of patients, each group administered a different treatment) into account to determine, if there is a true difference conferred by the two treatments. There are many statistical methods to compare groups and to test whether the differences appear to be real, and these methods are chosen depending on the type of clinical trial and on the outcome being studied (e.g., a change in a numerical value versus a discrete endpoint, such as death). Typically, the statistician hypothesizes that there is no difference in the signal produced by the two treatments, and then the statistical tests are performed to see if the signal appears to be present after all, suggesting that the initial hypothesis is false (i.e., that there appears to be a real difference between the treatments).

Given these definitions, signal detection methods may, in principle, give rise to three sets of clinical trial outcomes. If the hypothesis is that two treatments are the same, then a positive trial outcome is one in which the effect of Drug X is statistically different from and better than that of another treatment (better than Drug Y, or better than placebo). A negative trial outcome is one in which this statement cannot be made; namely, Drug X produces an effect that is not statistically different from that of another treatment, thereby suggesting no additional benefit of Drug X.

But how does one know that a negative outcome is truly due to a nonefficacious drug, and not simply to some arbitrary factor such as a poor clinical trial design or a poorly chosen patient population—one that does not allow detection of a positive signal even when a positive signal should be present? To answer this question, some clinical trials will add a third treatment that produces a well-known clinical effect. For example, a test drug, Drug X, will be compared to a placebo; and in the same trial, a known beneficial drug, Drug Z, is also compared to placebo. If Drug Z produces a signal that is significantly different from placebo, as it should, and then if the effect produced by Drug X is not different from placebo, then this is a well-designed clinical trial that yields a very interpretable result: namely, Drug X does not produce an effect. However, what if the trial is so poorly designed or conducted that Drug Z itself fails to separate from placebo? This is an example of a failed clinical trial—one that is poorly designed or conducted and which is therefore not even interpretable because even a drug with a known effect cannot be detected.

Repeated Measures

Sometimes companies conducting trials wish to explore the effect of different doses of a compound on multiple symptom domains. Although this sounds like an efficient way to conduct an experiment, it turns out that if there is a small chance that a measured effect of the drug on a single variable is due to chance (and therefore that this is a true effect); it is unfortunately also true that many measurements of many doses on many effects may appear to produce a measurable effect of the drug, and some of these effects are false. Why? Random chance has been multiplied across all these tests due to multiple measurements and a factor termed repeated measurements. Various statistical techniques have been constructed to overcome the effect of repeated measurements.

A more recent development in the statistical design of clinical trials has been alternative clinical trial designs such as adaptive designs and Bayesian methods, but there have been few examples of their adoption in the CNS clinical trials so far.

FUTURE DIRECTIONS FOR DRUG DEVELOPMENT

The cost of developing a new treatment has been increasing to the point that further increases are unsustainable. There is a realization in the pharmaceutical industry that the current model has to change. This will drive a much more synergistic model that includes a whole network of collaborators who, by fostering collective scientific creativity, will ensure future long-term success. Competitive and technologic changes in the pharmaceutical industry—from powerful new drug chemistries to innovative R&D partnerships—have the potential to improve productivity that will, in turn, increase output with greater efficiency, less cost, and more value.

Collaborative Networks

Some clinical trial goals require larger and longer trials, namely those designed for disease modification or for long-term follow-up, to examine the durability of effect. For example, in examining clinical trial outcome measures that target negative symptoms in schizophrenia, Laughren and Levin (2011) proposed that for FDA regulatory purposes, a 6-month or longer treatment period was desirable in order to examine the durability of the effect of the therapeutic molecule proposed for that indication.

Taking this trend toward longer trial durations much further, trials designed to examine the effects of treatments for prodromal Alzheimer disease may last 12 to 24 months and involve a large sample size in order to detect small changes in neuropsychological measures and/or in functional or clinical outcomes. To accomplish these large trials, involving many thousands of patients and years of treatment duration, collaborative networks have been established between one or more biopharmaceutical companies, academia, and government agencies. These large collaborations will not only facilitate the conduct of the trials but also decrease the per-entity cost and spread the risk of negative trial outcomes.

Globalization

There has been a marked trend toward drug development programs becoming more and more global. This is occurring for several reasons. Regulators and patient advocates are rightly demanding data that is obtained from ethnically relevant research subjects that address the possibility of differential efficacy and safety and optimal dosing in these populations.

This trend of globalization has its challenges, including more complexity, longer regulatory reviews in some regions, different standards of medical care in some countries, and the transcultural issues in ensuring consistent reliable administering of subjective outcome measures.

Personalized Technology

CNS trials often use outcome measures that require patient entries into diaries. These can include subjective judgments on sleep quality, symptoms of depression, quality-of-life, and functional measures. In the past, these entries were made using paper diaries, but increasingly trials have turned to electronic devices that patients use for their entries. Device-based entries offer a number of advantages over the older paper-based methods. For example, entries can be both time-stamped and immediately uploaded to a trial database. This speeds the acquisition of data, and it also allows the trial analysis staff to determine that entries were made at the times and dates specified by the protocol rather than just before a patient’s next clinic visit. Finally, electronic devices are increasingly familiar to people through their use of personal smartphones and other handheld devices; this tends to enhance patient interest in and familiarity with the use of these devices.

The introduction of Apple’s ResearchKit (http://www.apple.com/researchkit/) holds the promise of taking clinical research into the mainstream with the availability of clinical research to everyone via their mobile devices, with enabling modules such as eConsent. Digital Medicine is profoundly influencing health care delivery systems and disease management.

Personalized Medicine

Currently, most treatments available to patients were developed based on their effect on a patient population. It has been a long-held aspiration to develop treatments that are personalized to patients based on factors such as genetics. This goal can be termed personalized medicine.

An example of an efficacy-targeted success in personalized medicine is the use of HER2 screening in patients with metastatic breast cancer. Patients who over-express (are positive for) HER2 have a greater likelihood of responding to trastuzumab (Herceptin).

A similar example of a safety-targeted success is the use of CYP450 testing in order to predict drugs that should be dosed at different strengths in order to avoid excess or poor metabolism of drugs handled by those metabolic paths; and also, to avoid drug: drug interactions. How does this affect the use of psychiatric drugs? Fluoxetine inhibits the activity of CYP2D6, which metabolizes more than 100 commonly used medications. This means that patients taking fluoxetine may be at risk of receiving a functional overdose of other compounds that are metabolized by 2D6 unless the potential for a drug interaction is avoided by measuring 2D6 function in patients taking fluoxetine or by tailoring the doses of coadministered drugs that are metabolized by 2D6.

The field of personalized medicine extends much further than these two examples about its applicability to efficacy and safety. Two basic approaches will be considered: tailoring drugs for the individual and identification of genetic susceptibility to disease.

Numerous large clinical trials targeting CNS and other indications routinely collect genetic information on patients in those trials, allowing the retrospective identification of a gene variant profile that correlates with efficacy or safety signals observed during the trial. If identified, then prospective use of these gene variant profiles is feasible by screening patients for these gene variants in order to predict the efficacy or safety of the drug in question before it is administered to them.

Whole-genome sequencing (WGS) is now available through many avenues for the individual patient. WGS approaches conducted in large populations can help identify variants that confer susceptibility to a variety of conditions, and then individual WGS results can be correlated with these population databases to seek clues about that individual’s likelihood of harboring or developing those conditions. Caution must be exerted in applying this information to individual genomic data because of the ethical and interpretative issues that may be encountered. For example, if an asymptomatic individual has a variant profile consistent with schizophrenia, how likely is it that this person will ever be symptomatic? And how will this information influence: that person’s lifestyle; the extent to which this information is shared with family members, friends, work colleagues, or insurance providers; the advisability of parenting; or, if already a parent, the extent to which this information is communicated to the offspring? These issues and many others must be carefully considered by each individual in consultation with their health care providers.

Biomarkers

In the broadest sense, biomarkers are any measurable feature of a patient that can serve to predict disease likelihood and severity, disease course, disease response to treatment, or the current general health status of the patient. Optimally, a surrogate biomarker for disease severity with disease severity can be objectively measured; and so, can serve as a clinical trial outcome measure in place of a more subjective “soft” outcome measure such as a symptom rating scale. Biomarkers for psychiatric conditions have been sought for decades. Laughren discussed the last 50 years of psychiatric drug development and concluded that biomarker approaches were the most likely methods to help advance the field. This has been echoed by many in the research community. The challenge has been to identify and validate biomarkers meeting these criteria that can be used as trial outcome measures and that would also be accepted by regulatory authorities for drug approvals. To date, this has had no more than limited success in most indications and no success yet in psychiatric indications. Hopefully, the wide variety of imaging, genetic, and other approaches that are underway will soon change this landscape.

Patient Centricity

Traditional drug development processes have not been particularly patient-centric. As an example, research protocols had little, if any, input from patients, resulting in, at times, complex burdensome protocols that either excluded most patients or negatively impacted the recruitment and retention of research subjects and the success of the programs.

More recently, there has been a realization that this needs to change. Also, in some disease areas, especially rare and orphan diseases, patient advocacy groups are themselves funding clinical trials and having a stronger voice.

More “Real-World” Data

One of the criticisms of drug development programs has been that the data generated is not particularly helpful to clinicians in helping them make treatment decisions and choices. This is driven by a misunderstanding of the purpose of phase 2 and phase 3 clinical trials. These programs are conducted to meet the requirements of regulatory authorities, and traditionally it was the postapproval phase 4 studies that were conducted to address questions that would help prescribers understand the optimal use of the new treatment.

As payors across the world have increased their demand for data that justify pricing above generic treatments, it has become imperative to generate this data earlier in the development of new treatments. This will lead to a modification of the issues that are addressed in drug development in preapproval programs.

MORE RELEVANT OUTCOME MEASURES

Traditionally in psychiatric drug development, outcome measures are used, measuring symptom severity and the impact of a new treatment on these symptoms. However, it has become clear that reducing symptoms is not enough. Both payors and patients are looking for the impact on functioning and other novel domains. This will lead to new ways of assessing the impact of a new treatment on patients.

Disease Interception

Throughout the history of medicine, the main focus has been on treating diseases once they have become manifest. It is clear that in many conditions, this may be too late, the clearest example currently being Alzheimer disease. There are efforts underway across neuroscience to identify patients that may be at risk and develop interventions that would prevent the disease before it develops. This approach has implications both for regulatory bodies and payors, if physicians are to intervene that early in the natural history of a disease.

Regulation of Drug Approvals

When clinical trials yield results that demonstrate the potential benefits of efficacious and safe compounds, these data must be submitted to regulatory authorities for their scrutiny before approval for marketing is granted. It is important to understand the landscape in which regulatory guidelines have developed in order to put the challenges of drug development into perspective. This section reviews the history and the current U.S. regulatory environment for drug development.

Background

The FDA, according to its mission statement, “is responsible for protecting the public health by assuring the safety, efficacy, and security of human and veterinary drugs, biological products, medical devices, our nation’s food supply, cosmetics, and products that emit radiation. The FDA is also responsible for advancing public health by helping to speed innovations that make medicines and foods more effective, safer, and more affordable; and by helping the public get the accurate, science-based information they need to use medicines and foods to improve their health.” The FDA plays a role in overseeing and guiding the development of the medications and devices that are used in the practice of psychiatry and ensuring that they are effective, offer a favorable risk-to-benefit ratio, are accompanied by adequate directions for use, can be manufactured with reliable quality, and are not promoted in a misleading fashion.

HISTORY OF DRUG REGULATION

The history of American drug regulation features a series of legislative responses to public concerns. A century ago, there were no restrictions on what could be sold as medicine and what claims could be made about the products. *Patent* medicines were widely sold for a variety of ailments. Labels did not list ingredients. Some medicines contained ethanol, opium, morphine, heroin, or cocaine; others contained harmless (and ineffective) ingredients. Popular press accounts of the harms of these *tonics* sold to unsuspecting customers, along with concerns about food adulteration and the dangers of chemical additives, led to the passage of the Pure Food and Drug Act of 1906, championed by Harvey Wiley, head of the Bureau of Chemistry in the Department of Agriculture.

Pure Food and Drug Act of 1906

Wiley law, the Pure Food and Drug Act of 1906, was only six pages long. It prohibited the manufacture and interstate shipment of “adulterated” and “misbranded” foods and drugs. This gave the government the authority to go to court against illegal products but did not require any premarketing approval. Initially, the law also did not outlaw false therapeutic claims. A 1912 amendment declared such claims illegal but required that the government prove fraudulent intent. Thus, the only requirement was that the purveyor of a medication believes that the drug was effective. Regarding labeling, the law required only that a contents statement, if used, must be truthful. Not until a 1913 amendment was a statement of contents required.

Elixir Sulfanilamide Tragedy of 1937

The shortcomings of the 1906 law limited the government’s ability to ensure the safety of drugs. Although the need for reform of drug regulation was apparent by the early 1930s, legislation proposing an expansion of the regulatory authority of the FDA was held up by debate in Congress and opposition from the food and drug industries. Finally, in 1937, a public health scandal emerged that generated the political will to pass the bill. A chemist at Massengill, attempting to develop a palatable formulation for a sulfa antibiotic, created a product that used diethylene glycol as a solvent. No premarket testing, in humans or animals, was required under the laws then prevailing, and none was undertaken. As a result of exposure to this solvent (akin to antifreeze), more than 100 people died, and many of them were children. As it happens, the Pure Food and Drug Act gave the government the authority to take action against the manufacturer only because the product was labeled an *elixir.* By definition, an *elixir* uses ethanol as a solvent; had the manufacturer labeled the product a *solution,* it would not have been in violation of any law.

Food, Drug, and Cosmetic Act of 1938

Although there had been other crises associated with the marketing of untested, unsafe drugs, the elixir sulfanilamide tragedy spurred the passage of the Food, Drug, and Cosmetic Act of 1938. This law contained a number of important provisions concerning drug regulation. The 1938 law required that new drugs undergo safety testing, the results of which would be submitted to the FDA in a new drug application (NDA). Premarketing approval of the NDA was required before a drug could be sold. The law also required that drugs have adequate labeling for safe use. Furthermore, it clearly prohibited false therapeutic claims for drugs without the requirement that the government shows fraudulent intent. In addition, the FDA was given authority to conduct factory inspections.

Thalidomide Tragedy of 1962

By the early 1960s, Senate hearings on strengthening the regulation of drugs were once again underway. In the midst of these hearings, chaired by Senator Estes Kefauver, came the news of the devastating effects of thalidomide (Thalomid), which was being marketed in Europe. Although the NDA had been submitted to the FDA for review, the medical officer responsible for the review, Frances Kelsey, had requested additional safety information, and the drug had not yet been approved (of note, as a student, Kelsey assisted in the investigation of the elixir sulfanilamide tragedy). Her role in protecting the American public from the dangers of thalidomide was widely reported and generated support for the passage of what became known as the Kefauver–Harris Amendments in 1962. These amendments required drug manufacturers to prove to the FDA that their products were safe *and* effective before marketing. They established that the evidence of efficacy must be derived from “adequate and well-controlled trials.” Furthermore, the amendments also addressed the use of drugs in clinical trials. Under the Kefauver-Harris Amendments, the current investigational new drug (IND) process was established. Under the law, the FDA had to be provided with full details of the clinical investigations, including drug distribution, and the clinical studies had to be based on previous animal investigations to ensure safety.

After the addition of the efficacy requirement, the FDA commissioned a review by the Institute of Medicine of all drugs approved between 1938 and 1962 to determine whether they were effective, as well as safe. This process, known as the drug efficacy study implementation (DESI), resulted in the removal of some drugs from the market, whereas others were determined to be effective. Nonprescription drugs were also evaluated under a similar process, which resulted in the publication of monographs establishing which ingredients may be “generally recognized as safe and effective” for various conditions.

IND Regulations

After the 1962 amendments, the FDA issued and implemented regulations concerning the investigational use of drugs; these regulations were updated and revised in 1987. The revisions were intended to “ensure FDA’s ability to monitor carefully the safety of patients participating in clinical investigations, while also facilitating the development of new beneficial drug therapies.” The regulations serve to “focus FDA’s attention during the early phase of clinical research on protecting the safety of human test subjects and to give sponsors greater freedom to design, revise, and implement clinical research studies” and to “facilitate consultation between FDA and drug sponsors, once the preliminary human studies have been completed and a drug appears to have marketing potential, to help ensure that the design of major clinical trials is acceptable and will support marketing approval, if the test results are favorable.” And beginning in the mid-1970s, the FDA began giving advice on the kinds of studies and data that would be needed to support drug approval. Clinical guidances have been published, setting forth the data needed for drugs in various therapeutic categories.

PDUFA, FDAMA, and Pediatric Provision

Thirty years after the 1962 amendments, frustration over lengthy review times led to the passage of the Prescription Drug User Fee Act (PDUFA) of 1992, which charged fees for NDA review to facilitate the expansion of the FDA’s review resources and, in return, ensured action on an application in a specified period of time. PDUFA was renewed with the passage of the FDA Modernization Act (FDAMA) of 1997, which further shortened review times and made additional changes in the review process. The authority to collect user fees was again reauthorized as part of the Public Health Security and Bioterrorism Preparedness Act of 2002 and the Food and Drug Administration Amendments Act of 2007 (FDAAA).

Among the provisions of FDAMA were the pediatric exclusivity provisions. These offered incentives, in the form of an additional period of protection from generic competition, to sponsors for fulfilling an FDA request for pediatric data on their products. In 2002, the Best Pharmaceuticals for Children Act (BPCA) was passed, providing further clarification of the pediatric exclusivity program and establishing a system for conducting studies through the National Institutes of Health for high-priority drugs that were already off-patent or for which the commercial sponsor declined to respond to the agency’s request. In the first 10 years after the passage of FDAMA, over 150 submissions of data in response to written requests were evaluated to determine whether pediatric exclusivity should be granted. Among these were studies of a number of psychotropic agents, including antidepressants, antipsychotics, and anxiolytics. In many cases, the results of the studies led to labeling changes stating that safety and efficacy in children had not been established; yet pediatric exclusivity is awarded even in these cases. In addition, under the provisions of the BPCA, upward of 50 drugs have been included on a priority list, including lithium (Eskalith), bupropion (Wellbutrin), and clonidine (Catapres).

Although FDAMA and BPCA established incentives for pharmaceutical companies to conduct pediatric studies, such studies remained optional. Subsequently, in 2003, the Pediatric Research Equity Act (PREA) authorized the agency to require sponsors to conduct pediatric studies as part of its development program. The requirements may be waived, if the drug is not likely to be used in a substantial number of pediatric patients or does not represent a meaningful therapeutic benefit for pediatric patients. In addition, the agency may also waive the requirement, if “there is evidence strongly suggesting that the drug or biologic product would be ineffective or unsafe in all pediatric age groups” or on the grounds that the studies are impossible or highly impracticable. The agency may require that the pediatric studies are submitted as part of the initial application or may defer all or some of the pediatric studies until after the application is approved for use in adults.

The FDAAA of 2007 reauthorized PDUFA, PREA, and BPCA, as well as creating new programs to enhance postmarketing safety surveillance and the FDA’s ability to communicate safety information.

STRUCTURE OF THE FDA

There are five headquarters centers in the Washington, D.C., area: The Center for Drug Evaluation and Research (CDER), the Center for Biologic Evaluation and Research (CBER), the Center for Devices and Radiologic Health (CDRH), the Center for Veterinary Medicine (CVM), and the Center for Food Safety and Applied Nutrition (CFSAN). Staff in these offices work primarily on product review and regulatory policy. In addition, a network of field offices throughout the United States focuses on inspection and surveillance, laboratory work, and public and industry education.

Most relevant to psychiatric treatment is the work of CDER, where drugs and therapeutic biologic treatments are reviewed, and CDRH, which has jurisdiction over medical devices such as those used in ECT. This section will address the review of drugs and biologic therapeutic products as undertaken in CDER.

NEW DRUG DEVELOPMENT

The development of a new pharmacologic treatment for a psychiatric (or any other) disorder usually begins with an exploration of a compound's actions in vitro and animal investigations. These investigations may be carried out before involving the FDA. Indeed, initial safety testing in animals is required before beginning human studies that are subject to FDA review. Most compounds are abandoned in these early stages of development. However, those that appear promising and are reasonably tolerated in animal safety studies may go forward in development.

Definition of a Drug

The Food, Drug, and Cosmetic Act defines drugs as:

1. Articles recognized in the official U.S. Pharmacopeia, the official Homeopathic Pharmacopeia of the United States, or the official National Formulary
2. Articles *intended for* [emphasis added] use in the diagnosis, cure, mitigation, treatment, or prevention of disease
3. Articles (other than food) intended to affect the structure or function of the body
4. Articles intended for use as a component of any article specified previously.

Therefore, it is the *intended use* of the substance that defines it as a drug. As soon as a substance is intended for human use toward one of these objectives (i.e., when human studies begin), the substance is, by definition, a drug and is subject to regulation under the drug provisions of the Food, Drug, and Cosmetic Act.

Note that therapeutic biologic products are a subset of drugs. Although the term "drug" is often used to designate a typical small-molecule product and the terms "biologic" or "therapeutic biologic" to designate the large, biologically derived substances used therapeutically, such as monoclonal antibodies, both types of products are considered drugs outside the FDA. In this section, the term *drug* will be used to encompass all products meeting the statutory definition.

Phases of Drug Development

Drug development to support a commercial marketing application proceeds in somewhat overlapping phases. Nonclinical (also called *preclinical*) studies begin before human studies but continue throughout the clinical development period. Each phase of human testing is supported by animal tests that are appropriate to the extent of planned human exposure in the ensuing studies. Human studies proceed from basic characterization of the drug product to establishment of efficacy and elucidation of the adverse event profile in the intended population of use. Although the categorizations phase 1, phase 2, and phase 3 are used commonly, the development of a drug does not proceed linearly through these phases. For example, studies of drug–drug interaction and studies of pharmacokinetics in renal or hepatic impairment are both considered phase 1 studies but often occur late in drug development. Finally, unanswered questions or new indications may be explored after the drug has been approved for marketing and are commonly called phase 4 studies.

Nonclinical Studies

Nonclinical studies that are sufficient to establish a tolerable dose and to identify the target organs of toxicity for a new drug must be conducted before the first use of a new chemical entity in humans. A standard battery of animal studies and in vitro studies are required to support a marketing application, with the duration and timing of these studies related to the proposed duration of use of the drug in the clinical program. Thus, acute toxicology studies are needed to support single-dose human administration, whereas longer studies must be completed before using the drug in human studies involving chronic exposure.

Phase 1. Phase 1 studies represent the initial introduction of the new drug to humans. These studies, usually conducted in healthy volunteers, typically in closely monitored (often inpatient) settings, serve to characterize the absorption, distribution, metabolism, and excretion of the compound, to identify overt toxicities associated with drug administration, and to establish a tolerable dose for use in further studies. In some circumstances, phase 1 studies may involve patients and may be designed to provide early indications of the drug’s efficacy; examples include drugs intended as chemotherapeutic or anti-HIV agents. Phase 1 also includes studies in which investigational drugs are used as research tools to explore biologic phenomena or disease processes. The objective of phase 1 in commercial drug development is to obtain sufficient information about the drug’s pharmacokinetics and pharmacologic effects to permit the design of well-controlled, scientifically valid phase 2 studies. The total number of subjects and patients included in phase 1 studies varies with the drug but is generally in the range of 20 to 80.

Phase 2. Phase 2 includes the initial controlled clinical efficacy studies. These studies typically include carefully selected patients with the disease or condition under study and are usually well-controlled, closely monitored, and optimized for the collection of efficacy data. In phase 2, exploratory work is undertaken to help determine the optimal doses of the drug. In conditions for which no standard outcome measures have been defined, studies may explore different approaches to capturing the effect of the drug. Safety data are collected in this phase to determine the common short-term side effects and risks associated with the drug.

Phase 3. After preliminary evidence suggesting the effectiveness of the drug has been established in phase 2 trials, additional information about effectiveness and safety is needed to evaluate the overall risk-to-benefit relationship of the drug and to provide an adequate basis for product labeling. Phase 3 studies provide this information. The inclusion and exclusion criteria in phase 3 studies are usually broader than in phase 2 in order to select a population that is more representative of the patients who will be exposed to the drug during marketing. Phase 3 studies may include from several hundred to several thousand subjects, including individuals with a broad range of typical comorbid conditions and concomitant medications seen in the target population.

Phase 4. Once the drug has been approved, subsequent postmarketing activities may be conducted in phase 4. Studies to elucidate specific issues about a drug, which are considered important for optimal use but not essential to know before drug approval, may be performed. Studies to support approval for additional indications (e.g., use in pediatrics, if not previously established) may also be conducted. Programs of active surveillance for adverse events may be undertaken, as well as the required passive collection of adverse event reports from patients, health care providers, and other sources, which manufacturers must report to the FDA periodically.

IND Review Process. Before beginning human studies, the commercial drug sponsor must submit a request for an IND exemption from interstate shipping laws, or an IND, for review by the FDA. Once the IND is submitted, the sponsor must wait 30 calendar days before initiating any clinical trials. During this time, the FDA has an opportunity to review the IND for safety to ensure that research subjects are not subjected to unreasonable risk.

REVIEW EMPHASIS AND PROCEDURES. According to the Code of Federal Regulations:

FDA's primary objectives in reviewing an IND are, in all phases of the investigation, to assure the safety and rights of subjects and, in phases 2 and 3, to help assure that the quality of the scientific evaluation of drugs is adequate to permit an evaluation of the drug's effectiveness and safety. Therefore, although the FDA's review of phase 1 submissions will focus on assessing the safety of phase 1 investigations, the FDA's review of phases 2 and 3 submissions will also include an assessment of the scientific quality of the clinical investigations and the likelihood that the investigations will yield data capable of meeting statutory standards for marketing approval.

Thus, the reviewers pay close attention to all reviews to determine whether the risks of the study are appreciated, anticipated, communicated, and managed by the sponsor. For later phase studies, they also consider whether the study design is adequate to meet the scientific and regulatory objectives of the study. During the review of an initial IND submission (as well as the review of subsequent submissions of additional protocols and amendments), the interdisciplinary review team approaches these issues from various angles, including clinical study results; chemistry/product quality; pharmacology/toxicology information; clinical pharmacology data; and biostatistics use and results.

Clinical Hold. If no concerns are identified by the interdisciplinary review team, the IND goes into effect, and the studies proposed in the initial submission are permitted to proceed. Once the IND is in effect, additional protocols, study results, and safety information are submitted. For each new clinical study proposed, the review process is similar, but the sponsor is not required to wait 30 days before initiating subsequent studies.

However, at the time of initial IND review or at any time during drug development, the FDA can place human studies under an IND on *clinical hold.* A clinical hold, according to the Code of Federal Regulations, is "an order issued by FDA to delay a proposed clinical investigation or to suspend an ongoing investigation. The clinical hold order may apply to one or more of the investigations covered by an IND."

Under the Code of Federal Regulations, the FDA may impose a clinical hold on an IND for any of the following reasons:

- Human subjects are or would be exposed to an unreasonable and significant risk of illness or injury.
- The clinical investigators named in the IND are not qualified by reason of their scientific training and experience to conduct the investigation described in the IND.
- The investigator brochure is misleading, erroneous, or materially incomplete.
- The IND does not contain sufficient information required under §312.23 to assess the risks to subjects of the proposed studies.
- The plan or protocol for a phase 2 or 3 study is clearly deficient in its design to meet its stated objectives. When a clinical hold is issued, a sponsor must address the issues that are the basis of the hold before the order is removed, and the study may proceed.

Marketing Application Submission and Review. The culmination of the development program is the submission of the NDA for a small-molecule drug product or a biologics licensing application (BLA) for a therapeutic biologic product. FDA staff have 10 months (6 months for a drug designated as a priority application owing to its potential to offer a breakthrough in treatment) to assess the raw data thoroughly and to compare their own conclusions about the data to the applicant's conclusions presented in written reports. The objective is to determine whether the sponsor (now called the *applicant*) has met the regulatory burden for marketing: Demonstrating efficacy through "substantial evidence" from "adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved" and demonstrating safety through "all tests reasonably applicable." Therefore, in the course of the review, FDA staff assess whether the clinical studies were indeed appropriately designed and conducted, whether the data support the conclusion that the drug is effective, and whether the proper human and animal safety data ("all tests reasonably applicable") were collected and illustrate a favorable risk to benefit ratio and whether the manufacturing methods and controls are sufficient to ensure consistency of identity, quality, purity, and strength of the product.

SAFETY REVIEW. The purpose of the safety review is two-fold: first, to determine the adequacy of the extent of exposure and the specific tests carried out in the safety evaluation (such as laboratory tests, electrocardiograms [ECGs], and adverse event collection) and to assess whether it meets the standard of "all tests reasonably applicable." For example, approval of a drug in a class known to be associated with a prolongation of the QT interval might be withheld, if a safety review determined that the effects of the drug on cardiac conduction were not thoroughly evaluated. The second purpose is to characterize the safety profile so that the risks of the drug may be weighed against the benefits identified in the efficacy review. Characterization of the safety profile includes identifying and estimating the frequency of common adverse events that are, or maybe, causally related to the use of the drug; identifying factors that predict the occurrence of adverse events, such as age, gender, concomitant medication, or illness; and examining the data for events that suggest problems that are so serious that the use of the drug should be limited or approval withheld altogether. The clinical reviewer evaluates all available safety information and integrates it into an overall picture of the drug's effects.

EFFICACY REVIEW. The review of efficacy, conducted jointly or in parallel by the medical officer and biostatistician, uses raw data from the pivotal clinical trials in the application to confirm the applicant's conclusions about the study results. Any deviations from the protocol-specified study plan are evaluated for possible impact on the outcome of the studies. Reviewers analyze the applicant’s assumptions about the outcome for protocol violators, subjects who withdraw early from studies, or subjects with missing data. They test the impact of alternative assumptions on study outcomes. The objective of the efficacy review is to determine whether the studies submitted were “adequate and well-controlled” and whether they provide “substantial evidence” of efficacy. For the most important trials, the review also includes clinical trial site inspections by the Division of Scientific Investigation (DSI). By examining the trial site, documentation, source materials, and equipment, the DSI establishes whether there are important violations that call into question the results of the study. Such findings are unusual but do occur as a result of deliberate fraud, error, or incompetence.

Abuse Liability Assessment. For most psychotropics, the NDA also includes a section addressing the abuse potential of the product. Assessment of abuse potential is required for many drugs with CNS activity. This testing may include a battery of in vitro receptor-binding studies, various animal studies of the rewarding properties of the drug, and, if preliminary studies suggest the potential for abuse, human laboratory studies. Data from these studies, as well as other information concerning the chemical structure of the compound and any other relevant data from literature or epidemiologic reports, are reviewed as part of the NDA and are considered in writing the relevant sections of the product package insert. Importantly, these data are also taken into consideration in preparing a recommendation concerning the control of the substance under the Controlled Substances Act. This recommendation is prepared by a team within the Office of the Director of CDER at the FDA and is forwarded through the secretary of Health and Human Services to the Drug Enforcement Administration.

PHARMACOLOGY AND TOXICOLOGY. By the time the NDA is submitted, the nonclinical data include a variety of animal studies. There should be both acute and multidose toxicity studies, the duration of which is determined by the intended use of the product. Also included are reproductive toxicology studies, mutagenicity studies, carcinogenicity studies for chronically used products, and any special toxicity studies (such as dermal sensitization studies for a transdermal product and blood compatibility studies for an intravenous [IV] product). There also should be studies of the absorption, distribution, metabolism, and elimination of the drug in animals because these data are needed to compare the dose response in animals to the dose recommended for humans. In the course of the review, the pharmacologist examines the results of these studies and also critically evaluates the study design and conduct to determine if the data may be accepted as valid.

CLINICAL PHARMACOLOGY. The clinical pharmacology reviewer examines data relevant to understanding all aspects of the absorption, distribution, metabolism, and excretion of the drug. This includes information on any important influences on these parameters, such as food effects, drug interactions, or effects of age, gender, or hepatic or renal impairment. This information is included in the clinical pharmacology section of the label and often forms the basis for important precautions to ensure the safe use of the product. Explorations of exposure–response relationships for efficacy and safety endpoints are also an important part of the clinical pharmacology review. Some pharmacogenomic analyses (analyses of genetically based differences in pharmacology), such as explorations of the role of metabolic polymorphisms, may also be undertaken.

CHEMISTRY/PRODUCT QUALITY. In reviewing an NDA, the chemistry reviewer examines the manufacturing information to determine whether there are appropriate specifications and test methods in place to ensure identity, quality, purity, and strength, as well as consistency from batch to batch. For a BLA, the quality reviewer, similarly, is responsible for ensuring the quality, safety, efficacy, availability, and security of therapeutic protein and monoclonal antibody products. The reviewer also evaluates information about stability, including the adequacy of the stability data to support product use, expiration dating, and any concerns about packaging or shipping that might affect stability. In addition, the reviewer takes into account the findings of inspections of the manufacturing facilities, which are undertaken to ensure adherence to the principles of good manufacturing practices. Reviewers of biologic products participate in the actual preapproval and biennial facility inspections. The FDA may withhold approval, if inspections show significant procedural problems at the facilities that are responsible for manufacturing or packaging the product.

REGULATORY DECISION. All the discipline reviews are taken into consideration in making an assessment of whether the drug's benefits outweigh its risks for the intended use in the intended population. If a drug offers significant benefits for a narrowly defined population but carries significant risks, the FDA may decide to approve the drug subject to special conditions to ensure that physicians understand and comply with using the drug only for those in whom the anticipated benefits can be expected to be greater than the possible risks.

If there are unanswered questions or unresolved issues in an application or concerns about safety or lack of efficacy, the FDA may withhold approval. The applicant is issued a letter explaining the deficiencies precluding approval and instructions on how to address these deficiencies.

PRODUCT LABELING. For the physician and the patient, the most obvious product of the NDA review process (apart from the regulatory decision that makes the product available on the market) is the package insert or label. The multidisciplinary review team carefully examines language drafted by the NDA applicant and ensures that the statements throughout the label are supported by the submitted data. The format of the label is set forth in the Code of Federal Regulations (21 CFR §201).

Postapproval (Phase 4). Despite the careful premarket review, postmarketing surveillance of adverse drug effects is also essential. It is evident that all possible side effects of a drug cannot be anticipated based on preapproval studies. The total exposure to an investigational drug during its development depends, to some extent, on the novelty of the compound and on the presence of any specific safety concerns. However, for practical reasons, it is rarely possible to have sufficient premarket exposure to characterize the occurrence of rare adverse events. According to the current recommendations of the International Conference on Harmonization, for new chemical entities that are to be used chronically (or repeatedly but intermittently) in the treatment of conditions that are not immediately life-threatening, the number of patients exposed should be sufficient to detect serious adverse events occurring at a rate of approximately 1 per 1,000 in short-term use. It is recommended that the safety database include at least 1,500 subjects. Of these, 300 to 600 should be treated for at least 6 months, and 100 should be treated for 1 year. Applying the so-called rule of three, if a particular serious adverse drug effect is not observed in the 100 subjects exposed for 1 year, this can provide reasonable assurance that the true cumulative 1-year incidence is no greater than 3%. However, this rule of three represents only 95% confidence limits. It must be noted that even a database of 3,000 subject exposures stands a 5% chance of missing events that occur at a rate of 1 per 1,000, and such a rate may translate into a significant number of occurrences when the drug is introduced into a population of hundreds of thousands.

Therefore, the FDA maintains a system of postmarketing surveillance and risk assessment programs to identify adverse events that did not appear during the drug approval process. The agency monitors adverse reactions reported by drug companies, consumers, and physicians and seeks out signals suggestive of drug-related events. The agency uses this information to update drug labeling, identify unanticipated drug–drug or drug–disease interactions, spot common medication errors and take action to prevent them, and, on rare occasions, re-evaluate the approval or marketing decision.

POSTMARKETING SAFETY. Due to practical limitations on how many patients can be studied in clinical trials for a new drug, it is inevitable that some safety concerns will come to light until after a new drug is used in a broader population after approval. Unlike Frances Kelsey, the modern FDA reviewer often does not have foreign postmarketing data available to supplement the clinical trial data submitted in the drug application since drugs are now frequently approved first in the United States. In the post-PDUFA era, critics of the FDA raised concerns that the emphasis on bringing new drugs to market more rapidly had led to a failure to identify important safety issues. Several high-profile drug safety issues led the agency to examine its procedures and to launch several initiatives aimed at improving the safe use of medicines, including a dedicated Safety Policy and Communication staff that ensures that physicians, health provider groups, and patients are alerted to important new safety information in a timely manner. The Safety Policy and Communications staff includes a Drug Safety Oversight Board (known as the DSB) that was established to provide independent oversight and advice to the center's director on the management of important and emerging drug safety issues. The DSB provides perspective from scientists not involved in the initial review and approval of a particular marketing application and provides a venue for adjudicating organizational disputes concerning the management of drug safety issues. In 2007, the Safety Policy and Communications Staff launched a new tool for communicating safety issues, the FDA Drug Safety Newsletter, which is distributed electronically and viewable on the FDA's Web site.

FUTURE DIRECTIONS FOR FDA

In 2004, the FDA launched a new initiative known as the *Critical Path Initiative,* which is intended to "stimulate and facilitate a national effort to modernize the scientific process through which a potential human drug, biologic product, or medical device is transformed from a discovery or "proof of concept" into a medical product." By identifying key areas in need of attention and partnering with other agencies and organizations to initiate research, the FDA is working to improve the tools used to evaluate the safety and efficacy of human and veterinary products, as well as the safety and nutrition of food and food ingredients. A list of areas representing high-priority opportunities for improvement was disseminated, and work has been initiated on a variety of projects.

One project specific to psychiatry is an effort under the aegis of the *Biomarkers Consortium,* a public/private biomedical research partnership supported by the Foundation for the National Institutes of Health (FNIH). Among their first projects is an effort to identify genomic biomarkers for treatment response in major depressive disorder. Many other critical path projects are broadly relevant to psychotropic drugs, including work on examining the genetic basis of adverse drug events; work intended to broaden understanding of the biochemical and genetic basis of drug-induced liver injury; a database of digital ECGs collected in drug-approval studies to be used to improve methods of measuring drug effects on cardiac conduction, and efforts to identify nephrotoxic biomarkers that predict the effects of drugs on kidney function.

As drug development becomes more and more of a global endeavor, the need for regulatory authorities in different countries and regions to interact and discuss regulatory issues and challenges becomes more important.

CONCLUSION

Drug development in every field has become increasingly complex, costly, and challenging, and in no other therapeutic area is this more evident than in psychiatric conditions. The challenges are enormous in translating from the Gordian knot of poorly understood CNS pathophysiologic processes and nonpredictive models into clinical trials that will have a good probability of success. Nevertheless, there are new technologic and scientific approaches that provide a basis for increased hope as clinical trials for psychiatric conditions move into the 21st century. The goal is clear. Psychiatrists must discover, develop, and obtain regulatory approval for new medicines that will improve the treatment of many unmet or poorly served medical needs in the psychiatric field. Patients are waiting.

FURTHER READINGS

Apple ResearchKit. http://www.apple.com/researchkit/.

Brayanov JB, Smith MA. Bayesian and "anti-Bayesian" biases in sensory integration for action and perception in the size-weight illusion. *J Neurophysiol*. 2010;103(3):1518–1531.

Chapman SCE, Horne R. Medication nonadherence and psychiatry. *Curr Opin Psychiatry*. 2013;26(5):446–452.

Code of Federal Register Volume 71, No. 15, *Tuesday*, January 24, 2006, Rules and Regulations, Requirements on Content and Format of Labeling for Human Prescription Drugs Title 21: Food and Biological Products Drugs, Chapter 1—Food and Drug Administration, Department of Health and Human Services, Subchapter D—Drugs for Human Use, Part 314, Applications for FDA approval to market a new drug.

*Code of Federal Regulations, Title 21: Food and Drugs, Chapter 1—Food and Drug Administration, Department of Health and Human Services, Subchapter C—Drugs: General, Part 201, Labeling and Subchapter D—Drugs for Human Use: Part 312, Investigational new drug application.

Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: Few candidates, frequent failures. *Alzheimers Res Ther*. 2014;6(4):37.

Cutler AJ, Tran-Johnson T, Kalali A, Aström M, Brecher M, Meulien D. A failed 6-week, randomized, double-blind, placebo-controlled study of once-daily extended release quetiapine fumarate in patients with acute schizophrenia: Lessons learned. *Psychopharmacol Bull*. 2010;43(4):37–69.

Elenko E, Underwood L, Zohar D. Defining digital medicine. *Nat Biotechnol*. 2015;33(5):456–461.

Federal Food, Drug, and Cosmetic Act: Public Law No. 75–717, 52 Stat. 1040 (1938), 21 USC §§301 et. seq., Chapter II—Definitions, Sec. 201 [321]Definitions.

Federal Register Vol 52, No. 53. (1987): Rules and Regulations, Department of Health and Human Services, Food and Drug Administration, 21 CFR Parts 312, 314, 511, and 514 [Docket No. 82N-0394], New Drug, Antibiotic, and Biologic Drug Product Regulations, Final Rule.

Filiou MD, Turck CW. General overview: biomarkers in neuroscience research. *Int Rev Neurobiol*. 2011;101:1–17.

Foley KF, Quigley DI. Pharmacogenomic potential of psychiatric medications and CYP2D6. *MLO Med Lab Obs*. 2010;42(1):32–34.

*Guidance: Drug Safety Information—FDA's Communication to the Public: US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, 2007.

Hellewell JS. Patients' subjective experiences of antipsychotics: clinical relevance. *CNS Drugs*. 2002;16(7):457–471.

International conference on harmonisation; guidance on statistics principles for clinical trials; availability–FDA. Notice. *Fed Regist*. 1998;63(179):49583–49598. Accessed October 5, 2016. http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/statistical-principles-for-clinical-trials.html

Kemp AS, Schooler NR, Kalali AH, et al. What is causing the reduced drug-placebo difference in recent schizophrenia clinical trials and what can be done about it? *Schizophr Bull*. 2010;36(3):504–509.

Laughren T, Levin R. Food and drug administration commentary on methodological issues in negative symptom trials. *Schizophr Bull*. 2011;37(2):255–256.

Library of Congress. Project on Decade of the Brain. 2000. Accessed October 5, 2016. http://www.loc.gov/loc/brain/

Nutt DJ, Attridge J. CNS drug development in Europe—past progress and future challenges. *Neurobiol Dis*. 2014;61:6–20.

Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Durg Discov*. 2010;9(3):203–214.

Tufts Center for the Study of Drug Development, Volume 16, Number 6. 2014.

Wilson BJ, Nicholls SG. The Human Genome Project, and recent advances in personalized genomics. *Risk Manag Healthc Policy*. 2015;8:9–20.

33.1c Classification of Psychotropics: Neuroscience-Based Nomenclature (NbN)

JOSEPH ZOHAR, M.D., SASSON ZEMACH, M.D., AND DANIEL MINKIN LEVY, M.D.

Psychopharmacology, the scientific study of the effect drugs have on the mind, has redefined psychiatry over the last decades. This process began with a serendipity discovery of chlorpromazine and later on the monoamine oxidase inhibitors (MAO-I). These agents were found to be treatments for psychosis and depression, respectively, and led to the foundation of the concept that mental illness can be treated with psychotropics.

However, the use of the first modern psychotropics predated important revelations concerning pathologic disease processes. These include discoveries regarding dopamine and the D2 receptor in psychosis and the role of monoamines in depression. For years, while more pharmacologic agents were discovered, the knowledge of how they exert their effect was still missing. The gap between neuroscientific insights and clinical practice led to resulted in adapting a disease-based terminology. Soon, intuitive terms such as "antipsychotics," "antidepressants," "anxiolytics," and "mood stabilizers" were broadly used. Later, terms that are not indicative of pharmacologic activities, such as "typical," "atypical," "first-generation antipsychotics (FGA)," and "second-generation antipsychotics (SGA)," were added, in part due to commercial motives.

Currently, disease-based terminology is in use by several classification systems. Internationally, the anatomical therapeutic chemical (ATC), first published by the World Health Organization in 1976 as a monitoring tool, turned out to be a classification system, although this was not the primary intention. The ATC was built as a strict hierarchy, consisting of five levels, each level defined by a different property: anatomical site, therapeutic effect, indication, and chemical structure. The ATC uses terms such as "Psycholeptics" and "Psychoanaleptics," which are no longer in clinical use, to describe psychiatric medications alongside other central nervous system medications (e.g., "Anesthetics," "Analgesics," and "Anti-Parkinson Drugs").

Naming psychotropics following the indications they had been initially used for has unfortunate consequences. Physicians often prescribe "antidepressants" for the treatment of disorders other than depression (e.g., anxiety, obsessive-compulsive disorder [OCD], stress-related disorder). "Antipsychotics" are often prescribed for conditions other than psychosis (e.g., depression, tics, aggression). This discrepancy between terminology and practice is confusing for patients and professionals, opening the gate to stigma ("psychiatry is not a science") and negatively impacting adherence.

Psychiatric disorders affect different brain regions, different neural circuits, and different neurotransmitters. This complexity is reflected in the medication that is used for each disorder. For example, the "antidepressant" class holds 16 different pharmacologically different subgroups. Yet the term "antidepressants" blurs the pharmacologic, and consequently the clinical, differences between those medications. A deep understanding of the different tools we currently have is needed when customizing a treatment regime, planning the next pharmacologic step, and analyzing the past pharmacologic history of a patient.

To address the need for new terminology, in 2008, the “Nomenclature Taskforce” was launched with the mission to examine ways of improving the current nomenclature in psychopharmacology. This effort was headed by representatives from five major international organizations (the European College of Neuropsychopharmacology [ECNP], The American College of Neuropsychopharmacology [ACNP], the Asian College of Neuropsychopharmacology [AsCNP], the International College of Neuropsychopharmacology [CINP], and the International Union of Basic and Clinical Pharmacology [IUPHAR]). It was predefined as a nonprofit endeavor funded by the ECNP. The result was a *Neuroscience-based Nomenclature (NbN)*, relying on the growing neuroscience knowledge of recent decades. This pharmacology-driven nomenclature does not conflict with the clinical use of medications and aims to bridge neuroscience and clinical psychiatry, facilitating rational prescribing. It also offers a nomenclature that reflects contemporary neuroscience insights.

Table 33.1c–1.
NbN’s Two Domains: Pharmacology and Mode of Action

Pharmacology 10 domains	Mode of Action 9 domains
Acetylcholine	Enzyme inhibitor
Dopamine	Enzyme modulator
γ-aminobutyric acid (GABA)	Ion channel blocker
Glutamate	Neurotransmitter releaser
Histamine	Positive allosteric modulator (PAM)
Melatonin	Receptor agonist
Norepinephrine	Receptor antagonist
Opioid	Receptor partial agonist
Orexin	Reuptake inhibitor
Serotonin	

The nomenclature allows medications to be attributed to multiple domains, and the richness of psychopharmacology is better represented. These two domains reflect the state-of-the-art knowledge as of 2022.

PRINCIPLES OF THE NEUROSCIENCE-BASED NOMENCLATURE

As pharmacology provides better space to name, describe, summarize, and classify psychotropic medicines, it was decided to move from a disease-based to a pharmacology-driven nomenclature. The core concept of the new nomenclature is that a drug is attributed to two *domains*: (1) Pharmacology, which represents the system that is being modulated, and (2) Mode of Action, which details the effect an agent has on said system.

The ***Pharmacology*** domain includes 10 different domains referring to the neurotransmitter, molecule, or system being modified by the medication. These include but are not limited to the monoamine neurotransmitters, γ-aminobutyric acid (GABA), glutamate, and acetylcholine.

The Mode of Action domain includes nine different modes which specify the nature of the action exerted by the medications, such as receptor antagonist, reuptake inhibitor, enzyme modulator, and so forth. Table 33.1c–1 lists the 10 pharmacology and 9 modes of action domains in the nomenclature.

Based on clinical relevance, compounds can be assigned to more than one pharmacology or MoA domain to better reflect their versatility. Those 10 pharmacologic domains and 9 modes of actions listed in NbN can be used in different combinations to describe close to 150 compounds.

PHARMACOLOGICALLY DRIVEN GROUPING

One of the challenges for the new terminology is to find a way to address groups of medications. Even though there are almost no two drugs alike, the habit of grouping medications with common pharmacology and overlapping clinical applications is rooted deep in our vocabulary. The NbN glossary (Table 33.1c–2) lists each medication under a class term that should be used when referring to groups of medications. These classes are meant to be broad enough to encompass similar medications while remaining pharmacologically accurate.

DOPAMINE BLOCKERS AND THE TREATMENT OF PSYCHOSIS

Blockade of the dopamine receptor is the cornerstone upon which modern psychiatric treatment of schizophrenia is built, as nearly all medications targeting psychosis block the dopamine receptor with some affinity. The term *Dopamine Blockers* is used here as an umbrella term to describe medications with the common denominator of dopamine blockade, including dopamine receptor antagonists and dopamine-serotonin receptor antagonists.

Table 33.1c–2.
The NbN Glossary[a]

Current Terminologies	**NbN Classification**		**Drugs**
	Pharmacologic Domain	**Mode of Action**	
Antipsychotics	**Dopamine Blockers**		
Typical (1st generation)	Dopamine	Receptor antagonists (D2)	Fluphenazine, haloperidol, perphenazine, pimozide, sulpiride, zuclopenthixol
	Dopamine, serotonin	Receptor antagonists (D2, 5HT2)	Thioridazine, flupenthixol, trifluoperazine, pipotiazine
	Dopamine, serotonin	Receptor antagonists (D2, 5HT2)	Chlorpromazine
Atypical (2nd generation)	Dopamine	Receptor antagonist (D2)	Amisulpiride
	Dopamine, serotonin	Receptor antagonist (D2, 5-HT2)	Loxapine, lurasidone, olanzapine, perospirone, sertindole, ziprasidone, zotepine, blonanserin, iloperidone
	Serotonin, dopamine, norepinephrine	Receptor antagonist (5-HT2, D2, NE alpha-2)	Asenapine
	Dopamine, serotonin, norepinephrine	Receptor antagonist (D2, 5-HT2, NE alpha-2)	Clozapine, paliperidone
	Dopamine, serotonin, norepinephrine	Receptor antagonist (D2, 5-HT2, NE alpha-2)	Risperidone
	Dopamine, serotonin	Receptor antagonist (D2, 5-HT2) and reuptake inhibitor (NET)(metabolite)	Quetiapine
	Serotonin, dopamine	Receptor antagonist (5HT2, D2) and reuptake inhibitor (SERT)	Lumateperone

Dopamine partial agonists	**Dopamine Partial Agonists**		
	Dopamine, serotonin	Receptor partial agonist (D2, 5-HT1A) and receptor antagonist (5-HT2A)	Aripiprazole, brexpiprazole
	Dopamine, serotonin	Receptor partial agonist (D2, D3, 5-HT1A), receptor antagonist (5-HT2B)	Cariprazine
Stimulants	**Dopamine Enhancers**		
	Dopamine, norepinephrine	Reuptake inhibitors (DAT, NET) and releaser (DA, NA)	Amphetamine (D) and (D, L), lisdexamfetamine
	Dopamine, norepinephrine	Reuptake inhibitors (DAT, NET)	methylphenidate (D) and (D, L)
	Dopamine	Reuptake inhibitor (DAT)	Modafinil
	Dopamine	Receptor agonist (D2)	Pramipexole
Antidepressants	**Monoamine Enhancers**		
Tricyclic antidepressants	Norepinephrine	Reuptake inhibitor (NET)	Desipramine, protriptyline, nortriptyline, maprotiline
	Norepinephrine	Receptor antagonist (alpha-2), reuptake inhibitor (NET)	Mianserin
	Norepinephrine, serotonin	Reuptake inhibitor (NET and SERT)	Lofepramine, amoxapine
	Serotonin, norepinephrine	Reuptake inhibitor (SERT and NET)	Imipramine, dosulepin, clomipramine
	Serotonin, norepinephrine	Reuptake inhibitor (SERT and NET), receptor antagonist (5-HT2)	Amitriptyline, doxepine
MAO-I	Serotonin, norepinephrine, dopamine	Enzyme inhibitor (MAO-A and -B)	Isocarboxazid, phenelzine
	Serotonin, norepinephrine, dopamine	Reversible enzyme inhibitor (MAO-A)	Moclobemide
	Serotonin, norepinephrine, dopamine	Enzyme inhibitor (MAO-A and -B), releaser (DA, NE)	Tranylcypromine

	Dopamine, norepinephrine, serotonin	Enzyme inhibitor (MAO-B and -A)	Selegiline
SSRI	Serotonin	Reuptake inhibitor (SERT)	Citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline
	Serotonin	Reuptake inhibitor (SERT), receptor partial agonist (5-HT1A), receptor antagonist (5-HT3)	Vortioxetine
	Serotonin	Reuptake inhibitor (SERT), receptor partial agonist (5-HT1A)	Vilazodone
SNRI	Serotonin, norepinephrine	Reuptake inhibitor (SERT and NET)	Duloxetine, milnacipran, desvenlafaxine, levomilnacipran, venlafaxine
NRI	Norepinephrine	Reuptake inhibitor (NET)	Atomoxetine, reboxetine
NDRI	Norepinephrine, Dopamine	Reuptake inhibitor (NET, DAT), releaser (NE, DA)	Bupropion
Tricyclic Antidepressants	**Monoamine Blockers**		
	Serotonin, dopamine	Receptor antagonist (5-HT2 and D2)	Trimipramine
Antidepressants	**Monoamine Modulators**		
	Norepinephrine, Serotonin	Receptor antagonist (NE alpha-2, 5-HT2, 5-HT3)	Mirtazapine
	Serotonin	Receptor antagonist (5-HT2), receptor agonist (5-HT1A)	Nefazodone, trazodone
	Serotonin	Receptor agonist (5-HT1A), receptor antagonist (5-HT2A)	Flibanserin
	Norepinephrine Blockers		
	Norepinephrine	Receptor agonist (alpha-2)	Clonidine, guanfacine, lofexidine

	Norepinephrine	Receptor antagonist (alpha-1)	Prazosin
Antipsychotic (in Parkinson's Disease)	**Serotonin Blocker**		
	Serotonin	Receptor antagonist (5-HT2A)	Pimavanserin
Benzodiazepines/ Anxiolytics	**GABA Enhancers**		
	GABA	Positive allosteric modulator (GABA-A receptor, benzodiazepine site)	Alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, flunitrazepam, lorazepam, oxazepam
Benzodiazepines/ hypnotics	GABA	Positive allosteric modulator (GABA-A receptor, benzodiazepine site)	Estazolam, eszopiclone, flunitrazepam, flurazepam lormetazepam, midazolam, nitrazepam, quazepam, temazepam, triazolam
Z-Drugs/Hypnotics	GABA	Positive allosteric modulator (GABA-A receptor, benzodiazepine site)	Zaleplon, zolpidem, eszopiclone, zopiclone
	GABA	Positive allosteric modulator (GABA-A receptor, neurosteroid site)	Brexanolone
	GABA	Positive allosteric modulator (GABA-A receptor, alcohol site)	Chloral hydrate, chloral betaine
	GABA	Positive allosteric modulator (GABA-A receptor, barbiturate site)	Clomethiazole
	GABA	Receptor agonist (GABA-B)	Baclofen
	GABA	Receptor agonist (GABA-B, gammahydroxybutyrate [GHB])	Sodium oxybate (GHB)
	GABA Blocker		

	GABA	Receptor antagonist (benzodiazepine receptor)	Flumazenil
	Glutamate	Receptor antagonist (NMDA)	Esketamine, ketamine
	Glutamate	Receptor antagonist (NMDA)	Memantine
	Glutamate Enhancers		
	Glutamate	Yet to be determined	Acamprosate
Mood Stabilizers	**Cation Channel Blockers**		
	Glutamate/GABA	Voltage-gated sodium and calcium channel blocker	Carbamazepine, oxcarbazepine
	Glutamate/GABA	Voltage-gated sodium channel blocker	Lamotrigine
	Glutamate/GABA	Yet to be determined	Valproate
	Glutamate/GABA	Facilitation of GABA transmission, receptor antagonist on AMPA and KA	Topiramate
	Glutamate/GABA	Voltage-gated calcium channel blocker	Gabapentin, pregabalin
Mood stabilizer	**Lithium**		
	Lithium	Enzyme interactions	Lithium
	Histamine Blockers		
	Histamine	Receptor antagonist (H1)	Diphenhydramine, hydroxyzine
	Histamine, dopamine	Receptor antagonist (H1, D2)	Promethazine
	Histamine Enhancer		
	Histamine	Receptor antagonist (H3)	Pitolisant
	Opioid Blockers		
	Opioid	Receptor antagonist (μ, κ, d)	Naloxone
	Opioid	Receptor antagonist (μ, δ), receptor partial agonist (κ)	Nalmefene
	Opioid	Receptor antagonist (μ, κ, d)	Naltrexone
	Opioid Enhancers		
	Opioid	Receptor agonist (μ)	Methadone

	Opioid	Receptor partial agonist (μ), receptor antagonist (κ, d)	Buprenorphine
	Opioid, glutamate	Yet to be determined	Tianeptine
	Acetylcholine Enhancers		
	Acetylcholine	Enzyme inhibitor (acetylcholinesterase)	Donepezil
	Acetylcholine	Enzyme inhibitor (acetylcholinesterase), a-7 nicotinic Receptor positive allosteric modulator	Galantamine
	Acetylcholine	Enzyme inhibitor (acetylcholinesterase, butyrylacetycholinesterase)	Rivastigmine
	Acetylcholine	Receptor partial agonist (nicotinic receptors)	Varenicline
Hypnotics	**Melatonin Enhancers**		
	Melatonin	Receptor agonist (MEL1, MEL2)	Melatonin, ramelteon
	Melatonin	Receptor agonist (M1, M2), receptor antagonist (5-HT2B, 5-HT2C)	Agomelatine
	Orexin Blocker		
	Orexin	Receptor antagonist (OR1, OR2)	Suvorexant
	Enzyme Blocker		
	Alcohol	Enzyme inhibitor (mainly liver)	Disulfiram

[a] Each medication is attributed to two domains. Pharmacology and Mode of Action, alongside a class term (e.g., dopamine blocker, monoamine enhancer).

Despite their central role in the treatment of psychosis, dopamine blockers are broadly used for other indications; vice versa, psychosis can be treated, in cases, without dopamine blockers (as in the case of pimavanserin for Parkinson disease psychosis). Dopamine blockers are an essential component in the treatment of affective disorders. For example, in unipolar depression, aripiprazole, risperidone, and olanzapine are often used as augmenting agents. In bipolar disorder, dopamine blockers are important tools in the treatment of mania, depression, and relapse prevention alongside lithium, valproic acid, and carbamazepine, regardless of the presence of concordant psychotic symptoms. Clinical practice and guidelines suggest haloperidol, olanzapine, quetiapine, and risperidone can be especially useful for agitation in severely ill, but not psychotic patients.

The application of dopamine blockers is also indicated in disorders outside of the affective and psychotic spectrums. Tic disorders are often treated, albeit in small doses, with pimozide, haloperidol, risperidone, and aripiprazole. Some dopamine blockers are used for behavioral modification, as risperidone is frequently used to reduce aggression and conduct problems in children and adolescents. Interestingly, depending on the specific drug, treating OCD with a dopamine blocker can be beneficial or hazardous; low doses of risperidone or aripiprazole are proven helpful augmentation strategies for serotonergic enhancers, especially for OCD patients with a comorbid tic disorder, while clozapine was reported to worsen OCD symptoms.

Using NbN helps discriminate between the different dopamine blockers with a higher resolution. For example, olanzapine, previously classed as an “atypical antipsychotic/SGA” is now defined under the pharmacologic domain of “dopamine, serotonin,” and the “antagonist” mode of action. In comparison, brexpiprazole, another “atypical,” has the pharmacology domain “dopamine, serotonin” as well–but with two modes of action: “partial agonist and antagonist” (Table 33.1c–2). In contrast, sulpride and amisulpride, previously FGA and SGA, respectively, are now both dopamine antagonists. This approach better highlights the differences and similarities of the different dopamine blockers.

DOPAMINE ENHANCERS

Complimentary to the dopamine blockers, dopamine enhancers is a term for the class medications that primarily increase dopamine levels, indicated for the treatment of attention deficit hyperactivity disorder (ADHD) and narcolepsy. Medications in this group include amphetamine, lisdexamfetamine, methylphenidate, and modafinil (Table 33.1c–2). Some of these medications were previously known as “stimulants.” This term is a counterintuitive one since a large portion of patients suffering from ADHD actually displays hyperactivity. Since the term “stimulants” is not in line with the calming effect those medications have on patients, a more clinically accurate term for this class is dopamine enhancers.

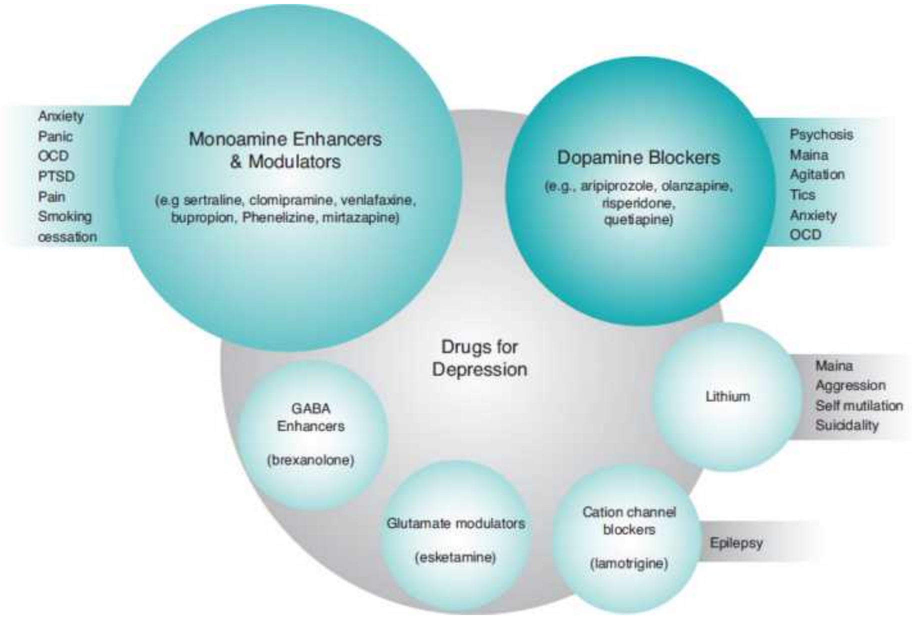
Anxiety
Panic
OCD
PTSD
Pain
Smoking cessation
Monoamine Enhancers & Modulators
(e.g sertraline, clomipramine, venlafaxine, bupropion, Phenelizine, mirtazapine)
Dopamine Blockers
(e.g., aripiprozole, olanzapine, risperidone, quetiapine)
Psychosis
Maina
Agitation
Tics
Anxiety
OCD
Drugs for Depression
GABA Enhancers
(brexanolone)
Lithium
Maina
Aggression
Self mutilation
Suicidality
Glutamate modulators
(esketamine)
Cation channel blockers
(lamotrigine)
Epilepsy

FIGURE 33.1c–1. Medications for Depression. According to the current nomenclature, only “monoamine enhancers” are considered to be “antidepressants,” while (a) many of them are used for other disorders (e.g., anxiety, OCD, PTSD) and (b) nonmonoamine enhancer drugs are also effective in the treatment of depression.

MONOAMINE ENHANCERS AND THE TREATMENT OF DEPRESSION

Monoamine enhancers are medications that raise monoamine levels by reuptake inhibition, receptor activity alteration, or metabolism inhibition. This group includes groups such as the “tricyclic antidepressants” and the MAO-I, SSRI, SNRI, and serotonin partial agonists. Trazodone, nefazodone, and mirtazapine are considered Monoamine *Modulators* due to their multimodal mode of action. While still considered the hallmark in the treatment of depression, monoamine enhancers and modulators are diverse and play a role in the treatment not only of depression but of other disorders such as generalized anxiety disorder (GAD), post-traumatic stress disorder (PTSD), and OCD.

A heterogeneous disorder, depression varies between individuals in several aspects. While the monoamine hypothesis of depression was the catalyst for the biologic study of the disorder, it is recognized today as a multifactorial one. Hence, the treatment of depression requires a tailored approach, combining the different clinical presentations, symptomatology, biologic/health parameters, concomitant medications, and neuroscientific insights (e.g., complicated neuronal interactions, connectivity, plasticity, and relevant circuitry). Often, optimal treatment of depression includes medications outside the monoamine enhancers class, as some dopamine blockers, cation channel blockers, and lithium (see below) have proven to possess independent efficacy in the treatment of depression.

For instance, aripiprazole, a partial dopamine D2 and serotonin 5HT1A receptor agonist, is approved by the FDA as adjunctive therapy for major depressive disorder. Likewise, quetiapine, first approved for the treatment of schizophrenia in 1996, can also be used in the treatment of depression and is considered a first-line treatment for bipolar depression in part for its activity as a serotonin 5HT1A partial agonist. Other examples of drugs that are widely used either as monotherapy or as augmenting agents for depression despite not being considered “antidepressants” (according to current nomenclature) are olanzapine, lithium, and lamotrigine (Fig. 33.1c–1).

Respectively, while some medications have an antidepressant activity while not being considered “antidepressants,” other medications termed “antidepressants” are often prescribed for conditions different from depression. Moreover, not all monoamine enhancers are the same; some effectively treat disorders while others do not. One example is related to the treatment of OCD; it is not treated with “antidepressants” but with serotonin reuptake inhibitors (SRI). Yet very effective “antidepressants” such as desipramine are inferior in treating OCD. In addition to being used in treating depression (e.g., the first-line therapy for depression), SRIs are also used as an approved treatment for anxiety and stress-related disorders. Similar is the class of serotonin-norepinephrine reuptake inhibitors (SNRI) (e.g., duloxetine, venlafaxine), a class of monoamine enhancers indicated for chronic pain. The norepinephrine–dopamine reuptake inhibitor (NDRI) bupropion is indicated both for depression and as an aid to smoking cessation.

Finally, inaccuracy is not limited only to the clinical indications but also to the neurobiologic functioning itself. For example, while the term SNRIs refers mostly to agents such as venlafaxine, duloxetine, or milnacipran, the “tricyclic antidepressants” imipramine, and clomipramine, although sharing similar properties, are not labeled as such.

To conclude, NbN replaces the term “antidepressants” with “monoamine enhancers.” This approach helps to group and differentiate the various medications based on pharmacology and MoA while remaining clinically relevant. Consequently, “antidepressants” such as clomipramine (a “tricyclic antidepressant”) and venlafaxine (“SNRI”) are now considered to be of the same group of SNRI, despite different molecular structures.

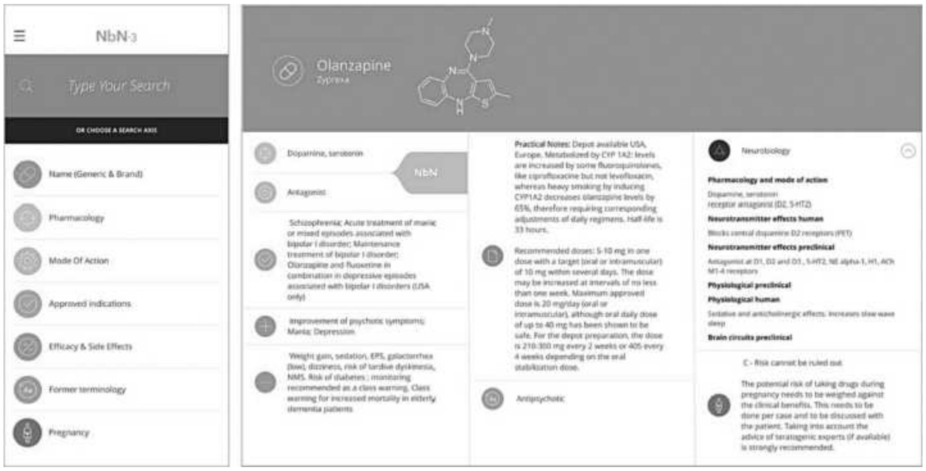

Name (Generic & Brand)
Pharmacology
Mode Of Action
Approved indications
Efficacy & Side Effects
Former terminology
Pregnancy

Olanzapine
Zyprexa

Dopamine, serotonin
Antagonist
NbN

Schizophrenia; Acute treatment of manic or mixed episodes associated with bipolar I disorder; Maintenance treatment of bipolar I disorder; Olanzapine and fluoxetine in combination in depressive episodes associated with bipolar I disorders (USA only)

Improvement of psychotic symptoms; Mania; Depression

Weight gain, sedation, EPS, galactorrhea (low), dizziness, risk of tardive dyskinesia, NMS. Risk of diabetes ; monitoring recommended as a class warning. Class warning for increased mortality in elderly dementia patients

Practical Notes: Depot available USA, Europe. Metabolized by CYP 1A2: levels are increased by some fluoroquinolones, like ciprofloxacine but not levofloxacin, whereas heavy smoking by inducing CYP1A2 decreases olanzapine levels by 65%, therefore requiring corresponding adjustments of daily regimens. Half-life is 33 hours.

Recommended doses: 5-10 mg in one dose with a target (oral or intramuscular) of 10 mg within several days. The dose may be increased at intervals of no less than one week. Maximum approved dose is 20 mg/day (oral or intramuscular), although oral daily dose of up to 40 mg has been shown to be safe. For the depot preparation, the dose is 210-300 mg every 2 weeks or 405 every 4 weeks depending on the oral stabilization dose.

Antipsychotic

Neurobiology

Pharmacology and mode of action

Dopamine, serotonin
receptor antagonist (D2, 5-HT2)

Neurotransmitter effects human

Blocks central dopamine D2 receptors (PET)

Neurotransmitter effects preclinical

Antagonist at D1, D2 and D3, 5-HT2, NE alpha-1, H1, ACh M1-4 receptors

Physiological preclinical

Physiological human

Sedative and anticholinergic effects; increases slow wave sleep

Brain circuits preclinical

C - Risk cannot be ruled out

The potential risk of taking drugs during pregnancy needs to be weighed against the clinical benefits. This needs to be done per case and to be discussed with the patient. Taking into account the advice of teratogenic experts (if available) is strongly recommended.

FIGURE 33.1c–2. NbN two main domains, pharmacology and mode of action, and additional dimensions. *Approved Indications* dimension states which indications are officially approved by major regulatory bodies (e.g., Food and Drug Administration [FDA], European Medical Agency [EMA]); the *Efficacy and Side Effects* dimension specifies whether there is evidence to support the compound’s use in indications additional to the ones approved by regulators (e.g., narcolepsy for methylphenidate, mirtazapine as a sleep promoter). Under “side effects,” only frequent or life-threatening side effects are listed; the Practical Notes dimension contains clinical knowledge including metabolism and dosing. The *Pregnancy* dimension displays current information on safety of use during pregnancy and lactation. Finally, the *Neurobiology* dimension summarizes empirical and clinically relevant data from preclinical and clinical studies.

CATION CHANNEL BLOCKERS, LITHIUM, AND THE TREATMENT OF BIPOLAR DISORDER

“Mood Stabilizers” is a term often applied to medications used in bipolar disorder. This class of medications includes agents that were first used for epilepsy (e.g., carbamazepine, valproate, lamotrigine) and lithium.

The term “mood stabilizers” is problematic for several reasons. For once, these drugs are commonly used for psychiatric indications other than bipolar disorder (e.g., unipolar depression, schizoaffective disorder, aggression, and impulsivity). Secondly, many other agents also possess mood-stabilizing properties, mainly dopamine blockers.

Replacing this term with NbN is challenging, as knowledge of how the therapeutic psychiatric effect is achieved is yet to be discovered. The term “cation channel blockers” emphasizes the important role of voltage-sensitive sodium channels (VSSC) and calcium channels in the basic function of those medications. However, it is not limiting the role of “mood stabilization” to these agents alone and does not limit the “stabilizing” effect to “mood” alone.

Due to its long-standing enigmatic MoA, lithium stands as its own pharmacologic domain (which is still unknown) as an “enzyme modulator.”

ADDITIONAL LAYERS TO NBN AND CLINICAL AIDS

While the pharmacology and MoA domains represent the main pharmacologic properties of each medication, they do not tell the whole story. Since finer details often separate one drug from the other, each drug in the new nomenclature has a profile with additional fields (dimensions) that hold relevant information–**Approved Indications, Efficacy and Side Effects, Practical Notes, Pregnancy, and Neurobiology** (see Fig. 33.1c–2 caption). Two clinical aids were made public to help utilize contemporary knowledge and make it accessible to clinicians and patients-the NbN application and website (see Fig. 33.1c–2). They demonstrate the "full picture" that each medication possesses.

Finally, two complementary versions of the NbN app aim to widen its relevance to child and adolescent (C&A) psychiatry and the general public (patient and family, P&F). NbN C&A contains information relevant to medications used for children and adolescents. To help demonstrate the scientific basis of psychiatric medication and consequently to decrease stigma, a third application and website, NbN P&F, was developed, providing information based on the NbN concept in a simple language.

CONCLUSION

Current neuroscience knowledge indicates that mental disorders are heterogeneous and need precise treatment. For instance, many patients do not respond to first-line treatment for depression, and there is often a need to select a specific treatment for their particular depression (considering genetics, medical history, symptoms, and course). At this point in therapy, overinformed or misinformed patients are likely deterred by the fact they are being prescribed medications labeled utterly different than what they suffer from (e.g., "antipsychotics" for nonpsychotic patients or "antidepressants" for anxiety).

During the initiation and management of psychiatric disorders, disease-based nomenclature often requires clarification, explaining to a confused patient that it is indeed the appropriate treatment despite the drug "class." This confusion may add further distress to an already tense situation, as the patient may feel their condition was not well-understood or their treatment does not match their symptoms, potentially harming the trust needed for treatment adherence and success. Furthermore, terms such as "mood stabilizer," "antipsychotic," and "sedative" may be perceived as judgmental. This may increase shame, especially in a social climate where the stigma on mental health and mistrust toward psychiatry still exists.

From the clinician's point of view, grouping drugs by indication has little use when prescribing. Whether they are called "antidepressants," "antipsychotics," "sedatives," or "hypnotics," this vocabulary does not represent differences or similarities between one another, either by neuroscientific attributes or by clinical application. The use of the current indication-based terminology neither encourages psychiatrists to utilize contemporary evidence-based knowledge for prescribing medications nor helps to change the habitual prescribing of a limited number of drugs.

An NbN presents a shift from disease-based to pharmacology-driven classification. It supports the application of scientific rationale to the prescription process for psychiatric medications, accommodates contemporary insights, and helps to avoid patient confusion. By bridging the gap between psychiatry and other fields of medicine, it presents the rich fabric of the neuroscience basis of psychotropics.

The NbN application is available on Android and iOS devices free of charge. It is constantly updated based on new scientific insights and feedback. It was designed for daily practice with an entirely searchable database, taking advantage of the multiple axes the classification is built on. Various versions, including NbN and NbN C&A, are suitable for clinicians, while NbN P&F was designed to "translate" the NbN "language" to the public. The NbN website (NbN.science) is also accessible, offering the same databases along with additional information.

FURTHER READINGS

Creese I, Burt DR, Snyder SH. Dopamine receptor binding: differentiation of agonist and antagonist states with 3H-dopamine and 3H-haloperidol. *Life Sci*. 1975;17(6):933–1001.

Fineberg NA, Hollander E, Pallanti S, et al. Clinical advances in obsessive-compulsive disorder: a position statement by the International College of Obsessive-Compulsive Spectrum Disorders. *Int Clin Psychopharmacol*. 2020;35(4):173–193.

Frazer A, Blier P. A neuroscience-based nomenclature (NbN) for psychotropic agents. *Int J Neuropsychopharmacol*. 2016;19(8):pyw066.

Gerasch S, Kanaan AS, Jakubovski E, Müller-Vahl KR. Aripiprazole improves associated comorbid conditions in addition to tics in adult patients with Gilles de la Tourette syndrome. *Front Neurosci*. 2016;10:416.

Goodwin GM, Haddad PM, Ferrier IN, et al. Evidence-based guidelines for treating bipolar disorder: revised third edition recommendations from the British Association for Psychopharmacology. *J Psychopharmacol*. 2016;30(6):495–553.

Krieser AE, Sanderson AG, Vik M, Myers JA. Effects of isonicotinic acid hydrazide in mentally ill patients. *Dis Chest*. 1953;23(1):28–35.

Laborit H, Huguenard P, Alluaume R. [A new vegetative stabilizer; 4560 R.P.]. *Presse Med*. 1952;60(10):206–208.

Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med*. 2005;353(12):1209–1223.

Loy JH, Merry SN, Hetrick SE, Stasiak K. Atypical antipsychotics for disruptive behaviour disorders in children and youths. *Cochrane Database Syst Rev*. 2017;8(8):CD008559.

Lykouras L, Zervas IM, Gournellis R, Malliori M, Rabavilas A. Olanzapine and obsessive-compulsive symptoms. *Eur Neuropsychopharmacol*. 2000;10(5):385–387.

Möller HJ, Schmitt A, Falkai P. Neuroscience-based nomenclature (jNbN) to replace traditional terminology of psychotropic medications. *Eur Arch Psychiatry Clin Neurosci*. 2016;266:385–386.

National Collaborating Centre for Mental Health (UK). Bipolar disorder: the NICE guideline on the assessment and management of bipolar disorder in adults, children and young people in primary and secondary care. *The British Psychological Society and The Royal College of Psychiatrists*. 2014.

Nutt DJ, Blier P. Neuroscience-based nomenclature (NbN) for journal of psychopharmacology. *J Psychopharmacol*. 2016;30(5):413–415.

Pringsheim T, Okun MS, Müller-Vahl K, et al. Practice guideline recommendations summary: treatment of tics in people with Tourette syndrome and chronic tic disorders. *Neurology*. 2019;92(19):896–906.

Schildkraut JJ. The catecholamine hypothesis of affective disorders: a review of supporting evidence. *Am J Psychiatry*. 1965;122(5):509–522.

Schirmbeck F, Zink M. Clozapine-induced obsessive-compulsive symptoms in schizophrenia: a critical review. *Curr Neuropharmacol*. 2012;10(1):88–95.

Sinyor M, Schaffer A, Levitt A. The sequenced treatment alternatives to relieve depression (STAR*D) trial: a review. *Can J Psychiatry*. 2010;55(3):126–135.

Snyder SH. The dopamine hypothesis of schizophrenia: Focus on the dopamine receptor. *Am J Psychiatry*. 1976;133(2):197–202.

Uchida H, Yamawaki S, Bahk WM, Jon DI. Neuroscience-based nomenclature (NbN) for clinical psychopharmacology and neuroscience. *Clin Psychopharmacol Neurosci*. 2016;14(2):115–116.

Vieta E, Sanchez-Moreno J. Acute and long-term treatment of mania. *Dialogues Clin Neurosci*. 2008;10(2):165–179.

Welsch P, Üçeyler N, Klose P, Walitt B, Häuser W. Serotonin and noradrenaline reuptake inhibitors (SNRIs) for fibromyalgia. *Cochrane Database Syst Rev*. 2018;2(2):CD010292.

WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC classification and DDD assignment. 2021. Oslo, 2020.

Wilkes S. The use of bupropion SR in cigarette smoking cessation. *Int J Chron Obstruct Pulmon Dis*. 2008;3:45–53.

Zohar J, Insel TR. Obsessive-compulsive disorder: psychobiological approaches to diagnosis, treatment, and pathophysiology. *Biol Psychiatry*. 1987;22(6):667–687.

Zohar J, Kasper S. Neuroscience-based nomenclature (NbN): a call for action. *World J Biol Psychiatry*. 2016;17(5):318–320.

Zohar J, Stahl S, Moller HJ, et al. A review of the current nomenclature for psychotropic agents and an introduction to the neuroscience-based nomenclature. *Eur Neuropsychopharmacol*. 2015;25(12):2318–2325.

33.2 Dopamine Blockers ("Antipsychotics")

33.2a Dopamine Antagonists

Martin T. Strassnig, M.D., and Philip D. Harvey, Ph.D.

Neuroscience Based Nomenclature (NbN)	
Amisulpiride, fluphenazine, haloperidol, perphenazine, pimozide, sulpiride	
Pharmacologic Domain	Dopamine
Mechanism of action	Receptor antagonists (D2)
Chlorpromazine, loxapine, thioridazine	
Pharmacologic Domain	Dopamine, serotonin
Mechanism of action	Receptor antagonists (D2, 5HT2)

HISTORY

There is now emerging agreement that antipsychotic drugs should be differentiated based on their individual side-effect profiles in addition to their efficacy. 5-HT$_{2A}$ and 5-HT$_{C}$ inverse agonism (or antagonism), in addition to dopamine D2 antagonism, has been a mainstay of dividing newer from older antipsychotics, but several of the (older) first-generation agents (e.g., clozapine [Clozaril], thioridazine [Mellaril], melperone [Buronil], and cyamemazine) have considerable 5-HT$_{2A}$ antagonism as well, making the distinction less useful. Moreover, it has become increasingly clear that SGAs differ from each other in many properties and are not a homogeneous class either. As a result, the recently proposed Neuroscience-based nomenclature (NbN) aims to realign psychotropics in terms of their pharmacologic properties. For antipsychotics, instead of the previously common first-versus second-generation antipsychotics division, their underlying mechanism is emphasized over a disease-specific nomenclature. That is, for “antipsychotics,” dopamine–receptor antagonists and dopamine–serotonin antagonists will henceforth be used. This allows for a mechanism-rather than disease-based classification while also rendering the previous dichotomy of first-versus second-generation antipsychotics obsolete.

Some of the dopamine D2 receptor antagonists are represented among the first group of effective agents for schizophrenia and other psychotic illnesses. They include fluphenazine, haloperidol, amisulpride, pimozide, and perphenazine. Although many of these drugs have gradually been replaced by dopamine–serotonin, serotonin–dopamine, and serotonin–dopamine–norepinephrine antagonists in the United States, dopamine antagonists remain the most commonly prescribed antipsychotics in many parts of the world, and as cost considerations have become more prominent, dopamine D2 receptor antagonists remain frequently prescribed in the United States as well.

These agents are highly effective for nearly all disorders that result in psychotic symptoms. However, these drugs have important limitations: They can cause acute and chronic neurologic symptoms that, especially at higher doses, significantly limit their effectiveness; they are ineffective or only partially effective for a substantial proportion of psychotic patients; the minimally effective dosage has frequently not been determined, and therefore patients may be over-or undermedicated; and they have limited efficacy for some of the most important signs and symptoms of the most common psychotic illness, schizophrenia. In higher dosages, the motor symptoms may induce “secondary” negative symptoms, such as akinesia, emotional blunting, apathy, and cognitive impairment.

Despite their limitations, dopamine blockers have played an important role in the management of psychosis. The modern era of psychopharmacology began with the discovery of chlorpromazine (Thorazine), the first effective antipsychotic, in the early 1950s. The annual increase in state and county mental hospital beds in the United States that had occurred since the beginning of the 20th century reversed sharply after the arrival of antipsychotics. This occurred because most psychotic patients with schizophrenia and other psychotic illnesses became clinically stable enough to be discharged from hospitals after they received drug treatment. Moreover, patients with stabilized illnesses became candidates, at least in theory, for psychosocial treatments and rehabilitation. There is also evidence suggesting that treatment with dopamine blockers can improve the long-term course of schizophrenia.

The dopamine blockers have wider applicability than for the treatment of schizophrenia. They are effective for treating psychosis regardless of its cause. As a result, dopamine blockers are utilized for many patients with psychoses associated with physical origins, bipolar illness, major depression with psychotic features, and schizoaffective disorder, predominantly in the acute phases of illness.

The first phenothiazines were the first effective antipsychotic drugs. Drugs from this class were originally used as antihelminthics in veterinary medicine and as urinary antiseptics in humans. In 1950, Paul Charpentier, at the Rhóne-Poulenc Laboratories in Paris, synthesized chlorpromazine, a drug that appeared notable as a mild antihistaminic that was also sedating. Henri Laborit, a surgeon with an interest in drugs that would decrease preoperative anxiety and prevent postsurgical shock, convinced the Rhóne-Poulenc Laboratories of the importance of identifying compounds that would relax patients and make surgical shock less likely. Chlorpromazine appeared to fulfill these criteria, but it also appeared qualitatively different from other sedating drugs. Patients appeared indifferent toward the environment and more tranquil. These properties led Laborit to convince psychiatrists, including Jean Delay and Pierre Deniker, to administer chlorpromazine to patients with mania and schizophrenia. In 1952, these authors reported their observation that chlorpromazine was effective in treating mania and schizophrenia. Within a year, the atmosphere in Paris’s psychiatric hospitals had improved substantially. The discovery of chlorpromazine’s effectiveness for psychosis represents the beginning of modern psychopharmacology in that it indicated that drugs could have specific effects on the symptoms of mental disorders that went beyond sedation. Moreover, the availability of drugs with direct effects on mental disorders became a powerful tool for laboratory studies with animals, leading to a greater understanding of how drugs affect the brain and information about the nature of the illnesses that were affected.

The use of chlorpromazine was reported by Lehman and Hanrahan from Montreal in 1954, and it rapidly captured the interest of U.S. psychiatrists. In 1955, Henry Brill, the assistant state mental health commissioner in New York, initiated the general use of chlorpromazine and reserpine in state hospitals in that state. Although reserpine was also an effective antipsychotic, its side effects led clinicians to prefer chlorpromazine, and, as a result, it is no longer used for treating psychosis because of its side effects, such as depression, suicidality, dysphoria, and orthostatic hypotension.

In 1953, reports surfaced indicating that chlorpromazine could cause neurologic adverse effects, predominantly parkinsonian symptoms. Although some individuals warned that these symptoms were evidence of neural toxicity, others suggested that the dosage should be increased until patients experienced a mild form of parkinsonism, termed the “neuroleptic threshold” dose. Later that year, clinicians noted that chlorpromazine could also result in forms of persistent dyskinesia, a syndrome that would later be called tardive dyskinesia (TD). Thus, information about the side effects of the family of dopamine-blocking medications has been available since their introduction.

The introduction of chlorpromazine, itself a dopamine-serotonin antagonist, was followed by the introduction of other prototypical dopamine D2 antagonists, including the phenothiazines perphenazine (Trilafon) and fluphenazine (Prolixin). In 1958, the first effective butyrophenone, haloperidol (Haldol), was introduced by Paul Janssen from Janssen Pharmaceutica in Beerse, Belgium. The first diphenylbutylpiperidine (Orap) was introduced in 1963. The first long-acting pure dopamine D2 blocker, fluphenazine enanthate, was introduced in the early 1960s. Although this and subsequent long-acting agents, such as haloperidol decanoate, had similar effectiveness and side effects to other antipsychotics, they provided a method for administering an antipsychotic without requiring patients to take daily oral medication, improving treatment adherence and reducing relapse and rehospitalization rates.

CHEMISTRY

The dopamine blockers have historically been referred to by a number of titles, including neuroleptics, major tranquilizers, and antipsychotics (typical or FGAs). The designation neuroleptic was suggested by Deniker, who noted the tendency of these agents to have neurologic side effects and thus appear to “take the neuron,” based on the observation that neurologic side effects are inevitably associated with antipsychotic activity. However, the effectiveness of drugs such as clozapine suggests that causing neurologic side effects are not necessary to treat psychosis. Thus, referring to a side effect of these agents rather than the primary clinical effect may be misleading. The term *major tranquilizer* referred to the strange quietness or blandness (or *ataraxia,* Latin) that had been associated with those agents. However, the suggestion that these agents function by their tranquilizing effects is also misleading.

In this section, the dopamine receptor antagonists, per NbN, are classified according to their pharmacologic mechanism (Table 33.2a–1). This new method is probably more useful to clinicians since this classification provides information about the mechanism rather than being entirely disease-based and potentially nonspecific, broadening the therapeutic armamentarium beyond psychosis (“antipsychotics”). The antipsychotic drugs come from a number of chemical classes, each with somewhat different pharmacokinetic properties. The differences can have important effects on the clinical use of the drugs. The phenothiazines benzamides have a somewhat different structure from the butyrophenones and the diphenylbutylpiperidines, which also share similar metabolic pathways. This overview will focus on haloperidol as a representative compound. Since the pharmacokinetics of the long-acting agents differs substantially from the short-acting agents, they are discussed separately.

Table 33.2a–1.
Dopamine D2 Receptor Antagonists: Adverse Effects

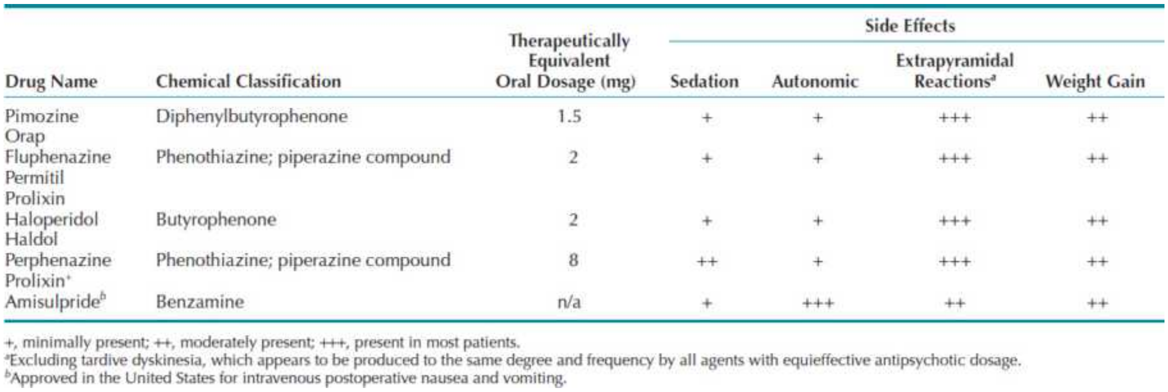

			Side Effects			
Drug Name	Chemical Classification	Therapeutically Equivalent Oral Dosage (mg)	Sedation	Autonomic	Extrapyramidal Reactions[a]	Weight Gain
Pimozine Orap	Diphenylbutyrophenone	1.5	+	+	+++	++
Fluphenazine Permitil Prolixin	Phenothiazine; piperazine compound	2	+	+	+++	++
Haloperidol Haldol	Butyrophenone	2	+	+	+++	++
Perphenazine Prolixin[a]	Phenothiazine; piperazine compound	8	++	+	+++	++
Amisulpride[b]	Benzamine	n/a	+	+++	++	++

+, minimally present; ++, moderately present; +++, present in most patients.

[a] Excluding tardive dyskinesia, which appears to be produced to the same degree and frequency by all agents with equieffective antipsychotic dosage.

[b] Approved in the United States for intravenous postoperative nausea and vomiting.

Phenothiazines

All the phenothiazines have the same three-ring structure with different side chains joined at the nitrogen atom of the middle ring (Fig. 33.2a–1). The activity of the group can be affected by substitutions at position 2 or

10. The phenothiazines are usually categorized into three classes based on substitutions at position 10. These substitutions have important effects on the pharmacologic characteristics of the phenothiazines. Thus, substituting an *aliphatic* side chain results in a drug with relatively low antipsychotic potency, such as chlorpromazine. These agents also tend to be sedating and cause anticholinergic and hypotensive effects at their effective dosages. Chlorpromazine has a chlorine atom at position 2. If the chlorine atom is eliminated, the resulting drug is promazine (Sparine), a weak antipsychotic. Substituting a *piperidine* ring at position 10 leads to an effective group of antipsychotics characterized by thioridazine and mesoridazine (Serentil). These drugs have similar potency and side effects when compared with the aliphatic phenothiazines. Fluphenazine (Prolixin) and trifluoperazine (Stelazine) are representative antipsychotics with a *piperazine* group substituted at the 10th position. Piperazine phenothiazines have fewer anticholinergic and autonomic effects but a greater affinity for D2 sites and, as a result, produce more EPS. A number of piperazine phenothiazines have been esterified at a free hydroxyl group with ethanoic or decanoic acid to produce injectable long-acting antipsychotics.

Butyrophenones

This class of antipsychotic drugs is characterized by a substituted phenyl ring, which is attached to a carbonyl group attached by a 3-carbon chain to a tertiary amino group. Most of the clinically useful butyrophenones have a piperidine ring attached to the tertiary amino group. Haloperidol is the representative drug for this class. It and other butyrophenones tend to be potent D2 antagonists and have minimal anticholinergic and autonomic effects, resulting in reduced sedation compared to phenothiazines. Haloperidol, a substituted piperidine, is the most commonly used (and oldest) drug in this class. Other important butyrophenones include droperidol (Dropletan), a short-acting, sedating compound that is commonly used to control agitation; and spiperone (Spiroperidol), an agent with very high affinity for D2, 5-HT_{2a}, and 5-HT_{2b} receptors that are commonly utilized for labeling receptors in animal studies and in PET scanning.

Diphenylbutylpiperidines

This group of agents is similar in structure to the butyrophenones. Pimozide (Orap), the only diphenylbutylpiperidine available in the United States (indication solely for Tourette syndrome), is derived from the butyrophenone benperidol. It has a long half-life: 20 to 26 hours. Doses of 20 mg or higher are associated with seizure risk and QTc prolongation. This agent has been removed from the market for psychotic conditions because of those reasons. Penfluridol (Semap) and fluspirilene (experimental) are antipsychotics from this group that are not available in the United States.

Benzamides

NbN–amisulpiride, sulpiride

Amisulpride (Solian), a sulfur-containing benzamide, has an unusual property in that at lower doses (100 mg to 300 mg a day), it tends to stimulate rather than block dopamine release, producing antidepressant effects, rendering it potentially useful for the treatment of schizophrenia patients with negative symptoms. At higher, standard doses, amisulpride may not have any direct effects on negative symptoms. Amisulpride was not approved by the Food and Drug Administration for use in the United States until February 2020, but it has been used in Europe, Israel, Mexico, India, New Zealand, and Australia to treat psychosis and schizophrenia. Its current U.S. approval is as an intravenous infusion for postoperative nausea and vomiting (10-mg dose).

Thioxanthenes

This group shares a similar three-ring structure with the phenothiazines but has a carbon atom substituted for nitrogen at position 10. Chlorprothixene (Taractan) is an effective low-potency *aliphatic* thioxanthene agent with an antipsychotic potency and a side-effect profile similar to its phenothiazine counterpart, chlorpromazine. Thiothixene (Navane), clopenthixol (Sordinol), and flupenthixol (Depixol) are thioxanthenes with a *piperazine* substitution. Again, their clinical characteristics resemble those of their phenothiazine counterparts. As a result of their double bond at position 10, these agents have geometric isomers, with the *cis* isomers demonstrating greater antipsychotic activity.

PHENOTHIAZINES

Aliphatic

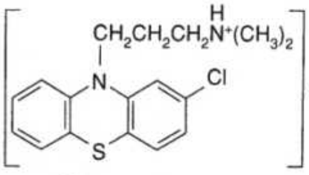
Chlorpromazine

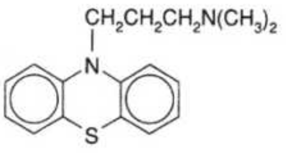
Promazine

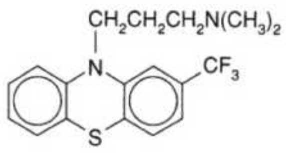
Triflupromazine

Piperazine

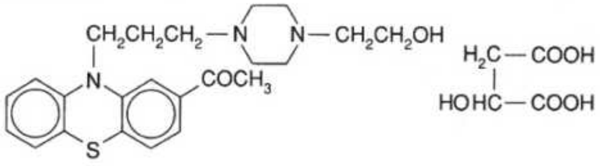
Acetophenazine

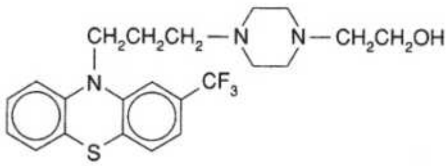
Fluphenazine

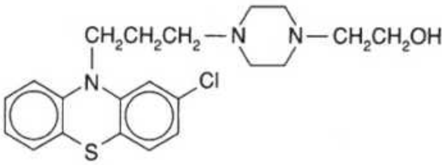
Perphenazine

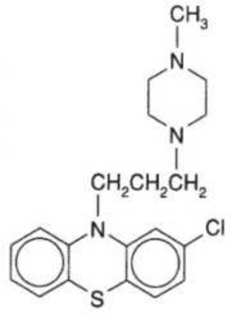
Prochlorperazine

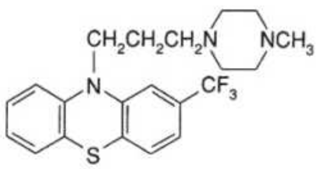
Trifluoperazine

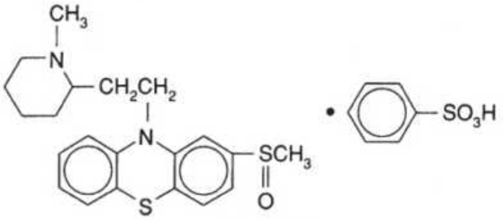
Mesoridazine

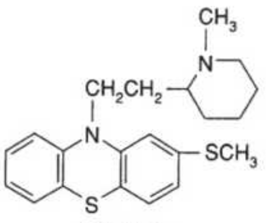
Thioridazine

THIOXANTHENES

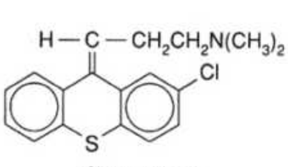
Chlorprothixene

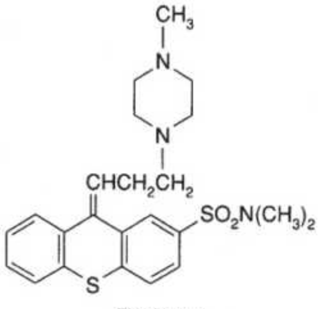
Thiothixene

A

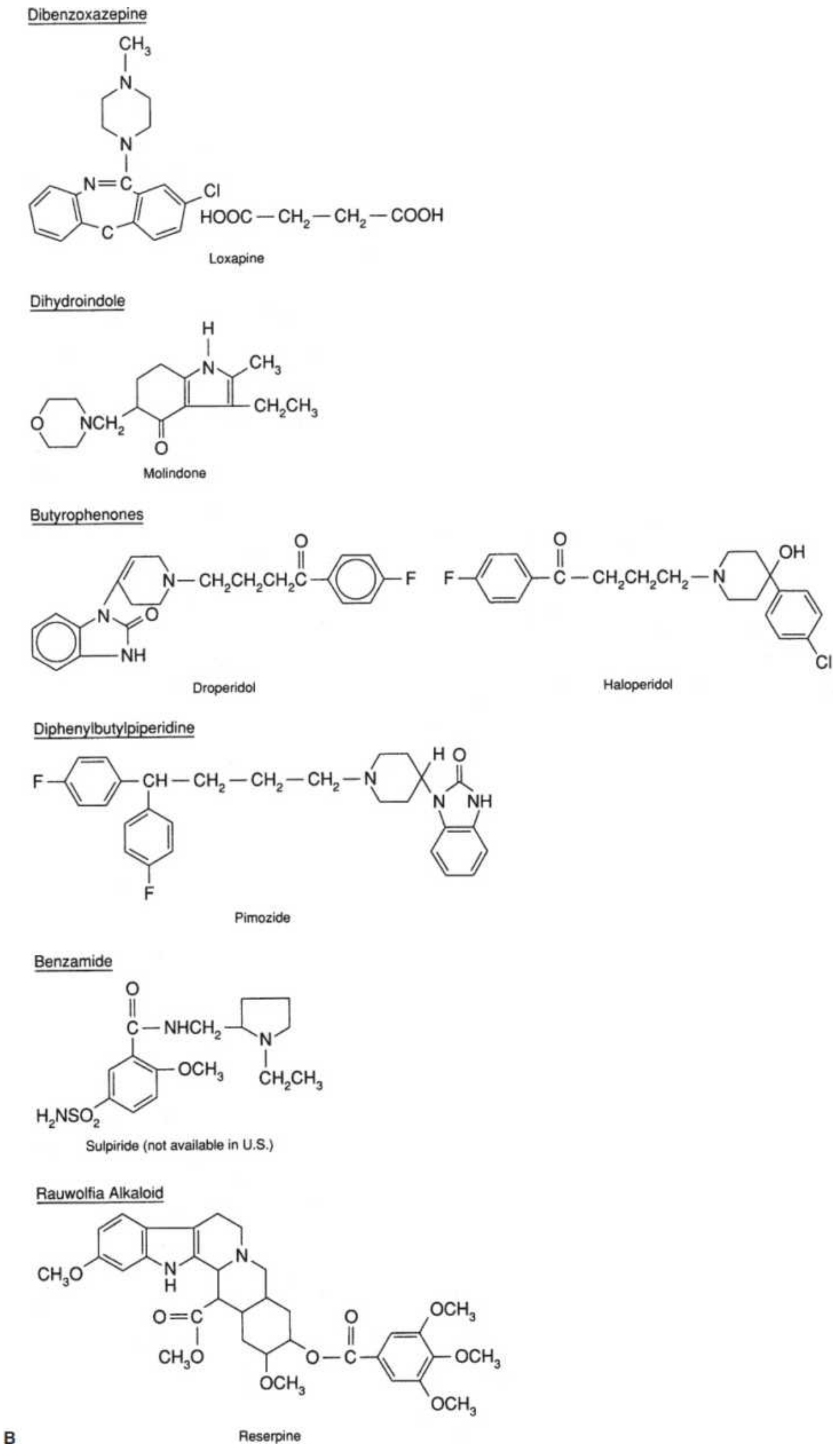
Dibenzoxazepine
CH3
N
N
N=C
Cl
C
HOOC—CH2—CH2—COOH
Loxapine
Dihydroindole
H
N
CH3
CH2CH3
O
NCH2
O
Molindone
Butyrophenones
O
N—CH2CH2CH2C
F
N
O
NH
Droperidol
F
C—CH2CH2CH2—N
OH
Cl
Haloperidol
Diphenylbutylpiperidine
F
CH—CH2—CH2—CH2—N
H O
N NH
F
Pimozide
Benzamide
O
C—NHCH2
N
CH2CH3
OCH3
H2NSO2
Sulpiride (not available in U.S.)
Rauwolfia Alkaloid
CH3O
N
H
N
O=C
CH3O
OCH3
O—C
O
OCH3
OCH3
OCH3
B
Reserpine

FIGURE 33.2a–1. Molecular structure of dopamine receptor antagonists and reserpine.

Dibenzoxazepines

This class of antipsychotic drugs has a three-ring structure with a seven-member center ring. Loxapine (Loxapac), a dibenzoxazepine, is the only drug from this group that is available in the United States. Clozapine, a dibenzodiazepine, differs from loxapine in having nitrogen instead of an oxygen atom in the middle ring, as well as differences in side chains.

Dihydroindoles

These derivatives are structurally related to serotonin, melatonin, and indole hallucinogens, such as dimethyltryptamine. The only dihydroindole available in the United States was molindone (Moban), the only antipsychotic agent without known weight gain liability. Molindone was temporarily discontinued because of a business decision by its generic manufacturer and was reintroduced in 2018.

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Plasma concentrations of short-acting drugs rise rather rapidly during what is referred to as the absorption phase and then decline during a distribution, metabolic, and elimination phase. Each of these three phases is influenced by different factors related to the chemical structure of the agent (Table 33.2a–2).

Absorption

In general, the dopamine D2-blockers are well-absorbed when they are administered orally or parenterally. As with most drugs, oral administration leads to less predictable absorption than parenteral administration. Liquid concentrates are absorbed slightly more rapidly than pills. Plasma concentrations of the drugs usually reach peak levels 1 to 4 hours after ingestion and 30 to 60 minutes after intramuscular (IM) administration. In general, IM preparations reach their peak concentrations sooner than oral drugs and, as a result, have an earlier onset of action. For example, IM administration of most antipsychotics results in peak plasma levels in about 30 minutes, with clinical effects emerging within 15 to 30 minutes. Oral administration of most FGAs results in a peak plasma level in 1 to 4 hours. A number of factors can interfere with the GI absorption of these drugs, including antacids, coffee, smoking, and fatty foods.

Drugs usually reach a steady state in about three to five half-lives. Thus, steady-state levels for haloperidol and many other dopamine blockers are reached in about 3 to 5 days since their half-lives are about 24 hours. The bioavailability (i.e., the amount of drug reaching the systemic circulation) is substantially greater—as much as 10-fold—when dopamine D2 blockers are administered parenterally. This difference may result from the incomplete absorption of the drug in the GI tract and from the extensive metabolism of oral drugs during the first pass, mainly through the liver.

Distribution

The initial decline in the plasma concentration of an antipsychotic (or any drug) results primarily from drug distribution into various bodily compartments. Since antipsychotics are highly lipophilic, they tend to accumulate in tissues such as fat, lungs, and brain. Brain levels of antipsychotics tend to be substantially higher than plasma levels. Since plasma and brain levels are in equilibrium, measuring plasma levels is sufficient for estimating brain levels. However, behavioral effects may last substantially longer than the blood levels indicate. For instance, sexual side effects may last for several weeks after the last dose of chronic treatment.

Most dopamine D2 blockers are highly protein bound. For example, more than 90% of drugs such as fluphenazine and haloperidol are bound to plasma protein. The remaining unbound portion is the drug that is available for passing through the blood–brain barrier. In theory, conditions that alter the amount of plasma protein, such as malnutrition, will also alter the amount of bioavailable antipsychotic drugs. This is potentially important for drugs that are protein bound for 98% or more.

Metabolism and Elimination

The metabolism of antipsychotic drugs is largely hepatic and occurs through conjugation with glucuronic acid, hydroxylation, oxidation, demethylation, and sulfoxide formation. Haloperidol has only one major metabolite, reduced haloperidol, which has substantially less antidopaminergic activity than the parent compound. However, reduced haloperidol is converted back to the parent compound and, in this manner, may contribute to the antipsychotic activity.

Most dopamine D2 blockers are metabolized by the cytochrome P450 2D6 and P450 3A subfamilies in the liver. Since these same isoenzymes also metabolize a number of drugs that are commonly combined with antipsychotics, a number of important drug–drug interactions are possible. These are discussed in the section on drug–drug interactions.

Patients demonstrate substantial differences in the extent to which they metabolize these compounds. This explains, in part, the marked differences among patients in the oral neuroleptic dosages required for treatment. This, along with differences in absorption, may also explain the 20-fold variation in plasma concentrations that can be found among patients treated with the same dosage of a dopamine D2 blocker.

Table 33.2a–2.
Pharmacokinetics of Selected Dopamine D2 Receptor Blockers

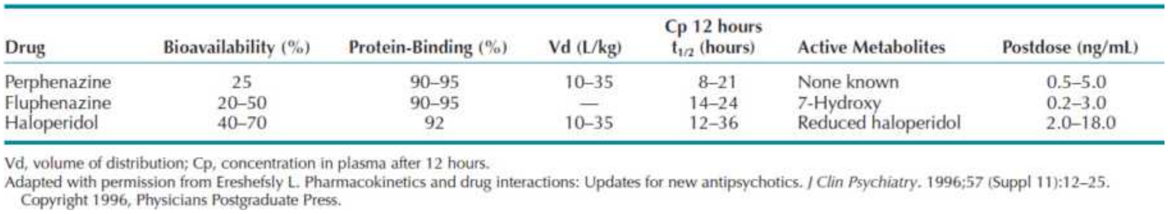

Drug	Bioavailability (%)	Protein-Binding (%)	Vd (L/kg)	Cp 12 hours $t_{1/2}$ (hours)	Active Metabolites	Postdose (ng/mL)
Perphenazine	25	90–95	10–35	8–21	None known	0.5–5.0
Fluphenazine	20–50	90–95	—	14–24	7-Hydroxy	0.2–3.0
Haloperidol	40–70	92	10–35	12–36	Reduced haloperidol	2.0–18.0

Vd, volume of distribution; Cp, concentration in plasma after 12 hours.

Adapted with permission from Ereshefsly L. Pharmacokinetics and drug interactions: Updates for new antipsychotics. *J Clin Psychiatry*. 1996;57 (Suppl 11):12–25. Copyright 1996, Physicians Postgraduate Press.

The systemic clearance of FGAs is high as the result of a high-hepatic extraction ratio. As a result, only negligible amounts of the unchanged drug are excreted by the kidneys. Phenothiazines and their metabolites are excreted in the urine and feces.

Pharmacokinetics of Long-Acting Antipsychotics

The pharmacokinetics of long-acting injectable antipsychotics differs markedly from those of short-acting oral and injectable drugs. Long-acting fluphenazine and haloperidol are administered as esters dissolved in sesame oil. The oil is injected into a muscle, and the drug gradually diffuses from the oily vehicle into the surrounding tissues. The rate-limiting step appears to be the rate of diffusion since once the drug enters the tissue, it is rapidly hydrolyzed, and the parent compound is released. Long-acting compounds, however, are absorbed continually during the interval between injections. Moreover, for patients who have received multiple injections, the drug will be absorbed from multiple injection sites simultaneously. As a result, it takes long-acting compounds much longer to reach a steady state, and they are eliminated much more slowly than their oral counterparts. For example, the decanoate forms of haloperidol and fluphenazine each require about 3 months to reach steady state, and substantial plasma concentrations can be detected up to several months after therapy has been discontinued. This has consequences for prolonged side effects and potentially for TD risk, even after discontinuation of therapy.

Older Adult Patients. Older adult patients require lower-than-usual dosages of dopamine D2-blockers for a number of reasons. These include decreased renal clearance, decreased cardiac output, decreased liver size, and decreased CYP450 enzyme. In addition, older adult patients may be more sensitive to the parkinsonian side effects of the drugs. Recently, dopamine and dopamine-serotonin blockers have been associated with increased mortality risk in older adult patients with dementia.

PHARMACODYNAMICS

The pharmacodynamics of an antipsychotic describes its activity at important sites of action, particularly at its brain receptors. Recent findings in neuroscience have provided important information about the effects of dopamine D2 blockers at a number of different receptor sites, how these effects influence groups of neurons, and how activity at receptors is translated into intracellular events.

MECHANISM OF ACTION OF DOPAMINE RECEPTOR ANTAGONISTS

The Dopamine Hypothesis

The dopamine hypothesis states that antipsychotics reduce psychotic symptoms by decreasing dopamine activity. It was originally proposed by Arvid Carlsson from Sweden and other basic scientists based on the observation that haloperidol and chlorpromazine raised the levels of dopamine metabolites in dopamine-rich areas of the mouse brain, interpreted as dopamine antagonism, and compensation of self-regulating systems by increasing dopamine production. Other investigators noted that dopamine D2 blockers (and dopamine-serotonin blockers) inhibit the activation and stereotypical behaviors of rodents induced by amphetamine. This effect has been found to be a reasonably reliable predictor of antipsychotic activity in patients. Moreover, the dose-response relationships in animals provide information about the likely clinical dosage range for patients. These behaviors are thought to be mediated by dopamine and have been used as screening devices to identify compounds that are likely to have clinical efficacy as antipsychotics. Possibly the most convincing but clearly indirect evidence supporting a role of dopamine in psychosis comes from the observation that all drugs—at least to this date—capable of reducing psychosis are also dopamine antagonists, including partial dopamine agonists. Moreover, agents that are associated with increased dopamine activity, such as amphetamine, methylphenidate, or cocaine, tend to worsen symptoms in patients with schizophrenia. At persistent high dosages, these drugs may cause psychosis in individuals without pre-existing psychiatric disorders.

Table 33.2a–3.
Dopamine Receptor Occupancy of Currently Used Dopamine D2 Receptor Antagonists in Relation to Clozapine

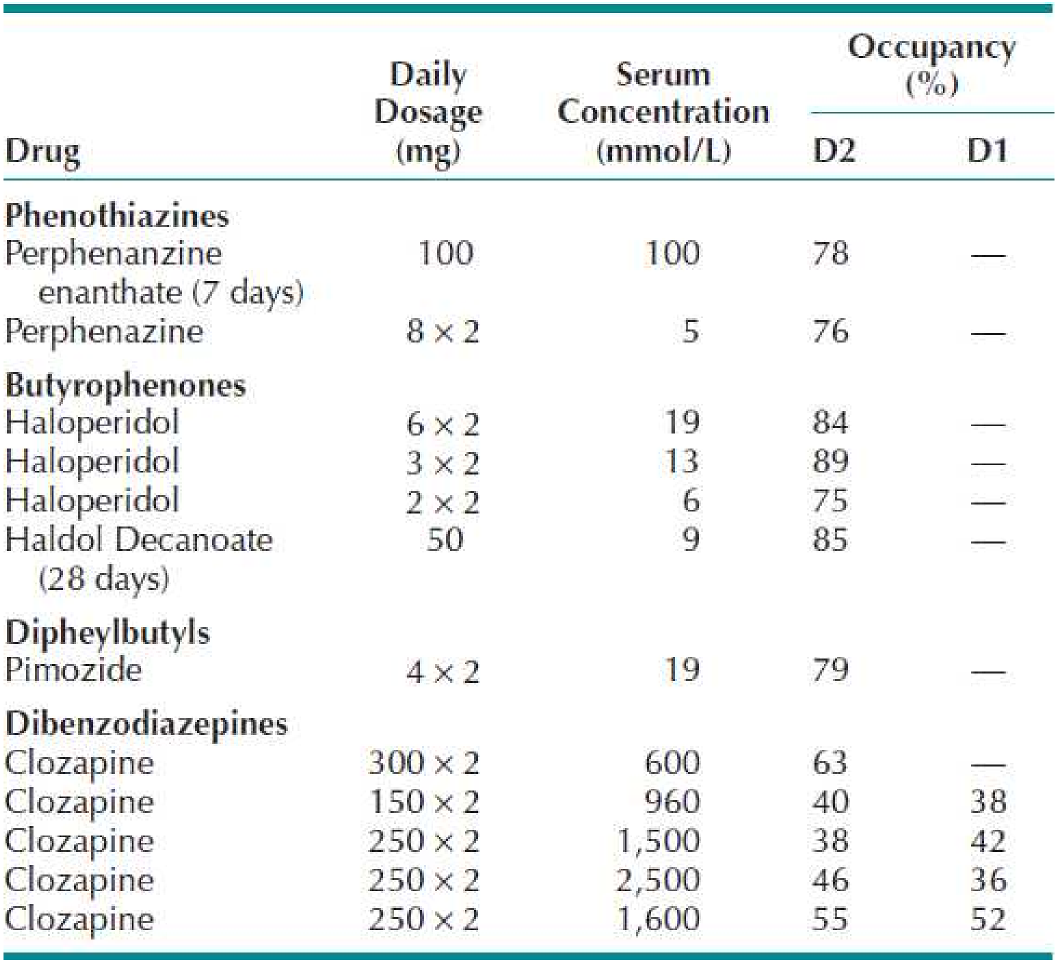

Drug	Daily Dosage (mg)	Serum Concentration (mmol/L)	Occupancy (%)	
			D2	D1
Phenothiazines				
Perphenanzine enanthate (7 days)	100	100	78	—
Perphenazine	8 × 2	5	76	—
Butyrophenones				
Haloperidol	6 × 2	19	84	—
Haloperidol	3 × 2	13	89	—
Haloperidol	2 × 2	6	75	—
Haldol Decanoate (28 days)	50	9	85	—
Dipheylbutyls				
Pimozide	4 × 2	19	79	—
Dibenzodiazepines				
Clozapine	300 × 2	600	63	—
Clozapine	150 × 2	960	40	38
Clozapine	250 × 2	1,500	38	42
Clozapine	250 × 2	2,500	46	36
Clozapine	250 × 2	1,600	55	52

A role for dopamine in their antipsychotic activity, originally proposed by Carlsson, the Nobel laureate, was given strong support by the work of Solomon Snyder from Johns Hopkins University and Philip Seeman from the University of Toronto in the mid-1970s. Using similar receptor binding methods, both found that the antipsychotic potency of a number of drugs was highly correlated with the drug's affinity for D2 receptors. More recent studies using highly selective D2 receptor ligands and PET indicate that dopamine receptor antagonists are effective when approximately 80% of D2 receptors in the striatum are occupied. Higher occupancy rates may be associated with more parkinsonian side effects without greater effectiveness. In contrast, clozapine is effective when only 40% to 60% of D2 receptors are occupied (Table 33.2a–3). Although it remains unclear why clozapine is effective at lower levels of D2 occupancy, it has been suggested that activity at other receptors, particularly 5-HT_{2A} receptors, may contribute to clozapine's clinical activity and to its major active metabolite, *N*-des-methyl clozapine (NDMC), which has partial D2/3 receptor agonist activity in addition to being a strong 5-$HT_{2A/2c}$ inverse agonist (or antagonist). The 5-HT_{2A} antagonism by itself has been suggested to have antipsychotic effects, albeit not as strong as D2/3 antagonism (Fig. 33.2a–2). NMDA receptor agonism at the glycine site may increase its efficacy as well.

Several studies using PET scanning have found that there is no difference in D2 receptor occupancy in the striatum between individuals who respond to an antipsychotic and those who do not. This finding suggests that processes beyond the dopamine receptor affect the response, that dopamine blockade in the striatum is not the mediator of behavior change, or that there are subgroups of patients whose psychosis has origins different from the majority who do respond to treatment.

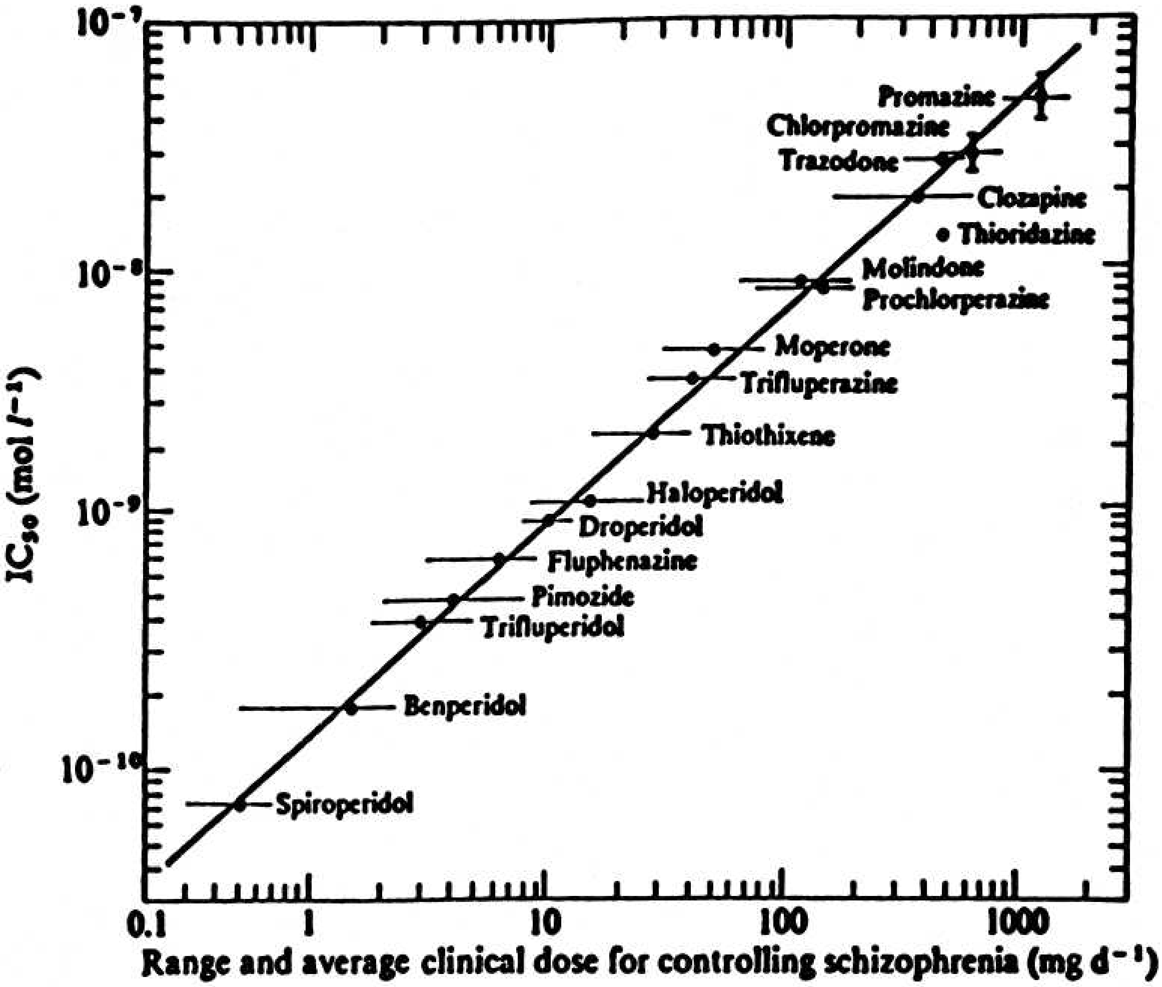

FIGURE 33.2a–2. Blockage of dopamine release. (From Seeman P, Lee T, Chau-Wong M, Wong K. Antipsychotic drug doses and neuroleptic/dopamine receptors. *Nature.* 1976; 261(5562):717–719, with permission from Springer Nature. doi: 10.1038/261717a0.)

The Effects of Dopamine D2 Blockers on the Dopamine System

Dopamine neurons with cell bodies in the ventral tegmental area (the A_{10} cell group in rat brain) project via the meso-limbo-cortical dopamine system to the nucleus accumbens, amygdala, and the neocortex. Dopamine cells in the substantia nigra (the A_9 cell group) project to the caudate putamen via the nigrostriatal system. It has been proposed that activity in the meso-limbo-cortical system explains the antipsychotic activity of drugs, whereas neurologic side effects such as EPS and TD result from activity in the nigrostriatal system. Thus, drugs with preferential activity at meso-limbo-cortical sites would result in clinical effectiveness with fewer neurologic side effects. However, the reason that the dopamine-serotonin blockers have fewer EPS has been attributed to having additional effects on the 5-HT$_{2A}$ receptors, which increases dopamine turnover.

Advances in molecular biology have led to the cloning of a number of dopamine receptor subtypes. The discovery of these subtypes has raised the question of their possible role in the treatment of psychosis. Dopamine receptors can be classified into two groups: D1-like (D1 and D5) and D2-like receptors (D2, D3, D4). Moreover, each of these receptor subtypes may have variants that have the potential to explain heterogeneity in drug responses. As mentioned previously, the antipsychotic potency of dopamine receptor antagonists is highly correlated with their affinity for D2 receptors. This relationship has not been established with other dopamine receptor subtypes. Nevertheless, certain characteristics of receptor subtypes suggest that they may contribute to clinical responses. For example, a role for D4 receptors has been suggested by clozapine's high affinity for this subtype. Unfortunately, multiple attempts to bring selective D4 antagonists to the market have failed for lack of antipsychotic efficacy. The localization of D3 receptors in limbic areas suggests that these subtypes may be clinically important for antipsychotic effects, and some antipsychotics have substantial binding at both D2 and D3. In contrast, D1 receptor blockade may be undesirable, as D1 stimulation in the cortex is associated with better cognitive functioning, and selective D1 antagonists have been reported to cause suicidality.

Studies of the dopamine system in animals and man have suggested important interactions between the subcortical and prefrontal dopamine systems that may affect responses to dopamine blockers. Lesions in the prefrontal cortex of rats' result in increases in dopamine activity in the striatum, suggesting that these two systems have important functional linkages. Similarly, increases in dopamine activity in the prefrontal cortex are associated with decreases in striatal activity. Also, a number of investigators propose that negative symptoms of schizophrenia are associated with decreased dopamine activity in the prefrontal cortex, whereas psychotic symptoms are generated by subcortical increases in dopamine activity in the striatum and temporal lobe. In this model, an effective antipsychotic covering both positive and negative symptoms would increase dopamine activity in prefrontal areas and decrease activity in subcortical areas.

Other clinical evidence for the role of dopamine in the antipsychotic effects of dopamine receptor antagonists comes from studies of homovanillic acid (HVA), a metabolite of dopamine. In these investigations, changes in plasma HVA after the introduction of an antipsychotic are used to reflect changes in dopamine activity in the brain. These studies indicate that plasma HVA rises soon after patients begin treatment with an antipsychotic and then falls below pretreatment levels in the subset of patients who improve clinically. Moreover, a number of studies indicate that patients with high levels of plasma HVA prior to drug treatment are more likely to improve than those with lower levels. These findings suggest that improvement in an antipsychotic agent is associated with an overall reduction in dopamine activity. Since subcortical areas have the greatest contribution to dopamine activity in the brain, this finding may not reflect cortical dopamine activity.

Studies that monitor the intracellular firing rates of dopamine neurons have also contributed to an improved understanding of the effects of antipsychotics. Chiodo, Grace, and Bunney have reported that the administration of haloperidol to rats resulted in a short-term increase in the firing of midbrain dopamine neurons followed by a prolonged decrease in firing that has been referred to as “depolarization inactivation.” It has been proposed that the antipsychotic effects of these drugs are associated with a decrease in the firing rate. Taken together with the studies of plasma HVA, these findings suggest that for an antipsychotic response to occur, the blockade of D2 receptors is an initial response that must be followed by a decrease in dopamine activity.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Design

Next to chlorpromazine, dopamine D2 blockers were the first specific compounds identified for treating a major mental illness. Their effectiveness, conclusively proven, challenged the views of some clinicians who believed that psychosis was a psychological reaction related to early experiences and that only psychological treatment could have antipsychotic effects. This initial skepticism was valuable in that it helped inspire the development of controlled clinical drug trials aimed at demonstrating the effectiveness of medication in treating psychosis. Early studies, such as those initiated by the Department of Veterans Affairs (VA) and the National Institute of Mental Health (NIMH), were influential in establishing objective research methods. One principle, now widely acknowledged, is for an antipsychotic (or any other new drug) to be considered effective; it should yield a better outcome than a placebo. That principle controls for nonspecific improvement that occurs because of a number of factors, such as (1) spontaneous fluctuations in symptoms; (2) the rate of spontaneous improvement for the illness; (3) the expectation of improvement, which is related to starting a new drug; (4) the subject's protection from environmental and psychosocial stresses that may occur with hospitalization; and (5) other environmental factors such as the hospital milieu and psychological treatments that can lead to improvement.

Two other principles that reduce sources of bias are random assignment to a treatment group and double-blind evaluation methods. Random assignment ensures that differences in outcome between the groups (usually between those receiving a drug and those receiving a placebo or between two different active groups) are not attributable to patients with different expected outcomes or drug responses being assigned to one group. An example would be a study in which patients who are more severely ill are assigned to a particular group or if patients who are predicted to improve spontaneously are assigned preferentially to the drug group rather than the placebo group. Double-blind methods were developed to ensure that bias by patients or by those who evaluate the outcome does not influence the study. For example, a person who performs a clinical rating may be prejudiced toward rating patients who are receiving a drug as more improved than patients who are receiving a placebo. Double-blind methods also protect against bias in the manner in which treatment is administered. For example, a treating clinician may give more attention to a patient who is receiving a certain drug.

A number of modifications in trial methods have been developed since the first studies aimed at treating psychosis, although the principles of random assignment and blindness remain important. For example, treating acutely ill patients with schizophrenia with dopamine D2 blockers has become a standard of care and has led some to question the ethical use of placebos when effective treatments are available. Alternative methods for demonstrating the efficacy of a compound have included showing that it is as effective as an antipsychotic agent with proven efficacy or showing that a dosage thought to be subtherapeutic is less effective than a higher dosage. Since maintaining unawareness of treatment assignments is often difficult to preserve when a drug has known side effects, active placebos with side effects but without CNS effects have been used. However, inexperienced evaluators tend to overinterpret side effects as being secondary to active treatment rather than to active placebo, which also is associated with similar side effects, even if less frequent. Recent research has indicated that high interrater reliability and centralized ratings may decrease variance even further, while the increased number of sites and more rating scales, raters, and rater changes may increase variance and therefore decrease the likelihood of separation from placebo.

INTERPRETATION

Acute Psychosis

Numerous double-blind clinical trials have demonstrated that dopamine D2 blockers are more effective than placebo for reducing symptoms of psychosis. The initial studies were carried out in the 1960s by the NIMH, reviewed by Donald Klein and John Davis, who concluded that with the exception of mepazine (Pacatal) and promazine, all of these agents were effective. These and subsequent studies also demonstrated that antipsychotics are effective for every subgroup of patients with schizophrenia. In 2009, the Schizophrenia Patient Outcomes Research Team (PORT) updated prior reviews and concluded that there is clear evidence that dopamine D2 blockers are effective for reducing positive symptoms during acute psychotic episodes. PORT recommendations also suggest that choosing antipsychotics should be based on individual preference, prior treatment response and adherence history, medical history and risk factors, the individual medication's side-effect profile, and long-term treatment planning. Studies found that amisulpride can be safely administered at an optimal dose immediately, not needing a titration phase and thereby aiding in acute symptom control.

Controlled clinical trials have demonstrated that dopamine D2 blockers are effective in reducing psychotic symptoms, such as hallucinations and paranoid ideation, to a greater extent than anxiety and depression. Improvements individual patient experiences on these agents vary considerably. For many patients, dopamine D2 blockers can lead to a complete remission of psychotic symptoms. The early studies from the NIMH indicate that 60% of drug-treated patients improved to the extent that they achieved a complete remission (albeit temporarily) or experienced only mild symptoms; the remaining 40% will continue to experience psychotic symptoms, and approximately 8% of them are not improved with drug treatment. A comprehensive meta-analysis published in 2019 in JAMA that compared 32 oral antipsychotics showed that amisulpride was the second most effective antipsychotic at reducing overall symptoms in schizophrenia patients (behind clozapine) and the most effective in terms of reducing positive symptoms. The analysis also ranked amisulpride better than clozapine in terms of tolerability and side effects.

Positive psychotic symptoms, which include hallucinations and delusions, are more responsive to dopamine D2 receptor antagonists than are negative symptoms, such as blunted affect, emotional withdrawal, and lack of social or vocational interest, or disorganized symptoms, such as impaired attention and impaired communication. Although these latter two groups of symptoms are less responsive than psychotic symptoms, double-blind studies indicate that these symptoms improve to some degree. Amisulpride, at lower doses, tends to stimulate rather than block dopamine release, producing antidepressant effects, rendering it potentially useful for the treatment of schizophrenia patients with negative symptoms. EPS can cause secondary negative symptoms, compounding symptom burden. Patients with parkinsonism may demonstrate fewer expressive gestures and diminished interest in their environment. Anticholinergic antiparkinsonian medications, aimed at mitigating parkinsonian and other EPS symptoms, sometimes recommended prophylactically when used with dopamine D2 receptor antagonists, have been consistently shown to impair memory in patients who require these drugs, while EPS may worsen cognitive test scores that require intact motor function.

As a group, women tend to respond better and to require lower dosages than men. Unfortunately, there are no biologic or clinical indicators that can be used to predict treatment response other than the patient’s prior responses.

The time course of treatment response in acute psychosis has been studied in some detail. Most patients who manifest an eventual positive response to dopamine D2 blocker treatment manifest improvements in their psychotic symptoms very early after the initiation of treatment. Studies have shown that by 3 or 7 days, eventual responders are showing detectable changes in symptoms. It has been reported that the maximum treatment gains with dopamine D2 blockers occur by 2 weeks, and there is little evidence of delayed onset.

Maintenance Therapy

A number of studies have compared dopamine D2 blockers to placebos for preventing relapse in patients who have been stabilized following a psychotic episode. These studies provide strong support for continuing these medications when patients have recovered from a psychotic episode. A review of 24 double-blind studies of maintenance antipsychotic therapy found that in all studies, many more patients relapsed when they were not receiving an active drug. A meta-analysis that included most of these studies reported that approximately 72% of patients will relapse in a year on placebo compared to only 23% on a conventional antipsychotic. A more recent review cited 66 studies in which antipsychotics were withdrawn. The mean relapse rate was 53% for patients withdrawn from medication, and 16% for patients maintained on an antipsychotic. The more robust the remission prior to discontinuation (the lower the psychosis ratings) is, the longer the time until relapse.

A number of investigations have focused on the usefulness of continuing medication for patients who have recovered from a first episode of schizophrenia. Again, maintenance medication showed a substantial advantage, although relapse rates were lower for remitted first-episode patients when compared to multiepisode patients. Other studies indicate that patients who have been well and stable on an antipsychotic for years have a high likelihood of relapse when their medications are discontinued.

Earlier studies by Donald Johnson suggest other advantages for maintenance of antipsychotics in schizophrenia. He found that patients who relapsed when they were receiving dopamine D2 blockers had episodes that were less severe than those in patients who had discontinued their drugs. Drug-maintained patients were less likely to have episodes of self-destructive behaviors, violence, and antisocial acts when compared with patients who were not receiving medication. Patients who were not receiving medication were also more likely to require involuntary hospitalization. Patients who discontinued their medications ended up, on average, receiving more total medication over time because the dosage needed to treat acute relapses is much higher than the dosages usually prescribed for preventing relapse. Many of these findings have since been replicated. The common denominator is that maintenance therapy with antipsychotics benefits patients with schizophrenia and substantially reduces relapse risk and reduces hospitalization rates. Depot antipsychotics appear to be more effective for relapse prevention than oral medications. This would suggest there is a more prominent place for depot antipsychotics in relapse prevention in schizophrenia, including haloperidol and fluphenazine decanoate. One recent study pointed out that there was no difference in relapse rates between the long-acting dopamine D2 blocker haloperidol and dopamine–serotonin blocker paliperidone over 24 months, although side effects rates differed. Haloperidol decanoate caused more and more frequent EPS, while paliperidone palmitate caused more hyperprolactinemia (McEvoy et al., 2014).

Medication compliance is a serious problem for patients with schizophrenia, with up to 50% of patients failing to comply with recommended (oral) treatments. Importantly, only a fraction of new patients adheres to their initial antipsychotic medication after their first hospitalization, with some estimates as high as 95% over a multiyear period. A substantial proportion never even collect their initial prescription; the initiation of depot antipsychotics during hospitalization significantly reduces rehospitalization rates as compared to oral antipsychotics.

A number of statistics also indicate the limitations of pharmacotherapy for chronic schizophrenia. Even among treatment responders, as many as 40% are permanently disabled, and most of the other patients fail to recover to a full extent. In addition, more than 80% of patients are not fully employed and require public support. In large cities, individuals with schizophrenia comprise 14% to 50% of the homeless population. The rate of completed suicide is about 10%, with suicidal ideation and suicide attempts occurring in 70% or more of cases, and the mortality rate is increased due to greater accident proneness.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Effects of Dopamine D2-Blockers on Motor Systems

Animals administered high doses of a dopamine receptor antagonist develop a syndrome consisting of immobility, increased muscle tone, and abnormal postures called catalepsy. In addition, these agents decrease spontaneous motor activity in rodents. These motor effects in animals and humans are caused by dopamine receptor blockade in the striatum and the substantia nigra.

In humans, all dopamine D2-blockers produce EPS, in line with their antipsychotic potency, including parkinsonism (i.e., stiffness, akinesia, bradykinesia, tremor, and shuffling gait), dystonia (i.e., abrupt onset, sometimes bizarre muscular spasms affecting mainly the musculature of the head and neck), and akathisia (with objective and subjective restlessness). These effects differ from antipsychotic effects in that they have a more rapid onset and are potentially amenable to adjunct treatment. In addition, prolonged treatment with dopamine receptor antagonists can lead to the development of late-appearing abnormal movements of TD. Both short- and long-term motor effects in patients are discussed in a later section.

Neuroendocrine Effects

Antipsychotic drugs influence the secretion of hormones in the pituitary and elsewhere mainly due to their blockade of dopamine receptors. Prolactin secretion is tonically inhibited by dopamine in the pituitary gland. As a result, dopamine receptor antagonists can increase serum prolactin concentration in the clinical dosage range, in line with their antipsychotic potency. This elevation sometimes produces gynecomastia, galactorrhea, and erectile dysfunction in men. Dopamine D2-blockers also suppress levels of luteinizing hormone (LH) and follicle-stimulating hormone (FSH). These changes can lead to amenorrhea and inhibition of orgasms in women. There is also evidence indicating that Dopamine D2-blockers such as Haloperidol can inhibit the release of growth hormone (GH).

THERAPEUTIC INDICATIONS

Indications for dopamine receptor antagonists are shown in Table 33.2a–4.

Acute Psychotic Episodes in Schizophrenia and Schizoaffective Disorder

Schizophrenia is usually an illness that is characterized by periods of relapse and remission. Acute psychotic episodes are usually identified as periods during which patients demonstrate the emergence or worsening of positive or disorganized symptoms of schizophrenia, such as hallucinations, delusions, and disorganized behavior and communication.

Attention has focused on whether dopamine D2 blockers should be considered the preferred antipsychotics for acute episodes of schizophrenia. This issue was considered by a group of experts who met in 2001 to consider important issues in the pharmacotherapy of schizophrenia. The consensus opinion at that time was that dopamine–serotonin blockers should be favored over dopamine D2 blockers for first-episode patients or when a history of drug responses was unavailable. The group supported the use of dopamine D2 blockers over dopamine–serotonin blockers for patients who have responded well to dopamine D2 blockers without experiencing EPS; for patients who have responded better to dopamine D2 blockers than dopamine–serotonin blockers; and for patients who respond better to depot forms of dopamine D2 blockers.

Table 33.2a–4.
Indications for Dopamine D2 Receptor Antagonists

Acute psychotic episodes in schizophrenia and schizoaffective disorder
Maintenance treatment in schizophrenia and schizoaffective disorders
Mania
Depression with psychotic symptoms
Delusional disorder
Borderline personality disorder
Substance-induced psychotic disorder
Delirium and dementia
Mental disorders due to a medical condition
Childhood schizophrenia
Pervasive developmental disorder
Tourette syndrome
Huntington disease

Between 2005 and 2007, the results of large multicenter clinical trials provided important information about the relative effectiveness and side effects of dopamine D2- and dopamine–serotonin blockers. These results challenged the view of some that the newer agents were more effective and had fewer adverse effects than their predecessors. The most important studies were a NIMH trial known as CATIE (Clinical Antipsychotic Trial of Intervention Effectiveness), a randomized open trial from the United Kingdom National Health Service known as CUtLASS (Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study), and EUFEST (European First Episode Schizophrenia Trial), another randomized open trial completed in Europe. These trials are important because they were each relatively large for a nonpharmaceutical drug study, and they were neither sponsored nor designed by pharmaceutical company sponsors.

The first phase of CATIE compared the dopamine–serotonin blockers, olanzapine (Zyprexa; 7.5 to 30 mg per day), ziprasidone (Geodon; 40 to 160 mg per day), quetiapine (Seroquel; 200 to 800 mg per day) and the dopamine—serotonin–norepinephrine antagonist risperidone (Risperdal; 1.5 to 6.0 mg per day), to one dopamine-D2 antagonist, perphenazine (8 to 32 mg per day), a midpotency phenothiazine. The subjects were 1,493 patients with schizophrenia who, on average, were in their 40s and who had been ill for more than 14 years. The primary outcome was the time that patients remained on their assigned drug. The results, reported by Jeffrey Lieberman and coworkers in 2005, demonstrated similar efficacy of all drugs in terms of time to discontinuation. However, patients with TD were excluded from the perphenazine group, which may have given this agent an advantage. Patients treated with perphenazine were more likely to discontinue their medications for EPS, even with the exclusion of participants with TD. Patients receiving clozapine and olanzapine experienced weight gain and lipid elevations. Despite modest prolactin elevation, risperidone was the best-tolerated antipsychotic. The trial was too brief to compare the risk of TD among the different agents.

CUtLASS randomized 227 patients to either a dopamine–blocker, dopamine–serotonin blocker, or dopamine–serotonin–norepinephrine blocker. Prior to randomization, the patient’s psychiatrist selected the drug that the patient would be assigned. The results were published in a report by Peter Jones and coworkers. There was no significant difference in scores on the Quality-of-Life Scale, the primary outcome instrument. Moreover, there were trends suggesting advantages for the dopamine D2 blockers. The size of the sample was too small to compare the effectiveness of individual antipsychotics.

EUFEST compared several dopamine–serotonin blockers, olanzapine, quetiapine, ziprasidone with a low dose (1 to 4 mg) of the high-potency Dopamine D2 blockers haloperidol and amisulpride in the treatment of first-episode schizophrenia. The pragmatic trial found high-discontinuation rates (highest with haloperidol), and symptom reductions and rehospitalization rates were virtually the same in all groups. Alongside olanzapine, amisulpride had the lowest rates of discontinuation for any reason across an entire year. The results of these trials do not mean that older and newer antipsychotics are equivalent in their side-effect profiles. The differences in EPS liability may have been underestimated in CATIE. First, eliminating patients with evidence of TD may have excluded those who were most vulnerable to discomforting EPS side effects. Moreover, while there were no differences in EPS ratings for parkinsonism and akathisia between perphenazine and the SGAs, it is interesting to note that the higher rate of discontinuations for EPS among the perphenazine-treated patients suggests that ratings may not have been adequately sensitive. However, most of the studies documenting large differences between these drugs may have used what would now be considered excessive doses of haloperidol, a drug with an incidence of EPS, which may have exaggerated the difference between the two types of agents.

The current body evidence does not support the earlier belief that the dopamine–serotonin or dopamine–serotonin–norepinephrine blockers are more effective than dopamine D2 blockers. However, the dopamine D2 blockers are more likely to cause EPS and, as noted below, are probably more likely to cause TD. Perhaps dopamine D2 blockers should be prescribed in lower dosages than historical uses. If patients are treated with dopamine D2 blockers, they should be regularly monitored for EPS, particularly drug-induced parkinsonism and akathisia, and they should be assessed for evidence of TD.

Maintenance Treatment in Schizophrenia and Schizoaffective Disorders

Dopamine D2 blockers are effective in decreasing the risk of psychotic relapse in patients who have recovered from a psychotic episode. National and international guidelines indicate that after the acute phase of treatment for a first episode, treatment should continue for at least 1 year, but there is no consensus regarding the total duration of treatment, if the patient remains symptomatic. In case of multiple episodes, maintenance treatment should be continued for at least 3 to 5 years and often indefinitely.

Mania

Dopamine receptor antagonists are effective for reducing excitement and psychotic symptoms in mania. Dopamine D2 receptor antagonists tend to have a more rapid onset of action than antimanic drugs, including lithium (Eskalith), valproic acid (Depakene) (divalproex [Depakote]), and carbamazepine (Tegretol). This observation has led to the widespread practice of combining antipsychotic and antimanic drugs during the first days of treatment of severe excited states before the antimanic compound has its onset of action. When the antimanic compound has become effective, the dosage of the antipsychotic can usually be reduced, or the drug can be discontinued. Although there is only limited evidence supporting the use of dopamine D2 receptor antagonists for the long-term management of bipolar mood disorder, it is common clinical practice to manage some patients with severe illness with maintenance antipsychotics combined with antimanic agents. Some studies indicate that patients with mood disorders are more vulnerable to developing TD than schizophrenia patients, suggesting the importance of using dopamine D2 receptor antagonists as briefly as possible and only when clearly indicated. The advent of dopamine–serotonin antagonists has preempted these arguments because some of these agents are approved for acute, as well as maintenance treatment of bipolar disorder and are also commonly available as generic medications.

Depression with Psychotic Symptoms

Patients with major depressive disorder may develop psychotic symptoms such as hallucinations or delusions. These patients tend to demonstrate a more robust improvement when an antipsychotic is added to an antidepressant agent. Although most studies have focused on perphenazine, fluphenazine, and haloperidol, there is no evidence that any single dopamine receptor antagonist is more or less effective than any other for this indication, barring side-effect considerations. The added benefits of dopamine D2 receptor antagonists are most apparent for those patients who are tormented by severe delusions. When the psychotic symptoms have remitted, the antipsychotic should be withdrawn.

Delusional Disorder

Delusional disorder is characterized mostly by “nonbizarre” delusions, that is, delusions that are within the realm of possibility for the patient. Moreover, patients rarely hallucinate, and their behavior is rarely bizarre. Patients with delusional disorder will often demonstrate substantial improvement when they are treated with a dopamine receptor antagonist.

Borderline Personality Disorder

Borderline personality disorder is characterized by a pervasive pattern of instability that affects mood, interpersonal relationships, and self-image. Some patients with severe forms of this disorder may experience transient psychotic states, particularly when they are under stress. These states may be characterized by suspiciousness, ideas of reference, impulsiveness, and aggressiveness.

Low dosages of haloperidol have been found to be effective in some of these patients, although there is no evidence indicating that other dopamine D2 blockers are less effective. Unfortunately, most trials have been relatively brief, and it is unclear how medications affect the overall course of this disorder and how medications interact with psychotherapeutic management that these patients usually require.

Substance-Induced Psychotic Disorder

A number of drugs, including amphetamines, cocaine, alcohol, and phencyclidine, can cause psychotic symptoms that occur while the patient is intoxicated or during drug withdrawal. These symptoms are usually time-limited and remit within hours or days, depending on the concentration of drug ingested and the rate of elimination. Dopamine D2 blockers may reduce these substance-induced psychotic symptoms, reducing dangerous behaviors or unusual suffering. Benzodiazepines may be helpful in calming the patient until psychotic symptoms remit. Some authors recommend prescribing benzodiazepines rather than dopamine receptor antagonists for phencyclidine intoxication. In some circumstances, it may be helpful to follow the patient’s untreated course in order to differentiate drug-induced syndromes from schizophrenia or bipolar disorder.

Dopamine D2 receptor antagonists are also useful for patients with hallucinations or delusions that occur with alcoholism, particularly for chronic alcoholic hallucinosis. Benzodiazepines are more appropriate than antipsychotics for uncomplicated alcohol withdrawal, and dopamine receptor antagonists can be prescribed as an adjunct, if withdrawal delirium develops. However, dopamine receptor antagonists should be prescribed carefully since these drugs may increase the likelihood of seizures that occur during alcohol withdrawal.

Impulse Control Disorders

There is a long-standing but controversial practice of prescribing dopamine receptor antagonists for individuals with extremely poor impulse control and a propensity for violent behavior. This practice is not supported by controlled clinical trials and may be appropriate only when other measures such as the use of SSRIs, carbamazepine, lithium, or β-blockers have failed. In such instances, non–weight gain–inducing dopamine–serotonin and dopamine–serotonin–norepinephrine antagonists may be more appropriate.

Delirium and Dementia

Low doses of high-potency dopamine D2 receptor blockers, especially if administered intravenously, can be useful in the treatment of the psychotic symptoms and agitation associated with delirium that may be due to organic conditions. Interactions with the anticholinergic properties of low-potency drugs may exacerbate or prolong toxic deliriums caused by anticholinergic agents.

Dopamine D2 receptor blockers can be useful for reducing agitation and psychotic symptoms that are related to dementia, Wernicke syndrome, and Korsakoff syndrome. A number of carefully done double-blind studies have found that dopamine D2 receptor blockers are superior to placebo for treating agitated older adult patients. A review by Janicak found that 66% of geriatric patients with “organic” psychosis improved on an antipsychotic compared to 35% who improved on a placebo. Low dosages of dopamine blockers are usually preferable for older adult patients with dementia. The treatment of older adult patients is complicated since many take multiple medications and are vulnerable to drug–drug interactions, hypotension, and falling episodes, as well as an added anticholinergic burden. Nevertheless, the adverse effects of these drugs should be weighed against the dangers of agitation in older adult patients. Clinicians should be aware that akathisia can present itself as agitation in these patients and that the risk of TD is markedly increased with extended dopamine-blocking treatment in older patients.

Childhood Schizophrenia

Schizophrenia infrequently has its onset during childhood (before age 17). Termed early onset schizophrenia (EOS) and associated with higher rates of premorbid abnormalities, children with schizophrenia may have symptoms that are similar to those in adults with schizophrenia, with a similar thought disorder and perceptual abnormalities but the reduced occurrence of systematized delusions. Clinical experience indicates that children improve with antipsychotics, and there are a number of controlled trials available to demonstrate this effect. It is particularly important to treat children with the lowest effective dosage of an antipsychotic because of increased vulnerability to side effects. Children are more prone to developing extrapyramidal symptoms associated with dopamine D2 receptor antagonists but are also more likely to experience weight gain and metabolic abnormalities with these agents. Antipsychotic drugs may impair learning in this population.

Pervasive Developmental Disorder

Patients with pervasive developmental disorders may demonstrate periods of hyperactivity, screaming, and agitation with combativeness. Although there are few controlled studies for this disorder, patients tend to improve with dopamine D2-blocking agents. Since these agents may impair learning, there is concern that they are overprescribed in some settings.

Tourette Syndrome

Tourette syndrome is a neurobehavioral disorder that is associated with motor and vocal tics. Motor tics may involve a single muscle group or several, thus appearing as purposeful movements. Similarly, vocal tics may involve an inarticulate sound or actual words. Often the words are socially inappropriate or even obscene. Milder tics may not require pharmacotherapy. However, when the tics are disabling, antipsychotic medication can be effective in reducing the severity of both motor and vocal tics. Although haloperidol and pimozide are the most commonly used agents, there is little evidence indicating that other agents are less effective.

Huntington Disease

Huntington disease is a genetically transmitted disease that follows an autosomal dominant type of heritability with 100% penetrance. The median age of onset is 35 to 40 years, but childhood onset (high number of CAG repeats) and much later onsets also occur. Once the diagnosis is made, death will follow in 15 to 20 years. Patients who eventually develop Huntington disease may display psychiatric symptoms and cognitive impairment years before the diagnosis of Huntington disease is made. The most common psychiatric symptoms are cognitive impairment, depression, irritability, apathy, obsessive-compulsive symptoms, and suicidality. These may progress to a psychotic delusional state or sometimes are followed by manic or hypomanic episodes. Some individuals may demonstrate symptoms that closely resemble schizophrenia with delusions and hallucinations (less than 10%).

Dopamine, dopamine–serotonin, and dopamine–serotonin–norepinephrine blockers are believed to be useful during the early stages of the illness to reduce chorea. However, double-blind studies are lacking. They also may be effective in managing irritability, paranoia, hallucinations, violent tendencies, and bizarre behaviors. Dopamine D2 blockers are usually preferred, but all agents appear to be effective. Clinicians should be aware of acute EPS, particularly in the rigid forms of the disorder. There is no dosing guidance. At present, it is believed that the long-term benefit does not merit their use, particularly as TD, other EPS, and akathisia may resemble Huntington disease’s later-stage symptomatology. Recently, the FDA-approved tetrabenazine (Xenazine, an amine depletor, like reserpine) for the treatment of chorea in Huntington disease. Tetrabenazine should not be used to prevent the occurrence of choreoathetotic symptoms, as there is evidence that tetrabenazine has no neuroprotective benefits. However, dysphoria, depression, suicidality, and dysphagia have been noted. Deutetrabenazine (Austedo) is currently approved for the treatment of Huntington disease-associated chorea. No long-term studies are available.

Parkinson Psychosis and Lewy Body Disease

Dopamine blockers are contraindicated in these conditions because they have the potential to markedly worsen parkinsonian syndromes and may lead to precipitous dystonic reactions.

Psychiatric Disorders Due to a Medical Condition

Dopamine receptor antagonists are effective for a broad range of medical conditions that result in psychosis. This category includes endocrinopathies and psychoses associated with temporal lobe epilepsy. Dopamine receptor antagonists are also helpful for patients with psychotic symptoms that can result from steroids.

Other Indications

Antipsychotics are also effective for treating ballismus, an illness characterized by abnormal activity of the axial and proximal musculature. Because of their antiemetic effects, dopamine receptor antagonists are sometimes prescribed for severe emesis associated with chemotherapy or postsurgical nausea. Patients with intractable hiccups are sometimes aided by dopamine receptor antagonists. Metoclopramide (Reglan), a dopamine receptor antagonist, is commonly prescribed for patients with gastroesophageal reflux and delayed gastric emptying. For obvious reasons, psychosis associated with Parkinson disease cannot be treated with these agents.

Table 33.2a–5.
Neurologic Side Effects of Dopamine D2 Receptor Antagonists

Acute extrapyramidal syndromes
Akathisia
Acute dystonia
Drug-induced parkinsonism
Neuroleptic malignant syndrome
Chronic extrapyramidal syndromes
Tardive dyskinesia
Perioral tremor

PRECAUTIONS AND ADVERSE REACTIONS

Neurologic Side Effects

The most serious side effects of dopamine receptor antagonists are neurologic and are largely confined to the extrapyramidal motor system (Table 33.2a–5). These EPS are, in turn, separated into acute effects that occur in the first days or weeks of treatment and chronic effects that occur after patients have received medications for months or years.

Acute Extrapyramidal Syndromes

All the dopamine receptor antagonists are associated with EPS. Since EPS, particularly parkinsonism, is believed to be caused by decreased dopamine activity in the basal ganglia, it suggests that this toxic effect is inescapable given that blockade of D2 receptors is associated with the antipsychotic effect of these agents. EPS correlate with their affinity for D2 receptors in the basal ganglia. Acute EPS, which includes dystonia, akathisia, and drug-induced parkinsonism, have their onset relatively soon after the initiation of antipsychotic drug treatment. They remit soon after the drugs are discontinued. In contrast, tardive EPS syndromes occur after months of treatment and commonly persist after medications are discontinued.

Akathisia. Akathisia comes from the Greek word *akathizein,* which means inability to sit still. It is the most common acute manifestation of EPS and often the most distressing. One study found that 41% of patients treated with a dopamine antagonist had mild akathisia, and an additional 21% had moderate to severe akathisia. The most common clinical manifestations consist of shifting the weight from foot to foot, walking on the spot, inability to keep the legs still, feelings of inner restlessness, and the shifting of body positions in a chair. Patients with akathisia may describe a compelling urge to walk or to initiate movement. In mild cases, patients may experience a subjective feeling of restlessness but not show increased motor activity. Akathisia may appear on the second or third day of antipsychotic treatment but more frequently has its onset after 5 days.

Differentiating akathisia from behaviors that are related to the psychotic illness may be difficult. Feelings of restlessness or agitation may result in irritability or anxiety, or even aggressive behavior. Psychotic patients with akathisia may appear more hostile or belligerent. This could lead clinicians to incorrectly conclude that the patient requires an increase in medication, which, in turn, could worsen this side effect. Akathisia is also thought to be a correlate of poor antipsychotic drug response. The use of these agents by inexperienced practitioners in nursing homes may lead to behavior deterioration.

Acute Dystonia. Acute dystonia consists of intermittent and sustained spasms of the muscles of the head and the neck, leading to involuntary movements. They can be experienced as the most frightening side effect of dopamine blockers, particularly when patients experience the sudden onset of what may appear to be a form of paralysis.

Ten percent of drug-induced dystonia occurs during the first hours of drug treatment, and 90% occurs within the first 3 days. Common types of dystonia include opisthotonos, a rigid contraction of the back muscles with arching; retrocollis and torticollis of the neck; oculogyric crisis, a spasm in which an eye or both eyes are turned upward; macroglossia and tongue protrusion, which can lead to choking; and laryngeal dystonia. Rarely, dystonia of laryngeal or pharyngeal muscles can lead to sudden death.

Younger patients, particularly young males, are more likely to develop dystonia. They are more common when patients are treated with large dosages of dopamine blockers. Dystonia almost always respond rapidly to antiparkinsonian medications and can usually be prevented by either pretreatment with antiparkinsonian medications or by limiting the antipsychotic dosage prescribed.

Drug-Induced Parkinsonism. Patients on dopamine receptor antagonists may experience all the common motor symptoms of idiopathic parkinsonism, including rigidity, bradykinesia, shuffling gait, and tremor. These side effects commonly occur during the first 5 to 30 days of treatment and may persist until the dosage is lowered, treatment is initiated, or dopamine blockers are discontinued. The examination will usually reveal a positive glabella tap. The first evidence of drug-induced parkinsonism may be a diminished arm swing or decreased facial expressiveness. Risk factors for drug-induced parkinsonism include increasing age, dosage, a history of parkinsonism, and underlying basal ganglia damage (e.g., vascular insult).

In its milder forms, drug-induced parkinsonism may appear as a decrease in spontaneous gestures, a masked facial expression, apathy, unspontaneous speech, and difficulty in initiating usual activities, so-called akinesia. These symptoms may be difficult to distinguish from the negative or deficit symptoms of schizophrenia. Patients with akinesia may also appear to be depressed. The tendency of patients with akinesia to not cross their legs can be helpful in diagnosing this side effect.

Management of Acute Extrapyramidal Syndromes

Anticholinergic agents are the most effective treatment of EPS. For acute dystonias, IM or intravenous (IV) diphenhydramine (Benadryl) (50 mg, 25 mg in children) or IM or IV benztropine (Cogentin) (0.5 to 2 mg IM or IV) or other anticholinergics alleviate this uncomfortable and often frightening side effect within minutes after the drug enters the bloodstream. Oral anticholinergic agents are usually prescribed for the management of parkinsonism and often for akathisia. Reducing the dosage of the antipsychotic is often effective, but it is common for patients to experience EPS at the minimum dosage that is needed to manage psychosis. Amantadine (Symmetrel), an antiviral agent that affects dopamine release, is also effective in reducing most EPS. This antiviral agent should not be used for this indication as a first-line agent. Direct dopamine agonists (e.g., bromocriptine [Parlodel]) and indirect dopamine agonists (e.g., levodopa [Dopar]) are not preferred, as they may aggravate psychotic symptoms.

Anticholinergics are frequently, but not always, effective for treating akathisia. β-Blockers, particularly propranolol (Inderal; 10 to 80 mg a day), have been found to be effective for treating akathisia and are preferred by many clinicians. The effectiveness of β-blockers may result from akathisia being a disturbance in the normal balance between dopamine and norepinephrine. Studies have shown that 5-HT_{2A} antagonists can reduce haloperidol-induced akathisia.

Anticholinergic antiparkinsonian medications should be prescribed with caution in combination with low-potency FGAs and some antidepressants since the combined anticholinergic activity may lead to a number of side effects, including hyperthermia and delirium. A sustained tachycardia may also occur. Further caution is warranted in the use of these agents in individuals with glaucoma, ileus, urinary retention, or prostatic hypertrophy. Because many over-the-counter cold remedies, sedatives, and sleep aids contain anticholinergic agents and antihistamines, patients should be warned about the interaction between these agents and anticholinergic antiparkinsonian drugs.

Some controversy has arisen about the appropriateness of prescribing antiparkinsonian medications prophylactically when dopamine blockers are started. Reports indicate that 30% to 50% of patients receiving long-term treatment do not need anticholinergics. In 1990, the World Health Organization (WHO) published a consensus statement on the use of anticholinergics in patients receiving long-term antipsychotic treatment. The statement argued against the prophylactic use of these agents because of some of their side effects (e.g., the possible worsening of TD, the risk of anticholinergic toxicity, the impairment in memory, and the possible pharmacokinetic interactions with the primary treatment agent). The consensus group concluded that anticholinergics are overprescribed, that they should be given only as needed, and side effects should be treated first by lowering the antipsychotic dosage or by using other agents. However, reports indicate that anticholinergics may be helpful in managing forms of EPS that may not be obvious to the clinician and side effects. That is, milder forms of akathisia may appear as anxiety or irritability, and milder forms of akinesia may appear as negative symptoms such as decreased motivation and a lack of affect. In these individuals, long-term anticholinergics may be useful for optimal pharmacotherapy.

There are other circumstances in which prophylactic antiparkinsonian medications may be helpful. Because of the discomfort and anxiety associated with acute dystonia, prophylaxis is prudent when there is a substantial risk of this side effect, for example, when young patients are treated with high-potency dopamine blockers. Some patients and staff may interpret dystonia as evidence that they are allergic to a particular drug, which may lead to drug refusal and an “allergy” label in the patient chart. Prudent practice may include using prophylactic antiparkinsonian drugs in high-risk patients, with a gradual lowering of the dosage of this drug with regular evaluations for obvious and subtle manifestations of EPS over a 2- to 3-month period.

The availability of serotonin–dopamine antagonists, provides an important alternative for patients with EPS. Symptoms of acute EPS that cannot be adequately controlled with dosage reduction or an adjunctive medication will probably improve, if they are changed to a serotonin dopamine antagonist.

Neuroleptic Malignant Syndrome. NMS is an uncommon but potentially fatal complication of dopamine-blocking agents. Its main clinical features include (1) hyperthermia; (2) severe muscular rigidity; (3) autonomic instability, including hyperthermia, tachycardia, increased blood pressure, tachypnea, and diaphoresis; and (4) changing levels of consciousness. It usually presents as muscular rigidity and progresses to elevated temperature, fluctuating consciousness, and unstable vital signs. These symptoms are often associated with elevations in creatine phosphokinase (CPK) and aldolase. Less common are elevations in liver transaminases, leukocytosis, myoglobinemia, and myoglobinuria. Acute renal failure may also occur. Mortality in well-developed cases ranges from 20% to 30% and may be higher when depot forms are used. More expedient early recognition of NMS has led to a substantially decreased mortality rate. Readers may find the Neuroleptic Malignant Syndrome Information Service (NMSIS) (nmsis.org) website helpful for specific questions or for their physician hotline to aid in the diagnosis of NMS.

NMS is more common when high-potency dopamine blockers are prescribed in high doses and when the dosage is escalated rapidly. The syndrome is twice as common in men as in women and is more likely to be present in younger patients. Clinicians should be concerned about any patient who demonstrates severe muscular rigidity and a rising body temperature because early diagnosis and treatment can be life-saving.

Management of NMS. When NMS is diagnosed or suspected, antipsychotics should be discontinued, and supportive and symptomatic treatment should begin. This may include treating EPS with antiparkinsonian medications, correcting fluid and electrolyte imbalances, treating fevers, and managing cardiovascular complications such as hyper- or hypotension. Dantrolene (Dantrium) may be effective for treating severe NMS. Treatment begins with IV administration of 0.8 to 2.5 mg/kg every 6 hours, with a maximum of 10 mg/kg daily. When the symptoms subside, and the patient can swallow, dantrolene is administered orally in dosages of 100 to 200 mg daily. Bromocriptine can then be added at doses of 20 to 30 mg daily in four divided doses. Amantadine may be helpful, if the other agents are not sufficient. The course of treatment is commonly 5 to 10 days unless a long-acting injectable agent has been used.

Although NMS is considered by some to be an extreme form of EPS, its cause and biologic risk factors are unknown. Most patients will require antipsychotic medications after they recover from NMS. Low-potency drugs may be preferable for patients who have previously had NMS. However, even clozapine has been demonstrated to be a potential cause of the syndrome. At this time, there are no data available to indicate whether clozapine or other serotonin–dopamine antagonists are associated with a lower risk of NMS than dopamine D2 antagonists.

CHRONIC EXTRAPYRAMIDAL SYNDROMES

Tardive Dyskinesia

TD is a movement disorder that can follow chronic treatment with antipsychotic medications. Patients with TD may have any or all of a number of abnormal movements, including mouth and tongue movements (lip smacking, sucking, and puckering), as well as facial grimacing. Other movements may include irregular movements of the limbs, particularly choreoathetoid-like movements of the fingers and toes, and slow, writhing movements of the trunk. Although seriously disabling TD is uncommon, a small proportion may affect walking, breathing, eating, and talking.

The differential diagnosis of TD includes a number of disorders of the basal ganglia (including Huntington disease, Wilson disease, and Sydenham chorea), striatal hypercalcifications, hyperthyroidism, hypoparathyroidism, tardive Tourette disorder, and dyskinesias related to other drugs, such as L-dopa and amphetamines. Abnormal movements were also described in patients with schizophrenia before the introduction of antipsychotic medications.

Although dopamine-blocking medications will frequently suppress or reduce TD, dose increases are only appropriate when the movements are severe and incapacitating. If patients have their dosages of antipsychotic medication either decreased or discontinued, abnormal movements may worsen temporarily or appear for the first time and then diminish, termed withdrawal emergent dyskinesias. Antiparkinsonian medications will often worsen TD. However, when patients suffer from combined TD and acute EPS, prescribing antiparkinsonian medication is often unavoidable.

Although the biologic mechanism underlying TD is still controversial, it is likely that the disorder is related to an increased sensitivity of dopamine receptors in the basal ganglia. This is supported by the observation that drugs that block dopamine, such as antipsychotic medications, suppress TD, whereas dopamine agonists worsen the disorder. These observations are consistent with findings from preclinical laboratories indicating that blocking dopamine receptors can result in their upward regulation.

Approximately 10% to 20% of patients treated with dopamine blockers for more than a year develop TD. In chronically institutionalized patients, the prevalence of TD is between 15% and 20% but increases with each year of treatment. Recent prospective studies indicate that approximately 4% to 5% of patients who are undergoing antipsychotic drug treatment, particularly with dopamine D2 receptor antagonists, will develop TD each year, with this risk markedly compounded in older adult individuals whose first treatment with antipsychotics is after age 65. Certain populations are at a greater risk than others for developing TD. Older adult women are particularly vulnerable. Moreover, TD in older adults is less likely to remit when dopamine blockers are discontinued. Patients with affective disorders may also be at a greater risk of TD when they are treated with dopamine receptor antagonists. Other possible risk factors for TD are the dosage of antipsychotic medications and the duration of treatment. There is also evidence indicating that the presence of acute EPS increases the risk of developing TD later.

TD does not appear to be a progressive disorder for most patients. The disorder seems to develop rather rapidly and then stabilize and often improves with time. The consensus of a number of long-term studies is that the severity of TD in most patients improves, even if their dopamine blockers are continued. Moreover, this improvement can be clinically meaningful in some patients. Individuals who are most likely to remit are those with recent onset and those with an onset prior to age 40. When dopamine blockers are discontinued, a substantial proportion of patients with TD will demonstrate remission. Reinstatement of a dopamine blocker does not mean an automatic reoccurrence. This is particularly likely for those with a rather recent onset of TD. Sometimes, this remission will take several months to occur.

Perioral Tremor

Perioral tremor is an uncommon side effect of dopamine receptor antagonists that appears after months or years of drug treatment. The peculiar tremors of the mouth have led to this disorder being called rabbit syndrome.

Management of TD. Patients receiving dopamine blockers for more than 6 months should have regular and systematic evaluations for TD every 6 months. Although instruments such as the Abnormal Involuntary

Movement Scale (AIMS) were originally developed as research tools, they also provide a structured system for evaluating and recording abnormal movements.

Once TD is diagnosed, the clinician should evaluate the need for continued antipsychotic treatment. This decision requires consideration of the dangers of both the psychotic illness and TD. It is a decision that, whenever possible, should be made in collaboration with the patient and their family. The results of this discussion should be documented in the patient's medical record. For most patients with chronic schizophrenia, the best decision will probably involve continuing the antipsychotic. If this is the case, there is some evidence that treating patients with the lowest effective dose may minimize the risk of TD. Moreover, prescribing the lowest effective dosage of antipsychotic medication may also decrease the patient's risk of developing the disorder. Patients with severe TD may also be candidates for a trial of clozapine or other dopamine–serotonin antagonists. Most clinical reports indicate that clozapine can suppress TD.

The American Psychiatric Association task force on TD recommended a number of steps for the prevention and management of the disorder, including (1) establishing objective evidence that antipsychotic medications are effective; (2) using the minimum effective dosage for long-term treatment; (3) exercising caution with children, the older adults, and patients with mood disorders; (4) examining patients on a regular basis for evidence of dyskinesia and noting the results of the examination in the medical record; (5) if TD is diagnosed, considering alternatives to antipsychotics, obtaining informed consent, and considering dosage reduction; and (6) if the dyskinesia worsens, considering the discontinuation of the antipsychotic, switching to a different drug, or considering a trial of clozapine.

In addition, there is substantial data indicating that clozapine and other dopamine–serotonin antagonists are less likely to cause the disorder than other antipsychotic medications, and, therefore, it is promising for the long-term management of patients who have developed TD. In addition, studies using animal models indicate that clozapine does not result in the increased sensitivity of striatal D2 receptors, which is associated with typical neuroleptics and has been theorized to result in TD.

Current evidence suggests that dopamine–serotonin antagonists are associated with a lower risk of TD than dopamine D2 receptor agonists. This has been proposed since the decreased tendency of these agents to cause acute EPS would predict a lower risk of TD. An accumulation of controlled and uncontrolled studies led a consensus group of experts in the pharmacotherapy of schizophrenia to conclude that dopamine–serotonin antagonists were associated with a lower risk of TD. Because most of these studies were not designed to focus on TD, this remains a controversial issue. A Cochrane review did not find sufficient evidence to conclude that there was an advantage for dopamine–serotonin antagonists, while other meta-analyses concluded that the available evidence supported an advantage. Still, despite the high penetration of dopamine–serotonin antagonists into clinical practice, the incidence and prevalence of TD appear relatively unchanged since the 1980s. Since it would be difficult to perform an adequate study of sufficient duration to prove that there is a difference, this issue remains unresolved.

There are currently two medications approved for the treatment of TD: deutetrabenazine and valbenazine. Both are VMAT2 inhibitors and can be administered in conjunction with antipsychotic treatments. Using these medications is suggested if TD symptoms cannot be controlled with dose reductions or medication changes.

BEHAVIORAL SIDE EFFECTS

Dopamine D2 antagonists may cause a number of subjective and behavioral effects that are not accompanied by obvious motor effects. Subtle parkinsonian side effects may lead to decreases in emotional expressiveness, apathy, and depression. These symptoms can easily be confused with the negative symptoms of schizophrenia. In addition, subjective akathisia may cause feelings of restlessness that can cause anxiety or even agitation. High dosages of dopamine D2 antagonists may also lead to cognitive impairment and attentional deficits. These behavioral symptoms can often be managed by decreasing the antipsychotic dosage.

CARDIOVASCULAR EFFECTS

Cardiac Side Effects

Animal studies have shown that antipsychotic agents may decrease cardiac contractility, disrupt enzyme activity in cardiac cells, decrease tissue levels and increase circulating levels of catecholamines, and prolong atrial and ventricular conduction time and refractory periods. Low-potency agents, particularly chlorpromazine and thioridazine, can induce electrocardiogram (ECG) abnormalities, such as prolongation of the QT and PR intervals, blunting of the T-waves, and depression of the ST segment. QT prolongation is a particular concern since this can lead to ventricular arrhythmia (known as torsades de pointes), which can progress to ventricular fibrillation and sudden death. There are indications that these arrhythmias have occurred in patients treated with thioridazine and chlorpromazine. A recent study monitored the QT interval in patients treated with a number of dopamine–serotonin antagonists, as well as haloperidol. The results led the FDA to conclude that the dopamine–serotonin antagonists thioridazine and mesoridazine (which is metabolized to thioridazine) were associated with substantial QTc prolongation and risk of torsades de pointes, with a subsequent high risk of sudden death. As a result, these agents are recommended only for patients who are refractory to other agents. Clinicians should consider obtaining an ECG before and during treatment with thioridazine and mesoridazine. These agents should not be prescribed for patients with prolonged QTc intervals or those who are receiving other agents that tend to prolong the QTc interval or that decrease the metabolism of such agents. Attempts are underway to identify patients at risk of QTc prolongation through pharmacogenomics. Recently the FDA and Johnson & Johnson have informed health care providers of a revision to the prescribing information for haloperidol, haloperidol decanoate, and haloperidol lactate to include a new cardiovascular subsection highlighting the risk of sudden death, QT prolongation, and torsades de pointes, particularly when given intravenously or at dosages higher than those recommended. Pimozide is also associated with QTc problems (and seizures) when given in doses of 20 mg or higher.

Older adult patients and those with dementia are at a greater risk of sudden cardiac death, presumably because of slower metabolism (higher blood levels) and concomitant polypharmacy for multiple aging-related illnesses and symptoms. This polypharmacy may lead to drug–drug interactions.

OTHER DOPAMINE RECEPTOR ANTAGONISTS EFFECTS

Orthostatic (Postural) Hypotension

Dopamine D2 blockers may cause orthostatic hypotension, particularly when these agents are first prescribed. This side effect can be dangerous when it results in fainting or falls. Both older adult patients and individuals with impaired balance should be monitored carefully for orthostatic changes. Since tolerance may develop gradually, patients should be initially treated with a relatively low dose of these agents (e.g., 25 mg orally twice daily), and the dose should be titrated upward over several days until a target dose is reached. When appropriate, patients should be warned of this side effect and told to sit or stand up slowly, to sit with their feet outside the bed when arising, to wait for as long as a minute, and to sit or lie down if they feel faint, particularly during the first 2 weeks of treatment. Some patients, particularly the older adults, may have persisting orthostatic hypotension and, as a result, may benefit from a change to a high-potency antipsychotic. Younger patients usually adapt to hypotension and may be asymptomatic, even though their blood pressure is low.

Orthostatic hypotension is more likely with IM administration. Thus, clinicians should consider measuring the patient’s blood pressure (both lying and standing) before and 1 to 2 hours after the first dose and before subsequent doses during the first days of treatment.

If hypotension does occur, it can usually be managed by having the patient lie down and instructing the individual to arise very slowly. Some patients may benefit from stockings, volume expansion, or pressor agents. The use of epinephrine paradoxically worsens antipsychotic-induced hypotension through β-stimulation and is, therefore, contraindicated. Pure α-adrenergic agents, such as metaraminol (Aramine), and norepinephrine, are the drugs of choice for this condition.

Sudden Unexplained Death

In spite of reports that dopamine receptor antagonists have been associated with sudden unexplained death, no evidence indicates that the incidence of such deaths changed in psychiatric patients after the introduction of antipsychotics. An American Psychiatric Association task force report concluded:

Although a relationship between the use of antipsychotic drugs and sudden death has not been firmly established, it has also not been disproved. From a neurocardiologic perspective, these drugs have the potential for both increasing and decreasing the risk of sudden death. The ultimate outcome is probably determined by a multitude of interacting factors, and the role played by the drug in a given individual is difficult, if not impossible to determine.

Reports of sudden death in patients treated with antipsychotics could be explained by cardiac arrhythmias such as torsades de pointes secondary to QTc prolongation or other causes. For example, seizures, asphyxiation, heat stroke, malignant hyperthermia, and NMS could result in relatively sudden death. Individuals who may be more vulnerable to sudden death include older adults, physically debilitated individuals, and agitated patients (particularly those who are physically restrained). Increasing evidence of QTc prolongation caused by antipsychotic agents is surfacing, presumably responsible for torsades de pointes.

Gastrointestinal Tract

The most common GI side effects of dopamine receptor antagonists, although uncommon, are related to their peripheral anticholinergic effects. These include dry mouth, constipation, occasional diarrhea, and urinary retention. Nausea and vomiting are less common. However, these effects can be exacerbated when other anticholinergic drugs such as tricyclic antidepressants are combined with dopamine blockers.

Patients who experience a severe dry mouth should be advised to rinse their mouth frequently and to chew sugarless gum or candy. Sugar-containing candies may lead to fungal infections in the mouth and increase the risk of dental caries. Constipation can be managed with stool softeners or laxatives. Clinicians should manage severe constipation carefully since it can progress to paralytic ileus, a potentially fatal condition.

Liver

Dopamine receptor antagonists may cause transient abnormalities in liver function tests. These are seldom of serious medical concern. Obstructive or cholestatic jaundice is a serious condition that was commonly associated with chlorpromazine treatment, a dopamine–serotonin antagonist, in the past. It is currently very uncommon.

Blood

Dopamine receptor antagonists may be associated with temporary and transient depression of leukopoiesis. Mild leukopenia may occur for prolonged periods, although it is more common for the leukocyte to eventually return to normal. Agranulocytosis is rare with dopamine receptor antagonists, occurring in less than 1 in 10,000 cases. Thrombocytopenia and pancytopenia may also occur.

Skin and Eye Effects

Cutaneous reactions seen during the first weeks of antipsychotic drug treatment include urticarial, maculopapular, petechial, and edematous reactions. These reactions usually improve when drugs are discontinued and sometimes clear even when drugs are continued. Although uncommon in patients receiving dopamine D2 blockers, chlorpromazine may develop photosensitivity reactions consisting of severe sunburn or rash. Patients should be instructed to avoid direct sunlight or to use sunscreens. Skin discoloration has also been observed, as well as granular deposits in the anterior lens and posterior cornea, especially after long-term treatment. Changing to a dopamine D2 receptor antagonist will usually lead to gradual improvement.

Endocrine Effects

Dopamine receptor antagonists elevate circulating levels of prolactin. In women, elevated prolactin can lead to breast enlargement, galactorrhea, and irregular menses, including anovulatory cycles and infertility, menses with abnormal luteal phases, or frank amenorrhea and hypoestrogenemia. The elevation of prolactin associated with dopamine blockers may also suppress testosterone in men. Prolactin elevation occurs as a direct result of D2 blockade in the tuberoinfundibular dopamine pathway and the pituitary. Since prolactin is under tonic inhibitory control by dopamine, blockade of dopamine receptors will elevate prolactin. Some patients, but not all, develop tolerance to the prolactin-elevating effects of these drugs after several weeks. Dosage reduction may relieve symptoms that are related to prolactin elevation.

Effects on Sexual Function

Dopamine blockers can have important effects on sexual functioning in both men and women. In men, these agents can result in decreased libido and ejaculatory and erectile disturbances. Erectile dysfunction has been reported to occur in 23% to 54% of men receiving an antipsychotic. This can be an important cause of medication noncompliance in men, and clinicians should inquire about it. Retrograde ejaculation has been reported. In addition, these agents can affect ejaculatory latency and the quantity of ejaculate. Since sexual functioning is affected by dopamine, serotonergic, and adrenergic systems, it is unclear how sexual dysfunction resulting from pure dopamine D2 antagonists is mediated. Dosage reduction may be helpful in some individuals.

Women who receive dopamine receptor antagonists may experience decreased libido, anorgasmia, and decreased lubrication. Although reports indicate that a number of adjunctive medications have been reported, there is little evidence that any are effective. As with men, reducing the dosage of the antipsychotic may be helpful. Presumably, agents with a high 5-HT_{2A} blockade should be able to counter these effects.

Pregnancy

Although dopamine receptor antagonists cross the placenta, there is very little evidence to suggest that there is an association between prenatal exposure and an increased incidence of congenital malformations. A meta-analysis by Altshuler and colleagues found that the baseline incidence of congenital anomalies is 2% and the incidence with antipsychotics is 2.4%. This meta-analysis found no evidence that dopamine D2-blocking drugs increase the risk to the fetus.

Prenatal exposure to dopamine receptor antagonists in rats may lead to a decreased number of brain dopamine receptors. Other animal studies indicate that prenatal haloperidol exposure may raise cholesterol levels in offspring. It is unclear if these animal findings indicate risks in humans.

In developing a treatment for pregnant patients with schizophrenia, the relatively low risk of congenital abnormalities associated with antipsychotics should be weighed against the dangers of psychotic episodes. Clinicians may attempt to discontinue these medications during the first trimester, if this is feasible and should consider carefully whether the risks of prescribing these drugs during the remainder of a patient's pregnancy are justified by the likely benefits. If patients have a history of relapse when dopamine blockers are discontinued, the drugs are relatively safe. If antipsychotics are prescribed, high-potency compounds are probably safer.

Lactation

Dopamine receptor antagonists are secreted in low concentrations with lactation, suggesting that patients should be advised against breastfeeding during antipsychotic treatment.

Seizure Threshold

Dopamine receptor antagonists tend to lower the seizure threshold. As a result, they are associated with an increased risk of seizures. This risk is usually clinically important in individuals with a prior history of seizures.

Temperature Regulation

Some patients become poikilothermic; that is, they become hypothermic in the winter and overheat in the summer with chlorpromazine and other low-potency agents. This indicates that special precautions should be taken for these patients during exposure to extremes in temperature.

Overdose

Overdoses with dopamine D2 receptor antagonists are usually characterized by exaggerations of their usual side effects. Thus, high-potency drugs are associated with severe extrapyramidal reactions, including dystonia and severe rigidity, as well as hypotension and sedation. The clinical picture may also include agitation, restlessness, convulsions, fever, autonomic reactions such as dry mouth and ileus, ECG changes, and cardiac arrhythmias. In severe overdoses with dopamine D2 blockers, the pupils are mydriatic, deep tendon reflexes are decreased, areflexia may occur, tachycardia and hypotension are present, and the EEG shows diffuse slowing and low voltage. This picture may progress to delirium and coma, with respiratory depression and hypotension.

The management of dopamine blocker overdose is symptomatic and supportive since overdoses with these agents are seldom lethal. However, the outcome may be unfavorable when these agents are combined with others, particularly CNS depressants such as alcohol, barbiturates, and benzodiazepines. If the overdose is recent, gastric lavage is helpful. If the patient can swallow, activated charcoal is indicated. Emetics should not be administered since their effects may be attenuated by the antipsychotic and because they are associated with a risk of aspiration pneumonia. Aspiration may be a serious risk, if the patient develops a dystonic reaction of the head or neck. Forced diuresis and hemodialysis are not effective because dopamine D2 blockers are highly protein bound and lipophilic.

When hypotension occurs, standard measures for managing circulatory shock should be implemented. The hypotension may occur when, during α-receptor blockade, β-receptors are unopposed. Therefore, the clinician should consider norepinephrine or dopamine. Epinephrine is contraindicated.

Tolerance, Dependence, and Withdrawal Symptoms

Some tolerance develops to the sedative, the anticholinergic, the hypotensive, and some other effects of dopamine blockers. Tolerance does not develop to the antipsychotic effects. Some patients experience a decrease in EPS over time, suggesting tolerance to the motor side effects of these agents. Dopamine receptor antagonists are not addictive. As a result, most patients can be withdrawn from these agents without difficulty, although manifestations of "cholinergic rebound," including nausea, vomiting, and diarrhea, can occur. Other reported withdrawal symptoms include headache, sweating, restlessness, and insomnia. These symptoms tend to occur within 2 to 3 days of drug withdrawal and may last as long as 2 weeks. However, most patients who are withdrawn from dopamine D2 antagonists do not show withdrawal symptoms.

DRUG INTERACTIONS

Clinicians should be aware of several types of interactions. These include drugs that interact with the metabolism of dopamine blockers; age and gender, which may affect metabolic pathways; medical conditions, which may affect drug-protein binding, absorption, and fecal or renal clearance; individuals who are dehydrated or malnourished or who have impaired liver and kidney function and are at risk of altered sensitivity and metabolism.

Metabolism

Dopamine D2 antagonists may have clinically important interactions with drugs that are metabolized by the same cytochrome P450 enzymes. Particularly for haloperidol, perphenazine, and others, this involves CYP 2D6. Since some SSRIs—particularly fluoxetine and paroxetine—are potent inhibitors of this enzyme, plasma levels of the dopamine antagonist are likely to be increased when an SSRI is added. This may result in an increase in a worsening of EPS. Other drugs that are substrates for 2D6, including heterocyclic antidepressants, β-blockers, and cimetidine (Tagamet), may also increase antipsychotic plasma concentrations.

Other drugs, particularly phenytoin (Dilantin), carbamazepine, and barbiturates, can reduce antipsychotic plasma concentrations by increasing the metabolism of the dopamine antagonist. Reports indicate that carbamazepine can reduce the plasma concentration of haloperidol by 50%. Cigarette smoking can also reduce the plasma levels of dopamine blockers.

There are other common interactions that should be of concern to clinicians. Antacids, activated charcoal, cimetidine, cholestyramine (Questran), kaolin, and pectin can decrease the absorption of antipsychotics from the gut. Increased absorption may occur with drugs that decrease GI motility, such as digoxin (Digitek) and steroids.

Coffee and tea may precipitate dopamine blockers. Although this has been demonstrated in vitro, its clinical significance is unclear.

Dopamine D2 blockers antagonize the effects of dopamine agonists or levodopa when these drugs are used to treat parkinsonism. Haloperidol can block the antihypertensive effects of guanethidine (Ismelin). Dopamine D2 blockers may also enhance the effects of CNS depressants such as analgesics, anxiolytics, and hypnotics.

Dopamine D2 blockers may enhance the effects of oral hypoglycemics. Phenothiazines enhance the effects of oral anticoagulants, although haloperidol may decrease these effects. Cigarette smoking may lower dopamine antagonist blood levels by enzyme induction.

Side-Effect Interactions

Narcotics, captopril (Capoten), trazodone (Desyrel), heterocyclic antidepressants, hydralazine (Apresoline), and minoxidil (Rogaine) may increase hypotension.

Amoxapine (Asendin), fluoxetine, chloroquine (Aralen), and lithium may increase extrapyramidal symptoms.

LABORATORY INTERFERENCES

Perphenazine may give false-positive or false-negative results on pregnancy tests and also give inaccurate reports of bilirubin (with reagent test strips) and urobilinogen (with Ehrlich reagent test). Phenothiazines may produce false-positive results in tests for phenylketonuria. Phenothiazines may also interfere with 17-ketosteroids (with the Haltorff-Koch modification of the Zimmermann reaction) and 17-OH corticosteroids (with the modified Glenn-Nelson technique).

TREATMENT OF ACUTE PSYCHOSIS

Initiation of Treatment

Treatment with a dopamine blocker is usually initiated soon after a diagnosis is made. Whenever possible, patients should receive a psychiatric evaluation, including a history and physical examination and a laboratory evaluation that includes complete blood cell count; blood electrolytes; glucose, liver, renal, and thyroid function tests; and other appropriate evaluations (including ECG, pregnancy tests, and screens for alcohol and drugs). The results of this evaluation can be useful in confirming the diagnosis and establishing a pretreatment baseline. Under most conditions, it is not necessary to delay treatment while awaiting the results of these tests. Moreover, because there are few contraindications to receiving FGAs, these evaluations can be postponed when patients refuse an evaluation or when treatment is urgent.

Choice of Agent

There is no evidence for differential effectiveness of dopamine D2 blockers, dopamine–serotonin blockers, or dopamine–serotonin–norepinephrine blockers, with the exception of clozapine, which is the most effective antipsychotic available. Amisulpride may also have a slightly lower EPS risk and does not appear to be as sedating as some of the low-potency SGAs. The choice among the dopamine D2 blockers is based on prior responses, both objective improvement and the patient's subjective response to the drug, available routes of administration and dosage forms, side-effect profiles, and cost.

The introduction of the dopamine–serotonin and dopamine–serotonin–norepinephrine antagonists (risperidone, olanzapine, quetiapine, ziprasidone, aripiprazole, iloperidone, lurasidone, zotepine, iloperidone, amisulpride, at this time) has complicated the decision-making process. These agents, in themselves quite heterogeneous, have different side-effect profiles when compared with dopamine blockers. The most important advantage of these “newer agents” is that they cause substantially less EPS. Patients who experience EPS at dosages of dopamine blocker needed to treat their psychosis may or may not experience EPS with an effective dosage of any of these other agents. This was evaluated in a phase of the CATIE trial in which patients who discontinued perphenazine were subsequently randomized to risperidone, olanzapine, or quetiapine.

There is also evidence from a number of well-designed clinical trials indicating that patients who have only partial responses to a dopamine blocker due to lack of efficacy or side effects may improve, if changed to a dopamine–serotonin and dopamine–serotonin–norepinephrine antagonist. The evidence is compelling for clozapine, but there are also some reports supportive of other agents as well. However, most of the studies reporting this advantage of newer drugs were carried out in the late 1980s and 1990s when patients had only been exposed to dopamine D2 blockers. More recently, there are clinical reports of patients who had only received dopamine–serotonin or dopamine–serotonin–norepinephrine antagonists who improved when they were changed to a dopamine D2 blocker. Moreover, cost considerations have become more important, especially in the United States. Many of the newer medications are patent protected and are considerably more expensive than dopamine blockers which are usually available as generics. Risks and benefits need to be considered carefully when choosing antipsychotics, especially in light of the different TD risks.

An additional advantage of some newer medications is that they are less likely to elevate prolactin. As a result, these agents may be helpful for women who experience galactorrhea and irregular menses with dopamine blockers.

Haloperidol and fluphenazine are still commonly prescribed. They are more strongly antidopaminergic and thus more likely to cause EPS than other available medications. However, EPS in some individuals may be easier to manage than the sedation and orthostatic hypotension associated with other medications. Haloperidol and Prolixin can also readily be administered intramuscularly and seldom cause problems with blood pressure, and oral doses can often be escalated over a few days without untoward sedation or hypotension.

Route of Administration

Dopamine D2 blocking drugs can be administered orally or as short- or long-acting injectables (Once patients are stabilized on a dosage, most patients can be managed with a single daily dose. Short-acting injectables are administered intramuscularly for calming acutely anxious patients or for administering a drug to patients who refuse oral medication. The clinical effects of an IM dose may appear as soon as 15 minutes after the injection. Since the dopamine D2 blockers are most effective when patients have reached a steady state for several days or even weeks, patients should be changed to oral administration as soon as possible.

Dopamine D2 blockers can also be administered as long-acting injectables. In the United States, haloperidol and fluphenazine are available in long-acting formulations. Long-acting antipsychotics are usually prescribed in stabilized patients during maintenance treatment after tolerance to the oral formulation has been established. It can take several months to reach a steady state, and oral supplementation may be necessary during that time, although the longer the oral medication treatment period and the better the adherence, the shorter time to steady state. For these reasons, long-acting depot dopamine D2-blocking drugs are clearly better suited than oral medications for long-term relapse prevention. They are cost-effective as well, but their TD risk may be higher.

Dosage

Patients demonstrate marked differences in the dosage of a dopamine receptor antagonist that is optimal. For example, some patients will find it difficult to tolerate the side effects of 1 mg of haloperidol, whereas others will tolerate 50 mg without suffering any side effects. Finding the best dosage for a dopamine D2 receptor antagonist is both important and difficult. The importance stems from the tendency of dopamine D2 blockers to cause side effects at their effective dosages. Oftentimes the clinician must weigh the therapeutic advantages of a particular dosage against discomforting or disabling side effects. As many of the dopamine blockers were developed prior to the required rigorous pharmacologic methodology of today, a minimally effective or lowest dose has usually not been established. As a result, the most reasonable practice may be to treat a patient with a dosage that is likely to be effective and to wait an adequate amount of time, 4 to 6 weeks at most, to determine if the patient is responding.

Dosage comparison studies provide guidance to clinicians who are making decisions regarding the dosages of dopamine blockers for most patients with schizophrenia or schizoaffective disorder. Doses below 2 mg of haloperidol or 5 mg of fluphenazine are likely to be too low for many psychotic patients. At the same time, doses above 10 mg of haloperidol or 20 mg of fluphenazine are seldom necessary and may lead to substantial side effects and even fewer antipsychotic effects. In other words, most patients will respond to doses between 5 and 20 mg of fluphenazine or 2 and 10 mg of haloperidol. Studies that focused on haloperidol suggest that doses below 8 mg are adequate for most patients and that these doses are associated with less EPS than 20 mg doses. First-episode patients should be treated with lower doses, and children and older adult patients should be treated with much lower doses.

The issue of dosing was recently highlighted by the FDA when it added a “Black Box” warning to the use of dopamine–serotonin and dopamine–serotonin–norepinephrine antagonists in older adult patients with dementia (increased mortality risk). A similar increased mortality risk has been associated with dopamine D2 blockers. Whether this is secondary to higher blood levels and the use of polypharmacy in older adults, exposing them to an increased risk of QTc prolongation, which may lead to torsades de pointes, arrhythmias, and subsequent death, has not been determined but seems likely. As with any condition that requires multiple medications, practitioners need to be aware of potential drug–drug interactions.

Some patients tolerate very high dosages of dopamine D2 blockers, leading some clinicians to raise the prescribed doses in the hope that “more is better.” This resulted in a substantial increase in the average dose of antipsychotics prescribed in the United States during the 1970s and 1980s. However, dosage comparison studies failed to support the routine use of higher doses. For example, when groups of patients were assigned to high doses such as 40 mg of haloperidol, the commensurate improvements were no greater than in those assigned to more moderate doses. Moreover, these higher doses were associated with more EPS. Current evidence does not support the use of excessive doses of antipsychotic medications.

As-Needed Medication

Prescribing medication as needed is recommended only for the initial phases of treatment. Although dopamine D2 blockers can result in nearly immediate calming for agitated individuals, this response appears to differ from the true antipsychotic effect of these medications, which has a delayed onset. Treating excitement and agitation with the addition of benzodiazepines such as lorazepam (Ativan) has the advantage of calming patients without causing EPS.

Blood Levels and Clinical Activity

Drug response is an interaction between dosage and time on drug and individual pharmacokinetic and dynamic variables. Finding the best dosage of an antipsychotic for an individual patient often requires weeks or months of adjusted treatment. The clinician usually selects a drug dosage that is effective for most patients with the same illness or a dosage that has been effective for the patient at an earlier time. Once a decision is made about the dosage and the plan implemented, the clinician is forced to wait days or weeks until it is clear whether the dosage is effective, toxic, or ineffective. If measuring a patient’s plasma or blood level enabled clinicians to select a drug dosage for an individual scientifically, this could decrease the amount of time it takes to arrive at an appropriate dosage.

Early studies of plasma levels indicated that patients who received the same dosage of a drug had large variations in their plasma levels. These findings suggested that studying individual differences in plasma levels would provide information that could be useful to clinicians. However, these early studies were limited to drugs such as chlorpromazine that have complex metabolisms. This creates a problem since some of the antipsychotic activity of these agents could result from metabolites rather than the parent drug. This and other problems, including flawed research methods, may explain why early studies failed to find a reliable relationship.

A number of more recent studies have focused on other dopamine receptor antagonists and have shown more promising results. Haloperidol has received the most attention for the study of relationships between plasma concentration and clinical response. An advantage of haloperidol for these studies is that its metabolism includes only a single important metabolite, reduced haloperidol, which does not have significant antipsychotic activity. As a result, measuring the plasma level of haloperidol may provide an accurate representation of the effective amount of the drug. At least five research groups, but not all, have reported a therapeutic window relationship between haloperidol levels and clinical response. Although all of these ranges are supported by one or more high-quality studies, the strongest evidence is for haloperidol.

Studies do not support the routine use of plasma level measurement with the exception of clozapine. Still, a level may be helpful in the patient who is receiving what appears to be an adequate drug dosage but who is not responding. A low level may suggest that the patient is not complying or that poor absorption or extensive drug metabolism is interfering with an adequate level. In other cases, the clinician may find it difficult to interpret anxiety or agitation and may consider whether psychotic agitation could come from an inadequate dosage or whether akathisia may result from an excessive dosage. A high or low plasma level could provide valuable information in these circumstances.

Plasma levels should only be obtained after steady-state levels have been reached, preferably during the late phase of the plasma level curve. This usually means after 5 days on a stable oral dose and 10 to 12 hours after the last dose.

Adjunctive Medications

Antipsychotic medications are frequently supplemented by other psychotropic medications. These are aimed at managing comorbid conditions that frequently occur with schizophrenia. There is substantial evidence indicating that depression is underdiagnosed and undertreated, while antidepressants may be effective for treating depressive symptoms in patients with schizophrenia and schizoaffective disorder. Mood-stabilizing drugs, including lithium, carbamazepine, and valproate (Depakote), are often used for patients with unstable moods. However, carbamazepine may induce the metabolism of antipsychotic agents, lowering plasma concentrations of the parent drug. Benzodiazepines are often effective for treating anxiety, as well as agitation and excitement. Newer sleeping agents may assist patients to sleep better with lower risks than older medications.

MAINTENANCE TREATMENT

Duration of Treatment

During the maintenance phase of treatment, the treatment goals include preventing psychotic relapse and supporting treatments aimed at improving the patient’s level of functioning and quality of life. Although dopamine D2 blockers are effective in delaying or preventing relapse, they are also associated with short- and long-term side effects. For most patients with schizophrenia or schizoaffective illness, the analysis of costs and benefits will favor continuing the patient on an antipsychotic, including depot antipsychotics. Patients who have recovered from a single psychotic episode have a substantial risk of relapse, if they are withdrawn from medications. As a result, first-episode patients should receive at least 1 year of maintenance therapy, if asymptomatic. This is often a substantial point of contention with parents and their families, but the evidence is very clear that relapse will occur in 95% of patients eventually. Patients who have had multiple episodes should continue to receive antipsychotic maintenance for at least 2 to 5 years, potentially indefinitely. Present data indicate that there is a 10% relapse rate per year, often precipitated by nonadherence with medication, suggesting that continuous, moderate dosage maintenance dopamine D2 blockers afford the best chance of avoiding relapse in psychotic patients. Patients with a history of violent, aggressive behavior or serious suicide attempts should receive antipsychotics indefinitely. If a decision is made to discontinue medications, the dosage of medication should gradually be reduced, and the patient should be carefully observed. However, as with many chronic medical diseases, patients do discontinue or take medications irregularly on their own.

Long-Acting Injectable Antipsychotics

There is substantial evidence indicating that long-acting injectable antipsychotics have advantages for patients, including first-episode patients during the maintenance stage of treatment. Depot antipsychotics should be considered for each patient, starting early in the course of illness, because they have been shown to reduce relapse more than oral antipsychotics, potentially improving long-term prognosis, as well as preventing both psychiatric readmissions and arrests. One advantage is that these agents are a partial solution to the problem of poor medication adherence in schizophrenia. Clinicians can administer a drug without depending on patients to take their pills. In addition, because depot drugs bypass variations in drug absorption and first-pass hepatic and gut metabolism, drug levels may be more consistent. A number of studies have compared the relative effectiveness of long-acting and oral drugs for preventing relapse in stabilized patients. Studies that closely resemble treatment in the community have found that patients are better protected by depot medications. A number of carefully controlled studies and meta-analyses have compared the risk of psychotic relapse in patients who were randomly assigned to receive either oral or depot medication, with depot antipsychotics reducing relapse rates to a greater extent than oral antipsychotics.

Treatment Strategies During Maintenance

Concerns about the side effects of antipsychotics and the risk of TD have led to a number of proposed strategies for safely managing patients on the lowest possible amount of a dopamine antagonist. A number of studies have found that patients can be safely managed with substantially lower dosages of a long-acting depot antipsychotic than are commonly prescribed. These studies found that most patients did well when their doses were reduced by 80% from doses that are usually prescribed for acute treatment. Patients treated with lower dosages experienced a small increase in the risk of relapse that was balanced in some studies by reductions in side effects, improved compliance, and suggestions of improved community adjustment.

Other studies have focused on a strategy called *targeted* or *intermittent therapy,* in which an antipsychotic is gradually reduced and eventually discontinued. Patients are monitored carefully for early or prodromal signs of relapse. The antipsychotic is reintroduced when these signs appear. A number of studies, including a NIMH collaborative study, have found that targeted treatment is associated with high relapse rates, indicating that this strategy is not advisable.

FURTHER READINGS

Agid O, Seeman P, Kapur S. The “delayed onset” of antipsychotic action–an idea whose time has come and gone. *J Psychiatry Neurosci*. 2006;31(2):93–100.

Buchanan RW, Kreyenbuhl J, Kelly DL, et al. The 2009 Schizophrenia PORT psychopharmacological treatment recommendations and summary statements. *Schizophr Bull*. 2010;36:71–93.

Correll CU, Kane JM. Ranking antipsychotics for efficacy and safety in schizophrenia. *JAMA Psychiatry*. 2020;77(3):225–226.

Farde L, Nordstrom AL, Wiesel FA, Pauli S, Halldin C, Sedvall G. Positron emission tomographic analysis of central D1 and D2 dopamine receptor occupancy in patients treated with classical neuroleptics and clozapine. Relation to extrapyramidal side effects. *Arch Gen Psychiatry*. 1992;49:538–544.

Gilbert PL, Harris MJ, McAdams LA, Jeste DV. Neuroleptic withdrawal in schizophrenic patients. A review of the literature. *Arch Gen Psychiatry*. 1995;52:173–188.

Hartling L, Abou-Setta AM, Dursun S, Mousavi SS, Pasichnyk D, Newton AS. Antipsychotics in adults with schizophrenia: comparative effectiveness of first-generation versus second-generation medications: a systematic review and meta-analysis. *Ann Intern Med*. 2012;157(7):498–511.

Jones PB, Barnes TRE, Davies L, et al. Randomized controlled trial of the effect on quality of life of second- vs first-generation antipsychotic drugs in schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1). *Arch Gen Psychiatry*. 2006;63:1079–1087.

Leucht S, Chaimani A, Krause M, et al. The response of subgroups of patients with schizophrenia to different antipsychotic drugs: a systematic review and meta-analysis. *Lancet Psychiatry*. 2022;9(11):884-893.

Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet*. 2013;382(9896):951–962.

Leucht S, Corves C, Arbter D, Engel RR, Li C, Davis JM. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. *Lancet*. 2009;373:31–41.

Leucht S, Pitschel-Walz G, Abraham D, Kissling W. Efficacy and extrapyramidal side-effects of the new antipsychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo. A meta-analysis of randomized controlled trials. *Schizophr Res*. 1999;35(1):51–68.

Lieberman JA, Stroup TS, McEvoy JP, et al. Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) investigators. *N Engl J Med*. 2005;353:1209–1223.

McEvoy JP, Byerly M, Hamer RM, et al. Effectiveness of paliperidone palmitate vs haloperidol decanoate for maintenance treatment of schizophrenia: a randomized clinical trial. *JAMA*. 2014;311(19):1978–1987.

Poyurovsky M. Acute antipsychotic-induced akathisia revisited. *Br J Psychiatry*. 2010;196(2):89–91.

Schooler NR, Keith SJ, Severe JB, et al. Relapse and rehospitalization during maintenance treatment of schizophrenia. The effects of dose reduction and family treatment. *Arch Gen Psychiatry*. 1997;54:453–463.

Sikich L, Frazer JA, McClellan J, et al. Double-blind comparison of first- and second generation antipsychotics in early-onset schizophrenia and schizoaffective disorder: findings from the treatment of early-onset schizophrenia spectrum disorders (TEOSS) study. *Am J Psychiatry*. 2008;165(11):1420–1431.

Tiihonen J, Haukka J, Taylor M, Haddad PM, Patel MX, Korhonen P. A nationwide cohort study of oral and depot antipsychotics after first hospitalization for schizophrenia. *Am J Psychiatry*. 2011;168:603–609.

Woods SW, Morgenstern H, Saksa JR, et al. Incidence of tardive dyskinesia with atypical versus conventional antipsychotic medications: a prospective cohort study. *J Clin Psychiatry*. 2010;71(4):463–474.

33.2b Dopamine-Serotonin Antagonists

STEPHEN R. MARDER, M.D., AND WALTER DUNN, M.D., PH.D.

Neuroscience-Based Nomenclature (NbN)	
Olanzapine, lurasidone, ziprasidone	
Pharmacologic Domain	Dopamine, serotonin
Mechanism of action	Receptor antagonists (D_2, 5-HT_2)
Clozapine, risperidone, paliperidone	
Pharmacologic Domain	Dopamine, serotonin, norepinephrine
Mechanism of action	Receptor antagonists (D_2, 5-HT_2, α2)
Quetiapine	
Pharmacologic Domain	Dopamine, serotonin, norepinephrine
Mechanism of action	Receptor antagonists (D_2, 5-HT_2), reuptake inhibitor (NET)
Iloperidone	
Pharmacologic Domain	Serotonin, dopamine
Mechanism of action	Receptor antagonists (5-HT_2, D_2)
Asenapine	
Pharmacologic Domain	Serotonin, dopamine, norepinephrine
Mechanism of action	Receptor antagonists (5-HT_2, D_2, α2)
Lumateperone	
Pharmacologic Domain	Serotonin, dopamine
Mechanism of action	Receptor antagonist (5-HT_2, D_2), reuptake inhibitor (SERT)

Dopamine-serotonin antagonists (DSAs) are the most widely prescribed agents for schizophrenia and other illnesses associated with psychotic symptoms.

This diverse group of compounds can be differentiated from the dopamine receptor antagonists based on their high affinity for serotonin 2A (5-HT_{2A}) as well as dopamine (D_2) receptors. In prior editions of this text, the term second-generation antipsychotics or SGAs was used to refer to these agents. However, this term is inaccurate since decades of studies have failed to support the contention that antipsychotics marketed after clozapine were more effective antipsychotics than older drugs. The adoption of neuroscience-based nomenclature (NbN) provides a preferred method for characterizing antipsychotic medications based on their pharmacology. It is notable that antipsychotics previously considered first-generation antipsychotics—including chlorpromazine, thioridazine, and flupenthixol—are categorized as DSAs, and sulpiride and amisulpride are classified as dopamine antagonists (DAs).

The differentiation of the DSAs from other antipsychotics emerged following the introduction of clozapine (Clozaril), an antipsychotic that was introduced in the United States in 1990, although it had been available in some countries in Europe since the early 1970s. Clozapine was considered an “atypical” antipsychotic because it was an effective agent at doses that caused negligible extrapyramidal symptoms (EPS). Moreover, it had broad effectiveness for a range of mood and anxiety symptoms that were common in schizophrenia. This led to a focus on understanding the mechanism of action of clozapine. Clozapine’s D_2 receptor blockade is lower than that of traditional antipsychotics, while its antagonism of the 5-HT_2 receptor is greater than that of haloperidol. The unique activity at these two receptors led Herbert Meltzer to propose that the distinctive antipsychotic effect of clozapine was related to its low receptor blockade at D_2 relative to 5-HT_2. Clozapine’s affinity for the 5-HT_2 receptor is among the highest for antipsychotic agents.

During the early 1980s, scientists at Janssen Pharmaceutica were interested in developing an antipsychotic that would have a clinical profile similar to clozapine but without a tendency to cause agranulocytosis. The company developed a number of drugs with 5-HT_2 antagonism with and without D_2 activity. One compound, setoperone, had weak dopamine activity and was a relatively weak antipsychotic. Ritanserin was developed as a 5-HT_2 antagonist (inverse agonist) that could be added to a D_2 antagonist to decrease the amount of extrapyramidal side effects. Although ritanserin had antinegative symptom effects, improved sleep, and some weak antipsychotic effects, it also had significant effects on QTc interval. The synthesis of risperidone in 1984 led the company to focus on the development of this agent since it combined strong D_2 and 5-HT_2 antagonism. Subsequent safety and efficacy studies with risperidone led to its approval in the United States in 1994. This was followed by the introduction of olanzapine in 1996 and quetiapine in 1997, and others during the subsequent years. This chapter will discuss the DSAs using the following NbN categories:

1. Dopamine, serotonin (chlorpromazine, flupenthixol, loxapine, lurasidone, olanzapine, perospirone, pipotiazine, sertindole, thioridazine, trifluoperazine, zotepine)
2. Dopamine, serotonin NE α2 (clozapine, risperidone, paliperidone)
3. Dopamine, serotonin, NET (quetiapine)
4. Serotonin, dopamine (iloperidone)
5. Serotonin, dopamine, norepinephrine (asenapine)
6. Serotonin, dopamine, serotonin transporter (lumateperone)

The partial agonists, aripiprazole, brexpiprazole, and cariprazine are discussed in the next chapter.

Comparisons of DSAs

A network meta-analysis of randomized controlled trials compared the effectiveness of 32 antipsychotics. The study found that clozapine, amisulpride, zotepine, olanzapine, and risperidone were more effective for improving overall symptoms than other antipsychotics. However, with the exception of clozapine, the differences between the drugs were relatively small, and it was not clear that the DSAs were more effective than the DAs. The differences among the antipsychotics in side effects were substantial. A reasonable conclusion from this and other meta-analyses is that—aside from clozapine—the selection of an antipsychotic will be based on side-effect concerns rather than efficacy.

DOPAMINE, SEROTONIN

This group of antipsychotics includes a number of older antipsychotics such as chlorpromazine, flupenthixol, thioridazine, trifluoperazine, zotepine, loxapine, perospirone, pipotiazine, as well as newer drugs, including lumateperone, lurasidone, olanzapine, and sertindole,) This section will focus on the most commonly prescribed drugs from this group, including chlorpromazine, olanzapine, ziprasidone, and lurasidone.

The discovery of chlorpromazine's effectiveness for psychosis represents the beginning of modern psychopharmacology in that it indicated that drugs could have specific effects on the symptoms of mental disorders that went beyond sedation. In 1950, Paul Charpentier, at the Rhóne-Poulenc Laboratories in Paris, synthesized chlorpromazine, a drug that appeared notable as a mild antihistaminic that was also sedating. Henri Laborit, a surgeon with an interest in drugs that would decrease preoperative anxiety and prevent postsurgical shock, convinced the Rhóne-Poulenc Laboratories of the importance of identifying compounds that would relax patients and make surgical shock less likely. Chlorpromazine appeared to fulfill these criteria, but it also appeared qualitatively different from other sedating drugs. Patients appeared indifferent toward the environment and more tranquil. These properties led Laborit to convince psychiatrists, including Jean Delay and Pierre Deniker, to administer chlorpromazine to patients with mania and schizophrenia. In 1952, these authors reported their observation that chlorpromazine was effective in treating mania and schizophrenia. Within a year, the atmosphere in Paris's psychiatric hospitals had improved substantially. Moreover, the availability of drugs with direct effects on mental disorders became a powerful tool for laboratory studies with animals, leading to a greater understanding of how drugs affect the brain and information about the nature of the illnesses that were affected. Although reserpine was also an effective antipsychotic, its side effects led clinicians to prefer chlorpromazine, and, as a result, it is no longer used for treating psychosis because of its side effects, such as depression, suicidality, dysphoria, and orthostatic hypotension.

Other DSA antipsychotics—including trifluoperazine, flupenthixol, thioridazine, and loxapine—were introduced prior to 1975. Along with chlorpromazine, these antipsychotics are not commonly prescribed—with the possible exception of the long-acting formulation of flupenthixol which is not available in the United States. This is partly due to their side-effect profiles. Both chlorpromazine and thioridazine can induce electrocardiogram (ECG) abnormalities, such as prolongation of the QT and PR intervals, blunting of the T waves, and depression of the ST segment. QT prolongation is a particular concern since this can lead to ventricular arrhythmia (known as torsades de pointes), which can progress to ventricular fibrillation and sudden death. Thioridazine (Mellaril) has the greatest risk of QTc prolongation among all currently available antipsychotics and is rarely prescribed.

Chlorpromazine is also associated with an uncommon discoloration of the skin. Skin areas that are exposed to sunlight, particularly the face and neck, develop blue–gray metallic discoloration. This skin reaction is usually associated with long-term treatment with high drug dosages. Chlorpromazine-treated patients may develop granular deposits in the anterior lens and posterior cornea. These deposits, visualized on slit lamp examination, seldom affect the patient’s vision. Changing the patient to another drug will usually result in a gradual improvement in the condition. High doses of thioridazine—above 1,000 mg daily—can result in altered retinal pigmentation (retinitis pigmentosa). This condition can lead to serious visual impairment or blindness. Moreover, this condition may not remit when thioridazine is discontinued. As a result, thioridazine should not be prescribed at doses over 800 mg daily.

Sertindole was approved and marketed in a number of E.U. countries in 1996. However, concerns about QTc prolongation and the risk of cardiac arrhythmias led to its suspension in 1998. This concern also prevented its approval by the U.S. FDA. Currently, sertindole is available in several E.U. countries with cautions regarding patient selection and recommendations for ECG monitoring.

Olanzapine

Eli Lilly and Company formulated olanzapine in Great Britain in 1982. Since olanzapine is similar in structure to clozapine, its early development was encouraged by the hope that it would have similar efficacy with milder side effects. Phase III clinical trials were completed in July 1995, and olanzapine became available in the U.S. and European markets in the fall of 1996.

Olanzapine reached the market in the United States in 1996. It has a mildly sedating, calming effect that has led to its acceptance in many acute psychiatric settings. Olanzapine has shown marginal efficacy advantages when compared with some other antipsychotics in a number of meta-analyzes. However, it is also associated with serious side effects, including weight gain, lipid elevation, and an increased risk of diabetes. This has led many clinicians to consider this a second-line antipsychotic.

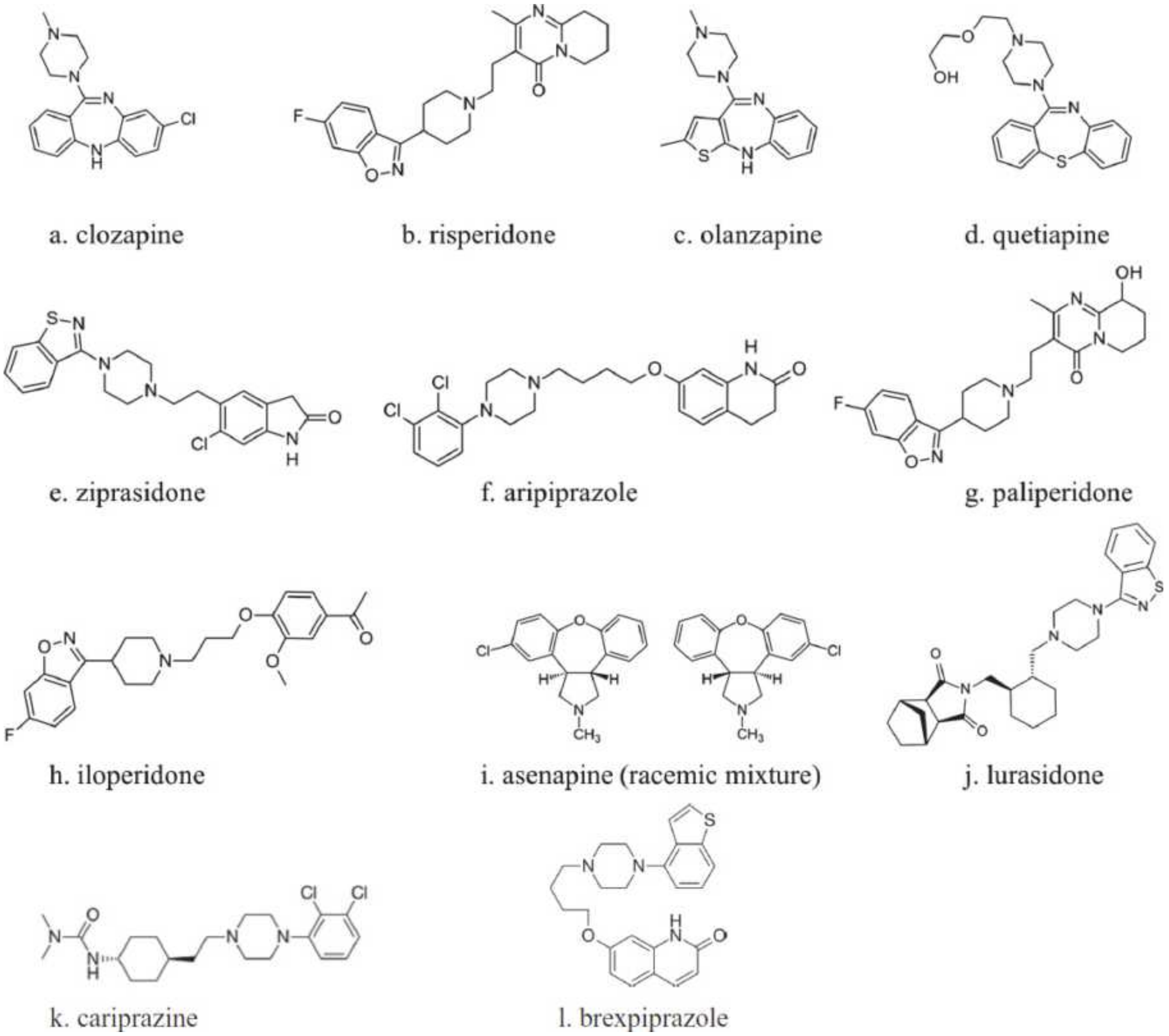

a. clozapine b. risperidone c. olanzapine d. quetiapine

e. ziprasidone f. aripiprazole g. paliperidone

h. iloperidone i. asenapine (racemic mixture) j. lurasidone

k. cariprazine l. brexpiprazole

FIGURE 33.2b–1. Molecular structures of SGAs discussed in the chapter. (Panels a–l by Fuse809, via Wikimedia Commons. Licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. Panel j in public domain.)

Chemistry. Olanzapine is a thienobenzodiazepine (Fig. 33.2b–1). It can be considered a derivative of clozapine with the substitution of a thieno ring for clozapine's carbonyl ring.

Pharmacologic Action

PHARMACOKINETICS

Absorption. Olanzapine is well absorbed when administered orally. Peak plasma concentrations are reached in about 6 hours. Food has relatively little influence on its absorption. Olanzapine's half-life is influenced by smoking, gender, and age, with a mean of about 31 hours (range of 21 to 54 hours). This allows for once-daily dosing. Patients reach steady plasma concentrations in about 7 days. IM administration of olanzapine leads to peak plasma concentrations within 15 to 45 minutes.

Distribution. Olanzapine is widely distributed throughout the body. The drug is 93% protein bound to plasma proteins.

Metabolism and Elimination. The primary metabolic pathways for olanzapine are direct glucuronidation and P450-medicated oxidation. None of olanzapine's metabolites are active. Although in vitro studies indicate that CYP1A2 and 2D6 and the flavin-containing mono-oxygenase system are involved in its oxidation, CYP1A2 appears to be the major pathway.

PHARMACODYNAMICS

Mechanism of Action. Olanzapine is a high-affinity antagonist at $5\text{-HT}_{2A/2C}$, 5-HT_6, $D_{1–4}$, H_1, and adrenergic α_1-receptors and a moderate affinity antagonist at $M_{1–5}$ and 5-HT_3 receptors. PET studies indicate that typical doses of olanzapine such as 10 to 20 mg result in 68% to 84% D_2 occupancy. Its 5-HT_2 activity is approximately eight times as strong as its dopamine receptor blockade. Compared with risperidone, quetiapine, and ziprasidone, olanzapine has greater muscarinic and histaminergic receptor antagonism. This histamine activity may explain its tendency to cause weight gain.

Blood Concentrations and Clinical Activity. Consensus guidelines from 2020 strongly recommend therapeutic-level monitoring for patients receiving olanzapine. The recommended therapeutic range is 20 to 80 ng/mL.

Design and Interpretation of Clinical Drug Studies

ACUTE TREATMENT OF SCHIZOPHRENIA. Double-blind studies have compared once-daily olanzapine at dosages ranging from 2.5 to 17.5 mg per day with placebo and haloperidol (10 to 20 mg per day). In one study, olanzapine was found to be superior to placebo and comparable to haloperidol in the treatment of overall psychopathology (mean Clinical Global Impression [CGI]-S score) and positive psychotic symptoms (positive Brief Psychiatric Rating Scale [BPRS] and Positive and Negative Syndrome Scale [PANSS] scores) at intermediate (7.5 to 12.5 mg per day) and high (12.5 to 17.5 mg per day) dosages. Olanzapine was also found to be superior to haloperidol in negative psychotic symptom response.

In another 6-week, fixed-dose, double-blind study involving 1,996 patients in North America and Europe, olanzapine showed significantly greater clinical improvement over haloperidol (total BPRS score, $p < .02$). Significant improvement was also seen in both positive symptom and negative symptom scores ($p = .03$) as well as depression rating ($p = .001$).

Double-blind studies have also evaluated the effect of a rapid-acting IM formulation of olanzapine in the treatment of acute agitation in schizophrenia. IM olanzapine was found to be effective against positive symptoms in acutely agitated patients, superior to placebo and similar to IM haloperidol.

OLANZAPINE IN CHILDREN AND ADOLESCENTS. There are relatively few data evaluating the effectiveness of olanzapine for children or adolescents with schizophrenia or bipolar illness. A study in adolescents reported by Kryzhanovskaya and coworkers found that olanzapine was more effective than placebo but associated with somnolence, liver enzyme elevations, prolactin elevation, and weight gain. An 8-week study that compared olanzapine, risperidone, and haloperidol in 50 patients aged 8 to 19 years with psychosis found that the three conditions were equally effective. However, weight gain and EPS on olanzapine and risperidone appeared to be greater than had been reported in adults. A trial by Linmarie Sikich and coworkers compared risperidone and olanzapine to molindone (Moban) in patients between 8 and 19 years of age. The National Institute of Mental Health (NIMH) Data and Safety Monitoring Board closed randomization to the olanzapine arm due to weight gain. These concerns are mirrored in a study of bipolar adolescents (aged 13 to 17 years) who were randomized in a 3-week comparison of olanzapine or placebo. Although olanzapine was effective in reducing symptoms, it was also associated with increases in weight, liver enzymes, fasting cholesterol, prolactin, and fasting glucose.

Taken together, all of these studies support very serious safety concerns related to treatment with olanzapine in children and adolescents. In the opinion of the authors, olanzapine should be considered for this population only when other antipsychotics have failed.

OLANZAPINE IN FIRST-EPISODE SCHIZOPHRENIA. A study of 263 patients with first-episode schizophrenia reported by Jeffrey Lieberman found that olanzapine had advantages over haloperidol in total PANSS scores and depression. An analysis of data over 2 years found that fewer patients discontinued olanzapine. However, olanzapine was associated with greater weight gain and elevations in liver enzymes. A 1-year double-blind study reported by Joseph McEvoy randomized 400 recent-onset patients to olanzapine, quetiapine, or risperidone. Discontinuation rates were similar on the three drugs. However, there were advantages for olanzapine on positive symptoms. As in other studies, weight gain was highest with olanzapine.

EFFICACY FOR MANIA AND MIXED STATES IN BIPOLAR DISORDER. Two studies found that olanzapine monotherapy at doses between 5 and 20 mg was superior to placebo at reducing manic symptoms as measured by the Young Mania Rating Scale (YMRS). A third study failed to find this advantage for olanzapine. These studies support the use of olanzapine in bipolar patients with mania or in mixed states. A study by Perlis and coworkers compared olanzapine and risperidone and found that the two SGAs had similar effectiveness in reducing manic symptoms. Two trials evaluated the effectiveness of adding olanzapine or placebo to either lithium or valproate when patients had symptoms that were inadequately controlled on the first drug. In both trials, the addition of olanzapine was more effective than the addition of a placebo in reducing manic symptoms.

AGITATION IN SCHIZOPHRENIA AND BIPOLAR DISORDER. Studies of IM olanzapine focused on the short-term reduction in agitation in patients with schizophrenia or bipolar disorder. They were double blind and consisted of patients being assigned to receive up to three injections in the first 24 hours.

Two studies compared IM olanzapine to placebo in patients with schizophrenia who demonstrated symptoms of excitement. Olanzapine at doses of 2.5, 5, 7.5, and 10 mg was more effective than placebo at decreasing levels of excitement. The higher olanzapine doses resulted in greater reductions in excitement. The advantage of olanzapine was apparent for all doses at 2 hours and for 5 to 10 mg at 30 minutes. These improvements were similar to those with IM haloperidol or IM lorazepam in separate studies. Another study found that 10 mg of IM olanzapine was more effective than placebo and oral olanzapine for reducing excitement in bipolar patients with mania or mixed states.

Effects on Specific Organs and Systems. Olanzapine is associated with orthostatic hypotension, tachycardia, and possible syncope. This probably results from its α_1-adrenergic antagonistic properties. IM olanzapine has been associated with a syndrome that includes hypotension with bradycardia.

The anticholinergic effects of olanzapine can lead to constipation in some patients.

Therapeutic Indications

SCHIZOPHRENIA. Olanzapine is an effective agent for treating acute psychotic symptoms in schizophrenia. Since it is mildly sedating, psychiatrists in acute settings often favor olanzapine for patients who will benefit from a calming agent.

Olanzapine is also effective for maintaining an antipsychotic response in stabilized patients. In large studies such as Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE), olanzapine patients tended to have lower discontinuation rates when compared with other first- and second-generation drugs. However, weight gain, lipid elevations, and insulin resistance can be serious problems in some patients. As a result, patients who continue in long-term treatment should be carefully monitored for these metabolic effects.

BIPOLAR DISORDER. Olanzapine is effective for the acute treatment of mania and mixed states, both as monotherapy and as adjunctive therapy to lithium or valproate. It is notable that olanzapine is effective for both psychotic and nonpsychotic patients.

Olanzapine is also effective as monotherapy for maintaining response in patients who have been treated for mania or mixed states. As with schizophrenia, patients receiving long-term treatment with olanzapine should be carefully monitored for its adverse metabolic effects.

Bipolar I Disorder with Acute Depression and Treatment-Resistant Major Depressive Disorder

Olanzapine combined with fluoxetine is indicated for both of these disorders.

AGITATION ASSOCIATED WITH SCHIZOPHRENIA AND BIPOLAR I MANIA. IM olanzapine is indicated for agitation that can be associated with schizophrenia or mania. The olanzapine product information defines psychomotor agitation as “excessive motor activity associated with a feeling of inner tension.” It should be noted that oral olanzapine in its different formulations, as well as other antipsychotics, are also effective for agitation. The important advantage of the IM route is a more rapid onset of action.

Precautions and Adverse Reactions

BLACK BOX WARNING

Increased Mortality in Older Adult Patients with Dementia. Olanzapine and the other SGAs carry a black box warning for increased mortality in older adult patients with dementia-related psychosis compared with placebo.

WEIGHT GAIN. A large number of studies indicate that olanzapine is associated with greater weight gain than other DSAs, with the possible exception of clozapine. In the NIMH CATIE study, olanzapine patients gained approximately 2 lb per week. Studies during long-term therapy with olanzapine found that 56% of patients gained greater than 7% of baseline weight. Weight gain occurred across all BMI levels indicating that low, normal, overweight, and obese patients were vulnerable to weight gain. As mentioned earlier, children and adolescents are more vulnerable to weight increase with olanzapine treatment.

HYPERLIPIDEMIA. Patients treated with olanzapine may experience serious increases in total cholesterol, low-density lipoprotein (LDL) cholesterol, and triglycerides. Triglycerides are most likely to be elevated, sometimes to levels above 500 mg/dL. In the NIMH CATIE study, the mean increase in triglycerides in patients taking olanzapine was 40.5 mg/dL. For comparison, quetiapine was associated with an increase of 21.2, risperidone with a decrease of 2.4, perphenazine with an increase of 9.2, and ziprasidone with a decrease of 16.5. Since some patients were receiving olanzapine at the start of the trial and may have already experience increases, this may be an underestimate of triglyceride elevation on olanzapine. Patients with no prior history of hyperlipidemia were particularly vulnerable to increases during olanzapine treatment. These concerns about hyperlipidemia suggest that patients who are treated with olanzapine should have lipid levels monitored before and after olanzapine treatment is initiated and at regular intervals.

HYPERGLYCEMIA, INSULIN RESISTANCE, DIABETES MELLITUS. Several studies suggest that hyperglycemia and diabetes mellitus are potential side effects that follow treatment with olanzapine. A number of studies, including CATIE, found greater elevations in blood glucose on olanzapine than on other antipsychotics. These effects were apparent in individuals with normal glucose regulation, as indicated by normal fasting blood glucose, as well as in individuals with diabetes mellitus or other evidence of glucose dysregulation. Other studies have found that olanzapine is associated with an increase in insulin resistance in both obese and nonobese patients. This suggests that the effects of olanzapine are not just secondary to weight gain.

There is also substantial evidence indicating that olanzapine treatment is associated with an increased risk of the development of type 2 diabetes mellitus. Using the U.K. General Practice Research Database, Koro and coworkers found that olanzapine was associated with a significantly increased risk of diabetes when compared with patients who were not on antipsychotics (odds ratio 5.8) or taking FGAs. This finding has been supported by other large epidemiologic studies. The increased risk could be related to weight gain, which is highly associated with the development of diabetes. However, there are descriptions of normal-weight patients on olanzapine who develop diabetes.

There is also evidence that olanzapine can result in elevations in blood glucose that are life-threatening. In a retrospective epidemiologic survey of spontaneously reported adverse events related to olanzapine therapy, 237 patients with olanzapine-associated diabetes or hyperglycemia were identified. Seventy-three percent of all cases of hyperglycemia appeared within 6 months of initiating olanzapine therapy. Eighty patients had metabolic acidosis or ketosis, 41 had glucose levels of 1,000 mg/dL or greater, and 15 patients died. When olanzapine was discontinued, or the dosage was decreased, 78% of patients had improved glycemic control. Hyperglycemia recurred in 8 of 10 cases with rechallenge.

These concerns regarding glycemic control with olanzapine indicate that other antipsychotics may be preferred for patients who already have diabetes as well as those with risk factors for diabetes such as obesity or a family history of diabetes. In addition, the American Psychiatric Association and Mount Sinai Guidelines recommend that clinicians evaluate fasting blood glucose before and after initiating treatment with olanzapine.

NEUROLEPTIC MALIGNANT SYNDROME AND TARDIVE DYSKINESIA. The product information for olanzapine has the same warnings regarding tardive dyskinesia and neuroleptic malignant syndrome (NMS) as other antipsychotics.

ELEVATED HEPATIC ENZYMES. Placebo-controlled trials of olanzapine found a small incidence of elevated liver enzymes. For example, 2% of olanzapine-treated patients had clinically significant alanine aminotransferase (ALT) elevations, which were defined as three times the upper level of normal. Patients with these elevations did not develop jaundice.

PREGNANCY. Olanzapine is considered pregnancy category C. In a study that compared exposure to all SGAs to no exposure during pregnancy, the authors reported no increased risk of malformation or neonatal complications, except for lower birth weight in the exposed group. A 2018 study using Medicaid data found that the use of olanzapine and quetiapine during pregnancy was associated with an increased risk of gestational diabetes.

NURSING MOTHERS. Olanzapine is excreted in breast milk, but the concentration of olanzapine in breast milk appears to be low. A study was conducted by Gardiner and colleagues that measured the transfer of olanzapine to breast milk and then infant plasma in seven mother–infant pairs. The measured concentration of olanzapine in breast milk was approximately 1% of the maternal dose. Olanzapine was undetectable in the plasma of exposed infants, and all infants were healthy without observable side effects. As with any medication used during breastfeeding, mothers should be advised of the risks and benefits of breastfeeding while on medication and the potential consequences of olanzapine exposure to their infant.

Drug Interactions. Agents that induce CYP1A2, such as omeprazole (Prilosec) and rifampin, may decrease olanzapine plasma concentrations. Smoking also induces CYP1A2 and may have similar effects. Carbamazepine can increase the clearance of olanzapine and may also decrease plasma concentrations. Fluvoxamine, an inhibitor of CYP1A2, can cause increases in olanzapine levels.

Laboratory Interference. Olanzapine does not interfere with common laboratory tests.

Dosage and Administration. Olanzapine is available as 2.5-, 5-, 7.5-, 10-, 15-, and 20-mg tablets; as 5-, 10-, 15-, and 20-mg orally disintegrating tablets (Zydis); as vials with 10 mg of olanzapine for IM injection; and as vials with 150-, 210-, 300-, or 450-mg olanzapine pamoate extended release for IM injection (Relprevv).

TREATMENT OF SCHIZOPHRENIA. Olanzapine is effective at dosages between 7.5 and 20.0 mg per day, administered once daily without regard to meals. A single starting dose of 5 to 10 mg is well tolerated in most adult patients. The recommended initial dose is then adjusted as necessary within the range of 5 to 20 mg per day. Although clinical studies suggest that there is no advantage to prescribing doses above 10 mg daily, the impression of many clinicians is that doses of 15 to 25 mg or more are optimal for acute schizophrenia. This is reflected in the CATIE study, where the mean modal daily dose of olanzapine was 20.1 mg daily. In other words, the mean dose was above the maximum dose of 20 mg that is recommended in the product information. In adolescents, the dose range of 2.5 to 20 mg/day demonstrated efficacy, with the mean dose being 11.1 mg/day.

MAINTENANCE IN SCHIZOPHRENIA. Long-term studies indicate that doses of 10 to 20 mg daily of olanzapine are sufficient to prevent psychotic relapse in stabilized patients.

TREATMENT OF BIPOLAR DISORDER. Patients with mania or mixed states can usually be started on 10 to 15 mg daily of oral olanzapine. Doses can be increased to 20 or 25 mg daily. Similar doses can be prescribed when olanzapine is added to lithium or valproate. Patients with bipolar disorder who have been stabilized can usually be maintained on 5 to 10 mg of olanzapine daily.

SPECIAL POPULATIONS. Patients who are debilitated or vulnerable to hypotension should be started on 5 mg or less. Doses of 5 to 10 mg orally may be effective for schizophrenia or bipolar disorder in these individuals.

IM OLANZAPINE IN AGITATION ASSOCIATED WITH SCHIZOPHRENIA OR BIPOLAR DISORDER. IM doses of 2.5 to 10 mg have been effective in reducing agitation. In most cases, 10 mg IM is an appropriate initial dose. Lower doses may be appropriate for some patients. Doses of 10 mg may be repeated after 2 hours if the patient remains agitated. As noted earlier, the advantage of IM administration is a more rapid onset of calming. Therefore, cooperative patients may be changed to oral administration with rapid dissolve or regular tablets.

OLANZAPINE PAMOATE LONG-ACTING INJECTABLE. In 2009, the FDA-approved olanzapine pamoate as a long-acting IM injectable agent was indicated for the treatment of schizophrenia. Its efficacy was demonstrated in an 8-week randomized controlled trial (*N* = 404), in which patients with schizophrenia were discontinued from their prior oral antipsychotics and underwent a 2- to 7-day washout, followed by initiation of olanzapine long-acting injectable 210 mg every 2 weeks, 405 mg every 4 weeks, 300 mg every 2 weeks, or placebo. The primary outcome measure was the change in PANSS total score from baseline to end point. Olanzapine long-acting injectable showed statistically significant improvement over placebo at all three dosages. A subsequent longer-term trial enrolled 1,065 adults with schizophrenia who were stable for 4 to 8 weeks on oral olanzapine. They were randomized to either continue their current oral olanzapine or switch to olanzapine in long-acting injectable form. The primary outcome measure was the time to exacerbation of schizophrenia, which was defined by prespecified increases in BPRS positive symptom score or hospitalization. In this study, olanzapine long-acting injectable at dosages of 150 mg every 2 weeks, 405 mg every 4 weeks, and 300 mg every 2 weeks were superior to the lower dose of 45 mg every 4 weeks.

Of particular importance, this medication has a black box warning for postinjection/sedation syndrome, which can present with signs and symptoms similar to an overdose of olanzapine (e.g., may include sedation, delirium, EPS, dysarthria, ataxia, aggression, dizziness, weakness, hypertension, and convulsions). In its most extreme presentation, it can lead to coma and require intubation. This syndrome generally occurs within the first 3 hours after injection but can also occur after this period. The frequency of this syndrome is reported to be <0.1% of injections and in roughly 2% of patients who received injections for up to 46 months.

The FDA mandates that this medication be administered at registered facilities with access to emergency services if needed. After injection, patients must be under observation for at least 3 hours before leaving the site. If this medication is used, the prescriber, health care facility, patient, and pharmacy must be enrolled in a patient care program to facilitate medication distribution.

Patients being prescribed olanzapine long-acting injection should already be taking olanzapine orally to ensure medication tolerability. This medication is recommended to be administered every 2 to 4 weeks by deep IM gluteal injection. The dosage schedules specify the initial dose during the first 8 weeks, followed by maintenance dosing thereafter. To achieve levels roughly equivalent to a 10 mg/day oral dose, patients should be initially administered 210 mg every 2 weeks or 405 mg every 4 weeks, followed by maintenance dosing of 150 mg every 2 weeks or 300 mg every 4 weeks. To achieve levels roughly equivalent to a 15 mg/day oral dose, patients should be initially administered 300 mg every 2 weeks, followed by maintenance dosing of 210 mg every 2 weeks or 405 mg every 4 weeks. To achieve levels roughly equivalent to a 20 mg/day oral dose, patients should be initially administered 300 mg every 2 weeks, followed by maintenance dosing of 300 mg every 2 weeks. Patients who are debilitated or more sensitive to olanzapine should be started at 150 mg every 4 weeks.

Ziprasidone

Ziprasidone was synthesized in 1987 by Harry Howard of Pfizer Pharmaceuticals. Tom Seeger screened the compound and identified the activity at D_2 and 5-HT_2 receptors. Early on, it was noted that ziprasidone differed from the other agents in this class by not being associated with weight gain and having relatively few sedative side effects. Its approval in the United States was delayed due to concerns regarding its tendency to prolong the QT interval on the ECG. The FDA required Pfizer to document the extent of the problem in a large trial that evaluated the effects of several antipsychotics on the QT interval. The results of this study led to ziprasidone's approval in the United States in 2001.

Concerns about weight gain, diabetes, and lipid elevations led to an increase in the use of ziprasidone. For example, in the NIMH CATIE trial, patients randomized to ziprasidone showed negligible changes in weight, glucose, or lipids. Moreover, patients who gained weight on other antipsychotics were likely to show substantial weight loss when they were switched to ziprasidone.

Chemistry. Ziprasidone is a benzisothiazolyl piperazine (Fig. 33.2b–1) with a structure unlike any of the other SGAs.

Pharmacologic Action

PHARMACOKINETICS OF ORAL ZIPRASIDONE

Absorption. Ziprasidone is well absorbed when orally administered. Bioavailability doubles when ziprasidone is administered with food. Therefore, patients should be instructed to take ziprasidone with a meal. After administration of multiple doses of ziprasidone under fed conditions, peak serum concentrations occur at 6 to 8 hours.

Distribution. Ziprasidone has a mean volume of distribution of 1.5 L/kg. It is more than 99% protein bound.

Metabolism and Elimination. Ziprasidone is extensively metabolized, with less than 1% excreted unchanged in the urine and less than 4% in feces. In vitro studies indicate that the formation of the major metabolites is mediated by CYP3A4, although there is a small contribution from CYP1A2. Ziprasidone's metabolites are, for the most part, inactive, representing less than 1% of the binding affinity of the parent compound for D_2 and $5\text{-}HT_{2A}$ receptors. Unmetabolized ziprasidone represents 44% of the circulating drug-related material.

Steady state is attained after 1 to 3 days of dosing. Overall exposure at steady state is relative to the dose over the range of 20 to 80 mg twice a day. The mean terminal plasma half-life for ziprasidone at steady state ranges from 5 to 10 hours in volunteers and in schizophrenia patients. The 0.5- to 1-hour lag in reaching peak plasma concentrations has been attributed to delayed gastric emptying. Age, gender, or mild to moderate renal or hepatic impairment have no significant effect on the pharmacokinetics of ziprasidone, making dosage adjustments unnecessary.

Pharmacokinetics of IM Ziprasidone

ABSORPTION AND DISTRIBUTION. Intramuscularly administered ziprasidone, as with all parenterally administered drugs, is 100% bioavailable. Serum concentrations of ziprasidone reach their peak approximately 1 hour after IM administration, with a half-life from 2 to 5 hours.

METABOLISM AND ELIMINATION. After IM administration, metabolism and elimination are anticipated to occur via the same pathways as with oral administration.

Pharmacodynamics

MECHANISM OF ACTION. In vitro studies show that while ziprasidone is a strong D_2 antagonist, it is an even more potent antagonist at the $5\text{-}HT_{2A}$ receptor, with a very high $5\text{-}HT_{2A}$ to D_2 ratio of 11. PET studies in healthy volunteers confirm that $5\text{-}HT_2$ occupancy substantially exceeds D_2 occupancy. Ziprasidone also has a potent affinity for the D_3 receptor and moderate affinity for the D_4 receptor, similar to haloperidol; the affinity for the D_1 receptor is 100-fold lower than for other dopamine receptors. Ziprasidone is an agonist at the $5\text{-}HT_{1A}$ receptor and a potent antagonist at $5\text{-}HT_{2C}$ (5 times greater than clozapine and 19 times greater than risperidone) and $5\text{-}HT_{1D}$ receptors. It also moderately affects norepinephrine and serotonin (and dopamine) uptake sites. Ziprasidone has only moderate affinity for the α_1 receptor and low affinity for the H_1

receptor. Ziprasidone is the only agent providing uptake inhibition of norepinephrine and serotonin. Stimulation of 5-HT_{1A} receptors is shared with clozapine.

BLOOD CONCENTRATIONS AND RELATION TO ACTION. With twice-daily administration of oral ziprasidone, estimates based on systemic exposure and PET scan data predict that 5-HT_2 and D_2 receptor occupancy ranges from 80% to 90% and 45% to 75%, respectively. Ziprasidone at serum concentrations of 30 to 40 ng/mL is likely to be associated with approximately 75% 5-HT_2 and more than 65% D_2 receptor occupancy, as determined from PET scan data.

A consensus statement recommends plasma level monitoring for ziprasidone with a recommended range of 50 to 200 ng/mL.

Design and Interpretation of Clinical Drug Studies

EARLY CLINICAL TRIALS IN SCHIZOPHRENIA. Results from 4- to 6-week, fixed-dosage, double-blind clinical trials indicate that oral ziprasidone is effective in the treatment of positive, negative, and affective symptoms in patients with schizophrenia and schizoaffective disorder. In two separate 4- to 6-week clinical trials in patients with acute psychotic exacerbation of schizophrenia, oral ziprasidone (80 to 160 mg per day) was significantly superior to placebo, and, in a 4-week trial, oral ziprasidone (160 mg per day) was comparable to haloperidol (15 mg per day) in reducing psychotic symptoms. Statistically significant, dose-dependent improvements in negative symptoms, as measured by the PANSS, have been observed with oral ziprasidone (80 and 160 mg per day) compared with placebo in a 6-week trial. Oral ziprasidone (160 mg per day) also significantly improved depressive symptoms in patients with clinically significant symptoms at baseline (Montgomery–Asberg Depression Rating Scale [MADRS] 14 or more; baseline mean, 23.5) compared with placebo at 6 weeks. A statistically significant improvement has also been observed in affective symptoms at 4 weeks in patients with a BPRS anxiety–depression cluster baseline score greater than or equal to 18. Further evidence for efficacy in negative symptoms comes from a 4-week, placebo-controlled trial in which a significant improvement in the BPRS anergia factor was observed with ziprasidone at 120 mg per day. A dose-dependent effect was not observed, which suggests that the negative symptom reduction effect may be related to its lack of extrapyramidal side effects.

A careful look at these short-term trials provides valuable information about the dosing of ziprasidone. In most of the fixed-dose studies, less than 40 mg twice daily was not superior to placebo; results for 40 mg twice daily were mixed. This suggests that for most patients, dosing of ziprasidone should be targeted at above 40 mg twice daily.

CATIE TRIAL AND ZIPRASIDONE. In phase I of CATIE, discontinuation rates on ziprasidone were similar to those for patients treated with the other SGAs and perphenazine. Ziprasidone had the best metabolic profile among the phase I agents. Ziprasidone-treated patients showed slight weight loss, as well as small reductions in cholesterol and triglycerides and trivial increases in glucose levels. Among patients who gained more than 7% of their baseline body weight during phase I, 42% of those assigned to ziprasidone during phase II lost 7% of their new weight.

LONG-TERM EFFECTIVENESS IN SCHIZOPHRENIA. Oral ziprasidone appears to be effective in the long-term treatment of stable patients with chronic schizophrenia according to double-blind, randomized studies. Results from a 28-week, fixed-dose comparative study indicate that ziprasidone is comparable to haloperidol in reducing overall psychopathology during long-term treatment of schizophrenia. Ziprasidone, however, demonstrated superior tolerability and control of negative symptoms.

Results from a prospective, 1-year double-blind, fixed-dose ziprasidone study in patients with chronic or subchronic schizophrenia showed that oral ziprasidone (40 to 160 mg per day) is associated with significantly lower relapse rates compared with placebo. This group of stable patients experienced a significant reduction in negative symptoms from baseline compared with placebo, as measured with the negative symptom subscale of the PANSS. This improvement in negative symptoms was maintained throughout 12 months of treatment, with a tendency for continuing improvement in those patients completing the study.

BIPOLAR MANIA. A study reported by Paul Keck and coworkers randomized 210 patients with mania or mixed states to a 3-week trial of ziprasidone or placebo. Dosing was flexible. From day 8 to 14, the mean ziprasidone dose was 139 mg; between days 15 and 21, it was 130.1.

Ziprasidone-treated patients demonstrated significant improvements compared to placebo-treated patients on a Mania Rating Scale and the CGI. Improvements on ziprasidone emerged by day 2. Advantages of the same instruments were found in a second 3-week controlled study.

AGITATION IN SCHIZOPHRENIA. Two double-blind studies evaluated the efficacy of IM ziprasidone for reducing agitation in schizophrenia during a 24-hour period. Patients could receive up to four injections 4 hours apart. Both studies used the Behavioral Activation Scales (BARS) to evaluate the outcome. One study found that 20 mg was more effective than the 2-mg dose, and the other found that 10 mg was superior to 2 mg. Two other studies were open label and compared 3 days of IM ziprasidone or haloperidol followed by 4 days on the same drug. A 6-week study administered IM ziprasidone or haloperidol for 3 days followed by oral treatment. Each of these studies found that IM ziprasidone had similar or better efficacy against agitation than haloperidol and caused fewer extrapyramidal side effects.

Effects on Specific Organs and Systems. Ziprasidone has substantial α_1-adrenergic antagonist effects, which may explain a tendency to cause orthostatic hypotension, tachycardia, and dizziness in some individuals. Ziprasidone can cause prolongation of the QTc interval (see Precautions and Adverse Reactions).

Therapeutic Indications

SCHIZOPHRENIA. Ziprasidone is effective for the treatment of schizophrenia. The evidence supporting its efficacy is from 4- to 6-week trials indicating greater efficacy than placebo. A large 52-week double-blind trial also found that ziprasidone was more effective than placebo for preventing relapse in stabilized patients with schizophrenia.

BIPOLAR I DISORDER. Ziprasidone is also indicated as monotherapy for the acute treatment of manic or mixed episodes in bipolar I disorder. It is also indicated for the maintenance treatment of bipolar I disorder as an adjunct to lithium or valproate. The efficacy of ziprasidone as monotherapy for maintenance therapy has not been systematically evaluated.

ACUTE TREATMENT OF AGITATION IN SCHIZOPHRENIA. IM ziprasidone is effective for managing agitation in patients with schizophrenia. The advantage of IM medication is a more rapid onset of effects. The product information for ziprasidone has the following description of the type of individual where ziprasidone may be useful: “Schizophrenic patients experiencing agitation often manifest behaviors that interfere with their diagnosis and care, for example, threatening behaviors, escalating or urgently distressing behavior, or self-exhausting behavior, leading clinicians to the use of IM antipsychotic medications to achieve immediate control of the agitation.”

Precautions and Adverse Reactions

BLACK BOX WARNING. Ziprasidone and the other SGAs carry a black box warning for increased mortality in older adult patients with dementia-related psychosis compared with placebo.

QT PROLONGATION. Ziprasidone is associated with a dose-related prolongation of the QT interval on the ECG. This is a concern since a prolonged QT interval can lead to the development of a characteristic ventricular tachycardia known as torsade de pointes. This arrhythmia can result in hypotension or, more seriously, ventricular fibrillation and death. There is a concern that if ziprasidone was administered in combination with other agents associated with QTc prolongation, the effects could be additive. As a result, the product information for ziprasidone warns against combining it with other drugs that may increase the QT interval, including dofetilide (Tikosyn), sotalol (Betapace), quinidine, other class Ia and III antiarrhythmics, mesoridazine (Serentil), thioridazine, chlorpromazine, droperidol (Inapsine), pimozide (Orap), sparfloxacin (Zagam), gatifloxacin (Tequin), moxifloxacin (Avelox), halofantrine (Halfan), mefloquine (Lariam), pentamidine (Nebupent), arsenic trioxide (Trisenox), levomethadyl acetate (Orlaam), dolasetron mesylate (Anzemet), probucol (Lorelco), or tacrolimus (Prograf). Ziprasidone should also not be prescribed for patients with existing QT prolongation. In addition, it should not be prescribed in patients who have conditions such as hypokalemia and hypocalcemia, which can prolong the QT interval.

An interesting study that was required prior to the approval of ziprasidone by the FDA gives a perspective on the amount of QT prolongation associated with ziprasidone and other antipsychotics. Patients receiving six antipsychotics had their QT interval measured at their maximal plasma concentration. This was done in the presence and absence of a metabolic inhibitor of the antipsychotic in question (in the presence of an inhibitor, the plasma concentration of the antipsychotic drug increases). Ziprasidone administration resulted in a mean increase in the QTc (i.e., the QT interval corrected for heart rate) of 9 to 14 ms greater than four of the comparator drugs (risperidone, olanzapine, quetiapine, and haloperidol) but 14 ms less than thioridazine. This finding was sufficiently reassuring to lead to the approval of ziprasidone.

USE IN PREGNANCY AND DURING LACTATION. Ziprasidone is considered category C for pregnancy. Although there are no well-controlled studies on the risk of ziprasidone to the fetus, animal studies suggest that there is a potential for teratogenic effects from ziprasidone. Unlike the other older SGAs, little is published on the effects of ziprasidone during pregnancy and lactation. Therefore, ziprasidone should only be taken during pregnancy if the clinical benefits justify the risk to the fetus.

OTHER SIDE EFFECTS. Double-blind comparisons of ziprasidone and placebo have not found a statistically significant difference in EPS or akathisia. However, case reports suggest that some patients experience an increase in the restlessness that may be mild akathisia.

Drug Interactions. Ziprasidone should not be combined with drugs that prolong the QT interval. Otherwise, ziprasidone has a low potential for clinically significant drug interactions. In vitro studies indicate that ziprasidone is not a substrate for CYP2D6, CYP2C9, CYP2C19, or CYP1A2 and has low potential to inhibit CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4. The lack of clinically relevant CYP2D6 inhibition has been confirmed in healthy volunteers, in whom ziprasidone did not inhibit the metabolism of dextromethorphan (Benylin) (a CYP2D6 model substrate) to its major metabolite, dextrorphan. Although ziprasidone is a substrate for CYP3A4, cimetidine (a CYP3A4 inhibitor) does not inhibit its metabolism, suggesting that significant metabolism occurs by an alternative pathway. Carbamazepine induces CYP3A4 and lowers ziprasidone plasma concentrations. In healthy subjects, clinically significant interactions have not been reported with lithium, cimetidine, antacids (aluminum and magnesium hydroxides), ketoconazole, or oral contraceptives. Ziprasidone did not affect steady-state lithium concentrations.

Laboratory Interference. Ziprasidone does not affect common laboratory tests.

Dosage and Administration

SCHIZOPHRENIA. For most adult patients, ziprasidone should be initiated at 20 mg twice a day with food. Dosage may be increased every 2 days. As mentioned earlier, trials in schizophrenia indicate that patients receiving a total daily dose of 100 mg or greater are more likely to respond. Daily dosage can be adjusted on the basis of clinical status up to a maximum of 80 mg twice a day.

Duration of Treatment. The benefits of ziprasidone in schizophrenia have been found to be maintained for periods of up to 1 year when given at dosages ranging from 20 to 80 mg twice a day, although, in these studies, no additional benefits were seen at dosages higher than 20 mg twice a day. Ziprasidone has been administered safely for up to 2 years.

BIPOLAR I DISORDER. For the acute treatment of manic or mixed episodes, ziprasidone should be given at an initial dose of 40 mg twice daily with food. It may be increased to 60 mg or 80 mg twice daily on day 2 and titrated according to efficacy and tolerance. In flexible-dose trials, the mean daily dose used was around 120 mg. For maintenance treatment as an adjunct to lithium or valproate, patients should be continued at the same dose as initial stabilization, within the range of 40 to 80 mg twice daily with food.

ACUTE TREATMENT OF AGITATION IN SCHIZOPHRENIA (IM FORMULATION). The recommended IM dosage is 10 to 20 mg administered every 2 hours for the 10-mg dose and every 4 hours for the 20-mg dose, up to a maximum dosage of 40 mg per day. IM administration has not been studied for longer than 3 consecutive days, and it is recommended that oral therapy be initiated if longer-term treatment is necessary.

Lurasidone

Lurasidone is an antipsychotic that is a potent antagonist at 5-HT_{2A}, D_2, and 5-HT_7 receptors and a partial agonist at the 5-HT_{1A} receptor. It is among the DSAs (along with aripiprazole and ziprasidone) that have a relatively low liability for metabolic side effects. Patients should be cautioned to take lurasidone with food. Its most common side effects include sedation, extrapyramidal side effects (including akathisia), and nausea.

Lurasidone was developed by Dainippon Sumitomo Pharma and marketed by Sunovion in the United States. It was approved by the FDA for the treatment of schizophrenia in adults in 2010. Subsequently, approval was granted in 2013 for the treatment of adults with bipolar depression (as both monotherapy and as an adjunct to lithium or valproate therapy).

Chemistry. Lurasidone belongs to the class of benzisothiazol derivatives, with the full chemical name of (3aR,4S,7R,7aS)-2-[((1R,2R)-2-{[4-(1,2-benzisothiazol-3-yl)-piperazin-1-yl]methyl}cyclohexyl)methyl]hexahydro-1H-4,7-methanisoindol-1,3-dione. Its molecular formula is $C_{28}H_{36}N_4O_2S$, and its molecular weight is 492.676 (free base). It is currently used in the form of its hydrochloride salt.

Pharmacologic Action

PHARMACOKINETICS

Absorption. Lurasidone has a relatively low oral bioavailability of 9% to 19%, and its absorption is affected by food intake. Administration of lurasidone with a meal of at least 350 calories increases drug exposure.

Distribution. Peak serum concentrations are reached in 1 to 3 hours. The volume of distribution is 6,173 L, and lurasidone is highly bound to serum proteins.

Metabolism and Elimination. Lurasidone is metabolized mostly by CYP3A4 and is not a substrate for CYP1A1, CYP1A2, CYP2A6, CYP4A11, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, or CYP2E1. Its major biotransformation pathways are oxidative *N*-dealkylation, hydroxylation of its norbornane ring, and S-oxidation. It has two major active and two major inactive metabolites. Lurasidone is eliminated primarily in the feces (80%), with 9% of labeled metabolites recovered in urine. The mean elimination half-life was found to be 18 hours.

PHARMACODYNAMICS

Mechanism of Action. Lurasidone is an antagonist with a high affinity for the dopamine D_2 receptor as well as serotonin 5-HT_{2A} and 5-HT_7 receptors. It has been found to occupy 60% to 70% of D_2 receptors at a 40-mg dose. In addition, it is an antagonist with moderate affinity for α_{2A} and α_{2C} adrenergic receptors and is a partial agonist at the 5-HT_{1A} receptor. It has a low affinity for the serotonin 5-HT_{2C} receptor and negligible affinity for histamine H_1 and muscarinic M_1 receptors.

Design and Interpretation of Clinical Drug Studies

SCHIZOPHRENIA. Efficacy was established for the treatment of schizophrenia by five 6-week randomized controlled trials, with two including active comparators. The studies were conducted on adult patients with DSM-IV schizophrenia. Two of these studies (*N* = 145 and B = 180) used the change in Brief Psychiatric Rating Scale derived (BPRSd) as the primary outcome measure, and lurasidone significantly outperformed placebo at 40, 80, and 120 mg/day dosages. A third study (*N* = 489) used the change in PANSS total score as the primary outcome measure, and lurasidone 80 mg/day was found significantly superior to placebo, but the 40 and 80 mg/day dosages did not separate from placebo. A fourth study (*N* = 473) used olanzapine 15 mg/day as an active comparator. Lurasidone 40 and 120 mg/day, as well as olanzapine 15 mg/day, were significantly more effective than placebo on the primary outcome measure of change in PANSS total score. In the fifth study (*N* = 482), patients were randomized to receive lurasidone 80 mg/day, lurasidone 160 mg/day, quetiapine extended release 600 mg/day, or placebo. The primary outcome measure was the change in PANSS total score. Both doses of lurasidone, as well as quetiapine extended release, were associated with significant improvement in the primary outcome measure as compared to placebo. A sixth study (*N* = 353) involved a random assignment to lurasidone 20, 40, and 80 mg/day or haloperidol 10 mg/day as an active comparator. This study was considered a failure because lurasidone at all dosages, as well as haloperidol, did not separate from placebo on the primary outcome measure (BPRSd).

The long-term effectiveness of lurasidone has been assessed in a 52-week controlled trial in which 639 stable outpatients with schizophrenia were randomized to receive lurasidone 40 to 120 mg/day or risperidone 2 to 6 mg/day. The two groups did not statistically differ in relapse rate (defined as a 30% increase in PANSS total score compared to baseline). Both groups also showed statistically equivalent improvement in PANSS total score over the 12 months.

DEPRESSIVE EPISODES ASSOCIATED WITH BIPOLAR I DISORDER

Monotherapy. Efficacy was established in a 6-week randomized controlled trial (*N* = 485) of adults with DSM-IV bipolar I disorder in a major depressive episode, with or without rapid cycling and without psychotic features. Participants were randomized to receive flexible-dose lurasidone 20 to 60, 80 to 120 mg/day, or a placebo. The primary outcome measure was the change in the MADRS score from baseline to week 6. Both dose ranges of lurasidone were superior to placebo, and the effectiveness of the lower- versus higher-range dosing of lurasidone was not statistically different.

Adjunctive Therapy. Efficacy was established in a 6-week randomized controlled trial (*N* = 340) of adults with DSM-IV bipolar I disorder in a major depressive episode, with or without rapid cycling and without psychotic features. Patients who were still symptomatic despite lithium or valproic acid treatment were randomized to receive flexible-dose lurasidone 20 to 120 mg/day or a placebo. The primary outcome measure was the change in MADRS from baseline to week 6. Lurasidone, as an adjunct to valproic acid or lithium, was statistically superior to placebo in the reduction of MADRS score.

Therapeutic Indications

SCHIZOPHRENIA. Lurasidone is indicated for the treatment of adults and adolescents (aged 13 to 17 years) with schizophrenia. Its efficacy was established in five short-term studies of adults with schizophrenia. Lurasidone is not indicated at the time of writing for longer-term use (greater than 6 weeks).

DEPRESSIVE EPISODES ASSOCIATED WITH BIPOLAR I DISORDER

Monotherapy. Lurasidone is indicated as monotherapy for the treatment of adults and pediatric patients (aged 10 to 17 years) with major depressive episodes associated with bipolar I disorder. Its efficacy for this indication was established in a 6-week monotherapy study of adults with bipolar depression.

Adjunctive Therapy with Lithium or Valproate. Lurasidone is indicated as adjunctive therapy (with either lithium or valproate) for the treatment of patients with major depressive episodes associated with bipolar I disorder. Its efficacy for this indication was established in a 6-week study of adults with bipolar depression who were being treated with lithium or valproate.

Precautions and Adverse Reactions. Lurasidone is contraindicated in individuals with hypersensitivity to lurasidone or any components in the tablets. Lurasidone is also contraindicated in patients using strong CYP3A4 inhibitors (e.g., ketoconazole, clarithromycin, ritonavir, voriconazole, and mibefradil) or CYP3A4 inducers (e.g., rifampin, avasimibe, St. John’s wort, phenytoin, and carbamazepine).

BLACK BOX WARNINGS

Increased Mortality in Older Adult Patients with Dementia. Lurasidone and the other SGAs carry a black box warning for increased mortality in older adult patients with dementia-related psychosis compared with placebo.

Suicidal Thoughts and Behaviors. Lurasidone also carries a black box warning for suicidal thoughts and behavior in children, adolescents, and young adults. This warning is associated with all antidepressant medications, including SGAs, with indications for the treatment of depression.

NEUROLEPTIC MALIGNANT SYNDROME. NMS has been associated with the use of antipsychotic medications, including lurasidone. If NMS is suspected, lurasidone and other antipsychotic medications should be discontinued immediately, and the patient should undergo intensive supportive care and medical monitoring.

TARDIVE DYSKINESIA. Tardive dyskinesia can develop in all patients treated with antipsychotic drugs. Its prevalence is highest among the older adults.

METABOLIC CHANGES. DSAs have been associated with weight gain, lipid elevations, and hyperglycemia, including severe cases with ketoacidosis, hyperosmolar coma, and death. Among available antipsychotics, lurasidone has a low to moderate risk of metabolic changes.

QT PROLONGATION. Lurasidone was evaluated in a randomized controlled QT study in 43 patients with schizophrenia or schizoaffective disorder. Participants were administered lurasidone 120 or 600 mg, and ECGs were assessed 2 to 4 hours after dosing. Patients who received the 120-mg dose experienced a 7.5 ms increase in QTc, and those who received the 600 mg experienced a 4.6 ms increase. There was no observed dose–response relationship. In short-term schizophrenia and bipolar depression clinical trials, no study participants experienced QTc prolongation to greater than 500 ms.

SEIZURES. In short-term premarketing schizophrenia studies, seizures occurred in 0.1% (2/1,508) of patients treated with lurasidone as compared to 0.1% (1/708) of patients treated with placebo. In short-term premarketing bipolar depression studies, no patients experienced seizures.

ORTHOSTATIC HYPOTENSION AND SYNCOPE. Lurasidone may cause orthostatic hypotension and syncope. In pooled premarketing short-term schizophrenia studies, orthostatic hypotension was reported in 0.3% of lurasidone and 0.1% of placebo patients; syncope was reported in 0.1% of lurasidone and no placebo patients. In the premarketing short-term bipolar depression study, there were no reported adverse events of orthostatic hypotension and syncope. However, orthostatic hypotension, as assessed by vital signs, occurred in 0.6% of lurasidone patients as compared to 0% receiving placebo. In the premarketing short-term adjunctive therapy bipolar depression study, there were no reported adverse events of hypotension and syncope. Orthostatic hypotension, as assessed by vital signs, occurred in 1.1% of lurasidone patients and 0.9% of placebo patients.

HYPERPROLACTINEMIA. As with most antipsychotic medications, lurasidone can increase prolactin levels. In pooled short-term schizophrenia studies, changes in prolactin level with placebo (*N* = 672) and lurasidone 20 (*N* = 70), 40 (*N* = 476), 80 (*N* = 495), 120 (*N* = 284), and 160 mg/day (*N* = 115) were −1.9, −1.1, −1.4, −0.2, +3.3, and +3.3 ng/mL, respectively.

POTENTIAL FOR COGNITIVE AND MOTOR IMPAIRMENT. Lurasidone can cause somnolence, which can impair judgment, thinking, and motor skills.

SERUM CREATININE. In short-term schizophrenia trials, serum creatinine levels increased 0.05 mg/dL in lurasidone-treated participants as compared to 0.02 mg/dL in placebo-treated participants. Changes in creatinine levels appeared to be dose dependent, with 7% of patients receiving lurasidone 160 mg/day (*N* = 121) shifting from normal to abnormally high creatinine levels as compared to 2% of those receiving placebo (*N* = 708).

EXTRAPYRAMIDAL SYMPTOMS. In premarketing schizophrenia studies, EPS (excluding akathisia) was reported in 13.5% of patients receiving lurasidone as compared to 5.8% of patients receiving placebo. Akathisia was reported in 12.9% of patients receiving lurasidone as compared to 3% of patients receiving a placebo. These effects were dose dependent, with the highest percentage of events occurring in the group receiving lurasidone 120 mg/day. EPS were also reported in short-term bipolar depression studies with similar frequencies. Dystonic symptoms occurred in 4.2% of patients in premarketing schizophrenia trials versus 0.8% of those receiving placebo, and 7/1,508 participants (0.5%) discontinued the clinical trial due to dystonic events, which occurred during the first few days of treatment. Dystonia occurred less frequently with lurasidone in bipolar depression studies though it still occurred more often than in the placebo groups. No participants in bipolar depression studies discontinued participation due to dystonic symptoms.

PREGNANCY. Lurasidone is classified as pregnancy category B, which indicates that animal reproductive studies did not identify a specific risk to the fetus, but there are no adequate and well-controlled studies on pregnant women. Lurasidone should only be used in pregnancy if the potential benefit justifies the potential risk to the fetus.

NURSING MOTHERS. Rat studies demonstrated excretion of lurasidone in milk during lactation. It is unknown whether this also occurs in humans. Mothers should consider discontinuation of nursing due to the potential for adverse reactions in infants.

COMMON AND DRUG-RELATED ADVERSE REACTIONS IN CLINICAL TRIALS. In short-term premarketing schizophrenia studies, the most common adverse reactions associated with lurasidone (defined as incidence ≥5% and at least twice the rate of placebo) were somnolence, akathisia, EPS, and nausea. In the bipolar depression monotherapy study, the most common adverse reactions associated with lurasidone were akathisia, EPS, somnolence, nausea, vomiting, diarrhea, and anxiety. In the bipolar depression adjunctive therapy study, the most common adverse reactions associated with lurasidone were akathisia and somnolence. In premarketing studies for both schizophrenia and bipolar disorder, there were no adverse reactions associated with lurasidone discontinuation (defined as ≥2% occurrence and at least twice the rate of placebo).

Drug Interactions. It is recommended that lurasidone not be used in conjunction with strong CYP3A4 inhibitors (e.g., ketoconazole, clarithromycin, ritonavir, voriconazole, and mibefradil). If lurasidone is used with moderate CYP3A4 inhibitors (e.g., diltiazem, atazanavir, erythromycin, fluconazole, and verapamil), the lurasidone dose should be reduced by half. Grapefruit and grapefruit juice should be avoided in conjunction with lurasidone use.

Lurasidone is contraindicated for use with strong CYP3A4 inducers (e.g., rifampin, avasimibe, St. John’s wort, phenytoin, and carbamazepine). If lurasidone is used in conjunction with a moderate CYP3A4 inducer, it may be necessary to increase lurasidone dosage after at least 7 days of treatment with the CYP3A4 inducer.

Lithium and valproic acid dosages do not need to be adjusted if used concomitantly with lurasidone. Dosages of P-glycoprotein substrates (i.e., digoxin) and other CYP3A4 substrates (e.g., midazolam or oral contraceptives) also do not need to be adjusted if used with lurasidone.

Dosage and Administration. Lurasidone is available in 20-, 40-, 60-, 80-, and 120-mg tablets. It is recommended that lurasidone is taken with at least 350 calories of food to improve absorption.

For the treatment of schizophrenia, it is recommended to start lurasidone at 40 mg once daily. Lurasidone has been found effective for schizophrenia in the dose range of 40 to 160 mg daily, and the maximum recommended dose is 160 mg daily.

For the treatment of bipolar I depression, the recommended starting dose is 20 mg once daily, given as either monotherapy or as an adjunct to lithium or valproate. Lurasidone has been found effective for bipolar I depression in the dose range of 20 to 120 mg daily, and the maximum recommended dose for both monotherapy and adjunct therapy is 120 mg daily.

DOSING IN SPECIAL POPULATIONS

Geriatric Use. There were insufficient numbers of patients aged 65 years and older in lurasidone trials to determine whether they respond differently to the medication than younger adults. In a limited sample of older adult patients with psychosis (65 to 85 years of age), plasma concentrations of lurasidone after 20 mg/day administration were similar to those in younger adults.

Hepatic Impairment. In patients with moderate (Child–Pugh Score of 7 to 9) or severe hepatic impairment (Child–Pugh score of 10 to 15), dose reduction in lurasidone is recommended. The recommended starting dose is 20 mg daily, and the maximum recommended dose in moderate or severe hepatic impairment is 80 and 40 mg/day, respectively.

Renal Impairment. In patients with moderate or severe renal impairment (creatinine clearance less than 50 mL/min), dose reduction in lurasidone is recommended. In this population, the recommended starting dose is 20 mg daily, and the maximum recommended dose is 80 mg daily.

DOPAMINE, SEROTONIN, NOREPINEPHRINE α2

Clozapine

Clozapine was the first of the DSAs. It was discovered in 1958 in Bern, Switzerland, and was first studied in animal experiments in 1960. Soon after clinical trials with clozapine began, the drug appeared to be an effective antipsychotic agent that did not cause extrapyramidal side effects. The first studies showed that patients became calmer in the first 5 days, and psychotic excitement decreased from day 5 onward. Day 20 effects on psychosis were substantial, and negative symptoms improved at 40 days. However, early studies with clozapine in Europe were met with skepticism because this agent was inconsistent with the dogma that effective antipsychotic agents were associated with extrapyramidal effects. This dogma was challenged by a number of European authorities, including Hans Hippius, and, as a result, clozapine was eventually approved in several countries in Europe in 1972.

The enthusiasm about clozapine was dampened when it was discovered that the drug was associated with significant hematologic toxicity. In 1975, clozapine was introduced in Finland, where 16 of 1,600 treated patients developed granulocytopenia (1,600 cells or less per mm^3). Eight of the 13 patients in whom granulocytopenia progressed to agranulocytosis died of infectious diseases. After 50 patients around the world had died, clozapine was withdrawn from most European markets, and research with this drug came to a virtual halt. Some patients who were receiving clozapine before its withdrawal from the market continued on the drug because of its clear advantages over prior treatment. Regular monitoring of white blood cell (WBC) counts in these individuals showed that clozapine-induced agranulocytosis was reversible when drug use was stopped. In addition, controlled trials in Europe showed that clozapine was more effective than conventional agents for treating schizophrenic patients who were either severely ill or resistant to other antipsychotic agents.

Discussions with the FDA led to a U.S. multicenter trial to determine whether clozapine was effective for treatment-refractory patients. The effectiveness of clozapine was compared with that of chlorpromazine (Thorazine) in 268 schizophrenia patients with thoroughly documented failure to respond to conventional antipsychotic agents. The study, reported in 1988 by John Kane, Herbert Meltzer, and others, found that clozapine was more effective than chlorpromazine in terms of a number of important outcomes, including psychosis, mood, and anxiety. The results from this study and others led to the approval of clozapine in the United States in 1990 for schizophrenia patients who are resistant to treatment with other antipsychotic drugs or who are unable to tolerate conventional drugs because of extrapyramidal side effects or severe tardive dyskinesia.

Treatment resistance in schizophrenia is a poorly defined concept that has evolved since clozapine's introduction. At that time, treatment resistance referred to a failure to respond to a conventional dopamine receptor antagonist. Because of clozapine's side-effect profile, it was hoped that other DSAs would replace it as the agent of choice for treatment-resistant schizophrenia. However, studies that compared clozapine to other DSAs for treatment-resistant patients nearly always found that clozapine was superior. Over the years, meta-analyses have consistently identified clozapine as the most effective antipsychotic for schizophrenia.

Chemistry. Clozapine is a five-membered heterocyclic compound, 8-chloro-11-(4-methyl-1-piperazinyl)-5*H*-dibenzo-[*b,e*][1,4-]diazepine. Its chemical formula is $C_{18}H_{19}N_4Cl$, with a molecular weight of 326.8. Substitutions for the A-ring have led to other clinically effective clozapine-like agents; however, loxapine (Loxitane), an antipsychotic with a clozapine-related molecular structure, has a clinical and side-effect profile similar to other traditional antipsychotic drugs. Other clozapine-derived antipsychotic agents are olanzapine and quetiapine, but being a derivative does not translate to similar efficacy. Chemical structures for clozapine and the other SGAs discussed in this chapter are depicted in Figure 33.2b–1.

Pharmacologic Actions

PHARMACOKINETICS. The pharmacokinetic profiles of clozapine and other DSAs are summarized in Table 33.2b–1.

Absorption. Clozapine is only available as an oral preparation. As with other antipsychotic agents, peak plasma levels are reached approximately 2 hours after oral administration (range, 1.1 ± 0.5 to 3.0 ± 1.5 hours). The elimination half-life is approximately 12 hours (between 10 and 16 hours; range, 6 to 33 hours). As a result, twice-daily dosing results in steady-state plasma concentrations in less than 1 week. Based on animal studies, coadministration of highly protein-bound drugs may increase free clozapine concentrations, although the total concentration (free plus bound) may be unchanged.

Distribution. Clozapine has a large volume of distribution (mean of 2.0 to 5.1 L/kg [range, 1.0 to 10.2 L/kg]), although this is lower than other antipsychotic drugs.

Metabolism and Elimination. Clozapine undergoes extensive first-pass metabolism in the liver and gut. Absolute bioavailability (percentage of clozapine reaching the systemic circulation unchanged) after oral administration varies from 27% to 47%. The primary metabolic routes are demethylation, *N*-oxidation, oxidation of the aromatic ring, and conjugation. Reportedly 80% of administered clozapine appears in the urine or feces as metabolites. Less than 5% of the parent compound is found unchanged in the urine.

The *N*-desmethylclozapine metabolite represents 40% to 80% of the concentration of clozapine. It is a muscarinic type 1 (M_1) receptor agonist, whereas clozapine itself has muscarinic antagonist effects. According to Herbert Meltzer, this metabolite may be a major contributor to clozapine's purported superior effects on cognition and negative symptoms compared to the other agents.

PHARMACODYNAMICS

Mechanism of Action. The explanation for clozapine's unique clinical profile remains controversial. Preclinical studies in rats have shown that clozapine blocks the conditioned avoidance response, indicating that the drug has potential antipsychotic activity. However, it did not produce catalepsy in animals, suggesting that it is unlikely to produce EPS in humans. This made it different from conventional antipsychotic medications, which all cause extrapyramidal side effects as a direct result of their D_2 receptor blockade in the forebrain.

The relatively low affinity for D_2 receptors may explain clozapine's lack of extrapyramidal side effects. In patients treated with most FGAs and SGAs, PET scan studies have shown that, with doses comparable to 10 mg of haloperidol (Haldol), more than 80% of striatal D_2 receptors are occupied. Higher doses do not increase receptor occupancy further. In clinically effective doses, clozapine only occupies 40% to 50% of D_2 receptors. This weak effect at the D_2 receptor site may be related to clozapine's unusual biologic profile that combines a low D_2 affinity with activity at a number of other receptor sites, including serotonergic (e.g., 5-HT_2 and 5-HT_{1C}), adrenergic, and cholinergic sites.

As mentioned above, the clozapine metabolite *N*-desmethylclozapine has M_1 agonist activity. Studies suggest that patients who have higher ratios of this metabolite to the parent compound tend to have better clinical outcomes. This suggests that M_1 agonism may contribute to clozapine's unique effectiveness.

Philip Seeman and Shitij Kapur have proposed that the lack of EPS with clozapine and other DSAs is related to their fast dissociation from the D_2 receptor. Conventional dopamine receptor antagonists tend to bind more tightly to dopamine receptors than dopamine itself. Clozapine and nearly all of the other SGAs bind more loosely and tend to come off the receptor more readily in the presence of dopamine. As a result, these newer agents may block these receptors more transiently, coming off the receptor to permit more normal dopamine transmission. Clozapine also has a 10-fold higher affinity for the D_4 receptor than do other antipsychotics. The D_4 receptor is present in the cortex and less so in the striatal areas. However, other agents with a high affinity for D_4 receptors have failed to demonstrate that they are effective antipsychotics.

Table 33.2b–1.
Changes in Pharmacokinetics: Special Populations

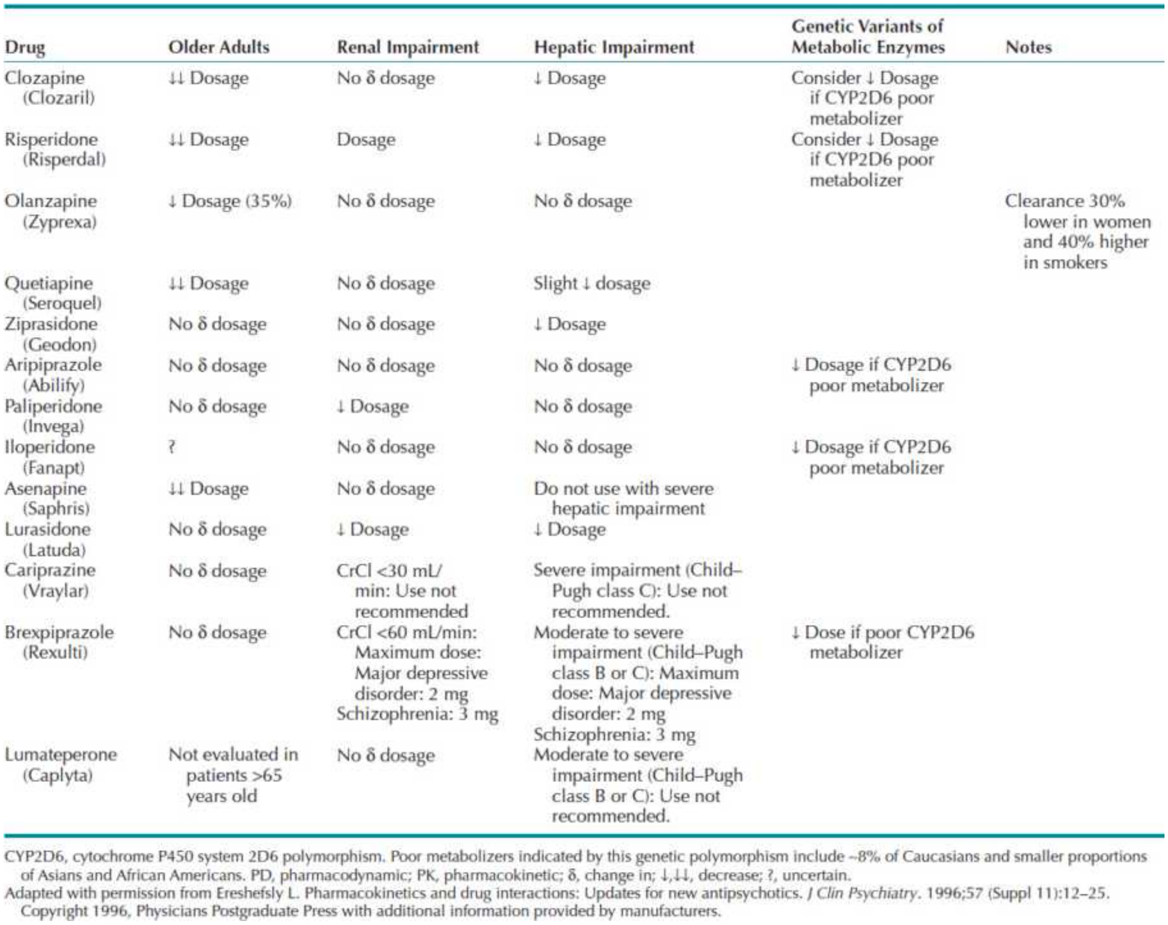

Drug	Older Adults	Renal Impairment	Hepatic Impairment	Genetic Variants of Metabolic Enzymes	Notes
Clozapine (Clozaril)	↓↓ Dosage	No δ dosage	↓ Dosage	Consider ↓ Dosage if CYP2D6 poor metabolizer	
Risperidone (Risperdal)	↓↓ Dosage	Dosage	↓ Dosage	Consider ↓ Dosage if CYP2D6 poor metabolizer	
Olanzapine (Zyprexa)	↓ Dosage (35%)	No δ dosage	No δ dosage		Clearance 30% lower in women and 40% higher in smokers
Quetiapine (Seroquel)	↓↓ Dosage	No δ dosage	Slight ↓ dosage		
Ziprasidone (Geodon)	No δ dosage	No δ dosage	↓ Dosage		
Aripiprazole (Abilify)	No δ dosage	No δ dosage	No δ dosage	↓ Dosage if CYP2D6 poor metabolizer	
Paliperidone (Invega)	No δ dosage	↓ Dosage	No δ dosage		
Iloperidone (Fanapt)	?	No δ dosage	No δ dosage	↓ Dosage if CYP2D6 poor metabolizer	
Asenapine (Saphris)	↓↓ Dosage	No δ dosage	Do not use with severe hepatic impairment		
Lurasidone (Latuda)	No δ dosage	↓ Dosage	↓ Dosage		
Cariprazine (Vraylar)	No δ dosage	CrCl <30 mL/min: Use not recommended	Severe impairment (Child–Pugh class C): Use not recommended.		
Brexpiprazole (Rexulti)	No δ dosage	CrCl <60 mL/min: Maximum dose: Major depressive disorder: 2 mg Schizophrenia: 3 mg	Moderate to severe impairment (Child–Pugh class B or C): Maximum dose: Major depressive disorder: 2 mg Schizophrenia: 3 mg	↓ Dose if poor CYP2D6 metabolizer	
Lumateperone (Caplyta)	Not evaluated in patients >65 years old	No δ dosage	Moderate to severe impairment (Child–Pugh class B or C): Use not recommended.		

CYP2D6, cytochrome P450 system 2D6 polymorphism. Poor metabolizers indicated by this genetic polymorphism include ~8% of Caucasians and smaller proportions of Asians and African Americans. PD, pharmacodynamic; PK, pharmacokinetic; δ, change in; ↓,↓↓, decrease; ?, uncertain.

Adapted with permission from Ereshefsly L. Pharmacokinetics and drug interactions: Updates for new antipsychotics. *J Clin Psychiatry*. 1996;57 (Suppl 11):12–25. Copyright 1996, Physicians Postgraduate Press with additional information provided by manufacturers.

Blood Concentrations and Relation to Action. As with most antipsychotic drugs, patients display considerable variability in their clozapine plasma concentrations among patients receiving the same daily dose. This probably results from differences in both absorption and metabolism. Variability is also observed within patients, which is presumably related to variable absorption rates rather than changes in hepatic metabolism and clearance. Some reports show that women have slightly higher and smokers slightly lower plasma levels (20% to 30%). The plasma levels of older adults may be twofold higher than those of young adults.

Several investigators have studied the relationship between plasma concentrations of clozapine and clinical response. Their findings suggest that monitoring clozapine plasma concentrations may be useful under certain conditions. A wide range of clozapine concentrations appears to be associated with a clinically meaningful response. Plasma concentrations of clozapine average approximately 10 to 80 ng/mL/mg of drug given per kilogram of weight. Thus, a typical daily dosage of 300 to 400 mg (approximately 5 mg/kg) is associated with plasma concentrations ranging from 200 to 400 ng/mL. Studies indicate that patients are more likely to respond when their clozapine plasma concentrations are >350 ng/mL. If patients with a plasma concentration of 250 ng/mL have not responded by 6 weeks, the clinician should consider adjusting the dosage to raise the concentration to approximately 350 ng/mL. The correlations between plasma concentration and clinical improvement or side effects are weak, as is observed with other antipsychotic agents.

Design and Interpretation of Clinical Drug Studies

CLOZAPINE FOR ACUTE SCHIZOPHRENIA. Open-label clinical trials conducted in the early 1970s suggested that clozapine was an effective antipsychotic drug that lacked extrapyramidal side effects. These findings were then reinforced with double-blind trials comparing clozapine with other agents. In 79% of the controlled trials that compared clozapine with another antipsychotic drug, clozapine was superior. In reviewing these studies, Ross Baldessarini cautions that the differences between clozapine and the comparator drug were relatively small. Nevertheless, certain patterns emerge. One comparison with chlorpromazine found a small advantage favoring clozapine when a heterogeneous group of patients was studied, but differences favoring clozapine were clearer when the most severely ill subgroup was studied.

EFFICACY IN TREATMENT-REFRACTORY PATIENTS. Early clinical experience with clozapine suggested that it was more effective than conventional drugs for patients who were unresponsive to conventional antipsychotic agents. Moreover, the studies cited in the previous section suggested that the differences between clozapine and other antipsychotics were greater in patients who were severely ill and refractory to treatment with conventional dopamine receptor antagonists. Before marketing clozapine to these patients, the FDA and the manufacturer of clozapine developed a study comparing clozapine with chlorpromazine plus benztropine (Cogentin). The 268 patients who participated in this trial had previously responded poorly in at least three prior antipsychotic trials and a prospective, single-blind trial of haloperidol. This study compared the effectiveness of chlorpromazine and clozapine in 248 patients who had thorough documentation of poor responsiveness to conventional antipsychotic agents. During this 6-week trial, the clozapine-treated patients did better on a range of measures of both positive and negative symptoms. Thirty percent of clozapine-treated patients met a priori criteria for significant improvement, in contrast to only 4% of chlorpromazine-treated patients. Other studies show that this 6-week trial may have significantly underestimated the proportion of patients who respond to clozapine and that up to 60% of patients may display substantial improvement after a 6-month trial. Clozapine in this trial was superior to chlorpromazine on 14 of 18 BPRS items, as well as on the CGI and nursing ratings such as social competence, social interest, personal neatness, irritability, manifest psychosis, and motor retardation. When patients were classified as having improved (by prospective criteria) to a clinically significant extent, 30% of clozapine patients improved within 6 weeks versus only 4% of those receiving chlorpromazine plus benztropine ($p < .001$). Based on this trial, the FDA approved clozapine for limited use in the United States.

The NIMH CATIE study compared clozapine (open label) to double-blind risperidone, olanzapine, or quetiapine in patients who had their phase I medication changed for lack of efficacy. Patients assigned to clozapine had the best outcome, with 56% of clozapine patients being discontinued compared to 71% on olanzapine, 86% on risperidone, and 93% on quetiapine. These results were also supported by a study from the United Kingdom by Shon Lewis that compared clozapine to an SGA selected prior to the randomization and found an advantage for clozapine.

The effectiveness of clozapine in treatment-refractory schizophrenia has also been compared to the use of antipsychotic polypharmacy, which is a nonrecommended strategy that is often used in this population. In this study, the utilization of emergency department and hospital services, as well as cost outcomes, was compared between Medicaid beneficiaries receiving clozapine monotherapy and those receiving two or more antipsychotics. It was found that Medicaid beneficiaries receiving clozapine monotherapy between 2006 and 2009 ($N = 479$) had a lower risk for emergency department use as well as lower Medicaid payments (both all cause and mental disorder related) than patients receiving SGA polypharmacy.

Altogether, these studies support the view that patients with schizophrenia who continue to demonstrate psychotic symptoms that are bothersome should not be considered refractory to treatment unless they have had an adequate trial with clozapine. An adequate trial is probably at least 3 months of clozapine at an adequate dose.

LONG-TERM EFFECTIVENESS. Long-term studies have supported the concept of maintaining patients on clozapine. A 29-week study reported in 2001 found that 60% of moderately refractory outpatients with schizophrenia treated with clozapine met improvement criteria compared with 12% of haloperidol-treated patients. In a study of patients with schizophrenia in Connecticut state hospitals, those who were discharged on clozapine treatment were more likely to remain out of the hospital when compared with individuals discharged on treatment with other agents. A Department of Veterans Affairs Cooperative Study comparing 1 year of treatment with clozapine to that of haloperidol also found that clozapine was associated with a decreased need for hospitalization. Moreover, patients treated with clozapine demonstrated a greater likelihood of participating in psychosocial treatments and rehabilitation programs. As mentioned in an earlier section, there is evidence that patients who are unresponsive to treatment in adequate trials of conventional dopamine receptor antagonists may respond when treated with SGAs other than clozapine. As a result, patients who respond poorly to conventional dopamine receptor antagonists may benefit from a trial of at least one of these agents before receiving a clozapine trial. Clinicians should also consider that prolonged sequential trials with a number of antipsychotics may take months or even years, thus delaying treatment with an agent that may be effective.

SEVERE TARDIVE DYSKINESIA. Patients with severe tardive dyskinesia should also be considered for a clozapine trial. Most clinical reports indicate that clozapine can suppress abnormal movements in tardive dyskinesia. This is not surprising, as conventional antipsychotic agents are effective for the short-term suppression of tardive dyskinesia. However, clozapine is probably less likely to cause the disorder and, therefore, is promising for the long-term management of patients, particularly those who already have tardive dyskinesia. Although rare cases of tardive dyskinesia have been found in patients who receive clozapine, the risk of developing this disorder is substantially reduced. Studies using animal models indicate that clozapine does not result in the increased sensitivity of striatal D_2 receptors that are associated with typical antipsychotic agents and has been theorized to result in tardive dyskinesia. It remains to be established whether a positive effect on tardive dyskinesia results from a direct suppressive effect of clozapine on certain transmitter systems or whether clozapine spares the striatal D_2 receptors, allowing spontaneous remission of tardive dyskinesia. The former interpretation is supported by the reports of a return of tardive dyskinetic movements after clozapine withdrawal.

CLOZAPINE IN HOSTILE AND AGGRESSIVE PATIENTS. Clozapine has also been studied in patients with schizophrenia and other psychotic disorders who display hostile and aggressive behaviors. A study of 157 inpatients by Citrome found that clozapine treatment led to greater reductions in ratings of hostility than other antipsychotics. Similarly, Chengappa found that clozapine led to a reduction in the need for seclusion and restraint. A recent study of physically assaultive patients with schizophrenia found that clozapine was more effective at reducing assaults than olanzapine and haloperidol. A within-subject study of a large population in Sweden found that clozapine was associated with a lower rate of violent offending than olanzapine. These findings suggest that clozapine may be of particular benefit to patients who demonstrate hostile and aggressive behaviors.

PATIENTS AT RISK FOR SUICIDE. Clozapine may also be a preferred agent for patients with schizophrenia who are at a higher risk of suicide. Large epidemiologic studies found what appeared to be a reduction in death from suicide in patients receiving clozapine. This and other observations led Herbert Meltzer and colleagues to design the International Suicide Prevention trial (or InterSePT). The study compared clozapine and olanzapine in 956 patients with schizophrenia or schizoaffective disorder who were considered at risk for suicide. Clozapine was more effective in delaying the time of recurrent suicidal behavior. In addition, over a 2-year period, 24% of clozapine and 32% of olanzapine patients were hospitalized for suicidal risk.

SCHIZOPHRENIA WITH SUBSTANCE ABUSE. Although the effectiveness of clozapine in patients with comorbid substance abuse has not been demonstrated by randomized controlled trials, there is some supporting evidence from naturalistic studies. A retrospective study by Green and coworkers found that patients treated with clozapine were more likely to abstain from alcohol and cannabis use than patients treated with risperidone. A prospective study by the same investigators found that patients treated with clozapine were better able to reduce their substance abuse.

TREATMENT-RESISTANT MANIA. Clozapine can be an effective treatment of manic symptoms that are not responsive to other agents. Early clinical case reports indicated that clozapine was effective both as monotherapy and as an adjunct to mood-stabilizing agents in this population. An open-label, randomized trial in acutely manic patients found that clozapine had a more rapid onset of effect than chlorpromazine. In another open-label trial involving patients with schizoaffective disorder or bipolar disorder who were randomized to either clozapine or treatment, as usual, clozapine-treated patients demonstrated greater improvements. Although the available data support clozapine's effectiveness in mania, none of these trials have used a double-blind design.

NEUROLOGIC ILLNESSES. Clozapine has also found uses in patients with neurologic disorders. It has been used in idiopathic Parkinson disease as a treatment of psychotic symptoms that are secondary to dopaminergic agents that are used to treat the disease. These patients may respond to relatively low dosages of clozapine (doses of 25 to 75 mg). Higher dosages (e.g., 100 to 250 mg a day) may exacerbate the parkinsonian symptoms. The effectiveness of clozapine in patients with idiopathic Parkinson disease is supported by at least two double-blind studies. A report from the American Academy of Neurology recommends clozapine as a preferred agent for psychosis in Parkinson disease. Patients with benign essential tremor or with Huntington disease have also responded positively to clozapine.

POLYDIPSIA–HYPONATREMIA SYNDROME. Case reports suggest that clozapine may have advantages over other antipsychotic agents in patients who tend to intoxicate themselves with water. This syndrome—called *polydipsia–hyponatremia syndrome*—can result in seizures.

Effects on Specific Organs and Systems

GENERAL. Sedation is the most common adverse effect of clozapine treatment. It occurs early in treatment, and tolerance usually develops after the first few days or weeks. Giving the larger part of the daily dose at bedtime may reduce daytime sedation. However, many individuals are still unable to tolerate the sedative effects and should, therefore, be treated with another antipsychotic. Dizziness or vertigo may be related to orthostatic hypotension.

DISTURBANCES IN TEMPERATURE REGULATION. Benign hyperthermia is occasionally observed during the first 3 weeks of clozapine treatment, with a peak incidence on the 10th day. The increase in temperature is usually not more than 1 °F to 2 °F and resolves over a few days. Occasionally, however, temperature readings above 101 °F have been seen, which require temporary interruption of clozapine treatment and frequent hematologic monitoring. Differential diagnoses include drug fever, intercurrent infection, infection secondary to agranulocytosis, dehydration, heat stroke, lethal catatonia, and, possibly, NMS. Patients who develop a high temperature can be given the drug again but with a more gradual increase in dosage. If this strategy does not work, clozapine may need to be discontinued. As is also seen with chlorpromazine, mild hypothermia is frequently observed (87%) after clozapine treatment. This common effect may indicate the poikilotherm effect of antipsychotic agents, presumably mediated by the hypothalamus.

CARDIOVASCULAR. The cardiovascular side effects most frequently observed with clozapine are tachycardia and postural hypotension. The tachycardia is probably a direct effect of the vagolytic properties of the drug. It can be present in the supine position and is thus not the result of orthostatic changes. Increased pulse rates of 20 to 25 beats per minute are encountered when clozapine reaches 300 mg per day or more over a 7-day period (e.g., sinus tachycardia). Tachycardia related to clozapine may be treated with a β-blocker such as atenolol (Tenormin), although patients should then be monitored for hypotension. Propranolol (Inderal) should not be prescribed because it may increase the risk of agranulocytosis.

Reversible nonspecific ST-T wave changes, T-wave flattening, or inversions (repolarization effects) are seen infrequently with clozapine and are usually of no clinical significance. These changes are similar to those seen with other antipsychotic agents and are dose dependent. Clozapine can increase the QT interval, although there are very limited data to determine if these increases are associated with a risk of ventricular arrhythmias. Orthostatic hypotension is more likely to occur during the initial titration of clozapine in association with rapid dose escalation but may occur after the first dose. In one report, initial dosages of clozapine as low as 12.5 mg were associated with orthostatic collapse and respiratory arrest. The risk of postural hypotension progressing to orthostatic collapse is substantially greater if the initial dosage of clozapine exceeds 75 mg per day. Tolerance to the hypotensive effects of clozapine often develops over time, and the effects can usually be managed if the initial dosage is low (e.g., 25 mg per day) and is increased gradually. When hypotension occurs, lowering the dosage to the previous level and slowing the rate of upward titration usually limits problems. Support stockings, increased sodium intake, or fludrocortisone (Florinef) may be helpful. Rarely (approximately 1 case per 3,000 or 4,000 patients), orthostatic collapse can be profound, accompanied by respiratory or cardiac arrest.

Clozapine has been associated with idiopathic myocarditis. The section on warnings and precautions below will discuss this further.

GASTROINTESTINAL. Constipation, most probably due to the antimuscarinic effects of clozapine, is the most common gastrointestinal adverse effect and, if left untreated, can progress to intestinal obstruction. Clinicians treating patients with clozapine should monitor the presence of constipation and intervene proactively at the earliest sign that this symptom has developed. Interventions may include modifications in the patient’s diet, use of stool softeners, increased fluid intake, and exercise.

Nausea and vomiting may occur during the initial stages of clozapine treatment. Patients who experience the side effects can often, but not always, be managed by dosage reduction.

UROGENITAL. Urogenital effects include enuresis, frequency and urgency, hesitancy, urinary retention, and impotence. All of these are independent of dose or blood concentrations of clozapine. Urinary incontinence or enuresis can be managed by setting an alarm clock to ring at night or administering intranasal desmopressin (DDAVP) or oxybutynin (Ditropan) (5 to 15 mg per day). So far, a few cases of clozapine treatment-associated priapism have been noted. Because patients may not volunteer these complaints, clinicians should ask about them.

Therapeutic Indications

TREATMENT-RESISTANT SCHIZOPHRENIA. Clozapine is indicated for patients who have failed trials with other antipsychotics. Given the serious side effects associated with clozapine treatment, including seizures and agranulocytosis, most guidelines recommend a trial of a minimum of two other antipsychotics, one of which should be a Second Generation Antipsychotic (SGA). Studies have found that clozapine's advantage, when compared with other antipsychotics, extends from patients who continue to experience severe symptoms to those who have moderate symptoms.

REDUCTION IN THE RISK OF RECURRENT SUICIDAL BEHAVIOR IN SCHIZOPHRENIA OR SCHIZOAFFECTIVE DISORDER. The InterSePT study, as well as epidemiologic data, indicates that clozapine is indicated for those patients who are at risk of suicidal behaviors. It is the only antipsychotic with this particular indication.

Precautions and Adverse Reactions. The only contraindications to the use of clozapine are a history of hypersensitivity to clozapine or tablet components, a WBC count below 3,500 cells per mm^3, a previous bone marrow disorder, a history of agranulocytosis during clozapine treatment, or the concomitant use of another bone marrow suppressant drug such as carbamazepine (Tegretol). The history of myocarditis is not a labeled contraindication, but the authors recommend against the use of clozapine in these patients. Although the mechanism of action is unknown, there is a slight risk of respiratory depression or collapse if treatment is initiated while patients are taking benzodiazepines.

BLACK BOX WARNINGS

Agranulocytosis. The risk of clozapine-induced agranulocytosis is 0.73% during the first year of treatment and 0.07% in the second year. A review of the clozapine registry indicates that the incidence may be as low as 0.038%. The risk of this adverse effect is greatest during the first 3 months of treatment, but a significant risk remains during the first year of treatment. For neutropenia, the risk is 2.32% in the first year and 0.69% in the second year.

Increased familiarity of clinicians with agranulocytosis (absolute neutrophil count [ANC] less than 500 cells per mm^3 or WBC less than 1,000 cells per mm^3) decreased the incidence of agranulocytosis by more than 50% from before 1995 to after 1997.

Agranulocytosis due to clozapine is a potentially fatal condition that requires immediate medical attention. The risk of agranulocytosis increases with age and is higher in women. Therefore, before treating patients with clozapine, clinicians must register patients with a clozapine monitoring system. The United Kingdom and Australia require, in addition to the 3,500 WBC per mm^3, that granulocyte count (ANC) is 2,000 cells per mm^3 or higher when clozapine is initiated.

Patients who are treated with clozapine in the United States must receive the following monitoring protocol:

- WBC is no longer required to be monitored—only ANC.
- Initial ANC must be ≥1,500/mm^3, and in patients with benign ethnic neutropenia, it must be ≥1,000/mm^3.
- Monitoring is weekly for the first 6 months, every 2 weeks for the second 6 months, and monthly after 12 months. The threshold minimum ANC that is not classified as neutropenia is now 1,500/mm^3 rather than 1,000/m^3.

The requirement for ANC monitoring is continued for as long as patients receive the drug and for at least 1 month after it is discontinued. Current guidelines specify that any fever or sign of infection (e.g., pharyngitis) is an immediate indication for a WBC count, particularly in the first 18 weeks of treatment. The revised requirements indicate that clozapine treatment should be interrupted for ANC <1,000/mm^3. Reports of clozapine re-exposure after hematologic recovery have shown that all such patients again experienced agranulocytosis, but sooner and at lower doses than during initial treatment. As a result, patients who develop agranulocytosis should not receive clozapine again. Although a WBC count below 3,500 cells per mm^3 is not a proven risk factor, clozapine should be administered cautiously in patients with such low counts because of the difficulty in the early identification of agranulocytosis. Rechallenge is not dictated by WBC any longer, only ANC. Rechallenge is allowed following recovery from moderate neutropenia (500 to 999/mm^3) but not after recovery from severe neutropenia (<500/mm^3) unless the prescriber determines benefits outweigh risks. Clozapine therapy should be discontinued whenever the total WBC count falls below 3,000 cells per mm^3, or the granulocyte count falls below 1,500 cells per mm^3.

Other hematologic changes have been reported with clozapine, including leukocytosis (0.6%), eosinophilia (1%), leukopenia, neutropenia, decreased WBC count (3%), and, rarely, thrombocytopenia. If the total eosinophil count rises above 4,000, clozapine should be discontinued until the count falls below 3,000.

Orthostatic Hypotension. On rare occasions, clozapine can cause severe episodes with orthostatic hypotension and possible syncope. This has been associated with cardiac or respiratory arrest. These serious events are more likely to occur early in treatment when the dose is increased. Also, these events have been associated with benzodiazepine use and suggest that combining clozapine and benzodiazepines early in treatment should be done cautiously. Dosage should be increased slowly, with divided doses, and clozapine should be used with caution in patients with cardiovascular or cerebrovascular disease or conditions predisposing them to hypotension.

Seizures. Five percent to 10% of patients treated with clozapine will develop seizures. The risk is clearly related to dosage, with an increased risk at dosages greater than 600 mg. The rapid escalation of dosage during the first weeks of clozapine treatment may increase the likelihood of seizures. A pre-existing seizure disorder or head trauma places the patient at even greater risk.

A number of strategies may be effective for managing the risk of seizures. Monitoring clozapine plasma concentrations in patients receiving higher dosages may ensure that concentrations do not become excessive. Patients with a history of head injury should have baseline electroencephalographs (EEG) performed before the initiation of clozapine. Clinicians may also consider an EEG before increasing clozapine dosages above 600 mg. In addition, when patients receive higher doses or have abnormal EEGs, adding an anticonvulsant such as sodium valproate (Depakene) may be helpful.

The occurrence of seizures in patients treated with clozapine should not be viewed as a clear reason to stop clozapine. Lowering the dosage and adding an anticonvulsant is usually sufficient to prevent further seizures. Sodium valproate, gabapentin (Neurontin), and topiramate (Topamax) have been effective in treating clozapine-induced seizures. Carbamazepine should not be prescribed because it can also cause neutropenia or granulocytopenia. Once patients have begun anticonvulsant therapy, the dosage of clozapine can be retitrated to an optimal level.

Myocarditis and Cardiomyopathy. Clozapine is associated with an increased risk of myocarditis. Although the incidence of myocarditis is very low (approximately 2.8/100,000 patient-years), it can result in death. Clinicians should consider the possibility of myocarditis when patients receiving clozapine present with unexplained chest pain, tachycardia, palpitations, dyspnea, fever, flu-like symptoms, or hypotension. Unfortunately, some of these symptoms are common in individuals who are in the early stages of clozapine treatment, and 80% of clozapine-induced cases of myocarditis occur during the first 6 weeks of treatment. Because mortality from myocarditis approaches 40%, early detection and treatment are important. If this syndrome is suspected, clozapine should be discontinued, and evaluations including an ECG, creatine kinase level, and erythrocyte sedimentation rate should be performed.

Increased Mortality in Older Adult Patients with Dementia. Clozapine and the other SGAs carry a black box warning for increased mortality in older adult patients with dementia-related psychosis compared with placebo.

OTHER ADVERSE EVENTS

Sialorrhea. Sialorrhea is an uncomfortable adverse effect that develops early in treatment. Hypersalivation is most profound during sleep, and patients may complain that their pillow is wet in the morning. Placing a towel over the pillow may be of some help. Clonidine (Catapres) (0.1- or

0.2-mg patch once a week), amitriptyline (Elavil, Endep), glycopyrrolate, and atropine sublingual drops at bedtime have been used to treat sialorrhea. Clinical reports also support the injection of botulinum toxin into the salivary glands. Studies suggest that clozapine does not increase saliva flow but decreases swallowing, which becomes more pronounced during sleep.

Peripheral Anticholinergic Effects. Dry mouth, blurred vision, constipation, and urinary retention are commonly observed with clozapine. Constipation can be severe, particularly when patients fail to exercise or drink sufficient fluids. Other side effects of the autonomic nervous system, such as increased sweating, have also been reported.

Weight Gain. Clozapine can be associated with substantial weight gain; a meta-analysis of short-term trials with clozapine found an average weight gain of 4.45 kg over 10 weeks. This exceeded the weight gain observed with all of the other medications evaluated in the trials. Other studies indicate that patients can continue to gain weight over a period of 5 years. The amount of weight that patients gain can be substantial and can result in some patients becoming obese. Younger patients, including adolescents, may be more vulnerable to gaining weight on clozapine.

These observations demonstrate the importance of preventing substantial weight gain for patients receiving clozapine. Physicians should weigh patients frequently during the first months of clozapine treatment and should encourage patients to monitor their own weight. If weight gain is observed, it is important for clinicians to intervene with counseling about nutrition, weight, and exercise. If these interventions are unsuccessful, changing medication to an agent less likely to cause weight gain should be considered.

Diabetes Mellitus. Clozapine, along with many other SGAs, can increase the risk of type 2 diabetes mellitus. In a naturalistic study by David Henderson that monitored 81 patients over a 5-year period, 36.6% developed diabetes. Case reports suggest that new cases of diabetes usually occur during the first 6 months of clozapine treatment. Michael Sernyak and his coworkers examined the records of more than 30,000 veterans and found that the prevalence of diabetes was higher in patients receiving clozapine, olanzapine, or quetiapine. It is also concerning that a number of case reports indicate that patients receiving clozapine developed diabetic ketoacidosis. Although these forms of evidence remain inconclusive, they strongly suggest that clinicians should be aware of the possible association of clozapine with an increased risk of diabetes. This may lead clinicians to monitor patients for diabetes, with fasting blood glucose determinations at regular intervals. It may also be helpful to inquire if patients are experiencing clinical signs of diabetes, such as polydipsia, polyuria, or weight loss.

Extrapyramidal Side Effects. Clozapine has a much lower incidence of extrapyramidal side effects than other antipsychotics. Akathisia (6%), tremors (6%), and rigidity (3%) are considerably less frequent with clozapine than with First Generation Antipsychotics (FGA). The authors are not aware of reports of acute dystonia with clozapine. Case reports suggest that tardive dyskinesia may improve with clozapine.

Neuroleptic Malignant Syndrome. Consistent with clozapine's low affinity for D_2 receptors, there have been few reports of NMS developing in patients who were treated solely with clozapine, although case reports of this syndrome developing in patients when concomitantly treated with lithium (Eskalith) exist. Although two patients had experienced NMS with conventional antipsychotics before, one case has been reported of a patient with a history of NMS that was successfully treated with clozapine. The authors recommend, therefore, that lithium should not be used as adjunctive treatment with clozapine in patients who have previously experienced NMS.

Because clozapine treatment is accompanied by side effects that resemble several signs and symptoms of NMS (e.g., hyperthermia, cardiovascular effects, delirium, increased sweating, elevations in creatine phosphokinase, and leukocytosis), clinicians should be alert for the development of a potential NMS during clozapine treatment. Some case reports have described NMS without rigidity in patients treated with clozapine, although the lack of rigidity could preclude the diagnosis of this syndrome.

Pregnancy and Lactation. Clozapine is rated as pregnancy category B. Studies have not revealed evidence that clozapine causes harm to the fetus. However, there are no adequate and well-controlled studies on pregnant women. Studies in animals suggest that clozapine is excreted in breast milk. As a result, women taking clozapine should not breastfeed.

Discontinuing Clozapine. If a decision is made to change a patient to a different antipsychotic, it should not be made abruptly unless patients have developed agranulocytosis. In some individuals, the sudden discontinuation of clozapine has led to a severe exacerbation of their underlying illness. In addition, withdrawal of a potent antimuscarinic agent such as clozapine can lead to a cholinergic rebound, with symptoms such as nausea and vomiting. For these reasons, the authors recommend that patients are gradually cross-tapered from clozapine to another antipsychotic agent over 2 to 4 weeks.

Drug Interactions. Because the cytochrome P450 (CYP) mitochondrial system is involved in the metabolism of clozapine, ethnic differences and interactions with drugs that inhibit or stimulate this system will affect clozapine plasma concentrations. For instance, caffeine *N*-3-demethylation, which is a measure of CYP isoenzyme 1A2 (CYP1A2) activity, accounts for 70% of the variance in clozapine clearance. This suggests that administration of fluvoxamine (Luvox), a potent inhibitor of CYP1A2, will elevate clozapine concentrations, whereas fluoxetine (Prozac) administration may only lead to changes in higher dosage ranges. Cimetidine (Tagamet), selective serotonin reuptake inhibitors (SSRIs), tricyclic drugs, and valproic acid decrease clearance, presumably through inhibition of CYP1A2 and CYP2D6, whereas phenytoin (Dilantin) and carbamazepine induce CYP2C and CYP3A4 hepatic isoenzymes and may decrease concentrations of clozapine.

Tobacco smoking induces the activity of CYP1A2, which reduces clozapine plasma concentrations by about one-third. When patients stop smoking, there can be a substantial increase in plasma levels resulting in increased side effects. This suggests that clozapine plasma levels should be checked whenever there is a change in smoking behavior.

Dosage and Administration

INITIATING TREATMENT. Clozapine treatment should begin with 12.5 to 25 mg per day and increased to 25 or 50 mg per day on the second day. The daily dose, which should be administered at least twice daily initially, can usually be increased by 25 mg every 1 or 2 days until a target dose of 300 to 450 mg per day is reached. The rate of increase is often limited by sedation and orthostatic hypotension. Because the titration phase for clozapine may require several weeks in some individuals, these patients should be continued on their current antipsychotic agent, which should be gradually decreased as the clozapine dosage increases. Patients who fail to respond in the 300 to 450 mg per day range and who do not have serious side effects can be increased gradually to doses as high as 900 mg per day. However, clinicians and patients should be aware of the risk of seizures at higher levels and when doses are escalated more rapidly.

DURATION OF TREATMENT. Once the appropriate dosage has been reached, and the patient improves, the medication should be continued. There is evidence that patients who receive clozapine will continue to improve over a 3- to 6-month period. If there are concerns that patients are experiencing excessive side effects or an inadequate response, measurement of plasma concentration may be helpful. Levels above 350 ng/mL are an indication that patients are receiving adequate dosages of clozapine.

Risperidone and Paliperidone

Risperidone was the first serotonin-dopamine antagonist (SDA) medication to be approved following clozapine. Whereas clozapine was reserved for patients with illnesses that were poorly responsive to available antipsychotic drugs, risperidone was a first-line antipsychotic that could be administered to nearly every patient with a psychotic illness. Following its introduction, it was hoped that risperidone would share clozapine's efficacy for a broad range of psychopathology in schizophrenia and that it would be more effective than the FGAs. Although this promise has not been met, risperidone has been widely accepted and is among the most prescribed antipsychotics worldwide.

Paliperidone is the major active metabolite of risperidone and shares risperidone's receptor profile. Paliperidone extended-release tablets use a pressure-based osmotic release oral system (OROS) delivery system to provide continual drug delivery for over 24 hours (see below). Because paliperidone extended release is gradually released into the plasma without peaks in concentration, it permits treatment initiation without titration. As the primary metabolite of risperidone, paliperidone does not undergo further significant hepatic metabolism. Therefore, it suffers from fewer potential drug–drug interactions.

Chemistry. Risperidone and paliperidone are benzisoxazole derivatives. Paliperidone is the major metabolite of risperidone.

Pharmacologic Action

PHARMACOKINETICS

Absorption and Distribution. Risperidone has a bioavailability of 70%, and studies show that all oral forms of risperidone are bioequivalent. Risperidone is metabolized in the liver to 9-hydroxyrisperidone (or paliperidone [Invega]), which generally has the same pharmacologic profile as the parent compound. After ingestion, peak plasma levels of the parent compound occur within 1 hour and within 3 hours for 9-OH-risperidone. Steady state is expected to be reached within 5 days (with a range of 1 to 5 days, depending on the CYP2D6 activity in the individual patient). Food does not affect the rate or extent of absorption in the gut.

The volume of distribution is 1 to 2 L/kg. Risperidone is 90% plasma protein bound, and its metabolite is 77% protein bound. However, there is only a slight increase in the free fraction of risperidone and paliperidone with high therapeutic doses of warfarin (Coumadin) and other highly protein-bound medications.

Peak plasma concentration of paliperidone extended release is reached approximately 24 hours after a single dose. The half-life of paliperidone is about 23 hours, which allows for once-a-day dosing. Absorption is increased by approximately 60% when paliperidone extended release is taken with food.

Metabolism and Elimination. Risperidone undergoes extensive metabolism, primarily by CYP2D6 (see the section on drug interactions). The major metabolite is 9-OH-risperidone (or paliperidone), which has similar pharmacologic activity as risperidone. The half-life of risperidone is 3 hours in rapid metabolizers and 20 hours in poor metabolizers. The half-life of paliperidone is 21 hours in rapid metabolizers and 30 hours in poor metabolizers. The combined half-life of risperidone and its metabolites has a mean of 20 hours. Seventy percent of risperidone is excreted in the urine.

PHARMACODYNAMICS

Mechanism of Action. Risperidone and paliperidone share similar receptor affinities. Both have a high affinity for dopamine D_2 receptors (Table 33.2b–2) and for serotonin 5-HT_{2A} receptors, and potent antagonism at these receptors is thought to mediate their therapeutic effects. They also demonstrate a high affinity for adrenergic α_1- and α_2-receptors and histaminergic H_1 receptors. Although risperidone and paliperidone have a high affinity for D_2 receptors, they have lower rates of EPS than the high-potency FGA medications. This is most likely due to the dopamine-promoting effects of 5-HT_{2A} antagonism. Risperidone blocks 65% of D_2 receptors (the lowest threshold percentage for antipsychotic efficacy) at an average dose of 2 mg per day. At an average of 6 mg per day, 80% of D_2 receptors are blocked, and EPS may occur. At the 2-mg dose, 5-HT_{2A} effects may not be optimal.

Blood Concentrations and Clinical Activity. The clinical activity of risperidone results from the combination of risperidone and 9-OH-risperidone. A consensus statement on therapeutic drug monitoring considers both risperidone and paliperidone as drugs where there is only limited evidence linking plasma concentrations to clinical response. The guidelines indicate that there is a lower likelihood of response when levels of risperidone plus paliperidone are below 20 ng/mL, and adverse effects are more likely above 60 ng/mL.

Table 33.2b–2.
Side Effects Associated with Receptor Antagonism

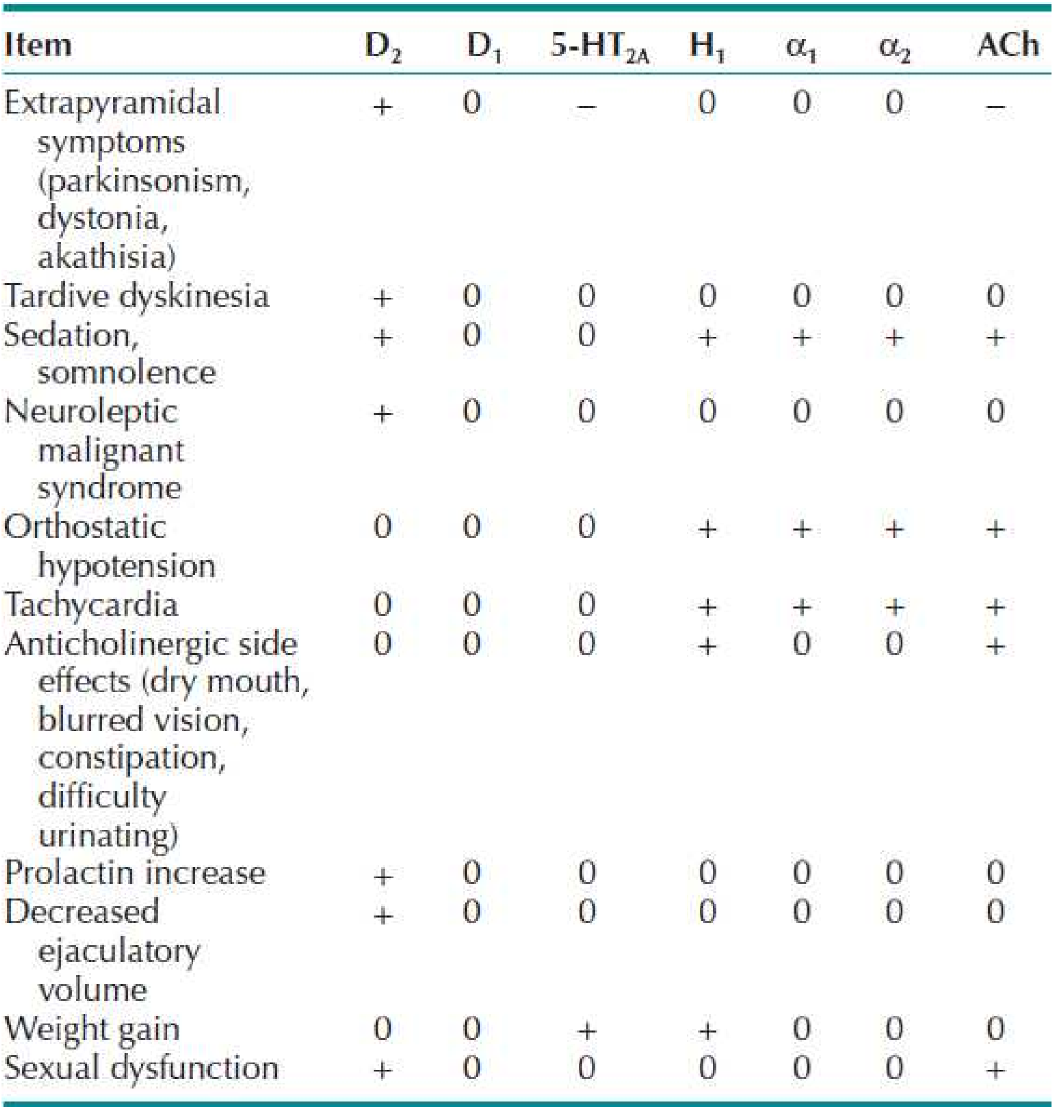

Item	D_2	D_1	5-HT_{2A}	H_1	α_1	α_2	ACh
Extrapyramidal symptoms (parkinsonism, dystonia, akathisia)	+	0	–	0	0	0	–
Tardive dyskinesia	+	0	0	0	0	0	0
Sedation, somnolence	+	0	0	+	+	+	+
Neuroleptic malignant syndrome	+	0	0	0	0	0	0
Orthostatic hypotension	0	0	0	+	+	+	+
Tachycardia	0	0	0	+	+	+	+
Anticholinergic side effects (dry mouth, blurred vision, constipation, difficulty urinating)	0	0	0	+	0	0	+
Prolactin increase	+	0	0	0	0	0	0
Decreased ejaculatory volume	+	0	0	0	0	0	0
Weight gain	0	0	+	+	0	0	0
Sexual dysfunction	+	0	0	0	0	0	+

Design and Interpretation of Clinical Drug Studies

DESIGN. Since risperidone and paliperidone became available, there have been numerous studies evaluating their efficacy, safety, and tolerability. In this section, the focus will mostly be on randomized double-blind, placebo- or active-controlled trials.

Schizophrenia and Psychosis: Acute Treatment. Multiple studies with brief treatment trials have confirmed risperidone's efficacy in treating the acute phase of psychosis associated with schizophrenia, schizoaffective disorder, schizophreniform disorder, and acute psychotic disorder. Risperidone has consistently been shown to be superior to placebo, superior or equivalent to FGAs such as haloperidol and perphenazine, and equivalent to other SGAs with the exception of clozapine. Early multicenter trials in the United States and Canada compared risperidone 2, 6, 12, or 16 mg, haloperidol 20 mg, or placebo in a combined total of 513 patients with schizophrenia. The results from the U.S. and Canadian studies were very similar. In the Canadian study, risperidone 2 mg per day was not shown to be superior to placebo on measures of efficacy. Risperidone was superior to placebo at all higher doses and superior to haloperidol on a variety of measures, most noticeably the PANSS. Risperidone 6 mg was as efficacious as the higher doses. In the U.S. sample, there was no significant difference in the rates of EPS between patients on risperidone 6 mg and placebo. There was a linear trend toward more EPS as risperidone doses increased. Another large multicenter, multinational study of 1,362 patients with schizophrenia compared risperidone 1, 4, 8, and 12 mg per day to haloperidol 10 mg for 8 weeks. At doses of 4 and 8 mg, the response rates were 63% and 66%, respectively. There was no significant difference between the response to haloperidol (response rate 59%) and the response to risperidone. Risperidone was associated with lower EPS side effects at all doses compared with haloperidol.

Taken together, these studies indicate that risperidone at doses of 4 to 8 mg has similar effectiveness to other SDAs. This is consistent with the product label. Using group data, the findings suggest that there is little advantage to increasing the dose above 4 to 6 mg. This observation, and the tendency of risperidone to be associated with dose-dependent EPS, has led many clinicians to target risperidone in the 3- to 5-mg range. A 2019 meta-analysis found that 6.3 mg of risperidone was the dose that was 95% effective for symptom reduction (ED95).

Three 6-week randomized, double-blind, placebo-controlled efficacy and safety studies were conducted using fixed doses of paliperidone extended release. Kane and colleagues administered a placebo, 6, 9, or 12 mg of paliperidone extended release to 628 patients with schizophrenia. Davidson and colleagues treated 618 subjects with schizophrenia with placebo or 3, 9, or 15 mg of paliperidone extended release. Marder and colleagues evaluated 444 patients administered placebo, 6, or 12 mg of paliperidone extended release. In all three studies, the PANSS total score decreased significantly in all paliperidone arms compared with placebo. In the study by Kane et al., there was a trend toward greater improvement in the 12-mg paliperidone group. Davidson's group reported that as early as day 4, the three paliperidone arms had separated significantly from the placebo on PANSS total score. The study by Marder and colleagues found significant improvement by day 4 for the 6-mg arm and by day 15 for the 12-mg arm. In all three studies, the three paliperidone arms showed significant or trend improvement from baseline compared with placebo on positive, negative, disorganized, hostility or agitation, and anxiety or depression factor scores. The above studies support the product label recommendation of 6 mg as the target dose for schizophrenia. There is evidence of increased efficacy at higher doses. The 2019 meta-analysis found an ED95 of 13.4 mg for paliperidone.

Schizophrenia and Psychosis: Maintenance Treatment. The effectiveness of risperidone for relapse prevention in schizophrenia was demonstrated in a double-blind comparison with haloperidol. Relapse rates were significantly lower in patients randomized to risperidone.

Kramer and colleagues treated 205 patients in a study with an 8-week open-label run-in phase, followed by a 6-week open-label stabilization period before entering a variable duration double-blind treatment phase where patients were randomized to either placebo or paliperidone 3 to 15 mg daily. The primary efficacy variable was time to the first recurrence of schizophrenia symptoms (e.g., hospitalization, changes in symptom scores, aggressive, self-injurious, or suicidal behavior). Time to recurrence was significantly longer in the paliperidone group; in the final analysis, 52% of patients on placebo had a recurrence versus 22% on paliperidone. Twenty-five percent of patients on placebo had a recurrence event by day 23, whereas 25% of paliperidone-treated patients had a recurrence event by day 83.

Schizophrenia and Psychosis: Adolescents. Two trials evaluating the efficacy of risperidone in adolescents aged 13 to 17 years with schizophrenia compared risperidone at doses ranging from 0.15 to 6 mg per day with a placebo. Risperidone treatment significantly improved PANSS total scores. Doses below 0.6 mg per day were less effective than higher doses. Doses above 3 mg per day were not more efficacious and resulted in a greater side-effect burden.

Bipolar Mania: Adults. In two 3-week studies comparing risperidone treatment with placebo in patients with acute mania, risperidone treatment was associated with significantly greater improvements on the YMRS versus placebo. The time to onset of improvement was as brief as 3 days in the first study. A recent meta-analysis by Scherk et al. was conducted on the effects of SGAs in the treatment of mania. It concluded that risperidone was significantly more efficacious than placebo in the treatment of mania, as efficacious as haloperidol, and more effective when used in combination with a mood stabilizer than was a mood stabilizer alone. A small study of 65 patients compared risperidone with haloperidol and lithium and found all medications were similarly efficacious.

Bipolar Mania: Children and Adolescents. An 8-week open-label trial of 30 children and adolescents (aged 6 to 17 years) found that risperidone monotherapy (mean dose 1.25 mg per day) resulted in clinically significant improvements in the YMRS scores, although at the end point, there were mild residual symptoms present. A double-blind trial of children and adolescents aged 10 to 17 years compared risperidone 0.5 to 2.5 mg per day, risperidone 3 to 6 mg per day, and placebo in the treatment of manic or mixed episodes of bipolar I disorder. Risperidone in both dosage groups significantly improved YMRS scores compared to placebo. Doses above 2.5 mg per day were not more efficacious than lower doses.

Irritability Associated with Autism: Children and Adolescents. A study of 101 children and adolescents aged 5 to 17 years compared risperidone treatment (dose range 0.5 to 3.5 mg per day) with placebo on measures of irritability and global improvement. Sixty-nine percent of risperidone-treated patients versus 12% of placebo-treated patients were classified as responders. Hyperactivity and stereotypy scores also improved significantly in the risperidone group. Patients treated with risperidone gained more weight (mean = 2.7 kg) than did patients on placebo (mean = 0.8 kg) in the 8-week trial. EPS rates were similar, although parents reported a higher incidence of tremor in the risperidone group. Of the risperidone responders who followed in open-label treatment for the next 6 months, 68% maintained those gains. A continuation of the same study followed 63 of the original patient sample for 4 months on open-label risperidone treatment, followed by an 8-week placebo-controlled discontinuation protocol. During the 4-month risperidone continuation phase, 83% of the patients maintained the gains in irritability diminution they had achieved during the original 8-week trial. Overall, the mean irritability ratings increased from 9.5 to 10.8, an increase the authors concluded was clinically insignificant. Thirty-eight subjects were enrolled in the placebo-controlled discontinuation phase. The relapse rate was 62.5% in the placebo group versus 12.5% in the risperidone group. A randomized, double-blind, placebo-controlled trial of 40 children with autism (aged 2 to 9 years) found that risperidone 1 mg per day produced improvements of 20% or greater on the Childhood Autism Rating Scale and the Children's Global Assessment Scale in 63% and 89%, respectively, versus 0% and 10% of children treated with placebo. Risperidone was also associated with improvements in social responsiveness, nonverbal communication, hyperactivity, and aggression.

Effects on Specific Organs and Systems

CARDIOVASCULAR SYSTEM. Tachycardia and palpitations are the most commonly reported cardiovascular side effects of risperidone and paliperidone, followed by hypotension and postural hypotension. This is most likely related to α-receptor blockade. Palpitations are noted to be dose related. There have been rare reports of atrioventricular and bundle branch block.

Risperidone and paliperidone are more strongly associated with an increase in prolactin levels than other antipsychotics. These elevations can result in gynecomastia, galactorrhea, menstrual disturbances, and sexual dysfunction.

Therapeutic Indications. Risperidone is indicated for schizophrenia in adults and adolescents and for maintenance therapy. It is also indicated for bipolar mania and irritability associated with autistic disorder.

Paliperidone is indicated for schizophrenia and schizoaffective disorder.

Precautions and Adverse Reactions

BLACK BOX WARNING. Risperidone, paliperidone, and the other SGAs carry a black box warning for increased mortality in older adult patients with dementia-related psychosis compared with placebo.

QT PROLONGATION. Both risperidone and paliperidone have only modest effects on the QT interval.

NEUROLEPTIC MALIGNANT SYNDROME. Risperidone and paliperidone have been associated with NMS, possibly at rates higher than most other antipsychotic medications. NMS is a serious and potentially life-threatening condition. Patients with suspected cases of NMS should be immediately discontinued from the antipsychotic agent and aggressively treated with supportive measures, symptomatic treatment, and medical monitoring. If the patient requires antipsychotic treatment after recovery from NMS, close monitoring and gradual reintroduction are essential.

CEREBROVASCULAR ADVERSE EVENTS. In placebo-controlled trials, older adult patients with dementia-related psychosis treated with risperidone had significantly higher incidences of cerebrovascular accidents and other related adverse events than did patients treated with placebo.

TARDIVE DYSKINESIA. All antipsychotic medications (with the possible exceptions of clozapine and quetiapine) carry a risk of tardive dyskinesia. Tardive dyskinesia risk is thought to be related to a combination of factors, the most important of which is unopposed D_2 blockade in the basal ganglia. Risperidone has a high affinity for D_2 receptors, but its action at the 5-HT_{2A} receptors may partially mitigate some of the striatal D_2 blockade. In one trial of risperidone treatment, 5 of 453 patients developed symptoms of tardive dyskinesia, and 14 patients with pre-existing tardive dyskinesia improved. In another study of risperidone microspheres, the rate of tardive dyskinesia emergence in 530 patients after 50 weeks of treatment was 1.19%. A review article that considered the second-generation medications together reported a rate of tardive dyskinesia with risperidone, amisulpride (Solian), olanzapine, and quetiapine ranging from 0.5% to 1.5%. The rates of tardive dyskinesia emergence are slightly higher in older adult populations, especially among women. Among 133 older adults administered risperidone, the incidence of persistent tardive dyskinesia was 2.6%, and at doses of 0.75 to 1.5 mg per day, the incidence dropped to 1%.

HYPERCHOLESTEROLEMIA, HYPERGLYCEMIA, AND DIABETES MELLITUS. Risperidone and paliperidone are associated with weight gain, dyslipidemias as well as an increased risk of diabetes mellitus. Results from the CATIE study indicate that risperidone produced changes in glucose level comparable to perphenazine and quetiapine, with an increase lower than that with olanzapine treatment and higher than that with ziprasidone treatment.

CHANGES IN WEIGHT. In the CATIE study, patients on risperidone gained an average of 0.8 lb during treatment, and 14% gained more than 7% of their body weight. There was no significant difference between risperidone, quetiapine, perphenazine, and ziprasidone on measures of body weight change. The percentage of patients gaining over 7% of their body weight during the 6- to 8-week trials was 18% versus 9% for placebo.

MOVEMENT SIDE EFFECTS AND DISORDERS. Double-blind trials indicate that risperidone and paliperidone are associated with a dose-dependent increase in extrapyramidal side effects. Although many adult patients may not experience these effects at the recommended daily dosage of 6 mg, dystonic reactions have occurred with dosages between 4 and 16 mg per day, and akathisia has been described. Many patients require only 2 to 4 mg per day of risperidone. In all studies, paliperidone extended release was noted to be associated with higher rates of movement side effects in dosages above 9 mg.

ORTHOSTATIC HYPOTENSION AND SYNCOPE. Risperidone and paliperidone have antagonistic activity at α-receptors, which contributes to their risk of orthostatic hypotension, dizziness, and syncope. In phase II and III trials among adult patients with schizophrenia, 0.2% of patients reported syncope on risperidone. Limiting the initial dose to 2 mg per day, either in a single or divided dose, may minimize these side effects in healthy adults. In patients with hepatic or renal insufficiency or older adult patients, the initial dose should be reduced to 0.5 mg twice a day to reduce the potential for hypotension and syncope.

HYPERPROLACTINEMIA. Risperidone and paliperidone increase prolactin levels. In the CATIE study, risperidone increased prolactin by a mean of 15.4 ng/dL, and it was the only study medication to significantly increase prolactin levels. Hyperprolactinemia suppresses gonadotropin-releasing hormone (GnRH), which reduces gonadotropin release and can impair gonadal steroidogenesis in men and women. This elevation is sometimes but not always associated with inhibited reproductive function and symptoms such as galactorrhea, amenorrhea, gynecomastia, erectile dysfunction, and anorgasmia. Long-term hyperprolactinemia may lead to decreased bone density in both sexes when associated with hypogonadism (although no direct evidence exists that would suggest that risperidone decreases bone density, regardless of prolactin level). Approximately one-third of breast cancers are prolactin dependent. A 2021 large cohort study from Finland found an association between prolactin-elevating antipsychotics and the risk of breast cancer in women. These findings suggest that clinicians treating patients with these two medications should inquire about prolactin-related side effects. If there is a concern, prolactin levels can be evaluated. In addition, clinicians should discuss the risk of breast cancer with women when they are treated with risperidone and paliperidone.

PREGNANCY. Risperidone and paliperidone are rated FDA pregnancy category C. A retrospective and prospective case series of 713 pregnancies with risperidone exposure found that in the prospective data, the percentage of congenital malformations (3.6%) and spontaneous abortions (16.9%) was consistent with background rates in the general population. In the retrospectively identified cases, the percentages of congenital malformation (7.6%) and spontaneous abortion (16.7%) were higher, but retrospective reporting is subject to bias. Thirty-seven of 197 pregnancies retrospectively had perinatal syndromes, 21 of which were behavioral or motor disorders. These syndromes involved symptoms of tremor, jitteriness, irritability, somnolence, and feeding problems, consistent with possible risperidone withdrawal or risperidone-induced movement side effects. If risperidone is used during pregnancy, the medication should be slowly titrated down prior to delivery.

LACTATION. Risperidone and paliperidone are excreted in breast milk. The milk/plasma concentration ratio for both agents was less than 0.5. Relative infant doses of risperidone ranged from 2.3% to 4.7% of the maternal dose. In infants who were breastfed, risperidone and paliperidone were not detected in the plasma of the infants. However, the possibility of infant exposure and possible EPS/NMS risk cannot be ruled out. The decision to breastfeed while on risperidone should be made only after a careful risk–benefit analysis.

FERTILITY. Risperidone and paliperidone increase prolactin, which can decrease fertility in men and women. Meta-analysis of 15 antipsychotic medications across 212 clinical trials found risperidone and its metabolite paliperidone caused the greatest elevation in prolactin level of the drugs studied.

SPECIAL POPULATIONS

Pediatric Use. Risperidone has been approved for use in children aged 5 to 16 years old for irritability associated with autism, aged 10 to 17 years for bipolar disorder, manic, or mixed episodes, and aged 13 to 17 years for schizophrenia. As a metabolite of risperidone, paliperidone has similar effectiveness. Overall, over 1,850 pediatric patients have been treated with risperidone in industry-sponsored clinical trials. The rate of tardive dyskinesia in this population was 0.1%, and it resolved with the discontinuation of risperidone. In a review of studies lasting more than 11 months, Correll and Kane reported an annual tardive dyskinesia incidence in children and adolescents of 0.3%. Weight gain among children and adolescents must be monitored with regard to expected weight gain for the appropriate age group. Risperidone is associated with larger-than-expected levels of weight gain in these populations. In a study of 103 adolescents with diagnoses of schizophrenia, the average weight gain was 9.0 kg over 8 months. In several studies of children with autistic disorder or other disorders, there was an average weight gain of 7.5 kg over 1 year, compared to the expected gains of 3 to 3.5 kg per year, depending on age. Risperidone increases prolactin in children and adolescents, as it does in adults. Rates of prolactin elevation vary from 49% to 87% of pediatric patients taking risperidone compared to 2% to 7% of patients on placebo. Galactorrhea was reported in 0.8% of pediatric patients on risperidone, and gynecomastia was reported in 2.3% of risperidone users in the pediatric population. The long-term effects of risperidone treatment on growth and sexual maturation have not been assessed. Transient somnolence lasting a median of 16 days after treatment initiation was reported frequently in the child and adolescent population.

Geriatric Use. Risperidone and paliperidone use in the older adults should be initiated cautiously. Older adult patients are more prone to orthostatic hypotension, which can put them at greater risk of falls.

Furthermore, older adult patients tend to have reduced pharmacokinetic clearance, higher rates of hepatic or renal insufficiency, and a greater likelihood of polypharmacy. Older adult patients started on risperidone should be given a lower initial dose of 0.5 once or twice daily, and titration should be slower than in the adult population. Potential drug interactions should be carefully monitored, and any medications significantly metabolized by CYP2D6 should be coadministered with caution. In particular, two placebo-controlled trials found that concomitant administration of risperidone and furosemide (Lasix) increased mortality in older adult patients with dementia-related psychosis compared to patients treated with risperidone or furosemide alone. Furosemide should not be combined with risperidone in the older adults.

Drug Interactions. Risperidone is extensively metabolized by the CYP 2D6 enzyme, where it is converted to its major metabolite, 9-OH-risperidone (paliperidone). There are identified genetic polymorphisms for the CYP2D6 enzyme, and some percentage of people (6% to 8% in Caucasians; higher in African Americans) have little to no 2D6 activity. These people are poor metabolizers of drugs hydroxylated through the 2D6 pathway. Similarly, some individuals are rapid metabolizers at this enzyme site. Risperidone metabolism can also be affected by drugs that are inhibitors (paroxetine [Paxil], fluoxetine, quinidine [Cardioquin]) or inducers (carbamazepine, phenytoin, rifampin [Rifadin], phenobarbital [Bellatal]) of the 2D6 enzyme. Medications that inhibit the 2D6 enzyme (Table 33.2b–3) increase plasma levels of risperidone and decrease the rate of conversion to 9-OH-risperidone. Dosing of 20 mg each of fluoxetine and paroxetine (strong inhibitors of 2D6) raised plasma levels of risperidone by 2.5 times and 3.9 times, respectively. Because risperidone and its major metabolite have similar if not identical pharmacologic effects, the result of 2D6 inhibition may not be clinically meaningful. However, risperidone dosage should be reassessed whenever a medication with 2D6 inhibition is initiated, discontinued, or changed. Medications that induce 2D6 metabolism can substantially decrease the steady-state levels of risperidone and 9-OH-risperidone. Carbamazepine (a strong 2D6 inducer) lowers the steady-state concentrations of risperidone and its metabolite by up to 50%. Patients coadministered risperidone and any 2D6 inducer, like carbamazepine, should be closely monitored when the inducing medication is initiated, titrated, or discontinued. In particular, if a 2D6-inducing medication is discontinued, risperidone may need to be decreased to avoid plasma concentration increases to toxic levels.

Other potential drug interactions include up to 20% higher levels of valproic acid with risperidone coadministration and the additive effects of medications with the potential to cause orthostatic hypotension.

Dosage and Administration

SCHIZOPHRENIA. For adults starting oral risperidone (available as tablets, fast-dissolve tablets, or a solution), the initial dose should be no higher than 2 mg daily in a single or divided dose. Higher starting dosages are associated with tachycardia and orthostatic hypotension. Dose increases of 1 to 2 mg per day are acceptable but should not be spaced less than 24 hours apart. Most patients are evaluated on 4 to 6 mg for the first 4 to 6 weeks. Many patients may be optimally treated with an average 4 mg daily dose, but doses ranging from 2 to 16 mg have been found effective. Several studies have concluded that doses above 6 mg per day in divided doses were not more effective than lower doses and were associated with higher rates of EPS and other dose-dependent side effects.

For paliperidone extended release, an initial dose of 6 mg daily is recommended. Some patients may require doses up to 12 mg, and others may respond to 3 mg.

Adolescents with schizophrenia should be started on no more than 0.5 mg once daily of risperidone. Dose increases should be made in increments of 0.5 to 1 mg per day, no less than 24 hours apart. The recommended daily dosage for adolescents with schizophrenia is 3 mg per day. Although a dose range of 1 to 6 mg per day has demonstrated efficacy, doses higher than 3 mg per day did not provide additional benefit and were associated with higher rates of side effects. If side effects are experienced, tolerability may be improved by dividing the dose into twice-daily dosing. There is only limited data on doses of paliperidone for children and adolescents. For 12 to 17 year olds, a starting dose of 3 mg is recommended.

Older adult and Parkinson disease patients should start with 0.5 mg once or twice daily of risperidone. Doses should be divided into twice-daily dosing to reduce the risk of orthostatic hypotension. Dose increases should be spaced 2 to 3 days apart, and dose increases above 1.5 mg twice a day should be made no sooner than 1 week after the last dosage change. This population may need maintenance doses of 1 mg or less to prevent other side effects. There is only limited information on paliperidone doses for older adult and Parkinson disease patients. As with risperidone, starting patients on a low dose such as 3 mg daily of paliperidone with monitoring for hypotension is recommended.

Long-Acting Formulations. Both risperidone and paliperidone are available as long-acting injectable formulations. The long-acting preparation of risperidone is an aqueous suspension, with risperidone contained in a matrix of glycolic acid–lactate polymer “microspheres” (25 to 150 μm). This preparation minimizes pain and inflammation at the injection site. Gradual hydrolysis of the copolymer at the site of injection ensures slow, steady, and smooth release of risperidone over several weeks. Paliperidone is available in three different formulations: Sustenna, a monthly formulation; Trinza, an every 3-month formulation; and Hafyera, an every 6-month formulation. The process for initiating these formulations is discussed in Table 33.2b–4.

BIPOLAR MANIA. Adults with acute bipolar mania should be started at a dose of 2 to 3 mg per day of risperidone with once-daily dosing. Dose increments should not be higher than 1 mg per day or spaced less than 24 hours apart. In this population, doses higher than 6 mg per day have not been fully evaluated. Paliperidone should be started at a dose of 6 mg daily. Doses can be increased to 12 mg or decreased to 3 mg depending on side effects and effectiveness.

Children and adolescents with bipolar mania should be started at doses of 0.5 mg once daily of risperidone and titrated up by 0.5 to 1 mg per day no less than 24 hours apart to a recommended dose of 2.5 mg per day. Doses as high as 6 mg per day have been studied in this population, but efficacy was no greater at doses above 2.5 mg per day. The dose can be divided to reduce side effects, should they emerge. Given the unknown long-term impact of risperidone on the neurologic, endocrinologic, and sexual maturation of children, clinicians should strive to find the lowest effective dose and regularly re-evaluate risperidone’s efficacy and appropriateness for each patient.

IRRITABILITY ASSOCIATED WITH AUTISM. Children and adolescents with autistic disorder should be initiated on 0.25 mg per day for patients under

20 kg and 0.5 mg per day for patients over 20 kg. No further dose increase should be initiated for a minimum of 4 days, after which time the dose can be titrated up to 0.5 mg per day for patients under 20 kg and 1 mg per day for patients over 20 kg. This dose should be maintained for 2 weeks. If there is no or partial response at that dose, increases of 0.25 mg per day for patients below 20 kg and 0.5 mg per day for patients over 20 kg should occur at intervals no briefer than 2 weeks. In clinical trials of children and adolescents with autistic disorder, 90% of the medication responders responded to doses between 0.5 and 2.5 mg per day. In patients over 45 kg, the optimal dose may be slightly higher, approximately 3 mg per day. As in the pediatric bipolar population, given the unknown long-term impact of risperidone on the neurologic, endocrinologic, and sexual maturation of children, clinicians should strive to find the lowest effective dose and regularly re-evaluate risperidone's efficacy and appropriateness for each patient.

OVERDOSE. Reports of an overdose of up to 360 mg of risperidone are associated with signs and symptoms such as sedation, tachycardia, hypotension, EPS, prolonged QT interval, and convulsions. A study evaluating the symptoms associated with overdose has shown that risperidone is associated with fewer side effects than clozapine and olanzapine (particularly anticholinergic side effects like agitation, confusion, and dysarthria).

Table 33.2b–3. Overview of Potential Cytochrome P450 Interactions with Antipsychotics

Drug	Major P450 Pathway[a]	Drugs that ↑ Antipsychotic Levels[b]	Drugs that ↓ Antipsychotic Levels[b]	Effect on Other Med Levels[b]	Comments
Clozapine (Clozaril)	1A2, 2D6, 3A4	Fluvoxamine (paroxetine, fluoxetine, bupropion, other 2D6 → modest ↑), cimetidine, caffeine, ciprofloxacin, erythromycin	Nicotine, carbamazepine, modafinil, St. John's wort, omeprazole, phenytoin, rifampin	Mild 2D6 inhibitor. Caution advised with other meds metabolized by 2D6. Levels of both may be ↑.	May decrease tobacco use, which can ↑ clozapine level; caution when patients change tobacco use; due to high toxicity and multiple P450 pathways, check all coadministered medicines for interactions and monitor clozapine levels with medicine changes.
Risperidone (Risperdal)	2D6	Fluoxetine, paroxetine, bupropion, quinidine	Carbamazepine, phenytoin, phenobarbital, rifampin	Valproate ↑ by 20%	Risperidone and its metabolite (paliperidone) are both active.
Olanzapine (Zyprexa)	1A2	Fluvoxamine, ciprofloxacin, fluoxetine (small)	Nicotine, carbamazepine, modafinil, omeprazole, rifampin	N/A	
Quetiapine (Seroquel)	3A4	Divalproex (small), cimetidine (small), nefazodone, ketoconazole, clarithromycin, erythromycin, ritonavir	Carbamazepine, thioridazine, phenytoin, modafinil, St. John's wort	Slight decrease in lorazepam clearance; Slight decrease in valproic acid exposure	
Ziprasidone (Geodon)	3A4	Nefazodone, ketoconazole, clarithromycin, erythromycin, protease inhibitors	Carbamazepine, phenytoin, modafinil, St. John's wort	N/A	Concomitant Rx of medicines that ↑ QTc is contraindicated or should be used with caution.
Aripiprazole (Abilify)	2D6, 3A4	Fluoxetine, paroxetine, bupropion, quinidine, nefazodone, ketoconazole, clarithromycin, erythromycin, protease inhibitors	Carbamazepine, rifampin, phenytoin, modafinil, St. John's wort	N/A	
Paliperidone (Invega)	N/A	Paroxetine (small), divalproex	N/A	N/A	As the metabolite of risperidone, paliperidone undergoes only limited P450 metabolism.
Iloperidone (Fanapt)	2D6, 3A4	Fluoxetine, paroxetine, ketoconazole, itraconazole	Carbamazepine, rifampin, phenytoin, modafinil, St. John's wort	N/A	Recommend not using with other drugs which prolong QTc interval.
Asenapine (Saphris)	1A2	Fluvoxamine, imipramine (small)	Paroxetine (small), carbamazepine (small)	Weak 2D6 inhibitor → double paroxetine exposure	
Lurasidone (Latuda)	3A4	Ketoconazole, clarithromycin, ritonavir, voriconazole, mibefradil, diltiazem, atazanavir, erythromycin, fluconazole, verapamil, grapefruit juice	Rifampin, avasimibe, St. John's wort, Phenytoin, carbamazepine	Small increase in midazolam exposure (CYP3A4 substrate)	
Cariprazine (Vraylar)	3A4, 2D6	Ketoconazole, erythromycin	Carbamazepine, phenytoin, modafinil, St. John's wort		
Brexpiprazole (Rexulti)	2D6, 3A4	Fluoxetine, paroxetine, bupropion, ketoconazole, quinidine, erythromycin	Carbamazepine, phenytoin, rifampin, modafinil, St. John's wort		
Lumateperone (Caplyta)	3A4, 2C8, 1A2	Amprenavir ciprofloxacin cyclosporin diltiazem erythromycin fluconazole fluvoxamine verapamil clarithromycin ritonavir nelfinavir itraconazole	Carbamazepine phenytoin efavirenz modafinil armodafinil pioglitazone nafcillin prednisone rifampin St. John's wort Bosentan		

Rx, prescription; N/A, not available.

[a] Medications primarily metabolized by only one cytochrome system, especially CYP2D6, are subject to more numerous and more significant potential drug–drug interactions than medications with multiple pathways.

[b] This is not a comprehensive list of potential interactions. Clinicians should check for possible drug interactions whenever patients' medication regimens are changed, including by other providers.

Table 33.2b–4.
Comparison of Usual Dosing[a] for the Commonly Used Antipsychotics in Schizophrenia

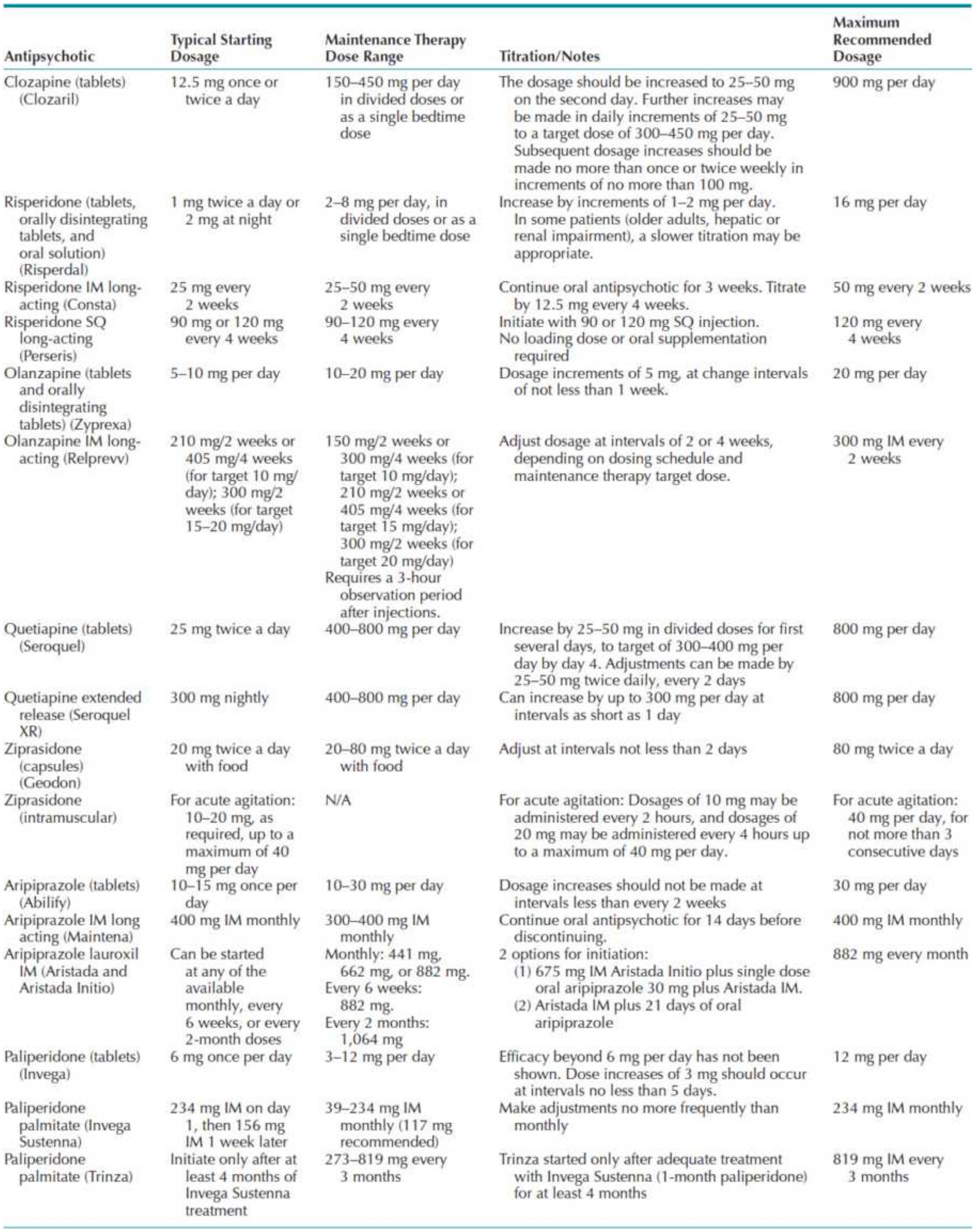

Antipsychotic	Typical Starting Dosage	Maintenance Therapy Dose Range	Titration/Notes	Maximum Recommended Dosage
Clozapine (tablets) (Clozaril)	12.5 mg once or twice a day	150–450 mg per day in divided doses or as a single bedtime dose	The dosage should be increased to 25–50 mg on the second day. Further increases may be made in daily increments of 25–50 mg to a target dose of 300–450 mg per day. Subsequent dosage increases should be made no more than once or twice weekly in increments of no more than 100 mg.	900 mg per day
Risperidone (tablets, orally disintegrating tablets, and oral solution) (Risperdal)	1 mg twice a day or 2 mg at night	2–8 mg per day, in divided doses or as a single bedtime dose	Increase by increments of 1–2 mg per day. In some patients (older adults, hepatic or renal impairment), a slower titration may be appropriate.	16 mg per day
Risperidone IM long-acting (Consta)	25 mg every 2 weeks	25–50 mg every 2 weeks	Continue oral antipsychotic for 3 weeks. Titrate by 12.5 mg every 4 weeks.	50 mg every 2 weeks
Risperidone SQ long-acting (Perseris)	90 mg or 120 mg every 4 weeks	90–120 mg every 4 weeks	Initiate with 90 or 120 mg SQ injection. No loading dose or oral supplementation required	120 mg every 4 weeks
Olanzapine (tablets and orally disintegrating tablets) (Zyprexa)	5–10 mg per day	10–20 mg per day	Dosage increments of 5 mg, at change intervals of not less than 1 week.	20 mg per day
Olanzapine IM long-acting (Relprevv)	210 mg/2 weeks or 405 mg/4 weeks (for target 10 mg/day); 300 mg/2 weeks (for target 15–20 mg/day)	150 mg/2 weeks or 300 mg/4 weeks (for target 10 mg/day); 210 mg/2 weeks or 405 mg/4 weeks (for target 15 mg/day); 300 mg/2 weeks (for target 20 mg/day) Requires a 3-hour observation period after injections.	Adjust dosage at intervals of 2 or 4 weeks, depending on dosing schedule and maintenance therapy target dose.	300 mg IM every 2 weeks
Quetiapine (tablets) (Seroquel)	25 mg twice a day	400–800 mg per day	Increase by 25–50 mg in divided doses for first several days, to target of 300–400 mg per day by day 4. Adjustments can be made by 25–50 mg twice daily, every 2 days	800 mg per day
Quetiapine extended release (Seroquel XR)	300 mg nightly	400–800 mg per day	Can increase by up to 300 mg per day at intervals as short as 1 day	800 mg per day
Ziprasidone (capsules) (Geodon)	20 mg twice a day with food	20–80 mg twice a day with food	Adjust at intervals not less than 2 days	80 mg twice a day
Ziprasidone (intramuscular)	For acute agitation: 10–20 mg, as required, up to a maximum of 40 mg per day	N/A	For acute agitation: Dosages of 10 mg may be administered every 2 hours, and dosages of 20 mg may be administered every 4 hours up to a maximum of 40 mg per day.	For acute agitation: 40 mg per day, for not more than 3 consecutive days
Aripiprazole (tablets) (Abilify)	10–15 mg once per day	10–30 mg per day	Dosage increases should not be made at intervals less than every 2 weeks	30 mg per day
Aripiprazole IM long acting (Maintena)	400 mg IM monthly	300–400 mg IM monthly	Continue oral antipsychotic for 14 days before discontinuing.	400 mg IM monthly
Aripiprazole lauroxil IM (Aristada and Aristada Initio)	Can be started at any of the available monthly, every 6 weeks, or every 2-month doses	Monthly: 441 mg, 662 mg, or 882 mg. Every 6 weeks: 882 mg. Every 2 months: 1,064 mg	2 options for initiation: (1) 675 mg IM Aristada Initio plus single dose oral aripiprazole 30 mg plus Aristada IM. (2) Aristada IM plus 21 days of oral aripiprazole	882 mg every month
Paliperidone (tablets) (Invega)	6 mg once per day	3–12 mg per day	Efficacy beyond 6 mg per day has not been shown. Dose increases of 3 mg should occur at intervals no less than 5 days.	12 mg per day
Paliperidone palmitate (Invega Sustenna)	234 mg IM on day 1, then 156 mg IM 1 week later	39–234 mg IM monthly (117 mg recommended)	Make adjustments no more frequently than monthly	234 mg IM monthly
Paliperidone palmitate (Trinza)	Initiate only after at least 4 months of Invega Sustenna treatment	273–819 mg every 3 months	Trinza started only after adequate treatment with Invega Sustenna (1-month paliperidone) for at least 4 months	819 mg IM every 3 months

Paliperidone Palmitate (Hafyera)	Initiate only after at least 4 months on Invega Sustenna or 3 months on Invega Trinza	1,092–1,560 mg every 6 months	Hafyera dose based on previous Invega or Trinza dose. Dose adjust every 6 months. Patient response to dose changes may not be apparent for several months.	1,560 mg IM every 6 months
Iloperidone (Fanapt)	1 mg twice daily	6–12 mg twice daily	Can make dose increases no more frequently than by 2 mg twice daily every day	12 mg twice daily
Asenapine (Saphris)	5 mg twice daily	5–10 mg twice daily	Can increase to 10 mg twice daily after 1 week	10 mg twice daily
Asenapine transdermal system (Secuado)	3.8 mg every 24 hours	3.8–7.6 mg every 24 hours	Can increase to 5.7 mg/24 hours or 7.6 mg/24 hours after 1 week at 3.8 mg	7.6 mg every 24 hours
Lurasidone (Latuda)	40 mg daily with food	40–160 mg daily with food	Not specified; tablets are available in 20, 40, 60, 80, and 120 mg strengths	160 mg daily
Cariprazine (Vraylar)	1.5 mg once per day	1.5–6 mg once daily	Start with 1.5 mg for 1 day with increases in 1.5 mg increments; due to the long half-life of cariprazine and its active metabolites, changes in dose will not be fully reflected in plasma for several weeks	6 mg
Brexpiprazole (Rexulti)	1 mg once per day	2–4 mg daily for schizophrenia; 2–3 mg daily for adjunctive treatment of depression	For schizophrenia, start 1 mg QD for 4 days, then 2 mg for 3 days, then 4 mg QD. For adjunctive treatment of depression, start 0.5–1 mg QD, then increase dose to 2 mg QD after 1 week.	4 mg for schizophrenia; 3 mg for adjunctive treatment of depression
Lumateperone (Caplyta)	42 mg once per day	42 mg daily	No titration required Start with 42 mg Only 42-mg dose approved	42 mg

Information taken from U.S. Prescribing Information for individual agents.

[a] Dosage adjustments may be required in special populations.

DOPAMINE, SEROTONIN, NET (NE TRANSPORTER) INHIBITOR (METABOLITE)

Quetiapine

Quetiapine was developed by Zeneca Laboratories. Its new drug application was filed with the FDA in 1996 and approved in 1998.

Chemistry. Quetiapine is a dibenzothiazepine with more potent 5-HT_2 than D_2 receptor-blocking properties. Its molecular formula is $C_{42}H_{50}N_6O_4S_2 \cdot C_4H_4O_4$. Quetiapine is related to the clozapine, olanzapine (Fig. 33.2b–1), and fluperlapine molecules.

Pharmacologic Action

PHARMACOKINETICS

Absorption. The time to maximum concentration (T_{max}) after oral intake is 1.5 hours, with an estimated half-life of 6 hours. No difference was found between twice- and three-times-daily dosing, in spite of a steady-state half-life of 6.9 hours.

Distribution. Steady-state levels are reached in 48 hours. Quetiapine has a wide volume of distribution, approximately 10 L/kg.

Metabolism and Elimination. Quetiapine is extensively metabolized by the liver. It has many metabolites, although only a small percentage of these are active. The primary metabolic pathway is CYP3A4. Ninety-five percent of [^{3}H] quetiapine is recovered as metabolites in urine and feces, but only 1% is recovered as quetiapine itself.

PHARMACODYNAMICS

Mechanism of Action. Quetiapine is a multitransmitter antipsychotic agent: It has a high affinity for 5-HT$_2$, histamine H$_1$, α_1, and α_2 receptors; a moderate affinity for 5-HT$_{1A}$ receptors; a moderate affinity for D$_2$; and a low affinity for D$_1$ receptors. Quetiapine has a very low affinity for cholinergic M$_1$ and D$_4$ receptors (despite its low affinity for cholinergic receptors, quetiapine is associated with several typically anticholinergic side effects, such as dry mouth and constipation). In addition, norquetiapine, the major metabolite of quetiapine, has a high affinity for the norepinephrine transporter (NET).

Quetiapine's low incidence of EPS may be explained by its relatively high 5-HT$_{2A}$ binding compared to its activity at D$_2$ receptors, according to Meltzer. An early PET study demonstrated that 2 hours after a single clinical dose, quetiapine occupies 44% of D$_2$ receptors and 72% of 5-HT$_2$ receptors. Other PET studies demonstrate rapid dissociation of quetiapine from the D$_2$ receptor. In one study, D$_2$ receptor occupancy 2 hours postdose ranged from 58% to 64% and fell to 20% to 0% 12 to 24 hours postdose for single doses of 400 and 450 mg, respectively. This supports the hypothesis of Kapur and colleagues, who propose that quetiapine's "fast dissociation" (i.e., a high D$_2$ occupancy which decreases to low levels by the end of the dosing interval) may be sufficient to induce antipsychotic response while minimizing EPS. Further, there is evidence that the clinical effectiveness of quetiapine is related to both the low occupancy of D$_2$ receptors and inhibition of the NET.

Design and Interpretation of Clinical Drug Studies

SCHIZOPHRENIA: ACUTE TREATMENT. The earliest clinical trials demonstrated that quetiapine is significantly superior to placebo at reducing positive, negative, and depressive symptoms of schizophrenia. A post hoc analysis of early placebo-controlled trials found that quetiapine separated from placebo on multiple clinical measures by the end of the first week of treatment. Later trials compared quetiapine to other FGAs and SGAs. Some of the early comparison trials suffered from quetiapine underdosing, with some study doses as low as 150 to 300 mg per day. These studies generally found quetiapine to have comparable efficacy to haloperidol in reducing positive symptoms but slightly less efficacy in reducing negative symptoms. Later studies began assessing quetiapine treatment at higher doses, up to 800 mg per day. Such studies mostly found comparable efficacy between quetiapine and haloperidol on positive and negative symptoms and virtually no EPS or persistent prolactin elevation with quetiapine. In two separate 6- to 8-week, double-blind comparative studies, quetiapine (300 to 600 mg per day) had comparable efficacy to haloperidol in reducing psychotic symptoms. Both quetiapine and haloperidol produced clear reductions in mean PANSS scores, with more quetiapine-treated patients showing a clinical response (greater than 20% reduction in total PANSS score, $p = .043$). In addition, patients treated with quetiapine had a greater reduction in EPS ($p = .005$) and fewer treatment-emergent extrapyramidal-related adverse events ($p < .001$) compared with haloperidol-treated patients. In another study comparing multiple fixed doses of quetiapine to haloperidol and placebo, a dose range of 150 to 750 mg per day of quetiapine was comparable to haloperidol 12 mg per day in reducing positive symptoms, and doses of quetiapine above 300 mg per day were comparable in reducing negative symptoms. A post hoc analysis of this data concluded that patients treated with quetiapine had significantly greater reductions of hostility and agitation compared to both placebo- and haloperidol-treated patients that were independent of the improvements in psychotic symptoms. An 8-week double-blind trial of 673 patients comparing quetiapine (mean dose 525 mg per day) and risperidone treatment (mean dose 5.2 mg per day) found that overall treatment efficacy did not differ between the two drugs, although there were some data to support greater positive symptom reduction with risperidone. Quetiapine had significantly lower rates of EPS and prolactin elevation but higher rates of somnolence. Another 6-week, double-blind trial in recently hospitalized patients with schizophrenia found that treatment with risperidone, but not quetiapine, was associated with statistically significant separation from placebo by week 2 of treatment.

Patients on quetiapine required higher rates of antipsychotic polypharmacy to control symptom exacerbation.

SCHIZOPHRENIA: MAINTENANCE TREATMENT. Quetiapine does not have FDA approval for the maintenance treatment of schizophrenia. However, quetiapine is frequently used off-label for maintenance treatment. Studies of quetiapine treatment in schizophrenia have followed patients for up to 18 months and found continued symptomatic improvement of positive, negative, and affective symptoms on par with improvements seen in acute treatment with quetiapine and the maintenance effects of other FGAs and SGAs. A small open-label study of 25 patients compared treatment with oral quetiapine (mean dose 493 mg per day) to treatment with haloperidol decanoate (mean dose 170 mg per injection). Quetiapine treatment was associated with comparable prevention of symptom exacerbation over 48 weeks, less rigidity and akathisia, and greater improvement in negative symptoms over haloperidol decanoate treatment. There have been few direct comparisons between quetiapine and the other SGA medications. A large multinational naturalistic prospective study of 5,883 participants compared the efficacy of haloperidol, olanzapine, risperidone, and quetiapine at 6 months of treatment. Quetiapine was found to have comparable efficacy to haloperidol but was less efficacious than olanzapine and risperidone. However, quetiapine was likely underdosed, with a mean study dose of 200 mg per day at baseline and 300 mg per day by 6 months.

An open-label study of 728 patients with schizophrenia compared quetiapine to risperidone treatment for 4 months. Quetiapine- and risperidone-treated patients showed similar improvements on positive and negative symptoms scales, and quetiapine-treated patients had greater improvements on depressive symptom scales. Quetiapine was not associated with any extrapyramidal side effects but had higher rates of dry mouth, somnolence, and dizziness. A 6-month study comparing treatment with quetiapine or olanzapine in 346 patients found that both medications demonstrated similar efficacy in the treatment of negative symptoms but that olanzapine therapy was associated with significantly greater positive symptom reduction, improvements in measures of functional outcome, and reduced rates of medication discontinuation. Finally, a 1-year study comparing olanzapine, risperidone, and quetiapine in the treatment of 400 patients with early-onset psychosis found no differences in the all-cause discontinuation rate among the drugs. However, positive symptom reductions were greater in the olanzapine and risperidone groups at some (but not all) time points.

These studies suggest that when appropriately dosed, quetiapine has equivalent or near equivalent efficacy compared with FGAs or other SGAs. Quetiapine appears to demonstrate a tendency toward slightly weaker symptom control but a better movement disorder profile than most other antipsychotic medications.

CATIE TRIAL. Quetiapine did not separate from fluphenazine, risperidone, or ziprasidone on the primary measure of time to discontinuation in the first phase of CATIE. Treatment with quetiapine (mean dose 543.4 mg per day) was associated with higher rates of side effects like dry mouth and constipation. In phase II, patients who had discontinued an SGA due to lack of efficacy had a significantly longer time to discontinuation on olanzapine and risperidone than on quetiapine or ziprasidone. However, in patients who had discontinued the first-generation medication due to lack of efficacy, the time to discontinuation was longer in the quetiapine and olanzapine groups than in the risperidone group. This latter finding suggests that quetiapine, because of its low incidence of EPS, may have advantages for patients who have difficulty tolerating FGAs.

BIPOLAR DEPRESSION. Two 8-week, placebo-controlled trials of identical design evaluated treatment response to quetiapine at 300 mg per day and 600 mg per day fixed doses in 542 and 509 outpatients with bipolar I or bipolar II depression. The results of both studies were similar. Both fixed doses demonstrated significant improvements in depressive symptoms, which separated from placebo by the first week of treatment. There were no statistically significant differences between the dosage groups. There was no difference between quetiapine- and placebo-treated patients in terms of manic switch rate. A subanalysis of one of the studies examined patients with a rapid-cycling course and concluded that quetiapine 300 or 600 mg per day was significantly more efficacious than placebo at reducing symptoms of depression, anxiety, and related symptoms in this patient population.

BIPOLAR MANIA. Two identical studies comparing adjunctive quetiapine treatment versus adjunctive placebo treatment in 191 and 402 bipolar manic patients on lithium or valproate found that adjunctive quetiapine treatment (mean dose 584 mg per day and 492 mg per day, respectively) was associated with significantly greater reductions in manic symptoms and significantly higher rates of remission than adjunctive treatment with placebo. In the second study, a statistically greater degree of improvement in manic symptoms was evident in the adjunctive quetiapine group by the end of the first week of treatment. A study of 302 patients randomized to quetiapine, lithium, or placebo monotherapy treatment concluded that quetiapine was significantly more effective at reducing manic symptoms than placebo. Although the study was not powered to detect differences between lithium and quetiapine treatment, a post hoc analysis found no significant differences in symptom reduction between patients treated with quetiapine versus lithium. Another placebo-controlled study using haloperidol as the active control found that quetiapine was slightly less effective at reducing symptoms and achieving remission than haloperidol. The comparative analysis was post hoc, and some differences between the medications were not statistically significant. A placebo-controlled trial of 403 patients with mania concluded that quetiapine was associated with higher rates of symptom remission than placebo at both 3 and 12 weeks.

OTHER DIAGNOSES. Quetiapine was found to be effective for treating generalized anxiety disorder in three randomized controlled trials.

Quetiapine is used off-label for a number of other diagnoses and symptoms, despite limited controlled data. Three small open-label studies have evaluated the effects of treatment with quetiapine for chronic post-traumatic stress disorder (PTSD). In all three studies, quetiapine significantly reduced the severity of PTSD symptoms in patients with and without psychotic features.

There are several small trials of quetiapine treatment of Obsessive-Compulsive Disorder (OCD). A small single-blind, placebo-controlled study found that adjunctive treatment with quetiapine and an SSRI was more efficacious at reducing symptoms of OCD than placebo and SSRI.

There are a few case reports and small open-label trials of quetiapine in the treatment of borderline personality disorder. The dose ranges in these studies varied from about 200 to 800 mg per day. Overall, the studies reported that treatment with quetiapine improved multiple symptom clusters, most notably mood, impulsivity, and aggression. There have been no randomized, double-blind, controlled trials in this population to date.

Anecdotally, quetiapine is considered the antipsychotic drug of choice to treat psychosis in Parkinson disease. Although quetiapine does not appear to worsen parkinsonism, two placebo-controlled trials failed to show efficacy in treating the psychotic symptoms associated with the dopaminergic treatment of Parkinson disease. One study comparing quetiapine to clozapine (the only antipsychotic medication with proven efficacy and tolerability in Parkinson psychosis) found comparable levels of efficacy between the two drugs. Because of the high 5-HT_{2A} blockade and relatively low D_2 binding, quetiapine has been used off-label as a first-line treatment for psychosis in Parkinson disease. Quetiapine is more easily administered than clozapine in patients with Parkinson disease, but treatment efficacy has not been sufficiently established. The sedating effects and potential EPS are treatment limiting in this population.

QUETIAPINE EXTENDED RELEASE FORMULATION. A study comparing quetiapine Extended-release (XR) (once daily dosing) to quetiapine immediate release (twice daily dosing) and placebo found comparable improvements with both forms of quetiapine preparation compared to placebo. There were no adverse events specific to the XR formulation.

Effects on Specific Organs and Systems

CARDIOVASCULAR SYSTEM. Postural hypotension, likely due to quetiapine's high affinity for adrenergic α_1 receptors, is the most common cardiovascular effect. Tachycardia also occurs at rates above placebo; whether the tachycardia is secondary to postural hypotension or a separate effect is not clear. Quetiapine is not associated with clinically important increases in the QTc interval.

RESPIRATORY AND EAR, NOSE, AND THROAT SYSTEM. Dyspnea, cough, pharyngitis, rhinitis, and nasal congestion have been reported with rates slightly above placebo, which is observed with all SGAs.

GASTROINTESTINAL SYSTEM. Dry mouth, constipation, and dyspepsia are commonly reported side effects of quetiapine, with dry mouth being an especially frequent side effect. Abdominal pain is less common, but it (along with dyspepsia) is a dose-related effect.

HEMATOLOGIC AND LYMPHATIC SYSTEM. Leukopenia and neutropenia in conjunction with quetiapine therapy have been observed. Very rare cases of agranulocytosis, including fatal cases, have also been described.

CENTRAL NERVOUS SYSTEM. Sedation, somnolence, and dizziness are frequently reported effects, with lethargy slightly less so. Headache and, to a lesser degree, asthenia are also commonly reported.

METABOLIC AND NUTRITIONAL SYSTEM. Peripheral edema has been described with quetiapine and other antipsychotics. Hyperlipidemia, hypertriglyceridemia, and hyperglycemia are common effects of quetiapine. These have been attributed to the H_1 and 5-HT_{2C} binding as in other multireceptor antipsychotics.

Therapeutic Indications

SCHIZOPHRENIA. Quetiapine is indicated for the acute treatment of schizophrenia, although in clinical practice, it is used both acutely and for maintenance treatment.

BIPOLAR DISORDER. Quetiapine is indicated for the acute treatment of manic episodes in bipolar I disorder, either as monotherapy or as adjunctive treatment with lithium or valproate. It is also indicated for monotherapy for the acute treatment of depressive episodes associated with bipolar I or II disorder. Quetiapine is indicated for the maintenance treatment of bipolar I disorder as an adjunct to lithium or divalproex.

OFF-LABEL USES. Quetiapine is used in clinical practice to treat a number of symptoms. Typically, in these cases, clinicians are taking advantage of quetiapine’s sedative and somnolent properties to treat insomnia, anxiety, and agitation. Usually, the doses used for such symptomatic indications are between 12.5 and 100 mg per day, and insomnia will often respond to a dose range of 25 to 50 mg at bedtime. Sleep induction may be mediated by H_1 blockade, while improvement in sleep maintenance insomnia (i.e., decreased slow-wave sleep) may be caused by 5-HT_{2A} antagonism. This mechanism in insomnia is similar to the effects of low-dose doxepin (Sinequan). Although quetiapine benefits from having limited abuse potential, the possible metabolic risks of quetiapine use must be weighed against the benefits of a nonaddictive medication. Having lower doses for these indications means that the metabolic risks are also lower but still require regular assessment.

Quetiapine is used to treat nonpsychotic disorders such as major depressive disorder, OCD, and borderline personality disorder.

Anecdotally, quetiapine is considered the antipsychotic drug of choice to treat psychosis in Parkinson disease, although efficacy failed to be established in two placebo-controlled trials.

Precautions and Adverse Reactions

BLACK BOX WARNINGS

Increased Mortality in Older Adult Patients with Dementia-Related Psychosis. Quetiapine and the other SGAs carry a black box warning for increased mortality in older adult patients with dementia-related psychosis compared with placebo.

Suicidal Thoughts and Behaviors. After quetiapine received an indication for the treatment of depression in bipolar disorder, it received the same black box warning as other antidepressant medications. In 2004, the FDA issued a warning about suicidal ideation and antidepressant therapy in the pediatric and young adult population. Although quetiapine was not one of the agents included in the 24 trials that led to the FDA black box warning, the impact of quetiapine on suicidal ideation in patients with depression or anxiety has not been assessed. Any patient, especially patients younger than 24 years, treated with medication for depression or severe anxiety, should be monitored for the emergence or worsening of suicidal ideation. Emergent or worsening symptoms of anxiety, panic attacks, agitation, irritability, hostility, aggressiveness, impulsivity, akathisia, insomnia, hypomania, and mania may precede the onset of suicidal ideation and should be evaluated carefully at each visit. Patients started on treatment for depression, or severe anxiety should be monitored frequently for the first 2 to 3 months of treatment or until remission occurs.

NEUROLEPTIC MALIGNANT SYNDROME. Despite quetiapine's very low rate of EPS, rare cases of NMS have been reported with quetiapine use. Physicians should remain aware that even medications with low to nonexistent EPS risk can still occasionally cause NMS. Furthermore, several cases of high creatine kinase levels in the absence of clear NMS have been reported with quetiapine.

TARDIVE DYSKINESIA. Quetiapine has a very low rate of tardive dyskinesia. Although there are some reported cases of presumed quetiapine-induced tardive dyskinesia, there are many more reports of quetiapine being an effective treatment of tardive dyskinesia, much like clozapine. However, the data supporting quetiapine as a treatment of tardive dyskinesia are equivocal.

MOVEMENT SIDE EFFECTS AND DISORDERS. Since its release in 1998, most studies suggest that quetiapine, along with clozapine, has the lowest rate of extrapyramidal side effects of all the SGAs. This is thought to be the result of quetiapine's low and transient D_2 occupancy and its relatively high 5-$HT_{2A/2C}$ binding.

HYPERCHOLESTEROLEMIA AND HYPERTRIGLYCERIDEMIA. In the industry-sponsored efficacy and safety trials of quetiapine, 16% of patients with schizophrenia and 9% of patients with bipolar disorder developed hypercholesterolemia, compared with 7% and 6% of placebo-treated patients. The rates of hypertriglyceridemia were higher; 23% of people with schizophrenia and 14% of people with bipolar disorder, versus placebo rates of 16% and 9%. In phase I of the CATIE trial, quetiapine-treated patients experienced exposure-adjusted mean increases in cholesterol and triglycerides of 6.6 and 21.2 mg/dL, respectively. These increases were second only to olanzapine and clinically significantly higher than those seen with risperidone or perphenazine. Severe cases of hypertriglyceridemia, with levels above 400 mg/dL, are not uncommon with quetiapine therapy. Patients on quetiapine should be monitored regularly (at baseline and every 6 months) for changes in lipid parameters.

HYPERGLYCEMIA AND DIABETES MELLITUS. Quetiapine is associated with increases in plasma glucose and an increased risk of diabetes, but the degree of glucose increase is less dramatic than that of lipid and triglyceride increases with quetiapine. In two long-term treatment studies, 10.7% of patients on quetiapine developed fasting glucose measurements greater than 126 mg/dL versus 4.6% of placebo-treated patients. During phase I of the CATIE trial, patients treated with quetiapine had exposure-adjusted mean increases in blood glucose of 7.5 mg/dL, an increase that was not significantly larger than that seen with risperidone or perphenazine. Patients treated with quetiapine should be monitored regularly for increases in fasting glucose above 100 mg/dL.

CHANGES IN WEIGHT. Quetiapine is associated with significant increases in weight, likely related in part to its high affinity for histaminergic H_1 receptors. In four 3- to 6-week trials of patients with schizophrenia, 23% of quetiapine-treated patients and 6% of placebo-treated patients gained more than 7% of their body weight. Similar rates were seen in patients with bipolar disorder treated with either quetiapine or placebo. However, the increases are not as extreme as those seen with olanzapine or clozapine. In phase I of the CATIE study, 16% of patients treated with quetiapine gained over 7% of their body weight. The mean weight gain for quetiapine-treated patients was 1.1 l from baseline to last observation.

ORTHOSTATIC HYPOTENSION AND SYNCOPE. Quetiapine is a potent antagonist at the α_1-adrenergic receptor, contributing to its risk of orthostatic hypotension and syncope. Other related effects include dizziness and tachycardia. These symptoms are most pronounced during the initial dose titration and are minimized by starting doses of 25 mg twice a day or less. In the older adults, even lower initial doses may be indicated. Patients with known cardiovascular disease, cerebrovascular disease, or conditions that increase the risk of hypotension should be initiated on quetiapine with caution and frequent monitoring.

CATARACTS. Studies on dogs at quetiapine doses four times the maximum human dose found associations between long-term (6- or 12-month) quetiapine use and cataracts. Studies in humans have failed to find similar associations between cataracts and quetiapine use. The product insert for quetiapine recommends slit lamp examinations at treatment initiation and every 6 months thereafter; it is unclear if these recommendations are followed regularly in clinical care or whether the health care costs of such frequent assessment are justified by the hypothetical human risk of cataracts.

THYROID HORMONE CHANGES. Quetiapine affects total and free thyroxine (T_4) levels in a dose-dependent manner. At the high therapeutic dose range of quetiapine, T_4 levels may be decreased by as much as 20%. These changes are generally not clinically significant, nor do they typically result in changes to thyroid stimulating hormone (TSH) or thyroid-binding globulin (TBG). Less than 1% of patients had changes in TSH while on quetiapine. T_4 levels generally return to baseline with discontinuation of quetiapine. Patients on lithium therapy who are started on adjunctive treatment with quetiapine are likely at greater risk of thyroid hormone fluctuation than those not started on quetiapine adjunctively.

ELEVATED TRANSAMINASES. Quetiapine has been associated with transient and asymptomatic elevations in serum transaminases, especially Alanine Transaminase (ALT). In 3- to 6-week trials of patients with schizophrenia, 6% of people on quetiapine had elevations in transaminases of over three times the normal limit, compared with 1% of people on placebo. The transaminase elevations typically occurred by the third week of treatment, and levels returned to baseline with continued quetiapine administration.

SOMNOLENCE AND SEDATION. Quetiapine can cause significant sedation and somnolence, especially during initial dose titration. In industry-sponsored studies, the rates of somnolence in the schizophrenia, bipolar mania, and bipolar depressed groups ranged from 18% to 34% of patients treated with quetiapine versus 7% to 9% of patients treated with placebo. The percentage of patients on quetiapine reporting somnolence in the CATIE trial was 31%, but this was not significantly different from the other antipsychotic treatment groups. The sedation and somnolence are likely secondary to quetiapine's high affinity for histaminergic H_1 receptors.

PREGNANCY. Quetiapine is rated pregnancy category C. Three case studies of pregnant women treated with quetiapine throughout all or part of their pregnancies reported no noticeable effects of quetiapine treatment on the neonates.

Dosage and Administration

SCHIZOPHRENIA. Quetiapine dosages should be titrated to avoid excessive somnolence, postural hypotension, and syncope. The manufacturer's recommended dosing schedule is as follows: 25 mg twice daily on day 1, 75 to 100 mg in divided doses by day 2 or 3, for a total dose of 300 to 400 mg in divided doses by day 4. Dose increments above day 4 should occur no more rapidly than every 2 days (however, other studies have shown that a more aggressive titration is well tolerated and associated with good response). The target dose range is 150 to 800 mg per day. All efficacy studies occurred with twice or three times a day dose schedules, but in the community, quetiapine is frequently administered once daily at bedtime. Dosages above 800 mg have not been systematically evaluated, but doses up to 1,200 mg per day have been used clinically.

MANIA. In acute mania, the manufacturer recommends an initial dose of 100 mg in a divided dose, with rapid titration of up to 100 mg increases per day in divided doses, to a total dose of 400 mg by day 4. The dose can be escalated by up to 200 mg increases on day 5 and 6, for a total dose of 800 mg by day 6. Children and adolescents (aged 10 to 17 years) should start with 25 mg twice daily and should reach a total dose of 400 mg by day 5. Patients should be monitored for orthostatic hypotension and syncope. The effective dose range for mania is 400 to 800 mg per day (400 to 600 mg per day in children and adolescents). Inpatients may benefit from the daytime sedative effects of twice-daily dosing, and as an outpatient, community practice is once a day at bedtime dosing.

DEPRESSION. The recommended titration schedule for bipolar depression is as follows: day 1, 50 mg at bedtime; day 2, 100 mg at bedtime; day 3, 200 mg at bedtime; day 4, 300 mg at bedtime. Patients being titrated beyond 300 mg can receive 400 mg by day 5 and 600 mg by day 8. In efficacy trials, no additional antidepressant benefit was noted above doses of 300 mg. The efficacy trials for bipolar depression dosed quetiapine once daily at night.

The manufacturer's recommended titration schedules should be modified for people who demonstrate significant orthostatic hypotension, somnolence, or have episodes of syncope.

QUETIAPINE EXTENDED RELEASE. The extended-release formulation of quetiapine is intended for once-daily dosing at bedtime. The recommended initial dose is 300 mg. The dose can be titrated up to 800 mg as rapidly as by day 3. The tablets should not be crushed or cut. It is recommended that quetiapine extended release be taken without food or with no more than 300 calories at the time of administration, as high calorie or high fat meals cause increased plasma concentrations.

DOSING IN SPECIAL POPULATIONS

Older Adults. Older adult patients are generally more sensitive to quetiapine's effect on blood pressure. In this population, orthostatic hypotension or syncope can increase the risk of falls. Furthermore, the metabolism of all drugs is slowed in geriatric populations. The titration schedule in the older adults should be slowed and the target doses lower. Initial doses should be 12.5 to 25 mg per day, and increments of no more than 25 mg should occur every 1 to 3 days, depending on tolerability. Quetiapine is not indicated for patients with psychosis due to dementia.

Hepatic Impairment. Patients with hepatic impairment should be initiated at doses no higher than 25 mg per day. The dose can be increased daily by 25 to 50 mg to an effective dose.

Drug Interactions. Quetiapine is metabolized by the CYP system. CYP 3A4 is the major metabolic pathway, and CYP 2D6 is a minor pathway. Quetiapine is neither an inhibitor nor an inducer of any CYP enzyme system. Cimetidine (a nonspecific enzyme inhibitor) did not inhibit quetiapine metabolism. Phenytoin, a strong inducer of CYP3A4, caused a fivefold increase in quetiapine clearance. Similarly, ketoconazole, a strong CYP3A4 inhibitor, increased quetiapine levels by 335%. Valproic acid increased quetiapine plasma levels by a clinically insignificant amount. Lithium, lorazepam (or other benzodiazepines), risperidone, haloperidol, fluoxetine, and imipramine (Tofranil) had no effects on the pharmacokinetics of quetiapine and vice versa (i.e., did not require dose adjustments). Concomitant administration of thioridazine (Mellaril), however, increased oral clearance of quetiapine by 65%.

Overdose. Quetiapine is relatively safe in overdose. Acute nonfatal overdoses up to 30 g have been described. One fatal overdose involved ingestion of over 13 g of quetiapine. The effects of overdose are exaggerations of the medication's side-effect profile: Sedation and somnolence, hypotension, and tachycardia.

SEROTONIN, DOPAMINE

Iloperidone

Iloperidone is the only antipsychotic in this NbN category. It has a high affinity for 5-HT_{2A} and D_2 receptors. Important clinical properties include a low liability to cause EPS but a risk of orthostatic hypotension requiring a titration period of several days.

Initial research and development of iloperidone occurred in the early to middle 1990s. FDA approval for the treatment of schizophrenia was finally granted in 2009 after an initial new drug application submitted in 2007 was denied due to the need for additional clinical trial support.

Chemistry. Iloperidone belongs to the class of piperidinyl-benzisoxazole derivatives (Fig. 33.2b–1). The chemical name of iloperidone is 4′-[3-[4-(6-Fluoro-1,2-benzisoxazol-3-yl)piperidino]propoxy]-3′-methoxyacetophenone. Its molecular formula is $C_{24}H_{27}FN_2O_4$, and its molecular weight is 426.48.

Pharmacologic Action

PHARMACOKINETICS. The mean elimination half-lives for iloperidone and its two major metabolites (P88 and P95) are 18, 26, and 23 hours, respectively, in CYP2D6 extensive metabolizers. In poor metabolizers, the half-lives are 33, 37, and 31 hours. Steady-state concentrations in plasma are reached after 3 to 4 days of dosing.

Absorption. Iloperidone is well absorbed with oral administration, with peak plasma levels occurring 2 to 4 hours after ingestion. Administration with a high-fat meal delayed $T_{\max}$ for iloperidone by 1 hour and its major metabolites, P88 and P95, by 2 and 6 hours, respectively. As a high-fat meal does not significantly affect the maximum concentration or area under the curve (AUC) of iloperidone or its major metabolites, product labeling indicates it can be administered without regard to meals.

Distribution. Iloperidone has an apparent volume of distribution of 1,340 to 2,800 L, and it, as well as its major metabolites, is approximately 95% serum protein bound.

Metabolism and Elimination. Iloperidone is metabolized by three pathways: CYP2D6-mediated hydroxylation, CYP3A4-mediated O-demethylation, and carbonyl reduction. There are two major metabolites of iloperidone, P95 and P88, with P95 exposure being greater in CYP2D6 extensive metabolizers and P88 being greater in CYP2D6 poor metabolizers. Elimination of metabolites is primarily through the urine, with feces accounting for a lesser proportion.

PHARMACODYNAMICS

Mechanism of Action. Iloperidone functions as an antagonist with high affinity at 5-HT$_{2A}$, D$_2$, D$_3$, α$_{1-}$, and α$_{2c}$-adrenergic receptors; intermediate affinity at D$_4$, 5-HT$_{6,}$ and 5-HT$_7$ receptors; and low affinity for 5-HT$_{1A}$, D$_1$, and H$_1$ receptors. Iloperidone possesses a negligible affinity for muscarinic receptors. Its metabolite P88 has receptor affinities "generally equal or less than that of the parent compound," according to product labeling. Its metabolite P95 possesses a high affinity for 5-HT$_{2A}$, α$_{1A}$, α$_{1B}$, α$_{1D}$, and α$_{2C}$ receptors only.

Design and Interpretation of Clinical Drug Studies

ACUTE TREATMENT TRIALS. The efficacy of iloperidone was supported by two large trials—one 4 weeks and one 6 weeks in duration. In the 4-week trial, 604 participants with DSM-IV schizophrenia were randomized to receive iloperidone 24 mg daily (titrated over the course of 1 week, with twice-daily dosing), ziprasidone (160 mg daily), or placebo. The primary endpoint was the change in PANSS total score from baseline to end of treatment (day 28). In this study, both iloperidone and ziprasidone were statistically superior to placebo in improving PANSS scores. Iloperidone was noted to have a different adverse event profile than ziprasidone, with iloperidone having less sedation, somnolence, EPS, akathisia, agitation, and restlessness and causing more weight gain, tachycardia, orthostatic hypotension, dizziness, and nasal congestion. In the 6-week trial, 706 participants with DSM-IV schizophrenia were randomized to receive either one of two flexible-dose ranges of iloperidone (12 to 16 mg/day or 20 to 24 mg/day), risperidone (6 to 8 mg/day), or placebo. The primary endpoint was the change in BPRS total score from baseline to end of treatment (day 42). Both dose ranges were superior to placebo in improving BPRS scores. While the efficacy of risperidone appeared greater than iloperidone within the first 2 weeks of treatment, the efficacy between treatments was similar in patients who remained on the treatment for at least 2 weeks. This differential efficacy in the first 2 weeks was attributed to the slower dose titration schedule necessary for iloperidone.

Two additional large short-term studies were performed to assess the efficacy of iloperidone for the treatment of schizophrenia or schizoaffective disorder. In the first study (6 weeks), iloperidone 12 mg/day (but not 4 mg or 8 mg/day) statistically improved the PANSS and BPRS total scores as compared to placebo. In the second 6-week study, both iloperidone 4 to 8 and 10 to 16 mg/day outperformed placebo for improving the BPRS total score. For the PANSS, the 4 to 8 mg/day group outperformed placebo only on the positive symptom subscale, while the 10 to 16 mg/day group outperformed placebo on the PANSS total as well as positive and negative symptom subscales.

MAINTENANCE TREATMENT TRIALS. While iloperidone is currently only indicated “for the treatment of adults with schizophrenia,” based on two large short-term studies, limited long-term maintenance studies have been performed. Kane et al. reported the data from three trials. Subjects from three short-term studies (N = 1,644), who had experienced at least a 20% reduction in PANSS total score in the initial 6-week study, were randomized to receive iloperidone 4 to 16 mg/day or haloperidol 5 to 20 mg/day for 46 weeks. The primary outcome measure was the time to relapse, defined as a ≥25% increase in PANSS total score, discontinuation due to lack of efficacy, aggravated psychosis with hospitalization, or a 2-point increase on the CGI-C after week 6. Iloperidone was found to be statistically equivalent to haloperidol for relapse prevention.

Therapeutic Indications

SCHIZOPHRENIA. Iloperidone is indicated for the treatment of schizophrenia, with efficacy established based on two short-term studies.

Precautions and Adverse Reactions

BLACK BOX WARNING. Iloperidone and the other SGAs carry a black box warning for increased mortality in older adult patients with dementia-related psychosis compared with placebo.

QT PROLONGATION. Iloperidone, at the dose of 12 mg twice daily, was found to increase the QTc interval by 9 ms in an open-label study of 160 individuals. This QTc prolongation was increased to 19 ms in the presence of a CYP2D6 inhibitor (paroxetine) or CYP3A4 inhibitor (ketoconazole). It is recommended that iloperidone be avoided in combination with other medications known to increase the QTc interval in individuals with congenital long QT syndrome, history of cardiac arrhythmias, or histories of significant cardiovascular illness. It is also recommended that individuals at risk of electrolyte disturbances have potassium and magnesium levels monitored periodically, as low levels can increase the risk of QTc prolongation and arrhythmia. Iloperidone should be discontinued if QTc is found to be greater than 500 ms.

NEUROLEPTIC MALIGNANT SYNDROME. NMS has been associated with the use of antipsychotic medications, including iloperidone. If NMS is suspected, iloperidone and other antipsychotic medications should be discontinued immediately, and the patient should undergo intensive supportive care and medical monitoring.

EXTRAPYRAMIDAL SYMPTOMS. Iloperidone is associated with a low risk of EPS, including akathisia, bradykinesia, dyskinesia, dystonia, parkinsonism, and tremor. When considering all EPS events in four short-term trials, 15.1% of subjects taking iloperidone 20 to 24 mg/day (*N* = 391), 13.5% of subjects taking 10 to 16 mg/day (*N* = 483), and 11.6% of subjects taking placebo (*N* = 587) reported any type of EPS. Parkinsonism was reported in 0.3% of subjects taking iloperidone 20 to 24 mg/day as compared to 0% in subjects taking placebo.

TARDIVE DYSKINESIA. The propensity of iloperidone to cause tardive dyskinesia is not known but is expected to be low due to the minimal extrapyramidal side effects of this medication.

METABOLIC CHANGES. Iloperidone has been associated with hyperglycemia, dyslipidemia, and weight gain, and patients should be monitored accordingly. In a pooled analysis of longer-term trials, iloperidone 10 to 16 mg/day was associated with a 1.8 mg/dL glucose increase at 3 to 6 months ($N = 773$) and a 5.4 mg/dL glucose increase at >12 months ($N = 425$). Iloperidone 10 to 16 mg/day was associated with a −3.9 mg/dL change in total cholesterol and −8.9 mg/dL change in triglycerides at 3 to 6 months ($N = 783$) and −7.7 mg/dL change in total cholesterol and −17.7 mg/dL change in triglycerides at >12 months ($N = 428$). Of note, the observed decrease in lipid parameters could be a consequence of the participants switching to iloperidone from another antipsychotic with a higher propensity for dyslipidemia. Iloperidone has also been associated with weight gain in a dose-dependent manner, with 18% of subjects taking iloperidone 20 to 24 mg/day and 12% of those taking 10 to 16 mg/day gaining ≥7% of body weight at 4 to 6 weeks.

SEIZURES. In 4- to 6-week placebo-controlled trials, iloperidone was associated with a seizure in 1 of 1,344 patients treated with iloperidone (0.1%) and 2 of 587 patients treated with placebo (0.3%).

ORTHOSTATIC HYPOTENSION AND SYNCOPE. Iloperidone has been associated with orthostatic hypotension with dizziness, tachycardia, and syncope, and this is thought to be related to its property as an α1-antagonist. In short-term studies where the dose was increased slowly, as per label recommendations, orthostatic hypotension occurred in 5% of patients taking iloperidone 20 to 24 mg/day, 3% taking iloperidone 10 to 16 mg/day, and 1% of patients taking placebo. It is noted that more rapid increases in iloperidone dose would be expected to increase the risk of orthostatic hypotension and syncope.

HYPERPROLACTINEMIA. In a 4-week placebo-controlled trial of iloperidone, prolactin increased by 2.6 ng/mL in the group taking iloperidone 24 mg/day, as compared to a decrease of 6.3 ng/mL in the placebo group. In longer-term trials ($N = 3{,}210$), 0.1% of males taking iloperidone developed gynecomastia as compared to 0% in the placebo group, and 0.2% of females taking iloperidone developed galactorrhea as compared to 0.5% in the placebo group.

PRIAPISM. Three cases of priapism were observed in the drug development process, according to the manufacturer's prescribing information. This is thought to be related to the α-antagonist properties of iloperidone.

PREGNANCY. Iloperidone is currently designated "category C" for use in pregnant women, meaning that animal studies have shown an adverse effect on the fetus, there are no adequate and well-controlled studies in humans, and potential benefits may warrant use despite potential risks. In pregnant rat studies, a dose of iloperidone 26 times the maximum recommended human daily dose (MRHD) (in mg/kg/day) caused increased intrauterine deaths, decreased fetal weight and length, decreased fetal skeletal ossification, and increased incidence of minor fetal skeletal anomalies. In pregnant rabbits, a dose of iloperidone 20 times the MRHD increased early intrauterine deaths and decreased fetal viability at term. In pregnant rats, doses of 1.6 to 4.8 times the MRHD caused no adverse effects on the pups except for an increase in stillbirth rates at the dose of 1.6 times MRHD. The iloperidone metabolite P95 was given to pregnant rats at 20, 80, and 200 mg/kg/day during the period of organogenesis, and no teratogenic effects were seen; however, delayed skeletal ossification occurred.

NURSING MOTHERS. While iloperidone has been found to be excreted in the milk of lactating rats, it is not known if this occurs in humans. It is currently recommended that women taking iloperidone should not breastfeed.

Drug Interactions. Iloperidone dose should be reduced by half when administered along with strong CYP2D6 inhibitors (e.g., fluoxetine and paroxetine), CYP3A4 inhibitors (e.g., ketoconazole and clarithromycin), or poor CYP2D6 metabolizers. If CYP2D6 or CYP3A4 inhibitors are discontinued, the iloperidone dose can be increased accordingly.

Dosage and Administration. Iloperidone is available in tablets of 1, 2, 4, 6, 8, 10, and 12 mg. Iloperidone should be uptitrated slowly to minimize orthostatic hypotension side effects, given its α-adrenergic antagonist properties. It is recommended to start at 1 mg twice daily, increasing the dose (by no more than 4 mg total/day) to a target dose of 6 to 12 mg twice daily (12 to 24 mg total/day). Efficacy in pivotal trials is demonstrated in the range of 6 to 12 mg twice daily. Iloperidone does not need to be administered with meals.

DOSING IN SPECIAL POPULATIONS

Pediatric Use. Safety and efficacy data supporting the use of iloperidone in child and adolescent patients have not been established.

Geriatric Use. Safety and efficacy data supporting the use of iloperidone in geriatric populations has not been established. Only 25 of 3,210 participants of iloperidone premarketing trials were 65 years old or greater.

Renal Impairment. Iloperidone undergoes almost total hepatic metabolism, with <1% excreted unchanged. Creatinine clearance <30 mL/min had minimal effect on C_{max} of iloperidone or its P88 and P95 metabolites. However, the AUC of iloperidone increased by 24%, P88 decreased by 6%, and P95 increased by 52%, so prescribers should monitor patients with renal impairment with greater vigilance.

Hepatic Impairment. No dosing adjustments to iloperidone are needed for patients with mild hepatic impairment. It is recommended that prescribers exercise caution when administering iloperidone to patients with moderate hepatic impairment and not prescribe iloperidone to patients with severe hepatic impairment. Specifically, in patients with mild hepatic impairment, there were no significant differences in the pharmacokinetics of iloperidone or its major metabolites as compared to healthy controls. With moderate hepatic impairment, a twofold higher exposure to the P88 metabolite was observed as compared to healthy controls. Adequate data is not available for patients with severe hepatic impairment.

SEROTONIN, DOPAMINE, NOREPINEPHRINE

Asenapine

Asenapine is an antipsychotic with a high affinity for 5-HT_{2A} and D_2 receptors, as well as several other serotonergic, dopaminergic, and adrenergic receptors. Important clinical properties include sublingual administration (with temporary oral hypoesthesia as a possible side effect) and minimal anticholinergic side effects. It is associated with mild to moderate sedation (roughly equivalent to olanzapine and haloperidol) and mild changes in body weight.

Asenapine was initially developed by Organon. It was approved by the FDA in 2009, with indications for the acute treatment of both schizophrenia and bipolar I disorder (manic or mixed episodes). In 2010, the approval indications were expanded to include maintenance treatment of schizophrenia as well as use as adjunctive therapy (to lithium or valproate) for the acute treatment of manic or mixed episodes in bipolar I disorder. In 2019, a once-daily transdermal asenapine patch was approved by the U.S. FDA. It is currently the only antipsychotic available in a transdermal formulation.

Chemistry. Asenapine belongs to the class of dibenzo-oxepino pyrroles (Fig. 33.2b–1). It was derived chemically by modifying the structure of the tetracyclic antidepressant, mianserin. The chemical name of asenapine is (3aRS,12bRS)-5-Chloro-2-methyl-2,3,3a,12b-tetrahydro-1Hdibenzo[2,3:6,7]oxepino[4,5-c]pyrrole (2Z)-2-butenedioate (1:1). Its molecular formula is $C_{17}H_{16}ClNO \cdot C_4H_4O_4$ and its molecular weight is 401.84.

Pharmacologic Action

PHARMACOKINETICS

Absorption and Distribution. Oral asenapine was developed as a sublingual formulation to avoid substantial first-pass metabolism when swallowed. It is 35% bioavailable by this route. Drinking water 2 to 5 minutes after asenapine administration decreases absorption, so it is recommended that eating and drinking are avoided for 10 minutes after administration. Consumption of food immediately prior to sublingual administration decreases asenapine exposure by 20%, and consumption of food 4 hours after administration decreases exposure by 10%; this is due to increased hepatic blood flow. Sublingual asenapine reaches maximal plasma concentration approximately 1 hour after administration and, with twice-daily dosing, reaches steady state within 3 days.

Transdermal asenapine is gradually released with a T_{max} of about 16 hours and a half-life of 30 hours. Absorption of the transdermal form is not affected by food or water intake.

Asenapine is Rapidly Distributed, Has a Volume of Distribution of Approximately 20 to 25 L/kg, and is Highly Protein Bound in Plasma

Metabolism and Elimination. The primary metabolic pathways for asenapine are glucuronidation by UGT1A4 and oxidative metabolism by CYP isozymes (primarily CYP1A2; CYP3A4, and CYP2D6 to a lesser extent). Asenapine is a weak inhibitor of CYP2D6. Though tobacco smoking induces CYP1A2, the clearance of asenapine in smokers has not been found to be altered. After single-dose administration, 50% of labeled asenapine metabolites were recovered in urine and 40% in the stool. The mean terminal half-life of asenapine is approximately 24 hours.

PHARMACODYNAMICS

Mechanism of Action. Asenapine has a high affinity for serotonin 5-HT_{1A}, 5-HT_{1B}, 5-HT_{2A}, 5-HT_{2B}, 5-HT_{2C}, 5-HT_{5A}, 5-HT_6, and 5-HT_7 receptors; dopamine D_{2A}, D_{2B}, D_3, D_4, and D_1 receptors; adrenergic α_{1A}, α_{2A}, α_{2B}, and α_{2C} receptors, and histamine H1 receptors, acting as an antagonist at all of these receptors. It has moderate affinity for histamine H2 receptors and minimal affinity for muscarinic receptors. Its activity is mostly due to parent asenapine rather than metabolites.

Design and Interpretation of Clinical Drug Studies

SCHIZOPHRENIA. The efficacy of sublingual asenapine for acute treatment of schizophrenia was supported by two of three 6-week randomized controlled trials of adults with DSM-IV schizophrenia who were having an illness exacerbation. The primary outcome measure in all three studies was the change in PANSS total score. In the first study (*N* = 174), asenapine 5 mg twice daily showed statistically significant improvement on the primary outcome measure, as well as the PANSS positive, negative, and general psychopathology subscales, as compared to placebo. The active comparator risperidone did not separate from placebo. In the second study (*N* = 448), asenapine 5 mg but not 10 mg was statistically superior to placebo on the primary outcome measure. In the third study (*N* = 386), which was considered negative, neither 5- nor 10-mg asenapine were superior to placebo, as compared to olanzapine 15 mg, which was superior to placebo.

The efficacy of transdermal asenapine was demonstrated in a double-blind trial (N = 616) that compared two doses of asenapine, 7.6 or 3.8 mg/24 hours and placebo. Both active doses separated from the placebo at week 2, and the effect persisted for the 6-week trial.

The efficacy of sublingual asenapine for the maintenance treatment of schizophrenia was demonstrated in a randomized withdrawal design trial. Seven hundred patients with schizophrenia entered open-label asenapine treatment for 26 weeks. Three hundred eighty-six patients who met stability criteria were randomized to either continue asenapine or withdraw to placebo. The primary outcome measure was time to relapse. Asenapine was statistically superior to placebo in increasing time to relapse as well as impending relapse (defined by prespecified changes in PANSS and CGI-S scores).

BIPOLAR DISORDER. The efficacy of sublingual asenapine in treating acute mania was established in two trials in adults with DSM-IV bipolar I disorder who were having acute manic or mixed episodes. After a 7-day placebo washout period, patients were randomized to receive asenapine 5 or 10 mg twice daily, olanzapine 5 to 20 mg/day, or a placebo. The primary outcome measure was the change in YMRS from baseline to day 21. Both asenapine and olanzapine were statistically superior to placebo on the primary outcome measure as well as the CGI-BP Severity of Illness score.

The efficacy of sublingual asenapine as adjunctive therapy in treating acute mania was established in a 12-week placebo-controlled trial. Patients experiencing manic or mixed episodes of bipolar I disorder with only partial response to lithium or valproic acid monotherapy were randomized to receive asenapine (5 to 10 mg twice daily, *N* = 154) or placebo (*N* = 162). Adjunctive asenapine was statistically superior to placebo at week 3 on the primary outcome measure (change in YMRS score).

Therapeutic Indications

SCHIZOPHRENIA. Sublingual asenapine is approved for the acute and maintenance treatment of schizophrenia. Transdermal asenapine is approved for schizophrenia.

BIPOLAR DISORDER

Monotherapy. Sublingual asenapine is indicated for the acute treatment of manic or mixed episodes associated with bipolar I disorder. Efficacy was established in two 3-week monotherapy trials.

Adjunctive Therapy. Sublingual asenapine is indicated as adjunctive therapy with either lithium or valproate for the acute treatment of manic or mixed episodes associated with bipolar I disorder. Efficacy was established in one 3-week trial.

Precautions and Adverse Reactions

BLACK BOX WARNING. Asenapine carries a black box warning for increased mortality in older adult patients with dementia-related psychosis.

QT PROLONGATION. Asenapine has been specifically assessed for its effect on QT interval at doses of 5, 10, 15, and 20 mg twice daily versus placebo ($N = 151$). Asenapine was associated with an increase in QTc of 2 to 5 ms as compared to placebo, and no patients showed greater than 60 ms increase or reached QTc ≥500 ms. In other asenapine premarketing clinical trials, there were no differences in QTc ≥500 ms between asenapine and placebo-treated patients, and there were no reports of Torsade de Pointes or other related events. It is recommended, however, that asenapine should not be combined with other medications known to increase the QTc interval.

NEUROLEPTIC MALIGNANT SYNDROME. NMS has been associated with the use of antipsychotic medications, including asenapine. If NMS is suspected, asenapine and other antipsychotic medications should be discontinued immediately, and the patient should undergo intensive supportive care and medical monitoring.

TARDIVE DYSKINESIA. The propensity of asenapine to cause tardive dyskinesia is not known.

METABOLIC CHANGES. In pooled 6-week trials, 5 to 10 mg asenapine twice daily ($N = 377$) was associated with a 3.2 mg/dL increase in glucose, as compared to −0.2 mg/dL in the placebo group ($N = 232$). In a 52-week trial of asenapine, with comparator control, the mean increase of fasting blood glucose was 2.4 mg/dL.

Dyslipidemia has also been associated with asenapine use. In pooled 6-week trials, 5- to 10-mg asenapine twice daily was associated with a 0.4 mg/dL increase in total cholesterol ($N = 539$), 1.3 mg/dL increase in LDL ($N = 465$), 0.5 mg/dL increase in high-density lipoprotein (HDL) ($N = 480$), and 3.8 mg/dL increase in triglycerides ($N = 380$). In the same group of pooled trials, the placebo group showed a 2.2 mg/dL decrease in total cholesterol ($N = 351$), 0.1 mg/dL increase in LDL ($N = 285$), 0.5 mg/dL increase in HDL ($N = 290$), and 7.6 mg/dL decrease in triglycerides ($N = 233$). In a 52-week trial of asenapine, with comparator control, asenapine was associated a 6 mg/dL decrease in total cholesterol and a 9.8 mg/dL decrease in triglycerides. Of note, these observed decreases in lipid parameters over 52 weeks could be a consequence of the study participants switching to asenapine from another antipsychotic with a higher propensity for dyslipidemia.

Asenapine may also affect body weight. In pooled 6-week studies, asenapine 5 to 10 mg twice daily ($N = 532$) was associated with a 1.1 kg increase in body weight as compared to a mean change of 0 kg in the placebo group ($N = 348$). In a comparator-controlled 52-week trial ($N = 887$), asenapine was associated with a 0.9 kg increase in body weight, with 14.7% of participants increasing ≥7% of their baseline body weight. The weight gain was most pronounced in participants with a baseline BMI of <23.

ORTHOSTATIC HYPOTENSION AND SYNCOPE. Asenapine possesses α1-adrenergic antagonist properties and has been reported to cause orthostatic hypotension and syncope in some patients, particularly early in treatment. In 6-week schizophrenia trials, 1/527 patients taking asenapine experienced syncope, as compared to 1/378 of patients taking placebo. In 3-week bipolar disorder trials, 1/379 patients taking asenapine experienced syncope, as compared to no patients taking placebo. When including all participants in premarketing clinical trials (including 52-week treatment), syncope was reported in 11/1,953 (0.6%) of patients taking asenapine.

HYPERPROLACTINEMIA. As with most antipsychotic medications, asenapine can increase prolactin levels. In premarketing clinical trials for schizophrenia and bipolar disorder, adverse events related to hyperprolactinemia were found in 0.4% of subjects in the asenapine group, as compared to 0% in the placebo group.

LEUKOPENIA/NEUTROPENIA. Leukopenia and neutropenia events have been infrequently reported to be temporally associated with asenapine use. It is recommended that patients with low WBC or a history of drug-induced leukopenia or neutropenia have their blood counts frequently monitored during the first few months of treatment and that asenapine be discontinued if a WBC decrease is observed that cannot be explained by other factors. Patients with ANC $<1{,}000/\text{mm}^3$ should discontinue asenapine until it has recovered.

POTENTIAL FOR COGNITIVE AND MOTOR IMPAIRMENT. In short-term premarketing schizophrenia trials, somnolence occurred in 15% of patients taking sublingual asenapine 5 mg twice daily (N = 274) and 13% of patients taking 10 mg twice daily (N = 208), as compared to 7% of patients taking placebo (N = 389). Somnolence was noted to be most significant during the first week of treatment and led to discontinuation in 0.6% of all patients in short-term placebo-controlled trials (N = 1,953).

DYSPHAGIA. Dysphagia was reported in 0.2% of patients with schizophrenia who were taking asenapine (N = 572), as compared to 0% of patients taking placebo (N = 378). There were no reported incidents of dysphagia in premarketing trials for bipolar mania.

PREGNANCY. Asenapine is currently designated category C, indicating that there are no adequate and well-controlled studies of asenapine in pregnant women. In animal studies, asenapine increased postimplantation loss and decreased rodent pup weight and survival though it was not found to cause other structural abnormalities. It was found to be nonteratogenic in rat and rabbit studies. It is recommended to be only used during pregnancy if the potential benefit justifies the potential risk to the fetus.

LACTATION. Asenapine is excreted in the milk of rats during lactation. It is unknown if this also occurs in humans. It is recommended that women who receive asenapine not breastfeed.

ORAL HYPOESTHESIA. Oral hypoesthesia is an adverse reaction associated with sublingual asenapine. This can occur immediately after sublingual administration and usually resolves within 1 hour. Oral hypoesthesia was associated with asenapine discontinuation in short-term premarketing trials for bipolar mania, occurring in 1.1% of the asenapine group patients (N = 379) and 0% of the placebo group (N = 203). In postmarketing monitoring of asenapine use, other oral reactions have been reported to have occurred with asenapine sublingual administration, including oral ulcers, blisters, peeling/sloughing, and inflammation.

SKIN REACTIONS. About one-sixth of patients receiving transdermal asenapine in a clinical trial developed mild to moderate skin reactions that included erythema, pruritus, and irritation. Fewer than 0.5% stopped the medication due to these reactions.

Drug Interactions. Asenapine is primarily metabolized by UGT1A4 (direct glucuronidation) and by several CYP isozymes (mostly CYP1A2). As such, administration of asenapine 5 mg with the CYP1A2 inhibitor fluvoxamine (25 mg twice daily) for 8 days increased the asenapine AUC by 29%; caution is indicated for coadministration of asenapine and fluvoxamine. Coadministration with the CYP2D6 inhibitor paroxetine was associated with a 9% decrease in the asenapine AUC. As asenapine is a weak inhibitor of CYP2D6 itself, it is recommended to be used with caution with drugs that are both substrates and inhibitors of CYP2D6 (such as paroxetine). Though valproic acid is an inhibitor of UGT1A4, coadministration of asenapine with valproic acid 500 mg twice daily did not significantly change asenapine levels.

Dosage and Administration. Asenapine is available as sublingual tablets of 5 or 10 mg. Patients should place the tablet under the tongue and allow it to completely dissolve, which occurs within seconds. Product labeling indicates asenapine tablets should not be split, crushed, chewed, or swallowed, and patients should not eat or drink for 10 minutes after administration.

For the treatment of acute schizophrenia, the recommended starting and target dose of asenapine is 5 mg twice daily. Based on clinical response, the dose can be increased to 10 mg twice daily if tolerated. For maintenance treatment of schizophrenia, the starting dose is recommended to be 5 mg twice daily, and in 1 week, it can be increased up to 10 mg twice daily if tolerated.

Transdermal asenapine is available as 3.8-, 5.7-, and 7.6-mg patches. Patients treated with transdermal asenapine for schizophrenia should be started on a 3.8-mg patch daily. The dose can be increased to 5.7 or 7.6 mg daily after 1 week. The patch can be applied to the abdomen, upper back, arms, or hips. The patch should be changed daily, and the location should be rotated. It should not be applied to broken skin.

For the acute treatment of manic or mixed episodes associated with bipolar I disorder, the recommended starting dose is 10 mg twice daily. If not tolerated, the dose can be decreased to 5 mg twice daily. For adjunctive therapy with either lithium or valproate, the starting dose of asenapine is recommended to be 5 mg twice daily and based on clinical response, the dose can be increased to 10 mg twice daily if tolerated. Asenapine is not currently indicated for the maintenance treatment of bipolar disorder.

Asenapine administration is contraindicated in patients with a known hypersensitivity to the medication. Hypersensitivity reactions in clinical trials have included anaphylaxis, angioedema, hypotension, tachycardia, swollen tongue, dyspnea, wheezing, and rash, and some of these cases occurred after the first dose.

DOSING IN SPECIAL POPULATIONS

Geriatric Use. Only 25 of 2,250 (1.1%) premarketing trial patients receiving asenapine were 65 years or older. In older adult patients, the asenapine AUC was on average 40% higher than younger than adults, so older adult patients taking asenapine should be monitored carefully.

Hepatic Impairment. Hepatic impairment has been found to increase the exposure to asenapine, according to the extent of hepatic impairment. Severe hepatic impairment (Child–Pugh C) was associated with a sevenfold increase in asenapine concentrations. Dosage adjustments are not required for mild (Child–Pugh A) or moderate (Child–Pugh B) hepatic impairment, but asenapine is not recommended for patients with severe hepatic impairment.

Renal Impairment. Renal impairment (mild, moderate, or severe, but not on dialysis) was not found to affect exposure to asenapine, so dosage adjustment is not required.

DOPAMINE, SEROTONIN, SEROTONIN TRANSPORTER

Lumateperone is the only antipsychotic in this group which is characterized by antagonism of D_2, 5-HT_{2A}, and inhibition of the serotonin transporter (SERT).

Lumateperone was developed by Intra-Cellular Therapies (ITC) as part of an agreement in 2005 to develop and commercialize a family of central nervous system (CNS) compounds from Bristol-Myers Squibb. The Investigational New Drug (IND) application for lumateperone was submitted in October 2007, with the first clinical study initiated as an open-label study on receptor occupancy. Discussions of phase III clinical trials for schizophrenia were held with the FDA in June 2014. In September 2017, ITC requested both Fast Track and Breakthrough Therapy Designations based on the claim that lumateperone was better tolerated than existing drug treatments for schizophrenia. Due to safety questions which will be discussed below, the FDA denied Breakthrough Status but granted the Fast Track request.

The clinical development of lumateperone was marked by several issues which complicated the approval process. In establishing efficacy, only two of the three registration trials demonstrated a statistical advantage of lumateperone over placebo. Several doses were studied across the three trials ranging from 14 mg to 84 mg. Only the 42-mg dose separated from placebo in the two positive trials. As a class, antipsychotics generally demonstrate a dose–response relationship, with higher doses achieving greater efficacy than lower doses. Therefore, for only the middle dose to achieve a positive signal was unusual. Failure of the higher dose could not be explained by a greater incidence of side effects.

Chemistry

Lumateperone was derived from a class of tetracyclic butyrophenones with potent binding at serotonin 5-HT_{2A} and dopamine D_2 receptors. The chemical name of lumateperone is 1-(4-fluoro-phenyl)-4-((6bR, 10aS)-3-methyl-2,3,6b,9,10,10a-hexahydro-1H,7H-pyrido[3′,4′:4,5]pyrrolo[1,2,3-de]quinoxalin-8-yl)-butan-1-one. Its chemical formula is $C_{24}H_{28}FN_3O$ and its molecular weight is 393.506.

Pharmacologic Action

Pharmacokinetics. Steady state is reached in 5 days with once-daily dosing. An increase in steady-state exposure is seen in a dose-proportional manner in the range of 21 mg to 56 mg. Half-life is 13 to 21 hours.

ABSORPTION. Absorption is rapid, with a bioavailability of 4.4%. C_{max} is achieved in 1 to 2 hours with once-daily dosing, with a mean T_{max} of 3 to 4 hours. Ingestion with food is recommended to decrease GI effects as high-fat meals can decrease by C_{max} 33%.

DISTRIBUTION. The volume of distribution following intravenous administration is about 4.1 L/kg.

METABOLISM AND ELIMINATION. Lumateperone is metabolized into more than 20 metabolites by multiple enzymes, including uridine 5′-diphospho-glucuronosyltransferases (UDP-glucuronosyltransferase, UGT) 1A1, 1A4, and 2B15, aldoketoreductase (AKR)1C1, 1B10, and 1C4, and CYP 3A4, 2C8, and 1A2.

In radioactive labeled oral dose studies, 58% of the dose was recovered in the urine and 29% recovered in feces. Less than 1% of the dose was excreted in the urine as unchanged lumateperone. There is a postmarketing requirement for a clinical trial evaluating the effect of UGT inhibitors on the pharmacokinetics of lumateperone and its metabolites. Presently, there is a recommendation to avoid the concomitant use of lumateperone and UGT inhibitors.

Pharmacodynamics. Lumateperone demonstrates activity at multiple receptors, with clinical activity purportedly due to binding at 5-HT_{2A}, D_2, and SERT sites. It has the highest affinity for 5-HT_{2A} and moderate affinity at D_2 receptors and SERTs. Moderate binding affinity is also demonstrated at D1, D4, and adrenergic α1 and α4 receptors. There is a low binding affinity for muscarinic and histaminergic receptors. Indirect modulation of glutamatergic activity is proposed through increasing phosphorylation of the GluN2B subunit of the N-methyl-D-aspartate channels from the mammalian target of rapamycin (mTOR) protein pathway.

Mechanism of Action

The precise mechanism of lumateperone in treating schizophrenia is unknown. It is proposed that efficacy is mediated through activity at several receptors, including antagonist activity at central serotonin 5-HT_{2A} receptors and postsynaptic antagonism at central dopamine D_2 receptors. There is also a relatively high affinity to the SERT, although how this directly contributes to antipsychotic activity is not clear. While several peer-reviewed articles from the manufacturer and their official drug application to the FDA claim partial agonism at presynaptic D_2 receptors and glutamate modulation, the final FDA postapproval review of the drug application data concluded there was not sufficient support for these claims. The FDA's evaluation concluded that while there is some *in vivo* evidence of partial agonism at presynaptic D_2 receptors, it was not sufficient to overcome the more compelling *in vitro* data that it is an antagonist at these receptors. The FDA also concluded there was no evidence to suggest lumateperone substantially modulates glutamate any differently compared to existing antipsychotics. Therefore, based on the NbN system, it is most similar to other DSAs, although with additional activity at the SERT.

Design and Interpretation of Clinical Studies

Acute Treatment of Schizophrenia. A total of three short-term, randomized, placebo-controlled phase II/III clinical trials were completed in the evaluation of lumateperone for the acute treatment of schizophrenia. Across the three trials, four doses of lumateperone were evaluated—14 mg, 28 mg, 42 mg, and 84 mg. Interestingly, only the 42 mg dose consistently demonstrated a statistical advantage over the placebo. Given that the 84 mg dose did not perform better than a placebo, the FDA review highlighted their concern about the lack of dose response. Antipsychotics as a class generally demonstrate a dose–response relationship, with higher doses yielding better responses than lower doses.

Special Populations

PREGNANCY AND LACTATION. Animal studies of lumateperone administered at multiple times the maximum doses to pregnant rats and rabbits did not result in malformations. However, when the pregnant animals were treated with a human metabolite of lumateperone, there was an increase in cleft palate and skeletal malformations. Presently, there is no data on lumateperone and its metabolite in animal or human mild. There have been no studies of human neonates or infants exposed to lumateperone, and therefore it is unclear if their pharmacokinetics are similar to those of adults. As cited earlier, given the observation of aniline metabolites accumulating in animal models, there is a theoretical concern that neonates and infants may not metabolize or eliminate lumateperone as adults, which could lead to the accumulation of these metabolites. Thus, breastfeeding is not recommended during treatment with lumateperone.

PEDIATRIC USE. The clinical trials did not include subjects under the age of 18. At the time of publication, there was no data regarding the safety and efficacy of lumateperone in pediatric patients. Postmarketing commitments and requirements outline three studies for patients aged 13 to 17 years to determine the safety, efficacy, and long-term safety of lumateperone. There are no known plans to evaluate lumateperone in children 12 years of age or younger.

GERIATRIC USE. No subjects aged 65 years or older were included in the clinical trials of lumateperone. Antipsychotics have a class warning of increased risk for death in older adult patients with dementia-related psychosis.

RENAL FAILURE. No dosage adjustments are required for chronic kidney disease or renal failure.

HEPATIC FAILURE. Patients with moderate or severe hepatic impairment (Child–Pugh B or C) should not use lumateperone as they experience higher exposure to lumateperone. Mild hepatic impairment (Child–Pugh A) does not require dose adjustment.

Therapeutic Indications

Schizophrenia. Lumateperone is indicated for the treatment of patients with schizophrenia. Its efficacy at the 42-mg dose was established in three short-term studies of adults with schizophrenia. Lumateperone is not indicated at the time of writing for longer-term use (greater than 6 weeks).

Dosing and Administration

Lumateperone is only available as a 42-mg capsule. No dose titration is required. A long-acting injectable form is under development.

For adults starting lumateperone, only the 42-mg dose has been shown to be effective for schizophrenia. Both lower, 28-mg, and higher, 84-mg, doses did not separate from placebo in clinical trials. Therefore, the patient should be started on and maintained at 42 mg. No dose titration is required.

Precautions and Adverse Reactions

Black Box Warning. Lumateperone and the other SGAs carry a black box warning for increased mortality in older adult patients with dementia-related psychosis compared with placebo.

QT Prolongation. The potential for QT prolongation was formally evaluated in a four-arm crossover study including placebo, active control (moxifloxacin 400 mg), 42 mg lumateperone, and 126 lumateperone. In that study, daily oral administration for 5 days yielded a placebo-corrected QTcF interval change from baseline of 4.9 ms for 42 mg lumateperone and 15.8 ms for the 126-mg dose. In the three 4- to 6-week placebo-controlled studies, no clinically meaningful change in QT interval was noted, although patients on the 84-mg dose did show a higher mean change from baseline in QTcB and QTcF. There appears to be a dose-related QTc increase based on the 84-mg and 126-mg data. The FDA does not recommend lumateperone in high-exposure situations, such as with concomitant CYP3A4 inhibitors or in patients with hepatic impairment. Therefore, the only FDA-recommended dose of 42 mg should not yield concerning QTc prolongation. However, if used in the context of reduced medication metabolism or doses higher than 42 mg, QT prolongation could become clinically relevant.

Neuroleptic Malignant Syndrome. Lumateperone has a drug class warning of NMS, although no cases occurred in the clinical trials or have been reported in the literature at the time of this writing.

Extrapyramidal Symptoms. EPS, including akathisia, dystonia, and parkinsonism, are known adverse reactions to antipsychotic drugs. There were no meaningful differences in EPS prevalence compared to baseline in any of the placebo-controlled short-term clinical studies.

Tardive Dyskinesia. Tardive dyskinesia can develop in all patients treated with antipsychotic drugs. Its prevalence is highest among the older adults. The propensity of lumateperone to cause tardive dyskinesia is not known. Two cases of tardive dyskinesia were reported in subjects taking 42 mg in a 6-week study. The FDA clinical reviewer concluded that it was highly unlikely that a new antipsychotic would cause tardive dyskinesia within 6 weeks of treatment. They concluded that these were more likely withdrawal dyskinesias that can occur with a change or discontinuation of antipsychotics. One case of treatment-emergent tardive dyskinesia was reported with the 42-mg dose in the 1-year uncontrolled study.

Metabolic Changes. Across the three short-term 4- to 6-week studies for lumateperone, there was no clear pattern of adverse metabolic effects. Parameters at baseline and end of treatment were collected for glucose, hemoglobin A1c, insulin, total cholesterol, high-density lipoproteins, LDLs, triglycerides, prolactin, ALT, aspartate aminotransferase, creatinine kinase, and body weight. Data was analyzed as both mean group level changes and shifts in change from normal to abnormal lab values. In regards to weight gain, 9.2% of subjects receiving placebo and 9.1% of subjects on 42 mg lumateperone gained ≥7% of their body weight from baseline in data pooled across all three studies.

Nursing Mothers. Breastfeeding is not recommended during treatment with lumateperone. While there is no data on the presence of lumateperone or its metabolites in human milk, infants may be exposed to antipsychotics through breast milk. Of concern is the accumulation of toxic aniline metabolites in animal models exposed to lumateperone. These metabolites could not be detected in adult humans due to different drug metabolism and elimination pathways.

However, it is unknown if human infants exposed to lumateperone would metabolize and eliminate the drug as observed in adults. A postmarketing lactation study is being required to assess the concentration of lumateperone and its metabolites in breast milk.

Pregnancy. There was no data from the registration trials on the effects of lumateperone in pregnancy, as women who were pregnant or unable to use contraception were excluded from the trials.

The collection of human data on the effects of lumateperone is ongoing through a postmarketing registry.

FURTHER READINGS

*Buchanan RW, Kreyenbuhl J, Kelly DL, et al; Schizophrenia Patient Outcomes Research Team (PORT). The 2009 schizophrenia PORT psychopharmacological treatment recommendations and summary statements. *Schizophr Bull*. 2010;36:71–93.

Citrome L, Walling DP, Zeni CM, et al. Efficacy and safety of HP-3070, an asenapine transdermal system, in patients with schizophrenia: a phase III, randomized, placebo-controlled study. *J Clin Psychiatry*. 2020;82(1):20m13602.

Cohen LS, Viguera AC, McInerney KA, et al. Reproductive safety of second-generation antipsychotics: current data from the Massachusetts general hospital national pregnancy registry for atypical antipsychotics. *Am J Psychiatry*. 2016;173:263–270.

Correll CU, Davis RE, Weingart M, et al. Efficacy and safety of lumateperone for treatment of schizophrenia: a randomized clinical trial. *JAMA Psychiatry*. 2020;77(4):349–358.

Davidson M, Emsley R, Kramer M, et al. Efficacy, safety and early response of paliperidone extended-release tablets (paliperidone ER): results of a 6-week, randomized, placebo-controlled study. *Schizophr Res*. 2007;93(1–3):117–130.

Durgam S, Earley W, Lipschitz A, et al. An 8-week randomized, double-blind, placebo-controlled evaluation of the safety and efficacy of cariprazine in patients with bipolar I depression. *Am J Psychiatry*. 2016;173:271–281.

Frieling H, Hillemacher T, Ziegenbein M, Neundorfer B, Bleich S. Treating dopamimetic psychosis in Parkinson's disease: structured review and meta-analysis. *Eur Neuropsychopharmacol*. 2007;17(3):165–171.

Harrigan EP, Miceli JJ, Anziano R, et al. A randomized evaluation of the effects of six antipsychotic agents on QTc, in the absence and presence of metabolic inhibition. *J Clin Psychopharmacol*. 2004;24(1):62–69.

Huhn M, Nikolakopoulou A, Schneider-Thoma J, et al. Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis. *Lancet*. 2019;394(10202):939–951.

Kahn RS, Fleischhacker WW, Boter H, et al; EUFEST study group. Effectiveness of antipsychotic drugs in first-episode schizophrenia and schizophreniform disorder: an open randomised clinical trial. *Lancet*. 2008;371(9618):1085–1097.

Kane J, Canas F, Kramer M, et al. Treatment of schizophrenia with paliperidone extended-release tablets: a 6-week placebo-controlled trial. *Schizophr Res*. 2007;90(1–3):147–161.

Kane JM, Meltzer HY, Carson WH Jr, McQuade RD, Marcus RN; Aripiprazole Study Group. Aripiprazole for treatment-resistant schizophrenia: results of a multicenter, randomized, double-blind, comparison study versus perphenazine. *J Clin Psychiatry*. 2007;68(2):213–223.

Kane JM, Skuban A, Ouyang J, et al. A multicenter, randomized, double-blind, controlled phase III trial of fixed-dose brexpiprazole for the treatment of adults with acute schizophrenia. *Schizophr Res*. 2015;164:127–135.

Kane JM, Zukin S, Wang Y, et al. Efficacy and safety of cariprazine in acute exacerbation of schizophrenia: results from an international, phase iii clinical trial. *J Clin Psychopharmacol*. 2015;35:367–373.

Keefe RS, Bilder RM, Davis SM, et al; CATIE Investigators, Neurocognitive Working Group. Neurocognitive effects of antipsychotic medications in patients with chronic schizophrenia in the CATIE Trial. *Arch Gen Psychiatry*. 2007;64(6):633–647.

Krakowski M, Tural U, Czobor P. The importance of conduct disorder in the treatment of violence in schizophrenia: efficacy of clozapine compared with olanzapine and haloperidol. *Am J Psychiatry*. 2021;178(3):266–274.

Kramer M, Simpson G, Maciulis V, et al. Paliperidone extended-release tablets for prevention of symptom recurrence in patients with schizophrenia: a randomized, double-blind, placebo-controlled study. *J Clin Psychopharmacol*. 2007;27(1):6–14.

Kumra S, Kranzler H, Gerbino-Rosen G, et al. Clozapine and “high-dose” olanzapine in refractory early-onset schizophrenia: a 12-week randomized and double-blind comparison. *Biol Psychiatry*. 2008;63(5):524–529.

Kumra S, Oberstar JV, Sikich L, et al. Efficacy and tolerability of second-generation antipsychotics in children and adolescents with schizophrenia. *Schizophr Bull*. 2008;34(1):60–71.

*Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet*. 2013;382:951–962.

*Leucht S, Corves C, Arbter D, Engel RR, Li C, Davis JM. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. *Lancet*. 2009;373(9657):31–41.

*Leucht S, Komossa K, Rummel-Kluge C, et al. A meta-analysis of head-to-head comparisons of second-generation antipsychotics in the treatment of schizophrenia. *Am J Psychiatry*. 2009;166(2):152–163.

Lewis SW, Barnes TR, Davies L, et al. Randomized controlled trial of effect of prescription of clozapine versus other second-generation antipsychotic drugs in resistant schizophrenia. *Schizophr Bull*. 2006;32(4):715–723.

Lieberman JA. Comparative effectiveness of antipsychotic drugs. A commentary on: Cost Utility of The Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1) and Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE). *Arch Gen Psychiatry*. 2006;63(10):1069–1072.

Mamo D, Graff A, Mizrahi R, Shammi CM, Romeyer F, Kapur S. Differential effects of aripiprazole on D(2), 5-HT(2), and 5-HT(1A) receptor occupancy in patients with schizophrenia: a triple tracer PET study. *Am J Psychiatry*. 2007;164(9):1411–1417.

Marder SR, Hakala MJ, Josiassen MK, et al. Brexpiprazole in patients with schizophrenia: overview of short- and long-term phase III controlled studies. *Acta Neuropsychiatr*. 2017:278–290.

McEvoy JP, Lieberman JA, Perkins DO, et al. Efficacy and tolerability of olanzapine, quetiapine, and risperidone in the treatment of early psychosis: a randomized, double-blind 52-week comparison. *Am J Psychiatry*. 2007;164(7):1050–1060.

McEvoy JP, Byerly M, Hamer RM, et al. Effectiveness of paliperidone palmitate vs haloperidol decanoate for maintenance treatment of schizophrenia: a randomized clinical trial. *JAMA*. 2014;311:1978–1987.

Novick D, Haro JM, Suarez D, Vieta E, Naber D. Recovery in the outpatient setting: 36-month results from the Schizophrenia Outpatients Health Outcomes (SOHO) study. *Schizophr Res*. 2009;108(1–3):223–230.

Petrides G, Malur C, Braga RJ, et al. Electroconvulsive therapy augmentation in clozapine-resistant schizophrenia: a prospective, randomized study. *Am J Psychiatry*. 2015;172:52–58.

Rajkumar AP, Horsdal HT, Wimberley T, et al. Endogenous and antipsychotic-related risks for diabetes mellitus in young people with schizophrenia: a Danish population-based cohort study. *Am J Psychiatry*. 2017;174(7):686-694.

Ray WA, Chung CP, Murray KT, Hall K, Stein CM. Atypical antipsychotic drugs and the risk of sudden cardiac death. *N Engl J Med*. 2009;360(3):225–235.

Sandarsh S, Bishnoi RJ, Shashank RB, Miller BJ, Freudenreich O, McEvoy JP. Monitoring for myocarditis during treatment initiation with clozapine. *Acta Psychiatr Scand*. 2021;144(2):194–200.

Schooler NR, Marder SR, Chengappa KNR, et al. Clozapine and risperidone in moderately refractory schizophrenia: a 6-month randomized double-blind comparison. *J Clin Psychiatry*. 2016;77:628–634.

Shim JC, Shin JGK, Kelly DL, et al. Adjunctive treatment with a dopamine partial agonist, aripiprazole, for antipsychotic-induced hyperprolactinemia: a placebo-controlled trial. *Am J Psychiatry*. 2007;164(9):1404–1410.

Sikich L, Frazier JA, McClellan J, et al. Double-blind comparison of first- and second-generation antipsychotics in early-onset schizophrenia and schizo-affective disorder: findings from the treatment of early-onset schizophrenia spectrum disorders (TEOSS) study. *Am J Psychiatry*. 2008;165(11):1420–1431.

Stroup TS, Gerhard T, Crystal S, Huang C, Olfson M. Comparative effectiveness of clozapine and standard antipsychotic treatment in adults with schizophrenia. *Am J Psychiatry*. 2016;173:166–173.

Stroup TS, Lieberman JA, McEvoy JP, et al; CATIE Investigators. Results of phase III of the CATIE schizophrenia trial. *Schizophr Res*. 2009;107(1):1–12.

Taipale H, Tanskanen A, Mehtälä J, Vattulainen P, Correll CU, Tiihonen J. 20-year follow-up study of physical morbidity and mortality in relationship to antipsychotic treatment in a nationwide cohort of 62,250 patients with schizophrenia (FIN20). *World Psychiatry*. 2020;19(1):61–68.

Tandon R, Belmaker RH, Gattaz WF, et al; Section of Pharmacopsychiatry, World Psychiatric Association. World Psychiatric Association Pharmacopsychiatry Section statement on comparative effectiveness of antipsychotics in the treatment of schizophrenia. *Schizophr Res*. 2008;100(1–3):20–38

33.2c Dopamine Blockers and Neuromotor Adverse Effects

PHILIP G. JANICAK, M.D., AND KUMAIL HUSSAIN, M.D.

All antipsychotics can induce movement disorders. Thus, the Diagnostic and Statistical Manual of Mental Disorders, fifth edition-text revision (DSM-5-TR) continues to characterize these agents as “neuroleptics” in recognition of this property. These agents are categorized as first-generation (or typical) and second-generation (or atypical) antipsychotics, with one potential differentiation being their propensity to produce movement disorders. In addition to antipsychotics, a variety of other psychotropic and nonpsychotropic agents can also produce movement symptoms (Table 33.2c–1). For example, lithium (e.g., Eskalith), anticonvulsants, or antidepressants can all induce tremors; dopamine-receptor agonists can induce dyskinesia; some antidepressants can induce akathisia or produce abnormal movements upon their discontinuation, and prochlorperazine (Compazine) or metoclopramide (Reglan) can induce pseudoparkinsonism. Based on the specific agent and type of movement problem, various strategies are employed to counter these adverse effects.

Table 33.2c–1.
Selected Medications Associated with Movement Disorders: Impact on Relevant Neuroreceptors

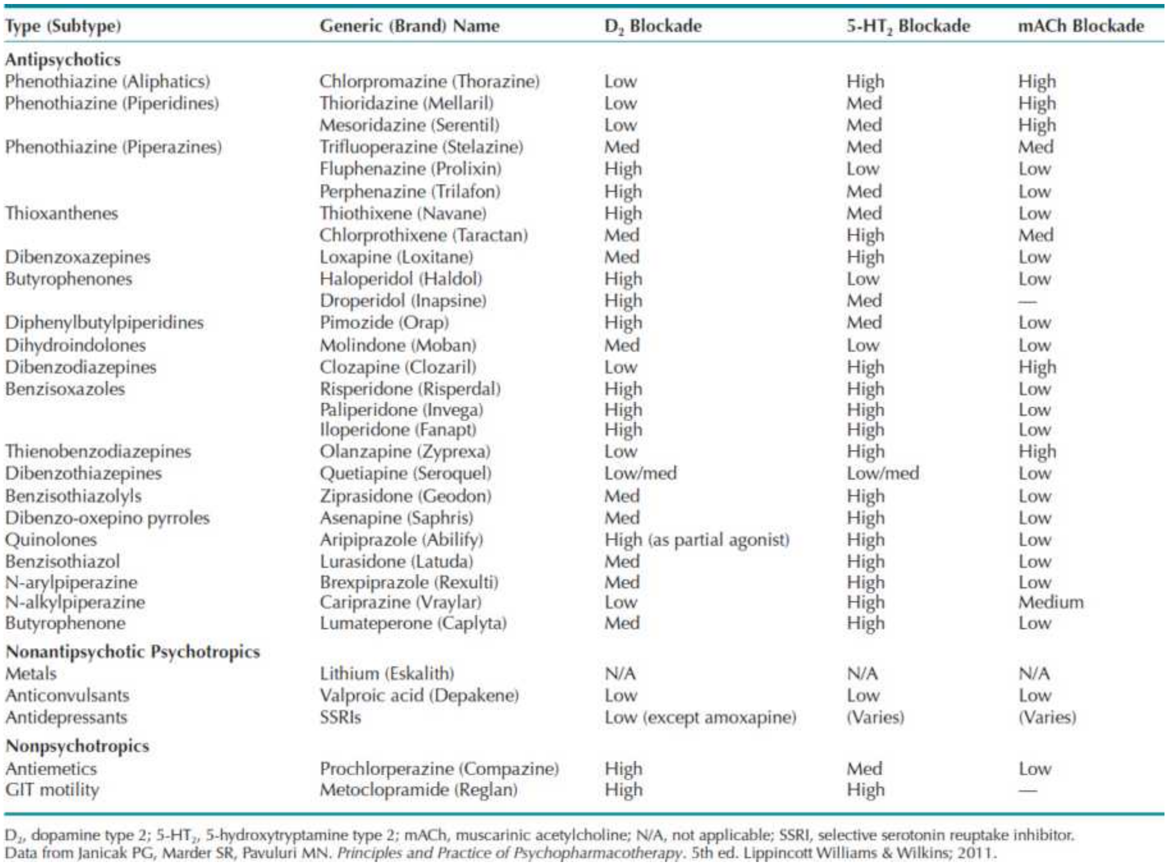

Type (Subtype)	Generic (Brand) Name	D_2 Blockade	5-HT_2 Blockade	mACh Blockade
Antipsychotics				
Phenothiazine (Aliphatics)	Chlorpromazine (Thorazine)	Low	High	High
Phenothiazine (Piperidines)	Thioridazine (Mellaril)	Low	Med	High
	Mesoridazine (Serentil)	Low	Med	High
Phenothiazine (Piperazines)	Trifluoperazine (Stelazine)	Med	Med	Med
	Fluphenazine (Prolixin)	High	Low	Low
	Perphenazine (Trilafon)	High	Med	Low
Thioxanthenes	Thiothixene (Navane)	High	Med	Low
	Chlorprothixene (Taractan)	Med	High	Med
Dibenzoxazepines	Loxapine (Loxitane)	Med	High	Low
Butyrophenones	Haloperidol (Haldol)	High	Low	Low
	Droperidol (Inapsine)	High	Med	—
Diphenylbutylpiperidines	Pimozide (Orap)	High	Med	Low
Dihydroindolones	Molindone (Moban)	Med	Low	Low
Dibenzodiazepines	Clozapine (Clozaril)	Low	High	High
Benzisoxazoles	Risperidone (Risperdal)	High	High	Low
	Paliperidone (Invega)	High	High	Low
	Iloperidone (Fanapt)	High	High	Low
Thienobenzodiazepines	Olanzapine (Zyprexa)	Low	High	High
Dibenzothiazepines	Quetiapine (Seroquel)	Low/med	Low/med	Low
Benzisothiazolyls	Ziprasidone (Geodon)	Med	High	Low
Dibenzo-oxepino pyrroles	Asenapine (Saphris)	Med	High	Low
Quinolones	Aripiprazole (Abilify)	High (as partial agonist)	High	Low
Benzisothiazol	Lurasidone (Latuda)	Med	High	Low
N-arylpiperazine	Brexpiprazole (Rexulti)	Med	High	Low
N-alkylpiperazine	Cariprazine (Vraylar)	Low	High	Medium
Butyrophenone	Lumateperone (Caplyta)	Med	High	Low
Nonantipsychotic Psychotropics				
Metals	Lithium (Eskalith)	N/A	N/A	N/A
Anticonvulsants	Valproic acid (Depakene)	Low	Low	Low
Antidepressants	SSRIs	Low (except amoxapine)	(Varies)	(Varies)
Nonpsychotropics				
Antiemetics	Prochlorperazine (Compazine)	High	Med	Low
GIT motility	Metoclopramide (Reglan)	High	High	—

D_2, dopamine type 2; 5-HT_2, 5-hydroxytryptamine type 2; mACh, muscarinic acetylcholine; N/A, not applicable; SSRI, selective serotonin reuptake inhibitor.
Data from Janicak PG, Marder SR, Pavuluri MN. *Principles and Practice of Psychopharmacotherapy*. 5th ed. Lippincott Williams & Wilkins; 2011.

HISTORY

An important observation is that spontaneous movement disorders (e.g., blunted affect, psychomotor slowing, buccal–lingual–masticatory movements, catatonia, alogia) can occur in patients with schizophrenia even in the absence of exposure to antipsychotic medications. For example, one review found a 4% to 7% prevalence of dyskinetic movements in patients with schizophrenia not on an antipsychotic. This risk increases with age and illness chronicity. Thus, an association between psychotic illness and movement abnormalities was described years before the introduction of antipsychotics.

The first antipsychotic, chlorpromazine (Thorazine), entered clinical practice in 1952. As a specific and effective medication for psychosis, it revolutionized treatment for millions of patients who, until then, had received little beyond custodial care. Additional agents were synthesized in the phenothiazine class, and eventually, several other first-generation antipsychotics (FGAs) became available. All appeared equally effective if given in sufficient bioequivalent doses.

Problems with movement disorders went relatively unnoticed with the early FGA, low-potency agents. Symptoms such as sedation, anticholinergic effects, and hypotension, however, encouraged the development of more potent and specific dopamine 2 (D_2) receptor antagonists, culminating with the introduction of haloperidol (Haldol) in 1968. Since this agent had little effect on other receptors, previous troublesome adverse effects could usually be avoided or substantially reduced. Before the introduction of the second-generation antipsychotics (SGAs), a common strategy was to use higher doses of a high-potency FGA with the belief that optimal efficacy could only be obtained at the cost of some extrapyramidal symptoms (EPS). This approach resulted in an increased incidence of acute and late-onset movement syndromes (e.g., pseudoparkinsonism; tardive dyskinesia [TD]). Not surprisingly, adherence was compromised because patients often felt the benefit of treatment was outweighed by such adverse effects. Adjunctive agents (e.g., benztropine [Cogentin]) were often used to control these problems but possessed their own set of uncomfortable (or even dangerous) anticholinergic effects. Coming full circle, older, lower-potency FGAs were then used as an alternative approach to minimize acute movement syndromes. Some psychiatrists also maintained that weaker D_2 antagonists could produce superior outcomes for certain patients while minimizing these risks. This position was supported by the release of the lower-potency D_2 antagonist, clozapine (Clozaril), in 1990.

Table 33.2c–2.

DSM-5 Categories for Medication-Induced Movement Disorders

Medication-induced parkinsonism	Tremor, muscular rigidity, bradykinesia, or akinesia developing within a few weeks of starting or raising the dose of a neuroleptic or after reducing a medication used to treat extrapyramidal symptoms.
Neuroleptic malignant syndrome	Severe muscle rigidity, elevated temperature, altered consciousness, autonomic dysregulation, and other related findings (e.g., diaphoresis, dysphagia, incontinence, mutism, elevated or labile blood pressure, elevated creatine kinase levels) developing in association with the use of certain medications.
Medication-induced acute dystonia	Abnormal positioning or spasm of the muscles of the head, neck, limbs, or trunk developing within a few days of starting or raising the dose of certain medications or after reducing a medication used to treat extrapyramidal symptoms.
Medication-induced acute akathisia	Subjective complaints of restlessness accompanied by observed movements (e.g., movement of the legs, rocking from foot to foot, pacing, or inability to sit or stand still) developing within a few weeks of starting or raising the dose of certain medications or after reducing a medication used to treat extrapyramidal symptoms.
Tardive dyskinesia	Involuntary choreiform, athetoid or rhythmic movements of the tongue, jaw, or extremities lasting at least a few weeks and developing in association with taking certain medications for at least a few months. Maybe for a shorter period of time in elderly individuals.
Tardive dystonia	Dystonia which occurs late in the course of treatment and persists for months to years.

Tardive akathisia	Akathisia which occurs late in the course of treatment and persists for months to years.
Medication-induced postural tremor	Fine tremor which develops during attempts to maintain posture in association with the use of certain medications.
Other medication-induced movement disorder	This category is for medication-induced movement disorders not explained by any of the disorders listed above. Examples include (1) parkinsonian, acute akathisia, acute dystonia, or dyskinetic movements associated with a nonantipsychotic medication; (2) a presentation that resembles NMS but is associated with a nonantipsychotic; and (3) tardive dystonia.
Antidepressant discontinuation syndrome	Movement symptoms which occur upon discontinuation or dose reduction of an antidepressant in use for at least 1 month.

NMS, neuroleptic malignant syndrome.

Data from the American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

DIAGNOSTIC CRITERIA

The DSM-5-TR lists 10 categories under “Medication-Induced Movement Disorders and Other Adverse Effects of Medication.” Clinical syndromes are divided into those associated with antipsychotics, other psychotropics, and nonpsychotropic agents. Antipsychotic-related syndromes are divided according to the usual time of onset in relation to starting medication or increasing the dose. In general, SGAs cause less frequent and less severe movement disorders than their first-generation counterparts. In sensitive individuals, however, these symptoms also occur with SGAs. Further, the reduction in incidence and severity of abnormal movements may delay recognition when they do occur. The specific categories include antipsychotic or other medication-induced parkinsonism; neuroleptic malignant syndrome (NMS); medication-induced acute dystonia and akathisia; TD; tardive dystonia and akathisia; medication-induced postural tremor; other medication-induced movement disorder; and antidepressant discontinuation syndrome. There is also a 10th category for other adverse effects of medication.

Table 33.2c–2 lists a brief description of the 10 categories provided in the DSM-5-TR.

CENTRAL NERVOUS SYSTEM CONTROL OF MOVEMENT

Cortical motor areas, the cerebellum, basal ganglia, and related subcortical nuclei control normal movement. Other central nervous system (CNS) areas that govern motivation are also important for the spontaneity and nature of movement. The classic antipsychotic-induced movement disorders are mediated primarily by their impact on D_2 receptors located in the extrapyramidal tract. Agents with high D_2 receptor affinity are associated with a greater risk of developing acute- and late-onset movement disorders.

These adverse effects are categorized into *hypokinetic disorders* (e.g., akinesia, bradykinesia), which involve a paucity of movement, or *hyperkinetic disorders* (e.g., dyskinesia, akathisia), which involve excessive movement. Such disorders are primarily mediated by the basal ganglia, which includes the corpus striatum (caudate nuclei, putamen); globus pallidus (internal, external segments); substantia nigra (pars compacta, pars reticulata); and the subthalamic nucleus.

These structures modulate motor activity by a set of complex mechanisms, ultimately balancing inhibitory and stimulatory impulses via dopamine, as well as other neurotransmitters. The latter include serotonin (5-HT), acetylcholine (Ach), γ-aminobutyric acid (GABA), and glutamate (Glu). In this context, the optimal range of D_2 blockade in drug-naive patients may be very narrow. For example, studies indicate that D_2 receptor occupancy in the corpus striatum with FGAs exceeding 78% is associated with an increased risk of EPS, while levels less than 65% are associated with suboptimal efficacy. An important finding is that these levels can be achieved in first-episode patients with haloperidol doses as low as 1 to 5 mg per day.

MEDICATION-INDUCED PARKINSONISM

Clinical Presentation

The clinical presentations of antipsychotic-induced parkinsonism and idiopathic primary Parkinson disease are indistinguishable. Further, patients with subclinical primary Parkinson disease often experience a worsening of the disorder when exposed to an antipsychotic. The classic triad of symptoms includes *muscular rigidity,* which can be continuous ("lead pipe") or cogwheeling (a discontinuous, ratchet-like muscle motion); *akinesia or bradykinesia,* which are characterized by psychomotor retardation, a minimization of spontaneous motor activity, slowed speech, decreased arm movements during walking, shuffling gait and masked facies. A *resting tremor* usually presents with rhythmic oscillations (3 to 6 Hz) affecting the head, tongue, jaw, legs, and extremities in a unilateral or symmetrical fashion. Although all three symptoms may be present, only one is required for the diagnosis. Symptoms can develop within days to weeks of starting or increasing the dose of medication or after the reduction of agents used to treat EPS. Not included in the criteria are other parkinsonian symptoms (e.g., poor balance, micrographia, shuffling gait), which are often seen in more severe cases and mimic the progression of Parkinson disease in a dose-related fashion. Women are more susceptible than men, and the incidence is higher for those older than 60 years. Symptoms are not better explained by a mental or generalized medical disorder or an adverse effect of another medication.

Pathophysiology

Drug-induced parkinsonism results from a decrease in nigrostriatal dopamine activity. This can be induced by various actions, including depletion of dopamine in presynaptic stores (e.g., reserpine); dopamine-blocking agents (e.g., antipsychotics); and atypical calcium-blocking agents such as cinnarizine (Stugeron).

Treatment Options and Implications

This syndrome usually subsides with dose reduction or medication cessation. Since this is not always an option, alternative strategies include switching from a high-potency D_2 blocking agent such as fluphenazine (Prolixin) to a lower-potency FGA such as perphenazine (Trilafon) or to an SGA such as quetiapine (Seroquel), lumateperone (Caplyta) or aripiprazole (Abilify). The last two agents are partial $D_{2,3}$ agonists.

Adjunctive anticholinergic agents such as benztropine (Cogentin) in divided doses of 2 to 8 mg per day; trihexyphenidyl (Artane) in divided doses of 6 to 10 mg per day; or diphenhydramine hydrochloride (Benadryl) in divided doses of 25 to 100 mg per day are usually effective in ameliorating this syndrome. Dopamine agonists such as amantadine (Symmetrel) in divided doses of 100 to 300 mg per day may also be helpful if anticholinergics are not tolerated or in elderly patients. Thus, adequate management is usually achieved by modification of the activity of dopamine, acetylcholine, or both. In addition, preliminary data indicate that the addition of oral dehydroepiandrosterone (DHEA) to existing antipsychotic regimens may also improve pseudoparkinsonism symptoms. After control of these adverse effects is achieved, the gradual withdrawal of the antiparkinsonian medication may be achieved without recurrence of EPS in some patients.

NEUROLEPTIC MALIGNANT SYNDROME

Clinical Presentation

The core symptoms of NMS include generalized “lead pipe” *muscle rigidity* and an *elevated temperature* which may range from modest to severe (e.g., 106 °F). *Altered consciousness* and *autonomic system dysregulation* are also common characteristics. These complications typically occur within days after exposure to an antipsychotic. NMS, however, can vary in terms of symptom presentation (e.g., core symptoms are not always present with certain SGAs); progression (e.g., rapid to delayed onset); and outcome (e.g., full recovery to death, particularly in unrecognized cases).

NMS can exhibit a rapid or gradual onset of the following symptoms:

- **Autonomic instability:**
    - Diaphoresis
    - Tachycardia
    - Elevated or labile blood pressure
    - Urinary incontinence
    - Pallor
    - Sialorrhea
- **Other symptoms:**
    - Dysphagia

  - Dysarthria
  - Trismus
  - Myoclonus
  - Akinesia
  - Tremor
  - Changes in the level of consciousness, ranging from confusion to stupor to coma
- **Laboratory findings:**
  - Leukocytosis
  - Metabolic acidosis
  - Evidence of muscle injury (e.g., creatine kinase elevation)
  - Liver enzyme elevation

NMS-like symptoms should not be due to a general medical (e.g., thyrotoxicosis, malignant hyperthermia [MH]) or neurologic (e.g., parkinsonian hyperthermia) disorder; the adverse effects of another type of medication (e.g., serotonin syndrome); or be primary to the underlying psychiatric illness (e.g., malignant catatonia) (Table 33.2c–3).

Table 33.2c–3.
Neuroleptic Malignant Syndrome: Differential Diagnosis

Primary CNS disorders	CNS vasculitis
	Infarctions
	Infections
	Parkinson disease
	Status epilepticus
	Trauma
	Tumors
Systemic disorders	Acute porphyria
	Autoimmune disorders
	Dehydration
	Heat stroke
	Hyperthyroidism
	Infections
	Pheochromocytoma
	Tetanus
Psychiatric disorders	Idiopathic lethal catatonia
Medication-related disorders	Anticholinergic syndrome
	Drug intoxication
	Levodopa syndrome
	Malignant hyperthermia
	Serotonin syndrome

CNS, central nervous system.

Data from Ware MR, Feller DB, Hall KL. Neuroleptic malignant syndrome: diagnosis and management. *Prim Care Companion CNS Disord*. 2018;20(1):17r02185.

Pathophysiology

A hypodopaminergic state which interferes with temperature regulation is the postulated mechanism for the development of this syndrome. NMS is associated with dopamine antagonists, the abrupt discontinuation of antiparkinsonian medications, and, rarely, the abrupt discontinuation of a dopaminergic agent. This suggests that fluctuation in dopamine binding may be important in its etiology. Because dopamine receptors are widely distributed in the CNS and spinal cord, it is likely that the various symptoms seen in NMS are accounted for by dopamine dysregulation in different areas. Thus, blockade of dopamine receptors in the corpus striatum causes muscle contraction and rigidity, which then initiates heat generation, whereas blockade in the hypothalamus may lead to impaired heat regulation. The autonomic dysfunction commonly seen in advanced cases may be caused by dopamine blockade of spinal neurons. The fluctuations in mental status (e.g., delirium; coma) may be secondary to hyperthermia and/or effects on the mesocortical dopamine tract.

Abnormal calcium availability in muscle cells may also contribute, given the potential for muscle damage reflected by high creatine kinase elevations. There are no direct data that link NMS and MH, although cautious use of antipsychotics is urged in patients with a history of MH. NMS may also be due to a dysregulation of the sympathetic nervous system.

The occasional familial clustering of NMS led to an analysis of genetic factors which may predispose to this condition. For example, the A1 allele of the dopamine D_2 receptor gene is observed to be overrepresented in patients who develop NMS compared with those who do not. This allele reduces the density and function of the dopamine receptor in the striatum and caudate nuclei, leading to reduced dopaminergic metabolism in those CNS areas with high concentrations of the neurotransmitter. By contrast, polymorphisms of the serotonin receptor gene do not appear to be associated with NMS but may increase the risk of developing acute EPS.

Treatment Options and Implications

NMS can be fatal or produce residual sequelae (e.g., cognitive or other neurologic deficits). Therefore, management involves the aggressive use of supportive measures, as well as more specific interventions. Early identification is important, and patients at higher risk should be carefully monitored.

Factors that may increase the risk of developing NMS include the following:

- **Concurrent medical and neuropsychiatric issues:**
  - Dehydration
  - Psychomotor agitation
  - Exhaustion

  - Comorbid neurologic disorders such as encephalitis and traumatic brain injury
  - Low serum iron levels
- **Psychiatric diagnosis and symptom history:**
  - Mood disorders
  - Pre-existing catatonia
  - History of NMS
  - History of MH
- **Medication related:**
  - Parenteral administration
  - High-potency FGAs (e.g., haloperidol)
  - High doses of an antipsychotic
  - Dose increases or intermittent nonadherence
  - Adjunctive lithium treatment
- **Demographic factors:**
  - Male gender
  - Younger age

One easily treatable risk factor is dehydration. Thus, as a preventative measure, severely ill or catatonic patients on an antipsychotic should receive oral fluids or intravenous (IV) hydration. In addition, patients should avoid excessively high temperatures. Very mild cases of low-grade temperature or mild creatine kinase elevations may be treated in a psychiatric unit, while a rapid onset or worsening of symptoms usually requires transfer to a medical intensive care unit. Re-exposure to an antipsychotic can be considered after full resolution of the episode. While switching to an SGA (e.g., clozapine) is often recommended, NMS also occurs with these agents. A past history of this syndrome is another reason patients are started or switched to SGAs. The relative risk of developing NMS between the two generations of antipsychotics, however, is difficult to compare.

While NMS is one of the few life-threatening psychotropic-induced complications, its infrequent occurrence and severe nature make controlled studies impossible. The best available data supports the following strategies:

1. *Prevention*: Antipsychotics should be used cautiously in patients with dehydration, catatonic symptoms, basal ganglionic disease, or a

history of NMS or MH.

2. *Early identification*: Patients with early signs should have their antipsychotic immediately discontinued, supportive measures initiated, and careful reintroduction of these agents considered only if absolutely necessary.
3. *Treatment*: The most important step is early recognition and prompt withdrawal of the offending agent. All cases require supportive interventions, which may include IV hydration, temperature control, and, in more severe cases, ventilator support or dialysis. In advanced cases, admission to an intensive care unit is required, with a fatal outcome occurring in up to 20% of cases. A variety of other treatments may also help, including the use of dopamine agonists (e.g., bromocriptine, amantadine); muscle relaxants (e.g., benzodiazepines, dantrolene); and, in selected patients, electroconvulsive therapy (ECT) (Table 33.2c–4). There has been some limited success reported with psychotropic agents such as bupropion (Wellbutrin), levodopa (Sinemet), carbamazepine (Tegretol), and apomorphine (Apokyn).
4. In patients who require an antipsychotic for their ongoing disorder after an episode of NMS, several steps may reduce the chance of recurrence. These include waiting at least 2 weeks after symptoms resolve before reintroduction is attempted, resolving prior risk factors (e.g., dehydration; avoiding concomitant lithium), and considering alternative therapies when appropriate (e.g., mood stabilizers; ECT). If rechallenge with an antipsychotic is necessary, choosing an alternate agent, starting with low doses and titrating slowly, can also reduce the risk of recurrence.

Table 33.2c–4.
Neuroleptic Malignant Syndrome: Treatment

Intervention	Dosing	Effectiveness
Amantadine	200–400 mg per day by mouth in divided doses	Beneficial as monotherapy or in combination; decreases death rate
Bromocriptine	2.5 mg twice a day or three times a day by mouth; may increase to a total of 45 mg per day	Mortality reduced when used as a single agent or in combination; may worsen psychosis, reduce blood pressure; induce emesis
Levodopa/carbidopa	Levodopa 50–100 mg per day intravenous (IV) as a continuous infusion	Case reports of dramatic improvement; may worsen psychosis, reduce blood pressure, or induce emesis
Dantrolene	1 mg/kg per day parenterally for 8 days; continue by mouth for 7 additional days	Benefits may occur in minutes or hours as a single agent or in combination
Benzodiazepines	1–2 mg intramuscular as a test dose; if effective, switch to oral route; consider if catatonic symptoms are present	Has been reported effective when other agents fail
Electroconvulsive therapy	Reports of good outcome with both unilateral and bilateral treatment; response may occur in as few as three treatments	Effective when medications have failed; also may treat underlying psychiatric disorder; may worsen cognitive function
Supportive measures	IV hydration and oxygenation	Often effective as an initial approach early in the episode

Cooling blankets	Can be used initially to lower temperature
Ice packs	A second-line measure if a more aggressive approach is needed to lower temperature
Ice water enema	Another second-line measure if a more aggressive approach is needed to lower temperature
Antipyretics	May be used to reduce fever, although its benefit is uncertain.

Data from Ware MR, Feller DB, Hall KL. Neuroleptic malignant syndrome: diagnosis and management. *Prim Care Companion CNS Disord*. 2018;20(1):17r02185.

MEDICATION-INDUCED ACUTE DYSTONIA

Clinical Presentation

Acute dystonia occurs in up to 10% of patients, especially young men receiving a high-potency FGA. Up to seven different symptoms may occur, although only one is required for the diagnosis. These symptoms include abnormal positioning of the head and neck (e.g., torticollis, retrocollis); spasms of the jaw muscles (e.g., trismus); impaired swallowing, speaking, or breathing; thickened or slurred speech; tongue protrusion or dysfunction; deviated eyes in any direction (e.g., oculogyric crisis); and abnormal positioning of limbs or trunk (e.g., opisthotonos). These symptoms are usually quite frightening to patients who may develop the syndrome within minutes of receiving parenteral, high-potency FGAs. More commonly, however, symptoms usually develop within days of starting or increasing the oral antipsychotic dose or reducing a medication used to treat EPS. Rarely, dystonia can be life-threatening if laryngeal muscles are affected.

Another complication with acute dystonic reactions is the frequent occurrence of long-term treatment nonadherence. After such an experience, patients commonly report being “allergic” to the offending medication. Careful history taking often reveals the true nature of the event, but patients are understandably reluctant to resume treatment with the implicated drug. Such a history may preclude future use of specific agents unless considerable time and energy are spent educating the patient regarding the probable cause of the reaction and how it can be prevented with anticholinergic agents. If a patient has a history of acute dystonia, the prophylactic use of such medications should be considered, even with the use of less potent antipsychotics.

The risk of acute dystonia is increased in young males. Recent use of cocaine, which depletes dopamine, is also associated with an increased risk. Some patients who are covertly nonadherent to oral tablets or capsules may develop acute dystonia when they are switched to a liquid concentrate or parenteral formulation. The risk of dystonia diminishes quickly after a steady dose is reached, and ongoing use of anticholinergic medication is usually not required. Patients at risk for developing dystonia, however, are also at risk for developing other forms of EPS, which may require continued anticholinergic management. A family history of movement disorders, as well as a prior history of dystonia, is predictive of a higher rate of acute dystonia and other forms of EPS.

Pathophysiology

Although the mechanism of acute dystonia is not understood, the sudden onset of symptoms and its limitation to the initiation of treatment or dose increase suggest that acute saturation of D_2 receptors plays a role. Preclinical studies suggest sigma 1 and sigma 2 receptors, found in the motor regions of the brain and involved in the control of posture and movement, may be involved in all forms of dystonia, including medication induced.

Treatment Options and Implications

Patients should always be informed of the risk of muscle spasms associated with these drugs and encouraged to report such events immediately. With severe reactions or laryngeal dystonia, which can compromise breathing, immediate recognition and treatment are essential. Parenteral (intramuscular [IM] or IV) benztropine (1 to 2 mg) or diphenhydramine (50 mg) typically produce rapid relief. Prophylactic use of oral anticholinergic medication, as well as the ready availability of the parenteral form, is strongly encouraged when the use of a high-potency FGA is considered.

Anticholinergic medication has the greatest effect on antipsychotic-induced acute dystonia, less of an effect on tardive dystonia, and little effect on idiopathic dystonia. Because tardive dystonia can be a particularly disabling disorder, there is some support for the use of clozapine which may even improve this condition. Botulinum toxin (Botox) injections also offer symptomatic improvement for tardive dystonia. The prevalence of this adverse effect diminishes when acute agitation is treated with a combination of intramuscular haloperidol (e.g., 2 to 5 mg) plus lorazepam (Ativan) (e.g., 2 mg). It is likely that both the lower haloperidol dose and lorazepam's sedative and peripheral muscle-relaxant effects contribute to the attenuation of this painful and frightening event. Ziprasidone (Geodon), olanzapine (Zyprexa), and aripiprazole all have approved acute IM formulations, which appear to carry a lower risk for acute dystonia and may be alternatives to haloperidol plus lorazepam.

MEDICATION-INDUCED ACUTE AKATHISIA

Clinical Presentation

Akathisia (from the Greek meaning "never to sit down") usually develops after starting or increasing the dose of medication or after decreasing the use of medications used to treat EPS. The incidence is estimated to be between 20% and 45% for patients on an antipsychotic. This syndrome is also reported with various antidepressants, particularly selective serotonin reuptake inhibitors (SSRIs), possibly mediated by serotonin agonism of the dopamine system. It is also associated with severe distress, suicidality, a higher risk for developing TD, and a poor treatment outcome.

Acute akathisia typically occurs within weeks of starting a medication, lasts less than 6 months, and is characterized by intense dysphoria and an inner sense of restlessness. In addition, at least one of the following observable movements should be present: fidgeting or swinging of the legs; rocking from foot to foot while standing; pacing; lifting the feet as if marching in place; crossing and uncrossing the legs when sitting; or the inability to sit or stand in one place for several minutes. *Chronic akathisia* involves persistent symptoms (i.e., 6 months or longer) after the introduction of or last change in medication dose. It is characterized by stereotypic movements, often in the presence of limb or orofacial dyskinesias. *Tardive akathisia* is often associated with TD and manifests with a delayed onset which may persist even after medication cessation. *Withdrawal akathisia* can occur within 6 weeks of switching or stopping an antipsychotic or upon stopping an anticholinergic agent. A common metric used to assess this problem is the Barnes Akathisia Scale (BAS) which quantifies both subjective and objective components.

Severe cases are easily recognized, whereas mild cases are more difficult to separate from symptoms of the primary psychiatric disorder, especially those in which anxiety and motor restlessness are prominent. Thus, patients who present with pacing or running may be misdiagnosed with manic excitement or psychotic agitation and receive additional doses of an antipsychotic, risking exacerbation of the movement disorder. By contrast, lowering the dose may unmask tardive akathisia. This adverse effect is also more common with high-potency FGAs and may lead to inadequate adherence to future treatment recommendations. The symptoms of akathisia should not be better explained as part of a mental, neurologic, or general medical disorder.

Pathophysiology

The pathophysiology of akathisia is uncertain. There is evidence. However, that decreasing dopamine tone in the ventral tegmental area and substantia nigra, which in turn alters interactions between the mesolimbic, mesocortical, and spinal dopamine/norepinephrine systems, is contributory. GABAergic, cholinergic, adrenergic, serotonergic, and neuropeptide mechanisms are also implicated.

Akathisia risk factors can be divided into:

- **Drug related:**
    - Higher doses
    - Rapid dose escalation
    - Higher-potency agents
- **Individual related:**
    - Older age
    - Female
    - Mood disorder
    - Preponderance of negative symptoms
- **Symptom related:**
    - Iron deficiency

- Cognitive dysfunction

Treatment Options and Implications

A slow dose titration schedule, dose reduction, or switching to a less potent FGA or an SGA are the preferred initial options. This should be done under close supervision to avoid exacerbation of psychosis or precipitating withdrawal symptoms, including akathisia. Clozapine may be an option for patients with chronic, persistent akathisia. In cases in which there is uncertainty, benzodiazepines are preferable to increasing the antipsychotic dose

β-*Blockers* such as propranolol (Inderal) are the most commonly prescribed medication for this problem. Propranolol's mechanism of action may relate to its alteration of both noradrenergic and serotonergic inputs into the Dopamine (DA) system. *5-HT*$_{2A/C}$ *antagonists* such as mirtazapine (15 mg per day); trazodone (100 mg per day); mianserin (15 mg per day); or cyproheptadine (8 to 16 mg per day) may produce benefits comparable to propranolol with possible better tolerability. *Anticholinergic* agents such as trihexyphenidyl (2 to 10 mg per day) or benztropine (1 to 8 mg per day)) and *benzodiazepines* such as lorazepam (1 to 2 mg per day) are second- and third-line choices, respectively.

In resistant cases, other *GABAergic agents* (e.g., gabapentin; 3,000 mg per day); *amantadine* (a weak glutamate receptor antagonist and DA-enhancing agent; 100 mg per day); or *clonidine* (a selective α2 adrenergic agonist; up to 0.15 mg per day) may help. In two randomized controlled trials (RCTs), the addition of *vitamin B6* (1,200 mg per day) for 5 days significantly improved subjective and global subscales of the BAS. In another RCT, 52 patients were treated with trazodone (50 mg daily for 5 days) and compared to a placebo group. At the end of the fifth day, those on trazodone demonstrated a significantly greater improvement in the objective and subjective symptoms of akathisia relative to placebo (i.e., $p < .01$; $p < .001$, respectively).

TARDIVE DYSKINESIA

Clinical Presentation

Symptoms of TD should develop after at least 2 months of exposure to a dopamine-blocking agent (or 1 month if the patient is 60 years or older);

while the patient is on medication or within 4 weeks of withdrawal from an oral agent (8 weeks from long-acting injectable medication); and be present for at least 4 weeks. The patient usually demonstrates involuntary movements of the tongue (e.g., twisting, protrusion); jaw (e.g., chewing); lips (e.g., smacking, puckering); trunk; or extremities. Patterns of movements may include rapid, jerky, and nonrepetitive movements (i.e., choreiform); slow, continuous, and sinuous movements (i.e., athetoid); or rhythmic movements.

The onset of the disorder may be gradual, occurring over a period of months to years. Symptoms are often seen in older patients who have poorly fitting dentures or other reasons for dyskinetic movements. In older patients (55 years or older), the onset of TD is about 25% after 1 year, 34% after 2 years, and 53% after 3 years. This rate of onset is three to five times higher than in younger patients. Patients with multi-infarct dementia and a psychiatric diagnosis are at greater risk than patients with Alzheimer disease or other organic mental disorders. As noted earlier, TD may also occur in never-medicated patients with schizophrenia.

Risk factors include increasing age, pre-existing parkinsonism, previous brain damage, antipsychotic treatment duration, and exposure to FGAs. These symptoms should not be due to a general medical or neurologic disorder or present as an adverse effect of another class of medication.

Pathophysiology

Sustained D_2 receptor blockade resulting in receptor hypersensitivity is the most common hypothesis explaining the development of TD. Clozapine's decreased propensity to induce TD and its multiple neuroreceptor effects (e.g., binding to 5-HT_{1C}, 5-HT_2, and 5-HT_3 receptors) suggest a role for serotonin as well. Serotonin-containing neurons interact with dopaminergic neurons in the substantia nigra and ventral tegmental areas. Genetic studies indicate a possible relationship with polymorphisms in the DA2 receptor, DA3 receptor, dopamine transporter (DAT1), and the serotonin 2A receptor genes. Oxidative stress and cell death secondary to increased glutamatergic neurotransmission caused by blockade of presynaptic dopamine receptors are also postulated. Other potential proposed theories involve GABA depletion and cholinergic deficiency.

Treatment Options and Implications

Since the advent of SGAs, fewer episodes and less severe symptoms of TD are now seen. For this reason, patients on FGAs are generally *switched to a newer agent if possible*. Further, there is evidence that agents such as clozapine may improve symptoms of TD without causing progression of the disorder. For example, some studies of clozapine report reduced symptom severity in tardive syndromes by up to 50%. Patients on an FGA who develop TD and are switched to an SGA have demonstrated improvement in their symptoms but which specific agents are preferable remains unclear. Although a small percentage of cases progress to florid and potentially disabling or life-threatening intensity, others remain stable, and some show spontaneous improvement.

Other strategies to address the problem include *gradual discontinuation* in patients whose need for antipsychotic medication is not clearly established (e.g., a psychotically depressed patient who is maintained on an antipsychotic after an apparent full recovery). Because a certain number of these patients will relapse, however, medication discontinuance requires close monitoring. Patients taken off high-potency FGAs usually require a cross-taper because TD symptom exacerbation may occur after an abrupt change to a low-potency agent. *Discontinuance, dose reduction, or switch* to another agent may improve, stabilize, or worsen symptoms.

The vesicular monoamine transporter 2 (VMAT2) is a protein embedded in the synaptic vesicle’s membrane and regulates dopamine uptake from the cytoplasm for storage and subsequent release. *Tetrabenazine*, an earlier generation VMAT2 inhibitor, was approved for the treatment of Huntington chorea and also considered as a potential therapy for TD. Due to its short half-life (2–10 hours), it needs to be dosed twice daily. In addition, common AEs can include body aches or pain, difficulty in breathing; difficulty with swallowing; drowsiness; fear; or nervousness.

Two new-generation VMAT2 inhibitors, *valbenazine* (Ingrezza) and *deutetrabenazine* (Austedo) were subsequently FDA approved specifically for TD and can differ from tetrabenazine in terms of pharmacokinetics, dosing requirements, tolerability, and safety. To date, however, there are no trials directly comparing valbenazine, deutetrabenazine, and tetrabenazine for the treatment of TD.

Valbenazine is the purified prodrug of the (+) α-isomer of tetrabenazine. It is a highly selective VMAT2 inhibitor with a half-life of 20 hours, allowing for once-daily dosing. It is the first medication approved by the FDA for the treatment of moderate to severe TD in adults. In one 6-week RCT, patients with moderate to severe TD (*n* = 339) were assigned to valbenazine (40 mg), valbenazine (80 mg), or placebo. At the study's end, those on valbenazine had a significantly greater (*p* < .001) reduction in their Abnormal Involuntary Movement Scale (AIMS) scores versus placebo, with a larger difference found for the 80-mg dose (i.e., –3.2 vs. –0.1) versus the 40 mg dose (i.e., –1.9 vs. –0.1). Common adverse effects of valbenazine can include drowsiness, dizziness, restlessness, and vomiting. It is also associated with the risk of QT prolongation and contraindicated in combination with CYP3A4/2D6 isoenzymes or monoamine oxidase inhibitors (MAOIs) due to increased monoamine levels in the synapse, potentially predisposing to the development of the serotonin syndrome.

Deutetrabenazine was subsequently FDA approved for moderate to severe TD. This agent is an isotopic isomer of tetrabenazine in which six hydrogen atoms are substituted with deuterium. Two 12-week RCTs involved patients with TD (n = 415) randomized to deutetrabenazine or placebo. AIMS change scores versus placebo for deutetrabenazine 18 mg BID (placebo-subtracted treatment difference −1.9, SE: 0.58, *p* < .001) and 12 mg BID (i.e., −1.8, SE: 0.60, *p* < .003) were significantly greater; but nonsignificant for 6 mg BID (i.e., −0.7, SE: 0.57, *p* < .22). Common AEs can include nasopharyngitis, insomnia, headaches, somnolence, diarrhea, dry mouth, and akathisia. Deutetrabenazine is contraindicated with medications that are strong CYP2D6 inhibitors and with the concurrent use of tetrabenazine, reserpine, or MAOIs. Finally, it differs from valbenazine in that it needs to be dosed BID and should be taken with food.

In addition, VMAT2 inhibitor doses should be lowered in patients with hepatic toxicity. Of note, VMAT2 inhibitors also carry a black box warning due to an increased risk of depression and suicidality.

B6 and *procholinergic agents* (e.g., donepezil [Aricept]). *Ondansetron* (Zofran), a selective 5-HT$_3$ receptor antagonist, may reduce symptoms of TD and also exert a beneficial effect on psychotic symptoms. Preliminary studies indicate *cyproheptadine* (Periactin), a 5-HT and histamine antagonist, and *levetiracetam* (Keppra) may improve TD. There are also two crossover studies that demonstrate the potential benefit of the use of amantadine (Symmetrel) to improve symptoms of TD.

The use of *deep brain stimulation* (DBS) for severe and refractory TD offers hope to those who are severely disabled. Systematic reviews and studies of galantamine (Razadyne), vitamin E, calcium channel blockers, benzodiazepines, lithium, estrogen, melatonin, buspirone, and essential fatty acids do not support their use for TD.

MEDICATION-INDUCED POSTURAL TREMOR

Clinical Presentation

Motor tremors can develop after starting a medication such as lithium, valproic acid (Depakene), α-adrenergic blockers, stimulants, dopaminergic agonists, caffeine, and theophylline (Bronkodyl), an antipsychotic or an antidepressant. Typically, it is a fine 8- to 12-Hz postural tremor affecting the limbs, head, mouth, or tongue. Other medical causes should be excluded, and the tremor should not be part of an antipsychotic-induced pseudoparkinsonism which usually occurs at rest, is slower, and, unlike true parkinsonian tremor, tends to worsen with movement.

Pathophysiology

The hypothesized mechanism of medication-induced postural tremor varies based on the class of drug implicated. For example, stimulants may cause tremors due to the resulting hyperadrenergic state.

Treatment Options and Implications

This diagnosis is commonly associated with the use of mood stabilizers such as lithium. For example, tremor is often present, may be dose related and can herald the onset of early lithium toxicity. Thus, checking the lithium plasma level and looking for other signs of toxicity is advisable. If the tremor is an isolated finding not associated with toxicity, it may improve with time. Although the dose of lithium may also be reduced, this may compromise its therapeutic effect. Tremor is often worsened by emotional factors such as anxiety, stress, or fatigue; medical issues such as alcohol withdrawal, hypoglycemia, thyrotoxicosis, pheochromocytoma, or hypothermia; or by the addition of other medications. In cases where a pre-existing tremor worsens with the addition of a medication, the diagnosis of other medication-induced movement disorder should be considered. *Reduction of dose or switching* to another agent in a different class may help. In some cases, a *benzodiazepine or β-blocker* can ameliorate the tremor.

TARDIVE DYSTONIA/TARDIVE AKATHISIA

Clinical Presentation

Tardive dystonia and tardive akathisia are movement symptoms that arise later in treatment with an antipsychotic and may persist for months or years despite the reduction or discontinuation of the patient's medication. Young male patients have a higher risk. This syndrome is differentiated from acute dystonia and acute akathisia based on the time to onset. Both tardive dystonia and akathisia are associated with a higher incidence of TD.

Pathophysiology

There is little known about the pathophysiology of tardive dystonia and tardive akathisia. Tardive dystonia appears to involve impairment of the basal ganglia, and tardive akathisia may involve dysfunction of the pontine region of the brain.

Treatment Options and Implications

Both of these movement disorders are preventable with the use of SGAs (e.g., clozapine; quetiapine). If the patient has symptoms while on or after discontinuing an antipsychotic, an increase in its dose or immediate reintroduction of the agent may be necessary to increase dopamine blockade. Such interventions can achieve short-term relief.

One potential treatment for refractory tardive dystonia is DBS. In one report, 60 patients with symptoms of various types of dystonia underwent DBS with leads placed in the globus pallidus pars interna. Using the Burke–Fahn–Marsden Dystonia Rating Scale (BFM), patients were assessed preoperatively, postoperatively, and for up to 5 years of follow-up. Improvement in the BFM severity and disability scores were 43% and 27%, respectively, by 6 months and were sustained thereafter. Thus, DBS may be a potential option for patients with severe, treatment-refractory, medication-induced tardive dystonia.

OTHER MEDICATION-INDUCED MOVEMENT DISORDER

Clinical Presentation

These disorders do not fit into more specific categories. For example, acute dystonia may occasionally follow the use of a benzodiazepine (e.g., midazolam) or antidepressants (e.g., amoxapine). Other forms of EPS which may occur as an idiosyncratic reaction to a nonantipsychotic medication are also coded under this category. NMS-like syndromes and tardive dystonia not associated with antipsychotic exposure may be coded under this category as well.

Pathophysiology

Genetic variability may account for idiosyncratic reactions seen in some patients. The ongoing research and rapid growth in knowledge about these variations in the CNS and their relationship to treatment response and adverse effects will help to clarify these questions.

Treatment Options and Implications

These disorders are infrequent, and discontinuation of the offending agent is always the preferred option. Careful history taking and collateral information may reveal exposure to more typical causes of movement disorders (e.g., the patient may have taken another's medication or is exposed to an antiemetic phenothiazine).

ANTIDEPRESSANT DISCONTINUATION SYNDROME

Clinical Presentation

While movement symptoms are not common withdrawal phenomena with antidepressants, abrupt discontinuation or rapid dose reduction may produce them. Usually, the patient is typically on the drug for a month or more, and movement symptoms arise within days after discontinuation or tapering. This may occur with various classes of antidepressants, including SSRIs (e.g., citalopram, fluoxetine); serotonin and noradrenergic reuptake inhibitors (SNRIs) (e.g., duloxetine, venlafaxine); tricyclic antidepressants (TCAs) (e.g., desipramine, imipramine); MAOIs (e.g., tranylcypromine, isocarboxazid); as well as other agents (e.g., mirtazapine). Common symptoms include *somatic* (e.g., dizziness, headache, nausea); *sensory* (e.g., numbness, ringing in the ears, “electric shocks,” flashes of light); *mood* (e.g., anxiety, irritability); and *abnormal movements* (e.g., EPS).

Some movement symptoms occur with all antidepressants, while others are specific to the class of agent and include the following:

- **SSRIs:**
    - Tremor
    - Akathisia
    - Parkinsonism
- **TCAs:**
    - Tremor
    - Akathisia
    - Parkinsonism
- **MAOIs:**
    - Catatonia
    - Myoclonic Jerks
- **SNRIs:**
    - Akathisia

Most of these symptoms occur upon discontinuation of an SSRI or TCA. The latter class has a higher tendency to cause parkinsonian symptoms, which are managed with anticholinergics (e.g., benztropine) if a patient does not wish to restart their antidepressant. MAOIs, especially tranylcypromine, may cause more severe movement symptoms such as myoclonic jerks and catatonia, which require acute management with benzodiazepines (e.g., lorazepam). Akathisia may occur with discontinuation from SSRIs, SNRIs, and TCAs and managed with β-blockers (e.g., propranolol), anticholinergics, and, in some cases, benzodiazepines.

Pathophysiology

While there are several hypotheses as to the cause of discontinuation symptoms with antidepressants, there is no clear consensus on their pathophysiology. There is agreement, however, that an increased propensity to develop these symptoms is associated with shorter half-life agents. For example, paroxetine is associated with the highest risk for SSRI discontinuation syndrome.

Treatment Options and Implications

Restarting the antidepressant typically resolves these symptoms. When no further treatment with an antidepressant is required, however, management of these symptoms depends on their severity. Mild symptoms often resolve spontaneously and require no treatment beyond counseling the patient. More moderate to severe symptoms may require reinstating the agent to alleviate symptoms quickly. The original antidepressant can be restarted, or another antidepressant with a similar mode of action may be substituted. Once stabilized, a more gradual tapering is then initiated. It is important to take various factors into account, such as the dose and type of antidepressant agent, to determine the appropriate tapering duration.

While the symptoms of the discontinuation syndrome are manageable, they can be prevented by educating the patient on the importance of gradually tapering the dose of medication. It is also essential to educate patients on the importance of adhering to their medication and the potential for adverse events upon their cessation.

SUMMARY

The impact of medication-induced movement disorders should not be underestimated. These conditions are often debilitating, even life-threatening; diminish the effectiveness of treatment by precluding an adequate trial; contribute to nonadherence; simulate or accentuate negative, cognitive, or mood symptoms; and increase the risk for irreversible, later-onset abnormal movements. Also, there is concern that some of these adverse effects (e.g., pseudoparkinsonism; TD; akathisia) are associated with an increase in suicidality (e.g., ideation, behavior, attempt). As a result, the choice of antipsychotic is often based in large part on its potential to produce such adverse effects. Since SGAs usually produce fewer acute and chronic movement symptoms than FGAs, patients are more likely to initially receive one of these agents. Further, patients with concurrent movement disorders (e.g., Parkinson disease), significant drug-induced movement adverse effects, or who are at increased risk for developing tardive syndromes are often switched to an

SGA. The substantial variation in the experience of such adverse effects, however, always requires an individualized risk–benefit assessment before choosing a specific agent. Figure 33.2c–1 depicts the relative risk of acute EPS with these agents in comparison to high-potency FGAs.

Diagnosis, phase of the illness, age, gender, and race affect the likelihood of developing movement disorders. For example, individuals with first- or early-onset psychosis are at the greatest risk for developing movement symptoms than more chronically ill patients; older White females with a mood disorder at are higher risk of developing TD; younger Black males are at the greatest risk for developing acute dystonia, and more severe acute EPS in the context of catatonic symptoms may predispose to developing NMS.

The management of medication-induced movement disorders includes the appropriate prescription of antipsychotic agents (e.g., accurate diagnosis); optimal dosing; medication change; or adjunctive treatments (e.g., anticholinergic agents, benzodiazepines, and β-blockers, which are most effective for antipsychotic-induced pseudoparkinsonism). Other agents may benefit NMS, including bromocriptine (Parlodel) and dantrolene (Dantrium). Newer generation VMAT2 inhibitors (e.g., valbenazine; deutetrabenazine) are now available for the treatment of TD. Management may also include the selection of an SGA with less likelihood of producing or even possibly improving these syndromes.

FUTURE DIRECTIONS

Medication-induced movement disorders require careful consideration of the risk–benefit ratio when initiating antipsychotics or other potential offending agents, limiting their use, and encouraging the development of safer alternatives. Developments in our understanding of brain function through advancements in neuroimaging and genetics will lead to safer and better-tolerated agents. For example, pharmacogenetics and pharmacogenomics are increasingly helping to explain how the various alleles of receptor-related genes impact response and/or influence the development of acute and long-term medication-related adverse effects. Since some medications are lifesaving for some patients but may be disabling or even life-threatening for others, such advances will permit a better risk assessment and a better selection process from the increasing number of available agents. For now, predicting and minimizing drug-related movement disorders are primarily based on multiple clinical factors (e.g., diagnosis; phase of the illness; race; age). Although the recognition and management of medication-induced adverse effects are important, it is always better to avoid or minimize them with a judicious choice of treatment(s). The goal is to improve adherence, thus maximizing efficacy, functionality, and quality of life.

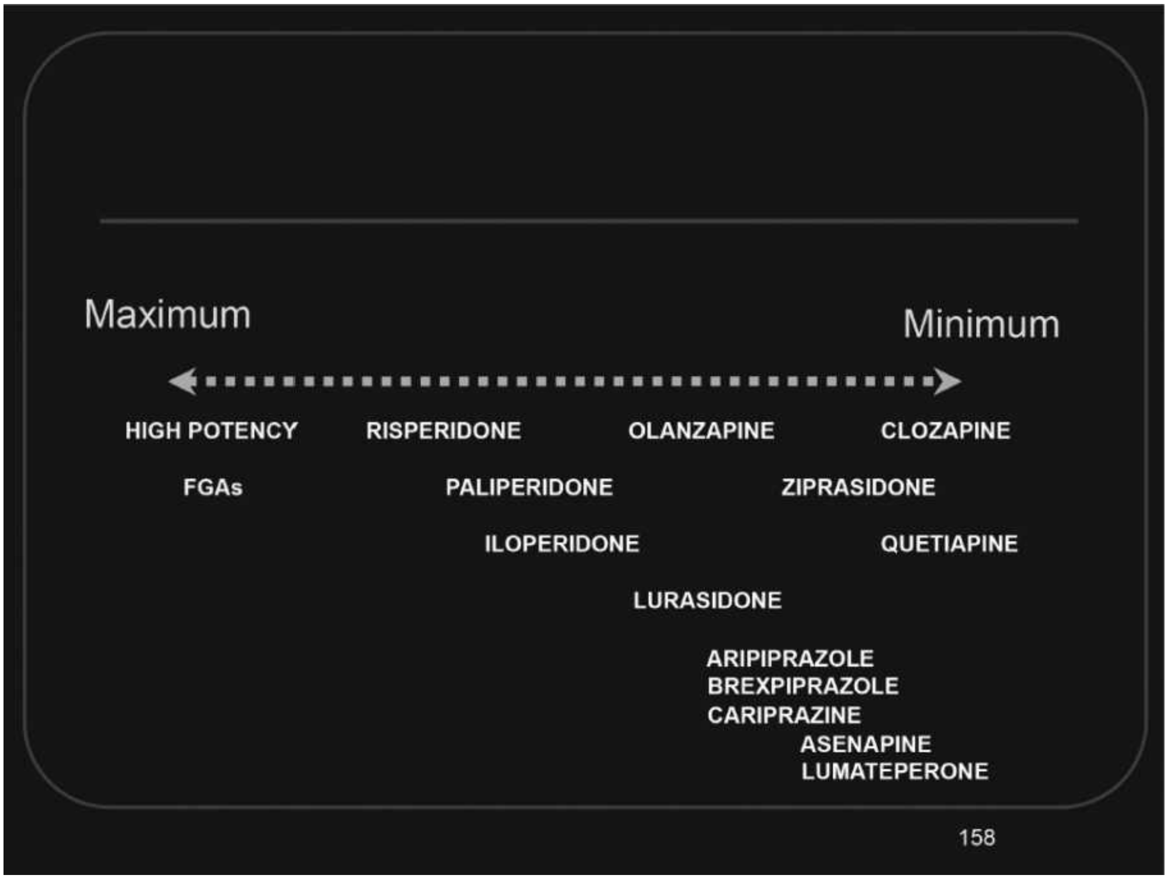

Maximum

Minimum

HIGH POTENCY
FGAs

RISPERIDONE
PALIPERIDONE
ILOPERIDONE

OLANZAPINE
LURASIDONE

ZIPRASIDONE

CLOZAPINE
QUETIAPINE

ARIPIPRAZOLE
BREXPIPRAZOLE
CARIPRAZINE
ASENAPINE
LUMATEPERONE

FIGURE 33.2c–1. Continuum of Acute Extrapyramidal Symptoms (EPS). (Data from Janicak PG, Marder SR, Pavuluri MN. *Principles and Practice of Psychopharmacotherapy*. 5th ed. Lippincott Williams & Wilkins; 2011.)

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Publishing; 2022.

Aquino CCH, Lang AE. Tardive dyskinesia syndromes: current concepts. *Parkinsonism Relat Disord*. 2014;20(Suppl 1):113–117.

*Atsariyasing W, Goldman M. Management of medication-related adverse effects. In: Janicak PG, Marder S, Tandon R, Goldman M, eds. *Schizophrenia: Recent Advances in Diagnosis and Treatment*. Springer; 2014:225–262.

Chang EF, Schrock LE, Starr PA, Ostrem JL. Long-term benefit sustained after bilateral pallidal deep brain stimulation in patients with refractory tardive dystonia. *Stereotact Funct Neurosurg*. 2010;88:304–310.

*Correll CU, Davis RE, Weingart M, et al. Efficacy and safety of lumateperone for treatment of schizophrenia: a randomized clinical trial. *JAMA Psychiatry*. 2020;77(4):349–358.

Damier P, Thobois S, Witjas T, et al; French Stimulation for Tardive Dyskinesia (STARDYS) Study Group. Bilateral deep brain stimulation of the globus pallidus to treat tardive dyskinesia. *Arch Gen Psychiatry*. 2007;64:170–176.

Davis JM, Janicak PG, Sakkas P, Gilmore C, Wang Z. Electroconvulsive therapy in the treatment of the neuroleptic malignant syndrome. *Convuls Ther*. 1991;7(2):111–120.

*Fagiolini A, Alcalá JÁ, Aubel T, et al. Treating schizophrenia with cariprazine: from clinical research to clinical practice. Real world experiences and recommendations from an International Panel. *Ann Gen Psychiatry*. 2020;19:55.

Fitzgerald JJ, Rosendal F, de Pennington N, et al. Long-term outcome of deep brain stimulation in generalized dystonia: a series of 60 cases. *J Neurol Neurosurg Psychiatry*. 2014;85(12):1371–1376.

Guzey C, Scordo MG, Spina E, Landsem VM, Spigset O. Antipsychotic-induced extrapyramidal symptoms in patients with schizophrenia: associations with dopamine and serotonin receptor and transporter polymorphisms. *Eur J Clin Pharmacol*. 2007;63:233–241.

Haddad PM, Anderson IM. Recognizing and managing antidepressant discontinuation symptoms. *Adv Psychiatr Treat*. 2007;13:447–457.

*Hauser RA, Factor SA, Marder S R, et al. KINECT 3: A Phase 3 Randomized, Double-Blind, placebo-controlled trial of valbenazine for tardive dyskinesia. *Am J Psychiatry*. 2017;174(5):476–484.

Hosenbocus S, Chahal R. SSRIs and SNRIs: a review of the discontinuation syndrome in children and adolscents. *J Can Acad Child Adolesc Psychiatry*. 2011;20:60–67.

*Kapur S, Zipursky R, Jones C, Remington G, Houle S. Relationship between dopamine D2 occupancy, clinical response, and side effects: a double-blind PET study of first-episode schizophrenia. *Am J Psychiatry*. 2000;157:514–520.

*Koning JP, Tenback DE, van Os J, Aleman A, Kahn RS, van Harten PN. Dyskinesia and parkinsonism in antipsychotic-naive patients with schizophrenia, first-degree relatives and healthy controls: a meta-analysis. *Schizophr Bull*. 2010;36(4):723–731.

Lencer R, Eismann G, Kasten M, et al. Family history of movement disorders as a predictor for neuroleptic-induced extrapyramidal symptoms. *Br J Psychiatry*. 2004;185:465–471.

*McAvoy J, Citrome L. Brexiprazole for the treatment of schizophrenia: a review of this novel serotonin-dopamine activity modulator. *Clic Schiophr Relat Psychoses*. 2016;9(4):177–186.

Miller D, Caroff SN, Davis SM, et al, Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators. Extrapyramidal side-effects of antipsychotics in a randomised trial. *Br J Psychiatry*. 2008;193:279–288.

*Niemann N, Jimenez-Shahed J. Deutetrabenazine in the treatment of tardive dyskinesia. *Neurodegener Dis Manag*. 2019;9(2):59–71.

Pappa S, Dazzan P. Spontaneous movement disorders in antipsychotic-naive patients with first-episode psychoses: a systematic review. *Psychol Med*. 2009:39(7):1065–1076.

Rana AQ, Chaudry ZM, HBlanchet PJ. New and emerging treatments for symptomatic tardive dyskinesia. *Drug Des Devel Ther*. 2013;7:1329–1340.

Renoir T. Selective serotonin reuptake inhibitor antidepressant treatment discontinuation syndrome: a review of the clinical evidence and possible mechanisms involved. *Front Pharmacol*. 2013;4:45.

*Ricciardi L, Pringsheim T, Barnes TRE, et al. Treatment recommendations for tardive dyskinesia. *Can J Psychiatry*. 2019;64(6):388–399.

Sethuram K, Gedzior J. Akathisia: case presentation and review of newer treatment agents. *Psych Annals*. 2014;44(8):391–396.

*Shams-Alizadeh N, Maroufi A, Asadi Z, Rahmani K, Hassanzadeh K. Trazodone as an alternative treatment for neuroleptic-associated akathisia: a placebo-controlled, double-blind, clinical trial. *J Clin Psychopharmacol*. 2020;40(6):611–614.

*Solmi M, Pigato G, Kane JM, Correll CU. Treatment of tardive dyskinesia with VMAT-2 inhibitors: a systematic review and meta-analysis of randomized controlled trials. *Drug Des Devel Ther*. 2018;12:1215–1238.

Suzuki T. Which rating scales are regarded as the “standard” in clinical trials for schizophrenia? A critical review. *Psychopharmacol Bull*. 2011;44(1):18–31.

Ware MR, Feller DB, Hall KL. Neuroleptic malignant syndrome: diagnosis and management. *Prim Care Companion CNS Disord*. 2018;20(1):17r02185.

*Witek N, Comella C. Valbenazine in the treatment of tardive dyskinesia. *Neurodegener Dis Manag*. 2019;9(2):73–81.

33.3 Dopamine Partial Agonists

33.3a Dopamine Partial Agonists

STEPHEN R. MARDER, M.D., AND WALTER DUNN, M.D., PH.D.

The dopamine-serotonin partial agonists–antagonists (DSPAs) were developed as agents offering balanced dopamine transmission modulation directly at the dopamine receptor site. This is in contrast to earlier dopamine-serotonin antagonists, which addressed aberrant dopamine activity by indirectly modulating dopamine levels via serotonin pathways. A partial agonist can function both as an agonist or antagonist based on the presence or absence of the endogenous ligand. It generally has lower intrinsic activity than the natural ligand, and thus in regions with reduced ligand, it will produce activity similar to the endogenous ligand but at a lower level. When there is excessive natural ligand present, it will function as an antagonist, blocking the effects of the excessive ligand. In schizophrenia, where there concurrently exists conditions of dopamine excess and deficit across different pathways, a partial agonist could target both types of dopamine imbalances. In the mesolimbic pathway, where excess dopamine causes positive symptoms, partial agonists would function as antagonists. In the mesocortical pathways where dopamine deficits are thought to underlie negative and cognitive symptoms, a partial agonist functioning as an agonist could improve these symptoms. In addition, partial agonists are also proposed to have reduced liability for extrapyramidal symptoms (EPS) and hyperprolactinemia when excessive dopamine receptor blockade occurs in the nigrostriatal and tuberoinfundibular pathways.

Presently, there are three DSPAs in clinical use for both schizophrenia and mood disorders. Aripiprazole was the first approved in this class of agents, followed by brexpiprazole and, most recently, cariprazine.

ARIPIPRAZOLE

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Dopamine, serotonin
Mechanism of Action	Receptor partial agonist (D_2, 5-HT_{1A}) and antagonist (5-HT_{2A})

History

The Nobel laureate Arvid Carlsson, most responsible for identifying the importance of dopamine in antipsychotic therapeutics, suggested that a dopamine partial agonist that acted at dopamine autoreceptors may be effective in treating psychosis. This led to a search by Tetsuro Kikuchi and Yasuo Oshiro at the Otsuka Pharmaceuticals laboratory in Tokushima, Japan, for partial dopamine agonists. Aripiprazole was synthesized in 1988 and entered phase I and early phase II studies in Japan in 1990, with later phase II trials in the United States beginning in 1993. In 1999, a collaborative arrangement was developed between Otsuka Pharmaceuticals and Bristol-Meyers Squibb for the worldwide development of aripiprazole. The oral formulation of aripiprazole was approved by the U.S. FDA in 2002. The first long-acting formulation, aripiprazole monohydrate once monthly (AOM), was approved in 2013; aripiprazole lauroxil (AL) was approved in 2015. A combination drug–device of oral aripiprazole and a digital tracking system to monitor drug adherence was approved in 2017.

Aripiprazole is among the most commonly prescribed dopamine modulators. This is likely due to its relatively benign side-effect profile, which includes fewer EPS and weight gain liability. The most prominent side effect is restlessness. In addition, aripiprazole seldom increases prolactin levels, and it can lower prolactin when it is added to prolactin, increasing dopamine blocker.

Chemistry

Aripiprazole is a highly lipid-soluble quinolinone derivative. Its chemical structure does not resemble any of the available dopamine blockers.

Pharmacologic Action

Pharmacokinetics

ABSORPTION. Aripiprazole is well absorbed after oral administration. Peak plasma concentrations are reached within 3 to 5 hours after administration. The bioavailability of aripiprazole after oral administration is 87%. Aripiprazole's absorption is not affected by administration with food.

IM administration of aripiprazole (short-acting form) leads to peak plasma concentrations at 1 to 3 hours. The bioavailability is 100%.

DISTRIBUTION. Aripiprazole is more than 99% bound to serum proteins. Studies of its steady-state volume of distribution indicate that aripiprazole has extensive extravascular distribution.

METABOLISM AND ELIMINATION. The activity of aripiprazole is primarily due to the parent compound, although some activity is related to dehydroaripiprazole metabolite. The mean elimination half-life of aripiprazole is approximately 75 hours, and the mean half-life for dehydroaripiprazole is 96 hours. These relatively long half-lives support once-daily dosing for aripiprazole and indicate that patients should attain a steady-state plasma concentration within 14 days. Aripiprazole is extensively metabolized in the liver by dehydrogenation, hydroxylation, and *N*-alkylation. Metabolism is primarily by CYP 3A4 and CYP 2D6 enzymes.

Pharmacodynamics

MECHANISM OF ACTION. As a partial D_2 agonist, aripiprazole offers the putative advantage of stabilizing the dopamine neurotransmitter system by decreasing the amount of signaling when there is excessive ligand (by occupying receptors and transmitting a reduced signal compared to the endogenous ligand) and by increasing the signal when there is reduced activity (by acting as an agonist). For aripiprazole, this could translate into an agent that acts as a functional antagonist under hyperdopaminergic conditions and as an agonist under hypodopaminergic conditions.

The evidence that aripiprazole acts as a partial agonist at D_2 receptors is substantial. In vivo binding studies have found that aripiprazole binds with high affinity to D_2 receptors. Studies in human volunteers have found that, at its clinically effective doses, aripiprazole occupies a similar proportion of D_2 receptors to that of other dopamine blockers. Aripiprazole blocks stereotypy induced by apomorphine in rodents, as do other dopamine blockers. Agonist effects of aripiprazole were identified in a study in which its administration resulted in a release of prolactin from rat pituitary slices. This property was supported in a model of dopamine hypoactivity in which aripiprazole decreased the reserpine-induced L-3,4-Dihydroxyphenylalanine (DOPA) accumulation that results from stimulation of presynaptic D_2 receptors.

Aripiprazole also appears to be a partial agonist at 5-HT_{1A} receptors. The importance of this activity in schizophrenia is unclear. Because buspirone (BuSpar), also a 5-HT_{1A} partial agonist, is an effective anxiolytic, it is conceivable that aripiprazole may also decrease anxiety. Aripiprazole also binds with high affinity to D_3, 5-HT_{2A}, 5-HT_{2C}, and H_1 receptors, with notable antagonist activity at 5-HT_{2A}.

Design and Interpretation of Clinical Drug Studies

Efficacy for Acute Schizophrenia. Double-blind trials have compared aripiprazole with placebo, risperidone, and haloperidol. The earliest published study evaluated 15 and 30 mg of aripiprazole, 10 mg of haloperidol, and placebo in 414 patients with acute schizophrenia and schizoaffective disorder. Over the 4 weeks of the trial, all three active groups were significantly superior to placebo and similar to one another on positive and negative syndrome scale for schizophrenia (PANSS) total and PANSS positive symptoms. For PANSS negative symptoms, 15 mg of aripiprazole and 10 mg of haloperidol resulted in significantly greater improvements compared with placebo. A subsequent study compared 20 and 30 mg of aripiprazole to placebo and 6 mg of risperidone in 406 patients with schizophrenia or schizoaffective disorder. All of the active conditions were superior to placebo for both positive and negative symptoms on the PANSS. Another study that evaluated 10, 15, and 20 mg of aripiprazole and placebo in 420 patients also found that all three active doses were more effective than placebo. Moreover, there were no significant differences among the three doses. The findings from these acute studies indicate that doses of 10, 15, 20, and 30 mg of aripiprazole are more effective than a placebo. There is also no indication that 30 mg is more effective than the lower doses. These results are consistent with a study indicating that 10 mg of aripiprazole results in 85% D_2 occupancy.

A study reported by John Kane and collaborators evaluated aripiprazole in patients who were treatment resistant. Treatment resistance was confirmed in an open-label trial of risperidone or olanzapine. After 6 weeks of double-blind treatment with aripiprazole or perphenazine, there were no significant differences in the number of patients who met improvement criteria. However, in both groups, a substantial proportion of patients, 27% of aripiprazole and 25% of perphenazine patients, met these criteria. This suggests that either strategy may be useful for patients who fail to respond well to risperidone or olanzapine.

Efficacy for Acute Mania. Aripiprazole at 30 mg daily was compared to placebo in two studies of bipolar disorder patients in an acute manic or mixed episode. Aripiprazole was superior to placebo in both improvements on the Young Mania Rating Scale (YMRS) and in the proportion of patients who demonstrated substantial improvement during the trial.

Long-Term Effectiveness in Schizophrenia. Two trials have evaluated the long-term effectiveness of aripiprazole in schizophrenia. In one study, stabilized patients (*N* = 310) were randomly assigned to either 15 mg of aripiprazole or placebo and observed for relapse over a 6-month period. Sixty-four percent of aripiprazole-treated patients survived the 6-month period without relapsing ($p < .01$) compared with 43% of placebo-treated patients. This study indicates that aripiprazole is an effective agent for preventing relapse in stabilized patients with schizophrenia.

The second study evaluated 1,294 patients who had experienced a psychotic relapse. Patients were randomly assigned to 30 mg of aripiprazole or 10 mg of haloperidol and followed for 1 year to evaluate their recovery from acute psychosis and the subsequent maintenance of the response. Aripiprazole-treated patients were more likely to remain on their medication for the full 52 weeks. Moreover, aripiprazole treatment was associated with greater improvements over the course of the study in both positive and negative symptoms, as well as depression as measured by the Montgomery-Asberg Depression Rating Scale (MADRS).

Long-Term Effectiveness in Bipolar Disorder. Patients with bipolar disorder who were stabilized on aripiprazole following a recent episode of acute or mixed mania were randomly assigned to continue on aripiprazole or change to a placebo. Aripiprazole was superior to placebo in delaying manic and depressed episodes.

Agitation in Schizophrenia or Bipolar Disorder. Clinicians often prescribe sedating dopamine blockers for patients who appear agitated. This common practice, as well as the tendency of aripiprazole to cause restlessness in some individuals, has led some to avoid prescribing this agent to agitated patients. A recent meta-analysis combining the data from several trials indicated that aripiprazole was effective in patients with relatively high and low levels of agitation.

A short-acting IM formulation of aripiprazole was evaluated in two studies of agitated patients with schizophrenia or schizoaffective disorder. In one study, patients (*N* = 357) were randomly assigned 1.0, 5.25, 9.75, or 15 mg of aripiprazole, 7.5 of haloperidol, or placebo. Although all of the treatment groups except 1 mg of aripiprazole were more effective than placebo, the 9.75 mg had the best response. Patients in this group demonstrated improvement by 45 minutes following their injection. A second study (*N* = 448) compared 9.75 mg of aripiprazole, 6.5 mg of haloperidol, and a placebo. Both aripiprazole and haloperidol were superior to placebo.

Adjunctive Treatment in Major Depression. Two double-blind, placebo-controlled trials evaluated adjunctive aripiprazole (2 to 15 mg per day) in patients with major depressive disorder (MDD) who had inadequate responses to escitalopram (Lexapro), fluoxetine, paroxetine controlled release, sertraline (Zoloft), or venlafaxine extended release (Effexor XR). Improvement on the MADRS was greater in the group that received aripiprazole. However, rates of akathisia and restlessness were higher in the aripiprazole group. The average doses in the two trials were 10.7 and 11.4 mg daily.

Other Studies of Aripiprazole. Aripiprazole was the first dopamine modulator shown to reduce levels of prolactin. An early study from South Korea found that elevated levels of prolactin in patients receiving haloperidol could be reduced by adding 30 mg daily of aripiprazole. Adding aripiprazole to a dopamine blocker to lower elevated prolactin has become an accepted practice. This is typically used for patients demonstrating a robust response to a dopamine blocker causing prolactin-related side effects, and switching agents risks a suboptimal response. Reduction of prolactin by aripiprazole can be seen with doses starting at 2 mg.

Effects on Specific Organs and Systems

Neurologic. While initial studies suggested aripiprazole was associated with only minimal amounts of extrapyramidal side effects, recent analyses and clinical reports suggest otherwise. In earlier comparisons of aripiprazole (15 or 30 mg), 10 mg of haloperidol, and placebo, only haloperidol resulted in a significantly greater increase in EPS as measured by the Simpson-Angus Scale compared with placebo. More recent literature report higher rates of akathisia with aripiprazole than initially believed. When aripiprazole is acutely intolerable, akathisia is often cited as the cause.

Cardiac. Aripiprazole is not associated with prolongation of the QT interval as assessed by ECG.

Therapeutic Indications

Schizophrenia. Aripiprazole is effective for the treatment of acute psychosis in schizophrenia. It is also effective for preventing relapse in stabilized patients with schizophrenia. This is supported by a 26-week trial that found higher relapse rates on placebo compared to aripiprazole. The 52-week comparison with haloperidol also provides evidence that patients tend to remain stable once they have responded to aripiprazole. Its effectiveness appears to be similar to that of haloperidol and risperidone. Aripiprazole is also indicated for adolescents aged 13 to 17 years with schizophrenia.

Bipolar I Disorder. Aripiprazole is effective for treating acute manic and mixed episodes as monotherapy or as adjunctive therapy with lithium or valproate. It is also indicated for children and adolescents (age 10 to 17 years).

Adjunctive Treatment of Depression. Aripiprazole is indicated as an adjunctive treatment for patients who have inadequate responses to monoamine enhancers.

Irritability Associated with Autistic Disorder. Aripiprazole is indicated in the treatment of irritability associated with autistic disorder. Trials demonstrating efficacy were conducted in patients 6 to 17 years of age.

Tourette Disorder. Aripiprazole is indicated in the treatment of Tourette disorder. Studies evaluating efficacy were conducted in pediatric patients aged 6 to 18 years.

Agitation Associated with Schizophrenia or Bipolar Mania. Short-acting IM aripiprazole has been approved for acute agitation. A dose of 10 mg of aripiprazole was effective after 45 minutes versus 7.5 mg IM haloperidol at 100 minutes.

Precautions and Adverse Reactions

Black Box Warning

INCREASED MORTALITY IN OLDER ADULT PATIENTS WITH DEMENTIA-RELATED PSYCHOSIS. Aripiprazole and the other dopamine modulators carry a black box warning for increased mortality in older adult patients with dementia-related psychosis compared with placebo.

SUICIDAL THOUGHTS AND BEHAVIORS. As with other dopamine modulators with indications for depression (and all other medications used for depression), aripiprazole has a black box warning indicating antidepressants may be associated with an increased risk of suicidal thinking and behavior in children, adolescents, and young adults.

Pregnancy and Lactation. Aripiprazole is designated pregnancy category C. There have been limited reports documenting neonatal and maternal outcomes to aripiprazole exposure during pregnancy. In one report, the mother and neonate had no pregnancy- or labor-related complications, and the infant seemed to be developing normally at a 6-month follow-up. However, the mother was unable to lactate, raising concerns about potential hypoprolactinemia. Aripiprazole has been associated with teratogenicity in animal studies at 3 to 10 times the maximum recommended dose for humans. Given the paucity of published data on the effect of aripiprazole on fetal and neonatal development, this medication should be prescribed during pregnancy only after a careful risk–benefit analysis. Currently, there is no published information on the effects of aripiprazole use in nursing mothers.

Drug Interactions

Aripiprazole is metabolized by CYP3A4 and CYP2D6. Because carbamazepine induces CYP3A4, coadministration with this agent reduces aripiprazole concentrations. Agents that inhibit CYP3A4, such as ketoconazole, may reduce the metabolism of aripiprazole, leading to an increase in concentrations. Similarly, agents that inhibit CYP2D6, such as quinidine, may raise aripiprazole concentrations (see Table 33.2b–3).

Dosage and Administration

Aripiprazole can be administered orally as a conventional tablet, an orally disintegrating tablet, or as an oral solution. It is also available as a short-acting IM formulation in single-dose vials containing 9.75 mg of aripiprazole and in a long-acting IM formulation (AOM) in both single-dose vials and prefilled syringes containing 300 or 400 mg of lyophilized aripiprazole. AL is available in prefilled syringes with 441, 662, 882, and

1,064 mg. A combination drug–device of aripiprazole tablets embedded with a digital tracking system is available for monitoring patient drug adherence.

Schizophrenia. Aripiprazole is administered at dosages between 10 and 30 mg daily. Two studies compared 10 mg of aripiprazole with placebo; in one study, aripiprazole was more effective than placebo, whereas, in the other, it was not. This suggests that 15 mg administered once a day is a reasonable dosage for initiating treatment. There is no evidence that higher dosages are associated with a greater likelihood of improvement. As a result, patients should probably be treated with 15 mg per day for a sufficient amount of time (e.g., 2 to 4 weeks) before increasing the dosage. Oral aripiprazole can be administered once daily without regard to meals.

Studies in adolescents found that aripiprazole was well tolerated when patients were started at 2 mg and raised to 5 mg in 2 days and to the target dose of 10 mg in 2 additional days.

Bipolar Disorder. Aripiprazole is effective at doses of 15 to 30 mg for patients with mania and mixed states. The recommended starting dose in adults is 15 mg daily if given as monotherapy or 10 to 15 mg daily if given as adjunctive therapy with lithium or valproate. In pediatric patients aged 10 to 17 years, it should be started at 2 mg/day, titrated to 5 mg/day after 2 days, and titrated to the target dose of 10 mg/day after 2 additional days. Further dose increases can be made by increments of 5 mg/day.

Adjunctive Treatment of Major Depressive Disorder. The recommended starting dose of aripiprazole in adults with inadequate responses to antidepressants is 2 to 5 mg daily. The effective dose appears to be between 10 and 15 mg.

Irritability Associated with Autistic Disorder. In patients aged 6 to 17 years, aripiprazole should be started at 2 mg/day, then increased to 5 mg/day, and subsequently increased to 10 or 15 mg/day based on efficacy and tolerability.

Tourette Disorder. For pediatric patients (age 6 to 18 years) weighing less than 50 kg, aripiprazole can be started at 2 mg/day, increased to 5 mg/day after 2 days, and subsequently increased to 10 mg/day based on efficacy and tolerability. For patients weighing more than 50 kg, aripiprazole should be titrated to a target dose of 10 mg/day on day 8, and the dose can be increased up to 20 mg/day if needed for control of tics.

Agitation Associated with Schizophrenia or Bipolar Mania (IM Formulation). The recommended IM dose for treating agitation in schizophrenia or bipolar mania is 9.75 mg. There is no indication that higher doses are more effective. As with other IM dopamine modulators, if repeated injections are required, patients should be treated with an oral formulation.

There is no indication that adjustments in dosage are required for age, gender, race, or smoking.

Long-Acting Injectable Aripiprazole. In 2013, a long-acting form of aripiprazole, AOM, was approved for IM injection, with an indication for the treatment of schizophrenia. As with all long-acting injectables, it is recommended to first assess medication tolerability with the oral formulation. The recommended starting and maintenance dose is 400 mg IM monthly. If this dose is not tolerated, it can be reduced to 300 mg IM monthly. The median $T_{\max}$ of this formulation is 5 to 7 days, and the terminal half-life of a 400 mg IM dose is 46.5 days. Thus, it is recommended that when starting this medication, oral treatment should be continued for 14 days before discontinuation, given the gradual increase in plasma levels after the initial injection.

The clinical efficacy of AOM was established in two clinical trials. The first trial (*N* = 339) was a 12-week randomized controlled trial in adults with schizophrenia who were having an acute relapse. The primary outcome measure used was the change in PANSS total score from baseline to week 10. Participants were randomized to receive aripiprazole long-acting injectable 400 mg or placebo on days 0, 28, and 56, with the dose of aripiprazole being allowed to be adjusted to 300 mg if clinically indicated. Aripiprazole showed statistically significant improvement on the primary outcome measure as compared to placebo. A second trial assessed longer-term efficacy. In this study, adults with schizophrenia were converted and stabilized on oral aripiprazole (10 to 30 mg daily). Next, participants were stabilized on aripiprazole long-acting injection (400 or 300 mg monthly) for at least 12 weeks, discontinuing oral aripiprazole after the first 2 weeks. In the final double-blind placebo-controlled randomized withdrawal phase, patients (*N* = 403) were randomized to either continue aripiprazole long-acting injection or switch to placebo injection. The primary outcome measure was the time to relapse, defined as worsening on prespecified Clinical Global Impression (CGI) scales, prespecified PANSS items, hospitalization, or violent behavior. Participants randomized to continue aripiprazole long-acting injectable had a longer time to relapse than those randomized to placebo.

In 2015, AL was approved for long-acting IM injections for the treatment of schizophrenia. AL is a prodrug of aripiprazole and marketed by Alkermes Inc. Safety and efficacy was demonstrated by a single phase III clinical trial where high (882 mg) and low (441 mg) were demonstrated to be safe in patients with schizophrenia. Additional safety and efficacy data were referenced from the previously approved listed drug aripiprazole. As there was sufficient evidence supporting similarity between AL and the listed drug aripiprazole, the FDA accepted reliance on safety and efficacy data previously demonstrated by aripiprazole. The primary advantage of AL over aripiprazole is increased dosing flexibility for the prescriber and patient. AL is available in strengths not available with long-acting IM aripiprazole and can be dosed every 4 to 8 weeks, compared to IM aripiprazole every 4 weeks (see Table 33.2b–4). In 2018, FDA approved aripiprazole initio. This formulation is administered with both oral aripiprazole and AL to achieve therapeutic plasma levels in the first days of treatment. As a result, there is no need for patients to continue on oral aripiprazole.

Drug–Device Combination of Aripiprazole. In 2017, a drug-device combination of aripiprazole and a digital tracking system called Abilify Mycite was approved. This was the first digital drug system of its kind to receive regulatory approval. The system is composed of aripiprazole tablets embedded with an ingestible event marker that signals to a sensor (Mycite Patch) worn by the user when the medication is ingested. This information is transmitted to a smartphone-based application that updates a web-based portal monitored by the health care provider. Notably, at the time of approval, there was no evidence of improved medication adherence or clinical outcomes compared to standard oral nondigital versions of aripiprazole.

BREXPIPRAZOLE

Neuroscience-Based Nomenclature (NbN)

Pharmacologic Domain	Dopamine, serotonin
Mechanism of Action	Receptor partial agonist (D_2, 5-HT_{1A}) and antagonist (5-HT_{2A})

Brexpiprazole was developed as a 5-HT_{1A} and dopamine D_2 partial agonist with antagonist activity at 5-HT_{2A} that would have similar efficacy to its predecessor, aripiprazole and a reduced tendency to cause akathisia, restlessness, and insomnia. Brexpiprazole has less agonist activity than aripiprazole which may result in less restlessness and insomnia.

History

Brexpiprazole was discovered by Otsuka Pharmaceutical and was codeveloped with Lundbeck. It received FDA approval in July 2015 for the treatment of schizophrenia and as an adjunctive treatment for depression. In September 2016, brexpiprazole was approved for the maintenance treatment of schizophrenia. In May 2023, brexpiprazole was approved for treatment of agitation associated with Alzheimer disease (AD). This is the first FDA approval of a medication for this indication.

Pharmacologic Action

Pharmacokinetics

ABSORPTION. After a single dose of brexpiprazole, peak plasma concentrations are reached within 4 hours. Steady-state concentrations are achieved within 10 to 12 days of multiple-dose administration. Levels are not affected by the ingestion of food.

METABOLISM AND ELIMINATION. Brexpiprazole is largely metabolized by CYP3A4 and CYP2D6 enzymes. Its main metabolite does not contribute to its clinical activity. After a single oral dose of brexpiprazole, approximately 25% of the dose is recovered in the urine and 46% in feces.

Pharmacodynamics and Mechanism of Action. Brexpiprazole is a partial agonist at serotonin 5-HT_{1A} and dopamine D_2 receptors and an antagonist at serotonin 5-HT_{2A} and adrenergic $\alpha_{1B/2C}$ receptors. The intrinsic activity of brexpiprazole at D_2 receptors is higher than that of pure antagonists but lower than that of aripiprazole. This limited intrinsic activity has the potential for reducing adverse reactions associated with other D_2 antagonists, such as extrapyramidal side effects and prolactin elevation. At the same time, brexpiprazole's lower intrinsic activity compared to aripiprazole may result in fewer adverse reactions related to D_2 receptor agonism such as akathisia, insomnia, restlessness, and nausea.

Design and Interpretation of Clinical Drug Studies

The efficacy of brexpiprazole in the treatment of acute exacerbations of schizophrenia was demonstrated in two pivotal 6-week, fixed-dose studies and a 6-week, flexible-dose trial with an active control, quetiapine XR.

In the fixed-dose studies, patients were randomized to brexpiprazole 2 or 4 mg once per day or placebo. One of the studies included a dose of 0.25 mg and the other 1 mg. The low doses were predicted to be ineffective and were not included in the analyses. Patients who were assigned to brexpiprazole were started on 1 mg for the first 4 days, increased to 2 mg on days 5 to 7, and then treated with 2 or 4 mg for the 5 remaining weeks. In both of the studies, 4 mg of brexpiprazole resulted in a statistically significantly greater improvement in PANSS total score than placebo. In one of the studies, 2 mg of brexpiprazole also demonstrated a greater improvement than placebo.

In the flexible-dose study, patients were randomized to flexible doses of 2- to 4-mg brexpiprazole, 400- to 800-mg quetiapine XR, or placebo. The mean doses of brexpiprazole and quetiapine at the last study visit were 3.5 mg/day and 674.4 mg/day, respectively. Quetiapine resulted in greater improvement compared to placebo, while brexpiprazole came very close to separating from placebo ($p = .056$).

The efficacy and tolerability of brexpiprazole as a maintenance treatment for schizophrenia were also demonstrated in a 52-week maintenance study. Patients who had been stabilized on brexpiprazole for at least 12 weeks were randomly assigned to receive brexpiprazole or placebo for 52 weeks. The primary efficacy endpoint was the time from randomization to the exacerbation of psychotic symptoms or an impending relapse. This study was terminated at an interim analysis when brexpiprazole demonstrated a significantly longer time to relapse.

The efficacy of brexpiprazole as an adjunctive treatment of MDD was evaluated in two 6-week, double-blind, placebo-controlled, fixed-dose trials. Subjects for both studies met the criteria for MDD and had an inadequate response to prior antidepressant therapy and an inadequate response to 8 weeks of prospective antidepressant treatment (with escitalopram, fluoxetine, paroxetine controlled release, sertraline, duloxetine delayed release, or venlafaxine extended release). In one study, patients were randomly assigned to 2 mg once a day or placebo, and in the other to 1 mg, 3 mg, or placebo (added to their antidepressant). For both studies, patients received 0.5 mg once daily during week 1, 1 mg during week 2, and the assigned dose for the 4 remaining weeks. After 6 weeks of treatment, adjunctive brexpiprazole 2 or 3 mg/day provided significantly greater improvements from baseline in mean MADRS total scores. Brexpiprazole is the first drug to receive an FDA approval for treatment of agitation associated with AD. Two phase III randomized, double-blinded, placebo-controlled studies, studies 331-12-283 and 331-12-284, were conducted as fixed- (1 mg and 2 mg) and flexible-dose studies (0.5 mg to 2 mg), respectively. Both had a treatment period of 12 weeks with the primary outcome as mean change from baseline to end of 12 weeks on the CMAI (Cohen-Mansfield Agitation Inventory) long form total score. The primary outcome of study 331-12-283 separated from placebo with the 2-mg but not the 1-mg dose. The active drug arm of study 331-12-284 did not achieve statistical separation from placebo. However, a post hoc analysis of subjects in study 331-12-284 who received the higher 2-mg dose did demonstrate improvement over the placebo. In 2018, the FDA provided additional guidance to Otstuka, upon which a third phase III, 12-week, placebo-controlled, fixed-dose study (2 mg and 3 mg) was conducted. The primary outcome of this study compared the grouped results of the 2-mg and 3-mg dose to placebo. A statistically significant separation from placebo was demonstrated, with post hoc analyses of the individual 2-mg and 3-mg doses both separating from placebo. Based on the results from these three phase III studies, brexpiprazole received FDA approval for agitation associated with AD in May 2023. The results from these three studies support 2 mg as the recommended target dose, with 3 mg as the maximum dose for this indication.

Therapeutic Indications

Brexpiprazole is FDA approved for acute and maintenance treatment of schizophrenia (2 to 4 mg/day), as adjunctive treatment for MDD (2 mg/day), and as treatment of agitation associated with dementia due to AD (2 to 3 mg/day).

Precautions and Adverse Reactions

Brexpiprazole is contraindicated in patients with a known hypersensitivity to brexpiprazole or any of its components.

Brexpiprazole and the other dopamine modulators carry a black box warning for increased mortality in older adult patients with dementia-related psychosis. Other precautions for all dopamine blockers include neuroleptic malignant syndrome, tardive dyskinesia, metabolic changes including hyperglycemia/diabetes mellitus, dyslipidemia, and weight gain. It also carries the same warning as other antidepressant medications for an increased suicidal risk in children, adolescents, and young adults with major depressive or other psychiatric disorders.

Metabolic Changes. In the 6-week trials for acute psychosis and depression, shifts in fasting glucose were similar for brexpiprazole and placebo. In the long-term open-label trials, 9% of subjects with depression and 10% of schizophrenia subjects experienced shifts to high fasting glucose from normal or borderline. The shifts to elevated levels of fasting total cholesterol, LDL, HDL, and triglycerides were also similar to placebo in the acute studies of schizophrenia and depression. In the 6-week trials in schizophrenia, 11% of patients who received 2 mg of brexpiprazole and 10% of those who received 4 mg had a weight gain equal to or greater than 7% compared to 4% on placebo. For the acute depression studies, 5% of 1 mg patients, 5% of 2 mg patients, and 2% of patients on 3 mg had a weight gain equal to or greater than 7% compared to 2% on placebo. This suggests a relatively modest risk for metabolic changes compared to other dopamine modulators.

Extrapyramidal Symptoms and Akathisia. Both objective ratings and reported events suggest that brexpiprazole is associated with a small risk for EPS and akathisia. For example, in depression trials, the incidence of reported EPS, excluding akathisia, was 6% for brexpiprazole patients and 3% for placebo patients. The incidence of akathisia events for brexpiprazole patients was 9% versus 2% for placebo patients. In schizophrenia trials, the incidences of reported EPS and akathisia were similar for brexpiprazole and placebo.

Other Side Effects. Other notable adverse effects which occurred in ≥1% of subjects and at greater frequency than placebo included blurry vision, nausea, prolactin elevation, myalgia, somnolence, and insomnia.

Pregnancy. Neonates exposed to brexpiprazole during the third trimester of pregnancy are at risk for EPS and/or withdrawal symptoms following delivery. There are insufficient data on brexpiprazole to determine if it is associated with a risk for birth defects or miscarriage.

Nursing Mothers. There is insufficient information about the presence of brexpiprazole in human milk, the effects on the breastfed infant, or the effects on milk production.

Drug Interactions

The dosage of brexpiprazole should be reduced when it is coadministered with strong CYP3A4 inhibitors or CYP2D6 inhibitors.

Dosage and Administration

Brexpiprazole is available as 0.25-, 0.5-, 1-, 2-, 3-, and 4-mg tablets.

The recommended dose range for schizophrenia is 2 to 4 mg administered as a single daily dose. Patients should be started on 1 mg daily for 4 days, increased to 2 mg for 3 days and then increased to 4 mg if necessary.

The recommended dose for the adjunctive treatment of depression is 2 mg daily. The recommended target dose for agitation associated with AD dementia is 2 mg/day with a maximum dose of 3 mg/day. Patients should be started at 0.5 mg/day and increased over the course of 15 to 28 days.

For both indications, the dose should be decreased by 50% in CYP2D6 poor metabolizers.

Dosing in Special Populations. The safety and effectiveness of brexpiprazole in pediatric populations have not been assessed. There are inadequate data regarding the safety of brexpiprazole in older adult patients. Patients with moderate to severe hepatic impairment (Child–

Pugh score ≥7) should be prescribed a maximum dose of 2 mg once daily for patients with MDD and 3 mg once daily for patients with schizophrenia. For patients with moderate, severe, or end-stage renal impairment (creatinine clearance CrCl <60 mL/minute), the maximum recommended dosage is 2 mg once daily for patients with MDD, 3 mg once daily for patients with schizophrenia, and 3 mg for treatment of agitation associated with AD.

CARIPRAZINE

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Dopamine, serotonin
Mechanism of Action	Receptor partial agonist (D_2, D_3, 5-HT_{1A}) and antagonist (5-HT_{2B})

Cariprazine is a partial agonist at dopamine D_3 and D_2 receptors and serotonin 5-HT_{1A} receptors and an antagonist at 5-HT_{2B} receptors. The binding affinity for D_3 is an order of magnitude higher than the affinity for D_2. It is the only dopamine modulator presently available in the United States with this higher affinity for D_3 over D_2 receptors. It is plausible that the preference for D_3 might result in improvements in cognition, mood, and negative symptoms in schizophrenia.

History

Cariprazine was developed by Gedeon Richter, a pharmaceutical company located in Budapest, Hungary. It was subsequently licensed to Allergan (which was previously known as Actavis) for the United States and Canada. Cariprazine, under the trade name Vraylar, was initially approved for the treatment of schizophrenia and bipolar disorder in September 2015. In 2022, it received approval as an adjunctive treatment for major depressive disorder.

Pharmacokinetics

ABSORPTION. Cariprazine is well absorbed following oral administration, reaching peak concentrations between 3 and 6 hours after dosing.

Absorption is not substantially affected by food.

PHARMACOLOGIC ACTION. The activity of orally administered cariprazine is mediated by the parent drug and its two major active metabolites, desmethyl cariprazine (DCAR) and didesmethyl cariprazine (DDCAR). The metabolites appear to be equipotent to cariprazine at target therapeutic receptors. After multiple-dose administrations, mean cariprazine and DCAR concentrations reach steady state between 1 and 2 weeks; mean DDCAR concentrations reach steady state in 4 to 8 weeks. The mean half-life of cariprazine is 2 to 5 days, whereas the half-life of DDCAR is considerably longer and results in substantial plasma accumulation. This very long half-life of the active metabolite indicates that patients who are started on cariprazine are likely to have a gradual increase in active drug over several weeks. Moreover, when the drug is discontinued, the active moiety is likely to remain in plasma for weeks.

METABOLISM AND ELIMINATION. Cariprazine is extensively metabolized by CYP3A4 and CYP2D6 to DCAR and DDCAR. Approximately 21% of the daily dose is found in urine. The excretion half-life is 2 to 4 days for the parent drug and 1 to 3 weeks for the active metabolite (see Table 33.2b–4).

Pharmacodynamics and Mechanism of Action. In vitro studies indicated that cariprazine has a greater affinity for D_3 compared with D_2 receptors. A recent positron emission tomography (PET) study found that subjects treated with 1 mg/day of cariprazine had mean D_3 and D_2 receptor occupancies of 76% and 45%, respectively, indicating selectivity for D_3 over D_2 receptors at low doses. It is a partial agonist at these two receptors in addition to 5-HT_{1A} receptors. It is an antagonist with high affinity at 5-HT_{2B} receptors with moderate affinity at 5-HT_{2A} receptors as an antagonist.

Cariprazine is unique among available dopamine modulators in having a much greater affinity for D_3 as compared to D_2 receptors. In addition, it has a greater affinity for D_3 receptors than dopamine itself. It is unclear whether this activity is related to its antipsychotic properties since studies in rodents indicate that D_3 antagonists are not effective in preclinical models of schizophrenia. It is possible that the combination of D_3 and D_2

activity of cariprazine results in antipsychotic activity and potentially confers additional activity for other domains of psychopathology, such as negative symptoms.

Design and Interpretation of Clinical Drug Studies

The efficacy of cariprazine in acute exacerbations of schizophrenia was demonstrated in three positive clinical trials. In a phase II trial, patients treated with 1.5, 3, or 4.5 mg daily of cariprazine or 4 mg of risperidone showed significantly greater improvement in total PANSS compared with placebo. Another trial with a similar population found that 3 and 6 mg of cariprazine, as well as 10 mg of aripiprazole, resulted in greater improvement than placebo. A third trial in acute schizophrenia found that patients dosed from 3 to 6 mg and 6 to 9 mg demonstrated greater improvements than placebo-treated patients.

A separate trial in schizophrenia randomized patients with predominantly negative symptoms to 26 weeks of double-blind treatment with cariprazine (mean dose, 4.2 mg/day) or risperidone (mean dose 3.8 mg/day). Patients assigned to cariprazine demonstrated significantly greater improvement in negative symptoms than patients assigned to risperidone, as measured by the PANSS negative symptoms factor score.

Cariprazine was also evaluated in 3-week trials for bipolar I disorder patients with manic or mixed states. In each of the trials, improvement was evaluated with the YMRS. In two flexible-dose studies (cariprazine doses of 3 to 12 mg/day), cariprazine was more effective than placebo. A third study compared cariprazine at 3 to 6 and 6 to 12 mg/day with a placebo. Both dose ranges of cariprazine were superior to placebo, and there was no efficacy advantage for the higher-dose range.

A study of patients with MDD and inadequate response to an antidepressant randomized subjects to 8 weeks of treatment with adjunctive cariprazine 1 to 2 mg/day, 2 to 4.5 mg/day, or placebo. Patients in the 2- to 4.5-mg group demonstrated greater improvement on the MADRS.

Finally, a study in patients with bipolar I depression randomized subjects to treatment with cariprazine 0.75, 1.5, or 3.0 mg/day or placebo. The 1.5- and 3-mg/day doses were associated with greater improvement in the MADRS, but the 0.75-mg/day dose was not more effective than the placebo.

Therapeutic Indications

Cariprazine is FDA approved in the United States for the treatment of adult patients with schizophrenia (1.5 to 6.0 mg/day). For bipolar I disorder, it is approved for acute manic or mixed episodes (3.0 to 6.0 mg/day) and depressive episodes (1.5 mg to 3.0 mg). As adjunctive therapy for treatment of MDD the recommended dose is 1.5 mg or 3 mg daily.

Precautions and Adverse Reactions

Cariprazine is contraindicated in individuals with hypersensitivity to the drug or components in the tablets. If cariprazine is prescribed with a strong CYP3A4 inhibitor, the dosage should be reduced.

Cariprazine and the other dopamine modulators carry a black box warning for increased mortality in older adult patients with dementia-related psychosis. Other precautions for all dopamine blockers include neuroleptic malignant syndrome, tardive dyskinesia, metabolic changes including hyperglycemia/diabetes mellitus, dyslipidemia, and weight gain. Due to cariprazine's long half-life (related to its active metabolites), patients should be monitored for adverse reactions and patient response for several weeks after starting this drug.

Metabolic Changes. In 3- and 6-week trials for acute psychosis and mania, shifts in fasting glucose were similar for cariprazine and placebo. In long-term open-label trials, 4% of patients with normal hemoglobin A1c baseline values developed elevated levels. In acute studies, the shifts in fasting total cholesterol, low-density lipoprotein (LDL), high-density lipoprotein (HDL), and triglycerides were also similar in patients treated with cariprazine and placebo. There was a small amount of weight gain associated with cariprazine treatment in acute studies. For long-term studies in schizophrenia, the mean changes in body weight from baseline at 12, 24, and 48 weeks were 1.2 kg, 1.7 kg, and 2.5 kg, respectively.

Extrapyramidal Symptoms and Akathisia. In clinical trials, EPS and akathisia were reported at rates ≥5% and at least twice the rate of placebo. Both of these effects were more common at higher doses of cariprazine.

Pregnancy. As with other dopamine-modulating drugs, neonates exposed to cariprazine during the third trimester of pregnancy are at risk for EPS and/or withdrawal symptoms following delivery. There are insufficient data on cariprazine to determine if it is associated with a risk for birth defects or miscarriage. There are animal data suggesting that cariprazine may cause fetal harm.

Nursing Mothers. There is insufficient information about the presence of cariprazine in human milk, the effects on the breastfed infant, or the effects on milk production.

Dosage and Administration

Cariprazine is available in 1.5-, 3-, 4.5-, and 6-mg capsules. It can be administered as a single daily dose.

For treating schizophrenia, it is recommended that patients be started on 1.5 mg on the first day. The dose can be increased in 1.5-mg increments. The recommended dose range is 1.5 to 6 mg/day. There is no indication that doses above 6 mg/day are associated with greater benefit.

For treating bipolar I mania and mixed states, it is also recommended that patients be started on 1.5 mg the first day, followed by an increase to 3 mg. The recommended dose range is 3 to 6 mg/day.

For depression associated with bipolar I, the maximum recommended daily dose is 3.0 mg/day (see Table 33.2b–4).

Dosing in Special Populations. The safety and effectiveness of cariprazine in pediatric populations have not been assessed. There are inadequate data regarding the safety of cariprazine in older adult patients. No dosage adjustment is recommended for patients with mild to moderate renal or hepatic disease (see Table 33.2b–1).

FURTHER READINGS

Correll CU, Skuban A, Ouyang J, et al. Efficacy and safety of brexpiprazole for the treatment of acute schizophrenia: a 6-week randomized, double-blind, placebo-controlled trial. *Am J Psychiatry*. 2015;172(9):870–880.

Durgam S, Earley W, Lipschitz A, et al. An 8-week randomized, double-blind, placebo-controlled evaluation of the safety and efficacy of cariprazine in patients with bipolar I depression. *Am J Psychiatry*. 2016;173:271–281.

Hard ML, Mills RJ, Sadler BM, Turncliff RZ, Citrome L. Aripiprazole lauroxil: pharmacokinetic profile of this long-acting injectable antipsychotic in persons with schizophrenia. *J Clin Psychopharmacol*. 2017;37(3):289–295.

Kane JM, Meltzer HY, Carson WH Jr, McQuade RD, Marcus RN, Sanchez R; Aripiprazole Study Group. Aripiprazole for treatment-resistant schizophrenia: results of a multicenter, randomized, double-blind, comparison study versus perphenazine. *J Clin Psychiatry*. 2007;68(2):213–223.

Kane JM, Skuban A, Ouyang J, et al. A multicenter, randomized, double-blind, controlled phase 3 trial of fixed-dose brexpiprazole for the treatment of adults with acute schizophrenia. *Schizophr Res*. 2015;164:127–135.

Kane JM, Zukin S, Wang Y, et al. Efficacy and safety of cariprazine in acute exacerbation of schizophrenia: results from an international, phase iii clinical trial. *J Clin Psychopharmacol*. 2015;35:367–373.

Kelly DL, Powell MM, Wehring HJ, et al. Adjunct aripiprazole reduces prolactin and prolactin-related adverse effects in premenopausal women with psychosis: results from the DAAMSEL Clinical Trial. *J Clin Psychopharmacol*. 2018;38(4):317–326.

Mamo D, Graff A, Mizrahi R, Shammi CM, Kapur S, Romeyer F. Differential effects of aripiprazole on D(2), 5-HT(2), and 5-HT(1A) receptor occupancy in patients with schizophrenia: a triple tracer PET study. *Am J Psychiatry*. 2007;164(9):1411–1417.

Marder SR, Hakala MJ, Josiassen MK, et al. Brexpiprazole in patients with schizophrenia: overview of short- and long-term phase 3 controlled studies. *Acta Neuropsychiatr*. 2017;29(5):278–290.

Nemeth G, Laszlovszky I, Czobor P, et al. Cariprazine versus risperidone monotherapy for treatment of predominant negative symptoms in patients with schizophrenia: a randomised, double-blind, controlled trial. *Lancet*. 2017;389(10074):1103–1113.

Shim JC, Shin JGK, Kelly DL, et al. Adjunctive treatment with a dopamine partial agonist, aripiprazole, for antipsychotic-induced hyperprolactinemia: a placebo-controlled trial. *Am J Psychiatry*. 2007;164(9):1404–1410.

Wong DF, Raoufinia A, Bricmont P, et al. An open-label, positron emission tomography study of the striatal D2/D3 receptor occupancy and pharmacokinetics of single-dose oral brexpiprazole in healthy participants. *Eur J Clin Pharmacol*. 2021;77(5):717–725.

33.4 Dopamine Enhancers

33.4a Dopamine Receptor Agonists

Carlos A. Zarate, Jr., M.D., and Lawrence T. Park, A.M., M.D.

The dopaminergic system has been extensively studied since Arvid Carlsson's discovery of the physiologic functions of dopamine (3-hydroxytryptamine) in 1957. Carlsson, who passed away in 2018, was awarded the 2000 Nobel Prize in Physiology and Medicine (along with

Paul Greengard and Eric Kandel) for his research into dopamine and for his "discoveries concerning signal transduction in the nervous system." Today, the dopaminergic system is among the best understood of all neurotransmitter systems. In humans, four major pathways have been identified: nigrostriatal (A9), mesolimbic (A10), mesocortical (A10), and tuberoinfundibular (A8). These pathways have been implicated in motor activity, affect, reward mediation, attention, cognitive functioning, appetite, and sleep. Dopaminergic dysfunction has been linked to several neurodegenerative diseases, including Parkinson disease (PD). Five dopamine (D) receptor subtypes have been identified: D_1, D_2, D_3, D_4, and D_5. All are G protein–coupled receptors that are classified into two classes: D_1 and D_2. The D_1 class receptors (D_1 and D_5) increase cyclic adenosine monophosphate (cAMP) production via positive coupling to adenylyl cyclase; the D_2 class receptors (D_2, D_3, D_4) are associated with adenylyl cyclase inhibition and with decreased cAMP production.

Dopamine receptor agonists (DAs) bind to and activate dopamine receptors to reproduce the activity of the endogenous neurotransmitter. Classic DAs were initially developed during the 1970s to treat advanced PD. They are typically classified into nonergot and ergot derivatives. Nonergot DAs include apomorphine (Apokyn), pramipexole (Mirapex), ropinirole (Requip), and rotigotine (Neupro). Though successfully used as monotherapy for new-onset PD, they are more commonly used as adjunctive treatments with levodopa (a dopamine precursor); as a result, smaller doses of levodopa are required and complications are delayed or reduced. Apomorphine has D_1 and D_2 agonist effects, and although it has demonstrated efficacy in PD, it is not used in the United States due to its short half-life. However, it is used in Europe for PD via subcutaneous and intravenous pump administration. Pramipexole and ropinirole are FDA approved for use in PD and restless legs syndrome (RLS) and have been studied as an adjunctive treatment for unipolar and bipolar depression. Amantadine (Symmetrel), an adamantane derivative, is FDA approved to treat PD, drug-induced extrapyramidal symptoms (EPS), and influenza A infection. It is also used off-label to treat neuroleptic malignant syndrome (NMS) and Cotard syndrome, a psychiatric condition that involves a delusional belief that one is dead. Rotigotine (Neupro), a continuous-delivery transdermal patch, is FDA approved for the treatment of PD and RLS. Rotigotine binds to all DA receptor subtypes but has the highest affinity at the D_3 receptor. Due to manufacturing issues, the Neupro patch was withdrawn from the U.S. market in 2008 but returned in 2012.

The ergot class of DAs includes bromocriptine (Parlodel), cabergoline (Dostinex, Caberlin), and pergolide (Permax). Bromocriptine is FDA approved for use in PD, galactorrhea (due to hyperprolactinemia), and acromegaly and is used off-label to treat RLS and NMS. Cabergoline is FDA approved for the treatment of hyperprolactinemic disorders, either idiopathic or due to pituitary adenomas. Several studies have also suggested that cabergoline may be effective in treating PD or preventing ovarian hyperstimulation syndrome. Pergolide, an ergot $D_{1/2}$ agonist, was FDA approved for PD and studied as an adjunctive treatment for treatment-resistant depression (TRD). It is important to note that the entire class of ergot derivatives has been associated with cardiac fibrosis. As a result, the FDA issued a warning about the risk of cardiac valvulopathy for cabergoline in 2011, and pergolide was withdrawn from the U.S. market in 2007 because of an increased risk of valvular heart disease.

Positive allosteric modulators of the dopamine D_1 receptor (D_1 PAMs) are currently under investigation in animal models as potential treatments for the motor symptoms of PD, cognitive impairment, somnolence, and immobility. However, none are currently FDA approved or available for clinical use.

Other classes of drugs with other modes of action also exert agonist effects at the dopamine receptor, though they may act at other receptors as well. For instance, the wakefulness-promoting agents modafinil (Provigil) and armodafinil (Nuvigil) have complex prodopaminergic activity, including dopamine transporter inhibition. They may also have dopamine (D_2) receptor partial agonist effects. These drugs are FDA approved to improve wakefulness in adult patients with excessive sleepiness associated with narcolepsy, obstructive sleep apnea (OSA), or shift work disorder (SWD). They have also been studied as adjunctive treatments for unipolar and bipolar depression as well as attention deficit hyperactivity disorder (ADHD). These drugs have also been used in nonpathologic settings to optimize normal performance.

Certain second-generation antipsychotic (SGA) drugs have also demonstrated complex partial agonist or agonist–antagonist properties at the dopamine receptor. Given their mixed activity at the dopamine receptor, these compounds—which include aripiprazole, brexpiprazole, cariprazine, and lumateperone—will not be reviewed further in this chapter.

This chapter describes the key therapeutic DAs with significant psychiatric applications: pramipexole, ropinirole, amantadine, rotigotine, bromocriptine, modafinil, and armodafinil.

CHEMISTRY

Pramipexole is a nonergot synthetic aminothiazole derivative. The chemical name is (*S*)-2-amino-4,5,6,7-tetrahydro-6-(propylaMino)benzothiazole dihydrochloride monohydrate (IUPAC nomenclature). The molecular formula is $C_{10}H_{17}N_3S.2HCl.H_2O$, and the molecular weight is 302.27. It is a white to off-white crystalline powder, freely soluble in water, soluble in methanol, sparingly soluble to slightly soluble in ethanol, and practically insoluble in methylene chloride. The molecular structure is shown in Figure 33.4a–1.

Ropinirole is a nonergot hydrochloride salt. The chemical name is 4-[2-(dipropylamino)ethyl]-1,3-dihydro-2*H*-indol-2-one. The molecular formula is $C_{16}H_{24}N_2O.HCl$, and the molecular weight is 296.84. It is a pale cream to yellow powder, soluble in water and methanol, and very slightly soluble in ethyl alcohol. The molecular structure is shown in Figure 33.4a–2.

Amantadine is a nonergot adamantane derivative. The chemical name is adamantan-1-amine. The molecular formula is $C_{10}H_{17}N$, and the molecular weight is 151.24868. It is a stable white or nearly white crystalline powder, freely soluble in water and soluble in alcohol and chloroform. The molecular structure is shown in Figure 33.4a–3.

Rotigotine is a nonergot tetralin class compound. The chemical name is (6*S*)-6-[propyl(2-thiophen-2-ylethyl)amino]-5,6,7,8-tetrahydronaphthalen-1-ol. The molecular formula is $C_{19}H_{25}NOS$, and the molecular weight is 315.5. It is a white to off-white powder, lipophilic and poorly soluble in water at neutral pH. The molecular structure is shown in Figure 33.4a–4.

Bromocriptine is a semisynthetic ergotamine alkaloid. The chemical name is (5′α)-2-bromo-12′-hydroxy-5′-(2-methylpropyl)-2′-(propan-2-yl)-3′,6′,18-trioxoergotaman. The molecular formula is $C_{32}H_{40}BrN_5O_5$, and the molecular weight is 654.5945. It is a white or slightly colored fine crystalline powder, soluble in methyl alcohol and practically insoluble in water, chloroform, and benzene. The molecular structure is shown in Figure 33.4a–5.

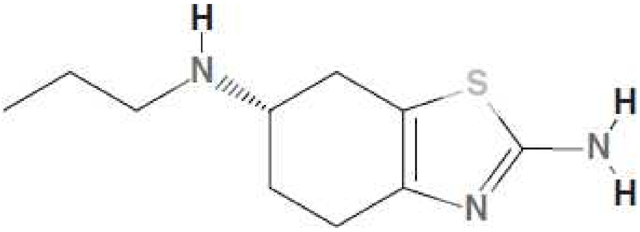

FIGURE 33.4a–1. Molecular structure of pramipexole. (From National Center for Biotechnology Information. PubChem Compound Summary for CID 119570, Pramipexole. 2023. Accessed July 11, 2023. https://pubchem.ncbi.nlm.nih.gov/compound/Pramipexole.)

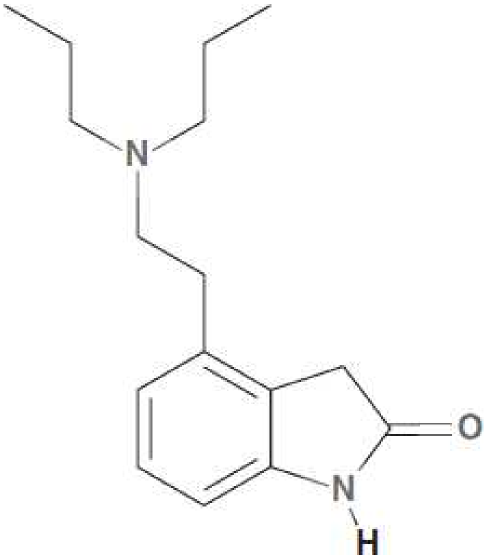

FIGURE 33.4a–2. Molecular structure of ropinirole. (From National Center for Biotechnology Information. PubChem Compound Summary for CID 119570, Ropinirole. 2023. Accessed July 11, 2023. https://pubchem.ncbi.nlm.nih.gov/compound/Ropinirole.)

Modafinil is an oxazolidinone compound, which is a racemic compound consisting of a 1:1 mixture of *R*- and *S*-isomers (enantiomers). Armodafinil is the (*R)*-isomer of the racemic mixture. The chemical name of modafinil is 2-[(diphenylmethyl)sulfinyl]acetamide, and armodafinil is (–)-2-[(*R*)-(diphenylmethyl)sulfinyl]acetamide. For both, the molecular formula is $C_{15}H_{15}NO_2S$, and the molecular weight is 273.35. Modafinil and armodafinil are white to off-white crystalline powders that are sparingly to slightly soluble in methanol and acetone and practically insoluble in water and cyclohexane. The molecular structure is shown in Figure 33.4a–6.

PHARMACOLOGIC ACTIONS

Pharmacokinetics

After oral administration, pramipexole is rapidly absorbed with >90% bioavailability. Absorption is not significantly affected by food. It undergoes little first-pass metabolism. The time to maximum concentration ($T_{\max}$) is approximately 2 hours, and the half-life ($T_{1/2}$) is approximately 8 hours (12 hours in older adult patients). It undergoes minimal metabolism, with 90% of the drug being excreted unchanged in the urine.

Ropinirole is rapidly absorbed after oral administration. Absorption is not significantly affected by food. The absolute bioavailability is 55%, indicating significant first-pass metabolism. The $T_{\max}$ is approximately 1 to 1.5 hours, and the $T_{1/2}$ is approximately 6 hours. Steady-state concentrations are expected within 2 days of dosing. It is primarily metabolized by cytochrome P450 (CYP)1A2 to inactive metabolites and excreted in the urine. Clearance is reduced by approximately 30% in patients above 65 years of age.

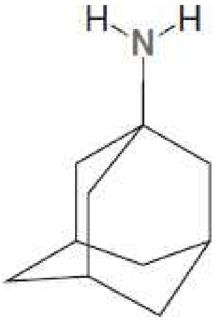

FIGURE 33.4a–3. Molecular structure of amantadine. (From National Center for Biotechnology Information. PubChem Compound Summary for CID 2130, Amantadine. 2023. Accessed July 11, 2023. https://pubchem.ncbi.nlm.nih.gov/compound/Amantadine.)

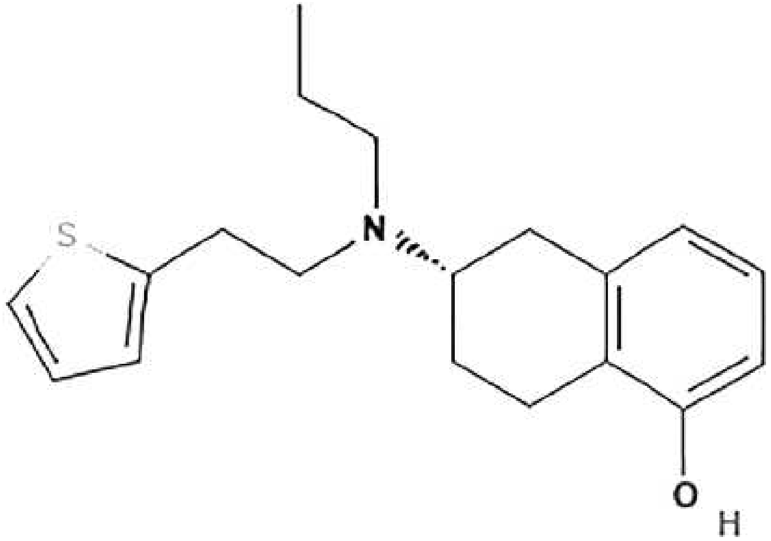

FIGURE 33.4a–4. Molecular structure of rotigotine. (From National Center for Biotechnology Information. PubChem Compound Summary for CID 59227, Rotigotine. 2023. Accessed July 11, 2023. https://pubchem.ncbi.nlm.nih.gov/compound/Rotigotine.)

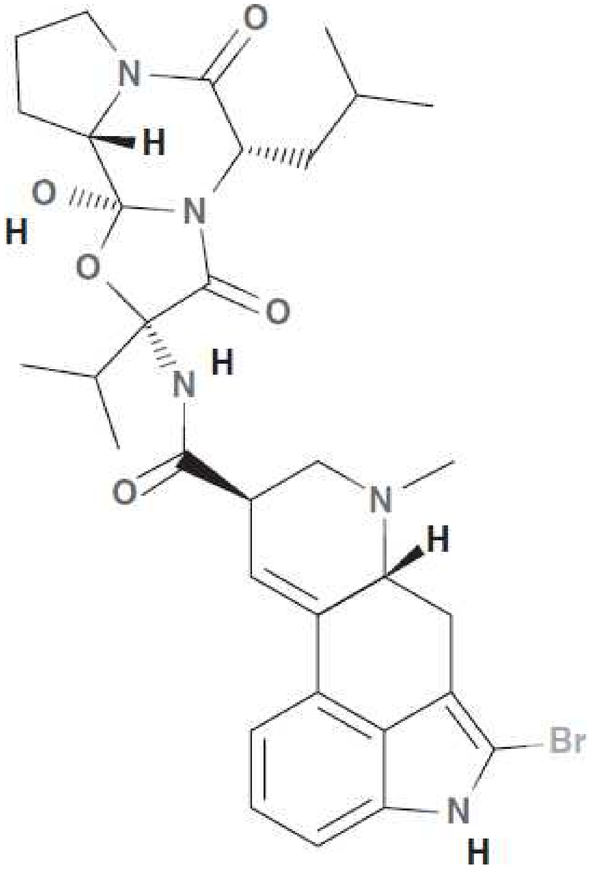

FIGURE 33.4a–5. Molecular structure of bromocriptine. (From National Center for Biotechnology Information. PubChem Compound Summary for CID 31101, Bromocriptine. 2023. Accessed July 11, 2023. https://pubchem.ncbi.nlm.nih.gov/compound/Bromocriptine.)

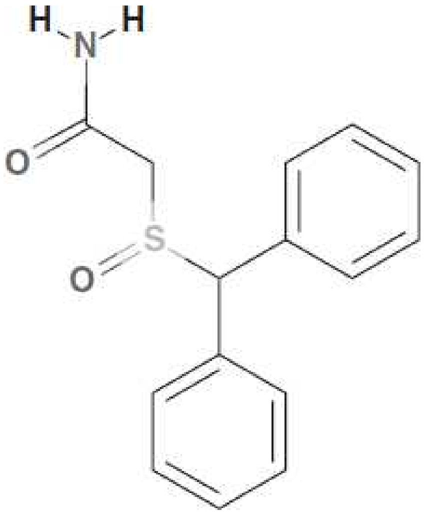

FIGURE 33.4a–6. Molecular structure of modafinil (armodafinil). (From National Center for Biotechnology Information. PubChem Compound Summary for CID 4236, Modafinil. 2023. Accessed July 11, 2023. https://pubchem.ncbi.nlm.nih.gov/compound/Modafinil.)

Amantadine is well absorbed after oral administration. The average $T_{\max}$ is 3.3 hours, and the $T_{1/2}$ is approximately 17 hours. It is primarily excreted unchanged in the urine. Amantadine clearance is significantly reduced in patients with renal insufficiency.

Rotigotine is delivered as a transdermal patch application. Bioavailability varies with the application site, with the highest absorption demonstrated on the shoulder. On average, approximately 45% of the drug is released in a 24-hour period. The average $T_{\max}$ is 15 to 18 hours, though no characteristic peak concentration has been observed. It is rapidly and extensively metabolized via multiple CYP isoforms (particularly 2C19), sulfotransferases, and UDP-glucuronosyltransferases. Rotigotine is primarily eliminated in the urine (71%) as an inactive conjugate. Initial $T_{1/2}$ is 3 hours, with a terminal $T_{1/2}$ of 5 to 7 hours.

The oral absorption rate of bromocriptine is approximately 28%. Absorption is not significantly affected by food, though it is recommended that the drug be administered with food to reduce possible nausea/vomiting. Absolute bioavailability is low (6%), indicating a significant first-pass effect. The $T_{\max}$ is approximately 1 to 2 hours, and the $T_{1/2}$ is 2.5 to 5 hours. It is primarily metabolized by CYP3A4 and excreted in the feces via biliary secretion; 6% is eliminated in the urine.

Modafinil is a racemic compound whose isomers have different kinetics. Armodafinil is the *(R)*-isomer of modafinil. Both isomers are rapidly absorbed after oral administration. The T_{max} is 2 to 4 hours for modafinil and 2 hours for armodafinil. Absorption is not significantly affected by food. The $T_{1/2}$ of armodafinil is approximately three times that of the (*S*)-isomer, so exposure to armodafinil is about three times that of the (*S*)-isomer. While the effective $T_{1/2}$ of both modafinil and armodafinil is approximately 11 to 15 hours after multiple doses, maximum plasma drug concentrations of armodafinil remain higher for a longer period of time because the (*S*)-isomer is cleared more quickly than the (*R*)-isomer. Modafinil is predominantly metabolized by the liver via glucuronide conjugation (as well as CYP3A4/5 metabolism) and is excreted in the urine.

Pharmacodynamics

Pramipexole and ropinirole are nonergot DAs with high specificity and full intrinsic activity at the D_2 subtype receptor and higher affinity binding at the D_3 receptor subtype (compared to D_2 and D_4). Both stimulate dopamine receptors in the striatum. Pramipexole may also exert robust neurotropic effects. Ropinirole demonstrates moderate in vitro affinity for opioid receptors.

Amantadine is a nonergot adamantane derivative that may have direct and indirect effects on the dopamine neuron, including DA, serotonergic antagonist, and glutamatergic (*N*-methyl-D-aspartate [NMDA]) antagonist properties. Its antiviral effects are thought to be mediated by interfering with the viral M2 protein. The mechanism of action for the treatment of PD and EPS is unknown but may involve effects on the dopamine and/or NMDA system.

Rotigotine is a nonergot tetralin compound and nonselective DA, though it has the highest affinity for the D_3 receptor. It is also an α_2 adrenergic antagonist and 5-HT_{1A} agonist. Rotigotine also inhibits dopamine uptake and prolactin secretion.

Bromocriptine is a semisynthetic ergotamine alkaloid with D_2 and D_3 receptor agonist activity. Postsynaptic D_2 stimulation is thought to be primarily responsible for its antiparkinsonian effects, while presynaptic D_2 stimulation is associated with its neuroprotective effects. It also exhibits agonist activity at the 5-HT_{1D}, 5-HT_{1A}, 5-HT_{2A}, 5-HT_{1B}, and 5-HT_{2C} receptors and antagonist activity at the α_{2A}, α_{2B}, α_{2C}, and D_1 receptors.

Bromocriptine also demonstrates partial agonist activity at 5-HT_{2B} receptors and inactivated D_4 and 5-HT_7 receptors. In the tuberoinfundibular pathway, bromocriptine inhibits prolactin secretion by pituitary lactotrophs. Hallucinations caused by bromocriptine are thought to arise due to 5-HT_{2A} agonism, while pulmonary fibrosis may be associated with 5-HT_{1B} and 5-HT_{2B} activities.

Modafinil and its corresponding (*R*)-isomer, armodafinil, are oxazolidinone compounds with complex neurotransmitter effects, including dopamine transporter blockade, dopamine reuptake inhibition, and potential DA properties (including in the nucleus accumbens). Modafinil activates orexin (hypocretin)-containing neurons, which are located exclusively in the lateral hypothalamus and send axons to numerous brain regions, including major nuclei implicated in sleep regulation. Animal studies have found that modafinil increases extracellular glutamate levels in the medial preoptic area and posterior hypothalamus and also decreases γ-aminobutyric acid (GABA) levels in these areas. Evidence suggests that modafinil acts robustly on catecholamines, glutamate, GABA, serotonin, orexin, and histamine systems in the brain.

CLINICAL DRUG STUDIES

Major Depressive Disorder

Several lines of evidence suggest that the dopaminergic system may be an effective target for the treatment of depression. Depressed participants demonstrated decreased levels of the primary dopamine metabolite, homovanillic acid (HVA), in cerebrospinal fluid and blood. Though results have been mixed, some single-photon emission computed tomography (SPECT) studies have observed higher in vivo availability of D_2 receptor binding in depressed participants compared to controls. In postmortem brains, lower levels of dopamine transporter binding were found in the basal and central amygdaloid nuclei of participants with major depressive disorder (MDD) versus healthy volunteers, whereas increased $D_{2/3}$ receptor binding was found in basal, central, and lateral amygdaloid nuclei. This finding supports the theory that a deficiency in mesolimbic dopamine may result in certain associated symptoms of depression (particularly anhedonia and loss of motivation). Finally, certain prodopaminergic agents such as bupropion (a dopamine reuptake inhibitor) have demonstrated efficacy and have been approved for the treatment of MDD.

Pramipexole has been the most studied DA for the treatment of MDD. Several open-label studies have reported the efficacy of pramipexole in augmenting the effects of tricyclic antidepressants and selective serotonin reuptake inhibitors (SSRIs) in TRD. One small, open-label trial of 10 TRD participants (at least 2 of whom had failed to respond to at least two prior antidepressants) demonstrated that pramipexole augmentation significantly improved depressive symptoms from baseline to posttreatment, and 6 participants met the criteria for treatment response (i.e., ≥50% reduction in Montgomery–Asberg Depression Rating Scale [MADRS] score). A randomized, double-blind, placebo-controlled trial of pramipexole augmentation in 60 outpatients with TRD found modest but statistically significant benefits ($p = .038$). In another double-blind, comparative study, 39 participants who had failed a standard antidepressant trial were randomized to receive pramipexole, escitalopram, or the combination. After 6 weeks of treatment, participants who received pramipexole monotherapy demonstrated significantly lower MADRS scores than the combination group ($p = .01$). In the combination group, however, only 15% of participants tolerated regularly scheduled increases in pramipexole. This study demonstrated that concurrent initiation and titration of SSRI and pramipexole resulted in side effects that often led to the cessation of treatment; in contrast, adjunctive use of pramipexole appeared to be better tolerated when it was added to ongoing SSRI treatment. In addition, a systematic review and meta-analysis that examined pramipexole use concluded that for MDD and bipolar depression, pramipexole was safe and well tolerated and that it demonstrated comparable efficacy in the short-term and perhaps greater overall long-term remission rates than traditional antidepressants.

Ropinirole has also been studied as an augmentation strategy for standard antidepressants. An open-label trial of 10 participants (7 with TRD) observed a statistically significant 4.5-point decrease in MADRS score from baseline to after 16 weeks of treatment. Another study demonstrated the effectiveness of ropinirole for antidepressant-induced sexual dysfunction.

Modafinil augmentation for MDD has also been extensively studied. A meta-analysis of six randomized controlled trials (four MDD trials) found that modafinil augmentation was well tolerated and effective for acute MDD, including for symptoms of fatigue.

Bipolar Depression

There is growing support for the use of pramipexole in bipolar depression. Specifically, several case reports, case series, and open-label trials have reported the efficacy of pramipexole in augmenting tricyclic antidepressants and SSRIs in TRD, and approximately half of the reported participants were diagnosed with bipolar disorder. One randomized, controlled trial found that pramipexole was associated with a superior response compared with placebo (60% vs. 9%) for bipolar II depression when added to therapeutic levels of lithium or valproate. Overall, pramipexole was found to have a large effect size (0.6 to 1.1) in the treatment of both MDD and bipolar depression and was associated with a low switch rate (1% mania, 5% hypomania). A systematic review and meta-analysis examining pramipexole concluded that, for unipolar and bipolar depression, pramipexole was safe and well tolerated and had comparable efficacy in the short-term and perhaps greater overall long-term remission rates than traditional antidepressants. In addition, a functional neuroimaging study of 15 participants with bipolar II depression treated with 6 weeks of pramipexole augmentation of mood stabilizers versus placebo found that pramipexole treatment was associated with reductions in normalized metabolism in the bilateral orbitofrontal cortex, left ventrolateral prefrontal cortex, and right anteromedial prefrontal cortex. Voxel-wise analyses also showed decreased normalized metabolism in the left inferior parietal cortex and medial frontopolar cortical (BA 10P) area of the anteromedial prefrontal cortex.

Ropinirole has also been reported to augment the efficacy of standard antidepressants. Limited evidence also supports its off-label use for augmenting antidepressant response in individuals with treatment-resistant bipolar depression or for alleviating antidepressant-induced sexual dysfunction. In one open-label trial of acutely depressed bipolar participants, ropinirole used in combination with lithium demonstrated a significant treatment effect. No switching into hypomania or mania was observed.

Modafinil and armodafinil have also been studied in the treatment of bipolar depression. A meta-analysis of six randomized controlled trials (four that included individuals with bipolar depression and two conducted exclusively in individuals with bipolar depression) demonstrated that modafinil augmentation was well tolerated and effective for acute bipolar depression, including for symptoms of fatigue. Another placebo-controlled trial found that modafinil was effective as monotherapy for bipolar depression. Finally, another randomized, double-blind, placebo-controlled, multicenter trial found that armodafinil was a safe and effective adjunctive treatment for bipolar depression.

Mood and Fatigue Related to PD (and Other Medical Conditions)

Some estimates suggest that depressive symptoms occur in up to 45% of individuals with PD, a factor that significantly affects their quality of life. Given the dysfunction of dopaminergic nigrostriatal pathways in PD, dopaminergic mesolimbic reward pathways may be similarly affected, resulting in depressive symptoms such as anhedonia or lack of motivation. Several studies support the efficacy of DAs in treating depressive symptoms associated with PD. In a large, open-label trial of 657 participants with depressive symptoms associated with PD, anhedonia and depression were reduced when pramipexole was used adjunctively with L-dopa. In addition, pramipexole was superior to sertraline in treating depression associated with PD and also reduced anhedonia in PD participants. Another randomized, double-blind, placebo-controlled 12-week trial of 323 individuals with mild-to-moderate PD found that pramipexole decreased depression scores (on the Beck Depression Inventory) more than placebo ($p = .01$). A meta-analysis of seven randomized controlled trials investigating the effect of pramipexole on mood and motivation symptoms in PD participants who did not have MDD concluded that the drug had beneficial effects on both. However, a systematic review of the literature by the same authors concluded that, given the quality of the data, further studies are required to determine if DAs can effectively treat depressive symptoms in PD.

DA agonists have also been used to improve cognitive recovery and rehabilitation in individuals with traumatic brain injury. In a systematic review of methylphenidate, amantadine, and bromocriptine, however, no positive trend could be established for any of the drugs.

Armodafinil has also been studied as a treatment for sleepiness and fatigue related to PD and HIV. In one study, armodafinil effectively treated 10 individuals with PD with excessive daytime sleepiness and was noted to increase tolerance for higher doses of DAs that may cause sleepiness as a side effect. Also, a placebo-controlled, randomized, double-blind trial of 70 HIV-positive participants found that armodafinil successfully reduced HIV-related fatigue. Although armodafinil did not improve depressive symptoms in the absence of fatigue improvement, it did improve depressive symptoms in those with an axis I depressive disorder who experienced an increase in energy.

Restless Legs Syndrome and Periodic Limb Movements of Sleep

Pramipexole and ropinirole are FDA approved for the treatment of moderate-to-severe primary RLS. Pramipexole was effective in four randomized controlled trials (two fixed-dose studies, one flexible-dose study, and one treatment withdrawal with survival analysis) across approximately 1,000 participants. Ropinirole was effective in three randomized, double-blind, placebo-controlled studies involving over 500 participants. A large Cochrane review of 35 placebo-controlled and three active-controlled randomized clinical trials concluded that DAs are superior to placebo for up to 7 months. Cabergoline and pramipexole were more effective than L-dopa for some but not all outcomes.

The transdermal rotigotine patch (Neupro) is FDA approved for the treatment of moderate-to-severe primary RLS. Support for this indication comes from two multicenter, randomized, controlled trials. In the first study, 505 participants with moderate-to-severe RLS were randomized to placebo or four different doses of rotigotine (0.5 mg/24 hours, 1 mg/24 hours, 2 mg/24 hours, or 3 mg/24 hours). Participants began at the lowest dose and were titrated to their assigned dose over 4 weeks. They were followed for 6 months and then tapered off treatment. Compared to placebo, participants receiving active treatment experienced a mean decrease in RLS symptoms (as measured by the International RLS Rating Scale [IRLS]) and improved Clinical Global Impression—Improvement (CGI-I) scores, with the higher dose groups demonstrating greater improvement. In the second study, 458 participants with moderate-to-severe RLS were randomized to receive placebo or three different doses of rotigotine (1 mg/24 hours, 2 mg/24 hours, or 3 mg/24 hours). Treatment began at the lowest dose and was titrated to the assigned dose over a 3-week period. Participants were followed for 6 months and then tapered off treatment. All three active treatment groups demonstrated statistically significant improvements in IRLS and CGI-I scores compared to placebo.

The American Academy of Neurology issued practice guidelines for the treatment of RLS in 2016. For moderate-to-severe primary RLS, strong evidence supports the use of pramipexole, rotigotine, cabergoline, and gabapentin enacarbil (Level A); moderate evidence supports ropinirole, pregabalin, and IV ferric carboxymaltose use (Level B); L-dopa may also be considered (Level C). Few head-to-head comparisons exist to suggest agents preferentially, but the guidelines suggest that when treating periodic limb movements of sleep, clinicians should consider prescribing ropinirole (Level A) or pramipexole, rotigotine, cabergoline, or pregabalin (Level B).

Extrapyramidal Symptoms

Amantadine is FDA approved for the treatment of drug-induced EPS. The mechanism of action of this effect is unclear. A double-blind study found that amantadine was as effective as benztropine for drug-induced EPS but had fewer side effects.

Neuroleptic Malignant Syndrome

There are no currently approved drug treatments for NMS. Treatment generally consists of supportive care and the immediate removal of the presumed offending agent as well as any other dopamine antagonists. Though data are limited, pharmacologic considerations may include benzodiazepines (for milder cases), DAs, dantrolene, and electroconvulsive therapy. Bromocriptine (starting at 2.5 mg b.i.d.–t.i.d., maximum total daily dose 45 mg) and amantadine (200 to 400 mg/day in divided doses) are DAs that may enhance the presynaptic release of dopamine to treat the presumed underlying cause of NMS. They might be particularly helpful if NMS was caused by withdrawal of medication to treat PD. However, bromocriptine may induce or exacerbate serotonin syndrome, so this diagnosis should be clearly ruled out before initiating a trial of bromocriptine for NMS.

Excessive Daytime Sleepiness

Modafinil has been approved for use in excessive daytime sleepiness secondary to narcolepsy, shift work sleep disorder, and OSA. Two large, placebo-controlled, randomized, double-blind, parallel-group studies found that modafinil was effective and well tolerated in the treatment of excessive daytime sleepiness secondary to narcolepsy and that effects were maintained for at least 40 weeks. At doses of 400 mg/day, modafinil was also effective in 10 individuals with excessive daytime sleepiness secondary to head injury and in 2 individuals with excessive somnolence due to sedating drugs; no side effects or evidence of loss of effect or tolerance were noted over 5 to 13 months of continued use. Another open-label study of 72 participants with fatigue associated with multiple sclerosis found that a dose of 200 mg per day of modafinil was well tolerated and significantly improved fatigue.

Armodafinil has also been approved to treat excessive daytime sleepiness secondary to narcolepsy, shift work sleep disorder, and OSA/hypopnea syndrome. One large, placebo-controlled, multicenter trial found that armodafinil significantly improved the ability to remain awake compared to placebo in individuals with narcolepsy and shift work sleep disorder. In addition, two large, placebo-controlled trials showed that armodafinil significantly improved the ability to remain awake compared to placebo in participants with OSA/hypopnea syndrome.

Schizophrenia-Associated Cognitive Dysfunction

In schizophrenia, modafinil and armodafinil have been studied as augmentation strategies targeting negative symptoms, fatigue/sleepiness, and cognition. A systematic review and meta-analysis concluded that both agents were well tolerated and reduced negative symptoms, though the effect size was small. The effects of these agents on fatigue/sleepiness are less clear. Likewise, though a variety of studies have investigated the ability of these agents to enhance certain aspects of cognition (e.g., alertness, attention, executive functioning, memory) in conditions such as depression, substance abuse/dependence, neurologic conditions (e.g., multiple sclerosis, dementia, brain tumors, stroke), and cancer, additional research is needed to fully assess their effects.

Performance Enhancement

While both modafinil and armodafinil were developed to ameliorate pathologic conditions, both have increasingly been used to enhance normal human functioning. Initially used to increase alertness and physical performance in military personnel (sometimes taking the place of psychostimulants), the off-label use of these agents is expanding to enhance performance in a variety of settings; reports have noted their use by military/rescue personnel, athletes, academicians, university students, and video-game participants. Improved performance has been investigated in a variety of settings, such as sleep deprivation, game performance (i.e., chess), and emotional recognition. The use of these agents as performance enhancers in nonpathologic conditions has prompted a public discussion of related ethical questions; these include which specific areas of functioning are affected, what risks accompany their use, how the absence of disease affects the benefit–risk assessment, what parameters are appropriate for their use, whether their use is fair in competitive settings, what the limits are to their enhancing effects, what oversight of this use is appropriate, and how enhancement might ultimately influence human experience.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Central Nervous System

DAs display prominent central nervous system effects. From their initial development as a treatment for PD, the prodopaminergic activity of the classic DAs underlies their effectiveness for RLS, mood disorders, EPS, and NMS. Central prodopaminergic activity also mediates many adverse events, including depression, irritability, confusion, hallucinations, dyskinesias, and impulsive/compulsive behavior. The increased wakefulness, energy, initiative, self-confidence, positive mood, and enhanced physical and mental performance associated with the wakefulness-promoting agents modafinil and armodafinil can often be dramatic and of rapid onset. As a result, the potential for abuse and the danger of dependence or addiction must be kept in mind.

Cardiovascular System

The entire class of ergot DA agonists has been associated with cardiac fibrosis. As a result, pergolide was withdrawn from the U.S. market in 2007 because of an increased risk of valvular heart disease. Cabergoline remains on the market, but in 2011, the FDA issued a warning regarding the risk of cardiac valvulopathy. Bromocriptine has been associated with an increased risk of myocardial infarction in postpartum women and with pericardial and pulmonary fibrosis, particularly with long-term, high-dose treatment. Congestive heart failure in conjunction with amantadine administration has also been reported. In general, most DAs can produce hypotension. However, modafinil and armodafinil are less commonly associated with adverse cardiovascular events.

Immunologic Effects

Modafinil and armodafinil have been associated with certain hypersensitivity reactions, particularly in pediatric participants. In premarket studies, 0.8% of pediatric participants developed a rash resulting in drug discontinuation; one case proceeded to Stevens–Johnson syndrome, while another developed multiorgan hypersensitivity reaction. With modafinil, one case of angioedema was reported, and with armodafinil, one case of hypersensitivity and one case of angioedema were reported. Based on the possibility of hypersensitivity reactions, modafinil did not receive FDA approval for the treatment of ADHD, despite good effectiveness data.

The rotigotine transdermal patch (Neupro) contains sodium metabisulfite; sulfite sensitivity may develop, including allergic-type reactions, anaphylactic symptoms, and life-threatening or less severe asthmatic episodes in certain susceptible people.

Endocrine Effects

Bromocriptine (as well as the other ergot class DAs) is a known suppressor of prolactin release and is thus used to treat galactorrhea (due to hyperprolactinemia), prolactin-secreting adenomas, and infertility due to hyperprolactinemia-related amenorrhea. In healthy individuals, bromocriptine increases growth hormone levels in the short term. In individuals with acromegaly, bromocriptine lowers growth hormone release and lowers adrenocorticotropic hormone (ACTH) release in patients with Cushing disease or Nelson syndrome.

THERAPEUTIC IMPLICATIONS

- Pramipexole and ropinirole are FDA approved for use in PD and RLS and have been studied as adjunctive treatments for MDD and bipolar depression.
- Rotigotine is FDA approved to treat the signs and symptoms of PD and moderate-to-severe RLS.

- Amantadine is FDA approved to treat PD, drug-induced EPS, and influenza A infection. It is also used off-label to treat NMS and Cotard syndrome, a psychiatric condition that involves a delusional belief that one is dead.
- Bromocriptine is FDA approved for use in PD, galactorrhea (due to hyperprolactinemia), and acromegaly and is used off-label to treat RLS and NMS.
- Modafinil and armodafinil are FDA approved to improve wakefulness in adults with excessive sleepiness associated with narcolepsy, OSA, or SWD. They have also been studied as adjunctive treatments for MDD, bipolar depression, ADHD, fatigue, and cognitive dysfunction associated with medical conditions. They also have been used in nonpathologic settings to optimize normal performance.

PRECAUTIONS AND ADVERSE REACTIONS

Each class of DAs has characteristic side effects that should be considered and discussed with patients before they are prescribed. Table 33.4a–1 lists common and potentially serious adverse events associated with each drug. As regards pregnancy risk, the previous U.S. FDA risk categories of A, B, C, D, and X have been phased out and replaced with narrative descriptions of risk.

Common side effects of the nonergot drugs pramipexole and ropinirole include somnolence, orthostatic hypotension, dyskinesias, and hallucinations. The use of these drugs has been associated with the possibility of sudden onset of sleep in the midst of daily activities, including driving. In addition, rare cases of impulsive/compulsive behavior (i.e., pathologic gambling, hypersexuality, compulsive eating [including binge eating], and increased spending) have been reported. Some of these side effects may be dose related, while some may be related to individual susceptibility. One meta-analysis of nonergot DAs in PD found an increased incidence of nausea and other symptoms compared to placebo. In addition, a greater incidence of side effects such as dizziness, somnolence, constipation, vomiting, and insomnia was only found in early PD, while a higher risk of dyskinesia and hallucinations was only found in advanced PD.

For pramipexole, there are no adequate data on the developmental risk of use in pregnant women. Ropinirole has been shown to have adverse effects on embryo–fetal development, including teratogenic effects, in animal studies; however, no adequate and well-controlled studies exist regarding its use in pregnant women. Thus, ropinirole is recommended for use during pregnancy only if the potential benefit outweighs the potential risk to the fetus (previous category C).

Common side effects of rotigotine include nausea, vomiting, skin reaction at the application site, dizziness, somnolence, fatigue, upper respiratory infections, and anorexia. In addition, its use has been associated with increased risk of hallucinations, psychotic-like behavior, symptomatic hypotension, syncope, impulsive/compulsive behavior (i.e., pathologic gambling, hypersexuality, compulsive eating [including binge eating], and increased spending), increased blood pressure and heart rate, weight gain, fluid retention, and dyskinesias. Because the rotigotine transdermal patch contains sodium metabisulfite, sulfite sensitivity may develop, with allergic-type reactions, anaphylactic symptoms, and life-threatening or less severe asthmatic episodes in certain susceptible people. In addition, the backing layer of the rotigotine patch contains aluminum and therefore should be removed prior to magnetic resonance imaging or cardioversion. While animal studies have demonstrated increased fetal risk, there are presently no adequate and well-controlled studies on pregnant women. Use during pregnancy is only recommended if the potential benefits outweigh the potential risks to the fetus (previous category C). Effective methods of contraception should be encouraged for women of childbearing potential.

Common side effects of amantadine include nausea, dizziness, and insomnia. Psychiatric side effects such as depression, anxiety, irritability, hallucinations, and confusion have also been reported. Care should be taken in individuals with seizure disorders, as amantadine may increase seizure activity. Several cases of congestive heart failure in conjunction with amantadine administration have also been reported. Deaths from an overdose using amantadine have also occurred, with the lowest reported acute lethal dose of 1 g. When the drug is tapered or discontinued, care should be taken to decrease the dose gradually because abrupt discontinuation may precipitate a parkinsonian crisis. In addition, sporadic cases of possible NMS have been reported. Animal studies suggest a potential risk for fetal harm with amantadine; however, there are no adequate data regarding its use in pregnant women (previous category C).

Table 33.4a–1.

Adverse Events Associated with Dopamine Receptor Agonists

Drug	Common Adverse Events	Other Potential Serious Adverse Events
Pramipexole Pregnancy risk: C Ropinirole Pregnancy risk: C	Dizziness, somnolence, orthostatic hypotension, dyskinesias, hallucinations, constipation, vomiting, and insomnia.	Episodes of sudden sleep onset, impulsive/compulsive behavior (i.e., pathologic gambling, hypersexuality, compulsive eating [including binge eating], and increased spending).
Amantadine Pregnancy risk: C	Nausea, dizziness, insomnia, depression, anxiety, irritability, hallucinations, and confusion.	Congestive heart failure, increased seizure activity. Abrupt discontinuation may result in parkinsonian crisis, neuroleptic malignant syndrome.
Rotigotine Pregnancy risk: C	Nausea, vomiting, skin reaction, dizziness, somnolence, fatigue, upper respiratory infections, anorexia, hallucinations, psychotic-like behavior, symptomatic hypotension, syncope, increased blood pressure and heart rate, weight gain, fluid retention, and dyskinesias	Allergic-type reactions, anaphylactic symptoms, and life-threatening or less severe asthmatic episodes in certain susceptible people. Episodes of impulsive/compulsive behavior (i.e., pathologic gambling, hypersexuality, compulsive eating [including binge eating], and increased spending). Due to metal in the patch, it

		should be removed prior to MRI or cardioversion.
Bromocriptine (ergot) Pregnancy risk: B	Nausea, headache, dizziness, fatigue, light-headedness, vomiting, abdominal cramps, nasal congestion, constipation, diarrhea, drowsiness, confusion, mental disturbance, hallucinations, impulse control (intense urges to gamble, increased sexual urges, intense urges to spend money uncontrollably, and other intense urges), somnolence, hypotension (in PD).	Pleural/pericardial effusion, pleural/pulmonary fibrosis, constrictive pericarditis, retroperitoneal fibrosis. Episodes of sudden sleep onset (PD). Cold sensitive digital vasospasm, tumor expansion, GI bleed from peptic ulcer (acromegaly).
Modafinil Pregnancy risk: C Armodafinil Pregnancy risk: C	Headache, nausea, nervousness, rhinitis, diarrhea, back pain, anxiety, insomnia, dizziness, dyspepsia, mania, delusions, hallucinations, suicidal ideation, and aggression.	Stevens–Johnson syndrome, multiorgan hypersensitivity reaction, angioedema, substance abuse.

GI, gastrointestinal; MRI, magnetic resonance imaging; PD, Parkinson disease.

Common side effects of bromocriptine include nausea, headache, dizziness, fatigue, light-headedness, vomiting, abdominal cramps, nasal congestion, constipation, diarrhea, and drowsiness. Somnolence and symptomatic hypotension may occur, especially in individuals with PD. Because episodes of sudden sleep onset may occur, individuals with PD taking or prescribed bromocriptine should be cautioned not to drive or operate machinery. Like other ergot derivatives, bromocriptine increases the risk of fibrotic complications. Pleural and pericardial effusions, pleural and pulmonary fibrosis, constrictive pericarditis, and retroperitoneal fibrosis have all been reported, especially with long-term, high-dose treatment. Caution should be taken in individuals with PD with a history of myocardial infarction who have a residual, atrial, nodal, or ventricular arrhythmia. In individuals with PD, there is also an increased risk of psychiatric complications such as confusion, mental disturbance, hallucinations, impulse control (with intense urges to gamble, increased sexual urges, intense urges to spend money uncontrollably, and other intense urges). Hallucinations typically resolve with dose reduction. In individuals with acromegaly, there have been reports of cold-sensitive digital vasospasm, possible tumor expansion, and severe GI bleeding from peptic ulcers (with some fatalities). Because bromocriptine use may restore fertility, women of childbearing age who do not wish to conceive should practice a reliable method of contraception. In humans, studies have not demonstrated an increased risk of abnormalities in pregnant women (previous category B). In addition, rare instances of myocardial infarction, seizures (including status epilepticus), and strokes have been reported in postpartum women receiving bromocriptine.

Modafinil and armodafinil are generally well tolerated. Common side effects include headache, nausea, nervousness, rhinitis, diarrhea, back pain, anxiety, insomnia, dizziness, and dyspepsia. Psychiatric complications of mania, delusions, hallucinations, suicidal ideation, and aggression have been reported. There have also been some reports of abuse; however, the abuse potential of this class of drugs is significantly lower than with psychostimulants. In pediatric patients, a rash resulting in drug discontinuation occurred in 13 of 1,585 participants (0.8%). One of these cases proceeded to Stevens–Johnson syndrome, and there was also one report of multiorgan hypersensitivity reaction. The median time to rash was 13 days. Rashes and Stevens–Johnson syndrome have not been observed in adults. In postmarket surveillance, modafinil was associated with one case of angioedema, and armodafinil was associated with one case of hypersensitivity and one case of angioedema. Previously, modafinil and armodafinil had been classified as pregnancy risk C. However, recent preliminary data demonstrated an increased incidence of major congenital malformations associated with modafinil compared to methylphenidate or with no exposure. Though these are preliminary results, it is recommended that modafinil and armodafinil not be used in pregnant women or women planning to conceive.

DRUG INTERACTIONS

For all DAs, coadministration with dopamine antagonists may result in decreased levels and decreased effect.

Because pramipexole is not significantly metabolized by CYP enzymes, it has no significant effects on these enzymes. It is renally cleared. Coadministration of cimetidine, an inhibitor of renal tubular secretions, may increase the maximum concentration (C_{max}) of pramipexole by 40% and its $T_{1/2}$ by 40%.

Ropinirole is a CYP1A2 inhibitor, and coadministration of CYP1A2 substrates may therefore result in higher concentrations of the substrate drug. For instance, in one study, coadministration with L-dopa increased C_{max} by 20%.

Rotigotine is primarily metabolized by conjugation and N-dealkylation by multiple CYP isoforms, sulfotransferases, and two UDP-glucuronosyltransferases. Given these multiple metabolic pathways, it is unlikely that inhibiting any one pathway would significantly alter rotigotine concentrations. There is a low risk that rotigotine inhibits CYP2C19 and CYP2D6 metabolism.

Bromocriptine is a CYP3A4 substrate, and CYP3A4 inducers such as macrolide antibiotics, azole antimycotics, and HIV protease inhibitors may thus increase serum levels of this agent. In one study, coadministration of bromocriptine and erythromycin increased bromocriptine levels by a factor of four.

Amantadine is primarily excreted unchanged in the urine, and clearance is thus significantly reduced in individuals with renal insufficiency. Coadministration of amantadine with quinine or quinidine may reduce the renal clearance of amantadine by up to 30%. When amantadine is administered with anticholinergics, the anticholinergic-like side effects of amantadine may be potentiated. Coadministration of thioridazine with amantadine may worsen tremors in older adults with PD. The use of amantadine may potentially interfere with intranasal live attenuated influenza vaccine (though this has not been evaluated). However, no interaction has been observed between amantadine and the trivalent inactivated influenza vaccine.

Modafinil and armodafinil are reversible moderate inhibitors of CYP2C19, a drug-metabolizing enzyme, and may increase blood levels of diazepam (Valium), phenytoin, propranolol (Inderal), omeprazole, and clomipramine. In individuals deficient in the CYP2D6 enzyme (roughly 7% to 10% of the Caucasian population and similar or lower for other populations), drug levels of tricyclic antidepressants, SSRIs, and bupropion may be increased. In addition, chronic administration of modafinil may lead to a modest induction of CYP3A4 that may reduce levels of steroid contraceptives, triazolam, cyclosporine (Neoral), and theophylline (Bronkodyl), and dose adjustments of these medications may be required; armodafinil is associated with moderate, rather than modest, CYP3A4 induction. Neither modafinil nor armodafinil has been studied in combination with monoamine oxidase inhibitors (MAOIs); caution should thus be used when concomitantly administering MAOIs and these drugs. Because modafinil and armodafinil are partially cleared by CYP3A enzymes, coadministration of potent CYP3A4/5 inducers (e.g., carbamazepine, phenobarbital, and rifampin) or inhibitors (e.g., ketoconazole, erythromycin) may significantly alter drug plasma levels.

LABORATORY INTERFERENCE

No significant laboratory interference has been reported.

DOSAGE AND ADMINISTRATION

Typical doses for the given indications are presented in Table 33.4a–2.

For pramipexole, the recommended dose range for PD is 0.125 mg t.i.d. (total daily dose 0.375 mg) to 1.5 mg t.i.d. (total daily dose 4.5 mg). The recommended dose range for RLS is 0.125 mg daily to 0.5 mg daily (2 to 3 hours before bedtime). For adjunctive treatment of depression, studies have used 0.375 to 3 mg (total daily dose).

For ropinirole, the recommended dose range for PD is 0.25 t.i.d. (total daily dose 0.75 mg) to 1 mg t.i.d. (total daily dose 3 mg). The recommended dose range for RLS is 0.25 mg to 4 mg at bedtime. For adjunctive treatment of depression, studies have investigated doses up to 200 mg total daily dose.

For amantadine, the recommended dose range for PD is 100 to 400 mg total daily dose. For EPS, the recommended dose range is 100 to 300 mg total daily dose. For influenza A, the recommended adult dose range is 100 to 200 mg total daily dose. For pediatric patients (1 to 9 years of age), the total daily dose should be calculated on the basis of 2 to 4 mg/lb/day (4.4 to 8.8 mg/kg/day), not to exceed 150 mg total daily dose. For NMS, off-label use has typically been 200 to 400 mg total daily dose.

Table 33.4a–2.

Dosing of Dopamine Receptor Agonist Drugs for Neuropsychiatric Indications

Drug	Indication	Dose
Amantadine	PD EPS NMS (off-label)	100–400 mg total daily dose 100–300 mg total daily dose 200–400 mg total daily dose
Armodafinil	All indications	150–250 mg total daily dose
Bromocriptine	PD NMS (off-label)	1.25–100 mg total daily dose 2.5 mg b.i.d.–15 mg t.i.d. (5–45 mg total daily dose)
Modafinil	All indications	200–400 mg total daily dose
Pramipexole	PD RLS Depression (investigational)	0.125–1.5 mg t.i.d. (0.375–4.5 mg total daily dose) 0.125–0.5 mg (2–3 hours before bedtime) 0.375–3 mg (total daily dose)
Ropinirole	PD RLS Depression (investigational)	0.25–1 mg t.i.d. (0.75–3 mg total daily dose) 0.25–4 mg at bedtime Up to 200 mg total daily dose
Rotigotine	PD RLS	1 mg/24 hours–8 mg/24 hours 1 mg/24 hours–3 mg/24 hours

EPS, extrapyramidal symptoms; MDD, major depressive disorder; NMS, neuroleptic malignant syndrome; PD, Parkinson disease; RLS, restless leg syndrome.

For bromocriptine, the recommended dose range for hyperprolactinemia is 1.25 to 15 mg total daily dose. For acromegaly, the recommended dose range is 1.25 to 100 mg total daily dose. For PD, the recommended dose range is 1.25 to 100 mg total daily dose. For NMS, off-label use has typically been 5 mg (2.5 mg b.i.d) to 45 mg total daily dose.

For rotigotine, the initial recommended dose for PD is 2 mg/24 hours for early-stage disease or 4 mg/24 hours for advanced-stage disease. The dose may be increased as needed by 2 mg/24 hours at weekly intervals, up to 6 mg/24 hours for early-stage disease and up to 8 mg/24 hours for advanced-stage disease. For RLS, the initial recommended dose is 1 mg/24 hours, increased as needed by 1 mg/24 hours at weekly intervals, up to 3 mg/24 hours.

Dose ranges for modafinil are 200 to 400 mg total daily dose, and dose ranges for armodafinil are 150 to 250 mg total daily dose.

FUNDING AND DISCLOSURES

Funding for this work was provided by the Intramural Research Program at the National Institute of Mental Health, National Institutes of Health (IRP-NIMH-NIH; ZIAMH002857). Dr. Zarate is listed as a co-inventor on a patent for the use of ketamine in major depression and suicidal ideation; as a co-inventor on a patent for the use of (2*R*,6*R*)-hydroxynorketamine, (*S*)-dehydronorketamine, and other stereoisomeric dehydroxylated and hydroxylated metabolites of (*R,S*)-ketamine metabolites in the treatment of depression and neuropathic pain; and as a co-inventor on a patent application for the use of (2*R*,6*R*)-hydroxynorketamine and (2*S*,6*S*)-hydroxynorketamine in the treatment of depression, anxiety, anhedonia, suicidal ideation, and PTSD. He has assigned his patent rights to the U.S. government but will share a percentage of any royalties that may be received by the government. Dr. Park has no conflict of interest to disclose, financial or otherwise.

ACKNOWLEDGMENTS

The authors thank the 7SE research unit and staff for their support. Ioline Henter (NIMH) provided invaluable editorial assistance.

FURTHER READINGS

Aiken CB. Pramipexole in psychiatry: a systematic review of the literature. *J Clin Psychiatry*. 2007;68(8):1230–1236.

Barone P, Poewe W, Albrecht S, et al. Pramipexole for the treatment of depressive symptoms in patients with Parkinson's disease: a randomized, double-blind, placebo-controlled trial. *Lancet Neurol*. 2010;9(6):573–580.

Battleday RM, Brem AK. Modafinil for cognitive neuroenhancement in healthy non-sleep-deprived subjects: a systematic review. *Eur Neuropsychopharmacol*. 2015;25:1865–1881.

Beaulieu JM, Gainetdinov RR. The physiology, signaling and pharmacology of dopamine receptors. *Pharmacol Rev*. 2011;63(1):182–217.

Bienvenu OJ, Neufeld KJ, Neendham DM. Treatment of four psychiatric emergencies in the intensive care unit. *Crit Care Med*. 2012;40(9):2662–2670.

Brooks DJ. Dopamine agonists: their role in the treatment of Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 2000;68(6):685–689.

Calabrese JR, Frye MA, Yang R, Ketter TA; Armodafinil Treatment Trial Study Network. Efficacy and safety of adjunctive armodafinil in adults with major depressive episodes associated with bipolar I disorder: a randomized, double-blind, placebo-controlled multicenter trial. *J Clin Psychiatry*. 2014;75(10):1054–1061.

Cole MG, Elie LM, McCusker J, Bellavance F, Mansour A. Feasibility and effectiveness of treatments for post-stroke depression in elderly inpatients: systematic review. *J Geriatr Psychiatry Neurol*. 2001;14(1):37–41.

DiMascio A, Bernardo CL, Greenblatt DJ, Marder JE. A controlled trial of amantadine in drug-induced extrapyramidal disorders. *Arch Gen Psychiatry*. 1976;33(5):599–602.

Dresler M, Sandberg A, Bublitz C, et al. Hacking the brain: dimensions of cognitive enhancement. *ACS Chem Neurosci*. 2019;10:1137–1148.

Frenette AJ, Kanji S, Rees L, et al. Efficacy and safety of dopamine agonists in traumatic brain injury: a systematic review of randomized controlled trials. *J Neurotrauma*. 2012;29:1–18.

Goss AJ, Kaser M, Costafreda SG, Sahakian BJ, Fu CHY. Modafinil augmentation therapy in unipolar and bipolar depression: a systematic review and meta-analysis of randomized controlled trials. *J Clin Psychiatry*. 2013;74(11):1101–1107.

Klimek V, Schenck JE, Han H, Stockmeier CA, Ordway GA. Dopaminergic abnormalities in amygdaloid nuclei in major depression: a postmortem study. *Biol Psychiatry*. 2002;52(7):740–748.

Kredlow MA, Keshishian A, Oppenheimer S, Otto MW. The efficacy of modafinil as a cognitive enhancer: a systematic review and meta-analysis. *J Clin Psychopharmacol*. 2019;39:455–461.

Kulkarni SK, Dhir A. Current investigational drugs for major depression. *Expert Opin Investig Drugs*. 2009;18(6):767–788.

Lam YW. Clinical pharmacology of dopamine agonists. *Pharmacotherapy*. 2000;20(1 Pt 2):17S–25S.

Lancranian I. The endocrine profile of bromocriptine: its application in endocrine diseases. *J Neural Transm*. 1981;51:61–82.

Leentjens AFG, Koester J, Fruh B, Shephard DTS, Houben JJG, Barone P. The effect of pramipexole on mood and motivational symptoms in Parkinson's disease: a meta-analysis of placebo-controlled studies. *Clin Ther*. 2009;31(1):89–98.

Leentjens AFG. The role of dopamine agonists in the treatment of depression in patients with Parkinson's disease: a systematic review. *Drugs*. 2011;71(3):273–286.

National Center for Biotechnology Information. PubChem Compound Database. http://pubchem.ncbi.nlm.nih.gov

Rakofsky JJ, Holtzheimer PE, Nemeroff CB. Emerging targets for antidepressant therapies. *Curr Opin Chem Biol*. 2009;13(3):291–302.

Rektorová I, Rektor I, Bares M, et al. Pramipexole and pergolide on the treatment of depression in Parkinson's disease: a national multicentre prospective randomized study. *Eur J Neurol*. 2003;10(4):399–406.

Sahakarian BJ, Bruhl AB, Cook J, et al. The impact of neuroscience on society: cognitive enhancement in neuropsychiatric disorders and in healthy people. *Philos Trans R Soc Lond B Biol Sci*. 2015;370(1677):20140214.

Svensson KA, Hao J, Bruns RF. Positive allosteric modulators of the dopamine D1 receptor: a new mechanism for the treatment of neuropsychiatric disorders. *Adv Pharmacol*. 2019;86:273–305.

Tundo A, de Filippis R, De Crescenzo F. Pramipexole in the treatment of unipolar and bipolar depression. A systematic review and meta-analysis. *Acta Psychiatr Scand*. 2019;140(2):116–125.

U.S. National Library of Medicine. Dailymed. Accessed May 28, 2015 http://dailymed.nlm.nih.gov/dailymed/index.cfm

Winkelman JW, Armstrong MJ, Allen RP, et al. Practice guideline summary: treatment of restless legs syndrome in adults: report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology*. 2016;87:2585–2593.

Zarate CA Jr, Payne JL, Singh J, et al. Pramipexole for bipolar II depression: a placebo-controlled proof of concept study. *Biol Psychiatry*. 2004;56(1):54–60.

33.4b Amphetamines, Methylphenidate, Atomoxetine, and Modafinil

MANPREET KAUR SINGH, M.D., M.S.

Sympathomimetic amines are compounds that activate the sympathetic central nervous system (CNS). Endogenous catecholamines, such as norepinephrine (NE) and dopamine (DA), are direct sympathomimetic amines and exert their effects by binding to DA and NE receptors.

Psychostimulants are indirect sympathomimetic amines with both central and peripheral effects. Psychostimulants enhance motivation, attention, concentration, wakefulness, and stimulate the cardiovascular system. They also suppress appetite and interfere with sleep. Psychostimulants achieve these effects by inhibiting the release or slowing the uptake of catecholamines, which leads to a buildup of DA and NE in the extracellular space.

Amphetamine (AMPH) was the first formulation of stimulants to be synthesized in 1927. Stimulants have been historically used for the treatment of “hyperkinetic” behavior in children, parkinsonism, epilepsy, and for some time, barbiturate poisoning. The second formulation of stimulants, methylphenidate (MPH), was synthesized in 1944 but was not identified as a stimulant until 1954. It was first used to treat barbiturate-induced coma and memory deficits in the older adults. Beginning in the 1960s, MPH was used to treat children with attention-deficit hyperactivity disorder (ADHD), known at the time as “hyperactivity or minimal brain dysfunction (MBD).” Together, the AMPH- and MPH-based formulations constitute the psychostimulant medication class, which is currently approved by the United States Food and Drug Administration (U.S. FDA) for the treatment of ADHD and narcolepsy for all age groups. Despite their clear therapeutic benefit, psychostimulants may be highly addictive and are classified as schedule II medications by the Drug Enforcement Agency (DEA). Although the immediate-release AMPH and immediate-release MPH (MPH-IR) are prototypic psychostimulants, intermediate and long-acting formulations are more frequently prescribed due to the convenience of increased adherence and improved pharmacologic coverage. See Table 33.4b–1 for a complete list of available psychostimulants and the nonstimulants atomoxetine and modafinil, as well as their strengths and dosing recommendations for ADHD. α_2-Agonists (guanfacine, clonidine), which have also been classified as nonstimulants, are covered in a separate chapter in this section.

ETIOLOGY AND PATHOGENESIS OF ATTENTION-DEFICIT HYPERACTIVITY DISORDER

Research in the pathogenesis and etiology of ADHD is evolving and may inform the pharmacologic properties of psychostimulants. Indeed, ADHD is known to be a heritable, chronic, but variable in its course neurobehavioral disorder that is frequently characterized by inattention, impulsivity, and hyperactivity. Clinical characteristics of ADHD correspond with aberrant catecholamine neurotransmission in the prefrontal cortex (PFC) and functionally associated cortical and subcortical areas. To sustain attention, regulate impulse control, and facilitate working memory and cognitive flexibility, the PFC exerts a “top-down” control of the temporal cortex, parietal cortex, basal ganglia, amygdala, cerebellum, and corpus callosum. Dysfunction of catecholamine neurotransmission in the PFC compromises these cognitive functions and may lead to some clinical manifestations of ADHD.

Table 33.4b–1.
Clinical Indications of Psychostimulants and Key Nonstimulants

Medication	Dosing	Supplied	Average Therapeutic Dose	Maximum Daily Dose	Key U.S. FDA Indication(s)
Generic names: Mixed amphetamine and dextroamphetamine salts **Brand name:** Adderall **Formulation:** Immediate-release tablet	**Adults** *ADHD:* Initial: 5 mg once or twice daily Increase daily dose by 5 mg weekly intervals ***Narcolepsy:*** Initial: 10 mg daily Increase daily dose in 10 mg weekly intervals **Pediatric** Not FDA approved for children ≤3 years of age *ADHD:* **3–5 years** Initial 2.5 mg once every morning Increase daily dose in 2.5 mg weekly intervals **6–12 years** Initial: 5 mg once or twice daily Increase daily dose in 5 mg weekly intervals **Adolescents 13–17 years** Initial: 5 mg once or twice daily Increase daily dose in 5 mg weekly intervals ***Narcolepsy:*** **6–12 years:** Initial: 5 mg daily Increase daily dose in 5 mg weekly intervals **Children >12 years: Refer to adult dosing.** **Children: <3 years not recommended**	5, 7.5, 10, 20, or 40 mg	**Narcolepsy:** 5–60 mg/day in 1–3 divided doses 2.5–20 mg b.i.d., daily	**Adults** *ADHD* 40 mg *Narcolepsy:* 60 mg **Pediatric** **3–17 years** *ADHD* 40 mg *Narcolepsy* 60 mg **Not FDA approved for children ≤3 years of age**	Monotherapy for ADHD and narcolepsy
Generic name: Dextroamphetamine sulfate **Brand names:** Dexedrine, ProCentra, Zenzedi **Formulation: immediate-release tablet or extended-release capsule**	**Immediate release** **Adults** **Narcolepsy:** Initial: 10 mg once daily Increase in increments of 10 mg at weekly intervals until optimal response is obtained **Pediatric** ***ADHD:*** **Children 3–5 years:** **Initial:** 2.5 once daily **Increase** daily 2.5 mg dose at weekly intervals *(Current AAP guidelines do not recommend use in children ≤5 years due to insufficient evidence) ***Children ≥6 years and adolescents:*** Initial: 5 mg once or twice daily; may increase in increments of 5 mg at weekly intervals **Narcolepsy** Children 6–12 years: Initial: 5 mg once daily Increase in increments of 5 mg at weekly intervals **Children >12 years and adolescents** Refer to adult dosing. **Extended-release capsules:** Initial: 5 mg once or twice daily Increase in increments of 5 mg weekly	5-, 10-, 15-mg capsules	**Adults, children 12 and older:** Narcolepsy: 5–60 mg in divided doses	**Immediate release** **Adults** **ADHD:** 40 mg **Narcolepsy:** 60 mg **Pediatric** **ADHD:** 40 mg **Narcolepsy:** 60 mg **Extended release** **Pediatric** **ADHD:** 40 mg (60 mg in divided doses can be tried in children >50 kg)	Monotherapy ADHD and narcolepsy
Generic names: Mixed amphetamine and dextroamphetamine salts **Brand name:** Adderall XR **Formulation:** Extended-release capsule	**Adults-** *ADHD:* Initial: 20 mg once daily in the morning Not FDA approved for children ≤6 years of age **Pediatric** *ADHD:* Initial: 5–10 mg once daily in the morning Increase daily dose in 5–10 mg increments at weekly intervals	5-, 10-, 15-, 20-, 25- or 30-mg capsule	20–30 mg	30 mg **Children: <6 years not recommended**	Monotherapy ADHD

Generic name: Lisdexamfetamine **Brand name:** Vyvanse **Formulation:** Capsule with inactive pro-drug compound	**Adults** **ADHD:** Initial: 30 mg once daily in the morning Increase in increments of 10 or 20 mg daily at weekly intervals **Binge eating disorder** (BED): **Pediatric** *ADHD:* Children 6–12 years: Initial: 20–30 mg once daily in the morning Increasing: weekly increases of 10 mg daily up to a maximum of 60 mg daily Adolescents: Refer to adult dosing	20-, 30-, 40-, 50-, 70-mg capsule	**ADHD:** 20–70 mg daily **Binge eating disorder:** 50–70 mg daily	**Adults** *ADHD* and *BED* 70 mg **Pediatric** *ADHD* 60 mg **Not FDA approved for children ≤6 years of age**	Monotherapy ADHD and BED
Generic name: Methylphenidate hydrochloride **Brand name:** Ritalin **Formulation:** Immediate-release tablet	**Adults** **ADHD:** Initial: 5 mg twice daily Increase by 5–10 mg daily at weekly intervals Narcolepsy: Initial: 5 mg twice daily Increase by 5–10 mg daily at weekly intervals Pediatric Not FDA approved for children ≤6 years of age Children ≥6 years and adolescents: Initial: 5 mg twice daily Increase by 5–10 mg daily at weekly intervals **Off-label use:** Depression in medically ill older adults, adult patients with terminal illness, receiving palliative care: 2.5–5 mg once daily or twice daily Increase by 2.5–5 mg daily every 1–3 days in divided doses	5, 10, 20 mg	20–30 mg Some patients might require 40–60 mg	**Adults** **ADHD** and **narcolepsy** 60 mg daily (in 2–3 divided doses) **Depression in medically ill older adults, adult patients with terminal illness, receiving palliative care:** 20–40 mg daily **Pediatric** ***ADHD:*** 60 mg daily (in 2–3 divided doses)	Monotherapy ADHD and narcolepsy
Generic name: Methylphenidate hydrochloride **Brand name:** Methylin **Formulation:** Immediate-release oral solution	Adults Children (6 years and over) Initial: 5 mg twice daily Increase by 5 mg daily at weekly intervals	Methylin oral solution 5 mg/5 mL Methylin Oral Solution 10 mg/ 5 mL	20–30 mg daily	60 mg Not FDA approved for children ≤6 years of age	Monotherapy ADHD and narcolepsy
Generic name: Methylphenidate hydrochloride **Brand name:** Ritalin SR **Formulation:** Immediate-release tablet or extended-release capsule	**Adults and pediatric** ***ADHD*** and ***narcolepsy:*** Equivalent to Ritalin immediate release	20 mg	20–60 mg	60 mg	Monotherapy ADHD and narcolepsy
Generic name: Methylphenidate hydrochloride **Brand name:** Concerta **Formulation:** Extended-release tablet	**Children 6–12 years, adolescents 12–17 years, adults** ***ADHD:*** Initial: 18 mg daily (18–36 mg adults) Increased by 18 mg/day at weekly intervals	18, 27, 45, 54 mg	**Adults** 18–72 mg daily **Pediatric** 18–54 mg daily	54 mg (6–12 years) 72 mg (adolescents and adults) **Not FDA approved for children ≤6 years of age or older adults**	Monotherapy ADHD
Generic name: Methylphenidate hydrochloride **Brand name:** Daytrana **Formulation:** Extended-release transdermal system	**6–12 years, adolescents-** ***ADHD:*** **week 1:** 10 mg/9 hours (1.1 mg/hour) **week 2:** (1.6 mg/hour), 15 mg/9 hours **week 3:** (2.2 mg/hour), 20 mg/9 hours **week 4–7:**30 mg/9 hours (3.3 mg/hour)	Transdermal patch: 10 mg/9 hours (1.1 mg/ hour) 15 mg/9 hours (1.6 mg/hour) 20 mg/9 hours (2.2 mg/hour) 30 mg/9 hours (3.3 mg/hour)	1.1–3.3 mg/hour	3.3 mg/hour **Not FDA approved for children ≤6 years of age or older adults** **No indication in adults**	Monotherapy ADHD

Generic name: Methylphenidate hydrochloride **Brand name:** Metadate CD **Formulation:** Expended-release capsule	Children 6–12 years, adolescents 12–17 years, adults- ADHD: Initial: 20 mg daily; Increase in 10–20 mg increments at weekly intervals	10, 20, 30, 50 mg	60 mg	**Children >6, adolescents, adults** **ADHD: 60 mg**	Monotherapy ADHD and narcolepsy
Generic name: Methylphenidate hydrochloride **Brand name:** Metadate ER **Formulation:** Extended-release tablets	**Adults and Pediatric** ***ADHD and Narcolepsy:*** Equivalent to Ritalin immediate release	20 mg	10–60 mg	60 mg	Monotherapy ADHD and narcolepsy
Generic name: Methylphenidate hydrochloride **Brand name:** Quillivant XR **Formulation:** Extended-release powder for suspension	Not approved by FDA for children <6 years of age Children 6–12 years, adolescents, and adults: ADHD: Initial: 20 mg daily; Increase in 10–20 mg increments at weekly intervals	Bottles of 300-, 600-, 750-, 900-mg powder (5 mg/mL)	60 mg	60	Monotherapy ADHD and narcolepsy
Generic name: Dexmethylphenidate **Brand names:** Focalin and Focalin XR **Formulation:** Immediate-release tablet and extended-release capsule	Adults: ADHD: Immediate release: Initial: 2.5 mg twice daily Increase in increments of 2.5–5 mg at weekly intervals Extended-release: Initial: 10 mg once daily. Increase in increments of 10 mg/day at weekly intervals (maximum dose: 40 mg/day) Pediatric ADHD: Children ≥6 years: Initial: 2.5 mg twice daily Increased in increments of 2.5–5 mg at weekly intervals Initial: 5 mg once daily; dosage may be adjusted in increments of 5 mg/day at weekly intervals	Focalin: 2.5, 5, 10 mg Focalin XR: 5, 10, 15, 20, 30, 35, 40 mg	Focalin: 2.5–10 mg twice a day (doses should be taken at least 4 hours apart) Focalin XR 20–40 mg daily (adults) and 10–30 mg daily (pediatric)	**Focalin immediate release** 20 mg/day **Focalin extended release** 30 (pediatric) and 40 (adults)	Monotherapy ADHD
Generic name: Atomoxetine HCl **Brand name:** Strattera **Formulation:** Immediate-release capsules	Adults ADHD Initial: 40 mg/day Increase to ~80 mg/day after minimum of 3 days. May increase to 100 mg/day in 2–4 additional weeks to achieve optimal response. Pediatric ADHD: *Children ≥6 years and ≤70 kg* Initial 0.5 mg/kg/day Increase to ~1.2 mg/kg/day after a minimum of 3 days	10, 18, 25, 40, 60, 80, 100 mg	Children and adolescents ≥70 kg: 100 mg Children and adolescents ≤70 kg: 1.2 mg/kg	1.4 mg/kg or 100 mg	Monotherapy ADHD
Generic name: Modafinil **Brand name:** Provigil **Formulation:** Immediate-release tablet	**Narcolepsy, obstructive sleep apnea/hypopnea syndrome (OSAHS):** Initial: 200 mg daily in 2 divided doses Increase: in 100-mg increments as needed and tolerated **Shift work sleep disorder (SWSD):** Initial: 200 mg as a single dose taken ~1 hour prior to start of work shift	Supplied: 100-, 200-mg tabs	Narcolepsy 200–300 mg/day **SWSD: 200** mg/day	400 mg	

ADHD, attention-deficit hyperactivity disorder; U.S. FDA, United States Food and Drug Administration.

AMPH and MPH change the dynamics of the catecholamine transport system, which allows for an increase of NE and DA levels in the PFC area and improves cognitive deficits associated with ADHD. Using animal models, and microdialysis techniques, Volkow and colleagues (2012) demonstrated that psychostimulants work by correcting the catecholamine dysfunction in the PFC and a related “attention” network. They have shown that MPH can increase DA levels in the ventral striatum, the PFC, and the temporal cortex, dysfunction which is implicated in ADHD. Although this small study in 20 adults study merits replication, it advances the mechanistic understanding of how stimulants work in humans. Further, MPH increases catecholamine levels in the fronto-striatal-thalamo-cerebellar circuit implicated in executive functions (EFs) and is positively associated with improvement in working memory, reaction time, and focus.

Neurobiology of ADHD

Neuroimaging studies have demonstrated aberrant and intercorrelated brain structural and functional network characteristics associated with key cognitive deficits in ADHD. Compared to healthy controls, individuals with ADHD across the lifespan have decreased volumes in frontal, parietal, and temporal cortices, as well as in the basal ganglia, limbic system (amygdala, hippocampus, nucleus accumbens), hypothalamus, cerebellum, and corpus callosum. Widespread but subtle differences in cortical surface area have been found in children but not adolescents and adults with ADHD, confirming prefrontal cortical involvement. Critically, impairments in these frontal-striatal-cerebellar networks that are known to regulate attention, impulse control, and other key EFs may lead to deficits in these cognitive functions in ADHD. Further, alterations present as endophenotypes in families and are linked to ADHD symptoms in the population, extending evidence that ADHD behaves as a continuous trait. Importantly, more granular findings have been demonstrated in key functional subregions of the PFC, including the orbitofrontal cortex (OFC; control and processing of emotion and motivation), the inferior frontal cortex (IFC; regulates inhibition of impulsive thoughts or feelings, and maintains cognitive flexibility essential for decision making), dorsolateral prefrontal cortex (DLPFC; key for sustained attention, planning, and working memory), dorsal anterior cingulate cortex (dACC; integrates information and controls executive and inhibitory functions; dACC facilitates sensory target detection, proper response selection, error detection, and action monitoring), and the anterior PFC (controls impulsivity of the motor behavior). Collectively, these regions function to filter irrelevant information or “noise” and amplify relevant signal needed to execute a specific task. Imaging studies have consistently demonstrated delays in brain maturation associated with ADHD but have inconsistently demonstrated conclusive beneficial effects of medication on brain volume.

Biochemistry of Catecholamine Neurotransmission

ADHD-related dysfunction may be related to disruption of PFC fibers that project catecholamines to cortical and subcortical regions. This model of ADHD has also been supported by animal and human lesion studies, as well as multimodal neuroimaging (e.g., diffusion tensor imaging and functional magnetic resonance imaging) studies that show that disruption of key PFC projections results in attention, emotion, and locomotor dysregulation. Further, mutating genes that encode for the dopamine transporter protein (DAT) and enzymes that metabolize DA and NE, have conferred behavioral traits in animals that mimic ADHD-like traits. Indeed, both DA and NE play an important role in the pathophysiology of ADHD and are candidate targets for treatment with psychostimulants. DA enhances working memory through its interaction with D_1 receptors. NE functions to increase arousal, excitability, and memory storage and retrieval, and to modulate EF, including cognitive organization and problem solving through its interactions of the postsynaptic α_{2A}-adrenoceptors. Thus, deficits in this catecholamine system within the PFC have been associated with decreased control over irrelevant information, leading to inattention, impulsivity, and hyperactivity associated with ADHD. As above, psychostimulants are thought to inhibit or reverse the direction of the NE/DA membrane transporter proteins (NET/DAT), which increase catecholamines in the synaptic space, leading to enhanced function in the PFC to restore attention.

Genetics

Family, adoption, and twin studies suggest that ADHD runs in families with a mean heritability estimate of ADHD is ~76% in children and adolescents, and ~30% in adults. Studies on genetic linkage show that the effects of DNA risk variants on ADHD at an individual level are very small. Several genetic loci in genome-wide association studies (GWAS) reach genome-wide level of statistical significance, suggesting that approximately a third of the heritability in ADHD is due to a polygenic component comprising many common variants that each have small effects. Studies of copy number variants also suggest that rare insertions or deletions may, in part, account for heritability in ADHD. Gene variants associated with the ADHD phenotype include mutations of genes that encode for the proteins involved in the catecholamine neurotransmission system. Unfortunately, the strength of the association between ADHD and likelihood of carrying mutations of most candidate genes that have been identified so far has not been adequately validated.

MECHANISMS OF ACTION OF PSYCHOSTIMULANTS

Neuroscience-Based Nomenclature (NbN)

Amphetamine (D) and (D.L), lisdexamfetamine, methylphenidate (D) and (D.L), dexmethylphenidate

Pharmacologic domain	Dopamine, norepinephrine
Mechanism of action	Reuptake inhibitor (DAT, NET) and releaser (DA, NE)

As above, psychostimulants exert clinical effects by raising extracellular catecholamine levels in the context of catecholamine deficit. The chemical compositions of psychostimulants AMPH and MPH are different. AMPH belongs to the phenethylamine chemical class. The asymmetric α-carbon atom gives rise to two AMPH enantiomers, the L-stereoisomer and the D-stereoisomer. MPH has four isomers but only L-MPH and D-MPH are part of the currently available MPH preparations, and only D threomethylphenidate exhibits clinically useful pharmacologic activity.

Just as chemical compositions vary, the mechanisms of action of various psychostimulants also differ significantly. Whereas MPH blocks the DAT and NET and increases DA and NE from the synaptic vesicle pool by inhibiting endocytosis, AMPH compounds use DAT and NET, along with several other transporters, to actively unload vesicular catecholamine stores. Using positron emission tomography (PET) and radioligands specific to the DAT, NAT, DA, and NE receptors, D-methylphenidate has been shown to rapidly cross the blood–brain barrier (BBB), and bind to the DAT and NET. To reach efficacy threshold, MPH needs to occupy approximately 70% to 80% of NET and 60% to 70% of DAT, in order to increase synaptic DA in the PFC leading to increased attention. The rate of extracellular DA rise is proportional to the rate at which MPH crosses the BBB.

In contrast to these MPH mechanisms, AMPH functions as a potent full agonist of trace amine-associated receptor 1 (TAAR1), which inhibits vesicular monoamine transporter function (VMAT), DAT (especially D_1, the predominant receptor type in the PFC), and NET. AMPH also functions by reducing the DA and NE vesicular pH gradient, leading to reversal of catecholamine transport and inhibition of the monoamine oxidase (MAO) and catechol-o-methyltransferase (COMT) enzymes that degrade catecholamines. This confers more potency in AMPH compared to MPH, which translates to more rapid extracellular increases in DA and NE concentrations compared to MPH. Indeed, microdialysis studies in rat brains have demonstrated that AMPH induces three times greater behavioral effects and greater synaptic DA and NE boost compared to the catecholamine rise caused by much higher doses of MPH. Thus, AMPH may be a more effective indirect sympathomimetic agent than MPH and may require lower effective doses to induce MPH-equivalent physiologic effects.

CLINICAL PHARMACOLOGY OF PSYCHOSTIMULANTS

Psychostimulants, AMPH, and MPH compounds are the most commonly used psychotropic medications nationwide and have U.S. FDA–indicated treatment for ADHD, narcolepsy, and in some agents, exogenous obesity in children, adolescents, and adults. Please refer to Table 33.4b–1 for a list of clinical indications of psychostimulants. This section will review the pharmacokinetics (see Table 33.4b–2), pharmacodynamics, interactions, efficacy, and tolerability (including adverse effects, contraindications, and precautions) of psychostimulants and then conclude with an overview of these characteristics for nonstimulant medications, modafinil and atomoxetine (ATM).

Pharmacokinetics

Immediate-Release Formulations. Due to their high degree of lipophilicity, psychostimulants rapidly cross the gastrointestinal (GI) tract and the BBB. The absorption phase parallels the rate of release of the catecholamine neurotransmitters into the CNS synaptic clefts.

AMPH-related compounds typically reach peak blood levels in 2 to 3 hours, with a variable effect of several hours after oral administration. AMPH is metabolized by deamination, oxidation, and hydroxylation.

Deamination produces an inactive metabolite, phenylacetone, which is further oxidized and then excreted in urine as hippuric acid and glucuronide conjugates. AMPH is also converted to norephedrine by oxidation and then this metabolite and the parent compound are p-hydroxylated. Several metabolites, including norephedrine, its hydroxymetabolite, and hydroxyamphetamine, are all pharmacologically active. The excretion of AMPH depends on urinary pH, which provides a basis for using ammonium chloride to acidify the urine to manage an AMPH overdose, which markedly increases excretion. Approximately 70% of the dose is excreted in a 24-hour urine collection with 30% excreted as unchanged drug. This increases to 74% under acidic conditions and decreases to 1% in alkaline urine.

The MPH tablet releases equal amounts of D-MPH and L-MPH, at a nearly identical, linear rate. The behavioral effects of MPH are exerted by its d-MPH enantiomer whereas L-MPH has little or no behavioral effect. MPH and related compounds reach peak blood concentration in 2 hours and have a short half-life of 1 to 2 hours. Concentrations of MPH in the brain exceed concentrations in the blood. Metabolism of MPH is highly enantio-specific, favoring L-MPH over D-MPH. Plasma half-life (t1/2) of D-MPH is markedly longer than that of L-MPH, and oral bioavailability is 22% for D-MPH and 2% to 8% for L-MPH. MPH undergoes hydrolysis and de-esterification by a liver and stomach enzyme, CES-1. Another route of metabolism is hepatic de-esterification and oxidation. Ritalinic acid is the major metabolite of MPH that is renally excreted.

Long-Acting Formulations. MPH and AMPH long-acting products extend the apparent plasma half-life of MPH and AMPH by altering the absorption kinetics. MPH sustained release (Ritalin SR) was the first available long-acting formulation of MPH. It delivers a single, prolonged pulse of MPH via a wax-matrix–based technology. This formulation has a problematic delay in release rate, onset of behavioral effects, and significant intrasubject variability and has been shown to have inferior clinical properties compared to MPH administered several times per day. Due to the availability of new multiple long-lasting release formulations of MPH, Ritalin-SR is rarely used. Controlled-release formulations of MPH include Metadate CD, Focalin XR, and Ritalin LA. Metadate CD is the first extended-release formulation to use beaded technology. The MPH compound is stored in porous beads, with pore diameter controlling its release. The porous beads produce a bimodal MPH distribution, with an immediate burst of a D- and L-methylphenidate mixture, and an extended release of the remaining MPH approximately 5 hours later. The Focalin XR formulation also demonstrates similar bimodal kinetics. It provides 8-hour coverage without significant troughs. Focalin XR releases D-threo-methylphenidate, which is approximately twice as potent as L-threo-methylphenidate.

Mixed salts of a single-entity AMPH product (Adderall XR) contain two types of beads in a 50:50 ratio. The immediate-release beads create an almost immediate burst of AMPH, and the delayed-release beads provide the second burst approximately 4 hours after oral ingestion. Extended-release methylphenidate, Ritalin LA, uses a spheroidal oral drug absorption system (SODAS) technology to evenly distribute MPH over approximately 8 hours. It has a bimodal distribution of a mixture of D- and L-threo-methylphenidate. Extended-release MPH (Concerta) uses an oral osmotic release system (OROS), which relies on an osmotic gradient to provide extended coverage. The capsule is composed of a medication layer next to a layer of inactive matrix material. A third layer contains active osmotic components, which regulate the rate of release. In an aqueous environment such as GI tract, the matrix layer dissolves within 1 hour, providing an initial dose of MPH. Water expands the osmotic layer, and then the remaining amount of MPH is released, controlled by the diameter of the osmotic layer orifices.

The MPH transdermal system (Daytrana) delivers MPH through a transdermal matrix. It does not undergo first-pass metabolism by CES-1 in the liver and stomach, and has a relatively constant release of both D- and L-methylphenidate throughout the entire patch over time, resulting in constant plasma MPH levels. Daytrana may be useful for patients who have problems tolerating oral medications. In addition, extended-delivery MPH (MEROS or Quillivant XR) is a powder formulation of MPH that forms an extended-release oral suspension for once-per-day oral administration. It contains approximately 20% immediate-release and 80% extended-release MPH. This formulation might be advantageous for patients who cannot tolerate pill and prefer a liquid form. Finally, lisdexamfetamine dimesylate (LDX; Vyvanse) is the first known inactive prodrug psychostimulant that is activated by a chemical modification in erythrocytes. LDX has high aqueous solubility, partially due to low lipophilicity. It is likely that in the small intestines, LDX is transported by peptide transport proteins such as PEPT1 and PEPT2. Unlike other AMPHs, it consistently increases the extracellular concentration of serotonin in the PFC and striatum.

Table 33.4b–2. Pharmacologic Properties of Psychostimulants and Key Nonstimulants

Medication	T_{max} (h)	Half-Life (Hours)	Duration of Action (Hours)	C_{max} (ng/mL)
Mixed (R,S)-amphetamine salts, short acting (Adderall) Racemic mixture of D:L-amphetamine isomers in a 3:1 ratio Each Adderall XR capsule contains a 1:1 ratio of immediate- to delayed-release mixed amphetamine salts beads.	2	For D-Isomer ~11–13 L-Isomer ~15–17 (if urine is acidic, both isomers' half-lives are reduced to 7 hours)	6	110
Mixed (R,S)-amphetamine salts, extended-release (Adderall XR) Racemic mixture: D:L amphetamine isomers in a 3:1 ratio Each Adderall XR capsule contains a 1:1 ratio of immediate- to delayed-release mixed amphetamine salts beads.	Children 6–12 years: D-amphetamine: 9 hours L-amphetamine: 11 hours Adolescents 13–17 years: D-amphetamine: 11 hours L-amphetamine: 13–14 hours Adults: D-amphetamine: 10 hours L-amphetamine: 13 hours An acidic diet: 8–11 hours An alkaline diet 16–31 hours	Mean 11.5	10–12	40.10
Dextroamphetamine sulfate extended-release (Dexedrine)	First peak: 1.6 hours Second peak: 6.5 hours	9.4–13	10	36.6
Dextroamphetamine sulfate **immediate release**	1–2	9.4–13	10	36.6
Lisdexamfetamine dimesylate (Vyvanse)	3.8–4.7	1	10–12	58.2 ± 28.1 ng/mL
Methylphenidate hydrochloride (Ritalin)	2	2–3	4	3.5 ± 0.4 ng/mL
Methylphenidate sustained release (Ritalin SR)	3–8	Unknown	~8	4.84
Methylphenidate extended-release oral suspension (MEROS) Long-acting liquid preparation with MEROS delivery system. MEROS consists of drug–polymer complexes 80% of drug–polymer complexes are coated with an extended-release coating of varying thicknesses, which allows for release of methylphenidate throughout the day 20% of drug–polymer complexes are uncoated and act as immediate-release methylphenidate	First peak ~2 Second peak 5	5–6	10–12	3.6 (± 5.8) ng/mL
Methylphenidate hydrochloride oral solution (Methylene)	2	2–3	4	3.5 ± 0.4 ng/mL
Extended-release methylphenidate hydrochloride (Metadate CD) Contains two types of coated beads, one type with immediate-release *methylphenidate* and second with extended-release *methylphenidate*. 30% of beads deliver d-threo- and l-threo-methylphenidate immediately The remaining 70% of beads are extended release, delivering the mixture of L-threo-methylphenidate content and D-threo-methylphenidate content several hours after the beads with methylphenidate have been released.	First peak ~2 Second peak 4–6	6.8	8–9	24.7–40.9 ng/mL
D-Threo-Methylphenidate (Focalin)	2	2–3	4	3.5 ± 0.4 ng/mL
Dextra-methylphenidate extended-release (Focalin XR) Formulation provides coverage of up to 12 hours. Focalin XR contains only the *d*-MPH which is twice as potent as Focalin immediate release. It delivers 50% of its d-threo-MPH initially, and 3–4 hours later. This creates an approximate 8–10 coverage without significant troughs. Thus, the closest concentration curve can be recreated by methylphenidate twice daily (50% of the dose).	First peak 1–2 Second peak 5–6	3.5	8–10	14–16

Methylphenidate transdermal system (Daytrana) Delivers methylphenidate into the systemic circulation directly through the skin, reducing first-pass metabolism by the liver.	10 hours after single application 8 hours after repeat patch applications	3–4	9	5.7 ± 9.39:12.5 CM^2– 542.9 ± 22.4:37.5
Extended-release methylphenidate (Ritalin LA) Delivers MPH via spheroidal oral drug absorption system (SODAS) technology to produce a bimodal distribution of 1:1 racemic mixture of d-and l-threo-MPH enantiomers. 50% is released initially, and another 50% approximately 3–4 hours later	First peak 1–2 Second peak 5–6	3.5	6–8	16.4
Extended-release methylphenidate (Concerta) It delivers methylphenidate via oral osmotic release system (OROS), which is composed of two drug layers containing the drug and excipients, and a push layer with active osmotic components. It delivers 1:1 racemic mixture of l:d-threo-methylphenidate. In the gastrointestinal tract, the drug overcoat dissolves within 1 hour, providing an initial dose or 22% of the medication. Water then permeates the tablet core, osmotically active polymer, releasing the remaining 78% of the mediation through the orifice	8–12	3.5	9–11	3.3–7.8
Extended-delivery methylphenidate (MEROS or Quillivant XR) is a powder that, after reconstitution with water, forms an extended-release oral suspension formulation of methylphenidate intended for once-daily oral administration. It contains approximately 20% immediate-release and 80% extended-release methylphenidate	5.0	5.2		13.6 (±5.8) ng/mL
Modafinil (Provigil)	2–3	D-modafinil: 4–5 hours L-modafinil: 13–16 hours L-modafinil constitutes 90% of C_{max}	6–14	6.4 μg/mL
Atomoxetine (Strattera)	1–2	5.2 (extensive metabolizers) 21.6 (poor metabolizers)	24	584 ± 268
C_{max}, maximum plasma methylphenidate concentration T_{max}, time to maximum plasma methylphenidate concentration				

MPH, methylphenidate.

Pharmacodynamics

Traditional stimulants are all considered indirect-acting sympathomimetics because they do not act directly on receptors, as do natural catecholamines. Stimulant action occurs through simultaneous neurotransmitter release at the synapse and block of reuptake, thereby slowing metabolic breakdown of active neurotransmitters. Dextroamphetamine is the most potent by virtue of its ability to release both DA and NE and its capacity to release newly formed soluble pools and granule-stored pools of catecholamines in nerve endings. MPH acts on DA (but not NE) through release of DA from granular pools and blockage of DA reuptake. This may explain why AMPH is a more powerful stimulant and is more likely to activate latent hypertension than is MPH. See Table 33.4b–3 for important psychostimulant drug–drug interactions.

Efficacy and Tolerability of Stimulants in the Treatment of ADHD and Other Conditions

Several randomized controlled trials (RCTs) have characterized the efficacy, tolerability, and adverse effects of stimulants for ADHD and for other psychiatric disorders. In total, approximately 5,899 children, adolescents, and adults have been involved in RCTs examining the efficacy and safety of stimulants in treatment of ADHD. Outcomes of these studies pooled together have demonstrated that 65% to 75% of individuals in the treatment arm improved in core ADHD symptoms (inattentiveness, hyperactivity, impulsivity), compared to 5% to 30% of individuals who improved when assigned to the placebo arm, with large effect sizes (Cohen's d = 0.7 to 1.5). Indeed, the efficacy of ADHD medication goes beyond symptom control and may help mitigate functional impairments and health-related quality-of-life deficits in patients with ADHD. Thus, psychostimulants are considered highly effective and relatively safe for the treatment of ADHD in all age groups. Based on such robust evidence, the American Academy of Pediatrics and the American Academy of Child and Adolescent Psychiatry (AACAP) have adopted guidelines that regard stimulant medications as a cornerstone treatment for ADHD.

Table 33.4b–3.
Important Psychostimulant Drug–Drug Interactions

Methylphenidate	**Inhibits Metabolism of:**
	Anticoagulants (e.g., Coumadin)
	Tricyclic drugs (e.g., imipramine, clomipramine, desipramine)
	Anticonvulsants (e.g., phenobarbital, phenytoin, primidone)
	CYP2D6 substrates (e.g., flecainide, haloperidol, metoprolol, propafenone, risperidone, thioridazine SSRIs, SNRIs)
	Cyclosporine
	Glipizide
	Induces metabolism of:
	Methylphenidate
	Metabolism is inhibited by:
	Bupropion
	CYP2D6 inhibitors (e.g., chlorpromazine, delaviridine, fluoxetine, miconazole, paroxetine, pergolide, quinidine, quinine, ritonavir, and ropinirole)
	Metabolism is induced by: Carbamazepine
Amphetamines	**Acetazolamide**
	Acetazolamide decreases the clearance of amphetamine
	Chlorpromazine
	Decreased amphetamine effectiveness
	Donepezil
	Increased norepinephrine availability lowers the seizure threshold
	Guanethidine
	Decreases guanethidine effectiveness
	Opipramol, **Pargyline**, **Phenelzine**,

Procarbazine, **Protriptyline**, **Tranylcypromine, Doxepin, Trimipramine, Toloxatone, Rasagiline**, **Selegiline**, **Sibutramine, Amitriptyline:** Increased norepinephrine availability can result in a hypertensive crisis (headache, hyperpyrexia, hypertension)

The concurrent use of amphetamine and monoamine oxidase inhibitors (MAOIs) is contraindicated. At least 14 days should elapse following the discontinuation of MAOIs before amphetamine therapy is instituted

Desvenlafaxine, Dolasetron, Lofepramine, Lorcaserin, Hydroxytryptophan, Imipramine, Iproniazid, Almotriptan, Clorgyline, Desipramine, Isocarboxazid, Meperidine, Moclobemide, Nialamide, Nortriptyline Tramadol, Trazodone, Vilazodone, Vortioxetine, Ziprasidone, Fentanyl, Fluoxetine, Furazolidone, Mirtazapine, Granisetron, Levomilnacipran, Palonosetron: Increased risk of serotonin syndrome from additive serotonergic effects

---

Data in Children and Adolescents with ADHD. The Collaborative Multisite Multimodal Treatment Study of Children with ADHD (MTA) is the first major RCT that demonstrated long-term efficacy and safety of MPH in the treatment of children diagnosed with ADHD, combined type. MTA investigators followed 579 children, 7 to 9.9 years of age. Participants were randomized to three carefully designed treatment groups and a standard-of-care (control) group that received treatment in the community. Treatment groups included one group of participants who were treated with MPH-IR monotherapy, a second group treated with MPH-IR in combination with behavioral and community interventions, and a third group treated with behavioral and community interventions only. Participants were assessed using parent- and teacher-reported ADHD rating scales at baseline and after 14 months of treatment. Adverse effects were also assessed and reported. MPH-IR was administered 3 times/day, with a mean total daily dose was 32.1 ± 15.4 mg. Subjects treated with MPH-IR with or without behavioral or community interventions showed significant improvement in core ADHD symptoms such as inattention, impulsivity, and hyperactivity. In addition, children randomized to the two groups with MPH treatment demonstrated significant improvement in family, social, and academic functioning, as well as in externalizing symptoms. The symptomatic and functional improvements persisted for the duration of the treatment.

A 93% retention rate was reported 24 months after the initiation of the MTA trial when ADHD symptoms and level of social, academic, and family functioning were reassessed. Most subjects continued to benefit from MPH or from the combination of MPH and behavioral interventions. The difference in ADHD symptoms and functioning among subjects treated with MPH, subjects receiving behavioral intervention alone, or subjects treated in a community setting decreased significantly over time. Improvement seen in the two groups treated with MPH remained, but the effect size decreased by 50%. Twenty-four months after initiation of treatment, 21.8% of youth in the combination MPH and behavioral treatment group stopped treatment versus 25.8% of subjects in the MPH-only group. In the behavioral therapy-only group, at 24-month assessment, 68.2% of subjects added MPH to their treatment regimen.

At the end of 14 months, 77% of the children in the MPH monotherapy treatment group maintained their original dose of MPH. In 23% of subjects who did not maintain the original dose, initial titration did not lead to an adequate response, and 12% did not maintain satisfactory long-term therapeutic gains. For the roughly 5% of youth who did not respond to initial titration of MPH, switching to dextroamphetamine was effective. MPH was later found to be effective in a minority of children who did not respond to the initial MPH titration, reflecting a false-negative titration outcome. Thus, some proportion of ADHD patients do not respond to initial stimulant treatment. However, a significant percentage of these patients will likely respond to another stimulant. Ninety percent of all MTA subjects assigned to MPH treatment (with or without behavioral intervention) responded well to either the first stimulant trial with MPH (79%) or a second stimulant trial with dextroamphetamine (11%).

The Preschool ADHD Treatment Study (PATS) is a double-blind, placebo-controlled, crossover trial that examined the effects of MPH-IR in 165 children, aged 3 to 5.5 years, diagnosed with severe ADHD. Subjects were randomly assigned to four treatment arms, each at a different MPH dose (1.25, 2.5, 5, or 7.5 mg, three times daily), or placebo. Significant improvements in ADHD symptoms were found with an average MPH dose of 14.2 ± 8 mg/day. This average effective dose was considerably lower than the effective dose reported in the MTA study of older children (30.5 ± 14.2 mg/day). Twenty-one percent of the total subjects in the trial were reported to be "excellent responders." During the 10-month maintenance phase of the PATS, 95 of the 140 children demonstrated continued improvement in global functioning, including improvement in social skills. However, parent and teacher symptom rating scales declined over the same time period for about one-third of these children. PATS had a relatively lower effect size range (Cohen's d = 0.22 to 0.87) compared to the MTA study (Cohen's d = 0.8 to 1.2).

Several adverse effects were reported in the PATS. Approximately 30% of PATS subjects developed moderate to severe side effects, and up to 11% of the subjects discontinued treatment due to medication intolerability. Side effects included insomnia, appetite loss, emotional outbursts, repetitive thoughts/behaviors, and irritability. Thus, based on findings from the MTA and PATS studies, it appears that compared to school-aged children, psychostimulants have lower short-term efficacy and lower tolerability in preschoolers. One-third of preschool-aged subjects discontinued stimulants due to adverse events such as worsening behaviors, while others discontinued to switch to a long-acting stimulant or discontinued due to inadequate benefit. Indeed, over a 6-year follow-up period in PATS, the diagnosis of ADHD in preschoolers was relatively stable, with a course characterized generally as chronic, with high symptom severity and impairment, especially in very young children with moderate-to-severe ADHD, despite treatment with medication.

Data in Adults with ADHD. A Cochrane review of seven RCTs, with 1,091 subjects combined, examined the efficacy and tolerability of three different AMPH agents (dextroamphetamine, lisdexamfetamine, and mixed AMPH salts [MAS]) in adults with ADHD. While dexamphetamine, lisdexamfetamine, and MAS, all improved core ADHD symptoms, MAS was associated with the lowest incidence of adverse effects and highest retention in treatment. In other pooled analyses of 19 trials in adults, effect sizes varied based on whether the psychostimulants were long-acting (pooled Cohen's d = 0.73) or immediate-release (pooled Cohen's d = 0.96) formulations. LDX had a significantly higher pooled effect size (Cohen's d = 1.5) based on several RCTs compared to other psychostimulants.

Other Clinical Indications of Psychostimulants

Beyond their use for the treatment of ADHD, psychostimulants have been used in several conditions, including oppositional defiant disorder (ODD), conduct disorder (CD), symptoms of inattention and hyperactivity associated with autism, cognitive deficits associated with traumatic brain injury (TBI), apathy, depression, and cognitive deficits associated with seizure disorder and epilepsy. Stimulant use in these and other disorders has been less extensively studied compared to ADHD, and the general recommendation to providers is to administer stimulants at lower doses than typically prescribed for the treatment of ADHD. The following sections consider the use of psychostimulants in each of these disorders.

Oppositional Defiant Disorder (ODD) and Conduct Disorder (CD) Comorbid with ADHD

Up to 70% of ADHD patients also carry a diagnosis of ODD or CD. In a 4-week, multicenter, randomized, double-blind, parallel-group, forced dose-escalation study, 308 children and adolescents diagnosed with ODD with and without ADHD were treated daily with 30 or 40 mg daily of MAS XR (Adderall XR). Independent of the ADHD diagnosis, treatment with Adderall XR led to a reduction of ODD symptoms measured with the SNAP-IV parent rating subscale. Subjects who received both 30-mg/day (Cohen's d = 0.52) and 40-mg/day (Cohen's d = 0.56) MAS XR showed efficacy in the medium effect size range. Adverse effects were mild to moderate in severity and included anorexia/decreased appetite (25.3%), insomnia (19.5%), headache (18.5%), and abdominal pain (10.7%).

In a similar trial of 84 children between the ages of 6 and 15 years with co-occurring CD in two-thirds of children with ADHD, participants were randomized to MPH hydrochloride (average MPH hydrochloride dose was 41.3 mg/day) or placebo for 5 weeks. Behavior was evaluated by parent, teacher, and clinician reports and by direct classroom observations. Compared with the placebo group, MPH showed significant a benefit for

ADHD symptoms and CD symptoms, including cheating and stealing, obscene language, attacks on others, vandalism of property, and deliberate cruelty.

Treatment of Neurodevelopmental Disorders with ADHD Symptoms. ADHD symptoms occur in 41% to 78% of children with autism spectrum disorders (ASDs). These symptoms often affect quality of life, interfering with learning or interventions that target primary ASD symptoms. Children who have developmental disabilities (including ASD, intellectual disabilities, fragile X syndrome), have lower effect sizes with these medications and are more sensitive to adverse effects, including agitation and emotional lability, compared with typically developing children with ADHD. One large RCT of MPH in children with ASD found that only 49% of children responded to stimulants compared to the 69% response rate in the MTA study. In addition, 18% of the children with ADHD and ASD discontinued participation due to adverse events (prominently irritability) compared to 1.4% in children with ADHD only.

NEUROPSYCHIATRIC DEFICITS SECONDARY TO TRAUMATIC BRAIN INJURY. The overall risk of TBI in the general population has been estimated at 2% to 5%. In the United States, approximately 230/100,000 of children under 15 years of age, and 294/100,000 adolescents between 15 and 19 years of age experience a TBI each year. A significant proportion of patients who sustained a TBI will experience disturbance in academic, emotional, and social functioning. In a research setting, the most common psychiatric complication of TBI was a cluster of symptoms consistent with ADHD. A double-blind, placebo-controlled, crossover-study design used to investigate the efficacy of MPH in children with ADHD symptoms secondary to a TBI has demonstrated that MPH was superior to placebo to benefit attention and concentration. This finding is in line with other RCTs that have demonstrated benefit of other stimulants, including pemoline and dextroamphetamine, for symptoms of ADHD secondary to TBI.

APATHY. Apathy is a common symptom of neurodegenerative diseases such as Alzheimer disease (AD), Parkinson disease (PD), as well as an adverse effect of the sedative effects of pain medications, or toxic effects of chemotherapy. Unfortunately, few studies have investigated the benefit of psychostimulants in treating apathy secondary to these conditions. One

RCT examining the efficacy of MPH for the treatment of apathy associated with AD demonstrated greater reduction of apathy with MPH 10 mg twice daily compared with placebo on Apathy Evaluation Scale change scores. Thus, efficacy studies of psychostimulants for the treatment of secondary neuropsychiatric complications are all limited by a small sample size and necessitate larger trials to guide evidence-based practice.

DEPRESSION. Antidepressants are first-line treatment for moderate or severe depression, but depression resistant to antidepressants presents a therapeutic challenge. A Cochrane review article assessed evidence from RCTs on the effectiveness and safety of stimulants, such as MPH and methylamphetamine, and modafinil and compared them to placebo in the treatment of major depressive disorder, either as monotherapy or as adjuncts to antidepressants. Several placebo-controlled parallel design studies have demonstrated benefit of adjunctive psychostimulants to reduce depressed mood in the short term. Further examination and pooling of longer-term outcomes for other symptoms of depression suggest that psychostimulants reduce fatigue associated with depression, but do not effectively target other neurovegetative symptoms in depression. Other smaller short-term trials of psychostimulants used adjunctively versus placebo have not found significant differences in overall depression scores. In addition, modafinil is no more effective than placebo in targeting symptoms of depression including fatigue or hypersomnia. The overall quality of the extant data is suboptimal due to small sample sizes, selection bias due to inadequate randomization, attrition, and performance bias.

Treatment of Comorbid ADHD and Seizure Disorder. A retrospective study of youth with comorbid epilepsy and ADHD found that being seizure free did not predict response to stimulant treatment. An MPH preparation had a 5.57-fold greater chance of treatment response compared to an AMPH preparation. Importantly, cognitive level may play a role in treatment response, but age, type of seizure, and success of previous stimulant trials were not associated with response to stimulants in this population. Stimulant discontinuation rates were comparable in seizure-free group (35%) and epilepsy group (53%). In the not-seizure-free group, 3 out of the 19 patients had a worsening of seizures while on a stimulant.

Currently, no specific treatment is recommended for ADHD associated with epilepsy. Rather, treatment is based on the usual treatments of ADHD, and reassurance about the safety of these treatments in epilepsy.

Tolerability of Psychostimulants: Adverse Effects, Contraindications, and Precautions

Psychostimulants are generally well tolerated. Different formulations of MPH and AMPH have similar side-effect profiles that present similarly in pediatric and adult populations, with some exceptions that will be noted below. Common side effects associated with stimulants include insomnia, decreased appetite, weight loss, headache, stomachache, and mild increases in heart rate and blood pressure (see Table 33.4b–4 for a comprehensive list of adverse effects). Short-term or acute stimulant side effects are more likely to be related to higher doses, and careful dose titration mitigates these acute adverse effects.

Certain patient groups appear to be more sensitive to stimulants at doses lower than the general population, including preschool children with ADHD, children with developmental delays, and the older adults. It is recommended that stimulants be used with caution in the following populations: children under 5 years of age with IQ ≤5 years old, severe tics, or in children with histories of severe anxiety, mood, or psychotic symptoms. Further consideration of lower doses should be given in the older adult population. Nevertheless, stimulants may be first-line treatment for ADHD symptoms in these populations, despite a relatively small number of RCTs, response rates of 50%, and higher rates of side effects. Absolute and relative contraindications for stimulant use are listed in Table 33.4b–5. Specific adverse reactions to stimulants will now be discussed in more detail.

Table 33.4b–4.
Comprehensive List of Adverse Events

Adverse Effects	
Common:	***Uncommon:***
MPH/AMPH	Psychosis
Anorexia, weight loss (13–20%)	Euphoria/mania Sadness/isolation
Insomnia (11–30%)	Major depressive episodes Cognitive impairment Growth retardation
Headaches (12–15%)	Tic disorders (i.e., Tourette syndrome) Increased heart rate (clinically significant)
Abdominal pain (6–12%)	Impaired liver functioning (pemoline only) Increased blood pressure
Emotional lability (2–10%)	Dizziness, lethargy, fatigue Nausea, constipation Rash/hives
Slight increase in HR, BP	Hyperacusis
***Severe side effects occur in 4–10% of children treated with stimulants**	Formication Necrotizing angiitis brain (IV amphetamine) Neuroleptic malignant syndrome?

Specific Adverse Events

Tics. Significant concerns regarding the development of tics and Tourette disorder due to use of psychostimulants in children and adolescents were initially raised in the 1980s. For many years, stimulants had been thought to induce or worsen tics in youth, and presence of tics constituted a contraindication to psychostimulant use. In the last decade, several studies have suggested that stimulants do not worsen or reactivate tics, and might even improve pre-existing tics in many patients.

A Cochrane review analyzed eight RCTs for a potential relation between psychostimulant use and the development of tic disorders. One RCT in children with co-occurring ADHD and Tourette syndrome found a positive treatment effect with MPH compared to placebo and with clonidine plus MPH on measures of ADHD and tic severity. In the MPH group, only 20% of participants reported worsening of tics as an adverse event compared to 22% receiving placebo. In a small placebo-controlled crossover study of dextroamphetamine that randomized 20 children to three dosage titrations of dextroamphetamine (target doses of 7.5, 15, or 22.5 mg twice daily) for 3 weeks, dextroamphetamine decreased hyperactivity but there was no significant interaction between drug and dose, indicating that additional improvements in hyperactivity were not observed for higher doses. Tic severity was significantly greater during the second (15 mg twice daily) and third weeks (22.5 mg twice daily) of dextroamphetamine treatment compared to weeks on placebo. Drug dose did not worsen tics, however. Larger studies are needed to definitively determine whether psychostimulants worsen tics or reduce pre-existing tics in youth with comorbid ADHD and a tic disorder.

Table 33.4b–5.
Contraindications to Psychostimulants

Absolute Contraindications	Relative Contraindications
1. Known hypersensitivity to stimulants 2. Structural cardiac defects 3. Narrow-angle glaucoma 4. Moderate to severe substance abuse disorder 5. Current use of MAOI or within 14 days after an MAOI has been discontinued	1. Presence of psychotic disorders like schizophrenia because of stimulants' potential to worsen or even precipitate psychotic symptoms 2. A severe tic or Tourette syndrome is a relative contraindication. This contraindication does not apply to mild tics in patients with severe ADHD symptoms 3. Unstable hypertension is a relative contraindication. Blood pleasure should be normalized and well controlled before administering stimulant medications

MAOI, monoamine oxidase inhibitor; ADHD, attention-deficit hyperactivity disorder.

The increased sensitivity of dopaminergic brain systems has been widely implicated in the pathophysiology of tics. To date, no clear genetic relation between tics and ADHD on the emergence of tics during stimulant trials has been established. Moreover, there is commonly a delay between stimulant exposure and tic onset, making it difficult to establish a causal association between stimulant exposure and tic outcome. In individuals with a personal or family history of tics or Tourette disorder and ADHD, stimulant medications may exacerbate motor and vocal tics. However, the data are not conclusive and, in most cases, the benefit of treating ADHD outweighs the risk of stimulants triggering tics.

Cardiovascular Effects. A risk of sudden cardiac death from psychostimulant use emerged around the time they became a mainstay treatment for ADHD. The U.S. FDA identified 11 sudden deaths in pediatric patients taking MPH from 1992 to 2005. However, the rate of sudden cardiac death among patients with ADHD is equivalent to that of the general population and comparable to mortality due to motor vehicle accidents in children. In large community samples of children and adolescents, stimulants have not been shown to increase the risk of cardiovascular complications, including a risk of sudden death. The risk for stimulant-induced adverse cardiovascular events appears to be elevated in children with ADHD and a family history significant for early cardiac death, arrhythmia, cardiogenic syncope, or a personal history of structural cardiac abnormalities, exercise-induced syncope, chest pain, or palpitations. According to the AACAP practice guidelines, in such cases, an assessment by a pediatric cardiologist prior to stimulant treatment is recommended. Finally, the authors of the large prospective MTA study investigated stimulant-related changes in vital signs in youth treated for ADHD for up to 10 years. Treatment was associated with increases in heart rate at 14-month, 3-year, and 8-year follow-ups, suggesting chronic induction of the sympathomimetic system by stimulant treatment. Stimulants were not associated with elevations in systolic or diastolic blood pressure.

Growth Velocity. Long-term stimulant use has been associated with interfering with typical growth trajectories in children and adolescents with ADHD. Specifically, MTA subjects reported significant decreases in height velocity with long-term stimulant treatment. Maximal reduction in annual growth rate occurred in the first year, decreased in the second year, and then subsided completely by the third year of treatment. Matched subjects with a history of treatment before entry into the MTA study did not show differences in height velocity compared to treatment-naive youth over 3 years of follow-up. Treatment-naive youth had a small growth acceleration at the 2-year follow-up that was equivalent to the untreated group, and the treatment-naive group on stimulants did not compensate for the slowdown of growth observed initially. Compared to typically developing healthy comparison youth without a diagnosis of ADHD, ADHD youth not treated with stimulants were taller and heavier and grew at a faster rate. It is hypothesized that stimulant-related reductions in growth may be related to a DA-induced decrease of prolactin secretion by the pituitary gland or suppression of growth hormone (GH) and insulin-like growth factors (IGFs).

The PATS study reported significant growth velocity suppression in preschoolers treated with stimulants for at least 1 year, with a 20% less than expected height gain (−1.38 cm/year) and 55% less than expected weight gain (−1.32 kg/year). Children under 6 years of age (mean age = 4.4 years) with ADHD were measured up to 29 times for at least 10 months in study follow-up. The height and weight of 140 stimulant-naive preschool-aged children with a diagnosis of ADHD were much larger (an average of 2.04 cm taller and 1.78 kg heavier) than referenced Centers for Disease Control (CDC) norms. After initiation of stimulant medication, the height and weight velocity slowed, and for the children who remained on medication, this resulted in annual gain that was 20% less for height and 55% less for weight than expected by CDC standards. These findings are consistent with the growth velocity findings reported in the MTA study. Moreover, greater stimulant-related reductions in height velocity have been observed with longer exposure to psychostimulants, in younger children compared to older adolescents, and for taller and heavier children. The magnitude of weight and height deficits does not appear to differ between MPH and AMPH.

In the long term, growth deficits normalize, so that final height is not affected. In some children with ADHD, deficits in physical growth constitute a developmental delay, much like core behavioral and cognitive signs of the disorder. Indeed, stimulant-associated growth deficits in children with ADHD may represent early temporary manifestations of ADHD itself. In addition, a failure to gain weight may impact height velocity, although no studies have directly examined this possibility in youth with ADHD or have followed patients into adulthood to investigate a complete height trajectory. Stimulant effects on child growth should be part of a risk/benefit discussion with the child and family as part of the informed consent, and clinicians should monitor growth parameters in all patients taking stimulants. Change in weight or height crossing two percentile lines on a standardized growth curve suggest a deficit in growth that should prompt a clinical response. Possible responses might include a stimulant drug holiday, nutritional supplementation, dose reduction, or change to a nonstimulant ADHD medication. Indeed, drug holidays, caloric supplementation, and increased monitoring demonstrably lead to increased weight but not increased height velocity in children taking CNS stimulants.

Abuse and Dependence. The risk of substance use, abuse, or dependence in individuals diagnosed with ADHD complicates the diagnosis and is of great interest to clinicians and researchers. In a 10-

year follow-up study, children and adolescents diagnosed with ADHD had an increased risk of a substance use disorder (SUD) compared to matched controls. Factors such as ADHD severity and duration, comorbid mood disorders, and CD are independently associated with a higher risk for SUD compared to controls. The presence or absence of treatment did not appear to change the risk of substance use, abuse, or dependence. In the MTA group, there were no protective or predisposing factors associated with SUD at the 3-year follow-up. At the 6-year follow-up (mean age = 15 years), 35% of the ADHD struggled with an SUD versus 20% of controls. Thus, according to the MTA study, stimulants neither protect from nor increase the risk for SUD. Finally, in a prospective study, the risk of SUD increased with age at the initiation of MPH, suggesting that delays in the treatment of ADHD may be associated with the development of SUD.

Misuse and Diversion. When taken as indicated, stimulants are generally safe and highly effective for the treatment of ADHD symptoms. However, stimulant-induced elevations in DA in key brain regions appear to be integral to both efficacy in ADHD and the potential for abuse. The nonmedical use of stimulants has received increasing attention and has prompted the consideration of nonstimulant options for the treatment of ADHD. Data from the National Survey on Drug Use and Health (NSDUH) Monitoring the Future database tracks the diversion and misuse of stimulants. In 2020, the NSDUH found that approximately 1.1% of 8th graders and 1.7% of 12th graders reported lifetime nonmedical misuse, with upward of 4.4% of 12th graders reporting past year nonmedical misuse of Adderall, with relatively stable rates reported between 2017 and 2020.

Benson and colleagues (2015) published a meta-analysis on the misuse and diversion in the college student population. They found that a substantial number of college students are involved in nonmedical misuse and diversion of stimulants. Diversion is defined as giving away, selling, or otherwise distributing one’s own medication. In 2014, 36% of stimulant prescription holders diverted their stimulant, and Adderall was the most frequently diverted stimulant. Fifty-six percent of prescription holders had been approached to divert their medication in the last 6 months. Factors associated with misuse of stimulants were male gender, being a member of a social college organization, having a diagnosis of CD and/or SUD, using immediate-release formulations, or using alcohol or illicit substances.

Thus, stimulant diversion and misuse among student populations are growing and are higher than that among general population samples. However, misuse appears to be most prevalent and problematic among adults with ADHD and illicit drug users. Based on these data, the AACAP practice parameters recommend that clinicians obtain a careful personal and family history of substance abuse or dependence prior to prescribing a stimulant and should educate populations most vulnerable to diversion, including high school or college students.

MODAFINIL (PROVIGIL)

Neuroscience-Based Nomenclature (NbN) **DRI**	
Pharmacologic domain	Dopamine
Mechanism of action	Reuptake inhibitor (DAT)

Modafinil is a wakefulness-promoting agent with behavioral actions similar to psychostimulants but is different pharmacologically not to be classified as a psychostimulant. Physiologically, modafinil can increase energy, concentration, and induce wake-promoting effects. Modafinil is approved for the treatment of excessive sleepiness associated with narcolepsy, obstructive sleep apnea/hypopnea syndrome (OSAHS), or shift work sleep disorder (SWSD). In addition, multiple studies have investigated the efficacy and safety of modafinil in the treatment of ADHD, mood disorders, schizophrenia, and SUD.

Mechanism of Action

Modafinil is a racemic compound. The chemical name for modafinil is diphenylmethyl sulfinyl acetamide, and its molecular formula is $C_{15}H_{15}NO_2S$. Modafinil and its R-enantiomer, armodafinil, exert clinical effects by enhancing catecholamine neurotransmission. Modafinil has been found to activate lateral hypothalamus orexin (hypocretin) neurons that send axons to numerous brain regions, including major nuclei that have been implicated in sleep regulation. However, because its pharmacologic profile is unique from stimulants, it is not classified as such. In addition to the wake-promoting effects, modafinil increases motor activity, euphoria, and alterations in perception, emotions, and cognition typical of CNS stimulants. Its optical enantiomer, armodafinil, has similar pharmacologic actions in animals.

Modafinil, like other psychostimulants, increases the release of monoamines but also elevates histamine, epinephrine, γ-aminobutyric acid, and glutamate levels. As a result, some researchers consider modafinil a “wakefulness-promoting agent” rather than a classic AMPH-like psychostimulant. For example, mice lacking DATs do not respond to the wake-promoting effects of modafinil, and this is also true for mice lacking D_1 and D_2 receptors. Further, microdialysis studies have reported that modafinil increases extracellular DA. PET with a D_2/D_3 radioligand (sensitive to changes in endogenous DA) and [11C] cocaine (DA transporter radioligand) was used to measure the effects of modafinil on extracellular DA and on DAT in 10 healthy males. A significant correlation between the percentage of DA transporter occupancy and plasma modafinil concentration in the caudate was demonstrated. Other studies have demonstrated that in mice engineered without DAT, modafinil failed to induce wake-promoting activity, suggesting that DAT was essential for its ability to exert its effects. Unlike those of AMPH, however, the function of modafinil was not inhibited by a DA receptor antagonist. These studies suggest that despite some functional mimicry, modafinil acts through different cellular mechanisms than through stimulants.

Studies have also demonstrated that modafinil has moderate interaction with NE and serotonin receptors. However, unlike stimulants, modafinil does not release NE at a high rate. It has been hypothesized that this difference in effect on the catecholamine system might be responsible for lower efficacy in reducing ADHD symptoms compared to psychostimulants.

Pharmacologic Properties of Modafinil

Modafinil is rapidly absorbed, and the absorption of the medication is not affected at a steady state; a peak plasma concentration of 6.4 mg/mL was reached in 2.7 hours with modafinil 200 mg once daily. Approximately 90% of a modafinil dose is metabolized hepatically, primarily to two inactive metabolites. Less than 10% of an administered dose is excreted as the parent compound. The pharmacokinetics of modafinil may be altered in older males and in patients with chronic hepatic impairment. In vitro, modafinil has been shown to inhibit the cytochrome P450 (CYP) isoenzyme CYP2C19 (moderately potent, reversible inhibition), modestly induce CYP1A2, CYP3A4/5, and CYP2B6, and suppress CYP2C9.

U.S. FDA–Approved Uses of Modafinil

Narcolepsy. Narcolepsy is a common, disabling sleep disorder characterized by excessive daytime sleepiness (EDS) and rapid eye movement (REM) sleep manifestations that intrude into wakefulness, including cataplexy (sudden loss of muscle tone triggered by strong emotions), sleep paralysis, hypnagogic/hypnopompic hallucinations (sensory events that occur at the transition from wakefulness to sleep), and sleep-onset REM periods. Narcolepsy affects approximately 0.02% of the population worldwide and causes disability in 24% of the affected subjects.

Currently, modafinil is considered a first-line therapy for EDS in narcolepsy. Four randomized, double-blind, placebo-controlled trials have assessed the usefulness of modafinil in the treatment of EDS in narcolepsy and have demonstrated that treatment with modafinil 200 to 400 mg/day had a significant positive impact on improved EDS. Modafinil does not appear to benefit cataplexy. Abrupt discontinuation may cause rebound sleepiness but does not appear to be associated with any AMPH-like withdrawal symptoms.

Obstructive Sleep Apnea. The efficacy and safety of modafinil treatment for obstructive sleep apnea (OSA) in patients who are not benefiting from continuous positive airway pressure (CPAP) have been demonstrated in a meta-analysis of three RCTs in this population. Modafinil is also effective in improving intentional or unintentional sleep in SWSD, which translates into fewer accidents and near accidents on the commute home but not during the night shifts.

Non–U.S. FDA–Indicated Use of Modafinil

Attention-Deficit Hyperactivity Disorder. Although sympathomimetic stimulants are the most common treatment for ADHD, a sizable minority of children and adolescents with ADHD do not respond adequately to psychostimulants or have intolerable side effects. Thus, additional treatment options are needed. Given the attention- and wakefulness-enhancing properties of modafinil, it has been considered a candidate medication for ADHD. By 2007, multiple trials had examined the clinical utility of modafinil for ADHD. More than 900 children and adolescents were examined and demonstrated the efficacy and tolerability of modafinil for the treatment of ADHD.

In 2005, 250 6- to 17-year-old youth, normalized for demographic factors, were involved in a 9-week, multicenter, randomized, double-blind, placebo-controlled, flexible-dose study of the film-coated tablet formulation of modafinil for the treatment of ADHD symptoms. Modafinil was titrated to an optimal dose based on efficacy and tolerability (range: 170 to 425 mg once daily). Efficacy was assessed by clinicians (who completed the Attention-Deficit Hyperactivity Disorder Rating Scale-IV [ADHD-RS-IV]), interviews with teachers (who completed the School Version of the ADHD-RS-IV), and parents (who completed the Home Version of the ADHD-RS-IV) and the Clinical Global Impression of Improvement scale. Safety assessment was based on adverse events reported, laboratory tests, and vital signs. The effect size for improvement of core ADHD symptoms was moderate to large (Cohen’s d = 0.69). Thus, once-daily modafinil was effective on the core symptoms of inattention, hyperactivity, and impulsivity, both at school and at home. Among the most common adverse events reported were insomnia and decreased appetite, which were mild to moderate in severity.

Cocaine Addiction. Several small RCTs have examined the use of modafinil to prevent cravings in patients with cocaine addiction. In a parallel-group study, 62 treatment-seeking cocaine-dependent subjects without significant medical or psychiatric problems were randomized to receive modafinil 200 to 400 mg/day or a placebo for 8 weeks. All subjects also received cognitive behavioral therapy twice a week. The efficacy measures were clean urine toxicology, subjective craving, and withdrawal reporting. Even though there were significantly more negative urine results in patients taking modafinil, the subjective reporting of cocaine use, dollars spent, and cocaine craving remained unchanged. At present, there is no consistent evidence that modafinil is beneficial for the treatment of cocaine cravings.

Cognitive Enhancement. A number of neurodevelopmental and neuropsychiatric disorders are associated with cognitive impairment. Impaired cognition can lead to poor functional outcomes and may limit the benefit gained by psychosocial interventions. Modafinil has recently emerged as a possible medication to improve cognition for certain patients. It has been shown to improve cognitive function, including memory, planning, and attention in animal studies, as well as reaction time, logical reasoning and short-term working memory, planning, and mental flexibility in non–sleep-deprived and sleep-deprived humans. Modafinil may offer advantages over compounds such as AMPH and L-dopa, particularly if repeated administration is needed, as it is associated with a low risk of abuse and has a relatively mild side-effect profile.

Adverse Effects. Modafinil is generally well tolerated in patients with excessive sleepiness associated with OSAHS or SWSD. Adverse events are generally of mild to moderate severity and include headache, nausea, and nervousness.

ATOMOXETINE (STRATTERA)

Neuroscience-Based Nomenclature (NbN)	
Atomoxetine	
Pharmacologic domain	Norepinephrine
Mechanism of action	Reuptake inhibitor (NET)

ATM was the first U.S. FDA medication approved for ADHD in children older than 6 years of age, adolescents, and adults. Similar to psychostimulants, ATM shows potent and very selective NE transporter binding. It selectively blocks NE reuptake at the noradrenergic neuron and is classified as a selective norepinephrine reuptake inhibitor (SNRI). However, it is not classified as a psychostimulant and is not a controlled substance.

Mechanism of Action

ATM has been shown to increase the extracellular availability of both NE and DA in the rat PFC. ATM improves measures of PFC cognitive function in rats, monkeys, and humans. In order to improve EF in primates, ATM has to produce an optimal dose of catecholamines in the synaptic space; higher than optimal doses may impair EF. The enhancing effects are blocked by the coadministration of either an NE α_2 or DA D_1 antagonist, suggesting that the NE and DA neurotransmission complement are essential for ATM function. Interestingly, optimal doses vary widely, which may explain the variability of drug response in ATM compared to psychostimulants.

Pharmacologic Properties and Drug Interactions of Atomoxetine

ATM has high aqueous solubility and lipophilicity that facilitates its rapid and complete GI absorption after oral administration. In the serum, 98% of the medication is bound primarily to albumin. Three oxidative metabolic pathways are involved in the hepatic metabolism of ATM: aromatic ring hydroxylation, benzylic hydroxylation, and *N*-demethylation. Aromatic ring hydroxylation results in the formation of the primary oxidative metabolite of ATM, 4-hydroxy atomoxetine, which is subsequently excreted in the urine. The formation of 4-hydroxy atomoxetine is primarily mediated by P450 (CYP) 2D6. This results in two distinct populations of individuals: those exhibiting active metabolic capabilities (CYP2D6 extensive metabolizers) and those exhibiting poor metabolic capabilities (CYP2D6 poor metabolizers) for ATM.

The oral bioavailability and clearance of ATM are influenced by the activity of CYP2D6. In extensive metabolizers, ATM has a plasma half-life of 5.2 hours. Poor metabolizers, ATM has a plasma half-life of 21.6 hours. As a result, the average steady-state plasma concentrations are approximately 10-fold higher in poor metabolizers compared with extensive metabolizers. ATM accumulates in the plasma of poor metabolizers but very little accumulation in extensive metabolizers. The volume of distribution in ATM suggests that it is distributed in total-body water, in both extensive and poor metabolizers. ATM is highly bound to plasma albumin (~99% bound in plasma). ATM does not inhibit or induce the clearance of other medications metabolized by the CYP system. In extensive metabolizers, potent and selective CYP2D6 inhibitors reduce ATM clearance; however, administration of CYP inhibitors to poor metabolizers has no effect on the steady-state plasma concentrations of ATM.

Albuterol used concomitantly with ATM may increase heart rate and blood pressure. ATM is contraindicated in patients currently receiving or having recently received (i.e., within 2 weeks) MAO inhibitor therapy. The starting dose of ATM should be reduced when potent CYP2D6 inhibitors, including fluoxetine, paroxetine, or quinidine, are administered concomitantly with ATM.

FURTHER READINGS

Atomoxetine/Methylphenidate Comparative Study Group; Newcorn JH, Kratochvil CJ, Allen AJ, et al. Atomoxetine and osmotically released methylphenidate for the treatment of attention deficit hyperactivity disorder: acute comparison and differential response. *Am J Psychiatry*. 2008;165(6):721–730.

Biederman J, Swanson JM, Wigal SB, Boellner SW, Earl CQ, Lopez FA; Modafinil ADHD Study Group. A comparison of once-daily and divided doses of modafinil in children with attention-deficit/hyperactivity disorder: a randomized, double-blind, and placebo-controlled study. *J Clin Psychiatry*. 2006;67(5):727–735.

Candy M, Jones L, Williams R, Tookman A, King M. Psychostimulants for depression. *Cochrane Database Syst Rev*. 2008;(2):CD006722.

Castells X, Ramos-Quiroga JA, Bosch R, Nogueira M, Casas M. Amphetamines for attention deficit hyperactivity disorder (ADHD) in adults. *Cochrane Database Syst Rev*. 2011;(6):CD007813.

Clemow DB, Walker DJ. The potential for misuse and abuse of medications in ADHD: a review. *Postgrad Med*. 2014;126(5):64–81.

Coghill DR, Banaschewski T, Soutullo C, Cottingham MG, Zuddas A. Systematic review of quality of life and functional outcomes in randomized placebo-controlled studies of medications for attention-deficit/hyperactivity disorder. *Eur Child Adolesc Psychiatry*. 2017;26(11):1283–1307.

Cortese S. Pharmacologic treatment of attention deficit-hyperactivity disorder. *N Engl J Med*. 2020;383(11):1050–1056.

Gonzalez-Heydrich J, Hsin O, Gumlak S, et al. Comparing stimulant effects in youth with ADHD symptoms and epilepsy. *Epilepsy Behav*. 2014;36:102–107.

Gould MS, Walsh BT, Munfakh JL, et al. Sudden death and use of stimulant medications in youths. *Am J Psychiatry*. 2009;166(9):992–1001.

Greenhill L, Kollins S, Abikoff H, et al. Efficacy and safety of immediate-release methylphenidate treatment for preschoolers with ADHD. *J Am Acad Child Adolesc Psychiatry*. 2006;45(11):1284–1293.

Faraone SV, Larsson H. Genetics of attention deficit with hyperactivity disorder. *Mol Psychiatry*. 2019;24(4):562–575.

Herrmann N, Rothenburg LS, Black SE, et al. Methylphenidate for the treatment of apathy in Alzheimer disease: prediction of response using dextroamphetamine challenge. *J Clin Psychopharmacol*. 2008;28(3):296–301.

Hoogman M, Muetzel R, Guimaraes JP, et al. Brain imaging of the cortex in ADHD: a coordinated analysis of large-scale clinical and population-based samples. *Am J Psychiatry*. 2019;176(7):531–542.

Kuczenski R, Segal DS. Effects of methylphenidate on extracellular dopamine, serotonin, and norepinephrine: comparison with amphetamine. *J Neurochem*.1997;68(5):2032–2037.

Mahajan R, Bernal MP, Panzer R, et al; Autism Speaks Autism Treatment Network Psychopharmacology Committee. Clinical practice pathways for evaluation and medication choice for attention-deficit/hyperactivity disorder symptoms in autism spectrum disorders. *Pediatrics*. 2012;130(Suppl 2):S125–S138.

Molina BSG, Hinshaw SP, Eugene Arnold LE, et al; MTA Cooperative Group. Adolescent substance use in the multimodal treatment study of attention-deficit/hyperactivity disorder (ADHD) (MTA) as a function of childhood ADHD, random assignment to childhood treatments, and subsequent medication. *J Am Acad Child Adolesc Psychiatry*. 2013;52(3):250–263.

MTA Cooperative Group. A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. The MTA Cooperative Group. Multimodal Treatment Study of Children with ADHD. *Arch Gen Psychiatry*. 1999;56(12):1073–1086.

MTA Cooperative Group. National Institute of Mental Health Multimodal Treatment Study of ADHD follow-up: 24-month outcomes of treatment strategies for attention-deficit/hyperactivity disorder. *Pediatrics*. 2004;113(4):754–761.

Mueller A, Hong DS, Shepard S, Moore T. Linking ADHD to the neural circuitry of attention. *Trends Cogn Sci*. 2017;21(6):474–488.

Pliszka S; AACAP Work Group on Quality Issues. Practice parameter for the assessment and treatment of children and adolescents with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*. 2007;46(7):894–921.

Posner J, Polanczyk GV, Sonuga-Barke E. Attention-deficit hyperactivity disorder. *Lancet*. 2020;395(10222):450–462.

Pringsheim T, Steeves T. Pharmacological treatment for attention deficit hyperactivity disorder (ADHD) in children with comorbid tic disorders. *Cochrane Database Syst Rev*. 2011;(4):CD007990.

Riddle MA, Yershova K, Lazzaretto D, et al. The Preschool Attention-Deficit/Hyperactivity Disorder Treatment Study (PATS) 6-year follow-up. *J Am Acad Child Adolesc Psychiatry*. 2013;52(3):264–278.e2.

Samea F, Soluki S, Nejati V, et al. Brain alterations in children/adolescents with ADHD revisited: a neuroimaging meta-analysis of 96 structural and functional studies. *Neurosci Biobehav Rev*. 2019;100:1–8.

Sáenz AA, Villemonteix T, Massat I. Structural and functional neuroimaging in attention-deficit/hyperactivity disorder. *Dev Med Child Neurol*. 2019;61(4):399–405.

Spencer RC, Devilbiss DM, Berridge CW. The cognition-enhancing effects of psychostimulants involve direct action in the prefrontal cortex. *Biol Psychiatry*. 2015;77(11):940–950.

Swanson JM, Elliott GR, Greenhill LL, et al. Effects of stimulant medication on growth rates across 3 years in the MTA follow-up. *J Am Acad Child Adolesc Psychiatry*. 2007;46(8):1015–1027.

Thorpy MJ, Bogan RK. Update on the pharmacologic management of narcolepsy: mechanisms of action and clinical implications. *Sleep Med*. 2020;68:97–109.

*Volkow ND, Wang GJ, Tomasi D, et al. Methylphenidate-elicited dopamine increases in ventral striatum are associated with long-term symptom improvement in adults with attention deficit hyperactivity disorder. *J Neurosci*. 2012;32(3):841–849.

Waxmonsky JG, Pelham WE 3rd, Campa A, et al. A randomized controlled trial of interventions for growth suppression in children with attention-deficit/hyperactivity disorder treated with central nervous system stimulants. *J Am Acad Child Adolesc Psychiatry*. 2020;59(12):1330–1341.

33.5 Monoamine Enhancers (“Antidepressants”)

33.5a Reuptake Inhibitors

33.5a.1 erotonin Reuptake Inhibitors (SRIs)

Simon Lafrenière, M.D., M.Sc., and Pierre Blier, M.D., Ph.D.

Neuroscience-Based Nomenclature (NbN)	
Citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline	
Pharmacologic domain	Serotonin
Mechanism of action	Reuptake inhibitors (SERT)

OVERVIEW

As a class, selective serotonin reuptake inhibitors (SSRIs) have been the most widely used psychopharmacologic agents for depression worldwide for three decades. They are recommended as a first-line treatment for depression in major guidelines. They are still extensively prescribed, in part because they have been found to be effective in treating a range of disorders and symptoms other than depression, including anxiety disorders. From a historic perspective, the most significant impact of fluoxetine and subsequent SSRIs may have been the attention they generated in the popular press, which helped to reduce the longstanding stigma and fear of taking medications for depression. Patients became more amenable to coming forward for treatment.

Clinicians also became less reluctant to treat depression with medications because the adverse event profile of the SSRIs was devoid of side effects associated with the earlier drugs—the tricyclic derivatives and monoamine oxidase inhibitors (MAOIs). These included dry mouth, constipation, sedation, orthostatic hypotension, and tachycardia. The ease of SSRI dosing with their initial posology being already in the minimal efficacious range led to clinicians other than psychiatrists prescribing SSRIs. Another very important consideration in the adoption of SSRIs as first-line treatments was their wide therapeutic index. Deaths from overdoses of tricyclic derivatives plummeted after SSRIs supplanted the former class of drugs as the mainstays of treatment.

It needs to be kept in mind that SSRIs have never been shown to be more effective than any other classes of medications for depression. All the SSRIs, starting with fluoxetine and followed by sertraline (Zoloft), paroxetine (Paxil), fluvoxamine (Luvox), citalopram (Celexa), and escitalopram (Lexapro), have been shown to be equally effective in the treatment of depression. There are, however, meaningful differences in their side effect profile, and taking these differences into consideration may improve compliance, thereby favoring clinical responses. This is of primary importance as a study showed that 42% of patients discontinued their medication for depression in the 30 days following the prescription, and only 28% were still taking the medication after 90 days.

SSRIs have proven more problematic in terms of side effects than the original clinical trials suggested. In clinical practice, quality-of-life-associated adverse effects such as sexual dysfunction, weight gain, and emotional blunting often mitigate the therapeutic benefits of the SSRIs. These side effects were not captured in the short-term pivotal trials of these agents. It should also be noted that some patients report a diminished response or total loss of response to SSRIs with a recurrence of depressive symptoms while remaining on a full dose of medication. The exact mechanism of this so-called “poop-out” is unknown but possibly includes desensitization of postsynaptic serotonin (5-hydroxy-tryptamine; 5-HT) receptors or a worsening of the underlying biologic anomalies that initially caused depression.

Not included in this section are vilazodone (Viibryd) and vortioxetine (Trintellix). Even though they inhibit 5-HT reuptake, they are also 5-HT_{1A} receptor agonists, which may be relevant to their therapeutic benefits.

THERAPEUTIC INDICATIONS

The extensive use of SSRIs also derives from the fact that they are approved by the FDA for the following indications: major depressive disorder (MDD), generalized anxiety disorder (GAD), obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), premenstrual dysphoric disorder (PMDD), panic disorder, and social anxiety disorder (SAD). Not all SSRIs have Food and Drug Administration (FDA) approval for all these indications, but it is commonly assumed that they have a common spectrum of action. All the SSRIs, except for fluvoxamine, were initially approved for the treatment of MDD. Fluvoxamine is unique in that it is only indicated for the treatment of OCD in the United States. This is a consequence of commercial considerations, rather than any significant difference in its clinical profile. It shares the same therapeutic spectrum of activity with the other SSRIs and is widely used as a treatment for MDD in other countries worldwide. The currently approved indications for the SSRIs are shown in Table 33.5a.1–1.

Table 33.5a.1–1.
Approved Indications of the Selective Serotonin Reuptake Inhibitors in the United States for Adult and Pediatric Populations

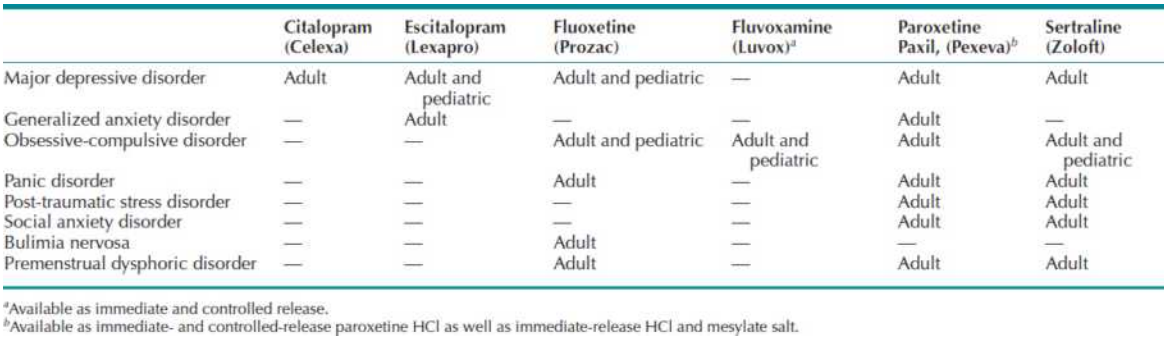

	Citalopram (Celexa)	Escitalopram (Lexapro)	Fluoxetine (Prozac)	Fluvoxamine (Luvox)[a]	Paroxetine Paxil, (Pexeva)[b]	Sertraline (Zoloft)
Major depressive disorder	Adult	Adult and pediatric	Adult and pediatric	—	Adult	Adult
Generalized anxiety disorder	—	Adult	—	—	Adult	—
Obsessive-compulsive disorder	—	—	Adult and pediatric	Adult and pediatric	Adult	Adult and pediatric
Panic disorder	—	—	Adult	—	Adult	Adult
Post-traumatic stress disorder	—	—	—	—	Adult	Adult
Social anxiety disorder	—	—	—	—	Adult	Adult
Bulimia nervosa	—	—	Adult	—	—	—
Premenstrual dysphoric disorder	—	—	Adult	—	Adult	Adult

[a] Available as immediate and controlled release.
[b] Available as immediate- and controlled-release paroxetine HCl as well as immediate-release HCl and mesylate salt.

Major Depressive Disorder

Thousands of patients have participated in dose-ranging studies and active comparator trials of SSRIs. They may well be the most-studied drugs in the history of psychiatry in terms of the number of patients who have taken them in controlled clinical trials. Most studies included outpatients between 18 and 65 years of age with MDD of at least moderate severity (Hamilton Depression [HAMD] Scale 17-item >17). These studies show that there is a wide range of doses effective in the treatment of moderate or severe depression.

Meta-analyses of studies that compared the SSRIs have found them to show equal effectiveness in MDD. In double-blind trials, about one-third of patients in short-term (6- to 8-week) placebo-controlled trials achieved full remission as determined by rating scales. About one-third had a partial response, and one-third had no response.

A recent network meta-analysis studied the efficacy of 21 medications for depression. All these medications were found to be more effective than placebo, with mostly modest effect sizes for responses. SSRIs had similar efficacy and acceptability compared to newer medications for depression.

One meta-analysis of SSRIs showed the drugs had little benefit over placebo in mild and moderate depression but clinically significant benefits over placebo in patients with more severe depression. On the other hand, there have been numerous studies demonstrating the effectiveness of SSRIs in the treatment of dysthymia and other forms of chronic mild depression.

An FDA systematic review of all SSRI maintenance trials submitted to the agency between 1985 and 2012 found that maintenance treatment reduced the risk of relapse by 52% compared to patients treated with a placebo.

Only a few predictors of response to SSRIs in the context of depression have been identified. A study showed that patients with MDD who have more anhedonia at baseline may have a poorer response to escitalopram monotherapy and may benefit more from aripiprazole augmentation versus those with less anhedonia at baseline. Also, evidence indicates that a serotonin transporter polymorphism may be associated with the antidepressant response to SSRIs and with the likelihood of side effects. Evidence that MDD patients who carry a long allele show better outcomes compared to those homozygous for the short allele has been reported, although some controversy remains.

Generalized Anxiety Disorder

Studies of SSRIs in the treatment of GAD have shown the effectiveness of these agents to be greater than the effectiveness of a placebo in treating this condition. SSRIs have also been demonstrated to be effective in preventing relapse after the successful treatment of GAD. There may be continuing improvement over a 6-month period in those receiving SSRIs, suggesting the usefulness of ongoing treatment.

Obsessive-Compulsive Disorder

Placebo-controlled studies have demonstrated SSRIs to be efficacious for the treatment of OCD in adults and children. In some but not most cases, doses higher than the maximum recommended for depression are needed. Full response to treatment also takes longer than for depressive disorders.

Panic Disorder

Overall, SSRIs usually decrease the frequency of panic attacks by about 80% and decrease associated anticipatory anxiety. The response rate may be lower for patients with more severe symptoms. Previous unsuccessful treatment with GABA-positive allosteric modulators (PAMs), formerly denoted as benzodiazepines, does not diminish the efficacy of SSRIs. When introduced, the daily doses of SSRIs should be half the usual starting doses used for the treatment of MDD. If not, the frequency of panic attacks may initially increase before abating, thereby possibly leading to early dropout from treatment. When stopping SSRIs in patients with panic disorder, a slower decrease in dose is recommended, as they may be more sensitive to discontinuation symptoms than patients with other disorders.

Social Anxiety Disorder (SAD)

SSRIs have been successfully used for the treatment of SAD. Improvement may be achieved slowly, becoming evident during the second month of treatment and continuing to increase until the third month. The combination of SSRIs and cognitive-behavior therapy (CBT) is more rapid and more complete than the use of an SSRI alone, as can be the case in other disorders.

Premenstrual Dysphoric Disorder (PMDD)

SSRIs are effective in alleviating the symptoms of PMDD. Improvement may be seen during the first week of treatment. Mood, irritability, and anxiety improve. Physical symptoms, such as swelling, bloating, and breast tenderness, are less responsive to treatment. Intermittent treatment (i.e., for the last 2 weeks of the menstrual cycle) can be as efficacious as continuous administration. Interestingly, this cyclical treatment does not produce SSRI discontinuation symptoms.

Post-Traumatic Stress Disorder

Placebo-controlled studies have confirmed the efficacy of SSRIs for PTSD. Discontinuation of the successful treatment after 6 months may result in the return of the PTSD symptoms in about half of patients.

Bulimia Nervosa

Fluoxetine (Prozac) 60 mg per day is FDA-approved for the treatment of bulimia nervosa. Trials have shown that fluoxetine is effective in decreasing episodes of binge eating and vomiting. No SSRI is FDA-approved for the treatment of anorexia nervosa.

Bipolar Depression

Fluoxetine, in combination with olanzapine (Symbyax), is FDA approved for the treatment of depression associated with bipolar I disorder.

Other Uses

SSRIs have been used off-label for several indications, including premature ejaculation. In this case, paroxetine may be the most useful of these drugs. However, continuous daily treatment is required to delay ejaculation significantly. SSRIs have also been used for the treatment of neuropsychiatric symptoms, agitation in neurocognitive disorders, and for the treatment of paraphilic disorders with inconsistent results.

PHARMACODYNAMICS

All SSRIs bind to and inhibit the action of 5-HT transporters. This raises 5-HT levels in the central nervous system (CNS) and throughout the body. The SSRIs act not only on the 5-HT transporter on 5-HT neurons but also at other sites, some well characterized, others not as clearly delineated (see below for individual medications). It is not clear whether the degree of SSRI selectivity offers any therapeutic benefit. One must be cautious, however, in extrapolating group effects to those of an individual patient, given that clinical experience reveals that patients who fail to respond to one may respond to another of these drugs. However, the equivalency of doses at the 5-HT transporter, drug compliance, and late responses should be taken into consideration when observing such differential responsiveness.

Medications other than the SSRIs that demonstrate significant inhibition of the 5-HT transporter include clomipramine (Anafranil), duloxetine

(Cymbalta), milnacipran (Savella, available in the United States only for the treatment of fibromyalgia), levomilnacipran (Fetzima), venlafaxine (Effexor), desvenlafaxine (Pristiq), vilazodone (Viibryd), vortioxetine (Trintellix), and trazodone (Desyrel). Most SSRIs are essentially devoid of any clinically significant degree of receptor antagonism properties at muscarinic, cholinergic, H_1, α_1, α_2, β-adrenergic, 5-HT_1, and 5-HT_2 receptors. Paroxetine has the greatest affinity for muscarinic cholinergic receptors of the available SSRIs and therefore produces, in a small percentage of patients, dry mouth, constipation, blurred vision, or urinary hesitancy. However, the affinity of paroxetine for muscarinic receptors as measured by inhibition of [^{3}H]quinuclidinyl benzilate binding. It is 15-fold less potent than that of amitriptyline, and its anticholinergic properties. This may thus explain its lack of side effects such as diarrhea and loose stools observed with other SSRIs, as well as its occasional sedating properties. Sertraline and fluvoxamine act as sigma-1 receptor agonists. The clinical relevance of this property remains to be fully understood, but some evidence suggests that it may contribute to therapeutic benefits in psychotic depression.

Positron emission tomography (PET) studies of the effect of SSRIs have been used for in vivo confirmation of in vitro and ex vivo derived hypotheses of mechanisms of action as well as for clinical dose findings. For example, PET imaging assessing occupancy of 5-HT reuptake sites by citalopram, fluoxetine, sertraline, paroxetine, or extended-release venlafaxine at minimum therapeutic doses revealed about 80% occupancy of 5-HT reuptake sites in the striatum. Similar quantitative results have been shown for the cingulate cortex. Higher regimens of SSRIs lead to greater occupancy of 5-HT transporters in the striatum. Notably, lower occupancy with a minimal effective dose of escitalopram has been reported in the orbitofrontal cortex, a brain region more intimately involved in OCD than for MDD. The latter observations help us understand why higher doses are commonly necessary for the treatment of OCD than depression. Taken together, these findings suggest that a threshold of 5-HT transporter occupancy is necessary to increase 5-HT transmission. That an increase in 5-HT transmission underlies the antidepressant action of SSRIs is indicated by the reappearance of depressive symptoms in patients who had responded to an SSRI when subjected to a tryptophan depleting paradigm (to decrease 5-HT synthesis). Further more, rapid return to their prior state with tryptophan administration (tryptophan is the amino acid precursor of 5-HT) confirmed the role of 5-HT in the therapeutic action of SSRIs.

Basis for the Delayed Onset of Therapeutic Effects of SSRIs

SSRIs penetrate the brain within hours, but their onset of action is delayed by at least 2 weeks. This can be accounted for by the initial decrease in the firing rate of 5-HT neurons because 5-HT$_{1A}$ autoreceptors on their cell bodies become overactivated as a result of the inhibition of the activity of 5-HT transporters. Since the release of 5-HT is directly proportional to the firing rate of 5-HT neurons, there is no sustained initial increase in 5-HT release in postsynaptic areas even though 5-HT reuptake is blocked. After prolonged SSRI administration, the 5-HT$_{1A}$ autoreceptors desensitize and allow recovery of the firing rate of 5-HT neurons in the presence of sustained 5-HT reuptake inhibition, at which time there is a net increase in 5-HT transmission. Also contributing to this enhanced 5-HT transmission is the desensitization of 5-HT$_{1B/D}$ autoreceptors on 5-HT terminals. The delay for the autoreceptors to desensitize (14 to 21 days) is consistent with the protracted onset of action of SSRIs in MDD. Importantly, the terminal autoreceptors in the orbitofrontal cortex desensitized after eight but not three weeks of SSRI administration, which is consistent with the longer delay of action of SSRIs in OCD than in MDD. These observations were made in rodent brains and can be translated to humans because the 5-HT system is nearly identical in mammalians.

Each of the SSRIs is structurally unique (Fig. 33.5a.1–1), although escitalopram is the *S*-enantiomer of the racemic mixture that is citalopram, and there are differences in their selectivity of reuptake inhibition, drug–drug interactions, and pharmacokinetic profiles. Studies of the reuptake inhibitory potencies of the SSRIs in vitro have been performed. Although different studies show different values, in general, the affinity (K_i) values are of the same relative ordering across studies. When interpreting in vitro data and attempting to extrapolate findings regarding relative degrees of reuptake inhibition to in vivo predictions, pharmacokinetic data such as steady-state concentrations, protein binding, molecular weights, and the presence or absence of active metabolites must all be accounted for.

TOLERABILITY

Both the therapeutic and adverse event profiles of SSRIs are the result of their blockade of the 5-HT transporter function. They eventually raise 5-HT levels in the CNS, as described above, but not necessarily to the same extent throughout the body, with a rapid action being particularly evident in the gastrointestinal (GI) tract. This is because 90% of the 5-HT in the body is in the periphery, where the degree of regulation is not as tightly controlled by autoreceptors. As a group, the SSRIs possess the adverse effects listed in Table 33.5a.1–2.

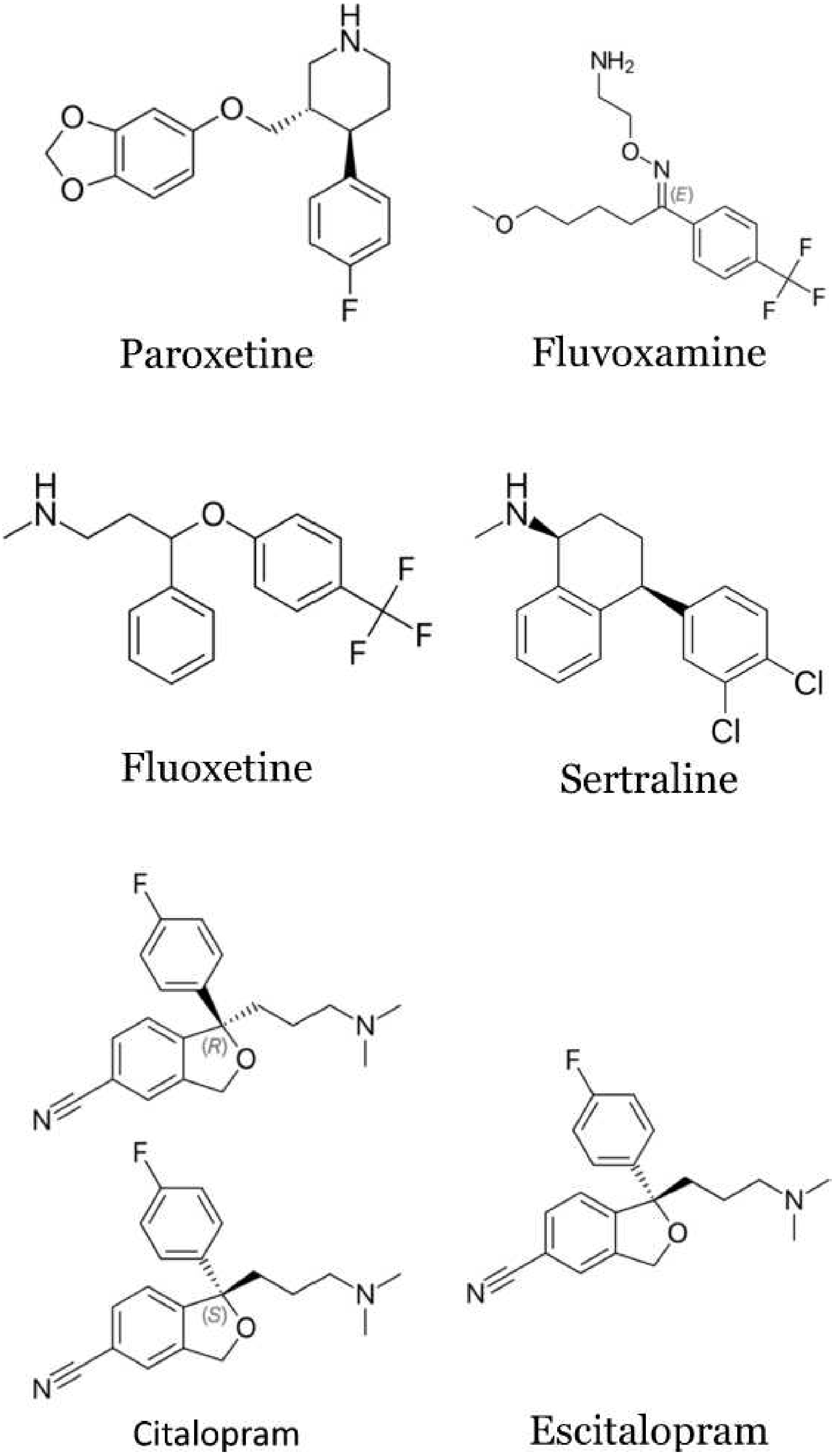
H
N
O
O
O
F
NH2
O
N
(E)
O
F
F
F
H
N
O
F
F
F
H
N
Cl
Cl
F
(R)
O
N
N
F
(S)
O
N
N
F
O
N
N

Paroxetine

Fluvoxamine

Fluoxetine

Sertraline

Citalopram

Escitalopram

FIGURE 33.5a.1–1. Structures of selective serotonin reuptake inhibitors.

When reviewing product information for the SSRIs, the frequency, severity, and duration of side effects need to be considered.

Gastrointestinal Effects

Nausea occurs in at least 25% of patients and can range from mild discomfort to, in some patients, vomiting. It is the most common early side effect of SSRIs and most often diminishes during the first to 2 weeks of treatment. Nausea at the initiation of treatment is generally dose-related. Accordingly, typical starting doses that are used to reduce this side effect are half of the usual starting dose. After 1 to 2 weeks at the lower doses, or when GI side effects resolve, the dose can be optimized. The disappearance of the GI side effects may be due to a desensitization of 5-HT_3 receptors. This is supported by the observations that GI side effects can be blocked by 5-HT_3 antagonists such as ondansetron (Zofran) and mirtazapine (Remeron).

Table 33.5a.1–2.
Selected Treatment-Emergent Adverse Events

- **Autonomic nervous system disorders**
  - Dry mouth
  - Increased sweating
- **Central and peripheral nervous system disorders**
  - Dizziness
  - Emotional blunting
  - Mild forgetfulness
- **Gastrointestinal disorders**
  - Nausea
  - Diarrhea
  - Constipation
  - Indigestion
  - Abdominal pain
- **Psychiatric disorders**
  - Insomnia
  - Somnolence
  - Decreased appetite
  - Decreased libido
- **Respiratory system disorders**
  - Rhinitis
  - Sinusitis
- **Urogenital**
  - Ejaculation disorder
  - Impotence
  - Anorgasmia
- **General**
  - Influenza-like symptoms
  - Fatigue
  - Yawning
  - Rash

Sexual Dysfunction

Sexual dysfunction, such as delayed ejaculation, delayed orgasm or anorgasmia, and diminished libido, is a common side effect of 5-HT reuptake inhibition. SSRI-associated sexual dysfunction is the most frequent side-effect–associated cause of premature medication discontinuation. Package insert data, almost always derived from phase III clinical trials, have underestimated the true incidence of sexual dysfunction, which occurs in clinical practice in about half of patients. Reliable estimates of the incidence and severity of untoward experiences involving sexual desire, performance, and satisfaction are difficult to obtain in part because patients and physicians may be reluctant to discuss them. Accordingly, estimates of the incidence of untoward sexual experience and performance cited in product labeling are underestimated in their actual incidence, as revealed by questionnaires on sexual function used in subsequent studies.

In some but not most instances, sexual dysfunction appears to be a dose-related effect, and occasionally decreasing the dose might be helpful. There have been a number of other strategies suggested to deal with this important adverse effect, including drug holidays for shorter half-life SSRIs (potentially associated with discontinuation symptoms or even a relapse) or the addition of buspirone, amantadine, yohimbine, cyproheptadine, stimulants, or bupropion. Anecdotally, selective inhibitors of cGMP-specific phosphodiesterase type 5, such as sildenafil (Viagra), have been reported by some patients to be helpful. A randomized controlled trial that used sildenafil as a treatment of sexual dysfunction in women taking SSRIs found that its use was associated with a reduction in sexual side effects. These observations lack solid replication, and, in fact, none of these approaches have been successful in completely reversing sexual side effects. Therefore, gradual switching to a medication not associated with sexual dysfunctions is often the most successful strategy. These include bupropion, mirtazapine, low-dose vortioxetine, vilazodone, or moclobemide (after an SSRI washout).

Sweating

Excessive sweating is not uncommon while being treated with SSRIs. The sweating is unrelated to ambient temperature. Nocturnal sweating may rarely drench bed sheets and require a change of night clothes. Terazosin, 1 or 2 mg per day, is often effective in counteracting sweating.

Discontinuation Syndrome

5-HT reuptake inhibitors have the potential to produce a discontinuation syndrome if stopped abruptly. However, it is important to emphasize that individuals treated with SSRIs do not develop addictions to these drugs. These drugs do not cause cravings, and individuals do not feel the need to increase the dose to get the same effect. Longer durations of treatment appear to be associated with a higher probability of discontinuation syndrome than shorter treatment periods (i.e., intermittent treatment in PMDD). In general, good clinical practice dictates a gradual tapering down of a therapeutic regimen after successful treatment.

Typical symptoms include asthenia, anxiety, agitation, GI distress, myalgias, sometimes a flu-like syndrome but with the absence of a fever, disequilibrium, and, in extreme cases, an experience described as electrical shocks coursing through the arms and legs. When discontinuing the medication, it should be done over the course of weeks rather than days. Drug taper can be problematic when the last dose is stopped. This problem can be managed by prescribing fluoxetine 20 mg per day for 1 to 2 weeks and then stopping that medication. The longer half-lives of fluoxetine and norfluoxetine help counter the discontinuation effects to a significant degree.

Bleeding

Both case reports and epidemiologic studies confirm that SSRIs alone may increase bleeding risk in some patients. There have been numerous case reports and case series describing abnormal bruising, bleeding, and heavy menses in patients taking SSRIs. Serotonin is released from platelets following vascular injury. It promotes vasoconstriction and a change in the shape of the platelets that lead to aggregation. Platelets cannot themselves synthesize 5-HT, so SSRIs deplete 5-HT in the platelets, which may decrease coagulation, theoretically leading to excessive bleeding diathesis in some patients. Other physiologic mechanisms may also be involved. There is an additive increase in bleeding risk when SSRIs are used concurrently with other drugs that can cause bleeding. SSRIs, for example, increase the incidence of serious bleeding in patients taking warfarin. Risk is also increased with concurrent use of nonsteroidal anti-inflammatory drugs, aspirin, or clopidogrel. A large general practice database search of data involving nearly 400,000 patients found that the number needed to harm (NNH) for upper GI bleeding with SSRI treatment in a low-risk population was 3,177, and in a high-risk population, it was 881. Thus, this risk associated with SSRI use is modest, requires monitoring and should not preclude their use in the presence of another predisposing factor.

Hyponatremia

There is an increased risk of hyponatremia in patients taking SSRIs. Risk factors include older age, female gender, volume depletion, and use of other medications or medical conditions causing hyponatremia like diuretics. SSRI-induced hyponatremia typically begins during the first weeks of treatment. This may be secondary to the syndrome of inappropriate antidiuretic hormone.

Cardiovascular Effects

All SSRIs can potentially lengthen the QT interval in otherwise healthy people and cause drug-induced long QT syndrome, possibly leading to torsade de pointes when taken in overdoses. In 2011, the FDA issued a safety warning about high-dose citalopram treatment. It made the following recommendations regarding citalopram use:

- 20 mg/day is the maximum recommended dose for patients with hepatic impairment, who are older than 60 years of age, who are CYP2C19 poor metabolizers, or who are taking concomitant cimetidine.
- No longer prescribe at doses greater than 40 mg/day.
- Do not use it in patients with congenital long QT syndrome.
- Correct hypokalemia and hypomagnesemia before administering citalopram.
- Monitor electrolytes as clinically indicated.
- Consider more frequent EKGs in patients with chronic heart failure, bradyarrhythmias, or patients on concomitant medications that prolong the QT interval.

This was based on a QT study assessing the effects of 20- and 60-mg doses of citalopram on the QT interval in adults compared with placebo. The study found a maximum mean prolongation in the individually corrected QT intervals of 8.5 milliseconds (ms) for 20 mg citalopram and 18.5 ms for 60 mg. For 40 mg, the estimated prolongation of the corrected QT interval was 12.6 ms.

Regulatory bodies in the United Kingdom and in Canada issued a similar safety warning about high-dose escitalopram treatment, with a maximal recommended dose of 10 mg/day in some populations considered at higher cardiac risk.

Antepartum use of SSRIs is sometimes associated with QTc interval prolongation in exposed neonates. A comparative study matched 52 newborns exposed to SSRIs in the immediate antepartum period with 52 controls. The mean QTc was significantly longer in the neonates exposed to SSRIs compared with the control group (409 ± 42 vs. 392 ± 29 ms; $p = .02$). Five neonates (10%) exposed to SSRIs had a prolonged QTc interval >460 ms compared with none in the controls. The longest QTc interval among exposed neonates was 543 ms. All the repolarization abnormalities associated with SSRIs normalized in subsequent electrocardiographic tracings, and no serious adverse effects occurred.

Comments

A 2013 study done at the Ann Arbor VAMC and the University of Michigan looked at the records of more than 600,000 VHA patients who received citalopram prescriptions between 2004 and 2009. This study did not show an increased risk for arrhythmias or death associated with daily doses of more than 40 mg. Citalopram daily doses greater than 40 mg were actually associated with slightly lower risks of all-cause mortality, noncardiac mortality, and ventricular arrhythmia compared with daily doses ≤20 mg. A subsequent study from the University of Minnesota in 35,848 patients whose regimens of citalopram were or were not reduced to 40 mg/day or less from a mean daily dose of 62 mg showed that reduction to the new safety limit was associated with a significantly higher rate of all-cause hospitalizations or death, as were hospitalizations for depression. Taken together, these findings raise questions about the legitimacy of the FDA warning.

Weight Gain

During early treatment, some patients may lose weight. Over time, many patients regain this weight, and it is not uncommon to see a significant increase in weight over the following months. The exact mechanism for this problem is unclear, but one should consider that appetite may increase as depression and/or anxiety abate. In some instances, it becomes difficult for patients to reverse this weight gain, even with rigorous dietary measures and exercise.

Diabetes

Population-based studies have shown that SSRIs may increase the risk of diabetes. An analysis was done based on three large well-known US datasets: the Nurses' Health Studies I and II and the Health Professionals Follow-up Study, a total of 139,000 women and 30,000 men. During over 1.6 million person-years of follow-up, the analysis documented a total of 6,641 new cases of type 2 diabetes and found that self-reported use of medications for depression was associated with an approximately 10% to 30% excess risk of developing diabetes. This association was robust despite adjustment for age, ethnicity, marital and living status, smoking habits, alcohol intake, physical activity, family history of diabetes, diet, body mass index, and major comorbidities (e.g., hypertension and dyslipidemia) and their treatment. There needs to be additional evidence that this increased risk is due to associated factors or SSRI treatment, but these findings suggest that it is prudent to monitor patients periodically for increased glycosylated hemoglobin (HbA1C).

Sleep Disturbances and Daytime Somnolence

Up to one-fourth of persons taking SSRIs note trouble sleeping, excessive somnolence, or overwhelming fatigue. Bruxism is another possible side effect. Nocturnal motor activity, including nocturnal myoclonus, may be increased during SSRI treatment. Some persons taking SSRIs report recalling extremely vivid dreams or nightmares. They describe sleep as "busy." Some persons benefit from taking their SSRI dose before going to bed, but others prefer to take it in the morning. SSRI-induced insomnia can be treated with Z drugs, trazodone, or other sedating medicines. Significant SSRI-induced somnolence often requires switching to another SSRI or bupropion.

Emotional Blunting

Emotional blunting or indifference are largely overlooked side effects associated with long-term SSRI use. Patients report an inability to cry in response to emotional situations, a feeling of apathy or indifference, or a restriction in the intensity of emotional experiences. This side effect is related to the inhibitory effect of 5-HT on the firing of dopamine and noradrenaline neurons. This side effect often leads to treatment discontinuation even when the drugs provide relief from depression or anxiety. The frequency of this side effect is unknown. It has a delayed onset, appears to be related to dosing, and may resolve with a dose reduction.

Extrapyramidal Symptoms

SSRIs rarely cause akathisia, dystonia, tremor, cogwheel rigidity, torticollis, opisthotonos, gait disorders, and bradykinesia. Rare cases of tardive dyskinesia have also been reported. Some people with well-controlled Parkinson disease may experience acute worsening of their motor symptoms when they take SSRIs, as 5-HT exerts an inhibitory effect on the firing of dopamine neurons.

Manic Switch

A switch in mood polarity has been reported following the initiation of SSRI in unipolar depression, although this adverse effect appears rare. Product monographs recommend screening for personal and family history of bipolar affective disorder before introducing an SSRI monotherapy. Except for the combination of fluoxetine with olanzapine, SSRIs are not FDA-approved for the treatment of bipolar depression.

Fractures

A recent meta-analysis found an increased risk of fracture among older adults treated with SSRIs (relative risk = 1.67, 95% confidence interval [CI] = 1.56–1.79). The authors reported an NNH of 19 for one additional fracture to occur over 5 years in SSRI-treated versus non-SSRI–treated individuals.

Laboratory Interferences

SSRIs do not interfere with any laboratory measures.

CHEMISTRY AND PHARMACOKINETICS OF INDIVIDUAL SELECTIVE SEROTONIN REUPTAKE INHIBITORS

The SSRIs as a class exhibit a broad range of chemical structures and serum half-lives ranging from hours to days. There is no evidence that the structure of the drug molecules differentiates the clinical benefits and risks of the drugs in any meaningful way. The wide range of serum half-lives does manifest clinically. This is true not only in terms of time to steady state and the time required for the drug to exit the body after it is stopped but also in differing potentials for discontinuation syndrome on stopping the SSRI or, more commonly, after missing occasional doses. The SSRIs demonstrate moderate to high protein binding. Table 33.5a.1–3 summarizes the pharmacokinetic properties of the SSRIs. The contribution of protein binding to potential drug–drug interactions is unclear. The only fraction of the drug that is available to exert biologic activity is the free drug in the plasma. The amount of free drug can enter the CNS but is also subject to being metabolized. SSRIs, such as paroxetine, sertraline, and fluoxetine, are tightly protein-bound, and only approximately 1/20 of the drug in the system exists as a free drug, the balance being nonspecifically bound to serum proteins. It remains unknown whether clinically significant displacement of protein binding by the SSRIs occurs during concomitant therapy with other medications, such as warfarin (Coumadin). It can be argued that clinicians should actively look for such drug–drug interactions following the addition of one drug with high protein-binding characteristics to another, especially when functional assays (such as with warfarin) are available. Another area that requires exploration is the determination of the level at which protein binding becomes saturated, and the accumulation of free drug becomes a nonlinear process. This may have important implications for agents that demonstrate different mechanisms of action across low-to-high dosing ranges. The chemical structures of the individual SSRIs are illustrated in Figure 33.5a.1–1, and their pharmacokinetic profiles are shown in Table 33.5a.1–3.

Table 33.5a.1–3.
Pharmacokinetic Profiles of the Selective Serotonin Reuptake Inhibitors

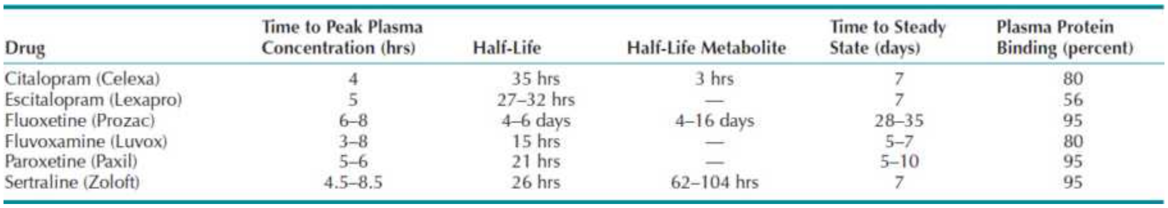

Drug	Time to Peak Plasma Concentration (hrs)	Half-Life	Half-Life Metabolite	Time to Steady State (days)	Plasma Protein Binding (percent)
Citalopram (Celexa)	4	35 hrs	3 hrs	7	80
Escitalopram (Lexapro)	5	27–32 hrs	—	7	56
Fluoxetine (Prozac)	6–8	4–6 days	4–16 days	28–35	95
Fluvoxamine (Luvox)	3–8	15 hrs	—	5–7	80
Paroxetine (Paxil)	5–6	21 hrs	—	5–10	95
Sertraline (Zoloft)	4.5–8.5	26 hrs	62–104 hrs	7	95

FLUOXETINE

Chemistry

Fluoxetine's empirical formula is $C_{17}H_{18}F_3NO \cdot HCl$, and it has a molecular weight of 345.79. Its chemical designation is (±)-*N*-methyl-3-phenyl-3-[(α–α–α-trifluoro-*p*-toylyl)-oxy]-propylamine hydrochloride.

Pharmacologic Actions

Absorption. Fluoxetine is well absorbed orally, with 72% to 90% systemic availability, reflecting the combination of almost complete absorption and first-pass metabolism by liver hepatocytes. Maximum plasma concentration is achieved within 6 to 8 hours of ingestion after a single 40-mg dose. Peak plasma concentrations are not affected by food, but absorption may be delayed 1 to 2 hours. The delayed-release, enteric-coated preparation of once-weekly fluoxetine also demonstrates peak plasma levels 6 to 8 hours after absorption, although absorption does not begin until entry into the portion of the GI tract where the pH exceeds 5.5. There is more variability of peak and trough levels with the once-weekly preparation compared to once-daily dosing. In addition, the C_{max} after administration of the 90-mg, once-weekly preparation one day after 20-mg, once-daily dosing was 1.7 times higher than the C_{max} observed with once-daily dosing. When the transition to once-weekly dosing was delayed by 1 week after the last once-daily dose was given, the observed C_{max} was approximately equal. Fluoxetine is available as pulvules (10, 20, or 40 mg), tablets (10 mg), an oral concentrate (20 mg/5 mL), and in an enteric-coated, delayed-release, once-weekly form (90 mg).

Distribution. In vitro, fluoxetine is approximately 95% bound to serum proteins (albumin and α_1–acid-glycoproteins) at concentrations of 200 to 1,000 ng/mL. The volume of distribution for fluoxetine and norfluoxetine, the biologically active metabolite, ranges from 20 to 45 L/kg. The rate of plasma clearance is 20 and 9 L/hr for fluoxetine and norfluoxetine, respectively. Fluoxetine is widely distributed, including excretion into breast milk.

Metabolism and Elimination. Fluoxetine is a racemic mixture of *R*-fluoxetine and *S*-fluoxetine in equal proportions. Both are approximately equipotent in 5-HT reuptake inhibition activity, although the *S*-fluoxetine enantiomer is more slowly eliminated and is, therefore, the predominant form in plasma at a steady state. Fluoxetine undergoes demethylation in the liver to norfluoxetine and to a number of other unidentified minor metabolites, although norfluoxetine is the only metabolite identified that possesses biologic activity. In animals, the *S*-enantiomer of norfluoxetine is essentially equivalent in activity to that of *R*- or *S*-fluoxetine, but *R*-norfluoxetine has significantly less activity. The kidneys excrete inactive metabolites produced by hepatic metabolism.

The elimination half-life of fluoxetine is 1 to 3 days after acute administration and 4 to 6 days after chronic administration. Norfluoxetine has an elimination half-life of 4 to 16 days, independent of the duration of administration. Steady-state levels of fluoxetine are higher than would be predicted based on single-dose administration, as the drug inhibits its own metabolism.

Individuals who are deficient in CYP 2D6 activity—so-called poor metabolizers—constitute 6% to 8% of the Caucasian population and metabolize *S*-fluoxetine at a slower-than-expected rate. In contrast, the metabolism of *R*-fluoxetine is normal. Owing to the slower metabolism of *S*-fluoxetine, it is present at higher-than-expected concentrations, but *S*-norfluoxetine is found at lower concentrations. The sum total of *S*-fluoxetine and *R*-fluoxetine and *S*-norfluoxetine and *R*-norfluoxetine are approximately equal between poor metabolizers and those with a full complement of CYP 2D6 activity. The percentage of individuals with ultra-rapid status at the 2D6 varies geographically in the world: between about 1% (Scandinavia) to 20% in northeast Africa and Arabian Peninsula. A major dose adjustment of fluoxetine is likely not necessary as it inhibits its own metabolism, and these individuals would theoretically be transformed into slow metabolizers.

Impairment of hepatic function is associated with impaired metabolism of fluoxetine. In a series of subjects with cirrhosis of the liver, the mean half-life of fluoxetine was increased to 7.6 days, compared to 2 to 3 days for healthy subjects. The half-life of norfluoxetine was similarly increased to a mean of 12 days, compared to 7 to 9 days for healthy subjects. This would indicate the necessity of a dose reduction in individuals with significant liver disease.

Renal disease does not appear to impair the metabolism of fluoxetine or norfluoxetine in single-dose studies. One small study (*N* = 12) of patients on dialysis receiving fluoxetine at 20 mg per day for 2 months found steady-state fluoxetine and norfluoxetine levels close to those observed in patients with intact renal function.

Patients 60 years of age and older who were depressed demonstrated normal fluoxetine levels of 209 ± 86 ng/mL after dosing 20 mg per day for 6 weeks.

Children with OCD or MDD (six to less than 13 years of age; *N* = 10), compared to adolescents (13 to less than 18 years of age; *N* = 11), receiving fluoxetine at doses of 20 mg per day showed steady-state (60 days) fluoxetine levels of 171 and 86 ng/mL, respectively. Norfluoxetine levels were also higher in children than in adolescents, 195 and 113 ng/mL, respectively. These differences appear to be primarily related to relative body mass differences and the levels normalized when this was considered.

SERTRALINE

Chemistry

The empirical formula of sertraline is $C_{17}H_{17}NCl_2 \cdot HCl$; it has a molecular weight of 342.7. The structure of sertraline is distinct from that of other 5-HT reuptake inhibitors, and it is a naphthylamine derivative. The chemical designation is (1*S*-*cis*)-4-(3,4-dichlorophenyl)-1,2,3,4-tetrahydro-*N*-methyl-1-naphthalenamine hydrochloride.

Pharmacologic Actions

Absorption. Sertraline is well absorbed orally, with equivalent absorption observed with tablets and oral concentrate. Maximum systemic concentration is achieved within 4.5 to 8.4 hours of ingestion after 14 days of dosing in the range of 50 to 200 mg daily. When administered with food, the mean peak plasma concentration is increased by 25% with the tablet preparation, and the peak plasma concentration is achieved in 5.5 hours after ingestion, compared to 8 hours when given without food. Whether this food-related increase in sertraline plasma levels is clinically significant in terms of adverse events is unclear, although many clinicians advocate the initial administration of sertraline, like other SSRIs, after a meal. Administration of the oral concentrate preparation of sertraline with food is associated with a delay in the time to peak plasma concentration of from 5.9 to 7.0 hours.

Distribution. In vitro, sertraline is approximately 98% bound to serum proteins at concentrations of 20 to 500 ng/mL. The volume of distribution for sertraline is approximately 20 L/kg, and the rate of plasma clearance is 96 L/hr. The extent of the distribution of sertraline in breast milk has been studied in detail. The data indicate that sertraline is present in breast milk and in infant serum, albeit in much lower concentrations.

Metabolism and Elimination. Sertraline undergoes demethylation in the liver to *N*-desmethylsertraline. The half-life of sertraline is 26 hours, and the half-life of *N*-desmethylsertraline is 62 to 104 hours; however, the metabolite is 5 to 10 times less potent in inhibiting 5-HT reuptake than the parent compound, but still in the range of that of fluvoxamine. Desmethylsertraline displays a five- to ninefold increase in area under the curve (AUC) and peak and minimum plasma concentrations after 14 days of sertraline administration. In healthy volunteers, sertraline demonstrates linear pharmacokinetics over the range of 50 to 200 mg per day. Based on the use of radiolabeled sertraline in two healthy male subjects, unchanged sertraline was found to constitute less than 5% of the plasma radioactivity. After nine days, 40% to 45% of the radioactivity was recovered in urine, but no unchanged sertraline was detected. Another 40% to 45% was recovered in feces, of which 12% to 14% was unchanged sertraline.

Liver disease is associated with impaired metabolism of sertraline. In a series of subjects with mild, stable cirrhosis of the liver, the mean half-life of sertraline was increased to 52 hours, compared to 22 hours for normal subjects. This suggests the necessity of dose reduction or increased intervals between doses in individuals with hepatic impairment.

The presence of renal impairment (mild, moderate, or severe) does not appear to affect the multiple-dose pharmacokinetics of sertraline. This is perhaps due to renal excretion providing a relatively minor pathway for the elimination of the unchanged drug.

Aging is associated with a decrease in sertraline clearance. In 16 older adult patients treated with sertraline (100 mg per day for 14 days), the clearance of sertraline was 40% lower than that observed in younger subjects (25 to 32 years of age). The establishment of steady-state drug levels is delayed in this group to a time frame of 14 to 21 days of medication administration. Desmethylsertraline clearance was decreased in older men but not women.

The pharmacokinetics of sertraline in children (6 to 12 years of age) and adolescents (13 to 17 years of age) are similar to those seen in adults when adjustments are made based on body mass. Pediatric patients with MDD or OCD who were administered sertraline (42 days of sertraline therapy, of which at least the final 11 days were at a dose of 200 mg per day) had a mean C_{max} (maximal concentration) and half-lives of 165 ng/mL and 26 hours and 123 ng/mL and 28 hours for children and adolescents, respectively. In comparison, a separate study involving healthy adults who received sertraline (200 mg per day for 30 days) had a mean C_{max} of 142 ng/mL and a half-life of 27 hours. Children and adolescents have a smaller (by approximately 22%) AUC with sertraline than adults, also suggesting the potential for more rapid metabolism in the younger age group. Even so, initial dosing strategies should typically be adjusted downward for younger patients, who typically have a lower body mass than adults.

PAROXETINE AND PAROXETINE CR

Chemistry

Paroxetine is the hydrochloride salt of a phenylpiperidine compound and is structurally distinct from other medications for depression. Paroxetine is designated as (–)-trans-4*R*-(4′-fluorophenyl)-3*S*-([3′,4′-methylenedioxyphenoxy]methyl) piperidine hydrochloride hemihydrate. The empirical formula is $C_{19}H_{20}FNO_3 \cdot HCl \cdot \frac{1}{2}H_2O$, and the molecular weight of the free base is 374.8. Paroxetine constant release (CR) is the same salt as the immediate-release (IR) formulation but is in a degradable polymeric matrix with an enteric coating that delays drug release until the tablet has left the stomach.

Pharmacologic Actions

Absorption. Paroxetine is well absorbed orally, and food does not alter absorption. Maximum systemic concentration is achieved approximately 5 hours and 6 to 10 hours after ingestion for the IR and CR formulations, respectively. The CR formulation demonstrates a decrease in bioavailability of approximately 25% compared to the IR preparation, and thus equivalent doses of paroxetine IR and CR would be 10.0- and 12.5-mg, 20.0- and 25.0-mg, and 30.0- and 37.5-mg tablets, respectively.

Distribution. Paroxetine, like sertraline and fluoxetine, is tightly bound to serum proteins in vitro, with binding of 95% and 93% at plasma levels of 100 and 400 ng/mL, respectively. Paroxetine has not been reported to alter the in vitro protein binding of warfarin or phenytoin (Dilantin). The volume of distribution is approximately 15 L/kg, and the rate of plasma clearance is close to 79 L/hr. Most individuals achieve steady-state plasma concentrations within 5 to 10 days of dosing of the IR formulation and within 2 weeks with paroxetine CR.

Metabolism and Elimination. The principal path of metabolism for paroxetine is oxidation and methylation. A number of major metabolites have been identified, although none appears to be any greater than 1/50 as potent as the parent compound in the in vitro reuptake inhibition of 5-HT. A single test dose of paroxetine (30 mg) was recovered as 64% in the urine (2% as the parent compound and 62% as metabolites over 10 days). In the feces, 36% was excreted, with less than 1% as the parent compound.

Paroxetine saturates the CYP 2D6 isozyme of the CYP system and therefore exhibits nonlinear pharmacokinetics. Steady-state plasma maximal and minimal concentrations were elevated 6 and 14 times, respectively, over single-dose values after reaching a steady state. This translates into a 24-hour AUC value close to eightfold higher at a steady state than would be predicted from a single dose. Pharmacokinetic data showed that the half-life of 2D6 substrates is prolonged in the presence of paroxetine in normal metabolizers but is not further prolonged in slow metabolizers. As for fluoxetine, a major increase in dose may not be necessary for ultra-rapid metabolizers.

In older adult patients, dosing of paroxetine at 20, 30, or 40 mg per day produced plasma concentrations approximately 70% to 80% greater than those observed in younger patients. Patients with a creatinine clearance of less than 30 mL/min demonstrated a fourfold elevation in plasma levels compared to healthy controls. Patients with moderate renal impairment (creatinine clearance of 30 to 60 mL/min), as well as patients with moderate hepatic impairment, had a twofold increase in plasma concentrations. All the previously mentioned groups should receive reduced starting doses of paroxetine, for example, 10 mg per day of paroxetine or 12.5 mg per day of paroxetine CR.

FLUVOXAMINE AND FLUVOXAMINE CR

Chemistry

Fluvoxamine is a member of the 2-aminoethyloxime ethers of the aralkylketone family. The empirical formula of fluvoxamine is $C_{15}H_{21}O_2N_2F_3 \bullet C_4H_4O_4$, and it has a molecular weight of 434.4. The chemical designation is 5-methoxy-4′-(trifluoromethyl)valerophenone-(E)-*O*-(2-aminoethyl)oxime maleate (1:1).

Pharmacologic Actions

Absorption. The bioavailability of fluvoxamine maleate is 53% by oral administration, and its absorption is essentially unaffected by food. At a steady state, maximal plasma concentrations are achieved 3 to 8 hours after dosing.

Distribution. In vitro assays demonstrate protein binding, predominantly albumin, of approximately 80% for fluvoxamine at concentrations of 20 to 2,000 ng/mL. The volume of distribution for fluvoxamine is approximately 25 L/kg. Breast milk concentration was reported to be about a third of the mothers’ plasma, thereby leading to low exposure to the babies.

Metabolism and Elimination. The primary route of metabolism for fluvoxamine is oxidative demethylation and deamination in the liver. The main human metabolite, as determined by radiotracer studies, is fluvoxamine acid, which, along with its *N*-acetylated analog, accounts for approximately 60% of urinary products. In this same assay, fluvoxethanol comprised approximately 10%. Fluvoxamine acid is 10- to 100-fold less potent in an in vitro assay of 5-HT reuptake inhibition than the parent compound.

Fluvoxamine demonstrates nonlinear pharmacokinetics. In 30 normal volunteers receiving fluvoxamine at 100, 200, or 300 mg per day over 10 days, the maximum plasma concentration was 88, 283, and 546 ng/mL, respectively. This represents increases of 160% and 207% for the 200-

and 300-mg doses, respectively, over what would have been predicted from the lowest dose. The mean plasma half-life at a steady state is 16 hours in young, healthy volunteers receiving 100 mg per day and 14 hours in those receiving 50 mg per day. After repeated dosing, the half-life increases to 17 to 22 hours. Steady-state is typically achieved in approximately 7 days.

Hepatic dysfunction is associated with an approximately 30% decrease in fluvoxamine clearance compared to healthy controls. In contrast, renal impairment (creatinine clearance of 5 to 45 mL/min) was not associated with decreased clearance.

Older adult patients show a prolongation in the half-life of fluvoxamine. Compared to younger patients, in the older adults, the half-life at 50 mg per day goes from 14 to 17 hours, and it goes from 16 to 26 hours for dosing of 100 mg per day.

Similarly to adults, children (6 to 11 years of age) and adolescents (12 to 17 years of age) display nonlinear, multiple-dose pharmacokinetics with fluvoxamine. Steady-state plasma concentrations of fluvoxamine were two- to threefold higher in children than in adolescents after administration of 100 mg twice a day. The AUC, C_{max}, and minimum drug concentration (C_{min}) were similarly elevated in children compared to adolescents. Female children, in a small sample ($N = 3$), displayed higher plasma levels than male children. Adolescents are similar to adults with regards to the pharmacokinetic profile of fluvoxamine and thus may require higher doses than those necessary for the treatment of children.

CITALOPRAM AND ESCITALOPRAM

Citalopram was a late addition to the U.S. market when it received approval for the treatment of MDD in 1998. Racemic citalopram is a 50–50% mixture of the S- and R- isomers. Escitalopram is the *S*-enantiomer of racemic citalopram. The S-enantiomer is the active component that blocks the 5-HT recapture. The R-citalopram binds to the allosteric site of the transporter and counteracts the therapeutic effects of the S-citalopram. Hence, doses of racemic citalopram more than twice those of S-citalopram are needed to produce similar therapeutic benefits.

Chemistry

Citalopram is the hydrobromide salt of a racemic mixture of a bicyclic phthalane derivative designated chemically as (±)–1-(3-dimethylaminopropyl)–1–(4-fluorophenyl)–1,3-dihydroisobenzofuran-5-carbonitrile, HBr. The molecular formula is $C_{20}H_{22}BrFN_2O$, and it has a molecular weight of 405.35.

Escitalopram is the oxalate salt of the *S*-enantiomer of racemic citalopram.

Pharmacologic Actions

Absorption. Citalopram is well absorbed orally, and absorption is unaffected by food. The bioavailability of oral citalopram (40 mg, single dose) is approximately 80% of that observed after intravenous (IV) administration. Maximum systemic concentration is achieved approximately four hours after ingestion. Peak serum levels of escitalopram are seen approximately 5 hours after ingestion.

Distribution. Citalopram has less protein binding than some of the other SSRIs, with citalopram and its two primary metabolites, demethylcitalopram (DCT) and didemethylcitalopram (DDCT), displaying in vitro protein binding of 80%. The volume of distribution of citalopram is 12 L/kg, and the clearance rate is 19.8 L/hr, of which approximately 20% is renal clearance. The half-life of citalopram is approximately 35 hours, and most patients should achieve steady-state concentrations in approximately 1 week.

Escitalopram has protein binding of approximately 56%, measured in vitro. The clearance rate of escitalopram is 36 L/hr, with approximately 7% due to renal clearance. The volume of distribution is not well characterized. The mean terminal half-life is approximately 27 to 32 hours, so steady-state concentrations should be achieved after approximately 7 days of dosing.

Metabolism and Elimination. The principal pathway of metabolism for citalopram is by *N*-demethylation to DCT, and the primary CYP isozymes involved are CYP 3A4 and CYP 2C19. Citalopram is the compound found in the highest concentrations in humans, with DCT and DDCT present at concentrations of approximately 50% and 10%, respectively, of that of citalopram. The metabolites DCT and DDCT possess only approximately one-eighth the potency of inhibition of the 5-HT transporter that citalopram does and likely contribute little, if at all, to the therapeutic activity. Two other compounds are products of citalopram metabolism in addition to DCT and DDCT; these are citalopram-*N*-oxide and a deaminated propionic acid derivative. Citalopram displays linear and dose-proportional pharmacokinetics for single and multiple doses in the dose range of 10 to 60 mg per day.

Citalopram pharmacokinetics have been studied in different population subgroups. Normal controls 60 years of age and older in a single-dose study exhibited increases in AUC and half-life by 30% and 50%, respectively. The same population had increases of 23% and 30%, respectively, thus, 20 mg is the recommended daily dose in the geriatric population.

Patients with reduced hepatic function have demonstrated a 37% decrease in clearance of orally administered citalopram and a doubling of half-life. The recommended dose for patients with hepatic impairment is 20 mg per day.

Mild to moderate renal impairment is associated with a decreased clearance rate of 17% compared to healthy controls. The effect of severe renal impairment on citalopram clearance is not well characterized.

Escitalopram is metabolized to *S*-demethylcitalopram (*S*-DCT) and *S*-didemethylcitalopram (*S*-DDCT). Similar to the metabolism of citalopram, escitalopram is the predominant compound found in human plasma, with *S*-DCT being present at approximately 30% of the concentration of the parent compound. The level of *S*-DDCT is typically below the limits of detection. Escitalopram inhibition of the 5-HT transporter in vitro is at least 7 and 27 times more potent than inhibition displayed by *S*-DCT and *S*-DDCT, respectively. The primary CYP isozymes thought to be involved in escitalopram metabolism are CYP 2C19 and, to a lesser extent, CYP 3A4. Escitalopram displays linear and dose-proportional pharmacokinetics for single and multiple doses in the dose range of 10 to 30 mg per day.

Examination of escitalopram clearance in subjects 65 years of age and older in single-dose and multiple-dose studies showed increases of approximately 50% in AUC and half-life. Thus, 10 mg is the recommended daily dose in the geriatric population. The effect of reduced renal or hepatic function on escitalopram clearance is poorly characterized, but dose adjustments are extrapolated from data available for the racemic compound citalopram.

DRUG INTERACTIONS

SSRIs can produce pharmacodynamic or pharmacokinetic interactions with other medications. Categories of potential pharmacokinetic interactions between the SSRIs and other agents are enzyme inhibition through the CYP isozyme system and protein binding. Fortunately, the majority of drug–drug interactions are not critical, but they may insidiously compromise the patient’s quality of life or overall tolerability of a therapeutic regimen. It is not simply an issue of whether drug–drug interactions are seen in a particular area of practice; it is often a question of whether they are recognized, given that they do occur on a regular basis. Ultimately, it is the subsequent pharmacodynamic consequence that dictates the importance of the pharmacokinetic interaction.

Interactions involving SSRIs are particularly relevant clinically because MDD often requires more than one step of treatment to elicit remission of symptoms. Frequently, a second medication is added to augment the first. A combination of an initial SSRI with several other agents, psychotropic and otherwise, is standard practice. The STAR*D study looked sequentially at the “next-step” benefits of switching medications or adding on to citalopram. The results found that among patients who do not have a remission during up to 14 weeks of an initial SSRI or who could not tolerate such therapy, switching to another SSRI, an SNRI, bupropion, or a combination of citalopram with another agent resulted in similar symptom remission within each step. Of note, clinical experience suggests that switching to another class after failing an SSRI trial is often more useful than switching to another SSRI. This is somewhat supported by a large meta-analysis that found modest benefits of switching classes, with a number needed to treat (NNT) of about 22 to achieve one additional remission compared to switching to another SSRI. Factors to consider when switching or combining treatments include efficacy, tolerability, the burden of side effects, interactions among drugs, dosing convenience, adherence, and cost. Nevertheless, paramount to any subsequent step following monotherapy is using an efficacious dose to ensure adequate target engagement. The question of whether to use combinations from treatment initiation in an attempt to achieve greater remission rates within a first trial still needs more documentation. Indeed, there have been both positive studies and a large negative trial, but the latter may have been hampered by suboptimal dosing in a single-blind design. A recent meta-analysis actually showed that combining a monoamine reuptake inhibitor with a presynaptic α_2-antagonist (such as mirtazapine) from treatment initiation was associated with better efficacy outcomes compared to monotherapy (Standardized mean difference = 0.64; 95% CI = 0.12–1.15) and did not result in more dropouts due to side effects.

Pharmacodynamic Interactions

The most dreaded outcome of pharmacokinetic interactions involving SSRIs is the 5-HT syndrome or 5-HT toxicity. Serotonin syndrome is a life-threatening syndrome that can lead to autonomic instability, abdominal pain, diarrhea, confusion or delirium, increased motor activity, myoclonus, hyperpyrexia, cardiovascular shock, or death.

Serotonin-related problems occur on a continuum. For instance, mild elevation of the synaptic 5-HT at a therapeutic dose of an SSRI, is associated with common serotoninergic side effects such as GI discomfort or nausea. A real 5-HT syndrome is rare and occurs when dramatic levels of 5-HT are in the synapse. This syndrome is mostly mediated by the postsynaptic 5-HT_{2A} and 5-HT_{1A} receptors. 5-HT levels in the synapse increase significantly via three major mechanisms most often involved in 5-HT toxicity: 5-HT reuptake blockade, MAO inhibition, and increased 5-HT release from presynaptic neurons. A 5-HT syndrome is expected when a combination of these mechanisms occurs. For instance, 5-HT toxicity is expected when SSRIs are combined with MAOIs. Hence, this combination should never be prescribed. In the vast majority of cases, it is not expected that SSRIs alone in overdose would trigger a real 5-HT syndrome, as the MAO would prevent dramatic 5-HT increases. A review found that SSRI overdoses up to 30 times the typical daily dose resulted in minor or no symptoms. Combining two SSRIs is not an evidence-based strategy for treatment-resistant depression. However, such an approach would not lead to serotonin toxicity. Tryptophan (5-HT precursor) in combination with SSRIs also has the potential to produce serotonergic side effects. Again, real 5-HT toxicity is very rare when the MAO is not inhibited.

SSRIs must not be administered within 14 days of discontinuing treatment with an irreversible MAOI. When switching from an SSRI to an irreversible MAOI, clinicians must wait for five to seven half-lives after the SSRI has been stopped before introducing the MAOI.

Treatment of the 5-HT syndrome consists of removing the offending agents and promptly instituting comprehensive support care. Some benefits of pharmacologic interventions such as cyproheptadine and chlorpromazine have been suggested, but further research on this topic is needed.

A frequent drug combination in clinical practice is the use of a triptan to treat migraine along with an SSRI. Concurrent use of a triptan and an SSRI is normally uneventful, but significant reactions may occur. In response to multiple reports of serious reactions, in 2006, the FDA issued a Public Health Advisory warning of the possibility of 5-HT syndrome when triptans and SSRIs are taken together. According to the FDA advisory, physicians prescribing a triptan and SSRI should weigh the potential risk of 5-HT syndrome against the expected benefit of using a triptan with an SSRI. They should follow patients closely if a triptan and an SSRI are used together, particularly during treatment initiation, with dose increases, or with the addition of another serotonergic medication. It is important to remember that migraines on their own can produce severe physical and neurologic symptoms. Furthermore, most triptans do not cross the blood–brain barrier, and even if they did gain access to the CNS during a migraine attack (presumably due to increased permeability), they would decrease 5-HT release because of their agonistic effect at inhibitory 5-HT$_{1D}$ autoreceptors on 5-HT terminals. Therefore, patients who take a triptan and an SSRI together should be advised to seek medical attention only if they experience symptoms that they do not usually experience.

Because SSRIs may reduce dopamine turnover, their use in combination with dopamine-receptor antagonists can raise prolactin levels, even causing breast enlargement and galactorrhea, and produce extrapyramidal side effects.

Pharmacokinetic Interactions

The ability of medications for depression to exert a pharmacokinetic effect by enzyme inhibition is dependent on the extent of inhibition that is produced. Classification schemes of medications for depression denote individual drugs as high, medium, or low or minimal potential for isozyme inhibition, as shown in Table 33.5a.1–4.

Predicting the potential clinical relevance of isozyme inhibition requires knowledge of not only the potential inhibitory properties of drugs but also the clinically relevant substrates for the isozymes. Examples of commonly prescribed medications and the isozymes that metabolize them are listed in Table 33.5a.1–4. The combination of a substrate and a potential inhibitor does not contraindicate their simultaneous use.

Fluoxetine is a relatively potent inhibitor of the CYP 2D6 and CYP 2C19 isozymes but with no appreciable activity in the CYP 3A4 isozyme system, although norfluoxetine does demonstrate some inhibition of the CYP 3A4 isozyme in vitro.

Table 33.5a.1–4.
Cytochrome P450 (CYP) Inhibitory Potential of Commonly Prescribed Medications for Depression

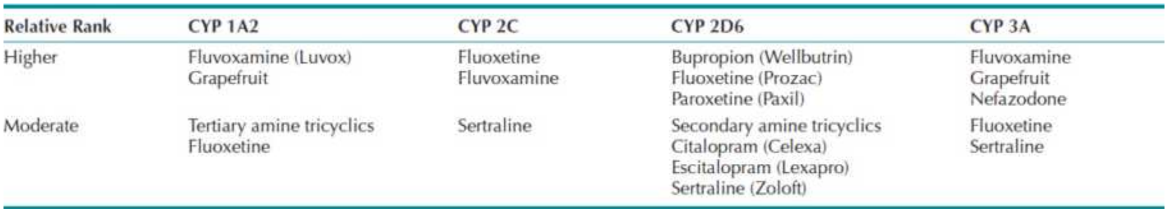

Relative Rank	CYP 1A2	CYP 2C	CYP 2D6	CYP 3A
Higher	Fluvoxamine (Luvox) Grapefruit	Fluoxetine Fluvoxamine	Bupropion (Wellbutrin) Fluoxetine (Prozac) Paroxetine (Paxil)	Fluvoxamine Grapefruit Nefazodone
Moderate	Tertiary amine tricyclics Fluoxetine	Sertraline	Secondary amine tricyclics Citalopram (Celexa) Escitalopram (Lexapro) Sertraline (Zoloft)	Fluoxetine Sertraline

Fluoxetine, paroxetine, sertraline, and citalopram inhibit the CYP 2D6 isozyme, albeit the latter do so only in their upper therapeutic ranges. Coadministration of these drugs with other medications that are substrates of the CYP 2D6 isozyme should be approached cautiously with an awareness of potential increases in serum concentrations of drugs that primarily use the CYP 2D6 pathway for metabolism. Examples of CYP 2D6 substrates include the type 1C antiarrhythmics—encainide (Enkaid), flecainide (Tambocor), propafenone (Rythmol), and mexiletine (Mexitil)—and β-blockers—labetalol (Normodyne), metoprolol (Lopressor), propranolol (Inderal), and timolol (Blocadren). Atenolol (Tenormin), a cardioselective β-blocker, is renally excreted and exhibits no pharmacokinetic interactions. Tamoxifen (Soltamox) is a prodrug that requires metabolism through the CYP 2D6 to be activated; hence coadministration with a drug that inhibits CYP 2D6 may reduces the effectiveness of Tamoxifen. Codeine is inactive as an analgesic and requires metabolism to morphine via the CYP 2D6 isozyme to be effective in reducing pain. Inhibition of the CYP 2D6 pathway can result in loss of analgesic properties.

There exists the potential for drug–drug interactions for an extended period after discontinuation, particularly with fluoxetine. Fluoxetine has a significantly longer half-life than the other SSRIs. A study was carried out to determine the onset and offset of CYP 2D6 inhibition. The 50% onset of

2D6 inhibition was 7 ± 2 days for fluoxetine 20 mg per day, 3 ± 2 days for paroxetine 20 mg per day, and 3 ± 3 days for sertraline 50 mg per day for 3 days, then 100 mg per day. The 50% offset of 2D6 inhibition after discontinuation was 63 ± 6 days (extrapolated time to baseline) for fluoxetine, 20 ± 6 days for paroxetine, and 25 ± 11 days for sertraline.

Fluoxetine may produce hypoglycemia in patients with diabetes who are treated with insulin or oral hypoglycemics. Dose reduction of the hypoglycemic or insulin may be required.

Paroxetine is a relatively potent inhibitor of the CYP 2D6 isozyme system but appears to be without clinically significant effects in the CYP 1A2, CYP 2C, or CYP 3A4 isozyme systems.

Citalopram and escitalopram are modest inhibitors of the CYP 2D6 isozyme based on in vivo testing. Coadministration of citalopram (40 mg per day) for 22 days resulted in a doubling of the plasma levels of metoprolol. In addition, coadministration of citalopram (40 mg per day) for 10 days along with a single dose of imipramine (Tofranil) (100 mg) did not alter the plasma concentrations of imipramine but did increase plasma levels of the primary metabolite, desipramine (Norpramin), by 50%. Similarly, escitalopram (20 mg per day) for 21 days resulted in increases in $C_{\max}$ (50%) and AUC (82%) of metoprolol given as a single dose of 100 mg. Likewise, escitalopram (20 mg per day) for 21 days resulted in increases in $C_{\max}$ (40%) and AUC (100%) of desipramine given as a single dose of 50 mg. In vitro enzyme assays suggest that citalopram is not an inhibitor for the CYP 3A4, CYP 2C9, or CYP 2E1 isozymes but is a weak inhibitor for CYP 1A2 and CYP 2C19, in addition to the data presented previously for the CYP 2D6 isozyme. Escitalopram does not appear to have in vitro inhibition of the CYP 3A4, CYP 1A2, CYP 2C9, CYP 2C19, and CYP 2E1 isozymes, although in vivo data are needed to confirm this.

Fluvoxamine is an inhibitor of the CYP 1A2, CYP 2C9, and CYP 3A4 isozyme systems. Theophylline, aminophylline (Phyllocontin), warfarin, propranolol, clozapine (Clozaril), and caffeine metabolism are inhibited through the CYP 1A2 isozyme by fluvoxamine. Warfarin metabolism is also inhibited through the CYP 2C9 isozyme system. GABA_A PAMs, including alprazolam (Xanax), clonazepam (Klonopin), triazolam (Halcion), and midazolam (Versed), are metabolized by hepatic oxidation via the CYP 3A4 isozyme. Approximately twofold elevations in serum concentrations of alprazolam have been observed after coadministration of fluvoxamine. Increased serum levels of carbamazepine (Tegretol), clozapine, methadone (Dolophine), propranolol, amitriptyline, clomipramine, and imipramine have been reported after the addition of fluvoxamine treatment. Thioridazine should not be combined with fluvoxamine owing to the potential drug–drug interaction and the potential for cardiac QTc interval prolongation secondary to the inhibition of thioridazine metabolism.

In premarketing trials, smokers demonstrated a 25% increase in fluvoxamine metabolism compared to nonsmokers, presumably through induction of the CYP 1A2 isozyme.

Fluoxetine, sertraline, and paroxetine are tightly bound to serum proteins and may displace other tightly protein-bound drugs, such as warfarin or digitoxin (Digitaline). Good clinical practice dictates checking serum levels or activity when medication with high protein binding is added to a tightly protein-bound drug with a low therapeutic index. Most assays do not differentiate between bound and free drugs, so functional assays, such as prothrombin time with warfarin, are a better indicator. Citalopram has a lower degree of protein binding (80%) than many other SSRIs, and escitalopram has an even lower degree of protein binding at approximately 56%. Thus, protein-binding interactions would not appear to be a problem with citalopram or escitalopram for most patients, although the clinical trials for these two compounds, as with other SSRIs, are typically done in a relatively healthy population, and the pharmacokinetics of SSRIs, when combined with multiple other agents is an under-researched area. The combination of citalopram with a highly protein-bound drug, such as warfarin resulted in a prolongation of prothrombin time by only 5%. Fluvoxamine has moderate protein binding and may be less likely to produce displacement of tightly protein-bound drugs than some of the others in this class.

DOSING AND ADMINISTRATION

The therapeutic dose ranges for the SSRIs have been well established, however, therapeutic effects are not dose related. There are only limited data that suggest that the presence of therapeutic levels of SSRIs correlates with plasma levels and subsequent clinical response. Nevertheless, patients may vary considerably in the amount of medication they need to derive clinical benefit. Consequently, dose optimization must be considered when the SSRI is well tolerated without inducing an adequate clinical response. Determination of the pharmacogenomic profile of patients with regards to transporters and receptors does not necessarily add to the effectiveness of SSRIs regimens, given their tolerability.

Fluoxetine

Fluoxetine is available in 10- or 20-mg pulvules, as 10-mg tablets, as 90-mg enteric-coated capsules for once-weekly administration, and as an oral concentrate of 20 mg/5 mL. The suggested starting dose of fluoxetine in patients with MDD is 20 mg per day. Effective doses for the treatment of MDD have been reported, ranging from 5 to 80 mg per day in dose-finding trials. Owing to the long half-life, achieving a new steady state after dose adjustment takes longer than that seen with most other medications. Pausing treatment for a few days to eliminate emerging side effects and then restoring the regimen at a lower dose, particularly if it has been several weeks since initiation of treatment or the last dose adjustment, is a reasonable approach to managing side effects.

Trials of fluoxetine in bulimia nervosa showed efficacy greater than placebo at 60 mg per day. Treatment of OCD is also usually at higher doses than those used in MDD, with commonly dosing up to 80 mg per day. However, SSRIs take longer to reduce OCD symptoms than symptoms of depression, so the drug should be given time on typical doses. The long half-life of fluoxetine and norfluoxetine does make interpretations of what may be the minimal therapeutic dose somewhat complicated. Fortunately, because fluoxetine is generally well tolerated, this becomes less of an issue than it might be with older medications.

Treatment of panic disorder is often initiated at 5 or 10 mg per day with gradual upward dose titration. Treatment of SAD or PMDD is often started at 10 to 20 mg per day. The dose is adjusted upward as needed.

A starting dose of 10 mg per day are often used in treating children, adolescents, or older adult patients. Dose increase is based on clinical response and tolerance of side effects.

Paroxetine

Paroxetine HCl is available as 10-, 20-, 30-, and 40-mg-size tablets. The 20-mg-size tablets are scored. Paroxetine CR is available in 12.5-, 25-, and 37.5-mg tablets. Paroxetine is also available in an oral suspension at a concentration of 10 mg/5 mL. Paroxetine mesylate is available as 10-, 20-, 30-, and 40-mg tablets.

The recommended starting dose for paroxetine is 20 mg per day for all approved indications, except in panic disorder, for which an initial dose of 10 mg per day is suggested. Starting paroxetine at 10 mg per day is also suggested for the older adults, medically ill, or those with significant renal or hepatic impairment. Paroxetine CR starting dose recommendations are similar, with 25 mg per day in MDD and 12.5 mg per day in patients with panic disorder, the older adults, the medically ill, or those with significant renal or hepatic impairment. Many clinicians find starting the more anxious or somatically preoccupied patient at a dose of 10 mg of paroxetine or 12.5 mg of paroxetine CR for a few days to be of benefit.

Taking paroxetine with meals, at least initially, reduces the incidence of nausea, and if present, the majority of patients describe nausea, to be a transient side effect. Patients with MDD who do not respond to 20 mg per day of paroxetine or 25 mg per day of paroxetine CR can have their doses increased by 10-mg (paroxetine) or 12.5-mg (paroxetine CR) intervals to as much as 50 mg per day (paroxetine) or 62.5 mg per day (paroxetine CR). In studies that established paroxetine to have efficacy in maintenance treatment, patients received a mean dose of approximately 30 mg per day.

The target dose of paroxetine in the treatment of OCD is 40 mg per day, with most responders receiving 40 to 60 mg per day. Presumably, doses of paroxetine CR need to be approximately 25% higher. The patient’s response to treatment of OCD is often slower than what is seen in MDD.

Patients with panic disorder usually tolerate the side effects of paroxetine if started at a lower dose than those used in treating MDD. Most panic disorder patients do better if started at an initial dose of 10 mg per day, with gradual titration to a target dose of 40 mg per day. Dose increases can be made at the rate of 10 mg per week. The maximal dose studied in premarketing clinical trials of paroxetine was 60 mg per day. The dosing strategy with paroxetine CR is similar, starting at 12.5 mg per week, and the maximal dose studied in clinical trials was 75 mg per day.

Dosing for other disorders is usually done with an initial dose of 10 to 20 mg per day (12.5 to 25.0 mg per day for paroxetine CR) followed by a gradual upward titration.

Among the SSRIs, paroxetine is associated with the most severe withdrawal syndrome and carries the strongest FDA warning about use in pregnancy.

Sertraline

Sertraline is available in scored tablets or capsules of 25-, 50-, and 100-mg strengths. The oral concentrate is formulated at a concentration of 20 mg/mL with 12% alcohol content and must be diluted before use. Owing to the alcohol content in the oral concentrate, it is contraindicated in conjunction with disulfiram (Antabuse). Initial dosing for MDD and OCD is typically at 50 mg per day, with escalation to 100 mg per day no earlier than after 14 days of treatment. Initial treatment of panic disorder, SAD, and PTSD is more typically 25 mg per day. Children in OCD studies were started at 25 mg per day, whereas adolescents in OCD studies were started at 50 mg per day. Treatment of PMDD is usually started at 50 mg per day given continuously daily or during the 2 weeks of the luteal phase. Individual patients may demonstrate improved tolerability starting at lower doses. Not uncommonly, patients require an upward titration of the dose.

The dose range in most patients with anxiety or affective disorders is typically 100 to 200 mg a day, although some patients respond to 50 mg per day. Patients with OCD often require higher doses for full therapeutic effect.

Citalopram and Escitalopram

Citalopram is available in 10-, 20-, and 40-mg-size tablets. The 20- and 40-mg-size tablets are scored. Citalopram is also available in an oral solution at a concentration of 10 mg/5 mL. Escitalopram is available in 10- and 20-mg-size tablets, both of which are scored. Disintegrating tablets are also available and may be useful when patients have difficulties swallowing. An oral solution of 5 mg/5 mL is available as well.

The manufacturer recommends a starting dose for citalopram of 20 mg per day, with the expectation of generally increasing the dose to 40 mg per day, which, in clinical practice, is typically done after at least 1 week on the initial dosing. Highly anxious patients or those with increased sensitivities to side effects of medications may benefit from starting with the 10-mg tablet. The recommended starting dose of escitalopram is 10 mg per day. If the dose is to be increased to 20 mg, this should generally occur no sooner than after one week on the initial dose. Because the 10-mg tablet is scored, 5 mg as a starting dose, if required, is available.

Many patients are observed to do well if citalopram or escitalopram is taken after meals for the first few days. After that period, they can take the medication with or without food.

Fluvoxamine

Fluvoxamine is available in 25-, 50-, and 100-mg scored tablets and 100- and 150-mg controlled-release capsules. Adults with OCD are often started at a dose of 50 mg administered at bedtime. The dose may be increased by 50 mg every 4 to 7 days. The effective dosing range in premarketing trials with OCD was 100 to 300 mg per day. The manufacturer suggests dividing the daily dose if it exceeds 100 mg per day, although many patients tolerate doses larger than 100 mg as a single dose. Initial upward titration is often easier with a split-dose regimen.

Treatment of OCD in adolescents and children should start at 25 mg per day with 25-mg-per-day increases every 4 to 7 days. The effective dosing range in premarketing studies was 50 to 200 mg per day. Adolescents often require a dose close to that used in adults, whereas children, perhaps owing to higher plasma fluvoxamine levels, may respond to the lower portion of the dosing range. Using the controlled-release formulation obviates this need for divided doses.

Dosing for disorders other than those for which fluvoxamine is approved in the United States is usually initiated at 25 to 50 mg per day. Patients with MDD typically show responses over the range of 100 to 300 mg per day, although the response is often seen in the range of 100 to 200 mg per day. Due to possible early activation of anxiety symptoms, treatment of panic disorder is best started at a dose of 25 mg per day for most patients, owing to increased sensitivity of side effects in this population. The typical effective dose is usually 100 to 200 mg per day.

SELECTIVE SEROTONIN REUPTAKE INHIBITORS AND SUICIDALITY

In the early 1990s, a debate arose as to whether fluoxetine treatment was associated with an increase in suicidal ideation and suicide attempts in some patients. A comparison of the emergence of suicidal ideation in patients receiving fluoxetine, tricyclic derivatives, or placebo suggested that fluoxetine-treated patients had less emergent suicidal ideation than those treated with the older drugs. Both groups on medication showed less emergent suicidal ideation than the group treated with a placebo. At times, fluoxetine produces symptoms, often described as restlessness, that are similar to akathisia in some patients. In the spring of 2004, the FDA asked manufacturers of the SSRIs, along with those of the other newer-generation medications for depression, to include a black box warning in their labeling recommending close observation of both adult and pediatric patients for signs or symptoms of worsening depression, including suicidal ideation, early on in therapy with these medications or on titration of dosing. This followed earlier recommendations concerning the potential of SSRIs to be associated with a higher rate of agitation, suicidal thoughts, or both in clinical trials when these rates were compared to treatment with placebo in studies of children and adolescents with MDD. Many clinicians believe that the recommendations to observe patients for worsening symptoms during the initiation of therapy represent a recommendation for sound clinical care. The issue regarding the treatment of children and adolescents remains somewhat unclear. One distinct possibility is that the 2% rate of agitation that was generally observed in studies that have been made public may represent pediatric patients who would ultimately go on to develop bipolar disorder.

A problem with the FDA advisory is that it was based on short-term trials and did not include suicidal patients. A study—the Adolescent Depression and Psychotherapy Trial (ADAPT)—actively included suicidal patients and found that the propensity for depressive relapse and self-reported suicidality in adolescents was reduced with SSRI administration. Furthermore, this study found that the reduction in clinical symptoms with SSRIs appeared to be a long-term effect. The authors of that study concluded that SSRIs are not detrimental to adolescent brain development and that therapeutic benefits outweigh the potential developmental risks.

A meta-analysis assessed the efficacy and suicidality risk of SSRIs, nefazodone, venlafaxine, and mirtazapine in a pediatric population. The authors found that these medications were effective for MDD, OCD, and anxiety disorders, as showed by an NNT of 10 (95% CI = 7–15), 6 (95% CI = 4–8), and 3 (95% CI = 2–5), respectively. Regarding the risk of reported suicidal ideation or attempt, the overall NNH was 143 (95% CI = 77–1,000). No completed suicide was identified.

Since the first press reports appeared, and especially since the new warning about suicidality in children and adolescents, there has been a drop in both SSRI prescriptions for that group and a decrease in the diagnosis of depression. Unfortunately, studies in the United States, Canada, and The Netherlands have shown a significant increase starting from 2003 to 2004 in suicide attempts and completed suicides. Analysis involving older adults shows that medications for depression reduce the risk of suicidal thoughts for 25- to 64-year-olds and may reduce the risk of both suicidal thoughts and behavior for people older than 65 years. The obvious conclusion is that although SSRIs might increase suicidality (suicidal ideation and behavior, but not death) in a small percentage of patients, the use of these drugs is associated with significant decreases in suicide attempts at the population level. Health advisories and warnings issued by the FDA may, in fact, have produced an unintended withholding of appropriate care and caused unexpected harm.

PREGNANCY

SSRIs are the most frequently used class of medications to treat depression among women of childbearing age. It is therefore not surprising that they are the most commonly used psychotropic drugs during pregnancy. Except for paroxetine, the SSRIs are considered relatively safe to take during pregnancy when deemed necessary for the treatment of the mother.

Women who are pregnant or planning a pregnancy and are taking SSRIs should consult with their physician about whether to continue taking their medication. As with all medications that treat serious disorders, such as those treated by SSRIs, the potential benefits and risks of SSRI treatment need to be weighed against the potential risks of not treating, given that pregnancy does not protect against the serious disorders these drugs treat.

A prospective study giving a choice to continue or stop medication for depression during pregnancy showed a higher rate of relapse following medication discontinuation, and women resuming their antidepressant regimen did not do as well as those who did not interrupt treatment.

Although studies have shown conflicting results, depression during pregnancy has been associated with a greater risk of obstetrical complications, premature delivery, pre-eclampsia, and a lower likelihood of breastfeeding initiation. Newborns of these women may be at an increased risk of low birth weight, and irritability, and may exhibit lower activity and interactive behaviors. Peripartum depression may also impact the maternal-fetal attachment process. A recent meta-analysis found a small-to-moderate association between antenatal anxiety and/or depression in mothers and performing worse offspring in social-emotional, cognitive, language, and adaptive behavior development. Interestingly, these findings were not limited to infancy but extended to childhood and adolescence.

Studies that assessed the risks associated with taking SSRIs during pregnancy have yielded conflicting results. This reflects the difficulty of studying this topic, as conducting randomized controlled trials on this matter poses ethical problems. Hence, the evidence is mostly derived from observational studies, and residual confounding is very likely. Among the possible risks identified, meta-analyses found an increased risk of preterm birth associated with the use of medications for depression (not limited to SSRIs) during pregnancy. The possibility of an association between SSRI use during pregnancy and autism has also been suggested but remains controversial.

A Centers for Disease Control and Prevention epidemiologic study on the causes of major birth defects among 9,622 case infants and 4,092 controls looked at the use of fluoxetine, sertraline, paroxetine, and citalopram during early pregnancy. It found that these drugs as a group did not significantly increase the risk for most birth defects, including congenital heart defects. The study did find an association between SSRI use and a slightly increased risk for three specific birth defects: anencephaly (adjusted odds ratio [OR] = 2.4; 95% CI 1.1–5.1), craniosynostosis (adjusted OR 2.5; 95% CI 1.5–4.0), and omphalocele (adjusted OR = 2.8; 95% CI, 1.3–5.7). Nevertheless, the authors emphasized that the absolute risks for these birth defects were small.

Concerns have been raised about the use of paroxetine during pregnancy and cardiovascular malformations in the newborn. In a study using Swedish national registry data, the risk of a cardiac defect was about 2% in infants exposed to paroxetine during pregnancy versus 1% among all registry infants. Another study using a U.S. insurance claims database compared the risk of congenital cardiac malformations in infants of women who received paroxetine versus other medications for depression in the first trimester. The results showed a trend toward increased risk with paroxetine (OR = 1.5; 95% CI 0.8–2.9). Most of the cardiac defects observed in these studies were atrial or ventricular septal defects, conditions in which the wall between the right and left sides of the heart is not completely developed. In general, septal defects are among the most common types of congenital malformations. They range from those that are symptomatic and may require surgery to those that are asymptomatic and may resolve on their own. It is generally recommended to avoid paroxetine during the first trimester of pregnancy or in women who plan to become pregnant in the near future.

Neonates exposed to SSRIs in the third trimester may develop poor neonatal adaptation syndrome (PNAS). This refers to short-term neurobehavioral effects in newborns. Symptoms such as poor muscle tone, trouble sleeping, shaking, weak or excessive crying, irritability, hypoglycemia, or seizures can be seen. It has been reported to affect up to 30% of exposed neonates and may represent discontinuation syndrome. The symptoms appear shortly after birth and may last up to two weeks. A meta-analysis found an increased risk of PNAS among neonates exposed to a medication for depression in utero (OR = 5.07; 95% CI, 3.25–7.90; P <.0001).

Another possible pregnancy-related side effect in offspring of SSRI-treated mothers is persistent pulmonary hypertension of the newborn (PPHN), a condition associated with substantial infant mortality and morbidity. A large cohort study examined the risk of PPHN associated with exposure to an SSRI in the 90 days before delivery versus no exposure. A total of 102,179 women had been exposed to SSRIs late in pregnancy. The authors found that 322 SSRI-exposed newborns (31.5; 95% CI, 28.3–35.2 per 10,000 births) were diagnosed with PPHN, compared to 7,630 newborns not exposed to medication for depression (20.8; 95% CI, 20.4–21.3 per 10,000 births). The unadjusted OR was 1.51 (95% CI, 1.35–1.69), and the OR became statistically nonsignificant when restricting to women diagnosed with depression and adjusting for confounders (OR = 1.10; 95% CI, 0.94–1.29). This data suggests that the risk of PPHN in newborns exposed to SSRIs late in pregnancy is low and may even be lower than previously thought.

Clinicians should not hesitate to offer a pharmacologic treatment for depression when meeting with pregnant women who struggle with moderate to severe depression, as the benefits most likely outweigh the potential risks in most cases.

Excretion of SSRIs in Breast Milk

SSRIs appear in breast milk to varying extents, and data from the literature differ in their findings regarding the relative dose being transmitted to the infant. Various reports about drug levels in milk are unreliable because of variations in milk collection procedures and because such medications tend to be lipophilic. Breast milk concentrations for some drugs may vary more than twofold between the watery fore milk and the fattier hind milk, whereas others may show little variation. Accordingly, clinical monitoring of maternal plasma or breast milk is not routinely indicated.

In some studies, concentrations are lowest for fluvoxamine and sertraline. Low levels of both sertraline and its metabolite are found in the sera of infants. The concentration of sertraline and desmethylsertraline in breast milk is highly variable and, on average, is of the same order of magnitude as their concentration in the blood plasma of the mother. As a result, more than half of breastfed babies receive less than 2 mg/day of sertraline and desmethylsertraline combined, and in most cases, these substances are undetectable in their blood. No changes occur in 5-HT uptake by the platelets of breastfed infants as measured by blood 5-HT levels before and after beginning sertraline treatment. Sertraline is more likely to be detected in an infant if the maternal daily dose is 100 mg or higher. Data on paroxetine are mixed, with some studies showing somewhat higher breast milk concentrations and others reporting paroxetine to have the lowest transfer into breast milk. According to a pooled analysis of available data, breastfed infants exposed to paroxetine or sertraline are least likely to demonstrate detectable or elevated plasma drug levels, whereas infants exposed to fluoxetine appear to be at higher risk of developing elevated levels. Most data show that the relative infant dose of citalopram is intermediate between those of paroxetine and fluoxetine. High doses of citalopram should also be used with caution. Escitalopram and its demethyl metabolite are present in breast milk, and unwanted effects in breastfed infant should be assessed.

If treatment with an SSRI is started during the postpartum period, fluoxetine should not be the first choice. Fluoxetine transfers to a greater extent, and its active metabolite, norfluoxetine, has a long half-life of 1 to 2 weeks and may accumulate in a breastfed infant.

Adverse effects are rarely reported in breastfed infants. Some short-term adverse effects that have been reported in association with fluoxetine include colic, seizures, irritability, withdrawal symptoms, and cyanosis. Caretakers should report irritability, poor feeding, poor sleep, or any other worrisome signs or symptoms in the infant. Premature breastfed infants, or those with impaired hepatic or other metabolite disturbances, need to be monitored more closely if the mother is taking an SSRI and breast-feeding.

When the use of an SSRI is clearly indicated in a breastfeeding woman, the data generally indicate that the positive effects of breastfeeding outweigh the risks for pharmacologic effects on the infant.

USE IN CHILDREN

The use of SSRIs in children and adolescents has been controversial. Few studies have shown clear-cut benefits from the use of these drugs. However, some children and adolescents do exhibit a beneficial response to these drugs in terms of depression and anxiety. Fluoxetine has most consistently demonstrated effectiveness in reducing symptoms of depression in both children and adolescents. This may be a function of the quality of the clinical trials involved. Given the potential negative effect of untreated depression and anxiety in a young population as well as the uncertainty about many aspects of how children and adolescents might react to medication, any use of SSRIs should be undertaken only within the context of a comprehensive management of the patient.

ACKNOWLEDGMENTS

This is a revised and updated version of the Selective Serotonin Reuptake Inhibitors chapter from Edition 10 of Kaplan & Sadock’s Comprehensive Textbook of Psychiatry by Dr. Norman Sussman (deceased).

FURTHER READING

Alwan S, Reefhuis J, Rasmussen SA, et al. Use of selective serotonin-reuptake inhibitors in pregnancy and the risk of birth defects. *N Engl J Med*. 2007;356(26):2684–2692.

Anglin R, Yuan Y, Moayyedi P, Tse F, Armstrong D, Leontiadis GI. Risk of upper gastrointestinal bleeding with selective serotonin reuptake inhibitors with or without concurrent nonsteroidal anti-inflammatory use: a systematic review and meta-analysis. *Am J Gastroenterol*. 2014;109(6):811–819.

Baldinger P, Kranz GS, Haeusler D, et al. Regional differences in SERT occupancy after acute and prolonged SSRI intake investigated by brain PET. *Neuroimage*. 2014;88:252–262.

Barbey JT, Roose SP. SSRI safety in overdose. *J Clin Psychiatry*. 1998;59(Suppl 15):42–48.

Beasley CM, Dornseif BE, Bosomworth JC, et al. Fluoxetine and suicide: a meta-analysis of controlled trials of treatment for depression. *BMJ*. 1991;303(6804):685–692.

Bech P, Tanghøj P, Cialdella P, Andersen HF, Pedersen AG. Escitalopram dose–response revisited: an alternative psychometric approach to evaluate clinical effects of escitalopram compared to citalopram and placebo in patients with major depression. *Int J Neuropsychopharmacol*. 2004;7(3):283–290.

Blier P, El Mansari M. Serotonin and beyond: therapeutics for major depression. *Philos Trans R Soc Lond B Biol Sci*. 2013;368(1615):20120536.

Bridge JA, Iyengar S, Salary CB, et al. Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: a meta-analysis of randomized controlled trials. *JAMA*. 2007;297(15):1683–1696.

Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Focus (Am Psychiatr Publ)*. 2018;16(4):420–429.

Cohen LS, Altshuler LL, Harlow BL, et al. Relapse of major depression during pregnancy in women who maintain or discontinue antidepressant treatment. *JAMA*. 2006;295(5):499–507.

Cole JA, Ephross SA, Cosmatos IS, Walker AM. Paroxetine in the first trimester and the prevalence of congenital malformations. *Pharmacoepidemiol Drug Saf*. 2007;16(10):1075–1085.

Dubnov-Raz G, Juurlink DN, Fogelman R, et al. Antenatal use of selective serotonin-reuptake inhibitors and QT interval prolongation in newborns. *Pediatrics*. 2008;122(3):e710–e715.

Goodyer I, Dubicka B, Wilkinson P, et al. Selective serotonin reuptake inhibitors (SSRIs) and routine specialist care with and without cognitive behaviour therapy in adolescents with major depression: randomised controlled trial. *BMJ*. 2007;335(7611):142.

Grigoriadis S, VonderPorten EH, Mamisashvili L, et al. The effect of prenatal antidepressant exposure on neonatal adaptation: a systematic review and meta-analysis. *J Clin Psychiatry*. 2013;74(4):e309–e320.

Henssler J, Alexander D, Schwarzer G, Bschor T, Baethge C. Combining antidepressants vs antidepressant monotherapy for treatment of patients with acute depression: a systematic review and meta-analysis. *JAMA Psychiatry*. 2022;79(4):300–312.

Huybrechts KF, Bateman BT, Palmsten K, et al. Antidepressant use late in pregnancy and risk of persistent pulmonary hypertension of the newborn. *JAMA*. 2015;313(21):2142–2151.

Källén BAJ, Olausson PO. Maternal use of selective serotonin re-uptake inhibitors in early pregnancy and infant congenital malformations. *Birth Defects Res A Clin Mol Teratol*. 2007;79(4):301–308.

Khanassov V, Hu J, Reeves D, van Marwijk H. Selective serotonin reuptake inhibitor and selective serotonin and norepinephrine reuptake inhibitor use and risk of fractures in adults: a systematic review and meta-analysis. *Int J Geriatr Psychiatry*. 2018;33(12):1688–1708.

Nurnberg HG, Hensley PL, Heiman JR, Croft HA, Debattista C, Paine S. Sildenafil treatment of women with antidepressant-associated sexual dysfunction: a randomized controlled trial. *JAMA*. 2008;300(4):395–404.

Olfson M, Marcus SC, Tedeschi M, Wan GJ. Continuity of antidepressant treatment for adults with depression in the United States. *Am J Psychiatry*. 2006;163(1):101–108.

Pan A, Sun Q, Okereke O, et al. Use of antidepressant medication and risk of type 2 diabetes: results from three cohorts of US adults. *Diabetologia*. 2012;55(1):63–72.

Papakostas GI, Fava M, Thase ME. Treatment of SSRI-resistant depression: a meta-analysis comparing within-versus across-class switches. *Biol Psychiatry*. 2008;63(7):699–704.

Rector TS, Adabag S, Cunningham F, Nelson D, Dieperink E. Outcomes of citalopram dosage risk mitigation in a veteran population. *Am J Psychiatry*. 2016;173(9):896–902.

Rogers A, Obst S, Teague SJ, et al. Association between maternal perinatal depression and anxiety and child and adolescent development: a meta-analysis. *JAMA Pediatr*. 2020;174(11):1082–1092.

Rush AJ, Trivedi MH, Wisniewski SR, et al. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med*. 2006;354(12):1231–1242.

Uher R, Frey BN, Quilty LC, et al. Symptom dimension of interest-activity indicates need for aripiprazole augmentation of escitalopram in major depressive disorder: a CAN-BIND-1 report. *J Clin Psychiatry*. 2020;81(4):9256.

Van Ameringen MA, Lane RM, Walker JR, et al. Sertraline treatment of generalized social phobia: a 20-week, double-blind, placebo-controlled study. *Am J Psychiatry*. 2001;158(2):275–281.

Voineskos AN, Wilson AA, Boovariwala A, et al. Serotonin transporter occupancy of high-dose selective serotonin reuptake inhibitors during major depressive disorder measured with [11C] DASB positron emission tomography. *Psychopharmacology (Berl)*. 2007;193(4):539–545.

Wikander I, Sundblad C, Andersch B, et al. Citalopram in premenstrual dysphoria: is intermittent treatment during luteal phases more effective than continuous medication throughout the menstrual cycle? *J Clin Psychopharmacol*. 1998;18(5):390–398.

Zivin K, Pfeiffer PN, Bohnert ASB, et al. Evaluation of the FDA warning against prescribing citalopram at doses exceeding 40 mg. *Am J Psychiatry*. 2013;170(6):642–650.

33.5a.2 Serotonin-Norepinephrine Reuptake Inhibitors (SNRIs)

PIERRE BLIER, M.D., PH.D., AND MICHAEL E. THASE, M.D.

Neuroscience-Based Nomenclature (NbN)

Desvenlafaxine, venlafaxine, duloxetine, milnacipran, levomilnacipran

Pharmacologic domain	Serotonin, norepinephrine
Mechanism of action	Reuptake inhibitor (SERT and NET)

The term *serotonin-norepinephrine reuptake inhibitor* (SNRI) describes a group of medications that have therapeutic effects that are presumably mediated by concomitant blockade of neuronal serotonin (5-HT) and norepinephrine (NE) uptake transporters. The SNRIs are also sometimes referred to as *dual reuptake inhibitors*, a broader functional class of medications that includes drugs sharing a tricyclic moiety such as clomipramine and, arguably, imipramine and amitriptyline. What distinguishes the newer SNRIs from the tricyclic derivatives is selectivity, which in this context refers to a relative lack of affinity for receptors, especially muscarinic, histaminergic, and the families of α- and β-adrenergic receptors. This distinction is an important one because the SNRIs have a more favorable tolerability profile than the older dual reuptake inhibitors.

Clinical interest in the SNRIs has largely stemmed from the hypothesis that direct modulation of these two monoamine systems conveys greater antidepressant effects than exerted by medications that selectively enhance only 5-HT or NE neurotransmission. Greater therapeutic benefits could result from: (1) an acceleration of postsynaptic adaptation to increased neuronal signaling, (2) simultaneous activation of two pathways for intracellular signal transduction, (3) additive effects on the activity of relevant neuronal factors such as brain-derived neurotrophic factor, (4) or, quite simply, broader coverage of depressive symptoms. Clinical evidence supporting this hypothesis first emerged in a pair of studies conducted by the Danish University Antidepressant Group, which found an advantage for the dual reuptake inhibitor clomipramine in comparison to the selective serotonin reuptake inhibitors (SSRIs) citalopram and paroxetine. A third report that compared the results of a group of patients prospectively treated with the combination of the tricyclic NE reuptake inhibitor desipramine and the SSRI fluoxetine, with a historical comparison group treated with desipramine alone, provided additional support. A meta-analysis of 25 inpatient studies comparing the efficacy of the tricyclic derivatives and SSRIs yielded the strongest evidence. Specifically, although the tricyclic medications were found to have a modest overall advantage, the superiority over SSRIs was almost entirely explained by the studies that used the drugs that are considered to be dual reuptake inhibitors—clomipramine, amitriptyline, and imipramine.

There are currently five SNRIs approved for use in the United States: Venlafaxine (Effexor and Effexor XR), desvenlafaxine succinate (Pristiq), duloxetine (Cymbalta), milnacipran (Savella), and levomilnacipran (Fetzima). The racemic agent milnacipran is indicated for major depressive disorder in a number of countries, including France and Japan, but in the United States, only for fibromyalgia.

VENLAFAXINE AND DESVENLAFAXINE

Venlafaxine (Fig. 33.5a.2–1) and desvenlafaxine are bicyclic phenylethylamine compounds that are structurally unrelated to other medications used to treat depression or anxiety. First synthesized in the late 1970s, venlafaxine was found to block uptake of 5-HT and NE in vitro, with a greater affinity for the 5-HT transporter. Importantly, venlafaxine was found to have virtually no affinity for muscarinic, histaminergic, or α- or β-adrenergic receptors. Venlafaxine thus was predicted to exert antidepressant activity comparable to the prototypic dual reuptake inhibitors (i.e., amitriptyline and clomipramine), yet affording a safety and tolerability profile that was more similar to the SSRIs.

After efficacy was established compared to placebo in the requisite phase II and phase III randomized controlled trials (RCTs), the original immediate-release (IR) formulation of venlafaxine hydrochloride was approved by the U.S. Food and Drug Administration (FDA) in 1994 for the treatment of major depressive disorder. A microencapsulated extended-release (XR) formulation was introduced in 1997, which permitted once-daily dosing. Generic formulations of venlafaxine are now available. An extended-release formulation of the succinate salt of O-desmethylvenlafaxine (ODV), the principal active metabolite of venlafaxine (DVS; Pristiq), was developed as a novel salt to improve its physicochemical and biologic characteristics. It was approved by the FDA for the treatment of major depressive disorder in early 2008. Generic formulations of desvenlafaxine are now available, but one of them is a free base.

Pharmacology

Venlafaxine and ODV inhibit neuronal uptake of 5-HT and, with significantly lower in vitro affinity, NE. In addition to an essential lack of affinity for muscarinic, histamine 1 (H_1), and adrenergic receptors, neither the parent drug nor the metabolite affects sodium fast channel activity. Venlafaxine and ODV do not inhibit the activity of monoamine oxidases.

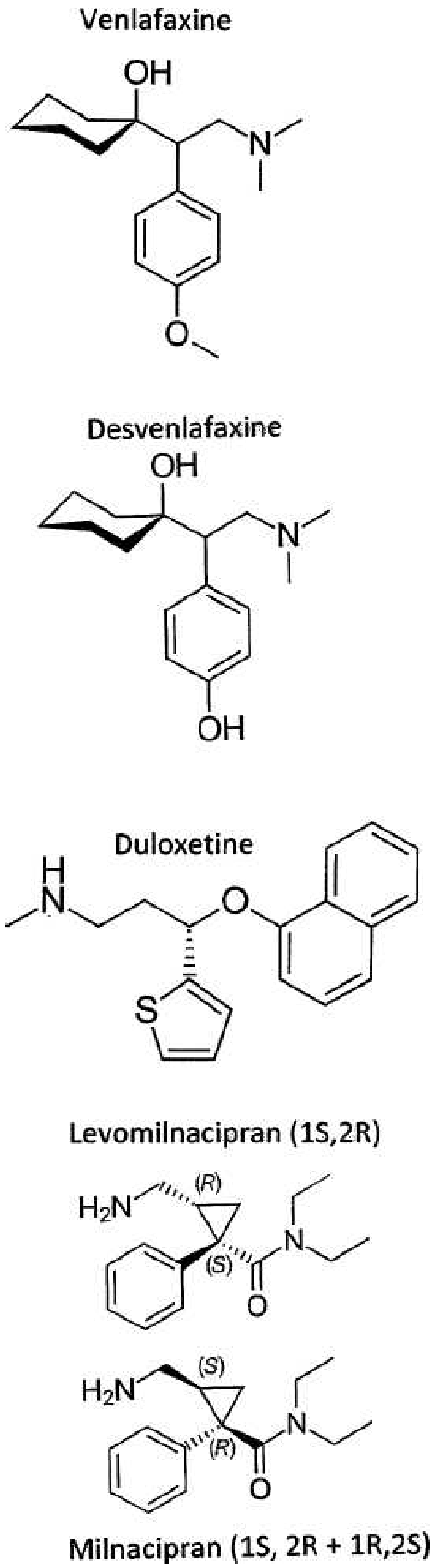
Venlafaxine

Desvenlafaxine

Duloxetine

Levomilnacipran (1S,2R)

Milnacipran (1S, 2R + 1R,2S)

OH
N
O
OH
N
OH
H
N
O
S
H2N
(R)
(S)
N
O
H2N
(S)
(R)
N
O

FIGURE 33.5a.2–1. Chemical structures of the serotonin-norepinephrine reuptake inhibitors (SNRIs) venlafaxine, desvenlafaxine, duloxetine, levomilnacipran, and milnacipran.

It is widely believed that the relative affinities of venlafaxine for inhibition of 5-HT and NE uptake underpin an ascending dose-response relationship. Specifically, venlafaxine is an SSRI at the lowest therapeutic doses (i.e., 75 mg per day), and NE transporters are consistently engaged at a dose of 225 mg/day. This hypothesis cannot be tested definitively until a specific ^{11}C-ligand for NE transporters is developed so that in vivo studies correlating clinical response with NE transporter occupancy can be undertaken. Venlafaxine at higher doses has also been proposed to be a triple reuptake inhibitor on the basis of in vitro work documenting weak inhibition of dopamine reuptake. Such a direct effect is probably not significant at usual therapeutic doses; however, considering that dopamine is taken back up by NE transporters in the frontal cortex and hippocampus, NE reuptake inhibitors can enhance synaptic levels of dopamine in some brain structures.

Orally administered venlafaxine and DVS are well absorbed from the gastrointestinal (GI) tract and undergo extensive first-pass metabolism in the liver. During venlafaxine therapy, ODV accumulates in higher concentrations (typically about twofold higher) than the parent drug. It thus can be said that ODV accounts for a majority of the therapeutic activity of venlafaxine. Coadministration with food decreases the rate, but not the extent, of absorption of the IR formulation and does not appreciably alter the absorption of venlafaxine XR or ODV. Peak plasma concentrations are achieved within 2 to 3 hours after ingestion of the IR formulation of venlafaxine. During therapy with the XR formulation, peak plasma concentrations are achieved within 5.5 hours for venlafaxine and 9 hours for ODV. When compared to the initial formulation, administration of venlafaxine XR results in lower peak plasma concentrations and a flatter plasma drug concentration curve when compared to venlafaxine IR and essentially the identical bioavailability. During the administration of DVS, peak plasma concentrations of ODV are achieved within about 7 to 9 hours. Formulation of ODV as a succinate salt significantly increases the bioavailability of DVS compared to venlafaxine ER.

Venlafaxine and ODV have linear kinetics over a dose range of 75 to 300 mg per day. Venlafaxine and ODV are primarily eliminated by the kidneys. The elimination half-lives of venlafaxine and its metabolite are short (i.e., 4 and 10 hours, respectively). Neither venlafaxine nor ODV is highly bound to plasma albumin (i.e., 25% to 30% protein binding at therapeutic concentrations).

Clearance of venlafaxine and ODV is reduced among patients with cirrhosis and severe renal disease, necessitating the use of lower doses for such patients. Without data from formal studies on the treatment of patients with severe hepatic or renal disease, a “half-dose–half-speed” approach is clinically reasonable. Adjustments in dosing on the basis of age alone typically are not necessary. However, a low starting dose and slow titration are recommended for the treatment of frail or medically compromised older patients.

Therapeutic Indications

Venlafaxine is approved by the FDA for the treatment of four disorders: major depressive disorder, generalized anxiety disorder, social anxiety disorder, and panic disorder. Major depressive disorder is currently the only FDA-approved indication for DVS.

Depression. The recommended starting dose of both formulations of venlafaxine, 75 mg per day, is the minimum effective dose for treatment of depression. It should be administered with meals on a two- or three-times-a-day basis for the IR. The XR formulation is available in 37.5-, 75-, and 150-mg capsules. The minimum therapeutic dose of DVS is 50 mg per day, and it is available in 50- and 100-mg tablets. Venlafaxine XR and DVS may be taken in the morning or evening, with or without food, as clinically indicated. In ambulatory clinical trials of venlafaxine using flexible doses, a modal daily dose of 150 mg per day was typically observed, regardless of the formulation used. When therapy at modest doses is ineffective, further increases up to 375 mg per day can be considered as tolerated. However, the maximum recommended daily dose of the XR formulation is 225 mg because of a lack of data on once-daily ingestion of higher dosages. However, higher mean doses are observed in studies of treatment-resistant depression. Despite some concerns that noradrenergic activity may exacerbate anxiety, efficacy has been established in studies of depressed people with significant concomitant symptoms of anxiety.

In placebo-controlled RCTs of DVS, doses ranging from 50 mg per day to 400 mg per day have shown efficacy, with no evidence of greater efficacy at higher doses, although higher regimens were introduced abruptly.

Venlafaxine is one of the most extensively studied newer medications for MDD. In published studies using active comparators, venlafaxine therapy has been shown to be at least as effective as treatment with the tricyclic derivatives imipramine, clomipramine, dothiepin, and amitriptyline (seven studies total). Research also suggests comparable overall efficacy with two other multi-action medications, mirtazapine (Remeron) and bupropion (Wellbutrin), although the former drug was found to have a better effect on sleep and the latter drug was found to have fewer sexual side effects. As described later in the chapter, a pair of RCTs has contrasted venlafaxine XR and duloxetine.

For both conceptual and commercial reasons, comparisons of venlafaxine and various SSRIs have garnered considerable attention. Looking at published fluoxetine (Prozac)-controlled studies, the differences in remission rates have ranged from nil to as much as 30% favoring venlafaxine; in aggregate, the difference in effect sizes was statistically significant. Although there are fewer published studies using other SSRIs, venlafaxine therapy similarly has ranged from comparable to more effective than citalopram, fluvoxamine, paroxetine, and sertraline. Only two studies have directly compared escitalopram and venlafaxine; the SNRI showed no efficacy advantage in these trials, and tolerability indices favored the SSRI.

A few comparative clinical trials have sufficient statistical power to distinguish between two effective medications for MDD, meta-analyses are increasingly used to examine relative efficacy. One such meta-analysis used the entire set of double-blind RCTs comparing venlafaxine IR and XR versus SSRIs conducted by the manufacturer of venlafaxine through 2005, focusing primarily on remission rates after 8 weeks of therapy. In this analysis of more than 7,000 depressed patients, those treated with venlafaxine were about one-third more likely to achieve remission than patients treated with SSRIs (odds ratio = 1.31), with an absolute advantage in remission rates of about 6%. Inspection of the results of all studies comparing venlafaxine and SSRIs, including those conducted by other sponsors, largely confirm this finding, although again, the average difference is small (odds ratio = 1.26), and the numeric difference favored the SSRI in approximately 20% of the studies (Fig. 33.5a.2–2).

Despite a noteworthy amount of comparative data, venlafaxine has not been studied extensively in MDD with psychotic features or bipolar depression. There are only three randomized studies of bipolar depression, including one comparison versus the SSRI paroxetine (bipolar I only), one versus the SSRI sertraline and bupropion (bipolar I and bipolar II), and one versus lithium (bipolar II only). In the first two studies, which required that patients take concomitant lithium or valproic acid, there were no significant differences in efficacy, although venlafaxine therapy was associated with a higher rate of treatment-emergent affective switches (TEAS) than the comparators. In the third study, which compared monotherapies, venlafaxine was significantly more effective than lithium and was not associated with an increased risk of TEAS.

Two long-term studies using more methodologically rigorous placebo-controlled discontinuation designs have confirmed efficacy for the prevention of relapse (i.e., 6 months after acute-phase treatment) and recurrence (12 and 24 months thereafter).

Generalized Anxiety Disorder. Venlafaxine XR received FDA approval for the treatment of generalized anxiety disorder in 1999. Efficacy and tolerability data for this indication are documented in four published randomized, controlled trials. In the one published study that used an active comparator, venlafaxine was found to be significantly more effective than the 5-HT$_{1A}$ agonist (BuSpar), which was not significantly more effective than placebo.

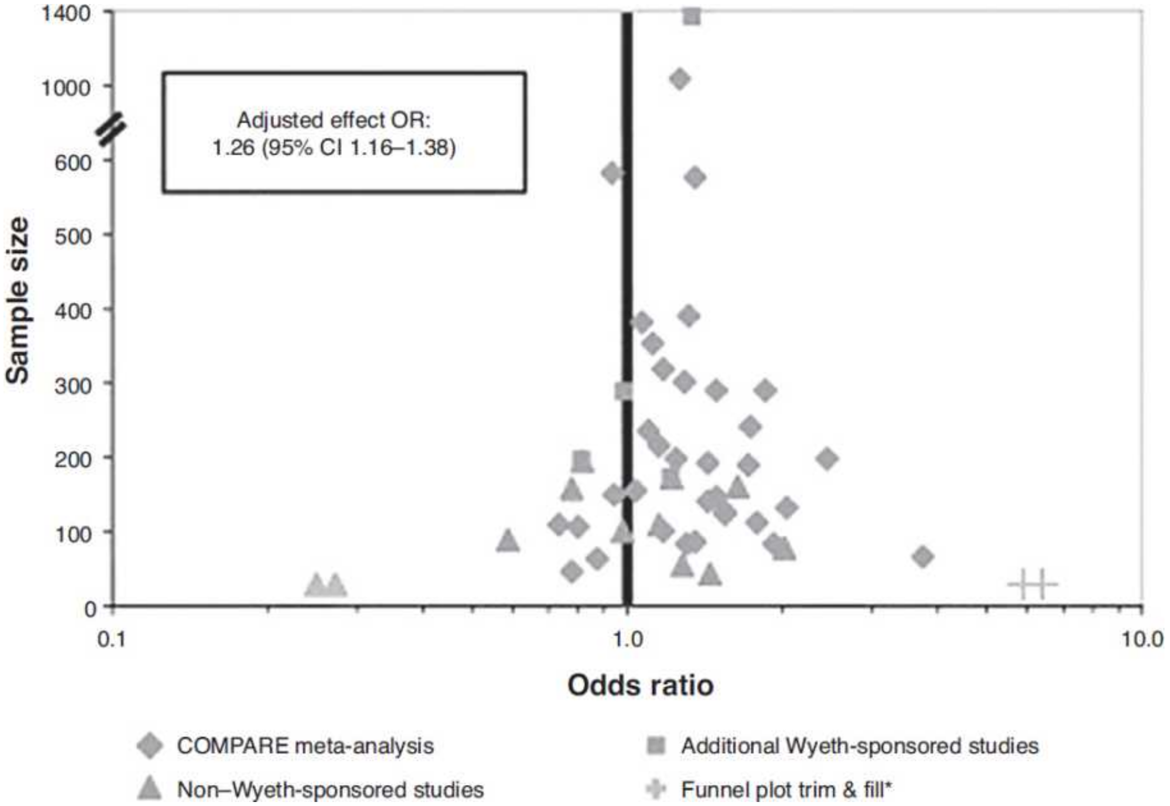

FIGURE 33.5a.2–2. Funnel plot of randomized controlled trials comparing venlafaxine and selective serotonin reuptake inhibitors. CI, confidence interval. (Reprinted from Nemeroff CB, Entsuah R, Benattia I, Demitrack M, Sloan DM, Thase ME. Comprehensive analysis of remission (COMPARE) with venlafaxine versus SSRIs. *Biol Psychiatry.* 2008;63(4):424–434. doi: 10.1016/j.biopsych.2007.06.027.)

To date, higher-dose venlafaxine therapy for generalized anxiety disorder has not been studied systematically, nor have there been any studies using another medication indicated for depression as the active comparator. The regulatory requirements of the first wave of double-blind studies also necessitated the exclusion of patients meeting criteria for MDD. Nevertheless, efficacy has been established in studies of anxious patients with a primary diagnosis of MDD.

Social Anxiety Disorder. The FDA approved venlafaxine for the treatment of social anxiety disorder in early 2003. To date, the results of three double-blind, placebo-controlled studies have been presented at scientific meetings. These trials demonstrated the efficacy and tolerability of venlafaxine at doses of 75 to 225 mg per day; both therapeutic effects and tolerability appear to be comparable to those of paroxetine.

Other Psychiatric Disorders. With respect to panic disorder, efficacy has been established in two RCTs, one including a 40 mg/day regimen of the SSRI paroxetine. In this specific study, 225 mg/day of venlafaxine outperformed 75 mg/day and the paroxetine treatment. Despite the obvious parallels with clomipramine, the antiobsessional effects of venlafaxine have not been as extensively studied. Two large RCTs have shown the efficacy of venlafaxine XR in post-traumatic stress disorder (PTSD), but it is an off-label use.

There are studies pertaining to the treatment of various other mood and anxiety disorders, attention-deficit/hyperactivity disorder (ADHD), and other psychiatric disorders. Premenstrual dysphoric disorder may be thought of as a sex-specific form of brief recurrent depression, and it is known to be responsive to SSRIs. A randomized, double-blind, placebo-controlled study evaluated flexible doses of venlafaxine therapy (ranging from 50 to 200 mg per day) across four menstrual cycles among 157 women with premenstrual dysphoric disorder. There was significantly greater improvement in the venlafaxine group compared to the placebo group across measures of mood, function, physical symptoms, and pain.

Tolerability

Venlafaxine has a safety and tolerability profile that approaches that of the more widely prescribed SSRI class. For example, in the pooled data set of Michael Thase and colleagues, 9% of venlafaxine-treated patients withdrew because of adverse events, as compared to 7% of SSRI-treated patients.

Nausea is the most frequently reported treatment-emergent adverse effect associated with venlafaxine therapy. The mechanism mediating this side effect is thought to be the stimulation of 5-HT type 3 (5-HT_3) receptors. With an incidence of upto 35% in short-term placebo-controlled trials, nausea is approximately one-third more likely than seen with SSRIs. This is presumed to be because low protein binding results in relatively high concentrations of free (biologically active) venlafaxine in the central nervous system (CNS) shortly after dosing. Nausea reliably diminishes in prevalence and severity during the first few weeks of therapy, presumably because of the desensitization of 5-HT_3 receptors. Use of the XR formulation somewhat reduces the incidence and intensity of nausea, as does taking the medication with meals. Initiating therapy at lower dosages (37.5 mg per day) also may help attenuate nausea. When extremely problematic, treatment-induced nausea can be controlled by prescribing a selective 5-HT_3 antagonist or mirtazapine.

As with other 5-HT reuptake inhibitors, venlafaxine therapy is associated with sexual side effects, predominantly decreased libido and a delay to orgasm or ejaculation. The incidence of these side effects is underreported in clinical trials and may exceed 30% to 40% when there is a direct, detailed assessment of sexual function. Although the exact mechanisms of the sexual side effects are not known, it is thought that 5-HT–mediated inhibition of NE and dopamine effects are implicated.

Other common side effects include headache, insomnia, somnolence, dry mouth, dizziness, constipation, asthenia, sweating, and nervousness. Although several side effects are suggestive of anticholinergic effects, the drug has no affinity for muscarinic or nicotinic receptors. Thus, noradrenergic agonism is likely to be the culprit.

Hypertension

Higher-dose venlafaxine therapy is associated with an increased risk of sustained elevations of blood pressure. Experience with the IR formulation in studies of depressed patients indicated that sustained hypertension was dose-related, increasing from 3% to 7% at doses of 100 to 300 mg per day and to 13% at doses greater than 300 mg per day. In this data set, venlafaxine therapy did not adversely affect blood pressure control of patients taking antihypertensives and even lowered the mean values of patients with elevated blood pressure readings before therapy. In controlled studies of the XR formulation, venlafaxine therapy resulted in only approximately a 1% greater risk of high blood pressure when compared to placebo. Arbitrarily capping the upper dose of venlafaxine used in these studies thus greatly attenuated concerns about elevated blood pressure. When higher doses of the XR formulation are used, however, monitoring of blood pressure is recommended.

Discontinuation Syndrome

Venlafaxine is one of the medications most commonly associated with discontinuation syndrome. This syndrome is characterized by the appearance of a constellation of adverse effects during a rapid taper or abrupt cessation, including dizziness, dry mouth, insomnia, nausea, nervousness, sweating, anorexia, diarrhea, somnolence, and sensory disturbances. As the XR formulation does *not* affect the elimination half-life of the compound, it does not decrease the potential for discontinuation symptoms. Patients need to be informed that venlafaxine therapy should not be stopped abruptly and that even missing a few days of therapy may provoke discontinuation symptoms. It is recommended that, whenever possible, a slow taper schedule should be used when longer-term treatment must be stopped (e.g., reducing the daily dose by no more than 75 mg each week). On occasion, substituting a few doses of fluoxetine may help to bridge this transition, especially at the last step of discontinuation.

Overdose

There were no overdose fatalities in premarketing trials, although electrocardiogram (ECG) changes (e.g., prolongation of QT interval, bundle branch block, QRS interval prolongation), tachycardia, bradycardia, hypotension, hypertension, coma, 5-HT syndrome, and seizures were reported. Fatal overdoses have been documented subsequently, typically involving venlafaxine ingestion in combination with other drugs or alcohol, or both. In an analysis of fatal medication overdoses in the United Kingdom from 1993 to 1999, Nicholas A. Buckley and Peter R. McManus found that venlafaxine had a significantly greater fatal toxicity index (a measure of overdose deaths per million prescriptions) than the SSRIs (13.2 vs. 0.7 to 3.0). These data could raise some concerns about the first-line use of venlafaxine for the treatment of suicidal patients. However, it should be noted that the authors did not take into account that psychiatrists wrote most of the venlafaxine prescriptions in the United Kingdom during the time in which these data were collected, whereas the SSRIs were widely prescribed by primary care providers. Thus, the higher fatality index may be describing the patient profile treated with venlafaxine (i.e., inpatients or those with more chronic, severe, or treatment-resistant mood disorders) rather than the toxicity of the drug.

Drug–Drug Interactions

Venlafaxine is metabolized in the liver primarily by the cytochrome P450 (CYP) 2D6 isoenzyme. Because the parent drug and principal metabolite are essentially equipotent, medications that inhibit this isoenzyme usually do not adversely affect therapy. In vitro and in vivo studies have shown venlafaxine to cause little or no inhibition of other P450 isoenzymes.

Venlafaxine is contraindicated in patients taking monoamine oxidase inhibitors (MAOIs) because of a potentially lethal pharmacodynamic interaction (i.e., serotonin syndrome). An MAOI must not be started for at least 7 days after stopping venlafaxine. There are now considerable data from trials examining the addition of partial dopamine agonists such as aripiprazole and brexpiprazole in inadequate response to SNRIs showing no complications. Regarding the combination of venlafaxine with lithium (Eskalith) and other medications for the treatment of bipolar disorder, clinical judgment should be exercised when combining medications, given the paucity of data.

DULOXETINE

Duloxetine hydrochloride is a propanamine compound that was synthesized in the 1980s by the same drug discovery team that identified the SSRI fluoxetine (Fig. 33.5a.2–1). Like fluoxetine, duloxetine is a potent 5-HT uptake inhibitor and lacks affinity for muscarinic, histaminic, and α- and β-adrenergic receptors. Unlike fluoxetine, duloxetine is also a relatively potent inhibitor of NE reuptake, and in in vitro studies, duloxetine was found to be a substantially more potent NE reuptake inhibitor than venlafaxine. If these studies predict effects in the brain, duloxetine should be expected to produce dual reuptake inhibition at lower absolute doses than observed with venlafaxine. Duloxetine was approved by the FDA for the treatment of major depressive disorder in 2004; other indications include generalized anxiety disorder, fibromyalgia, diabetic peripheral neuropathy, and chronic musculoskeletal pain.

Pharmacology

Orally administered duloxetine is well absorbed, and peak plasma levels are achieved approximately 3 hours after ingestion. Duloxetine has no active metabolites and exhibits linear pharmacokinetics within a dose range of 20 to 120 mg. Steady-state concentrations are achieved within 3 days of oral dosing, and the drug has an elimination half-life of approximately 12 hours (Fig. 33.5a.2–3).

In the plasma, approximately 90% of the drug is protein bound. After hepatic oxidation, metabolites of duloxetine are primarily eliminated by the kidneys. Duloxetine is metabolized by CYP 1A2 and is a moderately potent inhibitor of the CYP 2D6 isoenzyme.

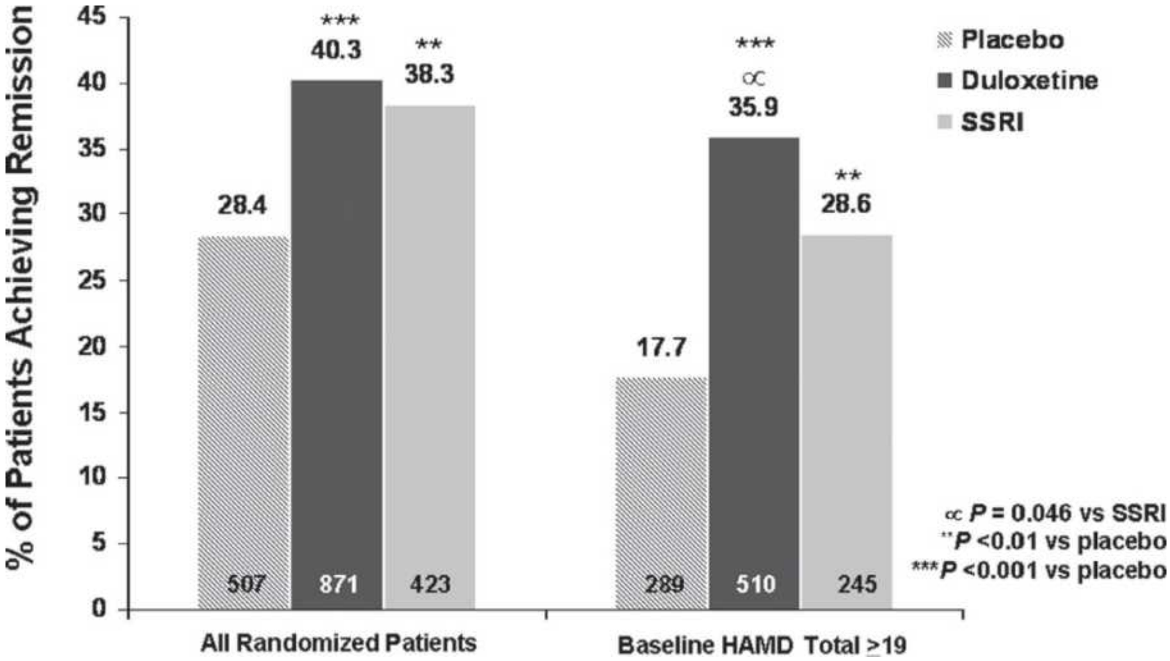

FIGURE 33.5a.2–3. Meta-analytic comparison of duloxetine (40 to 120 mg per day) and two selective serotonin reuptake inhibitors (SSRIs; fluoxetine and paroxetine; 20 mg per day) in six placebo-controlled studies of major depressive disorder. Remission is defined as Hamilton Depression (HAMD) Rating Scale Score <7 at the endpoint. Both antidepressants were more effective than placebo overall, with a significant advantage for duloxetine over the SSRIs among the subset of patients scoring 19 or higher on the 17-item HAMD. (Reprinted from Thase ME, Pritchett YL, Ossanna MJ, Swindle RW, Xu J, Detke MJ. Efficacy of duloxetine and selective serotonin reuptake inhibitors: comparisons as assessed by remission rates in patients with major depressive disorder. *J Clin Psychopharmacol*. 2007;27(6):672–676. doi: 10.1097/jcp.0b013e31815a4412.)

Antidepressant Efficacy

Eight phase II and phase III placebo-controlled, randomized, controlled trials were completed to establish efficacy and safety; these studies formed the basis for the FDA's approval of duloxetine therapy for major depressive disorder. No evidence of a dose-response relationship was observed between 60 mg per day and 120 mg per day; somewhat weaker antidepressant effects were observed at 40 mg per day. These clinical results are fully consistent with the observations that 40 mg of duloxetine reached the 80% occupancy level of the 5-HT transporter in the brain achieved with the minimal effective doses of SSRIs and venlafaxine, but that 60 mg once daily was necessary to maintain it. Consequently, given the lower affinity of duloxetine for the NE than the 5-HT transporter, higher doses than 60 mg/day are needed to reliably obtain a noradrenergic effect in the brain.

Six of the initial registration studies of duloxetine (40 to 120 mg per day) included an SSRI as an active comparator. A meta-analysis confirmed that both duloxetine and the comparators were effective and, among the subset of patients with moderate to severe depressive symptoms, there was a significant advantage favoring the SNRI in remission rates (Fig. 33.5a.2–3). However, none of these studies utilized the 60 mg per day dose of duloxetine and all of these studies used low (20 mg per day) doses of the comparators (fluoxetine or paroxetine). Three studies subsequently compared duloxetine (60 mg per day) versus escitalopram (10 to 20 mg per day). No advantage in efficacy was observed for the SNRI, and tolerability indices favored escitalopram. A pooled analysis of a pair of 6-week double-blind studies contrasting duloxetine (60 mg per day) and venlafaxine XR (150 mg per day) likewise found significant differences in efficacy. Duloxetine therapy was associated with a significantly higher incidence of nausea; venlafaxine therapy was associated with a higher incidence of treatment-emergent high blood pressure and, following the withdrawal of study medication, discontinuation-emergent symptoms.

Tolerability

The discontinuation rate due to adverse events in the studies summarized by Charles Nemeroff and colleagues was 15% in duloxetine-treated patients and 5% in placebo-treated patients. Side effect attrition was slightly higher in the duloxetine groups in the studies that used fluoxetine (10% vs. 6%) and paroxetine (14% vs. 10%) comparison groups, although these differences were not statistically significant in individual studies. The most commonly reported side effects during duloxetine therapy were nausea (22%), dry mouth (16%), fatigue (11%), dizziness (11%), and somnolence (8%). Incidence of sexual dysfunction appears to be comparable to that of the SSRIs. Of note, the incidence of nausea was similar for duloxetine (dosed twice daily) to fluoxetine and paroxetine (dosed once daily) in head-to-head double-blind trials. Despite the potential usefulness of the drug for the treatment of urinary incontinence, urinary hesitancy and urinary retention were not common occurrences in studies of depressed patients.

Like other 5-HT uptake inhibitors, duloxetine therapy has been associated with sexual dysfunction and, after abrupt cessation, a characteristic discontinuation syndrome. It is not yet possible to determine the relative risks of these phenomena during duloxetine therapy on a continuum with the other newer medications for MDD.

Duloxetine therapy resulted in a small increase in resting pulse rate (approximately two beats per minute) and an average increase in blood pressure of approximately 2 mm Hg in the studies reviewed by Nemeroff and colleagues. Of note, the risk of sustained high blood pressure during duloxetine therapy was approximately 1% greater than observed in placebo-treated patients. Curiously, the risk of increased blood pressure does not appear to be dose-dependent.

Although the CYP enzyme inhibition profile of the drug looks generally favorable, duloxetine levels are about 30% lower in smokers, likely as a result of CYP 1A2 induction, and increased four-fold by potent CYP 1A2 inhibitors, like fluvoxamine and some fluoroquinolone antibiotics. A lethal interaction with MAOIs is expected with duloxetine, and the guidelines suggested previously for venlafaxine therapy must be followed.

MILNACIPRAN AND LEVOMILNACIPRAN

The SNRI milnacipran is distinguished as a dual reuptake inhibitor that on the basis of in vitro and in vivo studies (Fig. 33.5a.2–1). Specifically, milnacipran is twice more potent for inhibition of NE uptake than for 5-HT reuptake inhibition. Milnacipran thus can be thought of as a preferential NE uptake inhibitor at the minimum therapeutic dosage (50 mg twice a day) that exerts progressively greater inhibitory effects on 5-HT reuptake at higher doses (i.e., 200 mg/day, its maximal recommended dose for fibromyalgia). Like the other SNRIs, milnacipran has almost no affinity for muscarinic, histaminic, or α- and β-noradrenergic receptors.

Pharmacology

Milnacipran has a half-life of approximately 8 hours and shows linear pharmacokinetics between doses of 50 and 250 mg per day. It is minimally metabolized in the liver and has no active metabolites. Milnacipran is primarily excreted by the kidneys. Most orally administered milnacipran is bioavailable, and only 13% of milnacipran is bound to plasma protein. Milnacipran is not an inhibitor of any of the CYP isoenzymes.

Antidepressant Efficacy

The countries that have approved milnacipran for general use in the treatment of depression have not required completion of a large number of placebo-controlled studies, therefore, efficacy is not as well established as in the case for venlafaxine or duloxetine. A meta-analysis of 16 RCTs (including all published and unpublished studies) encompassing 2,277 patients concluded that milnacipran did not have significantly different efficacy, acceptability, and tolerability when compared to other medications for MDD. However, when compared to TCAs, fewer patients dropped out due to the side effects of milnacipran.

Tolerability

In comparative clinical trials, milnacipran therapy was associated with significantly fewer anticholinergic and antihistaminic side effects than the tertiary amine tricyclic derivatives and fewer GI side effects than the SSRIs. Noteworthy side effects (i.e., incidence greater than an SSRI) included dizziness, sweating, and urinary hesitancy. Neither prospective data nor any expert opinion on the incidence of sexual dysfunction, especially when contrasted to the SSRIs, could be located. Likewise, the potential for lethality in overdose has not been systematically examined. Like 5-HT reuptake inhibitors, milnacipran should not be prescribed in proximity to an MAOI.

Characteristics of Levomilnacipran

Levomilnacipran is the active enantiomer of the racemic mixture of milnacipran. It is available in an extended-release formulation and has a higher maximal plasma concentration (peaks in 6 to 8 hours), a greater area under the curve, and a longer half-life (12 hours) than its much less active dextrorotary enantiomer. It is only 22% bound to plasma proteins and excreted by the kidneys 58% unchanged. It is recommended to initiate it at 20 mg/day for 2 days and gradually increase the dose to 40, 80, and 120 mg/day according to tolerability and response. It is effective in 40 to 120 mg/day doses based on a data set of five placebo-controlled studies on 2,598 patients aged 18 to 80. Interestingly, it was effective in patients aged 60 and older, a population that is generally resistant to other medications. Currently, there are no active comparator studies, but there is a positive double-blind maintenance trial that has not revealed evidence of discontinuation symptoms. In both acute and long-term studies, there was no evidence of weight gain. On average, there was approximately a 7 beat per minute increase in heart rate, a thumbprint of NE reuptake inhibition. Only prospective studies using specific sexual function scales, and not merely TEAEs, could determine whether it is less problematic than other more potent 5-HT reuptake inhibiting medications. As for other SNRIs, it is recommended to utilize a gradual taper to discontinue the medication.

SUMMARY

The SNRIs venlafaxine, desvenlafaxine, duloxetine, milnacipran, and levomilnacipran are effective medications for the treatment of MDD. There is evidence of an ascending dose-response relationship for venlafaxine and, possibly, milnacipran and duloxetine. A substantial body of research documents that venlafaxine therapy (at higher doses) has stronger antidepressant effects than the SSRIs; some data suggest a similar advantage for duloxetine and, for patients with severe depression, milnacipran. The efficacy of venlafaxine also has clearly been established in controlled studies of generalized anxiety, social anxiety, and panic disorders. Treatment with SNRIs is generally well tolerated. Tolerability and safety profiles are superior to the tricyclic derivatives and approach those of the SSRIs. A higher initial incidence of nausea and, at higher doses, elevated blood pressure may be distinguishing characteristics of venlafaxine therapy. Apart from duloxetine, the four other medications have a low potential for CYP-mediated drug–drug interactions.

FURTHER READINGS

Allgulander C, Hackett D, Salinas E. Venlafaxine extended release (ER) in the treatment of generalised anxiety disorder: twenty-four-week placebo-controlled dose-ranging study. *Br J Psychiatry*. 2001;179:15–22.

Amsterdam JD, Wang CH, Shwarz M, Shults J. Venlafaxine versus lithium monotherapy of rapid and non-rapid cycling patients with bipolar II major depressive episode: a randomized, parallel group, open-label trial. *J Affect Disord*. 2009;112:219–230.

Asnis GM, Henderson MA. Levomilnacipran for the treatment of major depressive disorder: a review. *Neuropsychiatr Dis Treat*. 2015;11:125–135.

Auclair AL, Martel JC, Assié MB, et al. Levomilnacipran (F2695), a norepinephrine-preferring SNRI: profile in vitro and in models of depression and anxiety. *Neuropharmacology*. 2013;70:338–347.

Buckley NA, McManus PR. Fatal toxicity of serotoninergic and other antidepressant drugs: analysis of United Kingdom mortality data. *BMJ*. 2002;325:1332–1333.

Bymaster FP, Dreshfield-Ahmad LJ, Threlkeld PG, et al. Comparative affinity of duloxetine and venlafaxine for serotonin and norepinephrine transporters in vitro and in vivo, human serotonin receptor subtypes, and other neuronal receptors. *Neuropsychopharmacology*. 2001;25:871–880.

Deecher DC, Beyer CE, Johnston G, et al. Desvenlafaxine succinate: a new serotonin and norepinephrine reuptake inhibitor. *J Pharmacol Exp Ther*. 2006;318:657.

Durgam, S, Chen C, Migliore R, Prakash C, Thase ME. Relapse prevention with levomilnacipran ER in adults with major depressive disorder: a multicenter, randomized, double-blind, placebo-controlled study. *Depress Anxiety*. 2019;36:225–234.

Frampton JE, Plosker GL. Duloxetine: a review of its use in the treatment of major depressive disorder. *CNS Drugs*. 2007;21:581–609.

Guelfi JD, Ansseau M, Timmerman L, Korsgaard S. The Mirtazapine-Venlafaxine Study Group. Mirtazapine versus venlafaxine in hospitalized severely depressed patients with melancholic features. *J Clin Psychopharmacol*. 2001;21:425–431.

Katzman MA, Bleau P, Blier P, Chokka P, et al. Canadian clinical practice guidelines for the management of anxiety, posttraumatic stress and obsessive-compulsive disorders. *BMC Psychiatry*. 2014;14 Suppl 1:S1.

Keller MB, Trivedi MH, Thase ME, et al. The prevention of recurrent episodes of depression with venlafaxine for two years (prevent) study: outcomes from the 2-year and combined maintenance phases. *J Clin Psychiatry*. 2007;68:1246–1256.

Kennedy SH, Eisfeld BS, Dickens SE, Bacchiochi JR, Bagby M. Antidepressant-induced sexual dysfunction during treatment with moclobemide, paroxetine, sertraline, and venlafaxine. *J Clin Psychiatry*. 2000;61:276–281.

Liebowitz MR, Manley AL, Padmanabhan SK, Ganguly R, Tummala R, Tourian KA. Efficacy, safety, and tolerability of desvenlafaxine 50 mg/day and 100 mg/day in outpatients with major depressive disorder. *Curr Med Res Opin*. 2008;24:1877–1890.

Montgomery SA, Baldwin DS, Blier P, et al. Which antidepressants have demonstrated superior efficacy? A review of the evidence. *Int Clin Psychopharmacol*. 2007;22:323–329.

Nemeroff CB, Entsuah R, Benattia I, Demitrack M, Sloan DM, Thase ME. Comprehensive analysis of remission (COMPARE) with venlafaxine versus SSRIs. *Biol Psychiatry*. 2008;63:424–434.

Owens MJ, Krulewicz S, Simon JS, et al. Estimates of serotonin and norepinephrine transporter inhibition in depressed patients treated with paroxetine or venlafaxine. *Neuropsychopharmacology*. 2008;33:3201–3212.

Pae CU, Lim HK, Ajwani N, Lee C, Patkar AA. Extended-release formulation of venlafaxine in the treatment of post-traumatic stress disorder. *Expert Rev Neurother*. 2007;7:603–615.

Papakostas GI, Thase ME, Fava M, Nelson JC, Shelton RC. Are antidepressant drugs that combine serotonergic and noradrenergic mechanisms of action more effective than the selective serotonin reuptake inhibitors in treating major depressive disorder? A meta-analysis of studies of newer agents. *Biol Psychiatry*. 2007;62:1217–1227.

Perahia DGS, Pritchett YL, Kajdasz DK, et al. A randomized, double-blind comparison of duloxetine and venlafaxine in the treatment of patients with major depressive disorder. *J Psychiatr Res*. 2008;42:22–34.

Pollack M, Mangano R, Entsuah R, Tzanis E, Simon NM, Zhang Y. A randomized controlled trial of venlafaxine ER and paroxetine in the treatment of outpatients with panic disorder. *Psychopharmacology (Berl)*. 2007;194:233–242.

Rynn M, Russell J, Erickson J, et al. Efficacy and safety of duloxetine in the treatment of generalized anxiety disorder: a flexible-dose, progressive-titration, placebo-controlled trial. *Depress Anxiety*. 2008;25:182–189.

Sagman D, McIntosh D, Lee MS, et al. Attributes of response in depressed patients switched to treatment with duloxetine. *Int J Clin Pract*. 2011;65:73–81.

Smith D, Dempster C, Glanville J, Freemantle N, Anderson I. Efficacy and tolerability of venlafaxine compared with selective serotonin reuptake inhibitors and other antidepressants: a meta-analysis. *Br J Psychiatry*. 2002;180:396–404.

Smith T, Nicholson RA. Review of duloxetine in the management of diabetic peripheral neuropathic pain. *Vasc Health Risk Manag*. 2007;3:833–844.

Takano A, Suzuki K, Kosaka J, et al. A dose-finding study of duloxetine based on serotonin transporter occupancy. *Psychopharmacology*. 2006;185:395–399.

Thase ME, Entsuah AR, Rudolph RL. Remission rates during treatment with venlafaxine or selective serotonin reuptake inhibitors. *Br J Psychiatry*. 2001;178:234–241.

Thase ME, Friedman ES, Howland RH. Venlafaxine and treatment-resistant depression. *Depress Anxiety*. 2000;12:55–62.

Thase ME, Pritchett YL, Ossanna MJ, et al. Efficacy of duloxetine and selective serotonin reuptake inhibitors: comparisons as assessed by remission rates in patients with major depressive disorder. *J Clin Psychopharmacol*. 2007;27:672–676.

Thase ME, Sloan DME. Venlafaxine. In: Schatzberg AF, Nemeroff CB, eds. *Textbook of Psychopharmacology*. 4th ed. American Psychiatric Publishing, Inc.; 2009:465.

Thase ME, Tran PV, Wiltse C, Pangallo BA, Mallinckrodt C, Detke MJ. Cardiovascular profile of duloxetine, a dual reuptake inhibitor of serotonin and norepinephrine. *J Clin Psychopharmacol*. 2005;25:132–140.

Turcotte JE, Debonnel G, de Montigny C, Hébert C, Blier P. Assessment of the serotonin and norepinephrine reuptake blocking properties of duloxetine in healthy subjects. *Neuropsychopharmacology*. 2001;24:511–521.

van den Broek WW, Mulder PGH, van Os E, Birkenhäger TK, Pluijms E, Bruijn JA. Efficacy of venlafaxine compared with tricyclic antidepressants in depressive disorder: a meta-analysis. *J Psychopharmacol*. 2009;23:708–713.

Vieta E, Martinez-Aran A, Goikolea JM, et at. A randomized trial comparing paroxetine and venlafaxine in the treatment of bipolar depressed patients taking mood stabilizers. *J Clin Psychiatry*. 2002;63:508–512.

Vincent S, Bieck PR, Garland EM, et al. Clinical assessment of norepinephrine transporter blockade through biochemical and pharmacological profiles. *Circulation*. 2004;109:3202–3207.

33.5a.3 Tricyclics and Tetracyclics

J. CRAIG NELSON, M.D.

Neuroscience-Based Nomenclature (NbN)	
Desipramine, Protriptyline, Nortriptyline	
Pharmacologic domain	Norepinephrine
Mechanism of action	Reuptake inhibitors (NET)
Lofepramine, Amoxapine	
Pharmacologic domain	Norepinephrine, Serotonin
Mechanism of action	Reuptake inhibitors (NET and SERT)
Imipramine, Dosulepin, Clomipramine	
Pharmacologic domain	Serotonin, Norepinephrine
Mechanism of action	Reuptake inhibitors (SERT and NET)
Amitriptyline	
Pharmacologic domain	Serotonin, Norepinephrine
Mechanism of action	Reuptake inhibitors (SERT and NET) and receptor antagonist (5-HT_2)
Doxepin	
Pharmacologic domain	Norepinephrine, Serotonin
Mechanism of action	Reuptake inhibitors (NET and SERT) and receptor antagonist (5-HT_2)
Trimipramine	
Pharmacologic domain	Serotonin, Dopamine
Mechanism of action	Receptor antagonist (5-HT_2 and D_2)

HISTORY

In 1957, Roland Kuhn, a Swiss psychiatrist, observed that imipramine (Tofranil) had antidepressant effects in a small sample of 40 patients. This serendipitous observation, like those before for lithium (Eskalith) and chlorpromazine (Thorazine), had far-reaching import. It demonstrated that medications could be used to treat depression. It led to the development of a new class of antidepressant compounds. This clinical observation, coupled with Julius Axelrod's report of imipramine's role in blocking the reuptake of norepinephrine, laid the foundation for the norepinephrine hypothesis of depression.

The study of the treatment of depression with imipramine and related compounds would drive the development of a new methodology. For example, objective measures were required to assess change during treatment. Max Hamilton developed the Hamilton Depression Rating Scale to monitor change during the administration of these new compounds. Other scales followed. Better methods of defining depression were needed, and the Washington University group proposed criteria for primary affective disorder, criteria that became the basis for the category of major depression in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III). Observation of substantial placebo effects in depression during clinical trials led to the appreciation of the need for placebo controls.

The field also learned the basics of pharmacologic management of depression with imipramine and other tricyclics. The necessity for administering an adequate dose for an adequate duration was described. The need to continue treatment to prevent relapse and recurrence was identified. The observation of the substantial variability of the tricyclic concentrations in plasma stimulated a whole area of study of the pharmacology of these compounds. This research introduced psychiatry to the concept of *genetic polymorphisms* and the possible consequences of drug interactions.

After the introduction of imipramine, several other antidepressant compounds were developed that shared a basic tricyclic structure and relatively similar secondary or side effects. Later, other heterocyclic compounds were also marketed that were somewhat similar in structure and that had relatively comparable secondary properties.

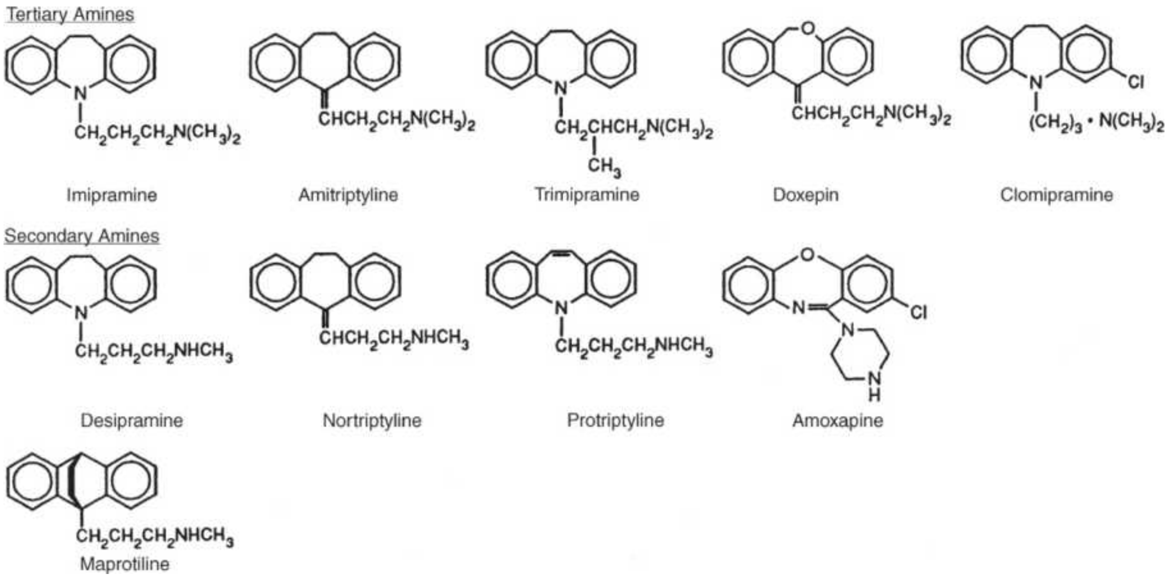
Tertiary Amines
N
CH2CH2CH2N(CH3)2
Imipramine
CHCH2CH2N(CH3)2
Amitriptyline
N
CH2CHCH2N(CH3)2
CH3
Trimipramine
O
CHCH2CH2N(CH3)2
Doxepin
N
Cl
(CH2)3 • N(CH3)2
Clomipramine
Secondary Amines
N
CH2CH2CH2NHCH3
Desipramine
CHCH2CH2NHCH3
Nortriptyline
N
CH2CH2CH2NHCH3
Protriptyline
O
N
Cl
N
N
H
Amoxapine
CH2CH2CH2NHCH3
Maprotiline

FIGURE 33.5a.3–1. Molecular structure of tricyclics and tetracyclics.

CHEMISTRY

Tricyclic and tetracyclic compounds are categorized primarily on the basis of their chemical structure (Fig. 33.5a.3–1). The *tricyclics* have a three-ring structure, hence, their name. The tertiary amine tricyclics, such as amitriptyline (Elavil) and imipramine, have two methyl groups at the end of the side chain. These compounds can be demethylated into secondary amines, such as desipramine (Norpramin, Pertofrane) and nortriptyline (Aventyl, Pamelor). The tetracyclic compounds maprotiline (Ludiomil) and amoxapine (Asendin) have a four-ring central structure. Five tertiary amines have been marketed in the United States—amitriptyline, clomipramine (Anafranil), doxepin (Sinequan), imipramine, and trimipramine (Surmontil). The three secondary amine compounds are desipramine, nortriptyline, and protriptyline (Vivactil).

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Absorption. Absorption of the tricyclic and tetracyclic drugs occurs in the small intestine and is reasonably complete and rapid. Peak levels after ingestion occur within 2 to 8 hours. Exceptions are protriptyline, which is absorbed more slowly, with peak levels obtained between 6 and 12 hours, and maprotiline, which has peak levels occurring in 8 hours or more. Because the antidepressant action of these drugs occurs over several weeks, the timing of peak levels is relatively unimportant for efficacy but does have implications for side effects, especially during initial dosing.

Distribution. The tricyclic and tetracyclic compounds are basic lipophilic amines. As a result, they are concentrated in a variety of tissues throughout the body and have a high volume of distribution. In some organs, the heart, for example, the drug is concentrated at levels exceeding those in plasma. Because of their lipid solubility, these compounds are extensively bound to plasma proteins. Binding of the tricyclic and tetracyclic drugs is generally 90% or greater. Exceptions are the hydroxymetabolites of the tricyclics, which have lower plasma protein binding.

The lipophilic properties of the tricyclic and tetracyclic compounds also contribute to their ability to cross the blood–brain barrier and reach their targets in the CNS. Previously it was thought that brain concentrations of antidepressants were dependent primarily on the free or unbound plasma concentrations. Recent work suggests that p-glycoproteins (P-GP) actively transport some antidepressant compounds out of the brain, thus lowering drug concentrations. P-GP is encoded by the ABCB1 gene and is expressed on capillary endothelial cells of the blood–brain barrier. The clinical relevance of P-GP is supported by studies finding single nucleotide polymorphisms of the ABCB1 gene associated with response to paroxetine, a P-GP substrate, but not all studies agree. Not all antidepressants are substrates of P-GP, but several of the tricyclics—amitriptyline, nortriptyline, imipramine, desipramine, doxepin, and trimipramine—are substrates.

Tricyclic compounds cross the placenta. Transfer of the tertiary compound amitriptyline is slightly higher than that for the secondary compound nortriptyline. Neonatal drug withdrawal can occur. The tricyclics are excreted in breast milk at concentrations similar to plasma. The actual quantity delivered, however, is small so drug levels in the infant are usually undetectable or very low.

Metabolism and Elimination. After absorption, the tricyclics are taken up in the circulation, pass first through the liver, and metabolism of the drug begins. As a result of this *first-pass effect,* the amount of the compound that enters the systemic circulation is reduced.

Clearance of the tricyclic and tetracyclic compounds is principally the result of hepatic metabolism. Renal clearance accounts for only a small portion of drug elimination. Rates of hepatic metabolism vary widely from person to person. Elimination half-lives for most of the tricyclic and tetracyclic compounds are approximately 24 hours or longer. Thus the drugs can be given once a day (Table 33.5a.3–1). Amoxapine, which has a shorter half-life, is an exception.

Table 33.5a.3–1.
Receptor Affinity or Potency of Cyclic Antidepressants in Human Brain

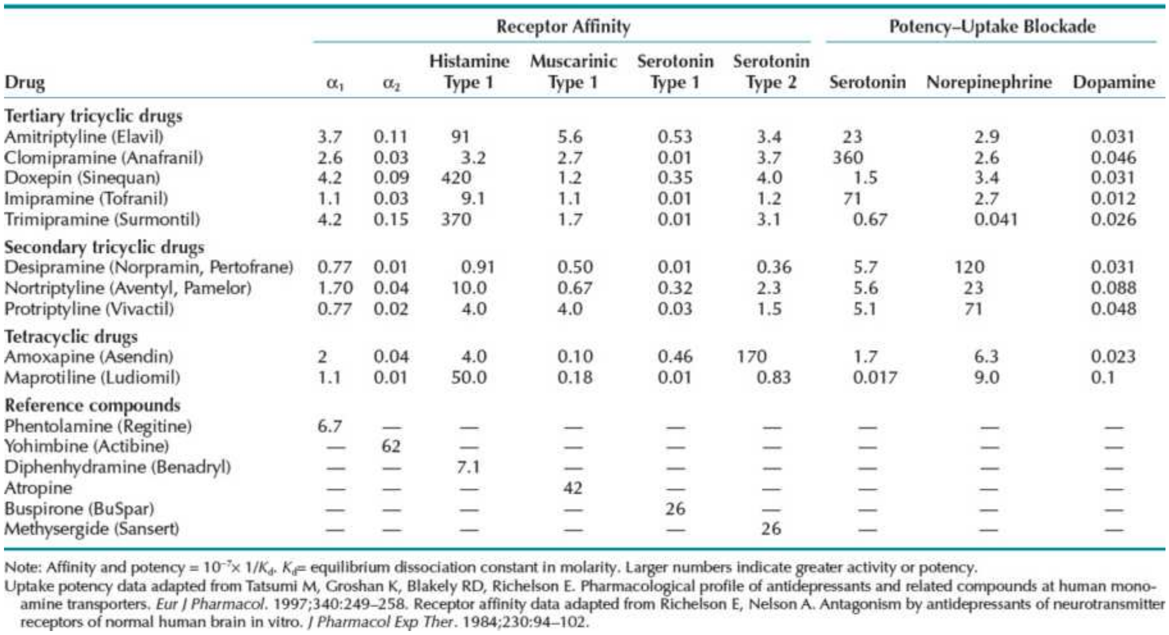

	Receptor Affinity						Potency–Uptake Blockade		
Drug	α_1	α_2	Histamine Type 1	Muscarinic Type 1	Serotonin Type 1	Serotonin Type 2	Serotonin	Norepinephrine	Dopamine
Tertiary tricyclic drugs									
Amitriptyline (Elavil)	3.7	0.11	91	5.6	0.53	3.4	23	2.9	0.031
Clomipramine (Anafranil)	2.6	0.03	3.2	2.7	0.01	3.7	360	2.6	0.046
Doxepin (Sinequan)	4.2	0.09	420	1.2	0.35	4.0	1.5	3.4	0.031
Imipramine (Tofranil)	1.1	0.03	9.1	1.1	0.01	1.2	71	2.7	0.012
Trimipramine (Surmontil)	4.2	0.15	370	1.7	0.01	3.1	0.67	0.041	0.026
Secondary tricyclic drugs									
Desipramine (Norpramin, Pertofrane)	0.77	0.01	0.91	0.50	0.01	0.36	5.7	120	0.031
Nortriptyline (Aventyl, Pamelor)	1.70	0.04	10.0	0.67	0.32	2.3	5.6	23	0.088
Protriptyline (Vivactil)	0.77	0.02	4.0	4.0	0.03	1.5	5.1	71	0.048
Tetracyclic drugs									
Amoxapine (Asendin)	2	0.04	4.0	0.10	0.46	170	1.7	6.3	0.023
Maprotiline (Ludiomil)	1.1	0.01	50.0	0.18	0.01	0.83	0.017	9.0	0.1
Reference compounds									
Phentolamine (Regitine)	6.7	—	—	—	—	—	—	—	—
Yohimbine (Actibine)	—	62	—	—	—	—	—	—	—
Diphenhydramine (Benadryl)	—	—	7.1	—	—	—	—	—	—
Atropine	—	—	—	42	—	—	—	—	—
Buspirone (BuSpar)	—	—	—	—	26	—	—	—	—
Methysergide (Sansert)	—	—	—	—	—	26	—	—	—

Note: Affinity and potency = $10^{-7} \times 1/K_d$. K_d= equilibrium dissociation constant in molarity. Larger numbers indicate greater activity or potency.

Uptake potency data adapted from Tatsumi M, Groshan K, Blakely RD, Richelson E. Pharmacological profile of antidepressants and related compounds at human monoamine transporters. *Eur J Pharmacol*. 1997;340:249–258. Receptor affinity data adapted from Richelson E, Nelson A. Antagonism by antidepressants of neurotransmitter receptors of normal human brain in vitro. *J Pharmacol Exp Ther*. 1984;230:94–102.

There are two principal metabolic pathways: demethylation of the side chain and hydroxylation of the ring structure. Demethylation converts tertiary amines to secondary amines. Examples are the conversion of amitriptyline to nortriptyline and imipramine to desipramine. This process alters the activity of the compound. The tertiary amines are relatively more serotonergic, whereas the demethylated amines are relatively more noradrenergic. Hydroxylation of the compounds also occurs and results in hydroxymetabolites. The concentration of 10-hydroxynortriptyline usually exceeds that of the parent. For desipramine, 2-hydroxydesipramine is present at levels of approximately 45% of the parent, but these ratios are quite variable. Hydroxyimipramine and hydroxyamitriptyline are present at low concentrations and are clinically unimportant. The hydroxymetabolites are then conjugated and excreted. The conjugated metabolites are not active.

Hydroxynortriptyline and hydroxydesipramine have antidepressant activity. Hydroxy-desipramine is comparable to the parent in terms of norepinephrine reuptake blockade. There are two isomers of hydroxynortriptyline: E- and Z-10-hydroxynortriptyline. E-10-hydroxynortriptyline is present at levels greater than the Z-isomer and the parent compound but is approximately 50% as potent as nortriptyline. The clinical significance of high levels of less potent hydroxynortriptyline is not entirely clear but may contribute to the therapeutic window described for this drug. The hydroxymetabolites of nortriptyline and desipramine are less anticholinergic than their parent compounds but may contribute disproportionately to the cardiac conduction abnormalities associated with tricyclic antidepressants.

The principal metabolic pathway for amoxapine is hydroxylation. 7-Hydroxyamoxapine and 8-hydroxyamoxapine are produced. These compounds differ in their clinical activity. 7-Hydroxyamoxapine has high-potency neuroleptic properties but has a short half-life. 8-Hydroxyamoxapine is metabolized more slowly and appears to contribute to the drug's antidepressant action.

The tricyclics are metabolized by the cytochrome 450 (CYP) system. The CYP 2D6 pathway is responsible for the hydroxylation of desipramine and nortriptyline. Demethylation of the tertiary amine compounds appears to involve several CYP isoenzymes, including CYP 1A2, CYP 3A4, and CYP 2C19. Approximately 5% to 10% of Caucasians are homozygous for the recessive autosomal CYP 2D6 trait, resulting in deficient hydroxylation of desipramine and nortriptyline. Approximately 50% of those of Asian descent carry the 2D6*10 allele, which results in reduced clearance by the 2D6 enzyme and is intermediate metabolizers. In addition, about 20% of Asians have a genetic polymorphism resulting in deficient CYP 2C19 metabolism of the tertiary tricyclic compounds.

The variability in plasma concentrations that results from these metabolic differences is substantial, often described as a 40-fold variation among individuals (Fig. 33.5a.3–2). Even among the extensive metabolizers, there can be moderate interindividual variability in the rates of metabolism.

The concept of *steady-state* is important for clinicians to understand if plasma concentration monitoring is used. A Steady-state is that point, on a fixed dose, at which plasma concentrations of the drug plateau. The steady-state is achieved after five times the half-life.

If blood level monitoring is used, a trough sample (before the next dose) is drawn when the patient has reached a steady-state. At steady-state, drug concentrations are relatively stable if the dose is constant, the patient is adherent, and no drugs with interactive effects are added.

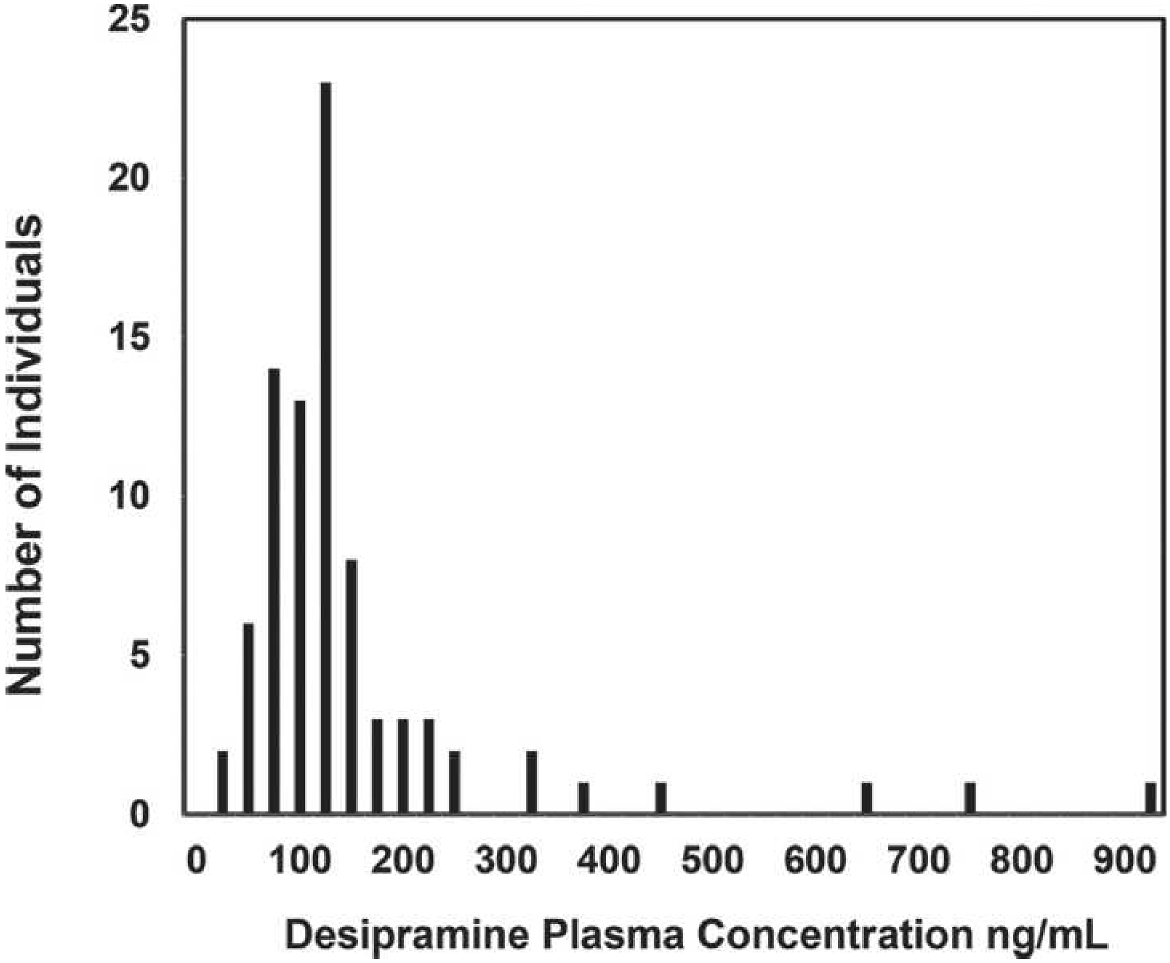

FIGURE 33.5a.3–2. Distribution of steady-state desipramine (Norpramin, Pertofrane) concentrations in 83 in-patients receiving desipramine (2.5 mg/kg/day). (Adapted from Nelson JC. Use of desipramine in depressed inpatients. *J Clin Psychiatry.* 1984;45(10 Pt 2):10–16.)

When the drug concentration is measured, the total of the free and bound tricyclic fractions are reported; yet it is the free levels that are directly related to the concentration of the drug in the cerebrospinal fluid (CSF). The free concentration is dependent on dose and hepatic clearance but not plasma protein binding. The latter is often misunderstood. If the binding protein concentration falls, the total (free and bound) concentration declines, but the absolute free concentration is unaffected.

The day-to-day biologic variability of drug concentrations is not frequently described. In outpatients, the day-to-day variability of nortriptyline levels (the coefficient of variation equals the standard deviation divided by the mean plasma concentration) has been reported to be 25%. In inpatients at steady-state, the coefficient of variation of desipramine is approximately 10% to 15%. (If the average plasma concentration is 150 ng/mL, two-thirds of samples obtained will be between 128 and 172 ng/mL.) For clinical purposes, usually, one sample is drawn, and the clinician needs to remember that, even if the laboratory error is low, the biologic variability is sufficiently great that blood levels are better viewed as estimates than precise measures.

Recently assays have become available that can genotype-specific isoenzyme pathways such as 2D6. Genotyping might be useful for predicting how an individual would metabolize a group of drugs that are cleared by a common pathway. Alternatively, a plasma drug concentration is more informative about the actual concentration of a specific drug present and accounts for variations in absorption, distribution, and adherence in addition to differences in metabolic rate.

For most tricyclics, kinetics are relatively linear, although there are some exceptions. Desipramine, for example, appears to have nonlinear kinetics in the usual dose range. The patients most vulnerable to nonlinear changes are the rapid metabolizers—patients who, at usual doses, have low blood levels. In cases of overdose, nonlinear changes are more likely to occur, and the clinician cannot assume that usual rates of drug elimination are maintained.

The pharmacokinetics of drug treatment in the older adults is complicated. A number of changes occur. For example, the ratio of fat to lean body mass increases, and cardiac output and hepatic blood flow decrease. There may be further changes associated with medical illness. Yet, the magnitude of these changes is usually relatively small in relation to the dramatic variability of hepatic metabolism. Metabolic changes with age appear to vary with the isoenzymes involved. Clearance by the CYP 3A4 pathway slows with age. Most studies of tertiary amines, such as imipramine, suggest that concentrations of these drugs are increased somewhat in the older adults. Concentrations of nortriptyline and desipramine, compounds metabolized primarily by CYP 2D6, are relatively unaffected by aging. Similarly, the relationship of drug concentrations to therapeutic effects appears to be relatively similar in younger and older adults for these agents. Renal clearance of the hydroxymetabolites does decrease with age. Concentrations of hydroxynortriptyline may be substantially elevated in older patients.

In children, the clearance of tricyclic compounds is increased. Half-lives of imipramine are shorter, and ratios of desmethylimipramine to imipramine are higher, consistent with more rapid metabolism.

Pharmacodynamics

Mechanisms of Action. The tricyclic and tetracyclic antidepressants block the transporter sites for serotonin and norepinephrine and thus reduce uptake of these amines into the presynaptic neuron (Table 33.5a.3–2). The tertiary amines have a greater affinity for the serotonin transporter. The secondary amines primarily block the uptake of norepinephrine. Because the tertiary amines are demethylated to secondary amines, during the administration of amitriptyline, imipramine, or clomipramine, secondary amines are also present; thus, serotonergic and noradrenergic effects occur. Because dopamine is inactivated by norepinephrine transporters in the frontal cortex, the tricyclics would be expected to increase dopamine concentrations in that region.

Table 33.5a.3–2.
Pharmacokinetic Parameters and Clinical Information for Tricyclic and Tetracyclic Drugs

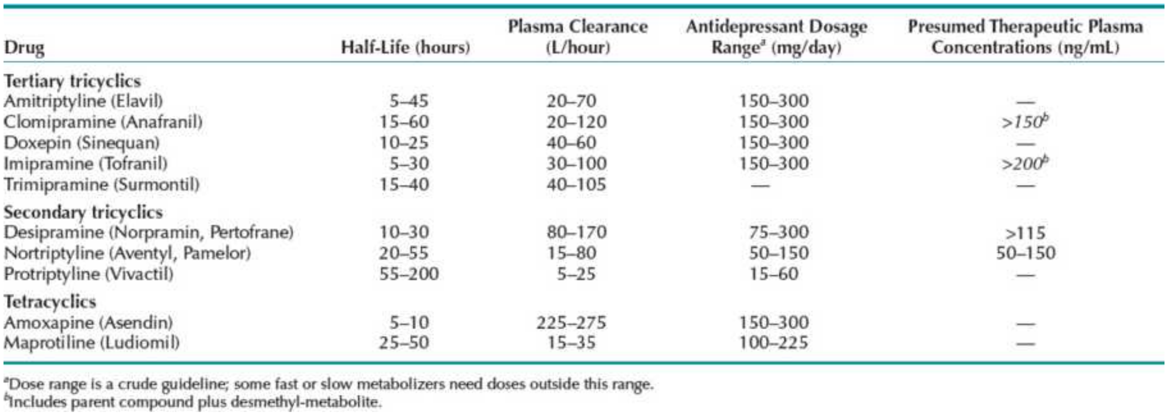

Drug	Half-Life (hours)	Plasma Clearance (L/hour)	Antidepressant Dosage Range[a] (mg/day)	Presumed Therapeutic Plasma Concentrations (ng/mL)
Tertiary tricyclics				
Amitriptyline (Elavil)	5–45	20–70	150–300	—
Clomipramine (Anafranil)	15–60	20–120	150–300	*>150*[b]
Doxepin (Sinequan)	10–25	40–60	150–300	—
Imipramine (Tofranil)	5–30	30–100	150–300	*>200*[b]
Trimipramine (Surmontil)	15–40	40–105	—	—
Secondary tricyclics				
Desipramine (Norpramin, Pertofrane)	10–30	80–170	75–300	>115
Nortriptyline (Aventyl, Pamelor)	20–55	15–80	50–150	50–150
Protriptyline (Vivactil)	55–200	5–25	15–60	—
Tetracyclics				
Amoxapine (Asendin)	5–10	225–275	150–300	—
Maprotiline (Ludiomil)	25–50	15–35	100–225	—

[a] Dose range is a crude guideline; some fast or slow metabolizers need doses outside this range.
[b] Includes parent compound plus desmethyl-metabolite.

After the initial reuptake blockade, a sequence of events occurs. The tertiary tricyclic compounds, particularly clomipramine, inhibit the uptake of serotonin and concentrations of serotonin rise. In response, the presynaptic somatodendritic serotonin autoreceptors reduce the firing rate of the presynaptic serotonin neuron, and concentrations of serotonin metabolites fall. Over a 2-week period, the presynaptic autoreceptor is desensitized, resulting in a normalization of the firing rate. At this point, serotonin transmission is enhanced.

The tricyclic agents also sensitize or upregulate *postsynaptic* 5-HT_{1A} receptors and downregulate the 5-HT_2 receptors. In addition, doxepin, clomipramine, amitriptyline, trimipramine, and nortriptyline act as antagonists at the 5-HT2 receptor. The 5-HT_2 receptors principally mediate excitatory effects, whereas the 5-HT_{1A} receptors have an inhibitory effect in the brain. Thus these two systems act in opposition. Thus, downregulation of 5-HT_2 receptors appears to facilitate the effects of serotonin at postsynaptic 5-HT_{1A} receptors. Changes in the sensitivity of postsynaptic receptors occur over a period of weeks and are more consistent with the timing of antidepressant response than the initial uptake blockade.

The sequence of events in the noradrenergic system is more complicated and less well-understood. Following reuptake inhibition, concentrations of 3-methoxy-4-hydroxyphenylglycol (MHPG), a metabolite of norepinephrine, tall rapidly within 24 hours. Presumably, this is mediated by the presynaptic somatodendritic α_2-autoreceptor, which lowers the firing rate of the presynaptic neuron. Different from the 5-HT system, the firing rate of the presynaptic noradrenergic neuron remains inhibited with chronic treatment. Norepinephrine concentrations, however, do increase at postsynaptic sites in the hippocampus and frontal cortex, and norepinephrine transmission is increased. This may be mediated by desensitization of the *terminal* α_2-autoreceptor.

During chronic treatment, the postsynaptic β-adrenergic receptor is downregulated or decreased in density. This appears to be a compensatory change. In fact, during continued administration of a noradrenergic reuptake inhibitor, the system is upregulated. This is reflected in the enhanced formation of the second messenger cyclic adenosine monophosphate (cAMP). Clinically, this is reflected as a persistent increase in heart rate.

Early theories of depression posited depletion of serotonin, norepinephrine, or both, and the tricyclics were thought to “correct” this imbalance through reuptake blockade. Subsequent theories suggested that the changes in presynaptic and postsynaptic receptors were a likely explanation for the delayed effects of these drugs in depression. Monoamine depletion studies indicated that the effects of antidepressant agents are, in part, mediated by serotonin and norepinephrine; however, they do not necessarily imply that tricyclics are correcting *deficits* in these neurotransmitter systems associated with depression.

Early theories of antidepressant action suggested that it might be possible to identify serotonergic and noradrenergic depressions based on the relative deficiencies of metabolites associated with each of these neurotransmitters and thus predict which type of antidepressant would be most effective. Several studies investigated the predictive value of MHPG levels pretreatment, but a clear predictive link with a type of antidepressant was not established. These studies were, in part, hampered by the use of agents such as amitriptyline and imipramine, which are not selective.

The possible role of the anticholinergic effects of tricyclics in the mechanism of antidepressant action has not received much attention. These effects usually have been associated with unwanted adverse effects. Yet, as early as 1972, theories about the role of the cholinergic system in depression were advanced. Scopolamine, an anticholinergic agent, has rapid antidepressant effects given intravenously. Physostigmine and donepezil, which have cholinergic effects, have been noted to exacerbate depression or increase the risk of relapse in patients with recurrent depression. Amitriptyline has anticholinergic effects at therapeutic doses, and it is likely these actions would contribute to antidepressant activity; however, it would be difficult to disentangle the contribution of these effects relative to the effects of monoamine reuptake inhibition.

Other possible mechanisms of action in depression have been explored, including the effects of tricyclics on postsynaptic signal transduction, glucocorticoid receptors, neurotrophic factors, and gene expression. The antidepressants, including tricyclics, appear to have neuroprotective effects. They raise levels of BDNF and promote neurogenesis in the hippocampus. A particularly intriguing finding is the observation of an association of decreased hippocampal volume with lifetime duration of *untreated* depression, implying that antidepressant treatment might protect against this outcome. A 2017 repositioning study screened 249 generic compounds that cross the blood–brain barrier for a variety of neurotrophic effects. Of these compounds, clomipramine appeared most promising. It inhibited iron-mediated neurotoxicity, had antioxidative properties, and reduced T-lymphocyte proliferation, among its effects. The use of clomipramine in multiple sclerosis or to reduce the “cytochrome storm” associated with COVID-19 has been suggested. Yet, at the present time, the role of these neuroprotective effects in mediating the actual antidepressant effects of tricyclics remains unclear.

The reuptake inhibition of the tricyclics and tetracyclic compounds also occurs in descending pathways of the spinal cord and contributes to analgesic effects. Animal studies suggest that serotonin and norepinephrine have independent synergistic effects on pain. In addition, the effects of tricyclics on sodium channels also contribute to their analgesic effects, although these same effects appear to be responsible for the dose-dependent delay in cardiac conduction.

The tricyclic and tetracyclic compounds have *secondary* effects on a variety of other receptors (Table 33.5a.3–2). These compounds block muscarinic receptors, producing anticholinergic effects. They block histamine-1 (H_1) receptors, and they block α_1- and α_2-adrenergic receptors. They affect fast sodium channels, and this explains their adverse cardiac effects. The potency of these various secondary effects varies considerably. Amitriptyline is the most anticholinergic. Doxepin is the most potent H_1 antagonist among the tricyclics, although it is less potent at this receptor than mirtazapine (Remeron).

Plasma Concentrations and Relation to Response. The wide interindividual variability of tricyclic plasma concentrations suggested that plasma level drug monitoring might ensure that therapeutic blood levels were achieved and might help to avoid toxic levels. In carefully selected endogenous or melancholic inpatients with major depression, treatment with adequate levels of imipramine or desipramine did result in response rates of approximately 85%. Alternatively, the careful selection of patients in these studies may result in samples not representative of typical clinical practice.

An American Psychiatric Association (APA) task force reviewing the evidence concluded that a relationship between blood levels and response had been demonstrated for imipramine, desipramine, and nortriptyline (Table 33.5a.3–1). For imipramine, drug levels greater than 200 ng/mL were more effective than lower levels. For desipramine, levels greater than 115 ng/mL were more effective. For desipramine and imipramine, blood levels of more than 300 ng/mL were more likely to be associated with serious side effects. Effective plasma concentrations were also established for nortriptyline, but the relationship appeared to be curvilinear: Plasma levels between 50 ng/mL and 150 ng/mL were more effective than levels outside this window.

For amitriptyline, it was difficult to establish a therapeutic relationship between levels and response. In part, this was related to the fact that during amitriptyline administration, amitriptyline, nortriptyline, and hydroxynortriptyline are all present and active. Amitriptyline responders usually have total amitriptyline and nortriptyline levels between 150 and 250 ng/mL, but there is no good agreement between studies. For clomipramine, blood levels of 150 to 300 ng/mL (total of clomipramine and desmethylclomipramine) have been suggested for antidepressant effectiveness. Higher doses and levels are usually used in the treatment of OCD. The data relating to blood levels and response are limited for the other tricyclic and tetracyclic compounds.

The clinical value of blood level monitoring has been the subject of much debate. Although relationships between blood levels and response have been best demonstrated in melancholic inpatients, it has proven difficult to demonstrate a relationship between blood levels and response in depressed outpatients. In this group, placebo rates are high, the difference between drug and placebo is modest, and the effect of drug treatment is harder to detect. Moderately depressed outpatients are a heterogeneous group and include patients who are not responsive to drug treatment. Finally, many studies are not designed to detect blood level–response relationships. It is logical to conclude that blood level relationships determined in severely depressed inpatients might be used as a guide for treating outpatients, but that has not been empirically confirmed.

The case study provided here gives an example of blood level monitoring.

A 74-year-old man was admitted to the hospital with severe depression. Irritability was prominent. Desipramine was started, but on 150 mg/day, he had no response whatsoever. He was adherent to treatment. Some of his symptoms seemed atypical, and the diagnosis was questioned. A blood sample revealed a very low drug concentration. In fact, the level suggested that if kinetics were linear, he would need a dose of 900 mg/day to achieve a therapeutic level. On 400 mg/day, the approved maximum dose, his plasma level was still subtherapeutic. On 600 mg, his level rose disproportionately. He achieved an adequate level and responded.

The case illustrates several issues. Blood levels can be useful for guiding dose adjustment in nonresponding patients. The extremely low plasma concentration on a conventional dose suggested he might be an ultrafast metabolizer. Nonlinear kinetics are most likely to occur in fast metabolizers. Ironically, this 74-year-old man had the lowest concentration-to-dose ratio encountered among over 400 patients treated with desipramine.

The alternative question is whether blood level monitoring might help avoid toxicity. A variety of data support this view. The risk of delirium is increased at amitriptyline plasma concentrations greater than 300 ng/mL. However, amitriptyline is the most anticholinergic tricyclic and is most likely to produce delirium. The risk of cardiac arrhythmia also is substantially increased at higher blood levels. First-degree atrioventricular block is increased with plasma concentrations of imipramine greater than 350 ng/mL. The risk of seizures also increases in patients at higher doses and, presumably, at higher blood levels, although a clear plasma level threshold has not been demonstrated for seizures. The risk of seizures for clomipramine is 1.66% with doses greater than 250 mg/day but is approximately 0.50% for doses less than 250 mg/day. In one large sample of patients with OCD at steady-state on 250 mg/day of clomipramine, the average blood level was 155 ng/mL for clomipramine and 335 ng/mL for desmethylclomipramine, or a total of 490 ng/mL. For the older drugs, amitriptyline and imipramine, the risk of seizures also appears to be increased at doses greater than 250 mg/day. For patients who are especially vulnerable to delirium, delayed cardiac conduction, or seizures, adverse events can occur at levels below those cited above. After an overdose, tricyclic blood levels greater than 1,000 ng/mL can be achieved, and the risks of delirium, stupor, cardiac abnormalities, and seizures are all substantially increased.

One of the problems in exploring the relationship of toxicity to increased blood levels is that, although the risk may increase, actual rates are still low, and large samples are required to demonstrate any difference in risk. For certain types of toxicity, there may be no early signs of impending seizures or cardiac arrhythmia to alert the clinician. Blood level monitoring might be most useful to reduce the risk of these side effects.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Design

Early in the history of tricyclics, there was interest in which patients or which conditions would respond to these medications. Yet, defining an adequate trial proved challenging. In early trials, high rates of response to placebo were noted; thus, open-label studies proved to be of limited value. Because of side effects, the tricyclics were started at a low dose and increased slowly, often resulting in subtherapeutic dosing. Subsequently, blood levels were employed to achieve adequate doses; however, the determination of effective blood levels itself proved challenging. Some early studies were quite brief (e.g., 3 to 4 weeks) and may have underestimated drug effects. As experience grew, randomized, double-blind, placebo-controlled, or comparison trials of adequate dose and duration became the yardstick by which efficacy was determined and “tricyclic responders” defined.

Interpretation

The clinical efficacy of the tricyclic and tetracyclic compounds in major depression is well-established and has been extensively reviewed by John Davis and Alexander Glassman in the fifth edition of this text and in the Agency for Health Care Policy and Research (AHCPR) Clinical Practice Guideline. Imipramine is the best-studied compound. In 44 placebo-controlled studies, imipramine was more effective than placebo in

30 studies. Overall, in studies reporting response rates ($N = 1{,}334$), 65% of the patients *completing* treatment with imipramine were substantially improved, whereas 30% of those on placebo were similarly improved. In a meta-analysis of *intent-to-treat* response rates for placebo-controlled studies of tricyclics in outpatients reported in the AHCPR guideline, response rates were 51% for imipramine and 30% for placebo. In most comparison studies, the other tricyclic and tetracyclic antidepressants have been found to be comparable to imipramine in efficacy.

In addition to efficacy in acute phase treatment of major depression, the tricyclic compounds are also effective for maintenance treatment. Several early studies demonstrated that maintenance treatment with a tricyclic would reduce the relapse rate associated with a placebo by approximately 50%. These studies, however, often used low doses. Subsequently, an important study by the Pittsburgh group found that if the full dose of imipramine used for acute treatment was continued, nearly 80% of the depressed patients maintained their improvement, whereas only 10% of those on placebo remained well at the end of a 3-year period. In this study, maintenance psychotherapy had an intermediate effect, with approximately 30% of the patients remaining well. This study again confirmed the value of maintenance treatment and underscored the importance of continuing the full dose of imipramine. The magnitude of the findings may, in part, relate to the sample selected, namely, patients with recurrent depression who had been symptom-free between prior episodes. In clinical practice, if patients with chronic depression, prior residual symptoms, or comorbid psychiatric disorders are included, the number of patients who remain in remission may be lower.

Efficacy in Subtypes

The efficacy of the tricyclic compounds varies in different subtypes of depression. Kuhn thought it was essential to establish the efficacy of imipramine in *endogenous* inpatients. The tricyclic compounds have been frequently studied in severely depressed inpatients and were found to be effective. It has been suggested that tricyclics may be more effective than SSRIs in severe major depressive disorders with melancholic features. The issue remains unsettled, in part, because of the paucity of studies on the SSRIs in severe melancholic depression. In addition, it was unclear if the suggested advantage of some tricyclics in severe depression was unique to tricyclics as a class or applied to those tricyclics, particularly clomipramine, that had potent effects on both serotonin and norepinephrine.

The tricyclic and tetracyclic compounds have often been studied in *anxious depression*. Three of the compounds—doxepin, amoxapine, and maprotiline—have FDA approval for use in patients with depression and symptoms of anxiety. For many years, clinical lore suggested that amitriptyline was most effective for anxious depression. Direct comparison studies, however, found little evidence that these compounds differ in their efficacy for the treatment of anxiety associated with depression. Alternatively, three studies found that, among depressed patients, those with high levels of anxiety responded less well to amitriptyline, imipramine, or desipramine than those with lower levels of anxiety. Yet, these drugs are still more effective than placebo in anxious-depressed patients, and it is not established that other classes of antidepressants are more effective in anxious-depressed patients.

In patients with *atypical features* (defined by DSM-IV criteria), imipramine is more effective than placebo but significantly less effective than the MAOI phenelzine. Little information is available regarding comparisons of the tricyclics and the SSRIs in atypical depression.

Major depression with *psychotic features* is less responsive than nonpsychotic depression to tricyclic antidepressants used alone. Several open studies and one prospective study found that the tricyclics, when combined with an antipsychotic, were effective in psychotic depression. Yet other trials did not find an antidepressant/antipsychotic combination more effective than nortriptyline, imipramine, or trimipramine monotherapy. One double-blind study compared the efficacy of amoxapine with the combination of perphenazine (Trilafon) and amitriptyline, finding amoxapine to have comparable efficacy in psychotic depression. It was suggested that the antipsychotic action of 7-hydroxyamoxapine might explain the efficacy of this compound in psychotic depression.

In controlled studies of patients with *bipolar depression*, imipramine was found to be less effective than tranylcypromine (Parnate) or fluoxetine (Prozac). These studies and others suggest that tricyclics appear to be more likely to induce mania than serotonergic agents or bupropion (Wellbutrin).

In the AHCPR review of drug trials in *late-life depression*, tricyclic antidepressants appear to be effective, but, overall, drug and placebo response rates appear to be lower than in non-older adult patients. Efficacy in the older adults may have, in part, reflected cautious dosing because of safety and tolerance concerns. In children and adolescents, tricyclic antidepressants have not demonstrated superiority over placebo. An FDA review of clinical trials of second-generation antidepressants revealed an increased risk of suicidal thinking and suicide-related behaviors in children, adolescents, and young adults under the age of 25. Although this risk has not been established for tricyclics, the warning has been extended to all antidepressants.

Tricyclic antidepressants are effective in *dysthymia*. Controlled trials suggest both imipramine and desipramine are effective for acute treatment of dysthymia and for maintenance treatment.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Somatic symptoms that occur during treatment are the result of the interaction of direct effects of the medication on specific organs and indirect effects of the medication on depression and its associated somatic symptoms. Depression is accompanied by a variety of somatic symptoms. Headache, constipation, poor memory, drowsiness, and difficulty sitting still have been observed in more than 50% of untreated inpatients with major depression. Other symptoms, such as palpitations, dry mouth, tremors, light-headedness, and poor coordination, are reported by more than 40% of depressed patients before treatment. The determination of whether a somatic symptom is an adverse drug effect or a symptom of depression involves a judgment about whether the symptom is new or has worsened during drug treatment. Severely depressed patients may not remember the severity of a symptom prior to treatment accurately. During treatment, the strongest predictor of overall somatic symptom severity is the level of depression at the time of assessment.

Antidepressant drugs do have adverse effects. The in vitro potency or affinity of antidepressant compounds for various receptor sites (Table 33.5a.3–1) is one method for comparing the likelihood that various agents will produce side effects. The other pharmacologic factor that needs to be considered is the bioavailability of the drug.

Central Nervous System

In the CNS, the principal action of the tricyclic and tetracyclic agents is to alleviate depression. These agents do not have a direct mood-elevating effect. They do not produce euphoria. Rather, they reduce the symptoms of depression. Apparently related to their effects on brain amines, these compounds can cause or aggravate psychotic symptoms and can induce mania.

The central anticholinergic and antihistaminic effects of tricyclics and tetracyclics can produce confusion or delirium. The incidence of delirium is dose-dependent and increases at blood levels greater than 300 ng/mL and may rise to 67% in patients with blood levels greater than 450 ng/mL and in patients receiving tertiary amines, particularly amitriptyline. Patients with dementia or early Alzheimer disease are more vulnerable to the anticholinergic effects of the tricyclics, and delirium can occur in these patients at lower drug concentrations.

Seizures can occur with all the tricyclic and tetracyclic agents. Seizures are dose- and blood-level related. For clomipramine, the risk for seizures is reported to be 0.5% in doses up to 250 mg/day. At doses greater than 250 mg/day, the seizure risk increases to 1.67%. For maprotiline, the overall risk of seizures is reported to be 0.4%, but, again, the risk of seizures increases at doses greater than the maximum recommended dose of 225 mg/day. Determination of seizure risk for some of the older compounds was hampered by less rigorous data collection during drug development. In one meta-analysis of data for imipramine, patients receiving less than 200 mg had a seizure rate of 0.1%. In patients receiving more than 200 mg/day, the rate was 0.6%. In patients receiving 250 to 450 mg/day, reported rates varied from 1% to 4%, but these samples were often small. Thus, these rates had wide confidence intervals. Seizure rates for the secondary amines have not been well-described. The mechanism by which tricyclics produce seizures is not well-understood, although it is proposed that antidepressant drugs may act at the GABA receptor chloride–ionophore complex, inhibiting chloride conductance and consequently inducing convulsions. The risk of convulsions is clearly increased in patients with predisposing factors, including a prior history of seizures, brain injury, and concurrent use of neuroleptics. The risk of seizures is substantially increased after overdose.

A fine rapid tremor can occur with the tricyclic agents, tends to be dose-dependent, and may be a clinical indication of an elevated blood level.

Because the 7-hydroxymetabolite of amoxapine has neuroleptic properties, administration of amoxapine carries the potential risk of a neuroleptic malignant syndrome, which has been reported, and tardive dyskinesia.

A 73-year-old male was referred for evaluation of his antidepressant. He had had a good response to desipramine, but it had taken some time to achieve it. He now had a fine rapid tremor that was quite bothersome. Both the patient and his physician were concerned that reducing the dose would compromise antidepressant efficacy. A blood sample revealed a desipramine concentration of over 300 ng/mL. His dose was adjusted to achieve a level of about 150 ng/mL. His tremor resolved, but the antidepressant effect was maintained.

Several of the tricyclic compounds and maprotiline have clinically significant antihistaminic effects. Doxepin is the most potent H_1-receptor blocker of the tricyclics, although mirtazapine and olanzapine are more potent. Central H_1-receptor blockade can contribute to sedation and delirium. Antihistaminic effects also appear to be related to increased appetite and associated weight gain.

Autonomic Nervous System

The tricyclics block muscarinic receptors and can cause a variety of anticholinergic side effects. These include dry mouth, constipation, blurred vision, urinary hesitancy, and an ocular crisis in patients with narrow-angle glaucoma. The tricyclic and tetracyclic compounds vary substantially in their muscarinic potency (Table 33.5a.3–2). Amitriptyline is the most potent, followed by clomipramine. Desipramine is the least anticholinergic. Amoxapine and maprotiline also have minimal anticholinergic effects. Anticholinergic effects can contribute to tachycardia, but tachycardia commonly occurs in patients receiving desipramine, which has minimal anticholinergic effects. It seems likely that noradrenergic effects also contribute to increased pulse. Increased sweating is another secondary effect of the tricyclic compounds that can occur. The mechanism for this symptom is unclear but may be associated with noradrenergic effects.

Cardiovascular Effects

Orthostatic hypotension is one of the most common reasons for discontinuation of tricyclic antidepressant treatment. Orthostatic hypotension can occur with all the tricyclics but is less pronounced with nortriptyline. It has been suggested that orthostatic hypotension is caused by α_1-adrenergic blockade; however, it is the postural reflex that is primarily affected. Resting supine blood pressure may be unaffected or may be elevated. Orthostatic hypotension is most likely to occur or is most severe in patients who have pre-existing orthostatic hypotension.

Desipramine has been reported to raise supine blood pressure in younger patients, although it is not clear that this effect is limited to that age group. This effect may be like that reported for venlafaxine (Effexor).

Tachycardia occurs with all tricyclics, not just the more anticholinergic agents. Supine and postural pulse changes can occur. Patients do not accommodate the pulse rise, which can persist for months. Tachycardia is more prominent in younger patients, who appear to be more sensitive to sympathomimetic effects, and is one of the most common reasons for drug discontinuation in this age group.

The exploration of the effect of tricyclic antidepressants on cardiac conduction has received much attention. *Cardiac arrhythmia* is the principal cause of death after overdose. As a result, for many years, there was great concern about the use of tricyclic antidepressants in patients with heart disease. The effect of these agents has now been well-described. These agents have type I antiarrhythmic qualities or quinidine (Cardioquin)-like effects. Apparently, through inhibition of Na^+:K^+ adenosine triphosphatase (ATPase), the tricyclics stabilize electrically excitable membranes and delay conduction, particularly His-ventricular conduction.

At therapeutic blood levels, tricyclics can have beneficial effects on ventricular excitability. Alternatively, in patients with pre-existing conduction delay, tricyclic antidepressants can further delay conduction and cause heart block. High-drug plasma levels further increase the risk of cardiac toxicity. For example, first-degree atrioventricular heart block is increased with imipramine plasma concentrations greater than 350 ng/mL and is increased more than 30-fold in patients with plasma concentrations greater than 450 ng/mL.

The tricyclic antidepressants do not appear to adversely affect cardiac contractility or cardiac output. Studies using radionuclide angiography indicate no adverse effect of imipramine or doxepin on cardiac output, even in patients with diminished left ventricular ejection fractions. However, it is noted that severe orthostatic hypotension was common in these patients.

Another potential hazard of tricyclic treatment was suggested by the Cardiac Arrhythmia Suppression Trial (CAST) studies, which found that type I antiarrhythmic drugs are given after myocardial infarction increased the risk of sudden death. Alexander Glassman and associates suggested that tricyclics may pose similar risks. Thus, the clinician considering the long-term use of tricyclics in patients with ischemic heart disease needs to weigh this possible risk against possible benefits.

Hepatic Effects

Drug-induced hepatitis can occur with these agents. Mild increases in liver enzymes are not uncommon and usually can be monitored safely over a period of days or weeks without apparent harmful consequences. However, acute hepatitis can occur quickly and is associated with high enzyme levels; for example, serum aspartate aminotransferase (AST) levels may be elevated 20-fold. The rise in enzymes can be hepatocellular or cholestatic in form and may precede clinical symptoms, especially in the hepatocellular form. This is a dangerous and potentially fatal condition. The antidepressant must be discontinued. The etiology is not well-established but, in some cases, appears to be a hypersensitivity reaction. The risk of severe drug-induced hepatitis is not well-established but appears to be between 1 in 1,000 and 1 in 100.

THERAPEUTIC INDICATIONS

Indications

The tricyclic and tetracyclic compounds have a variety of clinical uses. In fact, it is interesting to consider what these agents might have been called had their effectiveness in panic disorder, OCD, or pain been described before their antidepressant action was observed.

The FDA approved all tricyclic and tetracyclic compounds discussed in this section for use in the United States for the treatment of depression, with the exception of clomipramine. In Europe, clomipramine is also used for depression and, in fact, is regarded by many as the most potent antidepressant. Clomipramine acts on serotonin and norepinephrine, and the empirical data suggest that it is more effective than selective serotonin agents in severely depressed inpatients.

Although a variety of antidepressants appear to be effective in depressive disorders, this is not the case for OCD. In this instance, serotonergic antidepressants, such as clomipramine, appear to be substantially more effective than noradrenergic tricyclics, such as desipramine. Among the tricyclic drugs, clomipramine is the only agent approved by the FDA for use in OCD.

Other Indications

Tricyclic agents have been used for a variety of other clinical indications without FDA approval. Tricyclic antidepressants are effective in panic disorder; in fact, imipramine was the first drug found to be effective in these patients.

Amitriptyline has been a popular drug for the treatment of chronic pain syndromes and migraine headache prophylaxis. Over 60 placebo-controlled clinical trials of tricyclics or tetracyclics have been conducted in various chronic pain syndromes, including headache, back pain, fibromyalgia, gastrointestinal discomfort, idiopathic pain, and tinnitus. In fact, the magnitude of their effects on chronic pain appears greater than that in depression. For example, in meta-analyses of placebo-controlled pain trials, one patient out of three treated would benefit versus one out of six patients treated in depression trials. The analgesic effects appeared, at least in part, to be independent of antidepressant effects. In humans, comparison studies suggested tricyclics with effects on both serotonin and norepinephrine were more effective than agents selective for one neurotransmitter. During treatment of pain, the dose of amitriptyline ranged from 25 mg/day to 150 mg/day, with 75 mg/day being an approximate median dose. This dose is lower than that in depression, and effects occur more quickly, usually within the first 2 weeks. These differences suggest a different mode of action of amitriptyline in pain syndromes. The tricyclics appear to be of some benefit in chronic fatigue syndrome.

For many years, desipramine was commonly prescribed for ADHD in children. Reports of cases of sudden death, apparently of cardiac origin, in children younger than 12 years of age substantially reduced the use of this drug in younger patients. Imipramine has been used with FDA approval for the treatment of nocturnal enuresis in children. Usual doses are 25 to 50 mg at bedtime. It is doubtful, however, that the risk of cardiac problems would be substantially less with imipramine than with desipramine in young children.

Nortriptyline has been shown to be helpful for smoking cessation in controlled studies, but it is not approved for use in this syndrome.

Taking advantage of its antihistaminic effects, topical doxepin (Zonalon) is FDA-approved for use in pruritus associated with atopic dermatitis and lichen simplex chronicus. Doxepin (Silenor) is approved for use in insomnia at doses of 3 or 6 mg/day.

Tricyclic antidepressant drugs have been extensively studied in schizophrenic patients. However, in the absence of a major depressive syndrome, they appear to be of limited value, and their use for the treatment of negative symptoms has been eclipsed by the introduction of atypical neuroleptics.

PRECAUTIONS AND ADVERSE REACTIONS

Central Nervous System Effects

The central effects of tricyclics can produce delirium and seizures. These adverse effects are more frequent at blood levels greater than 300 ng/mL and are more common with amitriptyline. Recognition of delirium is important because the signs of delirium may be mistaken for worsening depression. Patients with concurrent dementia are particularly vulnerable to the development of delirium. Intramuscular (IM) or intravenous (IV) physostigmine (Eserine Salicylate) can reverse or reduce the symptoms of delirium. However, the short duration of action of this agent makes its continued use difficult. Because the risk of delirium and seizures increases at higher tricyclic plasma concentrations, some experts advocate routine blood level monitoring, especially in patients at increased risk, for example, older adults, patients with dementia, or patients with a history of seizures.

Tricyclic antidepressants can cause a fine rapid tremor. This adverse event is often associated with elevated plasma concentrations and may improve with dose reduction.

Because the 7-hydroxymetabolite of amoxapine has neuroleptic effects, this agent can potentially cause tardive dyskinesia and neuroleptic malignant syndrome. Although the occurrence of these adverse events is rare, the seriousness of the risk serves as a relative contraindication to this product unless the clinical condition warrants the use of an agent with antipsychotic properties.

Autonomic Effects

Tricyclics can cause a variety of anticholinergic side effects, including dry mouth, constipation, visual changes, urinary hesitancy, delirium, and an ocular crisis in patients with narrow-angle glaucoma. The use of nortriptyline or desipramine, which are less anticholinergic, can help reduce the likelihood of these problems.

Usually, anticholinergic effects are annoying but not serious. In some circumstances, however, these effects can be severe. An ocular crisis in patients with narrow-angle glaucoma is an acute condition associated with severe pain. Urinary retention can be associated with stretch injuries to the bladder. Constipation can progress to severe obstipation. (Paralytic ileus has been described but is rare.) Medication must be discontinued, and appropriate supportive measures must be instituted. Older adult patients appear most vulnerable to this adverse consequence. The frequency of severe anticholinergic adverse reactions is increased by concomitant neuroleptic administration.

Some anticholinergic effects may benefit from other interventions. Patients with urinary hesitancy may benefit from the use of bethanechol (Urecholine) at a dose of 25 mg three or four times daily. Patients with constipation often benefit from regular use of stool softeners. Patients with narrow-angle glaucoma who are receiving pilocarpine (Pilocar) eye drops regularly can be treated with a tricyclic, as can those who have had an iridectomy. Tricyclic agents do not affect patients with chronic open-angle glaucoma.

Cardiovascular Effects

Orthostatic hypotension is the most common reason for discontinuation of tricyclic antidepressant treatment. It is most likely to be problematic in patients with pre-existing orthostatic hypotension. Older adults are more likely to have pre-existing hypotension and are also more vulnerable to the consequences of orthostatic hypotension—falls and hip fractures. Orthostatic hypotension can occur at blood levels that are relatively low. Although the subjective experience of light-headedness usually improves with time, the actual orthostatic blood pressure change does not accommodate within a reasonable period (e.g., 4 weeks). Consequently, unless the plasma level is elevated and the dose can be reduced, patients who experience serious symptomatic orthostatic hypotension may not be treatable with a tricyclic antidepressant. Fludrocortisone (Florinef) has been used to raise blood pressure, but it is not very effective. If patients are taking an antihypertensive, it may be possible and helpful to reduce these agents. Orthostatic hypotension is less common with nortriptyline.

Tachycardia occurs with all tricyclics, not just the more anticholinergic agents. The supine pulse rate commonly increases by 8 to 10 beats/min. The standing pulse can be markedly elevated. Patients do not accommodate the pulse rise, which can persist for months. Although an 8 to 10 beat/min increase in heart rate may be subjectively tolerated by the patient, in the older adults, a sustained increase in cardiac work can be detrimental. Tachycardia is more prominent in younger patients, who appear to be more sensitive to the sympathomimetic effects of these agents, and is one of the most common reasons for drug discontinuation in this age group. Propranolol (Inderal) at low doses may help to reduce tachycardia unless it is associated with significant hypotension.

Cardiac Conduction Effects

Tricyclic antidepressants have type I antiarrhythmic effects. In patients without conduction delay, tricyclics are well-tolerated. They may be helpful in patients with premature ventricular beats. However, in patients with pre-existing conduction delay, tricyclic antidepressants can further delay conduction and can cause heart block. Patients with a QTc interval of 450 msec or greater are at increased risk and are not candidates for tricyclic antidepressant treatment. The risk of cardiac toxicity is increased in patients with high plasma levels and greatly increased with overdose.

Hepatic Effects

Mild increases in liver enzymes usually can be monitored safely over a period of days or weeks without apparent harmful consequences. However, acute hepatitis develops quickly, over 3 to 4 days, and is associated with high enzyme levels, for example, a 20-fold increase in AST levels. This can be life-threatening, and the antidepressant must be discontinued. If it is clear which drug caused hepatitis, that drug should not be introduced again because the next reaction may be more severe. However, it is not uncommon for the patients to be receiving several medications so the offending agent is hard to identify. If a random blood test indicates mildly elevated liver enzymes, enzyme levels can be followed for a few days to distinguish the more severe form of hepatitis from mild enzyme elevations.

Other Adverse Effects

Sexual dysfunction is a potential problem with a variety of antidepressant drugs but appears to be more frequent with clomipramine. Unlike many of the side effects that are more likely to be reported early in treatment, this side effect may be more bothersome to the patient in the second or third month of treatment after the depression has improved. Numerous pharmacologic antidotes have been described but without much success.

Tricyclic antidepressants can cause allergic skin rashes. They are sometimes associated with photosensitivity reactions. Various blood dyscrasias also have been reported, but fortunately, these are very rare.

Overdose

Because the risk of overdose is increased in depressed patients, the target population for which antidepressants are used, the lethality of antidepressant drugs in overdose is of great concern. A tricyclic overdose of 10 times the daily dose can be fatal. Death most commonly occurs because of cardiac toxicity. However, seizures and respiratory depression can occur. Amitriptyline is one of the leading causes of death by overdose for any single drug in the United States because of its high mortality risk and common use. A 15-year study (2000 to 2014) of single drug ingestions found more deaths from amitriptyline than all other TCAs combined and all second-generation antidepressants combined. The mortality risk with the TCAs was 25-fold higher than with SSRIs. The other tricyclic and tetracyclic compounds are also dangerous in overdose, although their use has diminished. Desipramine has the highest fatality rate (deaths per ingestion) among the tricyclics. Amoxapine has been reported to produce high rates of seizure in overdose. The morbidity and mortality risks for clomipramine were lower than for other TCAs. Because the second-generation antidepressants are safer in overdose, subtle differences in lethality between the tricyclic and heterocyclic compounds have become relatively unimportant.

Teratogenicity and Pregnancy-Related Risks

A definitive link between the tricyclic compounds and tetracyclic compounds and teratogenic effects has not been established, although isolated reports of morphogenesis have been reported. Tricyclics cross the placenta, and neonatal drug withdrawal can occur. This syndrome includes tachypnea, cyanosis, irritability, and poor sucking reflex. If possible, tricyclic and tetracyclic medications should be discontinued 1 week before delivery. Recently, norepinephrine and serotonin transporters have been identified in the placenta and appear to play an important role in the clearance of these amines in the fetus. The understanding of the effects of reuptake inhibitors on these transporters during pregnancy is limited, but one study compared intelligence and language development in 80 children exposed to tricyclics during pregnancy with 84 children exposed to other nonteratogenic agents and found no deleterious effects of the tricyclics. The tricyclics are excreted in breast milk at concentrations similar to plasma. The actual quantity delivered, however, is small so that drug levels in the infant are usually undetectable or very low. Because the risk of relapse is a serious concern in patients with recurrent depression, and these risks may be increased during pregnancy or the postpartum period, the risks and benefits of continuing or withdrawing treatment need to be discussed with the patient and weighed carefully.

DRUG INTERACTIONS

There are two principal types of drug interactions: pharmacodynamic and pharmacokinetic. Pharmacodynamic interactions are those in which the actions of two drugs interact. The effects may be deleterious. Because the tricyclics have acute effects on amine uptake, they are potentially dangerous when given with the MAOIs. The most dangerous sequence is to give a large dose of a tricyclic to a patient already on an MAOI. This can result in a sudden increase in catecholamines and a potentially fatal hypertensive reaction. If used cautiously, these two compounds can be used together to treat patients with refractory depression. Treatment is begun with lower doses, and the two compounds are started together, or the tricyclic is started first. Once begun, coadministration may reduce the risk of tyramine reactions; however, because the protective effect is variable and unpredictable, the usual MAOI diet is maintained.

Other pharmacodynamic interactions can occur. By blocking the transporters, the tricyclics block the uptake and thus interfere with the action of guanethidine (Ismelin). Quinidine poses two problems. First, the tricyclics have quinidine-like effects. Thus, their effects on cardiac conduction are potentially additive with those of quinidine. In addition, quinidine is a potent enzyme inhibitor that can raise tricyclic levels, further adding to the problem.

Pharmacokinetic Interactions

Several drugs can block the metabolic pathways of the tricyclics, resulting in higher and potentially toxic levels. Desipramine has been of particular interest because it is metabolized via the CYP 2D6 isoenzyme, and there are no major alternative pathways. Inhibition of CYP 2D6 can result in high desipramine plasma levels. Quinidine, mentioned previously, is a potent CYP 2D6 inhibitor. Several psychotropic drugs are 2D6 inhibitors, including fluoxetine, paroxetine (Paxil), duloxetine (Cymbalta), and bupropion. For example, fluoxetine 20 mg/day, on average, raises desipramine levels three- to fourfold in extensive metabolizers. In slow metabolizers, enzyme inhibitors have less of an effect because the patient is already deficient in this enzyme. At moderate doses, sertraline (Zoloft), citalopram (Celexa), and escitalopram (Lexapro) have less potent effects on CYP 2D6, but at high doses may have appreciable effects in some patients. Venlafaxine, nefazodone (Serzone), and mirtazapine appear to have minimal effects on CYP 2D6. It would be expected that CYP 2D6 inhibitors would also block nortriptyline metabolism, but the magnitude of this interaction has not been well-studied.

The other group of drugs commonly used in psychiatry, which interact with the tricyclics, is the antipsychotic drugs. Chlorpromazine and perphenazine, at antipsychotic doses, inhibit CYP 2D6. The magnitude of the effect appears to be somewhat less than that of fluoxetine. For example, perphenazine, on average, raises desipramine levels twofold, but this effect varies with dose and varies with the neuroleptic used.

Because the tertiary tricyclics are metabolized by several pathways (CYP 1A2, CYP 3A4, and CYP 2C19), inhibition of one of these pathways has less of an effect on these compounds. Drug interactions with the tertiary amines do occur but appear to be less robust than that for desipramine.

Enzyme induction also occurs. Unlike enzyme inhibition, which occurs quickly, enzyme induction requires the synthesis of a new enzyme. As a result, the full effect of an enzyme inducer may take 2 to 3 weeks to develop and, similarly, if the inducer is discontinued, it takes 2 to 3 weeks to dissipate. Barbiturates and carbamazepine (Tegretol) are two potent inducers commonly encountered in psychiatry. Although CYP 2D6 is a noninducible isoenzyme, phenobarbital (Luminal) reduces the availability of desipramine by approximately 40%. Apparently, CYP 3A4 is induced, becomes a more important pathway, and the metabolism of desipramine is increased. It can be difficult to attain an effective blood level with a tricyclic in the presence of barbiturates or carbamazepine. Phenytoin (Dilantin) is also an enzyme inducer, but the magnitude of its effects on the tricyclics appears to be less dramatic. Nicotine induces the CYP 1A2 pathway and may affect the tertiary tricyclics, but the secondary tricyclics, such as desipramine, are less affected.

Alcohol has a complicated interaction with tricyclics. Acute ingestion of alcohol can reduce first-pass metabolism, resulting in higher tricyclic levels. Because of the association of alcohol use with overdose, this can be an important factor resulting in higher tricyclic levels after overdose. Alternatively, chronic use of alcohol induces hepatic isoenzymes and may lower tricyclic levels.

Although most of the known drug-drug interactions have focused on interactions involving the CYP450 system, there is a growing body of literature examining the role of P-GP in drug interactions. P-GP plays an active role in the blood–brain barrier resulting in the efflux of some drugs from the CNS. Several of the tricyclics—amitriptyline, desipramine, doxepin, imipramine, nortriptyline, and trimipramine—are P-GP substrates. Drugs such as verapamil and quinidine inhibit P-GP, potentially leading to increased CNS concentrations of P-GP substrates. Some antidepressants, such as sertraline and paroxetine, are also P-GP inhibitors with effects equivalent to quinidine. The clinical consequences of these potential interactions in humans are not well established, but inhibition of P-GP may increase efficacy but also increase adverse effects.

LABORATORY INTERFERENCES

The tricyclic compounds are present at low concentrations (ng/mL) and are not likely to interfere with other laboratory assays. It is possible that they may interfere with the determination of conventional neuroleptic blood concentrations because of their structural similarity and the low concentrations of some neuroleptics.

Two types of assays are performed for the tricyclic and tetracyclic antidepressants. After an overdose, an immunoassay is commonly performed to determine if a tricyclic is present and to estimate the level. This assay is sensitive, simple to perform, and can be done quickly. The immune assays, however, are not specific, can cross-react with metabolites and other drugs, and thus are not precise quantitative measures. For therapeutic drug monitoring, a quantitative measure is needed, and a slower turnaround time for the laboratory is usually acceptable. For this purpose, high-performance liquid chromatography (HPLC) is sufficiently precise and is usually employed. In a competent laboratory, the coefficient of variation is less than 10%. This assay is relatively specific; however, other drugs can interfere. Because there are many modifications of the HPLC technique, and because the drugs that interfere depend on these modifications (the extraction method, the column used), the interfering drugs vary among labs. A clinician who frequently obtains tricyclic plasma levels is best advised to use one laboratory and to find out from that laboratory what agents interfere with their assay.

DOSAGE AND ADMINISTRATION

At one time, amitriptyline and imipramine were the two most commonly prescribed antidepressants in the United States, but because of their anticholinergic and antihistaminic side effects, their use declined, and nortriptyline and desipramine became more popular. Nortriptyline has the least effect on orthostatic hypotension, and desipramine is the least anticholinergic. Because therapeutic plasma concentrations have been established for nortriptyline and desipramine, blood level monitoring can be used to ensure that a therapeutic level is achieved.

Most studies of nortriptyline find that the average dose needed to reach a blood level of 100 ng/mL is approximately 75 mg/day; however, there is considerable variability. For desipramine, the average dose required to reach a target level of 150 ng/mL is 3.0 mg/kg or 180 mg/day for a 60-kg patient. Nortriptyline and desipramine can be adjusted to a full dose quickly, but the clinician needs to assess factors such as medical health, age, other medications present, and the severity of the depression when deciding how quickly to adjust the dose. The dose can be given once a day at bedtime and, if tolerated, can be increased every 3 days. The target dose can then be determined in one of two ways. An average dose can be targeted (e.g., nortriptyline 75 mg/day), and blood levels employed only if the patient fails to respond. Alternatively, a blood level can be drawn after 5 to 7 days on a moderate dose, 50 to 75 mg/day, and used to guide further dose adjustment for the trial. Because the kinetics of nortriptyline are linear, a proportional adjustment in dose can be made to reach a target level of 100 ng/mL. The advantage of the latter method is that one drug trial at an adequate blood level should establish if the drug is effective. Because nortriptyline has a therapeutic window, if the plasma level is near the upper limit of 150 ng/mL, a reduction in dose could be considered. Desipramine dose can also be adjusted relatively proportionately, with the exception that patients with low initial plasma levels (rapid metabolizers) are likely to have nonlinear changes. Plasma concentration–guided dose adjustment should reduce the time to achieve an effective dose (as opposed to a series of trials at gradually increasing doses) and should be of benefit in severe depression. Once the patient is at an adequate dose, treatment should then be continued for an adequate duration. At the time of the initial clinical trials of the tricyclics, trial duration was often short, 4 to 6 weeks. Recently, trials of up to 12 weeks have been advocated to achieve remission. Although it was suggested that tricyclics were more rapid-acting than the SSRIs, a 12-week comparison trial of nortriptyline and sertraline in late-life depression found no difference in the timing of the response. A retrospective analysis of that data concluded that for both drug classes, a 12-week trial was useful to achieve remission, but that patients showing minimal improvement after 4 to 6 weeks were not likely to remit, and other treatments should be considered. Doses and therapeutic plasma concentrations for other tricyclic and tetracyclic compounds are shown in Table 33.5a.3–1.

If patients experience side effects, this approach may be altered. If patients on desipramine report insomnia, the dose can be given early in the day. Sedation and anticholinergic effects may diminish with time, and gradual dose adjustment may help manage these effects. Although subjective light-headedness may improve with time, blood pressure and pulse changes do not usually accommodate during the first 4 weeks, and gradual dosing does not help avoid these effects. In patients at high risk for side effects, the dose can be increased gradually to assess tolerance. If anticholinergic side effects occur, allowing more time between dose increases allows the patient to accommodate these side effects.

Prospective Dosing Techniques

Conventional dosing requires the administration of a given dose for a long enough period to determine if that dose is effective. Sometimes two or even three trials are needed to determine the effective dose. Once therapeutic blood levels were established for the tricyclics, it was possible to adjust dosage quickly to achieve an effective plasma concentration and determine efficacy after one trial. In clinical practice, plasma level monitoring was most commonly employed in patients whose depression was not responding to a reasonable dose.

Exceptions

In older adult depressed patients, gradual dosing may be required to assess tolerance, but studies of desipramine and nortriptyline indicate that the final doses and blood levels required for response are similar to those needed in younger patients.

In panic patients, lower starting doses are used, and the dose is increased slowly to avoid exacerbation of panic attacks. In OCD, higher doses of clomipramine are used than those used in depression. In part, this is because OCD responds more slowly and less completely than depression, and, thus, the dose is raised to ensure that the most effective treatment has been given.

If amitriptyline is used for pain, lower doses (e.g., 75 mg/day) are used. When effective, the action of the drug is observed within the first 2 weeks.

Table 33.5a.3–3.
Tricyclic and Tetracyclic Drug Preparations

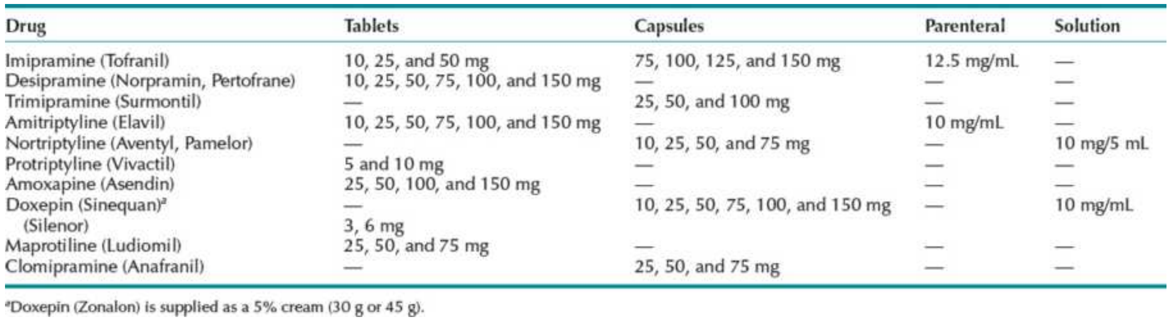

Drug	Tablets	Capsules	Parenteral	Solution
Imipramine (Tofranil)	10, 25, and 50 mg	75, 100, 125, and 150 mg	12.5 mg/mL	—
Desipramine (Norpramin, Pertofrane)	10, 25, 50, 75, 100, and 150 mg	—	—	—
Trimipramine (Surmontil)	—	25, 50, and 100 mg	—	—
Amitriptyline (Elavil)	10, 25, 50, 75, 100, and 150 mg	—	10 mg/mL	—
Nortriptyline (Aventyl, Pamelor)	—	10, 25, 50, and 75 mg	—	10 mg/5 mL
Protriptyline (Vivactil)	5 and 10 mg	—	—	—
Amoxapine (Asendin)	25, 50, 100, and 150 mg	—	—	—
Doxepin (Sinequan)[a]	—	10, 25, 50, 75, 100, and 150 mg	—	10 mg/mL
(Silenor)	3, 6 mg			
Maprotiline (Ludiomil)	25, 50, and 75 mg	—	—	—
Clomipramine (Anafranil)	—	25, 50, and 75 mg	—	—

[a] Doxepin (Zonalon) is supplied as a 5% cream (30 g or 45 g).

If doxepin is used for sleep, doses lower than those in depression are used, for example, 3 to 6 mg/day.

At this time, it is not clear when tricyclic antidepressants should be used in children. It is recommended that desipramine should not be used in children younger than 12 years of age because of possible adverse cardiac effects, but the same logic also applies to other tricyclics. Tricyclic antidepressants have not been shown to be more effective than placebo in adolescents with depression, although this probably relates, in part, to diagnostic problems. The lethality of tricyclics in overdose should be kept in mind. Higher doses, on a milligram-per-kilogram basis, are likely to be required in younger patients than in adults to reach the same plasma level, but effective plasma level ranges have not been established in children.

Tricyclic and heterocyclic preparations are shown in Table 33.5a.3–3.

FURTHER READINGS

Agency for Health Care Policy and Research. *Clinical Practice Guideline: Depression in Primary Care: Treatment of Major Depression*. Vol 2. U.S. Government Printing Office; 1993.

Akamine Y, Yasui-Furukori N, Ieiri I, Uno T. Psychotropic drug-drug interactions involving P-glycoprotein. *CNS Drugs*. 2012;26:959–973.

Anderson IM. Selective serotonin reuptake inhibitors versus tricyclic antidepressants: a meta-analysis of efficacy and tolerability. *J Affect Disord*. 2000;58:19–36.

APA Task Force. Tricyclic antidepressants–blood level measurements and clinical outcome: an APA Task Force report. Task force on the use of laboratory tests in psychiatry. *Am J Psychiatry*. 1985;142:155–162.

Chan CH, Janicak PG, Davis JM, Altman E, Andriukaitis S, Hedeker D. Response of psychotic and nonpsychotic depressed patients to tricyclic antidepressants. *J Clin Psychiatry*. 1987;48:197–200.

Charney DS, Delgado PL, Price LH, Heninger GR. The receptor sensitivity hypothesis of antidepressant action: a review of antidepressant effects on serotonin function. In: Brown SL, van Praag HM, eds. *The Role of Serotonin in Psychiatric Disorders*. Brunner/Mazel; 1991:29–56.

Danish University Antidepressant Group. Paroxetine: a selective serotonin reuptake inhibitor showing better tolerance, but weaker antidepressant effect than clomipramine in a controlled multicenter study. *J Affect Disord*. 1990;18:289–299.

Davis JM, Glassman AH. Antidepressant drugs. In: Kaplan HI, Sadock BJ, eds. *Comprehensive Textbook of Psychiatry*. Williams & Wilkins; 1989:1627–1655.

Drevets WC, Zarate CA Jr, Furey ML. Antidepressant effects of the muscarinic cholinergic receptor antagonist scopolamine: a review. *Biol Psychiatry*. 2013;73:1156–1163.

Duman RS, Heninger GR, Nestler EJ. A molecular and cellular theory of depression. *Arch Gen Psychiatry*. 1997;54:597–606.

Elkin I, Shea MT, Watkins JT, et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. *Arch Gen Psychiatry*. 1989;46:971–982.

Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science*. 1999;286:487–491.

Faissner S, Mishra M, Kaushik DK, et al. Systematic screening of generic drugs for progressive multiple sclerosis identifies clomipramine as a promising therapeutic. *Nat Commun*. 2017;8(1):1990.

Farahani A, Correll CU. Are antipsychotics or antidepressants needed for psychotic depression? A systematic review and meta-analysis of trials comparing antidepressant or antipsychotic monotherapy with combination treatment. *J Clin Psychiatry*. 2012;73:486–496.

Frank E, Kupfer DJ, Perel JM, et al. Three-year outcomes for maintenance therapies in recurrent depression. *Arch Gen Psychiatry*. 1990;47:1093–1099.

Glassman AH, Bigger JT Jr. Cardiovascular effects of therapeutic doses of tricyclic antidepressants. A review. *Arch Gen Psychiatry*. 1981;38:815–820.

Glassman AH, Perel JM, Shostak M, Kantor SJ, Fleiss JL. Clinical implications of imipramine plasma levels for depressive illness. *Arch Gen Psychiatry*. 1977;34:197–204.

Kuhn R. The imipramine story. In: Ayd FJ, Blackwell B, eds. *Discoveries in Biological Psychiatry*. JB Lippincott Co; 1970:205–217.

Nelson JC, Jatlow PI, Bock J, Quinlan DM, Bowers MB. Major adverse reactions during desipramine treatment: relationship to plasma drug concentrations, concomitant antipsychotic treatment, and patient characteristics. *Arch Gen Psychiatry.* 1982;39(9):1055–1061.

Nelson JC, Spyker D. Morbidity and mortality of medications used in the treatment of depression. *Am J Psychiatry.* 2017;174(5):438–450.

O’Malley PG, Jackson JL, Santoro J, Tomkins G, Balden E, Kroenke K. Antidepressant therapy for unexplained symptoms and symptom syndromes. *J Fam Pract*. 1999;48:980–990.

Potter WZ, Calil HM, Sutfin TA, et al. Active metabolites of imipramine and desipramine in man. *Clin Pharmacol Ther*. 1982;31:393–401.

Preskorn SH, Fast GA. Therapeutic drug monitoring for antidepressants: efficacy, safety, and cost effectiveness. *J Clin Psychiatry*. 1991;52 Suppl:23–33.

Roose SP, Laghrissi-Thode F, Kennedy JS, et al. Comparison of paroxetine and nortriptyline in depressed patients with ischemic heart disease. *JAMA*. 1998;279:287–291.

Rosenstein DL, Nelson JC, Jacobs JC. Seizures associated with antidepressants: a review. *J Clin Psychiatry*. 1993;54:289–299.

Ryan ND. The pharmacologic treatment of child and adolescent depression. *Psychiatr Clin North Am*. 1992;15:29–40.

Spiker DG, Weiss JC, Dealy RS, et al. The pharmacologic treatment of delusional depression. *Am J Psychiatry*. 1985;142:430–436.

Tremblay P, Blier P. Catecholaminergic strategies for the treatment of major depression. *Curr Drug Targets*. 2006;7:149–158.

White N, Litovitz T, Clancy C. Suicidal antidepressant overdoses: a comparative analysis by antidepressant type. *J Med Toxicol*. 2008;4:238–250.

Yan L, Wang Q, Fu Q, Ye Q, Xiao H, Wan Q. Amitriptyline inhibits currents and decreases the mRNA expression of voltage-gated sodium channels in cultured rat cortical neurons. *Brain Res*. 2010;1336:1–9.

Yoshimura M, Furue H. Mechanisms for the anti-nociceptive actions of the descending noradrenergic and serotonergic systems in the spinal cord. *J Pharmacol Sci*. 2006;101:107–117.

33.5a.4 Dopamine-Noradrenaline Reuptake Inhibitors (Bupropion)

Charles DeBattista, D.M.H., M.D., and Alan F. Schatzberg, M.D.

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic domain	Norepinephrine, dopamine
Mechanism of action	Reuptake inhibitor (NET, DAT) and releaser (NE, DA)

Originally introduced in the 1980s, bupropion (Wellbutrin) is one of the more commonly prescribed antidepressants in the United States. As a result of unique pharmacology, bupropion is often used as an alternative to SSRIs and serotonin-norepinephrine reuptake inhibitors (SNRIs). In addition to its FDA approval in the treatment of depression, bupropion is also approved for smoking cessation and is the only antidepressant to receive FDA approval for the preventive treatment of seasonal affective disorder (SAD). Additionally, bupropion is used for numerous off-label indications, including the treatment of ADHD, sexual dysfunction, obesity, and fatigue related to a nonpsychiatric medical condition. Currently, all formulations of bupropion are generically available, and the major metabolite of bupropion, hydroxybupropion, is being investigated as a potential antidepressant.

PHARMACOLOGY

Pharmacodynamics

The pharmacodynamics of bupropion remains in debate. Although frequently described as a noradrenergic agonist/dopamine reuptake inhibitor, the human in vivo effects of bupropion on dopamine appear rather modest. Animal studies have indicated that bupropion is associated with an increase in nucleus accumbens dopamine concentrations, and it appears to inhibit the reuptake of dopamine more potently than tricyclics. However, human studies have not consistently shown a dopamine effect.

For example, some PET studies have shown that bupropion increases occupancy of the dopamine transporter as a dopamine reuptake inhibitor should, but others have shown no effect. Likewise, concentrations of homovanillic acid (HVA), a major metabolite of dopamine, do not appear to rise in clinical responders, although they may rise in nonresponders.

The effects of bupropion on noradrenergic activity appear more consistent than its dopamine effects. Bupropion reduces the 24-hour excretion of norepinephrine and its metabolites. Bupropion and its active metabolites, hydroxybupropion, erythro-hydrobupropion, and threo-hydrobupropion, inhibit the reuptake of norepinephrine in rat cortex, although much less potently than does imipramine. Downregulation of beta-adrenergic receptors and decreased firing of norepinephrine cells in the locus ceruleus have been reported in animal studies. In addition, chronic administration of bupropion appears to modestly inhibit tyrosine hydroxylase activity. In addition to its noradrenergic and dopaminergic effects, bupropion is also known to interact with nicotinic receptors. Bupropion is a noncompetitive inhibitor of the nicotine $\alpha_3\beta_4$ receptor, while one of its metabolites, *S,S*-hydroxybupropion, is a noncompetitive inhibitor of the nicotine $\alpha_4\beta_2$ receptor in the rat brain. Along with the effects of bupropion on the reward circuits of the locus ceruleus, the inhibition of nicotinic receptors by bupropion and its metabolites provide a rationale for its use in smoking cessation.

Bupropion became the focus of investigation by medicinal chemists in the 1960s when it appeared to be effective in standard screening models of antidepressants but lacked many of the properties of the TCAs and MAOIs that make both classes problematic to use. These include muscarinic side effects, tyramine pressor effects, hypotension, and so forth. Bupropion is not a significant inhibitor of serotonin reuptake, an agonist of 5-HT$_{1A}$ receptors, or an antagonist of 5-HT$_2$ receptors. The lack of significant serotonergic effects is thought to be the major factor in bupropion’s lack of sexual side effects and its general lack of efficacy in anxiety disorders.

Pharmacokinetics

Bupropion is a racemic mixture of enantiomers and is rapidly absorbed. This absorption is not substantially affected by concurrent food of anticholinergic administration. The peak blood level occurs within 2 hours of administration, and bupropion is tightly bound to protein (80% to 90%). Protein binding for hydroxybupropion is similar to bupropion, while the other metabolites are about half as tightly bound as bupropion.

Elimination of bupropion is biphasic, consistent with what has been referred to as a “two-compartment model,” that is, after distribution, serum levels are in equilibrium between the intravascular and extravascular compartments. The terminal mean half-life is about 21 hours, while the distribution phase half-life is about 4 hours. Hydroxybupropion may have a half-life of as long as 45 hours.

Bupropion undergoes extensive first-pass metabolism, with only 1% of the drug excreted unchanged in the urine. Most elimination is fecal. Bupropion is a substrate for the CYP450 2B6 isoenzyme, which is not commonly inhibited or induced by other drugs. Oncology agents such as thiotepa (Thioplex) for meningeal neoplasms are among the rare, potent inhibitors of CYP 2B6, while barbiturates, such as phenobarbital (Bellatal), may induce the enzyme. There have been conflicting in vitro and in vivo studies of bupropion’s CYP 2D6 inhibition. While bupropion is a weak inhibitor of CYP2D6, a number of its metabolites are more potent inhibitors. Thus, there is a potential for bupropion to inhibit the metabolism of some CYP 2D6 substrates, including many antidepressants, some antipsychotics, lipophilic beta-blockers, tamoxifen, and some opiates.

CLINICAL USES

Depressive Disorders

Bupropion has long been known to be an effective antidepressant. It appears to be comparable in efficacy to both TCAs and SSRIs. Most studies with bupropion have involved outpatients with mild to moderate depression. However, inpatient studies with more severely depressed patients have also indicated benefits. Small open-label studies in children and adolescents suggest a benefit of bupropion in pediatric depression, but no controlled studies are available. Geriatric depression studies indicate that bupropion in lower doses may be effective and well-tolerated. Case reports suggest an occasional worsening of psychotic features in patients with delusional depression.

Unlike most antidepressants, the efficacy of bupropion in treating MDD with concomitant anxiety or anxiety disorders is not well-established. Some studies in depressed patients have reported comparable anxiolytic effects of bupropion to SSRIs. However, most studies of bupropion in the treatment of panic disorder have failed to show benefits. Likewise, no controlled trials have been completed that demonstrate the clear benefit of bupropion monotherapy for GAD, Social anxiety, or other anxiety disorders. On the other hand, anxiety may be a consequence of bupropion treatment on occasion.

Among the most common applications of bupropion is in the augmentation of response to other antidepressants. Thus, it is frequently combined with SSRIs, SNRIs, or mirtazapine in patients who have shown inadequate responses to monotherapy. It makes intuitive sense that the combination of a noradrenergic and perhaps dopaminergic molecule such as bupropion might.

Among the largest trials to evaluate the efficacy of bupropion augmentation to date is the Sequenced Treatment Alternatives in Resistant Depression (STAR*D) trial. In that study, patients who had failed citalopram (Celexa) treatment could opt to be randomized to either buspirone (BuSpar) or bupropion augmentation. Both treatments were equally effective, with about 30% of patients achieving remission after failing citalopram monotherapy. Bupropion was somewhat more effective than buspirone in an overall reduction of depressive symptoms. However, the lack of a placebo arm makes it difficult to determine whether either augmenting strategy was more effective than would be longer treatment with citalopram alone.

The other major study to evaluate the efficacy of bupropion augmentation was the VA Augmentation and Switching Treatments for Improving Depression Outcomes (VAST-D) study. In the 1,522-subject trial, patients with MDD who had not responded to an antidepressant were randomized to bupropion alone, bupropion augmentation, or augmentation with aripiprazole. The response rate appeared to be higher in the aripiprazole group than in either the switch or bupropion augmentation groups. Likewise, the remission rate in the aripiprazole group was higher than in the group switched to bupropion but not statistically different than bupropion augmentation. Thus, bupropion augmentation seemed slightly less effective than augmentation with aripiprazole overall.

Given the low remission rates of monotherapy of major depression with antidepressants, attempts have also been made to initiate combined treatment with bupropion and an SSRI as a way of enhancing the speed of onset and remission rates. A randomized controlled study of 245 MDD patients compared the combination treatment with escitalopram and bupropion to monotherapy with either escitalopram or bupropion (The Co-Med trial). No advantage was found for the combination treatment with bupropion to either monotherapy in the speed of response or overall remission rates.

Bupropion is commonly used adjunctively with antidepressants to treat SSRI-induced sexual side effects. Again, the rationale for using a dopaminergic agent, even a modest one, to impact sexual side effects appears sound. Stimulants and other dopamine-enhancing agents enhance libido. Although case reports and open studies supported the use of bupropion, several placebo-controlled trials have not demonstrated substantial benefit in depressed patients. Two studies at 150 mg per day of bupropion sustained-release (SR) added to an SSRI did not report significant benefit over placebo in treating sexual side effects; one study at 300 mg per day did demonstrate mild efficacy. The benefit of adding bupropion to a serotonergic antidepressant may be more evident in women than men. A randomized controlled study of 218 women with antidepressant-induced sexual side effects found that 300 mg per day of bupropion added to an existing antidepressant was significantly more effective than placebo in increasing libido and, to a lesser extent, arousal. Similarly, male patients with methadone-induced sexual dysfunction appeared to have a significant improvement in sexual function with the addition of bupropion in a placebo-controlled, randomized trial. Switching from a serotonergic antidepressant to bupropion also appears to be an effective strategy to reduce sexual side effects in both men and women. In 2022, the combination of bupropion with dextromethorphan was approved for the treatment of major depression. The combination of a relatively low dose of bupropion SR 105 mg and 45 mg of dextromethorphan 45 m, given twice daily, appears to increase the speed of on onset of the antidepressant response and may improve the overall efficacy relative to similar doses of bupropion alone.

Bupropion was approved for the preventive treatment of SAD in 2006. In a series of three trials, 1,042 patients who met the criteria for SAD were treated in the autumn while they were still well. Patients were randomized to receive either 150 to 300 mg per day of bupropion extended-release (XL) or placebo and then followed through the spring. On average, patients had 13 previous seasonal depressive episodes. Patients on placebo were significantly more likely to experience fall/winter depressive episodes than were bupropion-treated patients. Placebo-treated patients were also more likely to have a recurrence of depression sooner than were those on bupropion. To date, no other antidepressants have received an indication in the treatment of SAD.

Bupropion has been favored as an adjunctive treatment for bipolar depression secondary to the perception that it might be less likely to induce switches into mania or more rapid cycling. This perception, however, is based on very little data. Few large prospective studies have evaluated the utility and safety of bupropion in bipolar depression. The largest study to date to evaluate bupropion in bipolar disorder was a part of the NIMH-funded Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) study. In that study, patients with bipolar depression on a mood stabilizer were randomized to the addition of bupropion or paroxetine (Paxil) versus the addition of a placebo. Patients were treated for 26 weeks and evaluated to determine if they had at least 8 weeks of recovery. Neither bupropion nor paroxetine was significantly better than placebo in achieving a durable recovery from the bipolar depressive episodes. In fact, on some measures, placebo-treated patients fared somewhat better. The rate of switches into mania also did not differ between bupropion, paroxetine, and placebo in patients on therapeutic doses of standard mood stabilizers. A study supported by the Stanley Foundation likewise showed relatively low-switch rates of bupropion in an acute bipolar trial. The response rate in this study was about 50%, with about a 10% switch rate. Thus, the utility of bupropion in bipolar depression is still uncertain, but switch rates of bupropion, when combined with standard mood stabilizers, appear low.

Smoking Cessation

Another FDA indication for bupropion (SR) is for the promotion of smoking cessation. Early conceptualization about the mechanism for its efficacy in nicotine withdrawal and cravings included bupropion’s effects on dopamine reuptake and an influence on reward mechanisms in addictive states. Indeed, there is evidence that dopamine polymorphisms have some predictive value in determining the efficacy of bupropion in smoking cessation. However, more current thinking is that bupropion may have significant, although indirect, effects on nicotine receptors. For example, bupropion may mimic nicotinic effects on dopamine and norepinephrine, which may attenuate withdrawal symptoms. In addition, there is some evidence of direct antagonism of nicotine receptors, which might prevent relapse by diminishing the reinforcing properties of nicotine. In addition, bupropion appears to act as a noncompetitive α3β4 nicotinic antagonist. Despite additional understanding, the mechanism of bupropion for smoking cessation remains unclear, but bupropion appears unlike nicotine-specific therapies such as nicotine substitutes (nicotine patch/gum) or the partial nicotine receptor agonist varenicline (Chantix). There is evidence that the major metabolite of bupropion, hydroxybupropion, may mediate the smoking cessation effects of the drug. Since there are polymorphisms for the CYP2B6 isoenzyme that is responsible for the metabolism of bupropion to hydroxybupropion, slow metabolizers of CYP2B6 may require higher doses to achieve a serum level of 0.7 mcg/mL as these doses may predict better response to bupropion therapy for smoking cessation.

Overall, bupropion appears to be about twice as effective as placebo in reducing acute nicotine withdrawal symptoms, nicotine cravings, and maintaining abstinence. For example, bupropion has been associated with less mood instability and weight gain after acute withdrawal from tobacco. In one pivotal trial, only 27% of patients experienced nicotine cravings 1 month after smoking cessation on bupropion 300 mg per day versus 56% of those on placebo. About 30% of patients on bupropion are able to quit after 1 to 3 months versus about 15% of placebo-treated patients. After 1 year of treatment, patients who remain on bupropion are about 1.5 times as likely to remain abstinent as placebo-treated patients. Bupropion also appears effective in specific populations with nicotine addictions, including those with chronic obstructive pulmonary disease (COPD), cardiovascular illness, concurrent alcoholism, schizophrenia, and comorbid depression.

Direct comparisons with nicotine-specific therapies are few. Bupropion appears to be at least as effective as nicotine replacement with a nicotine patch, but the combination of bupropion and the nicotine patch may not offer benefit over either therapy alone. For example, the addition of bupropion to a nicotine patch did not offer benefit over the addition of placebo in a 26-week study of smoking cessation in adolescents. At least two head-to-head comparisons with varenicline tartrate 1 mg twice daily indicated that varenicline may be somewhat more effective in the acute and long-term management of smoking cessation. The combination of varenicline and bupropion may be more effective than either treatment alone.

The usual starting dose of bupropion SR for smoking cession is 150 mg per day, with a target dose of 300 mg per day within 7 days. An acute course of bupropion tends to last 8 to 12 weeks, with tobacco stopped within the first 2 to 3 weeks of treatment. Responders may continue bupropion treatment chronically to prevent relapse.

Obesity and Eating Disorders

The FDA approved the combination of bupropion and naltrexone for the treatment of obesity in 2014. Four large registration trials involving over 4,500 patients were completed in overweight BMI >27 to obese BMI >30 patients (mean BMI = 36). The general design of these studies was 3 weeks of dose titration followed by approximately 1 year of treatment. Subjects were randomized to the combination of naltrexone (15 to 60 mg) and bupropion (300 to 400 mg) along with dietary modifications (e.g., reducing daily calories by 500 kcal/day), behavioral counseling, and an increase in exercise or physical activity versus the control group included with these interventions with a placebo instead of the active drug. In all four trials, subjects randomized to the active group were more likely to lose a clinically meaningful amount of weight, defined as >5% or >10% of total body weight. In addition, subjects on the combination of bupropion (mean dose 360 mg per day) and naltrexone (mean dose 32 mg per day) lost about 5% more of total body weight versus subjects in the control group.

A large percentage of patients in both the active and control groups of the naltrexone/bupropion obesity studies dropped out before completing the 1-year endpoint. In fact, approximately 45% of both groups dropped out early, with subjects in the active group twice as likely to drop out as subjects in the control group (24% versus 12%). The most common side effects leading to discontinuation of the combination treatment were nausea, headache, and vomiting. Other common side effects of the combination of naltrexone/bupropion are constipation,

Bupropion alone also has mild anorexiant effects that have been evaluated in nondepressed obese patients. Approximately 25% of patients in acute depression trials lost 5 lb or more while taking bupropion. Several controlled studies have evaluated 300 to 400 mg per day of bupropion in the treatment of obesity. These studies have indicated that bupropion-treated patients tend to lose 1.1 to 4.5 kg at 6 to 12 months, and this is significantly greater than placebo-treated patients. Bupropion has also been investigated in the treatment of binge eating disorder. The largest randomized, placebo-controlled trial to date (*n* = 61) did not find the bupropion to reduce binging or to improve weight loss in women with the disorder.

Other Clinical Uses

In addition to the FDA-approved indications in the treatment of major depression and for smoking cessation, bupropion has been clinically tested for a number of other conditions. For example, as with other noradrenergic antidepressants such as nortriptyline (Aventyl), bupropion may be useful in the treatment of adult ADHD. In pediatric populations, the only large controlled trial did not show significant benefit of bupropion on the primary outcome measures. However, in this trial, teachers rated the bupropion-treated children as less aggressive and hyperactive relative to placebo-treated children. Clinicians and parents did not observe a significant difference. There are a few direct comparisons of bupropion with stimulants; however, these studies revealed that bupropion is either about as effective or somewhat less effective. A small randomized trial (*n* = 44) comparison of bupropion and methylphenidate found that both drugs were equally effective and well-tolerated. Studies of bupropion in adults with ADHD have been more consistently positive. For example, in a study of 162 adults with ADHD, 53% of patients responded to bupropion XL versus 31% of placebo-treated patients. Benefits over placebo were noted as early as week 2, and bupropion, with an average dose of 450 mg per day, was well-tolerated.

Finally, bupropion has been suggested to be of benefit in a number of small studies and case series for a broad spectrum of conditions. These have included sexual dysfunction in nondepressed patients, fatigue in cancer patients, the treatment of amphetamine addiction, lowering tumor necrosis factor-α, and improving concentration in patients without ADHD.

SIDE EFFECTS

The pharmacodynamic effects of bupropion result in a side-effect profile different from most antidepressants. Many of the adverse events reported with bupropion are associated with its noradrenergic and, to a lesser extent, dopaminergic effects. For example, CNS activation is associated with bupropion treatment. Insomnia is one of the common side effects of bupropion, affecting at least 11% of patients in clinical trials. Anxiety and agitation may occur as treatment-emergent effects. At least 6% of patients in clinical trials also experienced a mild-to-moderate tremor. Treatment-emergent psychotic symptoms have rarely been reported in depressed patients.

The risk of seizure with the immediate-release (IR) formulation of bupropion became evident in studies of eating disorder patients initially and led to the withdrawal of the drug in 1986. Patients in the eating disorder trials were treated with up to 600 mg per day of bupropion. Analysis of data sets revealed an enhanced risk of seizure in patients with a history of eating disorders, head trauma, or previous seizure disorder. Also, doses above 450 mg per day or 200 mg at one time (for the IR formulation) enhanced the risk of seizure. For XL formulations at doses of 450 mg per day or less, the seizure risk is estimated at 0.1% to 0.2%, which is comparable to the rates reported with SSRIs and may be less than the rates reported with some TCAs and maprotiline (Novo-Maprotiline).

Other common side effects reported in trials included nausea, dry mouth, excessive sweating, tinnitus, and rash. The most common reasons patients discontinued bupropion in acute clinical trials of depression were nausea and rash.

Bupropion has a favorable cardiovascular profile with few cardiac effects. Bupropion is not known to have clinically significant effects on blood pressure, heart rate, or cardiac conduction. Thus, bupropion is commonly used in patients with known cardiac pathology. However, there have been rare reports of significant hypertension associated with bupropion in patients with and without a preexisting history of high blood pressure.

Side effects common to many antidepressants (including sexual side effects and weight gain) appear uncommon with bupropion. As noted earlier, bupropion has been associated with weight loss, and there are many reports of improved sexual functioning, including increased libido and arousal, with the drug.

Other side effects such as hepatoxicity, hematopoietic changes, and severe headaches can rarely occur in bupropion-treated patients. Thus, the side-effect profile makes bupropion an important alternative for many patients who cannot tolerate other antidepressants.

DRUG INTERACTIONS

The limited cytochrome P450 interactions of bupropion contribute to its few drug interactions. Since bupropion and hydroxybupropion are metabolized via the cytochrome P2B6, there is a theoretical risk of an interaction between oncology agents such as IV thiotepa and the muscle relaxant orphenadrine (Norflex). Hydroxybupropion is a weak-to-moderate inhibitor of cytochrome P450 2D6 and has been associated with increased serum levels of some 2D6-dependent drugs such as venlafaxine (Effexor) and desipramine (Norpramin). Rare seizures that have occurred in the combination of fluoxetine (Prozac) and bupropion have sometimes been attributed to inhibition of the metabolism of hydroxybupropion by 2D6 inhibitors, but seizures can occur independent of these potential interactions, so it is unclear whether this mechanism was at play.

Since bupropion may lower the seizure threshold, particularly when the IR form is used in higher doses or in combination with other drugs that lower the seizure threshold, such as some tricyclics, amoxapine (Asendin), clozapine (Clozaril), and theophylline (Elixophyllin), it is sometimes associated with seizures.

As with other antidepressants, the concurrent administration with MAOIs is contraindicated. However, there have been few reports in the literature of serious interactions with MAOIs. Oral selegiline (Carbex) at doses of 10 mg per day or less and transdermal selegiline (Emsam) have sometimes been combined with bupropion without incident.

OVERDOSE

Bupropion has rarely been associated with fatalities in monotherapy overdoses. Overdoses of bupropion as low as 10 g have sometimes been associated with deaths. Causes of death in these overdoses have included complications of status epilepticus and arrhythmias. Although therapeutic doses of bupropion are rarely associated with cardiovascular effects, overdoses have sometimes been associated with tachycardia, intraventricular conduction delays, and QRS complex widening. Treatment for overdoses includes lavage, seizure treatment or prophylaxis, cardiac monitoring, and vital sign support. Most patients who overdose on bupropion recover.

DOSING AND FORMULATIONS

Bupropion hydrochloride is available in IR, SR, and XL forms. As of 2006, all of these formulations had become available as generics in the United States. The usual starting dose in all formulations is 150 mg per day taken in the morning, but the IR form was often started at 100 mg twice a day. The average therapeutic dose of bupropion for the treatment of depression is 300 to 450 mg per day. Doses above 200 mg at once are associated with greater seizure risk in the IR form. Thus, at maximum doses, compliance with bupropion IR prescriptions was somewhat difficult when dosed at 150 mg three times a day. In contrast, the typical therapeutic dose of the SR form was 150 or 200 mg twice a day. The XL form could be taken all in one morning dose, which was said to have more consistent blood levels and perhaps less sleep disruption. A cellulose coating in XL formulation allowed for slow absorption.

In smoking cessation, bupropion SR is typically started at 150 mg per day and then increased to the target dose of 300 mg per day. In the management of obesity (Contrave) comes in a combination tablet of 90 mg of bupropion with 8 mg of naltrexone. The typical therapeutic dose is 1 to 2 tablets, twice daily.

FURTHER READINGS

Caixàs A, Albert L, Capel I, Rigla M. Naltrexone sustained-release/bupropion sustained-release for the management of obesity: review of the data to date. *Drug Des Devel Ther*. 2014;8:1419–1427.

Cinciripini PM, Robinson JD, Karam-Hage M, et al. Effects of varenicline and bupropion sustained-release use plus intensive smoking cessation counseling on prolonged abstinence from smoking and on depression, negative affect, and other symptoms of nicotine withdrawal. *JAMA Psychiatry*. 2013;70(5):522–533.

Clayton AH, Croft HA, Horrigan JP, et al. Bupropion extended release compared with escitalopram: effects on sexual functioning and antidepressant efficacy in 2 randomized, double-blind, placebo-controlled studies. *J Clin Psychiatry*. 2006;67(5):736–746.

Clayton AH, Montejo AL. Major depressive disorder, antidepressants, and sexual dysfunction. *J Clin Psychiatry*. 2006;67(Suppl 6):33–37.

Davidson JR, Connor KM. Bupropion sustained release: a therapeutic overview. *J Clin Psychiatry*. 1998;59(Suppl 4):25–31.

DeBattista C. Augmentation and combination strategies for depression. *J Psychopharmacol*. 2006;20(3 Suppl):11–18.

DeBattista C, Solvason B, Poirier J, Kendrick E, Loraas E. A placebo-controlled, randomized, double-blind study of adjunctive bupropion sustained release in the treatment of SSRI-induced sexual dysfunction. *J Clin Psychiatry*. 2005;66(7):844–848.

DeBattista C, Solvason HB, Poirier J, Kendrick E, Schatzberg AF. A prospective trial of bupropion SR augmentation of partial and non-responders to serotonergic antidepressants. *J Clin Psychopharmacol*. 2003;23(1):27–30.

Ebbert JO, Hatsukami DK, Croghan IT, et al. Combination varenicline and bupropion SR for tobacco-dependence treatment in cigarette smokers: a randomized trial. *JAMA*. 2014;311(2):155–163.

Foley KF, DeSanty KP, Kast RE. Bupropion: pharmacology and therapeutic applications. *Expert Rev Neurother*. 2006;6(9):1249–1265.

Friedman ES, Davis LL, Zisook S, et al. Baseline depression severity as a predictor of single and combination antidepressant treatment outcome: results from the CO-MED trial. *Eur Neuropsychopharmacol*. 2012;22(3):183–199.

Jefferson JW, Pradko JF, Muir KT. Bupropion for major depressive disorder: pharmacokinetic and formulation considerations. *Clin Ther*. 2005;27(11):1685–1695.

Jha MK, Wakhlu S, Dronamraju N, Minhajuddin A, Greer TL, Trivedi MH. Validating pre-treatment body mass index as moderator of antidepressant treatment outcomes: findings from CO-MED trial. *J Affect Disord*. 2018;234:34–37.

Kharasch ED, Crafford A. Common polymorphisms of CYP2B6 influence stereoselective bupropion disposition. *Clin Pharmacol Ther*. 2019;105(1):142–152.

Killen JD, Robinson TN, Ammerman S, et al. Randomized clinical trial of the efficacy of bupropion combined with nicotine patch in the treatment of adolescent smokers. *J Consult Clin Psychol*. 2004;72(4):729–735.

Kuperman S, Perry PJ, Gaffney GR, et al. Bupropion SR vs. methylphenidate vs. placebo for attention deficit hyperactivity disorder in adults. *Ann Clin Psychiatry*. 2001;13(3):129–134.

Lee MS, Yang JW, Ko YH, et al. Effects of methylphenidate and bupropion on DHEA-S and cortisol plasma levels in attention-deficit hyperactivity disorder. *Child Psychiatry Hum Dev*. 2008;39(2):201–209.

Levin FR, Evans SM, Brooks DJ, Kalbag AS, Garawi F, Nunes EV. Treatment of methadone-maintained patients with adult ADHD: double-blind comparison of methylphenidate, bupropion and placebo. *Drug Alcohol Depend*. 2006;81(2):137–148.

Modell JG, Rosenthal NE, Harriett AE, et al. Seasonal affective disorder and its prevention by anticipatory treatment with bupropion XL. *Biol Psychiatry*. 2005;58(8):658–667.

Mohamed S, Johnson GR, Chen P, et al. Effect of antidepressant switching vs augmentation on remission among patients with major depressive disorder unresponsive to antidepressant treatment: the VAST-D randomized clinical trial. *JAMA*. 2017;318(2):132–145.

Morris DW, Budhwar N, Husain M, et al. Depression treatment in patients with general medical conditions: results from the CO-MED trial. *Ann Fam Med*. 2012;10(1):23–33.

Nissen SE, Wolski KE, Prcela L, et al. Effect of naltrexone-bupropion on major adverse cardiovascular events in overweight and obese patients with cardiovascular risk factors: a randomized clinical trial. *JAMA*. 2016;315(10):990–1004.

Onakpoya IJ, Lee JJ, Mahtani KR, Aronson JK, Heneghan CJ. Naltrexone-bupropion (Mysimba) in management of obesity: a systematic review and meta-analysis of unpublished clinical study reports. *Br J Clin Pharmacol*. 2020;86(4):646–667.

Rush AJ, Trivedi MH, Stewart JW, et al. Combining medications to enhance depression outcomes (CO-MED): acute and long-term outcomes of a single-blind randomized study. *Am J Psychiatry*. 2011;168(7):689–701.

Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry*. 2006;163(11):1905–1917.

Rush AJ, Trivedi MH, Wisniewski SR, et al. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med*. 2006;354(12):1231–1242.

Sachs GS, Nierenberg AA, Calabrese JR, et al. Effectiveness of adjunctive antidepressant treatment for bipolar depression. *N Engl J Med*. 2007;356(17):1711–1722.

Solhkhah R, Wilens TE, Daly J, Prince JB, Van Patten SL, Biederman J. Bupropion SR for the treatment of substance-abusing outpatient adolescents with attention-deficit/hyperactivity disorder and mood disorders. *J Child Adolesc Psychopharmacol*. 2005;15(5):777–786.

Taylor MJ. Strategies for managing antidepressant-induced sexual dysfunction: a review. *Curr Psychiatry Rep*. 2006;8(6):431–436.

Trivedi MH, Fava M, Wisniewski SR, et al. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*. 2006;354(12):1243–1252.

Upadhyaya HP, Brady KT, Wang W. Bupropion SR in adolescents with comorbid ADHD and nicotine dependence: a pilot study. *J Am Acad Child Adolesc Psychiatry*. 2004;43(2):199–205.

White MA, Grilo CM. Bupropion for overweight women with binge-eating disorder: a randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry*. 2013;74(4):400–406.

Wilens TE, Haight BR, Horrigan JP, et al. Bupropion XL in adults with attention-deficit/hyperactivity disorder: a randomized, placebo-controlled study. *Biol Psychiatry*. 2005;57(7):793–801.

Wilens TE, Spencer TJ, Biederman J, et al. A controlled clinical trial of bupropion for attention deficit hyperactivity disorder in adults. *Am J Psychiatry*. 2001;158(2):282–288.

Yee A, Loh HS, Ong TA, Ng CG, Sulaiman AH. Randomized, double-blind, parallel-group, placebo-controlled trial of bupropion as treatment for methadone-emergent sexual dysfunction in men. *Am J Mens Health*. 2018;12(5):1705–1718.

33.5b 5HT$_{1A}$ Receptor Partial Agonists (Buspirone)

PIERRE BLIER, M.D., PH.D., AND CARY KOGAN, PH.D., M.S.C.P.

Neuroscience-Based Nomenclature (NbN)	
Buspirone	
Pharmacologic domain	Serotonin
Mechanism of action	Receptor partial agonist (5-HT_{1A})

Several serotonin (5-hydroxy-tryptamine; 5-HT)$_{1A}$ agonists have been developed and tested in patients diagnosed with various medical conditions and psychiatric disorders. The most extensively studied is buspirone, an azaspirodecanedione derivative approved for use in many countries. Its second-generation compound, gepirone, has also been extensively studied as an extended-release formulation and deemed efficacious by the U.S. Food and Drug Administration (FDA) in the treatment of major depressive disorder (MDD). Two other related azapirones are ipsapirone and tandospirone, with only the latter being on the market (in Japan and China). Since its synthesis in 1968 in the laboratories of Mead Johnson, the azapirones have been studied for their potential antidepressant and anxiolytic effects. Buspirone was, however, initially developed as a drug to treat psychosis because it was thought to act preferentially on dopamine receptors. Indeed, studies conducted in the 1970s revealed that a single dose of buspirone decreased aggressive behavior in rhesus monkeys. In the early 1980s, buspirone showed marked anxiolytic effects in behavioral models of anxiety in rodents. It was then subsequently realized that its therapeutic effects were more likely mediated by the 5-HT system, more specifically the 5-HT_{1A} receptor subtype identified in the early 1980s. Eventually, placebo-controlled studies in patients with anxiety disorders showed improvement in symptoms, with buspirone leading to FDA approval for generalized anxiety disorder (GAD) in 1986. FDA approval for the treatment of GAD with buspirone reflected the new DSM-III classification whereby anxiety neurosis had been replaced by GAD. At that time, buspirone was the only non-benzodiazepines medication approved by the FDA for the treatment of this disorder. Other clinical uses have been investigated, such as treatment for other disorders (e.g., social anxiety disorder, obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), panic disorder), depressive disorders, eating disorders, and attention-deficit/hyperactivity disorder (ADHD). Buspirone has only been indicated for the treatment of GAD and anxiety associated with depressive symptoms. It is, however, listed in treatment guidelines for adjunctive treatment for depression (i.e., CANMAT) and some anxiety disorders.

PHARMACOLOGIC ACTIONS

Buspirone hydrochloride (BuSpar) has been shown to act on two types of receptors: 5-HT and, to a much lesser extent, dopamine. It has a high affinity for the 5-HT_{1A} receptor, acting as an agonist or partial agonist, and has a moderate affinity for the D_2-dopamine receptor, acting as both an agonist and antagonist. It was initially believed that the azapirones were a better alternative to the $GABA_A$ positive allosteric modulators (PAMs; formerly referred to as the benzodiazepines) since they are not sedating and do not possess the same potential for dependence. However, clinical experience with these drugs, especially buspirone, revealed that they produce significant undesirable side effects, more delayed onset of therapeutic action, and are not more effective than the $GABA_A$ PAMs.

The chemical structures for the azapirones (buspirone, gepirone, tandospirone, and ipsapirone) are presented in Figure 33.5b–1.

PHARMACOKINETICS

Absorption

Buspirone undergoes extensive first-pass metabolism when ingested orally, with an absolute bioavailability of 4%. A 20-mg dose of buspirone produces peak plasma concentrations of 1 to 6 ng/mL (C_{max}) from 0.7 to 1.5 hours following administration (T_{max}). Buspirone exhibits linear pharmacokinetics only over the 10- to 40-mg oral dose range in humans, which is generally within the usual therapeutic range. It has a plasma elimination half-life of 2.5 hours and thus, should be administered twice to thrice daily. The oral bioavailability of buspirone is increased when taken with food when compared to fasting conditions such that the area under the plasma concentration-time curve (AUC) and the maximal peak concentration (C_{max}) are increased by 84% and 116%, respectively. Patients with cirrhosis can have an AUC increased up to 13-fold, and therefore other anxiolytic alternatives that do not undergo hepatic metabolism should be considered (i.e., oxazepam).

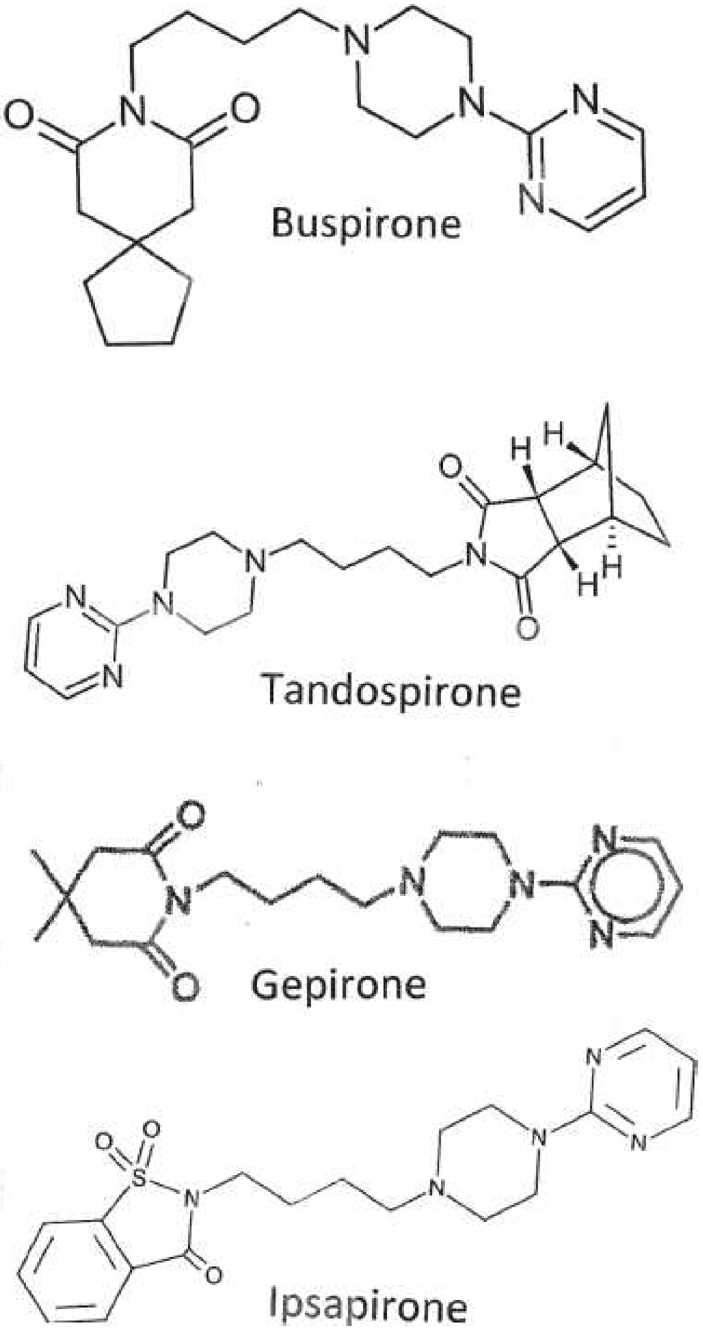
Buspirone
Tandospirone
Gepirone
Ipsapirone

FIGURE 33.5b–1. Chemical structures for the azapirones (buspirone, gepirone, tandospirone, and ipsapirone).

Distribution

Following first-pass liver metabolism, the remaining drug enters the systemic circulation. The volume of distribution is 5.3 L/kg as calculated by intravenous (IV) administration. Buspirone and its major active metabolites, 6-OH-buspirone and 1-pyrimidinylpiperazine (1-PP), both readily cross the blood-brain barrier, as demonstrated using IV administration in electrophysiologic recording experiments in the rat brainstem and forebrain. In vitro studies indicate that 86% to 95% of the drug is bound to plasma protein. In vivo studies in rats demonstrate that both buspirone and its metabolites are found in maternal milk and cross the placenta. It is not known to what extent both these events occur in humans.

METABOLISM AND ELIMINATION

Buspirone is metabolized through oxidation by cytochrome P450 3A4 (CYP3A4). CYP3A4, the most abundant form of cytochrome P450 in the human adult liver, is polymorphic and thus shows wide interindividual variation in its activity. Main metabolic pathways include hydroxylation on the 6′-position on the spiro ring system (6-OH buspirone) and the 5′-position on the pyrimidine ring of the drug complex (5-OH buspirone), as well as an *N*-dealkylation of a butyl side group, forming 1-PP. Approximately 26% of buspirone was found to be converted to 1-PP, and this metabolite had a 49-fold greater AUC than that of buspirone but conversely had 25% of the potency of the parent compound.

Another major active metabolite is 6-OH-hydroxybuspirone. This metabolite was found to have 12 times greater mean AUC in the brain compared to the parent compound. Similar to the parent compound, 6-OH-hydroxybuspirone had four times greater affinity for presynaptic 5-HT_{1A} receptors in the dorsal raphe nucleus compared to postsynaptic 5-HT_{1A} located in the hippocampus. As expected, given that CYP3A4 is polymorphic, the mean half-life ($t_{1/2}$) of buspirone varied greatly between healthy individuals, with a minimum of 2 to 3 hours and a maximum of 11 hours after a single dose of 10 to 40 mg. The mean half-life of 1-PP was 6 hours following a 20-mg dose. In subjects with renal impairment, defined as lower than normal range creatinine clearance, the half-life of buspirone increased on average by only 2 hours and would therefore not necessitate dose adjustment.

PHARMACODYNAMICS

Mechanism of Action

Buspirone has a unique mechanism of action from other medications used to treat anxiety, especially because it does not bind to the $GABA_A$ receptor complex. Although the exact mechanism of action of the azapirones has not been fully elucidated, it mainly relies on the 5-HT system with potential additional interactions with the dopamine and norepinephrine (NE) systems.

Serotonin System

Buspirone and the other azapirones act as full agonists at 5-HT_{1A} autoreceptors on 5-HT neurons and as partial agonists at postsynaptic 5-HT_{1A} receptors. The short-term activation of somatodendritic 5-HT_{1A} autoreceptors in the midbrain raphe nuclei has been demonstrated to decrease the firing rate of 5-HT neurons and the synthesis of 5-HT in rats. Under these conditions, such a decrease in 5-HT release does not result in decreased 5-HT_{1A} transmission in postsynaptic areas because the drug is also present in such regions, thereby counterbalancing the attenuated effect of 5-HT release. After sustained administration, however, firing rates of 5-HT neurons and 5-HT synthesis gradually return to normal as a result of desensitization of the 5-HT_{1A} autoreceptors. Subsequently, there is a net increase in 5-HT_{1A} transmission in the forebrain resulting from (1) normalized 5-HT release and (2) enhanced activation of postsynaptic 5-HT_{1A} receptors by the exogenous azapirone. The postsynaptic 5-HT_{1A} receptors in the forebrain do not desensitize with sustained administration of azapirones, unlike their presynaptic counterparts in the brainstem. Furthermore, the levels of the azapirones under such conditions, whereby the drug is administered systemically, do not attenuate the full agonistic effect of 5-HT. Therefore, such a net but delayed increase in 5-HT_{1A} neurotransmission in the forebrain is believed to underlie the therapeutic effect of buspirone in depression and anxiety beginning to manifest itself after at least 2 weeks.

Dopaminergic System

The dopaminergic system is an entity that can be considered to contribute to the clinical actions of buspirone. Buspirone has been categorized as both an agonist and antagonist to the D_2 subtype of the dopamine receptor. Buspirone acts as an antagonist to presynaptic dopamine autoreceptors, but some evidence has also pointed to the involvement of postsynaptic dopamine receptors. The inhibition of the autoreceptors on dopamine cell bodies and terminals increased dopamine synthesis and metabolism in the striatum and nucleus accumbens. However, given the much higher affinity of buspirone for 5-HT$_{1A}$ than for dopamine receptors, it is less likely that a dopamine action contributes significantly to the therapeutic action of buspirone.

Norepinephrine System

As mentioned above, buspirone and the azapirones generate 1-PP as a metabolite. This compound has been shown in in vivo electrophysiologic recordings in rats to be an effective antagonist of presynaptic α_2-adrenergic receptors on NE neurons. This could explain an increase in plasma levels of noradrenaline after acute administration of buspirone in humans, which are under an inhibitory influence of these receptors. However, 1-PP is also an effective postsynaptic α_2-adrenergic antagonist, which could counterbalance the impact of increased NE on these receptors. The α_1- and β-adrenergic receptors could, nevertheless, theoretically be under increased activation by 1-PP. However, when tandospirone was administered in a long-term fashion in rats as a precursor of 1-PP, there was no modification of the firing of NE neurons in the locus coeruleus, and the effects of the stimulation of NE fibers on the firing of hippocampus neurons mediated by α_1- and β-adrenergic receptors were unaltered. Therefore, it is unlikely that the NE system is involved in the therapeutic benefits of buspirone.

GABA System

As previously mentioned, buspirone has no affinity for the allosteric binding sites on the GABA$_A$ receptor, nor does it affect the binding of the endogenous ligand GABA to its receptor site. Despite this fact, there has been some suggestion that buspirone can indirectly affect the GABA system since buspirone was found to increase the level of binding of flunitrazepam in an in vivo model. Further studies are needed to explore the functional implication of this effect.

Acetylcholine System

Binding experiments have shown that buspirone has no binding affinity for muscarinic cholinergic receptors. However, some evidence indicates the possibility of an indirect relationship with this system. Acute administration of buspirone caused an efflux of acetylcholine in the rat hippocampus, a modest increase in motor activity, and was observed to have a trophic effect on septal cholinergic-containing neurons. Again, in order to link the delayed onset of action of buspirone to these observations in animal experiments with single-dose injections, repeated or sustained administration paradigms have to be carried out, and doses that mimic the plasma levels achieved in patients should be used.

Blood Concentration and Relation to Action

There is no definitive evidence linking plasma concentrations to the therapeutic effects of buspirone, although a study using tandospirone showed a robust anxiolytic effect with higher doses in patients with GAD.

CLINICAL TRIALS

Generalized Anxiety Disorder

Several double-blind, randomized, placebo-controlled trials have established the anxiolytic efficacy of buspirone. Most monotherapy studies have focused on the treatment of GAD, with the majority of trials finding buspirone to be significantly associated with a greater reduction in anxiety symptoms compared to placebo. Standard outcome measures of severity of anxiety symptoms (Hamilton Anxiety [HAM-A] scale) and global improvement (Clinical Global Impressions [CGI] scale) were generally employed. A clinical trial also examined the use of buspirone or placebo in conjunction with anxiety management therapy or nondirective therapy. All four treatment groups exhibited a significant reduction in anxiety symptoms, with no significant difference between any of the groups. Although eligibility for entry into each of these trials required a diagnosis of GAD, it is worth noting that the definition of GAD has changed over time, such that the required duration of anxiety symptoms was increased from 1 month (third edition of *Diagnostic and Statistical Manual of Mental Disorders* [DSM-III]) to 6 months (DSM-III-R to DSM-5). Initial buspirone clinical trials used DSM-III criteria for GAD, while more recent clinical trials utilized DSM-III-R or DSM-IV criteria. It is possible that those initial studies may have included many subjects with anxiety symptoms that would not have been sufficiently chronic to be currently classified as GAD.

Early studies found buspirone to be statistically inferior or equivalent to the GABA PAMs for the acute reduction of anxiety symptoms in GAD, for instance, an adjustment disorder with anxious mood. One clinical trial found buspirone to have significantly less anxiolytic efficacy than either lorazepam (Ativan) or alprazolam (Xanax), which is not surprising since the azapirones do not produce the sedation or muscle relaxation that is associated with the PAMs. In fact, they often produce unpleasant side effects such as nausea and light-headedness. However, two head-to-head studies did not find any statistical difference between buspirone and alprazolam/diazepam (Valium) with respect to the reduction of anxiety symptoms. Furthermore, it has been demonstrated that recent (<1 month) PAM treatment negatively predicts treatment response to buspirone. Specifically, there was no significant difference in response to buspirone or diazepam in patients with no prior exposure to GABA_A PAMs, but as expected, diazepam was significantly more effective than buspirone at reducing anxiety symptoms in patients with recent exposure to PAMs. It is conceivable that patients allocated to the buspirone arm may have experienced rebound phenomena, whereas those on PAMs did not.

A comparison of buspirone and the 5-HT/NE reuptake inhibitor (SNRI) venlafaxine XR (Effexor XR) in reducing anxiety symptoms in GAD did not reveal a significant difference. SSRIs are the first-line treatment for GAD, and the paucity of studies comparing buspirone and other azapirones to SSRIs remains a large gap in the literature. Further trials are needed to compare the use of other first-line medications that are the standard treatment for GAD (i.e., SSRIs, SNRIs, pregabalin) with buspirone to adequately understand its role in this indication. One study found no statistical difference between buspirone and kava-kava with respect to reducing anxiety symptoms in GAD. Kava-kava is a herbal medication with constituents that putatively act both as a monoamine oxidase B inhibitor and as a NE reuptake inhibitor. However, problems with such natural products include the variability in the dose and quality of the active ingredient(s). Another trial found buspirone to be equivalent to hydroxyzine (Vistaril) in reducing anxiety symptoms in GAD. Hydroxyzine is a first-generation antihistamine of the piperazine class that acts via antagonism of the H_1 receptor and is considered a second-line pharmacotherapy for GAD. Further studies comparing selective 5-HT_{1A}

agonists to other alternative pharmacologic and nonpharmacologic interventions are necessary to position these drugs adequately in our armamentarium.

Panic, Post-Traumatic Stress, and Obsessive-Compulsive Disorder

Double-blind, randomized, placebo-controlled trials found that buspirone was not significantly superior to placebo in treating patients with panic disorder. In addition, one of these trials also found alprazolam to be superior to buspirone and placebo with respect to its panicolytic and anxiolytic effects. With respect to PTSD, smaller, open-label trials found buspirone to be effective in alleviating symptoms such as anxiety, insomnia, flashbacks, and depression. In one of these small studies, seven of eight patients exhibited a significant reduction in symptoms as measured by the Structured Inventory for PTSD and the Beck Depression Inventory with buspirone monotherapy. In another case series trial, 12 combat-related PTSD patients exhibited a 40% reduction in re-experiencing symptoms after usage of buspirone as an adjunct to therapy programs, with or without other pharmacologic agents. The promising results from these smaller, open-label trials highlight the need for larger, placebo-controlled studies for these two disorders.

A few double-blind, randomized, placebo-controlled trials evaluated the addition of buspirone to SSRIs for patients with OCD. Two trials added buspirone to SSRIs (fluoxetine [Prozac] and fluvoxamine [Luvox]), and another used buspirone as an adjunct to an SNRI clomipramine (Anafranil). These trials found no significant improvement in Yale-Brown Obsessive Compulsive Scale scores associated with the addition of buspirone. The use of buspirone as an adjunctive treatment for OCD is therefore not recommended in the guidelines.

Table 33.5b–1 provides a summary of clinical trials of buspirone for psychiatric disorders.

Table 33.5b–1.
Buspirone Clinical Trials

	Number of Subjects	Duration	Dose	Significant Difference vs. Comparator
Generalized Anxiety Disorder (GAD)				
Buspirone vs. Placebo				
Ross and Matas, 1987	21	4 weeks	5–40 mg	NS
Enkelmann, 1991	62	6 weeks	18.7 mg (mean)	+
Gammans et al., 1992	Meta-analysis			+
Chiaie et al., 1995	44	8 weeks	15 mg	+
Sramek et al., 1996	162	6 weeks	5–45 mg	+
Lader and Scotto, 1998	163	4 weeks	20 mg	+
DeMartinis et al., 2000	487	4 weeks	30 mg (up to)	+
Bond et al., 2002	44	8 weeks	30 mg (up to)	+
Majercsik et al., 2003	52	6 weeks	30 mg	+
Buspirone vs. Benzodiazepine				
Cohn et al., 1986 (lorazepam)	40	4 weeks	15–50 mg	—
Cohn et al., 1986 (alprazolam)	40	4 weeks	15–50 mg	—
Ross and Matas, 1987 (diazepam)	19	4 weeks	5–40 mg	NS
Enkelmann, 1991 (alprazolam)	63	6 weeks	18.7 mg	NS
DeMartinis et al., 2000 (diazepam)	483	4 weeks	30 mg (up to)	—
Buspirone vs. Hydroxyzine				
Lader and Scotto, 1998	162	4 weeks	20 mg	NS
Buspirone vs. Venlafaxine XR				
Davidson et al., 1999	267	8 weeks	30 mg	NS
Buspirone vs. Kava-Kava				
Boerner et al., 2003	85	8 weeks	10 mg	NS
Panic Disorder				
Buspirone vs. Placebo				
Pohl et al., 1989 (+ imipramine)	60	8 weeks	29.5 mg (mean)	NS
Sheehan et al., 1990	52	8 weeks	57.2 mg	NS
Sheehan et al., 1993 (+ alprazolam)	92	8 weeks	61 mg (mean)	—
Post-Traumatic Stress Disorder (PTSD)				
Buspirone (open-label trial)				
Wells et al., 1991	3	NK	35–60 mg	+
Duffy and Malloy, 1994	8	NK	5–30 mg	+
Fichtner and Crayton, 1994	12	NK	46 mg (mean)	+
Obsessive-Compulsive Disorder (OCD)				
Buspirone vs. Placebo				
Piggott et al., 1992 (+ clomipramine)	14	10 weeks	57 mg (mean)	—
McDougle et al., 1993 (+ fluvoxamine)	33	6 weeks	60 mg (up to)	—
Grady et al., 1993 (+ fluoxetine)	13	4 weeks	60 mg	—
Major Depressive Disorder (MDD)				
Buspirone vs. Placebo				
Fabre, 1990	140	8 weeks	54 mg (mean)	+
Rickels et al., 1990	155	8 weeks	90 mg (up to)	+
Robinson et al., 1990	382	NK	90 mg (up to)	+
Appelberg et al., 2001 (+ citalopram or fluoxetine)	102	6 weeks	10–30 mg	+
Buspirone vs. Bupropion SR				
Trivedi et al., 2006 (+ citalopram)	565	6 weeks	41 mg (mean)	NS
Alcohol Dependence				
Buspirone vs. Placebo				
Malcolm et al., 1992	67	24 weeks	45–60 mg	NS
Malec et al., 1996	Meta-analysis	—	—	NS
Cannabis Dependence				
Buspirone vs. Placebo				
McRae-Clark et al., 2009	50	12 weeks	60 mg (up to)	NS
McRae-Clark et al., 2015	175	12 weeks	60 mg (up to)	NS
Premenstrual Dysphoric Disorder				
Landén et al., 1999	117	4 weeks	20–60 mg (flexible)	+
Landén et al., 2001	63	4 menstrual cycles	10–40 mg (flexible)	+
Nazari et al., 2013	100	8 weeks	10 mg	+

NS, no significant difference; (+), significantly greater improvement in primary efficacy measure for buspirone arm; (–), significantly greater improvement in primary efficacy measure for comparator arm; NK, not known (unpublished).

Major Depressive Disorder

Several large, double-blind, randomized, placebo-controlled trials have examined the efficacy of buspirone in treating patients with MDD. These clinical trials have generally shown buspirone to be significantly superior to placebo in reducing depressive symptoms. One such study compared buspirone to placebo as an adjunct therapy in patients who were treatment-resistant to citalopram. Buspirone was found to be significantly more effective than placebo at reducing depressive symptoms in these patients. Most of these trials used the HAM-D scale as the standard outcome measure of severity of depressive symptoms, although the Montgomery-Asberg Depression Rating Scale was also employed since the score on this measure is less affected by anxiety symptoms making it a better outcome of response to treatment for depression, specifically. The large NIMH-sponsored Sequenced Treatment Alternatives in Resistant Depression (STAR*D) trial included an arm of the study in which MDD patients who did not remit during a treatment course with citalopram could opt for the addition of buspirone or bupropion sustained release (Wellbutrin SR) in a double-blind manner. Both adjunctive medications resulted in similar remission rates, 39% versus 31%, respectively; however, buspirone was not as well-tolerated as bupropion.

Other Uses

Alcohol and Cannabis Dependence. Buspirone use has been examined in alcohol dependence. One clinical trial found that while buspirone reduced anxiety symptoms in male veterans dependent on alcohol, this change was not significantly different from the reductions associated with placebo. In addition, buspirone was found to be statistically equivalent to placebo with respect to standard measures of alcohol consumption, such as time to first drink and the number of standard drinks per drinking day. Similarly, a meta-analysis of five double-blind RCTs concluded that buspirone was not significantly superior to placebo in reducing alcohol consumption. That report did, however, highlight the positive effects of buspirone on treatment persistence and anxiety. The authors suggested that while buspirone did not affect alcohol consumption, it may have had a positive impact on associated psychiatric symptoms. Therefore, given its lack of interaction with alcohol, buspirone may have a role as an adjunct therapy in the treatment of co-occurring anxiety and alcohol dependence.

Treatment of cannabis dependence with buspirone was first examined in a placebo-controlled study of 50 individuals titrated to a target dose of 60 mg daily and treated for 12 weeks. There were no significant differences between the buspirone-treated individuals and those receiving a placebo on self-reported and urine drug detection measures of cannabis use. Similarly, in another controlled study of 175 participants by the same group, no statistically significant differences were identified in the buspirone arm. However, a greater number of female than male subjects taking buspirone had positive urine drug screens. Therefore, buspirone is not recommended for cannabis dependence.

Smoking Cessation. The efficacy of buspirone to help with smoking cessation has been evaluated in open-label trials and a small number of randomized controlled trials. The evidence of efficacy remains inconclusive. A comparison between buspirone and nicotine transdermal patches found both treatments had similar cessation rates, while another study looking exclusively at buspirone found that the greatest improvement was observed in patients reporting high anxiety. When buspirone was coadministered with sertraline (Zoloft) versus placebo in participants who were all receiving cognitive-behavioral therapy (*n* = 98), the combination of buspirone and sertraline improved abstinence rates by 44% compared to 17% for controls at the 26th week following smoking cessation.

Agitation. The possible role of buspirone in the management of behavioral and psychological symptoms of dementia in the context of Alzheimer disease (AD) has been examined. Buspirone was compared to trazodone (Desyrel) and placebo in AD patients with behavioral symptoms (*n* = 10) in a 12-week study. Only subjects taking trazodone showed improvement compared to placebo. In another study looking at an AD population, buspirone (*n* = 12) was also compared to haloperidol (Haldol) (*n* = 14) for the treatment of agitation. Subjects taking buspirone showed only a trend toward decreased tension when compared to subjects taking haloperidol.

Psychosis. Buspirone has not been shown to have intrinsic therapeutic activity in the treatment of psychosis. In the context of schizophrenia, there is no evidence of significant benefits in terms of alleviation of either positive or negative symptoms. This is supported by a meta-analysis of seven RCTs (six with buspirone and one with tandospirone) in 368 patients with schizophrenia. Such adjunctive buspirone was, however, associated with improvement in extrapyramidal symptoms and cognitive benefits while not affecting adverse drug reactions or discontinuation rates.

Somatic Dysfunction. Premenopausal patients with fibromyalgia (n = 22) and healthy female controls (n = 14) were challenged with a single dose of 60 mg of buspirone. Prolactin and growth hormone levels, as well as core body temperature, were measured. Only an increase in prolactin response was observed. The authors concluded that buspirone must only have altered the dopamine system since growth hormone levels and core body temperature after buspirone administration have been shown to be mediated by 5-HT$_{1A}$ receptors, whereas prolactin is mostly controlled by both systems.

Since 5-HT is an important neurotransmitter in the gastrointestinal tract where it plays a key role in the regulation of sensory and motor functions, buspirone has been investigated for its role in altering gastrointestinal functions. It appeared to be efficacious in functional dyspepsia, as it enhanced gastric relaxation and reduced postprandial symptoms in response to a challenge meal. Large, controlled studies are nevertheless needed to establish a clear benefit for such problem(s).

Sexual Dysfunction. Improvement in sexual function has been previously reported in patients with GAD and decreased sexual function after 4 weeks of buspirone. Sexual iatrogenic dysfunction occurs in at least half of individuals taking SSRIs or SNRIs. Buspirone has thus been investigated for its capacity to reverse this side effect. Patients being treated for a major depressive episode (MDE) with a stable dose of an SSRI for at least 4 weeks were given either buspirone or placebo for the primary purpose of evaluating for an additive antidepressant response and, secondarily, to determine if an improvement in sexual function would occur. An attenuation of sexual dysfunction was reported in 58% of patients and only 30% of those on placebo. The effect was more pronounced in women, and since the benefits in women were prominent in the first week, it was concluded that this was not simply attributable to an antidepressant action of buspirone. These results appear to be consistent with the finding that the SSRI/potent 5-HT$_{1A}$ agonist vilazodone produces low levels of sexual dysfunction. Furthermore, the approved indication of the 5-HT_{1A} agonist/5-HT_{2A} antagonist flibanserin in women with hypoactive sexual desire disorder would also appear to be consistent with buspirone having some beneficial action in sexual dysfunction. However, an 8-week double-blind trial in 61 females who had been treated successfully with fluoxetine for a depressive episode compared buspirone to the antiviral/dopamine enhancer amantadine (Symmetrel) and did not show any significantly improved sexual dysfunction compared to placebo.

Premenstrual Dysphoric Disorder. Buspirone has demonstrated some benefit in the treatment of premenstrual dysphoric disorder syndrome (PMDD) in three RCTs. The first RCT compared buspirone (*n* = 19; mean daily dose: 27 mg) to the 5-HT_2 antagonist/5-HT_{1A} agonist nefazodone (*n* = 22; mean daily dose 304 mg) and placebo (*n* = 22). Participants taking buspirone had greater global improvement and improvement in irritability compared to those taking placebo. The second RCT reported that subjects on buspirone (*n* = 17) had a significant improvement in related symptoms, including aches and pains, cramping, impaired social interaction, and irritability. In a third trial, 100 females were randomized to buspirone (10 mg/day) or fluoxetine (20 mg/day) for 2 months. Both drugs produced the same beneficial action.

Sleep Apnea. Buspirone as a treatment option for sleep apnea has been studied. A placebo crossover study in 16 male patients with heart failure (left ventricular ejection fraction <50% and moderate to severe central apnea >14 events/hour) demonstrated reduced carbon dioxide chemosensitivity and improved oxygen saturation across the 24-hour cycle.

Attention-Deficit/Hyperactivity Disorder. Four RCTs (*n* = 362) examined the therapeutic potential of buspirone in ADHD in children and adolescents. One head-to-head comparison with methylphenidate showed benefits, while the other did not. The design of the largest study (*n* = 241) included placebo, atomoxetine (a selective NE reuptake inhibitor), and atomoxetine + buspirone groups. The effect size of atomoxetine was 0.40, whereas that of the combination group was 0.51. This difference was not significant, but most measures showed a trend favoring the group receiving buspirone addition. An RCT (*n* = 57) administering buspirone via a transdermal patch did not report a significant decrease in ADHD-related symptoms in the buspirone group compared to placebo. In summary, the weight of the evidence is not in favor of using buspirone in ADHD.

Withdrawal from Other Medications/Substances of Abuse. Two placebo RCTs (*n* = 49) assessed the potential of buspirone to limit withdrawal symptoms from opiates after 5 days of stabilization on 30 mg of methadone. Both studies found that buspirone was as effective as a methadone taper in alleviating the withdrawal symptoms of subjects dependent on heroin. Another placebo RCT of 12 weeks in duration (*n* = 36) did not find a decrease in anxiety among individuals dependent on opioids receiving continuous methadone treatments. There was, however, a trend toward reduction of depressive symptoms and a longer period prior to substance use relapse in individuals receiving buspirone. Therefore, there seems to be some benefit to using buspirone in opioid dependence.

An RCT looked at the discontinuation of long-term use of alprazolam, diazepam, or lorazepam in 107 patients with GAD who had, on average, used these medications for 8.5 years. Patients received a placebo, buspirone (mean daily dose 38 mg), or imipramine (Tofranil; mean daily dose 180 mg) over a 5-week period during a gradual 4- to 6-week PAM discontinuation at a rate of 25% per week. The success rate of the taper was significantly higher for patients in the imipramine (83%) but not in the buspirone group (68%) as compared to the placebo group (38%) 3 months after the taper.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Nervous System

The most frequent adverse events affecting the nervous system reported by buspirone users included dizziness, headache, drowsiness, light-headedness, insomnia, fatigue, nervousness, decreased concentration, abnormal thinking, excitement, depression, confusion, vivid dreams, and hostility. Less frequent events included noise intolerance, euphoria, fear, apathy, seizures, and suicidal thoughts.

Gastrointestinal System

Users of buspirone report instances of nausea, GI distress, diarrhea, and vomiting. Less frequent events include anorexia and rectal bleeding.

Cardiovascular System

Tachycardia or palpitations was the most frequently listed cardiovascular effect with the use of buspirone. Less common adverse events included syncope, hypotension, and hypertension. Rare cardiovascular effects include congestive heart failure, cerebrovascular accident, myocardial infarction, cardiomyopathy, bradycardia, and ECG change. The causality of these adverse events is, however, difficult to ascertain.

Dermatologic Effects

The most frequent dermatologic side effect reported with buspirone use was skin rash, while facial edema, easy bruising, blisters, hair loss, and dry skin were reported infrequently. Rare events included acne and thinning of nails.

Endocrine System

Patients taking buspirone reported the following infrequent adverse events: change in libido, change in weight, and menstrual irregularity.

THERAPEUTIC INDICATIONS

The FDA has approved buspirone for the management of anxiety disorders or for short-term treatment of anxiety-related symptoms. It is recommended for use in individuals who present symptoms that correspond with GAD based on the former DSM criteria, that is, for at least 1 month. There are no data available pertaining to the long-term use of buspirone.

PRECAUTIONS AND ADVERSE REACTIONS

Patients with hypersensitivity to buspirone should not take this medication. Pharmacologic studies have indicated that liver and kidney impairment affect drug metabolism and drug clearance, respectively. As such, as indicated above, alternative treatments should be sought for patients with liver impairment, whereas those with kidney impairment should receive less than the recommended dose. Buspirone is not recommended for patients with severe liver or kidney impairment. In vivo animal studies have indicated a lack of teratogenic, carcinogenic, and fetal impairment effects of buspirone administration, but similarly, well-controlled studies in humans have not been performed, resulting in a designation of Pregnancy Category B according to the former U.S. FDA classification. In rats, buspirone and its metabolites are found in maternal milk and cross the placenta, although it is not known to what extent both these events occur in humans.

The most commonly listed adverse reactions to buspirone that were significantly higher when compared to placebo included dizziness, nausea, headache, light-headedness, excitement, and sweating or clamminess. A study in a child and adolescent population found that the most frequently reported side effects were light-headedness, headache, and dyspepsia, while studies in geriatric populations found no difference in adverse events reported in comparison to a younger population. Rare side effects include congestive heart failure, bradycardia, slurred speech, claustrophobia, tingling of limbs, burning tongue, and photophobia.

DRUG INTERACTIONS

Buspirone is a substrate for CYP3A4. Concurrent use of buspirone and MAOIs has been associated with a significant increase in blood pressure. Verapamil (Verelan), an inhibitor of CYP3A4, was found to increase the AUC and C_{max} of buspirone by 4.1- and 5.0-fold, respectively. Erythromycin (ERY-C), an inhibitor of CYP3A4, increased both AUC and C_{max} by 19- and 16-fold, respectively. Rifampin (Rifadin), an inducer of CYP3A4, decreased the C_{max} and AUC by 84% to 90% and the half-life from 2.8 to 1.3 hours. With the coadministration of haloperidol and buspirone in subjects with schizophrenia, buspirone increased the plasma concentration of both haloperidol and its metabolite, although this study did not examine the concentration of buspirone in the presence of haloperidol. Grapefruit juice, an inhibitor of CYP3A4, increased the AUC and C_{max} of buspirone by 9.2- and 4.3-fold, respectively. Other known inducers or inhibitors of CYP3A4 that may affect buspirone plasma concentration include ketoconazole (Nizoral), dexamethasone (Decadron), phenytoin (Dilantin), phenobarbital (Bellatal), and carbamazepine (Tegretol). Concurrent use of certain drugs has been found to alter the percentage of protein-bound buspirone. Acetylsalicylic acid (ASA) increased plasma levels of free buspirone by 23%, while flurazepam (Dalmane) decreased levels by 20%. Other drugs that alter the free plasma level of this drug include desipramine (Norpramin), diazepam, ibuprofen (Advil), propranolol (Inderal), thioridazine (Mellaril), and tolbutamide (Orinase).

DOSAGE AND ADMINISTRATION

Buspirone is available as an oral tablet and is provided in the following strengths: 5, 10, 15, and 30 mg. The recommended initial dose listed by the FDA is 15 mg daily (7.5 mg twice a day). Patients may be titrated to an optimal dose within 2 to 3 days by increasing the dose by 5 mg daily. The maximum daily dose should not surpass 60 mg.

FURTHER READINGS

Appelberg BG, Syvälahti EK, Koskinen TE, Mehtonen OP, Muhonen TT, Naukkarinen HH. Patients with severe depression may benefit from buspirone augmentation of selective serotonin reuptake inhibitors: results from a placebo-controlled, randomized, double-blind, placebo wash-in study. *J Clin Psychiatry*. 2001;62(6):448–452.

Bielski RJ, Cunningham L, Horrigan JP, Londborg PD, Smith WT, Weiss K. Gepirone extended-release in the treatment of adult outpatients with major depressive disorder: a double-blind, randomized, placebo-controlled, parallel-group study. *J Clin Psychiatry*. 2008;69(4):571–577.

Blier P, Bergeron R, de Montigny C. Selective activation of postsynaptic 5-HT1A receptors induces rapid antidepressant response. *Neuropsychopharmacology*. 1997;16(5):333–338.

Blier P, Curet O, Chaput Y, de Montigny C. Tandospirone and its metabolite, 1-(2-pyrimidinyl)-piperazine–II. Effects of acute administration of 1-PP and long-term administration of tandospirone on noradrenergic neurotransmission. *Neuropharmacology*. 1991;30(7):691–701.

Cantillon M, Brunswick R, Molina D, Bahro M. Buspirone vs. haloperidol: a double-blind trial for agitation in a nursing home population with Alzheimer's disease. *Am J Geriatr Psychiatry*. 1996;4(3):263–267.

Carrão JL, Moreira LB, Fuchs FD. The efficacy of the combination of sertraline with buspirone for smoking cessation: a randomized clinical trial in nondepressed smokers. *Eur Arch Psychiatry Clin Neurosci*. 2007;257(7):383–388.

Chessick CA, Allen MH, Thase M, et al. Azapirones for generalized anxiety disorder. *Cochrane Database Syst Rev*. 2006;3:CD006115.

Chial HJ, Camilleri M, Ferber I, et al. Effects of venlafaxine, buspirone, and placebo on colonic sensorimotor functions in healthy humans. *Clin Gastroenterol Hepatol*. 2003;1(3):211–218.

Cinciripini PM, Lapitsky L, Seay S, Wallfisch A, Meyer WJ 3rd, van Vunakis H. A placebo-controlled evaluation of the effects of buspirone on smoking cessation: differences between high- and low-anxiety smokers. *J Clin Psychopharmacol*. 1995;15(3):182–191.

DeMartinis N, Rynn M, Rickels K, Mandos L. Prior benzodiazepine use and buspirone response in the treatment of generalized anxiety disorder. *J Clin Psychiatry*. 2000;61(2):91–94.

Eison AS. Azapirones: history of development. *J Clin Psychopharmacol*. 1990;10(3 Suppl):2S–5S.

Giannoni A, Borrelli C, Mirizzi G, Richerson GB, Emdin M, Passino C. Benefit of buspirone on chemoreflex and central apnoeas in heart failure: a randomized controlled crossover trial. *Eur J Heart Fail*. 2021;23(2):312–320.

Godbout R, Chaput Y, Blier P, de Montigny C. Tandospirone and its metabolite, 1-(2-pyrimidinyl)-piperazine–I. Effects of acute and long-term administration of tandospirone on serotonin neurotransmission. *Neuropharmacology*. 1991;30(7):679–690.

Hilleman DE, Mohiuddin SM, Delcore MG. Comparison of fixed-dose transdermal nicotine, tapered-dose transdermal nicotine, and buspirone in smoking cessation. *J Clin Pharmacol*. 1994;34(3):222–224.

Imai H, Tajika A, Chen P, et al. Azapirones versus placebo for panic disorder in adults. *Cochrane Database Syst Rev*. 2014;(9):CD010828.

Katzman MA, Bleau P, Blier P, et al. Canadian clinical practice guidelines for the management of anxiety, posttraumatic stress and obsessive-compulsive disorders. *BMC Psychiatry*. 2014;14(Suppl 1):S1.

Kennedy SH, Lam RW, McIntyre RS, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 Clinical Guidelines for the Management of Adults with Major Depressive Disorder: Section 3. Pharmacological Treatments. *Can J Psychiatry*. 2016;61(9):540–560.

Landén M, Eriksson E, Agren H, Fahlén T. Effect of buspirone on sexual dysfunction in depressed patients treated with selective serotonin reuptake inhibitors. *J Clin Psychopharmacol*. 1999;19(3):268–271.

Landén M, Eriksson O, Sundblad C, Andersch B, Naessén T, Eriksson E. Compounds with affinity for serotonergic receptors in the treatment of premenstrual dysphoria: a comparison of buspirone, nefazodone and placebo. *Psychopharmacology (Berl)*. 2001;155(3):292–298.

Malec TS, Malec EA, Dongier M. Efficacy of buspirone in alcohol dependence: a review. *Alcohol Clin Exp Res*. 1996;20(5):853–858.

McRae-Clark AL, Baker NL, Gray KM, et al. Buspirone treatment of cannabis dependence: a randomized, placebo-controlled trial. *Drug Alcohol Depend*. 2015;156(1–2):29–37.

McRae-Clark AL, Carter RE, Killeen TK, et al. A placebo-controlled trial of buspirone for the treatment of marijuana dependence. *Drug Alcohol Depend*. 2009;105(1–2):132–138.

Nazari H, Yari F, Jariani M, Marzban A, Birgandy M. Premenstrual syndrome: a single-blind study of treatment with buspirone versus fluoxetine. *Arch Gynecol Obstet.* 2013;287(3):469–472.

Peroutka SJ. Selective interaction of novel anxiolytics with 5-hydroxytryptamine1A receptors. *Biol Psychiatry*. 1985;20(9):971–979.

Riblet LA, Taylor DP, Eison MS, Stanton HC. Pharmacology and neurochemistry of buspirone. *J Clin Psychiatry*. 1982;43(12 Pt 2):11–18.

Rickels K, DeMartinis N, García-España F, Greenblatt DJ, Mandos LA, Rynn M. Imipramine and buspirone in treatment of patients with generalized anxiety disorder who are discontinuing long-term benzodiazepine therapy. *Am J Psychiatry.* 2000;157(12):1973–1979.

Rickels K, Freeman E, Sondheimer S. Buspirone in treatment of premenstrual syndrome. *Lancet*. 1989;1(8641):777.

Salazar DE, Frackiewicz EJ, Dockens R, et al. Pharmacokinetics and tolerability of buspirone during oral administration to children and adolescents with anxiety disorder and normal healthy adults. *J Clin Pharmacol*. 2001;41(12):1351–1358.

Stahl SM, Kaiser L, Roeschen J, Keppel Hesselink JM, Orazem J. Effectiveness of ipsapirone, a 5-HT-1A partial agonist, in major depressive disorder: support for the role of 5-HT-1A receptors in the mechanism of action of serotonergic antidepressants. *Int J Neuropsychopharmacol*. 1998;1(1):11–18.

Sutherland SM, Adler LA, Chen C, Smith MD, Feltner DE. An 8-week, randomized controlled trial of atomoxetine, atomoxetine plus buspirone, or placebo in adults with ADHD. *J Clin Psychiatry*. 2012;73(4):445–450.

Trivedi MH, Fava M, Wisniewski SR, et al. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*. 2006;354(12):1243–1252.

33.5c Vilazodone and Vortioxetine

Pierre Blier, M.D., Ph.D., and Charles Desfossés, M.D.

Vilazodone and vortioxetine are two similar medications categorized as multimodal serotonin agents. They share common properties, namely inhibition of serotonin (5-hydroxy-tryptamine; 5-HT) reuptake transporters and agonism of 5-HT_{1A} receptors. For this reason, they are grouped together in this chapter, but vortioxetine also engages other 5-HT receptor subtypes.

VILAZODONE

Introduction and History

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic domain	Serotonin
Mechanism of action	Reuptake inhibitor (SERT) and partial agonist (5-HT_{1A})

Vilazodone (VIIBRYD) was approved by the FDA for the treatment of adults with major depressive disorder (MDD) on January 21, 2011, after more than a decade of development. It was first produced by Merck and then developed by GlaxoSmithKline in 2001, but after several failed trials in phase II, further clinical development was stopped in 2003. Clinical Data Inc purchased the compound and submitted a new drug application (NDA) in March 2010. The FDA accepted the New Drug Application for vilazodone for the treatment of MDD shortly thereafter. In March 2015, Actavis received approval for Viibryd 20 mg/day as a therapeutic dose. Actavis was purchased by Allergan, which in turn was bought by AbbVie, now marketing the medication in the United States and Canada.

Chemistry

The chemical structure of vilazodone is 2-benzofurancarboxamide, 5-[4-[4-[5-cyano-1H-indol-3-yl]butyl]-1-piperazinyl]-, hydrochloride (1:1) (Fig. 33.5c–1). It has a molecular weight of 441.535 g/mole. Given the therapeutic activity of selective serotonin reuptake inhibitors (SSRIs) and 5-HT_{1A} agonists in mood and anxiety disorders, the molecule was produced by combining the respective structural moieties responsible for such properties, an indole-butyl-amine and chromenonyl-piperazine, respectively, into a single compound.

Pharmacologic Actions

Pharmacokinetics. The pharmacologic activity of vilazodone resides in the parent compound, and its metabolites do not possess significant activity. Its pharmacokinetics are dose-proportional in a dose range of 5 to 80 mg, and peak concentrations occur at 4 to 5 hours after administration. When it is taken with food, vilazodone has an absolute bioavailability of 72%. The drug has a large volume of distribution and is 96% to 99% protein bound. It has a terminal half-life of about 25 hours and generally achieves a steady-state level within 4 days. With daily dosing of 40 mg with food, it reaches a maximum plasma concentration (C_{max}) of 156 ng/mL with an area under the curve (AUC) over 24 hours of 1,646 ng-hours/mL. The C_{max} and AUC of vilazodone are markedly lowered by about 50% when taken while fasting. Therefore, vilazodone should be taken with food to favor its absorption and, as with SSRIs, decrease gastrointestinal side effects. The drug is extensively metabolized, with only 1% and 2% recovered unchanged in urine and feces, respectively. Its metabolism primarily occurs by the cytochrome (CYP) 3A4, with CYP2C19 and CYP2D6 isoenzymes being minor catabolic pathways. Dose adjustments with mild to moderate hepatic and renal impairments are not necessary as its clearance is not affected in such individuals.

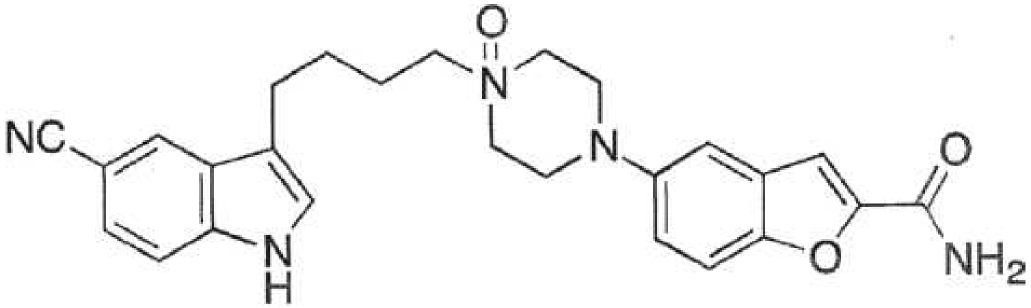

FIGURE 33.5c–1. Chemical structure of vilazodone.

Mechanisms of Action. Vilazodone is an agent that selectively targets the 5-HT system: (1) its affinity for 5-HT reuptake transporters is very high (K_i = 0.1 nM) without significant affinity for the norepinephrine (NE) and dopamine (DA) reuptake transporters (K_i = 56, and 37 nM, respectively), (2) its affinity for 5-HT_{1A} receptors is in the same range as for 5-HT reuptake transporters (K_i = 0.5 nM), and without any relevant affinity for any other 5-HT, NE, DA, and cholinergic receptors. In some secondary messenger systems coupled to 5-HT_{1A} receptors, vilazodone acts as a partial 5-HT_{1A} agonist.

Vilazodone has a delayed onset of therapeutic action similar to other SSRIs and 5-HT_{1A} receptor agonists. The potent 5-HT_{1A} property of the drug was demonstrated in vivo in anesthetized rats depleted of 5-HT:vilazodone injected intravenously was still able to inhibit the firing rate of 5-HT neurons through their 5-HT_{1A} autoreceptors. As expected, when administered in a sustained fashion to rats, the firing rate of 5-HT neurons is initially markedly attenuated, thereby decreasing endogenous 5-HT release. However, since vilazodone is present throughout the brain as an exogenous 5-HT_{1A} agonist, partial compensation of 5-HT transmission at postsynaptic 5-HT_{1A} receptors by the medication is expected. With prolonged administration, the firing rate of 5-HT neurons recovers to normal due to desensitization of their 5-HT_{1A} autoreceptors. Subsequently (i.e., after 14 days of administration through a constant release device), increased tonic activation of postsynaptic 5-HT_{1A} receptors occurs, as has been demonstrated with SSRIs and 5-HT_{1A} agonists separately using the same models. Despite its two combined molecular targets, there is no evidence indicating a more rapid onset of therapeutic action by vilazodone.

Therapeutic Indications

To date, the only FDA-recommended indication for vilazodone is for the treatment of MDD in adults. Its efficacy was initially established in two 8-week, randomized, double-blind, placebo-controlled trials in adult patients with a diagnosis of MDD. However, given that the 5-HT_{1A} agonist buspirone is FDA approved for generalized anxiety disorder (GAD) and that several SSRIs are also approved for the treatment of anxiety disorders such as GAD, social anxiety disorder, panic disorder, and separation anxiety disorder, as well as obsessive-compulsive disorder (OCD), posttraumatic stress disorder (PTSD), premenstrual dysphoric disorder, and eating disorders, it is reasonable to assume that vilazodone may also prove useful in treating these conditions. However, only controlled and replicated clinical trials can ascertain the potential benefits of vilazodone for these other disorders.

Clinical Drug Studies

Design. Twenty-four phase I studies were conducted in the clinical development program of vilazodone. For its development in the treatment of MDD, five placebo-controlled phase II trials, three phase III trials (two double-blind, placebo-controlled 8-week trials that led to its FDA approval), and three phase IV trials were conducted, including a relapse prevention study. Although there have been several positive controlled studies of vilazodone, initial trials were failed studies. A main reason for these failed trials relates to the dosing strategies used. Two phase II trials used a fixed low dose of 20 mg/day and only showed a trend toward efficacy. The other three phase II trials were comprised of flexible regimens using flexible dosing strategies. While two of the studies used doses up to 100 mg/day, they consequently did not provide comprehensive information on the dose–response relationship. Three of these phase II trials included an active comparator, which did not separate from the placebo, indicating a lack of sensitivity in the studies.

Interpretation of Drug Studies. There were two phase III 8-week, multicenter, randomized, double-blind, placebo-controlled studies in adults (aged 18 to 65 in one trial and 18 to 70 in the other) with MDD that led to its FDA approval (Table 33.5c–1). These trials encompassed standard inclusion and exclusion criteria and used standard scales to assess clinical responses. The primary outcome measure was the Montgomery–Åsberg Depression Rating Scale (MADRS). Secondary outcomes were the Hamilton Depression-17 item (HAM-D-17), Hamilton Rating Scale for Anxiety (HAM-A), Clinical Global Impressions Severity of Illness scale (CGI-S), and Clinical Global Impressions Severity of Improvement scale (CGI-I). Patients were titrated within 2 weeks to a daily dose of 40 mg/day, but they were allowed to reduce the doses to 20 mg/day if the higher dose was not tolerated. Interestingly, both studies included measures of sexual functioning, whereas most prior studies of medications for depression only relied on spontaneous reporting of sexual side effects. The latter methodology is known to markedly underestimate sexual side effects, especially with SSRIs. The first study used the Changes in Sexual Functioning Questionnaire (CSFQ), whereas the second study used the Arizona Sexual Experiences Scale (ASEX).

Most patients in more recent phase III and IV trials had moderate to severe depression at baseline, with a score of 22 or more on the HAM-D-17 for inclusion but with the MADRS to assess response. Using a scale to assess the antidepressant response that is different from the one used for the study entry likely contributed to signal detection, all the more so that the MADRS is less affected by co-occurring anxiety symptoms and is more sensitive to change. Both trials showed superiority over the placebo on the total MADRS score, with small effect sizes (0.23 and 0.30). Based on the observation that vilazodone separated from placebo at week 1 in the first trial but not until week 6 in the second trial, a claim of rapid onset of action cannot be made. In the absence of an active comparator, such a claim appears even more tenuous.

Vilazodone was developed in light of the evidence that the 5-HT$_{1A}$ agonist buspirone is efficacious for GAD. Therefore, secondary measures assessing anxiolysis were examined. By combining the results of the two 8-week phase III trials, a posthoc analysis classifying patients as having anxious depression or nonanxious depression showed significantly greater decreases in both depression and anxiety symptoms versus those taking placebo. Although these results suggest a possible role for vilazodone in the treatment of depression with prominent anxiety symptoms, active comparator studies are needed to determine whether vilazodone has any advantage over other medications. Such a possibility could be supported by the three positive placebo-controlled trials assessing the anxiolytic effect of vilazodone in GAD. All three trials included 200 or more patients in each treatment arm, two were flexible-dose trials (20 to 40 mg/day), whereas a third one had a 20 mg/day and a 40 mg/day group, and only the latter dose separated significantly from the placebo group. GAD is not an official indication in the United States or Canada.

The positive results of the two phase III trials in MDD are now supported by two large phase IV placebo-controlled trials, one of which included an active comparator for assay sensitivity. The first one had a flexible dose (20 to 40 mg/day) design, whereas the second used fixed doses of 20 and 40 mg/day of vilazodone with a 40 mg/day group of the SSRI citalopram, the latter group being titrated over 2 weeks. All three active arms separated significantly from placebo but were not different from each other. A double-blind placebo-controlled maintenance study did not show significantly lower relapse rates in the groups that continued their 20 or 40 mg/day regimen when compared to a switch to placebo after patients had been stabilized on vilazodone. This study may have failed because of the unexpectedly low relapse rates of 11% to 13% in all three groups. Indeed, relapse rates are usually twice as high in such studies in patients who responded to active medication and are then switched to placebo.

Precautions and Adverse Reactions

Vilazodone was not associated with any significant changes in laboratory tests, blood pressure, heart rate, or weight during placebo-controlled trials.

Table 33.5c–1.
Main Short-Term Randomized Controlled Trials of Vilazodone

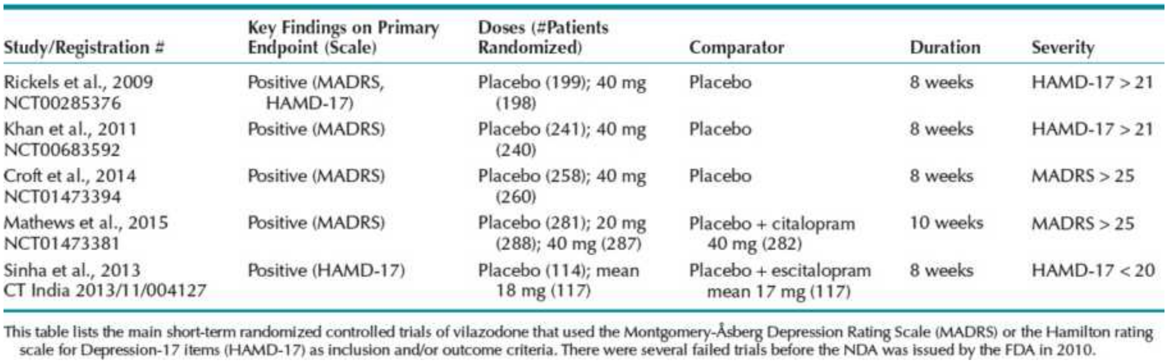

Study/Registration #	Key Findings on Primary Endpoint (Scale)	Doses (#Patients Randomized)	Comparator	Duration	Severity
Rickels et al., 2009 NCT00285376	Positive (MADRS, HAMD-17)	Placebo (199); 40 mg (198)	Placebo	8 weeks	HAMD-17 > 21
Khan et al., 2011 NCT00683592	Positive (MADRS)	Placebo (241); 40 mg (240)	Placebo	8 weeks	HAMD-17 > 21
Croft et al., 2014 NCT01473394	Positive (MADRS)	Placebo (258); 40 mg (260)	Placebo	8 weeks	MADRS > 25
Mathews et al., 2015 NCT01473381	Positive (MADRS)	Placebo (281); 20 mg (288); 40 mg (287)	Placebo + citalopram 40 mg (282)	10 weeks	MADRS > 25
Sinha et al., 2013 CT India 2013/11/004127	Positive (HAMD-17)	Placebo (114); mean 18 mg (117)	Placebo + escitalopram mean 17 mg (117)	8 weeks	HAMD-17 < 20

This table lists the main short-term randomized controlled trials of vilazodone that used the Montgomery-Åsberg Depression Rating Scale (MADRS) or the Hamilton rating scale for Depression-17 items (HAMD-17) as inclusion and/or outcome criteria. There were several failed trials before the NDA was issued by the FDA in 2010.

Safety and Tolerability. The safety of vilazodone was specifically evaluated using data from 2,177 patients who participated in phase III clinical studies, representing 552 patient years of exposure. The rate of discontinuation due to adverse events in clinical trials is considered a general indicator of tolerability. Vilazodone appears to be well tolerated overall. In pooled phase III data, discontinuation rates due to adverse effects were relatively low (7% vilazodone, 3% placebo). Rates of discontinuation due to lack of efficacy were not significantly different between the vilazodone and placebo groups. While no one single adverse event was responsible for discontinuation in >1% of the patients, the most common adverse event type leading to discontinuation were gastrointestinal complaints (2.3% vilazodone vs. 0.2% placebo). Adverse events significantly more common in vilazodone-treated patients compared with those receiving placebo were diarrhea, nausea, dizziness, insomnia, vomiting, abnormal dreams, gastroenteritis, restlessness or akathisia, palpitations, tremor, feeling jittery, and paresthesia; although in most cases symptoms were considered to be mild to moderate. Only 6.4% and 8.8% of participants in the two phase III trials reported a severe adverse event, compared to 5.6% and 5.4% in the placebo groups, respectively. In the vilazodone-treated group, symptoms in the first trial were diarrhea, fatigue, flatulence, gastroenteritis, palpitations, and paresthesias. In the second trial, symptoms reported were insomnia, nausea, vomiting, headache, and decreased libido.

Serious Adverse Events. There were no deaths in the vilazodone clinical trials in subjects taking vilazodone. Among all of the clinical trials, there were three deaths. These included a homicide in a phase I study and two suicides in a phase II study. Neither of the suicide victims was in the vilazodone arm. A total of 76 patients experienced a nonfatal serious adverse event in phase II and phase III trials, with similar rates for the vilazodone and placebo groups.

Overdose. Limited data exist regarding the effects of a vilazodone overdose. Four patients and the child of a patient experienced an overdose of vilazodone at doses between 200 and 280 mg. Symptoms included 5-HT syndrome, lethargy, restlessness, hallucinations, and disorientation. All patients recovered.

Sexual Dysfunction. Only head-to-head comparator studies using specific questionnaires on sexual function can delineate the impact of different medications. In the case of vilazodone, there is a 5-week study in 199 healthy male and female volunteers using doses of 20 and 40 mg/day in comparison to placebo and a 20 mg/day group on the SSRI paroxetine. Based on plasma level determinations, about 20% of participants did not have any detectable levels of active medication. When excluding these individuals, the CSFQ total score was not significantly different between either dose of vilazodone versus placebo. Nevertheless, the CSFQ total score decreased significantly for those participants on paroxetine versus those on placebo and vilazodone 20 mg/day but not versus vilazodone 40 mg/day. Another study assessed sexual dysfunction during the treatment of 1,147 patients with MDD. Sexual dysfunction, as assessed with the CSFQ, was present in 60% to 70% of women and 50% to 60% of men, with about 55% of the overall sample being female. The CSFQ score changes from baseline were small for all patients with normal baseline sexual function. In contrast, across treatment groups, baseline sexual function improved in men and women at the end of treatment in the MDD responders and in patients with baseline sexual dysfunction. Overall, these results suggest that sexual function improvement appears largely associated with the improvement of depressive symptoms rather than a specific effect of vilazodone.

Weight. Only mild but not clinically relevant changes in body weight were found compared to placebo in the four placebo-controlled 8- to 10-

week studies and in the open-label 52-week study. In the RCTs, vilazodone 20 mg was associated with an increase of +0.59 kg, 40 mg with +0.32 kg and placebo with +0.22 kg. The proportion of patients who had an increase of >7% body weight with 20 mg and 40 mg was 1.7% and 1.3%, respectively, compared to 1.0% for placebo. The proportion of patients who had a decrease of >7% body weight with 20 mg and 40 mg was 1.7% and 0.7%, respectively, compared to 0.7% for placebo.

Cardiac. Vilazodone was not associated with any clinically significant effects on electrocardiograms, arrhythmias, or QTc abnormalities within its recommended dose range (20 to 40 mg/day). Four patients treated with vilazodone in phase III trials discontinued treatment because of palpitations, compared with no patients on placebo. In a study of vilazodone in doses up to 80 mg/day, no QTc abnormalities were observed.

Platelets. Vilazodone may affect platelets, given that SSRI blockade of platelet transport proteins leads to a depletion of more than 90% of 5-HT within platelets and slight inhibition of platelet aggregation. However, in clinical trials, no increased risk of bleeding was observed in patients treated with vilazodone versus placebo, and rates of bleeding events were similar in both groups. As with all SSRIs, the product labeling advises caution when co-administering vilazodone with drugs that potentiate the risk of bleeding.

Hyponatremia. Although no cases of hyponatremia were reported in the clinical trials of vilazodone, hyponatremia has occurred with SSRIs and SNRIs. It is potentially lethal, and risk factors include older age, concomitant use of diuretics or other medications that can cause hyponatremia, volume depletion, and medical conditions causing hyponatremia. The mechanism of action is thought to be due to the development of the syndrome of inappropriate antidiuretic hormone secretion. In patients with symptomatic hyponatremia, vilazodone should be discontinued and appropriate medical intervention implemented.

Gastrointestinal. The most commonly observed adverse events of vilazodone are gastrointestinal in nature. Fortunately, they tend to be mild or moderate in severity and appear to be transient in nature. In the placebo-controlled phase III trials, diarrhea was reported by 28% of vilazodone-treated patients (vs. 9% placebo) and nausea by 23% of vilazodone-treated patients (vs. 5% placebo). Slow dose titration over 2 weeks is recommended to minimize gastrointestinal adverse events.

Suicidality. Vilazodone carries the same black box warning as other SSRIs regarding suicidal thoughts and behaviors in children, adolescents, and young adults. This was based on data showing an increased risk of suicidal thinking and behavior (although no completed suicides were observed) in children, adolescents, and young adults (ages 18 to 24) with MDD and other psychiatric disorders treated with SSRIs or other newer medications for MDD. In pooled phase III placebo-controlled vilazodone trials, suicidal ideation was experienced by one (0.2%) patient receiving vilazodone and three (0.7%) patients on placebo. Seven (1.2%) patients in the 52-week open-label study also experienced suicidal ideation. Suicidal behavior was identified by one (0.2%) vilazodone-treated patient and one (0.2%) patient in the placebo group in the pooled phase III placebo-controlled trials and by two (0.3%) patients in the 52-week open-label trial. In a recent phase IV trial, among patients receiving vilazodone, there was one reported suicidal ideation and one suicide attempt. All patients initiating an efficacious treatment for MDD should be monitored closely for clinical worsening or suicidality, especially upon drug initiation or dose adjustment.

Activation of Mania/Hypomania. As with all medications for the treatment of MDD, cautious monitoring should be exercised when prescribing vilazodone to patients with a history or family history of bipolar disorder, mania, or hypomania because of the risk of a switch. In clinical trials of vilazodone, six patients were identified as having probable treatment-emergent mania or hypomania, five receiving vilazodone and one placebo.

Discontinuation Syndrome. A discontinuation syndrome is a frequent occurrence upon stopping medications that are 5-HT reuptake inhibitors, especially when discontinued abruptly. This is largely linked to those that are rapidly eliminated, that is, within about 1 week, in contrast to fluoxetine, which is eliminated between 4 to 5 weeks and is the least likely to produce discontinuation symptoms. Therefore, abrupt discontinuation of vilazodone is not recommended. Rather, gradual discontinuation should be implemented to avoid a possible discontinuation syndrome.

Serotonin Syndrome. The development of a 5-HT syndrome is always something to consider when using medications that increase 5-HT availability. It is extremely rare when using a single agent within a recommended regimen. However, two subjects in the vilazodone clinical trials were identified as having 5-HT syndrome. The first was a patient in a phase II trial who received vilazodone 80 mg; the other was a patient in the 52-week, open-label, phase III study who took an overdose of approximately 240 mg of vilazodone. These two cases may have resulted from direct overstimulation of 5-HT$_{1A}$ receptors in the presence of optimal 5-HT reuptake inhibition. If a 5-HT syndrome is confirmed, there must be a determination of whether other substances that may stimulate 5-HT transmission, including drugs of abuse like ecstasy, have been ingested. Vilazodone must not be taken within 2, but preferably 3, weeks after stopping an irreversible monoamine oxidase inhibitor (MAOI). If symptoms occur, vilazodone and all other serotonergic agents should be discontinued immediately and supportive management initiated.

Nervous System. Although vilazodone has not been studied in patients with seizure disorders, the product label advises that caution be used when prescribing vilazodone to patients with a seizure disorder. In addition, headache was reported as an adverse event in 20% of patients in the 52-week open-label study.

Sleep. In the two phase III trials, insomnia and abnormal dreams were reported by 6% and 4% of patients on vilazodone, versus 2% and 1%, respectively, of those receiving placebo. Decreased REM and increased slow-wave and delta-wave activity were reported in a single-dose study of 10 healthy men.

Reproductive. There are no published data on vilazodone during pregnancy or during labor and delivery. In animal studies, however, vilazodone was not teratogenic in rats or rabbits at doses of 48 and 17 times, respectively, the maximum recommended human dose (MRHD). Developmental toxicity was observed, but only at doses much higher than the MRHD. The FDA package insert carries the same warning regarding potential neonatal toxicity as is listed with other medications used to treat MDD. Clinicians must consider the well-documented risks of untreated depression in pregnancy, including during the post-partum period, versus the associated risks of taking medications with serotonergic agents (see chapter on SSRI for more considerations).

Lactation. There are no published data regarding the excretion of vilazodone in breast milk. Animal studies have expectedly demonstrated that vilazodone is excreted into the milk of lactating rats. Risks and benefits should be carefully weighed and discussed with patients.

Fertility. There are no data on the effect of vilazodone on fertility in humans. In animal studies with rats, impairment of male fertility was observed with vilazodone at 30 times the MRHD but not at 6 times the MRHD. No effect on fertility was observed in female rats.

Carcinogenicity. In mice studies, the incidence of hepatocellular carcinomas in males was increased at 16 times the MRHD but not at 5 times the MRHD. Similarly, the incidence of malignant mammary gland tumors was significantly increased at 16 times the MRHD but not at 1.8 times the MRHD. Vilazodone was not carcinogenic in rats at doses up to 36 times the MRHD.

Mutagenicity. According to the package insert, vilazodone was not mutagenic in the in vitro bacterial reverse mutation assay and was negative in the in vitro V79/HGRPT mammalian cell forward mutation assay. Vilazodone was clastogenic in two in vitro mammalian cell chromosome aberration assays but was negative for clastogenic activity in an in vivo rat bone marrow chromosome aberration assay and a micronucleus test. Vilazodone was also negative in an in vivo/in vitro unscheduled DNA synthesis assay in rats.

Drug Interactions

Since vilazodone is metabolized primarily by CYP3A4, 3A4 inhibitors or inducers can potentially affect its plasma concentration. If co-administered with a strong inhibitor of CYP3A4 (perhaps the most common being the regular drinking of grapefruit juice), it is recommended that the vilazodone dose be limited to 20 mg daily. If co-administered with 3A4 inducers, like carbamazepine, the package insert states that the dose of vilazodone may be increased based on clinical response (not to exceed 80 mg/day). Co-administration of drugs interfering with other CYP isoenzymes will not affect the plasma levels of vilazodone to a clinically significant extent.

Since vilazodone does not modify the activity of other CYP isoenzymes, it will not alter the activity of other medications metabolized by such catabolic pathways. Given the high metabolic activity of the CYP3A4, co-administration of vilazodone with other substrates will not alter its plasma levels. An interaction of vilazodone with other highly protein-bound drugs has not been evaluated, but given its high protein binding, co-administration with other highly protein-bound drugs may result in increased free concentrations of the other drug. There is a low risk of increased bleeding if vilazodone is used with NSAIDs, aspirin, or oral anticoagulants (see SSRI chapter).

Dosage and Administration

Vilazodone is supplied in 10-, 20-, and 40-mg tablets. A 30-day starter kit is available, which contains (7) 10-mg tablets, (7) 20-mg tablets, and (16) 40-mg tablets. Vilazodone should be taken with food. The recommended dose for vilazodone is 40 mg daily and should be titrated over 2 weeks, starting with 10 mg/day for 7 days, 20 mg/day for the next 7 days, and then depending on the response to 40 mg/day. No dose adjustments are recommended for mild or moderate hepatic or renal impairment.

Geriatric Patients. No dose adjustment is recommended for use in geriatric patients. This is based on results from a study demonstrating similar pharmacokinetics in patients >65 years old and those ages 24 to 55 years. Nonetheless, the product labeling cautions that a greater sensitivity of some older individuals cannot be ruled out and that older adult patient may be at greater risk for hyponatremia.

Pediatric Patients. Vilazodone is not FDA approved for use in pediatric patients, and two studies failed to show an antidepressant effect relative to placebo in adolescents 12 to 17 years of age with MDD.

VORTIOXETINE

Introduction and History

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic domain	Serotonin
Mechanism of action	Reuptake inhibitor (SERT), receptor partial agonist (5-HT_{1A}), receptor antagonist (5-HT_3)

Vortioxetine (TRINTELLIX) was developed by Lundbeck with the aim of adding 5-HT_{1A} and 5-HT_3 receptor affinities to a 5-HT reuptake inhibitor. It is also a partial agonist at 5-HT_{1B} receptors and an antagonist at 5-HT_{1D} and 5-HT_7 receptors. Vortioxetine was initially denoted LU AA 21004, and its first phase II study was published in 2011. The clinical development program, which led to its FDA approval for the treatment of major depressive disorder on September 30, 2013, initially comprised eleven placebo-controlled studies, of which five were active-referenced. Vortioxetine was initially commercialized under the brand name Brintellix but subsequently changed to the brand name Trintellix in May 2016 to avoid confusion with the anticoagulant Brinlinta.

Chemistry

The chemical structure of vortioxetine, which is a piperazine derivative, is 1-[2-(2,4-dimethylphenyl-sulfanyl)-phenyl]-piperazine (Fig. 33.5c–2). It has a molecular weight of 298.4 g/mole.

Pharmacologic Actions

Pharmacokinetics. The pharmacologic activity of vortioxetine comes from the parent compound, its major carboxylic acid metabolite being pharmacologically inactive. At doses studied of 2.5 to 60 mg/day, the pharmacokinetics were linear and time-independent. It is well absorbed from the gastrointestinal tract with an absolute bioavailability of 75%, with no difference in fed or fasting conditions. Peak plasma levels are reached after 7 to 11 hours. After multiple doses of 20 mg/day, C_{max} reached 19.21 ng/mL, and AUC over 24 hours reached 361 ng-hours/mL. There is an extensive extravascular distribution and a high protein binding of 98% to 99%. The terminal half-life is of 66 hours, with a steady state being reached after 2 weeks. The drug is catabolized to its major inactive carboxylic acid metabolite mainly through the CYP 2D6 isoenzyme.

Plasma levels are about doubled when administered to poor 2D6 metabolizers or in the presence of 2D6 inhibitors, and doses should be halved in these situations. Plasma levels might be higher when administered with strong 3A4/2C9 inhibitors, and caution is advised in these situations. Vortioxetine does not inhibit or induce any major CYP isozymes, nor is it a substrate of the permeability glycoproteins. Two-thirds of the metabolites are excreted in the urine and one-third in the feces. Only very small amounts of the parent compound are found in the urine. No dose adjustment is necessary for renal or hepatic impairment, although caution is advised, particularly with severe impairment.

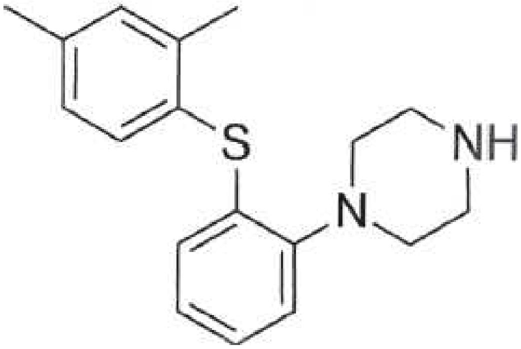

FIGURE 33.5c–2. Chemical structure of vortioxetine.

Mechanisms of Action. Although its exact mechanism of action is not yet fully understood, vortioxetine has been demonstrated to selectively target the 5-HT system, with multimodal 5-HT modulating properties. It was shown to produce the same adaptive changes as vilazodone on the 5-HT system in the rat brain, which led to a gradual increase in 5-HT_{1A} transmission with its sustained administration.

As with SSRI medications, vortioxetine binds to the 5-HT transporter with a high affinity (Ki = 1.6 nM). It also has high and clinically significant affinities for other 5-HT receptors, namely it is a 5-HT_3 antagonist (Ki = 3.7 nM), a full 5-HT_{1A} agonist (Ki = 15 nM), a 5-HT_7 antagonist (Ki = 18 nM), a 5-HT_{1B} partial agonist (Ki = 33 nM), and a 5-HT_{1D} antagonist (Ki = 54 nM).

In positron emission tomography (PET) studies, the mean 5-HT transporter occupancy reached 80% only at 20 mg/day, with an occupancy of 50% and 65% at doses of 5 and 10 mg/day, respectively. However, it is thought that the lower 5-HT transporter occupancy seen with the 5- and 10-mg doses still provides clinically significant effects through synergistic effects with the other 5-HT receptors vortioxetine engages.

Vortioxetine has no clinically significant affinity at the NE transporter (Ki = 113 nM) or at the DA transporter (Ki = >1,000 nM).

Therapeutic Indications

To this day, the only approved indication for vortioxetine is for the treatment of MDD in adults. As is the case for vilazodone, the 5-HT transporter blocker and the 5-HT_{1A} agonist properties of vortioxetine provide a rationale for its use in various anxiety, stress-related, OCD, eating and premenstrual dysphoric disorders. However, the dataset so far is not sufficient to support an indication for these other disorders.

Clinical Drug Studies

Clinical Drug Studies in Major Depressive Disorder. Although the initial dataset which led to the FDA approval of vortioxetine in 2013 included 11 placebo-controlled studies, the complete published dataset now includes 13 placebo-controlled studies which used a depressive symptom rating scale as the primary endpoint. Twelve of these studies were conducted in adult patients, and one study by Katona et al. included only older adult patients (≥65 years old) with a mean age of 71. In these various studies, doses used ranged between 1 and 20 mg/day for a duration of 6 to 8 weeks. All used the MADRS for inclusion in the study, with minimal scores ranging from ≥22 to ≥30, although most studies used the cut-off of ≥26. Most studies used the MADRS as a primary endpoint, and a few studies used the HAMD-24 items. Of these 13 studies, 7 were positive for the primary endpoint, and 6 were failed studies. It is noteworthy that until recently, the dataset included fourteen studies, but one study by Wang et al. 2019 was retracted by the authors, and the editors did not get an answer from the investigators about the source of vortioxetine since it had not been licensed for use in China at the time the study was conducted (Table 33.5c–2).

Overall, there are nine randomized studies that compared vortioxetine directly to another medication with antidepressant properties, using a change in depressive symptoms as the primary endpoint. Of the above 13 placebo-controlled studies, 6 included an active comparator, 5 with duloxetine (60 mg/day) and 1 with venlafaxine XR (225 mg/day). Of these six studies, the venlafaxine XR and three of the duloxetine-referenced studies were positive, whereas one duloxetine-referenced study was negative, and another was a failed study. The remaining three randomized studies were not placebo-controlled, thus comparing vortioxetine head-to-head with another medication. The first study examined the effectiveness of vortioxetine, with a flexible dose of 10 to 20 mg, with agomelatine, a melatonin receptor agonist and 5-HT$_{2B/2C}$ antagonist, in patients that had not responded to prior treatment with an SSRI or a serotonin and norepinephrine reuptake inhibitor (SNRI). Vortioxetine was found superior to agomelatine on the primary outcome (MADRS). In the second and third studies, vortioxetine 10 mg and 15 mg were found to be noninferior to venlafaxine XR 150 mg for 8 weeks in adults and sertraline 75 mg for 6 weeks in older adult patients, respectively.

Table 33.5c–2. Main Short-Term Randomized Controlled Trials of Vortioxetine

Study/Registration #	Key Findings on Primary Endpoint (Scale)	Doses (#Patients Randomized)	Comparator	Duration	Severity
Alvarez et al., 2012 NCT00839423	Positive (MADRS)	P (105); 5 mg (109); 10 mg (101)	P + venlafaxine 225 mg	6 weeks	MADRS > 30
Baldwin et al., 2012 NCT00635219	Failed; duloxetine did not separate (MADRS)	P (145); 2.5 mg (155); 5 mg (155); 10 mg (151)	P + duloxetine 60 mg	8 weeks	MADRS > 26
Henigsberg et al., 2012 NCT00735709	Positive (HAMD-24)	P (140); 1 mg (140); 5 mg (140); 10 mg (140)	P	8 weeks	MADRS > 26
Katona et al., 2012 NCT00811252	Positive (HAMD-24)	P (145); 5 mg (156)	P + duloxetine 60 mg	8 weeks	MADRS > 26
Jain et al., 2013 NCT00672958	Failed (HAMD-24)	P (300); 5 mg (300)	P	6 weeks	MADRS > 30
Mahableshwarkar et al., 2013 NCT00672620	Failed (HAMD-24)	P (153); 2.5 mg (153); 5 mg (153)	P + duloxetine 60 mg	8 weeks	MADRS > 22
Boulenger et al., 2014 NCT01140906	Positive (MADRS)	P (158); 15 mg (152); 20 mg (151)	P + duloxetine 60 mg	8 weeks	MADRS >26
McIntyre et al., 2014 NCT01422213	Positive (DSST 1° endpoint) Positive (MADRS 2° endpoint)	P (196); 10 mg (195); 20 mg (207)	P	8 weeks	MADRS > 26
Montgomery et al., 2014 NCT01488071	Positive (MADRS)	10–20 mg flexible dose (252)	Agomelatine 25–50 mg	8-12 weeks	MADRS > 22
Jacobsen et al., 2015 NCT01163266	Positive (20 mg only) (MADRS)	P (157); 10 mg (155); 20 mg (150)	P	8 weeks	MADRS > 26
Mahableshwarkar et al., 2015 NCT01153009	Positive (20 mg only) (MADRS)	P (161); 15 mg (147); 20 mg (154)	P +duloxetine	8 weeks	MADRS > 26
Mahableshwarkar et al., 2015 NCT01179516	Failed (MADRS)	P (160); 10 mg (157); 15 mg (152)	P	8 weeks	MADRS > 26
Mahableshwarkar et al., 2015 NCT01571453	Positive (DSST 1° endpoint) Positive (MADRS 2° endpoint)	P (194); 10–20 mg (198)	P +duloxetine	8 weeks	MADRS > 26
Wang et al., 2015	Noninferior (MADRS)	Venlafaxine (226); 10 mg (213)	Venlafaxine	8 weeks	MADRS > 26
Inoue et al., 2018	Failed (MADRS)	P (124); 5 mg (119); 10 mg (123)	P	8 weeks 52 weeks extension	MADRS > 26
Nishimura et al., 2018	Failed (MADRS)	P (152); 5 mg (144); 10 mg (150); 20 mg (154)	P	8 weeks	MADRS > 26
Borhannejad et al., 2020	Noninferior (HAMD)	Sertraline (30); 15 mg (30)	Sertraline	6 weeks	HAMD > 19
Inoue et al., 2020	Positive (MADRS)	P (164); 10 mg (165); 20 mg (164)	P	8 weeks	MADRS > 26

This table lists the main short-term randomized controlled trials of vortioxetine, which used a depression rating scale or the Digit Symbol Substitution Task (DSST) as its primary endpoint.
DSST, Digit Symbol Substitution Task; HAMD-24, Hamilton rating scale for Depression; MADRS, Montgomery-Åsberg Depression Rating Scale; P, placebo.

A Cochrane systematic review and meta-analysis by Koesters et al. concluded in favor of vortioxetine over placebo for response (RR 1.35, 95% CI 1.22–1.49; 14 studies, 6,220 participants), remission (RR 1.32, 95% CI 1.15–1.53; 14 studies, 6,220 participants) and depressive symptoms using MADRS (MD –2.94, 95% CI –4.07 to –1.80; 14 studies, 5,566 participants).

A meta-analysis by Baldwin et al. examined the effects of vortioxetine on anxiety levels in patients with MDD and high levels of anxiety. The analysis included data from 10 randomized, placebo-controlled trials, with a total of 1,497 vortioxetine-treated patients (5, 10, and 20 mg) and 860 placebo-treated patients who had a HAM-A baseline of ≥20. There was a significant effect of vortioxetine on MADRS as well as HAM-A total and psychic item subscales for all doses.

There are now two placebo-controlled relapse prevention studies, which are both positive. In the first study, remitted patients, 18 to 75 years old, were randomized to vortioxetine fixed dose (5 or 10 mg, *n* = 204) or placebo (*n* = 192) after an initial 12-week open-label treatment period (*n* = 639) of patients with a major depressive episode of ≥4 weeks duration and initial MADRS ≥26. The primary analysis of time to relapse favored vortioxetine, with a hazard ratio (HR) of 2.01 (95% confidence interval [CI] 1.26–3.21; *p* = .0035). In the second study, remitted patients, 18 to 75 years old, were randomized to vortioxetine fixed dose (5, 10 or 20 mg, *n* = 580) or placebo (*n* = 151) after an initial 16-week open-label treatment period (fixed dose 10 mg/day, *n* = 1,106) of patients with an MDE of 8 to 18 weeks duration and initial MADRS ≥26 and ≥2 prior depressive episodes. Analysis of the primary endpoint of time to relapse and cumulative relapse rates favored vortioxetine and were statistically significant at all doses studied: 32.5% in the placebo group versus 19.3% for vortioxetine 5 mg (HR 0.517; 95% CI 0.323–0.828; *p* = .006 vs. placebo), 17.9% for vortioxetine 10 mg (HR 0.476; 95% CI 0.296–0.767; *p* = .003 vs. placebo), and 17.4% for vortioxetine 20 mg (HR 0.483; 95% CI 0.298–0.782; *p* = .002 vs. placebo). There are also five open-label studies of 52 weeks that have assessed the maintained effectiveness of vortioxetine 5 to 10 mg (*n* = 1,313) and 15 to 20 mg (*n* = 1,144).

Clinical Drug Studies on Cognitive Function. The pharmacologic profile of vortioxetine led to the hypothesis that it could have beneficial effects on cognitive deficits seen in MDD. Thus, many studies of the effects of vortioxetine in MDD included change on various scales of cognitive function either as a primary or secondary endpoint. The most used scale is the Digit Symbol Substitution Task (DSST), which assesses attention, speed of processing and executive function. Other frequently used scales include the Ray Auditory Verbal Learning Test (RAVLT) and the University of California San Diego Performance-Based Skill Assessment (UPSA).

The initial findings on a battery of cognitive tests from three short-term placebo-controlled studies published between 2012 and 2015 led the European Medicine Agency and the FDA to add a cognition claim for patients with MDD to the label for vortioxetine. Of these three studies, two were active referenced with duloxetine 60 mg/day and used the change in the DSST as a primary endpoint. Vortioxetine, but not duloxetine, had a significantly greater effect than placebo on the DSST in all three studies, and the effect was shown through path analysis to be mostly a direct effect. Two meta-analyses (a network meta-analysis and a random effect meta-analysis) also confirmed the positive effect of vortioxetine on cognitive dysfunction versus placebo on the DSST.

Many other studies published since 2016 have assessed the effects of vortioxetine on the DSST or other cognitive, social or workplace functioning scales as a primary or secondary endpoint or through posthoc analysis. Study designs have varied, some studies having a small sample size and lacking power and some studies having a specific focus. Overall, results have been mixed.

Clinical Drug Studies of Other Disorders. Despite the rationale for using vortioxetine in the treatment of many other disorders for which SSRIs or buspirone have shown efficacy, very few studies have investigated this potential.

For the treatment of GAD, there are four short-term placebo-controlled studies, of which only one is positive and one large relapse prevention study which is positive. In meta-analyses of these studies, vortioxetine does not separate from placebo for doses studied of 2.5, 5, or 10 mg. As we know that anxiety disorders generally respond to higher doses of SSRIs and that the inhibition of vortioxetine of the 5-HT transporter is quite low at doses below 20 mg, it would be interesting to study the effects of vortioxetine in GAD at doses of 20 mg.

There is one positive open-label study for the treatment of panic disorder, one positive placebo-controlled RCT for improvement of negative symptoms in schizophrenia patients treated with risperidone 4 to 6 mg daily, and one positive double-blind RCT that showed an improvement in quality of life for patients with irritable bowel syndrome. Other studies for the treatment of social anxiety disorder, PTSD, binge-eating disorder, and attention-deficit disorder were all negative (one study on each).

Precautions and Adverse Reactions

Safety and Tolerability. The safety and tolerability of vortioxetine, as reported in the product monograph, are derived from data collected in patients with MDD treated in short-term clinical trials, with more than 6,700 patients, of whom 3,460 were treated with vortioxetine (5 to 20 mg/day). Also, an article published by Baldwin et al. analyzed the safety and tolerability data from 11 short-term MDD studies, 5 long-term MDD studies and 4 short-term GAD studies, for a total of 3,018 patients treated with vortioxetine in short-term studies and 2,457 patients treated with vortioxetine in long-term studies.

Overall, vortioxetine showed a very favorable profile. Across all studies, the most common treatment-emergent side effect that led to treatment discontinuation was nausea, occurring in about one-fourth to one-third of patients. It led to treatment discontinuation, in a dose-dependent manner, for about 4.5% to 7.8% of patients, reaching the same level as duloxetine 60 mg/day, compared to about 3.5% for placebo. It is noteworthy that with vortioxetine, many of the treatment-emergent adverse events (TEAEs) commonly found with SSRIs were seen at “placebo levels” with no dose effect: headache, dry mouth, dizziness, constipation, insomnia, somnolence, fatigue, sexual dysfunction, and hyperhidrosis.

Given the long terminal half-life of vortioxetine, if a side effect is problematic, it can be withheld for 3 days and restarted at half the dose.

Serious Adverse Events. Although five deaths have been reported in the short-term studies of vortioxetine, none seem to have been related or attributable to vortioxetine. Four of these deaths occurred in the short-term MDD trials. One 74-year-old woman on a vortioxetine 5 mg arm had a known history of cholelithiasis and died from gallbladder cancer 1 month after withdrawal from the study. One 63-year-old male receiving vortioxetine 2.5 mg died from falling from a fourth-floor balcony, an event not judged by the investigator to represent a suicidal act. One 46-year-old man, receiving a flexible dose of vortioxetine 5 to 10 mg, with a history of type II diabetes, died of a pancreatic carcinoma 8 months after withdrawal from the study. One 56-year-old man receiving vortioxetine 5 mg in an open-label long-term extension study died in a traffic collision caused by another driver. The fifth death occurred in a short-term GAD study: one 49-year-old woman randomized to vortioxetine 5 mg/day died from a morphine intoxication 3 days after her baseline visit, but no vortioxetine was found in her stomach or body fluids on autopsy.

Overdose. There are limited data in clinical and postmarketing reports on the effect of an overdose of vortioxetine. With accidental or intentional supratherapeutic doses of up to 40 mg/day for 4 days, adverse effects included increased incidence of nausea, postural dizziness, diarrhea, abdominal discomfort, generalized pruritus, somnolence, and flushing. With doses up to 80 mg, most patients reported absent or mild symptoms, most frequently nausea and vomiting. In doses above 80 mg, seizures and serotonin syndrome have been reported.

Sexual Dysfunction. Rates of sexual dysfunction in clinical trials are often underestimated when based on self-reporting. The use of standardized scales such as the ASEX and the CSFQ allow a much better estimation of the impact of various drugs on sexual function. Using the ASEX in seven short-term MDD studies, vortioxetine in doses of 5 to 10 mg/day had sexual adverse events at rates similar to placebo, whereas with doses of 20 mg/day, the incidence reached that of duloxetine 60 mg/day. With overall pooling of data from eleven short-term RCTs, sexual dysfunction was estimated at 1.6% to 1.8% for vortioxetine 5 to 20 mg versus placebo 1.0%, venlafaxine XR 225 mg 12.4%, and duloxetine 60 mg 4.5%.

A study by Jacobsen et al. examined the effects of switching responders to paroxetine, sertraline, or citalopram who were experiencing sexual dysfunction to either vortioxetine or citalopram (10 or 20 mg/day, with three-quarters of vortioxetine patients taking 20 mg/day). This study used the CSFQ short form and showed that vortioxetine had better rates than escitalopram on four of five dimensions and all three phases of sexual functioning.

Since sexual dysfunction with SSRIs and SNRIs is related to inhibition of the 5-HT transporter, these different findings are consistent with the lower occupancy of vortioxetine of the 5-HT transporter at doses inferior to 20 mg/day, with occupancy reaching 80% only at 20 mg/day.

Weight. No clinically relevant changes in body weight were found in short-term or long-term studies of vortioxetine. In long-term placebo-controlled studies, mean changes in weight were +0.4 kg for vortioxetine 5 to 10 mg, versus +0.1 kg for placebo. The proportion of patients with >7% weight gain was 6.5% for 5 to 10 mg, versus 5.8% for placebo. Proportion of patients with >7% weight loss was 3.5% for vortioxetine 5 to 10 mg, versus 2.6% for placebo.

Cardiac. In studies of vortioxetine, no clinically significant changes in cardiac parameters were found, including QT, QTc, PR and QRS intervals, blood pressure or heart rate, and no arrhythmogenic potential was found. In a thorough study of QTc in 340 healthy subjects, QTc was increased by 4.0 ms at 10 mg/day and 4.6 ms at 40 mg/day. A minor negative chronotropic effect was found of –4.7 beats per minute (BPM) for 10 mg/day and –5.4 BPM for 40 mg/day.

Laboratory Values. Studies of vortioxetine found no clinically significant changes versus placebo in clinical, biochemical values or hematologic parameters.

Platelets. Vortioxetine inhibits the 5-HT transporter and is expected to have similar effects on platelets, as seen with SSRIs or SNRIs. Caution is advised when administering vortioxetine with other drugs that potentiate the risk of bleeding, or with patients that have a history or risk factors of bleeding. No cases of bleeding have been reported in studies of vortioxetine.

Hyponatremia. Although no cases of hyponatremia have been reported in the clinical studies of vortioxetine, medications with serotonergic properties such as SSRIs and SNRIs have been shown to increase the risk of hyponatremia related to the syndrome of inappropriate antidiuretic hormone secretion. Thus, caution is advised when treating patients at higher risk of hyponatremia with vortioxetine. If symptomatic hyponatremia develops, vortioxetine should be discontinued and appropriate medical intervention implemented.

Gastrointestinal. When vortioxetine was developed, it was thought that the 5-HT_3 antagonism properties would eliminate nausea commonly seen with SSRI and SNRI medications. However, nausea has turned out to be the most reported TEAE in clinical trials, with rates between 20.9% and 31.2%, leading to treatment discontinuation in 1% to 4% of patients in a dose-dependent manner. This could be attributable to the 5-HT_{1A} agonist properties, as is seen with the 5-HT_{1A} agonist buspirone. Vomiting was also seen to a lesser extent, with rates of 2.9% to 6.5%.

Suicidality. Vortioxetine carries the same black box warning as other SSRIs regarding suicidal thoughts and behaviors in children, adolescents, and young adults, based on data showing an increased risk of suicidal thinking and behavior (although no completed suicides were observed) in children, adolescents, and young adults (ages 18 to 24) with MDD and other psychiatric disorders treated with SSRIs or other newer antidepressants. No completed suicides were reported in trials of vortioxetine. In short-term placebo-controlled trials of vortioxetine, suicide-related events (such as intentional overdose, self-injurious behavior, and suicide attempts) were reported by 0.4% of vortioxetine-treated patients versus 0.3% of placebo-treated patients, newly emergent suicidal ideation was reported by 2.7% of vortioxetine-treated patients versus 4.1% of placebo-treated patients, and treatment-emergent suicidal ideation in the 18- to 24-year-old group was reported by 15% of vortioxetine-treated patients versus 13.2% of placebo-treated patients.

Activation of Mania/Hypomania. In phase II/III clinical trials of vortioxetine, only 1 out of 3,904 (<0.1%) patients treated with vortioxetine-reported mania/hypomania. However, caution should always be used when initiating any treatment with antidepressant properties as bipolar disorder most often presents with an initial major depressive episode, and bipolar patients are at increased risk of developing mania or hypomania when treated with a medication indicated for MDD.

Discontinuation Syndrome. In studies of vortioxetine, no discontinuation symptoms were reported, which is consistent with its long terminal half-life, as complete elimination takes about 2 weeks. More specifically, three studies used the Discontinuation Emergent Signs and Symptoms (DESS) scale to specifically assess potential discontinuation symptoms when vortioxetine was abruptly stopped at the end of 8 weeks of treatment. No clinically significant differences were seen in this scale for patients who had been taking 10, 15, or 20 mg of vortioxetine at weeks 9 and 10 (thus after 1 and 2 weeks of abrupt cessation).

Serotonin Syndrome. Serotonin syndrome has not been reported in the short-term or long-term clinical trials of vortioxetine. However, one must always consider the risk of 5-HT syndrome when using medications that increase 5-HT availability. This risk is extremely low when the medication is taken alone in therapeutic doses. However, the risk increases in overdose or when co-administration occurs with illicit drugs (such as ecstasy and concentrated tetrahydrocannabinol) or other medications that increase 5-HT availability, such as MAOIs, triptans, opioids, and St. John wort. When serotonin syndrome is diagnosed, all medications increasing serotonin should be immediately stopped and supportive measures implemented.

Nervous System. According to the product label, seizures were reported in <0.1% of patients receiving vortioxetine versus 0% in patients receiving placebo in short-term clinical trials. Caution is advised in patients with epilepsy or a history of seizures, and treatment should be stopped if there is an increased frequency of seizures in these patients.

No significant driving or cognitive impairments relative to placebo were noted in a study of 21 healthy individuals who were administered a battery of neuropsychological tests the day after taking a single dose of 10 mg in the evening.

Sleep. No significant difference was seen between vortioxetine or placebo for insomnia or daytime somnolence. More specifically, insomnia occurred in 4.0% of patients taking placebo versus 2.0% to 5.1%, 15.9%, and 8.1% of patients taking vortioxetine, venlafaxine and duloxetine, respectively, in short-term MDD studies.

Reproductive. There are no published data on vortioxetine during pregnancy or during labor and delivery. Vortioxetine was not teratogenic in rats or rabbits for doses up to 160 and 60 mg/day, although in rats exposed to about six times the Cmax of the MRHD (20 mg/day), lower fetal weight and delayed ossification were observed.

The FDA package insert carries the same warning regarding potential neonatal toxicity as is listed with other medications used to treat MDD. Clinicians must consider the well-documented risks of untreated depression in pregnancy, including during the post-partum period, versus the associated risks of taking medications with serotonergic agents (see chapter on SSRI for more considerations).

Lactation. There are no human published data for breastfeeding while taking vortioxetine. Animal data have shown vortioxetine and its metabolites to be excreted in milk. As this is also expected to be the case in humans, breastfeeding is not recommended when taking vortioxetine. However, as with all medications in pregnancy, the decision should consider the risks and benefits.

Fertility. There are no data on the effect of vortioxetine on fertility in humans. In male and female rats, vortioxetine did not show any effect on fertility, with plasma levels of 24 and 17 times the MRHD.

Drug Interactions

As vortioxetine is metabolized primarily through CYP 2D6, blood levels will increase in 2D6 slow metabolizers and when co-administered with moderate (such as bupropion) or strong (such as fluoxetine) 2D6 inhibitors and decrease in the presence of 2D6 inducers. However, vortioxetine does not inhibit or induce any CYP isozymes and thus is not expected to alter other drug plasma levels through CYP interactions. As with all medications that inhibit 5-HT transporters, MAOIs should be stopped for a minimum of 2 weeks before starting vortioxetine, and vortioxetine should be completely eliminated (2 weeks) before initiating an MAOI or moclobemide. Efficacy of triptans might be reduced when co-administered with vortioxetine, as their therapeutic effects are mediated through activation of 5-HT_{1D} receptors, and vortioxetine is an antagonist of 5-HT_{1D} receptors. However, since the affinity of vortioxetine for the 5-HT_{1D} receptors is quite low (Ki = 54 nM), this pharmacokinetic interaction might not be significant at lower doses of vortioxetine.

Dosage and Administration

Based on the clinical development program, vortioxetine is recommended for the treatment of MDD at doses between 5 and 20 mg/day, usually given morning for convenience, although it can also be given at night. Considering the time to reach a steady state, each dose increase should be done at intervals of ≥2 weeks. For adults 18 to 65 years old, the starting dose is usually 10 mg. For older adult patients, poor 2D6 metabolizers or when co-administered with a 2D6 inhibitor, the starting dose should be 5 mg/day.

Switching Strategies to Vortioxetine. Switching from other medications to vortioxetine has been examined in three studies and seems well tolerated. However, some problems could occur in clinical practice if the other medication is stopped too abruptly since vortioxetine reaches steady state after only about 2 weeks, and it has low 5-HT transporter occupancy at its lower-dose range. Thus, when switching from an SSRI or SNRI medication with a short half-life, discontinuation symptoms could occur in the first weeks of treatment and be misinterpreted as side effects of vortioxetine, thus leading to premature discontinuation of vortioxetine. To prevent this, it is recommended to implement a gradual cross-over which considers the long half-life of vortioxetine.

Geriatric Patients. One trial by Katona et al. studied vortioxetine for the treatment of MDD in older adult patients, and a few other studies included patients aged 18 to 75, thus recruiting some patients >65 years old. So far, these studies have not shown any major safety issues in older adult patients treated with vortioxetine, and as with adult populations, nausea was the only TEAE that occurred with an incidence of ≥5% and >2× the placebo rate (21.8% vs. 8.3%).

Pediatric Patients. Vortioxetine is not recommended in pediatric patients, as its efficacy has not yet been demonstrated in adolescents 12 to 17 years, and its efficacy and safety have not yet been demonstrated in children 7 to 11 years.

ACKNOWLEDGMENTS

This chapter represents an update and revision of previous chapters in this series, most recently by Dr. Lourdes Illa-Sanchez from the last edition. We are indebted to Dr. Illa-Sanchez for her work and retain much of her prose in this current edition.

Further Readings

Alvarez E, Perez V, Dragheim M, et al. A double-blind, randomized, placebo-controlled, active reference study of Lu AA21004 in patients with major depressive disorder. *Int J Neuropsychopharmacol*. 2012;15:589–600.

Areberg J, Luntang-Jensen M, Sogaard B, Nilausen DO. Occupancy of the serotonin transporter after administration of Lu AA21004 and its relation to plasma concentration in healthy subjects. *Basic Clin Pharmacol Toxicol*. 2012;110:401–404.

Baldwin DS, Chrones L, Florea I, et al. The safety and tolerability of vortioxetine: Analysis of data from randomized placebo-controlled trials and open-label extension studies. *J Psychopharmacol*. 2016;30:242–252.

Baldwin DS, Florea I, Jacobsen PL, Zhong W, Nomikos GG. A meta-analysis of the efficacy of vortioxetine in patients with major depressive disorder (MDD) and high levels of anxiety symptoms. *J Affect Disord*. 2016;206:140–150.

Borhannejad F, Shariati B, Naderi S, et al. Comparison of vortioxetine and sertraline for treatment of major depressive disorder in elderly patients: A double-blind randomized trial. *J Clin Pharm Ther*. 2020;45:804–811.

Boulenger JP, Loft H, Olse CK. Efficacy and safety of vortioxetine (Lu AA21004), 15 and 20 mg/day: a randomized, double-blind, placebo-controlled, duloxetine-referenced study in the acute treatment of adult patients with major depressive disorder. *Int Clin Psychopharmacol*. 2014;29:138–149.

Clayton AH, Durgham S, Li D, et al. Effects of vilazodone on sexual functioning in healthy adults: results from a randomized, double-blind, placebo-controlled, and active-controlled study. *Int Clin Psychopharmacol*. 2017;32:27–35.

Clayton AH, Gommoll C, Chen D, Nunez R, Mathews M. Sexual dysfunction during treatment of major depressive disorder with vilazodone, citalopram, or placebo: results from phase IV clinical trial. Int Clin Psychopharmacol and antidepressant treatment. *Expert Opin Drug Saf*. 2014;13:1361–1374.

Croft HA, Pomara N, Gommoll C, et al. Efficacy and safety of vilazodone in major depressive disorder: a randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry*. 2014;75:e1291–e1298.

Durgam S, Gommoll C, Forero G, et al. Efficacy and safety of vilazodone in patients with generalized anxiety disorder: a randomized, double-blind, placebo-controlled, flexible-dose trial. *J Clin Psychiatry*. 2016;77:1687–1694.

Durgam S, Gommoll C, Migliore R, et al. Relapse prevention in adults with major depressive disorder treated with vilazodone: a randomized, double-blind, placebo-controlled trial. *Int Clin Psychopharmacol*. 2018;33:304–311.

El Mansari M, Crnic A, Oosterhof C, Blier P. Long-term administration of the antidepressant vilazodone modulates rat brain monoaminergic systems. *Neuropharmacology*. 2015;99:696–704.

El Mansari M, Lecours M, Blier P. Effects of acute and sustained administration of vortioxetine on the serotonin system in the hippocampus: electrophysiological studies in the rat brain. *Psychopharmacology (Berl)* 2015;232:2343–2352.

Gommoll C, Durgam S, Mathews M, et al. A double-blind, randomized, placebo-controlled, fixed-dose phase III study of vilazodone in patients with generalized anxiety disorder. *Depress Anxiety*. 2015;32:451–459.

Gommoll C, Forero G, Mathews M, et al. Vilazodone in patients with generalized anxiety disorder: a double-blind, randomized, placebo-controlled, flexible-dose study. *Int Clin Psychopharmacol*. 2015;30:297–306.

Guay DR. Vilazodone hydrochloride, a combined SSRI and 5-HT1A receptor agonist for major depressive disorder. *Consult Pharm*. 2012;27:857–867.

Henigsberg N, Mahableshwarkar AR, Jacobsen P, Chen Y, Thase ME. A randomized, double-blind, placebo-controlled 8-week trial of the efficacy and tolerability of multiple doses of Lu AA21004 in adults with major depressive disorder. *J Clin Psychiatry*. 2012;73:953–959.

Inoue T, Nishimura A, Sasai K, Kitagawa T. Randomized, 8-week, double-blind, placebo-controlled trial of vortioxetine in Japanese adults with major depressive disorder, followed by a 52-week open-label extension trial. *Psychiatry Clin Neurosci*. 2018;72:103–115.

Inoue T, Sasai K, Kitagawa T, Nishimura A, Inada I. Randomized, double-blind, placebo-controlled study to assess the efficacy and safety of vortioxetine in Japanese patients with major depressive disorder. *Psychiatry Clin Neurosci*. 2020;74:140–148.

Iovieno N, Papakostas GI, Feeney A, et al. Vortioxetine versus placebo for major depressive disorder: a comprehensive analysis of the Clinical Trial Dataset. *J Clin Psychiatry*. 2021;82:20r13682.

Jacobsen PL, Mahableshwarkar AR, Chen Y, Chrones L, Clayton AH. Effect of vortioxetine vs. Escitalopram on sexual functioning in adults with well-treated major depressive disorder experiencing SSRI-induced sexual dysfunction. *J Sex Med*. 2015;12: 2036–2048.

Jain R, Mahableshwarkar AR, Jacobsen PL, Chen Y, Thase ME. A randomized, double-blind, placebo-controlled 6-wk trial of the efficacy and tolerability of 5 mg vortioxetine in adults with major depressive disorder. *Int J Neuropsychopharmacol*. 2013;16:313–321.

Katona C, Hansen T, Olsen C. A randomized, double-blind, placebo-controlled, duloxetine-referenced, fixed-dose study comparing the efficacy and safety of Lu AA21004 in elderly patients with major depressive disorder. *Int Clin Psychopharmacol*. 2012;27:215–223.

Khan A, Cutler AJ, Kajdasz DK, et al. A randomized, double-blind, placebo-controlled, 8-week study of vilazodone, a serotonergic agent for the treatment of major depressive disorder. *J Clin Psychiatry*. 2011;72:441–447.

Koesters M, Ostuzzi G, Guaiana G, Breilmann J, Barbui C. Vortioxetine for depression in adults. *Cochrane Database Syst Rev*. 2017;7:CD011520.

Mahableshwarkar AR, Zajecka J, Jacobson W, Chen Y, Keefe RS. A randomized, placebo-controlled, active-reference, double-blind, flexible-dose study of the efficacy of vortioxetine on cognitive function in major depressive disorder. *Neuropsychopharmacology*. 2015;40:2025–2037.

Mathews M, Gommol C, Chen D, Nunez R, Khan A. Efficacy and safety of vilazodone 20 and 40 mg in major depressive disorder: a randomized, double-blind, placebo-controlled trial. *Int Clin Psychopharmacol*. 2015;30:67–74.

McIntyre RS, Lophaven S, Olsen CK. A randomized, double-blind, placebo-controlled study of vortioxetine on cognitive function in depressed adults. *Int J Neuropsychopharmacol*. 2014;17:1557–1567.

Montgomery SA, Nielsen RZ, Poulsen LH, Haggstrom L. A randomized, double-blind study in adults with major depressive disorder with an inadequate response to a single course of selective serotonin reuptake inhibitor or serotonin-noradrenaline reuptake inhibitor treatment switched to vortioxetine or agomelatine. *Hum Psychopharmacol*. 2014;29:470–482.

Qin B, Huang G, Yang Q, et al. Vortioxetine treatment for generalised anxiety disorder: A meta-analysis of anxiety, quality of life and safety outcomes. *BMJ Open*. 2019;9:e033161.

Rickels K, Athanasiou M, Robinson DS, et al. Evidence for efficacy and tolerability of vilazodone in the treatment of major depressive disorder: a randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry*. 2009;70:326–333.

Robinson DS, Kajdasz DK, Gallipoli S, et al. A 1-year, open-label study assessing the safety and tolerability of vilazodone in patients with major depressive disorder. *J Clin Psychopharmacol*. 2011;31:643–646.

Sanchez C, Asin KE, Artigas F. Vortioxetine, a novel antidepressant with multimodal activity: review of preclinical and clinical data. *Pharmacol Ther*. 2015;145:43–57.

Sinha S, Chary S, Thakur P, et al. A phase III prospective active and placebo-controlled randomized trial of vilazodone in the treatment of major depressive disorder. *Cureus*. 2021;13:e16689.

Spina E, Trifirò G, Caraci F. Clinically significant drug interactions with newer antidepressants. *CNS Drugs*. 2012;26:39–67.

Thase ME, Chen D, Edwards J, Ruth A. Efficacy of vilazodone on anxiety symptoms in patients with major depressive disorder. *Int Clin Psychopharmacol*. 2014;29:351–356.

Thase ME, Jacobsen PL, Hanson E, et al. Vortioxetine 5, 10, and 20 mg significantly reduces the risk of relapse compared with placebo in patients with remitted major depressive disorder: The RESET study. *J Affect Disord*. 2022;303:123–130.

Wang J, Liu XF, Feng C, Bao Q, Fu HR. Efficacy and safety of vortioxetine for the treatment of major depressive disorder: a randomised double-blind placebo-controlled study. *Int J Psychiatry Clin Pract*. 2019;23:245–250.

33.5d Monoamine Oxidase Inhibitors

Ethan F. Karsen, M.D., and Paul E. Holtzheimer, M.D., M.S.

HISTORY

Monoamine oxidase inhibitors (MAOIs) were among the first antidepressants found to be clinically effective in humans. Although MAOI use has declined over time due to the emergence of newer antidepressants with equivalent efficacy and more tolerable side effects, research interest in MAOIs has been rekindled by the development of selective MAOIs, new delivery systems, and the discovery of their neuroprotective role in neurodegenerative diseases.

Development of MAOIs

In 1957, at a regional meeting of the American Psychiatric Association (APA) in Syracuse, New York, two independent research groups presented findings showing that the antitubercular medication iproniazid had antidepressant effects in patients with tuberculosis and comorbid depression. These findings represented a culmination of years of laboratory and clinical research. Emil Fischer first isolated hydrazines, the class to which early MAOIs belong, in the 1870s. In the 1920s and 1930s, several research groups helped identify the enzyme monoamine oxidase (MAO) and showed that it is capable of metabolizing adrenaline through oxidation. In 1952, Albert Zeller demonstrated that iproniazid acts as an MAO inhibitor. The same year, physicians at Staten Island's Seaview Hospital observed that iproniazid had a more powerful stimulating effect than isoniazid in the patients they treated for tuberculosis. That first study of MAOI use in patients without tuberculosis was also presented at the APA meeting in 1957. The study showed substantial improvement of symptoms in 70% of a group of psychotically depressed patients taking iproniazid. By 1958, over 400,000 patients were being treated with iproniazid for depression, then marketed as Marsilid. The clinical use of iproniazid, however, was short-lived as reports of jaundice and hepatotoxicity accumulated, and in 1961 it was withdrawn from the U.S. market.

While iproniazid was entering and leaving the market, pharmaceutical companies were developing several new MAOIs. Hydrazine derivatives similar to iproniazid included: isocarboxazid (Marplan), phenelzine (Nardil), nyalamide, mebanazine, and pheniprazine. Like iproniazid, many of these agents showed hepatotoxicity that limited clinical use. In 1959, researchers discovered that tranylcypromine (Parnate), a cyclopropylamine structurally unrelated to the hydrazines, had MAOI action. Soon after its introduction, however, reports appeared of hypertensive crises and severe headaches associated with subarachnoid intracranial hemorrhage. Barry Blackwell, a psychiatric resident at Maudsley Hospital in London, traced these effects back to ingestion of certain cheeses in patients taking MAOIs, thereafter known as the "cheese effect." Further investigation showed that food rich in amines, especially tyramine, could precipitate hypertensive crises in patients treated with MAOIs. Due to the concern for hemorrhage, tranylcypromine was briefly removed from the U.S. market in 1964, only to be reintroduced a few months later after postmarketing analysis showed an incidence rate of less than 0.001%. Thereafter, all MAOIs included labeling on adverse reactions and suggested dietary restrictions.

The first-generation MAOIs are *irreversible* inhibitors of MAO; they covalently bond to and inactivate the enzyme. After the discovery of two isoforms of MAO, MAO-A and MAO-B (discussed in more detail later in this chapter), in the 1970s, researchers sought to synthesize MAOIs that would be reversible and selective inhibitors with fewer adverse effects. The use of platelets helped identify molecules selective for MAO-B, with evidence suggesting a correlation between platelet MAO-B inhibition and antidepressant response to phenelzine. Pargyline was the first FDA-approved selective MAO-B inhibitor, although it did not show significant antidepressant effects in clinical trials and is no longer available. In 1972, researchers described L-deprenyl (Eldepryl), an irreversible but selective MAO-B inhibitor. However, in clinical trials, L-deprenyl demonstrated antidepressant effect only at doses at which it loses its selectivity for MAO-B. This finding suggested that inhibition of MAO-A, alone or in combination with MAO-B inhibition, is necessary for the antidepressant effect. Interestingly, experimental findings from the 1970s found L-deprenyl to be an effective adjunct to L-dopa in the treatment of patients with Parkinson disease (PD). The FDA licensed the drug in 1996 and renamed it selegiline. Rasagiline (Azilect), with theoretically less adverse effects than selegiline, followed as another irreversible MAO-B inhibitor for use in PD.

Table 33.5d–1.
Classification of Monoamine Oxidase Inhibitors

Drug	Reversibility	Selectivity	Indication
Iproniazid[a] (Marsilid)	Irreversible	MAO-A MAO-B	Depression
Isocarboxazid (Marplan)	Irreversible	MAO-A MAO-B	Depression
Phenelzine (Nardil)	Irreversible	MAO-A MAO-B	Depression
Tranylcypromine (Parnate)	Irreversible	MAO-A MAO-B	Depression
Isoniazid (Nydrazid)	Irreversible	MAO-A MAO-B	Antituberculosis
Nialamide[a] (Niamid)	Irreversible	MAO-A MAO-B	Depression
Procarbazine (Matulane)	Weak irreversible	MAO-A MAO-B	Antineoplastic
Clorgyline[b]	Irreversible	MAO-A only	Depression
Selegiline (Deprenyl)	Irreversible	MAO-B only	Depression, Parkinson disease
Rasagiline (Azilect)	Irreversible	MAO-B only	Parkinson disease
Pargyline[a] (Eutonyl)	Irreversible	MAO-B only	Antihypertensive
Linezolid	Reversible	MAO-A MAO-B	Antibiotic
Lazabemide[c] (Pakio)	Reversible	MAO-B only	Parkinson disease
Moclobemide[c] (Aurorix/Manerix)	Reversible	MAO-A only	Depression
Brofaromine[c] (Consonar)	Reversible	MAO-A only	Depression
Befloxatone[c] (Synthelabo)	Reversible	MAO-A only	Depression
Pirindole[c] (Pirazidol)	Reversible	MAO-A only	Depression
Toloxatone[c]	Reversible	MAO-A only	Depression

(Humoryl)

[a] Discontinued.

[b] Used in research only.

[c] Available outside the United States only.

Following the suggested association between MAO-A inhibition and antidepressant effects, researchers in 1979 identified clorgyline, an irreversible but selective MAO-A inhibitor. It proved effective for depression but caused hypertensive crises as it irreversibly bound MAO in the gut. Moclobemide initially developed for hyperlipidemia, became the first reversible and selective MAO-A inhibitor; it is currently available for use in depression in over 50 countries worldwide. Reversible inhibitors of monoamine oxidase type A (RIMA) are less likely to cause hypertensive crises because they are readily displaced by tyramine in the gut. There are no RIMAs currently available in the United States (see Table 33.5d–1 for MAOI drugs available worldwide). The five MAOIs currently available in the United States are isocarboxazid, phenelzine, tranylcypromine, selegiline, and rasagiline (Fig. 33.5d–1).

CHEMISTRY

MAO is a flavoprotein enzyme system located on the outer mitochondrial membrane of both neurons and nonneuronal cells. In neurons, MAO metabolizes three major neurotransmitters: serotonin, dopamine, and norepinephrine. MAO also serves several other important roles in nonneuronal places, including the metabolism of trace amines in the brain, the peripheral inactivation of biogenic and dietary amines, phase I metabolism, and as an imidazoline receptor.

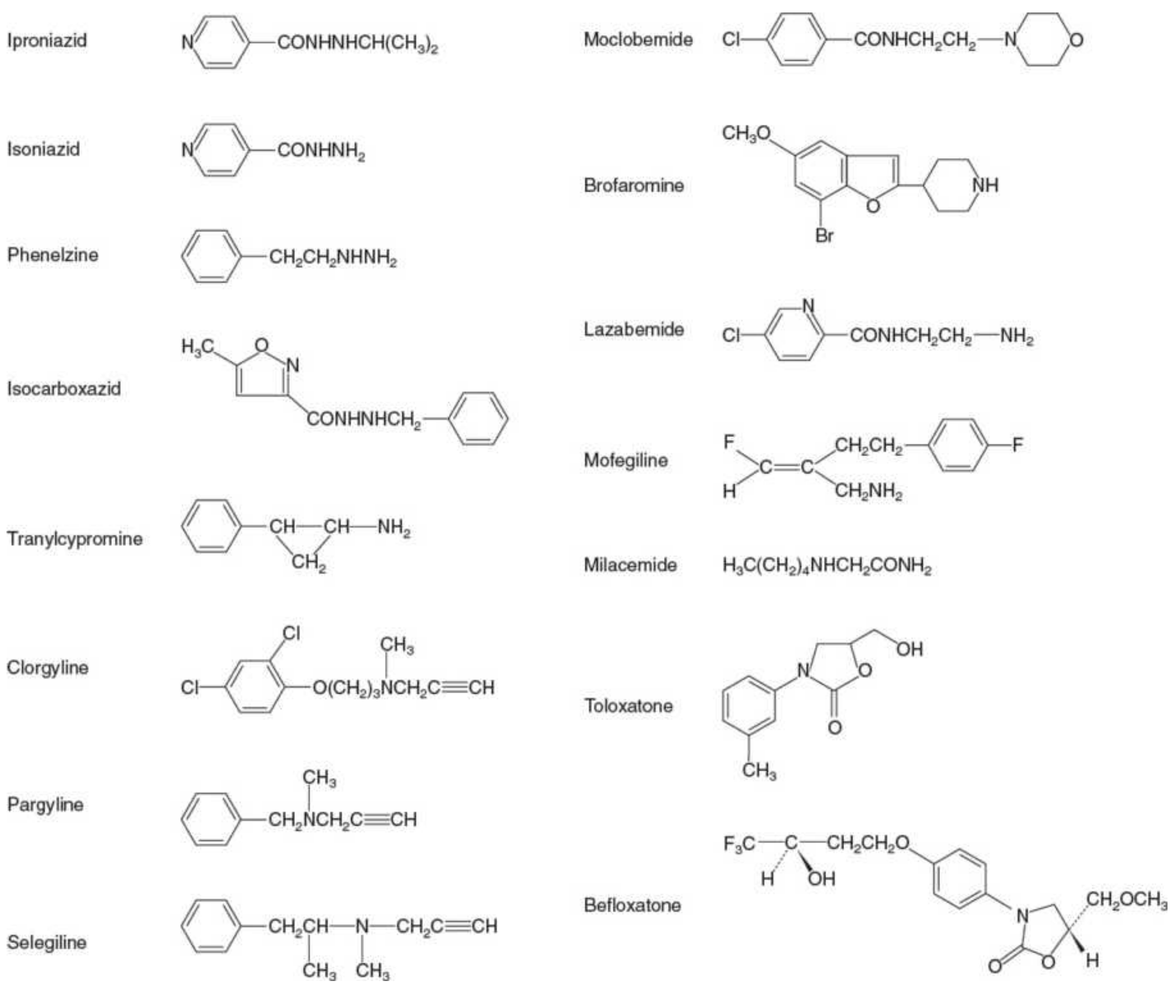
Iproniazid
CONHNHCH(CH3)2
Isoniazid
CONHNH2
Phenelzine
CH2CH2NHNH2
Isocarboxazid
CONHNHCH2
Tranylcypromine
CH—CH—NH2
CH2
Clorgyline
O(CH2)3NCH2C≡CH
Pargyline
CH2NCH2C≡CH
Selegiline
CH2CH—N—CH2C≡CH
Moclobemide
CONHCH2CH2—N
Brofaromine
CH3O
NH
Br
Lazabemide
CONHCH2CH2—NH2
Mofegiline
CH2CH2
CH2NH2
Milacemide
H3C(CH2)4NHCH2CONH2
Toloxatone
OH
Befloxatone
F3C—C—CH2CH2O
CH2OCH3

FIGURE 33.5d–1. Molecular structures of monoamine oxidase inhibitors.

MAO contains an intracellular active site that functions to metabolize amines through oxidation (Fig. 33.5d–2). MAO transfers two hydrogen atoms to a flavin cofactor (FAD). The reduced flavin is reoxidized by oxygen-producing hydrogen peroxide plus the imine, which spontaneously hydrolyzes to the aldehyde (or a substituted amine, in the case of secondary amines like adrenaline) plus ammonia. The aldehyde product is further metabolized for urine excretion.

MAO comes in two forms, MAO-A and MAO-B, encoded by separate genes on chromosome X, originally distinguished by their differing sensitivities to the inhibitors clorgyline and l-deprenyl. Researchers published structures of MAO-A and MAO-B (which share 70% DNA similarity) in 2005 and 2002, respectively.

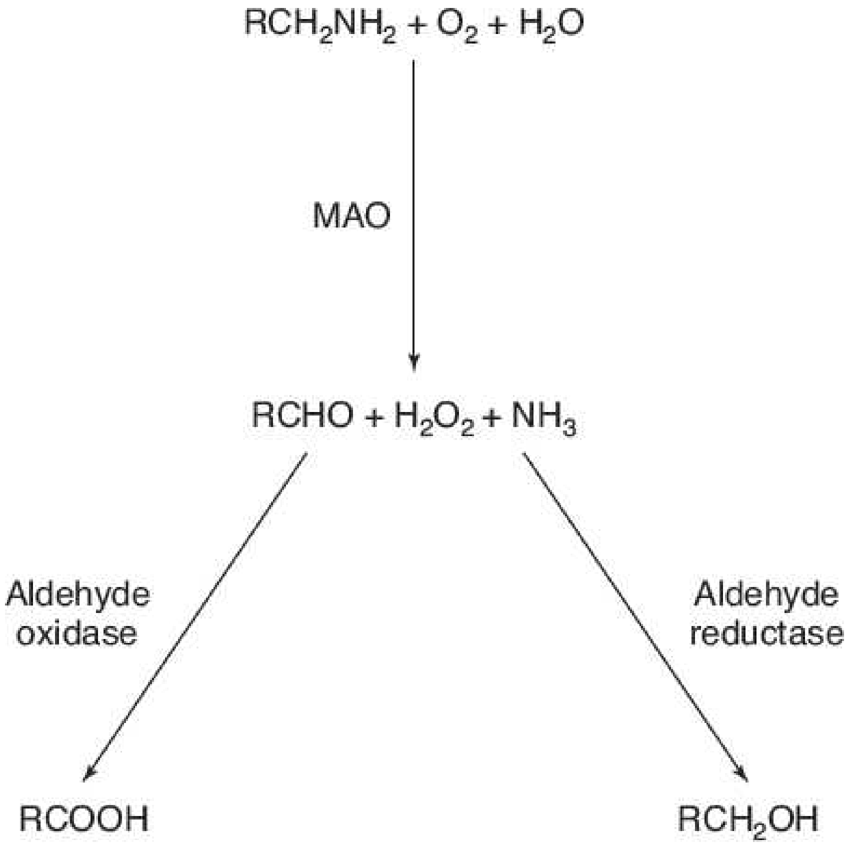

FIGURE 33.5d–2. The reaction catalyzed by monoamine oxidase (MAO), acting on a primary amine.

Researchers have identified four polymorphisms in the gene coding for MAO-A as possible risk factors for psychiatric disorders. The most studied of these polymorphisms is MAO-A-uVNTR, which has been associated with depression in male patients, aggression and impulsivity levels, autism spectrum disorders, alcoholism, and suicide attempts. Studies have also found polymorphisms in intron 13 and exon 14 of MAO-B to be associated with increased risk for PD.

In the central nervous system (CNS), both MAO-A and MAO-B serve to regulate the content of intracellular amine stores, terminate actions of amine neurotransmitters, and protect neurons from exogenous amines. In addition to MAO metabolism, neurotransmitter action may be terminated by catecholamine-o-methyl-transferase (COMT) or vesicular repackaging. Under normal conditions, neurotransmitters are preferentially stored in vesicles, where they are not exposed to MAO. As a result, MAO maintains low neurotransmitter concentration within the cytoplasm of neurons.

MAO-A preferentially deaminates serotonin and norepinephrine and predominates in catecholaminergic neurons in the substantia nigra, locus coeruleus, and periventricular regions of the hypothalamus. MAO-B preferentially deaminates phentylamine and benzylamine and predominates in serotonergic neurons in the dorsal raphe nucleus. Both forms metabolize dopamine, tyramine, and tryptamine. That MAO subtypes populate neurons with neurotransmitters of low affinity suggests a role in preventing the build-up of nonnative neurotransmitters.

In addition to the major monoamines, MAO also serves to metabolize other amines: phenylethanolamine, meta- and para-tyramine and octamine, which are found in the brain at lower levels and collectively known as trace amines. These trace amines can markedly affect the uptake and release of catecholamines and serotonin at nerve endings. There has been increased research into the signaling roles of trace amines with the discovery of specific G-protein–coupled receptors for trace amines in the mammalian brain, which may help regulate the activity of monoamine transporters.

In the periphery, MAO serves to rapidly inactivate biogenic and dietary amines. The most well-known example is MAO-A's role in the intestinal tract. MAO-A inactivates ingested tyramine from the diet, preventing its release into the general circulation. When absorbed, tyramine causes significant release of norepinephrine from peripheral adrenergic neurons leading to vasoconstriction, increased heart rate, and elevated blood pressure. MAO also deactivates peripheral serotonin, epinephrine, and norepinephrine, as it does in the CNS.

In addition to its role in amine regulation, MAO is involved in phase I metabolism of xenobiotics. Like the cytochrome P450 system, MAO can decrease bioactivity. For example, MAO-B in the liver oxidizes sertraline, citalopram, sumatriptan, and phenylephrine.

Finally, recent research has identified MAO as an imidazoline (I2) receptor (IR). The clinical relevance of this finding remains unclear. However, there is some emerging evidence that it may play a role in depression. While IRs are found throughout the brain and on platelets, compounds that have I2-selectivity appear to allosterically inhibit both MAO isozymes, which may explain some anxiolytic and antidepressant effects seen in mice. Additionally, rat models show that chronic treatment with irreversible MAOIs reduces the density of I2 sites. Finally, studies found an increased density of 45-kd I2 receptors in platelets of depressed patients and prefrontal cortices of suicide victims when compared to controls.

MECHANISM OF ACTION

The proposed mechanism of MAOIs efficacy in depression has evolved since its discovery and has expanded with use in other neuropsychiatric disorders.

The "catecholamine hypothesis" or "monoamine theory" of depression postulates that depression in humans is caused by decreased levels of primarily norepinephrine in the synaptic cleft. Several early observations led to this conclusion: the effectiveness of MAOIs in depression, the nearly concurrent discovery of TCAs and their mechanism of norepinephrine reuptake inhibition, and reserpine's depressogenic effect via monoamine depletion. Thus, it was thought that MAOIs alleviate depression by direct inhibition of MAO and an increase in catecholamines available within the synapse. Monoamines, especially dopamine, serotonin, and norepinephrine, have long been known to play an important role in the behavioral regulation of mammals. For instance, in humans, low levels of amines in the brain are associated with hyperactivity and depression, and elevated amines are associated with risk-taking behavior. However, increased monoamine levels can be measured within hours after the first dose of an MAOI, while antidepressant effects take two to four weeks.

Further studies suggest there are secondary adaptations to long-term medication therapy that may be responsible for the antidepressant effect, including the reduced density of alpha-2, beta-adrenergic, and 5-HT$_1$ and 5-HT$_2$ receptors; correction of hypothalamic–pituitary–adrenal (HPA) axis overactivation; increases in BDNF-mediated signaling; and induction of hippocampal neurogenesis. There is also evidence that tranylcypromine at sufficiently high concentrations may act as a monoamine reuptake inhibitor, and phenelzine may inhibit GABA transaminase, increasing levels of GABA within the brain.

Despite evidence challenging a simplistic monoamine-deficiency model, recent imaging studies have renewed interest in the monoamine theory of depression. Researchers using positron emission tomography (PET) have shown increased MAO-A binding (suggesting the decreased presence of monoamines) in the brains of medication-free patients during a major depressive episode compared with healthy controls. A follow-up study found that successful SSRI treatment did not result in normalization of MAO-A levels and that patients who experienced recurrent depressive episodes had increased MAO-A binding in the prefrontal cortex and anterior cingulate cortex compared to those who maintained remission. The authors go on to suggest an advanced monoamine model for depression whose primary mechanism is chronic monoamine loss and associated changes.

CLASSIFICATION

MAOIs may be classified according to chemical structure, reversibility, or selectivity (Table 33.5d–1). Isocarboxazid and phenelzine are hydrazine compounds containing a nitrogen–nitrogen pair on their side chain. All other MAOIs are classified as nonhydrazines. Irreversible MAOIs covalently bind to the enzyme, permanently inhibiting MAO function until new MAO can be produced, usually 2 to 3 weeks. Reversible MAOIs can be displaced by tyramine from the active site of MAO-B, allowing it to be metabolized. Tyramine is not present in high concentrations in the brain and therefore does not inhibit MAOI effects on neurotransmitters. Nonselective MAOIs target both MAO-A and MAO-B, while selective MAOIs target a single isozyme.

Pharmacologic Actions of Specific Agents

Tranylcypromine (Parnate)

Neuroscience-Based Nomenclature (NbN)	
MAO-I	
Pharmacologic domain	Serotonin, Norepinephrine, Dopamine
Mechanism of action	Enzyme inhibitor (MAO-A, MAO-B), releaser (DA, NE)

Tranylcypromine is Food and Drug Administration (FDA)-approved for use in major depressive episodes without melancholia. Off-label uses include other depressive disorders, panic disorder, social anxiety disorder, and posttraumatic stress disorder (PTSD).

PHARMACODYNAMICS

Mechanism of Action. Tranylcypromine is an irreversible and nonselective nonhydrazine MAOI. It covalently binds to both MAO-A and MAO-B. Of note, as a nonhydrazine, its antidepressant effects tend to have a more rapid onset.

PHARMACOKINETICS

Absorption. Tranylcypromine is readily absorbed from the GI tract.

Distribution. There are no clear data on the distribution of tranylcypromine.

Metabolism and Elimination. Tranylcypromine has peak plasma concentrations 1 to 3 hours after ingestion. Its elimination half-life is 2.5 hours. Metabolites are excreted in the urine.

DOSING. Dosing generally starts at 10 mg twice a day (see Table 33.5d–2). If there is no improvement within 2 to 3 weeks, then the dose can be increased by 10 mg every 2 to 3 weeks with a total maximum daily dose of 60 mg.

Tranylcypromine is contraindicated in patients with abnormal liver function tests, hepatic impairment, or a history of liver disease. Tranylcypromine is also contraindicated in patients with severe renal impairment. In patients with mild to moderate renal impairment, tranylcypromine should be used with caution due to the potential for accumulation of active metabolites.

MONITORING PARAMETERS. Complete metabolic panel at baseline, 1 and 3 months.

PREGNANCY AND LACTATION. Tranylcypromine is classified as FDA pregnancy risk category C. According to the manufacturer, tranylcypromine is excreted in human milk, and the potential benefit of the drug should be weighed against the potential hazards to the mother and infant during breastfeeding.

COMMON ADVERSE EFFECTS. Agitation, anxiety, constipation, diarrhea, dizziness, headache, impotence, insomnia, loss of appetite, mania, nausea, orthostatic hypotension, somnolence, weight gain, and xerostomia.

Table 33.5d–2.
Typical Dosage Forms and Recommended Dosages for Currently Available Monoamine Oxidase Inhibitors

Drug	Usual Dose (mg/day)	Maximum Dose (mg/day)	Dosage (Oral) Formulation
Tranylcypromine (Parnate)	20–60	60	10-mg tablets
Phenelzine (Nardil)	30–60	90	15-mg tablets
Isocarboxazid (Marplan)	20–40	60	10-mg tablets
Selegiline (Eldepryl)	10	30	5-mg tablets
Transdermal selegiline (Emsam)	6	12	6-, 9-, or 12-mg patch
Rasagiline	0.5–1.0	1.0	0.5- or 1.0-mg tablets

Phenelzine

Neuroscience-Based Nomenclature (NbN)	
MAO-I	
Pharmacologic domain	Serotonin, Norepinephrine, Dopamine
Mechanism of action	Enzyme inhibitor (MAO-A, MAO-B)

Phenelzine is FDA-approved for use in depressed patients clinically characterized as “atypical,” “nonendogenous,” or “neurotic.” Off-label uses include other depressive disorders, panic disorder, social anxiety disorder, and obsessive-compulsive disorder (OCD).

PHARMACODYNAMICS

Mechanism of Action. Phenelzine is an irreversible and nonselective hydrazine MAOI. It covalently bonds both MAO-A and MAO-B.

PHARMACOKINETICS

Absorption. Phenelzine is well absorbed following oral administration.

Distribution. There is insufficient data on phenelzine's distribution.

Metabolism and Elimination. Peak levels are reached in approximately 2 to 4 hours. The mean half-life is approximately 12 hours. Phenelzine is eliminated in the urine as metabolites.

DOSAGE. The initial dosage of phenelzine is generally 15 mg two to three times per day. It may be increased rapidly, as tolerated, to a maximum dose of 30 mg three times a day. Many patients do not show response until treatment at 60 mg has been continued for at least 4 weeks. Historically, it has been recommended that after the maximum effect is achieved, the dosage should be slowly reduced to the lowest dose needed to maintain clinical effectiveness.

Due to the potential for drug accumulation and associated effects, phenelzine is not recommended in patients with hepatic disease or severe renal impairment.

PREGNANCY AND LACTATION. Phenelzine is classified as FDA pregnancy risk category C. According to the manufacturer, the safe use of phenelzine during breastfeeding has not been established, and the potential benefit of the drug should be weighed against the possible hazards to the mother and nursing infant.

MONITORING PARAMETERS. Complete metabolic panel at baseline, 1 and 3 months.

COMMON ADVERSE EFFECTS. Constipation, disorders of ejaculation or orgasm, dizziness, edema, fatigue, headache, hyperreflexia, impotence, elevated values on liver function tests, orthostatic hypotension, sleep disorders, somnolence, tremor, weight gain, and xerostomia.

Isocarboxazid

Neuroscience-Based Nomenclature (NbN)	
MAO-I	
Pharmacologic domain	Serotonin, Norepinephrine, Dopamine
Mechanism of action	Enzyme inhibitor (MAO-A, MAO-B)

Isocarboxazid is FDA-approved for use in major depressive disorder. Off-label use includes panic disorder and social anxiety disorder.

PHARMACODYNAMICS

Mechanism of Action. Isocarboxazid is an irreversible and nonselective hydrazine MAOI. It covalently bonds both MAO-A and MAO-B.

PHARMACOKINETICS

Absorption. Isocarboxazid is readily absorbed from the GI tract.

Distribution. There is insufficient data on the distribution of isocarboxazid.

Metabolism and Elimination. Isocarboxazid reaches peak plasma concentrations in 3 to 5 hours. The half-life is approximately 2.5 hours. It is metabolized by the liver and excreted as metabolites in the urine.

DOSAGE. The initial dose of isocarboxazid is 10 mg twice a day. The dose may be increased from 10 mg daily up to a maximum dosage of 60 mg total daily in divided doses. If a therapeutic response is not seen in 3 to 6 weeks, it is unlikely to develop. After maximum benefit is achieved, the dose should be reduced to the lowest effective dose.

Isocarboxazid is not recommended in patients with hepatic impairment or renal impairment.

PREGNANCY AND LACTATION. Isocarboxazid is classified as FDA pregnancy risk category C. According to the manufacturer, the amount of excretion of isocarboxazid and its metabolites in human milk have not been determined; therefore, isocarboxazid should only be used during breastfeeding if clearly needed.

MONITORING PARAMETERS. Complete metabolic panel at baseline, 1 and 3 months.

COMMON ADVERSE EFFECTS. Anxiety, blurred vision, constipation, dizziness, headache, insomnia, mania, somnolence, weight gain, and xerostomia.

Selegiline

Neuroscience-Based Nomenclature (NbN)	
MAO-I	
Pharmacologic domain	Dopamine, Norepinephrine, Serotonin
Mechanism of action	Enzyme inhibitor (MAO-B, MAO-A)

Selegiline is FDA-approved for use in PD in combination with levodopa. Off-label use includes Alzheimer disease (AD), depression, and ADHD.

PHARMACODYNAMICS

Mechanism of Action. Selegiline is an irreversible and selective MAO inhibitor. At doses lower than 10 mg, it covalently binds MAO-B. At doses above 10 mg, it covalently binds both MAO-A and MAO-B. Of note, selegiline is metabolized to amphetamine compounds, which may explain its effectiveness in the treatment of attention-deficit hyperactivity disorder (ADHD).

PHARMACOKINETICS

Absorption. Selegiline is readily absorbed in the GI tract.

Distribution. Up to 85% of plasma selegiline is reversibly bound to proteins.

Metabolism and Elimination. Peak serum concentrations are found in 0.5 to 2 hours. At a steady state, the elimination half-life is approximately 10 hours. The drug is rapidly and completely metabolized to three active derivatives with the following half-lives: *N*-desmethyldeprenyl, 2 hours; 1-amphetamine, 17.7 hours; and 1-methamphetamine, 20.5 hours.

DOSAGE. Effective antidepressant doses of oral selegiline are not usually achieved until the range of 20 to 30 mg/day is used, at which point selegiline loses its selectivity and requires dietary monitoring.

For ADHD, trials have used 2.5 mg twice a day for children under 5 years old and 5 mg twice a day for children over 5 years old.

Specific guidelines for dosage adjustments of selegiline in patients with hepatic impairment and renal impairment are not available.

PREGNANCY AND LACTATION. Selegiline is classified as FDA pregnancy risk category C. According to the manufacturer, it is unknown if selegiline is excreted into breast milk, and a decision should be made whether to discontinue the drug or discontinue breastfeeding, taking into account the importance of the drug to the mother.

MONITORING PARAMETERS. None.

COMMON ADVERSE EFFECTS. Nausea, dizziness, and insomnia.

Transdermal Selegiline. Transdermal selegiline is FDA-approved for use in major depressive disorder.

PHARMACODYNAMICS

Mechanism of Action. Transdermal selegiline is an irreversible and selective MAOI. At doses less than 6 mg, it covalently binds MAO-B only. At doses above 9 mg, it covalently binds MAO-A and MAO-B.

PHARMACOKINETICS

Absorption. Approximately 23% to 30% of the medication is absorbed through the skin over 24 hours.

Distribution. Once absorbed, it is 90% protein bound, rapidly distributed throughout the body, and crosses the blood–brain barrier. Transdermal administration avoids first-pass metabolism and results in much higher bioavailability (74%) compared to the oral form (4%).

Metabolism and Elimination. Selegiline is metabolized via *N*-dealkylation or *N*-depropargylation by several CYP450 enzymes (2A6, 2B6, 2C9, 3A4, and 3A5). Under steady-state conditions, the median elimination half-life increases to 10 hours. The half-life is approximately 18 to 25 hours. The final metabolites are methamphetamine and amphetamine, which are excreted in the urine.

DOSAGE. The initial dose is 6 mg/24 hr. Although the maximum dose is 12 mg, there is little evidence for doses above 6 mg. If clinically indicated, dose increases should be made in 3 mg increments every 2 to 3 weeks. In older adult patients, the recommended dose is 6 mg per 24 hours. At a dose above 9 mg, the manufacturer suggests following a diet modified for reduced tyramine intake.

Dose adjustments for transdermal selegiline are not needed in patients with mild to moderate hepatic impairment or any degree of renal impairment.

PREGNANCY AND LACTATION. Transdermal selegiline is classified as FDA pregnancy risk category C. According to the manufacturer, it is not known if transdermal selegiline is excreted into human milk, and caution is advisable if administering selegiline to a nursing woman.

MONITORING PARAMETERS. None.

COMMON ADVERSE EFFECTS. Application site reaction, decreased systolic blood pressure, diarrhea, headache, indigestion, insomnia, orthostatic hypotension, weight loss, and xerostomia.

Rasagiline

Neuroscience-Based Nomenclature (NbN)	
MAO-I	
Pharmacologic domain	Dopamine, Serotonin
Mechanism of action	Enzyme inhibitor (MAO-B)

Rasagiline is FDA-approved for use in PD as monotherapy or an adjunct to levodopa.

PHARMACODYNAMICS

Mechanism of Action. Rasagiline is an irreversible and selective MAOI. It covalently binds MAO-B.

PHARMACOKINETICS

Absorption. Rasagiline is rapidly absorbed and reaches peak levels in 1 hour.

Distribution. Plasma protein binding ranges from 88% to 94%.

Metabolism and Elimination. Rasagiline is metabolized by the cytochrome P450 system in the liver, predominately CYP1A2, and is mainly excreted via urine.

DOSING. For monotherapy in PD, rasagiline may be started at 1 mg daily. In patients taking levodopa, the recommended initial dose is 0.5 mg daily, which, if tolerated, may be increased to 1 mg daily if needed.

In patients taking a CYP1A2 inhibitor such as ciprofloxacin, dosages should not exceed 0.5 mg daily.

In patients with mild hepatic impairment, dosage should not exceed 0.5 mg daily, and it should not be used in patients with moderate or severe hepatic impairment.

No dose adjustment is required for patients with mild or moderate renal impairment, and it has not been studied in patients with severe renal impairment.

At recommended doses, rasagiline retains MAO-B selectivity, and although the manufacturer does not recommend dietary restrictions, it suggests foods with high tyramine content may cause hypertensive cries.

PREGNANCY AND LACTATION. Rasagiline is pregnancy category C, and there are no adequate studies on the use of the medication in pregnant women. It is not known whether rasagiline is excreted in human milk.

MONITORING PARAMETERS. Liver function test at baseline, 1 and 3 months.

COMMON ADVERSE EFFECTS. Weight loss, orthostatic or postural hypotension, dry mouth, abnormal involuntary movements, and hallucinations.

Drug Level Monitoring

MAOI plasma levels may be measured, but there is little evidence that levels correspond to treatment response and may be difficult to interpret given the short half-life of MAOIs.

There is evidence that platelet MAO inhibition is a predictor of response. However, its use is not recommended. Several studies have shown that patients taking phenelzine who have more than 80% platelet MAO inhibition after 2 weeks of treatment showed higher response rates. However, researchers point out that the moderate effect size means that this accounts for only 9% to 16% of treatment outcomes. MAO-B platelet inhibition is most likely an epiphenomenon as it has been shown in studies of selegiline that 100% inhibition can occur in the absence of antidepressant effects.

Laboratory Interference

Laboratory interference is not a major problem with the use of MAOIs. They may lower blood glucose values. They have also been associated with a minimal false elevation in thyroid function test results.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Central Nervous System

Commonly reported CNS effects include: dizziness, headache, drowsiness, sleep disturbance, fatigue, weakness, twitching, myoclonus, and hyperreflexia. Uncommon but serious CNS effects include ataxia, shock-like coma, toxic delirium, acute anxiety reaction, psychosis, transient respiratory and cardiovascular depression following electroconvulsive therapy, confusion, memory impairment, hypomania, and mania. Rasagiline may cause increased impulsive or compulsive behaviors such as gambling, sexual activity, and binge eating.

Cardiovascular

Common side effects include orthostatic and postural hypotension, syncope, palpitations, and tachycardia.

Skin

Adverse dermatologic effects include rash, pruritus, and hyperhidrosis. Rasagiline may increase the risk for melanoma, and regular screening is recommended.

Gastrointestinal

Commonly reported GI effects include: constipation, diarrhea, nausea, xerostomia, and elevated hepatic enzymes. Severe but less frequent events include glottis edema, progressing necrotizing hepatocellular damage, and reversible jaundice, the last two more likely with hydrazine derivatives.

Genitourinary

Common urinary symptoms include urinary retention and hesitancy. Commonly reported sexual side effects include: impotence, ejaculation dysfunction, and anorgasmia. Clinical experience shows fewer sexual side effects with tranylcypromine. There have been case reports of priapism in patients taking phenelzine. Impaired water excretion compatible with the syndrome of inappropriate secretion of antidiuretic hormone (SIADH) has been associated with antidepressant use.

Optic

Glaucoma and blurred vision have been infrequently reported.

Hematologic

There are reports of rare occurrences of anemia, leukopenia, agranulocytosis, and thrombocytopenia.

Table 33.5d–3.
Relative Restrictions in Food and Beverages With Irreversible MAOI Use

Restriction	Foods
Absolute (high risk)	Aged cheeses; aged and cured meats; banana peel; broad bean pods; improperly stored or spoiled meats, poultry or fish; Marmite; sauerkraut; soy sauce and soybean condiments; tap beer
Moderate (moderate risk)	Red or white wine; bottle or canned beer
Unnecessary (low to no risk)	Avocados; bananas; beef/chicken bouillon; chocolate; fresh and mild cheeses; fresh meat, poultry and fish; gravy; monosodium glutamate; peanuts; properly stored pickled or smoked fish; soy milk; yeast extracts other than Marmite

Neurologic

Tremor is a commonly reported side effect of phenelzine. Peripheral neuropathy, sometimes associated with a pyridoxine deficiency, has been reported in patients receiving phenelzine. There are case reports that suggest an association between MAOI use and pseudoparkinsonism with symptoms of bradykinesia, tremor, cogwheel rigidity, and dysarthria.

WARNINGS AND PRECAUTIONS

Food–Drug Interactions

Nonselective and irreversible MAOIs, when combined with food rich in tyramine, can cause increased release of catecholamines, resulting in a hypertensive crisis characterized by an occipital headache, palpitations, nausea, vomiting, diaphoresis, and tachycardia. Severe reactions can lead to cardiac arrhythmias, stroke, and intracranial hemorrhage. Prior to recommendations on dietary restrictions, rates of hypertensive reactions were estimated to occur in 2.4% to 25% of patients treated with MAOIs.

After observation of hypertensive reactions in the 1960s, clinical use of MAOIs dropped dramatically. At the height of concern over drug–food interactions, more than 70 different types of food and beverages were on the list of foods to be avoided. However, over the past 20 years, a critical review of reports of adverse effects and analysis of tyramine levels in food and beverages has refined the list of dietary restrictions (Table 33.5d–3).

Drug–Drug Interactions

Hypertensive Crisis. Similar to tyramine-induced hypertensive crises, there are dangers in drug–drug interactions. Any medication that increases adrenergic stimulation by a mechanism other than MAO inhibition can lead to a hypertensive crisis (Table 33.5d–4). The list includes decongestants that contain alpha-1 agonists, stimulants that increase the release of norepinephrine, antidepressants that block norepinephrine reuptake, and appetite suppressants and analgesics with norepinephrine-reuptake inhibition properties.

Table 33.5d–4.
Medications Contraindicated With MAOI Use due to Risk of Hypertensive Crisis

Decongestants: phenylephrine, oxymetazoline, ephedrine, pseudoephedrine, phenylpropanolamine
Stimulants: methylphenidate, amphetamines
Antidepressants: SSRIs, SNRIs, TCAs, bupropion, buspirone
Appetite suppressants: sibutramine, phentermine
Analgesics: tramadol
Centrally acting antihypertensives: guanethidine, methyldopa
Antiparkinson agents: levodopa

Serotonin Syndrome. Concomitant administration of serotonergic agents with MAOIs, especially those with MAO-A inhibition, can result in dangerous levels of serotonin, leading to serotonin syndrome. Increased activation at postsynaptic serotonin receptors causes a disruption in thermoregulation, leading to a syndrome of fever, altered mental status, myoclonus, diaphoresis, and diarrhea. In extreme cases, symptoms may include hypothermia, seizures, coma, cardiovascular collapse, permanent brain damage, and death. MAOIs should not be prescribed in combination with SSRIs, SNRIs, TCAs, certain narcotics, and other medications with serotonergic properties (Table 33.5d–5).

In addition, care must be taken when changing antidepressants to or from MAOIs. A washout period of 2 weeks is required for all SSRIs prior to starting an MAOI, except for fluoxetine, which due to its long half-life, requires a 5-week washout period. A 2- to 4-week waiting period is also required when switching from an MAOI to another antidepressant or when discontinuing an MAOI prior to ending the restricted diet. The selegiline patch loses its selectivity at higher doses, so the manufacturer recommends dietary restrictions with doses of 9 mg or higher and for 2 weeks following any discontinuation.

Weight Gain

In a retrospective review, weight gain occurred in 35.7% of patients taking phenelzine and 10.5% taking tranylcypromine.

Orthostatic Hypotension

All nonselective MAOIs may reduce blood pressure, likely due to central alpha-adrenergic agonist effects of increased norepinephrine levels. Rates of orthostasis may be higher in phenelzine. Hypotension usually develops gradually over the first 2 to 3 weeks of treatment and correlates with peak plasma levels. The side effect may be reduced with divided dosing and resolves hours to days after discontinuation.

Insomnia

MAOIs, especially tranylcypromine, are associated with insomnia. MAOIs are known to reduce the time spent in REM sleep. Transdermal selegiline also shows this side effect in a dose-dependent relationship. Insomnia may be alleviated by avoiding evening dosing.

Table 33.5d–5.
Medications Contraindicated With MAOI Use due to Risk for Serotonin Syndrome

Antidepressants: SSRIs, SNRIs, TCAs
Other TCAs: cyclobenzaprine, carbamazepine
Appetite suppressant: sibutramine
Cough suppressant: dextromethorphan
Opioids: meperidine, tramadol, methadone, propoxyphene
Antihistamines: chlorpheniramine, brompheniramine

Application Site Reaction

The most common side effect of transdermal selegiline is an application site reaction, which occurs in a dose-dependent relationship, and occurs in 24% of patients.

Hepatotoxicity

Postmarketing monitoring showed an increased risk for hepatotoxicity for hydrazine MAOIs (isocarboxazid and phenelzine) compared to the nonhydrazine (tranylcypromine). However, the absolute risk appears to be low (less than 1 in 30,000).

Psychosis

The combination of MAOIs and dextromethorphan has been reported to cause brief episodes of psychosis and bizarre behavior.

Overdose

Signs and symptoms of overdose include dilated pupils, fever, diaphoresis, drowsiness, agitation, headache, hallucinations, trismus, opisthotonos, seizures, coma, fever, tachycardia, hypertension, hypotension, respiratory depression/failure, shock, and death. Signs and symptoms may be absent or minimal during the initial 12-hour period following ingestion and may develop slowly, reaching a maximum in 24 to 48 hours. In some cases, symptoms may persist for 8 to 14 days. Disseminated intravascular coagulation, rhabdomyolysis, and acute tubular necrosis can also occur. Treatment of overdose is mainly supportive. Inducing vomiting or gastric lavage with charcoal may be helpful in early poisoning.

Withdrawal

Common withdrawal symptoms include nausea, vomiting, and malaise. Abrupt withdrawal for patients on long-term treatment with MAOIs can result in a withdrawal syndrome characterized by increased arousal, mood disturbance, psychosis, and convulsions within 24 to 72 hours after drug discontinuation.

Contraindications

Contraindications to MAOI use include cardiac disease, cerebrovascular disease, hypertension, headache, hepatic disease, pheochromocytoma, renal disease, and general anesthesia with certain opioid analgesics.

THERAPEUTIC INDICATIONS AND CLINICAL DRUG STUDIES

A significant number of studies on MAOIs occurred prior to the publication of DSM-II, making interpretation difficult due to inconsistency in mood disorder diagnoses across studies. Reviewed here are studies with clear inclusion criteria and, when possible, a control group. Limitations are noted where they arise. Meta-analyses are generally not used as they usually include RIMA studies. Evidence for the efficacy of MAOIs not available in the United States, most notably moclobemide and brofaromine, is not discussed.

MAOIs were first used psychiatrically to treat patients with depression. However, given concerns over efficacy, side effects, drug–drug interactions, and dietary restrictions, MAOIs as a class generally fell out of favor. SSRIs have largely replaced MAOIs for psychiatric use. MAOIs are more frequently prescribed for atypical major depressive disorder and treatment-resistant depression (TRD). Less common uses include typical major depressive disorder, psychotic depression, dysthymia, bipolar depression, eating disorders, PTSD, social anxiety disorder, and panic disorder. Selective MAO-B inhibitors are used regularly in the treatment of PD and AD.

Atypical Depression

Early observations on groups of depressed patients who responded to MAOIs but not to TCAs sparked interest in identifying the characteristics of this subset of depressed patients. Studies reported that these "atypical" patients had depressed mood with anxiety and phobia or reactive mood with reversed vegetative symptoms. Subsequent controlled trials by a group at Columbia codified patients with atypical features of depression, defining criteria as reactive mood plus at least two of four associated features: hyperphagia, hypersomnolence, leaden paralysis, and sensitivity to rejection. They also identified a second group characterized by depression plus panic attacks for which MAOIs also appeared to show an advantage over TCAs. However, replication studies by the same group failed to show a difference between MAOIs and TCAs in mixed depressive/anxiety states.

Acute Treatment. The Columbia group summarized their findings in a 1993 paper. From the data of 269 patients, they compared response rates on Clinical Global Improvement scales for patients taking phenelzine (72%), imipramine (44%), and placebo (26%). They concluded that rather than an enhanced response to MAOIs, patients with atypical depression show substandard responses to TCAs. Through stratification, they also demonstrated that the presence of anxiety symptoms did not factor into response rates, despite previous suggestions of superior MAOI response in depression with anxiety.

Since that review, two controlled studies have been published, one comparing an MAOI to cognitive-behavior therapy (CBT) and another to an SSRI. A 1999 study randomized patients with atypical depression to phenelzine, CBT, or placebo for 10 weeks. Response rates for phenelzine (58%) and CBT (58%) were equivalent to each other and superior to placebo (28%). An analysis found similar results for reduction in Hamilton Rating Scale for Depression (HAM-D) scores. Patients taking phenelzine had higher rates of marked side effects (92%) compared to placebo (53%).

The only trial that has compared an MAOI available in the United States to an SSRI is a randomized, double-blind, non–placebo-controlled trial comparing fluoxetine and phenelzine. Forty patients with atypical depression randomized to fluoxetine or phenelzine showed no difference in response (80% vs. 85%) or remission (80% vs. 70%). The only patient who dropped out due to an adverse effect (confusion) was taking phenelzine. Dizziness, asthenia, arthralgias, somnolence, and blood pressure changes were more common in the phenelzine group. Tremor and weight gain were more common in the fluoxetine group.

Maintenance Treatment. In addition to acute atypical depression treatment, there is one long-term study of relapse prevention in atypical depression. In this study, 60 patients who had improved with phenelzine or imipramine over 6 months were randomly assigned to the same medication or placebo for an additional 6 months. Recurrence rates were 23% for those receiving phenelzine, 41% for imipramine, 47% for those who switched from imipramine to placebo, and 87% for those who switched from phenelzine to placebo.

Treatment-Resistant Depression

Recently, the use of MAOIs in typical depression has been limited to TRD. Most controlled trials have focused on tranylcypromine after early case studies found a benefit from tranylcypromine for patients with TCA-refractory depression. Response rates in these trials vary greatly, which may reflect the difficulty in defining this group, and it is a challenge to make firm conclusions about specific advantages of MAOIs in TRD.

In a 1988 study, 21 patients with major depressive disorder who had not responded to oxaprotiline or fluvoxamine were randomized to tranylcypromine or nomifensine. Compared to the nomifensine group, the tranylcypromine group showed statistical superiority in a reduction in HAM-D scores but not response rate (45% vs. 10%) at 4 weeks.

A 1993 study compared brofaromine and tranylcypromine in 39 patients with major depressive disorder who had previously not responded to 4 weeks of maprotiline or nortriptyline. Of note, 12 patients had psychotic features and 23 melancholic features. There was no difference in response rates for brofaromine (45.5, 59.1%) and tranylcypromine (52.9, 58.9%) as measured by Hamilton Depression Subscale and Clinical Global Improvement, respectively. Melancholic features did not predict response. Brofaromine was better tolerated.

A second study in 1993 also compared brofaromine and tranylcypromine. In this case, 93 inpatients who had failed two classes of antidepressants were treated for 6 weeks. There were no differences between the two groups on reduction in HAM-D scores or response rates (nearly 74% for both groups). There were no differences in tolerability.

A 38-year-old divorced mother of two teenage daughters presented at the Mood Disorders Clinic with a history of failed antidepressant treatment trials. Two years earlier, she had become increasingly irritable, exhausted, and tearful after the breakup of her marriage. She could not tolerate fluoxetine (Prozac) or sertraline and discontinued each of these SSRIs after only a few days because of GI symptoms. Although she could tolerate desipramine (Norpramin) up to 300 mg daily, she felt no benefit after 4 months and was reluctant to continue this TCA medication because of adverse effects (weight gain and sweating, in particular). Because of her increasing fatigue, increasing sleep needs, and generally atypical profile, she was offered a trial of phenelzine. After 4 weeks, at a dosage of 60 mg daily, she reported significant improvement in her energy level, mood, and sleep pattern. Two years later, she continues to take 45 mg of phenelzine daily with some adverse effects (occasional insomnia and anorgasmia).

A 2004 study enrolled 77 patients who had failed either of two phases: (1) a double-blind, randomized trial of imipramine or fluvoxamine, followed by augmentation with lithium or (2) an open trial with a TCA. They were randomized to either tranylcypromine or phenelzine for 5 weeks. Results showed no difference in response rate between tranylcypromine (44%) and phenelzine (47%) or remission rates (18% and 11%, respectively). About 21% of patients reported “severe” side effects, including dizziness, agitation, and insomnia, though there were no differences between the two groups.

Level 4 of the STAR*D trial compared tranylcypromine versus extended-release venlafaxine plus mirtazapine in 109 patients who had not responded to three prior medication trials. The results showed no difference in response (13.8% vs. 15.7%) or remission (6.9% vs. 13.7%) between the two groups. Patients in the tranylcypromine group were more likely to exit the study (41.4% vs. 21.6%), and almost half had less than 6 weeks of treatment, suggesting that MAOI treatment may be less tolerable than other antidepressant treatment. However, patients did not differ in ratings on side effect burden scales. Critics of the study point out that the average tranylcypromine dose was only 36.9 mg, less than in similar studies, and may explain the lower response rates.

Finally, a crossover study compared selegiline to placebo in 16 treatment-resistant (failed at least two clinical trials of antidepressant) older patients with depression. Patients received selegiline 60 mg or placebo in two 3-week phases. During the active treatment phase, patients showed significant improvement in HAM-D scores. Overall, 8 of 16 patients responded to treatment. Responders were more likely to be older, less depressed, less anxious, and more cognitively intact. Researchers also performed cognitive tests with no changes during active treatment. There were no differences in side effect rates in each group.

Major Depressive Disorder

Early clinical observation suggested that MAOIs tend to be more effective than TCAs in the treatment of atypical depression and less effective than TCAs in melancholic or endogenous depressions. A particularly large and influential early study from the United Kingdom showed phenelzine 45 mg to be marginally inferior to placebo in a group of depressed inpatients of unspecified type. Subsequent trials with higher doses showed more equivalent efficacy for MAOIs to TCAs in the treatment of endogenous or melancholic depression. Discussed here are trials investigating groups of patients with unspecified type of major depressive disorder or endogenous/melancholic subtype. Of note, there are no trials comparing MAOIs available in the United States with second-generation antidepressants.

Acute Treatment. Early studies of MAOI use in depression were characterized by limited size, short duration, and low dosages. A 1960 study crossed over 25 moderately or severely endogenously depressed patients to receive phenelzine 45 mg or placebo for 2 weeks each. Improvement rates were significantly higher for the second phenelzine group compared to placebo, but not the first. In a 1961 study, 21 consecutively admitted patients diagnosed with depression (13 with endogenous type) were given phenelzine or placebo for 3 weeks and then switched. On a seven-point improvement scale, a higher proportion of patients showed “complete recovery” or “marked improvement” in the phenelzine (50%) group versus placebo (15%). Fifty-three percent of the endogenous depressed patients improved or remitted.

Later trials appear to show equivalence for acute depression, in endogenous or mixed groups, when comparing MAOIs to TCAs, without clear evidence of a difference in side-effect burden. In 1981, researchers randomized 49 inpatients diagnosed with depression (33 nonendogenous) to phenelzine or imipramine for 3 weeks after a 1-week placebo-response period. Both groups showed an equal and statistically significant response as measured by the HAM-D. There was no subgroup analysis.

A 1986 study of 75 depressed patients over age 55 compared phenelzine and nortriptyline to placebo. Response rates found phenelzine and nortriptyline superior to placebo (59.1% vs. 60.0% vs. 10.7%) as measured by HAM-D. A majority of patients in each group had the endogenous subtype, but a subgroup analysis was not performed. Of note, only two patients from each active arm dropped out of the study due to side effects.

A 1987 study investigated the use of phenelzine or imipramine for 6 weeks after a 1-week placebo-response period in 83 patients with a diagnosis of major depression with melancholia or dysthymic disorder. In subgroup analysis for melancholia, reduction in scores on the HAM-D did not significantly differ between phenelzine (58.2%) and imipramine (71.4%) treatment. Side effects were not reported.

Finally, two short-term studies established the efficacy of transdermal selegiline for major depressive disorder. A 6-week trial of 6 mg versus placebo showed significant results (HAM-D reduction of 9 vs. 6.5 points), and an 8-week trial with doses from 6 to 12 mg also showed significant changes (HAM-D reduction of 10.9 vs. 8.6 points). Oral selegiline has been found to be effective in treating depression only at doses higher than 20 mg daily, a dose where selective enzyme inhibition is lost, and a tyramine-restricted diet is required.

Maintenance Treatment. Few studies have investigated the long-term use of MAOIs for the prevention of recurrent depressive episodes.

A 1991 study looked at maintenance on phenelzine versus placebo. After acute unblinded treatment with phenelzine, responders were entered into a 16-week continuation phase followed by double-blind randomization to phenelzine 60 mg, phenelzine 45 mg or placebo for a 2-year maintenance phase. Forty-seven patients entered the maintenance phase. Both phenelzine doses showed significant protection against relapse within 3 months but not against recurrence (symptom onset beyond 3 months of continuation of treatment). Generalization from this study is limited as 36% of patients who initially responded to treatment discontinued during continuation therapy prior to enrollment in the maintenance phase.

Fifty-two adults over 55 years old with depression who had responded to acute treatment with phenelzine or nortriptyline in the 1986 study were randomized to maintenance on the study drug or tapered to placebo. Results showed significantly less recurrence for patients taking phenelzine (13.3%) versus nortriptyline (53.8%) and placebo (65.2%) over 1 year. The phenelzine group had only two dropouts due to side effects.

A 52-week study investigated relapse rates for 322 patients who had responded to acute treatment with transdermal selegiline 6 mg daily and who were randomized to continued treatment or placebo. Significantly fewer patients on transdermal selegiline (16.8%) relapsed compared to placebo (30.7%).

Psychotic Depression

It is unclear that patients diagnosed with psychotic depression in early trials would meet contemporary criteria, and those early trials found no efficacy for MAOIs in the treatment of psychotic depression. There are no controlled studies. The only more recent study is an open-label trial examining the use of phenelzine in depressed inpatients. The researchers classified patients into three groups: definitely psychotic, probably psychotic, and nonpsychotic. Response rates were 22%, 43%, and 68%, respectively. This inpatient study showed increased effectiveness for phenelzine in nonpsychotic versus psychotic patients. MAOIs are not usually considered for this indication.

Dysthymia

There is limited evidence for the use of phenelzine in dysthymic disorder. Three prospective studies have investigated the use of MAOIs for dysthymia, but only one used an MAOI available in the United States. A 1993 study comparing phenelzine to imipramine in dysthymic disorders over 6 weeks found superior response in the phenelzine group when measured by the response on the HAM-D.

A retrospective analysis by the Columbia group studying atypical depression identified 37 patients meeting DSM-III criteria for dysthymic disorder. The researchers found response rates of 22%, 50%, and 17% for phenelzine, imipramine, and placebo, respectively. However, this study was limited by a small number of subjects, a short duration of only 6 weeks for a chronic illness, and the exclusion of patients with melancholic features. The same group found in a pooled group analysis that phenelzine was more effective than imipramine and placebo (83% vs. 45% vs. 21%) in double depression.

Bipolar Depression

There are three positive prospective studies for MAOI use in bipolar depression, one negative prospective study and one positive retrospective analysis.

A 1982 study on “anergic” depression, defined as depressed mood with motor retardation, fatigue, volitional inhibition, and reversed neurovegetative features, included 59 patients, 29 of whom met the criteria for bipolar disorder. The other 30 were classified as “pseudounipolar” or “borderline syndromes.” Researchers randomized patients to tranylcypromine or a placebo. They did not report subgroup analysis. Twenty of twenty-four patients (83%) on an active drug responded, and only 4 of 31 (13%) on placebo. The same group, in 1991, conducted a double-blind study comparing tranylcypromine to imipramine for patients with anergic bipolar depression. Fifty-six patients were treated with either tranylcypromine or imipramine without a mood stabilizer over the 6-week acute treatment phase and 10-week maintenance phase. Tranylcypromine outperformed imipramine in acute response (81% vs. 48%) and in sustained response (71% vs. 20%). Of note, 21% of tranylcypromine and 25% of imipramine patients developed mania or hypomania. Significantly more (7 vs. 2) patients taking imipramine dropped out than those taking tranylcypromine. They also conducted a crossover trial involving 16 patients who did not respond to the original treatment. Nine of twelve patients who failed imipramine responded to tranylcypromine and had categorical improvement at similar rates to the original tranylcypromine group (75% and 81%). Two developed hypomania. Only one of four patients who crossed over to imipramine responded; two withdrew due to side effects, and the other withdrew due to a mixed state.

The other prospective trial for bipolar disorder failed. Researchers compared tranylcypromine to lamotrigine in patients with bipolar depression who had previously failed a 6-week trial with an SSRI, TCA, venlafaxine, or bupropion added to an adequate dose of mood stabilizer. This study was severely limited by the small number who enrolled (20) and completed (11). However, response rates, though not statistically significant partly due to the small power of the study, were consistent with the response rates of trials discussed above: 62.5% for tranylcypromine and 36% for lamotrigine. Two patients in the lamotrigine group developed mania.

Finally, a retrospective analysis of data from the Maintenance Therapies in Bipolar Disorder (MTBD) study compared the addition of MAOIs or paroxetine in 52 patients with bipolar depression who had already failed to respond to a mood stabilizer alone (usually lithium). Rates of durable recovery (defined as a period of 8 weeks of euthymia starting within the first 16 weeks of treatment) were significantly higher in the MAOI group (70%) versus the paroxetine group (33%). The authors did not report adverse events.

Eating Disorders

Two controlled trials and a reanalysis have found efficacy for phenelzine and isocarboxazid in bulimia nervosa. Despite concerns about adherence to diet, the studies reported no hypertensive crises. There are no controlled trials of MAOI use in anorexia nervosa or binge eating disorder.

A 1986 crossover study compared isocarboxazid to placebo over 12 weeks in patients with bulimia nervosa. Isocarboxazid significantly decreased the number of binge episodes (7.0 to 2.8) compared to placebo (7.5 to 4.5). There were no adverse food or drug interactions, but 40% experienced insomnia or hypotension.

A 1987 study enrolled 82 patients with bulimia to receive phenelzine or placebo over 8 weeks following a 2-week placebo washout phase. Twenty dropped out before randomization, and another 12 did not complete at least 2 weeks of treatment. Although the number of binges per week at the study end did not differ between the two groups, the phenelzine group showed a statistically significant reduction in change in binge frequency (64%) versus placebo (5%). A covariate analysis showed that this response was not related to antidepressant effects. Nine of thirty-one patients taking phenelzine did not complete the 8-week trial due to side effects. There were no reports of hypertensive crisis.

Finally, the Columbia group reanalyzed their data on treatment with phenelzine, imipramine, or placebo for atypical depression. They identified 24 patients who met the criteria for comorbid bulimia. The study is limited by outcome measures that relied on chart review for notes indicating at least a 50% reduction in bingeing and purging and the overeating question on the Hopkins Symptoms Checklist (SCL-90). Phenelzine was found to be statistically superior to both imipramine and placebo on both endpoints.

Posttraumatic Stress Disorder

There are two controlled trials, one negative and one positive, of MAOI use for PTSD. A 1987 study enrolled 13 patients with PTSD in a double-blind crossover trial comparing phenelzine and placebo. Response to phenelzine and placebo did not differ as measured by either the PTSD Scale or the Impact of Events Scale. Only 10 patients completed treatment, and each phase lasted only 5 weeks. A second trial compared phenelzine and imipramine to a placebo. At the 8-week follow-up, phenelzine showed superiority to both imipramine and placebo on the Impact of Events Scale, with 45%, 25%, and 5% reductions, respectively. Of note, PTSD was combat-related for all patients in this trial.

Social Anxiety Disorder

Until the introduction of SSRIs, MAOIs were widely considered to be the most effective medication for the treatment of social anxiety disorder. Four of five controlled trials have found positive results for the use of phenelzine in social anxiety disorder or social phobia.

In a 1992 study, 74 patients with social phobia were randomized to phenelzine, atenolol, or placebo for 8 weeks. Response rates for phenelzine (64%) were statistically higher than atenolol (30%) and placebo (20%).

Another 1992 study compared phenelzine and moclobemide to placebo over three 8-week phases. At week 8, both phenelzine and moclobemide showed statistically superior results compared to placebo on the Social Phobia Scale. During the next 8 weeks, mean scores continued to improve in both active treatments with no difference between the two with remittance rates of phenelzine (91%), moclobemide (82%), and placebo (43%). Finally, in the last phase, during which patients were randomized to continue active treatment or switch to placebo, relapse rates were significantly higher for those who switched from either group to placebo. Finally, the phenelzine group showed the highest side effect rates and had two severe adverse events, urinary retention and hypertensive crisis, that led to discontinuation of treatment.

The negative result came in a 1991 study that randomized 65 patients with social phobia to weekly cognitive-behavioral group therapy (CBGT) or alprazolam, phenelzine, or placebo with instruction for self-exposure to phobic stimuli for 12 weeks. On self-rating scales of social avoidance, no difference was found between groups. On the physician-rated Work and Social Disability Scale, phenelzine and alprazolam were found to be superior to placebo (CBGT was not included as it could not be blinded).

Two more studies compared phenelzine to CBGT. A 1998 study randomized 133 patients with social phobia to CBGT, phenelzine, placebo, or education-supportive group therapy (ES) over 12 weeks. Independent ratings on the Social Phobic Disorders Severity and Change Form found response rates to be statistically significant for phenelzine (65%) and CBGT (58%) compared to placebo (33%) and ES (27%). On self-report measures, phenelzine was found to be superior on four of six anxiety rating scales, with no difference between the other three groups. No serious adverse effects occurred. Researchers then continued responders in the phenelzine and CBGT group in 6 months of maintenance treatment and 6 months of treatment-free follow-up. Of 28 patients in the maintenance phase, relapse rates did not differ significantly between groups: phenelzine (21%) and CBGT (14%). In the follow-up phase, there was a nonsignificant trend toward less relapse in CBGT (0%) versus phenelzine (30%).

The same researchers followed up with a 2011 study that randomized 166 patients to one of 4 groups: phenelzine, CBGT, phenelzine plus CBGT, or placebo. Remission rates on the Liebowitz Social Anxiety Scale (LSAS) were measured at 12 weeks and found that combined therapy (59.4%) was superior to placebo (11.1%), which also could not be distinguished from monotherapy phenelzine (20.0%) or CBGT (20.6%). Insomnia, light-headedness, dry mouth, weight gain, constipation, anorgasmia, and nervousness rates were all higher in patients taking phenelzine.

A 2003 meta-analysis of pharmacologic treatment for social anxiety disorder combined results from the two phenelzine trials. The moderate effect size (0.66) was not statistically different from SSRIs, gabapentin, or clonazepam. The authors concluded that given SSRIs tolerability and safety, they should remain the first-line treatment for social anxiety disorder.

Panic Disorder and Agoraphobia

Early studies on MAOI treatment for panic disorder are difficult to interpret as they often clumped phobias together or examined mixed anxiety and depressive states. However, the result of these early studies was that clinicians regarded MAOIs as more effective antipanic agents than TCAs.

The first prospective study on “endogenous anxiety” (characterized by spontaneous panic attacks and agoraphobia) in 1980 compared imipramine and phenelzine to placebo. All three groups received supportive group therapy. At 12 weeks, results showed both active groups to be superior to placebo. The phenelzine group showed trends toward superiority on all measures and a statistically significant improvement on the Work and Social Disability Scale and Symptom Severity and Phobic Avoidance Scale.

A double-blind study compared four treatment conditions: phenelzine/exposure, phenelzine/no exposure, placebo/exposure, and placebo/no exposure in a group of 40 patients with agoraphobia and social phobia. Exposure, with or without phenelzine, was found to be superior in reducing anxiety symptoms compared to no exposure groups.

The Columbia group compared the response to phenelzine, imipramine, and placebo in 104 patients with atypical depression with and without spontaneous panic attacks. In patients with panic, the response at 6 weeks was significantly higher for phenelzine (77%) compared to imipramine (2%) and placebo (15%) as measured by Clinical Global Improvement. It is unclear from the data whether the response is related to decreased panic or depression, although stratified subgroup analysis suggests that response to phenelzine was independent of pretreatment anxiety levels.

Since the introduction in 1981 of panic disorder to DSM-III, there have been no published prospective trials on MAOIs. However, there are two studies that may provide some information.

A 1987 open-label study treated 40 patients with panic disorder or agoraphobia with phenelzine. Thirty-four of thirty-five patients who completed the study showed statistically significant responses on the Hamilton Rating for Anxiety. This study is limited by its open-label design, and its high response rate suggests an unaccounted-for confounder. Patients also showed significant adverse effects with postural hypotension (64.3%), insomnia (44.3%), dry mouth (42.9%), sexual dysfunction (27.3% men, 16.7% women), and weight gain (17.1%).

A 2010 study compared tranylcypromine 30 mg and 60 mg without placebo control in a group of patients with panic disorder and comorbid social anxiety disorder. Both doses reduced panic attacks by about 70% from baseline, without statistical significance between the two. However, the higher dose reduced social anxiety symptoms by double the lower dose. Again, adverse effect rates were significant: orthostatic dizziness (58.8%), insomnia (47.1%), decreased libido (41.2%), and somnolence (35.3%).

Overall, the evidence is mixed, without clear antipanic effects from phenelzine. Studies are difficult to interpret given the inclusion of comorbidities and mixed treatment modalities. Still, guidelines suggest MAOIs as third-or fourth-line agents for panic disorder.

FUTURE DIRECTIONS AND INVESTIGATIONAL DRUGS

Despite evidence for MAOI effectiveness in a number of psychiatric illnesses and consensus guidelines that continue to recommend their use as third- or fourth-line agents, MAOIs have fallen out of favor as SSRI use has increased. A population-based study in Ontario showed that from 1997 to 2007, only 348 older adults were started on an MAOI, accounting for only 1/500 older adults started on an antidepressant. A 2021 review of active clinical trials showed a few dozen active trials on MAOIs, mostly tranylcypromine and selegiline; a majority investigated their use in treatment-refractory depression or bipolar depression.

Although MAOIs have fallen out of favor with prescribing clinicians, research continues in the area. Recent clinical trials have focused on new indications for MAOIs, including negative symptoms in schizophrenia, amyotrophic lateral sclerosis, metastatic breast cancer, hair regrowth, prostate cancer, muscular dystrophies, tobacco cessation, and cocaine addictions. Though not covered in this chapter, selective MAO-B inhibitors continue to be investigated for use in PD, AD, and other neurodegenerative disorders.

FURTHER READINGS

Amsterdam JD, Bodkin JA. Selegiline transdermal system in the prevention of relapse of major depressive disorder: a 52-week, double-blind, placebo-substitution, parallel-group clinical trial. *J Clin Psychopharmacol*. 2006;26(6):579.

Atiwannapat P, Arden PC, Stewart JW. Monoamine oxidase inhibitors for various psychiatric disorders and conditions. *Psych Ann*. 2014;44(12):567.

Birkenhäger TK, van den Broek WW, Mulder PG, Bruijn JA, Moleman P. Efficacy and tolerability of tranylcypromine versus phenelzine: a double-blind study in antidepressant-refractory depressed inpatients. *J Clin Psychiatry*. 2004;65(11):1505.

Blanco C, Heimberg RG, Schneier FR, et al. A placebo-controlled trial of phenelzine, cognitive behavioral group therapy, and their combination for social anxiety disorder. *Arch Gen Psychiatry*. 2010;67(3):286.

Blanco C, Schneier FR, Schmidt A, et al. Pharmacological treatment of social anxiety disorder: a meta-analysis. *Depress Anxiety*. 2003;18(1):29.

Buigues J, Vallejo J. Therapeutic response to phenelzine in patients with panic disorder and agoraphobia with panic attacks. *J Clin Psychiatry*. 1987;48(2):55.

Carradori S, Petzer JP. Novel monoamine oxidase inhibitors: a patent review (2012–2014). *Expert Opin Ther Pat*. 2015;25(1):91.

Citrome L, Goldberg JF, Portland KB. Placing transdermal selegiline for major depressive disorder into clinical context: Number needed to treat, number needed to harm, and likelihood to be helped or harmed. *J Affect Disord*. 2013;151(2):409.

Davidson JR, McLeod MN, Turnbull CD, Miller RD. A comparison of phenelzine and imipramine in depressed inpatients. *J Clin Psychiatry*. 1981;42(10):395.

Gardner DM, Shulman KI, Walker SE, Tailor SA. The making of a user friendly MAOI diet. *J Clin Psychiatry*. 1996;57(3):99.

Gelernter CS, Uhde TW, Cimbolic P, et al. Cognitive-behavioral and pharmacological treatments of social phobia. A controlled study. *Arch Gen Psychiatry*. 1991;48(10):938.

Georgotas A, McCue RE, Cooper TB. A placebo-controlled comparison of nortriptyline and phenelzine in maintenance therapy of elderly depressed patients. *Arch Gen Psychiatry*. 1989;46(9):783.

Georgotas A, McCue RE, Friedman E, Cooper TB. Response of depressive symptoms to nortriptyline, phenelzine and placebo. *Br J Psychiatry*. 1987;151:102.

Gillman PK. Advances pertaining to the pharmacology and interactions of irreversible nonselective monoamine oxidase inhibitors. *J Clin Psychopharmacol*. 2011;31(1):66.

Halaris A, Piletz JE. Relevance of imidazoline receptors and agmatine to psychiatry: a decade of progress. *Ann N Y Acad Sci*. 2003;1009:1.

Heimberg RG, Liebowitz MR, Hope DA, et al. Cognitive behavioral group therapy vs phenelzine therapy for social phobia: 12-week outcome. *Arch Gen Psychiatry*. 1998;55(12):1133.

Henkel V, Mergl R, Allgaier AK, Kohnen R, Möller HJ, Hegerl U. Treatment of depression with atypical features: a meta-analytic approach. *Psychiatry Res*. 2006;141(1):89.

Himmelhoch JM, Fuchs CZ, Symons BJ. A double-blind study of tranylcypromine treatment of major anergic depression. *J Nerv Ment Dis*. 1982;170(10):628.

Himmelhoch JM, Thase ME, Mallinger AG, Houck P. Tranylcypromine versus imipramine in anergic bipolar depression. *Am J Psychiatry*. 1991;148(7):910.

Janicak PG, Pandey GN, Davis JM, Boshes R, Bresnahan D, Sharma R. Response of psychotic and nonpsychotic depression to phenelzine. *Am J Psychiatry*. 1988;145(1):93.

Jarrett RB, Schaffer M, McIntire D, Witt-Browder A, Kraft D, Risser RC. Treatment of atypical depression with cognitive therapy or phenelzine: a double-blind, placebo-controlled trial. *Arch Gen Psychiatry*. 1999;56(5):431.

Kennedy S, Piran N, Garfinkel PE. Isocarboxazid in the treatment of bulimia. *Am J Psychiatry*. 1986;143(11):1495.

Kosten TR, Frank JB, Dan E, McDougle CJ, Giller EL Jr. Pharmacotherapy for posttraumatic stress disorder using phenelzine or imipramine. *J Nerv Ment Dis*. 1991;179(6):366.

Krishnan KR. Revisiting monoamine oxidase inhibitors. *J Clin Psychiatry*. 2007;68(Suppl 8):35.

Krishnan V, Nestler EJ. The molecular neurobiology of depression. *Nature*. 2008;455(7215):894.

Liebowitz MR, Heimberg RG, Schneier FR, et al. Cognitive-behavioral group therapy versus phenelzine in social phobia: long-term outcome. *Depress Anxiety*. 1999;10(3):89.

Liebowitz MR, Schneier F, Campeas R, et al. Phenelzine vs atenolol in social phobia. A placebo-controlled comparison. *Arch Gen Psychiatry*. 1992;49(4):290.

López-Muñoz F, Alamo C, Juckel G, Assion HJ. Half a century of antidepressant drugs: on the clinical introduction of monoamine oxidase inhibitors, tricyclics, and tetracyclics. Part I: monoamine oxidase inhibitors. *J Clin Psychopharmacol*. 2007;27(6):555.

Mallinger AG, Frank E, Thase ME, et al. Revisiting the effectiveness of standard antidepressants in bipolar disorder: are monoamine oxidase inhibitors superior? *Psychopharmacol Bull*. 2009;42(2):64.

Mann JJ, Aarons SF, Wilner PJ, et al. A controlled study of the antidepressant efficacy and side effects of (l)-Deprenyl. *Arch Gen Psychiatry*. 1989;46(1):45.

McGrath PJ, Stewart JW, Fava M, et al. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR*D report. *Am J Psychiatry*. 2006;163(9):1531.

Meyer JH, Ginovart N, Boovariwala A, et al. Elevated monoamine oxidase a levels in the brain: an explanation for the monoamine imbalance of major depression. *Arch Gen Psychiatry*. 2006;63(11):1209.

Meyer JH, Wilson AA, Sagrati S, et al. Brain monoamine oxidase A binding in major depressive disorder: relationship to selective serotonin reuptake inhibitor treatment, recovery, and recurrence. *Arch Gen Psychiatry*. 2009;66(12):1304.

Middlefell R, Frost I, Egan GP, Eaton H. A report on the effects of phenelzine (nardil), a monoamine oxidase inhibitor, in depressed patients. *J Ment Sci*. 1960;106:1533–1538.

Mirowska-Guzel D, Balkowiec-Iskra E. The role of monoamine oxidase in humans and its metabolism. *Psych Ann*. 2014;44(11):495.

Nakash-Eisikovits O, Dierberger A, Westen D. A multidimensional meta-analysis of pharmacotherapy for bulimia nervosa: summarizing the range of outcomes in controlled clinical trials. *Harv Rev Psychiatry*. 2002;10(4):193.

Nolen WA, Haffmans PM, Bouvy PF, Duivenvoorden HJ. Monoamine oxidase inhibitors in resistant major depression. A double-blind comparison of brofaromine and tranylcypromine in patients resistant to tricyclic antidepressants. *J Affect Disord*. 1993;28(3):189.

Nolen WA, Kupka RW, Hellemann G, et al. Tranylcypromine vs. lamotrigine in the treatment of refractory bipolar depression: a failed but clinically useful study. *Acta Psychiatr Scand*. 2007;115(5):360.

Nolen WA, van de Putte JJ, Dijken WA, et al. Treatment strategy in depression. II. MAO inhibitors in depression resistant to cyclic antidepressants: two controlled crossover studies with tranylcypromine versus L-5-hydroxytryptophan and nomifensine. *Acta Psychiatr Scand*. 1988;78(6):676.

Pande AC, Birkett M, Fechner-Bates S, Haskett RF, Greden JF. Fluoxetine versus phenelzine in atypical depression. *Biol Psychiatry*. 1996;40(10):1017.

Pohl R, Berchou R, Rainey JM Jr. Tricyclic antidepressants and monoamine oxidase inhibitors in the treatment of agoraphobia. *J Clin Psychopharmacol*. 1982;2(6):399.

Quitkin FM, McGrath PJ, Stewart JW, et al. Atypical depression, panic attacks, and response to imipramine and phenelzine. A replication. *Arch Gen Psychiatry*. 1990;47(10):935.

Quitkin FM, Stewart JW, McGrath PJ, et al. Columbia atypical depression. A subgroup of depressives with better response to MAOI than to tricyclic antidepressants or placebo. *Br J Psychiatry Suppl*. 1993;(21):30.

Ramsay RR. Monoamine oxidases: the biochemistry of the proteins as targets in medicinal chemistry and drug discovery. *Curr Top Med Chem*. 2012;12(20):2189.

Rapaport MH. Dietary restrictions and drug interactions with monoamine oxidase inhibitors: the state of the art. *J Clin Psychiatry*. 2007;68(Suppl 8):42.

Rees L, Davies B. A controlled trial of phenelzine (“Nardil”) in the treatment of severe depressive illness. *J Ment Sci*. 1961;107:560.

Robinson DS, Lerfald SC, Bennett B, et al. Continuation and maintenance treatment of major depression with the monoamine oxidase inhibitor phenelzine: a double-blind placebo-controlled discontinuation study. *Psychopharmacol Bull*. 1991;27(1):31.

Rothschild R, Quitkin HM, Quitkin FM, et al. A double-blind placebo-controlled comparison of phenelzine and imipramine in the treatment of bulimia in atypical depressives. *Int J Eat Disord*. 1994;15(1):1.

Shestatzky M, Greenberg D, Lerer B. A controlled trial of phenelzine in posttraumatic stress disorder. *Psychiatry Res*. 1988;24(2):149.

Shulman KI, Fischer HD, Herrmann N, Huo CY, Anderson GM, Rochon PA. Current prescription patterns and safety profile of irreversible monoamine oxidase inhibitors: a population-based cohort study of older adults. *J Clin Psychiatry*. 2009;70(12):1681.

Shulman KI, Herrmann N, Walker SE. Current place of monoamine oxidase inhibitors in the treatment of depression. *CNS Drugs*. 2013;27(10):789.

Solyom C, Solyom L, LaPierre Y, Pecknold J, Morton L. Phenelzine and exposure in the treatment of phobias. *Biol Psychiatry*. 1981;16(3):239.

Stewart JW, McGrath PJ, Quitkin FM, et al. Chronic depression: response to placebo, imipramine, and phenelzine. *J Clin Psychopharmacol*. 1993;13(6):391.

Sunderland T, Cohen RM, Molchan S, et al. High-dose selegiline in treatment-resistant older depressive patients. *Arch Gen Psychiatry*. 1994;51(8):607.

Thase ME, Mallinger AG, McKnight D, Himmelhoch JM. Treatment of imipramine-resistant recurrent depression, IV: A double-blind crossover study of tranylcypromine for anergic bipolar depression. *Am J Psychiatry*. 1992;149(2):195.

Thase ME, Trivedi MH, Rush AJ. MAOIs in the contemporary treatment of depression. *Neuropsychopharmacology*. 2015;12(3):185.

Versiani M, Nardi AE, Mundim FD, Alves AB, Liebowitz MR, Amrein R. Pharmacotherapy of social phobia. A controlled study with moclobemide and phenelzine. *Br J Psychiatry*.

1992;161:353–360.

Walsh BT, Gladis M, Roose SP, Stewart JW, Stetner F, Glassman AH. Phenelzine vs placebo in 50 patients with bulimia. *Arch Gen Psychiatry*. 1988;45(5):471.

Youdim MB, Edmondson D, Tipton KF. The therapeutic potential of monoamine oxidase inhibitors. *Nat Rev Neurosci*. 2006;7(4):295.

33.6 Monoamine Modulators

33.6a Trazodone

JOHN T. VERNON, M.D., AND SUSAN G. KORNSTEIN, M.D.

Neuroscience-Based Nomenclature (NbN)	
Serotonin Multimodal	
Pharmacologic domain	Serotonin
Mechanism of action	Receptor antagonist (5-HT_{2A}) and partial agonist (5-HT_{1A})

HISTORY

Developed in Italy in 1966, trazodone (Desyrel) is considered the first of the SGAs, possessing a novel triazolopyridine chemical structure that distinguished it from the TCAs that preceded it. This structural uniqueness was associated with pharmacologic properties that also differentiated it from the TCAs in its performance on animal antidepressant screens, proposed mechanism of action, and adverse effect profile. Trazodone was first marketed in Europe in the 1970s and received FDA approval in 1981 for the treatment of major depression in the United States. Its improved safety and tolerability compared to TCAs was most likely responsible for trazodone's rapid rise in clinical use after its introduction. However, with the increased popularity of SSRI antidepressants, trazodone has fallen out of favor as a first-line treatment for depression in the United States. A once-a-day formulation of trazodone (Oleptro—trazodone/Contramid) was introduced by Labopharm Inc. in 2010 and approved by the FDA for the treatment of major depressive disorder; however, it was withdrawn from marketing in the United States by 2015. Other formulations include trazodone immediate-release liquid formulation, trazodone injectable solution, and trazodone delayed-release RP-AC.

CHEMISTRY

Trazodone is a triazolopyridine derivative, designated as 2-[3-[4-(3-chlorophenyl)-1-piperazinyl] propyl]-1,2,4-triazolo[4,3-a]pyridin-3(2H)-one hydrochloride. Its empirical formula is C19H22ClN5O•HCl, and its molecular structure is depicted in Figure 33.6a–1. Trazodone's molecular weight is 408.3. It is soluble in water, possessing a pKa of 6.7.

PHARMACOLOGY

Pharmacokinetics

Trazodone is absorbed efficiently by the human digestive tract after oral administration, reaching peak plasma concentration in about 1 hour. When consumed with food, the total amount of drug absorption is increased, maximum plasma concentration is decreased, and time to peak concentration is doubled. Since some adverse effects such as dizziness are related to peak plasma levels, this strategy has been shown to increase tolerability. Trazodone exhibits a linear pharmacokinetic profile in the dosage range of 50 to 150 mg twice daily. When administered twice daily, initial steady-state concentrations are achieved in approximately 36 hours, with restabilization at new levels in 12 hours after dosage adjustment. There may be up to a fourfold difference in steady-state plasma levels between individuals maintained on identical dosages. Steady-state AUC of trazodone IR 100 mg three times daily and Oleptro 300 mg daily were similar, although steady-state C_{max} and C_{min} after 1 week were not found to be equivalent.

Trazodone is approximately 90% bound to plasma proteins and distributes without selective localization in any tissue. Trazodone and its major metabolite, meta-chlorophenylpiperazine (mCPP), both cross the blood–brain barrier. Trazodone is also excreted into breast milk, with one study reporting that the concentration is about 14% of the maternal plasma concentration.

Trazodone is primarily metabolized via hydroxylation and oxidation by the hepatic cytochrome P450 (CYP) system. Several studies implicate the CYP 3A4 microsomal enzyme in the production of its major, active metabolite, mCPP, while reports on the role of CYP 2D6 are conflicting. The elimination of trazodone is biphasic, with half-lives of 3 to 6 hours and 5 to 9 hours, respectively. It is eliminated primarily via the kidneys as metabolites, with only about 1% excreted as the parent compound. The half-life of mCPP is about 14 hours. The half-life of Oleptro is reported at 10 hours.

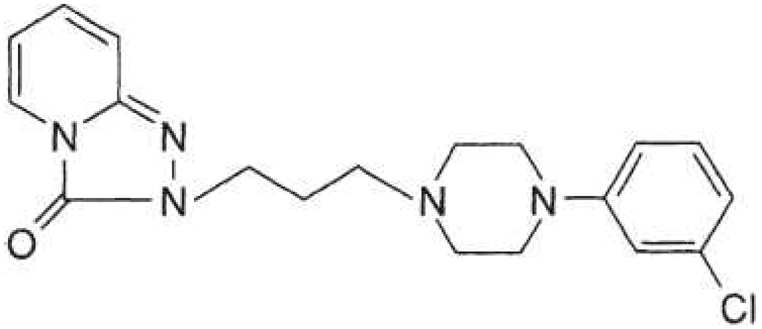

FIGURE 33.6a–1. Molecular structure of trazodone (Desyrel).

Pharmacodynamics

Trazodone's main pharmacologic effects are believed to be mediated through its actions at postsynaptic serotonin (5-hydroxytryptamine [5-HT]) 5-HT$_2$ receptors. Although it exhibits weak in vitro effects at the presynaptic serotonin transporter, which is responsible for the main therapeutic effects of the SSRIs, in vivo studies suggest that this is of negligible significance in humans. Trazodone is a potent 5-HT$_{2A}$ receptor antagonist, with 100-fold selectivity for the 5-HT$_2$ relative to the 5-HT$_1$ binding site. Chronic administration causes downregulation of 5-HT$_2$ as well as β-adrenergic receptors, similar to the actions of some other antidepressant compounds.

Trazodone's major metabolite, mCPP, is also active at serotonin receptors. It is an agonist at 5-HT$_{1B}$ and 5-HT$_{1C}$, and possibly, 5-HT$_{1A}$

and 5-HT_{1D} receptors. Research suggests that the parent compound, and not its metabolite, is responsible for the antidepressant effects. In fact, mCPP has been shown to be behaviorally activating and anxiogenic in experimental animals and human subjects.

Trazodone is an antagonist at peripheral α-adrenergic receptors, which is most likely responsible for its adverse effects of orthostatic hypotension and dizziness, dry mouth, and priapism. It has no significant activity at noradrenergic or dopaminergic receptors. Antagonism at acetylcholine and histamine type-1 (H_1) receptors is negligible. One in vitro study suggests that inhibition of T-type calcium channels may contribute to trazodone's psychiatric and neurologic effects.

There is no established consensus regarding the relationship between steady-state blood levels of trazodone and therapeutic response. One study in older adult subjects reported that a threshold concentration of 650 ng/mL predicts clinical efficacy, but studies in other populations produced conflicting results. An open-label, 3-week treatment study in a Japanese population of 26 adult patients with major depression reported a linear relationship between trazodone (but not mCPP) steady-state plasma concentration and both the final Montgomery Asberg Depression Rating Scale (MADRS) score and the percentage improvement at study completion. Also, the proportion of responders was significantly higher in the patients whose plasma concentration exceeded 714 ng/mL, consistent with the above study. No relationship between adverse effects and either trazodone or mCPP levels was seen. This finding suggests that, although not common in clinical practice, measuring trazodone levels to ensure that they exceed a threshold of 650 to 700 ng/mL may help improve outcomes in patients with inadequate response to the usual dosage regimen.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Major Depressive Disorder

The majority of clinical research on trazodone has focused on its use in the treatment of major depressive disorder. After a series of open trials suggesting the efficacy of trazodone in depressive syndromes, double-blind comparison trials of trazodone were carried out first in Europe and then in the United States. Some, but not all, of the studies included a placebo comparison arm, as some investigators considered it unethical to withhold active medication from seriously depressed patients. Comparator drugs included amitriptyline (Elavil), imipramine (Tofranil), doxepin (Sinequan), mianserin (Tolvon), dothiepin (Prothiaden), and diazepam (Valium). The studies were conducted in inpatient, outpatient, and general practice settings. Most, but not all, of these studies found trazodone to possess efficacy that is comparable to the TCAs. Several reviews and meta-analyses that attempted to overcome the limitations of the individual studies concluded that trazodone was indeed as efficacious as TCAs for the acute treatment of major depression. One such meta-analysis calculated the pooled odds ratio for response to trazodone compared to imipramine as 0.93 (95% confidence interval: 0.64 to 1.36), supporting the hypothesis of near-equivalent efficacy of the two medications. In addition, a 12-month, double-blind maintenance study found trazodone to be superior to imipramine on endpoint efficacy measures.

Later studies have compared trazodone to other newer antidepressants. Two double-blind, randomized, 6-week comparisons between trazodone and fluoxetine (Prozac) in 43 and 126 patients, respectively, reported equivalent response rates for the two medications, although trazodone demonstrated better HAM-D sleep subscale scores in the first study. One double-blind, placebo-controlled multicenter trial compared trazodone (50 to 300 mg) and venlafaxine (Effexor) (25 to 150 mg) in 225 patients in a 6-week acute treatment phase followed by a 1-year continuation phase for responders. Subjects in the venlafaxine group were more likely to respond sufficiently to enter the continuation phase and more likely to remain in the continuation phase. Trazodone proved more effective on measures of sleep, and venlafaxine proved more effective on measures of cognition and psychomotor slowing. Another 6-week, parallel, double-blind, randomized trial compared trazodone 50 to 300 mg to bupropion (Wellbutrin) 225 to 450 mg in 126 moderately to severely depressed outpatients, finding no significant differences in efficacy measures between the two treatments at the endpoint but greater improvement with trazodone at 7 days based on sleep measures. However, a 6-week, double-blind comparison of trazodone (150 to 450 mg per day) and mirtazapine (Remeron) (24 to 72 mg per day) in 200 inpatients with a *Diagnostic and Statistical Manual of Mental Disorders*, third edition (DSM-III) diagnosis of moderate to severe major depressive episode found mirtazapine to be superior to trazodone on most efficacy measures. Since the introduction of controlled-release trazodone in 2000, two small, 6-week, double-blind studies reported equal efficacy in response and remission rates between this and both paroxetine (Paxil) and sertraline (Zoloft). When Oleptro was studied in a 2009 randomized placebo-controlled trial, a statistically significant improvement in HAM-D 17 scores was demonstrated; antidepressant efficacy was independent of baseline insomnia or reduction in insomnia. A 2007 meta-analysis detected no significant differences between trazodone and its sister 5-HT$_2$ receptor antagonist, nefazodone (Serzone), compared with SSRIs for the treatment of major depression. A 2018 study with once-daily trazodone (Trazodone-XR-COAD) in 85 patients with moderate to severe depression showed efficacy based on reduction of MADRS and CGI scores.

Trazodone has been systematically studied in older adult depressed patients, as its low potential for anticholinergic and cardiovascular side effects suggested greater tolerability compared with tricyclics in this population. Trials comparing trazodone with various tricyclic and heterocyclic antidepressants found comparable efficacy and a generally lower side-effect burden for trazodone. The most frequently reported side effect in the older adults was dizziness, which may have been the result of orthostatic hypotension, expected from trazodone's pharmacologic profile. This symptom can be significant in the older adults for its potential to produce falls and, therefore, may limit the use of trazodone in those most sensitive to it, particularly if exacerbated by physical illness or concomitant medications such as antihypertensives. In addition, studies have compared antidepressants for their effects on cognitive performance in the older adults. One review of such studies found evidence of impaired ability to concentrate or sustain attention for sedating antidepressants, including trazodone, but little or no effects on memory with trazodone compared with tricyclics. In comparison, that review reported no cognitive adverse effects for MAOIs or SSRIs, suggesting that trazodone may not be the best choice for geriatric depression since the introduction of the SSRIs. Regarding the treatment of symptoms of depression in children, it should also be noted that a 2003 survey of 1,273 child and adolescent psychiatrists indicated that trazodone was the most frequently prescribed medication for insomnia for children with sleep disturbance associated with mood disorders despite very limited evidence for the efficacy and safety of trazodone for use in children.

During clinical studies of trazodone’s antidepressant properties, several double-blind trials compared the efficacy of single bedtime dosing and divided daily dosing. One smaller study compared single bedtime dosing with thrice-daily dosing in a 4-week trial of 56 patients with major depression. That study reported equivalent therapeutic efficacy and tolerability for the two dosing regimens; a trend for greater improvement in onset and duration of sleep was noted for the bedtime dose group, as well as more side-effect complaints in the thrice-daily dosing group. In a larger trial of 146 depressed patients, twice-daily dosing was the comparison with single bedtime dosing over a period of 6 weeks. That study similarly reported equivalent antidepressant efficacy for both regimens as well as improvements in sleep for both overall. During the early weeks of the study, the single bedtime dosing group reported significantly superior sleep, while the twice-daily dosing group reported significantly greater daytime fatigue.

In a systematic review and network meta-analysis evaluating the comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with MDD, trazodone was among the least effective medications (along with fluoxetine, fluvoxamine, and reboxetine) with OR 0.51 to 0.84; agomelatine, amitriptyline, escitalopram, mirtazapine, paroxetine, venlafaxine, and vortioxetine were comparatively better (OR 1.19 to 1.96).

Overall, trazodone is an effective treatment for major depression. There is no clear indication that its effectiveness differs from that of the TCAs or SSRIs. Some studies suggest that its tolerability, and perhaps its overall effectiveness, is maximized by limiting doses to the more moderate ranges (300 mg per day or less). Adverse effects of sedation can be minimized via bedtime dosing, which also utilizes its beneficial effects for early relief of depression-related insomnia.

Anxiety Disorders

Although most clinical trials with trazodone have been undertaken for major depression, many of those studies gave an early indication that trazodone possesses antianxiety as well as antidepressant effects, as evidenced by significant reductions in anxiety-related symptom scales. In addition, small, uncontrolled studies in patients with prominent anxiety symptoms or anxiety neurosis suggested its potential as an anxiolytic. However, few controlled trials of trazodone in samples meeting diagnostic criteria for anxiety disorders exist.

Generalized Anxiety Disorder. Two hundred and thirty patients with DSM-III GAD, in whom major depression and panic disorder were excluded, were treated with trazodone, imipramine, or diazepam in a randomized, double-blind, placebo-controlled 8-week trial. Although diazepam produced the greatest improvement in anxiety ratings during the first 2 weeks, both trazodone and imipramine attained comparable efficacy to diazepam in weeks 3 through 8, but only imipramine was significantly better than placebo at the endpoint. Among completers, 73%, 69%, and 66% of imipramine-, trazodone-, and diazepam-treated patients reported moderate to marked improvement, respectively, compared to only 47% of patients treated with placebo. The antidepressants tended to provide greater improvement in the psychic symptoms of worry, tension, and apprehension compared to diazepam.

Panic Disorder. Data regarding the efficacy of trazodone for panic disorder are limited and conflicting. A small, uncontrolled, single-blind 8-week trial of trazodone in 11 patients with DSM-III diagnoses of panic disorder or agoraphobia with panic attacks reported significant improvements in ratings of panic and generalized anxiety, phobias, avoidance, and depression. However, in a double-blind study of 74 panic disorder patients, comparing trazodone to imipramine and alprazolam (Xanax), only the latter two effectively reduced panic and anxiety symptoms, with a high dropout rate for trazodone-treated patients primarily due to nonresponse.

Obsessive-Compulsive Disorder. Similarly, the data for trazodone’s antiobsessional efficacy are mixed. Several case reports and small case series reported trazodone to be helpful for patients with obsessive-compulsive neurosis or disorder (OCD), including one report of nine patients who had failed to respond to clomipramine (Anafranil). However, a double-blind, placebo-controlled trial of trazodone in 21 patients with OCD reported no significant difference in OCD symptoms between those treated with trazodone (mean daily dose 235 mg) and those receiving a placebo.

Posttraumatic Stress Disorder. Several small trials of trazodone use in patients with PTSD have been published. One study retrospectively surveyed 74 chronic PTSD patients who had ever been prescribed trazodone for insomnia or nightmares. Of the 60 patients who had tolerated trazodone, 72% reported it to be helpful for nightmares, and 92% reported improvement in sleep. A correlation of 0.57 between these two measures suggests that trazodone’s beneficial effect on nightmares in PTSD may be partially independent of its sleep benefits. An open-label trial of trazodone in six patients with combat-related PTSD rated four of the six as much improved after 4 months of treatment. A 2011 literature review informing an algorithm for the treatment of PTSD proposed the use of low-dose trazodone following a trial of prazosin (Minipress) for nightmares and nighttime hyperarousal or for (persisting) initial insomnia (without nightmares) or after resolution of nightmares.

Adjustment Disorders

Two nearly identical studies of trazodone versus clorazepate (Tranxene) for the treatment of adjustment disorders in medically ill patients have been published. Both were randomized, double-blind, 28-day trials in patients meeting DSM-III-R criteria for adjustment disorder with anxious or depressed mood or mixed disturbance of emotion and conduct but without a significant history of more serious psychiatric disorders. One population consisted of 18 patients receiving treatment for breast cancer, and the other 21 HIV-positive subjects. The outcomes were similar, each reporting significant symptom benefits from both medications, with nonsignificant trends suggesting greater efficacy from trazodone.

Bulimia Nervosa

Like other serotonergic antidepressants, trazodone has been tested for the treatment of bulimia nervosa, with some success. Several case studies have reported benefits on a range of illness measures. In an open-label, flexible-dose study of 13 subjects diagnosed with DSM-III bulimia, trazodone produced a significant decrease in the number of bingeing and purging (vomiting) episodes, carbohydrate cravings, and urges to binge eat. Bingeing and purging episodes were reduced to zero in four of ten subjects available for analysis and decreased by more than 50% in two others after about 8 weeks on the medication. In a 6-week, placebo-controlled study of 42 women meeting DSM-III-R criteria for bulimia nervosa, trazodone was superior to placebo in decreasing the frequency of bingeing and purging. In both of these trials, adverse effects were relatively mild and consistent with other studies of trazodone (sedation and dizziness) and in no cases resulted in withdrawal from the study, despite some isolated case reports of delirium in bulimic patients treated with trazodone. No trials exist to date that compare trazodone to other agents, such as the SSRIs.

Schizophrenia

Although it has no significant effects on the dopamine system, trazodone's actions at 5-HT_2 receptors are similar to those of atypical antipsychotics, suggesting a potential for efficacy in psychosis. However, this has not been borne out in clinical studies. One research group tested the efficacy of trazodone in the management of chronic schizophrenia in two small double-blind, controlled trials of 20 patients each. The first reported that trazodone produced significantly less deterioration than placebo in the Excitement and Hallucinatory Behavior item scores of the Brief Psychiatric Rating Scale (BPRS) over a 12-week period. The second compared trazodone to chlorpromazine (Thorazine), finding that chlorpromazine was significantly superior to trazodone in total BPRS score and several individual factor scores, although trazodone did improve depression scores on the BPRS. Subsequently, another group examined the potential for adjunctive trazodone use to reduce negative symptoms in chronic, residual schizophrenia using a double-blind, placebo-controlled design. Those investigators found a modest (10% to 15%) but the beneficial effect for trazodone on two of three negative symptom scales, with only moderate clinical improvement seen in 3 of 26 trazodone-treated patients. Trazodone did not appear to worsen positive symptoms of schizophrenia. Of additional note, trazodone 100 mg has demonstrated some efficacy in reducing acute neuroleptic-induced akathisia in a small, placebo-controlled, double-blind crossover study in 2010.

Insomnia

From 1999 through 2010, trazodone was found to hold a stable (off-label) second position to the increased use of nonbenzodiazepine GABA-A omega-1 agonists as the most commonly prescribed medication for insomnia (from the National Health and Nutrition Examination Survey of 32,328 adults) despite limited availability of double-blind, placebo-controlled trials demonstrating clear efficacy and safety of trazodone for insomnia treatment. This reflects physician prescriptions of trazodone for the treatment of primary insomnia, depression/anxiety/withdrawal-related insomnia, and antidepressant-induced insomnia, especially for patients at higher risk of sedative/hypnotic/anxiolytic use disorders or patients with comorbid depression. While trazodone is not associated with physiologic dependence or withdrawal, a 2008 randomized, placebo-controlled trial of trazodone at 50 to 150 mg over the course of 12 weeks in patients with alcohol use disorder advised caution. The study found that trazodone was helpful for the treatment of insomnia during early recovery, with rates of relapse similar to placebo. However, at 6 months post detoxification, the prior trazodone-treatment group demonstrated an increased number of drinks per drinking day compared to placebo.

Trazodone is a 5-HT$_{2A}$ antagonist that increases slow-wave sleep, shortens sleep latency, improves sleep continuity and does not suppress REM sleep. In a 3-week, within-subjects randomized, double-blind trial, 16 patients with primary insomnia were given trazodone 50 mg or placebo 30 minutes before bedtime for 7 days and administered polysomnograms and MSLT on days 1 and 7. Trazodone produced small but significant impairments of short-term memory, verbal learning, equilibrium, and arm muscle endurance across time points. Compared to placebo, trazodone was associated with less middle insomnia, decreased stage 1 sleep, increased slow-wave sleep (by day 7), less daytime drowsiness (by day 7), and improved subjective experience of sleep by self-report. In another study of 14 patients with primary insomnia and Beck Depression Inventory (BDI) score >10 and 15 patients with primary insomnia and BDI score <10, trazodone CR 25 to 150 mg daily for 3 months improved sleep; however, for patients with elevated BDI scores, discontinuation of trazodone resulted in worsening sleep, while improvement in sleep quality was sustained for patients with lower BDI scores with trazodone. Compared to other off-label 5-HT$_{1/2A}$ antagonists, such as the atypical antipsychotic quetiapine, trazodone may have favorable efficacy with longer mean total sleep time hours and fewer nighttime awakenings (although more frequent GI upset) with trazodone compared to quetiapine based on a nonrandomized, unblinded 2013 study. A 2018 meta-analysis of 429 patients in seven randomized controlled trials (quality assessment low to moderate) between 1994 and 2014 indicated that trazodone did not improve sleep efficiency (although one trial showed improved perceived sleep quality) but did improve the number of awakenings (without improvement in sleep latency, total sleep time, or waking time after sleep onset) with no difference from placebo in discontinuation for adverse events or tolerability/acceptability measures. Trazodone has also been applied with some efficacy in the treatment of insomnia in the postpartum period. In 54 postpartum patients with insomnia during the third trimester of pregnancy but without depression, EPDS scores were lower in patients treated with trazodone 50 mg or diphenhydramine 25 mg compared to placebo, with no significant difference between the two active treatment groups.

Pain

Trazodone is among the various antidepressants that have been examined for their analgesic potential in a wide variety of chronic pain syndromes. Two case series of open-label trazodone trials for painful diabetic neuropathy, with sample sizes of 6 and 31 patients, respectively, reported fairly dramatic symptom relief in the majority of patients. Trazodone has also been reported to be beneficial in a double-blind, comparative study with amitriptyline for deafferentation pain and in a double-blind, placebo-controlled study of pediatric migraine prophylaxis, although two systematic literature reviews question these latter findings. Two uncontrolled open-label studies in 2010 and 2011 suggest that trazodone may help with aspects of fibromyalgia. In one study, pain interference with daily activities, sleep, and depression were improved without direct mitigation of pain intensity. In the other study, pregabalin (Lyrica) added to trazodone reduced overall fibromyalgia severity, depression, pain interference with daily activities, and pain intensity. Trazodone did not prove to be effective in double-blind, placebo-controlled trials for fibromyalgia, chronic low-back pain, dysesthetic pain following traumatic myelopathy, or rheumatoid arthritis. Thus, little evidence exists supporting the use of trazodone in most chronic pain syndromes.

Behavioral Disturbances

Trazodone and other serotonergic medications have been used in place of sedative-hypnotics and antipsychotics to manage aggression and other behavioral disturbances in certain patient populations, such as children with disruptive behavioral disorders and geriatric patients with cognitive disorders. It is postulated that trazodone targets disturbed serotonergic brain pathways that are believed to underlie such behaviors.

Childhood Disruptive Behavioral Disorders. One uncontrolled open trial studied the effectiveness of trazodone in the management of impulsive and aggressive behaviors in an inpatient group of 22 children diagnosed with DSM-III-R disruptive behavioral disorders (ADHD, conduct disorder, and oppositional-defiant disorder) comorbid with other psychiatric symptomatology. These children, aged 5 to 12, had been unresponsive to other inpatient and outpatient therapeutic interventions, including an average of 4.7 documented medication trials. Thrice-daily dosing of trazodone at an average dose of about 5 mg/kg per day for a duration of 3 to 4 weeks resulted in clinical improvement in 65% of the children, with most maintaining this benefit (per parental report) at 3 to 14 months follow-up. The medication was well tolerated by the majority of the children, with transient orthostatic hypotension being the most frequently noted adverse effect (50%). A larger (also uncontrolled) study examining the connection between sleep and behavioral disturbances in 51 young children with opsoclonus-myoclonus syndrome reported that trazodone improved sleep in 95% and reduced rage attacks in 33% of those children treated.

Neurocognitive Disorders. Several open-label case series and double-blind, controlled trials have been published examining the effect of trazodone on agitation in older adult patients with neurocognitive disorders. These case series, typically small and with a mixed population of older adult patients with various neurocognitive disorders, overall reported beneficial effects of trazodone. A Cochrane Database Review of the existing controlled studies found similarity in efficacy and tolerability of trazodone to placebo and haloperidol as a treatment of behavioral disturbances of neurocognitive disorders, and subsequently, the review concluded that trazodone might improve some measures of agitation and psychosis for some patients with neurocognitive disorders. Some studies have examined the potential for trazodone to decrease the risk of neurocognitive disorder; a 2019 study concluded that trazodone inhibits tau amyloidogenesis in vitro, although another 2019 U.K. population cohort study found no evidence of reduced risk of neurocognitive disorder in 4,596 primary care patients prescribed trazodone compared to 22,980 patients taking other antidepressants.

Sexual Dysfunction

Trazodone has been associated with an increased risk of priapism and also has been the subject of several case reports of enhanced sexual desire and functioning in both sexes. These findings have led researchers to evaluate its potential benefit for treating disorders of sexual function, particularly male erectile dysfunction. In general, studies of treatments for sexual disorders in women have been disproportionately sparse; however, trazodone is being studied in combination with bupropion for the treatment of female hypoactive sexual desire disorder; a phase 2a trial completed in 2019 showed favorable response and tolerability compared to bupropion alone. In 2009, one small open-label preliminary study of SSRI-induced sexual dysfunction in men and women with depressive/anxiety/somatic symptom disorders found significant improvements in both men and women for all domains of sexual function and satisfaction measured. In this study, the improvement was additionally noted to be independent of improvement in measures of depressive/anxiety symptoms. In vitro, trazodone has been shown to have a high affinity for human penile α_1 receptors and moderate affinity for α_2-adrenergic receptors, lending experimental support to this endeavor.

Trazodone reportedly improved objectively assessed erectile measures in several small, uncontrolled case series of patients with impotence. However, double-blind, placebo-controlled trials of trazodone use for erectile dysfunction present different results. Three separate double-blind, placebo-controlled trials, the first in 69 patients taking 150 mg daily for 4 weeks, the second in 51 patients with 50 mg daily dosing for 3 months, and the third in 34 patients taking 200 mg daily for 4 weeks, all failed to find significant differences between active medication and placebo. The results of a review and meta-analysis of the literature on trazodone treatment for erectile dysfunction reported a nonsignificant trend of improvement with trazodone, particularly at higher doses, with a subgroup analysis favoring benefit if the dysfunction is psychogenic in origin.

Other Disorders

Trazodone has been studied as a potential treatment for a wide range of other psychiatric and medical syndromes, mostly via case reports or open-label trials. These include dysthymia, depression following stroke, tic disorders, chest pain due to esophageal dysmotility, sleep apnea, menopausal symptoms, benzodiazepine, alcohol, and opiate use disorders or withdrawal. Although varying degrees of efficacy have been reported, the overall lack of data, particularly from trials with more stringent study methodology, prevents one from drawing any firm conclusions regarding trazodone’s potential use in these conditions.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

In both premarketing trials and postmarketing surveillance, the adverse effects seen with trazodone therapy have generally been benign, mostly limited to dry mouth, headache, nausea, drowsiness, fatigue, and dizziness, the latter in part secondary to orthostatic hypotension. Unlike TCAs, trazodone is virtually devoid of anticholinergic side effects, including major cardiac conduction abnormalities, thus explaining its benign profile following overdose. The adverse reaction for which trazodone is most widely recognized, although infrequent, is priapism.

Cardiovascular System

Trazodone appears to have a relatively benign cardiovascular profile. The most troublesome cardiovascular adverse effect is postural hypotension, which can be severe enough to produce syncope in susceptible patients, particularly the older adults or those with concurrent medical illness or taking other medications with cardiovascular effects. Adjustments in the dosage or dosing intervals of trazodone or concurrent medications, together with precautions regarding postural changes, can usually mitigate this problem if trazodone use is otherwise warranted.

No consistent effects of trazodone on cardiac contractility or the interventricular conduction system in humans or animals have been noted, except for possible nonspecific lengthening of the PR interval. However, several case reports of cardiac abnormalities associated with trazodone use exist in the literature, including ventricular arrhythmias, QT prolongation, and complete heart block. Most, but not all, of these effects were seen either associated with trazodone overdose or in individuals with pre-existing heart disease, suggesting caution in this latter category of patients.

Genitourinary System

The most concerning side effect attributed to trazodone is priapism, a painful unremitting erection that requires emergency medical attention. It can occur as a complication of various medical illnesses such as sickle cell disease, but a majority of cases have implicated medication use. The α-adrenergic receptor antagonists, such as trazodone, chlorpromazine, and prazosin, are hypothesized to induce priapism by inhibiting sympathetically controlled detumescence. The incidence of priapism during trazodone therapy is estimated at 1 in 1,000 to 1 in 10,000 cases. Although it is most likely to occur within the first 28 days of initiating treatment with trazodone, it has occurred as late as 18 months into treatment and at a wide range of therapeutic dosages. A past history of prolonged erections appears to be a general risk factor for priapism. At least two cases of clitoral priapism associated with trazodone use have also been reported in the literature.

All patients on trazodone therapy need to be warned of the risk and medical significance of priapism, as patients may be reticent to acknowledge it and seek treatment. Once priapism develops, it is considered a urologic emergency requiring immediate medical intervention. Patients should be advised to discontinue trazodone and seek emergency care in the event of an erection lasting more than 1 hour. The likelihood of morbid complications of impotence, cavernosal fibrosis, and gangrene is directly related to a delay in initiating treatment. Various medical interventions (ice packs, oral adrenergic medications) have produced inconsistent results but are usually attempted before more definitive surgical procedures are instituted. One of two surgical interventions (intracavernosal irrigation with an α-adrenergic agonist or shunting between the corpus cavernosa and spongiosum) will usually produce detumescence, but impotence may still be the outcome in as many as half of these cases.

Central Nervous System

Trazodone frequently produces CNS side effects of drowsiness (20% to 40%) and dizziness or light-headedness (20% to 30%), particularly early in treatment. Drowsiness has been the side effect most frequently leading to discontinuation, although it can be a welcome benefit for those patients with severe insomnia. Headache, fatigue, nervousness, and confusion have lower incidence. Isolated cases of delirium induced by trazodone have been reported in patients with pre-existing cerebral or metabolic abnormalities.

Trazodone shows little, if any, risk for lowering the seizure threshold in patients. There have been isolated case reports of increased seizure activity associated with trazodone use, although these are far less prevalent than with TCAs. In patients with a known history of seizure disorder, it is prudent to initiate therapy with trazodone at lower doses and titrate the dose more gradually.

As is the case with other antidepressants, trazodone has been reported to precipitate manic episodes in patients with diagnoses of both major depressive disorder and bipolar disorder. Trazodone's potential for increasing the risk for hypomania or mania appears to be lower than that of TCAs. Finally, trazodone also carries the black box warning (applied to all antidepressant medications) for increased risk of suicidal ideation, especially in children, adolescents, and young adults.

THERAPEUTIC INDICATIONS

Trazodone has an FDA indication for major depressive disorder, and research supports its general efficacy and tolerability for this indication. Currently, the most common off-label use of trazodone is for the treatment of insomnia, particularly insomnia associated with the use of other antidepressants such as the SSRIs, for which it remains a reasonable option for some patients. For other conditions, as described above, the findings are tentative and would require replication in larger samples before treatment recommendations can be made.

PRECAUTIONS AND ADVERSE REACTIONS

The most common adverse effect reported by patients using trazodone is drowsiness. Patients should be warned that this side effect may be exacerbated by alcohol or other concomitant medications, thus impairing their ability to safely engage in potentially hazardous tasks, such as driving. Orthostatic hypotension resulting in syncope can occur, particularly in the older adults or medically ill, and may cause significant injury through falls. Adjusting the dosage of trazodone or concurrent medications and ingesting trazodone with meals may diminish this effect, but continued caution should be exercised when prescribing trazodone to older adult patients. There are no reported withdrawal or discontinuation syndromes associated with trazodone. A 2014 review found some cases of dose-independent trazodone-induced liver injury (hepatocellular by immuno-allergic mediation), with one resulting in a liver transplant. A 2011 meta-analysis indicated that trazodone was correlated with higher all-cause mortality compared to SSRIs and TCAs in the treatment of elder depressed patients.

Priapism, although infrequent, is the side effect that is most concerning due to its potential for producing serious, irreversible injury. All patients should be strongly cautioned to discontinue trazodone and seek immediate medical attention in the event of prolonged or painful erections.

There is no evidence that trazodone possesses mutagenic or carcinogenic potential based on studies in rats receiving doses up to 300 mg/kg for 18 months. Trazodone is designated pregnancy category C, as it has produced fetal anomalies in animal studies at 15 to 50 times the maximum human dose, but it has not been adequately evaluated in human pregnancy. One multicenter study compared the pregnancy outcomes of 147 women who had exposure to either trazodone (n = 58) or the related phenylpiperazine antidepressant, nefazodone (n = 89), with the outcomes from women exposed to other antidepressants (n = 147) or nonteratogenic medications (n = 147). They found no statistically significant differences among the three groups for any of the endpoints of interest, including miscarriage rate, rate of major malformations, gestational age, and birth weight. As with all medications that have an unclear potential for fetal effects, use in pregnancy should only be considered if the potential benefit outweighs the risk. Trazodone is excreted into human breast milk, with one study in six lactating women reporting a milk/plasma ratio of about 15%, indicating that exposure to nursing infants would be low.

Trazodone has a very high therapeutic index, with an LD_{50} of around 500 mg/kg in experimental animals. Data from hundreds of documented cases of overdose support its safety in all but a small fraction of cases, with the majority of those involving other concurrent medications. The main symptoms that occur when trazodone is taken in overdose are excessive drowsiness and vomiting, although isolated cases of seizures, hyponatremia, priapism, respiratory arrest, QT prolongation, cardiac arrest, and coma have been reported. One postmortem toxicologic review of trazodone-related deaths concluded that trazodone played a primary role in the deaths by massive overdose of only three individuals who attained blood concentrations above 9 mg/L.

DRUG INTERACTIONS

Significant changes in prothrombin time (PT) and international normalized ratio (INR) in patients receiving warfarin (Coumadin) therapy and increased levels of digoxin (Lanoxin) and phenytoin (Dilantin) have been reported in isolated cases of patients receiving trazodone concurrently, most likely due to interactions via trazodone's high binding to plasma proteins. An in vitro study reported that the metabolism of trazodone to its principal metabolite (mCPP) via CYP3A-mediated clearance was significantly inhibited by ketoconazole (Nizoral), ritonavir (Norvir), and indinavir (Crixivan), suggesting caution in prescribing trazodone to individuals treated with CYP 3A4 inhibitors, including HIV-infected individuals who are commonly prescribed these medications. One patient taking fluoxetine and trazodone developed sedation and an unsteady gait, with the finding of a 30% increase in trazodone levels. However, this was an isolated report, as most patients tolerate this combination without difficulty.

As discussed earlier, trazodone may potentiate the sedative or hypotensive effects of other pharmacologic agents, requiring caution when used in such combinations. In combination with MAOIs to treat insomnia, trazodone appears to be safe. No cases of hypertensive crisis have been reported, and one reported case of nocturnal myoclonus resolved after discontinuation of trazodone. There have been several isolated reports of serotonin syndrome associated with trazodone use. Most of these occurred when trazodone was prescribed in combination with a variety of serotonergic psychotropic medications, including buspirone (BuSpar), fluoxetine, sertraline, paroxetine, venlafaxine, nefazodone, and amitriptyline with lithium (Eskalith). One case occurred when trazodone alone was prescribed to a 73-year-old patient who was later suspected of suffering from Lewy body dementia.

LABORATORY INTERFERENCE

Trazodone generally exhibits no interference with urinalysis, spectrometry, or other laboratory tests. There is one isolated report of trazodone-related decreases in hematocrit, hemoglobin, red blood cell count, serum cholesterol, calcium, and albumin during a 6-week comparison trial with fluoxetine, with the development of clinically insignificant pseudoanemia in 36% of the trazodone-treated and 20% of the fluoxetine-treated groups, respectively. Trazodone has also been reported to result in a false-positive urine drug screen for amphetamine.

DOSAGE AND ADMINISTRATION

The manufacturer's package insert recommends initiating trazodone at 150 mg per day in divided doses (after meals), increasing by 50 mg per day in intervals of 3 to 4 days up to a maximum dose of 400 mg per day in outpatients and 600 mg per day in inpatients. Elder patients should use a starting dose of 100 mg daily. There are not yet approved dosage recommendations for children or adolescents.

Trials of trazodone for the treatment of major depression suggest that the majority of patients begin to respond within the first 2 weeks of therapy, although this finding was in the context of objective, controlled studies, many with higher initial dosages. However, patients with significant insomnia or agitation may show an early response to trazodone's sedating effects. For the treatment of primary or medication-related insomnia, doses in the 50- to the 100-mg range are usually sufficient.

Trazodone is available in scored 50- and 100-mg tablets and in 150- and 300-mg tablets that can be divided into halves or thirds for convenient dosing adjustment and administration. Oleptro is available in 150- and 300-mg tablets.

FURTHER READINGS

Akbari V, Ghobadi S, Mohammadi S. The antidepressant drug; trazodone inhibits Tau amyloidogenesis: Prospects for prophylaxis and treatment of AD. *Archives of Biochemistry and Biophysics*. 2019;679:108218.

Ansari A. The efficacy of newer antidepressants in the treatment of chronic pain: A review of current literature. *Harv Rev Psychiatry*. 2000;7(5):257–277.

Bajor LA, Ticlea AN, Osser DN. The psychopharmacology algorithm project at the Harvard South Shore Program: An update on posttraumatic stress disorder. *Harv Rev Psychiatry*. 2011;19(5):240–258.

Belkin ZR, Krapf JM, Goldstein AT. Drugs in early clinical development for the treatment of female sexual dysfunction. *Expert Opin Investig Drugs*. 2014;24:1–9.

*Bossini L, Coluccia A, Casolaro I. Off-label trazodone prescription: evidence, benefits, and risks. *Current Pharmaceutical Design*. 2015;21:3343–3351.

Brahm NC, Yeager LL, Fox MD, Farmer KC, Palmer TA. Commonly prescribed medications and potential false-positive urine drug screens. *Am J Health Syst Pharm*. 2010;67(16):1344–1350.

Brauer R, Lau WCY, Hayes JF. Trazodone use and risk of dementia: a population-based cohort study. *PLoS Med.* 2019;16(2):e1002728.

Bruni O, Angriman M, Melegari MG. Pharmacotherapeutic management of sleep disorders in children with neurodevelopmental disorders. *Expert Opinion on Pharmacotherapy*. 2019;20(18):2257–2271.

Bryant SG, Hokanson JA, Brown CS. A drug utilization review of prescribing patterns for trazodone versus amitriptyline. *J Clin Psychiatry*. 1990;51:27–29.

Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391:1357–1366.

Coupland C, Dhiman P, Morriss R, Arthur A, Barton G, Hippisley-Cox J. Antidepressant use and risk of adverse outcomes in older people: population based cohort study. *BMJ*. 2011;343:d4551.

Cuomo A, Ballerini A, Bruni AC. Clinical guidance for the use of trazodone in major depressive disorder and concomitant conditions: pharmacology and clinical practice. *Rivista di Psichiatria*. 2019;54(4):137–149.

De Wit S, Cremers L, Hirsch D, et al. Efficacy and safety of trazodone versus clorazepate in the treatment of HIV-positive subjects with adjustment disorders: a pilot study. *J Int Med Res*. 1999;27(5):223–232.

*Fagiolini A, Comandini A, Dell'Osso MC, Kasper S. Rediscovering trazodone for the treatment of major depressive disorder. *CNS Drugs*. 2012;26:1033–1049.

Herr J, Caspi A. Oleptro™ (trazodone hydrochloride) extended-release tablets. *P T*. 2011;36(2):2–18.

Khazaie H, Ghadamia MR, Knight DC, Emamiana F, Tahmasian M. Insomnia treatment in the third trimester of pregnancy reduces postpartum depression symptoms: A randomized clinical trial. *Psychiatry Res*. 2013;210(3):901–905.

Martinez MA, Ballesteros S, Sanchez de la Torre C, Almarza E. Investigation of a fatality due to trazodone poisoning: Case report and literature review. *J Anal Toxicol*. 2005;29(4):262–268.

*Mittur A. Trazodone: properties and utility in multiple disorders. *Expert Rev Clin Pharmacol*. 2011;4(2):181–196.

Patten SB. The comparative efficacy of trazodone and imipramine in the treatment of depression. *CMAJ*. 1992;146(7):1177–1182.

Pyke RE, Clayton AH. Dose-finding study of lorexys for hypoactive sexual desire disorder in premenopausal women. *J Sex Med*. 2019;16(12):1885–1894.

Razavi D, Kormoss N, Collard A, Farvacques C, Delvaux N. Comparative study of the efficacy and safety of trazodone versus clorazepate in the treatment of adjustment disorders in cancer patients: a pilot study. *J Int Med Res*. 1999;27(6):264–272.

*Schatzberg AF. Trazodone A 5-year review of antidepressant efficacy. *Psychopathology*. 1987;20(Suppl 1):48–56.

Seitz DP, Adunuri N, Gill SS, et al. Antidepressants for agitation and psychosis in dementia. *Cochrane Database Syst Rev*. 2011;16(2):CD008191.

Thompson JW Jr, Ware MR, Blashfield RK. Psychotropic medication and priapism: A comprehensive review. *J Clin Psychiatry*. 1990;51(10):430–433.

Victor S, Ryan SW. Drugs for preventing migraine headaches in children. *Cochrane Database Syst Rev*. 2003;(4):CD002761.

Voican CS, Corruble E, Naveau S, Perlemuter G. Antidepressant-induced liver injury: a review for clinicians. *Am J Psychiatry*. 2014;171(4):404–415.

Wichniak A, Wierzbicka A, Jernajczyk W. Sleep and antidepressant treatment. *Curr Pharm Des*. 2012;18:5802–5817.

*Workman EA, Short DD. Atypical antidepressants versus imipramine in the treatment of major depression: A meta-analysis. *J Clin Psychiatry*. 1993;54(1):5–12.

Yi X-Y, Ni S-F, Ghadami MR, et al. Trazodone for the treatment of insomnia: a meta-analysis of randomized placebo-controlled trials. *Sleep Medicine*. 2018;45:25–32.

Zubieta JK, Alessi NE. Acute and chronic administration of trazodone in the treatment of disruptive behavior disorders in children. *J Clin Psychopharmacol*. 1992;12(5):346–351.

33.6b Nefazodone

AMIR A. KHAN, M.D., AND SUSAN G. KORNSTEIN, M.D.

Neuroscience-Based Nomenclature (NbN)

Serotonin Modulator

Pharmacologic domain	Serotonin
Mechanism of action	Receptor antagonist (5-HT$_{2A}$) and Partial Agonist (5-HT$_{1A}$)

Nefazodone (Serzone) is a serotonin modulator developed to improve the side-effect profile of the analogous, second-generation antidepressant trazodone (Desyrel) by using receptor-binding techniques. It has unique effects on the serotonergic system that are distinct from other antidepressants. It became available in the United States in early 1995. Following the issuance of a black box warning regarding rare but potentially life-threatening hepatotoxicity and the availability of generic nefazodone, the original manufacturer of Serzone (Bristol–Myers Squibb) discontinued the production of Serzone in June 2004, citing declining market share as a primary reason. Generic nefazodone remains available in the U.S. market, although its clinical utilization continues to decline. Sales of nefazodone in Canada were discontinued in 2003.

Nefazodone is a phenylpiperazine compound with some structural similarity to trazodone. It is structurally unrelated to other antidepressants. Nefazodone's mechanism of action is attributed to its antagonism of the 5-hydroxytryptamine, type 2A (5-HT$_{2A}$) receptor and inhibition of the serotonin transporter. It has a limited affinity for the α_1-adrenergic receptor, hence the decreased incidence of postural hypotension compared with trazodone.

CHEMISTRY

Nefazodone is a synthetically derived phenylpiperazine antidepressant. The chemical name for nefazodone hydrochloride is 2-[3-[4-(3-chlorophenyl)-1-piperazinyl]propyl]-5-ethyl-2,4-dihydro-4-(2-phenoxyethyl)-3H-1,2,4-triazol-3-one monohydrochloride. The molecular formula is $C_{25}H_{32}ClN_5O_2$•HCl with a molecular weight of 506.5 (Fig. 33.6b–1).

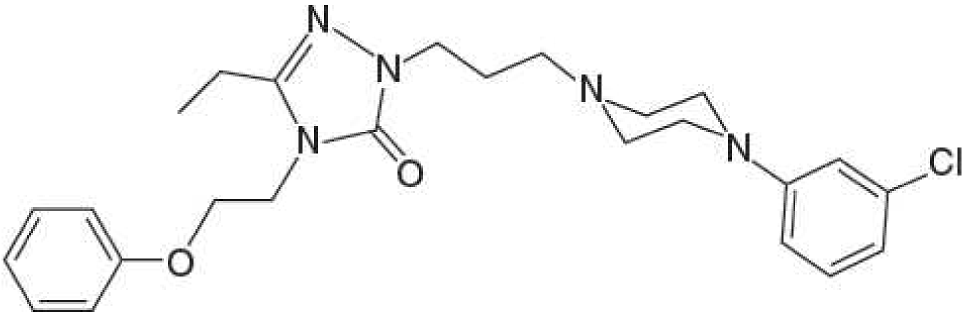

FIGURE 33.6b–1. Molecular structure of nefazodone (Serzone).

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Absorption. Nefazodone is rapidly and completely absorbed after oral intake. The absolute bioavailability, however, is low at about 20% and variable because of its extensive metabolism. The presence of food in the GI tract delays absorption and further reduces bioavailability. The clinical significance of this finding is unclear and probably not important. Peak plasma concentration occurs in about 1 hour.

Distribution. Nefazodone is widely distributed in body tissues, including the CNS. The volume of distribution for nefazodone is 0.22 to 0.87 L/kg. It readily crosses the blood–brain barrier. Nefazodone is extensively (>99%) but loosely bound to plasma proteins. It does not tend to displace other protein-bound drugs when studied in vitro, but this property in vivo is not well studied. It does decrease the protein binding of haloperidol by 5%, which is unlikely to be of any clinical importance.

Metabolism and Elimination. Nefazodone is extensively metabolized by n-dealkylation and aliphatic and aromatic hydroxylation. Less than 1% of the drug is excreted unchanged in the urine. Nefazodone’s three major pharmacologically active metabolites are hydroxynefazodone, meta-chlorophenylpiperazine (mCPP), and triazole-dione. Both nefazodone and hydroxynefazodone have a relatively short half-life (1.5 to 4 hours) and exhibit nonlinear kinetics. Compared to a single dose, multiple dosing, as well as dosage increases, produce a disproportional increase in the maximum concentration ($C_{\max}$) of the drug. This property signifies greater-than-predicted accumulation of the drug and its hydroxy metabolite with multiple dosing. Nefazodone and its metabolites attain steady-state concentration within 4 to 5 days of initiation of BID dosing or changes in previous regimens.

The pharmacologic profile of hydroxynefazodone is similar to that of nefazodone. mCPP is also somewhat similar to nefazodone, with the notable difference of possessing agonist activity at some serotonergic receptor subtypes. It has a half-life of 4 to 8 hours. The actions of triazole-dione are not yet well characterized. However, it has a longer half-life of 18 hours.

Nefazodone should not be used in patients with hepatic disease. Pharmacokinetic studies have shown reduced drug clearance in patients with hepatic cirrhosis; a dose reduction should be considered in such patients, as the steady-state concentrations of nefazodone and hydroxynefazodone are 25% higher than in normal controls. Exposure to mCPP is about two- to threefold higher in patients with hepatic cirrhosis. Patients with renal dysfunction do not show a significantly different pharmacokinetic profile. Older adult patients show a 10% to 20% increase in steady-state plasma levels; therefore, the starting dose for older patients, particularly women, should be half the usual dose. Nefazodone does not utilize the CYP 2D6 metabolic pathway; hence, dosage adjustment for poor metabolizers is not required.

Pharmacodynamics

Nefazodone has a multimodal, complex mechanism of action. It inhibits neuronal uptake of serotonin and norepinephrine; however, its antidepressant effects are presumably exerted only through the serotonergic system. Presynaptically, it inhibits 5-HT reuptake, which increases the 5-HT level within the synapse and results in increased availability of serotonin. Postsynaptically, it occupies 5-HT_{2A} receptors as an antagonist, hence downregulating the receptor. The latter property is believed to be responsible for the antidepressant action of nefazodone. Its ability to block 5-HT_{2A} receptors also provides a potential mechanism for the low likelihood of treatment-related activation and sexual dysfunction.

Hydroxynefazodone has an almost identical pharmacologic profile to the parent compound and probably contributes to the antidepressant action of nefazodone. mCPP is a minor metabolite, but it appears to have important pharmacologic properties. It binds potentially to all 5-HT_1, 5-HT_2, and 5-HT_3 receptors. It acts as an agonist at the 5-HT_{1A}, 5-HT_{2C}, and 5-HT_3 receptors and as a mixed agonist and antagonist at the 5-HT_{2A} site. It is believed to possess anxiogenic properties. Animal studies, as well as the direct administration of mCPP to patients with anxiety disorders and normal controls, generally support this finding.

Nefazodone has weak α_1-adrenergic antagonism that may cause postural hypotension, though this effect is much less pronounced compared to trazodone. Nefazodone does not have significant affinity for α_2-adrenergic, β-adrenergic, cholinergic, dopaminergic, 5-HT_{1A}, or benzodiazepine receptors. This receptor affinity profile offers distinct clinical advantages with regard to minimizing potential side effects.

THERAPEUTIC INDICATIONS

Nefazodone was approved for the treatment of major depression in the United States in 1994. This approval was granted based on the results of five well-controlled premarketing clinical trials. Since its approval, many small and large-scale studies, including both inpatient and outpatient populations, have documented its efficacy in the treatment of depression.

Major Depression

Nefazodone was evaluated in four phase III, double-blind, active (imipramine) and placebo-controlled trials to study its efficacy and safety as an antidepressant in outpatients with major depression. Primary efficacy measures were the 17-item Hamilton Rating Scale for Depression (HAM-D) and the Clinical Global Impression (CGI) scale. Three of the four studies showed nefazodone to be superior to placebo and comparable to imipramine in efficacy. The fourth trial failed to show a significant difference between both active drugs and placebo. The incidence of side effects and premature treatment discontinuation was higher with imipramine than with nefazodone. Two dose-finding studies, as well as other phase II/III studies, determined the optimal treatment dosage range to be between 300 and 500 mg/day.

Nefazodone has been directly compared with the SSRIs in both short-term (6 to 8 weeks) and maintenance studies (up to 46 weeks). A meta-analysis of six studies reported that nefazodone (n = 523) and SSRIs (fluoxetine, paroxetine, and sertraline; n = 513) were comparable in efficacy for the treatment of moderate to severe major depression. Differences were noted in the side-effect profile favoring nefazodone, particularly with regard to sleep disturbances, weight gain, and sexual dysfunction. An additional meta-analysis involving nine studies comparing nefazodone/trazodone with SSRIs for the treatment of major depression (n = 998) confirmed these findings.

There is considerable evidence for the efficacy of nefazodone in the treatment of various subtypes of depression. A meta-analysis of four clinical trials comparing nefazodone (n = 247), imipramine (n = 166), and placebo (n = 251) showed that nefazodone is effective for the treatment of severe depression as well as melancholic and recurrent depression. A placebo-controlled study evaluated the efficacy of nefazodone in 120 hospitalized patients with severe depression. Significant improvement was noted in patients treated with nefazodone compared to placebo (CGI-I response rates of 50% vs. 29%) in this 6-week study.

Nefazodone has been studied in direct comparison with psychotherapy and combination treatment in patients with chronic depression. In this study, 681 outpatients with chronic, nonpsychotic major depression were assigned to 12 weeks of treatment with nefazodone, The Cognitive Behavioral Analysis System of Psychotherapy, or their combination. In patients who completed the study (*n* = 519), the response rates were 55% in the nefazodone group, 52% in the psychotherapy group, and 85% in the combined treatment group. The response rates for the monotherapy groups are comparable to other studies of chronic depression. The level of response for combination treatment is the highest ever reported in this patient population. A subset of these patients who did not respond to the initial trial of either psychotherapy or nefazodone treatment showed an improved response when treatments were switched (from nefazodone to psychotherapy and vice versa).

Other Mood Disorders

Nefazodone has not been systematically evaluated for the treatment of bipolar depression. One small open-label trial studied nefazodone as an add-on treatment for bipolar depression in 13 patients and showed positive results. The reason to use nefazodone, despite its well-known hepatotoxicity, was to study the underlying mechanism of postsynaptic 5-HT(2A) antagonism as a potential mechanism to develop future therapies for bipolar depression. There are several case reports of nefazodone-induced mania. Caution should be exercised in using nefazodone for the treatment of bipolar depression.

Two small open-label studies reported the efficacy of nefazodone in the treatment of premenstrual dysphoric disorder (PMDD). One small randomized, controlled trial compared its efficacy in PMDD versus buspirone and placebo. This study did not find significant efficacy for nefazodone in the treatment of PMDD, except that it was modestly effective in controlling PMDD-associated irritability; however, the small sample size renders these findings inconclusive.

Symptoms of anxiety during depressive episodes are common. There is evidence from controlled trials that nefazodone helps the comorbid symptoms of anxiety in depressed patients. A meta-analysis of six randomized, controlled trials of nefazodone in patients with major depression showed that nefazodone produced earlier and more sustained improvement in the symptoms of agitation and somatic anxiety compared to imipramine and placebo. This finding was replicated by another retrospective analysis of over 300 patients with major depression. In this analysis, imipramine and nefazodone were both effective in alleviating depressive symptoms; however, nefazodone showed a superior response for comorbid anxiety symptoms. Additional evidence from the nefazodone chronic depression study also supports this beneficial effect of nefazodone on comorbid anxiety, particularly early (first or second week) in the course of treatment.

Anxiety Disorders

Although a number of studies have looked at anxiety symptoms during an episode of major depression, there has been little study of nefazodone for the treatment of primary anxiety disorders. Most of the evidence available is from small, open-label trials.

A meta-analysis of six open-label trials consisting of 105 patients with PTSD (both civilians and combat veterans) showed that nefazodone was effective for a broad spectrum of PTSD symptoms at a dosage of up to 600 mg/day. This finding has been replicated in four additional, small, open-label trials. Randomized controlled trials are needed to determine if nefazodone is superior to placebo and comparable to other treatments for PTSD. An interesting finding in one of the open-label trials was that although patients with chronic PTSD reported improvement in the symptoms of nightmares and other sleep disturbance as well as depression and other PTSD symptoms, the objective, polysomnographic sleep measures did not change, which is in contrast to the sleep studies of depressed patients on nefazodone.

There are two open-label studies of nefazodone use in panic disorder and an additional study in panic disorder comorbid with depression. The total number of patients enrolled in these studies was 39, with study durations of 8 to 12 weeks. Two of the three studies found nefazodone to be beneficial for the treatment of panic disorder, with or without agoraphobia and with or without depression. However, one study reported that the response rate of nefazodone was lower compared to some other antipanic medications such as paroxetine or low-dose venlafaxine.

A clinical case series of five patients suggested the potential efficacy of nefazodone for the treatment of social phobia. An open-label study consisting of 23 patients reported a response rate of about 70% (*n* = 16) for social phobia. However, a double-blind placebo-controlled study for the treatment of social phobia (*n* = 102) reported no significant difference in response rate between nefazodone and placebo-treated patients. Another open-label trial of nefazodone was conducted on 21 patients with generalized anxiety disorder (GAD). Of 15 subjects who completed the trial, 12 were considered significantly improved. An open-label study of nefazodone in patients with depression and comorbid OCD did not report significant antiobsessional effects. Another retrospective study evaluated the efficacy of nefazodone for the treatment of nonparaphilic compulsive sexual behavior. Of the 14 subjects studied, 6 reported good control of sexual obsessions and compulsions and 5 reported remission of those symptoms.

In summary, nefazodone can potentially be useful for a variety of anxiety disorders; however, the existing evidence is not rigorous. If a patient fails the standard treatment, then nefazodone can be used as a second- or third-line treatment, with the understanding that the data regarding its efficacy are very limited.

Substance Abuse and Dual Diagnosis

There is one double-blind, placebo-controlled study of nefazodone in patients with comorbid major depression and alcohol dependence. This study enrolled 64 subjects who were actively drinking alcohol and had a history of at least one episode of depression when abstinent. Patients were randomized to receive 12 weeks of treatment with nefazodone or placebo. All patients continued once-weekly psychoeducational group therapy. The results indicated that nefazodone is efficacious for the treatment of depression in this population, but it did not show any advantage over placebo with regard to alcohol outcome. There was also an increased incidence of nonserious, drug-related side effects (sedation, visual trails, dizziness, and dry mouth). Another larger study of 183 patients with alcohol dependence compared nefazodone with naltrexone and placebo. This study did not find any beneficial effects of nefazodone for the treatment of alcohol dependence. There is one case series reporting successful treatment with nefazodone for depression in four patients who were on methadone maintenance treatment and whose depression had not responded to other medications.

Two double-blind, placebo-controlled studies investigated the use of nefazodone in cocaine-dependent patients. The first study included 69 subjects with cocaine dependence and depressive symptoms, and the primary objective was to determine the efficacy of nefazodone in reducing cocaine use. The second study included 210 subjects. Both trials indicated no specific efficacy of nefazodone in treating cocaine dependence.

A small double-blind study investigated the efficacy of nefazodone and bupropion-sustained release for treating cannabis dependence. The sample size was relatively small. The study did not show any specific benefit of nefazodone or bupropion for the treatment of cannabis dependence in the studied population.

From these reports, it can be concluded that nefazodone does not possess specific efficacy for the treatment of alcohol or substance abuse. However, it can be used for the treatment of comorbid depression in this population. Extreme caution should be exercised in using nefazodone for patients who have alcohol use disorders because of the rare but life-threatening adverse effect of liver failure.

Pain

One open-label study evaluated the efficacy of nefazodone for the treatment of chronic daily headaches in 52 patients. Most of the patients enrolled (*n* = 48) had a history of migraine, based on the criteria established by the International Headache Society. The remainder of the patients had chronic tension-type headaches. The median dose used was 300 mg/day. At the end of the 12th week, 71% of the patients showed at least a 50% improvement in their headache index. Patients who were depressed or anxious showed significantly greater improvement in their headache as well as in the symptoms of depression and anxiety. There is another small, open-label study of 10 patients with diabetes mellitus type II with painful diabetic neuropathy but who did not meet the criteria for major depression. These patients received up to 450 mg/day of nefazodone for 8 weeks. At the end of the treatment period, the patients showed improvement in pain, parasthesias and numbness, as well as slightly improved diabetic control. Both of these studies indicate a potential analgesic effect of nefazodone. However, whether this effect is independent of antidepressant action remains unclear.

Other Uses

There are case reports and small, open-label studies of nefazodone use for the treatment of mood disorders in children and adolescents. The results show that nefazodone is effective for the symptoms of depression in this population and is well tolerated. However, these studies also indicate that the pharmacokinetic properties of nefazodone in children and adolescents are highly variable.

There are also small open-label trials of nefazodone use as an adjunct treatment for schizophrenia in patients who were on a stable dose of antipsychotic medication. There is also one study of nefazodone use in somatization disorder. These studies not only suffer from the limitations common to open-label trials, but they also do not have adequate power to establish the efficacy of nefazodone independent of the comorbid depression present.

A published analysis of nefazodone and cognitive behavioral analysis system of psychotherapy trial showed an interesting finding supporting patient preference as a potent moderator of treatment response for patients with chronic forms of MDD (i.e., patients who preferred psychotherapy responded better to psychotherapy and patients who preferred medication treatment responded better to that intervention).

From the above discussion, it is reasonable to conclude that nefazodone is comparable to other treatments for major depression. However, there is clearly a need for randomized, controlled trials before nefazodone can be recommended for routine use for conditions other than depression.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

General

Premarketing clinical trials with nefazodone in almost 3,500 patients showed that it has a relatively low incidence of adverse events. In these trials, about 16% of participating patients discontinued treatment due to adverse experiences. The commonly observed adverse events that differed significantly from placebo were somnolence (25% vs. 14%), dry mouth (25% vs. 13%), nausea (22% vs. 12%), dizziness (17% vs. 5%), constipation (14% vs. 8%), asthenia (11% vs. 5%), light-headedness (10% vs. 3%), blurred vision (9% vs. 3%), confusion (7% vs. 2%), and abnormal vision, consisting of scotoma and visual trails (7% vs. 1%). The most common reasons for treatment discontinuation were nausea, dizziness, asthenia, and agitation. The incidence of adverse events increases with the dosage of nefazodone. An analysis of three placebo-controlled trials in which two different dosage ranges were used (low ≤300 mg/day; high 300 to 600 mg/day) showed that the incidence of somnolence, nausea, dizziness, constipation, abnormal vision, confusion, and tinnitus was comparable to placebo in the low dosage range. However, the incidence of these adverse events increased significantly in patients treated at the higher dosage range (300 to 600 mg/day).

In 2004, FDA mandated that a boxed warning be placed on the labels of all antidepressants indicating an increased risk of suicidal thoughts and behavior in youth taking these medications and in 2006, it added young adults aged 24 years or less to this list. There are no data indicating a specific increase in this risk by nefazodone; however, caution should be exercised, particularly in children.

Central Nervous System

The most common adverse events pertaining to the CNS are somnolence, dizziness, confusion, insomnia, and lightheadedness. Less frequently reported events include memory impairment, paresthesias, flushing and feeling warm, abnormal dreams, ataxia, and tremors. There is at least one report of a nefazodone-induced acute dystonic reaction, which responded well to standard anticholinergic treatment.

There are reports of rare incidences of seizures after nefazodone administration, particularly in patients who have a seizure disorder; however, a causal relationship has not been established. Controlled data for combining antidepressant drug therapy and ECT are sparse. In a case report from Australia, one patient with treatment-resistant dysthymia underwent ECT while taking 600 mg/day of nefazodone and 15 mg of midazolam without significant problems. There was no adverse outcome, and the patient showed a good response.

Sleep disturbance is one of the most frequent and distressing symptoms of depression. Many antidepressants tend to cause or worsen sleep disturbances. Their effects include decreased REM sleep, increased REM latency, increased stage I sleep, and decreased sleep efficiency. Nefazodone, on the other hand, appears to possess unique and potentially beneficial effects on sleep architecture. It shows minimal effects on REM sleep, decreases stage I sleep and improves sleep efficiency. These effects have been observed in both normal as well as depressed subjects. In comparison with fluoxetine, nefazodone produced significantly greater improvement in sleep on both subjective and objective (sleep EEG) measures. The beneficial effects of nefazodone on sleep tend to appear early in treatment, during the first 2 weeks. The differences between nefazodone and SSRIs on sleep parameters have been shown to remain significant even after patients have responded to the antidepressant treatment. However, a study comparing nefazodone and paroxetine found no difference in subjective sleep improvement after 8 weeks of treatment.

Preclinical studies on human volunteers were carried out to determine the effects of nefazodone on daytime sleepiness, highway driving, and cognitive performance. Results of these early studies showed minimal impact of nefazodone on these measures. Clinical experience in patient populations shows that somnolence is frequently observed, particularly early in the treatment. However, it does not appear to result in an actual increase in daytime sleepiness. The incidence of adverse events related to motor performance and cognitive tasks is also minimal but has a dose–response relationship and tends to increase for doses over 300 mg/day.

Cardiovascular System

Nefazodone has limited effects on the cardiovascular system. It can cause a slightly increased incidence of postural hypotension and hypotension. These effects are much less pronounced compared to the TCAs or trazodone. In a pooled analysis of almost 800 patients, the incidence of postural hypotension with nefazodone was 4%, compared to 1% with placebo. In the same study, the incidence of hypotension was 2%.

There were no clinically significant ECG changes noted in placebo-controlled trials involving 1,153 patients taking nefazodone. Asymptomatic sinus bradycardia was noted in 1.3% of nefazodone-treated patients, compared to 0.4% of placebo-treated patients. The specific cardiac effects of nefazodone in patients with existing cardiovascular disease, particularly of recent onset, are not fully known. Nefazodone inhibited hERG channel activity in a concentration-dependent manner **in HEK-293 cells**, suggesting the theoretical possibility of QTc prolongation. There is a potential for QTc prolongation if nefazodone is combined with terfenadine, astemizole, cisapride, or pimozide.

Genitourinary System

Antidepressant medications are often associated with treatment-related sexual dysfunction, including effects on desire, arousal, and orgasmic function. These side effects are particularly common with SSRIs. In clinical studies, nefazodone has been associated with a low incidence of sexual side effects in both men and women. In one study comparing nefazodone and sertraline in depressed outpatients, 47% of women on sertraline reported difficulty achieving orgasm compared to 17% of women on nefazodone. In the same study, significantly more men taking sertraline (67%) reported ejaculatory difficulty than did men taking nefazodone (19%). The incidence of sexual side effects reported with nefazodone was also low (3.5%) in a study of patients with chronic depression.

Priapism, a recognized side effect of trazodone, has been very rarely reported with nefazodone. There is one case report of nefazodone-precipitated clitoral priapism.

It is believed that the relatively low affinity of nefazodone for α_1-adrenergic receptors is the reason for the low likelihood of treatment-related priapism or postural hypotension.

In premarketing clinical trials, the incidence of urinary frequency, retention, and infection was slightly increased with nefazodone compared to placebo. However, the overall incidence of these adverse events was low.

Gastrointestinal System

Rare cases of life-threatening hepatic failure have been reported in patients treated with nefazodone. The reported rate in the United States is about one case of liver failure resulting in death or transplant per 250,000 to 300,000 patient-years of nefazodone treatment. This rate is considered an underestimate by some because of potential under-reporting. This finding prompted a mandatory black box warning by the FDA and a revision of the product labeling and patient information packet in 2002. This warning is based upon the postmarketing experience of over seven million patients in the United States. The reported rate represents about three to four times the estimated background rate of liver failure. Most liver failures were reported between 1 and 8 months after drug initiation. The common clinical symptoms of hepatotoxicity included jaundice and increases in ALT (10x), AST (10x), total bilirubin (mostly conjugated), and prothrombin time. Dark urine and clay-colored stool were also reported. Available histologic evaluations of livers demonstrated centrilobular necrosis and bile-duct proliferation with cholestasis. Inhibition of bile acid transport by nefazodone has been proposed as an indicator of potential hepatotoxicity in animal models. The radiochemical in vitro methods are available to detect and quantify covalent binding (CB) to liver proteins for a drug or its metabolites and are being studied to predict the risk of liver injury. Nefazodone impairs oxygen consumption by inhibiting the complex I of the electron transport chain in hepatic mitochondria. Additionally, at higher extracellular concentrations, nefazodone could also inhibit anaerobic glycolysis resulting in cell death. During preclinical drug development, all new molecular entities are screened for possible mitochondrial damage by measuring ATP content and mitochondrial membrane potential in primary cultures of human hepatocytes to identify the risk of liver injury. The potential clinical application of these techniques remains to be seen. The risk factors for drug-induced hepatic injury include age and polypharmacy. Given the seriousness of hepatotoxicity, it is clinically important to obtain a complete pretreatment history of any underlying liver disease. Periodic serum testing of liver enzymes has not been shown to prevent serious liver injury. However, early detection of drug-induced hepatic injury and immediate withdrawal of the suspected drug is likely to improve the potential for recovery. Nefazodone should be discontinued in patients showing signs of hepatic toxicity. These patients should not be rechallenged with nefazodone.

More common GI problems associated with nefazodone use include dry mouth (25%), nausea (22%), and constipation (14%).

Other Systems

Nefazodone is generally considered to be weight-neutral. Evidence from both short-term and long-term studies comparing nefazodone with either placebo or other antidepressants supports this conclusion. A pooled analysis of six clinical trials comparing nefazodone (n = 523) and SSRIs (fluoxetine, sertraline, or paroxetine n = 523) as well as three trials comparing nefazodone (n = 224) with imipramine (n = 224) evaluated the impact of these antidepressants on weight. The criterion for clinically significant weight change was a 7% increase or decrease in weight from the baseline measurement. The results showed that in short-term trials, SSRI-treated patients showed more weight loss than nefazodone-treated patients; however, during long-term treatment, more patients gained weight on SSRIs compared to nefazodone. Imipramine was also found to be associated with more weight gain than nefazodone.

Abnormal vision, usually consisting of scotoma and visual trails (palinopsia), has been associated with nefazodone use. This adverse event was noted in up to 10% of the patients taking nefazodone, compared to 2% to 3% of patients on placebo. It is generally considered benign and is probably dose-dependent. It can usually be managed by lowering the medication dose or changing the time of the day the dose is given.

In a retrospective study, joint disorders were considerably more frequently reported adverse effects of 5-HT(2A)-blocking antidepressants, including nefazodone, than of other comparable drugs, suggesting a possible association between the use of 5-HT(2A)-blocking antidepressants and joint disorders.

PRECAUTIONS AND ADVERSE EFFECTS

Overdose

The limited information available regarding the effects of nefazodone overdose during premarketing trials consisted of seven reports of overdose ranging from 1,000 to 11,200 mg, alone or in combination with other drugs. Since its approval, one study has analyzed data on nefazodone from the American Association of Poison Control Centers. It contained over 1,300 human exposures to nefazodone overdose after excluding patients with concomitant agent overdose. Commonly reported symptoms of overdose included drowsiness (17.3%), nausea (9.7%), and dizziness (9.5%). The most common serious clinical effect was hypotension (1.6%). None of these patients died. Most of them were treated with GI decontamination. The study concluded that nefazodone appears to be of low toxicity during poisonings.

There are rare reports of fatalities, predominantly in patients taking an overdose of nefazodone in combination with alcohol and/or other substances. There is no specific antidote available to treat a nefazodone overdose. The mainstay of treatment consists of supportive measures, including adequate airway management and monitoring cardiac rhythm and vital signs.

Used in medically compromised patients during premarketing trials, the pharmacokinetics of nefazodone and its metabolites were studied in a small number of normal healthy volunteers, both younger and older, as well as in patients with biopsy-proven hepatic cirrhosis and patients with impaired renal function. The results of these studies indicate that older adult or cirrhotic patients may require lower doses of nefazodone to achieve and maintain therapeutic effectiveness. Patients with mild to moderate renal impairment (based upon creatinine clearance) appear to tolerate the usual dose without difficulty.

There are no systematic studies of nefazodone treatment of depression in patients with compromised hepatic or renal function. Given the potential for life-threatening hepatotoxicity, nefazodone should be avoided in patients with existing liver disease. It should be used with caution in patients with renal impairment and in patients on hemodialysis. There is one case report of the use of nefazodone to treat depression in a 59-year-old patient on hemodialysis who had not responded to an adequate trial of fluoxetine. Treatment was initiated with nefazodone at a dose of 50 mg BID, which was increased to 100 mg after 1 week and 150 mg after 4 weeks. This patient responded well to nefazodone treatment with minimal adverse events. Pre- and postdialysis nefazodone serum concentrations in this patient suggested that drug levels may be affected by this procedure. Additional studies are needed to confirm the efficacy and safety of nefazodone for the treatment of depression in such patients.

One open-label study evaluated the efficacy of nefazodone in 15 HIV-seropositive patients with major depression. Eleven patients completed the 12-week study. The results showed that HIV-seropositive outpatients respond well to nefazodone and comparably to other outpatient populations. Potential drug interactions with protease inhibitors should be considered before initiating nefazodone therapy in this patient population.

There are anecdotal reports of nefazodone use for depression in patients with cancer on chemotherapy and in patients with epilepsy and cardiac problems. However, systematic studies are not available. Aside from taking routine precautions, increased potential for drug–drug interactions as well as the need for lower dosage and slow titration should be important considerations in instituting nefazodone therapy in medically ill patients.

Use During Pregnancy and Lactation

Nefazodone is classified as a pregnancy category C medication. It has been shown to increase early pup mortality in rabbits at a dose approximately five times the human dose. A decrease in pup weight was also seen at this dose as well as at lower doses. There are no systematic human studies evaluating nefazodone during pregnancy. Nefazodone should be used in pregnancy only if the potential benefit justifies the potential risk to the fetus.

Very little information is available regarding the excretion of nefazodone in human breast milk. One small study of two lactating mothers reported that nefazodone is excreted in human milk; no metabolites were found. The levels of nefazodone in breast milk were not related to the maternal medication dose. Drug levels in the infants were not assayed; however, there was no indication of developmental difficulty in those infants. A case report of a premature, breast-fed infant described drowsiness, lethargy, poor feeding, and difficulty maintaining body temperature attributed to maternal use of nefazodone. Until more information becomes available, a careful risk–benefit analysis must be performed before recommending the use of nefazodone in lactating mothers.

Discontinuation Syndrome

Animal studies using monkeys trained for self-administration of cocaine as a reinforcer did not show this behavior when nefazodone was substituted for cocaine, suggesting that nefazodone has no abuse liability. Its potential for tolerance, physical dependence, or withdrawal has not been studied systematically in humans.

There are case reports suggesting possible withdrawal symptoms resulting from abrupt discontinuation of nefazodone. Typical symptoms reported include dizziness, nausea, vomiting, fatigue, insomnia, and paresthesias. Some unusual symptoms such as brief moments of disorientation and shock-like or electrical sensations have also been reported. Patients generally recover within a few days without any serious problems, and the symptoms can usually be prevented by tapering the medication. The incidence of discontinuation symptoms with nefazodone appears to be much less common than with SSRIs and venlafaxine. In a relapse prevention study of 259 patients who had been maintained on nefazodone treatment for 16 weeks and were randomly assigned to continuation treatment with nefazodone or placebo, there were no significant differences in new-onset or worsened adverse experiences between the nefazodone and placebo groups at 14 and 28 days after randomization.

DRUG INTERACTIONS

There are various mechanisms involved in the drug–drug interactions of nefazodone. Most of these interactions involve Cytochrome P450 (CYP) isoenzymes. Another potential mechanism involves drugs that are highly bound to plasma proteins, such as warfarin. Different metabolites of nefazodone also play a role in drug interactions; for example, the plasma concentration of mCPP is increased in individuals who have decreased activity of CYP2D6.

Psychotropic Agents

Nefazodone and its metabolites, hydroxynefazodone, mCPP, and triazoledione, inhibit CYP3A4, both in vitro and in vivo. This inhibition can result in increased bioavailability of medications metabolized by this pathway. Therefore, caution should be exercised when administering nefazodone with other compounds metabolized by CYP3A4. This interaction becomes even more important for drugs that are exclusively or predominantly metabolized by the CYP3A4 pathway. Two such benzodiazepines, alprazolam and triazolam, show a substantial and clinically important increase in blood levels when administered with nefazodone. When triazolam was coadministered with nefazodone, triazolam’s half-life increased by threefold, AUC increased by fourfold, and peak concentration increased by 1.7-fold. Likewise, there was a 1.6-fold increase in peak concentration and a twofold increase in the AUC of alprazolam. Nefazodone plasma concentration was unaffected by either medication. Given these interactions, the manufacturer recommends that triazolam be avoided during nefazodone treatment in most patients. If alprazolam is used with nefazodone, a 50% reduction in the initial dose of alprazolam is recommended. This effect has not been observed with lorazepam.

One of the metabolites of nefazodone, mCPP, is present in relatively small quantities and is metabolized by the CYP2D6 pathway. When a patient is switched from an antidepressant that is an inhibitor of CYP2D6 (e.g., fluoxetine) to nefazodone, the metabolism of mCPP may be inhibited for some time. This can result in an increased level of mCPP, which is an anxiogenic compound, and may lead to increased anxiety and agitation. This problem can be minimized by using an adequate washout period before switching antidepressants. The interaction between nefazodone and MAOIs has not been systematically studied in humans or animals. The existing evidence with SSRIs or TCAs, when combined with MAOIs, suggests a potential for a serious, sometimes fatal, reaction. This serotonin syndrome consists of hyperthermia, rigidity, myoclonus, autonomic instability with rapid fluctuation of vital signs, and mental status changes that may include extreme agitation progressing to delirium and coma. This reaction has also been reported in patients who have recently discontinued an SSRI or TCA and have been started on MAOI. Some cases present with symptoms resembling neuroleptic malignant syndrome, including severe hyperthermia and seizures. Given the existing evidence with other antidepressants, it is recommended that nefazodone not be used in combination with an MAOI or within 14 days of discontinuing treatment with an MAOI. In addition, a minimum of 1 week should be allowed after discontinuing nefazodone before starting an MAOI.

Coadministration of lithium and nefazodone resulted in only a slight decrease in the concentration of two metabolites of nefazodone, mCPP and triazoledione, which was not clinically significant. The combination of nefazodone and carbamazepine, although well tolerated in a small number of healthy volunteers, resulted in a very significant decrease in the steady-state C_{max} and AUC of nefazodone (86% and 93%, respectively), along with a similar but modest reduction in its metabolites. At the same time, the steady-state C_{max} and AUC of carbamazepine were each increased by 23%, while the steady-state C_{max} and AUC of its metabolite, 10,11-epoxycarbamazepine, decreased by 21% and 20%, respectively. Because the decrease in nefazodone and its metabolites is likely to result in a loss of antidepressant efficacy, this combination should be avoided. There are no systematic studies to evaluate the impact of coadministration of valproic acid and nefazodone. One case report suggested a potentially dangerous interaction of these two medications in a patient with bipolar disorder. This patient suffered from symptoms that were suggestive of serotonin syndrome. Whether this occurrence was incidental or represents a real interaction remains unclear. The impact of combining phenytoin and nefazodone has not extensively been studied. It appears that a single-dose administration of phenytoin with steady-state nefazodone does not alter the pharmacokinetic properties of phenytoin. However, due to the nonlinear pharmacokinetics of phenytoin, this combination should be avoided if possible, or the dose should be adjusted carefully, according to usual clinical practices.

Nefazodone tends to decrease the protein binding of haloperidol by 5%. However, this interaction is unlikely to be of any clinical significance. Human experience with clozapine and nefazodone in a small number of patients suggests that clozapine levels are increased by 4% and norclozapine levels by 16%. This relatively modest increase in levels is considered to be clinically insignificant. There are other small studies of concomitant antipsychotic and nefazodone use that have shown no significant difficulties. However, an important and potentially dangerous interaction may occur between nefazodone and pimozide, resulting in a prolonged QT interval, as described in detail in the following section.

Other Medications

Terfenadine, astemizole, cisapride, and pimozide are metabolized via the CYP3A4 pathway. Drugs such as ketoconazole or erythromycin inhibit this enzyme and block the metabolism of terfenadine, astemizole, cisapride, and pimozide, resulting in increased plasma concentrations of these medications. Increased plasma concentrations of these four medications are associated with QT prolongation and rare cases of cardiovascular adverse events, including death, due principally to ventricular tachycardia of the torsades de pointes type. Nefazodone is a potent inhibitor of CYP3A4. Consequently, nefazodone must not be used in combination with terfenadine, astemizole, cisapride or pimozide. At the present time, terfenadine, astemizole and cisapride have been removed from the U.S. market by the FDA.

Nefazodone was coadministered with digoxin to healthy volunteers who were genotyped as extensive CYP2D6 metabolizers. This combination resulted in an increased C_{max}, C_{min}, and AUC of digoxin by 29%, 27%, and 15%, respectively. There was no effect on the pharmacokinetics of nefazodone or its active metabolites. Since digoxin has a narrow therapeutic index, caution should be exercised in combining it with nefazodone. Plasma level monitoring of digoxin is also recommended. Coadministration of nefazodone and propranolol in healthy volunteers resulted in a decrease in the level of propranolol but did not impact the level of nefazodone or its active metabolites with the exception of mCPP, which was increased. Hence, initial dosage adjustment is not needed for either of the medications; however, subsequent dosage should be adjusted based upon the clinical response.

Simvastatin and atorvastatin are substrates of CYP3A4. When nefazodone was combined with either of these medications, it resulted in a 20-fold increase in the concentration of simvastatin and simvastatin acid and a 4-fold increase in atorvastatin and atorvastatin lactone. This effect appears to result from the inhibition of CYP3A4 by nefazodone. Nefazodone did not significantly affect the level of pravastatin, another HMG-CoA reductase inhibitor. Metabolic interactions are unlikely between nefazodone and HMG-CoA reductase inhibitors such as pravastatin, fluvastatin, or rosuvastatin that undergo little or no metabolism by the CYP3A4 isoenzyme; therefore, dosage adjustment for these agents should not be necessary. However, dose adjustment is needed for simvastatin and atorvastatin because the increased concentrations of these medications might result in a greater risk of rhabdomyolysis and other adverse events associated with HMG-CoA reductase inhibitors.

There have been reports of increased blood concentrations of cyclosporin and tacrolimus into toxic ranges in patients who received concomitant nefazodone. These are immunosuppressive agents that are substrates of CYP3A4. If nefazodone is combined with either of these medications, the level should be monitored, and the dose should be adjusted accordingly.

When nefazodone was administered with warfarin, there was no change in prothrombin or bleeding time despite a modest decrease in $C_{\max}$ of warfarin. This finding suggests a clinically insignificant interaction.

DOSAGE AND ADMINISTRATION

Nefazodone is indicated for the treatment of major depression. Based on the results of clinical trials, the recommended starting dose of nefazodone is 200 mg/day in divided doses (BID). The effective dosage range is usually between 300 mg to 600 mg/day in divided doses. The starting dose can be increased by 100 to 200 mg/day at intervals of no less than 1 week. In order to minimize adverse events, it is preferable to start at 100 mg/day and then increase by 100 mg/day every week. The usual therapeutic dosage range is 400 to 600 mg/day. Doses higher than 600 mg/day are not recommended. As with all antidepressants, several weeks of treatment at a reasonable therapeutic dose are required for antidepressant response. As with other antidepressants, a long-term study (up to 1 year) with nefazodone as a maintenance treatment suggests that the therapeutic dose should be continued as the maintenance dose. Once-daily dosing has been attempted in patients who were previously maintained on BID dosing as well as in patients starting nefazodone. The results did not show significant differences with once- versus twice-daily dosing of nefazodone. Since these were open-label trials, this finding needs to be replicated in double-blind studies.

The recommended starting dose for older adults or debilitated patients is 50 mg BID. The increase in dosage should be slower and tailored to the individual patient's symptoms and side effects. The final therapeutic range in older adult patients may be similar to younger patients.

Nefazodone is supplied in 50, 100, 150, 200, and 250 mg strengths.

FURTHER READINGS

Benson BE, Mathiason M, Dahl B, et al. Toxicities and outcomes associated with nefazodone poisoning: an analysis of 1,338 exposures. *Am J Emerg Med*. 2000;18(5):587–592.

Carpenter KM, McDowell D, Brooks DJ, Cheng WY, Levin FR. A preliminary trial: double-blind comparison of nefazodone, bupropion-SR, and placebo in the treatment of cannabis dependence. *Am J Addict*. 2009;18(1):53–64.

Ciraulo DA, Knapp C, Rotrosen J, et al. Nefazodone treatment of cocaine dependence with comorbid depressive symptoms. *Addiction*. 2005;100(Suppl 1):23–31.

Choi JM, Oh SJ, Lee JY, et al. Prediction of drug-induced liver injury in HepG2 cells cultured with human liver microsomes. *Chem res Toxicol*. 2015;28(5):872–885.

DeVane CL, Grothe DR, Smith SL. Pharmacology of antidepressants: focus on nefazodone. *J Clin Psychiatry*. 2002;63(1):10–17.

Dunner DL, Laird LK, Zajecka J, et al. Six-year perspectives on the safety and tolerability of nefazodone. *J Clin Psychiatry*. 2002;63(1):32–41.

Evelin K, Michele V, Zhibo G, et al. Untargeted metabolomics reveals anaerobic glycolysis as a novel target of the hepatotoxic antidepressant nefazodone. *J Pharmacol Exp Ther*. 2020;375(2):239–246.

Feiger AD, Bielski RJ, Bremner J, et al. Double-blind, placebo-substitution study of nefazodone in the prevention of relapse during continuation treatment of outpatients with major depression. *Int Clin Psychopharmacol*. 1999;14(1):19–28.

Feighner J, Targum SD, Bennett ME, et al. A double-blind, placebo-controlled trial of nefazodone in the treatment of patients hospitalized for major depression. *J Clin Psychiatry*. 1998;59(5):246–253.

Fontaine R, Ontiveros A, Elie R, et al. A double-blind comparison of nefazodone, imipramine, and placebo in major depression. *J Clin Psychiatry*. 1994;55(6):234–241.

Gillin JC, Smith-Vaniz A, Schnierow B, et al. An open-label, 12-week clinical and sleep EEG study of nefazodone in chronic combat-related posttraumatic stress disorder. *J Clin Psychiatry*. 2001;62(10):789–796.

Hicks JA, Argyropoulos SV, Rich AS, et al. Randomised controlled study of sleep after nefazodone or paroxetine treatment in outpatients with depression. *Br J Psychiatry*. 2002;180:528–535.

Keller MB, McCullough JP, Klein DN, et al. A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *N Engl J Med*. 2000;342(20):1462–1470.

Kocsis JH, Leon AC, Markowitz JC, et al. Patient preference as a moderator of outcome for chronic forms of major depressive disorder treated with nefazodone, cognitive behavioral analysis system of psychotherapy, or their combination. *J Clin Psychiatry*. 2009;70(3):354–361.

Kodesh A, Katz S, Lerner AG, Finkel B, Sigal M. Intermittent, luteal phase nefazodone treatment of premenstrual dysphoric disorder. *J Psychopharmacol*. 2001;15(1):58–60.

Kostrubsky SE, Strom SC, Kalgutkar AS, et al. Inhibition of hepatobiliary transport as a predictive method for clinical hepatotoxicity of nefazodone. *Toxicol Sci*. 2006;90(2):451–459.

Landen M, Eriksson O, Sundblad C, et al. Compounds with affinity for serotonergic receptors in the treatment of premenstrual dysphoria: a comparison of buspirone, nefazodone and placebo. *Psychopharmacology (Berl)*. 2001;155(3):292–298.

Marcus RN, Mendels J. Nefazodone in the treatment of severe, melancholic, and recurrent depression. *J Clin Psychiatry*. 1996;57(Suppl 2):19–23.

Papakostas GI, Fava M. A meta-analysis of clinical trials comparing the serotonin (5HT)-2 receptor antagonists trazodone and nefazodone with selective serotonin reuptake inhibitors for the treatment of major depressive disorder. *Eur Psychiatry*. 2007;22(7):444–447.

Passos SR, Camacho LA, Lopes CS, dos Santos MA. Nefazodone in outpatient treatment of inhaled cocaine dependence: a randomized double-blind placebo-controlled trial. *Addiction*. 2005;100(4):489–494.

Ren Z, Chen S, Zhang J, et al. Endoplasmic reticulum stress induction and ERK1/2 activation contribute to nefazodone-induced toxicity in hepatic cells. *Toxicol Sci*. 2016;154(2):368–380.

Roy-Byrne PP, Pages KP, Russo JE, et al. Nefazodone treatment of major depression in alcohol-dependent patients: a double-blind, placebo-controlled trial. *J Clin Psychopharmacol*. 2000;20(2):129–136.

Rush AJ, Armitage R, Gillin JC, et al. Comparative effects of nefazodone and fluoxetine on sleep in outpatients with major depressive disorder. *Biol Psychiatry*. 1998;44(1):3–14.

Schatzberg AF, Prather MR, Keller MB, et al. Clinical use of nefazodone in major depression: a 6-year perspective. *J Clin Psychiatry*. 2002;63(1):18–31.

Schatzberg AF, Rush AJ, Arnow BA, et al. Chronic depression: medication (nefazodone) or psychotherapy (CBASP) is effective when the other is not. *Arch Gen Psychiatry*. 2005;62(5):513–520.

Silva AM, Barbosa IA, Seabra C, et al. Involvement of mitochondrial dysfunction in nefazodone-induced hepatotoxicity. *Food Chem Toxicol*. 2016;94:148–158.

Sussman N, Ginsberg DL, Bikoff J. Effects of nefazodone on body weight: a pooled analysis of selective serotonin reuptake inhibitor- and imipramine-controlled trials. *J Clin Psychiatry*. 2001;62(4):256–260.

Van Ameringen M, Mancini C, Oakman J, et al. Nefazodone in the treatment of generalized social phobia: a randomized, placebo-controlled trial. *J Clin Psychiatry*. 2007;68(2):288–295.

Voican CS, Corruble E, Naveau S, Perlemuter G. Antidepressant-induced liver injury: a review for clinicians. *Am J Psychiatry*. 2014;171(4):404–415.

33.6c Mirtazapine

Mousa Sobhy Azer Botros, M.D., Radu V. Saveanu, M.D., and Michael E. Thase, M.D.

Neuroscience-Based Nomenclature (NbN)	
Monoamine Multimodal	
Pharmacologic domain	Norepinephrine, Serotonin
Mechanism of action	Receptor antagonist (Ne alpha-2, 5-HT_2, 5-HT_3)

Mirtazapine (Remeron) is a tetracyclic piperazinoazepine compound (Fig. 33.6c–1) that the FDA approved in 1995 for the treatment of depression. Unlike most newer antidepressants, mirtazapine has virtually no effect on monoamine uptake. Rather, therapeutic effects are thought to be mediated by inhibition of α_2-adrenergic receptors and blockade of postsynaptic serotonin type 2 (5-HT_2) and type 3 (5-HT_3) receptors. The former action diminishes the inhibitory effects of both α_2-autoreceptors (located on presynaptic norepinephrine [NE] neurons) and α_2-heteroreceptors (located on cell bodies of dorsal raphe 5-HT neurons). Inhibition of inhibitory cellular mechanisms thus increases the firing rates of noradrenergic and 5-HT neurons. This effect differs markedly from that of most other antidepressants, which tend to suppress cell firing rates until autoreceptors are downregulated.

Blockade of 5-HT_2 and 5-HT_3 receptors may enhance therapeutic effects directly (e.g., relief of insomnia or enhancement of appetite) or indirectly by dampening side effects attributable to disinhibited 5-HT release (e.g., agitation or nausea). Antagonism of 5-HT_2 receptors also may enhance the availability of neurotransmitters to interact with serotonin type 1 (5-HT_1) receptors. Mirtazapine is the only newer antidepressant that is a potent antagonist of histamine-1 receptors, which adds to sedative and appetite-enhancing properties. The sleep effects of mirtazapine in major depression have been studied. Mirtazapine therapy has been shown to lessen nocturnal awakenings and prolong total sleep time. Mirtazapine administration increased total slow wave sleep (SWS), REM latency, the duration of the first REM episode, and stage 3 sleep. It decreased the number of REM episodes. Simultaneously, it reduced wake time after sleep onset.

First synthesized in the early 1980s, mirtazapine is related to one older antidepressant, mianserin (Tolvon). Mianserin is available throughout most of the world (although not in the United States or Canada). Mianserin is a more potent antagonist of α_1-receptors than the newer compound, which is thought to explain why mianserin has less effect on 5-HT neurotransmission than mirtazapine.

PHARMACOLOGY

Mirtazapine has a half-life of approximately 30 hours and reaches steady state after 6 days of therapy. Linear pharmacokinetics is observed across the therapeutic dose range (i.e., 15 to 60 mg/day). Mirtazapine is essentially completely absorbed after oral dosing, and food has little effect on the rate or extent of absorption. Approximately 85% of circulating drug is bound to plasma protein.

Mirtazapine is metabolized in the liver via cytochrome P450 (CYP) 2D6, CYP 1A2, and CYP 3A4 pathways. The primary metabolite, *N*-

desmethylmirtazapine, is substantially less active than the parent compound and has trivial pharmacologic effects. Cimetidine, which inhibits CYP 2D6, CYP 1A2, and CYP 3A4, raises plasma levels of mirtazapine by more than 50%. Even greater increases in mirtazapine concentrations result from concomitant administration of the potent CYP 1A2 inhibitor fluvoxamine (Luvox).

The original formulation of mirtazapine is available in 15-, 30-, and 45-mg tablets. An alternative formulation, an orally disintegrating tablet (ODT), was introduced in 2001 to facilitate compliance and to aid therapy for patients who have difficulty swallowing.

INDICATIONS

Antidepressant Efficacy

The minimum effective dose of mirtazapine in placebo-controlled clinical trials of major depressive disorder (MDD) is 15 mg/day. The maximum approved dose ranges from 45 mg/day (United States) to 60 mg/day (some European countries). There is some evidence of an ascending dose–response relationship for antidepressant efficacy, although this has not been definitively established.

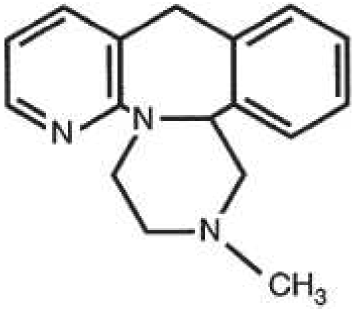

FIGURE 33.6c–1. Chemical structure of mirtazapine (Remeron).

The relative efficacy of mirtazapine has been established against older and newer antidepressants. Mirtazapine was found to be significantly more effective than trazodone (Desyrel) and as effective as the TCAs: amitriptyline (Elavil), doxepin (Sinequan), clomipramine (Anafranil), and nortriptyline (Aventyl). Therapy with the TCA, imipramine (Tofranil), was significantly more effective than mirtazapine treatment in one inpatient study. However, in this trial, dosages of both drugs were titrated in relation to plasma levels, which is a valid strategy for the TCA but of uncertain value for mirtazapine.

There are a number of published studies comparing mirtazapine and SSRIs, including citalopram (Celexa), paroxetine (Paxil), sertraline (Zoloft), fluvoxamine, and fluoxetine (Prozac). Several meta-analyses have been completed and have confirmed that, on average, the two types of antidepressants are comparably effective (less than 5% difference in response or remission rates) after 6 to 8 weeks of therapy, although patients treated with mirtazapine received an earlier therapeutic benefit. It is not clear if this temporal difference is primarily due to a relatively nonspecific advantage in improving sleep and appetite or if it is attributable to the difference in mechanism of action.

Two studies have contrasted mirtazapine with the serotonin-norepinephrine reuptake inhibitor (SNRI) venlafaxine (Effexor). In the first study, mirtazapine (mean dose: 50 mg/day) and the immediate-release formulation of venlafaxine (mean dose: 255 mg/day) were compared in 157 depressed inpatients with melancholic features. Trends favored the mirtazapine-treated group on most outcome measures, but only the difference on the Hamilton Rating Scale for Depression (HAMD) sleep disturbance factor reached statistical significance at the study endpoint. Significantly fewer patients did, however, withdraw from the mirtazapine group during the first 6 weeks of therapy because of side effects. The second study, which was conducted on 242 depressed outpatients, contrasted the ODT formulation of mirtazapine (45 mg/day) with the extended-release formulation of venlafaxine (225 mg/day). Although the efficacy of the study drugs again did not differ at the end of the trial, results consistently favored mirtazapine therapy during the first 2 weeks of double-blind therapy, with significant differences in overall symptom improvement, response rates, and changes in sleep.

Combining SSRIs/SNRIs and Mirtazapine

Almost from the time of introduction, it was speculated that mirtazapine could be combined with SSRIs to obtain synergistic or additive effects. Subsequently, a number of studies have examined the clinical utility—both effectiveness and tolerability—of combining mirtazapine with other antidepressants, including SNRIs and bupropion as well as SSRIs.

The first RCT to evaluate mirtazapine in combination with other antidepressants (predominantly SSRIs) studied patients with a documented history of nonresponse to antidepressant therapy in the current depressive episode. In this pilot study, 26 outpatients were randomly assigned to 4 weeks of additional double-blind therapy with either active mirtazapine (15 to 30 mg/day) or a matching placebo, in addition to ongoing therapy with the first antidepressant. The investigators found that the patient group receiving adjunctive therapy with active mirtazapine had significantly greater improvement in depressive symptoms than the group receiving adjunctive placebo. Although the difference in remission rates was not statistically significant in this small study, the magnitude of the difference (45% vs. 13%)—if replicable in a larger study—would be clinically meaningful.

Another study tested the hypothesis that starting patients on the combination of mirtazapine and an SSRI would deliver a more rapid and/or larger response than monotherapy. This outpatient RCT randomly assigned 61 participants with MDD to 6 weeks of double-blind therapy with one of three options: (1) mirtazapine (30 to 60 mg/day) plus placebo; (2) paroxetine (20 to 40 mg/day) plus placebo; or (3) mirtazapine plus paroxetine. The group receiving combined therapy of mirtazapine obtained a significantly larger reduction in depressive symptoms than the groups that received mirtazapine and paroxetine monotherapy groups. The remission rate for the combined therapy group at week 6 was 43%, as compared to 19% and 26% for the mirtazapine and paroxetine monotherapy groups, respectively. Although again not statistically significant, such a difference in remission rates would be considered clinically significant if confirmed in a larger study.

A second study on the value of combining antidepressants from the outset of therapy randomly assigned 105 MDD outpatients to receive 6 weeks of therapy with one of four strategies. One group received fluoxetine (20 mg/day) plus a placebo; this group served as the active comparator. A second group received fluoxetine (20 mg/day) in combination with mirtazapine (30 mg daily); this group directly tested the value of combining mirtazapine with the SSRI. Two other groups received mirtazapine (30 mg/day) in combination with either venlafaxine (225 mg/day) or bupropion (150 mg/day); these groups permitting a preliminary comparison of the utility of combining mirtazapine with other first-line antidepressants. This study again tested the hypothesis that patients receiving combination antidepressant therapy would show a more rapid or more complete remission than monotherapy from the outset of therapy. At the end of 6 weeks of double-blind therapy, the fluoxetine plus placebo group had a remission rate of 25%. Across all outcome measures, trends favored the groups that received mirtazapine in combination with fluoxetine (52% remission rate), venlafaxine XR (58% remission rate), or bupropion (46% remission rate).

The combination of mirtazapine and venlafaxine was evaluated in two other studies. The first study was part of the STAR*D project. In this open-label trial, 109 MDD outpatients who had not obtained adequate benefit from three rounds of antidepressant therapy were randomized to receive 12 weeks of therapy with mirtazapine (mean dose: 36 mg/day) and venlafaxine XR (mean dose: 210 mg/day) or the monoamine oxidase inhibitor tranylcypromine (mean dosage: 37 mg/day). Outcomes were assessed by self-report and by blinded independent evaluators. At the study endpoint, neither group did particularly well, though the remission rate in the mirtazapine plus venlafaxine XR group was twice that of the group that received the MAOI (14% vs. 7%). Patients receiving combination therapy were more likely to complete the trial and did show a statistically significant difference in improvement in depressive symptoms. A relatively few patients completed 12 weeks of treatment with tranylcypromine at a therapeutic dose (i.e., at least 40 mg/day); one implication is that physicians in contemporary practice are better able to implement a trial of combined therapy than execute a successful switch to an MAOI.

A subgroup of the STAR*D investigators completed a second, larger study of the value of combining antidepressants at the beginning of treatment. This trial, known by the acronym COMED, randomized 220 MDD outpatients to 12 weeks of treatment with escitalopram plus placebo (i.e., the active comparison group), mirtazapine (mean dose: 18 mg/day) plus venlafaxine XR (mean dose: 178 mg/day), or escitalopram plus bupropion (i.e., an alternative widely used combination that had performed reasonably well in STAR*D). This trial found no evidence to support the routine use of antidepressant combinations from the outset of therapy, which reported significantly greater side effects.

Prophylactic Therapy

Two studies have evaluated longer-term treatment with mirtazapine. The first report examined the outcomes of patients who received extended therapy after responding to double-blind, acute-phase therapy with mirtazapine, amitriptyline, or placebo. Results indicated that patients who remained on mirtazapine had a significantly lower risk of relapse than patients who continued to take a double-blind placebo. The amitriptyline-treated group had an intermediate outcome.

The second study was a formal relapse prevention trial in which 156 patients who had remitted during 8 to 12 weeks of acute-phase therapy were randomized to remain on active mirtazapine ($N = 76$) or to switch to placebo ($N = 80$). Across the subsequent 40 weeks of double-blind therapy, 44% of the placebo group relapsed, compared to 20% of the group receiving active continuation-phase therapy. The effects of continuation-phase mirtazapine therapy were of the same magnitude observed in studies of other effective antidepressants.

OTHER INDICATIONS FOR MIRTAZAPINE

Mirtazapine is not formally indicated for the treatment of any other disorder. Nevertheless, mirtazapine may be useful for the treatment of insomnia, panic disorder, generalized anxiety disorder, PTSD, and OCD, as well as autism and other pervasive developmental disorders. A number of small studies and case series have evaluated the utility of mirtazapine in treating anxiety and depression in cancer patients. Blockade of 5-HT_3 and histamine receptors may have particularly useful effects in the symptomatic management of anxiety and depression in cancer patients. Improvements in pain, nausea, appetite, insomnia, depression, and anxiety have been reported.

A study of 12 adult outpatient subjects with comorbid MDD/alcohol dependence was recently conducted. Subjects were given 15 mg of mirtazapine for the first 2 weeks and 30 mg for the last 6 weeks of the medication trial. Mirtazapine improved depression on the Beck Depression Inventory (BDI) and decreased the number of alcoholic drinks per week. These preliminary findings suggest that mirtazapine may be efficacious in reducing both depressive symptoms and excessive alcohol use in comorbid MDD and alcohol dependence. Two years later, 10 of the 12 patients who entered the acute-phase study participated in the follow-up study. The significant improvements initially observed in depressive symptoms (BDI), drinking, and sleep disturbance persisted at the follow-up evaluation. A study of depressed cocaine-dependent subjects who received either mirtazapine (11 subjects) or placebo (13 subjects) for 12 weeks suggested that mirtazapine is superior to placebo in improving sleep in patients with comorbid depression and cocaine dependence but is not more effective than placebo in reducing cocaine use.

In another study, 16 agitated patients with Alzheimer disease (AD) were enrolled in a 12-week open-label pilot study. Patients were given 15 to 30 mg/day of mirtazapine. The study found that agitated patients with AD could be treated with mirtazapine without significant adverse events or cognitive deterioration. Larger-scale studies are now warranted.

MIRTAZAPINE USE DURING PREGNANCY AND LACTATION

In a series of case reports, mirtazapine seems to be an effective treatment for mothers with hyperemesis gravidarum that failed to respond to conventional antiemetic treatment.

It is not known whether mirtazapine may cause harm to the unborn baby. Three reports (*n* = 266) referred to increased risk for preterm birth or shorter duration of pregnancy after exposure to mirtazapine. In an unadjusted analysis, mirtazapine was associated with spontaneous abortion (RR 2.23, 95% CI 1.34–3.70). The authors suggested this may be a reflection of disease severity. No correlation between mirtazapine exposure and spontaneous or therapeutic abortion, intrauterine death or infant death has been found when investigating the sole effects of mirtazapine exposure. Six papers reported no neurobehavioral developmental adverse effects in 22 infants exposed to mirtazapine when assessed from 1.5 months to 12 months, except for one case report that described an insignificant trend toward lower scores on both fine and gross motor function (p = .07 and .09).

Data on the neonatal risk of mirtazapine during lactation are scarce. In a study of eight women who were taking a median dose of 38 mg of mirtazapine per day while breastfeeding, concentrations of mirtazapine or desmethylmirtazapine were measured in milk and plasma. All infants were clinically examined, and four infants had blood drawn for analysis. No adverse effects were seen. Mirtazapine was detected in only one of four infants tested. The study suggested that mirtazapine use by lactating women is safe for breastfed infants. Four papers reported on 11 cases of mothers exposed to mirtazapine during lactation. The relative infant dose ranged from undetectably low to 2.9%. No adverse events were detected in three of the reports. One case report referred to possible prolonged sleep and weight gain when compared to older siblings. Given the poverty of evidence, the long-term safety and effects of mirtazapine in lactation require further assessment. Nevertheless, each decision to breastfeed should always be made on the basis of individual risk–benefit analysis.

TOLERABILITY, SAFETY, AND TOXICITY

Attrition from mirtazapine therapy because of adverse events generally was less than that observed in comparison groups treated with tricyclic antidepressants but somewhat greater than that observed in comparative studies versus SSRIs. Mirtazapine and venlafaxine have comparable attrition rates in controlled studies. The specific side-effect profile of mirtazapine is quite different, however, from that associated with SSRIs and SNRIs.

The most common side effects reported during mirtazapine therapy are sedation, increased appetite, weight gain, and dry mouth. The drug’s antihistaminic effects primarily cause these side effects. Approximately 50% of patients report sedation during the first week of mirtazapine therapy. This side effect can be clinically significant, as demonstrated in studies of reaction time or automotive driving performance. Sedation usually is not minimized by slow upward titration of the dose, and lower and higher starting doses of mirtazapine have similar rates of sedation. If early sedation can be tolerated, however, most patients adapt to this side effect without difficulty. When compared to SSRIs or SNRIs, mirtazapine therapy is associated with significantly lower rates of gastrointestinal disturbance and a lower incidence of treatment-emergent sexual dysfunction.

Mirtazapine-provoked periodic leg movements (PLMS) in 67% of young, healthy males. The effect was most pronounced in the first few days but gradually decreased in the first week. The indirect increase in central dopamine via 5-HT$_{1A}$ receptor activation may explain why mirtazapine-induced PLMS decreased over time. Further, diphenhydramine, another first-generation H$_1$ antagonist, provoked PLMS during wake time in subjects with restless legs syndrome, pointing to a possible antihistaminergic mechanism in the induction of PLMS.

Mirtazapine may cause withdrawal syndrome upon sudden discontinuation. Symptoms include depression, anxiety, restlessness, irritability, decreased appetite, insomnia, diarrhea, nausea, and vomiting. Therefore, reduction in dose should be gradual and slow.

Mirtazapine is not approved for use in children under the age of 18 years.

Perhaps 15% to 25% of patients gain an unacceptable amount of weight during mirtazapine therapy. This annoying adverse effect may occur more commonly among younger patients, and, interestingly, weight gain was not a problem in one study of depression in late life. The likelihood of problematic weight gain can usually be predicted within the first 4 to 6 weeks of therapy. A small study of seven women with MDD treated with mirtazapine for 6 weeks concluded that mirtazapine was associated with a significant increase in body weight, body mass fat and leptin concentration. Interestingly, glucose metabolism remained unaffected in this study, in contrast to other psychotropic medications inducing weight gain.

The effects of mirtazapine and venlafaxine on leptin plasma levels were recently contrasted. After 28 days of treatment, mirtazapine treatment was strongly associated with an increase in leptin concentrations. There was also a significant effect on the leptin/body mass index (BMI) ratio elevated in mirtazapine-treated patients and slightly declining in venlafaxine-treated patients. Blockade of histamine synthesis may contribute to leptin resistance, as well as via possible histamine H_1-receptor–mediated dysregulation of hypothalamic nuclei integrating central and peripheral signals relevant to energy balance.

The only significant effect of mirtazapine therapy on clinical chemistry profiles has been increased blood lipid levels. As this effect occurred with approximately twice the frequency seen in patients randomly assigned to treatment with amitriptyline, weight gain alone does not appear to be the cause.

The most controversial severe adverse effect possibly associated with mirtazapine therapy is agranulocytosis. Although concern has existed about this as a potential problem in the early (premarketing) studies, a definite association has not been established over the subsequent two decades.

Mirtazapine is remarkably safe in overdose. The principal problems following ingestion of as much as 5 g of mirtazapine are sedation, aspiration pneumonia, delirium, and tachycardia. No cases of seizure or abnormal heart rhythms have been reported. In the only reported fatal overdose, the patient also had ingested toxic levels of several other drugs.

FUTURE DIRECTIONS

Mirtazapine is distinctly different from the remainder of the newer antidepressants. It is at least as effective as the SSRIs and most likely conveys earlier symptomatic benefits. It also differs dramatically from the SSRIs and venlafaxine with respect to tolerability. Virtually free from causing side effects commonly associated with 5-HT uptake inhibitors (e.g., gastrointestinal disturbances, insomnia, and sexual dysfunction), mirtazapine therapy is complicated by sedation (in as many as half of all patients) and weight gain (in as many as a fourth of all patients). Mirtazapine is increasingly used for the treatment of older depressed people, for whom these effects are more likely to be therapeutic than problematic. Moreover, mirtazapine's antiemetic benefits may offer relief for mothers with treatment-resistant hyperemesis gravidarum.

One study demonstrated that the interaction between HTR1A+272G>A and 5-HTT VNTR genotypes may be involved in responding to mirtazapine treatment in MDD patients. The response was greater for the combination than for individual genotypes. This observation should be replicated and further expanded in larger-scale studies.

Further Readings

Aichhorn W, Whitworth, AB, Weiss U, Stuppaeck C. Mirtazapine and breast-feeding. *Am J Psychiatry*. 2004;161(12):2325.

Anttila SA, Leinonen EV. A review of the pharmacological and clinical profile of mirtazapine. *CNS Drug Rev*. 2001;7(3):249–264.

Benkert O, Szegedi A, Philipp M, et al. Mirtazapine orally disintegrating tablets versus venlafaxine extended release: a double-blind, randomized multicenter trial comparing the onset of antidepressant response in patients with major depressive disorder. *J Clin Psychopharmacol*. 2006;26(1):75–78.

Blier P, Ward HE, Tremblay P, et al. Combination of antidepressant medications from treatment initiation for major depressive disorder: a double-blind randomized study. *Am J Psychiatry*. 2010;167(3):281–288.

Bruijn JA, Moleman P, Mulder PG, et al. A double-blind, fixed blood-level study comparing mirtazapine with imipramine in depressed inpatients. *Psychopharmacology (Berl)*. 1996;127(3):231–237.

Chang HS, Lee HY, Cha JH, et al. Interaction of 5-HTT and HTR1A gene polymorphisms in treatment responses to mirtazapine in patients with major depressive disorder. *J Clin Psychopharmacol*. 2014;34(4):446–454.

Clayton AH, Montejo AL. Major depressive disorder, antidepressants, and sexual dysfunction. *J Clin Psychiatry*. 2006;67(Suppl 6):33–37.

Cornelius JR, Douaihy AB, Clark DB, et al. Mirtazapine in comorbid major depression and alcohol use disorder: a long-term follow-up study. *J Addict Behav Ther Rehabil*. 2013;3(1):1–7.

de Boer T. The pharmacologic profile of mirtazapine. *J Clin Psychiatry*. 1996;57(Suppl 4):19–25.

Djulus J, Koren G, Einarson TR, et al. Exposure to mirtazapine during pregnancy: a prospective, comparative study of birth outcomes. *J Clin Psychiatry*. 2006;67(8):1280–1284.

Fulda S, Kloiber S, Dose T, et al. Mirtazapine provokes periodic leg movements during sleep in young healthy men. *Sleep*. 2013;36(5):661–669.

Guclu S, Gol M, Dogan E, Saygili U. Mirtazapine use in resistant hyperemesis gravidarum: report of three cases and review of the literature. *Arch Gynecol Obstet*. 2005;272(4):298–300.

Guelfi JD, Ansseau M, Timmerman L, Kørsgaard S, Mirtazapine-Venlafaxine Study Group. Mirtazapine versus venlafaxine in hospitalized severely depressed patients with melancholic features. *J Clin Psychopharmacol*. 2001;21(4):425–431.

Kristensen JH, Ilett KF, Rampono J, Kohan R, Hackett LP. Transfer of the antidepressant mirtazapine into breast milk. *Br J Clin Pharmacol*. 2007;63(3):322–327.

Laimer M, Kramer-Reinstadler K, Rauchenzauner M, et al. Effect of mirtazapine treatment on body composition and metabolism. *J Clin Psychiatry*. 2006;67(3):421–424.

McGrath PJ, Stewart JW, Fava M, et al. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR*D report. *Am J Psychiatry*. 2006;163(9):1531–1541.

Montgomery SA, Reimitz PE, Zivkov M. Mirtazapine versus amitriptyline in the long-term treatment of depression: a double-blind placebo-controlled study. *Int Clin Psychopharmacol*. 1998;13(2):63–73.

Papakostas GI, Homberger CH, Fava M. A meta-analysis of clinical trials comparing mirtazapine with selective serotonin reuptake inhibitors for the treatment of major depressive disorder. *J Psychopharmacol*. 2008;22(8):843–848.

Rohde A, Dembinski J, Dorn C. Mirtazapine (Remergil) for treatment resistant hyperemesis gravidarum: rescue of a twin pregnancy. *Arch Gynecol Obstet*. 2003;268(3):219–221.

Rush AJ, Trivedi MH, Stewart JW, et al. Combining medications to enhance depression outcomes (CO-MED): acute and long-term outcomes of a single-blind randomized study. *Am J Psychiatry*. 2011;168:689–701.

Schilling C, Gilles M, Blum WF, et al. Leptin plasma concentrations increase during antidepressant treatment with amitriptyline and mirtazapine, but not paroxetine and venlafaxine: leptin resistance mediated by antihistaminergic activity? *J Clin Psychopharmacol*. 2013;33(1):99–103.

Schittecatte M, Dumont F, Machowski R, et al. Mirtazapine, but not fluvoxamine, normalizes the blunted REM sleep response to clonidine in depressed patients: implications for subsensitivity of alpha(2)-adrenergic receptors in depression. *Psychiatry Res*. 2002;109(1):1–8.

Shen J, Chung SA, Kayumov L, et al. Polysomnographic and symptomatological analyses of major depressive disorder patients treated with mirtazapine. *Can J Psychiatry*. 2006;51(1):27–34.

Smit M, Dolman KM, Honig A. Mirtazapine in pregnancy and lactation—A systematic review. *Eur Neuropsychopharmacol*. 2016;26(1):126–135.

Stahl S, Zivkov M, Reimitz PE, Panagides J, Hoff W. Meta-analysis of randomized, double-blind, placebo-controlled, efficacy and safety studies of mirtazapine versus amitriptyline in major depression. *Acta Psychiatr Scand Suppl*. 1997;391:22–30.

Thase ME, Nierenberg AA, Keller MB, Panagides J, Relapse Prevention Study Group. Efficacy of mirtazapine for prevention of depressive relapse: a placebo-controlled double-blind trial of recently remitted high-risk patients. *J Clin Psychiatry*. 2001;62(10):782–788.

Thase ME, Nierenberg AA, Vrijland P, et al. Remission with mirtazapine and selective serotonin reuptake inhibitors: a meta-analysis of individual patient data from 15 controlled trials of acute phase treatment of major depression. *Int Clin Psychopharmacol*. 2010;25:189–198.

Tonn P, Reuter SC, Hiemke C, Dahmen N. High mirtazapine plasma levels in infant after breast feeding: case report and review of the literature. *J Clin Psychopharmacol*. 2009;29(2):191–192.

Wheatley DP, van Moffaert M, Timmerman L, Kremer CM. Mirtazapine: efficacy and tolerability in comparison with fluoxetine in patients with moderate to severe major depressive disorder. Mirtazapine-Fluoxetine Study Group. *J Clin Psychiatry*. 1998;59(6):306–312.

Yaris F, Ulku C, Kesim M, et al. Psychotropic drugs in pregnancy: a case-control study. *Prog Neuropsychopharmacol Biol Psychiatry*. 2005;29(2):333–338.

33.7 Serotonin Antagonists/Inverse Agonists

33.7a Pimavanserin

JOAN M. STRIEBEL, M.D., JENNIFER A. O'DAY, M.D., NEKISA HAGHIGHAT, M.D., M.P.H., AND STEPHEN M. STAHL, M.D., PH.D., D.SC. (HON.)

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic domain	Serotonin
Mechanism of action	Receptor antagonist (5-HT_{2A})

Pimavanserin is a drug with antipsychotic actions that is a selective inverse agonist and antagonist (a partial inverse agonist) at 5-HT_{2A} and 5-HT_{2C} receptors. It is unique among drugs with antipsychotic actions in that it has negligible binding at dopamine receptors with no effect on motor function.

HISTORY

Parkinson's disease (PD) is a complex multisystem neurodegenerative disease. Patients suffer from motor symptoms—primarily bradykinesia, rigidity, and/or tremor—related to the loss of dopamine neurons in the substantia nigra pars compacta. Degeneration and/or aberrant activity of dopaminergic, cholinergic, noradrenergic, and serotonergic neurons result in non-motor symptoms. Parkinson's disease psychosis (PDP) is a non-motor symptom characterized by at least one of the following: illusions, false sense of presence, hallucinations, or delusions, which is present for at least 1 month in a PD patient after ruling out other etiologies. PDP is seen in up to 60% of PD patients and is associated with dementia, more hospitalizations, increased caregiver burden, nursing home placement, and death. With the exception of low-dose clozapine or low-dose quetiapine, treatment of PDP with other drugs that have antipsychotic actions frequently results in intolerable motor impairment due to dopamine D_2 receptor antagonism in the dorsal striatum. Clozapine's ability to treat PDP at mean doses of 25 to 36 mg daily and minute plasma levels of 4 to 6 ng/mL led to the recognition that its 5-HT_{2A} inverse agonism was the basis for its antipsychotic efficacy. Low doses of quetiapine also have anecdotal evidence of efficacy in PDP. Due to the need for mandatory hematologic monitoring, alternatives to clozapine were sought. In the late 1990s, Acadia Pharmaceuticals screened for potent, selective, and effective 5-HT_{2A} receptor inverse agonists and identified ACP-103 (pimavanserin). Subsequent safety and efficacy studies led to the FDA's approval of pimavanserin for the treatment of PDP in 2016.

Psychosis consisting of hallucinations and delusions is also a feature of other neurodegenerative disorders, including Alzheimer's disease, vascular dementia, frontotemporal dementia, and dementia with Lewy bodies. Although drugs with antipsychotic actions are prescribed off-label to treat dementia-related psychosis (DRP), efficacy is limited. Further, most drugs with antipsychotic actions demonstrate a negative effect on cognitive function and are associated with a higher risk of mortality compared with placebo. In contrast, early pimavanserin DRP studies did not demonstrate an adverse safety signal with regard to mortality risk, and pimavanserin appears not to impact cognition. These findings, along with the drug's receptor binding profile, built hopes that pimavanserin will also treat DRP.

CHEMISTRY

Pimavanserin is a member of the class of ureas in which three of four hydrogens are replaced by 4-fluorobenzyl, 1-methylpiperidin-4-yl, and 4-(isopropyloxy)benzyl groups. Pimavanserin tartrate's chemical name is urea, *N*-[(4-fluorophenyl)methyl]-*N*-(1-methyl-4-piperidinyl)-*N′*-[[4-(2-methylpropoxy)phenyl]methyl]-,(2*R*,3*R*)-2,3-dihydroxybutanedioate (2:1). It is an aromatic ether and tertiary amino compound. The molecular formula of pimavanserin tartrate is $(C_{25}H_{34}FN_3O_2)_2 \cdot C_4H_6O_6$, and its molecular weight is 1005.20. Its chemical structure is depicted in Figure 33.7a–1. Pimavanserin tartrate is water soluble.

PHARMACOLOGIC ACTION

Absorption

Pimavanserin is only available as an oral preparation. The median time to peak plasma concentration for pimavanserin and its major active *N*-desmethylated metabolite (AC-279) is 6 hours (range of 4 to 24 hours). AC-279 does not contribute to the drug's clinical effects as it does not cross the blood–brain barrier to any appreciable extent. High-fat meals do not significantly affect the rate or extent of drug exposure. The mean elimination half-life for pimavanserin and AC-279 is 57 hours and 200 hours, respectively.

Distribution

Pimavanserin is highly protein bound (~95%) in human plasma. Protein binding is dose-independent.

Metabolism and Elimination

Pimavanserin is predominantly metabolized by CYP3A4 and CYP3A5 and to a lesser extent by CYP2J2, CYP2D6, and various other CYP and flavin-containing monooxygenase enzymes. CYP3A4 is primarily responsible for the formation of AC-279. Pimavanserin and AC-279 do not cause clinically significant CYP inhibition or induction. In terms of excretion, less than 1% of the administered dose of pimavanserin or AC-279 was recovered in urine.

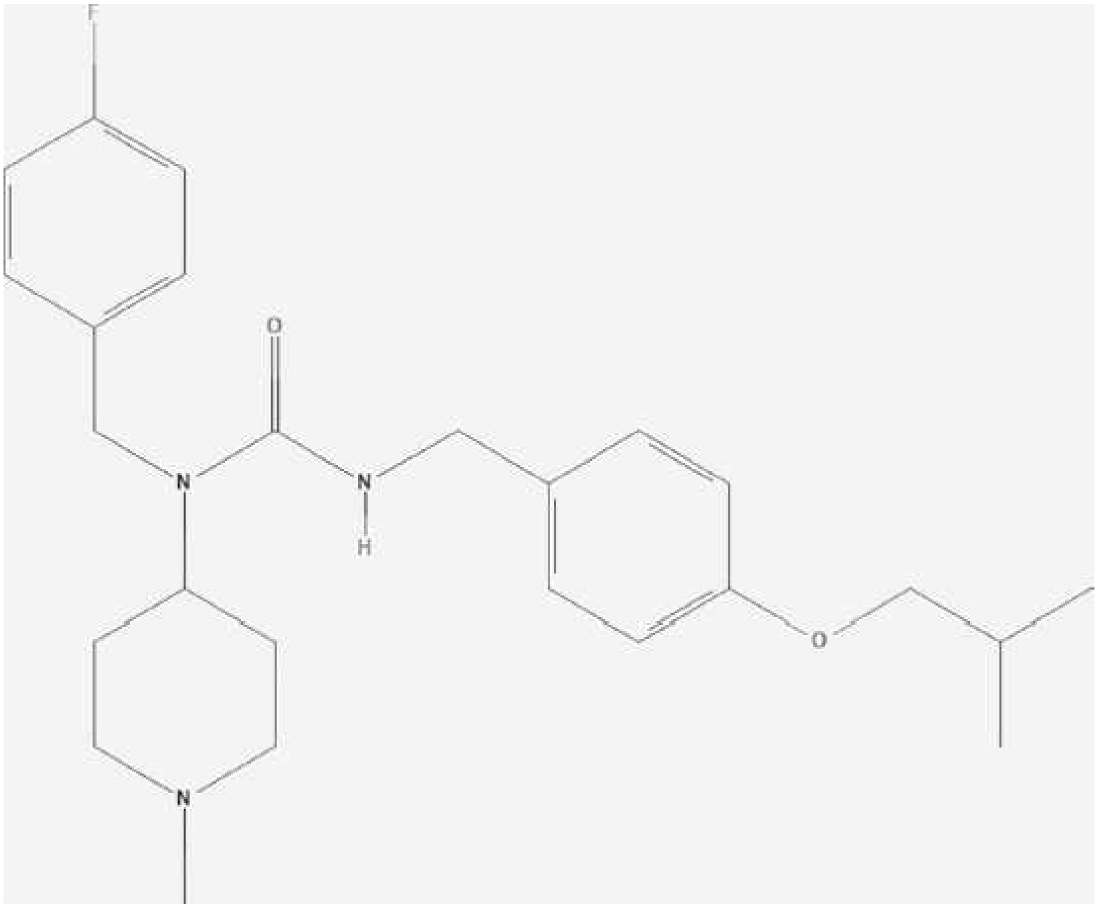

FIGURE 33.7a–1. Molecular structure of pimavanserin. (From National Center for Biotechnology Information. PubChem Compound Database; CID = 10071196, https://pubchem.ncbi.nlm.nih.gov/compound/Pimavanserin. Accessed September 15, 2021.)

Table 33.7a–1.
Binding Affinities of Pimavanserin

Receptor	Binding Affinity (K_i nM)	Binding Profile
5-HT_{2A}	0.087	Inverse agonist
5-HT_{2C}	0.44	Inverse agonist
D_2	>300	
H_1	>300	
M_1	>300	
α_1	>300	

Pharmacodynamics

Mechanism of Action. Pimavanserin is a potent, selective inverse agonist at serotonin 5-HT_{2A} receptors with lesser affinity at serotonin 5-HT_{2C} receptors. Pimavanserin has no appreciable affinity at serotonin 5-HT_{2B}, dopaminergic (including D_2), muscarinic, histaminergic, or adrenergic receptors, or calcium channels. See Table 33.7a–1 for specific binding affinities.

Pimavanserin exerts its antipsychotic effects through at least two pathways. In PDP, the loss of serotonergic neurons in the midbrain results in upregulation and increased sensitivity of postsynaptic 5-HT_{2A} receptors in the ventral visual pathway—an occipitotemporal network with projections to the ventrolateral prefrontal cortex and the orbitofrontal cortex. The stimulation of these receptors induces visual hallucinations, analogous to the visual phenomena occurring after ingestion of 5-HT_{2A} agonists such as lysergic acid diethylamide. As an inverse agonist, pimavanserin blocks the actions of serotonin and decreases the constitutive activity of the 5-HT_{2A} and 5-HT_{2C} receptors in the ventral visual pathway, reducing psychosis in patients with PDP.

The second pathway involves postsynaptic 5-HT_{2A} receptors located on a specific population of glutamatergic pyramidal neurons projecting from the prefrontal cortex. Excessive activity of these receptors leads to the downstream release of dopamine in the dopamine tract extending from the midbrain to the associative areas of the dorsal striatum and results in the positive symptoms of psychosis. Pimavanserin diminishes the excitatory tone of serotonin at these 5-HT_{2A} receptors resulting in an antipsychotic effect differing from that due to D_2 antagonism. Meltzer and colleagues (2012) provide support for this mechanism. In a population of acutely psychotic schizophrenia patients, pimavanserin added to risperidone 2 mg daily demonstrated comparable efficacy to risperidone 6 mg daily. When added to haloperidol 2 mg daily, the antipsychotic effect of pimavanserin's 5-HT_{2A} antagonism was not seen in the context of haloperidol's greater D_2 antagonism.

Table 33.7a–2.

Published Trials With Pimavanserin for Parkinson’s Disease Psychosis

Reference	Comments
Meltzer et al., 2010	**Sample:** Phase II double-blind, randomized, multicenter 28-day study of pimavanserin 20 mg daily titrating to 40–60 mg daily after 2 weeks vs. placebo in 60 patients with PDP as defined by Ravina et al. (2007). **Primary Endpoint:** Safety and tolerability as measured by the UPDRS Parts II and III. Primary efficacy measure was the SAPS total score and a UPDRS Part I psychosis score. **Results:** Pimavanserin did not separate from placebo with regard to motor function, sedation, hypotension, or other side effects. In terms of efficacy, there was a trend toward greater improvement with pimavanserin; statistical significance was not reached.
Cummings et al., 2014	**Sample:** Phase III double-blind, randomized, multicenter 6-week study of pimavanserin 40 mg daily vs. placebo, initiated after a 2-week lead-in period in 199 patients with PDP. **Primary Endpoint:** Change in total SAPS-PD score from baseline to week 6. **Results:** Pimavanserin-treated patients exhibited a significant improvement in the SAPS-PD compared to placebo (–5.79 [37% improvement] vs. –2.73 [14% improvement], respectively, effect size of 0.50, *p* = .0006. Treatment with pimavanserin resulted in significant decreases in other measures of

	psychosis, including the SAPS hallucinations and delusions score.

SAPS: Scale for the Assessment of Positive Symptoms; SAPS-PD: Scale for the Assessment of Positive Symptoms for Parkinson’s Disease Psychosis; UPDRS: Unified Parkinson’s Disease Rating Scale.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Parkinson’s Disease Psychosis

Given that there were no FDA-approved medications for the treatment of PDP, pimavanserin’s manufacturer received a Breakthrough Therapy designation in 2014, which lessened the standards required for review. Table 33.7a–2 summarizes the published placebo-controlled trials. The pivotal phase III trial was preceded by three trials that established pimavanserin’s safety and lack of effect on motor function. These early trials, including one published phase II trial, failed to reach statistical significance at the primary efficacy endpoint. When designing the phase III trial, investigators identified factors that were positively associated with effect size in failed studies, including study locations in North America, the use of centralized raters to reduce inter-rater variability, a 2-week lead-in period to minimize the placebo response, uniform pimavanserin dose of 40 mg daily, and use of a novel scale – the nine-item Scale for the Assessment of Positive Symptoms for Parkinson’s Disease Psychosis (SAPS-PD)—which measured the symptoms most relevant to PDP. The SAPS-PD was developed through posthoc analysis of questions on the Scale for the Assessment of Positive Symptoms (SAPS) that showed the most favorable change in the failed phase II trials. The original protocol for the phase III study listed the SAPS as the primary endpoint, and 16 months after data collection, the protocol was modified to include the SAPS-PD.

In the phase III trial, a 50% reduction in the SAPS-PD corresponded to a clinical presentation of much improved, was clinically meaningful, and was defined as a response. At the end of week 6, 37.2% of pimavanserin-treated patients and 27.8% of placebo-treated patients were responders. The number needed to treat (NNT) was 11; therefore, 11 patients must be treated with pimavanserin for one patient to demonstrate a 50% reduction in the SAPS-PD. Defining “harm” as any serious adverse event, including death, the number needed to harm (NNH) was 22. Since the NNH/NNT ratio was two, for every two patients with a 50% reduction in the SAPS-PD, one patient will experience a pimavanserin-related serious adverse event.

Two phase IV open-label extension studies evaluated the efficacy, safety, and tolerability of pimavanserin in 459 patients over varying lengths of treatment. In terms of safety, most adverse events were mild to moderate, and pimavanserin did not adversely affect motor function. Mortality was 6.45 deaths per 100 patient-years, which is in line with that expected for an older adult, medically fragile cohort. In terms of efficacy, response to pimavanserin was durable and was maintained for an additional 4 weeks of treatment.

Dementia-Related Psychosis

Given that pimavanserin resulted in a greater improvement in the SAPS-PD score in patients with cognitive impairment, including patients who were concomitantly treated with cognitive-enhancing medications, it was speculated that pimavanserin might treat psychosis occurring in other dementing illnesses. Pimavanserin was granted the FDA’s Breakthrough Therapy designation for the treatment of DRP in 2017. Table 33.7a–3 summarizes the published placebo-controlled trials. Based on the positive phase II study and data from the phase III HARMONY trial, the manufacturer submitted a supplemental new drug application for pimavanserin to treat hallucinations and delusions in patients with DRP. As of this writing, approval has not yet been granted.

Other Potential Indications Under Investigation

Pimavanserin had been under investigation as an adjunctive treatment for major depressive disorder. Although it demonstrated robust efficacy in a phase II trial (CLARITY), in two identical double-blind, placebo-controlled phase III trials, pimavanserin failed to achieve statistical significance on the Hamilton Depression Rating Scale, which was the primary endpoint. Case series examining the efficacy of adjunctive pimavanserin in treatment-resistant schizophrenia have yielded conflicting reports. Pimavanserin is currently under investigation as an adjunctive treatment for negative symptoms in schizophrenia and for insomnia in veterans with PTSD.

Table 33.7a–3. Published Trials With Pimavanserin for Dementia-Related Psychosis

Ballard et al., 2018	**Sample:** Phase II double-blind, randomized, 12-week study of 181 nursing home residents with possible or probable Alzheimer’s disease and psychotic symptoms who were treated with pimavanserin 34 mg daily or placebo. **Primary Endpoint:** Mean change from baseline to week 6 on NPI-NH psychosis score. **Results:** At week 6, pimavanserin-treated patients exhibited a significant improvement in the NPI-NH psychosis score compared to placebo (–3.76 [SE 0.65] vs. –1.93 [SE 0.63]) respectively, Cohen’s *d* effect size –0.32; *p* = 0.045. By week 12, no significant difference was observed between pimavanserin and placebo.
Tariot et al., 2021	**Sample:** Phase III double-blind, multicenter, randomized, discontinuation study involving 392 patients with psychosis related to Alzheimer’s disease (66.3%), Parkinson’s disease dementia (15.1%), vascular dementia (9.7%), dementia with Lewy bodies (7.1%), or frontotemporal dementia (1.8%). During the 12-week open-label lead-in phase, all patients were treated with pimavanserin 34 mg daily. Those who improved (61.8%) were randomized to pimavanserin or placebo for the double-blind treatment period of up to 26 weeks. **Primary Outcome:** Time from randomization to relapse of psychosis with a *p*-value below 0.0033. Relapse was

defined as hospitalization due to DRP, significant deterioration as measured by clinical scales, withdrawal from the study due to lack of efficacy, or the initiation of an off-label antipsychotic medication.

Results: The trial was stopped once it met the primary efficacy endpoint. A relapse occurred in 13% of the pimavanserin group and in 28% of the placebo group. There was no significant difference in mean MMSE scores between pimavanserin-treated patients and those receiving placebo.

MMSE, Mini-Mental State Examination; NPI-NH,: Neuropsychiatric Inventory-Nursing Home Version.

Therapeutic Indications

Pimavanserin is an FDA-approved treatment for patients with hallucinations and delusions associated with PDP. While studies treating patients with DRP have been positive, to date, the FDA has not approved pimavanserin for this indication.

Precautions and Adverse Reactions

Pimavanserin is contraindicated in individuals with hypersensitivity to pimavanserin or any components of the tablet or capsule.

Black Box Warning. Pimavanserin carries a black box warning regarding an increased all-cause risk of death in older adult patients with DRP. An FDA review of all postmarketing reports of deaths in patients with PDP treated with pimavanserin failed to identify new or unexpected safety findings and considered that patients with PDP have a higher mortality rate due to older age, advanced PD and other medical conditions.

QT Interval Prolongation. The effect of pimavanserin on the QT interval has been examined in patients with PDP, DRP, and healthy volunteers. The phase III PDP trial excluded subjects with QTc (Bazett correction formula) >460 msec for men or >470 msec for women, history of long QT, medical conditions that could prolong the QT interval, and medications known to prolong the QT interval. ECG data showed that pimavanserin 40 mg daily resulted in a mean QTc increase of 7.3 msec from baseline compared with no change for placebo. This increase was not associated with adverse cardiac events. In long-term safety studies, the mean maximum QTc (Fredericia correction formula) was <450 msec over the duration of the study, although individual QTc (Fredericia) values of >500 msec occurred in 2.6% of patients. In the phase III DRP trial, the mean QTc (Fridericia) increase was 5.4±0.9 msec. Five adverse events involving asymptomatic prolongation of the QT interval affected 1.3% of the study population over the trial's duration. The effect of pimavanserin at twice the maximal approved daily dose was evaluated in a randomized placebo- and positive-controlled double-blind, multiple-dose parallel QTc study in 252 healthy subjects. The maximum mean change from baseline (upper bound of the two-sided 90% CI) was 13.5 (16.6) msec.

Pimavanserin should be avoided in patients with known QT prolongation or those currently taking other medications known to prolong the QT interval. These medications include Class 1A antiarrhythmics (e.g., quinidine, procainamide), Class 3 antiarrhythmics (e.g., amiodarone, sotalol), certain antipsychotics (e.g., ziprasidone, chlorpromazine, thioridazine), and certain antibiotics (e.g., gatifloxacin, moxifloxacin). Pimavanserin should be avoided in patients with cardiac arrhythmias as well as in other circumstances that increase the risk of torsade de pointes, including symptomatic bradycardia, hypokalemia, hypomagnesemia, and the presence of congenital long QT syndrome.

Common and Drug-Related Adverse Reactions in Clinical Trials. In premarketing PDP studies, the most common adverse reactions associated with pimavanserin (defined as incidence ≥2% and greater than placebo) were nausea (7% pimavanserin vs. 4% placebo), constipation (4% pimavanserin vs. 3% placebo), peripheral edema (7% pimavanserin vs. 2% placebo), gait disturbance or fall (2% pimavanserin vs. <1% placebo), hallucinations (5% pimavanserin vs. 3% placebo), and confusional state (6% pimavanserin vs. 3% placebo). In the phase III PDP trial, 11% of pimavanserin-treated patients had a serious adverse event compared with 4% of subjects in the placebo group. Of the ten pimavanserin-treated patients who discontinued, six did so due to psychosis and of these, four discontinued within the first 10 days of the trial. The most common adverse events in the PDP open-label extension studies were falls, urinary tract infections, hallucinations, decreased weight, and confusional states. Postmarketing comparative pharmacovigilance assessment of pimavanserin versus treatment alternatives in a cohort of PD patients revealed no unforeseen adverse safety signal. When compared with other treatments, no excess mortality related to pimavanserin was identified.

In the phase II DRP study, adverse events occurred in 17% of pimavanserin-treated patients and 11% of placebo-treated patients. Specifically, agitation (19 [21%] vs. 13 [14%]), aggression (nine [10%] vs. four [4%]), and peripheral edema (seven [8%] vs. two [2%]) were more common with pimavanserin compared to placebo. There was no difference in the frequency of discontinuation due to adverse events or in the number of deaths. The most common side effects in the phase III DRP study that occurred more frequently with pimavanserin than with placebo were headache, constipation, and urinary tract infection.

Pregnancy. There is insufficient information on pimavanserin to determine if it is associated with a risk for birth defects or miscarriage. Animal data demonstrated that pimavanserin was not teratogenic. When pimavanserin was administered to pregnant or lactating rats at doses of 0.14 to 14-times the recommended dose, there was evidence of maternal toxicity as well as a decrease in pup survival, litter size, pup weight, and food consumption.

Nursing Mothers. There is insufficient information about the presence of pimavanserin in human milk, the effects on the breastfed infant, or the effects on milk production.

Drug Interactions

Concomitant use of pimavanserin with moderate CYP3A4 inducers (e.g., modafinil, nafcillin, efavirenz) or strong CYP3A4 inducers (e.g., rifampin, phenytoin, carbamazepine) is not recommended as treatment with a moderate or strong CYP3A4 inducer reduces pimavanserin exposure. Concomitant treatment with a strong CYP3A4 inhibitor (e.g., ketoconazole, clarithromycin, ritonavir) will increase pimavanserin exposure. If pimavanserin is used with a strong CYP3A4 inhibitor, the recommended dose is 10 mg once daily.

No dose adjustment of carbidopa/levodopa is required when administered with pimavanserin. Doses of P-glycoprotein substrates (e.g., digoxin, amiodarone) and other CYP3A4 substrates (e.g., clonazepam, warfarin) do not require adjustment when used with pimavanserin.

Dosage and Administration

Pimavanserin is available in 10-mg tablets and 34-mg capsules. Each capsule contains 40 mg of pimavanserin tartrate, which is equivalent to 34 mg of the pimavanserin-free base. The recommended dose is 34 mg taken orally once daily, without titration. Pimavanserin can be taken with or without food. Capsules can either be swallowed whole or opened with the contents sprinkled over a tablespoon of soft foods such as apple sauce, yogurt, or pudding.

Switching a patient with PDP from clozapine or quetiapine to pimavanserin should be commenced with attention to (1) preserving a stable level of 5-HT$_{2A}$ antagonism that allows symptomatic control throughout the switch and (2) avoiding cholinergic or histaminergic rebound effects by slowly tapering quetiapine or clozapine. The strategy developed by expert consensus involves continuing the quetiapine or clozapine while starting pimavanserin at 34 mg daily and allowing the medication to reach steady state. If tapering quetiapine, begin the taper four weeks after starting pimavanserin, reduce by 50% weekly if starting from a dose ≤100 mg/day or by 25%

weekly if starting from higher doses. Upon reaching 12.5 mg/day, discontinue the quetiapine. If tapering clozapine at doses ≤100 mg/day, begin the taper 6 weeks after starting pimavanserin, and taper clozapine by 6.25 mg weekly until discontinued.

Dosing in Special Populations

GERIATRIC USE. No dose adjustment is required for older adult patients. Of note, PDP primarily occurs in older individuals. The mean age of patients enrolled in the 6-week clinical studies was 71 years, with 49% of patients 65 to 75 years old and 31% of patients >75 years old. Neither the 6-week clinical studies nor the long-term open-label study identified significant differences in safety or effectiveness related to age.

HEPATIC IMPAIRMENT. No dose adjustment is recommended for patients with mild to severe hepatic impairment.

RENAL IMPAIRMENT. No dose adjustment is recommended for patients with mild to severe renal impairment or end-stage renal disease. However, pimavanserin should be used with caution in patients with severe renal impairment as defined as creatinine clearance <30 mL/min (Cockcroft-Gault) as the peak plasma level and total drug exposure were increased. Dialysis does not significantly affect the pimavanserin plasma concentration.

FURTHER READINGS

Acadia Pharmaceuticals Inc. Nuplazid (pimavanserin) package insert. 2020.

Acadia Pharmaceuticals Inc. Acadia Pharmaceuticals receives complete response letter from U.S. FDA for supplemental new drug application for pimavanserin for the treatment of hallucinations and delusions associated with dementia-related psychosis. https://ir.acadia-pharm.com/news-releases/news-release-details/acadia-pharmaceuticals-receives-complete-response-letter-us-fda. Published April 5, 2021. Accessed August 29, 2021.

Ballanger B, Strafella AP, van Eimeren T, et al. Serotonin 2A receptors and visual hallucinations in Parkinson's disease. *Arch Neurol*. 2010;67(4):416–421.

Ballard C, Banister C, Khan Z, et al. Evaluation of the safety, tolerability, and efficacy of pimavanserin versus placebo in patients with Alzheimer's disease psychosis: a phase 2, randomised, placebo-controlled, double-blind study. *Lancet Neurol*. 2018;17(3):213–222.

Ballard CG, Kreitzman DL, Isaacson S, et al. Long-term evaluation of open-label pimavanserin safety and tolerability in Parkinson's disease psychosis. *Parkinsonism Relat Disord*. 2020;77:100–106.

Ballard C, Youakim JM, Coate B, Stankovic S. Pimavanserin in Alzheimer's disease psychosis: efficacy in patients with more pronounced psychotic symptoms. *J Prev Alzheimers Dis*. 2019;6(1):27–33.

Black KJ, Nasrallah H, Isaacson S, et al. Guidance for switching from off-label antipsychotics to pimavanserin for Parkinson's disease psychosis: an expert consensus. *CNS Spectr*. 2019;24(5):574–575.

Brown JD, Cicali B, Henriksen C, et al. Comparative pharmacovigilance assessment of mortality with pimavanserin in Parkinson's disease-related psychosis. *J Manag Care Spec Pharm*. 2021;27(6):785–790.

Cummings J, Ballard C, Tariot P, et al. Pimavanserin: Potential treatment for dementia-related psychosis. *J Prev Alzheimers Dis*. 2018;5(4):253–258.

Cummings J, Isaacson S, Mills R, et al. Pimavanserin for patients with Parkinson's disease psychosis: a randomised, placebo-controlled phase 3 trial. *Lancet*. 2014;383(9916):533–540.

Espay AJ, Guskey MT, Norton JC, et al. Pimavanserin for Parkinson's Disease psychosis: Effects stratified by baseline cognition and use of cognitive-enhancing medications. *Mov Disord*. 2018;33(11):1769–1776.

Hacksell U, Burstein ES, McFarland K, Mills RG, Williams H. On the discovery and development of pimavanserin: a novel drug candidate for Parkinson's psychosis. *Neurochem Res*. 2014;39(10):2008–2017.

Isaacson SH, Ballard CG, Kreitzman DL, et al. Efficacy results of pimavanserin from a multi-center, open-label extension study in Parkinson's disease psychosis patients. *Parkinsonism Relat Disord*. 2021;87:25–31.

Lindenmayer JP, Rizvi S. Augmentation with pimavanserin in treatment refractory schizophrenia: case report. *J Clin Psychopharmacol*. 2021;41(3):337–339.

McGregor MM, Nelson AB. Circuit mechanisms of Parkinson's disease. *Neuron*. 2019;101(6):1042–1056.

Meltzer HY, Elkis H, Vanover K, et al. Pimavanserin, a selective serotonin (5-HT)2A-inverse agonist, enhances the efficacy and safety of risperidone, 2mg/day, but does not enhance efficacy of haloperidol, 2mg/day: comparison with reference dose risperidone, 6mg/day. *Schizophr Res*. 2012;141(2–3):144–152.

Meltzer HY, Mills R, Revell S, et al. Pimavanserin, a serotonin(2A) receptor inverse agonist, for the treatment of Parkinson's disease psychosis. *Neuropsychopharmacology*. 2010;35(4):881–892.

Meltzer HY, Roth BL. Lorcaserin and pimavanserin: emerging selectivity of serotonin receptor subtype-targeted drugs. *J Clin Invest*. 2013;123(12):4986–

4991.

Nasrallah HA, Fedora R, Morton R. Successful treatment of clozapine-nonresponsive refractory hallucinations and delusions with pimavanserin, a serotonin 5HT-2A receptor inverse agonist. *Schizophr Res*. 2019;208:217–220.

Parkinson's Study Group. Low-dose clozapine for the treatment of drug-induced psychosis in Parkinson's disease. *N Engl J Med*. 1999;340(10):757–763.

Pimavanserin. National Library of Medicine. https://pubchem.ncbi.nlm.nih.gov/compound/Pimavanserin. Accessed August 3, 2021.

Ravina B, Marder K, Fernandez HH, et al. Diagnostic criteria for psychosis in Parkinson's disease: report of an NINDS, NIMH work group. *Mov Disord*. 2007;22(8):1061–1068.

Schubmehl S, Sussman J. Perspective on pimavanserin and the SAPS-PD: novel scale development as a means to FDA approval. *Am J Geriatr Psychiatry*. 2018;26(10):1007–1011.

Stahl SM. Mechanism of action of pimavanserin in Parkinson's disease psychosis: targeting serotonin 5HT2A and 5HT2C receptors. *CNS Spectr*. 2016;21(4):271–275.

Stahl SM. Stahl's essential pharmacology. *Neuroscientific Basis and Practical Applications*. 5th ed. Cambridge University Press. 2021.

Tariot PN, Cummings JL, Soto-Martin ME, et al. Trial of pimavanserin in dementia-related psychosis. *N Engl J Med*. 2021;385(4):309–319.

United States Food and Drug Administration. FDA briefing document for the March29, 2016 meeting of the Psychopharmacologic Drugs Advisory Committee. https://www.fda.gov/media/96750/download. Published March 29, 2016. Accessed August 11, 2021.

United States Food and Drug Administration. FDA analysis finds no new or unexpected safety risks associated with Nuplazid (pimavanserin), a medication to treat the hallucinations and delusions of Parkinson's disease psychosis. https://www.fda.gov/drugs/drug-safety-and-availability/fda-analysis-finds-no-new-or-unexpected-safety-risks-associated-nuplazid-pimavanserin-medication. Published September 20, 2018. Accessed August 11, 2021.

Vanover KE, Weiner DM, Makhay M, et al. Pharmacological and behavioral profile of N-(4-fluorophenylmethyl)-N-(1-methylpiperidin-4-yl)-N′-(4-(2-methylpropyloxy)phenylmethyl) carbamide (2R,3R)-dihydroxybutanedioate (2:1) (ACP-103), a novel 5-hydroxytryptamine(2A) receptor inverse agonist. *J Pharmacol Exp Ther*. 2006;317(2):910–918.

Weiner DM, Vanover KE, Brann MR, Meltzer HY, Davis RE. Psychosis of Parkinson's disease: serotonin 2A receptor inverse agonists as potential therapeutics. *Curr Opin Investig Drugs*. 2003;4(7):815–818.

33.8 Noradrenaline Modulators

33.8a α_2-Adrenergic Receptor Agonists: Clonidine, Clonidine Extended Release, Guanfacine, and Guanfacine Extended Release

Eric Hollander, M.D., and Migena Dervishi, M.D.

Neuroscience-Based Nomenclature (NbN):	
Clonidine, guanfacine	
Pharmacologic domain	Norepinephrine
Mechanism of action	Receptor agonist (alpha-2 NE)

HISTORY

Clonidine (Catapres) was synthesized in the early 1960s and found to produce vasoconstriction mediated by α-receptors. The drug was initially developed as a topical nasal decongestant. During clinical testing, it was found to cause hypotension, sedation, and bradycardia, and it was subsequently used as a centrally active antihypertensive. Due to its effects on norepinephrine and other neurotransmitter systems and the functioning of the prefrontal cortex (PFC), clonidine has been studied in several psychiatric disorders, including tic disorders, ADHD, PTSD, and opiate, nicotine, and alcohol withdrawal. Guanfacine (Tenex) is also a α-adrenergic receptor agonist. It has a longer half-life and may have fewer side effects due to its differential affinity for certain receptor subtypes.

CHEMISTRY

Clonidine is an imidazoline derivative and exists as a mesomeric compound. Its chemical name is *N*-(2,6-Dichlorophenyl)-4,5–1H-imidazol-2-amine; the structural formula is shown in Figure 33.8a–1. Clonidine hydrochloride is an odorless, bitter, white, crystalline substance soluble in water and alcohol. Its molecular weight is 266.56.

The chemical name of guanfacine is *N*-(diaminomethylidene)-2-(2,6-dichlorophenyl)acetamide, and its molecular weight is 282.56. Guanfacine hydrochloride is a white to off-white powder sparingly soluble in water and alcohol and slightly soluble in acetone. Guanfacine's structural formula is shown in Figure 33.8a–2.

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Absorption. Clonidine immediate release (CLON-IR) is well absorbed after oral administration, and its bioavailability is nearly 100%. The plasma level peaks between 1 and 3 hours after an oral dose. Clonidine extended-release (Kapvay) (CLON-XR) is well absorbed after oral administration with a bioavailability of about 90% but is designed to be absorbed more slowly with peak plasma levels achieved at 6 to 7 hours. Due to differing pharmacokinetic characteristics, utilization of the same dose of CLON-IR and CLON-XR will result in different drug exposure. Comparisons across studies suggest peak plasma concentrations are 50% lower for CLON-XR compared with CLON-IR. The absorption of CLON-XR is not impacted by food intake. Guanfacine immediate release (GUAN-IR) is also well absorbed orally with a bioavailability of about 80%, and peak plasma levels are seen in 1 to 4 hours. Guanfacine extended-release (Intuniv) (GXR) has a relative bioavailability of 57% compared with the GUAN-IR formulation. Therefore the XR and IR formulations are not interchangeable. GXR reaches peak plasma levels in 5 to 6 hours after oral administration. Absorption of GXR is significantly affected by food intake, with a 75% increase in peak plasma levels when administered with a high-fat meal. Administration with a high-fat meal should be avoided.

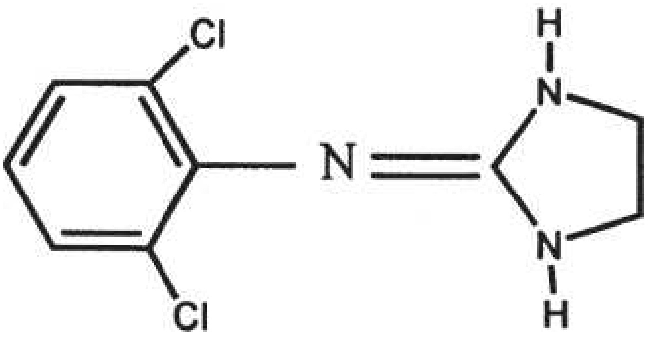

FIGURE 33.8a–1. Molecular structure of clonidine.

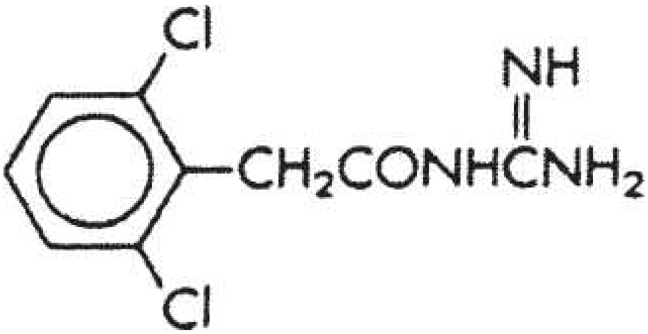

FIGURE 33.8a–2. Molecular structure of guanfacine.

Distribution. Clonidine and guanfacine formulations cross the blood–brain barrier, cross the placenta and are present in breast milk. Clonidine is between 20% and 40% protein bound and has a volume of distribution of 2.1 L/kg, whereas guanfacine is approximately 70% protein bound and has a volume of distribution of 6 L/kg. It is not known if guanfacine crosses the placenta or is excreted in breast milk, although guanfacine has been shown to cross the placenta in rats.

Metabolism and Elimination. The elimination half-life (time required for the concentration of the drug to reach half its original value) of both the immediate-release and extended-release clonidine range from 6 to 24 hours, with a mean of 12 hours. When renal function is severely impaired, the half-life can increase to approximately 41 hours. About 50% of the absorbed dose is metabolized in the liver. The elimination half-life of both guanfacine immediate release and guanfacine extended release is approximately 17 hours. Guanfacine is metabolized by hepatic cytochrome p450

(CYP) 3A4 microsomal enzymes. Exposure to guanfacine may potentially be increased by concurrent use of CYP3A4 inhibitors (clarithromycin, grapefruit juice, itraconazole) or decreased by concurrent use of CYP3A4 inducers (rifampin, St. John wort, phenytoin). If renal insufficiency clearance is reduced, drug levels are only slightly increased. Both clonidine and guanfacine are metabolized more rapidly in children and, therefore, the half-life is decreased.

Pharmacodynamics

Mechanism of Action. There are five adrenergic receptors (α_1, α_2, β_1, β_2, β_3). The α_1-receptors are located in the periphery, while the α_2-receptors are located centrally and peripherally. In the CNS, they are located both pre- and postsynaptically. The α_2-receptors have three distinct subtypes (A, B, and C) predominance. The α_2A-adrenoreceptors are concentrated in the PFC, locus ceruleus (LC), amygdala, septum, hippocampus, and hypothalamus. The α_2B-adrenoreceptor is mainly restricted to the thalamus. The α_2C-adrenoreceptor is present in the hippocampus, the olfactory system, the striatum, and the cerebral cortex. Each of these subtypes is a Gi/Go protein-coupled receptor, which, on a cellular level, when activated by endogenous agonists (epinephrine) or exogenous agonists (clonidine), suppresses adenylyl cyclase activity which then leads to decreased production of cAMP. This, in turn, decreases feedforward cAMP–calcium signaling and prevents the opening of potassium channels. Continued exposure to α_2-agonists leads to downregulation of all three receptor subtypes.

Clonidine behaves as a partial agonist at all three α_2-receptor subtypes. Guanfacine is a selective α_2A-adrenoreceptor agonist that has increased selectivity for the α_2A-receptor over α_2B- and α_2C-receptor subtypes. The majority of the pharmacologic and physiologic functions are attributed to the α_2A-adrenoreceptor subtype.

Presynaptic α_2A-adrenoreceptor activation leads to an autoinhibition of norepinephrine release with activation of presynaptic α_2-receptors in the LC and cortex and of α_2B-receptors in the thalamus, likely accounting for the sedative effects of the α_2-agonists. Guanfacine is less sedating than clonidine, which is likely due to a weaker action at α_2B- and α_2C-receptors. The majority of α_2A-receptors are postsynaptic, including receptors on the dendritic spines of pyramidal cells in the PFC. Activation of these receptors and the resulting decreases in cAMP enhances prefrontal neuronal firing and strengthens prefrontal network connectivity, allowing for greater control of behavior, including cognitive-enhancing effects. The activation of this postsynaptic receptor also has a stimulatory effect on growth hormone secretion. Less is known about the effects of agonists on the α_2B- and α_2C-adrenoreceptors.

Clonidine, an imidazoline derivative, also binds to imidazoline-binding sites (IBS) located throughout the body and in the CNS. There are three subtypes of IBS: I1, I2, and I3. Clonidine has a greater affinity for the I1 IBS, which is involved in the control of blood pressure, and there is evidence to support that the hypotensive effect of clonidine is mediated via the IBS I1. Hypotensive side effects are less common with guanfacine, likely due to its weaker affinity at the IBS I1.

Blood Concentrations and Relation to Action. Concentrations of both clonidine and guanfacine vary between individuals based on many factors, most notably age and renal function.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

The α_2-adrenergic receptor agonists have been evaluated in several psychiatric disorders with proposed underlying dysregulation of the adrenergic system or sympathetic nervous system. More recently, animal and neuroimaging research has led to further understanding of the neurobiology of PFC and the impact of the PFC and its connections on attention, emotion, and behavior. This has provided therapeutic targets to treat psychiatric disorders with a proposed underlying dysregulation of PFC. Various clinical drug trials have shown the α_2-adrenergic receptor agonists to have varying degrees of efficacy in the treatment of a variety of psychiatric disorders, including tic disorders, ADHD, PTSD, opiate, nicotine, and alcohol withdrawal. Over the past 5 years, with the development of the extended-release formulations of clonidine and guanfacine, there have been multiple larger randomized, placebo-controlled studies that have shown the extended-release α_2-agonists to be safe and efficacious for use in children and adolescents for the treatment of ADHD. In addition, the possible neuroprotective effects, especially in times of uncontrollable stress, of the α_2-adrenergic receptor agonists continue to be evaluated in humans because evidence of neuroprotective effects in animals has continued to expand.

EFFECT ON SPECIFIC ORGANS AND SYSTEMS

The α_2-agonists are active antihypertensive medications that directly affect the cardiovascular system by stimulating central α-adrenoreceptors, leading to reduced sympathetic nerve impulses from the vasomotor center to the heart and blood vessels. This results in a decrease in peripheral vascular resistance, renal vascular resistance, and a reduction in heart rate and blood pressure. Given these direct effects, measurement of heart rate and blood pressure prior to initiation, following dose increases and periodically while on therapy is recommended. Renal blood flow and glomerular filtration rate remain essentially unchanged. Normal hemodynamic responses to exercise are maintained. Both clonidine and guanfacine acutely stimulate growth hormone release in both children and adults but do not produce a chronic elevation of growth hormone with long-term use.

THERAPEUTIC INDICATIONS

Clonidine extended release and guanfacine extended release are FDA-approved in the treatment of ADHD as monotherapy and as adjunctive therapy to stimulant medication in children and adolescents. Immediate-release clonidine and guanfacine are FDA-approved only for the treatment of hypertension, as monotherapy, or in combination with other antihypertensive medications. These medications are also used in a number of off-label uses for the treatment of numerous psychiatric conditions.

Tourette Syndrome and Other Tic Disorders

A tic is a sudden, rapid, recurrent, nonrhythmic motor movement or vocalization. Tourette syndrome (TS) is a neuropsychiatric syndrome characterized by both multiple motor and one or more vocal tics that may wax and wane in frequency but have persisted for more than 1 year since first tic onset, with first tic onset before the age of 18 years.

Clonidine and guanfacine are considered by many to be the first-line treatment for tics and TS, with multiple positive double-blind, randomized placebo-controlled studies in the literature. Although haloperidol (Haldol) and pimozide (Orap) have been shown to be more effective than the α_2-agonists, concerns about their side-effect profiles, specifically NMS, tardive dyskinesia, and QTc prolongation (especially pimozide), have limited their use. Among the newer antipsychotics, the use of risperidone (Risperdal) has the strongest evidence for efficacy, yet its use is limited due to issues of weight gain, sedation, and risks of extrapyramidal symptoms and hyperprolactinemia. Preliminary data have shown clonidine and risperidone (Risperdal) to be equally efficacious. Preliminary data have also shown the clonidine transdermal patch to be an effective option for the treatment of TS and other tic disorders. When using clonidine for tics and TS, it is important to note that there can be a delayed response, sometimes taking 2 to 6 weeks to have a clinical effect.

A standard titration of clonidine for the treatment of tics is 0.05 mg before bed, increasing by 0.05 mg every 3 to 7 days until the dose is

0.05 mg in the morning and midday and 0.1 mg at bedtime. The total daily dose of clonidine for the treatment of tics is 0.1 to 0.3 mg. A standard titration of guanfacine for the treatment of tics is 0.5 mg before bed, increasing by 0.5 mg every 3 to 7 days until the dose is 1 mg twice per day. In studies of tics or TS and comorbid ADHD, the dose range was 3 to 5 µg/kg per day. Measurement of heart rate and blood pressure prior to initiation, following dose increases and periodically while on therapy is recommended.

There are no published studies on the use of the extended-release formulations of clonidine and guanfacine in TS or tic disorders.

Attention-Deficit/Hyperactivity Disorder

Dysregulation of the central noradrenergic network has long been hypothesized to be an important aspect of the pathophysiology of ADHD. Ongoing structural and functional imaging studies along with neuropsychological studies have demonstrated delayed or inadequate maturation of the PFC, particularly in the right inferior regions needed for inhibitory control and executive cognitive functioning. Moderate levels of norepinephrine are critical for proper PFC function, while high and low levels may impair function. Action at presynaptic α_2-receptors in the LC leads to decreased cell firing and release of norepinephrine modulating and optimizing catecholamine influences to remediate symptoms of ADHD. Ascending norepinephrine afferent projections from the LC to the PFC enhance functional connectivity between the PFC and other brain regions. Guanfacine mimics norepinephrine at the postsynaptic α_2A-receptors.

The α_2-agonist clonidine has been used in the treatment of ADHD for more than 20 years. Yet, the effect size for stimulant medications in ADHD is larger than that for the α_2-agonists, and clonidine and guanfacine are therefore considered second-line agents for the treatment of ADHD. The α_2-agonists are most often used for the augmentation of stimulants or when stimulant medications have been ineffective or poorly tolerated. In these patients, data have shown moderate efficacy for some symptoms of ADHD with clonidine and guanfacine monotherapy and with clonidine in combination with methylphenidate. There is increasing evidence that the addition of clonidine to ongoing psychostimulant treatment reduces symptoms of oppositional defiant disorder and conduct disorder, particularly symptoms of aggression.

A standard CLON-IR dose for the treatment of ADHD is approximately 3 to 5 µg/kg per day, with a daily dose range of 0.1 to 0.3 mg per day. It is recommended to start with low doses and increase by 0.05 mg every 3 days. A standard dose of GUAN-IR for the treatment of ADHD ranges from 0.5 to 4 mg per day. The dose can be increased by 0.5 mg every 3 to 4 days. CLON-IR is administered on a schedule of three to four times a day, while GUAN-IR, with its longer duration of action, is generally administered on a twice-a-day schedule.

For CLON-XR as monotherapy for ADHD or as adjunctive therapy to a psychostimulant, dosing should be initiated with 0.1 mg at bedtime and adjusted in increments of 0.1 mg per day at weekly intervals until the desired dose-response is achieved. Dosing should consist of twice a day with either an equal split or the higher dosage at bedtime to a maximal total dose of 0.4 mg per day. For GXR as monotherapy for ADHD or as adjunctive therapy to a psychostimulant, dosing should be initiated once daily, either morning or evening, with a dose of 1 mg per day and adjusted in increments of 1 mg per day at weekly intervals until the desired dose-response is achieved. Doses above 4 mg per day have not been systematically studied in controlled studies.

Concerns regarding the safety of the combination of stimulants and α_2-agonists remain after the sudden death of four children taking methylphenidate and clonidine was reported in 1995, although the specific contribution of clonidine in these deaths remains controversial. There have been no further reports of sudden death, and studies with children on this combination have been conducted with baseline, interval, and final visit ECGs and pulse and blood pressure monitoring. Before the initiation of an α_2-agonist, a cardiologist should evaluate patients with known cardiac problems or abnormal ECGs.

Opiate Withdrawal

Clonidine has been used for the treatment of opiate withdrawal since the 1980s and continues to be used in most detoxification protocols. Opiate withdrawal is manifested by symptoms of irritability, anxiety, sweating, sniffling, rhinorrhea, nausea, diarrhea, lacrimation, abdominal pain, headache, muscular pain, and insomnia. During withdrawal, neural activity in the LC is greatly increased, and this surge is thought to be responsible for many of the symptoms seen during withdrawal. Clonidine's ability, through its preferential action on presynaptic α_2-neurons, to inhibit noradrenergic activity has been consistently shown to be effective in decreasing these symptoms by counteracting catecholamine release and decreasing the sympathetic hyperactivity in opiate withdrawal.

Clonidine has been used as a single agent in the treatment of opiate withdrawal and in combination with methadone (Dolophine), or buprenorphine (Buprenex), a partial opiate agonist. Clonidine is also used in rapid opiate detoxification under anesthesia (RODA), which is a procedure using opiate antagonists, such as naloxone/naltrexone (Narcan/Revia), under general anesthesia to rapidly reverse opioid dependence. Preliminary evidence finds that clonidine may be particularly effective in managing the gastrointestinal (GI) side effects associated with RODA. Some authors have suggested that withdrawal-induced diarrhea may be mediated through α_2-adrenoreceptors located on intestinal smooth muscles.

Although studies have shown clonidine to be effective in the management of opiate withdrawal, in comparison to methadone and buprenorphine, the treatment retention rate of patients treated with clonidine only is markedly lower. This is believed to be because patients treated with clonidine, as compared to methadone or buprenorphine, experience the most intense symptoms of opiate withdrawal earlier in the detoxification process. In addition, these patients experience and report greater medication side effects. These side effects are sedation, lethargy, fatigue, dry mouth, and hypotension. Withholding doses or decreasing the clonidine dose according to blood pressure changes usually manages the hypotensive side effects. Although the treatment retention rates and the percentage of patients who complete detoxification is higher when methadone, buprenorphine, or RODA is used, some recent work has suggested that at 6- and 12-month follow-up, the relapse rates with these other medications are equal to or higher than, the clonidine only group.

Several standardized protocols for the use of clonidine in opiate withdrawal are commonly used. Generally, clonidine therapy is initiated at a dose of 0.1 to 0.2 mg orally up to six times per day and is titrated based on the severity of symptoms and side effects. Maximum doses of clonidine are usually approximately 1 mg per day. The maximal doses are used for 2 to 3 days during the most acute withdrawal period. When the patient is stabilized, the clonidine is tapered over a 7- to 10-day period. Some protocols involve a slower taper or transition to a transdermal patch in order to address chronic symptoms of withdrawal such as sleep disturbance.

Nicotine Withdrawal

Clonidine is considered a second-line treatment for nicotine withdrawal. Although it has been shown to be effective in promoting smoking cessation, there is a high incidence of adverse effects, such as hypotension, dry mouth, and sedation. In light of these adverse effects, clonidine is typically reserved for smokers in whom sedation would be a desirable effect, such as in a smoker who experiences extreme agitation or anxiety during nicotine withdrawal.

The recommended dose of clonidine for nicotine withdrawal from available studies is 0.1 mg twice a day with titration to 0.4 mg daily as tolerated. Clonidine should be started 48 to 72 hours prior to cessation of nicotine use to allow time to reach a steady state before the onset of nicotine withdrawal.

Autism Spectrum Disorders

In clinical practice, α_2-agonists are considered a second-line treatment for managing symptoms of hyperactivity, inattention, impulsivity, aggression, explosive outbursts, self-injury, and sleep dysfunction in autism spectrum disorders.

Overall empirical support for the use of the α-agonists in autism spectrum disorders is limited. There are three published trials of guanfacine in children with autism spectrum disorders. Preliminary evidence supports the efficacy and safety of guanfacine in the treatment of hyperactivity, inattention, and aggression in particular subsets of the autism spectrum disorders population. Clonidine has been studied as a sedative agent for autism spectrum disorder patients undergoing electroencephalogram (EEG), as a treatment for hyperactivity and inattention, and for sleep disturbance in these patients. Of note, individuals with autism spectrum disorders are generally more sensitive to the side effects of irritability and sedation when treated with α_2-agonists. There are case reports of the use of extended-release guanfacine in children with autism spectrum disorders for symptoms of inattention and hyperactivity.

Posttraumatic Stress Disorder and Other Anxiety Disorders

Increased and altered noradrenergic neurotransmission and reactivity in the CNS have been associated with a variety of hyperarousal symptoms and sleep disturbances in PTSD. Clonidine’s action at the postsynaptic α_2-adrenergic receptors in the CNS reduces sympathetic activity, possibly reducing the arousal symptoms of PTSD. Guanfacine, with the same mechanism as clonidine, has also been used for the treatment of the symptoms of PTSD, including nightmares. Clinically there are reports of effectiveness in child refugees and veterans, although there are no randomized, placebo-controlled trials of clonidine in PTSD, and there are reports that it loses its effectiveness over time.

Clonidine has also been studied in other anxiety disorders. In patients with panic disorder, clonidine appears to have an acute anxiolytic effect that is transient. Some debate exists as to whether clonidine has anxiolytic effects by proxy of its sedative side effects. The α_2-agonists do not appear to be effective in OCD.

Cognition and Neuroprotective Effects

The relationship between the noradrenergic system and executive and memory functions has been extensively investigated in animal models with positive results. Investigations of the cognitive enhancing potential of α_2-agonists via their effects on the PFC in humans have been inconclusive, and more recent studies found no significant clinical effects on healthy adult controls. There is evidence for the use of clonidine and guanfacine to enhance cognition in patients with schizophrenia and Korsakoff syndrome; however, its use remains limited by issues of tolerability.

Research on the neuroprotective effects of α_2-adrenergic receptor agonists began approximately 20 years ago after studies in rats indicated a protective effect of clonidine after cerebral ischemia. These neuroprotective effects have also been seen with the α_2-agonist brimonidine (Alphagan P) in glaucoma. At this time, despite extensive animal research, minimal investigation has been done to evaluate the neuroprotective effects of these drugs in humans.

Other Disorders

Clonidine has been used in the treatment of alcohol withdrawal to target the symptoms thought to be secondary to adrenergic hyperactivity and CNS excitation. There is no evidence suggesting that these medications are effective in reducing the incidence of seizures or delirium associated with alcohol withdrawal, thus reducing their use as an adjunctive treatment to benzodiazepines and typical antipsychotics. Animal studies have also looked at clonidine in decreasing voluntary alcohol drinking. There are also reports of using low-dose clonidine (0.025 mg three times per day) in hallucinogen persisting perception disorder, where symptoms may resolve within 2 months of initiating treatment.

There has also been a positive small double-blind, randomized placebo-controlled trial investigating the use of clonidine for hyperarousal symptoms in patients with borderline personality disorder with and without PTSD.

PRECAUTIONS AND ADVERSE REACTIONS

Adverse effects of clonidine include dry mouth (40%), drowsiness (33%), dizziness (16%), constipation (10%), and sedation (10%). There are also numerous, albeit less frequent, side effects that have been reported in which a causal relationship has not been established, including nervousness, agitation, sleep disturbance, impotence, loss of libido, weight gain, headache, nausea, and vomiting. There are isolated reports of transient abnormalities in liver function tests, urinary dysfunction, thrombocytopenia, transient elevation of blood glucose, muscle pain, and leg cramps. Abnormalities on ECG such as sinus node arrest, junctional bradycardia, high-degree AV block, and arrhythmias have also been reported. Marked bradycardia is observed in some patients. Localized contact dermatitis is an adverse reaction to the clonidine transdermal therapeutic system (TTS) for approximately 15% to 20% of patients, particularly when using the optional adhesive cover. This can be addressed by alternating the site of the patch. The clonidine TTS contains aluminum and, therefore, should be removed before undergoing magnetic resonance imaging (MRI). Adverse reactions with the use of extended-release clonidine include somnolence (30%), fatigue (15%), constipation (5%), and nightmares (3% to 9%), although there is some variability depending on the dose. For guanfacine, the adverse effects are similar to those for clonidine, although hypotension, sedation, and dry mouth are somewhat less frequent. Adverse reactions with the use of extended-release guanfacine include somnolence (35%), fatigue (14%), and hypotension (7%). For these medications, there is evidence that the severity of dose-dependent adverse effects lessens over time.

Clonidine is considered a class C drug in pregnancy, which means that the safety is uncertain and that well-controlled studies in humans do not exist. Clonidine is excreted in breast milk, and caution should be exercised when administering clonidine in pregnancy and during lactation. Studies have not found evidence to suggest carcinogenesis and mutagenesis as a result of clonidine. Guanfacine is considered a class B drug in pregnancy, meaning that studies in humans have shown no evidence of risk to the fetus, although there have been adverse findings in animal studies. The excretion of guanfacine in breast milk is unknown, but in light of the fact that most medications are eliminated in breast milk, caution is suggested.

Clonidine is known to have a withdrawal syndrome. With the sudden cessation of clonidine, there may be an associated increase in plasma and urinary catecholamines and associated symptoms such as nervousness, restlessness, insomnia, headache, increased heart rate, sweating, increased salivation, nausea, and flushing. There can also be episodes of rebound hypertension upon cessation of the medication. This syndrome is more pronounced after abrupt cessation of long-term therapy, in patients on a higher total dose, and in patients also receiving β-blockers. A taper of the clonidine is required when the duration of use has been greater than 1 week, with a reduction in dose of 0.05 mg every other day. When the duration of use has been longer than 4 weeks, the dose reduction should be every third day. If there are symptoms of withdrawal present, they can be alleviated with clonidine. There are less data regarding a withdrawal syndrome with guanfacine, and although it remains a possibility, its longer duration of action makes withdrawal less likely.

DRUG INTERACTIONS

Clonidine and guanfacine cause sedation, especially early in therapy, and when administered with other centrally active depressants, such as barbiturates, alcohol, and benzodiazepines, the potential for additive sedative effects should be considered, and dose reduction may be required. Due to the potential for additive effects, caution is required in patients receiving agents known to affect AV node and sinus node conduction, such as β-blockers, calcium channel blockers, and digitalis. This combination increases the risk of AV block and bradycardia. When utilized along with antihypertensive drugs, use caution as there is the potential for additive effects. Guanfacine is metabolized by CYP3A4 microsomal enzymes, and plasma concentrations can be affected significantly by CYP3A4 inhibitors or inducers. Guanfacine does not significantly affect exposures of methylphenidate and lisdexamfetamine when coadministered.

LABORATORY INTERFERENCE

Neither clonidine nor guanfacine is known to interfere with any laboratory tests.

DOSAGE AND ADMINISTRATION

Clonidine is supplied in 0.1-, 0.2-, and 0.3-mg tablets. It is also available in TTS that delivers 0.1, 0.2 or 0.3 mg daily (TTS-1, TTS-2, TTS-3). The dosing for the TTS is equivalent to the total oral daily dose. Multiple patches can be used at one time in order to administer the required dose. When administered for the treatment of hypertension, clonidine is initiated at 0.1 mg twice daily with increases of 0.1 mg at weekly intervals until the desired response is achieved. Therapeutic doses for the treatment of hypertension range from 0.2 to 0.6 mg per day in divided doses. The maximum effective daily dose is 2.4 mg. It is suggested to administer a larger portion of the oral daily dose at bedtime to minimize adverse effects of dry mouth and sedation. The clonidine TTS is applied once every 7 days to a clean, dry, hairless area of intact skin on the upper outer arm, chest, or back. In children, a new patch is applied every 5 days due to their more rapid hepatic metabolism. Each new application should be on a different skin site from the prior patch. If the clonidine TTS becomes loose from the skin, the adhesive cover can be placed over the system to ensure adhesion.

Clonidine extended release is available in 0.1- and 0.2-mg tablets. The tablets must be swallowed whole and never cut, crushed, or chewed.

Guanfacine is supplied in 1- and 2-mg tablets. When administered for the treatment of hypertension, guanfacine is initiated at 1 mg at bedtime to reduce sedation. The daily dose may be raised to 2 to 3 mg after 3 to 4 weeks of therapy. For the treatment of hypertension, the most clinical effect is seen in the 1- to 3-mg dose range, although the maximum recommended dose is 6 mg daily.

Guanfacine extended release is available in 1-, 2-, 3-, and 4-mg tablets. The tablets must be swallowed whole and never cut, crushed, or chewed.

FURTHER READINGS

Arnsten AF. The use of α$_2$A adrenergic agonists for the treatment of attention-deficit/hyperactivity disorder. *Expert Rev Neurother*. 2010;10:1595–1605.

Arnsten AF, Jin LE. Guanfacine for the treatment of cognitive disorders: a century of discoveries at Yale. *Yale J Biol Med*. 2012;85(1):45–58.

Blankenship K, Erickson CA, Stigler KA, Posey DJ, McDougle CJ. Guanfacine extended release in two patients with pervasive developmental disorders. *J Child Adolesc Psychopharmacol*. 2011;21(3):287–290.

Bukstein OG, Head J. Guanfacine ER for the treatment of adolescent attention-deficit/hyperactivity disorder. *Expert Opin Pharmacother*. 2012;13(15):2207–2213.

*Connor DF, Arnsten AF, Pearson GS, Greco GF. Guanfacine extended release for the treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Expert Opin Pharmacother*. 2014;15(11):1601–1610.

Connor DF, Findling RL, Kollins SH, et al. Effects of guanfacine extended release on oppositional symptoms in children aged 6–12 years with attention-deficit hyperactivity disorder and oppositional symptoms: a randomized, double-blind, placebo-controlled trial. *CNS Drugs*. 2010;24(9):755–768.

Connor DF, Grasso DJ, Slivinsky MD, Pearson GS, Banga A. An open-label study of guanfacine extended release for traumatic stress related symptoms in children and adolescents. *J Child Adolesc Psychopharmacol*. 2013;23(4):244–251.

Croxtall JD. Clonidine extended-release: in attention-deficit hyperactivity disorder. *Pediatr Drugs*. 2011;13(5):329–336.

Daviss WB, Patel NC, Robb AS, et al. Clonidine for attention-deficit/hyperactivity disorder: II. ECG changes and adverse events analysis. *J Am Acad Child Adolesc Psychiatry*. 2008;47(2):189–198.

Elbe D, Reddy D. Focus on guanfacine extended-release: a review of its use in child and adolescent psychiatry. *J Can Acad Child Adolesc Psychiatry*. 2014;23(1):48–60.

Findling RL, McBurnett K, White C, Youcha S. Guanfacine extended release adjunctive to a psychostimulant in the treatment of comorbid oppositional symptoms in children and adolescents with attention-deficit/hyperactivity disorder. *J Child Adolesc Psychopharmacol*. 2014;24(5):245–252.

Fiscella K, Moore A, Engerman J, Meldrum S. Management of opiate detoxification in jails. *J Addict Dis*. 2005;24(1):61–71.

Fresquez-Chavez KR, Fogger S. Reduction of opiate withdrawal symptoms with use of clonidine in a county jail. *J Correct Health Care*. 2015;21(1):27–34.

*Gamo NJ, Arnsten AF. Molecular modulation of prefrontal cortex: rational development of treatments for psychiatric disorders. *Behav Neurosci*. 2011;125(3):282–296.

Gourlay SG, Stead LF, Benowitz NL. Clonidine for smoking cessation. *Cochrane Database Syst Rev*. 2004;(3):CD000058.

Gowing L, Farrell M, Ali R, White JM. Alpha$_2$-adrenergic agonists for the management of opioid withdrawal. *Cochrane Database Syst Rev*. 2016;3:CD002024.

*Hirota T, Schwartz S, Correll CU. Alpha-2 agonists for attention-deficit/hyperactivity disorder in youth: a systematic review and meta-analysis of monotherapy and add-on trials to stimulant therapy. *J Am Acad Child Adolesc Psychiatry*. 2014;53(2):153–173.

Jain R, Segal S, Kollins SH, Khayrallah M. Clonidine extended-release tablets for pediatric patients with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*. 2011;50(2):171–179.

Kollins SH, Jain R, Brams M, et al. Clonidine extended-release tablets as add-on therapy to psychostimulants in children and adolescents with ADHD. *Pediatrics*. 2011;127:e1406–e1413.

Müller U, Clark L, Lam ML, et al. Lack of effects of guanfacine on executive and memory functions in healthy male volunteers. *Psychopharmacology (Berl)*. 2005;182:205–213.

Newcorn JH, Stein MA, Childress AC, et al. Randomized, double-blind trial of guanfacine extended release in children with attention-deficit/hyperactivity disorder: morning or evening administration. *J Am Acad Child Adolesc Psychiatry*. 2013;52(9):921–930.

Palumbo DR, Sallee FR, Pelham WE Jr, et al . Clonidine for attention-deficit/hyperactivity disorder: I. Efficacy and tolerability outcomes. *J Am Acad Child Adolesc Psychiatry*. 2008;47(2):180–188.

*Rizzo R, Gulisano M, Calì PV, Curatolo P. Tourette Syndrome and comorbid ADHD: current pharmacological treatment options. *Eur J Paediatr Neurol*. 2013;17:421–428.

*Sallee F, Connor DF, Newcorn JH. A review of the rationale and clinical utilization of α2-adrenoceptor agonists for the treatment of attention-deficit/hyperactivity and related disorders. *J Child Adolesc Psychopharmacol*. 2013;23(5):308–319.

Sallee FR, McGough J, Wigal T, et al. Guanfacine extended release in children and adolescents with attention-deficit/hyperactivity disorder: a placebo-controlled trial. *J Am Acad Child Adolesc Psychiatry*. 2009;48(2):155–165.

*Stigler KA. Psychopharmacologic management of serious behavioral disturbance in ASD. *Child Adolesc Psychiatr Clin N Am*. 2014;23:73–82.

Strawn JR, Geracioti TD Jr. Noradrenergic dysfunction and the psychopharmacology of posttraumatic stress disorder. *Depress Anxiety*.

2008;25(3):260–271.

*Waldon K, Hill J, Termine C, Balottin U, Cavanna AE. Trials of pharmacological interventions for Tourette syndrome: a systematic review. *Behav Neurol*. 2013;26:265–273.

Weber B, Steinfath M, Scholz J, Bein B. Neuroprotective effects of alpha 2-adrenergic receptor agonists. *Drug News Perspect*. 2007;20(3):149–154.

Ziegenhorn AA, Roepke S, Schommer NC, Merkl A, Danker-Hopfe H. Clonidine improves hyperarousal in borderline personality disorder with or without comorbid posttraumatic stress disorder: a randomized, double-blind, placebo-controlled trial. *J Clin Psychopharmacol*. 2009;29:170–173.

33.8b β-Adrenergic Receptor Antagonists

Migena Dervishi, M.D., and Eric Hollander, M.D.

HISTORY

The first nontoxic β-adrenergic receptor antagonist, propranolol, was isolated by the Scottish pharmacologist James W. Black, earning him the Nobel Prize in Medicine in 1988. β-Adrenergic receptors are ubiquitously distributed throughout the body, including the heart, lung, autonomic, and CNS. The five β-adrenergic receptor antagonists most frequently studied for psychiatric application are propranolol (Inderal), metoprolol (Lopressor, Toprol XL), atenolol (Tenormin), nadolol (Corgard), and pindolol (Visken).

CHEMISTRY

Propranolol is chemically described as 1-(isopropylamino)-3-(naphthalene-1-yloxy) propan-2-ol. Its molecular formula is $C_{16}H_{21}NO_2$. Propranolol is a stable, white, crystalline solid that is readily soluble in water and ethanol. Its molecular mass is 295.34. Its molecular structure is shown in Figure 33.8b–1.

Metoprolol tartrate is chemically described as (1-[4-(2-methoxyethyl) phenoxy]-3-propan-2-ylamino-propan-2-ol). It is a white, practically odorless, crystalline powder with a molecular mass of [267.36] 684.82. Its molecular formula is $C_{15}H_{25}NO_3$. It is very soluble in water, freely soluble in methylene chloride, chloroform, and alcohol; slightly soluble in acetone; and insoluble in ether. The molecular structure is shown in Figure 33.8b–1.

Atenolol may be chemically described as benzeneacetamide, 2-[4-[2-hydroxy-3-(1-methylethylamine)propoxy]phenyl]ethanamide. The molecular formula is $C_{14}H_{22}N_2O_3$. Atenolol has a molecular weight of 266.34. It is a relatively polar hydrophilic compound with a water solubility of 26.5 mg/mL at 37°C and a log partition coefficient (octanol/water) of 0.23. It is freely soluble in 1N hydrochloric acid (300 mg/mL at 25°C) and less soluble in chloroform (3 mg/mL at 25°C). The molecular structure is given in Figure 33.8b–1.

Nadolol is chemically described as 2R,3S 5-[2-hydroxy-3-3(tert-butylamino)-propoxy]tetralin-2,3-diol. Its molecular formula is $C_{17}H_{27}NO_4$. Nadolol has a molecular weight of 309.40. It is a white, crystalline powder that is freely soluble in alcohol, propylene glycol, and ethanol; slightly soluble in chloroform and water; very slightly soluble in sodium hydroxide; and insoluble in acetone, benzene, ether, and hexane. The molecular structure is shown in Figure 33.8b–1.

Pindolol is chemically described as 1-(1H-indol-4-ylox)-3-(1-methylethylamino)propan-2-ol. Its molecular formula is $C_{14}H_{20}N_2O_2$. Pindolol has a molecular mass of 248.32. Pindolol is a white to off-white crystalline powder with a faint odor. It is practically insoluble in water, slightly soluble in methanol, and very slightly soluble in chloroform. The molecular structure is shown in Figure 33.8b–1.

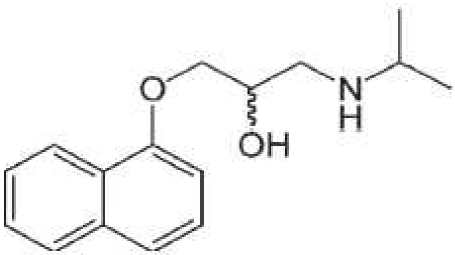

Propranolol

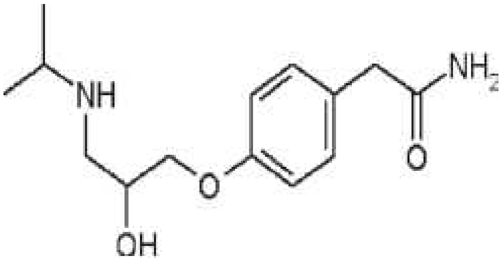

Atenolol

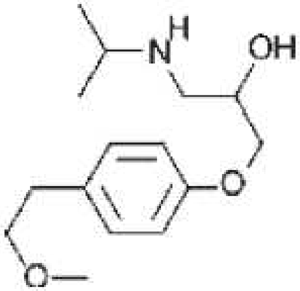

Metoprolol

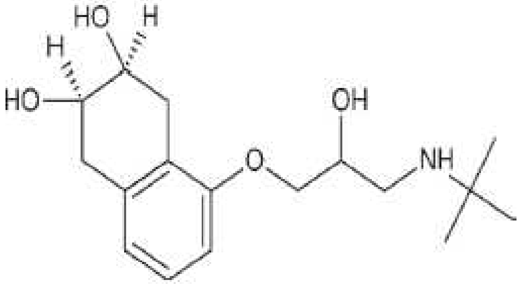
HO
H
H
HO
OH
O
NH

Nadolol

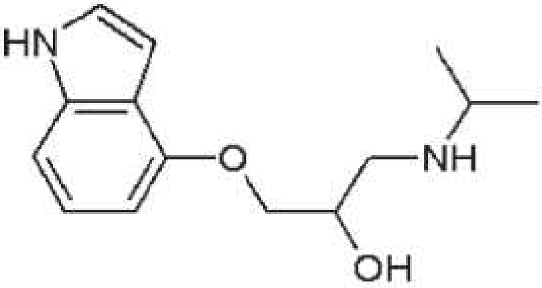

Pindolol

FIGURE 33.8b–1. Molecular structures of β-adrenergic receptor antagonists.

PHARMACOLOGIC ACTIONS

Pharmacokinetics

The pharmacokinetics of atenolol, metoprolol, nadolol, propranolol, and pindolol are listed in Table 33.8b–1.

Absorption. Absorption of propranolol (>95%) and pindolol (~95%) is extensive. Atenolol (~50%) and metoprolol (~50%) are intermediate, while absorption for nadolol is variable, averaging approximately 30%.

Table 33.8b–1. Pharmacokinetics

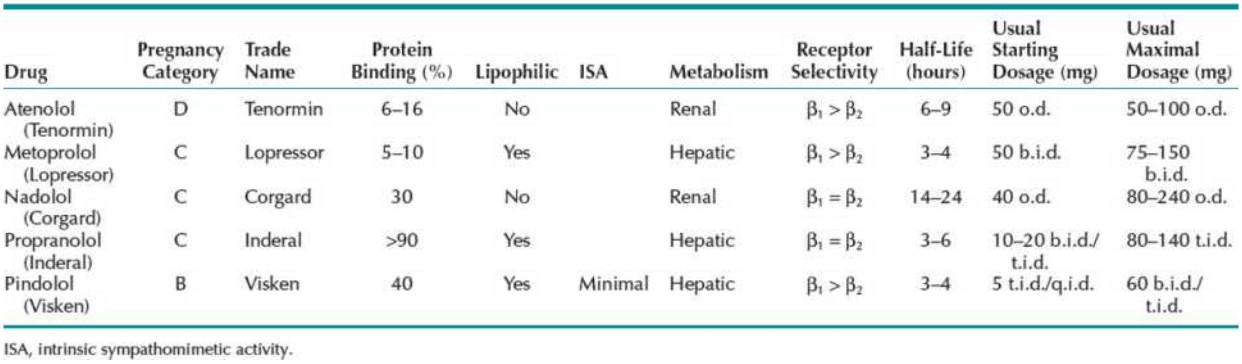

Drug	Pregnancy Category	Trade Name	Protein Binding (%)	Lipophilic	ISA	Metabolism	Receptor Selectivity	Half-Life (hours)	Usual Starting Dosage (mg)	Usual Maximal Dosage (mg)
Atenolol (Tenormin)	D	Tenormin	6–16	No		Renal	$\beta_1 > \beta_2$	6–9	50 o.d.	50–100 o.d.
Metoprolol (Lopressor)	C	Lopressor	5–10	Yes		Hepatic	$\beta_1 > \beta_2$	3–4	50 b.i.d.	75–150 b.i.d.
Nadolol (Corgard)	C	Corgard	30	No		Renal	$\beta_1 = \beta_2$	14–24	40 o.d.	80–240 o.d.
Propranolol (Inderal)	C	Inderal	>90	Yes		Hepatic	$\beta_1 = \beta_2$	3–6	10–20 b.i.d./t.i.d.	80–140 t.i.d.
Pindolol (Visken)	B	Visken	40	Yes	Minimal	Hepatic	$\beta_1 > \beta_2$	3–4	5 t.i.d./q.i.d.	60 b.i.d./t.i.d.

ISA, intrinsic sympathomimetic activity.

Distribution. β-Adrenergic receptor antagonists are categorized based on their lipophilicity. Propranolol and metoprolol are highly lipophilic and, as a result, have both potent central and peripheral effects. Pindolol is of intermediate lipophilicity; nadolol and atenolol are the least lipophilic and have minimal central effects.

Propranolol exhibits extensive first-pass metabolism, with approximately 25% of an orally administered dose reaching the systemic circulation. Metoprolol is extensively metabolized by cytochrome P450 enzymes of the liver. With repeated oral administration, it increases in a dose-dependent fashion. Metoprolol exhibits approximately 50% first-pass metabolism and relatively low protein binding (10% to 15%).

Atenolol plasma levels are proportional to the dose over the range of 50 to 400 mg with relatively low protein binding (5% to 16%).

Pindolol exhibits substantial interindividual variation in plasma concentrations. Pindolol has relatively low protein binding (~40%). In man, 35% to 40% is excreted unchanged in the urine and 60% to

65% is metabolized primarily to hydroxy-metabolites, which are excreted as glucuronides and ethereal sulfates. Approximately 30% of nadolol in serum is reversibly bound to plasma protein and is not metabolized by the liver but is excreted unchanged principally by the kidneys.

Elimination. The β-adrenergic receptor antagonists exhibit considerable variation in elimination half-life (see Table 33.8b–1).

PHARMACODYNAMICS

Mechanism of Action

β-Adrenergic receptor antagonists block the action of endogenous catecholamines (primarily norepinephrine, as well as epinephrine) on β-adrenergic receptors. There are three types of β-adrenergic receptors, each of which exhibits tissue selectivity. β_1-Adrenergic receptors are localized in the myocardium and kidney; β_2-adrenergic receptors are located in the gastrointestinal tract, lung, liver, uterus, and vascular smooth muscles; and β_3-adrenergic receptors are located in fat cells. Norepinephrine is produced in the CNS by neurons in the locus ceruleus, a phylogenetically old monoamine system located in the dorsal pons with diffuse projections to subcortical and cortical regions.

There are also three types of α_1- and α_2-adrenoreceptors. Both α- and β-adrenergic receptors are G-protein–coupled receptors. β-Adrenergic receptor antagonists are categorized as selective or nonselective on the basis of their affinity for β_1 and β_2 receptors, respectively. For example, propranolol and nadolol are nonselective, whereas atenolol, pindolol, and metoprolol exhibit differential affinity for the β_1-adrenergic receptor.

The β-adrenergic system participates in the formation and consolidation of emotional memories in humans. Pre- and postlearning stimulation of the β-adrenergic system enhances memory for the events that trigger stressful experiences, whereas β-adrenergic receptor antagonists administered prior to learning may inhibit declarative stressful memories. Propranolol affects context conditioning but has less impact on the acquisition or retention of cued fear conditioning.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

β-Adrenergic receptor antagonists are used in psychiatry to treat a variety of conditions. They are most frequently used as a treatment for performance anxiety, lithium-induced postural tremor, and neuroleptic-induced akathisia. β-Adrenergic receptor antagonists are also used occasionally as adjunctive agents for the treatment of major depressive episodes, anxiety disorders, notably performance-based anxiety and panic attacks, aggression, behavioral disorders in the older adults, posttraumatic stress, and alcohol and cocaine withdrawal.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

The ubiquity of β-adrenergic receptors throughout the body implies a panoply of physiologic effects on organs and systems (Tables 33.8b–2 and 33.8b–3). The systems most pertinent are cardiovascular, respiratory, and metabolic. Psychiatric populations commonly treated with β-adrenergic receptor antagonists (e.g., individuals with mood, anxiety, and psychotic disorders) are differentially affected by disorders in each of these systems. For individuals without pre-existing cardiovascular disease (e.g., hypertension), β-adrenergic receptor antagonists rarely cause hypotension at doses prescribed. Individuals with pre-existing hypertension, however, exhibit an increased propensity to the hypotensive effects of β-adrenergic receptor antagonists.

β-Adrenergic receptor antagonists minimally affect pulmonary function in healthy persons, but they may result in severe life-threatening bronchoconstriction in patients with chronic obstructive pulmonary disease and asthma. As β-adrenergic receptor antagonists may alter the metabolism of carbohydrates and lipids and mask the premonitory signs of hypoglycemia, they should be used with caution in individuals with diabetes mellitus.

TERATOGENIC EFFECTS

None of the available β-adrenergic receptor antagonists has been adequately tested with controlled trials in pregnant women. Perinatal complications (e.g., placental insufficiency and intrauterine growth retardation) are reported with propranolol. Some infants born to mothers administered propranolol during pregnancy exhibited hypoglycemia and/or bradycardia, hypotonia and intrauterine growth retardation.

Table 33.8b–2.

Side Effects and Toxicity of β-Adrenergic Receptor Antagonists

- Cardiovascular
  - Hypotension
  - Bradycardia
  - Congestive heart failure (in patients with compromised myocardial function)
- Gastrointestinal
  - Anorexia
  - Nausea
  - Vomiting
  - Diarrhea
  - Abdominal pain
- Metabolic
  - Weight gain
  - Worsened hypoglycemia in diabetic patients on insulin or oral agents
- Respiratory
  - Asthma (less risk with β_1-selective drugs)
  - Shortness of breath
  - Wheezing
- Sleep disturbances
  - Vivid dreams and nightmares with or without insomnia
- Central nervous system
  - Dizziness
  - Drowsiness
  - Lassitude
  - Fatigue
  - Dysphoria
  - Bronchospasm
  - Lethargy
  - Depression
  - Psychosis (rare)
- Dermatologic
  - Rashes
  - Urticaria
  - Sweating

Sexual function
    Impotence
Other (rare)
    Raynaud phenomenon
    Peyronie disease
Withdrawal syndrome
    Rebound worsening of pre-existing angina pectoris when β-adrenergic receptor antagonists are discontinued

---

β_1-Selective adrenergic antagonists such as atenolol or metoprolol may be preferred as they interfere less with β_2-receptor–mediated uterine relaxation.

Both metoprolol and pindolol cross the placental barrier; they are excreted in breast milk in small amounts. Transplacental passage of atenolol also occurs in pregnant women; fetal drug serum levels equal to the mother have been reported. Patients administered atenolol in the third trimester of pregnancy are at risk for low–birth-weight newborns and placental insufficiency. Nadolol crosses the placenta and is distributed into breast milk.

Table 33.8b–3.
β-Adrenergic Receptor Antagonist Preparations

	Tablets (mg)	Extended Release (mg)	Solution	Parenteral
Atenolol	25, 50, 100			
Metoprolol	50, 100	200		1 mg/mL
Nadolol	20, 40, 80, 120, 160			
Propranolol	10, 20, 40, 80, 120	60, 80, 120, 160		1 mg/mL
Pindolol	5, 10, 15			

From Arana GW, Santos AB. β-adrenergic receptor antagonists. In: Kaplan HI, Sadock BJ, eds. *Comprehensive Textbook of Psychiatry*. 6th ed. Williams & Wilkins; 1995.

Table 33.8b–4.
Indications for β-Adrenergic Receptor Antagonists in Psychiatric Practice

- Effective
  - Social phobia (performance type)
  - Obsessive-compulsive disorder
  - Lithium-induced postural tremor
  - Neuroleptic-induced acute akathisia
  - Accelerate antidepressant action
- Probably effective
  - Control of impulsive aggression or violence
  - Adjunct to benzodiazepines in treatment of alcohol withdrawal
  - Adjunct in treatment of cocaine and benzodiazepine withdrawal
- Possibly effective
  - Generalized anxiety disorder
  - Schizophrenia
  - Treatment-resistant depression
  - Panic disorder
  - Posttraumatic stress disorder

THERAPEUTIC INDICATIONS

Indications for β-adrenergic receptor antagonists are listed in Table 33.8b–4.

Lithium Tremor

Lithium may result in an action tremor at nontoxic therapeutic levels. A stable lithium tremor in the absence of toxicity may, however, be distressing and impairing to the individual and require intervention. Therapeutic guiding principles are to ensure adequate blood levels necessary to achieve efficacy and to ensure that no other concomitant agents are contributory. If tremor persists, propranolol is most often used at a dose of 20 to 160 mg per day given in b.i.d. or t.i.d. dosing.

Major Depressive Disorder

Pharmacologic blockade of 5-HT_{1A} autoreceptors is thought to mimic the effect of autoreceptor desensitization and hasten the antidepressant effect of SSRIs. Results from open and randomized controlled clinical trials indicate that pindolol may accelerate the onset of antidepressant action in individuals with major depressive disorder. Pindolol’s efficacy in treatment-resistant major depressive disorder, however, is less consistent. One possible explanation for inconsistent efficacy with pindolol may relate to drug–drug interactions as a consequence of a patient’s pharmacokinetic profile (i.e., slow vs. fast metabolizer).

The possibility that pindolol dosing in resistant populations was suboptimal (i.e., at 2.5 mg t.i.d.) has not been excluded. It is presumed that pindolol’s mechanism of action relates to its affinity for somatic 5-HT_{1A} autoreceptors. When combined with SSRIs, pindolol also increases rapid eye movement (REM) suppression, notably an increase in REM onset latency and reduction in slow-wave sleep.

Substance Use Disorders

Available data suggest that blockade of β-adrenergic receptors potentiates cocaine-induced elevation of dopamine transmission in the nucleus accumbens, which is associated with an increase in cocaine-induced motor activity and a decrease in cocaine self-administration. Propranolol may attenuate the responsivity of noradrenergic neurons, resulting in decreased anxiety from abstinence from cocaine, alcohol, and benzodiazepines. Propranolol may also selectively block memories associated with arousal in humans, which is salient to cue-triggered relapses in substance use disorders. Propranolol is superior to a placebo in reducing cocaine withdrawal symptoms but is not reliably efficacious in promoting abstinence from cocaine.

β-Adrenergic receptor antagonists have been used adjunctively for alcohol withdrawal and relapse prevention. β-Adrenergic receptor antagonists do not replace benzodiazepines for alcohol withdrawal. β-Adrenergic receptor antagonists are not administered during alcohol withdrawal in individuals with a heart rate of fewer than 50 beats per minute.

Violence

Results from case reports and randomized, placebo-controlled trials indicate that β-adrenergic receptor antagonists are modestly effective and well tolerated for individuals with impulsive violence and disruptive behaviors associated with impulsive control disorders, brain injury, neurodegenerative disorder, and general medical conditions affecting the CNS. In individuals with psychological impairment and severe aggression, propranolol 40 to 600 mg per day in divided doses has frequently been used as an adjuvant to antipsychotics, anticonvulsants, and lithium.

Performance Anxiety

The use of propranolol for performance anxiety as part of social phobia is the most familiar application of β-adrenergic receptor antagonists in anxiety disorders. Performance anxiety includes symptoms of dry mouth, coarse voice, increased heart rate, difficulty breathing, tremor, weakness, and dizziness. Single dosing of propranolol 10 to 40 mg per day given approximately 20 to 30 minutes before the performance is most frequently employed. It is generally recommended to consider a test dose prior to the performance event. The mechanism of action of β-adrenergic receptor antagonists relates to their effect on the peripheral sympathetic nervous system. In keeping with that view, less lipophilic agents may also be effective.

Anxiety

It has been postulated that disturbances in β-adrenergic receptor activity are associated with several anxiety disorders, notably panic disorder. Moreover, the anxiety behaviors are subserved by neural structures and circuits that involve the locus ceruleus. Several randomized, controlled trials have established the efficacy of β-adrenergic receptor antagonists in the treatment of disparate anxiety disorders. Pindolol 2.5 mg per day t.i.d. has been used successfully as an adjunct to SSRIs in the treatment of OCD. For generalized anxiety disorder and panic attacks, propranolol doses of 40 to 320 mg per day have been employed.

Neuroleptic-Induced Akathisia

β-Adrenergic receptor antagonists may be more efficacious than anticholinergics and benzodiazepines for medication-induced movement disorders such as akathisia. β-Adrenergic receptor antagonists are not considered efficacious for other neuroleptic-induced movement disorders such as acute dystonia and parkinsonism. In some resistant cases, β-adrenergic receptor antagonists may be combined with benzodiazepines and anticholinergics.

Precautions and Adverse Reactions

Symptoms of propranolol overdose include bradycardia, cardiac failure, hypotension, and bronchospasm. For patients with asthma or chronic obstructive pulmonary disease, selective β_1-adrenergic receptor antagonists should be used. Depression of the myocardium leading to cardiac failure has also been described. Abrupt discontinuation of propranolol therapy has been associated with severe exacerbation of angina and the occurrence of myocardial infarction. Skin rashes and conjunctival xerosis are rarely associated with β-adrenergic receptor antagonists. β-Adrenergic receptor antagonists may mask clinical signs of developing or continuing hyperthyroidism, as well as block premonitory signs and symptoms of acute hypoglycemia.

Other adverse events include CNS depression, fatigue, nightmares, insomnia, vivid dreaming, and hallucinations. β-Adrenergic receptor antagonists do not adversely affect cognitive function and may enhance performance in measures of memory, learning, and psychomotor speed. The incidence of CNS side effects is lowest with the most hydrophilic agents (e.g., atenolol) and highest with pindolol and propranolol, while metoprolol is intermediate. Other side effects attributable to β-adrenergic receptor antagonists are sexual dysfunction and gastrointestinal disturbances; hypersensitivity reactions are rare. Incident depressive symptoms are hypothesized to be due to the inhibitory actions of β-adrenergic receptor antagonists on β-adrenergic and serotonin receptors in the CNS. Individuals with sinus bradycardia, congestive heart failure, greater than first-degree atrioventricular block, severe chronic obstructive pulmonary disease, and bronchial asthma should not be prescribed, β-adrenergic receptor antagonists.

Drug Interactions

Coadministered drugs that induce hepatic oxidative enzymes may increase the clearance of propranolol and metoprolol. Individuals coadministered metoprolol with a cytochrome P450 2D6-inhibiting SSRIs, or those who are slow metabolizers of debrisoquine hydroxylase may exhibit elevated plasma concentrations of metoprolol.

Coadministered agents that alter hepatic blood flow are also predicted to increase plasma concentrations of orally administered metoprolol and propranolol. Agents that inhibit hepatic enzymes and/or reduce hepatic blood flow will exert an opposite effect.

β-Adrenergic receptor antagonists may alter the pharmacokinetics of other agents by reducing hepatic blood flow and inhibiting hepatic oxidative metabolism. Propranolol may increase the blood concentration of chlorpromazine by inhibiting its metabolism. Propranolol may also inhibit the metabolism of coumarin anticoagulants.

Pharmacodynamic interactions with β-adrenergic receptor antagonists include cardiac depression effects and potentiation of postural hypotension with coadministered antiarrhythmics and vasodilators.

Laboratory Interference

Pindolol may elevate liver transaminases, alkaline phosphatase, lactate dehydrogenase, and serum uric acid; pindolol rarely reduces bilirubin. Cases of thrombocytopenia and leucopenia have been described with β-adrenergic receptor antagonists.

DOSAGE AND ADMINISTRATION

For disorders in which a central effect is desired (e.g., violence), a more lipophilic agent (e.g., propranolol) should be considered. For peripheral side effects associated with psychotropic medication use, less lipophilic agents should be considered.

Propranolol is usually initiated orally at 10 mg t.i.d. or 20 mg b.i.d. increasing by approximately 20 to 30 mg per day until the desired effect is achieved. The patient's pulse and blood pressure readings should be taken regularly, and the agent should be withheld if the pulse is less than 50 beats per minute, the systolic blood pressure is less than 90 mm Hg, or the patient manifests CNS symptomatology. Propranolol should be tapered by approximately 40 to 60 mg daily until 60 mg per day is reached, after which time the drug should be tapered 10 to 20 mg per day every 3 to 4 days. β-Adrenergic receptor antagonists should not be discontinued abruptly as severe exacerbation of angina and the occurrence of myocardial infarction have been reported in patients with pre-existing cardiovascular disease following abrupt discontinuation of propranolol therapy.

Pindolol is initiated at 2.5 mg t.i.d., and increases are tolerated up to a maximum of 60 mg per day. Nadolol is initiated at 40 mg per day, increasing to 80 mg per day as tolerated. Atenolol is initiated at 50 mg per day and can be increased to 100 mg per day. Rarely an atenolol dose of up to 200 mg per day is used. The effective dosing range of metoprolol is 100 to 450 mg. The initiation dose is 100 mg; dosing beyond 450 mg has not been studied.

FURTHER READINGS

Amos T, Stein DJ, Ipser JC. Pharmacological interventions for preventing posttraumatic stress disorder (PTSD). *Cochrane Database Syst Rev*. 2014;7:CD006239.

Antonelli-Incalzi R, Pedone C. Respiratory effects of beta-adrenergic receptor blockers. *Curr Med Chem*. 2007;14(10):1121–1128.

*Arana GW, Santos AB. B-Adrenergic receptor antagonists. In: Kaplan HI, Sadock BJ, eds. *Comprehensive Textbook of Psychiatry.* 6th ed. Williams & Wilkins; 1995.

*Baker JG. The selectivity of beta-adrenoceptor antagonists at the human beta1, beta2 and beta3 adrenoceptors. *Br J Pharmacol*. 2005;144(3):317–322.

Ballesteros J, Callado LF. Effectiveness of pindolol plus serotonin uptake inhibitors in depression: A meta-analysis of early and late outcomes from randomised controlled trials. *J Affect Disord*. 2004;79(1–3):137–147.

Bell C, Wilson S, Rich A, Bailey J, Nutt D. Effects on sleep architecture of pindolol, paroxetine and their combination in healthy volunteers. *Psychopharmacology (Berl)*. 2003;166(2):102–110.

Boere E, Birkenhager TK, Groenland TH, van den Broek WW. Beta-blocking agents during electroconvulsive therapy: a review. *Br J Anaesth*. 2014;113(1):43–51.

*Compendium of Pharmaceuticals and Specialties. Ottawa: Canadian Pharmacist Association; 2007.

de Quervain DJ, Aerni A, Roozendaal B. Preventive effect of beta-adrenoceptor blockade on glucocorticoid-induced memory retrieval deficits. *Am J Psychiatry*. 2007;164(6):967–969.

Hurlemann R, Walter H, Rehme AK, et al. Human amygdala reactivity is diminished by the beta-noradrenergic antagonist propranolol. *Psychol Med*. 2010;40(11):1839–1848.

*Kaplan HI, Sadock BJ, Grebb JA. B-Adrenergic receptor antagonists. In: Kaplan HI, Sadock BJ, Grebb JA, eds. *Synopsis of Psychiatry.* 7th ed. Williams & Wilkins; 1994:891.

Kindt M, Soeter M, Sevenster D. Disrupting reconsolidation of fear memory in humans by a noradrenergic beta-blocker. *J Vis Exp*. 2014;(94):doi: 10.3791/52151.

Martiny K, Lunde M, Bech P, Plenge P. A short-term double-blind randomized controlled pilot trial with active or placebo pindolol in patients treated with venlafaxine for major depression. *Nord J Psychiatry*. 2012;66(3):147–154.

Peskind ER, Tsuang DW, Bonner LT, et al. Propranolol for disruptive behaviors in nursing home residents with probable or possible Alzheimer disease: A placebo-controlled study. *Alzheimer Dis Assoc Disord*. 2005;19(1):23–28.

Rasmussen DD, Beckwith LE, Kincaid CL, Froehlich JC. Combining the alpha1-adrenergic receptor antagonist, prazosin, with the beta-adrenergic receptor antagonist, propranolol, reduces alcohol drinking more effectively than either drug alone. *Alcohol Clin Exp Res*. 2014;38(6):1532–1539.

Tsai SY, Lee HC, Chen CC. Hyperinsulinaemia associated with beta-adrenoceptor antagonist in medicated bipolar patients during manic episode. *Prog Neuropsychopharmacol Biol Psychiatry*. 2007;31(5):1038–1043.

33.9 GABA Enhancers

33.9a Barbiturates and Similarly Acting Substances

Steven L. Dubovsky, M.D., and Dori Marshall, M.D.

Stories of substances that depress consciousness date back to antiquity. The first commercially available sedative-hypnotics were the bromides, which were introduced in 1853. These substances had unpredictable effects and were frequently taken in overdose. In 1864, Adolf Johann Friedrich von Baeyer (who eventually won the Nobel Prize for his contributions to the development of organic chemistry) synthesized barbituric acid. The name barbituric acid comes from a combination of uric acid, which Bayer had been studying, and the name Barbara. The name Barbara was apparently chosen because it was the name of a close friend of von Baeyer or because he celebrated the new compound by visiting a tavern on St. Barbara's Day. Barbituric acid had no sedative or hypnotic properties, but diethylbarbituric acid or barbital was introduced in 1903 by Fisher and Von Mering as the first sedative barbiturate. Barbital was marketed as Veronal, after Mt. Verona, where Fisher had been traveling. Many other barbituric acid derivatives (barbiturates) were subsequently developed, beginning with phenobarbital in 1912, followed by a number of related compounds such as chloral hydrate and paraldehyde.

Barbiturates were the first effective antiepileptic medications and the first injectable anesthetics. Barbiturates were widely used as sedative-hypnotic agents in the first half of the 20th century. However, by the 1950s and 1960s, these medications had become a common means of suicide, as was publicized in the deaths of celebrities such as Marilyn Monroe and Judy Garland. After additional problems such as high addiction potential, significant risk of respiratory depression, and a narrow therapeutic index became apparent, allegedly "safer" alternatives such as methyprylon, glutethimide, ethchlorvynol, and methaqualone were introduced. These drugs turned out to have the same problems, however, and the entire group of agents was eventually supplanted by the benzodiazepines, the first one of which (chlordiazepoxide) was synthesized in 1955 and marketed in 1960. Glutethimide, methaqualone, and other derivatives were withdrawn from the market because of significant toxicity and risks of misuse. Barbiturates are still used clinically, most frequently in anesthesia, including for ECT, for detoxification from alcohol or other sedative-hypnotics, as anticonvulsants and for promoting recovery from a cerebral hemorrhage.

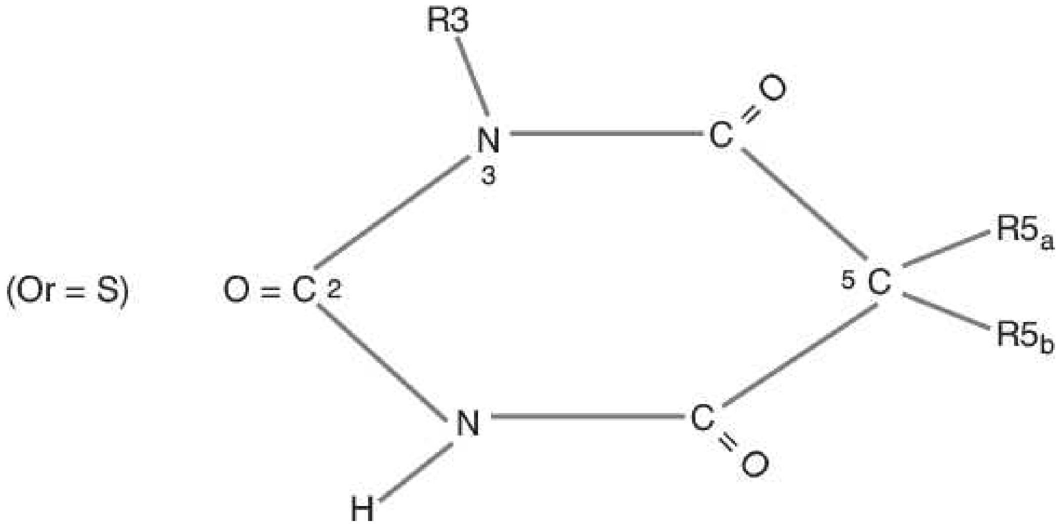

FIGURE 33.9a–1. Basic barbiturate molecular structure.

CHEMISTRY

Modern barbiturates (Fig. 33.9a–1) are derivatives of barbituric acid (2,4,6-trioxo-hexahydropyrimidine). Barbituric acid itself has no CNS depressant properties, but the presence of an alkyl or aryl group at position 5 confers sedative-hypnotic and anticonvulsant activity and various side chains alter lipophilicity, potency, and elimination. For example, a sulfur group at position 2 (Fig. 33.9a–1) produces thiobarbiturates, which have different lipid solubility and, therefore, different duration of action, latency of action, metabolism, and potency compared with other barbiturates.

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Absorption. The barbiturates are administered orally when used as sedative-hypnotics and chronic anticonvulsants and parenterally when used for anesthesia and for the treatment of status epilepticus. The intramuscular route is sometimes used for the diagnosis and early treatment of CNS depressant withdrawal in agitated patients. Oral doses of barbiturates are rapidly and completely absorbed from the small intestine. The onset of action after oral administration ranges from 10 to 60 minutes, depending on the preparation and stomach contents. Barbiturates with greater lipid solubility and protein binding and a high pKa, have a more rapid onset and a shorter duration of action. When administered intravenously (IV), the onset of action is immediate.

Distribution. Barbiturates are widely distributed and readily cross the blood–brain barrier. Barbiturates vary in their lipid solubility and their binding to plasma proteins. More lipophilic compounds such as thiopental and methohexital are 80% protein-bound, whereas less lipophilic compounds (such as phenobarbital) are only 20% to 40% protein bound. More lipid-soluble barbiturates are more rapidly distributed to the brain, resulting in a rapid onset of action. Rapid efflux from the brain along a concentration gradient then leads to a short duration of action. Some weak acids, such as warfarin or aspirin, can displace barbiturates from plasma proteins, augmenting their CNS effects.

Metabolism and Elimination. Most barbiturates are metabolized almost completely in the liver by cytochrome P450 microsomal enzymes. However, a significant amount of phenobarbital is excreted unchanged by the kidney. Urinary excretion of phenobarbital, which has a low pKa, can be increased five- to tenfold by alkalinization of the urine. This approach will not work for short-acting barbiturates because they have higher pKas, more protein binding, and more metabolism to inactive products in the liver. The biotransformation that is largely responsible for the termination of barbiturate activity is the oxidation of radicals at C5. N-glycosylation is also an important metabolic pathway that, in some instances, yields active metabolites. Metabolism and elimination are more rapid in young people than in the older adults, infants, and pregnant women. Metabolism is slowed, increasing the elimination half-life, in chronic liver disease. Elimination half-lives for some barbiturates are summarized in Table 33.9a–1.

Many barbiturates induce their own metabolism as well as the metabolism of other drugs. One consequence of this phenomenon is decreased elimination half-life and declining serum levels, and effect with repeated administration. Phenobarbital induces CYP2- and 3A isoenzymes, resulting in substantially increased metabolism and decreased effect of beta-blockers, corticosteroids, doxycycline, estrogen, phenothiazines, tricyclic and a number of newer antidepressants (e.g., fluoxetine, paroxetine, bupropion), quinidine, and other anticonvulsants.

Table 33.9a–1.
Dose Range, Elimination Half-Life, and Onset and Duration of Action of Some Barbiturates and Related Compounds

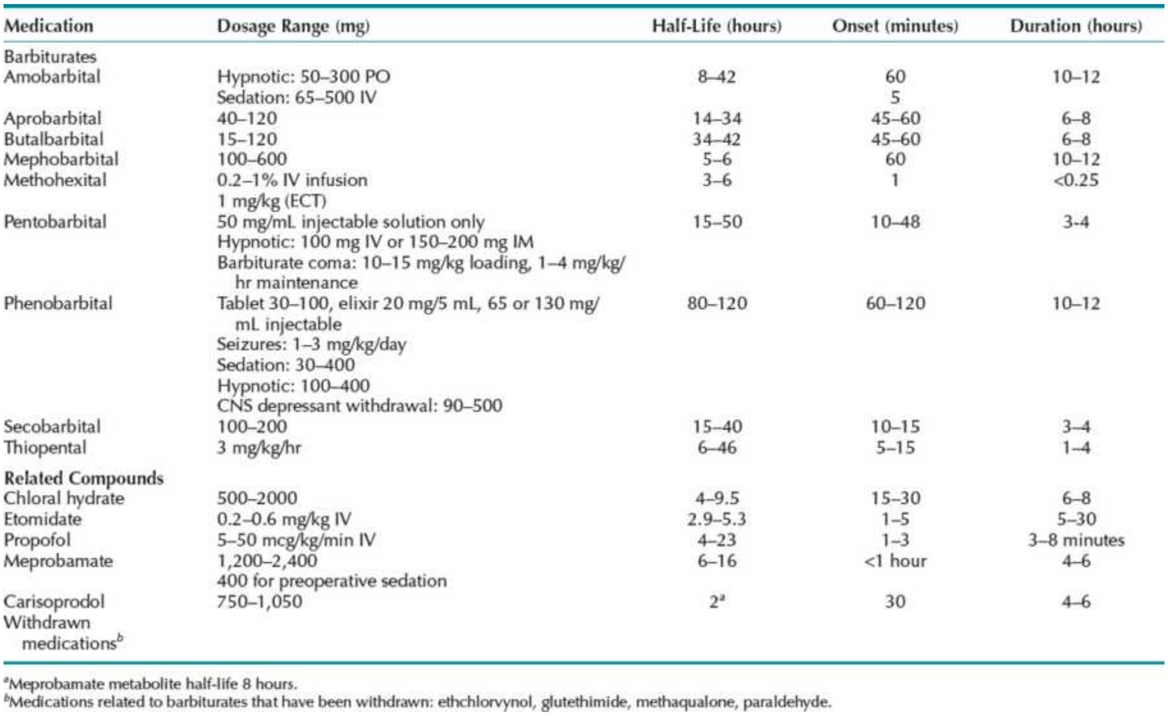

Medication	Dosage Range (mg)	Half-Life (hours)	Onset (minutes)	Duration (hours)
Barbiturates				
Amobarbital	Hypnotic: 50–300 PO Sedation: 65–500 IV	8–42	60 5	10–12
Aprobarbital	40–120	14–34	45–60	6–8
Butalbarbital	15–120	34–42	45–60	6–8
Mephobarbital	100–600	5–6	60	10–12
Methohexital	0.2–1% IV infusion 1 mg/kg (ECT)	3–6	1	<0.25
Pentobarbital	50 mg/mL injectable solution only Hypnotic: 100 mg IV or 150–200 mg IM Barbiturate coma: 10–15 mg/kg loading, 1–4 mg/kg/hr maintenance	15–50	10–48	3-4
Phenobarbital	Tablet 30–100, elixir 20 mg/5 mL, 65 or 130 mg/mL injectable Seizures: 1–3 mg/kg/day Sedation: 30–400 Hypnotic: 100–400 CNS depressant withdrawal: 90–500	80–120	60–120	10–12
Secobarbital	100–200	15–40	10–15	3–4
Thiopental	3 mg/kg/hr	6–46	5–15	1–4
Related Compounds				
Chloral hydrate	500–2000	4–9.5	15–30	6–8
Etomidate	0.2–0.6 mg/kg IV	2.9–5.3	1–5	5–30
Propofol	5–50 mcg/kg/min IV	4–23	1–3	3–8 minutes
Meprobamate	1,200–2,400 400 for preoperative sedation	6–16	<1 hour	4–6
Carisoprodol	750–1,050	2[a]	30	4–6
Withdrawn medications[b]				

[a] Meprobamate metabolite half-life 8 hours.

[b] Medications related to barbiturates that have been withdrawn: ethchlorvynol, glutethimide, methaqualone, paraldehyde.

Pharmacodynamics

Mechanism of Action. Barbiturate-binding sites are allosterically linked to other binding sites that comprise the GABAA–benzodiazepine complex on the anionic chloride ion channel on postsynaptic neuronal membranes. Whereas benzodiazepines alter GABA receptor affinity to increase the frequency of chloride ion channel openings, barbiturates increase the duration of openings of the chloride ion channel.

Most barbiturate sedative-hypnotics bind to GABAA receptors containing the alpha1 subunit, but etomidate produces sedation at the beta2 subunit and anesthesia at the beta3 subunit. In addition to binding to P- and Q-type calcium channels to inhibit the influx of depolarizing positive charges, barbiturates, along with etomidate, and propofol, inhibit NMDA and amino-3-hydroxy-5-methyl-4-isoxazole-propionic acid (AMPA)/kainate receptors as well as reducing glutamate release and increasing presynaptic glutamate reuptake. Barbiturates and related drugs inhibit pentameric ligand-gated excitatory cation channels such as those linked to nicotinic cholinergic receptors by acting at binding sites in the closed state of the cation channel, stabilizing the closed state and reducing neuronal excitability. At higher concentrations used for anesthesia, barbiturates appear to reduce the repetitive firing of neurons by inhibiting voltage-dependent sodium and possibly potassium channels. Barbiturates and related drugs may also affect the binding of other drugs to their binding sites. For example, propofol potentiates the binding of pentobarbital to its receptor on the GABAA–benzodiazepine complex.

Blood Concentrations. Although barbiturates have greater acute effects at higher plasma concentrations, tolerance is common; therefore, the same plasma level may produce drastically different actions at different times in the same individual and from one individual to another. Tolerance to barbiturates, which may not have a linear relationship to dose, is associated with downregulation and desensitization of receptors, structural changes in receptors, and uncoupling of the drug from the receptor, in addition to declining serum levels due to autoinduction of metabolism. Tolerance to anesthetics such as thiopental may occur within minutes. The Mellanby effect refers to acute tolerance that results in more clinical effects when serum levels are increasing than when they are decreasing. Cross-tolerance of barbiturates and all other drugs acting on the GABAA receptor is common. One mechanism of cross-tolerance is the reduced affinity of barbiturate-binding sites induced by tolerance to benzodiazepines.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Because barbiturates were introduced as hypnotics before the advent of validated instruments of assessment of insomnia and treatment response, there are few reliable controlled studies of this application. Although extensive clinical experience indicates that barbiturates are effective as sedative-hypnotics, the rapid development of tolerance, a frequent rebound of insomnia and REM sleep, multiple drug interactions, and a narrow therapeutic index make these drugs poor initial choices for insomnia. Barbiturates and related compounds are only approved for short-term treatment of insomnia and are at best second- or third-line treatment for severe acute insomnia after benzodiazepines, nonbenzodiazepine GABAA agonists, and sedating antidepressants or antipsychotic drugs.

Although the sedative and antiarousal actions of barbiturates may secondarily moderate anxiety, there is little evidence from any controlled studies of reliable benefit for any specific anxiety disorder. Benzodiazepines and antidepressants, especially SSRIs, are generally recommended as first-line agents in the treatment of anxiety disorders.

There is variable evidence supporting the use of barbiturates in the treatment of a number of neurologic conditions. Phenobarbital is considered a safe, economical, and effective treatment for generalized tonic–clonic seizures and simple and complex partial seizures. Because barbiturates reduce intracranial pressure (ICP), they are sometimes used as adjunctive acute treatments following head injury. Use of barbiturate infusion, usually pentobarbital, is sometimes used—along with various neurosurgical approaches—to manage increased ICP in pediatric severe traumatic brain injury (TBI) after first-line therapies have failed. However, a systematic review found no evidence that the reduction of ICP by barbiturates is associated with a reduction in morbidity or mortality. In addition, hypotension caused by the barbiturate can reduce cerebral perfusion pressure. Barbiturate-containing analgesic products are often prescribed for headaches (including vascular, tension, and mixed migraines) as well as premenstrual and postpartum pain. A Canadian review of the literature found that there was no evidence of a clinically important enhancement of analgesia by adding a barbiturate to an established analgesic, and the risks of dependence, misuse and withdrawal outweigh any benefit from sedation caused by the barbiturate. In addition, the use of barbiturates 5 days or more per month is associated with rebound headaches and the progression of headache disorders.

Barbiturates are still routinely used to induce anesthesia. Methohexital is the most frequently utilized anesthetic agent for ECT. Propofol, etomidate, and thiopental have also been found to be effective for this indication. Propofol has increasingly been used for outpatient anesthesia (e.g., for colonoscopy), but this application requires careful monitoring and benzodiazepines such as midazolam may be safer. Fentanyl appears to be an equivalent anesthetic agent to propofol, but it is associated with better postoperative pain control.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Barbiturates depress the activity of all excitable tissues. The CNS is more sensitive to these effects than are peripheral tissues. Barbiturates produce varying degrees of CNS depression, from mild sedation to general anesthesia. Although all barbiturates have anticonvulsant properties at anesthetic doses, only phenobarbital and mephobarbital have anticonvulsant properties at subhypnotic doses. Barbiturates depress the sensory cortex, decrease motor activity, alter cerebellar function, and produce drowsiness, sedation, and sleep. At subanesthetic doses, barbiturates may increase the reaction to painful stimuli by suppressing inhibitory pain centers.

Barbiturates decrease sleep latency and the number of awakenings, and they increase total sleep time, but they also suppress REM sleep as well as stages III and IV sleep. As a result, any increase in sleep is not necessarily restorative. Tolerance develops to the hypnotic effect, often within a few days. Within 2 weeks of continued use, the effects of barbiturates on prolonging sleep are virtually eliminated, and doses must be increased to obtain the same effect. Abrupt discontinuation of barbiturates results in rebound insomnia and REM rebound with nightmares.

In the peripheral nervous system, barbiturates selectively inhibit transmission in autonomic ganglia and reduce nicotinic excitation. This results in a reduction of blood pressure, at times with clinically significant hypotension.

Respiratory drive and sleep–wake rhythms are depressed by barbiturates. At therapeutic barbiturate doses, the neurogenic drive for wakefulness is reduced to the level that promotes normal sleep. However, at three times the hypnotic dose, the wakefulness drive is virtually eliminated. These doses also appear to reduce the hypoxic drive for breathing, exacerbating sleep apnea. Protective reflexes are not depressed until sedation produces severe respiratory depression, causing one of the most serious complications of barbiturate anesthesia, namely laryngospasm. These factors may have contributed to the death of the entertainer Michael Jackson after he was administered escalating doses of propofol in an attempt to suppress tolerance, rebound insomnia, and withdrawal.

Barbiturate effects on the cardiovascular system are generally mild. At sedative-hypnotic doses, barbiturates produce a slight decrease in blood pressure and heart rate but usually no more than that seen during normal sleep. This reduction in blood pressure and heart rate can be exaggerated in patients with congestive heart failure or hypovolemic shock. IV barbiturate administration may cause ventricular arrhythmias.

The primary effect of barbiturates on the liver is the induction of microsomal enzymes. The induction of cytochrome P450 isoenzymes increases the metabolism of a number of medications and other substances, including many anticonvulsants, antidepressants, antipsychotic drugs, and steroid hormones. In addition, the autoinduction of P450 enzymes accelerates the metabolism of barbiturates themselves.

THERAPEUTIC INDICATIONS

Many barbiturates, including amobarbital, butabarbital, aprobarbital, pentobarbital, and secobarbital, are indicated for oral use as sedatives or hypnotics (Table 33.9a–1). However, their use in insomnia should be limited to 2 weeks or less because they lose much of their effectiveness after this period of time, followed by frequent tolerance, rebound and misuse.

Seizure Disorders

Barbiturates are standard anticonvulsants used for prophylaxis of seizures and treatment of status epilepticus. These medications inhibit epileptic foci and reduce the spread of seizure activity. Phenobarbital, the most commonly used barbiturate, is often administered along with other anticonvulsant classes. Phenobarbital is effective for simple and complex partial seizures, as well as generalized tonic-clonic seizures. For status epilepticus, intravenous benzodiazepines (e.g., diazepam) and phenytoin are first-line treatments. However, if seizure activity continues, IV phenobarbital (20 mg/kg) at a rate no faster than 100 mg/min may be administered. A systematic review and meta-analysis concluded that propofol is superior to barbiturates in controlling refractory status epilepticus and shortening the average duration of tracheal intubation. Because it raises the seizure threshold, ECT has been used for refractory status epilepticus.

Anesthesia

The first barbiturate to be used intravenously was a combination of barbital and diallylbarbituric acid known as Somnifen, which was synthesized in 1920 for obstetrical use and introduced as an agent for surgical anesthesia in 1924. Some surgeons administered the barbiturate intravenously as preanesthesia prior to inhalational anesthesia. This and similar agents were limited by slow onset of action and long recovery times. The first short-acting barbiturate was hexobarbital (Evipan), which was introduced as an induction agent in 1932. Amobarbital (sodium amytal) was introduced into American anesthesia practice in the late 1920s. Thiopental came into widespread use in the 1930s, establishing barbiturates as popular anesthetic agents. The addition of a methyl group to the barbituric acid substrate led to the development of methohexital, which proved to be shorter acting than other barbiturates and therefore preferable for briefer procedures, including dental procedures and ECT.

In general, the brief duration of ECT requires shorter-acting agents. Methohexital is still the first-line agent for anesthesia during ECT. The rapid onset and offset of action of etomidate make it an appropriate choice. Thiopental is longer acting, and its anticonvulsant properties can interfere with the induction of therapeutic seizures. Propofol may also shorten seizure duration. The proconvulsant action of ketamine could make it useful when seizures are difficult to induce. Multiple reports of an acute antidepressant effect of ketamine led to the expectation that it might augment ECT, but an additive effect of ketamine and ECT has not been confirmed empirically, and side effects such as dissociation, psychosis and possibly excitotoxicity with ongoing use are precautions in its use in psychiatry. A randomized comparison of etomidate, propofol, and thiopental found all to be relatively free of important cardiovascular effects in otherwise healthy patients undergoing ECT. Etomidate raises the seizure threshold less than propofol, and its use is associated with the need for a lower stimulus dose to produce adequate seizures.

Cerebral Lateralization

The intracarotid amobarbital procedure (IAP or Wada Test) is used to determine the lateralization of memory and language in order to predict the risks of amnesia and aphasia following epilepsy surgery. First, amobarbital (100 to 200 mg) is administered directly into the internal carotid artery contralateral to the epileptic focus, and the next day it is administered ipsilateral to the focus. Temporarily anesthetizing each hemisphere separately interrupts all functions localized to that hemisphere, such as naming, repetition, following commands and remembering. By comparing the impact of injections on the two sides, it is possible to predict whether surgery on that side is likely to interrupt the functions tested. Methohexital is not as reliable as amobarbital for this purpose.

Narcoanalysis

Since the 1930s, interviews with amobarbital have been performed to help to distinguish between psychogenic and neurologic symptoms when symptoms disappear during the amobarbital session and to elicit a history from a mute patient. The amobarbital interview has been used as a diagnostic tool in cases of mutism, stupor, unexplained confusion, and suspected malingering and conversion disorder. The procedure involves administering IV amobarbital slowly in doses up to a maximum of 500 mg. Initial responses include dizziness, blurred vision, and fatigue, followed by drowsiness and mild euphoria. It is during this phase that the interview may be most fruitful. If the dose is increased further, the patient is likely to fall asleep. Controlled studies of the amobarbital interview are difficult to conduct, but clinical experience suggests that it may reduce inhibitions, help an uncommunicative patient has about communicating with the clinician, and clarify confusing psychogenic symptoms. The amobarbital interview should only be performed if personnel and equipment for resuscitation are present and increased intracranial pressure is not suspected.

Diagnosing Withdrawal From Sedative-Hypnotics

Because CNS depressants are cross-reactive through actions at different sites on the GABAA receptor complex, the barbiturate tolerance test can be used to determine the degree to which tolerance exists to barbiturates, related drugs, benzodiazepines and alcohol, alone or in combination. In the presence of signs and symptoms of CNS depressant withdrawal, a test dose is administered of 200 mg of pentobarbital or 30 to 100 mg of phenobarbital. If the patient is grossly intoxicated an hour later, tolerance is not significant, and a withdrawal syndrome is unlikely. For less marked intoxication after the test dose (e.g., dysarthria and ataxia without evidence of obtundation), further doses are administered every 1 to 2 hours for pentobarbital or every 1 to 4 hours for phenobarbital until signs and symptoms of withdrawal are completely suppressed, distinct intoxication emerges, or until a maximum dose (e.g., 480 mg of phenobarbital or 500 mg of pentobarbital) is reached. The barbiturate is then administered in a divided dose, and the dose is decreased by about 10 percent per day as determined each day by the emergence of signs of toxicity (e.g., gross sedation, nystagmus) or withdrawal (e.g., hypertension, hyperreflexia) charted before and after each dose of barbiturate. Such determinations can be difficult because features such as confusion, dysarthria and hypotension can occur with intoxication or withdrawal, although nystagmus is usually specific to intoxication. Prospective charting of specific features facilitates making the distinction. Because it is available in an injectable form and it has a longer duration of action, phenobarbital is preferred as the withdrawal agent (with 30 mg of phenobarbital being roughly equivalent to 100 mg of pentobarbital), whereas either drug can be used to diagnose tolerance.

Delirium Tremens

Oral or IV phenobarbital, sometimes in doses as high as 1,500 to 2,000 mg on the first day, has been used as alternatives to benzodiazepines in the treatment of alcohol withdrawal delirium. Propofol in doses up to 4 mg/kg/hr has been used for severe, intractable withdrawal delirium, especially if the patient is intubated. In such severe cases, ECT might also be considered since it is an effective treatment for delirium of any cause.

Neuroprotection

Because they reduce neuronal electrical activity and energy demands and antagonize NMDA and AMPA receptors, barbiturates alone or with hypothermia have been used for years to protect the brain during surgery and sometimes following severe CNS injury, especially intraoperative stroke. If damage is not severe, brain oxygen levels are increased by pentobarbital. However, any protective action is short-term because these medications also have systemic immunosuppressive properties related to inhibition of T-cell calcineurin/ calmodulin function, and their use for this purpose has been supplanted by newer anesthetic agents such as isoflurane. Barbiturates have been used in some settings to reduce intracranial pressure following TBI, but these medications can also cause hemodynamic instability that reduces cerebral perfusion. Despite some positive in vitro data, no compelling evidence has emerged of an actual clinical benefit of barbiturates following intraoperative stroke. Data on improved outcomes after barbiturate treatment following TBI have not been robust. Animal studies suggest that propofol and related medications may reduce beta amyloid-induced mitochondrial dysfunction and improve cognition. However, demented patients may be more susceptible to adverse cognitive effects, although they appear not to have increased vulnerability to respiratory depression after anesthesia. By combining nucleic acid bases, heterocyclic amines, and barbiturate derivatives, a group of nucleobase-based barbiturates have been developed that appear to have antioxidant properties that protect against DNA damage in laboratory studies. Magnesium has also been shown to have some promise in providing neuroprotection after surgical procedures associated with ischemia and also among patients treated within 12 hours of acute ischemic stroke, but there is conflicting evidence of clinical benefit of this effect.

PRECAUTIONS AND ADVERSE REACTIONS

Barbiturate overdoses can be dangerous: 1 g (10 pills) causes serious toxicity, and 2 g or more may be fatal. Acute overdose is manifested by dysarthria, ataxia, incoordination, CNS depression, respiratory depression, respiratory acidosis, tachycardia, hypotension, hypothermia, and oliguria. Overdoses of specific drugs are associated with additional features. For example, chloral hydrate and ethchlorvynol produce unique odors; meprobamate and chloral hydrate cause cardiac arrhythmias; methaqualone, propofol, and etomidate induce myoclonus; chloral hydrate and methaqualone can cause GI bleeding, and glutethimide produces anticholinergic signs. Intravenous overdoses can produce additional toxicity associated with the material in which the drug is diluted. Propylene glycol in intravenous formulations of etomidate (and lorazepam) can cause hypotension, hyperosmolar states, and metabolic acidosis.

Because death from barbiturate overdose is caused by cardiorespiratory collapse, treatment involves supporting the airway, oxygenation, and cardiovascular functions. Volume expansion is used to treat severe hypotension, with vasopressors only added when this approach does not work or causes volume overload. Beta-adrenergic antagonists may be helpful for tachyarrhythmias. Activated charcoal with a protected airway can be helpful, although controlled research has not confirmed the benefit of this practice. The benzodiazepine receptor antagonist flumazenil, which can reverse benzodiazepine overdoses, is not useful for overdoses of barbiturates because they act on a different receptor on the GABA receptor chloride ion channel complex. Severe barbiturate overdose can be successfully treated with hemodialysis, which is used earlier in the course of chloral hydrate overdoses with lifethreatening arrhythmias and prolonged intubation after ingestion of large quantities of phenobarbital.

The older adutls, children and those with brain damage may develop paradoxical excitation in response to barbiturates and other CNS depressants. Cases have been reported in which phenobarbital caused or exacerbated hyperactivity when used chronically in younger children, and the hyperactivity did not remit when the medication was withdrawn.

As noted earlier, barbiturates are often prescribed in combination with analgesics (Table 33.9a–2). However, they may exacerbate pain in some patients, and there is no evidence that they have analgesic properties. On the other hand, headache preparations containing butalbital are clearly implicated in medication overuse headaches. In this syndrome, which usually involves migraine headaches, continued use of the medication results in rebound headaches, which are then treated with more self-administered medication. Discontinuing the medication can be a challenge since rebound headaches may persist for some time after detoxification, and relapse of medication use is common.

Table 33.9a–2.
Barbiturate and Related Substance Containing Analgesics

Product Brand Name	Barbiturate Component	Other Constituents
Fiorinal with codeine	Butalbital 50 mg	Aspirin 325 mg, caffeine 40 mg, codeine 30 mg
Fioricet with codeine	Butalbital 50 mg	Acetaminophen 325 mg, caffeine 40 mg, codeine 30 mg
Esgic	Butalbital 50 mg	Acetaminophen 325 mg, caffeine 40 mg
Donnatal	Phenobarbital 16.2 mg	Atropine 0.02 mg, hyoscyamine 1 mg, scopolamine 6.5 mcg
Soma with codeine	Carisoprodol 200 mg	Aspirin 325 mg, codeine 16 mg

Hypersensitivity to barbiturates is not uncommon. Allergic reactions are most likely to occur in patients with asthma, urticaria, and other autoimmune disorders and range from local swelling to exfoliative dermatitis. Bullous lesions have occasionally been caused by phenobarbital, more frequently but not exclusively in comatose patients.

Barbiturates are contraindicated in patients with porphyria variegata or acute intermittent porphyria because they enhance porphyrin synthesis and aggravate the illness.

Barbiturate Dependence

Barbiturates rapidly induce tolerance and dependence. Dependence is most common with shorter-acting agents such as pentobarbital and secobarbital. Pre-existing substance use increases the risk of barbiturate dependence. The risk of dependence on barbiturates increases with higher doses taken for longer periods of time (e.g., 400 mg per day of secobarbital equivalent taken for 90 days or more). Patients who have become dependent on barbiturates take on average 1,500 mg per day.

As is true of benzodiazepines and other CNS depressants, rapid discontinuation of barbiturates induces withdrawal syndromes that include anxiety, tremor, myoclonus, unstable blood pressure, dizziness, nausea, vomiting, agitation, and insomnia. Seizures, delirium, coma, and death may occur in more severe cases. Signs and symptoms of barbiturate withdrawal begin sooner and abate more rapidly after discontinuation of shorter-acting agents. As noted earlier, withdrawal from barbiturates with or without other CNS depressants that act on the GABA receptor complex can be diagnosed with the barbiturate tolerance test and treated with a gradual taper of phenobarbital.

Teratogenicity

Barbiturates inhibit progesterone synthesis by inhibiting cholesterol transport, which can impair the maintenance of pregnancy. In addition, barbiturates readily cross the placental barrier, and there is some evidence that maternal consumption of barbiturates is associated with fetal abnormalities, especially cardiac anomalies. Newborns of mothers who use barbiturates in the third trimester may develop neonatal abstinence syndromes. Barbiturates are excreted in breast milk and may produce sedation and withdrawal syndromes in nursing infants. Concerns have been expressed about potential adverse effects of barbiturate and related anesthetics during pregnancy, but exposure is brief, and any minimal risk is balanced by the adverse impact on the pregnancy of the condition for which anesthesia is administered, including depression if anesthesia for ECT is planned. Otherwise, ongoing use of barbiturates should be avoided during pregnancy unless a compelling indication such as intractable epilepsy is present.

Psychomotor Impairment

As is true of benzodiazepines, barbiturates can impair driving in addition to causing sedation and an intoxicated feeling. The degree of impairment is similar to that of alcohol.

DRUG INTERACTIONS

Complications from the concomitant use of barbiturates and other medications most often result from additive CNS depressant effects or from induction by the barbiturate of hepatic microsomal enzymes. Additive

CNS depression occurs with alcohol, antihistamines, benzodiazepines, and other sedative-hypnotics. Metabolism of vitamins D and K are accelerated by barbiturates, which may reduce dietary calcium absorption. Induction of hepatic metabolism may decrease the effectiveness of oral anticoagulants such as warfarin, corticosteroids, and estrogen. The latter effect may result in the failure of oral contraceptives. Phenobarbital interferes with oral absorption of griseofulvin, thereby decreasing blood levels and possibly reducing therapeutic response. Phenobarbital induces metabolizing enzymes for doxycycline, phenytoin, TCAs, and some antipsychotic drugs. Valproate may decrease barbiturate metabolism, thus reducing the dose of barbiturate needed, while monoamine oxidase inhibitors may prolong the effects of barbiturates by inhibiting their metabolism. In view of the frequent induction of CYP450 enzymes by barbiturates, it is important to look for an altered efficacy of any P450 substrate to which barbiturates are added.

LABORATORY INTERFERENCES

Routine toxicology screening does not identify barbiturates. Barbiturates have been reported to produce a false-positive phentolamine test result. Acceleration of dexamethasone metabolism by phenobarbital results in a false-positive dexamethasone suppression test. Induction of metabolism by barbiturates of metyrapone, a cortisol synthesis inhibitor that normally increases the production of corticotropin-releasing hormone and corticotropin (ACTH) and is used to assess Cushing disease and other forms of hypercortisolism, can lead to false-negative tests due to reduced metyrapone levels and inadequate suppression of cortisol production.

DOSAGE AND ADMINISTRATION

Dosages and routes of administration are given in Table 33.9a–1. Table 33.9a–2 lists doses of some preparations that contain barbiturates in combination with other medications.

NONBARBITURATE, NONBENZODIAZEPINE SEDATIVE-HYPNOTIC AGENTS

Chloral Hydrate

Chloral hydrate was introduced in 1832. It was widely used in the late 19th century as a sedative and hypnotic. Chloral hydrate was also used to treat many psychiatric illnesses, such as anxiety, psychosis, and mania until more effective treatments were introduced. As reports of overdose deaths accumulated, by 1910, it was considered to be more dangerous than heroin or opium. The routine use of chloral hydrate as a sedative-hypnotic has mostly been replaced by benzodiazepines and related medications, but it is still used regularly to sedate children for medical procedures. In addition, some clinicians prefer to prescribe it because it is relatively inexpensive and it is a schedule IV medication. Probably as a matter of tradition, chloral hydrate is often used as a hypnotic agent during industry-sponsored clinical trials of novel psychotropic drugs.

Chemistry. Chloral hydrate is synthesized by adding a molecule of H_2O to the carbonyl group of chloral, 2,2,2-trichloroacetaldehyde. The hypnotic effect of chloral hydrate is attributable to an active metabolite (trichloroethanol).

Pharmacologic Actions

PHARMACOKINETICS. Chloral hydrate is quickly and completely absorbed from the GI tract after oral administration. It is rapidly converted by a hepatic dehydrogenase to trichloroethanol, which has an elimination half-life of 4 to 12 hours and is metabolized by alcohol and aldehyde dehydrogenases to the inactive trichloroacetic acid; it is also conjugated with glucuronide to urochloralic acid, which is excreted by the kidney. About 10% of the total chloral hydrate dose is excreted unchanged. Plasma concentrations reach their peak in 30 to 60 minutes. Chloral hydrate and trichloroethanol are highly lipid-soluble, readily penetrate the blood–brain barrier and the placenta, and are excreted in breast milk. In newborns, the elimination half-life of trichloroethanol is around 28 hours; it decreases to 10 hours in toddlers. Repetitive dosing in infants causes drug accumulation, and trichloroacetic acid has been detected 6 days after exposure of infants to the drug is discontinued.

PHARMACODYNAMICS. Most of the pharmacologic action of chloral hydrate is attributable to its major metabolite, tricholoroethanol, which increases GABAA receptor affinity in a manner similar to barbiturates, increasing chloride ion influx and reducing neuronal excitability.

Design and Interpretation of Clinical Drug Studies. Chloral hydrate was used at one time to treat alcohol withdrawal, but the effectiveness of this application was never supported by controlled trials. Comparisons of chloral hydrate and benzodiazepines suggested equivalent efficacy for short-term treatment of insomnia, but the benefits of chloral hydrate diminished after 2 weeks due to tolerance. Concerns about genotoxicity, carcinogenesis, and lethality in overdose have led to recommendations that chloral hydrate not be used routinely as a hypnotic, especially in older people.

Chloral hydrate is often used as a sedative during pediatric procedures such as echocardiography, MRI, and other radiologic examinations. A randomized study found that chloral hydrate was as effective as midazolam in producing adequate sedation during pediatric echocardiography.

Effects on Specific Organs and Systems

CENTRAL NERVOUS SYSTEM. The CNS depressant action of chloral hydrate produces drowsiness, sedation, lightheadedness, malaise, ataxia, and nightmares. Hangover effects seen with other hypnotics are rare with chloral hydrate. Chloral hydrate has few analgesic properties and occasionally associated causes agitation when taken by patients with pain. CNS effects of an overdose of chloral hydrate are similar to an overdose of other CNS depressants. However, the lethal dose (5 to 10 g) is only 2.5 to 10 times greater than the hypnotic dose (500 mg to 2 g).

CARDIOVASCULAR SYSTEM. Chloral hydrate and trichloroethanol reduce myocardial contractility, shorten the refractory period, and increase the sensitivity of the myocardium to endogenous catecholamines. As a result, most deaths from chloral hydrate poisoning are caused by cardiac dysrhythmias that are refractory to standard antiarrhythmics. Adrenergic-blocking agents may be helpful in this setting.

RESPIRATORY SYSTEM. In typical clinical doses, there is little effect of chloral hydrate on respiration. However, an overdose may cause respiratory depression.

GASTROINTESTINAL SYSTEM. Chloral hydrate may cause GI irritation, resulting in nausea and vomiting. Because GI toxicity is more likely when the medication is ingested on an empty stomach, the medication is usually taken with meals. Overdoses can cause hemorrhagic gastritis and occasionally gastric or intestinal necrosis with perforation, as well as esophagitis with stricture.

Therapeutic Indications

INSOMNIA. Chloral hydrate is indicated for the short-term treatment of insomnia. Most patients respond to a dose of 500 to 1,500 mg, with sleep taking place within 1 hour. Chloral hydrate may be helpful for some medically ill and younger patients and for some older adult patients who become agitated or confused with a benzodiazepine, but alternative treatments with a lower risk of toxicity such as suvorexant, ramelteon, sedating antidepressants and sedating antipsychotics should be considered first.

SEDATION FOR PEDIATRIC PROCEDURES. Chloral hydrate is often used to sedate children undergoing diagnostic and minor surgical procedures. In this population, it has a rapid onset of action when administered orally or rectally. Children younger than 2 do not respond as reliably. The usual dose in older children is 25 to 100 mg/kg, with a maximum dose of 1 g. Careful monitoring is necessary until the child has completely regained consciousness, especially if the dose of chloral hydrate is repeated.

PREOPERATIVE USE. Chloral hydrate can be used in low doses preoperatively to reduce anxiety associated with surgery and to induce sleep without excessively depressing respiration or the cough reflex, although close monitoring of respiratory status is important. There is no evidence that chloral hydrate is superior to benzodiazepines or z-drugs for this purpose.

Precautions and Adverse Reactions

CENTRAL NERVOUS SYSTEM. Chloral hydrate can cause sedation, dizziness, lightheadedness, confusion, and, occasionally, paradoxical excitement. The CNS depressant effect of chloral hydrate is additive to that of other CNS depressants, especially barbiturates and alcohol.

ABUSE AND DEPENDENCE. Similar to many other CNS depressants, chloral hydrate may be habit-forming, particularly in those who have a history of alcohol or other drug dependence. Tolerance and physiologic dependence may develop if chloral hydrate is used for long periods of time on a regular basis. Sudden withdrawal of chloral hydrate can precipitate a syndrome that is similar to barbiturate withdrawal and can be diagnosed and treated using the barbiturate tolerance test described above.

OVERDOSE. Signs and symptoms of overdose of chloral hydrate are similar to those of barbiturate overdose. Initial features are ataxia and lethargy, followed by stupor or coma. Pupils are initially constricted but then become dilated as the coma deepens. Toxic doses for adults are approximately 5 to 10 g. Treatment of overdose includes gastric lavage and supportive measures. Hemodialysis may be effective in promoting the clearance of trichloroethanol.

ALLERGIC REACTIONS. Allergic skin reactions have been reported with chloral hydrate, including hives, eczematoid dermatitis, and urticaria. Occasionally, its use has been associated with leukopenia and eosinophilia.

GASTROINTESTINAL SYSTEM. Chloral hydrate causes GI irritation, especially when taken on an empty stomach. Common symptoms include nausea, vomiting, diarrhea, and an unpleasant taste.

PORPHYRIA. Like the barbiturates, chloral hydrate may precipitate attacks of acute intermittent porphyria.

CARCINOGENICITY. Animal studies suggest carcinogenicity of trichloroethylene, which appears to increase the risk of hepatic and renal cancer. Research suggesting no compelling association between chloral hydrate and cancer in humans has been underpowered and has involved only short-term use of chloral hydrate.

Given these considerations, chloral hydrate is contraindicated in the presence of gastric ulcers, hepatic insufficiency, porphyria, respiratory insufficiency, use of anticoagulants, and hypersensitivity to the medication, and caution is necessary for patients who are prone to cancer.

Drug Interactions. Chloral hydrate competes with ethanol for alcohol dehydrogenase, resulting in an increase in the half-life of ethanol by concurrent chloral hydrate administration. At the same time, ethanol inhibits the conversion of chloral hydrate to trichloroethanol as well as the conjugation of trichloroethanol. The combination of ethanol and chloral hydrate (Mickey Finn or knockout drops) produces rapid unconsciousness and has been notorious for surreptitious use for criminal purposes. Chloral hydrate has an unpredictable effect on warfarin levels and may cause hypoprothrombinemia. Combining chloral hydrate with furosemide can lead to vasomotor instability due to displacement of furosemide from protein-binding sites resulting in increased free drug, resulting in vasodilation, flushing, tachycardia, sweating, and hypotension.

Laboratory Interference. Because they are radioopaque, chloral hydrate capsules may be detected on abdominal X-ray. Chloral hydrate may interfere with copper sulfate tests for glycosuria. It may also interfere with tests for urine catecholamines or 17-hydroxycorticosteroid.

Dosage and Administration. Chloral hydrate is available as a capsule, a liquid, and a suppository. Chloral hydrate capsules are formulated in 250- or 500-mg strengths. Chloral hydrate syrup is available in a 500 mg per 5 mL concentration. The oral syrup should be mixed with 4 to 8 oz of water or juice to reduce GI distress. Suppositories are formulated in 325- and 650-mg strengths. The usual adult dose for insomnia is 500 to 1,000 mg, taken 15 to 30 minutes before bedtime. Daily adult doses should not exceed 2 g. Pediatric doses vary from 25 to 100 mg/kg per dose, with a maximum dose of 1 g/day.

OTHER SEDATIVE-HYPNOTICS

Over the years, a number of medications with varying structures with actions similar to barbiturates have been used as sedative-hypnotics.

Glutethimide was introduced as an alternative to barbiturates. However, its absorption was erratic, and it had more than 14 metabolites, some contributing to toxicity. In addition, recirculation in the enterohepatic circulation contributed to fluctuating, prolonged toxicity alternating with withdrawal following overdose. Glutethimide was unique in causing anticholinergic side effects, leading to thick bronchial secretions that further compromised respiration after overdose. Psychosis, seizures, peripheral neuropathy, and ataxia occur with prolonged use. These problems led to the withdrawal of glutethimide in the United States.

Meprobamate was introduced in 1950 as a muscle relaxant. Carisoprodol (Soma), which was introduced 5 years later, is metabolized to meprobamate. Both of these propanediol carbamates have actions on the GABAA receptor and on NMDA receptors similar to the barbiturates and are used as muscle relaxants. Meprobamate can cause dysrhythmias, and it is more likely than other drugs in the group to cause euphoria. Meprobamate pills can agglutinate in the stomach to form concretions, leading to recurrent toxicity after overdose. Both meprobamate and carisoprodol can precipitate acute episodes of porphyria.

Paraldehyde was used parenterally for agitation in alcohol withdrawal, but it caused sterile abscesses, and the drug was withdrawn.

Ethchlorvynol was removed from the market because it caused severe withdrawal syndromes and was particularly lethal after overdose.

Anesthetic Agents

Propofol. Propofol is a GABAA that also induces the presynaptic release of GABA and dopamine (the latter possibly through action on GABAB heteroreceptors on dopaminergic neurons); it is also a partial agonist at NMDA and dopamine D2 receptors. Because it is very lipid-soluble, it readily crosses the blood–brain barrier, inducing anesthesia in less than 1 minute. Rapid redistribution out of the CNS terminates propofol action within 3 to 8 minutes of discontinuing the infusion. Propofol is well tolerated in conscious sedation, but it can cause acute adverse effects that include respiratory depression, apnea, and bradyarrhythmias. Prolonged infusion can cause acidosis and myopathies. The infusion liquid contains a soybean emulsion carrier that can serve as a culture medium for multiple organisms and that can impair macrophage function and cause anaphylactic reactions and hematologic and lipid abnormalities. As mentioned earlier, unmonitored use of propofol, especially for off-label indications such as insomnia, can be extremely dangerous.

Etomidate. Etomidate is a carboxylated imidazole that acts at the beta-2 and beta-3 subunits of the GABAA receptor. Etomidate has a rapid onset (1 minute) and short duration (less than 5 minutes) of action. As is true of parenteral lorazepam preparations, the propylene glycol vehicle used in etomidate infusions can cause hyperosmolar metabolic acidosis. Proconvulsant and anticonvulsant properties of etomidate can have unpredictable effects, and inhibition of cortisol release can alter hypothalamic–pituitary–adrenal axis function after long-term use.

Dexmedetomidine. Dexmedetomidine is an intravenous selective alpha-2 agonist, similar to clonidine but far more selective. While it does not act at the GABA receptor complex, it is often used as an adjunct or alternative to barbiturates and related medications. It is used for sedation in the ICU and for procedures because of sedative and analgesic actions. The former is mediated by dose-dependent postsynaptic alpha 2 receptor binding in the locus ceruleus, and the latter by alpha 2 receptor binding in the spinal cord. An open randomized controlled comparison of dexmedetomidine, superficial cervical plexus blockade or placebo prior to induction of anesthesia for thyroidectomy found both active treatments to be equivalently superior to placebo in reducing postoperative pain. Unlike agents acting on the GABA complex, dexmedetomidine does not produce respiratory depression. Low-dose intravenous dexmedetomidine can decrease mean arterial pressure; higher-dose intravenous dexmedetomidine loses alpha-2 selectivity and can cause hypertension. Common adverse effects include hypotension, hypertension, dry mouth, dizziness, and nausea; less common adverse effects include fever, cyanosis, and muscle weakness. More serious adverse involve cardiac arrhythmias. Dosage for premedication is 0.33 to 0.67 mg/kg/iv or 2.5 ug/kg IM given 15 minutes prior to surgery.

Although the safety and efficacy of dexmedetomidine in children are not established, moderately reliable evidence suggests that perioperative administration may reduce postoperative pain and agitation in juvenile patients. The mechanism of such benefits is not clear. Concerns remain with regard to bradycardia and hypotension among pediatric patients.

FURTHER READINGS

Alali AS, Naimark DM, Wilson JR, et al. Economic evaluation of decompressive craniectomy versus barbiturate coma for refractory intracranial hypertension following traumatic brain injury. *Crit Care Med*. 2014;42:2235–2243.

*Ball C, Westhorpe R. The history of intravenous anesthesia: the barbiturates. Part 1, Part 2, Part 3. *Anesthesia and Intensive Care*. 2001;29:3, 219, 323.

*Bigal ME, Lipton RB. Excessive acute migraine medication use and migraine progression. *Neurology*. 2008;71:1821–1828.

Browne ML, Van Zutphen AR, Botto LD, et al. Maternal butalbital use and selected defects in the national birth defects prevention study. *Headache*. 2014;54:54–56.

Caldwell JC, Keshava N. Key issues in the modes of action and effects of trichloroethylene metabolites for liver and kidney tumorigenesis. *Environmental Health Perspectives.* 2006;114:1457–1463.

Da Silva AN, Lake AE 3rd. Clinical aspects of medication overuse headaches. *Headache.* 2014;54:211–217.

Dhorajiya BD, Dholakiya BZ, Ibrahim AS, Badria FA. Nucleobase-based barbiturates: their protective effect against DNA damage induced by bleomyciun-iron, antioxidant, and lymphocyte transformation assay. *Bio Med Res Int*. 2014;2014:898670.

*Evans RW, Baskin SM: Why do migraineurs abuse butalbital-containing combination analgesics? *Headache.* 2010;50:1194–1197.

*Fourati Z, Ruza RR, Laverty D, et al. Barbiturates bind in the GLIC ion channel pore and cause inhibition by stabilizing a closed state. *J Biol Chemistry.* 2017;292:1550–1558.

*Giovannitti T. Alpha-2 adrenergic receptor agonists: A review of current clinical applications. *Anesthesia Progress.* 2015;62(1):31–38.

Haselkorn T, Whittemore AS, Udaltsova N, Friedman GD. Short-term chloral hydrate administration and cancer in humans. *Drug Safety.* 2006;29:67–77.

*Kochanek PM, Tasker RC, Bell MJ, et al. Management of pediatric Severe traumatic brain injury: 2019 consensus and guidelines-based algorithm for first and second tier therapies. *Pediatr Crit Care Med*. 2019;20(3):269–279.

Liu D, Pan L, Gao Y, et al. Efficaciousness of dexmedetomidine in children undergoing cleft lip and palate repair: a systematic review and meta-analysis. *BMJ Open.* 2021;11(8):e046798.

*Nelson LS, Howland MA, Lewin NA, et al. (eds). *Goldfrank's Toxicologic Emergencies*. 11th ed. McGraw-Hill; 2022.

*Peng L, Min S, Wei K, ZiemannGimmel P. Different regimens of intravenous sedatives or hypnotics for electroconvulsive therapy (ECT) in adult patients with depression. *Cochrane Anaesthesia, Critical and Emergency Care Group Cochrane Database of Systematic Reviews*. 2017;3.

*Rosa MA, Rosa MO, Marcolin MA, Fregni F. Cardiovascular effects of anesthesia in ECT: a randomized, double-blind comparison of etomidate, propofol and thiopental. *J ECT*. 2007;23:6–8.

Shao H, Zhang Y, Dong Y, et al. Chronic treatment with anesthetic propofol improves cognitive function and attenuates caspase activation in both aged and Alzheimer's disease transgenic mice. *J Alzheimers Dis.* 2014;41:499–513.

*Smith DE, Wesson DR. Phenobarbital technique for treatment of barbiturate dependence. *Arch Gen Psychiatry.* 1971;24:56–60.

Stephen MC, Matthew J, Varghese AM, Kurien M, Mathew GA. A randomized controlled trial comparing intranasal midazolam and chloral hydrate for procedural sedation in children. *Otolaryngol Head Neck Surg*. 2015;153:1042–1050.

Weaver MF. Prescription sedative misuse and abuse. *Yale J Biol Med*. 2015;88:247–256.

*Wheeler DS, Jensen RA, Poss WB. A randomized, blinded comparison of chloral hydrate and midazolam sedation in children undergoing echocardiography. *Clin*

Pediatr. 2001;40:381–387.

*Xing J, Wu R, Liang L, et al. Comparison of the analgesic effects of intravenous infusion of dexmedetomidine versus bilateral superficial cervical plexus block after thyroidectomy: a randomized controlled trial. *Clin J Pain*. 2021;37:623–628.

Yaekel JW, Logan BK. Butalbital and driving impairment. *J Forensic Sci*. 2013;58:941–945.

Yu Z, Cohen JB. Enantiomeric barbiturates bind distinct inter-and intrasubunit sites in a nicotinic acetylcholine receptor (nAChR). *J Biol Chem*. 2017;292:17258–17271.

Zeller A, Arras M, Jurd R, Rudolph U. Identification of a molecular target mediating the general anesthetic actions of pentobarbital. *Mol Pharmacol*. 2007;71:852–859.

*Zhang Q, Yu Y, Lu Y, Yue H. Systematic review and meta-analysis of propofol versus barbiturates for controlling refractory status epilepticus. *BMC Neurology.* 2019;19(1):55.

*Zwerus R, Absolom A. Update on anesthetic neuroprotection. *Current Opinion in Anesthesiology.* 2015;28:424–430.

33.9b Benzodiazepine Receptor Agonists and Antagonists

STEVEN L. DUBOVSKY, M.D., AND DORI MARSHALL, M.D.

Alcohol has been used as a sedative since biblical times. Bromides were introduced as hypnotics in the mid-19th century. Barbital, the first barbiturate, was introduced in 1903, and phenobarbital was synthesized in 1912. In the mid-1940s, mephenesin was found to have calming properties without sedation in rodents, but it was not effective in humans; its shortcomings were overcome by the development of meprobamate, which became a blockbuster in the mid-1950s even though it did not outperform placebo in clinical trials. Because it produced tranquil feelings, barbital was called a “tranquilizer” and later a “minor tranquilizer.” In the mid-1950s, competitive pressure facilitated the synthesis of the first benzodiazepine, chlordiazepoxide. Chlordiazepoxide was approved by the FDA in 1960, followed by the approval of diazepam in 1963. By 2020, all benzodiazepines currently marketed in the United States were available as generics.

Benzodiazepines were originally introduced in the 1960s as safe alternatives to barbiturates and related drugs as anxiolytics and hypnotics. In 1975, 60 million retail prescriptions were written in the United States, or 0.38 prescriptions per person, and by 1977, benzodiazepines were the most prescribed medications worldwide for anxiety, insomnia, agitation, seizures, muscle spasm, and anesthesia premedication. Subsequently, concerns about risks, particularly of misuse, led to the view of benzodiazepines as being potentially too dangerous for anything but short-term use, an attitude that in some settings led to extensive regulation and declining use for chronic anxiety. The year after a requirement for triplicate prescribing of benzodiazepines was introduced in New York State in 1989, benzodiazepine prescriptions decreased by 60%, with the largest reduction being for the treatment of seizure disorders, while prescriptions for the more dangerous meprobamate, chloral hydrate, and butabarbital increased by 1.25%, 136%, and 30%, respectively. As benzodiazepine discontinuation syndromes due to abrupt withdrawal increased, so did overdoses of the more dangerous alternative drugs.

At about the same time, an American Psychiatric Association task force review of over 2,000 studies concluded that benzodiazepines are rarely abused, and when they are, it is usually by people who abuse other substances. Although the task force concluded that most benzodiazepine prescriptions are for appropriate indications, many practitioners abandoned benzodiazepines in favor of antidepressants. As the use of SSRIs increased for anxiety and depression, prescriptions for benzodiazepines continued to decline even though there was no evidence for the superiority of SSRIs. The preferential use of SSRIs was also driven by studies of effectiveness in the new anxiety disorders that were introduced in DSM-III, whereas benzodiazepines were generally not studied for these indications because they were off patent by the time the nomenclature changed.

Despite these developments, clinicians have continued to prescribe benzodiazepines. A national cross-sectional survey of U.S. adults (*N* = 154,171) found that the self-reported rate of filling more than 2 benzodiazepine prescriptions in a year (i.e., repeated use) increased from 1.8% in 2002 to 2003 to 2.4% in 2015 to 2016. The increase in benzodiazepine prescriptions was largely attributable to an increase in repeated rather than single prescriptions, the latter remaining essentially the same at 1% to 1.8%. The implication is that the number of people chronically taking benzodiazepines increased, not that the use of benzodiazepines increased in the general population. Another serial cross-sectional national representative survey of 386,457 office-based physician visits between 2003 and 2015 found that visits for a benzodiazepine prescription essentially doubled (from 3.8% to 7.4%) over that time. Visits to psychiatrists stayed the same, the increase being attributable to visits to primary care physicians, who prescribed benzodiazepines slightly more frequently for anxiety, depression, and neurologic indications and much more for back pain and chronic pain, and who co-prescribed benzodiazepines with narcotics and other CNS depressants (nonbenzodiazepine sedative hypnotics, muscle relaxants, and antipsychotic drugs). In 2015 to 2016, 30.6 million persons reported past-year use of benzodiazepines. Not counting the “Z-drugs” (zolpidem, zaleplon, and eszopiclone), in 2017, there were 69.8 million prescriptions or 0.215 per person. In 2019, 92 million benzodiazepine prescriptions were filled, half of them for people taking a benzodiazepine for at least 2 months. Somewhere between 25% and 76% of patients who are prescribed benzodiazepines, or 2% to 7.5% of the general population, use benzodiazepines long-term. As many as half of chronic benzodiazepine users list insomnia as the primary reason for taking the medication.

Studies of the actions of the benzodiazepines and barbiturates have led to the elucidation of the structure and function of their primary target, the GABA–benzodiazepine receptor complex. This research has led to the development of other medications that act on this complex as agonists, antagonists, and inverse agonists. Understanding benzodiazepine receptors and their interaction with linked receptors makes it straightforward to predict the actions of a diverse group of medications.

GABA–BENZODIAZEPINE RECEPTOR COMPLEX

Gamma-aminobutyric acid (GABA), the major inhibitory neurotransmitter in the central nervous system (CNS), interacts with ionotropic GABA heteroreceptors. The pentameric GABA receptor complex is a member of the nicotinicoid superfamily and includes 6 subunit families (α, β, γ, ρ, δ, θ) and 16 specific subunits that are encoded by separate genes and are allosterically linked to form a chloride ion channel (Fig. 33.9b–1). Several GABA receptor subtypes have been identified.

GABA_A receptor activity is regulated by a number of transmembrane proteins, some of which are linked to other receptors such as the AMPA receptor. A GABA_B receptor, which is linked by a G protein to calcium and potassium channels, is a presynaptic autoreceptor that modulates the release of GABA, glutamate, and monoamines. A structurally and functionally distinct oligomeric GABA$_C$ (also called GABA$_P$) receptor composed of five subunits in a pentameric arrangement has been identified in the retina, spinal cord, superior colliculus, pituitary, and gut. These receptors appear to be involved in the regulation of vision, memory, and the sleep–wake cycle. Drugs targeting this receptor may be useful for visual, sleep, and cognitive disorders. The balance between inhibitory GABAergic, inhibitory glycinergic, and excitatory glutaminergic systems is fine-tuned to set the responsiveness of the CNS to a host of neurotransmitters and neuromodulators.

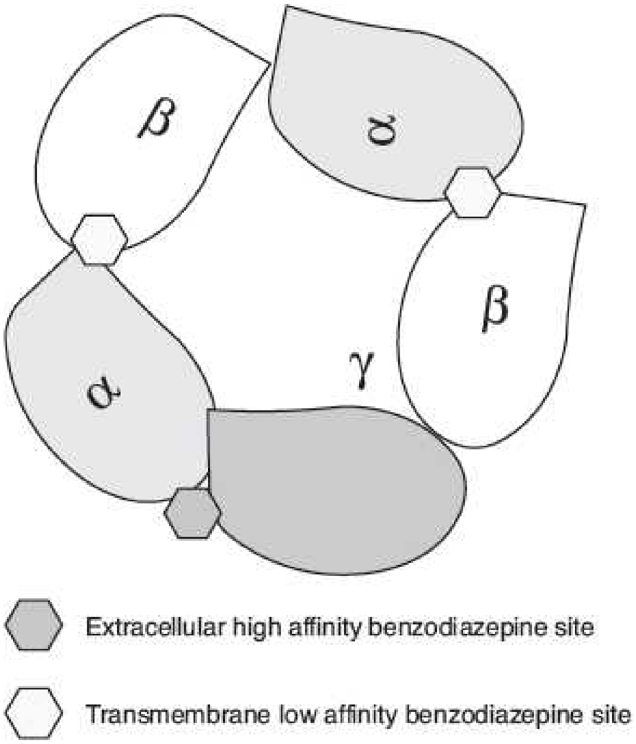
β
α
β
γ
α
Extracellular high affinity benzodiazepine site
Transmembrane low affinity benzodiazepine site

FIGURE 33.9b–1. Components of the benzodiazepine–GABA receptor complex.

The benzodiazepine receptor complex (Fig. 33.9b–1) consists of two α, two β, and one γ subunit from a pool of α, β, γ, δ, ε, π, and θ subunits allosterically linked to the GABA receptor to form a pentameric, GABA-gated chloride ion channel. Benzodiazepines bind to an extracellular site at the α/γ interface; there is also a low-affinity benzodiazepine-binding site in the transmembrane region of the β/α interface, which is also a binding site for propofol and etomidate. Many benzodiazepines bind at multiple sites. In addition to a receptor for GABA and for benzodiazepines, the

GABA receptor complex contains receptors for a number of ligands that influence the conformation of the central anion pore (Fig. 33.9b–2A).

Interaction of the postsynaptic $GABA_A$ receptor with GABA alters the conformation of the GABA receptor, increasing chloride channel openings and allowing more negative charges into the cytosol (Fig. 33.9b–2B), hyperpolarizing it and decreasing the excitability of neurons that regulate vigilance, emotions, cognition, arousal, and muscle activity.

Occupancy of the benzodiazepine receptor does not affect the chloride ion channel directly but changes the conformation of the GABA receptor to increase its affinity for GABA, resulting in more chloride ion influx per GABA molecule (Fig. 33.9b–2C). Benzodiazepine receptor agonists, therefore, have the effect of hyperpolarizing neurons on which the receptor complex is located. A picrotoxin site associated with the benzodiazepine–GABA receptor complex reduces GABA receptor affinity, reducing chloride ion influx and tending to depolarize the same neuron. A number of other substances, including inhaled and intravenous (IV) anesthetics, barbiturates, neurosteroids and alcohol, alter $GABA_A$ receptor affinity at various binding sites, not all of which have been identified.

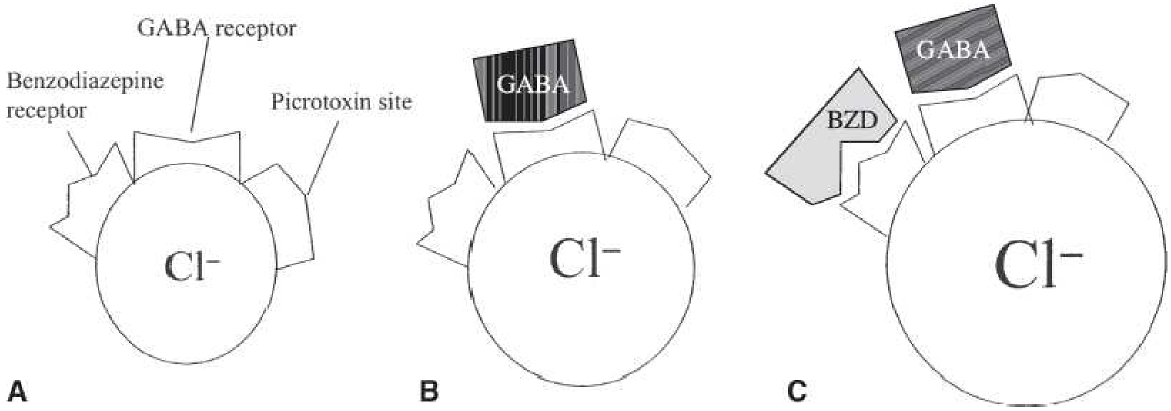

FIGURE 33.9b–2. The GABA–benzodiazepine receptor complex is comprised of multiple allosterically linked receptors that alter the conformation of the central chloride ion channel directly, or indirectly by modifying GABA receptor affinity (**A**). Occupancy of the GABA receptor increases chloride ion channel openings, hyperpolarizing the neuron (**B**). Benzodiazepine receptor occupancy increases the affinity of the GABA receptor, resulting in more chloride ion channel openings and more hyperpolarization, per molecule of GABA (**C**). The picrotoxin site has the opposite effect, reducing GABA receptor affinity and depolarizing the neuron.

Three major benzodiazepine receptor subtypes have been described, two of them called central benzodiazepine receptors, and one called a peripheral benzodiazepine receptor (PBR). Benzodiazepine-1 (Ω-1) receptors contain the α1 subunit and are located throughout the CNS, particularly the reticular activating system, cerebellum and cerebral cortex, mediating sleep and anti-anxiety effects. Benzodiazepine-2 receptors, which are a heterogeneous group of receptors containing α2, α3, or α5 subunits, are located in the cortex, hippocampus, striatum, and spinal cord and on pyramidal neurons and mediate anxiolysis, muscle relaxation, CNS depression, sedation, and psychomotor impairment and contribute to the anticonvulsant effect. Type 3 receptors in peripheral tissues may mediate dependence.

Benzodiazepine-3 (Ω-3) receptors (PBRs, also known as translocator proteins [TSPOs]) are 18-kDa proteins with a structure and function distinct from that of the central (i.e., benzodiazepine-1 and benzodiazepine-2) receptors. PBRs are found on glial and other brain cells as well as throughout the body. These receptors, which bind benzodiazepines and the endogenous inverse agonist, are primarily located in the outer mitochondrial membrane and may contribute to tolerance and withdrawal. They are also involved in a number of complicated processes, including the synthesis of neurosteroids implicated in some psychiatric disorders and the regulation of apoptosis. The mitochondrial PBR is linked to a mitochondrial permeability transition pore (MPTP) consisting of a nonselective ion channel, which alters the transmembrane potential, leading to the release of proapoptotic proteins from the mitochondrion and other actions. The PBR plays an important role in protecting neurons from oxidative damage, which is implicated in Parkinson disease, Alzheimer disease, and negative schizophrenia. At micromolar concentrations at the receptor, PBR ligands may promote apoptosis, whereas nanomolar concentrations may, under some circumstances, interfere with it. An agonist effect at this receptor could protect against oxidative damage, whereas an inverse agonist effect (described below) could play a role in cancer chemotherapy, which enhances apoptosis. Carbamazepine binds to PBRs in lymphocytes, which may contribute to altered chemotaxis by these cells in carbamazepine-treated patients.

BENZODIAZEPINE RECEPTOR LIGANDS

Compounds that bind to the benzodiazepine receptor can have five different kinds of action. Nonselective full agonists (e.g., benzodiazepines) cause all benzodiazepine receptors to increase the affinity of GABA receptors, which increases the frequency of chloride channel openings and hyperpolarizes neurons. Selective full agonists (e.g., zolpidem) have the same effect, but this action is selective for the benzodiazepine-1 receptor. Avoiding actions on benzodiazepine-2 receptors and PBRs reduces the risk of adverse effects such as psychomotor impairment, tolerance and withdrawal, although these still occur. Dobazam is the only benzodiazepine with more affinity for benzodiazepine-2 than for benzodiazepine-1 receptors. Partial agonists bind to all benzodiazepine receptors, but they are less effective than full agonists. If low levels of an endogenous agonist for a receptor are present, partial agonists occupy the receptor and stimulate it, although less so than the full agonist. If levels of the endogenous agonist are high, a partial agonist competes for receptors but is less effective than the full agonist, essentially reducing activity at the receptor. In the case of the benzodiazepine receptor, partial agonists have a less pronounced therapeutic effect at the receptor than full agonists, but they also have less severe adverse effects. Inverse agonists occupy receptors in the same manner as full agonists, but they have the opposite effect. An inverse agonist of the benzodiazepine receptor binds to the same site as the benzodiazepines but reduces GABA receptor affinity, decreasing the frequency of chloride channel openings and depolarizing neurons. Inverse agonists produce arousal and anxiety and sometimes cause seizures and psychosis. Antagonists occupy a receptor without affecting its activity, thereby blocking the action of agonists and inverse agonists. Flumazenil is an intravenous (IV) benzodiazepine receptor antagonist used to treat benzodiazepine overdose. It has a brain half-life of 30 minutes and a terminal half-life of 1 hour, necessitating repeated doses for treatment of overdose with longer-acting benzodiazepines.

The existence of benzodiazepine receptors would appear to imply the existence of an endogenous ligand for them. In fact, there are endogenous benzodiazepine receptor ligands called endozepines that act on central and peripheral benzodiazepine receptors to participate in the regulation of arousal, neurosteroid biosynthesis, neuropeptide expression, neurogenesis, and pro-inflammatory cytokine secretion. An 86-amino acid endozepine, the gene for which is found in astroglial cells and peripheral organs, has been identified that is called diazepam-binding inhibitor (DBI) because it displaces labeled diazepam from brain-binding sites. DBI, also called acyl-CoA-binding protein (ACBP), acts as a benzodiazepine inverse agonist, promoting arousal. The function of DBI/ACBP appears to fine-tune the set point of arousal in conjunction with other endogenous substances acting on different points on the benzodiazepine–GABA receptor complex to depress or activate the CNS.

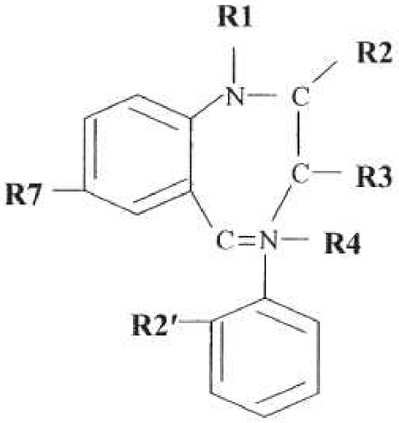

Benzodiazepine	R1	R2	R3	R7	R2′
Alprazolam	Fused triazolo ring		-H	-Cl	-H
Chlordiazepoxide	-	-$NHCH_3$	-H	-Cl	-H
Clonazepam	-H	=O	-H	NO_2	-Cl
Diazepam	-CH_3	=O	-H	-Cl	-H

FIGURE 33.9b–3. Basic benzodiazepine structure.

A number of medications and substances directly or indirectly alter the activity of the GABA–benzodiazepine receptor complex. For example, picrotoxin binds to the GABA receptor at a site distinct from the benzodiazepine receptor and reduces chloride channel openings to induce seizures. Alcohol and barbiturates have direct effects on GABA receptors and chloride channels and possibly on the picrotoxin site, reducing cortical and limbic neuronal activity. Valproate enhances GABA release, and gabapentin has unclear actions on GABAergic neurotransmission. Propofol, an IV anesthetic discussed in a separate chapter, acts at a site other than the benzodiazepine receptor to increase GABA affinity in a manner similar to benzodiazepines. Unidentified constituents of valerian, an herbal preparation used for centuries as a hypnotic and anxiolytic, seem to be GABA receptor agonists. The anticonvulsant tiagabine inhibits GABA reuptake, increasing GABA action at all GABA receptors, and the anticonvulsant vigabatrin reduces GABA breakdown. Baclofen is a $GABA_B$ agonist that reduces spasticity via an action on presynaptic receptors on excitatory spinal cord neurons. This chapter reviews medications that act primarily on the benzodiazepine-$GABA_A$ receptor complex.

BENZODIAZEPINES

Neuroscience-Based Nomenclature (NbN)	
Alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, estazolam, flurazepam, flunitrazepam, lorazepam, midazolam, oxazepam, temazepam, triazolam, quazepam	
Pharmacologic Domain	GABA
Mechanism of Action	Positive Allosteric Modulator (GABA-A receptor, benzodiazepine site)

Chemistry

Benzodiazepines contain two benzene rings and a seven-member diazepine ring. Most benzodiazepines in current clinical use contain a 5-aryl substituent ring and a 1,4-diazepine ring and therefore are 5-aryl-1,4-benzodiazepines (Fig. 33.9b–3). A fused triazolo ring at positions RI and R2 on the diazepine ring in triazolobenzodiazepines such as alprazolam, triazolam, and estazolam, increases potency; the same is true of a fused imidazole ring at these positions in midazolam. Replacement of ring C with a keto substituent in position 5 and a methyl substituent in position 4 in flumazenil confers benzodiazepine antagonist properties. Clobazam, the only 1-5-benzodiazepine, is approved for adjunctive treatment of seizures associated with Lennox-Gastaut syndrome. Benzodiazepines and other medications acting on the GABA–benzodiazepine receptor complex are listed in Table 33.9b–1.

Table 33.9b–1.
Benzodiazepines, Z-Drugs, and Benzodiazepine Antagonists

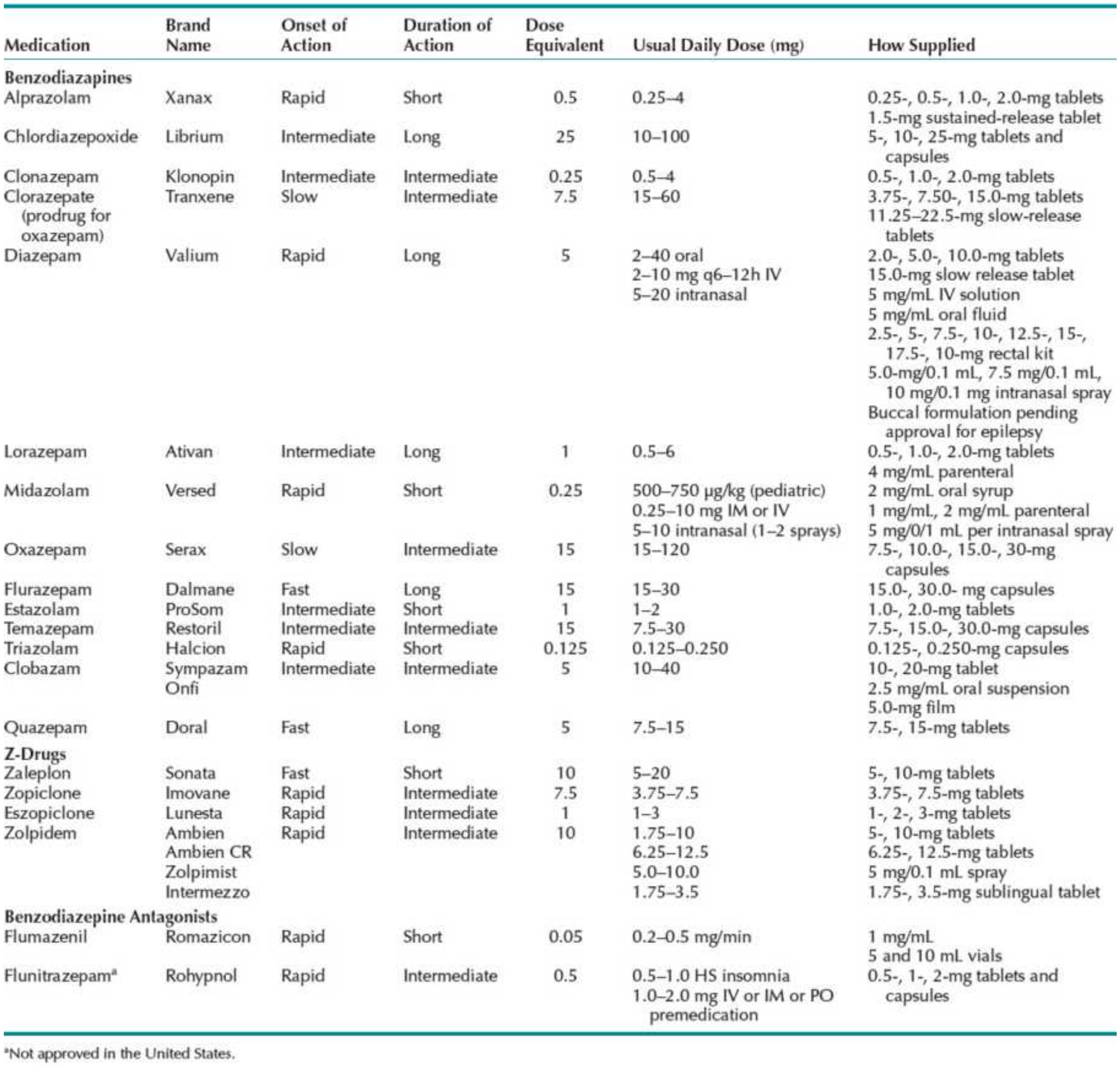

Medication	Brand Name	Onset of Action	Duration of Action	Dose Equivalent	Usual Daily Dose (mg)	How Supplied
Benzodiazepines						
Alprazolam	Xanax	Rapid	Short	0.5	0.25–4	0.25-, 0.5-, 1.0-, 2.0-mg tablets 1.5-mg sustained-release tablet
Chlordiazepoxide	Librium	Intermediate	Long	25	10–100	5-, 10-, 25-mg tablets and capsules
Clonazepam	Klonopin	Intermediate	Intermediate	0.25	0.5–4	0.5-, 1.0-, 2.0-mg tablets
Clorazepate (prodrug for oxazepam)	Tranxene	Slow	Intermediate	7.5	15–60	3.75-, 7.50-, 15.0-mg tablets 11.25–22.5-mg slow-release tablets
Diazepam	Valium	Rapid	Long	5	2–40 oral 2–10 mg q6–12h IV 5–20 intranasal	2.0-, 5.0-, 10.0-mg tablets 15.0-mg slow release tablet 5 mg/mL IV solution 5 mg/mL oral fluid 2.5-, 5-, 7.5-, 10-, 12.5-, 15-, 17.5-, 10-mg rectal kit 5.0-mg/0.1 mL, 7.5 mg/0.1 mL, 10 mg/0.1 mg intranasal spray Buccal formulation pending approval for epilepsy
Lorazepam	Ativan	Intermediate	Long	1	0.5–6	0.5-, 1.0-, 2.0-mg tablets 4 mg/mL parenteral
Midazolam	Versed	Rapid	Short	0.25	500–750 µg/kg (pediatric) 0.25–10 mg IM or IV 5–10 intranasal (1–2 sprays)	2 mg/mL oral syrup 1 mg/mL, 2 mg/mL parenteral 5 mg/0/1 mL per intranasal spray
Oxazepam	Serax	Slow	Intermediate	15	15–120	7.5-, 10.0-, 15.0-, 30-mg capsules
Flurazepam	Dalmane	Fast	Long	15	15–30	15.0-, 30.0- mg capsules
Estazolam	ProSom	Intermediate	Short	1	1–2	1.0-, 2.0-mg tablets
Temazepam	Restoril	Intermediate	Intermediate	15	7.5–30	7.5-, 15.0-, 30.0-mg capsules
Triazolam	Halcion	Rapid	Short	0.125	0.125–0.250	0.125-, 0.250-mg capsules
Clobazam	Sympazam Onfi	Intermediate	Intermediate	5	10–40	10-, 20-mg tablet 2.5 mg/mL oral suspension 5.0-mg film
Quazepam	Doral	Fast	Long	5	7.5–15	7.5-, 15-mg tablets
Z-Drugs						
Zaleplon	Sonata	Fast	Short	10	5–20	5-, 10-mg tablets
Zopiclone	Imovane	Rapid	Intermediate	7.5	3.75–7.5	3.75-, 7.5-mg tablets
Eszopiclone	Lunesta	Rapid	Intermediate	1	1–3	1-, 2-, 3-mg tablets
Zolpidem	Ambien	Rapid	Intermediate	10	1.75–10	5-, 10-mg tablets
	Ambien CR				6.25–12.5	6.25-, 12.5-mg tablets
	Zolpimist				5.0–10.0	5 mg/0.1 mL spray
	Intermezzo				1.75–3.5	1.75-, 3.5-mg sublingual tablet
Benzodiazepine Antagonists						
Flumazenil	Romazicon	Rapid	Short	0.05	0.2–0.5 mg/min	1 mg/mL 5 and 10 mL vials
Flunitrazepam[a]	Rohypnol	Rapid	Intermediate	0.5	0.5–1.0 HS insomnia 1.0–2.0 mg IV or IM or PO premedication	0.5-, 1-, 2-mg tablets and capsules

[a] Not approved in the United States.

Pharmacologic Actions

Absorption. All benzodiazepines except clorazepate are completely absorbed after oral administration and reach peak serum levels within 30 minutes to 2 hours. Metabolism of clorazepate in the stomach converts it to desmethyldiazepam, which is then completely absorbed. Intramuscular (IM) absorption of benzodiazepines other than lorazepam (Ativan) is slower than oral absorption. The onset of action is nearly immediate with intravenous administration of high-potency benzodiazepines such as midazolam.

Distribution. All benzodiazepines are lipid-soluble, but the degree of lipid solubility varies considerably between drugs. Benzodiazepines and their active metabolites are 70% to 99% protein bound, the extent of the binding being proportional to their lipid solubility. Benzodiazepines follow a two-compartment distribution, which involves a rapid central compartment phase followed by a redistribution phase to adipose tissue that ultimately determines the duration of action. As a result, a component of the long half-life of some highly lipophilic benzodiazepines such as diazepam is attributable to extensive storage in adipose tissue. At the same time, a single dose of a very lipophilic benzodiazepine (e.g., midazolam) redistributes rapidly to adipose tissue from CNS receptors, leading to a short duration of action while less lipophilic benzodiazepines like lorazepam remain at therapeutic concentrations at $GABA_A$ receptors longer because they do not redistribute as quickly. Reminazolam, a carboxylate benzodiazepine derivative with a high affinity for benzodiazepine receptors in the cerebral cortex, limbic system, midbrain and spinal cord, was developed as an alternative to midazolam for conscious sedation because it has a longer-acting metabolite that may reduce the risk of withdrawal and rebound seen with midazolam. This medication has not yet received regulatory approval. Because of high lipid solubility, high potency, and a short elimination half-life, alprazolam has a rapid onset and offset of action, with frequent interdose rebound and withdrawal that leads to increasing the dose which only intensifies symptoms when brain levels of the next dose fall.

Metabolism. Benzodiazepine pharmacokinetics vary. Oxazepam and lorazepam are conjugated directly by glucuronidation to a water-soluble form and are excreted without any intermediate metabolites. Most other benzodiazepines are oxidized by cytochrome P450 (CYP) 3A4 and 2C19, frequently to active metabolites. These metabolites may then be hydroxylated to another active metabolite (Fig. 33.9b–4). For example, diazepam is oxidized to desmethyldiazepam, which, in tum, is hydroxylated to oxazepam. These products undergo glucuronidation to water-soluble inactive metabolites that can be excreted in the urine. A number of benzodiazepines (e.g., diazepam, chlordiazepoxide, clorazepate) have the same active metabolite (desmethyldiazepam), which has an elimination half-life of more than 120 hours. Flurazepam, a lipid-soluble benzodiazepine used as a hypnotic that has a short elimination half-life, has an active metabolite (desalkylflurazepam) with a half-life greater than 100 hours. The hypnotic drug quazepam is selective for the benzodiazepine-1 receptor, but it is metabolized to desalkylflurazepam, a nonselective benzodiazepine metabolite of flurazepam that accounts for 40% of total drug activity. With chronic dosing, this medication, therefore, is similar to other longacting benzodiazepine hypnotics.

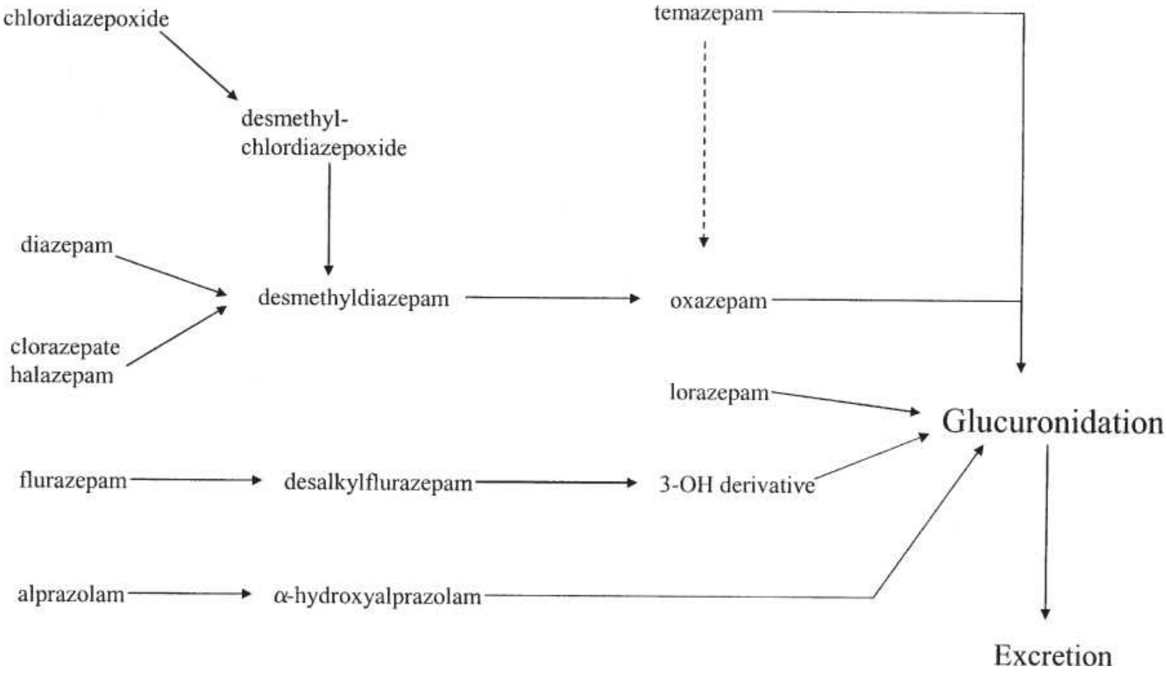
chlordiazepoxide
temazepam
desmethyl-chlordiazepoxide
diazepam
desmethyldiazepam
oxazepam
clorazepate
halazepam
lorazepam
Glucuronidation
flurazepam
desalkylflurazepam
3-OH derivative
alprazolam
α-hydroxyalprazolam
Excretion

FIGURE 33.9b–4. Simplified benzodiazepine metabolism.

The duration of action of a benzodiazepine, therefore, may not correspond to the half-life of the parent drug. Many benzodiazepines are substrates for CYP3A4 and CYP2C19. CYP3A inhibitors such as ketoconazole, nefazodone, and the antiretroviral lopinavir can substantially increase levels of CYP3A substrate benzodiazepines such as midazolam, diazepam and triazolam. Conversely, grapefruit juice, which induces CYP3A4, lowers levels of these benzodiazepines; however, even though alprazolam is a CYP3A4 substrate, it does not seem to be affected by grapefruit juice because of its high bioavailability.

Because administration of medications at shorter intervals than the elimination half-life leads to drug accumulation, medications such as diazepam and flurazepam accumulate with more than one daily dose, eventually resulting in increased daytime sedation. Clorazepate is not pharmacologically active but is a prodrug for desmethyldiazepam (nordiazepam), the primary active metabolite of diazepam and chlordiazepoxide. In contrast, temazepam, oxazepam, and lorazepam are conjugated directly and do not have active metabolites. Unlike barbiturates, benzodiazepines do not induce their own metabolism.

The combination of lipid solubility, potency, and elimination half-life of benzodiazepines (Fig. 33.9b–5) can be used to predict the onset and duration of action in clinical settings as well as the optimal frequency of drug administration. Benzodiazepines that are high in lipid solubility and potency and have a short half-life have a rapid onset but a brief duration of action and produce more intense and abrupt discontinuation syndromes. For example, midazolam, a high-potency and highly lipid-soluble nonselective full benzodiazepine receptor agonist with a short elimination halflife has an almost immediate onset of action when used intravenously for conscious sedation or anesthesia. As the medication leaves the CNS quickly, sedation diminishes rapidly, and after a single dose is followed by withdrawal symptoms (most frequently amnesia) in approximately 70% of patients. This profile makes midazolam useful for anesthesia and emergency treatment of agitation but impractical for chronic treatment. Alprazolam, which is high in lipid solubility and potency, reduces anxiety rapidly but can produce interdose withdrawal when it is administered in accordance with its intermediate half-life (three to four times per day) because it leaves the brain more rapidly than it is eliminated from the body. In contrast, chlordiazepoxide is low in lipid solubility and potency and has a long elimination half-life. This medication has a slow onset of action after a single dose and is administered less frequently in chronic treatment. Discontinuation of long half-life, low-potency, low lipidsoluble benzodiazepines such as chlordiazepoxide causes syndromes that are less intense but more prolonged than those associated with shorter-acting medications such as alprazolam.

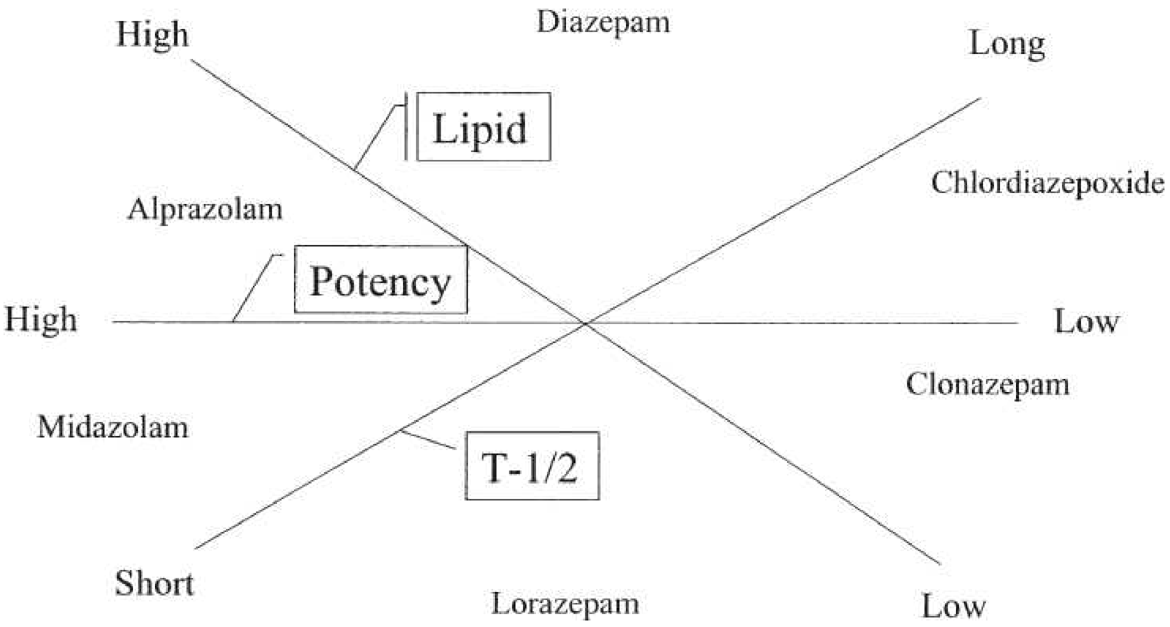

FIGURE 33.9b–5. Potency, lipid solubility, and half-life of some benzodiazepines.

Pharmacodynamics

Mechanism of Action. As noted earlier, the endogenous inverse agonist DBI may serve to upregulate the gain on the CNS by lowering the threshold for depolarization in arousal systems. DBI levels are increased in refractory epilepsy, suggesting that elevated DBI activity could be a mechanism of uncontrolled gross neuronal activation in this and other disorders associated with an excessive stress response. One cause of chronic anxiety, therefore, may be the excessive activity of DBI, which is antagonized when benzodiazepines compete for benzodiazepine receptor occupancy, a shift toward greater chloride ion influx hyperpolarizing neurons of arousal (Fig. 33.9b–2). Even if anxiety is linked to excessive excitability of arousal systems that is primary or secondary to some other influence, such as a hyperactive suffocation response (as in panic disorder), benzodiazepines can reduce anxiety by enhancing hyperpolarization of systems that underlie arousal.

The same action that results in the therapeutic action of benzodiazepines on benzodiazepine receptors in relevant locations can cause adverse effects. For example, the actions of benzodiazepines on benzodiazepine receptors in the cortex can cause cognitive and psychomotor impairment. Actions on pyramidal benzodiazepine receptors have an anticonvulsant effect and can also cause muscle relaxation and incoordination. Benzodiazepine hypnotics act on receptors in the reticular activating system, but if suppression of this system is prolonged, daytime sedation results. Suppression of the reticular activating system can induce delirium in demented patients and aggravate delirium if it is already present.

Blood Concentration and Relation to Action. Concentrations of benzodiazepines in the CSF are equal to plasma concentrations of the free drug. Because of the complex metabolism of long-acting benzodiazepines, serum levels vary considerably between individuals. Levels of parent compounds have not been correlated with therapeutic response, and because many of these medications have active metabolites, it is unlikely that such a correlation exists. Although many shorter-acting benzodiazepines do not have active metabolites, a correlation between blood levels and clinical response has not been demonstrated for these preparations either. There may be a correlation between levels of alprazolam and both adverse effects and efficacy for panic attacks, with a response more likely at plasma concentrations around 20 to 40 ng/mL, which correspond to doses of 1.5 to 6.0 mg per day. However, such levels are rarely measured in clinical practice, and levels would be expected to fluctuate substantially between doses, making the optimal time to measure the trough level difficult to identify. On the other hand, undetectable levels in an agitated patient who has been taking benzodiazepines raise the suspicion of a discontinuation syndrome. The primary uses of serum levels, therefore, are for toxicology screening and the identification of abstinence syndromes.

NONBENZODIAZEPINE AGONISTS OF THE BENZODIAZEPINE RECEPTOR

Neuroscience-Based Nomenclature (NbN)	
Zolpidem, zaleplon, zopiclone, eszopiclone	
Pharmacologic Domain	GABA
Mechanism of Action	Positive Allosteric Modulator (GABA-A receptor, benzodiazepine site)

Zolpidem, an imidazopyridine, and zaleplon, a pyrazolopyrimidine, are not benzodiazepines in that they lack a diazepine ring (Fig. 33.9b–6). Although it is technically accurate to describe these medications as nonbenzodiazepines, they act on benzodiazepine receptors, and their clinical effect can be blocked by the benzodiazepine receptor antagonist flumazenil. They are selective for the benzodiazepine-1 receptor and have no significant effect on other benzodiazepine receptor subtypes at usual doses; at higher doses, they are less selective. Alpidem is a selective benzodiazepine-1 agonist approved as an anxiolytic in some countries other than the United States. Selectivity for the benzodiazepine-1 receptor results in hypnotic and some anxiolytic action without muscle relaxant or anticonvulsant effects for zolpidem and zaleplon. Lack of effects on benzodiazepine-3 receptors appears to result in a lower incidence of withdrawal and rebound symptoms, although these may occur, especially at higher doses and with more chronic treatment. Zopiclone is a nonbenzodiazepine nonselective benzodiazepine receptor partial agonist that is in widespread use in Europe but is not available in the United States. Its S-enantiomer, eszopiclone, is approved in the United States. Zolpidem, zaleplon, and zopiclone/eszopiclone are collectively often referred to as Z-drugs. Currently available benzodiazepines and Z-drugs are listed in Table 33.9b–1, along with the benzodiazepine antagonist discussed below.

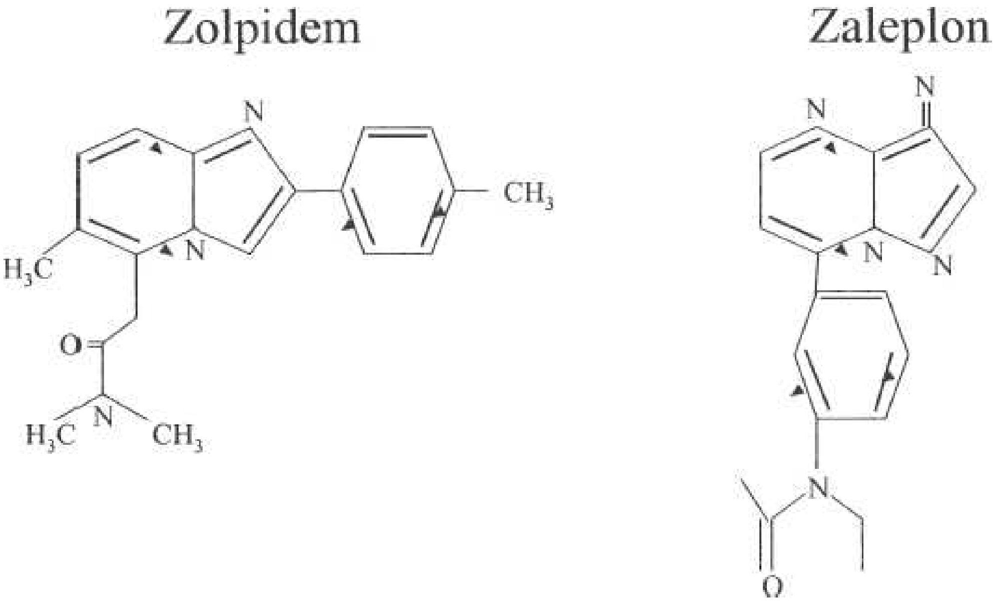

Zolpidem

Zaleplon

FIGURE 33.9b–6. Selective benzodiazepine-1 receptor agonists.

Zolpidem is rapidly absorbed after oral administration and reaches peak plasma levels in 2 to 3 hours, with 70% bioavailability; ingestion with food slows absorption. Zolpidem has an elimination half-life of approximately 2 hours and is oxidized to inactive metabolites. As a result, in most cases, a bedtime dose is completely eliminated during the night, and there is no daytime hangover, predicting greater usefulness for difficulty falling asleep than for interrupted sleep or early morning awakening. Zaleplon, which is rapidly absorbed and reaches peak plasma levels 1 to 2 hours after dosing with only 30% bioavailability due to extensive presystemic metabolism, has an elimination half-life of approximately 1.5 hours. Zaleplon is almost entirely metabolized to inactive metabolites. It, therefore, is primarily useful for difficulty falling asleep and usually does not cause a hangover. Eszopiclone is rapidly absorbed from the GI tract, with an elimination half-life of 5 to 7 hours. About 50% to 60% is weakly bound to plasma protein. Eszopiclone is extensively metabolized by CYP3A4 and CYP2E1; less than 10% of an orally administered dose is excreted in the urine, mainly as racemic zopiclone. It does not appear to stimulate or inhibit any CYP450 enzymes.

BREXANOLONE

Neuroscience-Based Nomenclature (NbN)	
Brexanolone	
Pharmacologic Domain	GABA
Mechanism of Action	Positive Allosteric Modulator (GABA-A receptor, neurosteroid site)

Approval has been granted to brexanolone (SAGE-547), a synthetic analog of the neuroactive steroid allopregnanolone that is a positive allosteric modulator of the GABA_A receptor. Used as an intravenous infusion to treat postpartum depression, brexanolone has been reported to produce rapid improvement of depressive symptoms. Phase III studies demonstrated superiority to placebo during a 60-hour infusion, with effect sizes as high as 1.2 and improvement lasting 7 to 30 days post-infusion. Brexanolone has not been compared with a benzodiazepine or any other benzodiazepine–GABA receptor complex agonist. It remains to be determined which features of postpartum depression would indicate the need for a 2.5-day continuous stay in an infusion center to receive an expensive drug but not hospitalization or electroconvulsive therapy (ECT). In addition, no data have been reported on appropriate maintenance treatment.

Zuranolone (SAGE-217) is an oral neuroactive steroid $GABA_A$ positive allosteric modulator that was granted breakthrough therapy designation for depression by the FDA. Zuranolone was well tolerated in Phase I studies for depression. In 2-week Phase II studies and a Phase III study comparing zuranolone to placebo and reference antidepressants, zuranolone was superior to placebo within a few days, with an NNT of 3 for response and remission at day 15, as well as superiority to SSRIs and SNRIs, with NNTs of 4 to 5. Zuranolone had a small negative NNH, suggesting few acute adverse effects. Some benefit was retained in studies at the last measurement on day 42. A 2-week industry-sponsored multicenter RCT of 150 women with postpartum depression (mean age 27 to 29) noted a significantly greater reduction at day 15 in HRSD-17 scores with zuranolone than placebo (–17.8 vs. –13.6 points, effect size 0.53), as well as a significantly greater reduction in anxiety. A difference in depression rating scale scores was still present at day 45. In all studies, the total sample size was small, the trial was short, and there was no comparison with a benzodiazepine, making superiority to cheaper and easier-to-use benzodiazepine–GABA agonists unclear. With these limitations, early experience suggests a rapid reduction of depressive symptoms. Long-term studies have not been performed. In an open-label study, zuranolone in the same doses used in depression studies (30 mg) improved tremor in Parkinson disease when added to a dopaminergic agent.

NONPRESCRIPTION, DESIGNER AND OTHER ILLICIT BENZODIAZEPINES

A number of substances with a 1,4-benzodiazepine/triazolo structure referred to as “designer benzodiazepines” have recently been synthesized but have never been licensed for medical use. These compounds have been designed to circumvent controlled substance regulations and are often sold on the internet. Examples include diclazepam, clonazolam, nitrazolam, adinazolam, cloniprazepam, diclazepam, etizolam, flubromazepam, flubromazolam, meclonazepam, nifoxipam, norflurazepam, pyrazolam metizolam, estazolam, bromazepam, deschloroetiozolam, ketazolam, phenazepam, and 3-hydroxyphenazepam. Many of these compounds are used illicitly in association with other substances.

Designer benzodiazepines all produce CNS depression that responds to the benzodiazepine antagonist flumazenil. Between 2014 and 2017, exposures to these substances, especially etizolam and clonazolam, increased by 330% in 40 U.S. states that were studied. Surveillance reports indicated that 36% of cases required hospitalization (22% to the ICU), and 3% needed intubation; one person died. The capacity of low concentrations of these drugs to produce pharmacologic effects allows them to escape detection by routine toxicology.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Placebo-controlled trials have clearly established the efficacy of benzodiazepines in the treatment of generalized anxiety disorder. Multicenter, double-blind studies have also demonstrated that alprazolam is effective for panic disorder. Smaller studies indicate that, in equivalent doses, other benzodiazepines can also reduce the frequency of panic attacks, although such doses are much higher than ordinary doses for generalized anxiety. For example, 60 mg of diazepam would be equivalent to the reported antipanic dose of alprazolam of around 6 mg per day in a pivotal multicenter trial. Although patients report substantially prolonged sleep with benzodiazepine hypnotics, and these medications shorten sleep onset in the sleep laboratory compared to placebo, total sleep time is increased by an average of only 30 minutes or so with these drugs. No recent therapeutic studies have been conducted in the absence of any potential return on investment for the industry to study medications that are all off-patent and prospective studies of preferential response to one or another benzodiazepine despite substantial differences in their pharmacology. In general, clinical experience suggests that factors listed in Table 33.9b–2 predict a good response to benzodiazepines in the treatment of acute anxiety.

Table 33.9b–2.
Predictors of a Good Response to a Benzodiazepine

Acute anxiety
Clear precipitant
Previous good response to a benzodiazepine
Patient awareness that symptoms are psychological
Patient desire for a benzodiazepine
Absence of history of substance misuse

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

The primary action of benzodiazepines is on the CNS, where they have anticonvulsant, anxiolytic, muscle relaxant, and CNS depressant effects. Hyperpolarization of brainstem neurons at high benzodiazepine doses can cause respiratory depression. Diazepam increases coronary blood flow and has negative inotropic effects. Benzodiazepines used as preanesthetics reduce peripheral vascular resistance, thereby decreasing blood pressure and increasing heart rate. As discussed below, benzodiazepines cross the placenta and enter breast milk.

Benzodiazepines such as diazepam bind to PBRs. The therapeutic potential of these interactions has not been extensively explored, but it is possible that stabilizing the mitochondrial ion channel could reduce neurotoxicity in some degenerative disorders.

THERAPEUTIC INDICATIONS

The primary advantages of benzodiazepines and Z-drugs are their effectiveness, ease of use, tolerability, and relative safety in the treatment of most forms of anxiety, and their rapid onset of action in acute anxiety. They can also facilitate exposure to phobic and other anxiety-provoking situations, promoting desensitization to such stimuli. However, like any medication, they have the potential to reinforce a passive approach to symptom relief that can reinforce relying on something outside oneself for symptom relief and undermine acquiring an internal sense of mastery of stressful situations.

Anxiety Disorders

Benzodiazepines have primarily been used to treat situational anxiety and anxiety disorders. Higher doses may be necessary for panic disorder more than other anxiety disorders. In recent years, prescribing trends and treatment guidelines have shifted from the use of benzodiazepines to SSRIs and SNRIs for anxiety disorders, and most treatment guidelines favor antidepressants over benzodiazepines for anxiety disorders. However, benzodiazepines and antidepressants have been found to have similar efficacy, side effects and withdrawal syndromes. A review and meta-analysis of randomized controlled trials (RCTs) of treatment for anxiety disorders, and maintenance studies lasting at least 13 weeks after RCTs, found no significant differences between antidepressants and benzodiazepines in any outcome after the initial 8 weeks of treatment, except that benzodiazepines had lower discontinuation rates than placebo. In the limited number of continuation studies of patients who responded to an initial 8-week RCT, continued treatment with benzodiazepines and antidepressants was equivalent in efficacy and safety. Another systematic review and meta-analysis examined comparisons between antidepressants and benzodiazepines. The authors identified 18 studies of tricyclic antidepressants (TCAs), phenelzine, and newer antidepressants. Due to the small number of studies in each category, a meta-analysis could only be performed on the 11 comparisons of TCAs to benzodiazepines in the treatment of panic disorder with or without agoraphobia. Panic disorder studies included 2,624 patients assigned to a benzodiazepine or a TCA. Heterogeneity of studies limited the finding of a slight advantage with benzodiazepines, especially for comparisons of a benzodiazepine with a newer antidepressant such as venlafaxine. However, benzodiazepines had significantly fewer adverse effects than TCAs. Overall, the authors found no support for the superiority of antidepressants for anxiety disorders.

A Cochrane review of 24 double-blind RCTs comparing benzodiazepines to placebo or other treatments in panic disorder in adults with or without agoraphobia (*N* = 4,233) concluded that most studies were of poor methodologic quality and 20 were at high risk of bias in at least one domain. The overall likelihood of response was 65% higher with benzodiazepines (RR 1.65; NNT = 4). Half as many benzodiazepine-treated patients withdrew (RR = 0.50), although dropouts because of adverse effects were higher with benzodiazepines. Benzodiazepines were superior for remission and social functioning, although the quality of evidence was low for all outcomes because of the likelihood of unblinding. A systematic review and meta-analysis of randomized controlled studies 4 to 12 weeks in duration of acute treatment of panic disorder with a benzodiazepine or an SSRI versus placebo found that benzodiazepines and SSRIs both caused somnolence; benzodiazepines were associated with more memory problems, decreased libido, constipation and dry mouth, and SSRIs caused more diaphoresis, abnormal ejaculation, fatigue, nausea, diarrhea, and insomnia. Benzodiazepines were protective against tachycardia, diaphoresis, fatigue, and insomnia. The small number of randomized trials directly comparing benzodiazepines and SSRIs made it impossible to draw definitive conclusions.

Benzodiazepines are used as adjuncts to reduce arousal in PTSD and are sometimes used intermittently to reduce social anxiety, although ongoing treatment, especially with antidepressants, is often preferred. β-Blockers can be more useful than benzodiazepines for performance anxiety because they are less likely to impair psychomotor function, and they directly reduce autonomic arousal that interferes with performance. Conversely, rapid reduction of anxiety and the arrhythmogenic effects of increased circulating levels of catecholamines on a damaged myocardium can be especially useful to patients with acute myocardial infarction in the coronary care unit.

Insomnia

As many as half of chronic benzodiazepine users list insomnia as the primary reason for taking the medication. Benzodiazepines promote sleep by a variety of mechanisms, including alteration of arousal systems in the ventrolateral preoptic area and inhibition of the orexin arousal-promoting system. Benzodiazepines increase sleep duration and perceived quality of sleep, decrease sleep latency, slightly reduce REM sleep, and suppress slow-wave sleep. Tolerance develops to the hypnotic effect after 1 to 2 months; after 24 weeks of continuous treatment, subjective sleep quality decreases below baseline, but efficacy can be retained by intermittent use. Rebound insomnia after discontinuing benzodiazepines has been noted in up to 71% of patients. Gaboxadol, an experimental neuroactive steroid that is a nonselective extrasynaptic $GABA_A$ agonist, reduces sleep latency and increases slow-wave sleep; however, it was withdrawn from further study because of a narrow therapeutic index. Benzodiazepines are particularly useful for insomnia caused by acute stress, such as hospitalization.

The benzodiazepine hypnotic flurazepam has a short elimination half-life, but its major active metabolite, desalkylflurazepam, has a half-life longer than 1 day. While high lipid solubility can lead to a rapid onset of action for people who have trouble falling asleep, the persistence of a drug effect the next day makes this medication most useful for patients with daytime anxiety. On the other hand, accumulation with repeated dosing often leads to daytime sedation. The relatively slow onset of action of temazepam, which is related to its lower lipid solubility, combined with an intermediate half-life, make it most appropriate for patients with middle or terminal but not initial insomnia. Triazolam, on the other hand, has a rapid onset of action and a short elimination half-life, making it useful for patients who have difficulty falling asleep and who want to avoid daytime impairment. However, the same features lead to a high frequency of discontinuation phenomena the next morning, especially amnesia. In addition, concerns have been raised about the efficacy of lower doses (0.125 mg) and the safety of higher doses (0.5 mg). Compared with placebo, 15 mg of temazepam decreased objective sleep latency by 37 minutes, while it increased objective total sleep time by 99 minutes; triazolam improved subjective sleep latency and total sleep time to clinically insignificant degrees.

Zolpidem accounts for approximately one-third of all hypnotic prescriptions, having eclipsed benzodiazepines because of its perceived more favorable side-effect profile, although psychomotor impairment has been noted with repeated dosing. Zolpidem improves sleep latency and total sleep time to an extent similar to benzodiazepines, without reduction of REM or delta sleep or other changes in sleep architecture. An extended-release form (Ambien CR) and a slow-release formulation (Intermezzo) were developed for the treatment of difficulty remaining asleep. A sublingual tablet (Edluar) and an oral spray (Zolpimist) are also available. As is true of most other hypnotics, it is usually recommended that zolpidem not be administered for more than 2 to 8 weeks. However, one study showed no loss of efficacy over 6 months of treatment with extended-release oral zolpidem. Zaleplon in a 10-mg dose reduces sleep latency to an extent similar to 0.25 mg of triazolam. Usual doses increase total sleep time by approximately 42 minutes. At higher doses (40 to 60 mg), slow-wave sleep increases, and REM percentage decreases.

Eszopiclone improves sleep efficiency, sleep latency, number of awakenings, total sleep time, and quality and depth of sleep in adults and elderly patients, in whom it also decreases the number of daytime naps. Because the manufacturer chose to conduct 6-month studies, eszopiclone is approved for long-term treatment of insomnia. Sales increased as a result of extensive advertising of this indication. From a pharmacologic standpoint, there is no reason to think that nonselective benzodiazepines that have been marketed for anxiety, such as lorazepam and oxazepam, would not be as effective as hypnotics as similarly acting benzodiazepines and Z-drugs such as temazepam and zolpidem.

A randomized, controlled trial assigned 63 patients with sleep-onset insomnia (mean age 49) to cognitive-behavioral therapy (CBT), zolpidem, the combination, or placebo. Subjects were blinded to the medication condition but not psychotherapy. CBT was more effective than other interventions in improving sleep-onset latency ($P = .003$), with effect sizes at the conclusion of treatment compared with placebo of 1.22 for CBT, 1.12 for combination treatment, and –0.1 for pharmacotherapy. CBT produced significantly greater increases in sleep efficiency than pharmacotherapy and produced the greatest number of normal sleepers (≥85% sleep efficiency), although the difference was not statistically significant. All treatments produced similar increases in total sleep time. The combination of CBT and zolpidem was no more effective than CBT alone.

Gamma-hydroxybutyrate (GHB), a $GABA_A$ receptor co-agonist that binds to GHB receptors on the benzodiazepine–GABA receptor complex, is approved for the treatment of narcolepsy. Prior to its approval in the United States, it was in increasing use as a “club drug” and for other recreational and illicit purposes. GHB has the capacity to induce absence seizures, dependence, abuse, severe withdrawal syndromes, dangerous intoxication, and amnesia as a result of complex actions on tegmental dopaminergic systems; the latter effects have led to its use as a date rape drug. GHB, therefore, is not recommended as a treatment for insomnia or any other purpose.

The “Z-drugs” and some short-acting benzodiazepines have been used to treat jet lag and to facilitate adjustment to work shift changes. However, while all hypnotics induce sleep, they have no effect on the sleep–wake cycle, which is determined more by the time of waking up, especially with exposure to bright light, than the time of going to sleep. Therefore, the best approach to preventing jet lag and sleep disruption associated with changes in work shift is to take melatonin or the melatonin agonist ramelteon at the desired time of going to sleep and have 30 minutes of exposure to daylight or artificial bright light at an intensity of 10,000 lux at the desired time of waking up.

Sedating antidepressants, especially trazodone, and sedating atypical antipsychotic drugs such as quetiapine and olanzapine are currently used frequently to treat chronic insomnia. Additional alternatives to benzodiazepines for insomnia include orexin antagonists (e.g., suvorexant, lemborexant), melatonin agonists (e.g., ramelteon), and other sedating antidepressants (e.g., low dose doxepin). A summary of 64 systematic reviews (35 with meta-analyses) found consistent evidence in more than one high or medium-quality study of effectiveness in multiple outcome domains for zolpidem, suvorexant, doxepin, melatonin, and CBT, and evidence of effectiveness across multiple outcomes based on one high-quality review with meta-analysis each for temazepam, triazolam, zopiclone, trazodone, and behavioral interventions; most studies were short-term.

In developing guidelines for the treatment of insomnia, the American Academy of Sleep Medicine regarded all of their recommendations as “weak,” which they defined as reflecting “a lower degree of certainty in the outcome and appropriateness of the patient-care strategy for all patients, but [also]… not… an indication of ineffectiveness.” These recommendations included zaleplon, triazolam, or ramelteon for initial insomnia, eszopiclone, zolpidem, or temazepam for initial and sleep maintenance insomnia and ramelteon or suvorexant for sleep maintenance insomnia. The Academy recommended against using trazodone, tiagabine, diphenhydramine, melatonin, tryptophan, or valerian for sleep onset or sleep maintenance insomnia. The Academy did not recommend quazepam, estazolam, or flurazepam because of a lack of sufficient data. CBT is considered by most sleep specialists to be the best first-line treatment for chronic insomnia.

Given these considerations, benzodiazepine and Z-drug hypnotics should probably be considered first for acute insomnia associated with hospitalization, acute illness, or other acute stress. The choice of agent would depend on whether medications with longer durations of action would be desirable to ameliorate acute daytime anxiety or whether daytime or psychomotor impairment would be problematic. CBT should be considered for more persistent insomnia. The choice of medication would depend on associated disorders. For example, insomnia associated with depression might be an indication for a sedating antidepressant, whereas insomnia associated with a chronic anxiety disorder could be treated with an antidepressant, benzodiazepine receptor agonist, or alternative agent recommended by the American Academy of Sleep Medicine.

Depressive Disorders

A systematic review and, where possible, meta-analysis of 38 RCTs of benzodiazepines in depression addressed the common assumption that benzodiazepines are not appropriate treatments for depression. The review included studies of a benzodiazepine versus a TCA and placebo, and studies of a benzodiazepine versus placebo, for depression with or without anxiety. All studies were conducted prior to the 1990s, and only one involved a newer antidepressant (fluvoxamine). The meta-analysis involved 22 RCTs of a benzodiazepine versus placebo and 20 separate analyses of a benzodiazepine versus a TCA. In about half the studies comparing a benzodiazepine to a TCA and/or placebo, benzodiazepines were significantly more effective than placebo and as effective as a TCA. In the review of 11 studies, TCAs were more effective than benzodiazepines, while benzodiazepines were more effective than TCAs in only one study. In 12 studies, benzodiazepines had a faster onset of action than TCAs. There were no significant differences in response to benzodiazepines between anxious depression and nonanxious depression. In the meta-analysis, there were no significant differences in response rates between benzodiazepines and TCAs or between benzodiazepines and placebo.

A Cochrane review evaluated 10 RCTs (*N* = 731) conducted through May 2019 of an antidepressant plus a benzodiazepine versus antidepressant monotherapy for patients with major depression. During the first 4 weeks but not after 5 weeks, the combination was more effective for reduction of depressive symptoms and for response and remission of major depression, but not for the reduction of anxiety. Dropouts for adverse effects were about half as frequent in the combination group. An advantage of early co-administration of a benzodiazepine with an antidepressant may be that it reduces the initial jitteriness that can occur with the initiation of antidepressant therapy, even in depressed patients without comorbid anxiety, reducing the risk of patients discontinuing the antidepressant before the therapeutic effect begins. Such benefits must be weighed against potential risks of the benzodiazepine, reports of which could be confounded by the more frequent use of benzodiazepine-antidepressant combinations in patients with more severe, complex, comorbid, and treatment-refractory depression that does not respond well to an antidepressant alone and is associated with more adverse events of treatment.

While the benefit of most benzodiazepines in the treatment of depression probably involves reduction of early antidepressant-induced jitteriness associated with initial increased noradrenergic neurotransmission and/or reduction of anxiety or insomnia, alprazolam has antidepressant properties, although it is not as effective as antidepressants as monotherapy for more severe depression. Alprazolam might be considered for patients with mild–moderate depression who cannot tolerate or do not respond to any antidepressant or psychotherapy. The long-term goal would be to transition to an antidepressant.

Bipolar Disorder

For many years, benzodiazepines, especially clonazepam and lorazepam, were frequently used to treat mania, alone or in combination with an antipsychotic drug or a mood stabilizer. In recent years, the use of benzodiazepines has declined, concomitant with the increased use of atypical antipsychotic drugs. However, because they can ameliorate agitation and promote sleep rapidly, with fewer adverse effects than antipsychotic drugs, benzodiazepines are still appropriate adjuncts in the treatment of mania. In longer-term treatment, nonspecific anxiety occurring in the context of an affective episode often remits with remission of the episode; for persistent anxiety or insomnia and for comorbid anxiety disorders, valproate or a benzodiazepine would be appropriate choices. One exception would be alprazolam, which has been reported to destabilize bipolar disorder, probably because of its antidepressant properties.

A secondary analysis of 482 patients with bipolar I or bipolar II disorder enrolled in a 6-month controlled comparison of lithium- versus quetiapine-based treatment suggested that the addition of a benzodiazepine did not improve clinical outcome over 6 months. As with other disorders, the need for continued use of benzodiazepines (and other forms of adjunctive treatment) may be a marker of more severe and refractory bipolar illness requiring more definitive treatment rather than promiscuous use of anxiolytics.

Schizophrenia

A Cochrane review of 34 short-term studies involving 2,657 schizophrenia patients, 8 of the studies comparing a benzodiazepine to placebo and 14 comparing benzodiazepine to antipsychotic monotherapy, found small sample sizes and incomplete data reporting. The only evidence that emerged for benzodiazepines alone or added to antipsychotic drugs was for short-term sedation (generally 30 to 60 minutes) of agitation in schizophrenia patients. A follow-up Cochrane report supported the impression that although they are well-tolerated, benzodiazepines do not augment antipsychotic drugs or improve overall outcomes in the treatment of schizophrenia, their only appropriate use being for very short-term treatment of agitation. In addition, although the quality of studies is limited, benzodiazepines reduce acute akathisia caused by antipsychotic drugs.

A meta-analysis of studies reported through 2017 found no difference in the outcome of schizophrenia between the benzodiazepine-1 agonist alpidem, eszopiclone, and placebo. However, one individual study reported superiority of alpidem to placebo in overall improvement of schizophrenia, and one study found that short-term treatment with eszopiclone was effective for insomnia in schizophrenia patients.

Catatonia

In case series, the remission rate of catatonia has been reported to range from 60% to 80% with benzodiazepines, the most commonly reported of which has been lorazepam. A common protocol begins with a lorazepam challenge test of 1 to 2 mg in adults and 0.5 to 1 mg in children and geriatric patients, administered orally (including via nasogastric tube) or parenterally. Following a response (≥50% improvement), the dose is increased to 6 mg/day in divided dose. The dose is further increased to 6 to 16 mg/day, and sometimes as much as 30 mg/day. Oral is less effective than sublingual or IM administration. At five times the lorazepam dose, diazepam has also been found to be effective. A zolpidem challenge test of 10 mg orally or via nasogastric tube has also been utilized. If the patient responds, lorazepam is substituted, although zolpidem up to 40 mg/day has been used for ongoing treatment.

Another benzodiazepine protocol begins with a 2-mg intravenous dose of lorazepam, repeated 3 to 5 times per day. The dose is increased to 10 to 12 mg/day if the initial doses are partially effective. A combined lorazepam/diazepam approach begins with 2 mg of IM lorazepam. If there is no effect within 2 hours, a second 2-mg dose is administered, followed by an intravenous infusion of 10-mg diazepam in 500 mL of normal saline at 1.25 mg/hr until catatonia remits.

Several studies suggest that a more robust response of catatonia to lorazepam, at least in lower doses, is more likely with a shorter duration of catatonia and waxy flexibility. Factors predicting a poorer response include passivity, mutism, auditory hallucinations describing the patient in the third person and marked retardation and mutism complicating schizophrenia, especially with chronic negative symptoms. Maintenance lorazepam can be effective in reducing relapse and recurrence. There are no controlled studies of maintenance treatment with benzodiazepines, but clinical reports suggest that lorazepam doses in the range of 4 to 10 mg/day are effective.

Gastrointestinal Disorders

Patients with unexplained gastrointestinal (GI) symptoms have an increased rate of panic disorder, agoraphobia, and major depression, while many GI disorders are aggravated by anxiety. A number of studies of benzodiazepines in GI disorders published in the 1970s and 1980s reported improvement of symptoms, as well as physical signs such as peptic ulcer. Benzodiazepines are better tolerated than antidepressants by some patients with GI disorders in that they do not cause constipation (like TCAs) or diarrhea (like SSRIs), and they can decrease gastric secretion and induce smooth muscle relaxation.

Cardiovascular Disorders

There is an increased prevalence of anxiety disorders and depression in patients with cardiovascular disease. Reduction by benzodiazepines of autonomic hyperactivity associated with anxiety and depression can attenuate myocardial ischemia and hypertension in these individuals. As noted earlier, reduction of arousal by benzodiazepines can be particularly helpful in the coronary care unit in decreasing high levels of circulating benzodiazepines that are arrhythmogenic in the presence of a damaged cardiac conduction system.

Treatment of Alcohol Withdrawal

Uncomplicated alcohol withdrawal symptoms begin within 4 to 12 hours after stopping or significantly reducing alcohol intake. Symptoms normally reach peak intensity on the second day and improve by the 4th or 5th day. Delirium tremens (DTs) usually has an onset 3 to 5 days after alcohol discontinuation and persists for about a week. The mortality rate in DTs remains about 8%.

The standard treatment of alcohol withdrawal syndromes is with benzodiazepines, most frequently lorazepam, diazepam, chlordiazepoxide, or clorazepate. Some protocols use standardized protocols with a progressively decreasing dose, while others adjust each dose based on the response to the previous dose. Alcohol withdrawal syndromes refractory to benzodiazepines are often treated with the addition of phenobarbital, propofol, or dexmedetomidine. The latter medication is an α_2-agonist that is more potent than clonidine, with an onset of action of 15 minutes and an elimination half-life of 2 hours, the antiadrenergic effect of which decreases sympathetic tone without causing respiratory depression.

A retrospective chart review comparing phenobarbital (N = 143) and benzodiazepines (N = 419) in treating alcohol withdrawal found equal outcomes in both, although 16 patients who did not respond to benzodiazepines did well when switched to phenobarbital. Amplifying this finding, a Cochrane review of 64 RCTs involving 4,309 subjects comparing benzodiazepines to placebo, other treatments, and each other in the treatment of alcohol withdrawal found benzodiazepines to be significantly superior to placebo for seizures and suggested superiority of benzodiazepines for many other outcomes, but the studies were too heterogeneous to draw definitive conclusions about comparative effectiveness and safety.

Agitation

Since the antimanic and antipsychotic effects of antipsychotic drugs are delayed, treatment of agitation associated with mania and schizophrenia with these medications relies on their sedative side effects. Adding or substituting a benzodiazepine has the advantage of rapid sedation without the risks of high or repeated doses of antipsychotics. In the treatment of mania, improvement of sleep with a benzodiazepine may contribute to a reduction of manic symptoms. The addition of a benzodiazepine to an antipsychotic drug may also ameliorate akathisia, which can be mistaken for agitation. IM lorazepam has been used most frequently for agitation in acutely ill psychiatric patients, but midazolam is also available in an IM formulation. High lipid solubility and potency and a very short elimination half-life make midazolam an appealing choice when a rapid onset of action without drug accumulation with repeated dosing is desired. Recent experience suggests that midazolam is rapidly effective and well tolerated in the prehospital treatment of agitation.

In one study, oral risperidone plus lorazepam was effective in the emergency treatment of acutely psychotic patients. Similarly, a multicenter controlled trial in acute psychosis reported that risperidone plus lorazepam more rapidly reduced aggression, produced more sedation, and was better tolerated than risperidone alone. Anticipating that an antipsychotic drug will be necessary anyway, most clinicians add a benzodiazepine as an adjunct. However, particularly in the treatment of mania, it may be reasonable to treat agitation and insomnia with a benzodiazepine with or without an antipsychotic drug and gradually transition to lithium or another mood stabilizing medication that does not have prominent sedative properties.

The treatment of agitation in unselected acutely ill patients in the hospital or emergency department who may be experiencing an acute psychiatric disorder, delirium, or a substance-induced condition frequently involves an antipsychotic drug alone or in combination with a benzodiazepine, especially lorazepam. A commonly used IM combination consists of 5-mg haloperidol and 2-mg lorazepam. This tradition is based on a randomized uncontrolled trial that found that the combination of 5 mg of haloperidol and 2 mg of lorazepam produced more rapid sedation than either medication alone, although hospitalization and the need for additional medication doses did not differ between the three conditions. Following a systematic review and consensus discussion, the American College of Emergency Physicians found no class I RCTs, prospective cohort studies or meta-analyses of RCTs supporting this or any other pharmacologic treatment for agitation. The consensus statement noted that multiple class II observational reports, case-controlled studies, and meta-analyses found equivalence of 2- to 4-mg lorazepam to 5-mg haloperidol, equivalence of clonazepam and flunitrazepam to haloperidol, and a shorter time to sedation with 5 mg of IM midazolam than with lorazepam or haloperidol; 15 mg of IM midazolam was found to be more effective than haloperidol plus promethazine for rapid sedation and sleep. Class III cross-sectional and case studies and consensus panels determined 2.5 to 3 mg of midazolam to be rapidly effective for agitation in acutely ill psychiatric patients, while there was no clear benefit of adding haloperidol to a benzodiazepine. The group had no level A recommendations. Their level B recommendation (moderate clinical certainty based on observational studies, decision analysis or strong consensus) suggested the use of monotherapy with lorazepam, midazolam, droperidol, or haloperidol, followed by benzodiazepine-antipsychotic drug combinations if the first approach was ineffective.

Delirium

Benzodiazepines are often used for agitation secondary to delirium; however, CNS depression from the benzodiazepine can aggravate delirium. A Cochrane attempt to review studies of the effectiveness and tolerability of benzodiazepines for non–alcohol-related delirium found only one study meeting the selection criteria. A study comparing lorazepam with the α_2-agonist dexmedetomidine for delirium in mechanically ventilated ICU patients found that dexmedetomidine was associated with 7 delirium- and coma-free days compared with 3 days for lorazepam ($P = .01$). One partially controlled study showed no advantage of alprazolam over neuroleptics for agitation associated with delirium, and another partially controlled study found that lorazepam was not as effective as haloperidol or chlorpromazine for acute confusion and had more adverse effects. The authors concluded that there are insufficient blinded, controlled studies of benzodiazepines for delirium and no support at this point for the use of benzodiazepines for delirium. In a later Cochrane review that addressed RCTs of benzodiazepines to treat delirium in adults not associated with alcohol or benzodiazepine withdrawal in settings other than an ICU, including hospitals, nursing homes, and long-term care settings, only two trials qualified for review. The Cochrane authors' review of a study of 58 delirious advanced cancer patients in an acute palliative care unit treated with haloperidol and then randomized to the addition of lorazepam or placebo for delirium of similar severity and duration, and a study of 30 AIDS patients (mean age 39) randomized to lorazepam, haloperidol, or chlorpromazine, revealed no superiority of one treatment over another. The authors concluded that there is still insufficient evidence to determine whether benzodiazepines are useful in the treatment of delirium.

Substance Use Disorders

In animal studies, pretreatment with experimental positive allosteric modulators of the GABAB receptor blocked neuroplasticity in the ventral tegmental area induced by alcohol and cocaine and attenuated self-administration of these substances. A review of human studies of alcohol dependence found that in lower doses, baclofen, an orthosteric $GABA_B$ agonist, was significantly superior to placebo in prolonging the time to relapse of drinking. However, the tolerability of baclofen was low.

Use With Electroconvulsive Therapy

The anticonvulsant action of benzodiazepines leads most clinicians to avoid benzodiazepines during ECT for several reasons. ECT increases diazepam binding in preclinical studies and shortens seizure duration in clinical series. Despite not taking benzodiazepines, pretreatment with flumazenil improved seizure duration and clinical response after declining benefit in a small study, possibly implicating a shift of the benzodiazepine receptor toward inverse agonism as a mechanism of ECT. For several reasons, the assumption that the anticonvulsant action of benzodiazepines increases seizure threshold has not always been supported by clinical experience.

Many studies suggesting that benzodiazepines impair the response to ECT have not used stimulus titration according to seizure threshold. In contrast, a retrospective chart review of 70 major depressive disorder patients (22 taking benzodiazepines) undergoing the dose titration method for ECT found significantly greater remission rates and decreases in MADRS scores in patients taking benzodiazepines; there were no differences between benzodiazepine and nonbenzodiazepine patients in seizure duration or seizure threshold. Changes in anxiety and cognitive function were not assessed. A retrospective review of records of 90 patients with major depressive disorder treated with bitemporal ECT found no impact of benzodiazepines on the efficacy of ECT. This finding was limited by the retrospective method, lack of information about seizure threshold, lack of randomization, absence of control for the number of ECTs, use of Hamilton Depression Rating Scale (HDRS) scores completed openly by treating clinicians, and use of a single approach to ECT.

If benzodiazepines increase the seizure threshold, interference with the therapeutic action of ECT should be more likely with unilateral electrode placement, for which stimulus intensity must exceed the seizure threshold to a greater degree than bilateral placement. Indeed, a review of the literature concluded that benzodiazepines may impair the response to unilateral more than bilateral ECT. This possibility was supported by a French retrospective study of 70 patients receiving biweekly ECT for major depressive disorder that compared outcomes in 22 patients (mean age 61) taking benzodiazepines (mean diazepam equivalent dose 17 mg) with outcomes for 48 patients (mean age 62) not taking benzodiazepines. All benzodiazepine patients and 88% of nonbenzodiazepine patients received bitemporal electrode placement. Stimulus intensity was titrated to twice the seizure threshold in bilateral and six times the seizure threshold in unilateral ECT. There were no significant differences between groups in the number of ECTs (mean 13 to 14), seizure threshold, or seizure duration. However, the benzodiazepine group had a significantly greater reduction in Montgomery-Asberg Depression Rating Scale (MADRS) scores (–28.9 vs. –21.9; $P < .02$) and significantly more remissions (MADRS score <10), but not responses (decrease in MADRS score ≥50%), than the nonbenzodiazepine group. The authors concluded that with appropriate stimulus titration and use of bilateral ECT, benzodiazepines may not impair and may even enhance the response to ECT, perhaps by treatment of comorbid anxiety or additive increases in seizure threshold to the effect of artificial seizures. A double-blind comparison of midazolam versus methohexital anesthesia for ECT found no difference in outcome as measured by MADRS scores. Such experiences suggest that if benzodiazepines are to be continued during ECT, they may be less problematic with bilateral than with unilateral ECT and with higher than with lower stimulus intensities. Lower benzodiazepine doses may be less likely than higher doses to impair the response to ECT.

USE IN PREGNANCY

Since 1.9% to 3.1% of pregnant women take a benzodiazepine, the question of benzodiazepine use during pregnancy is not trivial. Data on the topic are contradictory, and reports of associations of adverse outcomes and use of medications such as benzodiazepines during pregnancy often do not account for confounding variables such as the effects of increased hypothalamic–pituitary–adrenal activity in stress-related disorders for which benzodiazepines are prescribed, underreported psychiatric disorders, especially substance use disorders, use of substances including nicotine and alcohol during pregnancy not meeting criteria for a formal disorder, past history of problem pregnancies (which are more common in women taking benzodiazepines), and other lifestyle and environmental factors.

A risk of major congenital malformations with exposure during pregnancy to clobazam, a benzodiazepine used to treat epilepsy and anxiety in some countries other than the United States, was suggested but not proven in a network meta-analysis. A Norwegian questionnaire cohort study analyzed data from 82,030 singleton pregnancies among 69,434 women (mean age 30). Exposure to benzodiazepines or Z-drugs was reported as 0.8%. Lower birth weight, lower gestational age at birth, and higher risk of preterm birth were associated with the use of benzodiazepines or Z-drugs. The effect of benzodiazepine use on birth weight was probably related to earlier birth, not intrauterine growth retardation, and the magnitude of the association with gestational age was not considered clinically significant. In a meta-analysis of eight studies, benzodiazepine exposure in utero was not associated with an increased risk of congenital malformations, including in the five studies that specifically examined first-trimester exposure. There was no increased risk of cardiovascular malformations in four studies that examined this specific malformation. However, concurrent use of an antidepressant and a benzodiazepine in the first trimester was associated with an increased risk of major malformations (O.R. 1.40, $P = .008$). Specific malformations were not reported.

A systematic review and meta-analysis of 14 studies found significant associations of benzodiazepine use during pregnancy with spontaneous abortion (O.R. 1.86), preterm birth (O.R. 1.96), low birth weight (O.R. 2.24), low Apgar score (O.R. 2.19), neonatal intensive care unit (NICU) admission (O.R. 2.61), and induced abortion (O.R. 2.04). After adjusting for publication bias, early gestational age and small for gestational age continued to have a statistically significant association with antenatal benzodiazepine exposure. The authors concluded that the association with NICU admission was clinically relevant as well as statistically significant. It is unclear whether NICU admissions were driven by neonatal respiratory suppression or withdrawal syndromes.

The question of behavioral teratogenicity of benzodiazepines has also been raised. A Norwegian study of 36,401 children found no significant association between exposure to benzodiazepines and/or Z-drugs during pregnancy in the 273 children with this exposure and internalizing or externalizing behaviors measured on the Child Behavior Checklist at age 5. Another Norwegian study of 4,195 women with depression/anxiety, 5,260 with sleep disorders, and 26,631 with pain-related disorders included 36,086 children, 0.8% of whom (283) were exposed to benzodiazepines or Z-drugs in mid (weeks 17 to 28) or late (week 29 or later) pregnancy. At a mean age of 5.1 years, benzodiazepine/Z-drug–exposed children did not have an increased risk of ADHD symptoms or fine motor deficits with exposure at any time point during pregnancy. Children of women with a depressive or anxiety disorder who took benzodiazepines in late pregnancy had "moderately" more gross motor and communication deficits than unexposed children, suggesting confounding by indication or higher doses in women with anxiety/depression who had late pregnancy benzodiazepine/Z-drug exposure. The authors concluded that there was "no substantial detrimental risk on child fine motor and ADHD symptoms after prenatal benzodiazepine/z-hypnotic exposure alone or in combination with opioids or antidepressants."

Since benzodiazepines are excreted in breast milk, concerns arise about medication effects on the nursing infant. A Japanese study of 11 nursing mothers and their 1-month-old infants taking benzodiazepines including alprazolam, brotizolam, clonazepam, flunitrazepam, and lorazepam, among others, found that the ratio of milk to plasma benzodiazepine level was consistently less than one, and the relative infant dose was always less than 10% except for ethyl loflazepate, which was only slightly greater than 10%. No abnormalities were noted in the infants. The authors concluded that medication exposure is low in infants of mothers taking benzodiazepines. Other experts recommend that if a benzodiazepine is used during breastfeeding, shorter-acting agents such as lorazepam and oxazepam are preferred. Medication exposure can be further reduced by taking the medication after rather than before breastfeeding. The nursing infant should be monitored for sedation, lethargy, and poor sucking. Withdrawal syndromes can occur in the nursing infant if the mother abruptly discontinues the benzodiazepine.

BENZODIAZEPINES IN THE ELDERLY

Elderly patients are considered more vulnerable to benzodiazepine effects as a result of increased adipose/lean ratio, resulting in a greater volume of distribution, a longer elimination half-life, slowed metabolism, and decreased clearance, with increased steady-state concentrations. In addition, since most benzodiazepines are protein-bound, decreased plasma proteins with age can increase the free fraction with no change in the overall concentration but increased levels of active drug. There is also a greater pharmacodynamic effect secondary to changes in receptor function and signaling. These factors can result in greater drug effect and accumulation if the dose is not adjusted and monitored. Data are mixed, particularly for shorter-acting benzodiazepines, supporting the recommendation by the American Geriatrics Society that benzodiazepines and related drugs be avoided in geriatric patients, especially those with cognitive impairment, because such changes make the elderly more vulnerable to falls, automobile crashes, and impaired cognition, memory and psychomotor function.

PRECAUTIONS AND ADVERSE EFFECTS

Although they are effective for most forms of anxiety and generally well-tolerated, benzodiazepines, like other medications, have the potential to reinforce reliance on something outside oneself for relief of anxiety, potentially undermining the development of confidence in the ability to master stressful situations. As noted earlier, rapid reduction of anxiety can enhance confidence in emergency settings such as the CCU, and it can facilitate exposure to phobic and other anxiety-provoking situations, but therapies such as CBT, mindfulness, and desensitization may be more effective at enhancing a sense of personal agency in patients with chronic anxiety.

The most common physical adverse effects of benzodiazepines include sedation, slowed reaction time, and impaired memory and psychomotor function, although these are not as severe as with barbiturates. Especially in older people, benzodiazepines can cause falls, syncope and fractures due to slowed reaction time, impaired balance and gait, and impaired vision. Falls are more frequent with chronic benzodiazepine use. The propylene glycol vehicle in parenteral lorazepam preparations has been linked to hyperosmolar metabolic acidosis.

Cognitive Side Effects

Withdrawal from short-acting, high potency benzodiazepines such as midazolam and triazolam can cause retrograde amnesia after discontinuation or between doses. Withdrawal probably also accounts for amnesia following a single dose of the high potency, highly lipid-soluble “date rape drug” flunitrazepam, especially when it is combined with alcohol. All benzodiazepines can cause anterograde amnesia and can interfere with new learning, although they usually do not impair access to memories that are already acquired. Impaired verbal memory, visuospatial memory, working memory, attention and executive function independent of the degree of underlying anxiety and depression have been demonstrated in people chronically using high doses of benzodiazepines. In a study of older patients with normal cognition, those taking benzodiazepines had poorer cognitive performance, but there was no difference in the rate of cognitive decline between people who were or were not taking benzodiazepines, suggesting the possibility of subclinical or prodromal features of dementia in patients taking benzodiazepines.

In a multicenter prospective depression trial, 353 patients treated with an SSRI alone were compared to 49 patients treated with an SSRI plus a benzodiazepine at baseline, 8 weeks, and 48 weeks of treatment on measures of depression, anxiety and cognitive function. Cognitive function was similar for both groups at baseline. Controlling for differences in severity of depression and anxiety, scores on the Digit Symbol Coding Test (a measure of information processing speed) were slightly better in patients treated with an antidepressant and a benzodiazepine than an antidepressant alone at 8 weeks and distinctly and significantly better at 1-year follow-up. The results suggest that the short-term addition of a benzodiazepine does not impair cognitive function in patients with MDD, while it could improve processing speed after short-term treatment and 1 year later.

Delirium

Delirium is associated with globally impaired cortical functioning, altered network connectivity, and/or suppression of reticular activating system activity, any of which can be impaired by central nervous system depressants. Benzodiazepines, therefore, have the potential to cause or aggravate delirium, especially in patients who were cognitively impaired, to begin with. In one inpatient study, there was a significant association between ongoing benzodiazepine use and delirium (adjusted IRR 2.31 to 2.44). A study of all 2,446 patients admitted to a 32-bed pediatric intensive care unit over a 2 ½-year period found a dose-related contribution of exposure to benzodiazepines to the risk of delirium independent of other clinical factors. A systematic review of the association between benzodiazepine use in the ICU and delirium, depression, anxiety, PTSD and cognitive impairment included 49 studies of adult ICU patients (21 of them high quality) conducted over 10 years (total N 11,138). The data could not be pooled for meta-analysis because of the heterogeneity of samples and settings. Of these studies, 24 identified benzodiazepine use in the ICU as a risk factor for delirium; of this group of studies, 10 of 13 of high methodologic quality found benzodiazepine use to be a risk factor for delirium. Another 10 studies reported no evidence of an association between benzodiazepine use and delirium, and one study suggested a protective effect of midazolam. The risk of delirium was greater with high daily doses, the use of midazolam or lorazepam, and continuous administration. The possibility of confounding by indication (i.e., patients receiving benzodiazepines may have had illnesses that were more likely to predispose them to delirium) could not be excluded.

Benzodiazepines and Dementia

Debate continues about the potential of benzodiazepines to increase the risk of dementia. Proposed mechanisms of such an association include decreased cognitive reserve, downregulation of $GABA_A$ receptors resulting in reduced resilience of information processing systems, and decreased ability to utilize alternative neural networks because of reduced brain activation. However, data on whether benzodiazepines increase the risk of dementia have been conflicting. Reports of an increased risk of dementia with chronic benzodiazepine use may reflect reverse causality—in other words, benzodiazepines are more frequently prescribed for prodromal symptoms of dementia such as depression, which is often treated with benzodiazepines.

A number of studies have emerged to address this question. For example, using claims forms from the public drug plan, a Canadian case-control study correlated benzodiazepine prescriptions 6 years prior to diagnosis in 1,796 cases of dementia and 7,184 controls from 38,741 randomly selected patients over age 60 with a diagnosis of dementia or presumed dementia based on the prescription of a cholinesterase inhibitor or memantine, and 86,259 controls matched for age, sex, and duration of follow-up. Controlling for multiple comorbidities, there was no difference between Alzheimer disease patients and controls in prescriptions for ≤180 prescribed daily doses (6 months duration of use). Use of >180 prescribed daily benzodiazepine doses (>6 months duration) was noted in 32.9% of subjects with Alzheimer disease versus 21.8% of controls (adjusted odds ratio 1.51). The risk of Alzheimer disease increased with the increasing duration of benzodiazepine prescription and was higher with long-acting benzodiazepines. There was no significant effect on anxiety, insomnia, or depression. The data in this study were based on diagnoses and prescriptions in insurance claims submitted by a large number of clinicians and might not reflect consensus diagnoses or clear rationales for the prescription of medications usually associated with Alzheimer disease. Although the authors thought that benzodiazepines prescribed 6 years prior to Alzheimer disease diagnosis would not be likely to be prescribed for prodromes of dementia, the pathology of Alzheimer disease may have developed before patients took benzodiazepines, with apparent anxiety or insomnia either being early features of dementia or incidental symptoms. The possibility that a neurodegenerative process began prior to benzodiazepine prescription is supported by a higher incidence of stroke and hypercholesterolemia in dementia patients than in the controls. There was also no information about whether patients were still taking benzodiazepines at the time of diagnosis, an important issue since cognitive side effects of benzodiazepines could be confused with primary dementia. A comparison of 153 HIV patients taking benzodiazepines for an average of 2 years with 153 HIV patients matched for demographics and comorbidities and adjusted for premorbid functioning, current affective symptoms, and lowest CD4 count found more impairment of global cognitive function, processing speed, executive function, working memory and delayed recall in the benzodiazepine group (OR = 2.13), especially with diazepam, lorazepam, and alprazolam. This finding may have reflected exacerbation of HIV-induced cognitive impairment by the benzodiazepine or an interaction with antiretroviral drugs.

Lack of an impact of benzodiazepine use on the risk of dementia was found in an examination of the relationship between ongoing use of benzodiazepines and related drugs and cognitive function at 18-month follow-up in a European randomized multicenter study of the antihypertensive nilvadipine in 448 patients (mean age 72) with Alzheimer disease (Mini-Mental State Exam score 12 to 26; median duration of symptoms 3.7 years). In the 14% of patients who took a benzodiazepine or Z-drug during the study (most frequently alprazolam, lorazepam, oxazepam, bromazepam, and zolpidem), there was no relationship between ongoing use of the medication over 18 months and changes in ADAS-Cog scores. There was also no additive effect of APOE ε4 status and benzodiazepine use on dementia risk. The authors concluded that active benzodiazepine use did not hasten cognitive decline over 18 months in patients with mild–moderate Alzheimer disease.

A retrospective cohort study of a random sample of patients over 50 years of age using a national healthcare database in South Korea compared benzodiazepine users, nonusers, and users of SSRIs or SNRIs on the grounds that they have similar indications to the benzodiazepines. Among 1,576,452 subjects, there were 1,125,346 benzodiazepine users, 17,677 antidepressant users, and 433,429 nonusers. Each benzodiazepine user was matched with a nonuser or an antidepressant user. Over a mean follow-up of 5.6 years, benzodiazepine users (especially current users) had a 23% higher risk of dementia than nonusers. However, there was no difference in dementia risk between benzodiazepine and antidepressant users. Also, in comparisons of users and nonusers, the *longer* the duration of benzodiazepine (or antidepressant) use, the *lower* the risk of dementia, suggesting that the association between benzodiazepine or antidepressant use and dementia risk was attributable to something other than these medications. Similar findings emerged from a retrospective analysis of data from an open-label randomized trial in 116 family practices in the Netherlands involving 3,526 nondemented people (mean age 74.3) living in the community. After a median of 6.7 years after starting treatment, there was no significant association between the use of short or long-acting benzodiazepines and incidence of dementia and no interaction with any APOE genotype. Persistent high anticholinergic exposure was associated with an increased risk of dementia over 6 years of follow-up, but the association was driven by the use of antidepressants or antipsychotics, suggesting confounding by indication. Similarly, in a sample of 133 patients (mean age 78) with Alzheimer disease, anticholinergic drugs (most frequently quetiapine) but not benzodiazepines were associated with significant cognitive decline.

A number of other large international studies have not found an association between the use of benzodiazepines and the development of dementing illness. For example, a Canadian 10-year multicenter study of 10,263 older people living in the community or in institutions found that benzodiazepine use was associated with an increased risk of cognitive impairment not associated with dementia (presumably a medication side effect) but not with an increased risk of Alzheimer disease or other neurologic dementia. A British prospective follow-up of 8,216 nondemented patients found that 9.3% developed dementia over the next 8 years. Medications with more anticholinergic properties were associated with a higher risk of dementia, but less anticholinergic drugs and benzodiazepines were not. A 10-year Danish cohort and nested case-control registry study of 235,465 adults with a first hospitalization for a mood disorder found no meaningful association between use of any benzodiazepine or Z-drug and subsequent risk of dementia, even considering cumulative exposure and whether exposure was long- or short-term. There was even a suggestion of a protective effect in some analyses. In a British nested case-control database study comparing 40,770 people with dementia with 283,933 controls matched on age, sex, history, and other confounders, there was no association between new prescriptions for benzodiazepines/Z-drugs and development of dementia 4 to 20 years later. A PET amyloid uptake study in 268 nondemented older adults found significantly less amyloid uptake, implying a lower risk of Alzheimer disease, in patients taking benzodiazepines, especially shorter-acting benzodiazepines. Findings suggesting no increased risk of development of dementia with benzodiazepines do not negate the cognitive side effects of these medications, especially in demented patients, and most experts recommend avoiding combining benzodiazepines and anticholinergic drugs in patients prone to cognitive impairment unless absolutely necessary.

Driving

An important psychomotor function that can be impaired by CNS depressants is driving. Driving tests and neurocognitive batteries indicate clinically relevant impairment of executive function, vigilance, and reaction time in association with impaired driving in users of benzodiazepine anxiolytics and hypnotics, to which some degree of tolerance may develop after 3 years of treatment. In a series of 410 suspected drunk or drugged drivers over age 65, alcohol was found in 81% of blood samples, benzodiazepines in 15%, and Z-drugs in 13%; the most frequently detected medications were zopiclone and diazepam. Similarly, in an Australian sample of 616 individuals aged 60 and older hospitalized as a result of a motor vehicle crash, the risk of a crash requiring hospitalization was higher for those who were prescribed benzodiazepines (O.R. 5.3), especially men (O.R. 6.2); the risk of a crash was increased somewhat less in drivers taking an antidepressant (O.R. 3.4). In 72,685 French drivers involved in traffic accidents resulting in injury, the risk of being responsible for the accident was higher in users of benzodiazepine hypnotics (O.R. 1.39), with an even greater risk (O.R. 2.46) for drivers prescribed more than one pill/day of zolpidem. Zopiclone was not associated with an increased risk of accidents in this study.

A series of meta-analyses of 21 epidemiologic studies and 69 experimental studies found that benzodiazepines were associated with a 60% to 80% increased risk of being in a traffic accident and a 40% increased risk of being responsible for an accident; the benzodiazepine-associated risk was lower in people who were older than age 65. Taking benzodiazepines during the day impaired driving independent of drug half-life. Use of diazepam, flurazepam, flunitrazepam, nitrazepam, and zopiclone as nighttime hypnotics also impaired driving, at least during the first 2 to 4 weeks of treatment. The use of alcohol plus a benzodiazepine was associated with a much greater (7.7-fold) increased risk of a traffic accident. Other research supports the observation that the combination of a benzodiazepine and alcohol impairs driving significantly more than a benzodiazepine alone. This observation is important since studies of impaired drivers find that at least 20% take another substance along with alcohol. All patients taking benzodiazepines, therefore, should be warned about the potential for impaired driving, especially in combination with alcohol. Patients may not be aware of impaired driving, tolerance to which does not occur as reliably as does tolerance to sedation.

Disinhibition

Patients who are in tenuous control of their impulses may engage in impulsive behavior or speech in response to decreased ability to suppress behavior consciously when they take benzodiazepines and Z-drugs. Some hypervigilant people become openly suspicious or agitated when their ability to keep track of details of their environment and other people is dulled by the CNS depressant effect of these medications. Additive impairment of vigilance and behavioral control of benzodiazepine receptor agonists and alcohol can lead to impulsive suicidal or aggressive behavior in vulnerable individuals.

Suicide

Benzodiazepines by themselves have a low fatality rate in overdose, but the risk of death with overdose is increased when they are combined with opioids or alcohol. Nevertheless, benzodiazepines and opioids are often prescribed together. In fact, from 1993 to 2014, the U.S. rate of combined use of benzodiazepines and opioids increased from 9.8 to 62.5/100,000, possibly explaining a concomitant increase in benzodiazepine-related deaths.

Is suicide more common in benzodiazepine users? A Swedish study compared 154 psychiatric patients who committed suicide with controls matched for age, sex, and psychiatric diagnosis. Benzodiazepine prescriptions were more common in suicides (42%) than controls (28%) (O.R. 1.83). However, this difference may have reflected a greater likelihood of use of benzodiazepines by patients at higher risk of suicide (i.e., confounding by indication). In a retrospective case-control study of 6,960 patients with anxiety disorders (2,362 of whom filled benzodiazepine prescriptions) and 6,215 patients with sleep disorders (1,237 of whom filled benzodiazepine prescriptions), patients taking benzodiazepines for anxiety who had a shorter duration of benzodiazepine use with concomitant psychotherapy or antidepressant treatment had a 39% decreased odds of suicide (O.R. 0.611; $P = .03$), while differences in suicide rates were not significant between patients who did or did not use benzodiazepines as hypnotics. Controlling for ongoing use of benzodiazepines in patients with more severe illnesses, compelling evidence of increased risk of suicide in patients taking these medications is lacking.

Tolerance

Tolerance is common to certain benzodiazepine actions. One mechanism of tolerance to benzodiazepines is that prolonged benzodiazepine receptor occupation triggers an intracellular calcium ion signaling cascade downstream from the $GABA_A$ receptor, leading to gradual downregulation of receptors via removal from the postsynaptic membrane by endocytosis and uncoupling of benzodiazepine and $GABA_A$ receptors, rendering the system less responsive to benzodiazepines. Administering flumazenil along with diazepam blocks these events, indicating that initial binding to the benzodiazepine receptors is an essential step in the development of tolerance. It remains to be seen whether inhibiting these processes with calcium antagonists or flumazenil might reduce tolerance to benzodiazepines.

Tolerance to benzodiazepine effects develops most frequently to sedation and less frequently to psychomotor impairment. Tolerance to the anxiolytic effect is less common than escalating the dose because of interdose withdrawal or subtherapeutic dosing. With adequate treatment, long-term use of benzodiazepines is usually associated with maintenance of therapeutic benefit without requiring dose escalation.

Discontinuation Syndromes

Three syndromes can follow the cessation of benzodiazepine agonists and related substances. Relapse refers to the return of the original anxiety disorder or insomnia. Relapse begins over weeks to months and follows the inherently chronic or recurrent course of most anxiety disorders. Rebound is a more intense re-emergence of anxiety or insomnia that can occur after cessation of benzodiazepines and Z-drugs, even after short-term use. Rebound symptoms, which are worse than the original symptoms, usually start within days to weeks and last up to 3 weeks. Rebound anxiety can occur 3 to 4 hours after each dose of short-acting benzodiazepines such as alprazolam and is manifested by interdose anxiety or uneasiness and clock-watching that is often mistaken for drug dependence. Similarly, rebound insomnia is more frequent and severe with short-acting, high-potency hypnotics such as triazolam and can occur on the night following the use of the hypnotic.

Syndromes that involve new features of autonomic arousal such as sympathetic signs, myoclonus or seizures are variably called withdrawal or discontinuation syndromes (Table 33.9b–3). Withdrawal begins within hours to days of discontinuation of short-acting benzodiazepines but may not begin for 1 to 2 weeks after abrupt discontinuation of longer-acting preparations; interdose withdrawal may occur with high-potency, highly lipid-soluble, short half-life benzodiazepines such as alprazolam. Withdrawal may occur after discontinuation of 4 to 8 months of treatment with typical benzodiazepine doses or after half that duration of treatment with higher doses. Minor withdrawal symptoms from benzodiazepines include sweating, tachycardia, nausea, myoclonus, restlessness, visual changes, tremor, and confusion; major new withdrawal symptoms include seizures, psychosis, and delirium. Seizures are more common in patients with predisposing factors such as abnormal EEG or brain damage and with concomitant use of medications that lower the seizure threshold. Withdrawal symptoms usually last 2 to 4 weeks and are more common and severe with high-potency, short to medium half-life benzodiazepines, but attenuated, prolonged withdrawal symptoms have been noted after discontinuation of long half-life benzodiazepines. Some reports of persistent withdrawal syndromes after discontinuation of long-term benzodiazepine use include anxiety, depression, psychosis, cognitive impairment, insomnia, sensory phenomena (e.g., tinnitus, paresthesias, pain, numbness, tingling, vibration, strange skin sensations), motor phenomena (muscle pain, cramps, weakness, tremor, shaking, myoclonus, spasms, shaking attacks), and GI symptoms (e.g., food intolerance, abdominal distension). Such features are more common with long-term treatment with short and intermediate half-life drugs and may last 6 to 12 months and sometimes years. Conversely, withdrawal from very short-acting benzodiazepines such as midazolam can be manifested by amnesia following a single dose, while withdrawal from triazolam and zolpidem can be followed by automatisms at night and amnesia the next day. Withdrawal from bedtime doses of zolpidem, zaleplon, and eszopiclone has been associated in a small number of reported cases with complex sleep behaviors resulting in injury and death from automobile crashes, hypothermia, overdoses, falls, and gunshot wounds.

Table 33.9b–3.
Signs and Symptoms of Benzodiazepine Withdrawal

Anxiety
Depersonalization
Irritability
Insomnia
Hyperacusis
Diaphoresis
Hyperreflexia
Nausea
Difficulty concentrating
Confusion
Tremor
Hyperesthesia
Myoclonus
Seizures
Delirium

A confusing clinical picture may emerge when alcohol-dependent patients who have also been taking long-acting benzodiazepines are admitted to the hospital. The onset of withdrawal seizures or alcohol withdrawal delirium may be postponed up to several weeks because it is prevented by the benzodiazepine, which has a longer duration of action at the GABA-benzodiazepine receptor complex. When the serum level of the benzodiazepine eventually declines, typical alcohol withdrawal emerges, sometimes complicated by benzodiazepine withdrawal. The late onset of symptoms may result in withdrawal delirium being mistaken for postoperative psychosis or a medical or surgical complication.

Withdrawal can probably best be understood as a consequence of benzodiazepine receptor adaptation to chronic use. Benzodiazepine receptors downregulate in response to a large number of benzodiazepine molecules. At the same time, the inverse agonist DBI is upregulated to compensate for the suppression of CNS excitability. When the benzodiazepine is abruptly withdrawn, there are more DBI molecules than there are benzodiazepine molecules competing for fewer benzodiazepine receptors. This results in a shift of the benzodiazepine receptor toward inverse agonism, causing decreased, rather than increased, GABA receptor affinity, reduced, rather than enhanced, chloride influx, and membrane depolarization. CNS excitability increases abruptly, resulting in manifestations of arousal ranging from jitteriness and tremor to myoclonus and seizures. Slower benzodiazepine withdrawal allows the benzodiazepine receptor to upregulate, as DBI is gradually downregulated, but the unpredictability of this accommodation may lead to withdrawal symptoms with a slow taper or even with a simple reduction of benzodiazepine dose. An additional contributing factor to rebound insomnia after discontinuation of long-term benzodiazepine use may be dysregulation of endogenous melatonin rhythms.

Withdrawal can be minimized but not always prevented by slow tapering over 3 to 6 months or longer. Once it has developed, withdrawal from benzodiazepines and related CNS depressants is treated by reinstituting the previous dose and withdrawing the medication more slowly or by substituting a longer-acting benzodiazepine or another CNS depressant with a known pattern of withdrawal, preferably one that acts at the same receptor complex. Since barbiturates are cross-reactive with benzodiazepines, a barbiturate tolerance test can be used to more precisely determine the dose of barbiturate necessary to suppress withdrawal, after which the barbiturate is gradually withdrawn. Possible adjunctive treatments for benzodiazepine withdrawal include melatonin and melatonin agonists, carbamazepine, pregabalin, and CBT.

Management of alprazolam withdrawal can be particularly difficult. If the medication is not administered frequently, interdose symptoms are common. Furthermore, as chronic doses of alprazolam are reduced below approximately 1 to 1.5 mg, the benzodiazepine receptor undergoes paradoxical downregulation instead of continuing to upregulate, resulting in an intensification of withdrawal symptoms, even with a very slow taper. Some clinicians find that substitution of equivalent doses of clonazepam facilitates withdrawal of the final doses of alprazolam, though this benefit is not always reliable. There is no literature on the management of persistent withdrawal syndromes.

Dependence and Misuse

Some experts assume that benzodiazepines have a high risk of misuse. Others contend that anecdotal reports of benzodiazepine misuse and dependence are not generalizable and that data supporting the contention that prescription of benzodiazepines frequently leads to abuse of other substances have been mixed. A 2015–2016 survey found that 17% of people taking benzodiazepines reported misuse, which was defined as using without a prescription or not using as prescribed; rates of benzodiazepine misuse were highest in those aged 18 to 25 (51%, vs. 4% of those aged 65 or older). It is not known how many people reporting misuse were taking medications prescribed for them versus medications obtained on the street or from friends and family. The same limitation applies to the findings of a 2016 national drug adverse event surveillance protocol that reported 358,247 ED visits for harms from nonmedical use of pharmaceuticals. Illicit drugs were involved in 34% of visits, alcohol in 22%, benzodiazepines in 47%, and prescription opioids in 36%. Benzodiazepines were the only substance involved in 6.5% of cases, and prescription opioids were the only drug in 11.3%. The authors concluded that while benzodiazepines are frequently involved in ED visits for nonmedical drug use, they are most frequently used along with other substances. In such instances, a benzodiazepine is used to augment the “high” produced by other substances. A 2017 survey found that 17% of more than 2 million people being treated in publicly funded substance abuse programs identified benzodiazepines as a secondary or tertiary drug of abuse, whereas only 1% said that benzodiazepines were the primary drug of abuse. Data on 86,186 adults from the National Survey on Drug Use and Health confirmed that misuse of benzodiazepines was usually in combination with stimulants or opioids.

A retrospective cohort study using databases from five Massachusetts agencies reported that 24% of 63,345 adults who received buprenorphine in a 4-year period filled at least one benzodiazepine prescription during treatment with buprenorphine. Concomitant use of a benzodiazepine during buprenorphine treatment was associated with a 2.92 times increased risk of fatal opioid overdose, a 2.05 times increased risk of nonfatal opioid overdose, and a 1.90 increased risk of all-cause mortality, but a 13% decreased likelihood of buprenorphine discontinuation. The latter finding is somewhat consistent with a study of urine toxicology and prescriber information from a single payer in which patients receiving methadone maintenance taking nonprescribed benzodiazepines, but not those taking prescribed benzodiazepines, were more likely to drop out of treatment. There may therefore be a difference between prescribed and illicit benzodiazepine use in patients who misuse other substances. In fact, benzodiazepine dependence is relatively uncommon in patients without a history of substance use disorder when the medication is prescribed for an appropriate reason.

Misuse of benzodiazepines may be becoming more of a problem in younger people. Pediatric benzodiazepine use steadily increased from 2000 to 2015, and in 2018, 3.9% of 10th- and 12th-grade high school students reported nonmedical use of benzodiazepines. Some adolescents may be switching from recreational use of illicit drugs to prescription medications on the grounds that they are safer because they are approved by the government. The combination of increased misuse of benzodiazepines and increased distress associated with restrictions due to the COVID-19 pandemic has contributed to escalating overdoses involving benzodiazepines in adolescents: the number of emergency department visits for benzodiazepine overdoses increased by 34% with opioids and 21% without opioids.

Although there were fewer than 3,000 fatal overdoses involving benzodiazepines in 2020, the number increased by 43% (from 1,004 to 1,435) from April–June 2019 to the same period in 2020, including a 22% increase in deaths associated with prescription benzodiazepine overdoses and a 520% increase in deaths associated with nonprescribed benzodiazepines. Aside from alprazolam (LD_{50} 300 to 2,000 mg/kg), the lethality in overdose of benzodiazepines alone other than alprazolam in overdose is low (LD_{50}s in the thousands of mg/kg). Benzodiazepines increase the lethality of opioid and alcohol overdoses. In 2020, 92.7% of fatal benzodiazepine overdoses also involved opioids, especially illegal forms of fentanyl. Such data suggest that the bulk of benzodiazepine misuse and fatal benzodiazepine overdoses involves concomitant misuse of opioids and that administering opioid antagonists should be considered if a benzodiazepine overdose does not respond promptly to flumazenil.

In September 2020, the FDA announced that it would revise the boxed warning for benzodiazepines to include warnings about risks of abuse, addiction, dependence, and withdrawal syndromes. The FDA will also advise that benzodiazepines should primarily be used short-term, that co-prescribing with opioids should be avoided, and that alternative treatments should be considered. This advice may result in inappropriate withholding of benzodiazepines and, in some cases, the use of cannabis, alcohol, and other substances instead. Viewing taking more medication than prescribed as a sign of addiction may lead to unnecessarily discontinuing the medication when it is actually the result of suboptimal prescribing, attempts to eradicate all symptoms completely, interdose withdrawal and rebound, and poor doctor–patient communication rather than attempts to achieve euphoria.

Adverse Effects of Z-Drugs

Healthcare providers often overestimate the benefits and minimize the risks of Z-drugs compared to benzodiazepines. Reports of low rates of problems such as tolerance, withdrawal, and rebound insomnia with Z-drugs have been based on shorter-term studies and less experience than benzodiazepines, and more recent data suggest that tolerance and rebound insomnia do occur. Like benzodiazepines, Z-drugs can cause falls. Physical dependence has been associated with all of the newer benzodiazepine receptor agonist hypnotics, as have sleep-related behaviors, including sleep driving. Because of its short elimination half-life, zolpidem usually does not produce much daytime impairment, but memory impairment similar to that seen with triazolam can occur approximately 1.5 hours after dosing, and complex automatisms including driving and eating during sleep have been reported with amnesia for the behavior. Adverse effects are more common at doses exceeding 20 mg. Concerns about adverse effects and withdrawal have led to a reduction of the recommended dose of zolpidem to 5 mg for the immediate release form and 6.25 mg for the extended-release form.

In doses up to 10 mg, zaleplon has been found to cause dizziness, headache, somnolence, and psychomotor impairment. These side effects usually abate within 2 to 3 hours as a result of the short duration of action of the drug, minimizing daytime impairment. At doses of 60 mg, psychomotor impairment persists for 25 hours. The abuse potential of zaleplon and zolpidem appears to be low. In addition to expected side effects such as somnolence and dizziness, eszopiclone can also cause a bitter taste. An Irish study found that 23% of patients in a methadone maintenance program were abusing zopiclone, most frequently along with benzodiazepines or heroin. Although tolerance and rebound insomnia have been rare in clinical trials, rebound insomnia is more common than other discontinuation syndromes with eszopiclone.

DRUG INTERACTIONS AND LABORATORY INTERFERENCES

Benzodiazepines and Z-drugs all have additive sedative effects with each other and with other CNS depressants. As noted earlier, a number of other medication classes, including barbiturates, are cross-reactive with benzodiazepines and can be substituted for them in discontinuation protocols. Many benzodiazepines are substrates for CYP3A4 and CYP2C19. CYP3A inhibitors such as ketoconazole, nefazodone, and the antiretroviral lopinavir can substantially increase levels of benzodiazepines such as midazolam, diazepam, and triazolam. These medications should not be taken with grapefruit juice, which induces CYP3A4. However, even though alprazolam is a CYP3A4 substrate, it does not seem to be affected by grapefruit juice. L-Type calcium channel blockers and ritonavir may increase zolpidem effects.

No effects of benzodiazepines and Z-drugs on laboratory tests have been reported.

DOSAGE AND ADMINISTRATION

Usual doses of benzodiazepines and Z-drugs are summarized in Table 33.9b–1. Even benzodiazepines with long elimination half-lives are often given in divided doses to avoid oversedating peak levels. In chronic treatment, especially at higher doses, it is sometimes necessary to administer alprazolam frequently (e.g., every 2 to 3 hours) to prevent interdose withdrawal and rebound anxiety. An alprazolam sustained-release preparation (Xanax XR) has acceptable steady-state kinetics with three daily doses. Slow absorption of temazepam administered at bedtime makes this approach most useful for patients who have no trouble falling asleep but wake up early. If this medication is to be used for difficulty falling asleep, it should be taken 1 to 2 hours before bedtime.

Chronic and relapsing anxiety disorders such as GAD and panic disorder often require ongoing treatment, and withdrawing benzodiazepines after a few months of treatment often results in a relapse of chronic and relapsing anxiety disorders such as GAD and panic disorder. If maintenance pharmacotherapy with an antidepressant is planned, but the antidepressant induces intolerable early jitteriness, adding a benzodiazepine can suppress the initial increase in anxiety induced by the antidepressant. As noradrenergic activity is gradually downregulated over the first few months of treatment, the benzodiazepine can be gradually withdrawn. Conversely, for some patients who have been taking stable doses of benzodiazepines for years without evidence of impairment or misuse and who do not tolerate or respond well to an antidepressant, continuing the benzodiazepine is preferable to attempting multiple unsuccessful attempts at drug discontinuation.

The initial bedtime dose of oral zolpidem, sublingual Edluar, and oral spray Zolpimist is 5 mg. The dose can be increased to 10 mg. The controlled-release formulation of zolpidem (Ambien CR) is started at a dose of 6.25 mg, and the dose can be increased to 12.5 mg for males if necessary. The initial dose of Intermezzo is 1.75 mg for women and 3.5 mg for men. Zaleplon is started at a bedtime dose of 5 mg, and the dose can be increased to 10 to 20 mg. The short duration of action of zaleplon makes it applicable for patients who have to get back to sleep after waking up in the middle of the night. Eszopiclone is started at a bedtime dose of 1 mg, and the dose is increased as needed to 2 to 3 mg.

BENZODIAZEPINE RECEPTOR ANTAGONISTS

Neuroscience-Based Nomenclature (NbN)	
Flumazenil	
Pharmacologic Domain	GABA
Mechanism of Action	Receptor Antagonist (GABA-A receptor, α-1,2,3, and 5 subtypes)

As noted earlier, antagonists occupy receptors but have no effect on receptor signaling. The only available benzodiazepine antagonist is flumazenil, which has a benzodiazepine structure and is used to block the action of benzodiazepines and Z-drugs. The most common application of this effect is in the treatment of benzodiazepine overdose or to reverse conscious sedation with benzodiazepines such as midazolam.

Pharmacologic Actions

Flumazenil is administered intravenously because extensive presystemic metabolism results in low bioavailability after oral dosing, necessitating high oral doses that most patients cannot tolerate. Flumazenil has an elimination half-life of approximately 15 minutes and is metabolized to a free acid and a glucuronide, which are excreted in the urine. The drug is completely eliminated within 72 hours. Age and gender do not affect pharmacokinetics, but liver disease slows metabolism and prolongs the half-life and duration of action of flumazenil.

Flumazenil reverses CNS depression caused by benzodiazepines. Although alcohol and barbiturates act on the same receptor complex, their primary action is not at the benzodiazepine receptor, and sedation and other signs of CNS depression caused by these substances are not antagonized by flumazenil. The effects of opioids, which act at an entirely different site, are not affected by flumazenil. Emergency treatment of a known or presumed mixed narcotic/benzodiazepine overdose, or an apparent benzodiazepine overdose that does not respond to the benzodiazepine antagonist, may require administration of an opioid antagonist (naloxone) as well as flumazenil.

Therapeutic Indications

Flumazenil is indicated for the reversal of benzodiazepine-induced CNS depression caused by intentional overdose or reversal of anesthesia or conscious sedation. It has also been used experimentally to treat hepatic coma.

Precautions and Adverse Reactions

Common adverse effects of flumazenil include nausea, vomiting, dizziness, agitation, cutaneous vasodilation, headache, agitation, visual problems, memory impairment, and emotional lability. Benzodiazepine withdrawal may be precipitated by flumazenil, especially in patients taking benzodiazepines chronically. Seizures may occur in patients with epilepsy and in patients who have taken an overdose of TCAs.

The most important interaction of flumazenil is to antagonize the effect of benzodiazepines and Z-drugs. In benzodiazepine overdoses that include other medications, reversal of sedation from the benzodiazepine may uncover intoxication with the other substances. If a mixed overdose includes medications that can induce seizures, reversal of the anticonvulsant effect of the benzodiazepine can lead to the emergence of seizures driven by the other substance.

No interference with laboratory tests has been reported.

Dosing and Administration

The starting dose of flumazenil is 0.2 mg (2 mL), given intravenously over 30 seconds. A 0.3-mg (3-mL) dose is added if the patient does not regain consciousness within 1 minute. This dose can be repeated up to a maximum total dose of 5 mg over approximately 5 minutes. If unconsciousness is not reversed with this dose, it is probably not caused by benzodiazepines, and further treatment with flumazenil is not likely to be effective.

INVESTIGATIONAL DRUGS

Abecarnil is a β-carboline benzodiazepine receptor partial agonist that has anxiolytic and anticonvulsant actions with minimal sedation and psychomotor impairment. This medication may be useful in treating GAD and uncomplicated alcohol withdrawal syndromes. Abecarnil has not been released in the United States. Bretazenil and imepitoin are imidazolone partial benzodiazepine receptor agonists with potential applications in anxiety and seizure disorders that seem to be associated with minimal tolerance and potential for misuse. A benzodiazepine S-enantiomer, (S)-1-(2-fluorophenyl)-3-(2-oxo-5-phenyl-2,3-dihydro-1H-benzo[e] [1,4]-diazepin-3-yl)-urea (RSV604), inhibits respiratory syncytial virus replication by binding to the viral nucleoprotein.

Naturally occurring flavones bind to the benzodiazepine receptor, and multiple synthetic flavonoids have been designed to have agonist and inverse agonist effects. In animal studies, Gardenin A, a flavonoid found in medicinal plants, has been found to have anxiolytic, sedative, antidepressant, and anticonvulsant properties. Later studies contradicted early findings that the flavonoid 6-bromo-3’-nitroflavone, which has a structure derived from plants used in folk medicine as tranquilizers, had potent anxiolytic properties.

The PBR also referred to as the TSPO, is involved in apoptosis, inflammation and steroidogenesis. TSPO ligands have potential antiapoptosis actions that could be useful in Alzheimer disease. Conversely, the PBR ligand PK11195, but not diazepam, induces apoptosis in some cancer cell lines and is a potential adjunctive agent in the treatment of some forms of cancer. In some cell lines, however, PK11195 can prevent cytokine-induced apoptosis. An agonist effect at the

PBR could help protect the brain from oxidative damage and could possibly reduce the progression of some neurodegenerative diseases.

Indiplon, a potential hypnotic that is selective for the benzodiazepine-1 receptor, also inhibits platelet aggregation. Ocinaplon, a full agonist at benzodiazepine-1 receptors and a partial agonist at benzodiazepine-2 and -3 receptors, was withdrawn during Phase III studies because of elevated liver enzymes. Pagoclone is a cyclopyrrolone that is a partial agonist of benzodiazepine-1 and -2 receptors and a full agonist at PBRs that is under study as a treatment for generalized anxiety disorder and panic disorder. A placebo-controlled study followed by an open-label extension found significantly more improvement in stuttering with pagoclone and continued improvement during an open-label extension. Agonists of the $GABA_B$ receptor such as baclofen have been used to treat the misuse of substances such as alcohol, cocaine, and opioids. Experimental $GABA_B$ partial agonists, including CGP7930, GS39783, NVP-BHF177, and BHFF, are associated with less receptor desensitization and, therefore, less tolerance and have been found to have anxiolytic effects. Ongoing studies of $GABA_B$ antagonists appear to have antidepressant and possibly cognitive enhancement properties in preliminary studies. Ligands of peripheral benzodiazepine and $GABA_B$ receptors may be promising targets for the development of nonbenzodiazepine medications with novel applications.

FURTHER READINGS

Amato L, Minozzi S, Vecchi S, Davoli M. Benzodiazepines for alcohol withdrawal. *Cochrane Database Syst Rev.* 2010;3:CD005063.

Arbo BD, Ribeiro MF, Garcia-Segura LM. Development of new treatments for Alzheimer’s disease based on the modulation of translocator protein (TSPO). *Ageing Res Rev.* 2019;54(09):100943.

Bais B, Molenaar NM, Bijma HH, et al. Prevalence of benzodiazepines and benzodiazepine-related drugs exposure before, during and after pregnancy: A systematic review and meta-analysis. *J Affect Disord.* 2020;269:18–27.

Beach SR, Gomez-Bernal F, Huffman JC, Fricchione GL. Alternative treatment strategies for catatonia: a systematic review. *Gen Hosp Psychiatry.* 2017;48:1–1.

Benasi G, Guidi J, Offidani E, et al. Benzodiazepines as a monotherapy in depressive disorders: a systematic review. *Psychother Psychosom.* 2018;87:65–74.

Busto U, Sellers EM, Naranjo CA, et al. Withdrawal reaction after long-term therapeutic use of benzodiazepines. *N Engl J Med.* 1986;315:854–859.

Delamarre L, Galvao F, Gohier B, Poulet E, Brunelin J. How much do benzodiazepines matter for electroconvulsive therapy in patients with major depression? *J ECT.* 2019;35(3):184–188.

Dold M, Li C, Gillies D, Leucht S. Benzodiazepine augmentation of antipsychotic drugs in schizophrenia: a meta-analysis and Cochrane review of randomized controlled trials. *Eur Neuropsychopharmacol.* 2013;23(9):1023–1033.

Dujardin S, Pijpers A, Pevernagie D. Prescription drugs used in insomnia. *Sleep Med Clin.* 2020;15:133–145.

Dyer AH, Murphy C, Lawlor B, Kennelly SP. Cognitive outcomes of long-term benzodiazepine and related drug (BZDR) use in people living with mild to moderate Alzheimer's disease: Results from NILVAD. *JAMDA.* 2020;21:194–200.

Greenblatt DJ, Harmatz JS, Shader RI. Diazepam in the elderly: Looking back, ahead and at the evidence. *J Clin Psychopharmacol.* 2020;40(3):215–219.

Hafdi M, Hoevenaar-Blom MP, Beishuizen CRL, et al. Association of benzodiazepine and anticholinergic drug usage with incident dementia: a prospective cohort study of community-dwelling older adults. *JAMDA.* 2020;21:188–193.

Hirschtritt ME, Olfson M, Kroenke K. Balancing the risks and benefits of benzodiazepines. *JAMA.* 2021;325(4):347–348.

Hoffmann E, Nomikos GG, Kaul I, et al. SAGE-217, a novel GABAA receptor positive allosteric modulator: clinical pharmacology and tolerability in randomized Phase I dose-finding studies. *Clin Pharmacokinet.* 2020;59(1):111–120.

Jacobs GD, Pace-Schott EF, Stickgold R, Otto MW. Cognitive behavior therapy and pharmacotherapy for insomnia. *Arch Intern Med.* 2004;164:1888–1896.

Li Y, Ma J, Li N, et al. Benzodiazepines for treatment of patients with delirium excluding those who are cared for in an intensive care unit. *Cochrane Database Syst Rev*. 2020;2:CD012670.

Maust DT, Lin LA, Blow FC. Benzodiazepine use and misuse among adults in the United States. *Psychiatr Serv*. 2019;70(2):97–106.

Moosmann B, Auwarter V. Designer benzodiazepines: another class of new psychoactive substances. *Handb Exp Pharmacol.* 2018;252:383–410.

Moylan S, Staples J, Ward SA, et al. The efficacy and safety of alprazolam versus other benzodiazepines in the treatment of panic disorder. *J Clin Psychopharmacol.* 2011;31(5):647–652.

Nisavic M, Nejad SH, Isenberg BM, et al. Use of phenobarbital in alcohol withdrawal management- a retrospective comparison study of phenobarbital and benzodiazepines for acute alcohol withdrawal management in general medical patients. *Psychosomatics.* 2019;60(5):458–467.

Quagliato LA, Cocsi F, Shader RI, et al. Selective serotonin reuptake inhibitors and benzodiazepines in panic disorder: a meta-analysis of common side effects in acute treatment. *J Psychopharmacol.* 2019;33(11):1340–1351.

Richardson K, Mattishent K, Loke YK, et al. History of benzodiazepine prescriptions and risk of dementia: possible bias due to prevalent users and covariate measurement timing in a nested case-control study. *Am J Epidemiol.* 2019;188(7):1228–1236.

Rickels K, Schweizer E. Panic disorder: long-term pharmacotherapy and discontinuation. *J Clin Psychopharmacol.* 1998;18:12S–18S.

Rickels K, Moeller HJ. Benzodiazepines in anxiety disorders: Reassessment of usefulness and safety. *World J Biol Psychiatry.* 2019;20:514–518.

Sateia MJ, Buysse DJ, Krystal AD, Neubauer DN, Heald JL. Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: an American Academy of Sleep Medicine clinical practice guideline. *J Clin Sleep Med.* 2017;13(2):307–349.

Simoni-Wastila L, Ross-Degnan D, Mah C, et al. A retrospective data analysis of the impact of the New York triplicate prescription program on benzodiazepine use in Medicaid patients with chronic psychiatric and neurologic disorders. *Clin Ther.* 2004;26(2):322–336.

Tonon MC, Vaudry H, Chuquet J, et al. Endozepines and their receptors: structure, functions and pathophysiological significance. *Pharmacol Ther.* 2020;208(04):107386.

Tori ME, Larochelle MR, Naimi TS. Alcohol or benzodiazepine co-involvement with opioid overdose deaths in the United States, 1999–2017. *JAMA Network Open*. 2020;3(4):e202361.

van der Sluiszen NNJJM, Vermeeren A, Verster JC, et al. Driving performance and neurocognitive skills of long-term users of benzodiazepine anxiolytics and hypnotics. *Human Psychopharmacol.* 2019;34(6):e2715.

Zhang Y, Zhou X, Meranus DH, Wang LN, Kukull WA. Benzodiazepine use and cognitive decline in elderly with normal cognition. *Alzheimer Dis Assoc Disord.* 2016;30(2):113–117.

33.10 Cation Function Modulators and Lithium (“Mood Stabilizers”)

33.10a Calcium Channel Inhibitors

Steven L. Dubovsky, M.D., and Dori Marshall, M.D.

The development of novel mood-stabilizing medications has been hampered by the lack of a cohesive neurobiologic model that could explain the frequent mixtures and rapid alternations of elevated and depressed mood, the complex presentations of bipolar mood disorders, and their comorbidity with certain medical conditions. One new model involves intracellular messengers that directly regulate multiple aspects of cellular function in affective, cognitive, vegetative, and behavioral systems and can have enduring effects on those systems through altered gene expression. The same impact of changes in intracellular signaling can modify vascular, cardiac, endocrine, and other bodily systems. Treatments aimed at correcting dysregulated intracellular signaling could have potential therapeutic applications. The intracellular calcium ion (Ca^{2+}) is particularly interesting as a target for pharmacologic intervention. Within neurons, intracellular Ca^{2+} signaling regulates excitability, circadian and other rhythms, the sleep–wake cycle, information processing, sensory perception, cognition, consciousness, neurotransmission, synaptic proteins, regulation of expression of multiple genes, remodeling of neuronal architecture, and neuroplasticity, which is essential for neuronal resilience and stabilization of systems that regulate mood.

Free intracellular Ca^{2+} concentration ($[Ca^{2+}]_i$) is normally regulated tightly at around 100 nM, or 1/10,000 of the Ca^{2+} concentration in the extracellular fluid. Most intracellular calcium is inactive by being stored in organelles such as mitochondria and the endoplasmic reticulum or complexed with binding proteins. Free intracellular calcium levels are increased rapidly by release from intracellular stores, while influx through calcium channels replenishes intracellular stores, triggers the release of stored Ca^{2+} and help to regulate the baseline concentration of free intracellular Ca^{2+} ($[Ca^{2+}]_B$), which in turn influences neuronal excitability (Fig. 33.10a–1). Cellular activation, which may be initiated by agonists such as neurotransmitters or may be spontaneous, increases $[Ca^{2+}]_i$, resulting in stimulated free intracellular Ca^{2+} concentration ($[Ca^{2+}]_S$), activating multiple cellular processes. Resting $[Ca^{2+}]_i$ is restored by the active transport of Ca^{2+} into intracellular stores and out of the cell.

Intracellular calcium signaling has a biphasic effect (Fig. 33.10a–1). Moderate increases stimulate cellular activity, whereas greater elevations inhibit the same functions. Increased intracellular Ca^{2+} signaling in excitatory and inhibitory systems could change the balance between these systems, leading to alternations between excitatory and inhibitory states manifested as mania and depression, while these states could occur simultaneously if a hyperactive intracellular Ca^{2+} signal produced behavioral and emotional activation in some systems and suppression in others. This possibility is supported by observations that hypercalcemia can be associated with depression or mania. Enhancement of Ca^{2+} signaling by antidepressants could lead to antidepressant-induced mania, while attenuation by mood stabilizers of increased intracellular calcium signaling could improve both mania and bipolar depression. Increased intracellular calcium signaling can induce oxidative stress and apoptosis, contributing to the neuronal loss that has been observed with recurrent affective episodes. Hyperactive calcium signaling in other organs could play a role in the comorbidity of bipolar mood disorders with other conditions linked to cellular hyperactivity such as migraine headaches, hypertension, and coronary heart disease.

Empirical evidence has continued to accumulate in multiple cellular models of hyperactive intracellular Ca^{2+} signaling in bipolar disorder. Compared with controls, bipolar disorder patients have been found to have elevations of $[Ca^{2+}]_B$ and $[Ca^{2+}]_S$ in blood platelets, lymphocytes, B-lymphoblast cell lines (BLCLs), olfactory neurons, and neuroblastoma cells. Most but not all studies find increased $[Ca^{2+}]_i$ in mania and bipolar depression, but not in unipolar depression, as well as normalization of $[Ca^{2+}]_i$ with normalization of mood regardless of treatment (Fig. 33.10a–2). The same finding, regardless of cell type, along with a lack of increase of $[Ca^{2+}]_i$ in control platelets after incubation with a plasma ultrafiltrate from bipolar disorder patients with elevated platelet $[Ca^{2+}]_i$, suggests that hyperactive intracellular Ca^{2+} signaling represents a generalized change in intracellular calcium ion dynamics rather than the result of a circulating substance. The lack of hyperactive platelet intracellular Ca^{2+} signaling in schizophrenia suggests that the finding is related to a core feature of bipolar disorder and not a feature shared with other conditions such as psychosis.

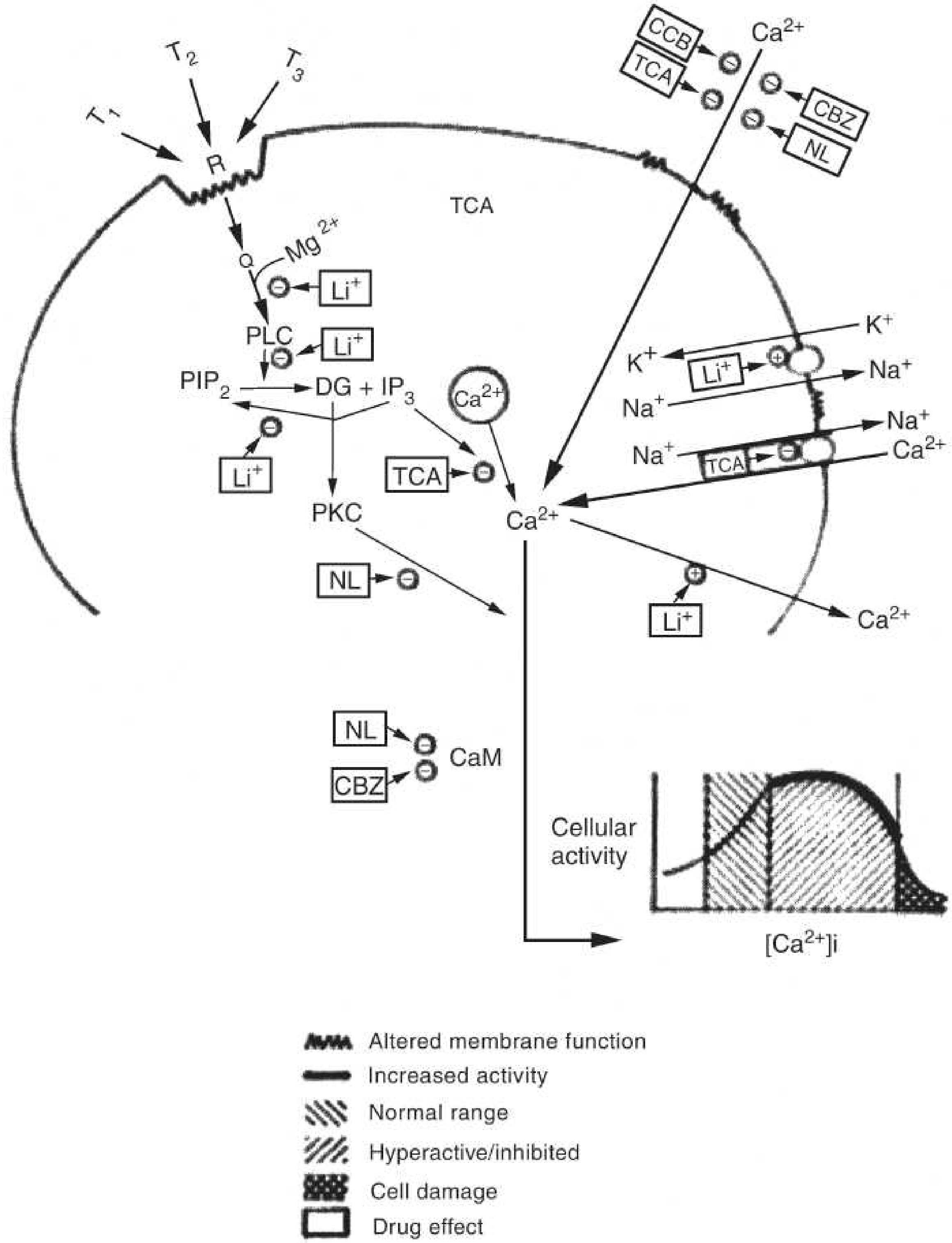
T1
T2
T3
R
G
Mg2+
Li+
PLC
Li+
PIP2
DG + IP3
Li+
PKC
TCA
Ca2+
TCA
NL
CCB
TCA
Ca2+
CBZ
NL
K+
K+
Li+
Na+
Na+
Na+
Na+
TCA
Ca2+
Ca2+
Li+
Ca2+
NL
CBZ
CaM
Cellular activity
[Ca2+]i
Altered membrane function
Increased activity
Normal range
Hyperactive/inhibited
Cell damage
Drug effect

FIGURE 33.10a–1. Intracellular calcium ion signaling. Intracellular calcium ion concentration ($[Ca^{2+}]_i$) is increased by entry into the extracellular space and by release from intracellular stores. Calcium enters the cytosol through receptor operated (e.g., *N*-methyl-D-aspartate) and potential-dependent ion-specific channels, as well as in exchange for intracellular sodium. Stimulation

of a variety of G-protein–associated receptors (e.g., serotonin type 2 and dopamine type 2) activates turnover of the membrane phosphoinositol system, with production of inositol triphosphate (IP3) and diacylglycerol (DG). IP3 releases calcium ions from intracellular stores, and DG draws protein kinase C (PKC) to the cell membrane, where it is activated by Ca^{2+}. Moderate increases in $[Ca^{2+}]_i$ stimulate cellular processes by activating PKC and a variety of other enzymes that are activated by the calcium–calmodulin (CaM) complex; greater increases suppress the same processes. CBZ, carbamazepine; CCB, calcium channel blocker; Li, lithium; NL, neuroleptic; PIP2, phosphatidylinositol biphosphate; PLC, phospholipase C; TCA, tricyclic antidepressant.

Animal studies support these findings. Heterozygous Myshkin (Myk/+) mice, which have a missense mutation of the neuron-specific sodium/potassium ATPase, exhibit mania-like behavior such as hyperambulation, altered sleep–wake cycle, and increased reward preference. The manic-like behavior, which is reduced by lithium and valproate, is associated with increased release of Ca^{2+} from intracellular stores in cortical neurons and elevated glutamate-induced Ca^{2+} currents. In infant pigtail and bonnet macaque monkeys separated from their mothers or peers, which produces observable agitation, distress, depression, and immune suppression that remit with a reunion, lymphocyte $[Ca^{2+}]_B$ and mitogen-stimulated $[Ca^{2+}]_S$ increased significantly, and levels returned to pre-separation values with a reunion.

Elevated resting $[Ca^{2+}]_i$ in bipolar disorder may be the result of an altered balance between homeostatic mechanisms such as calcium influx through calcium channels, efflux through ion exchange, and intracellular storage (Fig. 33.10a–1). Rapid increases in $[Ca^{2+}]_i$ associated with cellular activation are the result of the release of Ca^{2+} from intracellular stores. One clearly demonstrated the mechanism of release of stored Ca^{2+} involves turnover of the membrane phosphatidylinositol (PI) system, beginning with the hydrolysis of membrane-bound PI bisphosphate (PIP_2) through several intermediate metabolites to produce two end products of metabolism—inositol triphosphate (IP_3) and diacylglycerol (DG). IP_3 acts on mitochondrial and other intracellular receptors to promote the release of stored Ca^{2+}, while DG attracts the intracellular enzyme protein kinase C (PKC) to the cell membrane, where it is activated by free intracellular Ca^{2+} and phosphorylates intracellular proteins that regulate multiple cellular processes. PIP_2 is reconstituted from these metabolites through a number of steps for another round of cellular activation.

Dysfunction at various points in PI turnover has been suggested in bipolar disorder. For example, an allele of IMPA2, the gene for inositol monophosphatase (the rate-limiting step in the reconstitution of membrane PIP_2), has been reported to convey susceptibility to bipolar disorder probably by increasing IP_3 and PKC production. Additional proposed mechanisms of increased free intracellular Ca^{2+} signaling in bipolar disorder include hyperactivity of the calcium-binding protein neuronal calcium sensor-I (NCS-1), which is activated by IP_3, and decreased activity of B-cell lymphoma-2 (Bcl-2), an antiapoptotic and pro-neurotrophic intracellular protein that interacts with the IP_3 receptor to inhibit the release of stored Ca^{2+}. Decreased Bcl-2 activity, which has been reported in bipolar disorder, would contribute to increased Ca^{2+} release from intracellular stores. A single nucleotide polymorphism (SNP) of the Bcl-2 gene associated with decreased gene expression has been found in lymphoblasts of patients with bipolar disorder, along with increased $[Ca^{2+}]_B$ and increased IP_3-mediated $[Ca^{2+}]_S$, presumably associated with greater Ca^{2+} release from intracellular stores. Increased intracellular calcium signaling could exacerbate the loss of the antiapoptotic function, promoting the loss of cerebral grey matter in some bipolar disorder patients.

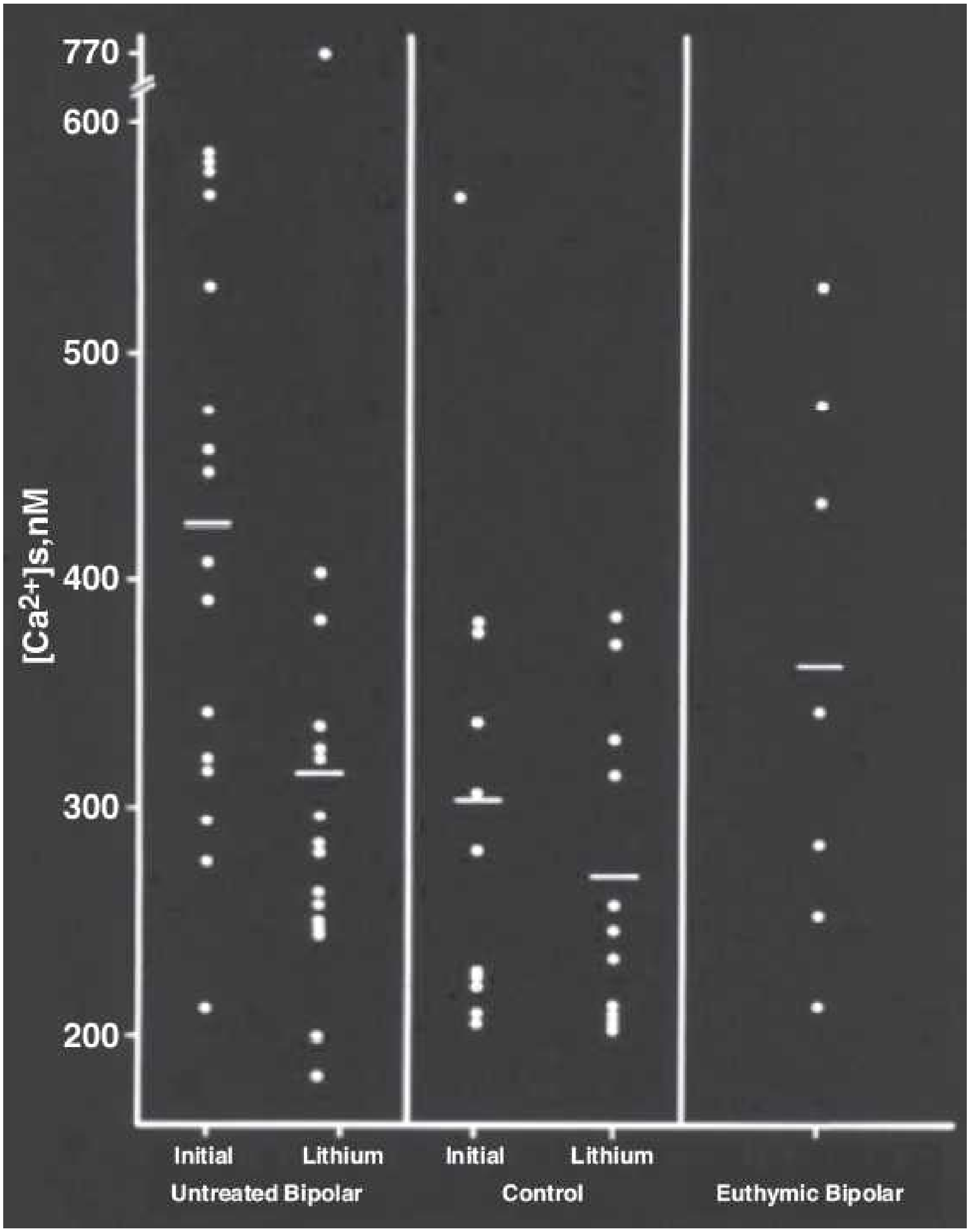

FIGURE 33.10a–2. Elevated platelet intracellular calcium ion concentration in bipolar disorder and normalization by lithium (Eskalith). Resting ($[Ca^{2+}]_B$) and, in this figure, stimulated ($[Ca^{2+}]_S$), as well as interindividual variability, are increased in mania and bipolar depression. Hyperactive intracellular calcium signaling is normalized by in vitro incubation with therapeutic concentrations of lithium, and bipolar patients who are euthymic after treatment with lithium or anticonvulsants have mean $[Ca^{2+}]_S$ levels that are statistically the same as in controls.

Lithium, the prototypical mood-stabilizing medication that is effective for both poles of bipolar mood disorders, reduces elevated $[Ca^{2+}]_B$ and $[Ca^{2+}]_S$ in peripheral cells of patients with bipolar disorder but does not affect normal $[Ca^{2+}]_i$ in most studies (Fig. 33.10a–2). Lithium inhibits inositol monophosphatase, the rate-limiting step in the reconstitution of PIP_2, which is upregulated in some cases of bipolar disorder; inhibition of the enzyme would eventually deplete IP3, decreasing both intracellular Ca^{2+} release and production of PKC. Mechanisms of calcium antagonism by lithium that could be relevant to the reduction of $[Ca^{2+}]_i$ include competitive inhibition of IP_3, inhibition of the calcium-binding protein NCS-1, interfering with calcium release from intracellular stores, decreased calcium influx, enhancement of calcium efflux by stimulating sodium-calcium countertransport, and inhibition of glycogen synthase kinase 3β.

Carbamazepine, which is also a well-established mood stabilizer, inhibits Ca^{2+} currents in a variety of cellular models, and the time course of suppression of calcium-dependent potentials is comparable to that produced by the calcium channel blocker verapamil. Similar to lithium, in vitro incubation with carbamazepine significantly lowers lymphocyte $[Ca^{2+}]_i$ in ill bipolar patients but not in controls or euthymic bipolar disorder patients. Valproate, another anticonvulsant with mood-stabilizing and especially antimanic properties, was found to reduce lysophosphatidic acid (LPA)-stimulated increased $[Ca^{2+}]_i$ in B lymphoblast cell lines from bipolar disorder patients and thapsigargin-stimulated release of Ca^{2+} from intracellular stores to an extent similar to lithium. Lithium, valproate, and ECT have been noted to upregulate Bcl-2 gene expression in some studies in conjunction with reducing elevated $[Ca^{2+}]_i$. Conversely, levetiracetam, an anticonvulsant that antagonizes *N*-type calcium channels, does not reduce hyperactive intracellular calcium ion signaling in bipolar disorder, and it is not effective as a mood stabilizer.

CALCIUM CHANNELS

Calcium ions enter the cytosol from the extracellular space through receptor-operated channels gated by receptors for hormones and neurotransmitters and through potential dependent channels (PDCs or voltage-gated channels), which are gated by membrane potential (Fig. 33.10a–1). In the brain, potential-dependent calcium channels are localized in regions rich in synapses, perhaps because high $[Ca^{2+}]_i$ must be produced rapidly to regulate neurotransmitter release. Additional pathways for calcium entry include leakage through an ungated channel and exchange of extracellular calcium ions for intracellular sodium ions (Na^+). Under physiologic conditions, PDCs can be regulated by receptor-mediated events such as the production of inositol triphosphate, and receptor-operated channels can be gated by changes in membrane potential. Furthermore, extracellular Ca^{2+} entering the cell may release calcium ions from intracellular stores, and trigger amounts of calcium ions released from intracellular stores may facilitate calcium channel opening. Endogenous regulators, which may be altered in disease states, appear to modulate PDC gating. Even subtle alterations of the function of one kind of calcium channel can therefore have significant effects on the overall balance of calcium-dependent cellular activity.

The PDC (Fig. 33.10a–3) consists of a central pore-forming CaV subunit and five allosterically linked units called α_1, α_2, β, γ, and δ. The PDC consists of a central pore-forming CaV subunit and five allosterically linked units: α1, α2, β, γ, and δ. There are 10 CaV-α1 subunits organized into various calcium channels. The mammalian nervous system contains 9 of the 10 major calcium channel types, with even more functional diversity related to alternative splicing events, as well as association with different subunits and regulatory proteins. Thus far, four separate genes coding for the α1 subunit have been identified. The α1C subunit of the CaV1.2 channel, a primary source of Ca^{2+} entry in neurons in the brain, is coded by the CACNA1C gene, an allele of which has been associated with bipolar disorder, as discussed below. The other subunits, and possibly other endogenous circulating factors, can alter the conformation of the α1 subunit, changing the activity of the calcium channel and its affinity for medications that bind to it at specific binding sites.

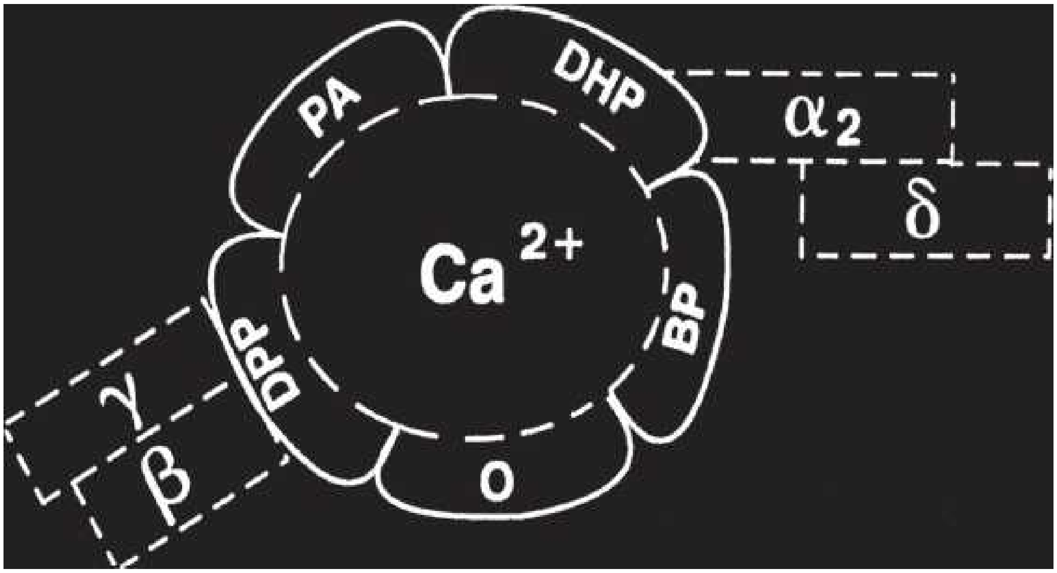

FIGURE 33.10a–3. Structure of the calcium channel blocker–binding site. BP, benzodiazepine-binding site; DHP, 1,4-dihydropyridine-binding site; DPP, diphenylpiperazine-binding site; O, other calcium channel blocking agent binding sites; PA, phenyl alkylamine–binding site.

Table 33.10a–1.
Calcium Channel Subtypes

Channel Type	L	T	N	P	Q	R
Locations	CV, neurons, endocrine, skeletal muscle	Neurons, smooth muscle, SA node	Neurons	Neurons	Neurons	Neurons
Functions	Excitation-contraction coupling, NT release, vascular tone	SA node activity, repetitive neuron depolarization, hormone release	NT release	NT release	NT release	NT release

CV, cardiovascular system; SA, sinoatrial node; NT, neurotransmitter.

Six types of potential-dependent calcium channels have been characterized (Table 33.10a–1). Different calcium channels in the nervous system perform different physiologic functions, and multiple copies of the same channel interact to fine-tune cellular regulation. The L (for long-lasting) channel requires significant depolarization for Ca^{2+} entry and inactivates slowly. Of the four members of the L-type channel family (CaV1.1–1.4), only CaV1.2 and CaV1.3 play a prominent role in the brain, the former in the hippocampus and the latter in the limbic system and striatum. L-Type calcium channels (LTCCs) do not participate directly in neurotransmitter release, but they play an important role in setting overall neuronal excitability; CaV1.2-based channels also influence gene transcription and gene expression.

The T (transient) channel is activated by small depolarizations and inactivates rapidly. T-type calcium channels may be involved in the action of the anticonvulsants ethosuximide, valproic acid, and divalproex in treating absence seizures. Calcium channels that are primarily found on central nervous system (CNS) neurons and participate in neurotransmitter release include N (neither L nor T) channels, P (Purkinje cell) channels, Q channels, and R (resistant to calcium channel blocker) channels. N channels are antagonized by the anticonvulsant levetiracetam and by ziconotide, a peptide toxin used to treat chronic pain, but not by calcium channel blocker drugs. The rapidly inactivating P channel is found in cerebellar Purkinje cells and is insensitive to a class (dihydropyridine) of calcium channel blockers.

L-Type calcium channel dysfunction resulting in increased calcium influx into cells in bipolar disorder could contribute to elevated $[Ca^{2+}]_B$, increasing resting neuronal excitability, as well as to increased intracellular Ca^{2+} storage with subsequent excessive elevations of $[Ca^{2+}]_i$ in response to stimulation. A polymorphism (rs1006737) of the CACNA1C gene, which encodes the α1c subunit of the L-type calcium channel $C_{av1.2}$, has consistently been associated in genome-wide association studies with bipolar disorder, especially early-onset bipolar disorder, as well as with vulnerability to depression and schizophrenia. In patients with bipolar disorder, the CACNA1C risk gene has been associated with crucial clinical aspects of the condition, such as altered circadian rhythms and deficits in executive functioning and cognition. Because this channel plays a central role in regulating gene transcription and the tonic excitatory drive, some cases of bipolar disorder could represent a channelopathy, similar to familial migraine and some seizure disorders.

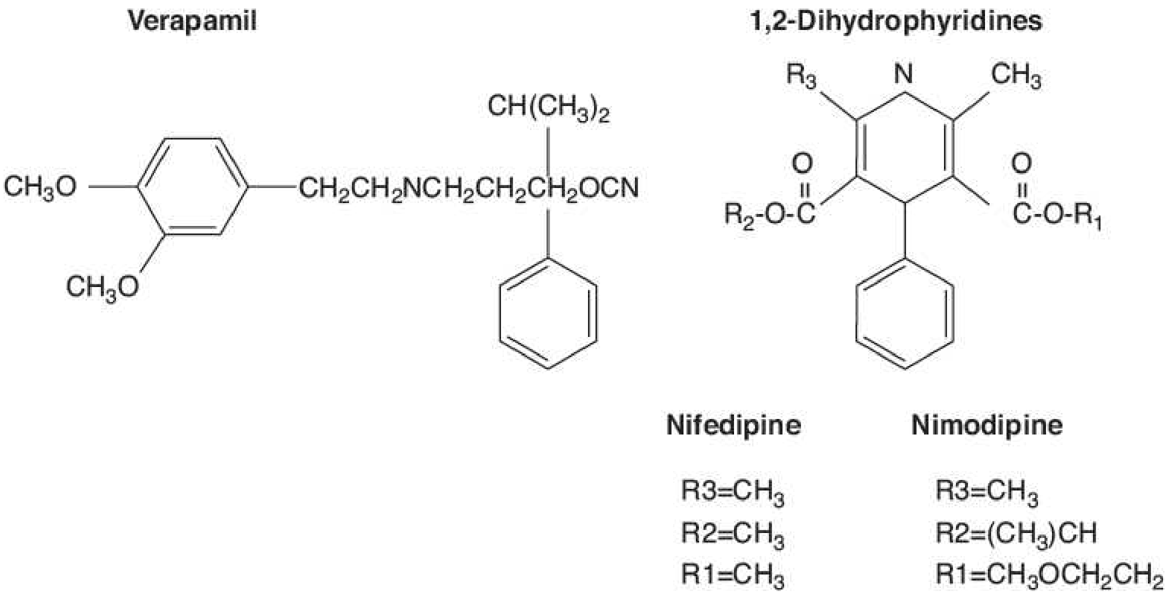
Verapamil
CH3O
CH3O
CH(CH3)2
CH2CH2NCH2CH2CH2OCN
1,2-Dihydrophyridines
R3
N
CH3
O
O
R2-O-C
C-O-R1
Nifedipine
R3=CH3
R2=CH3
R1=CH3
Nimodipine
R3=CH3
R2=(CH3)CH
R1=CH3OCH2CH2

FIGURE 33.10a–4. Structure of some CCBs.

CHEMISTRY

The first calcium channel inhibitor (calcium channel blocker or CCB) to be synthesized was verapamil, a derivative of papaverine (Fig. 33.10a–4). This medication was found to have applications in cardiovascular disease as a result of reduced excitation-contraction coupling caused by inhibition of calcium ion influx into cardiac cells. Subsequently, a total of 11 CCBs have been approved by the FDA, belonging to four chemical classes: phenylalkylamines (verapamil), benzothiazepines (diltiazem), dihydropyridines or DHPs (nifedipine, amlodipine, felodipine, isradipine, nicardipine, nisoldipine, nimodipine, and clevidipine), and diarylaminopropylamines (bepridil). Because of their heterogeneity of structure, binding site, and action, these medications are not interchangeable. However, they all have the capacity to reduce the excessive excitability of diverse cellular systems.

Verapamil and diltiazem are crystalline powders that are soluble in water and chloroform. Nifedipine is insoluble in water, but it is soluble in ethanol. Nimodipine, which is insoluble in water but is highly lipophilic, has two closely related congeners: isradipine and amlodipine. Amlodipine is a white crystalline powder that is slightly soluble in water and ethanol.

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Absorption and Distribution. The CCBs are nearly completely absorbed after oral administration, but bioavailability is decreased by extensive first-pass hepatic and probably intestinal metabolism. First-pass metabolism also results in considerable interindividual variability in blood level at a given dose and in oral bioavailability that is generally less than that of other drugs that are more frequently used to treat bipolar disorder. However, effective concentrations for their primary indications are an order of magnitude lower (Table 33.10a–2). Optimal serum concentrations for mood disorders or conditions other than cardiovascular disorders have not been investigated. The more lipophilic, uncharged nimodipine crosses the blood–brain barrier readily and is approved for the treatment of stroke. Verapamil and its active metabolite can be recovered from human cerebrospinal fluid (CSF) after oral administration, and the concentration of phenylalkylamines and other CCBs in the brain is sufficient to be protective after cerebral ischemia in animal models.

Elimination. Most CCBs have elimination half-lives in the range of 1.3 to 6.0 hours, requiring multiple daily doses, except for amlodipine, a once-daily preparation with a half-life of 35 to 50 hours. Because repeated administration saturates hepatic metabolizing enzymes, bioavailability and half-lives increase with chronic administration, and the medication can be given less frequently. For example, verapamil is administered four times per day initially, but after 1 to 2 months, the frequency of dosing can often be reduced to two to three times daily. Intestinal and hepatic metabolism of many CCBs by CYP450 3A4 results in increased bioavailability when CCBs are taken with CYP3A4 inhibitors such as grapefruit juice.

Pharmacodynamics

Mechanism of Action. Currently available CCBs bind to L channels (i.e., L-type channels or LTCCs), which are found in the cardiovascular, endocrine, and CNS and in muscle. Their actions, therefore, are not limited to the CNS. All CCBs inhibit calcium influx via activity-dependent binding to the α1 subunit of LTCCs. In addition to binding to L-channels, nimodipine, nicardipine, methoxyverapamil, flunarizine, and cinnarizine may also antagonize T-channels, and a phenylalkylamine-binding site exists on the inner mitochondrial membrane. Variations in CCB structure result in variations in potency at different L-channels in different tissues, and depending on the distribution of binding sites for a given chemical class of CCBs, that class may be more or less active in a particular tissue or region. For example, dihydropyridines are more selective for vascular tissue, whereas with more cardiac activity, verapamil and diltiazem have more prominent antiarrhythmic properties.

Table 33.10a–2.
Comparative Chemistry of Some Calcium Channel Blockers and Some Mood Stabilizers

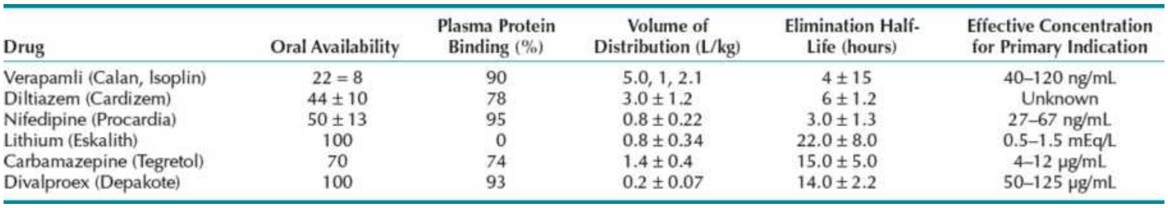

Drug	Oral Availability	Plasma Protein Binding (%)	Volume of Distribution (L/kg)	Elimination Half-Life (hours)	Effective Concentration for Primary Indication
Verapamli (Calan, Isoplin)	22 = 8	90	5.0, 1, 2.1	4 ± 15	40–120 ng/mL
Diltiazem (Cardizem)	44 ± 10	78	3.0 ± 1.2	6 ± 1.2	Unknown
Nifedipine (Procardia)	50 ± 13	95	0.8 ± 0.22	3.0 ± 1.3	27–67 ng/mL
Lithium (Eskalith)	100	0	0.8 ± 0.34	22.0 ± 8.0	0.5–1.5 mEq/L
Carbamazepine (Tegretol)	70	74	1.4 ± 0.4	15.0 ± 5.0	4–12 μg/mL
Divalproex (Depakote)	100	93	0.2 ± 0.07	14.0 ± 2.2	50–125 μg/mL

Along with allosteric regulation and the influence of circulating modulators, membrane potential determines the conformation of calcium channels and, therefore, the affinity of CCB-binding sites. CCB binding to L-channels is generally enhanced by depolarization and reduced by hyperpolarization, resulting in greater activity in hyperactive tissue, the CCB stabilizing the calcium channel in an inactivated state. Most DHPs are uncharged at physiologic pH and easily gain access to the receptor through the cell membrane if it is sufficiently depolarized. CCBs that are charged at physiologic pH, such as amlodipine, verapamil, and diltiazem, have slower onsets of action because of interactions with negatively charged phosphate head groups in the cell membrane. Because they gain access to the receptor through open channels, they are more active when channel openings are more frequent. The activity dependence of charged and uncharged CCBs makes these drugs more potent in hyperactive than in normally active tissue. Actions measured in cell preparations from normal subjects or animals, therefore, may not reflect activity in pathologic human states, a point exemplified by the finding that lithium attenuates elevated $[Ca^{2+}]_i$ in platelets of bipolar disorder patients but not in platelets of controls.

Although the finding of elevated free intracellular Ca^{2+} concentrations in bipolar disorder has been widely replicated and normalization of $[Ca^{2+}]_i$ has been shown to be associated with normalization of mood, there are several possible objections to the hypothesis that attenuation of excessive increases in $[Ca^{2+}]_i$ is directly relevant to mood stabilization or that any thymoleptic effect of CCBs is directly related to their action on intracellular Ca^{2+} signaling. For example, as noted previously, L-channels do not appear to be involved in the action potential-dependent release of neurotransmitters, and calcium channel subtypes that are specific to the CNS, such as N, P, Q and R-channels, are not affected by most L-type CCBs. However, L-channels are ubiquitous in the brain and are activated by neurotransmitters, making these channels an important component of neuronal interaction. Furthermore, it is not certain that neurotransmitter dysfunction is a primary element of the pathophysiology of mood disorders. Low concentrations of CCBs do not affect immediate neuronal Ca^{2+} currents in neurons containing L-channels because depolarization of these cells is normally too brief to permit sufficient CCB binding, but higher concentrations of CCBs may alter neuronal activity in other fundamental ways, and CCBs have greater effects on neurons with prolonged or marked depolarization. Reduction in tonic calcium influx may reduce the intracellular pool from which more immediate release results in rapid neuronal activation.

An alternative hypothesis is that the anticonvulsant property of some CCBs and not their effect on Ca^{2+} currents accounts for their activity in bipolar disorder. In favor of this possibility, prolonged elevations of $[Ca^{2+}]_i$ can contribute to the development and maintenance of seizure disorders by altering GABA receptor recycling, gene transcription, protein synthesis, and neuronal sprouting. In animal studies, nifedipine, diltiazem, and verapamil blocked the effect of subconvulsive doses of pentylenetetrazole (PTZ) to the same extent as valproate diltiazem enhanced the anticonvulsant effect of valproate, and amlodipine blocked PTZ-induced seizures in mice. Both verapamil and nimodipine inhibit amygdala-kindled seizures. Possibly as a result of T-channel antagonism, nimodipine has more of a potential for anticonvulsant actions. However, verapamil did not improve epilepsy control in phenobarbital-resistant canine epilepsy. The finding noted earlier that levetiracetam, which is clearly an anticonvulsant, does not normalize increased $[Ca^{2+}]_i$ in bipolar disorder and has little efficacy as a mood stabilizer tends to contradict the proposition that anticonvulsant properties are inherent to mood stabilization. Additional features of some CCBs that could be relevant to mood-stabilizing properties include antagonism of nicotinic cholinergic receptors and reduction of dopamine D_2 receptor signaling.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

The first published use of a calcium antagonist medication in psychiatry was a 1979 report of temporary reduction of agitation in three manic patients by subcutaneous injection of salmon calcitonin, which lowers $[Ca^{2+}]_i$ by driving Ca^{2+} into intracellular storage sites, but not placebo. Subsequently, the same group found that subcutaneous salmon calcitonin reduced agitation and manic behavior and improved sleep and Brief Psychiatric Rating Scale scores in 12 patients. Another approach to attenuating the impact of hyperactive intracellular Ca^{2+} signaling tamoxifen, an estrogen receptor antagonist used to prevent breast cancer recurrence that inhibits protein kinase C (PKC), had acute antimanic properties in small open and controlled Turkish trials.

In 1982, the first report emerged on the use of a CCB in the treatment of mania in a double-blind, placebo-controlled trial of verapamil in a single manic patient. The rationale for choosing this particular CCB was that it was the only CCB that could be obtained at a time at which none of these medications had yet been approved for use in the United States. Subsequently, case reports appeared of prevention by verapamil of antidepressant-induced hypomania and prevention of affective recurrence in bipolar disorder. Open trials of verapamil were effective in 4 of 7 manic patients, and in all patients with mania, ¾ of patients with mixed episodes, and about 1/3 of acutely depressed patients in a sample of 28 women.

Formal studies of CCBs have generally been small and relatively brief, with an emphasis on acute treatment of mania. In double-blind placebo-controlled trials, verapamil was superior to placebo in six manic patients and seven manic or schizoaffective-manic patients. Double-blind comparisons reported equivalent antimanic efficacy to lithium and clonidine. In open trials, the addition of verapamil or diltiazem improved the response to lithium in primary mania, but not mania secondary to CNS disease. A retrospective analysis of open administration of diltiazem suggested a benefit in treatment-resistant bipolar disorder, but no further investigation of this possibility has been published. A negative single-blind comparison of lithium and verapamil and a negative placebo-controlled double-blind trial of verapamil in mania were limited by verapamil doses that were lower and administered less frequently and for a shorter period of time than in positive trials. In the lithium comparison, patients treated with that medication were still very symptomatic at the end of the trial, and in the placebo-controlled study, patients only took verapamil for an average of 9.3 days, which would not be sufficient for established antimanic drug monotherapy.

Case reports have appeared of remission with nimodipine of chronic ultradian cycling refractory to all standard treatments in a 13-year-old patient and of the effectiveness of nimodipine for ultradian cycling after amygdalohippocampectomy. A 1-week open trial of nimodipine in six manic inpatients reported a significant response in all patients. Another open trial of nimodipine reported improvement in four manic patients, remission in one patient, and no response in two patients with mania secondary to CNS disease. In an 18-month open trial in 12 patients with rapidly cycling bipolar disorder, the combination of a low dose of nimodipine and lithium appeared to be significantly more effective than either medication alone in reducing recurrences of affective episodes, while anticycling effects of nimodipine were reported in two patients with rapid cycling when other treatments had failed. Careful longitudinal observations in small numbers of patients have suggested the possible usefulness of nimodipine for refractory recurrent mania and depression, ultradian cycling and recurrent brief depression, including in some patients with neurologic disorders. In double-blind studies using an off-on-off-on method, improvement has been noted with nimodipine, relapse on placebo, and then re-response on nimodipine. In doses sometimes exceeding the recommended maximum, several other small controlled studies have suggested the usefulness of nimodipine alone or combined with lithium, valproate, or carbamazepine in rapid and ultradian cycling bipolar mood disorders.

As is true of most putative mood stabilizers, the usefulness of CCBs in the treatment of depression is not as well studied as their potential applications in mania. Nifedipine reduces immobility in the behavioral despair test, which is believed to be an animal model of antidepressant activity; this effect is synergistic with antidepressants. Flunarizine was thought to reduce recurrences of bipolar depression in one patient, and one patient with recurrent brief depression responded to blind trials of nimodipine or verapamil. Improvement was reported in 27 of 30 patients with late-life depression when treated with nimodipine, and a significantly higher rate of improvement in patients with involutional depression occurred with nimodipine or amitryptiline versus placebo. Verapamil appeared to augment the antidepressant action of imipramine in a patient with unipolar depression who had previously responded to lithium augmentation; when imipramine was discontinued 6 months later, the patient remained euthymic on verapamil alone. An 8-month, double-blind placebo-controlled trial of the addition of nimodipine to fluoxetine in the treatment of major depression secondary to cerebrovascular disease (N = 101) found a number needed to treat of four both for remission and reduced risk of recurrence. In addition, a greater proportion of patients treated with fluoxetine-nimodipine (54% vs. 27%) exhibited full remission (P = 0.006). A Czech study found that overall, verapamil was no more effective than placebo, and was inferior to amitriptyline, in the treatment of depression, although a few patients did seem to have a genuine antidepressant response to verapamil. As the investigators did not differentiate between bipolar and unipolar depression, it is possible that, as would be expected with lithium, those with bipolar depression were more likely to respond to verapamil than were those with unipolar depression. Compared to placebo, treatment with nicardipine during ECT resulted in a better antidepressant response to ECT. In a single-blind pilot protocol involving 10 patients with bipolar depression generally of more than 6 months duration, acute treatment with isradipine resulted in significant improvement in six patients and remission in two. In open and controlled studies, nimodipine has appeared to be particularly useful in the treatment and prevention of both manic and depressive episodes. However, nifedipine was not effective for depression in a small study, and an animal study suggested that nifedipine interfered with the effects of antidepressants and electroconvulsive shock on the density of corticosteroid receptors.

The two negative antimanic trials of verapamil mentioned earlier led to conclusions by some reviewers that this and other CCBs are ineffective for bipolar disorder, a generalization that is not applied to contradictory experience with other medications with proposed antimanic and mood-stabilizing effects such as some anticonvulsants and atypical antipsychotic drugs. One factor may be the lack of industry sponsorship of larger CCB studies. In addition, studies supporting the thymoleptic properties of currently popular medications have relied on sample enrichment for patients with no comorbidities or complications who were shown a priori to be likely to respond to the experimental drug, while CCB studies have enrolled more heterogeneous samples of patients and have included subjects with complex and comorbid disorders. Since clinical testing of hypotheses of actions on intracellular processes has not been of as much interest as studies of neurotransmitter, receptor, and anticonvulsant hypotheses and development of “me too” products, conceptual as well as financial limitations on clinical research involving CCBs have limited the data available to reliably assess the place of CCBs in the treatment of bipolar disorder.

Bearing these limitations in mind, clinical experience suggests that verapamil is most likely to be useful for the treatment and prevention of mania and hypomania in patients who are responsive to lithium. It may be more likely to prevent bipolar depression recurrence than to treat it acutely. Nimodipine seems to have the potential for treating refractory and complex bipolar disorder, especially in combination with other thymoleptic medications. Nimodipine may be more useful acutely than verapamil for bipolar depression. However, larger controlled studies are necessary to assess its precise role. Very limited data suggest that isradipine, which is closely related to nimodipine, may have similar applications. Limited early research suggesting the possible benefit of verapamil, diltiazem, and nimodipine as adjuncts in the treatment of anxiety disorders has not been followed up in subsequent research.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Animal studies and clinical experience with verapamil in the treatment of fetal arrhythmias and premature labor, as well as with the prenatal treatment of hypertension with nicardipine and pre-eclampsia with isradipine, suggest relative safety of CCBs during pregnancy. A report of three patients for whom verapamil was effective for mania during pregnancy without any evidence of teratogenicity was amplified by a study of 28 pregnant women with acute episodes of mania or bipolar depression and 9 euthymic pregnant women who switched to verapamil maintenance from another mood stabilizer who had no complications of pregnancy and no teratogenicity, with a 100% response rate in mania, a 77% response rate in mixed states, and a 39% response rate in depression; six of eight patients who continued verapamil maintenance therapy remained well. A few case reports of similar outcomes with nimodipine during pregnancy have emerged. The number of reported cases is still too small to draw definitive conclusions about the use of nimodipine, verapamil, and isradipine in pregnancy, and there is virtually no information about other CCBs during pregnancy.

Concerns have been raised about an increased risk of reinfarction or mortality in elderly patients taking short-acting CCBs compared with long-acting agents following myocardial infarction (MI), possibly as a result of coronary vessel steal caused by diffuse vasodilatation. Such assumptions have been limited by the lack of control for the severity of hypertension. CCBs are usually used as second-line treatments for more severe hypertension, raising the possibility that patients treated with CCBs had worse cardiovascular outcomes because their hypertension was more severe. Furthermore, CCBs are known to reduce the risk of ventricular fibrillation after MI. Amlodipine and extended-release nifedipine were found not to increase the risk of acute cardiac events in older patients, and the large Danish Verapamil Infarction Trial (DAVIT) found that CCBs reduced the risk of reinfarction when used prophylactically after an index MI.

Arguing that CCBs could interfere with apoptosis in patients prone to cancer, early research suggested a slightly increased risk of cancer in patients taking CCBs. In support of this speculation, a meta-analysis suggested a statistically significant association between the use of CCBs and breast cancer but a decreased risk of prostate cancer nine years later. A review of observational studies suggested an increased risk of lung cancer with CCB use, although the lower limit of the 95% CI was only 1.01. However, uncontrolled cell growth is also calcium-dependent, and in studies suggesting increased cancer risk, antihypertensive use was assessed prior to the onset of cancer, without data that might inform whether these medications were being taken at the time when carcinogenesis was initiated or even whether CCBs were still being taken when cancer was diagnosed. Other cancer risk factors such as smoking, obesity and diet also were not controlled in these studies. A meta-analysis found no increased risk of breast cancer with long-term CCB use, and another large meta-analysis of trials involving a total of more than 334,000 subjects also found no increased risk of cancer with CCB use. A 5-year prospective RCT found no differences in cancer risk in older hypertensive patients treated with CCBs versus other antihypertensives, and the DAVIT trial found no increased risk of cancer in 10,474 patient-years experience in preventing MI. Furthermore, a mechanism of resistance to cancer chemotherapy is the induction of the MDR1 gene, which codes for Pglycoprotein, a drug extrusion protein that removes chemotherapy drugs from cancer cells. Verapamil, which inhibits P-glycoprotein, can reduce chemotherapy resistance. Since cell proliferation is a calciumdependent process, CCBs can inhibit tumor cell growth, and the ability of T-channels to generate rapid calcium currents at resting membrane potentials may account for findings that experimental T-channel antagonists reduce cancer cell proliferation in vitro.

THERAPEUTIC INDICATIONS

Because of their capacity to reduce cellular hyperactivity without suppressing normally active cells, CCBs have been used to treat angina, hypertension, Raynaud phenomenon, esophageal spasm, premature labor, and epilepsy. Verapamil and diltiazem, but not the DHPs, have class IV antiarrhythmic activity and are used to treat supraventricular arrhythmias. Verapamil can reduce migraine recurrence.

In view of the role of excessive elevations of $[Ca^{2+}]_i$ in neuronal apoptosis, nimodipine has the potential to reduce the progression of brain damage in neurodegenerative disease. Nimodipine is approved for the reduction of the risk of extension of neuronal damage by vasospasm following hemorrhaging stroke, and it has been reported to improve cognition following ischemic stroke. In an early study that randomized 87 elderly patients with cognitive decline to nimodipine and 88 patients to placebo for 12 weeks, there were more favorable changes in all cognitive measures and in depression with nimodipine. The drug was well tolerated and equally effective for “primary degenerative” and vascular dementia. There was a greater improvement in cognition after 3 months of treatment with nimodipine in 16 patients with mild cognitive impairment (MCI) compared with 15 controls. In a randomized trial in 230 patients (mean age 75) with vascular dementia to nimodipine or placebo, nimodipine was associated with better lexical production and Mini-Mental State Exam scores and had less deterioration and behavioral disturbances requiring intervention, as well as fewer dropouts and adverse events, than placebo. A randomized controlled study found nimodipine to be superior to placebo in slowing the deterioration of Alzheimer disease. A 2002 Cochrane review concluded that nimodipine “can be of some benefit” in dementia of various causes. Results of multicenter clinical trials of nimodipine for dementia have not been published. Verapamil has been reported to ameliorate tardive dyskinesia.

PRECAUTIONS AND ADVERSE REACTIONS

The most common CCB side effects involve consequences of excessive vasodilatation, such as dizziness, postural hypotension, headache, skin flushing, tachycardia, nausea, lightheadedness, weakness, and digital dysesthesia; aggravation of myocardial ischemia may rarely occur because of coronary steal related to vasodilatation elsewhere. Precapillary dilatation with reflex postcapillary constriction can increase capillary hydrostatic pressure and cause peripheral edema. Verapamil and diltiazem are more likely than other preparations to produce sinus bradycardia and atrioventricular block, which may aggravate heart failure. Coughing, shortness of breath, wheezing, rashes, somnolence, constipation, seizures, and, rarely, depressed level of consciousness or psychosis may be caused by the CCBs. Possibly because of inhibition of calcium signaling in postsynaptic dopamine receptors, akathisia, parkinsonism, and delirium have occasionally been associated with the use of calcium channel antagonists. Nimodipine sometimes causes bad dreams. Toxic serum concentrations of CCBs have not been defined. Possibly because of compensatory changes in calcium channels or CCB receptors, tachyphylaxis and withdrawal do not occur. CCBs do not alter exercise tolerance. Treatment of severe toxicity is with calcium, glucagon, volume expansion, and vasopressors.

Although inhibition of hyperactive intracellular calcium activity may reduce excitotoxicity in adult neurons, postnatal brain development depends on periodic increased intracellular calcium ion signaling, and animal studies suggest that neonatal (but not later) neurons are highly sensitive to reductions in intracellular calcium by CCBs, which can damage these neurons. The use of CCBs in very young children, therefore, requires substantial caution.

Drug Interactions

Adding CCBs to β-adrenergic blocking agents can result in additive cardiac slowing and atrioventricular block, whereas additive effects with α-adrenergic blocking agents may produce hypotension. Verapamil and nitrendipine increase plasma concentrations of digoxin and cause bradycardia, hypotension, or atrioventricular block when combined with quinidine. Because verapamil and nifedipine inhibit cytochrome P450 CYP3A4, they raise serum concentrations of CYP 3A4 substrates, such as carbamazepine and the triazolobenzodiazepines, which can result in carbamazepine neurotoxicity. Neurotoxicity also occurs occasionally when verapamil is combined with lithium or nifedipine due to inhibition of lithium clearance. Additive suppression of the spontaneous rate of depolarization of the sinus node and of conduction through the atrioventricular node can increase the risk of bradycardia when lithium is combined with verapamil.

Laboratory Interferences

As mentioned previously, verapamil and nifedipine raise serum levels of carbamazepine, digoxin, and other CYP 3.44 substrates. CCBs do not primarily affect other laboratory tests.

DOSAGE AND ADMINISTRATION

Verapamil is available in 40-, 80-, and 120-mg tablets. A 120-mg sustained-release preparation used to treat hypertension was effective for mania in one report of three cases, but it generally is not as reliably effective for mood disorders as the immediate release forms, possibly because higher peak levels are necessary for this indication. The usual daily doses of verapamil in the treatment of mania have been between 240 and 480 mg per day, although some patients may need higher doses because of low bioavailability resulting from extensive intestinal and hepatic first-pass metabolism. The short elimination half-life of verapamil requires three to four times per day dosing. Nimodipine is available as a 30-mg capsule and a 60 mg/20 mL solution and is administered in doses of 90 mg per day for dementia and stroke. However, doses for bipolar illness have ranged from 180 mg to as much as 720 mg per day, the latter dose being substantially greater than the maximum dose approved by the FDA. A frequency of dosing less than four times daily does not result in consistent levels, although less frequent dosing is sometimes effective. Isradipine, which is available as a tablet or capsule, is available in 2.5- and 5.0-mg immediate-release capsules and 5.0- and 10.0-mg extended-release tablets and is administered in doses of 2.5 to 20.0 mg per day in a frequently divided dose for hypertension. Amlodipine, the only CCB that is effective when it is administered once daily, is available in 2.5-, 5.0-, and 10.0-mg tablets and a 1 mg/mL oral suspension; the usual dose for hypertension and coronary heart disease is 5 to 10 mg per day. There is little experience with this medication in mood disorders.

INVESTIGATIONAL DRUGS AND PREDICTION OF RESPONSE

Research into psychiatric applications of CCBs has been limited by the relative lack of interest by manufacturers in developing and marketing them for such indications. Nimodipine and related dihydropyridine CCBs will probably turn out to have greater potential in psychiatric disorders than verapamil because they cross the blood–brain barrier more readily, and they act on T- as well as L-type channels. Future development of CCBs for psychiatric indications might profitably focus on selective antagonists of channels such as N-, P-, Q- and R-type channels that are primarily expressed in the CNS and would not be limited by cardiovascular and other systemic side effects, although the N-type channel blocker levetiracetam does not alter platelet $[Ca^{2+}]_i$ and is not effective for bipolar disorder. Further investigation of medications that antagonize other aspects of intracellular calcium ion dynamics, such as the estrogen receptor antagonist and PKC inhibitor tamoxifen, may also prove fruitful.

No mood stabilizer is effective for all patients, but few validated predictors have emerged of which medication may be most likely to be useful for which patients. In the case of CCBs, a few preliminary studies may provide future directions for research into predictors of response. A 2-week trial in which manic patients received 360 mg per day of verapamil found a 61% reduction in manic symptoms and a correlation between the degree of improvement of mania and increased plasma calcium levels. Since changes in extracellular calcium concentration are easy to measure, further investigation of a possible relationship between changes in plasma calcium levels and later response to a CCB may be justified. Similarly, a correlation was suggested between clinical response and the degree to which in vitro incubation of lithium with platelets of patients with bipolar disorder reduced elevated $[Ca^{2+}]_i$. Assessing the predictive value of incubation of a CCB with platelets could clarify whether reduction of

$[Ca^{2+}]_i$ might predict response to a CCB. Using this time-consuming and difficult assay, which is most valid when compared with a control, is not likely to be practical in clinical practice, but it could be used as a proxy for simpler and cheaper measures of potential response to a CCB. A positron emission tomography (PET) study finding that nimodipine responders had left insular hypometabolism compared with nonresponders, whereas carbamazepine responders had the opposite finding, may inform studies of neuroimaging changes as possible predictors of CCB response, either by themselves or as proxies to be used to study other possible laboratory predictors. Most of these lines of investigation have not been followed up due to a general lack of interest in psychiatry in the use of CCBs.

A more recent finding that suggests another line of research into possible predictors of CCB response is a study of healthy subjects reporting greater normalization by nimodipine of frontal metabolic activity measured by PET during a working memory task in those with the rs1006737 CACNA1C bipolar risk gene compared with those with a different allele. There was a "gene-dose effect" for the peak cluster of nimodipine action on working memory. Subjects with the AA genotype demonstrated a 62.32% decrease in BOLD activation with nimodipine compared to 37.42% for GA carriers and 16.05% for GG carriers. Since CCB binding is increased in hyperactive channels, such findings support the hypothesis that nimodipine may induce more blockade of a hyperactive CACNA1C variant that is linked to bipolar disorder than other CACNA1C genotypes. Some investigators, therefore, propose examining CACNA1C genotype as a predictor of CCB- or at least nimodipine-response. Imaging studies of the binding of labeled CCBs to central calcium channels may also turn out to predict clinical response.

The scatter in $[Ca^{2+}]_i$ in bipolar disorder patients (Fig. 33.10a–2) is one observable marker of the biologic heterogeneity of the disorder. Preliminary observations that patients with elevated platelet $[Ca^{2+}]_i$ that normalized with incubation with lithium may inform prospective studies of the use of this marker to predict CCB response. The heterogeneity of calcium channels and of CCB-binding sites may also provide for different spectra of action of different CCBs. Just as patients with one bipolar variant may do better with CCBs than those with another variant, it is possible that patients who do not respond to one CCB class may have a good response to another. For example, the rs1006737 variant of the CACNA1C gene may predict a preferential response to nimodipine or some other CCB. Combinations of CCBs with each other and with other antimanic drugs could have additive effects in treatment-resistant bipolar disorder. In view of the limited benefit from medications that are now commonly used as mood stabilizers and the tolerability of CCBs, more experience with this novel class of medications seems justified.

Further Readings

Berger SM, Bartsch D. The role of L-type voltage-gated calcium channels Cav1.2 and Cav1.3 in normal and pathological brain function. *Cell Tissue Res*. 2014;357:463–476.

Berridge MJ. Calcium signalling and psychiatric disease: bipolar disorder and schizophrenia. *Cell Tissue Res*. 2014;357:477–492.

Boccia ML, Reite ML, Kaemingk K, Held P, Laudenslager ML. Behavioral and autonomic responses to peer separation in pigtail macaque monkey infants. *Dev Psychobiol*. 1989;22:447–461.

Carman JSA, Wyatt RJ. Calcium: pacesetting the periodic psychoses. *Am J Psychiatry*. 1979;136:1035–1039.

Croarkin PE, Luby JL, Cercy K, et al. Genetic risk score analysis in early-onset bipolar disorder. *J Clin Psychiatry*. 2017;78(9):1337–1343.

Dubovsky SL, Christiano J, Daniell LC. Increased platelet intracellular calcium concentration in patients with bipolar affective disorders. *Arch Gen Psychiatry*. 1989;46:632–638.

Dubovsky SL, Daurignac E, Leonard KE, Serotte JC. Levetiracetam, calcium antagonism, and bipolar disorder. *J Clin Psychopharmacology*. 2015;35:422–427.

Dubovsky SL, Daurignac E, Leonard KE. Increased platelet intracellular calcium ion concentration is specific to bipolar disorder. *J Affect Disord*. 2014;164:38–42.

Dubovsky SL. Intracellular calcium ion signaling in bipolar mood disorders: new data and directions. *Internal Medicine Review*. 2019;5(2):1–18.

Dubovsky SL, Lee C, Christiano J. Lithium decreases platelet intracellular calcium ion concentrations in bipolar patients. *Lithium*. 1991;2:167–174.

Dubovsky SL, Franks RD, Allen S, Murphy J. Calcium antagonists in mania: A double-blind study of verapamil. *Psychiatry Res*. 1986;18:309–320.

Dubovsky SL. Why don't we hear more about the calcium antagonists? An industry-academia interaction. *Biol Psychiatry*. 1994;35:149–150.

Ferreira MA, O'Donovan MD, Meng YA, et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet*. 2008;40(9):1056–1058.

Garza-Trevino ES, Overall JE, Hollister LE. Verapamil versus lithium in acute mania. *Am J Psychiatry*. 1992;149:121–122.

Green EK, Grozeva D, Jones IR, et al. The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Mol Psychiatry*. 2010;15:1016–1022.

Hahn CG, Gomez G, Restrepo D, et al. Aberrant intracellular calcium signaling in olfactory neurons from patients with bipolar disorder. *Am J Psychiatry*. 2005;162:616–618.

Hoschl C, Kozeny J. Verapamil in affective disorders: a controlled, double-blind study. *Biol Psychiatry*. 1989;25:128–140.

Janicak PG, Sharma RP, Pandey G, Davis JM. Verapamil for the treatment of acute mania: a double-blind, placebo-controlled trial. *Am J Psychiatry*. 1998;155:972–973.

Klein E, Uhde TW. Controlled study of verapamil for treatment of panic disorder. *Am J Psychiatry*. 1988;145(4):431–434.

McCarthy MJ, Le Roux MJ, Wei H, Beesley S, Kelsoe JR, Welsh DK. Calcium channel genes associated with bipolar disorder modulate lithium's amplification of circadian rhythms. *Neuropharmacology*. 2016;101:439–448.

Pazzaglia PJ, Post RM, Ketter TA, et al. Nimodipine monotherapy and carbamazepine augmentation in patients with refractory recurrent affective illness. *J Clin Psychopharmacol*. 1998;18:404–413.

Sajadieh A, Hansen JF, Mortensen LS. Temporal pattern of the effect of verapamil on myocardial reinfarction. DAVIT Study Group. Danish Verapamil Infarction Trial. *Cardiovasc Drugs Ther*. 1998;12(4):405–408.

Soeiro-de-Souza MG, Dias VV, Figueira ML, et al. Translating neurotrophic and cellular plasticity: from pathophysiology to improved therapeutics for bipolar disorder. *Acta Psychiatr Scand*. 2012;126:332–341.

Thakur AA, Wang X, Garcia-Betancourt MM, Forse RA. Calcium channel blockers and the incidence of breast and prostate cancer: A meta-analysis. *J Clin Pharm Ther*. 2018;43(4):519–529.

Turner CP, Miller R, Smith C, et al. Widespread neonatal brain damage following calcium channel blockade. *Dev Neurosci*. 2007;29(3):213–231.

Wang X, Li Y, Fan GF, Zhang TY, Sun B, Fan PS. Effect of verapamil in the reversal of doxorubicin chemotherapy resistance in advanced gastric cancer. *Eur Rev Med Pharmacol Sci*. 2020;24(14):7753–7763.

Wisner KL, Peindl KS, Perel JM, Hanusa BH, Piontek CM, Baab S. Verapamil treatment for women with bipolar disorder. *Biol Psychiatry*. 2002;51:745–752.

Zink CF, Giegerich M, Prettyman GE, et al. Nimodipine improves cortical efficiency during working memory in healthy subjects. *Transl Psychiatry*. 2020;10:372.

33.10b Gabapentin, Topiramate, Tiagabine, Levetiracetam, Zonisamide, and Pregabalin

Hiroyuki Uchida, M.D., Ph.D.

Newer drugs for epilepsy, marketed over the past three decades, have generally been developed to have enhanced tolerability, simpler kinetics, and fewer drug–drug interactions. The chemical structures (Fig. 33.10b–1) and pharmacodynamic effects of these drugs are diverse. Some purported mechanisms of action include enhancing neuronal inhibition by increasing GABAergic function or decreasing neuronal excitation by decreasing glutamatergic function via blockade of AMPA receptors. These agents have potential psychotropic benefits, and these amino acid neurotransmitters have been implicated in psychiatric disorders. Thus, newer drugs for epilepsy appear to have distinctive efficacy profiles for the treatment of epilepsy and diverse psychotropic and analgesic effects (Table 33.10b–1). A few of these agents can provide psychotropic benefits and have received Food and Drug Administration (FDA) indications for acute mania (valproate and carbamazepine) and bipolar maintenance treatment (lamotrigine). To address this discrepancy between drug categories and indications, the use of Neuroscience-Based Nomenclature (NbN) is encouraged.

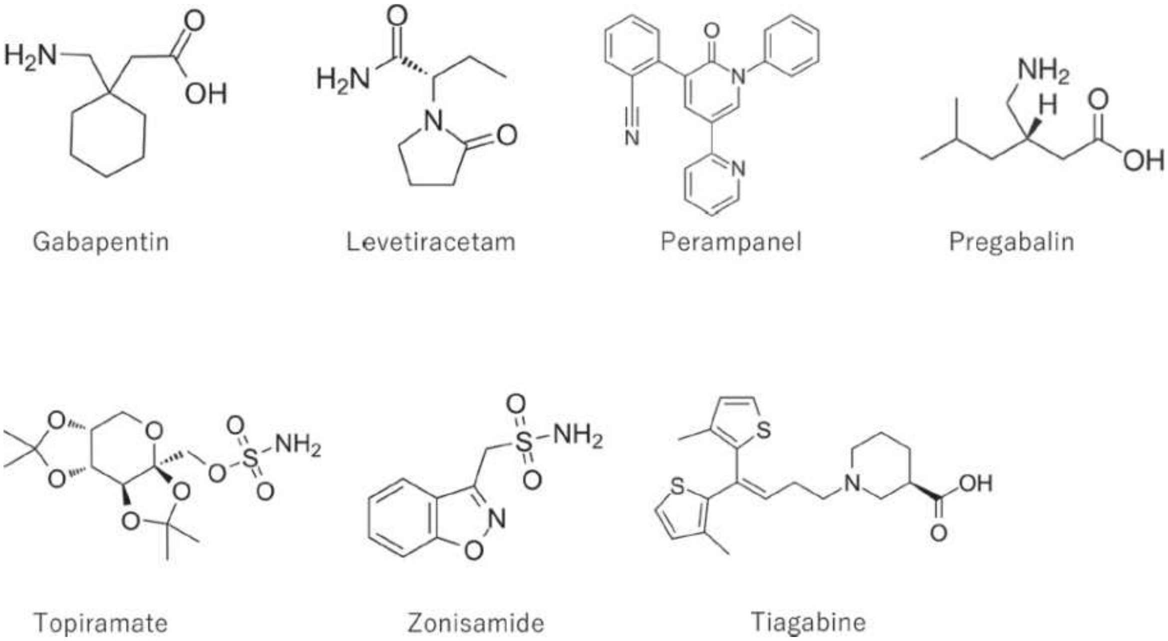
Gabapentin Levetiracetam Perampanel Pregabalin

Topiramate Zonisamide Tiagabine

FIGURE 33.10b–1. Chemical structures of selected glutamate and GABA modulators.

In 2008, the FDA announced a class warning regarding the increased risk of suicidality (suicidal behavior or ideation) in patients with epilepsy as well as psychiatric disorders based on data from diverse "anticonvulsants," six of which are described in detail in this chapter (gabapentin [Neurontin], topiramate [Topamax], tiagabine [Gabitril], levetiracetam [Keppra], zonisamide [Zonegran], and pregabalin [Lyrica]), and five of which are not described in this chapter (carbamazepine [Tegretol], felbamate [Felbatol], lamotrigine [Lamictal], oxcarbazepine [Trileptal], and valproate [Depakote]). This warning has been applied to subsequent anticonvulsants, including perampanel [Fycompa].

GABAPENTIN

Introduction and History

Gabapentin (Neurontin), which has a biochemical structure similar to the inhibitory amino acid neurotransmitter GABA, was developed by Warner-Lambert to create an agent for the treatment of spasticity. Gabapentin proved to be more effective as an anticonvulsant than an antispasticity agent. Gabapentin was found to be effective in several animal models of epilepsy, so it was developed as an antiepileptic drug and marketed for the adjunctive treatment of partial seizures in the United States in 1994 by Parke-Davis.

Due to its generally good tolerability, fewer clinically significant drug–drug interactions than older drugs for epilepsy, and initial, postmarketing, open, uncontrolled clinical experience suggesting efficacy in various neurologic and psychiatric conditions, gabapentin has become increasingly used in conditions other than seizure disorders. The use of gabapentin for other nonepileptic conditions substantially exceeded that in epilepsy. Double-blind randomized placebo-controlled trials later confirmed the efficacy of gabapentin for some, but not all, neurologic and psychiatric conditions suggested in open studies. Specifically, gabapentin appears to be effective in neuropathic pain and is marketed for the management of postherpetic neuralgia in adults. Gabapentin also appeared to be effective in anxiety disorders (social phobia and panic disorder) but not as the main intervention in mania or treatment-resistant mood disorders. Thus, gabapentin has emerged as a safe and effective agent in epilepsy and postherpetic neuralgia and has more widespread use adjunctively to relieve symptoms and pain and improve anxiety and insomnia in a dimensional fashion across diverse neurologic and psychiatric conditions. Moreover, gabapentin enacarbil (a prodrug of gabapentin) was found to be effective against moderate to severe restless legs syndrome. Finally, recent controlled data suggested that gabapentin could be effective for relapse prevention in people with alcohol dependence.

Table 33.10b–1.
Evidence-Based Uses of Selected Glutamate and GABA Modulators

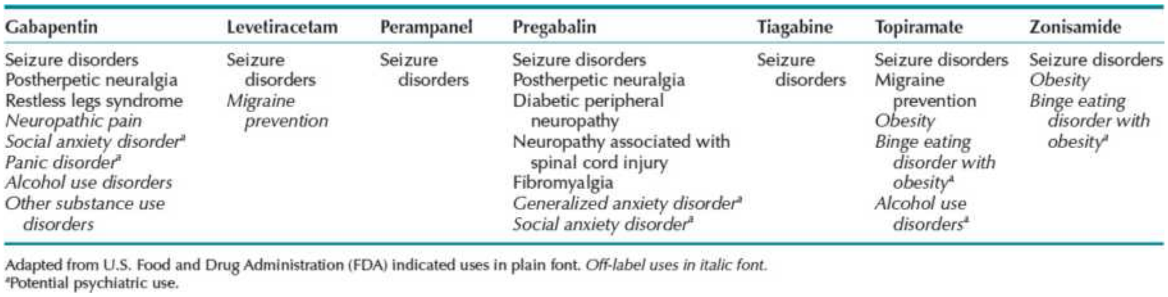

Gabapentin	Levetiracetam	Perampanel	Pregabalin	Tiagabine	Topiramate	Zonisamide
Seizure disorders Postherpetic neuralgia Restless legs syndrome *Neuropathic pain* *Social anxiety disorder*[a] *Panic disorder*[a] *Alcohol use disorders* *Other substance use disorders*	Seizure disorders *Migraine prevention*	Seizure disorders	Seizure disorders Postherpetic neuralgia Diabetic peripheral neuropathy Neuropathy associated with spinal cord injury Fibromyalgia *Generalized anxiety disorder*[a] *Social anxiety disorder*[a]	Seizure disorders	Seizure disorders Migraine prevention *Obesity* *Binge eating disorder with obesity*[a] *Alcohol use disorders*[a]	Seizure disorders *Obesity* *Binge eating disorder with obesity*[a]

Adapted from U.S. Food and Drug Administration (FDA) indicated uses in plain font. *Off-label uses in italic font.*

[a] Potential psychiatric use.

Chemistry

Gabapentin, 1-(aminomethyl)-cyclohexane acetic acid, is a novel amino acid derived from the addition of a cyclohexyl group to GABA (Fig. 33.10b–1). Immediate-release gabapentin preparations currently marketed in the United States include 100-, 300-, and 400-mg capsules; 600- and 800-mg tablets; and a 250 mg/5 mL oral solution. Generic immediate-release gabapentin is available in comparable formulations. Extended-release gabapentin and gabapentin enacarbil (a prodrug of gabapentin) are available in 300- and 600-mg tablets.

Pharmacologic Action

Pharmacokinetics. Gabapentin is available in immediate-release (Neurontin) and sustained-release (Gralise) formulations. In addition, gabapentin enacarbil is available in an extended-release formulation (Horizant). Immediate-release gabapentin has sublinear kinetics due to saturable gastrointestinal absorption, with oral bioavailability of approximately 60% at 300 mg three times a day, which decreases to 33% with 1,200 mg twice a day. In practice, the maximum individual doses of immediate-release gabapentin should not exceed 1,200 mg. Sustained-release gabapentin can be titrated up to 1,800 mg once daily with an evening meal. Gabapentin enacarbil was designed to have high-capacity absorption compared to gabapentin, even at doses greater than 1,200 mg, but the recommended dose is only 600 mg once or twice daily. Sustained-release gabapentin and gabapentin enacarbil are not interchangeable with other gabapentin products because of differing pharmacokinetic profiles that affect the frequency of administration. Food only modestly affects the rate and extent of immediate-release gabapentin absorption but more significantly affects the absorption of sustained-release gabapentin and gabapentin enacarbil, which needs to be taken with food. Gabapentin circulates in the blood largely unbound, with less than 3% bound to plasma proteins, and has a moderate volume of distribution of approximately 1 L/kg. Immediate-release gabapentin has a half-life of approximately 6 hours and is thus administered in three divided doses in patients with epilepsy. Sustained-release gabapentin and gabapentin enacarbil have an elimination half-life of approximately 6 hours, but due to differential absorption and indications, they are dosed once daily and once to twice daily, respectively. Gabapentin is not appreciably metabolized in humans, is eliminated unchanged by renal excretion, and can be removed by hemodialysis. Gabapentin clearance correlates with that of creatinine. Therefore, dosage adjustments are required in patients with impaired renal function. Clearance may be increased in children under the age of 5 years and decreased in the older adults, requiring dosage adjustments to be made.

Pharmacodynamics. Gabapentin has multiple biochemical actions, and the relationship between these and direct clinical effects remains to be established. Although gabapentin is a structural analog of GABA, it appears to lack direct GABAergic actions. Gabapentin appears to have indirect GABAergic effects, as both acute and chronic gabapentin increases cerebral GABA levels in humans. Gabapentin may also have antiglutamatergic actions, mediated by the inhibition of glutamate synthesis by branched-chain amino acid aminotransferase. It also modulates (but does not directly block) sodium channels and increases human whole-blood serotonin concentrations. Gabapentin has high-affinity regional (including neocortical and hippocampal) binding to $\alpha_2\delta$ subunits of voltage-activated calcium channels and thus may modulate calcium channels to reduce monoamine release. Several of these biochemical mechanisms may underlie the psychotropic effects of gabapentin.

Clinical Drug Studies

Seizure Disorders. Adjunctive immediate-release gabapentin was more effective than an adjunctive placebo in double-blind, randomized placebo-controlled trials involving adults and adolescents aged 12 years or older with refractory partial seizures and in one trial in children aged three to 12 years with refractory partial seizures.

Postherpetic Neuralgia. Gabapentin appeared to be more effective than a placebo in double-blind, randomized placebo-controlled trials in adults with postherpetic neuralgia, relieving pain and sleep interference and improving mood and quality of life. Moreover, gabapentin enacarbil appeared to be more effective than a placebo in several double-blind, randomized placebo-controlled trials in adults with postherpetic neuralgia. In contrast, one double-blind, randomized placebo-controlled trial found no significant relief from acute herpetic pain or prevention of postherpetic neuralgia with additional gabapentin to the usual treatment of herpes zoster within 72 hours of rash onset.

Other Pain Syndromes. Gabapentin appears to have analgesic effects in other pain syndromes, particularly in neuropathic pain. Thus, immediate-release gabapentin was more effective than placebo in relieving pain and, in some instances, yielded benefits on other measures such as sleep, mood, and quality of life in diabetic neuropathy, trigeminal neuralgia, neuropathic cancer pain when added to opioids, Guillain–Barré syndrome, fibromyalgia, painful human immunodeficiency virus (HIV)-associated sensory neuropathy, and diverse neuropathic pain. Immediate-release gabapentin was also more effective than a placebo in relieving post-amputation pain (in one of two studies), post-radical mastectomy pain, post-lumbar discectomy pain (when added to fentanyl but not when added to morphine), post-abdominal hysterectomy pain (when added to rofecoxib), and catheter-related bladder discomfort and postoperative pain in patients with percutaneous nephrolithotomy. Immediate-release gabapentin was more effective than placebo in chronic masticatory myalgia, complex regional pain syndrome type one (formerly known as reflex sympathetic dystrophy), and in suppressing nausea and vomiting and postoperative rescue analgesic requirements in laparoscopic cholecystectomy patients. Compared to a placebo, immediate-release gabapentin reduced migraine frequency and intensity in patients with migraines, relieved symptoms, and improved function and quality of life in patients with chronic daily headaches. It may attenuate hypobaric hypoxia-induced headache (a syndrome commonly experienced by newcomers to high-altitude areas). However, immediate-release gabapentin was no better than a placebo for chemotherapy-induced peripheral neuropathy.

Restless Legs Syndrome. Immediate-release gabapentin appeared to be more effective than a placebo in a small crossover design study of adults, reducing periodic leg movements, pain, sleep quality, sleep architecture, and overall clinical impression, and in another small crossover design study of adult hemodialysis patients. Gabapentin enacarbil was more effective than a placebo in double-blind, randomized placebo-controlled trials in adults with restless legs syndrome, relieving restless leg movements and improving sleep quality.

Other Neurologic Disorders. Immediate-release gabapentin compared to a placebo relieved essential tremor in two of three studies, primary orthostatic tremor in one study, and impairment of spasticity in multiple sclerosis in one study. Moreover, compared to a placebo, immediate-release gabapentin attenuated rigidity, bradykinesia, and tremor in Parkinson disease. Immediate-release gabapentin compared to no treatment improved lower, but not upper, extremity function in patients with spinal muscular atrophy. However, compared to a placebo, immediate-release gabapentin failed to slow the decline of arm strength in patients with amyotrophic lateral sclerosis in two studies and failed to increase muscle strength in patients with spinal muscular atrophy.

Alcohol and Substance Use Disorders. Data from numerous double-blind, randomized placebo-controlled trials on alcohol use disorders have found immediate-release gabapentin efficacious as monotherapy or in combination with either flumazenil or naltrexone, specifically associated with more abstinence days and longer time to relapse. Efficacy in opiate use disorders was found in two double-blind randomized placebo-controlled add-on (to methadone or buprenorphine) studies with less opiate use and fewer withdrawal symptoms, but not in one study of monotherapy with gabapentin. The efficacy of stimulant (methamphetamine or cocaine) use disorders is not well established for gabapentin, where only one double-blind, randomized placebo-controlled trial with immediate-release gabapentin to flumazenil had less drug use and cravings. However, three controlled studies (monotherapy and add-on) found no difference from placebo. One double-blind, randomized placebo-controlled trial of immediate-release gabapentin monotherapy for marijuana use disorder found less marijuana use and improved executive functioning on neuropsychological testing. Thus, gabapentin has the strongest evidence of efficacy for alcohol use disorders, with less robust evidence in other substances of abuse.

Anxiety Disorders. Controlled studies have confirmed immediate-release gabapentin utility in some anxiety disorders, with double-blind randomized placebo-controlled trials demonstrating efficacy in social phobia (social anxiety disorder), and in a post hoc analysis, in moderate to severe panic disorder.

Mood Disorders. Well-controlled studies found that immediate-release gabapentin was ineffective, despite early, open, uncontrolled reports suggesting that adjunctive immediate-release gabapentin was well tolerated and may offer benefits in patients with refractory mood disorders. In a double-blind, randomized placebo-controlled trial of patients with bipolar disorder suffering from mania or hypomania, while taking lithium, valproate, or lithium and valproate, adjunctive placebo tended to lower mania scores better than adjunctive immediate-release gabapentin. In treatment-resistant (primarily rapid-cycling bipolar) mood disorders, immediate-release gabapentin monotherapy was no more effective than a placebo, while lamotrigine monotherapy was more effective than both the immediate-release gabapentin and placebo. However, a small randomized clinical trial suggested that adjunctive immediate-release gabapentin compared to an adjunctive placebo might provide some prophylactic benefit in patients with bipolar disorder. Although the antidepressant efficacy of adjunctive immediate-release gabapentin in patients with bipolar disorder has been suggested in several open trials, controlled studies have not supported the primary use of gabapentin for mood disorders.

Therapeutic Indications

Seizure Disorders. Immediate-release gabapentin is indicated as an adjunctive treatment of partial seizures with and without secondary generalization in patients with epilepsy aged >12 years and as an adjunctive treatment of partial seizures in pediatric patients aged 3 to 12 years.

Postherpetic Neuralgia. Immediate- and extended-release gabapentin and gabapentin enacarbil are indicated for the management of postherpetic neuralgia in adults.

Restless Legs Syndrome. Gabapentin enacarbil is indicated for the treatment of moderate to severe restless legs syndrome in adults.

Precautions and Adverse Reactions

In general, the adverse effects of gabapentin are mild to moderate, can be attenuated by dosage rescheduling or decreases, and do not typically require discontinuation. Common adverse effects of adjunctive immediate-release gabapentin in adult patients with epilepsy include somnolence, dizziness, ataxia, fatigue, and nystagmus. Common adverse effects of immediate-release gabapentin in adult patients with postherpetic neuralgia are dizziness, somnolence, and peripheral edema. In pediatric (age 3 to 12 years) patients with epilepsy, the most common adverse effects of adjunctive gabapentin are viral infection, fever, nausea or vomiting, somnolence, and hostility. Interestingly, psychiatric adverse effects in pediatric patients with epilepsy have been sufficiently troublesome for the immediate-release gabapentin. The U.S. Prescribing Information (USPI) includes warnings about the risks of (1) emotional lability or behavioral problems in 6% (vs. 1.3% with placebo); (2) hostility or aggressiveness in 5.2% (vs. 1.3% with placebo); (3) thought disorder, concentration, or academic performance problems in 1.7% (vs. 0% with placebo); and (4) hyperkinesia, restlessness, or hyperactivity in 4.7% (vs. 2.9% with a placebo). Similar to other anticonvulsants, gabapentin has an increased risk of suicidality (suicidal behavior or ideation) in patients with epilepsy as well as psychiatric disorders.

Gabapentin is classified as pregnancy category C, as fetal loss and teratogenicity have been demonstrated in animals but to date not in humans. Gabapentin was classified as “Probably Safe” for lactation, although it is excreted in breast milk, with nursing infants receiving approximately 1 mg/kg per day. Therefore, gabapentin should only be administered to breastfeeding only if the potential benefits clearly outweigh the (unknown) risks.

Drug Interactions

Gabapentin bioavailability may decrease by as much as 20% when administered at the same time as antacids. Therefore, gabapentin should be taken at least 2 hours following antacids. Hydrocodone and morphine may increase blood gabapentin concentrations. However, as gabapentin is renally excreted, it generally lacks clinically significant hepatic drug–drug interactions. In addition, gabapentin may not have major drug–drug interactions with renally excreted drugs.

Dosage and Administration

The immediate-release gabapentin USPI recommends that adjunctive therapy in adolescents and adults with epilepsy should start at 300 mg three times a day and be increased up to 1,800 mg/day in three divided doses. However, in adults with epilepsy, immediate-release gabapentin is generally well tolerated at doses as high as 3,600 and 2,400 mg/day in short- and long-term studies, respectively. Adjunctive immediate-release gabapentin in pediatric seizures is recommended to start at 10 to 15 mg/kg per day in three divided doses and to increase over three days to 25 to 40 mg/kg per day in three divided doses. A long-term epilepsy study suggested that immediate-release gabapentin doses up to 50 mg/kg per day may be well tolerated. Plasma gabapentin concentrations in patients with epilepsy commonly range between approximately 12 and 20 µg/mL (70 and 120 µM/L) but did not appear to be systematically related to antiseizure effects and thus were not routinely monitored.

Immediate-release gabapentin for adults with postherpetic neuralgia or neuropathic pain is recommended to start with 300 mg on day one, increase to 600 mg on day 2, 900 mg on day 3, and subsequently increase as needed to relieve pain up to 1,800 mg/day in three divided doses. Extended-release gabapentin for postherpetic neuralgia is recommended to start at 300 mg daily, increased to 600 mg on day 2, and then further increased to 300 mg every 4 days, with a maximum dose of 1,800 mg. Gabapentin enacarbil for postherpetic neuralgia is recommended to start at 600 mg daily with food for 3 days, then increase to 600 mg twice daily, and for restless legs syndrome is suggested to start and maintain at 600 mg per evening with food.

Off-label gabapentin is typically used adjunctively at similar doses. In some (nonepilepsy) applications, clinicians may administer more of the dosage at bedtime to enhance sleep and decrease daytime somnolence. However, individual doses of immediate-release gabapentin above 1,200 mg are divided due to saturable absorption. In psychiatric outpatients, immediate-release gabapentin may be started at 300 mg at bedtime and increased by 300 mg every 4 days as necessary. Starting on a dose of 1,200 mg at bedtime is typically well tolerated. Occasionally, older adult patients or patients on complex combination therapies may require doses as low as 100 to 200 mg due to sedation. The final total daily doses tend to be between 1,200 and 2,400 mg/day but may range as high as 4,800 mg/day.

TOPIRAMATE

Introduction and History

Topiramate (Topamax) was developed by R. W. Johnson Pharmaceutical Research Institute. After effectiveness was observed in several animal models of epilepsy, topiramate was studied as an antiepileptic drug and then marketed for the adjunctive treatment of epilepsy in the United States in 1997 by Ortho-McNeil Pharmaceutical, Inc.

Double-blind, randomized placebo-controlled trials revealed that topiramate was useful for some, but not all, neurologic and psychiatric conditions reported in initial postmarketing, open, and clinical studies. Topiramate appeared to be effective in migraine prevention and was approved for this indication. In addition, topiramate appeared effective in obesity, bulimia, binge eating disorder with obesity, and alcohol dependence.

Chemistry

Topiramate, 2,3:4,5-bis-*O*-(1-methylethylidene)β-*D*-fructo-pyranose sulfamate, has a novel sulfamate-substituted monosaccharide structure (Fig. 33.10b–1). Two oral topiramate preparations are currently marketed in the United States: 25-, 50-, 100-, and 200-mg tablets, and 15- and 25-mg sprinkle capsules, which can be swallowed whole or opened and sprinkled on food.

Pharmacologic Action

Pharmacokinetics. Topiramate has an oral bioavailability of approximately 80% and linear kinetics that are not significantly altered by food. Topiramate is only 15% bound to plasma proteins and has a moderate volume of distribution of approximately 0.8 L/kg. Topiramate is approximately 70% excreted unchanged in the urine. Therefore, in the absence of enzyme inducers, clearance does not extensively involve hepatic mechanisms. However, in the presence of hepatic enzyme inducers (such as carbamazepine or phenytoin), hepatic metabolism increases, and renal clearance decreases by approximately 50%. Topiramate has a half-life of approximately 24 hours with monotherapy. The half-life decreases by approximately 50% with enzyme inducers, so topiramate is administered in two divided doses. Topiramate clearance is generally approximately 25 mL/min. Enzyme inducers can almost double clearance rates, and moderate to severe renal impairment decreases clearance rates by approximately 50%. Topiramate clearance may also be decreased in patients with hepatic impairment, requiring dosage adjustment.

Pharmacodynamics. Topiramate has multiple biochemical actions, and the relationship between these and clinical effects remains to be established. Topiramate is a state-dependent sodium channel blocker. Topiramate also has GABAergic effects, as it positively modulates $GABA_A$ receptors, and both acute and chronic administration appears to increase cerebral GABA in humans. Topiramate has antiglutamatergic effects, as it blocks kainite and α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptors. Topiramate also weakly inhibits carbonic anhydrase CA-II and CA-IV isoenzymes.

Clinical Drug Studies

Seizure Disorders. Topiramate was found to be superior to placebo as initial monotherapy in adults and children aged 10 years and older with partial-onset or primary generalized seizures in a double-blind, randomized placebo-controlled trial. Adjunctive topiramate appeared to be more effective than an adjunctive placebo in several double-blind, randomized placebo-controlled trials in patients with partial-onset seizures, one trial in children aged 2 to 16 years with partial-onset seizures, one trial in patients aged 2 years or older with primary generalized tonic–clonic seizures, and one trial in patients aged 2 years or older with Lennox–Gastaut syndrome.

Migraine Prevention. Topiramate was more effective than placebo in double-blind randomized placebo-controlled trials, decreasing the frequency of migraine headaches. Topiramate also appeared to improve daily activities and patient functioning.

Other Neurologic Disorders. Adjunctive topiramate was more effective than a placebo in attenuating essential tremor in two of three studies but was ineffective in patients with amyotrophic lateral sclerosis.

Weight Loss, Obesity, Bulimia, and Binge Eating Disorder With Obesity. Topiramate has been associated with weight loss in controlled studies of patients with epilepsy, psychiatric illness, and obesity and has shown efficacy in the treatment of adiposity in women receiving olanzapine, as well as bulimia and binge eating disorder associated with obesity in controlled trials. The combination of topiramate and phentermine was approved by the FDA as an adjunct to a reduced-calorie diet and increased physical activity for chronic weight management in adults with obesity.

Alcohol Use Disorder. Topiramate was superior to placebo in patients with alcohol dependence in numerous double-blind randomized placebo-controlled trials.

Smoking Cessation. In a controlled trial, topiramate promoted smoking abstinence among alcohol-dependent smokers.

Bipolar Disorders. Several double-blind, placebo-controlled trials have failed to demonstrate the efficacy of topiramate monotherapy in acute mania in adults. A similar trial in children and adolescents was inconclusive, perhaps because of premature termination. In a controlled trial, the addition of topiramate to lithium or valproate was no better than adding a placebo in patients with acute manic or mixed episodes. In one small single-blind study, adjunctive topiramate and adjunctive sustained-release bupropion had similar efficacy in bipolar depression, although topiramate appeared to yield more weight loss and adverse effects.

Schizophrenia. In one double-blind, placebo-controlled clinical trial, adjunctive topiramate compared to an adjunctive placebo in patients with schizophrenia yielded greater improvement in general psychopathological symptoms but not positive or negative symptoms.

Posttraumatic Stress Disorder. One double-blind, placebo-controlled clinical trial showed that topiramate was effective in improving reexperiencing and avoidance/numbing symptom clusters in patients with posttraumatic stress disorder. However, another controlled study in male veterans with posttraumatic stress disorder did not show any benefit with topiramate.

Therapeutic Indications

Seizure Disorders. Topiramate is indicated as an adjunctive treatment for adults and pediatric patients aged 2 to 16 years with partial-onset seizures or primary generalized tonic–clonic seizures and in patients aged 2 years or older with Lennox–Gastaut syndrome. Topiramate is also indicated as initial monotherapy in patients 2 years of age and older with partial-onset or primary generalized tonic–clonic seizures.

Migraine Prevention. Topiramate is indicated for adolescents (age ≥12 years) and adults for the prevention of migraine headaches.

Precautions and Adverse Reactions

The most common adverse effects of topiramate 400 mg/day monotherapy in patients with epilepsy were paresthesia, weight decrease, somnolence, anorexia, dizziness, and memory problems in adults, and weight decrease, upper respiratory tract infection, paresthesia, anorexia, diarrhea, and mood problems in children. The most common adverse effects of adjunctive topiramate at 200 to 400 mg/day in adult patients with epilepsy were somnolence, dizziness, ataxia, speech problems, psychomotor slowing, visual problems, memory problems, paresthesia, and diplopia. The most common adverse effects of adjunctive topiramate at 5 to 9 mg/kg per day in pediatric patients with epilepsy were fatigue, somnolence, anorexia, nervousness, concentration or attention problems, memory problems, aggression, and weight loss. The most common adverse effects of topiramate monotherapy in migraine prevention trials were paresthesia, fatigue, nausea, anorexia, upper respiratory tract infection, taste perversion, and dizziness. In many cases, the adverse effects of topiramate are mild to moderate and can be attenuated by dosage rescheduling or decrease.

Topiramate can cause appetite suppression and weight loss in patients with epilepsy, psychiatric disorders, and obesity. Although this is clearly a desirable effect in patients who are overweight or obese, it may be problematic in patients that are underweight. However, in epilepsy studies, weight loss with topiramate was more evident in patients with higher baseline weights. In these studies, topiramate-induced weight loss was more evident in females than in males with similar baseline weights.

The topiramate USPI includes a warning of the risk of hyperchloremic, non-anion gap metabolic acidosis (low serum bicarbonate without chronic respiratory alkalosis), which can occur in a dose-dependent manner with topiramate and yield hyperventilation, fatigue, anorexia, increased risk of nephrolithiasis, nephrocalcinosis, and osteomalacia, decreased growth rate, and rarely cardiac arrhythmias and stupor. Measurement of baseline and periodic serum bicarbonate is recommended, and if metabolic acidosis is detected, decreasing or discontinuing topiramate or treating it with alkali should be considered. Topiramate is rarely associated with acute myopia and secondary angle-closure glaucoma, characterized by acute onset of decreased visual acuity or ocular pain. Symptoms can occur in pediatric and adult patients, typically present within 1 month of starting topiramate and reverse with discontinuation. As with other anticonvulsants, topiramate is associated with an increased risk of suicidality (suicidal behavior or ideation) in patients with epilepsy as well as psychiatric disorders.

An important warning in the topiramate USPI concerns the risks of cognitive or neuropsychiatric adverse events. These are common and include cognitive problems (confusion, psychomotor slowing, concentration or attention, memory, speech, or language problems), psychiatric problems (depression, mood problems), sedation, and fatigue.

Approximately 1.5% of adults exposed to topiramate develop nephrolithiasis. Maintaining adequate fluid intake may help decrease the risk of nephrolithiasis. Topiramate is a weak carbonic anhydrase inhibitor that decreases urinary citrate excretion and increases urinary pH, which may contribute to the increased risk of nephrolithiasis. Carbonic anhydrase inhibition may also contribute to the increased risk of paresthesia with topiramate.

In patients with or without a history of epilepsy, topiramate should be gradually withdrawn to minimize the potential for seizures.

Acute oral overdoses of topiramate have been associated with convulsions, drowsiness, speech disturbance, blurred vision, diplopia, impaired mentation, lethargy, impaired coordination, stupor, hypotension, abdominal pain, agitation, dizziness, depression, and severe metabolic acidosis. In most instances, the clinical consequences were not severe, but deaths have been reported in polydrug overdoses involving topiramate.

Topiramate is classified as pregnancy category C, as fetal loss and teratogenicity have been demonstrated in animals, and use during pregnancy may cause cleft lip and/or palate in humans. It is unknown the degree to which topiramate is excreted in breast milk, so it was classified as “safety unknown” for lactation and should be administered to breastfeeding women with caution and only if the potential benefits clearly outweigh the (unknown) risks.

Drug Interactions

The large renal component of topiramate disposition is resistant to the effects of enzyme inhibitors, yet the limited hepatic metabolism component makes it susceptible to enzyme inducers. For example, in patients with epilepsy, carbamazepine yields clinically significant decreases in topiramate serum concentrations, but topiramate does not yield clinically significant changes in carbamazepine serum concentrations. In contrast, valproate and lamotrigine do not appear to have clinically significant pharmacokinetic interactions with topiramate in patients with epilepsy. However, concomitant administration of topiramate and valproate has been associated with hyperammonemia with and without encephalopathy, and hypothermia (with and without hyperammonemia), most likely due to pharmacodynamic drug interactions. Given its substantial renal clearance, topiramate can affect lithium levels. However, the clinical significance of such putative effects remains to be established. Although topiramate at doses over 200 mg/day may affect serum lithium concentrations, the clinical importance of this phenomenon remains to be adequately characterized. Nevertheless, the topiramate USPI recommends monitoring lithium concentrations when co-administered with high-dose topiramate. The effects of lithium on the pharmacokinetics of topiramate remain to be established. Although levetiracetam does not alter topiramate concentrations, this combination may yield adverse effects through a pharmacodynamic interaction. In addition, the combination of topiramate with other carbonic anhydrase inhibitors, such as acetazolamide and zonisamide, is not recommended because of concerns about increasing the risks of metabolic acidosis, nephrolithiasis or heat-related problems (oligohydrosis and hyperthermia). Topiramate is a mild enzyme inducer that can decrease the blood concentrations of hormonal contraceptives, potentially compromising their efficacy. This effect may be less problematic at doses of 200 mg/day or less. Because of the potential of topiramate to cause central nervous system (CNS) depression, as well as other cognitive and/or neuropsychiatric adverse reactions, combination with alcohol and other CNS depressants should only be performed with caution.

Dosage and Administration

The topiramate USPI recommends that topiramate monotherapy in adults and children aged 10 years or older with epilepsy be started at 25 mg twice daily for the first week, increased to 50 mg twice daily for week 2, 75 mg twice daily for week 3, 100 mg twice daily for week 4, 150 mg twice daily for week 5, and 200 mg twice daily from week 6 onward. In adult (age 17 years or older) patients with epilepsy, topiramate adjunctive therapy is recommended to start with 25 to 50 mg/day and be increased by 25 to 50 mg on a weekly basis, up to 400 mg/day in two divided doses. In pediatric (2 to 16 years) patients with epilepsy, topiramate adjunctive therapy is recommended to start with 1 to 3 mg/kg per day at bedtime and increase by 1 to 3 mg/kg on a weekly basis, up to 5 to 9 mg/kg per day. Plasma topiramate concentrations in patients with epilepsy commonly range between 5 and 20 μg/mL (15 to 60 μM/L), but they do not appear to be systematically related to antiseizure effects and thus are not routinely monitored. For migraine prevention, topiramate is recommended to start with 25 mg in the evening for the first week, increase to 25 mg twice daily for week 2, 25 mg in the morning plus 50 mg in the evening for week 3, and 50 mg twice daily for week 4 onward. Doses may need to be decreased by 50% in patients with moderate to severe renal impairment (creatinine clearance less than 70 mL/1.73 m^2 per minute) and doubled in patients receiving concurrent enzyme inducers such as carbamazepine and phenytoin. Topiramate doses may need to be decreased in patients with hepatic impairment.

Off-label topiramate is typically used adjunctively, starting with 25 mg at bedtime and increasing weekly by 25 mg, as necessary and tolerated. Dosage increases of 25 mg/day every 2 to 4 weeks during initiation may sometimes improve tolerability. Final doses to promote weight loss are often between 75 and 200 mg/day at bedtime. Lower doses appear to suppress appetite less consistently, and higher doses, at least early in therapy, tend not to be well tolerated due to sedation or cognitive difficulties. However, continuing to gradually increase topiramate by 25 mg/day on a monthly basis may allow some patients to eventually achieve doses of 400 mg/day. In some (non-epilepsy) applications, clinicians may administer most or all of the dosage at bedtime to enhance sleep and decrease daytime somnolence.

TIAGABINE

Introduction and History

Tiagabine (Gabitril), a lipophilic derivative of nipecotic acid, was developed by Novo Nordisk. The addition of a lipophilic group to nipecotic acid (a hydrophilic GABA reuptake inhibitor that does not cross the blood–brain barrier) allowed tiagabine to enter the CNS more easily. Tiagabine was found to be effective in several animal models of epilepsy. Abbott Laboratories developed it as an antiepileptic drug and initially marketed it for the adjunctive treatment of partial seizures in the United States in 1997. Tiagabine was later marketed by Cephalon, Inc.

Limited, postmarketing, open, and uncontrolled clinical experience has suggested efficacy in some psychiatric conditions. Safety concerns were raised paradoxically for this drug regarding the risk of seizures in patients with bipolar disorder, particularly if the drug was dosed aggressively. Limited double-blind, placebo-controlled studies have revealed that tiagabine may be useful for treating generalized anxiety disorder and primary insomnia. Safety concerns, along with a lack of controlled efficacy data, have limited the use of tiagabine in disorders other than epilepsy.

Chemistry

Tiagabine, (-)-(R)-1-[4,4-Bis(3-methyl-2-thienyl)-3-butenyl] nipecotic acid, is a lipophilic derivative of the hydrophilic cyclic amino acid nipecotic acid (comprising nipecotic acid linked to a lipophilic group by an aliphatic chain), which inhibits GABA reuptake in neurons and glial cells (Fig. 33.10b–1). Tiagabine is available in 2-, 4-, 12-, and 16-mg tablets for oral administration.

Pharmacologic Action

Pharmacokinetics. Tiagabine is well absorbed with a bioavailability of approximately 90%, extensively (96%) bound to plasma proteins, and has linear kinetics at doses between 2 and 24 mg/day. Giving tiagabine with meals decreases the rate of absorption and delays and lowers peak plasma concentrations but does not significantly affect the extent of absorption. In monotherapy, tiagabine has a half-life of approximately 8 hours and clearance of approximately 110 mL/min. In the presence of enzyme inducers (as in the pivotal epilepsy studies), the half-life decreased to approximately 4 hours, and clearance doubled to approximately 220 mL/min. Tiagabine is a CYP3A substrate that is extensively transformed into inactive 5-oxo-tiagabine and glucuronide metabolites, with only 2% excreted unchanged in the urine. The remainder is excreted as metabolites in the feces (65%) and the urine (25%).

Pharmacodynamics. Tiagabine has a putative specific single major biochemical action that may account for its clinical effects. Tiagabine acts on the GABA transporter, blocking the reuptake of the inhibitory amino acid neurotransmitter GABA into neurons and glia, thus enhancing the inhibitory action of GABA. Actions at both $GABA_A$ and $GABA_B$ receptors putatively yield anticonvulsant and antinociceptive effects, respectively. Tiagabine has antiseizure activity in the pentylenetetrazole mice, amygdala kindled rats, and other animal models of seizures but only has weak activity in the maximal electroshock mouse model.

Clinical Drug Studies

Seizure Disorders. Adjunctive tiagabine appeared to be more effective than an adjunctive placebo in double-blind, randomized placebo-controlled trials and placebo-controlled crossover studies involving patients with refractory partial seizures who were taking at least one enzyme-inducing anticonvulsant.

Pain Syndromes. Limited data suggest the possible utility of tiagabine in bruxism and associated temporomandibular joint pain, painful tonic spasms in multiple sclerosis, chronic pain, and chronic painful neuropathy.

Cocaine Use Disorder. Although preliminary studies were somewhat encouraging, in a larger controlled trial, tiagabine at 20 mg/day was no better than a placebo in treating cocaine dependence.

Therapeutic Indications

Seizure Disorders. Tiagabine is indicated as an adjunctive treatment of partial seizures in adults and adolescents aged 12 years and older with epilepsy.

Precautions and Adverse Reactions

In patients with epilepsy also taking enzyme-inducing anticonvulsants, tiagabine's adverse effects are mild to moderate, can be attenuated by dosage rescheduling or decreases, and do not typically require discontinuation. In patients with epilepsy, the most common adverse effects of adjunctive tiagabine were dizziness or lightheadedness, asthenia or lack of energy, somnolence, nausea, nervousness or irritability, tremor, abdominal pain, and abnormal thinking or difficulty with concentration or attention.

The tiagabine USPI dosing recommendations were based on patients with epilepsy taking enzyme-inducing anticonvulsants, which lower tiagabine plasma concentrations by approximately 50%. Administering tiagabine without enzyme inducers will approximately double tiagabine plasma concentrations compared to those attained in published epilepsy trials, which provided guidance for dosing. Thus, tiagabine administered without enzyme inducers needs to have doses of approximately 50% of those in the USPI. If this adjustment is not made, patients are at an increased risk of adverse effects, some of which are serious.

The tiagabine USPI includes a boldface warning regarding the risk of seizures in patients without epilepsy. As noted above, rapid loading of open primarily adjunctive tiagabine in patients with acute mania, starting at 20 mg/day, yielded unacceptable adverse effects. One patient had a seizure, and gradual initiation in outpatients with bipolar disorder at 4 mg/day and increased weekly as tolerated by 4 mg/day (mean final dose

8.7 mg/day) raised tolerability concerns, with two of 13 patients having seizures.

Other warnings in the tiagabine USPI include the risks of withdrawal seizures, cognitive or neuropsychiatric problems (impaired concentration, speech or language problems, somnolence, and fatigue), status epilepticus, and sudden unexpected death in epilepsy. As with other anticonvulsants, tiagabine is associated with an increased risk of suicidality (suicidal behavior or ideation) in patients with epilepsy as well as psychiatric disorders.

Tiagabine is classified as pregnancy category C, as fetal loss and teratogenicity have been demonstrated in animals, but to date, not in humans. The degree to which tiagabine is excreted in breast milk is unknown, so it was classified as “safety unknown” for lactation and should be administered to breastfeeding women with caution and only if the potential benefits clearly outweigh the (unknown) risks.

Drug Interactions

Tiagabine does not induce or inhibit hepatic enzymes, but it is a target for pharmacokinetic drug interactions. Thus, enzyme inducers such as carbamazepine can yield clinically significant decreases in serum tiagabine concentrations, possibly by inducing CYP3A. Valproate does not yield clinically significant changes in serum total tiagabine concentrations but appears to displace tiagabine from protein-binding sites, yielding an increase in free tiagabine and thus potentially higher tiagabine toxicity. Tiagabine does not yield clinically significant changes in the pharmacokinetics of carbamazepine, phenytoin, and valproate. The presence or absence of pharmacokinetic interactions between tiagabine, lamotrigine, or lithium remains to be established. Tiagabine does not yield clinically significant changes in the kinetics of hormonal contraceptives.

Dosage and Administration

Most importantly, tiagabine should not be rapidly loaded or rapidly initiated due to the risk of serious adverse effects, namely potential seizures. The tiagabine USPI recommends that in adults and adolescents aged 12 years of age or older with epilepsy who were also taking enzyme inducers, tiagabine be initiated at 4 mg/day and increased weekly by 4 mg/day during the first month, and then increased weekly by 4 to 8 mg/day for weeks 5 and 6, yielding 24 to 32 mg/day administered in two to four divided doses by week 6. In adults (but not adolescents), tiagabine doses may be further increased weekly by 4 to 8 mg/day to as high as 56 mg/day. Plasma tiagabine concentrations in patients with epilepsy commonly range between approximately 20 and 100 ng/mL (50 and 250 nM/L) but do not appear to be systematically related to antiseizure effects and thus are not routinely monitored.

In pivotal epilepsy trials, tiagabine was administered with meals, which decreased the rate (later, lower peak plasma concentration) but not the extent of absorption. Thus, to limit the risk of adverse effects, the USPI recommends that tiagabine be administered with meals.

If tiagabine is administered in the absence of enzyme-inducing anticonvulsants, the dosing needs to be even more conservative (50% lower doses) to limit the risk of adverse effects. Thus, it is prudent that tiagabine be initiated at no more than 2 mg/day and increased weekly by 2 mg/day during the first month and then increased weekly by 2 to 4 mg/day for weeks 5 and 6, yielding 12 to 16 mg/day administered in two to four divided doses by week 6. In adults (but not adolescents), tiagabine doses may be further increased weekly by 2 to 4 mg/day to as high as 28 mg/day.

LEVETIRACETAM

Introduction and History

Levetiracetam (Keppra), developed in an effort to create second-generation nootropic (memory enhancing) medications, proved to be a potent anticonvulsant in the audiogenic-susceptible mouse model of epilepsy. In 1999, UCB, Inc. began marketing levetiracetam as an adjunctive treatment for refractory partial seizures in the United States.

Postmarketing, open, uncontrolled clinical experience has suggested potential benefits in patients with depressive symptoms with seizure disorders, bipolar disorder, and premenstrual dysphoric disorder. On the other hand, there have also been reports of levetiracetam-related depression and aggressive behavior, and placebo-controlled studies did not find the efficacy of levetiracetam for behavioral or overall clinical improvement in children with autism. Controlled studies are lacking that assess the effects of levetiracetam in patients with bipolar disorder and other psychiatric disorders.

Chemistry

Levetiracetam, (-)-(*S*)-α-ethyl-2-oxo-pyrrolidine acetamide, is the (*S*)-enantiomer of the ethyl analog of the nootropic agent piracetam (Fig. 33.10b–1) and has a structure unrelated to other anticonvulsants. Levetiracetam preparations currently marketed in the United States include: 250-, 500-, 750-, and 1,000-mg immediate-release tablets, 500- and 750-mg extended-release tablets; a 100 mg/mL oral solution, and a 100 mg/mL injectable intravenous solution.

Pharmacologic Action

Pharmacokinetics. Levetiracetam has a favorable pharmacokinetic profile, with absorption unaffected by food, bioavailability greater than 95%, binding to plasma proteins of less than 10%, a moderate volume of distribution of approximately 0.6 L/kg, linear elimination pharmacokinetics at clinically relevant doses (between 500 and 5,000 mg/day), a half-life of approximately 8 hours, and clearance of 40 mL/min. Levetiracetam is 66% excreted unchanged in the urine and is 24% hydrolyzed (in the blood rather than in the liver) to an inactive metabolite. Dosage may need to be decreased in patients with impaired renal function. However, levetiracetam has a low potential for pharmacokinetic drug interactions because only a minor metabolic component involves hydrolysis (rather than cytochrome P450 [CYP] oxidation or glucuronidation), and levetiracetam does not yield clinically significant changes in hepatic oxidation or glucuronidation.

Pharmacodynamics. Levetiracetam has multiple biochemical actions, and the relationship between these and its clinical effects remains to be established. Levetiracetam has GABAergic and anti-kindling effects, reduces voltage-dependent potassium currents, and binds to the synaptic vesicle protein SV2A. Levetiracetam lacks antiseizure activity in the pentylenetetrazole and maximal electroshock acute seizure models but is potently active in chronic seizure models such as kindled seizures and genetic animal models. Levetiracetam also has antinociceptive activity in animal models of neuropathic pain and anxiolytic-like effects in animal models of anxiety disorders such as the rat conflict and elevated plus-maze.

Clinical Drug Studies

Seizure Disorders. Adjunctive levetiracetam appeared to be more effective than an adjunctive placebo in three double-blind, randomized placebo-controlled trials involving adults with refractory partial seizures with or without secondary generalization; one each in children (age 4 to 16 years) with refractory partial seizures; patients aged 12 years and older with juvenile myoclonic epilepsy; and patients aged 6 years and older with idiopathic primary generalized tonic–clonic seizures.

Other Neurologic Disorders. Levetiracetam was superior to placebo for migraine prevention in children aged 4 to 17 years in a double-blind randomized placebo-controlled trial. Levetiracetam was no better than a placebo for essential tremor and Tourette syndrome.

Psychiatric Disorders. Adequately powered controlled studies are lacking to assess the effects of levetiracetam on psychiatric disorders. A series of uncontrolled case reports and case series suggested that levetiracetam may be effective in patients with bipolar disorder. In a small open-label study, patients with premenstrual dysphoric disorder showed some improvement with levetiracetam. However, in controlled pilot studies, levetiracetam was no better than a placebo in patients with social anxiety disorder and autism. There are reports of levetiracetam-related depression and aggressive behavior.

Therapeutic Indications

Seizure Disorders. Levetiracetam is indicated as an adjunctive treatment for partial-onset seizures in adults and children aged 4 years and older with epilepsy; myoclonic seizures in adults and adolescents aged 12 years and older with juvenile myoclonic epilepsy; and primary generalized tonic–clonic seizures in adults and children aged 6 years and older with idiopathic generalized epilepsy.

Precautions and Adverse Reactions

Levetiracetam is generally well tolerated but can cause somnolence, fatigue, coordination problems, mood, anxiety, and adverse behavioral effects. In adult patients with epilepsy, somnolence was observed in 14.8% (vs. 8.4% with a placebo), asthenia occurred in 14.7% (vs. 9.1%

with a placebo), and coordination problems were evident in 3.4% (vs. 1.6% with a placebo). In adult patients with epilepsy, psychosis occurred in 0.7% (vs. 0.2% with a placebo), and other behavioral problems were observed in 13.3% (vs. 6.2% with a placebo). Behavioral problems included aggression, agitation, anger, anxiety, apathy, depersonalization, depression, emotional lability, hostility, and irritability. Behavioral adverse effects occurred early in treatment (often in the first month of therapy) and led to discontinuation in 1.7% (vs. 0.2% with a placebo). Suicide attempts occurred in 0.5% (vs. 0% with a placebo).

In pediatric patients with epilepsy, somnolence was observed in 22.8% (vs. 11.3% with a placebo), and asthenia occurred in 8.9% (vs. 3.1% with a placebo). In pediatric patients with epilepsy, behavioral problems were observed in 37.6% (vs. 18.6% with a placebo). Behavioral problems in pediatric epilepsy included agitation, anxiety, apathy, depersonalization, depression, emotional lability, hostility, hyperkinesia, nervousness, neurosis, and personality disorder, leading to discontinuation in 10.9% (vs. 6.2% with placebo).

As with other anticonvulsants, levetiracetam is associated with an increased risk of suicidality (suicidal behavior or ideation) in patients with epilepsy as well as psychiatric disorders. Acute oral overdoses of levetiracetam have been associated with somnolence, agitation, aggression, respiratory depression, and coma.

Levetiracetam is classified as pregnancy category C, as fetal loss and teratogenicity have been demonstrated in animals but to date not in humans. Levetiracetam was classified as “probably safe” for lactation, though it is excreted in breast milk, so levetiracetam should be administered to breastfeeding women only if the potential benefits clearly outweigh the (unknown) risks.

Drug Interactions

Levetiracetam’s high degree of renal clearance and hydrolytic metabolism yields a low potential for pharmacokinetic drug interactions. Hence, levetiracetam does not appear to have clinically significant pharmacokinetic interactions with valproate or lamotrigine in patients with epilepsy. Levetiracetam does not significantly alter the serum concentrations of carbamazepine or carbamazepine epoxide in patients with epilepsy. However, levetiracetam plus carbamazepine may still yield adverse effects through pharmacodynamic interactions. Carbamazepine may decrease serum levetiracetam concentrations, but with variable reports of potential clinical significance. Oxcarbazepine may also modestly decrease the serum concentrations of levetiracetam. Levetiracetam does not appear to have clinically significant pharmacokinetic interactions with gabapentin or topiramate in patients with epilepsy, although a combination of levetiracetam and topiramate combination may yield adverse effects through a pharmacodynamic interaction. Despite being renally cleared, levetiracetam and lithium pharmacokinetic interactions remain to be established.

Dosage and Administration

In adult patients with epilepsy, adjunctive levetiracetam is commonly started at 500 to 1,000 mg/day and increased as necessary and tolerated every 2 weeks by 1,000 mg/day, up to a maximum of 3,000 mg/day in two divided doses. In pediatric patients with epilepsy, adjunctive levetiracetam is commonly started at 20 mg/kg per day and increased as necessary and tolerated every 2 weeks by 20 mg/kg per day, up to a maximum of 60 mg/kg per day in two divided doses. Plasma levetiracetam concentrations in patients with epilepsy commonly range between approximately 6 and 20 µg/mL (35 and 120 µM/L) but do not appear to be systematically related to antiseizure effects and thus are not routinely monitored. In view of substantive renal clearance, levetiracetam dosage needs to be adjusted in patients with impaired renal function based on the degree of renal insufficiency.

ZONISAMIDE

Introduction and History

Zonisamide (Zonegran) was synthesized by Dainippon Pharmaceutical Company while screening arylsulfonamides for anticonvulsant effects, putatively related to carbonic anhydrase inhibition. Although zonisamide only weakly inhibited carbonic anhydrase, it was found to be effective in several animal models of epilepsy. Zonisamide was subsequently developed as an antiepileptic drug and marketed for over a decade in Japan and South Korea prior to being marketed for the adjunctive treatment of partial seizures in the United States in 2000 by Elan Pharmaceuticals and more recently by Eisai, Inc.

Limited, open, uncontrolled clinical experience and controlled trials suggest potential efficacy in other neurologic and psychiatric conditions.

Chemistry

Zonisamide (3-[sulfamoylmethyl]-1,2-benzisoxazole) is a 1,2-benzisoxazole-3-methanesulfonamide, which is structurally different from other marked anticonvulsants (Fig. 33.10b–1). Zonisamide is available in 25- and 100-mg capsules for oral administration.

Pharmacologic Action

Pharmacokinetics. Zonisamide generally has favorable pharmacokinetic properties, with good (close to 100%) bioavailability, only 40% to 60% binding to plasma proteins (but extensive binding to erythrocytes), a clearance of 20 mL/min, and a moderate volume of distribution of 1.5 L/kg. Zonisamide has linear pharmacokinetics between 200 and 400 mg/day but has disproportionate serum concentrations at 800 mg, presumably related to saturable binding to carbonic anhydrase in erythrocytes. As a monotherapy, zonisamide has a half-life of 60 hours, which decreases to 30 hours when administered with enzyme inducers. Zonisamide is reduced to 2-sulfamoylacetylphenol (SMAP), primarily by CYP3A4, but is also excreted unchanged in the urine by 30%. Marked renal impairment (creatinine clearance less than 20 mL/min) increased plasma zonisamide concentrations by 35%.

Pharmacodynamics. Zonisamide has multiple biochemical actions, and the relationship between these and its clinical effects remains to be established. Zonisamide blocks sodium channels, reduces T-type calcium currents, facilitates dopaminergic and serotonergic neurotransmission, and is a weak carbonic anhydrase inhibitor. Although zonisamide binds to the GABA/benzodiazepine receptor complex, it does not alter chloride currents and does not appear to have major GABAergic activity.

Clinical Drug Studies

Seizure Disorders. Adjunctive zonisamide appeared to be more effective than an adjunctive placebo in numerous double-blind, randomized placebo-controlled trials in patients with refractory partial seizures, with and without secondary generalization.

Diabetic Neuropathy. Zonisamide was no better than a placebo for diabetic neuropathy in a small double-blind, randomized placebo-controlled trial.

Other Neurologic Disorders. A randomized, double-blind, placebo-controlled study suggested that adjunctive zonisamide was effective in treating motor symptoms in patients with Parkinson disease. However, zonisamide was no better than a placebo for essential tremor in a small double-blind, randomized placebo-controlled trial.

Eating Disorders. In double-blind, randomized placebo-controlled trials, zonisamide was superior to placebo for obesity and for binge eating disorders associated with obesity.

Bipolar Disorder. Open, uncontrolled reports suggest that zonisamide may have some utility in bipolar disorder, including for the treatment of weight gain.

Therapeutic Indications

Seizure Disorders. Zonisamide is indicated as an adjunctive treatment of partial seizures in adults with epilepsy.

Precautions and Adverse Reactions

In epilepsy registration trials, adverse effects with adjunctive zonisamide were generally mild to moderate, commonly attenuated by dosage rescheduling or decreased, and did not typically require discontinuation (adverse effect discontinuation rate 12 vs. 6% with a placebo). The most common adverse effects of adjunctive zonisamide in epilepsy trials were somnolence, anorexia, dizziness, headache, nausea, and agitation or irritability. The appetite suppression effect of zonisamide has been used in patients with obesity and binge eating disorders.

In patient populations other than epilepsy, zonisamide tolerability may be more challenging, with reports of higher adverse effect discontinuation rates in patients with binge eating disorder with obesity (27% vs. 13% with a placebo), diabetic neuropathy (38% vs. 0% with a placebo), essential tremor (30% vs. 0% with a placebo), and bipolar disorder (32% due to worsening mood symptoms with open treatment).

Zonisamide has been associated with nephrolithiasis in approximately 4% of patients (but only symptomatic in approximately 1% of patients), presumably due to carbonic anhydrase inhibition. Zonisamide increases serum creatinine and blood urea nitrogen by approximately 8% but does not alter the glomerular filtration rate (GFR) or yield progressive renal damage. Due to toxicity risks, zonisamide is not recommended in patients with renal failure (GFR less than 50 mL/min). Sudden unexplained death in epilepsy, status epilepticus, and rare creatine phosphokinase elevation and pancreatitis have been reported.

The zonisamide USPI includes a warning regarding the risks of severe reactions (Stevens–Johnson syndrome, toxic epidermal necrolysis, fulminant hepatic necrosis, agranulocytosis, aplastic anemia, and other blood dyscrasias) seen with sulfonamides. Therefore, if signs of such reactions occur, zonisamide should be discontinued immediately. Rash was seen in 2.2% of patients taking zonisamide in pivotal trials, resulting in discontinuation in 1.4% of patients. Other warnings include the risks of metabolic acidosis and nephrolithiasis, oligohydrosis and hyperthermia in pediatric patients, withdrawal seizures, teratogenicity, cognitive or neuropsychiatric problems (depression, psychosis, psychomotor slowing, impaired concentration or speech, language, word finding), somnolence, fatigue, and rare serious hematologic events (two cases of aplastic anemia and one case of agranulocytosis in the first 11 years of marketing in Japan, and no case of aplastic anemia but two cases of agranulocytosis in the United States, European, and Japanese development programs). As with other anticonvulsants, zonisamide is associated with an increased risk of suicidality (suicidal behavior or ideation) in patients with epilepsy as well as psychiatric disorders.

The experience with zonisamide overdose is limited. Acute oral overdoses of zonisamide have been associated with coma, bradycardia, hypotension, and respiratory depression.

Zonisamide is classified as pregnancy category C, as fetal loss and teratogenicity have been demonstrated in animals but to date not in humans. The degree to which zonisamide is excreted in breast milk is unknown, so it was classified as “safety unknown” for lactation and should be administered to breastfeeding women with caution, and only if the potential benefits clearly exist and outweigh the (unknown) risks.

Drug Interactions

Zonisamide does not inhibit CYP450 isoenzymes, does not induce its own metabolism, and is a target rather than an instigator of drug interactions. Carbamazepine decreases serum zonisamide concentrations, while zonisamide does not appear to yield clinically significant changes in the pharmacokinetics of carbamazepine. Zonisamide, valproate, zonisamide, and lamotrigine have clinically significant pharmacokinetic interactions in patients with epilepsy. The presence or absence of pharmacokinetic interactions between zonisamide and lithium remains to be established. Phenytoin and phenobarbital decrease the serum zonisamide concentrations. Combining zonisamide with other carbonic anhydrase inhibitors such as acetazolamide and topiramate is not recommended because of concerns about increasing the risk of nephrolithiasis or heat-related problems (oligohydrosis and hyperthermia).

Dosage and Administration

In patients with epilepsy, adjunctive zonisamide is started at 100 mg/day and increased every 2 weeks (no faster due to a 60 hours half-life) as necessary and tolerated by 100 mg/day, with final doses commonly ranging between 300 and 600 mg/day in two divided doses. Adverse effects were more common at doses of 300 mg/day or higher. Plasma, and zonisamide concentrations in patients with epilepsy, commonly range between approximately 10 and 38 µg/mL (45 and 180 µM/L) but do not appear to be systematically related to antiseizure effects and thus are not routinely monitored. In controlled studies, the mean final zonisamide doses were 160 mg/day for essential tremor, 427 mg/day for obesity, 436 mg/day for binge eating disorder with obesity, and 540 mg/day for painful diabetic neuropathy. As zonisamide undergoes both hepatic metabolism and renal excretion unchanged, patients with hepatic or renal disease may require more conservative dosing and titration as well as more frequent monitoring.

PREGABALIN

Introduction and History

Pregabalin (Lyrica) was developed in a fashion similar to that of gabapentin as a compound with a structure similar to the inhibitory amino acid neurotransmitter GABA, which easily crosses the blood–brain barrier.

Although these agents lack direct GABAergic effects, they were found to be effective in animal models of pain and epilepsy (with pregabalin being more potent) and developed as an analgesic and antiepileptic drugs. Pregabalin was marketed for the treatment of neuropathic pain associated with diabetic peripheral neuropathy in the United States in 2005 by the Parke-Davis division of Pfizer, Inc. Pregabalin subsequently received approval for the treatment of postherpetic neuralgia, adjunctive treatment of adults with partial-onset seizures, for the treatment of fibromyalgia, and for the treatment of neuropathic pain associated with spinal cord injury. Pregabalin also appeared effective in controlled trials for generalized anxiety disorder and social anxiety disorder, but as of late 2014, it had not been approved for these indications.

Chemistry

Pregabalin, *S*-(+)-3-isobutyl GABA, is a novel amino acid derived from GABA (Fig. 33.10b–1), with a structure and mechanisms that overlap those of gabapentin and some structural similarities to other anticonvulsants that resemble GABA, such as valproate. Pregabalin is currently marketed in the United States as an oral preparation in 25-, 50-, 75-, 100-, 150-, 200-, 225-, and 300-mg capsules and as a 20 mg/mL oral solution.

Pharmacologic Action

Pharmacokinetics. Pregabalin appears to have generally favorable pharmacokinetic properties, similar to gabapentin. Pregabalin is not bound to plasma proteins, has a moderate volume of distribution of 0.5 L/kg, a half-life of 6 hours (requiring twice-daily dosing), is greater than 95% excreted unchanged in the urine, and has a clearance of 80 mL/min that varies with creatinine clearance. Pregabalin clearance is highly dependent on renal function, and dosage reduction may be necessary for patients with renal impairment. Pregabalin lacks hepatic metabolism, does not induce or inhibit hepatic metabolism, and lacks metabolic drug interactions. Unlike gabapentin, pregabalin has higher (greater than 90%) bioavailability and absorption that is proportional to dose, without the gastrointestinal absorption saturation of gabapentin. Food has no clinically relevant effect on pregabalin absorption.

Pharmacodynamics. Pregabalin has multiple biochemical actions, and the relationship between these and its clinical effects remains to be established. Pregabalin, like gabapentin, is a structural analog of GABA, but these agents appear to lack direct GABAergic actions. However, both of these agents have selective inhibitory effects on voltage-gated calcium channels containing the $\alpha_2\delta$ subunit, reducing depolarization-induced calcium influx at nerve terminals, resulting in decreased release of excitatory neurotransmitters such as glutamate and noradrenaline.

Clinical Drug Studies

Seizure Disorders. Adjunctive pregabalin was more effective than an adjunctive placebo in double-blind, randomized placebo-controlled trials in adults with partial-onset seizures.

Postherpetic Neuralgia. Pregabalin was more effective than a placebo in double-blind, randomized placebo-controlled trials in adults with postherpetic neuralgia. Mood, sleep, and quality of life measures improved, in addition to the improvement of pain problems.

Diabetic Peripheral Neuropathy. Pregabalin was more effective than a placebo in double-blind, randomized placebo-controlled trials in adults with diabetic peripheral neuropathy. Similar to postherpetic neuralgia, pain, mood, sleep, and quality of life measures improved.

Neuropathic Pain Associated With Spinal Cord Injury. In adults with neuropathic pain associated with spinal cord injury, pregabalin was more effective than a placebo in double-blind, randomized, placebo-controlled trials. Similar to postherpetic neuralgia, pain, mood, sleep, and quality of life measures all improved.

Fibromyalgia. Pregabalin was more effective than placebo for fibromyalgia in one double-blind, randomized placebo-controlled trial (with superior relief of pain, disturbed sleep, and fatigue) and in a 6-month randomized withdrawal study.

Other Pain Syndromes. Pregabalin was effective for central neuropathic pain associated with spinal cord injury, improving pain, sleep, anxiety, and overall patient status. Pregabalin appeared to be effective for pain associated with chronic pancreatitis, postoperative mastectomy, and neuropathic pain due to severe burns. Pregabalin was not found to be more effective than a placebo for pain associated with prostatitis and HIV.

Other Neurologic Disorders. In a pilot, double-blind, randomized placebo-controlled trial, pregabalin was superior to placebo in the treatment of essential tremor, whereas no significant improvement was found in a double-blind, randomized placebo-controlled, crossover trial.

Anxiety Disorders. In double-blind, randomized placebo-controlled trials, pregabalin was effective as both short- and long-term treatment for adult and older adult patients with generalized anxiety disorder and social anxiety disorder.

Therapeutic Indications

Diabetic Peripheral Neuropathy. Pregabalin is indicated for the treatment of diabetic peripheral neuropathy.

Postherpetic Neuralgia. Pregabalin is indicated for the treatment of postherpetic neuralgia.

Neuropathic Pain Associated With Spinal Cord Injury. Pregabalin is indicated for the treatment of neuropathic pain associated with spinal cord injury.

Seizure Disorders. Pregabalin is indicated for the adjunctive treatment of adults with partial-onset seizures.

Fibromyalgia. Pregabalin is indicated for the treatment of fibromyalgia.

Precautions and Adverse Reactions

In general, the adverse effects of pregabalin are mild to moderate, can be attenuated by dosage rescheduling or decreases, and do not typically require discontinuation. The most common adverse effects in patients with diabetic neuropathy include dizziness, somnolence, peripheral edema, dry mouth, and asthenia. The most common adverse effects in patients with postherpetic neuralgia were dizziness, somnolence, peripheral edema, infection, and headache. In patients with partial-onset seizures, the most common adverse effects of adjunctive pregabalin were dizziness, somnolence, ataxia, weight gain, blurred vision, diplopia, and accidental injury. The most common adverse effects in patients with fibromyalgia were dizziness, somnolence, headache, weight gain, blurred vision, and dry mouth. The pregabalin USPI includes warnings regarding the risks of several of the above adverse effects (dizziness, somnolence, peripheral edema, weight gain, and blurred vision), as well as angioedema, hypersensitivity, withdrawal symptoms, tumorigenic potential (hemangiosarcoma in mice), elevated creatine kinase, thrombocytopenia, and PR interval prolongation. Like other anticonvulsants, pregabalin is associated with an increased risk of suicidality (suicidal behavior or ideation) in patients with epilepsy and psychiatric disorders.

Pregabalin is a schedule V-controlled substance, with abuse and dependence potential concerns based on its potential to yield euphoria and the potential for withdrawal reactions.

Pregabalin is classified as pregnancy category C, as fetal loss and teratogenicity have been demonstrated in animals but to date not in humans. It is not unknown the degree to which pregabalin is excreted in breast milk, so it was classified as “safety unknown” for lactation and should be administered to breastfeeding women with caution and only if the potential benefits clearly outweigh the (unknown) risks.

Drug Interactions

Similar to gabapentin, in view of the absence of hepatic metabolism, pregabalin lacks metabolic drug interactions. Thus, in patients with epilepsy, pregabalin lacks clinically significant pharmacokinetic drug interactions with carbamazepine, valproate, and lamotrigine. The presence or absence of pregabalin pharmacokinetic interactions with lithium remains to be established. In patients with epilepsy, pregabalin lacks clinically significant pharmacokinetic drug interactions with gabapentin and topiramate. Although tiagabine does not affect the pharmacokinetics of pregabalin, the effect of pregabalin on the pharmacokinetics of tiagabine remains to be established. The presence or absence of pregabalin pharmacokinetic interactions with oxcarbazepine, levetiracetam, and zonisamide remains to be established.

Dosage and Administration

The pregabalin USPI recommends that therapy for diabetic neuropathy, postherpetic neuralgia, fibromyalgia, partial-onset seizures, and neuropathic pain associated with spinal cord injury should start at 150 mg/day in two to three (two for fibromyalgia and neuropathic pain associated with spinal cord injury) divided doses, and be increased within 1 week (based on efficacy and tolerability) to 300 mg/day in two or three (two for fibromyalgia and neuropathic pain associated with spinal cord injury) divided doses. Patients with insufficient efficacy may be further increased to 450 mg/day in divided doses for fibromyalgia and 600 mg/day in divided doses for partial-onset seizures, postherpetic neuralgia, and neuropathic pain associated with spinal cord injury. Plasma pregabalin concentrations in patients with epilepsy do not appear to be systematically related to antiseizure effects and thus are not routinely monitored. In controlled trials, pregabalin doses ranged from 150 to 600 mg/day in generalized anxiety disorder and social anxiety disorder, although pregabalin lacks a specific indication for anxiety disorders. In patients with renal impairment, pregabalin doses may need to be decreased, depending on the degree of renal insufficiency.

PERAMPANEL

Introduction and History

Perampanel (Ficompa), which acts as a highly selective, noncompetitive antagonist of α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors, was developed by Eisai to create an agent for the treatment of epilepsy. The AMPA receptor is one of the four ionotropic transmembrane receptors for glutamate, and it plays a pivotal role in fast synaptic transmission in the CNS. This compound is postulated to reduce excessive neuronal excitation associated with seizures by blocking AMPA receptors expressed on postsynaptic membranes. Perampanel was first marketed in 2012 in Europe and the United States and is approved for monotherapy and adjunctive use in the treatment of partial-onset seizures with or without secondarily generalized seizures in patients with epilepsy aged 4 years and older and adjunctive therapy for primary generalized tonic–clonic seizures in patients with epilepsy aged 12 years and older.

Chemistry

Perampanel, 2-(2-oxo-1-phenyl-5-pyridin-2-yl-1,2-dihydropyridin-3-yl) benzonitrile, has a 2,3’-bipyridin-6’-one core structure, which distinguishes it from other AMPA receptor antagonists. Perampanel preparations marketed in the United States, Europe, and Japan include 2-, 4-, 6-, 8-, 10-, and 12-mg tablets; a 0.5 mg/mL oral suspension; and 1% fine granule.

Pharmacologic Action

Pharmacokinetics. The half-life of perampanel is approximately 105 hours, and its steady state is reached in 2 to 3 weeks. Perampanel is rapidly and completely absorbed after oral administration with negligible first-pass metabolism. The drug is approximately 95% to 96% bound to plasma proteins. Perampanel is extensively metabolized via primary oxidation, followed by glucuronidation. The oxidative metabolism is primarily mediated by CYP3A4. Moreover, 70% of the dose is excreted in the feces and 30% in the urine. The area under the curve (AUC) of perampanel increased in a dose-proportional manner after single-dose administration of 0.2 to 12 mg tablets and after multiple-dose administration of 1- to 12-mg tablets once daily.

Its oral suspension and fine granules have comparable bioavailability to tablets. No significant effect of age on the apparent clearance of perampanel was found.

Pharmacodynamics. The precise mechanism by which perampanel exerts its antiepileptic effects in humans remains to be fully elucidated. Activation of AMPA receptors by glutamate is considered to be primarily responsible for fast excitatory synaptic transmission in the brain. In in vitro studies, perampanel did not compete with AMPA for binding to the AMPA receptor, but perampanel binding was displaced by noncompetitive AMPA receptor antagonists, indicating that perampanel is a noncompetitive AMPA receptor antagonist. In vitro, perampanel inhibited the AMPA-induced increase in intracellular calcium levels. In vivo, perampanel significantly prolonged seizure latency in an AMPA-induced seizure model.

Clinical Drug Studies

Seizure Disorders. Both monotherapy and adjunctive use of perampanel in partial-onset seizures, with or without secondary generalization, were more effective than a placebo in patients who were not adequately controlled with one to three drugs for epilepsy in patients aged 12 years and older in double-blind, randomized placebo-controlled trials. Adjunctive therapy for primary generalized tonic–clonic seizures also showed superiority over placebo in patients with epilepsy aged 12 years and older. Although one systematic review suggested the potential of perampanel as a broad-spectrum antiseizure medication, more data are needed.

Therapeutic Indications

Seizure Disorders. Perampanel was first marketed in 2012 in Europe and the United States and is approved for monotherapy and adjunctive use in the treatment of partial-onset seizures with or without secondarily generalized seizures in patients with epilepsy aged 4 years and older and adjunctive therapy for primary generalized tonic–clonic seizures in patients with epilepsy aged 12 years and older.

Precautions and Adverse Reactions

Hostility- and aggression-related adverse reactions occurred in 12% and 20% of patients randomized to receive perampanel at 8 mg and 12 mg/day, respectively, compared to 6% in the placebo group. These effects were dose-dependent and generally appeared within the first 6 weeks of treatment, although new events continued to be observed for more than 37 weeks. Other adverse events included irritability, aggression, anger, and anxiety, which occurred in 2% or more of perampanel-treated patients and twice as frequently as in placebo-treated patients. Patients, their caregivers, and families should be informed that perampanel may increase the risk of psychiatric events.

Like all other drugs for epilepsy, perampanel increases the risk of suicidal thoughts or behavior in patients taking these drugs for any indication. Patients taking these medications should be monitored for the emergence or worsening of depression, suicidal thoughts or behavior, and/or any unusual changes in mood or behavior.

Common dose-related adverse reactions in patients receiving 8 or 12 mg/day in clinical trials (≥4% and occurring at least 1% higher than the placebo group) included dizziness (36%), somnolence (16%), fatigue (10%), irritability (9%), falls (7%), nausea (7%), ataxia (5%), balance disorder (4%), gait disturbance (4%), vertigo (4%), and weight gain (4%).

Perampanel was classified as pregnancy category C. It is unknown the degree to which perampanel is excreted in breast milk, so it was classified as “safety unknown” for lactation and should be administered to breastfeeding women with caution and only if the potential benefits clearly outweigh the (unknown) risks.

Drug Interactions

Concomitant use of perampanel at 12 mg/day reduced the C_{max} and AUC of levonorgestrel by approximately 40%. Concomitant use of moderate and strong CYP3A4 inducers, including carbamazepine, phenytoin, or oxcarbazepine with perampanel, decreased the plasma concentrations of perampanel by approximately 50% to 67%. Concomitant use of perampanel with CNS depressants, including alcohol, may increase the depressant effects.

Dosage and Administration

For monotherapy or adjunctive therapy for partial-onset seizures, the starting dose in adults and pediatric patients aged 4 years and older is recommended to be 2 mg/day once at bedtime. Dose increases should be performed no more frequently than at weekly intervals in increments of 2 mg/day based on individual clinical response and tolerability. The recommended maintenance dose ranges from 8 to 12 mg/day, although some patients may respond to 4 mg/day. Dosage adjustment is recommended in cases where moderate or strong CYP3A4 enzyme-inducing drugs are concomitantly used.

For adjunctive therapy for primary generalized tonic–clonic seizures, the recommended starting dose in adults and pediatric patients aged 12 years and older is 2 mg/day once at bedtime, and dose increases should be performed no more frequently than at weekly intervals in increments of 2 mg/day based on individual clinical response and tolerability. The recommended maintenance dose was 8 mg/day at bedtime. Dosage adjustment is recommended with the concomitant use of moderate or strong CYP3A4 enzyme-inducing drugs.

FURTHER READINGS

Ahmed M, Hays R, Steven Poceta J, Jaros MJ, Kim R, Shang G. Effect of gabapentin enacarbil on individual items of the international restless legs study group rating scale and post-sleep questionnaire in adults with moderate-to-severe primary restless legs syndrome: pooled analysis of 3 randomized trials. *Clin Ther*. 2016;38(7):1726–1737.

*Bray GA, Hollander P, Klein S, et al. A 6-month randomized, placebo-controlled, dose-ranging trial of topiramate for weight loss in obesity. *Obes Res*. 2003;11(6):722–733.

Coyle H, Clough P, Cooper P, Mohanraj R. Clinical experience with perampanel: focus on psychiatric adverse effects. *Epilepsy Behav*. 2014;41:193–196.

Frampton JE. Pregabalin: a review of its use in adults with generalized anxiety disorder. *CNS Drugs*. 2014;28(9):835–854.

*Gadde KM, Kopping MF, Wagner HR 2nd, Yonish GM, Allison DB, Bray GA. Zonisamide for weight reduction in obese adults: a 1-year randomized controlled trial. *Arch Intern Med*. 2012;172(20):1557–1564.

Johnson BA, Rosenthal N, Capece JA, et al. Topiramate for Alcoholism Advisory Board, Topiramate for Alcoholism Study Group. Topiramate for treating alcohol dependence: a randomized controlled trial. *JAMA*. 2007;298(14):1641–1651.

Knapp CM, Ciraulo DA, Sarid-Segal O, et al. Zonisamide, topiramate, and levetiracetam: efficacy and neuropsychological effects in alcohol use disorders. *J Clin Psychopharmacol*. 2015;35(1):34–42.

Kushner SF, Khan A, Lane R, Olson WH. Topiramate monotherapy in the management of acute mania: results of four double-blind placebo-controlled trials. *Bipolar Disord*. 2006;8(1):15–27.

Lee JJ, Song HS, Hwang YH, Lee HW, Suh CK, Park SP. Psychiatric symptoms and quality of life in patients with drug-refractory epilepsy receiving adjunctive levetiracetam therapy. *J Clin Neurol*. 2011;7(3):128–136.

Mason BJ, Quello S, Goodell V, Shadan F, Kyle M, Begovic A. Gabapentin treatment for alcohol dependence: a randomized clinical trial. *JAMA Intern Med*. 2014;174(1):70–77.

*McElroy SL, Hudson JI, Capece JA, Beyers K, Fisher AC, Rosenthal NR. Topiramate Binge Eating Disorder Research Group. Topiramate for the treatment of binge eating disorder associated with obesity: a placebo-controlled study. *Biol Psychiatry*. 2007;61(9):1039–1048.

McElroy SL, Kotwal R, Guerdjikova AI, Welge JA, Nelson EB. Zonisamide in the treatment of binge eating disorder with obesity: a randomized controlled trial. *J Clin Psychiatry*. 2006;67(12):1897–1906.

Molokwu OA, Ezeala-Adikaibe BA, Onwuekwe IO. Levetiracetam-induced rage and suicidality: two case reports and review of literature. *Epilepsy Behav Case Rep*. 2015;4:79–81.

Pollack MH, Tiller J, Xie F, Trivedi MH. Tiagabine in adult patients with generalized anxiety disorder: results from 3 randomized, double-blind, placebo-controlled, parallel-group studies. *J Clin Psychopharmacol*. 2008;28(3):308–316.

*Rickels K, Pollack MH, Feltner DE, et al. Pregabalin for treatment of generalized anxiety disorder: a 4-week, multicenter, double-blind, placebo-controlled trial of pregabalin and alprazolam. *Arch Gen Psychiatry*. 2005;62(9):1022–1030.

Sadeghian H, Motiei-Langroudi R. Comparison of levetiracetam and sodium valproate in migraine prophylaxis: a randomized placebo-controlled study. *Ann Indian Acad Neurol*. 2015;18(1):45–48.

Shin JH, Gadde KM. Clinical utility of phentermine/topiramate (Qsymia™) combination for the treatment of obesity. *Diabetes Metab Syndr Obes*. 2013;6:131–139.

Silberstein S, Lipton R, Dodick D, et al. Topiramate treatment of chronic migraine: a randomized, placebo-controlled trial of quality of life and other efficacy measures. *Headache*. 2009;49(8):1153–1162.

Sun Y, van Valkenhoef G, Morel T. A mixed treatment comparison of gabapentin enacarbil, pramipexole, ropinirole and rotigotine in moderate-to-severe restless legs syndrome. *Curr Med Res Opin*. 2014;30(11):2267–2278.

*Trinka E, Lattanzi S, Carpenter K, et al. Exploring the evidence for broad-spectrum effectiveness of perampanel: A systematic review of clinical data in generalised seizures. *CNS Drugs*. 2021;35(8):821–837.

Vieta E, Manuel Goikolea J, Martínez-Arán A, et al. A double-blind, randomized, placebo-controlled, prophylaxis study of adjunctive gabapentin for bipolar disorder. *J Clin Psychiatry*. 2006;67(3):473–477.

Wang PW, Yang YS, Chandler RA, et al. Adjunctive zonisamide for weight loss in euthymic bipolar disorder patients: a pilot study. *J Psychiatr Res*. 2008;42(6):451–457.

33.10c Carbamazepine

ROBERT M. POST, M.D., AND MARK A. FRYE, M.D.

Neuroscience-Based Nomenclature (NbN)	
Glutamate Channel Blocker	
Pharmacologic Domain	Glutamate
Mechanism of Action	Voltage-gated sodium and calcium channel blocker

Carbamazepine (Tegretol) is an iminostilbene drug possessing some structural similarity to the tricyclic antidepressant imipramine (Tofranil) but having a very different clinical spectrum of efficacy. It was approved for use in the United States for the treatment of trigeminal neuralgia in 1968 and for temporal lobe epilepsy (complex partial seizures) in 1974. Interestingly, carbamazepine was first synthesized as a potential antidepressant, but because of its atypical profile in a number of animal models, it was initially developed for use in pain and seizure disorders. It is now recognized in most guidelines as a first- or second-line mood stabilizer useful in the treatment and prevention of both phases of bipolar affective disorder. A long-acting sustained-release formulation (Equetro) was approved by the FDA for the treatment of acute mania in 2002.

The structure and chemistry of carbamazepine are closely related to the more recently introduced anticonvulsant oxcarbazepine (Trileptal). As illustrated in Figure 33.10c–1, whereas carbamazepine has a double bond in its middle ring, oxcarbazepine has a keto-oxygen attached. These differences result in very different routes and profiles of drug metabolism, yet, apparently, a generally similar clinical spectrum of efficacy in seizure, pain, and mood disorders. Oxcarbazepine is not FDA approved for acute mania. Although double-blind comparisons showed equal response to oxcarbazepine and haloperidol (Haldol) or lithium (Eskalith), these results were contaminated by the concomitant use of high-potency benzodiazepines. One study suggested better efficacy of oxcarbazepine in less severe mania compared with more severe mania, while carbamazepine is effective in severe forms of mania.

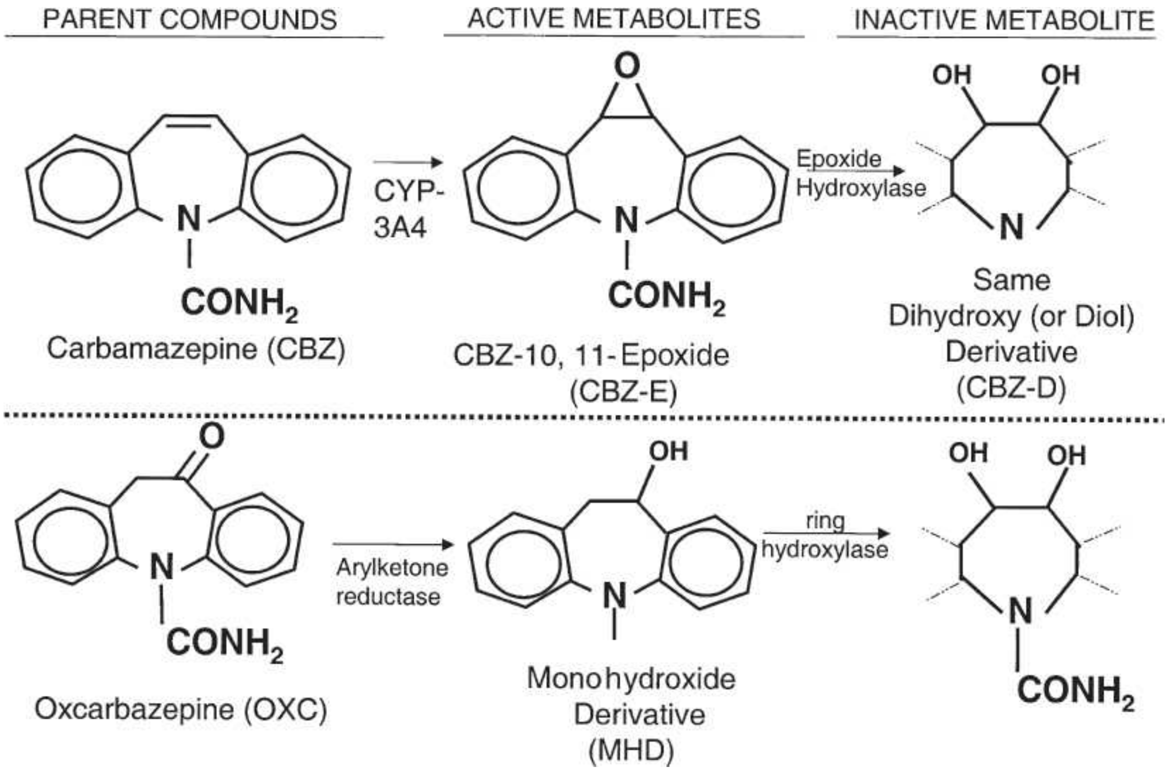
PARENT COMPOUNDS
ACTIVE METABOLITES
INACTIVE METABOLITE
N
CONH2
Carbamazepine (CBZ)
CYP-3A4
O
N
CONH2
CBZ-10, 11-Epoxide (CBZ-E)
Epoxide Hydroxylase
OH
OH
N
Same Dihydroxy (or Diol) Derivative (CBZ-D)
O
N
CONH2
Oxcarbazepine (OXC)
Arylketone reductase
OH
N
Monohydroxide Derivative (MHD)
ring hydroxylase
OH
OH
N
CONH2

FIGURE 33.10c–1. Comparative metabolism of carbamazepine and oxcarbazepine. Both parent and first metabolite are active anticonvulsants—the top two are potent enzyme inducers; the bottom two are not.

CHEMISTRY

Carbamazepine is oxidized by cytochrome P450 oxidase of the 3A4 subtype to an active 10,11-epoxide metabolite that also possesses both anticonvulsant and antinociceptive properties. This epoxide is stable but can be further hydrolyzed to an inactive dihydroxy compound via an enzyme called epoxide hydroxylase. In contrast, oxcarbazepine is metabolized directly to its monohydroxy metabolite, which is also metabolically and physiologically active. Neither oxcarbazepine nor its active monohydroxy metabolite is readily oxidized by P450 liver enzymes, resulting in a lesser number of pharmacokinetic interactions with other drugs compared with carbamazepine. These and other differences are highlighted in Table 33.10c–1.

Carbamazepine is highly insoluble in water and difficult to dissolve in most preparations. It is commonly given in preclinical studies to animals in a solution of ethanol and propylene glycol, but no clinically available intravenous formulation exists.

Two long-acting preparations are now available. One of these preparations, extended-release carbamazepine, needs to be taken with food, as its absorption depends on normal gastrointestinal transit time, and it leaves a residual capsule ghost. Another long-acting preparation (Carbatrol or Equetro) uses a microtrol delivery system with 25% uncoated intermediate release beads, which are absorbed almost as soon as the pill is swallowed; 40% extended-release beads, which dissolve gradually over a period of 8 to 12 hours because of a polymer coating that allows carbamazepine to diffuse slowly through tiny micropores; and 35% enteric release beads, which release the drug very slowly as a result of special pH-sensitive coating.

Such a formulation makes the long-acting forms suitable for single bedtime administration in affectively ill patients (although this is not recommended in epilepsy).

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Carbamazepine is absorbed slowly and somewhat erratically through the gastrointestinal tract. Bioavailability is 80%, and the drug should not be exposed to high humidity because this can decrease bioavailability. Peak plasma levels are reached 2 to 8 hours after a single dose. Carbamazepine is 70% to 80% bound to plasma protein. The average half-life after a single dose is 26 hours, with a range of 18 to 54 hours. However, with long-term administration in humans, the half-life decreases to approximately 12 hours (range, 10 to 25 hours). This change occurs with the induction of hepatic P450 3A4 enzymes, which increase the rate of metabolism of carbamazepine itself (autoinduction) and a variety of other agents, as referred to in Table 33.10c–2.

Table 33.10c–1.
Comparative Effects of Carbamazepine (CBZ) and Oxcarbazepine (OXC)

	CBZ	OXC	Comment on OXC
Hepatic enzyme 3A4 induction	+++	+	No autoinduction
Decrease WBC	+++	±	No WBC suppression
Rash	6.5%	2.5%	Lower rate of rash[a]
Hyponatremia	+	++	More hyponatremia
LFT↑	+	±	Less liver enzyme elevation
Teratogenic	++	±	Lesser incidence of spina bifida
Agranulocytosis Aplastic anemia	1/10,000	±	Lower incidence of hematologic reaction

[a] With benign rash on CBZ did not rerash on OXC. 3/3 with exfoliative rash did, however.

Table 33.10c–2.

Drugs Whose Serum Concentrations Are Decreased by Carbamazepine Via Prominent Induction of Cytochrome P450 3A4

Drug Group	Drugs
Analgesics	Alfentanil; buprenorphine; fentanyl (?); levobupivacaine; methadone; tramadol
Anticoagulants	Dicoumarol (?); phenprocoumon; warfarin[a]
Antiepileptics	Carbamazepine[a]; ethosuximide; felbamate; lamotrigine[a]; oxcarbazepine; phenytoin; primidone; tiagabine; topiramate; valproic acid[b]; zonisamide
Antidepressants	Bupropion; citalopram; mirtazapine; tricyclic antidepressants
Antimicrobials	Caspofungin; doxycycline
Antipsychotics	Aripiprazole; clozapine; fluphenazine (?); haloperidol[a]; lurasidone[a]; olanzapine; quetiapine; risperidone; thiothixene (tiotixene); ziprasidone; brexpiprazole; cariprazine
Antivirals	Delavirdine; protease inhibitors
Anxiolytics/	Alprazolam (?); buspirone
Sedatives	Clonazepam; midazolam; zopiclone (?)
Immunosuppressants	Cyclosporine (?); sirolimus (rapamycin); tacrolimus
Muscle relaxants	Doxacurium chloride; pancuronium bromide; rapacuronium bromide; rocuronium bromide; vecuronium bromide
Estrogen; steroids	Estrogen in hormonal contraceptives[a]; dexamethasone[b]; mifepristone; prednisolone
Stimulants	Methylphenidate[b]; modafinil
Others	Bepridil; oxiracetam (?); paclitaxel;

	quinidine; remacemide (?); repaglinide; theophylline (?); thyroid hormones

[a] Clinical most important.

[b] Important.

(?) Signifies doubtful evidence/clinical significance.

The degree of carbamazepine autoinduction is somewhat dose-dependent but is usually complete after 3 to 5 weeks of treatment. Thus, doses that are not initially tolerated may become so after several weeks, and blood levels usually fall despite the use of the same daily dose, necessitating later dose increases to maintain similar blood levels and clinical effect. Long-term use of carbamazepine is associated with a markedly increased ratio of the amount of the 10,11-epoxide to carbamazepine. The half-life of the 10,11-epoxide is approximately 6 hours.

Carbamazepine is metabolized in the liver and then excreted by the kidneys, as only 1% is eliminated by biliary excretion. Oxcarbazepine is a white to faintly orange crystalline powder, and, like carbamazepine, it is practically insoluble in ethanol, ether, or water. It is slightly soluble in chloroform, acetone, and methanol.

Because there is no or minimal autoinduction with oxcarbazepine, its half-life remains relatively stable over time. The half-life of oxcarbazepine is approximately 2 hours, and because of the rapid conversion to the monohydroxide, which has a half-life of approximately 9 hours, the monohydroxide is responsible for most of the anticonvulsant activity. Dosing is usually twice or three times daily. There is an FDA-approved long-acting preparation of oxcarbazepine (Oxtellar) for complex partial seizures, but efficacy in bipolar disorder has not been demonstrated. However, single nighttime dosing is possible with it. Oxcarbazepine is 96% absorbed and 60% bound to plasma proteins. Its monohydroxy derivative is 40% bound to plasma proteins. The monohydroxy derivative, eslicarbazepine, was FDA approved for seizures in 2013, although the results of earlier trials in mania have not been reported.

PHARMACODYNAMICS

Mechanisms of Antinociceptive, Antimanic, and Antidepressant Action

The onset of anticonvulsant and antinociceptive effects of carbamazepine is rapid and can occur after a single dose in appropriate animal model studies or after 24 to 48 hours in most clinical studies of paroxysmal pain syndromes, such as trigeminal neuralgia (Fig. 33.10c–2). Its antimanic actions can also be relatively rapid in onset, beginning within the first several days of treatment, but several weeks are typically required for maximum treatment response in patients. In contrast, the antidepressant effects of carbamazepine are often delayed, with 3 to 4 weeks required before maximum effects are observed. However, because carbamazepine can be relatively sedating, an improvement in sleep may occur much earlier, yet as with other antidepressant modalities, the degree of improvement in sleep is not correlated with the degree of subsequent antidepressant response.

In assessing the mechanism of action related to clinical effects, the actions of carbamazepine at clinically relevant blood levels and in the appropriate time frame should be considered, as schematized in Figure 33.10c–2. Immediate effects would be more likely related to the anticonvulsant and antinociceptive profile of the drug, whereas effects requiring more chronic administration would be those more likely linked to the parallel time course of antidepressant effects of carbamazepine.

The anticonvulsant effects of carbamazepine have been most closely linked to the blockade of type 2 or batrachotoxin-sensitive sodium channels. When these channels are inactivated, they tend to decrease the release of excitatory amino acids such as glutamate and inhibit sustained rapid neuronal firing in a variety of in vitro models. This action of sodium channel blockade is shared by oxcarbazepine, as well as a variety of other anticonvulsant compounds, including phenytoin (Dilantin), lamotrigine (Lamictal), and topiramate (Topamax), and, to a lesser extent, valproate (Depakene; Depakote). Although these are acute effects of the drug and are associated with decreased neurotransmitter release, Robert Belmaker and colleagues have postulated that the potency of sodium channel blockade is related to the antimanic efficacy of the anticonvulsant compounds as a group. This potential relationship, which is discordant with the time frame of clinical action and violated by topiramate (which is not antimanic), remains to be further investigated.

The acute anticonvulsant effects of carbamazepine have been closely linked to those that can be readily studied in a variety of animal models, including amygdala-kindled seizures. Carbamazepine appears to exert its actions on amygdala-kindled seizures through the so-called peripheral-type or mitochondrial benzodiazepine receptor. It is at this receptor that the peripheral-type ligand Ro 5–4864 is proconvulsant and an agonist, whereas PK1195 is an anticonvulsant and an antagonist. Carbamazepine appears to share PK1195-like antagonist activities at the peripheral-type benzodiazepine receptor.

In contrast to the anticonvulsant effects of diazepam (Valium), which are reversed by a pure central benzodiazepine antagonist such as Ro 15–1788 or flumazenil (Romazicon), carbamazepine actions are reversed by Ro 5–4864 in a PK1195-blockable fashion, strongly implicating this peripheral-type benzodiazepine receptor in carbamazepine's anticonvulsant actions on amygdala-kindled seizures (as illustrated in Fig. 33.10c–3A). Whether this peripheral-type effect is related to any of carbamazepine's psychotropic properties remains to be further investigated.

The peripheral-type benzodiazepine receptors have been linked to cholesterol transport and calcium influx through mitochondrial membranes. Cholesterol is converted to pregnenolone in mitochondria, and this compound subsequently becomes transformed into neuroactive steroids: allopregnanolone and allotetrahydro-DOC (which enhance γ-aminobutyric acid type A [$GABA_A$] receptors); pregnenolone sulfate (which inhibits TBPS [t-(35)S]butyl-bicyclophosphorothionate] binding and voltage-gated Ca^{++} channels, but enhances NMDA receptors); and dehydroepiandrosterone sulfate (which inhibits the barbiturate binding site). Thus, it remains possible that some of carbamazepine's actions initiated at peripheral-type benzodiazepine receptors may be achieved indirectly at other receptors through these effects on neurosteroid metabolism.

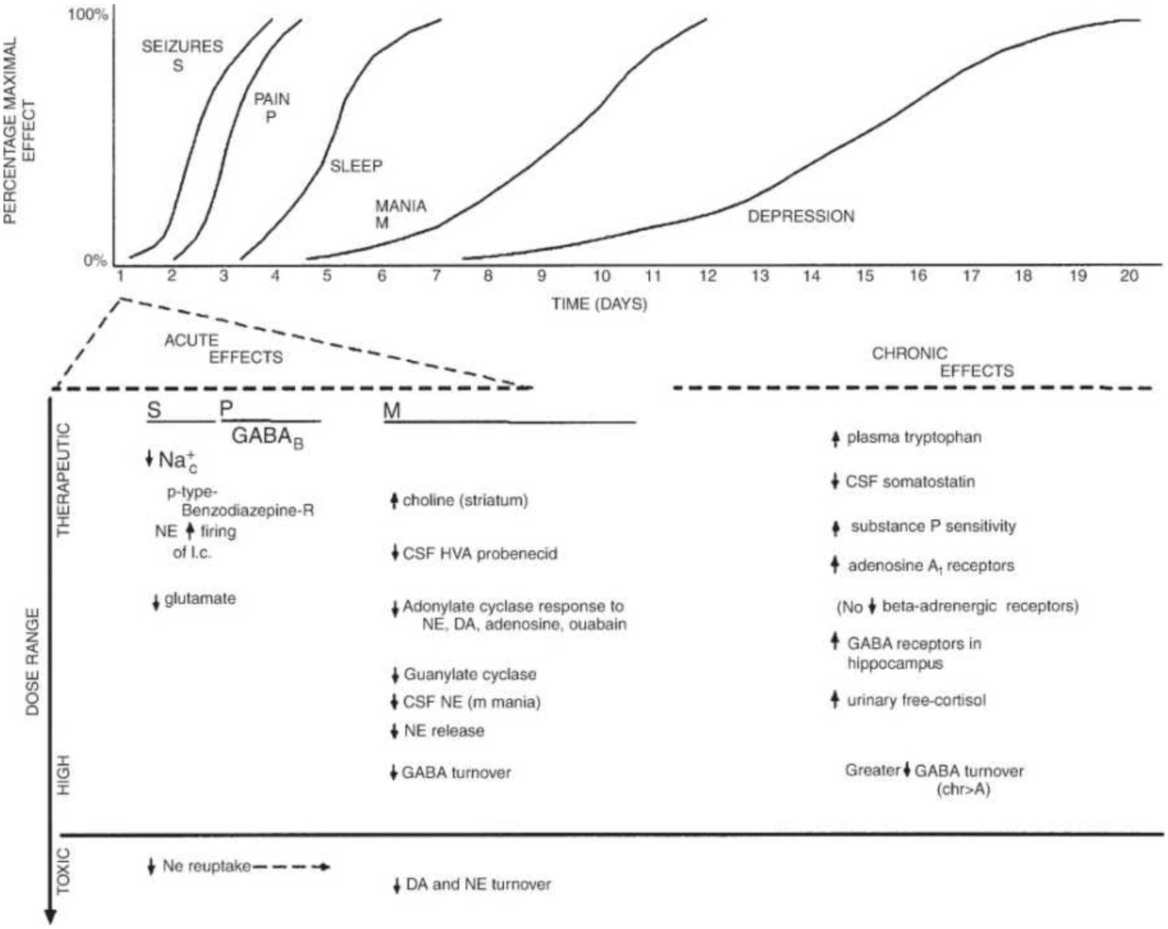
PERCENTAGE MAXIMAL EFFECT
100%
0%
SEIZURES S
PAIN P
SLEEP
MANIA M
DEPRESSION
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
TIME (DAYS)
ACUTE EFFECTS
CHRONIC EFFECTS
DOSE RANGE
THERAPEUTIC
HIGH
TOXIC
S
P
GABA_B
M
↓ Na⁺c
p-type-Benzodiazepine-R
NE ↑ firing of l.c.
↓ glutamate
↑ choline (striatum)
↓ CSF HVA probenecid
↓ Adonylate cyclase response to NE, DA, adenosine, ouabain
↓ Guanylate cyclase
↓ CSF NE (m mania)
↓ NE release
↓ GABA turnover
↑ plasma tryptophan
↓ CSF somatostatin
↓ substance P sensitivity
↑ adenosine A1 receptors
(No ↓ beta-adrenergic receptors)
↑ GABA receptors in hippocampus
↑ urinary free-cortisol
Greater ↓ GABA turnover (chr>A)
↓ Ne reuptake
↓ DA and NE turnover

FIGURE 33.10c–2. Schematic and time course of clinical and biochemical effects of carbamazepine. Seizure(s) and paroxysmal pain(p) syndromes respond rapidly to carbamazepine, whereas maximal effects in mania(m) and depression are delayed 2 to 3 weeks or more, respectively. The acute effects observed in animal and humans are probably related to anticonvulsant and antinociceptive efficacy, whereas chronic effects are more likely to be related to antidepressant mechanisms. Candidate mechanisms for each clinical effect are listed according to their occurrence at clinically relevant concentrations. Chr>A, chronic greater than acute; CSF, cerebrospinal fluid; DA, dopamine; GABA, γ-aminobutyric acid; HVA, homovanillic acid; l.c., locus ceruleus; NE, norepinephrine.

In addition to its blockade of glutamate release, carbamazepine is also potent at adenosine A_1 receptors. Its agonist properties at these receptors could account for not only some of its anticonvulsant properties but also its sedative properties.

It is also of interest that a variety of mood stabilizers beyond carbamazepine, such as lithium, valproate, and lamotrigine, are able to weakly block calcium influx through glutamate receptors of the NMDA subtype. Thus, the possibility remains that carbamazepine could exert relatively antiglutamatergic effects both by decreased release of glutamate and by relative decreases in glutamate’s postsynaptic efficacy by inhibiting calcium influx. Glutamate NMDA receptor antagonist actions, in general, have been increasingly linked to potential antidepressant (ketamine) and possible mood-stabilizing effects (memantine), but further investigation of this possible linkage to carbamazepine’s actions is required.

In common with both lithium and valproate, carbamazepine blocks inositol transport at the myoinositol transporter. Other common and differential effects of carbamazepine compared with lithium and valproate are illustrated in Figure 33.10c–4.

The antinociceptive effects of carbamazepine have been most closely linked to effects at $GABA_B$ or baclofen-like receptors. Carbamazepine and baclofen share structural similarities, and the antinociceptive actions of both drugs are specifically inhibited by $GABA_B$ antagonists. It is also of interest that chronic but not acute administration of lithium, carbamazepine, and valproate are all associated with upregulation of $GABA_B$ receptors in the hippocampus, yielding this action as a potential convergent mechanism for mood stabilization. In contrast, unimodal antidepressants appear to upregulate $GABA_B$ receptors in the frontal cortex.

Carbamazepine also exerts a variety of interesting effects on aminergic and indolaminergic systems. Carbamazepine is a very weak blocker of norepinephrine reuptake, which is probably not clinically relevant, but it does decrease the induced release of norepinephrine at clinically relevant concentrations. These preclinical findings converge with carbamazepine’s ability to decrease CSF norepinephrine in manic patients.

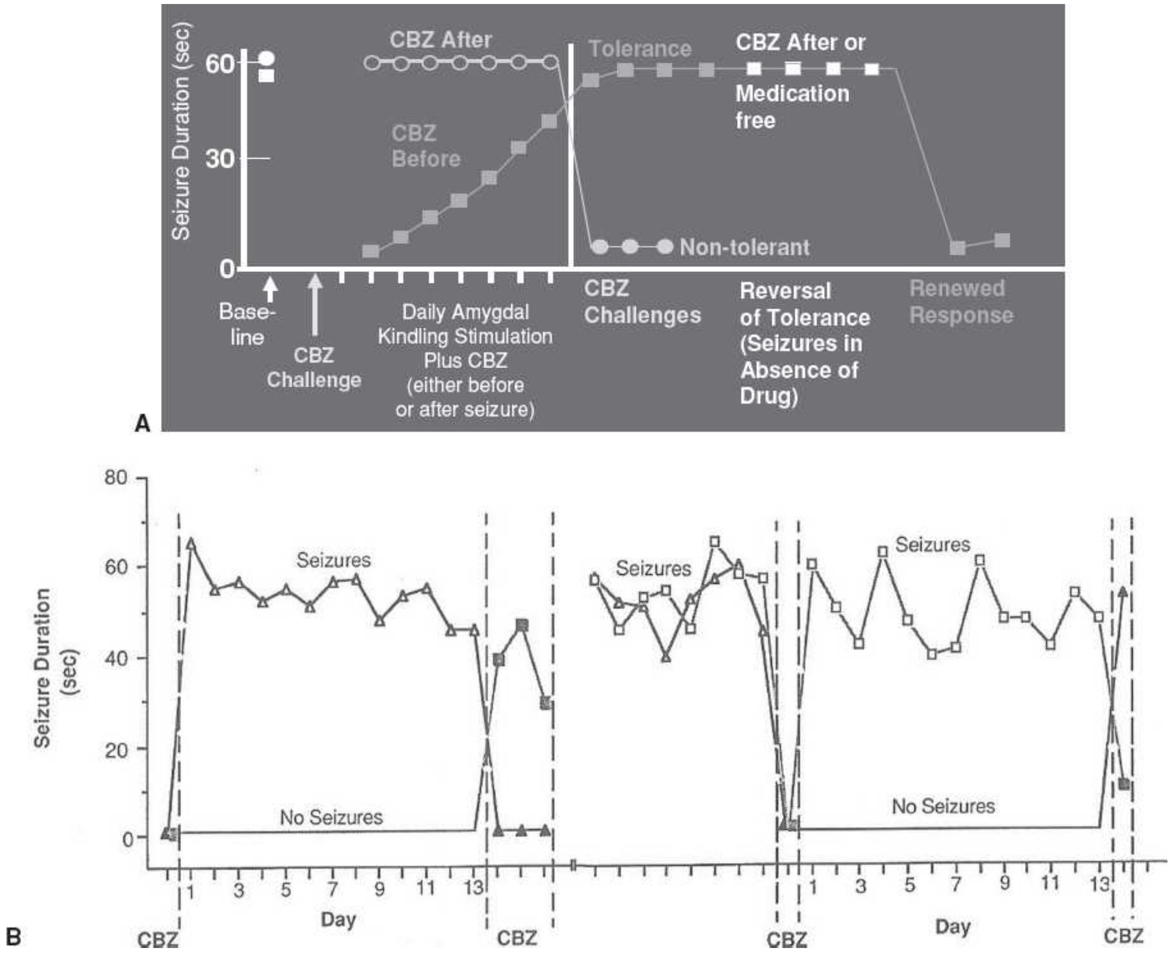

FIGURE 33.10c–3. **A:** Schematic illustration of contingent tolerance to carbamazepine (CBZ): development and reversal. In fully kindled animals (*open circles, squares*), carbamazepine treatment inhibits kindled seizures (*filled squares*). Repeated drug administration before (*filled squares*), but not after (*filled circles*), stimulation results in tolerance development. Tolerance induced in this matter (*filled squares*) can be reversed by a period of kindled seizures with drug administration after each seizure (*open squares, dotted lines*). **B:** Seizures enable carbamazepine's anticonvulsant effect. *Left side:* In fully kindled animals, CBZ was administered (15 mg/kg, i.p.) and produced complete seizure suppression (CBZ-first day). Animals were then divided into two groups that received either 13 daily amygdala-kindled seizures (at 800 μA [*open triangles*]) or no treatment. Following this, both groups of animals received 3 days of kindling stimulation preceded by CBZ administration on each day (15 mg/kg). The rats that did not experience seizures for the 13-day interval showed only a partial response to CBZ (*closed squares*), whereas the rats that underwent seizures again showed complete seizure suppression by CBZ. Both groups were then given repeated seizures (*open squares and triangles*) to reinstate the original CBZ effect. *Right side:* A crossover design was used to replicate the initial finding. Animals that previously were given no treatment now received 13 days of seizures (*open squares*) and vice versa.

> Again, and even more robustly, the animals that did not experience seizures for 13 days showed no anticonvulsant response to CBZ administration (day 14), whereas the animals experiencing seizures showed almost complete seizure suppression.

Actions at noradrenergic α_2 receptors are of special interest in that the α_2 agonist in clonidine (Catapres) reverses some of carbamazepine's effects in some seizure models, whereas, paradoxically, the α_2 antagonist yohimbine reverses carbamazepine's anticonvulsant effects in other models. In epilepsy-prone animals, carbamazepine's ability to increase serotonin in the hippocampus appears to be important, as depletion of serotonin by several different mechanisms is associated with a loss of carbamazepine's anticonvulsant effects in this model.

Consistent with other mood-stabilizing drugs (lithium and valproate), carbamazepine is also able to decrease dopamine turnover, an action that could be related to its antimanic and mood-stabilizing properties. In contrast to the antipsychotics that directly block dopamine receptors, carbamazepine is not a direct antagonist at dopamine receptors and exerts its actions on the dopamine system through some other series of mechanisms. Dopamine levels have been reported to be increased in the prefrontal cortex after chronic carbamazepine administration. Carbamazepine decreases levels of the dopamine metabolite homovanillic acid (HVA) in the CSF of affectively ill patients following probenecid administration, consistent with its ability to decrease dopamine turnover in animals.

Taken together, these data suggest that different mechanisms of carbamazepine's actions could be related differentially to its anticonvulsant efficacy in different seizure models. To the extent that this is clearly delineated in epilepsy research, as, for example, in relationship to α_2 receptors, it raises the possibility that different components of carbamazepine's actions could also be related to different aspects of its efficacy and side effect profile in the affective disorders.

In addition to those mentioned above, carbamazepine's actions on a variety of neuropeptides appear to require chronic administration and, thus, could be pertinent to antidepressant actions. Chronic, but not acute, drug administration increases substance P levels and sensitivity in the striatum, findings of particular interest, as they are shared by lithium.

Substance P antagonists have been studied in depression but have a mixed record for effectiveness.

Carbamazepine decreases levels of somatostatin in CSF, in contrast to nimodipine (Nimotop; Nymalize), which increases CSF somatostatin. These differential effects of carbamazepine and nimodipine exist not only in terms of their blockade of calcium influx via NMDA-type glutamate receptors versus L-type calcium channels, respectively but also in terms of their physiologic effects on brain metabolism, as assessed by PET. Ketter and associates have found a differential relationship of metabolism in paralimbic areas, particularly in the left insula, to the degree of antidepressant response to carbamazepine, with higher baseline levels of activity being associated with better antidepressant response. In contrast, low levels of metabolism in the left insula are associated with a better antidepressant response to the dihydropyridine calcium channel blocker nimodipine. In both instances, response to the drug was associated with a relative normalization of the baseline abnormalities in the activity of the insula and related paralimbic areas of the brain.

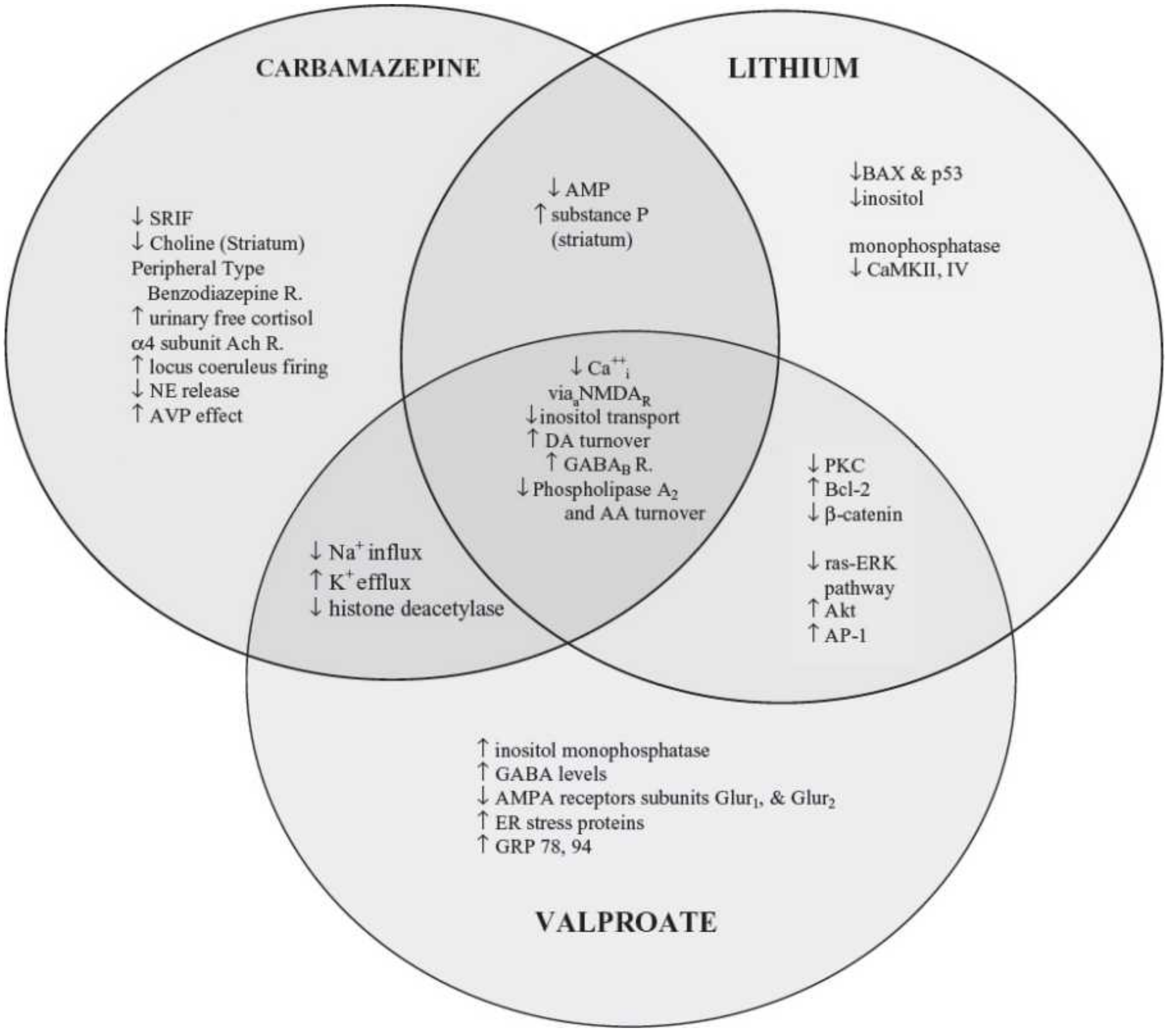
CARBAMAZEPINE
LITHIUM
VALPROATE
↓ SRIF
↓ Choline (Striatum)
Peripheral Type Benzodiazepine R.
↑ urinary free cortisol
α4 subunit Ach R.
↑ locus coeruleus firing
↓ NE release
↑ AVP effect
↓ AMP
↑ substance P (striatum)
↓BAX & p53
↓inositol
monophosphatase
↓ CaMKII, IV
↓ Ca++i via NMDA R
↓inositol transport
↑ DA turnover
↑ GABAB R.
↓ Phospholipase A2 and AA turnover
↓ PKC
↑ Bcl-2
↓ β-catenin
↓ ras-ERK pathway
↑ Akt
↑ AP-1
↓ Na+ influx
↑ K+ efflux
↓ histone deacetylase
↑ inositol monophosphatase
↑ GABA levels
↓ AMPA receptors subunits Glur1, & Glur2
↑ ER stress proteins
↑ GRP 78, 94

FIGURE 33.10c–4. Overlapping, differential, and unique effects of mood stabilizers. SRIF, somatostatin; R, receptors; NE, norepinephrine; AVP, arginine vasopressin; NMDA, *N*-methyl-D-aspartate; cAMP, cyclic adenosine monophosphate; PKC, protein kinase C; GABA, γ-aminobutyric acid; ER, endoplasmic reticulum.

Carbamazepine's antidepressant effects may be mediated by those that are atypical of many other antidepressant modalities. For example, most antidepressants decrease β-adrenergic receptors in the prefrontal cortex, whereas carbamazepine upregulates them. Most antidepressants upregulate mineralocorticoid and glucocorticoid receptors and, thus, decrease cortisol, whereas carbamazepine increases plasma and urinary free cortisol. This unusual profile of carbamazepine may explain why it is effective in some depressed patients with affective syndromes that are resistant to more traditional antidepressant modalities. A possible link of carbamazepine to antidepressant response in depressed and PTSD patients with low baseline cortisol or increased dexamethasone suppression has not been adequately studied, however.

As noted earlier, carbamazepine has both convergent and contrasting effects with other mood stabilizers, such as lithium and valproate. Whereas lithium and carbamazepine block cAMP and G proteins, their effects on phosphoinositide turnover diverge. Lithium inhibits several inositol phosphatases, whereas valproate does not, and carbamazepine may enhance their activity. This contrasts with the common effects of the three drugs in blocking the inositol transporter and in decreasing arachidonic acid turnover.

Carbamazepine, as well as the mood stabilizers lithium, valproate, and most antidepressant modalities, increase BDNF. This action may be of considerable importance because the Val66 met allele of pro-BDNF and early life adversity can reduce hippocampal volume and yield low levels of BDNF in the hippocampus and cortex in adulthood. Stressors, affective episodes (both mania and depression), and substances of abuse all can lower BDNF in the hippocampus or serum of bipolar patients. As such, effective prophylaxis will indirectly prevent episode-related BDNF decrease, and mood stabilizer treatment may also directly increase BDNF and help correct persisting BDNF decreases.

Whereas lithium is a blocker of vasopressin effects via its inhibition of the second messenger cAMP, carbamazepine appears to exert indirect facilitatory effects at the vasopressin receptor. Therefore, because carbamazepine’s effects are at the level of the receptor and lithium’s blockade is below the receptor (in the transduction system), carbamazepine is unable to reverse the effects of lithium on the development of diabetes insipidus.

Carbamazepine, like phenytoin and lamotrigine, is not effective in blocking the initial developmental phase of electrical kindling of the amygdala but is a highly effective anticonvulsant in the midphase of full-blown kindled seizures. In contrast, chronic treatment with carbamazepine does block the initial or developmental phase of pharmacologic kindling with the local anesthetics lidocaine (Xylocaine) or cocaine. Chronic carbamazepine should thus make the development of cocaine-related seizures in humans less likely as well. This effect could contribute to carbamazepine’s utility in helping with cocaine abstinence, an effect that has only been demonstrated compared with placebo in those with concomitant affective disorders.

Mechanisms of Tolerance Development

It is of interest that carbamazepine is acutely effective against completed amygdala-kindled seizures, but tolerance rapidly develops to this effect with repeated daily administration (Fig. 33.10c–3A). This form of tolerance is called contingent tolerance, as it only develops when the drug is given before the amygdala stimulation, but not when the drug is given immediately after the stimulation. In this way, tolerance is contingent on having the drug present during the seizure-inducing stimulation and is clearly based on pharmacodynamic and not pharmacokinetic mechanisms. This tolerance that develops to the anticonvulsant effects of carbamazepine may provide an indirect but useful model for considering mechanisms potentially associated with clinical tolerance development in seizure and mood disorders and ways of slowing its development or reversing it once tolerance has occurred.

A gradual loss of efficacy after many years of successful treatment has been reported in a subgroup of patients via an apparent pharmacodynamic tolerance mechanism in patients with: paroxysmal pain syndromes such as trigeminal neuralgia, the prophylaxis of bipolar illness (Fig. 33.10c–5), and to a lesser extent in clinical epilepsy syndromes. Some of the ways of slowing tolerance development in the preclinical seizure model could be tested clinically for their relevance to human therapeutics, either in epilepsy patients specifically or in other syndromes such as bipolar disorder as well.

As noted above, the tolerance that develops to carbamazepine's anticonvulsant effects on amygdala-kindled seizures is not related to pharmacokinetics or a drop in carbamazepine blood levels. This can be demonstrated by the fact that tolerance only develops when the drug is given to the animal just before each amygdala stimulation, but not when the same dose of the drug is administered immediately after each kindled seizure has occurred. Animals administered daily carbamazepine just after each seizure has occurred (and drug levels again become negligible prior to the next stimulation the next day) never become tolerant. In this fashion, tolerance development is contingent on the drug's being in the brain concurrently with the amygdala stimulation. This type of contingent tolerance can be reversed either by a period off the drug (and the occurrence of seizures in the medication-free state) or, remarkably, by a period of continuing to give the drug daily, but now after (rather than before) each of a series of several once-daily seizures has occurred.

The neurobiology of this anticonvulsant tolerance development has been preliminarily explored in the kindled seizure model and appears to involve differential regulation of a variety of seizure-induced adaptations at the level of gene expression. Seizures ordinarily induce a variety of neurotransmitter and receptor effects that have been linked either to the primary pathology of kindled seizure progression or to more short-lived secondary compensatory or endogenous anticonvulsant effects (Table 33.10c–3, left column). For example, thyrotropin-releasing hormone (TRH) peptide and its messenger ribonucleic acid (mRNA) are increased for approximately 4 days after an amygdala-kindled seizure. However, TRH does not increase after a seizure that occurs when animals are tolerant to carbamazepine.

TRH is likely an anticonvulsant adaptation based on the finding that its bilateral injection of TRH into the hippocampus decreases amygdala-kindled seizures. Further supporting this perspective are the observations that after a period of 4 or more days after the last amygdala-kindled seizure, endogenous TRH levels and other putative endogenous anticonvulsant adaptations dissipate and carbamazepine loses its anticonvulsant effectiveness (even in nontolerant animals) (Fig. 33.10c–3B; day 14). These data suggest that TRH is one example of many endogenous substances that are induced postictally in an attempt to prevent further seizures from occurring and, at the same time, are also required for carbamazepine's efficacy. Thus, the failure of seizures to induce TRH and other anticonvulsant peptides in carbamazepine tolerant animals could be relevant to this loss of efficacy. TRH also has antidepressant effects when administered intravenously or intrathecally to depressed patients. Therefore, it remains a possibility to be further studied that loss of depressive-episode induced increases in TRH could be related to clinical tolerance development in mood disorders.

Similarly, in carbamazepine-tolerant animals, the breakthrough seizures selectively fail to induce other endogenous adaptations, such as increases in $GABA_A$ receptors, although increases in benzodiazepine receptors continue to occur (Table 33.10c–3, right columns). The inability of GABA receptors to show their normal increase after a seizure is also observed at the level of the $GABA_A$ receptor subunit mRNAs. The mRNA for the α_4 subunit is one of several subunits that ordinarily increase after seizures, but it fails to do so in the carbamazepine-tolerant state. Thus, in addition to an anticonvulsant neuropeptide such as TRH, transient seizure-induced increases in other inhibitory neurotransmitter receptors, such as those of $GABA_A$, and its α_4 subunit are also selectively suppressed during carbamazepine tolerance, which could contribute to its loss of anticonvulsant efficacy. That is, some of the needed episode-induced adaptations fail to occur.

These are just two examples of several neural systems that are selectively impaired after seizures in carbamazepine-tolerant animals but are consistently induced by seizures in nontolerant animals that have been exposed to the same number of prior seizures and the same amount of carbamazepine (but given after the seizure has occurred) (Fig. 33.10c–3A). Also supporting this view that seizure-induced adaptations are important to carbamazepine's efficacy is the observation that carbamazepine again becomes effective in tolerant animals after a series of seizures are given in the absence of medication (which then re-induce the full range of adaptive effects on gene expression).

Clinical case studies suggest that this renewal of efficacy may also be seen after a period of time off medications in the instance of clinical tolerance development in seizure or mood disorders, but it would be preferable to find a new treatment agent that did not show cross-tolerance to carbamazepine so as to avoid the necessity of experiencing further episodes while off medication.

In the preclinical model of anticonvulsant tolerance to carbamazepine on amygdala-kindled seizures, cross-tolerance is observed to the anticonvulsant effects of PK11195 (a peripheral type of benzodiazepine antagonist), lamotrigine, and, surprisingly, valproate and levetiracetam (Keppra), but not to diazepam or phenytoin. How these cross-reactivities for seizure tolerance in the rodent relate to those in the clinical condition of recurrent affective disorders remains to be further studied. Finding ways to address clinical tolerance development is potentially important because a substantial group of treatment-refractory patients (perhaps some 20% to 40%) who initially experience an excellent prophylactic response to carbamazepine begin to show breakthrough episodes consistent with a tolerance process, which often begins after a period of 2 to 3 years of effective treatment (see years 1986 to 1988 in Fig. 33.10c–5).

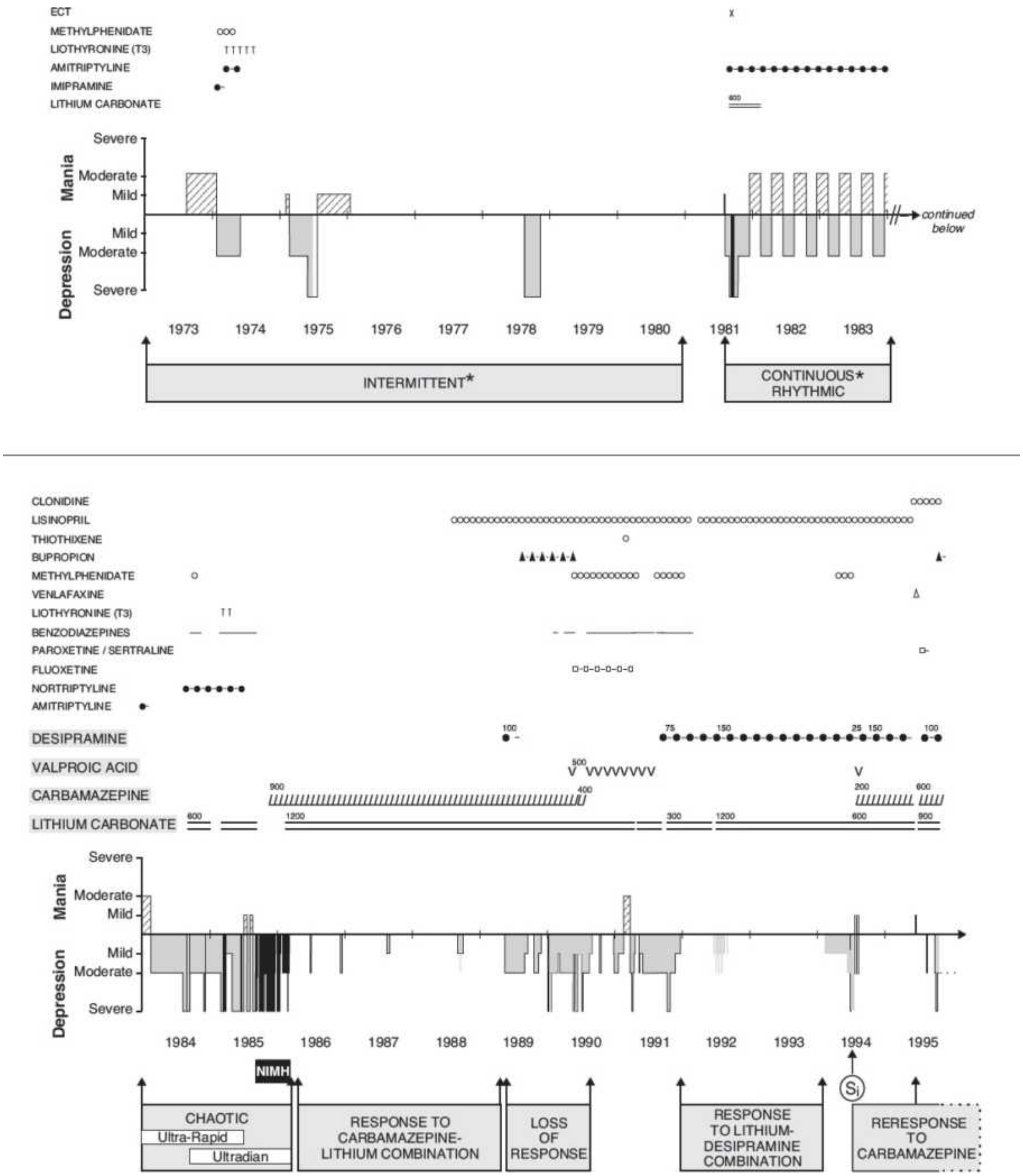

FIGURE 33.10c–5. Life chart showing the phases in evolution and treatment response in a female patient with bipolar II disorder. This life chart illustrates the phenomena of illness progression, probable antidepressant-induced cycle acceleration (m 1981), lack of response to lithium in a rapid cycler (1984–5), response to the combination therapy of lithium and carbamazepine, development of loss of responsiveness to this treatment regimen via tolerance (1989), and cross-tolerance to valproate (1990). Key: Lithium (= = = =); carbamazepine (//////); valproic acid (VVVV); tricyclics (-.-.-.); benzodiazepines (–.–.–.); thyroid (TTTT); other (oooo). CBZ, carbamazepine; Dep, depression; ECT, electroconvulsive therapy; VPA, valproic acid.

Interestingly, as in the seizure model of cross-tolerance between carbamazepine and valproate, the patient illustrated in Figure 33.10c–5 failed to respond to valproate when she was becoming tolerant to carbamazepine. Further exploration of ways to circumvent tolerance development and ascertain which drugs do not possess cross-tolerance is of considerable interest. A number of approaches have been examined that slow tolerance development in the amygdala-kindled seizure model. These include: initially using sustained higher rather than marginal doses of the drug; combining carbamazepine with valproate; treating earlier rather than later in the course of kindling, and reducing the stimulation intensity (or lowering illness drive). Whether some of these are also pertinent to slowing tolerance development during clinical treatment of seizure and mood disorders remains to be studied systematically, but they are listed here for their potential heuristic value.

Table 33.10c–3.
Selective Failure of Some Kindled Seizure-Induced Neurochemical Changes in CBZ-Contingent Tolerance

Seizure-Induced Adaptation (in Nontolerance Animals)	Seizure-Induced Adaptations Present in CBZ-Tolerant Animals	Seizure-Induced Adaptations Absent in CBZ-Tolerance Animals
↑ c-fos mRNA		No ↑ c-fos
↑ DZP receptors	↑ DZP-R	
↑ GABA receptors (3H) muscimol		No ↑ GABA-R
↑ Subunit		No ↑ subunit
↑ Subunit	↑ Subunits	
↑ TBPS binding		No ↑ TBPS
↑ Glucocorticoid-R mRNA		No ↑ Glucocorticoid-R
↑ Mineralocorticoid-R mRNA	↑ Mineralocorticoid-R	
↑ BDNF mRNA		No ↑ BDNF
↑ NT_3 mRNA	↑ NT_3	
↑ TRH mRNA		No ↑ TRH
↑ CRH mRNA		No ↑ CRH
↑ CRH-BP mRNA		No ↑ CRH-BP
↑ NPY mRNA		No ↑ NPY
↑ Enkephalin mRNA		No ↑ Enkephalin
↑ Dynorphin mRNA	↑ Dysmorphia	

Treated with CBZ after each daily amygdala kindling stimulation; nontolerance animals were matched for amount of drug and number of seizures seen in tolerant animals.

CBZ, carbamazepine; DZP, diazepam; GABA γ-aminobutyric acid; TBPS (35S) t-butylbicyclophosphorothionate; BDNF, brain-derived neurotrophic factor; NT_3, neurotrophin-3; TRF, thyrotropin-releasing hormone; CRH, corticotropin-releasing hormone; CRH BP, corticotropin-releasing hormone-binding protein.

CLINICAL DRUG STUDIES

Acute Mania

Numerous double-blind studies of various designs have compared carbamazepine to placebo or other active agents that indicate antimanic efficacy. Two large placebo-controlled, parallel-group design trials have recently revealed that the long-acting formulation of carbamazepine (Equetro) was significantly more effective than a placebo in acute mania. Similar acute antimanic effects of carbamazepine that roughly parallel those of lithium and antipsychotics are illustrated in Figure 33.10c–6.

Moreover, acute antimanic responsivity has been documented in double-blind, off-on–off-on study designs, as illustrated in Figure 33.10c–7, in a patient who was a clear nonresponder to lithium, as revealed by a breakthrough episode during carefully monitored lithium treatment upon hospital discharge. The patient was readmitted in a full-blown manic episode despite lithium treatment and showed a third excellent re-response to active carbamazepine and, subsequently, has been able to remain out of the hospital during long-term prophylaxis with carbamazepine as supplemented by other agents. Response to carbamazepine but not to valproate has also been seen in some patients (Figs. 33.10c–3 and 33.10c–8). Other patients with bipolar disorder have shown the opposite pattern of clear-cut responses to valproate but not carbamazepine, emphasizing the principle that response to one anticonvulsant is not predictive of response to another.

Acute Depression

Fewer studies have addressed the acute antidepressant effects of carbamazepine, but several are highly suggestive of efficacy. One large parallel-group study of carbamazepine compared with placebo has been completed recently by Zhang and associates (Table 33.10c–4). These authors found 50% reductions in Hamilton Depression Rating Scale scores in 34.8% on placebo (*N* = 23), 63.8% on carbamazepine (*N* = 47), and 84.8% on carbamazepine plus a Chinese herbal preparation (Free and Easy Wanderer Plus [FEWP]) (*N* = 46; *P* < .05), further supporting the acute antidepressant effects of the drug. Zhang et al. have also demonstrated acute antidepressant effects of carbamazepine compared with placebo in unipolar depression.

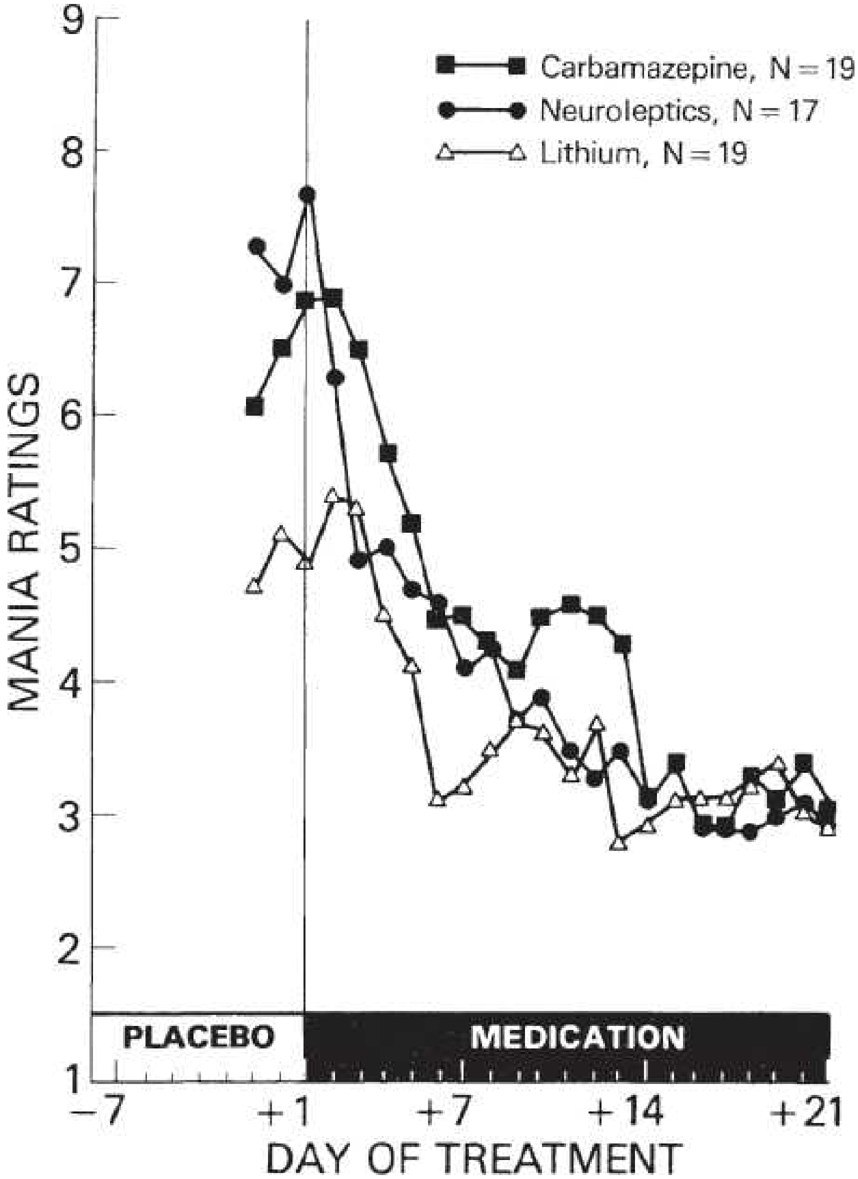

FIGURE 33.10c–6. Time course of double-blind acute antimanic effects of carbamazepine compared to lithium and neuroleptics. The incidence and time course of antimanic effects of carbamazepine appear comparable with those achieved by neuroleptics, although several studies have suggested a slightly faster onset of efficacy of neuroleptics in the first 3 days of treatment, a difference that was gone by 7 days. Even patients requiring seclusion for acute, fulminant, psychotic manic episodes may rapidly improve after initiation of treatment to carbamazepine alone.

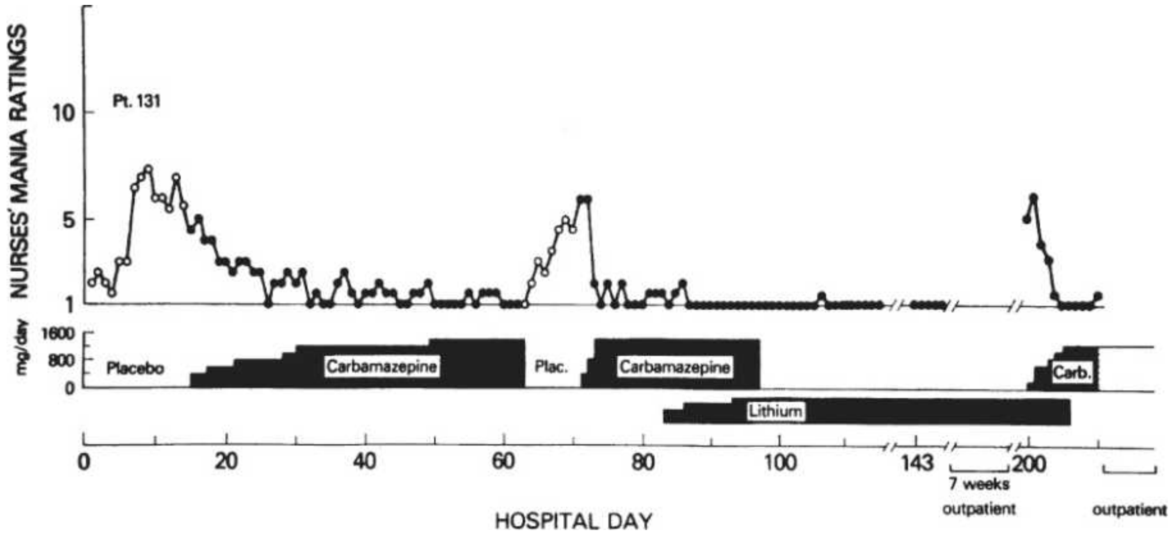

FIGURE 33.10c–7. Unequivocal antimanic response to carbamazepine in a double-blind, off-on–off-on clinical trial, a subsequent manic episode on lithium, and a third "on" to carbamazepine.

Moreover, the authors have observed that a group of 10 patients at the National Institute of Mental Health (NIMH) who were initial good acute antidepressant responders to carbamazepine on a double-blind basis subsequently relapsed on placebo and then re-responded during a second blind treatment trial with carbamazepine (Fig. 33.10c–9). This off-on-off–on documentation of responsivity indicates that carbamazepine has unequivocal antidepressant effects in a subgroup of highly treatment-refractory depressed patients. While good acute antidepressant effects occurred in about one-third of patients exposed to carbamazepine, the percentage of antidepressant response to this agent and the particular characteristics and predictors of response remains to be further delineated.

Initial data summarized in Table 33.10c–5 suggest that those with more severe episodic and less chronic depressions were among the best responders to the drug, as were those with low baseline CSF HVA, the major metabolite of dopamine. Carbamazepine responders also showed a pattern of baseline hypermetabolism on PET scans that is unusual and contrasts with the more typical hypofrontality of major depression. Some data also suggest that those with a history of depression in association with prior alcoholism are among those who respond well to the drug. This is of considerable interest because carbamazepine can block the effects of alcohol withdrawal in experimental animals, and the drug is widely used in the acute and long-term treatment of those with alcohol-related dysphoria in Scandinavian countries.

Prophylaxis

More than a dozen partially controlled studies (involving either randomization or double-blind comparison with other agents such as lithium) reveal that carbamazepine is approximately as effective as lithium in the prevention of affective episodes, although several caveats remain. In these controlled studies of carbamazepine prophylaxis, 224 of 382 (59%) showed a moderate to marked response to carbamazepine. Similarly, lithium response in these same studies occurred in 241 of 395 patients (61%). However, in several studies, there was better antimanic efficacy of lithium compared with carbamazepine.

The German collaborative study indicated several different predictors of response to these two agents. Lithium was substantially more effective than carbamazepine in those with a classic bipolar I diagnosis, with euphoric mania, no thought disorder or comorbid substance abuse, as well as (in other studies) a positive family history of affective disorder. Conversely, there was a trend toward better prophylactic efficacy of carbamazepine in those with bipolar II illness, schizoaffective diagnosis, dysphoric mania, and substance abuse comorbidity (current or prior) as well as (in other studies) a negative family history of bipolar disorder in first-degree relatives.

Case studies performed in a double-blind fashion have also documented that some patients respond to carbamazepine, and they fail to respond to divalproex sodium (Depakote) (Fig. 33.10c–8B) or vice versa. These observations are consistent with the notion that response to the anticonvulsants does not occur as a class effect and that further elucidation of the clinical and biologic predictors of response to carbamazepine would be of great clinical importance.

It is noteworthy that within individual patients, there appears to be dose-related antimanic and antipsychotic responsiveness (Fig. 33.10c–8C). However, there is little evidence of a relationship between blood levels and the degree of clinical antidepressant or antimanic response across individuals (Fig. 33.10c–10).

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Central Nervous System

Typical of many anticonvulsant drugs, carbamazepine can exert a variety of untoward effects on CNS function. Most commonly, these include dizziness, ataxia, and diplopia (characteristically occurring approximately 2 hours after a given dose of the regular immediate-release formulation). These common side effects can largely be avoided with slow upward titration of the dose to just below a patient's side-effects threshold, which can vary greatly, as illustrated in Figure 33.10c–10. Given this wide range of individual proclivity to side effects (Fig. 33.10c–11) and the lack of a clear anticonvulsant dose–response relationship, the importance of individualizing dose titration is emphasized as preferable to attempting to achieve a set blood level within the conventional 4 to 12 μg/mL "therapeutic" window.

In overdose, carbamazepine produces drowsiness, stupor, coma, sinus tachycardia, atrial ventricular conduction blocks, hypotension or hypertension, seizures, nystagmus, and hyporeflexia or hyperreflexia. Hypothermia, oral dyskinesias, and respiratory depression can also occur, yet few fatalities have been reported even after ingestion of as high as 10 to 20 g of carbamazepine.

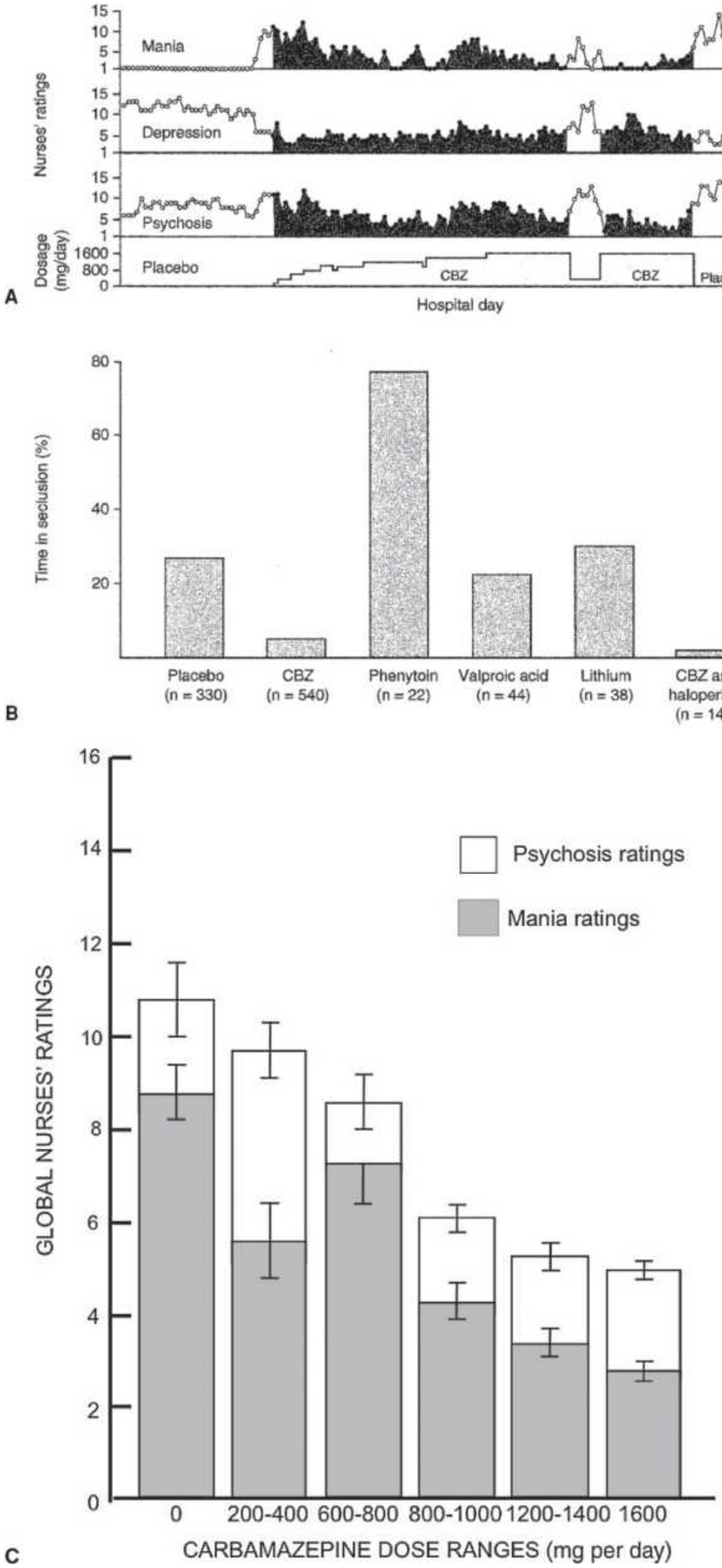

FIGURE 33.10c–8. Efficacy of carbamazepine (CBZ) in preventing manic and depressive episodes in a patient whose symptoms were previously refractory to lithium and antipsychotic treatment **(A).** CBZ administered on a blind basis

acutely suppressed manic and psychotic symptoms. During periods of CBZ dose reduction or placebo substitution, exacerbation of manic and psychotic symptoms was consistently noted and reinstitution of active CBZ treatment led to improvement on three different occasions. Each rating circle is one hospital day **(B).** The prophylactic effect of CBZ during double-blind, clinical trials in a treatment-refractory, rapidly cycling patient who required seclusion 26% of the time (based on each 24-hour day) while taking placebo or 30% of the time while taking lithium, but only 4% of the time while taking CBZ (*n* = 540 days) at doses greater than 800 mg per day. The patient was not responsive to two other antiepileptic drugs, phenytoin (seclusion 77%) or valproic acid (seclusion 22%), but did exceptionally well and rarely required seclusion (1.6%) while treated with the combination of CBZ and haloperidol. *n,* number of days in treatment. **C:** Dose-dependent therapeutic effects of carbamazepine during a prolonged manic episode in the same patient.

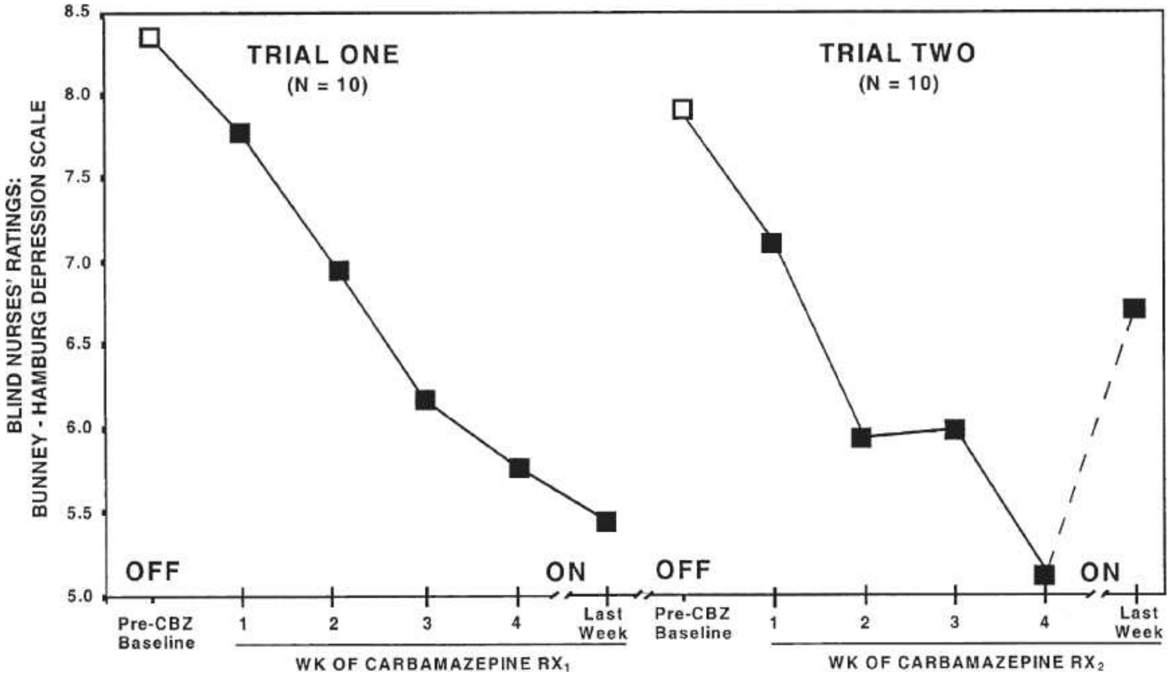

FIGURE 33.10c–9. Confirmation of antidepressant response to carbamazepine (CBZ) in an off-on–off-on design in a subgroup of ten treatment-refractory patients. Seventeen of the 54 depressed patients studied in a blind fashion (31.5%) showed an initial good response (≥2 points on the Bunney–Hamberg Depression scale).

In overdose, supportive techniques include gastric lavage and instillation of activated charcoal (10 to 100 g and then 12.5 g per hour) until the patient is symptom-free. Consideration of charcoal hemoperfusion before the patient becomes hypotensive has also been recommended. Although flumazenil (Ro 15–1788) has been recommended by some, the preclinical pharmacology noted above would not support such a suggestion. Instead, the GABA_A receptor partial inverse agonist Ro 15–4513, which has been shown to reverse alcohol intoxication and is also effective in reducing the sedation and lethality of carbamazepine overdose in rodents, should be considered.

Carbamazepine's effects on memory and psychomotor performance have generally been minor, and improvement in a few indices has even been reported during treatment. However, higher doses or those that exceed an individual's threshold for neurotoxicity can be associated with impairment. A similar range of side effects can occur with oxcarbazepine, although it is generally better tolerated than carbamazepine, with the exception of an increased incidence of hyponatremia (Fig. 33.10c–1).

Table 33.10c–4.
Response[a] Rates in Bipolar Depression (Zhang et al., 2007)

LOCF Analysis

Placebo	CBZ[b]	CBZ + FEWP[b,c]
34.8%	63.8%	84.8%
(*N* = 23)	(*N* = 47)	(*N* = 46)

Observed Cases

Placebo	CBZ	CBZ + FEWP[b]
40.0%	61.1%	82.5%
(*N* = 15)	(*N* = 36)	(*N* = 40)

[a] ≥50% reduction in Ham-D scores after 12 weeks.

[b] $p < 0.05$ vs. PBO.

[c] Preparation of 11 herbal materials: 4 dry roots; 1 fresh root; 1 dry ripe fruit; 1 dry root bark; 1 peeled dry root; 1 dry root and stem; 1 dry root and leaf; and 1 dry sclerotium.

Data from Zhang ZJ, Kang WH, Tan QR, et al. Adjunctive herbal medicine with carbamazepine for bipolar disorders: a double-blind, randomized, placebo-controlled study. *J Psychiatr Res.* 2007;41(3–4):360–369.

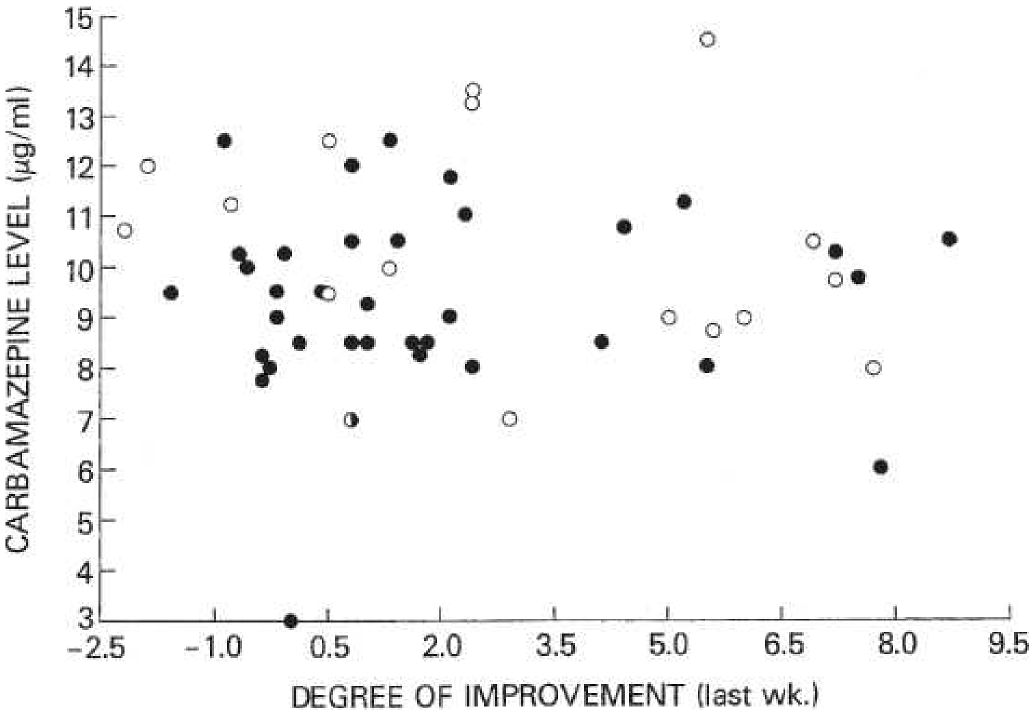

FIGURE 33.10c–10. The lack of relationship between plasma levels and degree of clinical improvement across patients in mania (*open circles*) or depression (*filled circles*).

Table 33.10c–5.
Possible Correlates of Carbamazepine Response to Treatment in the Affective Disorders

Depression

Greater acuity and severity

Low homovanillic acid levels in CSF

Higher epoxide/carbamazepine ratio

Greater decreases in thyroxine and free thyroxine decrease during carbamazepine treatment

Frontal and paralimbic (especially left insula) hypermetabolism on positron emission tomography

History of alcoholism

Comorbid posttraumatic stress disorder

Prophylaxis

Lack of family history of bipolar illness in first-degree relatives

Lack of prior rapid cycling

Nonclassical presentations:

Bipolar II, non-euphoric mania

Schizoaffective (especially depressive with mood incongruent delusions)

Substance abuse comorbidities

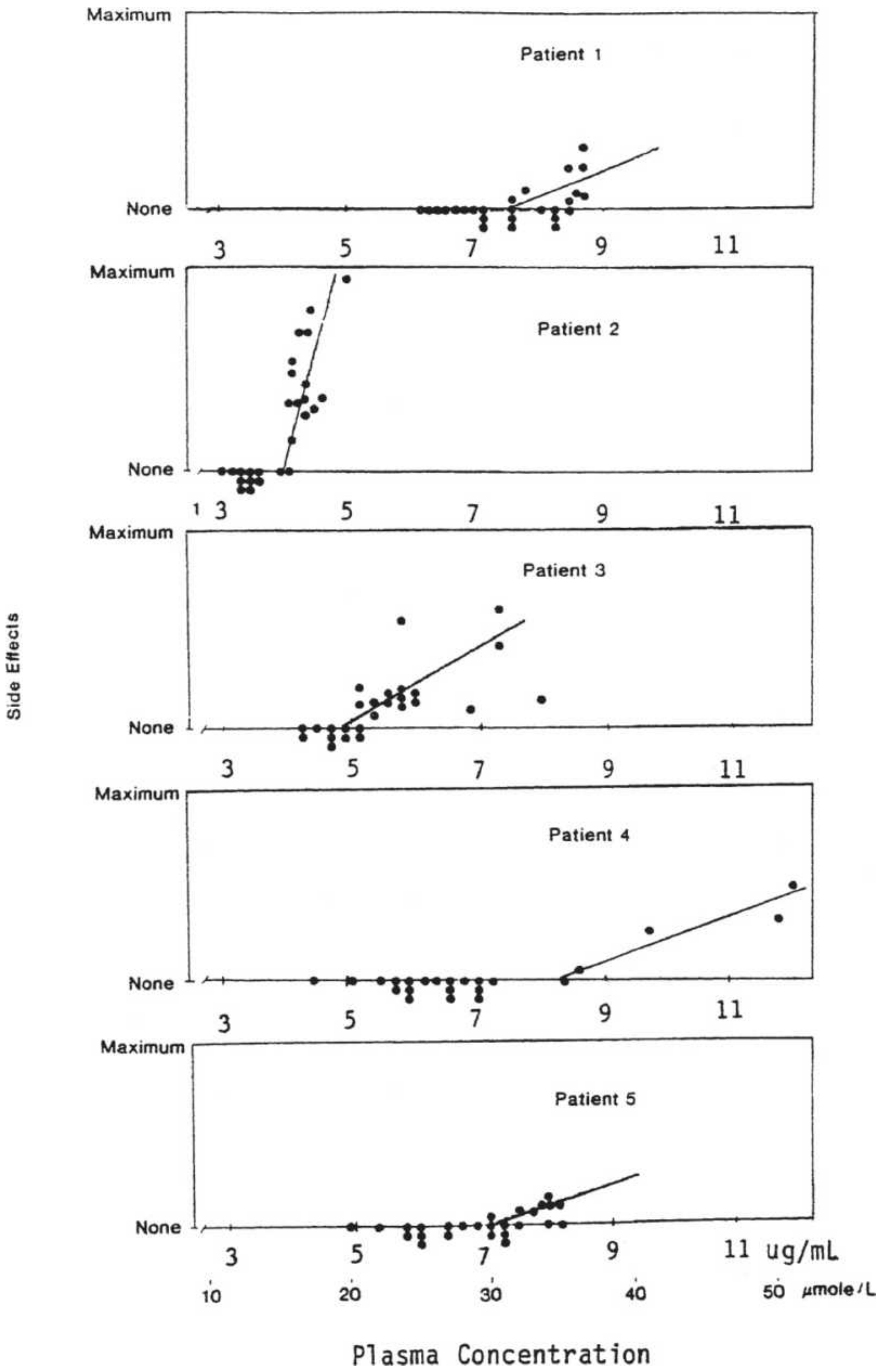

FIGURE 33.10c–11. The lack of relationship between carbamazepine concentrations and adverse effects emergence (i.e., highly individualized blood concentration to adverse-effects relationships exist within different

patients). (Adapted from Tomson T. Interdosage fluctuations in plasma concentration determine intermittent adverse effects. *Arch Neurol*. 1984;41(8):830–834. Copyright 1984, the American Medical Association.)

Table 33.10c–6.

Some Side Effects of Carbamazepine and Their Possible Management

Side Effect (Frequency, Type)	**Possible Approach (Strength of Evidence or Practice)**	**Comment**
Dizziness, ataxia, diplopia (C, D)	Reduce peak dose effects by ↓ dose (+ + +); switch to all HS sustained-release formulation (+)	Will occur 2–3 hours after IR dose if patient is near their side effects threshold (+ + +)
Fatigue, sedation (C, D)	Switch more of IR dose to HS (+ + +) or all of ER dose to HS (+); ↓ Dose or wait for adaptation or autoinduction (+ +)	Start with low HS dose; given wide individual variability in side effects threshold, titrate slowly to clinical effect or to side effects limitation (+ + +)
Benign rash (VC, I)	Discontinue drug (+); antihistamines (+); steroids (+)	Possible switch to oxcarbazepine, 2/3 no rash (+ +); or can restart CBZ slowly under coverage of prednisone (+ +)
Severe rash (R, I, S)	Discontinue drug (+ + +)	Obtain medical/dermatologic consultation (+ + +); re-challenge not recommended (even with steroids)
Benign WBC suppression (VC, D) vs. Agranulocytosis Aplastic anemia	Lower dose (+); add lithium (+ + +) Discontinue drug (+ + +) Discontinue drug (+ + +)	Lithium increases CSF (+ + +), whereas CBZ suppresses CSF (+ + +); 2,500–3,000 WBC lower limit of tolerability (±) Medical support and consultation required

		(+ + +) Medical support and consultation required (+ + +)
Weight gain (UC, I)	Consider topiramate or zonisamide augmentation (+ +) or switch to weight-neutral drug such as LTG, OXC (+)	Rarely a problem
Hyponatremia (O, D)	↓ Dose (+ +); add lithium (+ +); add demeclocycline (+ +)	Women, older age, and use of higher doses are vulnerability factors (+ + +)
Thyroid hormone suppression (VC, D)	None needed as ↓ T_4 may correlate with better response (+ +)	No ↑ TSH; no change BMR with CBZ (+); thyroid supplements rarely needed (+ + +); ↓ T_4 and ↓ T_3 may be addictive with lithium (+ +)
Tremor (R, D)	↓ Dose (+)	Check ammonia levels when combined with VPA if tremor is coarse or flapping (±)
Teratogenic (O, D) Spina bifida ≈1% Cranial facial abnormality	Pretreat with ↑ folate (+)	Avoid combination Rx if possible, especially with VPA (+); minor developmental delays do not persist
Hepatitis (R, I)	Discontinue if enzymes ↑ to three times normal (+ +) Discontinue if patient is symptomatic (+ + +)	
Memory disturbance	↓ Dose (+); switch to all	CBZ potentiation of

	HS with long-acting preparation (±)	vasopressin receptors may prevent memory loss (+/−)
High cholesterol	Treat with statins (+)	Switch to non-enzyme inducer (±)

C, common; CSF, colony stimulating factor; D, dose related; hs, at night; I, idiosyncratic; O, occasional; R, rare; S, sensitivity may cross to other antiepileptic drugs; T_3, triiodothyronine (liothyronine); T_4, thyroxine; UC, uncommon; VC, very common; WBC, white blood cell; ↑, indicates increase; ↓, indicates decrease; (+) likely works; (+ +) many case reports; (+ + +) well supported, controlled data; (±) questionable or hypothetical.

Blood

As summarized in Table 33.10c–6, the hematopoietic system is a major target of carbamazepine's benign and serious adverse actions. A statistically significant but usually clinically insignificant suppression of the white blood cell (WBC) count occurs in a substantial proportion of patients and is thought to be based on the drug's ability to inhibit colony-stimulating factor. Lithium enhances colony-stimulating factor and will reverse the benign WBC suppression of carbamazepine.

In contrast, the potentially life-threatening occurrence of agranulocytosis, pancytopenia, or aplastic anemia is a rare phenomenon and thought to occur in approximately 1 in 25,000 to 100,000 patients. Because of the rarity of these idiosyncratic events, many authorities recognize the inadequacy of routine blood count monitoring and suggest advising patients that, in the face of clinical signs of potential hematologic dysfunction (such as fever, sore throat, rash, petechiae, bruising, or overt bleeding), they should immediately contact their physician and obtain a complete blood cell count (CBC). It may be wise to monitor the CBC during the initial phases of carbamazepine administration so as to assess the degree of benign white count suppression or any evidence of more problematic changes, such as drops in platelet or red cell counts which deserve closer monitoring and attention.

In contrast to carbamazepine, the routine benign suppression of the WBC has not been reported with oxcarbazepine, and the incidence of more serious blood dyscrasias does not appear elevated over baseline.

Hormonal Effects

Carbamazepine increases the 24-hour excretion of urinary-free cortisol and notably induces a robust escape from dexamethasone suppression. These effects are thought to be not entirely related to heightened metabolism of steroid compounds (such as lowering of dexamethasone levels) and may suggest actions of carbamazepine that are contrary to those of more traditional antidepressants that downregulate the hypothalamic–pituitary–adrenal axis and increase glucocorticoid hormone receptors upon chronic administration. Whether this side effect of increasing cortisol could be clinically useful in the treatment of patients with PTSD who have abnormally low basal secretion of urinary free cortisol and show enhanced sensitivity to dexamethasone suppression compared with normal or depressed patients remains to be systematically explored.

Carbamazepine decreases circulating levels of thyroxine (T_4), free T_4, and triiodothyronine (T_3) without substantially altering thyroid-stimulating hormone. Studies indicate that it does not affect basal metabolic rate and that thyroid replacement is almost never required. In fact, patients with the largest decrements in T_4 or free T_4 showed the greatest degree of antidepressant response to carbamazepine in one series. For example, the degree of improvement in depression with carbamazepine was inversely related to the degree of decrease in free T_4 from baseline ($r = -0.56$; $N = 36$; $P = .0004$).

Renal Effects (Hyponatremia)

Carbamazepine has direct or indirect agonist-like effects at the vasopressin receptor, accounting for its occasional use in the treatment of diabetes insipidus. It will not reverse the lithium-induced variety of this syndrome, however. This carbamazepine-increased resorption of water predisposes patients to the development of hyponatremia and, in rare instances, water intoxication. Older adult patients, women, and those treated with high doses are more prone to this side effect. Several small case series suggest that carbamazepine-related hyponatremia can be prevented or reversed by the administration of lithium or demeclocycline (Declomycin).

Hyponatremia is the one area in which oxcarbazepine has a higher incidence of side effects than carbamazepine. Approximately 3% to 5% of patients treated with oxcarbazepine develop hyponatremia, and whether lithium or demeclocycline would be effective treatments is not known. The elimination half-life of oxcarbazepine is prolonged in those with renal impairment.

Skin

Carbamazepine can induce a variety of dermatologic syndromes—from a common benign macular papular pruritic rash typically emerging in the first 9 to 23 days after drug initiation to more serious conditions. Drug cessation usually results in regression of the rash, but, in rare instances, it can progress to exfoliative dermatitis, erythema multiforme, Stevens–Johnson syndrome (SJS), or toxic epidermal necrolysis (TEN). Given the inability to distinguish which rashes might evolve toward these serious endpoints, most clinicians recommend stopping the drug in the face of even the benign presentation of a carbamazepine-related rash, which can be expected to occur in some 5% to 15% of individuals. However, a substantial group of patients re-exposed to carbamazepine will not have a re-emergence of the rash. The rash rate may be lower with the new extended-release preparations.

In those individuals who appear to respond well to carbamazepine but no other agents, one can attempt to suppress the benign dermatologic reaction with prednisone (Cordrol) (40 mg per day). However, several case reports have indicated that, in the face of evidence of more serious systemic allergy (such as fever, lymphadenopathy, and pneumonitis), such steroid co-treatment may not be successful.

Oxcarbazepine appears to have a reduced frequency of inducing both the benign and serious forms of rash. Moreover, the majority of patients who have a rash on carbamazepine will not show a recurrence on oxcarbazepine, but approximately 25% to 30% will show such a cross-sensitivity.

Considerable progress has been made in studying the pharmacogenomics of serious rash (SJS/TEN) on carbamazepine. The human lymphocyte antigen (HLA-B*1502 has been associated with this syndrome in patients of Asian descent (including Han Chinese, Thai, Malaysian, Filipino, and Taiwan), where the frequency of HLA-B*1502 is about 10% to 15% in the general population. If one has this antigen, the odds ratio (OR) is 79.8. A direct interaction of HLA-B*1502 with cytotoxic T lymphocytes has been reported.

In contrast, in those of northern European descent, the risk of SJS/TEN is lower and of the order of 1 to 6 per 10,000 and is associated with HLA-A*3101 (which occurs in 2% to 5% of Europeans) with an OR of 25.9. Nine percent of Japanese have the HLA-A*3101, and it is associated with SJS/TEN OR = 10.8. In addition, preliminary evidence links HLA-B*1511 to severe rash in Japanese. There is some controversy about the clinical utility of doing these genetic screens for potential carbamazepine-treated patients, but the consensus is emerging that it would be worthwhile for those of Asian descent. In Thailand, the positive predictive value (PPV) was 1.92, while the negative predictive value (NPV) was 99.96.

There is another hypersensitivity syndrome (also called DRESS for drug reaction with eosinophilia and systemic symptoms) which occurs with a maculopapular rash (77%), increased liver enzymes (71%), eosinophilia (64.3%), fever (61%), lymphadenopathy (55%), and increased WBC (43%). Treatment with *N*-acetylcysteine (NAC) and/or intravenous immunoglobulin (IVIG) has anecdotally been reported effective, as has treatment without the use of steroids. In those with the CBZ-induced hypersensitivity syndrome, CBZ-specific CD4(+) and CD8(+) T cells have been shown to persist some years after the resolution of the syndrome.

Liver

Like other anticonvulsants, carbamazepine has been associated with enzyme elevations. However, the outright appearance of hepatitis is extremely rare. As with oxcarbazepine, routine monitoring does not appear indicated.

Carbamazepine increases total serum cholesterol by an average of 20 points. It increases both high- and low-density lipoproteins, such that deleterious cardiovascular effects have not been unambiguously demonstrated. However, in those at high risk, preventive treatment with statins may be indicated. Homocysteine increases and folate decreases have been reported, so treatment with folate may also be helpful. Vitamin D levels may also decrease on carbamazepine with or without associated increases in osteoporosis.

Weight

In epilepsy patients, long-term treatment with carbamazepine is associated with small degrees of weight gain, which are significantly less than those observed on valproate. Oxcarbazepine is reported to be weight neutral. Carbamazepine decreases plasma levels of ghrelin, which also decreases with obesity.

Heart

The cardiac effects of carbamazepine are generally not prominent, although the drug does decrease atrial-ventricular conduction and is relatively contraindicated in patients with heart block. Carbamazepine has been used to treat digitalis-induced arrhythmias, including ventricular tachycardia. Bradycardia and syncopal episodes have been reported in rare instances, with older adult women apparently at increased risk. Effects on blood pressure are typically minimal, and orthostatic hypotension is not a common accompaniment (as it is with several structurally related tricyclic antidepressants).

THERAPEUTIC INDICATIONS

Bipolar Disorder

Carbamazepine has been approved by the FDA for use in seizure disorders and trigeminal neuralgia. Like many treatments of bipolar disorder, carbamazepine is widely used in the treatment of acute mania and in long-term prophylaxis but is only FDA-approved as the long-acting preparation Equetro in acute mania. Carbamazepine is approved in many European countries, Japan, and Canada, however. It is considered a first- or second-line treatment in many guidelines and can be used as a substitute for lithium or as an adjunct to it for the partial responder (Table 33.10c–4). Use in combination with lithium appears especially helpful in rapid cyclers.

Oxcarbazepine is thought to have effects similar to carbamazepine, but these have not yet been documented in large placebo studies, and the drug is not FDA approved for bipolar illness. One study reported that overall, oxcarbazepine was no more effective than placebo in children with mania, although it did exceed placebo in the younger subgroup of children, but not in those with later adolescent-onset acute mania.

Carbamazepine can also be used in conjunction with unimodal antidepressants and antipsychotic agents, although several instances of drug–drug interactions are of some importance—for example, blood levels of nefazodone (Serzone), haloperidol, aripiprazole (Abilify), lurasidone

(Latuda), and quetiapine (Seroquel) and lamotrigine are substantially reduced. This would also occur with several of the newer atypical antipsychotics, including brexpiprazole (Rexulti), cariprazine (Vraylar), and lumateperone (Caplyta). Carbamazepine is not recommended for use in conjunction with the atypical antipsychotic clozapine (Clozaril) for fear of an additive risk of agranulocytosis, although the mechanisms for WBC suppression appear to differ. The use of carbamazepine in conjunction with valproate has not been as systematically explored in refractory mood disorders as it has in seizure disorders. The combination usually requires a decrease in the dose of carbamazepine (because valproate displaces carbamazepine from its protein-binding sites and also increases the levels of the 10,11-epoxide by virtue of its inhibition of the epoxide hydroxylase). Case reports show the efficacy of the combination of carbamazepine and valproate when either drug alone was not efficacious.

As antimanic effects are often observed within the first several days of treatment with carbamazepine, a several weeks trial in acute mania is usually sufficient to ascertain whether the patient is going to be a responder. Longer periods of evaluation may be required to assess its potential antidepressant efficacy, however. Exploration of its potential efficacy in long-term prophylaxis requires the observation of a consistent reduction of episode severity or the prevention of episodes for a duration several times longer than the previous usual “well interval.” Thus, the faster the cycling at baseline, the more rapid can be the assessment of prophylactic effectiveness.

Although 4 to 12 μg/mL has been identified as a therapeutic range for anticonvulsant effects, there is no relationship between blood levels to the degree of anticonvulsant, antimanic or antidepressant response. One study suggested that 1,200 mg per day was associated with an antimanic response, although trough serum levels of 8 to 12 μg/mL are generally associated with good anticonvulsant and antimanic effects (Fig. 33.10c–10).

However, there is very considerable variability in the relationship of blood levels to an individual’s threshold for and slope of the rate of rise of side effects. That is, some patients may begin to have considerable side effects at low blood levels of 4 ug/mL, while others may have no side effects at levels of 12 ug/mL. This indicates that in most instances, upward dose titration can be pursued as a function of clinical responsiveness and an absence of emergence of side effects rather than with close attention to blood level monitoring. If fact, most patients can be very well managed without blood level monitoring except in instances of unexpected toxicity, absence of effectiveness at high doses, or suspicion of noncompliance.

Schizophrenia and Schizoaffective Disorder

A series of double-blind studies with few exceptions have indicated that the addition of carbamazepine to antipsychotics is more effective than placebo, particularly in those patients with more florid positive symptoms and problems with aggression and dyscontrol. It does not appear useful as an adjunct in chronic, negative, symptom-prominent schizophrenia. Carbamazepine is helpful in schizoaffective disorder, particularly in depressed phases. One study suggested that, in contrast to carbamazepine, oxcarbazepine was less effective in those with more fulminant psychotic manic symptoms.

Impulse Control Disorders

A very substantial case series literature indicates that carbamazepine is highly effective in a wide range of aggressive syndromes and impulse-dyscontrol disorders that can manifest in those with diagnoses of borderline and other personality disorders, affective disorders, and schizophrenia. Its efficacy in these dyscontrol syndromes appears independent of the presence or absence of an associated EEG abnormality.

Other Disorders

Carbamazepine is as effective in acute alcohol withdrawal syndromes as diazepam, and it does not have the abuse liability of benzodiazepines. Greater anxiolytic and global distress decreases have also been reported in withdrawing alcoholics treated with carbamazepine compared with benzodiazepines. In addition, placebo-controlled prophylactic studies of alcohol dependence have suggested greater relapse prevention of heavy drinking with carbamazepine. Although carbamazepine inhibits the development of cocaine-kindled seizures and their associated lethality in rodents, controlled clinical studies in cocaine abusers did not indicate that it is more successful in maintaining abstinence or drug-free urine than placebo, except in a subgroup with a concomitant mood disorder.

Open studies suggest the potential use of carbamazepine in the treatment of PTSD, particularly the paroxysmal recurrent components and the associated sleep disorder. Isolated case reports, but no controlled clinical trials, indicate its adjunctive use in other anxiety disorders and in OCD.

PRECAUTIONS AND ADVERSE REACTIONS

The *Physicians' Desk Reference* indicates that aplastic anemia and agranulocytosis occur in conjunction with the use of carbamazepine at a rate five to eight times greater than that in the general population. That incidence is approximately 6 patients per 1,000,000 of the population per year for agranulocytosis and 2 patients per 1,000,000 population per year for aplastic anemia. If, in the course of treatment, a patient exhibits low or decreased WBC or platelet counts, that individual should be monitored closely in the face of stable, benign WBC suppression; the usual minimum safe total WBC count is approximately 3,000 and absolute neutrophil count approximately 1,000 to 1,500 neutrophils/mm^3. The addition of lithium may be sufficient to reverse the benign WBC suppression of carbamazepine based on their opposite effects on stimulating or inhibiting colony-stimulating factor. However, if there is a concomitant suppression of the red cell or platelet series or other indications of bone marrow depression, carbamazepine should be stopped immediately.

Carbamazepine will increase the metabolism of estrogen in birth control formulations, resulting in occasional unwanted pregnancies. Therefore, a higher dosage form of the estrogen component or other contraceptive devices should be used.

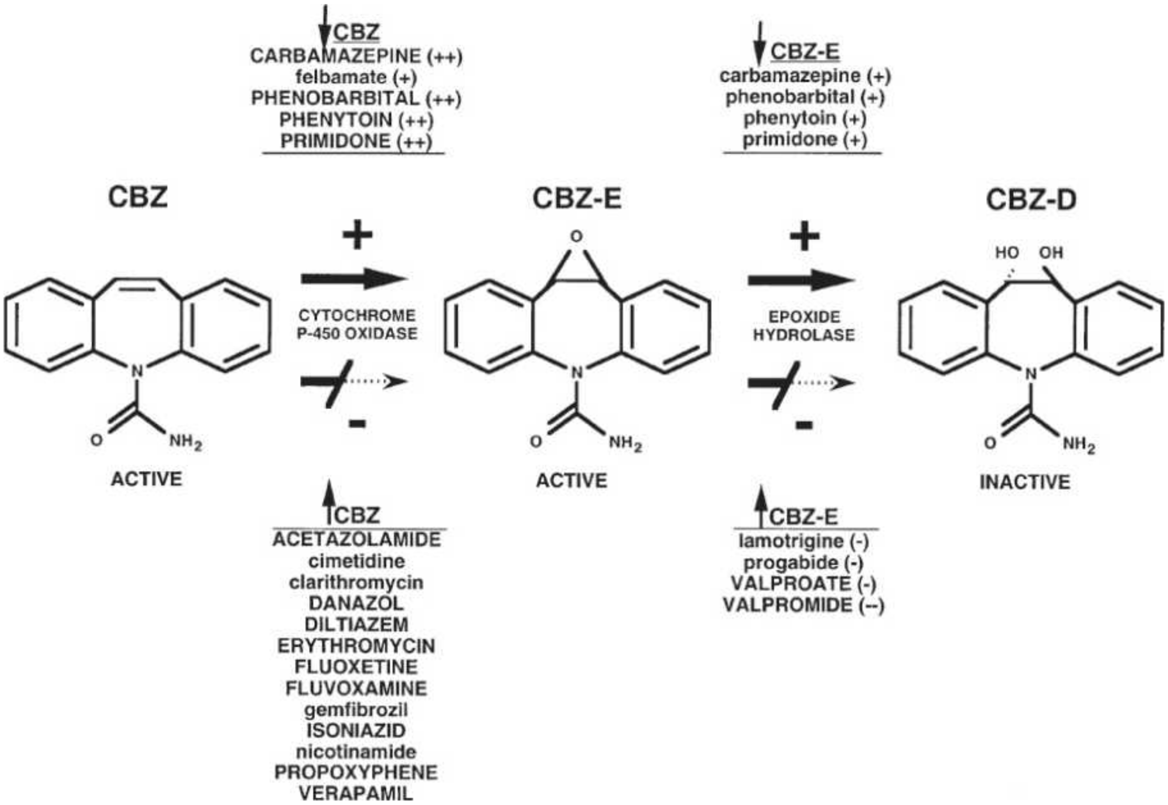
CBZ
CARBAMAZEPINE (++)
felbamate (+)
PHENOBARBITAL (++)
PHENYTOIN (++)
PRIMIDONE (++)
CBZ-E
carbamazepine (+)
phenobarbital (+)
phenytoin (+)
primidone (+)
CBZ
CBZ-E
CBZ-D
+
+
O
HO OH
CYTOCHROME P-450 OXIDASE
EPOXIDE HYDROLASE
-
-
N
O NH2
ACTIVE
ACTIVE
INACTIVE
CBZ
ACETAZOLAMIDE
cimetidine
clarithromycin
DANAZOL
DILTIAZEM
ERYTHROMYCIN
FLUOXETINE
FLUVOXAMINE
gemfibrozil
ISONIAZID
nicotinamide
PROPOXYPHENE
VERAPAMIL
CBZ-E
lamotrigine (-)
progabide (-)
VALPROATE (-)
VALPROMIDE (--)

FIGURE 33.10c–12. Carbamazepine (CBZ) metabolism. *Bottom left:* Drugs that inhibit cytochrome P450 increase CBZ. ++, drugs that decrease the level of the active metabolite carbamazepine through induction of its oxidative metabolic pathway cytochrome P450; +, drugs that decrease the level of the active metabolite carbamazepine-10, 11-epoxide (CBZ-E) through induction of the epoxide hydroxylase and production of the inactive dihydroxy metabolite (CBZ-D); −, drugs that increase the level of carbamazepine-10,11-epoxide through blockade of the epoxide catabolism. Drug names in capital letters indicate common or more important interactions.

Carbamazepine is a category D drug and should be avoided during pregnancy. There is a several percentage risk of spina bifida with the use of carbamazepine, and this appears to increase with anticonvulsant combination treatment. In one series, carbamazepine was reported to decrease the head size and body length, although the developmental lag was completely compensated for after a period of several months. Although carbamazepine and its epoxide metabolite are transferred to breast milk during lactation, their concentrations are approximately 50% of those in maternal plasma. There are no adequate studies of the clinical effects of oxcarbazepine in pregnant women; given its similar structure to carbamazepine and evidence of problems in animal studies, one might consider it to be teratogenic in humans until more systematic data to the contrary are acquired.

DRUG INTERACTIONS

In contrast to oxcarbazepine, carbamazepine is a potent inducer of cytochrome P450 oxidase enzymes, particularly of the 3A4 category. This leads to its own autoinduction after 2 to 3 weeks, as well as a reduction in the levels of a variety of other compounds listed in Table 33.10c–2.

If a patient is switched from carbamazepine to oxcarbazepine, deinduction of the 3A4 enzyme should be expected, and substances that had their levels reduced by carbamazepine may show increases. This is particularly important for lamotrigine, as carbamazepine approximately halves lamotrigine levels, and substantial increases in lamotrigine levels would occur with carbamazepine discontinuation. This could also occur with dicumarol and warfarin (Coumadin), increasing proclivity for excess anticoagulant activity and increased bleeding.

Of potential importance for the generation of carbamazepine's toxicity are a series of agents that inhibit either cytochrome oxidase 3A4 or the epoxide hydroxylase, as schematized in Figure 33.10c–12. Many commonly used agents markedly increase blood levels of carbamazepine (Table 33.10c–7). These include the antibiotic erythromycin and all of its related congeners, as well as the calcium channel blockers verapamil (Verelan) and diltiazem (Cardizem), but not nifedipine (Adalat), nimodipine, or amlodipine (Norvasc). The antituberculosis drug isoniazid (Nydrazid) will also markedly increase carbamazepine levels, but more minor effects are observed with dextropropoxyphene (Darvon), fluoxetine (Prozac), and nicotinamide (Nicomide), and minimal effects occur with cimetidine (Tagamet).

Table 33.10c–7.
Drugs That Increase Concentrations of Carbamazepine

Drug Class	Drug
Antidepressants	Fluoxetine, flavoxamine, nefazodone[a,b]
Antiepileptics	Valproate (increases CBZ-10, 11-epoxide[c])
Antimicrobials	Isoniazid[d], quinupristin/dalfopristin
Ca^+ channel blockers	Diltiazem[c], verapamil[c,d], (but not nifedipine or nimodipine)
Lipid active	Gemfibrozila, nicotinamide[a]
Macrolide antibiotics	Clarithromycin[c,d], erythromycin[c,d], flurimycin[c,d], josamycin[c,d], midecamycin acetate, ponsinomycin[c,d], troleandromycin, triacetyleandronyion[c,d]
Anti-HIV	Delavirdine[c,d], ritonavir[c,d]
Others	Acetazolamide[a], cimetidine[a], danazol, delavirdine[c,d], dextropropoxyphene[a], gestodene[b], mifepristone, omeprazole[a], ticlopidine[a]

[a] Weak effect.

[b] Mild to moderate.

[c] Potential for toxicity.

[d] Important or severe.

Table 33.10c–8.
CBZ in Bipolar Illness: A Brief Users Guide

Start with low (200 mg) H.S. doses in depression, euthymia; higher doses (600–800 mg/day in nighttime and divided doses) in manic inpatients.
All H.S. dosing is reasonable with CBZ ER (Equetro) preparation.
Titrate slowly to an individual's response or side effects threshold (as there is no correlation with blood levels and degree of response).
Hepatic enzyme CYP 450 (3-A-4) induction and autoinduction occurs in 2–3 weeks; slightly higher doses of CBZ may be needed/tolerated at that time.
Employ adequate birth control methods, including higher-dosage forms of estrogen (as CBZ will lower estrogen levels).
Major increases in CBZ blood levels and potential toxicity are caused by erythromycin, verapamil, isradipine.
Warn regarding benign rash 5–10%; rare progression to severe rash, so drug should be discontinued. HLA-A*3101 predicts severe rash in northern Europeans and Japanese. HLA-B-1502 predicts in those of Asian descent.
Benign WBC decreases occurs regularly (and are usually inconsequential). Lithium will prevent or reverse these.
Rare agranulocytosis and aplastic anemia (several per million new exposures); warn regarding appearance of fever, sore throat, petechiae, bleeding gums, and to check with physician to obtain immediate CBC.
Avoid CBZ in pregnancy: Spina bifida 0.5%; other adverse outcomes about 8%.
Some people will response well to CBZ and no other mood stabilizers (lithium) or anticonvulsants (VPA).
Combination treatment often required to maintain remission and prevent loss of effect via tolerance.

H.S., nighttime or hours of sleep.

There are fewer drug–drug interactions of importance for oxcarbazepine. However, strong inducers of cytochrome P450 enzymes (such as carbamazepine) decrease plasma levels of the monohydroxy derivative metabolite by 29% to 40%. As noted in Table 33.10c–1, there is no autoinduction, and few major drug–drug interactions have been reported. However, oxcarbazepine reduces estrogen levels in oral contraceptives, and its use, as with carbamazepine, will necessitate higher dosage forms or alternative contraceptive devices.

The use of carbamazepine fell off in the 1990s after a decade of considerable use in the 1980s. One reason was the advent of valproate and the notion that it was easier to use. However, now that it is clear that some patients are relatively selective responders to carbamazepine, it becomes increasingly important to demystify some of the important details about its use. A primer employing a dozen key principles is included in Table 33.10c–8.

FURTHER READINGS

Alvarez G, Marsh W, Camacho IA, Gracia SL. Effectiveness and tolerability of carbamazepine vs. oxcarbazepine as mood stabilizers. *Clin Res Reg Affairs*. 2003;20:365.

Benedetti A, Lattanzi L, Pini S, Musetti L, Dell'Osso L, Cassano GB. Oxcarbazepine as add-on treatment in patients with bipolar manic, mixed, or depressive episode. *J Affect Disord*. 2004;79:273–277.

*Bertilsson L, Tomson T. Clinical pharmacokinetics and pharmacological effects of carbamazepine and carbamazepine-10,11-epoxide: an update. *Clin Pharmacokinet*. 1986;11:177–198.

*Denicoff KD, Smith-Jackson EE, Disney ER, Ali SO, Leverich GS, Post RM. Comparative prophylactic efficacy of lithium, carbamazepine, and the combination in bipolar disorder. *J Clin Psychiatry*. 1997;58(11):470–478.

Emrich HM. Studies with (Trileptal) oxcarbazepine in acute mania. *Int Clin Psychopharmacol*. 1990;5[Suppl 1]:83.

Fattore C, Cipolla G, Gatti G, et al. Induction of ethinylestradiol and levonorgestrel metabolism by oxcarbazepine in healthy women. *Epilepsia*. 1999;40(6):783–787.

Ghaemi NS, Berv DA, Klugman J, Rosenquist KJ, Hsu DJ. Oxcarbazepine treatment of bipolar disorder. *J Clin Psychiatry*. 2003;64:943–945.

Glauser TA, Nigro M, Sachdeo R, et al. Adjunctive therapy with oxcarbazepine in children with partial seizures. The Oxcarbazepine Pediatric Study Group. *Neurology*. 2000;54(12):2237–2244.

*Greil W, Kleindienst N, Erazo N. Muller-Oerlinghausen B. Differential response to lithium and carbamazepine in the prophylaxis of bipolar disorder. *J Clin Psychopharmacol*. 1998;18(6):455–460.

Hartong EG, Moleman P, Hoogduin CA, Broekman TG, Nolen WA, LitCar Group. Prophylactic efficacy of lithium versus carbamazepine in treatment-naive bipolar patients. *J Clin Psychiatry*. 2003;64:144–151.

Hebeisen S, Pires N, Loureiro AI, et al. Eslicarbazepine and the enhancement of slow inactivation of voltage-gated sodium channels: a comparison with carbamazepine, oxcarbazepine and lacosamide. *Neuropharmacology*. 2015;89:122–135.

Isojarvi JI, Turkka J, Pakarinen AJ, Kotila M, Rattya J, Myllyla VV. Thyroid function in men taking carbamazepine, oxcarbazepine, or valproate for epilepsy. *Epilepsia*. 2001;42(7):930–934.

Joca SR, Skalisz LL, Beijamini V, Vital MA, Andreatini R. The antidepressive-like effect of oxcarbazepine: Possible role of dopaminergic neurotransmission. *Eur Neuropsychopharmacol*. 2000;10(4):223–238.

Ketter TA, Kimbrell TA, George MS, Willis MW, Benson BE. Baseline cerebral hypermetabolism associated with carbamazepine response, and hypometabolism with nimodipine response in mood disorders. *Biol Psychiatry*. 1999;46(15):1364–1374.

Pazzaglia PJ, Post RM, Ketter TA, et al. Nimodipine monotherapy and carbamazepine augmentation in patients with refractory recurrent affective illness. *J Clin Psychopharmacol*. 1998;18(5):404–413.

Post RM, Altshuler L, Leverich GS, et al. Clinical correlates of sustained response to individual drugs used in naturalistic treatment of patients with bipolar disorder. *Compr Psychiatry*. 2016;66:146–156.

Post RM, Berrettini W, Uhde TW, Kellner C. Selective response to the anticonvulsant carbamazepine in manic-depressive illness: a case study. *J Clin Psychopharmacol*. 1984;4(4):178–185.

Post RM, Leverich GS. *Treatment of Bipolar Illness: A Casebook for Clinicians and Patients, Chapters 14–20*. W.W.Norton & Company, Inc; 2008:133–190.

Post RM, Uhde TW, Roy-Byrne PP, Joffe RT. Antidepressant effects of carbamazepine. *Am J Psychiatry*. 1986;143:29–34.

Post RM, Weiss SR. Tolerance to the Prophylactic Effects of Carbamazepine and Related Mood Stabilizers in the Treatment of Bipolar Disorders. *CNS Neurosci Ther*. 2011;17(6):649–660. .

Post RM, Yatham LN, Vieta E, Berk M, Nierenberg A. Beyond evidence-based treatment of bipolar disorder: Rational pragmatic approaches to management: *Bipolar Disorders*. 2019;21:650–659.

Post RM, Goldstein B, Birmaher B, et al. Toward prevention of bipolar disorder in at-risk children: Potential strategies ahead of the data. *J. Affective Disorders*. 2020;272:508–520.

Raja M, Azzoni A. Oxcarbazepine vs. valproate in the treatment of mood and schizoaffective disorders. *Int J Neuropsychopharmacol*. 2003;6:409–414.

Rattya J, Turkka J, Pakarinen AJ, et al. Reproductive effects of valproate, carbamazepine, and oxcarbazepine in men with epilepsy. *Neurology*. 2001;56(1):31–36.

Rosa FW. Spina bifida in infants of women treated with carbamazepine during pregnancy. *N Engl J Med*. 1991;324:674–677.

Shimoyama R, Ohkubo T, Sugawara K. Monitoring of carbamazepine and carbamazepine 10,11-epoxide in breast milk and plasma by high-performance liquid chromatography. *Ann Clin Biochem*. 2000;37(Pt 2):210–215.

Wagner KD, Kowatch RA, Emslie GJ, et al. A double-blind, randomized, placebo-controlled trial of oxcarbazepine in the treatment of bipolar disorder in children and adolescents. *Am J Psychiatry*. 2006;163(7):1179–1186.

Wang PW, Ketter TA, Post RM. Carbamazepine and oxcarbazepine. In: Schatzberg AF, Nemeroff CB, eds: *Textbook of Psychopharmacology*. American Psychiatric Press;

2004:581.

Weiss SR, Clark M, Rosen JB, Smith MA, Post RM. Contingent tolerance to the anticonvulsant effects of carbamazepine: relationship to loss of endogenous adaptive mechanisms. *Brain Res Brain Res Rev*. 1995;20:305–325.

*Weisler RH, Kalai AK, Ketter TA. SPD417 Study Group. A multicenter, randomized, placebo-controlled trial of extended-release carbamazepine capsules as monotherapy for bipolar disorder patients with manic or mixed episodes. *J Clin Psychiatry*. 2004;65(4):478–484.

Zhang ZJ, Kang WH, Tan QR, et al. Adjunctive herbal medicine with carbamazepine for bipolar disorders: a double-blind, randomized, placebo-controlled study. *J Psychiatr Res*. 2007;41(3–4):360–369.

Zhang ZJ, Postma T, Obeng K, Russell S, Weiss SR, Post RM. The benzodiazepine partial inverse agonist Ro-15-4513 alters anticonvulsant and lethal effects of carbamazepine in amygdala-kindled rats. *Neurosci Lett*. 2002;329:253–256.

Zhang ZJ, Xing GQ, Russell S, Obeng K, Post RM. Unidirectional cross-tolerance from levetiracetam to carbamazepine in amygdala-kindled seizures. *Epilepsia*. 2003;44:1487–1493.

33.10d Lamotrigine

ROBERT M. POST, M.D.

Neuroscience-Based Nomenclature (NbN)	
Glutamate channel blocker	
Pharmacologic Domain	Glutamate
Mechanism of Action	Voltage-gated sodium channel blocker

HISTORY

Lamotrigine (Lamictal) was developed as a result of screening folate antagonists as anticonvulsants. The observations that folates could produce seizures in animals and some anticonvulsants such as phenytoin (Dilantin) and phenobarbital (Solfoton) could impair folate metabolism led to the hypothesis that folate antagonists could be anticonvulsants. Further research has failed to establish antifolate as a treatment for epilepsy in humans, and lamotrigine also appears to lack clinically significant antifolate effects in humans. Nevertheless, lamotrigine did prove effective in several animal models of epilepsy. Burroughs Wellcome (later Glaxo Wellcome and currently GlaxoSmithKline) developed lamotrigine as an antiepileptic drug, and it was marketed for the adjunctive treatment of partial seizures in the United States in 1995.

Initial, postmarketing, open, clinical experience suggested efficacy, coupled with generally good tolerability (aside from the risk of rash), in a variety of neurologic and psychiatric conditions. Later double-blind, placebo-controlled studies revealed that lamotrigine was useful for some of the neurologic and psychiatric conditions suggested in open studies. Lamotrigine demonstrated efficacy and was approved for maintenance treatment of bipolar I disorder in 2003, with effects more robust in preventing depression than mania. Lamotrigine also had some efficacy in acute bipolar depression, but the magnitude of the effect was only modest. Limited data also suggested lamotrigine had potential utility in rapid cycling bipolar disorder. However, lamotrigine did *not* receive approval for the treatment of acute bipolar depression or rapid cycling bipolar disorder. Lamotrigine was *not* effective as the main intervention in acute mania. In summary, lamotrigine emerged as an agent that appears to “stabilize mood from below,” such that it may maximally impact the depressive component of bipolar disorders and have some mood-stabilizing effects as well.

CHEMISTRY

Lamotrigine, 3,5-diamino-6-(2,3-dichlorophenyl)-*as*-triazine, is a structurally novel phenyltriazine anticonvulsant (Fig. 33.10d–1). Three oral lamotrigine preparations are currently available in the United States: 25-, 100-, 150-, and 200-mg immediate- or extended-release tablets; 50-, 250-, and 300-mg extended-release tablets; and 2-, 5-, 25-, 50-, 100-, and 200-mg chewable dispersible tablets. By mid-2016, generic formulations of all of the above (with the exception of the 2- and 5-mg chewable dispersible tablets) were available.

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Lamotrigine has an oral bioavailability of about 98% and linear kinetics that are not significantly altered by food. Lamotrigine is 55% bound to plasma proteins and has a moderate volume of distribution of about 1 L/kg. A large percentage (approximately 85%) of lamotrigine is conjugated to yield inactive glucuronide metabolites, with only about 10% excreted unchanged in the urine. In monotherapy, its half-life is around 28 hours and clearance is about 40 mL per minute.

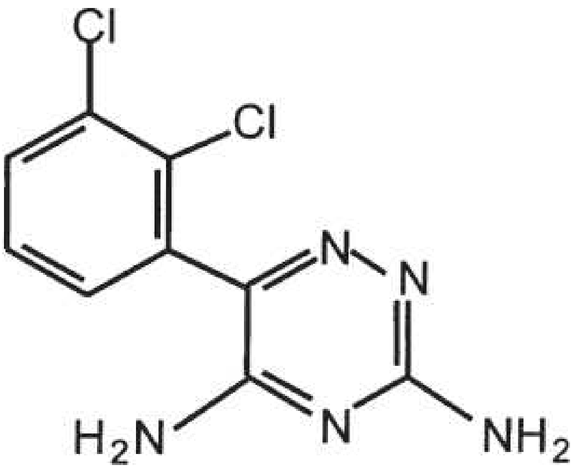

FIGURE 33.10d–1. Molecular structure of lamotrigine.

Combined with hepatic enzyme inducers such as carbamazepine (Tegretol, Equetro) or phenytoin (Dilantin), lamotrigine's half-life falls about 50% to approximately 14 hours and clearance doubles to approximately 80 mL per minute. Estrogen-containing hormonal contraceptives also induce lamotrigine metabolism.

Combined with the enzyme inhibitor valproate (Depakote), lamotrigine's half-life doubles to approximately 56 hours and clearance falls approximately 50% to around 20 mL per minute. Thus, in patients with epilepsy, immediate-release lamotrigine is given in two divided doses unless valproate is present, which may allow single daily dosing. In bipolar disorder patients where stable blood levels are less critical, single daily dosing of immediate-release lamotrigine is used unless an enzyme inducer (such as carbamazepine [Tegretol]) is present in which case two divided doses are recommended. An extended-release formulation of lamotrigine has been approved for once-daily administration in patients with epilepsy but not bipolar disorder. Lamotrigine clearance is decreased with hepatic insufficiency, which may necessitate decreased dosage. Clearance is decreased to a lesser extent with renal insufficiency, but dosage decreases may still be needed.

Pharmacodynamics

Like other anticonvulsants, lamotrigine has multiple biochemical actions, and the relationships between these and clinical effects remain to be established. Lamotrigine blocks voltage-sensitive sodium channels, modulating the release of the excitatory amino acids glutamate and aspartate, and may have more limited effects on calcium channels.

Lamotrigine also weakly blocks 5-HT_3 serotonergic receptors, has weak effects at σ-opioid receptors, and modestly inhibits serotonin uptake in human platelets and rat brain synaptosomes and noradrenaline and dopamine uptake in rat brain synaptosomes. In contrast, lamotrigine does not bind with high affinity to GABAergic, NMDA, glutamatergic, adrenergic, 5-HT_2 serotonergic, dopaminergic, cholinergic, histaminergic, adenosinergic, or κ-opiate–binding sites, and does not affect dihydropyridine calcium channels. Although lamotrigine inhibits dihydrofolate reductase in vitro, it does not appear to have clinically significant antifolate effects in humans.

Lamotrigine has antiseizure activity in maximal electroshock, pentylenetetrazol, amygdala-kindled, and other animal models of seizures; cerebroprotective actions in several animal models of ischemia; analgesic effects in animal models of neuropathic pain; and activity in the conflict animal model of anxiety. It is noteworthy that lamotrigine, like carbamazepine, does not block the initial stages of the development of amygdala-kindled seizures, while it is highly effective in inhibiting the middle phase of fully developed kindled seizures. Like carbamazepine, repeated administration of lamotrigine against fully kindled seizure results in tolerance development. This contingent tolerance phenomenon depends on the drug being present at the time of amygdala stimulation and can be reversed by a period of time when seizures are induced in the absence of the drug. There are some clinical observations that a similar tolerance process could develop in the long-term prevention of affective episodes.

Lamotrigine blood concentrations are not systematically related to antiseizure effects or efficacy in bipolar disorders.

CLINICAL DRUG STUDIES

Seizure Disorders

Immediate-release lamotrigine monotherapy appeared effective in a multicenter, double-blind controlled study of converting adults with partial seizures taking carbamazepine, phenytoin, phenobarbital, or primidone monotherapy to lamotrigine monotherapy. In a single study with a less rigorous methodology, extended-release lamotrigine monotherapy also appeared effective for conversion to monotherapy in adolescents and adults with partial seizures. Adjunctive immediate-release lamotrigine appeared more effective than adjunctive placebo in three multicenter, double-blind, placebo-controlled trials in adults with refractory partial seizures; in one multicenter, double-blind, placebo-controlled trial in pediatric (age 2 to 16 years) partial seizure disorder patients; another in adult and pediatric (age 3 to 25 years) patients with Lennox–Gastaut syndrome; and in pediatric and adult patients with primary generalized tonic–clonic seizures. Extended-release adjunctive lamotrigine appeared effective for primary generalized tonic–clonic seizures and partial-onset seizures in adolescents (age ≥13 years) and adults.

Pain Syndromes

Lamotrigine may have analgesic effects, as it was effective in placebo-controlled studies of post-transurethral prostatectomy pain, central poststroke pain, and HIV painful distal sensory polyneuropathy. However, in placebo-controlled studies, lamotrigine had inconsistent efficacy in diabetic neuropathy and was ineffective in neuropathic pain and spinal cord injury pain.

Other Neurologic Disorders

Although open reports were encouraging, a double-blind, placebo-controlled study failed to show benefits for migraine. Lamotrigine was also ineffective in placebo-controlled studies in amyotrophic lateral sclerosis, Huntington disease, and Parkinson disease.

Bipolar Disorders

In two 18-month multicenter, randomized, double-blind, placebo-controlled trials (one after depressive episodes, one after manic, or hypomanic episodes), both lamotrigine and lithium, compared with placebo, delayed time to intervention for any mood episode in bipolar I patients. Lamotrigine mood prophylaxis effect was driven by a more robust delay of depression, so lamotrigine was predominantly effective against depressive relapse. Lithium mood prophylaxis was driven by a more robust delay of mania, so lithium was predominantly effective against mania.

Although an early double-blind, placebo-controlled study suggested that lamotrigine monotherapy had some efficacy in the acute treatment of bipolar depression (based on a secondary outcome measure), four subsequent trials were negative. However, a meta-analysis including all five controlled trials suggested a statistically significant but only clinically modest efficacy in acute bipolar depression. One double-blind, placebo-controlled, industry-sponsored study suggested that the olanzapine (Zyprexa) plus fluoxetine (Prozac) combination compared to lamotrigine monotherapy had modestly superior efficacy for acute bipolar depression but also more adverse effects. A small controlled trial suggested that adjunctive lamotrigine compared to adjunctive risperidone (Risperdal) might have utility in bipolar disorder patients who remained depressed despite taking a mood stabilizer plus an antidepressant.

In treatment-resistant (primarily rapid cycling bipolar) mood disorder inpatients, about 50% responded to lamotrigine monotherapy, compared to only about 25% with gabapentin (Neurontin) monotherapy or placebo. Moreover, in patients who were responsive to lamotrigine in the first randomized phase, the response was again confirmed in a blind re-exposure to lamotrigine, such that antidepressant effects were affirmed by the initial increased percent response and by a renewed response to a second exposure.

In a less treatment-resistant community sample of patients with bipolar disorder and a rapid cycling course of illness, lamotrigine monotherapy appeared more effective than placebo bipolar II disorder, but not bipolar I disorder. Added to lithium, adjunctive lamotrigine was superior to placebo in acute bipolar depression and also appeared to provide continued benefit in longer-term treatment. Adding lamotrigine to quetiapine also yielded acute and longer-term benefits in patients with quetiapine-resistant bipolar depression. However, in that study, a subgroup that was treated with adjunctive folate surprisingly did not show a significant effect compared to those not treated with folate. Lamotrigine monotherapy has not been shown efficacious in controlled studies of acute mania. In summary, these controlled studies support the notion that lamotrigine may “stabilize mood from below” in the sense that it may impact the depressive more than the manic component of bipolar disorders. Although lamotrigine is not FDA-approved for acute bipolar depression, the meta-analysis of the four RCTs, the superiority of gabapentin and placebo, and its efficacy as an adjunct to both lithium and quetiapine provide substantial evidence of its efficacy for acute as well as a preventive treatment of bipolar depression.

Family history may offer hints to the effectiveness of the mood stabilizers. While lithium is more effective in those with a positive family history of mood disorder, carbamazepine appears more effective in those without a positive family history of bipolar disorder. In contrast, lamotrigine appears to have good responses in those with a family and personal history of anxiety disorders. As comorbid anxiety disorders are a consistent predictor of a poor response to lithium, lamotrigine, and other mood-stabilizing anticonvulsants valproate and carbamazepine may offer better effects in this large subgroup of patients with bipolar disorders.

Major Depressive and Other Disorders

There are fewer studies with mixed results suggesting lamotrigine efficacy in (unipolar) major depressive disorders. In controlled studies, chronic lamotrigine was no better than placebo in children with autism or adults with depersonalization disorder and was only numerically but not statistically superior to placebo in adults with PTSD. In hospitalized treatment-resistant schizophrenia patients on clozapine (Clozaril), adding lamotrigine compared to placebo reduced general psychopathologic and positive (but not negative) symptoms. A subsequent similar trial of adding lamotrigine to antipsychotics failed to replicate this in the primary (last observation carried forward) analysis but was positive in a post hoc observed case analysis. Interestingly, in a small controlled trial, lamotrigine appeared superior to placebo in the treatment of anger in women with borderline personality disorder. Lamotrigine was also superior to adding a placebo to SSRI treatment for obsessive-compulsive disorder. A small controlled study in patients with obesity suggested that lamotrigine compared to a placebo might yield modestly greater decreases in BMI, but not weight.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Lamotrigine has been associated with common benign rashes, uncommon serious rashes, and rare hypersensitivity reactions (see Precautions below). Lamotrigine binds to melanin-containing tissues and thus has the potential to accumulate in the eyes. However, to date, lamotrigine has not been associated with ocular problems.

THERAPEUTIC INDICATIONS

Seizure Disorders

Immediate-release lamotrigine is indicated as an adjunctive treatment for adults and children over age 2 years with partial seizures and for pediatric (over age 2 years) and adult patients with Lennox–Gastaut syndrome. It is also approved for conversion to monotherapy in adults with partial seizures who are receiving valproate or a single enzyme-inducing antiepileptic drug such as carbamazepine, phenytoin, phenobarbital, or primidone in patients ≥16 years old. Extended-release lamotrigine is indicated in adolescents and adults as adjunctive therapy for primary generalized tonic–clonic seizures and partial-onset seizures, as well as for conversion to monotherapy in patients with partial seizures who are taking a single anticonvulsant.

Bipolar Disorders

Lamotrigine is indicated for the maintenance treatment of bipolar I disorder to delay the time to occurrence of mood episodes in patients treated for acute mood episodes with standard therapy. Limited controlled data suggested lamotrigine might have at least some efficacy in the acute treatment of bipolar depression and in rapid cycling bipolar disorder. In view of such data, lamotrigine was considered to be an important treatment option for bipolar maintenance treatment and acute bipolar depression in the 2013 revision of the Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders' update of the CANMAT guidelines for the management of patients with bipolar disorder. However, as of late 2021, lamotrigine continued to lack FDA approval for the treatment of acute bipolar depression or for bipolar II prophylaxis, where its spectrum of greater efficacy in depression and mania prevention would suggest utility for this condition.

PRECAUTIONS AND ADVERSE REACTIONS

The prescribing information for lamotrigine includes a boxed warning due to the risk of serious rashes (including Stevens–Johnson syndrome) requiring hospitalization and discontinuation of treatment. The incidence of serious rashes was estimated in 1997 to be 1% in prospectively monitored pediatric (age less than 16 years) and 0.3% in premarketing clinical trials of adult patients. Rare cases of toxic epidermal necrolysis and rash-related death have been reported, but numbers are too few to allow precise estimates of rates.

Other than age, possible factors increasing the risk of serious rash include increasing lamotrigine doses too quickly, concurrent administration of valproate (which doubles lamotrigine blood concentrations), and exceeding the recommended initial lamotrigine dose or the rate of lamotrigine dose escalation. Nearly all life-threatening rashes have occurred between 2 and 8 weeks after lamotrigine initiation, but isolated cases have been reported after prolonged treatment. If patients do not take lamotrigine for 5 days or more (approximately five times the half-life of lamotrigine), the drug needs to be reintroduced according to the gradual initial escalation regimen, as serious rashes can occur with rapidly restarting the medication.

Benign rashes can occur in as many as 10% of patients taking lamotrigine, and the prescribing information recommends that lamotrigine be ordinarily discontinued at the first sign of rash unless the rash is clearly not drug related. Thus, any rash in a patient on lamotrigine requires immediate evaluation. Some clinicians report that “dermatology precautions” and initial titration even slower than that described in the prescribing information may decrease the rate of benign rash.

Dermatology precautions include not starting lamotrigine within 2 weeks of having a rash, viral syndrome, or vaccination. It is advised during the first 3 months on lamotrigine to avoid ingesting other new medicines or new foods or utilizing new cosmetics, conditioners, deodorants, detergents, or fabric softeners or developing sunburn or poison ivy/oak exposure. However, the utility of these dermatology precautions was not confirmed in a controlled study. The utility of initial titration, even slower than that described in the prescribing information, is widely practiced but remains to be established in a controlled study.

Any patient developing a rash accompanied by eye, mouth, or bladder discomfort, fever, or lymphadenopathy should go immediately to the emergency room for dermatologic assessment as such symptoms could herald the onset of a serious rash. Rashes with more benign presentations require medical evaluation as soon as possible (and before taking the next dose of lamotrigine) for assessment of the risk of continuing lamotrigine treatment. Rare hypersensitivity reactions with multiorgan failure/dysfunction have been reported so that patients who develop fever or lymphadenopathy require emergent evaluation. Rare blood dyscrasias have been reported. In view of the potential risk of withdrawal seizures in patients with epilepsy, lamotrigine should be tapered over a period of at least 2 weeks unless safety concerns require more rapid lamotrigine discontinuation. Stopping lamotrigine immediately for a rash in patients with affective disorders would appear prudent.

With the exception of the risk of rash, lamotrigine is generally well tolerated. Although transient difficulties with headache may be encountered during the initiation phase, other difficulties such as sedation, neurocognitive adverse effects, sexual problems, or weight gain are not typically seen.

The most common lamotrigine adverse effects during dose escalation (added to standard therapy) phase in adult bipolar disorder patients were headache, rash, dizziness, diarrhea, dream abnormality, and pruritus. In the lamotrigine monotherapy phase in maintenance treatment studies in adult bipolar disorder patients, the most common adverse effects were nausea, insomnia, somnolence, back pain, fatigue, rash, and rhinitis, and the adverse effects most commonly leading to discontinuation were rash and manic/hypomanic/mixed mood symptoms.

In late 2008, the FDA announced a class warning regarding increased risk of suicidality (suicidal behavior or ideation) in patients with epilepsy as well as psychiatric disorders for lamotrigine as well as 10 other anticonvulsants (carbamazepine, felbamate [Felbatol], gabapentin [Neurontin], levetiracetam [Keppra], oxcarbazepine [Trileptal], pregabalin [Lyrica], tiagabine [Gabitril], topiramate [Topamax], valproate, and zonisamide [Zonegran]). In the FDA’s analysis, anticonvulsants compared to placebo yielded approximately twice the risk of suicidality (0.43% vs. 0.22%). The relative risk for suicidality was higher in the patients with epilepsy compared to psychiatric disorders, suggesting that the risk in psychiatric patients was quite small. Acute oral overdoses of lamotrigine up to 15 g have caused central nervous system and cardiac adverse effects, and some have been fatal.

Lamotrigine was classified as pregnancy category C. Limited initial data suggested the possibility of a modest association between lamotrigine and cleft lip or palate, but this has not been replicated in later studies. Specifically, the North American Antiepileptic Drug Registry found a higher risk of oral cleft, which was not found in other pregnancy registries. Lamotrigine is excreted in breast milk, considered “unsafe” for lactation, and thus is contraindicated with breastfeeding.

DRUG INTERACTIONS

Lamotrigine is mainly metabolized by glucuronidation with minimal involvement of oxidative metabolism and is primarily a target rather than an instigator of drug–drug interactions. Thus, enzyme inducers such as carbamazepine, phenytoin, phenobarbital, primidone (Mysoline), rifampin (Rifadin), and estrogen-containing (but not progesterone) hormonal contraceptives can decrease blood lamotrigine concentrations by 40% to 50%. The carbamazepine derivatives, oxcarbazepine (Trileptal), and eslicarbazepine acetate (Aptiom) appear to less robustly decrease lamotrigine levels (so that dosage adjustment of lamotrigine administered with these agents is not recommended). The enzyme inhibitor valproate approximately doubles blood lamotrigine concentrations. Also, sertraline (Zoloft) may increase lamotrigine levels.

Lamotrigine appears to have at most modest effects on the pharmacokinetics of other drugs. For example, lamotrigine can yield modest (25%) decreases in valproate levels. Although lamotrigine can exacerbate the neurotoxic effects of carbamazepine, this may be due to a pharmacodynamic interaction, as only equivocal evidence suggests that lamotrigine could increase levels of the epoxide metabolite of carbamazepine. Lamotrigine does not alter lithium pharmacokinetics, which is important, given lamotrigine’s utility in bipolar disorders. Lamotrigine does not appear to have clinically significant drug interactions with clozapine, olanzapine, risperidone, or aripiprazole, although a therapeutic drug monitoring database study suggested that lamotrigine might decrease quetiapine concentrations. Lamotrigine does not appear to have clinically significant antifolate effects in humans, although folate may interfere with the benefit of adding lamotrigine to quetiapine.

LABORATORY INTERFERENCES

Lamotrigine has not been associated with interference with laboratory tests.

DOSAGE AND ADMINISTRATION

It is crucial to gradually introduce lamotrigine in order to attenuate the risk of rash.

Because valproate increases and enzyme inducers (such as carbamazepine, phenytoin, phenobarbital, primidone, and rifampin) decrease lamotrigine blood concentration, the lamotrigine prescribing information includes detailed instructions for gradual lamotrigine introduction (Table 33.10d–1).

Table 33.10d–1.
Gradual Introduction of Lamotrigine in Adults With Bipolar Disorders

	Lamotrigine With Valproate (mg/day)	Lamotrigine With Enzyme Inducers (mg/day)	Lamotrigine With Neither Valproate Nor Enzyme Inducers (mg/day)
Weeks 1 and 2 dose	12.5	50	25
Weeks 3 and 4 dose	25	100	50
Week 5 dose	50	200	100
Week 6 dose	100	300	200
Week 7 and thereafter (FDA target dose)	100	400	200
Typical final dose range	100–200	400–800	200–400

Adapted from FDA, U.S. Food and Drug Administration.

In maintenance treatment of adults with bipolar I disorder, when given without valproate or enzyme-inducing anticonvulsants, lamotrigine is started at 25 mg per day for weeks 1 and 2, increased to 50 mg per day for weeks 3 and 4, increased to 100 mg per day for week 5, and then increased to the 200 mg per day target dose. It may be wise, as some clinicians do, to use even more gradual initial titration in efforts to further decrease the risk of rash, starting with 25 mg per day for the first 2 weeks, increasing to 50 mg per day for the next 2 weeks, and thereafter increasing by 25 mg per day on a weekly basis.

When added to valproate, lamotrigine is started at lower doses and titrated upward more slowly. It is started at 25 mg every *other* day (or 12.5 mg per day) for weeks 1 and 2, increased to 25 mg per day for weeks 3 and 4, increased to 50 mg per day for week 5, and then increased to 100 mg per day target dose. Many clinicians use even more gradual initial titration in efforts to further decrease the risk of rash, starting with 12.5 mg per day for the first 2 weeks, increasing to 25 mg per day for the next 2 weeks, and thereafter increasing by 12.5 mg per day on a weekly basis.

When added to enzyme-inducing anticonvulsants (such as carbamazepine), lamotrigine is started at 50 mg per day for weeks 1 and 2, increased to 100 mg per day in divided doses for weeks 3 and 4, increased to 200 mg per day in divided doses for week 5, increased to 300 mg per day in divided doses for week 6, and then increased to 400 mg per day in divided doses to the target dose. Again, clinicians who wish to be more conservative use even more gradual initial titration in efforts to further decrease the risk of rash, starting with 50 mg per day for the first 2 weeks, increasing to 100 mg per day for the next 2 weeks, and thereafter increasing by 50 mg per day on a weekly basis.

Based on controlled trials, the prescribing information describes target doses of lamotrigine as 200 mg/day in monotherapy, 100 mg per day (with valproate), and 400 mg per day (with enzyme inducers). However, many clinicians choose to gradually titrate lamotrigine to a good clinical response rather than continuing to a presumed target dose. In the absence of a good response once target doses are achieved, some clinicians will further increase doses as necessary and tolerated to about twice the target figures. Although a therapeutic plasma concentration range has not been established for lamotrigine for bipolar disorder or for epilepsy, if patients attain high doses (e.g., 400 mg per day without inducers), it may be helpful to assess a 12-hour trough lamotrigine level to determine if there is a pharmacokinetic cause to apparent nonresponse.

While lamotrigine use in bipolar disorders had generally been limited in patients under the age of 16 years, due to the lack of an FDA indication and the potential for increased risk of rash, this has begun to change with the appreciation of the use of lamotrigine instead of antidepressants in adolescents with or at high risk for bipolar depression. Treatment with lamotrigine of young children with bipolar not otherwise specified (BP-NOS) and related comorbidities is increasingly also being considered given early evidence of effectiveness and the fact that the drug is used in pediatric epilepsy patients aged 2 to 12. Starting lower doses and slower rates of titration than even used in adult patients on valproate is advisable.

The prescribing information describes weight-proportional gradual adjunctive lamotrigine introduction in pediatric (age 2 to 12 years) epilepsy patients, with maintenance doses of 1 to 5 mg/kg per day (maximum 200 mg per day in one or two divided doses) in patients taking valproate, and 5 to 15 mg/kg per day (maximum 400 mg per day in two divided doses) in patients on enzyme-inducing anticonvulsants.

Although estrogen-containing hormonal contraceptives can decrease blood lamotrigine concentrations by as much as 50%, no adjustment to the lamotrigine dose-escalation guidelines is recommended solely based on the use of estrogen-containing hormonal contraceptives. However, the maintenance dose of lamotrigine may need to be increased as much as twofold.

When estrogen-containing hormonal contraceptives are added to lamotrigine, lamotrigine doses may be increased by 50 to 100 mg per day on a weekly basis. If adverse effects related to increased lamotrigine concentrations such as dizziness, ataxia, and diplopia consistently occur during the week of inactive hormonal preparation (i.e., the “pill-free” week), it may be necessary to decrease the overall lamotrigine maintenance dosage, but dose adjustments confined to the “pill-free” week are not recommended.

When estrogen-containing hormonal contraceptives are discontinued, lamotrigine doses may be decreased by 25% in the first week and by another 25% in the second week.

In patients with hepatic impairment, lamotrigine doses may need to be decreased by 25% in patients with moderate to severe hepatic impairment without ascites and 50% in patients with severe hepatic impairment with ascites. Lamotrigine dosage may need to be decreased in patients with impaired renal function.

If patients do not take lamotrigine for 5 days or more (i.e., for five times the half-life of lamotrigine or more), the drug needs to be reintroduced according to the gradual initial escalation regimen as serious rashes can occur with rapidly restarting the medication.

In conditions other than seizure disorders, lamotrigine is commonly given in one or two divided doses. Single daily dosing is convenient, enhances adherence, and in contrast to epilepsy, breakthrough symptoms due to the resulting lower trough blood concentrations are less of a concern. In some (nonepilepsy) applications, clinicians may administer most or all of the dosage in the morning or at bedtime if individual patients find the drug activating or sedating, respectively.

FURTHER READINGS

Agarwal N, Joshi M. Effectiveness of amitriptyline and lamotrigine in traumatic spinal cord injury-induced neuropathic pain: a randomized longitudinal comparative study. *Spinal Cord*. 2017;55(2):126–130.

Barbee JG, Thompson TR, Jamhour NJ, et al. A double-blind placebo-controlled trial of lamotrigine as an antidepressant augmentation agent in treatment-refractory unipolar depression. *J Clin Psychiatry*. 2011;72(10):1405–1412.

Biederman J, Joshi G, Mick E, et al. A prospective open-label trial of lamotrigine monotherapy in children and adolescents with bipolar disorder. *CNS Neurosci Ther*. 2010;16(2):91–102.

*Bowden CL, Calabrese JR, Sachs G, et al; Lamictal 606 Study Group. A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently manic or hypomanic patients with bipolar I disorder. *Arch Gen Psychiatry*. 2003;60(4):392–400.

Brown E, Dunner DL, McElroy SL, et al. Olanzapine/fluoxetine combination vs. lamotrigine in the 6-month treatment of bipolar I depression. *Int J Neuropsychopharmacol*. 2009;12(6):773–782.

*Calabrese JR, Bowden CL, Sachs G, et al. Lamictal 605 Study Group. A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently depressed patients with bipolar I disorder. *J Clin Psychiatry*. 2003;64(9):1013–1024.

Calabrese JR, Huffman RF, White RL, et al. Lamotrigine in the acute treatment of bipolar depression: results of five double-blind, placebo-controlled clinical trials. *Bipolar Disord*. 2008;10:323–333.

*Calabrese JR, Suppes T, Bowden CL, et al. A double-blind, placebo-controlled, prophylaxis study of lamotrigine in rapid-cycling bipolar disorder. *J Clin Psychiatry*. 2000;61(11):841–850.

Findling RL, Chang K, Robb A, et al. Adjunctive maintenance lamotrigine for pediatric bipolar I disorder: a placebo-controlled, randomized withdrawal study. *J Am Acad Child Adolesc Psychiatry*. 2015;54(12):1020–1031.

*Geddes JR, Calabrese JR, Goodwin GM. Lamotrigine for treatment of bipolar depression: independent meta-analysis and meta-regression of individual patient data from five randomised trials. *Br J Psychiatry*. 2009;194:4–9.

Geddes JR, Gardiner A, Rendell J, et al. Comparative evaluation of quetiapine plus lamotrigine combination versus quetiapine monotherapy (and folic acid versus placebo) in bipolar depression (CEQUEL): a 2 × 2 factorial randomised trial. *Lancet Psychiatry*. 2016;3(1):31–39.

Goldberg JF, Bowden CL, Calabrese JR. Six-month prospective life charting of mood symptoms with lamotrigine monotherapy versus placebo in rapid cycling bipolar disorder. *Biol Psychiatry*. 2008;63:125–130.

Ittasakul P, Johnson KR, Srivastava S, et al. Effectiveness of quetiapine plus lamotrigine maintenance therapy in challenging bipolar disorder patients. *J Affect Disord*.

2012;137(1–3):139–145.

Kessing LV, Hellmund G, Andersen PK. An observational nationwide register based cohort study on lamotrigine versus lithium in bipolar disorder. *J Psychopharmacol*. 2012;26(5):644–652.

Ketter TA, Brooks JO, Hoblyn JC, et al. Effectiveness of lamotrigine in bipolar disorder in a clinical setting. *J Psychiatr Res*. 2008;43:13–23.

*Ketter TA, Calabrese JR. Stabilization of mood from below versus above baseline in bipolar disorder: a new nomenclature. *J Clin Psychiatry*. 2002;63(2):146–151.

Ketter TA, Manji HK, Post RM. Potential mechanisms of action of lamotrigine in the treatment of bipolar disorders. *J Clin Psychopharmacol*. 2003;23(5):484–495.

Khalkhali M, Aram S, Zarrabi H, Kafie M, Heidarzadeh A. Lamotrigine augmentation versus placebo in serotonin reuptake inhibitors-resistant obsessive-compulsive disorder: a randomized controlled trial. *Iran J Psychiatry*. 2016;11(2):104–114.

Normann C, Hummel B, Scharer LO, Horn M, Grunze H, Walden J. Lamotrigine as adjunct to paroxetine in acute depression: a placebo-controlled, double-blind study. *J Clin Psychiatry*. 2002;63(4):337–344.

Passmore M, Garnham J, Duffy A, et al. Phenotypic spectra of bipolar disorder in responders to lithium versus lamotrigine. *Bipolar Disord* 2003;5:110–114.

Sajatovic M, Gildengers A, Al Jurdi RK, et al. Multisite, open-label, prospective trial of lamotrigine for geriatric bipolar depression: a preliminary report. *Bipolar Disord*. 2011;13(3):294–302.

Solmi M, Veronese N, Zaninotto L, et al. Lamotrigine compared to placebo and other agents with antidepressant activity in patients with unipolar and bipolar depression: a comprehensive meta-analysis of efficacy and safety outcomes in short-term trials. *CNS Spectr*. 2016;21(5):403–418.

Suppes T, Marangell LB, Bernstein IH, et al. A single blind comparison of lithium and lamotrigine for the treatment of bipolar II depression. *J Affect Disord*. 2008;111:334–343.

Tiihonen J, Hallikainen T, Ryynanen OP, et al. Lamotrigine in treatment-resistant schizophrenia: a randomized placebo-controlled crossover trial. *Biol Psychiatry*. 2003;54(11):1241–1248.

Tritt K, Nickel C, Lahmann C, et al. Lamotrigine treatment of aggression in female borderline-patients: a randomized, double-blind, placebo-controlled study. *J Psychopharmacol*. 2005;19(3):287–291.

van der Loos ML, Mulder PG, Hartong EG, et al; LamLit Study Group. Efficacy and safety of lamotrigine as add-on treatment to lithium in bipolar depression: a multicenter, double-blind, placebo-controlled trial. *J Clin Psychiatry*. 2009;70(2):223–231.

van der Loos ML, Mulder P, Hartong EG, et al. LamLit Study Group. Long-term outcome of bipolar depressed patients receiving lamotrigine as add-on to lithium with the possibility of the addition of paroxetine in nonresponders: a randomized, placebo-controlled trial with a novel design. *Bipolar Disord*. 2011;13(1):111–117.

Vestergaard K, Andersen G, Gottrup H, Kristensen BT, Jensen TS. Lamotrigine for central poststroke pain: a randomized controlled trial. *Neurology*. 2001;56(2):184–190.

Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) collaborative update of CANMAT guidelines for the management of patients with bipolar disorder: update 2013. *Bipolar Disord*. 2013;15(1):1–44.

Zoccali R, Muscatello MR, Bruno A, et al. The effect of lamotrigine augmentation of clozapine in a sample of treatment-resistant schizophrenic patients: a double-blind, placebo-controlled study. *Schizophr Res*. 2007;93(1–3):109–116.

33.10e Valproate

ROBERT M. POST, M.D., AND MARK A. FRYE, M.D.

Neuroscience-Based Nomenclature (NbN)	
Glutamate Modulator	
Pharmacologic Domain	Glutamate
Mechanism of Action	Yet to be determined

Valproic acid (Depakene), and its formulation derivatives divalproex sodium (Depakote), extended-release (ER), divalproex sodium (Depakote ER), and valproate sodium injection (Depacon), represent one of several first-generation mood-stabilizing anticonvulsants.

Valproate is currently approved for seizure disorders by the U.S. Food and Drug Administration (FDA) as (1) monotherapy or adjunctive therapy for complex partial seizures that occur in isolation or in conjunction with other types of seizures, (2) monotherapy and adjunctive therapy of simple and complex absence seizures, and (3) adjunctive therapy for patients with multiple seizures that include absence seizures.

Divalproex has an FDA indication in these same epilepsy syndromes but additionally has indications for (4) prophylaxis of migraine and (5) treatment of acute manic episodes associated with bipolar disorder. Intravenous (IV) valproate sodium is currently FDA approved as an alternative in patients for whom oral administration is temporarily not feasible (i.e., unresponsive or otherwise on nothing-by-mouth [NPO] status). ER divalproex sodium is currently FDA approved for migraine prophylaxis, monotherapy, and adjunctive therapy in complex partial seizures and acute mania.

HISTORY

The first clinical observations of valproate's anticonvulsant activity were in Europe in the 1970s. Shortly thereafter, the first reported evaluation of valpromide (the prodrug of valproate) in bipolar disorder was that of O. Lambert in France. Since then, valproate has been extensively evaluated in bipolar disorder (particularly the manic phase as exemplified by the 1995 indication for acute mania). Valproate has been studied to a much lesser extent in other psychiatric conditions and has not been approved for indications such as schizophrenia, schizoaffective disorder, bipolar depression, panic disorder, post-traumatic stress disorder (PTSD), alcohol withdrawal, and target symptoms of impulsivity, agitation, and mood lability.

CHEMISTRY

Valproic acid (*N*-dipropylacetic acid) is a simple-chain branched-chain carboxylic acid (Fig. 33.10e–1).

The commercial formulation derivatives available include valproic acid, which is available in syrup and capsules; valpromide (amide prodrug of valproic acid used in Europe); valproate sodium injection; divalproex sodium, which is an enteric-coated delayed-release complex of valproic acid and sodium valproate in a 1 to 1 ratio that is available in tablet and sprinkle formulation (can be opened and spread on food); and an ER formulation of divalproex sodium.

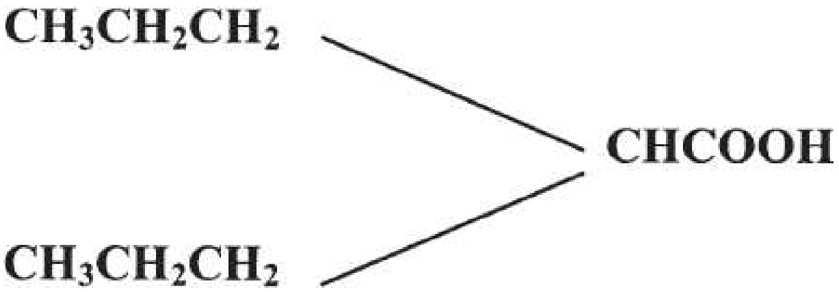

FIGURE 33.10e–1. Molecular structure of valproate.

PHARMACOLOGIC ACTIONS

Pharmacokinetics

All formulations of valproate are rapidly absorbed, with peak plasma concentrations obtained 3 to 4 hours after oral administration. The plasma half-life of valproate is 10 to 16 hours. However, when dosed in combination with drugs that induce cytochrome P450 enzymes (i.e., carbamazepine [Tegretol]), the half-life is shorter.

The ER formulation of divalproex sodium was designed for once-daily dosing, thereby maintaining lower maximum drug concentration (C_{max}) and higher minimum drug concentration (C_{min}) between dosing intervals. The ER formulation has 80% bioavailability of direct release (DR) divalproex sodium, and increasing the DR dose by 20% to an approximately equivalent ER dose with clinical monitoring would be a reasonable switch strategy.

Valproate is highly protein bound, primarily to albumin. It is generally thought that at serum levels less than 45 to 50 μg/mL, these binding sites are unsaturated, while at higher serum levels, these binding sites become saturated, with the unbound valproate level increasing. The unbound portion of valproate is considered to be pharmacologically active and can cross the blood–brain barrier. When used in combination with other psychotropic agents, valproate has a higher degree of plasma protein binding such that it can displace these drugs and increase their free levels (i.e., lamotrigine [Lamictal] and carbamazepine).

Valproate is metabolized primarily by hepatic glucuronidation and mitochondrial β-oxidation. Several of valproate's metabolites, 2-propyl-2-pentenoic acid and 2-propyl-3-oxopentanoic acid, have potent anticonvulsant activity.

Pharmacodynamics

Valproate's anticonvulsant and mood-stabilization mechanisms of action have not been specifically identified and may not necessarily be the same. Its anticonvulsant effects are rapid in onset, while antimanic and antidepressant effects are slower in onset and appear to require chronic administration. Several hypotheses for its mechanisms of action include: enhancement of brain γ-aminobutyric acid (GABA) levels via multiple actions of increased synthesis and decreased degradation; modulation of other neurotransmitters; blockade of voltage-sensitive Na^+ channels; and effects on extra hypothalamic neuropeptides, secondary messenger systems, and neuroprotection. Comparative effects with two other antimanic mood stabilizers, carbamazepine and lithium (Eskalith), are listed in Table 33.10e–1.

Preclinical studies have shown that valproate enhances GABAergic function in specific regions of the brain by a variety of presynaptic or postsynaptic mechanisms, or both, such as enhancement of GABA synthesis through activation of glutamate decarboxylase, prevention of GABA catabolism through inhibition of GABA transaminase, DR of GABA in response to calcium influx, reversal of the GABA transporter, and postsynaptic GABA type A ($GABA_A$) receptor activation effects. However, whether GABA enhancement is directly related to the antimanic effects of valproate remains controversial, and indirect inferences from the effects of other potent anticonvulsants that enhance brain GABA further suggest that this is unlikely. For example, other potent GABAergic drugs, tiagabine (Gabitril Filmtabs), gabapentin (Neurontin), and topiramate (Topamax), are not effective antimanic treatments.

Table 33.10e–1.

Common and Differential Effects of Valproate (VPA) Compared to Carbamazepine (CBZ) or Lithium (Li)

	CBZ	VPA	Li
Common Effects			
GABA turnover	↓	↓	↓
$GABA_B$ receptors in hippocampus (chronic)	↑	↑	↑
GABA levels	(↑)	↑	(↑)
↓ GSK-3B	+	+	+
Brain tryptophan and 5-HT	↑	↑	↑
Dopamine turnover	↓	↓	↓
NMDA-mediated Ca^{++} currents	↓	↓	↓
Arachidonic acid turnover	↓	↓	↓
Myoinositol transport (SMIT)	↓	↓	↓
Stroke neuroprotection	+	+	+
BDNF	↑	↑	↑
Lithium-Shared Effects			
Inhibition of ERK MAP kinase path	−	+	+
↓ PKC	−	+	+
↓ MARCKS levels	−	+	+
↑ Bcl-2	−	+	+
↑ Neurogenesis	−	+	+
↓ β-Catenin	−	+	+
↑ AP-1 Binding	−	+	+
Carbamazepine-Shared Effects			
Na_i+ block	+	(+)	−
K efflux	↑	↑	−
[^{3}H] picrotoxinin binding	↓	↓	−
Release of aspartate, glutamate	↓	↓	−
Somatostatin levels	↓	↓	−

Histone, deacetylase inhibition	+	+	−
Block completed kindled seizures	+	+	−
Unique Effects of Each Drug			
Inositol Monophosphatases	−	↓	↑
Absence seizures	↑	↓	−
Blocks kindling development	−	+	−
Inhibit after-discharge spread in	−	+	−
hippocampus in cortex	+	−	−
Potential anticancer effects (AML)	−	+	−
Forms ester with coenzyme A	−	+	−
↑ PPARs	−	+	−

GABA, γ-aminobutyric acid; GSK, glycogen synthase kinase; 5-HT, serotonin; NMDA, *N*-methyl-d-aspartate; SMIT, *myo*-inositol transport system; BDNF, brain-derived neurotrophic factor; ERK, extracellular-signal-regulated kinases; PKC, protein kinase C; MARCKS, myristoylated alanine-rich C kinase substrate; AP, activator protein; AML, acute myelogenous leukemia; PPARs, peroxisome proliferator-activated receptors.

In addition, GABA deficits, as measured in cerebrospinal fluid (CSF) by magnetic resonance spectroscopy (MRS), or plasma analysis, have not been consistently reported in clinical studies of affectively ill patients. GABA assessments after drug treatment have also produced inconsistent results. Valproate treatment increased GABA in the CSF of 31 pediatric epileptic patients and 6 schizophrenic patients; by MRS analysis, valproate increased occipital GABA in 6 euthymic bipolar patients. However, no increase in GABA, as measured by CSF analysis in six schizophrenic patients or by MRS analysis in five refractory epileptic patients treated with valproate, has also been reported.

The discordance between the CSF, MRS, and plasma studies remains unresolved; different study populations, collection and storage times of samples, lengths of treatment, and methodologies of GABA assay need to be standardized to better elucidate the role of GABAergic mechanisms in the therapeutic effects of valproate.

Valproate's effect on less well-studied neurotransmitters and neuropeptides has included glutamate (increased release in mouse brain), serotonin (increased extracellular level in rat hippocampus and striatum), dopamine (increased extracellular level in rat), somatostatin (decreased CSF level), and corticotropin-releasing factor (CRF) (decreased CRF in the median eminence, raphe nuclei, and decreased CRF messenger ribonucleic acid [mRNA] in the central nucleus of the amygdala and paraventricular nucleus of the hypothalamus). Although these neurotransmitters and neuropeptides have all been implicated in the pathophysiology of mania, depression, or psychosis, it is not known whether valproate’s modulation of these neurotransmitters and neuropeptides contributes to its mood-stabilizing properties.

Intracellular signaling systems similar to those of lithium have also been observed with valproate. Like lithium, valproate stimulates glutamate release and, through activation of the *N*-methyl-D-aspartate (NMDA) receptor, has been shown to increase inositol-1,4,5-triphosphate. Additionally, lithium, valproate, and carbamazepine, but not amitriptyline (Elavil), have been shown to inhibit the high-affinity *myo*-inositol transport system in astrocytes; this inhibition of inositol and subsequent reduced intraneuronal level of inositol (*inositol depletion*) may be related to the mood-stabilization mechanism of valproate. However, although lithium decreases inositol monophosphatase activity, valproate increases it.

Similar actions have also been observed with lithium and valproate in the downregulation of protein kinase C (PKC) activity and downregulation of myristoylated alanine-rich C kinase substrate (MARCKS) activity (a major substrate of PKC) and increased level of Bcl-2 (a neuroprotective protein). PKC inhibition is a likely contributor to antimanic efficacy. In one open and two double-blind, placebo-controlled studies, the potent PKC inhibitor tamoxifen (Nolvadex), which also has antiestrogen receptor effects, has shown antimanic efficacy.

Like lithium, valproate increases Bcl-2, brain-derived neurotrophic factor (BDNF), and neurogenesis, but the increases in hippocampal volume and cortical gray matter seen with lithium have not been consistently observed with valproate. However, patients on valproate had increased volumes of their cingulate cortices and increased *N*-acetyl-aspartate (NAA) to creatine (Cre) or choline (Cho) ratios in the hippocampus compared with those not treated with valproate. Since BDNF decreases during episodes of depression and mania, valproate’s ability to increase BDNF could be important to some of its clinical actions.

Valproate is also positive in animal models of stroke and, surprisingly, is effective not only when given before middle cerebral artery ligation but (like lithium) when given shortly after the ligation as well. In contrast to lithium, valproate inhibits histone deacetylase, which could account for many of its effects on gene transcription. It is also possible that this effect contributes to valproate's teratogenicity (as described below) or to its possible therapeutic actions in conjunction with hydralazine (an inhibitor of DNA methylation) in reversing treatment resistance to chemotherapy of malignancies. In contrast to valproate, valpromide does not inhibit histone deacetylase, such that a role for this mechanism could be revealed in comparative studies.

CLINICAL DRUG STUDIES

Acute Mania

The early uncontrolled studies of valproate (divalproex or valpromide) in acute mania suggested efficacy. Subsequent controlled studies of valproate in acute mania confirmed this view; these studies are reviewed in Table 33.10e–2.

Table 33.10e–2.
Controlled Studies of Valproate or Divalproex in Acute Mania

Study	Design (N)	Primary Outcome Variable	Results	Comment
Pope et al., 1991	3-week, randomized, db, pc, DSM-III-R mania, unresponsive to Li ($N = 36$)	% ↓ YMRS	VAL 54% ↓ vs. PLC 5% ↓ ($P = .003$), ITT ($P = .013$)	Response 1–4 days after achieving serum VAL >50 μg/mL
Bowden et al., 1994	3-week, multisite, randomized, db, pc, RDC confirmed mania ($N = 179$) DVPX vs. Li vs. PLC in 2:1:2 ratio. Dosage titrated to DVPX (150 μg/mL) and Li (1.5 mmol/L), except if adverse event was present	50% ↓ MRS	Mean dose/level DVPX: 2,000 mg/93.2 μg/mL) Li: 1,950 mg/1.2 mmol/L ITT: DVPX, 46% ($P = .004$ vs. PLC); Li, 49% ($P = .025$ vs. PLC); PLC 25%	All rapid-cycling patients ($N = 8$) randomized to DVPX; therefore, no comparative analysis possible 48% of current Li responders reported past Li inefficacy or intolerance
Freeman et al., 1992	3-week, randomized, db, DSM-III-R mania ($N = 27$) DVPX (titrated to 3,000 mg—lower dose for response or side effect) vs. Li (titrated to response, side effect, 1,800 mg q.d., or level 1.5 mmol/L)	50% ↓ SADS-C mania	Li levels (0.8–1.4 mmol/L), mean DVPX level (98 ± 17 μg/mL) Li 12/13 (92%) vs. DVPX 9/14 (64%)	Pretreatment depression significantly higher in DVPX responders vs. nonresponders ($P < .05$) Post-treatment mania scores higher in mixed vs. nonmixed manics treated with Li ($P < .05$)
McElroy et al., 1996	1-week, randomized, sb, psychotic mania, rapid load DVPX ($N = 21$, 20 mg/kg/day) vs. HAL ($N = 15$, 0.2 mg/kg/day)	Mean change YMRS	DVPX 43% YMRS ↓; HAL 35% YMRS ↓	Equal efficacy Greater use of benztropine in HAL group
Hirschfeld et al., 1999	10-day, multisite, randomized, db, DSM-IV acute mania (YMRS ≥ 14) DVPX load ($N = 20$, 30 mg/kg/day on days 1–2, 20 mg/kg/day on days 3–10) vs. DVPX nonload ($N = 20$) vs. Li ($N = 19$)	Safety and tolerability DVPX load	84% DVPX load had level >50 μg/mL by day 3 vs. 30% DVPX nonload	No difference in adverse event or YMRS change
Muller-Oerlinghausen et al., 2000	3-week, multisite, randomized, db, pc adjunct VAL (20 mg/kg/day) vs. PLC to neuroleptic ($N = 136$) Step-wise dose reduction of neuroleptic	Mean dose ↓ neuroleptic at study end	Day 21 VAL level (80.3 ± 22.3 μg/mL) 43% ↓ of neuroleptic in adjunct VAL vs. 14% ↓ in adjunct PLC ($P = .0007$)	Decreased neuroleptic burden in VAL patients
Vasudev et al., 2000	4-week, randomized, db, CBZ vs. VAL ($N = 30$)	Mean change in YMRS	VAL (75.4 μg/mL) greater YMRS ↓ vs. CBZ (6.2 μg/mL) ($P = .023$)	Correlation between VAL level and ↓ YMRS at week 2
Tohen et al., 2002	3-week, multisite, randomized, db, parallel DSM-IV mania (YMRS ≥ 20) ($N = 251$) DVPX ($N = 125$, flexible dose of 500–2,500 mg q.d.) vs. OLZ ($N = 123$, 5–20 mg q.d.)	Mean change YMRS Response (50% ↓ YMRS) Remission (YMRS ≤ 12)	DVPX (mean dose 1,401 mg q.d., serum level 79.4 μg/mL), OLZ (mean dose 17.4 mg) YMRS ↓ DVPX 10.4 points vs. OLZ 13.4 points ($P =$ N.S.) Response and remission: DVPX 42.3% and 34.1% vs. OLZ 54.4% and 47.2% ($P = .058$; $P = .039$)	Weight gain 2.5 kg OLZ vs. 0.9 kg DVPX Somnolence, dry mouth, ↑ appetite, tremor, neck rigidity more common with OLZ Nausea, diarrhea more common with DVPX
Zajecka et al., 2002	3-week, multisite, randomized, db, parallel DSM-IV mania (MRS ≥ 25) ($N = 120$) DVPX ($N = 43$; loading 20 mg/kg/day) vs. OLZ ($N = 57$; start dose 10 mg with ↑ for persistent symptoms)	Mean MRS change Response (50% ↓ MRS)	DVPX, mean dose of 2,115 mg; OLZ mean dose of 15 mg MRS ↓ DVPX 14.8 points vs. OLZ 17.2 points ($P =$ N.S.) ITT: Response (DVPX 53%, OLZ 62%, $P =$ N.S.)	Weight gain, somnolence, rhinitis, edema, ↑ cholesterol more common with OLZ ↑ Monocytes, ↑ eosinophils, ↓ neutrophils, ↓ platelets more common with DVPX
Bowden et al., 2006	3-week, multisite, randomized db, pc, DSM-IV acute mania DVPX started at 25 mg/kg day 1 ↑ 500 mg on day 3 → 85–125 μg/mL (Dosing once daily in AM)	Response (50% ↓ YMRS)	ITT: Response DVPX 48% PLC 34% ($P = .02$) Mean final dose 3,057 mg per day (35.4 mg/kg) → mean serum level 95.9 μg/mL	On DVPX: somnolence, dizziness, complaints; weight gain: 1.8 kg DVPX 0.3 kg PLC Total cholesterol −13.5 mg/dL DVPX +2.5 mg/dL PLC

CBZ, carbamazepine (Tegretol); db, double blind; DVPX, divalproex sodium (Depakote); HAL, haloperidol (Haldol); HAMD, Hamilton Depression Scale; ITT, intent-to-treat analysis; Li, lithium (Eskalith); MRS, Mania Rating Scale; N.S., not significant; OLZ, olanzapine (Zyprexa); pc, placebo controlled; PLC, placebo; RDC, Research Diagnostic Criteria; SADS-C, Schedule for Affective Disorders and Schizophrenia-Change; sb, single blind; VAL, valproate; YMRS, Young Mania Rating Scale.

The first placebo-controlled studies by Pope and colleagues (1991) and Bowden et al. (1994) were pivotal in receiving an FDA indication for acute mania. Post hoc analyses of the Bowden study have suggested that, although there was no difference from lithium in the response rate for patients with euphoric mania, patients with concomitant depressive symptoms (i.e., dysphoric mania) or a cumulative lifetime episode burden of greater than 4 prior depressive episodes or 11 prior manic episodes responded more poorly to lithium. Furthermore, four distinct manic subtypes (anxious-depressed, psychotic, classic, and irritable) based on pretreatment symptom cluster analyses confirmed that the classic manic subtype responded similarly to divalproex and lithium, but the irritable manic subtype responded significantly better to divalproex than lithium ($P < .004$) or placebo ($P < .04$). The three treatments did not differ in response rates to depressive or psychotic subtypes.

Two studies have suggested that a serum valproate level of greater than 45 to 50 μg/mL is associated with an antimanic response. A multicenter study of the divalproex ER preparation in acute mania was again highly positive compared with placebo for both mania and mixed patients and in those with and without psychosis. Starting dose and rate of titration may have been too high because the side-effect profile was not diminished compared with the acute release studies. Serum levels of 71.3 to 85 μg/mL or higher appeared to be the therapeutic target since those under this concentration show fewer good responses (smaller effect sizes [0.4 to 0.8]) compared with those at higher blood levels (effect sizes 0.9 to 1.1). The ER preparation was dosed on a once-daily basis in the morning, but clinically, one would recommend single bedtime dosing to enhance sleep and reduce side effects during awake time and increase compliance. Many clinicians would similarly recommend that the IR preparation could be given all at night.

Because of valproate's robust effects in acute mania in adults, it was surprising that it failed to separate from placebo in a study of childhood- and adolescent-onset mania. Concomitant stimulant medication was allowed in that study, which may have confounded the results. In another study on childhood mania, response to valproate was inferior to that of risperidone.

Given that there are many FDA-approved conventional mood-stabilizing agents for the treatment of acute mania (lithium, valproate, long-acting carbamazepine [Equetro], olanzapine [Zyprexa], risperidone [Risperdal], quetiapine [Seroquel], aripiprazole [Abilify], ziprasidone [Geodon], and cariprazine [Vraylar]), the few controlled comparison studies provide only preliminary information regarding the selection of an antimanic agent. Side-effect profile and long-term tolerability would appear to provide an important rationale for the choice of antimanic agents.

Although the typical antipsychotic agents have been extensively used for the treatment of acute mania in the past, they represent a somewhat outmoded treatment because they are associated with more postmanic depressive symptom severity and breakthrough episodes of depression (vs. lithium) and an adverse side-effect profile. Typical antipsychotics are associated with considerable side-effect liabilities, including neuroleptic-induced dysphoria, acute dystonia, extrapyramidal side effects, and, most importantly, long-term risk of tardive dyskinesia in some 20% to 40% of bipolar patients (even those treated intermittently).

As reviewed previously, the post hoc analyses of comparison studies of lithium versus divalproex would suggest that depressive or irritable symptoms in mania are more effectively treated with divalproex. Other possible correlates of a relatively positive response are listed in Table 33.10e–3. There is one large open trial in rapid-cycling patients of valproate prophylaxis as monotherapy in lithium-refractory patients (N = 43) or as add-on therapy (N = 101) with a greater than 70% antimanic or nonmixed response. The only comparison study of valproate and carbamazepine suggested greater efficacy and dose-dependent antimanic response with valproate.

Table 33.10e–3.

Potential Correlates of Response to Valproate

	Strength of Association
Diagnosis-Related Factors	
Diagnosis of bipolar disorder (vs. schizoaffective or schizophrenia disorder)	+++
Dysphoric (or mixed mania)	+++
Bipolar II disorder	++
Clinical Course Variables	
Presence of rapid cycling	+
Decreasing or stable episode frequency	++
Later age of onset of illness (but see Fig. 33.10e–2A,B)	+
Shorter duration of illness	+
Long and severe course of illness (see Figs. 33.10e–2 and 33.10e–3)	++
Comorbid Psychiatric Diagnosis	
Comorbid alcohol abuse	+++
Comorbid mental retardation	+
Comorbid anxiety and panic attacks	+++
Comorbid PTSD	+
Treated hypothyroidism	+
Marked sleep disturbance	+
Explosive dyscontrol and aggression	++
Associated Neurologic Factors	
Presence of nonparoxysmal EEG abnormalities (not subsequently replicated)	±
Organic affective disorder	±
History of closed head trauma	+
Comorbid migraine	+++
Family History	

Positive family history of mood disorder	+
Treatment-related variables:	
Patients who are lithium naïve (especially BP-II)	+
Higher serum valproate concentrations at days 3–6 of treatment	
Higher serum valproate concentrations with more severe illness: cyclothymia < BP-II < BP-I	++
Brain Imaging	
Low activity in the frontal and anterior cingulate and medial frontal gyrus	+
Plasma GABA	
Not low (near normal)	+

PTSD, post-traumatic stress disorder; EEG, electroencephalogram; BP, bipolar; GABA, γ-aminobutyric acid; +++, well documented in several open studies or one controlled study; ++, good support in literature; +, fair support; ±, weak support or equivocal.

Data from West SA, Strakowski SM, Sax KW, McElroy SL, Keck PE Jr, McConville BJ. Phenomenology and comorbidity of adolescents hospitalized for the treatment of acute mania. *Biol Psychiatry*. 1996;39(6):458–460 and Post RM, Speer AM, Obrocea GV, Leverich GS. Acute and prophylactic effects of anticonvulsants in bipolar depression. *Clin Neuroscience Res*. 2002;2(3–4):228–251.

Two comparative studies of divalproex versus olanzapine were similar in design, with the exception of dose and titration. In one smaller study that started with a rapid loading dose of divalproex (mean dose of 2,115 mg vs. olanzapine mean dose of 15 mg), the mean reduction in manic symptoms and response rates were not statistically different. In the larger-powered study that started with a nonloading and lower divalproex dose (mean dose of 1,401 mg vs. olanzapine mean dose of 17.4 mg), response and remission rates were significantly in favor of olanzapine. A 12-week comparative study in acute mania suggested the superiority of quetiapine to valproate for efficacy but with a higher side-effect burden.

Table 33.10e–4.
Controlled Studies of Valproate and Divalproex Sodium in Bipolar Disorder (Acute Depression)

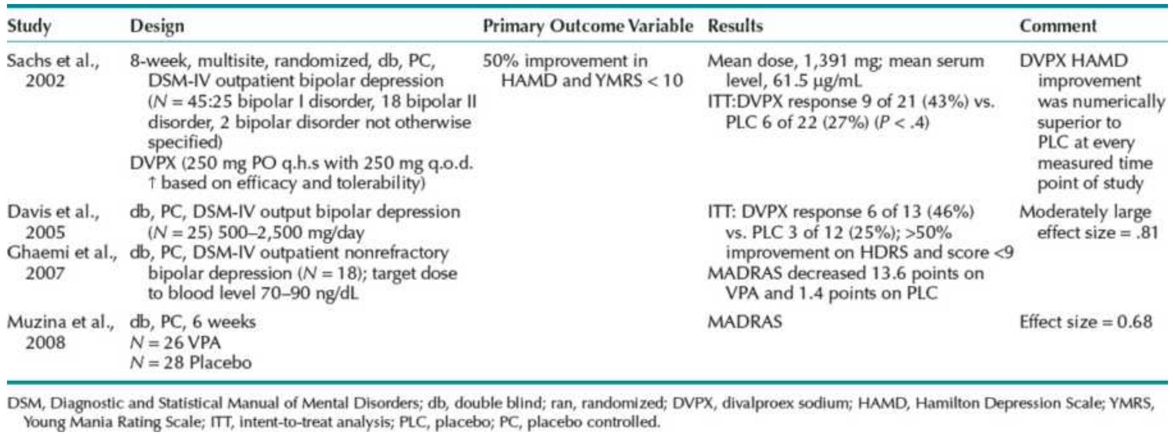

Study	Design	Primary Outcome Variable	Results	Comment
Sachs et al., 2002	8-week, multisite, randomized, db, PC, DSM-IV outpatient bipolar depression (*N* = 45:25 bipolar I disorder, 18 bipolar II disorder, 2 bipolar disorder not otherwise specified) DVPX (250 mg PO q.h.s with 250 mg q.o.d. ↑ based on efficacy and tolerability)	50% improvement in HAMD and YMRS < 10	Mean dose, 1,391 mg; mean serum level, 61.5 μg/mL ITT:DVPX response 9 of 21 (43%) vs. PLC 6 of 22 (27%) (*P* < .4)	DVPX HAMD improvement was numerically superior to PLC at every measured time point of study
Davis et al., 2005	db, PC, DSM-IV output bipolar depression (*N* = 25) 500–2,500 mg/day		ITT: DVPX response 6 of 13 (46%) vs. PLC 3 of 12 (25%); >50% improvement on HDRS and score <9	Moderately large effect size = .81
Ghaemi et al., 2007	db, PC, DSM-IV outpatient nonrefractory bipolar depression (*N* = 18); target dose to blood level 70–90 ng/dL		MADRAS decreased 13.6 points on VPA and 1.4 points on PLC	
Muzina et al., 2008	db, PC, 6 weeks *N* = 26 VPA *N* = 28 Placebo		MADRAS	Effect size = 0.68

DSM, Diagnostic and Statistical Manual of Mental Disorders; db, double blind; ran, randomized; DVPX, divalproex sodium; HAMD, Hamilton Depression Scale; YMRS, Young Mania Rating Scale; ITT, intent-to-treat analysis; PLC, placebo; PC, placebo controlled.

Acute Bipolar Depression

Conventional mood stabilizers have been less well studied in the depressive phase of bipolar disorder. Several open trials have clearly suggested antidepressant properties with valproate, generally at lower doses than studies in acute mania. In a 12-week open trial of 19 patients with bipolar II depression, divalproex sodium monotherapy (with a mean dose of 882 mg every night and a mean level of 80.7 μg/mL), 63% of the cohort responded to treatment as defined by a 50% reduction in the Hamilton Rating Scale for Depression.

As presented in Table 33.10e–4, a controlled study of divalproex monotherapy in acute bipolar I or II depression did not statistically separate from placebo at the 8-week study endpoint. The small sample size (*N* = 45) does not allow definitive conclusions about the acute antidepressant efficacy of divalproex. However, in another small randomized-controlled trial including 25 patients, valproate not only showed superior antidepressant effects compared with placebo but showed even more clinically and statistically robust antianxiety effects in monotherapy of bipolar depression. A more recent double-blind placebo-controlled study showed marked superiority of valproate over placebo with a moderately large effect size of 0.81.

Current guidelines emphasize the importance of maximizing mood stabilizer or atypical antipsychotic treatment and revising these agents if needed before starting treatment with conventional unimodal antidepressant therapy.

Prophylactic Mood Stabilization

There are six controlled studies that have evaluated the efficacy of valproate or divalproex in long-term mood stabilization (Table 33.10e–5). These studies include one open randomization comparing valpromide versus lithium monotherapy; one open randomization comparing valproate versus lithium monotherapy in a child, adolescent, and young adult population; one placebo-controlled, adjunctive divalproex trial to lithium-maintained bipolar I patients; one 52-week, parallel design comparing lithium, divalproex, and placebo monotherapy in recently recovered manic patients; and one controlled comparison study between divalproex and olanzapine monotherapy in those patients who had presented in an acute manic episode.

Several conclusions can be drawn regarding these studies. First, one study suggested that valproate had better prophylaxis against mania compared to depressive episodes. Conversely, the largest controlled study to date did not find any difference between divalproex, lithium, or placebo in time to recurrence of a mood episode or, specifically, a manic episode. However, almost 200 patients, who were initially enrolled as acutely manic, were unable to fully recover and thus were not randomized into the prophylactic phase; this methodologic limitation may have yielded a type II error in a cohort of only randomized mildly ill patients, thereby making it more difficult to show differences between active drugs and the low rate of manic relapse even on placebo. A secondary outcome measure, however, suggested greater depression prophylaxis with divalproex than placebo in the monotherapy analysis and in a subgroup of patients who were subsequently treated with adjunctive antidepressant therapy (i.e., sertraline [Zoloft] or paroxetine [Paxil]) added to divalproex versus placebo.

In the 47-week acute mania comparison study of divalproex versus olanzapine, the median time to remission (YMRS ≤12) took significantly longer for divalproex (62 days) than for olanzapine (14 days, $P = .047$). Similarly, there was greater improvement in the YMRS with olanzapine versus divalproex at weeks 2 to 15 and week 23; however, from weeks 30 to 47, there was no difference in the two groups. Once remitted, there was no difference in relapse rates between the two groups.

In a second 18-month maintenance study of lithium versus valproate, those patients who had deteriorated or who were intolerant of lithium treatment had a higher response rate with a subsequent crossover to valproate in comparison to those patients who were initially treated with valproate with a subsequent crossover to lithium.

In a maintenance study comparing lithium plus valproate versus lithium alone, the combination was more effective than lithium alone; however, the adverse event profile was higher with the combination treatment. In the largest practical clinical trial (using a randomized open design of x months duration), Geddes and colleagues found clear superiority of lithium compared to valproate, with the combination marginally better than lithium alone. Despite an indication only in acute mania and some evidence of inferiority compared to lithium, there is wider use of valproate than lithium in the United States in the maintenance phase of bipolar disorder, case examples of which are illustrated in Figures 33.10e–2A,B and 33.10e–3. The latest Cochrane analysis in 2013 concluded that limited evidence supported the efficacy of valproate in the maintenance treatment of bipolar disorder.

Valproate in combination with lithium was evaluated in adults with rapid-cycling disorder in preparation for a blind randomization to monotherapy. Disappointingly, only 25% of the evaluable sample responded to the combination acutely, and only 19% of the intact to treat (ITT) sample responded. After randomization, relapses were about 50% on either lithium or valproate monotherapy, suggesting a very low valproate monotherapy response rate (about 12%) in this difficult-to-treat but common subtype of rapid cycling (which makes up 40% to 60% of outpatients in some series). A return to the combination of lithium plus valproate produced a high reresponse rate.

Table 33.10e–5. Controlled Long-Term Evaluation of Divalproex Sodium in Bipolar Disorder

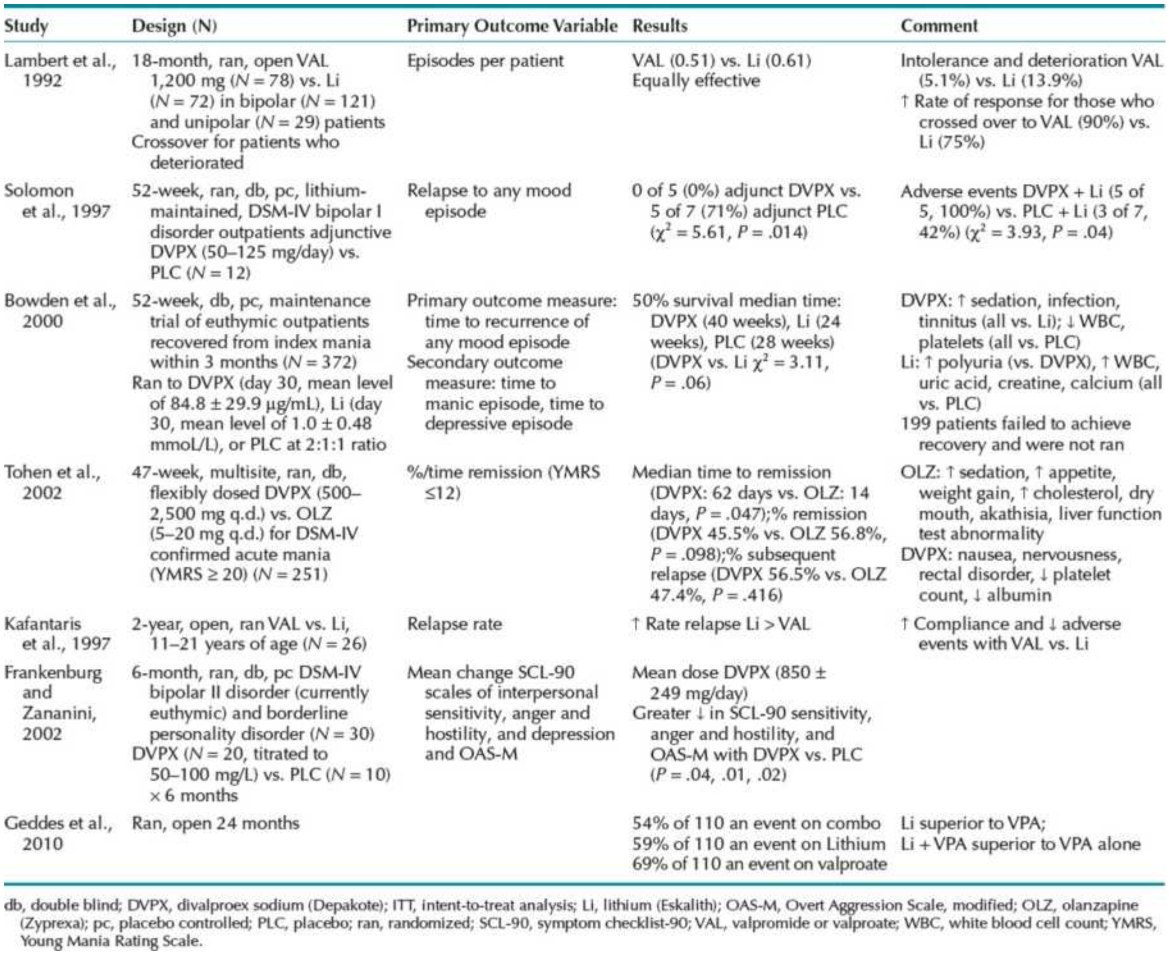

Study	Design (N)	Primary Outcome Variable	Results	Comment
Lambert et al., 1992	18-month, ran, open VAL 1,200 mg ($N = 78$) vs. Li ($N = 72$) in bipolar ($N = 121$) and unipolar ($N = 29$) patients Crossover for patients who deteriorated	Episodes per patient	VAL (0.51) vs. Li (0.61) Equally effective	Intolerance and deterioration VAL (5.1%) vs. Li (13.9%) ↑ Rate of response for those who crossed over to VAL (90%) vs. Li (75%)
Solomon et al., 1997	52-week, ran, db, pc, lithium-maintained, DSM-IV bipolar I disorder outpatients adjunctive DVPX (50–125 mg/day) vs. PLC ($N = 12$)	Relapse to any mood episode	0 of 5 (0%) adjunct DVPX vs. 5 of 7 (71%) adjunct PLC ($\chi^2 = 5.61$, $P = .014$)	Adverse events DVPX + Li (5 of 5, 100%) vs. PLC + Li (3 of 7, 42%) ($\chi^2 = 3.93$, $P = .04$)
Bowden et al., 2000	52-week, db, pc, maintenance trial of euthymic outpatients recovered from index mania within 3 months ($N = 372$) Ran to DVPX (day 30, mean level of 84.8 ± 29.9 μg/mL), Li (day 30, mean level of 1.0 ± 0.48 mmol/L), or PLC at 2:1:1 ratio	Primary outcome measure: time to recurrence of any mood episode Secondary outcome measure: time to manic episode, time to depressive episode	50% survival median time: DVPX (40 weeks), Li (24 weeks), PLC (28 weeks) (DVPX vs. Li $\chi^2 = 3.11$, $P = .06$)	DVPX: ↑ sedation, infection, tinnitus (all vs. Li); ↓ WBC, platelets (all vs. PLC) Li: ↑ polyuria (vs. DVPX), ↑ WBC, uric acid, creatine, calcium (all vs. PLC) 199 patients failed to achieve recovery and were not ran
Tohen et al., 2002	47-week, multisite, ran, db, flexibly dosed DVPX (500–2,500 mg q.d.) vs. OLZ (5–20 mg q.d.) for DSM-IV confirmed acute mania (YMRS ≥ 20) ($N = 251$)	%/time remission (YMRS ≤12)	Median time to remission (DVPX: 62 days vs. OLZ: 14 days, $P = .047$);% remission (DVPX 45.5% vs. OLZ 56.8%, $P = .098$);% subsequent relapse (DVPX 56.5% vs. OLZ 47.4%, $P = .416$)	OLZ: ↑ sedation, ↑ appetite, weight gain, ↑ cholesterol, dry mouth, akathisia, liver function test abnormality DVPX: nausea, nervousness, rectal disorder, ↓ platelet count, ↓ albumin
Kafantaris et al., 1997	2-year, open, ran VAL vs. Li, 11–21 years of age ($N = 26$)	Relapse rate	↑ Rate relapse Li > VAL	↑ Compliance and ↓ adverse events with VAL vs. Li
Frankenburg and Zananini, 2002	6-month, ran, db, pc DSM-IV bipolar II disorder (currently euthymic) and borderline personality disorder ($N = 30$) DVPX ($N = 20$, titrated to 50–100 mg/L) vs. PLC ($N = 10$) × 6 months	Mean change SCL-90 scales of interpersonal sensitivity, anger and hostility, and depression and OAS-M	Mean dose DVPX (850 ± 249 mg/day) Greater ↓ in SCL-90 sensitivity, anger and hostility, and OAS-M with DVPX vs. PLC ($P = .04, .01, .02$)	
Geddes et al., 2010	Ran, open 24 months		54% of 110 an event on combo 59% of 110 an event on Lithium 69% of 110 an event on valproate	Li superior to VPA; Li + VPA superior to VPA alone

db, double blind; DVPX, divalproex sodium (Depakote); ITT, intent-to-treat analysis; Li, lithium (Eskalith); OAS-M, Overt Aggression Scale, modified; OLZ, olanzapine (Zyprexa); pc, placebo controlled; PLC, placebo; ran, randomized; SCL-90, symptom checklist-90; VAL, valpromide or valproate; WBC, white blood cell count; YMRS, Young Mania Rating Scale.

In a parallel study in children and adolescents with mania, some 43% responded acutely to the lithium plus valproate combination, but upon randomization to either monotherapy, some two-thirds failed to maintain their improvement. However, the great majority reresponded to the combination when it was reintroduced. Most of these youngsters also required stimulant comedication, and several required an atypical antipsychotic as well. These systematic prospective data in a double-blind maintenance study suggest that as in many adult patients, valproate monotherapy is very often insufficient, and combination treatment with lithium or an atypical antipsychotic is typically required in children for sustained mood stabilization and in adults with rapid cycling bipolar disorder.

A recent randomized multicenter study of childhood mania, comparing valproate, lithium, or the atypical antipsychotic risperidone, supported this viewpoint. Geller and colleagues found that risperidone was superior to both lithium and valproate. The percentages with CGI-BP improvement of 2 (much improved) or less (very much improved) were risperidone, 68.5%; lithium, 35.6%; and valproate, 24.0%. However, risperidone had more side effects, such as weight gain and prolactin increases. When nonresponders were rerandomized to the addition of one of the other two treatments, those receiving risperidone as part of the regimen again fared the best.

In a secondary prevention study of children at high risk for bipolar I illness by virtue of both the presence of (1) subthreshold symptoms or a bipolar not otherwise specified presentation and (2) a positive parental history of bipolar illness, valproate was no more effective than placebo. Whether other monotherapies would fare better in this early attempt at secondary prevention remains to be established. Thus, because of the documented inefficacy of valproate monotherapy in a large double-blind study in childhood and adolescent acute mania, its inferiority to risperidone in another study, and its lack of prevention of the disorder in children at high risk, the combination of valproate with lithium or with an atypical such as risperidone may be of critical importance in the management of these children. The role of valproate in conjunction with the better tolerated atypical antipsychotics in long-term treatment also needs to be further explored.

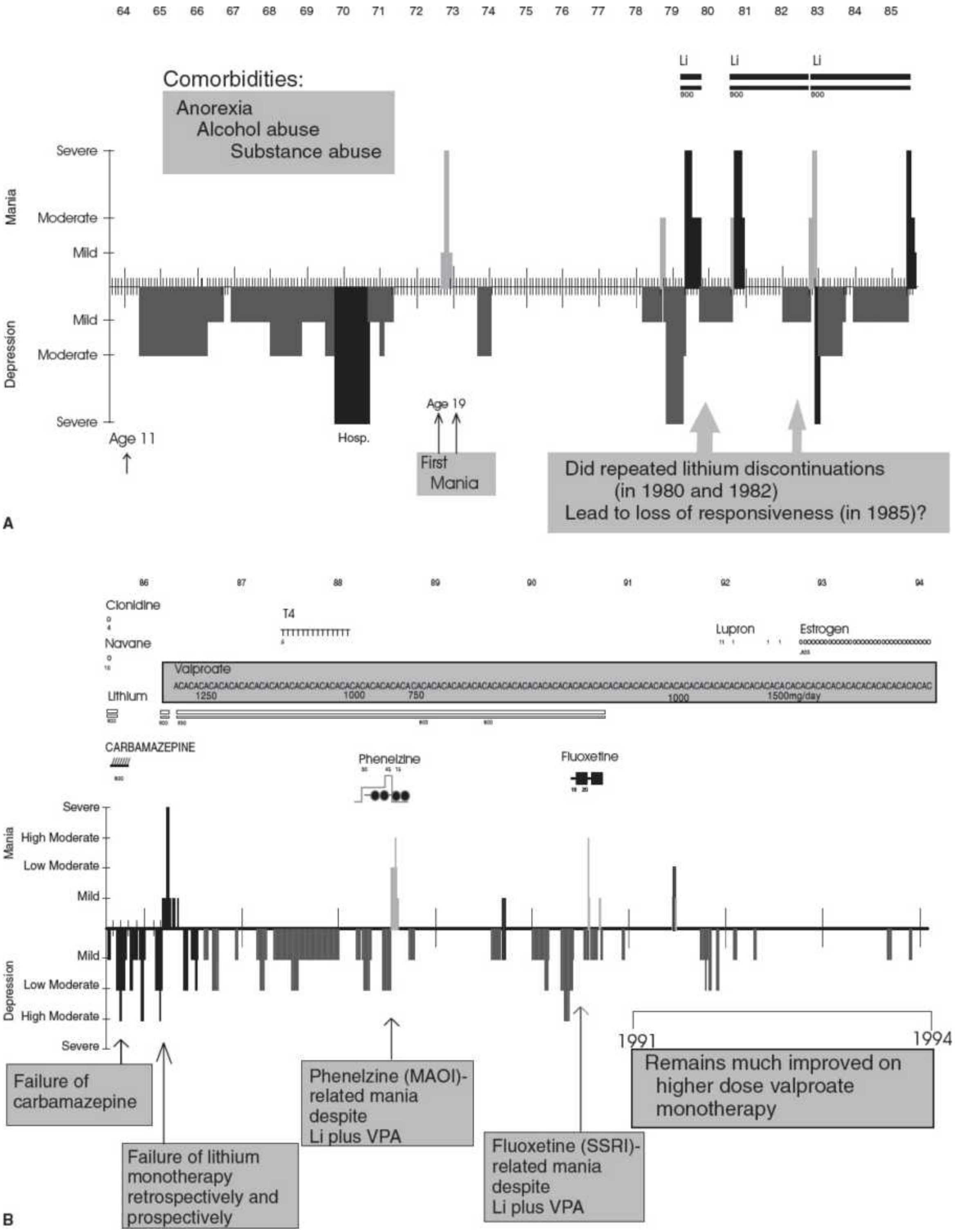

FIGURE 33.10e–2. **A:** Early-onset illness (at age 11) with depression and multiple comorbidities and ultimate failure to respond to lithium in 1985. **B:** Good prophylactic response to valproate in this lithium and carbamazepine-nonresponder. She did well on the combination of lithium and valproate and then on valproate monotherapy since 1990. She had no further hospitalizations for mania (from 1986 to the present), but had occasional

depressive recurrences. Note shift in time scale from (A). (From Post RM, Leverich GS. *Treatment of Bipolar Illness: A Casebook for Clinicians and Patients. Treatment of Bipolar Illness*. W.W. Norton & Co.; 2008:198–200.)

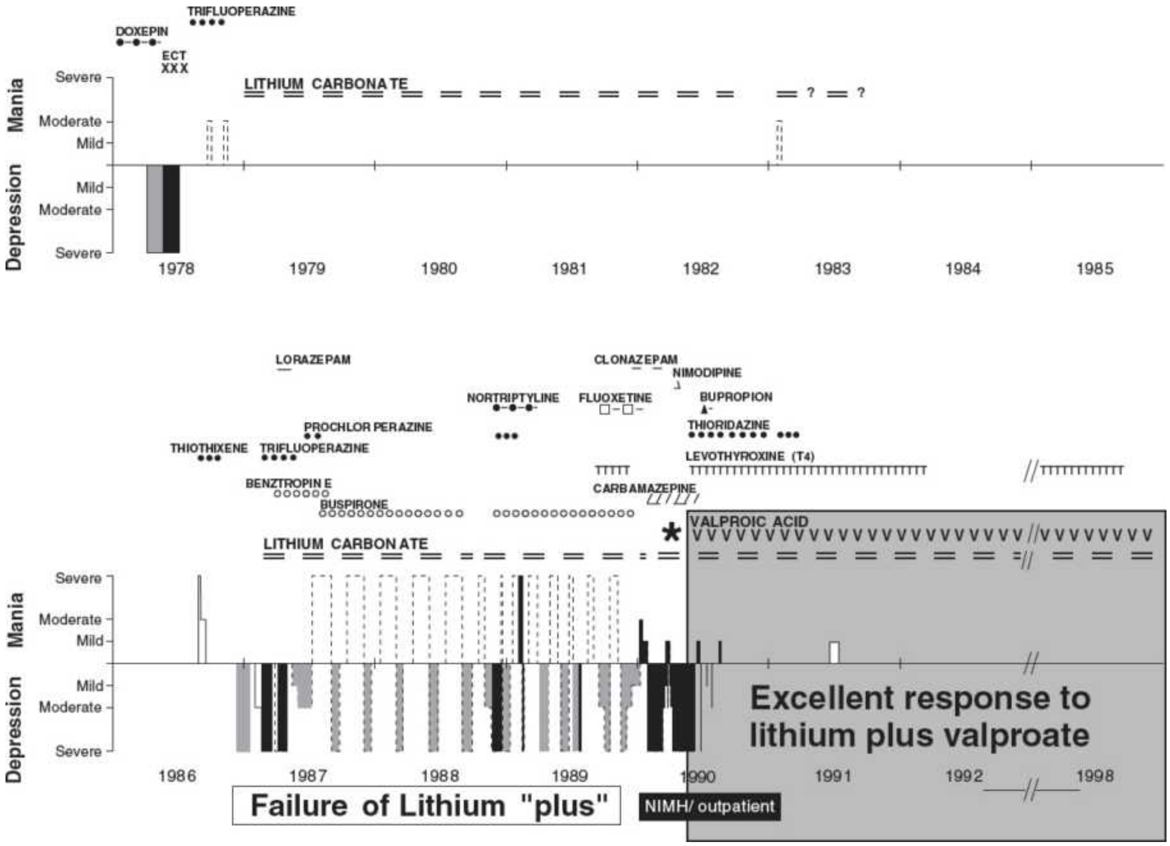

FIGURE 33.10e–3. Valproate augmentation in a lithium-refractory bipolar patient with profound recurrent psychotic depression. Reemergence of the illness in 1986 was unresponsive to the reinstitution of lithium, buspirone, a variety of antidepressants, adjunctive antipsychotics, and minor tranquillizers. Carbamazepine, alone and in combination with lithium, was effective in reducing the amplitude of the manias, but had no impact on the patient's severe psychotic, delusional depressions. When carbamazepine was discontinued and valproate was added to lithium, the patient soon started to improve and reached a complete and sustained remission for the first time in 5 years. Her remission, illustrated in the figure from 1990 to 1998, persisted through 2008. ECT, electroconvulsive therapy.

Schizophrenia and Schizoaffective Disorder

There are no controlled studies of valproate in the treatment of schizoaffective disorder. Clinically, the treatment regimen of schizoaffective disorder usually includes a mood stabilizer, such as valproate, carbamazepine, or lithium, and an atypical antipsychotic.

Recent prevalence data have suggested that the use of valproate alone and in combination with atypical antipsychotics has substantially increased, and the use of lithium has decreased, but whether these trends reflect relative effectiveness or some other variable is not known.

Although mood stabilizers are not identified as first-line agents in the treatment of schizophrenia, valproate may be useful as an accelerating agent to antipsychotic therapy. In a 28-day, double-blind, parallel-group, multicenter trial of 249 patients who were hospitalized with an acute psychotic exacerbation of schizophrenia, subjects were randomized to one of four treatments: (1) olanzapine monotherapy (15-mg fixed dose by day 15), (2) risperidone monotherapy (6-mg fixed dose by day 15), (3) olanzapine plus divalproex (15 mg/kg/day on day 1, titrated as clinically indicated to a maximum dose of 30 mg/kg/day by day 12), or (4) risperidone plus divalproex. Statistically significant differences favoring the combination groups were noted as early as day 3 and maintained significance through day 21 in positive but not negative symptoms. Moreover, at the final day-28 visit, there was no significant difference in PANNS total score reduction in the monotherapy groups versus the combination groups. Weight gain was not significantly different between the olanzapine–divalproex group (8.3 lb) versus the olanzapine group (7.7 lb); however, there was significantly more weight gain in the risperidone–divalproex group (7.5 lb) versus the risperidone alone group (4.2 lb).

This study suggests that the time course of treatment response of antipsychotic therapy (specifically, positive symptom severity) can be accelerated with divalproex, but the potential pharmacoeconomic and the utility of the short-term benefits remain uncertain.

Alcohol Withdrawal and Relapse Prevention

Divalproex, carbamazepine, and gabapentin have all been studied in the treatment of alcohol withdrawal. Although the intermediate (diazepam [Valium]) and long-acting benzodiazepines (lorazepam [Ativan] and clonazepam [Klonopin]) are the standard established agents for alcohol detoxification, comparative studies have shown that carbamazepine decreases global distress more than oxazepam and anxiety more than lorazepam.

In an open-label, 5-day study involving patients undergoing a standard protocol for alcohol detoxification, as needed dosing of lorazepam, with or without divalproex (500 mg by mouth three times a day), was given based on a Clinical Institute Withdrawal Assessment for Alcoholism, revised (CIWA-A-R) score. Although the CIWA-A-R score reduction was not significantly different, the cumulative lorazepam dosage was lower in the group that received divalproex. The treatment was well tolerated without significant side effects or hepatocellular enzyme elevation. In a second study of 23 inpatients hospitalized for dysphoric mania complicated by substance use (primarily alcohol), treatment of the comorbidity with divalproex resulted in not only dysphoric symptom reduction (as measured by the YMRS and Hamilton Rating Scale for Depression) but also effective alcohol withdrawal and abstinence over the course of 20-week prospective follow-up. Mild hepatocellular enzyme elevation (not greater than three times normal) at the time of admission, presumably related to alcohol ingestion, decreased over the study period.

Alcohol relapse prevention (i.e., maintaining sobriety after alcohol detoxification) has also been evaluated with the mood-stabilizing anticonvulsants carbamazepine and valproate. In a 12-month, placebo-controlled study in 29 alcohol-dependent patients, carbamazepine treatment was associated with a decrease vs. increase in the time to return to heavy drinking and a decrease in the number of drinking measures at 2- and 4-month follow-up. Recent data have suggested similar effectiveness with divalproex. Twenty-nine outpatients with alcohol dependence participated in a 12-week, double-blind, placebo-controlled pilot trial of divalproex (mean divalproex level at week 12 was 88.2 μg/mL). There was a greater reduction in irritability in the divalproex treatment group, and a smaller percentage of divalproex-treated patients relapsed to heavy drinking (37%) compared with placebo-treated patients (63%; $P \le .05$). Given the extensive comorbidity of alcoholism with other axis I disorders, particularly bipolar I and II disorders, treatment with valproate may stabilize mood and help treat components of alcoholism (withdrawal and relapse prevention). These results have now been replicated and extended.

In a 24-week, double-blind, placebo-controlled study of bipolar patients with comorbid active alcoholism maintained on lithium and psychosocial treatment as usual, valproate showed greater effects than placebo on several measures of drinking. Of the 25 patients randomized to placebo, 68% returned to heavy drinking compared with 44% of the 27 randomized to divalproex ($P = .02$). Liver function tests, ALT, and AST did not differ, while GTP (γ-67P) was higher in those on placebo and was correlated with the amount of alcohol consumed per week. Of those who did return to heavy drinking (four or more drinks per day for women; five or more drinks per day for men), time to relapse was also significantly delayed for those on divalproex (93 days) compared with 62 days on placebo ($P = .02$).

Although other agents are effective in primary alcohol dependence (i.e., naltrexone [ReVia] and acamprosate [Campral] [which are FDA approved], and now effectiveness is seen with topiramate in two large randomized controlled trials and also gabapentin), none have been specifically studied in bipolar patients with comorbid alcoholism with the exception of valproate, making it a particularly good agent for such dual diagnosis patients. It is noteworthy that in patients with bipolar illness, women are 7.35 times and men 2.8 times more likely than those in the general population to have problems with alcohol use and abuse. Greater amounts of depression and anxiety were correlated with the women who abused alcohol, suggesting that it was a form of self-medication and that earlier and more effective intervention on these target symptoms potentially could have lessened the incidence of alcohol comorbidity.

Panic Disorder

Several open-label studies suggest that valproate may be effective in treating panic attacks. In one open-label study of lactate-induced panic attacks, valproate effectively blocked the reintroduction of panic attacks on lactate rechallenge in 10 of 12 patients; furthermore, 71% of patients had a greater than 50% reduction in panic attack frequency, with 43% achieving complete remission. These data are consistent with the marked decreases in anxiety when valproate is used in the treatment of bipolar depression. However, in a double-blind, placebo-controlled, crossover study, 35 patients with panic disorder responded equally well to valproate and placebo.

Post-Traumatic Stress Disorder

Valproate has been investigated in several open trials (total $N = 53$) of patients with combat-related PTSD. The first open trial noted a decrease in avoidance and hyperarousal but no change in intrusive symptoms. In the first 8-week study, adjunctive valproate (mean dose of 1,365 mg, mean serum level of 80 μg/mL) was associated with a significant decrease in intrusion and hyperarousal but no change in avoidant symptoms. In the second 8-week open trial, which included a 2-week washout with a majority of patients on monotherapy, valproate (mean dose of 1,840 mg ± 270 mg, the mean level of 69 μg/mL ± 29 μg/mL) was associated with a significant reduction in all symptom clusters of intrusion, avoidance, and hyperarousal.

Further clinical studies of valproate in the extinction of conditioned fear would appear appropriate in light of preclinical data that valproate specifically increases the promotor of a BDNF variant that is involved in the new learning that is necessary for the extinction of conditioned fear. It is likely that the effects of valproate as a histone deacetylase inhibitor are involved in this effect.

Impulse Control Disorder

Several impulse control disorders, such as intermittent explosive disorder, pathologic gambling, and kleptomania share some clinical features with mania. Valproate has been investigated in several case reports and open trials with generally positive results, particularly if manic or mixed symptoms were present. One hundred and sixteen patients with intermittent explosive disorder were part of a 12-week, randomized, placebo-controlled study evaluating valproate in the treatment of impulsive aggression; in this study, only patients with impulsive aggression associated with borderline personality disorder responded to valproate. A second controlled, single-blind study randomized patients with pathologic gambling, those taking lithium or valproate showed significant improvement.

Highly significant effects of valproate compared with placebo in a crossover design were seen in children and adolescents with an explosive temper and a disruptive behavioral disorder (oppositional defiant disorder or conduct disorder). These children were also severely mood-dysregulated, and some may have met criteria for bipolar not otherwise specified, but they were not bipolar II or bipolar I and did not have a high incidence of a family history of bipolar disorder in first-degree relatives.

These findings are of considerable importance because they suggest the utility of valproate in children with explosive aggression (86% response on valproate vs. 25% on placebo), even in those who do not meet the criteria for bipolar disorder. The fact that these irritable, hostile, and aggressive children show marked improvement over placebo and that in another study of children with more classic bipolar presentations meeting DSM-IV criteria, valproate did not separate from placebo suggests the usefulness of valproate in just this area of greatest diagnostic controversy, that is in those who are the highly impaired by their irritability and dyscontrol but may or may not end up in the bipolar diagnostic spectrum.

Further concerted study in this area and in those with conduct disorder would be of great clinical importance.

Borderline Personality Disorder

Several open trials have suggested valproate's effectiveness in some symptom domains in patients with borderline personality disorder (i.e., mood lability, irritability). In one 10-week, placebo-controlled study, 12 patients treated with valproate (mean dose of 750 mg) had nonsignificant improvement in global symptom severity and functioning compared to patients treated with placebo. A second, 6-month, placebo-controlled study of 30 euthymic bipolar II disorder patients who also met the criteria for borderline personality disorder were randomized to divalproex (mean dose of 850 mg) versus placebo. The primary outcome variable was the mean change in symptom checklist-90 (SCL-90) scales of interpersonal sensitivity, anger and hostility, and depression and the Overt Aggression Scale, modified (OAS-M). In a last-observation-carried-forward analysis, controlling for baseline severity, divalproex was more effective than placebo in reducing interpersonal sensitivity, anger and hostility, and aggression.

Another 12-week, randomized, placebo-controlled study evaluated the effectiveness of divalproex in the treatment of impulsive aggression in patients with borderline personality disorder ($N = 96$), intermittent explosive disorder ($N = 116$), and PTSD ($N = 34$). The primary outcome variable, the mean score reduction in the OAS-M obtained during the last 4 weeks of treatment, did not show a significant treatment difference in those patients treated with divalproex versus those treated with placebo. However, when the data were analyzed by the diagnostic subgroup, there were significant treatment differences in patients with borderline personality disorder ($\chi^2 = 3.95$; $P = .047$) but not in those with an intermittent explosive disorder or PTSD, in measures of aggression, verbal assault, assault against objects, and assault against others but not in an assault against self. There was a significant reduction in irritability, as measured by the OAS-M Irritability score, and severity of illness, as measured by the Clinical Global Impression (CGI) Scale. The mean dose of divalproex was 1,404 mg (with a range between 500 and 2,250 mg), with a corresponding serum level of 65.5 μg/mL (with a range between 0 and 147 μg/mL).

Behavioral Agitation of Dementia

Several open-trial investigations had suggested valproate's effectiveness in reducing behavioral agitation associated with dementia. The dose was generally lower than that used in other conditions, most likely reflecting an older patient population and a pharmacokinetic profile characterized by decreased protein-binding capacity. However, the one double-blind, placebo-controlled, 6-week study of higher doses of divalproex in the treatment of manic symptoms in older adult demented patients was suspended prematurely because of a disproportionate rate of adverse events (primarily sedation and thrombocytopenia) and dropouts (noncompliance) reported with divalproex compared to placebo. At the time of study suspension, 85 patients had been randomized to placebo, and 87 had been randomized to divalproex (median dose of 1,000 mg daily or 18 mg/kg/day, mean weekly serum valproate levels ranging from 55.3 to 68.9 μg/mL). Although there was no significant difference in the primary outcome as measured by the Bech–Rafaelsen Mania Scale, there was a greater reduction in the Cohen–Mansfield Agitation total score (noted as early as week 2 and maintained throughout the 6-week period) and the verbal agitation subscale score at week 6. This study suggests that, although agitation can be reduced with valproate, many older adult patients may not be able to tolerate an aggressive dosing strategy (18 mg/kg/day).

THERAPEUTIC INDICATIONS

Neurologic Conditions

Valproate is currently approved by the FDA for (1) monotherapy or adjunctive therapy of complex partial seizures that occur in isolation or in conjunction with other types of seizures, (2) monotherapy and adjunctive therapy of simple and complex absence seizures, and (3) adjunctive therapy for patients with multiple seizures that include absence seizures. Divalproex has an FDA indication in these same epilepsy syndromes but additionally has indications for (4) prophylaxis of migraine. The ER preparation is approved for seizures, acute mania, and migraine.

Psychiatric Conditions

Valproate and carbamazepine ER are currently the only mood-stabilizing anticonvulsants that are FDA approved for the treatment of acute mania associated with bipolar disorder. Based on open studies and case series in treatment-refractory and rapid-cycling patients, valproate has also been widely used in the prophylactic treatment of recurrent mania and depression, although the database for this is meager. Moreover, caution is urged in the use of valproate in women of childbearing age, as emphasized below.

Valproate has been shown to be particularly effective in acute mania with depressive, dysphoric, and mixed symptoms, acute mania associated with a rapid-cycling course of illness specifier (four or more episodes in a 12-month period), and acute dysphoric mania complicated by substance abuse, particularly alcohol abuse and dependence (Table 33.10e–3). Given its current neurologic FDA indications, valproate may be particularly useful in bipolar disorder complicated by seizure disorder, electroencephalogram (EEG) abnormalities, head trauma, and migraine. Outside of formal DSM-IV-TR diagnoses, valproate appears to be effective on target symptoms of irritability, aggressiveness, mood lability, and impulsivity.

PRECAUTIONS

Serious Adverse Reactions

Rare serious side effects—as noted in the FDA black box warnings—include hepatic failure, pancreatitis, and hyperammonemic encephalopathy associated with patients with urea cycle disorders. However, baseline or serial laboratory monitoring does not necessarily predict severe hepatotoxicity or pancreatitis, and close clinical observation appears superior to routine clinical monitoring in this regard. Additional warnings include sedation in older adults and thrombocytopenia.

Valproate is rated pregnancy category rating D, as noted by increased risk (1% to 4%) of neural tube defects, such as spina bifida (Table 33.10e–6). Recent studies have also highlighted the notable occurrence of other teratologic effects. The rate of any adverse outcome (a major congenital malformation or fetal death) was higher with valproate (20.3%) than with phenytoin (Dilantin) (10.7%), carbamazepine (8.2%), or lamotrigine (1.0%), and, the adverse events with valproate were dose-related.

Table 33.10e–6.
Black Box Warnings and Other Warnings

More Serious Side Effect	Management Considerations
Hepatotoxicity	Rare, idiosyncratic event Estimated risk 1:118,000 (adults) Greatest risk profile (polypharmacy, younger than 2 years of age, mental retardation) → 1:800
Pancreatitis	Rare, similar pattern to hepatotoxicity Incidence in clinical trials data is 2 of 2,416 (0.0008%) Postmarketing surveillance shows no increased incidence Relapse occurs with rechallenge Asymptomatic amylase not predictive
Hyperammonemia	Rare—more common in combination with carbamazepine (Tegretol) Associated with coarse (flapping) tremor and may respond to L-carnitine administration
Associated with urea cycle disorders	Discontinue valproate and protein intake Assess underlying urea cycle disorder Divalproex is contraindicated in patients with urea cycle disorders
Teratogenicity	Neural tube defect: 1–4% with valproate; high incidence of other congenital anomalies; developmental delay substantial and IQ decreases prominent, averaging 10 IQ points. Abnormalities may be more frequent in women with epilepsy than those treated for bipolar disorder. Preconception education and folate plus vitamin B complex supplementation and relative avoidance of valproate for women of childbearing potential
Somnolence in older	Slower titration than conventional doses

adults	Regular monitoring of fluid and nutritional intake
Thrombocytopenia	Decrease dose if clinically symptomatic (i.e., bruising, bleeding gums) Thrombocytopenia more likely with valproate levels ≥110 μg/mL (women) and ≥135 μg/mL (men)

Adapted from *Physician's Desk Reference.* Medical Economics Company; 2002.

In a cohort study of 913,302 Danish children, prenatal valproate exposure was associated with 4.5-fold increased risk of intellectual disability and a 6-fold increased risk of combined outcome of intellectual disability and delayed childhood milestones compared with children with no prenatal valproate exposure. Other antiepileptic drugs, including carbamazepine, clonazepam, and oxcarbazepine, but not lamotrigine, were also found to be associated with an increased risk of intellectual disability. Of note, however, in the small number of women exposed to valproate who did not have an epilepsy diagnosis, the risks were not increased and were equal to those of unexposed controls.

In addition, behavioral teratologic effects have also been reported. A syndrome of moderate to severe developmental delay has been identified, with a small percentage having severe retardation and an intelligence quotient (IQ) below 70. Thus, it would appear that avoiding pregnancy during valproate treatment or switching to other less problematic agents when planning for pregnancy is of very great importance. However, two studies have suggested that women with epilepsy are particularly vulnerable to these adverse outcomes, and the magnitude of the relationship to women with other diagnoses exposed to valproate remains uncertain.

Although neural tube defects can be monitored for and detected with modern techniques (i.e., fetal ultrasound in weeks 18 to 20; α-fetoprotein; and, if needed, amniotic fluid at 16 weeks), there is currently no way of predicting the behavioral deficits which, upon follow up of children exposed to valproate in utero, appear substantially more severe than with other anticonvulsants. For example, in those exposed to valproate, the IQ measured at 3 years of age averaged 9 points lower than those exposed to lamotrigine. Valproate's use in combination with other agents also appears to be an additional risk factor.

Valproate has also been associated with a many-fold increased risk for autism of childhood or lifetime onset. This increased clinical risk has been validated in animal models of fetal exposure to valproate.

Given the evidence of a highly elevated risk for neural tube and other teratologic defects and for lower IQ, most neurologists attempt to avoid the use of valproate in women with epilepsy who wish to become pregnant. If the use of valproate is unavoidable because of failure of other agents, most would recommend the use of folic acid and vitamins B_6 and B_{12} in the hope of ameliorating the known risks to the fetus, in addition to careful attention to the optimal forms of birth control.

These cautions would also appear appropriate to women of childbearing age who have bipolar disorder. The characteristics of mania itself, such as hypersexuality and poor judgment, make the risk of an unplanned or unwanted pregnancy even more likely. Thus avoidance of the use of valproate in these women would appear indicated, and in the absence of other alternative treatments, special emphasis on the importance of birth control and adjunctive use of folate, B_6 and B_{12}. It is noteworthy, however, that the high risks for birth defects and developmental delay appear to be more prominent in women with epilepsy than in those with bipolar disorder. It has been suggested that the analog of valproate valnoctamide has both anticonvulsant and antimanic efficacy and may have a lesser likelihood of congenital malformations based on studies in laboratory animals. This drug may have lesser potency as a histone deacetylase inhibitor which some have posited as the mechanism of valproate's teratogenic effects.

Valproate has been classified as compatible with breastfeeding, as case reports have consistently reported low concentrations (1% to 10% of the mother's serum level) of valproate in breast milk.

Several preclinical studies have suggested that valproate can stimulate human immunodeficiency virus 1 (HIV-1) and hepatitis C virus (HCV) replication. Although small clinical studies in HIV-1–infected and HCV-infected patients do not confirm these preclinical observations, close clinical monitoring is recommended.

In long-term toxicity studies in rats, valproate was associated with an increase in pulmonary adenomas and fibrosarcomas. There is no evidence to suggest a carcinogenic effect in humans, and on the contrary, there is considerable evidence of an antineoplastic effect in several tumor types, including chronic myelogenous leukemia (CML).

On the side of possible benefit in the future, valproate can reopen critical periods of synaptic plasticity, rendering the possibility of allowing new input in adulthood to reverse some effects of earlier experience.

Minor Adverse Reactions

Common side effects include gastrointestinal (GI) distress, tremor, sedation, benign hepatic transaminase elevation, leukopenia, thrombocytopenia, hair loss, increased appetite, and weight gain (Table 33.10e–7). Despite weight gain, there are no increases in triglycerides or total cholesterol. Generally, most of the common side effects are dose-related and can often be minimized by dose reduction, dose consolidation, or adjunctive pharmacotherapy.

Table 33.10e–7.

Some Valproate Side Effects and Their Potential Management

Side Effect (Frequency)	**Potential Treatments**	**Support**	**Comment**
GI distress (O)	Switch to enteric coated preparation Add H_2 blocker Give with meals or all H.S.	++ + +	Effect may be additive with lithium
Tremor (O)	↓ dose Switch to ER formulation Add Propanolol	+ + ±	Additive with lithium?
Weight gain (O)	Early diet and exercise instructions Augment with topiramate, zonisamide	++ ++ ++	Additive with lithium?
Alopecia (hair loss) (UC)	Prophylaxis with zinc and selenium supplements	+	Straight hair may grow back curly
Thrombocytopenia (low platelets) (UC)	Discontinue valproate	++	Monitor for bleeding
PCO syndrome (UC)	Treatment with oral contraceptives will prevent PCOs Switch to lamotrigine	++ ++	Menstrual irregularities may precede use of VPA and may increase on VPA Testosterone not ↑ in 4 of 5 studies Hirsutism rare
Hepatic enzyme elevation; <3	Monitor direction and rate of	+	Caution patient to advise you if RUQ

times normal (O)	change		pain, fever, malaise, fatigue, cola-colored urine, or jaundice occurs. Risk (1:600) in children <2 years of age for fatal hepatitis.
Hepatitis (VR)	Discontinue VPA	++	
Pancreatitis (VR)	Discontinue VPA, monitor amylase	++	Patient to advise you if severe GI pain, nausea, vomiting occurs
Hypothermia (VR)	Discontinue VPA	+	Especially with topiramate
Metabolic: A. Asymptomatic ↑ ammonia (UC) B. Coarse, “flapping” tremor (UC) C. Encephalopathy (VR) D. ↑ homocysteine (C) E. ↓ carnitine and acylcarnitine (?)	↓ Dose, add l-carnitine ↓ Dose, add l-carnitine Discontinue VPA Discontinue VPA, add l-carnitine, folate Add folate	+ + ++ ++ ++	Is there a history of urea cycle abnormalities in patient or family? EEG background slowing and high voltage theta–delta activity 75% occurs in first week of Rx VPA may lower folate
Teratogenicity: A. Neural tube defects; Spina bifida in 1–2% of exposed (UC) B. Developmental delays (mild to severe C, D)	Avoid pregnancy while on VPA. Use oral contraceptives, other devices. Use folate ≥1 mg/day, vitamin B_6 and B_{12} prophylactically	+++ +++ ++	Avoid VPA in planned pregnancies Especially avoid VPA in combination with other anticonvulsants

in women of
childbearing age
Avoid pregnancy

O, occasional; UC, uncommon; VR, very rare; C, common; D, dose related; PCO, polycystic ovaries; PCOs, polycystic ovarian syndrome; H_2, histamine; ER, extended release; VPA, valproate; RUQ, right upper quadrant; GI, gastrointestinal; EEG, electroencephalogram; ?, ambiguous or uncertain.

Valproate causes dose-related increases in homocysteine, and for this reason, most clinicians would recommend routine cotreatment with folate to reduce this risk.

The relationship between valproate, polycystic ovaries, and polycystic ovarian syndrome (menstrual irregularity and clinical evidence of hyperandrogenism) remains controversial because patients with epilepsy and obesity appear to be at increased risk for the syndrome. Studies in women with bipolar disorder taking valproate or lithium have reported conflicting results. In one study of 22 outpatients, most patients reported menstrual irregularity, but none of them had evidence of significant hirsutism on physical examination or hormonal evidence consistent with polycystic ovarian syndrome. In another study of 32 bipolar outpatients, 7 patients currently on valproate and experiencing menstrual problems had evidence of hyperandrogenism (hirsutism and elevated testosterone).

Rasgon and colleagues again found a high incidence of menstrual irregularities in women with bipolar illness prior to treatment, and these increased further with valproate therapy. Hirsutism and significant elevation of serum testosterone were not observed, however.

In contrast, data from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) group did find statistically significant testosterone increases in valproate and, rarely, hirsutism. Thus, it would appear that there is a clear pretreatment increased risk of polycystic ovarian syndrome (menstrual irregularities, weight gain, increases in testosterone, even without polycystic ovaries themselves) and polycystic ovaries in women with epilepsy, which is further exacerbated by valproate. However, in women with bipolar disorder, polycystic ovarian syndrome at baseline pretreatment is relatively uncommon, and exacerbation by valproate is less apparent and more inconsistent.

Previous studies in epilepsy have been unable to fully assess baseline endocrine profile before valproate treatment, the impact of nonspecific weight gain as a risk factor for the development of polycystic ovarian syndrome, and the extent to which epilepsy and bipolar disorder are independent risk factors unrelated to medication exposure. All aspects of the syndrome are reversible in patients with epilepsy, especially if valproate is discontinued and lamotrigine is substituted. Menstrual history and weight should be followed carefully in women of reproductive potential who begin treatment with valproate. Since polycystic ovarian syndrome can be mitigated by birth control pills, this side effect should be discussed but should not preclude women with bipolar illness from treatment with valproate. More pressing is the need to inform women, particularly those with epilepsy, of the major risks to the fetus should they become pregnant, as discussed above.

DRUG INTERACTIONS

Valproate and Other Psychotropics

Lithium and valproate are often used in combination, and additive side effects of tremor or cognitive dulling can often be treated with dose reduction. Additive problems with weight gain can be more daunting. Recent use of the combination of bupropion ER 150 to 300 mg with naltrexone 50 mg has produced dramatic weight loss in many individuals, but this regimen has not been systematically studied for valproate or its combination with lithium. Additive GI problems and dyspepsia can be treated with adjunctive agents histamine 2 (H_2) antagonists.

More recently, atypical antipsychotics and valproate have been used together, with increased sedation (olanzapine) and weight gain (risperidone) reported as side effects. It is noteworthy, however, that valproate may prevent the increases in cholesterol induced by some atypical antipsychotics. This is thought to be attributable to valproate’s influence on peroxisome proliferator-activated receptors (P-PARs), which increase β-oxidation and are transcription factors in steroid–thyroid hormone–retinoic receptor superfamily.

Valproate has also been used as an adjunctive mood stabilizer and anticonvulsant in patients treated with clozapine. In fact, for patients who have had a seizure related to high-dose clozapine, valproate is the preferred agent over carbamazepine, given its risk of agranulocytosis. The combination is generally well tolerated, with the exception of sedation and weight gain.

There are clinical data to suggest greater efficacy and good tolerability of the lamotrigine–valproate combination for the treatment of complex-partial epilepsy compared with either drug alone. Given valproate's efficacy in acute mania and lamotrigine's efficacy in acute bipolar depression, it is likely that the two will be used together in psychiatric practice. Given valproate's plasma protein–binding capacity and its metabolic inhibition of lamotrigine, free levels of lamotrigine are doubled when used in combination with valproate; to reduce the risk of serious rash, the initial dose, rate of titration, and final dose of lamotrigine should be reduced by approximately 50% of the standard monotherapy recommendations when combined with valproate.

Two case series suggesting efficacy of the valproate–carbamazepine combination for refractory rapid-cycling bipolar disorder have been reported. The pharmacokinetics of this combination warrants clinical anticipation and monitoring, however. Adjunctive carbamazepine, as well as its cytochrome P450 3A4 autoinduction and heteroinduction, generally reduces levels of valproate, which are typically not consequential. However, valproate increases levels of free carbamazepine and its active metabolite by displacement from plasma protein. Valproate also increases levels of carbamazepine-10,11-epoxide by inhibiting its metabolism via epoxide hydroxylase to the inactive diol metabolite. Thus the increases in the carbamazepine-10,11 epoxide may not be evident if only carbamazepine levels alone are measured. Given these pharmacokinetic interactions, side effects of the combination generally are related to increased levels of carbamazepine and its active carbamazepine-10,11 epoxide, potentially yielding diplopia, ataxia, sedation, and confusion in those already near their carbamazepine side-effect threshold. Lowering the carbamazepine dose in anticipation of adding valproate will generally avoid this difficulty (Table 33.10e–8).

Antidepressants are often used in combination with valproate, although the efficacy of this approach has recently been questioned. There are case reports of the tricyclic antidepressant amitriptyline increasing valproate concentrations and, conversely, valproate increasing the levels of tricyclics. Additionally, adjunctive fluoxetine (Prozac) has been reported to increase serum levels of valproate.

Table 33.10e–8.
Pharmacokinetic Interactions of Valproate

Metabolic Interactions		
VPA → ↑ Drug	**Drug → ↑ VPA**	**Drug → ↓ VPA**
Benzodiazepines	ASA	CBZ
CBZ-E	Cimetidine	± Lamotrigine
Ethosuximide	Fluoxetine	Phenobarbital
Lamotrigine[a]	Felbamate	Phenytoin
Phenobarbital	Erythromycin	Rifampin
Phenytoin	Phenothiazines	Folate
TCAs	Risperidone	
Binding interactions	Nimodipine	
	Discontinuation of enzyme inducers	
VPA → ↑ Free Drug	**Drug → ↓ Free VPA**	
CBZ	ASA	
Diazepam	NSAIDs	
Phenytoin		
Tiagabine		
Tolbutamide		
Warfarin		
Zidovudine		

[a] VPA doubles lamotrigine levels and increases risk of serious rash.

VPA, valproate; CBZ-E, carbamazepine-10,11-epoxide; TCAs, tricyclic antidepressants; ASA, aspirin; CBZ, carbamazepine; NSAIDs, nonsteroidal anti-inflammatory drugs.

Valproate and Other Drugs

In contrast to carbamazepine, the primary pharmacokinetic liability of valproate is related to its plasma protein–binding capacity and the inhibition of other drugs metabolized by the liver. For example, valproate increases the free concentration of diazepam by plasma protein–binding site displacement and also has been shown to decrease the serum level of a diazepam metabolite *N*-desmethyldiazepam. Close clinical monitoring of this coadministration is indicated to avoid any potential for diazepam toxicity. Similar observations (i.e., binding displacement or inhibition of metabolism, or both) have been noted with ethosuximide (Zarontin), phenobarbital (Barbital), chlorpromazine, and aspirin.

Because valproate has been shown to inhibit the secondary phase of platelet aggregation, close clinical monitoring is indicated when using the drug with other drugs that affect coagulation, such as aspirin and warfarin (Coumadin).

VALPROATE AND LABORATORY INTERFERENCE

Valproate occasionally interferes with the results of laboratory tests. As a fatty acid, serum levels, if measured, reflect an increase. Valproate metabolites may produce a false-positive result in urine analysis for ketones.

DOSAGE AND ADMINISTRATION

In the outpatient setting, one should obtain a baseline hepatic panel and complete blood cell count before starting a patient on divalproex, although this can be ordered concomitantly with or after starting medications in more urgent treatment indications for hypomania or full-blown mania. An appropriate starting dose is 250 mg three times per day, with increases of 250 to 500 mg every several days to a therapeutic threshold (serum level greater than 50 mg/L). Once a steady-state dosage has been achieved, consolidating to bedtime dosing maximizes treatment adherence. The divalproex-ER preparation also lends itself well to single nighttime dosing. Rapid loading of divalproex (20 mg/kg/day) in acute mania has produced therapeutic serum levels more quickly than standard titration and has generally been well tolerated. However, high-dose IV valproate (20 mg/kg) over 30 minutes was not effective, and aggressive titration of the ER preparation was not more effective than less aggressive regimens. These findings add to the suggestion that actions induced by chronic administration of the drug are necessary for its antimanic and other psychotropic effects and suggest that more modest dosing strategies are likely to be as effective, produce fewer side effects, and be more clinically prudent.

FURTHER READINGS

Albanese MJ, Clodfelter RC Jr, Khantzian EJ. Divalproex sodium in substance abusers with mood disorder. *J Clin Psychiatry*. 2000;61(12):916–921.

Baetz M, Bowen RC. Efficacy of divalproex sodium in patients with panic disorder and mood instability who have not responded to conventional therapy. *Can J Psychiatry*. 1998;43(1):73–77.

BALANCE investigators and collaborators; Geddes JR, Goodwin GM, et al. Lithium plus valproate combination therapy versus monotherapy for relapse prevention in bipolar I disorder (BALANCE): a randomised open-label trial. *Lancet*. 2010;375(9712):385–395.

Banach R, Boskovic R, Einarson T, Koren G. Long-term developmental outcome of children of women with epilepsy, unexposed or exposed prenatally to antiepileptic drugs: a meta-analysis of cohort studies. *Drug Saf*. 2010;33:73–79.

Bersudsky Y, Applebaum J, Gaiduk Y, et al. Valnoctamide as a valproate substitute with low teratogenic potential in mania: a double-blind, controlled, add-on clinical trial. *Bipolar Disord*. 2010;12(4):376–382.

*Bowden C, Göğüş A, Grunze H, Häggström L, Rybakowski J, Vieta E. A 12-week, open, randomized trial comparing sodium valproate to lithium in patients with bipolar I disorder suffering from a manic episode. *Int Clin Psychopharmacol*. 2008;23(5):254–262.

Bowden CL, Mosolov S, Hranov L, et al. Efficacy of valproate versus lithium in mania or mixed mania: a randomized, open 12-week trial. *Int Clin Psychopharmacol*. 2010;25(2):60–67.

Brady KT, Myrick H, Henderson S, Coffey SF. The use of divalproex in alcohol relapse prevention: a pilot study. *Drug Alcohol Depend*. 2002;67(3):323–330.

Bredy TW, Wu H, Crego C, Zellhoefer J, Sun YE, Barad M. Histone modifications around individual BDNF gene promoters in prefrontal cortex are associated with extinction of conditioned fear. *Learn Mem*. 2007;14(4):268–276.

Carr RB, Shrewsbury K. Hyperammonemia due to valproic acid in the psychiatric setting. *Am J Psychiatry*. 2007;164(7):1020–1027.

Chaudron LH, Jefferson JW. Mood stabilizers during breastfeeding: a review. *J Clin Psychiatry*. 2000;61(2):79–90.

Chen PS, Wang CC, Bortner CD, et al. Valproic acid and other histone deacetylase inhibitors induce microglial apoptosis and attenuate lipopolysaccharide-induced dopaminergic neurotoxicity. *Neuroscience*. 2007;149(1):203–212.

Christensen J, Grønborg TK, Sørensen MJ, et al. Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *JAMA*. 2013;309(16):1696–1703.

Cipriani A, Reid K, Young AH, Macritchie K, Geddes J. Valproic acid, valproate and divalproex in the maintenance treatment of bipolar disorder. *Cochrane Database Syst Rev*. 2013;(10):CD003196.

Davis LL, Bartolucci AL, Petty F. Divalproex in the treatment of bipolar depression: a placebo-controlled study. *J Affect Disord*. 2005;85(3):259–266.

Donovan SJ, Stewart JW, Nunes EV, et al. Divalproex treatment for youth with explosive temper and mood lability: a double-blind, placebo-controlled crossover design. (Brief Report). *Am J Psychiatry*. 2000;157(5):818–820.

Daugaard CA, Pedersen L, Sun Y, Dreier JW, Christensen J. Association of prenatal exposure to valproate and other antiepileptic drugs with intellectual disability and delayed childhood milestones. *JAMA Netw Open*. 2020;3(11):e2025570.

Findling RL, Frazier TW, Youngstrom EA, et al. Double-blind, placebo-controlled trial of divalproex monotherapy in the treatment of symptomatic youth at high risk for developing bipolar disorder. *J Clin Psychiatry*. 2007;68(5):781–788.

Geller B, Luby JL, Joshi P, et al. A randomized controlled trial of risperidone, lithium, or divalproex sodium for initial treatment of bipolar I disorder, manic or mixed phase, in children and adolescents. *Arch Gen Psychiatry*. 2012;69(5):515–528.

Ghaemi SN, Gilmer WS, Goldberg JF, et al. Divalproex in the treatment of acute bipolar depression: a preliminary double-blind, randomized, placebo-controlled pilot study. *J Clin Psychiatry*. 2007;68(12):1840–1844.

Gomes FV, Zhu X, Grace A. Correction: The pathophysiological impact of stress on the dopamine system is dependent on the state of the critical period of vulnerability. *Mol Psychiatry*. 2020;25(12):3449.

Langosch JM, Drieling T, Biedermann NC, et al. Efficacy of quetiapine monotherapy in rapid-cycling bipolar disorder in comparison with sodium valproate. *J Clin Psychopharmacol*. 2008;28(5):555–560.

Loscher W. Valproate: a reappraisal of its pharmacodynamic properties and mechanisms of action. *Prog Neurobiol*. 1999;58(1):31–59.

*Manji HK, Moore GJ, Chen G. Clinical and preclinical evidence for neurotrophic effects of mood stabilizers: Implications for the pathophysiology and treatment of manic-depressive illness. *Biol Psychiatry*. 2000;48(8):740–754.

O'Donovon C, Kusumaker V, Graves GR, Bird DC. Menstrual abnormalities and polycystic ovary syndrome in women taking valproate for bipolar mood disorder. *J Clin Psychiatry*. 2002;63(4):322–330.

Post RM, Leverich GS. *Treatment of Bipolar Illness: A Case Book for Clinicians and Patients*. WW Norton; 2008.

Rasgon NL, Altshuler LL, Fairbanks L, et al. Reproductive function and risk for PCOS in women treated for bipolar disorder. *Bipolar Disord*. 2005;7(3):246–259.

Salloum IM, Cornelius JR, Daley DC, Kirisci L, Himmelhoch JM, Thase ME. Efficacy of valproate maintenance in patients with bipolar disorder and alcoholism: a double-blind placebo-controlled study. *Arch Gen Psychiatry*. 2005;62(1):37–45.

Smith LA, Cornelius VR, Azorin JM, et al. Valproate for the treatment of acute bipolar depression: systematic review and meta-analysis. *J Affect Disord*. 2010;122:1–9.

Thomas SV, Ajaykumar B, Sindhu K, Nair MK, George B, Sarma PS. Motor and mental development of infants exposed to antiepileptic drugs in utero. *Epilepsy Behav*. 2008;13(1):229–236.

Tohen M, Baker RW, Altshuler LL, et al. Olanzapine versus divalproex in the treatment of acute mania. *Am J Psychiatry*. 2002;159(6):1011–1017.

Winsberg ME, DeGolia SG, Strong CM, Ketter TA. Divalproex therapy in medication-naive and mood-stabilizer-naive bipolar II depression. *J Affect Disord*. 2001;67:207–212.

33.10f Lithium

Robert H. Belmaker, M.D.

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Yet to be determined
Mechanism of Action	Enzyme interaction

The discovery of the use of lithium in medicine is usually attributed to John Cade in Australia in the late 1940s, whose work based on a theory that increased blood uric acid is the cause of mania. This theory is almost certainly false. However, historical records find that Hammond used lithium for acute mania in the mid-1800s in New York, and in the late 1800's Lange in Denmark used lithium for periodic depression. The message of this historical record for the clinical psychiatrist may be a more general truth about how ideas often have multiple sources and origins and can rarely be ascribed to a single individual. The finding that lithium salts are helpful in mania was developed into a concept that the lithium ion is a specific treatment for manic–depressive illness. This concept was highly heuristic but is not supported by the empirical clinical data available today (see below).

The use of lithium in mania was greatly delayed by the widespread toxicity of lithium chloride as a table salt substitute in congestive heart failure patients. Mogens Schou's classic controlled clinical trials of lithium in bipolar patients showed safety and efficacy in the prevention of both manic and depressive attacks in the disorder using monitoring of blood levels by flame photometry and led to the introduction of lithium treatment widely in Europe and then in the United States, the first limited treatment beginning in 1970. Sam Gershon described lithium as "the first salvo in the psychopharmacologic revolution."

Lithium was not and is not a patentable discovery, and no pharmaceutical company promoted its research, distribution, and marketing. Its development as a psychotropic compound was overshadowed by the introduction of chlorpromazine and the early neuroleptics in the 1950s, which were easier to use, did not require blood tests, and had a wider therapeutic index. However, the use of lithium salts in psychiatry developed a fascinating mystique: A simple ion, which is not metabolized anywhere in the body and is excreted the same way it is ingested, a compound that is mined as an ore in the earth rather than being synthesized in a factory, a compound that seemed so closely related to the body's natural physiologic mechanisms of electrical neurotransmission using sodium and potassium—these facts all created a scientific excitement around the hypothesis that lithium might be the key to understanding the basic biochemistry and neurophysiology of bipolar disorder (then known as manic–depressive illness). Moreover, the introduction of chlorpromazine and its congeners for the treatment of schizophrenia and imipramine and its congeners for the treatment of depression supported the concept that lithium is "specific" for bipolar disorder. This made the scientific opportunities for studying it even more tantalizing, and it also fits into the emerging desire for psychiatry in the second half of the 20th century to see itself as a medical discipline with specific illnesses and specific treatments. The goal of the current chapter is to provide necessary information to the clinical psychiatrist to choose and use lithium for the treatment of appropriate patients and not to subscribe to any ideology of treatment.

CHEMISTRY

Lithium is the third lightest element after hydrogen and helium and exists in the periodic table in the same column as sodium and potassium, which are physiologically critical substances for all life but particularly important in the propagation of the nerve signal. Lithium is a metal used in aeronautics, and its ion is extremely important in batteries as well (Fig. 33.10f–1). Most PubMed searches of lithium find far more articles about lithium in airplane wings and batteries than lithium in bipolar disorder. However, these uses are unrelated, and it should be clear that only the lithium ion is medically useful.

PHARMACOKINETICS AND PHARMACODYNAMICS

Lithium is completely absorbed from the gastrointestinal tract within a few hours of oral administration without regard to simultaneous food ingestion. It is absorbed into the bloodstream and is gradually eliminated by the kidney into the urine. It is water-soluble and lipophobic. It does not seem to interact directly with extracellular receptors but needs to cross the cell membrane into the cytoplasm in brain cells to have its effect. Lithium is absorbed equally well whether it is administered as the carbonate, orotate, sulfate, citrate, or chloride because these salts, of course, dissolve in the aqueous medium of the stomach. Preparations that have been commercially claimed to be “slow-release” with the aim of reducing the peak lithium level in the plasma and thereby reducing side effects are of questionable chemical validity as it is difficult to make a simple salt into a reliable slow-release preparation. It is far easier to tell a patient who has mild gastric acidity or nausea a few hours after ingestion to simply divide his lithium tablet into two half-doses and take them a few hours apart. At the other extreme are those who have advocated that patients take their whole daily dose as a single before-bed dose, the average dose being four tablets of 300 mg lithium carbonate each. Some small studies have found such administration to have equal clinical efficacy but fewer side effects. However, such administration leading to a single peak blood level with a long trough throughout most of the day is pharmacologically counterintuitive for a hydrophilic ion that must pass through a lipid cell membrane in order to achieve therapeutic levels in the cytoplasm. Most clinicians recommend giving lithium three times a day to minimize side effects and maximize the achievement of steady intracellular blood levels. Lithium salts should never be given intravenously as this can cause cardiac arrhythmias.

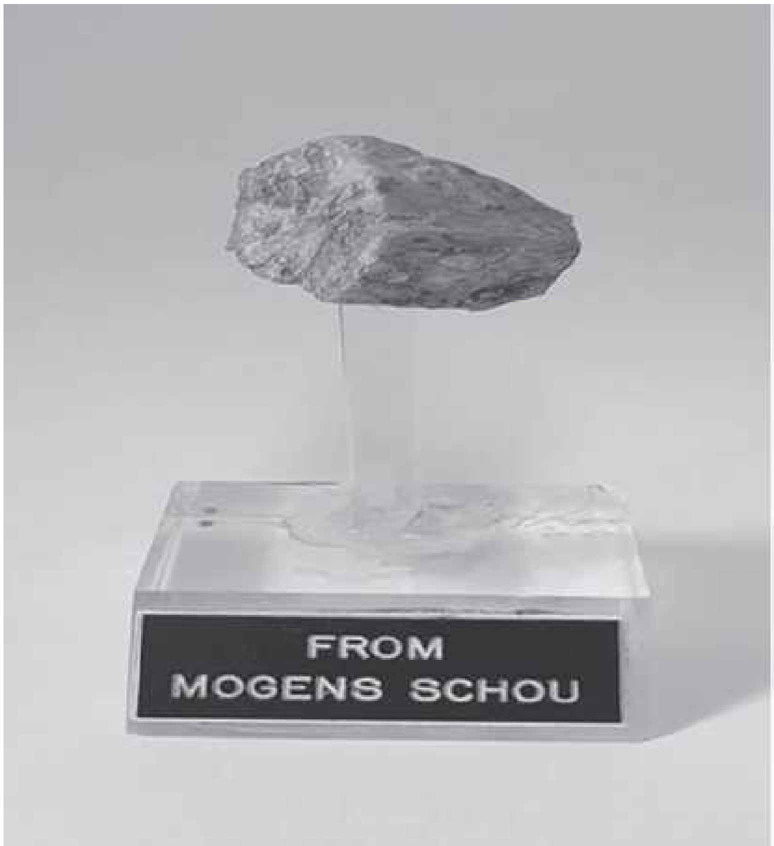

FIGURE 33.10f–1. Lithium ore, a gift from Mogens Schou to the author in 1976.

Lithium enters cells by complex mechanisms, including several distinct sodium channels. Intracellular lithium levels in the red blood cell are about two-thirds of those in plasma and were claimed at one point to be predictive of brain levels and clinical response. This seems not to be the case. Excitable cells (neurons) propagate the nerve signal by a successive opening of the voltage-activated sodium channel and an inrush of sodium ions. Some have claimed that lithium enters nerve cells in the brain by this mechanism, but the issue has not been decided. There may be differences in intracellular lithium blood levels between different areas of the brain, and this is currently being studied in the human brain, measuring lithium with advanced magnetic resonance imaging techniques.

Lithium, when started in acute mania at average doses of 300 mg lithium carbonate three or four times daily, accumulates gradually in plasma and distributes increasingly into the intracellular space reaching an equilibrium state after about 10 days. There is no metabolism of the lithium ion, which is excreted by the kidney with a half-life of about a day if renal filtration is normal. Dehydration, hyponatremia, and other conditions can decrease lithium excretion markedly. Blood levels should be determined in the morning before the first daily lithium dose of the day as these are felt to be most predictive of intracellular levels. Blood levels should be determined after 10 days of treatment at a specific dose. Blood levels taken earlier than this may deceive the clinician into thinking that the level is within the therapeutic window while the patient goes on over the coming days to accumulate lithium and go into toxicity. Thus, the pharmacokinetics of lithium is a very practical topic for every psychiatrist who uses it. Adjustment of dose should rarely be done at intervals of less than 10 days. The exception might be at the very beginning of lithium treatment when the first doses might be administered as one or two tablets to establish tolerance and patient acceptance.

The pharmacodynamic effect of lithium in the treatment of acute mania usually becomes apparent after 10 days, a somewhat shorter interval than the 3 weeks usually reported for the onset of treatment effects with monoamine enhancers or dopamine blockers. It has been speculated that the delay of 10 days in the case of lithium is related to the delay in achieving intracellular pharmacologically active cytoplasmic levels. Some clinicians have tried to give large loading doses of lithium in the first few days in order to speed up the response to treatment and drive the lithium ion into brain cells. This should be done only in carefully controlled conditions where daily blood levels are available and internal medicine consultation is accessible.

The pharmacodynamics of lithium in the prophylaxis of bipolar disorder suggest that a lithium-treated group shows fewer relapses compared to a control group by the end of the first period of measure, usually 2 weeks or 1 month. Upon lithium cessation after long prophylaxis, some rebound may occur where the lithium-treated group has increased relapse compared to the control group. Tapered discontinuation is advised, usually with overlapping initiation of a different mood stabilizer.

Overdose of lithium can be pharmacokinetically confusing. An overdose patient might arrive at the emergency room with a history of ingesting a large number of tablets and have a blood level in the emergency room of 10 mM, known to be highly toxic. He could be completely asymptomatic even at this blood level. However, if this patient is sent home, he will probably arrive the next morning comatose. The explanation is that the lithium in the stomach is absorbed into the blood but takes many hours, even at this high blood level, to get into the cells, where its pharmacologic effects then become apparent. The opposite case is that of the patient who has taken lithium chronically, sometimes for years, at stable blood levels and who neglects his follow-up and continues the same dose while he ages or starts a low sodium diet or increases his lithium dose because of low mood, without adequate blood level monitoring. He may gradually develop lithium toxicity with severe symptoms (see below), including diarrhea and seizures even at a blood level of 2.5 mM since the intracellular level is in full equilibrium with the blood level.

Therapeutic drug monitoring of blood levels is necessary with lithium because of the narrow therapeutic index and also the large variability of oral dose that leads to a given blood level because of differences in renal clearance. Target blood levels in lithium treatment are 0.5 to 1.0 mM. For prophylaxis, levels of 0.6 to 0.8 mM have been reported to be more effective than levels of 0.4 to 0.6 mM. However, levels of 0.4 to 0.6 have fewer side effects than levels of 0.6 to 0.8 mM. Thus, there is no right answer, only an individualized trade-off in clinical care. The rare case of monotherapy in mania can be treated for a few weeks with levels of 1.0 to 1.5 mM and may benefit more than with lower levels. Some have reported that in bipolar depression, where other options are limited, lithium levels of 1.0 to 1.5 mM may also be more antidepressant, but well-controlled data are lacking. Poor oral hydration and sodium depletion by undernutrition can complicate lithium treatment of bipolar depression.

THERAPEUTIC INDICATIONS

Efficacy in Bipolar Disorder

The findings of Schou's classic studies have been generally confirmed. Lithium is efficacious in acute mania, in the prophylaxis of mania and depression, and to a smaller extent in the treatment of acute bipolar depression. Early studies of lithium were carried out before the year 2000 without pharmaceutical company support. They could be and have been criticized by those who accept FDA-based standards for pharmaceutical research in terms of requiring large numbers of subjects, highly quantitative rating scales with predefined endpoints, and hard-core assurances of being double-blind. Such criticisms of the bedrock studies of lithium in bipolar disorder were useful in marketing newer drugs with efficacy in bipolar disorder. The first to be discovered was carbamazepine, originally an antiepileptic with chemical similarities to chlorpromazine and completely unrelated structurally to lithium.

The original discovery of carbamazepine by Okuma in Japan was serendipitous but led to many further studies by Post and others postulating an epilepsy-like etiology of bipolar disorder and a kindling model of the course of bipolar disorder. These theories are still speculative, lithium itself is not antiepileptic at all, and not all antiepileptics are useful in bipolar disorder. The next drug discovered efficacious in bipolar disorder by Emrich in Germany was valproic acid, a 5-carbon branched-chain carboxylic acid whose only similarity to lithium is that it might be found naturally in geologic samples. Pharmaceutical support for valproic acid as a new drug for bipolar disorder was massive due to a special patent for divalproex sodium (which dissolves into valproate in the stomach).

Many of the key studies of valproic acid and carbamazepine in mania and the prophylaxis of bipolar disorder in the last 20 years used lithium as a positive control. These large, modern pharmaceutical company-funded studies provided good data for the use of lithium in bipolar disorder. They did not find a superiority of lithium over valproate or vice versa. Some clinical markers of specific Li responses were identified but are not generally replicable. These drugs, discussed in more detail elsewhere in this section, bear a clinical relationship to lithium roughly as described in Figure 33.10f–2.

That is, these drugs are also efficacious in acute mania and efficacious in the prophylaxis of both mania and depression in bipolar disorder but perhaps less efficacious in the treatment of bipolar depression. Some patients who have been lithium responders but must stop lithium because of side effects can be switched to carbamazepine and vice versa. The same is true of lithium and valproic acid or valproic acid and carbamazepine themselves. That is, it has not been possible to define patient groups who clearly respond only to lithium and not to valproic acid, and it is not possible to describe patients who respond both to valproic acid and to lithium and not to carbamazepine. Lamotrigine is again chemically unrelated to the other lithium-like medications we have described. Studies of lamotrigine using lithium as a positive control in large pharmaceutical company-sponsored multicenter trials again supported lithium’s efficacy in bipolar disorder.

The group of medications useful in bipolar disorder, including lithium, have been termed “mood stabilizers.” It is important to state immediately that this does not mean that they affect the range of normal human mood variability on a day-to-day or hour-to-hour basis. Moreover, all evidence suggests that they do not affect the mood instability of borderline personality disorder, and their effect on mood labile unipolar depressive disorder or dysthymic disorder is limited or nonexistent. The term “mood stabilizer” in psychiatric literature is often misleading to the public and to patient families. The key distinction should be that lithium and the other drugs in this class seem to be effective in preventing the pathologic mood disorders of mania and depression in bipolar disorder in a medical disease model. One can explain to patients that aspirin does not reduce body temperature unless there is a fever (but this does not make it specific to any particular cause of inflammation).

Partly because of the success of lithium in psychopharmacology and its popularity in psychopharmacology research in the 1970s and 1980s, manic–depressive illness has expanded as a clinical diagnosis. On the one hand, psychotic states with mood components have been included in the DSM-IV and 5 as psychotic mood disorders; on the other hand, recurrent depressions with occasional episodes of hypomania or some hypomanic symptoms have been included under bipolar disorders as bipolar II. This great expansion of the conceptual limits of bipolar disorder has also led to an expansion of the epidemiologic incidence and prevalence of bipolar disorder in population studies. However, lithium is not effective in most cases of bipolar II disorder, and as a treatment requiring blood level monitoring and risking short-term and long-term toxicity, its use should be carefully considered when used in cases unlike those in whom its efficacy has been clearly demonstrated.

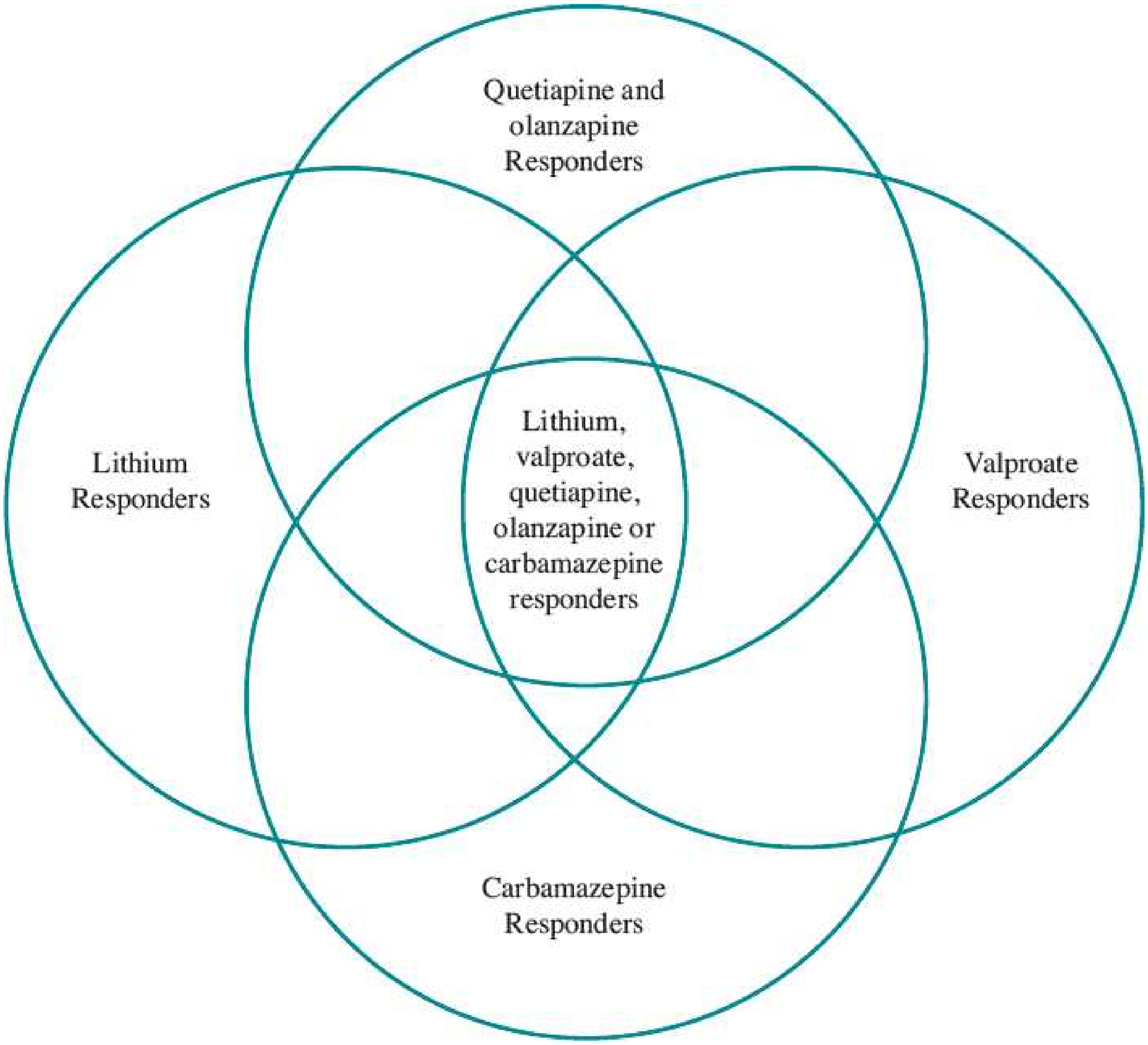
Quetiapine and olanzapine Responders
Lithium Responders
Lithium, valproate, quetiapine, olanzapine or carbamazepine responders
Valproate Responders
Carbamazepine Responders

FIGURE 33.10f–2. The above illustrates roughly the fact that some people respond to lithium and another group responds to valproate, another group responds to carbamazepine, and another group responds to some dopamine blockers; many patients with bipolar disorder respond to any of these four classic mood stabilizer groups, and some respond to only one or two. There is good clinical data to call all of these treatments mood stabilizers, but each is distinct chemically and biochemically.

While controlled studies of lithium find it more effective than placebo in bipolar disorder, these studies represent the average result for large groups of patients. Many patients with clear bipolar disorder do not respond to lithium or respond with a reduction, but not elimination, in frequency and severity of episodes. For those who do not respond to lithium, other treatments are available, and continued physician follow-up and monitoring are essential. Conversely, lithium should be considered as an option in patients who fail to achieve good remission with other mood stabilizers.

Lithium is often called “the gold standard of treatment of bipolar disorders”: Some have called it a “first-line” treatment for bipolar disorder. However, these are misrepresentative statements. Not every newly diagnosed bipolar patient is put on lithium treatment, and no guideline suggests that to be the desirable situation. Rather than the gold standard of therapy, lithium should be called the classic therapy, but it is clearly not for every patient.

Is lithium underutilized? Those who study it and use it regularly often feel this, but they may have an understandable bias not dissimilar to that of a company selling its product. It is not reasonable to estimate 1% of the adult U.S. population as bipolar, accept lithium as the gold standard for bipolar treatment, and conclude that two million Americans should be benefitting from lithium treatment for their expensive and disabling illness. These estimates must be adjusted enormously for those who do not want blood tests, do not respond to lithium, have side effects with lithium, or have better prophylaxis or acceptability with carbamazepine, valproate, olanzapine, quetiapine or other mood stabilizers.

POLYPHARMACY AND ADMINISTRATION OF LITHIUM

The medical workup before beginning lithium therapy includes complete blood count and blood chemistry, including renal function and thyroid-stimulating hormone (TSH) and calcium, and physical and neurologic examination within the last 6 months, including electrocardiogram (EKG) in patients over 60.

Lithium monotherapy is rare today for several reasons. First, in the acute hospitalized manic patient, it is often difficult to get cooperation for blood level testing, and nutrition and hydration may be unstable in uncooperative patients. Lithium is appropriate only when hydration is stable and blood testing can be routinely achieved. Second, most patients receive antipsychotic treatment as their first-line treatment to achieve rapid behavioral control, and lithium is sometimes added as a polypharmacy treatment in order to allow discharge to the community as soon as possible. However, a patient should not be started on lithium treatment in the acute hospital phase if he has not had exposure to lithium in the past and commits realistically to using it prophylactically in the future. The danger of a patient being discharged on a dopamine antagonist plus lithium who then stops his lithium after discharge is great because the sudden discontinuation of lithium can cause a rebound relapse. Third, lithium monotherapy is also rare nowadays in prophylaxis because many more patients with a clinical diagnosis of psychotic bipolar disorder require the continuation of dopamine-serotonin antagonists with their lithium. If side effects of lithium develop, it is easy nowadays to replace it with a dopamine-serotonin antagonist, not requiring blood testing and without danger of fatal toxicity with the medicolegal responsibility that goes along with prescribing lithium without blood monitoring.

Lithium does not seem to have significant pharmacokinetic or drug–drug interactions with other psychoactive compounds and can be used in combination. Because many patients respond only partially to lithium (or other mood stabilizers as well), it is often clinical practice to add another antimanic compound or mood stabilizer rather than stopping lithium when some, even partial, response has been observed. This leads to polypharmacy. In the early post–World War II days of modern clinical psychopharmacology, Lasagna was greatly influential in American medicine in promoting monotherapy as the only way to achieve true knowledge in medical pharmacology but also the only way to avoid impossibly complex interactions in individual patient care. This has proved increasingly difficult to maintain in clinical practice. It is estimated today that 80% of patients taking lithium are taking other major psychoactive drugs as well, usually dopamine blockers but antiepileptic mood stabilizers as well. Initially and especially in academic centers and in textbooks, this practice was considered embarrassing, rarely talked about, and even more rarely studied. The FDA demands monotherapy in psychopharmacologic trials so that lithium must be compared to a placebo (with perhaps a third group such as valproate or quetiapine). This standard has been stricter, perhaps unfairly so, with psychiatry than with other disciplines. In the neurology of epilepsy, for instance, it is common practice for the FDA to allow add-on trials of one antiepileptic agent to another current ongoing treatment in partially responsive epilepsy patients. Robert Post and his group have given considerable intellectual backing to the reality of rational polypharmacy in psychopharmacology in general and in bipolar disorder in particular. Since lithium, the anticonvulsant mood stabilizers, and the dopamine-blocking mood stabilizers seem to have no obvious common denominator biochemically, it makes intuitive sense that their therapeutic modes of action might be at least additive and perhaps synergistic with each other. The goal will be for each patient to have an accurate record of what drug combinations he responded to best over his own history of illness as the best predictor so far of what he might respond to in the future. Charts such as Figures 33.10f–3 and 33.10f–4 may give important information regarding the individual patient to complement the information available from controlled large group studies.

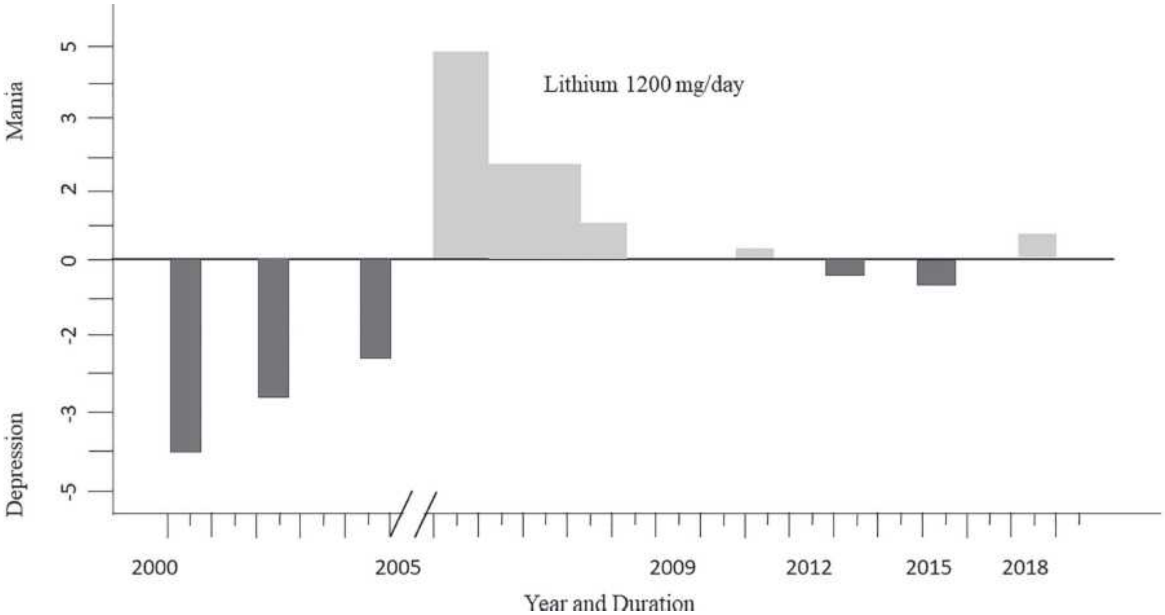

FIGURE 33.10f–3. Lithium responder. The patient is a 43-year-old married woman who had three episodes of depression in her 20s that lasted about 6 months each and resolved without treatment. At age 31, she had a clear manic attack with a few days of hospitalization in a general hospital and was started on lithium 1,200 mg per day. Since then, 12 years have elapsed, and she has had only subclinical cycling of mild depression or hypomania. Positive scores indicate mania and negative scores depression, with higher scores indicating greater severity of illness and zero indicating no illness.

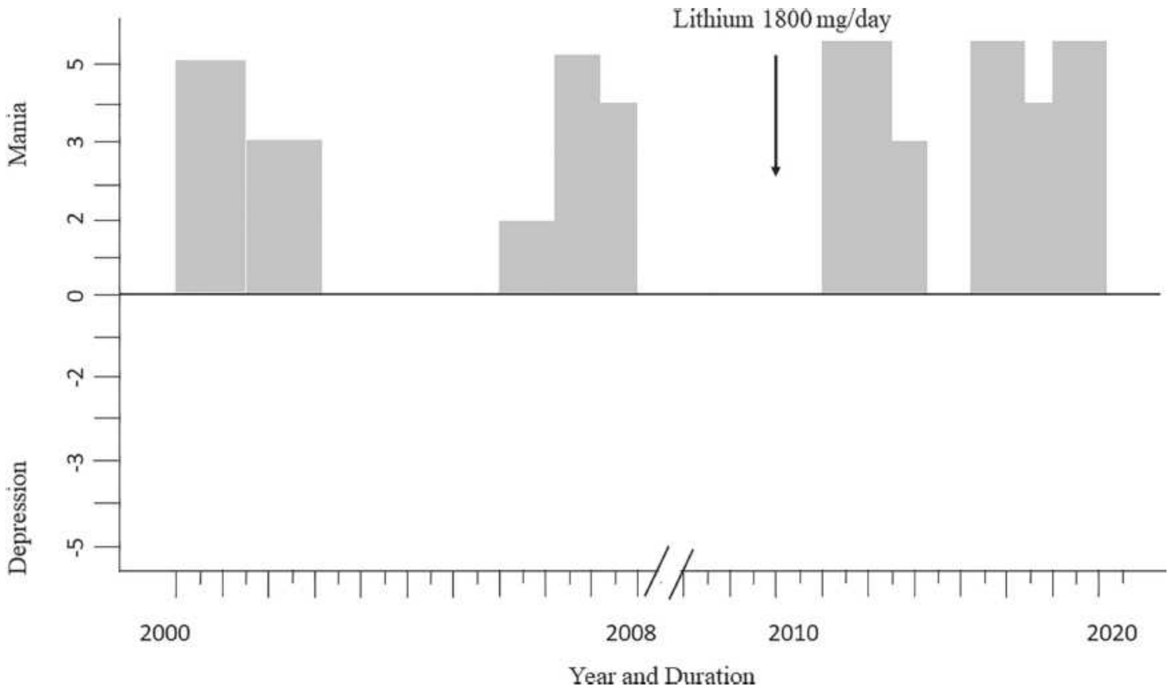

FIGURE 33.10f–4. Lithium nonresponder. The patient is a 56-year-old divorced man who had two long and difficult involuntary hospitalizations with mania separated by about 5 years in his 30s. He was begun on lithium after recovery from the second long hospitalization, and despite genuine motivation and apparent adherence to the treatment regimen, he has had two severe manias in the subsequent decade. Positive scores indicate mania, and negative scores depression, with higher scores indicating greater severity of illness and zero indicating no illness.

Patients treated in the hospital for mania with a dopamine antagonist often begin a mood stabilizer such as lithium toward the end of their hospitalization if they meet the criteria described above. If they have a past history of breakthroughs during lithium prophylaxis, or if they find the neuroleptic treatment to be oppressive to their cognition, motivation, or hedonic state, they will often demand reductions in their neuroleptic dosage as they reach euthymia. This is a good reason to add lithium. However, if, in the past, lithium was not enough for such a patient, it is often best to continue the dopamine-serotonin antagonists in a low dose while continuing on lithium prophylaxis. Patients who have only a partial response in prophylaxis on lithium but breakthroughs not severe enough to require hospitalization on dopamine blockers may often have their lithium prophylaxis augmented with an anticonvulsant mood stabilizer such as carbamazepine or valproate. There are no physiologic contradictions to such polypharmacy. Clinical experience, along with some small controlled studies, suggest that it is often effective in some patients who do not respond to lithium alone. Clozapine addition to lithium is powerfully effective in some very severe rapid cyclers or patients with highly violent manias. ECT can be safely given to a lithium prophylaxis patient who has had a severe breakthrough depression.

THE IMPORTANCE OF PATIENT HISTORY

Most cases appearing for treatment in an acute manic state have had prior treatment, and the results of this prior treatment should be assiduously sought out by the clinician before making his choice. For example, if a computerized record is available, it should be read carefully. If a family member accompanies the patient or is available by phone, he should be queried carefully. The patient should be asked more than once if he has any previous medical records in his wallet, purse, or at home where a family member might be able to find them and read them out on the phone. A patient with a previous history of response to valproic acid should not be started on lithium out of the assumption of its therapeutic equivalency. Similarly, a patient who has done well on lithium for years and is now undergoing a manic attack because he stopped his lithium should not be changed to carbamazepine in the emergency room merely because of its appearance as an equivalent drug in the latest guidelines of the resident’s clinical seminar. Patient history is the best predictor of patient response.

EFFECTS OF LITHIUM IN NORMALS

Lithium has rarely been studied in normal people without psychiatric disease. Evidence suggests that there might be some slowing of thinking and cognitive impairment, and some patients do complain of reductions in creativity. However, patients who experience both lithium monotherapy treatment and neuroleptic monotherapy treatment almost always prefer lithium treatment and feel that drugs that block dopamine receptors impair their hedonic experiences, motivation, and emotional life far more than lithium. In recent years, there has also been increasing awareness that monoamine enhancers, including serotonin specific reuptake inhibitors (SSRIs), may cause over time an “emotional blunting,” although this has been difficult to define in controlled objective tests. Nonetheless, it is important that the treating physician not dismiss all patient complaints of subjective mental side effects of lithium as if they were irrelevant. There may be some psychological price that some or many patients pay for the benefits of reduction of risk of mania and depression.

In animal studies, chronic lithium blocks the increase in activity, aggressiveness, and motivational behavior induced by amphetamine but much less effectively than neuroleptic dopamine-blocking drugs. It has been difficult to define chronic lithium effects on normal rats or mice, especially since the drug cannot be given by injection without causing pain and abscess in the intestines, and chronic administration in food is often associated with a bad taste in the food for the rodents and reduction in weight. Studies of acute single-dose lithium injections of animal behavior are of little relevance to understanding human effects. Returning to humans, several famous studies of lithium in highly creative individuals have not found negative effects of lithium on their creativity.

MECHANISMS OF ACTION INCLUDING ANIMAL MODELS

The mechanism of action of lithium is not known. This is true despite 70 years of intensive research and thousands of reports of biochemical effects in animal brain, human blood, and in vitro cell culture experimental studies. While “antidepressants” could be reasonably claimed to work by inhibiting the reuptake of noradrenalin or serotonin at the synapse, “antipsychotic” drugs could reasonably be claimed to work by inhibiting the dopamine D2 receptor at the synapse, there is no explanation for lithium’s mechanism of action. In over seven decades of psychopharmacologic study, every area that has been developed as a frontier of study has found effects of lithium; this is summarized in Table 33.10f–1. When monoamine metabolites were the frontier of study and the reuptake inhibition effects of “antidepressants” were discovered, lithium was also found to have effects. When the receptor concept was developed and neuroleptics were found to inhibit dopamine receptors, lithium was found to have effects. When the importance in the brain of the transcription of many active genes into mRNA was discovered, lithium was found to have effects. When neurogenesis was found to occur in discrete areas of the human brain and to be important in memory and perhaps mood, lithium was found to have effects. Many of these effects did not replicate in further laboratories, and this is a well-known problem in modern psychopharmacology. However, some of the findings are replicable but may be epiphonema or unimportant. An example might be the effect of lithium on increasing levels of choline in human red blood cells dramatically. This effect does not occur in brain cells in humans or in animals but seems to be a peculiar effect that, while replicable and dramatic, is not important. More recent is the story of lithium’s effect on second messenger systems that carry the synaptic signal that has been transmitted across the synapse into the cells that receive the message. These second messengers are clearly extraordinarily important. One of them is called the phosphatidylinositol cycle. A key enzyme in that cycle, inositol monophosphatase, is inhibited by lithium. If lithium works by reducing brain levels of inositol and choking off the phosphatidyl inositol second messenger system, then inositol should be a kind of “antidote” for lithium effects. However, it was not found to be an antidote for lithium toxicity. Inositol could stereospecifically reverse one behavioral effect of lithium in rodents. This behavioral effect in rodents may be an artifact of the fact that rodents have much lower brain inositol than humans, and a new inositol monophosphatase inhibitor, ebselen, that was hoped to be a new lithium congener was found in 2020 to have no effect in this behavioral system.

Table 33.10f–1.
History of Psychopharmacology and the Mechanism of Action of Lithium[a]

Period	Focus	Finding about Lithium
1967–1975	Monoamine metabolites	Increased deaminated norepinephrine metabolites
1976–1985	Neurotransmitter receptors	Prevents dopamine-receptor supersensitivity
1986–1995	Second messengers cAMP Phosphatidylinositol cycle GSK-3 beta	Inhibits adenylate cyclase Inhibits inositol monophosphatase Inhibits glycogen synthase kinase
1996–2000	Third messengers Immediate early-response genes Response elements	Affects c-fos, c-jun Increases CREB DNA binding
2000–2010	Neuroprotection and neurogenesis	Increases levels of neuroprotective protein bcl-2
2010–present	Gene transcription patterns	Multiple circuit-based and system-wide effects

[a] cAMP denotes cyclic AMP, and CREB cyclic AMP-responsive element-binding protein.

From Belmaker RH. Bipolar disorder. *N Engl J Med.* 2004;351(5):476–486. doi: 10.1056/NEJMra035354

Glycogen synthase kinase is a key enzyme in another second messenger system, and it is clearly lithium inhibitable in the test tube. The details of lithium inhibition are difficult to define, and of course, the devil is in the details: If the inhibition occurs only at 2 millimolar in the cell, then this could be a mechanism of lithium’s toxicity. If the inhibition occurs at 1 millimolar, it could be a mechanism of lithium’s therapeutic benefit. The lithium inhibitability depends on the magnesium concentration, and if the experimenter uses a magnesium concentration equivalent to that in the blood plasma and that is used routinely as the medium in biochemical experiments, results are different than if he chooses the magnesium concentration that is prevalent inside most cells including the brain. While glycogen synthase kinase 3 beta (GSK-3 beta) heterozygous knockout mice show behavioral traits reminiscent of lithium treatment, according to some reports, synthetic GSK-3 beta inhibitors have major effects on the body outside the brain, including on the pancreas, and none have been developed at this writing that would be brain penetrable, brain selective and available for testing as possible lithium substitutes.

A modern line of lithium research uses the hypothesis-free huge data sets that have become available using gene chips. An example would be Akkouh and colleagues, who studied the effect of lithium on the increases and decreases of mRNA levels simultaneously on thousands of different genes in brain tissues. Some levels went up, and some levels went down, and perhaps some sense could be made by looking at which system lithium seemed to affect. This would be a bit like saying that lithium is affecting the performance of a great orchestra, and you can see the lithium effect because some of the wind instruments get faster, some of the percussion instruments get slower, and some of the violins get higher in tone. However, this kind of data really begs the question of what the lithium is doing to the conductor himself. In order to develop a new lithium-like drug, we would need to find the proximal mechanism of the simultaneous effects on hundreds of gene expression changes. Stern and colleagues also present a paradigm shift in research on the mechanism of lithium action. They used cultures of human stem cells differentiated in vitro into neurons after originating in skin biopsies of bipolar patients or controls and also immortalized lymphocytes from bipolar patients and controls. Thus, they could use human tissue without invasive practices and address the question of whether lithium has different effects in bipolar lithium-responsive patients than in those free of this illness or lithium nonresponsive. Complex effects were found in cell membrane excitability, but it is difficult to know yet which are relevant or specific.

Some researchers have accepted the idea that lithium, valproate, and carbamazepine are all mood stabilizers, and it might be possible to look for which biochemical effects, among the numerous biochemical effects of these three compounds, are shared between them. While Manji and colleagues have done considerable work, there is clearly no common dominator that yet explains the biochemical basis of calling these three compounds mood stabilizers. The situation has become even more complicated as data has accumulated suggesting that many dopamine-serotonin antagonists are effective and convenient both in preventing mania and preventing depression and can be reasonably called mood stabilizers. Their common element is dopamine blockade, which is certainly not a primary effect of lithium. The honest answer to the textbook question of how lithium works is that we don’t know.

SUICIDALITY

One of the fascinating aspects of lithium’s clinical usage is its proven correlation with reduction both in serious suicide attempts and in completed suicide. Bipolar disorder is associated with high suicide rates, and thus this antisuicidal effect is potentially very important. Some scientists have considered suicidality to be a psychiatric syndrome in its own right and that it may have biochemical pathophysiology and treatment unrelated to the disorder in which it occurs. However, this evidence of lithium’s antisuicidal effects outside of bipolar disorder is not strong enough for lithium to be a recommended treatment in suicidal patients without bipolar disorder, such as recurrently suicidal borderline patients. In planning treatment options in bipolar patients with lithium versus other mood stabilizers, certainly, the evidence for its antisuicidal action can and should be taken into account. A mechanism for lithium’s antisuicide action is not known, although it is practical and commonsensical to note that clozapine is also proven to have antisuicidal effects different from the other neuroleptic drugs. Both lithium and clozapine require frequent blood level monitoring. It is certainly possible that the increased physician attention and discipline required to agree to monthly blood levels and choose patients who are willing to have monthly blood levels is the real reason why these two compounds seem to be associated with antisuicidal effects.

LITHIUM IN PREGNANCY, CHILDBIRTH, AND LACTATION

Lithium has been considered a classic teratogen and is clearly teratogenic in frog embryos and bird embryo models. However, human registers of patients born to mothers treated with lithium through their pregnancy have revealed rather modest and somewhat specific teratogenicity; Ebstein abnormality of the heart occurs in lithium-treated pregnancies at elevated rates compared to control pregnancies but still occurs in only a very small percentage of total pregnancies of “lithium babies.” If a woman must continue on lithium because her psychiatric history clearly requires it, and if she very strongly wants to become pregnant, a reasonable option would be to follow her throughout conception and delivery with regular ultrasound of the fetus with special attention to the fetal heart. Ebstein anomaly, while a serious cardiac birth abnormality, can be surgically corrected.

While this data on lithium babies is encouraging for those cases of women who must continue lithium and very much want to become pregnant, it is important for the physician to realize that studies of teratogenicity usually examine the newborn at birth and for limited time periods thereafter. The possibility of “behavioral teratogenicity,” meaning possible alterations in important substrates of control of mood in the brain or other systems, could be revealed in lithium babies only as they become older and has not yet been studied conclusively.

The pregnant patient being treated with lithium should have her lithium levels checked frequently and her dose adjusted as glomerular filtration rates change during the three trimesters of pregnancy. A day or two before the expected date of delivery or upon signs of impending delivery, such as breaking of the membranes, lithium should be stopped. This is because the massive intercompartmental shifts of electrolytes before, during, and after delivery can make sudden lithium toxicity possible. Lithium can be renewed a day or two after delivery as oral feeding and drinking resume. Lithium passes into breast milk, so breastfeeding should be avoided.

Alternatives to lithium in patients needing treatment who want to become pregnant should be introduced gradually at least 6 months to a year before the planned pregnancy. Valproate and carbamazepine are both teratogens and are not alternatives for lithium-stabilized bipolars. Olanzapine, quetiapine, or lamotrigine (in the case of patients whose main problem is depression) may be reasonable alternatives.

CHILDHOOD USE

The diagnosis of bipolar disorder in children has expanded exponentially in recent decades. Some have claimed that a large portion of these children, in fact, have no clear relationship to adult bipolar disorder and should be categorized as disruptive mood dysregulation disorder for which lithium has no demonstrable efficacy. Lithium is rarely used in prepubertal children, and controlled studies have been generally unimpressive in pediatric bipolar disorder (or disruptive mood dysregulation disorder). Bipolar children of a lithium-responsive bipolar parent may be an exception to this overall pessimistic evaluation. Lithium is of no use in the hyperactivity of ADHD. Side effects of lithium in children are not dramatically different than in adults, but dosage should be re-evaluated yearly as body mass increases. Target blood levels are the same as in adults.

GERIATRIC USE

Lithium is effective in late-onset mania after the age of 60, even in patients with no family history of bipolar disorder, and even if the mania seems secondary to a small stroke. Generalized cardiovascular disease, hypertension, or renal disease should advise caution. Bipolar patients who reach old age can be continued safely on lithium, preferably with lower target blood levels, more frequent monitoring, and careful warnings about nonsteroidal anti-inflammatory drug (NSAIDs) and diuretics. Such patients should not be routinely switched to another mood stabilizer as this can be a complicated retirement present with unintended negative consequences.

LITHIUM IN CLUSTER HEADACHE AND ALZHEIMER’S

The use of lithium in cluster headache, a rare disorder related to migraine but with definably different features, is proven to be effective prophylactically. The mechanism is speculative and unrelated to those biochemical mechanisms currently studied as the possible basis of lithium use in bipolar disorder. Then there is the inverse case of use of lithium in a disease where there may be good theoretical reasons but no good clinical data: Lithium might be expected to be useful in the prevention or treatment of Alzheimer disease because lithium inhibits GSK-3 beta in test tube biochemical studies. This effect might be expected to reduce the phosphorylation of tau and thereby reduce the deposition of amyloid, which is one popular pathophysiologic theory of Alzheimer disease. Good studies of this possible treatment of Alzheimer disease are scarce, and the majority of those that have been done are negative. The claims that lithium levels in drinking water predict the incidence of Alzheimer's in various regions or countries are highly speculative and suspect since the amount of lithium that could possibly accumulate in the body from such natural lithium in drinking water is far from the levels necessary to inhibit glycogen synthase kinase or to have any of the other biochemical effects of lithium that we know of today. Psychiatry has an obligation to point out to our biochemical colleagues and research colleagues in other fields that lithium can do almost anything biochemically if the concentration is high enough, but most of these effects are unlikely to have clinical implications.

LITHIUM IN AMPHETAMINE, METHYLPHENIDATE, OR COCAINE ABUSE

Chronic lithium in rodent models and in some human volunteer studies partially blocks the response to amphetamine, methylphenidate, or cocaine. However, lithium treatment has not been found useful in human abuse of these stimulant drugs, although adherence issues may be part of the reason. In bipolar patients with comorbid drug or alcohol abuse, improvement of the affective disorder with lithium or other treatments may improve the comorbid substance or alcohol abuse.

INDIVIDUAL DIFFERENCES AND PREDICTION OF LITHIUM RESPONSE IN BIPOLAR DISORDER

Despite numerous efforts of clinicians over the years, no replicable clinical syndrome has been shown to be strongly predictive of lithium response in bipolar disorder. The presence of psychosis in the manic phase or in the depressive phase is not an absolute negative predictor. Positive family history of bipolar disease is not a strong positive predictor, although a positive family history of lithium-responsive bipolar disorder has not been ruled out and should be considered by the clinician as a provisional predictor of lithium response. Positive family history of valproate response or carbamazepine response in other family members does not rule out a lithium response in the individual patient. Considerable numbers of studies have attempted to identify genomic variants by GWAS that might predict lithium response. None of these are currently proven or useful at a clinical level, despite efforts by several companies to market genetic tests for lithium response directly to patient consumers.

Studies comparing lithium with placebo find statistically significant advantages for lithium treatment in mania and bipolar prophylaxis, as said above. However, looking at the standard deviations shows that, as in almost all medicine, including psychiatry, the rule of thumb is that about one-third of the patients respond to a placebo, two-thirds respond to the drug or treatment, and one-third do not respond even with the appropriate drug or treatment. Often this gross rule of thumb conceals an even more complex reality where some individuals improve completely while others improve only partly, and yet others have an improvement in only one symptom rather than their whole disease syndrome. Thus, the prediction of the individual patient response is a major concern of modern medicine.

TOXICITY

Lithium is a drug with a very narrow therapeutic index and a high potential for severe toxicity. As discussed above, therapeutic plasma levels of lithium range from 0.5 mM to 0.8 mM; levels above 1.5 mM in the plasma show toxic effects, often beginning with polyuria, abdominal pain, vomiting, and diarrhea. These effects create a vicious cycle of dehydration and increasing plasma lithium levels because lithium is not excreted in vomitus, diarrheal stool, or in the dilute urine of lithium nephrotoxicity. Therefore, once it begins, lithium toxicity is liable to advance rapidly. As the blood level increases above 2.0 mM, cardiac abnormalities are frequent, and grand mal seizures often occur. Mortality is high. While lithium inhibits numerous magnesium-dependent enzymes, intravenous magnesium sulfate is ineffective in lithium toxicity. Dialysis is often done, but its value has been questioned because the gradient of lithium at 2 mM to 3 mM in the blood is not strong enough to make for efficient dialysis. Most important in cases of toxicity are oral rehydration when possible with liquids containing salt, an intravenous rehydration with isotonic saline if oral rehydration is not possible or effective, along with careful monitoring of renal function.

The prevention of lithium toxicity involves, first and foremost, patient education. Every patient should receive a written information sheet with a list of common lithium side effects that occur at therapeutic doses in contrast with symptoms of early toxicity. The lithium patient information sheet should also contain warnings to maintain normal hydration, avoid diets that limit sodium intake, and inform the treating psychiatrist when any new medication such as a thiazide diuretic for hypertension is introduced. The most common cause of lithium toxicity today is the onset of use of a thiazide antihypertensive medication prescribed by a family physician, unaware that the patient is under psychiatric care with lithium. Many NSAIDs also change lithium excretion and could induce toxicity unless patients are warned, and blood levels are monitored. The public and the medical profession should be aware that lithium is analogous to digitalis: Extremely useful compounds both in the history of medicine and in modern practice but which have by the nature of their effects on basic physiologic systems the capacity to induce severe and even fatal toxicity.

SIDE EFFECTS

Lithium treatment has a variety of side effects, many of which are unfamiliar to the clinical psychiatrist whose practice has included mostly the use of “antidepressants,” neuroleptics, and benzodiazepines. Lithium side effects should be studied carefully by the clinician prescribing lithium and discussed with the patient who is beginning treatment. Prominent among lithium side effects are several side effects outside of the central nervous system on physiologic systems where hormone receptors and second messengers are involved. It was therefore hypothesized for a time that lithium’s effects on second messengers in the brain might be paralleled by similar effects on second messengers in the thyroid, parathyroid, and kidney that could be responsible for the side effects in a manner parallel to the brain effects. However, this has not turned out to be the case, and lithium’s effects in each of the organs that mediate side effects are complex and not fully understood.

Thyroid

Almost all patients that start lithium treatment have a decrease in serum T3 and T4 and an increase in consequent physiologically compensatory TSH. This is transient in most patients, where these blood levels return to equilibrium. In about 10% to 15% of patients, TSH levels remain elevated, but serum T3 and T4 return to normal. Since TSH is the primary screening measurement in primary practice today for a thyroid disorder, many of these patients are flagged by the family physician as having thyroid disorder and perhaps, after various tests, advised to stop their lithium. The psychiatrist should be careful to advise the family physician and the patient in advance that elevation of TSH in the presence of normal serum T3 and T4 is not an indication for discontinuation of lithium. This is an indication that lithium is interfering with the secretion of T3 and T4 from the thyroid gland, but since T3 and T4 are the hormones necessary for physiologic function in all of their target organs, the elevation of TSH alone should be seen as a healthy compensated state. There is a minority view among endocrinologists to the effect that TSH itself may have effects on the body and that its elevation could be problematic. The present state of knowledge suggests that a patient who needs lithium should not be discontinued on the basis of TSH elevation alone.

In perhaps 1% or 2% of patients, the lithium-induced blockade of proper thyroid secretion of T3 and T4 will not be compensated by moderate increases in TSH. In these cases, TSH continues to elevate, and the thyroid gland cells multiply, resulting in a goiter. Psychiatrists should get a clear view of the patient’s neck in every follow-up of a lithium-treated patient and should also ask the patient about changes in neck size. Early detection of lithium-induced goiter should be treated by T3 or T4 oral replacement. Goiter enlargement will stop, but if the size of the thyroid has increased greatly, it will not recede to its previous normal size. The dose of thyroxine is usually about 100 mcg. Blood levels of TSH, T3, and T4 should then be followed until they return to normal. TSH levels alone are not adequate follow-up.

In a smaller but not completely defined percentage of patients, a lithium blockade of normal T3 and T4 secretion by the thyroid is not compensatable by elevated TSH nor by an increase in the number of thyroid cells, and lithium-induced hypothyroidism develops. This can have the classical symptoms of hypothyroidism or appear as a depression. This is not an indication for the discontinuation of lithium treatment. Such patients should have oral thyroid supplementation while continuing lithium therapy. Thyroid supplementation to achieve normal serum T3 and T4 levels along with normal TSH levels should not be confused with high-dose thyroid hormone augmentation treatment for resistant unipolar depression. Thyroid hormone supplementation to achieve normal thyroid function in lithium-treated patients should also not be confused with thyroid disease, including Hashimoto thyroiditis, thyroid cancer, and idiopathic hypothyroidism, all of which exist in elevated incidence in bipolar patients and their families, unrelated to lithium treatment (see relevant sections on bipolar disorder in this textbook). Thyroid hormone abnormalities discovered in bipolar patients unrelated in time to the onset of lithium treatment should be evaluated by an endocrinologist, including an appropriate radiologic scan of the thyroid gland.

Polyuria and Polydipsia

The most common kidney side of lithium is reversible nephrogenic diabetes insipidus, a syndrome caused by blockade of the effect of antidiuretic hormone (ADH) in the kidney. As with TSH in the thyroid, there is no effect of lithium at therapeutic concentrations on the ADH receptor itself. The blockade seems to occur distal to the receptor but is much more complex than early formulations of effects on the second messenger system. The blockade of the ADH, as might be expected, leads to a diuresis of dilute urine. Almost all patients starting lithium treatment will have polydipsia (thirst leading to increased drinking) and polyuria (an increased necessity for urination). Patients should be warned not to satisfy their thirst with sweet drinks, or they will increase weight in a manner not directly caused by the lithium treatment itself, although lithium is commonly blamed. Patients should also be warned in advance of the increased urination, especially if they have benign prostatic hyperplasia (BPH) or other reasons why they have a frequent need for urination and perhaps suffer from nocturia. The polyuria–polydipsia from lithium is not responsive clinically or experimentally to ADH, either by injection or in the sublingual or intranasal spray forms (DDAVP) used to treat primary enuresis in children or adults before bed or used to treat pituitary derived forms of diabetes insipidus. This is because, as said above, the block induced by lithium occurs distal to the receptor. The polyuria or polydipsia induced by lithium often disappears or becomes imperceptible or a nuisance side effect over a period of months but can return, sometimes severely, at any time in the course of years of prophylactic lithium treatment. The reasons for these reoccurrences and fluctuations are usually not clear. Sometimes reducing the dose of lithium while monitoring blood levels is helpful. The patient should not receive the impression that polyuria–polydipsia is a sign of a serious kidney disorder. If he is concerned about kidney damage resulting from lithium that he has read about on the internet, the diabetes insipidus should be clearly distinguished from the rare but far more serious nephrotoxicity of lithium.

Paradoxically, thiazide diuretics or the potassium-sparing diuretic amiloride are quite helpful in the relief of lithium-induced polyuria–polydipsia. The nephrologic mechanisms of this effect are partially understood but far too complex for this chapter. Every psychiatrist, though, should know that the thiazide treatment of lithium-induced polyuria-polydipsia can raise lithium levels which must be followed closely. The psychiatrist should do the thiazide treatment himself as with the thyroxine treatment of lithium-induced hypothyroidism as discussed above.

Chronic Renal Failure

In very rare cases, lithium can cause progressive anatomical damage to the kidney leading to chronic renal failure (CRF) and a high risk for the necessity of renal dialysis and even renal transplant. This is often traumatic for the patient and for the family but also difficult for the psychiatrist emotionally and even practically in terms of his medical malpractice insurance payments. It is difficult to determine the frequency of this catastrophic progression of events, although there is no question today that it does happen. CRF occurs with increasing age, above age 60 in 1% to 2% of the general population. An average of the epidemiologic evidence suggests that chronic use of lithium may double this percentage. The risk is associated with the years of lithium treatment, and so clearly, physicians who carefully follow up on their patients over years of successful prophylaxis of manic–depressive illness as is clinically indicated, will, unfortunately, as the decades pass, see more of these cases of CRF leading to dialysis and renal transplant. There is some association between this rare syndrome with increases in creatinine over years of treatment, and this has been called “creeping creatinine syndrome.” Increases in creatinine over a level of 1.4 mg/dL within a space of 5 years should be a reason to consider gradual crossover to alternative mood stabilizers. Kidney biopsy should be done in patients who have rapid increases in creatinine but whose psychiatric clinical necessity for lithium is strong. Some evidence today suggests that a lithium-specific pathologic diagnosis can sometimes be made, which would be an indication for the cessation of lithium. Surprisingly, patients with very severe CRF will not benefit nephrologically from lithium cessation (it is too late) and should be continued with lithium treatment if it is psychiatrically necessary, along with concomitant nephrologic follow-up for dietary treatment (protein restriction) and consideration of the necessity for dialysis. Some have claimed that careful lithium level monitoring, avoidance of even short periods of mild toxicity, preference for low doses of lithium, and avoidance of NSAIDs, will make lithium treatment possible without any risk of CRF.

Tremor

Lithium induces in most patients a mild resting tremor. It should be examined regularly by the psychiatrist in the office by observation. This is not a Parkinsonian tremor, and the patients do not show akinesia, mask faces, festinating gait, or cogwheel rigidity unless they are simultaneously taking neuroleptics or have been confusing their own lithium pills with their brother’s neuroleptics. The lithium tremor is more akin to what is described in neurology as a MacDonald Critchley essential tremor. Good clinical practice does not require neurologic evaluation of a tremor from lithium. If it is not severe, it is not a reason for the discontinuation of lithium. Interestingly, it is often treatable with a small dose of propranolol, e.g., 10 mg three times a day. It is not responsive to anticholinergic anti-Parkinson medication.

Parathyroid

Lithium has some effects on the parathyroid gland that are not entirely defined and are uncommon. Blood calcium levels should be checked yearly in patients on lithium. Unfortunately, surgical removal of benign parathyroid adenomas associated occasionally with lithium treatment is a dangerous operation and has resulted in death.

Dental Side Effects

Lithium reduces salivary gland secretion and can lead to an increase in dental cavities. Preventative dental care must be encouraged in lithium-treated patients.

Leukocytosis

Lithium causes a mild but consistent leukocytosis. There is no need for referral to an internist to identify the cause of inflammation. The ability of lithium to cause leukocytosis may be related to its ability to enhance stem cell regeneration. But lithium has not been found useful in clozapine-induced leukopenia or leukopenia induced by cancer chemotherapy. Thus, the hopes of its usefulness in neurodegenerative disorders based on the laboratory finding must soberly await clinical confirmation.

Dermatologic Side Effects

Lithium exacerbates psoriasis, and lithium treatment should be avoided in such patients. Lithium can cause de novo psoriasis, and if this occurs at the beginning of lithium treatment, lithium should be stopped because psoriasis will probably worsen with time. If psoriasis is discovered during ongoing lithium treatment (often the patient discovers it late or ignores it for months) and if the patient needs lithium for psychiatric indications, lithium in such cases can be continued. Often lithium patients live with their psoriasis for their whole lives. Clinical reports suggest that if such patients stop lithium later (i.e., not at the very beginning of treatment), psoriasis usually continues. A dermatologist should make the clinical diagnosis of psoriasis visually or by biopsy, and nonspecific skin rashes from lithium do not carry the above implications but are benign. Psoriasis treatments are not contraindicated with lithium treatment. The mechanism of lithium effects in psoriasis was often ascribed to the second messenger systems in the skin cell, but as with other side effects of lithium, the current data is not clear and multiple, complex mechanisms seem to be involved.

Lithium can exacerbate acne in adolescents or cause acne de novo. The usual treatments for acne are not contraindicated, but acne does not constitute a reason for lithium cessation.

LITHIUM IN SCHIZOAFFECTIVE DISORDER

The definition of schizoaffective disorder has changed frequently from DSM-III to DSM-IV to DSM-5 and between the American DSM and the international classification of diseases (ICD) definitions used over these periods. Research genetic studies have also often used different definitions than those used by clinicians in various countries and psychiatric traditions. One of the exciting developments after the introduction of lithium in the 1970s in the United States was the finding that patients who had been diagnosed as schizophrenic because of psychosis at the peak of their mania often responded to lithium. This led to an expansion of the diagnosis of bipolar disorder to include, first, mood-congruent delusions and hallucinations in both depression and mania and then later, in many diagnostic systems, to include even cases with mood-incongruent delusions and hallucinations in depression or mania. Most studies of bipolar disorder today find that more than one-third deteriorate cognitively, and more than 80% are receiving disability insurance. These facts represent the great expansion of bipolar disorder diagnosis to include patients with features formerly considered to be schizophrenic. In addition, many clinical case reports document responses to lithium in schizophrenic patients, even those with chronic forms of the illness, but who suffered from excited or depressed periods that constituted the most burdensome part of their illness. One study looked at "excited psychosis" in sequential admissions to a catchment area psychiatric hospital in western Jerusalem, Israel. Patients were all administered haloperidol. They randomly received lithium or a placebo as an addition. They were later divided into those that met the criteria for nonpsychotic mania, psychotic mania, schizoaffective disorder, or chronic schizophrenia with exacerbation. Surprisingly, the significant lithium benefit was not restricted to any one group, nor could it be demarcated at any point along this continuum from mania to chronic schizophrenia. Lithium seemed to always add benefit to neuroleptic treatment of these excited psychotic patients although the total treatment response was less the further along the continuum the patient fell between nonpsychotic mania and schizophrenia. In this tradition, lithium enhances response whenever affective symptoms are present in psychosis.

However, another completely contradictory tradition exists in psychiatry. This second tradition has defined the lithium responsive bipolar patient as one who has manias that are euphoric (rather than irritable as in the majority of manic patients today), nonpsychotic, patients without comorbidity and with a nondeteriorating life course, no cognitive impairment and full restitutio ad integrum after each episode. The latter tradition claims that lithium responsiveness can be predicted by these clinical characteristics. Clearly, lithium is valuable in the classic manic patient but should also be considered in those schizoaffective or even schizophrenic patients who have an affective component to their illness and who have not responded adequately to other therapies.

LITHIUM AND AGGRESSIVITY

A considerable number of small studies and case reports indicate that chronic lithium treatment reduces the severity of episodic impulsive aggression. Only a small percentage of persons who have committed aggressive or violent acts are found by history to have had repeated perhaps cyclical or perhaps unprovoked episodes of aggression. Not every patient with antisocial behavior disorder derives benefit from lithium treatment. However, there does seem to be such a small group, although it has not been confirmed in large well designed controlled trials. The effect of lithium in such cases seems to be similar to that described for antiepileptic drugs such as carbamazepine and valproate, which are also useful for bipolar disorder. It is hard to know whether coincidence is involved here or psychopathophysiology. Speculation is of little value and certainly does not constitute proof.

LITHIUM AUGMENTATION OF ACUTE ANTIDEPRESSANT TREATMENT

Lithium can augment the effects of tricyclic and SSRI monoamine enhancers in nonresponders to those medications, even when they are clearly unipolar depressive patients. The initial discovery by Blier and deMontigny found dramatic effects within a few days in the great majority of such patients. More recent and controlled studies suggest more modest effects that can take several weeks to develop. There does not seem to be clinical value in adding lithium to every antidepressant-treated patient from the outset of treatment.

LITHIUM IN UNIPOLAR DEPRESSION MAINTENANCE THERAPY

Lithium is not useful in the monotherapy maintenance treatment of unipolar depressive disorder, which occurs at least once in the lifetime of about 20% of the general population, according to the DSM-based epidemiologic studies of major depressive disorder. However, it is important to note that the definition of major depressive disorder and the concept of depression has expanded tremendously over the last 50 years. Some evidence suggests that a bipolar family history may predict responsiveness to lithium prophylaxis in such recurrent unipolar depressive patients. The inappropriate use of lithium in unipolar depressive disorder may be as much of a problem as its underuse in bipolar disorder.

IS LITHIUM IN ANY SENSE SPECIFIC?

Lithium has strong preventive actions against euphoric mood in mania and against depressive mood in bipolar depression but little if any effect on mood in normal volunteers. In this sense, it is “specific.” However, if specificity implies that lithium helps only in bipolar disorder and in every case of bipolar disorder, then we have seen in the above discussions that this is not the case. Lithium is helpful in many cases of schizoaffective disorder, in augmentation of the response to monoamine enhancers in unipolar disorder, in some cases of episodic aggression unrelated to mood disorders and in cluster headaches. The concept of lithium specificity generated a large number of scientific studies over the last 70 years and, in that sense, was heuristic and positive for scientific development. However, as a useful concept for the psychiatric clinician, the ideology of lithium specificity has passed its time. It should be clear today that there are multiple treatment options for mania, bipolar prophylaxis, and for bipolar depression and that lithium is sometimes useful also in conditions unrelated to bipolar disorder.

SUMMARY

Lithium is a useful treatment for bipolar disorders and several other psychiatric conditions. It has a simple molecular composition, but this in itself makes it no more unique clinically than gold treatment of rheumatoid arthritis or arsenic treatment of syphilis. Lithium responsiveness is not specific to bipolar disorder diagnosed by DSM-5 or any other explicit criterion. Elements of cycling or fluctuation in a psychiatric disorder such as described in bipolar II, borderline disorder, or recurring depression do not make them a priori lithium responsive. Attempts to define a specific “classical” picture of bipolar disorder that is always lithium responsive have failed. Attempts to understand a molecular or biochemical cause of bipolar disorder using the effects of lithium biochemically as a starting point has not yielded conclusive answers to this date. Genome-wide association studies (GWAS) or other genetic methods have not yet yielded a replicable genetic profile predictive of lithium response in a subgroup of manic-depressive patients. Lithium remains a viable treatment option among other treatment options available to the knowledgeable and flexible clinical psychiatrist for appropriate patients used with appropriate care in appropriate settings.

ACKNOWLEDGMENT

The author would like to gratefully acknowledge the help and inspiration of Thomas A. Ban, M.D., who read and commented on every draft and inspired a history-based version of the place of this treatment and its evolution over time.

Further Readings

Akkouh IA, Skrede S, Holmgren A, et al. Exploring lithium's transcriptional mechanisms of action in bipolar disorder: a multi-step study. *Neuropsychopharmacology*. 2020;45:947–955.

Azab AN, Shnaider A, Osher Y, Wang D, Bersudsky Y, Belmaker RH. Lithium nephrotoxicity. *Int J Bipolar Disord*. 2015;3(1):28.

Baldessarini RJ, Tondo L, Vazquez GH. Pharmacological treatment of adult bipolar disorder. *Mol Psychiatry*. 2019;24(2):198–217.

Ban TA. *Lithium in Psychiatry in Historical Perspective*. INHN Publisher; 2021.

Belmaker RH. A saga of lithium research 1975–2020. In: Ban TA, ed. *Lithium in Psychiatry in Historical Perspective*. INHN Publisher; 2021:343–349.

Belmaker RH, Bersudsky Y, Agam G. Individual differences and evidence based psychopharmacology. *BMC Med*. 2012;7015:10–110.

Biederman J, Lerner Y, Belmaker RH. Combination of lithium carbonate and haloperidol in schizoaffective disorder: a controlled study. *Arch Gen Psychiatry*. 1979;36(3):327–333.

Brown WA. *Lithium: A Doctor, a Drug, and a Breakthrough: Liveright Publishing Co., New York*. 2019.

Decker BS, Goldfarb DS, Dargan PI, et al. Extracorporeal treatment for lithium poisoning: systematic review and recommendations from the EXTRIP workgroup. *Clin J AM Soc Nephrol*. 2015;10(5):875–887.

Duffy A, Goodday S, Keown-Stonema C, Grof P. The emergent course of bipolar disorder: observations over two decades from the Canadian high-risk offspring cohort. *Am J Psychiatry*. 2019;176(9):720–729.

Fornaro M, Maritan E, Ferranti R, et al. Lithium exposure during pregnancy and the postpartum period: a systematic review and meta-analysis of safety and efficacy outcomes. *Am J Psychiatry*. 2020;177(1):76–92.

Geddes JR, Burgess S, Hawton K, Jamison K, Goodwin GM. Long-term lithium therapy for bipolar disorder: systematic review and meta-analysis of randomized controlled trials. *Am J Psychiatry*. 2004;161(2):217–222.

Goodwin FK, Jamison KR. *Manic Depressive Illness: Bipolar Disorders and Recurrent Depression*. 2nd ed. Oxford University Press; 2007.

Hou L, Heilbronner U, Degenhardt F, et al. Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. *Lancet*. 2016;387:1085–1093.

Healy D. *Mania: A Short History Of Bipolar Disorder*. Johns Hopkins University Press; 2008.

Jamison KJ. *Touched with Fire: Manic-Depressive Illness and the Artistic Temperament*. Free Press; 1996.

Kishi T, Ikuta T, Matsuda Y, et al. Mood stabilizers and/or antipsychotics for bipolar disorder in the maintenance phase: a systematic review and network meta-analysis of randomized controlled trials. *Mol Psychiatry*. 2021;26(8):4146–4157.

Kishi T, Matsuda Y, Sakuma K, Okuya M, Mishima K, Iwata N. Recurrence rates in stable bipolar disorder patients after drug discontinuation v. drug maintenance: a systematic review and meta-analysis. *Psychol Med*. 2021;51(15):2721–2729.

Licht RW. Lithium: still a major option in the management of bipolar disorder. *CNS Neurosci Ther*. 2012;18:219–226.

Lin Y, Mojtabai R, Goes FS, Zandi PP. Trends in prescriptions of lithium and other medications for patients with bipolar disorder in office-based practices in the United States: 1966–2015. *J Affect Disord.* 2020;276:883–889.

Manji HK, Bowden CL, Belmaker RH. *Bipolar Medications: Mechanisms of Action*. APA Press; 2000.

Miura T, Noma H, Furukawa TA, et al. Comparative efficacy and tolerability of pharmacological treatments in the maintenance treatment of bipolar disorder: a systematic review and network meta-analysis. *Lancet Psychiatry*. 2014;1(5):351–359.

Papiol S, Schulze TG, Alda M. Genetics of lithium response in bipolar disorder. *Pharmacopsychiatry*. 2018;51(5):206–211.

Scherf-Clavel M, Treiber S, Deckert J, Unterecker S, Hommers L. Drug-drug interactions between lithium and cardiovascular as well as anti-inflammatory drugs. *Pharmacopsychiatry*. 2020;53(5):229–234.

Smith KA, Cipriani A. Lithium and suicide in mood disorder: updated meta-review of the scientific literature. *Bipolar Disorders*. 2017;19(7):575–586.

Stern S, Santos R, Marchetto MC, et al. Neurons derived from patients with bipolar disorder divide into intrinsically different sub-populations of neurons, predicting the patients' responsiveness to lithium. *Mol Psychiatry*. 2018;23(6):1453–1465.

Tondo L, Baldessarni RJ. Antisuicidal effects in mood disorders: are they unique to lithium? *Pharmacopsychiatry*. 2018;51(5):177–188.

Yatham LN, Kennedy SH, Parikh SV, et al. Canadian network for mood and anxiety treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for management of patients with bipolar disorder. *Bipolar Disorders*. 2018;20(2):97–170.

Yildiz A, Nikodem M, Vieta E, Correll CU, Baldessarini RJ. A network meta-analysis on comparative efficacy and all-cause discontinuation of antimanic treatments in acute bipolar mania. *Psychol Med*. 2015;45(2):299–317.

Young AH, Hammond JH. Lithium in mood disorder: increasing evidence base, declining use? *Br J Psychiatry*. 2007;191:474–476.

33.11 Acetylcholine Blockers and Enhancers

33.11a Anticholinergics and Amantadine

Samoon Ahmad, M.D.

HISTORY

Anticholinergics inhibit the binding of acetylcholine to acetylcholine receptors located in the peripheral and central nervous system (CNS). They are used to treat numerous conditions, including motion sickness, nausea and vomiting, incontinence, parkinsonian symptoms like involuntary muscle movements, asthma, chronic obstructive pulmonary disease, and depressive disorders. In nature, anticholinergic substances exist as alkaloids and can be found in perennial herbaceous plants such as *Atropa belladonna* (deadly nightshade), *Mandragora officinarum* (mandrake), *Hyoscyamus niger* (henbane), and species of the genus *Datura*. Belladonna, for example, contains several compounds with anticholinergic properties, including atropine, hyoscine (scopolamine), and hyoscyamine.

There are two types of cholinergic receptors, nicotinic and muscarinic. Consequently, anticholinergics fall into two categories: antinicotinic and antimuscarinic. The focus of this chapter will be on antimuscarinic agents. Antimuscarinic agents affect both the central and peripheral nervous systems with varying selectivity among different agents. Antimuscarinic agents that block central cholinergic activity are used to treat symptoms associated with Parkinson disease, while those with nonselective antagonism have various clinical uses, including bronchospasm, antispasmodic, urinary incontinence, and anesthesia. The prototype of these agents is atropine which is the most common antimuscarinic agent and has been used for millennia in cosmetics, for medicinal purposes and homeopathy, and as a recreational drug. In addition to being hallucinogenic, *Atropa belladonna* is one of the most toxic plants found in the Eastern hemisphere.

ATROPINE

Atropine is a tropane alkaloid extracted from *Atropa belladonna* and other plants of the family *Solanaceae*. It is a secondary metabolite of the plants noted above and serves as a drug with a wide variety of effects. Being potentially deadly, it derives its name from *Atropos*, one of the three fates who, according to Greek mythology, chose how a person was to die. It was historically used to dilate the eyes either for ophthalmic examination or for aesthetic reasons, particularly among Italian women (*Bella Donna* is Italian for beautiful lady). Atropine also produces cardiac and gastrointestinal effects. Even small doses can cause a marked increase in heart rate, so it can help reverse reflex vagal cardiac slowing or asystole caused by various agents. Administered in small doses before general anesthesia, atropine diminishes oral and air-passage secretions. This same effect has made it popular as an ingredient in preparations commonly used for symptomatic relief of colds. Blockade of muscarinic receptors in the CNS by atropine has been found to be therapeutically useful in treating Parkinson disease but can lead to restlessness, excitement, and even hallucinations in large enough doses.

Atropine also serves as an antidote for poisoning by cholinesterase inhibitors (particularly certain kinds of mushrooms and organophosphate insecticides) and reverses the effects of nerve agents. Troops who are likely to be attacked with chemical weapons often carry autoinjectors with atropine, which can be quickly injected into the thigh (Fig. 33.11a–1).

Other commonly used anticholinergics include benztropine mesylate (Cogentin), diphenhydramine (Benadryl), trihexyphenidyl (Artane), and procyclidine (Kemadrin).

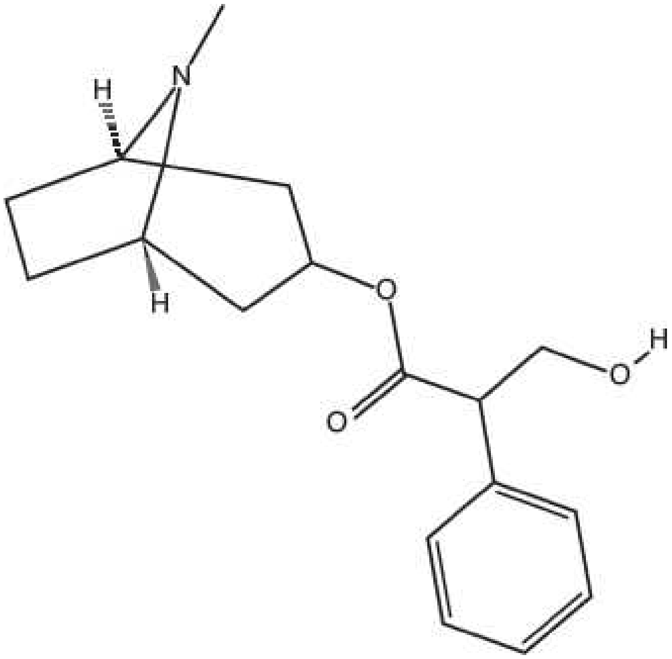

FIGURE 33.11a–1. Molecular structure of atropine.

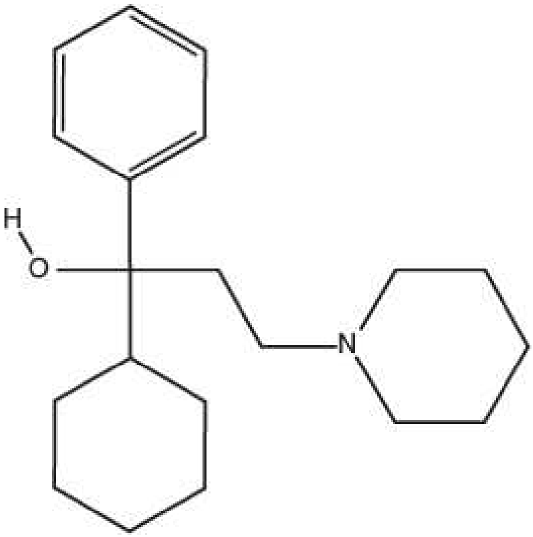

FIGURE 33.11a–2. Molecular structure of trihexyphenidyl hydrochloride.

ANTICHOLINERGIC AGENTS

Chemistry

Artane (trihexyphenidyl hydrochloride) is the substituted piperidine salt, 3-(1-piperidyl)-1-phenyl-cyclohexyl-1-propanol hydrochloride. It exerts a direct inhibitory effect on the parasympathetic nervous system. It is used in the treatment of parkinsonism and extrapyramidal disorders. Artane exhibits a direct spasmolytic action and weak mydriatic, antisialagogue, and cardiovagal blocking effects (Fig. 33.11a–2).

Benztropine mesylate is designated chemically as 8-azabicyclo[3.2.1]octane, 3-(diphenylmethoxy)-*endo*, methanesulfonate. Its empirical formula is $C_{21}H_{25}NO\bullet CH_4O_3S$. It is a crystalline white powder that is very soluble in water and has a molecular weight of 403.54. It possesses both anticholinergic and antihistaminic effects, although only the former has been established as therapeutically significant in the management of parkinsonism (Fig. 33.11a–3) or in the treatment of drug-induced dystonia. Cui and colleagues (2017) reported that benztropine mesylate may inhibit the growth of breast cancer stem cells, while Sogawa et al. (2020) found that it "can suppress tumor progression, circulating tumor cells, and metastasis" in murine models.

Diphenhydramine hydrochloride is an antihistamine that occurs as a white, crystalline powder and is freely soluble in water and alcohol. It has a molecular weight of 291.82 (Fig. 33.11a–4). It is regularly used to treat symptoms associated with the common cold and seasonal allergies, particularly runny nose, sneezing, and itching of the nose or throat.

Procyclidine hydrochloride (Fig. 33.11a–5), known chemically as α-cyclohexyl-α-phenyl-1-pyrrolidinepropanol hydrochloride, is a white, tasteless, crystalline powder that is soluble in water. It is a synthetic antispasmodic compound that has proven effective in the treatment of Parkinson disease and extrapyramidal symptoms. Compared to other antiparkinsonian treatments, some find fewer side effects with this agent. The dose is begun at 2.5 mg per day and increased gradually to a maximum of approximately 30 mg per day.

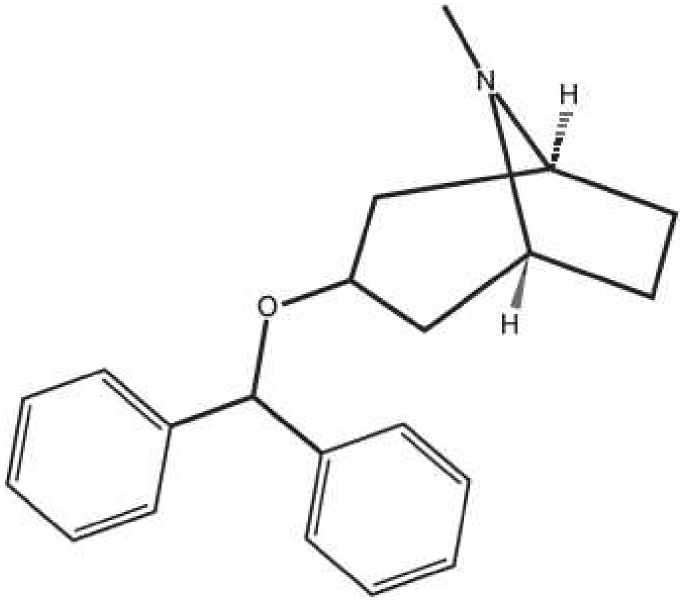

FIGURE 33.11a–3. Molecular structure of benztropine mesylate.

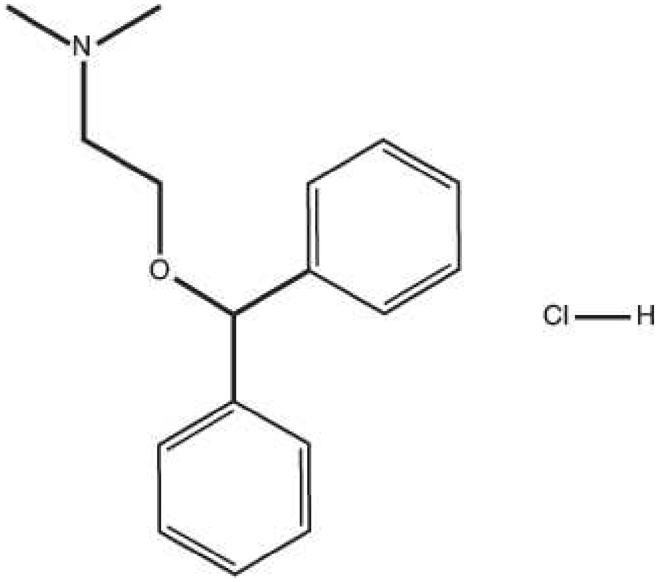

FIGURE 33.11a–4. Molecular structure of diphenhydramine hydrochloride.

Biperiden is an analog of trihexyphenidyl. It is approved for the treatment of parkinsonism and other movement disorders. Biperiden has greater pharmacologic anticholinergic activity than trihexyphenidyl. Recent research by Zhou and colleagues (2017) suggested that it may possess antidepressant effects in murine models.

Pharmacologic Actions

There is limited information available regarding the pharmacokinetics of these drugs, especially among older adults. This may limit the safe and effective use in this population. There are numerous factors contributing to this limited data, including newer and more effective treatments, loss of patent for older drugs, and very low dosages used for treatment.

Absorption and Distribution. All anticholinergic agents readily cross the blood–brain barrier (BBB) and are widely distributed after absorption. They achieve significant levels in the CNS within 30 minutes to 1 hour of administration, and this high concentration in brain tissue may be related to their high lipophilicity. In humans, the bioavailability of these drugs is not reported, and plasma protein binding is also poorly understood. These drugs are extensively metabolized and all unchanged drugs, as well as metabolites, are excreted in the bile and urine.

Trihexyphenidyl is rapidly absorbed from the GI tract after oral administration and is excreted in the urine. The onset of action typically occurs within 1 hour and reaches peak levels within 2 to 3 hours. The full duration of action can last anywhere between 6 and 12 hours.

Benztropine is well absorbed by the GI tract and has an onset of action of between 1 to 2 hours when given orally, but its pharmacologic effects may not be fully apparent for 48 to 72 hours following administration. Benztropine exhibits a more rapid onset of action when administered either by intravenous (IV) or intramuscular (IM) route. There is no difference in onset of action between these two parenteral routes. It is excreted through bile and urine.

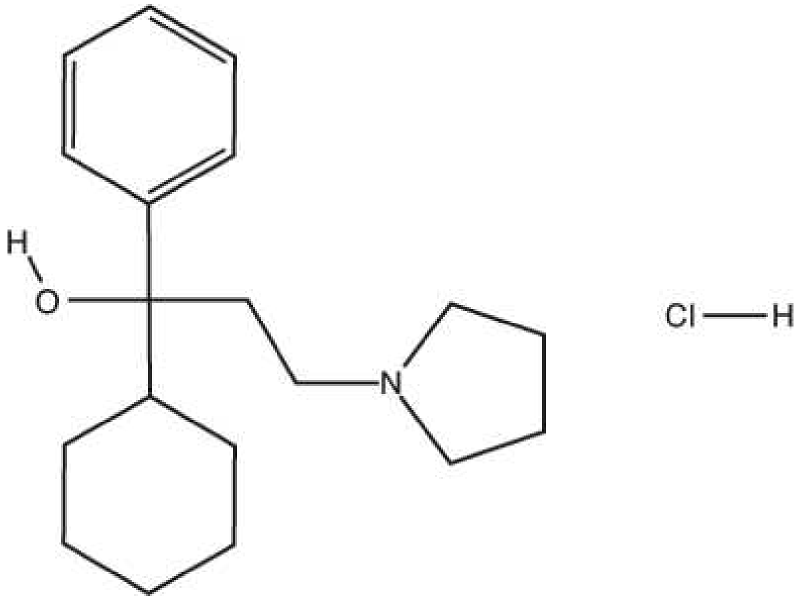

FIGURE 33.11a–5. Molecular structure of procyclidine hydrochloride.

Pharmacodynamics

Anticholinergic drugs act by blocking cholinergic receptors, which respond to the neurotransmitter acetylcholine. Within the nervous system, there are two major classes of cholinergic receptors: nicotinic and muscarinic. The distribution of the two receptors is varied.

Muscarinic receptors are G-protein–coupled receptors located in the CNS, smooth muscle, cardiac muscle, and parasympathetically innervated glands, such as the salivary glands. There are numerous muscarinic receptor subtypes (M1 to M5), and all are present in the brain. These receptor subtypes have selectivity for various brain regions, with the M1 receptor being the most abundant subtype in the CNS and are predominantly located in the neocortex, neostriatum, and hippocampus. The M2 and M3 receptors are also widely distributed throughout the brain, but the M3 receptor has a relatively lower density. The M4 receptors are mostly present in the neostriatum, while M5 receptors are localized in the hippocampus and within the projection neurons of substantia nigra, pars compacta, and ventral tegmental nuclei. Anticholinergic medications, such as benztropine, trihexyphenidyl, procyclidine, and biperiden, act as competitive antagonists at muscarinic receptors. The beneficial effect of agents in Parkinson disease is mediated by CNS muscarinic receptor blockade.

Nicotinic receptors are ligand-gated ion channels located on striated muscle at the neuromuscular junction (motor end plate), the autonomic ganglia, and in the CNS. There are multiple subtypes of nicotinic receptors.

The anticholinergic activity is independent of the serum drug levels of anticholinergic medications because drug levels vary up to 100-fold and hence, are not routinely measured. There is evidence that there is a threshold serum level for anticholinergics, and once this threshold is achieved, it is unlikely for extrapyramidal side effects to occur. This threshold level is a result of muscarinic receptor blockade. Consequently, the decrease in extrapyramidal side effects is not related to the dose of anticholinergic medication nor the dose or serum level of antipsychotic medication, but only to the serum level of anticholinergic activity.

Clinical Studies

Antimuscarinic drugs cause a wide variety of CNS side effects, which are largely dependent on their ability to successfully pass through the BBB, as well as their level of efficacy at receptor binding sites. Cognitive impairment is associated with antimuscarinic drugs, as is memory loss, which may result from antagonism of M1 and, to some extent, M2 or M4 receptors in the CNS. Clinical studies suggest that older patients are more vulnerable to these cognitive side effects, particularly if they are in the early stages of dementia, have a pre-existing cognitive impairment, or are struggling with age-associated memory decline.

Many studies have looked at the impact of these agents on cognitive functioning. A double-blind, crossover trial showed that agents that inhibit cholinergic transmission can impair the storage of new information in new long-term memory but that these agents do not disrupt memory retrieval or hinder tasks involving only immediate memory. Meanwhile, discontinuation of anticholinergic medications was associated with significant improvement in a group of older adult schizophrenia patients whose total cognitive score improved, as did the ideational praxia and orientation subscales.

Anticholinergics have been implicated in worsening psychoses, but there is conflicting evidence on the subject. A few recent small double-blind, placebo-controlled crossover studies have looked at anticholinergic medications and their effects on both positive and negative symptoms of schizophrenia. When patients are stabilized, the agents appear to have no effect, possibly because the presumed dopamine (DA) hyperactivity has been corrected. In schizophrenic patients who are acutely ill, studies consistently find that anticholinergic medications worsen positive symptoms (hallucinations, delusions, disorganized speech, etc.).

Limited data exist regarding the utility of measuring serum anticholinergic levels while assessing schizophrenic patients. One study compared the verbal memory and reaction time of ten schizophrenic patients at two different serum anticholinergic levels. The verbal recall was worse at higher drug levels, while reaction time tended to be improved by anticholinergic treatment.

The effect of anticholinergics on tardive dyskinesia (TD) is also a controversial subject, though studies have revealed that there is cause for concern even if the exact mechanism of effect is unknown. It has been postulated that the condition may be related to an abnormal influence of DA on striatal dopaminergic receptors. In one double-blind, placebo-controlled study, 10 schizophrenic patients with pronounced TD were withdrawn from treatment with anticholinergic medication because of increased symptom severity. Within 2 weeks of cessation, TD decreased significantly in nine patients. This improvement, most pronounced in the oral region, persisted during a 6-week placebo period.

TD is believed to arise because acetylcholine and DA exert opposite effects on the striatum and a delicate balance is necessary to maintain normal functioning. The neuroleptic-induced dopaminergic blockage is considered to be chemical denervation and analogous to pathologic anatomical denervation of idiopathic Parkinson disease. Studies have shown the incidence of TD to be higher in patients receiving both neuroleptic and anticholinergic medications. This hypothesis suggests that anticholinergic medications may predispose patients to the development of TD. Consequently, it is recommended anticholinergic medications should be avoided if possible or kept to a minimum in patients receiving long-term neuroleptic therapy. Discontinuation of anticholinergic medication is a possible therapeutic approach in patients with TD.

Effects on Specific Organs and Systems

Central Nervous System. The centrally acting antimuscarinic agents reduce the tremors associated with Parkinson disease and, when used in combination with a dopaminergic drug (levodopa [Dopar]), can have a more effective response.

In the usual recommended dosages, anticholinergic drugs have a mild stimulant effect and a slower, longer-lasting sedative effect. Initial central vagal stimulation causes bradycardia in the short term, but the antimuscarinic effects eventually stimulate the vagal nucleus in the medullary center and the sinoatrial node, which leads to tachycardia.

Eye. Anticholinergic agents can cause cycloplegia (paralysis of the ciliary muscle), leading to loss of accommodation and the inability of the eye to focus for near vision. These medications also cause dilation of the pupil by blocking the response of the sphincter muscle of the iris.

Anticholinergic medications have little effect on intraocular pressure, except in patients with narrow-angle glaucoma, in whom the pressure may rise to dangerous levels.

Cardiovascular System. In the heart, the atria are innervated by the parasympathetic (vagal) nerve fibers. As noted above, anticholinergic agents cause tachycardia when moderate to high therapeutic doses are used, while bradycardia may result in extremely high doses. In toxic doses, these agents can cause intraventricular conduction block.

Respiratory System. The secretory glands and the smooth muscle of the airways are innervated by muscarinic receptors. These agents inhibit secretions of the nose, mouth, pharynx, and bronchi and may manifest clinically as dry mouth.

Gastrointestinal Tract. The GI tract is innervated by both cholinergic and noncholinergic fibers and is also modulated by local hormones. The net effect of anticholinergic agents is decreased motility via inhibition of parasympathetic control, thereby resulting in constipation or paralytic ileus. These agents also cause dry mouth (as noted above) and delay gastric emptying time.

Genitourinary Tract. Anticholinergics cause relaxation of the smooth muscle in the ureter and the bladder wall and are used in the treatment of spasms caused by inflammatory conditions. Anticholinergics have also been prescribed to treat overactive bladder. Clinicians should also note that anticholinergics can cause urinary hesitancy, particularly in older adult men with prostatic hypertrophy.

Therapeutic Indications

Extrapyramidal Side Effects From Medications. Extrapyramidal side effects (acute dystonias, parkinsonism, and akathisia) are caused by neuroleptic blockade of dopaminergic nigrostriatal tracts, which leads to augmented cholinergic activity. Additionally, decreased serum calcium levels may affect the cholinergic and dopaminergic systems, thereby resulting in extrapyramidal side effects. Anticholinergics reduce cholinergic activity and restore the cholinergic and dopaminergic equilibrium, leading to improvement in these side effects. Acute dystonic reactions usually appear at the onset of most treatments and may occur within 1 to 4 days of neuroleptic therapy, while parkinsonism and akathisia may not manifest until the first 1 to 3 months of treatment.

ACUTE DYSTONIA. Acute dystonia is characterized by a spastic contraction of muscle groups and often affects the neck, eyes, and trunk. These involuntary muscle contractions can be painful and emerge suddenly, thereby frightening the patient. In rare cases, the contractions can be life-threatening (e.g., laryngospasm). The various forms of dystonia include torticollis, retrocollis, trismus or lockjaw, grimacing, dysarthria, blepharospasm, oculogyric crisis, and opisthotonos. Acute dystonia frequently develops after rapid dose escalation of antipsychotic medication or at the beginning of regular treatment, oftentimes within just a few days.

Dystonic reactions are typically treated with benztropine (1 to 2 mg) or diphenhydramine (25 to 50 mg), either IM or IV. The onset of action of these medications, IM or IV, is rapid, with a resolution of the symptoms often within 10 to 15 minutes. As dystonic reactions may recur, prophylaxis with oral medications such as benztropine (1 to 2 mg twice daily) or trihexyphenidyl (2 mg three times a day) is recommended. Tardive dystonia may develop after months or years of treatment with antipsychotic medications and usually does not respond rapidly to anticholinergic medication.

AKATHISIA. Akathisia is characterized by an internal sense of motor restlessness, often described as an inability to resist the urge to move. The most common forms of akathisia involve compulsive fidgeting, rocking back and forth, and pacing, and it most often occurs during the first few months of conventional antipsychotic use. At times, patients may have difficulty clearly describing the sensation of akathisia, which can lead to the misdiagnosis of agitation and an increase in antipsychotic medication—hence worsening the akathisia. Anticholinergics have limited utility in the treatment of akathisia; they should be considered second-line therapy and used when first-line agents, such as adrenergic receptor antagonists (propranolol), either fail or are contraindicated.

PARKINSONISM. Parkinson disease and secondary parkinsonism arising from the use of antipsychotics can be indistinguishable. Secondary parkinsonism is more common with high-potency antipsychotics and is seen much less frequently with atypical antipsychotics. These symptoms may manifest at any time during antipsychotic treatment but usually begin within the first 3 months of treatment and can be managed by lowering the antipsychotic dose to the lowest effective dose. Symptoms include generalized slowing of movement (akinesia), rigidity, resting tremor, hypersalivation, cogwheeling, and hypomimia (masked facies), and are caused by a blockade of DA receptors of nigrostriatal neurons, which terminate in the caudate nucleus. In some cases, these symptoms may also be mistaken for the negative symptoms of schizophrenia.

Anticholinergics have been found to be useful in treating the symptoms of neuroleptic-induced parkinsonism. The additions of benztropine (0.5 to 6.0 mg per day), trihexyphenidyl (2 mg, two to four times daily), procyclidine (5 to 20 mg per day), and biperiden (2 to 24 mg per day) have been found to be effective therapies.

Long-Term Use of Anticholinergics

Patients with schizophrenia often require long-term treatment with conventional antipsychotics. As noted above, these medications have serious adverse effects such as extrapyramidal side effects, including acute dystonic reactions, akathisia, parkinsonism, and TD, and anticholinergics may mitigate their severity. Despite this, the prophylactic use of anticholinergic medication is controversial because the long-term use of anticholinergics may predispose patients to TD. Additionally, anticholinergics can be abused and may have serious side effects. Long-term use may affect memory, and cognitive impairment can be seen as early as 60 days if these medicines are taken continuously. In the case of overdose, anticholinergic agents can create a dangerous condition with tachycardia, hyperthermia, mydriasis, hallucinations, seizure, and coma. Consequently, the World Health Organization (WHO) released a consensus statement that “the prophylactic use of anticholinergics in patients on neuroleptic treatment is not recommended, and may be justified only early in treatment (after which it should be discontinued and its need should be reevaluated).”

Despite the consensus statement from WHO, subsequent studies have found that there is evidence to support the prophylactic use of anticholinergic agents, especially in young men receiving high-potency antipsychotics and who remain at high risk for developing acute dystonia. They also reviewed the literature for continuing prophylactic treatment and concluded that continuing prophylaxis is helpful for many patients because of the relatively high risk of developing parkinsonism upon cessation of anticholinergic medication. The prophylactic use of these agents is not routinely indicated and should be used on a case-to-case basis for patients with a high risk of developing acute dystonic reactions.

Table 33.11a–1.
Adverse Effects of Anticholinergics

- Cardiovascular system
    - Tachycardia
    - Digestive system
    - Constipation
    - Dry mouth
    - Nausea
- Nervous system
    - Confusion
    - Disorientation
    - Memory impairment
- Visual system
    - Blurred vision
    - Dilated pupils
    - Worsening of narrow-angle glaucoma
- Urogenital system
    - Urinary retention

Precautions and Adverse Reactions

Anticholinergic medications can produce adverse reactions in the CNS, as well as in the peripheral nervous system, due to their widespread antagonism of muscarinic receptors in the periphery. These agents produce clinically meaningful effects on the cardiovascular, GI, autonomic, and urogenital systems. The major adverse reactions of these medications are summarized in Table 33.11a–1.

Anticholinergic agents block muscarinic acetylcholine receptors throughout the body and produce system-wide side effects. Dry mouth is a common side effect at low doses, while higher dosages can cause nausea, drowsiness, mild tachycardia, abdominal cramps, urinary retention (especially in men), and severe constipation, which can progress to intestinal pseudo-obstruction and bowel dilation. Dry skin and impaired sweating have also been reported, and patients may risk heat stroke if precautions are not taken. Anticholinergics are contraindicated in patients with closed-angle glaucoma. In the CNS, anticholinergics can cause confusion and memory loss, and in the more advanced and demented patients, hallucinations and disorientation can occur. Lowering or discontinuing the drug can treat the side effects.

Toxic doses or an overdose of anticholinergic medication may lead to an atropine-like poisoning, which may manifest as confusion, nervousness, psychosis, ataxia, blurred vision, tachycardia, elevated blood pressure, dry mouth, GI discomfort, anhidrosis, hyperthermia, and respiratory collapse. In general, supportive measures are used after an overdose. However, 1 to 2 mg of physostigmine salicylate (Antilirium), an acetylcholinesterase inhibitor, may be used to reverse the anticholinergic symptoms of overdose. If required, a second dose of physostigmine may be administered 2 hours after the first dose.

Anticholinergic Abuse

Anticholinergic agents have an abuse potential on account of their euphoriant and hallucinogenic effects, especially when combined with other street drugs. The prevalence of anticholinergic abuse varies from 0% to 34%. The most common reason for abuse is the drug-induced euphoria. Trihexyphenidyl is the most commonly abused anticholinergic for the reason that it tends to be the most stimulating of all anticholinergics.

Table 33.11a–2.
Anticholinergic Preparations and Dosages

Generic Name	Brand Name	Preparations	Daily Oral Dose
Benztropine	Cogentin	0.5, 1.0, 2.0 mg tablets	0.5–2.0 mg b.i.d. or t.i.d.
		1 mg/mL injection	
Trihexyphenidyl	Artane	2, 5 mg tablets	2 mg b.i.d. to q.i.d.
		2 mg/5 mL elixir	
Biperiden	Akineton	2 mg tablets	2 mg t.i.d. to q.i.d.
		5 mg/mL injection	

Drug Interactions

Any medication with anticholinergic or antihistaminic properties may be pharmacodynamically additive to the effects of anticholinergic medications, thereby leading to side effects and potential issues of toxicity. Pharmacokinetic interactions with these agents occur as well. For example, the coprescription of benztropine and digoxin (Digitek) has been reported to increase digoxin levels (Table 33.11a–2).

Any medication with anticholinergic properties may also potentiate the anticholinergic effects of amantadine (see below). Medications that decrease renal clearance may result in higher plasma concentrations of amantadine and, consequently, a greater risk for side effects. These include triamterene hydrochlorothiazide (Maxzide), trimethoprim sulfamethoxazole (Bactrim), and quinidine.

Laboratory Interference

Anticholinergic medications and amantadine are not known to alter the results of any standard laboratory tests and values.

Dosage and Administration

Anticholinergic doses can vary widely among individuals and must be empirically determined. In general, prophylactic doses are the same as the doses usually needed to treat parkinsonism. For acute dystonias, an IM or IV medication such as benztropine (1 to 2 mg IM) or the antihistamine diphenhydramine (25 to 50 mg IM or IV) is usually given. The starting dose for trihexyphenidyl is 0.5 to 1 mg per day, and the dose should be increased no faster than every 7 to 14 days. A dose increase of 1 mg per day is reasonable and a dose of 2 mg three to four times per day is the usual goal.

The dose of benztropine is begun at 0.5 mg per day and can be raised to 4 to 8 mg per day and requires dosing two to three times per day.

The dose of procyclidine is begun at 2.5 mg per day and increased gradually to a maximum of approximately 30 mg per day.

The starting dose of diphenhydramine is 25 mg per day, and it can be titrated to 150 to 200 mg per day in divided doses.

SCOPOLAMINE

Introduction and Chemistry

Scopolamine is a tropane alkaloid drug and a competitive acetylcholine antagonist with a significant affinity for the M3 receptor and has a minimal effect on nicotinic receptors. Scopolamine has a high affinity for the M3 receptor and is 30 times more potent than amitriptyline. Plants containing scopolamine have been in use since prehistoric times, but the compound has only been used in its purified forms (including hydrobromide, hydrochloride, hydroiodide, and sulfate) since 1880. It has a more potent antimuscarinic effect on the eye and the salivary glands than atropine but exhibits less potent effects on the heart, respiratory system, and intestinal smooth muscle (Fig. 33.11a–6). Scopolamine is used for motion sickness and produces drowsiness, and can cause amnesia in sensitive individuals. Toxic doses can cause agitation, hallucinations, excitement, and coma.

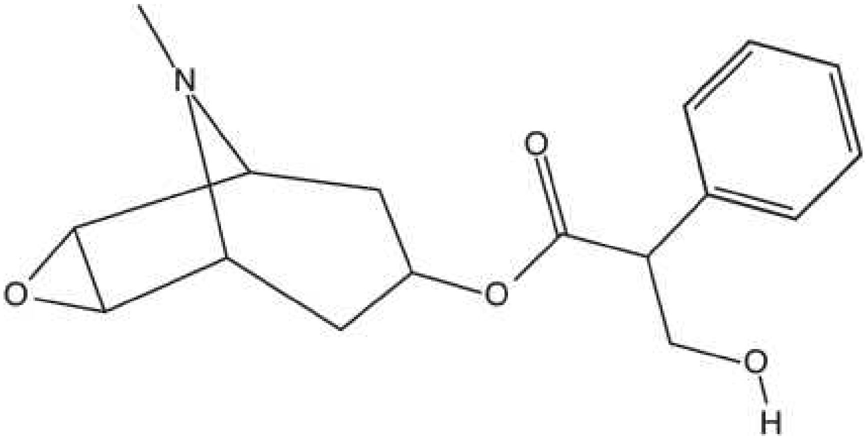

FIGURE 33.11a–6. Molecular structure of scopolamine.

Pharmacologic Actions

Scopolamine is well absorbed from the GI tract, with absorption primarily occurring in the upper small intestine. Its onset of action varies depending on the formulation but ranges from 15 to 30 minutes for an injectable form and up to 4 hours for the transdermal form. The therapeutic effects last up to a maximum of 72 hours with the transdermal patch. Scopolamine crosses the BBB, as well as the placenta, and is secreted in breastmilk. Consequently, it is classified as a pregnancy category C drug. It is reversibly bound to plasma proteins and principally eliminated in the urine. The transdermal formulation has fewer side effects and a longer duration of action compared to the oral form. It is also used preoperatively to inhibit gastric and respiratory tract secretions.

Dosage and Administration

Scopolamine can be administered orally, as a transdermal patch or injected IM, direct IV or sub-Q. The oral adult dosage ranges from 0.4 to 0.8 mg initially and up to three times daily as needed. The injectable dose ranges from 0.32 to 0.65 mg and can be repeated up to three times daily.

Therapeutic Indications

Scopolamine is mostly used for motion sickness and postoperative nausea and vomiting. Recent studies suggest additional uses for scopolamine. Jeraisy and colleagues (2020) report that transdermal patches of scopolamine may help control pediatric drooling (sialorrhea), while ongoing research seems to support scopolamine's efficacy in treating unipolar and bipolar depression, as well as anxiety, particularly in women.

Precautions and Adverse Reactions

Adverse reactions to scopolamine can range from mild drowsiness, somnolence, dizziness, and confusion to more severe effects like disorientation, amnesia, delirium, and acute toxic psychoses. Scopolamine may also cause dryness of the mouth, dilated pupils, increased intraocular pressure requiring adjustment of therapy, increased pulse rate, decreased GI motility, urinary retention, and decreased bronchial secretions with the formation of bronchial plugs.

Design and Interpretation of Clinical Drug Studies

During the last few years, the role of anticholinergic medications has expanded to include the treatment of depressive disorders. The hypothesis that an adrenergic–cholinergic imbalance could lead to depression was proposed as early as 1972, and though initially ignored, most of the recent body of research is based on this theoretical model. Increasing cholinergic activity using physostigmine (an anticholinesterase inhibitor) is known to exacerbate acute depressive symptoms in patients with major depressive disorder (MDD). In addition, increased cholinergic signaling may also induce depressive symptoms in currently manic patients with bipolar disorder (BD). Numerous studies have also demonstrated that the inverse is true: inhibiting cholinergic activity is associated with symptom improvement in major depression. It is also well established that REM sleep changes are associated with MDD, and REM sleep deprivation as induced by anticholinergic medications leads to a reduction in these symptoms.

Scopolamine has generated the most interest as a potential anticholinergic agent that could treat symptoms associated with moderate to severe depression. The first evidence of this finding was suggested in 1991 research conducted on a small group of patients with moderate depression using IM scopolamine and showed a statistical difference. Furey and colleagues (2006) then established the efficacy of scopolamine in depressive disorders while studying the cognitive effects of IV scopolamine and incidentally found antidepressant effects of this drug. To validate these findings, the investigators designed two crossover randomized controlled trials and showed a significantly robust and rapid antidepressant response of IV scopolamine compared to placebo. A later study by the same researchers in 2010 validated these findings specifically in patients with unipolar depression.

Another study conducted at the National Institute of Mental Health enrolled patients with MDD who received multiple sessions of placebo or scopolamine hydrobromide (4 μg/kg) and were randomized to a placebo/scopolamine or scopolamine/placebo sequence of infusions 3 to 5 days apart. Patients were evaluated using the Montgomery–Åsberg Depression Rating Scale and the Hamilton Anxiety Rating Scale to assess their responses to scopolamine. Patients showed reductions in depression and anxiety from baseline during scopolamine infusions compared to placebo infusions. The antidepressant and anxiolytic effects occurred in less than 3 days compared to conventional antidepressants, which may take up to 4 weeks. The group receiving scopolamine infusions first followed by placebo infusions continued to show response, which suggests alternative mechanisms beyond scopolamine’s known effect on neuroreceptors. This persistence of improvement in mood beyond the initial infusion period also negates the hypothesis of anticholinergic euphoria and suggests a sustained antidepressant response. This study showed a rapid antidepressant response to scopolamine in depressed patients that is well tolerated and safe relative to other rapid antidepressant treatments, including ketamine and high-dose TCA therapy. Another study at the National Institute of

Mental Health showed that IV scopolamine had a significant effect on unipolar and bipolar depression.

Recognizing the fact that a significant number of patients with MDDs do not respond to antidepressant medications, the need for augmentation strategies from the beginning is being increasingly entertained. Considering the role of scopolamine as discussed above, it is no surprise that it may have a substantial role in such treatments. One major limitation is that treating MDD with scopolamine requires either an IM or IV route of administration, which presents logistical issues for its use in an everyday practice setting.

To overcome this barrier and assess the tolerability and efficacy of oral scopolamine, investigators designed a randomized, placebo-controlled, double-blind, parallel-group study to evaluate the role of oral scopolamine as an adjuvant therapy to citalopram in moderate to severe MDD. Forty patients were randomized into two groups, the active group receiving citalopram and scopolamine hydrobromide (containing 0.5-mg scopolamine) with the second group receiving citalopram and placebo. The total dose of scopolamine used was 1 mg/day (0.5 mg b.i.d.), while citalopram was adjusted from an initial dose of 20 mg in the first week to 40 mg/day for the next 5 weeks. The patients were assessed using 17-item Hamilton Depression Rating Scale (HDRS) for 6 weeks, with assessments on days 4, 7, 14, 28, and 42. By week 6, all patients in the scopolamine group and most patients in the placebo group had achieved a 50% reduction in the HDRS, while the remission rate by week 6 was significantly higher (65%) in the scopolamine group compared to the placebo group (20%).

The results of this study validated the role of scopolamine as an adjunct to antidepressant therapy when initiated from the beginning of the treatment with a high response and remission rate. It resulted in few adverse effects. These studies suggest that scopolamine as monotherapy, as well as when used as an adjunct, is a safe and effective strategy in MDDs.

More recently, researchers at the Medical University of Lublin in Poland (Kruk-Slomka et al., 2015) studied the antidepressant efficacy of cannabinoids, including oleamide, AM 251, JWH 133, and AM 630, alone and in conjunction with nicotine hydrogen tartrate and scopolamine hydrochloride in murine models. Among other things, the studies found that scopolamine, in conjunction with noneffective concentrations of cannabinoids, did have antidepressant effects.

Various theories abound regarding the antidepressant efficacy and precise mechanism of action of scopolamine in MDDs, and to date, none have been established. Some of these theories suggest alterations in type 2 and 3 muscarinic receptors, as well as polymorphisms in the type 2 muscarinic receptor gene, and that scopolamine has a high affinity for the type 3 muscarinic receptor. Other possible mechanisms include cholingeric hyperactivity, REM sleep deprivation, as well as hyperactivity of the glutamatergic system, which has been implicated in the development of MDD. Scopolamine decreases glutamatergic transmission, thereby alleviating depressive symptoms. Given that activation of the endocannabinoid system is known to inhibit glutamatergic transmission, this could potentially explain observed antidepressant effects following the coadministration of cannabinoid receptor ligands and scopolamine.

Scopolamine seems to be an appropriate treatment option for difficult-to-treat patients with major and bipolar depression. Its relatively simple dosing and minor side effect profile with no significant impact on cognition based on the low doses used in the studies suggest its place as an adjunct to antidepressant treatment. Moreover, it should be noted that combination therapy from the beginning of treatment results in greater improvement than monotherapy. Some early studies reported less robust antidepressant effects, but those results may have been due to the uncontrolled nature of the studies as well as the lower-dose range utilized. Despite the positive implications of the studies, numerous limitations remain regarding its impact on cognition (confusion and delirium), dosing, and safety, and more research is needed to study its efficacy in the long term and rapid antidepressant response as an adjunct when added later in the treatment.

Finally, scopolamine’s rapid antidepressant effect in some patients has made it a candidate for the study of biomarkers. A study in 2012 showed that neural responses in the visual cortex, especially to emotional stimuli, maybe a potential biomarker for patients with a favorable response to scopolamine. The study conducted at the National Institute of Mental Health suggests differences in treatment response based on brain-based variables and requires a more personalized approach to treatment for depressive disorders.

AMANTADINE

Amantadine (Symmetrel; 1-aminoadamantane hydrochloride) is an antiviral drug that was developed in the 1960s for the prophylaxis and treatment of influenza A, but was found to have antiparkinsonian properties. In 1973, it received FDA approval for the treatment of Parkinson disease and is also used to treat drug-induced extrapyramidal symptoms such as parkinsonism, akinesia, and the rabbit syndrome (focal perioral tremor of the choreoathetoid type).

Chemistry

Amantadine is a tricyclic symmetric amine and is a stable, white, or nearly white, crystalline powder that is freely soluble in water and soluble in alcohol and in chloroform. Its molecular weight is 187.71 with a molecular formula $C_{10}H_{18}NCl$ (Fig. 33.11a–7).

Pharmacologic Actions

ABSORPTION AND DISTRIBUTION. Amantadine is well absorbed orally, reaching peak plasma concentrations of 0.2 to 0.4 μg/mL within 1 to 4 hours and attains a steady-state in 4 to 7 days. Doses above 200 mg per day may result in a greater than proportional increase in maximum plasma concentrations. It is approximately 67% bound to plasma protein.

METABOLISM AND ELIMINATION. Amantadine is primarily excreted unchanged in the urine by glomerular filtration and tubular secretion. Increasing plasma concentrations can lead to toxicity, but no absolute values of amantadine concentrations associated with adverse effects have been fully established. The mean half-life ranges from 10 to 14 hours. One study involving 19 healthy volunteers showed the plasma half-life averaged 16 ± 6 hours (range 9 to 31 hours).

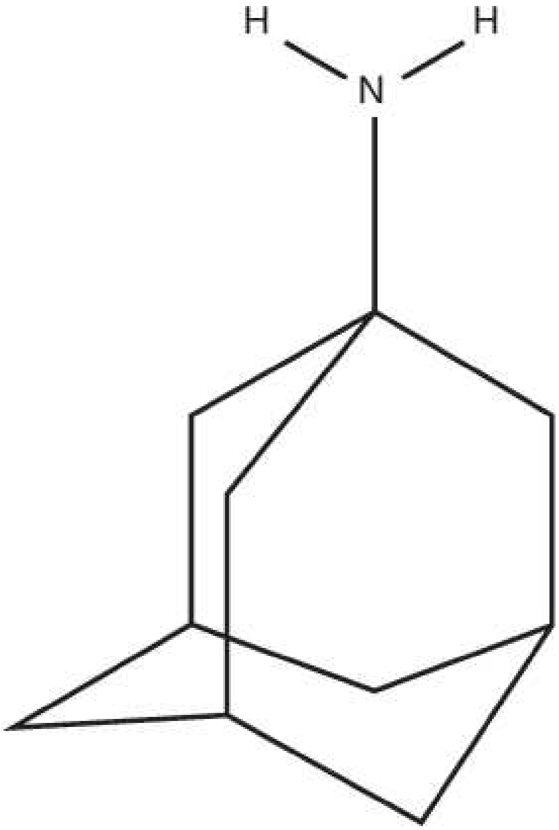

FIGURE 33.11a–7. Molecular structure of amantadine.

SPECIAL CONSIDERATION IN OLDER ADULTS AND RENALLY IMPAIRED. In healthy older adult individuals aged 60 and older, the oral plasma clearance of amantadine is reduced, while the plasma half-life and plasma concentrations are increased. It is unclear whether these changes are due to a decline in renal function or other age-related factors, though evidence suggests that impaired renal function can significantly reduce the clearance of amantadine. The elimination half-life increases two- to threefold or greater when creatinine clearance is less than 40 mL/1.73 m^2 per minute. Acidic urine increases the excretion rate of amantadine, and urine-acidifying drugs may increase the elimination of the drug from the body. Amantadine is removed in negligible amounts by hemodialysis.

Pharmacodynamics

MECHANISM OF ACTION. The exact mechanism of action of amantadine in the treatment of Parkinson disease and drug-induced extrapyramidal reactions is not known. It has preferential selectivity for central catecholamine neurons and acts by potentiating the release of endogenous catecholamines and DA from presynaptic vesicles, blocking the reuptake of DA and having an agonist effect on postsynaptic DA receptors. Amantadine is a weak, noncompetitive NMDA receptor antagonist, and some studies suggest that it may have direct and indirect effects on dopaminergic cell groups. It has no anticholinergic activity, and though it is only 1/209,000 as potent as atropine, it does cause side effects like other anticholinergics, including constipation, dry mouth, and urinary retention. When used alone or in combination with other treatments for mild parkinsonism, many patients derive only transient benefits, if any at all.

In addition to showing NMDA receptor antagonism, amantadine is also a sigma-1 agonist, which may potentially play a role in its effects on parkinsonian symptoms and its antiviral properties. Amantadine has also shown neuroprotective properties via induced expression of glial-cell–derived neurotrophic factor (GDNF) or other antioxidative or anti-inflammatory avenues.

Design and Interpretation of Clinical Trials

Amantadine has been used as an antiviral drug for more than 50 years and was initially developed to fight influenza caused by type A influenza viruses. However, widespread drug resistance has reduced its popularity to such an extent that its use is no longer recommended. Reports of antiparkinsonian effects were observed a few years later, and the first clinical study (n = 163) using amantadine to treat patients with Parkinson disease was published by Schwab and colleagues in 1969 when they reported a positive outcome. Four years later, in 1973, the FDA approved the use of amantadine to treat Parkinson disease.

Since that time, many studies have compared the effectiveness of amantadine and anticholinergic agents, but to date, there are no conclusive data to establish its efficacy as a first-line agent. A double-blind study compared the effects of amantadine and biperiden in treating extrapyramidal side effects in 42 schizophrenic subjects treated with haloperidol (Haldol). Of note, the treatment effect of amantadine and biperiden was similar. This study supports the notion that amantadine is a useful alternative to anticholinergics.

In another double-blind, placebo-controlled study, amantadine was compared to trihexyphenidyl in treating drug-induced parkinsonism. The study showed that amantadine not only produced far fewer adverse effects but also was equally effective as trihexyphenidyl and is a useful alternative to atropine-like drugs. Some case reports show that amantadine may alleviate symptoms of TD. This is a controversial issue, and there are no controlled studies to replicate these results.

A relatively new form of an extended-release formulation of amantadine is available for the treatment of Parkinson disease. The initial starting dose is 137 mg and can be increased to 274 mg per day. The extended-release formulation of amantadine is effective in decreasing motor fluctuations and has a side-effect profile similar to the regularly available form of amantadine. The new extended-release formulation is taken at bedtime and is slowly released into the bloodstream overnight. As a result, the peak levels of the drug are present when it is time to wake up.

Recent studies have also looked at weight-loss potential, reversal of sexual dysfunction, and improvement in cognitive functioning by amantadine, as well as its potential neuroprotective properties and management of sequelae following cerebrovascular events like stroke or traumatic brain injuries. There is also research currently testing to see if amantadine is effective at treating COVID-19, possibly by disrupting viral replication through the viroporin E channel of the virus or agonistic behavior at sigma-1 receptors. One study assessed the weight loss potential of amantadine in patients who were treated with olanzapine (Zyprexa). Significantly fewer subjects taking amantadine gained weight compared to the placebo group. Amantadine stabilized weight in patients treated with olanzapine but did not improve fasting insulin or lipid levels. Recent studies have also found that amantadine may enhance the therapeutic effects of antidepressants. A preliminary study involving four patients with acute bipolar depression found depressive symptoms improved after 1 week of treatment. More studies are needed to confirm these results. Amantadine has a predilection for the nigrostriatal dopaminergic system compared to the mesolimbic system and enhances the CNS dopaminergic activity. This can lead to delusion, paranoia, alterations and elevation of mood, confusion, and disorientation. A few case reports suggest psychotic decompensation in patients treated with amantadine. The vast majority who are treated for extrapyramidal symptoms show a lack of psychotic relapse. Clinicians should be alerted to the fact that patients treated with amantadine are vulnerable to psychotic decompensation.

A double-blind study showed that amantadine did not cause memory impairment in patients treated with the medication, and, therefore, it is a good alternative for older adults who are more prone to these side effects.

Effects on Specific Organs and Systems

Amantadine relieves parkinsonian symptoms by increasing dopaminergic function and increasing the release of catecholamines. These effects can be manifold and manifest as adverse reactions. These include restlessness, depression, irritability, insomnia, agitation, excitement, and occasionally psychotic decompensation. Amantadine is excreted unchanged by the kidneys and patients with renal dysfunction should be monitored closely.

Therapeutic Indications

DRUG-INDUCED EXTRAPYRAMIDAL REACTIONS. Amantadine is indicated in the treatment of drug-induced extrapyramidal reactions, but there are no data that support the use of this drug as a first-line agent, particularly for acute dystonic reactions. Some studies suggest that it is as effective as anticholinergics for extrapyramidal signs and symptoms but is not effective in treating TD and akathisia. Although side effects have been noted with amantadine, the incidence of anticholinergic-type side effects is lower than that observed with the anticholinergic antiparkinsonian drugs. This is particularly useful in older adults who are more sensitive to the anticholinergic effects, both in the CNS, such as anticholinergic delirium, and in the peripheral nervous system, such as urinary retention.

PARKINSON DISEASE AND PARKINSONISM. The usefulness of amantadine in the treatment of parkinsonism is controversial because its efficacy in comparison with the anticholinergic antiparkinsonian drugs has not yet been established. In a controlled trial of 17 patients with parkinsonism, amantadine hydrochloride in a dose of 100 mg b.i.d. showed a minor beneficial effect on physical signs but failed to demonstrate any significant effect on functional disabilities. This improvement in physical signs was only sustained for 8 weeks, but numerous patients noticed deterioration in functional performance after 2 to 5 weeks of treatment.

Amantadine has been shown to be less effective than levodopa in treating idiopathic Parkinson disease (paralysis agitans), postencephalitic parkinsonism, and symptomatic parkinsonism, which may follow injury to the nervous system by carbon monoxide intoxication. It appears to be more effective in treating older adult patients believed to have developed parkinsonism in association with cerebral arteriosclerosis.

Amantadine has been reported to be beneficial in the treatment of selective serotonin reuptake inhibitor-associated side effects, such as lethargy, fatigue, anorgasmia, and ejaculatory inhibition.

BRAIN INJURY. Traumatic brain injury (TBI) is one of the leading causes of death and disability in persons between 15 and 30 years of age. Currently, there are no interventions or treatments to change the prognosis of this condition, and it leads to chronic disability in more than 90,000 survivors of brain injury. The trauma to the brain usually impacts the frontal region, disrupting neurotransmitters, including DA, and causing frontal lobe dysfunction. This syndrome includes problems with aggression, disinhibition, attention, impulsivity, planning, perseveration, amotivation, apathy, memory deficits, mood dysregulation, and problem solving. Other causes of brain injury with similar sequelae include stroke, encephalitis, and aneurysm. The role of DA and lower levels of its metabolites is well established in brain injury, specifically in the frontal lobe regions. Multiple studies have shown that DA is involved in the regulation of cognition, mood, and behavior. In patients with brain injury, there is deficiency and dysregulation of DA in the frontal lobes.

Amantadine was first used in 1988 in two patients for the treatment of TBI. Subsequently, a study assessed 30 patients aged 5 to 59 with TBI with behavior problems including agitation, mood lability, and inattention. Treatment with amantadine led to improvement in 63% of the patients. These positive results led to the hypothesis that since amantadine impacted the dopaminergic system (presynaptic and postsynaptic), it would be an appropriate treatment for patients with TBI with compromised dopaminergic systems. Numerous studies since then have validated these results, and amantadine is now considered safe and beneficial for arousal, agitation, and problems with executive functioning in patients with a head injury. Amantadine has been used frequently for the treatment of head injuries for decades.

Amantadine is one of the most frequently used medications for patients undergoing neurorehabilitation. Various authors have reported the use of amantadine alone or in combination therapy with a positive response in patients with brain injury, specifically in the frontal lobe syndrome. Recent studies suggest that amantadine and other dopaminergic agents may substantially improve the level of executive functioning in these patients. The clinical effects can be seen within a few days to a week. In one study of 184 patients with TBI who were minimally conscious 1 to 4 months after the brain injury, amantadine use after 4 weeks showed a significantly faster recovery compared to the placebo group. Of note, despite the high risk of seizures in this population, amantadine did not increase the risk of adverse effects, including any medical or neurologic events.

Another study showed improvements in attention span, speed of information processing, learning, and alertness. Other studies have shown improvements in apathy, amotivation, and perseveration. The greatest improvements have been seen in a reduction of agitation, physical aggression, distractibility, and mood swings. More recently, amantadine has shown promising results in improving clinical status of dementia patients. According to one report, it is also useful in treating fatigue and depression from multiple sclerosis. Despite the fact that these case reports show promising results, there are limitations to these small and uncontrolled studies, which underscores the need for double-blind, large, and controlled studies.

The exact mechanism of action remains unclear, but amantadine appears to act as an NMDA antagonist and indirect DA agonist. Amantadine may increase dopaminergic activity by increasing presynaptic release and blocking reuptake postsynaptically, targeting the nigrostriatal, mesolimbic, and frontostriatal circuits. The final result is increased arousal, drive, and attention.

These results suggest that patients with severe TBI can be treated with amantadine safely at doses between 200 and 400 mg.

Treatment in Children

Head injury is a common problem in children: nearly 300 out of 100,000 children suffer TBI by the age of 18. Most children are able to survive these injuries yet are left with deficits, including inattention, difficulties with executive functioning, and academic performance. The use of agents to treat these behavior disturbances secondary to head injuries in the pediatric population is less well known, but a small number of studies suggest the use of amantadine. Despite the limited body of literature on the use of amantadine in the pediatric population, it is a useful and beneficial treatment in children with a head injury and symptoms of arousal and alertness.

Pathologic Gambling and Drug Addiction

Pathologic gambling (PG) is a persistent gambling behavior with devastating consequences and is identified as one of the substance abuse and addictive disorders. Despite significant research, there have been no FDA-approved treatments for this condition. Considering the fact that PG and addictive disorders may have similar clinical and biologic underpinnings, it may be appropriate to use drugs for the treatment of PG that are used for drug addiction. Numerous studies suggest the role of DA in addictive, reinforcing, and rewarding behaviors. The pleasure and reward-seeking behavior present in drug addiction is linked to the hypodopaminergic state and may also be responsible for gambling behavior triggering the release of DA and, consequently, a sense of pleasure. In addition, an imbalance in glutamate (Glu) homeostasis may also be responsible for reward-seeking behaviors.

One study suggested that amantadine may be effective in reducing PG and showed symptoms resolving after 4 days of treatment. Another case study of a male patient with PG, comorbid depression, and alcohol dependence showed that amantadine 100 to 150 mg daily had a 43% to 64% reduction in the Gambling Symptom Assessment Scale (G-SAS).

The data suggest that the glutamatergic system may play an important role in PG, and amantadine could be a potential treatment option for this condition.

Amantadine exerts its effects on DA neurotransmission by increasing DA release and DA reuptake inhibition and acts as an antagonist at NMDA glutamatergic receptors. These effects may explain its role as an antidepressant. Numerous studies have suggested that its dopaminergic properties may be beneficial in cocaine-dependent patients with withdrawal symptoms, but more studies are required to validate these results.

Precautions and Adverse Reactions

The most frequently reported CNS adverse reactions (5% to 10%) are insomnia, dizziness, and impaired concentration. CNS adverse reactions that are reported less frequently (1% to 5%) include restlessness, irritability, agitation, excitement, depression, anxiety, and confusion. More severe CNS side effects are seizures and psychotic symptoms. Overdose may produce acute toxic psychosis.

For patients already receiving medications for Parkinson disease, the addition of amantadine to their treatment may lead to intense urges to engage in sex acts, gamble, or spend money in wanton and reckless ways because medications that increase central dopaminergic tone, as amantadine does, may impair a patient's ability to exhibit self-control. Prescribers should routinely inquire about the development of these urges (gambling, sexual, uncontrolled spending, or other urges) and lower the dose of amantadine or completely stop the medication if a patient reports the development of such urges.

Other less frequently (1% to 5%) reported adverse reactions are anorexia, dry mouth, peripheral edema, orthostatic hypotension, congestive heart failure, headache, somnolence, nervousness, dream abnormality, dry nose, diarrhea, and fatigue. Livedo reticularis of the legs (purple discoloration of the skin caused by dilatation of the vessels) may occur in some patients. The phenomenon usually clears within a month after cessation of amantadine.

Rare (less than 0.1%) occurring adverse reactions are instances of convulsion, leukopenia, neutropenia, eczematoid dermatitis, oculogyric episodes, suicidal attempt, suicide, and suicidal ideation.

Amantadine is relatively contraindicated in patients who have a known history of seizures or renal disease. It should also be used only with caution in persons with congestive heart failure or edema. There is evidence to suggest that it is teratogenic. Consequently, it should not be used by pregnant women or women who are breastfeeding.

Amantadine should not be discontinued abruptly in patients with Parkinson disease since a few patients have experienced a parkinsonian crisis (i.e., a sudden marked clinical deterioration upon abrupt cessation of this agent) (Table 33.11a–3).

Deaths

Amantadine has been associated with deaths secondary to overdose in amounts as low as 2 g. Anticholinergic effects may be responsible for these deaths secondary to cardiac, respiratory, renal, or CNS toxicity.

Table 33.11a–3.
Adverse Effects of Amantadine

Cardiovascular system
- Arrhythmia
- Tachycardia

Nervous system
- Delirium
- Paranoia
- Paresthesias

Drug Interaction

Coadministration of amantadine with phenelzine (Nardil) or other MAOIs may result in a significant increase in blood pressure. There is an increased likelihood of side effects such as confusion, hallucinations, nightmares, dry mouth, and blurred vision if administered with anticholinergics. The coadministration with CNS stimulants can result in insomnia, irritability, nervousness, and possibly seizures.

Laboratory Interferences

Amantadine is not known to alter the results of any standard laboratory tests and values.

Dosage and Administration

ADULT. The usual starting dose of amantadine hydrochloride is 100 mg twice a day when used alone and can be cautiously increased to 200 mg twice a day in patients whose responses are not optimal. Amantadine has an onset of action, usually within 48 hours.

The initial dose of amantadine hydrochloride is 100 mg daily for patients with serious associated medical illnesses or who are receiving high doses of other drugs to treat Parkinson disease. After one to several weeks at 100 mg once daily, the dose may be increased to 100 mg twice daily, if necessary.

Some patients have reported a notable decrease in efficacy after a few months and may benefit by increasing the dose to 300 mg daily or temporarily discontinuing amantadine for several weeks, followed by reintroduction of the drug.

Dosage for Concomitant Therapy

When patients do not respond to monotherapy with anticholinergic anti-Parkinson drugs, monotherapy or concomitant therapy with amantadine may be beneficial. In some cases, when amantadine and levodopa are initiated concurrently, the patient can exhibit rapid therapeutic benefits. Amantadine should be held constant at 100 mg daily or twice daily while the daily dose of levodopa is gradually increased to optimal benefit.

Table 33.11a–4.
Dosage for Impaired Renal Function

Creatinine Clearance (mL/min/1.73 m^2)	Amantadine Hydrochloride Dosage
30–50	200 mg 1st day and 100 mg each day thereafter
15–29	200 mg 1st day followed by 100 mg on alternate days
<15	200 mg every 7 days

The recommended dosage for patients on hemodialysis is 200 mg every 7 days.

Dosage for Drug-Induced Extrapyramidal Reactions

The usual starting dose of amantadine hydrochloride is 100 mg twice a day. Occasionally, patients whose responses are not optimal with amantadine hydrochloride at 200 mg daily may benefit from an increase of up to 300 mg daily in divided doses (Table 33.11a–4).

Further Readings

Comings DE, Wu S, Rostamkhani M, McGue M, Iacono WG, MacMurray JP. Association of the muscarinic cholinergic 2 receptor (CHRM2) gene with major depression in women. *Am J Med Genet*. 2002;114(5):527–529.

Cui J, Hollmén M, Li L, et al. New use of an old drug: inhibition of breast cancer stem cells by benztropine mesylate. *Oncotarget*. 2017;8(1):1007–1022.

Damusz W, Dekundy A, Scheschonka A, Riederer P. Amantadine: reappraisal of the timeless diamond—target updates and novel therapeutic potentials. *J Neural Transm*. 2021;128:127–169.

Drevets WC, Furey ML. Replication of scopolamine's antidepressant efficacy in major depressive disorder: a randomized, placebo-controlled clinical trial. *Biol Psychiatry*. 2010;67(5):432–438.

*Drimer T, Shahal B, Barak Y. Effects of discontinuation of long-term anticholinergic treatment in elderly schizophrenia patients. *Int Clin Psychopharmacol*. 2004;19(1):27–29.

Furey ML, Drevets WC. Antidepressant efficacy of the antimuscarinic drug scopolamine: a randomized, placebo-controlled clinical trial. *Arch Gen Psychiatry*. 2006;63(10):1121–1129.

Furey ML, Drevets WC, Hoffman EM, Frankel E, Speer AM, Zarate CA Jr. Potential of pretreatment neural activity in the visual cortex during emotional processing to predict treatment response to scopolamine in major depressive disorder. *JAMA Psychiatry*. 2013;70(3):280–290.

Furey ML, Khanna A, Hoffman EM, Drevets WC. Scopolamine produces larger antidepressant and anti-anxiety effects in women than in men. *Neuropsychopharmacology*. 2010;35(12):2479–2488.

Giacino JT, Whyte J, Bagiella E, et al. Placebo-controlled trial of amantadine for severe traumatic brain injury. *N Engl J Med*. 2012;366(9):819–826.

Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 2020;583:459–468.

Graham KA, Gu H, Lieberman JA, Harp JB, Perkins DO. Double-blind, placebo-controlled investigation of amantadine for weight loss in subjects who gained weight with olanzapine. *Am J Psychiatry*. 2005;162(9):1744–1746.

Hodgins DC, Stea JN, Grant JE. Gambling disorders. *Lancet*. 2011;378(9806):1874–1884.

Hughes S, Colantonio A, Santaguida PL, Paton T. Amantadine to enhance readiness for rehabilitation following severe traumatic brain injury. *Brain Inj*. 2005;19(14):1197–1206.

Janowsky DS. Serendipity strikes again: scopolamine as an antidepressant agent in bipolar depressed patients. *Curr Psychiatry Rep*. 2011;13(6):443–445.

Jeraisy MA, AlFuraih M, AlSaif R, AlKhalifah B, AlOtaibi H, Abolfotouh MA. Efficacy of scopolamine transdermal patch in children with sialorrhea in a pediatric tertiary care hospital. *BMC Pediatrics*. 2020;20:437.

Ji C-X, Zhang J-J. Effect of scopolamine on depression in mice [in Chinese]. *Yao Xue Xue Bao*. 2011;46(4):400–405.

Krystal JH. N-methyl-D-aspartate glutamate receptor antagonists and the promise of rapid-acting antidepressants. *Arch Gen Psychiatry*. 2010;67(11):1110–1111.

Krzystanek M, Palasz A. Possibility of a new indication for amantadine in the treatment of bipolar depression-case series study. *Pharmaceuticals (Basel)*. 2020;13(10):326.

Pettorruso M, Martinotti G, Di Nicola M, et al. Amantadine in the treatment of pathological gambling: a case report. *Front Psychiatry*. 2012;3:102.

Potenza MN. The neurobiology of pathological gambling and drug addiction: an overview and new findings. *Philos Trans R Soc Lond B Biol Sci*. 2008;363(1507):3181–3189.

Rejdak K, Grieb P. Adamantanes might be protective from COVID-19 in patients with neurological diseases: multiple sclerosis, parkinsonism and cognitive impairment. *Mult Scler Relat Disord*. 2020;42:102163.

Rodríguez-Muñoz M, Sánchez-Blázquez P, Merlos M, Garzón-Niño J. Endocannabinoid control of glutamate NMDA receptors: the therapeutic potential and consequences of dysfunction. *Onctotarget*. 2016;7(34):55840–55862.

Schatzberg AF. Drugs to treat extrapyramidal side effects. In: Schatzberg AF, Nemeroff CB, eds. *The American Psychiatric Publishing Textbook of Psychopharmacology*. 3rd ed. American Psychiatric Publishing; 2004:284.

Sharma VD, Lyons KE, Pahwa R. Amantadine extended-release capsules for levodopa-induced dyskinesia in patients with Parkinson’s disease. *Ther Clin Risk Manag*. 2018;14:665–673.

Sogawa C, Eguchi T, Tran MT, et al. Antiparkinson drug benztropine suppresses tumor growth, circulating tumor cells, and metastasis by acting on SLC6A3/DAT and reducing STAT3. *Cancers(Basel)*. 2020;12(2):523.

Tandon R, Dequardo JR. Treatment of schizophrenia with anticholinergic medications. *Am J Psychiatry*. 1995;152(5):814–815.

Wang JC, Hinrichs AL, Stock H, et al. Evidence of common and specific genetic effects: association of the muscarinic acetylcholine receptor M2 (CHRM2) gene with alcohol dependence and major depressive syndrome. *Hum Mol Genet*. 2004;13(17):1903–1911.

Williams SE. Amantadine treatment following traumatic brain injury in children. *Brain Inj*. 2007;21(9):885–889.

Zhou D, Zhang Z, Liu L, et al. The antidepressant-like effects of biperiden may involve BDNF/TrkB signaling-mediated BICC1 expression in the hippocampus and prefrontal cortex of mice. *Pharmacol Biochem Behav*. 2017;157:47–57.

33.11b Cholinesterase Inhibitors and Memantine

Samir A. Sabbag, M.D., Zelde Espinel, M.D., M.P.H., Luminita Luca, M.D., Ph.D., and Elizabeth A. Crocco, M.D.

In 2015, the total number of individuals affected by dementia worldwide was 47 million. It is predicted that this number will increase to 75 million in 2030 and 132 million by 2050. With the aging of the population, particularly in the United States, the annual number of new cases due to Alzheimer and related dementias is expected to double by 2050. The incidence of Alzheimer disease (AD) increases with age. Epidemiologic studies in the United States report its prevalence to be about 3% at age 65, which increases to greater than 30% by age 85. In 2021, it was estimated that 6.2 million Americans aged 65 and older were living with Alzheimer dementia. As the population ages, the prevalence of the disease will increase and significantly impact society and its health care systems.

Currently, cholinesterase inhibitors and memantine (Namenda) are the medications approved by the U.S. Food and Drug Administration (FDA) for the treatment of AD and other dementias (Table 33.11b–1). Since their introduction in the early 1990s, cholinesterase inhibitors were indicated only for mild to moderately ill patients with AD. Their indication has expanded over time to include patients with moderate to severe AD and patients with dementia associated with Parkinson disease (PD). These agents are also being studied as a potential treatment for other types of non-AD dementias, but they presently lack an FDA indication. Memantine, an *N*-methyl-D-aspartate (NMDA) receptor antagonist, has been approved for moderate to severe AD. It is routinely used with a cholinesterase inhibitor in clinical practice, as studies have demonstrated how this combination provides a potentially better response compared to monotherapy with a cholinesterase inhibitor.

Table 33.11b–1.
Medications Used in Neurocognitive Disorders

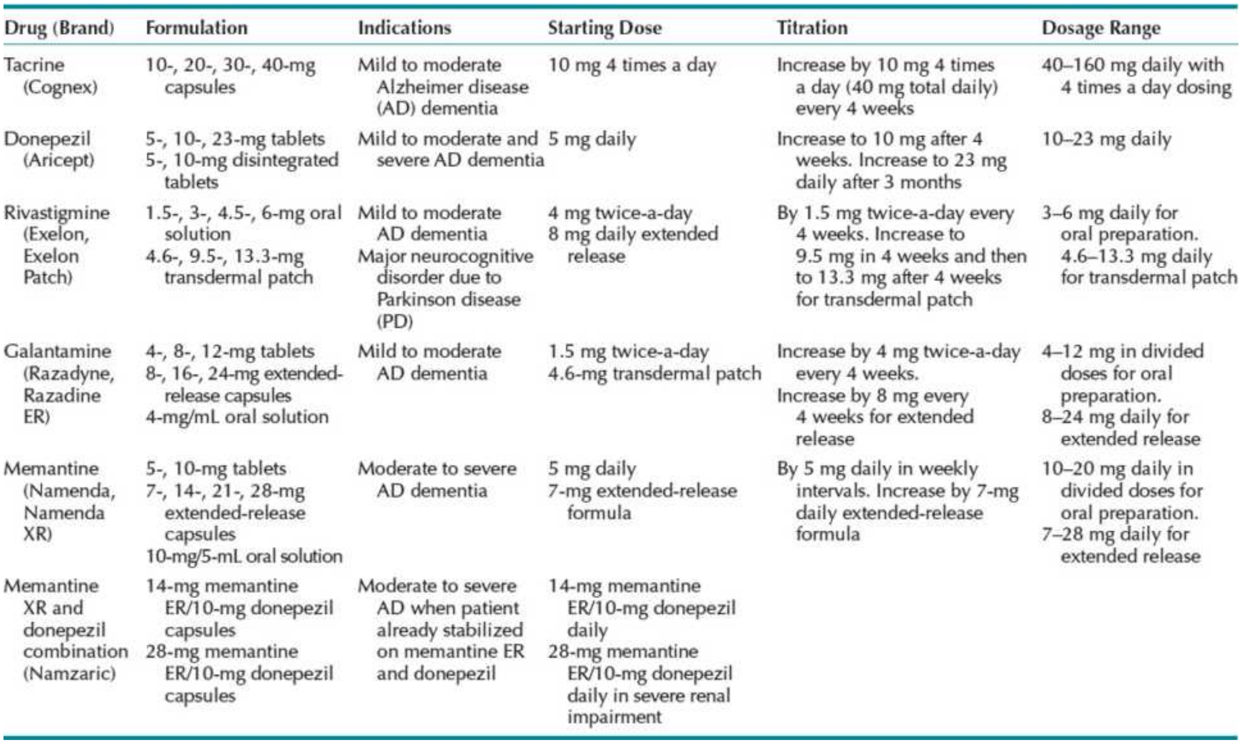

Drug (Brand)	Formulation	Indications	Starting Dose	Titration	Dosage Range
Tacrine (Cognex)	10-, 20-, 30-, 40-mg capsules	Mild to moderate Alzheimer disease (AD) dementia	10 mg 4 times a day	Increase by 10 mg 4 times a day (40 mg total daily) every 4 weeks	40–160 mg daily with 4 times a day dosing
Donepezil (Aricept)	5-, 10-, 23-mg tablets 5-, 10-mg disintegrated tablets	Mild to moderate and severe AD dementia	5 mg daily	Increase to 10 mg after 4 weeks. Increase to 23 mg daily after 3 months	10–23 mg daily
Rivastigmine (Exelon, Exelon Patch)	1.5-, 3-, 4.5-, 6-mg oral solution 4.6-, 9.5-, 13.3-mg transdermal patch	Mild to moderate AD dementia Major neurocognitive disorder due to Parkinson disease (PD)	4 mg twice-a-day 8 mg daily extended release	By 1.5 mg twice-a-day every 4 weeks. Increase to 9.5 mg in 4 weeks and then to 13.3 mg after 4 weeks for transdermal patch	3–6 mg daily for oral preparation. 4.6–13.3 mg daily for transdermal patch
Galantamine (Razadyne, Razadine ER)	4-, 8-, 12-mg tablets 8-, 16-, 24-mg extended-release capsules 4-mg/mL oral solution	Mild to moderate AD dementia	1.5 mg twice-a-day 4.6-mg transdermal patch	Increase by 4 mg twice-a-day every 4 weeks. Increase by 8 mg every 4 weeks for extended release	4–12 mg in divided doses for oral preparation. 8–24 mg daily for extended release
Memantine (Namenda, Namenda XR)	5-, 10-mg tablets 7-, 14-, 21-, 28-mg extended-release capsules 10-mg/5-mL oral solution	Moderate to severe AD dementia	5 mg daily 7-mg extended-release formula	By 5 mg daily in weekly intervals. Increase by 7-mg daily extended-release formula	10–20 mg daily in divided doses for oral preparation. 7–28 mg daily for extended release
Memantine XR and donepezil combination (Namzaric)	14-mg memantine ER/10-mg donepezil capsules 28-mg memantine ER/10-mg donepezil capsules	Moderate to severe AD when patient already stabilized on memantine ER and donepezil	14-mg memantine ER/10-mg donepezil daily 28-mg memantine ER/10-mg donepezil daily in severe renal impairment		

Both the histopathology and clinical features of AD were first described over 100 years ago by Alois Alzheimer. Even today, its underlying pathophysiology and genetic basis continue to be investigated. Amyloid-β peptide and phosphorylated tau proteins make up the neuritic plaques and neurofibrillary tangles seen in AD. These proteins play a central role in the two leading theories behind the pathogenesis of the disease. Amyloid positron emission tomography (PET) is now FDA-approved as a diagnostic biomarker in the diagnosis of AD. Additionally, Tau PET imaging is also being investigated for its role in AD as well as other related dementias such as frontotemporal dementia (FTD).

Loss of cholinergic neurons and impairment in cholinergic function have for many years been demonstrated in AD. Early subcortical changes occur in the nucleus basalis of Meynert, which provides cholinergic innervation to the hippocampus and cortex. The decline of acetylcholine in this area has been linked to the cognitive impairment seen in AD and has been the basis for cholinergic therapies.

Currently, the most effective approach to increasing acetylcholine activity is achieved with the use of cholinesterase inhibitors. Cholinesterase inhibitors deter the activity of the enzymes that break down acetylcholine at the synapse. Two enzymes are known to hydrolyze acetylcholine: acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE). AChE has two binding sites—anionic and esteratic—and is present in several areas of the brain, as well as in erythrocytes, platelets, lymphocytes, and skeletal muscle. BuChE has a similar structure to that of AChE but lacks the peripheral anionic binding site. It is synthesized in the liver, secreted into the plasma, and found in the gastrointestinal tract and heart, comprising approximately 10% of the total amount of cholinesterase in the central nervous system (CNS) present in glial cells. In vitro studies indicate that cholinesterase inhibitors can selectively inhibit AChE and BuChE to varying extents. Currently, three cholinesterase inhibitors are available in the United States: donepezil (Aricept), rivastigmine (Exelon), and galantamine (Razadyne). The first cholinesterase to be approved for the treatment of dementia, tacrine (Cognex), was removed from the U.S. market due to its high side-effect profile.

Tacrine inhibits BuChE 18 times more potently than AChE, while donepezil inhibits AChE with higher selectivity. Rivastigmine inhibits both AChE and BuChE, and galantamine selectively inhibits AChE while also producing allosteric modulation of nicotinic receptors. Although the actions of AChE have been thoroughly studied, the relevance of BuChE and nicotinic receptor modulation has become increasingly apparent to the overall function of cholinergic neurons. BuChE has been shown to be important in the development of amyloid plaques. When a polymorphic mutation of BuChE, known as the K variant, is combined with the apolipoprotein E4 (APOE4), this may result in an increased risk factor for developing AD. Nicotinic modulation results in increased ion channel opening, enhancing cholinergic transmission independent of enzymatic inhibition. Nicotinic receptors are found in presynaptic and postsynaptic neurons. Presynaptic nicotinic receptors regulate the release of acetylcholine, glutamate, and γ-aminobutyric acid (GABA), while postsynaptic nicotinic receptors mediate cholinergic transmission in the hippocampus and cerebral cortex. Both receptor types play an important role in memory and learning.

Cholinesterase inhibitors can be classified into three categories: reversible, pseudo-irreversible, and irreversible. These categories are based on the duration of enzyme inhibition. Reversible inhibitors bind to the enzyme for a very short time (seconds to minutes). Pseudoirreversible inhibitors bind for hours, but new enzymes are not developed to assist in the body's biochemical functions. Irreversible inhibitors bind to the enzyme and prevent further activity, prompting the body to synthesize new enzymes.

Prior to the appearance of medications for cognitive impairment, the estimated total direct and indirect costs for the treatment and care of Alzheimer patients in the United States was nearly $100 billion in 1990. Their introduction was associated with a moderate reduction in costs of medical and social services, as well as a slight increase in functioning and a delay in nursing home placement.

In addition to AD, there are other neurodegenerative illnesses that can commonly lead to a major neurocognitive disorder. Major neurocognitive disorders due to PD, Lewy body disease (LBD), vascular dementia (VD), as well as FTD, have distinct pathophysiologic mechanisms, although some overlap may occur. Dementia can occur in up to 40% of idiopathic PD patients who demonstrate a decline in cholinergic function. The cholinesterase inhibitor, rivastigmine, is indicated by the FDA for the treatment of cognitive impairment in PD, and there is some evidence that other cholinesterase inhibitors like donepezil may also improve major neurocognitive disorder due to PD. Diffuse LBD, although a distinct illness clinically, has similar pathophysiologic mechanisms to PD and may possibly benefit from cholinesterase inhibitors. VD may produce cerebrovascular damage-causing cholinergic dysfunction. Additionally, the added presence of AD pathology rather than pure VD may explain why some studies demonstrate higher rates than others of improvement in VD with the use of cholinesterase inhibitors. Currently, there are no adequate randomized controlled trials (RCTs) showing the efficacy of cholinesterase inhibitors in the treatment of

FTD and similar dementias with tau neuropathology. Given the heterogeneity of neurodegenerative dementias in older adults as well as the lack of adequate disease-modifying treatments for these progressive conditions, cholinesterase inhibitors are commonly used off-label in clinical practice in several select non-AD dementia conditions. Ultimately, tolerability and adverse reactions must be carefully taken into account during off-label use as quality clinical research evidence is lacking, and the benefits must outweigh the risks.

Patients with mild neurocognitive disorders, including mild cognitive impairment (MCI), are a group of patients who display a modest decline in cognition. They do not meet the criteria for dementia, as this decline does not preclude them from the independent performance of everyday activities. These individuals pose a challenging dilemma, as cholinesterase inhibitors are not approved for the treatment of their illness, even though the annual progression into a major neurocognitive disorder is approximately 10% to 15%. Trials with these medications have not been found to decrease the rate of progression to dementia, and at this time, there is weak support for using them.

Very few studies have looked into the efficacy of cholinesterase inhibitors in treating behavioral and psychiatric symptoms, such as agitation, visual hallucinations, or apathy in patients with AD. The evidence is limited due to methodologic problems with these studies. Currently, despite not having FDA approval for the treatment of behavioral or psychiatric problems, their use could be an alternative in the absence of other safe and effective options.

Cholinesterase inhibitors and memantine are moderately effective in treating the symptoms of AD, including cognitive, functional, and behavioral impairment. Unfortunately, these agents are not disease modifying, and risks versus benefits must be carefully weighed by the clinician before starting treatment. Clinicians, caregivers, and patients should be made aware of and educated on realistic goals of therapy. The overall results of the pharmacologic intervention should be to maintain for a longer period of time the patient’s quality of life and functional ability. By detecting, diagnosing, and treating AD early during its development, there is an opportunity for clinicians and caregivers to optimize drug therapy and other psychosocial interventions that are important in the disease process. Cholinesterase inhibitors are indicated for all stages of the disease, and early therapeutic interventions can maximize pharmacologic efficacy with these agents.

TACRINE

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Acetylcholine
Mechanism of Action	Enzyme inhibitor (Butyrylcholinesterase [BuChE] and AchE)

The first cholinesterase inhibitor approved by the FDA for the treatment of AD, tacrine, was introduced to the market in 1993. Tacrine is a reversible cholinesterase inhibitor that works by making intrasynaptic acetylcholine readily available in the brain of patients with AD. Despite being the first agent developed for AD, the use of tacrine was inconvenient due to its multiple daily dosing of three to four times a day and the likelihood of hepatotoxicity, which required constant monitoring of liver function during its use. In 2013, tacrine was removed from the U.S. market due to concerns with safety and is no longer available.

DONEPEZIL

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Acetylcholine
Mechanism of Action	Enzyme inhibitor (acetylcholinesterase)

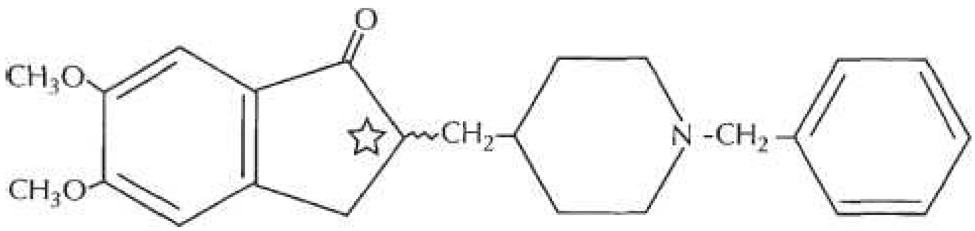

FIGURE 33.11b–1. Molecular structure of donepezil.

Originally developed in 1986, Donepezil was approved by the FDA in 1997 for its use in mild to moderate AD. It was later approved in 2006 for the treatment of severe AD as well. Unlike its predecessor, tacrine, this medication immediately appealed to clinicians thanks to its once-a-day administration, lack of hepatotoxicity, and not having to measure serum liver transaminase levels constantly.

Chemistry

Figure 33.11b–1 shows the molecular structure of donepezil hydrochloride. It is a piperidine compound that consists of (±) –2,3 dihydro-5,6-dimethoxy-2-[[1-(phenylmethyl)-4-piperidinyl]methyl]-1H-inden-1-one hydrochloride.

Pharmacologic Actions

Pharmacokinetics. Donepezil is not affected by food intake in its regular or oral disintegrating tablet presentations. The relationship between donepezil dose and its area under the plasma concentration–time curve (AUC) is linear at a dose range of 1 to 10 mg. It is bound to plasma proteins by approximately 96% (75% to albumin and 21% to α_1-acid glycoprotein). The donepezil 10-mg formulation reaches peak plasma concentration at approximately 3 hours, compared to 8 hours in the 23-mg tablet. Its volume of distribution at a steady state is between 12 and 16 L/kg, reaching steady state after 2 weeks of use.

Donepezil undergoes extensive hepatic metabolism by the CYP 2D6 and 3A4. 6-O-desmethyl-donepezil is its major active metabolite and possesses equal potency. It is also converted into an inactive *cis*-*N*-oxide metabolite and two glucuronide conjugates. The elimination half-life of donepezil is approximately 70 hours, allowing for once-daily administration. Age, gender, hepatic disease, and renal impairment have not been shown to affect donepezil disposition to any significant degree. Due to this, dose adjustments based on renal or liver function are not necessary. Donepezil is classified as a category C drug. There are currently no adequate trials of donepezil in pregnant women; however, animal trials revealed that it may cause fetal harm. It, therefore, should not be used on pregnant women. It is unknown if it is excreted in breast milk.

Pharmacodynamics. Donepezil enhances cholinergic neurotransmission by increasing the concentration of acetylcholine through reversible, noncompetitive inhibition of AChE, the main enzyme involved in the hydrolysis of acetylcholine. Unlike tacrine, it is not highly selective for inhibition of BuChE. It is a potent, selective inhibitor that acts on the CNS, and it does not act on the peripheral nervous system.

Design and Interpretation of Clinical Drug Studies

A substantial number of randomized, double-blind, placebo-controlled studies have provided evidence of the effectiveness of donepezil as a treatment for AD at all its stages. The results of these studies confirmed that donepezil was effective and well-tolerated in treating the symptoms of mild to moderate AD, leading to its initial FDA approval. Additionally, the effectiveness of donepezil for the treatment of severe AD has been demonstrated by studies using both 10-mg daily doses and, more recently, 23 mg.

To establish donepezil's effect on mild to moderate AD, two large multicenter clinical trials were conducted, both using similar study methods. One of them was a U.S. study that took place during a 15-week period (12 weeks of active treatment followed by a 3-week blind placebo washout period), and the other one was a multinational study done during a 30-week period (24 weeks of treatment followed by a 6-week washout period). Patients were randomized into three groups: placebo, donepezil 5 mg, and donepezil 10 mg. In both studies, at the end of each treatment period (weeks 12 and 24), participants treated with placebo had a decline in their ADAS-cog score, while donepezil-treated patients showed statistically significant dose-dependent improvement. A scale measuring clinician’s assessment of functionality with the help of information provided by a caregiver called the Clinician’s Interview-Based Impression of Change with caregiver input (CIBIC-plus) revealed scores at weeks 12 and 24 in each study that showed statistically significant improvement in the donepezil-treated groups. The 12-week study presented no difference between the 5-mg and 10-mg groups. On the other hand, the international study group demonstrated a statistically significant difference between the two doses, with the 10-mg donepezil group showing greater benefits in all outcome measures when compared to the 5-mg group. After the placebo washout periods, ADAS-cog scores declined in both donepezil-treated groups of the two studies, reaching scores that could not be differentiated from the placebo groups.

A 6-month randomized, double-blind, parallel-group, placebo-controlled study of 248 nursing home-dwelling patients with severe AD (Mini-Mental State Examination [MMSE] score of 1 to 10) evaluated the effect of donepezil on cognition and daily functioning. Patients were assigned to donepezil or placebo. Donepezil dose was 5 mg/day for 30 days and then increased to 10 mg/day for the remainder of the study. The primary study endpoints were changed from baseline to 6 months on the Severe Impairment Battery (SIB) and the Alzheimer’s Disease Cooperative Study–Activities of Daily Living–Severe Impairment Version (ADCS-ADL-SIV). The SIB uses a validated 40-item scale developed for the evaluation of the severity of cognitive dysfunction in more advanced AD patients. ADCS-ADL-SIV is a caregiver-based ADL scale composed of 19 items developed for use in clinical studies of dementia. When compared to baseline at 6 months, patients treated with donepezil displayed more improvement in SIB scores and less decline in the ADCS-ADL scores than placebo in severe AD subjects.

The use of donepezil 23 mg in moderate to severe AD was approved by the FDA in 2010, based on the conclusions of a multinational, randomized, double-blind study that included patients who were already being treated with a 10-mg dose for at least 3 months. 1,467 patients were randomly assigned to receive either a 23 mg once-a-day dose of donepezil or a standard dose of 10 mg for 24 weeks. At week 24, changes in cognition and global functioning were assessed with the SIB and the CIBIC-plus, respectively. This study revealed a statistically significant benefit in cognition with the 23-mg dosing but did not find a significant change in global functioning. Side effects were also monitored, revealing an increased incidence of gastrointestinal symptoms, which included nausea, vomiting, and diarrhea, in the higher 23-mg dose group.

One randomized placebo-controlled study was conducted over a 1-year period, evaluating the preservation of function in 431 patients with probable AD who received either placebo or donepezil. Participants randomized into the active treatment arm were started on 5 mg with a posterior dose increase to 10 mg on day 28. The function was assessed by evaluating basic activities of daily living (ADLs) and instrumental ADLs through the AD Functional Assessment and Change Scale. The results indicated that after 1 year, donepezil had a 38% reduction in the risk of functional decline when compared to placebo. Moreover, this study showed how donepezil delayed functional decline by 5 months in patients with AD. At 48 weeks, the probability of survival with no evidence of functional loss was 51% for those taking donepezil, compared to only 35% for those on placebo. These results demonstrated that continued use of donepezil benefits patients over a period of 1 year and that untreated patients tend to decline in their functional ability.

Effects on Specific Organs and Systems

In older adults, hepatic function is frequently decreased, leading to reduced tolerability of certain medications. Donepezil does not induce elevation of liver transaminases or cause hepatotoxicity and is, in general, well tolerated. Even though kidney function may also decrease with aging, donepezil may be safely administered in patients with moderate to severe renal impairment. This medication may produce vagotonic effects in the sinoatrial and atrioventricular nodes by acting on the muscarinic receptors, causing a slowing of the heart rate. Like all cholinesterase inhibitors, it should be used with caution in patients with cardiac arrhythmias and significant cardiac dysfunction. Due to its increased cholinergic activity, donepezil may also augment gastric acid secretion in a dose-dependent fashion, causing gastrointestinal complications that potentially may increase the risk of gastric disease and gastrointestinal bleeding.

Therapeutic Indications

Donepezil is indicated for the treatment of mild, moderate, and severe AD. Additionally, select randomized, double-blind, placebo-controlled studies have demonstrated some benefit in cognition and functioning when treating VD. Donepezil has not been approved by the FDA for this use, as there is a concern for its safety in this population due to cardiovascular and cerebrovascular risks. Similarly, small studies have shown improvement in cognition in patients with LBD, traumatic brain injury, and dementia associated with PD, but the use of donepezil in these populations remains off-label.

Precautions and Adverse Reactions

The most common adverse reactions leading to discontinuation and poor tolerability of dose optimization in clinical settings include diarrhea, nausea, vomiting, insomnia, muscle cramps, fatigue, and anorexia. These side effects are usually mild and transient and have not resulted in a significant dropout rate during clinical trials with donepezil. Less commonly, donepezil can potentially cause more severe reactions such as seizures, cerebrovascular disease, or exacerbation of chronic obstructive pulmonary disease (COPD).

The incidence of gastrointestinal adverse side effects did not differ between patients with mild to moderate dementia when compared to severely ill patients. Patients who suffer from peptic ulcer disease, or are taking NSAIDs, need to be carefully monitored for gastrointestinal bleeding. Caution should also be exerted when using the 23-mg dose in patients who weigh less than 55 kg, as there have been increased reports of nausea, vomiting, and weight loss in this group. Due to donepezil’s effect on cardiac conduction, donepezil should be used with caution in patients with cardiac arrhythmias, especially sick sinus syndrome. Patients may be at risk of developing bradycardia. It is recommended to titrate donepezil slowly to minimize the incidence of side effects. Additionally, donepezil is contraindicated in patients with known hypersensitivity to the drug or to piperidine derivatives.

Drug Interactions

Although metabolized mainly through the CYP 2D6 and 3A4 and highly protein-bound at 96%, donepezil has minimal influence on drugs metabolized through these pathways. Pharmacokinetic studies have been conducted assessing the induction effects of donepezil on cimetidine, ketoconazole, risperidone (which is also metabolized through 2D6 and 3A4), warfarin, and digoxin, and no effects have been observed on the pharmacokinetics of these drugs.

In vitro studies of inhibitors of CYP 3A4, such as ketoconazole, and 2D6, such as quinidine, have shown inhibition of donepezil metabolism. These studies have not translated into clinically significant drug–drug interaction. Phenytoin, carbamazepine, dexamethasone, rifampin, and phenobarbital, which are inducers of 2D6 and 3A4 pathways, can potentially increase donepezil’s elimination rate. Donepezil is not affected by the concomitant use of cimetidine or digoxin. As with all cholinesterase inhibitors, caution is advised when patients are taking anticholinergic medications, as donepezil may interfere with their activity due to its mechanism of action. Donepezil may also lengthen the action of succinylcholine-type muscle relaxants during anesthesia.

Laboratory Interferences

Donepezil has not been shown to interfere with any laboratory test, including biochemical, urinalysis, or spectrometry.

Dosage and Administration

Donepezil is available in 5-, 10-, and 23-mg tablets and in an orally disintegrating tablet form of 5 and 10 mg. The recommended starting dose is 5 mg once a day at bedtime. After 4 to 6 weeks, the dose can be increased to 10 mg depending on tolerability. For mild to moderate AD, 5 and 10 mg are the recommended effective doses. In patients presenting with moderate to severe dementia, clinical trials have found doses of 10 and 23 mg effective. In the moderate to severe group, for those who are tolerating it well, the dose may be increased to 23 mg after a minimum of 3 months on a 10-mg dose. The medication should be administered at bedtime so that peak plasma concentrations occur during the night and potential adverse effects are minimized. As food intake does not affect its absorption, donepezil may be taken with or without food.

RIVASTIGMINE

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Acetylcholine
Mechanism of Action	Enzyme inhibitor (AchE, Butyrylcholinesterase [BuChE])

Rivastigmine has been approved for the treatment of mild to moderate AD since 2000 by the FDA. It is also the only cholinesterase inhibitor approved, since 2006, for the treatment of mild to moderate dementia due to major neurocognitive disorder related to PD. Additionally, in 2013, rivastigmine was approved for severe AD. Rivastigmine is a second-generation cholinesterase inhibitor and has a different pharmacologic profile in comparison to the earlier cholinesterase inhibitors, tacrine and donepezil.

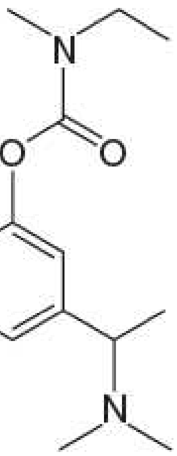

FIGURE 33.11b–2. Molecular structure of rivastigmine.

Chemistry

Rivastigmine is composed of the chemical compound (S)-3-[1-(dimethylamino)ethyl]phenylethyl(methyl)carbamate. Its molecular structure is shown in Figure 33.11b–2.

Pharmacologic Actions

Pharmacokinetics. Rivastigmine is rapidly absorbed in its oral formulation, with a mean time to maximum plasma concentration (T_{max}) of 1 hour (range 0.8 to 1.7 hours). Additionally, it has a mean absolute bioavailability of 60% in comparison to intravenous drug administration. When ingested with food, its C_{max} is decreased by 30% and increases the total amount of drug in the body (AUC) by 30%. Because its bioavailability is unaffected, it is recommended that oral rivastigmine always be given with food to potentially prevent the drug's most common side effects, which are nausea, vomiting, and other forms of gastrointestinal distress. Rivastigmine is only 40% bound to plasma proteins with a volume of distribution between 1.8 and 2.7 L/kg. The relationship between drug dose and AUC is linear, with a mean elimination half-life of 1.6 hours. The drug easily crosses the blood–brain barrier. Cerebrospinal fluid (CSF) rivastigmine concentrations are rapidly reached with a T_{max} of 1.4 to 3.8 hours, and it is quickly eliminated from the CSF with a half-life of 0.31 to 2.95 hours. Rivastigmine is converted after activity to an inactive metabolite which undergoes *N*-demethylation and sulfate conjugation prior to elimination. Rivastigmine is mainly eliminated through renal excretion, and it is not metabolized by hepatic oxidative CYP isozymes. Age, hepatic impairment, and renal insufficiency do not significantly affect rivastigmine disposition, and dosage adjustments are not necessary for those individual patients. Additionally, the drug is not affected by inducers or inhibitors of CYP450 metabolism. There are currently no adequate trials of rivastigmine in pregnant women; however, animal trials have revealed no evidence of teratogenicity. It is unknown if it is excreted in breast milk.

A rivastigmine transdermal patch (5 to 20 cm^2) was developed and approved by the FDA in 2007. The 10-cm^2 patch pharmacokinetic profile is comparable to the oral 6-mg twice-daily dose, while the 20-cm^2 patch is similar to a 9- to 10-mg twice-daily oral dose. The patch showed a 30% lower C_{max}, a longer T_{max} (8 hours vs. 1 hour), and a 1.8-fold greater increase in AUC compared to the oral dose. Markedly less fluctuation in plasma drug concentrations between peak and trough levels was found with the patch. The 10-cm^2 patch showed adequate efficacy with fewer side effects than the 20-cm^2 patch and therefore was initially recommended as the primary target dose in mild to moderate AD and major neurocognitive disorder due to PD. A higher, 15-cm^2 transdermal patch formulation is currently available and is the FDA-recommended dose for patients with severe AD. Drug exposure is optimal when the rivastigmine patch is applied to the upper arm, chest, or back. These locations are the recommended application sites.

Pharmacodynamics. Rivastigmine is a noncompetitive, pseudo-irreversible cholinesterase inhibitor due to its prolonged enzyme inhibition activity, where a single dose was shown to produce inhibition for as long as 10 to 12 hours. Rivastigmine displays equal potency in inhibiting both AChE, and BuChE reported from both in vitro and in vivo data. This period of activity inhibition indicates that twice-a-day dosing is optimal. A linear relationship was reported between rivastigmine dose and AChE and BuChE inhibition measured in CSF. Plasma BChE inhibition rose slowly with the patch compared to two distinct peaks with the oral twice-daily dosing.

Design and Interpretation of Clinical Drug Studies

There have been a number of randomized, placebo-controlled, double-blind studies that have demonstrated the efficacy of rivastigmine in the treatment of AD. The majority of trials have been conducted on community-dwelling subjects with mild to moderate AD and were of 26-week duration or shorter. Doses ranged from 6 to 12 mg daily for oral doses and 5 to 20 cm^2 for the transdermal patch. The more efficacious oral dose was found to be 6 to 12 mg daily, with higher doses demonstrating increased side effects and worse tolerability. For the transdermal patch, doses ranged from 5 to 20 cm^2, with 10 cm^2 having better tolerability than higher doses with significant clinical outcomes. Moderate improvement in cognitive and functional scales occurred, as well as an impression of improvement as assessed by the subjects' clinician. These improvements demonstrated equivalent efficacy with donepezil and other cholinesterase inhibitors.

Specifically for the oral drug form, two placebo-controlled multicenter clinical studies with rivastigmine (N = 1,424) were conducted prior to FDA approval. Subjects were randomized into the placebo group, rivastigmine 1 to 4 mg/day, and rivastigmine 6 to 12 mg/day. Drug doses were adjusted weekly for the first 7 weeks as subjects reached the minimum assigned dose range. From weeks 8 to 26, subjects could be maintained on a flexible dose but in their assigned range to maximize therapeutic benefit and minimize side effects. Both studies showed a dose-dependent improvement effect on ADAS-cog scores from baseline to the study endpoint. The placebo group showed a decline of two to four points in the ADAS-cog, while the rivastigmine 6- to 12-mg/day group had about a one-point increase in ADAS-cog score. The rivastigmine 1- to 4-mg/day group had approximately a one- to two-point decline in ADAS-cog scores. Both rivastigmine groups showed significant benefit in the CIBIC-plus scores compared to placebo.

Additionally, the subjects in one of these large trials continued into an open-label extension study where both, the placebo and rivastigmine groups, were treated with rivastigmine 6 to 12 mg/day for an additional 26 weeks. In this study, both the placebo and rivastigmine groups showed improvement at the study endpoint compared to the projected rate of decline in untreated subjects. However, their scores were lower than the rivastigmine 6- to 12-mg/day group, with about a three- to four-point decrease in ADAS-cog scores at week 52. Subjects treated with rivastigmine 6 to 12 mg/day had almost maintained their ADAS-cog scores from baseline with an approximate two-point decrease after 52 weeks. These results imply that with a 6-month delay in treatment or inadequate treatment with low doses, subjects may not be able to demonstrate the same level of improvement as patients with therapeutic treatment from the start. For maximum benefit, when a diagnosis of AD is established, treatment with cholinesterase inhibitors such as rivastigmine, when indicated, should be started as soon as possible. Dosing at suboptimal levels for extended periods can also hinder drug efficacy.

In a 24-week, randomized, double-blind trial, patients with severe AD were randomized, receiving a 4.6-mg/24-hour patch versus a 13.3-mg/24-hour rivastigmine patch. There was an 8-week titration phase and a 16-week maintenance phase. As in the donepezil severe AD clinical trial, the SIB, as well as the ADCS-ADL-SIV, were used as clinical endpoints. The decline in both the SIB and the ASCS-ADL-SIV were both significantly less at 16 and 24 weeks for the 13.3-mg/24-hour dose.

In a 24-week, double-blind, placebo-controlled trial, rivastigmine with doses ranging from 3 to 12 mg was evaluated in 541 patients with PD dementia. The diagnosis was based on a 2-year or greater history of idiopathic PD before the onset of dementia, and patients did not meet the criteria for other etiologic causes of dementia. Based on ADAS-cog scores comparing baseline to week 24, rivastigmine-treated PD dementia patients had a mean improvement of 2.1 points, whereas the placebo group declined by a mean of 0.7 points. The Alzheimer’s Disease Cooperative Study Clinical Global Impressions of Change (ADCS-CGIC) scores also showed benefit for rivastigmine, as 19% had improved compared to 13% for placebo. Clinically meaningful worsening was reported to be only 13% for rivastigmine compared to a higher occurrence of 23% with placebo. Of note, the most frequent adverse side effects for rivastigmine were gastrointestinal, with nausea in 29% and vomiting in 16%, as well as tremors in 10% of patients. Worsening tremor may signify increased clinical relevance in PD dementia patients and warrants closer monitoring with rivastigmine treatment. Rivastigmine has also proved to be somewhat effective in LBD. In a randomized controlled study on 120 patients with LBD, patients receiving oral rivastigmine up to 12 mg daily were less apathetic and less anxious than the placebo-treated subjects, as 63% of patients in the treatment group showed a significant improvement in their symptoms measured by the Neuropsychiatric Inventory compared to 30% of control. The rivastigmine–treated group demonstrated a trend toward improvement in the Mini-Mental Status Examination (MMSE) scores, yet this trend was not found to be statistically significant.

Rivastigmine had little or no effect on cognition in VD, according to a recent Cochrane review. Rivastigmine transdermal did, however, improve or maintain cognitive status on ADAS-cog scores on 52 out of 100 patients with AD and small vessel cerebrovascular disease in a 24-week, open-label study.

Effects on Specific Organs and Systems

Rivastigmine does not produce hepatotoxicity or liver transaminase elevations, and therefore no monitoring of liver function tests are required with its use. Like other cholinesterase inhibitors, rivastigmine can produce vagotonic cardiovascular effects and should be used cautiously in patients with cardiac arrhythmias and cardiac conduction delays. Gastrointestinal problems can occur with all cholinesterase inhibitors but are particularly seen with rivastigmine use in a more dose-dependent manner. Rivastigmine can also increase gastrointestinal acid production. Patients on NSAIDs with peptic ulcer disease or other gastrointestinal illnesses should be carefully monitored, as they may be at greater risk of gastrointestinal bleeding.

Therapeutic Indications

Rivastigmine is indicated for the treatment of mild to moderate AD. It is also the only cholinesterase inhibitor with an FDA indication for mild to moderate dementia associated with PD. Similarly, rivastigmine has demonstrated benefits in LBD.

Precautions and Adverse Reactions

The most common adverse side effects of rivastigmine are typical of all cholinesterase inhibitors. These include diarrhea, nausea, vomiting, insomnia, muscle cramps, fatigue, and anorexia. These side effects are usually mild and transient, but in clinical trials, rivastigmine-treated groups had a higher dropout rate than placebo, usually due to GI side effects. Additionally, in a study comparing rivastigmine with donepezil in AD patients, oral rivastigmine-treated participants had a significantly higher dropout rate due to side effects. Gastrointestinal side effects with rivastigmine are not different when comparing mild to moderate individuals with those with moderate to severe AD. The transdermal patch formulation demonstrates much fewer gastrointestinal problems, most likely because of reduced plasma drug concentration, due to longer T_{max} and lower C_{max} when compared to oral preparation. Less commonly, rivastigmine can potentially cause adverse reactions such as seizures, cerebrovascular disease, or worsening COPD. It can also worsen urinary obstruction. Reanalysis of the electrocardiograph data from the clinical trial data revealed no significant differences between rivastigmine and placebo in heart rate, PR, QRS, and QTc interval. However, patients with cardiac conduction delays and significant heart disease should use rivastigmine with caution, given the potential for bradycardia with cholinesterase inhibitors. The transdermal patch formulation reported skin erythema barely visible or mild that self-resolved within 24 hours. The upper back site, arms, and chest had a lower incidence of erythema compared to the abdomen and thigh sites.

Drug Interactions

Because rivastigmine is not metabolized by hepatic oxidative CYP isozymes, it does not interact with the substrates of CYP450. Subsequently, there have been no drug interactions reported with digoxin, warfarin, diazepam, or fluoxetine. A retrospective analysis of the clinical trials did not find an increase in adverse side effects in patients who received various medications, including β-blockers, antihypertensives, and other drugs that may potentially produce pharmacodynamic interactions such as bradycardia. As with all cholinesterase inhibitors, caution is advised when patients are taking anticholinergic medications. Rivastigmine may also increase the action of succinylcholine-type muscle relaxants during anesthesia.

Laboratory Interferences

Rivastigmine has not been shown to interfere with any laboratory test, urinalysis, or spectrometry.

Dosage and Administration

Oral rivastigmine is available in 1.5-, 3.0-, 4.5-, and 6.0-mg capsules. Based on clinical trials, the suggested dosage range is 3 to 12 mg/day. Patients should be initiated at 1.5-mg twice-a-day and gradually increased every 2 to 4 weeks by 1.5 mg daily until the maximum daily dose of 12 mg/day is reached or titrated to the patient's tolerability. It is highly recommended to take food with rivastigmine to minimize side effects. The transdermal patch preparation (5 to 15 cm^2) is able to be applied once a day due to its pharmacokinetic and pharmacodynamic profile for cholinesterase inhibition, reported producing a slow and consistent release of the medication. The average exposure of transdermal patch dose of 10 cm^2 was comparable to an oral dose of 9.5 mg/24 hr. Patients should be initiated at a dose of 5 cm^2/4.6 mg daily and increased to 10 cm^2/9.5 mg in 4 weeks. Once this dose is maintained for 4 weeks and the patient tolerates it well, the maximum effective dose of 15 cm^2/13.3 mg/24 hour may be given. The slow release of rivastigmine from the patch should improve its gastrointestinal tolerability.

GALANTAMINE

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Acetylcholine
Mechanism of Action	Enzyme inhibitor (acetylcholinesterase), alpha-7 nicotinic receptor positive allosteric modulator

Galantamine was approved by the FDA for the treatment of AD in 2001 and is the fourth cholinesterase inhibitor made available in the United States. Galantamine is extracted from daffodils of the snowdrop plant *Galanthus nivalis* and is a phenanthrene alkaloid similar to codeine.

Chemistry

Galantamine is a (4aS,6R,8aS)–4a,5,9,10,11,12-hexahydro-3-methoxy-11-methyl-6H-benzofuro [3a,3,2–ef][2]benzazepin-6-ol hydrobromide compound. Its molecular structure is shown in Figure 33.11b–3.

Pharmacologic Actions

Pharmacokinetics. Galantamine immediate release (IR) has an 85% to 100% bioavailability in comparing tablets to the oral solution and is readily absorbed with a T_{max} from 0.5 to 2.0 hours. C_{max} is decreased by 25% with a slight delay of T_{max} by 1.5 hours with food intake. Coadministration with food does not affect galantamine IR AUC. The volume of distribution at a steady state was 2.6 L/kg with a very low protein binding of 18%. The elimination half-life of galantamine IR is about 7 hours. The bioequivalence of galantamine extended-release (ER) 24 mg was reported to be within the 90%

confidence intervals for AUC_{24} and C_{min} of galantamine IR 12-mg tablets twice daily. The mean time to peak plasma concentration was 4.4 hours, with an elimination half-life of 8.3 hours after fasting and 8.0 hours after food intake. Galantamine ER formulation does not differ in its pharmacokinetic profile, whether given it with food or while fasting.

Galantamine is mainly metabolized by hepatic CYP 2D6 and CYP 3A4 isozymes. Sanguinine (*O*-desmethylgalantamine) is formed by CYP 2D6 and accounts for 20% of the parent drug. Sanguinine is its major active metabolite and was shown to be three times more potent than galantamine in producing AChE inhibition. Other metabolites include galanthaminone and *N*-desmethylgalantamine, but they are much less potent than sanguinine. Although a 25% decrease in drug clearance of galantamine IR was found between poor metabolizers and extensive metabolizers of CYP 2D6, and a 50% higher drug exposure in galantamine ER, these findings are probably not significant due to the large interpatient variability in galantamine disposition. The doses can be titrated to individual tolerability.

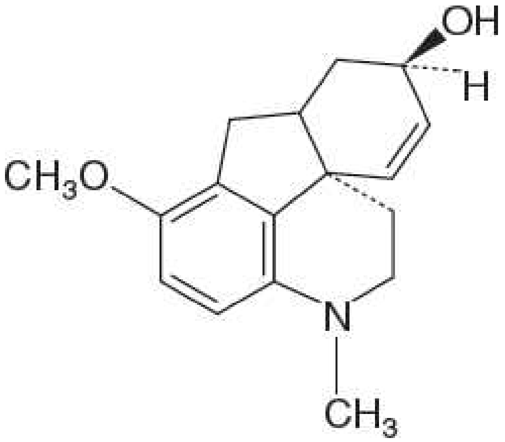

FIGURE 33.11b–3. Molecular structure of galantamine.

Age and mild hepatic impairment (defined as a Child–Pugh score of 5 to 6) were shown not to affect galantamine pharmacokinetics. Patients with moderate hepatic impairment (Child–Pugh score of 7 to 9) had galantamine AUC, and elimination half-life increased by 30%. Patients with mild renal impairment had galantamine concentrations similar to Alzheimer patients without renal impairment. No dosage adjustments are needed for galantamine in older adult patients or in patients with mild hepatic impairment. A maximum dose of 16 mg/day is recommended for patients with moderate hepatic impairment. Galantamine is contraindicated in patients with severe renal (creatinine clearance less than 9 mL/min) and severe hepatic impairment (Child–Pugh score greater than 9). Galantamine is classified as a category C drug with no trials in pregnant women and animal trials revealing potential teratogenicity. Additionally, it is unknown if it is excreted in breast milk.

Pharmacodynamics. Galantamine is a competitive, reversible inhibitor that is significantly more selective for AChE over BuChE based on in vitro data. Galantamine also produces allosteric modulation of nicotinic receptors that are distinct from acetylcholine and nicotinic agonists or antagonists. Allosteric modulators upregulate or potentiate ion channel activity of nicotinic receptors in response to acetylcholine. This mechanism of action may play a unique clinical role as AD is associated with a deficit in nicotinic transmission. By enhancing the release of acetylcholine, this action can additionally modulate the release of other neurotransmitters such as GABA, serotonin, and glutamate.

Design and Interpretation of Clinical Drug Studies

A number of large, multicenter clinical trials with galantamine have been conducted. 978 subjects were evaluated in the main clinical trial that established the minimum effective galantamine dose. After a 4-week single-blind placebo treatment, patients were randomized into four treatment groups: placebo, galantamine 8, 16, and 24 mg/day. Patients were initiated on 8 mg/day, and the dose was increased by 8 mg/day after 4 weeks until the target dose was achieved. The total study duration for the randomized groups was 5 months. All three groups showed improvement in ADAS-cog scores from baseline to the study endpoint when compared to placebo. During the intent-to-treat analysis, the 16- and 24-mg/day groups were found to be superior to placebo, but not the 8-mg group. The CIBIC-plus evaluations showed that 64% to 68% of the patients in the 16- and

24-mg/day groups remained stable or improved after 5 months versus 47% of the placebo group. The 8-mg/day group had only a 53% reported incidence of stability or improvement.

Several multicenter double-blind, placebo-controlled clinical trials have evaluated galantamine at higher doses of 18, 24, and 32 mg/day. Each study employed a different dosing strategy to reach these three dose groups, with a starting dose of 8 mg/day. Doses were either increased weekly or every 2 to 3 days until the assigned target dose was reached. The duration of the studies ranged from 12 weeks to 6 months. Each trial revealed that the 18- and 24-mg/day dose groups were superior to placebo. The 32-mg/day group also showed improvement compared to placebo but did not reveal increased benefit when compared to the 24-mg group. The incidence of gastrointestinal adverse effects was highest in the 32-mg/day group, thereby limiting its usefulness. The results of these studies suggested that the optimal dose of galantamine is 24 mg daily.

One of these clinical trials reported a 6-month extension where patients assigned to the placebo group were placed on galantamine 24 mg/day. After the subsequent 6 months, the placebo/galantamine group had ADAS-cog scores that resembled the galantamine 32-mg group, which had its dose decreased to the 24-mg/day dose group and above the projected rate of decline in untreated patients. The original 24-mg/day dose group at 6 months had ADAS-cog scores above baseline and at the 12-month time slightly decreased to baseline scores. The two other groups (placebo/galantamine and galantamine 32 mg or 24 mg) had ADAS-cog scores below the original 24-mg dose group at the 12-month time period. These results suggest that a 24-mg dose is the optimal galantamine dose and that a delay of 6 months by the placebo group could worsen the outcome in AD.

A 24-week double-blind, placebo-controlled study compared galantamine IR to galantamine ER in 971 patients with mild to moderate AD. Galantamine IR and ER were started at 8 mg/day and to 16, and then to 24-mg by 8-mg increments every 4 weeks as tolerated. The IR and ER were significantly better than a placebo in regard to cognition, measured by the ADAS-cog, but not in global change (CIBIC-Plus), where they did not differ from placebo. Although these findings with CIBIC-Plus differ from the original IR studies, the secondary ADCS-ADL measure favored ER over placebo.

A recent long-term study evaluated the effects of cholinesterase inhibitors on cognition and the progression to severe dementia. This longitudinal cohort study followed over 17,000 patients with AD, cholinesterase inhibitor-users and cholinesterase inhibitor-nonusers who were enrolled in the Swedish national dementia registry from 2007 to 2018. Cognitive trajectories were measured with MMSE. While each cholinesterase inhibitor showed an association with higher MMSE scores on follow-up, galantamine alone was found to reduce the risk of developing severe dementia (MMSE < 10). Cholinesterase inhibitors also reduced the overall mortality risk within the cohort in a dose-dependent manner, although galantamine at any dose was protective.

Effects on Specific Organs and Systems

Galantamine does not elevate hepatic transaminases and is not known to cause hepatotoxicity. Like the other cholinesterase inhibitors, this medication may cause vagotonic cardiovascular effects and should be used cautiously in patients with cardiac arrhythmias. Gastrointestinal problems may occur in a dose-dependent manner, and patients with peptic ulcer disease and other gastrointestinal problems should be carefully monitored.

Therapeutic Indication

Galantamine is indicated for the treatment of mild to moderate AD.

Precautions and Adverse Reactions

The most commonly reported side effects of galantamine are dizziness, headache, nausea, vomiting, diarrhea, anorexia, and weight loss. These side effects are mild and transient and did not produce a significant dropout rate during clinical studies. As previously stated, patients with renal and hepatic impairment may have substantially increased levels of galantamine. A maximum daily dose of 16 mg/day is the recommended dose for patients with moderate hepatic impairment. Galantamine is contraindicated in patients with severe renal and severe hepatic impairment. Less commonly, galantamine can potentially cause more severe reactions such as seizures, cerebrovascular disease, or exacerbation of COPD. The adverse side-effect profile of galantamine ER is similar to that of the IR formulation.

Drug Interactions

Galantamine is metabolized by hepatic CYP 2D6 and CYP 3A4 isozymes and could have significant drug–drug interactions. Galantamine AUC was reported to be increased by 40% when coadministered with paroxetine, 30% when given with ketoconazole, and 12% when erythromycin was coadministered. Patients should be monitored closely when known CYP 2D6 and CYP 3A4 inhibitors are given along with galantamine. As with all cholinesterase inhibitors, caution is advised when patients are taking anticholinergic medications, as galantamine may interfere with their activity due to its mechanism of action. It may also lengthen the action of succinylcholine-type muscle relaxants during anesthesia.

Laboratory Interferences

Galantamine has not been shown to interfere with any laboratory test, including urinalysis, spectrometry, and biochemical.

Dosage and Administration

Galantamine IR is available in 4-, 8-, and 16-mg tablets. The suggested dose range is 16 to 32 mg/day given twice-a-day in a divided dose. It should be given with food. The 32-mg dose is less well tolerated than lower doses, and clinical trial data do not indicate significant benefit. If patients tolerate the 24-mg dose, an increase to the 32-mg dose can be warranted. The initial dose is 8 mg/day, and after a minimum of 4 weeks, the dose can be increased. Further dose increases should be based on tolerability, with a minimum time period of 4 weeks on the previous dose. An oral solution is also available in a 4 mg/mL presentation. It should be taken as the IR form. Galantamine ER is available in 8-, 16-, and 24-mg tablets. The ER formulation should be taken with food with an initial dose of 8 mg daily in the morning and increasing the dose to 16 mg daily after a minimum of 4 weeks.

MEMANTINE

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Glutamate
Mechanism of Action	Receptor antagonist (NMDA)

Memantine is the only noncholinesterase inhibitor indicated for the treatment of AD. It was first synthesized in the early 1960s as a possible hypoglycemic agent but was found to be devoid of such activity. It was not until 2003 after multiple studies revealed its utility in dementia, that the FDA approved memantine for the treatment of moderate to severe AD. This agent is not indicated for mild to moderate AD, as are the cholinesterase inhibitors.

Chemistry

Memantine is a 1-amino-3,5-dimethyladamantane hydrochloride, a derivative of amantadine (Symmetrel) (1-amino-adamantane). Its molecular structure is shown in Figure 33.11b–4.

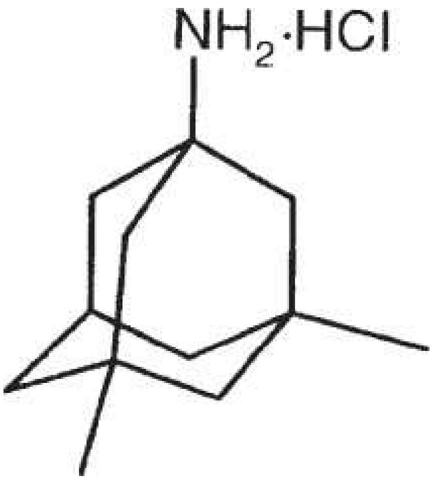

FIGURE 33.11b–4. Molecular structure of memantine.

Pharmacologic Actions

Pharmacokinetics. Memantine is absorbed completely from the gastrointestinal tract and achieves peak T_{max} within 3 to 7 hours. Protein binding is low, from 42% to 45%, and readily crosses the blood–brain barrier with a CSF serum concentration of 0.52. The elimination half-life of memantine is between 60 and 80 hours and is excreted unchanged in the urine. Steady-state conditions are reached in about 21 days. In a study comparing 28 mg once daily of memantine XR to 10 mg twice daily of memantine, C_{max}, and AUC0–24 values were 48% and 33% higher for the XR dosage regimen, respectively. Memantine is largely excreted unchanged in the urine, and the remaining is converted to three minimally active polar metabolites. Memantine is a weak base with a pKa of 10.27, and urinary pH can affect renal clearance. A reduced renal clearance occurs at alkaline urinary pH (pH 8) and results in increased plasma concentrations of the drug. Memantine has linear pharmacokinetics and, with doses of 10 to 20 mg/day, has produced minimal inhibitions of CYP enzymes, having no significant interactions with CYP inhibitors. Pharmacokinetic studies in healthy young adults and AD patients indicate that memantine can be safely coadministered with cholinesterase inhibitors. With repeated administration, total serum concentrations reach 0.5 μM, which corresponds to the Ki value of memantine at the MK-801 binding site in the human frontal cortex, indicating that NMDA receptor antagonism occurs at therapeutic concentrations. There are currently no adequate trials of memantine in pregnant women, and animal trials have revealed minimal evidence of teratogenicity. It is unknown if it is excreted in breast milk.

Pharmacodynamics. The pharmacologic effects of memantine are believed to result from the selective uncompetitive inhibitor blockade of the NMDA receptor class of glutamate receptors. Memantine does block the binding of glutamate to the receptor, but it blocks the ion current (Ca^{+2}, Na^{+}) through the receptor channel activated by glutamate. Memantine blocks the NMDA receptor with a strong voltage dependence and rapid blocking and unblocking kinetics, dissociating from the receptor in response to membrane depolarization. Blockade of the NMDA receptor channel during chronic, low-level excitation is thought to be a therapeutic benefit in neurodegenerative diseases such as AD.

Design and Interpretation of Clinical Drug Studies

Memantine indication for moderate to severe AD was based on two multicenter clinical trials. The first was a double-blind, placebo-controlled study comparing memantine 20 mg/day versus placebo in moderately to severely ill Alzheimer patients ($N = 252$). They had MMSE scores between 3 and 14 or a stage 5 or 6 on the Global Deterioration Scale (GDS). The study duration was 28 weeks, with an optional 24-week open-label extension phase. Patients receiving memantine had a significantly better outcome than placebo patients measured by the CIBIC-plus, ADCS-ADLsev, and SIB at 28 weeks. The second study was a double-blind, placebo-controlled trial, but it had a single-blind, placebo lead-in period of 1 to 2 weeks to assess compliance. Afterward, patients were randomized to drug or placebo for the next 24 weeks. The memantine dose was initiated at 5 mg/day, with weekly increases to 20 mg/day at the beginning of week 4. The primary outcome measures were the SIB, CIBIC-plus, and the ADCS-ADLsev at week 24. Patients were allowed to remain on stable donepezil doses of 5 to 10 mg/day throughout the trial if the medication was prescribed for at least 6 months prior to study entry.

The change in SIB, CIBIC-Plus, and ADCS-ADLsev outcome scores were all statistically significant in favor of memantine versus placebo. The SIB scores had a slight increase above baseline at week 12 and then remained above baseline by a mean of 1 point at 24 weeks. ADCS-ADLsev scores showed a gradual decline from baseline to 24 weeks for the placebo group, while the drug group displayed a slower rate of decline.

A 24-week open-label extension trial with moderate to severely ill patients reported continued benefit and a much slower decline in ADCS-ADL scores in those patients who were initially treated with memantine. Patients assigned to the placebo group were then placed on memantine. The newly treated patients showed less worsening of SIB, CIBIC-plus, and ADCS-ADL scores compared to the projected decline in untreated patients. At the end of 52 weeks, both groups had similar SIB, CIBIC-Plus, and ADCS-ADL scores and significantly higher scores than the projected scores for untreated patients. These findings indicate that memantine-treated patients continued to benefit from medication and had significantly delayed impairment compared to projections in untreated patients.

Although memantine is indicated only for moderate to severely ill patients, trials have been conducted on mild to moderate AD patients. These trials did not result in FDA indication for the mild stage of AD. A 24-week double-blind, placebo-controlled trial (N = 403) compared 20-mg memantine per day to placebo in mild to moderately ill AD patients, defined by MMSE scores of 10 to 22 points. The outcome measures included the ADAS-cog, CIBIC-plus, and ADCAS-ADL rating scales, comparing baseline to week 24. The study results indicated that only a modest treatment effect in favor of memantine had occurred. The mean ADAS-cog score for the placebo group increased by 1.1 points, while the memantine group score decreased by 0.8 points (P = .003). The mean end point CIBIC-plus scores for placebo and memantine groups were 4.52 and 4.20, respectively (P = .004). A significant difference was not found in ADCS-ADL scores between the two groups.

A more recent meta-analysis examined the efficacy and safety tolerability of memantine versus placebo and memantine +

cholinesterase inhibitors versus cholinesterase inhibitors alone in subjects with all levels of AD severity. As previously shown, memantine was superior to placebo in moderate to severe disease status for improving cognitive function but also demonstrated a reduction in behavioral disturbances as well. Memantine + cholinesterase inhibitors also had improved efficacy in behavioral disturbances over cholinesterase inhibitors alone. Memantine was well tolerated in both combination and monotherapy.

Effects on Specific Organs and Systems

Memantine does not cause an increase in hepatic transaminases or hepatotoxicity. Memantine XR terminal elimination half-life increased by 16% in patients with moderately impaired liver function compared to controls. It is recommended that patients with severe hepatic impairment take memantine ER with extreme caution. Memantine is mainly excreted through the kidneys. Even though there is scarce data about renally impaired patients taking memantine, drug doses should be reduced in patients with mild to moderate renal impairment. For patients with severe renal impairment, memantine is not recommended.

Therapeutic Indications

Memantine is indicated for the treatment of moderate to severe AD.

Precautions and Adverse Reactions

The most common side effects observed with memantine are constipation, dizziness, headache, and confusion. These effects are usually mild and transient and did not result in a significant dropout rate during clinical trials. With the use of memantine ER, the most common side effects are headache, diarrhea, and dizziness.

Drug Interactions

The use of memantine along with other NMDA antagonists such as amantadine and dextromethorphan has not been evaluated, and this combination should be used with caution. In vitro studies with CYP substrates 1A2, 3A4, 2A6, 2C9, 2D6, 2E1, and 3A4 showed minimal inhibition by memantine. Therefore, no pharmacokinetic interaction by these enzymes should occur. Using memantine with donepezil did not affect the pharmacokinetics of either compound. The adverse side-effect profile of the combination of memantine and donepezil was similar to that of patients treated with donepezil alone. Memantine is eliminated by renal tubular secretion, and if administered with other drugs that are excreted through the same mechanism, the plasma concentration of both agents could be altered. Examples of these drugs include hydrochlorothiazide, triamterene, cimetidine, ranitidine, quinidine, and nicotine. Concomitant use of memantine, hydrochlorothiazide, and triamterene did not affect the disposition of memantine or triamterene, but the bioavailability of hydrochlorothiazide was decreased by 20%. Memantine clearance was reduced by 80% with alkaline urine conditions at pH 8. If the urine pH is altered, drug accumulation may occur. This condition could be caused by diet, drugs like carbonic anhydrase inhibitors and sodium bicarbonate, and medical conditions such as renal tubular acidosis. Memantine should be cautiously used in this situation.

Laboratory Interferences

Memantine has not been shown to interfere with any laboratory test, including biochemical, urinalysis, and hematology. No clinically significant electrocardiographic changes were noted during clinical trials.

Dosage and Administration

Memantine is available in 5- and 10-mg tablets, with a recommended starting dose of 5 mg daily. The recommended target dose is 20 mg/day. The drug is administered twice daily in separate doses with 5-mg increment increases weekly depending on tolerability. Memantine ER is available in 7-, 14-, 21-, and 28-mg capsules. The starting dose is 7 mg daily, with a target dose of 28 mg daily. The ER formulation should be increased to no greater than 7 mg weekly. Patients with renal impairment should have a target dose of 14 mg daily. A drug is now available as a combination of memantine XR and donepezil (Namzaric). It is available in 14-mg/10-mg and 28-mg/10-mg formulations. It is only indicated for patients with severe AD already stabilized on memantine ER 14 or 28 mg and donepezil 10 mg.

FUTURE DIRECTIONS

In more recent years, drugs have been actively developed with different pharmacodynamic mechanisms to modify and delay the onset of AD. Multiple treatment trials with the goal of slowing or even reversing the course of cognitive and functional decline among AD subjects continue. They include secretase inhibitors, amyloid-β inhibitors, tau and amyloid-β-aggregation inhibitors, as well as vaccines against amyloid-β and tau. These potentially disease-modifying therapies focus mainly on agents that will decrease amyloid-β content and tau hyperphosphorylation. To date, trials have not demonstrated any significant improvement in the decline of patients already suffering the full clinical picture of AD. Current research has shifted to better recognition of AD in its earliest clinical phases. The aim is to enlist those at-risk populations with the mildest symptoms, such as those with MCI or even those still clinically asymptomatic, such as preclinical AD, into active clinical trials to potentially reverse or even prevent AD as well as other neurodegenerative disorders.

Aducanumab injection is the first amyloid beta-directed antibody approved by the FDA in 2021 for the overall treatment of AD. The indication proceeded under accelerated approval based on a reduction in amyloid-beta plaques in the brains of study participants; however, the clinical trials were limited to patients with very early disease, and clinically significant improvement was not robust. The continued approval for this very broad indication may be altered upon verification of clinical benefit in future confirmatory Phase 4 trials.

FURTHER READINGS

Auchus AP, Brashear HR, Salloway S, et al. Galantamine treatment of vascular dementia: a randomized trial. *Neurology*. 2007;69(5):448–458.

*Battle CE, Abdul-Rahim AH, Shenkin SD, Hewitt J, Quinn TJ. Cholinesterase inhibitors for vascular dementia and other vascular cognitive impairments: a network meta-analysis. *Cochrane Database Syst Rev*. 2021;2:CD013306.

Birks JS, Harvey RJ. Donepezil for dementia due to Alzheimer's disease. *Cochrane Database Syst Rev.* 2018;6(6):CD001190.

Black SE, Doody R, Li H, et al. Donepezil preserves cognition and global function in patients with severe Alzheimer disease. *Neurology*. 2007;69(5):459–469.

Boinpally R, Chen L, Zukin SR, McClure N, Hofbauer RK, Periclou A. A novel once-daily fixed-dose combination of memantine extended release and donepezil for the treatment of moderate to severe Alzheimer's disease: two phase I studies in healthy volunteers. *Clin Drug Investig*. 2015;35(7):427–435.

Brodaty H, Corey-Bloom J, Potocnik FC, Truyen L, Gold M, Damaraju CR. Galantamine prolonged-release formulation in the treatment of mild to moderate Alzheimer's disease. *Dement Geriatr Cogn Disord*. 2005;20:120–132.

Corey-Bloom J, Anand R, Veach J; ENA 713 B352 Study Group. A randomized trial evaluating the efficacy and safety of ENA 713 (rivastigmine tartrate), a new acetylcholinesterase inhibitor, in patients with mild to moderately severe Alzheimer's disease. *Int J Ger Psychopharmacol*. 1998;1:55–65.

Cummings J, Lefevre G, Small G, Appel-Dingemanse S. Pharmacokinetic rationale for rivastigmine patch. *Neurology*. 2007;69(4 Suppl 1):S10–S13.

Emre M, Aarsland D, Albanese A, et al. Rivastigmine for dementia associated with Parkinson's disease. *N Engl J Med*. 2004;351(24):2509–2518.

Grossberg GT, Manes F, Allegri RF, et al. The safety, tolerability, and efficacy of once-daily memantine (28 mg): a multinational, randomized, double-blind, placebo-controlled trial in patients with moderate-to-severe Alzheimer's disease taking cholinesterase inhibitors. *CNS Drugs*. 2013;27(6):469–478.

Kavirajan H, Schneider LS. Efficacy and adverse effects of cholinesterase inhibitors and memantine in vascular dementia: a meta-analysis of randomized controlled trials. *Lancet Neurol*. 2007;6(9):782–792.

*Kishi T, Matsunaga S, Oya K, Nomura I, Ikuta T, Iwata N. Memantine for Alzheimer's disease: an updated systematic review and meta-analysis. *J Alzheimers Dis*. 2017;60(2):401–425.

Lefèvre G, Sedek G, Jhee SS, et al. Pharmacokinetics and pharmacodynamics of the novel daily rivastigmine transdermal patch compared with twice-daily capsules in Alzheimer's disease patients. *Clin Pharmacol Ther*. 2008;83(1):106–114.

Matsunaga S, Kishi T, Iwata N. Combination therapy with cholinesterase inhibitors and memantine for Alzheimer's disease: a systematic review and meta-analysis. *Int J Neuropsychopharmacol*. 2014;18(5):pyu115.

McKeith I, Del Ser T, Spano P, et al. Efficacy of rivastigmine in dementia with Lewy bodies: a randomised, double-blind, placebo-controlled international study.

Lancet. 2000;356(9247):2031–2036.

Miranda LF, Gomes KB, Tito PA, et al. Clinical response to donepezil in mild and moderate dementia: relationship to drug plasma concentration and CYP2D6 and APOE genetic polymorphisms. *J Alzheimers Dis*. 2017;55(2):539–549.

Panek D, Wieckowski A, Wichur T, et al. Design, synthesis and biological evaluation of new phthalimide and saccharin derivatives with alicyclic amines targeting cholinesterases, beta-secretase and amyloid beta aggregation. *Eur J Med Chem*. 2017;125:676–695.

Qassem A, Snow V, Cross JT Jr, et al. Current pharmacologic treatment of dementia: a clinical practice guideline from the American College of Physicians and the American Academy of Family Physicians. *Ann Intern Med*. 2008;148(5):370–378.

Raskind MA, Peskind ER, Wessel T, Yuan W. Galantamine in AD. A 6-month randomized placebo controlled trial with a 6-month extension. The Galantamine USA-1 Study Group. *Neurology*. 2000;54(12):2261–2268.

Reisberg B, Doody R, Stöffler A, et al. Memantine in moderate-to-severe Alzheimer’s disease. *N Engl J Med*. 2003;348(14):1333–1341.

Rogers SL, Farlow MR, Doody RS, Mohs R, Friedhoff LT. A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer’s disease. *Neurology*. 1998;50(1):136–145.

*Rolinski M, Fox C, Maidment I, McShane R. Cholinesterase inhibitors for dementia with Lewy bodies, Parkinson’s disease dementia and cognitive impairment in Parkinson’s disease. *Cochrane Database Syst Rev*. 2012;3:CD006504.

*Russ TC, Morling JR. Cholinesterase inhibitors for mild cognitive impairment. *Cochrane Database Syst Rev*. 2012;9:CD009132.

Tariot PN, Solomon PR, Morris JC, Kershaw P, Lilienfeld S, Ding C. A 5-month, randomized placebo-controlled trial of galantamine in AD. *Neurology*. 2000;54(12):2269–2276.

The Alzheimer’s Association. 2021 Alzheimer’s disease facts and figures. *Alzheimer’s Dement*. 2021;17(3):327-406.

Winblad B, Grossberg G, Frölich L, et al. IDEAL: a 6-month, double-blind, placebo-controlled study of the first skin patch for Alzheimer disease. *Neurology*. 2007;69(4 Suppl 1):S14–S22.

Winblad B, Kilander L, Eriksson S, et al. Donepezil in patients with severe Alzheimer’s disease: double-blind, parallel group, placebo-controlled study. *Lancet*. 2006;367(9516):1057–1065.

World Health Organization. Global action plan on the public health response to dementia 2017–2025. Accessed 29 July, 2021. https://apps.who.int/iris/bitstream/handle/10665/259615/9789241513487-eng.pdf;jsessionid=4DA480FA93471AC53988E52B35F416D8?sequence=1

Xu H, Garcia-Ptacek S, Jönsson L, Anders W, Nordström P, Eriksdotter M. Long term effects of cholinesterase inhibitors on cognitive decline and mortality. *Neurology*. 2021;96(17):e2220–e2230.

33.12 Histamine Modulators

33.12a Histamine Blockers and Enhancers

Sasson Zemach, M.D., and Abraham Weizman, M.D.

HISTORY

Histamine is a phylogenetically old amine, widely distributed throughout the central nervous system (CNS) of many species. The role of histamine in immunomodulation and smooth muscle relaxation was originally described in the early 1900s. Charles Best and colleagues were the first to isolate histamine (H substance) in 1927. The antihistamines were first discovered by the Swiss-born Italian pharmacologist Daniel Bovet in 1937. The first nontoxic histamine-blocking drug, pyrilamine (Poly-Histine-D), was developed in 1944. In the 1950s, diphenhydramine (Benadryl) and tripelennamine (Pyribenzamine) were introduced as highly effective histamine antagonists. The sedative and anticholinergic activity of antihistamines constituted the basis for their frequent prescription for a panoply of psychiatric symptoms. The more recent nonsedating antihistamines are less frequently prescribed in psychiatric practice.

Table 33.12a–1.
Pharmacokinetics of Selected H_1 Receptor Antagonists Used in Psychiatry[a]

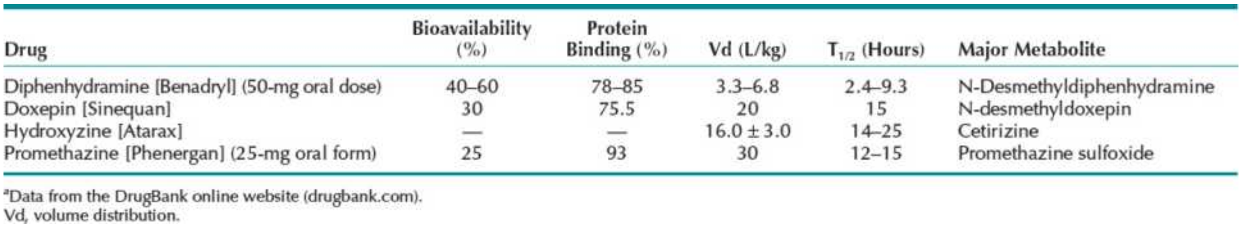

Drug	Bioavailability (%)	Protein Binding (%)	Vd (L/kg)	$T_{1/2}$ (Hours)	Major Metabolite
Diphenhydramine [Benadryl] (50-mg oral dose)	40–60	78–85	3.3–6.8	2.4–9.3	N-Desmethyldiphenhydramine
Doxepin [Sinequan]	30	75.5	20	15	N-desmethyldoxepin
Hydroxyzine [Atarax]	—	—	16.0 ± 3.0	14–25	Cetirizine
Promethazine [Phenergan] (25-mg oral form)	25	93	30	12–15	Promethazine sulfoxide

[a] Data from the DrugBank online website (drugbank.com).
Vd, volume distribution.

CHEMISTRY

With approximately 50 H_1 receptor (H_1R) antagonists currently available worldwide, antihistamines are among the most widely utilized class of agents. Although the antihistamines are structurally unrelated, they share a common ethylamine moiety. The typology, first- and second-generation antihistamines refer to sedating and nonsedating agents, respectively. More recently, nonsedating antihistamines (i.e., substrates of P-glycoprotein) with less arrhythmogenic potential (e.g., fexofenadine [Allegra], loratadine [Claritin], desloratadine [Clarinex], and cetirizine [Zyrtec]) have been referred to as third-generation antihistamines.

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Some general properties of important histamine antagonists commonly used in psychiatry are shown in Table 33.12a–1.

Absorption

The first-generation H_1R antagonists are readily absorbed via the gastrointestinal (GI) tract. Most agents achieve maximum plasma levels within 1 to 3 hours (T_{max}).

Distribution

Antihistamine binding of plasma protein varies from approximately 50% to 98%. The H_1R antagonists' volume of distribution varies more than 100-fold from 0.33 L/kg (i.e., levocetirizine [Xyzal]) to greater than 100 L/kg (i.e., loratadine [Claritin]). Sedating antihistamines are widely distributed in the body and CNS. The antihistamine CNS adverse event profile is a consequence of its relative ability to traverse the blood–brain barrier (BBB). Factors that influence BBB permeability are lipophilicity, molecular weight, and recognition by the P-glycoprotein efflux pump expressed on luminal surfaces of nonfenestrated endothelial cells in the vascular wall of the CNS. Radioligand studies utilizing carbon-11 doxepin (Sinequan) indicate that first-generation H_1R antagonists occupy approximately 50% to 90% of the H_1Rs in the frontal and temporal cortices, hippocampus, and pons. Second-generation H_1R antagonists occupy approximately 0% to 30% of CNS H_1Rs, largely due to their minimal lipophilicity and affinity for P-glycoproteins.

METABOLISM

Most antihistamines are subject to phase I metabolism via hepatic cytochrome P450 systems. In vitro studies evaluating human liver microsomes transfected with cytochrome P450 2D6 (CYP2D6) complementary deoxyribonucleic acid (cDNA) indicate that several antihistamines (e.g., diphenhydramine, chlorpheniramine, clemastine [Dayhist-1], perphenazine [Trilafon], hydroxyzine, tripelennamine) inhibit the metabolism and are hepatically biotransformed by CYP2D6 and CYP3A4. Second- and third-generation antihistamines are neither potent inhibitors nor inducers of the cytochrome P450 system. The H_1R antagonist terminal elimination half-life ($t_{1/2}\beta$) ranges from approximately 2 to 27 hours. Pediatric populations rapidly metabolize antihistamines; individuals who are slow metabolizers may exhibit drug accumulation.

ELIMINATION

Most antihistamines are excreted largely unchanged in the urine and feces.

PHARMACODYNAMICS

Histamine is a pleiotropic, low–molecular-weight amine synthesized endogenously in diverse cell types by the enzyme histidine decarboxylase from the semiessential amino acid L-histidine. Histamine is produced in the CNS, in neurons of the tuberomammillary nucleus of the posterior hypothalamus with projections to both cortical and subcortical regions. Although no precise topographical organization of the histaminergic projection field has been elucidated, the cerebral cortex, amygdala, substantia nigra, and striatum receive moderate to dense histaminergic innervation. Innervations to the hippocampus, thalamus, retina, and spinal cord vary. Histaminergic networks interact and modulate classical monoamine networks. Afferent projection to tuberomammillary neurons includes the infralimbic cortex, lateral septum, preoptic nucleus, and brainstem (from the adrenergic cell group C1 to C3, the noradrenergic group A1 to A3, the ventral tegmental area [VTA] and the serotonergic group B5 to B9).

In addition to histamine, the tuberomammillary neurons also contain the gamma-aminobutyric acid (GABA)-synthesizing enzyme —glutamic acid decarboxylase (GAD), as well as galanin, thyrotropin-releasing hormone, proenkephalin-derived peptides, substance P, and adenosine deaminase. The CNS histamine pool is also comprised of mast cells, glial cells, and vascular endothelium, which are activated by proinflammatory agents.

To date, there are four known histamine receptors, all of which are G-protein–coupled receptors (GPCR): The H_1Rs and H_2Rs exert mostly excitatory actions on neurons, while the H_3Rs are high-affinity auto- and heteroreceptors, inhibiting tuberomammillary neurons and reducing neurotransmitter release. H_4Rs are localized in immunocompetent cells (mast cells, dendritic cells, T cells) and are involved in the release of chemokines and cytokines. H_1Rs participate in synaptic plasticity and long-term potentiation (LTP) via their effect on intracellular Ca^{2+} and protein kinase C (PKC). H_2R activation potently stimulates the cyclic adenosine monophosphate (cAMP) pathway, a critical substrate for late-phase N-methyl-D-aspartate (NMDA) receptor-dependent LTP. H_2R activation also participates in the induction of mossy-fiber LTP. High histamine concentration in the CA1 region of the hippocampus facilitates the induction of LTP in the presence of H_1R and H_2R antagonists, suggesting direct action at the NMDA receptors.

Histamine plays a critical role in CNS arousal. Lesioning of the posterior hypothalamus results in hypersomnia and decreased psychomotor activity. Histamine also regulates cortical activity by stimulating cholinergic neurons of the basal forebrain and serotonergic neurons of the dorsal raphe nucleus. The wakefulness-promoting effects of orexins may be mediated by increased histamine release as inactivation mutations of the H_1R decrease the orexin-mediated arousal effect. Histamine is also involved in vasopressor effects, temperature regulation, and analgesia.

In addition, histamine regulates energy homeostasis via effects on leptin networks. Central administration of leptin increases histamine turnover in the hypothalamus. Leptin-induced suppression of food intake and upregulation of uncoupling protein 1 expression in brown adipose tissue is attenuated in H_1R knockout (KO) mice. H_1R as well as proopiomelanocortin (POMC) melanocortin 4 receptor (MC4-R) regulate energy homeostasis in the brain. The effect of histamine on drinking behavior and osmotic homeostasis is mediated by its stimulatory effect on vasopressin-producing neurons in the supraoptic nucleus. Histamine also regulates neuroendocrine systems (e.g., prolactin, growth hormone, adrenocorticotropin, and thyrotropin-releasing hormone).

Unlike other aminergic systems, histamine and the levels of its metabolites increase with age in the cerebrospinal fluid (CSF). In Alzheimer disease, neurofibrillary tangles occur prominently in the tuberomammillary nucleus, with a corresponding decrease in the number of tuberomammillary neurons. Histamine levels and histidine decarboxylase activity are diminished in brain regions affected by Alzheimer disease and Down syndrome. In Parkinson patients, histamine levels are selectively increased in the putamen, substantia nigra, and external globus pallidus. Levels of tele-methylhistamine (t-MeHA), the primary metabolite of histamine, are unchanged in the substantia nigra. In contrast, t-MeHA is elevated in the CSF of individuals with schizophrenia, particularly those with prominent negative symptoms. These observations indicate that increased histamine release and turnover may be pathophysiologic features in schizophrenia, providing the basis for evaluating the efficacy of H_2R antagonists. Interestingly, results from relatively small placebo-controlled clinical trials in schizophrenia indicate that adjunctive famotidine to ongoing antipsychotic treatment was superior to placebo in reducing overall objectively measured negative symptom severity. In addition, global psychopathology in famotidine-treated subjects may exhibit significantly greater reduction when compared to placebo treatment. However, a meta-analysis of studies with three H_2R antagonists did not demonstrate a significant difference from placebo.

During the past several decades, the human H_1Rs, H_2Rs, H_3Rs, and H_4Rs have been cloned and characterized, each differing in structure, function, localization, and pharmacology. Immunocytochemical studies have isolated H_3R throughout the CNS, and it was found to be a heptahelical protein that exists in both active and inactive forms in dynamic equilibrium. Several isoforms of each of the histamine receptors have been identified with reverse transcription polymerase chain reaction. The four major histamine receptors transduce extracellular signals via G proteins to intracellular second-messenger systems. Histamine receptors are active, and, following stimulation with specific agonists, each one independently is capable of signaling, both in vitro and in vivo.

The human H_1Rs have approximately 45% homology with muscarinic receptors. Mutations of the H_1R gene associated with inactivation result in deficits in learning, memory, and locomotor activity, as well as in impaired immunologic activity. The H_1Rs also mediate the increased permeability of small vessels associated with allergic responses. Extravascular smooth muscle (e.g., bronchiole smooth muscle) contraction is also mediated by H_1R. H_3Rs are presynaptic autoreceptors on histamine-containing neurons; they mediate the synthesis, release, and inhibition of histamine. They are also involved in modulating the release of several other monoamines and neuropeptides.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

A substantial body of evidence has established the efficacy of H_1R antagonists for allergy-related conditions as well as motion sickness. There are no recent studies that establish the efficacy of H_1R antagonists for any of the mental disorders defined in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5). These antagonists are frequently prescribed to achieve ataraxia and to reduce psychomotor agitation and sleep latency. The anticholinergic effects of H_1R antagonists have provided the basis for prescribing these agents in Parkinson disease and for dopamine blocker–induced movement disorders. Data on the use of H_2R antagonists in the treatment of Alzheimer disease, medication-associated weight loss, anxiety disorders, and autism are equivocal. H_3R antagonists/inverse agonists have been shown to have some beneficial effects in narcolepsy.

The involvement of histamine receptors both in immune activity regulation and monoamine release raises the possibility of it playing a role in the pathophysiology of depression. It was suggested that the interaction between the histaminergic system that affects cytokine release (neuroinflammation) and the monoamine release is relevant to the pathophysiology of depression. However, these suggestions have not been studied as yet.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

The anxiolytic and sedative effects of antihistamines are related to their direct antagonism at the H_1Rs. In some cases, paradoxical excitation characterized by restlessness, agitation, and nervousness has also been reported. The stimulating effects of H_1R agonists have been particularly apparent in tasks that require a high level of attention (e.g., divided attention tasks, vigilance, and driving tasks). In some cases, however, there are dose-dependent stimulating effects of some antihistamines. The precise biochemical mechanism subserving the excitatory effects of antihistamines has not been elucidated; it is hypothesized that neurotransmitters such as dopamine and GABA are salient or that the effect is mediated by the H_3R.

Antihistamines also have prominent autonomic nervous system effects. They exert a drying effect on mucous membranes in the respiratory system, and their effects on smooth muscles result in bronchodilation and the inhibition of vasoconstriction. H_1R antagonists do not significantly affect blood pressure, but parenteral administration of antihistamines may result in a transient lowering of blood pressure.

COMMONLY USED HISTAMINERGIC AGENTS IN PSYCHIATRY

Diphenhydramine

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Histamine
Mechanism of Action	Receptor antagonist (H1)

Oral diphenhydramine is efficacious in the treatment of dopamine blocker–induced movement disorders. Diphenhydramine is approved by the U.S. Food and Drug Administration (FDA) for the treatment of dopamine blocker–induced parkinsonism, allergic and anaphylactic reactions, as well as motion sickness. When prescribed for parkinsonism, diphenhydramine is often combined with anticholinergic agents. The parenteral availability of diphenhydramine presents a viable treatment option for Parkinson patients with difficulty swallowing and for acute dystonic reactions. It has historically been prescribed as a nonspecific sedative for children and older adults. Reports of paradoxical excitation, as well as cognitive impairment (e.g., anticholinergic effect), reduce its therapeutic index.

Hydroxyzine

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Histamine
Mechanism of Action	Receptor antagonist (H1)

Hydroxyzine is indicated for the short-term treatment of anxiety and tension due to psychological factors. A Cochrane review of 39 studies found the efficacy of hydroxyzine in the treatment of generalized anxiety disorder (GAD) to be similar to that of benzodiazepines and buspirone. However, it was associated with a larger rate of sleepiness/drowsiness. In addition, hydroxyzine trials were associated with significant study bias. Preliminary evidence suggests that hydroxyzine may be superior to placebo at reducing the overall severity of bruxism in children. It is often included in numerous over-the-counter preparations for the treatment of anxiety or insomnia. Hydroxyzine is also indicated for use as an antipruritic and as a sedative as part of a premedication regimen in conjunction with general anesthesia.

Promethazine

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Histamine, dopamine
Mechanism of Action	Receptor antagonist (H1, D2)

Promethazine is indicated before or after surgery as an antiemetic, as well as an adjuvant for analgesia. It is also indicated for sedation in both pediatric and adult populations, and similar to diphenhydramine, it is also used for allergic reactions and to treat motion sickness. In 2016, the British National Institute for Health and Care Excellence (NICE) approved the use of the haloperidol and promethazine combination as a treatment for acute aggression in patients with psychosis. It was suggested that the synergy between the sedative effect and the anticholinergic properties of promethazine that attenuate haloperidol-induced extrapyramidal side effects results in aggression reduction.

Cyproheptadine

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Histamine, serotonin
Mechanism of Action	Receptor antagonist (H1, H2, 5HT2A)

Cyproheptadine is indicated for allergic reactions. It is a potent H_1R and H_2R antagonist, and is also a serotonin (5-HT) 2AR antagonist. Cyproheptadine is used for various purposes in psychiatry (e.g., to counter antidepressant-associated sexual dysfunction, stimulant-induced weight loss, or insomnia) and for the symptomatic treatment of schizophrenia (e.g., agitation, akathisia), autism spectrum disorder (ASD), serotonin syndrome, or migraine.

OTHER THERAPEUTIC INDICATIONS IN PSYCHIATRY

H_1R antagonists are widely prescribed in the treatment of insomnia (including during pregnancy), analgesia, akathisia, serotonin syndrome, anxiety, motion sickness, vertigo, emesis, and bruxism.

Insomnia and Sedation

Although benzodiazepines and other classes of psychotropic agents have largely replaced antihistamines in clinical practice, H_1R antagonists are still considered effective for sedation and sleep promotion. The rationale for considering antihistamines is supported by evidence that histaminergic stimulation is critical for wakefulness.

It is of note that agents with H_1R antagonistic activity at low doses, like the monoamine transmission modulators/enhancers mirtazapine (Remeron), doxepin (Sinequan), and amitriptyline (Elavil), as well as the dopamine blocker quetiapine (Seroquel), are effective for insomnia. Similar to diphenhydramine, doxepin has also been reported to induce non-REM sleep, while mirtazapine at a low dosage is also effective for the treatment of dopamine blocker–induced akathisia.

Narcolepsy

Narcolepsy is a chronic disorder characterized by excessive daytime sleepiness (EDS), which leads to difficulties in concentration. It was proposed that narcolepsy is related to the denervation of hypothalamic hypocretin (orexin)-producing neurons. Dopaminergic stimulators (mainly modafinil and armodafinil) are still the first choice of treatment. Pitolisant (Wakix), a selective H_3 autoreceptor blocker, was approved for the treatment of narcolepsy by the european medicines agency (EMA) in 2016 and by the FDA in 2020. Clinical trials demonstrated significant improvements in EDS and cataplexy, with adverse events including headaches, insomnia, and nausea. Serious safety concerns were not detected. Another study comparing pitolisant to placebo and modafinil found no significant differences between the beneficial effects of the two active agents. However, post hoc analyses did not show pitolisant to be noninferior to modafinil, possibly due to the small sample sizes ($N =$ 30 in each group). Hence, pitolisant may be an optional therapeutic strategy when stimulants fail.

FUTURE DIRECTIONS

Attenuation of Iatrogenic Weight Gain

H_2R antagonist use in psychiatric practice has primarily been for weight loss in overweight individuals and persons experiencing dopamine blocker–associated weight gain. Initial studies indicated that the H_2-receptor antagonist nizatidine (Axid) was efficacious as a secondary prevention strategy for olanzapine (Zyprexa)-associated weight gain. Subsequent placebo-controlled studies with this agent as a primary or secondary prevention strategy for olanzapine- and quetiapine-associated weight gain have produced more equivocal results. The mechanism mediating nizatidine’s modest weight loss effect is hypothesized to be via central appetite–regulating centers.

Aside from H_2R antagonists, other histaminergic agents were also investigated for the prevention of dopamine blocker–induced weight gain. A double-blind, randomized placebo-controlled trial assessed the effects of a combination of betahistine (H_1R agonist/H_3R antagonist) and reboxetine (a norepinephrine reuptake inhibitor previously demonstrated as having a positive effect on weight gain) versus placebo among olanzapine-treated patients. At the end of the trial, patients in the betahistine/reboxetine group gained significantly less weight than those in the placebo group (2.02 ± 2.37 and 4.77 ± 3.16 kg, respectively). Nonetheless, it is of note that metformin may be superior to betahistine in the treatment of weight gain induced by antipsychotic dopamine-blocking agents.

Neurocognitive Disorders

Data regarding the role of histamine in Alzheimer disease is inconsistent. Yet, ecologic studies suggested that H_2R antagonists may decrease the emergence of Alzheimer disease. Subsequent randomized, double-blind, placebo-controlled trials with nizatidine, however, failed to demonstrate a therapeutic effect in the treatment of cognitive deterioration that is part of Alzheimer disease. Notwithstanding, several lines of preclinical studies suggest that antihistaminergic therapy may facilitate the reduction in amyloid-beta neuropathology and prevention of autophagic failure. Although, to date, there is no clear-cut evidence for procognitive effects of H_3R antagonists, the approval of pitolisant for narcolepsy and its increasing use may offer an opportunity to collect new data on its cognitive effect as well.

PRECAUTIONS AND ADVERSE EVENTS

The most common side effect of first-generation H_1R antagonists is CNS depression, manifesting as sedation, somnolence, cognitive impairment, and reduction in psychomotor activity. Simple and complex tests of psychomotor performance indicate that second-generation antihistamines are less impairing than first-generation antihistamines. Although second- and third-generation antihistamines are less sedating when coadministered with other agents that alter P-glycoprotein function, their propensity for CNS adverse events increases.

The coadministration of alcohol, benzodiazepines, or other CNS depressants may result in additive sedation, requiring dose adjustment. Other reported CNS side effects include blurred vision, nervousness, and tremors. Other non-CNS side effects associated with antihistamines are dry mouth, increased appetite, nausea, vomiting, diarrhea, or constipation. H_1R antagonists should be administered with caution in patients treated with anticholinergics since the combination may aggravate side effects such as dry mouth, blurred vision, and bladder neck obstruction, especially in patients with an enlarged prostate.

The use of antihistamines in older adults could result in confusion and disorientation. Pediatric populations and individuals overdosing with H_1R antagonists are at risk for excitatory effects, which include ataxia, agitation, incoordination, and seizures. Other adverse events resulting from overdose resemble atropine poisoning and include flushed face, dilated pupils, tachycardia, and fever. Treatment of overdose is generally through supportive medical therapy.

Table 33.12a–2.

Dosage and Administration of Common Histamine Antagonists

Medication	**Route**	**Preparation**	**Common Dosage**
Diphenhydramine (Benadryl)	Oral	Capsules and tablets: 25 mg, 50 mg	Adults: 25–50 mg three to four times per day
		Liquid: 12.5 mg/5.0 mL	Children: 5 mg/kg three to four times per day, not to exceed 300 mg/day
	Deep IM or IV	Solution: 10 or 50 mg/mL	Same as oral
Hydroxyzine Hydrochloride (Atarax)	Oral	Tablets: 10, 25, 50, and 100 mg	Adults: 50–100 mg three to four times daily
		Syrup: 10 mg/5 mL	Children younger than 6 years of age: 2 mg/kg per day in divided doses
			Children older than 6 years of age: 12.5–25.0 mg, three to four times daily
	IM	Solution: 25 or 50 mg/mL	Same as oral
Hydroxyzine Pamoate (Vistaril)	Oral	Suspension: 25 mg/mL	Same as dosages for hydrochloride
		Capsules: 25, 50, and 100 mg	

Promethazine (Phenergan)	Oral	Tablets: 15.2, 25.0, and 50.0 mg	Adults: 50–100 mg three to four times daily for sedation
		Syrup: 3.25 mg/5 mL	Children: 12.5–25.0 mg at night for sedation
	Rectal	Suppositories: 12.5, 25.0, and 50.0 mg	
	IM	25 and 50 mg/mL	
Cyproheptadine (Periactin)	Oral	Tablets: 4 mg	Adults: 4–20 mg/day
		Syrup: 2 mg/5 mL	Children 2–7 years of age: 2 mg two to three times daily (maximum of 12 mg/day)
			Children 7–14 years of age: 4 mg two to three times daily (maximum of 16 mg per day)

IM, intramuscular; IV, intravenous.

Of particular concern with H_1R antagonists is the possibility of cardiotoxicity, particularly in cases of overdose or when coprescribed with other agents. The arrhythmogenic potential of several second-generation H_1R antagonists is mediated by nonhistaminergic receptors. Reports in the late 1980s of cardiac arrhythmia associated with intentional overdose and therapeutic use of terfenadine (Seldane) drew attention to the hazards of drug–drug interactions via cytochrome P450 oxidative enzymes. The first-generation H_1R antagonists exhibit both muscarinic and alpha-adrenergic blocking effects as well as dose-dependent prolongation of QT intervals in the electrocardiogram (ECG).

Second-generation H_1R antagonists astemizole (Hismanal) and terfenadine (Seldane) are known to block the rapid component of the delayed rectifier potassium current (I_{Kr}) with a 50% inhibitory effect in the nanomolar range. This results in prolongation of the monophasic cardiac action potential and the QT interval, engendering early depolarization and dispersion and eventually increasing the risk for torsades de pointes. Terfenadine is similar in potency to quinidine at inhibiting the I_{Kr} in cardiac tissue.

Second-generation antihistamines that were introduced later exhibit less potency in blocking the I_{Kr}, bestowing these agents with less cardiac toxicity potential.

Intentional or unintentional overdose of antihistamines can be fatal. Fortunately, the abuse liability of H_1R antagonists is generally low, although some first-generation agents may be used for recreational purposes (e.g., dimenhydrinate [Dramamine]). Some first-generation H_1R antagonists have been associated with sudden infant death syndrome. Other adverse events include eruption, photosensitivity, fever, hepatotransaminase elevation, and agranulocytosis.

Antihistamines are frequently used by pregnant women for the treatment of nausea, vomiting, and insomnia. First- and second-generation H_1R antagonists are FDA category B or C. Despite there being no clear association between antihistamines and fetal malformations, cardiac abnormalities, or obstetric outcomes, and despite the data being generally reassuring, antihistamines during pregnancy should be used only under medical supervision.

DRUG INTERACTIONS

Pharmacodynamic drug interactions are not uncommon and usually have additive CNS depressant and anticholinergic effects. Many over-the-counter agents contain antihistamines. Coprescribing antihistamines with monoamine oxidase inhibitors (MAOIs) may aggravate the anticholinergic effects.

LABORATORY INTERFERENCES

Promethazine has been reported to cause false results on pregnancy tests and increase glucose levels. Hydroxyzine may falsely elevate urinary 17-hydroxycorticosteroid levels. All available H_1R antagonists may interfere with allergy skin tests, resulting in a false-negative test.

DOSAGE ADMINISTRATION

Table 33.12a–2 outlines the commonly prescribed dosage of first-generation H_1R antagonists. Superficial administration of intramuscular formulations may result in local irritation. Hence deep injection is recommended. Most antihistamines are prescribed in doses that are divided to be administered over the day. However, when used as a hypnotic, a single evening dose is recommended.

FURTHER READINGS

Armstrong SC, Cozza KL. Antihistamines. *Psychosomatics*. 2003;44(5):430–434.

Barateau L, Dauvilliers Y. Recent advances in treatment for narcolepsy. *Ther Adv Neurol Disord*. 2019;12:1756286419875622.

Brown RE, Stevens DR, Haas HL. The physiology of brain histamine. *Prog Neurobiol*. 2001;63(6):637–672.

Chaudhry SK, Susser LC. Considerations in treating insomnia during pregnancy: a literature review. *Psychosomatics*. 2018;59(4):341–348.

Cheng L, Liu J, Chen Z. The histaminergic system in neuropsychiatric disorders. *Biomolecules*, 2021;11(9):1345.

Cookson J. Histamine in psychiatry: promethazine as a sedative anticholinergic. *BJPsych Adv*. 2019;25(4):265–268.

Dauvilliers Y, Bassetti C, Lammers GJ, et al; HARMONY I study group. Pitolisant versus placebo or modafinil in patients with narcolepsy: a double-blind, randomised trial. *Lancet Neurol*. 2013;12(11):1068–1075.

Fernández-Novoa L, Cacabelos R. Histamine function in brain disorders. *Behav Brain Res*. 2001;124(2):213–233.

Garakani A, Murrough JW, Freire RC, et al. Pharmacotherapy of anxiety disorders: current and emerging treatment options. *Focus (Am Psychiatr Publ)*. 2021;19(2):222–242.

Guaiana G, Barbui C, Cipriani A. Hydroxyzine for generalised anxiety disorder. *Cochrane Database Syst Rev*. 2010;(12):CD006815.

Haas H, Panula P. The role of histamine and the tuberomamillary nucleus in the nervous system. *Nat Rev Neurosci*. 2003;4(2):121–130.

Hersey M, Hashemi P, Reagan LP. Integrating the monoamine and cytokine hypotheses of depression: is histamine the missing link? *Eur J Neurosci*. 2022;55(9–10):2895–2911.

Kishi T, Iwata N. Efficacy and tolerability of histamine-2 receptor antagonist adjunction of antipsychotic treatment in schizophrenia: a meta-analysis of randomized placebo-controlled trials. *Pharmacopsychiatry*. 2015;48(1):30–36.

Kollb-Sielecka M, Demolis P, Emmerich J, Markey G, Salmonson T, Haas M. The European Medicines Agency review of pitolisant for treatment of narcolepsy: summary of the scientific assessment by the Committee for Medicinal Products for Human Use. *Sleep Med*. 2017;33:125–129.

Krief S, Berrebi-Bertrand I, Nagmar I, et al. Pitolisant, a wake-promoting agent devoid of psychostimulant properties: preclinical comparison with amphetamine, modafinil, and solriamfetol. *Pharmacol Res Perspect*. 2021;9(5):e00855.

Krystal AD, Prather AA, Ashbrook LH. The assessment and management of insomnia: an update. *World Psychiatry*. 2019;18(3):337–352.

Linnet K, Ejsing TB. A review on the impact of P-glycoprotein on the penetration of drugs into the brain. Focus on psychotropic drugs. *Eur Neuropsychopharmacol*. 2008;18(3):157–169.

Meskanen K, Ekelund H, Laitinen J, et al. A randomized clinical trial of histamine 2 receptor antagonism in treatment-resistant schizophrenia. *J Clin Psychopharmacol*. 2013;33(4):472–478.

Owczuk R, Twardowski P, Dylczyk-Sommer A, et al. Influence of promethazine on cardiac repolarisation: a double-blind, midazolam-controlled study. *Anaesthesia*. 2009;64(6):609–614.

Poyurovsky M, Weizman A. Beneficial effect of betahistine, a structural analog of histamine, in clozapine-related sedation. *Clin Neuropharmacol*. 2019;42(4):145.

Poyurovsky M, Fuchs C, Pashinian A, Levi A, Weizman R, Weizman A. Reducing antipsychotic-induced weight gain in schizophrenia: a double-blind placebo-controlled study of reboxetine–betahistine combination. *Psychopharmacology (Berl)*. 2013;226(3):615–622.

Rosse RB, Kendrick K, Fay-McCarthy M, et al. An open-label study of the therapeutic efficacy of high-dose famotidine adjuvant pharmacotherapy in schizophrenia: preliminary evidence for treatment efficacy. *Clin Neuropharmacol*. 1996;19(4):341–348.

Shan Y, Gao Y, Zhang L, Ma L, Shi Y, Liu X. H4 receptor inhibits lipopolysaccharide-induced NF-κB activation by interacting with tumor necrosis factor receptor-associated factor 6. *Neuroscience*. 2019;398:113–125.

Simons FER. Advances in H1-antihistamines. *N Engl J Med*. 2004;351(21):2203–2217.

Simons FER. Comparative pharmacology of H1 antihistamines: clinical relevance. *Am J Med*. 2002;113(Suppl 9A):38S–46S.

Tiligada E, Ennis M. Histamine pharmacology: from Sir Henry Dale to the 21st century. *Br J Pharmacol*. 2020;177(3):469–489.

Wang YQ, Takata Y, Li R, et al. Doxepin and diphenhydramine increased non-rapid eye movement sleep through blockade of histamine H1 receptors. *Pharmacol Biochem Behav*. 2015;129:56–64.

Welch MJ, Meltzer EO, Simons FER. H1-antihistamines and the central nervous system. *Clin Allergy Immunol*. 2002;17:337–388.

Drug information was retrieved from the Drugbank online database—https://go.drugbank.com/ and from the Neuroscience based Nomenclature (NbN) official website—https://nbn2r.com/

33.13 Pharmacological Treatments of Substance Abuse

33.13a Opioid Receptor Modulators

33.13a.1 Opioid Maintenance Treatments: Methadone and Buprenorphine

Jonathan R. Buchholz, M.D., Kelly Barth, D.O., and Andrew J. Saxon, M.D.

It has long been recognized that after ongoing, regular use of opioids, a use disorder including maladaptive use, functional impairment, withdrawal symptoms, and tolerance may develop. Commonly misused opioids include heroin, fentanyl, morphine, hydromorphone (Dilaudid), codeine, meperidine (Demerol), oxycodone, tramadol, hydrocodone (Robidone), and kratom. The prevalence of opioid misuse, use disorder, and opioid-related overdoses has risen exponentially since the 1990s. The 2019 National Survey on Drug Use and Health reported that approximately 9.7 million Americans used prescription pain relievers for nonmedical reasons in the past year, a decrease from previous estimates. However, starting in the early 2010s, heroin use started to rise, with estimates of 745,000 having used in the last year in 2019, and since 2013, the use of fentanyl and its analogs has become a central contributor to the exponential rise in opioid-related overdoses in North America. Whereas early reports suggested unintentional or unknown use of illicit fentanyl through drugs laced with fentanyl or counterfeit prescription opioid tablets, recent reports suggest more individuals are knowingly using fentanyl and may prefer fentanyl over other opioids. Regardless of the primary opioid used, opioid use disorder (OUD) is associated with several negative outcomes, including early mortality, increased rates of hepatitis, HIV infection, sexually transmitted diseases, and other health problems, as well as criminal justice system involvement.

PHYSICAL DEPENDENCE AND WITHDRAWAL

If an individual is physiologically dependent on opioids, opioid discontinuation or diminution in the amount consumed often leads to the development of an opioid withdrawal syndrome. Physical signs and symptoms that frequently occur include sweating, piloerection, yawning, rhinorrhea, sneezing, lacrimation, pupillary dilation, nausea, vomiting, abdominal cramps, diarrhea, tremor, muscle aches or cramps, agitation, dysphoric mood, fever, and sleep disturbances. Although opioid withdrawal is not usually life-threatening, other than times where it exacerbates underlying serious medical conditions, it is highly physically and subjectively uncomfortable, such that an individual continues to use opioids to avoid withdrawal symptoms.

The goal of most medically supervised withdrawal (detoxification) protocols is to limit patient discomfort by decreasing or ameliorating withdrawal symptoms. Medications commonly used in the treatment of acute opioid withdrawal include α_2-adrenergic agonists clonidine (Catapres) and lofexidine (Lucemyra), cyclobenzaprine (Flexeril) or benzodiazepines for muscle cramps, the antispasmodic dicyclomine (Bentyl), antidiarrheals such as loperamide, antiemetics such as prochlorperazine (Compazine) or ondansetron (Zofran), and sedatives such as zolpidem (Ambien) or benzodiazepines. Another option is to convert the patient to an equivalent dose of another opioid, typically either methadone or buprenorphine, and then gradually reduce the dose to minimize withdrawal. If methadone is used, it must be done under the auspices of a federally licensed program, and if buprenorphine is used, it must be prescribed by a qualified, trained physician with a special Drug Enforcement Agency

(DEA) number. If the goal is to expedite the withdrawal process, especially in the case of a long-acting opioid, administration of an opioid antagonist such as naloxone (Narcan) or naltrexone (Revia) hastens the onset of the withdrawal syndrome. Caution must be taken when using this approach because of the abrupt onset of symptoms that can appear in precipitated withdrawal. Again, adjunctive medications, such as clonidine, cyclobenzaprine, and dicyclomine, may also be used for the symptomatic treatment of some opioid withdrawal symptoms. The use of ultrarapid opioid detoxification, which involves the administration of general anesthesia or heavy sedation in individuals undergoing withdrawal, although effective in reducing physiologic dependence, has been demonstrated to lead to a greater number of serious adverse events than the other approaches without superior outcomes, and so it is not recommended. Once a patient has been completely withdrawn from opioids, naltrexone, an orally and parenterally active μ-opioid antagonist, may also be used to help prevent relapse. Prior to starting naltrexone, a naloxone challenge test can be performed in situations where the clinician is unsure if residual physiologic dependence remains to avoid a potential prolonged, precipitated withdrawal episode induced by the inherently longer acting naltrexone. Oral naltrexone may be dosed at 50 mg/day or 100 to 150 mg two to three times a week. Positive results have been found with using naltrexone in highly motivated patients, but many patients experiencing OUD tend to drop out of naltrexone therapy. The long-acting injectable form of naltrexone (Vivitrol) provides blockade of the μ-opioid receptor for about 30 days, is also available for the treatment of OUD, and has been shown to retain people in treatment more effectively than the oral formulation in a recent randomized trial. Two randomized control trials have compared the effectiveness of injectable naltrexone to sublingual buprenorphine for OUD.

OPIOID AGONIST THERAPIES FOR OPIOID USE DISORDER

Outcomes for medically supervised opioid withdrawal are generally suboptimal, so the generally preferred approach is utilizing one of the three Food and Drug Administration (FDA)-approved medications for the treatment of OUD. The other two FDA-approved medications for OUD include the μ-receptor opioid agonist agents methadone or buprenorphine. The goal of treatment with opioid agonist therapy is to provide a physician-ordered or prescribed opioid (either methadone or buprenorphine) to deter and ideally eliminate the use of illicit opioids. Early researchers in this area speculated that an effective pharmacologic treatment for OUD must eliminate the euphoric appeal of heroin, withdrawal symptoms, and cravings that draw those using opioids back to use; it must be sufficiently free from toxic dysphoric effects that patients will continue with the treatment; it must be orally effective, long acting, medically safe, and compatible with a normal performance at work and at school and with responsible behavior in society.

Methadone (Dolophine) which came into widespread use in the 1970s, was the first medication to meet these requirements. Levomethadyl (ORLAAM) was available from 1993 until 2003 as opioid maintenance therapy; however, production of levomethadyl was discontinued by the manufacturer due in part to reports of prolonged QT intervals and risk of torsades de pointes. In 2002, buprenorphine and buprenorphine/naloxone (Suboxone, Zubsolv, Bunavail, Cassipa, Probuphine, Sublocade) became the first opioid maintenance medications approved for office-based treatment of OUD.

As of 2019, among the more than 1.6 million people aged 12 years or older with a past-year OUD, only 294,000 were receiving treatment for OUD. Barriers to patients entering licensed opioid treatment programs include state and federal regulations for program admittance, treatment costs, and inadequate expansion of treatment centers since the 1980s. The social stigma associated with methadone treatment also may deter individuals experiencing OUD from entering treatment. Office-based treatment with buprenorphine, which appears to have nearly comparable effectiveness to clinic-based treatment, is making opioid maintenance therapy more accessible. As of March 2021, 98,076 providers had obtained waivers to prescribe buprenorphine, but according to a recent study, only half of the waivered clinicians prescribed buprenorphine and most were treated at levels below their patient limits. No clear data exist on the total number of patients treated since FDA approval, but it certainly exceeds 1,000,000. The following discussion reviews the history and current status of methadone and buprenorphine. A discussion on the use of opioid agonists in the treatment of chronic pain is also included.

METHADONE

Neuroscience-Based Nomenclature (NbN)	
Opioid Receptor Agonist	
Pharmacologic Domain	Opioid
Mechanism of Action	Receptor antagonist (μ)

In response to high rates of heroin-related deaths in the 1960s, researchers began investigating the feasibility of opioid maintenance therapy. Vincent Dole, Marie Nyswander, and Mary Jeanne Kreek evaluated several short-acting opioids as possible replacement medications but found that patients were overly sedated by the narcotic effects and still craved heroin. Moreover, patients required frequent dosing—every 4 to 6 hours—to prevent the emergence of withdrawal symptoms. Searching for a longer-acting alternative, the researchers administered methadone, which had been previously shown to be useful in heroin withdrawal. They discovered that, with methadone, it was possible to maintain a patient experiencing heroin dependence "between the limits of narcosis and abstinence [where] there is a functional zone."

The rapid expansion of methadone maintenance programs occurred in the 1970s during the presidency of Richard Nixon. At this time, politicians were concerned about the high rate of heroin-related crimes and were encouraged by the decrease in illegal activities reported by individuals receiving methadone. However, initial rapid expansion was soon curtailed by government regulation in an attempt to maintain some control over the distribution of methadone. Today, the DEA and the Center for Substance Abuse Treatment, which is part of the federal Substance Abuse and Mental Health Services Administration, both have roles in regulating licensed opioid treatment programs. There remains substantial community resistance to the placement and existence of these programs, which limits the availability of services in some areas. Critics often suggest that methadone maintenance is, in essence, trading one chemical dependency for another and is therefore not an appropriate treatment. Finally, there is concern that methadone diversion may occur, further worsening OUD epidemics or contributing to opioid overdoses. In reality, evidence accrued thus far suggests that diverted methadone comes mainly from methadone prescribed to treat pain, not from methadone dispensed in licensed opioid treatment programs.

Chemistry

Methadone is a synthetic opioid of the phenylheptylamine class structurally unrelated to morphine (Fig. 33.13a.1–1). Methadone is comprised of a racemic mixture of two stereoisomers. R-(-)-methadone (levomethadone) is responsible for the drug's analgesic and other opioid agonist effects, while S-(+)-methadone (dextromethadone) exhibits substantially less analgesic activity and does not have respiratory depression activity or addiction liability, but does have antitussive effects.

Pharmacologic Actions

Pharmacokinetics. Methadone is rapidly absorbed after oral administration. It is highly plasma protein bound, with an approximate volume of distribution of 4 L/kg. Methadone is primarily metabolized in the liver by cytochrome P450 (CYP) enzymes, predominately by CYP2B6 and CYP3A4, while its inactive metabolite, 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine (EDDP), is excreted in the urine. Single nucleotide polymorphisms

(SNPs) located within the CYP enzymes may alter methadone metabolism. Carriers of CYP2B6*6/*6 may have a greater risk for adverse effects given the decreased metabolism and clearance of methadone in these individuals, while carriers of CYP2B6*4 have been observed with decreased plasma concentrations of methadone secondary to increased methadone clearance. Of interest, women may metabolize methadone more rapidly than men. After oral administration, methadone has a biexponential half-life with an initial phase range of 12 to 24 hours and a secondary phase of 55 hours. Methadone obviously crosses the blood–brain barrier as well as the placenta and is found in maternal milk. The benefits of breastfeeding far outweigh concerns about exposing infants to the methadone in breast milk, and mothers on methadone maintenance are encouraged to breastfeed.

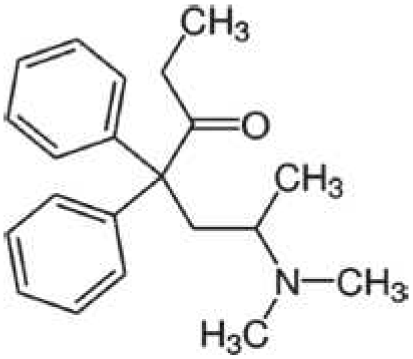

FIGURE 33.13a.1–1. Molecular formula for methadone.

Pharmacodynamics. The major pharmacodynamic action of methadone is agonism at the μ-opioid receptor, but methadone, unlike most other opioids, also acts as an antagonist at the N-methyl-D-aspartate receptor (NMDA) receptor and blocks serotonin and norepinephrine transporters. In addition, methadone inhibits the cardiac potassium channel, hERG, which can cause a prolonged QT interval on the Electrocardiogram (ECG). There is often not a clear linear relationship between methadone blood levels and its clinical effects. Therefore, it is rarely necessary to obtain blood levels during clinical care of patients on methadone maintenance. Suggested minimally therapeutic trough serum levels of methadone range from

100 to 400 ng/mL. Peak serum levels of methadone vary considerably in stable patients. A peak-to-trough serum ratio of greater than 2:1 is often considered indicative of inadequate pharmacologic coverage throughout the 24-hour dosing interval. Patients who display such a pattern may be rapid metabolizers of methadone who would ideally do best on a split-dosing regimen rather than on a higher single daily dose.

Design and Interpretation of Clinical Studies on Methadone

Although methadone maintenance treatment (MMT) was introduced in the 1970s, few randomized, controlled trials have evaluated its efficacy. Although these studies had small sample sizes and rarely followed patients for longer than 1 year, they reported positive results. The majority of data supporting the use of methadone maintenance comes instead from the many observational studies that have been conducted to evaluate this treatment modality. These studies support the findings of clinical trials in demonstrating that methadone treatment reduces heroin or other opioid use and criminal activity. Furthermore, MMT has been shown to decrease mortality among patients with OUD. Mortality estimates indicate that individuals who use heroin are one-fourth as likely to die if they are treated with methadone as those not in treatment. MMT also appears to decrease drug use practices, such as needle sharing, that are associated with the transmission of HIV and viral hepatitis. Methadone treatment appears to be cost-effective in that the cost per year of life saved is less than that incurred for other common medical therapies such as coronary bypass surgery or renal dialysis. A review that included seven randomized controlled trials and two quasi-experimental studies of MMT from 1995 to 2012 confirmed all the benefits described above, particularly when the dose of methadone exceeded 60 mg/day, and also confirmed improved maternal and fetal outcomes for pregnant women with OUD when provided MMT.

Adverse Effects

General. Because methadone shares the adverse effects and toxic potential of other opioid agonists, the usual precautions of opioid agonist therapy should be observed. Common side effects of opioids include sedation, constipation, sweating, nausea, dizziness, and hypotension. During methadone maintenance, most adverse effects disappear over the course of several weeks. However, constipation and excessive sweating often persist even with long-term methadone administration. Recent case reports also indicate that, as noted above, methadone, especially in high doses, although occasionally in moderate doses, can impair cardiac conduction, prolong the QT interval, and, in rare instances, lead to torsades de pointes. Like other opioids, methadone does suppress the hypothalamic–pituitary–gonadal axis and thus can cause hypogonadism with low testosterone levels in men and anovulation and lack of menstruation in women. A potential downstream effect of hypogonadism could be osteoporosis. Methadone does not cause chronic adverse effects on the heart, lungs, liver, or kidneys.

Overdose and Prevention. Severe intoxication or overdose with opioids constitutes a medical emergency. Miosis (known as pinpoint pupils) is a recognized effect of most opioids and can be used to identify opioid intoxication. Of most concern in opioid overdose is potentially fatal respiratory depression from direct suppression of respiratory centers in the midbrain and medulla. For severe cases, the administration of the pure opioid antagonist naloxone in combination with general supportive measures is indicated. Owing to the long half-life of methadone, repeat administration of naloxone may be necessary. An overdose of methadone may also involve other sedative agents that may affect treatment decisions. In 2006, the FDA issued a public safety advisory leading to a black box warning regarding risks of respiratory depression and QT prolongation associated with methadone. Center for Disease Control and Prevention opioid prescribing guidelines released in 2016 recommend considering naloxone and overdose prevention education for patients and household members of patients with a history of overdose, substance use disorder, those taking higher opioid dosages, or those with concurrent benzodiazepine use. Patients taking methadone or buprenorphine for OUD should be given naloxone rescue kits and counseled on proper use. There are multiple FDA-approved formulations of naloxone that are used in naloxone rescue kits: (1) injectable naloxone that is drawn up out of a vial with a needle into a syringe with a dose concentration of 0.4 mg/1 mL, (2) an autoinjector with audio prompts that administers a 0.4 mg intramuscular (IM) dose via a retractable needle, (3) a single-step nasal spray that administers a dose concentration of 4 mg/0.1 mL into one nostril, and (4) a multistep nasal spray assembled by combining a prefilled syringe with a nasal atomizer, that administers a dose concentration of 2 mg/2 mL, where 1 mL is administered to each nostril. Opioid overdose education and community naloxone distribution (OEND) programs have preferred intranasal, and IM autoinjector formulations as these are easier for a layperson to use. IM formulations tend to be more expensive and have insurance coverage variability.

Drug Interactions

Pharmacokinetic interactions are mainly related to methadone being a substrate for CYP450 2B6 and 3A4. Inducers of P450 enzymes such as the anticonvulsants, phenytoin and carbamazepine, the antituberculous medication rifampin, and antiretroviral medications, nevirapine and efavirenz increase the metabolism of methadone so that the patient could go into opioid withdrawal suddenly and become destabilized. Newer antiviral agents are generally well tolerated in patients taking methadone with routine monitoring. Inhibitors of the enzymes such as the antidepressant fluvoxamine, the antibiotic, erythromycin, and antifungal medications, like ketoconazole, have the potential to reduce methadone metabolism and lead to side effects or toxicity, though clinically meaningful effects are rarely observed with inhibitors. Pharmacodynamic interactions obviously can occur with opioid antagonists, such as naloxone or naltrexone, which would cause a patient on methadone maintenance to go into severe, precipitated withdrawal. Other opioids, sedative–hypnotics, and alcohol also have a pharmacodynamic interaction with methadone, leading to additive excess sedation and a greater risk of suppression of respiratory drive and overdose.

Laboratory Interference

None known.

Dosing

Federal eligibility requirements for MMT stipulate that individuals entering treatment have a minimum 1-year history of OUD. The minimum age requirement is 18 years of age; however, individuals between 16 and 18 years of age may enter methadone treatment under certain conditions. There is no upper age limit on program entry. Pregnant women with OUD may be accepted for methadone treatment. Individuals with major medical conditions are also eligible for treatment, as are those who use multiple substances where OUD is one component of their addictive disease.

Methadone dose determination should always be performed by an experienced physician based on their clinical assessment of the patient. Data indicate that adequate methadone doses are correlated with reduced illicit drug use and improved patient retention in treatment. A comparison by Eric C. Strain and colleagues of moderate daily doses of methadone (40 to 50 mg) and high doses (80 to 100 mg) showed significantly lower illicit opioid use in the high-dose group. Similar findings have been reported in regard to treatment retention, with one study showing that patients receiving 80 mg/day of methadone were twice as likely to remain in treatment as those receiving 60 to 79 mg, who were, in turn, twice as likely to remain in treatment as patients receiving less than 60 mg. The lowest effective dose of methadone is now widely accepted to be approximately 60 mg, with an optimum average dose for most patients of approximately 80 mg (±20 mg), although some patients require a dose higher than this range.

In general, methadone can be initiated in an individual with well-established current physiologic dependence at a dose of 20 to 30 mg. If there is any doubt as to the degree of an individual's tolerance to opioids, an initial dose of 10 mg is indicated. Under federal regulations, the maximum allowable first dose is 30 mg, and the maximum total dose for the first day of treatment is 40 mg unless the physician documents the fact that 40 mg is insufficient to suppress withdrawal symptoms. Dose escalation should be done in small increments (usually 5 to 10 mg at a time) under close physician supervision until an optimum maintenance dose is achieved. Given methadone's long half-life, it is advisable to observe the patient for at least 3 to 4 days before an additional dose increase is provided. The optimum dose is achieved when (1) the patient evidences no withdrawal signs or symptoms throughout the 24-hour dosing period, (2) the patient reports an absence of craving for opioids, (3) adequate tolerance is obtained such that the patient experiences little or no reinforcement from use of other opioids, and (4) self-report and urine testing indicate the absence of illicit opioid use. Ideally, these beneficial effects need to be balanced against side effects so that small dose decrements may be required if unpleasant or intolerable side effects occur.

The duration of maintenance therapy may range from a few months to an indefinite period. In general, methadone maintenance should be continued as long as the patient is continuing to benefit from and wishes to remain in treatment. It should be noted that methadone administration should ideally be combined with comprehensive counseling and rehabilitative services.

BUPRENORPHINE

Neuroscience-Based Nomenclature (NbN)

Opioid Receptor Modulator

Pharmacologic Domain	Opioid
Mechanism of Action	Receptor partial agonist (µ) and antagonist (κ, d)

Reckitt and Coleman Ltd. developed buprenorphine in the 1970s as an analgesic because of the widespread interest in developing mixed agonist and antagonist opioid agents for pain relief. It was hoped that these agents would offer analgesia with greater safety and less abuse liability than the pure opioid agonists. Buprenorphine is available in many countries, including the United States, as an analgesic. It was first marketed in the United Kingdom in 1978 for injection and in 1981 as a sublingual tablet.

The potential for the use of buprenorphine in the treatment of OUD was first recognized in the late 1970s when Donald Jasinski and colleagues demonstrated that buprenorphine substituted for morphine and suppressed opioid withdrawal while having few withdrawal symptoms. Human laboratory studies have demonstrated that buprenorphine suppresses heroin self-administration. A number of published clinical observations have also supported the usefulness and general safety of buprenorphine in clinical settings. However, in clinical practice, buprenorphine has been widely used as an injectable analgesic, and intravenous (IV) buprenorphine has been reported as a significant drug of misuse in Europe, New Zealand, Australia, and the Republic of Georgia. In 2002, the FDA approved the use of buprenorphine for the treatment of OUD (then termed opioid dependence). It is commercially available as a generic sublingual tablet containing buprenorphine only or a combination of buprenorphine and naloxone, which can come as a sublingual tablet, a sublingual film, or a buccal film. More recently, implantable and injectable formulations of buprenorphine have become available. In 2016, the FDA approved a 6-month subdermal formulation of buprenorphine requiring surgical implantation, but this formulation was discontinued in 2020 primarily due to low utilization. In 2017, a monthly subcutaneously injectable formulation of buprenorphine was approved for the treatment of moderate to severe OUD for individuals stable on transmucosal buprenorphine for at least 7 days, and another subcutaneous injectable formulation with weekly and monthly injection options is currently in the FDA-approval process.

Of note, there are also several formulations of buprenorphine that are specifically FDA approved for the treatment of pain, including an IV formulation for acute severe pain and transdermal patch and buccal film formulations for severe, chronic pain that requires long-term opioid treatment.

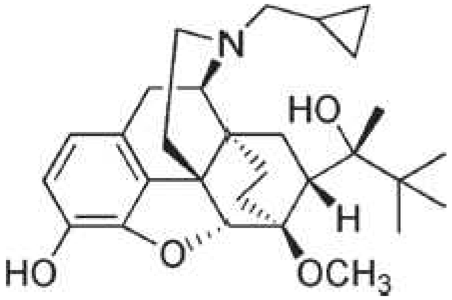

FIGURE 33.13a.1–2. Molecular formula for buprenorphine.

Chemistry

Buprenorphine is a semisynthetic derivative of the opioid alkaloid thebaine (Fig. 33.13a.1–2).

Pharmacologic Actions

Pharmacokinetics. Because of extensive metabolism by the intestine and liver, buprenorphine is poorly absorbed by the oral route and is, therefore, best administered using IV or sublingual, buccal, subcutaneous, or transdermal dosing. When administered by the IV route, buprenorphine is readily distributed into tissues, resulting in a large volume of distribution. The elimination is biphasic, with a rapid distribution phase ($T_{1/2}$ = 2 minutes) and a slower elimination phase ($T_{1/2}$ = 3 to 5 hours). The rate of absorption is slower by the sublingual route, with a time to reach maximum drug concentration (T_{max}) of approximately 60 minutes. It takes 3 to 5 minutes to dissolve, and sublingual bioavailability is approximately 55%. There is considerable between-subject variability in bioavailability, but the data suggest greater bioavailability of buprenorphine/naloxone versus buprenorphine alone. Buprenorphine has a variable range (3 to 44 hours) but typically has a long half-life. Following three-times-per-day dosing, steady-state levels of buprenorphine are reached at approximately 4 days. For the monthly extended-release buprenorphine injection, upon subcutaneous injection into the abdomen, a solid depot is created from which buprenorphine slowly diffuses as the depot undergoes biodegradation, resulting in sustained therapeutic plasma concentrations over the entire monthly dosing interval. Buprenorphine is 96% bound to plasma proteins. Buprenorphine undergoes both glucuronidation and N-dealkylation, with N-dealkylation catalyzed by the liver enzyme CYP450 3A4 producing an active metabolite, norbuprenorphine. The main route of elimination for both parent drug and metabolites is fecal, with lesser amounts excreted by the kidneys. Buprenorphine crosses the blood–brain barrier as well as the placenta and is found in maternal milk. The benefits of breastfeeding far outweigh concerns about exposing infants to the buprenorphine in breast milk, especially since buprenorphine has such low bioavailability by the oral route, so mothers on buprenorphine maintenance are encouraged to breastfeed.

Pharmacodynamics. Pharmacologic studies have shown that buprenorphine combines a strong affinity for the μ-opioid receptor with low intrinsic activity. As such, it is considered a partial μ-agonist. Buprenorphine is also an antagonist at the κ-receptor. Because buprenorphine is a partial agonist, it has a considerable margin of safety with little chance of lethal overdose.

While blood levels of buprenorphine and norbuprenorphine are generally proportional to dose, to date, no specific correspondence between blood levels and therapeutic effect has been demonstrated.

Design and Interpretation of Clinical Studies on Buprenorphine

A series of well-controlled clinical trials have clearly established the safety and clinical efficacy of buprenorphine for OUD. These trials have demonstrated the efficacy of buprenorphine, as compared to placebo, in decreasing illicit opioid use in individuals with OUD. Subsequent trials have established the safety and efficacy of buprenorphine in the maintenance treatment of OUD in doses ranging from 2 to 32 mg/day.

Controlled clinical trials have demonstrated that buprenorphine and moderate-dose methadone (50 to 60 mg/day) have similar effects on retention and illicit drug use in maintenance programs. Comparisons of buprenorphine with higher doses of methadone suggest that high-dose methadone may have a slight advantage in terms of treatment retention and suppression of illicit opioid use. One recent study suggests that higher buprenorphine doses are associated with better treatment retention. Studies by Walter Ling and colleagues have demonstrated successful implementation and use of buprenorphine in private physicians' offices. Further studies have indicated that buprenorphine works as well or better for patients who primarily misuse prescription opioids as opposed to heroin; that for patients without severe psychiatric illness, outcomes for buprenorphine are as good with quality medical management alone as compared to addiction-focused counseling or psychotherapy added to medical management; that outcomes with maintenance buprenorphine treatment are superior to outcomes when buprenorphine is tapered; and that initiating buprenorphine is more readily achievable than initiating extended-release naltrexone, which leads to lower relapse rates.

The Drug Addiction Treatment Act, which was signed into law in October 2000, marked the beginning of a process designed to allow qualified physicians to prescribe approved narcotic drugs for the treatment of OUD in office-based settings. There are a number of approved training programs in place to teach physicians about the use of buprenorphine in an office-based setting. This is the first time in modern medicine that physicians in the United States practicing in a variety of clinical settings, including office-based practice, are able adequately to treat OUD with pharmacotherapy. This initiative has increased access to treatment for some individuals with OUD, but in the face of an ongoing opioid overdose crisis that has worsened during the COVID-19 pandemic, a treatment gap still remains.

Adverse Effects

Because buprenorphine is a partial agonist rather than a full agonist at the μ-receptor and it is a weak antagonist at the κ-receptor, this agent produces a milder withdrawal syndrome and has a wider margin of safety than the full μ-agonist compounds generally used in treatment. Buprenorphine has a ceiling effect beyond which dose increases prolong the duration of action of the drug without further increasing the agonist effects. Because of this, buprenorphine has a high clinical safety profile, with limited respiratory depression, therefore decreasing the likelihood of lethal overdose. Buprenorphine does have the capacity to cause typical side effects associated with opioids, including sedation, nausea and vomiting, constipation, dizziness, headache, and sweating. Buprenorphine may have the potential to suppress the hypothalamic–pituitary–gonadal axis, but this possible adverse effect requires further study. Buprenorphine does not cause chronic adverse effects on the heart, lungs, or kidneys. Some small early studies and case reports raised a concern that it could be hepatotoxic, but more recent, larger studies show no evidence of a hepatotoxic effect.

Drug Interactions

Theoretical pharmacokinetic interactions have been proposed mainly related to buprenorphine also being a substrate for CYP450 3A4. However, to the extent that these theoretical interactions have been studied in the human laboratory, no clinically meaningful effects have been observed, and no isolated clinical case reports have surfaced. It has been proposed that because buprenorphine has a high affinity for and slow dissociation from the μ-opioid receptor, changes in plasma levels have little effect on efficacy. Pharmacodynamic interactions obviously can occur with opioid antagonists, such as naloxone or naltrexone, although because of buprenorphine's high affinity and slow dissociation, there is competition for the receptor, with the likelihood of precipitated withdrawal being low. Other opioids, sedative–hypnotics, and alcohol could also have a pharmacodynamic interaction with buprenorphine leading to additive excess sedation and a greater risk of suppression of respiratory drive and potential overdose. This risk is less than with full-agonist opioids, and overdoses in adults on buprenorphine alone are nearly nonexistent.

Dosing

To reduce the likelihood of misusing buprenorphine via the IV route, buprenorphine has been combined with the opioid antagonist naloxone for sublingual administration. Because naloxone is poorly absorbed by the sublingual route, when the combination drug is taken sublingually, there is no effect of the naloxone on the efficacy of buprenorphine. If an individual with OUD injects the combination medication, the naloxone may precipitate a withdrawal reaction and/or block the intended effects of the buprenorphine being sought, therefore reducing the likelihood of illicit injection use of the sublingual preparation.

Inducting and stabilizing a patient on buprenorphine is analogous to inducting and stabilizing a patient on methadone, except that, as a partial agonist, buprenorphine has the potential to cause precipitated withdrawal in patients who have recently taken full-agonist opioids. Thus, a patient must abstain from the use of short-acting opioids for 12 to 24 hours prior to starting buprenorphine and from longer-acting opioids such as methadone for 24 to 48 hours or longer. The physician must assess the patient clinically and determine that the patient is in mild to moderate opioid withdrawal with objectively observable withdrawal signs prior to initiating buprenorphine. In most instances, a relatively low dose of buprenorphine (2 to 4 mg) can then be administered with additional doses given in 1 to 2 hours if withdrawal signs persist. Alternatively, the induction can be done at the patient's home after sufficient education about when to begin the medication.

Certain groups can be at higher risk for precipitated withdrawal during buprenorphine induction, including those being transitioned from high-dose full-opioid agonists (including methadone and fentanyl) and those unintentionally or intentionally exposed to illicit fentanyl. There has been a growing literature showing early promise for several different clinical approaches to decrease the risk of precipitated withdrawal in these populations, including buprenorphine "microinductions." Buprenorphine "microinduction" entails introducing very low doses of buprenorphine (e.g., 0.5 mg sublingually) while an individual is still on full-agonist opioids or coming off high-affinity, slowly cleared full agonists such as fentanyl, and gradually tapering the full opioid agonists (as applicable) while gradually uptitrating buprenorphine, a process that generally occurs over several days. Additional prospective studies would clarify the early clinical and anecdotal success in this area.

Once induction is accomplished, the subsequent goal for the first 24 hours is to suppress withdrawal signs and symptoms, and the total 24-hour dose to do so can range from 2 to 16 mg on the first day. In subsequent days, the dose can be adjusted upward or downward to resolve withdrawal fully and, as with methadone, to achieve an absence of craving, adequate tolerance to prevent reinforcement from the use of other opioids, and ultimately abstinence from other opioids while minimizing side effects. Dose-ranging studies have demonstrated that doses of 6 to 16 mg/day are associated with improved treatment outcomes, as compared to lower doses of buprenorphine (1 to 4 mg). Sometimes, patients seem to need doses higher than 16 mg/day, although there is no evidence for any benefit of doses beyond 32 mg/day. For the treatment of OUD, a dose of approximately 4 mg of sublingual buprenorphine is the equivalent of a daily dose of 40 mg of oral methadone. It has also been demonstrated that daily, alternate-day, or three-times-per-week administration, all have equivalent effects in suppressing symptoms of opioid withdrawal in physiologically dependent individuals. The buprenorphine/naloxone combination formulations are recommended for most clinical purposes, including induction and maintenance. The buprenorphine monoproduct should be used only for pregnant patients or for patients who have a documented anaphylactic reaction to naloxone.

The newer forms of buprenorphine delivery include a long-acting monthly subcutaneous injection that provides therapeutic plasma levels for 28 to 30 days. This delivery system obviates the need for taking medications daily while virtually eliminating the risk of medication nonadherence or diversion. Administered in a health care provider's office, this formulation is indicated for those with moderate to severe OUD who have been stable on a transmucosal buprenorphine product for at least 7 days. The approved dosing regimen is 300 mg administered subcutaneously for the first 2 months, followed by maintenance doses of 100 mg monthly with the option to continue 300 mg monthly if clinically indicated.

OPIOID AGONISTS IN THE TREATMENT OF PAIN IN INDIVIDUALS WITH ADDICTIVE DISORDERS

Opioid agonists are efficacious agents for the treatment of acute pain and possibly for some individuals with chronic pain syndromes, although in randomized controlled trials, their long-term use in chronic pain has not been demonstrated to be superior to nonopioid medications, and the use of opioid agonists in the treatment of pain in individuals with substance use disorders is controversial. Untreated pain can be a powerful trigger for relapse. On the other hand, exposure to opioid analgesics for some individuals with addictive disorders could trigger reminders of past use and place them at risk for relapse. In addition, some of the physiologic aspects of addictive disorders may make pain more difficult to treat in the individual with a substance use disorder than in individuals without such a disorder. Finally, it can be difficult to distinguish distress related to pain from distress related to drug craving in individuals with OUD. In spite of these potential problems, individuals with substance use disorders develop both acute and chronic pain disorders that warrant treatment, and total prohibitions against the use of opioids are neither practical nor good medical practice. However, pain management in individuals with addictions must attend to the treatment of both the pain and the addictive disorder to be successful. Ongoing support for maintaining recovery is essential for patients with OUD taking opioids for pain, including nonpharmacologic treatments like recovery-oriented meetings and/or regular follow-up with a treatment professional who understands the issues involved in the treatment of pain in individuals with addiction.

Acute Pain Management

Undertreatment of pain associated with trauma and surgery has been demonstrated to increase morbidity and lengthen hospital stays. In individuals with addictive disorders, acute pain conditions increase distress and anxiety and may become a risk factor for relapse. Physicians treating patients in acute pain who have addictive disorders should reassure the patient that their addiction is not an obstacle to the appropriate and effective treatment of the pain condition.

For patients with OUD, the prescriber should not consider tapering or discontinuing opioid medications until the acute pain situation is resolved. The use of nonmedication pain treatment approaches, such as ice, heat, physical therapy, acupuncture, or a transcutaneous electrical nerve stimulation (TENS) unit, as indicated, should be optimized. When medications are necessary, it is best to use those least likely to alter mood or to produce physical dependence if they are effective. Nonsteroidal anti-inflammatory drugs, tricyclic antidepressants, anticonvulsants, and topical agents, all have demonstrated efficacy in certain pain conditions. However, when opioid agents are necessary to manage pain conditions, they should be used effectively. In the management of moderate to severe acute pain, opioids are generally the mainstay of treatment in addicted or nonaddicted individuals. Scheduled administration of opioids is preferred to as-needed administration in individuals with addictive disorders, providing several advantages. If the patient needs to ask for medications, this may be misconstrued as drug-seeking behavior. In addition, when drug administration is time dependent rather than symptom dependent, reinforcement of pain symptoms is minimized.

Individuals on methadone maintenance or physically dependent on other opioids should have their baseline opioid requirements met in addition to the medications provided for pain treatment. At times these individuals require more opioid pain medication than the average individual because of their opioid tolerance and associated hyperalgesia. When possible, patients on methadone should receive a different opioid medication for the control of acute pain or have additional methadone prescribed on a three- or four-times-daily basis. Although an increase in methadone in divided doses throughout the day is generally effective in acute pain management, this strategy makes it easy to confuse issues of pain treatment with issues of addiction treatment. When the acute pain condition resolves, the taper of methadone back to the original maintenance dose can at times prove difficult for the patient.

The management of patients on buprenorphine who develop an acute pain condition is somewhat more complex, with fewer data and clinical experiences to guide it. Although buprenorphine is an effective analgesic, it may not have the analgesic potency of a full-agonist opioid. There are essentially three options. One is to increase the dose of buprenorphine and give it in divided doses throughout the day. Alternatively, even on stable doses of buprenorphine, additional μ-opioid receptors still remain available for full-agonist opioids to bind and activate, so another option is to try adding to the regimen a full-agonist opioid with high intrinsic activity at the μ-opioid receptor, such as morphine, hydromorphone, or fentanyl. A third option is to discontinue buprenorphine and switch temporarily to a full-agonist opioid during the time acute pain treatment is needed. When the acute pain (whether from injury, illness, or postoperatively) subsides, the patient can be transitioned back to buprenorphine, with care taken not to administer buprenorphine too soon after ingestion of the full agonist to avoid precipitating withdrawal.

Similar considerations apply to perioperative pain management of patients on buprenorphine, and rather than discontinuing buprenorphine perioperatively, many medical centers now recommend continuing or dose-adjusting buprenorphine during the perioperative period.

Management of Chronic Pain

The treatment of chronic pain in individuals with addictive diseases also does not differ significantly from the treatment of these pain conditions in nonaddicted individuals. The goals of treatment are the reduction of pain, improvement in associated features, and restoration of function. In general, these goals are approached initially through the use of nonpharmacologic strategies, such as physical therapy/therapeutic exercise, TENS units, peripheral nerve blocks, and cognitive behavioral therapy. The mainstay of medication treatment for chronic, nonmalignant pain includes nonsteroidal anti-inflammatory drugs, tricyclic antidepressants, and anticonvulsants. Anticonvulsants (pregabalin) and antidepressant agents (duloxetine) have received FDA approval for specific chronic nonmalignant pain syndromes.

Even when medications, including opioids, are deemed necessary in the treatment of chronic pain syndromes, the treatment plan should be multidimensional. Specifically, it should include a careful history of the development of the pain problem, factors that contribute to the patient’s pain and distress, and a plan for dealing with each of these contributing factors. The treatment plan may include the physical treatments of pain mentioned previously (i.e., TENS unit, nerve blocks), as well as behavioral interventions, such as relaxation training, biofeedback, and changes in behavioral responses to pain. The use of long-term opioids as one component in the treatment of chronic pain is generally a reasonable approach when other treatment regimens have failed. Clinicians should be aware that long-term opioid therapy use may be associated with the development of opioid tolerance as well as an abnormal sensitivity to pain or hyperalgesia, conditions which can be related and manifest clinically as needing increased doses of opioids to maintain the same level of analgesia. Opioid dose escalation should be approached with caution in these scenarios, as, by nature of tolerance, a dose increase may temporarily provide relief and meet the goals of treatment (decreased pain, improved functionality), but several weeks, months to years later, the patient and provider may find themselves in the same predicament (inadequate pain relief despite previous positive response to increased opioid dose). When tolerance development is noticed after multiple dose adjustments, a discussion should be had regarding options to address tolerance because continued dose escalation is generally not the safest long-term option. Alternative approaches include opioid rotation to an agent with less potential for tolerance, such as buprenorphine or a trial of opioid tapering, though much more research in the area of opioid tolerance is needed. For individuals with or without addiction, when opioids are used for pain, it is important to document their benefit by observing that their functional status is improving. Particularly for those with a history of opioid or other substance use disorder, when considering long-term opioid therapy, one may consider an agent such as buprenorphine (patch or buccal film FDA approved for pain) even as an initial long-term opioid choice. The majority of chronic pain patients, with or without addiction, achieve relief of pain through multimodal approaches that usually do not include the use of opioids. For most patients, a trial period with the use of coordinated alternative treatments should be tried before long-term opioid therapy is used.

Most recently, Health and Human Services convened an interagency task force that released an updated guide for clinicians: “Pain Management Best Practices,” which reviews evidence-based medications, restorative therapies, interventional procedures, behavioral health approaches, and complementary and integrative approaches to chronic pain. This report highlights the importance of multimodal care, risk assessment, safe opioid stewardship, and regular re-evaluation for all patients with chronic pain to ensure pain is managed as effectively as possible while mitigating risks of undertreated pain and long-term opioid use.

FUTURE DIRECTIONS

The efficacy of opioid agonist medications in the treatment of OUD has been clearly established. In spite of the data demonstrating that maintenance on buprenorphine or methadone is effective in decreasing opioid use, improving employment and prosocial behaviors, and enhancing public health and safety, only a small percentage of individuals with OUD in the United States are in treatment. It is hoped that legislation making it possible to prescribe opioid agonist therapies in a variety of settings and the development of agonist agents with a greater margin of safety and less abuse potential will increase the number of individuals with OUD receiving treatment.

FURTHER READINGS

Ahmad T, Valentovic MA, Rankin GO. Effects of cytochrome P450 single nucleotide polymorphisms on methadone metabolism and pharmacodynamics. *Biochem Pharmacol*. 2018;153:196–204.

Ahmed S, Bhivandkar S, Lonergan BB, Suzuki J. Microinduction of buprenorphine/naloxone: a review of the literature. *Am J Addict*. 2021;30(4):305–315.

Antoine D, Huhn AS, Strain EC, et al. Method for successfully inducting individuals who use illicit fentanyl onto buprenorphine/naloxone. *Am J Addict*. 2021;30(1):83–87.

Ball JC, Ross A. *The Effectiveness of Methadone Maintenance Treatment: Patients, Programs, Services, and Outcome*. Springer-Verlag; 1991.

Center for Substance Abuse Treatment. Clinical Guidelines for the Use of Buprenorphine in the Treatment of Opioid Addiction. *Treatment Improvement Protocol (TIP) Series 40. DHHS Publication No. (SMA) 04-3939*. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2004.

Center for Substance Abuse Treatment. Medication-Assisted Treatment for Opioid Addiction in Opioid Treatment Programs. *Treatment Improvement Protocol (TIP) Series 43. DHHS Publication No. (SMA) 12-4214*. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2005.

Dole VP, Nyswander M. A medical treatment for diacetylmorphine (heroin) addiction. A clinical trial with methadone hydrochloride. *JAMA*. 1965;193:646–650.

Doran CM, Shanahan M, Mattick RP, Ali R, White J, Bell J. Buprenorphine versus methadone maintenance: a cost-effectiveness analysis. *Drug Alcohol Depend*.

2003;71:295–302.

*Effective medical treatment of opiate addiction. National Consensus Development Panel on Effective Medical Treatment of Opiate Addiction. *JAMA*. 1998;280:1936–1943.

Ehret GB, Voide C, Gex-Fabry M, et al. Drug-induced long QT syndrome in injection drug users receiving methadone: high frequency in hospitalized patients and risk factors. *Arch Intern Med*. 2006;166(12):1280–1287.

Elkader A, Sproule B. Buprenorphine: clinical pharmacokinetics in the treatment of opioid dependence. *Clin Pharmacokinet*. 2005;44(7):661–680.

Fiellin DA, Moore BA, Sullivan LE, et al. Long-term treatment with buprenorphine/naloxone in primary care: results at 2–5 years. *Am J Addict*. 2008;17:116–120.

Fiellin DA, Schottenfeld RS, Cutter CJ, Moore BA, Barry DT, O'Connor PG. Primary care-based buprenorphine taper vs maintenance therapy for prescription opioid dependence: a randomized clinical trial. *JAMA Intern Med*. 2014;174:1947–1954.

Fullerton CA, Kim M, Thomas CP, et al. Medication-assisted treatment with methadone: assessing the evidence. *Psychiatr Serv*. 2014;65:146–157.

Gibson A, Degenhardt L, Mattick RP, Ali R, White J, O'Brien S. Exposure to opioid maintenance treatment reduces long-term mortality. *Addiction*. 2008;103:462–468.

Hser YI, Saxon AJ, Huang D, et al. Treatment retention among patients randomized to buprenorphine/naloxone compared to methadone in a multi-site trial. *Addiction*. 2014;109:79–87.

Jones AK, Ngaimisi E, Gopalakrishnan M, Young MA, Laffont CM. Population pharmacokinetics of a monthly buprenorphine depot injection for the treatment of opioid use disorder: a combined analysis of phase II and phase III trials. *Clinical Pharmacokinetics*. 2021;60(4):527–540.

Kerensky T, Walley AY. Opioid overdose prevention and naloxone rescue kits: what we know and what we don't know. *Addict Sci Clin Pract*. 2017;12(1):4.

Lee JD, Nunes EV Jr, Novo P, et al. Comparative effectiveness of extended-release naltrexone versus buprenorphine-naloxone for opioid relapse prevention (X:BOT): a multicentre, open-label, randomised controlled trial. *Lancet*. 2018;391(10118):309–318.

*Ling W, Charuvastra C, Collins JF, et al. Buprenorphine maintenance treatment of opiate dependence: a multicenter, randomized clinical trial. *Addiction*. 1998;93:475–486.

Mattick RP, Kimber J, Breen C, Davoli M. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. *Cochrane Database Syst Rev*. 2014;(2):CD002207.

Saxon AJ, Ling W, Hillhouse M, et al. Buprenorphine/naloxone and methadone effects on laboratory indices of liver health: a randomized trial. *Drug Alcohol Depend*. 2013;128:71–76.

Sigmon SC, Moody DE, Nuwayser ES, Bigelow GE. An injection depot formulation of buprenorphine: extended bio-delivery and effects. *Addiction*. 2006;101(3):420–432.

Strain EC, Bigelow GE, Liebson IA, Stitzer ML. Moderate- vs high-dose methadone in the treatment of opioid dependence: a randomized trial. *JAMA*. 1999;281:1000–1005.

Strain EC, Moody DE, Stoller KB, Walsh SL, Bigelow GE. Relative bioavailability of different buprenorphine formulations under chronic dosing conditions. *Drug Alcohol Depend*. 2004;74:37–43.

Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: Results from the 2019 National Survey on Drug Use and Health (HHS Publication No. PEP20-07-01-001, NSDUH Series H-55). Rockville, MD: *Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration*; 2020. Retrieved from https://www.samhsa.gov/data/

Suzuki J, El-Haddas S. A review: fentanyl and non-pharmaceutical fentanyls. *Drug Alcohol Depend*. 2017;171:107–116.

Tanum L, Solli KK, Latif ZEH, et al. Effectiveness of injectable extended-release naltrexone vs daily buprenorphine-naloxone for opioid dependence: a randomized clinical noninferiority trial. *JAMA Psychiatry*. 2017;74(12):1197–1205.

Thomas CP, Fullerton CA, Kim M, et al. Medication-assisted treatment with buprenorphine: assessing the evidence. *Psychiatr Serv*. 2014;65:158–170.

Verebely K, Volavka J, Mulé S, Resnick R. Methadone in man: pharmacokinetic and excretion studies in acute and chronic treatment. *Clin Pharmacol Ther*. 1975;18:180–190.

Weiss RD, Potter JS, Fiellin DA, et al. Adjunctive counseling during brief and extended buprenorphine-naloxone treatment for prescription opioid dependence: a 2-phase randomized controlled trial. *Arch Gen Psychiatry*. 2011;68:1238–1246.

33.13a.2 Opioid Receptor Antagonists: Naltrexone and Nalmefene

Oluwole O. Jegede, M.D., M.P.H., Stephanie S. O'Malley, Ph.D., Tiffany Arnold, B.S., M.D., Katherine S. Pier, M.D., and Srinivas B. Muvvala, M.D., M.P.H.

INTRODUCTION AND HISTORY

Naltrexone is an opioid antagonist—synthesized in 1965 by H. Blumberg and H. B. Dayton—which proved to have several advantages over earlier opioid antagonists, including cyclazocine, nalorphine, and naloxone. Specifically, naltrexone had a relatively long half-life, was orally effective, was not associated with dysphoria, and could be administered once a day. Whereas earlier narcotic antagonists—principally naloxone—were used to reduce narcotic overdose, these characteristics of naltrexone made it suitable for preventing relapse to opioid use in detoxified individuals with opioid use disorder (OUD). In 1984, the U.S. Food and Drug Administration (FDA) approved naltrexone (under the original trade name Trexan) for providing blockade of the effects of exogenously administered opioids after detoxification.

Over the ensuing years, naltrexone has been tried for the treatment of a wide range of disorders, including, among others, alcohol use disorders, eating disorders, autism, self-injurious behavior, cocaine use disorder, gambling, and chronic pain syndromes.

Naltrexone (ReVia; Depade) was approved as an oral formulation to treat alcohol use disorder in 1994. Generic formulations of oral naltrexone have been available since then. In addition, an extended-release, once-a-month injectable suspension (XR-NTX, Vivitrol) was approved in 2006 to treat alcohol use disorder and in 2010 to treat OUD.

In 2014, the FDA approved an extended-release (XR) formulation combining naltrexone hydrochloride and bupropion (Contrave) to treat obesity.

Nalmefene (Revex) is indicated for the complete or partial reversal of opioid drug effects and managing known or suspected opioid overdose and alcohol use disorder (Europe). An oral formulation of Nalmefene is available in some countries but not in the United States. Naloxone, an opioid antagonist, has been FDA approved for reversal of opioid overdose. It is an effective medication in reversing respiratory depression and preventing fatalities secondary to opioid overdose.

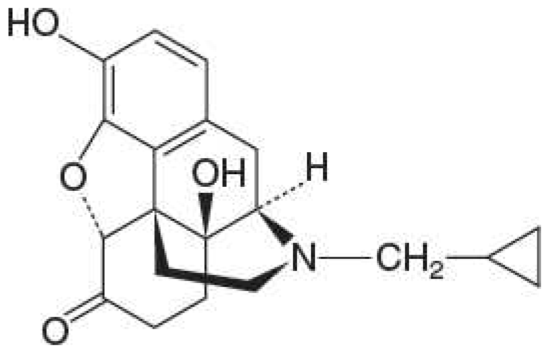

FIGURE 33.13a.2–1. Molecular structure of naltrexone (ReVia).

CHEMISTRY

Naltrexone, or 17-(cyclopropylmethyl)-4,5a-epoxy-3,14-dihydroxymorphinan-6-one hydrochloride, is a competitive opioid antagonist developed by substituting an N-methyl group on the opioid agonist oxymorphone with a methylcyclopropyl moiety. Naltrexone is a white, crystalline hydrochloride salt soluble in water, with a melting point of 275 °C and a pKa of 8.13 at 37 °C. Its molecular formula is $C_{20}H_{23}NO_4$•HCl. Nalmefene is a 6-methylene analog of naltrexone with a molecular formula $C_{21}H_{25}NO_3$, and it is also available as a hydrochloride salt with a pKa of 7.6 (Fig. 33.13a.2–1).

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Naltrexone is rapidly and completely absorbed following oral administration of a 100-mg dose, with peak plasma concentrations of 19 to 44 mg/L achieved within 1 hour. The oral bioavailability of naltrexone is 60%, with 20% of the drug bound to plasma proteins. A single dose of naltrexone has a volume of distribution of 16.1 L/kg, which decreases to 14.2 L/kg following repeated administration.

Naltrexone is primarily recovered in plasma, urine, and feces as conjugated naltrexone and conjugated and unconjugated 6β-naltrexol along with minor quantities of 2-hydroxy-3-O-methyl-6β-naltrexol and 2-hydroxy-3-O-methylnaltrexone. Approximately 60% of the drug is recovered in the urine 48 hours following oral administration and 72 hours following intravenous (IV) administration. Naltrexone has a plasma half-life of 10.5 hours regardless of whether it is administered via the oral or IV route. Conjugated 6β-naltrexol has been shown to have a plasma half-life of 14 to 19 hours. This metabolite may contribute to the longer duration of action of naltrexone. Studies indicate that after a 100-mg dose, naltrexone can be detected in plasma for 30 hours, whereas 6β-naltrexol can be detected for up to 125 hours. The XR injectable formulation (Vivitrol), containing 380 mg of naltrexone, was created using Medisorb drug-delivery technology for once-a-month intramuscular (IM) injections. Following administration of this formulation, an initial peak is seen within 2 hours, followed by a second peak 2 to 3 days later, with a slow decline in concentrations after 14 days. Thus, total naltrexone exposure is three- to fourfold higher than that observed with daily oral dosing over 28 days. The elimination half-life of both naltrexone and β-naltrexol in this form is 5 to 10 days.

Nalmefene, following administration of a 1-mg IV dose, has a volume of distribution of 8.6 L/kg and a terminal half-life of 10.8 hours. Nalmefene is completely bioavailable following IM and subcutaneous (SC) routes, with time-to-peak plasma concentrations of 2.3 (±1.1) and 1.5 (±1.2) hours, respectively. Because of variability in the speed of absorption of IM and SC doses, repeated dosing should be done with great care. Nalmefene is primarily metabolized through glucuronide conjugation in the liver and is excreted in the urine. Both the glucuronide and the dealkylated metabolite (formed in trace amounts) have minimal pharmacologic activity. Therefore, when administered orally, Nalmefene has a pharmacokinetic profile similar to that of naltrexone. In addition, a metabolite, nornalmefene, has been identified that also exists in glucuronidated and alkylated forms. Hepatic or renal impairment decreases the clearance of both naltrexone and nalmefene.

Pharmacodynamics

Naltrexone is a nonspecific opioid antagonist that binds to all three opioid receptor sites (μ, κ, δ) due to the dose administered. Naltrexone has minimal opioid agonistic effects, and its relative opioid antagonistic potency is 12 times that of nalorphine and 2.5 times that of naloxone or cyclazocine. The time course of action of oral naltrexone is about three times longer than naloxone in rodents. No significant species differences in the binding of [^{3}H] naltrexone have been documented. An oral 100-mg dose of naltrexone was found to almost completely antagonize the subjective and objective responses of 25 mg of IV heroin.

Similarly, following a 1-mg IV dose, nalmefene blocks greater than 80% of brain opioid receptors within 5 minutes after administration. Whereas nalmefene has been shown to have greater antagonistic potency at the δ-opioid receptors, it has also been shown following oral administration of a dose of 20 mg to produce complete and prolonged blockade of the μ-opioid receptors. Recent evidence based on the rapid onset of action, long half-life (>7 hours), and the high affinity of nalmefene at μ-opioid receptors has led to the development of an intranasal (IN) formulation as a rescue medication that could be useful in the context of synthetic opioid overdose. The pharmacokinetic profile of nalmefene was enhanced by the addition of the absorption enhancer dodecyl maltoside.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Naltrexone and nalmefene have demonstrated efficacy in blocking exogenously administered opioids and were used in protocols for rapid opioid detoxification. Both agents have also been extensively examined as aids in preventing relapse in patients with alcohol use disorder. Both agents, particularly naltrexone, have shown benefits in reducing heavy drinking in numerous studies. XR formulations of naltrexone have been evaluated to prevent relapse in both alcohol and OUDs. Naltrexone has also been evaluated to treat self-injurious behaviors in patients with intellectual disability and autism, based only on small studies and case reports. Case reports and small open-label trials have also examined naltrexone's utility in treating compulsive sexual behavior and kleptomania. Open-label and randomized controlled trials using relatively small samples have examined the efficacy of naltrexone in reducing binge eating. Randomized placebo-controlled trials have evaluated the extended-release combination of naltrexone and bupropion (NB) to treat obesity, resulting in FDA approval for this indication. Some randomized clinical trials support the use of nalmefene and naltrexone for pathologic gambling disorders. Naltrexone has been evaluated for cocaine use disorders without much success. Although randomized clinical trials suggest the possible utility for naltrexone as adjunctive therapy for smoking cessation, the effects are inconsistent and may vary by gender. Small Randomized Controlled Trials (RCTs) have also demonstrated a limited anti-inflammatory efficacy of low-dose naltrexone (LDN) in patients with chronic pain syndromes.

THERAPEUTIC INDICATIONS

Alcohol Use Disorders

The efficacy of naltrexone was initially demonstrated in the 1990s in two double-blind, placebo-controlled trials of naltrexone in alcohol-dependent patients. In each of these studies, treatment with an oral dose of 50 mg daily was initiated following a minimum of 7 days of abstinence or completion of medical detoxification and was provided with concurrent psychosocial interventions. In a combined analysis of these studies, naltrexone-treated patients were more likely to remain abstinent (54% vs. 31%), to experience fewer days of drinking (3.1% vs. 7.5%), and to report lower rates of relapse to heavy drinking (25% vs. 52%). Furthermore, participants who lapsed (i.e., were nonabstinent) were less likely to experience heavy drinking if they took naltrexone rather than a placebo. The COMBINE Study, a large multisite trial combining pharmacotherapy and behavioral interventions, conducted in 11 academic sites across the

United States showed that a 100-mg dose of naltrexone, with medical management, was more effective than medical management placebo in increasing percent days of abstinence and reducing heavy drinking days. In a meta-analysis of pharmacotherapies for alcohol use disorder, 53 trials studied the efficacy of naltrexone (*n* = 9,140), and treatment duration ranged from 12 to 52 weeks. Naltrexone was shown to have benefits in various consumption outcomes. Naltrexone is particularly shown to have efficacy in preventing the return to heavy drinking and reducing percent drinking days and percent heavy drinking days.

An XR formulation of naltrexone has also shown similar efficacy in reducing the rate of heavy drinking. A large multisite double-blind placebo-controlled study evaluated the efficacy and tolerability of IM long-acting naltrexone injections (one group was administered 380 mg and the other was administered 190 mg) versus placebo in alcohol-dependent patients. The long-acting preparation was well tolerated over 6 months, and the 380-mg injectable naltrexone preparation showed a 25% reduction in the number of heavy drinking days. In the subset of patients who were abstinent 4 or more days prior to treatment initiation, the 380-mg preparation also prolonged abstinence and reduced the number of heavy drinking days and drinking days. Subjects who received 380 long-acting naltrexone mg versus placebo and had lead-in abstinence were shown to have a greater effect size (HR, 0.20, *p* = .005) when compared to subjects who drank during the lead-in period (HR, 0.79, *p* = .05). This increase in effect size was also observed in subjects who received 190-mg long-acting naltrexone versus placebo (HR, 0.79, *p* = .05). Thus, while naltrexone effectively reduced the number of heavy drinking days in those using alcohol prior to commencing naltrexone, the treatment effects were greater in patients abstaining from alcohol consumption prior to treatment.

The dosing interval recommended for the administration of XR-NTX in the United States is 4 weeks. However, it should be noted that there may be variation in efficacy with this recommendation, and flexibility in dosing may need to be considered on a case-by-case basis. For example, a case report on a patient with cyclical relapse showed significant improvement when the dosing interval was shortened from 4 to 3 weeks, highlighting that flexibility in dosing intervals may be beneficial for some patients whose cravings and drinking behaviors return before 4 weeks have elapsed.

Despite substantial literature supporting the use of naltrexone to reduce drinking behavior, this agent's efficacy is modest. This evidence has motivated researchers to examine predictors of response to naltrexone. Recent evidence suggests that the efficacy of naltrexone may be more pronounced in specific subgroups of patients, such as those for whom alcohol use is primarily driven by the positive, rewarding effects of alcohol and those who smoke cigarettes. Individuals with high levels of alcohol cue–induced craving also seem to benefit from naltrexone. The predictive utility of the Asn40Asp polymorphism of the µ-opioid receptor (OPRM1) gene in treatment response is being debated, with some studies showing a stronger effect of naltrexone in Asp40 carriers while studies prospectively randomizing patients based on genotype do not. As a result, the data are still insufficient to recommend selecting patients for treatment with naltrexone based on these moderators.

Nevertheless, because of the good safety profile of naltrexone and support for its effectiveness in a range of patient populations, naltrexone should be offered as one of the treatment options to all patients with alcohol use disorder who do not have contraindicating conditions. Interestingly, there are small clinical studies that demonstrate increased responsiveness among patients that have an affinity for sweet taste (sweet-liking phenotype [SL phenotype]) versus patients who dislike sweet take (sweet-disliking phenotype [SDL]). It is also noted that the SL phenotype and high craving for alcohol both separately and especially in combination are associated with a positive response to naltrexone.

In 2013 a four-site randomized controlled trial known as the PREDICT study postulated that individuals whose drinking was driven by positive reinforcement (i.e., reward drinkers) would be more responsive to naltrexone, as a reward is thought to be partially mediated by the opioidergic system. Using the Inventory of Drinking Situations, a 30-item self-report questionnaire, subjects were categorized into high-reward/high-relief, high-reward/low-relief, low-reward/high-relief, and low-reward/low-relief phenotypes. The resulting data showcased that using naltrexone with predominantly reward drinkers produced significantly larger effect sizes than prescribing the medication to a more heterogeneous sample. In addition, the interaction between naltrexone and the reward subgroup significantly predicted any heavy drinking ($p = .03$). This finding was not replicated when the data from the COMBINE study were analyzed, however, which may be because, in this study, naltrexone was generally more effective for most patients with a significant main effect of medication.

Nalmefene was approved in 2013 by the European Medicines Agency to reduce alcohol consumption in patients with high-risk drinking who do have alcohol withdrawal symptoms and who do not require alcohol detoxification. The studies supporting the approval tested nalmefene taken as needed (i.e., in high-risk situations, when drinking is imminent, or as soon as possible after drinking) compared to placebo. Although fewer studies have been on nalmefene, a recent meta-analysis concluded that nalmefene led to a reduction in heavy drinking days per month and the number of drinks per drinking day.

Several small trials have also examined the combined use of opioid antagonists with other agents. For example, preliminary small placebo-controlled studies suggest that a combination of the 5-HT$_3$ antagonist ondansetron and naltrexone, when compared with placebo, may produce greater suppression of drinking outcomes in individuals with early-onset alcohol use disorders but not in those with late-onset alcohol use disorder.

Combining naltrexone with disulfiram has shown modest efficacy in increasing abstinence rates in concurrent psychiatric disorders and alcohol use disorder patients. A 12-week randomized clinical was conducted to evaluate the efficacy of disulfiram and naltrexone, alone and in combination, in participants with dual diagnoses. However, a large, multisite trial indicated no significant advantage of using naltrexone in combination with acamprosate to treat alcohol use disorder. A combination of LDN and prazosin has shown some promise in patients with comorbid post-traumatic stress disorder (PTSD). The addition of gabapentin to naltrexone has been shown to improve drinking outcomes than naltrexone alone in the first 6 weeks of abstinence, possibly secondary to beneficial effects with symptoms such as insomnia and mood instability.

Opioid Use Disorder Treatment

Although naltrexone was initially developed to treat opioid addiction, the value of oral naltrexone for this indication has been limited primarily by two issues: (1) the need for opioid detoxification prior to induction and (2) poor adherence compared to agonist therapy. The newer XR formulations address to a large extent the issue of adherence at least for the 30 days between injections.

As an opioid antagonist, naltrexone selectively competes with exogenous opioids for the central nervous system (CNS) and non-CNS opioid receptors and blocks their activity. This blockade lasts for up to 72 hours after oral naltrexone administration and prevents patients from achieving sought-after reinforcement from opioid use. This blockade can have positive psychological consequences, including reduced craving related to the reduced immediate availability of opioid effects and deconditioning of cue-related craving as patients are exposed to people, places, and things formerly associated with opioid use. The opioid antagonist properties of naltrexone do not appear to be subject to the development of tolerance because potency has been demonstrated after maintenance for 1 year or more. Thus, naltrexone does not have opioid agonist properties, and patients do not obtain a psychoactive reward ("high") from using naltrexone. It is also not associated with significant withdrawal symptoms if use is discontinued. However, patients may be more sensitive to lower doses of opioids once naltrexone is discontinued due to the loss of tolerance to exogenous opioids. This decrease in tolerance has been found in animal studies where chronic opioid antagonism with naltrexone results in an increased density of opioid receptors in the brain, thereby increasing sensitivity to opioid agonists relative to the pretreatment naltrexone state.

In certain highly motivated patients with OUD, it can be used as an alternative to maintenance treatment with opioid agonists, such as methadone or the partial agonist buprenorphine. As a nonscheduled medication without psychoactive or addictive properties, naltrexone can be administered as a treatment in various medical settings without any special licensing requirements. Because diversion is not an issue, take-home medications can be provided. As a drug without agonist effects that block the effects of opioids, naltrexone can be used only after completion of a 7- to 10-day period of complete opioid abstinence or after a rapid detoxification procedure to avoid precipitating opioid withdrawal. Hence, patients must start treatment to achieve total opioid abstinence and a willingness to undergo detoxification. Unlike patients maintained on opioid agonists, patients on naltrexone achieve no rewarding effects of taking medications and experience no withdrawal effects if doses are missed. It should be noted that oral naltrexone is not FDA approved for OUD, and maintenance treatment with naltrexone for OUD is with the long-acting formulation only. Retention in treatment is at least two times higher for subjects given XR preparations compared to oral.

Indeed, oral naltrexone has been associated with high rates of discontinuation. A Cochrane review of randomized controlled clinical trials of oral naltrexone maintenance treatments for opioid addiction suggested that oral naltrexone was not more effective than placebo or no pharmacologic agent in preventing relapse, mainly due to low adherence and retention rates. Greater success has been achieved with patients motivated by powerful contingencies such as opioid-dependent health care professionals for whom opioid abstinence is a condition for regaining medical licensure or incarcerated individuals for whom treatment is a requirement for probation or parole. Other strategies to enhance compliance include the use of voucher-based contingency management and the dispensing of naltrexone that is directly observed by either a medical professional or a significant other. The importance of retention in therapy for opioid addiction is highlighted by a review of coronial data in Australia in which those receiving oral naltrexone were 7.6 times more likely than those receiving opioid agonist treatments to experience an overdose. XR formulations, both injectable and surgically implantable, address some of the shortcomings of the oral naltrexone related to adherence and retention. The FDA approved the XR injectable formulation of naltrexone (XR-NTX) in 2010 to treat and prevent relapse after patients with OUD have undergone detoxification treatment. A double-blind placebo-controlled multicenter trial in Russia showed that XR-NTX in conjunction with psychosocial treatment is efficacious in preventing relapse on opioids in detoxified patients. The median proportion of weeks of abstinence was 90% in the 380-mg XR-NTX group versus 34% in the placebo group. In addition, treatment retention was much higher, and opioid craving was much lower in the active medication group. The 1-year open-label extension phase of this study showed that 62.3% of clients remained compliant, out of which 50.9% of patients remained abstinent from opioid use, and the safety of XR-NTX is further established during this phase. A recent study demonstrated the usefulness of XR-NTX as a therapeutic option in treating OUD in individuals involved with the criminal justice system. Opioid-dependent patients maintained on XR-NTX were shown to have lower inpatient substance use–related utilization and total cost of treatment than methadone. It should also be noted that where there is an external motivation for adherence, naltrexone may compare to opioid agonist medications.

Newer sustained-release formulations have been developed, such as naltrexone implants, but further research is needed to establish their safety and efficacy in clinical use.

The XR-NTX versus buprenorphine–naloxone for opioid relapse prevention (X-BOT) trial shows that XR-NTX was more difficult to initiate than sublingual buprenorphine–naloxone and associated with higher relapse rates in the overall sample; however, the two treatments were equally effective for those who were successfully inducted onto treatment. In addition, XR-NTX appears generally well tolerated, and concerns about hepatotoxicity have not been realized. No differences have been found between XR-NTX and placebo on the proportion of patients showing elevated liver enzymes, and XR-NTX is safe in patients with mild to moderate chronic Hepatitis C Virus (HCV) or HIV infections. Opioid detoxification is critical prior to naltrexone induction to avoid precipitating opioid withdrawal. This detoxification can be done by a gradual opioid or buprenorphine taper or by more rapid discontinuation with the help of adjunctive nonopioid medications for symptomatic relief. The latter is more suitable for patients with low physiologic dependence. For patients at risk of moderate to severe withdrawal symptoms, detoxification is best carried out in an inpatient detoxification setting or a partial hospital program with inpatient backup.

Other protocols that have shown promise in the induction of patients with OUD onto XR-NTX use small ascending doses of oral NTX in addition to adjunctive medications, including clonidine and benzodiazepines. In one example of a transition from buprenorphine, along with a brief buprenorphine taper, patients were started with 0.25 mg oral naltrexone twice daily on day 1, increasing gradually to 15 mg twice daily on day 7 and administering XR-NTX on day 8 following negative naloxone challenge. RCTs show that these methods can mitigate the period of relapse risk prior to XR-NTX by obviating the traditional requirement to maintain opioid abstinence before induction. For example, a 2017 randomized trial assigned 150 opioid-dependent adults in a 2:1 ratio to naltrexone-assisted detoxification or buprenorphine-assisted detoxification, followed by an injection of XR-NTX. Naltrexone-assisted detoxification lasted 7 days and included a single day of buprenorphine followed by ascending doses of oral naltrexone along with adjunctive medications for opioid withdrawal symptoms. Buprenorphine-assisted detoxification included a 7-day buprenorphine taper followed by a week-long delay before administration of XR-NTX, consistent with official prescribing information for XR-NTX. Participants from both groups received behavioral therapy focused on medication adherence and the second dose of XR-NTX.

Opioid Detoxification Protocols in Combination with Naltrexone

As a method for rapid detoxification, outpatient protocols have been developed combining naltrexone with clonidine, a selective α2-adrenergic antagonist initially used as an antihypertensive agent, and other adjunctive medication. This combined treatment allows patients to be completely opioid free from the first day of treatment, to have completed detoxification within 48 to 72 hours, and to be able to commence naltrexone maintenance at that time. As a method of initiating maintenance naltrexone treatment, this detoxification mode removes the need for an unprotected 7- to 10-day opioid-free period prior to starting naltrexone, a time when vulnerability to relapse is high. Because blood pressure may be substantially lowered by clonidine administration, patients must be monitored for up to 8 hours on the initial detoxification day, limiting this procedure to settings in which treatment rooms and adequate nursing staff are available. As this procedure may result in significant withdrawal symptoms, this method is best used in patients with a low level of physiologic dependence on opioids.

A variant method of “ultrarapid” inpatient detoxification involves administering naltrexone to opioid-dependent patients who are placed under light sedation or general anesthesia (including mechanical ventilation) followed by detoxification using full doses of naltrexone. However, this method was associated with more incidences of life-threatening adverse events, and therefore this form of “ultrarapid” detoxification cannot be recommended at this time.

Opioid Overdose Prevention

Deaths from opioid overdoses have reached epidemic levels in the United States. Naloxone hydrochloride (Narcan), a potent opioid antagonist, has been widely used in various health care settings and by emergency medical personnel to reverse an opioid overdose and the associated respiratory depression. Multiple routes can administer naloxone in a hospital setting, but it can be administered intramuscularly or intranasally in the community. Some community-based overdose prevention programs have been distributing naloxone to patients at risk of overdose since 1996. Subsequently, several agencies have trained clinicians to screen patients for overdose risk and educate them and their significant others in recognizing and responding to overdoses. With an increase in these initiatives focusing on naloxone distribution, there have been reports of thousands of overdose rescues in the United States. As of August 2020, all 50 states and the District of Columbia have some form of a naloxone access law, allowing individuals to obtain naloxone without a prescription in some capacity. A total of 17 states have a provision related to the co-prescribing of naloxone with an opioid. In nine of these states, the co-prescribing of naloxone is required in certain situations, such as prescribing opioids over a specific morphine milligram equivalent (MME), concurrent benzodiazepine use, or patient history of OUD or overdose. In addition, the FDA announced it would require labeling and prescribing information for opioid pain medicine and opioid agonist treatment to be updated to recommend that health care professionals discuss the availability of naloxone with patients and caregivers, both when beginning and renewing treatment. The FDA also recently approved higher doses of naloxone hydrochloride nasal spray to combat the escalating opioid overdose rates associated with highly potent fentanyl and its analogs.

Cocaine Use Disorders

Based on preclinical studies showing that opioid antagonists can attenuate many effects of cocaine, including acquisition and maintenance of self-administration, a randomized, placebo-controlled clinical trial evaluated its efficacy in an oral 50-mg/day dose alone and combination with relapse prevention therapy. Naltrexone was associated with modestly improved cocaine use outcomes only when combined with relapse prevention therapy. However, trials in patients with cocaine and alcohol use disorders found no benefit of oral naltrexone in reducing cocaine use. In addition, a prior open-label study and another double-blind placebo-controlled study found disulfiram to be superior to naltrexone for patients with combined cocaine and alcohol use disorders. Hence, the evidence does not justify a recommendation for oral naltrexone use in cocaine dependence.

Methamphetamine Use Disorder

Emerging evidence suggests the efficacy of XR-NTX in combination with XR bupropion for the treatment of individuals with moderate and severe methamphetamine use disorder. A randomized, double-blind trial designed to evaluate the efficacy and safety of XR injectable naltrexone (380 mg every 3 weeks) combined with oral XR bupropion (450 mg per day) as compared with placebo in adult patients (18 to 65 years old) with methamphetamine use disorder. This combination was associated with a small but significant negative methamphetamine urine drug screen compared to placebo. Adverse events with naltrexone–bupropion included gastrointestinal disorders, tremor, malaise, hyperhidrosis, and anorexia. However, serious adverse events occurred in only 3.6% who received naltrexone–bupropion during the trial.

Another randomized control trial investigated the use of XR-NTX on its own in reducing methamphetamine use and high-risk sexual behaviors among methamphetamine-dependent men who have sex with men compared with placebo. While initial studies showed possible promise for using naltrexone alone in treating methamphetamine use disorder, this study and others have cast significant doubt on these findings. Therefore, naltrexone in the absence of bupropion is not currently recommended for treating methamphetamine use.

Nicotine Use Disorders

Several lines of evidence suggest that the opioid system may be involved in tobacco use disorder. Nicotine administration is known to stimulate the release of endogenous opioids and attenuate pain perception via opioid mechanisms. Clinical laboratory studies examining the effects of opioid antagonists, including naloxone and naltrexone, on smoking behavior have been inconclusive, with some studies showing reductions in smoking and others finding no effect.

Prior clinical studies suggested that naltrexone may reduce urges or the desire to smoke and reduce weight gain following smoking cessation. The clinical utility of naltrexone as an aid for smoking cessation is supported by two randomized clinical trials, both of which tested the efficacy of naltrexone when used in combination with the nicotine patch for smoking cessation. One study found preliminary efficacy for 50-mg naltrexone primarily in women. The other trial, which combined naltrexone (25 to 100 mg/day) with a 21-mg nicotine patch, found some advantages for the 100-mg dose of naltrexone as an augmentation strategy for smoking cessation. However, a recent meta-analysis of opioid antagonists for smoking cessation concluded that naltrexone alone or as an adjunct to nicotine replacement has no beneficial effect in short- or long-term smoking abstinence. The two clinical trials mentioned above found significant reductions in weight gain following smoking cessation. A third trial in weight-concerned smokers found a modest weight reduction but somewhat lower smoking cessation success with 25 mg naltrexone combined with the nicotine patch.

Several studies have tested the effectiveness of naltrexone in combination with existing approved smoking cessation medications (bupropion or nicotine replacement therapy) to decrease the amount of weight gain following smoking cessation. The studies all measured outcomes within 52 weeks of cessation, a period consistently associated with the most significant postcessation weight gain. In addition, five out of seven studies looking at naltrexone as an adjunct to bupropion or nicotine replacement therapies found a reduction in post–smoking cessation weight gain, while this reduction was not statistically significant in two other studies. Thus, naltrexone may hold promise in moderating weight gain associated with smoking cessation.

Self-Injurious Behavior

Naltrexone has been proposed to treat self-injurious behavior and autistic behaviors in patients with mental retardation and autism. Clinically significant symptom reduction has been reported from numerous open case studies and trials with small sample sizes. However, more extensive studies using more rigorous experimental designs have failed to replicate these reports. Hence, naltrexone's utility for these symptoms is not established.

Eating Disorders/Obesity

Numerous studies in rodents have suggested an association between altered endogenous opioid activity and obesity and other eating disorders. Exogenous and endogenous opioid administration increases food intake, whereas opioid antagonists block the increase in food intake. The clinical data on the use of opioid antagonists in eating disorders is controversial, and the literature is replete with studies that demonstrate differential effects of these agents based on the type of eating disorder. For example, a laboratory study demonstrated that naloxone reduces the consumption of sweet, high-fat foods in both obese and lean binge eaters compared with obese and lean nonbingers suggesting a role of the opioid system in binge eating but not in obesity. In bulimia nervosa, a handful of small sample-controlled clinical trials have been conducted, with some, but not all, reporting reductions in binge symptomatology. Regarding anorexia nervosa, naloxone infusion produced a 10-fold increase in weight gain per week, an effect that was ended when the naloxone infusion was stopped in patients who were also concurrently treated with the antidepressant amitriptyline. Therefore, the role of naltrexone and other opioid antagonists in treating eating disorders remains unclear, and more extensive and definitive, controlled clinical trials are needed.

Naltrexone, when used alone, has failed to produce clinically significant weight loss in controlled studies. However, in translational studies, the combination of NB produced a synergistic reduction in rodent food consumption. This reduction was further proven by studies in overweight and obese individuals in which weight loss produced by the combination of bupropion and naltrexone was shown to be greater than the sum of individual monotherapies. A large, multicenter, randomized, double-blind, placebo-controlled trial showed that NB is a useful therapeutic option for treating obesity. The mean change in body weight was −6.1% in the sustained-release naltrexone 32 mg plus sustained-release bupropion 360-mg/day group (NB32) and −5% in the sustained-release naltrexone 16 mg plus sustained-release bupropion 360-mg/day group (NB16) as compared to −1.3% in placebo. Similarly, 48% of participants assigned to the NB32 group and 39% of participants assigned to the NB16 group had greater than 5% weight loss compared to only 16% in the placebo group. Another large 56-week, randomized, double-blind, placebo-controlled study in obese or overweight participants with dyslipidemia or hypertension compared NB32 to placebo. 55.6% of the participants treated with NB32 had greater than 5% weight loss compared to 17.5% in the placebo group. This study also showed that NB32 improved weight-related quality of life, control of eating, and various cardiometabolic risk markers without any increase in depression or suicidality. Another double-blind placebo-controlled study in overweight/obese type 2 diabetics showed that NB was associated with improvements in glycemic control and select cardiovascular risk factors, and promoting weight loss. FDA recently, in 2014, approved the combination of naltrexone hydrochloride and bupropion hydrochloride in an XR formulation for the treatment of obesity.

Naltrexone showed promise in reducing long-term smoking cessation–related weight gain in women. In addition, in recent pilot studies, naltrexone has shown promise in arresting antipsychotic-associated weight gain in patients diagnosed with schizophrenia and schizoaffective disorder. It is hypothesized that naltrexone could help reduce food cravings in this population by interacting with the dopamine–opioid system.

Few evidence-based psychopharmacologic treatments of eating disorders in adolescents exist. However, a case series in 2019 demonstrated that 50- to 100-mg naltrexone in 32 adolescents undergoing outpatient eating disorder treatment was well tolerated and reduced vomiting and urges to vomit in 52% of the patients. In addition, 18% of patients experienced a decrease in urge to binge/purge. Most importantly, naltrexone was well tolerated without significant adverse effects in this small sample treated for a median of 90 days. Given the paucity of effective pharmacotherapies for eating disorders in adolescence, the potential of naltrexone for this indication warrants further investigation.

Pathologic Gambling

Opioid antagonists can decrease dopamine release in the nucleus accumbens by disinhibiting GABAergic input to dopamine neurons in the ventral tegmental area. As such, it has been hypothesized that these agents might be effective in treating impulse control disorders in general and pathologic gambling in particular. Two open-label case series and a moderately sized, placebo-controlled, randomized, double-blind study supported the efficacy of naltrexone in the short-term treatment of pathologic gambling. In these studies, naltrexone was used at higher doses than generally used in treating alcohol or OUDs. For example, in the placebo-controlled, double-blind study, the dose of 250 mg/day was targeted, and completers had an end-of-study dose of 188 mg/day. Consistent with the use of these higher doses, adverse effects were not infrequently reported, the most significant of which included elevated liver function tests (LFTs). An examination of the individuals experiencing naltrexone-induced increases in LFTs suggested that concurrent use of nonsteroidal anti-inflammatory drugs (ibuprofen, aspirin) was associated with LFT increase. A double-blind placebo-controlled trial demonstrated that 50-, 100-, and 150-mg doses of naltrexone were equally efficacious in reducing gambling urges and behavior in pathologic gambling.

There has been more research examining the use of nalmefene for the treatment of pathologic gambling. Two multicenter investigations suggested that the use of 25 mg of nalmefene daily may effectively reduce the severity of gambling in those with pathologic gambling disorder; there was a 59% improvement in those who received nalmefene versus 34% in those who received placebo. However, higher doses of nalmefene (50 and 100 mg/day) were associated with intolerable side effects. Another double-blind placebo-controlled showed that at doses of 20 and 40 mg, nalmefene was not superior to placebo in the treatment of pathologic gambling.

Currently, there is no FDA-approved medication for the treatment of gambling disorders. A recent meta-analysis investigating the efficacy of pharmacologic treatments found that opioid antagonists have a small but significant benefit than placebo in treating pathologic gambling. However, the study did not support the use of the opioid antagonists, citing that some of the studies have not adhered to intent-to-treat principles. Further studies are needed to examine the safety and efficacy of nalmefene and naltrexone in treating pathologic gambling, especially in individuals with co-occurring substance use disorders and in combination with nonpharmacologic treatment modalities.

Chronic Pain

Long-term opioid treatment leads to a compensatory response from the brain, which increases pain drivers such as glutamate, dynorphin, corticotrophin-releasing factor, and substance P to compensate for chronic agonism of μ-opioid receptors. This brings about a paradoxical worsening of pain and diminished pain tolerance, resulting in the syndrome of opioid-induced hyperalgesia (OIH). Although this phenomenon of sensitization of nociceptors is known to be caused by exposure to opioids, the mechanism of action is still not fully understood. Consequently, although there have been numerous proposed theories about compensatory/allostatic changes of the pain receptors, no consensus has been reached about the exact mechanism. Some clinical research to date suggests that the mechanism of action is thought to involve inflammatory processes.

While the mechanism by which naltrexone restores endogenous opioid tone remains unknown, it is thought that LDN may impact neuroimmunomodulation by intermittent blockade of opioid growth factor receptor (OGFr), resulting in increased levels of opioid growth factor (OGF). The increase of OGF is associated with decreasing neuronal damage, inflammation, and proliferation of T and B cells. In addition, LDN may correct a disruption to the OGF-OGFr axis rather than simply increasing enkephalin levels alone. Case studies have shown that after 3 months of treatment with LDN, individuals with OIH had improved pain tolerance.

LDN has shown promise as a potential treatment for patients with fibromyalgia (FM). FM is a syndrome of chronic pain that is generalized yet intensified at multiple sensitive points and accompanies other somatic and cognitive symptoms. It is hypothesized that the endogenous opioid system plays a role in FM, as cerebrospinal fluid (CSF) studies of FM patients have shown elevated κ-opioid peptide, dynorphin, and met-enkephalin-Arg-Phe. These studies suggest that receptor modulation and desensitization may be contributory. Furthermore, Positron Emission Tomography (PET) scans of patients with FM showcase significantly decreased μ-opioid binding in the bilateral nucleus accumbens, left amygdala, and right anterior dorsal cingulate. Thus, FM involves an autoimmune process in which μ-opioid receptor dysfunction may cause a compensatory increase in endogenous opioids. Despite this compensation, the diminished binding of μ-opioid receptors results in brain-mediated pain experienced by patients as occurring diffusely over the body.

A recent case series found that low doses of naltrexone, up to 4.5 mg twice a day, restores endogenous opioid tone in OIH and improves it in FM to a lesser degree. The difference in response noted between OIH and FM may be due to FM having a distinct neuroimmunologic component from OIH. In addition, a case series suggested that the variation of an inflammatory component may indicate why FM patients have a significant but more minor response to LDN compared to OIH patients.

Immune System

Studies suggest that within a specific dosage window, LDN can act as an immunomodulator in malignant tumors and multiple autoimmune diseases, such as multiple sclerosis (MS), Crohn disease, FM, acquired immunodeficiency syndrome (AIDs), and type

1 diabetes. Currently, evidence supports the safety and tolerability of LDN in these diseases, but fewer studies support the efficacy of LDN, and most efficacy studies focus on subjective measures such as quality of life or self-reported pain.

The suggested mechanisms of action for LDN in autoimmune diseases are mainly related to endogenous opioid activity. Thus, naltrexone is thought to broadly stimulate immune cells by blocking the binding of endogenous opioids and opioid receptors, which may have a role in tissue repair and immunologic effects.

In a randomized controlled trial, 40 participants with moderately to severely active Crohn disease were randomized to receive naltrexone (4.5 mg daily) or a placebo. Eighty percent of patients treated with naltrexone had a 70-point decrease in Crohn’s Disease Activity Index (CDAI) score compared to 40% in the placebo group. Although promising, more extensive studies are needed before naltrexone can be recommended in patients with CD.

In MS, it is hypothesized that LDN reduces apoptosis in oligodendrocytes. A retrospective chart review of 215 patients’ charts found that almost 60% of the patients reported a reduction in fatigue, and 75% reported a subjective increase in quality of life. In addition, two placebo-controlled double-blinded crossover studies found significant effects on quality of life measures; however, a similar study from Iran found no effect.

A retrospective study was conducted that followed patients with relapsing–remitting MS on LDN versus placebo for 10 years and helped establish through serial MRIs and laboratory testing that LDN appeared to be nontoxic and safe in managing symptoms of MS without exacerbating the disease course. At this time, randomized clinical studies are needed to determine whether LDN has any beneficial effects in modifying the course of MS or not, but it may help reduce symptoms without causing significant deleterious effects.

A single 9-month randomized clinical trial compared the results of patients with AIDS on three antiretroviral treatments alone versus these three antiretrovirals plus LDN. This trial demonstrated that subjects randomized to treatment with LDN had a 37.06% higher recovery rate of CD4 count at 6 months (*p* = .041), and a 31.62% higher recovery rate of CD4 count at 9 months (*p* = .067) compared to placebo. Despite these promising results, further research is needed.

There is currently no FDA approval for this indication.

Anxiety and Depression

Currently, data on ultralow-dose naltrexone (ULDN) or LDN are scarce, but anecdotal evidence and clinical trials of LDN in different conditions have suggested beneficial mood effects. For instance, LDN reduces FM symptoms, many of which overlap with core symptoms of major depressive disorder (MDD). In cancer patients, naltrexone reduced depression-like side effects of treatment with interferon-α. In addition, preliminary evidence suggests that ULDN may prevent desensitization to D2/3 agonists, which may explain how ULDN, when combined with dopamine agonists in the treatment of restless leg syndrome, allows equivalent control of limb movements at half the last dose of D2/3 agonists.

In 2016, a small proof-of-concept randomized trial showed that in 12 depressed subjects who relapsed on prodopaminergic antidepressants, LDN reduced depression severity.

In 2018, a randomized clinical trial was conducted comparing the efficacy of XR-NTX versus buprenorphine–naloxone in alleviating symptoms of anxiety, depression, and insomnia among patients diagnosed with OUD. The study results demonstrated that both drug treatments were equally effective in reducing symptoms of anxiety and depression, but symptoms of insomnia were significantly further reduced by the XR-NTX treatment. In 2019, a meta-analysis of 26 studies demonstrated contradictory results, and naltrexone was found to have a higher risk of insomnia and somnolence than placebo.

A double-blind, placebo-controlled trial combining sertraline and naltrexone for the treatment of co-occurring depression and alcohol use disorder found that more patients taking the sertraline and naltrexone combination achieved abstinence from alcohol, delayed relapse to heavy drinking reported fewer serious adverse events, and tended not to be depressed by the end of treatment compared to placebo or either medication alone.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Although this section is explicitly written for naltrexone, the effects of nalmefene are similar except where noted.

Respiratory System

In opioid-dependent individuals, the administration of naltrexone has been shown to produce various nasal symptoms, such as nasal congestion, sneezing, rhinorrhea, sore throat, cough, and shortness of breath. In addition, in clinical trials with the injectable formulation Vivitrol, one diagnosed case and one suspected case of eosinophilic pneumonia were observed.

Heart

Naltrexone has increased blood pressure in opioid-dependent individuals, and it can also produce palpitations and tachycardia. Naltrexone has also been shown to decrease fibrinogen levels in opioid users.

Smooth Muscles

Naltrexone has produced muscle tension, tremors, and twitching in opioid-dependent subjects. In addition, incidences of back pain/stiffness, muscle cramps, and joint stiffness have also been reported with Vivitrol.

Liver

While naltrexone is not hepatotoxic when administered in the recommended dose of 50 mg/day, hepatotoxic potential has been documented for a sixfold higher dose of 300 mg/day in obese individuals and patients with Alzheimer disease. In an open-label safety study in which the typical dose was 50 mg, naltrexone was not associated with increased risk of hepatotoxicity in more than 500 alcohol-dependent patients treated for up to 12 weeks. However, in another trial, a daily dose of 100 mg was associated with a small but statistically significant increase in the number of patients who experienced elevated LFTs during 16 weeks. Thus, naltrexone is contraindicated in patients with acute hepatitis, liver failure, or significant liver disease. However, mild to moderate transaminase elevations associated with alcohol or drug use is not a contraindication. Similarly, XR-NTX is reported to have a lower likelihood of hepatotoxicity than previously suspected, even in individuals with HIV and HCV co-infection. Moreover, naltrexone has also shown some utility in the treatment of pruritus observed in cholestatic patients.

Endocrine System

Changes in baseline levels of some hypothalamic, pituitary, and gonadal hormones have been reported with naltrexone. In addition, naltrexone has been shown to produce significant increases in plasma luteinizing hormone (LH) and testosterone in opioid-dependent individuals and reverse heroin-induced decreases in such individuals. In healthy adult men, naltrexone has been shown to increase plasma prolactin, LH, testosterone, adrenocorticotropin (ACTH), and cortisol levels.

Studies have demonstrated that naltrexone has efficacy in treating different reproductive pathologic conditions, such as hypothalamic amenorrhea and polycystic ovary syndrome (PCOS). The studies note that the administration of naltrexone leads to the resumption of pulsatile gonadotropin secretion followed by restoration of menstrual cycles and pregnancies in patients with hypothalamic amenorrhea. In addition, pilot studies administering 50-mg daily naltrexone to women with PCOS and insulin resistance showed a dramatic decline in insulin levels after 3 to 7 weeks of treatment.

Gastrointestinal Symptoms

Abdominal pain and cramps have been documented following naltrexone administration in both opioid-dependent and nondependent individuals. Some individuals also report feeling nauseated or vomiting and suffering a loss of appetite and diarrhea or constipation. Naloxegol, methylnaltrexone, and naldemedin are peripherally acting μ-opioid receptor antagonists—PAMORA approved for opioid-induced constipation (OIC) in adults with chronic noncancer pain.

Carcinogenesis and Mutagenesis

Carcinogenicity studies with high doses in rats have documented a small increase in mesotheliomas and vascular tumors. No evidence of mutagenicity has been documented with naltrexone or nalmefene.

Pregnancy

Naltrexone, when administered at doses 140 times the therapeutic dose in humans, has been shown to decrease pregnancy rates in rats and to have an embryocidal effect in rats and rabbits. Doses of 100 mg per kg in rats and 60 mg per kg in rabbits have been shown to have similar effects. In preclinical studies, naltrexone has been shown to pass through the placenta to the fetus. There are no adequate or controlled studies of naltrexone in oral or injectable form in pregnant or nursing women. A 2017 retrospective cohort study compared the obstetric outcomes in women with OUD who were treated with either implant naltrexone, oral methadone, or sublingual buprenorphine. The study highlighted that rates of ectopic pregnancies in naltrexone-treated women occurred at more than four times the rates of the three other treatment groups. Similarly, preclinical reproduction studies with nalmefene indicate no evidence of impaired fertility or fetal harm. However, there are no adequate trials in pregnant women, and this drug should be used during pregnancy or in nursing women only if critically needed.

Genitourinary

Less than 10% of individuals have reported increased frequency of, or discomfort during, urination, increased frequency of erections, or altered sexual interest.

Psychiatric/Neurologic

Non–opioid-dependent individuals have reported feeling fatigued, tired, and listless following the administration of naltrexone. Some individuals have reported increasing irritability, restlessness, mental confusion, and slowing following naltrexone administration. Depression and dysphoria have been cited as adverse effects of naltrexone; however, a review of clinical studies concluded that naltrexone treatment does not produce dysphoria as a serious side effect. A recent meta-analysis of medications for alcohol use disorders found that both naltrexone and nalmefene have a higher risk of withdrawal due to adverse events like dizziness, nausea, and vomiting. However, the latter has a higher risk difference for incidence of dizziness than placebo and has higher rates of incidence of headache. In addition, substance users have an increased risk of suicide, and this risk is not reduced by naltrexone.

Skin/Local Injection Site

Vivitrol (XR-NTX) injections can result in pain, tenderness, induration, swelling, erythema, bruising, and pruritus at the injection site. Proper injection technique is essential as inadvertent SC administration may increase the likelihood of severe injection site reactions.

General Problems

Some individuals have also reported blurred eyes and increased light sensitivity, as well as tinnitus.

THERAPEUTIC INDICATIONS

FDA-approved indications for naltrexone include maintenance treatment for detoxified patients with OUD and preventing relapse in patients with alcohol use disorder. Naltrexone has established utility when combined with clonidine as part of outpatient detoxification protocols. An XR injectable formulation of naltrexone has been approved for maintenance treatment of both opioid and alcohol use disorder. In addition, a fixed-dose combination of NB in an XR formulation has been approved to treat obesity. In addition, the FDA has approved naloxone hydrochloride injection for opioid overdose reversal or suspected acute opioid overdose.

PRECAUTIONS AND ADVERSE REACTIONS

One of the most critical risks with this class of agents is the risk of precipitated withdrawal in those who have received high doses of opioids for therapeutic purposes or those who use opioids recreationally. As a result, patients should abstain from short-acting narcotics (e.g., heroin) for at least 5 to 7 days and from longer-acting opioids (e.g., methadone) for 10 days or more based on self-report and urine toxicology. Suppose there is a question of occult OUD in an individual who has a negative urine toxicology screen for opioids and is not showing signs of withdrawal. In that case, a naloxone challenge test is recommended: the patient receives an SC injection of up to 0.8-mg naloxone and is observed for withdrawal symptoms for 20 minutes. Otherwise, naltrexone is generally well tolerated, and most patients who do not react to a test dose of naloxone report no side effects. The profile of reported side effects has considerable similarity to the symptoms of opioid withdrawal and includes gastrointestinal distress (nausea, vomiting, diarrhea, and abdominal pain), anxiety and restlessness, dysphoria, mild hypertension, headache, and insomnia. These side effects are most prominent in the first several days of use and improve rapidly for patients remaining in treatment. However, a substantial minority of those with a diagnosis of OUD (5% to 10%) may refuse a second dose, and an additional 10% to 20% may discontinue medication in the first month of treatment, often citing continued discomfort reminiscent of opioid withdrawal throughout the first 1 to 3 weeks of treatment. Side effects can be minimized by ensuring an adequate interval between cessation of opioid use and initiation of naltrexone and by a graduated initial dose schedule starting with 12.5 or 25 mg daily for the first several days. A prior study that used naltrexone depot treatment showed that 32% of opioid-dependent patients discontinued treatment after the first injection. Six-month retention rates of oral naltrexone in opioid-dependent patients are generally less than 20% following induction. However, injectable naltrexone has been shown to have higher 6-month retention rates than placebo in a recent sizeable Russian study. Studies examining the medication refill persistence showed that injectable naltrexone is consistently associated with better adherence rates in alcohol-dependent patients than the oral agents.

Patients should be advised that taking large quantities of exogenous opioids to overcome the blockade provided by opioid antagonists can be extremely dangerous. Life-threatening opioid intoxication may result, including respiratory arrest or circulatory collapse. In addition, patients should be advised that once these agents are discontinued, they may be more sensitive to the effects of opioids due to loss of tolerance.

Nausea is the most common adverse reaction to naltrexone among alcohol-dependent patients. It can be mild or severe and typically occurs within 90 minutes of administration. In a large multisite safety study, side effects that occurred in more than 2% of patients included nausea (10%), headache (7%), dizziness (4%), nervousness (4%), fatigue (4%), insomnia (3%), vomiting (3%), anxiety (2%), and somnolence (2%). A similar side-effect profile was observed in a large clinical trial for alcohol use disorder with a 100-mg daily dose of naltrexone. The effect of naltrexone on acute alcohol withdrawal has not been investigated. The rate of early termination due to adverse events has been somewhat higher in studies in which naltrexone was administered to individuals with active alcohol use disorder. Women and younger patients appear to be at somewhat increased risk of side effects.

The potential for hepatotoxicity has been identified when naltrexone is administered at high doses. However, convincing reports of elevated LFT results have been limited to patients treated for eating disorders with doses of 250 to 300 mg daily, five to six times higher than the recommended maintenance dose for OUD. In a clinical trial of alcohol use disorder with a 100-mg/day dose, a 2% rate of increases in liver enzymes (aspartate or alanine aminotransferases) to greater than five times the upper limit of normal was observed; this increase resolved on discontinuation of naltrexone. Therefore, as a precaution, patients should receive a full battery of LFTs prior to receiving naltrexone, and naltrexone is contraindicated in patients with liver failure or acute hepatitis. The moderate transaminitis often seen in alcoholism and narcotic addiction is not a contraindication; however, baseline bilirubin, serum transaminases, and periodic monitoring of transaminases are indicated. If persistent elevations in liver enzymes occur, naltrexone should be discontinued. The possibility of medication noncompliance and continued alcohol or drug use should be ruled out when mild transaminase elevations (less than threefold) occur.

Because this class of agents is excreted through the kidneys and metabolized by the liver, caution should be used in patients with renal impairment or severe liver disease. Among patients with compensated and decompensated liver cirrhosis, a 5- to 10-fold increase in naltrexone levels has been found relative to individuals with normal liver function. The increased systemic availability of naltrexone, which is believed to be due to lower conversion rates of naltrexone to 6β-naltrexol, appears to be related to the severity of the liver disease.

Naltrexone has not been studied in pregnant or nursing women, and an embryocidal effect has been demonstrated in rats and rabbits when given a dose approximately 140 times the human recommended dose. As a result, naltrexone should be used during pregnancy only when the potential benefit outweighs the potential risks to the fetus. Similarly, there are no studies of nalmefene in pregnant women, and hence this agent should be used in pregnant women only when needed.

DRUG INTERACTIONS

Because naltrexone and nalmefene are opioid antagonists, they block the pain-relieving effects of opioid analgesics. Therefore, in patients maintained on these agents, nonopioid pain medications or approaches should be tried first when possible. For elective surgery, cessation of oral naltrexone treatment 72 hours prior to surgery is recommended. Naltrexone does not reverse the effects of nerve root blocks or spinal anesthesia and does not antagonize agents used for general anesthesia. IV nalmefene can be used to reverse respiratory depression produced by these agents without blocking the analgesic effects of subsequently administered opioids. In patients treated with opioid antagonists, if opioid medications are necessary to treat severe pain, the blockade can be overridden with a rapidly acting analgesic. However, treatment requires close monitoring in a setting equipped and staffed for cardiopulmonary resuscitation because the resulting respiratory depression may be deep and prolonged. Patients should be given a card identifying that they are on naltrexone to alert medical personnel in an emergency.

Recent trials of alcohol use disorder have combined naltrexone and disulfiram without any significant increase in adverse effects or hepatic dysfunction, and some clinicians use the two drugs concomitantly for brief periods. However, given the potential for hepatotoxicity with both these agents, frequent monitoring of hepatic function is recommended. Although formal drug interaction studies have not been conducted with antidepressant medications, the incidence of adverse events reported by patients who started naltrexone while ongoing antidepressant therapy was similar to that of patients receiving naltrexone only in a large safety study. However, simultaneous initiation of antidepressant therapy and naltrexone has resulted in higher rates of nausea than using naltrexone alone. In addition, increased lethargy and somnolence have been reported with the combined use of naltrexone and thioridazine. Combined use of high-dose naltrexone (e.g., 250 mg daily) with nonsteroidal anti-inflammatory agents (ibuprofen, aspirin) may be associated with elevations in LFTs.

IV nalmefene has been administered after benzodiazepines, inhalational anesthetics, muscle relaxants, and muscle relaxant antagonists administered in conjunction with general anesthetics without any adverse reactions. However, care should be taken when using flumazenil and nalmefene together, as both agents have induced seizures in preclinical studies.

LABORATORY INTERFERENCES

Thin-layer, gas–liquid, and high-pressure liquid chromatographic methods used to detect opioids in the urine do not interfere with naltrexone. However, the potential for false-positive urine for opioids using less specific urine screens such as enzyme multiplied immunoassay technique (EMIT) may exist, given that naltrexone and nalmefene are derivatives of oxymorphone.

DOSAGE AND ADMINISTRATION

Oral naltrexone is prepared in scored, 50-mg, capsule-shaped tablets. The recommended dose for adults is 50 mg/day for alcohol use disorder and maintenance of abstinence from opioids. Lower initial doses can be considered over the first few days for patients at increased risk of adverse effects, with the dose increased as tolerated. In a large multisite study of alcoholism treatment, effect sizes for the 100-mg dose (compared to placebo) were not larger than those observed in earlier studies with a 50-mg daily dose. Morning administration is recommended for improved medication compliance.

Although daily oral administration is effective for treating alcohol use disorder, there is some evidence that targeting naltrexone use to the days when drinking is anticipated may also be helpful for some patients. The injectable formulation Vivitrol is available in single-use kits that contain 380 mg of Vivitrol microspheres and a diluent. The microspheres must be suspended in the provided diluent before administration and administered as an IM gluteal injection once every 4 weeks or once a month.

Naltrexone should not be administered until there is no reasonable possibility of opioid use within the past 7 to 10 days or unless the patient has completed detoxification and abstinence has been verified with urine testing for opioids. If there is a question about continued opioid use, opioid-free status should be confirmed by a negative response to an IM naloxone test dose of 0.8 mg. If substantial opioid withdrawal symptoms do emerge, initiation of naltrexone should be delayed at least 1 additional day.

Nalmefene is available as a sterile solution for IV, IM, and SC administration in two concentrations, containing 100 µg/mL or 1.0 mg/mL of nalmefene. The dose of this agent is titrated to achieve the desired reversal of the opioid effects. The oral form of nalmefene is not available in the United States. Naloxone hydrochloride injection is available to be administered intravenously, intramuscularly, or subcutaneously. An initial dose of 0.4 to 2 mg can be administered in adults and may be repeated at 2- to 3-minute intervals to reverse opioid effects. If a reversal is not achieved after administering 10 mg of naloxone, the diagnosis of opioid overdose should be questioned. Although it is not yet FDA approved, naloxone hydrochloride is also available in IN preparations for opioid overdose reversal.

INVESTIGATIONAL DRUGS

Naltrexone and nalmefene are nonspecific opioid antagonists that have been shown to dose-dependently bind to all three types of opioid receptors. The highest antagonistic potency is at the μ-receptor, which is 5 to 10 times greater than that at the δ-receptor and 13 to 33 times greater than that at the κ-receptor. In addition, opioid antagonists with a specific affinity of specific opioid receptor types have been developed, which have enormous use as selective tools to identify specific receptor types involved in specific pharmacologic endpoints. Most of the new specific opioid antagonists are nonpeptides that can easily penetrate the blood–brain barrier and, therefore, be administered via peripheral routes. All of these are investigational and remain in the different phases of drug testing.

FURTHER READINGS

Anton RF, Latham PK, Voronin KE, et al. Nicotine-use/smoking is associated with the efficacy of naltrexone in the treatment of alcohol dependence. *Alcohol Clin Exp Res*. 2018;42(4):751–760.

Anton RF, Myrick H, Wright TM, et al. Gabapentin combined with naltrexone for the treatment of alcohol dependence. *Am J Psychiatry*. 2011;168(7):709–717.

*Anton RF, O'Malley SS, Ciraulo DA, et al; COMBINE Study Research Group. Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study: a randomized controlled trial. *JAMA*. 2006;295(17):2003–2017.

Bartley CA, Bloch MH. Meta-analysis: pharmacological treatment of pathological gambling. *Expert Rev Neurother*. 2013;13(8):887–894.

Bisaga A, Mannelli P, Yu M, et al. Outpatient transition to extended-release injectable naltrexone for patients with opioid use disorder: a phase 3 randomized trial. *Drug Alcohol Depend*. 2018;187:171–178.

Coffin PO, Santos GM, Hern J, et al. Extended-release naltrexone for methamphetamine dependence among men who have sex with men: a randomized placebo-controlled trial. *Addiction*. 2018;113(2):268–278.

Comer SD, Mannelli P, Alam D, et al. Transition of patients with opioid use disorder from buprenorphine to extended-release naltrexone: a randomized clinical trial assessing two transition regimens. *Am J Addict*. 2020;29(4):313–322.

David SP, Chu IM, Lancaster T, Stead LF, Evins AE, Prochaska JJ. Systematic review and meta-analysis of opioid antagonists for smoking cessation. *BMJ Open*. 2014;4(3):e004393.

Garbutt JC, Kranzler HR, O'Malley SS, et al; Vivitrex Study Group. Efficacy and tolerability of long-acting injectable naltrexone for alcohol dependence: a randomized controlled trial. *JAMA*. 2005;293(13):1617–1625.

Gormez A, Rana F, Varghese S. Pharmacological interventions for self-injurious behaviour in adults with intellectual disabilities: abridged republication of a Cochrane systematic review. *J Psychopharmacol*. 2014;28(7):624–632.

Grant JE, Odlaug BL, Potenza MN, Hollander E, Kim SW. Nalmefene in the treatment of pathological gambling: multicentre, double-blind, placebo-controlled study. *Br J Psychiatry*. 2010;197(4):330–331.

*Greenway FL, Fujioka K, Plodkowski RA, et al; COR-I Study Group. Effect of naltrexone plus bupropion on weight loss in overweight and obese adults (COR-I): a multicentre, randomized, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2010;376(9741):595–605.

*Jonas DE, Amick HR, Feltner C, et al. Pharmacotherapy for adults with alcohol use disorders in outpatient settings: a systematic review and meta-analysis. *JAMA*. 2014;311(18):1889–1900.

Kelty E, Hulse G. A retrospective cohort study of obstetric outcomes in opioid-dependent women treated with implant naltrexone, oral methadone or sublingual buprenorphine, and non-dependent controls. *Drugs*. 2017;77(11):1199–1210.

Krieter P, Gyaw S, Crystal R, Skolnick P. Fighting fire with fire: development of intranasal nalmefene to treat synthetic opioid overdose. *J Pharmacol Exp Ther*. 2019;371(2):409–415.

*Krupitsky E, Nunes EV, Ling W, Illeperuma A, Gastfriend DR, Silverman BL. Injectable extended-release naltrexone for opioid dependence: a double-blind, placebo-controlled, multicentre randomised trial. *Lancet*. 2011;377(9776):1506–1513.

Krystal JH, Cramer JA, Krol WF, Kirk GF, Rosenheck RA; Veterans Affairs Naltrexone Cooperative Study 425 Group. Naltrexone in the treatment of alcohol dependence. *N Engl J Med*. 2001;345(24):1734–1739.

Latif Z, Šaltyte· BJ, Solli KK, et al. Anxiety, depression, and insomnia among adults with opioid dependence treated with extended-release naltrexone vs. buprenorphine-naloxone: a randomized clinical trial and follow-up study. *JAMA Psychiatry*. 2019;76(2):127–134.

Lee JD, Nunes EV Jr, Novo P, et al. Comparative effectiveness of extended-release naltrexone versus buprenorphine-naloxone for opioid relapse prevention (X: BOT): a multicenter, open-label, randomised controlled trial. *Lancet*. 2018;391(10118):309–318.

Mann K, Roos CR, Hoffmann S, et al. Precision medicine in alcohol dependence: a controlled trial testing pharmacotherapy response among reward and relief drinking phenotypes. *Neuropsychopharmacology*. 2018;43(4):891–899.

McLaughlin PJ, Zagon IS. Duration of opioid receptor blockade determines biotherapeutic response. *Biochem Pharmacol*. 2015;97:236–246.

Mitchell MC, Memisoglu A, Silverman BL. Hepatic safety of injectable extended-release naltrexone in patients with chronic hepatitis C and HIV infection. *J Stud Alcohol Drugs*. 2012;73(6):991–997.

O'Malley SS, Corbin WR, Leeman RF, et al. Reduction of alcohol drinking in young adults by naltrexone: a double-blind, placebo-controlled, randomized clinical trial of efficacy and safety. *J Clin Psychiatry*. 2015;76(2):e207–e213.

O'Malley SS, Garbutt JC, Gastfriend DR, Dong Q, Kranzler HR. Efficacy of extended-release naltrexone in alcohol-dependent patients who are abstinent before treatment. *J Clin Psychopharmacol*. 2007;27(5):507–512.

Schacht JP, Randall PK, Latham PK, et al. Predictors of naltrexone response in a randomized trial: reward-related brain activation, OPRM1 genotype, and smoking status. *Neuropsychopharmacology*. 2017;42(13):2640–2653.

Tanum L, Klemmetsby K, Latif ZEH. Effectiveness of injectable extended-release naltrexone vs daily buprenorphine-naloxone for opioid dependence: a randomized clinical noninferiority trial. *JAMA Psychiatry*. 2017;74(12):1197–1205

Toljan K, Vrooman B. Low-dose naltrexone (LDN)—review of therapeutic utilization. *Med Sci (Basel)*. 2018;6:82.

Trivedi MH, Walker R, Ling W, et al. Bupropion and naltrexone in methamphetamine use disorder. *N Engl J Med*. 2021;384(2):140–153.
Witkiewitz K, Roos CR, Mann K, Kranzler HR. Advancing precision medicine for alcohol use disorder: replication and extension of reward drinking as a predictor of naltrexone response. *Alcohol Clin Exp Res*. 2019;43(11):2395–2405.
Younger J, Parkitny L, McLain D. The use of low-dose naltrexone (LDN) as a novel anti-inflammatory treatment for chronic pain. *Clin Rheumatol*. 2014;33:451–459.

33.13b Disulfiram and Acamprosate

KATHERINE S. PIER, M.D.

DISULFIRAM

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Alcohol
Mechanism of Action	Enzyme inhibitor (mainly liver)

History

In 1937, Dr. E. E. Williams, a rubber company's family physician, serendipitously discovered disulfiram (DSF) for medical use. Dr. Williams observed that workers in the vulcanized rubber industry exposed to tetraethylthiuram disulfide became physically ill after alcohol consumption. The tetraethyl analog of the vulcanization chemical had also been used topically since 1942 to treat scabies. In 1945, while investigating this chemical to treat parasitic stomach infections, Erik Jacobsen and Jens Hald tested the compound on themselves, only to become ill after drinking alcohol. They observed and reported symptoms such as nausea, vomiting, facial flushing, headache, and dizziness. Two years later, they collaborated with Dr. Oluf Martensen-Larsen and designed a study for DSF as a treatment for alcohol dependence. In 1951, it became the first medication approved by the Food and Drug Administration (FDA) for that indication.

Over the years, it has been recognized that the very same properties that constitute DSF's main therapeutic effect (i.e., the ability to produce unpleasant symptoms after alcohol intake, also known as a DSF–alcohol reaction) may also have potentially dangerous consequences. These are frequently related but not limited to dehydration and electrolyte imbalance secondary to vomiting, a symptom of the DSF–alcohol reaction. In the most severe cases, serious clinical conditions such as respiratory depression, myocardial infarction, arrhythmias, heart failure, seizure, loss of consciousness, and death have been documented. These potential complications and the more recent development of alternative alcohol use disorder medications have limited some of DSF's initial popularity.

Chemistry

DSF, when it is in its pure state, is a white to off-white, odorless, almost tasteless powder soluble in water and alcohol. Its chemical name is bis(diethylthiocarbamoyl) disulfide; its structural formula is shown in Figure 33.13b–1.

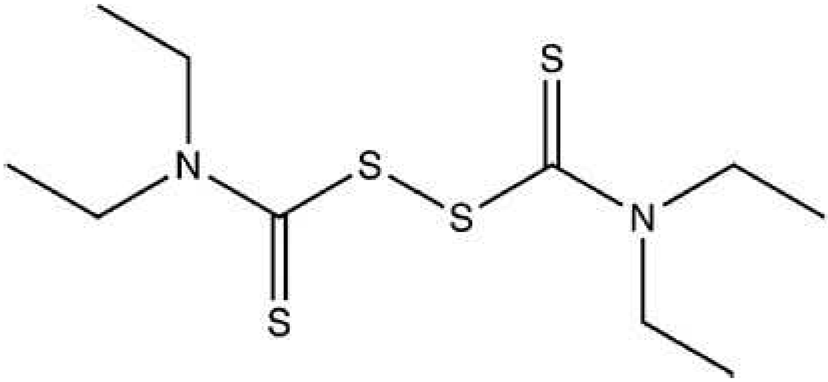

FIGURE 33.13b–1. Chemical structure of disulfiram.

Pharmacologic Actions

Pharmacokinetics

ABSORPTION. After ingestion, DSF rapidly converts to its diethyldithiocarbamate (DDC) copper complex. Subsequent absorption and distribution via the gastrointestinal (GI) mucosa into the bloodstream involve both the parent drug and the DDC copper complex. Absorption of DSF from the GI tract is rapid but incomplete (e.g., approximately 80% to 95% is absorbed through the GI tract). The concentration of DSF's active metabolite diethylthiocarbaminic acid methyl ester (Me-DTC) reaches its plasma peak about 4 hours after absorption. It is hypothesized that DSF tablets may be more bioavailable when taken with food. It can be crushed and mixed with water, coffee, tea, milk, soft drinks, or fruit juice.

DISTRIBUTION. Because of its high lipid solubility, DSF is widely distributed and accumulates in various fat deposits. Its safety during pregnancy has not been established, and it is labeled a Class C drug. In experimental animals, administration of DSF with ethanol potentiates the embryotoxic effects of ethanol. It is not clear whether the same potentiation occurs in humans. There are several case reports of normal pregnancy outcomes in humans exposed to DSF without exposure to ethanol. It is not known whether it is excreted in breast milk.

METABOLISM AND ELIMINATION. Due to partial absorption, approximately 5% to 20% of DSF is excreted in the feces. In the blood, DSF is rapidly metabolized to DDC, which is partly excreted as carbon disulfide in the expired air and is partly metabolized in the liver to diethyldithiomethylcarbamate (Me-DDC). Further metabolic changes result in the formation of the active metabolite Me-DTC. This metabolic cascade is not appreciably affected by a mild to moderate decrease in hepatic function. DSF's metabolites are chiefly excreted in the urine, and a part is expired as carbon disulfide. Overall, elimination of DSF is slow, and up to 20% of a single dose may remain in the body for 7 days or more.

Pharmacodynamics

MECHANISM OF ACTION. DSF is a prodrug. It has been shown to inhibit the hepatic enzyme aldehyde dehydrogenase (ALDH) in vitro.

However, in vivo, the drug is too short lived to inhibit the enzyme directly. The net effect of irreversible inhibition of ALDH is achieved by DSF's active metabolite Me-DTC.

The pharmacologic action of DSF and its metabolites is based on inhibiting enzymes involved in ethanol catabolism. DSF is a sulfhydryl (-SH, thiol) group reagent and inhibits enzymes concerned with oxidation of active (-SH group) sites on enzymes' protein molecules. Normally, ethanol is metabolized to carbon dioxide and water, but in the presence of DSF, the metabolic chain of reactions stops after the production of acetaldehyde, leading to the accumulation of this substance. The acetaldehyde concentration in the blood may increase 5 to 10 times more than without DSF. Acetaldehyde is an important factor for developing a DSF–alcohol reaction, which occurs within 10 to 30 minutes after the ingestion of alcohol. The clinical picture of the DSF–alcohol reaction includes several unpleasant symptoms: intense flushing of the face, a feeling of difficulty in breathing, palpitations, a throbbing headache, nausea, and vomiting. Ingestion of large amounts of alcohol may cause a significant fall in blood pressure, resulting in fainting and the risk of cardiovascular shock.

Re-establishment of the ALDH enzymatic activity depends on new synthesis, which occurs gradually for up to 2 weeks. Although the plasma half-life of DSF's active metabolite Me-DTC is about 10 hours, its inhibiting effect on ALDH may last considerably longer, persisting for 7 to 14 days after discontinuation. Prolonged use of DSF does not result in the development of tolerance; conversely, it has been shown to produce heightened sensitivity to alcohol.

In addition to its effect on ALDH, DSF and some of its metabolites (e.g., DDC) also inhibit the enzyme dopamine-beta-hydroxylase (DBH), thus reducing the synthesis of noradrenaline. This mechanism is suggested to play a role in the DSF–alcohol reaction. DSF also diffuses readily into cells and raises intracellular -SH levels; therefore, it can affect intracellular oxidation reactions. Notably, DSF has been shown to inhibit the enzymes xanthine oxidase and succinoxidase. DSF also possesses antithyroid properties as it reacts with free iodine to form a stable molecular complex.

BLOOD CONCENTRATIONS AND RELATION TO ACTION. Plasma concentrations of DSF may vary between individuals due to several factors, most notably age and hepatic function. In general, the severity of the DSF–alcohol reaction is proportional to the amount of the ingested DSF and alcohol. However, DSF plasma levels are rarely utilized in clinical practice. Mild effects may occur at blood alcohol concentrations of 5 to 10 mg per 100 mL. Symptoms will fully develop when alcohol levels reach 50 mg per 100 mL. Unconsciousness may occur when alcohol levels reach 125 to 150 mg per 100 mL.

Design and Interpretation of Clinical Drug Studies

To date, the Veterans Administration Cooperative Study remains the largest DSF treatment study, conducted in a cohort of 605 males diagnosed with alcoholism. Participants were randomly assigned to three treatment arms: 202 received DSF 250 mg daily, 204 were administered DSF 1 mg daily, and 199 were prescribed riboflavin. After 1 year of treatment, there was no difference in the percentage of patients from the three groups in terms of total abstinence, time to first drink, employment, or social stability. There was a significant relationship between adherence to drug regimens and complete abstinence. Among those who did drink, those in the DSF 250 mg group reported significantly fewer drinking days. Also, in the subpopulation of patients who attended all the scheduled follow-up research appointments, treatment with DSF 250 mg daily significantly decreased drinking days compared to low-dose DSF and riboflavin. In this large trial, adherence to the study medication was a notable barrier, with only 19% of patients taking DSF consistently. Ultimately, the study suggested that DSF was unlikely to be necessary for subjects who went on to achieve total abstinence but may be helpful in older males, who were socially stable, who relapse in terms of reducing the number of relapse days.

The first systematic meta-analysis of DSF's effectiveness was published in 2011. Eleven randomized controlled trials were included with a total of 1,527 patients. Studies were included from 1979 to 2010 in India, Denmark, the United States, Austria, Finland, and Italy. This review concluded that DSF had some short-term clinical benefits, including short-term abstinence, a fewer number of drinking days, and prolonged days until relapse. However, possible confounds were related to the fact that five of the six efficacious studies also had supervised DSF administration and that all studies were short-term treatment protocols. Others have pointed to possible selection bias since the authors included one nonefficacy study while five eligible trials were absent.

Moreover, the meta-analysis above utilized both open and blind studies. This issue was addressed by a second meta-analysis that compared blind versus open-label studies of DSF. The analysis, which included 2,414 participants from 22 studies, showed no difference between DSF and control groups when open studies were included. The authors explained this finding on the premise that the clinical effect of DSF is directly proportional to the expectation of a DSF–alcohol reaction, meaning a blinded placebo group has the same psychological threat as the treatment arm. Moreover, they point out that even in a blinded trial, the participants can unmask themselves easily with minimal alcohol. When analyzing data from only open-label studies, they found DSF had a significant effect size and, in fact, a greater one than naltrexone, acamprosate, and control group treatment conditions. Finally, the meta-analysis found that in randomized controlled trials where DSF administration was supervised, DSF performed significantly better than controls, while in unsupervised studies, DSF was not superior to controls on the main outcome.

There are no randomized controlled studies on long-term supervised DSF administration, and there have not been randomized controlled trials. One study, the Outpatient Long-Term Intensive Therapy for Alcoholics (OLITA), integrated supervised DSF administration into a 2-year comprehensive biopsychosocial approach to addiction treatment and achieved a 52% cumulative 9-

year abstinence rate. This does not isolate the effects of DSF from the effects of the psychosocial treatment but provides a model in which supervised DSF over 2 years may have contributed to long-term abstinence in some individuals.

By expert consensus, optimal DSF response is achieved when treatment is conducted in a specialized facility for substance use disorders (SUDs). Aspects of care that can increase DSF efficacy by adhering include using incentives, contracting with the patient and a significant other, providing regular reminders, patient behavioral training, and social support networks. The need for adherence and supervised administration of this medication provides some framework to understand that DSF has documented superior efficacy over placebo in treating alcohol use disorders in adolescents or in combination with spouse-observed administration. This also helps conceptualize why court-ordered DSF therapy promotes efficacy by increasing adherence to the regimen. Additional research indicates that, generally, older men involved in a structured social network like alcoholics anonymous (AA) are more likely to remain compliant to treatment regimens and thus, derive clinical benefit.

Spousal administration of DSF can be advantageous in increasing adherence to the treatment. Several studies have documented the benefit of behavioral marital therapy as an adjunct to DSF treatment. A pivotal component in this form of combined therapy is the DSF contract—a time-limited, renewable agreement by the couple to practice a behavioral interaction of observed DSF administration daily. It is believed that the contract aims to increase patient adherence to the medicine on an individual level. On the couple's level, the goal is to restructure the relationship dynamics of early sobriety behaviorally, often characterized by negative emotions like distrust, shame, and occasionally violence. Several trials have evaluated the efficacy of this approach as the couples received the combined treatment for periods of 10 to 24 weeks and were followed as long as 104 weeks after that. Overall, the combined treatment has produced greater DSF adherence, better drinking outcomes, higher marital satisfaction, and less domestic violence.

There are few comparisons between DSF and the other FDA-approved medications for the treatment of alcoholism, naltrexone, and acamprosate. Some of these trials are also confounded by the effects of frequent supervision for that administered DSF, which is not the case for those administered the other pharmacotherapies.

One randomized, open-label trial aimed to compare the effects in alcohol-dependent patients of three pharmacotherapies: DSF, naltrexone, and acamprosate. It included 243 voluntary treatment-seeking alcohol-dependent adult outpatients. It had two phases: a 12-week continuously supervised medication phase, followed by up to a year of unsupervised medication. Participants were followed for up to 2.5 years. All treatment arms included a brief manual-based cognitive behavioral therapy. The investigators found that all three medication interventions demonstrated a reduction in drinking from baseline to the end of the study. DSF was significantly superior to the other two agents (i.e., naltrexone and acamprosate) during the phase of supervised treatment; it demonstrated a greater reduction in heavy drinking days, reduction in weekly consumption, and increased abstinence days, and an increased time to the first drink. During the unsupervised period of the study, there were no significant differences between the groups in the heavy drinking days or time to first drink, although DSF displayed a significantly greater efficacy in the number of abstinence days.

Another study analyzed retrospectively collected data from 2002 to 2007 of 353 alcohol-dependent individuals in Germany and attempted to understand the relative efficacy of acamprosate and DSF. On the last week of 3-week inpatient detoxification, medical providers and patients mutually selected patients to be started on DSF versus acamprosate. Those receiving DSF were required to attend outpatient appointments three times weekly for supervised DSF, while those receiving acamprosate were required to attend a one-time, weekly appointment without the supervision of acamprosate administration. Abstinence was assessed using a breathalyzer, self-report, the reports of patients’ loved ones, urine and serum assays, and physician assessment.

Time before the first alcohol relapse, attendance in outpatient treatment, and cumulative alcohol abstinence achieved during a 12-month period were significantly higher in the DSF group. A longer duration of alcohol dependence predicted favorable treatment outcomes in the DSF group, while for the acamprosate group, the chances for a successful treatment increased with a shorter duration of alcohol dependence. The study has numerous limitations and confounding factors limiting the ability to draw true comparisons between acamprosate and DSF. That being said, the study provided a model wherein supervised DSF administration three times weekly contributed significantly to cumulative abstinence, adherence to outpatient treatment, and increased time to first relapse in patients with severe alcohol dependence compared to another treatment.

Efficacy of DSF has also been studied in individuals with addiction and co-occurring psychiatric disorders, including major depression, post-traumatic stress disorder (PTSD), bipolar disorder, schizophrenia, and schizoaffective disorder. One study enrolled 254 patients with alcoholism and a co-occurring Axis I disorder and randomized participants into four groups: (1) naltrexone alone, (2) placebo alone, (3) DSF and naltrexone, and (4) DSF and placebo. There was open randomization to DSF or no DSF and double-blind randomization to naltrexone or placebo. All patients were treated for 12 weeks in an outpatient setting. Participants treated with DSF or naltrexone as compared to placebo had significantly more consecutive weeks of abstinence and less craving than those treated with placebo, but there were no significant group differences in other measures of alcohol consumption (percent of days abstinent, percent of heavy drinking days, achieving total abstinence). There was no advantage to the combination of both medications. The study found that in subjects who chose DSF, medication adherence rates were as high as with naltrexone. Surprisingly, DSF-treated subjects reported lower levels of a craving than the naltrexone-treated subjects. Of interest, subjects taking any medication reported lower rates of paranoid ideation compared to placebo, contradicting other reports that DSF can actually precipitate psychosis.

DSF may be more effective in promoting short-term abstinence and treatment retention after detoxification than long-term relapse prevention. As such, some have suggested that this medication be solely used to assist in the initial period of abstinence to facilitate the client's involvement in psychosocial interventions. Because of the severe adverse effects of DSF in the event of relapse and some of the side effects of DSF itself, it has largely fallen out of favor in the United States. In general, DSF is considered an inappropriate treatment option for patients with severely impaired judgment, significant cognitive impairment, or those who are highly impulsive.

Effects on Specific Organs and Systems

Deaths from DSF treatment have become rare due to treatment with lower dosages and excluding patients with significant cardiac disease. DSF has a black box warning to never administer to a patient in a state of alcohol intoxication or without the patient's full knowledge.

GI and nervous system adverse effects can be prominent. DSF has been associated with hepatotoxicity. The symptoms of hepatotoxicity may be mild and reversible (e.g., jaundice, cholestasis, and altered liver function tests) or more serious and include acute hepatitis, hepatic necrosis, and liver failure, requiring transplantation. These effects may occur after many months of DSF treatment and have been documented in both individuals with and without prior history of abnormal liver function. Estimates of DSF-induced hepatitis are between 1 in 25,000 and 1 in 30,000 patients treated yearly.

DSF's effects on the nervous system may involve optic neuritis, resulting in eye pain or tenderness and vision changes, as well as peripheral neuropathy (i.e., paresthesias—numbness, tingling, pain, or weakness in hands or feet). The latter is mainly reported in patients on a relatively high dose of DSF (e.g., 500 mg daily). Peripheral neuropathy tends to improve with the discontinuation of DSF treatment. More rare side effects include the antithyroid effect of

DSF and skin eruptions that are usually controlled with antihistamine treatment.

Severe reactions (when DSF is used at very high doses or in individuals with cardiovascular disease) can include **myocardial infarction, cardiac arrhythmias, and congestive heart failure (CHF)**.

Therapeutic Indications

Alcohol Use Disorder. In clinical practice, DSF is recommended to aid in managing selected patients with alcohol use disorder. Selected patients should be highly motivated to maintain a state of sobriety. The agent has little effect on long-term drinking patterns once it is discontinued. It is most appropriate and efficacious in the short term after detoxification to assist in the initial period of abstinence to facilitate the client’s involvement in psychosocial interventions, including psychotherapy. Optimum response to DSF seems to be achieved in facilities specializing in treating SUDs.

The DSF–ethanol reaction is an aversive physiologic state that functions through negative reinforcement and behavioral counterconditioning to extinguish an unwanted behavior (alcohol consumption). When an individual takes this medication, they acknowledge that possible alcohol consumption will entail a severe adverse reaction, and this knowledge is the motivation for abstinence. Baseline blood tests include blood alcohol or breathalyzer (to confirm abstinence), liver function tests, a complete blood count, serum chemistries including blood urea nitrogen (BUN) and creatinine, and a pregnancy test. Follow-up liver function tests are recommended 10 to 14 days after therapy initiation, then monthly for 6 months, to monitor for possible hepatic dysfunction. After the first 6 months, liver function testing should be obtained every 3 months. Monthly pregnancy tests are indicated for women of childbearing age. An ECG should be obtained at baseline if clinically indicated in people with heart disease.

DSF should be prescribed for individuals abstaining from alcohol for at least 12 hours. On occasions, patients who appear to be on an adequate maintenance dose of DSF may report that they can consume alcohol without any negative reactions. In such cases, the physician should presume that patient’s compliance with the treatment has been compromised. Before the patient is observed to reliably administer the medicine (preferably crushed and mixed with water), it may not be concluded that DSF is ineffective.

Disulfiram Test Reaction. During the earlier use of DSF, it was thought advisable to have patients experience at least one supervised DSF–alcohol reaction, a so-called “alcohol challenge.” This practice is currently not recommended; moreover, it should never be considered for patients over the age of 50 years. It is believed that a clear, convincing, and detailed description of the DSF–alcohol reaction is considered sufficient.

Management of Disulfiram–Alcohol Reaction. The DSF–alcohol reaction is diagnosed when a patient with a history of DSF use and possible exposure to ethanol presents with symptoms of the characteristic hypotensive flushing reaction. Blood DSF levels are not of value in diagnosis or treatment. Blood acetaldehyde levels may be elevated during the DSF–ethanol reaction, but this information is of little value in acute management. In a severe reaction, all measures should be taken to stabilize the patient’s respiratory and cardiovascular systems. This may include supplemental oxygen, intubation, intravenous (IV) fluids, and/or pressor agents. Other interventions include IV use of vitamin C, ephedrine sulfate, and antihistamines. Since hypokalemia has been reported, potassium levels should be monitored, particularly in patients treated with digitalis. Other useful laboratory studies include electrolytes, glucose, BUN, creatinine, liver aminotransferases, and ethanol level. Some protocols include treating vomiting with metoclopramide. There is no specific antidote. A case series found that IV fomepizole safely and rapidly reversed vasodilation and toxicity related to alcohol ingestion with DSF.

Management of Disulfiram Overdose (Without Alcohol). DSF overdose (without alcohol ingestion) is rare, and its clinical presentation is one of the neurologic adverse effects of the medication: confusion, lethargy, headache, ataxia, seizures, and coma. Brain lesions in the basal ganglia have also been reported. Patients may also have a garlic-like breath odor, vomiting, and hypotension. The diagnosis of overdose is made based on history and presenting symptoms. In some circumstances, activated charcoal may be clinically indicated. The treatment is focused on maintaining a patent airway, stabilizing the cardiovascular system, and treating seizures.

Other Disorders. There is evidence that DSF could be efficacious for cocaine use disorder alone, co-occurring cocaine and alcohol use disorders, and cocaine and opioid use disorders. Initial research into DSF for cocaine addiction emerged based on evidence that alcohol and cocaine are often co-ingested. Researchers hypothesized that withholding alcohol from cocaine users through DSF administration would reduce cocaine use, which was found to be true in one initial study from 1993. A subsequent randomized 12-week pilot trial compared the effects of DSF administration to the effects of naltrexone. Both treatments were accompanied by cognitive behavioral therapy (CBT). DSF proved more effective than naltrexone at lowering the frequency of cocaine use. In a larger clinical trial, DSF treatment improved abstinence from cocaine compared to no treatment; DSF treatment in this larger trial appeared more effective for those outpatients who also received CBT. The beneficial effect of DSF on cocaine use was still evident a year later in a follow-up study.

Subsequent trials assessed the effects of DSF on cocaine in patients who were addicted to both cocaine and opioids and were maintained on methadone or buprenorphine. Those treated with DSF were better able to reduce their alcohol, cocaine, and opioid intake. DSF shortened the time necessary for patients to reach continuous cocaine, but not heroin abstinence. In these trials, alcohol consumption was minimal for all subjects, and baseline alcohol use did not predict responses to DSF. A 2004 randomized, placebo-controlled trial investigated cocaine use with and without concurrent alcohol use disorders, showing that the benefits of DSF therapy were most pronounced in patients who either were not alcohol dependent at baseline or who fully abstained from alcohol during treatment.

Whereas ALDH is the primary target in treating alcoholism, human laboratory, genetic, and preclinical animal studies indicate that its beneficial effects on cocaine use result from the inhibition of DBH, resulting in increased dopamine activity in the synaptic cleft. Given that cocaine also increases dopamine levels by blocking the reuptake of the dopamine transporter, the combination may produce a state of hyperstimulation, causing discomfort and aversion. In laboratory studies, people experiencing a DSF–cocaine interaction demonstrated increased heart rate and blood pressure and reported anxiety, paranoia, and restlessness.

People with low levels of the enzyme DBH have increased dopamine activity and tend to be more responsive to DSF. Hormones, as well as genes, may influence DBH levels. Researchers suspect that estrogen hormones increase DBH, attenuating the effect of DSF, which could explain why women seem to benefit less than men. A recent study showed that patients with genetically higher dopamine transporter levels had better treatment outcomes with DSF pharmacotherapy of cocaine dependence than those with lower DAT levels.

A 2010 Cochrane review included seven controlled studies with 492 participants receiving DSF, a placebo, no pharmacologic treatment, or naltrexone in addition to psychosocial treatment. One out of four studies comparing DSF to placebo found that DSF significantly affected the number of weeks of cocaine abstinence compared to placebo. Another study found DSF significantly affected maximum weeks of consecutive abstinence and the number of people achieving 3 or more weeks of consecutive abstinence compared to no treatment. Ultimately, the authors concluded that the evidence for DSF for cocaine dependence was low and warranted further randomized controlled trials.

The review showed “low evidence” in support of DSF for the treatment of cocaine dependence. Current investigations are examining safety profiles for the combination of DSF with alcohol and cocaine.

Starting in the 1970s, evidence accumulated that DSF had anticancer properties both through its direct actions and by reversing resistance to chemotherapeutic drugs. Its antitumor effect has been reported in many preclinical studies and on seven types of cancer in humans: non–small-cell lung cancer (NSCLC), liver cancer, breast cancer, prostate cancer, pancreatic cancer, glioblastoma multiforme (GBM), and melanoma. The mechanisms, particularly the intracellular signaling pathways, have yet to be fully elucidated.

Studies have found that DSF inhibits NF-kB signaling, proteasome activity, and ALDH activity. It induces endoplasmic reticulum (ER) stress and autophagy and has been used as an adjuvant therapy with irradiation or chemotherapy drugs. DSF kills the normal cancer cells and can target cancer stem cells, providing a new approach to preventing tumor recurrence and metastasis. Furthermore, other researchers have reported the ability of DSF to bind to nuclear protein localization protein 4 (NPL4), inducing its immobilization and dysfunction, ultimately leading to cell death.

Using Denmark’s cancer registry, examining more than 240,000 cases diagnosed between 2000 and 2013, along with data on the medications each patient took, a population-based cohort study found that of more than 3,000 patients taking DSF, the cancer death rate was 34% lower for the 1,177 patients who stayed on the drug compared to those who stopped taking it. Its benefits are held for prostate, breast, and colon cancer, as well as cancer overall. Studies have revealed that treatment with DSF and chemotherapy extends the lives of patients with metastatic breast and colon cancers. It has also been shown to slow the progression of small-cell lung cancer and GBM in small trials.

Cancer research into DSF’s mechanism has revealed that it may have anti-inflammatory properties more broadly by inhibiting nuclear factor-κB (NF-κB). Other research has also demonstrated that DSF can regulate FOXO1, leading to impaired hepatic gluconeogenesis, suggesting it may be involved in the hepatic regulation of metabolic function. Recent research has sought to investigate whether these properties of DSF might hold promise in treating obesity and metabolic syndromes. One study found that in mice fed high-fat diets, those treated with DSF had a 40% decrease in body weight and an increase in energy expenditure without a significant change in diet. DSF improved glucose tolerance and insulin sensitivity while preventing fibrosis and steatosis of the liver and hypertrophy of pancreatic islet cells. A follow-up study found that in cell culture, DSF was a potent activator of autophagy as well. While there have not been studies on the metabolic effects of DSF in humans yet, DSF may hold promise for treating human obesity and its metabolic sequelae.

Precautions and Adverse Reactions

Precautions. All patients prescribed DSF should be fully informed of the DSF–ethanol reaction and cautioned against the possible consequences of taking alcohol either surreptitiously or unwittingly. Patients receiving DSF should be strongly advised not to use any alcohol-containing preparations, for example, certain cough syrups, sauces, vinegar, tonics, foods prepared with wine, and should avoid the use of aftershave lotions and moisturizers containing alcohol. In addition, patients should be advised to obtain and carry an identification card describing the symptoms of DSF–alcohol reaction and containing contact information in emergency cases. DSF should not be coadministered with paraldehyde or metronidazole.

Candidates for DSF should be evaluated for hypersensitivity to other thiuram derivatives used in pesticides and rubber vulcanization. Concurrent conditions such as severe myocardial disease or ischemic heart disease, pregnancy, advanced hepatic and renal disease, psychotic states, and serious organic brain damage are contraindications of the use of DSF.

Adverse Reactions. The most common side effects associated with DSF include drowsiness, headache, fatigue, allergic dermatitis, stomach upset, dizziness, acneiform eruptions, halitosis, and impotence. A metallic or garlic-like aftertaste has been reported during the first 2 weeks of therapy. These complaints disappear spontaneously or by reducing the dosage.

Sensorimotor peripheral neuropathy, frequently associated with combined treatment with metronidazole or isoniazid, may present with numbness, tingling, pain, or weakness in hands or feet. In addition, symptoms of optic neuritis may occur, indicated by eye pain and vision changes. Most of these symptoms improve when DSF is discontinued.

More serious but rare side effects include DSF-induced hepatitis, with an incidence between 1 in 25,000 and 1 in 30,000 patient-years of treatment. Varying degrees of hepatotoxicity have been reported, including altered liver function tests, jaundice, and hepatic necrosis. Psychiatric side effects like mood liability and mental status changes have also been reported with the administration of DSF. DSF may also block DBH and increase dopamine concentrations, contributing to psychosis, particularly in patients with pre-existing psychotic disorders.

Drug Interactions

Concomitant ingestion of antacids containing divalent cations and large doses of ferrous salts may reduce the absorption of DSF. Other clinically salient drug–drug interactions of DSF may occur through several separate physiologic mechanisms (i.e., inhibition of hepatic enzymes). For instance, DSF may affect the metabolic rate of other drugs, in particular phenytoin and theophylline. The concomitant use of DSF with either of these two agents should be used with caution since it may result in elevated phenytoin or theophylline plasma level and possibly lead to phenytoin or theophylline toxicity. It is recommended that baseline phenytoin and theophylline levels be obtained before the initiation of DSF with subsequent regular monitoring of phenytoin and theophylline levels.

DSF may prolong prothrombin time when coadministered with oral anticoagulants (e.g., warfarin), as DSF inhibits warfarin’s metabolism, and regular assessments of anticoagulation profiles are recommended.

Another clinical interaction with a less understood mechanism occurs between isoniazid and DSF. Their coadministration may cause ataxia and even mental status changes, which usually resolve after discontinuation of DSF. Another antituberculosis agent, rifampin, also interacts with DSF, resulting in inhibition of rifampin's oxidation and renal excretion. Mental status changes (e.g., psychosis and confusion) have also been demonstrated with concurrent use of metronidazole.

DSF can differentially potentiate and prolong benzodiazepines' suppressing effect on the CNS—for instance, this effect is pronounced with coadministration of chlordiazepoxide and diazepam but not seen with oxazepam or lorazepam.

Animal research has documented that in the GI tract, DSF may interact with nitrates and form nitrosamine in the GI tract, which has been shown to cause stomach tumors. The toxicity of certain centrally acting drugs (e.g., morphine, pethidine, amphetamine, and barbiturates) has been potentiated by DSF in rats; however, the importance of this observation for humans is unclear.

Laboratory Interference

DSF is not known to interfere with any laboratory tests.

Dosage and Administration

Dosing for the first 1 to 2 weeks is 250 mg in the morning or the evening (based on whether the patient has sedation). The average maintenance dose is 250 mg daily, ranging from 125 to 500 mg. The dose of DSF should never exceed 500 mg per day.

The duration of DSF therapy should be determined on an individual basis.

Administration. DSF should be taken preferably upon waking in the morning, although in patients who experience a sedating effect, it may be taken at bedtime.

ACAMPROSATE

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Glutamate
Mechanism of Action	Yet to be determined

History

Acamprosate was developed in response to a growing literature demonstrating that chronic alcohol consumption and withdrawal fundamentally change neurochemistry and electrophysiology, thought to underlie some of the difficulty achieving initial abstinence from alcohol and regaining abstinence after relapse. Over 30 years ago in Europe, various pharmaceutical companies started synthesizing amino acid neuromediators, hoping to ameliorate the hyperglutamatergic state resulting from protracted alcohol use. In collaboration with several universities in France, pharmaceutical companies investigated several taurine and homotaurine compounds in the 1980s. An animal model of alcohol relapse conducted at the University of Rouen in France supported the utility of calcium acetyl-homotaurinate in reducing alcohol consumption in dependent rats. This initial preclinical trial was again validated in another animal model, which revealed a significant reduction in voluntary alcohol consumption.

These preclinical studies catalyzed the first multicenter, double-blind placebo-controlled study in 569 detoxified alcohol-dependent male patients in Europe, again showing decreased alcohol consumption in patients receiving calcium acetyl-homotaurinate compared to placebo. This compound, now known as acamprosate, was approved for the treatment of alcohol dependence in 1989 in France. Extensive clinical studies involving over 4,000 subjects throughout Europe demonstrated the safety and efficacy of acamprosate across clinical settings. By 1996, the medication became available in all European Union countries to treat alcohol use disorder, promulgating additional investigations into its utility in the pharmacologic treatment of alcohol use disorder.

In July 2004, the FDA approved acamprosate to maintain abstinence from alcohol in detoxified patients with alcohol use disorder. It became available for use in January 2005 under the trade name Campral. Since its introduction in 1989, acamprosate has been used to treat over 1.5 million patients worldwide.

Chemistry

Acamprosate is the calcium salt of *N-acetyl-homotaurine* (Ca-AOTA), a synthetic analog of the endogenous amino acid taurine with additional carboxyl and methyl groups in its chain. Acamprosate calcium is a white, nearly odorless powder readily soluble in water and practically insoluble in absolute ethanol. Its chemical name is calcium 3-acetylaminopropane-1-sulfonate, and its structure is shown in Figure 33.13b–2.

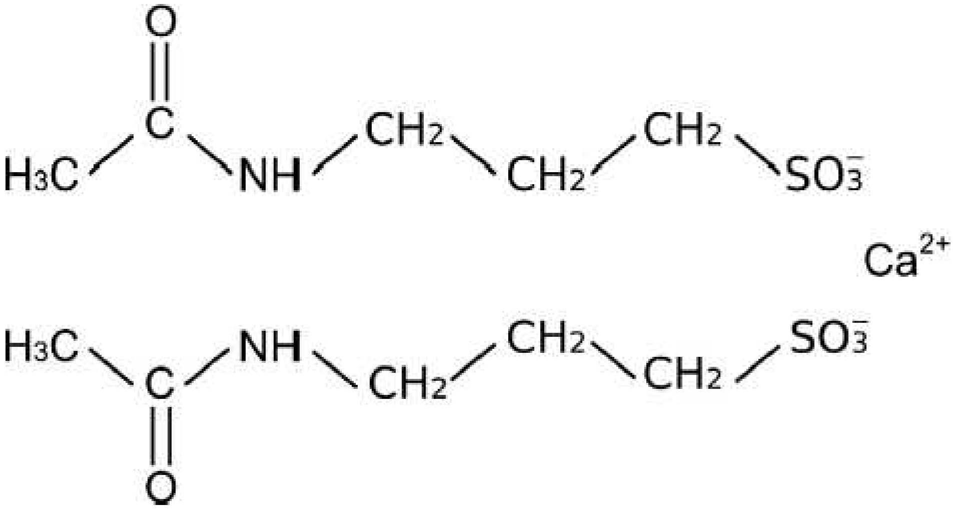

FIGURE 33.13b–2. Chemical structure of acamprosate.

Pharmacologic Actions

Pharmacokinetics

ABSORPTION. Acamprosate is a dimer of acetyl-homotaurine linked by a calcium salt formulated to assist in its absorption across the GI tract. However, the amount of acamprosate absorbed is quite low, with a bioavailability of ~11% in the fasting state. Food consumption reduces acamprosate absorption but not sufficiently to justify dose adjustments. Acamprosate is so slowly absorbed that it acts as if it were a modified-release medication. Peak plasma concentrations are usually reached within 3 to 8 hours after dosing, and steady-state levels are achieved by the fifth to the seventh day of dosing. Its terminal half-life ranges from 20 to 33 hours after administering a standard 666-mg enteric-coated formulation.

DISTRIBUTION. The amount of acamprosate that is protein bound is negligible, such that the active drug that permeates the intestinal mucosa is entirely dissociated in plasma. While its volume of distribution following IV acamprosate administration is high, estimated to be 72 to 109 L (approximately 1 L/kg), its oral bioavailability is relatively low. Only 6.5% to 11% of what is ingested crosses the GI tract, which explains why it must be given in high doses. In contrast to orally administered acamprosate, its concentrations are quite high when administered intravenously, but its half-life is approximately 3 hours compared to 20 to 33 hours when taken orally. This is an example of "flip-flop" pharmacokinetics in which the concentration and half-life are limited by the rate and extent of absorption rather than the volume of distribution or excretion.

Acamprosate has been assigned a Category C labeling in pregnancy. Teratogenicity has been demonstrated in rat and rabbit fetuses though there have been no large-scale studies in humans to date. It is also unknown whether the medication is excreted through breast milk, but in lactating rats given oral doses of acamprosate, the concentration in milk compared to the drug in blood was 1.3.

METABOLISM AND ELIMINATION. Importantly, acamprosate bypasses the liver and is eliminated unchanged in the urine and feces. Most of the acamprosate does not permeate the intestinal mucosa and is therefore removed through the feces. Once the medication is available via paracellular transport, renal clearance is expedient, occurring at approximately 6 mL/minute. The portion of the absorbed drug (ranging from 6.5% to 11%) does not undergo metabolism and is excreted in the urine as acamprosate. There is a linear relationship between renal clearance and plasma half-life of acamprosate and creatinine clearance values when taken orally. No significant pharmacokinetic differences between male and female subjects have been found for acamprosate.

Pharmacodynamics

MECHANISM OF ACTION. Acamprosate bears structural similarities to several amino acids, including gamma-aminobutyric acid (GABA), glycine, aspartate, and glutamate; its mechanism of action is therefore thought to be mediated, at least in part, through numerous neurotransmitter systems.

Initial studies focused on acamprosate's in vitro affinity for $GABA_A$ and $GABA_B$ receptors, which appears to decrease electrical excitability without changing the membrane potential of the neuron. It is postulated that the therapeutic effect of acamprosate is partially due to its potentiation of the GABA network, the main neuronal inhibitory system. Unlike other GABA agonists, however, acamprosate doesn't exert anxiolytic, hypnotic, or muscle relaxant properties.

Other research has shifted focus to the structural similarities between acamprosate and the excitatory neurotransmitter glutamate. It has been proposed that acamprosate may decrease the excitatory effects of glutamate, especially during the period after abrupt discontinuation of alcohol. Since alcohol inhibits the activity of the *N*-methyl-D-aspartate (NMDA) receptors, chronic alcohol consumption leads to the overexpression of these receptors. Chronic alcohol intake also affects the release of glutamate—in order to compensate for the alcohol-induced inhibition at the NMDA receptors, the amount of glutamate released into the synaptic cleft increases over time. Thus, abrupt discontinuation of alcohol results in unopposed neuronal excitation due to supranormal levels of glutamate acting on upregulated NMDA receptors, which may contribute to the pathophysiology of the alcohol withdrawal syndrome, including delirium tremens and seizures.

Other research supports the hypothesis that acamprosate interacts with NMDA receptors in a more complex way; for example, studies show that the drug is an NMDA partial co-agonist. This means that at lower concentrations, acamprosate activates NMDA receptors when receptor activity is low. Conversely, at higher concentrations and when NMDA receptor activity is high, acamprosate is inhibitory, thereby helping to restore equilibrium in this disrupted neuronal system.

Furthermore, it has been shown that acamprosate induces the release of taurine, an inhibitory neuromodulator, which also decreases neuroexcitability. In early alcohol withdrawal, the drug also decreases dopamine hyperexcitability in the nucleus accumbens, as well as the expression of brain *c-fos*, a gene expressed during early alcohol withdrawal.

Other research supports that by activating glycine receptors in the nucleus accumbens and nicotinic acetylcholine receptors in the ventral tegmental area, acamprosate elevates extracellular dopamine levels in the nAc. In rat models, acamprosate appears to elevate dopamine and simultaneously prevent ethanol from further increasing dopamine output. In these animal models, 10 days of treatment with acamprosate produced a loss of the ethanol intake–reducing as well as the dopamine-elevating effects of acamprosate, and the dopamine-elevating property of ethanol was restored, suggesting possible tolerance to the dopaminergic properties of acamprosate. In multiple studies, a significant shortcoming with acamprosate is about 50% of patients discontinue the medication before 6 months. It is possible that this is not only due to the three times daily dosing schedule but also due to the diminishing effects of acamprosate over time.

While it was previously thought that *N-acetyl-homotaurine* was the biologically active compound mediating acamprosate’s effects, recent literature has demonstrated that calcium may be the moiety responsible for acamprosate’s role in the treatment of alcohol use disorder. In fact, there is compelling evidence that acamprosate's efficacy does not stem from its direct interactions with glutamate receptors but rather from increasing calcium concentrations. Studies have demonstrated that treatment with calcium-acamprosate decreases voluntary drinking, while sodium-acamprosate is not superior to placebo in attenuating alcohol consumption. Sodium- and calcium-acamprosate have the same bioavailability, such that differences in pharmacokinetics are unlikely to render this finding. Human studies have shown that higher brain and plasma calcium levels correlate with acamprosate's efficacy parameters, such as time to first drink, time to severe relapse, and cumulative abstinence. This has led researchers to question whether *N-acetyl-homotaurine* is biologically active or whether the efficacy of naltrexone could be realized with calcium salts alone.

It is clinically important to emphasize that despite acamprosate's modulation of glutamate dysregulation in rodents, there is no clinical improvement in patients who begin acamprosate during acute detoxification. On the contrary, available research suggests that patients who receive acamprosate during the detoxification period drink more alcohol statistically in 10-week follow-up than those administered a placebo**.** Acamprosate does not eliminate or diminish withdrawal symptoms and is not indicated for the treatment of delirium tremens.

Acamprosate is most beneficial in maintaining abstinence and helping patients regain abstinence after relapse. The medication is thought to alleviate the negative affect associated with early stages of alcohol withdrawal, reverse alcohol-related changes in sleep architecture, and also decrease alcohol cravings once abstinence is achieved. It additionally decreases cravings in response to conditioned environmental cues. Unlike DSF, which relies on decreasing alcohol use by inducing distinctly unpleasant effects or naltrexone, which decreases the pleasurable effects of alcohol, acamprosate's primary mechanism appears to lie in its neuromodulatory effect—its ability to normalize disturbed brain chemistry and electrophysiology during acute and protracted withdrawal. Some studies suggest that it may reduce the rewarding effects of alcohol.

The medication has not traditionally been associated with the development of tolerance or dependence; however, in rodents, there does appear to be some tolerance toward the ethanol intake–reducing effect of the compound following prolonged exposure. The mechanism for this is discussed above. The clinical significance of this in humans awaits further research. Acamprosate does not cause alcohol aversion or produce a DSF-type reaction in the event of relapse.

Preclinical research provides evidence that acamprosate has neuroprotective properties in neurons exposed to alcohol. The hyperglutamatergic imbalance that takes place during alcohol withdrawal is associated with neuronal excitotoxicity, which acamprosate has been shown to mitigate. Results from cell culture and animal studies have demonstrated decreased cell death in alcohol-exposed neurons in the cerebral cortex and hippocampus. Research supports that excessive glutamate activity continues after acute withdrawal, inducing ischemia, as measured by lactate dehydrogenase levels. Furthermore, acamprosate has been shown to inhibit trans-ACPD, a metabotropic glutamate-type receptor (mGluR) agonist responsible for some of the neurotoxicity that glutamate exerts. The hyperglutamatergic syndrome that occurs during alcohol withdrawal is also thought to be mediated by calcium entry into neurons, which acamprosate alleviates. The relevance of these findings for treatment in humans awaits further research.

BLOOD CONCENTRATIONS AND RELATION TO ACTION. Plasma concentrations of acamprosate may significantly vary between individuals due to multiple factors, most notably individual rates of absorption. Pharmacokinetic studies indicate that the concomitant use of acamprosate and alcohol doesn’t result in any noticeable changes in the pharmacokinetics of either substance. At least one double-blind placebo-controlled randomized controlled trial in a sample of 56 patients seeking treatment for alcohol use disorder found a negative correlation between alcohol craving following a priming drink (i.e., first drink after 21 days of continuous acamprosate treatment) and acamprosate plasma levels. This suggests that higher acamprosate blood concentrations may increase the chances of regaining abstinence after mild relapse.

Design and Interpretation of Clinical Drug Studies

Since its inception, there have been more than 20 double-blind, placebo-controlled protocols investigating the utility and safety of acamprosate in the pharmacologic treatment of alcohol use disorders. The length of any individual trial ranged from 12 to 52 weeks, with some protocols offering additional follow-ups. With few exceptions, patients were detoxified prior to treatment; dosing was based on weight and ranged between four and six 333-mg tablets. The treatment took place in outpatient programs with concurrent psychosocial interventions. The subjects across trials were predominantly male, and the mean age fell in the fourth decade of life.

The cumulative sample size from a meta-analysis of eight positive randomized-controlled trials was 2,430, sufficiently powered to support some clinically relevant conclusions. Six of these protocols followed patients for over 52 weeks, providing a relatively long duration to study adverse effects over time. Acamprosate treatment separated from placebo most consistently when acamprosate was initiated after the acute withdrawal phase and in patients receiving concurrent psychosocial interventions. Acamprosate outperformed placebo in terms of continuous abstinence duration (time to first drink) as well as cumulative abstinence proportion (number of days abstinent out of the total days patients were in treatment) measured at 6 months. These results support the notion that acamprosate may help patients achieve complete abstinence and regain abstinence if relapse occurs. Another meta-analysis assessed acamprosate in 27 studies across 7,519 subjects. Sample sizes ranged from 21 to 1,383, treatment durations were 12 to 52 weeks, and the mean age was in the 40s. The participants met the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for alcohol dependence, and the vast majority of studies included concurrent psychosocial interventions. This meta-analysis found that the number needed to treat to prevent a return to any drinking for acamprosate was 12, suggesting modest efficacy of acamprosate treatment. In this meta-analysis, the number needed to harm for withdrawal from trials due to adverse effects was 48 for naltrexone and not significantly increased for acamprosate.

In protocols that included patients who had not undergone detoxification or were acutely withdrawing from alcohol, acamprosate failed to demonstrate superiority over placebo. Further, in another study where patients initiated treatment 25 days after recovery from physical withdrawal symptoms, acamprosate was not more effective than placebo. One meta-analysis was designed to study differential outcomes as a consequence of the duration of detoxification and abstinence prior to initiating treatment. Nineteen trials were analyzed that used acamprosate alone and in combination with various other interventions (naltrexone with and without adjunctive psychotherapies). In the 5,559 subjects included in this part of the meta-analysis, detoxification prior to treatment was associated with larger medication effects. These data indicate that the outcome of acamprosate treatment is time sensitive and may depend on the timing of intervention in relation to acute detoxification.

Several studies have attempted to evaluate and compare the usefulness of different psychotherapies (e.g., dynamically oriented individual psychotherapy, cognitive behavioral individual coping skills therapy, group therapy, or various brief interventions) in the treatment of abstinence from alcohol when used in combination with acamprosate. To date, it remains unclear whether any particular psychotherapy is superior to another. The literature does substantiate that acamprosate is more effective when administered concurrently with psychotherapy. Brief *motivational enhancement therapy* (MET) is readily learned and can be implemented in psychiatric and primary care settings alike. This intervention focuses on improving a person’s motivation to achieve and sustain abstinence. Given that acamprosate has been shown to be more effective when abstinence is the goal of treatment, MET holds promise as a psychotherapeutic adjunct to acamprosate.

Additional studies have investigated the efficacy of acamprosate in comparison, or in conjunction with, naltrexone as compared to acamprosate alone, acamprosate + psychosocial interventions, and placebo. These studies haven't shown any efficacy of acamprosate over placebo with or without psychotherapy. Results from a 16-week randomized controlled trial, including 1,383 recently alcohol-abstinent volunteers, demonstrated no evidence of the efficacy of acamprosate alone or together with *combined behavioral intervention* (CBI) when compared to other treatments (e.g., naltrexone, CBI, and placebo alone or in various combinations). The largest limitation of this study was that subjects were only given 4 days for detoxification in an outpatient program. Studies that reported no additional benefit from the combined acamprosate + psychotherapy treatment had additional limitations, including the following: (i) lack of randomization in one open-label naturalistic study of 753 alcohol-dependent patients; and (ii) presenting the participants with the option to be randomized to the no psychotherapy arm, which may have created a selection bias against participants with a particular interest in receiving psychotherapy as documented in one randomized, multicenter study of 248 patients.

Using the structural clinical interview for DSM-IV and *International Statistical Classification of Diseases, Tenth Revision* (ICD-10), another protocol studied 426 subjects meeting DSM-IV criteria for alcohol dependence. All subjects were recruited from inpatient wards for medical detoxification to which their general practitioners had referred them. Following a mean of 18 days' inpatient detoxification, eligible patients were treated for 12 weeks in the outpatient setting with acamprosate, naltrexone, acamprosate + naltrexone, or placebo in conjunction with brief psychotherapy. Exclusion criteria included dependence or abuse of another substance other than cannabis or nicotine, a comorbid psychiatric disorder for which subjects were receiving medication, and unstable medical conditions. The study showed no effect on abstinence in those who received treatment as compared to those who received placebo. Limitations were related to higher acuity since participants were recruited from an inpatient setting, in addition to a relatively long detoxification period followed by a relatively short outpatient treatment duration. Furthermore, these data likely are not generalizable as patients with alcohol dependence often have comorbid psychiatric and medical conditions. Out of all the patients screened, for example, only 43% met inclusion criteria, and only 31% of all eligible patients chose to participate, which is suggestive that selection bias may have influenced the results.

One study investigated the efficacy of acamprosate with respect to gender using a gender-specific meta-analysis of individual patient data obtained from double-blind, randomized controlled trials with inpatients carrying diagnoses of alcohol dependence based on DSM-III or IV criteria. Records were obtained from 1,317 women and 4,793 men from 22 studies across 18 countries. Exclusion criteria were pregnancy, lactation, or refusal to use birth control in women of childbearing age and potential, significant comorbid psychiatric disorders for which patients were prescribed medication, medical disorders such as renal failure, and concurrent use of another pharmacologic intervention for alcohol use. Fourteen trials had durations of at least 6 months, and 5 lasted longer than 1 year. Patients needed to have at least one quantitative measure of drinking behavior to assess treatment efficacy. The four endpoint variables included the percentage of abstinent days, percentage of no heavy drinking days, rates of complete abstinence, and rates of no heavy drinking over the study's duration. There was an incremental gain of 10.4% in the percentage of abstinent days (95% confidence interval [CI] 7.1 to 13.7, $p < .001$), a gain of 11% in the percentage of no heavy drinking days, an odds ratio of 1.9 for complete abstinence rates (CI 1.6 to 2.2, $p < .001$), and an odds ratio of 1.9 (CI 1.6 to 2.3, $p < .001$) for a rate of no heavy drinking over the studies' durations. Importantly, outcomes in acamprosate efficacy, safety, or tolerability did not differ between men and women.

More recent studies have set out to investigate whether certain subtypes of persons with alcohol use disorder may respond preferentially to naltrexone or acamprosate. The central hypothesis of these studies is that individuals with significant alteration of the opioidergic system, who drink alcohol primarily in the context of positive reinforcement (reward drinkers), would benefit more from naltrexone, while individuals with significant glutamatergic dysfunction, who primarily drink alcohol in the context of negative reinforcement (relief drinkers) may benefit more from acamprosate. One of these studies utilized the data from a 16-week randomized controlled trial, including 1,383 recently alcohol-abstinent volunteers, comparing the efficacy of various combinations of pharmacotherapy and CBI. While the initial study did not find any significant difference in efficacy outcomes when combined with placebo alone, when subjects were stratified based on drinking for reward versus drinking for relief (i.e., preventing aversive effects from alcohol withdrawal), the data showed a significant effect for acamprosate. Specifically, the study found a significant interaction effect between the "high-relief/moderate-reward" class and acamprosate in the prediction of the percentage of drinking days at the end of treatment (week 16) and the percentage of heavy drinking days 10 weeks post-treatment (week 26) compared to placebo. These findings did not generalize to other "high-relief" classes, specifically those in the "high-relief/high-reward" class, suggesting that acamprosate may be more effective in treating persons who have higher relief temptation relative to reward temptation. Importantly, inpatient detoxification prior to initiating the medication in this study was 4 days.

Another study used data from a randomized controlled trial conducted in Germany testing the efficacy of naltrexone and acamprosate in the treatment of alcohol dependence. 426 participants were randomly assigned to naltrexone, acamprosate, or placebo over the course of 12 weeks. All participants were recruited from inpatient medical detoxification programs of 2 to 3 weeks' duration (on average, participants were abstinent for 22 days prior to initiating the study). In contrast to the data above, this study did not find an effect of acamprosate for the high-relief subgroup. The authors postulated that this discrepancy might be explained by the fact that after an average of 22 days' abstinent, hyperglutamatergic states may have already started to normalize. These studies may further highlight the importance of initiating acamprosate shortly after acute detoxification when glutamate dysregulation remains high.

Other studies have established that pretreatment commitment to abstinence could be an important moderator of treatment response to acamprosate. Abstinence prior to initiating acamprosate appears to be an important moderator of treatment, and shorter abstinence (usually 1 week or less) appears to be another moderator of responsiveness to acamprosate. Finally, it appears that acamprosate's effect may be dose dependent and that patients with low to normal body masses indices do better with acamprosate compared to their overweight or obese counterparts. This may, in part, explain the differential success of acamprosate in Japanese and European studies over ones conducted in the United States.

A 2018 meta-analysis set out to confirm the efficacy of acamprosate in reducing insomnia in subjects with alcohol use disorder. Twelve studies were found, including 3,508 patients. The main endpoint of the study was a change in Short Sleep Index (SSI), which was derived from answers subjects had provided in the Hamilton Depression and Hamilton Anxiety Scales. The questions included in the SSI are: (a) insomnia early (difficulty falling asleep), (b) insomnia middle (being restless and disturbed during the night), (c) insomnia late (waking in the early hours of the morning), and (d) difficulty in falling asleep, broken sleep, unsatisfying sleep and fatigue on waking, dreams, nightmares, and night terrors.

After 6 months of treatment with acamprosate, the mean SSI decrease over baseline was −26% and −45% for placebo and acamprosate, respectively (treatment effect = 19%, 12.5–25.5; $p < .001$). Using a univariate mediation model, the authors found that the mediating effect of abstinence on insomnia accounted for 55.7%. The remaining unexplained effect of acamprosate was 44.3%. Limitations of this study include that one of the included studies did not require subjects to have stopped drinking prior to initiating acamprosate and that many subjects had returned to alcohol use by the end of the 6-month investigation period. Other studies have shown that acamprosate improves sleep continuity and architecture measured by polysomnography. DSF, naltrexone, and nalmefene have been shown to have deleterious effects on sleep, so there is only low-level evidence thus far that acamprosate may improve insomnia in persons with alcohol use disorder; it may hold advantages for persons in whom insomnia is a primary concern.

Mild diarrhea or loose stool was the only statistically significant adverse event documented in a large meta-analysis. It is estimated that up to 10% of patients may experience mild diarrhea; importantly, those who suffered from diarrhea were no more likely than their counterparts to discontinue the medication.

These data are limited in that the randomized controlled trial with the highest power failed to demonstrate the medication's efficacy as compared to placebo or naltrexone and that many studies supporting acamprosate's efficacy are short. Taken cumulatively, however, acamprosate appears to be a safe and effective adjunct to psychosocial alcohol rehabilitation programs. It is most useful for patients receiving concomitant psychotherapy, and whose stated goal upon initiating treatment is abstinence.

Effects on Specific Organs and Systems

The most serious adverse effect of acamprosate may be related to the urogenital system since there have been several documented cases of patients developing acute renal failure while receiving acamprosate treatment. However, no causal relationship has been established between acamprosate and any specific physiologic mechanism that may result in acute kidney impairment. Less severe but more frequent are acamprosate's effects on the cardiovascular (palpitation and syncope), respiratory (increased cough, rhinitis, dyspnea), digestive (diarrhea, nausea, vomiting, and abdominal pain), musculoskeletal (myalgias and arthralgias), nervous (somnolence, amnesia, tremor), and dermatologic (pruritus) systems.

Therapeutic Indications

Acamprosate is an FDA-approved agent for maintenance of abstinence from alcohol in patients with alcohol dependence who are abstinent at treatment initiation. The efficacy of acamprosate in promoting abstinence has not been demonstrated in subjects who have not undergone detoxification and have not achieved alcohol abstinence prior to beginning acamprosate treatment. The efficacy of acamprosate in promoting abstinence from alcohol in patients with polysubstance abuse has not been adequately assessed. Treatment with acamprosate should be a part of a comprehensive management program that includes psychosocial support.

Other Disorders. Acamprosate is only FDA approved for the treatment of alcohol use disorder.

The literature on other uses for acamprosate over the last few years is expanding. The medication appears to hold promise in the treatment of *fragile-X-syndrome* (FXS)-associated *autistic spectrum disorders* (ASDs) and tinnitus.

Excessive glutamatergic and deficient GABAergic neurotransmission is found in patients with both FXS and with idiopathic ASD. Extensive literature supports a correlation between excessive mGluR5 and deficient $GABA_A$ neurotransmission in FXS-associated ASD. Preliminary studies showed that three adults with FXS and ASD who were treated with a mean dose of 1,221 mg/day of acamprosate for a mean duration of 21.3 weeks achieved clinical and statistical improvement in social–communicative language. A subsequent 10-week prospective open-label trial of acamprosate in 12 patients (mean age 11.9 years old) with comorbid FXS and ASD found improvement in social behavior and inattention/hyperactivity in 75% of the subjects. Moreover, the medication was well tolerated with no severe or serious adverse effects. It is currently the preferred agent in patients with FXS and co-occurring alcohol use disorders.

A derivative of amyloid-beta precursor protein (APP) secreted APPα (sAPPα), has neuroproliferative effects, and has been shown to be elevated in the plasma of persons with ASD compared to control subjects. A pilot study of 15 subjects treated with acamprosate for 10 weeks showed a significant decrease in sAPP-total and sAPPα in subjects with FXS-associated or idiopathic ASD ($p < 0.05$), proposing a mechanism by which acamprosate might treat these disorders on the molecular level. The largest limitation of this study was its very low power and that 13 of the patients enrolled were receiving other pharmacologic therapies. There were no adverse effects reported in patients receiving acamprosate. The medication may benefit patients with FXS and ASD, though no high-powered randomized controlled trials have been conducted to date.

A prospective, randomized, placebo-controlled, double-blind, cross-over study investigated the use of acamprosate in treating 40 subjects with sensorineural tinnitus. The patients were randomized to receive three 333-mg tablets of acamprosate daily or placebo for 45 days, followed by a washout period of 7 days and then crossed over to acamprosate or placebo for another 45 days. 92.5% of subjects receiving acamprosate demonstrated objective and subjective improvement in tinnitus as opposed to 12.5% subjective improvement seen with placebo. Objective improvement of tinnitus was measured by tinnitus matching for loudness by decibel. On this objective measure of tinnitus, 80% of the treatment group showed improvement as compared to 0.0% of the placebo group. Another randomized controlled trial divided 20 subjects with chronic tinnitus into either acamprosate or placebo. This study indicated subjective relief of tinnitus as well as some degree of electrophysiologic improvement at the level of the cochlear and the distal portion of the auditory nerve in the 9 subjects who received acamprosate versus the 11 who received placebo. Numerous case reports also reflect subjective improvement in tinnitus with acamprosate. Larger studies are needed for FDA approval, but evidence suggests some efficacy of acamprosate in the treatment of tinnitus.

There is insufficient data to determine acamprosate’s utility in the treatment of other SUDs. This may be because the medication is relatively new, and small pilot studies have failed to demonstrate superiority to placebo in the treatment of stimulant (cocaine) use disorder, while preliminary studies in rodents were promising. This topic awaits further research.

Precautions and Adverse Reactions

Precautions. The dose of acamprosate will need to be adjusted for patients with moderate renal impairment (i.e., creatinine clearance of 30 to 50 mL/minute). Patients with severe renal impairment (creatinine clearance of ≤30 mL/minute) should not be prescribed acamprosate. Because acamprosate does not rely on hepatic metabolism, it can be used in patients with mild to moderate hepatic impairment (Child–Pugh groups A and B). It has not been used in Child–Pugh C liver disease and is not approved for use in this population.

In placebo-controlled trials, adverse events such as suicidal ideation, suicide attempts, and even completed suicides were reported more frequently in acamprosate-treated patients than in patients who received placebo (1.4% vs. 0.5% in studies of 6 months or less; 2.4% vs. 0.8% in year-long studies). Completed suicides in the pooled acamprosate group were 0.13% (3 out of 2,272 patients) and 0.10% (2 out of 1,962) in the placebo group. Depression was reported at similar rates in both treatment groups. The interrelationship between alcohol dependence, depression, and suicidality is well recognized and complex, and since many of these adverse events occurred during the periods of addiction relapse, it was hard to determine any direct link between the clinical course of recovery from alcoholism and the emergence of suicidality. Patients with alcohol use disorders, regardless of whether or not they are treated with acamprosate, should be monitored for the development of symptoms of depression and suicidality since suicide attempts and completions occur more frequently in patients with alcohol use disorders.

Although the safety of acamprosate during pregnancy has not been determined, the agent has been shown to have a teratogenic effect on the fetus in rodents. The benefit of acamprosate use during pregnancy should be carefully weighed against possible adverse effects. It is also not recommended for use in nursing mothers.

Adverse Reactions. More frequent and less serious side effects may include headaches, insomnia, stomach upset with nausea, vomiting, and abdominal pain, as well as sexual dysfunctions (e.g., decreased libido, frigidity, and impotence). In clinical trials, the most commonly reported side effects were diarrhea and pruritus, though these adverse effects were usually transient, and in recent meta-analyses, patients with these adverse effects are no more likely than those receiving a placebo to discontinue the medication. Those who developed anxiety and depression while taking acamprosate were most likely to discontinue the medication, but it is difficult to parse out whether these mood symptoms are directly attributable to acamprosate or to underlying untreated mood disorders or mood disorders that developed as a result of prolonged alcohol use. More serious side effects include allergic reactions to the drug, arrhythmias, and fluctuations in blood pressure.

Drug Interactions

Because acamprosate does not rely on hepatic metabolism, it has relatively few drug–drug interactions. Coadministration of benzodiazepines, DSF, antidepressants, or alcohol does not change acamprosate's pharmacokinetic properties, which is fortunate, and these medications are commonly prescribed in patients with alcohol use disorders. However, coadministration of acamprosate with naltrexone may result in a 33% increase in the C_{max} of acamprosate, which requires no adjustment of the dosage. Similarly, the pharmacokinetics of naltrexone and its major metabolite 6-beta-naltrexol are unaffected when given concomitantly with acamprosate. Patients taking acamprosate with antidepressants more commonly reported both weight gain and weight loss in comparison to patients taking either medication alone.

The absorption of several agents (tetracyclines, quinolones, levothyroxine, atenolol, iron, alendronate, sodium fluoride, and zinc) may be decreased by the calcium component of acamprosate during concurrent administration. The absorption of acamprosate can also be decreased by food. Acamprosate has shown no inhibiting or inducing potential both in vitro and in vivo on the enzymes of the cytochrome P-450 system.

Laboratory Interference

Acamprosate is not known to interfere with any laboratory tests.

Dosage and Administration

Initiation of therapy should begin after alcohol detoxification has taken place and the patient has abstained from alcohol for 5 to 7 days. Just as it should not be initiated during acute detoxification, a review of the literature substantiates that it may not be superior to placebo if the medication is initiated more than 25-days post acute alcohol detoxification. Acamprosate treatment should not be interrupted if relapses occur and are recommended to proceed for 12 months.

Acamprosate is dosed by weight: patients >60 kg should receive 1,998 mg/day in divided doses with meals, while patients <60 kg should receive 1,332 mg/day in three divided doses with meals. The drug has not been sufficiently studied in pediatric and geriatric populations.

ACKNOWLEDGMENTS

The author would like to acknowledge Dr. Iliyan Ivanov and Dr. Alison Ungar for their contributions to earlier editions of this chapter.

FURTHER READINGS

Bernier M, Mitchell SJ, Wahl D, et al. Disulfiram treatment normalizes body weight in obese mice. *Cell Metab*. 2020;32(2):203–214.e4.

Diehl A, Ulmer L, Mutschler J, et al. Why is disulfiram superior to acamprosate in the routine clinical setting? A retrospective long-term study in 353 alcohol-dependent patients. *Alcohol*. 2010;45:271.

Erickson CA, Ray B, Maloney B, et al. Impact of acamprosate on plasma amyloid-β precursor protein in youth: a pilot analysis in fragile X syndrome-associated and idiopathic autism spectrum disorder suggests a pharmacodynamics protein marker. *J Psychiatr Res*. 2014;59:220.

Fuller RK, Branchey L, Brightwell DR, et al. Disulfiram treatment of alcoholism. A Veterans Administration cooperative study. *JAMA*. 1986;256(11):1449.

Gaval-Cruz M, Weinshenker D. Mechanisms of disulfiram-induced cocaine abstinence: antabuse and cocaine relapse. *Mol Interv*. 2009;9(4):175–187.

Hammarberg A, Jayaram-Lindstöm N, Beck O, Franck J, Reid MS. The effects of acamprosate on alcohol-cue reactivity and alcohol priming in dependent patients: a randomized controlled trial. *Psychopharmacology (Berl)*. 2009;205(1):53–62.

Jørgensen CH, Pedersen B, Tønnesen H. The efficacy of disulfiram for the treatment of alcohol use disorder. *Alcohol Clin Exp Res*. 2011;35:1749–1758.

*Jonas DE, Amick HR, Feltner C, et al. Pharmacotherapy for adults with alcohol use disorders in outpatient settings: a systematic review and meta-analysis. *JAMA*. 2014;311(18):1889–1900.

Kampangkaew JP, Spellicy CJ, Nielsen EM, et al. Pharmacogenetic role of dopamine transporter (SLC6A3) variation on response to disulfiram treatment for cocaine addiction. *Am J Addict*. 2019;28(4):311–317.

*Karila L, Reynaud M, Aubin HJ, et al. Pharmacological treatments for cocaine dependence: is there something new? *Curr Pharm Des*. 2011;17(14):1359–1368.

Kiefer F, Mann K. Acamprosate: how, where and for whom does it work? Mechanism of action, treatment targets and individualized therapy. *Curr Pharm Des*. 2010;16(9):2098–2102.

Krampe H, Stawicki S, Hoehe MR, Ehrenreich H. Outpatient Long-term Intensive Therapy for Alcoholics (OLITA): a successful biopsychosocial approach to the treatment of alcoholism. *Dialogues Clin Neurosci*. 2007;9:399–412.

Kufahl PR, Watterson LR, Olive MF. The development of acamprosate as a treatment against alcohol relapse. *Expert Opin Drug Discov*. 2014;9(11):1355–1369.

Laaksonen E, Koski-Jännes A, Salaspuro M, Ahtinen H, Alho H. A randomized, multicentre, open-label, comparative trial of disulfiram, naltrexone and acamprosate in the treatment of alcohol dependence. *Alcohol Alcohol*. 2008;43:53–61.

Lu C, Li X, Ren Y, Zhang X. Disulfiram: a novel repurposed drug for cancer therapy. *Cancer Chemother Pharmacol*. 2021;87(2):159–172.

Maisel NC, Blodgett JC, Wilbourne PL, Humphreys K, Finney JW. Meta-analysis of naltrexone and acamprosate for treating alcohol use disorders: when are these medications most helpful? *Addiction*. 2013;108(2):275–293.

Mann K, Kiefer F, Smolka M, Gann H, Wellek S, Heinz A; PREDICT Study Research Team. Searching for responders to acamprosate and naltrexone in alcoholism treatment: rationale and design of the PREDICT study. *Alcohol Clin Exp Res*. 2009;33(4):674–683.

Mann K, Lemenager T, Hoffman S, et al; PREDICT Study Team. Results of a double-blind, placebo-controlled pharmacotherapy trial in alcoholism conducted in Germany and comparison with the US COMBINE study. *Addict Biol*. 2013;18(6):937–946.

Mann K, Roos CR, Hoffmann S, et al. Precision medicine in alcohol dependence: a controlled trial testing pharmacotherapy response among reward and relief drinking phenotypes. *Neuropsychopharmacology*. 2018;43(4):891.

Mason BJ, Goodman AM, Chabac S, Lehert P. Effect of oral acamprosate on abstinence in patients with alcohol dependence in a double-blind placebo-controlled trial: the role of patient motivation. *J Psychiatr Res*. 2006;40(5):383–393.

Mason BJ, Lehert P. Acamprosate for alcohol dependence: a sex-specific meta-analysis based on individual patient data. *Alcohol Clin Exp Res*. 2012;36(3):497–508.

O'Farrell TJ, Fals-Stewart W. Family-involved alcoholism treatment. An update. *Recent Dev Alcohol*. 2001;15:329–356.

Oka M, Hirouchi M, Tamura M, Sagahara S, Oyama T. Acamprosate {monocalcium bis(3-acetaminopropane-1-sulfonate)} reduces ethanol-drinking behavior in rats and glutamate-induced toxicity in ethanol-exposed primary rat cortical neuronal cultures. *Eur J Pharmacol*. 2013;718:323.

Pani PP, Trogu E, Vacca R, Amato L, Vecchi S, Davoli M. Disulfiram for the treatment of cocaine dependence. *Cochrane Database Syst Rev*. 2010;(1):CD007024.

Perney P, Lehert P. Insomnia in alcohol-dependent patients: prevalence, risk factors and acamprosate effect: an individual patient data meta-analysis. *Alcohol Alcohol*. 2018;53(5):611.

Petrakis IL, Nich C, Ralevski E. Psychotic spectrum disorders and alcohol abuse: a review of pharmacotherapeutic strategies and a report on the effectiveness of naltrexone and disulfiram. *Schizophr Bull*. 2006;32(4):644–654.

Roache JD, Kahn R, Newton TF, et al. A double-blind, placebo-controlled assessment of the safety of potential interactions between intravenous cocaine, ethanol, and oral disulfiram. *Drug Alcohol Depend*. 2011;119(1–2):37.

Roos CR, Mann K, Witkiewitz K. Reward and relief dimensions of temptation to drink: construct validity and role in predicting differential benefit from acamprosate and naltrexone. *Addict Biol*. 2017;22(6):1528.

Sharma DK, Kaur S, Singh J, Kaur I. Role of acamprosate in sensorineural tinnitus. *Indian J Pharmacol*. 2012;44(1):93.

*Skinner MD, Lahmek P, Pham H, Aubin HJ. Disulfiram efficacy in the treatment of alcohol dependence: a meta-analysis. *PLoS One*. 2014;9(2):e87366.

Spanagel R, Vengeliene V, Jandeleit B, et al. Acamprosate produces its anti-relapse effects via calcium. *Neuropsychopharmacology*. 2014;39(4):783.

33.14 Fast-Acting Pharmacologic Treatments

33.14a Esketamine and Ketamine

CARLOS A. ZARATE, JR., M.D., AND LAWRENCE T. PARK, A.M., M.D.

Neuroscience-Based Nomenclature (NbN)	
Pharmacologic Domain	Glutamate
Mechanism of Action	Receptor antagonist (NMDA)

HISTORY

The discovery that racemic (*R*,*S*)-ketamine (hereafter referred to as ketamine) has antidepressant properties has been described as the most significant development in the treatment of depression in the past 60 years. While the monoamine era of pharmacotherapy yielded a host of moderately effective, relatively slow-acting, mechanistically similar drugs, ketamine offers the prospect of rapid relief of symptoms via a novel mechanism of antidepressant action. Our understanding of this underlying mechanism has the potential to serve as a basis for reconceptualizing the neurobiology of depression as well as a platform for developing innovative pharmacotherapeutic strategies.

Ketamine was originally synthesized in 1962 as a safer anesthetic alternative to phencyclidine (PCP), which was associated with significant emergence delirium and other adverse effects. As a structurally related analog, ketamine shares PCP's neurochemical activity as an *N*-methyl-D-aspartate (NMDA) antagonist but is less potent and has a shorter duration of action. Ketamine received FDA approval in 1970 as an anesthetic for procedures that do not require skeletal muscle relaxation, as an induction agent for short procedures, and as a supplement to low-potency anesthetic agents. In contrast to other anesthetic agents, ketamine produces a dissociative state with relatively little sedation. It has excellent analgesic properties and a relatively large therapeutic index. Given these characteristics, it remains widely used in adult and pediatric anesthesia as well as in veterinary medicine. Ketamine is on the World Health Organization's list of essential medicines.

Trullas and Skolnick (1990) first postulated that substances capable of reducing neurotransmission at the NMDA receptor complex might have potential as a new class of antidepressants. Based on the theory that NMDA mediates behavioral changes associated with inescapable stress, they demonstrated that ketamine reduced depressive-like symptoms in stress-related animal models of depression. Berman and colleagues (2000) subsequently conducted the first controlled human study, which administered a single intravenous (IV) ketamine infusion (0.5 mg/kg) to a small group of individuals primarily diagnosed with major depressive disorder (MDD). Compared to saline placebo, ketamine demonstrated antidepressant effects within 72 hours of infusion. Zarate and colleagues (2006) subsequently conducted a larger, saline placebo-controlled trial using a randomized, within-subjects crossover design. The study confirmed ketamine's rapid antidepressant effects (onset within 24 hours) in individuals with treatment-resistant MDD (TRD) as well as those with treatment-resistant bipolar depression. Additional studies also confirmed that these rapid antidepressant effects peaked within 1 day of infusion and lasted 3 to 7 days postinfusion, as well as ketamine's effectiveness in both TRD and bipolar depression and the feasibility of a repeated-dose strategy to extend the duration of these effects.

The discovery of ketamine's antidepressant properties renewed interest in pharmacologic research of compounds with related mechanisms of action. Investigators examined ketamine's enantiomers in an attempt to reproduce ketamine's antidepressant properties but minimize its adverse secondary effects. Researchers initially focused on (*S*)-ketamine (esketamine), which demonstrates greater potency at the NMDA receptor than (*R*)-ketamine (arketamine). Developed for intranasal administration, esketamine was found to be safe and effective for treating TRD. Intranasal esketamine (Spravato) received FDA approval for TRD in March 2019 (ketamine itself does not have an FDA indication for depression) and for use in adults with MDD with acute suicidal ideation or behavior in August 2020. Because of concerns about dissociative adverse events and the potential for abuse, the FDA approval required that esketamine use be managed through a risk evaluation and mitigation strategy (REMS). REMS requires that esketamine be dispensed and administered only in a REMS-certified, medically supervised health care setting where patients are monitored for at least 2 hours postadministration. Several recent treatment guidelines now incorporate ketamine and esketamine into the treatment algorithm.

Interestingly, research with animal models suggests that (*R*)-ketamine has greater and longer-lasting antidepressant-like effects. In addition, a metabolite of the (*R*)-enantiomer, (2*R*,6*R*)-hydroxynorketamine (HNK), appears to be essential for achieving antidepressant-like effects without apparent dissociative-like symptoms or abuse potential. Phase I human trials with (2*R*,6*R*)-HNK have recently been completed, and Phase II trials are planned.

CHEMISTRY

Ketamine (and ketamine hydrochloride) is an arylcycloalkylamine and a synthetic derivative of cyclohexanone. Its chemical name is 2-(2-chlorophenyl)-2-(methylamino)cyclohexan-1-one (IUPAC nomenclature). The molecular formula is $C_{13}H_{16}C1NO$, and the molecular weight is 237.72. Ketamine is a white to off-white crystalline powder soluble in water and lipid solutions, freely soluble in methanol, soluble in alcohol, and sparingly soluble in chloroform. It is often supplied as a clear liquid. Ketamine has an asymmetric carbon at the C2 position and is made up of a mixture of (*S*) and (*R*)-enantiomers. (*S*)-ketamine has four times greater affinity at the NMDA receptor than (*R*)-ketamine. The molecular structure is shown in Figure 33.14a–1.

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Ketamine can be delivered orally, intranasally, rectally, IV, intramuscularly, subcutaneously, and epidurally. Bioavailability and time to maximum concentration ($T_{\max}$) differ by route of administration, with IV demonstrating 100% bioavailability ($T_{\max}$ 3 minutes), intranasal 8% to 45% bioavailability ($T_{\max}$ 10 to 20 minutes), and oral 17% to 29% bioavailability ($T_{\max}$ 30 minutes). Ketamine absorption is very rapid. The distribution half-life ($T_{1/2a}$) is on the order of 2 to 4 minutes, and 17% is absorbed after first-pass metabolism. Ketamine is 53.5% plasma protein bound. The elimination half-life ($T_{1/2b}$) of ketamine in adults is 2 to 4 hours, and the elimination $T_{1/2b}$ of esketamine is 5 hours. Ketamine and esketamine undergo mainly hepatic metabolism via cytochrome P450 (CYP) 3A4 and 2B6. Most of the elimination is via the urine in the form of metabolites. Norketamine is the principal initial metabolite and is further metabolized to 18 other distinct compounds, including (2*R*,6*R*)-HNK (see Fig. 33.14a–2).

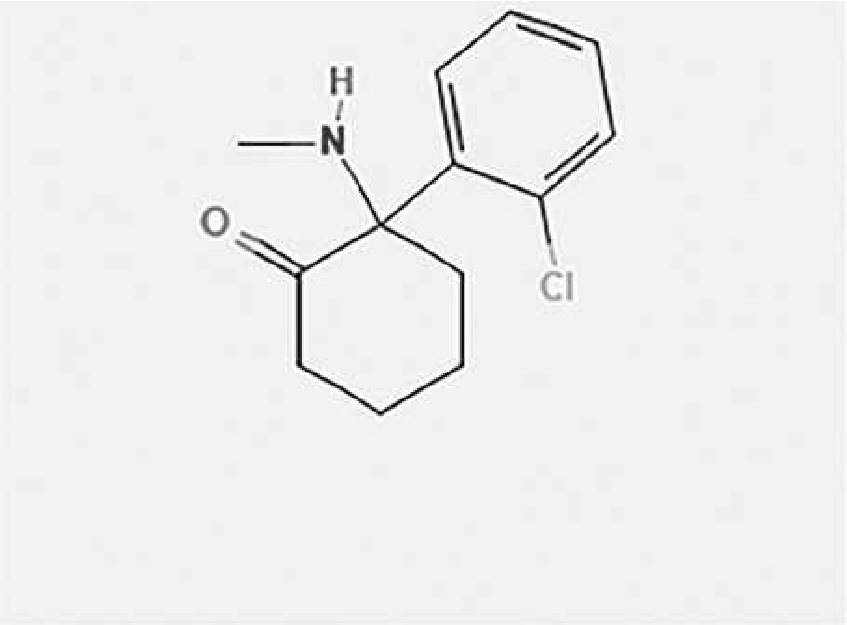

Enantiomers

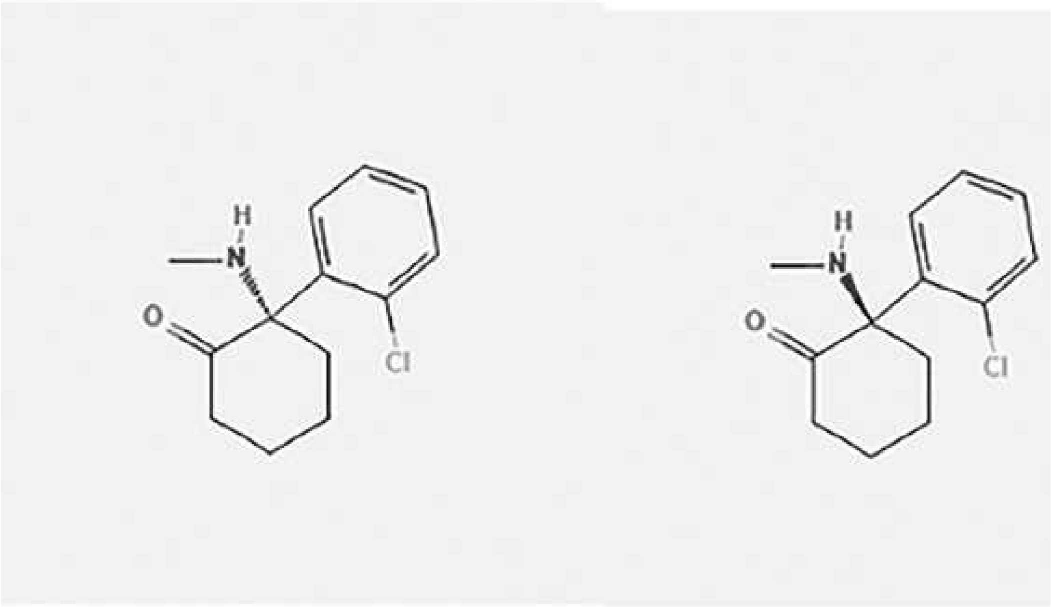

(*S*)-ketamine **(*R*)-ketamine**

FIGURE 33.14a–1. Molecular structure of ketamine and its enantiomers. (From National Center for Biotechnology Information. PubChem Compound Summary for CID 15851, Ketamine. Accessed August 30, 2021. https://pubchem.ncbi.nlm.nih.gov/compound/Ketamine)

Pharmacodynamics

Ketamine and its metabolites bind to a variety of biologic targets. Although its primary activity is as a noncompetitive NMDA receptor antagonist, ketamine also interacts with opioid receptors (e.g., kappa, mu, and delta subtype agonist; mu opioid 2 receptor antagonist), with the monoaminergic system (e.g., dopamine 2 receptor agonism), and with the serotonergic system (e.g., 5-HT1, 5-HT2 receptor antagonism, 5-HT3 receptor weak antagonism). It also weakly inhibits the dopamine- and sodium-dependent norepinephrine transporters, has noradrenergic effects on the cholinergic system (e.g., alpha-7-nicotinic receptor subunit antagonist, binds at the muscarinic receptor, cholinesterase inhibitor), is a glycine receptor agonist, and is a sigma receptor 1 and 2 agonist. Finally, ketamine also binds the voltage-sensitive calcium and voltage-gated sodium ion channels, is an antagonist at the neurokinin 1 receptor, an agonist at the D-serine site, and inhibits nitric oxide synthase receptors. Ketamine has no significant direct effects on gamma aminobutyric acid (GABA) receptors.

The precise mechanism of rapid antidepressant response associated with ketamine remains unclear, and this is a rapidly evolving area of research. A proposed schema of glutamatergic mechanisms of interest (not comprehensive) is presented in Figure 33.14a–3. Early theories postulated a “glutamate surge” resulting from NMDA receptor antagonism as a necessary component of the antidepressant response, and mounting evidence suggests that ketamine promotes neuroplasticity by facilitating synaptogenesis and synaptic potentiation. Increased synaptic glutamate release is thought to result from NMDA antagonism at the GABAergic inhibitory interneuron that decreases inhibitory tone on the presynaptic pyramidal neuron. This “disinhibition” of the presynaptic neuron increases synaptic glutamate release and also increases glutamatergic tone on the postsynaptic neuron.

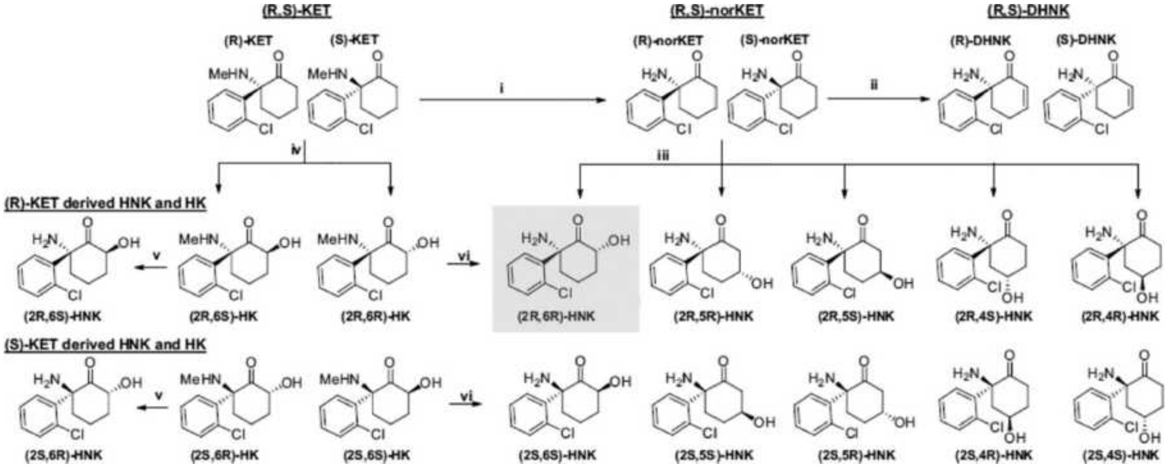
(R,S)-KET
(R)-KET
(S)-KET
(R,S)-norKET
(R)-norKET
(S)-norKET
(R,S)-DHNK
(R)-DHNK
(S)-DHNK
i
ii
iii
iv
v
vi
(R)-KET derived HNK and HK
(2R,6S)-HNK
(2R,6S)-HK
(2R,6R)-HK
(2R,6R)-HNK
(2R,5R)-HNK
(2R,5S)-HNK
(2R,4S)-HNK
(2R,4R)-HNK
(S)-KET derived HNK and HK
(2S,6R)-HNK
(2S,6R)-HK
(2S,6S)-HK
(2S,6S)-HNK
(2S,5S)-HNK
(2S,5R)-HNK
(2S,4R)-HNK
(2S,4S)-HNK

FIGURE 33.14a–2. The metabolism of ketamine (KET). Ketamine is metabolized in vivo via P450 enzymatic transformations. **i**: (*R*,*S*)-KET is selectively demethylated to (*R*,*S*)-norketamine (norKET). **ii**: NorKET can then be dehydrogenated to (*R*,*S*)-dehydronorketamine (DHNK). **iii**: Alternatively, norKET can be hydroxylated to the hydroxynorketamines (HNKs). **iv**: (*R*,*S*)-KET can also be hydroxylated at the 6 position to give either the *E*-6-hydroxyketamine ([2*S*,6*R*;2*R*,6*S*]-HK) or *Z*-6-hydroxyketamine ([2*S*,6*S*;2*R*,6*R*]-HK). **v**: Demethylation of (2*S*,6*R*;2*R*,6*S*)-HK produces (2*S*,6*R*;2*R*,6*S*)-HNK. **vi**: Demethylation of (2*S*,6*S*;2*R*,6*R*)-HK further gives (2*S*,6*S*;2*R*,6*R*)-HNK. (Reproduced from Zanos P, Moaddel R, Morris PJ, et al. NMDAR inhibition-independent antidepressant actions of ketamine metabolites. *Nature*. 2016;533:481–486. doi: 10.1038/nature17998.)

Several other NMDA-related processes have also been implicated in promoting neuroplasticity and, thus, mediating antidepressant response. These include blockade of spontaneous synaptic NMDA receptor activation, which inhibits eukaryotic elongation factor 2 kinase (eEF2K), leading to brain-derived neurotrophic factor (BDNF) translation and protein synthesis; inhibition of extrasynaptic NMDA receptors, which suppresses mechanistic target of rapamycin (mTOR) function and induces protein synthesis; and inhibition of NMDA receptor-dependent excessive burst firing activity of lateral habenula neurons. Other non-NMDA glutamate-related mechanisms of ketamine's rapid antidepressant effect have also been proposed. For instance, increased postsynaptic α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPAR) activation has been shown to mediate antidepressant-like response in animal models. AMPAR activation may result from increased synaptic glutamate release and/or ketamine metabolites (i.e., [2*R*,6*R]*-HNK). AMPAR activation may also increase BDNF release, activation of the tropomyosin receptor kinase B (TrkB) receptor, and activation of the mTOR complex, with the subsequent promotion of protein synthesis. Animal models and human trials also suggest that the presynaptic group 2 metabotropic glutamate receptor (mGlurR2) may play a role in antidepressant response; mGluR2 antagonism has been shown to enhance glutamate release and subsequent AMPAR activation, and mGluR2 antagonists are currently in Phase II trials. Another recently proposed mechanism of action concerns focal glutamatergic hyperactivity in the ventral anterior cingulate cortex (vACC), which is thought to potentially lead to global reductions in monoaminergic activity, resulting in depressive symptoms.

Nonglutamatergic mechanisms have also been hypothesized to contribute to ketamine's antidepressant properties. Recent attention has focused on the opioid system. In particular, mu opioid receptor agonism may mediate antidepressant response to ketamine based on observations that naltrexone, a primarily mu opioid antagonist, blocked ketamine's effects in animals and humans. Kappa opioid receptor involvement has also been implicated, given evidence that kappa receptor antagonists have been found to block the behavioral effects of ketamine in rodents.

CLINICAL DRUG STUDIES

Ketamine for MDD

Ketamine is available on the U.S. market with an anesthesia indication but is not FDA-approved for a psychiatric indication. While multiple formulations and routes of administration are available, IV ketamine administration has been the most widely studied in TRD and has generated the greatest body of evidence; to date, ketamine administered via other routes has been studied in only limited fashion, and insufficient data currently exist to support its safety and effectiveness. In addition, evidence for its use for indications other than TRD and suicidal thoughts and behaviors is preliminary and in need of replication.

Adequate scientific evidence supports the rapid, though short-term, effectiveness of ketamine for TRD. Zarate and colleagues (2006) found that a single dose of IV ketamine (0.5 mg/kg) had rapid antidepressant effects (within 24 hours of administration). This first adequately-powered, placebo-controlled trial examined single-dose IV ketamine (0.5 mg/kg) versus saline placebo infused over 40 minutes. The study used a within-subjects crossover design where participants were randomized to receive ketamine infusion then placebo infusion, or vice versa, 1 week apart. Eighteen individuals with TRD (defined as nonresponse to at least two previous antidepressant trials) were tapered off psychotropic medications and underwent a 2-week washout period. Depressive symptoms were measured with the 21-item Hamilton Depression Rating Scale (HAM-D). Ketamine infusion significantly reduced depressive symptoms within 2 hours compared to placebo, with maximal effects seen at 24 hours. Improvement lasted 3 days and symptoms typically returned to baseline by 7 days. Subsequent single-infusion ketamine trials reproduced this initial finding.

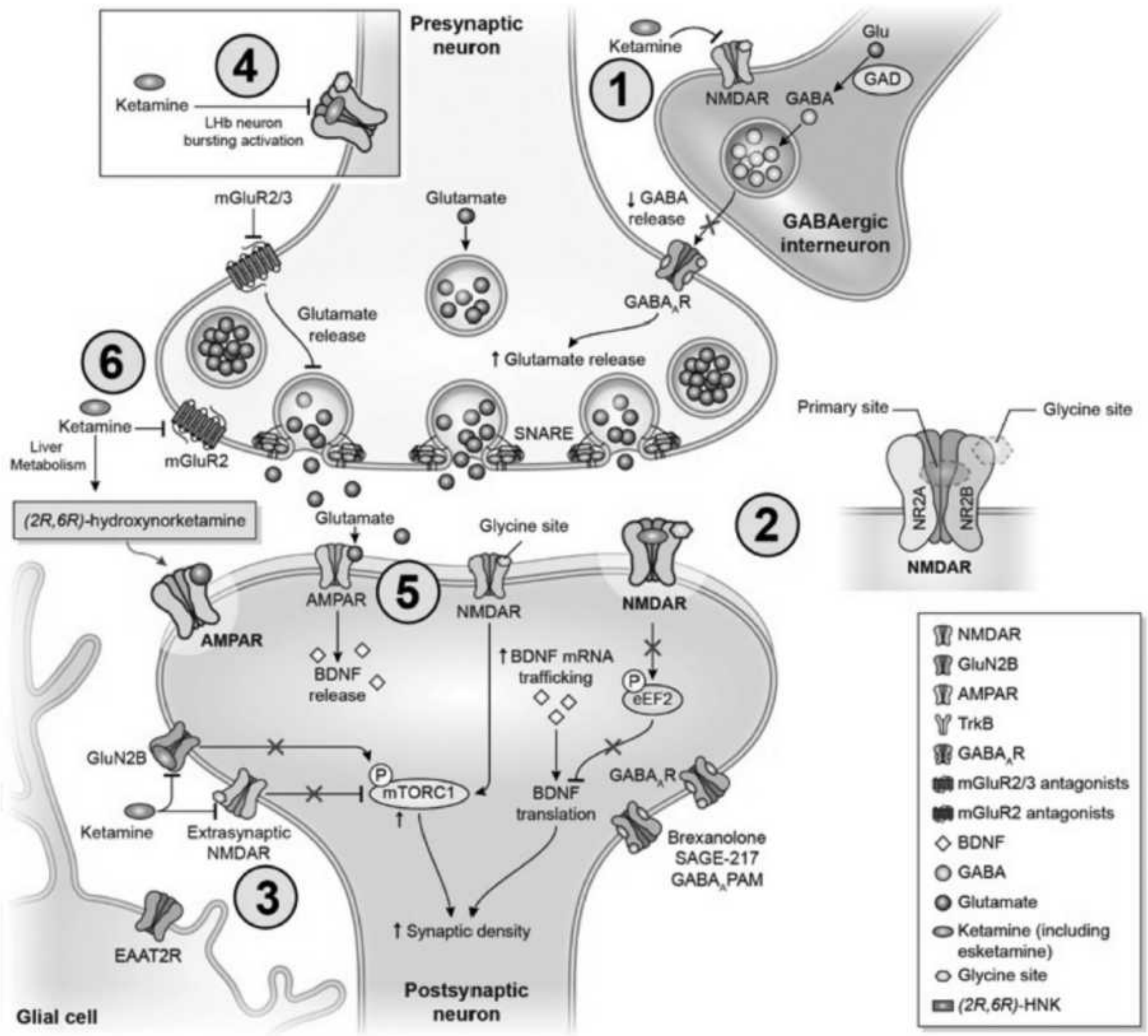
Presynaptic neuron
4
Ketamine
LHb neuron bursting activation
Ketamine
1
Glu
GAD
NMDAR
GABA
GABAergic interneuron
mGluR2/3
Glutamate
↓ GABA release
GABAAR
Glutamate release
↑ Glutamate release
6
Ketamine
mGluR2
Liver Metabolism
SNARE
Primary site
Glycine site
NR2A
NR2B
NMDAR
(2R,6R)-hydroxynorketamine
Glutamate
Glycine site
2
AMPAR
AMPAR
5
NMDAR
NMDAR
BDNF release
↑ BDNF mRNA trafficking
P
eEF2
GluN2B
P
mTORC1
BDNF translation
GABAAR
Ketamine
Extrasynaptic NMDAR
Brexanolone
SAGE-217
GABAAPAM
3
EAAT2R
↑ Synaptic density
Glial cell
Postsynaptic neuron
NMDAR
GluN2B
AMPAR
TrkB
GABAAR
mGluR2/3 antagonists
mGluR2 antagonists
BDNF
GABA
Glutamate
Ketamine (including esketamine)
Glycine site
(2R,6R)-HNK

FIGURE 33.14a–3. Proposed Mechanisms of Action of Ketamine and Esketamine; (1) **Disinhibition hypothesis**: i.e., blockage of N-methyl-D-aspartate (NMDA) receptors on gamma aminobutyric acid (GABA)-ergic inhibitory interneurons, leading to disinhibition of pyramidal neurons and enhanced glutamatergic firing; (2) **Inhibition of spontaneous synaptic NMDA receptor activation** inhibits eukaryotic elongation factor 2 kinase (eEF2K), leading to brain-derived neurotrophic factor (BDNF) translation and protein synthesis; (3) **Inhibition of extrasynaptic NMDA receptors** suppresses mechanistic target of rapamycin (mTOR) function and induces protein synthesis; and (4) **Inhibition of NMDA receptor-dependent excessive burst firing activity of lateral habenula (LHb) neurons**. Other proposed glutamate-related mechanisms of ketamine's rapid antidepressant effect include: (5**) Increased postsynaptic alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPAR) activation** from increased synaptic glutamate release and/or ketamine metabolites ([2*R*, 6*R*]-hydroxynorketamine [HNK]); (6) **Presynaptic metabotropic glutamate receptor (mGluR) (group 2) antagonism** leading to enhanced glutamate release and subsequent AMPAR activation. AMPAR activation leads to

increased release of BDNF, increased activation of the tropomyosin receptor kinase B (TrkB) receptor, increased activation of the mTOR complex, and subsequent promotion of protein synthesis. (Adapted from Henter ID, Park LT, Zarate CA Jr. Novel glutamatergic modulators for the treatment of mood disorders: current status. *CNS Drugs.* 2021;35(5):527–543. doi: 10.1007/s40263-021-00816-x.)

Multiple ketamine infusions have also been studied to address the short duration of ketamine's antidepressant effects. Several studies demonstrated the feasibility of up to six repeated infusions of IV ketamine administered two to three times per week. These studies found a comparable antidepressant response to ketamine as the single-dose studies, noting that effects typically lasted 2.5 to 3 weeks; ketamine was generally well tolerated. One study that compared twice-weekly versus thrice-weekly IV ketamine infusions over 4 weeks found no significant difference in response rates over a 15-day follow-up period. Another study that reviewed two Cochrane reviews, 14 meta-analyses, and 15 clinical trials concluded that 0.5 mg/kg of IV ketamine had rapid and robust antidepressant effects in MDD and bipolar depression. Lower doses were not efficacious. Interestingly, the treatment effect was observed to last longer for individuals with MDD rather than those with bipolar depression and for those who received repeated doses compared to a single dose.

Esketamine for MDD

In 2019, esketamine, administered in conjunction with oral antidepressants, received FDA approval for TRD in adults. The pivotal trials supporting the FDA application comprise a short-term and a long-term study. In the short-term study, 223 adult patients with TRD (defined as inadequate response to at least two different antidepressants in the current episode) were evaluated in a randomized, placebo-controlled, double-blind fashion over a 4-week period. Patients were tapered off their existing antidepressant medications and randomized to receive twice weekly intranasal esketamine (flexible dosing of 56 or 84 mg) or placebo in conjunction with open-label concomitant treatment with a newly initiated daily oral antidepressant (duloxetine, escitalopram, sertraline, or extended-release venlafaxine). As assessed using the Montgomery-Asberg Depression Rating Scale (MADRS), esketamine plus the newly initiated oral antidepressant significantly improved depressive symptoms (mean difference of four points) compared to placebo plus the newly initiated oral antidepressant. In the esketamine groups, most improvement was realized within 24 hours and continued over time through Day 28.

In the long-term study, 297 adults who were known remitters and responders to intranasal esketamine were enrolled in a randomized, double-blind, parallel-group, maintenance-of-effect study. In this 16-week initial trial of intranasal esketamine and an oral antidepressant, stable remitters (MADRS ≤ 12) and stable responders (MADRS reduction ≤ 50%) combined were randomized to receive intranasal esketamine or placebo maintenance treatment (weekly or every other week) in addition to an oral antidepressant. Of 176 patients who achieved stable remission, 24 (26.7%) in the esketamine group and 39 (45.3%) in the placebo group experienced relapse by week 80.

Three other Phase III trials have been conducted with intranasal esketamine. One randomized, double-blind, placebo-controlled trial of 346 adult patients with TRD receiving standard oral antidepressants found that twice-weekly intranasal esketamine was superior at the 4-week time point for the 56 mg dose but not the 84 mg dose. The other large, randomized, double-blind, placebo-controlled trial of 138 older adult TRD patients (≥65 years) receiving standard oral antidepressants found no significant change in the primary outcome measure, change in MADRS score at Day 28. In a long-term safety and efficacy study assessing treatment-emergent adverse events in patients treated with intranasal esketamine, 90% of patients reported such events over a period of 56 weeks, including dizziness (22%), headache (19%), dissociation (19%), renal and urinary tract symptoms (17%), somnolence (14%), and viral upper respiratory tract infections (12%).

Bipolar Depression

Two initial studies of single-dose IV ketamine (0.5 mg/kg) infused over 40 minutes were conducted in adult patients with treatment-resistant bipolar I or II depression who had not responded to an adequate trial of lithium or valproate (study one, *n* = 18; study two, *n* = 15). Both studies were randomized, saline placebo-controlled, crossover, add-on trials with a therapeutic mood stabilizer. Patients were tapered off psychotropic medications other than lithium or valproate and underwent a 2-week washout period. Order of condition was randomized, with a 2-week period between each condition. The MADRS was the primary outcome measure. In both studies, depressive symptoms improved significantly in the ketamine condition compared to the placebo condition and, in both studies, over 70% of participants responded to ketamine (compared to 0% to 6% to placebo). Significant improvement was seen within 40 minutes and lasted through Day 3.

In addition, a recent systematic review examined 6 published studies of 135 individuals with bipolar depression (including the two studies discussed above) who received add-on IV ketamine (0.5 mg/kg over 40 minutes). The authors found an average response rate of 61% for IV ketamine versus 5% for placebo, and the overall response rates ranged from 52% to 80% across the six studies. Notably, two participants in the review (one receiving ketamine and one receiving placebo) developed manic symptoms. The authors concluded that there is preliminary evidence that IV ketamine administration has rapid, though short-term, effects as an adjunctive treatment to conventional mood stabilizers. Additional studies are needed to firmly establish this finding and explore longer-term outcomes.

Suicidality Associated With TRD

One systematic review and meta-analysis of 10 controlled trials (n = 167) of single-dose IV ketamine versus saline placebo or active midazolam control concluded that ketamine rapidly (within 1 day) reduced suicidal ideation with moderate to large effect sizes (Cohen's d = 0.48–0.85) up to 1 week posttreatment. Another systematic review of 25 reports from 15 independent trials (n = 572) concluded that there was moderate to low evidence that ketamine had antisuicidal effects, though the study noted significant heterogeneity in the results and the lack of data on actual suicidal behavior.

Intranasal esketamine also received an FDA indication for depressive symptoms in adults with MDD with acute suicidal ideation or behavior. Two randomized, double-blind, placebo-controlled studies of adults with moderate to severe MDD with active suicidal ideation and intent compared intranasal esketamine (56 or 84 mg) versus placebo twice weekly for 4 weeks, in addition to comprehensive standard of care. In both studies, the intranasal esketamine group demonstrated statistically greater improvement in MADRS scores than the placebo group (study one [n = 223] average difference in improvement = 3.8 points on MADRS; study two [n = 226]) average difference in improvement = 3.9 points on MADRS). Notably, neither study demonstrated the superiority of intranasal esketamine over placebo in a clinician assessment of suicide severity (Clinical Global Impression of Suicidal Severity–Revised [CGI-SS-r]).

Subtypes and Comorbidities

Investigations into the effectiveness of ketamine and esketamine for treating certain subtypes of depression, comorbid symptom constellations, or other diagnoses have only recently begun. Initial evidence suggests that some subtypes of depression (i.e., anxious, anhedonic, or atypical), as well as social anxiety, posttraumatic stress disorder (PTSD), obsessive compulsive disorder (OCD), and some types of substance use disorders may respond to ketamine infusion. However, these data should be considered preliminary and require further study and replication.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

While serious adverse effects may occur with any ketamine dose, they are generally dose-dependent, being more frequent and more severe at higher dosages and with longer durations of use. Indeed, much of the existing information regarding ketamine's effects is drawn from over 50 years of its use as an anesthetic agent, and the adverse effects reviewed in this section are thus associated with doses that are typically several times higher than those recommended for psychiatric indications. The lower ketamine doses associated with antidepressant use (e.g., 0.5 mg/kg IV infused over 40 minutes, administered twice daily over 4 to 6 weeks) are associated with less severe and less frequent adverse effects that are generally transient in nature.

One theoretical concern with ketamine (and other glutamatergic agents) is the possibility of neurodegenerative changes. Neurodegenerative changes, specifically neuronal vacuolization or "Olney lesions," have been reported in animal models administering high doses of NMDA receptor antagonists such as PCP, MK-801, or ketamine. However, a recent animal study found dose-dependent induction of "Olney lesions" for MK-801 (up to 0.8 mg/kg subcutaneous) but not for ketamine (up to 20 mg/kg IV or 60 mg/kg intraperitoneally) or its metabolite (2*R*,6*R*)-HNK (160 mg/kg IV). Olney lesions from ketamine exposure have not been documented in human studies.

Central Nervous System

Ketamine and esketamine's most common adverse effects are neuropsychiatric in nature. Dissociative symptoms, which over 70% of patients report experiencing, are the hallmark of IV ketamine administration. Dissociative symptoms include perceptual disturbances, abnormal and vivid sensory experiences, hallucinations, derealization (feeling detached from your surroundings), and depersonalization (loss of identity, feeling separate and/or outside of one's body). They may be accompanied by emotional dysregulation, sedation, agitation, confusion, and memory disturbance. If dissociative symptoms are severe, they may lead to a condition of complete disconnection and nonresponsiveness. Dissociative symptoms are generally experienced at the time of administration, are transient, and typically resolve with 2 hours postadministration. Psychotomimetic effects (nondissociative psychotic symptoms) may also emerge, particularly in patients with a history of schizophrenia.

At anesthetic doses, emergence reactions have been reported in approximately 12% of patients. Emergence reactions include confusion, severe psychomotor agitation, and amnesia that may begin shortly after administration and last 2 to 4 hours postadministration, though recurrence has been reported up to 24 hours postadministration. When emergence reactions occur, they may be treated with low-dose benzodiazepines or barbiturates.

Increased intracranial pressure has also been reported following ketamine administration. Finally, in pediatric populations, long-term cognitive deficits have been noted when ketamine is used for longer than 3 hours in children 3 years old or younger.

Cardiovascular System

At antidepressant doses, both ketamine and esketamine have been associated with transient hypertension and tachycardia. Rare reports of hypotension, bradycardia, arrhythmias, and cardiac decompensation have also been associated with ketamine administration. Thus, it is recommended that vital signs and cardiac function be monitored during ketamine and esketamine administration; both agents are contraindicated in patients with uncontrolled significant hypertension.

Gastrointestinal System

Gastrointestinal symptoms of nausea, vomiting, and anorexia are relatively common with ketamine and esketamine administration, including with antidepressant doses. If gastrointestinal symptoms are present, postponing treatment should be considered. Symptoms that develop over the course of treatment may require antinausea or antiemetic intervention. Airway management to prevent aspiration may be required if vomiting occurs during treatment.

Hepatic System

Short-term use of ketamine for anesthesia has been associated with rare instances of transient and clinically nonsignificant serum enzyme elevations. Longer-term recurrent use has been associated with hepatobiliary dysfunction (most often cholestatic pattern) with dilation and irregularity of the bile ducts. Patients may develop right upper quadrant discomfort and tenderness along with elevated serum alkaline phosphatase and aminotransferase levels and minimal or no increases in bilirubin levels. Monitoring baseline and follow-up liver function panels during treatment is recommended. If issues persist, discontinuation of ketamine administration is recommended.

Pulmonary System

Though rarely seen with antidepressant doses, risk of respiratory depression may occur with ketamine overdosage or too rapid a rate of administration. Maintaining adequate oxygenation and ventilation is recommended.

Urologic System

Long-term chronic ketamine use has been associated with inflammation and irritation of the urinary bladder and urethra. Symptoms may include dysuria, increased urinary frequency, and increased urinary urgency, urge incontinence, and hematuria. Diagnostic studies have identified cystitis, hydronephrosis, and reduced bladder capacity.

Immunologic Effects

Anaphylactic reactions are rare but have been reported with ketamine and esketamine administration.

Pregnancy

No controlled studies have examined the safety of ketamine use in human pregnancy. In rats and nonhuman primates, ketamine use during pregnancy was associated with altered neurodevelopment and neurodegeneration. In rat studies, esketamine similarly demonstrated adverse effects in offspring. Animal studies also suggest that ketamine is excreted in breast milk. Ketamine and esketamine have not been assigned a pregnancy risk category by the FDA, and their use is not recommended during pregnancy or breastfeeding.

Therapeutic Implications

IV ketamine is FDA-approved as a sole anesthetic agent for diagnostic and surgical procedures that do not require skeletal muscle relaxation, for the induction of anesthesia prior to the administration of other general anesthetic agents, and as a supplement to other anesthetic agents.

Intranasal esketamine is FDA-approved in adults to treat TRD or MDD with acute suicidal ideation or behavior.

Precautions and Adverse Reactions

Adverse effects associated with ketamine and esketamine are largely dose-dependent. Given that doses for psychiatric indications are typically several times lower than anesthetic doses, the adverse effects seen with anesthetic or illicit use of ketamine are greatly reduced. Common adverse events associated with anesthetic ketamine include emergence reactions and elevated blood pressure and pulse. The primary concerns with ketamine and esketamine administered at doses suggested for psychiatric indications include: dissociative symptoms, confusion, memory dysfunction, hypertension, tachycardia, emergence reactions, and risk of abuse and dependence. The most common adverse events at doses suggested for TRD include: dissociation, visual distortions, difficulty speaking, numbness, confusion, dizziness, faintness, drowsiness/sedation, euphoria, blurred vision, and hypertension. For intranasal esketamine, the most common adverse events include dissociation, dizziness, nausea, sedation, vertigo, hypoesthesia, anxiety, lethargy, increased blood pressure, vomiting, and feeling drunk. Common and other potentially serious adverse events are listed in Table 33.14a–1.

Table 33.14a–1.
Adverse Events Associated With Ketamine and Esketamine

Drug	Common Adverse Events at Antidepressant Doses	Other Potential Serious Adverse Events (including at higher doses)
Ketamine	Dissociation, visual distortions, difficulty speaking, numbness, confusion, dizziness, faintness, drowsiness/sedation, euphoria, blurred vision, and hypertension.	Hemodynamic instability, emergence reactions, risk of respiratory depression, risk of use of ketamine alone for procedures of the pharynx, larynx, or bronchial tree (as pharyngeal and laryngeal reflexes are not suppressed by ketamine), and pediatric neurotoxicity.[a]
Esketamine	Dissociation, dizziness, nausea, sedation, vertigo, hypoesthesia, anxiety, lethargy, increased blood pressure, vomiting, and feeling drunk.	Risk of sedation and dissociation, potential for abuse and misuse, availability only through REMS, and increased risk of suicidal thoughts and behaviors in pediatric and young adult patients taking antidepressants.

[a] Rates and type of side effects vary depending on dose and duration of administration. These serious adverse events are rare at doses and durations used for psychiatric indications.

The FDA label for ketamine (approved for anesthetic indications) has warnings and precautions for hemodynamic instability, emergence reactions, risk of respiratory depression, risk of use of ketamine alone for procedures of the pharynx, larynx, or bronchial tree (because pharyngeal and laryngeal reflexes are not suppressed by ketamine), and pediatric neurotoxicity. It is contraindicated in patients for whom a significant elevation of blood pressure would be a serious hazard and for those with known hypersensitivity to ketamine or to any excipient.

Esketamine (approved for psychiatric indications) is contraindicated in aneurysmal vascular disease (including thoracic and abdominal aorta, intracranial, and peripheral arterial vessels) or arteriovenous malformation, intracerebral hemorrhage, and hypersensitivity to esketamine, ketamine, or any of their excipients. Esketamine has a boxed warning for risk of sedation and dissociation, potential for abuse and misuse, availability only through REMS, and increased risk of suicidal thoughts and behaviors in pediatric and young adult patients taking antidepressants. The risk/benefit ratio for patients at higher risk of abuse should be considered prior to prescribing, and patients should be monitored for signs and symptoms of abuse and misuse as well as for clinical worsening and emergence of suicidal thoughts and behaviors. Additional warnings and precautions for esketamine include increased blood pressure, cognitive impairment, impaired ability to drive and operate machinery, and embryo–fetal toxicity.

Ketamine is classified as a Schedule III controlled substance under the Controlled Substance Act. According to the Surgeon General’s Report on Alcohol, Drugs, and Health, recreational misuse of ketamine has been increasing. In 2015, an estimated 3 million individuals (1.1%) aged 12 or older had tried ketamine in their lifetime. Ketamine, also known as K, Special K, Vitamin K, or Cat Valium, can be injected, snorted, smoked (added to tobacco or marijuana cigarettes), or ingested to induce a feeling of detachment and altered perception and consciousness. In abuse, ketamine is typically administered in higher doses than those recommended for psychiatric indications, often approaching anesthetic doses of 50 to 100 mg. These higher doses may generate a state of noncommunicativeness (termed the “K-hole”) that is experienced as extreme disconnectedness, lack of awareness of self and surroundings, and unresponsiveness to external stimuli; accidental overdoses may also occur. Repeated use of ketamine may result in the development of tolerance and a dependency syndrome, and psychological withdrawal symptoms may occur when use is curtailed.

For both ketamine and esketamine, side effects may be monitored during administration using a questionnaire such as the Ketamine Side Effect Tool (KSET).

DRUG INTERACTIONS

Ketamine and esketamine interact with many other drugs. Both undergo hepatic metabolism via CYP 3A4 and 2B6, and inhibitors of these isoenzymes may increase ketamine and esketamine drug levels while inducers may decrease levels. Both ketamine and esketamine also act as CPY 3A4 inducers and may decrease levels of drugs metabolized by this route. It is recommended that interactions be checked before using ketamine or esketamine in combination with other drugs.

Notable interactions include:

- Benzodiazepines and other GABA agonists: Administration of concomitant benzodiazepine and other GABA agonist drugs (such as barbiturates) may attenuate or delay antidepressant response.
- Lamotrigine: Concomitant lamotrigine administration may also attenuate antidepressant response, given that it decreases glutamate release.
- Sympathomimetics and vasopressin: Vital signs should be closely monitored when these agents are coadministered with ketamine. Consider dose adjustment individualized to the patient's clinical situation.
- Opioid analgesics or other central nervous system (CNS) depressants: Concomitant use of ketamine with opioid analgesics, benzodiazepines, or other CNS depressants, including alcohol, may result in profound sedation, respiratory depression, coma, and death. Concomitant use of opioid analgesics may prolong recovery time.
- Monoamine oxidase inhibitors (MAOIs): Given the risk of hypertension associated with MAOIs, coadministration of these agents with ketamine should be done with caution and close monitoring of vital signs.
- Theophylline or aminophylline: Concomitant administration of ketamine and theophylline or aminophylline may lower the seizure threshold.

Table 33.14a–2.
Dosing of Ketamine and Esketamine

Drug	Indication	Dose
Ketamine	Anesthesia induction	IV: 1–4.5 mg/kg slow IV, single dose IM: 6.5–13 mg/kg IM, single dose
	Anesthesia maintenance	50% of IV ketamine induction dose PRN, or 0.1–0.5 mg/minute IV continuous infusion
	TRD (off label)	IV: 0.5 mg/kg over 40 minutes up to twice weekly; not to exceed 6 weeks
Esketamine	TRD	Intranasal
	Induction	Weeks 1–4: 56 mg (starting dose), may titrate to 84 mg twice weekly
	Maintenance	Weeks 5 to 8: 56 or 84 mg once weekly Week 9 and thereafter: 56 or 84 mg once weekly or once every 2 weeks

TRD, treatment-resistant major depressive disorder; IV, intravenous; IM, intramuscular; PRN, as needed.

DOSAGE AND ADMINISTRATION

Typical doses for the given indications are presented in Table 33.14a–2. For psychiatric indications, the IV ketamine dose is 0.5 mg/kg, given over a 40-minute infusion. Other dosages and routes of administration have not been adequately studied.

The intranasal esketamine dose is either 56 or 84 mg. In the induction phase (weeks 1 to 4), dosing begins twice weekly with a starting dose of

56 mg. After the first administration, the dose may be titrated up to 84 mg as clinically indicated. In the maintenance phase, dose administration is reduced to once weekly for 4 weeks and then continues at once weekly or decreases to once every other week. As noted above, due to concerns about sedation, dissociation, and abuse potential, intranasal esketamine is currently being administered only under an FDA-required REMS program. The drug must only be distributed to REMS certified hospitals and clinical settings and administered by patients under the direct observation of a health care provider who must observe the administration, and the patient must be monitored for at least 2 hours postadministration. Intranasal esketamine may impair the ability to drive or operate machinery; patients should not engage in these activities on the day of administration or the following day.

FUNDING AND DISCLOSURES

Funding for this work was provided by the Intramural Research Program at the National Institute of Mental Health, National Institutes of Health (IRP-NIMH-NIH; ZIAMH002857). Dr. Zarate is listed as a co-inventor on a patent for the use of ketamine in major depression and suicidal ideation; as a coinventor on a patent for the use of (2*R*,6*R*)-HNK, (*S*)-dehydronorketamine (DHNK), and other stereoisomeric dehydroxylated and hydroxylated metabolites of (*R,S*)-ketamine metabolites in the treatment of depression and neuropathic pain; and as a coinventor on a patent application for the use of (2*R*,6*R*)-HNK and (2*S*,6*S*)-HNK in the treatment of depression, anxiety, anhedonia, suicidal ideation, and PTSD. He has assigned his patent rights to the U.S. government but will share a percentage of any royalties that may be received by the government. Dr. Park has no conflict of interest to disclose, financial or otherwise.

ACKNOWLEDGMENTS

The authors thank the 7SE research unit and staff for their support. Ioline Henter (NIMH) provided invaluable editorial assistance.

Further Readings

Acevedo-Diaz EE, Cavanaugh GW, Greenstein D, et al. Comprehensive assessment of side effects associated with a single dose of ketamine in treatment-resistant depression. *J Affect Disord*. 2020;263:568–575.

Artigas F. Brain circuitry in major depressive disorder: the critical role of ventral anterior cingulate cortex. *Eur Neuropsychopharmacol*. 2021;51:134–137.

Bahji A, Zarate CAJ. Ketamine for bipolar depression: a systematic review. *Intl J Neuropsychopharmacol*. 2021;24:535–541.

Berman RM, Cappiello A, Anand A, et al. Antidepressant effects of ketamine in depressed patients. *Biol Psychiatry*. 2000;47(4):351–354.

Castrén E, Monteggia LM. Brain-derived neurotrophic factor signaling in depression and antidepressant action. *Biol Psychiatry*. 2021;90:128–136.

Corriger A, Pickering G. Ketamine and depression: a narrative review. *Drug Des Devel Ther*. 2019;13:3051–3067.

Daly EJ, Trivedi MH, Janik A, et al. Efficacy of esketamine nasal spray plus oral antidepressant treatment for relapse prevention in patients with treatment-resistant depression: a randomized clinical trial. *JAMA Psychiatry*. 2019;76:893–903.

Diazgranados N, Ibrahim L, Brutsche NE, et al. A randomized add-on trial of an N-methyl-D-aspartate antagonist in treatment-resistant bipolar depression. *Arch Gen Psychiatry*. 2010;67:793–802.

Drugbank Online. Ketamine. 2021. Accessed https://go.drugbank.com/drugs/DB01221

Duman RS, Aghajanian GK, Sanacora G, Krystal JH. Synaptic plasticity and depression: new insights from stress and rapid-acting antidepressants. *Nat Med*. 2016;22:238–249.

Feder A, Parides MK, Murrough JW, et al. Efficacy of intravenous ketamine for treatment of chronic posttraumatic stress disorder: a randomized clinical trial. *JAMA Psychiatry*. 2014;71(6):681–688.

Fedgchin M, Trivedi M, Daly EJ, et al. Efficacy and safety of fixed-dose esketamine nasal spray combined with a new oral antidepressant in treatment-resistant depression: results of a randomized, double-blind, active-controlled study (TRANSFORM-1). *Int J Neuropsychopharmacol*. 2019;22:616–630.

Highland JN, Zanos P, Riggs LM, et al. Hydroxynorketamines: pharmacology and potential therapeutic applications. *Pharmacol Rev*. 2021;73:763–791.

Ionescu DF, Luckenbaugh DA, Niciu MJ, et al. Effect of baseline anxious depression on initial and sustained antidepressant response to ketamine. *J Clin Psychiatry*. 2014;75:e932–e938.

Jones JL, Mateus CF, Malcolm RJ, Brady KT, Back SE. Efficacy of ketamine in the treatment of substance use disorders: a systematic review. *Front Psychiatry*. 2018;9:277.

Lally N, Nugent AC, Luckenbaugh DA, Niciu MJ, Roiser JP, Zarate CA Jr. Neural correlates of change in major depressive disorder anhedonia following open-label ketamine. *J Psychopharmacol*. 2015;29:596–607.

Lener MS, Niciu MJ, Ballard ED, et al. Glutamate and gamma-aminobutyric acid systems in the pathophysiology of major depression and antidepressant response to ketamine. *Biol Psychiatry*. 2017;81:886–897.

Lin P, Ma ZZ, Mahgoub M, Kavalali ET, Monteggia LM. A synaptic locus for TrkB signaling underlying ketamine rapid antidepressant action. *Cell Rep*. 2021;36:109513.

Morris PJ, Burke RD, Sharma AK, et al. A comparison of the pharmacokinetics and NMDAR antagonism-associated neurotoxicity of ketamine, (2R,6R)-hydroxynorketamine and MK-801. *Neurotoxicol Teratol*. 2021;87:106993.

Murrough JW, Perez AM, Pillemer S, et al. Rapid and longer-term antidepressant effects of repeated ketamine infusions in treatment-resistant major depression. *Biol Psychiatry*. 2013;74:250–256.

Olney JW, Labruyere J, Price MT. Pathological changes induced in cerebrocortical neurons by phencyclidine and related drugs. *Science*. 1989;244:1360–1362.

Park L, Luckenbaugh DA, Pennybaker SJ, et al. The effects of ketamine on typical and atypical depressive symptoms. *Acta Psychiatr Scand*. 2020;142:394–401.

Popova V, Daly EJ, Trivedi M. et al. Efficacy and safety of flexibly dosed esketamine nasal spray combined with a newly initiated oral antidepressant in treatment-resistant depression: a randomized double-blind active-controlled study. *Am J Psychiatry*. 2019;176:428–438.

Singh JB, Fedgchin M, Daly E, et al. Intravenous esketamine in adult treatment-resistant depression: a double-blind, double-randomization, placebo-controlled study. *Biol Psychiatry*. 2016;80:424–431.

Swainson J, McGirr A, Blier P. et al. The Canadian Network for Mood and Anxiety Treatments (CANMAT) task force recommendations for the use of racemic ketamine in adults with major depressive disorder: Recommandations du groupe de travail du réseau Canadien pour les traitements de l’humeur et de l’anxiété (Canmat) concernant l’utilisation de la kétamine racémique chez les adultes souffrant de trouble dépressif majeur. *Can J Psychiatry*. 2021;66:113–125.

Trullas R, Skolnick P. Functional antagonists at the NMDA receptor complex exhibit antidepressant actions. *Eur J Pharmacol*. 1990;185:1–10.

Wilkinson ST, Ballard ED, Bloch MH, et al. The effect of a single dose of intravenous ketamine on suicidal ideation: a systematic review and individual participant data meta-analysis. *Am J Psychiatry*. 2018;175:150–158.

Williams NR, Heifets BD, Blasey C, et al. Attenuation of antidepressant effects of ketamine by opioid receptor antagonism. *Am J Psychiatry*. 2018;175:1205–1215.

Wulf HA, Browne CA, Zarate CA, Lucki I. Mediation of the behavioral effects of ketamine and (2R,6R)-hydroxynorketamine in mice by kappa opioid receptors. *Psychopharmacology (Berl)*. 2022;239:2309–2316.

Zanos P, Moaddel R, Morris PJ, et al. Ketamine and ketamine metabolite pharmacology: insights into therapeutic mechanisms. *Pharmacol Rev*. 2018;70:621–660.

Zanos P, Moaddel R, Morris PJ, et al. NMDAR inhibition-independent antidepressant actions of ketamine metabolites. *Nature*. 2016;533:481–486.

Zarate CA Jr, Singh JB, Carlson PJ, et al. A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. *Arch Gen Psychiatry*. 2006;63:856–864.

33.14b MDMA, Psilocybin, Ayahuasca, and Ibogaine

DAVID NUTT, F.R.C.P., F.R.C.PSYCH., F.B.PH.S., F.MED.SCI., D.LAWS., AND DAVID ERRITZOE, M.D., PH.D., M.R.C.PSYCH.

In this chapter, we cover the resurgence of clinical research in the use of serotonergic psychedelic drugs such as psilocybin and the synthetic entactogen MDMA (methylenedioxymethamphetamine). These are currently classified as “illegal” under the UN conventions on psychotropic drugs on the grounds that they are harmful and have no special medical value. These assertions are not evidence-based, as psilocybin, LSD, and MDMA all had significant evidence of medical benefit before being made illegal. The ban has very significantly impeded research for several decades, but a few centers have begun to revisit clinical research with these compounds. This research is underpinned by neuroimaging studies that have revealed their unique modes of action on the brain. This chapter focuses on the new clinical data with some reference to the earlier research and relevant preclinical and neuroimaging studies of mechanisms. Note: Ketamine can have psychedelic properties, but its pharmacologic target is glutamate and is covered in a separate chapter.

BACKGROUND AND EARLY STUDIES PREBANNING

Psychedelics

Most psychedelics are agonists of the 5-HT2A receptor and many are structural analogues of serotonin (see Fig. 33.14b–1). They include plant products such as ergot, mescaline (peyote cactus), DMT (ayahuasca) and ibogaine, plus psilocybin (magic mushrooms), and some of animal origin, for example, 5-MEO-DMT. They have been used for millennia in different cultures across the globe. LSD a derivative of ergot was made in 1943 by Albert Hofmann at Sandoz. Sandoz made LSD available to researchers across the world under the trade name *Delysid*. Research with LSD (and to a lesser extent psilocybin) flourished with the U.S. NIH funding over 130 grants for clinical LSD studies (though none since the 1967 ban). This research led to hundreds of papers. LSD trips were used to model psychosis as they mimicked some of the symptoms of acute psychosis, particularly ego-dissolution thought disorder and misperceptions. Lower doses were used to “unblock” psychological resistance in psychotherapy. One interesting area was in the treatment of alcohol use disorder. Bill Wilson, the founder of Alcoholics Anonymous, reported he “escaped” from the chains of his addiction as a result of a trip. Wilson’s own experience led him to promote the use of LSD in the treatment of alcohol use disorder, and six trials were conducted before the drug was banned. A modern meta-analysis of these data by Krebs and Johansen found this to be more efficacious than any current therapies.

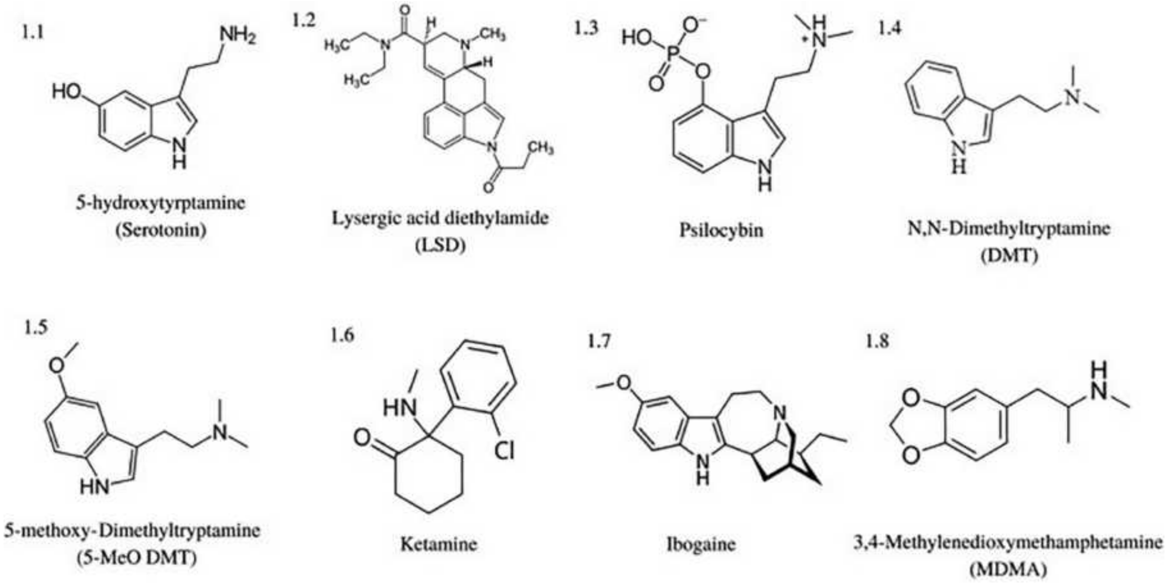

1.1 5-hydroxytyrptamine (Serotonin)

1.2 Lysergic acid diethylamide (LSD)

1.3 Psilocybin

1.4 N,N-Dimethyltryptamine (DMT)

1.5 5-methoxy-Dimethyltryptamine (5-MeO DMT)

1.6 Ketamine

1.7 Ibogaine

1.8 3,4-Methylenedioxymethamphetamine (MDMA)

FIGURE 33.14b–1. Chemical structures.

For political reasons, rather than real evidence of harm, in 1967, the U.S. government banned LSD and related hallucinogens question ban consolidated internationally under the 1971 UN convention on narcotic drugs. They were all put into Schedule 1, which is for drugs defined as having no accepted special medical use plus significant potential for harm and dependence. This scheduling effectively stopped research on psychedelics for over 40 years, and only in the past decade have attempts been made to reverse this.

Methylenedioxymethamphetamine

MDMA is a synthetic molecule related to methamphetamine that in the 1960s was synthesized and tested by Alexander Shulgin, who reported it to have unique proempathy (entactogenic) properties. MDMA was widely used on the west coast of the United States by psychotherapists, especially in couples counseling, where its entactogenic effects allowed couples to regain the affection that was present when they fell in love and so helped them rebuild their failing relationships. At that time, it was called empathy and was legal, whereas other amphetamine-like drugs were banned. Because it was legal, MDMA became sold for recreational purposes, especially in clubs and raves where its name was changed to ecstasy. This led to a media and political backlash that resulted in MDMA being banned, and treatment stopped.

THE TWENTY-FIRST–CENTURY RESURRECTION OF PSYCHEDELICS AND MDMA IN PSYCHIATRY

Over the past two decades, research on psychedelics has resumed due to persistence from a small group of researchers. Two themes drove this. One was the discovery, using neuroimaging techniques, of their very novel effects on the brain. The other was from some small exploratory clinical studies. The neuroimaging results with both fMRI and MEG were unexpected and surprising, revealing a reduction in brain activity in many regions but particularly the default mode network (DMN), including in the subgenual cingulate cortex (SGC). Many different treatments for depression have been found to reduce SGC activity, and other imaging studies have found the DMN overactive in depression.

PSILOCYBIN TRIALS

The first modern clinical trial of a psychedelic was by Moreno and colleagues, who gave nine subjects with treatment-resistant OCD up to four different doses of psilocybin in an open-label design (Table 33.14b–1). Then Grob and colleagues gave 12 subjects with advanced cancer diagnoses that resulted in DSM-IV-defined acute stress or anxiety disorders a moderate (0.2 mg/kg) dose of psilocybin and an active placebo (niacin 250 mg) in a double-blind within-subject design. The Johnson/Griffiths team at Johns Hopkins has subsequently conducted several clinical trials. The first was an open study on 15 people trying to quit tobacco smoking. They were given psilocybin 25 mg three times during a 20-week tobacco quitting course. Outcomes were assessed by urinary cotinine level plus smoking craving and urge for up to 6 months. Twelve and 10 of the 15 subjects became and remained smoking-free at 6 and 12 months, respectively, a very significant outcome.

The other addiction study with psilocybin was conducted by Bogenschutz and colleagues at the University of New Mexico, who gave two psilocybin sessions to 10 alcohol-dependent patients in addition to a standard psychotherapy course. They found a marked reduction in the percentage of heavy drinking days, and the psilocybin was well-tolerated.

Table 33.14b–1.
Psilocybin Studies

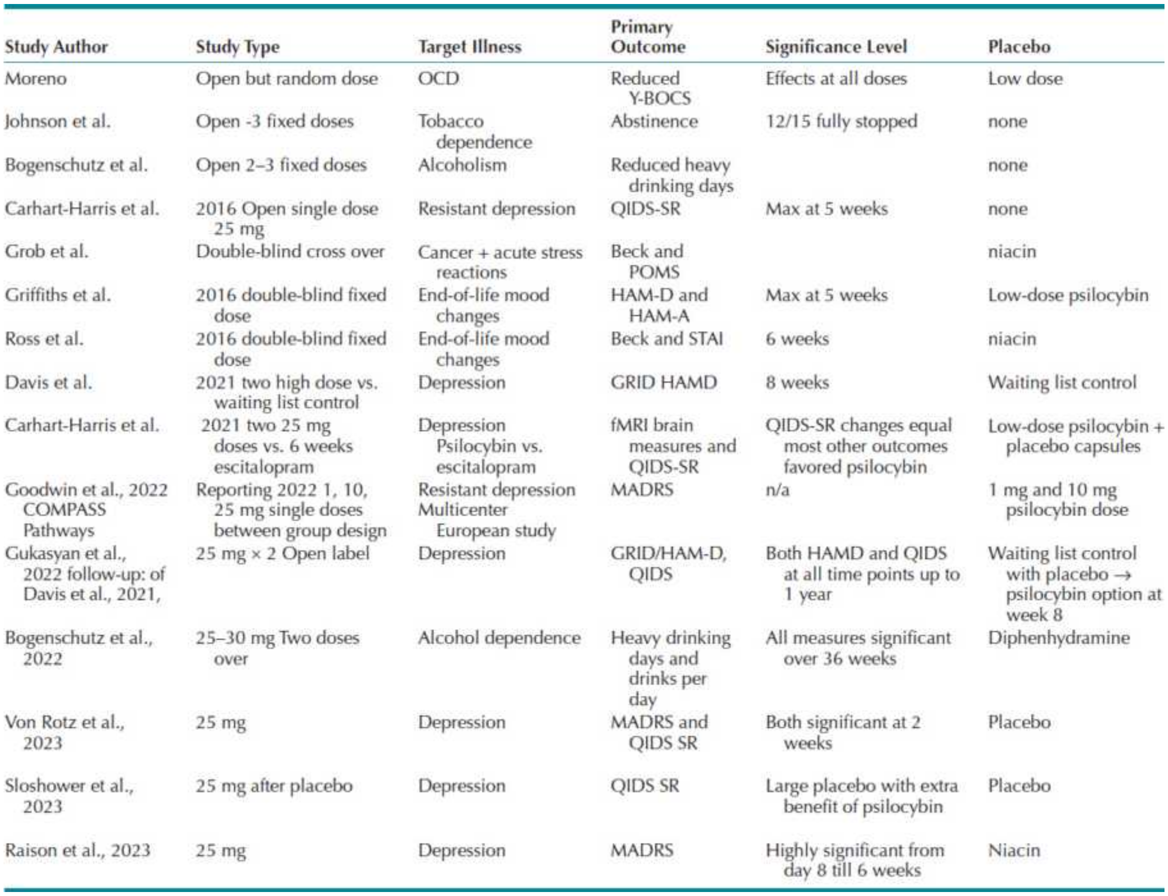

Study Author	Study Type	Target Illness	Primary Outcome	Significance Level	Placebo
Moreno	Open but random dose	OCD	Reduced Y-BOCS	Effects at all doses	Low dose
Johnson et al.	Open -3 fixed doses	Tobacco dependence	Abstinence	12/15 fully stopped	none
Bogenschutz et al.	Open 2–3 fixed doses	Alcoholism	Reduced heavy drinking days		none
Carhart-Harris et al.	2016 Open single dose 25 mg	Resistant depression	QIDS-SR	Max at 5 weeks	none
Grob et al.	Double-blind cross over	Cancer + acute stress reactions	Beck and POMS		niacin
Griffiths et al.	2016 double-blind fixed dose	End-of-life mood changes	HAM-D and HAM-A	Max at 5 weeks	Low-dose psilocybin
Ross et al.	2016 double-blind fixed dose	End-of-life mood changes	Beck and STAI	6 weeks	niacin
Davis et al.	2021 two high dose vs. waiting list control	Depression	GRID HAMD	8 weeks	Waiting list control
Carhart-Harris et al.	2021 two 25 mg doses vs. 6 weeks escitalopram	Depression Psilocybin vs. escitalopram	fMRI brain measures and QIDS-SR	QIDS-SR changes equal most other outcomes favored psilocybin	Low-dose psilocybin + placebo capsules
Goodwin et al., 2022 COMPASS Pathways	Reporting 2022 1, 10, 25 mg single doses between group design	Resistant depression Multicenter European study	MADRS	n/a	1 mg and 10 mg psilocybin dose
Gukasyan et al., 2022 follow-up: of Davis et al., 2021,	25 mg × 2 Open label	Depression	GRID/HAM-D, QIDS	Both HAMD and QIDS at all time points up to 1 year	Waiting list control with placebo → psilocybin option at week 8
Bogenschutz et al., 2022	25–30 mg Two doses over	Alcohol dependence	Heavy drinking days and drinks per day	All measures significant over 36 weeks	Diphenhydramine
Von Rotz et al., 2023	25 mg	Depression	MADRS and QIDS SR	Both significant at 2 weeks	Placebo
Sloshower et al., 2023	25 mg after placebo	Depression	QIDS SR	Large placebo with extra benefit of psilocybin	Placebo
Raison et al., 2023	25 mg	Depression	MADRS	Highly significant from day 8 till 6 weeks	Niacin

Two studies of psilocybin in major depressive disorder have been conducted by the Imperial College group of Nutt and Carhart-Harris. In the first, an open study, patients with resistant depression were given two doses of psilocybin (a 10 mg “test” dose and a 25 mg therapeutic dose) 1 week apart with psychological support before and after the experience. In practice, participants were usually multiply-treatment resistant (mean number of failed medications was 4.6 +/− 2.6), and the illness was very long-standing (mean duration of illness was 17.7 years +/− 8.4 years). The primary outcome measure was the mean change in the participant-rated quick inventory of depressive symptoms (QIDS-SR) rating scale with follow-up for 6 months. Highly significant improvements in depression ratings were seen at all time points, with the maximal effect at 5 weeks (Cohen’s *d* = 2.3). The trial established the feasibility and initial evidence of safety, but efficacy interpretations are limited by the open-label design.

Subsequently, they conducted a comparative trial of two 25 mg doses of psilocybin 3 weeks apart versus the SSRI escitalopram (10 mg for 3 weeks, then 20 mg for a further 3 weeks). The escitalopram group was also given two 1 mg (placebo equivalent) treatments of psilocybin with psychological support, so both groups had the same level of psychotherapy. The primary outcomes were mood and well-being changes from baseline at 6 weeks. Well-being was significantly improved by both treatments but more with the high-dose psilocybin. Depression scores also improved with both treatments, and though the primary measure—change in QIDS SR—did not differentiate the groups, all other mood measures favored psilocybin, as did remission rates.

fMRI imaging of both groups revealed fundamental differences in the brain effects of psilocybin compared with escitalopram (Daws et al., 2022).

Several further studies have been conducted—see Table 33.14b–1. Of particular importance is the Goodwin et al.'s 2022 study as it was the first by a pharmaceutical company. This was a large multicenter, multicountry RCT in patients with treatment-resistant depression in which three doses (1, 10, 25 mg) of psilocybin were used. The 25-mg dose produced remission rates of 25% at 3 months—so replicating the Carhart-Harris studies. The lower (nontrip) doses had very little therapeutic impact. Based on this, a phase 3 study of 25 mg in treatment-resistant depression has been started. In 2023, Raison et al. reported a major U.S. multicenter study in which a single 25-mg psilocybin dose produced profound and enduring reductions in depression.

An important consequence of these positive trials is that in July 2023 the Australian regulators approved the use of psilocybin on a compassionate basis for patients with treatment-resistant depression.

Finally, three separate controlled studies have been published on the use of psilocybin for anxiety and depression associated with life-threatening illnesses. Griffiths and his Johns Hopkins team treated 51 patients with anxiety and depressive symptoms occasioned by a life-threatening cancer diagnosis. Placebo was a very low (ineffective) dose of psilocybin (1 mg or 3 mg/70 kg) compared with the much higher active treatment dose (22 mg or 30 mg/70 kg). The administration was counterbalanced with 5 weeks between sessions and a 6-month follow-up. The Hamilton Depression and Anxiety scales were the primary outcome measures and showed superiority for the high dose versus the low dose at 5 weeks. The placebo group was then crossed over to a high dose, and these experienced significant improvements also.

In a similarly designed study, Ross and colleagues at NYU gave 29 patients a single dose of 0.3 mg/kg psilocybin or 250 mg niacin as an active placebo with crossover at 7 weeks. The treatment was delivered safely, with no reports of serious adverse events. The group receiving psilocybin showed clinical benefits as measured by both clinician and participant-rated scales that lasted for 7 weeks prior to crossover and were also sustained approximately 8 months after dosing. The group that received niacin showed transient reductions that were not sustained at 7 weeks. After they crossed over to psilocybin, they also showed immediate and sustained reductions in anxiety and depression scores that were sustained at 6 months.

Additionally, Gasser et al. in Switzerland performed a trial with LSD, either a low (20 ug) or a high (200 ug) dose in depression/trauma patients, and these data led to this drug being made available for compassionate use by a few psychiatrists in that country.

Other controlled trials of psilocybin in OCD anorexia and addictions are underway.

DMT AND AYAHUASCA

A number of countries in South and Latin America allow the use of ayahuasca drink for religious and therapeutic purposes. This drink contains a mixture of plant products, one of which contains the psychedelic dimethyl tryptamine (DMT) and the other a monoamine oxidase inhibitor harmaline. The harmaline blocks the activity of MAO in the gut and liver and so allows the DMT in the drink to get to the brain. Human neuroimaging studies of ayahuasca have shown it to have similar effects to psilocybin and LSD, and several open clinical trials have been conducted in depressed patients. The first RCT was conducted by Palhano-Fontes, who found rapid and highly significant reductions in depressive symptoms. However, moving ayahuasca into Western medicine is challenging because it is a plant mixture of two active ingredients, so concentrations of these are hard to control, plus the need for coadministration of an MAOI is problematic for some patients. To overcome these issues, several pharmaceutical companies are now developing infusions of synthetic DMT as a potential antidepressant and antiaddiction treatment. Also, the related compound 5-MEO-DMT, a very potent short-acting psychedelic, is being developed for clinical testing.

IBOGAINE

Ibogaine has developed in a number of countries as an off-license treatment for addictions, especially heroin though there are no high-quality studies. Some deaths have been reported probably due to QTc prolongation. This cardiac effect has limited research, and its effects on the brain are not yet characterized. An active metabolite noribogaine is being developed as a possibly safer alternative.

METHYLENEDIOXYMETHAMPHETAMINE

Though MDMA isn't a psychedelic in the true sense of the word, it does have profound effects that allow people insights into their psychiatric problems and has proved especially helpful in the treatment of PTSD. Several studies run by the Mithoefers for MAPS (Multidisciplinary Association for Psychedelic Studies) have demonstrated that a couple of sessions of exposure therapy under MDMA (usually 125 mg + a 62.5 mg booster after 2 hours) can have powerful therapeutic effects (Table 33.14b–2). Based on these positive findings, both the FDA and EMA gave their approval for multicenter studies in the United States and Europe and the first one of these was reported in 2021 by Mitchell et al. and the second in a broader patient group including some less severely ill patients by Mitchell et al. (2023). Both reported very positive clinical outcomes and good safety so it seems likely that MDMA will also become an approved medicine sometime in the 2020s. As already mentioned with psilocybin, these positive trials with MDMA led to the Australian regulators in July 2023 approving the use of MDMA on a compassionate basis for patients with treatment-resistant PTSD.

Table 33.14b–2.
MDMA Trials

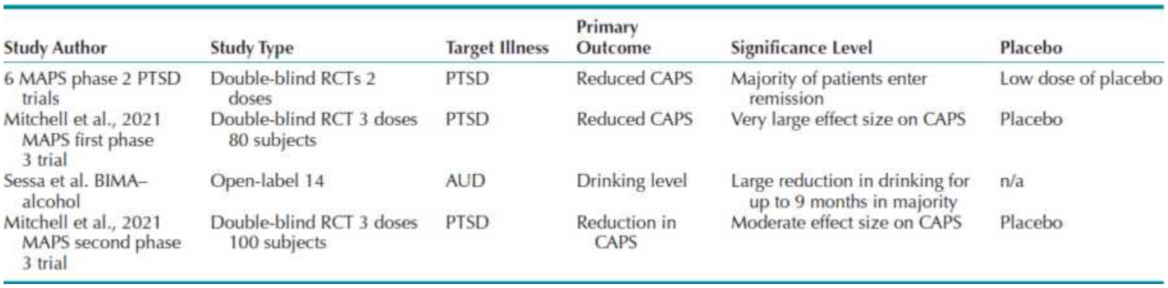

Study Author	Study Type	Target Illness	Primary Outcome	Significance Level	Placebo
6 MAPS phase 2 PTSD trials	Double-blind RCTs 2 doses	PTSD	Reduced CAPS	Majority of patients enter remission	Low dose of placebo
Mitchell et al., 2021 MAPS first phase 3 trial	Double-blind RCT 3 doses 80 subjects	PTSD	Reduced CAPS	Very large effect size on CAPS	Placebo
Sessa et al. BIMA–alcohol	Open-label 14	AUD	Drinking level	Large reduction in drinking for up to 9 months in majority	n/a
Mitchell et al., 2021 MAPS second phase 3 trial	Double-blind RCT 3 doses 100 subjects	PTSD	Reduction in CAPS	Moderate effect size on CAPS	Placebo

Key Differences Between Psychedelic/MDMA Therapy and Conventional Treatments in Psychiatry

Treatment with psychedelics and MDMA is more complex than conventional antidepressant therapies as it requires psychotherapy as well as drug administration. It is, therefore, best considered a new form of psychiatric treatment—drug-assisted psychotherapy (or psychotherapy-assisted drug treatment). Patients have to be properly prepared for the powerful psychological impact of the psychedelic/MDMA session, and this involves at least one dedicated prep session with a trained therapist (often called a guide) before the drug session. In this prep session, the patient is educated about the rationale, purpose, and procedure of the drug administration session and is reassured that they will be looked after throughout it.

The drug session itself is usually given in an environment that optimizes patient comfort and security—often a pleasantly decorated room (rather than a traditional medical clinic) with soft ambient lighting and a comforting soundtrack (as music may contribute to the therapeutic value as well). Ideally, two, though at least one, therapists (often called guides) are present in the room (optimally one male and one female) who are there to provide reassurance, medical cover, and care. They only talk with the patient if the patient wants to, which they generally do not. It is important to note that there is no expectation of conversation during the “trip” and no direction by either therapist of the patient’s speech or thought. It is the next day in the “integration” session that the contents and experiences of the trip are discussed with the same guides(s), and psychotherapeutic insights are developed and promoted. Often additional integrations visits (typically 2 to 4 in total) are part of the study design, and in addition, different psychotherapeutic methods, namely Acceptance and

Commitment Therapy (ACT), are being tested as an integrated part of the psychedelic-assisted therapy program. An important difference between psychedelic- and MDMA-assisted therapy is that therapists/guides are more actively engaged during a therapeutic session with MDMA than with psychedelics. Whereas therapeutic material, such as trauma, is discussed actively during the presence of MDMA, the approach with psychedelics is more “hands off” during the exposure period to the drug (i.e., during the trip). In psychedelic therapy sessions, patients typically spend most of the time wearing eyeshades with music in their ears and with the guide(s) providing support to ensure that the patients feel safe during the experience rather than engaging in or exploring therapeutic content. With psychedelics, the active involvement of the therapist occurs before and after the actual experience in the preparation and integration of the sessions.

For some patients, the impact of just one or two dosing sessions can be very long-lasting—allowing the concept of a “cure” to be considered. For others, the beneficial effects though powerful, wane over time probably as the underlying disorder has become so deeply entrenched in the brain. Top-up treatments are now being explored for patients who relapse, but the still-illegal status of these drugs in most countries makes their use outside of clinical trials impossible.

What about safety? Although the banning of psychedelics and MDMA was based on largely fictitious claims of harm, there is no doubt that psychedelics, in particular, are powerful mind-altering drugs that can leave deep negative as well as positive memories. There is also the risk of precipitating or worsening psychosis, and so they are to be avoided in patients with a history of psychosis in themselves (or in first-degree family members) and most studies excluded patients with bipolar depression in case mania might be precipitated. With these precautions, modern psychedelic research has confirmed the data from the 1950 and 1960s, which found adverse effects to be rare. Note that a bad trip can be terminated with a benzodiazepine given either IM or IV or with a serotonin 5-HT2A receptor blocker (e.g., risperidone or olanzapine).

One common clinical challenge is the fact that so many patients are already on antidepressants and other drugs that block the effects of psychedelics. 5-HT2A receptor-blocking antipsychotic drugs such as quetiapine and olanzapine, as well as some antidepressants such as mirtazapine, block the access of the psilocybin to the 5-HT2A receptor, so it cannot work; the effects of MDMA are also attenuated. Also, SSRIs desensitize these receptors and so reduce the psychedelic effects. To optimize therapeutic effects, it’s best to stop these, but this has to be done with care to avoid the risk of worsening the underlying depression and producing withdrawal or discontinuation reactions.

HOW DO THESE DRUGS WORK?

The current thinking about psychedelics and depression is that they disrupt the dysregulated brain circuits that underpin depression, PTSD, and addictions. In the introduction, we already mentioned the switching down of activity in the SGC by psilocybin, in common with other treatments for depression. Another interesting idea is that psilocybin reduces brain activity—and communication between cortical mid-line regions—within the DMN. Several neuroimaging studies have demonstrated that a much greater amount of the brain is engaged in the DMN in depressed people than in healthy controls, likely a manifestation of the intense self-deprecatory rumination that these patients experience. Psychedelics are thought to completely disrupt this DMN overactivity, so for the duration of the trip, these thinking processes stop. Recent data also suggest some of these brain effects are lasting beyond the presence of the drug. This in itself may give the patients a view of a depression-free world, but more than that, the increase in brain flexibility produced by 5-HT2A receptor stimulation may allow the brain to reset itself into a different, depression-free state. Another possibility shown by fMRI connectivity analysis is that psilocybin may alter the pattern of dominant connections between the frontal cortex and subcortical regions.

There is now good conceptual evidence that the antidepressant effects of psychedelics are mediated in quite a different way to those of serotonin-acting antidepressant drugs such as SSRIs. The latter seems to provide a buffer against stress by enhancing serotonin function at the 5-HT1A receptor and so making the person more resilient. In contrast, we believe that psychedelics, via stimulation of the 5-HT2A receptor, work to reset the brain processes that underpin depressive thinking and so allow the patient to work through their issues and so overcome their depression.

MDMA works differently; it is a serotonin-releasing agent. fMRI studies have revealed its main activity being in the limbic system, where it suppresses activity, especially in the amygdala and hippocampus. This damping down of the emotional memory circuits helps patients relive their traumas and, at the same time, overcome and eventually extinguish the intense emotions that go with these memories.

MICRODOSING

In recent years the use of subpsychedelic and even subdetectable doses of psychedelics, typically 5% to 10% of full psychedelic doses, has become popular, especially among people working in the creative industries who believe it improves creativity. Some claim it helps keep depression and/or anxiety at bay. These doses are taken regularly, usually a couple of times a week over longer periods (weeks/months). However, since these drugs are illegal, there is no way of knowing for sure what they are and what exact doses are being taken. The best-controlled trial, using a novel citizen science "self-blinding" approach conducted by Szigeti et al. found that the effect of LSD and psilocybin microdosing was likely mostly due to expectation.

FURTHER READINGS

Bogenschutz MP, Forcehimes AA, Pommy JA, Wilcox CE, Barbosa P, Strassman RJ. Psilocybin-assisted treatment for alcohol dependence: a proof-of-concept study. *J Psychopharmacol*. 2015;29:289–299.

Carhart-Harris RL, Bolstridge M, Rucker J, et al. Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *Lancet Psychiatry*. 2016;3:619–627.

Carhart-Harris RL, Bolstridge M, Day CMJ, et al. Psilocybin with psychological support for treatment-resistant depression: six-month follow-up. *Psychopharmacology (Berl)*. 2018;235:399–408.

Carhart-Harris R, Giribaldi B, Watts R, et al. Trial of psilocybin versus escitalopram for depression. *N Engl J Med*. 2021;384:1402–1411.

Carhart-Harris RL, Murphy K, Leech R, et al. The effects of acutely administered MDMA on spontaneous brain function in healthy volunteers measured with arterial spin labelling and BOLD resting-state functional connectivity. *Biol Psychiatry*. 2015;78:554–562.

Carhart-Harris RL, Nutt DJ. Serotonin and brain function: a tale of two receptors. *J Psychopharmacol*. 2017;31:1091–1120.

Davis AK, Barrett FS, May DJ, et al. Effects of psilocybin-assisted therapy on major depressive disorder a randomized clinical trial. *JAMA Psychiatry*. 2021;78:481–489.

Daws RE, Timmermann C, Giribaldi B, et al. Increased global integration in the brain after psilocybin therapy for depression. *Nat Med.* 2022;28(4):844–851.

Gasser P, Kirchner K, Passie T. LSD-assisted psychotherapy for anxiety associated with a life-threatening disease: a qualitative study of acute and sustained subjective effects. *J Psychopharmacol*. 2015;29:57–68.

Goodwin GM, Aaronson ST, Alvarez O, et al. Single-dose psilocybin for a treatment-resistant episode of major depression. *N Engl J Med*. 2022;387(18):1637–1648.

Green AR, Mechan AO, Elliott JM, O'Shea E, Colado MI. The pharmacology and clinical pharmacology of 3,4-methylenedioxymethamphetamine (MDMA, "ecstasy"). *Pharmacol Rev*. 2003;55(3):463–508.

Griffiths RR, Johnson MW, Carducci MA, et al. Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: a randomized double-blind trial. *J Psychopharmacol*. 2016;30:1181–1197.

Grinspoon L, Bakalar J. The psychedelic drug therapies. *Curr Psychiatr Ther*. 1981;20:275–283.

Grob CS, Danforth AL, Chopra GS, et al. Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. *Arch Gen Psychiatry*. 2011;68:71–78.

Gukasyan N, Davis AK, Barrett FS, et al. Efficacy and safety of psilocybin-assisted treatment for major depressive disorder: prospective 12-month follow-up. *J Psychopharmacol*. 2022;36(2):151–158.

Johnson MW, Garcia-Romeu A, Cosimano MP, Griffiths RR. Pilot study of the 5-HT2AR agonist psilocybin in the treatment of tobacco addiction. *J Psychopharmacol*. 2014;28:983–992.

Kaelen M, Giribaldi B, Raine J, et al. The hidden therapist: evidence for a central role of music in psychedelic therapy. *Psychopharmacology (Berlin)*. 2018;235(2):505–519.

Krebs TS, Johansen PO. Lysergic acid diethylamide (LSD) for alcoholism: meta-analysis of randomized controlled trials. *J Psychopharmacol*. 2012;26:994–1002.

Mitchell JM, Bogenschutz M, Lilienstein A, et al. A randomized, double-blind, placebo controlled phase 3 study assessing efficacy and safety of MDMA-assisted therapy for the treatment of severe PTSD. *Nat Med*. 2021;27(6):1025–1033.

Mitchell JM, Ot'alora GM, van der Kolk B, et al. MDMA-assisted therapy for moderate to severe PTSD: a randomized, placebo-controlled phase 3 trial. *Nat Med*. 2023;29(10):2473–2480.

Mithoefer MC, Mithoefer AT, Feduccia L, et al. MDMA-assisted psychotherapy for post-traumatic stress disorder in military veterans, firefighters, and police officers: a randomised, double-blind, dose-response, phase 2 clinical trial. *Lancet Psychiatry*. 2018;5(6):486–497.

Mithoefer MC, Wagner TM, Mithoefer AT, Jerome L, Doblin R. The safety and efficacy of ±3,4-methylenedioxymethamphetamine-assisted psychotherapy in subjects with chronic, treatment-resistant posttraumatic stress disorder: the first randomized controlled pilot study. *J Psychopharmacol*. 2011;25(4):439–452.

Moreno FA, Wiegand CB, Taitano EK, Delgado PL. Safety, tolerability, and efficacy of psilocybin in 9 patients with obsessive-compulsive disorder. *J Clin Psychiatry*. 2006;67:1735–1740.

Nutt D, Carhart-Harris R. The current status of psychedelics in psychiatry. *JAMA Psychiatry*. 2021;78:121–122.

Nutt DJ, de Wit H. Putting the MD back into MDMA, *Nat Med*. 2021;27:950–951.

Nutt D, Erritzoe D, Carhart-Harris R. Psychedelic psychiatry's brave new world. *Cell*. 2020;181:24–28.

Palhano-Fontes F, Barreto D, Onias H, et al. Rapid antidepressant effects of the psychedelic ayahuasca in treatment-resistant depression: a randomized placebo-controlled trial. *Psychol Med*. 2019;49(4):655–663.

Raison CL, Sanacora G, Woolley J, et al. Single-dose psilocybin treatment for major depressive disorder: a randomized clinical trial. *JAMA*. 2023;330:843–853.

Ross S, Bossis A, Guss J, et al. Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: a randomized controlled trial. *J Psychopharmacol*. 2016;30:1165–1180.

Sessa B, Higbed L, O'Brien S, et al. First study of safety and tolerability of 3,4-methylenedioxymethamphetamine-assisted psychotherapy in patients with alcohol use disorder. *J Psychopharmacol*. 2021;35:375–383.

Sessa B, Nutt D. Making a medicine out of MDMA. *Brit J Psychiatry*. 2015;206:4–6.

Sloshower J, Skosnik PD, Safi-Aghdam H, et al. Psilocybin-assisted therapy for major depressive disorder: an exploratory placebo-controlled, fixed-order trial. *J Psychopharmacol*. 2023;37:698–706.

Szigeti B, Kartner L, Blemings A, et al. Self-blinding citizen science to explore psychedelic microdosing. *Elife*. 2021;10:e62878.

von Rotz R, Schindowski EM, Jungwirth J, et al. Single-dose psilocybin-assisted therapy in major depressive disorder: a placebo-controlled, double-blind, randomised clinical trial. *EClinicalMedicine*. 2023;56:101809.

Watts R, Day C, Krazanowski J, et al. Patients' accounts of increased "connectedness" and "acceptance" after psilocybin for treatment-resistant depression. *J Humanist Psychol*. 2017;57(5):520–564.

33.15 Hormone Receptor Enhancers

33.15a Melatonin Receptor Agonists

TAWNY L. SMITH, PHARM.D., B.C.P.P., AND D. JEFFREY NEWPORT, M.D., M.S., M.DIV.

Neuroscience-Based Nomenclature (NbN)	
Melatonin, Rameleton, Tasimelteon	
Pharmacologic Domain	Melatonin
Mechanism of Action	Receptor agonists (MT1, MT2)
Agomelatine	
Pharmacologic Domain	Melatonin, Serotonin
Mechanism of Action	Receptor agonist (M1, M2), receptor antagonist (5-HT2B, 5-HT2C)

Insomnia is the most common sleep disorder, affecting approximately one-third of U.S. adults (NHLBI 1999), with as many as one in five regularly experiencing insomnia. *DSM-5-TR* characterizes Insomnia Disorder as difficulty initiating or maintaining sleep, or early morning awakening with an inability to return to sleep, occurring 3 or more nights per week for at least 3 months that results in significant psychosocial impairment, affecting physical and mental health, productivity, and quality of life. While therapeutic guidelines recommend cognitive behavioral therapy as a first-line intervention for insomnia, pharmacotherapies are commonly employed. Historically, many classes of pharmacologic agents have been used to treat insomnia, including classic benzodiazepines, other benzodiazepine receptor agonists (i.e., “z-drugs”), antihistamines, naturopathic agents, and even sedating antidepressants and antipsychotics. Potential side effects of such drugs include residual daytime sedation, gait instability, memory impairment, weight gain and metabolic complications, and abuse and withdrawal potential. Consequently, the availability of safe, well-tolerated medications with low side-effect profiles for insomnia is imperative.

The pineal hormone melatonin (N-acetyl-5-methoxytryptamine; 5-methoxy-N-acetyltryptamine) was first isolated in the 1950s. Nearly 40 years would elapse before human melatonin receptors were characterized. Melatonin is released from the pineal gland in accordance with a circadian rhythm producing high nocturnal levels (400 pM) peaking around 2 AM and low daytime levels (40 pM). In most vertebrates, nocturnal melatonin secretion varies inversely with the length of the day, thereby regulating not only sleep but also seasonal variation in reproductive behavior in some species. Interestingly, the circadian clock in humans is slightly longer than 24 hours, but the suprachiasmatic nucleus (SCN) of the hypothalamus synchronizes melatonin synthesis to the light/dark cycle, transmitting neural signals from retinal neurons to the pineal gland. In the absence of environmental light–dark cues, circadian desynchronization can occur. In fact, 50% to 70% of totally blind individuals experience non–24-hour sleep–wake phase disorders.

Melatonergic dysfunction has garnered heightened attention as a putative pathophysiologic mechanism underlying circadian disruptions associated with insomnia, depression, and other disorders. Accordingly, the medicinal use of melatonin, categorized in the United States as a nutritional supplement, has rapidly expanded and has been accompanied by the development and introduction of several melatonin receptor agonists. Agents currently approved for use in the United States, Europe, or Australia include melatonin, agomelatine, ramelteon, and tasimelteon.

CHEMISTRY

Melatonin is derived from serotonin within the pineal gland and other sites. As a result, the molecular structure of melatonin (Fig. 33.15a–1) closely resembles that of serotonin. Both molecules are comprised of an indole aromatic ring (i.e., a 6-member benzene ring fused with a 5-member nitrogen-containing pyrrole ring) with a side chain at position 3. To form melatonin, the terminal amine group on serotonin’s side chain is acetylated, producing an acylaminoethyl side chain, and serotonin’s hydroxyl group is methylated, producing a methoxy group at position 5. These changes render melatonin a more lipophilic molecule that, unlike serotonin, readily crosses the blood–brain barrier.

All currently available melatonin receptor agonists are structurally similar to melatonin. Most similar is agomelatine, the only structural difference being the replacement of melatonin’s indole (fused benzene and pyrrole rings) aromatic ring with a naphthalene (two fused benzene rings) aromatic ring (Fig. 33.15a–2). Ramelteon and tasimelteon preserve melatonin’s acylaminoethyl side chain but replace the indole with a benzofuran aromatic ring in which a 6-member benzene ring is fused with a 5-member oxygen-containing furan ring (Figs. 33.15a–3 and 33.15a–4).

PHARMACOLOGIC ACTIONS

Pharmacokinetics

Pharmacokinetic profiles for melatonin and the three commercially available melatonin receptor agonists are remarkably similar (Table 33.15a–1). Each is readily and rapidly absorbed when administered orally, reaching peak plasma concentrations within an hour of administration. Concomitant intake of a high-fat meal lowers the absorption of tasimelteon (44%), ramelteon (22%), and melatonin (13%), delaying the time to maximum concentration, but does not affect agomelatine absorption.

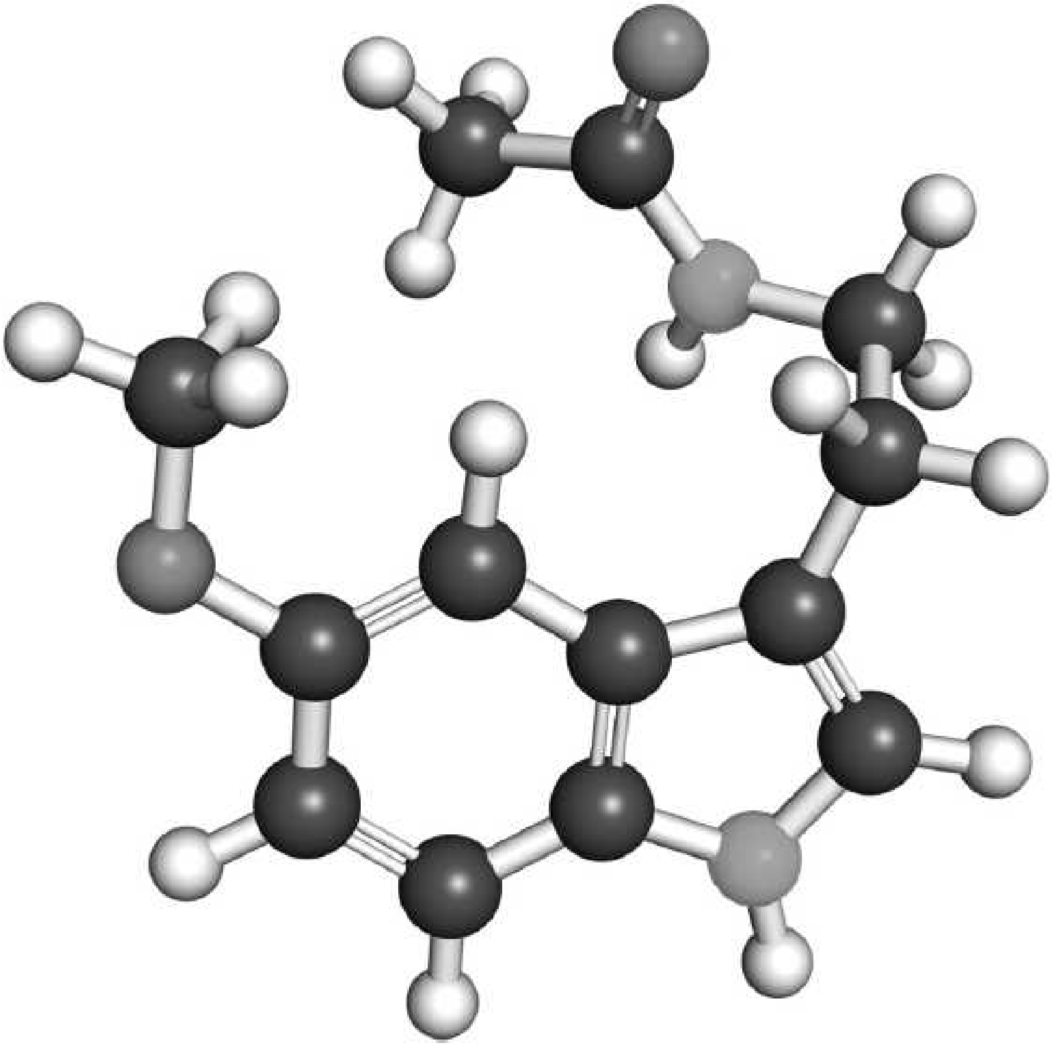

FIGURE 33.15a–1. Molecular structure of melatonin. Light gray, nitrogen; gray, carbon; medium gray, oxygen; white, hydrogen.

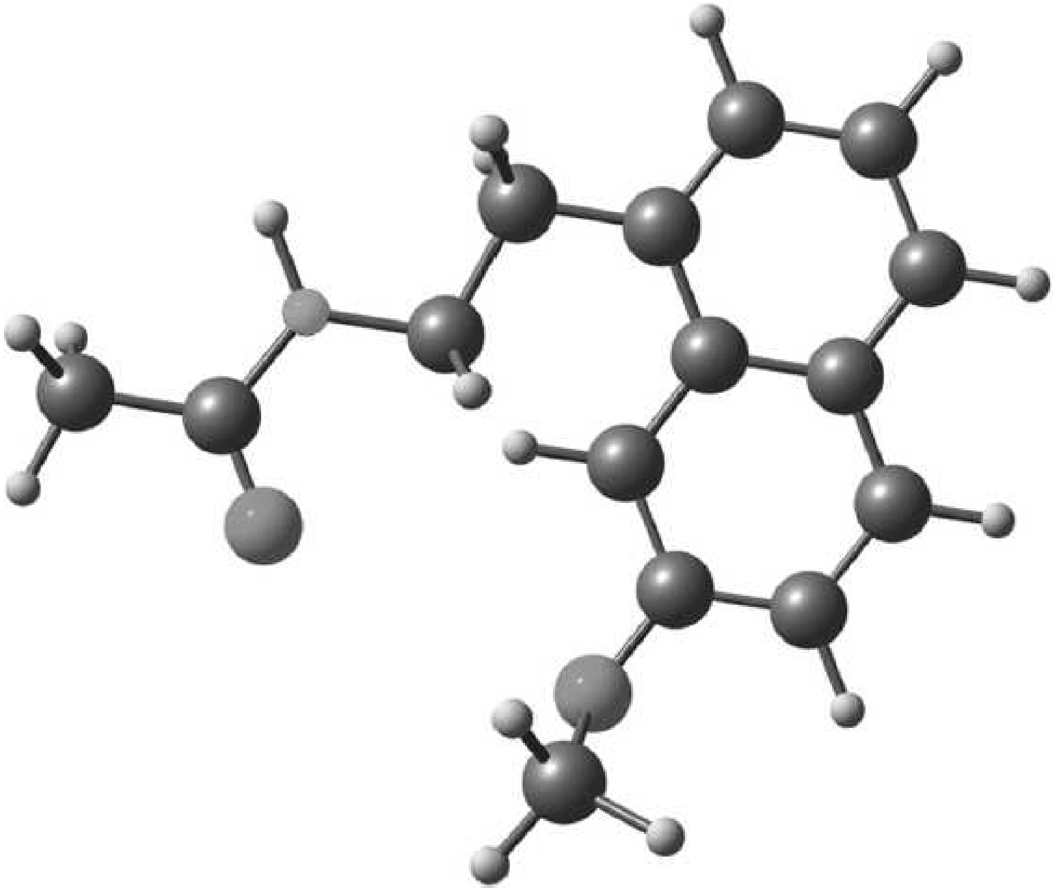

FIGURE 33.15a–2. Molecular structure of agomelatine. Light gray, nitrogen; gray, carbon; medium gray, oxygen; white, hydrogen.

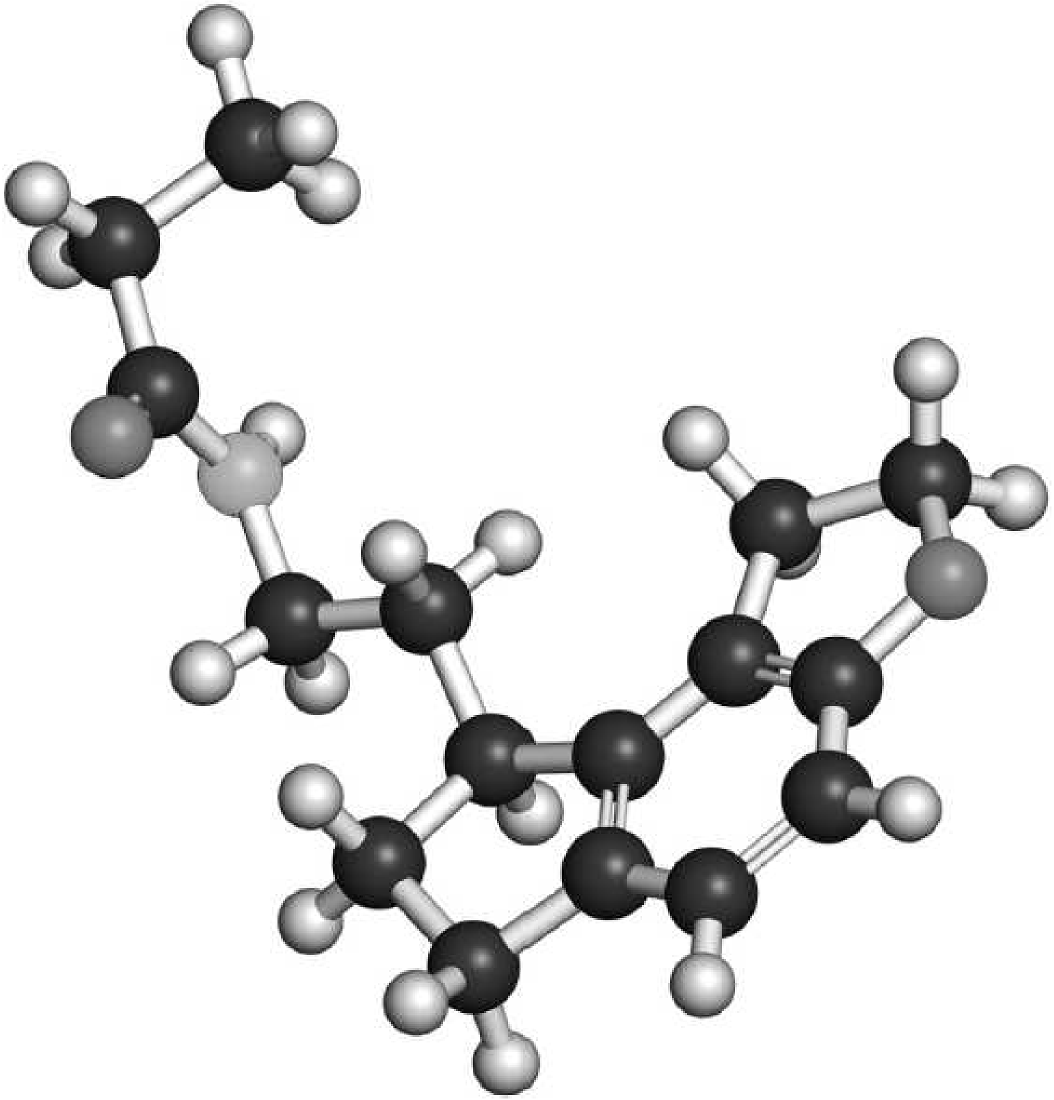

FIGURE 33.15a–3. Molecular structure of ramelteon. Light gray, nitrogen; gray, carbon; medium gray, oxygen; white, hydrogen.

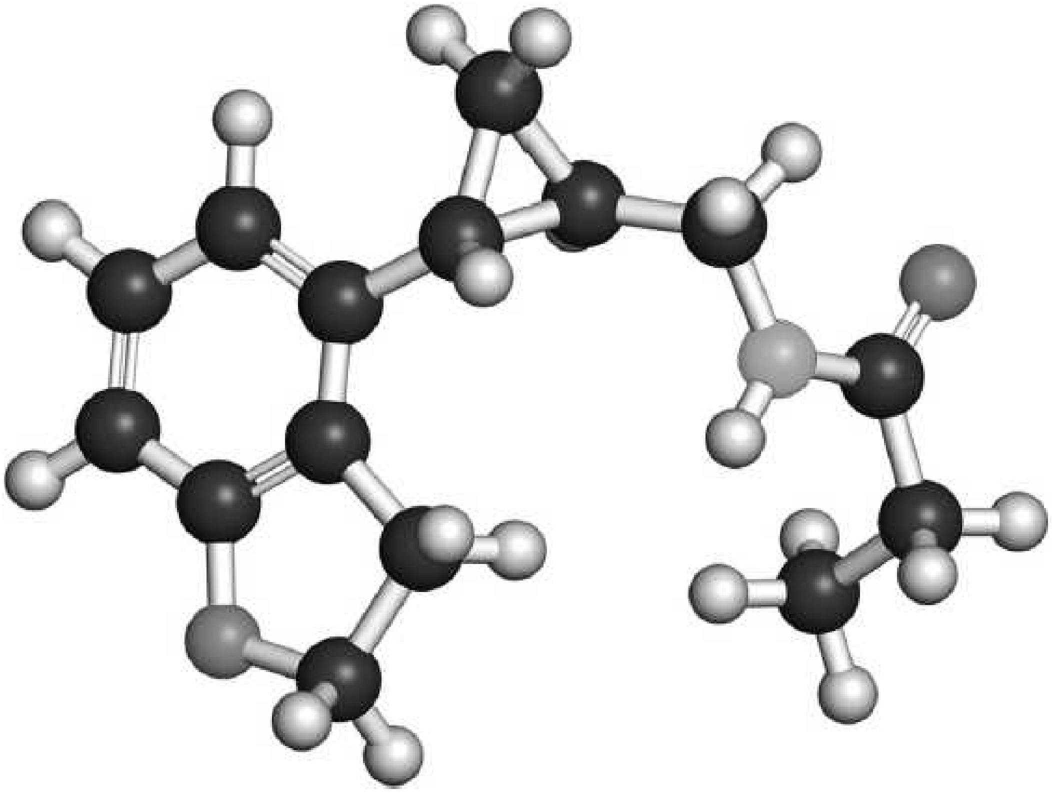

FIGURE 33.15a–4. Molecular structure of tasimelteon. Light gray, nitrogen; gray, carbon; medium gray, oxygen; white, hydrogen.

Despite uniformly high absorption rates (≥80%), absolute oral bioavailability is only 38% for tasimelteon, 15% for melatonin, and <5% for the others due to extensive hepatic first-pass metabolism and probable tissue uptake. All four are rapidly and extensively metabolized with elimination half-lives less than 2 hours and no more than 2% of the parent compound excreted unchanged in the urine.

All four agents undergo extensive oxidative metabolism primarily by the cytochrome P450 (CYP) 1A2 isoenzyme. Ramelteon and tasimelteon are also minor substrates for CYP3A4, whereas melatonin, agomelatine, and ramelteon are minor substrates for the CYP2C9 and CYP2C19 isoenzymes.

Renal impairment does not change the rate of drug clearance for any of the four melatonin receptor agonists. Conversely, mild and moderate hepatic impairment, respectively, produce 70-fold and 140-fold increases in agomelatine exposure and 4-fold and 10-fold increases in ramelteon exposure. Alterations in tasimelteon clearance have not been observed in mild to moderate hepatic impairment. Although the effect of hepatic impairment upon the clearance of exogenous melatonin has not been formally evaluated, it has been reported that clearance of endogenous melatonin is reduced by hepatic impairment.

The pharmacokinetic profiles for melatonin, ramelteon, and tasimelteon are significantly affected by age, with older adult patients exhibiting higher peak concentrations, diminished clearance rates, and longer elimination half-lives. However, these differences are not thought to be clinically relevant. Age-associated pharmacokinetic changes have not been observed with agomelatine.

Table 33.15a–1.
Pharmacokinetic Profiles of Melatonin Receptor Agonists

Generic Name	Oral Absorption (%)	Time to Peak Concentration (Hours)	Absolute Oral Bioavailability (%)	Protein Binding (%)	Elimination Half-Life (Hours)	Renal Excretion of Parent Compound (%)	Metabolism (Cytochrome P450 Substrate)		
							1A2	2C9/19	3A4
Agomelatine	≥80%	1–2	<5%	95%	1.0–2.0	<1%	+++	+	
Melatonin	100%	0.75	15%	60%	0.8	2%	+++	+	
Ramelteon	84%	0.75	1.8%	82%	1.3	<1%	+++	+	+
Tasimelteon	≥90%	0.5-3	38%	90%	1.3	<1%	+++		+

There is a distinct gender difference in pharmacokinetic profile, with higher peak concentrations and/or longer elimination half-lives reported among women for all four agents, arguably a consequence of CYP1A2 inhibition by estradiol. If so, gender differences may be most pronounced among premenopausal women.

Finally, all four agents are highly lipophilic compounds that readily traverse the blood–brain barrier. Of note, melatonin is less tightly protein-bound (60%) than the other melatonin receptor agonists.

Pharmacodynamics

Melatonin's physiologic activity is primarily mediated by two membrane-bound G protein–coupled receptors, melatonin receptor 1 (MT_1) and melatonin receptor 2 (MT_2), which are located in numerous brain regions and peripheral tissues. Both MT_1 and MT_2 receptors are expressed in the SCN of the hypothalamus, a region adjacent to the optic chiasm that is well recognized for its role in regulating circadian rhythms. Additional MT_1 receptors are located in the *pars tuberalis* of the pituitary and retina. MT_2 receptors are also found in the hippocampus, cerebellum, and retina. Although their respective roles are not clearly delineated, MT_1 receptors in the SCN are thought to be involved in sleep promotion, whereas MT_2

receptors are believed to play a major role in melatonin's phase-shifting activity.

Melatonin receptor 3 (MT_3), a quinone reductase 2 enzyme present in multiple peripheral tissues, may explain the antioxidant properties of melatonin but has a negligible impact on the sleep–wake cycle. Other physiologic roles ascribed to melatonin include regulation of metabolic, cardiovascular, and immune systems with reported antioxidant, antinociceptive, and anticancer effects. These may be mediated, at least in part, by activation of MT_1 and MT_2 receptors, which like MT_3, are present in various peripheral tissues.

Melatonin and the other commercially available agents (i.e., agomelatine, ramelteon, tasimelteon) are all considered nonselective MT_1/MT_2 receptor agonists, albeit with distinct affinity profiles (Table 33.15a–2). None has a pharmacologically meaningful affinity for benzodiazepine, histaminic, or cholinergic receptors. The affinities of agomelatine, ramelteon, and tasimelteon for the MT_3 receptor are obscure. Among the current agents, ramelteon has the greatest MT_1 selectivity, with six times melatonin's affinity for the MT_1 receptor. Alternatively, tasimelteon has a two to fourfold higher affinity for the MT_2 than the MT_1 receptor, which would theoretically explain its particular utility in remedying circadian rhythm disturbances.

Agomelatine, approved in Europe and Australia for the treatment of depression, is not only an MT_1 and MT_2 agonist but also a serotonin-2C ($5HT_{2C}$) receptor antagonist. It remains debated whether the antidepressant efficacy of agomelatine is a consequence solely of its action on $5HT_{2C}$ receptors or of synergy between its effects on MT_1, MT_2, and $5HT_{2C}$ receptors.

Table 33.15a–2.
Receptor Affinities of Melatonin Receptor Agonists

Generic Name	**Receptor Affinity (K_i)**		
	MT_1	**MT_2**	**MT_1/MT_2 K_i Ratio**
Agomelatine	62 pM	268 pM	0.231
Melatonin	81 pM	383 pM	0.211
Ramelteon	14 pM	112 pM	0.125
Tasimelteon	350 pM	170 pM	2.059

MT_1, Melatonin 1 Receptor; MT_2, Melatonin 2 Receptor.

CLINICAL DRUG STUDIES

Numerous randomized clinical trials have demonstrated the beneficial effects of melatonin and other melatonin receptor agonists for sleep disturbances. For example, a recent umbrella review encompassing 18 meta-analyses and systematic reviews concluded that short-term use for insomnia improved sleep quality and sleep latency.

Ramelteon is approved in the United States for the treatment of insomnia characterized by difficulty with sleep onset. A meta-analysis encompassing 13 trials and 5,812 participants concluded, ramelteon improves sleep quality and reduces sleep latency by 4.30 minutes in a nondose-dependent manner, with somnolence as the only significant side effect. A similar meta-analysis including eight trials reported improvements in subjective and polysomnographic sleep latency, total sleep time, and latency to REM with ramelteon. Moreover, improvements in sleep latency produced by ramelteon may be enduring, with one study demonstrating reduced sleep latency for up to 6 months.

Tasimelteon is approved in the United States for the treatment of Non–24-Hour Sleep–Wake Disorder (Non-24) in blind adults. In a controlled trial of 84 Non-24 patients, tasimelteon significantly outperformed placebo in circadian rhythm entrainment to a 24-hour day (20% vs. 3%) with longer nighttime sleep duration and less daytime nap time. Moreover, a subsequent Non-24 withdrawal study by the same group reported greater maintenance of 24-hour synchronization among those maintained on tasimelteon than those who transitioned to placebo (90% vs. 20%). Two reports drawn from three randomized trials collectively encompassing 769 participants experiencing insomnia subsequent to a sleep-time shift, for example, jet lag, reported greater sleep efficiency among those receiving tasimelteon than placebo. Finally, a clinical trial (NCT00548340) for primary insomnia (*n* = 321) demonstrated significantly greater reductions in sleep latency for tasimelteon at both 20 mg and 50 mg doses than placebo.

Melatonin receptor agonists have consistently outperformed placebo in clinical trials of insomnia and other circadian sleep disturbances. Unfortunately, tolerance to these agents has not been systematically assessed, nor have they been scrutinized in head-to-head comparisons against other classes of hypnotic agents. Nevertheless, one meta-analysis reported that polysomnographically determined reductions in sleep latency and wakefulness after sleep onset for extended-release melatonin (2 mg) and ramelteon (16 mg) were less than that of benzodiazepines but comparable to changes produced by other benzodiazepine receptor agonists.

Because of the negative consequences associated with long-term benzodiazepine use, there has been considerable interest in the use of melatonin to aid in reducing the frequency of prescribing benzodiazepines as hypnotics. Unfortunately, in a meta-analysis and systematic review of six trials, melatonin did not outperform placebo in enabling benzodiazepine discontinuation. In 2020, a retrospective observational study (*n* = 170) by An and colleagues reported a successful switch from chronic zolpidem or benzodiazepine use to control-release melatonin in nearly half of the study participants.

The utility of melatonin receptor agonists in the management of mood disorders has also been a focus of inquiry, albeit with mixed results. A controlled trial (NCT00705003) evaluating melatonin (3 mg) augmentation of buspirone for depression demonstrated no evidence of therapeutic efficacy. Similarly, a controlled trial of ramelteon (8 mg) augmentation of escitalopram for depression (NCT00642694) was terminated due to lack of efficacy. Finally, in an 8-week depression trial (NCT01428661) including 507 subjects, 20 mg tasimelteon failed to separate from placebo in reducing depressive symptom severity.

Among melatonin receptor agonists, the strongest evidence for efficacy in the treatment of depression is found with agomelatine, which, in Europe and Australia, is indicated for the treatment of depression. It is suggested that the antidepressant efficacy of agomelatine stems both from its effects on mood and on circadian resynchronization. A meta-analysis conducted by Taylor and colleagues, incorporating 20 placebo-controlled agomelatine trials and 7,460 subjects, demonstrated a significant odds ratio for the antidepressant therapeutic response of 1.25 (95% CI: 1.11–1.40). Interestingly, in a meta-analysis of 21 antidepressants, which included 522 double-blind trials and 116,477 subjects, agomelatine had a therapeutic response rate comparable to other antidepressants and the highest acceptability rate.

The foundation for the role of chronobiotic therapies for mood disorders has largely emerged from the study of seasonal affective disorder (SAD). Most patients with SAD experience a phase delay, or later secretion of dim light melatonin, due to the absence of early morning light during the winter months in the northern hemisphere. Both the dose and timing of chronobiotic administration are arguably key to successful treatment. Proper treatment results in phase advancement with improvements in mood symptoms, while incorrect dosing or timing may actually worsen insomnia and mood disturbances. Unfortunately, a recent Cochrane Review, limited to only one trial, reported inconclusive evidence with wide confidence intervals encompassing both prevention and worsening of SAD symptoms with agomelatine. No SAD studies of other melatonin agonists were identified.

Controlled trials of melatonin receptor agonists for bipolar disorder have produced mixed results. A trial of ramelteon (8 mg) as a maintenance therapy augmentation agent for euthymic bipolar patients reported a significantly longer time to relapse among those receiving ramelteon than placebo (188 vs. 84 days, respectively). For bipolar mania, two augmentation trials reported that neither melatonin nor ramelteon outperformed placebo; however, a third study ($n = 60$) in manic patients treated with lithium and risperidone demonstrated a modest but statistically significant greater reduction in manic symptoms for those receiving melatonin 6 mg than those receiving placebo.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Cardiovascular System

Endogenous melatonin is likely involved in the circadian regulation of the cardiovascular with decreases in cardiac output, blood pressure, and heart rate when melatonin levels peak during the evening hours. Rodent studies indicate similar effects of exogenous administration of melatonin.

Clinical trial data have not indicated any adverse effects on cardiovascular function among the four commercially available melatonin receptor agonists. In fact, it has been speculated that melatonin receptor agonists may have beneficial cardiovascular effects.

Central Nervous System

As a group, the melatonin receptor agonists compare favorably to benzodiazepines and other hypnotics with respect to adverse effects on the central nervous system. Indeed, melatonin receptor agonists may promote sleep and aid in the prevention of delirium in the intensive care setting, though their benefit for preventing delirium has been questioned. Moreover, the risk of abuse liability for this class of medications has not been found in small cross-over trials.

Endocrine System

Preclinical studies indicate that supraphysiologic administration of exogenous melatonin may disrupt secretory profiles for numerous hormones, including corticosteroids, gonadal steroids, luteinizing hormone, insulin, prolactin, and thyroid hormone.

In a controlled trial of adults with chronic insomnia, ramelteon 16 mg nightly for 6 months was not associated with alterations in indices of adrenal or thyroid function; however, prolactin concentrations increased 34% in female participants, although no clinical effects of elevated prolactin were noted. Endocrine function has not been specifically examined in clinical trials of the other commercially available melatonin receptor agonists.

Hepatic and Renal Systems

There is no evidence that clearance of any of the four available melatonin receptor agonists is altered by preexisting renal impairment. Nor is there evidence that these agents impair renal function. In fact, some preclinical studies suggest that melatonin may protect the kidneys from renal ischemia and reperfusion injury, exposure to carcinogens, hemorrhagic shock, or other injuries.

Agomelatine, unfortunately, has been associated with hepatic dysfunction. During clinical trials, 1.4% of patients receiving agomelatine 25 mg and 2.5% of those receiving agomelatine 50 mg experienced elevations in serum transaminase concentrations greater than three times the upper limit of normal. Rare occurrences of hepatitis, γ-glutamyltransferase concentrations three times the upper limit of normal, jaundice, and hepatic failure have been reported in association with agomelatine therapy. Hepatic injury has not been reported in association with any of the other three approved melatonin receptor agonists.

Respiratory System

In contrast to benzodiazepines and other hypnotics, melatonin receptor agonists appear to be devoid of respiratory depressant effects. For example, in studies of patients with obstructive sleep apnea and chronic obstructive pulmonary disease, ramelteon did not adversely affect apnea-hypopnea indexes or arterial oxygen saturation levels.

THERAPEUTIC INDICATIONS

As shown, melatonin agonists provide a novel approach to the treatment of insomnia and other circadian sleep disturbances (Table 33.15a–3). In addition to its effect on the circadian rhythm, melatonin has also been reported to possess anti-inflammatory, antitumor, vascular, retinal, antioxidant, neuroprotective, and pain-modulatory effects, perhaps providing other avenues for which to explore its therapeutic potential. For example, because of its potential anti-inflammatory and antioxidant properties, the use of adjuvant ramelteon is being studied in individuals with SARS-CoV-2 infection to evaluate whether it can modulate protective effects against acute lung injury (NCT04470297).

Melatonin receptor agonists also hold potential utility in the treatment of depression and other mental disorders associated with disturbances of circadian rhythmicity. Indeed, agomelatine is indicated for the treatment of major depression in Europe. It is speculated that agomelatine's antidepressant properties may be due to a synergism between its effects on melatonin receptors and $5HT_{2c}$. Although studies of other melatonin receptor agonists for depression have produced disappointing results, one study of tasimelteon, which overall produced negative results, suggested tasimelteon might offer benefit to African Americans experiencing depression. Further study is required before melatonin receptor agonists can be recommended for bipolar disorder and the prevention of delirium.

Table 33.15a–3.
Formulations of Commercially Available Melatonin Receptor Agonists

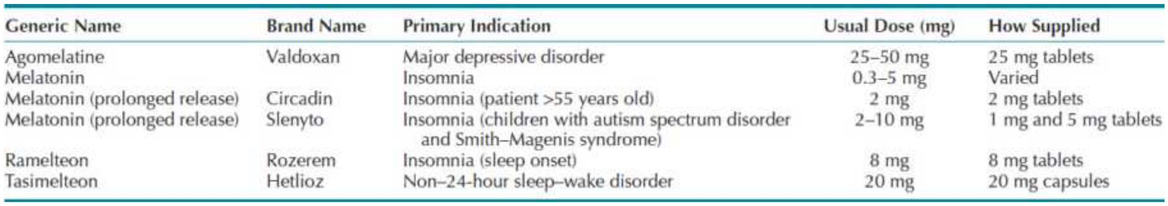

Generic Name	Brand Name	Primary Indication	Usual Dose (mg)	How Supplied
Agomelatine	Valdoxan	Major depressive disorder	25–50 mg	25 mg tablets
Melatonin		Insomnia	0.3–5 mg	Varied
Melatonin (prolonged release)	Circadin	Insomnia (patient >55 years old)	2 mg	2 mg tablets
Melatonin (prolonged release)	Slenyto	Insomnia (children with autism spectrum disorder and Smith–Magenis syndrome)	2–10 mg	1 mg and 5 mg tablets
Ramelteon	Rozerem	Insomnia (sleep onset)	8 mg	8 mg tablets
Tasimelteon	Hetlioz	Non–24-hour sleep–wake disorder	20 mg	20 mg capsules

PRECAUTIONS AND ADVERSE REACTIONS

In many respects, the side effect profile of the melatonin receptor agonists compares favorably to other sleep agents. Unlike benzodiazepine hypnotics, melatonin agonists have not been associated with prominent cognitive impairment, fall risk, discontinuation-associated withdrawal syndromes, or rebound insomnia. The melatonin agonists are also devoid of prominent anticholinergic (e.g., dry mouth, constipation, blurred vision) or antihistaminic (e.g., carbohydrate craving, weight gain) side effects associated with other hypnotic agents.

Pregnancy

Endogenous maternal melatonin plays an important role during pregnancy. Although the fetal pineal gland does not yet produce melatonin (infant melatonin production does not begin until 2 to 3 months after delivery), there are more melatonin binding sites in both the fetal pituitary and fetal hypothalamus than in the adult. It is speculated that maternal melatonin, known to cross the placenta, establishes periodicity and entrains the circadian pacemaker. For example, in evening deliveries, significantly elevated melatonin levels are detected via sampling of the umbilical artery or vein at birth. One hypothesis is that maternal melatonin might communicate important information about maternal sleep–wake cycles and entrain the fetal pineal gland so that fetal circadian rhythms may be found as early as 3 months. Preclinical evidence suggests other roles for endogenous melatonin during fetal life, including influencing the development of reproductive organs, regulating gonadal hormone activity, and protecting the fetal brain from oxidative damage during episodes of fetal hypoxia.

Despite the growing evidence that endogenous melatonin benefits the fetus, exogenous melatonin administration, which likely produces markedly higher fetal circulating concentrations, should not carry a presumption of safety. For example, in a series of rodent studies, administering exogenous melatonin during pregnancy at doses as low as 2 to 10 times recommended human doses have demonstrated decreased luteinizing hormone levels, decreased ovarian and uterine size, and delayed vaginal opening in the offspring. Following delivery, offspring exposed to exogenous melatonin exhibit disruption in estrous cycles and blunted responses to testosterone, suggesting impairment in the maturation of the hypothalamic–pituitary–gonadal and prolactin axes.

There are no controlled data addressing the reproductive safety of exogenous administration of melatonin or other melatonin receptor agonists in humans.

In summary, there is no clear evidence of teratogenicity, though there is preclinical evidence that increased levels of melatonin may interfere with the regulation of sex hormone levels and cycles, as well as delay sexual maturation. Pending further data, exogenous administration of melatonin and other melatonin receptor agonists should be avoided during pregnancy.

Lactation

Existing studies have examined levels of endogenous, but not exogenous, melatonin in breast milk. Cyclic levels of melatonin in breast milk replicate maternal patterns, including elevations in the evening and decreases during daylight, with endogenous melatonin breast milk levels ranging from 35% to 80% of maternal serum levels. Circadian fluctuations in breast milk melatonin concentrations may dictate nursing schedules for the infant.

In one case report, agomelatine was detectable in breast milk, with peak concentrations occurring 2 hours after the drug was administered with undetectable levels 4 hours after administration. The milk:plasma ratio was 0.19. Clinical breast milk excretion data have not been published for other melatonin receptor agonists.

In a recent edition of Hale's standard lactation reference, both melatonin and ramelteon are assigned an LE (no data-probably compatible) rating.

DRUG INTERACTIONS

All currently available melatonin receptor agonists, including melatonin itself, are predominantly metabolized by the CYP1A2 isoenzyme and, to a lesser extent, the CYP2C9/2C19 and/or CYP3A4 isoenzymes. Thus, there are potential interactions with drugs that inhibit or induce the activity of these P450 isoenzymes (Table 33.15a–4).

Of greatest concern are drug interactions mediated by CYP1A2. For example, coadministration with fluvoxamine, a potent CYP1A2 inhibitor, produces a 6.5-fold increase in total tasimelteon exposure, a 17-fold increase in melatonin exposure, a 60-fold increase in agomelatine exposure, and a whopping 160-fold increase in ramelteon exposure. By comparison, coadministration with ketoconazole, a potent CYP3A4 inhibitor, produces a 0.84-fold increase in ramelteon exposure and a 0.54-fold increase in tasimelteon exposure, although the clinical significance of this is unknown. Fluconazole, a potent CYP2C9 inhibitor, produced a 1.5-fold increase in ramelteon exposure and was not observed to alter agomelatine exposure.

The magnitude of the drug interaction observed between fluvoxamine and the four melatonin receptor agonists raise concern for interactions with other agents that alter 1A2 activity. For example, coadministration with estrogen-containing compounds (e.g., oral contraceptives, hormone replacement therapies) may produce similar increases in levels of exposure to these agents. Unfortunately, no data are available to clarify the magnitude of this potential interaction. Tobacco, a known inducer of CYP1A2, may produce the opposite effect, which is lowering exposure and thus the efficacy of melatonergic agents. Indeed, smoking has been associated with a 40% reduction in exposure to tasimelteon and a 3.7-fold lowering of cumulative exposure to agomelatine.

Table 33.15a–4.
Potential Drug interactions for Melatonin Receptor Agonists

Pharmacokinetic Activity	Cytochrome P450 Isoenzyme		
	1A2	2C9/2C19	3A4
Substrates[a]	Agomelatine	Agomelatine	
	Melatonin	Melatonin	
	Ramelteon	Ramelteon	Ramelteon
	Tasimelteon		Tasimelteon
Inhibitors[b]	Estrogen[c] Fluvoxamine Quinolones[d]	Amiodarone Azole antifungals[e] Fluoxetine Fluvoxamine Omeprazole, Esomeprazole	Antiretrovirals[f] Azole antifungals[e] Conivaptan Diltiazem Fluvoxamine Macrolides[g] Nefazodone Verapamil
Inducers[b]	Carbamazepine Montelukast Phenobarbital Phenytoin Tobacco (Smoking)	Carbamazepine Phenytoin Rifampin	Carbamazepine Efavirenz Etravirine Modafinil Nafcillin Oxcarbazepine Phenobarbital Phenytoin Rifampin St. John's Wort Topiramate

[a] Substrate status for melatonin receptor agonists.

[b] Representative lists of strong and moderate potency inhibitors and inducers.

[c] Estrogen-containing agents include oral contraceptives and hormone replacement therapies.

[d] Quinolone antibiotics include ciprofloxacin and ofloxacin.

[e] Azole antifungals include fluconazole, itraconazole, ketoconazole, miconazole, and voriconazole.

[f] Antiretroviral agents include atazanavir, boceprevir, darunavir, fosamprenavir, indinavir, lopinavir, nelfinavir, ritonavir, saquinavir, and telaprevir.

[g] Macrolide antibiotics include clarithromycin, erythromycin, and telithromycin.

None of the four melatonergic agents, at therapeutic concentrations, has been observed to alter CYP enzyme activity; therefore, they are unlikely to alter the clearance of other medications. However, *in vitro* research indicates that melatonin, at supratherapeutic concentrations, induces the activity of CYP3A4. The clinical relevance of melatonin's impact on CYP3A4 activity, if any, is unknown. An additive impact upon psychomotor performance when taken with alcohol has been reported for the melatonin receptor agonists.

LABORATORY INTERFERENCE

None of the four approved agents (i.e., agomelatine, melatonin, ramelteon, tasimelteon) have been shown to have significant laboratory interferences. They also have not been shown to interfere with drug urine toxicology screens or to produce false-positive results.

DOSAGE AND ADMINISTRATION

Extended-release melatonin is administered using a 2 mg bedtime dose. Exogenous melatonin should not be administered to patients with moderate or severe hepatic impairment.

Ramelteon is usually administered at an 8 mg dose 30 minutes prior to bedtime, but there have been reports of efficacy between 4 mg and 64 mg. No dosing adjustment is needed for renal impairment. Ramelteon should be used cautiously in those with moderate hepatic impairment, and use is not recommended in severe hepatic impairment.

Tasimelteon is administered at a 20 mg dose 1 hour prior to bedtime. It may take several weeks or months before the full benefit is achieved. There is no dosing adjustment necessary for renal impairment or mild-to-moderate hepatic impairment; however, it is not recommended for those with severe hepatic impairment.

Agomelatine is administered for depression using a 25 mg bedtime dose. No dose adjustment is needed for renal impairment. Agomelatine should not be administered to patients with even mild hepatic impairment.

INVESTIGATIONAL DRUGS

Other novel melatonin receptor agonists are in various phases of development. TIK-301, formerly known as LY-156735 and PD-6735, was initially developed by Eli Lilly and is currently licensed to Tikvah Therapeutics, where Phase 3 trials were planned. Unfortunately, further testing and development of this product appear to have been abandoned. The pharmacodynamic profile of TIK-301 is similar to agomelatine, a nonselective MT_1/MT_2 agonist (Table 33.15a–3), as well as a $5HT_{2B}$ and $5HT_{2C}$ antagonist. It would theoretically have utility for insomnia, circadian rhythm disorders, and depression.

Piromelatine, also known as Neu-P11, is a nonselective MT_1/MT_2 agonist and $5HT_{1A}/5HT_{1D}$ agonist under development by Neurim Pharmaceuticals for sleep and cognitive function in association with insomnia and Alzheimer Disease. Piromelatine outperformed placebo in a 4-week phase II trial (NCT01489969) for insomnia released in 2013. A subsequent Phase II study specifically in individuals with Alzheimer disease (NCT02615002) was completed in 2019; however, the results of that trial are not yet available. Preclinical research indicates that piromelatine may improve cognitive function in animal models of Alzheimer Disease and prevent insulin resistance and metabolic derangements associated with sleep deprivation.

A number of other melatonin receptor ligands are in various stages of development and/or preclinical testing. Melatonin agonists in development exhibit one of four pharmacodynamic profiles: (1) nonselective MT_1/MT_2 agonists; (2) nonselective MT_1/MT_2 agonists with agonist or antagonist activity at one or more serotonin receptor subtypes; (3) MT_1 selective agonists, and (4) MT_2 selective agonists. By convention, a selective agonist must display at least 50 times greater affinity for one receptor subtype.

Two MT_2 selective partial agonists, UCM765 and UCM924, have demonstrated antinociceptive properties in animal models of neuropathic pain. In other preclinical studies, UCM765 showed anxiolytic effects and promoted non-REM sleep. Production of MT_1 selective ligands has proven more challenging. Thus, progress toward the development of MT_1 selective agonists is lagging.

FURTHER READINGS

An JH, Park MJ, Fava M, et al. Comparisons between successful versus unsuccessful switches from benzodiazepines or zolpidem (Z-drug) to controlled-release melatonin in patients with major depressive disorder and insomnia. *Front Psychiatry*. 2020;11:444.

Boutin JA, Witt-Enderby PA, Sotriffer C, Zlotos DP. Melatonin receptor ligands: a pharmaco-chemical perspective. *J Pineal Res*. 2020;69:e12672.

ClinicalTrials.gov [Internet]. National Library of Medicine (US); 2000–21.

Emet M, Ozcan H, Ozel L, Yayla M, Halici Z, Hacimuftuoglu A. A review of melatonin, its receptors and drugs. *Eurasian J Med*. 2016;48:135–141.

Gahr M, Freudenmann RW, Connemann BJ, Hiemke C, Schonfeldt-Lecuona C. Agomelatine and hepatotoxicity: implications of cumulated data derived from spontaneous reports of adverse drug reactions. *Pharmacopsychiatry*. 2013;46:214–220.

Gooneratne NS, Gehrman P, Gurubhagavatula I, Al-Shehabi E, Marie E, Schwab R. Effectiveness of ramelteon for insomnia symptoms in older adults with obstructive sleep apnea: a randomized placebo-controlled pilot study. *J Clin Sleep Med*. 2010;6:572–580.

Guaiana G, Gupta S, Chiodo D, Davies SJC, Haederle K, Koesters M. Agomelatine versus other antidepressive agents for major depression. *Cochrane Database Syst Rev*. 2013;12:CD008851.

Johnsa JD, Neville MW. Tasimelteon: a melatonin receptor agonist for non-24-hour sleep-wake disorder. *Ann Pharmacother*. 2014;48:1636–1641.

Kryger M, Roth T, Wang-Weigand S, Zhang J. The effects of ramelteon on respiration during sleep in subjects with moderate to severe chronic obstructive pulmonary disease. *Sleep Breath*. 2009;13:79–84.

Kuriyama A, Honda M, Hayashino Y. Ramelteon for the treatment of insomnia in adults: a systematic review and meta-analysis. *Sleep Med*. 2014;15:385–392.

Lewandowska K, Malkiewicz MA, Sieminski M, Cubala WJ, Winklewski PJ, Medrzycka-Dabrowska WA. The role of melatonin and melatonin receptor agonist in the prevention of sleep disturbance and delirium in intensive care unit—a clinical review. *Sleep Med*. 2020;69:127–134.

Lockley SW, Dressman MA, Licamele L, et al. Tasimelteon for non-24-hour sleep–wake disorder in totally blind people (SET and RESET): two multicentre, randomised, double-masked, placebo-controlled phase 3 trials. *Lancet*. 2015;386:1754–1764.

Low TL, Choo FN, Tan SM. The efficacy of melatonin and melatonin agonist in insomnia—an umbrella review. *J Psychiatr Res*. 2020;121:10–23.

Moghaddam HS, Bahmani S, Bayanati S, Mahdavinasa M, Rezaei F, Akhondzadeh S. Efficacy of melatonin as an adjunct in the treatment of acute mania: a double-blind and placebo-controlled trial. *Int Clin Psychopharmacol*. 2020;35:81–88.

Nishimon S, Nishimon M, Nishino S. Tasimelteon for treating non-24-h sleep-wake rhythm disorder. *Exp Opin Pharmacotherap*. 2019;20:1065–1073.

Norris ER, Burke K, Correll JR, et al. A double-blind, randomized, placebo-controlled trial of adjunctive ramelteon for the treatment of insomnia and mood stability in patients with euthymic bipolar disorder. *J Affect Disord*. 2013;144:141–147.

Ogilvie BW, Torres R, Dressman MA, Kramer WG, Baroldi P. Clinical assessment of drug-drug interactions of tasimelteon, a novel dual melatonin receptor agonist. *J Clin Pharmacol*. 2015;55:1004–1011.

Polymeropoulos CM, Mohrman MA, Keefe MS, et al. Efficacy of tasimelteon (HETLIOZ®) in the treatment of jet lag disorder evaluated in an 8-h phase advanced model; a multicenter, randomized, double blind, placebo-controlled trial. *Front Neurol*. 2020;11:611.

Posadzki PP, Bajpai R, Kyaw BM, et al. Melatonin and health: an umbrella review of health outcomes and biological mechanisms of action. *BMC Med*. 2018;16:18.

Rajaratnam SM, Polymeropoulos MH, Fisher DM, et al. Melatonin agonist tasimelteon (VEC-162) for transient insomnia after sleep-time shift: two randomised controlled multicentre trials. *Lancet*. 2009;373:482–491.

Richardson G, Wang-Weigand S. Effects of long-term exposure to ramelteon, a melatonin receptor agonist, on endocrine function in adults with chronic insomnia. *Hum Psychopharmacol*. 2009;24(2):103–111.

Schmidt FM, Lichtblau N, Uribe MM, Kirchherr H, Himmerich H. Agomelatine in breast milk. *Int J Neuropsychopharmacol*. 2013;16:497–499.

Taylor D, Sparshatt A, Varma S, Olofinjana O. Antidepressant efficacy of agomelatine: meta-analysis of published and unpublished studies. *BMJ*. 2014;348:g1888.

Tordjman S, Chokron S, Delorme R, et al. Melatonin: pharmacology, functions and therapeutic benefits. *Curr Neuropharmacol*. 2017;15:434–443.

Voiculescu SE, Zyfouropoulos N, Zahiu CD, Zagrean AM. Role of melatonin in embryo fetal development. *J Med Life*. 2014;7:488–492.

Wright A, Diebold J, Otal J, et al. The effect of melatonin on benzodiazepine discontinuation and sleep quality in adults attempting to discontinue benzodiazepines: a systematic review and meta-analysis. *Drugs Aging*. 2015;32:1009–1018.

33.15b Thyroid Hormones

RENA COOPER-KAZAZ, M.D., AND BERNARD LERER, M.D.

Neuroscience-Based Nomenclature (NbN)	
Levothyroxine, liothyronine	
Pharmacologic Domain	Thyroid hormones
Mechanism of Action	Thyroid hormone receptor agonists (multiorgan systems)

Various hormonal systems have been implicated in the biology of the stress response and in the pathophysiology of affective disorders. The hypothalamic–pituitary–adrenal (HPA) axis has been the focus of most attention, but the thyroid axis has also been established as an important endocrine system, particularly in depression.

Thyroid hormones have long been of interest to those who study and treat mood disorders, mainly unipolar major depression and bipolar disorder. There have been two major areas of research. The first is whether thyroid hormones are implicated in the etiology of mood disorders and whether there is a specific abnormality of the thyroid axis in primary mood disorders. The second, following the first, is whether thyroid hormones may be useful in the treatment of unipolar or bipolar affective disorders.

For more than a century, depressive symptoms have been known to be one of the most prominent features of clinical thyroid disease, particularly hypothyroidism. However, a recent meta-analysis suggested that the association of hypothyroidism with depression may be less definitive than previously thought, especially for subclinical hypothyroidism, and may be more prominent in females than in males. Clinical hyperthyroidism has a much more heterogeneous psychiatric presentation with a wider variety of symptoms reported. Earlier studies and clinical observations noted that when the thyroid deficiency or excess was corrected, depressive or other psychiatric symptoms frequently, but not invariably, resolved. As a result of the inverse relationship between thyroid function and depression observed in endocrine patients, it was suggested that thyroid hormones—particularly thyroid hormone deficiency—might be a factor in the etiology of unipolar and bipolar mood disorders.

Although thyroid hormone deficiency has been excluded as a major contributor to the biologic basis of primary mood disorders, particularly depression, its potential role in the cascade of biologic events that leads to depressive and bipolar disorders remains to be clarified. Thyroid hormones have extensive effects on various cellular mechanisms and may still be implicated in the biology of these disorders.

USE OF THYROID HORMONES IN TREATMENT

The early link established between abnormal mood symptomatology and thyroid deficiency in endocrine patients prompted an investigation into the use of thyroid hormones to treat depression. Earlier studies yielded ambiguous findings, which may have been contributed to methodologic limitations, especially a lack of controlled designs and heterogeneous groups of patients. In the 1960s, synthetic forms of thyroxine (levothyroxine, T_4) and triiodothyronine (liothyronine, the synthetic levorotatory isomer of triiodothyronine, T_3), were developed and replaced desiccated thyroid as a primary replacement therapy for thyroid disease. This allowed a more refined approach to examining the efficacy of thyroid hormones in the treatment of mood disorders, and the therapeutic use of thyroid hormones in the treatment of depression became the subject of study. Although thyroid hormones have been used to treat a range of psychiatric illnesses, systematic research data have been collected primarily on unipolar mood disorders and, to a lesser extent, bipolar mood disorders.

THYROID HORMONE BIOCHEMISTRY

The source of thyroid hormones in the body is the thyroid gland. The self-regulated hypothalamic–pituitary–thyroid axis (HPT) is responsible for the homeostasis of thyroid function. The two main active hormones, T_4 and T_3 are tyrosine-based molecules produced by the thyroid gland and conjugated to iodide, an essential dietary element. T_4 and T_3 are stored in the follicular lumen of the thyroid gland bound to the iodinated protein, thyroglobulin, and are released into the circulation upon stimulation (T_4—80%, T_3—20%). The circulating lipophilic hormones are bound to transport proteins leaving approximately 0.03% T_4 and 0.3% T_3 unbound and active (free T_4/T_3). T_4 is derived almost entirely from the thyroid gland, whereas the main source of T_3 is the monodeiodination of T_4 in extrathyroidal tissue by deiodinase enzymes.

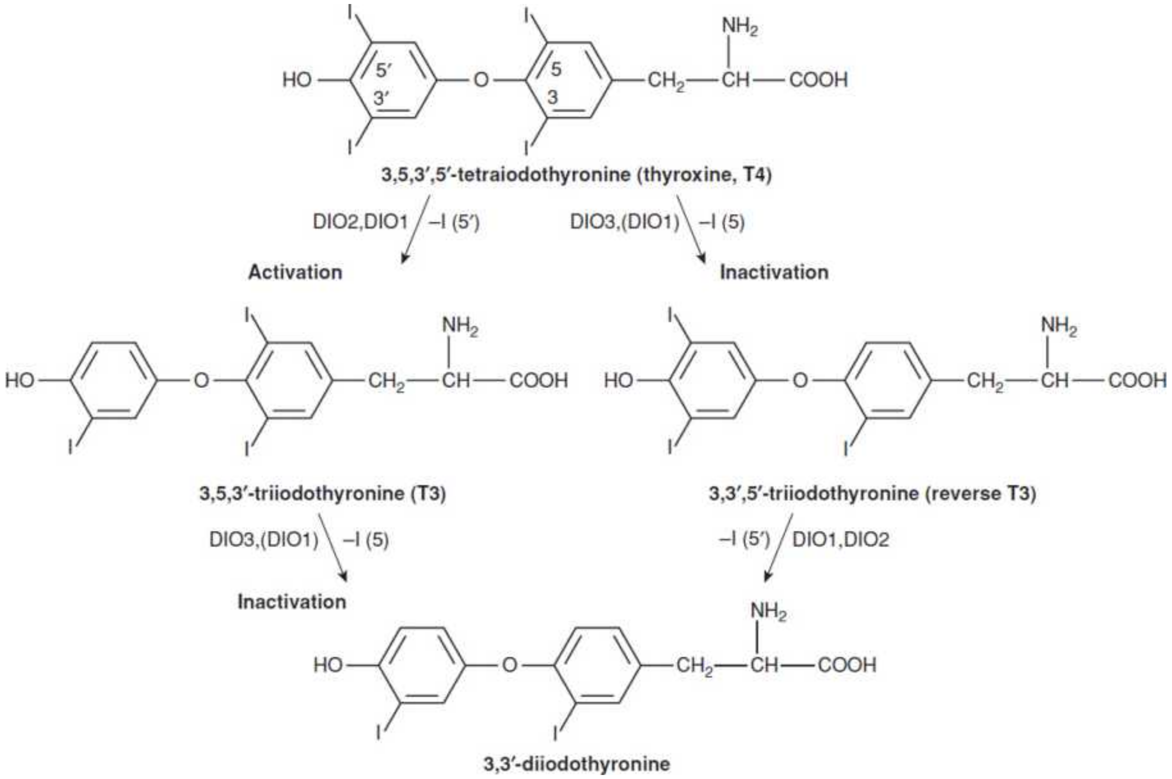
HO
5′
3′
O
5
3
CH2
CH
COOH
NH2
I
3,5,3′,5′-tetraiodothyronine (thyroxine, T4)
DIO2,DIO1
–I (5′)
Activation
DIO3,(DIO1)
–I (5)
Inactivation
3,5,3′-triiodothyronine (T3)
3,3′,5′-triiodothyronine (reverse T3)
DIO3,(DIO1)
–I (5)
Inactivation
–I (5′)
DIO1,DIO2
3,3′-diiodothyronine

FIGURE 33.15b–1. Metabolism of thyroid hormones by deiodinase enzymes. (Reprinted from Guo T-W, St. Clair D, He L. The deiodinase type 2 (DIO2) gene and mental retardation in iodine deficiency. In: Preedy VR, Burrow GN, Watson R. *Comprehensive Handbook of Iodine: Nutritional, Biochemical, Pathological and Therapeutic Aspects*. Academic Press; 2009:635–641, with permission from Elsevier.)

There are three types of deiodinases: type I, II, and III (Fig. 33.15b–1). Types I (DIO1) and II (DIO2), located in the liver, kidneys, muscles, and thyroid gland, convert T_4 to the active form, T_3, via deiodination of T_4's outer ring. Type III (DIO3) deiodinase is located in the CNS, skin, and placenta. Deiodinase 1 (*DIO1*) and *DIO3* inactivate both T_4 and T_3 and terminate thyroid hormone action via deiodination of T_4's inner molecular ring.

Polymorphisms in the three deiodinase genes exist, and several were found to be associated with (1) autoimmune thyroid diseases, (2) symptoms of anxiety and depression after ischemic stroke, (3) executive function and personality traits, and (4) cardiac-related mortality after acute myocardial infarction.

Other thyroid hormone products include reverse T_3, 3-diiodothyronine, 3-monoiodothyronines, and the acetate metabolites of T_4 and T_3—tetrac and triac, respectively. These other thyroid hormone products appear in very small amounts, and their physiologic significance is limited.

The secretion of thyroid hormones by the thyroid gland is regulated by the HPT axis. This involves the classic negative-feedback regulatory mechanism, which applies to most of the major endocrine systems. The cells of the hypothalamus secrete thyrotropin-releasing hormone (TRH), which stimulates the secretion of thyrotropin (thyroid-stimulating hormone [TSH]) from the cells of the anterior pituitary via the portal venous blood system, which connects the hypothalamus and the pituitary gland. The TSH released from the pituitary then acts on the thyroid gland to stimulate the production of thyroid hormones, including T_4 and T_3. A negative feedback system operates at the levels of both the pituitary and the hypothalamus to modulate circulating levels of T_4 and T_3 and thereby regulate the activity of the thyroid axis.

The physiologic action of T_4 is largely dependent on its extrathyroidal conversion to T_3, which then acts at T_3 receptors in various tissues to initiate the physiologic actions of thyroid hormones. T_4 thus fulfills the definition of a prohormone. In the brain to a greater extent and pituitary to a lesser extent, the main source of thyroid hormone for action at the receptor is T_4, which is converted locally within the neuron or cell to T_3 before exerting its hormonal action. T_3 receptors are located within the nucleus and are members of the nuclear receptor superfamily. T_3 receptors (TR) are encoded by two different genes and exist in a number of different isoforms, predominantly TRα1 and TRβ, that vary among tissues. Nuclear receptors such as T_3 influence gene expression by binding to specific DNA elements, which, in the case of T_3, are called thyroid response elements (TREs) and are located in the promoter region of T_3-responsive genes. Thus, the physiologic effects of thyroid hormone are brought about by the changes it causes in the expression of T_3-responsive genes in target tissues.

PHARMACOLOGIC ACTION

Pharmacokinetics

T4 and T3 are both relatively small molecules that are easily absorbed across the gastrointestinal (GI) tract. Both hormones are strongly protein-bound and are widely distributed in tissues. There is an active mechanism for the transportation of thyroid hormones across the blood–brain barrier. There is also a transport mechanism at the choroid plexus to ensure the wide distribution of thyroid hormones within the CNS. The deiodinase enzymes, particularly DIO3, convert T_3 to smaller molecules, including reverse T_3, which are then excreted in both urine and feces. T4 has a long half-life of approximately 7 days, whereas T3 has a half-life of approximately 8 to 12 hours.

Pharmacodynamics

Thyroid hormones have widespread metabolic effects in virtually every organ or system in the body, including the CNS, heart, lungs, bones, muscles, GI, reproductive system, and more. They activate genes that increase metabolic rate and thermogenesis. They are also critically important in pre- and postnatal brain development and influence key neurodevelopmental processes such as neurogenesis and neural migration. Thyroid hormones continue to play an active role in the mature brain in processes such as adult neurogenesis; therefore, their levels remain highly regulated within narrow limits to ensure optimal neuronal functioning. Indeed, hypothyroidism is a known cause of reversible cognitive impairment and even dementia syndrome.

THYROID FUNCTION IN PRIMARY AFFECTIVE DISORDERS

The vast majority of patients with primary major depression are euthyroid. A large number of studies evaluating baseline thyroid function tests have produced inconsistent findings. The most consistent findings to emerge are higher measures of T_4 or free T_4 in the depressed state as compared with recovery or healthy controls, with much more inconsistent findings for measures of circulating T_3 or free T_3. Up to one-third of depressed patients may have a blunted TSH response to TRH, which often normalizes with treatment. However, due to low sensitivity and moderate specificity for major depression, the TRH test is not diagnostic for depression.

In a recent Mendelian randomization study, no associations were found between genetically predicted TSH and fT_4 levels and major depressive disorder (MDD) risk, nor MDD subtypes and minor depressive symptoms. A slight increase in free T_4 levels was nominally associated with a decrease in the overall bipolar disorder risk.

Suicide attempters with MDD had more pathologic psychiatric rating scale scores (depression, anxiety, and psychosis) along with higher serum levels of TSH, antithyroglobulin antibodies (TgAb), and thyroid peroxidase antibodies (TPOAb) compared to nonsuicide attempters. This suggests the importance of regular assessment of thyroid function parameters for suicide prevention and possible biomarkers of suicide risk in MDD.

Indeed, evidence points to a shared pathophysiology of immune dysfunction and mood disorders. High rates of thyroid dysfunction have been found in patients with bipolar disorder compared to the general population. Thyroid autoimmunity may, in fact, be an independent risk factor for bipolar disorder with no clear association with lithium treatment.

Neither hypothyroidism nor hyperthyroidism was significantly associated with rapid-cycling bipolar disorder; however, TPOAb were significantly related to rapid-cycling bipolar disorder. This finding warrants further investigation into a possible biomarker, an underlying mechanism, and even treatment for rapid-cycling bipolar disorder.

There is limited data on baseline thyroid function and the relationship to antidepressant response. In some studies, higher mean T_4 levels and in other studies, higher mean TSH levels have been associated with an antidepressant response. A reduced TSH response to TRH has also been associated with treatment response. There are preliminary data that polymorphisms of the deiodinase enzymes, which are responsible for the extrathyroidal conversion of T_4 to T_3 (type I and type II deiodinase) and for the inactivation of T_3 by conversion to reverse T_3 (type III deiodinase), may be associated with antidepressant response in euthyroid-depressed patients.

AFFECTIVE DISORDERS IN PRIMARY THYROID DISORDERS

Approximately 15% of patients taking T4 (LT_4) for hypothyroidism report impaired quality of life and depressive symptoms compared to controls and are dissatisfied with their treatment. Although T4 is a good treatment for hypothyroidism, the prevalence of anxiety and depression among T4-

treated hypothyroid women was higher compared to women without hypothyroidism. Thus, psychiatric symptoms can remain despite adequate hormone replacement and should be a focus of interest for treating physicians. Indeed, when T4 doses were increased in older adults undergoing thyroid hormone replacement, their depressive symptoms improved without significant hyperthyroid symptoms or signs. These findings suggest the potential for varying the treatment target for primary hypothyroidism to include mood status. The setting of TSH cutoff values for hypothyroid patients should, therefore, also include associated symptoms such as mood. Some suggest a TSH cutoff value of 2.5 mIU/L to be optimal.

Recent studies showed promising advantages for the combined treatment of hypothyroidism with T4 (T_4) and T3 (T_3) in symptomatic patients. Further investigation is needed to assess the benefits and safety of this combination. Finally, adherence to T4 among hypothyroidism patients was associated with significantly lower rates of comorbid diagnoses, including depression and bipolar disease.

Subclinical hypothyroidism (SCH), defined as elevated TSH levels with normal circulating T_4 and T_3 levels, occurs in a minority of patients, at most 10% to 15%, with primary mood disorders. SCH is more common with advanced age and in women. SCH is associated with a lower response rate to antidepressants, and there is recent evidence that even the presence of thyroid autoimmunity as measured by positive anti-TPOAb with normal TSH may be associated with lower rates of antidepressant response. There are no data to guide clinical decision-making about whether thyroid replacement would enhance antidepressant response, so the decision on thyroid replacement should be consistent with current guidelines for the treatment of SCH. Several clinical guidelines recommend treatment of SCH with thyroid hormones for adults with TSH levels >10 mIU/L and for people with lower TSH values who are young, symptomatic, or have specific indications for treatment. However, most recent studies and guidelines raise doubts as to the benefits and risks associated with treatment. This controversy remains unresolved.

CLINICAL HORMONE STUDIES

All peptides and hormones of the HPT axis have been studied for the treatment of mood disorders. These include TRH and TSH, as well as the peripheral hormones T_4 and T_3. The best evidence for therapeutic efficacy is for the use of T3 in unipolar major depression; there are fewer data on the role of T4 in the treatment of depression. Preliminary data suggest that T4 may have a role as a mood stabilizer in bipolar disorder.

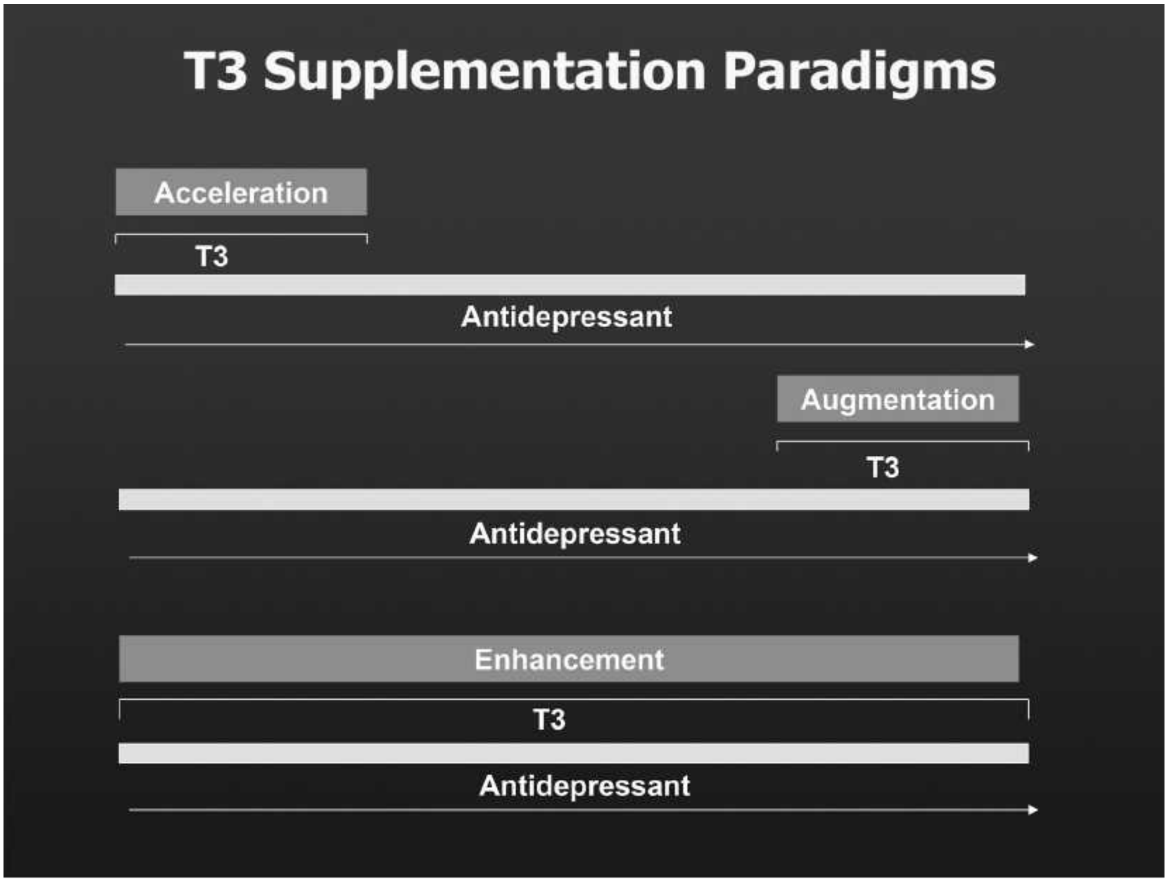
T3 Supplementation Paradigms
Acceleration
T3
Antidepressant
Augmentation
T3
Antidepressant
Enhancement
T3
Antidepressant

FIGURE 33.15b–2. Thyroid hormone supplementation strategies in major depression.

In major depression, thyroid hormones have been used in several ways (Fig. 33.15b–2). In very early studies, they were used as *monotherapy* for the treatment of depression with uncertain benefits. More recent work has focused on the *supplementation* of antidepressants (tricyclics or selective serotonin reuptake inhibitors [SSRIs]) with thyroid hormone. The first type of supplementation is *acceleration,* in which thyroid hormones have been used at the outset of an antidepressant trial, primarily with tricyclic antidepressants, to accelerate the response to antidepressant treatment. A second type of supplementation is the addition of thyroid hormones to antidepressants in patients who have failed to have a satisfactory response to an antidepressant trial, termed *augmentation*. A third supplementation approach is to initiate thyroid hormone from the start of an antidepressant trial to achieve both *acceleration* and *enhancement* of antidepressant response in depressed patients starting antidepressant therapy. Finally, thyroid hormones, particularly thyroxine, have been used as mood stabilizers in bipolar disorder, especially in rapid-cycling forms.

Thyrotropin-Releasing Hormone

TRH is a peptide released from the hypothalamus that regulates thyroid hormone secretion by the thyroid gland through its effects on pituitary TSH release. In addition, TRH has direct behavioral effects on the brain. TRH also has effects on cardiorespiratory, GI, and neurologic functions. Due to its widespread effects, TRH was tested as an antidepressant both as monotherapy and together with electroconvulsive therapy (ECT). TRH was administered both intravenously (single dose or repeated) and orally, but only a transient antidepressant effect was demonstrated. In conclusion, the administration of TRH in its various forms had little if any antidepressant effect, and in any case, this may be a nonspecific activating effect of the peptide rather than a direct effect on the HPT axis. When TRH was administered intravenously to eight patients before ECT, this yielded greater arousal and improved cognitive functioning compared with placebo.

Thyrotropin

TSH has also received limited attention as an antidepressant. The rationale for its use is based on its stimulation of the thyroid gland and, thus, would achieve an effect similar to the administration of peripheral thyroid hormones for the treatment of depression. One study has examined the antidepressant effect of TSH by administering 10 IU of TSH intravenously to 20 depressed women 1 day before beginning an antidepressant trial with the tricyclic imipramine. When compared to a placebo control group, the TSH-treated patients had a rapid antidepressant response. These data, while of interest, require replication with larger samples, a more current methodology for clinical trials, and the use of SSRIs. Such study designs would allow a more appropriate evaluation of the efficacy of TSH in accelerating the response to antidepressants.

T3

T3 is the most widely studied thyroid hormone for the treatment of depression. In early studies, it was used as *monotherapy* for the treatment of depressed patients. The data from these studies were largely inconclusive, as they involved small patient samples, inadequate clinical trial designs by current standards, and the use of heterogeneous patient groups who, by today's diagnostic criteria, would not necessarily be diagnosed with major depression. There have been no well-designed studies of T3 monotherapy to date; therefore, its use as a single treatment for depression has not gained any clinical ground.

T3 was used in earlier studies to *accelerate* the response to tricyclic antidepressants. T3 was administered at the outset of a tricyclic antidepressant trial, and a shorter lag in onset of therapeutic effect as compared with placebo was noted in some but not all studies. This acceleration effect was observed particularly in women as compared to men. All these studies had major methodologic flaws, including small numbers of patients, poorly defined patient groups, and relatively brief duration of treatment. These earlier studies involved tricyclic antidepressants only, usually at suboptimal doses by current standards, as SSRIs were not available. Nonetheless, a meta-analysis of these earlier studies showed a significant acceleration of antidepressant response during the initial 4 weeks of tricyclic treatment for T3 over placebo.

Table 33.15b–1.
T3 Augmentation of Antidepressants

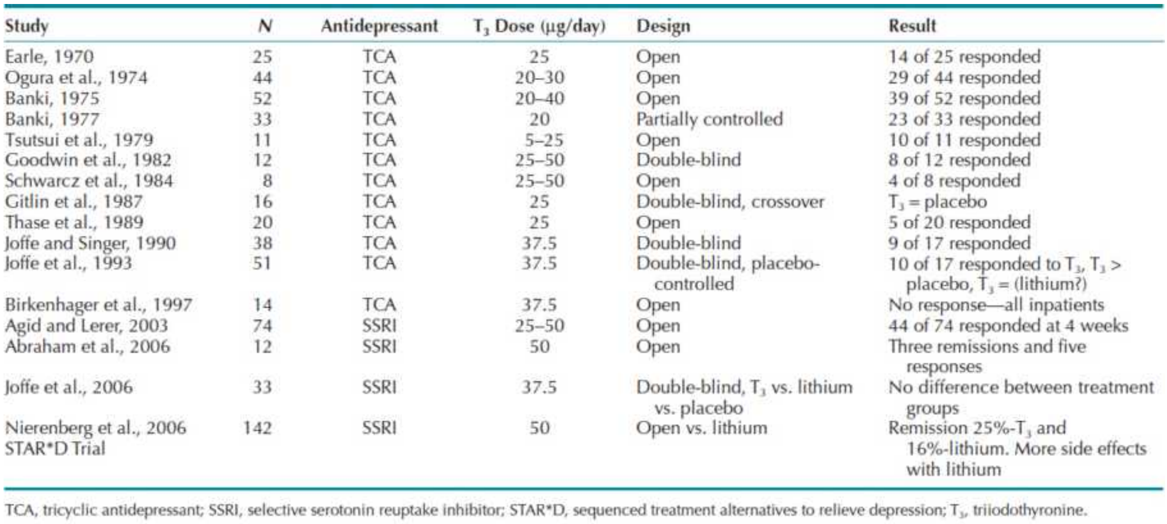

Study	N	Antidepressant	T_3 Dose (μg/day)	Design	Result
Earle, 1970	25	TCA	25	Open	14 of 25 responded
Ogura et al., 1974	44	TCA	20–30	Open	29 of 44 responded
Banki, 1975	52	TCA	20–40	Open	39 of 52 responded
Banki, 1977	33	TCA	20	Partially controlled	23 of 33 responded
Tsutsui et al., 1979	11	TCA	5–25	Open	10 of 11 responded
Goodwin et al., 1982	12	TCA	25–50	Double-blind	8 of 12 responded
Schwarcz et al., 1984	8	TCA	25–50	Open	4 of 8 responded
Gitlin et al., 1987	16	TCA	25	Double-blind, crossover	T_3 = placebo
Thase et al., 1989	20	TCA	25	Open	5 of 20 responded
Joffe and Singer, 1990	38	TCA	37.5	Double-blind	9 of 17 responded
Joffe et al., 1993	51	TCA	37.5	Double-blind, placebo-controlled	10 of 17 responded to T_3, T_3 > placebo, T_3 = (lithium?)
Birkenhager et al., 1997	14	TCA	37.5	Open	No response—all inpatients
Agid and Lerer, 2003	74	SSRI	25–50	Open	44 of 74 responded at 4 weeks
Abraham et al., 2006	12	SSRI	50	Open	Three remissions and five responses
Joffe et al., 2006	33	SSRI	37.5	Double-blind, T_3 vs. lithium vs. placebo	No difference between treatment groups
Nierenberg et al., 2006 STAR*D Trial	142	SSRI	50	Open vs. lithium	Remission 25%-T_3 and 16%-lithium. More side effects with lithium

TCA, tricyclic antidepressant; SSRI, selective serotonin reuptake inhibitor; STAR*D, sequenced treatment alternatives to relieve depression; T_3, triiodothyronine.

A third approach to the use of T3 in depressive disorders has been to *augment* response to antidepressants in those who failed to respond to an antidepressant trial. These studies are summarized in Table 33.15b–1. Open and controlled studies show that approximately 50% to 60% of patients who do not respond to an antidepressant trial with tricyclics will respond within 2 to 3 weeks after the addition of T3. Notwithstanding the limitations of the studies of T3 augmentation to date, a meta-analysis of these studies concluded that T3 may increase response rates and decrease the severity of depression scores in patients who are refractory to tricyclic antidepressant treatment. Patients with T3 augmentation were approximately twice as likely to respond as were controls.

In the past few years, there has been emerging data on the use of T3 to augment SSRIs in antidepressant nonresponders. The findings with the SSRIs are consistent with those for the tricyclics and are included in Table 33.15b–1. Both open and controlled studies have generally been positive and indicate that T3 may be an effective augmentation agent for SSRI nonresponders. In data from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial, T3 augmentation had comparable remission rates to other augmentation options and a more favorable adverse event profile and dropout rate.

There are also preliminary data that if T3 is administered at the outset of a course of ECT, a reduced number of treatments are required to achieve a therapeutic response, and there may be a greater sparing of cognitive difficulties.

In recent studies, there has been another use of T3; administration from the outset of an SSRI trial to determine whether this will reduce delay in therapeutic onset (*acceleration*) and increase rates of treatment response (*enhancement*) in patients not necessarily refractory to antidepressant treatment. There have been four randomized controlled trials which are summarized in Table 33.15b–2. One study showed clearly significant results for enhancement of sertraline treatment by T_3 (Cooper), and one showed moderately positive results for enhancement of several different SSRIs (Posternak). Two trials showed no difference in the effect of T_3 to enhance paroxetine (Appelhof) and sertraline (Garlow), respectively. A meta-analysis of these four trials did not show a significant effect of T3 administered from the outset of treatment to enhance the antidepressant effect of SSRIs. However, the response may be specific to subgroups of patients and may also be influenced by genetic polymorphism of the *DIO1* receptor gene. The dose of SSRI may also be a factor, with patients receiving a lower SSRI dose more likely to show enhancement of response with T3 supplementation from the onset of treatment.

Table 33.15b–2.
T3 Enhancement of SSRI

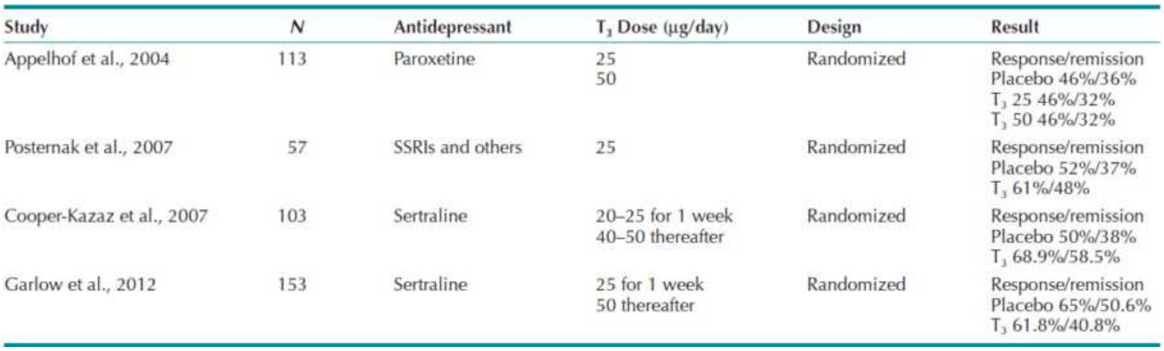

Study	N	Antidepressant	T_3 Dose (µg/day)	Design	Result
Appelhof et al., 2004	113	Paroxetine	25 50	Randomized	Response/remission Placebo 46%/36% T_3 25 46%/32% T_3 50 46%/32%
Posternak et al., 2007	57	SSRIs and others	25	Randomized	Response/remission Placebo 52%/37% T_3 61%/48%
Cooper-Kazaz et al., 2007	103	Sertraline	20–25 for 1 week 40–50 thereafter	Randomized	Response/remission Placebo 50%/38% T_3 68.9%/58.5%
Garlow et al., 2012	153	Sertraline	25 for 1 week 50 thereafter	Randomized	Response/remission Placebo 65%/50.6% T_3 61.8%/40.8%

Table 33.15b–3.
Antidepressant Augmentation With T4

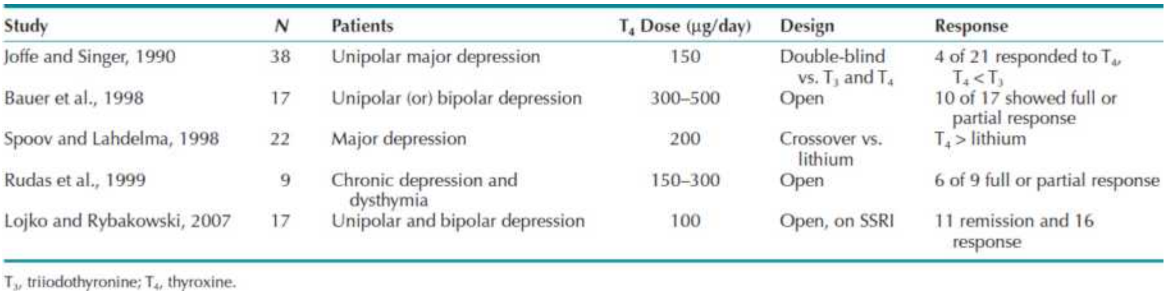

Study	N	Patients	T_4 Dose (µg/day)	Design	Response
Joffe and Singer, 1990	38	Unipolar major depression	150	Double-blind vs. T_3 and T_4	4 of 21 responded to T_4, $T_4 < T_3$
Bauer et al., 1998	17	Unipolar (or) bipolar depression	300–500	Open	10 of 17 showed full or partial response
Spoov and Lahdelma, 1998	22	Major depression	200	Crossover vs. lithium	T_4 > lithium
Rudas et al., 1999	9	Chronic depression and dysthymia	150–300	Open	6 of 9 full or partial response
Lojko and Rybakowski, 2007	17	Unipolar and bipolar depression	100	Open, on SSRI	11 remission and 16 response

T_3, triiodothyronine; T_4, thyroxine.

In summary, there is evidence suggesting that T3 augmentation may be effective in treatment-resistant patients. Larger studies are required to address several issues. First, further studies of T3 augmentation of SSRIs with large-scale, well-controlled designs are needed. Second, issues such as dosing and duration of treatment need to be addressed to provide solid data to guide the use of T3 in a broad-based safe clinical approach to treatment-resistant depression.

T4

This thyroid hormone has been used in two ways in the treatment of mood disorders. First, it has been used as an augmentation agent for the treatment of tricyclic-refractory depression and, second, as a mood stabilizer for rapid-cycling bipolar disorder.

Data on the use of T4 in refractory depression are summarized in Table 33.15b–3. These studies have generally involved open designs and various diagnoses, including both unipolar and bipolar subjects in one study and chronic depression or dysthymic patients in another study. All involve small sample designs. In each of these studies, T4 showed an augmentation effect in antidepressant nonresponders.

T4 has also been used in high doses as a mood stabilizer. There are now five case studies or case series, four involving rapid-cycling bipolar patients and one involving nonrapid-cycling bipolar patients. The total patient sample in the 5 studies consists of 30 patients. These studies are summarized in Table 33.15b–4. Doses of T4 ranged from 150 to 500 µg/day. Of the 30 patients, 27 were reported to respond to T4 or augmentation of mood stabilizers. Clearly, these data are of considerable interest, especially because rapid-cycling bipolar patients represent a substantial therapeutic challenge. A large-scale systematic controlled trial would clarify the efficacy of T4 in bipolar disorder and its role relative to other mood-stabilizing interventions.

Therapeutic Indications

None of the peptides or hormones of the HPT axis have received formal approval by regulatory authorities in North America or worldwide for use in the treatment of depression. Evidence suggests that these hormones may have clinical use in the treatment of different types of mood disorders. Indeed, the use of thyroid hormones (mainly T_3) appears in various clinical practice guidelines for the treatment of major depressive depression. T3 may be useful as an augmentation agent in refractory depression and may also have some use in accelerating antidepressant response. The use of T4 in refractory depression requires further study, and T4 may be an adjunct to mood-stabilizer treatment, particularly in rapid-cycling patients. For thyroid hormones to have a broad clinical application, empirical data are required to clarify their efficacy and to determine their role relative to other options for these various types of mood disorders. Until such data are obtained, T3 remains one of many possibilities to consider for use in patients who do not respond to their first or subsequent antidepressant trial. Moreover, T4 may be considered an option for the refractory bipolar patient, especially when conventional mood stabilizers have proven to be ineffective or only partially effective. Finally, the presence of depressive symptoms in SCH patients could warrant the initiation of T4.

Precautions and Adverse Reactions

T3 administered in the usual doses for augmentation treatment should not lead to symptoms of hyperthyroidism or thyrotoxicosis. A dosage of 25 µg/day is below the replacement dosage for this hormone; thus, symptoms of thyroid toxicity are extremely rare. A dose of 50 µg/day is closer to the daily T_3 production range; thus, symptoms of thyroid excess could occur. We recommend continuous clinical monitoring and TSH assessment while administering T3. It should be noted that there are no well-demonstrated dose–response studies for T3; therefore, doses in excess of 25 µg/day should be used with caution. Long-term administration of T3 to euthyroid individuals is not advisable; in depressed patients who have responded, the treatment should be tapered and stopped by 12 weeks. T3 is generally well tolerated when administered to depressed patients. A transient headache is the most commonly reported adverse effect, and it usually subsides after 1 or 2 days, either spontaneously or with the administration of mild analgesics. Allergic reactions are extremely uncommon. T3 should be used with great caution in older adults and in patients with various medical disorders, including endocrine and cardiovascular disease, and should be avoided in pregnancy.

Table 33.15b–4.
T4 as Adjunctive Mood Stabilizer in Bipolar Disorder

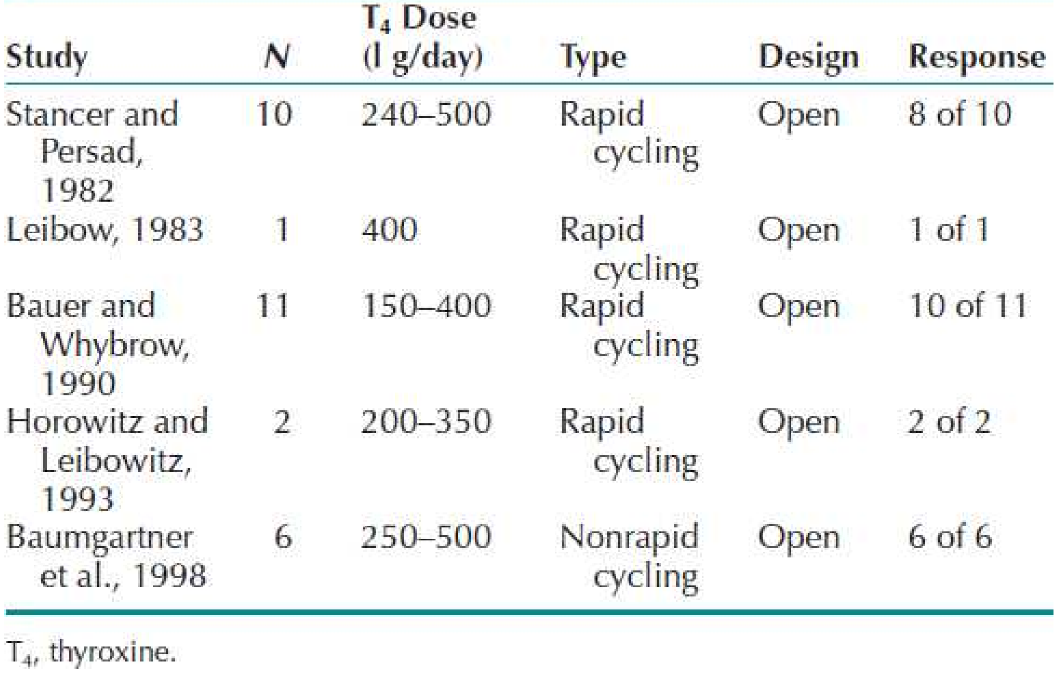

Study	N	T_4 Dose (l g/day)	Type	Design	Response
Stancer and Persad, 1982	10	240–500	Rapid cycling	Open	8 of 10
Leibow, 1983	1	400	Rapid cycling	Open	1 of 1
Bauer and Whybrow, 1990	11	150–400	Rapid cycling	Open	10 of 11
Horowitz and Leibowitz, 1993	2	200–350	Rapid cycling	Open	2 of 2
Baumgartner et al., 1998	6	250–500	Nonrapid cycling	Open	6 of 6

T_4, thyroxine.

The same precautions as noted for T3 apply to T4, especially since this compound is commonly used in supraphysiologic doses. A recent study has demonstrated that depressed patients are able to tolerate supraphysiologic doses of T4 much better than healthy controls. Long-term thyroid hormone treatment may be associated with bone demineralization, resulting in osteoporosis in women. Although this association is still controversial, one should approach the long-term use of thyroid hormones with caution.

Dosage and Administration

T3 is generally administered as a single dose of 25 to 50 mcg in the morning to avoid insomnia. There is a good rationale for divided administration of T3 because of its short half-life, but a single daily dose is commonly used to improve compliance. T4 has been administered in dosages that produce a circulating T4 concentration, which exceeds or is at the upper range of normal for patients with bipolar mood disorder. These dosage recommendations apply to healthy adults and should be modified for older patients or those with significant medical illnesses.

There are no data to support the efficacy of thyroid hormone treatment for depression in children.

FURTHER READINGS

Altshuler LL, Bauer M, Frye MA, et al. Does thyroid supplementation accelerate tricyclic antidepressant response? A review and meta-analysis of the literature. *Am J Psychiatry*. 2001;158:1617–1622.

Appelhof BC, Brouwer JP, van Dyck R, et al. Triiodothyronine addition to paroxetine in the treatment of major depressive disorder. *J Clin Endocrinol Metab*. 2004;89:6271–6276.

Azizi F, Amouzegar A, Mehran L, Abdi H. LT4 and slow release T3 combination: optimum therapy for hypothyroidism? *Int J Endocrinol Metab*. 2020;18(2):e100870.

Bauer M, Baur H, Berghöfer A, et al. Effects of supraphysiological thyroxine administration in healthy controls and patients with depressive disorders. *J Affect Disord*. 2002;68:285–294.

Baumgartner A. Thyroxine and the treatment of affective disorders: an overview of the results of basic and clinical research. *Int J Neuropsychopharmacol*. 2000;3:149–165.

Bekkering GE, Agoritsas T, Lytvyn L, et al. Thyroid hormones treatment for subclinical hypothyroidism: a clinical practice guideline. *BMJ*. 2019;365:l2006.

Bevins EA, Peters J, Léger GC. The diagnosis and management of reversible dementia syndromes. *Curr Treat Options Neurol*. 2021;23:3.

Bode H, Ivens B, Bschor T, Schwarzer G, Henssler J, Baethge C. Association of hypothyroidism and clinical depression: a systematic review and meta-analysis. *JAMA Psychiatry*. 2021;78:1375–1383.

Cooper-Kazaz R, Apter JT, Cohen R, et al. Combined treatment with sertraline and liothyronine in major depression: a randomized, double-blind, placebo-controlled trial. *Arch Gen Psychiatry*. 2007;64:679–688.

Cooper-Kazaz R, van der Deure WM, Medici M, et al. Preliminary evidence that a functional polymorphism in type I deiodinase is associated with enhanced potentiation of the antidepressant effect of sertraline by triiodothyronine. *J Affect Disord*. 2009;116:113–116.

Day GS. Reversible dementias. *Continuum (Minneap Minn)*. 2019;25(1):234–253.

Eller T, Metsküla K, Talja I, Maron E, Uibo R, Vasar V. Thyroid autoimmunity and treatment response to escitalopram in major depression. *Nord J Psychiatry*. 2010;64:253–257.

Gabriel FC, de Melo DO, Fráguas R, Leite-Santos NC, Mantovani da Silva RA, Ribeiro E. Pharmacological treatment of depression: a systematic review comparing clinical practice guideline recommendation. *PLoS One* 2020;15(4):e0231700.

Garlow SJ, Dunlop BW, Ninan PT, Nemeroff CB. The combination of triiodothyronine (T3) and sertraline is not superior to sertraline monotherapy in the treatment of major depressive disorder. *J Psychiatr Res*. 2012;46:1406–1413.

Hepp Z, Lage MJ, Espaillat R, Gossain VV. The association between adherence to levothyroxine and economic and clinical outcomes in patients with hypothyroidism in the US. *J Med Econ*. 2018;21(9):912–919.

Hoon Moon J, Han JW, Jung Oh T, et al. Effect of increased levothyroxine dose on depressive mood in older adults undergoing thyroid hormone replacement therapy. *Clin Endocrinol (Oxf)*. 2020;93(2):196–203.

Joffe RT, Sokolov STH, Levitt AJ. Lithium and triiodothyronine augmentation of antidepressants. *Can J Psychiatry*. 2006;51:791–793.

Kuś A, Kjaergaard AD, Marouli E, et al. Thyroid function and mood disorders: a mendelian randomization study. *Thyroid*. 2021;31:1171–1181.

Łojko D, Rybakowski JK. L-thyroxine augmentation of serotonergic antidepressants in female patients with refractory depression. *J Affect Disord*. 2007;103(1–3):253–256.

Nierenberg AA, Fava M, Trivedi MH, et al. A comparison of lithium and T(3) augmentation following two failed medication treatments for depression: a STAR*D report. *Am J Psychiatry*. 2006;163:1519–1530.

Peeters RP, Brito JP. Subclinical hypothyroidism: to treat or not to treat? *Eur J Endocrinol*. 2020;183(6):D15–D24.

Peterson SJ, Cappola AR, Castro MR, et al. An online survey of hypothyroid patients demonstrates prominent dissatisfaction. *Thyroid*. 2018;28(6):707–721.

Pirahanchi Y, Tariq MA, Jialal I. In: StatPearls [Internet]. Treasure Island (FL). *Physiology, Thyroid*. StatPearls Publishing; 2023.

Posternak M, Novak S, Stern R, et al. A pilot effectiveness study: placebo-controlled trial of adjunctive L-triiodothyronine (T3) used to accelerate and potentiate the antidepressant response. *Int J Neuropsychopharmacol*. 2008;11(1):15–25.

Romero-Gómez B, Guerrero-Alonso P, Carmona-Torres JM, Notario-Pacheco B, Cobo-Cuenca AI. Mood disorders in levothyroxine-treated hypothyroid women. *Int J Environ Res Public Health*. 2019;16(23):4776.

Talaei A, Rafee N, Rafei F, Chehrei A. TSH cut off point based on depression in hypothyroid patients. *BMC Psychiatry*. 2017;17(1):327.

van de Ven AC, Muntjewerff JW, Netea-Maier RT, et al. Association between thyroid function, thyroid autoimmunity, and state and trait factors of depression. *Acta Psychiatr Scand*. 2012;126:377–384.

Wiersinga WM. Paradigm shifts in thyroid hormone replacement therapies for hypothyroidism. *Nat Rev Endocrinol*. 2014;10:164–174.

Wiersinga WM, Duntas L, Fadeyev V, Nygaard B, Vanderpump MPJ. 2012 ETA Guidelines: the use of L-T4 + L-T3 in the treatment of hypothyroidism. *Eur Thyroid J*. 2012;1:55–71.

33.16 Combination Pharmacotherapy

CHARLES DEBATTISTA, D.M.H., M.D., AND ALAN F. SCHATZBERG, M.D.

Historically, the combination of multiple psychotropic medications for the treatment of psychiatric disorders has been frowned upon. The practice of “polypharmacy” was often seen as evidence of poor clinical judgment. A patient who was on three or four psychotropics was one whom many clinicians and impartial observers would characterize as overmedicated and perhaps improperly diagnosed and cared for.

However, the significant limitations of monotherapy with antidepressant, antipsychotic, and mood-stabilizing agents have become increasingly apparent. For many patients, monotherapy results in insufficient benefit.

There are three primary reasons for combining medications in psychiatric practice. Perhaps the most common reason for polypharmacy is that many patients do not respond adequately to a single agent. For example, in acute trials with antidepressants, most unipolar depressed patients are either partial or nonresponders to monotherapy. Medications are thus frequently combined to enhance antidepressant response. A second related reason for combining therapy is that not all aspects of a given illness may respond to a single agent. This may be particularly true in bipolar disorder, where monotherapy with a single mood stabilizer has become “the exception rather than the rule.” Treating the manic phase or the depressed phase may require different combinations of medications. Finally, medications are sometimes added to counteract the side effects of the original medication. Thus, patients on first-generation antipsychotics are usually given an anticholinergic agent to mitigate extrapyramidal side effects.

In addition to the advantages of combination treatment, there can be important downsides to combining psychotropic agents. The most important of these are the risks of pharmacodynamic and pharmacokinetic drug interactions. Cytochrome P450 interactions can result in either substantially higher or lower serum levels which might increase side effects or render one or more agents less effective. For example, the combination of paroxetine and imipramine can substantially raise the serum levels of imipramine and potentially increase antimuscarinic, hypotensive, cardiac conduction, or sedating effects of the imipramine. Likewise, pharmacodynamic interactions can be equally serious. For example, the combination of an monoamine oxidase inhibitor (MAOI) and a serotonergic agent may result in a potentially lethal serotonin syndrome. Thus, clinicians must carefully weigh the risks and benefits of all combination treatments.

COMBINATION STRATEGIES IN THE TREATMENT OF MAJOR DEPRESSION

Antidepressant–Antipsychotic Combinations

Given the limited efficacy of antidepressant monotherapy in the treatment of major depression, augmenting agents or combination strategies are frequently employed in clinical practice to enhance response. Conventionally, *augmentation* is used to describe the addition of an agent that is not viewed as a primary antidepressant by itself (i.e., thyroid hormone). *Combination* has been used when a known antidepressant is added to another (i.e., bupropion [Wellbutrin] and fluoxetine [Prozac] combination).

Although antipsychotics have been used in combination with antidepressants for many years in the treatment of psychotic depression, the second-generation antipsychotics were the first group of agents food and drug administration (FDA) approved in the adjunctive treatment of depression that had not responded to monotherapies. The pharmacology of the second-generation drugs, including their 5-hydroxytryptamine type 2 (5-HT2) antagonism, has long been thought to imbue them with antidepressant properties. In addition, they clearly help a number of symptom domains in depressed patients, including sleep, appetite, and agitation. As of this writing, several second-generation antipsychotics have FDA approval in the adjunctive treatment of resistant depression, and several more are in late-stage clinical trials.

Olanzapine (Zyprexa) was the first of the second-generation antipsychotics to seek approval in the adjunctive treatment of depression. The combination of 10 to 20 mg/day of olanzapine with fluoxetine in the treatment of various types of depression, including resistant depression. In 2007, the olanzapine–fluoxetine combination (Symbyax) was approved for the treatment of resistant depression. Combination pills, however, tend not to be popular with clinicians. Many clinicians prefer the more flexible dosing options of combining two different agents. Since both olanzapine and fluoxetine are generically available, there is also a cost advantage of using two generic agents over one branded drug.

Aripiprazole (Abilify) has long been of interest in the treatment of depression because it is a partial agonist of dopamine and a 5-HT1A agonist in addition to its 5-HT2 effects. Randomized controlled trials have demonstrated that the addition of aripiprazole to an selective serotonin reuptake inhibitor (SSRI) or venlafaxine (Effexor) is more effective than the addition of a placebo in relieving depressive symptoms. As a result, aripiprazole was FDA approved in the adjunctive treatment of major depression. In addition to the FDA registration trials demonstrating the efficacy of aripiprazole in major depressive disorder (MDD), a 1,500-patient multicenter trial (VAST-D) found that aripiprazole augmentation was somewhat more effective in achieving remission in MDD patients than either switching to bupropion or augmentation with bupropion. The relative lack of weight gain, sedation, or metabolic effects in aripiprazole compared to other second-generation agents such as olanzapine has made it among the most frequently prescribed antidepressant-augmenting agents. However, akathisia and some weight gain were noted in the aripiprazole augmentation trials.

Quetiapine (Seroquel), the first monotherapy approved for bipolar depression, also appears to be useful in many unipolar patients. Several registration trials demonstrated that the addition of 150 to 300 mg/day of quetiapine to standard antidepressants improved antidepressant response. Thus, quetiapine was approved as an adjunctive treatment for major depression in 2009. In 2015, brexpiprazole, which is chemically related to aripiprazole, was also FDA approved for the adjunctive treatment of depression. Other second-generation antipsychotics, including lurasidone (Latuda), ziprasidone (Geodon), cariprazine (Vraylar), and risperidone (Risperdal), also appear effective in some open-label and controlled studies as adjunctive agents in unipolar depression. While lurasidone is approved as a treatment for bipolar depression and schizophrenia, it also has demonstrated efficacy in at least four registration trials for the treatment of depressive symptoms associated with schizophrenia. As with other combination strategies, the potential benefits of augmenting with a second-generation antipsychotic must be weighed against the possible side effects of weight gain, extrapyramidal symptoms (EPS), and metabolic effects.

Lithium–Antidepressant Combinations

After antipsychotics, lithium has been among the best-studied augmenters to standard antidepressants in MDD. However, lithium augmentation has become less popular with clinicians in the past 25 years, secondary to the potential toxicity of lithium, the requirement for laboratory monitoring, and the availability of other augmenters. Nonetheless, lithium is still used by many clinicians to boost antidepressant response in partial and nonresponders to monotherapy with an antidepressant.

The early reports of lithium augmentation indicated that the addition of lithium to a tricyclic antidepressants (TCA) often resulted in clinical improvement within 72 hours in patients who had failed to respond to a TCA alone. Since then, at least 80 studies have been completed on lithium augmentation, and the vast majority have suggested some acute benefits of the addition of lithium to an antidepressant. However, many of these reports suggest maximum benefits may not be seen for weeks. Furthermore, the addition of lithium also appears to prevent relapses in unipolar patients more effectively than an antidepressant alone. The response often occurs at low doses (600 to 1,200 mg/day) and at relatively low serum levels. Current data would suggest that serum levels in the range of 0.5 to 0.8 mEq/L are effective in lithium augmentation of antidepressant effects.

Although most studies of lithium augmentation have employed TCAs, a number of studies have evaluated the efficacy of lithium augmentation of SSRIs. The National Institute of Mental Health (NIMH)-funded Sequenced Treatment Alternatives for Resistant Depression (STAR*D) trial is among the most frequently cited study that included augmentation strategies. In patients who had failed monotherapy with citalopram (Celexa) and an augmentation trial with either buspirone (BuSpar) or bupropion, only 15.9% of patients achieved a remission level response with the addition of lithium. This was a numerically but not statistically lower level of response than those patients who were augmented with thyroid hormone at that point in treatment. Since most lithium augmentation trials do not indicate the degree of treatment resistance in the study population, the STAR*D data may be the most relevant to date on the limitations of lithium augmentation in resistant–depressed patients. There are likely certain characteristics in depressed patients that may predict response to the addition of lithium to their antidepressant regimen. For example, there is some suggestion that lithium augmentation may be more successful in MDD patients who exhibit psychomotor retardation, have significant anorexia and weight loss, have suicidal thoughts, or have a low serum cortisol level.

Thyroid Supplement–Antidepressant Combinations

The addition of thyroid hormone to augment antidepressant response has been employed since the 1960s. As with lithium, it has also faded in popularity with the advent of new augmentation options.

Trials that have compared thyroid to lithium augmentation directly have generally suggested that thyroid augmentation is at least as effective as lithium augmentation. The few head-to-head trials of thyroid to lithium augmentation have tended to favor thyroid augmentation.

Although most trials employing thyroid augmentation have involved the use of TCAs, a number of trials also suggest efficacy when thyroid supplements are added to SSRIs. Multiple case reports and small studies have suggested that 25 to 50 mcg of liothyronine (Cytomel) (triiodothyronine or T3), added to various SSRIs, potentiate their antidepressant effects. Among the largest and most systematic thyroid augmentation of an SSRI trial to date was the STAR*D trial. In this trial, patients who had failed two other trials (monotherapy with citalopram and augmentation with bupropion or buspirone) were treated with triiodothyronine or lithium augmentation. About 25% of patients on T3 augmentation of citalopram achieved a remission level response compared to about 15% of the lithium-augmented patients. Among the limitations of the STAR*D trial was that it was an open-label trial, and patients had the option of not participating in randomization. Not all placebo-controlled studies have shown the benefit of thyroid augmentation.

The question of whether T3 is more effective than supplementation with levorotatory thyroxine (T4) has not been adequately addressed in the literature. Most studies have employed T3, and the only head-to-head study favored T3 over T4. Since the half-life of T4 is relatively long and the study was short (2 weeks), it cannot be conclusively interpreted for the superiority of T3. Indeed, there are positive studies employing T4 in thyroid augmentation. Although side effects were not common in this trial, the use of supraphysiological doses of thyroxine for an extended period might be problematic for some patients.

Patient characteristics of responders to thyroid augmentation include females and those with a history of thyroid abnormalities. There is growing evidence that patients with mild thyroid abnormalities (slightly elevated thyroid stimulating hormone (TSH), normal T3 and T4) are less likely to respond to antidepressants alone but may respond to the addition of a thyroid supplement. Women over the age of 50, perhaps because they are most susceptible to hypothyroidism, also appear to have higher levels of response to thyroid augmentation. One study has reported that responders to T3 could be identified by their having lower T3 levels and by being carriers for a specific allele on an single nucleotide polymorphism (SNP) for the deiodinase-converting enzyme for T3 to T4. This finding is in need of replication but could explain why some patients appear to derive great benefit from thyroid augmentation.

Stimulants/Dopamine Agonist–Antidepressant Combinations

There are a variety of stimulants and stimulant-like agents that have been useful augmenters in depression. Amphetamines have been used since the early 1930s for the treatment of depression. In addition, dopamine agonists such as pramipexole (Mirapex), dopamine reuptake blockers including bupropion, nondopamine and wake-promoting agents such as modafinil (Provigil) have shown some promise as augmenters in the treatment of depression.

The use of stimulants as augmenting agents in the treatment of depression has been applied to special populations, including medical and geriatric patients. Amphetamines tend to be used to target fatigue, hypersomnia, and cognitive difficulties in depressed patients. When combined with TCAs, methylphenidate and amphetamines inhibit the cytochrome P450 CYP isoenzyme 2D6 and artificially increase TCA serum levels.

Although most amphetamine augmentation studies have involved TCAs, a few have shown benefits in combination with MAOIs and SSRIs. Pharmacodynamic interactions between the stimulants and MAOIs can be problematic. Most controlled trials with amphetamine in the treatment of major depression have not shown benefits as a monotherapy. By far, the largest and most comprehensive studies to evaluate a stimulant to augment standard antidepressants involved lisdexamfetamine combined with an SSRI or serotonin norepinephrine reuptake inhibitor (SNRI). Subjects who failed to respond to 8 weeks of antidepressant monotherapy were randomized to an additional 8 weeks of lisdexamfetamine or placebo combined with the antidepressant. The two phase III trials involving over 800 patients failed to show an advantage of adding the lisdexamfetamine over adding a placebo to the antidepressant monotherapy.

There are now several dopamine agonists approved for the treatment of Parkinson disease, including pramipexole. Preliminary work with pramipexole as an adjunctive agent to SSRIs has suggested that it may have a role in treatment-resistant patients in doses of 1 mg/day (starting doses are lower). Ropinirole (Requip) is also actively being investigated as an adjunctive treatment in resistant depression. In addition, pramipexole and ropinirole clearly treat restless legs, including restless legs exacerbated by the use of SSRIs.

Modafinil and armodafinil are nonstimulants that are approved by the FDA for the treatment of excessive sleepiness associated with narcolepsy, obstructive sleep apnea, and shift work sleep disorder. The lack of a more significant abuse potential for modafinil and armodafinil has led to their application as adjunctive agents in the treatment of depression with the same symptom targets as the stimulants: fatigue and hypersomnia. Several studies have indicated that modafinil is well tolerated and appears to be useful in depressed partial responders to SSRIs. While meta-analyses of completed randomized trials with combination treatment with modafinil and armodafinil have generally suggested the efficacy of these drugs in bipolar and unipolar patients, registration trials of armodafinil for the treatment of bipolar depression failed to demonstrate adequate efficacy.

Bupropion has weak dopamine reuptake–blocking properties but may act primarily as a noradrenergic agent. It is among the most common augmenting agents used today in resistant depression. The combination of bupropion with SSRIs is simple and usually well tolerated. Studies completed to date have been retrospective or open. The largest of these open trials, the STAR*D trial, reported a remission rate of about 30% in patients who had failed to achieve remission on citalopram alone. As with many augmentation trials, it is not clear whether longer treatment with citalopram alone may have resulted in the improvement. In fact, the CO-MED trial demonstrated that starting with the combination of escitalopram and bupropion was no more effective than treating with escitalopram alone in achieving antidepressant remission or response. Nonetheless, the experience of U.S. clinicians supports the use of bupropion augmentation.

Sex Steroid–Antidepressant Combinations

It is known that estrogen-deficiency states, such as the postpartum period, may increase the risk of depressive symptoms in some women. In addition, estrogen possesses some antidepressant properties, including

MAO inhibition and serotonin agonist properties that are dose and duration specific. However, estrogen supplementation has been inconsistently useful in the treatment of depressed women. Some studies found that estrogen had no significant antidepressant effect when used alone for perimenopausal depression. In addition, other studies have found estrogen to be an effective prophylaxis against recurrent affective illness in the postpartum period. However, there are few adequately controlled studies, and a number of studies have not shown substantial benefit.

The combination of estrogens with antidepressants has sometimes been useful. For example, there have been case reports of oral contraceptives augmenting the effects of antidepressants in patients with refractory depression. In addition, postmenopausal women may respond better to fluoxetine if they are on hormone replacement therapy than if they are not. Recent studies of estrogen augmentation of SSRIs in both pre- and postmenopausal women have not shown a consistent benefit. However, 0.1 mg of transdermal estradiol (Alora) was suggested to accelerate SSRI response in some perimenopausal women. The potential benefits of estrogen augmentation must be balanced against the defined risks to long-term treatment with unopposed estrogen, and the benefits have not been clearly established in the treatment of depression. Overall, the data suggest maximal benefit during the perimenopausal phase.

Dehydroepiandrosterone (DHEA), a corticosteroid precursor of both androgens and estrogens, is used over the counter in many older patients with claims of enhanced mood, energy, and sexual vigor. There are a few controlled trials that indicate that DHEA may be efficacious in the treatment of major depression, minor depressive symptoms, and dysthymia. The safety of DHEA has been called into question because of its potential to enhance the growth of sex-steroid tumors and possible virilizing effects in women.

Esketamine/Ketamine–Oral Antidepressant Combination

Racemic ketamine and its S-enantiomer, esketamine (Spravato), are increasingly used in the treatment of resistant depression. In addition to second-generation antipsychotics, esketamine is the only other drug currently FDA approved as an adjunctive treatment with oral antidepressants in resistant depression. Esketamine was approved as a rapidly acting adjunctive agent in 2019 and, in 2020, received another indication for treating depressive symptoms in patients with suicidal ideation or actions. Racemic etamine has demonstrated utility as a treatment for resistant depression in smaller controlled trials since 2006 and is used off-label for this indication. Intranasal esketamine was shown to be effective in one-third of registration trials in patients who had failed two or more antidepressants and was near positive in the other two trials. In addition, adjunctive esketamine was shown to effectively reduce the relapse rate of patients who had responded to or remitted with acute, adjunctive treatment of esketamine in combination with an oral antidepressant. The antidepressant effect of ketamine and esketamine is often seen in 2 to 24 hours after administration and tends to last about 1 week. Thus, repeated, intermittent dosing is generally required to sustain a response. Between 30% to 60% of patients unresponsive to monotherapies for MDD may respond to an acute trial of ketamine or esketamine.

Ketamine and esketamine are Schedule 3 drugs administered in the clinic and, as such, are different than other antidepressant combinations. As controlled substances, they require a drug enforcement administration (DEA) registration and protocols for administration, inventory, security, disposal, etc. Ketamine is a common drug of abuse in many parts of the world, particularly in Asia. Thus, patients with a significant history of abuse may not be ideal candidates. In addition, disassociation is a common side effect of ketamine, and some patients with dissociative symptoms, including some post traumatic stress disorder (PTSD) patients, may be sensitive to these drugs. Patients with a history of psychosis may also be sensitive to the psychotomimetic effects of ketamine and esketamine.

Other Antidepressant Combinations

A variety of other antidepressant combinations are used in the treatment of resistant depression with varying degrees of scientific support. Given the poor response to antidepressant monotherapy, the question of whether combining antidepressants from the onset of treatment would improve remission or response rates was evaluated in a large multicenter trial. The CO-MED trial (2011) evaluated whether the combination of escitalopram and bupropion or the combination of venlafaxine and mirtazapine from the onset of treatment was superior to monotherapy treatment with escitalopram or venlafaxine alone in achieving remission or response. There appeared to be no benefit from starting treatment with combination treatment versus monotherapy.

Trazodone (Desyrel) has been used primarily as a hypnotic to treat depression or emergent insomnia. Although some small open studies have suggested an augmenting role for the use of trazodone combined with SSRIs, there is little evidence of benefit beyond the hypnotic effects of the drug. Likewise, buspirone, a generically available anxiolytic, has been applied as an augmenter to SSRIs with inconsistent benefits. Double-blind trials for buspirone augmentation of SSRIs in the treatment of obsessive compulsive disorder (OCD) have been negative, and only open studies in depression suggest some benefit of adding buspirone to an SSRI. The STAR*D study found that among patients who had failed to achieve remission to citalopram alone, about 30% of patients achieved remission with the addition of buspirone. This is the same rate found to remit with bupropion augmentation in the trial.

The beta-blocker pindolol (Visken) has some support in the literature as an augmenter to SSRIs but little support among clinicians.

There have been multiple controlled trials of pindolol augmentation, with most being positive. The relative lack of use among clinicians may be related to the perception that pindolol is not efficacious, or to concerns about the implication of beta-blockers in inducing depression, and some wariness of possible pharmacokinetic interactions between beta-blockers and CYP 2D6 inhibiting SSRIs such as fluoxetine and paroxetine (Paxil). Mirtazapine, like trazodone, can treat depression-associated insomnia and may help stimulate appetite in depressed older adult patients. The combination of venlafaxine and mirtazapine in patients who had failed three other trials in the STAR*D trial and the combination of venlafaxine and mirtazapine showed more of an improvement in depressive symptoms and was far better tolerated than the MAOI tranylcypromine (Parnate). However, the overall response rates in the mirtazapine SNRI combination and tranylcypromine groups were not statistically different. Only about 10% of patients achieved remission.

The combination of bupropion and dextromethorphan was approved for the treatment of MDD in 2022. Dextromethorphan has been available as an over-the-counter antitussive agent since the 1950s but studied in the treatment of mood disorders since at least 2014. For example, an open-label study of quinidine/dextromethorphan for the treatment of major depression showed had 45% response rate over 10 weeks of treatment.

While subsequent studies with quinidine in combination with dextromethorphan for major depression did not demonstrate efficacy, the combination of bupropion with dextromethorphan has proven effective. Like quinidine, bupropion is also a CYP2D6 inhibitor which inhibits the normally rapid metabolism of dextromethorphan allowing for adequate and consistent blood levels. In an early comparison study, the combination of bupropion 105 mg/dextromethorphan 45 mg twice daily was compared to bupropion alone 105 mg twice daily in 80 patients with MDD. The combination showed rapid response by week 2 and by week 6, was 46% for the combination and 17% for bupropion monotherapy. The phase III program has also shown consistent efficacy of the combination tablet compared to placebo. The combination appeared to be well tolerated with anxiety being the only side effect that leads to discontinuation of the combination tablet at higher rates than placebo and was seen in 2% of patients.

COMBINATION PHARMACOTHERAPY IN BIPOLAR DISORDER

Combination of Two Mood Stabilizers

Mood stabilizers typically refer to agents that are used to treat one or more poles of bipolar illness and might also have utility in the maintenance treatment of bipolar disorder. Mood stabilizers have historically included agents such as lithium, divalproex (Depakote), carbamazepine (Tegretol), and perhaps other anticonvulsants. Although agents such as second-generation antipsychotics might also be appropriately referred to as mood stabilizers, in this section, mood stabilizers will refer to lithium and anticonvulsant agents that have demonstrated utility in bipolar disorder. Approximately two-thirds of patients with bipolar affective disorder are treated with two more medications as maintenance therapy. In addition, second-generation antipsychotics are now frequently added during acute manic and depressive episodes, and sedative/anxiolytics are used to manage sleep, anxiety, and agitation. Thus, monotherapy is the exception in the treatment of bipolar disorder rather than the standard of care.

Lithium is commonly combined with anticonvulsants such as valproate, carbamazepine, and lamotrigine in the treatment of patients with refractory mania, in maintenance treatment, and in the treatment of bipolar depression. In most, although not all, studies, the combination of two mood stabilizers or a mood stabilizer and an antipsychotic is somewhat more effective than monotherapy with a single mood stabilizer. A number of reports have indicated that lithium–carbamazepine combinations are effective in patients who have failed to respond to these two agents given separately. Retrospective reviews have generally found the combination of lithium and carbamazepine to be useful and synergistic. There is also evidence that this combination may be particularly useful in rapid-cycling bipolar disorder. The combination of lithium and carbamazepine appears to be well tolerated. There is no evidence of increased neurotoxicity or blood dyscrasias with the combination. Dosing schedules of the lithium and carbamazepine should parallel regimens used for prescribing each drug alone. Therapeutic serum levels of both drugs should be monitored and maintained.

Perhaps the most common mood stabilizer combination in recent years has been the addition of valproate (Depakote) to lithium in patients who have manic breakthroughs or are unresponsive to lithium monotherapy. There are still relatively few prospective studies that have examined the efficacy of this combination. One naturalistic study indicated that the addition of valproate to lithium is more likely to be effective if the lithium level is above 0.6 mEq/L. Open-label studies have suggested that the combination of valproate and lithium may be more effective than either agent alone in the treatment of rapid-cycling bipolar disorder. For example, one study reported that eight of nine rapid-cycling patients with bipolar disorder, most of whom failed to respond to the combination of lithium and carbamazepine, responded when lithium was combined with valproate. However, other studies have not necessarily supported the advantage of adding valproate to lithium in the treatment of rapid-cycling bipolar disorder. Likewise, while the combination of valproate and lithium is commonly thought to be more effective in preventing cycles of mania and depression than monotherapy, there is little data to support this idea. The largest trial to date to evaluate the efficacy of the valproate/lithium combination in the maintenance treatment of bipolar 1 disorder, the BALANCE study (2010), failed to show an advantage of the combination treatment over lithium alone. However, the combination was more effective than valproate monotherapy. The combination of lithium and valproate seems to be well tolerated, and the dosing should aim to achieve adequate serum levels for both drugs.

The combination of valproate with carbamazepine has fallen out of favor in the past decade. Unfortunately, only anecdotal data support the combination of valproate and carbamazepine in the treatment of resistant bipolar disorder. In addition, a CYP450-based interaction increases the risk of carbamazepine toxicity. Also, the two taken together may result in even greater leucopenia than either alone.

A variety of other anticonvulsants such as gabapentin (Neurontin), topiramate (Topamax), lamotrigine (Lamictal), and tiagabine (Gabitril Filmtabs), levetiracetam (Keppra), and oxcarbazepine (Trileptal) have been combined with lithium and divalproex in the treatment of resistant bipolar patients. With the exception of lamotrigine, most of the reports detailing the benefits of combining standard mood stabilizers with these other anticonvulsants in the treatment of bipolar disorder have been anecdotal. For example, many open-label studies have supported the role of gabapentin in the treatment of bipolar disorder. At this time, no controlled trials have demonstrated a clear benefit of gabapentin in any phase of bipolar disorder as either an adjunct or a monotherapy. In one large-scale, double-blind study, gabapentin was no more effective than placebo when added to lithium. In that study, the lithium dosage could be adjusted during the trial, and this may have limited the likelihood of finding statistically significant differences. However, gabapentin may still benefit some symptoms in bipolar patients. For example, one small controlled trial has suggested that while gabapentin does not show acute benefits, it may augment the effect of standard mood stabilizers in preventing relapse in bipolar patients. Gabapentin is well tolerated in combination with most mood stabilizers and appears to benefit anxiety and agitation associated with a variety of conditions. A chemically related compound, pregabalin (Lyrica), has shown utility in the treatment of generalized anxiety but has not been extensively evaluated in the treatment of bipolar disorder.

Lamotrigine has shown efficacy in the maintenance treatment of bipolar disorder and in some studies in the treatment of acute bipolar depression, but its efficacy in the treatment of acute mania has not been demonstrated. Open-label augmentation with lamotrigine for bipolar depression in the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) did not show significant benefits on the primary outcome measures when compared to either risperidone or inositol. However, the patients treated with adjunctive lamotrigine had lower depression ratings than patients treated either with adjunctive lamotrigine or inositol. The benefits of lamotrigine make it likely that it will be increasingly employed as an adjunctive treatment to lithium, an atypical antipsychotic, or valproate in the treatment of the depressed phase of the illness. There do not appear to be significant pharmacokinetic interactions between lamotrigine and lithium. However, the risk for interactions does exist in the combination of lamotrigine with either valproate or carbamazepine. Valproate doubles the serum level of lamotrigine, thus increasing the risk of rash, while carbamazepine, on the other hand, cuts the lamotrigine serum levels in half.

The addition of oxcarbazepine, tiagabine, levetiracetam, and topiramate to lithium or valproate has not been extensively studied in the treatment of bipolar disorder. Oxcarbazepine is thought to carry fewer risks of pharmacokinetic interactions than does its chemical cousin carbamazepine. Thus, oxcarbazepine may be easier to combine with standard mood stabilizers than carbamazepine. However, some enzyme induction still occurs with the drug, and at least one controlled study did not find any benefit of adding oxcarbazepine to lithium. Earlier studies have supported the efficacy of oxcarbazepine in the treatment of mania and hypomania.

Topiramate has been studied primarily as an adjunctive agent in the treatment of bipolar disorder. Open trials suggest that the addition of topiramate to a standard mood stabilizer may help the depressed phase of the illness and improve symptoms in refractory patients. However, most controlled trials have failed to show a benefit of topiramate in bipolar disorder. The most common current rationale for combining topiramate with a mood stabilizer is to mitigate the weight gain associated with standard therapy. Topiramate has been associated with weight loss averaging up to 5 kg in 8 weeks as adjunctive treatment in bipolar disorder.

Tiagabine has not been proven effective or practical in the treatment of acute mania. The drug requires a slow titration secondary to potential side effects, including sedation and risk of seizure, with loading doses. However, open trials of tiagabine as an add-on to a standard mood stabilizer in refractory bipolar disorder have reported benefits at low doses (1 to 8 mg/day).

Levetiracetam was approved as an adjunctive treatment for partial seizures in 2000. A number of small case series have suggested some benefits of adding 1,000 to 4,000 mg/day of levetiracetam to standard mood stabilizers in the treatment of resistant mania. Likewise, adjunctive treatment with zonisamide (Zonegran) may have benefits for both the manic and depressed phase of bipolar disorder in open-label studies. Like topiramate, zonisamide has been associated with weight loss.

Mood Stabilizer–Antipsychotic Combinations

Antipsychotics have long been used as adjunctive treatments in bipolar disorder and are now used as monotherapies as well. Historically, an antipsychotic was used to treat acute mania or the psychotic symptoms and agitation associated with mania. Most of the second-generation antipsychotics are now approved for the treatment of acute mania, and a number are now approved in the depressed and maintenance phase of bipolar disorder. Many patients who are hospitalized with a manic episode will be treated with an antipsychotic soon after admission because of their more rapid onset of action relative to mood stabilizers. Multiple controlled studies have demonstrated that the addition of an antipsychotic to a mood stabilizer, particularly lithium and valproate, may be more effective in treating both manic and depressive symptoms than a mood stabilizer alone. For example, controlled trials employing risperidone, quetiapine, lurasidone, ziprasidone, and olanzapine have shown benefits in augmenting the antimanic effects of mood stabilizers. In a pooled analysis of five large controlled trials of second-generation antipsychotics combined with lithium in the treatment of acute mania, there appeared to be about a 20% advantage of the combination treatment over monotherapy with lithium or the antipsychotic.

Given that the second-generation antipsychotics appear effective in preventing mania, antipsychotics are commonly added to lithium or valproate when there are breakthrough symptoms during monotherapy. However, to date, there is very little controlled data to evaluate the efficacy of the antipsychotic/mood stabilizer combination in the prevention of bipolar cycling. Likewise, while several second-generation antipsychotics are now approved in the treatment of acute bipolar depression, relatively few studies have been completed to demonstrate that lurasidone, cariprazine, quetiapine, or olanzapine added to a mood stabilizer increases antidepressant efficacy. Still, the combination of one of the approved antipsychotics with either lithium or lamotrigine makes intuitive sense.

The combination of lithium clozapine (Clozaril) is occasionally used in the most treatment-resistant bipolar patients. A potential disadvantage of lithium combined with clozapine is the risk of neuroleptic malignant syndrome (NMS) as well as other side effects.

Carbamazepine–antipsychotic combinations may result in a pharmacokinetic interaction that reduces the serum levels of concurrent antipsychotics. A higher dose of the antipsychotic may be required.

Mood Stabilizer–Antidepressant Combinations

The treatment of bipolar depression remains a vexing problem. The mood stabilizers and antipsychotics, even in combination, are often inadequate to prevent or treat the depressed phase of the illness alone. On the other hand, the addition of an antidepressant may increase the risk of a switch into mania or rapid cycling. It is even unclear whether the addition of an antidepressant has any benefit in most patients. The few controlled studies to date have not necessarily supported the benefit of adding an antidepressant to a mood stabilizer. For example, the STEP-BD did not find that the addition of an antidepressant to lithium or valproate was more effective than an adjunctive placebo in the treatment of bipolar depression. In fact, there was a nonsignificant trend favoring the placebo group in a number of secondary measures. However, the rates of switching were also not different between the groups.

TCAs may be more problematic than other classes in exacerbating the illness, but all effective antidepressants, including electroconvulsive therapy, have the potential to switch a bipolar patient into a manic episode. Other SNRIs, including venlafaxine, may share a greater proclivity for inducing switches than bupropion. Bupropion may be less likely to induce rapid cycling or mixed states than either TCAs or venlafaxine, but there have been reported cases of bupropion-induced mixed states and manic episodes. In addition, the data on switch rates with bupropion is limited, although some studies, including the Stanley Foundation study on antidepressant treatment in bipolar disorder, have suggested that the switch rate into hypomania may be lower with bupropion than with some agents, including SNRIs. The addition of antidepressants, including TCAs, to lithium, does not appear to increase the risk of switching any more than lithium alone. Thus, most antidepressants can probably be added to lithium without substantially increasing the risk of a switch. By extrapolation, it would be expected that the addition of an antidepressant to valproate or carbamazepine might also mitigate the risk of a switch to mania. However, prospective trials are lacking on antidepressant–anticonvulsant combinations in the treatment of bipolar depression.

An important question is whether the addition of a second mood stabilizer is more efficacious for the treatment of bipolar depression than the addition of an antidepressant. Although antidepressants may destabilize bipolar cycling in the long run, at least one acute study has suggested that adding an SSRI may be as useful as adding a second mood stabilizer. This finding may be more a reflection of the lack of efficacy of adding a second agent to a mood stabilizer rather than an indication that both antidepressants and a second mood stabilizer are equally effective.

The Combination of Thyroid Hormone and Mood Stabilizers

Abnormalities in the thyroid axis have long been known to destabilize bipolar disorder. Thyroid dysfunction has been associated with rapid cycling, mixed states, depression, and refractory mania in bipolar patients.

The use of thyroid supplements in the treatment of bipolar has primarily been in bipolar disorder in two states: rapid cycling and bipolar depression. In the 1990s, small studies of supraphysiologic doses of levothyroxine (100 to 300 mcg/day) were effective in treating rapid cycling disorder. Most of the studies supporting this approach have been open-label or small case series. However, there is one more recent (2018) randomized controlled trial (RCT) that has evaluated high-dose levothyroxine (T4), triiodothyronine (T3), or placebo added to lithium. That study dosed the thyroid supplements to the point of suppressing TSH levels (<0.1) and found the T4 but not T3 or placebo-treated patients spent less time depressed or in a mixed state and great time feeling euthymic.

As with rapid cycling, most studies of thyroid hormone combined with mood stabilizers for bipolar depression have been small, uncontrolled, and difficult to draw conclusions from. The largest study to date is a retrospective of 159 patients with insufficient response to previous medications. This analysis included bipolar 1, bipolar II, and bipolar not otherwise specified (NOS) patients who had failed an average of 14 previous medication trials. T3 was added to the existing regimen to an average dose of 90.4 mcg. Eighty-four percent of patients were reported to have had significant improvement on a Clinical Global Impression (CGI) scale, and 33% were reported to be in remission.

While the existing data on thyroid augmentation in bipolar disorder does suggest benefits when combined with mood stabilizers or antipsychotics, the data remain sparse. Potential side effects of high-dose thyroid supplements include osteoporosis, tachycardia, palpitations, and tremor. However, thyroid supplementation has generally been well tolerated in most of the studies to date.

COMBINATION STRATEGIES IN THE TREATMENT OF SCHIZOPHRENIA

Combining Two Antipsychotics

The combination of two or more antipsychotics in a schizophrenic patient who has inadequately responded to one agent is commonly attempted but has been rarely studied. Earlier onset of illness, a longer duration of illness, more frequent hospitalizations, and the presence of violence as a symptom of schizophrenia are more likely to be associated with treatment with a combination of antipsychotics. The augmentation strategy that has been perhaps most investigated in resistant schizophrenia has been the combination of clozapine with another antipsychotic. Since 40% to 70% of patients fail to achieve an adequate response to clozapine, clinicians have typically added a second antipsychotic. For example, a number of small studies have suggested a modest benefit of adding risperidone to clozapine-nonresponsive patients. Case reports and small open studies have also suggested a potential role for amisulpride (Solian) and aripiprazole in patients who do not respond to clozapine alone. A meta-analysis of 19 studies examining the combination of antipsychotics in the treatment of resistant schizophrenia concluded that there may be an advantage of combining antipsychotics in some resistant patients. However, the combination of antipsychotics has also been associated with more side effects, including an increased risk of parkinsonism and elevated prolactin levels.

Risperidone has been combined with typical antipsychotics to produce some improvement in chronic schizophrenic patients studied prospectively. In addition, risperidone has been added to clozapine with a benefit for refractory schizophrenic patients. The primary risks of combining antipsychotics may be the increased risk of EPS or the risk of hypotension with low-potency and atypical antipsychotics.

Antidepressant–Antipsychotic Combinations in Schizophrenia

Antidepressants are frequently combined with antipsychotics in the treatment of schizoaffective disorder and psychotic depression. In addition, antidepressants have shown benefits in the treatment of negative symptoms. Trazodone produces significant but modest improvements in negative symptoms compared to placebo when added to a standard antipsychotic. Likewise, the addition of a low-dose MAOI or moderate doses of nefazodone (Serzone) to a typical antipsychotic has been found to benefit negative symptoms. Controlled trials of adding SSRIs and mirtazapine to antipsychotics have been reported to benefit negative symptoms. Furthermore, the addition of mirtazapine has also been suggested to help antipsychotic-associated akathisia.

Mood Stabilizer–Antipsychotic Combinations in Schizophrenia

Some reports have suggested that lithium can augment standard antipsychotics in the treatment of schizophrenia. However, most controlled studies have not shown many benefits. In addition, the combination of lithium with standard antipsychotics may increase the risk of EPS. Lamotrigine has been effective in some studies as an augmenter to clozapine in treatment-resistant schizophrenia. For example, in a double-blind study, 200 mg of lamotrigine was more effective than placebo in improving negative symptoms as well as general psychopathology in treatment-resistant schizophrenia patients who had not responded to clozapine alone. However, a more recent study of the addition of lamotrigine to other second-generation antipsychotics did not show an advantage of adding lamotrigine in the treatment of positive or negative symptoms of schizophrenia. In addition, lamotrigine has been observed to reduce alcohol cravings in at least one study when added to standard antipsychotics in the treatment of schizophrenics with alcohol dependence.

The addition of carbamazepine to standard antipsychotics has not been clearly beneficial. On the other hand, a clear disadvantage of adding carbamazepine to many antipsychotics is that carbamazepine can induce the metabolism of the concurrent antipsychotic and reduce serum levels. Oxcarbazepine is less likely than carbamazepine to have a pronounced effect on the CYP450 system.

Valproate has been added to typical and atypical antipsychotics in resistant schizophrenia and catatonia. The data suggest that there may be a role for the addition of valproate in controlling agitation and perhaps impacting some positive and negative symptoms in schizophrenia. In addition, valproate augmentation of risperidone and olanzapine has been associated with more rapid improvement in schizophrenia. However, most studies have not shown a significant advantage of adding valproate to an antipsychotic in the treatment of schizophrenia.

The use of newer anticonvulsants in combination with antipsychotics is largely untested in schizophrenic patients. The use of gabapentin has anecdotally been beneficial in treating anxiety and agitation in some schizophrenic patients. Likewise, there are anecdotal reports of tiagabine being used for agitation in combination with both typical and atypical antipsychotics in some patients. Depressed schizoaffective patients have sometimes benefited from the addition of lamotrigine. Valproate has been applied for the treatment of agitation in schizophrenia.

Other Augmentation Strategies in Schizophrenia

Abnormalities in glutamate transmission have been proposed to play a role in the pathophysiology of deficit symptoms in schizophrenia. As a result, studies have examined the role of glycine and D-cycloserine as augmenters to antipsychotics. Glycine is an amino acid that is an agonist of the N-methyl-D-aspartate (NMDA) receptor. Early open trials have suggested that glycine helps some schizophrenic patients but worsens symptoms in others. D-cycloserine, another agonist of the NMDA receptor, may improve negative symptoms and neuropsychological functioning in patients on standard antipsychotics but not clozapine.

Cyproheptadine (Periactin) is a general serotonin antagonist used primarily to treat migraines and allergies. Its serotonin antagonist properties have made it of interest to clinicians treating negative symptoms since all atypical antipsychotics are also 5-HT2 antagonists. Cyproheptadine also has the advantage of being anticholinergic and can help with EPS. The problem with cyproheptadine is that most patients find it quite sedating, and 24 mg/day is a large dose for many patients. It is unclear whether smaller doses might work just as well.

FUTURE DIRECTIONS

The monotherapy of major depression, bipolar disorder, and schizophrenia is often insufficient to achieve an adequate therapeutic response. Thus the combination of psychotropic agents is often attempted in order to improve outcome. Although augmentation strategies have been best demonstrated in the treatment of resistant depression, combination treatment is common in the treatment of bipolar disorder and increasingly common in the treatment of resistant schizophrenia. Studies such as STAR*D and STEP-BD have provided additional insights into the utility and limitations of combination pharmacotherapy. However, much remains to be learned about optimal strategies for combining agents in the treatment of resistant psychiatric disorders.

FURTHER READINGS

*BALANCE investigators and collaborators, Geddes JR, Goodwin GM, Rendell J, et al. Lithium plus valproate combination therapy versus monotherapy for relapse prevention in bipolar I disorder (BALANCE): a randomised open-label trial. *Lancet*. 2010;375(9712):385–395.

Bauer M, Dopfmer S. Lithium augmentation in treatment-resistant depression: meta-analysis of placebo-controlled studies. *J Clin Psychopharmacol*. 1999;19:427–434.

Bowden CL, Singh V, Weisler R, et al. Lamotrigine vs. lamotrigine plus divalproex in randomized, placebo-controlled maintenance treatment for bipolar depression. *Acta Psychiatr Scand*. 2012;126(5):342–350.

Carpenter LL, Jocic Z, Hall JM, Rasmussen SA, Price LH. Mirtazapine augmentation in the treatment of refractory depression. *J Clin Psychiatry*. 1999;60:45–49.

Correll CU, Rummel-Kluge C, Corves C, Kane JM, Leucht S. Antipsychotic combinations vs monotherapy in schizophrenia: a meta-analysis of randomized controlled trials. *Schizophr Bull*. 2009;35(2):443–457.

Dé Montigny C, Grunberg F, Mayer A, Deschenes JP. Lithium induces rapid relief of depression in tricyclic antidepressant drug non-responders. *Br J Psychiatry*. 1981;138:252–256.

*Fava M, Rush AJ, Wisniewski SR, et al. A comparison of mirtazapine and nortriptyline following two consecutive failed medication treatments for depressed outpatients: a STAR*D report. *Am J Psychiatry*. 2006;163(7):1161–1172.

Geller B, Luby JL, Joshi P, et al. A randomized controlled trial of risperidone, lithium, or divalproex sodium for initial treatment of bipolar I disorder, manic or mixed phase, in children and adolescents. *Arch Gen Psychiatry*. 2012;69(5):515–528.

Goodwin FK, Prange AJ Jr, Post RM, Muscettola G, Lipton MA. Potentiation of antidepressant effects by L-triiodothyronine in tricyclic nonresponders. *Am J Psychiatry*. 1982;139:34–38.

Grosso G, Pajak A, Marventano S, et al. Role of omega-3 fatty acids in the treatment of depressive disorders: a comprehensive meta-analysis of randomized clinical trials. *PLoS One*. 2014;9(5):e96905.

Iosifescu DV, Jones A, O'Gorman C, Streicher C, Feliz S, Fava M, Tabuteau H. Efficacy and safety of AXS-05 (dextromethorphan-bupropion) in patients with major depressive disorder: a phase 3 randomized clinical trial (GEMINI). *J Clin Psychiatry*. 2022;83(4):21m14345.

*Joffe RT, Levitt AJ, Bagby RM, MacDonald C, Singer W. Predictors of response to lithium and triiodothyronine augmentation of antidepressants in tricyclic non-responders. *Br J Psychiatry*. 1993;163:574–578.

Johns CA, Thompson JW. Adjunctive treatments in schizophrenia: pharmacotherapies and electroconvulsive therapy. *Schizophr Bull*. 1995;21:607–619.

Juruena MF, Ottoni GL, Machado-Vieira R, et al. Bipolar I and II disorder residual symptoms: oxcarbazepine and carbamazepine as add-on treatment to lithium in a double-blind, randomized trial. *Prog Neuropsychopharmacol Biol Psychiatry*. 2009;33(1):94–99.

Keck PE Jr, McElroy SL, Vuckovic A, Friedman LM. Combined valproate and carbamazepine treatment of bipolar disorder. *J Neuropsychiatry Clin Neurosci*. 1992;4:319–322.

Loebel A, Cucchiaro J, Silva R, et al. Lurasidone as adjunctive therapy with lithium or valproate for the treatment of bipolar I depression: a randomized, double-blind, placebo-controlled study. *Am J Psychiatry*. 2014;171(2):169–177.

Murrough JW, Wade E, Sayed S, et al. Dextromethorphan/quinidine pharmacotherapy in patients with treatment resistant depression: a proof of concept clinical trial. *J Affect Disord*. 2017;218:277–283.

*Nierenberg AA, Ostacher MJ, Calabrese JR, et al. Treatment-resistant bipolar depression: a STEP-BD equipoise randomized effectiveness trial of antidepressant augmentation with lamotrigine, inositol, or risperidone. *Am J Psychiatry*. 2006;163(2):210–216.

Nierenberg AA, Price LH, Charney DS, Heninger GR. After lithium augmentation: a retrospective follow-up of patients with antidepressant-refractory depression. *J Affect Disord*. 1990;18:167–175.

Ouyang WC, Hsu MC, Yeh IN, Kuo CC. Efficacy and safety of combination of risperidone and haloperidol with divalproate in patients with acute mania. *Int J Psychiatry Clin Pract*. 2012;16(3):178–188.

Palinkas LA, Barrett-Connor E. Estrogen use and depressive symptoms in postmenopausal women. *Obstet Gynecol*. 1992;80:30–36.

Peselow ED, Fieve RR, Difiglia C, Sanfilipo MP. Lithium prophylaxis of bipolar illness. The value of combination treatment. *Br J Psychiatry*. 1994;164:208–214.

Post RM, Leverich GS, Nolen WA, et al; Stanley Foundation Bipolar Network. A re-evaluation of the role of antidepressants in the treatment of bipolar depression: data from the Stanley Foundation Bipolar Network. *Bipolar Disord*. 2003;5(6):396–406.

*Sachs GS, Nierenberg AA, Calabrese JR, et al. Effectiveness of adjunctive antidepressant treatment for bipolar depression. *N Engl J Med*. 2007;356:1711–1722.

Sagud M, Vuksan-Ćusa B, Zivković M, et al. Antipsychotics: to combine or not to combine? *Psychiatr Danub*. 2013;25(3):306–310.

Shelton RC, Tollefson GD, Tohen M, et al. A novel augmentation strategy for treating resistant major depression. *Am J Psychiatry*. 2001;158:131–134.

Szegedi A, Verweij P, van Duijnhoven W, Mackle M, Cazorla P, Fennema H. Meta-analyses of the efficacy of asenapine for acute schizophrenia: comparisons with placebo and other antipsychotics. *J Clin Psychiatry*. 2012;73(12):1533–1540.

Tabuteau H, Jones A, Anderson A, Jacobson M, Iosifescu DV. Effect of AXS-05 (dextromethorphan-bupropion) in major depressive disorder: a randomized double-blind controlled trial. *Am J Psychiatry*. 2022;179(7):490–499.

Thase ME, Howland RH, Friedman ES. Treating antidepressant nonresponders with augmentation strategies: an overview. *J Clin Psychiatry*. 1998;59(Suppl 5):5–12.

*Trivedi MH, Fava M, Wisniewski SR, et al; STAR*D Study Team. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*. 2006;354(12):1243–1252.

Wolkowitz OM, Reus VI, Roberts E, et al. Dehydroepiandrosterone (DHEA) treatment of depression. *Biol Psychiatry*. 1997;41:311–318.

Woo YS, Bahk WM, Chung MY, et al. Aripiprazole plus divalproex for recently manic or mixed patients with bipolar I disorder: a 6-month, randomized, placebo-controlled, double-blind maintenance trial. *Hum Psychopharmacol*. 2011;26(8):543–553.

Young LT, Joffe RT, Robb JC, MacQueen GM, Marriott M, Patelis-Siotis I. Double-blind comparison of addition of a second mood stabilizer versus an antidepressant to an initial mood stabilizer for treatment of patients with bipolar depression. *Am J Psychiatry*. 2000;157:124.

33.17 Sex and Psychopharmacology

33.17a Sex Differences in Psychopharmacology

Anna E. Bauer, Ph.D., M.P.H., and Samantha Meltzer-Brody, M.D., M.P.H.

There are well-known sex-related differences in the incidence and prevalence of psychiatric conditions. The prevalence of anxiety and mood disorders is greater in women, while schizophrenic psychoses and substance use disorders (SUDs) are more prevalent in men. The difference in depression is particularly profound; the ratio of women to men with major depressive disorder (MDD) is 2:1. Because of the greater burden of psychiatric illness, women use psychotropic agents more frequently than men, particularly antidepressants and anxiolytics. Women are also more susceptible to misuse and adverse effects of psychotropic drugs than men. Differences in the pathogenesis of psychiatric illness are affected by both physiologic and social factors, and sex is an important variable that is often neglected.

Although the precise mechanisms of psychiatric illness are still under investigation, it is well accepted that the underlying pathogenesis of psychiatric illness is associated to some extent with deficiency and dysregulation of neurotransmitter systems (i.e., norepinephrine, dopamine, serotonin, glutamate, and gamma-aminobutyric acid [GABA]) and changes in synaptic strength. These systems may be modulated by sex-specific factors such as fluctuations of neurosteroid hormones or dysregulation of the hypothalamic–pituitary–adrenal (HPA) axis due to gender-based violence and other stressors. There is evidence in both preclinical and clinical studies of sex differences in the functioning of monoaminergic systems that are implicated in the pathogenesis of mental illness. For example, women have more dopamine in the striatum than men and also greater availability of serotonin transporter. Estrogen also enhances presynaptic function and increases neurotransmitter release.

There is considerable evidence for sex-specific physiologic differences that affect drug metabolism, efficacy, and safety. Approximately 7% of new drug applications to the U.S. Food and Drug Administration (FDA) include evidence of robust pharmacokinetic sex differences. Factors that differ between sexes and might affect the pharmacokinetics of psychiatric medications are numerous and include body fat and weight distribution, liver metabolism rates, changes in physiology and hormone levels during times of reproductive change (i.e., puberty, menstruation, pregnancy, menopause), gastric emptying, gastric acid production, splanchnic blood flow, plasma volume, protein levels, enzyme activity, drug transport and clearance rates, medication adherence, side-effect profile differences, interactions between reproductive hormones and neurotransmitters, and brain monoamine functioning.

Despite the potential of physiologic sex differences to influence the efficacy, safety, and acceptability of psychopharmacologic agents, prior to the early 1990s, women were underrepresented in clinical trials. Researchers worried including women in clinical trials would increase interindividual variability due to menstrual cycles and other factors and would require larger sample sizes to detect significant effects. In 1993, the FDA issued its Guideline for the Study and Evaluation of Gender Differences in the Clinical Evaluation of Drugs and now recommends separate analyses of men’s and women’s responses to pharmaceutical agents.

Although data regarding sex differences in pharmacodynamics and pharmacokinetics of psychopharmacologic agents are still sparse, and current prescribing guidelines and recommendations do not generally differentiate sex, the current evidence provides cause to consider sex as an important factor when prescribing and dosing psychotropic medications. This chapter will describe the modulation of synaptic pathways by gonadal steroid hormones, sex differences in pharmacokinetic processes, sex differences in the pharmacokinetics of specific psychotropic drug classes, sex differences in efficacy and safety of psychotropic drugs, and periods of reproductive and hormonal change that may affect psychopharmacology.

SEX DIFFERENCES IN PHARMACOKINETICS AND PHARMACODYNAMICS

The efficacy and safety of a pharmacologic agent are primarily dependent on pharmacokinetics, which describes how a drug is absorbed, distributed, metabolized, and excreted in the body. Each of these pharmacokinetic phases is potentially affected by biologic sex differences. Sources of sex differences in pharmacokinetics include differences in body weight, plasma volume, gastric acid production, gastrointestinal (GI) motility, regional blood flow, transport protein levels, enzyme activity, basal metabolic rate, and drug clearance rate.

Anatomical differences alone account for higher average body weight, height, body surface area, total body water, and intra- and extracellular water volume in men compared with women. In contrast, women have a greater proportion of fat mass relative to body size compared to men. Women are also characterized by lower gastric acidity, lower levels of plasma binding proteins, differential gastric enzyme levels, and more efficient renal function than men. In combination with the specific pH, lipophilicity, and binding capacity of a drug, these factors change absorption, distribution, metabolism, and excretion, which subsequently alter the effective dose and tolerability of a drug.

Pharmacokinetic sex differences exist broadly for many pharmaceutical agents due to physiologic differences. These physiologic differences and the specific drug profiles of many psychotropic medications lead to the potential for divergent efficacy and safety profiles in men and women.

Absorption

Absorption of a drug is dependent on its physical and chemical properties, formulation, and route of administration. If given orally, absorption is highly dependent on the GI environment, including pH within the lumen of the GI tract, surface area per luminal volume, blood perfusion, presence of bile and other secretions, and epithelial membrane structure. Similarly, if given dermally, intravenously, or via aerosol, absorption is dependent on surface area and regional blood and fluid flow. Although sample sizes were small and differences were attenuated when correcting for body weight, a 20-year review of sex differences in pharmacokinetics demonstrated that women broadly have greater drug absorption and higher area under the concentration (AUC) time curve than men. GI motility, bile and gastric acid production, gastric enzymes, and transporter proteins, all affect drug absorption and differ between males and females.

On average, men have greater basal and maximal gastric fluid flow, faster gut motility, and faster gastric emptying. Mean transit times are shorter in men (44.8 hours) compared to women (91.7 hours). This leads to increased absorption and gastric hydrolysis in males. Bile acid composition and gastric acid secretion affect ionization and drug solubility, depending on the pH of the drug. Men have higher concentrations of cholic acid, while women have a higher concentration of chenodeoxycholic acid. Gastric fluids are generally more acidic in males than females (pH 1.92 vs. 2.59, respectively). A lower and more acidic pH increases the absorption of basic compounds, resulting in higher drug concentrations. Differential levels of gastric enzymes and transporter proteins by sex also affect absorption. Increased gastric enzyme activity can increase plasma drug concentrations. For example, alcohol dehydrogenase and aldehyde dehydrogenase are more active in men than women. These enzymes are involved in oxidative degradation, which subsequently makes ethanol more bioavailable in women and leads to greater effects of alcohol on women than men. Differential intestinal expression of proteins that modulate gut transport of drugs can also lead to differential plasma drug concentrations by sex. Transporter proteins like H^+ di- and tripeptide enhance drug absorption, while others, such as P-glycoprotein, inhibit drug absorption.

Distribution

The volume of distribution of a drug is determined by factors such as body composition, regional blood flow, and protein-binding capacity. Water and adipose tissue content affect drug concentration and distribution. On average, total body water, intra- and extracellular water, total blood volume, plasma volume, and red blood cell volume are greater in men than women. Therefore, a water-soluble drug will have an increased volume of distribution in a man, decreasing the overall drug concentration. Conversely, although women have lower body weight and lower blood volume on average, they have a higher percentage of adipose tissue. Thus, a lipid-soluble drug will have an increased volume of distribution in a woman, resulting in lower overall plasma concentrations.

Cardiac output and regional blood flow can impact drug distribution. When adjusted for body surface area, the cardiac index is similar between males and females. Regional blood flow is similar between sexes for some organs (adrenal, bone, brain, lung, skin, thyroid) but quite different for others (adipose, heart, kidney, liver, muscle). Differences in body fat and organ blood flow have been implicated in the speed of onset and duration of action.

Once absorbed, drugs typically bind to plasma proteins that are structurally specific to the drug, such as albumin, alpha-1-acid glycoprotein (AGP), and alpha-globulins. The concentration of the drug is dependent on the amount of protein-bound compound, and generally only the free quantity is active and can pass through the blood–brain barrier. AGP and alpha-globulins are lower in women and influenced by estradiol, which theoretically increases proportion of free drug in women.

Metabolism

Metabolism is one of the strongest sources of sex-dependent variability in pharmacokinetics. Men typically have a higher basal metabolic rate than women, which is partially attributed to differences in the body composition of adipose and muscle tissue. Metabolism of most psychotropic drugs is catalyzed by enzymatic systems in the liver, and hepatic metabolism is divided into two sequential reactions: drug modification (phase I) and conjugation (phase II). Lipophilic compounds are most susceptible to phase I types of metabolism.

Much of the hepatic metabolism of psychotropic drugs is regulated by the phase I cytochrome P450 (CYP) family of isozymes, particularly CYP2D6, CYP2A4, CYP1A1/2, and CYP2C19. Hydroxylation is slower in women than men, increasing drug plasma concentrations in women. Some of the CYP enzymes demonstrate clear sex-related differences. The most important and well established of these sex differences is with CYP3A4. CYP3A4 is responsible for the metabolism of benzodiazepines, painkillers, calcium channel blockers, and steroids. CYP3A4 is known to be up to 50% more active in young adult women than men and declines with age, resulting in lower plasma levels of these drugs in women of reproductive age.

Tricyclic antidepressants (TCAs), fluvoxamine, and clozapine are all metabolized by CYP1A2. CYP1A2 activity is slower in women than men, which can increase drug concentrations and the possibility of adverse events in women.

CYP2C19 contributes to the metabolism of diazepam and mephenytoin. Women show greater enzymatic activity in CYP2C19 than men. and activity is quite variable by ethnicity and contraceptive use.

CYP2D6 is responsible for the metabolism of most antidepressants, haloperidol, and some analgesics. Inhibition of CYP2D6 leads to increased plasma drug concentrations. CYP2C9 contributes to the metabolism of diazepam, imipramine, amitriptyline, and phenytoin. However, there have not been any significant sex differences found in CYP2D6 or CYP2C9 metabolism.

Enterocytes also express significant levels of phase II isoenzymes such as uridine diphosphate glucuronosyltransferase, and as with hydroxylation, glucuronidation is slower in women than men, making conjugation reactions less efficient in women than men.

Elimination

Elimination is responsible for drug inactivation, and drugs are eliminated primarily by hepatic, renal, and pulmonary routes. They can also be eliminated through sweat, tears, and breast milk. Some elimination occurs through hepatic elimination, in conjunction with metabolism, and is a function of liver blood flow and enzyme activity. Hepatic blood flow is lower in women than in men, and some hepatic enzymes differ by sex. Expression of p-glycoprotein regulates biliary excretion and is twofold lower in women than men. Most elimination occurs via the kidneys, and renal function, including glomerular filtration rate (GFR), tubular secretion, and tubular reabsorption, is typically more efficient in men than women, even when adjusted for body weight and surface area. Thus, on average, the overall elimination of compounds is typically lower and slower in women than men.

GONADAL STEROID MODULATION

Reproductive steroid hormones are regulated by the hypothalamic–pituitary–gonadal (HPG) axis. The hypothalamus releases gonadotropin-releasing hormone, which stimulates the pituitary to produce and release gonadotropins, luteinizing hormone (LH), and follicle-stimulating hormone (FSH). LH then stimulates the release of steroid hormones from the gonads. These hormones include estrogen and progesterone, which both regulate reproductive and cognitive function. Estrogen and progesterone both increase cerebral blood flow, and both can have neuroprotective effects. Within the brain, the prefrontal cortex, amygdala, hippocampus, and thalamus, all contain estrogen receptors, and estrogen can act both as a neurotransmitter and as a modulator of other neurotransmitters in synaptic pathways. Progesterone reduces the expression of proapoptotic proteins and apoptotic DNA fragmentation, which leads to reduced cell death and gliosis. Estrogen influences neurotrophic factors and facilitates the formation of dendritic spines.

Gonadal steroid receptors bind to enhancer elements to regulate genes and modify the expression of neurotransmitters and neuropeptides (e.g., serotonin by affecting tryptophan hydroxylase; GABA by affecting glutamic acid decarboxylase; acetylcholine by affecting choline acetyltransferase), their receptors (e.g., 5-HT1A, 5-HT2A), receptor conformation (e.g., GABA_A receptor), and neurotransmitter reuptake (e.g., serotonin transporter and postreceptor signal transduction). Within the cell, gonadal steroids can also modulate ion channels and activate signal transduction cascades.

Estrogen’s specific mechanism of action is unclear, but it is thought to be involved in both the pathogenesis of depression and the effectiveness of antidepressant medications. Estrogen also specifically interacts with the serotonergic system. Estrogen may separately play a role as both a mood enhancer and an antidepressant enhancer. Numerous studies have evaluated the association between absolute estrogen concentration and depression, but none have found a direct correlation. However, in perimenopausal women who have had recurrent episodes of depression, withdrawal of estrogen replacement therapy rapidly induced a new depressive episode. Further, in women at high risk for postpartum depression (PPD), treatment with conjugated estrogen prevented depression symptoms. Similarly, transdermal estradiol prevented depression symptoms in one study of perimenopausal women, and in another, treatment with estradiol resolved symptoms of depression, even in the absence of antidepressant use. In a study of postmenopausal women, those receiving estrogen replacement therapy and a selective serotonin reuptake inhibitor (SSRI) had a better response compared to those only treated with an SSRI. Several other studies of serotonergic agents showed a poorer response in postmenopausal women compared to premenopausal women, supporting the role of estrogen in enhancing serotonergic response.

LH may also interact with the serotonergic system. Serotonin levels tend to vary inversely with LH. Therefore, lower LH levels suggest higher baseline serotonin level for antidepressants to work upon, and low LH levels may predict an improved response to antidepressant therapy.

Gonadal steroid hormones, particularly LH, also influence HPA axis activity, which is disrupted in depression and other psychopathology. A hyperexcitable HPA axis is a common feature of depression, and antidepressants act to reduce HPA axis activity. Low LH levels are associated with HPA axis hyperactivity, and, thus, may identify those with hyperactive HPA axes who may be good responders to antidepressant treatment.

The GABA system is important in inhibiting the HPA axis, and GABA acts as the dominant inhibitory neurotransmitter at the level of the paraventricular nucleus of the hypothalamus. Gonadal steroid hormones are allosteric modulators of the $GABA_A$ receptor, and in particular, the progesterone metabolite, allopregnanolone, is a positive allosteric modulator of the $GABA_A$ receptor and has strong anxiolytic and tranquilizing effects. When administered in rats, allopregnanolone reduces the release of stress hormones and attenuates HPA axis activity.

In addition to modulating neurotransmitters, gonadal steroid hormones also alter protein binding and enzymatic activity, which affect the metabolism and excretion of substrates. Estrogen increases levels of serum-binding globulins: sex hormone–binding globulin, corticosteroid-binding globulin, and thyroxine-binding globulin. Increased levels of estrogen and progesterone alter hepatic enzyme activity. Reproductive hormones may lower the expression of p-glycoprotein in women.

All of the CYP family enzymes are coded for by autosomal (i.e., nonsex) chromosomes, but differences in regulation of expression and enzyme activity may be affected by endogenous hormones. Gonadotropin hormone secretion profiles are different in males and females and are preprogrammed during development by estrogens. Gonadotropin hormone secretion profile strongly controls expression and activity of CYP enzymes in rats and moderately in humans. Several CYP family enzymes are inhibited (e.g., CYP1A2) or increased (e.g., CYP2D6) by gonadal steroid hormones.

Hormonal variations due to the menstrual cycle and periods of hormonal variability can affect GI physiology and function. Estrogens inhibit gastric acid and reduce GI mobility, and progesterone decreases gastric emptying. Absorption time of antidepressants and other agents is increased by exogenous estrogens and progesterone.

Nearly all psychotropic drugs act via modulation of neurotransmitter-gated ion channels or signal transduction systems, which are affected by gonadal steroid levels. Thus, the same psychotropic agents would be expected to produce differential responses by sex. For most pharmacologic agents, absorption, distribution, metabolism, and excretion are all influenced to some extent by hormones. Thus, prescribing must take into account menstrual cycle phase, reproductive stage, and concomitant use of oral contraceptives or hormone replacement therapy.

REPRODUCTIVE PERIODS THAT AFFECT PHARMACOKINETICS AND PHARMACODYNAMICS

Gonadal steroid hormones change throughout the life course, and those changes may result in differences in the pharmacokinetics of psychotropic drugs. In particular, periods of reproductive transition, most notably pregnancy and menopause, can significantly alter hormone levels to increase susceptibility to psychiatric illness and alter pharmacokinetics. Gonadal steroid hormones also change throughout the menstrual cycle and with the use of oral contraceptives or hormone replacement therapy. Each of these stages may result in changes in baseline psychiatric symptoms and the efficacy, safety, and tolerability of psychotropic medications.

Menstrual Cycle

Levels of FSH, LH, estrogen, and progesterone fluctuate throughout the menstrual cycle which can affect all of the pharmacokinetic phases. Gut motility and esophageal emptying are slower during the luteal phase, and gastric emptying is slower during the follicular phase. Both may lead to decreased drug concentration during the luteal and follicular phases. Absorption rates may vary widely throughout the menstrual cycle and are perhaps highest premenstrually when progesterone levels are highest. However, results are conflicting. Heart rate and systolic blood pressure also increase during the luteal phase and may lead to greater absorption. The volume of distribution may be affected by weight gain that typically occurs during the early menstrual phase and after ovulation. AGP has a peak concentration during menses, which affects the amount of unbound compound and overall plasma concentration. Estrogen, which is highest in the late luteal phase, is a substrate for some of the CYP isozymes, particularly CYP3A4, which affects drug metabolism. Finally, there is some evidence of changes in renal function during the luteal phase. Creatinine clearance increases as does vasopressin, aldosterone, and renin activity, which may either increase or reduce elimination during the luteal phase.

Use of Oral Contraceptives and Hormone Replacement Therapy

Oral contraceptives contain a combination of synthetic estrogen and progesterone to suppress ovulation and LH secretion. Similarly, hormone replacement therapy uses similar hormones to increase and normalize fluctuating hormone levels. These exogenous hormones change metabolic enzyme activity, particularly for the CYP family of isozymes involved in the metabolism of psychotropic medications. Oral contraceptives are moderate inhibitors of CYP1A2 and CYP2C19 and weak inhibitors of CYP2B6 and CYP3A4. Lower metabolism and clearance rates subsequently increase drug plasma concentration. Oral contraceptive-induced changes in metabolism and drug concentration have been observed for bupropion, imipramine, and escitalopram.

Progesterone-containing oral contraceptives also increase the receptor binding of benzodiazepines. Metabolism of oxidized benzodiazepines (triazolam, alprazolam) is inhibited by oral contraceptives, and metabolism of conjugated benzodiazepines (lorazepam, temazepam) is enhanced.

Pregnancy and Postpartum

Estrogen and progesterone increase substantially during pregnancy, particularly during the last trimester, where they increase to 10 and 50 times prepregnancy levels, respectively. The HPA axis is activated, cortisol levels increase substantially, and high levels of reproductive steroids create an immunosuppressive response. Similar to the synthetic estrogens and progestins in oral contraceptives, endogenous increases in steroid hormones during pregnancy induce or inhibit enzyme activity and can increase or decrease the metabolism of psychotropic compounds. In addition to direct induction or inhibition of CYP enzymes and UDP-glucuronosyltransferase (UGT) isoenzyme changes, numerous other changes occur during pregnancy that affect the pharmacokinetics, including increases in plasma volume, total body water, cardiac output, heart rate, body fat proportion, and regional blood flow to uterine, renal, skin, and mammary tissue. Renal plasma flow and GFR increase about 1.5-fold during pregnancy, increasing elimination. Thus, faster metabolism during pregnancy may result in lower drug plasma concentration and potentially lower efficacy, requiring an increase in dose. Conversely, however, lower CYP1A2 activity has been observed during pregnancy, resulting in decreased metabolism and clearance. Combined with higher body fat mass in pregnancy that can increase the accumulation of lipid-soluble drugs, other drugs may require a reduction in dose.

Following labor and delivery, there is a precipitous and sudden drop in estrogen and progesterone levels. Reproductive steroids are involved in the regulation of both the HPA axis and nearly every neurotransmitter, including GABA, dopamine, glutamate, serotonin, and norepinephrine. Some women are particularly sensitive to the rapid change in hormone withdrawal, as has been demonstrated in experimental studies in which euthymic women became depressed after blinded exposure and subsequent withdrawal of high doses of progesterone and estradiol. Thus, these rapid hormonal changes may affect psychopharmacology through the regulation of neurotransmitters and changes in metabolism and clearance.

Perimenopause and Menopause

Evidence of estrogen’s effect on CYP enzymatic activity supports an important role in the menopausal transition in psychopharmacologic sex differences. However, data on the pharmacokinetic differences in menopausal women are less consistent. Although some studies have found no difference in pharmacologic response, most have found that response to psychotropic medication is poorer in postmenopausal women than premenopausal women and that elevated FSH is possibly associated with poorer response. As fat mass also increases with age, concentrations of lipophilic drugs accumulate more in older adult women.

In addition, there are also important racial and ethnic differences in treatment response and metabolism of antidepressant medications. Clinical trials of antidepressant treatments have historically included few Black and Hispanic women, and as such, the optimal treatment for non-White women represents a major gap in the literature.

SEX DIFFERENCES IN THE PHARMACOKINETICS OF SPECIFIC DRUG CLASSES

Numerous studies examine pharmacokinetic sex differences of specific psychotropic drugs. Antidepressants are the most widely studied, but there is also evidence for antipsychotics, benzodiazepines, anticonvulsants, and others. Details of each of these drug classes are described below. Drugs that show sex differences in pharmacokinetics include alcohol, beta-blockers, benzodiazepines, clorazepate, clozapine, diazepam, ketamine, lithium, metoprolol, midazolam, naratriptan, olanzapine, phenobarbital, phenytoin, propranolol, rizatriptan, SSRIs, and valproic acid. Table 33.17a–1 describes metabolic routes, sex-specific enzyme activity, modulation by steroid hormones, and potential drugs affected.

Antidepressants

Most antidepressants present some extent of pharmacokinetic sex differences. Women generally have higher exposure to equal doses of antidepressants than men, indicated by higher maximal or AUC concentration, metabolism and formation of active metabolites, elimination half-life, and other pharmacokinetic indices. There are substantial and consistent differences in pharmacokinetics, but data on how much they affect clinical outcomes is limited. Clinical findings indicate that women might respond better to newer antidepressants than men, particularly to SSRIs. However, drug serum levels do not correlate well with the therapeutic effect, and evidence of clinical differences is still quite limited.

Physiologic differences in males and females lead to differences in all pharmacokinetic phases. As discussed, women have lower secretion of gastric acid, slower gut transit times, and a higher proportion of adipose tissue. There are also differences in transport proteins and metabolic enzymes between men and women, all resulting in differences in the absorption, distribution, metabolism, and excretion of antidepressants.

Most antidepressants are weak bases and more effectively absorbed under basic conditions. Because women secrete less gastric acid, antidepressants are absorbed more quickly and reach higher concentrations in women than men. This is particularly true for TCAs that have a more basic pH. Higher plasma levels and lower clearance of TCAs have been found in females compared to males.

Most antidepressants are lipophilic and have an affinity for adipose tissue. Because women have a higher percentage of body fat than men, lipophilic antidepressants may have a larger volume of distribution and lower clearance. Bupropion and trazodone are particularly lipophilic antidepressants and have a longer half-life and lower plasma concentrations in women than men.

Protein binding affects how much free drug is available and therapeutically active. AGP and alpha-globulins are lower in women, increasing the proportion of active drug in women. Many antidepressants are bound by AGP and include triazolam, amitriptyline, desipramine, imipramine, dioxepine, nortriptyline, reboxetine, chlorpromazine, thioridazine, and olanzapine. SSRIs are typically highly bound (>95%), TCAs are moderately bound (75%, 77%), and citalopram and venlafaxine have low levels of binding (<50%).

As mentioned, most psychotropic drugs are metabolized through the CYP enzyme superfamily, and expression of these CYP isoenzymes differs by sex. CYP3A4 helps metabolize SSRIs and TCAs, and CYP2C19 helps metabolize clomipramine, imipramine, and citalopram. Both CYP3A4 and CYP2C19 are more active in females than males, leading to faster clearance and lower plasma concentrations in women. Conversely, substrates of CYP1A2 and CYP2D6 clear faster in males than in females. CYP1A2 metabolizes escitalopram, TCAs, and fluvoxamine, and CYP2D6 metabolizes desipramine and mirtazapine. CYP2C9 metabolizes imipramine and amitriptyline but does not seem to vary by sex.

Table 33.17a–1.
Sex Differences in the Pharmacokinetics of Psychotropic Drugs

Pharmacokinetic Factor	Sex-Specific Activity	Modulated by Sex Hormones or Reproductive Stage	Potential Drugs Affected
Absorption			
Gastric acidity	More acidic in M than F; increases absorption of basic compounds in F	Yes, reduced acidity during pregnancy	Compounds with alkaline pH; benzodiazepines, tricyclic antidepressants
Gastric fluid flow	Greater in M than F, increases absorption in M		Any oral compound
Gastric emptying	Faster in M than F; increases plasma levels in M	Yes, gastric emptying is slower during pregnancy and slower during follicular phase	Any oral compound
Intestinal transit time	Faster in F than M, increases absorption in M and reduces plasma levels in F	Yes, transit time is slower during pregnancy and during luteal phase	Any oral compound
Distribution			
Plasma, blood, and water volume	M > F, so greater volume of distribution of water-soluble compounds in men	Yes, fluid volume increases during pregnancy, early menstrual phase, and after ovulation	Water-soluble compounds
Proportion adipose tissue	F > M, so greater volume of distribution of	Yes, increases during pregnancy	Lipophilic compounds

	lipophilic compounds in women; may increase half-life from bioaccumulation		
Cardiac output	M = F	Yes, cardiac output increases during pregnancy and during luteal phase	Any compound
Regional blood flow	M = F for adrenal, bone, brain, lung, skin, thyroid; M > F for renal, muscle; F > M for adipose, liver, heart	Yes, uterine, renal, skin, mammary blood flow increase during pregnancy	Any tissue-specific compound
Plasma Binding Proteins			
Albumin[14]	M = F	Yes, decreases during pregnancy	
Alpha-1-acid glycoprotein	M > F, thus more free (active) drug in women	Yes, decreases during pregnancy and is greatest during menses, modulated by estradiol	Highly protein-bound compounds; amitriptyline, benzodiazepines, chlorpromazine, desipramine, doxepin, fluoxetine, imipramine, nefazodone, nortriptyline, olanzapine, paroxetine, reboxetine, sertraline, thioridazine, triazolam, tricyclic antidepressants
Metabolism			

Basal metabolic rate	M > F	Yes, hepatic metabolism increases during pregnancy	Any compound
Phase I enzymes			
CYP1A2	M > F	Yes, decreased in pregnancy and with oral contraceptives (OCs)	Clomipramine, clozapine, duloxetine, fluvoxamine, olanzapine, tacrine
CYP2A6	F > M	Yes, increased with OCs	Ketamine
CYP2B6	F > M		Bupropion, ketamine
CYP2C9	M = F	Yes, increased in pregnancy	Fluoxetine, imipramine, phenytoin
CYP2C19	M = F	Yes, decreased in pregnancy and with OCs	Diazepam, citalopram, fluoxetine, fluvoxamine, hexobarbital, imipramine, methylphenobarbital, propranolol, topiramate
CYP2D6	F > = M	Yes, increased in pregnancy	Clomipramine, desipramine, fluoxetine, haloperidol, hydrocodone, imipramine, metoprolol, mirtazapine, paroxetine, propranolol, sertraline, testosterone, venlafaxine
CYP3A4	F > M	Yes; stimulated by testosterone	Alfentanil, alprazolam, carbamazepine, cortisol, diazepam, estradiol, fentanyl,

			midazolam, zolpidem
Phase II enzymes			
UDP-glucuronosyltransferases (UGTs)	M > F; M = F		Oxazepam, lemotrigine
N-acetyltransferases	F > M; M = F		Catecholamine derivatives
Methyltransferases	F > M; M > F		Dopamine, epinephrine, levodopa, norepinephrine
Elimination			
Renal blood flow and glomerular filtration rate	M > F	Yes, higher in pregnant women, increased clearance in luteal phase	Any compound

F, females; M, males.

Although antidepressants in the same drug class may be similarly metabolized, each substance has its own pharmacokinetic profile and may exhibit different sex-dependent effects depending on the specific affinity for various CYP isozymes. Pharmacokinetic sex differences for each drug are described below and in Table 33.17a–1.

Selective Serotonin Reuptake Inhibitors

CITALOPRAM. Marketing data suggests AUC concentrations of citalopram are generally higher in females compared to males. Although some studies failed to replicate, most subsequent studies have demonstrated that concentrations of citalopram and its metabolite, desmethylcitalopram, are higher in women than men. Results for ***S***-citalopram/escitalopram have been more variable.

FLUOXETINE. According to the manufacturer, sex differences have not been observed for fluoxetine. Although experimental studies show female rats metabolize fluoxetine faster than males and have higher levels of its active metabolite, giving support for the possibility of sex differences in pharmacokinetics, only one human study has shown any evidence of higher concentrations of fluoxetine in women compared to men. Although strong evidence for sex differences has not been reported, the manufacturer's dosing guidance suggests a lower starting dose if used in combination with olanzapine for those who may have factors that slow metabolism, including female sex.

FLUVOXAMINE. In phase 1 trials of fluvoxamine, maximal and AUC concentrations were robustly increased in women compared with men. At lower doses, fluvoxamine levels are significantly higher in females than males; however, this sex difference is attenuated at higher doses. This relationship suggests CYP1A2 saturation or autoinhibition by fluvoxamine at higher doses.

PAROXETINE. Higher plasma concentrations of paroxetine in women have been demonstrated in some populations, but other populations show no sex differences in concentration. Men have been shown to have a larger volume of distribution and greater dose–concentration correlation.

SERTRALINE. Few sex differences in plasma concentrations of sertraline have been demonstrated, particularly for adolescents and young adults. Women do have a better dose–concentration correlation than men for sertraline and its metabolite, desmethylsertraline, as well as longer half-lives and decreased clearance.

Serotonin and Norepinephrine Reuptake Inhibitors

DULOXETINE. Research on the sex differences of duloxetine is very sparse. According to the manufacturer, the elimination half-life of duloxetine is similar for males and females, but a meta-analysis of five studies (***n*** = 2,000 participants) found that oral clearance was lower in females, which is consistent with lower CYP1A2 activity in females. In older adult females, AUC was about 25% higher, and half-life was 4 hours longer compared to reproductive-aged females.

VENLAFAXINE. Aside from one larger study (***n*** = 1,780), which showed higher serum venlafaxine concentrations in women than men, there have not been demonstrated sex differences in plasma levels of venlafaxine. However, women do tend to have slightly higher concentrations of venlafaxine's metabolites, ***O***-desmethylvenlafaxine and ***N***-desmethylvenlafaxine, consistent with known associations of CYP3A4- and CYP2D6-mediated metabolism.

Noradrenaline and Dopamine Reuptake Inhibitors

BUPROPION. Pooled marketing data of bupropion generally found higher systemic concentration in males (***n*** = 90 females, 90 males); however, later studies have generally demonstrated slightly higher plasma concentrations, a larger volume of distribution, and longer elimination half-life in females compared to males. Bupropion is primarily metabolized by CYP2B6, and this relationship is consistent with higher CYP2B6 activity in women compared to men.

Tricyclic Antidepressants

AMITRIPTYLINE AND NORTRIPTYLINE. No pharmacokinetic sex differences were found in early studies of amitriptyline and nortriptyline, but later studies have been consistent in finding higher plasma levels and lower clearance rates of amitriptyline and nortriptyline in women compared to men. This sex difference becomes more pronounced with increasing age. Sex differences in amitriptyline and nortriptyline are primarily attributed to CYP2D6 activity.

CLOMIPRAMINE. Several studies in both humans and animals have found no pharmacokinetic sex differences in clomipramine. A few studies have found higher clomipramine plasma levels in women compared to men. However, most consistent are the studies of clomipramine metabolites. Desmethylclomipramine concentrations are higher in women, and studies of metabolic ratios support greater demethylation for women and greater hydroxylation for men.

IMIPRAMINE. Both animal and human studies of imipramine have demonstrated higher imipramine plasma levels and lower and slower clearance in females compared to males. Female rats treated with imipramine also have higher brain levels of its metabolite desipramine than male rats.

Tetracyclic Antidepressants

MIRTAZAPINE. Studies of mirtazapine consistently show that concentrations of mirtazapine and its metabolite, desmethyl mirtazapine, are higher in women than men. Women also have longer elimination half-life (37 hours in females vs. 26 hours for males), and all results have been replicated in large studies.

Serotonin-2 Antagonists and Reuptake Inhibitors

NEFAZODONE. An early study of nefazodone found no evidence for sex differences, but later studies found support for sex variation when age was taken into account. Older adult women consistently have higher concentrations of nefazodone and its metabolite, hydroxynefazodone. Nefazodone has been largely discontinued due to declining use.

TRAZODONE. Evidence for sex differences in trazodone is conflicting and may be dependent on age and menopausal status. An underpowered observational study found no sex differences in plasma levels of trazodone or its metabolites, but several additional studies have found some evidence for higher plasma concentrations and higher distribution of volume in women than men. Trazodone clearance seems to be affected by age, and elimination half-life is longer in young women than young men but shorter in older adult women than older adult men.

MONOAMINE OXIDASE INHIBITORS. Very few sex differences in monoamine oxidase inhibitors (MAOI) pharmacokinetics have been reported. Sex differences have not been reported for any of the nonselective irreversible hydrazines, isocarboxazid, and phenelzine. Manufacturing data for selegiline, a highly selective reversible MAOI indicated for Parkinson disease, showed no differences in overall AUC concentration, time to maximum exposure, or elimination half-life, but did find that female subjects have approximately a 25% decrease in maximum concentration compared to male subjects.

GABA_A Receptor–Positive Allosteric Modulators. Brexanolone was the first ever FDA-approved treatment with specific indications for PPD. It has only been studied in women for the indication of PPD but is notable as it is a novel and innovative treatment that has had a powerful response in clinical trials. Brexanolone has also been investigated for essential tremor and treatment-refractory status epilepticus. Brexanolone is a proprietary formulation of the GABA allosteric modulator and progesterone metabolite, allopregnanolone, administered intravenously over the course of 60 hours to women with PPD. Three double-blind, randomized, placebo-controlled trials displayed a swift remission of symptoms (by hour 60) that was sustained (they remained in remission at 1-month follow-up). The effect was strong even among those being treated with other antidepressant therapy, supporting consideration of brexanolone for monotherapy or adjunct treatment with other pharmacotherapies. Zulranolone, a structurally different compound from brexanolone of the same drug class, is an oral treatment for PPD approved by the FDA in August 2023. Zulranolone also has a rapid onset of action, with clinical improvement observed by day 3 of a 15 day outpatient treatment trial, and demonstrated efficacy and safety compared with placebo in phase 3 trials.

Rapid-Acting Antidepressants. Ketamine (administered intranasally as esketamine) is another novel therapeutic with rapid-acting antidepressant efficacy, which is indicated specifically for treatment-resistant depression and depressive symptoms with acute suicidal ideation or behavior. Ketamine is a noncompetitive N-methyl-D-aspartate (NMDA) receptor antagonist, and preliminary data indicate sex differences in ketamine metabolism and clearance may be dose dependent. Female rodents had greater concentrations of ketamine and norketamine in both brain and plasma and exhibited a more sensitive response to lower doses. Ketamine is moderated by estrogen via induction of CYP2A6 and CYPB6.

Antipsychotics

As with antidepressants, antipsychotics tend to be lipophilic, so the volume of distribution is greater in women than in men. Adipose tissue proportion, in conjunction with the lipophilicity of antipsychotics, may also prolong the half-life of antipsychotics and accumulate over time. Sex differences in pharmacokinetics have been observed for haloperidol and for atypical antipsychotics (clozapine, olanzapine, and sertindole). Data for pharmacokinetic sex differences of aripriprazole and quetiapine are lacking, however, limited data showing greater increases in prolactin levels in females compared to males after administration of these drugs suggest there may be differential effects.

Haloperidol. CYP2D6 is responsible for the metabolism of haloperidol. Inhibition of CYP2D6 leads to increased plasma concentrations in women compared to men.

Clozapine and Olanzapine. Clozapine is metabolized by CYP1A2. CYP1A2 activity is higher in men than women, which affects the metabolism of clozapine and olanzapine. Plasma levels of clozapine and its major metabolite, norclozapine, are higher in women than men, particularly at the beginning of treatment. Similarly, plasma levels of olanzapine are higher for women than men at the beginning of treatment. In addition, manufacturer dosing guidance for olanzapine suggests a lower starting dose if used in combination with fluoxetine for those who may have factors that slow metabolism, such as female sex.

Benzodiazepines

Benzodiazepines are highly lipophilic and, as with other psychotropic drugs, accumulate in adipose tissue, which may decrease in half-life. Benzodiazepines are highly bound to plasma proteins, with a percentage of 99%. Differences in body fat content and protein binding cause larger volumes of distribution in females than in males. CYP3A4 is highly sex dependent and responsible for the metabolism of benzodiazepines. Increased fat proportion in older adult women combined with the declining activity of CYP3A4 and decreased oxidation in older adult men mean that older women generally have lower benzodiazepine levels than older men at comparable doses.

Diazepam. Diazepam is metabolized by CYP2C19, of which women show greater activity than men. Thus, women have a faster metabolism and a lower plasma concentration of diazepam. CYP2C9 also contributes to metabolism of diazepam, but sex differences have not been observed for CYP2C9 metabolism.

Lorazepam, Temazepam, and Oxazepam. The benzodiazepines lorazepam and oxazepam are metabolized by conjugation (phase II) reactions, which are generally slower in women than men resulting in lower clearance in women than men. Oral contraceptives also facilitate conjugation metabolism, so the coadministration of benzodiazepines with hormonal oral contraceptives needs to be considered carefully, and the dose should likely be adjusted.

Triazolam and Alprazolam. The clearance of benzodiazepines metabolized by oxidation, such as triazolam and alprazolam, are similar in both sexes. Some studies show alprazolam concentrations are higher in women than men, but this is likely due to weight differences rather than sex. Oral contraceptives inhibit the metabolism of oxidized benzodiazepines, and progesterone-containing oral contraceptives enhance the receptor binding of benzodiazepines, making triazolam much more potent.

Anticonvulsants and Antiepileptics

Anticonvulsants and antiepileptic drugs show some pharmacokinetic sex differences. Higher CYP3A4 isozyme activity in women compared to men may lower drug concentration of anticonvulsants such as carbamazepine. CYP2C9 metabolizes phenytoin, another anticonvulsant, but variation by sex has not been demonstrated in CYP2C9 metabolism. Gabapentin and pregabalin are drugs used to treat epilepsy and are highly sensitive to factors such as sex. Plasma levels obtained at the usual doses can vary more than a 100-fold depending on age, gender, and coadministered medications. The concentration-to-dose ratio of pregabalin is up to 42% higher in women than in men, and significant care should be taken in dose adjustment.

EFFICACY AND SAFETY

Efficacy

Evidence that there are sex differences in the pharmacokinetics of psychotropic drugs and that these drugs are modulated by gonadal steroid sex hormones is very well established. Evidence of sex differences in clinical outcomes, including efficacy and safety profiles of specific drugs, is much more sparse. Most of the work of sex differences in efficacy of psychotropic drugs evaluated antidepressants, but there is limited evidence of differences in antipsychotics and benzodiazepines, and it is likely many more are yet to be discovered. Table 33.17a–2 describes sex differences in efficacy and safety of psychotropic drugs.

Table 33.17a–2.
Sex Differences in Exposure, Efficacy, and Safety of Psychotropic Drugs

		Reported Sex Differences		
Drug	**Drug Type**	**Exposure**	**Efficacy**	**Safety**
Antidepressants				
Selective serotonin reuptake inhibitors (SSRIs)		F > M	F may respond better than M	F at greater risk of QT prolongation
Citalopram	SSRI	F > M	Premenopausal F responded better than M to citalopram	F at greater risk of QT prolongation
Fluoxetine	SSRI	M > F; F > M for metabolites	F responded better than M to fluoxetine vs. tricyclic antidepressant (TCA)	F at greater risk of QT prolongation; reduced dose recommended in F when adjuvant to olanzapine
Fluvoxamine	SSRI	F > M, but attenuated at high doses	M responded better than F for compulsive symptoms in obsessive-compulsive disorder (OCD)	
Paroxetine	SSRI	Conflicting data		
Sertraline	SSRI	F > M	F responded better than M to sertraline vs. imipramine	F at greater risk of QT prolongation
Serotonin and norepinephrine reuptake inhibitors (SNRIs)			F responded slightly better than M	
Duloxetine	SNRI	F > M		
Reboxetine	SNRI	F > M	Premenopausal F responded better than M to reboxetine	
Venlafaxine	SNRI	F > M for metabolites	Comparable response in M and F	F at greater risk of QT prolongation
Noradrenaline and dopamine reuptake inhibitors (NDRIs)				
Bupropion	NDRI	F > M		
Tricyclic antidepressants (TCAs)			M may respond better than F, but many studies show comparable response in M and F	Protein binding may make F more susceptible to toxic effects
Amitryptiline and nortriptyline	TCA	F > M		F at greater risk of QT prolongation
Clomipramine	TCA	Conflicting data	M responded better than F for compulsive symptoms in OCD	F at greater risk of QT prolongation
Dothiepin/dosulepin	TCA	Inadequate data		
Imipramine and desimpramine	TCA	F > M	M responded better than F; M had more sustained response	F at greater risk of QT prolongation
Tetracyclic antidepressants				
Mirtazapine	Tetracyclic antidepressant	F > M		
Serotonin-2 antagonists and reuptake inhibitors (SARIs)				
Nefazodone	SARI	F > M for older ages		
Trazodone	SARI	Conflicting data		F at greater risk of QT prolongation
Monoamine oxidase inhibitors (MAOIs)			F responded better than M	
Isocarboxazid	MAOI			
Phenelzine	MAOI		Older M responded better than older F	
Selegiline	MAOI	F > M maximum concentration		No consistent differences in M and F
Gamma-aminobutyric acid type A receptor–positive allosteric modulators ($GABA_A$ RPAM)				
Brexanolone and zulranolone	$GABA_A$R PAM	Currently only indicated for postpartum depression (PPD) in women		
Rapid-acting N-methyl-D-aspartate (NMDA) antagonist				
Ketamine and esketamine	NMDA antagonist	F > M	F mice responded better and at lower doses than M mice; F and M similar in human nasal esketamine trials	Nausea, dissociation, dizziness, and vertigo more frequent in F than M; F may be more susceptible to QT prolongation, tardive dyskinesia, and hyperprolactinemia

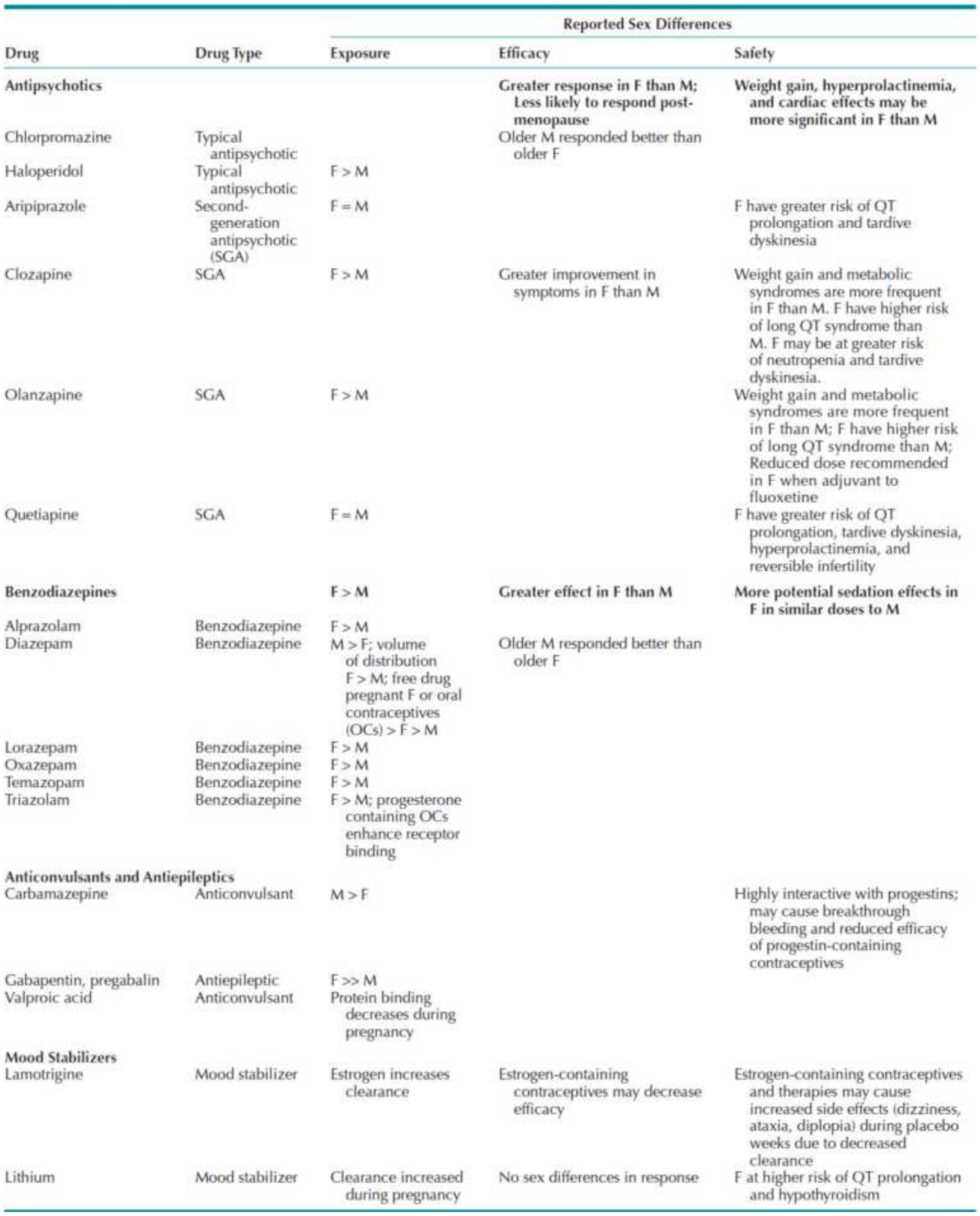

		Reported Sex Differences		
Drug	**Drug Type**	**Exposure**	**Efficacy**	**Safety**
Antipsychotics			**Greater response in F than M; Less likely to respond post-menopause**	**Weight gain, hyperprolactinemia, and cardiac effects may be more significant in F than M**
Chlorpromazine	Typical antipsychotic		Older M responded better than older F	
Haloperidol	Typical antipsychotic	F > M		
Aripiprazole	Second-generation antipsychotic (SGA)	F = M		F have greater risk of QT prolongation and tardive dyskinesia
Clozapine	SGA	F > M	Greater improvement in symptoms in F than M	Weight gain and metabolic syndromes are more frequent in F than M. F have higher risk of long QT syndrome than M. F may be at greater risk of neutropenia and tardive dyskinesia.
Olanzapine	SGA	F > M		Weight gain and metabolic syndromes are more frequent in F than M; F have higher risk of long QT syndrome than M; Reduced dose recommended in F when adjuvant to fluoxetine
Quetiapine	SGA	F = M		F have greater risk of QT prolongation, tardive dyskinesia, hyperprolactinemia, and reversible infertility
Benzodiazepines		**F > M**	**Greater effect in F than M**	**More potential sedation effects in F in similar doses to M**
Alprazolam	Benzodiazepine	F > M		
Diazepam	Benzodiazepine	M > F; volume of distribution F > M; free drug pregnant F or oral contraceptives (OCs) > F > M	Older M responded better than older F	
Lorazepam	Benzodiazepine	F > M		
Oxazepam	Benzodiazepine	F > M		
Temazopam	Benzodiazepine	F > M		
Triazolam	Benzodiazepine	F > M; progesterone containing OCs enhance receptor binding		
Anticonvulsants and Antiepileptics				
Carbamazepine	Anticonvulsant	M > F		Highly interactive with progestins; may cause breakthrough bleeding and reduced efficacy of progestin-containing contraceptives
Gabapentin, pregabalin	Antiepileptic	F >> M		
Valproic acid	Anticonvulsant	Protein binding decreases during pregnancy		
Mood Stabilizers				
Lamotrigine	Mood stabilizer	Estrogen increases clearance	Estrogen-containing contraceptives may decrease efficacy	Estrogen-containing contraceptives and therapies may cause increased side effects (dizziness, ataxia, diplopia) during placebo weeks due to decreased clearance
Lithium	Mood stabilizer	Clearance increased during pregnancy	No sex differences in response	F at higher risk of QT prolongation and hypothyroidism

Most antidepressants showed some evidence of greater efficacy in women than men, especially those that are serotonergic. This may be due to higher plasma levels in women, protein-binding specificity, or metabolic differences. There may also be pharmacodynamic differences, as women have steeper cortisol responses when exposed to SSRIs. A recent study of hospitalized patients treated for MDD (***n*** = 234) found women were more likely to achieve remission but also experience more comorbid metabolic abnormalities (i.e., hypercholesterolemia and low plasma uric acid). Greater response in females than males has been reported for SSRIs, particularly sertraline and fluoxetine, and also for serotonin and norepinephrine inhibitors (SNRIs) and MAOIs.

TCAs have been found to be more efficacious in men than women in some cases. Males responded better to imipramine and desipramine and had a more sustained response than females in both rodents and humans. Men also responded better than women with fluvoxamine and clomipramine for compulsive symptoms of obsessive-compulsive disorder.

Females are also more likely than males to experience improvement in symptoms with typical antipsychotics and the second-generation antipsychotic (SGA), clozapine, but there is not yet substantial evidence.

Most studies, however, particularly large observational studies and meta-analyses of randomized controlled trials (RCTs), find a comparable response in men and women. A large prospective sertraline utilization study (***n*** = 1,594 males, 3,858 females) found no sex differences in treatment response and also no differences in adverse effects or treatment termination. Other large observational studies have also found no sex differences in the efficacy of TCAs, MAOIs, fluoxetine, clomipramine, citalopram, paroxetine, and moclobemide. A meta-analysis of RCTs of SSRIs versus venlafaxine (***n*** = 2,045 patients with MDD) found response in men and women was comparable, and a large review of 30 RCTs of antidepressant efficacy (***n*** = 1,555 males, 2,331 females) found that TCA response was independent of sex.

There are significant challenges when evaluating sex differences in the efficacy of psychotropic drugs. As discussed, pharmacokinetics may vary across the life course and can be modulated by hormonal fluctuations. Differences have been demonstrated in the literature in that premenopausal and younger females have a better response to some antidepressants than postmenopausal females and males. Other studies have demonstrated a better response to diazepam in older men compared to older women and a better response to chlorpromazine in older men compared to women.

Studies of efficacy also generally fail to take adverse events and early discontinuation into account. Adverse effects are reported more frequently in women, which may result in early dropout. If a greater proportion of women who stayed in the study experienced a better response to a certain drug, then differences in effect by sex could be attenuated. Finally, current studies of sex differences in psychotropic drug efficacy are underpowered. Because genetic polymorphisms lead to much of the interindividual variability in pharmacokinetics, these studies are significantly underpowered to identify sex differences. When a weak effect of sex is identified, it may be dismissed as not being clinically significant, as there is also limited statistical power to detect sex differences in pharmacokinetic and safety profiles.

Safety

Adverse effects of pharmaceuticals, and particularly psychotropic drugs, are typically greater in women than men. Data from the national Adverse Events Reporting System administered by the U.S. FDA suggest that women experience more adverse events than men and that the adverse events are more serious. In a review by the U.S. Government Accounting Office of 10 drugs removed from the market over a 3-year period from 1997 to 2000, 8 of the 10 drugs were withdrawn due to a greater risk of adverse events in women. Since that time, the FDA has significantly improved surveillance of sex differences in adverse outcomes, particularly for potential QT-interval prolongation and drug-induced liver injury, for which women are at greater risk.

Women may experience more adverse drug events for a number of reasons. Due to differences in pharmacokinetics, females often experience a greater volume of distribution of lipophilic drugs, slower elimination, and differences in metabolism depending on the specific drug. Due to protein binding, the free fraction of many compounds is often greater in women, leaving a greater proportion of the active drug in the body. All of these differences may lead to exposure to higher drug concentrations in women compared to men, despite similar prescribed dosages. Most antidepressants are highly bound compounds and therefore produce a higher free fraction of drugs in women and may increase the risk of QT-interval prolongation, for which women are more susceptible. Benzodiazepines and other sedatives/hypnotics, which are lipophilic, typically have a more potent effect in women, and zolpidem (a nonbenzodiazepine hypnotic) was the first drug for which the FDA recommended sex-based dosing.

Women may also be more likely to experience adverse drug interaction effects. Overall, women take more medications and are more likely than men to be taking antidepressants, mood stabilizers, analgesics, contraceptives, and hormone replacements. In a large European study of psychotropic drug use by inpatients diagnosed with MDD between 2001 and 2017 (***n*** = 44,418), the use of monotherapy was more common among male patients, while female patients had a higher risk of being treated with four or more medications. Polytherapy of two or more antidepressants or the combination of an antidepressant and antipsychotic was similar between men and women, but women were more likely to be treated with antidepressants in combination with tranquilizers, hypnotics, or antiepileptics. Differences in protein binding and metabolism may make women more susceptible to drug interactions, particularly for psychotropic drugs, in which one may be used as adjuvant therapy to another. CYP2D6 inhibition is particularly of concern with psychotropic drugs. In addition, simultaneous administration of multiple drugs with elevated protein binding may cause one to be displaced, increasing the free amount of a drug and greater possibility of toxicity. Most SSRIs are highly bound compounds, as are anticonvulsants. Coadministration of psychotropic drugs with hormonal contraceptives and hormone replacement therapies must also be carefully considered, as oral contraceptives inhibit oxidative metabolism and enhance conjugation metabolism. For example, progesterone-containing oral contraceptives enhance the receptor binding of benzodiazepines and can significantly increase the effective dose of triazolam.

Overall, reported side effects are greater in women, but research investigating sex differences in the safety profiles of psychotropic drugs is still limited. Most comprehensive is the evaluation and review of SGAs, which found that weight gain, hyperprolactinemia, cardiac effects, metabolic syndromes, and long QT interval are more frequent in women taking clozapine and olanzapine than men.

CONCLUSION

There is strong evidence for sex differences in psychopharmacology, yet biologic sex is a variable that is far too often neglected, both in research and in clinical care. In general, women are more likely to be prescribed psychotropic medications, use a greater number of medications, have higher drug concentrations, and experience more adverse effects. Currently, most drugs lack sex-differentiated dosage data, and except for zolpidem, current guidelines and pharmaceutical manufacturer prescribing information do not provide dose adjustment for men and women, despite evidence that women should frequently receive lower doses than men. In addition to biologic sex, other factors such as body weight, oral contraceptive use, hormonal therapy, reproductive stage, and race/ethnicity may all contribute to differences in efficacy, safety, and tolerability of medications by sex. Continued research in sex differences in psychopharmacology is needed so that prescribing guidance and clinical care can evolve to meet the needs of each individual, rather than a one-size-fits-all approach.

FURTHER READINGS

Bigos KL, Pollock BG, Stankevich BA, Bies RR. Sex differences in the pharmacokinetics and pharmacodynamics of antidepressants: an updated review. ***Gend Med***. 2009;6(4):522–543.

Ceresa A, Esposito CM, Surace T, et al. Gender differences in clinical and biochemical parameters of patients consecutively hospitalized for unipolar depression. ***Psychiatry Res***. 2022;310:114476.

Damoiseaux VA, Proost JH, Jiawan VCR, Melgert BN. Sex differences in the pharmacokinetics of antidepressants: influence of female sex hormones and oral contraceptives. ***Clin Pharmacokinet***. 2014;53(6):509–519.

Duffy KA, Epperson CN. Evaluating the evidence for sex differences: a scoping review of human neuroimaging in psychopharmacology research. ***Neuropsychopharmacology***. 2022;47(2):430–443.

Farkouh A, Riedl T, Gottardi R, Czejka M, Kautzky-Willer A. Sex-related differences in pharmacokinetics and pharmacodynamics of frequently prescribed drugs: a review of the literature. ***Adv Ther***. 2020;37(2):644–655.

Hamilton J, Yonkers K. Sex differences in pharmacokinetics of psychotropic medication, part I: physiological basis for effects. In: Jesnvold M, Halbreich U, Hamilton J, eds. ***Psychopharmacology and Women: Sex, Gender, and Hormones***. American Psychiatric Press; 1996:11–42.

Jenkins MR, Munoz MA, Bak D, et al. Food and drug administration beyond the 2001 government accountability office report: promoting drug safety for women. ***J Womens Health (Larchmt)***. 2021;30(7):927–934.

Khan A, Brodhead AE, Schwartz KA, Kolts RL, Brown WA. Sex differences in antidepressant response in recent antidepressant clinical trials. ***J Clin Psychopharmacol***. 2005;25(4):318–324.

Kokras N, Dalla C, Papadopoulou-Daifoti Z. Sex differences in pharmacokinetics of antidepressants. ***Expert Opin Drug Metab Toxicol***. 2011;7(2):213–226.

Kornstein SG, Schatzberg AF, Thase ME, et al. Gender differences in treatment response to sertraline versus imipramine in chronic depression. ***Am J Psychiatry***. 2000;157(9):1445–1452.

Labbé L, Sirois C, Pilote S, et al. Effect of gender, sex hormones, time variables and physiological urinary pH on apparent CYP2D6 activity as assessed by metabolic ratios of marker substrates. ***Pharmacogenetics***. 2000;10(5):425–438.

LeGates TA, Kvarta MD, Thompson SM. Sex differences in antidepressant efficacy. ***Neuropsychopharmacology***. 2019;44(1):140–154.

Loebstein R, Lalkin A, Koren G. Pharmacokinetic changes during pregnancy and their clinical relevance. ***Clin Pharmacokinet***. 1997;33(5):328–343.

Marazziti D, Baroni S, Picchetti M, et al. Pharmacokinetics and pharmacodinamics of psychotropic drugs: effect of sex. ***CNS Spectr***. 2013;18(3):118–127.

Meltzer-Brody S, Colquhoun H, Riesenberg R, et al. Brexanolone injection in post-partum depression: two multicentre, double-blind, randomised, placebo-controlled, phase 3 trials. ***Lancet***. 2018;392(10152):1058–1070.

Meltzer-Brody S, Kanes SJ. Allopregnanolone in postpartum depression: role in pathophysiology and treatment. ***Neurobiol Stress***. 2020;12:100212.

Nicolson TJ, Mellor HR, Roberts RRA. Gender differences in drug toxicity. ***Trends Pharmacol Sci***. 2010;31(3):108–114.

Pitychoutis PM, Zisaki A, Dalla C, Papadopoulou-Daifoti Z. Pharmacogenetic insights into depression and antidepressant response: does sex matter? ***Curr Pharm Des***. 2010;16(20):2214–2223.

Quitkin FM, Stewart JW, McGrath PJ, et al. Are there differences between women's and men's antidepressant responses? ***Am J Psychiatry***. 2002;159(11):1848–1854.

Riecher-Rössler A. Sex and gender differences in mental disorders. ***Lancet Psychiatry***. 2017;4(1):8–9.

Rubinow DR, Moore M. Sex-dependent modulation of treatment response. ***Dialogues Clin Neurosci***. 2004;6(1):39–51.

Rubinow DR, Schmidt PJ. Sex differences and the neurobiology of affective disorders. ***Neuropsychopharmacology***. 2019;44(1):111–128.

Schiller CE, Johnson SL, Abate AC, Schmidt PJ, Rubinow DR. Reproductive steroid regulation of mood and behavior. ***Compr Physiol***. 2016;6(3):1135–1160.

Schiller CE, Schmidt PJ, Rubinow DR. Allopregnanolone as a mediator of affective switching in reproductive mood disorders. ***Psychopharmacology (Berl)***. 2014;231(17):3557–3567.

Schweizer-Schubert S, Gordon JL, Eisenlohr-Moul TA, et al. Steroid hormone sensitivity in reproductive mood disorders: on the role of the GABAA receptor complex and stress during hormonal transitions. ***Front Med (Lausanne)***. 2021;7:479646.

Seifert J, Führmann F, Reinhard MA, et al. Sex differences in pharmacological treatment of major depressive disorder: results from the AMSP pharmacovigilance program from 2001 to 2017. ***J Neural Transm (Vienna)***. 2021;128(6):827–843.

Soldin OP, Mattison DR. Sex differences in pharmacokinetics and pharmacodynamics. ***Clin Pharmacokinet***. 2009;48(3):143–157.

Songtachalert T, Roomruangwong C, Carvalho AF, Bourin M, Maes M. Anxiety disorders: sex differences in serotonin and tryptophan metabolism. ***Curr Top Med Chem***. 2018;18(19):1704–1715.

Sramek JJ, Murphy MF, Cutler NR. Sex differences in the psychopharmacological treatment of depression. ***Dialogues Clin Neurosci***. 2016;18(4):447–457.

U.S. Government Accounting Office. Drug Safety: Most Drugs Withdrawn in Recent Years Had Greater Health Risks for Women. Published 2001. Accessed June 22, 2022. https://www.gao.gov/products/gao-01-286r

Wisner KL, Perel JM, Wheeler SB. Tricyclic dose requirements across pregnancy. ***Am J Psychiatry***. 1993;150(10):1541–1542.

Yonkers KA, Brawman-Mintzer O. The pharmacologic treatment of depression: is gender a critical factor? ***J Clin Psychiatry***. 2002;63(7):610–615.

Yonkers KA, Kando JC, Cole JO, Blumenthal S. Gender differences in pharmacokinetics and pharmacodynamics of psychotropic medication. ***Am J Psychiatry***. 1992;149(5):587–595.

33.17b Reproductive Hormonal Therapy: Theory and Practice

DAVID R. RUBINOW, M.D., AND PETER J. SCHMIDT, M.D.

The 19th-century medical literature contained several presumptions about the pathophysiology of mood disorders in women, largely based on anecdotal observations of reproductive endocrine dysfunction (e.g., amenorrhea) in psychiatrically ill women. These inferences, in turn, were translated into therapeutics. Thus, numerous reports also documented the beneficial effects on mood and behavior associated with medical or surgical manipulations of a woman's reproductive function. In addition to their interest in the role of reproductive function in psychiatric disorders, medical researchers in the late 19th century also developed a general interest in glandular secretions that potentially could modify physiology. Early experiments in humans with preparations derived from animal glands, including the thyroid, adrenal, ovary, testes, and spleen, provided the directions for the developing fields of endocrinology and immunology. In his seminal lecture in 1889, Charles-Édouard Brown-Séquard not only originated the formal study of endocrinology, he also reported the potential psychotropic effects of gonadal secretions. His description of the invigorating effects of testicular extracts resulted in a brief but widespread use of these compounds (as well as other interventions intended to increase the body's testicular or seminal fluid levels) in a variety of therapeutic settings. In addition to extracts of both testes and ovaries, investigators experimented with extracts of the thyroid and adrenal glands in psychiatric patients. This interest in the therapeutic use of thyroid and adrenal extracts enjoyed more enduring success with the recognition of their beneficial effects in the treatment of myxedema and Addison disease, respectively. However, the widespread use of gonadal extracts ultimately met with criticism due to some disappointing results and the realization that these extracts were not the "fountain of youth." Their subsequent clinical use was largely restricted to the treatment of menopause-related hot flushes (ovarian extracts). Nonetheless, the observation that ovarian extracts inhibited ovulation in rabbits was sufficient to maintain medical interest in the potential role of gonadal extracts in the control of human fertility. This process of promise and disappointment, passionate advocacy and evidence-based caution, is the story of both the birth and the ongoing existence of endocrinology. Many of the claims of the organotherapists were refuted by observations that most of the endocrine preparations were inactive when administered orally. Nonetheless, the so-called "practical men," the practitioners whose therapeutic zeal was not about to be extinguished by study results to the contrary, continued to claim the efficacy of endocrine preparations for all manner of illnesses. An additional source of the pendular swing of practice, and one particularly noteworthy for reproductive preparations, was the recognition of previously undiscovered side effects, a phenomenon that is particularly relevant for the therapeutic use of estradiol, the primary active substance in ovarian extracts. Efforts to improve techniques of extraction and purification of biologically active substances from the gonads were fueled by the hope that factors regulating reproductive function would be identified. By the late 1920s and early 1930s, many gonadal steroids were isolated and characterized, including estrone, estradiol, progesterone, and several androgens. Moreover, during the next 10 years, chemists identified modifications of the steroids that could alter their absorption (i.e., acetylation) and potency (e.g., the addition of an ethinyl group or removal of C-19 methyl groups). These findings initiated a resurgence in the medical use of gonadal steroids. Estrogen replacement therapy was used to treat menopausal symptoms in the 1950s, and oral contraceptives (i.e., mestranol and norethynodrel [Enovid]) were developed and first approved by the FDA in 1960. The use of gonadal steroids in women once again was widespread, and several papers were published reporting the therapeutic benefits of these compounds in involutional melancholia, premenstrual syndrome, and postpartum depression (PPD). However, their widespread usage was restricted—the pendular swing—after reports in the 1970s of increased rates of endometrial cancer secondary to unopposed estrogen replacement and increased rates of thrombosis and pulmonary emboli in women taking oral contraceptives. In the last two decades of the 20th century, the popularity of gonadal steroid therapy resumed due in part to the reports of the enhanced safety of both hormone replacement therapy (HRT) (due to the protective effect on the endometrium of the addition of progesterone) and oral contraceptives (due to the lower doses of estradiol and new progestins) and to the reported beneficial/disease protective effects of gonadal steroids on several tissues including the musculoskeletal system, cardiovascular system, and central nervous system (CNS).

The discovery of several other factors involved in the control of reproduction also led to new drug development. The decapeptide gonadotropin-releasing hormone (GnRH) was isolated and sequenced in the 1970s, and the observation that continuous GnRH infusion resulted in the downregulation of pituitary GnRH receptors led to the development of several GnRH agonists. These agonists were used to suppress reproductive endocrine function in a variety of medical conditions, including hormone-dependent cancers and endometriosis. In combination with gonadal steroids, preparations of GnRH agonists provided physicians with a strategy to control reproductive function and regulate the exposure to specific gonadal steroids without resorting to surgery. Thus physicians could selectively eliminate and/or replace reproductive factors considered to be the potential source of a medical or psychiatric problem.

More recently, however, reports of increased risk of breast cancer and, particularly, the Women's Health Initiative (WHI) demonstration of increased, not decreased, risks of both heart disease and dementia after hormone therapy (HT) have resulted once more in a mass exodus from reproductive HT. Subsequent findings suggest that the pendulum again may have swung too far (this time, in the direction of avoiding HT). Nonetheless, the conclusions that follow about the practical and potential applications of HT in psychiatric disorders must be viewed in the evolving context in which HT now resides.

Reproductive steroids are powerful and complex neuroregulators. If one is to knowledgeably, responsibly, and creatively employ reproductive steroids in the psychotherapeutic armamentarium, some familiarity with the structure and related complexity of activities of these compounds is required.

STRUCTURE AND ACTIONS

Gonadal steroids, like all steroid hormones, are derivatives of cholesterol, and all share a core four-ring structure. Despite the wide array of steroid hormones, the synthesis of individual steroids is performed by a much smaller number of enzymes with multiple sites of action in the hormone synthetic cascade. Indeed, during the 1980s, the cloning of the complementary DNAs (cDNAs) and genes of several steroidogenic enzymes led to the realization that multiple different steroids could serve as substrates for the same steroidogenic enzyme in most tissues. In the past two decades, there has been an explosion of findings that have demonstrated the enormous complexity of steroid synthesis. These findings include the identification of multiple isozymes for steroid synthetic enzymes (including at least 14 different isozymes of 17-beta hydroxysteroid dehydrogenase encoded by several different genes) and the presence of a range of alternate synthetic pathways leading to steroid production in both health and disease states (e.g., the "backdoor pathway" from progesterone to dihydrotestosterone without the usual androgen intermediates). In humans, plasma low-density lipoproteins are the major source of tissue cholesterol used for synthesizing steroids. The first step in steroid synthesis takes place in the mitochondria. The mechanism by which cholesterol is transported and deposited onto the outer mitochondrial membrane remains only partially understood. However, in steroidogenic cells, steroidogenic acute regulatory protein (STAR) and other members of the STAR-related lipid transfer family of proteins control the availability of cholesterol within the mitochondria, where it is cleaved in a series of steps by the P450 side chain cleavage enzyme (P450scc) to produce pregnenolone. Pregnenolone then serves as the precursor of the family of steroid hormones, individual members of which, as stated above, are generated through the actions of a relatively small number of enzymes with multiple sites of action (see Fig. 33.17b–1). The resulting end products of this cascade are determined by the tissue in which the metabolism is occurring and the enzymes present in that tissue. For example, testosterone in both men and women may be the end product and act directly at the androgen receptor, or it may be reduced to a form with greater affinity for the androgen receptor (dihydrotestosterone), converted to a form with less affinity for the androgen receptor (androsterone—also a neurosteroid active at the GABA receptor), or aromatized to estradiol and act through the estrogen receptor (ER). Recently, a new level of complexity has been introduced by the discovery of tissue-specific variations in the activity of steroidogenic enzymes. For example, 17-beta hydroxysteroid dehydrogenase exists in both oxidative and reductive forms: the type 2 isoform in the gut is oxidative and converts estradiol to estrone (thus “inactivating” orally administered estradiol), while the type 1 isoform in the ovary and placenta is reductive and “activates” estradiol via its conversion from the less potent estrone. Further, a genetic variation in 5-alpha reductase type 1, an enzyme that converts testosterone to dihydrotestosterone and is localized in skin and the CNS, has been associated with enhanced hirsutism in women with polycystic ovarian syndrome. Thus, polymorphic variants of steroidogenic enzymes can produce tissue-specific alterations in exposure to steroids with accompanying alterations in phenotype. Moreover, as an example of the multifunctional roles of steroidogenic enzymes, 5-alpha reductase also is the rate-limiting step in the conversion of progesterone to its neurosteroid metabolite allopregnanolone, which is active at the GABA receptor. Finally, given the structural similarities shared across different types of steroid hormones, steroids are classified by the particular receptor that mediates their actions rather than the structural elements of the hormone or the tissue of origin (see bioidentical hormone section below). By this classification, reproductive steroids are those that act via the androgen receptor, ER, or progesterone receptor (PR) systems.

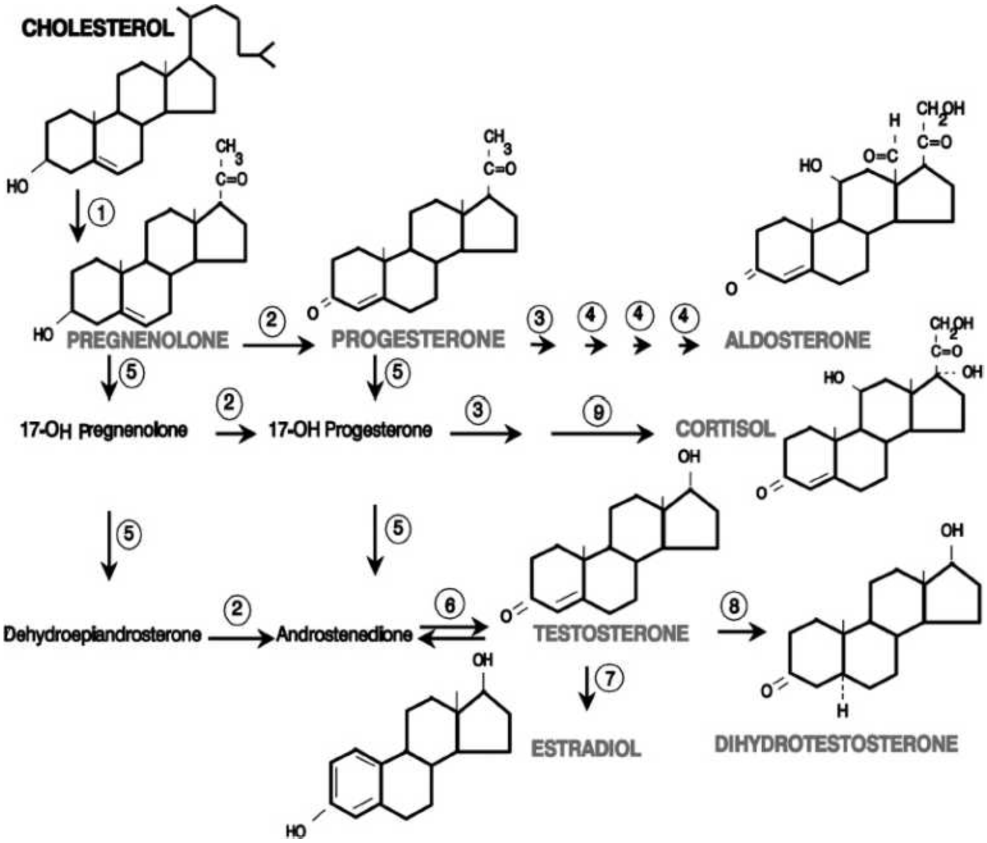
CHOLESTEROL
HO
①
PREGNENOLONE
②
PROGESTERONE
③ ④ ④ ④
ALDOSTERONE
⑤
17-OH Pregnenolone
17-OH Progesterone
③
⑨
CORTISOL
Dehydroepiandrosterone
Androstenedione
⑥
TESTOSTERONE
⑦
ESTRADIOL
⑧
DIHYDROTESTOSTERONE

FIGURE 33.17b–1. Synthetic pathways for steroid hormones. Circled numbers identify synthetic enzymes: 1, cytochrome P450 (CYP) 11A (P450 side chain cleavage enzyme); 2, 3β-hydroxysteroid dehydrogenase; 3, CYP21 (21-hydroxylase); 4, CYP11B2 (11β-hydroxylase, 18-hydroxylase, 18-oxidase); 5, CYP17 (17α-hydroxylase, 17,20-lyase); 6, 17β-hydroxysteroid dehydrogenase (or oxidoreductase); 7, aromatase; 8, 5α-reductase; 9, CYP11B1 (11β-hydroxylase). (Reprinted with permission from the American Journal of Psychiatry, 153:8 (Copyright © 1996). American Psychiatric Association. All Rights Reserved.)

The classic genomic action of steroid hormones occurs after the steroid binds (activates) its intracellular receptor, which, after undergoing phosphorylation, binds (usually as a dimer) to a hormone response element on a gene and directs or modifies transcription of that gene (see Fig. 33.17b–2). Note that in contrast to the membrane-bound receptors of neurotransmitters and neuropeptides, which act on the genome through a signal transduction cascade, the steroid hormone receptors are intracellular and are transcription factors that can directly bind to DNA. Further, once the ligand–receptor complex binds to the hormone response element (which is located on or near the promoter of the gene), the result can be either the initiation or repression of transcription of messages coding for an enormous array of proteins, including synthetic and metabolic enzymes for neurotransmitters, neuropeptides, receptor proteins, transporters, and second messengers; literally thousands of genes are so regulated.

Alas, while accurate, this mechanistic description is glaringly inadequate, and it is in the missing complexity that the explanation for so many ostensibly idiosyncratic responses to hormones lies. First, an additional source of transcriptional variance is found in the different forms of the gonadal steroid receptors. For example, isoforms (structurally varied products of the same gene) of the androgen and progesterone receptors exist that have different transcriptional effects, and two separate ERs, alpha and beta, have been identified and coded for by genes on the 6th and 14th chromosomes, respectively. These ERs have different distributions in the brain, different transcriptional actions, and even possess additional isoforms (e.g., insertional and deletional variants of ER beta). In addition, recent studies have identified splice variants of steroid receptors present in some disease states, including a splice variant in the androgen receptor that lacks a ligand-binding domain but is constitutively active (and, therefore, acts on the androgen receptor in the absence of steroid) and contributes to the development of hormone-refractory prostate cancer. In addition to the intracellular location of steroid receptors, receptors have been identified on the cell membrane; again, multiple forms of steroid membrane receptors have been identified. For example, membrane progesterone receptors exist in the following forms: the progesterone-adiponectin Q receptor (five subtypes), the membrane component progesterone receptor (three subtypes), and the classical PR. Second, the activated steroid receptor regulates transcription by binding protein intermediaries called coregulators. Some coregulators stimulate transcription while others inhibit it. The activities of these coregulators are tissue specific, and different types of coregulators interact with specific isoforms of the steroid receptors. Thus, the relative expression of coregulators and receptor isoforms within different tissues could underlie, in part, opposite effects of the same hormone in different tissues. Approximately 350 of these coregulators have been identified, with coregulators usually part of assemblies consisting of up to eight different regulators. Since the coregulator function is altered by covalent modification of its constituent amino acid residues, it has been estimated that there are 10^{13} possible different functional states that can be created in response to hormone activation. Third, steroid receptors can also indirectly associate with DNA through protein: protein interactions with other transcription factors—a process called **tethering**—to regulate the transcription of genes that do not have classical receptor binding sites—response elements—in the DNA.

Fourth, hormones display acute nongenomic actions; that is, they can, in a receptor isoform-specific fashion, allosterically modify the activity of neurotransmitter-gated ion channels, can directly bind and regulate ion channels, and can initiate signal transduction through the binding membrane rather than cytoplasmic or nuclear receptors. Fifth, hormone receptors can be activated, even in the absence of hormones, by a variety of neurotransmitters and growth factors. In this fashion, environmental stressors may amplify the response to even minimal levels of a hormone signal. Sixth, hormones may influence each other's activity by competing for cofactors or by producing opposite effects on genes regulated through integrator proteins. What are the practical implications of this biologic complexity? First, hormones of a given class (i.e., different hormonal preparations) are not going to produce similar effects on all tissues. For example, medroxyprogesterone (MPA) differs from progesterone in at least three ways: first, MPA interferes with the activity of the neurosteroid, $GABA_A$ receptor modulating metabolite of progesterone—allopregnanolone; second, MPA lacks the anti-inflammatory/neuroprotective effects of both progesterone and allopregnanolone; and third, MPA inhibits the enzyme that converts pregnenolone to progesterone (i.e., 3-beta hydroxysteroid dehydrogenase) thereby further decreasing substrate for the production of endogenous neurosteroids. At a cellular level, it is unlikely that the equine estrogens used so frequently in HRT produce the same conformational changes in the ER as estradiol and hence would have a different profile of actions. Indeed, the genetic effects of conjugated equine estrogens are not identical to those of estradiol. Second, hormone plasma levels are only part of the story. Hormone effects are context specific and reflect the constituents of the milieu that bathes the cell as well as the specific constellation of expressed proteins within the cell. Depending on one's genetic background, then, one might expect that the same hormone level would elicit a range of possible effects. This has been clearly demonstrated in animal studies in which exogenous estradiol exerts 100-fold differences in effects on reproductive development across rodent strains, accounting for far more of the variance than the plasma level of estradiol achieved. Third, recognition of the critical nature of intracellular proteins and hormone signaling/transcription suggests a strategy for future drug development, namely designing drugs that will facilitate or disrupt binding interactions between the receptor and specific cofactors. This strategy has already borne fruit in the development of selective estrogen receptor modulators (SERMs), as well as those for progesterone (SPRMs) and androgen (SARMs) receptors.

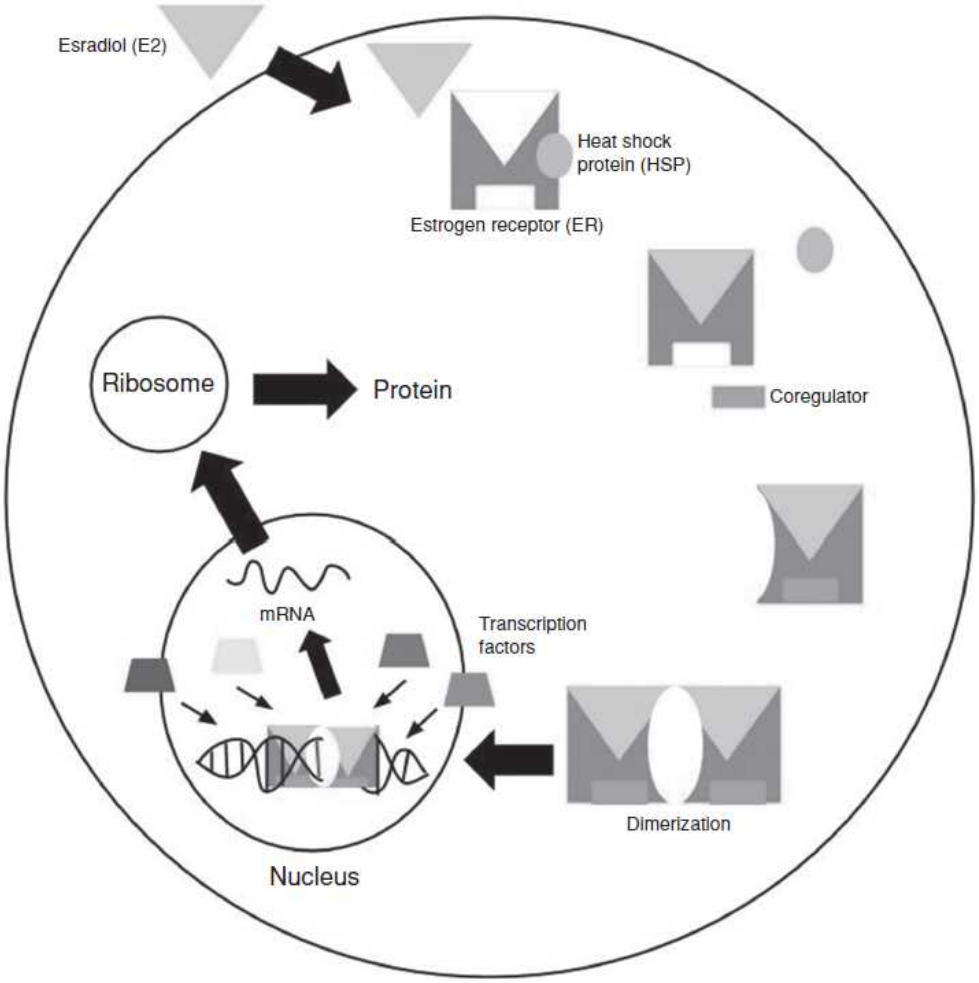

FIGURE 33.17b–2. Genomic action of steroid hormones. Estradiol acts by binding (activating) its intracellular receptor, which dissociates from heat shock protein. The activated receptor undergoes a conformational change that facilitates the binding of a coregulator (in this case, a coactivator) with enzymatic properties. The receptor dimerizes, and the receptor/coregulator complex both binds the DNA hormone response element and recruits a variety of transcription factors. The DNA is then remodeled to facilitate transcription of mRNA, which is exported to the cytoplasmic ribosome for protein synthesis.

WHY REPRODUCTIVE HORMONAL THERAPY?

It is biologically plausible that reproductive steroids might display psychotropic properties, as their pattern of neuroregulatory effects is highly convergent with those presumed to underlie psychotropic efficacy.

Indeed, wherever one finds a system believed to play a role in the etiology or treatment of depression, modulatory effects of gonadal steroids are observed as well. The neurotransmitter systems implicated in depression—serotonin, norepinephrine, dopamine, acetylcholine, gamma-amino butyric acid (GABA), glutamate—all are regulated by estradiol, as are other relevant processes like neuroplasticity, neuronal/glial survival (e.g., brain derived neurotrophic factor [BDNF], B-cell lymphoma 2 [Bcl-2]), signal transduction (e.g., mitogen-activated protein kinases [MAPK], Protein kinase B

[AKT]), and neuronal excitability (e.g., calcium and potassium ion channels). The effects of estradiol on these systems are similar to those of antidepressants and mood stabilizers (e.g., increasing the growth factors [BDNF] and signal transduction proteins [extracellular signal-regulated kinases (ERK), AKT] that promote neuroplasticity/neuronal survival and that are decreased in depression). Similarly, many areas of the brain that have been implicated in depression, including the prefrontal cortex (PFC), amygdala, hippocampus, striatum, and thalamus, contain ERs and are modulated by estradiol. More to the point, reproductive steroids regulate activity in multiple networks implicated in depression (e.g., salience, affective control, reward circuitry). Finally, the importance of gonadal steroids in affective regulation is suggested by studies demonstrating the role of estradiol in stress amplification, ER beta in anxiety in female rats, and ER alpha in arousal.

Progesterone is metabolized to the neurosteroid allopregnanolone, the levels of which have been shown to be decreased in depression and increased during treatment with a selective serotonin reuptake inhibitor (SSRI). Indeed, SSRIs increase the activity of the neurosteroid synthetic enzyme 3-alpha-hydroxysteroid oxidoreductase, and it has been hypothesized that the resultant increase in allopregnanolone may underlie the therapeutic efficacy of SSRIs in premenstrual dysphoria. Progesterone, through allopregnanolone, modulates both the conformation and the ligand sensitivity of $GABA_A$ receptors, with a role for allopregnanolone implicated in both the genesis and treatment of anxiety (depending upon the time course of administration) and in the sensitivity of $GABA_A$ receptor to the anxiolytic benzodiazepines. Notably, allopregnanolone, unlike benzodiazepines, modulates extrasynaptic (as well as synaptic) $GABA_A$ receptor activity, thus regulating tonic inhibition, critical for the “tuning” of neuronal excitability in the brain. Finally, allopregnanolone also exerts neuroprotective effects by reducing the expression of proapoptotic proteins and apoptotic DNA fragmentation, thereby reducing cell death and gliosis, which have been associated with depression. Neuroprotection may be mediated by immune regulation in depression, and both progesterone and allopregnanolone reduce the expression of the “proinflammatory” cytokine tumor necrosis factor (TNF)-alpha, which is elevated in individuals with major depression. Further, allopregnanolone has been shown to dramatically inhibit proinflammatory signaling through toll-like receptor pathways both peripherally and in the brain. Thus, allopregnanolone (or progesterone) may exert antidepressant effects by reducing the physiologic impact of stress, promoting neuroprotection, and protecting against the immune activation and cytokine hypersecretion associated with major depressive disorder. As such, both progesterone and therapies that target progesterone metabolism might be expected to modulate mood.

TYPES OF REPRODUCTIVE HORMONE THERAPY IN AFFECTIVE ILLNESS

Reproductive hormone therapies serve two principal functions. First, they potentially may replace traditional pharmacotherapies if the latter have proved ineffective or are accompanied by unacceptable side effects. Second, some reproductive hormone therapies can also serve as short-term trials to inform patients and their health care providers about possible links between the secretion of a particular hormone and the presence or exacerbation of symptoms (e.g., GnRH agonist–induced ovarian suppression in the menstrual cycle–modulated affective disorders, as described below).

Hormone Suppression in Premenstrual Dysphoric Disorder

A variety of hormonal therapies have been employed in premenstrual dysphoric disorder (PMDD). Currently, the best evidence of efficacy exists for GnRH agonists, compounds that reversibly suppress the pituitary secretion of the gonadotropins, follicle stimulating hormone (FSH), and luteinizing hormone (LH), and hence turn off ovarian steroid secretion. GnRH-induced ovarian suppression works in about two-thirds of women with PMDD, the same percentage (although not necessarily the same individuals) as are successfully treated by SSRIs. GnRH therapy should be considered for those women who do not respond to SSRIs, particularly if they have severe symptoms. While few predictors of therapeutic response to GnRH agonists exist, one report suggests that the presence of symptoms in the follicular phase predicts a lack of efficacy (although the women in this report would not meet the criteria for PMDD and more likely experienced a premenstrual exacerbation of a chronic depressive illness). Because of the largely unknown risk of compromising fertility and of nonreversible reduction in bone mass, this treatment should be reserved for those who do not intend to subsequently have children and should be administered only in consultation with a gynecologist and with full consideration of risks and benefits. Recently, concerns over the safety of long-term GnRH agonist–induced gonadal suppression in both fertility and bone health have been offset in part by reports of the return of normal reproductive function and bone mass in the majority of children with precocious puberty who were treated with these agents for several years. Nonetheless, prudent clinical practice would entail the use of GnRH agonists for not greater than 3 months without consultation with a woman's gynecologist and without consideration of the addition of bone-sparing adjunctive estradiol (which, if given continuously, would not be expected to reverse the efficacy of the GnRH agonist). One additional advantage to a therapeutic trial of GnRH agonist is that the results can determine the role of cyclical variation of steroid secretion in PMDD. If the patient does not respond to GnRH agonist–induced suppression (nonresponse is reported in about 30% of women with PMDD despite adequate ovarian suppression), it is unlikely that a reproductive hormone etiology is present or that a hormonal intervention will be successful. Other hormones studied as potential treatments for PMDD include progesterone and estrogen. Despite early claims to the contrary, the majority of studies overwhelmingly refute the efficacy of progesterone in PMDD.

Recently published data suggest that it is the change in hormone levels—rather than the levels themselves—that trigger the symptomatic state in PMDD. This is consistent with the possibility that it is the shortened pill-free interval that accounts for the efficacy of the drospirenone-containing oral contraceptive Yaz in PMDD. It should be noted, however, that several reports fail to demonstrate the efficacy of even continuous administration of oral contraceptives in PMDD. Three other studies are consistent with the hypothesis that differential sensitivity to changing hormone levels precipitates symptoms in women with PMDD. First, stabilizing neurosteroid levels across the menstrual cycle by inhibiting 5-alpha reductase type 1 (the rate-limiting enzyme for the conversion of progesterone to allopregnanolone) with the medication dutasteride results in the elimination of PMDD symptoms. Second, PMDD symptoms were significantly reduced following luteal phase administration of the isomeric antagonist of allopregnanolone (isoallopregnanolone; Sepranolone), which would prevent the effects of the rise in allopregnanolone during the luteal phase. Finally, efficacy specifically for mood (rather than physical) symptoms of PMDD was recently demonstrated for ulipristal acetate, an SPRM, which reduces both progesterone signaling and levels (through anovulation). These findings collectively hold great potential for understanding the role of progesterone and neurosteroids in PMDD but must be viewed with caution until larger trials with longer treatment duration can confirm these results.

While estradiol is described as superior to placebo in several studies, evidence to date must be regarded as preliminary and insufficient to support the recommended use of estradiol for the treatment of PMDD. Indeed, the doses of estradiol required to effectively treat PMDD suggest that in this condition, estradiol is acting to inhibit ovulation and interfere with normal cyclic ovarian hormone secretion. No studies have examined the therapeutic potential of GnRH antagonists, which have recently become available in an oral preparation.

Although not unrelated to the treatment of women with PMDD, a trial of GnRH agonist–induced ovarian suppression can be undertaken in some women who present with a recurrent affective disorder and in whom the clinical presentation suggests a menstrual cycle recrudescence of affective symptoms. In this clinical scenario, the results of the trial of ovarian suppression inform future decisions to include (or not) hormonal therapies to augment existing mood-stabilizing pharmacotherapy.

Estradiol Administration in Perimenopausal Depression and Postpartum Depression

Perimenopausal Depression

A 48-year-old gravida 2 para 2 married Caucasian female presented with an 8-month history of depressed mood, anhedonia, increased anxiety, tearfulness, avoidance of social activities, disturbed sleep, and decreased concentration. She did not feel that she was sad so much as, paradoxically, both more worried about even relatively trivial matters and less engaged by events in her life. She felt herself to be more emotionally reactive, finding herself becoming tearful in response to almost any emotionally laden situation. On questioning, the patient reported hot flushes, worse at night; however, daytime flushes were associated with marked anxiety, a racing heartbeat, and an enhanced sense of worry. Her menstrual cycle had been somewhat erratic in length for 2 years, and in the last year, she had skipped several menses with periods of amenorrhea lasting 2 to 3 months. The patient had no past personal or family history of depression. Laboratory testing revealed plasma FSH levels ranging from 25 to 40 IU/L and estradiol levels of 50 to 60 pg/mL. This typical case provides a springboard for highlighting elements of the diagnosis and treatment of perimenopausal depression (PMD).

For diagnostic purposes, FSH levels greater than two standard deviations above the mean for women of reproductive age combined with irregular menstrual cycles and amenorrhea characterize the menopause transition and the first year postmenopause (i.e., perimenopause). The most recent revision (2012) of the Stages of Reproductive Aging Workshop (STRAW +10) criteria emphasizes menstrual cycle variability with ≥7-day differences in length (i.e., either shortened or lengthened) of consecutive menstrual cycles (i.e., early menopause transition) and intervals of amenorrhea >60 days (i.e., late menopause transition) as the criteria for the perimenopause, with elevated plasma FSH levels >25 IU/L as confirmatory evidence only.

Several randomized, placebo-controlled trials of estradiol therapy (ET) in PMD (as well as those in perimenopausal women with depressive symptoms who may or may not have met diagnostic threshold criteria for major or minor depression) have demonstrated both short-term antidepressant efficacy and preventative benefits. In addition, low or declining levels of estradiol were shown to play a causal role in the appearance of depression in the subset of women with PMD; that is, under blinded and controlled conditions, acute estradiol withdrawal induced depressive symptoms in asymptomatic postmenopausal women with a past PMD but not in those without such a history. These findings support a role in maintaining stable levels of estradiol in women with or at risk for PMD. The small samples and short duration of treatment in trials of ET in perimenopausal depression notwithstanding, the best case for ET in the treatment of depression (either as monotherapy or as an adjunctive therapeutic agent) can be made in peri- rather than postmenopausal women. Of note, in a prospective placebo-controlled study, ET significantly decreased the risk of subsequent depression in perimenopausal women, with this effect most notable in those with a history of recent life stressors. Promising results notwithstanding, there are at present no clear predictors (including estradiol levels) of response to estradiol in women with perimenopausal depression. Further, the long-term risks of estradiol to endometrial and breast tissues continue to deter many women from its use. While initial reports suggested beneficial effects of estradiol on both cognitive performance (e.g., verbal memory) and cognitive pathology (i.e., Alzheimer dementia), consistent, perhaps, with its neuroprotective effects, subsequent reports demonstrate a far more restricted impact on cognition if any (e.g., limited to women with perimenopausal symptoms). The timing of starting ET relative to menopause has been proposed as one factor explaining the otherwise discordant observations (of the effects of ET on cognition) between the observational studies and the recent randomized controlled trials from the WHI. Thus, in several studies, women who take ET during the late menopause transition or early postmenopause have a lower risk of dementia than those starting ET later. This construct of a reproductive stage–dependent differential "risk" for both the benefits of ET on cognitive performance and the adverse effects in young peri- and postmenopausal women compared with older postmenopausal women has recently been tested. Two primary prevention, randomized, controlled trials examining the effects of ET on cognitive performance in younger and older postmenopausal women have recently been completed. Results of both the Kronos Early Estrogen Prevention Study (KEEPS) and the Early versus Late Intervention Trial with Estradiol (ELITE) showed that the effects of ET on measures of cognitive performance (including verbal and working memory) did not differ from those of placebo, nor did young postmenopausal women show more beneficial effects of ET on cognitive performance than older women. Thus, neither the therapeutic nor the prophylactic benefits of ET on cognition in postmenopausal women are supported by currently available data. Further, beliefs about other indications for the use of estrogen in the treatment of perimenopausal depression have been complicated by the HERS (Heart and Estrogen/Progestin Replacement Study) and WHI reports, which contradict the long-held view that estrogen protects against the development of heart disease in women. Since the cardioprotective effects previously reported were so substantial as to offset the increased risk of breast cancer after >4 years of use of estradiol, particularly given the efficacy of estradiol in the prevention of osteoporosis, many physicians recommended HT as prophylaxis against future heart and bone diseases. With the demonstration that, to the contrary, estrogen combined with the progestin MPA may increase the risk of coronary heart disease and stroke (during the first year of treatment), estradiol prophylaxis can no longer be recommended. However, estradiol treatment may still be indicated in women with perimenopausal symptoms (hot flushes, vaginal dryness) and depression, particularly in those who fail standard antidepressant treatment, who have a strong risk of osteoporosis, or who are intolerant of other acute treatments of perimenopausal symptoms. Indeed, recent analyses of the WHI demonstrate that the adverse effects of ET on cardiovascular outcomes were largely confined to older women (and those at a greater distance from menopause) compared with younger perimenopausal women, thus supporting the critical window or “gap” effect for the effects of estradiol on the cardiovasculature and providing some reassurance regarding the health effects of ET during the perimenopause.

The relationship between estrogen, both endogenous and exogenous, and the development of endometrial carcinoma has been suggested by several different lines of investigation. Numerous retrospective case-control studies published since 1975 have indicated that postmenopausal exposure to unopposed estrogens for more than 1 year results in a 2- to 12-fold increased relative risk for endometrial cancer. However, the addition of progesterone to ET decreases the risk of endometrial hyperplasia and endometrial cancer to equal or below that of women receiving no hormonal treatment. The optimal regimen to prevent hyperplasia and, thus, inferentially, the risk of carcinoma consists of 12 to 13 days of progestin treatment each month when estrogens are administered. There is an increased incidence of thromboembolism and stroke in women receiving ET and some SERMs; however, stroke is less likely in younger-aged perimenopausal women. Blood pressure, on average, appears to be unaffected by ET, although both increases and decreases have been reported. High doses of oral estrogens have been reported to elevate hepatocellular enzyme levels and, less commonly, cause cholestatic jaundice. The risk for gallstones and hepatocellular adenomas has been reported to be increased in association with oral contraceptive use, and although uncommon, these complications may also occur with the use of more physiologic doses of estrogen. Most studies have suggested an increased relative risk of breast cancer after 4 or 5 years of use, similar to the risk expected if the onset of menopause is delayed for a comparable length of time; recent analyses, however, suggest that the increased risk is related to current use and the risk normalizes after cessation of HT. In addition, it is recognized that ET (estrogen alone) is associated with considerably less risk of breast cancer than combined estrogen and progestin formulations. Indeed, a recent 20-year follow-up in the WHI reported that the use of estradiol alone (for an average of 7 years) was accompanied by a significant decrease in the risk of breast cancer. These data suggest that for some older women in whom estradiol stimulation of the endometrium is less of a clinical concern (e.g., in the presence of endometrial atrophy), continuous ET with appropriate endometrial monitoring (e.g., periodic vaginal ultrasounds) could be an acceptable longer-term therapy. Finally, the risk for breast cancer using HT for 5 years is associated with an increased "absolute risk" of only 0.67%, which some authors compare to the increased risk conferred by obesity and many other common exposures. For a more detailed review of the benefits and risks of menopausal ET, the reader is referred to recent updates from the Endocrine Society and the North American Menopause Society.

Estrogens may precipitate migraine headaches, and depression has also been reported to occur with the use of estrogens, but these reports are usually in the context of concurrent progestin use as in the oral contraception pill. If estradiol is to be used in the treatment of depression (see Table 33.17b–1), then one should consider using the 50-μg estradiol patch (transdermal estradiol [TE]). The patch delivers estradiol, the form of estrogen that mediates most of its physiologic effects, and does not cause an undesirable increase in the plasma proteins that mediate some of the adverse effects of estrogen. Levels achieved with this method are equivalent to levels in the early follicular phase and are slightly above the threshold necessary to preserve bone (60 pg/mL). In women without a uterus, estradiol may be given by itself and does not appear to have the risks associated with the combination of estradiol and progestin. Although ET does increase the risk of venous thromboembolism (VTE) and stroke, TE and low-dose estrogen are not accompanied by an increased risk of VTE. If a progestin is needed, micronized progesterone tablets may be used, which deliver progesterone rather than MPA. Of note, progestins may differ markedly in their molecular effects, physiologic effects, and side effects (as discussed above). For example, second-generation progestins (e.g., norethisterone or levonorgestrel) are more androgenic than third- (e.g., desogestrel) or fourth-generation (e.g., drospirenone) progestins but, as a consequence, are more able to oppose the adverse (inducing) effects of estradiol on the liver and are less likely to result in VTE. Finally, individuals can show 20-fold differences in the blood level achieved with the same form of estrogen, so one must be prepared to change doses or preparations if necessary after soliciting reports from the patient about side effects, adverse effects, or lack of efficacy.

Table 33.17b–1.
The Following Conditions Constitute Risk Factors or Contraindications* to Initiating Treatment With Estrogen

Women with a history of carcinoma of the breast*, or women with a family history of the following: premenopausal breast cancer or bilateral breast cancer in a first-degree relative; multiple family members (greater than three relatives) with postmenopausal breast cancer

Women with a history of uterine cancer*, endometriosis, ill-defined pelvic lesions*, particularly undiagnosed ovarian enlargement, undiagnosed vaginal bleeding*

(Active thrombophlebitis or thromboembolic disorders, or history of ischemic cardiac disease, pulmonary embolism, retinal thrombosis, or thrombophlebitis)*; (***risk factors for thromboembolic phenomena include cigarette smoking, varicose veins, prolonged periods of immobilization [including prolonged travel], and active heart disease***).

Known or suspected pregnancy*: Estrogens may cause fetal harm when administered to a pregnant women

Known hypersensitivity to estrogen or progestin (if prescribed) or the excipients (inactive compounds) contained within these medications. For example, Provera–calcium stearate, corn starch, lactose, mineral oil, sorbic acid, sucrose, talc; Alora–sorbitan monooleate, acrylic adhesive

History of cerebrovascular disease (stroke), epilepsy, hypertension, hypercalcemia

Cholecystitis or pancreatitis

Recurrent migraine headaches

Diabetes mellitus

History of familial hyperlipoproteinemia

Porphyria

Despite current concerns regarding ET and the requirement of adjunctive progestins for endometrial protection in women with an intact uterus, the specificity of actions of some of the newer SERMs and SARMs (and hence the promise of separating therapeutic from toxic effects) suggests that ER activators may occupy a useful niche in the future psychopharmacologic armamentarium. The recent introduction of tissue-selective estrogen complexes (TSECs), which combine estrogen with a SERM (some of which may be unable to enter the CNS), also holds promise for the delivery of individualized ET for those women who choose to employ hormonal therapies. Indeed, TSECs may have advantages over several SERMs (i.e., raloxifene and tamoxifen) since recent trials have documented that some TSECs (e.g., conjugated estrogen + bazedoxifene) are associated with a lower risk of VTE. These newer therapies could avoid the need for the addition of progestins (and the potential associated adverse effects) and could lead to therapies designed to optimize ER activity.

In the wake of the WHI report, interest in compounded bioidentical hormones as treatments for perimenopausal symptoms soared. Bioidentical means that the hormone is structurally and chemically indistinguishable from the same hormone in the human body. The purported benefits of this treatment class are increased safety and improved efficacy, both associated with the belief that these hormones are “natural” rather than "synthetic." There are several problems with these widely proclaimed benefits and inferences. First, as noted above, the effects—both therapeutic and adverse—are mediated through hormone receptors, which are sensitive only to the structure of the ligand, not to its source. Consequently, "bioidentical" progesterone that comes from a yam is, at the receptor level, indistinguishable from endogenous progesterone or FDA-approved micronized progesterone. If a hormone is bioidentical, the "natural" source is functionally irrelevant. Second, "natural bioidenticals" often begin with a chemical substrate from a plant that is subsequently modified in a lab to produce the final product (e.g., progesterone from diosgenin in the yam). Hence, even the "natural" hormone is synthetic. Third, bioavailability data for compounded products are exiguous at best, and results are inconsistent. As such, patients cannot reliably know the concentration of hormone to which their tissues are exposed. Fourth, FDA-approved bioidentical hormones exist, including micronized forms of estradiol and progesterone as well as a combination of estradiol and estriol. Many therapeutic benefits have been attributed in particular to estriol, a weak metabolite of estradiol and estrone, but confirmation of these benefits must await confirmation in adequately designed, controlled, head-to-head comparison trials with estradiol.

Postpartum Depression. Despite the absence of endocrine abnormalities in this disorder, women with PPD show an enduring abnormal mood response to reproductive hormones. A previous study demonstrated that exposure of euthymic women with a history of PPD to high doses of estradiol and progesterone followed by abrupt hormone withdrawal produced depression not seen in women lacking a history of PPD. This suggested that PPD occurred in a subgroup of women who were susceptible to reproductive hormone–linked affective disturbances. In parallel, supplemental reproductive steroids have been administered during the immediate postpartum in an attempt to prevent or treat postpartum affective disorders. An open trial in women at risk for puerperal psychosis demonstrated that high-dose estrogen treatment resulted in a lower-than-expected 1-year recurrence rate (9% compared with an expected 35% to 60% without prophylaxis). Varying doses of estrogen (Premarin ranging in dose from 0.625 to 10 mg/day or intravenous [IV] estradiol 25 mg q8h) were administered immediately postpartum and then tapered over 4 weeks. It was suggested that estrogen administration could attenuate the rapid puerperal drop in estradiol levels, thereby reducing the negative impact of the postpartum "estrogen withdrawal state" on mood. In a follow-up study, investigators tested the suggestion that estradiol withdrawal caused PPD in a double-blind, placebo-controlled study of estradiol in 61 women who developed major depression within 3 months of delivery. Eighty percent of the patients receiving estradiol (estrogen patch) experienced a significant reduction in depression severity after 3 months of treatment, compared with 31% of the placebo-treated group. Decreased mood symptoms on ET were observed in women regardless of concurrent antidepressant use, and estrogen's antidepressant effects were rapid and observed after 2 to 3 weeks of treatment. A similar rapid response to estradiol was also recently reported in an open-label trial of sublingual estradiol, similar to the timing of the response to estradiol in perimenopausal depression. Double-blind, placebo-controlled trials of estradiol in the treatment of PPD are ongoing and will help define the role of estradiol monotherapy in PPD.

An intriguing animal model suggested underlying mechanisms for the critical observations in PPD: the triggering of symptoms during a specific reproductive context (the puerperium) in a select genetic context conferring differential sensitivity to hormonal changes. In this model, a partial knockout of the delta-subunit of the $GABA_A$ receptor resulted in behavioral changes (increased sensitivity to stress and depressive-like behavior) that appeared only after the dams delivered. The extrasynaptic effects of allopregnanolone are mediated through the delta-subunit–containing $GABA_A$ receptors, and administration of allopregnanolone prevents adverse behavioral changes. These data were foundational in conducting a trial of a modified form of allopregnanolone, brexanolone, in PPD. The efficacy of a 60-hour infusion of brexanolone IV was demonstrated in a multicenter trial, with therapeutic separation from placebo seen by 24 hours and persisting for at least 30 days. These findings resulted in rapid FDA approval of the first treatment specifically for PPD. As oral forms of progesterone-derived neurosteroids are being developed, this class of compounds may represent a significant, practical addition to our therapeutic armamentarium.

Estradiol and SERMS in Acute Psychosis

Animal studies suggest potential neuromodulatory interactions between the dopaminergic system and estradiol signaling pathways. Estradiol priming has been reported to increase dopamine synthesis, affect dopamine turnover, and alter both pre- and postsynaptic D_1 and $D_{2/3}$ receptor binding affinity. Indeed, estradiol was reported to change $D_{2/3}$ receptors from high- to low-affinity states.

Since dysregulated dopamine action serves a pivotal role in the pathophysiology of psychosis, the supplementation of estradiol could potentially impact central dopaminergic function and mitigate the psychotic state. Interestingly, several trials have demonstrated therapeutic effects of physiologic doses of estradiol (i.e., 100 mcg per day) or SERMs in acute psychoses (double-blinded). In many of these studies, clinical effects have been evident within 2 weeks of administering the first dose of estradiol. However, the role of dopaminergic system dysfunction in the treatment response to ET or SERMS has not been fully characterized.

Androgen Administration in Midlife Depression

Several investigations have suggested that the change in mood, cognition, and libido during perimenopause/postmenopause may reflect deficiencies in androgens rather than (or in addition to) estrogens. Indeed, midlife in both men and women is characterized by a steady decline in the production of androgens such as dehydroepiandrosterone (DHEA) and androstenedione, which are mostly of adrenal origin yet contribute up to 50% of circulating androgen levels (testosterone, dihydrotestosterone) in women. Nonetheless, the definition of low testosterone levels is severely compromised by the current inability to accurately measure peripheral levels of testosterone in women or hypogonadal men (similar to the methodologic obstacles for measuring low estradiol levels). These assay limitations preclude the clear identification of men or women in whom androgen secretion is deficient (and in whom testosterone therapy would be indicated). Indeed, recent clinical practice guidelines recommend against making a diagnosis of androgen deficiency in women because of both the lack of a well-defined syndrome and the unavailability of reliable data correlating peripheral androgen levels with specific signs or symptoms. Moreover, the practice guidelines discourage the prescription of either DHEA or testosterone therapies in women because there are only limited data supporting therapeutic benefits for the targeted symptoms, and there are no long-term studies examining the risks of androgen therapy. In healthy men, there is considerable variability in the blood levels of testosterone below which symptoms develop, and the emergence of hypogonadal symptoms (e.g., decreased libido, fatigue, and depression) in aging men are frequently confounded by obesity and other comorbidities. Indeed, results from a recent hormone manipulation trial showed that the level of testosterone below which symptoms emerge both varies across men and is considerably below levels normally reported to accompany hypogonadism in men. Thus, testosterone therapy in older men also should be employed cautiously, and the initial management of age-related hypogonadal symptoms should focus on the achievement of a more normal body mass index (BMI) and management of comorbidities. That said, androgen therapy has anecdotally been reported to benefit both mood and sexual function/libido in males who are clinically hypogonadal. Further, several investigators and a recent meta-analysis have shown that androgens improve a variety of elements of sexual function (desire, arousal, fantasies) when administered as sole treatment (ovariectomized women) or in combination with estradiol (postmenopausal women). Medical literature dating back to the isolation of testosterone in 1935 is replete with reports of the antidepressant efficacy of testosterone, androgen/anabolic steroids, and adrenal (and ovarian) androgens like DHEA. While most of these reports were uncontrolled, one study in ovariectomized women again demonstrated the superiority of testosterone to placebo in the mitigation of depressive symptoms. If androgens can improve mood, are they effective only in the context of clinically demonstrable androgen deficiency (which, as noted above, is difficult to reliably determine)? Further, even if shown to “improve well-being” (as was DHEA in older men and women), is androgen treatment effective in clinically significant affective disorders? Indeed, a recent meta-analysis concluded that in women with normal adrenal function, the administration of DHEA is not indicated for complaints of either depressed mood or decreased sexual interest or arousal. This meta-analysis, however, does not answer the question of DHEA’s efficacy in clinically significant depression since the majority of the reviewed studies were not sufficiently powered to evaluate DHEA’s effects on depression. As should, by now, be clear, this, like other forms of HT, has attendant nonuniform benefits and risks. DHEA, for example, may improve bone mineral density with chronic administration, thus helping to prevent age-related osteopenia (although, with the exception of small improvements in lumbar spine density in women, a recent meta-analysis found little evidence that DHEA had significant effects on bone density in women). It has been suggested, however, that DHEA administration, while not increasing total plasma testosterone levels in men, might be concentrated in the prostate, where local conversion to testosterone could accelerate the growth of androgen-dependent tumors.

FUTURE DIRECTIONS

There is little question that reproductive hormones can regulate mood and that manipulations of reproductive hormones can, in some instances, dramatically alter the course and expression of affective illness. It is equally clear that these hormonal effects are not universal and appear in subgroups of individuals who appear differentially sensitive to these central actions of reproductive steroids. A major task for the future will be to identify the mediators of this differential sensitivity. Identification of the genetic or other predictors of vulnerability to reproductive endocrine-related mood disorders would assist in the identification of those who would respond to reproductive therapies and those who might be at risk for adverse effects. Indeed, the main constraint on the effective incorporation of reproductive hormonal modulators into the psychotherapeutic armamentarium as primary psychotropic agents or adjunctive therapies (e.g., in those who experience incomplete responses to other psychotropic agents) is the inability to predict the likelihood of adverse medical outcomes (e.g., breast cancer). In addition, it is now clear that direct assays of sex steroids in populations with low levels of sex steroids (i.e., postmenopausal women, older adult men, and young children) are no longer acceptable evidence to confirm endocrine “deficiency” states, although even if accurate, single plasma hormone levels will not detect the pharmacodynamic differences across individuals that may result in marked phenotypic variation. The specificity and complexity of the actions of individual hormonal agents preclude generalization about the likely therapeutic or adverse outcome of exposure to a given hormone. Slight structural differences between the hormonal agents available in a given class can translate into dramatic differences in cellular and physiologic effects (and hence efficacy and risk), as can the timing of administration of these agents (i.e., their continuous or discontinuous administration). While generating current uncertainty, the specificity of actions of individual hormonal preparations, steroidogenic enzyme inhibitors, and steroid receptor antagonists creates an extraordinary opportunity for the development of tissue-specific pharmacologic agents. The SERMs already available, for example, not only possess the beneficial effects of estradiol in some tissues (e.g., bone) but additionally block the potentially harmful effects of estradiol in other tissues (e.g., breast). Current knowledge, then, of tissue-specific coregulators will increasingly permit the design of reproductive agents that select and restrict the tissues affected and the actions elicited. In this fashion, the goal of psychopharmacologic precision will be greatly advanced.

FURTHER READINGS

Bassuk SS, Manson JE. The timing hypothesis: do coronary risks of menopausal hormone therapy vary by age or time since menopause onset? ***Metabolism***. 2016;65(5):794–803.

Bixo M, Ekberg K, Sundström Poroma I, et al. Treatment of premenstrual dysphoric disorder with the GABAA receptor modulating steroid antagonist Sepranolone (UC1010)—Arandomized controlled trial 2021. ***Psychoneuroendocrinology***. 2017;80:46–55.

Bloch M, Daly RC, Rubinow DR. Endocrine factors in the etiology of postpartum depression. ***Compr Psychiatry***. 2003;44(3):234–246.

Clarkson TB. Estrogen effects on arteries vary with stage of reproductive life and extent of subclinical atherosclerosis progression. ***Menopause***. 2007;14(3):373–384.

Comasco E, Kopp Kallner H, Bixo M, et al. Ulipristal acetate for treatment of premenstrual dysphoric disorder—a proof-of-concept randomized controlled trial. ***Am J Psychiatry***. 2021;178(3):256–265.

Dreher JC, Schmidt PJ, Kohn P, Furman D, Rubinow D, Berman KF. Menstrual cycle phase modulates reward-related neural function in women. ***Proc Natl Acad Sci USA***. 2007;104(7):2465–2470.

*Freeman EW, Sammel MD, Lin H, Nelson DB. Associations of hormones and menopausal status with depressed mood in women with no history of depression. ***Arch Gen Psychiatry***. 2006;63(4):375–382.

Gordon JL, Rubinow DR, Eisenlohr-Moul T, Xia K, Schmidt PJ, Girdler SS. Progesterone in the prevention of depressive symptoms in the menopause transition: A randomized clinical trial. ***JAMA Psychiatry***. 2018;75:149–157.

Lonard DM, O'Malley BW. Nuclear Receptor coregulators: modulators of pathology and therapeutic targets. ***Nat Rev Endocrinol***. 2012;8(10):598–604.

Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. ***Endocr Rev***. 2011;32(1):81–151.

Nilsson S, Koehler KF, Gustafsson JÅ. Development of subtype-selective oestrogen receptor-based therapeutics. ***Nat Rev Drug Discov***. 2011;10:778.

North American Menopause Society. The 2012 hormone therapy position statement of the North American Menopause Society. ***Menopause***. 2012;19(3):257–271.

*Remage-Healey L. Frank Beach Award Winner: steroids as neuromodulators of brain circuits and behavior. ***Horm Behav***. 2014;66(3):552–560.

Rosner W, Hankinson SE, Sluss PM, Vesper HW, Wierman ME. Challenges to the measurement of estradiol: an endocrine society position statement. ***J Clin Endocrinol Metab***. 2013;98(4):1376–1387.

Rossouw JE, Prentice RL, Manson JE, et al. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. ***JAMA***. 2007;297(13):1465–1477.

Rubinow DR, Schmidt PJ. Androgens, brain and behavior. ***Am J Psychiatry***. 1996;153(8):974–984.

*Rubinow DR, Schmidt PJ, Meltzer-Brody S, Harsh V. Gonadal hormones and behavior in women: concentration vs. context. In: Pfaff D, ed. ***Hormones, Brain and Behavior***. Academic Press; 2007.

Rubinow DR, Johnson SL, Schmidt PJ, Girdler S, Gaynes B. Efficacy of estradiol in perimenopausal depression: so much promise and so few answers. ***Depress Anxiety***. 2015;32(8):539–549.

Rubinow DR, Schmidt PJ, Craft CD. Gonadal Steroids and Mood Disorders. In: Charney DS, Buxbaum JD, Sklar P, Nestler EJ, eds. ***Neurobiology of Mental Illness***, 4th ed. Oxford University Press; 2013.

Santen RJ, Kagan R, Altomare CJ, Komm B, Mirkin S, Taylor HS. Current and evolving approaches to individualizing estrogen receptor-based therapy for menopausal women. ***J Clin Endocrinol Metab***. 2014;99(3):733–747.

Scheffers CS, Armstrong S, Cantineau AEP, Farquhar C, Jordan V. Dehydroepiandrosterone for women in the peri- or postmenopausal phase. ***Cochrane Database Syst Rev***. 2015;1:CD011066.

Schiller CE, Schmidt PJ, Rubinow DR. Allopregnanolone as a mediator of affective switching in reproductive mood disorders. ***Psychopharmacology (Berl)***. 2014;231(17):3557–3567.

*Schmidt PJ, Ben Dor R, Martinez PE, et al. Effects of estradiol withdrawal on mood in women with past perimenopausal depression: a randomized clinical trial. ***JAMA Psychiatry***. 2015;72(7):714–726.

Schmidt PJ, Nieman L, Danaceau MA, et al. Estrogen replacement in perimenopause-related depression: a preliminary report. ***Am J Obstet Gynecol***. 2000;183(2):414–420.

Schwartz TB. Henry Harrower and the turbulent beginnings of endocrinology. ***Ann Intern Med***. 1999;131(9):702–706.

Soares CD, Almeida OP, Joffe H, Cohen LS. Efficacy of estradiol for the treatment of depressive disorders in perimenopausal women: a double-blind, randomized, placebo-controlled trial. ***Arch Gen Psychiatry***. 2001;58(6):529–534.

Spearow JL, Doemeny P, Sera R, Leffler R, Barkley M. Genetic variation in susceptibility to endocrine disruption by estrogen in mice. ***Science***. 1999;285(5431):1259–1261.

Taylor HS, Manson JE. Update in hormone therapy use in menopause. ***J Clin Endocrinol Metab***. 2011;96(2):255–264.

Werner AA, Johns GA, Hoctor EF, Ault CC, Kohler LW, Weis MW. Involutional melancholia: probable etiology and treatment. ***JAMA***. 1934;103:13–16.

Wierman ME, Arlt W, Basson R, et al. Androgen therapy in women: a reappraisal: an endocrine society clinical practice guideline. ***J Clin Endocrinol Metab***. 2014;99(10):3489–3510.

Wisner KL, Sit DKY, Moses-Kolko EL, et al. Transdermal estradiol treatment for postpartum depression: a pilot, randomized trial. ***J Clin Psychopharmacol***. 2015;35(4):389–395.

*Writing Group for the Women's Health Initiative Investigators; Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. ***JAMA.*** 2002;288(3):321–333.

33.17c Pharmacologic Treatment of Sexual Dysfunction

CARIDAD PONCE MARTINEZ, M.D., SARIA EL HADDAD, M.D., AND ANITA H. CLAYTON, M.D.

Satisfactory sexual functioning is an important part of the human experience and contributes to the overall quality of life. According to the World Health Organization, sexual health is "a state of physical, emotional, mental and social well-being in relation to sexuality; it is not merely the absence of disease, dysfunction, or infirmity." Sexual dysfunction is a syndrome comprised of pathologies that can affect women and men across several phases of the sexual response cycle, desire, arousal, and orgasm, plus sexual pain. It is frequently multifactorial, with biologic, psychological, sociocultural, and interpersonal/relational factors. There are important sex differences in the neurobiology, etiology, and comorbid sexual/medical/psychiatric conditions, including associated medications, that highlight the need for different strategies in the evaluation and management of sexual dysfunction.

The DSM-5 categorizes sexual dysfunctions into gender-specific disorders of impaired desire or excitation (male hypoactive sexual desire disorder [HSDD], erectile disorder [ED], female sexual interest/arousal disorder [FSI/AD]), disorders involving orgasm (female orgasmic disorder [FOD], delayed ejaculation, premature or early ejaculation [PE]), disorders of pain/penetration (genitopelvic pain/penetration disorder—female only), and others (substance/medication-induced sexual dysfunction, other-specified sexual dysfunction, and unspecified sexual dysfunction). The diagnosis of FSI/AD resulted from an integration of HSDD and female sexual arousal disorder (FSAD) in the previous editions of the DSM. Clinical treatment and research have focused on desire and arousal separately, as essentially all studies/publications utilized the DSM-IV diagnoses. As such, in this chapter, we will focus on pharmacologic treatment of HSDD and FSAD rather than the combined FSI/AD, FOD, ED, and PE. The 10th revision of the ***International Statistical Classification of Diseases and Related Health Problems*** (ICD-10) codes corresponding disorders as gender neutral and includes lack or loss of sexual desire, failure of genital response, orgasmic dysfunction, and PE. The ICD-11 has a separate chapter for sexual health, which includes parallel sexual dysfunctions for women and men. Table 33.17c–1 displays the diagnoses across DSM and ICD versions for comparison.

Although each of these disorders has its specified diagnostic criteria, it is important to note that they frequently coexist, highlighting the importance of a full assessment once sexual dysfunction is identified. Furthermore, the DSM-5 requires symptoms to be present for at least 6 months for a diagnosis of a disorder. Lifelong sexual dysfunctions are present following initiation of sexual activity, while acquired sexual disorders occur after a period of normal sexual function per the individual. Distress about sexual dysfunction is a required criterion for diagnosis. The disorder can be situational in that it is specific to certain types of stimulation, situations, or partners, or it can be generalized. The severity of the disorder can be mild, moderate, or severe, depending on the level of distress the symptoms cause in the individual.

Table 33.17c–1.
Classification of Sexual Dysfunctions in Recent Versions of DSM and ICD

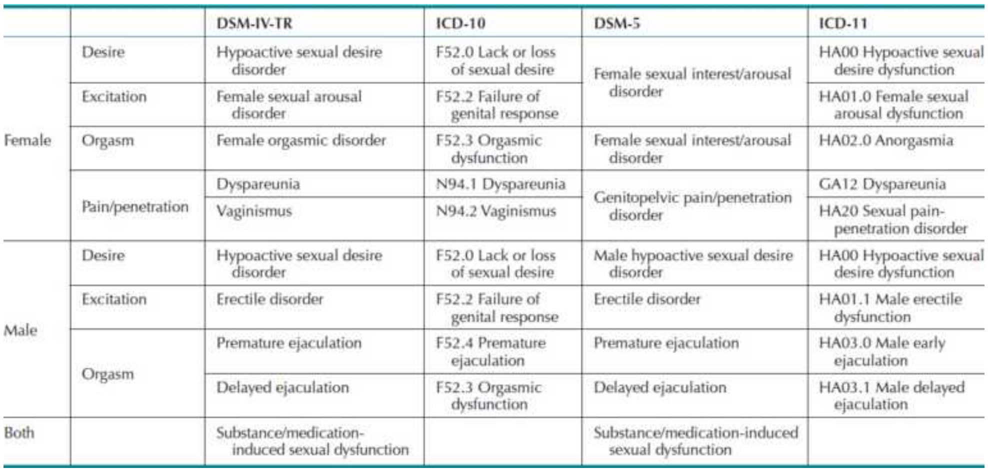

		DSM-IV-TR	ICD-10	DSM-5	ICD-11
Female	Desire	Hypoactive sexual desire disorder	F52.0 Lack or loss of sexual desire	Female sexual interest/arousal disorder	HA00 Hypoactive sexual desire dysfunction
	Excitation	Female sexual arousal disorder	F52.2 Failure of genital response		HA01.0 Female sexual arousal dysfunction
	Orgasm	Female orgasmic disorder	F52.3 Orgasmic dysfunction	Female sexual interest/arousal disorder	HA02.0 Anorgasmia
	Pain/penetration	Dyspareunia	N94.1 Dyspareunia	Genitopelvic pain/penetration disorder	GA12 Dyspareunia
		Vaginismus	N94.2 Vaginismus		HA20 Sexual pain-penetration disorder
Male	Desire	Hypoactive sexual desire disorder	F52.0 Lack or loss of sexual desire	Male hypoactive sexual desire disorder	HA00 Hypoactive sexual desire dysfunction
	Excitation	Erectile disorder	F52.2 Failure of genital response	Erectile disorder	HA01.1 Male erectile dysfunction
	Orgasm	Premature ejaculation	F52.4 Premature ejaculation	Premature ejaculation	HA03.0 Male early ejaculation
		Delayed ejaculation	F52.3 Orgasmic dysfunction	Delayed ejaculation	HA03.1 Male delayed ejaculation
Both		Substance/medication-induced sexual dysfunction		Substance/medication-induced sexual dysfunction	

Establishing the prevalence of sexual dysfunction (SD) has been complicated by variations in definitions (sexual complaints, SDs, sexual dysfunctions), variations in how SD is assessed (self-report vs. validated/nonvalidated measures), variations in populations studied, the inclusion of distress, cultural effects, and co-occurring sexual dysfunctions, among other factors. Men more commonly seek evaluation for sexual dysfunction. However, women experience a high prevalence of sexual dysfunction.

The National Health and Social Life Survey (NHSLS) was a study of sexual behavior in U.S. adults aged 18 to 59 years. In 1992, 1,749 women and 1,410 men were randomly selected; the evaluation assessed the risk of experiencing SD as well as associated negative outcomes but did not evaluate distress. SD was more prevalent in women (43%) than men (31%). The most common problems in women were desire (32%), orgasm (26%), and lubrication (21%). In men, PE was the most common complaint (30%), compared to erectile dysfunction (10%) and low desire (14%). SD was associated with demographic characteristics, including age and educational attainment. Sexual concerns were more likely among men and women with poor physical and emotional health and were highly associated with negative experiences in sexual relationships and overall well-being.

ED is the most common type of sexual dysfunction in men, with an estimated prevalence in the U.S. male population aged 20 years or older of 18.4%, equivalent to more than 18 million men. In the Massachusetts Male Aging Study, which surveyed approximately 1,700 men longitudinally, the combined prevalence of ED was 52%, with 39% in 40-year-old men and 67% in men aged 70 years or older. After adjusting for age, the strongest correlates for ED were markers of cardiovascular disease—hypertension, diabetes, and heart disease.

HSDD is the most common type of sexual dysfunction in women, with prevalence ranging between 17% to 50% depending on the population studied, methodology, and definition. In the Prevalence of Female Sexual Problems Associated with Distress and Determinants of Treatment Seeking (PRESIDE) study, which included 31,581 U.S. female respondents from 50,002 households (63% response rate) between the ages of 18 to 102 years, low desire was the most common sexual problem (38.7% of participants); low arousal and orgasm difficulties had prevalence rates of 26.1% and 20.5%, respectively. Low desire associated with distress (HSDD) was present in 10% of women and was more common than personally distressing problems with arousal or orgasm (5.4% and 4.7%, respectively). Any sexual problem associated with personal distress occurred in about 12% of women. Depression was positively associated with problems with desire and orgasm.

In men, increased age is associated with a higher frequency of erection problems and decreased sexual desire. In women, the prevalence of sexual problems tends to peak in the middle years, then decrease with increasing age, except for changes associated with menopause (e.g., trouble lubricating); however, this has not been a consistent finding across studies. Poor health is associated with increased risk for all categories of sexual dysfunction in men, whereas in women, it is only associated with increased sexual pain.

PATHOPHYSIOLOGY

Normal sexual function involves psychological, social, and biologic processes. Biologic or organic factors that include medical conditions, surgical procedures, medications, recreational substances, and lifestyle factors can negatively impact sexual function and lead to sexual dysfunction. Psychosocial factors such as depression, anxiety, traumatic experiences, relationship problems, prior experiences, moral prohibitions, and psychosocial stressors can also play an important role. Many cases of SD have both biologic and psychogenic features. Even in cases of clear organic etiology like perineal trauma, impaired sexual function can result in psychological consequences such as low self-esteem, low self-confidence, and even anxiety or depressed affect, which can further exacerbate the dysfunction. Involvement of sexual partners during evaluation and treatment is important in education about the diagnosis and in examining patient and partner expectations for treatment outcomes. Studies have identified a bidirectional association in sexual dysfunction between partners, with a threefold increase in male SD when the female partner had SD and a significant increase in female SD and infrequency of sex associated with male SD.

The neurobiology of sexual functioning is mediated by neurotransmitters, sex steroids, and other hormones with actions in the central nervous system and peripherally in the urogenital organs. Sexual excitation involves dopamine, norepinephrine, nitric oxide, oxytocin, vasopressin, and melanocortins. Inhibition involves serotonin, prolactin, endocannabinoids, and opioids. Normal levels of sex steroids such as testosterone and estrogen increase the sensitivity of an individual toward sexual stimuli centrally and peripherally.

In generalized HSDD, increased inhibition and/or decreased excitation appear to mediate low desire. In women, sex hormones play a crucial role in all phases of the sexual response cycle. Estradiol affects sexual desire and genital structures. In hypoestrogenic states, whether natural (e.g., perimenopause, postmenopause, and postpartum) or acquired (e.g., hormonal chemotherapies, surgical menopause), the resulting physiologic changes may affect sexual functioning due to the peripheral effects of atrophy of urogenital tissues, inadequate vaginal lubrication, or reduced capacity for arousal and orgasm due to reduced vaginal blood flow. In addition, hormonal contraceptives containing estrogen may increase sex hormone–binding globulin (SHBG) which binds testosterone, lowers free or bioavailable testosterone, and may impact sexual desire and arousal. In females, despite the importance of testosterone in normal sexual function, testosterone serum levels do not correlate with sexual desire, nor do they reflect intracrine sex hormone production. This is likely because testosterone levels in females are much lower than in men, and current assays do not accurately measure these low free and total testosterone levels.

In addition to the neuroendocrine factors described above, medications such as antidepressants, antiandrogens, and oral contraceptives can contribute to female sexual dysfunction. The most common psychological factors affecting female sexual function are depression, anxiety, distraction, negative body image, and trauma. In medically healthy women, problems with sexual arousal are often the result of insufficient sexual stimulation rather than organic causes. Vascular disease, associated with diabetes mellitus (DM) and atherosclerosis, as well as many of its treatments, can cause arousal problems in women by inhibiting smooth muscle relaxation and vasocongestion, necessary for the engorgement of the clitoris and vagina, and resulting lubrication. There is conflicting data, however, regarding the relationship between metabolic syndrome and HSDD.

In general terms, ED can be classified as psychogenic, organic, or mixed. Psychogenic factors such as performance anxiety and others mentioned previously can negatively impact erectile function. This type of ED tends to occur on a situational basis, has a sudden onset, and concurrently, men can experience normal erections with masturbation, at night or upon awakening. Organic processes can be vascular (e.g., arterial insufficiency, venous leakage), neurologic, endocrine (e.g., hypogonadism, hyperprolactinemia), substance induced, or due to a systemic illness. In these cases, ED tends to be more insidious in its onset, and erections are deficient in most contexts (masturbation, nocturnal).

Healthy vascular and neurologic functions are necessary for an erection to occur. There is a neurologically mediated increase in penile arterial inflow. Nitric oxide is released from the endothelium and parasympathetic nerve terminals, allowing relaxation of the cavernosal smooth muscle, vasocongestion, and engorgement of penile tissue, resulting in an erection. A vasculogenic etiology is the most common among organic etiologies of ED. Current evidence suggests a strong correlation between cardiovascular disease and ED, with a common pathophysiologic basis of endothelial dysfunction, inflammation, and low testosterone. ED is also associated with coronary heart disease, stroke, and all-cause mortality. The presence of vasculogenic ED in young men can be a particularly early indicator of future cardiovascular disease, and these patients should receive an appropriate cardiovascular workup for risk stratification.

Risk factors for ED can be examined in accordance with etiology. Some neurologic disorders (e.g., multiple sclerosis, stroke, temporal lobe epilepsy, Parkinson disease, Alzheimer disease, spinal cord injury) are associated with ED, and the risk of nerve injury leading to ED is very high in patients undergoing radical pelvic surgery. Endocrine risk factors include hypogonadism, hyperprolactinemia, and hypothyroidism. Some risk factors for vasculogenic ED are atherosclerosis, hypertension, hyperlipidemia, obesity, smoking, DM, and pelvic irradiation. Among men with DM, 50% to 75% develop ED; the prevalence of ED in diabetics is three times that seen in nondiabetics. Lower urinary tract symptoms have also been linked to the development of ED. The most common medications that can cause or contribute to ED are antiandrogens, antihypertensives (thiazide diuretics, beta-blockers, calcium channel blockers), antiarrhythmics (e.g., digoxin), and many psychotropic drugs (tricyclic antidepressants, selective serotonin reuptake inhibitors [SSRIs], phenothiazines, and butyrophenones).

The use of recreational, illegal, or controlled substances may contribute to sexual dysfunction in both men and women, although the evidence for most substances is very limited, particularly in regard to women. Table 33.17c–2 summarizes some of these findings. An umbrella review of 98 meta-analyses sought to synthesize all of the available meta-analytic research on the risk factors, treatment, and prevalence of SD; substances evaluated included tobacco, alcohol, caffeine, cannabis, and opioids. A systematic review and meta-analysis including 8,829 men from 10 studies indicated that the use of opioids was significantly associated with an increased risk of ED (RR = 1.96, 95% confidence interval [CI] 1.66 to 2.32, $p < .001$), although the overall quality of evidence was rated as low due to observational study designs. This study included methadone, heroin, opium, and other opioids, but there remains limited knowledge on the effects of medications used in the treatment of opioid use disorder, buprenorphine and naltrexone. Heroin appears to raise testosterone-binding globulin levels and thus, reduce plasma levels of free testosterone; heroin and other opioids also affect the mesolimbic dopamine reward pathway, which may contribute to reduced reward with sex. Cocaine use may initially enhance sexual functioning by increasing desire, but prolonged use may affect desire and orgasm. Amphetamines are popularly considered potent aphrodisiacs, and methamphetamines are perceived to be sexual enhancers due to increased social confidence, sexual disinhibition, and a sense of physical energy. There is no scientific evidence linking it to direct action on specific receptors that can enhance sexuality.

Table 33.17c–2.
Effect of Recreational/Illegal Substances on Sexual Dysfunctions

	Effect on Sexual Function
Tobacco	↑ ED (linear dose-response association)
Alcohol	Moderate (***mode*** = ***1–3 drinks/day; mean*** = ***7–8 drinks/ week***): ↓ risk of ED High (***mode*** >***3 drinks/day; mean*** = ***23.1 drinks/week***): ↑ risk of ED
Cannabis	Inconclusive (Results likely due to intermittent vs. chronic daily use)
Opioids	↑ risk of ED (unclear if applicable to buprenorphine, naltrexone)
Cocaine	↑ desire initially Chronic daily use: ↓ desire, ↓ orgasm
Amphetamines	Perceived enhancement of sexual function (↑ confidence/enjoyment and ↓ inhibition) but no objective improvement Chronic use: ED, delayed ejaculation, delayed orgasm in women

ED, erectile disorder.

PE can occur due to organic causes (e.g., urologic or thyroid dysfunction) but is most often the result of a normal variation of ejaculatory performance or due to psychological/relationship problems. PE is present in about a third of patients complaining of ED. As a result, patients may often respond to psychoeducation, reassurance, or psychotherapy.

ASSESSMENT

Assessment of sexual dysfunction is best performed using a biopsychosocial approach and should include psychiatric history, medical and surgical history, sexual history, family history, and social history. Detailed substance abuse and medication histories can help identify substance-induced SD.

Screening questions in women need to include HSDD in reproductive-age women and pain/dyspareunia in peri- and postmenopausal women. When evaluating HSDD in women, contributing conditions such as vaginal atrophy, vaginitis, diabetes, and lactation need to be ruled out.

In men, questions include erectile dysfunction and PE. A brief screening tool can be used to look at every phase of the sexual response cycle in both men and women. The Decreased Sexual Desire Screener (DSDS) can be used for HSDD and the abridged version of the International Index of Erectile Function (IIEF-5) for ED.

In addition to the interview, an evaluation for ED includes a general physical examination with particular emphasis on the nervous, vascular, and genital systems. Diagnosis is based on clinical presentation, not laboratory values. Laboratory studies can be utilized to rule out endocrine causes and identify associated comorbid conditions. The International Society for Sexual Medicine recommends ordering tests to measure fasting glucose, hemoglobin A1c (HbA1c), lipids, and total testosterone levels.

TREATMENT

Prior to starting specific interventions, initial management should address modifiable risk factors for sexual dysfunction. These include minimizing medication side effects and addressing the use of illicit/recreational substances, including tobacco. Management of cardiovascular risk factors includes weight management, exercise, and the adoption of a Mediterranean diet. In men with obstructive sleep apnea (OSA), treatment with continuous positive airway pressure (CPAP) is associated with some improvement in sexual function. Finally, gastric bypass surgery can increase testosterone levels and therefore improve sexual function in men with morbid obesity.

Involvement of the sexual partner may be useful in order to explore psychosocial or relational issues and evaluate the benefit of psychotherapy.

Psychotherapy in the form of individual cognitive behavioral therapy and couples therapy is utilized, randomized, adequately controlled clinical trials on its efficacy are lacking.

Specific Interventions

Hypoactive Sexual Desire Disorder. There are two Food and Drug Administration (FDA)-approved agents for the treatment of acquired, generalized HSDD in premenopausal women, flibanserin and bremelanotide.

Flibanserin, approved in 2015, is a 5-HT1A agonist and a 5-HT2A antagonist, which in the prefrontal cortex causes a decrease in serotonin and an increase in norepinephrine and dopamine activity. Efficacy trials in pre- and postmenopausal women indicate that flibanserin can cause an increase in sexual desire and satisfying sexual events and reduce distress related to sexual desire.

Dosed at 100 mg nightly, flibanserin is generally well-tolerated with side effects including dizziness, sedation, and fatigue. Initially, the FDA required a Risk Evaluation and Mitigation Strategy (REMS) program with a boxed warning to avoid alcohol with flibanserin due to concerns for hypotension and syncope. Three postapproval compulsory studies demonstrated no clinically significant risk if the use of flibanserin was separated by at least 2 hours from alcohol consumption. As a result, the FDA removed the REMS.

The melanocortin-4 receptor modulates sexual arousal and behavior, as supported by evidence in studies of mice, rats, and primates. Bremelanotide, a novel melanocortin receptor agonist, approved in 2019, is available in an autoinjector for subcutaneous formulation to be administered 45 minutes prior to anticipated sexual activity. Dosing should be limited to once a day (effect lasts around 24 hours) and no more than eight times a month. The most common side effect is nausea. It can cause transient hypertension so avoid in those with hypertension or cardiovascular illness.

Off-label agents include bupropion and transdermal testosterone. Sustained-release bupropion in premenopausal women was shown to improve sexual arousal, orgasm completion, and sexual satisfaction. Transdermal testosterone administration was shown to have favorable outcomes in all domains of sexual function and reduction in personal distress in postmenopausal women.

Female Sexual Arousal Disorder. Mindfulness-based therapy has been studied for FSAD and was found to have a beneficial effect on sexual desire and sexual distress. The efficacy seemed to be more pronounced in women with a history of sexual trauma. Nevertheless, mindfulness-based cognitive therapy (MBCT) and other psychological interventions are the mainstays of treatment when pharmacologic treatment is ineffective or causes intolerable side effects.

Female Orgasmic Disorder. The approach to treating FOD is largely psychoeducational. A variety of approaches have been tested, with the most consistent support emerging for directed masturbation, sensate focus, and psychotherapy.

EROS is a medical device approved in 2000 that consists of a small vacuum pump applied over the clitoris to stimulate blood flow, aiding in sexual arousal and orgasm.

Adequate stimulation is required for orgasm; enhancing arousal with off-label drugs such as bupropion, bremelanotide, and Phosphodiesterase-5 inhibitors (PDE-5Is) may facilitate achieving orgasm and enhance the frequency, intensity, and duration of orgasm.

Erectile Disorder. Nitric oxide (NO) mediates tumescence by increasing the levels of cyclic guanosine monophosphate (cGMP), which in turn leads to smooth muscle relaxation in the corpus cavernosum, allowing for engorgement with blood, resulting in an erection. Dilatation of the penile arteries leads to a more prolonged erection.

Phosphodiesterase-5 (PDE-5) is an enzyme that catabolizes cGMP. PDE-5 Inhibitors (PDE-5Is) competitively inhibit the PDE-5 enzyme and increase intracavernosal cGMP, thereby leading to an increase in both the number and duration of erections in men with ED. PDE-5Is improve endothelial function and reduce apoptosis of vascular smooth muscle cells in the corpus cavernosum.

Adequate testosterone levels and sufficient psychological and environmental cues that lead to sexual arousal and stimulation are required for PDE-5Is to work.

An evaluation of the underlying cause of ED is necessary prior to the initiation of PDE-5I treatment. These agents are contraindicated in men treated with nitrates and are to be used cautiously in men treated with alpha-adrenergic blockers. Vardenafil is known to cause minor QT prolongation and is, therefore, contraindicated in patients taking type 1A and type 3 antiarrhythmics and patients with congenital prolonged QT syndrome.

Treatment response seems to be largely dependent on proper use and administration (Table 33.17c–3).

SILDENAFIL. A meta-analysis of 27 trials in 6,659 men with ED showed a marked difference in the percentage of successful sexual intercourse between sildenafil and placebo (57% vs. 21%, respectively).

Sildenafil was also found effective for the treatment of ED in the setting of various conditions, including DM, prostate cancer treatment, parkinsonism, and depression treated with SSRIs.

Sildenafil could also have psychological benefits in men with ED, as shown by improvement in the mean self-esteem subscale score of the Self-Esteem and Relationship (SEAR) questionnaire (a psychometric instrument specific to ED).

Sildenafil is also available in an orodispersible form that disintegrates within seconds of ingestion leading to quicker absorption. This formulation offers an added advantage in patients with dysphagia.

VARDENAFIL. Vardenafil shares a similar structure and pharmacokinetic profile with sildenafil. It is efficacious in ED secondary to DM and nerve-sparing prostatectomy. It is also available in an orally disintegrating form and may have a more rapid onset of action.

TADALAFIL. Tadalafil is structurally distinct from sildenafil and vardenafil. It has a longer half-life than other PDE-5Is, thereby allowing the user greater flexibility in participating in sexual activity. Low dose may be taken daily.

AVANAFIL. Avanafil is a new agent currently approved in the United States and Europe for the treatment of ED. It is more selective to PDE-5 and has a more rapid onset of action.

Adverse Effects and Precautions. Common side effects include flushing, headache, and dyspepsia.

Table 33.17c–3.
Comparison of PDE-5 Inhibitors

Medication	Onset of Action and Duration	Dosing
Sildenafil	Effective for as-needed treatment of organic, psychogenic, or mixed-type ED Taken 1 hour before sex and effective up to 4 hours	50–100 mg on an empty stomach Administration with a high-fat meal may delay absorption 25 mg if coadministered with a strong CYP3A4 inhibitor or an alpha-1 blocker and in patients with renal impairment (CrCl <30 mL/minute) or moderate hepatic impairment
Vardenafil	Same as sildenafil	10–20 mg on an empty stomach Patients ≥65 years old: 5 mg initially Orally disintegrating tablet (ODT): 10 mg on an empty stomach (no titration) Administration with a high-fat meal may delay absorption 2.5 or 5 mg if coadministered with a strong CYP3A4 inhibitor or an alpha-1 blocker, respectively Dose adjustment for moderate hepatic impairment needed
Tadalafil	Similar onset of action as sildenafil Duration of action up to 36 hours Effective for as-needed **or** daily treatment of organic, psychogenic, or mixed-type ED	10–20 mg if as-needed 10 mg not more than once every 72 hours if coadministered with a strong CYP3A4 inhibitor Dose adjustment for renal or hepatic impairment is necessary when used as needed **or** Daily treatment: 2.5–5 mg once daily (if normal renal and hepatic function) 2.5 mg once daily if coadministered with a strong inhibitor of CYP3A4 or alpha-1 blocker Administration with a high-fat meal does not alter the absorption
Avanafil	Onset 15–30 minutes is more rapid than sildenafil	50–200 mg, 15 minutes to 30 minutes before sexual activity

	50 mg if coadministered with an alpha-1 blocker or moderate inhibitor of CYP3A4 Administration with a high-fat meal may delay absorption

Sildenafil can cause transient “blue vision” in approximately 3% of men. This is caused by the cross-reactivity of the drug with the PDE-6 inhibitor present in the retina.

Rare cases of nonarteritic anterior ischemic optic neuropathy have been reported in men taking sildenafil and tadalafil.

Sporadic reports have indicated a risk of sudden unilateral hearing loss with sildenafil, vardenafil, and tadalafil.

Drug Interactions. All PDE-5Is are contraindicated in men on nitrates. Concurrent use with alpha-adrenergic antagonists causes symptomatic hypotension. Drugs that block the CYP3A4 metabolic pathway can prolong the half-life of sildenafil and thus should be avoided.

Choice of Drug. Current data indicates that all PDE-5Is are of equivalent efficacy and have a similar side-effect profile. Therefore, the choice of agent is based on patient preference, cost, and desired duration of action.

TESTOSTERONE. Testosterone regulates sexual desire and modulates almost every component involved in erectile function. There is a significant overlap between hypogonadism and ED. The diagnosis of testosterone deficiency is made on the basis of the clinical symptoms of hypogonadism, and consistently subnormal serum testosterone concentrations (below 8 nmol/L or 231 ng/dL) at 8 to 10 AM on three occasions.

Testosterone improves libido in men with low testosterone levels but does not necessarily improve ED. As adequate testosterone levels are necessary for NO function, the addition of testosterone to PDE-5Is may improve erectile function in hypogonadal men who did not respond to initial treatment with PDE-5Is alone.

Testosterone is available in different formulations, including oral tablets, transdermal gel/solution or patch, buccal application, subcutaneous implants, and intramuscular injections. Dosing varies depending on the formulation. Dosing is gradually titrated to restore the serum testosterone concentration to a normal range of 13.9 to 27.0 nmol/L or 400 to 700 ng/dL. Monitoring serum testosterone should be done 2 to 3 months after treatment initiation and after changing the dose. Monitoring is then extended to every 6 to 12 months once the dose is stabilized.

Testosterone is contraindicated in men with a history of prostate cancer.

Common side effects of testosterone therapy include exacerbation of benign prostatic hyperplasia, worsening of OSA, erythrocytosis, venous thromboembolism, and skin irritation in patients using the transdermal formulation.

SECOND-LINE THERAPIES. If PDE-5Is are ineffective, second-line therapies include vacuum erection devices, penile injections with vasodilating agents, or intraurethral alprostadil.

Vacuum-Assisted Erection Devices. Vacuum-assisted erection devices are usually tried first as they are less expensive and noninvasive, and some are available over the counter. They apply continuous negative pressure to the penis, drawing blood into the corpora cavernosa, which is then retained by an elastic band applied at the base of the penis. The vacuum erection device may be combined with a PDE-5I to augment the effect of the PDE-5I. Use should be limited to 30 minutes, and caution needs to be exerted in patients who receive anticoagulation treatment. The success rate varies between 60% to 70%, and patient satisfaction ranges between 25% to 49%.

Penile Self-Injection. Intracavernosal injection therapy with alprostadil (prostaglandin E1 [PGE1]) is the only FDA-approved drug for penile self-injection. Alprostadil has vasodilatory properties, which cause inhibition of sympathetic vasoconstriction and direct smooth muscle relaxation, which causes an increased blood flow to the penis. A combination of increased arterial inflow and decreased venous outflow from the corpora cavernosa creates an erection. Off-label compounded mixtures are available in other countries; these include combinations of vasoactive intestinal peptide (VIP) and phentolamine (Invicorp), phentolamine and papaverine (Bimix) with PGE1 sometimes added as a third component (Trimix).

Compounded mixtures have excellent efficacy. Side effects include penile plaque development, penile pain, and priapism.

Intraurethral Alprostadil. Intraurethral administration of alprostadil is a less invasive alternative to intrapenile injection; however, it is more costly and less effective. The alprostadil pellet is inserted into the urethra, and then the penis is massaged for 1 minute. Doses include 125, 250, 500, and 1,000 mcg. Side effects include penile pain and bleeding.

SURGICAL OPTIONS. Surgical management of ED should be reserved for men who cannot use, or who have not responded to, first- and second-line therapies.

Vascular surgery, including arterial bypass and venous ligation surgery, can improve vasculogenic ED. Given that such procedures target the etiology of the ED, in appropriate candidates, the surgery can be curative.

Surgical implantation of an inflatable or malleable (semirigid) penile prosthesis is an invasive and irreversible form of treatment, generally reserved for patients who have failed other types of treatment (third-line treatment).

Premature (Rapid/Early) Ejaculation. Three essential criteria define PE: (1) brief ejaculatory latency; (2) loss of control; and (3) psychological distress in the patient and/or partner. Subtypes include global versus situational, lifelong versus acquired, and coexisting with other SD, including ED. Management depends on the etiology, and approved treatments include SSRIs, topical anesthetics, and psychotherapy for psychogenic factors.

SSRIs are considered first-line treatment: paroxetine (10 to 40 mg/day), sertraline (50 to 200 mg/day), fluoxetine (20 to 40 mg/day), citalopram (20 to 40 mg/day), and escitalopram (10 to 20 mg/day), with paroxetine suggested to be the most effective. The full effect of SSRIs may require 2 to 3 weeks to occur. Clomipramine (12.5 to 50 mg/day) may be tried if SSRIs are ineffective. PDE-5Is may help patients with co-occurring ED. The usual order of treatment in men with coexisting ED and PE is to start with PDE-5Is, then add SSRIs if the patient still has PE.

Tramadol, an analgesic with actions on opioid, serotonin, and norepinephrine receptors, can be used as a third-line treatment but caution needs to be exercised, given the risk of physiologic dependence and respiratory depression.

Topical anesthetics like lidocaine–prilocaine spray have been found to improve ejaculatory latency, ejaculatory control, and sexual satisfaction. Topical alprostadil cream was also found to be beneficial.

Men who have psychogenic PE and/or couples' issues benefit from combined pharmacologic and behavioral treatment as compared to pharmacotherapy alone.

The squeeze technique involves stimulation of the penis until just before orgasm, followed by exerting pressure where the head meets the shaft of the penis to inhibit ejaculation for about 30 seconds and then repeating the steps as desired.

PSYCHOTROPIC-INDUCED SEXUAL DYSFUNCTION AND ITS MANAGEMENT

Psychotropic-induced sexual dysfunction is common but frequently not recognized/addressed in clinical practice. Patients may be reluctant to report this as a side effect, and physicians may not routinely screen for sexual dysfunction or have limited knowledge about its management. Lack of recognition and treatment of sexual dysfunction can negatively impact a patient's quality of life, be an additional source of distress for them and/or their partners, and may ultimately lead to medication nonadherence, which can further exacerbate the psychiatric condition.

The most studied psychotropics with regards to impact on sexual function are antidepressants, given serotonin's role in sexual function. SSRIs and serotonin–norepinephrine reuptake inhibitors (SNRIs) can negatively affect all three sexual response phases. Men more frequently report problems with desire and orgasm, while women more frequently report decreased arousal. A meta-analysis by Serretti and Chiesa suggests that sertraline, venlafaxine, citalopram, paroxetine, and fluoxetine are associated with the highest rates of treatment-emergent sexual dysfunction (with prevalence rates between 70% to 80%); imipramine, phenelzine, duloxetine, escitalopram, and fluvoxamine have lower rates of sexual dysfunction (25% to 45%), but higher than that of placebo. No significant difference from placebo was observed for bupropion, vilazodone, and vortioxetine. Clinical and research use of nefazodone has markedly declined in the United States as a result of the risk of hepatotoxicity.

Bupropion, a dopamine and norepinephrine reuptake inhibitor; buspirone, an agonist of 5-HT1A receptors; and trazodone, a serotonin transporter inhibitor plus 5-HT2A and 5-HT2C receptor antagonist, are not associated with sexual dysfunction. Other atypical antidepressants, including vilazodone and vortioxetine, appear to have lower rates of sexual dysfunction compared to SSRI/SNRIs. Studies on sexual dysfunction related to tricyclic antidepressants and monoamine oxidase inhibitors are scarce, although studies have found that transdermal selegiline has rates of sexual dysfunction comparable to placebo.

When there is evidence of sexual dysfunction in patients experiencing depression, one recommendation is to select a medication with a lower incidence of sexual dysfunction. If sexual dysfunction emerges during the course of treatment with an antidepressant, strategies include switching to a medication with a lower risk of sexual side effects, use of an antidote or medication that will help reverse the sexual dysfunction, waiting for potential resolution/tolerance to the sexual side effect, and lowering the dose of medication. Any of these strategies must be balanced with the risk of exacerbation of psychiatric symptoms. In men, there is evidence supporting the use of PDE-5Is for the treatment of erectile dysfunction induced by antidepressant medications. Bupropion, mirtazapine, and buspirone can be used as augmenting agents to treat antidepressant-induced sexual dysfunction. In a 2013 Cochrane review of the management of antidepressant-induced sexual dysfunction, bupropion (at doses of at least 300 mg daily) was the only medication, aside from the as-needed use of sildenafil in men found to be statistically significant in reversing treatment-emergent SD in both men and women.

Antipsychotic medications can affect all three phases of the sexual response. Sexual dysfunction is thought to occur via dopaminergic antagonism within the tuberoinfundibular system, which causes an elevation in plasma prolactin levels. Head-to-head comparisons are unavailable, but studies suggest that drugs that cause more hyperprolactinemia like risperidone and its metabolite, paliperidone; first-generation antipsychotics (e.g., chlorpromazine, pimozide, thioridazine, thiothixene); and haloperidol are most likely to cause sexual dysfunction. Olanzapine, clozapine, ziprasidone, and quetiapine have slightly lower rates; for olanzapine, this may be dose dependent, with doses >15 mg/day associated with greater sexual dysfunction. Aripiprazole appears to have the lowest incidence of sexual dysfunction compared to other antipsychotics; it may, to some degree, reverse the sexual adverse effects caused by antipsychotics and antidepressants. There is limited but promising data on newer antipsychotics like lurasidone, iloperidone, brexpiprazole, and cariprazine. There is insufficient evidence regarding the treatment of antipsychotic-induced sexual dysfunction. Some authors have proposed dose reduction and switching to a prolactin-sparing second-generation antipsychotic as possible strategies. PDE-5Is may also be of benefit in men with ED. Other treatment strategies include waiting for spontaneous remission or tolerance to the sexual side effects or switching to an antipsychotic less likely to cause sexual dysfunction (aripiprazole). Drug holidays and dose reductions are not recommended due to the risk of psychosis or medication discontinuation.

Limited data is available on sexual dysfunction associated with mood stabilizers. Most of the available data on anticonvulsants originates from studies in epilepsy, which itself can be associated with sexual dysfunction. There is evidence that anticonvulsants that lower bioavailable testosterone (carbamazepine, phenytoin) are more likely to cause sexual dysfunction; this is likely related to the increased metabolism of sex hormones and increased hepatic synthesis of SHBG. Lithium has also been associated with sexual dysfunction, although the mechanism is currently not well understood. Lamotrigine does not appear to cause sexual dysfunction.

There is mixed evidence regarding the use of benzodiazepines and sexual dysfunction. In the treatment of attention deficit hyperactivity disorder (ADHD), there is some evidence that atomoxetine has a greater association with SD compared to placebo.

Mr. J is a 30-year-old man with no active medical problems and a history of Major depressive disorder (MDD) whose initial treatment with a maximum dose of SSRI achieved a partial response. Risperidone 1 mg was added as an inexpensive augmenting agent resulting in full remission of his depressive symptoms. Within a month, he developed ED. No change in the relationship with his wife preceded the ED, but he began to question his attractiveness to his wife and began avoiding sex. Despite his wife's supportive response to his ED, he developed anxiety associated with sexual activity and behavioral avoidance. He did not spontaneously report this side effect to his psychiatrist until questioned directly. The onset of ED seemed to correlate with the addition of risperidone to the high dose of SSRI. Interpersonal issues appeared to be the result of the ED rather than the cause. Medical workup revealed no substance use disorders or recent changes in the use of substances, no cardiovascular risk factors, normal lipid profile and HbA1c, normal testosterone, and slightly elevated prolactin. Sildenafil was added at 50 mg prior to sexual activity. At follow-up, Mr. J reported complete resolution of the ED and remained on the same medication regimen.

Mrs. A is a 32-year-old married woman who complains of a lack of sexual desire since the birth of their second child. She initially thought her lack of interest would recover once the baby started sleeping through the night, but it has persisted for 3 years. She denies relationship difficulties saying she loves her husband, with whom she shares values and communicates well. However, she is worried that her lack of desire might cause a problem due to avoidance behaviors such as avoiding casual intimacy like kissing, going to bed later than her husband, and turning him down about 50% of the time when he approaches her for sex, otherwise participating out of a sense of obligation. When sexually active with her husband, she is able to become aroused and sometimes experiences an orgasm, but this does not increase her desire to have sex the next time. She reports frustration with this problem and is seeking a remedy. She has two children delivered vaginally but no ongoing medical problems or use of daily medications. She has a glass of wine on weekend nights and denies the use of drugs (prescription or illicit drugs). Her husband has had a vasectomy. She exercises regularly, sleeps 7 to 8 hours per night, and is functional at work. She denies depressed or anxious mood. She completed a DSDS which indicates HSDD without current modifiable factors.

If there were identified difficulties that needed to be modified (i.e., relational problems, medical conditions, substances including prescription medications, psychiatric conditions), then these could be addressed, but otherwise, laboratory workup is not warranted in this case. Management can include flibanserin, bremelanotide, and/or mindfulness-based CBT, depending on patient preference.

RECOMMENDATIONS

The evaluation of SD includes a thorough interview with a focus on medical, surgical, sexual, and psychosocial history. Management follows a biopsychosocial approach.

- The treatment for SD should be individualized, taking into account patient and partner's preferences, comorbid conditions, likely etiology of SD, and concomitant medications.
- Treatment starts with addressing modifiable factors (if any).
- There are fewer FDA-approved drugs for female SD versus male SD. Choice of treatment is based on risks versus benefits and patient preference.
- Dosing needs to be adjusted based on the response, and if symptoms persist despite adequate dosing, an additional treatment option should be explored.
- Treatment must include addressing co-occurring depressive disorders.
- Augmentation with bupropion or buspirone may be an option in patients whose depression responded well to an SSRI/SNRI and are experiencing significant SD.
- In patients who did not fully respond to SSRI/SNRI treatment and have significant SD, switching to a medication with less ED risk (mirtazapine, bupropion, vortioxetine, vilazodone) may be the appropriate next step in treatment.

FURTHER READINGS

Allen MS, Walter EE. Erectile dysfunction: an umbrella review of meta-analyses of risk-factors, treatment, and prevalence outcomes. ***J Sex Med***. 2019;16(4):531–541.

Bakr AM, El-Sakka AA, El-Sakka AI. Considerations for prescribing pharmacotherapy for the treatment of erectile dysfunction. ***Expert Opin Pharmacother***. 2021;22(7):821–834.

Basson R, Wierman ME, van Lankveld J, Brotto L. Summary of the recommendations on sexual dysfunctions in women. ***J Sex Med***. 2010;7(1 Pt 2):314–326.

Burnett AL, Nehra A, Breau RH, et al. Erectile dysfunction: AUA guideline. ***J Urol***. 2018;200(3):633–641.

Chew PY, CL, Sidi HB, et al. The association between female sexual dysfunction and sexual dysfunction in the male partner: a systematic review and meta-analysis. ***J Sex Med***. 2021;18(1):99–112.

Clayton AH, Alkis AR, Parikh NB, Votta JG. Sexual dysfunction due to psychotropic medications. ***Psychiatr Clin North Am***. 2016;39(3):427–463.

Clayton AH, El Haddad S, Iluonakhamhe JP, Martinez CP, Schuck AE. Sexual dysfunction associated with major depressive disorder and antidepressant treatment. ***Expert Opin Drug Saf***. 2014;13(10):1361–1374.

Clayton AH, Goldstein I, Kim NN, et al. The international society for the study of women's sexual health process of care for management of hypoactive sexual desire disorder in women. ***Mayo Clin Proc***. 2018;93(4):467–487.

Clayton AH, Montejo AL. Major depressive disorder, antidepressants, and sexual dysfunction. ***J Clin Psychiatry***. 2006;67(Suppl 6):33–37.

Feldman HA, Goldstein I, Hatzichristou DG, Krane RJ, McKinlay JB. Impotence and its medical and psychosocial correlates: results of the Massachusetts male aging study. ***J Urol***. 1994;151(1):54–61.

Fooladi E, Davis SR. An update on the pharmacological management of female sexual dysfunction. ***Expert Opin Pharmacother***. 2012;13(15):2131–2142.

Harsh V, Clayton AH. Sex differences in the treatment of sexual dysfunction. ***Curr Psychiatry Rep***. 2018;20(3):18.

Invernizzi RW, Sacchetti G, Parini S, Acconcia S, Samanin R. Flibanserin, a potential antidepressant drug, lowers 5-HT and raises dopamine and noradrenaline in the rat prefrontal cortex dialysate: role of 5-HT1A receptors. ***Br J Pharmacol***. 2003;139(7):1281–1288.

Kingsberg SA, Simon JA. Female hypoactive sexual desire disorder: a practical guide to causes, clinical diagnosis, and treatment. ***J Womens Health (Larchmt)***. 2020;29(8):1101–1112.

McMahon CG. Current diagnosis and management of erectile dysfunction. ***Med J Aust***. 2019;210(10):469–476.

Montejo AL, de Alarcón R, Prieto N, Acosta JM, Buch B, Montejo L. Management strategies for antipsychotic-related sexual dysfunction: a clinical approach. ***J Clin Med***. 2021;10(2):308.

Montejo AL, Prieto N, de Alarcón R, Casado-Espada N, de la Iglesia J, Montejo L. Management strategies for antidepressant-related sexual dysfunction: a clinical approach. ***J Clin Med***. 2019;8(10):1640.

Montorsi F, Adaikan G, Becher E, et al. Summary of the recommendations on sexual dysfunctions in men. ***J Sex Med***. 2010;7(11):3572–3588.

Morgado A, Tomada N, Sarquella J. Surgical treatment of erectile dysfunction: historical perspective and overview. In: Martins FE, Kulkarni SB, Köhler TS, eds. ***Textbook of Male Genitourethral Reconstruction***. Springer; 2020:547–559.

Nappi RE, Cucinella L. Advances in pharmacotherapy for treating female sexual dysfunction. ***Expert Opin Pharmacother***. 2015;16(6):875–887.

Parish SJ, Cottler-Casanova S, Clayton AH, McCabe MP, Coleman E, Reed GM. The evolution of the female sexual disorder/dysfunction definitions, nomenclature, and classifications: a review of DSM, ICSM, ISSWSH, and ICD. ***Sex Med Rev***. 2021;9(1):36–56.

Segraves RT, Clayton A, Croft H, Wolf A, Warnock J. Bupropion sustained release for the treatment of hypoactive sexual desire disorder in premenopausal women. ***J Clin Psychopharmacol***. 2004;24(3):339–342.

Selvin E, Burnett AL, Platz EA. Prevalence and risk factors for erectile dysfunction in the US. ***Am J Med***. 2007;120(2):151–157.

Serretti A, Chiesa A. A meta-analysis of sexual dysfunction in psychiatric patients taking antipsychotics. ***Int Clin Psychopharmacol***. 2011;26(3):130–140.

Serretti A, Chiesa A. Treatment-emergent sexual dysfunction related to antidepressants: a meta-analysis. ***J Clin Psychopharmacol***. 2009;29(3):259–266.

Shamloul R, Ghanem H. Erectile dysfunction. ***Lancet***. 2013;381(9861):153–165.

World Health Organization. ICD-10 Version: 2019. Chapter V, Section F52. Published 2019. Accessed April 5, 2021. https://icd.who.int/browse10/2019/en

Zaazaa A, Bella AJ, Shamloul R. Drug addiction and sexual dysfunction. ***Endocrinol Metab Clin North Am***. 2013;42(3):585–592.

Zhao S, Deng T, Luo L, et al. Association between opioid use and risk of erectile dysfunction: a systematic review and meta-analysis. ***J Sex Med***. 2017;14(10):1209–1219.

33.18 Psychopharmacology Across the Life Cycle

33.18a Pediatric Psychopharmacology

JEFFREY R. STRAWN, M.D., AND LAURA B. RAMSEY, PH.D.

Psychotropic medications play a critical role in treating many psychiatric disorders in children and adolescents. However, using these medications effectively and appropriately requires an understanding of pediatric pharmacokinetics and pharmacodynamics in addition to the nuances of response that are revealed by numerous double-blind, placebo-controlled trials. These concepts are reviewed in Chapter 53.5, and we review the developmental pharmacology relevant to psychotropic medication use in pediatric populations and the evidence for multiple psychotropics across internalizing and externalizing disorders.

Briefly, pharmacokinetics and pharmacodynamics in children and adolescents differ considerably from adults, and these differences give rise to differences in dosing, differences in response, and differences in tolerability. Importantly, pharmacokinetic and pharmacodynamic factors interact and are mutually influenced by development. These pharmacokinetic factors and pharmacodynamic factors subtend differences in side effects (e.g., extrapyramidal symptoms), dosing (e.g., lithium), and efficacy (e.g., benzodiazepines) between children, adolescents, and adults.

ROLE OF PHARMACOTHERAPY IN CHILDREN AND ADOLESCENTS

While psychotropic medications often play an integral part in the treatment of children and adolescents with psychiatric disorders, they are usually used in combination with other modalities. Depending on the individual, a treatment plan may warrant individual, group, and/or family therapy; educational interventions; parent training; substance use disorder treatment; or any of several additional interventions.

PEDIATRIC ANXIETY AND DEPRESSIVE DISORDERS

Multiple meta-analyses suggest that selective serotonin reuptake inhibitors (SSRIs) are superior to other medication classes for pediatric anxiety disorders. To date, sertraline, fluoxetine, escitalopram, fluvoxamine, and paroxetine have been evaluated in pediatric anxiety disorders. In addition, meta-analyses of SSRIs and serotonin norepinephrine reuptake inhibitor (SNRIs) in youth with anxiety disorders suggest that SSRIs produce faster and greater improvements compared to SNRIs.

For youth with major depressive disorder, multiple randomized controlled trial (RCTs) support the efficacy of fluoxetine and escitalopram in youths with major depressive disorder (MDD)—and both are food and drug administration (FDA) approved for MDD in pediatric patients. Beyond fluoxetine and escitalopram, several other SSRIs have demonstrated efficacy in youth with MDD. However, the use of antidepressant medications in youth must consider the FDA boxed warning that any antidepressant use in children and adolescents and young adults was accompanied by an increased risk for suicidal thinking and behavior, despite most recent meta-analyses failing to replicate the original analyses from 2004.

PHARMACOTHERAPIES FOR BIPOLAR DISORDER AND SCHIZOPHRENIA IN YOUTH

Placebo-controlled studies suggest that multiple mixed dopamine and serotonin receptor antagonists are efficacious in pediatric patients with acute mania, and six agents have received FDA approval for pediatric mania (lithium, risperidone, aripiprazole, quetiapine, olanzapine, and asenapine). In general, antimanic effects are observed within the first 2 weeks of treatment and mood stabilizers may be used as second-line treatments or in combination with mixed dopamine receptor antagonists. While the anticonvulsants likely have primary roles in the treatment of adults with bipolar disorder, these agents have produced very mixed results in youths.

Aripiprazole, lurasidone, olanzapine, quetiapine, risperidone, and paliperidone have received FDA approval for the treatment of adolescents with schizophrenia aged 13 to 17 years. The mixed dopamine–serotonin receptor antagonists are considered first-line medications for adolescents with schizophrenia. Studies involving clozapine (Clozaril) in treatment-refractory early-onset schizophrenia spectrum disorders have demonstrated the superiority of clozapine compared to haloperidol and olanzapine, with regard to reduction in negative symptoms and, to a certain degree, with the overall response.

ATTENTION DEFICIT HYPERACTIVITY DISORDER (ADHD)

Dopamine/norepinephrine reuptake inhibitors—including both amphetamine- and methylphenidate-based agents are highly efficacious in treating youth with attention deficit hyperactivity disorder (ADHD). In general, 65% to 75% of stimulant-treated youth response and response rates are similar for methylphenidate and amphetamine in immediate-release as well as extended-release preparations.

In addition to stimulants and norepinephrine reuptake inhibitors (e.g., viloxazine and atomoxetine), multiple RCTs support the efficacy of extended-release preparations of clonidine and guanfacine as monotherapy in children and adolescents with ADHD.

CONSIDERATIONS FOR EFFECTIVE PHARMACOTHERAPY IN CHILDREN AND ADOLESCENTS

There is substantial evidence for pharmacologic treatments for psychiatric disorders in children and adolescents. Pharmacokinetics and pharmacodynamics in children and adolescents differ considerably from adults, and these differences give rise to differences in dosing, differences in response, and differences in tolerability. As such, safe and effective use of pharmacotherapy requires an understanding of pediatric pharmacokinetics and pharmacodynamics in addition to the nuances of response that are revealed by numerous double-blind, placebo-controlled trials.

33.18b Biologic Therapies: Geriatric Psychopharmacology

Ari B. Cuperfain, M.D., M.Sc., and Bruce G. Pollock, M.D., Ph.D., F.R.C.P.C.

Sixteen percent of the U.S. population who are older than age 65 consume 40% of prescription drugs. This polypharmacy is ubiquitous, and indeed, the most common treatable illness in older adults is due to the numerous medications (including prescribed, over the counter, and herbal) that they take. Not infrequently, prescribing cascades occur when additional medications are added to treat the side effects of prior medications. The Beers criteria lists medications that are inappropriate for older adults. Wherever possible, physicians should deprescribe inappropriate or less essential medications. This is often difficult to do, and deprescribing.org can provide some guidance on this process.

Pharmacokinetic and pharmacodynamic changes associated with aging and illness render older patients more sensitive than younger adults to the adverse effects of medications. The most common age-associated changes affecting pharmacokinetics are related to distribution and excretion.

With age, adipose tissue increases proportionally compared to body water in an older adult due to a loss of muscle mass. Thus, the volume of distribution and half-life for lipid-soluble drugs (most psychotropics) increases while it decreases for water-soluble drugs (e.g., lithium, digoxin [Lanoxin]). Lipid-soluble drugs, especially those with long half-lives (e.g., diazepam [Valium], clonazepam [Klonopin], aripiprazole [Abilify]), accumulate in fat and clear more slowly from the body. For water-soluble drugs, the therapeutic window shrinks as transient increases in plasma concentrations are less effectively buffered. Plasma proteins (e.g., albumin) can also decrease with age, meaning drug concentrations for highly protein-bound drugs (e.g., valproic acid [Depakote]) are less reliable.

For excretion, the average glomerular filtration rate decreases by 50% between the ages of 20 and 70 years. Decreased renal clearance may require dose adjustment for drugs (or metabolites) that are renally cleared, most notably lithium, but also others, including gabapentin (Neurontin), pregabalin (Lyrica), topiramate (Topamax), amantadine (Symmetrel), paliperidone (Invega), memantine (Namenda), rivastigmine (Exelon), and desvenlafaxine (Pristiq). Moreover, excretion can be further impaired by commonly used nonsteroidal anti-inflammatory drugs (NSAIDs) and angiotensin-converting enzyme (ACE) inhibitors.

Although there is no uniform age-associated decline in cytochrome P (CYP) 450 enzymes, reductions in hepatic mass and blood flow place greater emphasis on interindividual differences in drug metabolic capacity as well as reducing clearance of normally high-clearance CYP3A4 metabolized drugs. Psychotropics metabolized by this isozyme include alprazolam (Xanax), sertraline (Zoloft), and mirtazapine (Remeron). Inhibition or induction of enzymes, often due to drug–drug interactions, cause wide variation in drug concentrations. It is important to recognize isozyme specificity in CYP drug metabolism and then to know which medications may interact with a given isozyme. For example, CYP1A2 metabolizes clozapine (Clozaril), asenapine (Saphris), olanzapine (Zyprexa), and fluvoxamine (Luvox). 2D6 metabolizes duloxetine (Cymbalta), venlafaxine (Effexor), paroxetine (Paxil), brexpiprazole (Rexulti), risperidone (Risperdal), and many tricyclic antidepressants. 2C19 metabolizes diazepam and phenytoin (Dilantin). Phase II metabolism in older patients is understudied, although it may also decrease with age.

Pharmacodynamic considerations must also factor into clinical decision making. The density of drug targets, such as the dopaminergic and cholinergic receptors, is generally reduced, resulting in increased sensitivity to D2 antagonists and antimuscarinic agents. Reduced β-receptor sensitivity contributes to diminished baroreceptor reflex responses, exacerbating orthostatic hypotension in response to α2-agonists and other drugs. Atypical antipsychotics have been shown to increase mortality in patients with dementia. Constipation is a common side effect of many psychotropic medications and should be monitored. Risks of bleeding and hyponatremia with selective serotonin reuptake inhibitor (SSRI) treatment are more pronounced in older adults. All psychotropics may increase the risk of falls and may increase cognitive impairment in older patients. In addition to prescribed medications, older adults are also more sensitive to recreational substances and over-the-counter nutraceuticals, and these may figure in drug interactions. Future areas for innovation include disease-modifying agents for Alzheimer disease and increased representation of older adults in clinical trials.

33.19 Biologic Therapies Around the Globe

33.19a Essential List of Medicines

Sasson Zemach, M.D., Daniel Minkin Levy, M.D., and Joseph Zohar, M.D.

Similar to all other fields of medicine, therapeutic options in psychiatry differ greatly worldwide. While the biologic therapies section reflects the current treatment options in the North American and Western European continents, the case is naturally different in the developing world and in regions of distinct cultural differences. The “Biologic Therapies Around the Globe” subsection aims to give a brief introduction to the current tools available in three central regions: Africa, India, and Latin America (mostly Brazil).

THE ESSENTIAL LIST OF MEDICINES

In order to obtain a basic “package” of vital medications, health care systems create national lists of essential medicines (nEML). These commonly rely on the infrastructure of the World Health Organization (WHO)’s Essential Medicines List (EML). First issued in 1977, The EML model is a recommendation document published biannually and includes the basic medications needed to “satisfy the priority health care needs of the population” (from the WHO website). According to the WHO website, “essential medicines are selected with due regard to disease prevalence and public health relevance, evidence of clinical efficacy and safety, and comparative costs and cost-effectiveness.”

The 22nd version was published in September 2021. It is divided into a core list, namely a basic list of the most efficient, safe, and affordable drugs, and a complementary list which includes medications that are too essential but require specific care of specialized facilities/require special training/consistently more costly or maybe less cost-effective.

Table 33.19a–1.

A Modified Table of the WHO’s Essential List of Medicines Recommended for Mental and Behavioral Disorders

Drug	Dosage and Preparation	Comments
Medicines Used in Psychotic Disorders		
Chlorpromazine[a,b]	**Injection:** 25 mg/mL (hydrochloride) in 2 mL ampoule **Oral liquid:** 25 mg/5 mL (hydrochloride) **Tablet:** ***10 mg; 25 mg; 50 mg; 100 mg (hydrochloride)***	Restricted use in children and adolescents
Fluphenazine[a]	**Injection:** 25 mg (decanoate or enantate) in 1 mL ampoule	
Haloperidol[a,b]	**Injection:** 5 mg in 1 mL ampoule **Tablet:** 2 mg; 5 mg	Restricted use in children
Paliperidone[a]	**Injection (prolonged-release):** 25 mg; 50 mg; 75 mg; 100 mg; 150 mg (as palmitate) in prefilled syringe	Therapeutic alternatives: - risperidone injection
Risperidone[a]	**Solid oral dosage form:** 0.25–6.0 mg	
Clozapine[b]	***Solid oral dosage form: 25–200 mg***	
Medicines Used in Mood Disorders		
Medicines used in depressive disorders		
Amitriptyline[a]	**Tablet:** 25 mg; 75 mg (hydrochloride)	
Fluoxetine[a,b]	**Solid oral dosage form:** 20 mg (as hydrochloride)	Use in children requires special training; age restriction >8 years Therapeutic alternatives: - citalopram - escitalopram - fluvoxamine - paroxetine - sertraline
Medicines used in bipolar disorders		
Carbamazepine[a]	**Tablet (scored):** 100 mg;	

	200 mg	
lithium carbonate[a]	**Solid oral dosage form:** 300 mg	
valproic acid (sodium valproate)[a]	**Tablet (enteric coated):** 200 mg; 500 mg	***Avoid use in pregnancy and in women and girls of childbearing potential, unless alternative treatments are ineffective or not tolerated because of the high risk of birth defects and developmental disorders in children exposed to valproate in the womb***
Medicines for Anxiety Disorders		
Diazepam[a]	**Tablet (scored):** 2 mg; 5 mg	
Medicines Used for Obsessive-Compulsive Disorders		
Clomipramine[a]	**Capsule:** 10 mg; 25 mg (hydrochloride)	
Medicines for Disorders due to Psychoactive Substance Use		
Bupropion[a]	**Tablet (sustained release):** 150 mg (hydrochloride)	
Nicotine replacement therapy (NRT)[a]	**Chewing gum:** 2 mg; 4 mg (as polacrilex) **Transdermal patch:** 5–30 mg/16 hours; 7–21 mg/24 hours	
Varenicline[a]	**Tablet:** 0.5 mg, 1 mg	
Methadone[b]	***Concentrate for oral liquid: 5 mg/mL; 10 mg/mL (hydrochloride)*** ***Oral liquid: 5 mg/5 mL; 10 mg/5 mL (hydrochloride)***	Therapeutic alternatives should be considered, ***Therapeutic alternatives:*** ***- buprenorphine***

---

[a] Core list.

[b] Complimentary list. Medications may be included both in the core list and complimentary list. Information is retrieved from the WHO's full Essential List of Medicines issued at Sep. 2021.

Overall, of the 479 medications included (September 2021), 17 medications for mental and behavioral disorders are included (Table 33.19a–1).

Parallel to the call for a broadening of the list to meet the global burden of psychiatric disorders (mainly depression and attention deficit hyperactivity disorder [ADHD]), the criteria for inclusion of new agents remain complex and multifactorial. In 2018, the American Psychiatric Association (APA) approved an action plan aimed to help revise the EML in collaboration with the WHO to help provide well-founded recommendations. Nevertheless, the list is updated gradually, with risperidone and clozapine added in 2013 and paliperidone long-acting injectables (LAIs) added in 2021, along with bupropion and varenicline for the indication of smoking cessation.

In summary, EML helps to ensure a sufficient level of health and is related to the increased understanding that a feasible standard of health, including mental health, is a global necessity and a global responsibility.

FURTHER READINGS

American Psychiatric Association (APA). November 2018 Assembly Action Papers. http://www.area3apa.net/uploads/1/8/4/8/18487482/actions_papers-_nov_2018.pdf

*Executive summary. The Selection and Use of Essential Medicines 2021. https://apps.who.int/iris/handle/10665/345554

Machado IK, Lippert C, Ostacher MJ. Incorporating clinical research into recommendations for addressing global mental health needs: updating the World Health Organization's Essential Medicines List. ***Compr Psychiatry***. 2021;109:152245.

*WHO Model List of Essential Medicines. 22nd List (2021). https://www.who.int/publications/i/item/WHO-MHP-HPS-EML-2021.02

World Health Organization (WHO). General information. https://web.archive.org/web/20140527003625/http://www.who.int/medicines/events/fs/en/

33.19b Biologic Therapies in Africa

OYE GUREJE, PH.D., D.SC., F.R.C.PSYCH., F.W.A.C.P., AND OLATUNDE AYINDE, M.B.B.S., M.S., F.W.A.C.P.

The major mental disorders are present wherever they have been sought in Africa. Even though rates may vary within the continent and sometimes differ from what has been reported in other parts of the world, most of the disorders described in both DSM-5 and ICD-11 have been identified in populations across the continent. These similarities with other parts of the world would suggest that considerations of biologic therapies as they apply in those parts will also be relevant to Africa. However, features that may affect drug metabolism and that may differ between population groups could produce substantially different profiles of efficacy and side effects and therefore require a need for specific consideration of biologic therapies in Africa. Examples of such features with a potential effect on the mechanisms of action of biologic therapies are genetics, lifestyles, diet, and the age structure of the population. Beyond these potential differences in the metabolism or mechanisms of action of biologic therapies, factors that may be of far greater relevance are those associated with cost, availability, and access. These factors may have a profound impact on the utility of specific therapies in ways that go beyond their biologic profile.

Africa is diverse in regard to culture, social norms and practices, and also in respect to the organization of the health systems of the countries on the continent. Any discussion of the continent about therapeutics and health care will therefore have to be considered as providing broad generalizations. This chapter is thus an attempt to provide such a broad picture that is nevertheless informative. In doing so, it should be made clear that no allusion is being made to any form of racial similarity between the populations resident on the continent or to possible differences between them and the rest of the world. Our current knowledge cautions against any such simplistic use of race, a concept that has essentially been rendered obsolete by genetics.

OVERVIEW OF THE EFFICACY AND SIDE EFFECTS OF BIOLOGIC THERAPIES IN AFRICA

Only a few studies have compared the efficacy and side-effect profile of different forms of biologic therapies among Africans. An early study of the correlation between chlorpromazine metabolism and clinical response among Nigerian patients with schizophrenia identified three groups: responders, nonresponders who subsequently improved with haloperidol, and nonresponders who improved with electroconvulsive therapy (ECT). A study in South Africa that compared oral antipsychotics, specifically haloperidol or risperidone, with risperidone long-acting injection among patients with early psychosis found the latter to have advantages in terms of efficacy and fewer extrapyramidal symptoms but was associated with more weight gain. An open-label study conducted in Nigeria comparing patients treated with a combination of ECT and pharmacotherapy with first-generation antipsychotics with those who received pharmacotherapy alone found that the former had a shorter duration of hospitalization even though they were more likely to report subjective but short-lived memory impairment. A study conducted in Malawi found adverse drug reactions to be common among patients receiving chlorpromazine or haloperidol, with more severe reactions among males, patients aged 35 years or more, and those receiving chlorpromazine. Thus, and in general, the findings of these studies have generally been broadly like what has been reported from other groups across the world. This general observation is supported by the results of a systematic review of 26 studies of antipsychotic medication in Africa, which found improvement in psychiatric symptoms with use, frequent occurrence of side effects, including metabolic syndrome, and a common difficulty with adherence by patients with severe mental disorders.

A few studies have sought to explore the use of biologic therapies in the context of some peculiarities of the continent. An important consideration is the concurrent use of psychotropics with medication for physical health conditions that are common in Africa. In this regard, an early study found marked increases in the plasma levels of chlorpromazine when used concurrently with antimalarial agents, thus raising the possibility of an elevated risk of toxicity. Unfortunately, no subsequent exploration of this interaction has been reported in the literature. Along the same line, and given the common co-occurrence of psychiatric illness and HIV and evidence suggesting increased incident neurocognitive impairment among a subset of persons with this comorbidity, studies exploring the effectiveness of psychotropic medications in persons living with HIV are needed. The potential interactions between cytochrome P450 metabolism and the nonnucleoside reverse-transcriptase inhibitors provide a good justification for such studies.

SPECIFIC EXAMPLES OF PSYCHOTROPIC MEDICATIONS

Both first-generation antipsychotics and newer medications in oral (haloperidol, quetiapine, ziprasidone) and depot preparations (flupentixol, risperidonem, and ziprasidone) have been shown in a systematic review of cohort studies and randomized trials to be effective in reducing symptoms and improving functioning in African patients with psychosis. A better response was predicted by a shorter duration of illness, less comorbidity, less cognitive impairment, fewer negative symptoms, good premorbid functioning, and medication adherence. Relapses appeared to be associated with reduced effectiveness of antipsychotic medications. Extrapyramidal side effects (EPSE) were more likely in patients taking first-generation antipsychotics, while metabolic side effects and hyperprolactinemia were more associated with second-generation medication use. In a study conducted in Ethiopia among a cohort of 312 patients with bipolar disorder, being on both antipsychotics and antidepressants for more than 50% of the average follow-up period (2.5 years) or longer was a strong predictor of remission. Another study found that treatment-emergent EPSE was associated with spontaneous neurologic soft signs in antipsychotic-naive Yoruba Nigerians and Xhosa South Africans with schizophrenia.

The most commonly used antidepressants in most African settings are tricyclic antidepressants (TCAs) and selective serotonin reuptake inhibitors (SSRIs). For example, a study from South Africa indicates that TCAs and SSRIs are the most prescribed medications in the child and adolescent population, with amitriptyline and fluoxetine being the drugs of choice. In regard to comparative efficacy, a Nigerian trial comparing moclobemide (max dose 600 mg/day) to amitriptyline (max dose 150 mg/day) in 60 subjects found both to be safe and effective in treating depression, with no statistically significant difference between the two.

Anticonvulsants in common use in African settings include carbamazepine, sodium valproate, phenytoin, and lithium, with the commonest indication being epilepsy and mood disorders. A common difficulty with the use of lithium salt is the lack of the required laboratory backup to monitor patients or the inability of the patients to pay for the service. Payment out of pocket, a common way to access service in much of Africa, is indeed a major reason why adherence to treatment and related care remains a major problem in many settings in Africa.

ELECTROCONVULSIVE THERAPY

ECT is widely used in Africa. In Nigeria, for example, its use dates to the 1970s. A study published in 2015 reports the widespread availability of ECT in South Africa and indicates that, despite lingering challenges with access and training, its use follows international guidelines. Studies show a good attitude to ECT by psychiatrists of African origin. ECT has been used to manage patients with schizophrenia, mania, and severe depression, with those with poor response to medication or who have psychomotor retardation being the most likely to be prescribed the treatment. An early controlled study in Nigeria did not find any significant difference in symptom remission between two groups of patients with schizophrenia on antipsychotics (n = 20) who took six sessions of real or sham ECT, with both groups recording improvement over a 20-week follow-up period. An uncontrolled study in Malawi, however, found up to 90% improvement in a mixed population of 47 patients. The most common side effects reported in the study included cognitive problems (confusion and amnesia), muscle aches, headache, and lesions of the oral cavity.

Modified bilateral ECT has largely replaced unmodified administration that was common in the early days of ECT use on the continent. The average number of treatments in one Nigeria tertiary center is six. Patients appear to accept ECT, even when unmodified, and acknowledge its benefits, despite complaints of not being well informed about the procedure and deficiencies in ECT service delivery.

TRANSCRANIAL MAGNETIC RESONANCE STIMULATION

Transcranial Magnetic Stimulation is very rarely used on the continent, with only limited use in South Africa and in Nigeria, where it has only recently become available in one private clinic in Lagos. A recent study found accelerated theta-burst repetitive transcranial magnetic stimulation led to improvement in depressive symptoms by the eighth day of treatment in nine patients who underwent the procedure. Mild headache was the most commonly reported adverse effect.

ATTITUDES OF PATIENTS AND CLINICIANS TO PHYSICAL THERAPIES

Studies from different parts of Africa report good attitude toward antipsychotics among persons with psychosis, with better attitude reported when the formulation of choice matches the current formulation the patient is on, there is a relative absence of side effects, the patient has good insight and comes from a higher income household. On the other hand, a short duration of illness and the presence of side effects are associated with poor attitude. There are suggestions that patients may not be well educated by their physicians on the role of their medications in treatment. One study from Nigeria found that out of 200 patients, about three-quarters did not know the role of Benzhexol (Artane) in their treatment. Lack of information may partly explain the poor medication adherence that has been reported among patients on oral antipsychotics in some settings.

In the 1980s, the most widely prescribed antipsychotics in Nigeria were chlorpromazine, thioridazine, and trifluoperazine; antidepressants, amitriptyline and imipramine; anxiolytic, diazepam and chlordiazepoxide. This pattern has only changed to a limited extent in recent times and may also reflect the practice in many other parts of Africa. With the availability of generic forms, the use of second-generation antipsychotics has now become common on the continent. Still, polypharmacy is common among many clinicians. One study from Nigeria found that more than 50% of 320 outpatients were prescribed a combination of depot first-generation antipsychotics and oral second-generation medications. The commonest depot medications in use are fluphenazine and Flupentixol. As highlighted by Adeponle et al., polypharmacy encourages waste and has the potential to increase the financial burden on patients receiving care in low-income settings.

PHARMACOECONOMICS

Even though psychotropic drugs are becoming more available in sub-Saharan Africa, there are still several hurdles to cross before they are widely affordable to the majority of end users. Newer medications often take very long to be licensed for widespread availability and are very costly and therefore not cost-effective. In many parts of Africa, out-of-pocket payment is still the predominant mode of payment for services and medications, and even when these medications are covered by health insurance, newer, costly drugs are not stocked by most health facilities. There is also a high prevalence of counterfeiting. For example, in one report, more than 50% of all drugs on sale in Nigeria were either fake or substandard. Similar experiences have been reported in Gabon, Kenya, and Madagascar concerning anticonvulsants specifically. This had been attributed to drugs being manufactured in China, as well as poor storage in some cases, among other factors. In some settings, essential psychotropic medications are not available in public health facilities.

Cost-effectiveness studies of interventions for major psychiatric disorders are still very few in Africa. Recent studies have sought to compare total mental health care costs between task-shared care and routine care rather than focusing on the pharmacoeconomics of specific biologic therapies. In general, packages of intervention consisting of first-generation antipsychotics and TCAs have been reported to be more affordable in many African settings. In regard to what proportion of the total cost of receiving care might be due to the direct cost of medication, an early study in Nigeria by Suleiman et al. suggests that this could be about 53% in the case of schizophrenia. A more recent study in Nigeria found that the cost of medications constituted 51.2% of the direct cost and 21.9% of the total cost of Schizophrenia.

Further Readings

Adeponle AB, Obembe AO, Nnaji F, Adeyemi SO, Suleiman GT. Psychotropic drugs prescription at two regional psychiatric hospitals in Northern Nigeria. ***West Afr J Med***. 2008;27(2):106–110.

Adewuya AO, Ola BA, Mosaku SK, Fatoye FO, Eegunranti AB. Attitude towards antipsychotics among outpatients with schizophrenia in Nigeria. ***Acta Psychiatr Scand***. 2006;113(3):207–211.

Benson-Martin JJ, Milligan PD. A survey of the practice of electroconvulsive therapy in South Africa. ***J ECT***. 2015;31(4):253–257.

Bröcker E, van den Heuvel L, Seedat S. Accelerated theta-burst repetitive transcranial magnetic stimulation for depression in South Africa. ***S Afr J Psychiatr***. 2019;25(0):1346.

Burger JR, Van Der Westhuizen E, Lubbe MS, Serfontein JHP. Prescribing patterns of medicine classified as ‘antidepressants’ in South African children and adolescents. ***Health SA Gesondheid***. 2009;14(1):67–74.

Chisholm D, Gureje O, Saldivia S, et al. Schizophrenia treatment in the developing world: an interregional and multinational cost-effectiveness analysis. ***Bull World Health Organ***. 2008;86(7):542–551.

Dixon PA, Oforah E, Makanjuola R. Predictive indices for chlorpromazine therapy in schizophrenics. ***Br J Clin Pharmacol***. 1982;14(2):273–277.

Eaton J. Ensureing access to psychotropic medication in Sub-Saharan Africa. ***Afr J Psychiatry (Johannesbg)***. 2008;11(3):179–181.

Emsley R, Oosthuizen P, Koen L, Niehaus DJH, Medori R, Rabinowitz J. Oral versus injectable antipsychotic treatment in early psychosis: post hoc comparison of two studies. ***Clin Ther***. 2008;30(12):2378–2386.

Emsley RA, Oosthuizen PP, Joubert AF, Hawkridge SM, Stein DJ. Treatment of schizophrenia in low-income countries. ***Int J Neuropsychopharmacol***. 1999;2(4):321–325.

Fadare JO, Sunmonu TA, Bankole IA, Adekeye KA, Abubakar SA. Medication adherence and adverse effect profile of antiepileptic drugs in Nigerian patients with epilepsy. ***Neurodegener Dis Manag.*** 2018;8(1):25–36.

Fekadu A, Kebede D, Alem A, et al. Clinical outcome in bipolar disorder in a community-based follow-up study in Butajira, Ethiopia. ***Acta Psychiatr Scand.*** 2006;114(6):426–434.

Gureje O, Appiah-Poku J, Bello T, et al. Effect of collaborative care between traditional and faith healers and primary health-care workers on psychosis outcomes in Nigeria and Ghana (COSIMPO): a cluster randomised controlled trial. ***Lancet***. 2020;396(10251):612–622.

Gureje O, Chisholm D, Kola L, Lasebikan V, Saxena S. Cost-effectiveness of an essential mental health intervention package in Nigeria. ***World Psychiatry***. 2007;6(1):42–48.

Gureje O, Oladeji BD, Montgomery AA, et al. Effect of a stepped-care intervention delivered by lay health workers on major depressive disorder among primary care patients in Nigeria (STEPCARE): a cluster-randomised controlled trial. ***Lancet Glob Health***. 2019;7(7):e951–e960.

Gureje O, Oladeji BD, Montgomery AA, et al. High- versus low-intensity interventions for perinatal depression delivered by non-specialist primary maternal care providers in Nigeria: cluster randomised controlled trial (the EXPONATE trial). ***Br J Psychiatry***. 2019;215(3):528–535.

Ikeji OC, Ohaeri JU, Osahon RO, Agidee RO. Naturalistic comparative study of outcome and cognitive effects of unmodified electroconvulsive therapy in schizophrenia, mania and severe depression in Nigeria. ***East Afr Med J***. 1999;76(11):644–650.

Ioannidis JPA, Powe NR, Yancy C. Recalibrating the use of race in medical research. ***JAMA***. 2021;325(7):623–624.

Kumar S, Sudhakar S, Sajatovic M, Levin JB. Antipsychotic medication in sub-saharan Africa: a systematic literature review. ***J Clin Psychopharmacol***. 2020;40(6):541–552.

Ojagbemi A, Chiliza B, Bello T, et al. Neurological soft signs, spontaneous and treatment emergent extrapyramidal syndromes in black africans with first episode schizophrenia. ***Front Psychiatry***. 2018;9:172.

Oloniniyi IO, Akinsulore A, Aloba OO, Mapayi BM, Oginni OA, Makanjuola R. Economic cost of schizophrenia in a Nigerian teaching hospital. ***J Neurosci Rural Pract***. 2019;10(1):39–47.

Ononye F, Sijuola OA, Chukwuani CM, Mume OC, Makanjuola RO. Open comparative randomised study of moclobemide versus amitriptyline in major depressive illness (DSM IIIR) in Nigeria. ***West Afr J Med***. 2000;19(2):148–153.

Oyewumi LK, Kazaria SS. Electroconvulsive therapy in Nigeria: psychiatrists attitudes, knowledge and skills. ***West Afr J Med***. 1994;13(1):43–47.

Roopun KR, Tomita A, Paruk S. Attitude and preferences towards oral and long-acting injectable antipsychotics in patients with psychosis in KwaZulu-Natal, South Africa. ***S Afr J Psychiatr.*** 2020;26:1509.

Selis MA, Kauye F, Leentjens AFG. The practice of electroconvulsive therapy in Malawi. ***J ECT***. 2008;24(2):137–140.

Somoye EB, Onifade PO, Oluwaranti AO, Adeniji AA. A 10-year descriptive study of electroconvulsive therapy at the neuropsychiatric hospital, aro, abeokuta. ***J ECT***. 2014;30(4):315–319.

Suleiman TG, Ohaeri JU, Lawal RA, Haruna AY, Orija OB. Financial cost of treating outpatients with schizophrenia in Nigeria. ***Br J Psychiatry***. 1997;171:364–368.

Taylor RB, Shakoor O, Behrens RH, et al. Pharmacopoeial quality of drugs supplied by Nigerian pharmacies. ***Lancet***. 2001;357(9272):1933–1936.

Ukpong DI, Makanjuola ROA, Morakinyo O. A controlled trial of modified electroconvulsive therapy in schizophrenia in a Nigeria teaching hospital. ***West Afr J Med***. 2002;21(3):237–240.

33.19c Biologic Therapies in India

E. MOHANDAS, M.B.B.S., M.D., SANDEEP GROVER, M.B.B.S., M.D., AND AJIT AVASTHI, M.B.B.S., M.D.

India is a large country of 28 states and eight union territories with a population of 1.39 Billion (139 Crores). India has less than 10,000 psychiatrists (although 900 psychiatrists graduate every year), short of 27,000 to reach the desirable minimum of 3 psychiatrists per 100,000 population. According to the recent National Mental Health Survey (NMHS), the lifetime prevalence of mental disorders is 13.7%, and the current prevalence is 10.6%. India ranks 3rd worldwide for pharmaceutical production by volume and 14th by value. The domestic pharmaceutical industry has a strong network of 3,000 drug companies and about 10,500 manufacturing units. Newer antipsychotics like lumateperone and brexiprazole are still not available in India. Almost all psychotropics are available at affordable rates. Long-acting antipsychotic preparations, including risperidone, olanzapine, and paliperidone, are available with the expectation of aripiprazole and once-in-3 months paliperidone reaching the market soon. The use of agomelatine, lurasidone, vilazodone, vortioxetine, and levomilnacipran is minimal due to the cost factor or lack of expected efficacy. Neuroscience-based nomenclature (NbN), although initiated in some academic sessions, needs to be disseminated further for a better understanding of prescription principles.

Considering somatic treatments, electroconvulsive therapy (ECT) is widely used in India, and over the last decade, many centers have provided other brain stimulation techniques, such as repetitive transcranial brain stimulation (rTMS), transcranial direct stimulation (tDCS), and deep brain stimulation.

Despite the availability of various treatment modalities at an affordable price, there is a wide treatment gap (70% to 92%) for various mental disorders. The cost of illness studies suggests that the cost of illness of disorders like schizophrenia and bipolar disorder is significantly lower in India when compared to the data from Western countries, and the cost of treatment of schizophrenia is comparable to the cost of illness of diabetes mellitus. Treatment guidelines by any professional organization for the management of various psychiatric disorders were nonexistent two decades back. However, the Indian Psychiatric Society has initiated publishing management guidelines on various psychiatric disorders since 2005. Multicentric studies on prescription patterns suggest that among the various antidepressants, escitalopram and sertraline are the most commonly prescribed antidepressants. The multicentric study that evaluated the prescription patterns found that olanzapine was the most commonly prescribed antipsychotic medication, followed by risperidone. On the other hand, a survey of psychiatrists suggested that risperidone, olanzapine, and haloperidol were the three most commonly prescribed antipsychotics. Clozapine is, in general, underutilized due to the fear of side effects and the need for monitoring. The cross-national studies, however, suggest that prescription rates of clozapine are comparable to other Asian countries. Long-acting injectable (LAI) antipsychotics are also used less often, with prescription rates for the same during the acute phase of illness being 9.3% and 18.4% during the stable/stabilization phase of schizophrenia. First-generation LAIs are used more often than second-generation antipsychotic LAIs. The LAIs are mostly used in patients who have a history of medication and treatment nonadherence. The cross-national studies, which involved centers from India, suggest that prescription rates of LAI antipsychotics in India are much lower than in countries like Malaysia, Singapore, Sri Lanka, and Pakistan.

The studies evaluating prescription patterns of patients with bipolar disorder seen for the first time suggest that valproate is more often used when compared to lithium, whereas studies that have included clinically stable patients of long duration of illness suggest a small difference in the prescription pattern for valproate and lithium. Across the various disorders, there is a high prescription rate for the prescription of benzodiazepines during the acute phase of illness. Available data also suggest low prescription rates for polypharmacy (i.e., prescription of more than one agent from the same class of medications).

With regard to the therapeutic drug monitoring (TDM) facilities, serum lithium level estimation is most widely available. Estimation of serum valproate and carbamazepine, although available across the country, is not performed frequently as estimation of serum lithium levels. Serum clozapine level estimation facilities are limited, and estimation of other antipsychotic levels is nonexistent.

Till recently, both modified and unmodified ECT was used for patients with severe mental disorders in India. However, the Mental Health Care Act of 2017 banned the use of unmodified ECT in the country. A survey published in 2005 showed that ECT services were available in a majority of the teaching institutions, with brief pulse being used in more than half of the institutes and bilateral ECT being used in almost all the institutes. Available data on indications for ECT revealed that the most common diagnostic group was schizophrenia (36.5%), followed by depression (33.5%) and mania (17.9%). Adult patients in the age group of 45 to 64 years formed the largest proportion of patients receiving ECT in contrast to the older adult group (14.7%, aged ≥65 years). Researchers from India have evaluated the various techniques to improve the safety of the use of ECT in patients with mental disorders. Emerging data from India also suggest that combining clozapine with ECT may be beneficial in the management of patients with clozapine-resistant schizophrenia.

In the last decade, neurostimulations like rTMS and tDCS have been added to the therapeutic armamentarium in selected centers. The preliminary studies cite the safety and efficacy of rTMS in the management of negative symptoms in patients with schizophrenia. Indian studies have found the usefulness of tDCS in the management of treatment-resistant obsessive-compulsive disorder and refractory auditory verbal hallucinations in schizophrenia.

Overall, the biologic therapies available for the management of various psychiatric disorders in India are at par with most developed countries, and the overall cost of treatment of mental disorders is much lower than in developed countries. Currently there are no registries for proper guidance about the effect of psychotropics in special population like children, elderly, pregnancy and postpartum. In India, the private sector contributes to a significant proportion of mental health services. However, limited data is available in terms of the use of psychotropics and various somatic treatments in the private sector.

FURTHER READINGS

Bose A, Shivakumar V, Agarwal SM, et al. Efficacy of fronto-temporal transcranial direct current stimulation for refractory auditory verbal hallucinations in schizophrenia: a randomized, double-blind, sham-controlled study. ***Schizophr Res***. 2018;195:475–480.

Chanpattana W, Kramer BA, Kunigiri G, Gangadhar BN, Kitphati R, Andrade C. A survey of the practice of electroconvulsive therapy in Asia. ***J ECT***. 2010;26(1):5–10.

Chanpattana W, Kunigiri G, Kramer BA, Gangadhar BN. Survey of the practice of electroconvulsive therapy in teaching hospitals in India. ***J ECT***. 2005;21(2):100–104.

Girish K, Mayur PM, Saravanan ES, et al. Seizure threshold estimation by formula method: a prospective study in unilateral ECT. ***J ECT***. 2000;16(3):258–262.

Grover S, Avasthi A. Anti-psychotic prescription pattern: a preliminary survey of Psychiatrists in India. ***Indian J Psychiatry***. 2010;52(3):257–259.

Grover S, Avasthi A, Chakravarty R, et al. Prescription patterns in clinically stable patients with bipolar disorder: Findings from the Bipolar Disorder Course and Outcome from India (BiD-CoIN) study. ***Asian J Psychiatr***. 2021;57:102549.

Grover S, Avasthi A, Chakrabarti S, Bhansali A, Kulhara P. Cost of care of schizophrenia: a study of Indian out-patient attenders. ***Acta Psychiatr Scand***. 2005;112(1):54–63.

Grover S, Avasthi A, Sinha V, et al. Indian psychiatric society multicentric study: prescription patterns of psychotropics in India. ***Indian J Psychiatry***. 2014;56(3):253–264.

Grover S, Balachander S, Chakarabarti S, Avasthi A. Prescription practices and attitude of psychiatrists towards clozapine: a survey of psychiatrists from India. ***Asian J Psychiatr***. 2015;18:57–65.

Grover S, Sahoo S, Mehra A. Perceptions of psychiatrists toward the use of long-acting injectable antipsychotics: an online survey study from India. ***J Clin Psychopharmacol***. 2019;39(6):611–619.

Grover S, Shouan A, Chakrabarti S, Sahoo S, Mehra A. Comparative study of effectiveness of augmentation with ECT in clozapine resistant schizophrenia (CRS) and non-clozapine resistant schizophrenia (Non-CRS). ***Nord J Psychiatry***. 2021;75(6):397–405.

Gururaj G, Varghese M, Benegal V, et al. ***National Mental Health Survey of India, 2015–16: Prevalence, patterns and outcomes***. National Institute of Mental Health and Neuro Sciences; 2016, NIMHANS Publication No. 129.

Hiremani RM, Thirthalli J, Tharayil BS, Gangadhar BN. Double-blind randomized controlled study comparing short-term efficacy of bifrontal and bitemporal electroconvulsive therapy in acute mania. ***Bipolar Disord***. 2008;10(6):701–707.

Kumar N, Vishnubhatla S, Wadhawan AN, Minhas S, Gupta P. A randomized, double blind, sham-controlled trial of repetitive transcranial magnetic stimulation (rTMS) in the treatment of negative symptoms in schizophrenia. ***Brain Stimul***. 2020;13(3):840–849.

Mental Healthcare Act, 2017. ***Gazette Notification***. 7th April, Government of India; 2017.

Somaiya M, Grover S, Chakrabarti S, Avasthi A. Comparative study of cost of care of outpatients with bipolar disorder and schizophrenia. ***Asian J Psychiatr***. 2014;12:125–133.

Tang CT, Chua EC, Chew QH, et al. Patterns of long acting injectable antipsychotic use and associated clinical factors in schizophrenia among 15 Asian countries and region. ***Asia Pac Psychiatry***. 2020;12(4):e12393.

Thamby A, Seshachala K, Sharma L, et al. Transcranial direct current stimulation for treatment-resistant obsessive-compulsive disorder-a large case series. ***Asian J Psychiatr***. 2021;60:102625.

Thimmaiah R, Thirthalli J, Ramesh VJ, et al. Effect of a course of electroconvulsive therapy on interictal bispectral index values: a prospective study. ***J ECT***. 2012;28(1):20–23.

Xu SW, Dong M, Zhang Q, et al. Clozapine prescription pattern in patients with schizophrenia in Asia: the REAP survey (2016). ***Psychiatry Res***. 2020;287:112271.

33.19d Biologic Therapies in Latin America

RODOLFO FURLAN DAMIANO, M.D., PH.D., AND ANDRÉ RUSSOWSKY BRUNONI, M.D., PH.D.

Even though there are several well-known Latin American researchers both currently and in the history of medicine, the economic status of the continent impairs a greater development of the field. Furthermore, limited resources impact the development and availability of new drugs to the whole continent, being both unavailable or unaffordable to the majority of the population. This heterogeneity in the medication market throughout the globe justifies chapters outlining the biologic therapies available in each continent. Moreover, due to the markedly cultural and economic differences among countries in Latin America, as well as in drug regulation policies within some countries, we chose to focus, in this chapter, on biologic therapies available in Brazil (due to the citizenship of the authors) in order to serve as a model to other Latin American countries, despite knowing that this pretentious task is almost unreachable.

PREVALENCE AND GAP IN TREATMENT OF MENTAL DISORDERS

The 12-month prevalence of any mental disorder in Latin America is 17%, varying from 7.2% in Guatemala to 29.6% in Brazil, which is higher than in the United States (26.2%). Severe mental disorders follow the same pattern, with a 12-month mean prevalence of 4.9%, varying from 2% in Guatemala to 10% in Brazil. Suicide follows a different pattern, with Guyana (26.07/100.000), Suriname (24.78/100.000), Uruguay (16.82/100.000), Argentina (10.61/100.000), Chile (10.14/100.000), Ecuador (9.21/100.000), Venezuela (8.68/100.000), and Bolivia (6.41/100.000) having an incidence greater than Brazil (6.09/100.000), despite the large number of mental health disorders found in Brazil.

The treatment gap is another important issue that might also explain this contrast between mental disorders versus suicide among countries. The weighted mean for treatment gap in Latin America is 79.0% for any mental disorder, with a smaller gap found in Chile (61.5%) and a larger gap in Colombia (86.1%). For severe mental disorders, the treatment gap is a little smaller (69.9%), with Guatemala representing a large number of untreated individuals (86.9%). Due to the social inequality in Latin America, special attention should be given to minorities (e.g., indigenous) that represent important vulnerable populations usually neglected. Also, several countries do not have reliable data, representing an urgent area for research and public policies.

PARTICULARITIES IN DRUG TREATMENT AND DRUG RESPONSE

Brazil and Latin America represent a “world in a continent” due to the remarkable cultural differences. It might impact several different areas, such as pharmacotherapy, patient care, and ethics, for example. Noteworthy, pharmacogenetic polymorphisms represent one important and usually forgotten therapeutic challenge in Latin America, impacting drug response and side effects. Pharmacogenetic studies should be conducted in Brazil, and the generability of North American and European studies should be viewed with caution.

Another particularity of drug treatment in Latin America is the presence of diverse religious backgrounds and medicinal plants. Religious affiliation might impact the way the biologic treatment is seen and conducted (e.g., faith healings; electroconvulsive therapy; ethics), sometimes influencing the treatment choice. Moreover, the diversity of medicinal plants also represents an important feature, as many patients use it without prescription, having important (positive and negative) short- and long-term mental health consequences that deserve closer attention from researchers.

OTHER BIOLOGIC THERAPIES

Electroconvulsive Therapy and Magnetic Seizure Therapy

Electroconvulsive therapy (ECT) has been used in Brazil since the 1930s, with the first publication of clinical results by the founder of the Institute of Psychiatry—Clinics Hospital, Prof. Pacheco Silva, following an increased use in academic institutions and asylums and less frequently after the introduction of antipsychotics in the 1970s. ECT in Brazil is restricted to a few current university services (14 academic centers), leaving critically ill patients without access to this important therapeutic resource. There is an important lack of public assistance for ECT, which affects mainly poor and severely ill patients, resulting in overcrowded services handled by brave colleagues offering their best skills even with poor resources and sometimes enduring unfair criticism on their important work. In early 2021, the first magnetic seizure therapy (MST) device was introduced in São Paulo, Brazil, at the Institute of Psychiatry, being solely available for research purposes.

Repetitive Transcranial Magnetic Stimulation

Transcranial magnetic stimulation (rTMS) was approved by ANVISA in 2010, just after the U.S. FDA approval, and by the National Medical Board (CFM) in 2012, making Brazil one of the first countries in the world to have rTMS introduced as a clinical (not experimental) treatment. Currently, MagVenture (Denmark), Neurosoft (Russia), and Magstim (U.K.) are ANVISA cleared and are used in almost all clinical and research centers in Brazil. In fact, there are more than 100 private practice centers that perform rTMS in many locations in Brazil. Unfortunately, rTMS is covered by neither health insurance nor the Brazilian Universal Health System (SUS), which is a huge limitation for access.

Transcranial Direct Current Stimulation (tDCS)

Several transcranial direct current stimulation (tDCS) devices have been ANVISA approved between 2017 and 2021 in Brazil, such as the Depression LTE system (SoterixMedical, USA), DC-Stimulator Plus (Neuroconn, Germany), and Flow (FlowNeuroscience, Sweden). tDCS might represent a nonexpensive neuromodulatory strategy to be used both by patients and providers in a low/middle-income country (LMIC) scenario such as Brazil; therefore, there are more than 200 private practice centers that offer tDCS services. However, here the issues of off-label use and low level of evidence are even more of concern, which is summed by an unclear regulatory status by the Medical Board that has not yet legislated on the indications and contraindications of tDCS in medical practice, hindering all tDCS applications essentially off-label and decided on an individual medical basis.

FURTHER READINGS

Brandão MGL, Acúrcio FA, Montemor RLM, Marlière LDP. Complementary/alternative medicine in Latin America: use of herbal remedies among a Brazilian metropolitan area population. ***J Complement Integr Med***. 2006;3(1).

Ciocca DR, Delgado G. The reality of scientific research in Latin America; an insider's perspective. ***Cell Stress Chaperones***. 2017;22(6):847–852.

Homedes N, Ugalde A. Availability and affordability of new medicines in Latin American countries where pivotal clinical trials were conducted. ***Bull World Health Organ***. 2015;93(10):674–683.

Kohn R, Ali AA, Puac-Polanco V, et al. Mental health in the Americas: an overview of the treatment gap. ***Rev Panam Salud Publica***. 2018;42:e165.

Ribeiro RB, Melzer-Ribeiro DL, Rigonatti SP, Cordeiro Q. Electroconvulsive therapy in Brazil after the "psychiatric reform": a public health problem–example from a university service. ***J ECT***. 2012;28(3):170–173.

Silva ACPE. O eletrochoque no tratamento das doenças mentais. ***Revista de Medicina***. 1941;25(95-96):15–28.

Suarez-Kurtz G. Pharmacogenetics in the Brazilian population. ***Front Pharmacol***. 2010;1:118.

33.20 Brain Stimulation

33.20a Electroconvulsive Therapy

Joan Prudic, M.D.

Electroconvulsive therapy (ECT), also known as electroshock therapy, is a standard psychiatric treatment in which electricity is delivered through the scalp to induce a therapeutic seizure. First conducted in 1938 by two Italian physicians, Ugo Cerletti and Lucio Bini, on patients with schizophrenia, ECT has become the oldest somatic treatment for many psychiatric conditions, including depression, catatonia, as well as bipolar disorder and schizophrenia. ECT is generally reserved for patients who have failed to respond to pharmacotherapy or have adverse reactions to standard pharmacotherapy, as well as those with severe symptoms such as suicidality, psychosis, or grave functional impairment, resulting in a need for a rapid and definitive response. It is estimated that about 5–8/100,000 residents in the United States receive ECT annually. Contrary to the sometime negative depiction of ECT, its safety and efficacy has been much improved with modification of various aspects of the technique such as optimization of the stimulus energy, electrode placement, the use and optimization of general anesthesia, as well as improvement in cardiopulmonary management. The application of ECT is regulated by the FDA, and there are regulatory and professional guidelines for proper consent procedures including situations where the patient lacks capacity. The mechanism of action of ECT remains an active area of hypothesis and research, although a great deal of information has been obtained with the availability of newer research tools.

HISTORICAL OVERVIEW

Convulsive therapies for major psychiatric illnesses predate the modern therapeutic era with the use of camphor reported as early as the 16th century, and several accounts of camphor convulsive therapies from the late 1700s to the mid-1800s. With the success of malarial fever therapy for general paresis, interest in biologic therapies for psychiatric illnesses increased in the early 20th century. By the 1930s, insulin coma and psychosurgery had also been introduced.

Unaware of the history of camphor convulsive therapy, Meduna made the observation that the brains of epileptics had greater than normal numbers of glial cells, while those of schizophrenics had fewer, and hypothesized that there might be a biologic antagonism between convulsions and schizophrenia. Following animal experimentation, camphor was (again) selected as the appropriate agent to use for the therapeutic induction of seizures. In 1934, the first catatonic psychotic patient was successfully treated using intramuscular injections of camphor in oil to produce therapeutic seizures. Thirteen of a series of 26 patients treated similarly were reported to have had some degree of remission of psychosis. In 1937, a conference on convulsive therapies and insulin coma was held in Switzerland and the use of convulsive treatments spread.

Although effective, camphor was relatively long acting and was soon replaced by pentylenetetrazol. This agent was also associated with problems, principally patient rejection of the noxious sensations associated with GABA antagonism while conscious, for example, increased heart rate, feelings of terror, and myoclonus preictally.

Bini and Cerletti were interested in the use of electricity to induce seizures, and, after a series of animal experiments and observation of the use of electricity commercially, were able to safely apply current across the heads of animals for this purpose. In 1938, the first electroconvulsive treatment (ECT) course was administered to a delusional and incoherent patient, who improved with one treatment and remitted after 11 ECT. Electrical induction of convulsive therapy could be made more reliable and shorter acting than chemically induced convulsive therapies, and, by the early 1940s, had replaced them. In 1940, the first use of ECT occurred in the United States.

Throughout the 1940s, ECT was a first-line therapy for a variety of major psychiatric illnesses. In the ensuing decades, efforts to improve its acceptability and safety profile continued. In 1940, A. E. Bennett developed curare for use in the prevention of ECT-associated fractures, which had previously resulted from the severe muscle contractions associated with the tonic phase of the seizure induction. By the 1950s, brief general anesthesia with succinylcholine replaced curare. Other improvements in anesthesia, most notably oxygenation throughout the treatment session and cardiac and oxygen saturation monitoring became standard. Efficacy in disorders other than schizophrenia was demonstrated, most notably major depression.

In an effort to reduce the retrograde memory problems that persisted for some patients after the initial recovery period post-ECT, explorations of nondominant electrode placement and alternative, more efficient, waveforms were undertaken in subsequent decades. The practice of ECT also benefited from the introduction of controlled trials methodology, which demonstrated its safety and efficacy, and from refinements made in diagnostic systems and the process of informed consent. In the 1980s and 1990s, efforts to assure uniformly high standards of practice were underway with the publication of recommendations for treatment delivery, education, and training by professional organizations in the United States, England, Scandinavia, and Canada among others.

With the current widespread use of pharmacologic agents as first-line treatments for major psychiatric disorders, ECT is now more commonly used for patients with resistance to those treatments, except in the case of life-threatening illness due to inanition, severe suicidal symptoms, or catatonia. While the failure of subconvulsive stimulation during ECT to induce remission of psychiatric illness and the effectiveness of chemical convulsive therapy suggested that the seizure was necessary and sufficient for therapeutic benefit with ECT, it is now known that there is a dose–response relationship with right unilateral ECT and that bifrontotemporal ECT may be ineffective with ultrabrief pulse widths. Work continues to explore the underlying mechanisms and biologic characteristics of effective ECT treatments with a focus on making the treatment focus on appropriate neural networks with a more efficient stimulus as a method of reducing cognitive side effects. With the growing understanding that depression is a chronic disease for many patients, more emphasis has been placed on continuation and maintenance treatments following an acute course of ECT. Work has been published demonstrating the efficacy of both psychotropics and ECT in reducing relapse post-ECT.

The utilization of ECT has diminished since the middle of the last century. ECT is relatively safe and remains the most effective treatment for major depression, and it is a rapidly effective treatment for life-threatening psychiatric conditions such as catatonia. Hence, ECT has remained in the active treatment portfolio of modern therapeutics for over 80 years. Its use has shifted from public to private institutions, and it is estimated that approximately 5–8/100,000 patients have received ECT annually over the last few decades in the United States (Table 33.20a–1).

ELECTRICAL INDUCTION OF SEIZURES

Factors in the Generation and Measurement of Electrical Stimulation for ECT

ECT is a form of brain stimulation in which induction of a seizure occurs when an applied transcranial electrical stimulus, consisting of a train of rectangular pulses that alternate in direction, creates an electrical field with a flow of current in excitable brain tissue sufficient to depolarize cell membranes of neurons synchronously. Fundamentally, the generation and behavior of the electrical stimulus have been conceptualized by Ohm's law:

$$V(\text{voltage}) = I(\text{current}) \times R(\text{resistance})$$

Voltage is the electromotor force (volts) causing current to flow; current is the rate of flow of electrons (amperes); and resistance is the relative ability to pass current (ohms). Ohm's law assumes a direct current circuit, but ECT devices function on alternating current. In this environment, it is necessary to replace resistance with the term impedance (Z). Impedance involves capacitance and inductance as well as resistance. Since biologic tissues, being nonferrous, have low inductive properties, the inductance term is usually assumed to have little relevance to impedance during ECT. Capacitance (F) is the ability to store charge (farads). Neural tissue has relatively low capacitance (10^{-12} to 10^{-10} F), and the electrode skin interface (estimated 0.4 μF), is the source of relatively more of this term.

Although it is possible to trigger a seizure with a single DC pulse of long duration, voltage and current may vary over time in the electrical stimulus as clinically used. A variety of stimulus waveforms have also been used to deliver the stimulus, including sine wave, rectified partial sine wave, brief pulse square wave, and ultrabrief pulse square waves. These different waveforms are not of equal efficiency in depolarizing neural tissue. By examining mammalian neurons, it has become evident that fast-rising current results in firing at a lower current intensity, and hence, lower dosage. The slow rise of sine wave results in accommodation and increases in the threshold for neuronal firing, estimated to be threefold. This observation has led to a preference for brief pulse waveforms that "instantly" achieve peak intensity (on the order of microseconds) over the relatively slower rising sine waveform, typically reaching peak intensity in several milliseconds.

Chronaxie, optimal current duration to depolarize human neuronal tissue is on the order of 0.1 to 0.2 ms. Furthermore, once nerves have fired, they enter a refractory period, and continued stimulation will not elicit further response. This has led to a reexamination of the need for longer duration stimulation per phase where much of the stimulus is delivered when the neuronal tissue is refractory. Even brief pulse waveforms with pulse widths on the order of 0.5 to 1.5 ms, commonly used in modern devices, may be expected to deliver excessive energy in the refractory period, although it is an order magnitude lower than with sine wave stimulation. Ultrabrief pulse waveforms, 0.25 to 0.5 ms available on modern devices, most closely approximate the chronaxie of human neuronal tissue. Finally, it has been observed that increasing the frequency of pulses is less efficient in eliciting depolarization than increasing train duration of pulses when using a brief pulse device. This phenomenon is probably also related to the complex relationship of frequency to depolarization, a U-shaped curve in primates, resulting in maximal efficiency in depolarization around 20 Hz. Duration of a pulse train has a more linear covariation with depolarization.

Table 33.20a–1.
Milestones in the History of Convulsive Therapy

1500s	Paracelsus induces seizures by administering camphor (po) to treat psychiatric illness.
1785	First published report of the use of seizure induction to treat mania, again using camphor.
1934	Von Meduna begins the modern era of convulsive therapy using intramuscular injection of camphor for catatonic schizophrenia. Camphor is soon replaced with pentylenetetrazol.
1938	Cerletti and Bini, conduct the first electrical induction of a series of seizures in a catatonic patient and produce a successful treatment response.
1940	ECT is introduced to the United States.
	A. E. Bennett develops curare for use as a muscle relaxant at ECT.
1951	Introduction of succinylcholine.
1958	First controlled study of unilateral ECT.
1960	Ottosson demonstrates that attenuation of seizure expression with an anticonvulsant agent (lidocaine) reduces the efficacy of ECT. In concert with other work suggesting that subconvulsive treatment produces only weak clinical responses, the hypothesis that seizure activity is both necessary and sufficient for efficacy is upheld.
1960s	Randomized clinical trials of the efficacy of ECT vs. medications in the treatment of depression yield response rates that are significantly higher with ECT.
	Comparisons of neuroleptics and ECT show that neuroleptic medication is superior for acute treatment, although ECT may be more effective long term.
1970	D’Elia develops what has become the most common electrode positioning for right unilateral ECT.
1976	A constant current, brief pulse ECT device, the prototype for modern devices, is developed.
1978	The American Psychiatric Association publishes the first

	Task Force Report on ECT with the aim of establishing standards for consent and the technical and clinical aspects of the conduct of ECT.
Late 1970s to Early 1980s	Randomized controlled trials demonstrate ECT is more effective than sham treatment for major depression.
1985	The National Institutes of Health/National Institute of Mental Health Consensus Conference on ECT endorses a role for the use of ECT and advocates research and national standards of practice.
1987	The belief that the seizure in itself is sufficient for clinical response is challenged by a report that the combination of dosage just above seizure threshold and right unilateral electrode placement, while producing a seizure of sufficient duration, is ineffective.
1988	Randomized controlled clinical trials of ECT vs. lithium demonstrate them to be equally effective in mania.
2000	In controlled trials, the dose/response relationship for right unilateral ECT is validated; high-dose right unilateral and bilateral ECT show equal response rates in major depression, but right unilateral electrode placement is associated with fewer adverse cognitive effects.
	Convulsive treatment is induced with magnetic stimulation by Lisanby and colleagues.
2001	The largest modern controlled trial of relapse prevention post-ECT with continuation pharmacotherapy demonstrates a significantly better outcome for combined treatment with a tricyclic antidepressant (nortriptyline) plus lithium compared to nortriptyline alone or placebo during the first 6 months post-ECT.
2006	In a large controlled trial, continuation ECT, given in a fixed schedule without concurrent psychotropic medications, is demonstrated to be effective, but a fixed schedule is only equivalent to combination lithium/nortriptyline.

2008	Ultrabrief right unilateral ECT is shown to be equivalent to brief pulse right unilateral and bilateral ECT in efficacy, but difficult to distinguish from a healthy never depressed control group in cognitive performance, including retrograde memory. In the same controlled trial, ultrabrief bilateral ECT at 150% above seizure threshold was ineffective, confirming earlier observations that the seizure itself is not sufficient for efficacy.
2013	Feasibility study of Focal Electrically Assisted Seizure Therapy published.

Once the electrical stimulus is generated, current is distributed inversely to the resistance of the tissue compartments traversed. There are three major compartments: scalp (about 200 Ω); skull (about 18,000 Ω); and brain tissue (about 200 Ω). Because current flow is greatest across the lowest resistance, it has been estimated that 80% of current is shunted through the scalp. The skull obviously provides the greatest component of resistance, and skull anatomy, for example, the position and size of sutures, skull thickness, etc., greatly influences the proportion of remaining electrical dosage that enters brain tissue. Differences in scalp and skull anatomy are thought to be a major source of the wide variation among individuals in the amount of electrical dosage required to elicit a generalized seizure or seizure threshold. These differences have been observed to be of the order of 50-fold. It has been theorized and modeled that interindividual anatomic variation in tissue layer thickness and conductivity could be compensated by varying the amplitude of the stimulus to produce a more focal and uniform electric field exposure, but the degree of adjustment required has not been determined thus far. CSF volume also affects the distribution of the electrical stimulus, with higher volume of fluid being associated with a weaker, more focal distribution and field strength.

Another influence on current shunting is interelectrode distance: The closer electrodes are placed on the skull, the more shunting occurs. It is thought that right unilateral electrode placements can result in the creation of a virtual electrode occurring in the area between the two electrodes and, hence, in reduced amounts of electrical dosage required for seizure elicitation. This increased shunting may also result in lower current density in brain tissue.

The quantity of the electric stimulus delivered has been expressed as energy (joules) traditionally, and, in the modern era, as charge (millicoulombs). Substantial evidence indicates that current density in neural tissue is the critical factor responsible for seizure production and the subsequent neurobiologic effects of the electrical stimulation. Although it is not presently possible to measure charge density directly in human neural tissue during ECT, dynamic impedance can be used to approximate charge density because of its relationship to current shunting during stimulus delivery. When electrode site preparation and positioning are standardized, anatomic factors become the major determinants of impedance. If these factors are also constant during the treatment, impedance can then reflect current density in brain tissue.

With the development of other forms of electrical and also magnetic brain stimulation, there has been revived interest in understanding and exploring the impact of electrical stimulation on brain tissue. The various elements of brain have differing electrical properties and vary markedly in excitability with myelinated axons being most excitable, and neuronal cell bodies and dendrites least. Tissue response also varies with distance from electrical source, in this case, the stimulating electrode. Although direct recording of the effects of electrical stimulation is not available for ECT, except in animal models, functional neuroimaging and computational modeling are expected to expand the current understanding of the physiology of the treatment.

Devices Used to Deliver ECT

Although constant current, constant voltage, and constant energy devices have been in use, modern devices are constant current with voltage adjusted by the device to maintain the desired current level. The range of constant current is in the order of 0.5 to 0.9 A in these devices. Dosage is manipulated by varying exposure to a fixed current (total charge). With brief pulse devices, the most commonly used constant current devices, frequency of pulses, pulse width, and duration of pulse train are parameters that can be used to vary the stimulus dosage. Because impedance varies directly with voltage, particular attention must be paid to impedance in a constant current environment. Commonly, impedance and the integrity of the electrical circuit are tested prior to stimulus delivery by passage of a low-amplitude, high-frequency test stimulus. If impedance exceeds manufacturer limits, treatment stimuli are not permitted by the device, although some devices are equipped with override features. If too high, impedance can be lowered by improving skin–electrode contact, and other parts of the circuit, for example, the cable, contacts with the electrode and device, may also need examination and adjustment. By keeping impedance within safe guidelines, local tissue damage to skin is avoided. When impedance is too low, voltage will also be very low, and may not be sufficient to drive enough current flow to brain tissue where resulting charge density may not be great enough to cause synchronous neuronal depolarization and the desired biologic effects. Low impedance is usually the result of current shunting in the scalp and is corrected by interrupting any alternative current pathway at the skin surface. For example, gel used to prepare the electrode site, perspiration, or hair products, can run together and form a continuous current pathway on the scalp, and can be corrected by removal.

Calculation of stimulus energy in joules requires information about interelectrode impedance, which is not readily available. Charge is commonly used to measure the intensity of the stimulus used in treatment with brief pulse devices. Charge is the quantity of electrons flowing through a conductor over a given time period (coulombs) and can easily be computed from the parameters of a brief pulse constant current device.

Constant voltage devices allow current to vary inversely with impedance. The typical device generates a sine wave stimulus and has a range of 70 to 170 V. With these devices, high impedance will result in low current flow, which in turn, theoretically, may be inadequate to induce sufficient charge in neuronal tissue to result in neuronal depolarization. Because the practitioner does not have access to crucial information about impedance when using a constant voltage device, and it may be difficult to assess current intensity in this setting, and judgments about the adequacy of parameters can be difficult to make. The use of these devices in the United States has become outmoded.

ECT device output in the United States has been capped at 100 J, which corresponds to a charge of 500 to 576 mC (assuming 220 Ω impedance), depending on the particular device specifications. With ultrabrief pulse width, this range is generally adequate to deliver right unilateral ECT to almost all patients. Output limitations can be addressed by using double stimulation, which provides for summation. ECT devices in the European Union and elsewhere can have an extended range up to 200 J to allow delivery of adequate dosage with brief pulse right unilateral electrode placement, which is mandated in some E.U. nations.

MECHANISMS OF ACTION

The relative inaccessibility of the central nervous system (CNS) has made understanding the pathophysiology of psychiatric diseases a continually evolving process. As a consequence, the mechanisms of action of ECT, the oldest somatic treatment currently in use in the field of psychiatry, continue to be an active area of hypothesis and research. Past and present, theories abound, and a wide range of data, particularly at the preclinical level, is available on the effects of electroconvulsive stimulation (ECS).

Proposed mechanisms have involved theories in almost every domain. These include psychological constructs; structural explanations, for example, neural plasticity; interpretations based on electrophysiology and neurophysiology; biochemical theories encompassing neurochemistry, including neuroendocrine phenomena; and other molecular rationales, including genetic and epigenetic mechanisms.

Among psychological theories, there is no evidence that fear, regression, or increased medical attention account for the efficacy or side effects of ECT. As many of the most recent advances in the field have reduced the adverse cognitive effects of ECT while maintaining efficacy, there is little reason to believe that memory loss has much mechanistic validity. Extensive reviews of the preclinical and clinical literature have refuted the assertion that brain damage occurs, let alone, accounts for ECT’s effects. Furthermore, current anesthesia practice assures that hypoxia does not occur. Sham ECT (anesthesia alone) is ineffective; hence, anesthesia is unlikely to play a role in mechanism of action.

Distinguishing the relevance of observed phenomena with ECS constitutes a major challenge in the area of mechanisms. Considerations include questions about the relevance of animal data, much of which has been derived from rodents, although there has been recent interest in primate work. This work suffers from the current lack of adequate animal models of the psychiatric syndromes for which ECT is utilized. In preclinical work, ECS is often given in a time course that does not closely follow the treatment course given to patients, that is, two to three times weekly. For example, stimulation is administered once or several times in 1 day in preclinical investigations.

ECT has a very broad spectrum of efficacy and action, as will become evident from discussions of indications for its use, of efficacy in a variety of syndromes, and of its side effects. The treatment is often studied in comparison with antidepressants or to a lesser degree other psychotropics. It is known that there is overlap in mechanistic effects, but ECT has more therapeutic effects than any studied medication. Coupled with the fact that ECT has more indications for treatment, the challenge of ascertaining critical similarities and differences is significant. Because of the perceived invasiveness of the treatment, the severity of illness among treated patients, and the necessity for anesthesia, little work is currently being done with sham ECT. There are limited data available from weaker forms of the treatment, for example, right unilateral ECT at low or modest dosage which is now known to be a less effective form of the treatment.

Biochemical Actions

There is a substantial body of work comparing ECS with psychotropic (and other) medications whose actions and mechanisms have been assayed in vitro and in vivo, most often preclinically, but also in human studies. Not surprisingly, the major areas of interest have been neurotransmitter and peptidergic systems. A primary focus has been on monoamine effects, many of which are shared by ECS in animals, ECT in humans, and psychotropics, especially antidepressants. Although it appears that both ECS and antidepressants downregulate β-adrenergic receptors, human studies of broader changes in noradrenergic pathways have been characterized by inconsistent outcomes. There are also variable data on the effects of antidepressants and ECT on serotonergic systems: Antidepressants typically downregulate 5-hydroxytryptamine receptors ($5HT_2$), whereas ECT leads to increased density of $5HT_2$ receptors (in rodents) and decreased binding (in primates). Although serotonergic transmission may be enhanced by both antidepressants and ECT, the mechanisms appear distinct. ECT is more effective than antidepressant medications, so defining what possible mechanisms are responsible for this phenomenon is of pharmaceutical as well as heuristic interest. Currently, the critical differences are not known. Both animal and clinical studies indicate increased dopaminergic functioning following ECS, particularly D1 and D3 receptor functions. Although this effect may be an underlying mechanism for ECT's therapeutic effects in Parkinson disease, it does not help clarify the underlying process for the benefits conferred by ECT on patients with schizophrenia and other psychotic illnesses, where dopaminergic antagonism, at least in some brain regions, has been presumed to be critical.

Another major area of interest has been the variety of different neurotransmitter systems, which may form the biochemical foundation for the anticonvulsant effects of ECS. Because of the efficacy of anticonvulsant medications in mania, there is also speculation that the anticonvulsant effects of ECT contribute to its efficacy in mania. Most of this work is preclinical, and indicates mixed GABAergic activity, upregulation of adenosine receptors, and blocking of anticonvulsant effects with opiate antagonists. Currently, GABAergic effects in humans are being explored neurophysiologically. It has been difficult to duplicate anticonvulsant effects of opiate antagonists in humans. The proconvulsant effects of caffeine, aminophylline, and similar compounds are observable in humans and are thought to be due at least in part to alterations in adenosine transmission, even though these agents have other biochemical effects that may also contribute. With increased interest in the exploration of the glutaminergic system in the etiology and possible treatment for depression, preclinical studies demonstrate increased glutamate levels in hippocampus after ECS in a depressive rat model. Unfortunately, it is not currently possible to conclude that alterations in any transmitter system are associated with the efficacy of ECT in any syndrome for which it is used.

Most recent preclinical research focuses on growth factors and molecular mechanisms. It appears that BDNF and NGF are both increased by a single ECS and chronic ECS, whereas neurotrophin is decreased. There is interest in linking these observations to those made of marked neural plasticity in some brain areas, for example, hippocampus, due to ECS, as will be discussed below. Work continues with a wide variety of other neuropeptides, for example, somatostatin, CCK, precursors of opiates, and thyroid hormones, etc. Other areas of interest concern the effects of ECS on second messenger systems, genetic expression, and transcription regulation. The work, focused largely on molecular phenomena related to cell proliferation—both neuronal and vascular, plasticity, and resiliency—including antiapoptotic effects, is expanding steadily, but its clinical and functional significance remains to be defined.

Hypotheses related to neuroendocrine mechanisms were generated early in the history of the use of ECT and have persisted to the present. Most hypotheses reason that stimulation of diencephalic structures, especially the hypothalamus, is involved and perhaps critical to the efficacy of the treatment. Experimental designs generated by these hypotheses have involved indirect quantification of function, specifically various peripheral neuroendocrine measures. Examples include CRH, ACTH, and cortisol, both acutely and subacutely; studies of oxytocin and vasopressin; and measures related to growth and thyroid hormones. Currently, there is no consistent evidence that acute or subacute changes in neuroendocrine factors impact efficacy or cognitive effects. Similarly, the involvement of immune system in the therapeutic effect of ECT has also been investigated. The available data suggest that a single session of ECT induces an acute, transient immune activation as measured by cellular immune function (lymphocyte subpopulations and NK cell activity) and levels of circulating cytokines, whereas repetitive ECT treatment results in long-term downregulation of immune activation. More prospective, controlled studies with larger sample sizes are needed to confirm this finding.

Neurophysiology

With advances in the area of brain imaging, neurophysiology has become a focus of interest in investigating the mechanisms of action of ECT. It is known that large increases in global cerebral blood flow (CBF) and cerebral metabolic rate (CMR) occur along with increased blood–brain barrier permeability during a generalized seizure, including those electrically induced. In the postictal period, there is functional suppression with decreases in CBF and CMR as well as varying degrees of bioelectric suppression on EEG. Changes are also topographically distributed. These reductions have been correlated with treatment outcome.

The magnitude of reductions in CBF in specific prefrontal regions is related to efficacy of ECT, and responders are more likely to exhibit postictal suppression and develop slow-wave activity on EEG than nonresponders. As techniques have been refined, the possibility of identifying specific functional networks and associating them to efficacy has improved. Related work in field modeling has yielded information concerning the way seizures are generated with various electrode placement and electrical dosing techniques as well. For example, studies comparing effective forms of bifrontotemporal, bifrontal, and right unilateral electrode placements indicate that the prefrontal cortex, particularly on the right, the anterior cingulum, and associated subcortical areas including the right medial thalamus, brainstem, and midbrain tegmentum are involved. Ineffective forms of ECT, for example, low-dose right unilateral, trigger seizures from the motor cortex, and involve the prefrontal area less. Positron emission tomography (PET) scan investigations point to these areas as well, particularly Brodmann area 25, the anterior cingulum, and the thalamus. These observations are important to clarifying which brain regions are critical to the efficacy of the treatment, as well as to understanding the dose–response relationship of electrical stimulation above seizure threshold to efficacy that has been observed for right unilateral ECT. The examination of the functional connectivity using functional MRI before and after bilateral ECT treatment revealed a significant increase in and around the left dorsolateral prefrontal cortical region (Brodmann areas 44, 45, and 46), where the average global functional connectivity was considerably decreased after ECT treatment, corresponding to the improvement in depressive symptoms. There has been expanding work using brain imaging technologies to investigate networks related to the cognitive effects of the treatment as well. Although preliminary, this work confirms the involvement of the medial temporal lobe in anterograde amnesia and points to prefrontal cortex involvement in retrograde amnesia.

Although the neurophysiologic inquiries are continuing, they are hindered by the limited localizing value and coarse resolution of currently available methods. The degree of covariance among various regions and the spatial distribution and magnitude of association are yet to be defined. There is also a need to better characterize topographies and to determine whether normalization, abnormalization, or independence of observations applies. Currently, investigational samples are small and heterogeneous as well.

Genetic and Epigenetic Mechanisms

In addition to focus on cellular level changes relevant to the antidepressant effect of ECS, advances in genomics have provided new insights about the gene expression and epigenetic changes associated with ECS. The first studies of ECS-induced gene expression were published in the early 1990s. Further studies have revealed regulation of distinct neurotrophic signaling pathways (BDNF, neuropeptide Y, vascular endothelial growth factor, fibroblast growth factor), synaptic vesicle proteins (synaptosomal-associated protein, *Syt 3 and Syn 1*), and immediate early genes (*c-fos, Egr 1, and Nrn 1*). BDNF has received extensive attention because of its involvement in the etiology of depression and response to antidepressant treatment. It has been shown that the BDNF expression in the hippocampi is elevated immediately (in the few hours) after acute ECS and decreased afterward in a temporal-dependent manner, while BDNF expression was decreased in the ventral tegmental area. The expression patterns highlight the widespread action of ECS and the spatial and temporal diversity of expression changes for each gene type.

A few studies also attempted to examine possible epigenetic mechanisms, that is, the process of modifying gene expression independent from the primary DNA sequences in the therapeutic effect of ECT. Evidence indicates a strong effect of ECS on the different epigenetic mechanisms in the rodent brain. One of the earliest studies by Tsankova et al. in 2004 reported histone modification at gene promoter regions in hippocampi of rats undergoing acute and chronic ECS in response to electrical stimulation. Increased levels of histone H4 acetylation at the promoters of the *c-fos, BDNF,* and *CREB* genes were observed after acute and chronic seizures, which, in general, corresponded with increased mRNA levels of these genes. Demethylation at the *BDNF-IX* and *FGF-IB* promoter regions, increased DNA methylation at the *Arc* promoter, and demethylation at the *c-fos* promoter region were also reported. Generalizing this work to clinical populations remains to be carried out.

Neuroplasticity

Recently, there has been increased interest in structural changes in the brain associated with psychiatric syndromes and response to treatment. This has been particularly so for microscopic changes associated with ECS, as well as antidepressant and other medications. It is observed that the delayed behavioral response to common antidepressant medications in rodents coincides with the delayed upregulation of neurogenesis, at approximately 3 weeks. Similarly, the rapid antidepressive action of electroconvulsive shock is associated with a rapid stimulation of neurogenesis, producing immature cells demonstrated to be highly excitable compared to mature neurons. Synaptic plasticity, in addition to cellular plasticity (neurogenesis), for example, sprouting of mossy fibers, alteration of cytoskeletal structure, and increased connectivity in performant pathways have been observed in animals after ECS. In addition to neurogenesis after ECS, microenvironment changes have been observed: for example, upregulation of activated microglia cells, the immune cells of the CNS, and increased proliferation of astrocytes, angiogenesis in certain layers of hippocampus, which may potentially modulate neurogenesis process after ECS. These reports have also galvanized controversy over various aspects of technical validity of the observations. It is unknown whether such changes occur clinically, and, if they do, what significance to efficacy and cognitive side effects might be discovered.

INDICATIONS

Although, historically, the first diagnostic indication for ECT was schizophrenia, practice has evolved. Approximately two-thirds of patients receiving ECT in the United States currently have major depression as the diagnostic indication for its use. The remainder has catatonia, schizophrenia spectrum disorders, mania, and a few other diagnoses, such as Parkinson disease. The clinical literature demonstrating the short-term efficacy of ECT in the major syndromes for which it is used is considerable, and includes randomized controlled trials with sham treatment or with medications as alternatives, observational data from studies involving various treatment techniques, as well as case reports/series, and surveys of expert opinion.

Beyond diagnostic indications, there are other clinical considerations commonly brought to bear in the prescription of ECT. Historically, ECT was a first-line treatment for the major psychiatric syndromes, but, with the development of psychopharmacologic treatments, use as a primary treatment has decreased, and resistance to medications has become the most common clinical indication. Nonetheless, when patients require rapid clinical response because of the severity and danger of psychiatric symptoms or medical problems resulting from psychiatric causes, ECT may be given without a prior drug trial. This is also the case when there has been prior optimal response with ECT or the patient prefers the treatment over medications (Table 33.20a–2).

Principal Indications

Major Depression. ECT is the most effective treatment for major depression currently available. It is commonly used as the standard against which other treatments, including innovative treatments such as novel forms of brain stimulation, are compared. No treatment has ever been found to be superior to ECT in the treatment of major depression in a controlled trial.

Evidence for ECT's efficacy derives from several different sources. The introduction of the treatment occurred prior to the development of clinical trials methodology in the postwar era. In the 1940s, open clinical series and case studies cited response rates of 80% to 90% with the use of ECT as a first-line treatment. Decreased chronicity of illness, decreased morbidity, and the suggestion of decreased mortality were frequently cited as benefits of ECT. From the 1950s through the 1980s, ECT was subjected to the randomized clinical trials with sham ECT and psychotropic medications as comparators. In the sham versus real ECT trials, anesthesia alone was used for comparison groups because of a concern that there were benefits from its repeated use. ECT was shown to be consistently superior, and the further use of sham conditions has been abandoned in studies of major depression. From the 1950s onward, there was also interest in the relative efficacy of antidepressants and ECT. In these reports, ECT was found to be 20% to 40% more effective than medications, and there has not been a study that found medication to be superior to ECT. Most of this work was conducted in the earlier decades of this period and is limited by the possibility that modern pharmacologic standards would dictate more aggressive dosage and duration of treatment in the medication conditions. In this type of study, adequate blinding is not possible, and concerns about bias cannot be completely dispelled even if clinical raters do not know the treatment conditions. Many of the antidepressants investigated in the past have been superceded by newer compounds, such as the specific serotonin reuptake inhibitors, and pharmacologically unique agents such as bupropion and mirtazapine. Unfortunately, there is little data to inform comparisons of ECT and modern pharmaceuticals.

Table 33.20a–2.
Indications for Use of Electroconvulsive Therapy

- Diagnostic indications
  - Principal diagnostic indications
    - Major depression, both unipolar and bipolar
    - Catatonia
  - Other diagnostic indications
    - Schizophrenia, especially treatment refractory
    - Mania, including mixed episodes
  - Less common diagnostic indications
    - Neuroleptic malignant disorder
    - Super refractory status epilepticus
- Clinical indications
  - Principal clinical factors
    - Need for rapid response required on medical or psychiatric grounds
    - Failure to respond to pharmacotherapy in the current episode
  - Other clinical factors
    - Risks of alternative treatments outweigh benefits
    - Intolerance of pharmacotherapy in the current episode
    - Patient preference

Although in an earlier era prior to modern advances in defining the phenomenology of depression, it was thought that various clinical criteria conferred a negative prognosis to the use of ECT, more recent observations have found ECT to be effective for all subtypes of major depression including unipolar and bipolar depression, melancholia, and psychotic depression. There have been suggestive data that psychotically depressed patients have a superior response to ECT, whereas there is suboptimal response with antidepressants alone and the recommended regimens of adequate doses of both an antidepressant and an antipsychotic are seldom given, perhaps due to poor tolerance of the combination. ECT may be particularly useful for psychotically depressed patients because the syndrome is often accompanied by severe symptoms and greater risk of suicide.

ECT can also be used safely in all populations of patients, including those often considered to be at special risk: the older adults, the medically ill, pregnant women, and adolescents, for whom ECT is now a labeled FDA indication. Each of these groups may require specific considerations and extra preparation and/or precautions to be taken prior to treatment. ECT has even been used to treat preadolescent children, although rarely.

Dilemmas persist among experts in the field about when to offer ECT in the course of treatment for major depression, and the speed and efficacy of ECT are often the most salient factors in selection. Severity of symptoms and refractoriness to psychotropics are the most common clinical factors for recommendation of ECT. The hierarchy of factors that need to be considered has not been resolved, but it is important to gauge the risk of not terminating an episode of depression while giving sequential trials of antidepressants along with augmentation and combination medications against the risk of the principal side effects of ECT, namely persistent retrograde memory deficits of varying significance. With recent improvements in technique, these cognitive effects have been reduced to the point where they can be difficult to detect. Even with these improvements, an acute course of ECT will often disrupt the routines of daily life because of the recovery requirements accompanying general anesthesia. Another important consideration relates to the fact that major depression is a subacute and often chronic disease with a need for continuation and maintenance treatment. For ECT, acute response is followed by the need to determine adequate continuation and maintenance treatments, most often pharmacologic, although continuation/maintenance ECT is also an option. This circumstance can also affect the timing of the choice of ECT versus pharmacology.

Finally, the response rates cited early in the history of ECT must be adjusted to the changing makeup of patient populations receiving the treatment. It appears that prior medication nonresponse within an episode, extended lengths of episode, and comorbidity with other psychiatric conditions may lower the expected response rate. Comorbidity with personality disorder, that may independently contribute to mood changes, can be especially problematic since patients with personality disorder can also suffer major depression, and it is difficult to define what phenomenology belongs to each syndrome and to determine expectations. This does not automatically mean that ECT should be withheld from these populations. Response rates to alternative treatments may be even lower, while response to ECT in antidepressant nonresponders can be expected to be 50% to 70%.

Catatonia. Catatonia is a syndrome characterized by stupor, mutism, and a variety of motor symptoms, including posturing, waxy flexibility, and catalepsy. Schizophrenia and bipolar disorder are the most frequent underlying diagnoses, but with the increased occurrence of autism in modern populations, this illness is becoming a more common diagnosis comorbid with catatonia. One of the more commonly reported uses of ECT in adolescence is for patients with developmental disorder accompanied by pharmacologically refractory dangerous behaviors thought to be on the catatonic spectrum. Currently, benzodiazepines are the primary treatment for catatonia along with a primary pharmacologic approach to the comorbid psychiatric disorder. When catatonia is resistant to benzodiazepines, ECT may be used. The data supporting this approach are largely uncontrolled, but quite positive. The syndrome of lethal catatonia is characterized by stupor or excitement, hyperthermia, autonomic dysregulation, rigidity, and changes in consciousness. ECT can be quite effective and, in fact, lifesaving.

Other Indications

Schizophrenia. Although schizophrenia is the most common diagnostic indication for ECT worldwide, there is a lack of expert consensus about its use and wide disparity in practice. Especially since the advent of atypical antipsychotic medications, pharmacologic treatment remains the first choice in the treatment of schizophrenia. ECT was introduced for use in schizophrenia, and the earliest reports, consisting of case series and comparisons with historical cohorts and nonsomatic treatments, were quite positive. As with other observations from this era, there was a lack of scientific methodology in diagnostic and outcome criteria, which would not conform to modern standards. Real versus sham ECT studies were carried out with chronic patients in the 1950s to 1960s and more acutely ill samples on neuroleptics in the 1980s. The earlier studies showed no advantage for ECT over anesthesia, but the later studies showed ECT to be superior to anesthesia alone. The source of the disparate outcomes is not clear, but in addition to differences in chronicity, the presence of neuroleptics in modern samples may be critical. Because of the positive reports of efficacy in the first decades of use, ECT was also used as the standard against which to compare newer somatic therapies, particularly neuroleptics, when they were first introduced. Although subject to the usual limitations of such early clinical trials, it soon became apparent that, at best, ECT was equivalent to neuroleptics in short-term efficacy. At this point, pharmacology superseded ECT for the treatment of schizophrenia. One observation from this era has remained intriguing and might merit further consideration. In long-term follow-up, patients who had received ECT in the first episode had superior outcomes.

In modern practice, ECT is often used as an augmentation to psychotropics for treatment-refractory psychosis, especially for those with particularly severe symptoms. In recent meta-analyses, an advantage has been shown for the combination of ECT with psychotropics, particularly clozapine. It appears that lower doses of neuroleptic are sufficient in combination treatments. Recent reports confirm the safety of using ECT and atypical antipsychotics, with the clozapine as the oldest and most reported agent. Most professional bodies issuing guidelines for the treatment of schizophrenia reserve ECT for patients who are nonresponsive or intolerant to clozapine.

In general, good prognosis patients for pharmacologic treatment are also those who do best with ECT. Evidence to support the traditional idea that the presence of affective symptoms improves the likelihood of response to ECT is inconsistent, and there is no controlled evidence concerning its use for schizoaffective patients. Uncontrolled observational evidence suggests efficacy of ECT in schizoaffective illness, but response falls short of that obtained in straightforward affective disorder and resembles response rates obtained for schizophrenia.

One last area of interest is the impact of ECT on the movement disorders resulting from neuroleptic treatment of schizophrenia. Extrapyramidal symptoms induced by neuroleptics are reversible by pharmacologic intervention, but also respond to ECT. Furthermore, there is evidence from several case series that ECT may protect against the development of tardive dyskinesia in neuroleptic-treated patients, independent of the observation that combination treatments with ECT and neuroleptic require lower doses of neuroleptic. This observation is consistent with preclinical observations of the effects of ECS on the dopaminergic system and the effects of ECT on patients with Parkinson disease.

Mania; Mixed States and Rapid Cycling. Historically, ECT was a first-line treatment for mania as well, but pharmacologic strategies, particularly lithium, anticonvulsants, and atypical neuroleptics, have become the mainstay of treatment for the manic episodes. In the modern era, the use of ECT is generally reserved for patients who are resistant or intolerant to the usual medication treatments, including those with mixed states or rapid cycling, or who have very severe symptomatology, for example, manic delirium. Compilations of the evidence for the use of ECT in mania have reported 72% to 92% response rates. The bulk of the evidence comprises uncontrolled case series, undoubtedly because of the great difficulty in conducting usual clinical trials in this population, due to uncontrolled behavior, difficulties in assuring capacity to consent, etc. Sham versus real ECT studies, with patients on neuroleptics in both conditions, demonstrated superior outcomes for ECT patients both in response rates and in lower neuroleptic requirements. Controlled retrospective comparisons have found ECT to be equivalent to lithium, and at least as efficacious as typical neuroleptics, if not more so. Prospective work demonstrated equivalence to lithium in studies where all patients received neuroleptics as well as the study treatments. One small study comparing ECT to lithium and neuroleptic in patients who failed lithium or neuroleptic found ECT to be noticeably superior to combined pharmacology. There have been no comparisons of ECT to anticonvulsants. Early work that suggested the need for increased frequency in treatment schedule and increased numbers of sessions to complete a treatment course has not been born out in modern studies, which have used standard treatment schedules.

Prospective observations and a retrospective controlled reviews have found that patients in mixed states are responsive to ECT, although increased numbers of treatments may be required compared to pure manic or depressed states. Case reports have indicated that ECT is also potentially effective in managing rapid cycling, where finding effective pharmacologic approaches can be particularly complex and difficult. Response rates are lower than for acute manic states.

Less Common Diagnostic Indications

In general, ECT treatment for other indications is supported by case material only. The use of ECT will be rare and only when standard treatments cannot be tolerated or fail.

Psychiatric. There is only very limited information about the impact of ECT on other disorders, and no evidence for clinically significant efficacy without the presence of another primary diagnostic indication. There has been interest in anxiety disorders because of the overlap in symptomatology and pharmacologic therapeutics between affective and anxiety disorders, but the only published data concern obsessive-compulsive disorder (OCD). Prospective work demonstrated acute response in OCD, but patients soon relapsed, and ECT is not recommended for this disorder. Comorbidity with post-traumatic stress disorder (PTSD) and depression has provided an opportunity to observe the effects of ECT on PTSD. A minority of patients with PTSD may be vulnerable to increased symptoms related to the syndrome, but there are also case series where PTSD improved independent of the response of depression to ECT. Patients with double depression (dysthymia and major depression) have similar response rates to those with episodic major depression alone, but are left with higher levels of mild depressive symptoms at the end of treatment, suggesting that dysthymia is more difficult to ameliorate with ECT even in the setting of comorbid major depression. The delirium that can accompany withdrawal and toxic states induced by some substances, that is, alcohol, phencyclidine, and stimulants, can be treated with ECT, but the underlying substance abuse syndrome is not significantly affected. This use of ECT is largely of theoretical interest.

Another area of concern is the issue of comorbidity with personality disorders. Here most of the work concerns cluster B disorders, especially borderline personality disorder. A secondary area of interest should be cluster A disorders because of their hypothesized overlap with axis I anxiety disorders. In general, comorbidity with one or more personality disorders is associated with poorer outcomes, especially higher relapse rates. The challenge for clinicians is to discriminate between affective symptoms that result from the personality disorder and those that constitute an ongoing affective episode. ECT is not considered to be a treatment for any personality disorder.

Neurologic. ECT has been used rarely for the treatment of several neurologic disorders, usually in situations where they are comorbid with a psychiatric disorder that is an indication for ECT. Neuroleptic malignant syndrome (NMS) is thought to be an iatrogenically induced form of lethal catatonia and is often treated with dopaminergic strategies as well as sedatives. Outcomes with ECT are equivalent to those obtained pharmacologically, but ECT has the advantage of often being a treatment for the comorbid psychiatric syndrome for which neuroleptics must be discontinued. There is evidence that NMS is less frequent with atypical antipsychotics, and the use of ECT for NMS may become less necessary as practice changes.

Open- and sham-controlled data have demonstrated that the motor symptoms of Parkinson disease, especially rigidity and bradykinesia, are lessened by ECT, particularly in the “off” phase of treatment that characterizes the phase of the disease treated with L-dopa. These effects are of 4- to 6-week duration generally, but there are reports that maintenance ECT may be helpful as well. ECT is seldom used as a primary treatment for Parkinson disease and is more frequently used to treat the comorbid depression that can occur with the disorder. The combination of ECT with levodopa/carbidopa may require decreasing medication dosage in order to stave off dyskinesia or delirium as complications of combined treatment. With the increasing use of Clozaril and other atypical neuroleptics to manage the side effects of levodopa/carbidopa, the use of ECT for Parkinson disease has diminished.

The anticonvulsant effects of ECT are well characterized, and ECT has been used in epilepsy, although effects are time limited, and ECT has no demonstrated superiority over anticonvulsant medications. When such patients are treated, including when they are treated for comorbid psychiatric conditions more commonly treated with ECT, the dosage of anticonvulsant medications is usually not changed, except when therapeutic seizures cannot be elicited. Recent reviews have suggested that ECT might be of use in intractable seizures or status epilepticus, particularly severe superrefractory status epilepticus unresponsive to pharmacologic treatment, including immunotherapy and plasma exchange. Outcomes vary with the associated underlying pathology and the degree of brain tissue damage sustained before ECT.

There are reports of lessening of pain syndromes when comorbid depression is treated with ECT. ECT has also been used to mitigate other movement disorders, for example, Tourette syndrome. Historically, there have been cases of delirium associated with both medical and psychiatric illnesses that respond to ECT, although the treatment is very, very seldom used for this reason currently.

The remainder of ECT use in neurologic syndromes is as a treatment for comorbid affective and, less commonly, psychotic syndromes. There are case reports and series of its use in poststroke depression and mood disorders accompanying dementia, mental retardation, and even head trauma and brain tumors. In all of these diagnoses, the risk of cognitive side effects of ECT can be greater, and modifications of the treatment may be needed, for example, to avoid skull defects found with trauma or to minimize the ictal blood pressure changes for patients with modest, stable brain tumors.

PATIENT MANAGEMENT

Pretreatment Evaluation

Evaluation of patients for treatment with ECT should be done as close to the first treatment session as possible. Both psychiatric and medical assessments are needed in order to conduct a risk–benefit analysis for the advisability of ECT treatment of an individual patient. In general, there are no absolute medical contraindications to ECT. Although a cochlear implant is labeled as a contraindication, ECT has been reported to be administered safely to the contralateral side of an implant, and the implants themselves were not damaged by ipsilateral ECT administration in cadavers. There is also literature on ECT use in the presence of deep brain stimulation and vagal nerve stimulation implants. Safe use requires attention to the location of ECT treatment electrodes with respect to those devices and, in fact, any implanted electrical device and consideration of temporarily turning off devices during ECT sessions. It may not be possible to administer ECT in every situation involving implanted stimulation devices.

The psychiatric component of the pretreatment evaluation is generally conducted by a practitioner privileged to administer ECT, but there are often contributions from the patient's attending psychiatrist, if different from the ECT provider. Establishing the indications for ECT will include a careful diagnosis as well as consideration of other clinical factors contributing to the recommendation for ECT, for example, severe suicidal risk. There may be local requirements and regulations for second opinions, especially in special populations such as adolescents and children.

Past treatment with ECT should be reviewed not only for response, but also to identify technical aspects and tolerance of the treatment. Knowledge of prior ECT treatments, including review of specific ECT records, will guide recommendations for the current course. Pharmacologic treatments should be evaluated for number, adequacy, tolerance, and degree of response. Psychotropic history can contribute to the indication for ECT as well as a basis for recommendations concerning post-ECT continuation treatment. Decisions regarding the adjustment of psychotropic medications pre-ECT should be based on the degree to which they may enhance response, for example, maintaining an antipsychotic; the possibility that they may interfere with ECT, for example, anticonvulsants used as psychotropics; the possibility that medications may contribute to adverse effects, for example, lithium; and the need to maintain medications for patient management while ECT takes effect.

As part of the pretreatment evaluation, there should be documentation of the target symptoms to facilitate evaluations during the treatment course and decisions concerning the termination of acute treatment. In addition to behavioral assessments, cognitive evaluation pretreatment is also critical. At a minimum, those aspects of cognitive function most likely to be affected should be documented at baseline: orientation and memory, including recall of recent and remote events. Evaluation of language and visual motor and perceptual abilities can also be valuable as well. Finally, the patient’s subjective assessment of their cognitive and memory functioning should also be documented, particularly since the predominant diagnostic indication for ECT, major depression, is often accompanied by both subjective and objective memory problems.

The medical component of the pre-ECT evaluation focuses on the patient’s ability to tolerate anesthesia as well as ECT and should be conducted by an ECT practitioner as well as an anesthetist. The general medical evaluation will be similar to that conducted for patients undergoing outpatient surgery and will focus on, although not be limited to, cardiac (e.g., congestive heart failure [CHF], symptoms of coronary artery disease), pulmonary (e.g., Chronic obstructive disease), and neurologic systems (e.g., acute stroke, increased intracranial pressure), those most likely to be impacted by ECT. The patient’s past experience with anesthesia should be elicited, as well as a family history of capacity to tolerate anesthesia. A complete list of medications, including herbal and over-the-counter medications, as well as medication allergies, is also needed. Most providers document a complete blood count and electrolytes as well as an electrocardiogram. In addition, a chest x-ray, thyroid function tests, a complete metabolic panel, and a pregnancy test in women of childbearing age may be needed. A brief dental examination for loose or missing teeth and the presence of bridges and dentures is recommended. Testing of plasma pseudocholinesterase or dibucaine number is done as indicated by patient and family history. The risk of fracture of the long bones or spine has been almost markedly diminished by the use of muscle relaxants. Hence, spinal x-rays are done as indicated by the patient’s history of skeletal injury or disease. Brain imaging is recommended when a questionable neurocognitive history, focal findings on neurologic or funduscopic examinations, or altered neurocognitive examinations are present. The basic medical evaluation should reveal whether there may be a need for further specialty consultations, testing, or labs. Many institutions involve an internist routinely, especially when conditions that can be affected by or interact with ECT, for example, diabetes, will need management during the course of treatment.

Taken as a whole, the pretreatment evaluation will result in recommendations for the modification of medications the patient may be receiving, both on a medical and a psychiatric basis; the modification of ECT technique; the modification of anesthesia technique; and the management of both psychiatric and medical conditions during the ECT course.

Consent

Written informed consent is standard in the practice of ECT and involves a discussion of the risks and benefits of the treatment with the patient and/or legally sanctioned surrogate, if the patient lacks capacity to consent. Procedures for consent are locally regulated, particularly at the state level, and familiarity with these regulations is essential. Consent can include both the anesthetic procedures and the electrical stimulation, although institutions typically provide separate consents for each of these aspects of treatment. Consent forms generally cover several different areas and can be lengthier than the type of document used for other comparable medical or surgical procedures. Descriptions of the procedure and expected benefits are detailed. Risks of the treatment including possible cognitive and medical adverse effects, including death, should be documented. The possibility of relapse and nonresponse should be included. Expected participation by or behavioral restrictions of the patient, for example, NPO for several hours before treatment, are spelled out. Alternatives to treatment, including no treatment, should be discussed. In addition to a generic description, patients should understand the individual indications and rationale for ECT. They should know who can answer any questions about the treatment that may occur during the course of ECT. It should also be clear that consent may be withdrawn during the course of treatment. The amount of information given should be tailored to each patient, recognizing that too much information can obscure the process just as too little. Guidelines and sample consent documents are available from professional organizations.

Most patients referred for ECT have the capacity to consent and can give it voluntarily. In general, capacity is assumed to be present unless there is strong evidence to the contrary. Where doubt exists as to the capacity, an independent consultation may be prudent. It is important to weigh the protection of autonomy and the opportunity to receive effective treatment in the consent process. While patients with capacity should not be coerced to receive treatment, seriously ill patients who lack capacity should not be unnecessarily obstructed from receiving it, especially when ECT is being recommended to prevent severe adverse outcomes such as increased morbidity or even death. Although regulations vary by jurisdiction, most allow for involuntary treatment with surrogate consent and/or legal review, which can include a formal judicial review. Some jurisdictions mandate not only judicial review but also review by the department of health for patients who lack capacity and object to ECT.

Surrogates should proceed through the consent process in a similar way to patients with capacity. Family members and other significant associates or companions are encouraged to become educated about ECT. Audiovisual aids and written material may be provided to patients, families, and others as indicated. Finally, it should be remembered that consent is a process throughout the course of ECT, and clinicians should maintain an open exchange about procedures, response, side effects, and prognosis. Professional organizations recommend separate consent be obtained for continuation or maintenance ECT as the indication has changed from acute treatment to relapse or recurrence prevention, and the risk–benefit assessment may have changed as well. With the increased use of tapering an acute course of ECT to continuation treatments, many hospitals have instituted consents that include a specific number of ECT sessions and duration of consent, regardless of whether they may be classified as acute or continuation ECT.

Parameters of the ECT Stimulus

Controlled trials in ECT over the past two decades have involved measuring the impact of varying aspects of ECT technique, most frequently electrode placement, pulse width, and stimulus dosing strategies, on response rates and adverse cognitive effects. Other aspects of the ECT stimulus include the direction and amplitude of the current and pulse frequency. Research from a prior era had led to a belief that seizure elicitation was necessary and sufficient for efficacy because subconvulsive stimuli were associated with lower response rates in several studies. When attenuation of electrically induced seizures with lidocaine was also associated with reduced efficacy, adequate seizure duration was also postulated to be necessary for therapeutic response. These tenets have been revised in light of observations from modern studies, and current practice has been directly impacted.

A series of studies since the 1980s demonstrated that, depending on the combination of electrode placement and stimulus dosage, seizures could vary in eliciting clinical response from 20% to 70%. Recent studies utilizing an ultrabrief pulse stimulus configured the waveform in a more physiologic way in order to minimize excess electrical dosage, similar to the substitution of brief pulse waveform for the original sine waveform. Right unilateral electrode placements were particularly affected by dosage, and a dose–response relationship exists up to a stimulus intensity of 8 to 12 times the seizure threshold. At electrical dosage six times the seizure threshold, right unilateral ECT is equivalent to bilateral ECT in clinical efficacy but maintains an advantage in producing fewer adverse cognitive effects. For electrical dosage above six times the threshold, the advantage in acute cognitive effects may be eroded, but there is evidence that an advantage in long-term cognitive effects remains. These findings make dosage relations with seizure threshold particularly critical for right unilateral ECT. Ultrabrief right unilateral ECT at six times seizure threshold has been shown to be equally effective to both brief pulse bilateral ECT and brief pulse right unilateral ECT; ultrabrief bilateral ECT at 2.5 times seizure threshold was an ineffective treatment in at least one randomized trial, and brief pulse waveform is recommended for bilateral ECT. Manipulation of the waveform in this manner had even more powerful effects on cognitive outcomes than had electrode placement, with results for ultrabrief right unilateral ECT statistically indistinguishable from those obtained in cognitive testing of age-, gender-, and IQ-matched medically and psychiatrically healthy controls on measures of retrograde amnesia. Combined with the results of neurophysiologic and brain imaging studies of the effects of ECT, the clinical impact of the interaction of electrode placement and dosage also suggests that the antidepressant effects of ECT are dependent on the anatomic specificity of the effects of current density rather than on seizure elicitation alone. It was also found that the speed of clinical response can be enhanced as stimulus intensity was increased above the seizure threshold for both unilateral and bilateral electrode placements.

Although there has long been a theoretical interest in greater localization of the effects produced by ECS in brain tissue, the neurophysiologic and clinical findings related to the interaction of electrode placement and stimulus intensity have rekindled interest in concentrating current density in prefrontal areas by manipulating electrode placement.

Electrode Placement. Historically, most practitioners have used bifrontotemporal electrode placement because of its reliability in producing efficacy and its ease of use. This electrode placement is also associated with more short- and long-term adverse cognitive effects, and is more likely to produce delirium, which may require interrupting a course of ECT and perhaps even terminating it before optimal therapeutic effects have been obtained. Hence, when bifrontotemporal ECT is used, attention should be paid to restricting dosage to a moderately suprathreshold to attenuate adverse cognitive effects as much as possible. It should be emphasized that the combination of ultrabrief pulse and bifrontotemporal electrode placement has not been demonstrated to be consistently effective. Treatment with bilateral electrode placements, particularly bifrontal configuration, is more likely to manifest EEG seizure without motor seizure, and EEG monitoring is particularly useful in detecting the occurrence of a seizure at treatment.

Newer electrode placements include bifrontal configuration and asymmetric placements. There are limitations to these strategies, imposed by the fact that the high impedance of the skull and scalp causes spreading of the electrical stimulus and restricts possibilities for localization of the stimulus. Bifrontal electrode placement, with positioning far enough laterally to minimize interference with impedance relations, has been investigated and there have been several demonstrations that bifrontal electrode placements are equally effective to bifrontotemporal and adequately dosed right unilateral electrode configurations. Evidence of advantages in sparing of cognitive effects is preliminary, and adequately powered investigations with more extensive and sensitive cognitive batteries are needed. Seizure threshold is likely to be relatively higher with bifrontal ECT.

The relatively better cognitive side-effect profile of right unilateral ECT should encourage wider utilization now that the efficacy of this electrode placement can be assured with adequate dosing strategies. In contrast to bilateral ECT, a dosage closer to 500% above the seizure threshold for right unilateral ECT is more likely to assure efficacy. ECT devices in the United States are restricted to an output in the range of 504 to 576 mC. Approximately 90% of patients have seizure thresholds that can accommodate optimal dosing with brief pulse right unilateral ECT, and the combination of right unilateral electrode placement with ultrabrief pulse width extends the range of U.S. devices so that most patients can be treated within these constraints. Maximizing interelectrode distance by using the d’Elia placement for right unilateral ECT may also be optimal. Many other right unilateral placements have been described, but there is little work to support their use (Fig. 33.20a–1).

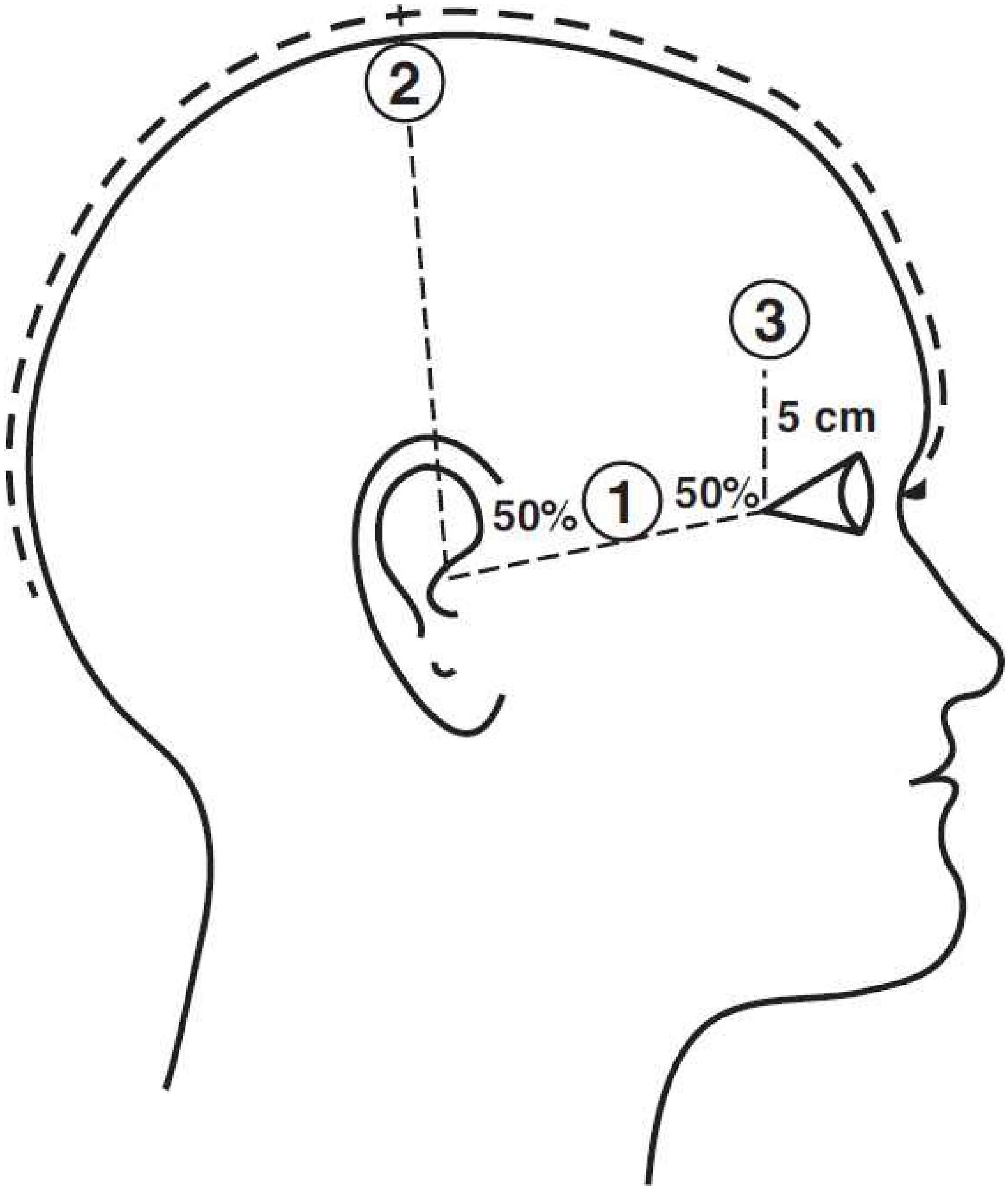

FIGURE 33.20a–1. Electrode placements. Position no. 1 represents the frontotemporal position, used for both electrodes, one on each side of the head, in conducting traditional bilateral ECT. For right unilateral ECT, one electrode will be in the right frontotemporal position, and the other will be just to the right of the vertex at position no. 2. Position no. 3 represents the frontal placement, used for both electrodes in conducting bifrontal ECT.

There has been some concern that left-handed patients may require different electrode placement from right-handed patients, especially if a unilateral placement is desired. There is evidence for independent lateralization of affect, with the right hemisphere involved in sustaining depressed mood regardless of handedness. In addition, no differences in efficacy have been found to be related to handedness when examined in studies. Because of limited indications that affective function and efficacy of ECT are associated with handedness, it is not generally used to guide the choice of electrode placement.

Stimulus Waveform. The original ECT devices used in the 1930s generated the electrical stimulus as a sine wave, with a rather wide pulse width, approximately 8.33 ms (60 Hz). Modification of the waveform to a bipolar rectangular pulse, 0.5 to 2.0 ms wide, represented an advance: diminishing nonessential electrical dose, dramatically reducing adverse cognitive effects, but preserving efficacy. Modern ECT devices retain this capability. However, neurophysiologic observations suggest that the optimal pulse width for neuronal depolarization is actually in the range of 0.1 to 0.2 ms. Thus, the brief pulse width is inefficient and delivers unnecessary stimulation. While reductions of pulse width to this extent may not be practical with the current engineering of ECT devices, several reports, including randomized controlled trial evidence, employing a pulse width of 0.25 to 0.3 ms (known as “ultrabrief”), have demonstrated that right unilateral ECT, typically at six times seizure threshold, is equivalent in efficacy to both right unilateral and bilateral ECT using brief pulse widths.

In one report, patients treated with the ultrabrief pulse width showed significantly less impairment in acute, subacute, and long-term cognitive side effects of anterograde and retrograde memory than patients receiving standard pulse width stimulation for both right unilateral and bifrontotemporal ECT. In fact, those who received ultrabrief right unilateral ECT did not deteriorate in any cognitive measure relative to pre-ECT performance. A smaller randomized trial confirmed the equivalent efficacy of ultrabrief and brief pulse right unilateral ECT. In effectiveness, open-label prospective and retrospective studies, findings have been more mixed, although predominantly positive for the efficacy of ultrabrief right unilateral ECT, but the lower quality of these designs must be factored into judging the outcomes presented. The possibility of differences in speed of response among electrode placements has been raised but not resolved. It is likely that there are a very small minority of patients who require specific electrode placements. for example, bifrontotemporal, and pulse widths, for example, brief pulse. It is not recommended that pulse wide exceeds 1 ms in modern ECT.

Seizure Threshold and Dosing. As indicated above, the intensity of the electrical stimulus, that is, the degree to which electrical stimulus exceeds the seizure threshold, affects the efficacy of right unilateral ECT, the speed of clinical response, and the cognitive side effects produced, regardless of electrode placement. The effects are in the same direction for each of these domains: the higher the stimulus dosage above the seizure threshold, the better the efficacy of right unilateral ECT up to 8 to 12 times the seizure threshold, the quicker the clinical response by one or two treatments over a course, and the greater the cognitive side effects. These observations have led to the development of dosing methods to refine stimulus intensity.

Strategies for determining stimulus dose include dosing based on empirical determination of seizure threshold (titration), formula-based dosing, and fixed dosing. Empirical titration involves an approximation of seizure threshold in the first treatment session, followed by a calculation of optimal dosage based on the results obtained. The first electrical stimulus in the titration is set low enough that a seizure results in only a minority of patients. Restimulations at intervals of 30 to 60 seconds are given in the same session using progressively greater amounts of charge until a seizure is produced. The stimulus intensity estimated on this basis (seizure threshold) is then generally used to compute dosage for the ensuing ECT sessions.

The dosage at which a seizure is elicited (seizure threshold) has been found to vary at least 50-fold in patients undergoing ECT. Because of the wide variation in seizure threshold, dosing based on empirical titration is thought to be the most precise method of stimulus determination. Furthermore, it is safe for most patients when conducted with adequate physiologic monitoring and the use of anticholinergic agents as appropriate to prevent the brief periods of bradycardia that may accompany subconvulsive stimulation. In controlled trials where the dosing was based on empirical titration, patients benefited clinically beginning with the first treatment session even though dosage was given near the threshold, a theoretical concern with right unilateral electrode placement. The seizure threshold, as empirically determined, increases by 20% to 100% over the course of ECT but returns to levels observed in the first treatment session within a week of ending treatment. Because of increases in seizure threshold over the course, retitration and dose adjustment may be needed if clinical symptoms plateau.

Formula-based dosing uses factors known to influence seizure threshold, such as age, electrode placement, stimulus pulse width, gender, and concomitant medications including anesthetics, to estimate adequate stimulus intensity. Seizure threshold is generally higher in men than women, with bilateral electrode placement than with unilateral electrode placement, and with increasing age. It is currently unresolved whether bifrontal electrode placement results in differences in seizure threshold when compared with bifrontotemporal placement. The basis for these observations has not been extensively investigated. The simplest formulae use age alone, but even the most detailed and elaborate formula can only account for about 40% of the variance in seizure threshold observed in patients. Hence, there has been limited success obtaining precision in seizure threshold approximation and dosing using formulae.

Fixed dosing occurs when the same dosage of electrical stimulus is applied to all patients receiving ECT. It is obviously the easiest technically and may maximize the chances of clinical response, especially for right unilateral electrode placement. Unfortunately, this dosing strategy can cause patients with the lowest seizure thresholds to receive up to 20 times an adequate electrical stimulus, especially with bilateral electrode placement. For low-threshold patients, fixed dosing enhances the possibility of increased adverse cognitive side effects. If fixed dosing is used with right unilateral electrode placement, the expected benefit in adverse cognitive effects could be lost. It is felt that fixed dosing may be appropriate only for patients with such serious medical conditions that even one subconvulsive stimulus, as is still a possibility even with formula-based dosing, must be assiduously avoided. Fixed dosing at the maximum output of the device should be avoided.

Seizure threshold and subsequent dosage determination can also be influenced by which stimulus parameters are adjusted. Comparisons of adjusting frequency versus stimulus train duration on brief pulse devices have indicated that longer duration is more efficient in eliciting seizures than increased frequency, although the effect sizes are small. Recent work has also shown that frequency does not exhibit a linear function in the depolarization, an observation that complicates understanding this relationship. Efficiency in seizure elicitation can also be optimized by using narrower pulse widths in setting the parameters on brief pulse devices. This strategy has the additional effect of extending the range of devices which are capped in stimulus output. There has been recent interest in optimizing pulse amplitude as well with initial indications that further efficiencies may be obtained. Use of unidirectional stimuli is under investigation.

Anesthesia and Treatment Procedures

Since ECT is given under brief general anesthesia, the treatment is usually performed in a specific treatment area that generally meets standards for the performance of day surgery, including the capacity to manage medical emergencies and to recover patients postictally. The staffing commonly consists of a psychiatrist privileged to administer ECT, an anesthetist, and one or more nursing personnel. Although the anesthetist maintains responsibility for anesthesia procedures, particularly airway management, the treatment procedures are conducted collaboratively.

Before anesthesia is induced, appropriate physiologic monitoring is instituted and continued until the patient is adequately recovered. Cardiac monitoring and pulse oximetry are standard; monitoring end-tidal carbon dioxide is also standard at many institutions. Automated noninvasive blood pressure monitoring, set at a short frequency, is common. Monitoring of seizure activity consists of observing both motor and EEG seizure duration. Motor seizure duration can be determined optimally using the “cuff method.” A blood pressure cuff is applied to the forearm or just above the ankle and inflated to a level above the anticipated maximum systolic pressure (>250 mmHg) to prevent anesthetics from affecting muscle contraction in the distal limb. Motor movements resulting from seizure induction can then be directly observed. The cuff should be ipsilateral to electrode placement for unilateral ECT. Ictal movements may continue longer in the face and neck. Use of the cuff method may not be appropriate for patients with severe peripheral vascular conditions, especially in the lower limbs. ECT devices have incorporated technology to allow at least 2-lead EEG monitoring with a frontal–mastoid montage preferred. Care must be taken to prepare EEG sites adequately and to take into account artifacts, for example, heartbeat and movement. Current FDA labeling requires EEG monitoring.

Oxygenation is provided using 100% O_2 at 15 to 20 breaths/minute via mask, and administered under positive pressure when respirations have stopped secondary to anesthesia medications, except during electrical stimulation. Respiratory support is discontinued when there is return of spontaneous respiration adequate to maintain clinically acceptable hemoglobin saturation on room air. Preoxygenation begins 2 to 5 minutes prior to induction and even earlier in the sequence of treatment procedures when medical conditions such as obesity or pulmonary disease pose a risk to respiratory function. In rare circumstances, such as third trimester pregnancy or anatomic airway anomalies, patients may be intubated.

There is variation in the use of anesthetic medications. Anticholinergic medications such as atropine and glycopyrrolate are usually unnecessary but may be used to dry secretions that may interfere with respiration under anesthesia and to decrease the vagally mediated bradycardia and asystole. The cardiac effects of anticholinergic agents are most useful in the following clinical situations: during seizure threshold titration when subconvulsive stimuli will not be accompanied by compensatory catecholaminergic output for every stimulation; in the presence of agents that promote bradycardia, for example, β-blockers; and when cardiac conditions characterized by some degree of the electrical blockade are pre-existent. Anticholinergic agents may be contraindicated when they might provoke pre-existing arrhythmias or threaten cardiac compensation through increases in workload from higher heart rates. Although glycopyrrolate does not cross the blood–brain barrier under normal circumstances and could have an advantage over atropine, controlled comparisons with atropine have not detected differences in cognitive or cardiac adverse effects or postictal nausea, perhaps because ECT disrupts the blood–brain barrier and diminishes this advantage.

Ultrabrief general anesthetics are used to induce unconsciousness while muscle relaxants are in effect and during seizure elicitation. Methohexital, a short-acting barbiturate, has been the preferred agent in the United States, due to near-ideal kinetics and only moderate anticonvulsant effects. Alternative anesthetics include short-acting agents from a variety of classes, including propofol, etomidate, ketamine, and short-acting opiates remifentanil and alfentanil. Propofol reduces hemodynamic changes accompanying ECT and typically produces a rapid, smooth recovery. Its antiemetic properties can be useful when patients develop nausea or vomiting. This agent also attenuates seizure duration, but studies comparing it to more standard agents have not detected differences in clinical efficacy or cognitive recovery. Agents that attenuate seizure duration, generally are associated with slightly shorter emergence and recovery although a recent review concluded that the differences while statistically significant may not be of sufficient magnitude to affect the choice of agent. Etomidate does not affect seizure threshold and is associated with fewer cardiac side effects, although recovery may be longer than with other agents. Etomidate can also produce myoclonus, and repeated use in animals has been associated with reductions in adrenal function, although no clinically relevant adrenal suppression has been associated with the intermittent dosing schedule associated with its use during ECT. Ketamine, an NMDA receptor antagonist has been associated with longer seizures in animals given ECS. With the potential involvement of the glutamatergic system in the etiology of major depression and schizophrenia, as well as the potent independent antidepressant effect of ketamine, and the theoretically neuroprotective effect of ketamine against cognitive impairment associated with ECT, the use of ketamine in ECT treatment has resurged. To date, studies do not support a synergistic antidepressant effect or mitigation of cognitive side effects of ketamine and ECT. Unfortunately, ketamine can result in hallucination and unusual states of consciousness as well as increases in heart rate and blood pressure. In general, both etomidate and ketamine have been reserved as alternative or adjunctive agents for situations where seizure elicitation is difficult. Greater attention to adequate dosing of these anesthetic agents has become increasingly important as the increased use of cognitively sparing techniques, such as ultrabrief right unilateral ECT have been accompanied by reports of patients regaining awareness and remembering the effects of succinylcholine before full respiratory capacity returns to baseline. Numerous reports have explored using short-acting opioids, either alone or in combination with standard agents. In combination, less of the standard agent is required and seizure durations can be longer; apnea may be longer than with other anesthetics.

Before the development of muscle relaxants, musculoskeletal injury, particularly fracture of the long bones or vertebrae, was not uncommon, as it is today. Muscle relaxants are part of the standard of care with ECT. Curare was developed to reduce the tonic–clonic movements produced by the electrical induction of a generalized seizure, but its use has been superseded by succinylcholine, a depolarizing agent with rapid onset and brief duration. Before administration, the airway must be secure, and unconsciousness achieved. Although modification of movement is the usual goal of using muscle relaxants, complete paralysis may be desirable when there is a history of skeletal disease, such as osteoporosis or a recent fracture. Myalgia can be a side effect of succinylcholine. The standard treatment of these myalgias, which are often present in only the first few treatments, is the use of common analgesics. If the use of succinylcholine is contraindicated, as in pseudocholinesterase deficiency, neuromuscular disease, prolonged immobilization, or risk of hyperkalemia, then nondepolarizing agents can be employed, although very low–dose succinylcholine is sometimes also an option. Agents available in the U.S. market, for example, rocuronium, are rather longer acting, even when the reversal is produced with an agent such as sugammadex.

Although complications at ECT have involved the cardiovascular system most commonly, the routine use of agents, such as β-blockers, which reduce the increases in hemodynamic activity that accompany ECT, is not recommended. Controlled studies of routine use do not demonstrate advantages for active intervention, perhaps because the patients who benefit are a minority of the entire population of ECT recipients, most of whom easily tolerate the observed increases in heart rate and blood pressure for a few minutes. There has also been some concern that the hypermetabolic cerebral state induced by the treatment requires a corresponding increase in cardiovascular functional level to maintain adequate oxygenation and nutrient supply to the CNS. The conservative position is to restrict the use of pharmacologic intervention to those conditions where the need is unambiguous, for example, aneurysm, sustained hypertension, and/or tachycardia.

Because the application of the electrical stimulus directly stimulates contraction of the jaw muscles, a flexible foam or rubber bite block is placed in the mouth just before the delivery of the electrical stimulus to protect the teeth and oral soft tissues. Stimulus duration is typically only a few seconds.

Patients are recovered with monitoring of cardiovascular and respiratory functioning until physiologically stable. If not awake and oriented at that time, further recovery is needed until the patient is alert and reoriented. Patients with medical comorbidities, that is, morbid obesity may require a lengthy period of recovery supported with prolonged oxygen supplementation. Many facilities require that the patient be fully ambulatory before leaving the treatment and recovery areas. Bed rest may be needed for a period of time. Side effects, such as headache and nausea, are monitored and treated as needed. Headache frequently responds to analgesics, especially COX inhibitors; patients with migraine may require more specific antimigraine treatments. Postictal delirium with agitation, disorientation, and lack of full consciousness, occurs in a minority of patients. Decreasing environmental stimulation and securing protection may be adequate management, but not infrequently, patients will require intravenous (IV) pharmacologic intervention with sedatives and prophylaxis at subsequent treatments.

ECT is considered an urgent, not an elective, procedure, both for acute and continuation and maintenance courses. In the COVID-19 era, patient screening for both urgency and infection status, adequate personal protective equipment for staff, and modifications of anesthesia, an aerosolizing process, should be implemented. Suggested modifications include maximal preinduction oxygenation, using bag-mask ventilation only when desaturation begins, viral filters in anesthesia circuits, limiting use of AMBU, and disinfection of nose and mouth.

During a treatment session, seizure duration is monitored for motor and EEG duration, with seizure duration commonly less than 1 minute. Motor duration is measured by observation of a cuffed limb with the cuff applied prior to muscle relaxant, and of the entire body as well; EEG is measured by application of one or two channel front-mastoid EEG leads prior to induction of anesthesia with visual and possibly auditory output on the ECT device. The relationship of duration to efficacy is tenuous at best, and it is clear that duration alone is insufficient to insure clinical adequacy. Seizures of less than 15 seconds can occur close to seizure threshold and are often considered abortive. Seizures of similar duration can also be observed in older adults, late in a treatment course, and with markedly suprathreshold dosage, but such short-duration seizures may not be associated with lack of adequacy. These circumstances highlight the usefulness of empirical titration techniques to inform the practitioner more exactly about seizure threshold and dosage, and to guide what actions, if any, to take when short seizures occur. Hyperventilation pre-ECT, decreasing dose of methohexital or propofol and substituting short-acting opiates (remifentanil) for part of the dose, switching anesthetic to etomidate or ketamine, and reversal of benzodiazepines with flumazenil are strategies used to lengthening seizure duration. Seizures longer than 3 minutes, with either motor or EEG monitoring, are considered prolonged and should be actively controlled using IV anticonvulsants, such as the short-acting barbiturate used to induce anesthesia or benzodiazepines. The occurrence of long-duration seizures is unusual, except in the youngest patients, and the cause of such long duration should be investigated, especially if it was not easily terminated.

Treatment Course

Typically, ECT is given two to three times weekly on nonconsecutive days. The latter frequency is common in the United States. Twice-weekly treatment is equally effective compared to thrice-weekly treatment, can take longer to reach efficacy, and is associated with fewer acute cognitive side effects. Other frequencies of treatment are more problematic. Daily treatment is seldom done now, but, at one time, was believed to speed efficacy dramatically and was used in the most urgent situations, for example, severe mania. This treatment schedule was also characterized by a risk of prolonged severe cognitive side effects, especially when a course known as regressive ECT was used. In addition to being logistically difficult to obtain in many places, the risks appear to outweigh the benefits, and daily treatment is no longer advised. Another approach to increasing frequency was eliciting more than one generalized seizure per session under continuous anesthesia, known as multiple-monitored ECT (MMECT). Similar to daily ECT, the goal was to accelerate clinical response with the additional advantage of decreasing the number of times anesthesia is delivered. Although there have been both reports of uncontrolled observations and one controlled trial concerning MMECT, the literature is clear that severe adverse events accompany this treatment approach with only modest gains in efficiency of response. Of most concern are reports of cardiovascular morbidity and mortality, as well as serious adverse cognitive effects including prolonged confusion. Although there may be a place for the use of MMECT, it is only in extremely unusual circumstances, for example, intractable seizures, and should be restricted to double stimuli/session. At the other end of frequency, once-weekly ECT is associated with weaker clinical response and more suitable for continuation ECT schedules where the goal is to maintain a patient’s status as responder or remitter—as opposed to acute treatment, which is used to induce response.

The number of treatment sessions in a course of ECT should be individualized, as patients vary in this requirement from a few sessions to more than 15, especially when changes in treatment technique have been made. The course length cannot be predicted, although the average is from 6 to 12 sessions. Fixed course lengths are not appropriate for these reasons. Patients should be evaluated every one to two treatments for clinical response and adverse side effects, particularly cognitive effects. These assessments are best done between treatment days when the acute effects of anesthesia and ECT will interfere less with evaluation. Acute treatment should end when there is remission of target symptoms or when further improvement has not occurred over two to three sessions with optimal technique.

When response is inadequate, modifications can be considered. Such modifications can include intensification of stimulus energy by increasing charge through manipulations of the various parameters of the stimulus or by changing electrode placement. Seizure threshold increases over the course of treatment, so higher stimulus intensity may be useful later in a treatment course if response is inadequate, particularly if response was more robust earlier in the course but waned. If any medications with anticonvulsant properties, including barbiturate anesthetics or propofol, are being used, discontinuation or decreasing dosage should be evaluated, particularly when the patient has high seizure threshold. At least 10 treatments with optimal electrode placement and dosing should be given before nonresponse is declared. When course length exceeds 15 to 20 sessions, especially with modifications of technique, reconsideration of the use of ECT, with careful balancing of the risk of an unresolved episode of illness against the risk of accumulating adverse effects, should be undertaken. Additional consultation, with reassessment of primary and comorbid diagnoses, may be advisable.

ECT is the only treatment for major psychiatric syndromes that is discontinued when remission occurs. Hence it is not surprising that a significant proportion of patients who relapse do so within the first few weeks of termination of ECT and then require further ECT. Often additional ECT is given along with medications that will be used for continuation and maintenance of remission. Although not yet examined in a controlled trial, tapering the termination of ECT by decreasing the frequency of sessions while medications are stabilized and producing their effects may be a more effective strategy for relapse prevention.

Continuation Treatment

Continuation treatment, that is, treatment for 6 months beyond remission of an acute episode of illness to prevent relapse, is standard practice for the major syndromes that are somatically treated. It is recommended and routine post-ECT as well. Populations that receive ECT are particularly likely to be characterized by risk factors for relapse: medication resistance and the presence of psychotic features. At least half and upward of 85% of ECT responders in some samples will relapse without continuation treatment particularly in the first several weeks post treatment. Aggressive continuation treatment should be initiated as soon as remission is evident, although initiation of pharmacology during the ECT course itself has not been shown to be of additional benefit for depression.

Continuation treatment with medications is the predominant practice post-ECT. This has been based on studies conducted in the 1960s in Britain and on clinical experience. The class of agent advised matched that used in the absence of ECT: antidepressants for unipolar depression, antidepressants and mood stabilizers for bipolar depression, mood stabilizers with or without antipsychotics for mania, and antipsychotics for schizophrenia. Studies in the 1970s found little support for the use of lithium alone as relapse prevention following depression, but some support for lithium as maintenance treatment after 6 months of remission to prevent recurrence once remission is firmly established. The first randomized prospective controlled study of continuation treatment post-ECT in unipolar patients confirmed substantial relapse rates on placebo (84%), but only moderate improvement in relapse rate on the antidepressant nortriptyline alone (60%). The relapse rate in unipolar depression was significantly reduced in this study by the addition of lithium to nortriptyline (39%) with most of the relapse in this arm of the study occurred in the first 5 weeks. Randomization had been stratified by medication resistance and (mood congruent) psychosis, but mood congruent psychotic patients did not relapse more frequently. Medication dosage was in the range used for acute treatment. A prospective trial of combination antidepressant and antipsychotic continuation treatment post-ECT did not find an advantage for combination treatment over antidepressant alone in an older adult population, and the combination treatment was found to be associated with greater frequency of side effects. A recent randomized controlled trial found no advantage for beginning nortriptyline or venlafaxine during the course of ECT in preventing relapse. Observations of community practice post-ECT yield relapse rates similar to those found in the RCT assignments using antidepressants alone. Little is known about the use of second- and third-generation agents or anticonvulsant mood stabilizers in relapse prevention, although preliminary evidence suggests that the combination of lithium and antidepressant is more effective than other medication combinations in maintaining remission. There is also preliminary work indicating that combining medications with continuation ECT or cognitive therapy is more useful than monotherapy. There are preliminary studies suggesting that relapse prevention is more effective when using a class of antidepressants that was not shown ineffective prior to ECT treatment. Modern practice suggests sparing use of antidepressants in bipolar patients, but in continuation treatment post-ECT, antiepileptic drugs or antipsychotics have not been shown to be all that effective in preventing relapse to bipolar depression. Little work in pharmacologic relapse prevention post-ECT for any other diagnosis than depression has been done in the modern era.

ECT was introduced before the psychopharmacologic era; thus, administering ECT after an index episode was not uncommon early in the history of ECT. With ECT populations becoming increasingly medication resistant and various pharmacologic strategies still characterized by relapse rates in the range of 40%, there is renewed interest in continuation ECT for patients who relapse on continuation pharmacology. The first randomized controlled comparison of continuation ECT in a fixed schedule with the combination of nortriptyline and lithium did not show an advantage for ECT and confirmed relapse rates observed in prior studies; relapse escalated when the schedule of administration dropped below once weekly. Other prospective data have confirmed these observations. Recent prospective, controlled, nonrandomized data as well as post hoc analyses from the RCT support the practice of tapering ECT after an acute course, combining ECT and pharmacology as continuation treatments, and using a symptom-driven, flexible schedule of ECT administration to improve relapse prevention. A recently completed RCT has confirmed this practice.

Recommendations from professional organizations have guided various aspects of the practice of continuation ECT although these bodies use differing thresholds for evidence when recommending strategies, a circumstance most notable when the National Institute for Health and

Clinical Excellence in the United Kingdom issued guidelines in conflict with the Royal College of Psychiatrists as well as other professional bodies in the Commonwealth and globally. Suitable patients for continuation/maintenance ECT should have recurrent illness, demonstrated response to ECT, intolerance of or relapse on continuation medications, and ability to comply with continuation ECT protocols, particularly behavioral prescriptions, such as fasting prior to the procedure, and restrictions, such as refraining from important decision making for a period of time following each treatment. In general, continuation ECT utilizes the electrode placement and dosage which produced remission. A taper of ECT from the acute phase to continuation phase has become common. Treatment frequency is gradually extended between treatment sessions, as tolerated, to monthly or even less often. It is thought that if patients remain well 2 months or more, they are out of episode, and the discontinuation of continuation ECT can be entertained. Continuation ECT is almost always given in an outpatient setting. Cognitive functioning should be monitored as well as the status of psychiatric symptoms and medical fitness for each session. There has been one promising but uncontrolled pilot study of the use of cognitive behavioral therapy to augment somatic continuation treatments post-ECT.

Maintenance treatments are those given beyond the 6-month continuation treatment period and the purpose of these treatments is prevention of recurrence. Because patients receiving ECT in current practice are almost always suffering from highly recurrent illness, maintenance treatment can be entertained for most ECT responders who do not relapse with continuation treatments. There are now several controlled observations of the use of maintenance ECT up to 4 years with positive outcomes in relapse prevention for depression. Japanese investigators have published a case series on the use of ECT as a maintenance strategy for treatment-resistant schizophreniform and catatonic illness. Individual decisions about maintenance treatment will be informed by the number, frequency, and intensity of prior episodes; the ability to tolerate continued treatment; the ability to comply with treatment; the experience when attempts to taper treatment are made; and patient preference. Again the standard will be maintenance medications. Maintenance ECT is also available, especially for those patients undergoing continuation ECT when medications have proven inadequate. Maintenance ECT is similar in many aspects to continuation ECT, and it is performed at the longest intervals to sustain remission. There is no evidence that a lifetime maximum number of ECT applies.

INTERACTIONS OF MEDICATIONS AND ECT

ECT can interact with medications given for psychiatric as well as medical conditions. As such, patients' medication regimen should be reviewed for risk and benefit during an ECT course to determine whether it is necessary to make changes. Changes can include tapering, withdrawing, augmenting, or adjusting the timing of administration of medications. Some psychotropics and many drugs for medical indications are continued. Medications that are already a part of the patient's treatment regimen and are expected to mitigate physiologic changes that accompany ECT or protect against adverse reactions to ECT or anesthesia are often given the morning of treatment. Examples are antihypertensives, antireflux agents, inhalers for pulmonary conditions, and steroids, although anesthetists may wish to substitute IV steroids for oral steroids pre-ECT. It appears that stress doses of steroids may not be necessary at ECT. Lidocaine and its analogs can interfere with seizure induction and should be withheld. Other medications that should be held until after recovery are diuretics, hypoglycemics, including insulin, depending on an individual evaluation of the patient's needs, and long-acting cholinesterase inhibitors for glaucoma that may interfere with anesthesia recovery.

Medications that can result in severe adverse reactions include theophylline and lithium. Theophylline is a member of a class of agents that are known to increase seizure duration, and theophylline can result in status epilepticus, even at therapeutic levels, when administered during an ECT course. In fact, caffeine, a short-acting member of the same class of medications as theophylline, had been used to increase seizure duration when seizure length was thought to be a critical parameter to ECT efficacy. Inhaled bronchodilators and steroids, as well as other oral and inhaled agents, have replaced theophylline in modern pulmonary practice to a great extent. If theophylline is part of the therapeutic regimen and ECT is undertaken, consultation to evaluate the possibility of a substitute regimen is generally sought. Lithium combined with ECT has been shown to result in confusional states, serotonin syndrome, and prolonged and/or focal seizures. Although it is not neurotoxic in all patients and the toxicity may be dose related, discontinuation of lithium is frequently undertaken in an acute course of ECT. Some observations in bipolar mania have concluded that lithium may be used with ECT, particularly at low dose. History of cerebrovascular disease and head injury raises the risk of combining lithium and ECT. For continuation and maintenance ECT, doses are often withheld for 24 hours prior to treatment, allowing for washout, and markedly diminishing the chance of negative interaction.

Anticonvulsants can increase seizure threshold, decrease seizure expression, and may interfere with the efficacy of ECT, particularly in depressed patients. Interference with clinical outcome appears less marked in manic patients administered ECT with anticonvulsants, although there is increased frequency of missed seizures. When used for control of a seizure disorder, anticonvulsants are continued but held until after treatment sessions, unless poor seizure expression along with weak response occurs. In this circumstance, levels of drug may be re-evaluated in consultation with neurology. If anticonvulsants are used for psychiatric indications, they are generally tapered and discontinued, particularly in bipolar depression, to prevent a negative effect on clinical response.

Benzodiazepines are also anticonvulsants, but more often used for psychiatric indications. There are reports of diminished efficacy when benzodiazepines are used in conjunction with ECT, especially right unilateral electrode placement. As a result, this class of medications is tapered to the least necessary dose, discontinued, or substituted with other agents, for example, melatonin, low-dose trazodone, antihistamines, and agents with short half-lives, for example, zalephon. If use of benzodiazepines is considered necessary for management of severe behavioral syndromes, agents with shorter half-lives are preferable and should be withheld several hours prior to treatment. Use of flumazenil to reverse benzodiazepines has been reported, but requires further study and benzodiazepines may need to be administered during recovery to abort possible withdrawal. Little is known about the interaction of other sedative/hypnotics with ECT.

Anesthetists have been averse to allowing the use of monoamine oxidase inhibitors (MAOIs) concurrently with elective anesthesia, due to a potential for interference with management of cardiovascular abnormalities that may arise at treatment. ECT seldom results in a situation where the use of indirect-acting vasopressors, which would not be metabolized because of interference by the recent use of MAOIs, is necessary. Hence, the literature includes many reports of the safe performance of ECT in the presence of MAOIs. Current practice allows for consultation between the psychiatric and anesthetic practitioners as to the best recommendation for any particular individual or practice setting.

An augmenting effect for neuroleptics when combined with ECT in schizophrenia can be accepted with some confidence. As discussed previously, the combination treatment is more effective than ECT alone and may be more effective than medications alone. A few older studies even suggest that the long-term outcome of schizophrenia may be enhanced by combination treatment. Whether the acute or long-term effects carry over to other psychotic conditions is not clear. Most of the published data involve the use of typical neuroleptics, It appears that low-potency typical antipsychotics may be proconvulsant; haloperidol and fluphenazine appear to have little impact on parameters of seizure quality. The use of ECT combined with atypical neuroleptics appears to be safe. Clozapine and olanzapine have been observed to have proconvulsant properties and quetiapine, anticonvulsant properties. Atypical antipsychotics are also reported to have U-shaped functions with respect to convulsant properties. Because of these circumstances, dosing of neuroleptics should generally be in the moderate range. The only antipsychotic that cannot be combined safely with ECT is reserpine.

Earlier work with tricyclics (tricyclic antidepressants [TCAs]) and MAOIs provided some evidence to recommend the use of antidepressants to augment ECT. A recent RCT found that nortriptyline increased the response rate to ECT by about 15%; nortriptyline was also associated with diminished adverse cognitive effects. The combination of ECT and antidepressants is common practice in many parts of the world. Increasingly, in the United States, psychotropics, including antidepressants, are continued during a course of ECT because of expanding limitations placed byz insurers on utilization of mental health resources. It appears that TCAs are safe in noncardiac populations, and SSRIs in most populations, although they can increase seizure duration. After a series of cases where asystole occurred at doses of venlafaxine greater than 225 mg even with the concurrent use of atropine, it may be advisable to limit the dosage of venlafaxine if it is combined with ECT. The well-known proconvulsant effect of bupropion does not appear to pose safety issues with ECT, which has notable anticonvulsant properties. It is unknown whether newer medications, including selective MAOIs, now used more frequently for depression have augmenting effects. There is a limited data on stimulants, but there is risk for hypertension, dysrhythmias., and effects on seizure threshold. Cholinesterase inhibitors used in cognitive disorders appear safe, but do not enhance ECT including cognitive adverse effects. Finally, there has been limited interest in pharmacologic approaches to mitigating the cognitive side effects of ECT. Preliminary data using physostigmine, thyrotropin-releasing hormone, and naloxone have been published, but whether clinically significant effects could be obtained remains in doubt.

PHYSIOLOGY, ADVERSE EFFECTS, AND CONTRAINDICATIONS TO ECT

ECT is generally low risk and one of the safest procedures performed under general anesthesia. Estimated risk of serious complications, occurring in about 1/1,000 patients, is similar to that of general anesthesia for minor medical procedures. Risk of death, likewise in the range of that observed for such procedures, was estimated in a recent pooled analysis of 15 studies (766,180 treatments) at about 2.1/100,000 patients in modern populations undergoing ECT. Some believe that these rates of serious complications and death are lower than those related to the use of antidepressant medications, for example, TCAs. Causes of death in the first 30 days post ECT: falls and pneumonia most common, but not clearly related to ECT. Long-term mortality rates for hospitalized patients receiving ECT are lower than for patients receiving other forms of treatment.

Cardiovascular System

The ECT procedure routinely results in parasympathetic outflow during and immediately after the electrical stimulus. Unopposed by anticholinergic medications, sinus bradycardia and even brief, clinically insignificant, asystole can occur as a result. In patients over 50 years old, supraventricular ectopic beats are common, but clinically insignificant. In vulnerable patients, vagally related arrhythmias can occur such as atrial, junctional, and nodal rhythms as well as atrial flutter, and atrial fibrillation. Following parasympathetic stimulation, sympathetic output increases, mainly during the electrical stimulation and postictally when the adrenal gland also releases catecholamines. As a consequence, heart rate, blood pressure, and, perhaps, most importantly, rate pressure product increase. These increases peak immediately postictally and drop off within minutes to pre-ECT values, although in those over 50 years, normalization takes up to 1 hour for blood pressure and longer for heart rate. There can be a short period of decreased heart rate at about 5 minutes post-ECT. Related abnormalities of cardiac rhythm can include bigeminy and trigeminy, which are generally benign. Repolarization abnormalities such as T-wave changes and ST segment depression can appear infrequently. Ventricular tachycardia, usually nonsustained, and fibrillation can occur rarely. About 55% of those with pre-existing cardiovascular abnormalities will manifest some abnormality in the course of ECT. About 10% of one sample manifested transient increase in troponins of uncertain clinical significance.

With judicious pretreatment evaluation of both the risks and benefits of ECT, cardiac consultation, and the stabilization of any underlying disease, many of the more serious cardiac complications can be avoided. Uncontrolled hypertension, recent myocardial infarction (MI) (4 to 6 weeks), unstable angina, decompensated CHF, severe aortic stenosis, uncontrolled atrial fibrillation, and abdominal aortic aneurysm (>5 cm) pose enough risk to require close evaluation for suitability. About 90% of these patients will be able to complete a course of ECT safely. For example, ECT has been administered to patients with histories of pulmonary hypertension, subacute MI, abdominal aortic aneurysm, and Takatsubo syndrome. CHF patients with ejection fractions over 30% have been able to tolerate ECT. In patients with MI, a period of time must be allowed for healing, but the risk relates not only to the time elapsed but also to the severity of resulting cardiac compromise. As with arrhythmias and CHF, stress testing can be informative in deciding whether a patient can tolerate ECT. The presence of a pacemaker or device to abort arrhythmias is not a contraindication to ECT. ECT has been administered without complication to stable patients with transplanted hearts.

Respiratory System

Along with cardiac complications, adverse events related to pulmonary function are a leading cause of morbidity and mortality associated with ECT. The use of muscle relaxants leads to the cessation of breathing and oxygen is delivered by positive pressure through mask substitutes. Where patients have underlying pulmonary disease, exacerbation of asthma or COPD is possible, as well as excess secretions and aspiration. Pulmonary function should be optimized pre-ECT, and inhalers used before each session. Excess secretion can be addressed with anticholinergic medication. Aspiration can be mitigated by strict adherence to NPO pre-ECT, but obesity, pregnancy, neurologic compromise, DM, hiatal hernia, tube feeding, and medications causing hypersalivation, for example, Clozaril, increase risk. Immobilized and catatonic patients are at increased risk for peripheral venous thrombosis and pulmonary embolism. Risk can be mitigated by evaluation for anticoagulation, compression stockings, and prescribed physical activity. Rare cases of negative pressure pulmonary edema have been attributed to inspiration against an obstructed airway or mechanical irritation precipitating laryngospasm; it can also be neurogenically induced, for example, as a complication of status epilepticus. With proper preparation pre-ECT and careful anesthetic technique, these events are uncommon. Obstructive sleep apnea poses risk or may even be discovered during ECT anesthesia; it can generally be managed acutely with preoxygenation, and the use of nasal pharyngeal airway or laryngeal mask. Another rarely observed complication is prolonged apnea. This can result from decreased respiratory drive related to COPD, central apnea, medullar infarct, and slow metabolism of succinylcholine, due to cholinesterase or pseudocholinesterase deficiency. Supportive therapy suffices to manage this complication, and subsequent sessions of ECT can be given with very low–dose succinylcholine or a nondepolarizing agent. Cholinesterase inhibitors used for the treatment of Alzheimer disease have not been associated with prolonged paralysis or apnea.

Central Nervous System

Central Nervous System Seizure induction results in a hypermetabolic state characterized by increased blood flow with significant elevation in oxygen and glucose consumption. Increased blood–brain barrier permeability has been hypothesized due to capillary hypertension. Postictally, metabolic suppression develops. These physiologic changes have not been demonstrated to produce structural brain changes or damage in well-designed and controlled animal and human studies or by brain imaging or analyses of cerebrospinal fluid. There is insufficient energy, even at maximal output of a modern brief pulse device, to cause brain tissue injury by temperature elevation.

ECT can be complicated by prolonged seizure duration, defined as greater than 3 minutes by motor or EEG measurement, in 1% to 2% of treatment sessions. The risk is greater in adolescents and young adults. There is a risk of structural brain injury along with cardiovascular and pulmonary complications if the seizure persists beyond 30 minutes, even when oxygenation is adequate. Termination at 2 to 3 minutes is undertaken using parenteral anticonvulsant medications, for example, parenteral benzodiazepines, propofol, or the short-acting barbiturate used for anesthesia. If this fails, a standard protocol for status epilepticus treatment should be instituted and the patient transferred to a setting for emergency treatment if not already in one. Patients who experience prolonged seizures need to be reassessed for underlying etiology. Prolonged seizures are more common in the presence of proconvulsant medications such as theophylline or lithium or when more than one seizure is induced in a single session. They are also associated with medical conditions that decrease seizure threshold, such as electrolyte abnormalities and epilepsy. Nonconvulsive status should be suspected if there is lack of recovery from anesthesia or abnormalities in mental status persist. This state may be detectable by ECT device EEG monitoring, but can be present when the typical 2-lead EEG does not exhibit abnormalities. Risks for NCSE include history of epilepsy, CNS pathology especially frontal lobe disorders, abrupt taper of antiepileptic medications, and medications that decrease seizure threshold. Treatment with IV benzodiazepines is generally the first approach. Tardive seizures are rare. The rate of onset of new seizure disorder after ECT does not exceed baseline rates in the general population. Because of its anticonvulsant properties, ECT can be given safely to patients with epilepsy.

Before the development of brain imaging, deaths and neurologic deterioration were reported when ECT was given to patients with occult brain tumors and brain metastases. Evaluation, including radiologic procedures, of patients with neurologic findings or abnormal mental status now most often leads to the identification and treatment of such diseases prior to any administration of ECT. It is possible for ECT to be given safely to patients with small, slow-growing tumors without associated increased intracranial pressure, such as meningiomas. ECT has also been given to patients with healed skull and brain trauma, including craniotomy, as long as any skull defect is avoided. Patients may also experience transient CNS symptoms of almost any variety, for example, various cortical sensory abnormalities.

Depression not uncommonly complicates CNS disease such as stroke and dementia. If patients with stroke are neurologically stable 30 days after a cerebrovascular event, ECT may be given. Attention should be paid to the treatment and stabilization of the underlying cause of the stroke as well as to the current neurologic status. Patients with dementia receive ECT for complicating depression, but are at risk for greater adverse cognitive effects, such as delirium or prolonged confusion. There is interest in using the treatment for agitation complication dementia, but little supporting, controlled evidence is currently available. Treatment using techniques focused on mitigating cognitive effects, such as right unilateral electrode placement, specific tailoring of dosage, brief or ultra-brief pulse waveform, and twice weekly frequency, are often employed. Signs and symptoms of CNS lesions or increased intracranial pressure, such as papilledema, focal neurologic deficits, and abnormal brain scan require careful consideration prior to prescription and administration of ECT.

Dental

Due to direct stimulation of jaw muscles during delivery of the electrical stimulus, patients bite down, sometimes strongly, and unstable teeth may be broken or dislodged completely. Also vulnerable are front teeth with crowns or other dental modifications such as laminates. The approach to this complication is routine prophylactic screening and may necessitate dental consultation and procedures prior to ECT. If teeth are injured during treatment, care must be taken to remove any broken fragments before respiration resumes, when they may obstruct the patient’s airway.

Musculoskeletal

With the consistent use of muscle relaxants, the most serious of these adverse effects is only of historical interest. Before the development of curare and later succinylcholine, fracture of the long bones or spine occurred in up to 10% of patients as a consequence of unmodified movements associated with seizure induction and expression. The modern patient, who receives enough muscle relaxant to modify movement rather than fully abolish it, has no significant musculoskeletal complications. Those with skeletal disease, for example, osteoporosis, are more fully paralyzed for treatment. Although this complication has been almost completely abolished, older adult patients remain at risk for falls during a course of ECT and appropriate fall precaution protocols should be used.

More common in the modern era and, ironically, often an adverse effect of depolarizing agents, is myalgia. These muscle pains are most often a consequence of the fasciculations produced during depolarization and are most prominent after the first treatment, but seldom persist after a few more sessions. Myalgia can be managed symptomatically with analgesics.

Those patients with temporomandibular joint problems may experience an exacerbation of pain, resulting from the unmodified contraction of jaw muscles during stimulus delivery. Spasm can also be induced. Prophylactic approaches may be advisable, such as the use of dental appliances already in the patient’s possession or the administration of prophylactic analgesics such as nonsteroidal anti-inflammatory medications. Symptomatic post-ECT approaches may also suffice.

Neurocognitive

Cognitive side effects are a limitation to the use of ECT, although in recent years the syndrome of postoperative cognitive dysfunction has been detected in patients undergoing medical procedures requiring either regional or general anesthesia with up to 10% to 20% incidence, particularly those over 60 years. With ECT, these effects have been best described for depressed populations. Most of the reports of cognitive function in schizophrenic and even autistic populations show stable cognitive function after a course of ECT. The phenomena are made more difficult to discriminate because depression itself is accompanied by impairments in attention, concentration, and learning. Post-ECT, with remission of depression, attention, and concentration are improved. Several factors influence the likelihood and extent to which impairments develop, but response to treatment is not a variable in the development of the cognitive adverse effects of ECT. Furthermore, functions such as reasoning, creativity, and the ability to form memories, are not affected in any enduring fashion. Nondeclarative memory, that is, procedural memory and priming, does not appear to be affected either. It is also important to note that objective measurement and subjective report of cognitive side effects do not correlate for many measures of subjective memory with the exception of global assessment.

The first factor in the development of adverse cognitive side effects is time point in treatment, which influences not only severity, but the type, of effect observed. Disorientation, diminished processing speed, decreased anterograde and retrograde memory, and errors in visual–spatial function and word finding are greatest immediately after a treatment session. These effects diminish fairly quickly once the treatment course ends. There is preliminary evidence that reactivating and reconsolidating a semantic memory may make it particularly vulnerable to impairment. Except for retrograde memory impairment, other cognitive effects generally return to pre-ECT baseline or even improve over baseline with right unilateral ECT. Diminished processing speed can persist when sine wave ECT has been administered. Retrograde memory improves more gradually, but spotty deficits may persist, with recent memories for public information more vulnerable than older memories for personal experiences. Retrograde memory deficits may persist several months after treatment ends. In a community setting, 12.4% of patients manifested marked and persistent retrograde amnesia, defined as deterioration in autobiographical memory of at least two standard deviations from pre-ECT baseline at 6 months post-ECT. The only significantly correlated technical factor was the number of bilateral ECT; female sex was the only demographic or clinical factor to achieve significance.

The likelihood of developing adverse cognitive effects is influenced by technical factors in the treatment: bilateral electrode placement; inefficient stimulus waveforms, most particularly, sine waveform, but including brief pulse; markedly suprathreshold dosing; greater frequency and number of treatments; and larger doses of anesthesia are more likely to result in deficits during the acute and subacute course of treatment. Bilateral electrode placement is associated with greater likelihood of persistent retrograde memory deficits long term. There is mixed data on the relative cognitive profiles of bifrontal versus bifrontotemporal electrode placement. In recent decades, sine waveform, which are no longer recommended, have been largely replaced by brief pulse and, most recently, ultrabrief pulse stimulation, and much has been learned about optimizing right unilateral electrode placement and restricting dosage while maintaining efficacy. The advantage of ultrabrief pulse stimulation over brief pulse stimulation in cognitive side-effect profile has become more apparent in the past decade of increasing use of ultrabrief pulse waveforms (Table 33.20a–3).

Patient characteristics and clinical factors also impact the frequency and severity of adverse cognitive effects. Those with baseline neurologic disease, MRI abnormalities, and baseline impairments in global cognitive functioning are more vulnerable to developing deficits, as well as older and female patients. The concurrent use of certain medications such as lithium or medications with anticholinergic effects also increases the risk of adverse cognitive effects. There have been no well-established pharmacologic approaches to attenuating adverse cognitive effects, although there is interest in using NMDA antagonists, such as ketamine anesthesia, antiglucocorticoid agents such as mifepristone, and thyroid hormone.

Table 33.20a–3.
Treatment Factors Influencing the Degree of Adverse Cognitive Effects From ECT

Treatment Factor	Technique Associated With Fewer Adverse Cognitive Effects
Electrode placement	Right unilateral
Stimulus waveform	Ultrabrief (and brief) pulse
Stimulus intensity	Threshold determination; dosage closer to threshold
Number of treatments	Fewest necessary to achieve remission or plateau of target symptoms
Frequency of treatment sessions	Less frequent, e.g., twice weekly for acute treatment course
Number of seizures per session	One
Concomitant psychotropic medications	Discontinue lithium, sedatives; moderate dose of neuroleptics
Anesthetic dosage	Adjust dosage to produce lighter anesthesia; use ketamine

Perhaps because subjective memory correlates best with mood, and the majority of patients respond to ECT in systematic research, most patients studied in these settings report improved cognitive function post-ECT. It has been difficult to distinguish these patients from controls in the subjective assessment of memory months after ECT in randomized controlled trials. In community settings, which are characterized by inclusion of a broader range of patients and ECT techniques, subjective complaints of memory impairment are more frequent and correlate to the number of bilateral ECT received. Nonetheless, generally favorable attitudes are expressed by patients who have received ECT, especially compared to attitudes of the general public, and the majority of patients would have treatment again if necessary.

Other Adverse Effects

Postictal agitation has an undefined etiology, but can be associated with various factors, including electrode placement, electrical dose over seizure threshold, seizure duration, anesthetic dose, concomitant use of medications, and anxiety or agitation pre-ECT. Although ranging in severity, the more severe forms of agitation require emergent management with decreased stimulation, medications such as short-acting benzodiazepines, and ECT anesthetics, and even restraints. Prophylaxis may include switching electrode placement to unilateral, attention to electrical dosing, increasing the dose of anesthetic agent pre- and post-ECT or switching to propofol, modifying the dose of succinylcholine to prevent awakening while paralyzed or to decrease lactate generation, and alteration of medication regimens to address agitation by the use of neuroleptics. Lithium and carbidopa are associated with increased frequency of postictal agitation and should be evaluated for discontinuation or reduced dose.

ECT may be associated with a number of less serious, although nonetheless, uncomfortable side effects as well. Headaches and nausea can occur postictally. The etiology of postictal headache is unknown, and ascertaining cause is complicated by the presence or withdrawal of medications and the fact that anxiety often accompanies target illnesses for which ECT is used, such as depression. The throbbing quality of the headache suggests a vascular component, and migraineurs certainly seem more vulnerable to developing this side effect. Symptomatic management with analgesics, including nonsteroidal anti-inflammatory drugs, is often sufficient, but use of these agents prophylactically, for example, IV ketorolac, may be required. Migraine sufferers may need to use migraine specific treatments.

Nausea may accompany headache or occur alone in a minority of patients. Again the etiology is complex, including the use of anesthetic agents. Antiemetics, including $5HT_3$ antagonists, such as ondansetron or dopamine antagonists such as compazine, can be employed symptomatically. Change of anesthetic to propofol may also be effective.

FUTURE DIRECTIONS

There is interest in continued refinements of ECT technique. Common themes in these approaches are as follows: focusing the treatment spatially to optimize dosing in brain areas associated with putative neural networks involved in depression and other psychopathologies that are indications for ECT and to diminish dosing in areas associated with adverse cognitive effects and improving the efficiency of a noninvasive electrical stimulus in direction and amplitude, even to a subconvulsive level. Methodologies such as magnetic induction of convulsive treatment and focal electrically assisted seizure therapy are examples of innovation designs under investigation. This research is parallel to investigations in magnetic stimulation, for example, rTMS, and to the renaissance of invasive electrical techniques, for example, vagal nerve stimulation and deep brain stimulation.

FURTHER READINGS

American Psychiatric Association. *The Practice of Electroconvulsive Therapy: Recommendations for Treatment, Training, and Privileging—A Task Force Report*. 2nd ed. American Psychiatric Press; 2001.

Bergsholm P. Patients' perspectives on electroconvulsive therapy: a reevaluation of the review by Rose et al on memory loss after electroconvulsive therapy. *J ECT*. 2012;28(1):27–30.

Bouckaert F, Sienaert P, Obbels J, et al. ECT: its brain enabling effects: a review of electroconvulsive therapy-induced structural brain plasticity. *J ECT*. 2014;30(2):143–151.

Boylan LS, Haskett RF, Mulsant BH, et al. Determinants of seizure threshold in ECT: benzodiazepine use, anesthetic dosage, and other factors. *J ECT*. 2000;16(1):3–18.

Ghaziuddin N, Kutcher SP, Knapp P, et al; Work Group on Quality Issues, AACAP. Practice parameter for use of electroconvulsive therapy with adolescents. *J Am Acad Child Adolesc Psychiatry*. 2004;43(12):1521–1539.

Haskett RF, Loo C. Adjunctive psychotropic medications during electroconvulsive therapy in the treatment of depression, mania, and schizophrenia. *J ECT*. 2010;26(3):196–201.

*Kellner CH, Knapp RG, Petrides G, et al. Continuation electroconvulsive therapy vs pharmacotherapy for relapse prevention in major depression: a multisite study from the consortium for research in electroconvulsive therapy (CORE). *Arch Gen Psychiatry*. 2006;63(12):1337–1344.

*Kellner CH, Knapp R, Husain M, et al. Bifrontal, bitemporal, and right unilateral electrode placement in ECT: a randomised trial. *B J Psychiatry*. 2010;196:226–234.

Mayur P, Byth K, Harris A. Autobiographical and subjective memory with right unilateral high-dose 0.3 millisecond ultrabrief-pulse and 1 millisecond brief-pulse electroconvulsive therapy: a double-blind randomized controlled trial. *J ECT*. 2013;29(4):277–282.

McCall WV, Reboussin D, Prudic J, et al. Poor health-related quality of life prior to ECT in depressed patients normalizes with sustained remission after ECT. *J Affect Disord*. 2013;147(1–3):107–111.

McCall WV, Reboussin DM, Weiner RD, Sackeim HA. Titrated moderately suprathreshold vs fixed high-dose right unilateral electroconvulsive therapy: acute antidepressant and cognitive effects. *Arch Gen Psychiatry*. 2000;57(5):438–444.

McClintock SM, Choi J, Deng ZD, Appelbaum LG, Krystal AD, Lisanby SH. Multifactorial determinants of the neurocognitive effects of electroconvulsive therapy. *J ECT*. 2014;30(2):165–176.

Østergaard SD, Bolwig TG, Petrides G. No causal association between electroconvulsive therapy and death: a summary of a report from the Danish Health and Medicines Authority covering 99,728 treatments. *J ECT*. 2014;30(4):263–264.

Perera TD, Luber B, Nobler MS, Prudic J, Anderson C, Sackeim HA. Seizure expression during electroconvulsive therapy: relationships with clinical outcome and cognitive side effects. *Neuropsychopharmacology*. 2004;29(4):813–825.

Petrides G, Malur C, Braga RJ, et al. Electroconvulsive therapy augmentation in clozapine-resistant schizophrenia: a prospective, randomized study. *Am J Psychiatry*. 2015;172(1):52–58.

*Prudic J, Haskett RF, McCall WV, et al. Pharmacological strategies in the prevention of relapse after electroconvulsive therapy. *J ECT*. 2013;29(1):3–12.

Prudic J. Strategies to minimize cognitive side effects with ECT: aspects of ECT technique. *J ECT*. 2008;24(1):46–51.

Prudic J, Olfson M, Marcus SC, Fuller RB, Sackeim HA. Effectiveness of electroconvulsive therapy in community settings. *Biol Psychiatry*. 2004;55(3):301–312.

Rasmussen KG. Do patients with personality disorders respond differentially to electroconvulsive therapy? A review of the literature and consideration of conceptual issues. *J ECT*. 2015;31(1):6–12.

Sackeim HA, Haskett RF, Mulsant BH, et al. Continuation pharmacotherapy in the prevention of relapse following electroconvulsive therapy: a randomized controlled trial. *JAMA*. 2001;285(10):1299–1307.

Sackeim HA, Long J, Luber B, et al. Physical properties and quantifications of the ECT stimulus: I. Basic principles. *Convuls Ther*. 1994;10(2):93–123.

*Sackeim HA, Prudic J, Devanand DP, et al. A prospective, randomized, double-blind comparison of bilateral and right unilateral electroconvulsive therapy at different stimulus intensities. *Arch Gen Psychiatry*. 2000;57(5):425–434.

Sackeim HA, Prudic J, Fuller R, Keilp J, Lavori PW, Olfson M. The cognitive effects of electroconvulsive therapy in community settings. *Neuropsychopharmacology*. 2007;32(1):244–254.

*Sackeim HA, Prudic J, Nobler MS, et al. Effects of pulse width and electrode placement on the efficacy and cognitive effects of electroconvulsive therapy. *Brain Stimul*. 2008;1:71–83.

Schoeyen HK, Kessler U, Andreassen OA, et al. Treatment resistant bipolar depression: a randomized controlled trial of electroconvulsive therapy versus algorithm-based pharmacological treatment. *Am J Psychiatry*. 2015;172(1):41–51.

Sundsted KK, Burton MC, Shah R, Lapid MI. Preanasthesia medical evaluation for electroconvulsive therapy: a review of the literature. *J ECT*. 2014;30(1):35–42.

*Tharyan P, Adams CE. Electroconvulsive therapy for schizophrenia. *Cochrane Database Syst Rev*. 2002;CD000076.

Verwijk E, Comijs HC, Kok RM, Spaans HP, Stek ML, Scherder EJ. Neurocognitive effects after brief pulse and ultrabrief pulse unilateral electroconvulsive therapy for major depression: a review. *J Affect Disord*. 2012;140:233–243.

Zhuo C, Yu C. Functional neuroimaging changes subsequent to electroconvulsive therapy in unipolar depression: a review of the literature. *J ECT*. 2014;30(4):265–274.

33.20b Brain Stimulation Methods

STEFAN B. ROWNY, M.D., M.F.A., AND LAWRENCE AMSEL, M.D., M.P.H.

Brain stimulation in psychiatric practice and research uses electric current, or changing magnetic field, to induce currents within the brain to alter neuronal firing. There is a growing list of tools capable of eliciting such neuromodulation, each with a different spectrum of action and therapeutic utility in psychiatric disorders. These tools either apply electrical or magnetic fields transcranially or involve the surgical implantation of electrodes to deliver electrical currents to a cranial nerve or to the brain tissue directly. The transcranial techniques include cranial electrical stimulation (CES), electroconvulsive therapy (ECT), transcranial direct current stimulation (tDCS, also called direct current (DC) polarization), transcranial magnetic stimulation (TMS), and magnetic seizure therapy (MST). The surgical techniques include implanted cortical stimulator (ICS), deep brain stimulation (DBS), and vagus nerve stimulation (VNS). These methods are briefly described in Table 33.20b–1. ECT and DBS will be discussed in separate chapters.

Table 33.20b–1.
Brain Stimulation Methods: Definitions

CES	**Cranial Electrical Stimulation** Weak alternating electrical current is applied via saline-soaked felt-covered electrodes, usually to the earlobes or bifrontal scalp location.
cTMS	**Controllable pulse shape Transcranial Magnetic Stimulation (TMS)** A transcranial magnetic stimulator with a modified circuit that can induce nearly rectangular pulses, in contrast to conventional TMS that induces cosine waveforms.
DBS	**Deep Brain Stimulation** A wire electrode with multiple stimulating sites that is surgically implanted into the brain and chronically administers electrical stimulation.
ECT	**Electroconvulsive Therapy** Electricity is applied transcranially via electrodes placed on the scalp to induce a seizure under anesthesia.
MST	**Magnetic Seizure Therapy** High intensities of TMS are applied to induce focal seizures under anesthesia, as a less invasive form of ECT.
PAS	**Paired Associative Stimulation** TMS is paired with peripheral nerve stimulation, resulting in Hebbian plasticity, altering functioning of the corticospinal tract.
tDCS	**Transcranial Direct Current Stimulation, also called DC Polarization** Weak direct electrical current is applied to the scalp via sponge electrodes to polarize underlying brain tissue.
TBS	**Theta Burst Stimulation** High-frequency bursts (e.g., 3 pulses at 50 Hz) of TMS are applied at a regular frequency (e.g., 5

	bursts per second). TBS can be given continuously (cTBS) or intermittently (iTBS).
TES	**Transcranial Electrical Stimulation** Alternating electrical currents applied to the scalp.
TMS	**Transcranial Magnetic Stimulation** Rapidly alternating magnetic fields are applied to the scalp to induce small, focused electrical currents in superficial cortex. Remote sites may be activated trans-synaptically.
TNS	**Trigeminal Nerve Stimulation** Mild electrical current stimulate branches of the largest cranial nerve—the trigeminal nerve—either externally or subcutaneously aimed at modulating activity in brain areas such as the anterior cingulate cortex, nucleus solitarius, locus coeruleus, vagus nerve, and the cerebral cortex.
VNS	**Vagus Nerve Stimulation** An implanted pacemaker administers electrical pulses to the vagus nerve in the neck, activating the vagal afferents.

The historical roots of modern brain stimulation in psychiatry can be traced to the introduction of ECT in 1938 by Ugo Cerletti and Lucio Bini. Since then, electrical methods for neuromodulation have been in continuous use in psychiatry. ECT, which uses alternating electrical current applied to the scalp to induce a seizure, has been substantially improved over the intervening decades, both in refinements to the electrical stimulus characteristics and the placement of electrodes, as well as in advancements to the accompanying anesthesia, particularly with the use of muscle relaxants (e.g., succinylcholine, atracurium). In 1949, another alternating current brain stimulation method, CES, was developed in the former Soviet Union. It was originally called “Electrosleep” and, as the name suggests, was initially used to treat insomnia. This method uses much lower levels of alternating current than ECT and is thus subconvulsive. In 1985, nearly 50 years after the first use of ECT, Anthony and colleagues published the first use of pulsed magnetic fields to stimulate the brain with a procedure called TMS. TMS was initially used in neurology for studies of nerve conduction but quickly caught the attention of psychiatrists eager to explore other, less invasive alternatives to ECT. This nonconvulsive, well-tolerated stimulation method through TMS has been approved by the FDA for the treatment of depression. It also continues to be under active study, with some promising results in the treatment of various other psychiatric disorders like bipolar disorder, post-traumatic stress disorder (PTSD), obsessive-compulsive disorder (OCD), schizophrenia (as described by Lefauchear and colleagues in 2014), addiction disorders, particularly nicotine dependence, and even depersonalization disorder. MST is also a convulsive treatment in which the seizure is induced by the application of rapidly alternating powerful magnetic fields. It was first tested in an animal in 1998 and in a human in 2000. Since then, MST has been under investigation in nonhuman primates and in humans as a more focal means of inducing seizures in an attempt to retain the unparalleled efficacy of ECT but with fewer side effects.

Two more invasive brain stimulation methods, DBS and VNS, were introduced about a decade following the first trials of TMS. Both DBS and VNS were first approved by the U.S. Food and Drug Administration (FDA) in 1997 with indications for the treatment of neurologic conditions: VNS was originally approved as adjunctive therapy in reducing the frequency of epileptic seizures in adults and children aged 12 years and over; while DBS was initially approved for the treatment of essential tremor and Parkinson tremor. Currently, indications for DBS include treatment of all symptoms of Parkinson disease, including tremor, slowness, stiffness, involuntary movements induced by medications, as well as symptoms of OCD that are not adequately controlled with medications or other therapies. At present, indications for VNS include adult treatment-resistant depression and epilepsy.

Frequently, initial developments in brain stimulation have come from neurology. For example, TMS, DBS, and VNS originated as tools to treat neurologic conditions, but psychiatrists quickly saw the potential of those tools in the treatment of psychiatric conditions. It was not clear, however, whether those technologies and doses designed to treat neurologic indications could be directly coopted for the purposes of treating mental illness. Today, brain stimulation techniques in psychiatry are undergoing considerable research to answer those questions and many others that have emerged, and stimulation technologies are gaining momentum in psychiatry as they continue to evolve.

THERAPEUTIC NEUROMODULATION: TREATING PSYCHIATRIC DISORDERS THROUGH BRAIN STIMULATION

Mechanism of Action

Electrical Stimulation—Common Pathway. The brain stimulation tools considered in this chapter generate either electrical or magnetic pulses. However, both of these share a common final pathway—they affect the neurons electrically. That electrical effect may either be through the direct application of electricity and the resultant electrical field that reaches neurons or through the indirect induction of electricity via magnetic stimulation. The direct forms of electrical stimulation are exemplified in either *transcranial* delivery, as with ECT, CES, TNS, and tDCS, or surgical *intracerebral* delivery, as in the case of DBS, or direct cortical stimulation (epidural or subdural). The indirect forms of electrical stimulation include TMS and MST, which induce electric fields in the brain through the application of alternating magnetic fields (Fig. 33.20b–1). Of note, both the epidural and intracerebral modalities are more localized, referred to as focal, than the transcranial application of electricity because electrodes are placed directly in the neuronal tissue, bypassing the impedance of the scalp and skull. Magnetic stimulation methods also bypass the impedance of the scalp and skull. Hence they are also more focal when compared to the transcranial direct application of electricity. Magnetic stimulation is considered an indirect method of electrical brain stimulation in that changing the magnetic field from these devices induce electric currents in the brain according to the principle first described by Faraday in a law that bears his name and later incorporated into Maxwell’s equations, which unify all of electromagnetism. The magnetic modalities achieve their enhanced spatial resolution (focality) noninvasively, in contrast to the intracerebral and epidural methods. Therefore these modalities are at the center of intensive research as they offer a promise of an unparalleled degree of spatial specificity without the need for surgery.

All but one of the brain stimulation modalities described here act by stimulating neurons; the one exception is tDCS, which does not stimulate but rather it polarizes. In this sense, the “S” of tDCS is a misnomer. It is more accurate to conceptualize tDCS as exerting a polarizing effect aimed at altering the likelihood of neuronal firing.

The action of the subconvulsive modalities of stimulation relies on the effects of the repeated stimulation of the targeted neural circuitry. In contrast, in the case of the convulsive modalities (ECT and MST), the therapeutic action is dependent upon the seizure induced by the stimulation, and the subsequent effects of repeated seizures on brain processes. This is not to say that the form of stimulation that triggers the seizure has no impact on outcome itself (see later discussions of nonconvulsive electric stimulation). Indeed, it is well replicated that electrode placement and electrical stimulus parameters have a profound impact on the efficacy and side effects of ECT. Whether the same will be true for MST is under active investigation. Importantly, MST has the capability of inducing seizure in a more localized manner and thus targeting specific brain regions and their neuronal networks thought to be implicated in depression, while avoiding those responsible for untoward side effects.

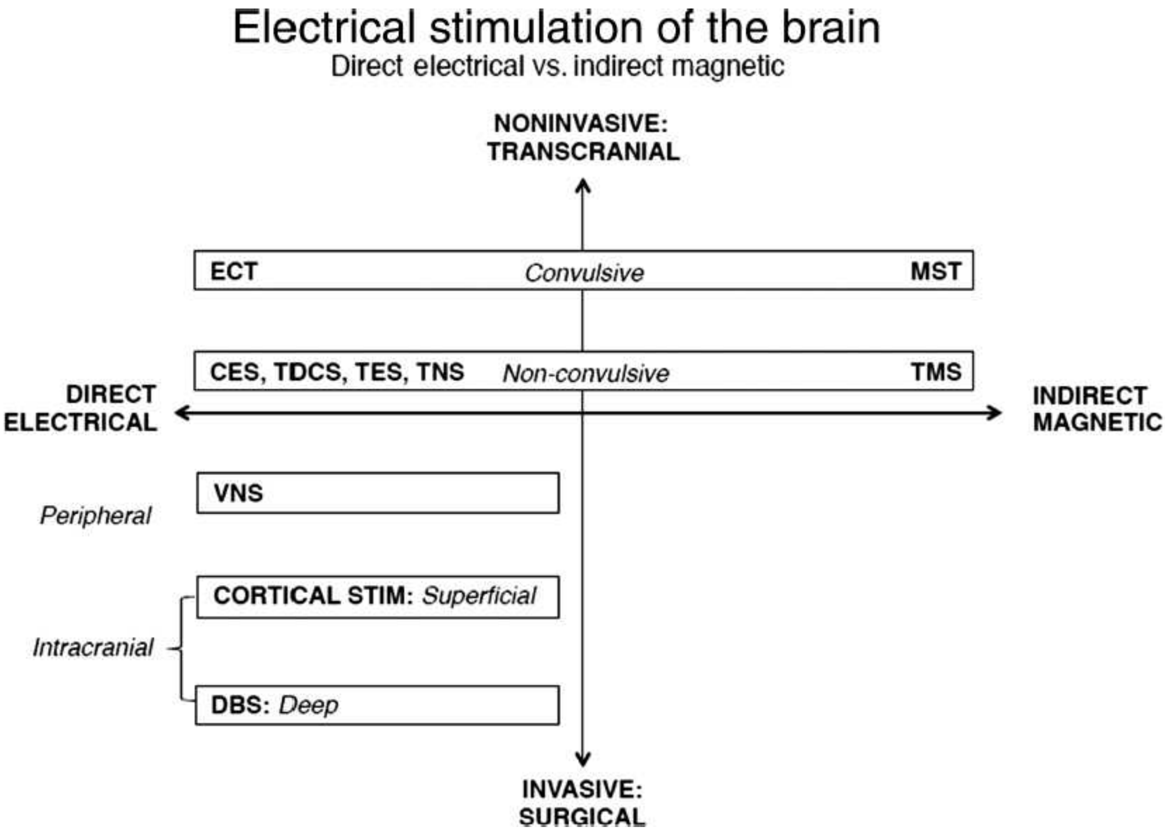
Electrical stimulation of the brain
Direct electrical vs. indirect magnetic
NONINVASIVE: TRANSCRANIAL
ECT
Convulsive
MST
CES, TDCS, TES, TNS
Non-convulsive
TMS
DIRECT ELECTRICAL
INDIRECT MAGNETIC
Peripheral
VNS
Intracranial
CORTICAL STIM: Superficial
DBS: Deep
INVASIVE: SURGICAL

FIGURE 33.20b–1. Classification of tools that act to stimulate the brain electrically using direct (electrical) or indirect (magnetic) means.

Comparison of Brain Stimulation With Psychopharmacology. Psychopharmacology and brain stimulation differ in several clinically relevant respects. Pharmacologic agents rely on molecular action at synapses or at intracellular targets. They bind to either presynaptic or postsynaptic receptors or to intracellular targets in order to affect their function. In contrast, brain stimulation induces action potentials via electrical depolarization of axons or cell bodies independent of receptor activity. Thus, not only does brain stimulation have the potential to work when medications have failed, but it may also act synergistically with medications, as their respective sites and mechanisms of action are different. Because the chief mechanism of action in all modalities of brain stimulation is direct electrical modulation of the brain's neuronal activity, this approach avoids the systemic side effects seen with medications which are orally introduced exogenous compounds distributed throughout the body to sites that have little to do with the intended target. These compounds may be secondarily metabolized to other biologically active metabolites. In contrast, brain stimulation involves no exogenous compounds that circulate systemically in the bloodstream, does not undergo metabolism and clearance, is not protein bound, and has no known interactions with the metabolism or clearance of pharmacologic agents.

Brain stimulation is applied in a phasic fashion, which simply means that in contrast to medications that can remain biologically available in a patient's body for hours to weeks, or even months in the case of injectable depots, electricity can be turned on or off, with a flip of a switch. Most of the contemporary brain stimulation modalities that administer alternating electric current—either directly or via electromagnetic conduction (e.g., ECT, MST, DBS, VNS)—deliver pulses, or collections of pulses otherwise referred to as trains of pulses, with each pulse typically lasting 1 ms or less and trains of pulses lasting up to 30 seconds. The phasic nature of brain stimulation provides immediate washout that is not possible for pharmacologic agents, which must be first metabolized and then cleared. While the transient on–off nature of brain stimulation has proved advantageous in research probing functional roles of certain brain areas, it poses a challenge to the induction of lasting effects that will persist beyond the period of the therapeutic stimulation (see Table 33.20b–2).

Table 33.20b–2.
Brain Stimulation Versus Psychopharmacologic Therapy

	Brain Stimulation	Psychopharmacology
Therapeutic Agent	Electrical or electromagnetic	Neurochemical
Temporal characteristics	Phasic	Tonic
Delivery	Verifiable on electronic or clinical record	Dependent on patient compliance
Site of action	Axonal depolarization	Molecular target
Distribution	Focally applied to the brain	Systemic, throughout the body
Duration of availability	Milliseconds to seconds	Typical half-life of hours to weeks
Dosing parameters	Intensity, frequency, intertrain interval, number of pulses, site of stimulation	Milligrams, blood level
Metabolism	None	Hepatic
Clearance	None	Renal
Interactions with metabolism of other agents	None known	Possible

Table 33.20b–3.
Acute and Prolonged Mechanisms of Action

Acute effects
- Phasic activation of neural circuits
- Observable motor responses (e.g., twitch)
- Temporary disruption (e.g., speech arrest) or facilitation of ongoing processing (e.g., speeds reaction time)

Prolonged effects
- Neuroplasticity
  - Changes in synaptic efficacy akin to LTP or depression
  - Alteration in neurotropic factors
  - Modulation of cortical excitability
  - Modulation of functional connectivity

Acute Versus Prolonged Effects. Brain stimulation can have immediate or lasting effects (see Table 33.20b–3). A single electrical pulse delivered at sufficient intensity can induce depolarization, trigger an action potential, release neurotransmitters at the synapse, and result in trans-synaptic propagation with subsequent activation of a functional circuit. For example, brain stimulation applied to the hand area of the primary motor cortex may activate the corticospinal tract and induce a motor-evoked potential in the contralateral hand. Such stimulation can result acutely in the induction of either a positive effect, as in the case of a muscle twitch or visualization of phosphenes, or a disruptive effect as in the case of visual masking. Although inducing a muscle twitch is and immediate, short-lasting effect, it allows one to examine the latency of this effect from the time of the TMS single-pulse stimulation. One can thus determine the nerve conduction delay, a finding which may have value in the study of central motor pathways in clinical cases such multiple sclerosis and spinal injury.

Repetitive pulses delivered at fixed frequencies can exert even more powerful effects. Epstein and colleagues described in 1999 how repetitive TMS (rTMS) applied to the language-dominant hemisphere-induced arrest of speech. Following termination of the stimulation, speech returned to normal. More recently, in a study with cats, researchers used interleaved functional neuroimaging with rTMS to examine network-wide effects, and demonstrated behavioral impairments to visual attention with rTMS over the left intraparietal sulcus. Interestingly, they also found downstream adaptive compensatory activity in areas of the visual attention network that were not stimulated.

Some of the more invasive brain stimulation modalities, such as DBS or VNS, are programmed to operate chronically, thus extending the acute action for as long as the stimulation is turned on. In the case of DBS, the pulses are typically given continuously at a high frequency, whereas in the case of VNS the pulses are given in trains lasting up to 30 seconds and typically repeated every 5 minutes. The less invasive modalities, such as rTMS, tDCS, CES, and even ECT, may require the induction of some form of neuroplasticity for their effect to become lasting.

Neuroplasticity. Persistent activation of a brain circuit can induce prolonged changes in the subsequent activities of that circuit. For example, repeated stimulation of the prefrontal pathway is implicated in spatial memory learning and consolidation at a high electrical frequency; the stimulation induces a lasting increase in the efficacy of that circuit; conversely, repeated stimulation with low frequencies depresses it. These phenomena are termed long-term potentiation (LTP) and long-term depression (LTD), respectively, and are collectively known as neuroplasticity. Neuroplasticity is thought to occur through dynamic alterations in the synaptic architecture and efficacy. Establishing evidence of longer-lasting neuronal effects (i.e., LTP or LTD) from brain stimulation is crucial for establishing both the clinical utility and for understanding the mechanism of action. Recent research in animal models offers evidence of the lasting effects of high-frequency rTMS, and there is evidence emerging in humans that rTMS can produce changes in synaptic efficacy that outlast stimulation. Initially described by Wang and colleagues, low-frequency rTMS (1 Hz or less) is believed to induce a sustained inhibitory effect on the excitability of the motor cortex, whereas high-frequency rTMS (5 Hz or greater) produces facilitatory after-effects in excitability of the motor cortex. Esser and colleagues coupled TMS with topographical EEG measurements of TMS-evoked potentials. This work provided the first evidence that the changes in excitability may come about as a result of an enhancement of a cortical (voltage) potential. While such in vivo evidence is consistent with the hypothesis that TMS causes LTP (Esser et al., 2006) without ex vivo experiments on excised human tissue, it is difficult to say with complete certainty whether these observations are true measurements of synaptic plasticity.

Effect of Illness and Early and Late Brain Development on Brain's Response to Stimulation

Because brain stimulation exerts its acute effects by depolarization of neurons and over the longer term possibly through induction of lasting changes in synaptic efficacy, factors that influence pretreatment neural transmission and plasticity may influence the brain's response to therapeutic stimulation. For example, investigators have found that infants and children have higher thresholds for activation of functional circuits in response to single pulses of TMS, presumably because the myelination in their brains has not yet been completed. Diseases of the central nervous system can also impact the response to stimulation. Fitzgerald and colleagues (Fitzgerald et al., 2004) reported that patients with schizophrenia exhibited reduced plasticity in brain response to TMS when compared to healthy controls. Specifically, the patients failed to show inhibitory effects of low-frequency rTMS over motor cortex, possibly reflecting abnormalities in glutamatergic functioning in schizophrenia.

Cortical atrophy due to normal aging or disease-related can reduce the total dosage of magnetic stimulation delivered because the strength of the magnetic field drops rapidly as a function of distance from the stimulating coil. Aging can also alter the brain's response to stimulation in a manner independent of the distance issue. Some investigators have demonstrated an age-related differential response to rTMS over the motor cortex in older adults compared to younger subjects (45 years delineation of matched subjects), suggesting normal age-related changes in synaptic plasticity and the capacity for LTP and LTD that may reduce the efficacy of rTMS. These findings are in line with the generally accepted view that the efficacy of antidepressant pharmacotherapy is also highly age dependent.

Relationship Between Spatial Resolution and Invasiveness in Brain Stimulation

Each of the brain stimulation techniques can be separately classified with respect to the degree to which it can achieve precise focal targeting, that is, focality, and with respect to the degree of invasiveness, as seen in Figure 33.20b–2. A high level of focality is important when brain stimulation is a tool being used to map regional brain functions and is also advantageous in clinical practice provided that a specific target is known, or if stimulation of spatially adjacent regions might be counter therapeutic. On the other hand, if the target involves broad areas or is not precisely defined, then narrow focality may not offer benefits.

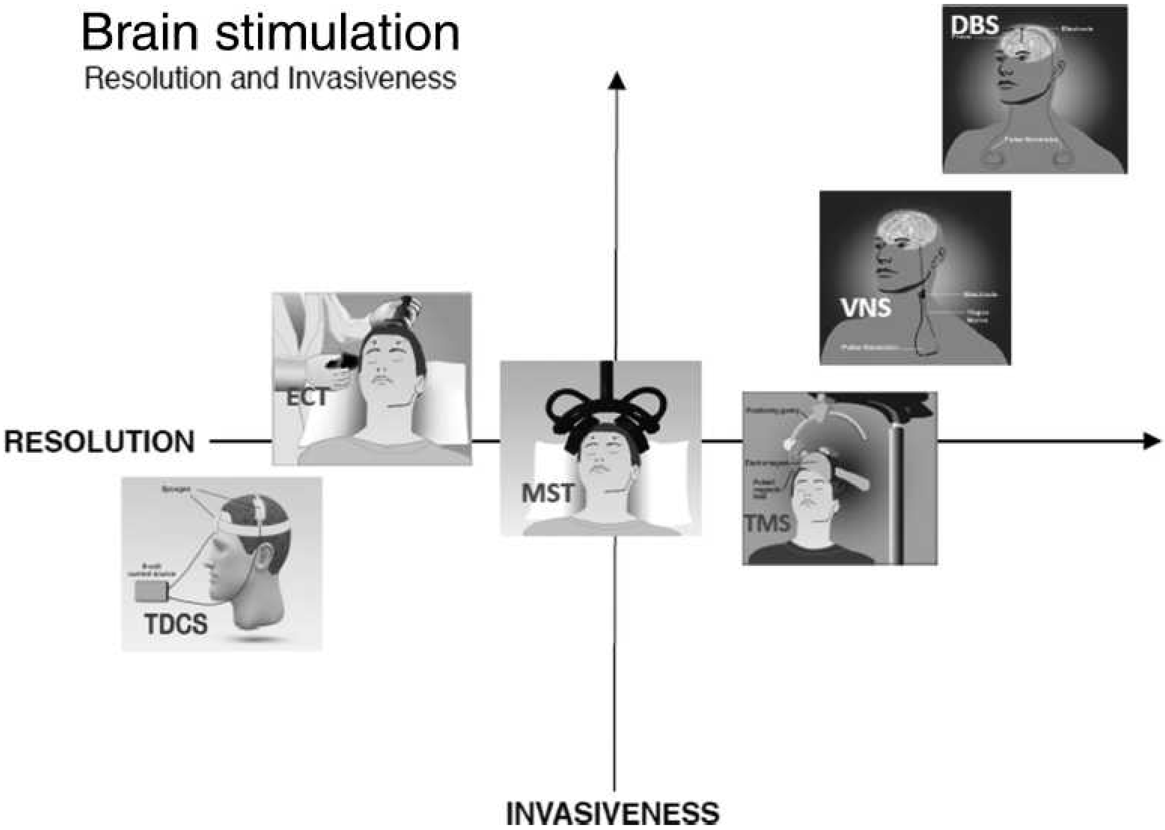

FIGURE 33.20b–2. Relationship between spatial resolution and invasiveness of the brain stimulation tools. DBS, deep brain stimulation; ECT, electroconvulsive therapy; MST, magnetic seizure therapy; tDCS, transcranial direct current stimulation; VNS, vagus nerve stimulation. (From National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services.)

The least focal modalities are the noninvasive electrical stimulation methods such as tDCS, CES, and ECT. While ECT is more invasive by virtue of the need to use general anesthesia during the procedure and due to the induction of seizures, it is also nonfocal for the reasons it shares with the previous two methods—the electric current needs to traverse through the scalp and skull, which considerably impede the electric current, and which vary substantially from one individual to another.

The most focal or precisely targeted method is DBS, where a thin lead usually containing a number of wires connecting electrodes placed close to its tip, is implanted directly into the brain region being targeted. However, it is also the most invasive of all methods, requiring neurosurgical insertion of the stimulator lead, usually bilaterally, via a burr hole for each of the leads approximately the size of a quarter. Somewhat less invasive but also less focal is VNS, which stimulates the vagal afferents on their way to the brain. The magnetic modalities like TMS and MST are considered more focal because the magnetic field passes through the scalp and skull without impedance, in contrast to stimulation methods that directly administer electrical currents to the scalp. Magnetic flux has limited penetration (approximately between 2 and 4 cm), which is an important aspect of the stimulation's spatial resolution.

PSYCHIATRIC APPLICATIONS OF BRAIN STIMULATION

Dual Roles for Brain Stimulation: As a Probe of Brain Function in Health and Disease and as a Clinical Treatment of Psychiatric Disorders

Brain stimulation methods are used in psychiatric clinical and research practice along several parallel pathways. By virtue of their noninvasive nature, brain stimulation tools offer unique versatility as tools to examine normal brain function. By transiently perturbing brain activity, it creates temporary "virtual lesions" and allows for the monitoring of the subsequent effects on cognition and behavior. Such investigation is particularly valuable for probing brain areas implicated in the pathophysiology of psychiatric illness. Brain stimulation modalities are also used in the treatment of mental illness as novel therapeutic agents, as shown in Figure 33.20b–3.

Brain stimulation actively changes brain function and can induce measurable effects on brain activity that can be used to test the causal relationship between the brain and behavior. This is in contrast to neuroimaging or other methods of studying brain function that passively acquires information about brain function and can only provide correlates of behavior. To demonstrate the utility of brain stimulation in testing the causal relationship between the brain and behavior, Pascual-Leone and colleagues in 1991 created a virtual lesion in Broca area with rTMS, which consequently produced speech arrest in healthy participants. More recently, using brain stimulation to investigate pathology, investigators compared the impact of rTMS on participants with autism spectrum disorder (ASD) to healthy matched volunteers to demonstrate differential effects of stimulation. These variances suggest language-related neural network differences in those with ASD compared to neurotypical individuals.

In addition to examining brain–behavior relationships, the various stimulation modalities have the potential for delivering a therapeutic effect in the treatment of psychiatric disorders, primarily because brain stimulation modulates neuronal activity in a way that may be lasting, as reviewed in other parts of this chapter. The focal and temporal precision of some forms of brain stimulation poses a challenge in defining the disease-relevant anatomical and physiologic targets of psychiatric disorders. Improving the therapeutic use of neuromodulation in psychiatry will also require knowledge of the focal and distributed effects of the various stimulation paradigms.

Integrating Brain Stimulation With Neuroimaging and Neurophysiology

The combination of brain stimulation and functional neuroimaging is a powerful research paradigm for exploring and understanding neural circuitry. TMS, DBS, and ECT have been successfully coupled with positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and electroencephalography (EEG) to investigate the modulatory effects of brain stimulation in both healthy and pathologic populations. Commonly, physiologic measures are performed sequentially *offline* before and after brain stimulation to assess modulatory effects. However, for brain stimulation modalities such as TMS and tDCS, a simultaneous or *online approach* is now possible, whereby stimulation and imaging occur in an interleaved manner. Whereas functional neuroimaging alone can indicate correlations between brain and behavior, simultaneous neuroimaging and brain stimulation can indicate causal brain–behavior relations.

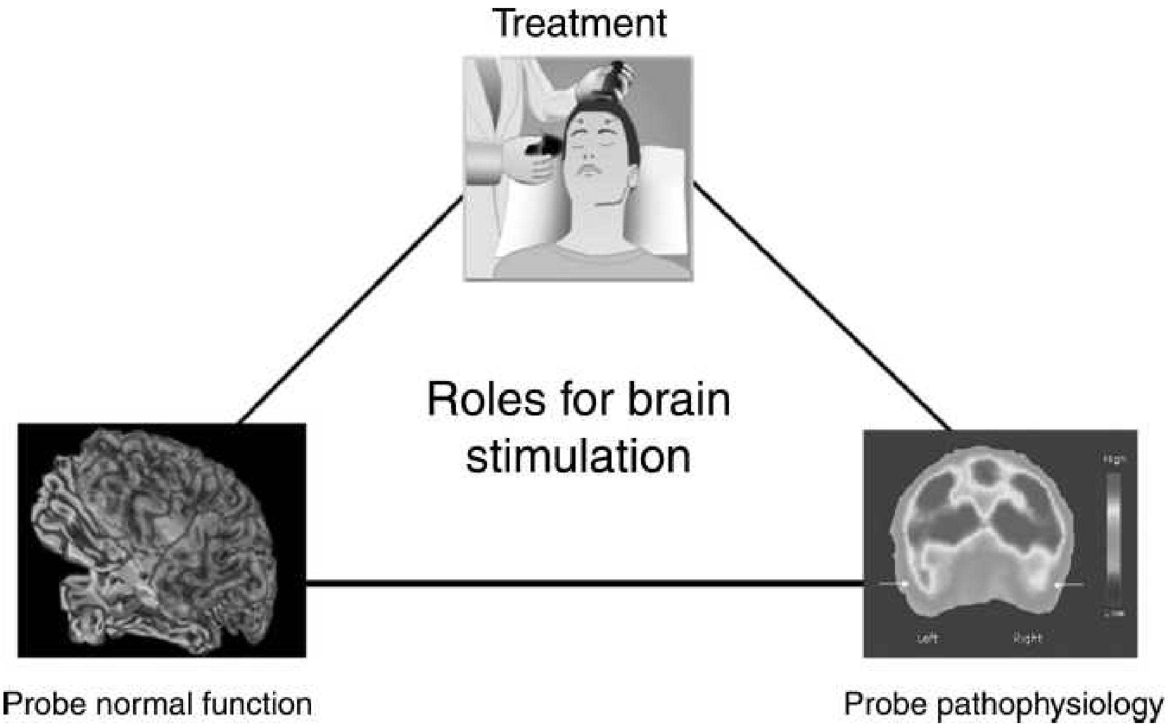

FIGURE 33.20b–3. Research and clinical applications of brain stimulation in psychiatry. (From National Institute of Mental Health, National Institutes of Health, Department of Health and Human services.)

With psychopathology increasingly conceptualized as dysfunction not of one isolated brain area but rather a dysfunction of a network of brain areas, the combination of brain stimulation and neuroimaging is useful for examining the dynamics of these brain networks and the functional roles of network components, which may be target sites for therapeutic stimulation. This is achieved by using brain stimulation to perturb or enhance one node of a network while monitoring activity in the entire brain. Results from such research elucidate the functional connectivity of brain networks. For example, researchers have found that stimulation with TMS of the angular gyrus resulted in modulated activity in remotely connected nodes of the visuospatial attention network (enhanced activity in the contralateral V1 and V2), in addition to the behavioral effects of TMS. In another example, investigators delivered low-frequency rTMS to the intraparietal sulcus, followed immediately by an attention fMRI task to demonstrate a decrement in activity in the targeted region and changes in activation of the corresponding contralateral region, as well as other regions in the visual attention network. In this experiment, rTMS appears to have produced changes in activity not only at the site of stimulation but also disrupted the interhemispheric balance of activity.

Interleaved brain stimulation and neuroimaging are also useful in guiding and localizing stimulation with much-improved accuracy rather than relying on gross structural anatomy as reference, which is susceptible to interindividual variability. In turn, accurately navigated TMS may reduce variability in clinical response.

These examples illustrate how combining brain stimulation with neuroimaging and physiologic measures can inform the exploration of mechanisms of action of various brain stimulation modalities. This can lead to effective screening of stimulation paradigms hypothesized to have therapeutic effects based on their ability to modulate the activity of specific circuitry implicated in the illness. However, combining neuroimaging methods with brain stimulation is not without limitations. For instance, the hemodynamic response as measured in fMRI is only an indirect measure of neural activity, and neuroimaging with interleaved brain stimulation is susceptible to artifacts.

BRAIN STIMULATION TOOLS

Transcranial Magnetic Stimulation

Definition. TMS is the application of a rapidly changing magnetic field to the superficial layers of the cerebral cortex, which locally induces small electric currents, technically referred to as “eddy-or Foucault-currents.” These are circular electric currents induced within a given conductor by a changing magnetic field in that conductor. Michael Faraday originally discovered this induction process through his experiments in 1831 which were later quantified in Maxwell’s equations that encompass all of electromagnetism. Thus, TMS may be referred to as electrical stimulation without an electrode, as it uses magnetic fields to indirectly induce electrical pulses in the targeted cerebral tissue. A TMS device delivers strong magnetic pulses via a coil that is held on the scalp. Since magnetic fields are unaffected by the electrical impedance of the scalp and skull, this method of stimulation enables the focal stimulation of smaller areas of the brain than is possible with other noninvasive devices which use either alternating (ECT, CES) or direct (tDCS) electrical current for primary stimulation. TMS is an example of noninvasive stimulation of focal regions of the brain delivered at levels well below the seizure threshold and, as such, can be used for research or therapeutics without the need for anesthesia. It is a relatively new treatment modality in psychiatric practice. The original device was approved by the FDA in 2008 for the treatment of depression in patients who have failed a trial of one, but not more than one, antidepressant medication like an SSRI. This initial indication has been since expanded to include another device and to treat adult patients with major depressive disorder (MDD) who have failed to benefit from any number of antidepressant medications.

TMS, when given one pulse at a time without regular rhythm and at low rates of delivery, is referred to as single-pulse TMS. When repeated at regular rhythmic intervals that usually range from 1 to 25 Hz, the method is described as rTMS, which can be further functionally divided into "low-frequency rTMS" when stimulation frequency is 1 Hz or less or "high-frequency rTMS" when higher frequencies are used, usually 5 to 20 Hz. Theta burst stimulation (TBS) refers to high-frequency bursts (e.g., 3 pulses at 50 Hz) applied at a regular frequency (e.g., 5 bursts per second). TBS can be given continuously (cTBS) or intermittently (iTBS). cTBS is reported to induce inhibitory effects, while iTBS is reported to exert excitatory effects.

History. In 1985, Barker, Jalinus, and Freeston from the University of Sheffield in the United Kingdom reported the first use of TMS to stimulate the human brain. Early stimulators were capable of producing only low frequencies, typically less than 0.3 Hz, but it was the single-pulse TMS that, in those early explorations, proved to be very useful in the study of motor conduction and mapping of basic brain processes. While these studies used large round coils that lacked the focality of today's technology, even this early work with low-frequency stimulation was promising enough in the experimental treatment of depression and schizophrenia to encourage development of TMS as a promising treatment modality in psychiatry.

With the subsequent development of devices capable of much higher repetitive rates up to 20 Hz—higher brain functions like language, mood, and memory became accessible to stimulation. In 1996, Pascual-Leone and colleagues (Pascual-Leone et al., 1996) reported that TMS stimulation for 5 days to the dorsolateral prefrontal cortex (DLPFC) might have antidepressant properties that are prolonged and maintained beyond the acute period of stimulation. Studies such as this motivated an intense interest in the potential therapeutic value of focally applied rTMS targeted to the neuronal circuitry underlying not only depression but also other psychiatric disorders like anxiety disorders, schizophrenia, and others.

Mechanisms of Action. By Faraday's Law, the induced electromotive force in a closed loop is directly proportional to the time rate of change of magnetic flux through the loop. As applied in TMS, a changing magnetic field induces an electrical current in the circuitry of the superficial cortex underlying the TMS coil. This electrical current is commonly referred to as an "eddy current" for its circulating motion in a plane perpendicular to the plane of the magnetic field (also known as Foucault currents, named after the French physicist who described the phenomenon). At sufficient intensity, electrical currents across the cerebral cortex affect cell electrochemical gradients and produce local neuronal depolarization, which may result in a propagated action potential. For example, when the TMS coil is positioned over the hand area of the cerebral cortex's motor strip, the changing magnetic field, generated by the repetitive pulses, induces local currents immediately below the site of stimulation, which cause the neurons in area M1 to fire. In turn, this action potential propagates through the polysynaptic corticospinal tract and results in a twitch in the contralateral hand muscle, which can be used to study the excitability of the corticospinal system. In summary, TMS uses changing magnetic fields to indirectly induce focal electrical currents in the brain, thereby triggering the firing of functional neuronal circuits that can lead to observable behavioral effects. This effect can be easily demonstrated by simply moving the TMS coil across the cortical representation of neighboring muscle groups (the homunculus) with a series of single TMS pulses and then observing different muscle groups twitching.

Single TMS pulses can exert other effects when moved to different cortical areas. When positioned over the primary visual cortex (V1), scotomas, or "blind spots," are often elicited. This effect was referred to as "visual masking" or creating a "virtual lesion," and it illustrates that TMS can transiently disrupt neural functions.

Activation of motor neurons resulting in a muscle twitch or disruption of visual perception with single-pulse TMS represents examples of the acute effects of TMS-induced neuronal depolarization, as shown in Table 33.20b–3. The effects of single TMS pulses are believed to be immediate and short-lived. The muscle twitch as induced by TMS to area M1 is nearly instantaneous, with the hand movement occurring approximately 20 ms after the TMS pulse is applied. Both the motor-evoked potential and visual masking operate on a similar timescale measured in milliseconds. While producing such a motor-evoked potential does not require the patient to engage in activity, a single magnetic pulse does not typically affect cognition. Rather, interference or improvement of performance in a cognitive task is observable when a train of magnetic pulses are delivered immediately preceding or during the time when the participant is engaged in a cognitive task.

Traditionally TMS coils were limited to direct stimulation of the cortical surface, activating neurons at 1.5 to 3.0 cm beneath the scalp rather than deeper subcortical limbic structures. However, recent development and research of coil designs for stimulation of deeper brain areas, such as the H-Coil, suggest TMS may stimulate subcortical structures up to 6 cm beneath the scalp. The cost of deeper stimulation, however, may be a spatially wider electrical field distribution in the brain (i.e., reduced focality).

TMS can exert longer-lasting effects when the pulses are repeated at regular intervals in a process of rTMS, or when TMS pulses are paired with other forms of stimulation, such as electrical stimulation of a peripheral nerve (as in Paired Associative Stimulation), or when TMS is coupled with audiovisual stimuli, as in the example of classical conditioning of the brain response to TMS. The mechanisms underlying these lasting effects of TMS have been described by various researchers and are thought to be related to neuroplasticity and alterations in synaptic efficacy.

Treatment of psychiatric disorders with rTMS has been informed by attempts to focally alter pathologic cortical excitability where irregular excitability has been linked to a specific illness. For example, reduced activity in the left DLPFC has been implicated in several studies as a physiologic correlate of affective disorders. To correct this, numerous studies have applied high-frequency rTMS, which purportedly increases excitability, to the left DLPFC in an attempt to normalize activity in this region. In a related approach, some investigators who believe that abnormal interhemispheric balance in activation between the right and left DLPFC is related to mood psychopathology, have applied low-frequency rTMS to the right DLPFC, in an attempt to normalize this balance with stimulation trains believed to be inhibitory.

The potential therapeutic role of rTMS to focally modulate cortical excitability has been validated in animal studies, which suggests that rTMS, like ECT and antidepressant medication, can exert effects on neurotransmitter release, receptor expression, and gene expression. However, animal models of the effects of rTMS do not perfectly replicate the effects of rTMS on the human brain due to size differences and consequently how they impact the strength and distribution of the induced electric fields.

Side Effects, Interactions With Medications, and Other Risks. Administration of TMS is a noninvasive, relatively benign procedure when applied by a knowledgeable professional to a subject who has been properly evaluated. However, it is not entirely without risk. The most serious known risk of TMS is an unintended seizure, but this risk varies depending on stimulation parameters and patient history. Namely, single-pulse stimulation is less likely to result in a seizure compared to rTMS and variations in dosage—the combination of frequency, power, train duration, and intertrain interval—may also change the risk of unintended seizure. In addition, subject factors such as the presence of a neurologic disorder (epilepsy or a focal brain lesion) or use of seizure-lowering medications may also result in greater likelihood of an unintended seizure. The FDA's safety guidelines for avoiding seizure with rTMS are summarized in Table 33.20b–4.

Table 33.20b–4.
Safety Guidelines for rTMS

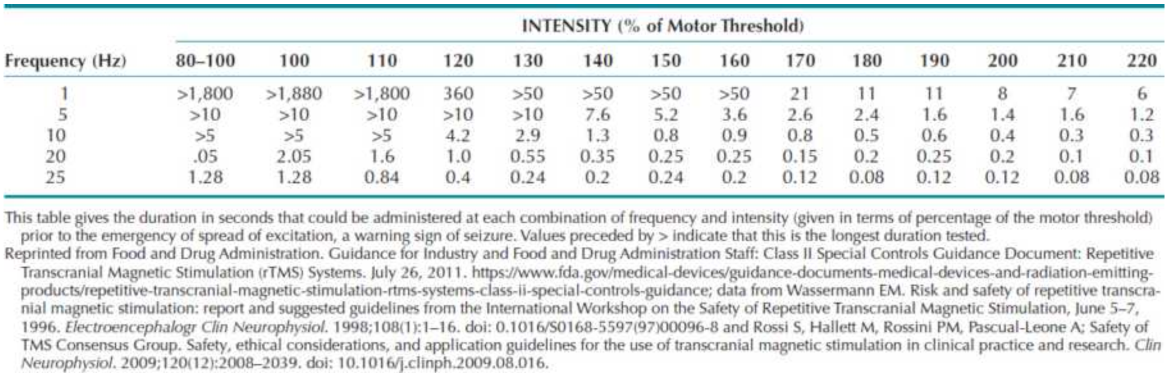

	INTENSITY (% of Motor Threshold)													
Frequency (Hz)	80–100	100	110	120	130	140	150	160	170	180	190	200	210	220
1	>1,800	>1,880	>1,800	360	>50	>50	>50	>50	21	11	11	8	7	6
5	>10	>10	>10	>10	>10	7.6	5.2	3.6	2.6	2.4	1.6	1.4	1.6	1.2
10	>5	>5	>5	4.2	2.9	1.3	0.8	0.9	0.8	0.5	0.6	0.4	0.3	0.3
20	.05	2.05	1.6	1.0	0.55	0.35	0.25	0.25	0.15	0.2	0.25	0.2	0.1	0.1
25	1.28	1.28	0.84	0.4	0.24	0.2	0.24	0.2	0.12	0.08	0.12	0.12	0.08	0.08

This table gives the duration in seconds that could be administered at each combination of frequency and intensity (given in terms of percentage of the motor threshold) prior to the emergency of spread of excitation, a warning sign of seizure. Values preceded by > indicate that this is the longest duration tested.

Reprinted from Food and Drug Administration. Guidance for Industry and Food and Drug Administration Staff: Class II Special Controls Guidance Document: Repetitive Transcranial Magnetic Stimulation (rTMS) Systems. July 26, 2011. https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/repetitive-transcranial-magnetic-stimulation-rtms-systems-class-ii-special-controls-guidance; data from Wassermann EM. Risk and safety of repetitive transcranial magnetic stimulation: report and suggested guidelines from the International Workshop on the Safety of Repetitive Transcranial Magnetic Stimulation, June 5–7, 1996. *Electroencephalogr Clin Neurophysiol*. 1998;108(1):1–16. doi: 0.1016/S0168-5597(97)00096-8 and Rossi S, Hallett M, Rossini PM, Pascual-Leone A; Safety of TMS Consensus Group. Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin Neurophysiol*. 2009;120(12):2008–2039. doi: 10.1016/j.clinph.2009.08.016.

Basic Studies. TMS has helped advance the basic understanding of normal brain function, as well as the pathophysiology of psychiatric illness. The combination of noninvasiveness, high temporal resolution, and particularly the ability to make use of frequency-dependent effects are significant advantages of TMS. Furthermore, the ability for TMS to be coupled with neuroimaging to explore the functional role of areas of activation seen with PET or fMRI imaging makes it a valuable tool for understanding brain function (see Fig. 33.20–4). For example, Luber and colleagues reported on the use of fMRI-guided stereotaxically applied rTMS to test the functional role of brain regions implicated in working memory. They found that rTMS applied to the precuneus facilitated working memory performance (Luber et al., 2007). Such an approach informs our understanding of the brain networks involved in normal brain function. Investigators have also delivered high-intensity TMS to participants during functional magnetic resonance image to demonstrate that stimulation of the left motor cortex was correlated with distributed cortical and subcortical activity in the bilateral premotor cortex, bilateral primary somatosensory cortex, bilateral second somatosensory cortex, left medial pontine nuclei, bilateral cerebellum, deep cerebellar nuclei, thalamic nuclei, and the basal ganglia (caudate and putamen). From such correlated changes in activity researchers are able to postulate brain areas that form distributed networks.

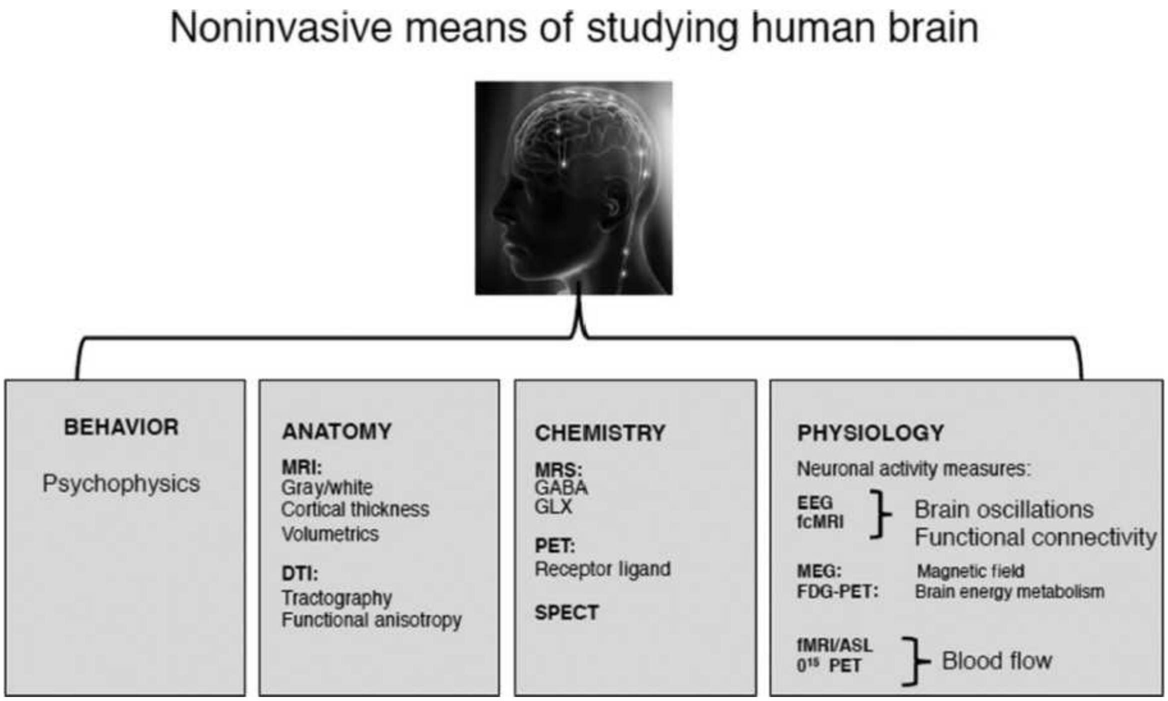
Noninvasive means of studying human brain
BEHAVIOR
Psychophysics
ANATOMY
MRI:
Gray/white
Cortical thickness
Volumetrics
DTI:
Tractography
Functional anisotropy
CHEMISTRY
MRS:
GABA
GLX
PET:
Receptor ligand
SPECT
PHYSIOLOGY
Neuronal activity measures:
EEG
fcMRI
Brain oscillations
Functional connectivity
MEG:
Magnetic field
FDG-PET:
Brain energy metabolism
fMRI/ASL
0^{15} PET
Blood flow

FIGURE 33.20b–4. Noninvasive means for studying human brain function. The human brain can be *passively* studied in terms of its anatomy, chemistry, physiology, and behavior. However, only brain stimulation can *actively* change brain function, and this can serve as a problem to test casual relationships. ASL, arterial spin labeling; DTI, diffusion tensor imaging; EEG, electroencephalography; ERP, event-related potentials; FA, fractional anisotropy; FDG, flurodeoxyglucose; fMRI, functional magnetic resonance imaging; MEG, magnetoencephalography; MRI, magnetic resonance imaging; MRS, magnetic resonance spectroscopy; PET, positron emission tomography; SPECT, single photon emission computer tomography. (Reprinted with permission from Columbia University, Department of Psychiatry.)

Clinical Studies. The use of rTMS in the treatment of MDD has been the most-studied therapeutic application of this technique. Evidence from a growing number of randomized controlled trials suggests a significant antidepressant effect with rTMS applied to the DLPFC. The largest study to date has been a multicenter trial conducted at four U.S. university hospital clinics reported by George and colleagues that included approximately 860 antidepressant medication-free patients with unipolar MDD who had not benefited from prior treatment. A total of 190 participants randomized to either active or sham treatment were included in the final analysis. Participants in the active treatment arm received daily left prefrontal rTMS for 3 weeks, with treatment standardized at 120% of the magnetic field intensity relative to the patient’s resting motor threshold, at 10 pulses per second (10 Hz) for 4 seconds, and an intertrain interval of 26 seconds. The authors reported that TMS was effective in treating major depression with minimal side effects. Not all studies have been positive, however, and the average effect size in depression is moderate—less than with ECT, but comparable to that seen with traditional pharmacotherapy.

Use of TMS in the treatment of schizophrenia has yielded some promising results in experimental studies that used either low-frequency rTMS to inhibit temporal–parietal areas of the cortex linked to hyperactivity during auditory or verbal hallucinations, or high-frequency rTMS to the prefrontal cortex aimed at modulating dopamine hypofrontality linked to negative symptoms. In the first double-blind randomized trial of deep TMS for the negative symptoms of schizophrenia, investigators found statistically significant reduction in negative symptoms after treatment, although there was no statistically significant difference between the active treatment and sham groups. Ongoing research has also demonstrated some positive results of TMS as a treatment modality in anxiety disorders and in OCD, panic, and Tourette syndrome.

Treatment Algorithms. In 2010, the APA included TMS within their Practice guidelines for the treatment of patients with MDD (3rd edition), recommending TMS as an initial treatment modality in the acute phase of depression. There are two FDA-approved TMS devices for use in adult patients with MDD: the NeuroStar TMS device manufactured by Neuronetics Inc. was approved in 2008 with the following FDA-approved indication: “for the treatment of MDD in adult patients who have failed to achieve satisfactory improvement from one prior antidepressant medication at or above the minimal effective dose and duration in the current episode”; and the Brainsway Deep TMS device that excites or inhibits neurons deeper inside the brain by using an H-Coil, was approved in 2013. The FDA-approved indication appears to be broader: “for the treatment of depressive episodes in adult patients suffering from MDD who failed to achieve satisfactory improvement from previous antidepressant medication treatment in the current episode.” The Neuronetics device delivers 10-Hz frequency pulses for trains of 4 seconds duration with an intertrain interval of 26 seconds and a typical session duration of 37.5 minutes (total of 5 daily sessions of 3,000 pulses for 6 weeks), while the Brainsway device delivers pulses with 18-Hz frequency for 2 seconds with an intertrain interval of 20 seconds and total trains duration of 20.2 minutes (5 daily sessions of 1,980 pulses for 4 weeks, followed by optional additional 12 weeks of biweekly maintenance treatments). A treatment course with either device is applied to the prefrontal cortex. At this time, rTMS treatments would most likely be considered as an alternative treatment modality to medications prior to more invasive methods like ECT.

Patient Selection. Patients who have failed a trial of one or more antidepressant medications are medication intolerant, and pregnant or postpartum patients avoiding the use of medications may be appropriate candidates for TMS. Notice that treatment of patients with a Neuronetics device who failed two or more medication trials in the current episode would be considered off-label based on its current FDA-approved indication. However, given the smaller effect size of TMS, for urgent or severely refractory cases, ECT would remain the ultimate “gold” standard treatment.

Table 33.20b–5.
Transcranial Magnetic Stimulation Dosing Parameters

- Scalp position, target localized method
- Coil shape and orientation; number of coils
  - Circular; Figure-eight; Double-cone; H-coils; B-shaped; cloverleaf
- Single pulse vs. paired pulse vs. repetitive vs. convulsive
- Train: Frequency, duration, intensity, total pulses
  - fast/high frequency at rates >1 Hz;
  - slow/low frequency at rates <1 Hz;
  - quadripulse repeated trains of four monophasic pulses separated by interstimulus interval
- Intertrain interval
- Number of trains/session
- Number of session/day
- Number of days/course
- Taper/maintenance/reintroduction

Dosing. The major characteristics of TMS dosing are shown in Table 33.20b–5. Each device’s technical parameters, such as stimulation intensity, pulse frequency, intertrain interval, and total number of pulses administered are the most basic adjustable elements of TMS dosing. In addition, coil shape and position relative to the target cortical area as well as the waveform shape and polarity, are also of important specific consideration (see Fig. 33.20b–5). In general terms, the delivered intensity of rTMS is related to the distance of the coil from the target area of the cortex. Thus, increasing the distance decreases the intensity of the stimulation able to reach the brain. Such decreased intensity has been found to be negatively correlated with the antidepressant response and with the degree of brain activation as evidenced by functional imaging. It is important to note that despite the many aspects that go into each device's dosing characteristics, considerable individual variability in the response to TMS has been reported.

Future Directions. TMS and other forms of magnetic stimulation hold a tremendous promise in psychiatric treatment due to their focality and noninvasiveness. However, much research is still needed to replicate preliminary findings with improved stimulators, optimize dosing, establish the patient characteristics that predict response, and examine the influence of concomitant medications on TMS effect. Research on the durability of clinical benefits of TMS is ongoing with some promising results. Using a large cohort of patients, investigators recently followed patients after their treatment for depression with TMS for 12 months post-treatment, and they found statistically and clinically meaningful durability of clinical benefits.

Other pursued directions are modification of pulse shapes to optimize stimulation as seen in controllable pulse shape TMS (cTMS), a device type that can generate rapid rate (<50-Hz) trains of mono/bi/polyphasic electrical pulses that are near rectangular in shape and vary in width, directionality, and amplitude. This work takes after the discovery in the 1990s, in the realm of ECT, of the benefits of square or rectangular waves of variable width compared with sinusoidal waveforms, which are generally now abandoned due to their inefficiency and creation of excessive cognitive side effects. In the first study examining cTMS use in human subjects, investigators found the cTMS device generated a broader range of pulse widths compared to conventional TMS devices, which these authors suggest could enable a more accurate study of the pulse width dependence of cortical excitability.

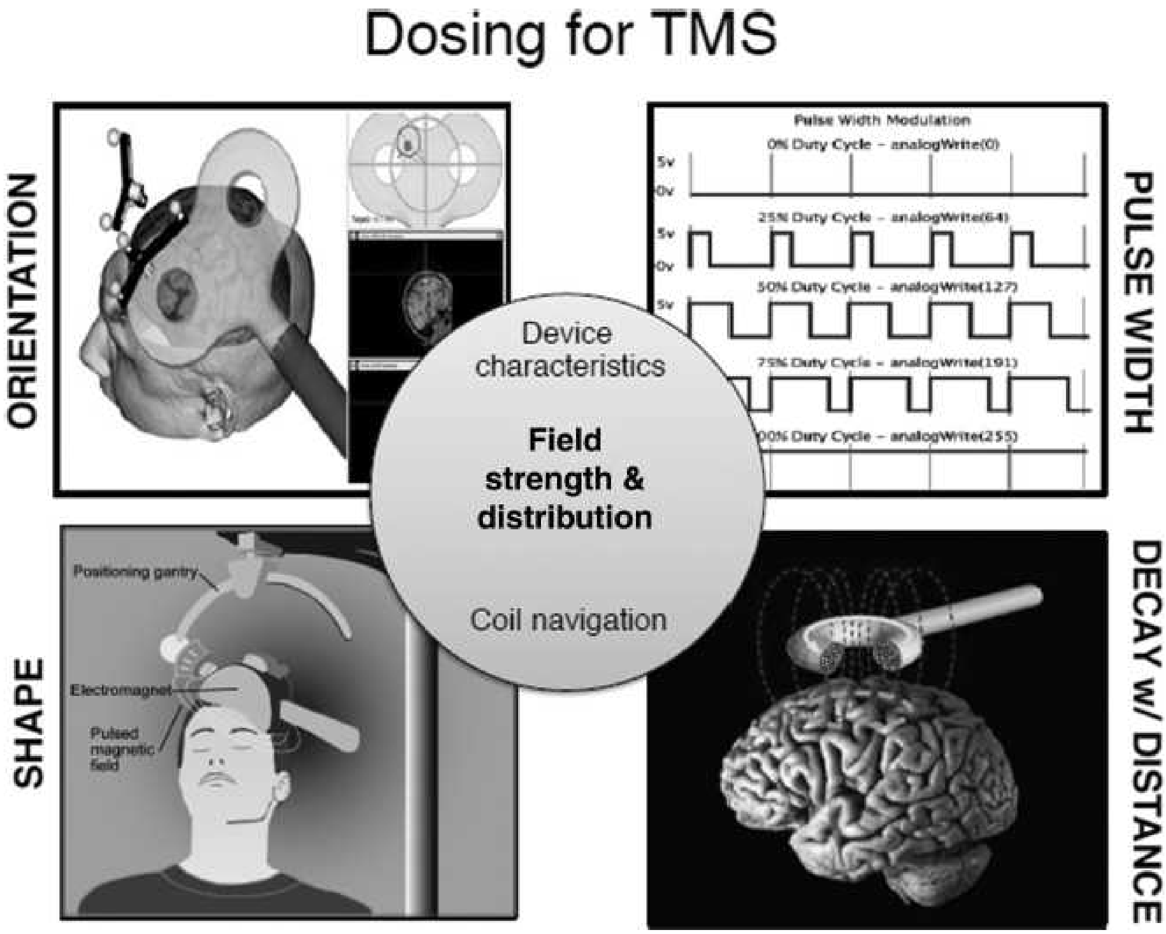
Dosing for TMS
ORIENTATION
PULSE WIDTH
SHAPE
DECAY w/ DISTANCE
Device characteristics
Field strength & distribution
Coil navigation
Pulse Width Modulation
0% Duty Cycle - analogWrite(0)
25% Duty Cycle - analogWrite(64)
50% Duty Cycle - analogWrite(127)
75% Duty Cycle - analogWrite(191)
100% Duty Cycle - analogWrite(255)
5v
0v
Positioning gantry
Electromagnet
Pulsed magnetic field

FIGURE 33.20b–5. Transcranial magnetic stimulation (TMS) dosing parameters. (Images clockwise from top left: NITRC (Online image) 2023. Accessed July 16, 2023. http://www.nitrc.org/project/list_screenshots.php?group_id=499&screenshot_id=374. Width Modulation (Online image). 2023. Accessed July 15, 2023. http://arduino.cc/en/Tutorial/PWM (https://creativecommons.org/licenses/by-sa/4.0/). Wasserman, E. via Wikimedia Commons (Online image). Accessed July 16, 2023. https://commons.wikimedia.org/wiki/File:Transcranial_magnetic_stimulation.jpg. National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services.)

Another Innovation, which cleared FDA approval at the end of 2021, is poised to impact the delivery of TMS and is discussed below.

Stanford Accelerated Intelligent Neuromodulation Therapy

Definition. Intermittent theta-burst stimulation (iTBS), which obtained FDA approval in 2021, is a more efficient form of TMS, that can significantly shorten the duration of TMS treatment sessions from 37 minutes to 3 minutes and produces equivalent antidepressant responses. Building on this success a team at Stanford sought to further improve the efficacy and efficiency of iTMS through the following: (1) treating patients with multiple sessions per day at optimally spaced intervals, (2) applying a higher overall pulse dose of stimulation, and (3) using functional connectivity MRI (fcMRI) for individualized precision targeting of the left DLPFC. This approach is 6 to 10 times faster than conventional TMS.

History. The approved iTBS courses involve daily stimulation sessions (600 iTBS pulses) for 6 weeks and demonstrated remission in 32% of patients and response in 49%. The team at Stanford believed that this could be improved by accelerated, spaced delivery of stimulation sessions, higher overall pulse doses, and individualized targeting of the Left DLPFC. The Stanford Accelerated Intelligent Neuromodulation Therapy (SAINT) is a therapy that incorporates all three of these improvements and has undergone two trials. These will be described below.

Mechanism of Action. As there are several components to the SAINT improvements, it is difficult to disentangle their effects. However, some educated hypotheses are available. Instead of a single treatment per day, spread out over weeks, the SAINT protocol uses 10 treatments per day with a 50-minute spacing. This approach is believed to have an enhanced effect compared with the same number of single daily sessions. The 50-minute spacing in the SAINT protocol mirrors evidence that multiple-spaced sessions produce an accumulating nonlinear improvement in clinical symptoms, that is, they are potentially synergistic with each other when administered close together. The SAINT protocol also used higher doses, and it administers five times the overall pulse dose of the FDA-approved iTBS protocol, as well as a higher density of stimulation (90,000 pulses in 5 days, compared with the standard 18,000 iTBS pulses in 6 weeks). Previous studies found that individuals who do not respond to a TMS treatment course will respond with a higher overall pulse dose. And the higher pulse doses are generally associated with higher TMS efficacy. Finally, the authors conjecture that the use of the functional connectivity–guided targeting method used in the SAINT study may have contributed to the high remission rate because the left DLPFC is a large brain area that consists of several subregions, some of which are correlated and some anticorrelated with sgACC activity, and it is believed that stimulating the subregion of the left DLPFC that is anticorrelated with the sgACC is what reduces melancholic symptoms. So the authors believe this may be an important factor that accounted for their high response rate compared to other TMs studies. However, they fully recognize the need for further studies to confirm these mechanistic hypotheses.

Clinical Studies. There have been two studies on the SAINT protocol. In the first, the researchers demonstrated efficacy in a small cohort of participants with severe and treatment-refractory depression. The second study builds on that by testing SAINT in a larger and more generalizable cohort of participants with treatment-resistant depression. In the second study, they were able to demonstrate the feasibility, safety, and preliminary efficacy of this approach. There was one subject excluded. Thus, 19 of 21 who actually participated (90.5%) met remission criteria, and in the intent-to-treat analysis, 19 of 22 participants (86.4%) met remission criteria. This is an impressively high response rate for depression and higher than many existing treatments. If it stands up to further research, it will impact the field substantially.

Side Effects. Neuropsychological testing demonstrated no negative cognitive side effects, which will be a major factor in patients selecting this therapy if it becomes more available.

Future Directions. The impressive early results of the research into SAINT make it a serious contender for more broad use, especially for in-patients, who can conveniently tolerate the need to be available for long days of treatment and who would enjoy the much shorter total days of treatment involved. The cost of SAINT will no doubt come down as its effectiveness is recognized, making it more available and scalable across mental health systems.

Transcranial Direct Current Stimulation

Definition. Transcranial Direct Current Stimulation (tDCS) is a noninvasive and nonconvulsive form of brain stimulation in which a very weak (0.5 to 2 mA) direct electrical current is applied using at least two surface scalp electrodes in which optimal placement continues to be investigated. This DC form of electrical stimulation can be distinguished from most other devices that stimulate with alternating currents, such as

CES, TMS, ECT, MST, VNS, and DBS, and it does not directly result in neuronal firing by inducing suprathreshold neuronal membrane depolarization, Rather tDCS elicits polarity-dependent alterations in the resting membrane potentials. These facilitating or inhibiting effects can last for up to 1 hour after stimulation and are thought to induce neuronal plasticity. The device is very portable and usually operated by 9V DC batteries.

History. There is history of ancient Egyptian, Roman, and Middle Eastern use of live electric catfish and torpedo fish as crude therapeutic interventions that involved use of DC stimulation for headaches and epilepsy. It was not until the 1960s that tDCS was empirically investigated. In Great Britain, Lippold and Redfearn (1964) reported that positive (*anodal*) currents induced elevated mood and motor activity and reduced anhedonia, whereas negative (*cathodal*) currents induced withdrawal and apathy. While subsequent attempts to replicate these findings failed, since the 1980s there has been a growing body of evidence of positive effects from tDCS in cognitive and motor function. Of late, tDCS has experienced resurged interest due to its ease of application, small device size, low cost, and relative lack of risks and side effects as compared with treatments like ECT and TMS.

Side Effects. There are no known serious adverse effects of tDCS. It is well tolerated with a low prevalence of mild adverse effects reported in the literature, which include mild tingling sensations, mild itching sensations, skin irritation, mild burns, and headache.

Mechanism of Action. The mechanism of tDCS is assumed to be modification of the synaptic microenvironment, with weak DC either lowering or raising the membrane potential and, thus, the threshold of activation, in turn, this affects the firing and conductance of neurons. Specifically, anodal electrodes excite the neuron and shift the resting membrane potential toward depolarization, while cathodal electrodes inhibit the neuron by shifting its potential toward hyperpolarization. Moreover, these effects may be dependent on the subtype of cortical neuron being stimulated. Because tDCS works via polarization and does not affect action potential firing in cortical neurons, the term “transcranial DC polarization” is favored by some investigators and both terms appear interchangeably in the literature.

Clinical Studies. Much of the experimental research on tDCS has been conducted with healthy controls and demonstrates the effectiveness of this stimulation method in exciting and inhibiting brain activity. Preliminary research has also examined the effects of tDCS on neuropsychiatric disorders such as depression, schizophrenia, pain disorders, and epilepsy. While early research on the application of tDCS for depression produced mixed results, methodologic improvements have reignited interest in research of tDCS as a therapeutic technique.

Current clinical psychiatric research focuses on the potential effectiveness of tDCS in reducing depressive symptoms. A recent meta-analysis of seven randomized sham-controlled studies demonstrated significant antidepressant effects after left prefrontal anodal tDCS. There was a reduction of symptoms as measured by the Hamilton Rating Scale for Depression (HRSD), ranging from 14.6% to 80% in the active treatment group versus approximately 10% symptom reduction in the sham tDCS group. In terms of the durability of this clinical response, four of the seven studies reported that antidepressant effects were sustained for 1 month after the end of treatment in the active tDCS group. Further research is needed to confirm these findings.

Current Status in Treatment Algorithms. tDCS is not approved by FDA for psychiatric treatment and is thus considered an experimental research technology with much work needed to demonstrate its efficacy. Should sufficient evidence be attained and the device approved by the regulatory authorities, tDCS could represent an inexpensive and relatively safe alternative form of treatment to medications, or prior to the more invasive stimulation methods like ECT or VNS. Within research, the typically applied stimulation protocols are two cephalic electrodes positioned below the cranial foramina and fissures at 1 to 2 mA intensity, with stimulation up to 20 minutes per session.

Future Directions. Most of the current tDCS devices utilize large, saline solution–soaked electrodes. The recent development of high-definition tDCS (HD-tDCS) devices that use small arrays of high-definition electrodes are under investigation for their increased focality compared to standard bipolar montages of traditional tDCS devices. The results of an HD-tDCS 4 × 1 montage that consists of one center electrode over the target and four surrounding return electrodes are promising in terms of increased excitation of a target area.

In general, current knowledge of tDCS safety, efficacy, indications, and dose–response relationship, as well as predictors of response, is limited and further exploration is needed.

Cranial Electrical Stimulation

Definition. CES, like tDCS, uses a weak (1 to 4 mA) current. However, with CES, the current is alternating. It is traditionally applied via saline-soaked, felt-covered electrodes clipped onto the earlobes. Other placement strategies are also currently being investigated. For example, a study at Columbia University directed by Peter Bulow utilized a cathode placed over the scalp area immediately in the vicinity of the DLPFC and an anode placed over the right arm. The current doses used in CES are in a similar range to that of tDCS, with the major difference being that the current is alternating, creating either sinusoidal or modified square waves with very widely varying frequencies ranging from 0.5 to 15,000 Hz, depending on the type of the stimulator.

History. CES was originally developed in 1949 in the former Soviet Union and used as an aid in inducing sleep. It was thus called initially "Electrosleep." In addition to the treatment for insomnia, it was also utilized as an aid in anesthesia because of the observation that even when the current was reduced to the point of not causing excessive sleepiness, some patients seemed to benefit from the anesthetic effect of the stimulation. The device was brought to the United States, and early research in America dates back to the 1960s. It was grandfathered in by the FDA as a Neurologic Therapeutic Device for the treatment of "insomnia, anxiety, depression, and stress," and the name CES, for Cranial Electrotherapy Stimulator, was suggested. In the 1970s CES was used by many U.S. clinicians to treat substance abuse and substance abstinence syndrome, in the belief that patients with a history of abuse were more susceptible to cross-addiction to psychotropic medications and because they were thought to be more resistant to the effects of such medications when compared to nonaddicted persons.

Mechanism of Action. The exact mechanism of action has not been ascertained at this time, and there is no agreement among researchers on its mode of action. Previous hypotheses proposed that the stimulation with the alternating microcurrent affects the thalamic and hypothalamic brain tissue and facilitates the release of neurotransmitters. Claims have been made that through interaction with cell membranes, the stimulation produces changes in signal transduction cascades associated with classical second messenger pathways involving calcium channels and cyclic AMP. The most promising research on the mechanism of action comes from a neuroimaging study where researchers concurrently stimulated participants with CES. Investigators found that CES at both 0.5 and 100 Hz (two commonly used frequencies in clinical practice) deactivated frontal and parietal brain areas and significantly altered connectivity in the Default Mode Network. These researchers suggest CES, an oscillating current, may alter intrinsic brain oscillation patterns, which is in line with previous EEG research on CES. While such research is promising, possible therapeutic effects are still not well understood.

Side Effects. It is believed that the CES stimulation is not harmful, primarily due to its low-voltage power supply (9V battery) and lack of any reported adverse event by the FDA. However, local skin irritation, headache, nausea, as well as a general feeling of dizziness have been reported. Device manufacturers do not advise the use of the device during pregnancy, in those with low blood pressure, or people who have arrhythmias or pacemakers.

Clinical Studies. While there are over 40 years of research on CES, many studies lack a sham comparator or control measure, and many are open-label studies that lack rigor. A recent Cochrane Review suggests that at the current time, there are no high-quality studies comparing the use of CES with a sham treatment that warranted inclusion in the review, thus asserting that there is insufficient evidence to support the use of CES in treating acute depression.

In generalized anxiety disorder, a pilot study using a small sample of 12 patients receiving CES for 6 weeks at 0.5-Hz frequency and 300-μA intensity resulted in significant decreases in anxiety scores as measured by the Hamilton Rating Scale for Anxiety. While promising, much more research is needed.

Current Status in Treatment Algorithms, Patient Selection, and Dosing. Although the FDA approves a number of CES devices, the use of CES has not been studied sufficiently in the United States and it does not have a specific place in any algorithm of standard U.S. psychiatric practice. Routinely, CES used for depression is delivered daily for a minimum of 3 weeks followed by a reduced schedule as needed for up to a year. It is reported to be a generally well-tolerated, noninvasive treatment.

Future Directions. As with tDCS, basic issues of indications, patient selection, dose–response relationship, and efficacy are currently under active research and remain to be optimized.

Magnetic Seizure Therapy

Definition. MST is a novel form of a convulsive treatment, currently under development in several research institutions in the United States, Canada, England, Germany, and Australia. The treatment employs an alternating magnetic field that crosses the scalp and the calvarium, unaffected by their high electrical impedance, to induce a more-localized electric current in the targeted regions of the cerebral cortex than is possible with ECT. The aim is to produce a seizure whose focus and pattern of spread may be controlled. Numerous publications characterize MST as an attempt to combine the potential benefits of unimpeded focality provided by magnetic fields with the powerful antidepressant efficacy of ECT.

MST is a convulsive treatment, in many ways similar to ECT. It requires approximately the same preparation and infrastructure as ECT. However, MST is administered using a modified TMS device, one that can administer higher output than the conventional TMS devices and thus relies on magnetic stimulation to induce a seizure, unlike external electrode electric stimulation in ECT. The MST procedure is performed under general anesthesia with a muscle relaxant. MST is still at the stage of clinical trials and is not presently FDA-approved. Current studies concentrate on optimizing the efficacy by focusing on the putative depressive centers of the brain while avoiding those brain areas involved in memory formation and consolidation (medial temporal lobes), with the aim of lessening the acute cognitive side effects possible with ECT.

History. MST was first described as a convulsive treatment by Lisanby and colleagues in 2001. In Bern, Switzerland, a 20-year-old inpatient with a medication-resistant depressive episode was reported to have been successfully treated. The treatment resulted in a 50% reduction in depression scores, and it was well tolerated with no adverse effect on Mini-Mental State Exam Scores. Prior to this human treatment, magnetically induced seizures were first performed in rhesus monkeys in 1998 with a custom-modified TMS device, and later more extensive safety testing was completed with nonhuman primates.

The first trial of MST in the United States was performed at Columbia University in New York and reported fewer acute cognitive side effects with MST than with ECT in a within-subject, cross-over, blinded trial. The second center to perform MST in the United States was the University of Texas, Southwestern Medical Center. Internationally, small human trials are either ongoing or have been recently conducted at centers in Canada, Australia, Wales, Scotland, and Germany.

Mechanism of Action. Induction of a seizure is hypothesized to be the underlying event responsible for the mechanisms of action of MST. As in ECT, these are not fully understood. However, due to its focality, MST appears to represent a tool better suited than ECT to study the mechanisms of action of convulsive therapy through its potential to induce seizures that are initiated in different regions of the brain. This focality of MST allows systematic investigation of the coupling between the site of seizure initiation and outcome. Furthermore, the quality of seizures induced by MST is believed to be distinct from those induced by ECT. MST-induced seizures show less impact on parasympathetic outflow, less generalization to the hippocampus and deeper brain structures, and result in less serum prolactin surge.

Side Effects. Adverse effects from MST are largely connected to the risks associated with anesthesia and generalized seizure. In addition, MST magnetic coil produces a clicking noise that may potentially affect hearing. To mitigate that risk and prevent any cumulative damage, both the patient and members of the treating team should wear earplugs. Studies suggest MST results in less retrograde and anterograde amnesia than ECT. Specifically, investigators have demonstrated that there is significantly less acute memory disruption (i.e., anterograde amnesia on treatment days) with MST compared to right unilateral ECT. Though promising, this result needs replication in a larger trial.

Currently, MST seizures are less generalized and tend to produce less robust motor convulsions. For this reason, the dosage requirement of the muscle relaxant (usually succinylcholine) is lower than with conventional

ECT. Shortening the period of paralysis could reduce the risks of the procedure by reducing the period of respiratory suppression. It should be noted, however, that the attenuated robustness and generalization of the MST-induced seizures may be a function of cautious dosing with new technology. Further studies are needed, both on device development and the efficacy of dosing.

Clinical Studies. As mentioned, MST was used to treat a human patient for the first time in 2000 in Bern. A depressed patient received four convulsive sessions induced by a prototype MST device by Magstim (Magstim, Whitland, Wales, UK) using a frequency of 40 Hz at 100% of that prototype machine's power intensity. Soon after, a second human treatment case was reported by the same center, with both patients describing post-treatment relief of their depressive symptoms without experiencing any significant side effects. In 2003 after the first blinded, randomized, within-subject, crossover trial in 10 human patients, Lisanby and colleagues reported that there were fewer acute side effects of MST compared to ECT (Lisanby et al., 2003). A custom-modified Magstim device was used in this trial with the frequency increased to 50 Hz, a peak magnetic field of 1.2 T produced for 8 seconds, and a pulse width of 500 µs. One of the more important observations made in this trial was that although seizures were induced in all 10 patients, the seizure threshold was attained in several cases only at the very maximum of that device's power output. Since this experiment, there have been technologic improvements such that MST devices are now capable of continuous stimulation at 100 Hz for up to 10 seconds (Magstim device) and even up to 250 Hz for a limited stimulation time (MagVenture device). It has been shown that antidepressant effects of convulsive therapy require a suprathreshold stimulation requiring at least doubling and, in some cases, even six times more than the seizure-threshold power in order to achieve an optimal antidepressant result. Thus, the technologic improvements in MST more reliably ensure the induction of a seizure and may optimize clinical effects.

Convulsive treatment at 100 Hz was first used in 2004 in an ongoing study of rhesus monkeys and, starting in 2007, in human trials. Published results that compared MST to ECT in nonhuman primates continue to show a superior safety profile of MST, even with the newer high-dose MST device prototypes. Pilot trials investigating the efficacy and safety of the 100-Hz MST in human patients demonstrate promising results. Investigators at Australia’s Monash University and Alfred Psychiatry Research Centre reported on an open-label trial of 13 patients receiving three MST per week (a maximum of 18 treatments were provided) with a MagVenture device. These researchers found 5 out of the 12 patients who completed the study experienced a 50% reduction in depressive symptoms, and none of the 13 patients receiving the treatment demonstrated changes in cognitive functioning after treatment.

A randomized controlled trial comparing the efficacy of MST to ECT in 20 patients with treatment-resistant depression was reported by Kayser and colleagues at the University of Bonn. Six of 10 patients receiving MST in this study demonstrated a clinical response to treatment with a 50% reduction in depressive symptoms as measured by the Montgomery Åsberg Depression Scale and the HRSD, and there was no significant difference in efficacy or cognitive side effects when compared to the 10 patients who received ECT. Notably, the patients in this study receiving MST were significantly quicker to recover and reorient immediately after treatment and did not experience nausea, headache, or muscle pains that were evident in the ECT treatment group.

More recently, Lisanby and colleagues have been conducting multicenter, double-blind, randomized, controlled trials comparing the efficacy of right unilateral ECT and MST. They found that ECT and MST had the greatest effect on the factor consisting of core depressive mood symptoms: hopelessness (loading score = 0.77), depressed mood (0.61), work and activities (0.56), helplessness (0.53), worthlessness (0.44), psychic anxiety (0.40), and suicide (0.34). There was no difference between ECT and MST on the magnitude or speed of score reduction across these factors. They concluded that MST exerts similar antidepressant action across all symptom dimensions compared to ECT. In a related article, her team discussed the advantage of MST as potentially a personalized therapy based on the greater localization of seizure activity possible with MST compared to ECT.

Current Status and Treatment Algorithms. No clinical algorithms exist for MST, as it is still an investigational protocol, and treatments outside of research are not FDA-approved. There are two MST devices being used in research: the MagPro MST device with Twin Coil and Dual Coil manufactured by MagVenture that delivers a biphasic waveform and pulse width of 370 μs allowing a maximal stimulation with 2,000 pulses at 100 pulses per second at 100% energy output (effective 70% to 75% energy due to power roll off at those levels) or 1,000 pulses at approximately 67 Hz at 100% energy output; and a MagStim MST device manufactured by MagStim in the United Kingdom which has a maximum capacity of 1,000 pulses at 100 Hz with a single round coil, as in ECT, where parameters such as electrode placement, pulse width, and dose are essential to maximize clinical benefit while minimizing risks of cognitive side-effects, stimulation parameters, and coil placement are also important considerations in MST. Assuming the hypothesis that MST can approach the efficacy of ECT with fewer side effects is correct, this magnetically induced convulsive treatment will play an important role in the referral for convulsive therapy. Furthermore, should future technologic advances in the delivery of MSTAND combine with our improved understanding of the brain's circuitry implicated in depression, the result could be a treatment that is equally efficacious to ECT but with a more favorable side effect profile. In that case, MST could ultimately replace ECT altogether. However, much more research is needed both in the realm of clinical treatment efficacy and technologic device development.

Patient Selection. MST is investigated as a treatment alternative for those patients who have failed several adequate medication trials or whose symptom severity requires a fast treatment response, traditionally requiring ECT treatment. Since general anesthesia is an integral part of the procedure, MST will most likely continue to be reserved for treatment-resistant patients after several classes of medications have been already tried.

Future Directions. MST is a novel treatment in early phases of clinical testing. Clinical treatment variables, including dosing, optimal coil placement, patient selection, and mechanisms of action are the topics of ongoing and future studies.

Vagus Nerve Stimulation

Definition. VNS is a direct, intermittent electrical stimulation of the left cervical vagus nerve via an implanted pulse generator. The stimulating electrode is wrapped around the left vagus nerve in the neck and is connected to the battery/generator, usually implanted subcutaneously in the left chest wall. George and colleagues (George et al., 2000) described VNS in a paper in 2000 as an important new tool for brain research and therapy. As a result of electrode stimulation of the left vagus nerve, afferent signals are sent to the nucleus tractus solitarius and the connected limbic and cortical areas. Implantation surgery is completed under brief general anesthesia, usually performed as a day procedure, and involves two incisions—one in the chest for the battery/generator and another for the electrode in the left neck. Electrical parameters affecting the stimulation characteristics can be controlled by professional staff externally with a programming wand. Patients have the ability to switch off the pulse generator by holding a magnet over the skin directly above the area where it is implanted.

History. In 2005, VNS was approved by the FDA for the adjunctive treatment of chronic, treatment-resistant depression. It had been previously approved for the treatment of resistant epilepsy. More than 50,000 patients worldwide have received VNS.

Mechanisms of Action. The majority of the fibers contained in the left vagus are afferents. It is estimated that as many as 80% of these fibers are upgoing afferents, and thus chronic stimulation of these nerve fibers predominantly changes activity in the brainstem nuclei, such as the nucleus of the tractus solitarius and other neighboring nuclei (e.g., Raphe) that alter serotonergic activity in cortical and limbic structures. Additionally, persistent stimulation of the vagal afferents is anticonvulsant, an effect that appears to be dependent upon the norepinephrine-producing locus coeruleus.

Side Effects and Contraindications. Implantation of VNS is a minor procedure but carries a few risks. Besides the risk of perioperative infection, surgical implantation carries a small risk of vocal cord paralysis, bradycardia, or asystole. Stimulation also appears to be safe and generally well-tolerated. The adverse events that have been reported are coughing, hoarseness dyspnea, reversible bradyarrhythmias, obstructive sleep apnea, neck pain, and, most commonly, voice alteration. A long-term naturalistic study of VNS for major depression estimates less than 1% of patients have terminated VNS because of adverse events. Generally, there are no adverse cognitive side effects with VNS; on the contrary, research demonstrates performance on neurocognitive tests is significantly improved with VNS, likely due to the improvement in depression. However, modest cognitive impairment has been reported with higher stimulation parameters.

Outright contraindications to VNS include bilateral or left cervical vagotomy, use of short-wave diathermy and microwave diathermy, or therapeutic ultrasound diathermy. A special consideration may be that, in patients with implanted VNS devices, MRI imaging is not permitted, unless a special “send/receive” coil is utilized.

Clinical Studies. The first VNS study by Rush and colleagues (Rush et al., 2000) in 30 treatment-resistant depressed patients reported in 2000 showed response rates of 40% after 10 weeks of open-label, adjunctive VNS treatment and an initial remission rate of 17%. The response rate was sustained (46%) after an additional 9 months of VNS, while a further significant increase in remission rates, from 17% to 29%, was observed. Likewise, considerable improvements in patient functioning were also reported with long-term VNS treatment. Subsequently, in a report on an extended sample (*n* = 59), the short-term (3 months) and longer-term (24 months) effects of adjunctive VNS treatment revealed 10-week response rates of 31% (after 3 months), 44% (after 1 year), and 42% (after 2 years). Remission rates were 15%, 27%, and 22%, respectively, demonstrating continuous and stable improvement. By 2 years, 81% were still receiving VNS. However, a large 10-week randomized, sham-controlled, blinded, multicenter trial of VNS for treatment-resistant unipolar or bipolar depression later reported by Rush and colleagues (Rush et al., 2005) failed to show a significant difference in benefit on the primary outcomes between active and sham treatment after 10 weeks. On the other hand, an open-label long-term follow-up of the same cohort (*n* = 202) demonstrated increased response rates (27.2% response and 15.8% remission rates at 12 months), which was higher than response rates in a nonrandomized comparison group receiving treatment as usual (i.e., not VNS). The researchers did not attribute improvement in the VNS group to medication changes or a delayed placebo effect in such chronically ill patients.

Current Status in Treatment Algorithms. The FDA indication for VNS is for the adjunctive long-term treatment of chronic or recurrent depression in patients 18 years or older experiencing a major depressive episode in the setting of unipolar or bipolar disease, who have not had an adequate response to four or more adequate antidepressant treatments. Consultation with another clinician experienced with treatment-resistant depression and VNS is recommended.

VNS treatment success rates are lower than those with ECT. Response to VNS is relatively slow—typically only 15% of patients respond in the first 3 months, increasing to approximately 30%, after 12 months. VNS may be worth considering, therefore, when patients have failed to respond to less invasive treatments, ECT was ineffective, or post-ECT relapse cannot be prevented with less invasive means. VNS might be helpful with longer-term relapse prevention, but results of controlled trials would be needed to guide that practice.

Patient Selection. A stepwise approach to patient evaluation for VNS is outlined in Table 33.20b–6. VNS is approved as a long-term adjunctive treatment for chronic or recurrent depressive episodes in adults with a major depressive episode that has not had a satisfactory response to four or more adequate antidepressant trials.

Table 33.20b–6.
Evaluation for Vagus Nerve Stimulation

- Does this patient meet the U.S. FDA labeling criteria?
  - Unipolar or bipolar disorder in a major depressive episode
  - Long-term major depressive episode (≥2 years)
  - Recurrent depression
  - Inadequate response to four or more adequate treatments
- Have other causes of medication resistance been ruled out?
  - Medical, substance, personality disorders
  - Compliance with adequate doses and durations
- Can the patient tolerate a slow-acting therapy?
- Does the patient need acute response now?
  - Electroconvulsive therapy may be a better option for the acutely suicidal
- Get a second opinion

Table 33.20b–7.
Clinical Scenarios: Vagus Nerve Stimulation (VNS) or Electroconvulsive Therapy (ECT)

Acutely suicidal, in need of rapid response	ECT
Psychotic depression	ECT
Acutely depressed, medication resistant, never had convulsive therapy	ECT
Chronically depressed, able to tolerate slow onset of action	VNS
Had convulsive therapy before but always relapsed quickly despite adequate medication	VNS
Cannot be tapered off maintenance convulsive therapy, and it is not effective	VNS

Consideration of what constitutes an "adequate" trial requires careful assessment of not only whether adequate dosages were prescribed for an adequate period of time, but also whether the patient was reliably compliant. Optimally, such assessment should be substantiated by careful review of treatment records, pharmacy reports, and blood levels where appropriate. Failed trials should include different medication classes, including selective serotonin reuptake inhibitors (SSRIs) and selective serotonin/norepinephrine reuptake inhibitors (SNRIs), as well as tricyclic antidepressants, and monoamine oxidase inhibitors (MAOIs), and various augmentation strategies with or without ECT. A second opinion from a psychopharmacologist specializing in treatment-resistant depression and knowledgeable about VNS is advisable.

Clinical scenarios when ECT might be favored over VNS, and vice versa, are outlined in Table 33.20b–7. For example, VNS does not exert rapid antidepressant action and thus acutely suicidal patients and others in need of rapid response would be more appropriately treated with other strategies. VNS is indicated as a long-term treatment option only for those with chronic or recurrent courses of illnesses.

ECT and TMS have been safely used in patients with an implant when the VNS generator was turned off during the treatments, which is necessary for ECT because of the anticonvulsant effects of VNS. It remains to be studied whether VNS could be useful in relapse prevention post-ECT.

Dosing. There are varying stimulation parameters (i.e., current, pulse width, frequency, and duty cycle) that are determined by the treating physician, which can be programmed into the pulse generator via telemetry. Optimum dose–response parameters are still being researched, although a large multicenter clinical trial did not find significant differences in efficacy in treating depression between “low” (output current of 0.25 mA, pulse width of 130 ms), “medium” (0.5 to 1.0 mA, pulse width 250 ms), or “high” (1.25 to 1.5 mA, pulse width 250 ms) at acute phases (8 to 22 weeks) or chronic phases of treatment (25 to 50 weeks).

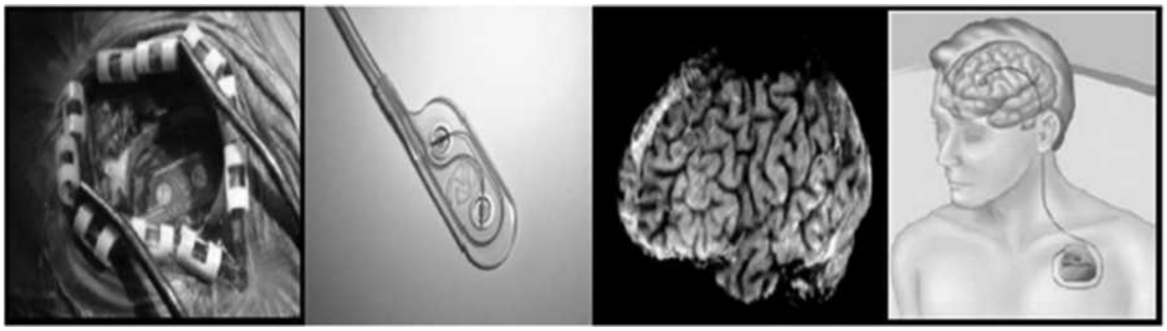

FIGURE 33.20b–6. Implanted cortical stimulator.

The epilepsy literature suggests that there is a threshold current for efficacy. Because the adverse effects of VNS are known to be dose-dependent, treatment parameters are often chosen to mitigate specific side effects. For example, lowering pulse width reduces neck pain, allowing patients to tolerate higher currents.

Future Directions. More research is required to establish the dose–response relationships for VNS. Future studies may explore optimal medication strategies to augment responses, test the potential role of VNS for long-term relapse prevention (e.g., after ECT), and study its mechanisms of action.

Implanted Cortical Stimulation

Definition. Cortical brain stimulation is a novel neurosurgical approach in which electrodes are implanted directly over the surface of the cortex to provide electrical brain stimulation in a targeted superficial region. This approach is being studied for the treatment of conditions such as stroke, tinnitus, and treatment-resistant depression. Because of its novelty, different investigators use different names to describe it, ranging from implanted cortical stimulation (ICS), through cortical stimulation system, to cortical brain stimulation, and direct cortical electrical stimulation (DCS). The stimulating electrode can be implanted relatively superficially in either epidural or subdural layers, but the specific topographic placement of the stimulator varies depending on the condition being treated (see Fig. 33.20b–6). For example, in psychiatric applications, investigators have reported on a trial involving an ICS in the superficial brain cortex area overlying the left DLPFC. The electrodes are powered by a pacemaker-like device that is implanted in the upper chest area.

History. ICS is an invasive but reversible form of brain modulation and, like DBS and other less invasive neuromodulation methods, stems in part from the psychosurgical treatments for severe mental disorders performed during the 1940s and 1950s. These early forms were highly experimental and often resulted in significant adverse side effects due to physical surgery in the brain tissue. Today, with the advancement in our understanding of the specific functional brain regions and their interconnections, as well as the relative ease of access to neuroanatomical and brain imaging studies, direct electrical stimulation is a safer way of either temporarily disrupting faulty function, enhancing deficient functions, or otherwise reversibly modulating abnormal brain activity.

Trigeminal Nerve Stimulation

Definition. Research of trigeminal nerve stimulation (TNS) in humans has been ongoing since 2001 and has been most commonly studied as a therapeutic intervention for drug-resistant epilepsy. However, TNS has been explored as an adjunct to pharmacotherapy in depression with positive results, and clinical trials of TNS for ADHD are currently underway. At this time, TNS is not approved for use outside of research, however.

Mechanisms of Action. Due to the anatomy of the three bilateral afferent cutaneous branches of the trigeminal nerve with connections to the nucleus tractus solitarius, locus coeruleus, and raphe nuclei, noninvasive stimulation of the V1 branch of the trigeminal nerve has therapeutic potential in the treatment of mood disorders and drug-resistant epilepsy.

Clinical Studies. Studies exploring the psychiatric applications of TNS are ongoing. Within epilepsy research, typical TNS treatment utilizes stimulation for at least 12 hours a day for 18 weeks. In a recent multicenter trial comparing high (experimental active treatment) and low frequency (active control) external TNS (eTNS) for the treatment of drug-resistant epilepsy, investigators reported that the response rates for high-frequency eTNS at 120 Hz with a pulse duration <250 μs and maximum off-time of 90 seconds were significantly greater (30.2% reduction in seizures) than low-frequency eTNS at 2 Hz with a duty cycle of 2 seconds on and 90 seconds off and 50-μs pulse duration (21.1% reduction in seizures).

Side Effects and Contraindications. TNS is generally well tolerated, with a low incidence of headache and skin irritation. Some patients report sensory experiences of tingling or vibration, and there have been reports of a significant increase in heart rate but no significant changes in blood pressure.

Future Directions. The simplicity of administering eTNS by applying a gel electrode on the forehead is a unique advantage over surgically implanted devices. Thus far, research demonstrates that bilateral TNS is more effective than unilateral stimulation and high-frequency TNS is more efficacious compared to low-frequency stimulations. As clinical benefits may increase with treatment course, longer treatment length may increase therapeutic effects.

Nonconvulsive Electrotherapy for Treatment-Resistant Unipolar and Bipolar Depression

Definition. While traditional ECT, in which a seizure is induced in the patient, is still the treatment with the most evidence for being an effective therapy for treatment-resistant major depression, many patients are reluctant to engage in this option due to their concerns about having seizures and the associated memory loss. Moreover, while repetitive transcranial magnetic stimulation (rTMS), VNS, and other approaches mentioned above are either FDA-approved, or have evidence for efficacy, the brain stimulation alternatives to ECT mentioned so far have lower efficacy for treatment-resistant patients than traditional ECT. Thus, in addition to the alternatives mentioned above, an emerging innovation to traditional ECT involves a treatment that is almost the same as ECT, but without the seizure induction, namely nonconvulsive electrotherapy (NET). NET involves similar electrode placement and use but the stimulation is kept below the doses that induce seizures.

Mechanisms of Action. While the mechanism of action is not known, clinical experience has shown that some patients who received subconvulsive electrical stimulation that failed to cause seizures never the less had a therapeutic effect. Moreover, both older human studies from the 1950s and more recent animal studies (showed that subconvulsive electrical stimulation elicited antidepressant-like effects. This led to the hypothesis that, at least for some individuals, the active ingredient in ECT was not the seizure but the stimulation.

Clinical Studies. While this is a new field in need of more research, a study of NET for the treatment-resistant unipolar and bipolar depressive disorder was conducted. This pilot, proof-of-concept study used an ECT device with a dose of electric stimulation below the dose that induces a seizure. The procedure still requires sedation with anesthesia but avoids the seizure, and the treatment was delivered using a bifrontal electrode placement. The study was performed on a small number of participants (n = 13), and the depression response and remission rates were 73% and 55%, respectively. Thus, this new adaptation of ECT could provide an additional treatment option for patients who struggle to respond to treatment.

Side Effects and Contraindications. Additionally, those results showed that NET could avoid the cognitive effects associated with ECT, as assessed by Mini-Mental State Exam and the Autobiographical Memory Inventory-Short Form. While formal cognitive testing was not performed, these preliminary results indicate that NET is likely to have fewer cognitive side effects.

Future Directions for Treatment of Resistant Depression. Currently, additional studies are being performed with NET in patients with treatment-resistant depression. In addition, it is being tested in other patient populations, including those with chronic pain, as described below.

Future Directions for Treatment of Pain. In a currently active study, Diana Martinez at Columbia University is investigating whether NET can be helpful in treating chronic pain. In this context, the technique is called transcranial electric stimulation therapy (TEST). In the procedure, electric stimulation is applied to the brain, and anesthesia is required, but, as in NET, the dose of stimulation is lower and does not cause a seizure. In the current trial, treatments are performed three times per week, just like traditional ECT, and measurements of pain, mood, and memory are collected before and after the sessions (ClinicalTrials.gov identifier [NCT number]: NCT05113563). It should be noted that there is prior literature on the use of traditional ECT for pain. However, traditional ECT involving seizures did not seem appropriate to many pain patients and their clinicians. Given the current crises in balancing the needs of chronic pain patients with risks of addiction and overdose, a nonseizure form of electrical stimulation, one without cognitive side effects, may add an important tool to the pain management armamentarium. The results of the study by Martinez are, therefore, of great interest to psychiatrists and pain management clinicians alike.

ACKNOWLEDGMENT

This chapter represents an update and revision of previous chapters in this series, most recently by Dr. Sarah Hollingsworth Lisanby from the last edition. We are indebted to Dr. Lisanby for their work and retain much of their prose in this current edition.

FURTHER READINGS

Alexander L. Nonconvulsive electric stimulation therapy; its place in the treatment of affective disorders, with notes on the reciprocal relationship of anxiety and depression. *Am J Psychiatry.* 1950;107:241–250.

American Psychiatric Association. *Practice Guideline for the Treatment of Patients With Major Depressive Disorder.* 3rd ed. American Psychiatric Association; 2010.

Barker AT, Jalinous R, Freeston IL. Non-invasive magnetic stimulation of human motor cortex. *Lancet*. 1985;1(8437):1106–1107.

Blumberger DM, Vila-Rodriguez F, Thorpe KE, et al. Effectiveness of theta burst versus high-frequency repetitive transcranial magnetic stimulation in patients with depression

(THREE-D): a randomized non-inferiority trial. *Lancet.* 2018;391:1683–1692.

Cole EJ, Stimpson KH, Bentzley BS, et al. Stanford accelerated intelligent neuromodulation therapy for treatment-resistant depression. *Am J Psychiatry.* 2020;177(8):716–726.

Deng ZD, McClintock S, Husain M, Lisanby S. Depressive symptom dimensions and response trajectories in electroconvulsive therapy and magnetic seizure therapy. *Biol Psychiatry*. 2022;91(9):S21.

Esser SK, Huber R, Massimini M, Peterson MJ, Ferrarelli F, Tononi G. A direct demonstration of cortical LTP in humans: a combined TMS/EEG study. *Brain Res Bull*. 2006;69(1):86–94.

Fitzgerald PB, Brown TL, Marston NA, et al. Reduced plastic brain responses in schizophrenia: a transcranial magnetic stimulation study. *Schizophr Res*. 2004;71(1):17–26.

George MS, Lisanby SH, Avery D, et al. Daily left prefrontal transcranial magnetic stimulation therapy for major depressive disorder: a sham-controlled randomized trial. *Arch Gen Psychiatry*. 2010;67(5):507–516.

George MS, Sackeim HA, Rush AJ, et al. Vagus nerve stimulation: a new tool for brain research and therapy. *Biol Psychiatry*. 2000;47(4):287–295.

Gersner R, Toth E, Isserles M, Zangen A. Site-specific antidepressant effects of repeated subconvulsive electrical stimulation: potential role of brain-derived neurotrophic factor. *Biol Psychiatry*. 2010;67:125–132.

Huang YZ, Rothwell JC. The effect of short-duration bursts of high-frequency, low-intensity transcranial magnetic stimulation on the human motor cortex. *Clin Neurophysiol*. 2004;115:1069–1075.

Kallioniemi E, McClintock SM, Deng ZD, Husain MM, Lisanby SH. Magnetic seizure therapy: towards personalized seizure therapy for major depression. *Pers Med Psychiatry*. 2019;17–18:37–42.

Kayser S, Bewernick BH, Grubert C, Hadrysiewicz BL, Axmacher N, Schlaepfer TE. Antidepressant effects of magnetic seizure therapy and electroconvulsive therapy in treatment-resistant depression. *J Psychiatr Res*. 2011;45(5):569–576.

Lisanby SH, Luber B, Schlaepfer TE, Sackeim HA. Safety and feasibility of magnetic seizure therapy (MST) in major depression: randomized within-subject comparison with electroconvulsive therapy. *Neuropsychopharmacology*. 2003;28(10):1852–1865.

Lisanby SH, Schlaepfer TE, Fisch HU, Sackeim HA. Magnetic seizure therapy of major depression. *Arch Gen Psychiatry*. 2001;58(3):303–305.

Luber B, Kinnunen LH, Rakitin BC, Ellsasser R, Stern Y, Lisanby SH. Facilitation of performance in a working memory task with rTMS stimulation of the precuneus: frequency- and time-dependent effects. *Brain Res*. 2007;1128(1):120–129.

Pascual-Leone A, Rubio B, Pallardó F, Catalá MD. Rapid-rate transcranial magnetic stimulation of left dorsolateral prefrontal cortex in drug-resistant depression. *Lancet*. 1996;348(9022):233–237.

Polster JD, Kayser S, Bewernick BH, Hurlesmann R, Schlaepfer TE. Effects of electroconvulsive therapy and magnetic seizure therapy of acute memory retrieval. *J ECT*. 2015;31:13–19.

Rasmussen KG, Rummans TA. Electroconvulsive therapy in the management of chronic pain. *Curr Pain Headache Rep*. 2002;6(1):17–22.

Regenold WT, Noorani RJ, Piez D, Patel P. Nonconvulsive electrotherapy for treatment-resistant unipolar and bipolar major depressive disorder: a proof-of-concept trial. *Brain Stimulation*. 2015;8(5):855–861.

Rush AJ, George MS, Sackeim HA, et al. Vagus nerve stimulation (VNS) for treatment-resistant depressions: a multicenter study. *Biol Psychiatry*. 2000;47(4):276–286.

Rush AJ, Marangell LB, Sackeim HA, et al. Vagus nerve stimulation for treatment-resistant depression: a randomized, controlled acute phase trial. *Biol Psychiatry*. 2005;58(5):347–354.

Wang H, Wang X, Sackeim H et al. LTD and LTP induced by transcranial magnetic stimulation in auditory cortex. *Neuroreport*. 1996;7(2):521–525.

Wassermann EM. Risk and safety of repetitive transcranial magnetic stimulation: report and suggested guidelines from the international workshop on the safety of repetitive transcranial magnetic stimulation, June 5–7, 1996. *Electroencephalogr Clin Neurophysiol*. 1998;108(1):1–16.

Williams NR, Sudheimer KD, Bentzley BS, et al. High-dose spaced theta-burst TMS as a rapid-acting antidepressant in highly refractory depression. *Brain*. 2018;141:e18.

33.20c Neurosurgical Treatments

BENJAMIN D. GREENBERG, M.D., PH.D., AND NICOLE C.R. MCLAUGHLIN, PH.D.

Neurosurgical treatments for psychiatric illness continue to receive increasing attention. Clinically or in research, they are used in small, highly selected groups of patients with disabling illnesses proven intractable to conventional medication and behavioral therapies. Current techniques are vastly different from the crude freehand “psychosurgery,” especially prefrontal lobotomy, of the mid-20th century, which saw wide, often indiscriminate use. Stereotactic ablation is now MRI guided, vastly improving its accuracy. MRI-guided ablation using radiosurgery or the newer high-intensity focused ultrasound (HIFU) avoids craniotomy, while radiofrequency or newer laser ablation requires small, usually bilateral, craniotomies. Deep brain stimulation (DBS) also requires craniotomy for implantation. DBS can be stopped or modified if it causes adverse effects or is ineffective but requires indefinite follow-up by (relatively rare) specialized teams. In the United States, DBS for intractable OCD received FDA humanitarian approval in 2009 and is an approved treatment in the European Union. Other uses of DBS in psychiatric populations remain investigational. This chapter focuses on neurosurgery for otherwise intractable OCD, which also serves to illustrate general principles and procedures for potential uses of neurosurgery in other conditions. These include intractable major depression, eating disorders, addiction, Tourette syndrome, and other illnesses where research is ongoing; recent publications on those topics are cited in this chapter.

A large body of open-label clinical data supports the effectiveness and relative safety of modern neurosurgical interventions. An increasing body of controlled data has been reported, usually in relatively small studies, given the relative rarity of the “surgical” population. In tandem, an increasing body of systematically collected data on adverse events allows risk/benefit judgments to be better informed. Despite intriguing hypotheses and, recently, more physiologic data from intracranial recordings, the mechanisms of action of psychiatric neurosurgeries remain to be established, a situation that still parallels functional neurosurgery for movement disorders. While these treatments are sometimes criticized for this reason, neurosurgical interventions are similar to other therapies insofar as the most relevant issues are safety and efficacy, rather than correction of still poorly understood pathophysiologic processes.

Standards of care for psychiatric neurosurgery use strict criteria and a formalized practice. For Humanitarian Use Devices (i.e., DBS), there is oversight by a local Institutional Review Board (IRB). More generally, DBS or ablation is not approved unless a multidisciplinary committee reaches a consensus that a given candidate is appropriate, based on diagnostic evaluations and a detailed assessment of prior clinical course and treatments received. The informed consent process unfolds over time in the context of a thorough medical and psychiatric evaluation. At the time of this writing, surgical intervention is mainly reserved for patients with severe, incapacitating OCD who have exhausted conventional behavioral and medication therapies. The ethical context of psychiatric neurosurgery has increasingly seemed similar to that of any higher-risk procedure or research, for example, movement disorder neurosurgery.

Lesion procedures at first developed largely empirically and were then influenced by theories implicating cortical–subcortical systems in disordered behavior. DBS was, in turn, influenced by the experience with stereotactic ablation. For both kinds of procedures, neuroscience tools allow hypothesis generation with respect to potential neurocircuit and physiologic mechanisms of action, as does translational research in animal models. Such multidisciplinary work may inform other device-based techniques, both invasive and noninvasive. Brain stimulation using implantable devices that can respond selectively to patterns of brain activity putatively associated with symptoms (e.g., closed-loop or “responsive” DBS) has recently become a special focus of interest.

HISTORY

Trephination in prehistoric times was likely the earliest surgery for psychopathology. The first modern neurosurgical treatment in psychiatry was reported in 1891. These bilateral cortical excisions in patients with dementia and depression were described as yielding mixed results. In 1935, John Fulton and Charles Jacobsen reported that chimpanzees showed less fear in the “experimental neurosis” paradigm after frontal cortical ablation (lobectomy). The animals retained an ability to perform complex tasks, but fear expression was reduced. António Egaz Moniz, a Portuguese neurologist, attended Fulton’s 1935 lecture. Prompted by these findings, Moniz pioneered prefrontal leukotomy in collaboration with his neurosurgeon Almeida Lima. First, by using alcohol injections and subsequently mechanically using a leucotome (a kind of surgical knife), Moniz and Lima performed “psychosurgery” on 20 severely ill institutionalized patients; 14 were said to have exhibited worthwhile improvement. In an era of overflowing asylums and few effective treatments for debilitating illnesses, surgery was at first enthusiastically embraced. Moniz won a Nobel Prize in 1949 for this contribution.

From the mid-1930s until phenothiazines emerged in the 1950s, these techniques proliferated globally. Walter Freeman, a neuropsychiatrist, was perhaps the most zealous promoter of psychosurgery in the United States. Initially, he worked closely with neurosurgeon James Watts. They developed freehand procedures to achieve prefrontal lobotomy (i.e., severing white matter connections between the prefrontal cortex and the rest of the brain). Freeman and Watts reported on their first 200 cases in 1942. Although the benefits of the surgery were highlighted, others acknowledged a significant rate of complications, such as frontal lobe syndromes (including apathy and disinhibition), seizures, and even deaths. At its peak, lobotomy was being performed on approximately 5,000 patients per year in the United States alone. A review of 10,365 prefrontal lobotomies performed from 1942 to 1954 in Britain concluded that while 70% showed improvement, adverse effects included 6% mortality, seizures in 1%, and disinhibition in 1.5%. There were widespread reports of blunted personality and socially inappropriate behavior.

In the late 1940s and early 1950s, recognition of these risks prompted attempts to develop modified stereotactic surgical procedures that might yield better results. For example, Spiegel and Wycis, who began stereotactic neurosurgery in humans, reported in the 1940s that dorsomedial thalamotomies improved obsessive-compulsive symptoms. However, with the introduction of chlorpromazine (Thorazine) in 1954, medical management of psychiatric illness became newly possible. Thus, despite the advent of stereotactic neurosurgical techniques and a continued high prevalence of severe, treatment-refractory psychiatric illness, psychiatric neurosurgery was all but abandoned in favor of nonsurgical therapies. Later, as the limitations of neuroleptics and electroconvulsive treatment became known, there was a renewed interest in psychiatric neurosurgery at a few academic medical centers in the 1960s. Refinement of stereotactic methods and emerging theories regarding the role of the limbic system in emotion shaped a newer age of psychiatric neurosurgery. The pioneers of these contemporary procedures faced harsh opposition from political groups, as well as the general population, based largely on the results of crude and indiscriminately applied operations of an earlier era. Political and social pressure dissuaded all but a handful of neurosurgeons and psychiatrists from offering these treatments. Despite these societal influences, in the late 1970s, the authors of a U.S. National Commission report found, to their own surprise, that available evidence indicated that many patients benefited after lesion procedures. They advocated continued systematic research and a national data registry. Although the registry never materialized, gradually, over nearly five decades, research and education regarding modern neurosurgery led to its legitimacy as a treatment for severe, otherwise intractable psychiatric disorders, now with more refined practices and surgical procedures. Indications, expectations of benefit, and risks have become better defined.

PATIENT SELECTION

Indications and Contraindications

The best-established indication for psychiatric neurosurgery remains otherwise intractable OCD; there are also multiple reports of lesion procedures or DBS used in major depressive disorder (MDD). Factors considered during patient selection include the following:

1. *Primary diagnosis:* The patient must meet clinical criteria for the diagnostic indication, and this disorder should be a primary cause of the patient's debility and suffering.
2. *Severity and functional impact:* The patient must have a chronic, severe, and debilitating illness; the duration of the primary illness must exceed 1 year and typically exceeds 5 years. Severity is gauged on standardized instruments. Surgical candidates with OCD typically have Yale-Brown Obsessive-Compulsive Scale (YBOCS) scores of a minimum of 26 to 30 (of a possible 40). Debility should be indicated by standardized ratings of a low level of functioning and quality of life.
3. *Adequacy of previous treatment:* Patients must have already tried an exhaustive array of other available established treatments delivered by competent practitioners, who should typically be interviewed by the evaluating team after available medical records are reviewed. A thorough assessment of the adequacy of previous medication and psychotherapy trials is equally important and often requires interviews with treating psychotherapists as well as reviews of medical records. Medication trials must similarly be judged adequate; where appropriate, high-dose antidepressant trials (for OCD in particular) should have been attempted, with duration of each medication trial no less than 2 months.
4. *Psychiatric comorbidity:* Appropriate treatment must also have been rendered for any comorbid psychiatric disorder. Strong

relative contraindications include active psychoactive substance abuse or severe personality disorders, which may interfere with adherence to presurgical treatment trials or with the exhaustive evaluation and consent processes, or with the recommended postsurgical treatments and follow-up.

5. *Medical comorbidity and surgical fitness:* Structural brain lesions or central nervous system injuries of clinical significance are strong contraindications. Medical conditions that increase neurosurgical risks (e.g., cardiopulmonary disease) and age over 70 years represent a relative contraindication for lesion procedures, while DBS may be used, as for movement disorders, in somewhat older cases. A history of past seizures is a risk factor for perioperative seizures after lesion procedures and must factor into the overall risk–benefit assessment.
6. *Access to postoperative care:* The extensive evaluations and subsequent psychiatric neurosurgery mark the beginning of a new episode of care. It is crucial that patients have access to adequate postoperative treatment, including a psychiatrist (typically the referring physician) and a psychologist who will accept responsibility for managing the case after discharge. Arrangements for postoperative care (e.g., intensive behavior therapy) should be confirmed ahead of time. Importantly, after lesion procedures, care can generally be delivered in standard treatment settings without the need for highly specialized psychiatric neurosurgery teams. For DBS, access to specialized teams is essential over the long term. Once implanted, DBS patients require clinical monitoring and device adjustment, which can be intensive and time-consuming, especially but not only early in treatment. Device monitoring and replacements may need to occur on a relatively urgent basis. The continuing costs incurred can be substantial, and the adequacy of third-party reimbursement needs to be ensured in advance to the extent possible. After either lesion procedures or DBS, family or significant others may be needed to support and accompany patients to follow-up care, similar to the level of support that is usually necessary during the intensive evaluation process. Currently, DBS for intractable OCD is eligible for reimbursement

under an FDA HDE approval in the United States. However, actual reimbursement can vary. The procedure has full approval from the European Union. Other psychiatric uses of DBS remain investigational, an important consideration for reimbursement.

7. *Informed consent:* Under no circumstances should psychiatric neurosurgery be performed on an unwilling patient. The patient must be able and willing to render informed consent. Consent monitoring, including the use of structured consent feedback tools, may be considered to help ensure that the consent process is adequate. In rare instances, these procedures are performed with the assent of the patient and formal consent from a legal guardian. In this context, age less than 18 years represents at least a strong relative contraindication.

Multidisciplinary Case Review

Assessing candidacy differs somewhat from center to center and between nations, as governed by law. Across centers, particular importance is placed on case review by a group with the appropriate breadth of expertise. As noted above, decisions regarding candidacy must be made after a thorough review of the patient’s case, taking into account the written record, current clinical examination, and formal referral from the patient’s treating psychiatrist. In most U.S. expert centers, the decision is typically made by an interdisciplinary committee; in other countries, psychiatric neurosurgery may require additional formal review and approval from duly constituted national commissions or other procedures.

Available evidence strongly supports the idea that the group of patients for which neurosurgery might be appropriate represents a small proportion of the total of treatment-seeking individuals. For OCD, less than 1% of individuals in a large naturalistic study would meet current surgical criteria. Even when referred by clinicians, up to half of the patients in a different study were not accepted for surgery on the basis of either a questionable diagnosis or, most often, a conventional treatment that was judged inadequate. Another, more recent study of DBS for OCD found that of nearly 300 individuals clinician or self-referred for initial consideration for study entry, less than 10% eventually underwent DBS implantation.

PREOPERATIVE WORKUP

In addition to the above clinical evaluation and formalized review, the preoperative workup for psychiatric neurosurgery includes physical and neurologic examinations, standard presurgical blood and urine laboratory tests, as well as an ECG, plus specialized tests of brain structure and function. A comprehensive preoperative assessment should include brain MRI, psychometric testing using neuropsychological batteries that assess multiple domains, and in some cases, an EEG. This workup is broadly similar to that for DBS or lesion procedures for movement disorders.

POSTOPERATIVE CARE

Immediate postoperative care includes the usual medical and surgical management following any stereotactic neurosurgical procedure. Special attention is paid to signs or symptoms of potential surgical complications, including infection, hemorrhage, seizures, or altered mental status. For ablative procedures, a postoperative MRI should be obtained to document the placement and extent of lesions, timed to optimally visualize the evolution of the lesions; for DBS, imaging by CT assesses potential postimplantation hemorrhage. Over the longer term, intensive postoperative psychiatric treatment is recommended since the efficacy of the intervention may rely on the synergy between the neurosurgical intervention itself and enhanced response to pharmacologic or behavioral therapies. Although dosages of psychotropic medications might be reduced during the immediate perioperative period, the medication regimen should be readjusted as tolerated postoperatively. Moreover, in the case of OCD, intensive behavior therapy should be continued or resumed early, for example, as soon as the first postoperative month for DBS or by 3 months for lesion procedures (which include radiofrequency ablation, Gamma Knife radiosurgery, and the newer procedures of laser or high-intensity ultrasound ablation) where the clinical benefit is expected to evolve more slowly. Longitudinal assessment of patient adherence to medication and behavioral treatments is essential so that treating clinicians can intervene to optimize care if patients change their regimens on their own (as has happened in some cases postoperatively despite extensive preoperative advice that patients not change medications on their own).

For DBS, as noted above, electrode implantation is followed by immediate postoperative imaging (e.g., computerized tomography; CT scan) to assess for potential hemorrhage and evaluate general device placement. This is followed by a several-week recovery period to enable resolution of local edema and stabilization of other factors (e.g., pneumoencephaly) that might influence the response to stimulation. Postoperative imaging to more accurately visualize device locations (e.g., CT scan for fusion with preoperative MRI) may take place at that time. Then, a systematic electrode survey and adjustment of stimulation parameters are performed before initial settings are determined. This is usually done in the outpatient setting and is often a time-consuming process, lasting hours over 1 or more days. Ongoing protocols for DBS entail frequent follow-up, especially during the initial 6 months after implantation, to enable optimization of stimulation parameters, patient monitoring, and management, including coordination of pharmacologic and behavioral therapies.

DESIGN AND INTERPRETATION OF CLINICAL STUDIES

It is critical to appreciate the inherent challenges which continue to surround research in this domain. The patient populations in question are very challenging and can require intensive management over the months and years of a clinical trial. These populations can be very heterogeneous. High rates of comorbidity are the rule. The total numbers are small. As noted above, less than 1% of treatment-seeking OCD patients may be eligible for neurosurgery, and proportions of appropriate patients with major depression or other conditions are also likely to be small. As a result, the major U.S. psychiatric neurosurgical centers collectively treat a small number of cases yearly. Thus, statistical power is limited, as are estimates of efficacy and adverse effect rates. The major academic centers might have their own surgery of preference (sometimes more than one). Historically, centers have employed somewhat different measures for assessing treatment response, although methods are converging. Selection criteria have also varied somewhat across centers, especially for the earliest studies of a particular technique. Consequently, it is often difficult to contrast or combine data across different groups. Moreover, determining the cause of long-term postoperative improvement can be confounded by exposure to other new therapies as they emerge or even by the longitudinal course of illness. And improvement in function and quality of life have not been weighted as highly as changes in symptom severity, which are at least equally important.

LESION PROCEDURES

Although numerous approaches have been tried from the mid-20th century to the present, four lesion procedures have been the most used. All four entail bilateral lesions, performed using stereotactic methods. Two of them, anterior cingulotomy and anterior capsulotomy, are most used currently. For all ablative procedures, the outcome cannot be fairly assessed for a considerable postoperative period, which could extend up to 2 years. While effects of DBS may develop sooner, for example, within several months postoperatively, long-term follow-up remains essential to assess the durability of responses.

Anterior Cingulotomy

The first cingulotomies were performed for intractable pain, a practice that continues along with newer research into DBS of the anterior cingulum for pain. The clinical feature targeted is the affective component of pain rather than nociception. It was noted that pain patients with comorbid anxiety or depressive conditions had better responses. H. Thomas Ballantine Jr. and colleagues later found that anterior cingulotomy was relatively safe and studied its effectiveness for psychiatric indications. Since 1962, the group at Massachusetts General Hospital has performed well over 1,000 anterior cingulotomies. The surgery is conducted under local anesthesia, and two or three approximately 1-cc lesions are made, under MRI guidance, on each side by thermocoagulation through bilateral burr holes. The target is within the anterior cingulate cortex (Brodmann areas 24 and 32), at the margin of the white matter cingulum bundle. A subgroup of patients may undergo a second procedure months later, either a cingulotomy extension or the addition of subcaudate tractotomy. There has also been investigational use of DBS at the ventral capsule (VC)/ventral striatum (VS) target after cingulotomy.

In an early (1987) retrospective review of cingulotomy outcomes (*n* = 198, average follow-up 8+ years), from 56% to 79% of individuals with affective disorders, OCD, or anxiety disorders were judged improved on a global measure. A later prospective study of 64 consecutive patients undergoing cingulotomy for intractable OCD between 1989 and 2009, using the YBOCS for OCD severity and Beck Depression Inventory (BDI) for associated depressive symptoms, found full or partial responses (≥35% or 25% to 34% YBOCS severity improvement, respectively) in 35% or 7% a mean of close to 11 months postoperatively. Almost half had a repeat cingulotomy or subcaudate tractotomy by 5 years, when full or partial response rates improved to 47% and 22%, respectively. The severity of comorbid major depression decreased modestly from baseline. Psychiatric adverse events included suicide in 2 of the 64 patients (3%), one of whom had made a prior suicide attempt and the other had bipolar depression. Other adverse events, including those after repeat procedures, were mainly transient: memory impairment (8%), abulia (6%), urinary retention (3%), intraoperative or postoperative single seizures (6%), or infection requiring repeat surgery (1.5%). More recent reviews have demonstrated that about 38% to 47% of OCD patients who have received cingulotomy are full responders. As in the above study, nonresponders may also respond to a repeated procedure.

Anterior Capsulotomy

Anterior capsulotomy (or capsulotomy) was first performed by Tailarach and colleagues in France. Later, Lars Leksell and colleagues further developed the procedure in Sweden. Initially mainly done by radiofrequency thermocoagulation, Gamma Knife radiosurgery, with laser ablation (laser interstitial thermal therapy [LITT], see below) and HIFU are more recent developments. While earlier in capsulotomy, the entire dorsoventral extent of the anterior limb of the internal capsule was lesioned, modern procedures place lesions selectively within its ventral part impinging on the adjacent ventral striatum. This primarily interrupts fibers of passage between the ventral prefrontal cortex and subcortical nuclei, including dorsomedial and anterior thalamus. The term “ventral capsulotomy” is used to describe the newer procedure. Ventral capsulotomy may be performed as an outpatient procedure, with an overnight hospital stay usually the most that is required. Anterior capsulotomy is primarily used for OCD, though it has also seen use in major depression and other severe anxiety disorders.

In the 1970s, Leksell and colleagues reported on their first 116 cases of thermocapsulotomy (again, lesions included the entire dorsoventral capsule). They noted favorable responses in 50% for OCD and 48% for major depression. A subsequent larger review found adequate data in 213 of 362 cases. Of those 213, 64% were deemed to have benefitted substantially on global scales (i.e., were judged much or very much improved). Transient side effects included headache, incontinence, and confusion, the latter lasting up to 1 week. In a subset of 24 patients followed prospectively, there was one intracranial hemorrhage and one episode of seizure. Fatigue was seen in 29%, 17% had a poor memory, and 8% exhibited a lack of attention to appearance. Weight gain was a common enduring side effect with a mean increase in body mass of about 10%.

A 2008 study reported on 25 OCD patients who underwent thermocapsulotomy, Gamma Knife anterior capsulotomy, or both combined. It used disorder-specific rating scales and measures of adverse effects. Even though these patients had failed to respond to treatments unavailable to the earlier cohorts (thus, their illnesses might be expected to be more refractory than in the early experience), effectiveness was, in fact, comparable to that in the early work: 48% response and 36% remission using YBOCS OCD severity ratings. This series also demonstrated a notable adverse effect burden. Forty percent were judged to have a clinically significant frontal impairment (problems with executive functioning, apathy, or disinhibition). Most in this group had repeated thermocapsulotomy and/or high-dose Gamma Knife radiation resulting in larger lesions and which today would be avoided. In contrast to the early work, MRI was available and showed that smaller lesions, less lateral or posterior in extent, were associated with not only fewer adverse effects but also greater symptom reduction.

A 2011 study is an example of a more recent use of capsulotomy in intractable depression. Twenty patients treated from 1992 to 1999 were followed for a mean of 7 years. At follow-up, 50% were responders. Neurocognitive and personality testing were not significantly different. Significant adverse effects were infrequent, and there were no deaths. A 2019 prospective study of capsulotomy for depression ($n = 10$) by a different group found a 60% response rate and no pattern of adverse neurocognitive events. The response was associated with changes in ventromedial but not dorsolateral prefrontal pathways on diffusion MRI.

Gamma Ventral Capsulotomy

Starting in the 1990s, radiosurgery has been used in intractable OCD for intractable OCD. The procedure which developed was ventral capsulotomy, placing two lesions (“shots”) in the ventral portion of the anterior limb of the internal capsule under MRI guidance using the Leksell Gamma Knife. A study reported in 2018 by Rasmussen and colleagues describes the largest cohort of gamma ventral capsulotomy (GVC) participants to date (*n* = 55), where the first patient had the procedure in 1993. Like almost all studies of lesion procedures, this was open-label. It showed a significant reduction in OCD symptom severity (a YBOCS severity drop of 35% or more) in 56% of the group. Patients also had significant improvements in depression, anxiety, quality of life, and global functioning.

Five-year outcome data illustrate the delayed nature of the response. The therapeutic benefit was achieved over 1 to 2 years and was essentially stable by 3 years. Full responses (YBOCS OCD severity improvement by >35%) developed in over half of cases by 2 to 3 years after surgery. Partial responses averaged approximately 10% at 2 to 3 years postop. Psychiatric adverse events included manic episodes during follow-up in 6% of patients, all of whom had a prior history of hypomania or mania; all improved after mood stabilizers were started. One patient developed substance dependence years after the procedure, and another developed a mild frontal syndrome which later improved. The most serious adverse effect after Gamma Knife ventral capsulotomy was radionecrotic cyst formation in 5% of the group, which in one case had severe sequelae. Cyst formation is a risk also associated with Gamma Knife radiosurgery for other indications, as is cerebral edema and headache (onset after months to years). And small asymptomatic caudate infarctions in 11%. In general, studies have shown no changes in personality and either stable or improved cognitive functioning following Gamma Knife anterior capsulotomy.

A controlled trial of GVC carried out at the University of São Paulo, Brazil is a rare example in the literature on ablative psychiatric neurosurgery. Sixteen patients with rigorously defined intractable OCD were enrolled; 8 received active GVC, and 8 received sham surgery in which they entered a specially constructed antechamber attached to the Leksell Gamma Knife. The groups were followed for 1 year, after which individuals in the sham group were offered actual surgery. There was no change in the sham group during the 1-year follow-up period. In contrast, two patients (25%) in the active GVC group showed reductions in OCD severity during the masked phase of 35% or greater on the YBOCS. Given the small sample, the effect reached the trend level of statistical significance (p = .54). In the group receiving active surgery first, five of the eight patients (62.5%) went on to respond during follow-up after the masked phase. Four of the eight patients who initially received the sham procedure later underwent the active procedure. Two of these later responded, for an overall response rate of 58% after active radiosurgery in the total sample. In this study, a longer masked sham period, which would have been optimal scientifically given the delayed nature of the response, was not chosen given the ethical concern that severely ill individuals not undergo unduly long follow-up after sham surgery.

Newer Methods: Laser Interstitial Thermal Therapy and High-Intensity Focused Ultrasound

More recently, given technologic advances in the use of real-time neuroimaging in targeting brain structures, a new technique, LITT (or laser ablation), has been used for capsulotomy. This procedure is a minimally invasive ablative approach using laser thermal energy transmitted via fiber optic catheters fed through cranial burr holes. In recent open-label research, responses were comparable to other capsulotomy techniques: McLaughlin et al. found 77.8% full responders (seven of nine patients) after LITT for treatment-resistant OCD. Adverse effects included transient apathy/amotivation postsurgery (two patients), hemorrhage (one patient), and persistent insomnia (one patient).

A different emerging technology, in this case, used to noninvasively create ventral capsulotomy lesions is MRI-guided HIFU, which utilizes high-power sonication to create targeted thermal lesions in neural tissue. HIFU can make targeted intracranial lesions without the need for burr holes or craniotomy. One group, Kim and colleagues, in South Korea recently demonstrated that 54.5% of 11 patients had >35% reduction in YBOCS OCD severity 2 years after the procedure, which was well tolerated. An issue in using this method is that adequate lesions are not made after HIFU in all individuals, based on factors including skull density which differ across individuals. Thus, in one study, while four of six OCD patients who had adequate lesions responded, 25% of those undergoing the procedure were excluded from the analysis due to inadequate lesions.

Subcaudate Tractotomy

Subcaudate tractotomy was introduced by Geoffrey Knight in 1964 as an early attempt to limit adverse effects by restricting lesion size compared to prefrontal lobotomy. Targeting the substantia innominata (just inferior to the head of the caudate nucleus), its goal was to interrupt white matter tracts connecting the orbitofrontal cortex and subcortical structures. The surgery involved placement of radioactive yttrium-90 seeds at the desired centroid, yielding lesion volumes of approximately 2 cc on each side. Subcaudate tractotomy has been used primarily in OCD and major depression. It has sometimes been used as a procedure subsequent to anterior cingulotomy when the latter has been judged to have a suboptimal outcome. Subcaudate tractotomy is now rarely done as a single procedure.

A 1975 study (predating MRI-guided surgery or modern psychopharmacology) reported follow-up data (at 2.5 years on average) for 134 patients (of an initial 208) after subcaudate tractotomy in a mixed patient population (OCD, depression, or anxiety disorders). The subset was chosen since adequate data, including structured interviews, were available. Global improvement was judged clinically meaningful in 50%, 68%, and 63% of those with OCD, depression, and anxiety disorders, respectively. Individuals with schizophrenia, substance abuse, or personality disorders did poorly. Adverse events typically included transient headache, confusion, somnolence, or disinhibition. Longer-term adverse effects included “mild” undesirable personality changes in 7%, seizures in 2%, and suicide in 3%.

In 1994, a review of 1,300 subcaudate tractotomy cases concluded the procedure enabled 40% to 60% of patients to lead normal or near-normal lives. The suicide rate was 1% versus 15% in a contemporary comparison group with major affective disorders. There were no deaths due to other causes in this series, and the seizure rate was comparable to that seen in previous reports. In contrast to the earlier report, the 1994 review did not find enduring personality changes. The investigators proposed predictors of positive response, including MDD diagnosis, sudden onset of symptoms, positive family history of psychiatric disorder, and previous positive response to electroconvulsive therapy. Physiologic responses elicited during intraoperative electrical stimulation of the target area before it was lesioned were also thought to predict better responses.

Limbic Leucotomy

Limbic leucotomy (also spelled leukotomy) combines subcaudate tractotomy and anterior cingulotomy. It was introduced by Desmond Kelly and colleagues in 1973. The lesions have typically been made via thermocoagulation or with a cryoprobe. Prior to MRI guidance, intraoperative stimulation was used to guide the placement of the lesions—eliciting autonomic responses was thought to mark the target. As for other procedures, limbic leukotomy has been used for OCD, major depression, severe anxiety disorders, and, less often, for otherwise treatment-intractable repetitive self-injurious behaviors or tic disorders.

An initial report of prospective outcomes in 66 patients reported global improvement, at a mean of 16 months of follow-up, in 89%, 78%, and 66% of patients with OCD, major depression, and anxiety disorders, respectively. A subsequent report found that 84% of those with OCD were improved at 20 months. Transient side effects included headache, lethargy or apathy, confusion, and lack of sphincter control. There were no reported seizures and no deaths; however, enduring lethargy was present in 12%, and one patient suffered severe memory loss due to improper lesion placement.

Much later, in 2002, Montoya and colleagues reported on 21 patients who underwent limbic leukotomy for OCD or major depression at Massachusetts General Hospital. Using stringent criteria, they reported that 36% and 50% responded, respectively, at a mean follow-up of 26 months. Another 2002 report described five consecutive cases that suggested that limbic leucotomy might be beneficial for patients with severe self-mutilation in the context of repetitive self-injurious, tic-like behaviors. A subsequent study from a different group of 16 patients with affective disorders found that 69% responded, 19% had a partial response, and 13% did not improve or became worse.

More recent reviews of limbic leucotomy have noted short-term adverse effects on memory, headache, hallucinations, extrapyramidal signs, perseveration, confusion/delirium, scalp infection, and urinary incontinence with lethargy and seizure (approximately 3%) adverse effects that have the greatest impact. Patients often had longer postsurgical hospitalizations as compared to subcaudate tractotomy or cingulotomy alone. Overall, the adverse effect burden of limbic leucotomy appears greater than those of other lesion procedures.

DEEP BRAIN STIMULATION

DBS for psychiatric illness is an old idea, although the devices, surgical and stimulation techniques, and theoretical models of relevant neurocircuitry, as well as how to define structural and functional networks, continue to advance. In 1948, Pool and colleagues used caudate nucleus stimulation in a patient with depression and anorexia, and there was a line of research by Heath and colleagues that included stimulation to “septal nuclei” in the 1950s and 60s. Modern DBS began 50 years after the earliest work when in 1998, Nuttin and colleagues described a case series of patients who received DBS initially based on the anterior capsulotomy target for intractable OCD, with positive results. The procedure involves the placement of small-diameter brain “leads” with multiple electrode contacts into subcortical nuclei or specific white matter tracts. The surgeon drills burr holes in the skull under local anesthesia and then places the leads, guided by multimodal imaging and stereotactic landmarking. Usually, this is done bilaterally. The patient is typically sedated but awake during surgery. Later, the “pacemakers” (an implantable neurostimulator or pulse generators) are implanted subdermally (e.g., in the upper chest) and connected via extension wires tunneled under the skin to the brain leads. Since various combinations of electrodes can be activated at adjustable polarity, intensity, and frequency, DBS permits a considerable range of ways one can modulate brain function. Thus, parameters can be optimized for individual patients. This might achieve improved efficacy and more favorable adverse effect profiles in comparison with ablation. The process of optimizing stimulation parameters for individual patients, typically performed by a specially trained psychiatrist in an outpatient setting, is quite time-consuming. Since device programming and troubleshooting need to continue indefinitely, providing DBS requires expert, attentive, multidisciplinary, and long-term follow-up. In cases where no beneficial settings can be identified despite extensive efforts, the electrodes can be inactivated, and device components may be surgically removed. The surgical risks of device removal include infection and hemorrhage, as for implantation. There are important psychiatric risks of device removal, which patients may view as the end of a “last resort” episode of care and could become despondent or suicidal as a result.

Here we focus on DBS for OCD, where there is regulatory approval in multiple countries (although health care reimbursement varies). There has also been considerable research on DBS for intractable depression, which has been reviewed in a number of places, including by Sullivan and colleagues (cited at the end of this chapter). Similarly, DBS for medically refractory Tourette syndrome is a focus on increasing interest and work; see the referenced article by Martino and colleagues for a discussion of the state of the art. There are multiple other potential uses of DBS in neuropsychiatric illness, ranging from addiction to eating disorders to others, which will not be further discussed here. We also do not discuss closed-loop or “adaptive” DBS, except to note that intriguing work in this area is ongoing and has the potential to not only advance this particular field but also to reveal physiologic signals underlying disabling symptoms in unprecedented ways. That, in turn, might inform advances in noninvasive brain-circuit–focused therapies, which could be applied to a vastly greater proportion of patients versus surgical approaches.

As noted, collaborative work in the 1990s led to the development of DBS at the ventral anterior limb of the internal capsule and adjacent ventral striatum (the VC/VS) for treatment-intractable OCD. A collaborative study by Greenberg et al. reported open-label data (*n* = 26); some later underwent masked DBS discontinuation. Across centers, 61% had a full response (≥35% YBOCS improvement), and 12% were partial responders (25% to 34% improvement). In the group as a whole, responses tended to stabilize after about 3 months of stimulation, although this varied across individuals. Cognitive functioning was not adversely affected by DBS. Adverse events included hemorrhage after implantation in two patients (~8%); in one of them, apathy resulted, which resolved over 3 months. One patient (~4%) had a single seizure intraoperatively without recurrence. One patient developed a superficial surgical wound infection that responded to antibiotics. In two patients (~8%), a DBS system wire broke, requiring surgical replacement. Complications due to stimulation itself or stimulation interruption were the most common, including four (~16%) patients with worsened depression or suicidal thinking similar to events in their presurgical course of illness. In three of them, this worsened depression occurred during a blinded crossover to sham stimulation. Mild hypomania, judged as stimulation related but nonserious, was seen in eight patients (31%). In all cases, hypomania was either reversible with a change in stimulation or resolved over days without a DBS parameter change.

In a study of DBS at an anatomical target closely related to the VC/VS, the nucleus accumbens (NAcc), 16 patients underwent 8 months of open-label treatment which included (as did the VC/VS study) exposure-based behavior therapy and ongoing pharmacotherapy. The open phase was followed by a double-blind crossover phase with randomly assigned 2-week periods of active or sham stimulation, followed by open maintenance DBS. In the open phase, 56% of patients responded. In the double-blind, sham-controlled phase (*n* = 14), YBOCS scores were significantly lower for active versus sham stimulation. The most prominent stimulation-related adverse was elevated mood or hypomania in 50% of patients, judged nonserious in all cases. One patient (6%) had a surgical wound infection, while five (31%) reported forgetfulness, and three (19%) reported word-finding difficulties. However, objective testing in this, as in other studies of psychiatric DBS, found either a lack of cognitive deterioration or improvement. This is unlike the situation in DBS for movement disorders, where stimulation at the subthalamic nucleus may be associated with cognitive decline in Parkinson disease.

A controlled crossover study investigated the effects of DBS at a different target within the general cortical–basal ganglia–thalamic circuitry implicated in OCD, the subthalamic nucleus. Of a total group of 16 patients, half were randomized to active DBS and half to sham stimulation, after which each group crossed over to the other condition in a double-masked fashion. There was a significant difference between active and sham DBS in favor of the active treatment. Adverse events included an intracerebral hemorrhage on device insertion in one patient (6%), leading to a permanent motor deficit in one hand, and infection leading to pulse generator removal in two patients (12%). One patient was described as having stimulation-related hypomania, and one with mania related to DBS. Both effects were reversible.

Most recently, there have been a number of relatively small-scale controlled trials of DBS for intractable OCD, focusing on targets including the ventral anterior limb of the internal capsule, the VS/NAcc, and the (also adjacent) bed nucleus of the stria terminalis. The studies have had different designs, including a sham stimulation–controlled “delayed start” or alternatively open-label stimulation optimization followed by masked stimulation withdrawal. Irrespective of these differences, this group of studies found improvement rates roughly comparable to each other, that is, above 50% “full response” with respect to OCD severity on the YBOCS. It is worth emphasizing that the YBOCS is “nonlinear,” meaning that at the most severe end (where all surgical candidates fall by definition), a relatively small change in score can mean a relatively large gain in symptom-free time, which is the single most important clinical feature related to improvement. Thus, YBOCS changes in groups of this type may underestimate actual improvement.

Major Depression

There is continuing research into DBS for intractable major depression. One line of work began with observations from functional neuroimaging pointing to the subgenual cingulate cortex as a node in circuits involved in the normal experience of sadness, symptoms of depressive illness, and responses to depression treatments. A small open-label study found that patients with unipolar depression refractory to four or more treatment trials improved after stimulation of white matter tracts adjacent to the subgenual cingulate cortex. A larger prospective trial had similar findings. However, a large industry-supported controlled trial of subgenual DBS was stopped early based on a “futility analysis.” The stimulation target continues to be refined based on updated hypotheses concerning which neural pathways, amenable to identification on diffusion MRI prospectively, might be most important to modulate.

Another line of research on DBS for depression was prompted by the VC/VS DBS OCD research discussed above and on observed antidepressant effects of anterior capsulotomy on which the VC/VC stimulation target was initially based. The OCD patients in both the lesion and DBS experience had high rates of comorbid depression, as expected in that population. Responders displayed rapid improvements in mood and motivation, which preceded reductions in core OCD symptoms. An open-label study of DBS at the VC/VS target for depression without comorbid OCD enrolled patients meeting a very high threshold for treatment refractoriness and found clinically significant antidepressant responses in roughly half of the patients. Induction of transient, reversible mood elevation, occasionally reaching the diagnostic threshold for hypomania, was the most significant adverse effect of active stimulation. However, similarly to the subgenual cingulate DBS industry-supported trial, an industry-led controlled trial of VC/VS stimulation for intractable depression was also stopped early due to futility, reflecting low response rates in both the active and sham stimulation groups.

CURRENT STATE OF PSYCHIATRIC NEUROSURGERY

Despite progress and the promise of advances in conventional medication and behavioral therapies, there remains a small proportion of individuals with intractable OCD, major depression, or other conditions who are potential candidates for neurosurgical treatments. These individuals remain very ill and chronically disabled despite sustained and conscientiously applied standard care. As with any treatment, the potential risks and benefits of psychiatric neurosurgery must be weighed against the potential risks and benefits of foregoing these approaches in individuals likely to remain burdened with severe illness and very poor quality of life. What remains true across eras in which stereotactic psychiatric neurosurgery has been applied is that half or more of carefully selected patients may show clinically meaningful benefit for a period of years after neurosurgeries, whether lesions or intracranial stimulation and that a subset might show outstanding improvement. Although data remain incomplete, and cases of attempted and completed suicide have continued to occur after psychiatric neurosurgery, these procedures might reduce mortality, as evidenced by data on comparative suicide rates. Nonetheless, patients who undergo and fail to benefit from these “last resort” procedures are at particularly high risk for completed suicide.

Given the current landscape, clinicians are faced with a number of issues when considering a particular clinical or research procedure for a given patient who might be judged a candidate for neurosurgery. The first, for OCD in particular, is consideration of a lesion procedure versus DBS. DBS may be acceptable to patients who would not choose to undergo lesion procedures, although the reverse is also true, given the long-term treatment context of DBS. Despite its advantages of reversibility and adjustability, DBS implantation carries risks and, importantly, requires that patients be treated by highly specialized teams willing and able to provide long-term care. The logistics and expenses involved, including inconsistent coverage by insurers, can represent significant barriers. In contrast, psychiatric care can be delivered in standard treatment settings after lesion procedures. However, although the relative risks of enduring adverse effects after psychiatric DBS remain to be clearly established, at this stage, ablative methods appear to carry a greater potential for them. Rates of adverse outcomes are relatively low when modern lesion procedures are performed by highly experienced multidisciplinary teams, something also seen with DBS for movement disorders, where there has been the most experience. This represents a particularly strong rationale for the referral of appropriate patients to expert psychiatric neurosurgery centers.

RESEARCH AND FUTURE DIRECTIONS

The best evidence on which one could base treatment recommendations would be adequately powered, long-term, prospective comparison studies using random assignment of individuals to different surgical methods. That remains unlikely to become available. As noted above, the small number of appropriate patients, the experience and preferences of expert centers for particular methods, and the preferences of patients themselves for one method over another are all impediments in this regard. In contrast, long-term follow-up studies, systematic reviews, and ongoing research pointing toward individualized predictors of response have occurred and should be encouraged.

An example is a 2015 systematic review by Pepper et al., comparing outcomes of anterior capsulotomy to those of DBS for OCD. A total of 170 cases were identified across 20 studies. DBS targets were the VC/VS or the closely related NAcc. After DBS (*n* = 62) there was a 40% decrease in YBOCS scores, while after capsulotomy (*n* = 108), YBOCS OCD severity decreased by 51%. There was a 9% higher likelihood of remission after capsulotomy than after DBS. The authors noted no overall difference in complication rates, although the pattern of complications was different (lasting adverse effects tend to be more common after capsulotomy). The follow-up period was very different: 19 months for DBS versus 61 months for capsulotomy, which may allow for further development of improvement after capsulotomy, and, conversely, might not allow assessment of how many DBS patients experienced device failures or were no longer receiving stimulation for a variety of reasons.

Another 2015 systematic review by Brown and colleagues, also based on observational data, compared outcomes of anterior cingulotomy to those of anterior capsulotomy for OCD. Ten studies involving 193 total participants were included; the capsulotomy dataset was similar to that in the review above. Mean follow-up was 47 months for cingulotomy and 60 months for capsulotomy. At the last follow-up, 41% were responders after cingulotomy, while 54% responded after capsulotomy. Serious or permanent adverse events, characterized as the number of events per procedure, was 5.2% across cingulotomy studies and 21.4% across capsulotomy studies, although in the latter, half of those events were attributable to very high-dose, multiple target Gamma Knife procedures (which no longer would be performed) across three patients. Excluding those, the rate of serious complications in the capsulotomy group dropped to 12.8%.

There are several key research directions as of this writing. One is the use of increasingly high-resolution structural and functional imaging methods that can assess changes in the relevant networks, for example, their connectivities, prospectively after lesion procedures to test and generate hypotheses about which brain network alterations are most relevant to therapeutic improvement and also to adverse effects. As is possible in studies of ablative methods, work in DBS can make use of multiple modalities of neuroimaging at baseline to test for morphometric, fiber tract, and local and network activity as individual predictors of outcomes. This is occurring as brain science advances toward an understanding that functional brain networks may be less segregated than previously thought and could represent a “mosaic” of functional networks where small differences in, for example, lesion or stimulation within a given individual could make a relatively large difference in clinical responses. Moreover, there is early but promising work in both intractable OCD and major depression, where electrophysiologic correlates of symptoms or clinically relevant behaviors are recorded and analyzed. Such advances could make closed-loop or adaptive stimulation a viable option, which could enhance symptomatic responses, patient function, and quality-of-life outcomes. That line of research could also lead to physiologically informed targeting of ablation in individual patients.

In summary, current neurosurgical treatment continues to represent a viable option for appropriately selected patients with severe and treatment-refractory illnesses. Evidence remains most robust for OCD. As evidence has accumulated, it has become increasingly clear that the patterns of risks, burdens, and benefits, while they overlap, are distinctly different for DBS versus ablation. Research intended to inform psychiatric neurosurgery is proceeding along a number of fronts. Neuroimaging techniques might be a potential aid in individual patient selection by predicting treatment response. Structural and/or functional neuroimaging might be used to tailor surgical interventions on an individual basis. Observations suggest that ablation or DBS might enhance the benefits of subsequent behavioral or medication therapies that proved ineffective before surgery. Both kinds of treatments are being studied in the context of an increasing focus on the neurocircuitry underlying both syndromic illness and dimensions of psychopathology. Insights gained from invasive methods have the potential to inform further development of noninvasive, circuit-based treatments for psychiatric illness, such as transcranial magnetic or electrical stimulation or low-intensity focused ultrasound, which, uniquely, promises to noninvasively modulate deep brain structures.

FURTHER READINGS

Avecillas-Chasin JM, Hurwitz TA, Bogod NM, Honey CR. An analysis of clinical outcome and tractography following bilateral anterior capsulotomy for depression. *Stereotact Funct Neurosurg*. 2019;97(5–6):369–380.

Boccard SGJ, Fitzgerald JJ, Pereira EAC, et al. Targeting the affective component of chronic pain: a case series of deep brain stimulation of the anterior cingulate cortex. *Neurosurgery*. 2014;74(6):628–635.

Brown LT, Mikell CB, Youngerman BE, Zhang Y, McKhann GM 2nd, Sheth SA. Dorsal anterior cingulotomy and anterior capsulotomy for severe, refractory obsessive-compulsive disorder: a systematic review of observational studies. *J Neurosurg*. 2016;124(1):77–89.

Cho DY, Lee WY, Chen CC. Limbic leukotomy for intractable major affective disorders: a 7-year follow-up study using nine comprehensive psychiatric test evaluations. *J Clin Neurosci*. 2008;15(2):138–142.

Christmas D, Eljamel MS, Butler S, et al. Long term outcome of thermal anterior capsulotomy for chronic, treatment refractory depression. *J Neurol Neurosurg Psychiatry*. 2011;82(6):594–600.

Davidson B, Hamani C, Rabin JS, et al. Magnetic resonance-guided focused ultrasound capsulotomy for refractory obsessive compulsive disorder and major depressive disorder: clinical and imaging results from two phase I trials. *Mol Psychiatry*. 2020;25(9):1946–1957.

Denys D, Mantione M, Figee M, et al. Deep brain stimulation of the nucleus accumbens for treatment-refractory obsessive-compulsive disorder. *Arch Gen Psychiatry*. 2010;67(10):1061–1068.

Dougherty DD, Rezai AR, Carpenter LL, et al. A randomized sham-controlled trial of deep brain stimulation of the ventral capsule/ventral striatum for chronic treatment-resistant depression. *Biol Psychiatry*. 2015;78(4):240–248.

Fox MD, Buckner RL, Liu H, Chakravarty MM, Lozano AM, Pascual-Leone A. Resting-state networks link invasive and noninvasive brain stimulation across diverse psychiatric and neurological diseases. *Proc Natl Acad Sci U S A*. 2014;111(41):E4367–E4375.

Garnaat SL, Greenberg BD, Sibrava NJ, et al. Who qualifies for deep brain stimulation for OCD? Data from a naturalistic clinical sample. *J Neuropsychiatry Clin Neurosci*. 2014;26(1):81–86.

Greenberg BD, Gabriels LA, Malone DA Jr, et al. Deep brain stimulation of the ventral internal capsule/ventral striatum for obsessive-compulsive disorder: worldwide experience. *Mol Psychiatry*. 2010;15(1):64–79.

Greenberg BD, Price LH, Rauch SL, et al. Neurosurgery for intractable obsessive-compulsive disorder and depression: critical issues. *Neurosurg Clin N Am*. 2003;14(2):199–212.

Greenberg BD, Rauch SL, Haber SN. Invasive circuitry-based neurotherapeutics: stereotactic ablation and deep brain stimulation for OCD.

Neuropsychopharmacology. 2010;35(1):317–336.

Haber SN, Yendiki A, Jbabdi S. Four deep brain stimulation targets for obsessive-compulsive disorder: are they different? *Biol Psychiatry*. 2021;90(10):667–677.

Hariz M. Renaissance for anterior capsulotomy for obsessive-compulsive disorder? *J Neurol Neurosurg Psychiatry*. 2022;93(3):229.

Holtzheimer PE, Kelley ME, Gross RE, et al. Subcallosal cingulate deep brain stimulation for treatment-resistant unipolar and bipolar depression. *Arch Gen Psychiatry*. 2012;69(2):150–158.

Kim SJ, Roh D, Jung HH, Chang WS, Kim CH, Chang JW. A study of novel bilateral thermal capsulotomy with focused ultrasound for treatment-refractory obsessive-compulsive disorder: 2-year follow-up. *J Psychiatry Neurosci*. 2018;43(5):327–337.

Lehman JF, Greenberg BD, McIntyre CC, Rasmussen SA, Haber SN. Rules ventral prefrontal cortical axons use to reach their targets: implications for diffusion tensor imaging tractography and deep brain stimulation for psychiatric illness. *J Neurosci*. 2011;31(28):10392–10402.

Lopes AC, Greenberg BD, Canteras MM, et al. Gamma ventral capsulotomy for obsessive-compulsive disorder: a double-blind, randomized controlled trial. *JAMA Psychiatry*. 2014;71(9):1066–1076.

Mallet L, Polosan M, Jaafari N, et al; STOC Study Group. Subthalamic nucleus stimulation in severe obsessive-compulsive disorder. *N Eng J Med*. 2008;359(20):2121–2134.

Malone DA Jr, Dougherty DD, Rezai AR, et al. Deep brain stimulation of the ventral capsule/ventral striatum for treatment-resistant depression. *Biol Psychiatry*. 2009;65(4):267–275.

Martino D, Deeb W, Jimenez-Shahed J, et al. The 5 pillars in Tourette syndrome deep brain stimulation patient selection: present and future. *Neurology*. 2021;96(14):664–676.

McLaughlin NCR, Lauro PM, Patrick MT, et al. Magnetic resonance imaging-guided laser thermal ventral capsulotomy for intractable obsessive-compulsive disorder. *Neurosurgery*. 2021;88(6):1128–1135.

Miguel EC, Lopes AC, McLaughlin NCR, et al. Evolution of gamma knife capsulotomy for intractable obsessive-compulsive disorder. *Mol Psychiatry*. 2019;24(2):218–240.

Nuttin B, Cosyns P, Demeulemeester H, Gybels J, Meyerson B. Electrical stimulation in anterior limbs of internal capsules in patients with obsessive-compulsive disorder. *Lancet*. 1999;354:1526.

OCD-DBS Collaborative Group. Deep brain stimulation for psychiatric disorders. *Neurosurgery*. 2002;51(2):519.

Pepper J, Hariz M, Zrinzo L. Deep brain stimulation versus anterior capsulotomy for obsessive-compulsive disorder: a review of the literature. *J Neurosurg*. 2015;122(5):1028–1037.

Price BH, Baral I, Cosgrove GR, et al. Improvement in severe self-mutilation following limbic leucotomy: a series of 5 consecutive cases. *J Clin Psychiatry*. 2001;62(12):925–932.

Rasmussen SA, Noren G, Greenberg BD, et al. Gamma ventral capsulotomy in intractable obsessive-compulsive disorder. *Biol Psychiatry*. 2018;84(5):355–364.

Rück C, Karlsson A, Steele JD, et al. Capsulotomy for obsessive-compulsive disorder: long-term follow-up of 25 patients. *Arch Gen Psychiatry*. 2008;65(8):914–921.

Spofford CM, McLaughlin NCR, Penzel F, Rasmussen SA, Greenberg BD. OCD behavior therapy before and after gamma ventral capsulotomy: case report. *Neurocase*. 2014;20(1):42–45.

Sullivan CRP, Olsen S, Widge AS. Deep brain stimulation for psychiatric disorders: from focal brain targets to cognitive networks. *Neuroimage*. 2021;225:117515.

Synofzik M, Clausen J. The ethical differences between psychiatric and neurologic DBS: smaller than we think? *AJOB Neurosci*. 2011;2(1):37–39.

van Wingen G, Bergfeld I, de Koning P, et al. Comment to: Deep brain stimulation for refractory obsessive-compulsive disorder (OCD): emerging or established therapy? *Mol Psychiatry*. 2022;27(3):1276–1277.

Yang JC, Papadimitriou G, Eckbo R, et al. Multi-tensor investigation of orbitofrontal cortex tracts affected in subcaudate tractotomy. *Brain Imaging Behav*. 2015;9(2):342–352.

34

Child Psychiatry

34.1 Introduction and Overview

Caroly Sue Pataki, M.D.

UPDATES IN CHILD AND ADOLESCENT PSYCHIATRY

The past decade has demonstrated major advances in evidence-based knowledge in child and adolescent psychiatry, with investigation of new disorders and significant gains in modalities to provide effective treatments to youth. Efforts to link scientific advances and psychosocial insights to more individualized treatments for childhood psychiatric disorders have been met with increasing promise. Five new chapters are included in the child and adolescent psychiatry section of this edition of the textbook. These include Internet Gaming Disorder in Youth; Social Media and Youth; Gay, Bisexual, Transgender, Queer/Questioning, Intersex, and Asexual (LGBTQIA) Youth; Racism and Discrimination's impact on the health and well-being of Developing Children and Adolescents; Forensic Evaluations of Child and Adolescent Migrants.

Since the COVID-19 pandemic, starting in 2020, there has been an explosion of telehealth utilization to provide uninterrupted treatment to children and adolescents remotely. This necessity has led to an ability to provide treatment to youth in regions that are difficult to reach in person, as well as overcoming barriers of transportation. The success of telehealth over the past 2 years has reinforced the utility of this modality and will likely have a long-term impact on the options for providing psychiatric treatment to children and adolescents in the future.

Data from large randomized controlled studies and their follow-up over the past 10 years support the efficacy of medications, cognitive behavioral therapy, and especially their combination, for youth manifesting obsessive-compulsive disorder (OCD); anxiety disorders including generalized anxiety disorder, separation anxiety disorder, social anxiety disorder; and major depressive disorder (MDD). Those youth with the most robust positive responses within the first 3 months of treatment are the most likely to maintain these gains over the long term.

Genetic research, including genome-wide linkage scans seeking to identify gene loci, has focused on several childhood disorders, including attention-deficit hyperactivity disorder (ADHD), autism spectrum disorders (ASDs), Tourette syndrome (TS), and OCD. Neuroimaging studies have elucidated the brain development of typically developing children, as well as in children with psychiatric disorders. For example, neuroimaging studies have revealed that whereas total brain volume remains virtually stable after the age of 5 years, there are two dynamic processes continue through adulthood. These processes, synaptic pruning and myelination. profoundly influence cognitive and social decision making and behavior. Furthermore, neuroimaging studies have led to the conclusion that girls and boys differ in the timing of maximal gray matter volume and its normal decline after puberty, in that the process occurs about 1 year earlier in boys than in girls. A variety of neuroimaging techniques, such as magnetic resonance imaging (MRI), functional MRI (fMRI), diffusion tensor imaging (DTI), computerized tomography (CT), single photon emission computerized tomography (SPECT), and positron emission tomography (PET) have been increasingly employed to uncover both morphology and function of various parts of the brain to further understand childhood psychiatric disorders such as ADHD, mood disorders, OCD, and anxiety disorders, including post-traumatic stress disorder (PTSD).

The *Diagnostic and Statistical Manual of Psychiatric Disorders*, Fifth Edition Text Revision (DSM-5-TR), published in 2022 integrates the criteria for the childhood disorders published in in 2013 in the DSM-5, and makes some further modifications mainly for clarity in an effort to refine and match the criteria with available scientific evidence and expert consensus. It is well known that youth with the same psychiatric disorder, as defined by the DSM-5-TR, may present with heterogeneous symptom constellations. Diagnostic categories have played an important role in providing explanations of symptoms to patients, communication between clinicians, and providing a method for researchers to devise systematic studies. A limitation of a categorical (disorder) approach, rather than a dimensional (symptom or neurobiologic domain) approach in child psychiatric treatment research is that youth often present with overlapping symptoms of several disorders or present with multiple comorbid disorders. In fact, it is quite common for children to present with more than one psychiatric disorder. This clinical heterogeneity among childhood psychiatric disorders makes it a challenge to determine which constellation of symptoms will be most responsive to a given treatment.

An in-depth description of the clinical presentation, course, and treatment as well as updates on the scientific advances in childhood psychiatric disorders is presented in the following chapters.

CLASSIFICATION OF CHILDHOOD PSYCHIATRIC DISORDERS AND HIGHLIGHTS OF MODIFICATIONS IN DSM-5-TR AND DISORDERS UNDER CONSIDERATION

The DSM-5-TR provides a categorical classification system reflecting evidence-based and expert consensus on current childhood disorders. Children and adolescents who meet criteria for the same psychiatric diagnoses often present with heterogeneous clinical pictures. DSM-5-TR defines a psychiatric (mental) disorder as *a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning. Mental disorders are usually associated with significant distress or disability in social, occupational, or other important activities.*

The first edition of the World Health Organization's *International Statistical Classification of Diseases (ICD)* to contain a section for mental illness was the sixth (ICD-6). The first formal draft of a U.S. variant of this document, the DSM, was developed shortly thereafter, in 1952, by the Committee on Nomenclature and Statistics of the American Psychiatric Association.

Dennis Cantwell, MD, along with many psychiatric colleagues, was instrumental in advancing the classification of child psychiatric disorders through suggested revisions to the DSM in an effort to facilitate empirical investigation. These efforts were first realized in the development of the DSM-III in 1980. The DSM-III included explicit diagnostic criteria for each disorder, use of the multiaxial system, and a diagnostic focus on phenomenology rather than etiology. The DSM-5-TR does not include the multiaxial system; however, it allows for specifiers to denote important psychosocial factors.

Although both categorical and dimensional approaches have merit in describing psychiatric syndromes in youth, the DSM system uses only categorical criteria to identify psychiatric disorders. The overlap of behavioral and mood symptoms across multiple childhood and adolescent disorders made defining the boundaries challenging. In addition, it soon became evident that in more severely psychiatrically impaired youth, comorbidity is the rule rather than the exception.

Highlights of changes in childhood psychiatric disorders and diagnostic criteria first added in 2013 in DSM-5 and maintained in DSM-5-TR, published in 2022, include the following:

Autism Spectrum Disorder (ASD) was first added to the DSM-5 as a new disorder 2013, and maintained in the DSM-5-TR. ASD is characterized by core deficits in two areas: (1) social communication/social interaction, and (2) restricted pattern of interests, behaviors, or activities. ASD subsumes the following five pervasive developmental disorders contained in previous versions of the DSM: autistic disorder, Asperger disorder, childhood disintegrative disorder, Rett disorder, and pervasive developmental disorder not other-wise specified.

Social (pragmatic) communication disorder—a new disorder first added in DSM-5 is characterized by persistent deficits in the social uses of verbal and nonverbal communication, however, without a restricted pattern of behaviors or interests.

Specific learning disorder—new disorder to DSM-5 in 2013 and maintained in DSM-TR encompasses the previous disorders: reading disorder, mathematics disorder, disorder of written expression, and learning disorder not otherwise specified. Specifiers indicate which skills are affected.

Disruptive mood dysregulation disorder—a new disorder in the DSM-5 and maintained in DSM-5-TR characterized by children and adolescents with persistent irritability and frequent episodes of severe tantrums.

DEVELOPMENTAL PERSPECTIVE: IDENTIFYING THE EMERGENCE OF A CHILDHOOD PSYCHIATRIC DISORDER

Onset

Onset of a psychiatric disorder is a challenging concept in child and adolescent psychiatry because it is often difficult to pinpoint reliable landmarks for sufficient symptoms and impairment. Defining the onset of a childhood psychiatric disorder must include a balance between too narrow and too broad a set of diagnostic criteria. In other words, defining the onset of a disorder by a single impairing symptom might lead to the inclusion of a heterogeneous group of disorders, whereas inclusion of a rigid set of symptoms might exclude important subtypes. Age at onset of a childhood or adolescent psychiatric disorder is a variable of considerable importance because the emergence of a known constellation of symptoms at a younger-than-expected age might reflect increased vulnerability genetically or environmentally. For example, a younger childhood onset of major depressive disorder is associated with an increased probability of depressive disorders in first-degree relatives and implies an increased genetic vulnerability.

Onset of Child and Adolescent Psychiatric Disorders. The onset of a psychiatric disorder in a child or adolescent is often difficult to pinpoint. Investigations of age at onset of a psychiatric disorder can be applied to models of delayed onset of illnesses. Developmental failure (i.e., a lack of physiologic development within the nervous system at the appropriate time) is another model of delayed onset of a disease process.

A theory of delayed onset of a disorder, the two-hit model, hypothesizes that the emergence of a disorder is related to a “primary” genetic vulnerability triggered by exposure to a second event that may be environmental. For example, the rate of emergence of bipolar disorders increases markedly during adolescence, is maximal in young adulthood, and then gradually diminishes. One model to explain the emergence of a disorder with this distribution is a genetic predisposition (first hit) leading to vulnerability for the disorder and a second exposure, possibly the process of puberty itself or an adverse life event (second hit), that results in the emergence of the disorder.

Genetic anticipation is defined as the emergence of more severe forms of a genetic disorder in successive generations. This mechanism has been shown to contribute to the severity of pathology in a number of genetic disorders that are caused by mutations of DNA triplet repeat nucleotide sequences. Increases in repeat patterns of three nucleotides (e.g., GAC GAC GAC) increase the length of DNA sequences and result in pathologic genes. Triplet repeat expansion has been shown to be the mechanism of a number of genetic disorders, including Huntington disease and fragile X syndrome. Anticipation has been suggested as a possible explanation for increasingly severe psychopathology in successive generations in psychiatric disorders. The investigation of possible molecular bases of anticipation in psychiatric disorders, such as the increased density of genetic triplet repeat expansions shown in Huntington disease and fragile X mental retardation, is on the horizon.

Neuroimaging Studies in Child and Adolescent Psychiatry

Neuroimaging studies over the past 10 years have provided increased evidence of the dynamic changes that occur in developing brains. Neuroimaging has further increased an understanding of multiple childhood psychiatric disorders including ADHD, mood disorders such as MDD and bipolar I disorder (BP-I), anxiety disorders, OCD, PTSD, TS, childhood-onset schizophrenia, substance use disorders, feeding and eating disorders, ASD, and specific learning disorders including dyslexia. For example, recent studies have found that methylphenidate, the first-line treatment for ADHD, changes frontal–striatal activation in pediatric patients with ADHD, with greater prefrontal cortex (PFC) and striatal activation as measured by fMRI demonstrated in ADHD patients.

fMRI, in studies by Swartz and colleagues, has shown, for example, that children with ASDs have hyperaroused amygdalae leading to deficits in habituating to a variety of stimuli including sounds, sad faces, and certain tactile feelings. This may contribute to the typical sensory sensitivities in children. Singh and colleagues employed fMRI to show that adolescents with bipolar disorder have less activation in superior frontal brain regions after a negative experience, compared to healthy adolescents, which may be indicative of deficits in executive control over emotional decisions. Child and adolescent brains have been described by clinicians and researchers as “works in progress,” with a series of predictable patterns of biologic processes that extend into and beyond adolescence. The neuroimaging work of Giedd, Rapoport, Rosenberg, Castellanos, and others has revealed the typical wave of high production of gray matter among developmentally normal adolescents emerging just before puberty. Furthermore, investigations have documented the links between the pruning and myelination processes in various sites in the brain and the subtle increases in social and emotional problem-solving abilities that typically emerge in later adolescence. Other investigations have identified areas of the brain targeted in children who learn first and second languages early in life. fMRI scans have determined the spatial relationship of language centers in children who have first and second languages. In such children, both language centers appear in the same cortical region whereas when a second language is acquired in adult life, the new language center is not represented in the same cortical region as the first language. MRI studies have also provided evidence that adolescents process emotional stimuli differently than adults. Using fMRI and presenting pictures of faces depicting various emotions, Yurgelun-Todd and colleagues showed that in young teens, who typically have difficulties with the task, the amygdala (a brain center for fear reactions) is activated during the task of identifying the emotions presented; whereas in older teens, who perform better on the task, the frontal lobes are more active in the process of identifying emotions in the stimuli. There were also additional findings indicating that language tasks correlated with activation of temporal lobes in young teens and of frontal lobes in older teens.

GENE X ENVIRONMENT INTERACTIONS: RISK AND PROTECTIVE FACTORS

There is a convergence of evidence regarding the interactional effects of genetic vulnerability and environmental events on the development of many psychiatric disorders. Much research has focused on the alleles of the serotonin transporter gene and sensitivity to adverse events and stress. The impact of age and developmental stage that a stressful event is experienced adds another factor to the equation, and finally, positive relational and environmental variables may increase resiliency. Risk factors include all variables that increase the probability that a given child or adolescent will develop psychopathology. Protective factors decrease the risk of developing psychopathology. Demographics, psychosocial factors, biologic factors, genetics, family environment, and external environment interact to produce risk and protection from psychiatric disorders. Rarely can a single risk factor account for the entire variance between the emergence and inhibition of a psychiatric disorder. Thus, the study of psychiatric risk is complicated and multifactorial. Risk factors often studied for influence on the emergence of psychiatric disorders among children and adolescents include family history of psychiatric disorders, social class, intellectual function, adverse life events, temperamental factors, and the quality of family and peer relationships. Developmental delays, child maltreatment, and high levels of family conflict increase the risk of psychiatric disorders. Protective factors that have been identified and studied include individual temperamental predisposition, positive family relationships, and success experiences academically and socially.

INTERVENTIONS FOR PSYCHIATRIC DISORDERS IN YOUTH

Interventions to address psychiatric symptoms and disorders in youth include clinical, targeted, and universal interventions.

Clinical interventions provide treatment to a family with a child or adolescent identified as having a psychiatric problem. Psychosocial, psychopharmacologic, and other environmental interventions can be compared with a placebo condition and with each other to assess efficacy. Approximately one-half of families who begin psychiatric treatment terminate prematurely, thus, the nature and delivery of the treatments also merit investigation.

Targeted interventions are designed for children identified as at increased risk of a psychiatric disorder but whose families are not seeking treatment. Children at risk may be identified either by an external factor, such as a family characteristic (e.g., offspring of a parent known to have a psychiatric disorder) or through a behavior (e.g., defiance or aggression in the classroom). A given child may be monitored or receive a psychosocial intervention on the basis of the risk factor—whether psychiatric disorder is already present or not;

however, these interventions may not be welcomed by the family, or threshold for interventions may not be correlated with a clinically relevant situation.

Universal *interventions,* also termed *primary prevention,* are received by all children and families within a specific geographical distribution. They may occur throughout a targeted school or community or on a citywide, statewide, or national basis. This strategy obviously reaches more children than any other method of intervention, but questions arise about the benefit and need of the intervention for the population at large, as well as the cost–benefit ratio.

Study of the risk factors and protective factors is leading the field to develop the most effective combination of prevention strategies for children and adolescents. More effective treatments are being investigated in order to match the treatment to the individual. Future goals include mitigating risk factors, enhancing protective factors, and when prevention is not possible, to provide efficacious treatments.

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Barr CL. Genetics of childhood disorders: XXII. ADHD, Part 6: the dopamine D4 receptor gene. *J Am Acad Child Adolesc Psychiatry*. 2001;40:118–121.

Blakemore SJ, Choudhury S. Development of the adolescent brain: implications for executive function and social cognition. *J Child Psychol Psychiatry*. 2006;47:296–312.

Blakemore SJ, den Ouden H, Choudhury S, Frith C. Adolescent development of the neural circuitry for thinking about intentions. *Soc Cogn Affect Neurosci*. 2007;2:130–139.

Court MH. A pharmacogenomics primer. *J Clin Pharmacol*. 2007;47:1087–1103.

Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med*. 2013;11:126.

Ernst M, Rumsey JM, eds. *Functional Neuroimaging in Child Psychiatry*. Cambridge University Press; 2000.

Franklin ME, Sapyta J, Freeman JB, et al. Cognitive behavior therapy augmentation of pharmacotherapy in pediatric obsessive-compulsive disorder: the Pediatric OCD Treatment Study II (POTS II) randomized controlled trial. *JAMA*. 2011;306:1224–1232.

Giedd JN, Rapoport JL. Structural MRI of pediatric brain development: what have we learned and where are we going? *Neuron*. 2010;67:728–734.

Gold AL, Brotman MA, Adleman NE, et al. Comparing brain morphometry across multiple childhood psychiatric disorders. *J Am Acad Child Adolesc Psychiatry*. 2016;55(12):1027–1037.

Goodyer IM. Genes, environments and depressions in young people. *Arch Dis Child*. 2015;100(11):1064–1069.

Kaufman J, Gelernter J, Hudziak JJ, Tyrka AR, Coplan JD. The research domain criteria (RDoC) project and studies of risk and resilience in maltreated children. *J Am Acad Child Adolesc Psychiatry*. 2015;54:617–625.

Kavanaugh BC, Dupont-Frechette JA, Tellock PP, Maher ID, Haisley LD, Holler KA. Neurocognitive phenotypes in severe childhood psychiatric disorders. *Nerv Ment Dis*. 2016;204(10):770–777.

Krueger RF, Hopwood CJ, Wright AGC, Markon KE. Challenges and strategies in helping the DSM become more dimensional and empirically based. *Curr Psychiatry Rep*. 2014;16:515.

Lenroot RK, Giedd JN. Annual research review: developmental considerations of gene by environment interactions. *J Child Psychol Psychiatry*. 2011;52:429–441.

Molenaar D, Middeldorp C, van Beijsterveldt T, Boomsma DI. Analysis of behavioral and emotional problems in children highlights the role of genotype × environment interaction. *Child Dev*. 2015;86(6):1999–2016.

National Institute of Mental Health. *The National Institute of Mental Health Strategic Plan.* National Institute of Mental Health; 2008. http://www.nimh.nih.gob/about/strategic-planning-reports/index.html

Nigg JT. Temperament and developmental psychopathology. *J Child Psychol Psychiatry*. 2006;47:395–422.

Piacentini J, Bennett S, Compton SN, et al. 24- and 36-week outcomes for the Child/Adolescent Anxiety Multimodal Study (CAMS). *J Am Acad Child Adolesc Psychiatry*. 2014;53:297–310.

Rauh VA, Margolis AE. Research review: environmental exposures, neurodevelopment, and child mental health – new paradigms for the study of brain and behavioral effects. *J Child Psychol Psychiatry*. 2016;57(7):775–793.

Rettew DC, Althoff RR, Dumenci L, Ayer L, Hudziak JJ. Latent profiles of temperament and their relations to psychopathology and wellness. *J Am Acad Child Adolesc Psychiatry*. 2008;47:273–281.

Singh MK, Chang KD, Kelley RG, et al. Reward processing in adolescents with bipolar I disorder. *J Am Acad Child Adolesc Psychiatry*. 2013;52:68–83.

Swartz JR, Wiggins JL, Carrasco M, Lord C, Monk CS. Amygdala habituation and prefrontal functional connectivity in youth with autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry*. 2013;52:84–93.

Teicher MH, Samson JA. Annual research review: enduring neurobiological effects of childhood abuse and neglect. *J Child Psychol Psychiatry*. 2016;57(3):241–266.

Treatment for Adolescents With Depression Study (TADS) Team; March J, Silva S, Curry J, et al. The Treatment for Adolescents With Depression Study (TADS): outcomes over 1 year of naturalistic follow-up. *Am J Psychiatry*. 2009;166:1141–1149.

Uher R. Gene-environment interactions in severe mental illness. *Front Psychiatry*. 2014;5:48.

Walsh ND, Dalgleish T, Dunn VJ, et al. 5-HTTLPR-environment interplay and its effects on neural reactivity in adolescents. *Neuroimage*. 2012;63:1670–1680.

White TJH. A window into the neurobiology of childhood and adolescent psychopathology. *J Am Acad Child Adolesc Psychiatry*. 2013;52:9–11.

Yurgelun-Todd D. Emotional and cognitive changes during adolescence. *Curr Opin Neurobiol*. 2007;17:251–257.

Züchner S, Roberts ST, Speer MC, Beckham JC. Update on psychiatric genetics. *Genet Med*. 2007;9:332–340.

34.2 Child Development

GABRIELLE SHAPIRO, M.D., D.F.A.A.C.A.P., D.F.A.P.A., ROBERT J. JAFFE, M.D., ALEXANDER KOLEVZON, M.D., AND TRESHA A. GIBBS, M.D.

Imagine an infant reciting Shakespeare, a toddler calculating trigonometry angles, or an 8 year old driving to grade school each morning. The thought of children performing mundane adult tasks highlights the obvious: humans are born with a very limited skillset and develop a more complex one over time. In essence, development is the acquisition of new skills and the multiplex reactions to these additions. Whether it is the capacity to walk, talk, empathize, or think about the future, suddenly, a capacity is there that previously was not. At no time in life is the explosion in newfound possibilities more pronounced than in early childhood, the focus of this section.

With a basic understanding of development, clinicians of every discipline can benefit their patients. Pathology can frequently be traced back to childhood, offering clues as to its inception and psychological sequelae. Often, adult patterns of symptom presentation or behavioral interaction with the environment mimic patterns from childhood.

Until the 1600s, children were viewed as “small adults” and expected to work and socialize in an adult manner. Children were sources of cheap labor during the Industrial Revolution. This era’s end, coupled with compulsory school laws in the United States, meant that by the early 20th century, the vast majority of children were attending schools. John Dewey was essential in viewing schools as a vehicle through which childhood development could be fostered. G. Stanley Hall formally studied children and adolescents and blazed a trail for the relatively recent surge in psychological theories by Sigmund Freud, Erik Erikson, Jean Piaget, and John Watson, to name a few.

Though the field remains ripe for further exploration, some critical concepts have emerged over the past 100 or so years. Normal development can be conceived of as what a typical child will go through in an average expectable environment. This section will highlight the ages when different changes are typically seen. It is imperative to understand that each new advance has a range of when its existence should be expected. Significant lags are thought of as developmental delays (with the exception of premature babies), whereas marked abnormalities in development are developmental disorders.

There is a debate as to whether development is a continuous or discontinuous process. What is clear is that the direction of the curve is not always positive, as there is a normal degree of progression and regression. It is normal during the acquisition phase of new skills for them to come and go. An example is a fluidly speaking 3 year old who, upon the birth of a younger sibling, will revert to baby talk. This is distinct from a frank loss of a solidified skill, a symptom which should always prompt inquiry.

Current research offers evidence that the mechanism for growth comes from the interplay of genetics and the environment. Though the relative contribution of each remains an area of debate, this transactional model has become the established paradigm. Genetic background primes an individual and provides potential, but can be altered positively and negatively based on environmental stimulation. Identical twins reared in the same household do not develop identical personalities. There is evidence that children shape their environment and elicit different responses from their environment, affecting how they develop. Some aspects of development are more genetically determined, whereas the environment may play a greater role in others. When pertinent, this section will highlight relative contributions.

This chapter will cover nine areas of development from birth through adolescence. The domains of brain, language, cognitive, social, emotional, motor, sensory, gender, and psychosexual development will each be addressed. Developmental lines are also interdependent, maturing on their own yet influenced by one another. For example, language fosters advances with emotional and social development.

Finally, the scope of this chapter is on normal childhood development only. Clinical correlates will be interspersed throughout to highlight the importance of understanding normal development in practice. There is also a comprehensive master table (Table 34.2–1) arranged by age on one axis and developmental domain on the other. This is intended as a quick reference to identify skills typically developed at a certain point in time or to follow one domain longitudinally.

BRAIN DEVELOPMENT

Unlike some other species, humans are born with an immature central nervous system (CNS). An infant is entirely dependent on others and is governed by innate reflexes, yet in time becomes a creature capable of complex cognitive, social, and emotional capabilities and remarkable physical and verbal skills. How does one go from someone who cannot hold his own head up to an athlete who can sense his orientation in space and manipulate his body accordingly? This section will focus on the physical changes to the brain during infancy and childhood which allow for all the capacities described in the rest of the chapter.

Table 34.2–1.
Child Development Resource Table

Periods:		Gestational–Infancy: 0–2 Years				Early Childhood (2–5 Years)			Middle Childhood (5–11 Years)			Adolescence	
	Theories	0–2 Months	2–7 Months	7–12 Months	1–2 Years	2–3 Years	3–4 Years	4–5 Years	5–7 Years	7–9 Years	9–11 Years	11–14 Years	14 Years on
Emotional development	(Integrated theory) Acquisition of basic emotions	Stage 1: Infant has innate emotions and emotional responsiveness—emotions are available and easily stimulated at birth			Stage 2: Self-conscious emotional reactivity	Stage 3: Connecting emotion with effects; emotional self-awareness			Stage 4: Develops an emotional responsiveness to the feelings of others				
	Emotional skills	Touching evokes love, loud noises evoke fear. Restrictions evoke rage. The brain pathways for the forming of emotions begin to develop	Emotions begin to become self-regulated as the brain pathways of emotion grow	Self-regulation of emotion continues to develop	Shame, pride, envy, and embarrassment appear. Displaces onto other children	Understands the sources of many emotions and begins to develop techniques for regulating emotions and expressing them Begins to identify with adults as a coping mechanism			Displays ability to react to the feelings of others, indicating a heightened sense of awareness and empathy				
	Emotional behavior	Infants have a "social smile" in response to human voices. They display the ability to respond to emotions of others	Begin to laugh and display more control over smiles. Anger is able to be shown as well	Ability to induce responsiveness increases	Begins to display empathy, as seen by expressions of feeling. The infant enjoys attention and approval and likes to play alone and next to peers	An increase in understanding leads to increased empathy. Self-regulation increases and the child becomes more responsive than reactionary. Aggression develops as a competition and by the age of 5, the child exhibits sensitivity to criticism and compassion			Until the age of 6, ego rules. At 7, empathy transforms into altruism, and the superego dominates				
Social development	John Bowlby's attachment theory	Preattachment (0–6 weeks)	Attachment making (6 weeks to 8 months)	In month 8, clear-cut attachment begins to develop >> stranger anxiety		2 years on, reciprocal relationships begin to develop							
	Margaret Mahler's attachment theory	Autistic phase (0–6 weeks)	Develops a symbiotic relationship with the caregiver (6 weeks to 8 months), and the beginnings of differentiation	Differentiation continues	Rapprochement (18 months to 2 years)	The development of consolidation and individuation							
	Robert Selman's stages of perspective taking						Level 0 (years 3–6): Beginning at age 3, child recognizes that the self and others can have different thoughts and feelings, but frequently confuses the two. (undifferentiated perspective taking)		Level 1 (years 5–9): After age 4, can understand social information; can understand that different perspectives may result because people have access to different information. (Social-informational perspective taking)		Level 2 (years 7–12): Begins to exhibit self-reflection. Now aware that they have ideas and attitudes, they look to friends to guide their self-discovery. Child can understand other's perspectives; can "step into another's shoes" to view their own thoughts, feelings, and behavior from the other person's perspective. (Self-reflective perspective-taking)	Level 3 (years 10–15): Can understand how the self is viewed from the perspective of a third-party; Progression from two-party situation. (Third-party perspective-taking)	Level 4 (14 to adult): Individuals understand that third-party perspectives are influenced by larger societal values. (Societal perspective-taking)
	Erik Erikson's social stages (psychosexually influenced)	Stage 1: Crisis: Basic trust vs. mistrust Virtue: Hope			Stage 2: Crisis: Autonomy vs. shame and doubt Virtue: Will		Stage 3: Crisis: Initiative vs. guilt Virtue: Purpose		Stage 4: Crisis: Industry vs. inferiority (6–11 years) Virtue: Competency			Stage 5: Crisis: Identity vs. role confusion Virtue: Fidelity, ability to live by society's standards and expectations.	
	Social behaviors	Infant acknowledges the caregiver as "different," recognizing caregiver's smell, voice, touch	Infant is able to differentiate self from caregiver and begins to manifest separation anxiety		Child revises constancy of caregiver and develops trust based on caretaker's ability to provide	Child begins to display independence, separating from the caregiver when given clear goals. Parents and caregivers help children develop a sense of autonomy by allowing the child to make choices	Child begins to assert their power and control over the world through directing play and other social interactions. By making comparisons to their peers, children develop a sense of purpose	Child begins to understand different perspectives	Continues to understand foreign views and preferences	Through social interactions, both at school and home, child begins to develop a sense of pride in their accomplishments and abilities. This helps develop a sense of competency	Although truly beginning at year 7, the child's ability to get motivation from others continues to develop from years 9–11	Can understand the point of view for third party, and from this understanding develops a secure personal identity and desires independence	Understands personal values and goals and can see issues through social value context
Psychosexual development	Freudian psychosexual stages	Oral stage			Anal stage		Phallic stage		Latency stage			Genital stage	
	Gender identity				Gender labeling and beginning of gender stability	Gender is usually fixed	Gender consistency			Gender constancy			
	Gender role				Stereotyping begins to appear	Emerges		Consolidates	Stereotyping fixed but becomes less rigid with increases understanding	Child looks to social cues to recognize gender			
	Gender relationship				Sex not an issue with peer choice	Sex play common	Same-sex peers favored			Increased masturbation and heterosexual play		Sexual activity begins	
	Psychosexual skill	Ego directs sucking needs			Child labels stereotypes through objects: "girls like pink" (1–1.5 years). Child can identify the sex of self, a picture, or other children by using clothes, toys, hair, activities, as proxies (1.5–2.5 years). Child lacks control over bowels	Child knows that gender label is a part of self but does not see it as constant. I.e., "I can't do that or I'd be a girl"	The child's id transferred to genitals and begins to view activities as related to gender and not causing gender. At this point, gender consistency develops, the child understands gender cannot be changed by objects or activities		Superego consolidates, and gender constancy develops, gender is constant regardless of clothing or behavior			Sexual activity increases	
Moral development	Piaget's stages of moral behavior related to cognitive phases							Egocentric stage (preoperational cognitive)		Incipient cooperation (7–11)			
	Piaget's two-stage theory of moral judgment (cognitively based)							Heteronomous morality: Moral realism: morality imposed from the outside morality comes from an authority figure and disobeying the rules results in immediate and severe punishment (immanent justice). Rules are fixed, divine like, view a large amount of accidental damage as worse than a small amount of deliberate damage (no concept of motive)		Autonomous morality (moral relativism): Rules no longer come from a divine-like source. Child can negotiate rules with peers (concept of reciprocity). Child is able to account for motives. Punishment becomes restitution vs. retribution. Child begins to argue about punishment because they understand justice is imperfect, punishment should be adjusted to crime			
	Kohlberg's six-stage theory of moral reasoning (cognitively, socially, and emotionally influenced)							Preconventional level: Judge the morality of an action by its direct consequences		Stage 1: (Obedience and punishment driven): Bases moral decisions on fear or self-interest. Stage 2: Relativistic orientation (self-interest driven). Child develops a concern for fairness and seeks rewards obtained by bartering "If I do my chores, what's in it for me?" (sees morality as rules for society's order). Stage 3: (social consensus determines good intentions): Tries to live up to societal notions of good and bad. Conforms to societal norms. Stage 4: (Social order obedience driven): It is important to obey the laws of system in order to maintain a functioning society. Teen and adult morals replace the preoccupation with what is "fair" for the individuals to what is necessary for society's needs; social conscience		Stage 5: Sees laws as social contracts. What is good for society is paramount Stage 6: Develops universal ethics and focus on fairness and justice. Recognizes others' moral standards	

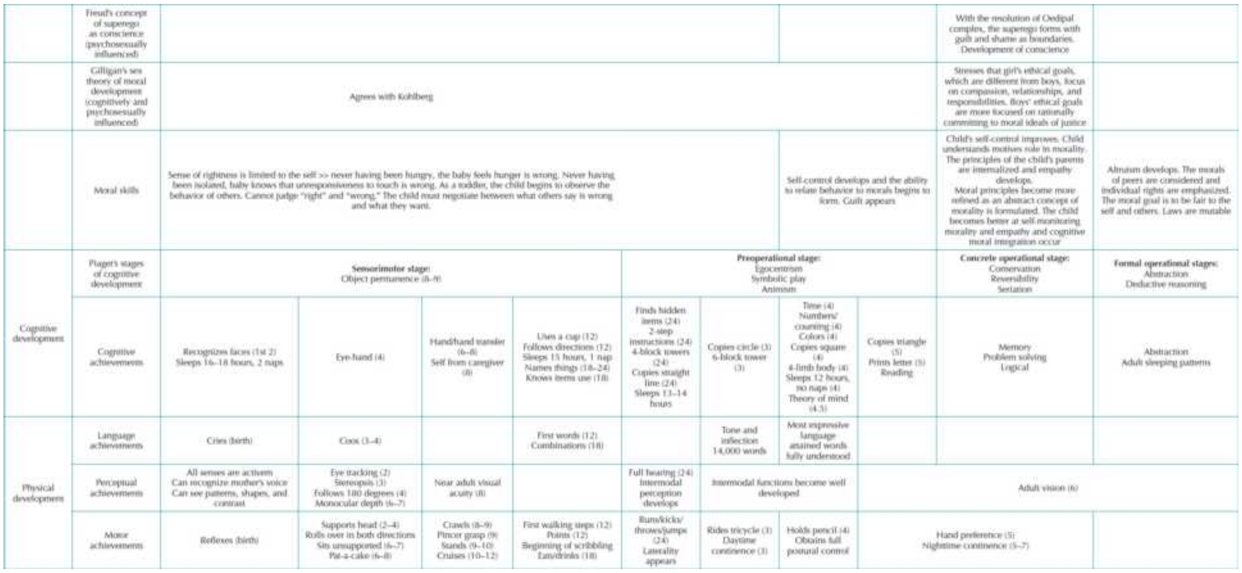

	Freud's concept of superego as conscience (psychosexually influenced)									With the resolution of Oedipal complex, the superego forms with guilt and shame as boundaries. Development of conscience	
	Gilligan's sex theory of moral development (cognitively and psychosexually influenced)	Agrees with Kohlberg								Stresses that girl's ethical goals, which are different from boys, focus on compassion, relationships, and responsibilities. Boys' ethical goals are more focused on rationally committing to moral ideals of justice	
	Moral skills	Sense of rightness is limited to the self >> never having been hungry, the baby feels hunger is wrong. Never having been isolated, baby knows that unresponsiveness to touch is wrong. As a toddler, the child begins to observe the behavior of others. Cannot judge "right" and "wrong." The child must negotiate between what others say is wrong and what they want.						Self-control develops and the ability to relate behavior to morals begins to form. Guilt appears		Child's self-control improves. Child understands motives role in morality. The principles of the child's parents are internalized and empathy develops. Moral principles become more refined as an abstract concept of morality is formulated. The child becomes better at self-monitoring morality and empathy and cognitive moral integration occur	Altruism develops. The morals of peers are considered and individual rights are emphasized. The moral goal is to be fair to the self and others. Laws are mutable
Cognitive development	Piaget's stages of cognitive development	**Sensorimotor stage:** Object permanence (8–9)				**Preoperational stage:** Egocentrism Symbolic play Animism				**Concrete operational stage:** Conservation Reversibility Seriation	**Formal operational stages:** Abstraction Deductive reasoning
	Cognitive achievements	Recognizes faces (1st 2) Sleeps 16–18 hours, 2 naps	Eye-hand (4)	Hand/hand transfer (6–8) Self from caregiver (8)	Uses a cup (12) Follows directions (12) Sleeps 15 hours, 1 nap Names things (18–24) Knows items use (18)	Finds hidden items (24) 2-step instructions (24) 4-block towers (24) Copies straight line (24) Sleeps 13–14 hours	Copies circle (3) 6-block tower (3)	Time (4) Numbers/counting (4) Colors (4) Copies square (4) 4-limb body (4) Sleeps 12 hours, no naps (4) Theory of mind (4.5)	Copies triangle (5) Prints letter (5) Reading	Memory Problem solving Logical	Abstraction Adult sleeping patterns
Physical development	Language achievements	Cries (birth)	Coos (3–4)		First words (12) Combinations (18)		Tone and inflection 14,000 words	Most expressive language attained words fully understood			
	Perceptual achievements	All senses are activem Can recognize mother's voice Can see patterns, shapes, and contrast	Eye tracking (2) Stereopsis (3) Follows 180 degrees (4) Monocular depth (6–7)	Near adult visual acuity (8)		Full hearing (24) Intermodal perception develops	Intermodal functions become well developed		Adult vision (6)		
	Motor achievements	Reflexes (birth)	Supports head (2–4) Rolls over in both directions Sits unsupported (6–7) Pat-a-cake (6–8)	Crawls (8–9) Pincer grasp (9) Stands (9–10) Cruises (10–12)	First walking steps (12) Points (12) Beginning of scribbling Eats/drinks (18)	Runs/kicks/throws/jumps (24) Laterality appears	Rides tricycle (3) Daytime continence (3)	Holds pencil (4) Obtains full postural control	Hand preference (5) Nighttime continence (5–7)		

There are two important and related concepts affecting development: *critical periods* and *plasticity.* Critical periods are windows of opportunity for growth in certain areas. This window offers an opportunity for tremendous gains but is also vulnerable to disruption. For example, the first few years appear to be important for language, in particular, the first year for a perceptual map of phonemes. Plasticity enables the ability of one part of the brain to compensate for another. An example is the primary auditory cortex in congenitally deaf individuals which assists in processing visual and somatosensory stimuli.

Early in life, there are many changes occurring within the brain. Along with the rest of the body, the brain is growing substantially during the first few years. By age 5, it has tripled in size and is at 95% of its adult mass (1.2 kg for adult women and 1.3 kg for men). Growth occurs bottom-up in a caudocranial manner with the pons, medulla, and spinal cord ahead of the cerebellum, which is ahead of the cortex. This explains why infants rely on reflexes and instincts and lack coordination.

Histologically, the adult brain consists of billions of neurons configured into trillions of synapses. Early on it is a hotbed of activity. Neuronal *proliferation* and *differentiation* explode in utero and continue into the first year of life, particularly in the cortex. Though neurogenesis slows and was previously thought to halt, recent evidence has shown that two areas in adult brains possess the ability for new neurons, the subventricular zone and hippocampal dentate gyrus. These newly formed fetal and infantile neurons must get from their place of origin to their destination, which they accomplish in a task called *migration.* This journey is what lays down cortical columns and continues through the first 6 to 7 months of life, with the later neurons settling in the cortex. Once the neurons have reached their destination (and in certain cases even during migration), the cells experience dendritic and axonal outgrowth. This extension is particularly heavy during the first year and continues over the following 18 months.

The settled and grown neurons rapidly interact with one another with an explosion of new synapses during the first 2 to 3 years of life. By age 2, the toddler has roughly twice the number of synapses an adult will have. This number gets reduced as the child's synapses are *pruned* from 18 months to age 20, with the majority occurring between 7 and 15 years. During this process, the brain becomes more energy efficient, strengthening the connections that are needed while ridding itself (and the metabolic demand) of those that are not. Just which synapses are deemed necessary is a direct reflection of one's environment, a nature–nurture interaction which is fundamental to brain development.

To communicate with each other, the nerve cells need receptors and neurotransmitters. Serotonin appears first, followed by norepinephrine and dopamine, with acetylcholine and GABA produced later in fetal life.

Finally, the neurons become myelinated from birth to adulthood, a process of insulation which increases transmission speed. In general, *myelination* occurs first in the spinal cord and lower brain, with the cerebellum fully myelinated by age 4, the corpus callosum at 8 to 12 years of age, and the cortex from age 10 into young adulthood.

The manner in which these brain changes occur is a result of complex interactions between genes and environment. Biologically, genes initiate the process of brain development through organizer genes and homeoboxes such as *Sonic hedgehog, Delta,* and the *Hox* genes among many others. The endocrine system is also involved with insulin-like growth factor (IGF) and thyroid hormones playing roles. Environmentally, illnesses and toxic exposures can alter development. Diet and nutrition is thought to have an effect, as industrialized nations demonstrate faster growth than do developing ones. For example, an increase in docosahexaenoic acid may be linked to greater visual acuity. By kindergarten and even earlier, some children from disadvantaged backgrounds have been shown to lag behind in language and vocabulary and may have heard 30 million fewer words in the home. Neglect in infancy has been linked to a significantly reduced midsagittal area of the corpus callosum. Clearly, genes alone do not account for brain development.

Genetic programing allows for brain development so that the infant can learn new skills. These capabilities facilitate interactions with the environment, which then precipitates further genetic modification. For example, motor skills afford one the chance to throw a ball and learn what happens to it. Repetition stimulates the same neural pathways, strengthening the synapses. Neurotrophins and prostaglandins affect synaptogenesis and long-term potentiation, a process that underlies learning and memory. Underuse of a neuronal pathway due to lack of exposure leads to its pruning and atrophy, mediated by genes. In short, genes lead to abilities which feedback and control other genes, affecting synapses and pathways. Thus, both the skills governed by genetics as well as the environment in which they evolve are integral to development.

LANGUAGE DEVELOPMENT

Language is so essential and omnipresent in everyday life, it is hard to imagine that it does not exist at birth. Though the ability to *communicate* is present from day one, language development is a process that takes years to develop. Speech delay is a common cause for clinical presentation, and language delays may be associated with later psychiatric comorbidities. Moreover, an undiagnosed speech delay is present in patients presenting for other reasons. Therefore, an understanding of normal language development is indispensable for all providers.

The development of language begins in utero, where a fetus has been shown to recognize its mother's voice. On delivery, an infant clearly lacks the ability to speak, but yet still requires a way to demonstrate needs. This is achieved through facial expressions such as cries. An attuned parent can often tell what a baby needs (such as a feeding or diaper change) based on the particular type of cry. Though the exact timing of each step may differ, the overall process is quite constant. At about 3 to 4 months, typical infants make cooing (vowel) sounds followed by single-syllable consonant babbling. These are learned by imitating others and carry no meaning at this age. Repetitive vowel–consonant sounds (such as mama) may appear, but without meaning these are not considered first words. Meaningful words are produced at about 12 months, with single words intended to convey entire phrases. It is not unusual for words to come and go at this stage, and loss of a spoken word should not be cause for alarm. At 18 months, word combinations appear and about 25% of words are intelligible to strangers. At 24 months, there are about 300 words, 50% of which are intelligible. By 30 months there are 500 words. At 36 months, 75% are intelligible to strangers.

Grammar develops significantly from months 20 to 36. As children learn the rules of speech, incorrect verb tenses are common ("I throwed the ball"). The same child may have used this correctly in the past, which likely reflected an imitative phrase rather than a novel one, and is considered normal. By age 4, most expressive language is developed and appropriately used. Nearly 100% of words should be intelligible to strangers, even though pronunciation mistakes may be present. Over the next 2 years, children are able to connect sentences with one another logically. At ages 5 to 6, language centers slow their development, highlighting the importance of early language and the challenges of learning a second language later in life.

From a theoretical perspective, Noam Chomsky emphasized the universality of linguistic structure and that all humans develop the ability to understand and produce language. This stands in contrast to a behavioral theory of language, championed by B. F. Skinner. Skinner believed that children imitate others and have their language reinforced through the environment. Factors such as large family size and lower social class correlate with later development and fewer words. Expanding on these environmental variables, the quality of verbal stimulation plays an important role. Hearing words from reading is more effective than just hearing spoken words; significantly more “unique words” are perceived through books, likely exposing children to words they would not otherwise encounter. Recent fMRI studies show that children with a higher reading exposure had greater activation in the left parietal–temporal–occipital association cortex, integrating the auditory with the visual and potentially stimulating imagery and narrative comprehension.

A 24-month-old boy is referred from the pediatrician for severe behavioral outbursts. Upon taking a history, you discover that he only knows about 25 words and cannot make word combinations. You astutely hypothesize that the tantrums are the manner in which this child is expressing himself, and aim to identify the reason for the speech delay. Your differential includes hearing and vision difficulties, and vision and hearing tests are ordered. Deficits in other areas are explored for signs of a broader developmental delay including intellectual disability and autism spectrum disorder. Environmental contributors such as neglect or trauma must be ruled out. An auditory processing disorder is also a possibility. Finally, you conclude the delay is due to a speech or language learning disorder. A speech and language referral helps identify the nature of the delay and breaks speech down into further subcomponents: articulation (pronouncing sounds), syntax (the arrangement of words in a sentence), semantics (meaning of words), and pragmatics (appropriate social use of words). With speech therapy, the child makes linguistic gains and his behaviors subsequently improve as he now has other means to regulate and express himself.

COGNITIVE DEVELOPMENT

Cognition can be conceived of as what happens to sensory information once it reaches an individual. How this occurs is substantially more detailed and involves perception, understanding, memory, learning, reasoning, problem solving, and language. A discussion on cognitive development would be incomplete without mentioning *Jean Piaget.* He proposed four stages of cognitive development, with changes occurring through *assimilation* and *accommodation.* Assimilation, he argued, is taking in new information and fitting it in with an existing understanding, or *schema.* It is fitting new data into old ideas. This contrasts with accommodation, in which the new experience or sensory data causes the old schema to be altered so that it is in accord with this new information. For example, a child has an understanding of a tree, which is comprised of a trunk and branches. In the summer when leaves appear, the basic concept of tree remains the same but it has expanded to include this new knowledge. Later on, the child encounters a bush and calls this a tree as well. The child is taught the differences between a bush and a tree and a new bush schema is formed. Though parts of Piaget's theories have come under question, he added tremendous insight into how children think differently from adults. This has consequences for everyone interacting with children, and is especially noteworthy for those in the educational and health care systems. Given his significance to the field, this section will be structured around Piagetian stages.

The first of Piaget's stages is *sensorimotor,* occurring between birth and approximately age 2. As the name implies, at this stage, the infant is mostly limited to sensory information and simple motor responses. Within 2 months, the infant can pay attention to faces and even recognize them. They can follow things with their eyes. By 4 months, there is eye–hand coordination, as the infant can reach for something he sees. Recognizing familiar faces and important objects improves and extends to greater distances. At 6 months, they show curiosity for things out of reach, and once within reach, they tend to put them in their mouths. Eye tracking further develops over the next few months, as the infant can even follow the path of something as it falls. Picked-up objects can be transferred from hand to hand smoothly. At 9 months, peek-a-boo is a particularly fun game, notably because at this age, the infant is grappling with *object permanence.* To an infant lacking this skill, anything that cannot be perceived cannot exist. By the age of 1, there are clear signs that schemata are forming and are being used. A 12 month old typically can follow directions and use things correctly, such as a cup. These schemata strengthen and broaden over the next 6 months and at 18 months, they often clearly show knowledge of what a typical item is used for, such as a brush. As schemata are forming, a sense of humor can be seen when these expectations are playfully breached. Following simple instructions such as “sit down” implies an ability to understand language and know what to do with it. This implicates gains in memory and learning. By the age of 2, the individual can find hidden items, clearly demonstrating the achievement of object permanence. Make-believe games begin to take on a pleasurable role. Rudimentary organizational skills are demonstrated by sorting shapes and colors. Two-step instructions can be followed and four-block towers can be built.

The second stage Piaget described is *preoperational,* ranging from about 2 to 7 years of age. Symbolic play is a key feature of this time period. Pretend play and make believe, role-playing, and symbolization (such as a stick representing an airplane) appear. A hallmark of this stage according to Piaget is that toddlers exhibit *egocentrism,* meaning they lack the ability to take another’s point of view. Children in this stage lack the understanding of *conservation.* The classic example is if equal amounts of water are poured into a tall skinny glass and a short wide glass, those who have not mastered conservation yet will say the tall skinny glass has more water in it. This skill develops from ages 6 to 11 and may differ for various units of measurement.

Clinical Correlate

With egocentrism, children cannot take on another’s point of view and are susceptible to misattribution. This normal stage in development leaves one vulnerable should a major stressor occur. For example, loss of a parent, divorce, or terrorist attacks demand explanation, and typical children with egocentrism are likely to blame themselves for the event. Though with age they may cognitively realize the situation was not their fault, the initial memory and experience was recorded in that way and guilt and self-blame may be present. Whether treating a child in real time or an adult decades later, recognizing where the patient was developmentally at the time of tragedy is paramount for processing.

At age 3, children can copy a circle, complete three- to four-piece puzzles, and build a six-block tower. By age 4, they can understand the idea of same and different. They can start to understand time, begin to understand numbers and counting, draw two- to four-limbed bodies, name colors, and copy a square. At 5, they can count to 10, draw six-part bodies, copy a triangle, and start printing some letters.

The third Piagetian stage is *concrete operations,* from roughly ages 7 to 11. Order, rules, and organization are hallmarks of this time period. As the name implies, children at this age have difficulty with abstract and hypothetical thinking but are learning how to logically think about concrete things. Children implement inductive reasoning, generalizing from the specific (an example would be dad drives a red car, therefore all dads drive red cars). Deductive reasoning remains a challenge, requiring the child to understand something specific based on a generalized rule. Two other important cognitive skills emerge during this stage: *conservation and reversibility.* Children who have mastered conservation now realize that, as in the example above, there is the same amount of water in both glasses regardless of height. In addition, children in this age group develop the capacity to see things from someone else’s point of view, marking the end of egocentrism. This contrasts somewhat with theory of mind research. *Theory of mind* is the ability to recognize someone else as another entity with their own thoughts and feelings and emerges between ages 4 and 5. Learning *that* someone else thinks differently than oneself is one thing, anticipating *what* that person may be thinking or feeling is more complex and develops later.

This ability and others further develop in Piaget’s final stage *formal operations* from age 12 and older. Here, deductive reasoning develops, with many math and science implications. The ability to think abstractly appears, allowing for better planning and organization and problem solving. Trial and error is no longer necessary to come to a conclusion, results can be anticipated. Though cognitive milestones have been discussed in the context of Piagetian stages, it is worth noting other developmental theories which help predict how and why children acquire the aforementioned skills. *Information processing theory* holds that children are goal-directed problem solvers who work to achieve certain ends. Improvements in memory, processing speed, planning, and logic help facilitate learning and solving which leads to more efficient approaches and greater knowledge. Change occurs continuously, gradually, and more subtly than in a stage-based paradigm as Piaget proposed. Similarly, *Lev Vygotsky* postulated that children continuously grow as social creatures heavily influenced by their social and cultural environments. The *dynamic systems theories* attempt to incorporate many factors viewing children as constantly evolving products of their past and present adapting to the context of their physical abilities and social environment.

Sleep wields tremendous influence on cognitive and physical development and has its own unique process. At birth, infants sleep 16 to 18 hours per day, including nighttime sleep and their two or more daytime naps. By age 1, there is typically just one nap, and daily totals are down to 15 hours. This is further reduced to 13 to 14 hours by age 2 and 12 hours by age 4, with naps typically ceasing. Common adult patterns of 7 to 10 hours are seen by ages 10 to 14. On polysomnography, rapid eye movement (REM) sleep dominates the first year. By ages 3 to 4, deep non–rapid eye movement (NREM) sleep achieves a more adult-like share. It is worth noting that sleep can be highly individualized and influenced by culture. The importance of sleep cannot be overstated, with impacts on memory, learning, and physical growth. Recent evidence has shed light on the glymphatic system, which eliminates neurotoxic waste (proteins including beta-amyloid and metabolites) from the CNS during sleep. Finally, some sleep phenomena which in and of themselves are not pathologic, including nightmares, are often reported, typically at age 5. Sleep terrors (also called night terrors) and sleepwalking occur in

NREM stage 3 and are not uncommon in healthy children. Other normal sleep phenomena include hallucinations while falling asleep (hypnagogic hallucinations) and on waking (hypnopompic).

SOCIAL DEVELOPMENT

Social maturation is traditionally divided into four phases beginning with the attachment period and ending with late adolescence. There are many theories that attempt to explain social development (see Table 34.2–1).

Attachment theories explain the development of the sense of self. “Attachment is a deep and enduring emotional bond that connects one person to another across time and space” (Ainsworth, 1973; Bowlby, 1969). Infants have their primary needs met by a caretaker, typically a parent, and the emotional bond that forms between them becomes the foundation for the child’s social, emotional, and cognitive development. This *affective bond* provides the context for the “emerging pattern of dyadic regulation of emotion” (Stroufe, 2000).

The infant’s behavior directed toward the caregiver becomes organized based on needs; thus, the quality of this attachment impacts their survival both biologically and psychologically. A healthy attachment supports the infant’s exploration and mastery of the world. There are phases in the development of attachment that begin with *preattachment* at ages 0 to 6 weeks. From 6 weeks to 8 months, attachment progresses based on the responsiveness of the caretaker, and from 8 months to 2 years, *reciprocal* relationships begin and the child organizes around their acquisition of mobility and the “secure base experience.”

The historical leaders in attachment theory include John Bowlby, Harry Harlow, Rene Spitz, Mary Ainsworth, and Mary Main. Disorders of attachment are included in the *Diagnostic and Statistical Manual for Mental Disorders,* fifth edition (DSM-5), and its text revision (DSM-5-TR) and their criteria will be reviewed in the chapter on abnormal child development.

Bowlby developed his theory by borrowing from evolutionary thinking and derived from the idea that the attachment of mother and infant was “due to a species-typical behavioral pattern to promote infant survival.” He also incorporated ideas from Piagetian cognitive theory and the concept that “with increasing schema comes increasing complexity of thought” and applied them to the intensifying connection of the infant’s attachment behaviors.

Two researchers in the 1940s and 1950s provided the seeds for the idea that an infant has more than basic physical needs. Psychologically, the infant needs and craves human contact. This was demonstrated in the Harlow studies of rhesus monkeys’ preference for cloth monkeys and in Spitz’s *hospitalism* research which confirmed a failure to thrive in infants who had their biologic needs met but did not have a primary attachment figure with whom they had physical contact.

Spitz described the concept of *anaclitic depression*, where separation from the caregiver leads to depression, withdrawal, and loss of milestones.

Ainsworth’s *Strange Situation Experiment* identified four types of infant attachment:

Avoidant—needs frequently not being met by the caretaker
Securely Attached—needs were generally met by the caretaker
Ambivalent—needs inconsistently met
Disorganized–Disoriented—correlated with a history of maltreatment

Margaret Mahler’s more complex theory proposes a sequence of attachment which, if successful, enables an individual to establish a secure sense of self. An infant who is securely attached and has established *basic trust* also experiences appropriate separation anxiety when away from the primary attachment figure.

Mary Main has developed an attachment interview for adults to assess the quality of their childhood attachments. She also described a new attachment classification to apply to Ainsworth’s Strange Situation test, the disorganized/disoriented attachment classification.

Advances in biotechnology support the possibility that attachment begins in utero. Thomas Verny, MD, in his book *The Secret Life of the Unborn Child* (1981), began the discussion of this and the idea of the “attachment pregnancy,” and guides for parents to foster healthy in utero attachments have been popular recently.

The changing dynamics of gender roles for parents as well as the changing identity of the nuclear family have led to new research which has generally supported the concept that there are no great differences in quality of attachments in lesbian, homosexual, or transgender families with their children.

Successful secure attachments in infancy correspond to later psychosocial functioning and are characterized by better peer relationships, more cooperation with parents and teachers, and greater security in unfamiliar social situations and in the exploration of new environments. A child’s sense of self is generally more secure with positive early attachments, which are the foundation of emotional development. There are many consequences of abnormal attachments that will be discussed in the sections on psychopathology including feeding problems, disruptive behavior disorders, and impaired ability to sustain lasting relationships.

Morality

There are societal and cultural norms that guide one’s behaviors and actions. The internalization of these standards is a process that begins at birth when infants become aware that their behavior elicits a corresponding behavior in their caretaker, such as when a parent hugs a smiling happy infant. This cause-and-effect relationship is observed by infants and young children and becomes the basis for the rules that developing children assimilate and accommodate into their moral foundation. Children feel sadness and shame when they are punished or see that their behaviors bring displeasure to their parents or other important adult figures. Therefore, behavior is shaped by the fear of punishment, and life can be explained and organized within the framework of cultural rules and a sense of fair play. Children tend to go from the fear of punishment to the desire to confess. As children enter adolescence and reach puberty, they develop guilt and the ability to be critical of their own behaviors and transgressions. In many cultures, this is the age when children's lives become their responsibility and their conscience must be their guide.

There are many different theories of moral development but the most famous is that of *Lawrence Kohlberg,* which built on the foundations of Piaget's theory of cognitive development. Kohlberg's work was criticized for being too culturally biased toward the middle-class male and, despite revisions over the decades, is still controversial. *Carol Gilligan* in 1982 offered an alternative theory, which corresponds to morality growing in parallel to empathy. Her theory had a gender basis, with males' moral development developing into a system dominated by "the ethics of justice and the assertion of rights" whereas "woman's ends in compassion and an ethic of caring."

EMOTIONAL DEVELOPMENT

The ability to acquire and express emotions and feelings is essential to adult functioning. Learning to identify, understand, express, communicate, and modulate affect is key to successful relationships. An infant's task is to learn to be emotionally responsive and not react to internal feelings and anxiety. This is achieved through the maturation and the development of coping skills. There are many theorists who have attempted to describe emotional maturation, with achievements highlighted in Table 34.2–1.

More recently, theories integrating the neurobiology of brain maturation with social interaction have focused on the interaction between genetics, environment, and behavioral patterns to explain emotional development and maturation.

All of the theories to describe emotional maturation can basically be divided into four stages with basic milestones. Stage I is the acquisition of basic emotions that are learned through the interaction of the infant with others, including the use of mimicry. Mirror neurons, which are neurons that fire when one acts or observes an action, are proposed as being an essential pathway for this acquisition of a repertory of emotions by the infant. Stage II involves the development of self-conscious emotions, which are derived from approval or disapproval experienced and include shame, envy, embarrassment, pride, and guilt. The ability of emotional self-regulation is the task of stage III, which continues into middle childhood. Stage IV is marked by true empathy, the ability to understand and share the feelings of another.

Temperament

Temperament has fascinated scientists since ancient times when Galen of Pergamum made his prophecy “that much of human behavior can be predicted according to an individual’s temperament.”

Stella Chess and Alexander Thomas made important contributions to the study of temperament in the 1970s. They defined three major clusters which are dependent on the perceived sense of ease with which children interact with their environment. The three different styles were assessed through the categorization of nine types of behavior that they defined as dimensions of temperament, including activity level, regularity/rhythmicity, approach or withdrawal, adaptability, threshold of responsiveness, intensity of reaction, quality of mood, distractibility, and attention span and persistence. Their longitudinal study sample showed the following distribution which was generalized to apply to the general population:

One in 10 children had a *difficult* temperament.
One in six was categorized as *slow to warm up.*
Four in 10 were characterized as having an *easy temperament.*

The *goodness of fit* between parent and child as described in their book, is an important factor in working with parent–child relational problems therapeutically. Jerome Kagan, inspired by Galen, also did work on temperament and described two main types of temperament: inhibited and uninhibited. He stated that “temperament refers to stable behavioral and emotional reactions that appear early and are influenced in part by genetic constitution.” Behavioral geneticists are now exploring specific genetic polymorphisms that contribute to temperament. Some research cites polymorphisms in serotonin transporter and promoter region genes as having an effect on personality development. In addition, polymorphisms in the dopamine, serotonin, and norepinephrine transporter genes affect temperament.

MOTOR DEVELOPMENT

Perhaps more than any other developmental domain, gross and fine motor development is highly intertwined with progress in other areas. Motor development begins in utero, and is experienced by a pregnant mother as kicking. At birth, a premature CNS renders infants with little control over their actions. At this stage, reflexes are prominent, such as the grasping, rooting, sucking, swallowing, startle, and Moro. These reflect CNS functioning and disappear at predictable time periods.

Cortical columns are being laid down during the first 6 to 7 months of life and synapses are exploding in number, allowing for greater complexity and control of actions. At 2 to 4 months of age, infants can support their own head and make smoother movements in their extremities, demonstrated by shaking their toys. Some infants may be able to roll over from their stomach to back, and by 6 months, they can roll over in both directions. Also at 6 to 7 months, they can sit without support and may start to rock back and forth as if they are trying to crawl, which they will be doing over the next 2 to 4 months. A developing sensory system involving all modalities but particularly hearing, proprioception, and vision facilitate the emergence of new skills, and will continue to help them hone. Hand-to-hand transfer occurs at 7 to 8 months, first grossly then more finely, aided by a *pincer grasp.* This is when something is held between the thumb and index finger, first seen at about 9 months, and with practice develops strongly over the next year.

Clinical Correlate

Infants are unable to roll over from their stomachs to backs before 6 months, a primary reason the American Academy of Pediatrics has recommended babies be placed to sleep on their backs in order to reduce the chance of sudden infant death syndrome (SIDS).

A 9- to 10-month-old infant can stand with support and may start walking while holding onto furniture, a practice called *cruising.* Walking is typically reported at 12 months, with a normal range of 9 to 15 months. Highlighting the interdependent nature of development, the newfound ability to walk corresponds with Margaret Mahler's separation–individuation subphases of practicing and rapprochement.

By 18 months, toddlers can drink from a cup, eat with a spoon, and help dress themselves. Athletic abilities emerge at age 2, with running, kicking, throwing, jumping, climbing, and copying straight lines. Simultaneous cognitive development is crucial as these skills require a coordinated effort involving attention and planning. A 3 year old's athletic skills become more refined, and the child can climb stairs and show off the skill of riding a tricycle. By the age of 4, movements continue to improve, with catching skills and hopping abilities. Visuomotor coordination is needed and spurs development through activities such as sports and art, the latter aided by an ability to hold a pencil appropriately. These actions strengthen synapses in the developing cerebellum, which is fully myelinated by age 4, and in turn facilitates further coordination. This level of dexterity allows pouring and simple food-cutting behavior to occur, which is much improved by age 5. At this time, the child can engage in basic gymnastics, such as somersaults, and can hop and begin to skip. Hand preference is typically established by age 5. The ability to tie shoelaces generally appears within the next year or 2.

Toilet training is influenced by culture, but myelination of pyramidal tracts allows for sphincter control by 18 months, and children can start toilet training as early as age 2. They typically develop bowel control before the bladder, and the vast majority of Western children are toilet trained by ages 3 to 4. Nighttime bladder control is typically achieved months to years later between ages 5 and 7.

As with all development, it is imperative to keep in mind that the listed ages should be considered normal ranges. There is a high degree of variability, and the amount of delay, as well as pervasiveness of delays, should inform clinical suspicion of abnormality. Typically a 3- to 5-month delay in the early milestones (sitting without support, standing, holding on, first steps) is granted before there is a concern, and a larger 10- to 15-month window for later complex achievements (such as climbing stairs, jumping). Family history is important, as gross motor development is considered to be highly influenced by genes. Environment and stimulation are, of course, important too, particularly in fine motor development, such as utensils or pencil use, and in sharpening gross motor sporting skills. Inquiring whether a skill has been lost or never developed is also a crucial piece of clinical history.

SENSORY DEVELOPMENT

Imagine walking on a crowded city street at rush hour: neon lights flashing, sirens wailing, mysterious scents wafting, and fellow pedestrians obliviously making contact. In a mere instant, the body is flooded with a barrage of stimuli. That the mind can perceive, organize, and make sense of it all to the point where such a moment would be considered unremarkable is in itself quite remarkable.

Infants are born with all of their primary senses working in some capacity, some more developed than others. Within the first year, however, there is tremendous growth across the sensory modalities. As with other domains, there is a split into how important nature or nurture is in sensory development. For senses fairly functional at birth, genetics play a large but not exclusive role. For those areas that require more development after birth, any predisposition becomes dependent as well as modifiable via environmental factors.

Hearing is functional and fairly developed by birth, and in fact, there is evidence that even in the womb that infants can distinguish their mother’s voice. Despite this functionality, there is evidence that the softest sound an adult can hear would need to be amplified four times for an infant to perceive it. Children eventually reach adult levels of acuity as early as age 2. At birth, infants will turn their heads toward a sound in a phenomenon known as *auditory localization* and demonstrate a preference for voices. That infants can recognize music should come as no surprise, given the melodic nature of lullabies. Infants can discern consonant tones from dissonant ones and can be seen bouncing up and down to the music's beat.

Given how much humans rely on vision, it is surprising that this sense is the least developed at birth. In the early months, infants require great contrasts in color to make distinctions. Their color-sensing retinal cones are immature and inefficient at capturing light. Their visual acuity is quite poor, with vision estimates ranging from 20/120 to 20/800 at birth. There is an explosion of improvements made as the cells that will make up the visual cortex proliferate, reach their destination, and then make and break connections. During the first 6 months, 18,000 new visual synapses have the potential to form. Infants have near-adult levels of visual acuity by 8 months, and by age 6, their vision is full.

A number of visual experiences appear to be present at birth. Research has shown that infants possess *perceptual constancy,* where an object is known to be a constant size or shape. Infants also possess the depth cue known as *optic expansion,* where an object getting closer is perceived to be approaching. They can also differentiate human faces. Eye tracking begins at 2 months, and infants are drawn to moving stimuli. Along with this newfound skill comes the ability to take in more of an object and process it. This phenomenon is particularly noteworthy when it comes to looking at faces, as 2 month olds will scan a face much more comprehensively and will spend significantly more time looking at eyes and mouths. This has implications for social, emotional, and language development.

At 3 to 4 months, depth perception is improved via *stereopsis,* where the quickly maturing visual cortex uses the slightly different images perceived by each eye to calculate depth. At 6 to 7 months, infants can demonstrate monocular depth, perceiving depth with one eye.

Infants (much like adults) have a preference for sweets, beginning in utero. An ingenious study by de Snoo (1937) demonstrated that when saccharin was injected into the amniotic fluid, the fetus would drink more of it and excrete it via their mothers, helping to decrease high amniotic fluid volumes. Odor perception is inferred to take place in the womb as well, based on the fact that amniotic fluid has discernible scents and that this fluid comes into contact with olfactory receptors. After birth, infants show clear preferences for their mother’s scent.

Touch is inferred to begin in utero, when a fetus may be seen grasping the umbilical cord, touching other parts of its developing body, and even sucking its thumb. Learning one’s orientation in space helps develop proprioceptive skills. Initially, after birth, infants will explore new objects orally, putting everything in their mouths. They eventually develop stronger motor and visual skills and learn their environment through touch.

The human mind is able to integrate stimuli from various senses (and even multiple stimuli of just one sense) into a cohesive explanation through what is called *intermodal perception.* This ability allows one to successfully navigate the busy city street, all while listening to music.

GENDER IDENTITY

There are multiple theories of gender identity development, including biologic, cognitive, developmental, and social learning theory. There are differences between gender identity, sexual identity, and sexual orientation. Gender development begins in utero. During the first trimester, fetuses with a Y chromosome have hormonal masculinization produced by the release of androgens, and feminine hormones and development are inhibited. The fetus with only X chromosomes develops ovaries at the end of this trimester. This is essentially how infants are born, with most of the development of secondary sex characteristics beginning in latency for girls and at preadolescence for boys.

A person’s *gender identity* can be the same or different from their sexual assignment at birth.

The concept of gender identity is defined as the person’s personal experience of whether they are male or female. A person comfortable with their gender assignment at birth is termed *cisgender.* People with gender identity disparity and discomfort can be gender nonconforming or *transgender,* and some societies define a third or other gender. Societal gender standards are typically binary and impact a developing child's social identity.

Most theories and research support that core gender identity is formed by age 3. Gender dysphoria is a marked incongruence between one's experienced or expressed gender and assigned gender. Gender dysphoria can be caused by attempting to reassign a child's gender and can lead to serious consequences, including depression and severe anxiety. It is thought that, between the ages of 6 and 8, gender identity has coalesced. However, recent evidence suggests that a percentage of people may later shift identities after puberty. Due to the historical changes in Western societal attitudes toward sexuality, research is being done on gender identity development and disorders as well as parenting styles such as gender-neutral parenting that are helpful for the gender nonconforming child.

The *S.A.G.E. Test,* the Sex and Gender Explorer, is one of many new tests developed to assess gender identity.

Gender roles and gender relationships in society have changed drastically within the last century, as have sexuality and the acceptance of nonbinary sexuality; as social identities change with this new acceptance of different gender and sexual roles, so will parenting and child-rearing practices. The days of defining gender and sex roles in infant toys, clothes, and play are changing. The United States military now allows women in combat, and men are staying home to care for children. Mental health professionals need to assist in maximizing a child's development by supporting an understanding and acceptance of new forms of family, gender, and sexual identities.

RACIAL AND ETHNIC IDENTITY DEVELOPMENT

A child's development in the area of racial and ethnic identity is a process that begins early and evolves over time. Since race is a social construct that is primarily based on the child's phenotype and systematized societal hierarchies, some may wonder what place it has in childhood development. Ethnicity and culture, on the other hand, refers to the shared customs, norms, beliefs, and history that contribute to child’s sense of belonging to a community as they develop their sense of self. The development of a child’s racial and ethnic identity involves the intersection of healthy childhood social development with established social hierarchies, values, and preferences. In the social context that is characterized by the privilege of majority groups, there is often prejudice, bias, and discrimination toward minority groups. These norms are learned and internalized by children of majority and minority groups.

The developing brain is intricately involved in processing race-based information. The identification of in-group and out-group faces is a critical function played by the fusiform face area, anterior cingulate, amygdala, and anteromedial prefrontal cortex, which mediate and interpret this information. There is evidence that the developing brain already begins to categorize faces into groups. In the first year of life, it is developmentally appropriate for infants to demonstrate a preference for in-group faces and tend to maintain an in-group preference for the first 2 years of life. As the child approaches age 3, there is more explicit labeling of color groups, and children are able to draw a self-portrait reflective of their particular skin color. As toddlers, some children have already inferred social hierarchies and will demonstrate majority racial group preference if given a choice. In the United States, this is evidenced by a pro-White bias. While not pathologic, it reflects an acquired internalization of the societal hierarchy. During ages 6 to 10, children, in their effort to apply logic to the social hierarchy, may begin to create rules around why racial hierarchy exists and thus give them some rational basis (this racial group is not hardworking; therefore, they don’t have money or another racial group is better off because they work harder). Positive and negative social stereotypes are thereby reinforced in their young minds, particularly if not challenged. Also, during ages 6 to 10, explicit racialized language may be subdued as youth learn to avoid talking about race explicitly. During late childhood and adolescence, as they work to consolidate their own social identity, many of the early messages about social position, stereotypes, and hierarchies become more fixed ideologic positions. At the same time, it may be more difficult to discern implicit biases which are not endorsed and are out of awareness.

There are proactive strategies to promote healthy racial/ethnic identity development. For children from majority racial/ethnic groups, proactive discussions about human diversity, race, and ethnicity can lead to discussions about individuality, fairness, and the negative impact of bias. Developing relationships with children from diverse racial groups can diminish established social stereotypes. Avoiding majority group children's exposure to biased and prejudiced language and ideology which does not encourage critical thinking about established racial hierarchy can be impactful. For children from a majority race/ethnicity, the social context already privileges their identity. However, specifically for minority youth, efforts to imbue positive racial/ethnic identity through positive or neutral representations of the racial group in media, role modeling, and other areas of the social context can diminish stereotypes and promote a positive racial/ethnic identity. Also, efforts to promote a more equitable society overall can reduce race-based social hierarchies.

PSYCHOSEXUAL DEVELOPMENT

Sigmund Freud's contributions to psychiatry and understanding of the mind are the subjects of numerous textbooks and even institutions in and of themselves. As this chapter is dedicated to development, this section will solely focus on a portion of Freud's theory, that of psychosexual development and its contributions to character formation, from his 1905 piece *Three Essays on the Theory of Sexuality.*

Freud postulated that children go through successive stages in childhood: oral, anal, phallic, latency, and genital. Their names correspond to the erogenous zone from which Freud believed *psychic energies* attach to (or cathect) at different times of life. Psychic energies are instinctual drives that motivate the individual, and at the time, this theory proposed that drive was sexual (aggressive motives would later be emphasized as well).

Unresolved conflicts can arise at any time during any phase, leading to a variety of personality types. In the *oral phase* (from birth until about age 1), an infant's main desire, according to Freud, is to breastfeed (or bottle feeding—a warm caregiver is more important than the mode of feeding). Toward the end of the first year comes the emergence of teething and biting, aggressive tools. Undergratified suckling needs (such as neglect or emotional indifference) can lead to jealousy, envy, and a seeking and consuming nature. The typical example is the oral fixation, someone constantly chewing or smoking. A clinging dependency may also be seen and manifested by being unwilling or afraid to let go of someone or something for fear there will not be more. On the contrary, one whose needs have always been met may expect things will always work out and maintains an optimistic nature. This belief may lead to passivity and inactivity, on the other hand, and the potential for fewer coping skills when a need is not met. An infant's sexual (and aggressive) drives can be met with numerous responses by his caretaker, which can lead to a fear of losing the caretaker. To placate these fears, the baby will implement a defense such as denial of aggression, which, while adaptive in infancy, can be destructive later in life.

In the *anal stage* (ages 1 to 3), the psychic energy has shifted, and control becomes a central theme: control over one's autonomy, control over one's impulses, control over one's body, and sphincters. Parental demands increase, and compliance and rebellion with these requests are at odds here, contributing to the "terrible twos." Whereas the oral stage is a symbiotic one between the caregiver and the infant, the newly developed abilities to walk and talk make the anal stage marked by separateness, or the *dyad.* Frustration and aggression abound as the toddler struggles to meet expectations and master control. Tolerating ambivalence toward others, learning to titrate aggression, and accepting that caregivers will not leave or withhold their love is crucial for a successful resolution. It is in this stage where the anal-compulsive character originates, with the Freudian hallmark triad of pedantry, parsimony, and miserliness. A love of order and cleanliness embodies this person, who translates into the obsessive-compulsive personality type. The opposite is one who lives in disordered chaos.

The phallic phase (ages 3 to 6) is marked by the child's sexual desires for their opposite-sex parent. An obstacle to this achievement is the same-sex parent, who elicits murderous wishes from the child. In boys, this is known as the *Oedipus complex,* and in girls, *Electra.* These forbidden thoughts evoke fears of bodily injury, notably castration anxiety in boys. Fear of injury is indeed typically seen during this time. These fears, as well as impractical realities, eventually force the child to give up his quest and alternatively identify with the same-sex parent. In doing so, he internalizes his parents' rules and mores, ending the phallic phase with the formation of a superego. There are now three entities involved (child plus mom and dad), and conflicts in this phase lead to issues in triadic relationships. Unresolved conflicts in this stage may lead to myriad traits, from self-evident Oedipal ones to difficulties with following rules.

Latency follows from the resolution of the Oedipal complex until puberty. Freud believed this time was one of relative calmness sexually, though not entirely static. Psychic energy is invested in academic and social avenues as the child goes through grade school. This phase is dominated by school and peers, learning socialization, and rules. Values and ideals are solidified as the newly formed superego evolves from an external presence to an internalized one.

Though outside the scope of child development, following latency, Freud described the genital phase, where the genitals (as in the phallic stage) again become the source of libidinal energy, but now the pubertal adolescent has biologic potency. This newfound ability confers its own set of conflicts, and the teenager moves from within the family to outside the family.

It is important to bear in mind that these character types are simplifications. In addition, this theory is psychological and does not account for innate tendencies, which were indeed noted by analysts as well. Finally, when conflicts occur, they are from repeated behavior patterns between a child and the important caretakers. Though many are skeptical about the veracity of Freud's theories, his contributions are irrefutable. The impact of early experiences remains crucial to understanding an individual, as does the relationship with family members. The significance of an unconscious—that feelings occur outside of one's awareness—has an enormous bearing on how one conceptualizes the mind and treats patients clinically.

CONCLUSION

In summary, development is the acquisition of new skills or abilities and the various responses both intra- and interpersonally. Humans are born with a genetic background, and the environment contributes to which pathways get further strengthened or pruned. Changes in the brain allow for new possibilities in the developing child. Small advances begin to accumulate, and more possibilities arise. By the end of childhood, a person who started as mostly innate reflexes has become capable of writing essays about the future, performing complex athletic activities, establishing a sense of identity, and engaging in deep reciprocal relationships. However, although this section has focused on child development and all the significant achievements that are made during that period, it is important to note that development continues throughout the lifespan.

ACKNOWLEDGMENTS

The authors would like to acknowledge Maureen Fulchiero Gordon, MD, who authored the last edition of this chapter, and Jean Thomas, MD, for their contributions to this chapter.

FURTHER READINGS

Adler SA, Gerhardstein P, Rovee-Collier C. Levels-of-processing effects in infant memory. *Child Dev*. 1998;69:280–294.

Arbel S, Benjamin J, Golin M, Kremer I, Ebstein RP, Belmaker RH. Relation of shyness in grade school children to the genotype for the long form of the serotonin transporter promoter region polymorphism. *Am J Psychiatry*. 2003;160:671–676.

Bangoo RK, Leibenluft E. Affective neurons and the state of normal and abnormal emotional regulation. *Child Adolesc Psychiatr Clin N Am*. 2002;11(3):519–532.

Berenbaum SA, Bailey JM. Effects on gender identity of prenatal androgens and genital appearance: evidence from girls with congenital adrenal hypertrophy. *J Clin Endocrinol Metab*. 2003;88(3):1102–1106.

*Brainard MS, Knudsen EI. Images in neuroscience. Brain development, V: experience affects brain development. *Am J Psychiatry*. 1998;155:1000.

Cassidy J, Jones JD, Shaver PR. Contributions of attachment theory and research: a framework for future research, translation, and policy. *Dev Psychopathol*. 2013;25(4 Pt 2):1415–1434.

Coley RL. Children's socialization experiences and functioning in single-mother households: the importance of fathers and other men. *Child Dev*. 1998;69:219–230.

Demiris Y, Meltzoff A. The robot in the crib: a developmental analysis of imitation skills in infants and robots. *Infant Child Dev*. 2008;17:43–53.

Eisenberg N, Fabes RA, Shepard SA, Guthrie IK, Reiser M, Murphy BC. Parental reaction to children's negative emotions: longitudinal relation to quality of children's social functioning. *Child Dev*. 1999;70:513–534.

*Fisher SE, Lai CSL, Monaco AP. Deciphering the genetic basis of speech and language. *Annu Rev Neurosci*. 2003;26:57–80.

Galliese V, Keysers C, Rizzolatti G. A unifying view of the basis of social cognition. *Trends Cogn Sci*. 2004;8:396–403.

Giedd J. Brain development, IX: human brain growth. *Am J Psychiatry*. 1999;156(1):4.

Huffman LC, Bryan YE, del Carmen R, Pedersen FA, Porges SW, Doussard-Roosevelt JA. Infant temperament and cardiac vagal tone: assessments at twelve weeks of age. *Child Dev*. 1998;69:624–635.

Kaufman AS. Generics of childhood disorders II: reports I and II. Genetics and intelligence. *J Am Acad Child Adolesc Psychiatry*. 1999;38:626–628.

Kubota JT, Banaji MR, Phelps EA. The neuroscience of race. *Nat Neurosci*. 2012;15(7):940–948.

Kurtz-Costes B, Defreitas SC, Halle TG, Kinlaw CR. Gender and racial favouritism in black and white preschool girls. *Br J Dev Psychol*. 2011;29(Pt 2):270–287.

Merriman J. Linking risk-taking behavior and peer influence in adolescence. *Neuropsychiatry Reviews*. 2008;9(1).

Muris P, Meesters C, Merckelbach H, Sermon A, Zwakhalen S. Worry in normal children. *J Am Acad Child Adolesc Psychiatry*. 1998;37:703–710.

Pleak RR. Gender-variant children and transgender adolescents. *Child Adolesc Psychiatr Clin N Am*. 2011;20(4):xv–xx.

*Plomin R, Rutter M. Child development, molecular genetics, and what to do with genes once they are found. *Child Dev*. 1998;69:1223–1242.

Quinn PC, Lee K, Pascalis O. Face processing in infancy and beyond: the case of social categories. *Annu Rev Psychol*. 2019;70:165–189.

*Rizzolatli G. The mirror neuron system and its function in humans. *Anat Embryol (Berl)*. 2005;210:419–421.

*Robinson JL, Zahn-Waxler C, Emde RN. Patterns of development in early empathic behavior: environmental and child constitutional influences. *Soc Dev*. 1994;3:125–146.

Schore AN. Effects of a secure attachment relationship on right brain development, affect regulation, and infant mental health. *Infant Mental Health J*. 2001;22:7–66.

Stillwell BM, Galvin MR, Kopta SM, Padgett RJ. Moral volition: the fifth and final domain leading to an integrated theory of conscience understanding. *J Am Acad Child Adolesc Psychiatry*. 1998;37:202–210.

Tangney JP, Miller RS, Flicker L, Barlow DH. Are shame, guilt and embarrassment distinct emotions? *J Pers Soc Psychol*. 1996;70:1256–1269.

Tettamanti M, Buccino G, Sacculman MC, et al. Listening to action-related sentences activates fronto-parietal motor circuits. *J Cogn Neurosci*. 2005;17:273–281.

Xiao NG, Wu R, Quinn PC, et al. Infants Rely More on Gaze Cues from own-race than other-race adults for learning under uncertainty. *Child Dev*. 2018;89(3):e229–e244.

Yip T, Cheon YM, Wang Y. The diversity paradox: opportunities and challenges of “contact in context” across development. *Res Hum Dev*. 2019;16(1):51–75.

Zanh-Waxler C, Radke-Yarrow M, Wagner E, Chapman M. Development of concern for others. *Dev Psychol*. 1992;28:126–136.

34.3 Adolescent Development

Ardis C. Martin, M.D., and Caroly Sue Pataki, M.D.

Adolescence is a critical period of neurobiologic and physical maturation leading to increased psychological awareness and more sophisticated social, cognitive, and emotional engagement with peers and adults. Striking features of social and emotional growth in adolescence include (1) increased affiliation with a peer group, (2) striving for more autonomy from parents, (3) increasing ability to “mentalize,” that is, understand what others are thinking and feeling, (4) an increase in risk-taking behaviors, and (5) greater sensation seeking. A convergence of data from neuroimaging studies, including the work of Blackmore and Mills (2014) have been able to identify certain regions of the adolescent brain specialized to produce increased mentalizing, and other regions of the brain that remain immature, but are adaptive to changing environmental demands through their “plasticity.” The varying levels of maturation in diverse regions of the adolescent brain lead to a seeming paradox, for example, between an adolescent’s increased ability for empathic social thought process and at the same time reckless risk-taking behaviors. This apparent “mismatch,” that is, varying degrees of maturity across different regions of the adolescent brain has been referred to by some researchers as the “Dual System Model.” The work of Strang, Chein, and Steinberg (2013) supports a view that increased risk-taking in midadolescence may be, in part, due to an increased reactivity of the social and emotional regions of the brain in the context of less mature cognitive control brain development. The social and emotional reactivity system is believed to be localized mainly in the ventral striatum and ventromedial prefrontal cortex (PFF), whereas the cognitive control system is believed to be localized in the lateral prefrontal, parietal, and anterior cingulate cortices. Emerging evidence from neuroimaging studies supports a correlation between increased adolescent social awareness and mentalizing abilities and activity in three specific brain regions including the medial prefrontal cortex (mPFC), the posterior superior temporal sulcus (pSTS), and the temporoparietal junction (TPJ). However, as older adolescents transition into young adulthood, there is a decrease during mentalizing tasks in the activation of the mPFC. Thus it appears that during the development of mentalizing abilities, the mPFC is a critical brain region, however, once the process is developed, in adulthood, this region no longer predominates, and other brain regions may take over this function. Giedd (2014), and other colleagues, using neuroimaging to track adolescent brain development find that gray matter in the brain, such as the PFC, does not grow larger after puberty, on the contrary, it matures by becoming more specialized, that is, by a process termed “pruning,” which ultimately leads to a diminished size. Gray matter is at its peak volume in girls at approximately age 11 years, and in boys, at approximately 13 years. On the other hand, white matter, representing myelinated brain regions continues to increase throughout childhood and adolescence. The continued increase in white matter reflects increasing "connectivity" and continued ability of the brain to adapt to new cognitive demands throughout adolescence. Thus, the increasing and less specialized white matter growth allows for greater cognitive adaptive brain function, with a resulting delay in its maturation in adolescents. Consequently, adolescents experience a longer period to learn and respond to novel cognitive information and develop new responses, however, this delays the brain maturation guiding cognitive control and decision making.

DEFINITION

Adolescence is the period of maturation between childhood and adulthood heralded by the physiologic signs and surging hormones of puberty. It is a time when peer relationships deepen, autonomy grows, and intellectual pursuits intensify and a sense of social belonging is sought. Adolescence is a time of accelerated social and psychosexual exploration, culminating in an integrated concept of "self." Adolescents can be viewed as neurobiologic "works in progress," often consumed by academic, interpersonal, and emotional challenges, and exploring new territories using their talents, and experimenting with social identities.

WHAT IS TYPICAL IN ADOLESCENCE?

The conceptualization of adolescence as a continual struggle, with its defining feature of intense psychic pain and severe turmoil, was a historical notion. The hypothesis that most adolescents experience enduring severe psychological distress, intense depressive symptoms, severe mood lability, and a disruption of their usual conduct was not borne out by epidemiologic studies. Although certain psychiatric disorders, such as major depression, schizophrenia, and bipolar disorders, emerge with greater frequency in adolescents than in younger children, the majority of adolescents are free of psychiatric disorders. For up to approximately 75% of youth, adolescence is a period of successful adaptation to physical, cognitive, and emotional changes, largely continuous with their previous functioning. Thus, the majority of adolescents display reasonable autonomy in decision making, increasing independence in completing academic work, and successful peer relationships without significant alienation from their families or the emergence of conduct-disordered behaviors. Psychological maladjustment, self-loathing, disturbance of conduct, substance abuse, affective disorders, and other impairing psychiatric disorders emerge in approximately 20% of the adolescent population. Adolescent adjustment is, in general, continuous with previous psychological function; thus, psychologically disturbed children are at greater risk for psychiatric disorders during adolescence. Adolescents with psychiatric disorders are at increased risk for greater conflicts with families and for feeling alienated from their families. Although up to 60% of adolescents endorse occasional distress or a psychiatric symptom, this group of adolescents functions well-academically and with peers and describes themselves as generally satisfied with their lives. Studies conducted since the 1980s indicate that most adolescents endorse more acceptance than rejection of their parents' values and lifestyles, and few report feelings of isolation or alienation from their families.

The developmentalist Erik Erikson characterized the normative task of adolescence as identity versus role confusion. In other words, adolescents, faced with the task of applying childhood talents, strengths, and experiences to their evolving physical and emotional selves, must be able successfully to change “who they are” so as to build a more mature identity. The integration of past experiences with current changes takes place in what Erikson termed ego identity. Adolescents express various aspects of their psychological selves by becoming fans of heroes or other well-known musical or political idols. Some adolescents appear consumed by their identification with a particular idol, whereas others are more moderate in their expression. Adolescents who feel accepted by a peer group are less likely to become consumed by adoration of an idol. Adolescents who are socially isolated, feel socially rejected, and become overly identified with an idol to the exclusion of all other activities are at higher risk for serious emotional problems and require psychiatric evaluation.

Adolescence is a time to search for ideals and social values that can be carried into adulthood. Erickson used the term moratorium to describe the interim period between the concrete thinking of childhood and a more evolved complex abstract thinking and ethical development. He defined identity crisis as a normative task of adolescence in which adolescents pursue alternative behaviors and styles and then successfully mold these different experiences into a solid identity. A failure to do so results in identity diffusion, or role confusion, in which there is a lack of cohesive or confident sense of identity. Successful development creates the primary psychological virtue of fidelity, characterized by the ability to relate to others and form genuine relationships—setting the stage for the next developmental phase, intimacy versus isolation. Adolescence is the time to bond with peers, experiment with new beliefs and styles, fall in love for the first time, figure out who one wants to be and where one fits in, and explore creative ideas for future plans. The literature leads to the conclusion that the majority of adolescents go through this developmental process with basic optimism and good self-esteem, maintain good peer relationships, and sustain basically harmonious relationships with their families.

CULTURAL CONSIDERATIONS

Racial and Ethnic Identity Development During Adolescence

According to the 2018 to 2019 U.S. Census Bureau report, teenagers age 10 to 19 make up 13% of the U.S. population and are rich in ethnic and racial diversity. Of that 13%, 50% are white, 25% are Latinx, 14% are Black, 5% are Asian, 4.5% are multiracial, 1% are American Indian or Alaska Native, and 0.5% are Native Hawaiian or Pacific Islander. First- and second-generation immigrants make up 26% of all children ages 0 to 17 years.

Erickson posited that the goal of adolescence is the formation of a solid and positive sense of self. He believed the process of self-exploration could prove challenging for adolescents whose cultural identity is not supported by their own culture nor validated by their community and foresaw the potential problems faced by youth from diverse racial and ethnic backgrounds.

When looking at adolescents' identity development, providers would be remiss if they don't look at the role of ethnic and racial identity development. Ethnic and racial identity is a crucial part of adolescents' discovery of their sense of self-worth and self-image–how others see them, how they see themselves, and for teens of color, how their race becomes a salient component of their lives. Researchers have found that ethnic and racial attitudes start to emerge around ten years of age, and examining one's racial and ethnic identity begins in middle school and junior high. Racial and ethnic identity is essential to all youth, but especially for teens of color. In addition to the tasks of separating and individuating from their families, establishing a sense of autonomy, coming to terms with their sexuality/gender, identifying their own religious and political views, adolescents of color must also develop their identity as it corresponds to their racial and ethnic identity. They are trying to navigate where they belong, identify within a particular race or ethnicity, share values, and develop a sense of self.

William Cross's racial identity development model has been a robust framework for understanding the process that teens of color, particularly African Americans, go through when developing their racial identity. The model includes the following phases: pre-encounter, encounter, immersion–emersion, and internalization. Adolescents in the pre-encounter stage may consciously or unconsciously devalue or have little to no awareness of their race and how it impacts their life, concurrently valuing the majority culture's ways and values. In the encounter stage, the adolescent experiences an event, positive or negative, that makes them evaluate themselves from a racial perspective; for teens of color, this is typically caused by their first-time encounter with discrimination or racism, which challenges their worldview and sense of belonging. In the immersion/emersion stage, teens of color begin to explore their racial identity, engage more within their respective cultures, and develop a sense of pride. In the internalization phase, adolescents can incorporate into their self-concept, a sense of self-confidence and secure identity.

Several other models were created to examine the racial and ethnic identity development of people of color, immigrants, and their majority counterparts. Janet Helms looks at white identity development, W.S. Carlos Poston looks at biracial identity development, and John Berry looks at ethnic/immigrant identity development. Jean Phinney examined the various models and, finding commonalities across them developed her three-stage model to describe racial and ethnic identity development: Unexamined Ethnic Identity, Ethnic Identity Search (Moratorium), and Achieved Ethnic Identity. In the unexamined stage, teens of color do not associate much personal meaning with their race or ethnicity and prefer the majority culture. The ethnic identity search stage is often triggered by an experience of discrimination forcing the adolescent to become aware of their racial and ethnic identity. This negative experience leads them to immerse themselves in their own culture more intensely–reading, attending cultural events, grouping with others of their background, and sometimes rejecting the dominant group. In the final stage, achieving identity, adolescents can internalize their race and ethnicity, seeing the differences between their group and the majority group and the racism and prejudice that exist for their group. These stages aren’t necessarily linear and can occur at any time during early or late adolescence.

As seen across these models, the role of racism and discrimination complicates the exploration of a teen’s racial and ethnic identity development. Adolescent youth of color are forced to see themselves in terms of race because society does. Racism and bigotry create a significant amount of stress for youth of color and can hinder their ability to develop a positive self-identity and harm their mental health. It is important for adolescents of color to have access to culturally positive information, positive family interactions, and supportive relationships that convey positive regard for their racial and ethnic identity creating resilience. These experiences must be considered as they undertake the normative tasks of adolescent development. Not all teens of color will go through the process of exploring their racial and ethnic identity, but it must be considered to ensure that they have a healthy transition to adulthood.

STAGES OF ADOLESCENCE

Early Adolescence

Early adolescence, from 12 to 14 years of age, is the period in which the most striking changes are noticed—physically, attitudinally, and behaviorally; encompassing the biologic changes of puberty as well as psychological and sexual awakenings. Growth spurts often begin in these years for boys, whereas girls may have had rapid growth for 1 to 2 years prior. At this stage, many adolescents express a sense of identification with a particular peer group, often a same-sex group of peers join together, and may identify belonging to a peer group with markers including clothing, hair styles, or affiliation with certain activities. Both boys and girls begin to identify themselves as distinct members of their families, often critique usual family habits, insist on spending time with peers with less supervision, and may question previously accepted family values.

In the early stage of adolescence, there may be some degree of withdrawal from family activities, increased interest in spending time with peers, greater awareness of style and appearance, a stated goal of becoming more independent, and claims of needing less parental involvement. A new awareness of sexuality may be displayed by increased modesty and embarrassment with their physical development or may exhibit itself in an increased interest in the opposite sex.

Early adolescents engage in subtle or overt displays of their growing desire for autonomy, sometimes with challenging behaviors toward authority figures, including teachers and school administrators, and exhibit disdain for rules. At this age, some adolescents begin to experiment with cigarettes, alcohol, and marijuana. The Youth Risk Behavior Surveillance System (YRBSS) surveyed all public and private schools from grades 9 to 12 throughout the United States (more information is available at https://www.cdc.gov/healthyyouth/data/yrbs/index.htm). The 2019 report indicated that 18.1% of ninth graders had tried smoking cigarettes, 8.8% tried their first cigarette before age 13, and 0.8% of male ninth graders and 0.3% of female ninth graders had smoked at least one cigarette every day consecutively for 1 month. The report also showed that 40.2% of ninth graders have used electronic vapor products (e.g. vapes, vape pens, e-cigarettes, hookah pens, e-cigars); 40.7% of ninth-grade females, and 39.8% of ninth-grade males had vaped at least 1 day during the month prior to the survey; 3.1% of ninth-grade females, and 4.0% of ninth-grade males vaped daily during the month prior to the survey. The YRBSS reported that 21.9% of ninth-grade girls and 16.3% of ninth-grade boys consumed at least one alcoholic drink in the month preceding the survey, 18.7% had their first drink of alcohol before age 13, and 7.3% of ninth-graders were binge drinking–8.2% of females and 6.4% of males. The 2019 National Survey of Drug Use and Health (NSDUH) showed that 1.3% of adolescents males and 2.1% of adolescent females age 12 to 17 had an alcohol use disorder. Findings also showed a decline in underage drinking from 2002 to 2019–prevalence in the past 30 days of alcohol use decreased 41.1% for 16- to 17-year-olds, 54.7% for 14- to 15-year-olds, and 61.9% for 12- to 13-year-olds. The YRBSS reported that 22.6% of ninth-grade girls and 26.0% of ninth-grade boys have used marijuana at least once, and current use of marijuana was present among 14.6% of ninth graders. Overall, the prevalence of one or more time use of marijuana and daily current use has decreased over the last 8 years among 9th to 12th graders. With regard to other illicit drug use (cocaine, heroin, hallucinogens, ecstasy, and amphetamines) among 9th to 12th grades, the prevalence of all has decreased between 2011 and 2019. Of note, use of electronic vaping devices among adolescents has increased substantially between 2015 and 2019. According to the National Youth Tobacco Survey (NYTS), current electronic vapor product use among high school students increased 78% (11.7% to 20.8%) from 2017 to 2018.

During early adolescence, there is significant variation in the timing of acquiring new behaviors. For example, some 12-year-olds identify with popular music or wearing “adolescent” clothing, whereas other adolescents are seeking to reinvent themselves with a more mature appearing image. Epidemiologic surveys have not provided evidence for this notion that adolescence is characterized by overwhelming alienation, profound angst, and significant disruption of family relationships. In this age group, most adolescents spend at least twice as much time with their families as with their peers.

Suzanne, a 14-year-old freshman in high school concerned about being compared to her older sister, who was known as a troublemaker. Suzanne felt that her teachers didn’t value her for being her “own” person and automatically assumed that she was a poor student, always looking for a way to avoid schoolwork. Suzanne secretly smoked cigarettes, occasionally drank alcohol, and had tried smoking marijuana, however, she found it extremely annoying that her parents were always checking up on her. She rationalized that her parents’ intrusion was just because her older sister had been suspended several times for going to school intoxicated. Suzanne spent most of her time out of school in her room with her door closed, doing homework while watching TV, instant messaging her friends on her computer, and text messaging, snap chatting, and using Instagram with her friends. Although Suzanne had enough friends, and enjoyed singing in the school choir, she found herself feeling anxious much of the time and fearing that she had forgotten to finish all of her homework or studied enough for a test. She found going to class more and more unpleasant, and decided that classes were a waste of time. Suzanne began to find the school rules increasingly irritating, especially being “locked” out of the classroom if she was a few minutes late. Suzanne felt that the teachers were too strict, and never gave her the benefit of the doubt. To avoid getting into trouble for being late, Suzanne began to cut classes then she began to cut the whole day of school if she thought she was late in the morning. Before long, she had fallen behind in her work but was too embarrassed to tell her parents. She told her friends, and they helped her get some of the schoolwork, but she had missed so much work that it was difficult to catch up. Since it was getting close to her first progress report and Suzanne could not talk to her parents, she finally went to her school counselor to say that she was not going to many classes because it was too stressful to get there on time and that she was in danger of failing. The school counselor convinced Suzanne to tell her parents about her academic problems, and, begrudgingly, she agreed to a meeting with the counselor and her parents. Suzanne's parents were shocked to find out that she was so stressed, and near failing, since she consistently said that everything was fine as long as everyone left her alone. After the meeting, Suzanne realized that instead of trying to be so independent she was actually being self-destructive. Suzanne agreed to accept her parents' help with a ride to school in the morning, and despite some initial resistance tutoring sessions were started. Suzanne worked hard with her tutor, improved her grades, and began to leave her bedroom door open more often at home. She also turned to her parents and sisters for help if she did not understand something and found that they were actually helpful. Suzanne ended up with three A's and one B for the first grading period and was proud of herself. Suzanne felt less stressed and more confident after the first grading period, she was more connected to her family, and looked forward to spending more time with friends for the remainder of her freshman year.

Middle Adolescence

During the middle phase of adolescence (roughly between the ages of 14 and 16 years), peer groups begin to shift to coed, and become more predominant in their importance. Concrete signs of greater autonomy are beginning to emerge, such as learning to drive, and in some states, obtaining a driver's license by the age of 16 years. Along with a greater lifestyle of autonomy adolescents are also presented with increased responsibilities, especially in the area of social judgment. Conflict may arise when empathy and social awareness are paired with poor impulse control and decision making, potentially leading to health risks, car accidents, and alcohol and drug use. Sexual behavior intensifies in this phase of development, initiating romantic relationships is more complicated, and self-esteem becomes a pivotal influence on positive and negative risk-taking behaviors.

The 2019 YRBSS reported more frequent use of electronic vapor products among white students (14%) and Hispanic/Latinx students

(7.5%) than among Black and Asian students (5.0% and 5.1%, respectively). Overall, frequent use of cigarettes among all races of students decreased substantially since 2011 while electronic vapor product use has steadily been on the rise. Use of alcohol, that is, report of at least one alcoholic drink within 1 month of the survey time, was 26.7% for 10th graders and 32.3% for 11th graders. The prevalence of heavy drinking—five drinks or more in a row—increased with each increasing grade from 9th to 12th grade. Tenth graders reported rates of heavy drinking, 10.6%, which increased in 11th graders, 15.3%; significant decrease seen in heavy drinking across all grades since 2011. Marijuana use was found to be somewhat higher among boys than girls across all high school grades, except 10th graders, and the overall prevalence of current marijuana abuse was 28.3% among 12th graders. The prevalence of hallucinogenic drug use (LSD, mushrooms, acid) was reported to be higher in boys than girls, and lifetime hallucinogenic drug use was 8.2% among 11th graders and 9.9% among 11th-grade boys.

In this phase of development, adolescents tend to identify with a group of peers who become highly influential in their choices of activities, styles, music, idols, and role models. Adolescents' underestimation of the risks associated with a variety of recreational behaviors and their sense of "omnipotence," mixed with their drive to be autonomous, frequently cause some conflict with parental requests and expectations. These conflicts, however, for the most part, do not result in the destruction of meaningful relationships with parents or siblings. For most teens, the process of defining themselves as unique and different from their families can be achieved while they still maintain alliances with family members.

Olivia was a 16-year-old starting her junior of high school, who had just gotten her driver's license. She realized how lucky she was to have been given her older sister's old car as soon as she had turned age 16 years, but she was overwhelmed when her friends called and constantly wanted rides. Olivia drove her younger sister to school each morning, and this created tremendous friction between them since Olivia's sister always threatened to tell their parents that Olivia was an unsafe driver because she was always late and would text her friends while in the car. Olivia was a confident driver but she would become distracted when her friends were in the car talking and laughing while she was trying to concentrate on the road. Olivia felt unable to say no to giving her friends rides, even when they smoked marijuana in her car. Olivia would half-heartedly object, however, she had never been popular, and believed that she would lose the few friends she had, if she did not take risks. Olivia decided not to tell her parents about this, and so she continued to allow her friends to smoke marijuana in her car until she had a "close call" in which a police car was tailing her but did not pull him over. Olivia began to panic and had to stop his car on the side of the road. Olivia's friends congratulated her for not getting caught, however, they didn't understand how much stress this caused her. After this, Olivia told her parents that she no longer wanted to drive on the weekends, and she told her friends that her parents had taken away the car. Although Olivia was very hurt because some of her friends dropped her and began to take rides from other friends without her, she felt a great sense of relief. Olivia began to spend more time at home drawing, painting, and creating a portfolio. Olivia went through a period of social isolation and sadness, but over time, she developed new friendships with peers who took art lessons and enjoyed drawing and painting. Olivia looked forward to applying to college and was glad to be leaving the social situation in his high school. Olivia put new energy into studying and taking challenging courses in the second semester of junior year including several advanced placement classes, and found that she had more in common with classmates in these courses than her previous group of friends. Olivia felt excited and enthusiastic about entering an art program at a competitive and popular college.

Late Adolescence

Late adolescence (between the ages of 17 and 19 years) is a time when the adolescent cognitive brain processes, impulse control, and decision making begin to catch up with the social brain and behavior. During this phase, exploration of academic pursuits, musical and artistic tastes, athletic participation, and social bonds leads an older teen toward greater definition of self and a sense of belonging to certain groups or subcultures within mainstream society. As Erickson aptly described, through this exploration, an adolescent gradually moves toward a solid “identity formation.” Well-adjusted adolescents can be comfortable with current choices of activities, tastes, hobbies, and friendships yet remain open to the idea that their “identities” will continue to be refined during young adulthood. Peer interactions are primary, and some bridges may be formed with family members to increase communication and social repartee. This transition spans the completion of high school and at least the first year of college, where exploration of new relationships and interests, often in a new setting away from home, is initiated. Prior to finishing high school, however, the use of cigarettes, alcohol, and marijuana is higher than in the lower high school grades. The 2019 YRBSS reported that 7.8% of 12th graders had had the occasion to drive while they had been drinking alcohol, and 17.0% of 12th graders had been passengers with drivers who had been drinking alcohol. The prevalence of current frequent electronic vapor use was reported to be 40.4% among 12th graders. Alcohol use within 1 month of the survey was reported by 39.9% of 12th graders, 37.5% of 12th-grade boys and 42.0% of 12-grade girls, with episodic heavy drinking of more than four consecutive drinks in a row occurring in 22.4% of 12th graders. Nationwide, current marijuana use was reported by 31.1% of 12th-grade boys and 25.1% of 12th-grade girls. The prevalence of having ever used ecstasy was reported by 6.5% of 12th-grade boys and 3.4% of 12th-grade girls.

Neuroimaging studies provide evidence suggesting that during late adolescence and even into adulthood, the PFF neural networks continue to facilitate working memory, prospective memory (the ability to keep in mind an intention to do something in the near future while multitasking), perspective taking (understanding another person's point of view), and responding to others while keeping one's feelings in mind. This phase of adolescence is an exciting time for the development of finely tuned cognitive abilities, social skills, and enduring relationships.

Damien, 18 years-old, was in his second semester of college. He loved being on his own away from home and was gratified that his new college relationships seemed more real and important to him than his high school peers. Damien had no idea about what he might want to choose as a future career, and during his first semester, he had opted for random classes in many different subjects. In high school, Damien was naturally good in languages, and math without studying much. He enjoyed travel and never gave much thought to what it might be like to be settled down in a career. Damien's parents had encouraged him to explore interests during high school, however, they made it clear that since they were paying a lot of money, once he was in college, he was expected to pick a career goal as soon as possible, so that his college years wouldn't be wasted. So far, Damien had found that manty of his classmates came to college with well-defined ideas about future careers, and he began to feel pressure. Damien agonized about the infinite possibilities and wondered why many peers already seemed sure of their future goals and he felt so confused. Damien began to feel alienated by his lack of direction and developed insomnia. Damien mentioned his insomnia to his friend Emma, and she was surprised by his level of stress, and told him she viewed him as an open and social person who made friends easily and got along with everyone. Damien felt that his open and nonjudgmental style was now getting in the way of making a decision about a major where he fits best. Damien confided this to Emma, and she then challenged the rigid expectations and timeline that Damien's parents had imposed on him. Emma suggested that he take at least two more quarters of classes in any subjects that seemed interesting before feeling compelled to determine the best match for him. Damien realized that Emma's approach was more reasonable than his parents' timeline and he felt the burden of those pressures diminish. Damien had been influenced to believe that quitting or changing can be a sign of weakness. Emma's approach supported the idea that life decisions should be based on positive experiences rather than on a timeline that didn't seem to fit. After much internal deliberation, Damien embraced Emma's suggestion and decided to allow himself at least another year to decide on a major and a potential career path.

COMPONENTS OF ADOLESCENCE

Physical Development

Physical growth, along with the onset of sexual development, is identified as the "hallmark" of adolescence. Puberty is the process by which adolescents develop physical and sexual maturity, along with reproductive ability. Pubertal development begins in response to a surge of gonadotropin-releasing hormone from the hypothalamus. The hypothalamus is influenced by a multitude of physiologic and psychological factors, and the onset of puberty may be delayed or altered in situations of severe nutritional deprivation. Gonadotropin-releasing hormone acts on the pituitary gland to release the gonadotropin-luteinizing hormone and follicle-stimulating hormone. These two hormones, in turn, act to stimulate the production of a variety of androgens, testosterone, estradiol, and other peptides. Physical development during puberty is also regulated by two additional hypothalamic peptides, growth hormone–releasing hormone and somatostatin, both of which promote the release of growth hormone from the pituitary gland. Puberty unfolds over time in conjunction with the pulses and feedback loops of these gonadotropins. The first signs of the pubertal process are an increased rate of growth in both height and weight. This process begins in girls by approximately 10 years of age. By the age of 11 or 12 years, many girls noticeably tower over their male classmates, who do not experience a growth spurt, on average, until their 13th year. Even more striking are the differences in the development of secondary sexual characteristics in girls versus boys approximately 13 years old. By the age of 13 years, many girls have experienced menarche, and the majority have developed breasts and pubic hair. The initial appearance of pubic hair in both genders may occur before puberty, during adrenarche, which is dependent on hormones secreted by the adrenal glands, rather than by gonadal development. Some components of pubertal development are identical in boys and girls and include an increase in height, weight, muscle, and fat, whereas others, such as breast development or growth of facial hair, are specific to only one gender. There is wide variation in the normal range of onset and timing of pubertal development and its components. There is a set sequence, however, in the order in which pubertal development proceeds. Thus, secondary sexual characteristics in boys, such as increased length and width of the penis, for example, will occur after the release of androgens from developed enlarged testes.

Sexual maturity rating (SMR), also referred to as Tanner stages, ranges from SMR 1 (prepuberty) to SMR 5 (adult). The SMR ratings include stages of genital maturity in boys and breast development in girls, as well as pubic hair development. Pubic hair development is a less reliable measure of sexual maturity, however, because it may emerge in response to androgens from the adrenal glands before the pubertal process has actually begun. Table 34.3–1 outlines SMRs for boys and girls.

Table 34.3–1.
Sexual Maturity Rating (SMR) of Girls and Boys

SMR	Girls	Boys
Stage 1	Preadolescent, papilla elevated	Penis, testes, scrotum preadolescent
	No pubic hair	No pubic hair
Stage 2	Breast bud, small mound; areola diameter increased	Penis size same, testes, and scrotum enlarged, with scrotal skin reddened
	Sparse, long pubic hair, mainly along labia	Sparse, long pubic hair, mainly at the base of the penis
Stage 3	Breast and areola larger; no separation of contours	Penis elongated, with increased size of testes and scrotum
	Pubic hair darker and coarser, spread over the pubic area	Pubic hair darker and coarser, spread over the pubic area
Stage 4	Breast size increased	Penis increased in length and width
	Areola and papilla raised	Testes and scrotum larger
	Pubic hair coarse and thickened; covers less area than in adults, does not extend to thighs	Pubic hair coarse and thickened; covers less area than in adults, does not extend to thighs
Stage 5	Breasts resemble adult female breast; areola has	Penis, testes, scrotum appear mature Pubic hair increased in density; area

	recessed to breast contour Pubic hair increased in density, area; extends to thighs	extends to thighs

The primary female sex characteristic is ovulation, or the release of eggs from ovarian follicles approximately once every 28 days. When adolescent girls reach SMR 3 to 4, ovarian follicles are producing enough estrogen to result in menarche, the onset of menstruation. When adolescent girls reach SMR 4 to 5, an ovarian follicle matures on a monthly basis and ovulation occurs. Estrogen and progesterone promote sexual maturation, including further development of fallopian tubes and breast development.

For adolescent boys, the primary sex characteristic is the development of sperm by the testes. In boys, sperm development occurs in response to follicle-stimulating hormone acting on the seminiferous tubules within the testes. The pubertal process in boys is marked by the growth of the testes stimulated by luteinizing hormone. There is no discrete sign of the formation of viable sperm, but it must be assumed that at the time that ejaculatory capacity occurs, viable sperm are likely to be present. An adolescent boy's ability to ejaculate generally emerges within 1 year of reaching SMR 2. Secondary sexual characteristics in boys include thickening of the skin, broadening of the shoulders, and the development of facial hair.

In general, the majority of adolescents adjust well to the changes in their physical growth and maturation. Inasmuch as overall identity formation is a gradual process, adolescents incorporate their developing physical identities into their evolving adult selves.

Cognitive Maturation

Cognitive maturation in adolescence encompasses a wide range of expanded abilities that fall within the global category of "executive functions" of the brain. These include the transition from concrete thinking to more abstract thinking; an increased ability to draw logical conclusions in scientific pursuits, with peer interactions, and in social situations; and new abilities for self-observation and self-regulation. These result in an adolescent’s capacity to develop hypotheses, express altruism, create idealistic scenarios, and develop hopes and dreams for the future. Adolescents acquire increased awareness of their intellectual, artistic, and athletic gifts and talents, yet it often takes many more years into young adulthood for them to establish a practical application for these abilities.

The cognitive-developmental theorist Jean Piaget has been a central influence on the way in which clinicians and researchers understand the cognitive strategies of children throughout development and the changes that occur during adolescence. The central cognitive change that occurs gradually during adolescence is the shift from concrete thinking (concrete operational thinking, according to Piaget) to the ability to think abstractly (formal operational thinking, in Piaget’s terminology). This evolution occurs, according to Piaget, as an adaptation to stimuli that demand an adolescent to produce hypothetical responses, as well as in response to the adolescent’s expanded abilities to provide generalizations from specific situations. Recent magnetic resonance imaging (MRI) studies showing evidence of a preadolescent increase in cortical gray matter in the brain, possibly due to a wave of increased production of synapses, may be a physical correlate of the adolescent’s ability for expanded abstract reasoning. The development of abstract thinking is not a sudden epiphany but, rather, a gradual process of expanding logical deductions beyond concrete experiences and achieving the capacity for idealistic and hypothetical thinking based on everyday life. Younger adolescents may be able to verbalize an understanding of complex abstract constructs and still react to situations in rigid and concrete ways. For example, a young adolescent may be able to verbalize an understanding of the link between drug and alcohol use and health risks but cannot imagine it happening to themselves. Similarly, adolescents often believe that they can “beat the odds,” even when they understand the mathematical meaning of the probability of an event occurring. Adolescents often use an omnipotent belief system that reinforces their sense of immunity from danger, a false sense of immortality, even when confronted with logical risks. Some degree of child-like magical thinking continues to coexist with more mature abstract thinking in many adolescents. Despite the persistence of magical thinking into adolescence, adolescent cognition departs from that of younger children with regard to an increased ability for self-observation and the development of strategies to promote strengths and compensate for weaknesses. For example, an adolescent who becomes aware of a weakness in completing academic work or tests during the time allotted in school may realize that they know the material but cannot perform well in timed examinations and will make a request to the teacher to receive additional time on examinations. The ability to develop strategies for survival and use proactive methods of making them happen increases during the period of adolescence. One of the important cognitive tasks in adolescence is to identify and gravitate toward those pursuits that seem to match the adolescent's cognitive strengths in academic courses and in thinking about future aspirations. Piaget believed that cognitive adaptation in adolescence was profoundly influenced by social relationships and the dialogue between adolescents and peers, making social cognition an integral part of cognitive development in adolescence.

Socialization

Socialization in adolescence encompasses the ability to find acceptance in peer relationships and in turn the development of more mature social cognitions enabling the adolescent to accept others. It is common knowledge that, for early adolescents, the skills for developing a sense of belonging to a peer group, are of central importance to a sense of well-being. Being viewed as socially competent by peers is a critical component in building good self-esteem for most early adolescents. Peer influences are powerful and can foster positive social interactions, as well as apply a pressure in less socially accepted behaviors or even high-risk behavior. Belonging to a peer group is, in general, a sign of adaptation and a developmentally appropriate step in separating from parents and turning the focus of loyalty toward friends. Social cognition—the way in which adolescents use logic and reason in their assumptions about peers and in their peer interactions—has been explored by Robert Selman (see Asher and Gottman, 1981). After interviewing children and adolescents with respect to their interpretation of various social interactions, he identified stages of social cognition. In school-aged children and preadolescents, Selman described social cognitions as "fair-weather cooperation." That is, the friendships are characteristically on-again/off-again, depending on the level of affiliation or disagreement in the last social interaction. Thus, children between the ages of 6 and 12 years are able to engage in exchanges of ideas and opinions and acknowledge feelings of peers, but the relationships often wax and wane in a discontinuous way on the basis of altercations and good times. Friendships deepen with repeated good times, but, for some school-aged children, a variety of peers are often interchangeable—that is, a companion is sought when a given child has free time, rather than out of a desire to spend time with a specific friend. As adolescence ensues, friendships become more individualized, and personal secrets are likely shared with a friend rather than a family member. A comfort level is achieved with one or several early adolescent peers, and the group may "stick together," spending most free time together. In early adolescence, a blend of these two social modes may emerge, small "cliques" may arise, and, even within the cliques, competition, and jealousies regarding which dyads are "preferred" or higher ranked within the clique may result in some discontinuities in the relationships. In later adolescence, the peer group solidifies, leading to increased stability in the friendships and a greater mutuality in the quality of the interactions.

Moral Development

Moral reasoning among adolescents depends on the maturation of cognitive strategies for problem solving, societal conventions, exposure to familial attitudes, and personal values adopted. Morality is a set of values and beliefs about codes of behavior that conform to those shared by others in society. Adolescents, like younger children, tend to develop patterns of behaviors characteristic of their family and educational environments and by the imitation of specific peers and adults whom they admire. Moral development is not strictly tied to chronologic age but, rather, is an outgrowth from cognitive development. Piaget described moral development as a gradual process parallel to cognitive development, with expanded abilities in differentiating the best interests of society from those of individuals occurring during late adolescence. The development of trends in moral judgment has been divided into levels by Lawrence Kohlberg, who posed “moral dilemmas” to children and adolescents and subsequently characterized their responses. According to Kohlberg, morality can be divided into preconventional, conventional, and postconventional. Before adolescence, decisions regarding actions that are loosely defined as “right” and “wrong” are dependent on the anticipated consequence of those actions. School-aged children, for example, denoted as using level 1 of preconventional morality, may define actions that result in a negative consequence or punishment as “wrong” simply due to the consequence that may ensue. The most important defining feature of “right” then becomes being able to “get away with it,” and it only becomes “wrong” if there is punishment associated with it. Some young adolescents, still functioning within the category of preconventional morality but one level higher (level 2), may define “right” and “wrong” as synonymous with what is perceived as fair and reciprocal. Thus, the philosophy of “an eye for an eye, and a tooth for a tooth” may prevail in this stage of morality. During adolescence, conventional morality, level 3, becomes prevalent, with the view that there is a “right” and a “wrong” determined by society, often without the possibility of mitigating or extenuating circumstances. Adolescents espousing this level of morality are often rigid in their concept of “good” and “bad” and unable to find a role for motivation in their judgment of behaviors. Thus, adolescents may often take the viewpoint of applying the harshest punishments that society can offer when asked to judge complex moral situations—for example, a poor man robbing a pharmacy to obtain a life-saving medication for his child. Later in adolescence, at level 4, a greater understanding of societal rules may be espoused within the context of conventional morality. It is generally not until late adolescence or early adulthood that the complexities of moral decision making are recognized and incorporated into their thinking. Postconventional morality (level 5), according to Kohlberg, implies the ability to interpret complex moral dilemmas in which the best moral decisions of an individual may come into conflict with the moral standards of society. At this level of thinking, adolescents are more flexible in their understanding of “right” and “wrong.” The concept of extenuating circumstances is understood at this level of moral thinking, allowing exceptions to rules that are ordinarily followed. Kohlberg postulated a final, higher level of postconventional moral development that he labeled level 6, in which an individual is able to recognize a universal moral imperative that might take priority over standard societal moral values.

Although Kohlberg’s and Piaget’s notions of moral development focus on a unified theory of cognitive maturation for both sexes, Carol Gilligan emphasizes the social context of moral development leading to divergent patterns in moral development. Gilligan points out that, in women, compassion and the ethics of caring are dominant features of moral decision making, whereas, for men, predominant features of moral judgments are related more to a perception of justice, rationality, and a sense of fairness.

Self-Esteem

The proliferation of social media since the 1990s has provided opportunities for adolescents to engage in social interactions with peers across the globe. These interactions have enabled adolescents to broaden their social connections when they feel isolated, to seek support when they are experiencing distress, and to learn about other adolescents and their interests. This has been a positive experience for most adolescents as they engage in social and interpersonal challenges. In some cases, however, particularly for adolescents who are vulnerable, at-risk for emotional problems, or have pre-existing mental health conditions, their self-esteem may be compromised. Epidemiologic surveys suggest that the majority of adolescents meet the challenges of their physical, social, and emotional development with optimism and solid self-esteem. Self-esteem is a conglomerate measure of one’s sense of self-worth based on perceived success and achievements, as well as a perception of how much one is valued by peers, family members, teachers, and society in general. The most important correlates of good self-esteem are one's perception of positive physical appearance and high-value to peers and family. Secondary features of self-esteem relate to academic achievement, athletic abilities, and special talents. Adolescent self-esteem is mediated, to a significant degree, by the positive feedback received from a peer group and family members, and adolescents often seek out a peer group that offers acceptance, regardless of negative behaviors associated with that group. There are also developmental factors that influence adolescents' cognitions about themselves that affect global ratings of self-esteem in this age group. Studies comparing reported self-esteem across the lifespan indicate that there are trends in self-esteem ratings that are age-related. Starting in adolescence, there are also gender differences in reported self-esteem. A study investigating global measures of self-esteem in school-aged children—adolescents, 13 to 17 years old, and older teens to young adults, 18 to 22 years old—found group differences depending on age. Overall, reported self-esteem was found to be highest in school-aged children as compared to adolescents and young adults, and boys and girls were identical in their self-reports. During adolescence, global reports of self-esteem decreased throughout, with adolescent girls experiencing a drop twice that of adolescent boys. Self-esteem, at a group level, continued to drop during the high school years, rose in young adulthood during the college years, and continued to rise in early to midadulthood. Girls continued to rate themselves with generally lower self-esteem into adulthood. The question arises as to how most adolescents surveyed report relatively good adjustment and yet, at the same time, experienced a drop in self-esteem. These data appear to reflect the "inflated" self-esteem of school-aged children, who may be seeing themselves as better, smarter, and more capable than they really are. During adolescence, it is quite possible that even well-adjusted individuals realize that they are not as exceptional as they used to believe they were. Adolescents seem to be more aware of their shortcomings and, therefore, had lower self-reported scores on self-esteem measures. In young adulthood, however, and into midadulthood, along with expanded cognitive abilities for self-regulation, reported self-esteem increases, and continues to increase well into midadult life.

ADOLESCENT BRAIN DEVELOPMENT: CORRELATIONS WITH EMOTIONAL, SOCIAL, AND COGNITIVE CHANGES

It is now widely accepted that emotional, social, and cognitive brain processes continue to develop throughout childhood and early and late adolescence. The structural and functional development of various regions of the adolescent brain interacts with the social environment to influence social sensitivity, cognitive decisions, and risk-taking behaviors.

The work of Blakemore and Mills and other researchers have consistently suggested that a network of specific brain regions is involved in social cognitive processes. For example, an adolescent's increasing ability to understand the mental states and, thereby, the behaviors of others, "mentalizing," also called "theory of mind," has been associated with a network of brain regions, including the dorsal medial prefrontal cortex (dmPFC), the TPJ, pSTS, and the anterior temporal cortex (ATC). These brain gray matter regions have been shown to continue to develop in volume, cortical thickness, and surface area throughout adolescence before stabilizing in young adulthood. During adolescence, the overall gray matter does not mature by growing larger. Instead, it is pruned to be more specialized. However, the PFF, parts of which are involved in mentalizing and other portions with judgment, impulse control, and long-range planning, continue to undergo dynamic changes into the early 20s.

A variety of tasks-related to mentalizing, such as recognizing the meaning of facial expressions, have been shown to activate the mPFC, and in fact, this brain region has also been shown to become diminished in size and more specialized between late adolescence and adulthood. Social emotions such as embarrassment, shame, guilt, and pride require an understanding of what others may be thinking or feeling (mentalizing) about someone, whereas fear does not. In adolescence, increased sensitivity to peer opinions is a hallmark trait, and the corresponding brain regions in adolescents may be activated. Functional MRI (fMRI) study investigating neural activation during tasks related to viewing or reading social-emotional scenarios found that adolescents recruited the dmPFC even more intensely than adults given the same tasks.

In a recent study of an online communication game requiring the subject to understand the mental state of the online player in order to make correct decisions, a midadolescent group made less mistakes than a midchildhood group, but more mistakes than a young adult group. This result supports the notion that social development is still in play and not maximized even in late adolescence and that social sensitivity and social skills increase into young adulthood.

On the other hand, brain regions associated with risk-taking and sensation-seeking include the amygdala, nucleus accumbens (NAcc), and PFF. A study investigating the time course of maturation of the brain regions associated with risk-taking behaviors found that relatively earlier maturation of the amygdala and/or the NAcc compared to the PFF suggested a mismatch between these structures. The amygdala has been found to increase in volume from late childhood until mid to late adolescence with less growth after the age of 16 years. On the other hand, the PFF showed little change in gray matter volume in childhood and continues to decline into young adulthood. More research is needed in order to draw conclusions about adolescent risk-taking behaviors and sensation-seeking in view of these differentially developing brain regions.

In contrast to gray matter, white matter volumes increase throughout childhood and adolescence and reflect a developing integration or “connectivity” of brain function regions that are myelinated.

During the course of adolescence significant qualitative changes in thinking are demonstrated in greater ability to discriminate emotional cues, increased self-awareness, selective attention to information, and increased response inhibition. Perhaps the most dramatic changes in cognitive abilities are an increased capacity for abstract reasoning, planning, and affect modulation. Over the past two decades, MRI and fMRI have become important modalities for studying both the structure and the function of adolescent brains. Two areas of the brain consistently shown by MRI to evidence structural changes during adolescence are the PFF and the parietal cortex. Neuroimaging studies indicate that the PFF matures more slowly than other areas of the brain and that its development is correlated with increased capacity in adolescents to inhibit behavioral responses. The PFF is known to mediate “executive function,” that is, skills involving working memory, selective attention, and voluntary response inhibition. In addition to executive function, the PFF is believed to be involved in social and cognitive abilities such as self-awareness and the development of an understanding of another person’s perspective. Recent studies have provided evidence of a linear improvement in performance by age on certain tasks within executive function, such as those involving selective attention, working memory, and problem solving. This lends support for an increase in performance in these tasks as a function of the pruning and synaptic reorganization of the PFF.

Social cognition—in particular, recognition of facial expressions of emotion—has been studied in healthy adolescents. The amygdala, a region of the brain known to be associated with emotion processing, was found by fMRI to be activated in adolescents shown fearful facial expressions and in response to happy faces compared to neutral faces. Among adolescents, left amygdala activity when presented with fearful faces decreased in girls but not in boys. Female adolescents also demonstrated greater activation of the dorsolateral PFF, whereas male adolescents did not show this pattern. It was hypothesized that the pattern of diminished amygdala and increased prefrontal activation in female adolescents reflects increased maturity in the regulation of emotional experiences in them compared to male adolescents. Structural and fMRI studies implicate puberty as a period of synaptic reorganization in the prefrontal and parietal cortex, which may place the adolescent brain in a sensitized state with respect to the development of executive functions and social cognitive skills. Clearly, the adolescent brain continues to develop structurally, and the relationships of these structural changes to behavioral and emotional responses warrant further investigation.

CURRENT ENVIRONMENTAL INFLUENCES AND ADOLESCENCE

Adolescent Sexual Behavior

Sexual interest and experimentation are critical components of adolescent development for boys and girls. Sexual experimentation in adolescents often begins with fantasy and masturbation in early adolescence, followed by genital touching with the opposite sex or, in some cases, same-sex partners, oral sex with partners, and initiation of sexual intercourse at a later point in development. In one study of urban adolescents, at least one-third reported some form of genital sexual activity in the last year with a partner, and approximately 10% reported oral sexual behavior with a partner. The balance between healthy adolescent sexual experimentation and emotionally and physically safe sexual practices has been one of the major challenges for society in the new millennium.

Trends in adolescent sexual activity are monitored by the YRBSS, which measures the self-reported prevalence of health-risk behaviors, including sexual activity. Surveys obtained for the years 2018 to 2019 provide detailed information regarding sexual activity in adolescents and the risks associated with these behaviors. Among high school students across the nation, 38.4% reported that they had had sexual intercourse during their life, and 27.4% of high school students reported that they were currently sexually active. Among those who were currently sexually active, 45.7% reported that they or their partner hadn’t used a condom during their last sexual intercourse. The prevalence of having had sexual intercourse at least once was higher among black males (50.1%) than Hispanic males (43.6%), followed by Hispanic females (40.3%), and black females (34%). The overall rate of having had sexual intercourse in high school among white students was 38.2%, with white females (39.2%) reporting it more frequently than white males (36.7%). A total of 8.6% of students report that they had had sexual intercourse with at least four persons during their lifetime. Among the 27.4% of currently sexually active students nationwide, 45.7% reported that either they or their partner hadn't used birth control pills as a contraceptive before their last sexual intercourse. Prevalence of condom use at last sexual intercourse was 54.3%, and condoms were the most prevalent primary contraceptive method, a decline from the recent high of 63% in 2003. According to the CDC, the percentage of schools where students are required to receive instruction in HIV prevention decreased from 64% in 2000 to 41% in 2014. Nationwide, 9.4% of students had been tested for HIV. In 2018, 21% of new HIV cases were among youth. Statistics show that by the end of 2016, 51% of youth living with HIV were unaware that they had HIV.

Factors That Influence Adolescent Sexual Behavior. Factors impacting sexual behavior in adolescents include peer groups, psychosocial factors, temperamental and personality traits, externalizing and internalizing psychiatric symptoms, gender, cultural and religious background, racial factors, family attitudes, and sexual education and prevention programs.

A recent study has associated the role of low self-control, specifically impulsivity, with early sexual initiation and risky sexual behaviors, including sex while intoxicated, infrequent use of condoms, engaging in sex with strangers, a history of sexually transmitted diseases, and having multiple partners.

Longitudinal studies have also shown that a childhood history of aggressive-disruptive behaviors is associated with early sexual initiation and increased rates of adolescent sexual activity among boys and early initiation, increased teen pregnancies, and multiple sexual partners among girls.

Historically, male adolescents have initiated sexual intercourse at a younger age than female adolescents, and the 2019 YRBSS report revealed that 3.9% of boys and 2.1% of girls had sexual intercourse for the first time by age 13; the median age of sexual initiation was found to be 16 years of age for both males and females. Child Trends showed that the age at which sexual activity occurs among teens increases with grade level; 16% of ninth graders reported they were sexually active compared to 46% of 12th graders. In general, adolescents who initiate sexual intercourse at younger ages are more likely to have a greater number of sexual partners.

There is a suggestion from surveys of adolescents that early initiation of sexual intercourse before the age of 15 years is more frequent among African-American boys than in Hispanic boys and more frequent in Hispanic boys than in white boys.

Communities with high levels of trust and that offer youth opportunities like after-school programs, job training, and sports help young people avoid risky sexual behaviors, pregnancy, and support positive sexual health. Additional Protective factors include school and family connectedness, academic achievement, parental monitoring, a supportive and responsive parenting style, higher family income, spirituality, and religious affiliation, having plans for a positive future, and believing in the ability to control one’s life. Students who participate in well-designed sexual health education programs delay the initiation of sexual intercourse, have fewer sex partners, have fewer experiences of unprotected sex, and increase their use of protection, specifically condoms.

Risk-Taking Behavior

Risk-taking behaviors are typically the product of cumulative contributions from brain development in particular regions of the adolescent brain and social and environmental influences. These contributions include the possibility of overactive reward-related circuitry, as in the ventral striatum, as well as the influences of peer acceptance. In experimental circumstances, adolescents may rate the potential reward of a risky action as inflated. However, at times adolescents also rate the relative risks of actions as inflated.

Reasonable risk-taking is a necessary response to increasing autonomy in adolescence, leading to confidence in forming new relationships and competence in sports and in social situations. However, high-risk behaviors among this generation of adolescents are associated with serious negative consequences and can take many forms, including drug and alcohol use, unsafe sexual practices, self-injurious behaviors, and reckless driving. Alcohol is the most commonly used substance among adolescents and is associated with a multitude of coexisting high-risk behaviors, such as motor vehicle accidents resulting in injury and death, academic deterioration, aggressive behavior, and criminal behaviors. According to the National Institute on Alcohol Abuse and Alcoholism, in 2019, 4.2 million young people reported binge drinking (defined as at least five consecutive drinks at one time) at least once in the past month, and 8,25,000 reported binge drinking on 5 or more days over the past month. While the current use of alcohol has decreased over the last decade, binge drinking continues to be a problem. Historically, adolescent boys were more likely to drink alcohol and binge drink, but there has been a significant increase in both behaviors among teen girls compared to teen boys in recent surveys. Adolescents frequent and heavy use of alcohol was associated with future antisocial behavior, as well as increased levels of depression. Furthermore, even lower levels of alcohol use predicted similar problems as teens entered young adulthood. This implies the need for preventive interventions during adolescence for those teens who use alcohol, even if drinking behavior does not appear to be frequent or serious.

Illicit drugs, including marijuana, cocaine, heroin, hallucinogens, “ecstasy,” and amphetamines, also pose serious health risks among adolescents. Monitoring for the Future National Survey indicated that 8.7% of 8th graders used an illicit substance in a given month, and 18.22% and 22.22% of 10th and 12th graders, respectively, used illicit drugs in a given month. In younger adolescents, boys and girls had similar rates of drug use, whereas older adolescent boys had higher rates, especially among 11th and 12 graders. Although there are vast differences between adolescents who had experimented on rare occasions with marijuana and adolescents who engaged in daily drug use, the large numbers of adolescents engaged in these behaviors at every level is cause for public health concern. Furthermore, the use of illicit substances is also associated with youth crimes and violent behaviors. According to Child Trends Data Bank, in 2016, the rates of violent crime victimization (which include aggravated and simple assault, robbery, and rape) among adolescents aged 12 to 20 had fallen to 28 victimizations per 1,000 from 181 per 1,000 in the mid-1990s. Adolescents are most often the victim of simple assault; in 2016, 76% of adolescents aged 12 to 14 and 67% aged 15 to 17 were victims of simple assault. Male adolescents, ages 12 to 17, are more likely than female adolescents to be the victims of serious violent crimes, with the exception of sexual assault and rape. Younger male adolescents aged 12 to 14 are also more likely to be victims of violent crime than their older counterparts, 15 to 17 years old. These grave statistics indicate a need to identify risk factors and find protective factors for today's youth.

Identifying adolescents with high-risk behavior is not always feasible until the high-risk behaviors have resulted in academic or legal consequences. Once this occurs, a given adolescent can be identified as being at "high risk" for future difficulties and, optimally, should be monitored by a counselor through school, an outside therapist, and, in some cases, through the legal system. It is not always easy for an outsider, at first glance, to separate adolescents engaging in normal experimentation from those embroiled in high-risk behaviors. When there is evidence that an adolescent is engaged in frequent alcohol or drug use, reckless behavior, or criminal activities, urgent psychosocial interventions are needed. In less obvious cases, adolescents who have ongoing academic difficulties, do not participate in any athletic or other extracurricular structured activities and are socially isolated may also be in need of psychiatric evaluation.

A study in Pediatrics showed that high-risk behaviors (which include sexual intercourse, substance use, violent behaviors, and school problems) were strongly associated with the presence of tattoos and body piercings in adolescents. Both piercings and tattoos have become more mainstream among adolescents over the years. Recent studies show that approximately 10% to 23% of adolescents have tattoos, and 2% to 42% have body piercings; 27% of adolescents aged 16 to 18, and 11% aged 13 to 15 have asked their parents for a tattoo. Both tattoos and body piercings were more common in girls than in boys (16.6% vs. 8.1% for tattoos, 36.7% vs. 10.1% for body piercing). Adolescents who endorsed possession of at least one tattoo or body piercing were more likely to endorse the use of gateway drugs (cigarettes, alcohol, marijuana), as well as experience with hard drugs (cocaine, crystal methamphetamine, and “ecstasy”). Female adolescents who had tattoos or body piercings, especially younger adolescents, were more likely to endorse a history of suicidal ideation or behaviors. Male adolescents with tattoos and female adolescents with body piercings were more likely to report a history of involvement in physical fights, carrying a weapon, or bodily injury as a result of physical fighting. Adolescents with tattoos or piercings were more frequently sexually active than adolescents without them and at greater risk for unsafe sexual behavior.

Adolescents may choose clothing, hairstyles, and/or body decorations as a means of fitting in with a selected peer group or to attract attention and distinguish themselves as special or different from others. The presence of tattoos or body piercings does not necessarily indicate high-risk behaviors in adolescents. However, in certain peer groups, it is associated with risky lifestyles. Conversely, many adolescents with serious high-risk behaviors do not possess tattoos or body piercings. Given the serious consequences of alcohol and drug use and violence among adolescents, inquiry into these behaviors should be an integral part of every health care evaluation.

Adolescent Use of the Internet and Social Media

The internet and social media are ubiquitous among adolescents and serve as platforms to socialize, communicate, and share talents, as well as a myriad of personal information, affecting all aspects of life for today’s adolescents. The “digital revolution” the explosion and integration of multiple technologies into daily life) among adolescents, as discussed by Giedd (2012), has highlighted the adaptive nature of the adolescent brain in its ability to embrace and incorporate these technologies. According to Giedd, the amount of time it took for 50 million users to develop (pace of “penetration”) with respect to older technologies such as television was 13 years. However, for newer technologies, the pace of “penetration” is moving at record speed. For the internet, it was 4 years; for Facebook, it took about 3½ years; for

Twitter, about 3 years; for iPads, 2 years; and for Google+, it took only 88 days to gain 50 million users!

Recent reports indicate that about 90% of children and adolescents use the internet in myriad ways, including instant messaging, visiting social networking sites, blogging, chatting in chat rooms, posting photos, videos, and stories, and obtaining information by visiting numerous websites. Positive effects of the internet identified for adolescents include exponentially increased communication with both friends and family and increased access to academic information, and familiarity with the social and cultural habits of people around the world. Concerns raised regarding teens' access to a constant and overlapping barrage of stimuli, however, include the ubiquitous "multitasking" that seems to characterize adolescent practices in the United States. While adolescents may be under the impression that doing their homework while chatting with a friend, listening to music, instant messaging, and viewing Facebook seems like an efficient way to get more done in the same amount of time, however, the brain is actually rapidly shifting between tasks leading to interruption of attention to potentially vital tasks. For example, an MRI study of a simulated driving task with or without competing stimuli showed significant impairment in driving performance during competing conversation, in addition to a sizable reduction in activation in the parietal lobe.

Some adolescents use the internet to learn about health-related issues and for self-help. In this regard, several health-related, adolescent-friendly websites are available that provide information on a variety of psychological and medical topics. One such site is http://www.teenshealth.org, a U.S.-based website, provides information about general and healthy living information for kids, teens, and parents, including information on mental health, sexual health, causes of illness, a question and answer section, and information on different coping skills. Mind Your Mind (http://mindyourmind.ca) is a Canadian website for youth by youth, which provides resources to help teens manage their stress and mental health issues and includes personal stories, music, videos, toolkit, and a blog. Safe Place (https://www.nationalsafeplace.org/for-teens) is a national youth outreach and prevention program for young people under age 18 who need immediate help and safety. Mental Health Literacy (https://mentalhealthliteracy.org/) is another Canadian website where, friends, parents, health professionals, and educators can help support teens. There are significant risks associated with Internet use, especially for adolescents who share personal information and communicate with strangers on line. According to the National Cyber Security Alliance (NCSA), a national survey that sampled 804 online teen users between the ages of 13 and 17 years showed that American teens have had hurtful and complicated online experiences which include negative messages about things they have said or done, their appearance, sexual orientation, and gender or race or ethnicity. Thirty-nine percent of teens reported that someone had been mean or cruel to them while online or while using apps over the past year. Many teens were "very concerned" about having someone post untrue things about them online (35%), receiving unwanted communications that made them uncomfortable (32%), and being pressured to participate in harassing or bullying someone else (32%). Online teens also said that they were very concerned" about being harassed or bullied for a sustained period themselves (29%), being approached by people they don't know online (25%), and being called offensive names (25%). The 2014 Teen Internet Safety Survey revealed 54% of teens had witnessed cyberbullying online, 24% reported being bullied online, 60% who were bullied told an adult. One study showed that cyberbullying and victimization begin as early as second grade, peaking in middle school when the online exchange of sex-related content begins; and behaviors typically increase into the high school years. Internet safety programs urge youth to avoid sharing personal information and not to communicate with strangers online. It is also crucial that parents play an active role in keeping their children and adolescents safe online, emphasizing supervision.

It is clear that adolescent culture in the United States is dominated by technology in a variety of ways and is a vital part of teen social interactions. Social media technologies have enabled adolescents to connect with a vastly larger scope of other teens than they could meet in daily life. Teens are not necessarily more or less social due to social media, however, the opportunities are broader to interact with more people using more diverse modalities.

FURTHER READINGS

Ballentine KL. Understanding racial differences in diagnosing ODD versus ADHD using critical race theory. *Fam Soc*. 2019;100:282–292.

Blakemore SJ, Mills KL. Is adolescence a sensitive period for sociocultural processing? *Annu Rev Psychol*. 2014;65:187–207.

Cenat JM. How to provide anti-racist mental health care. *Lancet Psychiatry*. 2020;7:929–931.

Child Trends. Violent Crime Victimization. 2018. Accessed in 2021.

Cookingham LM, Ryan GL. The impact of social media on the sexual and social wellness of adolescents. *J Pediatr Adolesc Gynecol*. 2015;28(1):2–5.

Coverdale GE, Long AF. Emotional wellbeing and mental health: an exploration into health promotion in young people and families. *Perspect Public Health*. 2015;135(1):27–36.

Douglas K, Poletti A. *Life Narraties and Youth Culture: Representation, Agency, and Participation (Studies in Childhood Youth)*. Palgrave MacMillan; 2016.

Fox J, Warber KM. Romantic relationship development in the age of Facebook: an exploratory study of emerging adults' perceptions, motives, and behaviors. *Cyberpsychol Behav Soc Netw*. 2013;16(1):3–7.

Gentzke AS, Creamer M, Cullen KA, et al. Vital signs: tobacco product use among middle and high school students—United States, 2011–2018. *MMWR Morb Mortal Wkly Rep.* 2019;68:157–64.

Giedd JN. The digital revolution and adolescent brain evolution. *J Adolesc Health*. 2012;51(2):101–105.

Giedd JN, Raznahan A, Alexander-Bloch A, Schmitt E, Gogtay N, Rapoport JL. Child psychiatry branch of the national institute of mental health longitudinal structural magnetic resonance imaging study of human brain development. *Neuropsychopharmacology*. 2015;40(1):43–49.

Huang CY, Stormshak EA. A longitudinal examination of early adolescence ethnic identity trajectories. *Cultur Divers Ethnic Minor Psychol*. 2011:17(3):261–270.

Kohlberg L. *Essays on Moral Development: The Philosophy of Moral Development*. Harper & Row; 1981.

Magnusson BM, Crandall A, Evans K. Early sexual debut and risky sex in young adults: the role of low self-control. *BMC Public Health.* 2019;19:1483.

Mills KL, Goddings AL, Clasen LS, Giedd JN, Blakemore SJ. The developmental mismatch in structural brain maturation during adolescence. *Dev Neurosci*. 2014;36(3–4):147–160.

Mitchell KJ, Jones L, Finkelhor D, Wolak J. Trends in unwanted online experiences and sexting. Final report 2014. Accessed July 2023. https://scholars.unh.edu/cgi/viewcontent.cgi?article=1048&context=ccrc

Morgan PL, Staff J, Hillemeier MM, Farkas G, Mczuga S. Racial and ethnic disparities in ADHD diagnosis from kindergarten to eighth grade. *Pediatgrics.* 2013;132:85–93.

Murty VP, Calabro F, Luna B. The role of experience in adolescent cognitive development: integration of executive, memory, and mesolimbic systems. *Neurosci Biobehav Rev*. 2016;70:46–58.

National Institutes of Health. National Institute on Drug Abuse. Monitoring the Future Survey, 2013. *Behavior and Social Environment Indicators*. Accessed 014. http://www.childstats.gov

National Security Cyber Alliance. Keeping Up With Generation App: NCSA Parent/Teen Online Safety Survey. 2016. Accessed in 2021. https://enough.org/objects/Keeping_Up_With_Generation_App_Findings_Summary_2.pdf

Paus T, Wong AP, Syme C, Pausova Z. Sex differences in the adolescent brain and body: findings from the Saguenay youth study. *J Neurosci Res*. 2017;95:362–370.

Rothstein R. *The Color of Law: A Forgotten History of How Our Government Segregated America*. 1st ed. Liveright Publishing; 2017.

Rousseau A, Beyens I, Eggermont S, Vandenbosch L. The dual role of media internalization in adolescent sexual behavior. *Arch Sex Behav*. 2016;46(6):1685–1697.

Selman RL The child as a friendship philosopher. In: Asher SR, JM Gottman JM, eds. *The Development of Friendships*. Cambridge University Press; 1981:242–272.

Strang NM, Chein JM, Steinberg L. The value of the dual systems model of adolescent risk-taking. *Front Hum Neurosci*. 2013;7. https://doi.org/10.3389/fnhum.2013.00223

Sue DW, Sue D. *Counseling the Culturally Diverse: Theory and Practice*. 6th ed. Wiley; 2012.

Telzer EH, Fuligni AJ, Lieberman MD, Miernicki ME, Galvin A. The quality of adolescents' peer relationships modulates neural sensitivity to risk taking. *Soc Cogn Affect Neurosci*. 2015;10(3):389–398.

Trent M, Dooley DG, Douge J, Section on Adolescent Health, Council on Commnity Pediatrics, Committee on Adolescence. The impact of racism on child and adolescent health. *Pediatrics*. 2019;144:e20191765.

Umaña-Taylor AJ, Quintana SM, Lee RM, et al. Ethnic and racial identity during adolescence and into young adulthood: an integrated conceptualization. *Child Dev*. 2014; 85(1):21–39.

Van Rensburg EJ, Raubenheimer J. Does forgiveness mediate the impact of school bullying on adolescent mental health? *J Child Adolesc Ment Health*.

2015;27(1):25–39.

Willoughby T, Good M, Adachi PJC, Hamza C, Tavernier R. Examining the link between adolescent brain development and risk taking from a social-developmental perspective. *Brain Cogn*. 2014;89:70–78.

Yonker LM, Scirica CV, Jethwani K, Kinane TB. “Friending” teens: systematic review of social media in adolescent and young adult health care. *J Med Internet Res*. 2015;17(1):e4.

35

Child Psychiatric Examination

35.1 Psychiatric Examination of the Infant, Child, and Adolescent

ROBERT A. KING, M.D., PRUDENCE W. FISHER, PH.D., MICHAEL H. BLOCH, M.D., M.S., MARY SCHWAB SELIGMAN, M.D., M.A.R., WALTER S. GILLIAM, PH.D., AND MICHAEL VITULANO, PH.D.

Clinical diagnostic assessment of youth includes identification of the reasons for referral, evaluation of the nature and extent of the child's emotional or behavioral difficulties, and determination of factors in the child, family, and environment that cause, exacerbate, or potentially ameliorate them. Although information gathering and differential diagnosis are the primary aims of assessment, an ancillary goal of the evaluation process is to facilitate the development of an alliance between the clinician and the family.

Psychiatric assessment of the child or adolescent is undertaken for a variety of purposes, including outpatient office or clinic evaluation to determine the need for treatment; emergency department assessment of safety issues and need for hospitalization; consultation with pediatric, school, or legal services; as a component of family or other treatment modalities; and for research purposes.

Although the structure and style of the assessment vary to some extent depending on the purpose and setting, certain elements are common to all such assessments.

Decisions about intervention follow from evaluation, so clinical assessment may rightly be regarded as a cornerstone of child and adolescent psychiatry. Clinical evaluation of the child or adolescent may be challenging and complex in that the clinician must draw on multiple sources and a range of clinical techniques for eliciting information on various aspects of the child's functioning. This information must then be integrated using the clinician's knowledge of development and psychopathology, cultural sensitivity, and context of the symptoms to arrive at a diagnostic formulation that can guide effective treatment planning.

The initial portion of this chapter focuses on the comprehensive clinical assessment of the child and adolescent; subsequent subsections discuss more specialized types of assessment, such as the assessment of infants and adolescents, psychological testing, laboratory assessment, and clinical research assessment instruments.

CLINICAL INTERVIEW OF THE CHILD

The psychiatric assessment of the child requires a comprehensive approach that evaluates the child's developmental progress in various domains and positive adaptive capacities, as well as the presence of the pathognomonic symptoms of specific disorders. A developmental approach to the assessment of the child is essential because children differ from adults in certain key respects.

Distinctive Features of Childhood Psychopathology

First, psychiatric disturbances in children often consist of a lack of developmental progress in one or more domains rather than the presence of specific symptoms that are pathognomonic of adult disorders. For example, a nursery school child's failure to develop useful social language or interactions or a school-aged child's inability to meet the developmental expectation of separating from parents and settling into the school day may prompt an assessment.

Second, the child’s developmental status may affect the clinical presentation of various symptoms and syndromes. For example, in children, depression often presents with irritability and somatic complaints, whereas excessive guilt or depressive delusions are rare.

Third, development brings expectable periods in which distressing emotions or behavior may occur as part of a normal transition, for example, the separation anxiety of a child starting preschool or the oppositionality of the adolescent.

In many cases, clinical conditions represent severe forms of symptoms found in milder form in most children. Fears, tantrums, moodiness, or restlessness are relatively common in childhood and occur transiently at different stages. Assessment may be sought by concerned parents seeking guidance on how to understand and manage these developmental manifestations. Thus, the clinician must judge whether the behavior is likely to resolve with time and without substantial deleterious effect on the child or family or whether, instead, the level of distress, compromised functioning, or symptom persistence indicates the need for clinical intervention.

To distinguish transient or normative difficulties from those that are more clinically worrisome, the evaluator must have a solid knowledge of typical and atypical child development. This developmental frame of reference includes an understanding of what behaviors can be expected in children of different ages, the time frame within which various behaviors normally wax and wane, the natural history of psychiatric disturbances at different stages in development, and at what age particular syndromes are likely to emerge.

Many children coming to clinical attention have symptoms that cannot be neatly subsumed under the rubric of a single diagnostic label. Comorbidity is usually not the exception but the rule in childhood disorders. Even in epidemiologic studies of children and adolescents, as many as one-half of those who meet diagnostic criteria for one disorder also meet criteria for at least another disorder. This high rate of comorbidity, found even in nonreferred populations, has several likely sources. Many traditional nosologic entities draw their definition from clinical experience with adults; although childhood analogs clearly exist, the appropriate descriptive boundaries may not be the same. Furthermore, impairing symptoms in childhood tend to have a ramifying effect by interfering with the ongoing acquisition of key developmental skills in multiple areas. Finally, regardless of whether the origins of a given form of childhood psychopathology lie in biologic factors, family or social environment, or interactions among these realms, these forces often produce symptoms that cut across many diagnostic lines.

Distinctive Aspects of Child Assessment

Most adults coming for psychiatric assessment acknowledge at least some degree of self-perceived distress and wish for help. In contrast, for most children, the responsibility for initiating and following through with a psychiatric evaluation lies with the parents. In many cases, the child's behavior is a greater source of concern for others—parents, teachers, and peers—than to the child him- or herself. Other children, even while acknowledging distress, may view the locus of the troubles as purely external and hence see little need for evaluation, even if they comprehend the process. In any event, the lack of any stated wish for help or active role in seeking the evaluation colors the child assessment process from the start. As a result, explicit attention must be paid to assessing the child's view of the problem and cultivating the child's engagement in the process.

A related aspect of the child psychiatric assessment involves the need for the clinician to assess and accommodate to the limitations in the child's ability to reflect, to conceptualize, and to report experiences and internal states. These capacities emerge only gradually with maturation and are influenced by normal and pathologic factors. Thus, the clinician's communication must be attuned to the child's developmental level and may require play, stories, drawing, and other alternative modes of interaction in addition to direct discourse.

Context and the Need for Multiple Informants

Even more so than for an adult, the child’s functioning and psychological well-being are strongly related to the social context—the family, school, and community settings—in which the child lives and develops. Thus, to evaluate the nature and severity of the child’s psychiatric status, information is needed about the child’s relationship to, and functioning in, these diverse settings. The fit of the child and social environment can be appraised from what parents or primary caregivers say about their relationships with the child and from the child’s direct report, as well as from play and drawings; however, direct observation of the content and tone of parent–child interactions in the office setting also provides much useful additional data. In some cases, a home visit may provide invaluable information.

It is also important to gather information on the child’s functioning in school, a task that optimally includes speaking with the child’s teachers in addition to reviewing school reports and educational assessments. Here, too, direct classroom observation can provide invaluable in vivo information. For children involved with child welfare, juvenile justice, or other institutional systems, information from caseworkers, probation officers, and residential treatment caregivers is essential.

Special Types of Assessment

The main emphasis of this section is on the assessment of children and adolescents, especially those capable of participating verbally in the assessment process. Because developmental considerations shape the evaluation process, infants and toddlers, on the one hand, and adolescents, on the other, require specific modifications of the general approach outlined in this section. Nonverbal or children with language deficits, such as those with developmental disorders or intellectual disabilities, require specialized assessment techniques.

Although the prototype for clinical assessment of the child is the office consultation or clinic visit, evaluations occur in a variety of settings, including schools, pediatric wards, residential treatment centers, detention centers, and hospital emergency departments. Procedures necessarily vary with the setting and reasons for carrying out the assessment. The following general model for the outpatient evaluation can be adapted to fit more specific clinical situations. More specialized types of evaluation, such as those for epidemiologic or research purposes, are discussed later in separate subsections.

Preliminary Contact

In the initial phone call to the clinician or clinic intake office, it is helpful to gather some basic information about the nature of the chief complaint; to set the time, date, and place of the first appointment; and, if relevant, to discuss issues of fee or insurance coverage.

During the evaluation, the clinician will optimally see the parents alone, the child alone, and the parents and child together. The order in which these should be done and who attends the first meeting should be discussed at the initial contact. This decision is usually made with reference to the child's age. When evaluating a young child, the parents are often seen first, without the child. This allows the clinician to obtain a history and discuss how to prepare the child for the evaluation. The adolescent should usually be included in the initial interview or seen first so that the youngster does not identify the clinician as an agent of the parents, which may interfere with their willingness to talk and impair the formation of a treatment alliance with the adolescent from the start. If children or adolescents are to attend the initial appointment, the clinician should discuss how they are to be prepared for the appointment in the initial call.

In the initial phone contact, the issue of which adults will attend should also be discussed. When the family is intact, the presence of both parents or caregivers is optimal. When this is not the case, the clinician and parent should discuss whether a noncustodial parent, stepparent, or other adult should be present.

Parent Interview

In the initial parent interview, the clinician seeks to understand the parents' view of the problems that have led to the referral, including the nature of the child's current difficulties, the explicit and implicit reasons for the timing of the referral, and the effect of the child's difficulties on the parents, individually and as a couple, and on the well-being and functioning of the family as a whole. A detailed history of the child's development and a review of family medical and psychiatric disorders are obtained. Finally, the clinician seeks to understand how the family is functioning in its community and cultural setting. This includes a careful appraisal of the strengths of the child and the family, in addition to the history of the presenting symptoms and areas of impairment.

Child's Current Difficulties and Reasons for Referral. In child psychiatry, referrals are usually initiated by the parent or parents rather than the child. Similarly, parents are key in determining whether a recommendation for treatment is followed. Over and above the need to gather information about the child's symptoms and functioning, the parent interview provides an opportunity to develop an alliance with the parents around the goals of identifying and helping the child's difficulties. Success of this process requires understanding the parents' expectations and concerns regarding the evaluation and its outcome. How these issues are addressed depends on the nature of the child's problems and the parents' knowledge of and previous experience with psychiatric evaluations and treatment. Attention to the parents' view of the evaluation process allows the clinician to address unrealistic fears or excessive expectations about what can be accomplished in a given time and to reduce family stress when parental anxiety is too high.

In taking a history of the current problem, the clinician obtains a detailed picture of the parents' current emotional and behavioral concerns about the child. Parents should be asked to describe specific instances of the child's problematic behavior, frequency, intensity, duration, circumstances in which the behavior occurs, their responses, and the child's responses to it. The extent to which the symptoms cause functional impairment is ascertained by asking about the level of the child's distress, interference with social and academic activities, effect on ongoing development, and the impact of the child's behavior on others. Parental responses to manifestations of the child's distress also require careful inquiry. Do the parents respond supportively, judgmentally, or punitively and are the parents in accord or at cross-purposes in responding to the child's symptoms? Do the parents' responses encourage the child's active coping or, say in the case of overly anxious or somaticizing children, do parents over-accommodate or inappropriately truncate expectations to spare the child distress, but in such a way that inadvertently increases avoidance and apprehension on the child's part? The clinician should also inquire directly about behaviors or symptoms not reported by the parents but known to be commonly associated with the presenting problem or with disorders that the presenting problem may suggest. Thus, one goal of the initial parent interview is to develop a detailed picture of the presenting problems, their course, their impact on the child's and family's functioning, and what has been tried to help alleviate them.

The clinician is also interested in the implicit and explicit reasons that the child has been brought for help at this particular moment. Often, the reason lies in an escalation of troublesome behavior or in the deleterious effect of a deficit emerging into sharper relief (e.g., a learning disability or attention deficit becoming more apparent as the child faces the increased demands of a new grade). In other cases, the child's behavior may not have changed, but a stressor within the family may have made it more difficult for the parents to manage bearable behavior. The factors propelling families to seek clinical attention are often complex. In surveys of the general child population, rates of diagnosable psychiatric disorders approach 20%, yet the number of children actively seeking or receiving clinical services is far lower. Multiple barriers may delay or inhibit psychiatric evaluations, including cultural ones, such as different understandings of emotional problems, a lack of diversity among mental health providers, or feelings of stigma or embarrassment; lack of funds or insurance may also contribute.

Throughout the history taking, the clinician strives to understand the meaning of the symptoms in relation to the specific factors in the child and the environment that influence them. A given symptom (e.g., a theft, a period of anxiety, or hallucinations) may have quite different meanings and clinical implications in different children and in children at different developmental periods. Thus, detailed information on precipitating circumstances, behavioral concomitants, and consequences and alleviating factors is needed, much as is the case in taking a medical history. However, this information must be interpreted within the broader developmental and family contexts in which the symptoms occur; thus, while gathering history, the clinician simultaneously takes in information about these contextual elements. History taking and diagnostic formulation are not distinct processes; during the interview process, clinicians continually formulate and test tentative hypotheses that guide further questions and the diagnostic possibilities under consideration.

To understand the child's presenting problems, developmental status, and family and social context, the clinician must learn about the child's strengths, interests, talents, and areas of adequate or superior adjustment. Such information is essential for a full diagnostic understanding of the child and helps the clinician to appraise the child's overall adjustment and functioning. Moreover, in planning for treatment, factors that may help to counterbalance or remediate the child's vulnerabilities must be identified. Finally, the clinician's appreciation and discussion of the child's strengths and areas of good functioning provide support and reassurance at a time when the child's problems and limitations have been the parents' major focus of attention.

The parent interview should also cover the practical and administrative aspects of the diagnostic assessment, including discussion of fees, scheduling, confidentiality, and, if applicable, permission to obtain information from school personnel and other clinicians. If the child was not included in the initial interview, the clinician should discuss appropriate preparation of the child for a meeting.

Developmental History. The developmental history is a detailed accounting of the child's development across physical, cognitive, linguistic, social, and emotional domains. One task of this part of the parent interview is to ascertain the timing and sequence of developmental events from pregnancy to the present. Parents vary in their ability to provide precise information on the timing of specific milestones; however, an adequate sense of the course of development can often be obtained by having the parents compare the patient's progress in a particular domain with that of siblings or peers. Asking parents about the timing of changes in the family, such as moves or other memorable life events, may help to provide a chronologic framework for remembering features of the child's development. Besides gathering descriptive information about the child's development, this part of the interview often allows one to tap the emotional aspects of the parent–child relationship by querying the parents' hopes, worries, expectations, and life circumstances in relation to different developmental events.

It is useful to conceptualize the child's past and current development in the following domains: *cognitive and academic development; family relationships; peer relationships; physical development and medical history; emotional development and temperament; development of conscience and values; interests, hobbies, talents, and avocations; and unusual circumstances.*

COGNITIVE AND ACADEMIC DEVELOPMENT. The child's pattern of cognitive strengths and weaknesses should be traced from early childhood, including verbal and attentional skills. Although academic performance is important (including areas of particular strength and weakness), important information about aspects of emotional and social development also flows naturally from the discussion of how the child progressed from one school year to the next. Thus, in tracking cognitive development and academic success, one should ask about the child's ability to separate from parents and to attend school regularly, interpersonal relationships with peers and teachers, motivation to learn, ability to function independently, tolerance for frustration and delay of gratification, attitudes toward authority, and ability to accept criticism, among others. A grade-by-grade history of the schools attended can be obtained during this part of the interview, as well as the reasons for any changes. Whether the child has ever received a special education classification should be queried and the supporting evaluations, specific resulting classification, and accommodations and services agreed upon should be reviewed.

When the child’s behavior or progress at school is among the reasons for the psychiatric evaluation, one should obtain permission to communicate with the child’s teachers, counselors, or other school personnel and to review the school records, including results of standardized tests.

FAMILY RELATIONSHIPS. The parent interview should include assessment of how the child relates to each family member and how the child fits into the overall family system. The child’s reactions to major family changes should be noted (e.g., deaths; birth of siblings; marital separation, divorce, or remarriage; and changes in caretaking arrangements, custody, or visitation), as well as parental responses to those reactions. Compliance with family rules and standards should be queried, as should consequences when the child does not comply (i.e., the parents’ usual modes of discipline or limit-setting) and the child’s response to such interventions.

PEER RELATIONSHIPS. The clinician should gather information about how the child relates to peers, including the number of friends, preferences regarding age and gender of friends, and any major changes in peer group; the child’s satisfaction with these relationships; their relative reciprocity and stability; activities and interests shared with peers; and parents’ feelings about the child’s close peers or lack thereof. The parents’ perspective on the child’s social skills and deficits should be assessed, including their sense about any difficulties the child has in this domain. Episodes of bullying or being bullied, and the consequences thereof, should be specifically queried. For adolescents, this part of the history includes issues such as the capacity for intimate relationships, romantic interests, sexual activity, and concerns about sexual orientation and gender identity. The clinician should gather information about the child’s social media usage and peer engagement through interactive online gaming.

PHYSICAL DEVELOPMENT AND MEDICAL HISTORY. Assessment of physical development includes fine and gross motor development, toilet training and its lapses, eating behavior and attitudes, and sleep patterns. Precocious development or delays in physical growth or pubertal maturation should be noted, as well as the child's concerns and reactions. In this context, parents should be asked systematically about medication, illnesses, hospitalizations, emergency room visits, operations, episodes of loss of consciousness, seizures, head injury, or other serious injuries, as well as the child's reactions to these and their effect on subsequent health and activities. Inquiry about tics, difficulties with hearing or vision, lead exposure, and other specific conditions may be suggested by the clinical presentation. When necessary, medical reports should be obtained and reviewed.

EMOTIONAL DEVELOPMENT AND TEMPERAMENT. Although some aspects of emotional development and temperament will have been covered in taking the history of the presenting problems, the clinician should fill in gaps with a systematic review. This domain of assessment includes the child's present and past mood and capacity for emotional regulation, anxieties, and ability to adapt to changes and situations that are new, challenging, or frustrating. Prevailing mood as perceived by the parents should be assessed; the clinician should ask specifically about self-injury and suicidal ideation, gestures, or attempts, especially if the child has shown depressive symptoms, including irritability. Worries, fears, and other manifestations of anxiety should be covered, and the level of any distress and impairment should be explored. The generalized or specific nature of symptoms, situational evokers, historical precipitants, and effects on the child in terms of impairment or distress are relevant. It is also important to note unusual fears, excessive shyness or withdrawal, and obsessive or compulsive symptoms.

Emotional regulation (ER) describes the child's ability to manage emotional arousal (including negative affects related to frustration), an ability that develops over time, has both heritable and environmental components, and is influenced by attachment behaviors and caregiver responses. Irritability (elevated anger-proneness and persistence of negative mood relative to peers) is a form of *emotional dysregulation* that is emerging as a robust trans-diagnostic trait or dimension that is moderately persistent over the life-span and confers risk for future anxiety, depression, and impaired role functioning. Various research paradigms and instruments have been developed to assess and study this concept. Clinically, however, it can be assessed by inquiring about: (1) the types (and thresholds) of stimuli or situations that elicit negative affect on the child's part (what has been termed the *reactive* component of ER); (2) what sorts of strategies the child employs to reduce, balance, or recover from negative emotions and how successful these efforts are or not (what has been termed the *regulatory* component); and (3) the parents' role in helping the child with these coping efforts.

The child's capacity to regulate aggression falls within this domain and should be reviewed. The clinician should ask about circumstances in which the child becomes angry or aggressive, the child's mode of expressing anger (i.e., verbal, physical, or both), how this affects others, and the child's reaction to and processing of such feelings and behavior. In this context, the clinician seeks to evaluate how the child manages aggression and whether the child is too aggressive or too fearful of their anger or that of others. Specific aggressive symptoms (e.g., bullying, fire setting, vandalism, or cruelty to animals) should also be inquired after.

DEVELOPMENT OF CONSCIENCE AND VALUES. The clinician should assess the development of conscience to gauge whether it is too harsh, lax, or overly focused on particular issues. The effectiveness of conscience in helping the child conform to family and societal expectations is important. Religious or ethical concerns and their concordance with those of the family can be addressed in this phase of history taking. The family's expectations with regard to the child's values and future life choices should be ascertained from the parents' perspective, and areas of potential discord should be noted.

INTERESTS, HOBBIES, TALENTS, AND AVOCATIONS. Although this line of inquiry is routinely pursued with the child, it is useful to obtain the parents' perspectives on the child's interests and activities and to assess the parents' approval, involvement, and support for them. Indications of parent–child conflict should be probed and provide useful information about family relationships. The presenting difficulties may have affected the child's ability to focus on or to engage in previously necessary endeavors, such as completing schoolwork, and enjoyable areas of interest or activity, such as sports or music.

The quantity and type of mobile phone use and screen time approved by parents or caregivers—including social media, texting, television, and videogames—and the extent of any resulting conflicts about parent approval provide important information on the child's interests and the quality of parental supervision, monitoring, and limit setting.

CHILDHOOD ADVERSE EXPERIENCES. The clinician should assess for the child's experience of cultural discrimination or isolation, immigration, or exposure to traumatic events. The clinician should assess for a history of sexual or physical abuse, domestic or community violence, race-based trauma, natural disaster, or armed conflict. If a history of trauma exists, the child's immediate and subsequent reactions and the nature of the parents' or caregivers' response should be assessed. If a child or adolescent discloses a history of child maltreatment, such as physical or sexual abuse, neglect, or emotional abuse, which has not previously been reported, the clinician must report it to the appropriate local child protective service agency.

Family and Community Background. Children develop in the contexts of the family and community, and these provide opportunities and challenges that influence and shape the course of that development. The clinician should gain a picture of the child's functioning in these contexts and their influences on the child. Although the clinician learns a great deal about these areas from taking the history of the presenting problem and the developmental history, direct questioning may also be needed.

PARENTS. The clinician should assess the parents as individuals and as a couple, including strengths, weaknesses, and areas of conflict or difficulty. If the child resides with someone other than the parents (e.g., a relative or foster parent), the clinician should obtain information on the history and circumstances of that relationship, including how the child came to be in that adult's care. It is also necessary to determine what the child's previous experiences were with the parents and the nature of the bond with them. The clinician should similarly assess the current caregivers and their parenting relationship with the child.

If the child is adopted, information is needed on the circumstances, the parents' feelings and expectations about the adoption, the child's age at adoption, the child's adjustment at the time and later, and how the child and family members currently view the fact of the adoption. The clinician should also determine how the adoption has been discussed with the child and any concerns or questions about the biologic parents.

When taking the developmental history, it is often natural to inquire about the parents' feelings and involvement with the child at different stages. Parents' past attitudes and behavior are likely to shed light on their responses in the current situation; hence, one wants to know how they felt about past difficulties and what they have done to help and to support the child. The parents' individual views on the causes of the child's difficulties, and the extent to which they agree or disagree on this and other aspects of the child's care and management, affect how they respond to the clinician's recommendations.

The clinician should ask about the parents' cultural, ethnic, and religious backgrounds and traditions. These may be a source of shared identity and values to be imparted to the child or a source of conflict between the parents or between an adolescent and the parents. If the clinician is not very familiar with the culture or language of the family to understand the cultural context of clinically relevant aspects of the child and family's situation, an attitude of "cultural humility" is needed, and a respectful desire to know more about their culture should be conveyed. The clinician should guard against making assumptions about race, ethnicity, appearance, or preconceived notions about culture.

The parents' education and occupation may bear on how they understand and respond to the evaluation. How knowledgeable and worried they are may depend on their past experiences and familiarity with psychiatric issues. Financial resources and insurance coverage should be ascertained as these may be a source of stress or worry and thus pose practical constraints on the treatment options available to the family.

The clinician should gain an understanding of how the family system functions, with whom the child lives, and the nature of relationships between household members and the child. Other issues to explore are boundaries and alliances within the family and the child's affinities and conflictual relationships with parents, siblings, and other household members. Family communication and problem-solving styles should be assessed, including how issues of separation or disagreement are handled. The emotional tone of the family (e.g., tense, anxious, critical, competitive, warm, or supportive) can be ascertained by observing interactions and by direct inquiry. Stresses on the family as a whole or individual members (e.g., moves, migration, illness, accidents, job loss, or economic or legal difficulties) should also be addressed.

Adverse childhood experiences, beginning in utero, including separation, abuse, neglect, exposure to violence, poverty, and parental substance use or depression, have deleterious cumulative effects on the child's development over time. Long-term effects of adverse childhood experiences cross traditional diagnostic boundaries, impact development, and are known risk factors for multiple forms of medical, psychiatric, and social adversity in later life. These factors often operate across generations. Hence, information on the parents' backgrounds and histories with their own families of origin is essential to understanding their relationships with and responses to their child.

COMMUNITY. The clinician should understand the community or neighborhood in which the family lives and their feelings about and identification with it. Information on the family's involvement with civic, community, and religious activities (and the extent to which the child participates) should be ascertained. Is the family socially isolated by reason of discrimination, ethnicity/race, immigration status, or other forms of stigma (e.g., related to mental health or chronic illness)?

Learning from parents and adolescents about their neighborhood is critical, especially for a deeper understanding of the child and adolescent's environment. What is the family's housing like, and what options are realistically available? What is the level of residential segregation? How serious is crime or community violence, and how has it impacted the family directly or indirectly? What steps do the parents take to try to keep their child or teen safe? How are the police regarded in the neighborhood—as protective, potentially dangerously adversarial, or some admixture thereof? What has been the family's and child's history with law enforcement and the justice system? How do other neighborhood factors, such as lack of childcare, grocery stores, or access to transportation impact the family?

Numerous studies point to the centrality of social determinants of mental health issues for children and adolescents, especially the burdens on youth from marginalized groups, including racial and ethnic minorities. Many of these factors are community-based and reflect multi-generational histories of residential and educational segregation. These result in economic, educational, and healthcare disparities whose potential impact on the child and family must be considered and need to be assessed. These disparities can be considered under several headings: Economic instability; educational disadvantage; access to health care; neighborhood environment. Although these factors can be examined at the level of the individual parents and family, broader social determinants impact how a given family got there.

For example, inadequate household income, parental unemployment, and housing and food insecurity are fundamental problems with immense mental health implications for any given family. However, ameliorating the perpetuation of these risk factors for a family involves addressing broader community and social factors that are often difficult or impossible for a family to surmount, at least without local resources and advocacy that transcend those usually available in most clinical programs. Educational inequality is also often multi-generational, truncating parents' economic, vocational, and housing opportunities. These in turn impact their children, beginning with lack of access to early childhood education programs, adequate neighborhood schools, availability of tutoring and other academic assistance programs, and the absence of role models for academic success.

Stresses related to racial and/or ethnic discrimination can have a cumulative negative impact on parents and children. In addition to influencing factors such as housing, occupational, and educational discrimination noted above, racial traumas also include microaggressions, various forms of in-person and online harassment, and exposure to racial violence, either directly or vicariously via news or other media. Inquiring about the occurrence and impact of race-related stressors on the parents and youngster requires tact, training, humility, and clinicians' self-awareness of how their own racial identity, position of privilege, and potential unconscious biases influence how they interact with patients.

Beyond the parents' and family's own prior experiences with mental health services, community attitudes may strongly influence how the child and family approach a psychiatric evaluation. In some communities, mental health clinicians and agencies are viewed positively; in others, their use is stigmatized or viewed with suspicion as potentially racist or coercive, which may adversely affect the child. These issues should be assessed and understood because they may affect how and where the child receives any indicated treatment, how the treatment goals are formulated, definition of treatment type appropriate of the particular patient, and selection of provider best matched to the patient's needs. In some systems, it is unfortunately not possible to choose the ideal match for referral due to lack of resources.

The DSM-5 Cultural Formulation Interview (CFI) and its supplemental modules provide a cultural assessment tool that may help clinicians with this important task (https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/APA_DSM5_Cultural-Formulation-Interview.pdf;

https://mn.gov/dhs/assets/supplementary-modules-cultural-formulation-interview_tcm1053-413785.pdf).

Family Medical and Psychiatric History. A careful family medical and psychiatric history is needed to identify any medical or psychiatric disorders with potential genetic or environmental implications for the child. Among those to be routinely considered are psychotic disorders, mood disorders, anxiety disorders, tic and obsessive-compulsive spectrum disorders, alcohol or substance use disorders, attention-deficit disorders, learning disorders, and antisocial personality disorder, as well as metabolic and neurologic disorders. The impact of any family members' symptoms, hospitalizations, or incarcerations should be probed, with particular reference to the child.

Special Issues in the Parent Interview

INFORMANT DIFFERENCES. Although parents' concerns usually play a major role in initiating the evaluation, parents may perceive the child's problems differently and may disagree about their severity and the need for treatment. Parents' accounts of the child's difficulties and of various aspects of the developmental history may not agree completely with each other or with those of the child, teachers, or records of past events. To note these discrepancies is not to denigrate the reliability of parents as informants but rather to emphasize the necessity of obtaining information from multiple informants.

Informants provide differing accounts of children's problems for various reasons and discrepancies should be noted and regarded as potentially useful data regarding what they reflect about the clinical situation. Informants may interact with the child in different settings that vary in the extent to which the child expresses disturbing feelings and problematic behavior. This is commonly true for behaviors expressed most dramatically in settings that challenge the child in vulnerable areas (e.g., the school's demand that a child with attention-deficit/hyperactivity disorder [ADHD] sit still and pay attention). Information about such situation-specific symptoms is best provided by those who spend the most time with the child in that setting (in this instance, teachers). These different perspectives on the child's behavior may also have led to tension between informants who see the child discrepantly. Discrepancies are often found between divorced or separated parents, whose child may show quite different behavior in the context of each parent's home. The clinician should be aware of these different perspectives and the feelings that they generate so that the focus is maintained on evaluating and providing services for the child, with adults working together rather than at cross-purposes.

Another reason for discrepant reporting involves differences in how the informants evaluate the child's behavior. Observers may judge the same behavior by differing standards or expectations for many reasons. The clinician's appreciation of the informant's psychology or cultural background is useful for understanding and working with such differences. For example, when one parent is overly worried about the child having a serious disorder and the other is too ready to chalk up symptoms as something the child will "just grow out of," the clinician must navigate carefully to allow both parents to move to a reasonable middle ground that best serves the child's needs. In such instances, understanding each parent's defensive style helps the clinician to avoid appearing to collude with one parent and alienating the other.

Discrepant reporting often results from differences between informants in their willingness or ability to report their views to the clinician. For example, parents or caregivers may be quick to report a behavior that they find disturbing or annoying, whereas a child may be unable to describe the problematic behavior verbally, may not appreciate the level of disturbance that it creates for others, or may minimize or omit it due to feelings of shame or fear of being blamed or punished.

In general, methodologic studies and clinical experience suggest that parents are more likely than children to report disruptive or externalizing child behaviors, such as restlessness, impulsiveness, opposition, or aggression. Conversely, children may be more likely to report anxious or depressive feelings and symptoms, including suicidal thoughts and acts, of which the parents may be unaware. Because children may be immature in their capacity to report information, such as frequency and duration of symptoms, parents are usually more accurate informants about chronologic details. The child or adolescent, however, is often the only source of information on traumatic events, such as sexual abuse, exposure to violence, or behavior that occurred in peer contexts (e.g., bullying). In general, children become more reliable reporters of specific symptoms and time frames with age, with prepubertal children tending to be less reliable reporters of behavioral symptoms than their parents.

Mothers and fathers often differ in their experiences with their children, depending on the nature of their caregiving roles, the child's age and gender, shared interests, and the respective temperaments and personalities of the parents and child. Although mothers have been reported to be more reliable informants than fathers and traditionally spend more time with their children, it is important to gather information from all adult caregivers and to involve and form an alliance with both parents from the outset. When appropriate, other informants, such as siblings, grandparents, or childcare providers, may be called on to contribute information, but only after the implications for the family system and the child's specific relationships with household members are carefully thought out.

Parents and teachers may view the nature and severity of the child's difficulties differently because of the situation-specific nature of some behaviors and symptoms. When present at home and school, symptoms such as hyperactivity and inattention may carry a different clinical significance and prognosis than those noted in only one of the settings. Parents and teachers may also differ in their appraisals of the child's behaviors. For example, a stressed and preoccupied parent or caregiver may be less attentive to her child's behavior, whereas an observant teacher with years of experience with similar-aged children may readily identify the child as distressed, inattentive, or disruptive.

For the clinician conducting the assessment, the issue is not which informant is right but how to understand the differences in viewpoint most effectively for the patient's benefit. Thus, the clinician must elicit and then synthesize the information about the child's behavior and feelings in multiple settings, like school, home, on sports teams, and with peers, in order to provide an optimal clinical evaluation.

INTERVIEW TECHNIQUES WITH PARENTS. The clinician may need to use a variety of interview techniques to elicit information from parents. The type and quality of information obtained are, in part, determined by the methods used and the clinician's skill and comfort in using them. A series of empirical studies confirm that systematic factual information and details about onset and timing of symptoms are generally most easily elicited by specific and direct questions, whereas an understanding of feelings and relationships in the family is usually gained most effectively by means of more open-ended, indirect questioning strategies.

In the initial interview, parents should have the opportunity to tell their story in their own way. For rapport to develop, they must feel that they have been able to express their concerns in their own words and were heard and empathically understood. An open-ended beginning also avoids narrowing down too quickly on a single aspect of the child's situation and conveys the clinician's interest in many different aspects of the child's situation. Subsequent probing can clarify details and help shape the clinician's understanding of diagnostic and potential treatment issues.

Parents are most likely to be nondefensive and comfortable about sharing personal information when they feel that the clinician is trustworthy, interested, understanding, and nonjudgmental. Clinicians must thus convey respect and empathy and avoid seeming either aloof or excessively familiar. They should discuss the child and family's situation in a natural way, using words and concepts that the parents can understand. The range of questions and the tone of the inquiry convey the clinician's interest in the whole child (including strengths, interests, and talents) and an appreciation that the child is not merely the bearer or repository of a set of symptoms.

EXPECTATIONS AND ANXIETIES. Parents may hold strong preconceived ideas about the nature of their child's problem and the kind of help needed. Because of the extensive coverage of issues relating to psychotropic medications in the popular press, many parents have drawn conclusions about the appropriateness of medications in general and in relation to their child's needs. For example, parents may arrive convinced that their child needs fluoxetine (Prozac) or that methylphenidate (Ritalin) is an addicting and inherently harmful drug.

Preconceptions or parental anxiety or conflict may also be played out around the issue of blame for the child's problems. In such situations, a parent may maneuver to have the clinician arbitrate or agree that the teacher, school, or other parent is responsible for the child's behavior and distress. The clinician should be alert to such situations and should handle them without making either parent feel defensive or discounted. If ignored, such hidden or overt agendas can have a deleterious effect on the assessment process and the likelihood that treatment recommendations will be accepted or successfully implemented.

Parents may bring other entrenched, preconceived agendas to the clinical evaluation process. A parent may insist that the evaluation confirm that the child has been abused (e.g., by a former spouse) or does not have a psychiatric problem but a medical one. Alternatively, the parent may demand that the clinician agree that there is nothing wrong with the child. In such instances, clinicians must explain carefully and clearly that they will conduct a thorough assessment and provide thoughtful professional advice, adding that the assessment must be done openly, without preordained conclusions. This may not satisfy some parents, and some situations may simply not be workable.

Bringing a child for psychiatric evaluation is a charged and emotionally difficult step for nearly all parents, and often it is only taken after much worry and disappointment that various accommodations, home remedies, or school-based adjustments have not sufficed. Children are the repositories of parents' dearest hopes (and often darkest fears) for themselves. Although a parent may rationally conclude that seeking consultation is reasonable and necessary, admitting that one's child or oneself needs help is often painful. Submitting one's child and one's parenting to professional scrutiny stirs many anxieties, as well as positive expectations of help, in most parents.

Parents often bring unspoken anxieties about the child's possible diagnosis or prognosis to the evaluation. These anxieties may be based on the parent's personal or family experience with psychiatric disturbance or may derive from accounts in the popular media. For example, is the child's uneven development a sign of autism or intellectual deficiency? Do the facial grimaces mean that the child is developing a disabling tic disorder? Is the child's moodiness a harbinger of a serious mood disorder or suicide? Other parental anxieties may be guilty ones. Is failed parental rearing or transmitted heredity somehow to blame for the child's difficulties? Is the child fated to share the parent's (or other relative's) painful struggle with depression, anxiety, or psychosis? Does the parent's relationship to the child seem to recapitulate some distressing aspect of the parent's own early family experience?

The clinician must be aware of such parental anxieties and thus remain alert for hints of them as the evaluation progresses; some parents are too anxious or ashamed to be forthcoming about them. By and large, it is best to have these concerns aired, even if they cannot be settled at the initial stage of evaluation and perhaps cannot be settled for some time. When such anxieties are revealed or suggested indirectly, the clinician must express an interest in them with sensitivity. Respect and concern for these anxieties can be conveyed without dismissing them or providing premature or unfounded reassurance. The clinician may need to acknowledge the limits of knowledge about cause or prognosis in individual cases. Parents often seek psychiatric assessment after years of worry, having been given well-meaning (but inappropriate) reassurances that the child would grow out of it. Frank discussion of these issues may provide some relief to the parents, albeit perhaps not definitive answers to their questions.

PREPARATION OF THE CHILD. If the child or adolescent does not attend the initial interview, the clinician and parents should discuss the preparation of the child for their first appointment. Children have many reactions to a first visit to the psychiatric clinician, and parents must anticipate their child’s possible reaction and prepare their response. Children may feel anxious because they cannot envision what will occur or because the evaluation may feel like a punishment. Some children fear that the assessment will reveal some profound or irremediable problem (this may be projected onto the clinician or parent with the reproach, “You think I’m crazy”). Many younger children fear shots or having to get undressed for a physical examination. In most cases, the child is probably aware of many of the parental or school concerns that have led to the evaluation and may feel a sense of relief (which can nevertheless coexist with anxiety) about the possibility of getting some help. The child’s anxiety or defensiveness may be reduced if the purpose of the visit is put in supportive (rather than pejorative or accusatory) terms. The parents might be encouraged to tell the child something like, “We know you’ve been having a hard time at school [or feeling upset and sad or having trouble with friends], so we are going to see Dr.—, who is used to helping kids with that sort of problem and can maybe figure out what would help you.” For younger children, explanatory terms, such as a *feelings doctor* or a *talking doctor* who helps children with problems and worries, may be useful. The parents must give permission and encourage children to tell the clinician about their concerns, even those that might seem to involve private family matters not otherwise spoken about to others.

Child Interview

In the child interview, the clinician directly assesses the child’s psychological functioning, developmental status across different domains, and perceptions and experiences of the presenting problems. Consideration of the child’s age and development is critical for the clinician to approach the child in a child-centered manner. Like the parent interview, the child interview also provides an important opportunity to establish rapport. Such rapport reduces the level of stress experienced by the child during the assessment process, enhances the child’s engagement and willingness to share information, and facilitates the child’s cooperation with subsequent recommendations for intervention. Thus, from the outset, the clinician should convey empathy and interest in understanding and helping the child.

There has been debate regarding the reliability, validity, and clinical usefulness of various approaches to the child interview, but it is generally agreed that the direct child interview is an essential component of a comprehensive clinical child assessment. It provides observational data and diagnostic and historical information that are not accessible from other sources. In particular, information about feelings and moods, level of distress, and the child’s view of significant events is essential for a full diagnostic understanding of the child’s problems. For their own reasons, parents may not understand or fully appreciate the importance of these issues, or the child or adolescent may not have divulged them. The child may be guarding important information, such as secrets about abuse, bullying, sexual or antisocial activities, substance use, suicide attempts, pathologic patterns of eating, or collusive family dynamics. Certain symptoms or forms of distress, such as hallucinations, compulsions, premonitory tic urges, obsessive or suicidal thoughts, or concerns over sexual orientation or gender identity, may be apparent only to the child. Other symptoms may escape or mislead the untrained eye; detection of tics or a thought disorder may require direct clinical observation.

There are many varieties of child interviews, depending on the nature of the presenting problem, the clinical setting, and the purpose of the assessment. Thus, the scope, focus, duration, setting, and predominant techniques used vary, depending on, for example, whether the assessment is an emergency department evaluation of a suicidal threat, a pediatric ward consultation of a regressed or uncooperative child, a forensic custody evaluation, an investigation of alleged child abuse, or the evaluation of a child who is academically underachieving, anxious, or manifesting tics.

The clinician approaches the child interview with a developmentally based understanding of child and adolescent functioning and psychopathology. Within this framework, the clinician expects to obtain information relevant to the diagnostic questions raised by the presenting problems and so systematically reviews the major domains of symptomatology. The order and manner of conducting the child interview vary with the child’s age and developmental status, the nature of the presenting problems, the setting and context, and the clinician’s style. The constant elements in the interview process are the orienting framework chosen and the clinician’s clinical skill in engaging and eliciting relevant data from children of varying ages and capacities with diverse problems and concerns. Some data may emerge spontaneously; some require questioning or other deliberate means of eliciting information.

Components of the Child Interview. The components of the child interview are often conceptualized under two broad headings: History taking and the mental status examination. *History taking* involves inquiry into the significant areas of the child’s life and functioning, past and present, including the presenting problem. The *mental status examination* consists of an assessment and description of the child’s appearance and functioning as manifested in the interview. Although these aspects of the child interview can be conceptually distinguished, they usually proceed concurrently. For example, the interviewer may ask a specific question, such as who lives at home with the child or how the child gets along with a sibling or a teacher. From a history-taking point of view, the child’s response provides some information about these facts, as well as how the child feels about them. Although the child’s response is thus of interest for the historical information it reveals, the child is also providing the clinician with data relevant to the mental status examination (e.g., how the child conceptualizes the response, how verbally fluent the child is, how readily the child engages, how confiding or suspicious the child is toward the interviewer). Similarly, while engaging a young child in play with figures or puppets, the clinician can simultaneously gather information for the mental status examination (e.g., speech, relatedness, and conceptual skills) and inferential data about the child’s feelings, fantasies, and conflicts.

Several distinctive characteristics of clinical work with children bear on the manner of conducting the child interview and mental status examination and on interpreting the data that they yield. The focus of history taking and the mental status examination of the child is developmental, that is, it seeks to describe the child's current presentation in various domains and to compare the child's functioning in these areas with that expected for the child's age and phase of development. Another distinctive aspect of the child interview involves the child's potential lability and propensity to react in more immature ways when tired, sick, anxious, or stressed by an unfamiliar situation. Although one can sometimes gather sufficient information for an assessment from a single interview, usually more than one interview is needed to assess the child's optimal or characteristic level of functioning.

Mental Status Examination. For the mental status examination, the clinician observes and assesses the following areas:

Physical appearance: Includes physical development, presence of minor congenital anomalies, style of hair and dress, cleanliness, and other indicators of the quality of self-care or parental attentiveness (or its lack) to the child's physical appearance and hygiene.

Manner of relating to examiner and parents: Includes ease of separation, how confiding or guarded the child seems toward the clinician, eagerness to please or to impress, flirtatious behavior, readiness to make excuses or to minimize, passive withholding, or open defiance. The child's behavior toward the examiner is compared to that observed or reported toward the parents or other adults. Reactions to the clinician's physical appearance, gender, or ethnicity (especially if different from that of the child) should be noted.

Affect: The child's predominant mood and range of emotion during the interview and appropriateness of mood.

Coping mechanisms: As manifested toward the clinician, expressed in play, or described by the child in relation to important others; age-appropriateness of the child's dependent longings, sexual interests and impulses, and aggressive feelings, with respect to intensity, object, and mode of expression; child's control or modulation of such urges (e.g., finding alternative or socially permissible means of satisfying them); strategies used for coping when anxious or frustrated.

Orientation to time, place, and person: Whether children are accurately aware of the date, where they are, and who they are. This can be evaluated by asking the child, "What is the year? season? date? day? month?" or "Where are we?" (name(s) of building, town, state, country).

Motor behavior, including activity level: Coordination, presence of unusual postures or motor patterns (e.g., tics, compulsions, or stereotypies, such as hand flapping or twirling).

Quality of thinking and perception: Includes presence of hallucinations, delusions, thought disorder, flight of ideas, and adequacy of hearing and vision.

Speech and language: Includes reading and writing; articulation, inflection, pitch, rhythm, and fluency of the child's speech; richness, limitation, or idiosyncratic aspects of vocabulary and syntax; presence of echolalia or persistent neologisms; misuse of pronouns and gender; delayed or deviant social use of language; and lack of accompanying nonverbal communicative behavior (e.g., lack of facial expression or eye contact with interlocutor). When evidence of delays or problems in these realms exists, formal speech and language evaluation by a qualified specialist may be warranted.

Overall intelligence and fund of knowledge: Capacity for classification, abstraction, and inference appropriate to age; counting, alphabet, days of the week, months of the year, reading, writing, computational abilities; geographical and historical facts.

Attention, concentration, and impulsivity: Level of attention to an activity or topic of discussion, degree to which child jumps from activity to activity, need for physical limits, distractibility (e.g., by outside noises).

Memory: Immediate short- and long-term memory (e.g., repeat three items and recall at 5 and 30 minutes).

Neurologic functioning: Fluctuating alertness, tremor, nystagmus, choreiform movements, abnormal gait, and cerebral dominance (preferred hand, foot, eye, or ear); testing for neurologic soft signs can be incorporated in a game-like fashion (e.g., ball throwing, walking on a line, heel-to-toe walking, evidence of hand overflow when walking on outside of feet, repetitive or sequential finger tapping, and alternating pronation–supination of the hand are relatively reliable components of a standardized exam for neurologic soft signs).

Judgment and insight: Especially concerning the presenting problem; this is most usefully judged after the child has had the opportunity to develop some rapport with the clinician because the child's initial impulse may be to deny or to minimize the parents' presenting concerns.

Preferred modes of communication: Is the child open to talking directly about the presenting problems and the significant aspects of their life, or is the child more comfortable with indirect modes of expression, such as play or drawing?

Reliable detailed assessment of a child's speech and language capacities, intelligence, academic attainment, attentional and executive functioning, and memory requires standardized psychometric testing. Some simple tasks, such as those used in the Mini-Mental State Examination (as adapted for children by R. A. Ouvrier and colleagues), can be incorporated in a game-like fashion into the interview of the child to provide a rough screen of higher mental functions, such as orientation, attention, memory, language, and constructional ability.

Although the clinician should usually evaluate all areas of the mental status assessment, the emphasis and detail given to the different elements vary with the presenting problem and the type of assessment. For example, in an agitated, potentially delirious child on a pediatric ward, particular attention is paid to the child's orientation, alertness, memory, verbal coherence, neurologic functioning, and the possible presence of hallucinations.

Child Interview Techniques. A range of techniques is used in the psychiatric interview of the child or adolescent. The choice and timing of techniques depend on the child's developmental, cognitive, and linguistic level; the emotional difficulty of the issue being addressed; and the degree of rapport between child and clinician. Being able to gauge the potential usefulness of a technique at a particular moment and feeling comfortable with shifting from one technique to another as the occasion demands are important acquired skills for those conducting child and adolescent clinical assessments.

PLAY TECHNIQUES. Children younger than 8 years of age have less mature capacities to recount their feelings or interpersonal interactions verbally. For these younger children, as well as many older ones, play is a useful adjunct to direct questioning and discussion and is often a less challenging mode for the child. Some children find it easier to communicate in displacement; thus, imaginative play with puppets, small figures, and the like, either solo or in interaction with the interviewer, can provide useful inferential material about the child's concerns, perceptions, and characteristic modes of regulating affects and impulses.

The skilled interviewer can facilitate the child's engagement in play without prematurely introducing speculations or reactions that might distort or cut short the presentation of certain types of material. During the course of play, the clinician follows the sequences of play content, noting themes that emerge, points at which a child backs away from the storyline or shifts to a new sequence or activity, and situations in which the child gets stuck or falls into a repetitive loop. To facilitate the play component of an interview, the interview room should have a supply of human and animal figures or dolls and appropriate props. These should be relatively simple because elaborate toys can become distractions rather than serve as vehicles for the expression of the child's fantasies. Stock figures (such as Barbies, superheroes, or Disney characters) may impose specific storylines and themes and thus may limit access to the child's own concerns.

Not only is the content of the child's play a rich source of information, but careful observation of the form of play also provides important information for the mental status examination. During imaginative play, the clinician can observe the child's coordination and motor capacities, speech and language development, attention span, readiness to engage the interviewer, capacity for complex thought, and affective state. Absence of imaginative play or very limited, concrete, noninteractive play may indicate an autism spectrum disorder.

Games, such as cards or board games, are also useful for putting the child at ease and developing rapport. These provide opportunities to observe the degree of the child's engagement in and enjoyment of the shared activity; how the child manages competition, including reactions to winning or losing; or whether the child is prone to cheating. Some play activities (e.g., throwing a ball back and forth or easy card games, such as War) are simple enough to permit ongoing conversation while helping to discharge tension and diminish the pressures of the interview situation. As with toys, elaborate games (e.g., chess) should be avoided; games that demand much cognitive energy and concentration usually preclude discussion of issues relevant to the assessment and lend themselves to avoiding issues involved with the child's difficulties.

PROJECTIVE TECHNIQUES. In addition to imaginative play, projective techniques often help provide an indirect picture of concerns that the child may be reluctant or unable to report directly. These techniques can help the child to feel more comfortable with the clinician, are often experienced as fun, and may provide access to concerns that are important to the diagnostic formulation.

One commonly used technique is picture drawing. The child can be asked to draw a picture, leaving the choice of subject up to the child or, alternatively, can be given a specific request (e.g., to draw a person or the child's family doing something). When the picture is finished, or nearly so, it is helpful to compliment the child's effort and to express interest in what is happening in the drawing. The child's elaboration provides information that may not be readily apparent from the drawing itself. The content and form of the drawing offer a window on the child's emotional concerns and aspects of intellectual and visuomotor development. For example, the relative size and placement or omission of family members in a family drawing may be important nonverbal indicators of the child's perceptions or feelings about the family. Aggressive or sexual themes may be reflected more readily in drawings than in words. Self-image may be indicated through depictions of the child as grotesque, inconsequential, superhumanly invulnerable, or different than the child's birth gender. The clinician should become familiar with the developmental progression and norms for human figure details, such as limbs, joints, facial features, and clothing, which can provide a useful rough estimate of intellectual maturity. Various systems have been developed for systematically assessing the cognitive and emotional aspects of children's drawings. In addition, the child's behavior and speech while drawing may yield useful information (e.g., throwing the picture away unfinished and saying that it is no good).

Frequently used verbal projective techniques include asking what animal the child would most or least like to be or whom the child would pick to take along to a deserted island. It is useful to ask what the child would do with three magic wishes; if elaboration is needed, the clinician can explain that the wishes could be to have anything, to have the world be different in any way, or to change oneself in any way. Children's responses are often revealing. Some may needily or impulsively wish for material possessions, such as a video game or a billion dollars. Other responses may reveal longings to change distressing circumstances, such as “for my mom and dad to get back together again,” “not to have tics anymore or to get teased about them,” or “to have a dad who doesn't yell at me all the time.” Still other children appear uncomfortable wishing for something for themselves, preferring instead seemingly altruistic responses, such as “no more poverty or wars.” Children's responses can be used as the starting point for further exploration. For example, the child who wishes for “a big house and lots of money” may be asked who else would live there and what they would do together. Children who wish for “no more fighting in the world” can be asked if there are some particular fights that they would especially like to stop.

Some interactive imaginative techniques useful to interviewers comfortable with them provide elements of playing a game, which may appeal to the child. The squiggle drawing game developed by Donald Winnicott consists of the clinician drawing a curvy line and asking the child to turn it into a picture of something; the child then draws a curvy line that the therapist elaborates, and so on, taking turns. The Despert Fables are a series of incomplete, affectively evocative stories that the child is asked to complete.

Asking the child to tell about a dream or a book, movie, or television program can provide information about the child’s interests and preoccupations. (If the clinician is familiar with the plot, the distortions the child may introduce can be informative regarding the child’s cognitive and emotional style.) Inquiring about what the child would like to do for a living when grown up can provide insight into the child’s aspirations, values, and concerns, as well as those of the family.

DIRECT QUESTIONING. Direct questioning about the current problems and other aspects of the child’s life requires tact, attention to the child’s level of cognitive and linguistic development, and regard for the effect of the questions on the child’s self-esteem. The goal of such discussion is to determine how children see their world and their functioning in it, including attachments and antipathies, pleasures and anxieties, and strengths and weaknesses, which may include the difficulties that led to the evaluation.

It is important to phrase questions in language that is appropriate to the child’s developmental level and verbal capacities. Children may become confused or made more anxious by technical words or abstract or complicated questions. Closed questions often yield little information. For example, “Do you get along OK with your brother?” may bring forth a simple “yes” or “no,” whereas open-ended or descriptive questions are more likely to elicit a richer picture of how a child thinks and feels (e.g., “What sort of guy is your brother?” or “Tell me about your last play-date with Danny.”). Some usefully evocative questions include “What sorts of things make you mad (or make you feel afraid)?” “What is your saddest (or happiest) memory?” “What do you like to daydream about?” If it is important to establish the order or duration of events, school-aged children do better if the questions are anchored to markers in their lives (e.g., “Did that happen before or after your birthday?” or “Have you had trouble with that since school got out?”). Several empirical studies by Michael Rutter and colleagues have examined the effects of various interview styles on the information elicited from children and parents.

Although direct questions are often needed to initiate or to focus discussion, settling into a direct question format may constrict the interview process and give children the impression that the clinician just wants to hear facts or chronology and is not so interested in the child’s feelings and experiences. Young children frequently respond in a manner that they believe to be socially desirable, whereas older children are often reluctant to acknowledge feelings of sadness or vulnerability, even to themselves. Clinicians should convey, by demeanor and tone of voice, their interest in the child’s real feelings.

When discussion of feelings seems difficult for the child, it may be useful for the clinician to acknowledge the link between a situation and the child’s affective response (“Some kids might have felt mad if that happened to them; how about you?”). However, the clinician must avoid leading a child whom the clinician does not yet know well by assuming experiences or feelings. Probing a bland, equivocal response may prove useful; for example, in response to “Okay, I guess,” the clinician may say, “Sounds like you’re not quite sure.”

The assessment should convey respect for the child and should protect the child’s self-esteem. Many children feel ashamed or guilty about the difficulties that have brought them to the evaluation. When the clinician conveys a nonjudgmental understanding, using language that suggests that the difficulties are problems to be helped rather than transgressions, the child is more likely to be forthcoming. The child has come to the evaluation because of problems, but one should also talk with the child about strengths. For example, in inquiring about school, it may be useful to say, “Tell me about the parts of school you like best" before asking, "Are there subjects [classmates, teachers] that give you trouble?" or "What are the rough spots at school?" It may be better to ask the socially isolated child, "Do you sometimes have trouble making or keeping friends as much as you'd like?" and then "What do you think the trouble is?" rather than "Do you have any friends?" (especially if the truthful answer to this question is "no").

Structure and Sequence of the Child Interview. When interviewing the child, the clinician has in mind an overall map of the functional and historical areas to be covered; however, the format and sequence of any given interview are responsive to the individual child, the presenting problems, the unfolding pattern of the interview, and the clinician's personal style. For example, in every assessment, the clinician should ask about the cardinal symptoms of depression, anxiety, and disruptive behavior disorders, but when, how, and in what order this is done vary from one interview to another.

BEGINNING THE INTERVIEW. There are various ways to start the interview. The one chosen should be comfortable for clinician and child and should set the stage for a productive interchange. Many clinicians prefer not to address the presenting problem(s) immediately at the outset. Some clinicians, for example, remain quiet and let the child look about the room and begin playing or talking without prompting or guidance. This approach imposes a minimum amount of structure; the clinician can see how the child reacts to the materials present in the playroom and to the clinician. It also provides information about how the child handles the situational anxiety that is often evoked by the first meeting with the clinician. However, this unstructured approach may leave many children disoriented and uncomfortably uncertain as to the clinician's goals and intentions. Other clinicians are more direct in attempting to place the child at ease and begin by inviting the child to explore the available play materials ("What looks interesting?"). Others open the interview by inquiring about neutral or pleasurable topics ("Tell me about what you like to do for fun."). The clinician must read the child's cues regarding level of comfort and urgency in deciding when to address the problems that led to the evaluation. Some children prefer to first talk about nonthreatening topics, whereas others expect to get down to business and talk about their problems right away.

Although it may be useful to establish some initial rapport around neutral or interesting topics, most children understand that they have not come for a conventional social encounter. Hence, relatively early in the interview, it is useful to ask what the child's ideas are about the purpose of the meeting. It is usually best to do this in a matter-of-fact manner by asking, "What did your mom and dad tell you [or what do you know] about why they wanted you to come and see me today?" If the child has already met with the clinician together with the parents as part of an earlier contact, the clinician will have asked something similar at that time and may have heard about the child's view of the presenting problem and its history. If so, the clinician can follow up by saying, "Last time, mostly your mom and dad said what they thought about things. Today, we have a chance to talk more about how things look to you. Are there any things we talked about then that we should talk about today?" Asking what the child knows or thinks about the purpose of the evaluation shows the clinician's interest in the child's view of the situation and allows the clinician to hear how the child describes or assumes responsibility for the problem (or repudiates it as someone else's concern). Asking for the child's view up front, before the child has become deeply engaged or focused on some aspect of the interview process, allows the clinician to address misapprehensions. It also provides an opportunity for the clinician to explain, in terms that can be understood and potentially acknowledged by the child, the issues (at least as the clinician understands them) that have prompted the assessment.

ASSESSMENT OF MAJOR REALMS OF FUNCTIONING. Throughout the assessment, the clinician elicits and assimilates information to develop a picture of the child's functioning in the realms of family, school, peers, and recreation. Discussion of the presenting problem reveals information, particularly with respect to impairments, in these domains; however, additional discussion of the child's experiences in these areas enhances the picture of the child’s life and level of adaptation. It is often useful to begin exploration of each of these realms by asking for objective facts (e.g., “Who lives at home with you?” “Whom do you like to play with [or hang out with]?” “What school do you go to? What grade are you in?” or “What do you like to do for fun?”). This can be followed up with probes for more subjective views: “How do people in your family get along? Are there any problems at home? Things people fight about? Things you get in trouble for? Things people at home do that you don’t like?” This approach is intended to open up discussion of the major realms of functioning, not simply to elicit rote responses.

It is often helpful to begin by inquiring about the child’s recreational activities (“What do you like to do for fun?”). This allows the clinician to learn about the child’s interests, preoccupations, and talents and may provide a less threatening or problematic starting point than the areas of peer and academic functioning. Family functioning can be approached by asking who lives at home, how the child gets along with each person, and how family members get along with each other. One can also inquire more specifically, “What do you like to do with your dad [or other relative]?” or “What is your brother like?” In the case of nontraditional family structures or families of divorce or separation, the clinician may inquire about the history of the current arrangement and what sort of relationship is maintained with the noncustodial parent(s).

INQUIRY ABOUT PSYCHIATRIC SYMPTOMS. Just as a medical evaluation includes a review of key symptoms in the major systems, the child psychiatric evaluation should include a set of questions that screens for symptoms of common or major psychiatric disorders. The family or child may not see a link between one behavior or area of distress and another and may not volunteer symptoms that seem unrelated to the presenting problem. Comorbidity is relatively common in children with psychiatric disorders. Thus, the child psychiatric clinician must routinely question across the major areas of psychiatric disorders. Using developmentally appropriate language, this inquiry should cover depressive symptoms (e.g., low self-esteem, anhedonia, and suicidal ideation or behavior), self-harm (e.g., cutting), excessive anxiety or fears, obsessions and compulsions, hallucinations and delusions, and antisocial behaviors. For example, one should ask, “Do you feel sad or grumpy a lot of the time? Do you often feel like crying?” If the answer is positive, one can follow up concerning persistence (“Does that last for more than a few minutes?”), frequency (“Does that usually happen every week [or most days]?”), and context (“What sorts of things make you feel that way?”). To inquire about suicidal ideation, one can begin by asking, “Do you ever feel that life is so hard, you wish you weren’t alive [or would be better off dead]?” following up with the more explicit “Have you ever tried to hurt yourself [or done anything to try to make yourself dead]?” Exposure to potentially traumatic experiences should be routinely covered, and developmentally appropriate language should be used (e.g., “Have any really scary things ever happened to you?” “Has anyone ever touched you in a way that made you uncomfortable?”). The child should also be given an open-ended opportunity to talk about concerns that have not been specifically queried or anticipated by the interviewer. (“Are there any other worries that you have or things that make you unhappy that we haven’t talked about yet?”)

MENTAL STATUS EXAMINATION. Throughout the clinical interview, the clinician observes and notes information relevant to the domains of the mental status examination. Some abnormalities or deficits may be evident through this process; however, other areas require direct examination, which can usually be incorporated in a playful way into the clinical interview. Although intensive developmental, psychometric, language, and neurologic assessment usually cannot be subsumed within the clinical interview, a brief semi-structured evaluation requires only approximately 5 to 10 minutes and provides useful information. This generally involves drawing, copying a design, and engaging in simple tests of orientation, recall, general knowledge, and language (e.g., naming three objects; repeating a phrase such as “No ifs, ands, or buts”; a three-stage command, such as “Take a piece of paper in your right hand, fold the paper in half, put the paper on the floor”; a writing sample [such as asking the child to write a sentence of their choice]; and a brief screen of neurologic soft signs).

CLOSING THE INTERVIEW. In winding up the interview, the clinician can check for concerns that have not been covered and that also signal that the interview is coming to a close by asking, “We’ve talked about a lot of things; are there any other things that would be important for me to know about you?” Similarly, one might ask, “I’ve asked you a lot of questions; do you have any questions you want to ask me?” The clinician should also give the child a general idea of what will be done with all the information the child has shared and some sense of what happens next. For example, one might explain, “I’m going to think over what I’ve heard from you and your folks and meet again with them (and you) to talk about what might help with—. Are there any particular things I should be sure to mention?” If the clinician already has a sense of what interventions might be helpful, this may be tentatively broached with the child, keeping in mind the parents’ right to hear and decide on those recommendations.

Clinician and System Factors

In the past decade, there has been an increased awareness of the potentially distorting effects of implicit biases in the clinician as well as systemic cultural and racial issues that can impact the child and adolescent mental health treatment system. These range from implicit (or explicit) biases in the individual clinician and extend to cultural, socio-economic, and racial issues that may impact the family’s encounter with the clinic setting. It is important for clinicians to educate themselves and become aware of their own potential biases in order to be cognizant of how and when these biases might affect how they view their patients and their families. Systemic biases in how seemingly “objective” diagnostic criteria or psychometric norms are applied can influence the diagnoses assigned children from racial or ethnic minorities. The article “The Essence of Innocence: Consequences of Dehumanizing Black Children,” by Goff and colleagues (2014), identified the following example of implicit racial bias: black male youth are often assumed to be older than their chronologic age so that typical adolescent behaviors are often viewed as more menacing or potentially dangerous than they should. These types of biases, in conjunction with disparities in access to mental health services, contribute to the striking differences between groups in assigned treatments and outcomes.

SPECIAL ISSUES IN THE ASSESSMENT OF ADOLESCENTS

The clinical assessment of adolescents requires familiarity with the distinctive dynamics of adolescents and their characteristic patterns of managing conflicts, anxiety, and interactions with adults.

The adolescent's attitude toward the evaluation has important implications for how the assessment process unfolds. In contrast to the adult patient (who usually takes some responsibility and initiative in seeking help) and the child patient (who is simply brought by parents), many adolescents are cautiously ambiguous as to whether there is a problem with which they want help or whether they are merely passively (or reluctantly) complying with their parents' initiative.

One important initial decision involves who is to be seen in the initial interview: the parents, the adolescent, or all together. Seeing adolescents alone first has the advantage of emphasizing their active participation in the process and underlines the clinician's interest in their perspective; however, it may mean that crucial elements of the history are not available to the assessor. The clinician may opt to divide the initial interview into an allotment of time for the adolescent alone, the parents alone, and all of them together. Some adolescents may protest that they are being singled out as the identified patient or bearer of a problem that they feel resides in their parents or in the family as a whole. The clinician may need to explain explicitly to the entire family that the evaluation is not a judicial proceeding to assess blame but rather an attempt to understand what the problems are.

In some cases, when the adolescent balks at being seen at all, the clinician may need to meet with the parents alone to facilitate the youngster's participation in the evaluation. Regardless of when the adolescent is seen, the parents should, at some point, be seen alone to explain their concerns about the child, to obtain a full developmental and family history, and to hear about aspects of their spousal and parenting relationship that may be clinically significant but that may not be appropriate to discuss initially in the adolescent's presence.

Meeting with the adolescent and parents together is usually helpful. This provides an opportunity to explore how family members interact, their convergences or divergences of perspective, and their capacity to work together therapeutically. Such conjoint interviews usually require active structuring by the clinician, so that the purpose is framed as an occasion to talk together. Without such active intervention, the family session can easily deteriorate into an adversarial proceeding in which the parents complain about the adolescent, who, in turn, retreats into hostile or guarded silence.

Confidentiality

The assessment of adolescents often raises complex issues of confidentiality. Although the parents and the adolescent must know that the adolescent's communications are kept confidential, they must all understand from the onset that this does not extend to situations that pose a clear danger to the patient or others. For example, if the adolescent reveals significant suicidal ideation or a recent suicide attempt of which the parents are unaware, parents must be informed for the youngster's safety. Having made clear that the parents need to know, the task is then to negotiate whether the adolescent prefers to tell the parents directly or to have the clinician inform them. It is usually best to have the adolescent undertake this responsibility, but sometimes, after discussion with the youngster, the clinician must be the one to speak with the parents, preferably in the adolescent's presence. Less clear-cut situations in which the adolescent reveals problematic but not imminently dangerous antisocial or sexual behavior or substance use must be decided on an individual basis. Treatment may not be possible without some confidentiality; however, confidentiality must not become collusion.

Developmental Issues

Adolescents' developmentally appropriate push toward autonomy is often linked to a wariness about feeling vulnerable, dependent, or controlled. Tolerating feelings of ambivalence, guilt, or other painful affects or internal conflicts does not come easily to adolescents. In addition, adolescents are prone to externalize their concern or embarrassment over self-perceived shortcomings or difficulties. As a result, even adolescents who consciously desire help often approach a clinical assessment with the fear that revealing what they perceive as shameful weaknesses leaves them vulnerable to criticism, control, or regressive dependency. Fear of acknowledging vulnerability or dependency may result in bland denials of any problems or insistence that "I can handle it by myself" or "I don't need a shrink; I can talk about my problems with my friends." Externalizing their conflicted feelings about autonomy versus dependence, adolescents may fluctuate between grievances about parental overprotectiveness and complaints about perceived lack of care. Adolescents who adamantly reject the notion that they are experiencing difficulties or need help may be masking anxiety, shame, or other painful feelings of vulnerability. The skilled clinician learns to look beyond surface behavior and to notice when an adolescent "doth protest too much."

Although many adolescents seem intolerant of self-reflection, an unrealistic belief in the omnipotence of thought often helps to support denial and avoidance. Even when confronted with a pattern of long-standing maladaptive behaviors, the adolescent may insist, "I can stop any time I really want to," or "I don't do that anymore" (i.e., not since Saturday).

The lability of adolescents' moods and their truncated time perspective often make it useful to extend the evaluation over time to distinguish the transient issues from those that are more enduring.

Technical Issues in the Adolescent Interview

Interviewing adolescents requires tact, flexibility, a sense of humor, and clinicians who have come to terms with their own adolescent issues. Condescension, sarcasm, competitiveness, or excessive passivity on the part of the interviewer is likely to result in failure. However, although conveying a genuine interest in the youngster is essential, seductiveness, overfamiliarity, or the failure to maintain an adult role is likely to frighten the adolescent. The interviewer should resist the temptation to be too “with it” or too “cool”—nothing is more dated than last year’s slang. Because adolescents can so easily feel defensive, criticized, or embarrassed, the interviewer must learn to talk frankly yet tactfully about areas of apparent difficulty. Many embattled adolescents see adults as either for them or against them. Hence, the clinician needs to convey a genuine empathic interest in the adolescent’s view of a situation without implying collusive, uncritical acceptance of it or entering into an adversarial dispute.

The general observation that it is important to learn about youngsters’ strengths and interests, as well as their problems and vulnerabilities, is especially true of adolescents. Without appearing squeamish or avoidant of the problem areas, the clinician should inquire about the adolescent’s interests, hobbies, recreations, and friendships. Knowledge about the areas of accomplishment, talent, and gratification is required for a full picture that includes the youngster’s adaptive and compensatory strengths and assets. Furthermore, such an inquiry helps to facilitate rapport and conveys an interest in the adolescent as a whole person rather than simply as a bearer of problems. A too-exclusive focus on pathology puts many youngsters on the defensive, and an opportunity to talk about more neutral areas or those of greater mastery or confidence helps the adolescent to feel less on the spot. Instead of the interviewer straining to demonstrate familiarity with the latest band, cool video game, or winning team, letting adolescents teach the interviewer about the details of their interests lets youngsters enjoy a sense of mastery, control, and a degree of parity with the adult interviewer that helps to put them at ease. Indeed, a frequently useful technique in the adolescent interview is what Hilde Bruch called *the constructive use of ignorance* (and others have termed “the Colombo” technique after the eponymous detective’s seemingly guileless interrogatory style). This consists in being able to ask about some missing piece in the story or about some puzzling impasse with an ingenuous curiosity that invites the adolescent’s self-reflection (rather than cross-examining with confrontational or prosecutorial zeal).

Although interviewing and taking the history of the adolescent cover the same general areas outlined previously, certain domains are of particular salience with the adolescent. These relate to the core developmental tasks of adolescence: developing increased autonomy and individuation from parents; taking over the tasks of physical self-care; developing a satisfying, realistic body image in the face of pubertal changes; mastering burgeoning sexual and aggressive impulses; consolidating a social identity, including ethical values and vocational goals; delineating a gender identity; developing a network of satisfying peer relations; and in later adolescence, developing a capacity for romantic intimacy with an appropriate partner (Tables 35.1–1 and 35.1–2).

Table 35.1–1.
Developmental Growth Tasks of Adolescence

1. Gradual development as an independent individual
2. Mental evolution of a satisfying, realistic body image
3. Harnessing of appropriate control and expression of sexual drives
4. Expansion of relationships outside of the home
5. Implementation of a realistic plan to achieve social and economic stability
6. Transition from concrete to abstract conceptualization
7. Integration of a value system applicable to life events

Reprinted with permission of Elsevier from Gross RT, Duke PM. Adolescence. In: Levine MD, Carey WB, Crocker AC, Gross RT, eds. *Developmental Behavioral Pediatrics*. WB Saunders; 1983.

In the realm of values, ideals, and aspirations, the clinician should explore the adolescent’s values and the models whom the adolescent wishes to emulate or to reject. To what extent do these values complement or conflict with those of the parents? What expectations and aspirations (realistic or not) does the adolescent have concerning the future?

The world of peers and friends should also be explored with adolescents. With whom do they hang out? What crowd do they see themselves part of? How do they get along with peers, and what do they do for fun? Asking teenagers to describe a close friend provides a chance to learn how they think about people and relationships and to assess their capacity for empathy.

It is also important to inquire about the adolescent's use of the Internet and social media. How much time does the youngster spend online or on a mobile device? What is the balance of in-person (*IRL* or "in real life") interactions with local flesh-and-blood acquaintances versus purely electronic ones? Has the youngster had any upsetting experiences online, such as cyber-bullying, sexting, or sexual harassment?

The topic of friends leads naturally to the topic of romantic and sexual interests and relationships and gender identity. Exploration of this area requires tact, and the clinician cannot always be sure of receiving a candid response. Nonetheless, as with questions concerning depression or substance use, the clinician's tone can convey the sense that the inquiry into these sensitive areas is motivated by a genuine and legitimate wish to understand the adolescent. One can broach the topic of gender identity by asking "what pronouns do you like to go by" or "how do you like to be called?" If indicated, this can lead to a discussion of whether youngsters view themselves as male, female, or something else. As for sexual orientation, one can begin by asking whether the teenager dates (or "goes out with") anyone and whether there is anyone of either sex with whom the youngster is close. If so, who are they, and what attracted the adolescent to them? Are the adolescent's interests reciprocated? How has the relationship gone? Does the teenager note any recurrent patterns in terms of the type of person to whom they are attracted or how such relationships have unfolded in the past? Have any of the teenager's relationships developed into sexual ones? Has the adolescent had other sexual experiences? (Here one must be open to hearing about possible episodes of sexual abuse or non-consensual sex).

Table 35.1–2.
Growth Tasks of Adolescence by Developmental Phase

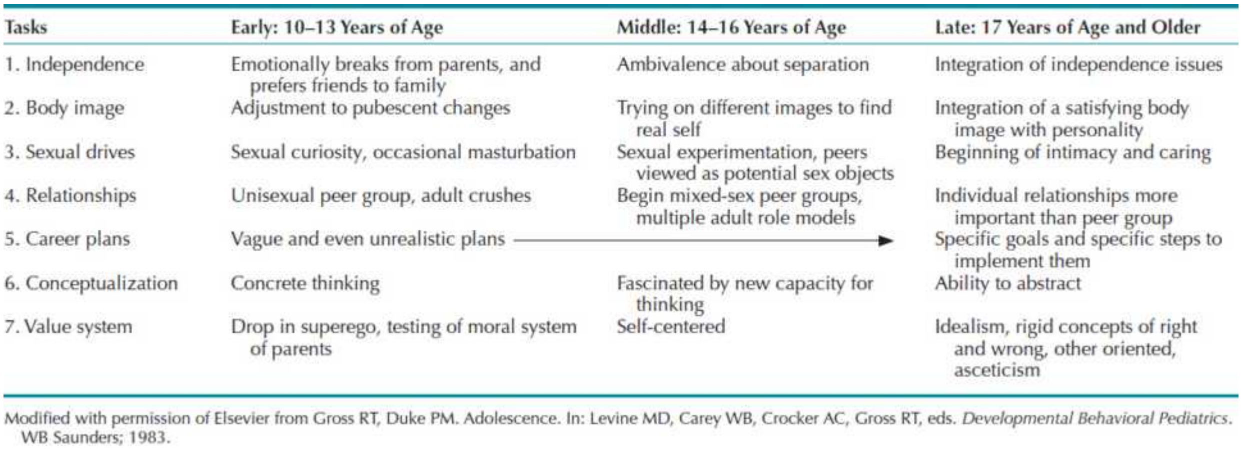

Tasks	Early: 10–13 Years of Age	Middle: 14–16 Years of Age	Late: 17 Years of Age and Older
1. Independence	Emotionally breaks from parents, and prefers friends to family	Ambivalence about separation	Integration of independence issues
2. Body image	Adjustment to pubescent changes	Trying on different images to find real self	Integration of a satisfying body image with personality
3. Sexual drives	Sexual curiosity, occasional masturbation	Sexual experimentation, peers viewed as potential sex objects	Beginning of intimacy and caring
4. Relationships	Unisexual peer group, adult crushes	Begin mixed-sex peer groups, multiple adult role models	Individual relationships more important than peer group
5. Career plans	Vague and even unrealistic plans →		Specific goals and specific steps to implement them
6. Conceptualization	Concrete thinking	Fascinated by new capacity for thinking	Ability to abstract
7. Value system	Drop in superego, testing of moral system of parents	Self-centered	Idealism, rigid concepts of right and wrong, other oriented, asceticism

Modified with permission of Elsevier from Gross RT, Duke PM. Adolescence. In: Levine MD, Carey WB, Crocker AC, Gross RT, eds. *Developmental Behavioral Pediatrics*. WB Saunders; 1983.

Because of the high prevalence of depression, self-harm, suicidal ideation, and even suicide attempts in the general population, specific inquiry about these areas as outlined previously is particularly important. The clinician should also inquire matter-of-factly about whether the adolescent smokes, drinks, or uses drugs illicitly, and, if so, with what frequency, in what amounts, in what contexts, and with what consequences.

ASSESSMENT OF INFANTS AND TODDLERS

The assessment of infants and toddlers poses special challenges for the clinician. First, the evaluation of young children requires specific and detailed familiarity with the cognitive, linguistic, motor, and adaptive developmental landmarks from birth to 6 years of age, as well as the specialized techniques needed to assess these competencies. Second, although direct observation of the child is an important source of data, the preverbal or limited verbal capacities of the child severely restrict the historical material that can be obtained from the direct interview of the child. Third (and perhaps most important), infants and young children cannot be assessed, diagnosed, or treated apart from their caregiving environment, which includes not only the parents but also other members of the extended family and other care providers, such as paid and unpaid childcare providers.

The developmental tasks of infancy include achieving physiologic homeostasis; developing stability of state and learning to negotiate transitions between states; organizing basic physiologic rhythms of wakefulness and feeding; coping with novel stimuli; perceiving coherent patterns of experience; mastering simple sensorimotor schema; establishing a specific and secure relationship with a caretaker; maintaining a positive affective tone; developing expressive and receptive channels of communication (including vocalizations and facial and gestural cues); and forming schemas of causality, intentionality, object permanence, and affect. The older infant or toddler must grapple with regulating and socializing bodily functions (e.g., eating, toileting, and sleeping), learning to concentrate and regulate shared attention, acquiring the capacity for imaginative play and more complex forms of imitation, dealing with the challenge of socialization (including development of more complex forms of social perception, communication, and empathy and negotiating conflicts of will with caregivers and peers), and regulating negative and positive mood and aggression. Development can proceed at a different pace in each area and may go awry or be delayed in any or all of them.

The goal of the developmental assessment of infants and toddlers is to depict their functional capacities in each of these areas, the relationships among the various domains of their development, their ability to adapt, and the range of coping strategies that they use in doing so, as well as the caregiving environment as it affects them, the quality of caregiving interactions, and the effect of their developmental status on the caregiver. As with older children, the purpose of the assessment is to define what, if any, interventions are necessary, while at the same time helping to form an alliance with the parents that facilitates implementation of any recommendations that may come out of the assessment. The assessment process itself, properly done, increases parents' sensitivity, awareness, and reflectiveness about their child's development, and in this respect the assessment of an infant or toddler may be an effective parenting intervention, as the parents observe their own child and learn more about his or her capacities and challenges.

The most common referral problems for infants and toddlers involve the following:

1. Developmental delays (e.g., specific delays, such as delayed, absent, or deviant motor or speech development, or general delays in multiple areas of development).
2. Psychophysiological dysregulation (fussy behavior, disturbances of sleep or feeding, rumination, eczema, self-stimulatory behaviors [e.g., head banging, rocking], and failure to thrive).
3. Behavioral disturbances (excessive tantrums, excessive negativism, aggression, and hyperactivity).
4. Disturbed social development (lack of apparent awareness of or interest in others, response to abuse, neglect, or multiple placements or repeated or prolonged separations).
5. Medical or genetic problems that may compromise a child's development (e.g., prematurity and genetic syndromes), exposure to toxins (such as lead or fetal alcohol syndrome).
6. Exposure to environmental stressors (such as maternal depression, change of caretaker, domestic violence or other trauma, reaction to transitions, e.g., beginning preschool or out-of-home childcare, parental marital separation).

The source of these referrals are often pediatricians, clinical social workers, child protection workers, parents and other adult family caregivers, and increasingly extra-familial childcare providers who may have concerns about the child's functioning within a group congregate care setting. The referral also may be prompted by consideration of early intervention services provided by a state agency through Part C of Individuals with Disabilities Education Act (IDEA), the Program for Infants and Toddlers with Disabilities.

Because all aspects of early development are interdependent, the significance of difficulties in one functional domain cannot be understood without reference to the child's developmental status in other areas. For example, speech, language, and hearing difficulties profoundly affect personal and social development. Assessment of motor or adaptive competency that involves interaction with the caregiver or assessor is affected by disturbances in relatedness and hence yields findings that appear erratic or uneven. Infants and toddlers are particularly sensitive to environmental stresses, which may delay the appearance of phase-appropriate skills or may lead to transient regressions.

Technical Issues in the Development of the Infant and Young Child

The assessment of the young child usually requires several brief sessions. In addition to sessions spent with the parents alone to gather history and to present the results of the assessment, familiar adults must be present during the assessment of the child; this is necessary to place the child at ease in the assessor's presence and to permit direct observation of child and caregiver interaction. Furthermore, the behavior of infants and young children is highly state dependent (and sensitive to fatigue or illness), and one usually needs to see the child on more than one occasion to have a sufficient sample of characteristic behavior.

The general principles of history taking from the parents and assessment of the caregiving environment follow those outlined previously. However, particular emphasis is given to the parental perception of the pregnancy with the child—whether the conception was planned or came at a good time in the life of the family and what the effect of the child's arrival was on the parents individually, on older siblings, and on the family as a whole. The history should explore parental (and grandparental) expectations for the child, the meaning that the child has for them, whom the infant reminds them of, what they perceive as the child's best and most challenging traits, any trauma or stressful disruptions of caregiver relationships, and what a typical day in the life of this child is like.

The developmental history for the young child includes prenatal, perinatal, and postnatal complications, including maternal illness, medication and substance use, and neonatal status. In addition to the areas outlined previously, assessment of the child's behavioral organization should include the regulation of physiologic state, alertness, and activity patterns. Children's ability to console themselves and to adapt to new situations and their response to separations are also key areas of evaluation. The history should also include the history of child care arrangements (nannies, relatives, out-of-home child care) and the child's reactions to new care environments or changes in these arrangements. (Many young children spend about as many waking hours in the care of a paid caregiver as they do with their parent(s), hence, in many cases, it may be very useful to hear the report of these extra-familial caretakers who often have a much larger store of observed normative child development to which this child's development may be compared and contrasted.)

The child's mastery of various gross motor, fine motor, language, adaptive, and personal-social milestones should be systematically assessed by history and in the direct evaluation of the child. In taking such a history, the assessor may be aided by a variety of structured or semi-structured caregiver report instruments, such as the Vineland Adaptive Behavior Scales, Third Edition (Vineland-3). It is useful to inquire about the timing of the child's mastery of these milestones relative to those of any siblings; this also helps to gauge parental expectations and to learn more about sibling dynamics. The Vineland-3 provides a semi-structured interview format that can be easily woven into a developmental history.

Within the first few months of life, children begin to manifest distinctive and often enduring behavioral styles or temperament. Table 35.1–3 shows the nine temperamental categories and three distinctive constellations of temperament as defined by Stella Chess and Alexander Thomas. The goodness of fit, or degree of consonance between the child's temperament and the parents' expectations or demands of the child, is an important predictor of future development and parent–child interactions.

Table 35.1–3.
Temperament Categories

Activity level: The motor component present in a given child’s functioning and the diurnal proportion of active and inactive periods.

Rhythmicity (regularity): The predictability or unpredictability in time of any function; it can be analyzed in relation to the sleep–wake cycle, hunger, feeding pattern, or elimination schedule.

Approach or withdrawal: The nature of the initial response to a new stimulus, be it a new food, a new toy, or a new person. Approach responses are positive, whether displayed by mood expression (e.g., smiling and verbalizations) or motor activity (e.g., swallowing a new food, reaching for a new toy, and active play). Withdrawal reactions are negative, whether displayed by mood expression (e.g., crying, fussing, grimacing, and verbalizations) or motor activity (e.g., moving away, spitting new food out, and pushing new toy away).

Adaptability: Responses to new or altered situations. One is not concerned with the nature of the initial responses, but with the ease with which they are modified in a desired direction.

Threshold of responsiveness: The intensity level of stimulation needed to evoke a discernible response, irrespective of the specific form of the response or the sensory modality affected. The behaviors used concern reactions to sensory stimuli, environmental objects, and social contacts.

Intensity of reaction: The energy level of response, irrespective of its quality or direction.

Quality of mood: The amount of pleasant, joyful, and friendly behavior (as contrasted with unpleasant, crying, and unfriendly behavior).

Distractibility: The effectiveness of extraneous environmental stimuli in interfering with or altering the direction of the ongoing behavior.

Attention span and persistence: *Attention span* concerns the length of time a particular activity is pursued by the child. *Persistence* refers to the continuation of an activity in the face of obstacles to the maintenance of the activity direction.

Qualitative analysis, supplemented by factor analyses, led to the formulation of, in addition, three constellations of temperament made up of various combinations of the individual categories, which appeared to have functional significance. These are as follows:

Easy temperament: The combination of biologic regularity, approach tendencies to the new, quick adaptability to change, and a predominantly positive mood of mild or moderate intensity. This group comprised approximately 40% of the study population.

Difficult temperament: The opposite of easy temperament, namely, biologic irregularity, withdrawal tendencies to the new, slow adaptability to change, and frequent negative emotional expressions of high intensity. This group comprised approximately 10% of the study population.

Slow-to-warm-up temperament: Withdrawal tendencies to the new, slow adaptability to change, and frequent negative emotional reactions of low intensity. Such individuals are often labeled shy. This group comprised approximately 15% of the study population.

Adapted with permission from Chess S, Thomas A. Temperament and its clinical applications. In: Lewis M, ed. Child and Adolescent Psychiatry: *A Comprehensive Textbook*. 3rd ed. Lippincott Williams & Wilkins; 2002:221.

Particularly salient aspects of the parental interview include a detailed history of the parents' childhoods and history of care, out-of-home placement, separations, and experience in their families of origin. The effect of the child on each parent and their marital relationship should be assessed, as well as the appropriateness and degree of congruity in their respective expectations of the child.

Adequate caregiving for the young child requires a delicate balance among gratification, comfort, support, and frustration that is appropriate to the child's developmental phase. It is important to hear (and to observe) which aspects of parenting and which features of the child's behavior the parent feels most comfortable with or most challenged by. The clinician should assess the parents' overall level of psychological functioning; the parental difficulties that adversely affect a young child's development do not always involve such gross failures as serious neglect or abuse but may involve problems more difficult to identify, such as persistent maternal or paternal depression. To help obtain a concrete picture of the parent, child, and their interaction, it is useful to ask for a detailed description of a typical day in the life of the child. In addition to providing data regarding the child's capacities, responses, and temporal organization, such a description provides clues to the parents' attunement and affective responses to the various elements of the child's behavior.

Although much of the clinical observation is focused on the child's behavior and capacities, direct observation of the parent–child relationship is essential. Inviting the family to interact in the playroom as they usually would at home allows observation of their interaction in unstructured parent–child or family play; assigning various structured tasks may also be useful. Observing the child in interaction with each parent separately helps to reveal relationship-specific patterns of behavior. Among the dimensions of parental behavior to be assessed are the ability to engage the child (verbally and nonverbally); the quality and affective tone of this engagement; the parent's attunement to the child's cues; parental vigilance, protectiveness, and limit setting; the parent's facilitation or restriction of autonomous play by the child; thematic content; and the degree of mutual pleasure, joint attention, and social referencing during play. Instruments and methodologies exist for formally assessing parental responsivity and parent–young child interaction, such as the Home Observation for Measurement of the Environment (HOME) and the Parent–Child Early Relational Assessment (PCERA). Some of these use standardized protocols or structured video recording and analysis. Some yield impressionistic global scores, whereas others provide detailed frame-by-frame coding of parent–child interactions. Although these methods are used primarily for research purposes, familiarity with the various facets of such instruments is clinically useful.

Observation and Examination of the Infant and Toddler

The direct examination of the young child draws on three levels of observation: The child's response to structured assessment items during formal testing, the child’s response to other interactions with the assessor and the assessment environment, and the spontaneous interactions of the child and parent in both clinical and naturalistic settings. When observation of the child and parents takes place in the clinic, it is essential to remember that the unfamiliar setting may elicit behavior that is not characteristic of the home setting; for this reason, it is important for the clinician to continuously inquire of the accompanying parent or caregiver as to whether the behaviors being observed are typical for the child.

Structured Assessment Instruments. A variety of instruments exists for the structured assessment of infants and young children, and each has somewhat different goals, theoretical orientation, and psychometric properties (Table 35.1–4). These instruments do not yield diagnoses; they detail the child’s development in various areas relative to a normative population. Some of these developmental tests are screeners designed to help pediatricians or trained paraprofessionals identify children with significant motor, social, or language delays who need fuller evaluation, whereas others are complete batteries evaluating a full range of developmental competencies. For example, the Bayley Scales of Infant and Toddler Development, Fourth Edition (Bayley-4), which is administered by a trained assessor, can be used to assess children 16 days to 42 months of age; it includes scales yielding standard scores for cognitive abilities, gross and fine motor skills, receptive and expressive communication, social-emotional, and adaptive (self-help) competencies, as well as a behavior rating scale for assessing qualitative aspects of the child’s behavior during the assessment. The assessment now includes special guidance for telepractice.

Although tests such as the Bayley-4 may show good reliability and concurrent validity, their ability, when administered to children younger than 3 years, to predict later performance on intelligence quotient (IQ) assessments or later adaptive functioning is weak. Among the reasons for this weakness of prediction are sources of error, such as the rapid pace of development during this period and the child’s susceptibility to state or environmental effects (e.g., fatigue and unfamiliar test setting), as well as the intervening effects of social and family environment and the heavy emphasis infant tests place on perceptual and motor skills that may have relatively little to do with information-processing capacities. However, very low scores on repeated measures are highly suggestive of continuing difficulties and warrant periodic follow-up assessment within the context of early intervention services to track development and to evaluate and refine intervention goals, intensity, and effectiveness.

The mental status examination of the infant and young child can be organized using a schema such as that shown in Table 35.1–5.

Diagnostic Formulation of the Assessment of Infants and Young Children

Although intellectual disability and autism spectrum disorders can be at least tentatively diagnosed in infants and toddlers, the most widely used official diagnostic system—the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) and the eleventh revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-11)—generally lack a developmental perspective and diagnostic criteria or categories suitable for infants and young children. Zero to Three's *DC:0-5 Diagnostic Classification of Mental Health and Development Disorder of Infancy and Early Childhood, Version 2.0* (DC:0-5 Version 2.0) provides a diagnostic schema for this age group that complements the DSM and ICD.

The appropriate outcome of a thorough clinical assessment of the young child thus extends far beyond an attempt at a specific categorical diagnosis. The diagnostic formulation should provide a detailed picture of the strengths and challenges in the child's social, emotional, cognitive, linguistic, motor, and perceptual development; the quality of the child's temperament, attachment, and regulatory capacities; hypotheses regarding the etiology of the child's difficulties; and potential exacerbating or risk factors and protective or ameliorative factors in the child and family. This formulation, which should be shared with the family in a comprehensible and usable form, should provide the basis for recommendations regarding intervention or further assessment.

SPECIAL TYPES OF ASSESSMENT

This section describes the comprehensive clinical assessment of the child. Although the clinician must be aware of the many facets of history taking, interview, and observation described here, certain clinical settings or tasks may call for a narrower focus of inquiry or distinctive emphasis for the evaluation. For example, an urgent situation may require therapeutic intervention before a comprehensive assessment can be completed; in such cases, the immediate evaluation focuses on defining the crisis and selecting the most suitable immediate intervention. Thus, the emergency room or crisis center evaluation focuses on precipitants and extent of the current decompensation, predisposing vulnerabilities, the exacerbating factors affecting the child and family and the ameliorative resources available to the family in such situations, the risk of danger to self or others, and other issues germane to making an immediate disposition (including assessing the potential need for hospitalization). The assessment of suicidal ideation, threats, or attempts or aggressive outbursts is perhaps the most common clinical context for such urgent evaluations. Children exposed to acute trauma (e.g., natural catastrophe or violence) are often screened outside of the traditional clinic setting in a combination of emergency assessment and crisis intervention. The noncompliant, acutely agitated, or delirious child in the pediatric setting raises yet another set of assessment issues (Table 35.1–6).

Various specialized consultative purposes also require distinctive techniques or procedures or specific areas of inquiry. (The American Academy of Child and Adolescent Psychiatry has prepared valuable practice parameters providing guidance for assessment and treatment of children and adolescence in forensic settings, detention or correctional facilities, schools, community systems of care, and pediatric settings, as well as youngsters presenting with certain specific conditions, such as substance abuse, suicidal behaviors, aggression, exposure to trauma, etc.)

For example, forensic evaluations raise special issues of role definition, confidentiality, and privilege and require familiarity with the legal issues involved. Clinicians must be clear about whether they are serving as neutral impartial advocates for the best interests of the child at the behest of the court or as consultants to one party or another. Although the clinician treating or assessing the child for clinical purposes may have information valuable to a forensic evaluation, it is usually in the best interests of the therapy and the legal proceedings for the evaluator to be someone other than the therapist.

The assessment of children or adolescents who may have been physically or sexually abused is a common form of forensic evaluation. In addition to assessing the child's competency and credibility, the examiner may also be asked for recommendations regarding whether the child should testify in court (and, if so, how) and whether removal of the child or alleged perpetrator from the home is needed, as well as any therapeutic interventions that may be required. A forensic evaluation requires careful documentation of parent and child interviews. The suggestibility of younger children raises the danger of false accusations and false denials and requires meticulous safeguards against repetitive, suggestive questioning.

Table 35.1–4.
Selected Formal Tests of Infant and Toddler Development

	Age	Domains[a]	Norm Sample	Reliability/Validity	Comments
Neonatal					
Assessment Tests					
Graham/Rosenblith Behavioral Examination of Neonates		M, TA, V, A, MT	Not applicable	Encouraging early findings	Currently seldom used
Brazelton Neonatal Behavioral Assessment Scale—2	37–44 weeks gestational age	NI, B, PS	Not applicable	High interrater reliability, weak test-retest, poor predictor beyond first year	Most widely used neonatal test
Infant/Toddler					
Development Tests					
Bayley Scales of Infant Development-4	16 days–42 months	Cg, GM/FM, EC/RC, Ad, SE, B	Large, representative	Excellent for mental, adequate for psychomotor, varied for behavior	Most widely used and validated infant test
Mullen Scales of Early Learning	0–68 months	GM, FM, VR, RC, EC	Large, representative	Adequate to exceptional	Mostly useful for toddlers and preschoolers
Battelle Developmental Inventory—3	0–95 months	Comm, SE, Ad, M, Cg (and sub-domains)	Small, representative	Numerous psychometric concerns cast significant doubts (see text)	Used considerably, despite many test flaws
Griffiths Abilities of Babies	0–24 months	Lm, PS, HS, EH, P	Small, all from London	Mixed results	Mostly used in Europe
Infant–Toddler Developmental Assessment—2	0–36 months	GM, FM, RIO, Lang/Comm, Ad, RP, SE, Cop	Not applicable	Excellent reliability	Multi-informant, criterion referenced
Screening Tests					
Battelle Developmental Inventory—3 Screening Test	6–95 months	Comm, SE, Ad, M, Cg	Small, representative	Strong correlation with full BDI	Suffers same limitations as full BDI
Bayley Scales of Infant and Toddlers Development Screening Test—4	16 days–42 months	Cg, Lang, M	Large; representative	Excellent reliability; promising validity	
Child Development Inventory	15–72 months	GM, FM, EC, RC, Ad, PS, L, N	Small; from St. Paul, MN	Little evidence of reliability and validity	300 parent report items
Developmental Activities Screening Inventory—II	0–60 months	15 sensory and problem-solving areas	>200 disabled children	Little evidence of reliability; valid for severely delayed	Can be used with language- and visually impaired children
Developmental Indicators for the Assessment of Learning—4	30–71 months	M, Lang, Cg	Large; representative	Acceptable reliability and validity	One of the better screeners for age level
Diagnostic Inventory for Screening Children—4	0–60 months	FM, GM, RC, EC, AM, VM, Ad, PS	Small; all from southwest Ontario	Excellent reliability; limited validity	
Developmental Observation Checklist System	0–72 months	Cg, EC/RC, PS, FM/GM, Adj, PS&S	Adequate; representative	Sound reliability and concurrent validity	Parent report only
Developmental Profile—4	0–21 years, 11 months	M, Ad, PS, AS, EC/RC	Large	Adequate reliability and validity	Parent report, teacher checklist, clinician rating
Early Screening Profile	24–71 months	Cg, EC/RC, M, Ad/PS, Ar	Adequate; representative	Good reliability; exceptional validity	One of better screeners for age level
Kent Infant Development Scales	0–12 months	Cg, M, EC/RC, Ad, PS	All from Northeast Ohio	Adequate reliability and validity	Primarily parent report

[a] A, auditory responsiveness; Ad, adaptive, self-help or daily living; AM, auditory attention and memory; Ar, articulation; AS, academic or pre-academic skills; B, behavior; Cg, cognitive or problem-solving; Cop, coping; EC, expressive communication/language; EH, eye-hand coordination; FM, fine motor; GM, gross motor; HS, hearing and speech; L, letters; Lm, locomotor; M, motor or physical; MT, muscle tone; N, numbers; NI, neurologic intactness; P, performance; PS, personal-social; PS & S, parental stress and support; RC, receptive communication/language; RIO, relationship to inanimate objects; RP, relationship to persons; SE, social-emotional; TA, tactile-adaptive; V, visual responsiveness; VM, visual attention and memory; VR, visual reception. A slash mark (/) indicates that multiple domains are assessed in the same scale or subtest.

Adapted with permission from Gilliam WS, Mayes LC, Tobon AL. Clinical assessment of infants and toddlers. In: Martin A, Bloch MH, Volkmar FR, eds. *Lewis's Child and Adolescent Psychiatry: A Comprehensive Textbook*. 5th ed. Wolters Kluwer; 2018:291–299.

Table 35.1–5.
Infant and Toddler Mental Status Examination[a]

I. Appearance

Size, level of nourishment, dress and hygiene, apparent maturity compared to age, dysmorphic features (e.g., facies, eye and ear shape and placement, epicanthal folds, and digits), abnormal head size, cutaneous lesions.

II. Apparent reaction to situation

Note where and with whom the evaluation takes place.

- A. Initial reaction to setting and to strangers: Explores; freezes; cries; hides face; acts curious, excited, apathetic, or anxious (describe).
- B. Adaptation
  1. Exploration: When and how child begins exploring faces, toys, stranger.
  2. Reaction to transitions: From unstructured to structured activity; when examiner begins to play with infant; cleaning up; leaving.

III. Self-regulation

- A. State regulation: An infant's state of consciousness ranges from deep sleep through alert stages to intense crying. Predominant state and range of states observed during session, patterns of transition (e.g., smooth vs. abrupt), capacity for being soothed and self-soothing, capacity for quiet alert state. (Some of these categories also apply to toddlers.)
- B. Sensory regulation: Reaction to sounds, sights, smells, light and firm touch; hyperresponsiveness or hyporesponsiveness (if observed) and type of response, including apathy, withdrawal, avoidance, fearfulness, excitability, aggression, or marked behavioral change; excessive seeking of particular sensory input.
- C. Unusual behaviors; mouthing after 1 year of age; head banging; smelling objects; spinning; twirling; hand flapping; finger flicking; rocking; toe walking; staring at lights or spinning objects; repetitive, perseverative, or bizarre verbalizations or behaviors with objects or people; hair pulling; ruminating; or breath holding.

D. Activity level: Overall level and variability (note that toddlers are often incorrectly called hyperactive). Describe behavior; for example, squirming constantly in parent’s arms, sitting quietly on floor or in infant seat, constantly on the go, climbing on desk and cabinets, exploring the room, pausing to play with each of six to eight toys.

E. Attention span: Capacity to maintain attentiveness to an activity or interaction, longest and average length of sustained attention to a given toy or activity, distractibility. Infants: Visual fixing and following at 1 months of age, tracking at 2–3 months of age, attention to own hands or feet and faces, duration of exploration of object with hands or mouth.

F. Frustration tolerance: Ability to persist in a difficult task, despite failure; capacity to delay reaction if easily frustrated—e.g., aggression, crying, tantrums, withdrawal, and avoidance.

G. Aggression: Modes of expression, degree of control of or preoccupation with aggression, appropriate assertiveness.

IV. Motor

Muscle tone and strength, mobility in different positions, unusual motor pattern (e.g., tics and seizure activity), intactness of cranial nerves (e.g., movement of face, mouth, tongue, and eyes, including feeding, swallowing, and gaze [note excessive drooling]).

A. Gross motor coordination. Infants: Pushing up, head control, rolling, sitting, standing. Toddlers: Walking, running, jumping, climbing, hopping, kicking, throwing, and catching a ball. (It is useful to have something for the child to climb on, such as a chair.)

B. Fine motor coordination. Infants: Grasping and releasing, transferring from hand to hand, using pincer grasp, banging, throwing. Toddlers: Using pincer grasp, stacking, scribbling, cutting. Fine motor and visual-motor coordination can be screened by observing how the child handles puzzles, shape boxes, a ball and hammer toy, small cars, and toys with connecting parts.

V. Speech and language

A. Vocalization and speech production: Quality, rate, rhythm, intonation, articulation, volume.

B. Receptive language: Comprehension of others' speech as seen in verbal or behavioral response (e.g., follows commands), points in response to "where is" questions, understands prepositions and pronouns (include estimate of hearing, especially in child with language delay, e.g., response to loud sounds and voice; ability to localize sound).

C. Expressive language: Level of complexity (e.g., vocalization, jargon, number of single words, short phrases, and full sentences); overgeneralization (e.g., uses *kitty* to refer to all animals); pronoun use, including reversal; echolalia, immediate or delayed; unusual or bizarre verbalizations. Preverbal children: Communicative intent (e.g., vocalizations, babbling, imitation, and gestures, such as head shaking and pointing), caregiver's ability to understand infant's communication, child's effectiveness in communication.

VI. Thought

The usual categories for thought disorder almost never apply to young children. Primary process thinking, as evidenced in verbalizations or play, is expected in this age group. The line between fantasy and reality is often blurred. Bizarre ideation, perseveration, apparent loose associations, and the persistence of pronoun reversals, jargon, and echolalia in an older toddler or preschooler may be noted in a variety of psychiatric disorders, including pervasive developmental disorders.

A. Specific fears: Feared object, worry about being lost or separated from parent.

B. Dreams and nightmares: Content is sometimes obtainable in children 2–3 years of age. Child does not always perceive it as a dream; for example, "A monster came in the front door."

C. Dissociative state: Sudden episodes of withdrawal and inattention, eyes glazed, "tuned out," failure to track ongoing social interaction. Dissociative state may be difficult to differentiate from an absence seizure, depression, autism, or deafness. The context may be helpful (e.g., a child with a history of neglect freezes in a dissociative state as mother leaves the room). Neurologic or audiologic evaluation may be warranted.

D. Hallucinations: Extremely rare, except in the context of a toxic or medical disorder, then usually visual or tactile.

VII. Affect and mood

The assessment of mood and affect may be more difficult in young children because of limited language, lack of vocabulary for emotions, and use of withdrawal in response to a variety of emotions, from shyness and boredom to anxiety and depression.

A. Modes of expression: Facial, verbal, body tone and positioning.

B. Range of expressed emotions: Affect, especially in parent–child relationship.

C. Responsiveness: To situation, content of discussion, play, and interpersonal engagement.

D. Duration of emotional state: Need history or multiple observations.

E. Intensity of expressed emotions: Affect, especially in parent–child relationship.

VIII. Play

Play is a primary mode of information gathering for all sections of the Infant and Toddler Mental Status Examination. In young children, play is especially useful in the evaluation of the child’s cognitive and symbolic functioning, relatedness, and expression of affect. Themes of play are helpful in assessing older toddlers. The management and expression of aggression are assessed in play as in other areas of behavior. Play may be with toys or with the child’s own or another’s body (e.g., peek-a-boo, roughhousing), verbal (e.g., sound imitation games between mother and infant), or interactional or solitary. It is important to note how the child’s play varies with different familiar caregivers and with parents versus the examiner.

A. Structure of play (ages are approximate).
   1. Sensorimotor play
      a. 0–12 months of age: Mouthing, banging, dropping and throwing toys or other objects.
      b. 6–12 months of age: Exploring characteristics of objects (e.g., moving parts, poking, and pulling).
   2. Functional play
      a. 12–18 months of age: Child's use of objects shows understanding and exploration of their use or function (e.g., pushes car, touches comb to hair, and puts telephone to ear).
   3. Early symbolic play
      a. 18 months of age and older: Child pretends with increasing complexity; pretends with own body to eat or to sleep; child pretends with objects or other people (e.g., "feeds" mother); child uses one object to represent another (e.g., a block becomes a car); child pretends a sequence of activities (e.g., cooking and eating).
   4. Complex symbolic play
      a. 30 months of age and older: Child plans and acts out dramatic play sequences, uses imaginary objects. Later, child incorporates others into play with assigned roles.
   5. Imitation, turn taking, and problem solving as part of play.

B. Content of play. The toddler's choice and use of toys often reflect emotional themes. It is desirable to have on hand toys that tap different developmental and emotional domains. An overfull playroom may be overwhelming or overstimulating and may reduce meaningful observations. Young toddlers of both sexes often gravitate to dolls, dishes, animals, and moving toys, such as cars. The examiner's choice of specific materials may facilitate the expression of pertinent emotional themes. For example, a child traumatized by a dog bite may more likely re-enact the trauma if a dog and doll figures are available. The child's reaction to scary toys, such as sharks, dinosaurs, or guns, should be noted, especially if they are avoided or dominate the session. Does aggressive pretend play become real and physically hurtful? By 2.5–3 years of age, a child's animal or doll play can reveal important themes about family life, including reactions to separation, parent–child and sibling relationships, experiences at day care, quality of nurturance and discipline, and physical or sexual abuse. The examiner must use caution in interpreting play, viewing it as a possible combination of re-enactment, fears, and fantasy.

IX. Cognition

Using information from all previously mentioned areas, especially play, verbal and symbolic functioning, and problem solving, roughly assess the child's cognitive level in terms of developmental intactness, delays, or precocity.

X. Relatedness

A. To parents: How in tune do the child and parent seem? Does the child make and maintain eye, verbal, or physical contact? Is there active avoidance by child? Note the infant's level of comfort and relaxation being held, fed, molding into caregiver's body. Does the toddler move away from the caregiver and check back or bring toys to show, to put into their lap, to play with together or near the caregiver? Comment on physical or verbal affection, hostility, reaction to separation and reunion, and use of transitional objects (blanket, toy, caregiver's possession). Describe differences in relating if more than one caregiver is present.

B. To examiner: Young children normally show some hesitancy to engage with a stranger, especially after 6–8 months of age. Appropriate wariness in young children may result in a period of watching the examiner, staying physically close to a familiar caregiver before engaging, or showing some constriction of affect, vocalization, or play. After initial wariness, does the child relate? Does the child engage too soon or not at all? How does relatedness with a stranger compare to that with a parent? Is the child friendly versus indiscriminately attention seeking or guarded versus overanxious? Can the examiner engage the child in play or structured activities to a degree not seen with caregiver? Does the child show pleasure in successes if the examiner shows approval?

C. Attachment behaviors: Observe for showing affection, seeking comfort, asking for and accepting help, cooperating, exploring, controlling behavior, and responding to reunion. Describe age-related disturbances in these normative behaviors. Disturbances often are seen in abused and neglected children; for example, fearfulness, clinginess, overcompliance, hypervigilance, impulsive overactivity, and defiance; restricted or hyperactive and distractible exploratory behavior; and restricted or indiscriminate affection and comfort seeking.

---

[a] By Anne L. Benham, M.D.

Reprinted from Thomas JM, Benham AL, Gean M, et al. Practice parameters for the psychiatric assessment of infants and toddlers (0–36 months). American Academy of Child and Adolescent Psychiatry. *J Am Acad Child Adolesc Psychiatry.* 1997;36(10 Suppl):21S–36S. doi: 10.1097/00004583-199710001-00003.

Child custody evaluations also require careful definition of the assessor’s role with all concerned. Specific attention must be given to the best interests of the child in terms of continuity and stability of care, the quality of parent–child attachment and attunement, the child’s stated preferences, the mental health of each parent, and the level of parental conflict and its effect on the child. Consultations for the juvenile court regarding youthful offenders or children in need of services and probate or commitment hearings represent other forms of forensic evaluation. Other specialized forms of evaluation include medication consultations and consultation to day care centers, schools, or pediatric specialty clinics (e.g., pain or oncology clinics).

STANDARDIZED ASSESSMENT INSTRUMENTS FOR CHILDREN AND ADOLESCENTS

While, as discussed earlier, psychosocial and environmental factors must inform clinical decision-making, the systematic assessment of symptoms, behavior and impairment is an essential part of clinical evaluation. Numerous tools—diagnostic interviews and rating scales—have been developed that elicit comprehensive information about psychiatric symptoms and diagnoses. The use of such tools provides some protection against the risks to which traditional clinical approaches may be prone, for example, tendencies to make diagnoses too quickly, collecting information selectively that confirms initial impressions, ignoring contradictory information, and applying diagnostic criteria idiosyncratically. The relatively structured diagnostic measures are likely to yield more reliable diagnoses than those made from less systematic approaches, and these features in turn improve the likelihood of obtaining meaningful findings from research efforts, which depend on assessing diagnoses in a consistent manner across a large number of study participants. However, although structured interviews and scales can be useful for certain purposes, as discussed below, for clinical purposes they cannot take the place of an individualized clinical psychiatric interview by an experienced clinician that takes into account individual, family, and situational contexts that are crucial for assessment and treatment planning.

Table 35.1–6.
Differential Psychosocial Diagnosis of the Uncooperative Patient

Inadequate understanding of the nature of illness or injury, treatment details, or staff expectations and intentions
Anxiety and regression
Separation anxiety
Concerns over bodily integrity
Response to immobilization
Denial
Idiosyncratic sources of anxiety
Inadequate pain control
Depression
Exacerbation of pre-existing psychopathology, especially oppositional or conduct difficulties
Family issues
Parental anxiety or hostility toward staff
Parental psychopathology: Depression, psychosis, or factitious disorder by proxy
Prior family experiences with illness
Cultural issues
Developmental issues (especially adolescence)
Emphasis on autonomy, ambivalence regarding dependency
Emphasis on activity, peer involvement
Self-consciousness about body
The body as a means of expressing psychological issues
Mature decision to refuse treatment

Reprinted from King RA, Lewis M. The difficult child. *Child Adolesc Psychiatr Clin N Am.* 1994;3(3):531–541. doi: 10.1016/S1056-4993(18)30485-1.

The advent of many of these tools can be traced to two related developments in psychiatric clinical thinking and research since 1980. First, the introduction of explicit, criterion-based diagnostic systems beginning with the Research Diagnostic Criteria and the DSM-III, followed by the DSM-III-R, the DSM-IV, and the DSM-5, as well as the research version of the ICD-10 facilitated the development of standardized interviews and questionnaires, which specify the relevant elements of a diagnosis. Second, at the same time, there has been a rapid expansion in psychiatric research, creating a need for tools for research assessment. Thus, working synergistically, major changes in the diagnostic system, the impetus provided by emerging research findings, and the increasingly apparent advantages of using standardized diagnostic tools in research (and, subsequently, in clinical settings) are key developments that have presented the field with an array of psychiatric assessment techniques and thus with a corresponding growing body of research literature on their characteristics and utility.

Because the inventory of symptomatic and diagnostic tools for children and adolescents has rapidly expanded over recent years, it is important for those evaluating and treating youth to become familiar with the literature and with the clinical and research relevance of these tools.

Issues in Developing Diagnostic Methods

The task of developing diagnostic tools for psychiatric and mental health assessment is a complex and challenging one, given the complicated nature of the symptoms and their clusters and patterns over time. In addition, developing measures to assess psychopathology in children and adolescents is more challenging than the comparable task with adults. These challenges include the need to incorporate an understanding of aspects of development into the framing and wording of questions about symptoms and into the process of discriminating normative from pathologic behaviors. Many behaviors only become “symptoms” when they occur at too high a frequency or at the wrong age (e.g., separation problems and oppositional behaviors), or cause functional impairment. In addition, many symptoms reflect an underlying substrate of psychopathology with phenotypic manifestations that differ during different developmental stages. Moffit referred to these variations as a function of the “heterotypic continuity” of disorders. Severe irritability provides an illustrative example: what is difficult temperament in early months and years may progress to the characteristics of disruptive mood dysregulation disorder in middle childhood and adolescence, and eventually become manifest as adult anxiety and depression.

Another challenge in the development of structured interviews or scales is the necessity of obtaining information from both the child and another informant, usually the parent and at times the teacher, paralleling what is done in good clinical practice, because developmental stage can limit the capacity of children to report on their own behaviors/symptoms. Not surprisingly, investigators have repeatedly found poor agreement between different informants with respect to symptom presence and hence diagnosis; yet, as has long been realized by clinicians, there is general agreement that information obtained from any informant is diagnostically useful. There is no consensus, however, in using these tools, on how best to combine the information across informants or as to whether to combine the information at all. Some kinds of information may be uniquely accessible to one or another informant—for example, teachers may be better reporters of attention problems than parents, whereas parents are likely to be better reporters of eating behaviors and irritable mood. For alcohol or other substance use, youth consistently report more symptoms than parents do, and similar results have been found for sexual behavior, suicidality and self-injurious behaviors, and conduct disorder.

Characteristics of informants may influence how they report on symptoms and behavior. Parents who have anxiety disorders tend to report greater levels of psychiatric symptoms in their children than do independent observers. Mothers who are depressed tend to over-report behavior problems and symptoms in their children, even when contextual variables and child gender have been controlled for. Complicating this issue, children of depressed mothers have been found to have more behavioral problems than children of nondepressed mothers. At younger ages, children often do not have the ability to compare themselves to others, which can affect their reports on oppositional behavior or hyperactivity.

Further challenges to the development of diagnostic measures include the problem of how to integrate concepts of social context and adaptive functioning into a symptom-based assessment. These factors are difficult to isolate conceptually and to operationalize without producing an overly long and complex measure.

In the history of diagnostic methodologic development, the psychometric properties of reliability and validity have been touted as the basis for determining what measure was best. It is important for users to understand these concepts and to be familiar with a proposed instrument's performance in these respects. Briefly, the *reliability* of an assessment method is often conceptualized as the extent to which the results of an assessment can be repeated or reproduced, whereas *validity* is conceptualized as the extent to which the instrument measures what it purports to measure.

There are various types of reliability. *Inter-rater reliability* involves the stability of an assessment when it is administered by different raters (in the case of psychiatric diagnostic interviews, by different interviewers). *Test–retest reliability* reflects the stability of results across administrations, that is, when an instrument is administered on two occasions, separated by a period of time in which the degree of real change in the condition being measured is expected to be trivial, and seeing if the two administrations agree with one another. To assess the test–retest reliability of child psychiatric diagnostic instruments, the measure is often administered twice, 1 to 2 weeks apart. The 1-to-2-week interval is chosen to minimize precise memory effects (i.e., the child or parent is not likely to remember, and thus repeat, exact responses to a great number of the questions previously asked). The relatively long duration of many child psychiatric conditions makes it likely that changes between the first and second administrations reflect unstable aspects of the measure rather than real changes in the disorder. For instruments which use interviewers, assessment of test–retest reliability is particularly complicated as a different interviewer typically administers the "retest" interview. This is necessary for two reasons. First, taking the same assessment from the same interviewer a second time may lead the respondent to leave out symptom reports, assuming that the interviewer already knows about the respondent's experience with an area of symptomatology or, conversely that the phenomena should be reported differently because the interviewer is going over the same issues a second time. Second, the interviewers themselves may recall the first interview and this knowledge might lead them (consciously or unconsciously) to conduct the second interview differently than they would with a respondent they did not already know. Thus, test–retest reliability for diagnostic interviews will necessarily be somewhat limited by inter-rater reliability. In addition, discrepancies between the initial administration and the subsequent interview usually result from the informant reporting fewer symptoms at the second administration—the "attenuation effect." This has been found in all forms of repeated psychiatric inquiry and likely represents a systematic bias in reporting. *Internal consistency* is a third type of reliability and refers to the extent to which items designed to assess aspects of a known disorder (e.g., depression) are measuring the same underlying construct. In omnibus measures, the internal consistency of different diagnoses or domains may be assessed separately because these instruments by design assess a wide range of behaviors and symptoms rather than a single construct. Thus, the component items for each diagnosis would not be expected to correlate with the items from other diagnoses.

The *validity* of a measure is more difficult to determine than its reliability. There are various ways to conceptualize and assess validity. To assess what is termed *construct validity,* the psychiatric diagnoses or scores generated by a particular instrument are usually compared to those generated by an expert, or criterion, assessment. Statistical methods are then applied to quantify the degree of agreement about whether diagnoses are present or absent. (In the methodologic evaluation of the validity of an omnibus diagnostic instrument, agreement is determined diagnosis by diagnosis, so that it can be revised in areas of poor agreement with the criterion.) The essential challenge in determining the construct validity of a measure is to compare its performance to some standard that is known to be diagnostically correct. A central problem in the field of psychiatric assessment is the lack of a readily available "gold standard" with which to compare results because there is no laboratory test or other agreed upon standard to provide objective evidence for the presence of a disorder. Clinical assessments of child psychiatric diagnoses are notorious for varying from one clinician to another; thus, the so-called “standard” is a moving target. Indeed, structured interviews and scales were developed in large part to improve the poor psychometric properties of nonstandard assessments. This situation has led to an approach that considers an instrument’s performance within the nomologic net of theory and findings about the condition being assessed. Stated briefly, this means that, to the degree that an instrument’s findings converge with accumulated thinking and knowledge in the field, one may have relative confidence that, at least at this point, the instrument validly measures what it claims to assess. The major diagnostic assessments have shown considerable convergence in their evidence for reliability and validity; thus, in the complex process of instrument development, it is important to be aware when a new measure is falling below existing standards in these respects.

Despite—or perhaps because of—these various challenges, extensive research efforts have been directed toward refining psychiatric assessment methods in child and adolescent psychiatry. These endeavors have yielded a number of measures for use in research contexts; in addition, some of those described in the following discussion have also found appropriate adjunctive roles in clinical settings. The following discussion of methods first examines interview methods, which yield categorical diagnoses, and then reviews scales that cover symptom dimensions, often with respect to a specific form of psychopathology, for example, depressive symptoms.

Diagnostic Interviews

There are numerous “omnibus” diagnostic interviews for children and adolescents. In clinical settings, such interviews might be included as a supplementary part of the diagnostic evaluation to ensure that major symptoms domains are assessed and to guide treatment, triage or planning. Most of these diagnostic interviews are designed to be administered to the youth’s parent (about the child/adolescent) and, depending on the youth’s age, separately to the young person. Although the exact contents of the various diagnostic interviews differ, each covers a wide range of the more common psychiatric disorders of childhood and adolescence, although many do not include the autism spectrum disorder. Each version addresses the DSM diagnostic criteria (and often the ICD criteria as well), and updates for DSM-5 have either been prepared or are underway. (Of note, when moving from DSM-IV to DSM-5, with the exception of Autism, few changes were made to the DSM criteria for disorders that typically affect youth, so it is not necessary to restrict oneself to considering only DSM-5 measures.)

The omnibus diagnostic interviews differ from one another in several important respects, which should be taken into account when choosing among them. The first issue to consider is the *purpose* of the assessment (e.g., to identify a participant for research or to address a need for clinical treatment or triage? to establish diagnosis(es)? to identify co-morbidity? to measure severity?). Consideration of the purpose of the assessment also involves the time frame of interest (current only? past year? ever? current and ever?). Practical issues also inform plans for implementation, including: how long it takes to administer the interview; whether the same interviewer must interview both parent and youth; qualifications for administrators or interviewers (is it necessary to have a clinician administer the assessment?); the standard training procedures and supervision; scoring procedures (are there specific rules or algorithms? is a computer program necessary; if so, what are the available reports? is scoring based on clinical judgment? is an expert editor/reviewer necessary? how transparent is the scoring? what data will the interview yield—diagnoses only or symptoms and scales as well?); infrastructure considerations (necessary staff, computer/analytic capabilities; office space to conduct private interviews); and, finally, costs (including license fees, copying, storage).

Also, if the sample contains cultural subgroups, one must consider whether the style of the interview, especially language used in asking about symptoms and behavior, is suitable for the group(s) to be interviewed. In such instances, a pretest on a subsample with a post-administration focus group may suggest expansion of question(s) about particular symptoms or additional culturally more sensitive wording to cover the same symptoms. *In all instances*, it is strongly recommended that before choosing an instrument one obtain a copy of the interviews under consideration and critically review their contents to understand how symptoms and criteria are conceptualized by their authors and whether one agrees with the conceptualizations.

Respondent-Based versus Interviewer-Based Interviews. Perhaps the most fundamental distinction between the different diagnostic interviews is the degree of structure they impose. Terms often used to describe this are “highly structured” versus “semi-structured,” but a more useful way to think about this is a “respondent-based” versus an “interviewer-based” approach. The respondent-based interview (RBI) has been described by Shaffer and colleagues as a “precise script” to obtain clinical information from the informant. The RBI has precisely worded and ordered questions and equally structured response options; thus, the authors essentially transformed the diagnostic criteria into questions that should be easily understood. The RBI is totally dependent upon how clearly the questions are written and are understood by respondents. A primary question that is answered positively leads to a branching sequence of nested questions that specify symptom frequency, intensity, duration, and degree of associated impairment. These questions help determine the significance of the reported behavioral-emotional state and aim to determine whether it reaches a certain clinical threshold. Alternatively, an initial negative response allows “skip out” of these specific questions. The relative advantages of the RBI include the potential for methodologic development and also issues of economy and public health applicability. Because this type of interview is given in the same way each time, it is possible to study its performance and then modify it with a relative degree of confidence in the rationale for changes.

Another advantage is the potential for relatively economic data gathering and management. Because the structure of this type of interview drives its administration, a range of administrators (typically "lay" [clinically untrained] interviewers with little psychiatric/medical background) are used and some of these interviews can even be "self-administered" via computer where there is no interviewer and the informant either reads the questions themselves or the computer "asks" the questions. While it is more cost-effective to use self-administration with visual or auditory prompts or computer administration that bypasses the need for and the expense of interviewers, the limitations and potential risks of the RBIs must be considered. These include the risk of the child or youth finding the process impersonal and therefore not freely reporting symptoms, rushing through the evaluation by entering responses before thinking about what was is being asked (especially if the interview is lengthy), and not understanding questions, but still entering a response. However, the potential of RBIs as screening tools in community settings is another attractive aspect of this method.

A further cautionary word about RBI administration involves the use of nonclinically trained interviewers. While this is a real asset in terms of the available pool for hiring potential interviewers, there are important caveats. Nonclinically trained interviewers need to be coached in the importance of establishing appropriate rapport, often more difficult with children than parents. Good administration of an RBI involves more than just reading the questions as written. Interviewers need to be coached in how to establish an appropriately relaxed and welcoming attitude and rapport consistent with the sense that "we are here to do something important that means you will be asked a lot of questions about yourself. Do your best to answer them without worrying about the questions. If you need a stretch or short break to get a drink of water or use the restroom, just tell me…that's OK. Any questions? Good, let's get started!" If the interview process leads to the child/adolescent revealing something obviously painful, an understanding, empathic expression or sympathetic nod and slight pause may help the child/youth regroup. "I understand" may be appropriate, but the interviewer should not probe and should definitely report periods of distress (e.g., revelations of child abuse, breakdowns in composure) to the clinical supervisor. In addition, the interviewers must be trained in a protocol of what to do if clinically emergent information, such as suicidality, potential child abuse, etc., is reported.

In contrast to RBIs, interviewer-based interviews (IBIs) rely on the skills of the clinically trained interviewer to make sound judgments about the presence and clinical significance of reported phenomena. Thus, the IBI is totally dependent upon the interviewer's clinical skill and quality of training. IBIs contain some degree of structure, in that definitions are usually provided about what constitutes a symptom, but although questions are provided to prompt the clinician-administrator, it is expected that the interviewer will pursue inquiry until their clinical judgment is satisfied about whether a symptom is present. This allows for a degree of flexibility and responsiveness to the body of clinical data that are ordinarily used to determine the significance of reported emotions, moods, and behavior—for example, observation of affective state in response to a query or other observational data, such as eye contact, posture, and voice inflection.

Because the choice of interview type and then of the specific interview depend on the intended use, before selecting a measure, it is worthwhile to examine some of the extensive literature developed on these methodologies. Briefly, RBIs probably do not do well in diagnosing extreme and atypical phenomena (e.g., psychosis), because this method "assumes" that respondents (youth or parents) interpret the questions as intended by the interview developer and also that the respondent has the capacity to reflect and to self-observe about the presence or absence of the symptom. IBIs can be more flexible and responsive to the state of the respondent and hence could take less time to administer. For example, an astute clinician can usually assess manic symptomatology fairly quickly and efficiently in a previously high functioning youth. In practice, however, these interviews typically take as long or even longer as an RBI. In addition, when using an IBI the careful clinician should catch most instances of the respondent misunderstanding the intent of the question, whereas the RBIs have little capacity to self-correct for this inevitable source of error. IBIs, although more costly than RBIs to administer, are particularly appropriate for clinically based research in which subtle diagnostic distinctions may be critical for defining relatively diagnostically “pure” samples and in which sample sizes are small enough to justify the expense of hiring interviewers trained in making these judgments.

RBIs, however, as noted previously, have many flexible and innovative applications in research and public health contexts. They have also found recent application in an increasing number of clinical contexts, for example, as efficient aids in collecting comprehensive symptom and diagnostic information for outpatient clinic or even inpatient intake. Because they systematically review symptoms across all domains, this type of screen sometimes uncovers significant symptomatology, such as suicidal tendency, that was not queried or not revealed on initial clinical interview. But, while the RBIs are, to some extent, mechanical and lack the personal element of the skilled clinical interview, it has been shown that many adolescents, in particular, appreciate the privacy of the computer-based, self-report methods (a familiar modality for many youth).

RBIs are particularly useful for epidemiologic studies in which large sample sizes (i.e., in the thousands) require many interviewers either in person or, more recently, via telemeeting, thus making the use of trained clinicians prohibitively expensive. These studies investigate issues of the prevalence of disorders, developmental patterns of psychopathology, their distribution within psychosocial subgroups in the population, risk and protective factors, access and use of mental health services in relation to types of psychopathology, and other topics that require large samples to provide enough statistical power to examine the study hypotheses. The ultimate choice of a measure requires careful weighing of the goals that it is intended to serve, the resources available, and the pragmatics of the situation.

RBIs and IBIs have been developed with the goal of reducing error in the collection of information relevant to making diagnoses. Error in the counting and combining of symptom data to determine a diagnosis can be minimized with RBIs or IBIs by the use of specified algorithms, usually computer applied, to establish whether sufficient symptoms are present for the diagnosis and whether they meet DSM criteria for intensity, frequency, and duration. Ultimately, it remains an informed judgment call as to the method that best serves the ends of general appropriateness, including error reduction, logistics, and economy (relative to the intended goal). As opposed to period of the early 1980s following the advent of DSM-III when the development of structured and semi-structured instruments began, there now exists considerable theoretical and empirical information on which to base decisions about instrument choice.

Respondent-Based Interviews. The Diagnostic Interview Schedule for Children (DISC) is the most extensively developed and studied of the RBIs. The DISC was specifically designed for use in large epidemiologic and community studies where it could be administered by “lay” (clinically untrained) interviewers who would read questions verbatim to parent (concerning children ages 6 to 18) and to child (ages 9 to 18), with most structured so as to be answered “yes” or “no.” Since it appeared, the DISC has also been used in clinical research studies, screening projects, clinical treatment settings, and juvenile justice settings. Initial work on the DISC began in 1980 and since then it has gone through numerous updates to maintain alignment with the DSM system in its iterations, as well as with methodologic testing in both clinical and community settings. The recently completed DISC-5 is compatible with DSM-5, DSM-IV, and ICD-10 (research version), and, like the DISC-IV, assesses diagnoses over three different time frames—past 4 weeks, past year, and lifetime, although the accuracy of recall for the lifetime frame is questionable and hence those questions are optional. DISC diagnostic modules are self-contained, so that in a particular context the user might selectively administer only those modules of interest. In the DISC, a uniform question structure has been applied in that for each criterion being assessed, an initial stem question is asked, which if endorsed, is then followed by contingent questions that elicit increasingly more specific types of information about characteristics of the potential symptom. This structure allows probes to be shorter and contain few concepts, making them more easily understood by respondents. Given the complexity of information required for assigning DSM diagnoses, administration of the whole interview using paper protocols is not practical, so the DISC is given with the aid of a computer. Computer-assisted versions ensure fidelity to the intended structure of the questioning, as well as reducing the time needed for the interviewer to consider what the next question should be. In addition to symptom questions, the DISC includes questions in each diagnostic section that inquire about associated impairment and perceived severity of impairment. These questions are asked if a respondent meets a certain base level of symptomatology for a given diagnosis—about half of what would be required for a DSM diagnosis. For the DISC-IV, there was also a computerized version, the Voice-DISC, which can be self-administered; a comparable version is under development for DISC-5.

DISC diagnoses are derived by application of a set of computer algorithms. The interview has few symptom skips so that it is possible to generate symptom and DSM criterion scale scores, further increasing its flexibility and general usefulness, especially in research settings. Of importance, respondent acceptability has been demonstrated to be quite good, despite some (but not an overwhelming) perception that it is a bit lengthy. For the DISC-IV, a modified version, the Young Child DISC, is administered to parents only and is used to assess childhood symptomatology down to age 4. The Young Adult DISC-IV, with parallel self and “other informant” versions, was also available, and is being updated for DISC-5. A Spanish-language DISC-IV version was developed and extensively tested. The DISC-IV was translated into a number of other languages and found various innovative applications around the world. Particularly exciting applications are those in which it can assist in clinical contexts where trained mental health clinicians are in short supply (e.g., its use in juvenile justice facilities, orphanages, and regions of the world with few mental health practitioners per capita). Modules from the DISC-5 have also undergone translations into other languages and are currently in use in international studies. Of course, no RBI can substitute for thoughtful clinical practice, and the developers of the DISC have been clear that its use is solely adjunctive and facilitative, and not diagnostically definitive when applied in clinical settings (pwf1@columbia.edu [for information, copies, translations]; https://telesage.com/netdisc-5 [current computerized version]).

The Children's Interview for Psychiatric Syndromes (ChIPS) is a highly structured interview that was designed for use with lay interviewers to screen for current psychiatric disorders. The authors recommend that interviewers have at least a bachelor's degree in a field associated with mental health (e.g., psychology, social work, and psychiatric nursing). The first version of the ChIPS was specifically designed for interviewing children ages 6 to 12 years (unlike the DISC, which is not given to children under age 9); for later versions the age range was extended to include adolescents. Development of the parallel parent version (P-ChIPS) followed the child version. For each diagnosis, "cardinal" symptoms—those that are most frequently endorsed by youth who may have a particular disorder—are assessed first and, if not endorsed, the interviewer skips to the next diagnosis, making the interview briefer to administer than the DISC. The interview is scored by the interviewer using a scoring booklet that provides rules for scoring DSM diagnoses. The ChIPS has not been updated to DSM-5, but given the few changes in the diagnostic criteria, it should still be considered for use (https://www.appi.org/chips-childrens_interview_for_psychiatric_syndromes).

The Development and Well Being Assessment (DAWBA), which is widely used throughout Europe, although less so in the United States, assesses current diagnoses (the exact time frame varies by diagnosis) and is DSM-5 and ICD-10 compatible. The DAWBA was designed to provide a shorter alternative to typical diagnostic interviews by including many "skips" and screening questions. It also makes use of a companion screening questionnaire, the Strengths and Difficulties Questionnaire (SDQ). The DAWBA employs a two-part procedure beginning with an RBI comprising both closed-ended, precisely specified questions (answered "yes" or "no" or using a choice of two to three relevant options) as well as open-ended questions (to which the answers are typed verbatim into an open text field by the interviewer or, in a self-administered version, directly by the respondent). The open-ended questions obtain supplementary information about the symptoms and accompanying functional impairment. Thus, like the DISC, and the CHiPS, the DAWBA does not require much training to administer. There are parent and youth versions and a teacher questionnaire. When the interviews and questionnaires are completed, all information is forwarded to a clinical rater who has been trained by the authors of the DAWBA. This clinician reviews and assigns diagnoses based on the transcripts of the interviewer notes and answers to the structured questions from all available informants (usually parent and child, and sometimes the teacher). Preliminary scoring yields “likelihood” estimates, and provisional diagnosis is done by computer. The scoring rules used for the DAWBA have not been made available for review. The DAWBA and the SDQ have been translated into many languages (https://www.dawba.info/).

Kiddie Schedule for Affective Disorders and Schizophrenia—Computerized version (KSADS-COMP) is the newest version of the popular KSADS, the most widely used IBI (see below) and comprises three web-based versions—clinician administered (an IBI described below), and self-administered youth and parent versions (all RBIs). All KSADS-COMP versions follow the same structure as the most recent noncomputerized K-SADS, beginning with an introductory module, followed by a screening interview and then supplementary modules for diagnoses identified by the screening interview as being probable. KSADS-COMP youth and parent version contain the same structure, probes (one or two per symptom), response options, scoring, and branching options as the clinician version The KSADS-COMP parent and youth version differ significantly from all earlier versions of the KSADS, in which all probes were “suggested,” relying on the clinician to decide how to elicit information and to apply clinical judgment to decide when a symptom was present; in these earlier versions it was unclear after the fact how the particular probes included were selected Also, in the KSADS-COMP, a common frequency metric is used to rate presence of current (last two week) symptoms. (The parent version is text based, where the parent reads the probes, while the youth version includes an option to view videos of an interviewer reading the probes, instead of reading the probes themselves). Unfortunately, none of the three KSADS-COMP versions have been compared to the standard clinician-administered K-SADS, nor is such a comparison planned; results from ongoing research in community samples suggests that the RBI versions may overdiagnose some disorders (e.g., bipolar disorder, obsessive compulsive disorder) (https://www.ksadslogin.net/ksads-comp/aboutus.aspx).

The Flexible Interview for ICD-11 (FLII-11-Epidemiology) will soon be released from the World Health Organization. It was prepared by an international team and was specifically designed for administration by trained lay interviewers for use in international settings to assess the more common psychiatric disorders specified in the ICD-11. The impetus behind developing the FLII-11 was to make widely available a free instrument that could be used in large national studies that would find clinical utility in many settings. A downward extension of this interview for adolescents is being prepared and should be available in 2023.

Interviewer-Based Interviews. Various IBIs, developed to meet different research needs, are finding adjunctive use in clinical settings. These include the Schedule for Affective Disorders and Schizophrenia for Children (K-SADS) as mentioned above; the Anxiety Disorders Interview Schedule for Children (ADIS-C); the MINI International Neuropsychiatric Interview for Children and Adolescents (MINI-KID); the Child and Adolescent Psychiatric Assessment (CAPA); and a youth version of the Structured Clinical Interview for DSM-5 (Kid-SCID), a widely used adult measure, is in development, as is a youth version of the Structured Clinical Interview for ICD-11 (SCII-11). Finally, the Diagnostic Interview for Anxiety, Mood and Obsessive Compulsive and Related Neuropsychiatric Disorders—Child and Adolescent Version (DIAMOND-KID), which was initially developed for use in clinical settings, is finding use in research. As would be done in a typical clinical evaluation, most of these interviews begin with some sort of introductory module during which information about the child’s presenting problem(s) and a short clinical history are obtained, along with demographic information and information about current functioning which can be used to set the context for further inquiry. The length and content of this varies widely between instruments. With the exception of the SCII-11, which is keyed to ICD-11 diagnoses, all have been updated for DSM-5 and most can be used to assess ICD-11 diagnostic criteria. For each of the existing measures, a literature is available that defies summary here. Following are some brief comments about the currently available IBIs.

The K-SADS, originally developed for use in studies of child/adolescent depression was based on the SADS, an adult research instrument, and was the first widely used IBI for children and adolescents. Unlike the SADS, the K-SADS has kept pace with changes to the DSM. There are a number of different versions of the K-SADS, reflecting the modification of this measure by researchers at different sites with differing specific interests and assessment philosophies. Most assess current and lifetime diagnoses. Some versions incorporate severity scales for the criteria assessed, allowing their use in examining change, whereas others use more categorical scoring. With the exception of the KSADS-COMP parent and youth versions (which should be considered an IBI), all versions of the K-SADS contain probes that are “suggested,” but there are no questions that must be asked, because all versions of the K-SADS assume that the interviewer has been clinically trained and is experienced in assessing youth psychopathology. In the KSADS-COMP, clinician version, the clinician is also expected to use clinical judgment when asking or adapting questions and making ratings. Nonetheless, many research studies use trained research assistants to administer the KSADS interview, a practice which some have criticized. Following an overview (and, for the most recent KSADS-PL, a screening module), the interviewer proceeds through a series of modules to assess different diagnoses. For most of the K-SADS versions, the same person interviews both the parent and the child and, when necessary, works to resolve differences. The authors of the KSADS have recommended that for younger children the parent be interviewed first, whereas for adolescents, the recommendation is to interview the youth first; in practice, however, many have found that it makes more clinical sense to interview the parent first, regardless of the young person's age. Consensus ratings for each diagnostic criterion are made, based on the interviewer's clinical judgment using information from both informants and from any other source (e.g., clinical records). Diagnoses are then assigned by the interviewer using a series of score sheets; for the KSADS-COMP, the computer yields a diagnostic report using the clinician ratings. An early childhood version (K-SADS-EC) (parent informant only) has been prepared and used in studies of preschool depression (https://www.pediatricbipolar.pitt.edu/resources/instruments [for traditional KSADS]; https://www.ksadslogin.net/ksads-comp/aboutus.aspx [for KSADS-COMP]).

The ADIS-C, derived from the A-DIS, an adult interview, was originally developed for use in studies of childhood anxiety. The time frame for inquiry is current and lifetime; the interview provides extensive coverage of anxiety disorders and is particularly strong in this area, with less extensive coverage of other symptom areas. "Thermometers" are used to obtain severity, fear/avoidance, distress, and interference information and a clinician severity rating is made for each diagnosis (which makes possible tracking symptom change). The ADIS-C is designed to be administered by a clinician with experience in assessing child psychopathology. Like the K-SADS, the ADIS-C relies on a reconciliation procedure to sort out disagreements between parent and youth and the scoring is done by the clinician administering the interview, rather than by computer algorithm (https://global.oup.com/academic/product/anxiety-and-related-disorders-interview-schedule-for-dsm-5-child-and-parent-version-9780197621950?lang=en&cc=us).

The MINI-KID is a youth version of the MINI, a widely used research interview for adults, and was designed to quickly assess the presence of current and past psychiatric disorders of youth. Like the MINI, the MINI-KID incorporates a branching structure with each diagnostic section containing screening questions, which if not endorsed, allows the interviewer to skip the rest of the section. Unlike the administration procedure for all of the other diagnostic interviews for youth, parents and children are typically interviewed together, with question directed to the child and the parent asked to speak up if the information the child provides is not accurate. Discrepancies between informants are addressed by the clinician interviewer during the interview. For adolescents, sometimes only the adolescent is interviewed, depending on setting and user choice. There is a separate parent version available as well. The authors also have a “MINI Kid Screen” that comprises all the screening questions that can be used to screen for probable disorders prior to using the whole interview. Unlike the other IBSs in this section, the MINI-KID does not include an introductory module to set the stage for the rest of the diagnostic inquiry. Of all the IBIs, the MINI-KID takes the least time to administer (the authors estimate half an hour), due to the many skips and the efficiency of combining parent/youth interviews; however, some have questioned whether the screening questions are sensitive enough for some disorders. Diagnostic criteria are documented within each of the disorder sections and on a summary sheet by the clinician administering the interview (https://harmresearch.org/product/mini-international-neuropsychiatric-interview-mini-kid-kid-parent-version-7-0-2-10/).

The CAPA is a very comprehensive measure that, in addition to psychiatric symptomatology and impairment, covers family, environmental, and psychosocial problems, life events, the effect of the child’s problems on the family, and mental health service use. The time period assessed is last three months. Unlike the other interviews, the CAPA is not organized into obvious diagnostic modules. The CAPA contains a number of questions that should be asked exactly as written, with follow-up questions that are less specified. Unlike the K-SADS, the CAPA was specifically designed for administration by non-clinicians who undergo extensive training in how to administer the interview. This is accomplished by providing detailed definitions and rules that govern the assessment of reported phenomena and determination of whether they should be rated as symptoms. Thus, the onus is on the interviewer to assure that the questions are understood and the information provided is relevant. A clinically experienced supervisor reviews each CAPA interview before final ratings are made. Young Adult (YAPA) and Preschool Versions (PAPA) of the instrument are available. Like the DISC, the CAPA is scored according to well-specified computer algorithms. The CAPA has recently been computerized, easing administration and substantially reducing data entry costs. Training on the instrument typically takes 4 days (https://devepi.duhs.duke.edu/measures/the-child-and-adolescent-psychiatric-assessment-capa/).

The Diagnostic Interview for Anxiety, Mood and Obsessive Compulsive and Related Neuropsychiatric Disorders—Child and Adolescent Version (DIAMOND-KID), derived from an adult instrument (DIAMOND), was recently developed (post publication of the DSM-5) and is designed to address Anxiety, Obsessive-Compulsive, and related disorders more comprehensively than other currently available diagnostic interviews. The timeframe of inquiry is the past month, except when DSM-5 requires a longer timeframe (e.g., for bipolar disorder it is necessary to inquire about a lifetime history of mania). The DIAMOND-KID also includes some items that are rated based on observation of the youth or on the clinician's synthesis of all available information about the child. Like the MINI-KID, questions are directed to the child (if the parent is present) or the child is interviewed alone—there is no corresponding parent interview, although it is flexible enough that the clinician can decide how best to administer it. The DIAMOND-KID adheres closely to what a clinician would ask or probe in a comprehensive clinical evaluation. Handily, the interview includes information about differential diagnoses, common associated features, and specifiers for many diagnoses, which help to improve accuracy. This interview can be downloaded at no cost (https://giving.hartfordhospital.org/diamond/) and training costs are minimal.

Rating Scales

A large repertoire of questionnaires and rating scales have been developed to assess symptomatology and various aspects of behavior and adaptive functioning. Different measures have been designed for completion by parents, teachers, and/or children. Some of these questionnaires/scales include a broad range of symptoms (“omnibus” measures) while others focus on a more specific area (e.g., depression and anxiety). The development of each scale ideally involves testing its psychometric properties, that is, the functioning of the scale with respect to its reliability and validity as well as administering it to large and diverse samples to determine norms and clinically relevant cutoffs.

A number of specific aspects of scale functioning must be considered in selecting a scale in practice or research. Again, a basic issue is the nature of the intended use, that is, to aid in clinical assessment, treatment or planning or to address a research question or clinic/agency purpose (e.g., describe patients being treated or quality improvement). Intended use determines which domains of symptomatology and functioning must be covered and is tied to the nature of the sample or population with which it will be used. For example, the major scale requirements may vary for studies or clinical settings that assess clinical patients receiving an intervention designed to change symptom occurrence over time in contrast to studies estimating community population in which the prevalence of a disorder(s) would be relatively stable or else likely to change in response to an imposed condition (e.g., intervention) or the changing developmental stage of a cohort tracked over time. If one wants to implement a measure to track symptom change, which could be part of an effort to guide treatment and implement “measurement-based” care, it is also important to consider the timeframe for symptom assessment. In considering which scale to use for a particular purpose, the psychometric characteristics of various scales should be compared and factored into a decision, together with specific characteristics of the children or adolescents to be assessed. One obvious example is age. Given that developmental stage and not solely chronologic age determines the young person’s capacity to abstract at the level required by the particular scale’s questions, the appropriateness of such questions to the youngster’s activities and opportunities are all factors very important to accurate assessment. The young person’s cultural context also may bear on symptom reporting for various reasons, such as acceptability of reporting certain types of symptoms, culture-specific theories of etiology, or subjects' expectations of what is normative in behavioral or emotional expression. For self-completion questionnaires, literacy and reading level of parents and youth who will be completing the measure is also an important factor to consider.

Thus, although a particular scale may appear to provide a convenient tool for clinical or research use, many features of its demonstrated functioning, of the young people themselves, and of their social contexts must be taken into account when choosing a measure. Similarly, the critical reader will want to consider such factors in interpreting findings from the research literature. As for the diagnostic interviews, there are many issues to be considered when deciding which measures to use and, indeed whether to introduce measures into clinical practice at all. These include how, when and where to administer them, storage, and, for self-completion measures particularly, a procedure which assures that the clinician/clinical staff is provided with information obtained. For symptom measures, it is helpful if, in addition to a total score, the clinician is given the measure itself to see which items were endorsed to get a fuller picture of the youth. Finally, there is the issue of cost, as many of these measures must be purchased, which can be prohibitively expensive in low resource settings.

Comprehensive (Omnibus) Rating Scales. Probably the most widely used comprehensive symptom rating scales are those that constitute the Achenbach System of Empirically Based Assessment (ASEBA) family of instruments (http://www.aseba.org). There are forms for school-aged children—the Child Behavior Checklist for Ages 6 to 18 (CBCL/6 to 18), the Teacher Report Form (TRF/6 to 18), and the Youth Self Report (YSR/11 to 18); forms for younger children —the Child Behavior Checklist for Ages 1½–5 (CBCL/1½–5) and the Caretaker-Teacher Report Form (C-TRF); as well as forms for young and older adults to report on themselves. The ASEBA forms average 99 to 120 items and generate several syndrome scales. In addition, there are clinician-completed forms, including the Direct Observation

Form (DOF), which is completed after observing classroom behavior. Important features of this system of measures include its empirical development and the large population sample on which it was normed, as well as testing on large samples of clinically referred children. The availability of age-based and gender-based norms provides a convenient way to screen for overall levels of symptomatology, and when scored, the CBCL yields a profile of problem areas and competencies. The CBCL is stable over relatively brief intervals (i.e., good test–retest reliability) but also reflects changes resulting from treatment. The CBCL, TRF, and YSR have been widely used in epidemiologic and clinical research, in many clinical settings and in translated forms in numerous countries. These measures can be a useful clinical screening tool, as well as an efficient way for clinics to gather uniform information for characterizing the populations served as well as changes in those groups over time. Each of these measures includes DSM-oriented subscales for common youth diagnoses. Computerized data entry and scoring have greatly eased the use of the ASEBA measures and allow for comparisons of reports across informants (www.aseba.org).

The Behavior Assessment System for Children (BASC-3) is another widely used assessment system for evaluating aspects of behavior and personality in children and adolescents aged 2 to 21 years. It consists of parent and teacher rating forms, a Structured Developmental History form and a Student Observation System (for rating classroom behavior), and, for youngsters 8 to 21, a Self-Report of Personality (for thoughts and feelings). The parent and teacher measures have forms appropriate for ages 2 to 5 years, 6 to 11 years, and 12 to 21 years (the latter also has a youth form); each form has between 139 and 175 questions. A distinctive feature of the BASC is its emphasis on *positive* aspects of the child's adjustment, in addition to rating symptoms and problem behaviors. An additional feature of this system is its coverage of problems related to school functioning, such as difficulties in social or study skills, and it has built-in validity checks. Norms have been established for children by age, gender, and clinical status. This system has been designed for application in clinical, educational, research, and program evaluation contexts. Use and interpretation of the BASC are eased by the availability of computer software that generates profiles and allows consideration of different informants' viewpoints (https://www.pearsonassessments.com/store/).

The Conners Rating Scales are another family of instruments that are widely used in research and clinical settings. These include short and long forms of scales for parent administration CPRS-RS, CPRS-R-L, teacher administration (CTRS-R-S, CTRS-R-L), and self-report (CASS-S, CASS-L). These scales collect information on child oppositional behavior, cognitive difficulties, hyperactivity–impulsivity, anxiety and shyness, perfectionism, social problems, and psychosomatic complaints. The Conners scales have been developed and tested on samples of children and youth 3 to 17 years of age, for parent and teacher forms, and 12 to 17 years on the adolescent forms. These measures have shown excellent psychometric properties, and norms are available, as is computer software to aid in administration and scoring. The Conners scales have long been the standard instruments for the assessment of ADHD in children and adolescents, and their use for ADHD and Disruptive behaviors is described below (https://pearson.my.site.com/usclinical/s/article/Conners-3-Forms).

In contrast to the rather lengthy omnibus scales described above, the Strengths and Difficulties Questionnaire (SDQ) has been gaining in popularity in research, is very brief and is made freely available. The SDQ was designed to be used as a short screen to identify children (ages 2 to 17 years) with behavioral problems. A drawback from clinical use is that it does not cover any specific area in depth. The SDQ comprises 5 subscales (emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behaviors), for a total of 25 items: it has cross informant forms (parent, teacher, and self) and has been translated into over 75 languages Versions that include additional questions about "impact" and questions about given interventions are also available (www.sdqinfo.org).

DSM-5 also includes the "Parent/Guardian-Rated Level 1 Cross Cutting Symptom Measure—Child Age 6–17" which some clinical settings have found useful. This measure was created to help clinicians identify diagnostic domains that might not have been considered in a typical unstructured clinical assessment but may be important for a particular youth. The measure was found to be clinically useful and have good reliability in the DSM-5 field trials. A child rated version (ages 11 to 17) is also available (www.psychiatry.org/dsm5).

Other omnibus scales, which are available but have found less use in recent years include the Revised Behavior Problem Checklist (RBPC), the Louisville Behavior Checklist (LBC), and the Personality Inventory for Children (PIC). The RBPC focuses on symptoms of conduct disorder, socialized aggression, attention problems and immaturity, anxiety and withdrawal, with auxiliary scales for psychotic behavior and motor tension-excess. It was designed for use by parents and teachers of children and youth ages 5 to 17 years. Although generated using factor-analytic studies of clinical samples, it has also been normed on nonreferred children and youth. The LBC and the PIC are both comprehensive rating scales for parent report that do not have corresponding versions for other informants. The LBC consists of 19 scales that cover a wide range of behaviors and symptomatology. The PIC was developed on the model of the Minnesota Multiphasic Personality Inventory (MMPI) and also comprises a variety of component scales.

Disorder-Related (Domain Specific) Scales. Each of the scales described above was designed to collect information across a range of symptoms and competencies. More narrowly focused symptom rating scales have been developed to permit a more detailed quantitative assessment of specific symptom realms. These latter scales may be useful in quantifying the presenting severity of certain symptoms to establish a baseline against which response to a therapeutic intervention can be compared. A number of these scales are self-report measures. These have the advantage of providing an opportunity for the child to report on symptoms that might be difficult to conceptualize or verbalize. Sometimes parents or children (especially adolescents) are initially more willing to report an area of difficulty in the seemingly more anonymous context of a questionnaire than in a face-to-face interview. However, self-reports also carry the risk of the child misunderstanding questions or feeling compelled to provide socially desirable responses rather than acknowledge behaviors and feelings that the child believes might meet with disapproval. Fundamentally, the focused symptom approach may also ignore important interactions across domains of psychopathology, because of the focus on one particular area of symptomatology. With these potential limitations in mind, child self-report measures can provide useful adjunctive information in clinical and research settings.

SCALES FOR ASSESSING INTERNALIZING SYMPTOMATOLOGY—DEPRESSION. The Revised Child Anxiety and Depression Scale (RCADS, RCADS-25), is a self-completed measure suitable for screening and assessing *depression and anxiety* in youth ages 8 to 18, which has found wide use in research and more recently in clinical settings. The RCADS has good psychometric properties and was recently selected as a “core measure” for inclusion by the National Institute of Mental Health and the Wellcome foundation in their funded studies of youth. There are parallel youth and parent/caretaker forms. Each symptom item is rated on a four-point scale (“never” to ‘always”) and there are six subscales (major depression, generalized anxiety, social anxiety, separation anxiety, panic disorder, obsessive-compulsive disorder). The RCADS has been translated into many languages and is free to use (https://www.childfirst.ucla.edu/resources/).

The Patient Health Questionnaire (PHQ-9), another “core measure” selected by NIMH and Wellcome, includes nine items, corresponding to the nine DSM-5 criteria for Major Depression. A short two-item form is also available. Each item is rated on a frequency scale for the past 2 weeks. The PHQ-9 has found wide use for adults in clinical and research settings, with some use with adolescents, down to age 12. There is no corresponding parent/caretaker form. This free-of-charge measure can be obtained from many websites.

The Beck Depression Inventory (BDI) is a widely used scale for depressive symptoms in cognitive, behavioral, affective, and somatic domains. It has found wide use with adolescents. Because it is an "adult" measure, there is no corresponding parent version. The BDI was originally developed to measure the severity of depressive symptoms in patients who had been diagnosed with depression; it is also used to screen patients to identify depression and to monitor treatment outcomes. The BDI has been widely studied and used in clinical and nonclinical populations, with generally good psychometric properties. As with other scales, a fine-grained examination of the literature shows specific areas in which it falls short of ideal; for example, different cut-points are needed, depending on the population (inpatients vs. outpatients; male vs. female). Unlike the earlier version of the BDI, which did not provide complete coverage of the DSM depression criteria, the newest version, BDI-II, covers all the depression symptom criteria in the DSM system and extends the time frame of inquiry to 2 weeks, again to reflect the DSM duration requirements for Major Depressive Episode. Nevertheless, varying degrees of stability that are inherent to psychopathology in adolescence must also be kept in mind in the overall determination of the functioning of any scale. Generally speaking, positive results from various psychometric studies, ease of use, scoring and interpretation have made the BDI an attractive instrument in both clinical and research contexts (https://www.pearsonassessments.com/store/).

The Children's Depression Inventory (CDI) is a self-report measure for assessing symptoms of depression in children 7 to 17 years of age. It shares a conceptual basis with the BDI and comprises five factors: dysphoric mood, acting out, loss of personal and social interest, self-deprecation, and vegetative symptoms. As with many measures of child psychopathology, various psychometric properties vary by characteristics of the sample used, especially patient versus community status, gender, and age. As noted for the BDI, participant characteristics play important roles in interpreting psychometric results because, for example, the capacity to conceptualize and report on symptoms improves with age. Hence, scales reaching into the younger school-aged years are likely to show more variability in assessments of reliability and in examinations of agreement between child and adult informants. Levels of agreement with other measures related to depressive psychopathology and its known correlates have provided support for the validity of the CDI. It has been widely used across a range of populations; thus, as with the BDI, the availability of considerable information relevant to its performance with different groups of children and youth makes it attractive to those considering it for use in various research and clinical populations. Parent and teacher informant forms are available. Many useful results have emerged from studies using the CDI, further adding to its attractiveness as a measure of depressive symptomatology. One criticism of the scale is that adolescents may find it too “babyish” (https://storefront.mhs.com/collections/).

The Center for Epidemiologic Studies Depression Scale (CES-D), was developed first as a rating scale for depressive symptoms in epidemiologic studies of adults, but it has also been widely used with adolescents, and a modification (CES-DC) was developed specifically for use with younger individuals. Both are freely available from many websites. There is no corresponding parent measure. Psychometrically, the CES-DC functions sufficiently well in some domains for adolescents but, overall, does not hold up well for children. Both instruments are brief and easy to use, and, particularly in adolescents, some aspects of reliability and validity meet performance standards. However, for both measures, problems with discriminant validity pose limitations, suggesting that they may be most appropriately considered as screening tools for general psychopathology in adolescent populations. An alternative version of the scale, Center for Epidemiologic Studies Depression Scale Revised (CESD-R-20), adheres more closely to the DSM diagnostic criteria for Major depression, compared to the original scale, and includes questions about suicide ideation (not assessed in the original scale) as well as an additional frequency rating (whether present nearly every day for 2 weeks).

The Mood and Feelings Questionnaire (MFQ) is another freely available brief self-report measure designed to screen for and assess core depressive symptoms for children and adolescents. Originally designed for use in epidemiologic studies, the MFQ has also found use in clinical settings and screening programs. There are long (32-

item) and short (13-item) versions of the scale, and parallel parent and youth reports. Although these questionnaires have demonstrated good psychometric properties and discrimination for depression appears better than for other self-report measures, they are not as widely used as the other scales (https://psychiatry.duke.edu/research/research-programs-areas/assessment-intervention/developmental-epidemiology-instruments-0).

The Quick Inventory of Depressive Symptomatology (QIDS) and QIDS-A (a version for adolescents that includes an extra item assessing irritability) are short forms of the Inventory of Depressive Symptoms (IDS) and are confined to the symptoms of Major Depressive Episode. In addition to a self-completed questionnaire, there is a corresponding clinician rated version, as well as a structured interview guide. The QIDS has strong psychometric properties and is useful for tracking treatment response. These tools are free and available from many websites.

Two related scales, the Reynolds Adolescent Depression Scale (RADS, RADS-2) (for ages 11 to 20) and the Reynolds Child Depression Scale (RCDS, RCDS-2) (for ages 7 to 13), are widely used measures to assess a wide range of symptoms associated with depressive disorder. Symptoms are rated on a 4-point Likert-type response scale that yields scores on four subscales (dysphoric mood, anhedonia/negative affect, negative self-evaluation, and somatic complaints). Both the RADS and the RCDS function very well in terms of reliability (both internal consistency and test–retest). The validity of the RADS is supported by relationships with other measures of depressive symptomatology and related constructs. It has been used with diverse populations, providing a broad base of evidence for its functioning, as well as normative comparison data. The strong psychometrics, basis in the highly specified symptomatology of the DSM system, and the linkage of the two scales for use in longitudinal studies or across a broad age range in a single study are advantageous features of these measures. The RADS and RCDS are often used as screening measures to identify depressed youth (there is no corresponding parent measure), and there is some evidence that these measures may be sensitive to clinical change, so they hold promise for evaluating treatment outcomes (https://www.parinc.com/Products/).

The Hamilton Rating Scale for Depression (HRDS), another freely available measure, differs from the aforementioned measures in being a clinician-rated scale. It is the most commonly used clinician rating scale for depression. Developed for assessing depressive symptoms in adults, the HRDS has also been used in studies of adolescents. The HRDS was initially developed to measure the severity of symptoms in patients who had a depressive illness and was widely used in psychopharmacological and other treatment studies evaluating outcomes of different treatments for depression. The original scale had 21 items, of which only the first 17 were scored; there are alternative versions with 23 items and also versions with fewer items. Not all of the versions have been psychometrically evaluated. Although internal consistency and inter-rater reliabilities have generally yielded acceptable to excellent results, specific training in its administration and the use of a structured or semi-structured protocol for administering the HDRS improves the psychometric properties. The validity of the HRSD has not been as convincingly demonstrated in adolescent populations as for adults. The main application has been in treatment outcome studies, in which findings have supported its sensitivity to the effects of certain psychopharmacologic agents, as well as to other specific therapeutic modalities. The HRDS is free for download from many websites.

The Children's Depression Rating Scale-Revised (CDRS, CDRS-R) is a clinician-rated symptom inventory related to the HRDS but designed for the assessment of school-aged children. It uses clinician expertise by including their ratings of behavior observed during the interview and by having clinicians generate scores based on information from each parent and child considered alone, as well as a combined score that integrates information from both informants. Psychometrically, aspects of reliability have ranged from adequate to excellent, depending on the study and type of reliability considered. There is some support for validity, and the CDRS has found substantial use in clinical research generally and treatment outcome studies specifically. Overall, the unique features of the CDRS, clinician rating, and use of multiple informants have made it attractive for studies conducted in settings in which the use of clinicians is feasible and their skills in eliciting and integrating information are advantageous for the study design (https://www.wpspublish.com/cdrs-r-childrens-depression-rating-scale-revised).

SCALES FOR ASSESSING INTERNALIZING SYMPTOMATOLOGY—ANXIETY. The Revised Child Anxiety and Depression Scale (RCADS, RCADS-25), a self-completed measure assessing anxiety and depression in youth ages 8 to 18, was described above.

The Screen for Child Anxiety-Related Emotional Disorders (SCARED) is a self-rated scale developed for children and adolescents ages 9 through 19 years, with a parallel form to be completed by parents. It maps onto the criteria for DSM-IV anxiety disorders of five types: somatic/panic, generalized anxiety disorder, separation anxiety disorder, social phobia, and school phobia. The overall scale and subscales show reliabilities comparable to those of other well-functioning scales, and there is evidence for convergent and discriminant validity of the SCARED with respect to other assessment measures and types of psychopathology. The SCARED has been shown to be sensitive to treatment effects. There are long (41-item) and short (5-item) versions of the scale, and the instrument has been translated into many languages. The SCARED is available at no cost at www.pediatricbipolar.pitt.edu under resources/instruments.

The Multidimensional Anxiety Scale for Children (MASC) is an empirically derived self-rated scale for children and adolescents that comprises the four major factors of anxiety, namely physical symptoms, social anxiety, harm avoidance, and separation anxiety. With respect to scale construction, the MASC maps onto the diagnosis of generalized anxiety disorder and individual factors capture the specific diagnoses of separation anxiety disorder and social phobia. With respect to psychometrics, various types of reliability and validity have been moderate to excellent. It has attractive structural features in capturing and distinguishing specific types of anxiety symptoms, and some results indicate that it distinguishes the constructs of anxiety and depressive symptoms in young people as well as being sensitive to treatment effects. The MASC has been translated into over 15 languages. The most recent form, MASC-2, has a corresponding parent version (https://storefront.mhs.com/collections/masc-2).

The Generalized Anxiety Disorder Scale (GAD-7) is another selected measure included in NIMH and Wellcome funded studies of youth. This self-completed questionnaire comprises seven items (with a 2-item short form) addressing GAD symptoms and has found use with adults and some use with youth (down to age 12). As with the PHQ, mentioned above, there is no parallel parent/caretaker form. The GAD-7 (and GAD-2) is free of charge and can be downloaded from many websites.

The Revised Children's Manifest Anxiety Scale (RCMAS), a downward extension of the adult Manifest Anxiety Scale, was developed to assess trait anxiety in school-aged children, although it has also been used with adolescents. Unlike many of the other scales, it does not have a parent informant version. The RCMAS has three component subscales: physiologic, worry/oversensitivity, and fear/concentration. It has been in use longer than the MASC and the SCARED, and thus experience has yielded a number of studies that document the psychometric properties of the RCMAS. Concurrent validity has been supported through comparisons with other measures of anxiety, although examination of discriminant validity has found that ADHD symptomatology may be rated for those who are primarily anxious. The many studies using the RCMAS have contributed to the understanding of assessment issues in anxiety disorders among the young and the role of anxiety in children and youth from various ethnic/cultural groups and diverse stressful social contexts. Some construct and other methodologic limitations have been noted which may have their origin in the inception of the RCMAS prior to the newer, DSM-linked measures (https://www.wpspublish.com/rcmas-2-revised-childrens-manifest-anxiety-scale-second-edition).

The State-Trait Anxiety Inventory for Children (STAIC), also based on an adult instrument, is another youth-only self-completed measure. The STAIC comprises two subscales, each of 20 items; one quantifies “state” with the other quantifying “trait” anxiety. It thus distinguishes between a general proneness to anxious behavior and anxiety as a fleeting emotional state. The subscales can be used independently. The STAIC is designed to be used with upper elementary or junior high school aged children youth but can be administered verbally to younger children. The STAIC has demonstrated acceptable psychometric properties and has somewhat better specificity than sensitivity. It has been shown to be sensitive to change and useful in assessing response to stress (https://www.mindgarden.com/146-state-trait-anxiety-inventory-for-children).

The Pediatric Anxiety Rating Scale (PARS) is another clinician rating scale that is freely available. The PARS, modeled on the Children’s Yale-Brown Obsessive Compulsive Scale (CY-BOCS), was developed by the NIMH funded Research Units on Pediatric Psychopharmacology, for a study on child anxiety to target symptoms that the researchers thought were not adequately covered by the HARS. In administering the PARS, the clinician first reviews the child’s specific symptoms of anxiety and then makes severity ratings on seven dimensions using a six-point scale. The psychometrics of the PARS are good and the scale is useful for tracking treatment response. However, administration can be lengthy. The scale can be downloaded for free.

The Hamilton Anxiety Rating Scale (HARS) focuses on anxiety symptoms and, like its depressive symptom counterpart, the HDRS, is a freely available clinician-rated scale, which has been used with adolescents. The HARS is generally administered to both adolescent and parent to yield an overall rating. It covers both psychological and somatic components of anxiety (focusing heavily on the latter), with psychological symptoms showing better internal consistency than somatic ones. Other aspects of reliability are moderate to excellent, depending on the type of reliability examined, while validity is not strongly supported. The use of clinicians for the assessment and the use of multiple informants are potentially advantageous features, although more study is needed to elucidate the psychometric performance of the HARS and its utility as a treatment outcome measure. The HARS can be downloaded from many websites.

SCALES FOR ASSESSING EXTERNALIZING SYMPTOMS—ATTENTION-DEFICIT/HYPERACTIVITY DISORDER WITH AND WITHOUT RELATED SYMPTOMS. The NICHQ Vanderbilt Assessment Scales, Parent and Teacher forms (VADPRS and VADTRS, respectively) are self-completed screening questionnaires that assess frequency of DSM ADHD, Oppositional Defiant Disorder and Conduct Disorders symptoms in the last 6 months on a 4-point scale ("never" to "very often"), has screening questions for Anxiety and Depression, and also questions concerning performance problems or impairment (55 items). Follow-up versions (26 items) designed for tracking change in ADHD symptoms (with a user-defined timeframe and potential side effects from medications/interventions are also available. These scales have good psychometric properties and have found wide use in clinical as well as research settings and are freely available. Although originally developed for children aged 6 to 12, it is possible to use these scales in older youth (https://www.nichq.org/sites/default/files/resource-file/NICHQ_Vanderbilt_Assessment_Scales.pdf and numerous other websites).

Several scales developed by C. Keith Conners and colleagues are commonly used to assess and monitor ADHD and other externalizing disorders. The Conners (third edition) has parent, teacher, and self-rated versions, all with good psychometric properties. There are long and short versions for each form, and each assesses symptoms of ADHD, Conduct Disorder and Oppositional Defiant Disorder, with each symptom being rated on a 4-point scale; in addition to symptom counts and scales, it also yields a Conners ADHD index and a Global Index. There is sophisticated computer software that aids in administration and scoring. The more recent Conners 4 is confined to ADHD and related problems (it does not include Conduct and Oppositional symptom assessments). Other commonly used Conners instruments include the Conners Abbreviated Symptom

Questionnaire, Hyperactivity Index (CASQ-HI), the Conners ADHD Index and Inattention/Overactivity with Aggression (IOWA) Conners Scale, and the Conners, Loney, and Milich (CLAM) Scale. The CASQ-HI is a brief (10 item) checklist of the 10 most "medication-sensitive" hyperactivity items from the longer Conners' scales. In addition to assessing treatment response, the CASQ-HI has often been used to screen for ADHD; however, it has been criticized as potentially missing children whose primary problem is inattention. The Conners ADHD Index addresses this possible weakness in identifying inattentive symptomatology in that the 12 items that comprise the index were derived to discriminate between children with ADHD and without ADHD. The IOWA Conners Scale has two subscales: Inattention/Overactivity (IO) and Oppositional Defiant (OD). For each subscale, items were selected to have the least overlap with the other scale. In practice, however, the scales correlated moderately in clinical samples. The CLAM Scale combines the CASQ with the IOWA Conners. All of these scales are brief, widely applied, have demonstrated good psychometric properties, and have been used to identify children in need of intervention and to track treatment response. In addition, Conners has developed scales for preschool children and for identification of adults who have ADHD (https://storefront.mhs.com/collections/clinical#adhd).

A second "family" of scales focusing primarily on ADHD has been developed by Swanson and colleagues, all of which are made freely available. The most widely used of these is the short version of the Swanson, Nolan, and Pelham (SNAP-IV) Rating Scale, which comprises a checklist of the DSM-III, DSM-III-R, and DSM-IV ADHD and ODD criteria rated on a 4-point scale and allows comparison between the different diagnostic formulations. The scale can be completed by either parent or teacher. The longer SNAP-IV contains items for other DSM diagnoses as well, and a more abbreviated version of the SNAP-IV contains only the ADHD items. The Strengths and Weaknesses of ADHD-symptoms and Normal-behavior (SWAN) Scale includes the same DSM ADHD symptom types as the SNAP-IV, but each is written so that it can be considered as an area of strength or weakness (e.g., the SNAP-IV "often is distracted by extraneous stimuli” appears as “ignores extraneous stimuli” on the SWAN). The parent or teacher is asked to consider each symptom as a dimension and rate the strengths or weaknesses of the child using a 7-point metric that compares how the child does in comparison to other children, ranging from “far below average” to “far above average. Finally, the Swanson, Kotkin, Agler, M-Flynn, and Pelham (SKAMP) rating scale was created to monitor the classroom behavior of a child, rather than DSM symptoms; it contains 10 items about typical target behaviors addressed by psychosocial interventions. The SKAMP is completed by the child’s teacher and is incorporated in the full teacher version of the SNAP-IV. The main disadvantage of the Swanson family of scales is the limited psychometric or normative data on their functioning. (These instruments are free and readily available for download from numerous websites.)

In addition to the Vanderbilt, Conners, and Swanson scales, there are a number of other brief symptom checklists keyed to the DSM criteria that can be used to assess ADHD and other related behaviors. These include the ADD-H Comprehensive Teacher Rating Scale (ACTeRS); the Attention Deficit Disorders Evaluation Scale (ADDES, home and school versions), the Yale Children’s Inventory (YCI), a parent rating scale designed for children with learning disabilities); the Vanderbilt AD/HD Diagnostic Teacher Rating Scale; the ADHD Rating Scale; and the ADHD Symptom Checklist.

OTHER SCALES FOR EXTERNALIZING SYMPTOMATOLOGY. The Eyberg Child Behavior Inventory (ECBI) is a parent-completed rating scale, with its teacher-completed companion, the Sutter–Eyberg Student Behavior Inventory (SESBI-R), was designed to assess a wide range of disruptive behavior symptoms in children aged 2 to 16 years. These easy-to-use scales have found wide use in treatment studies and in clinical settings and must be purchased. For each of the behaviors assessed (36 on the ECBI and 38 on the SESBI-R), the informant rates the frequency of the behavior on a 7-point scale (never to always) and then rates whether the behavior is “currently problematic.” Two scores are generated—the Intensity Score (based on the frequencies) and the Problem Score (calculated using ratings on how problematic each behavior is). Normative data for these scores are provided. The scale scores have good internal consistency for gender, ethnic, and age groups, and although they are not tied to DSM diagnoses, the ECBI has shown good discriminant and concurrent validity, strong temporal stability, and evidence of construct validity (https://www.parinc.com/products/pkey/97).

The Antisocial Process Screening Device (APSD) was designed to identify psychopathy and can be completed by either parents or teachers about children ages 6 to 13 years. This 20-item scale is based on Hare's Psychopathy Checklist-Revised. Items on the APSD address callousness, narcissism, and conduct problems that are related to poor impulse control. The scale has acceptable internal consistency, there is evidence for construct validity, and norms are available for the scale. Sensitivity and specificity have not yet been established, and there are no recommended "cut offs" for diagnosing antisocial psychopathology (https://www.wpspublish.com/apsd-antisocial-process-screening-device).

The Children's Aggression Scale (CAS), based on the Overt Aggression Scale (OAS) and the Buss–Durkee Hostility Index (BDHI), was developed with a specific focus on aggression, rather than a DSM diagnosis, as exhibited by youth ages 5 to 18 years. There are parallel parent and teacher versions (but no self-report version); the CAS has shown good psychometric properties and has been normed (https://www.parinc.com/Products/).

In addition, several other scales have been developed to monitor aggression and hostility. These include: The OAS, which is used widely in inpatient settings; modifications of the BDHI; the Vitiello Aggression Questionnaire (VAQ), which differentiates proactive from reactive aggression; and the Direct and Indirect Aggression Scale (DIAS). Unfortunately, normative data for most of these scales are not available for youth, although initial psychometric analyses look promising.

SCALES FOR OTHER SYMPTOM AREAS, IN BRIEF. For Autism spectrum disorders there are several good and widely used screening measures: the Social Communication Questionnaire (SCQ), the Autism Spectrum Screening questionnaire, the Childhood Autism Rating Scale, 2nd Edition (CARS 2) and, for very young children, the Modified Checklist for Autism in Toddlers (M-CHAT). The state-of-art diagnostic measures–the Autism Diagnostic Interview Revised (ADI-R) and the Autism Diagnostic Observation Schedule (ADOS-2)–are both clinician-administered and require specific training. For Obsessive Compulsive Disorder, the Leyton Obsessional Inventory–Child Version (LOI-CV) is useful as a screen and has good psychometric properties. The clinician administered Children's Yale-Brown Obsessive-Compulsive Scale CY-BOCS (adapted from an adult measure) is the most widely used research and clinical measure for tracking treatment change; there are also self- and parent-report versions. Widely used screening measures for Posttraumatic stress symptoms include the UCLA Post-traumatic Stress Disorder Reaction Index (UCLS-PTSD-RI) and the Children's Revised Impact of Event Scale (CRIES), both downloadable and free of charge. There are also many freely available screening measures for substance problems, including the CAGE, the Drug abuse screen test for Adolescents (DAST-A), and the Michigan Alcoholism Screening Test (MAST).

SCALES ASSESSING ADJUSTMENT AND IMPAIRMENT. Instruments also exist for assessing levels of adaptive functioning and impairment, globally or with respect to specific domains. The Children's Global Assessment Scale (CGAS) is an easy-to-use quantitative scale for the clinician's rating of overall impairment. The CGAS has good psychometric properties and has been translated into over 20 languages. A parent-completed version of the CGAS is also available. The Columbia Impairment Scale (CIS) and the Brief Impairment Scale (BIS) are both self-completed rating scales of impairment that tap four major areas of functioning: interpersonal relations, broad psychopathological domains, functioning in school or at a job, and use of leisure time with parent and youth (ages 9 and older) versions. The WHO Disability Assessment Schedule (WHO-DAS) was adapted for youth as part of the DSM-5 revision process. This scale includes parent-rated and youth-rated questionnaires as well as a clinician-rated component. The CGAS, CIS, BIS, and WHO-DAS are all free of charge.

The Vineland Adaptive Behavior Scales 3rd Edition (Vineland-3) provide a comprehensive age-normed framework for systematically recording and assessing the development of children and adults (multiple versions for birth to age 90) with respect to various realms of adaptive functioning. For youth, there are self-report, parent-report, and teacher rated forms. The Vineland has been widely used in both clinical and research settings.

Various structured and semi-structured formats for performing or recording portions of the child mental status examination have also been developed. Some, such as the child version of the Mini-Mental State Examination, contain specific standardized items for screening cognitive functions. Other instruments, such as the Mental Health Assessment Form, provide a format for organizing mental status data derived from a semi-structured clinical interview.

Limitations of Structured Measures in Clinical Assessment

Interviews and scales, such as those described previously, can, at times, be helpful adjuncts in clinical assessment if thoughtfully chosen and administered with sensitivity to their meaning in the overall process of the evaluation. For example, they can help to ensure a comprehensive review of symptoms or provide a standardized baseline measure of severity for later determination of intervention efficacy. They cannot, however, replace an individualized child psychiatric interview or be relied on as the sole basis for establishing diagnoses or planning treatment. Most standardized interview schedules focus on symptoms and behaviors, with little attention paid to coping mechanisms, situational context, and adaptive strengths that the clinical interview affords. Such factors may be as crucial to the clinical assessment and treatment planning as the presence or absence of a given pathognomonic symptom or categorical diagnosis. (Indeed, the limitations of standardized instruments in the clinical setting are acknowledged by the authors of many of these instruments, some of which require the user to seek permission from the authors and sign a license or complete a qualification form before being allowed to administer or order the instrument to ensure proper professional use. Some, like the DISC, require that the user agree that the interview not be used in place of thorough clinical evaluation in making treatment decisions.)

Another caveat in the clinical setting is to be wary of over-reliance or “reification” of a measure, especially with less experienced clinicians, who may assume a score on an inventory is more “valid” than their clinical observations, even though none of these measures should be considered as perfect. For example, a depressed child who has become withdrawn, has insomnia, and difficulty concentrating might not endorse many symptoms on a depression inventory, but that doesn’t mean the child isn’t depressed. The child’s clinician must take account of features such as the child’s affect, rate of speech, and reports of how the child is interacting with peers and family before concluding that depression is not present. Often the issue isn’t *whether* a standardized instrument might be used, but rather *how* it is used. For example, youngsters and their parents will sometimes more readily admit symptoms on a self-completed questionnaire than in a face-to-face interview. Similarly, self-completion forms can be clinically useful in giving a clinician a starting place with a noncommunicative patient. In such instances, it is useful to provide the questionnaire itself to the clinician rather than just a score, so that the clinician can review it with the patient or parent. When using any of the measures in a clinical setting, it is important to consider carefully the procedures for generating the most useful information from the measure and how best to share that information with the clinical team.

The researcher interested in selecting measures for a study protocol must carefully review the existing measures in the area of interest and consider the psychometric characteristics of potential measures as well as the appropriateness of the measures for the specific study sample and goals of the study. It is useful to begin by examining up-to-date critical reviews of these measures to evaluate the interviews/scales being considered and to actually test them out on volunteers, or if not available, with clinically savvy colleagues who can role play in developmentally and symptomatically authentic ways.

If the researcher is not experienced with biostatistics, it is highly recommended that the basic psychometric papers be reviewed by a statistician colleague. Combining information from these different perspectives will maximize the chances that the time and effort put into data collection will yield information that is interesting and ultimately useful to the fields of child psychopathology, treatment, psychosocial intervention, and ultimately prevention.

PSYCHOLOGICAL AND NEUROPSYCHOLOGICAL ASSESSMENT

Psychological tests of personality, intelligence and various neurocognitive and behavioral abilities can be a valuable tool in answering a specific referral question following an initial psychiatric evaluation described above. These tests provide a unique and more in-depth perspective on the child's functioning by generating data complementary to that obtained from history, observation, and interview. Combined with clinical interviews, psychological testing data may clarify relationships among various etiologic factors and/or may help direct the selection of appropriate intervention. Advantages of testing include identifying neuropsychological strengths and deficits, identifying learning styles, understanding underlying reasons for academic difficulty and being less influenced by certain reporting biases (e.g., social desirability, malingering) than self-report. Psychological testing also provides data that are more quantitative than the qualitative information provided by the child's history, presenting complaints, and clinical presentation.

The quantitative character of test findings is more important in the assessment of cognitive abilities than in personality assessment because the child's relative strengths and weaknesses in mental processing can be estimated only grossly from history and clinical observation. Such clinical estimates are also necessarily less accurate. The ability to discriminate meaningful differences between various domains of functioning helps clarify the particular pattern or syndrome evidenced by the child. The assessment of cognitive functioning also reveals cognitive impairments that may underlie or influence the psychological condition under investigation (e.g., whether poor social judgment in a child with conduct disorder also results from a nonverbal learning disability). Furthermore, the manner in which any cognitive impairments or strengths interact with the child’s psychological disorder can be delineated. These findings can be useful in helping students find success in school with appropriate levels of academic support and classroom accommodations. Findings from tests of personality identify important dynamics that affect the child’s clinical condition, which help inform clinical treatment recommendations for therapists, medication providers, and school counselors.

Referral for Psychometric Assessment

A referral for a psychometric assessment of the child is a request for a professional consultation to provide an opinion regarding the referral questions. The referral for psychological examination is analogous to that made to a neurologist or radiologist, who uses their assessment methods and particular expertise to contribute potentially valuable information that can be integrated into the psychiatric evaluation. For example, a referral may be made to understand underlying reasons for a teenager exhibiting sudden academic decline. Another example is a referral to understand whether attention and processing difficulties are best explained by an adolescent-onset mood disorder or a primary attention disorder dating to earlier in childhood. Referrals are, therefore, best worded in terms of the information that is needed (e.g., what is the current level of academic functioning relative to cognitive potential?), the issues to be resolved (e.g., intactness of reality testing, differentiating between potential diagnoses), or the decisions that must be made (e.g., psychiatric treatment recommendations, academic placement/accommodations), rather than requesting that specific tests be administered.

Less commonly, referral questions can be broad in scope. Requests of this nature may include a second opinion regarding the child’s condition or follow-up to previous neuropsychological testing. Regular follow-up of neuropsychological testing at fixed time intervals (e.g., every 3 years) may be indicated in medical conditions such as head injury/concussion or following neurosurgery or radiotherapy to the head for treatment of a malignancy. While it is best practice to state a specific referral question, referrals can be made for a general type of evaluation: psychological versus neuropsychological. A request for a *psychological* evaluation is commonly understood by the psychologist to mean assessment of personality, symptoms, intelligence, and achievement. In contrast, a request for a *neuropsychological* evaluation is assumed to mean a more comprehensive assessment of cognitive functioning, including intelligence, memory, complex mental processing, executive functioning (i.e., planning, organizing, time management), attention, language, visual perception, and motor functioning, with only a screening of social and emotional behavior.

Referrals for psychological examination should be made to professionals who are properly trained to use the appropriate instruments and to interpret the results. Just as psychiatrists may specialize in the treatment of adult, child, or geriatric patients or may develop a special expertise with certain disorders, psychologists also develop specialized skills in assessing certain groups of children or syndromes. Just as in medicine, board certification in clinical psychology or pediatric neuropsychology is one method for identifying psychologists with specialty qualifications. The American Psychological Association directory or various directories of certifying boards are useful references in the search for an appropriate examiner.

The primary purpose of neuropsychological evaluations is not to detect brain damage. Rather, comprehensive assessment of cognitive functioning and sensory-motor functioning permits the identification of deficits that cause academic, social, or developmental disabilities or that impede psychiatric treatment. Such assessments also identify the child’s areas of strength and weaknesses and help provide the basis for prescribing rehabilitative interventions that may be delivered as special education, occupational therapy, physical therapy, or vocational or cognitive rehabilitation. By parsing the psychological and cognitive impairments that contribute to the child’s functional disability, such assessments also help delineate the interaction between the child's psychological disorder and cognitive impairments. Although neuropsychologists are familiar with the cognitive sequelae of various neurologic conditions and can be particularly helpful in the assessment of children with neurologic disorders, any child requiring a detailed assessment of cognitive functioning is appropriate for referral.

Psychological Test Batteries

Table 35.1–7 contains a list of commonly used psychological measures for various domains of functioning. A comprehensive examination samples all of the domains, whereas a more focused evaluation selects a subset. In the most general sense, a battery of tests consists merely of a collection of tests to measure a variety of mental functions. There are loosely defined batteries that are routinely administered in psychological examinations. For example, the psychological battery for a child often includes the Wechsler Intelligence Scale for Children-V (WISC-V) or the Kaufman Assessment Battery for Children-III (K-ABC-III) as the measure of intelligence. In addition, there may be a measure of achievement in basic linguistic skills and math, such as the Test of Achievement from the Woodcock–Johnson Psychoeducational Battery—Fourth Edition (W/JPB-IV)—or the Wechsler Individual Achievement Test—Third Edition (WIAT-III). A test of visuomotor integration or visual construction, such as the Beery–Buktenica Test of Visual Motor Integration (VMI) or the Rey–Osterreith Complex Figure, is also usually administered. Finally, a variety of personality tests and behavior rating scales are used.

Some published test batteries consist of measures of various mental functions that have been developed together and can usually be combined into composite or factor scores. For infants and preschool children, psychometric instruments primarily focus upon development, including motor, sensory, perceptual, and language development and the development of conceptual reasoning. Comprehensive developmental quotients are derived from many of these instruments. The Bayley Scales of Infant and Toddler

Development 4 (Bayley-4) and the Mullen Scales of Early Learning are two notable examples. The Woodcock-Johnson Psychoeducational Battery-Fourth Edition (W/JPB-IV) and the NEPSY-2 are appropriate for school-aged children and examine a variety of smaller domains of functioning.

Personality Measures

Personality tests include both objective and projective assessment tools. For adolescents, the Minnesota Multiphasic Personality Inventory-Adolescent (MMPI-A), the Millon Adolescent Personality Inventory (MAPI) and the Millon Adolescent Clinical Inventory (MACI) are three well-known examples. These tests involve between 160 and 478 true–false self-report items and yield personality profiles based on scales of various psychological attributes such as emotional patterns, clinical symptoms, and apprehensions in teen life. Objective measures of personality for younger children are usually limited to behavior rating scales or structured reports by parents or caregivers. The Personality Inventory for Children—Second Edition (PIC-2) and the Behavior Assessment System for Children—Second Edition (BASC-2) are good examples of these measures.

For younger school-aged children, personality measures consist primarily of projective tests. Projective tests contrast with objective tests of personality in that projective tests use limited structure and ambiguous stimuli to which the child can respond in a virtually unlimited variety of ways. Perhaps the best-known example of a projective technique is the Rorschach Test. Other projective tests commonly used for the young, school-aged child include the Robert’s Thematic Apperception Test (RTAT) and the Children’s Apperception Test (CAT), which requires the child to compose a story in response to a picture depicting a scene that is suitable for children. A variety of tests administered to this age group require the child to draw, such as the Draw-A-Person (DAP) test, the House-Tree-Person (HTP) test, and the Kinetic Family Drawing (KFD). (Methods exist for scoring children’s drawings to assess cognitive and visual motor development, as well as affective content.) For the most part, the interpretation of responses to projective tests of personality is inferential; that is, the meaning is not self-evident in the response. The inferences are based on accumulated clinical experience reported in the literature, symbolic interpretation or analogous reasoning based on a theory (e.g., psychoanalytic theory), and, to a much lesser extent, empirically established norms. The notable exceptions are the scores derived from the RTAT and the Exner and Rorschach-Performance Assessment System (R-PAS) scoring systems for the Rorschach test, which have norms and have been extensively studied in empirical research.

Table 35.1–7.
Commonly Used Psychological and Neuropsychological Tests for Children and Adolescents

General Ability/Intelligence Tests	For Ages (years)
Differential Ability Scales, Second Edition (DAS-II)	2:6–17:11
Kaufman Assessment Battery for Children, 2nd ed. (KABC-2)	3:0–18:11
Stanford-Binet, Fifth Edition (SB-5)	2:0–89:11
Wechsler Preschool and Primary Scale of Intelligence, 4th ed. (WPPSI-IV)	2:6–7:7
Wechsler Intelligence Scale for Children, 5th ed. (WISC-V)	6:0–16:11
Wechsler Adult Intelligence Scale, 4th ed. (WAIS-IV)	16:0–90:11
Wechsler Abbreviated Scale of Intelligence 2nd ed. (WASI-II)	6:0–90:11
Woodcock–Johnson Tests of Cognitive Abilities, 4th ed. (WJ-IV)	2–90+ years
Achievement Test Batteries	
Kaufman Test of Educational Achievement, 3rd ed. (KTEA-3)	4:6–25:11
Peabody Individual Achievement Test—Revised (PIAT-R)	4:0–50:11
Wechsler Individual Achievement Test, 3rd ed. (WIAT-III)	4:0–50:11
Wide Range Achievement Test 4 (WRAT4)	5–94 years
Woodcock–Johnson Tests of Achievement, 4th ed. (WJ-IV)	2–90+ years
Neuropsychological Test Batteries	
Delis-Kaplan Executive Function System (DKEFS)	8–89 years
Halstead–Reitan Neuropsychological Test	5–8, 9–14, 15+

Batteries	years
Luria-Nebraska Neuropsychological Battery—Children's Revision	8–12 years
Luria-Nebraska Neuropsychological Battery (LNNB)	15+ years
NEPSY-II: A developmental neuropsychological assessment, 2nd ed.	3–16 years
Reitan-Indiana Neuropsychological Test Battery	5–8 years
Nonverbal Tests[a]	
Comprehensive Test of Nonverbal Intelligence, 2nd Ed. (C-TONI-II)	6:0–89:11
Leiter International Performance Scale Revised, 3rd ed. (Leiter-3)	3:0–75+
Naglieri Nonverbal Ability Test-Individual Administration, 3rd ed. (NNAT-3)	5–18 years
Test of Nonverbal Intelligence, 4th ed. (TONI-4)	6:0–89:11 years
Adaptive Behavior	
AAMR Adaptive Behavior Scale-School, 2nd ed. (ABS-S2)	3–21 years
Adaptive Behavior Assessment System, 3rd ed. (ABAS-3)	0–89 years
Scales of Independent Behavior-Revised (SIB-R)	infancy–80+ years
Vineland Adaptive Behavior Scales, 3rd ed.	0–90 years
Vineland Social-Emotional Early Childhood Scales (SEEC)	0–5:11
Diagnostic Instruments	
Anxiety Disorders Interview Schedule for Children (ADIS-C)	5–17
Child and Adolescent Psychiatric Assessment	8–17

(CAPA)	
Children's Interview for Psychiatric Syndromes (ChIPS)	6–18
MINI International Neuropsychiatric Interview for Children and Adolescents (MINI-KID)	6–17
Diagnostic Interview Schedule for Children (DISC-5)	6–17
Schedule for Affective Disorders and Schizophrenia for School-aged Children (K-SADS) and computerized version of KSADS (KSADS-COMP)	6–18
Projective Tests	
Children's Apperception Test	3–10
Draw-a-Person Technique	5–17
Kinetic Drawing System for Family and School	5–20
Rorschach Inkblot Test	5–adult
Roberts Apperception Technique for Children: (Roberts-2)	6–18
Rotter Incomplete Sentence Blank	Child–adult
Thematic Apperception Test	6–adult

[a] The KABC-2, DAS-II, and SB5 also yield nonverbal composite scores.

Modified with permission from Tsatsanis K, Hart L, Powell KK. Psychological and neuropsychological assessment of children. In: Martin A, Bloch MH, Volkmar FR, eds. Lewis's Child and Adolescent Psychiatry: *A Comprehensive Textbook.* 5th ed. Wolters Kluwer; 2018:321–335.

Objective personality tests are usually more structured than are projective tests, in that the stimuli have limited ambiguity, and only a select number of responses are permitted. In most objective measures, individuals are asked whether various statements are true or false for them.

Validity and Reliability

Although the validity and reliability of individual tests, as measures of specific mental functions, are established with group data, the reliability and validity of findings for the individual child are also a concern for the examining psychologist. Test batteries provide some advantages in determining the reliability and validity of individual test results. An integrated battery contains some overlap in the testing of mental functions, permitting comparative assessment of the consistency of the child's performance across measures. In addition, if there is a theoretical or empirical basis for knowing that some mental abilities are associated with each other, then consistent findings across measures of those abilities strengthen the confidence in the findings for the individual child. Examinations that focus on a few mental functions, by administering a limited number of tests, do not permit detection of transient performance factors that may cause spurious scores. In addition, no test is a pure measure of the target variable; therefore, performance on a single test may be impeded by impairment of a mental function necessary for performing the task but not the function that the test purports to measure. If previous psychometric evaluations have been performed, reliability of findings for a particular child can be estimated from the consistency of findings across examinations. While reliability is limited by various performance factors on the day of testing, data can still provide very useful insight as a snapshot of the child's current functioning (e.g., current depressive episode). Implications of testing biases are taken into account in the interpretation of results, discussed below.

Finally, the assessment of children with motor, language, hearing, or visual impairments presents a special challenge. Every psychological measure requires a modality of perception, mental processing, and modality of expression. To assess children with impairments in any of these functions, tests are selected with forms of input and expression that avoid the obviously impaired functions. Children with pervasive disorders of mental processing require assessment methods appropriate for their level of functioning. In such cases, it is frequently more helpful to characterize the impairments in terms of criterion-referenced measures, which list abilities mastered, or in terms of the child’s standing among children of the same diagnostic group.

Interpretation of Psychometric Findings

The interpretation of psychometric findings is predicated on standardized administration of the tests. In addition, the proper interpretation of psychometric findings generally requires knowledge of parametric and nonparametric statistics, mastery of the literature concerning test performance and test construction, familiarity with the construction and psychometric properties of the specific tests employed, and knowledge of the literature regarding those instruments.

The results of psychological examination are often best integrated in a concluding discussion of the findings as they pertain to the referral question. The results of a neuropsychological examination are organized around a discussion of various functions with a concluding integration. The interpretation of test results includes consideration of the immediate context in which the tests were administered (e.g., performance factors, such as fatigue, anxiety, rapport, level of cooperation and motivation of the child and examiner bias). There are generally three levels of analysis for the test scores: (1) pathognomonic symptoms, (2) levels of performance, and (3) syndromes or patterns. There is no direct interpretation for single scores. During syndrome analysis, scores are interpreted within the context of the results of other measures, which is the most powerful form of analysis in terms of the ability to extrapolate beyond the immediate tests data. This capacity to extrapolate is based on known characteristics of children evidencing the syndrome. Finally, when testing children, one must bring a developmental perspective to findings, so that future difficulties can be anticipated with the changes that occur as the child grows older.

Psychometric findings are communicated directly to the referring provider and to the family along with the data provided in a comprehensive report. These data are more easily understood if the psychological examiner first helps those receiving the findings understand some basic principles of psychometrics. First, the psychological examiner often explains (often by drawing a graph for visual representation) that most people cluster around some average score for standardized measures of mental abilities. This implies that fewer individuals score in the high or low ranges compared to the average range. The distribution approximates a binomial distribution in the form of an inverted U, known as the *normal curve.* Raw scores are frequently the number of correct responses during a test and are not usually reported. Percentile scores indicate the percentage of persons obtaining raw scores equal to or less than the raw score obtained by the individual. In the normal distribution, small changes in raw scores close to the average (where many people are grouped) result in relatively large shifts in the percentile rank, whereas large shifts in raw scores at the high and low ends of the distribution may change the person's percentile rank little.

To compensate for the variations in the rate of change of percentile scores, and to make performance levels comparable across tests, raw scores are converted to standard, or deviation scores that indicate the extent to which the individual deviates from the mean or average. By convention, the *average standard score for most psychological tests is 100*, and the *standard deviation is set at 15* IQ points or standard score points. By the use of standard scores, performance levels for measures of specific mental abilities can be compared with IQ or with other test results. Standard scores also give the best indication of significant differences among scores, because the intervals between scores of the same magnitude are equal anywhere along the distribution of scores. Statistical significance occurs when the probability is low that a difference between scores could have occurred by chance, as the result of inherent error variance in the test. By convention, a level of .05 probability is set as the standard for statistical significance. Although two scores may significantly differ, that does not mean that the difference is meaningful or has clinical significance. Many statistically significant differences occur frequently in the normal population. For example, a difference of 15 points between the Verbal and Performance IQ scores of the WISC-IV is statistically significant; however, a difference of this magnitude or greater occurs in approximately one-fourth of the normal subjects used to standardize the test. Of course, a difference of this magnitude could have clinical significance in a particular case, if the adverse effect of the difference was exacerbated by other factors.

Some tests for children allow calculating standard scores from age norms or grade norms, for example, the Woodcock–Johnson Psychoeducational Battery and the Kaufman Tests of Educational Achievement. The difference between these two methods of calculation illustrates the importance of understanding tests scores in relation to the normative reference group. A standard score based on age norms compares the performance of a child to that of other children the same age. A standard score based on grade norms compares the child's performance with that of other children in the same grade. If the child has been retained in grade to compensate for learning problems, the standard score based on grade norms is artificially inflated, masking the extent of the child's impairment. However, an achievement score based on grade norms indicates how well the child has mastered the curriculum at the grade level. In general, grade scores (grade and months) and age scores (years and months), as opposed to standard scores, should be interpreted cautiously, because their psychometric properties are poor, and the implications of the scores vary across levels. Clinical decisions should not be based upon these types of scores.

There has traditionally been a concern about reporting IQ scores, because people do not understand that the scores are not absolute and immutable. Reporting the standard error of measurement or confidence intervals helps others realize that scores have inherent error and vary to some extent over successive administrations of the test, even when the test is valid and reliable by present standards. Confidence intervals are the range of scores believed to contain the true score at a specified level of probability. The WISC-IV and WJPB-III scoring programs, for example, will calculate confidence intervals for standard scores, at 90% and 95% probability. Higher levels of confidence necessitate larger ranges of scores.

Report of the Psychometric Findings

Ordinarily, the child's parents are informed of the results of a psychological or neuropsychological evaluation during a feedback session, typically without the child. Children may receive a separate session to communicate relevant information from the testing, such as relevant clinical information and treatment recommendations that apply to the child. During this feedback session, the psychological examiner presents and discusses the results of testing and the clinical implications. Because tests scores are easily misinterpreted, reports of psychological test findings should never be given to the family without opportunities to answer questions.

A written psychological report is typically provided for the family during the feedback session. The report provides a detailed description of the tests administered, scored data, and the impressions and conclusions based on the data. The report is designed to be shared with whomever the child's parents choose, usually the referring provider, other members of the child's treatment team, and the school. It is important to note that confidentiality guidelines state that it is the *parents' choice* whether or not to share the results of their child's evaluation. For example, parents may not wish to share the report with the school because they do not want it included in the child's academic record. Thus, it is best practice to provide the written report directly only to the parents so that they have a chance to read the report thoroughly and share it with those whom they choose. Of course, clinical information is to be never released or shared without obtaining a written permission for the release of information. Furthermore, some examiners and parents may agree that it is best to provide a separate written report for the school containing test finding and relevant academic recommendations, while limiting the amount of other sensitive information about the family and the child.

PEDIATRIC AND OTHER ADJUNCTIVE EVALUATIONS

Pediatric Evaluation

Close collaboration with the child's pediatrician is usually an important component of any evaluation. The pediatrician is a valuable source of information about the child's developmental and medical history and can contribute a unique long-term perspective on the child and family. Especially when the pediatrician has had a longitudinal relationship with the family, they can also serve as an important source of support.

In addition to a general review of systems, the pediatrician should assess the child's current and past physical growth, including head circumference, height, weight, and pubertal status. Screening of hearing and vision, with more detailed testing as indicated, is important because deficits in these areas can manifest themselves in learning, language, and social delays or seeming inattention or oppositionality. History taking includes abnormal movements or absence spells suggesting seizures, abnormal postures, and tics. The physical examination notes difficulties in muscle tone and strength; asymmetrical, abnormal, or primitive reflexes; tremor; and gait or postural disturbances.

Especially in children with intellectual disability or autism spectrum disorder, a thorough history and physical examination can provide important clues to an underlying detectable medical cause. For children with developmental difficulties, a careful history of perinatal risk factors is taken; the pediatric examination includes careful evaluation for major or minor congenital anomalies, dysmorphic features, dermatologic abnormalities (tubers, fibromas, and areas of hyperpigmentation), or ophthalmologic findings (lens abnormalities and retinal pigmentation). Loss of previously acquired developmental milestones or the presence of congenital anomalies, dysmorphic features, microcephaly or macrocephaly and other growth problems, abnormal skin pigmentation or lesions, and symptoms suggesting seizure activity or other neurologic abnormality warrant more detailed genetic, metabolic, or neurologic evaluation.

Diagnostic Laboratory Tests

The clinical usefulness and cost-effectiveness of routine laboratory tests for children who present with psychiatric problems has not been substantially demonstrated. Most guidelines for performing these tests for children have historically been developed using data from studies of adults. Adult studies generally suggest that routine laboratory tests are not clinically useful in typical psychiatric settings, such as outpatient clinics and most inpatient units, although laboratory screening tests may be of some use in psychiatric settings in which patients are at high risk of medical illness, such as the emergency room, substance abuse treatment centers, AIDS clinics, and geriatric clinics, as well as for those who have new-onset psychosis, depression, or dementia or when evaluating children with developmental disturbances, such as intellectual disability or autistic spectrum disorder. Similarly, routine screening laboratory measures are more likely to yield clinically significant information when clinical symptoms of physical illness are present.

Evidence suggests that if anything, routine laboratory testing is less useful in pediatric compared to adult populations, where medical comorbidity that causes psychiatric manifestation is less common. The few studies of the use of routine laboratory tests in child psychiatry patient populations have yielded similar conclusions. One review of routine laboratory screenings (e.g., thyroid function tests, electroencephalogram [EEG], chest x-ray, chemistry panel, urinalysis, complete blood count, electrocardiogram [ECG]) in 100 consecutive adolescent inpatient admissions reported variable rates of abnormal values, depending on the specific test, but in only 1 of these 100 patients did the tests produce a change in diagnosis from functional to organic—and even then the diagnostic information proved to have little clinical relevance. Most laboratory abnormalities in this study were regarded as minor and did not indicate a need for clinical follow-up.

Similarly, routine use of more specialized diagnostic laboratory evaluations (EEG, computed tomography [CT], magnetic resonance imaging [MRI], or chromosomal analyses) also provide a relatively low yield of clinically useful information in most cases. In a study of 200 consecutive child psychiatric inpatients, these evaluations were performed only when clinically indicated; despite their judicious use, the tests provided clinically relevant information in only 7 patients

(3.5% of the total patient sample) or in 7 of 136 tests (5.1% of all tests performed). Chromosomal analyses proved to be the most informative of all these tests, yielding new medical diagnoses in 5 of 32 (15.6%) selected children for whom they were performed. A study of 111 putatively high-risk inpatients with new-onset adolescent psychoses produced similar conclusions. In this population, routine endocrine and neuroimaging screening evaluations failed to provide any information of diagnostic usefulness in any patient (although inconsequential laboratory abnormalities were present in 15.4% of the neuroendocrine screens and 11% of the neuroimaging screening tests). Structural neuroimaging assessments are typically of greater utility in children presenting with significant deficits in neurologic or cognitive function, atypical or abrupt presentation of psychiatric illness, or a history of head trauma. More specialized neuroimaging technologies, such as positron emission tomography (PET), single-photon emission computerized tomography (SPECT), and functional MRI (fMRI), currently have no routine clinical or diagnostic usefulness in child and adolescent psychiatric populations.

Two useful screening diagnostic tests in child psychiatric practice include (1) urine drug screening when indicated, and (2) testing for sexually transmitted diseases in high-risk patients. Substance use can cause a wide range of neuropsychiatric symptoms and mimic a large spectrum of psychiatric conditions. The high prevalence of substance abuse in adolescent populations has prompted recommendations that toxicology screens be obtained for (1) all adolescents who have psychiatric symptoms or who exhibit acute behavioral changes; (2) high-risk adolescents, such as delinquents and runaways; and (3) adolescents who have recurrent accidents or unexplained somatic symptoms. The possibility of substance use must always be considered as a cause of unexplained behavior in adolescents and it is likely that urine drug screening is often underutilized in clinical practice (especially emergency settings). Similarly, children and adolescents should be screened for risk factors associated with sexually transmitted diseases. Children and adolescents at risk due to either a history of sexual activity (or sexual abuse) or intravenous substance use should be evaluated for sexually transmitted diseases, including HIV infection and syphilis.

In contrast to routine screening, targeted laboratory testing in specific child psychiatric syndromes can be quite useful in (1) detecting the etiology of conditions or (2) uncovering treatable, high-risk medical comorbidities. Specific characteristics of a child's presentation can often trigger targeted diagnostic testing. For instance, testing thyroid functioning should be considered in the evaluation of a depressed adolescent, especially one presenting with prominent low energy, weight gain, menstrual irregularities, and/or brittle hair. Similarly, targeted diagnostic testing may be triggered when a child's psychiatric presentation places them at increased risk for specific medical conditions. For instance, children with intellectual disability or autism spectrum disorder may particularly benefit from targeted diagnostic testing. In general, the more severe the child's disability, the greater is the likelihood of an identifiable (albeit not necessarily treatable) cause. For example, one study found that a medical cause could be identified in 60% to 70% of children with severe intellectual disability versus only 35% to 55% of children with mild intellectual disability. Chromosomal abnormalities, especially trisomy 21 (Down syndrome) or fragile X syndrome, account for more than one-half of cases of intellectual disability. Children with autism spectrum disorder have relatively high rates of intellectual disability, seizure disorder (20% to 40%), and tuberous sclerosis (3% to 9%). Most experts would therefore advocate a Wood lamp examination to search for tubers, an EEG to exclude seizures, and high-resolution chromosome analysis to exclude known genetic syndromes in children who display features of autism spectrum disorder or intellectual disability. Identification of genetic causes of autism is now possible in more than 10% of affected children and, as current knowledge of specific genetic syndromes advances, high-resolution DNA sequencing is likely to increase the ability to discern specific subtypes of neurodevelopmental disorders.

Lead testing is also indicated in cases of intellectual disability because chronic lead ingestion due to pica can cause substantial retardation; conversely, many children with intellectual or pervasive developmental disabilities from other causes indulge in pica and may accumulate a substantial lead burden.

Finally, periodic laboratory testing (as well as tracking height and weight) is also important when starting certain psychiatric medications or monitoring for potential adverse during ongoing administration. For instance, height and weight must be tracked in youngsters on long-term stimulant medication and weight, hemoglobin A_{1C}, and fasting glucose and lipid profile in patients starting and continuing antipsychotic medications.

DIAGNOSTIC FORMULATION AND RECOMMENDATIONS

In the diagnostic formulation, the clinician attempts to integrate the data gathered through the assessment process into a coherent account of the child's difficulties, the factors that appear to have predisposed the child to develop the problem, the concomitants and consequences of the problem, and the factors that maintain the problem or that might ameliorate it. It is not always possible to provide a complete formulation at the conclusion of the initial assessment; under such conditions, the best that the clinician may be able to provide is a differential diagnosis that can point toward the appropriate treatment options and the subsequent steps needed to clarify or establish a diagnosis.

One product of the assessment may be the assignment of a diagnosis of one or more categorical psychiatric disorders using the diagnostic schema of DSM-5-TR or ICD-11. Although these nosologic systems have done much to standardize diagnostic practices and to improve the quality of research, one must be aware of their clinical limitations as applied to children. Many children experience sufficient symptoms, distress, or impairment to be regarded as psychiatrically disturbed and to warrant therapeutic intervention without meeting the full criteria of a categorical psychiatric disorder. Furthermore, as currently conceptualized, the DSM-5 classification is descriptive and atheoretical; it specifies neither presumed etiology nor treatment for the various diagnostic categories. If the assessment process is to provide a picture of the child that suffices to guide intervention, the resulting diagnostic formulation must go beyond simply assigning a categorical psychiatric diagnosis (if one is warranted). As noted in the introduction, comorbidity is often the rule rather than the exception in childhood, and many children have complex patterns of developmental and behavioral difficulties that are not adequately captured by a list of multiple categorical diagnoses. Children who share a given categorical diagnosis may differ in crucial respects that influence current severity, natural history, and treatment response. Examples of potential risk (or exacerbating) factors and protective (or ameliorative factors) include intelligence, compensatory skills, family and community resources and supports, and vulnerabilities, such as comorbid pathologic personality traits, neurologic impairments, coexisting general medical conditions, or family pathology.

The clinician's diagnostic formulation should thus supplement the assignment of any formal categorical diagnoses by attempting to identify, as far as possible, the potential causes, predisposing factors, and current determinants of the child's difficulties. On the basis of this information and the clinician's expertise regarding the various interventions available for the diverse forms of childhood psychopathology and developmental difficulties, the clinician formulates appropriate recommendations for the treatment of the child's problems.

Communicating Findings and Recommendations

Communicating the findings and recommendations that result from the assessment to the parents and child is an essential part of the evaluation process and may require one or more sessions. Depending on the problem and the child's age and level of comprehension, presenting the findings usually involves meeting with the child and parents, separately or together.

Several principles must be observed to maximize the likelihood that the parents and child hear, understand, and experience the clinician's findings and recommendations as helpful. First, the clinician must communicate a sense of the child as a whole person, with strengths and abilities, as well as problems and vulnerabilities. This perspective conveys a sense of the clinician's appreciation and empathic understanding of the child, reduces defensiveness, and helps to mitigate the fear that there is only bad news.

The findings and recommendations are most useful to the family and likely to be implemented if the interpretive session is a dialogue rather than a lecture. The clinician must use language comprehensible to the parents and child, must avoid jargon, and must keep technical terms to a minimum. When technical terms are necessary, they should be explained clearly, so the parents and child understand them; terms that are merely descriptive to the clinician may hold frightening or negative connotations for the child or family. The clinician should allow ample opportunity for the parents and child to discuss the recommendations and to discuss any questions. This provides a chance to judge whether the clinician's formulation made sense to the family, to address any differences of opinion, and to explore the feasibility and acceptability of the recommendations offered. Remaining areas of uncertainty or ambiguity about the assessment or recommendations should be clearly noted and discussed.

If the assessment is a consultation requested by an agency, a school, or another clinician, the findings and recommendations should be communicated to the referring party in appropriate terms after being discussed with the parents and child and obtaining their consent. If further evaluation or treatment is indicated but is best done by someone else, the clinician should offer to assist in finding an appropriate referral.

FURTHER READINGS

Angold A, Erkanli A, Copeland W, Goodman R, Fisher PW, Costello EJ. Psychiatric diagnostic interviews for children and adolescents: a comparative study. *J Am Acad Child Adolesc Psychiatry*. 2012;51(5):506–517.

Angold A, Fisher PW. Interviewer-based interviews. In: Shaffer D, Lucas CP, Richters JE, eds. *Diagnostic Assessment in Child and Adolescent Psychopathology*. Guilford Press; 1999:34–64.

Bird HR, Canino GJ, Davies M, et al. The Brief Impairment Scale (BIS): a multidimensional scale of functional impairment for children and adolescents. *J Am Acad Child Adolesc Psychiatry*. 2005;44(7):699–707.

Bostic JQ, Potter MP, King RA. Clinical assessment of children and adolescents: content and structure. In: Martin A, Bloch M, Volkmar FR, eds. *Lewis Comprehensive Textbook of Child and Adolescent Psychiatry*. 5th ed. Wolters Kluwer; 2018:299–320.

Canino GJ, Fisher PW, Alegria M, Bird HR. Assessing child impairment in functioning in different contexts: implications for use of services and the classification of psychiatric disorders. *Open J Med Psychol*. 2013;2(1):29–34.

Carlson GA, Singh MK. Emotion dysregulation and outbursts in children and adolescents: Part I & II. *Child Adolesc Psychiatr Clin N Am*. 2021;30(2–3).

Cotton NK, Shim RS. Editorial: social determinants of health, structural racism, and the impact on child and adolescent mental health. *J Am Acad Child Adolesc Psychiatry*. 2022;61(11):1385–1389.

*Fisher PW, Chin EM, Vidair HB. Use of structured interviews, rating scales, and observational methods in clinical settings. In: Tharper A, Pine DS, Leckman JF, Scott S, Snowling MJ, Taylor E, eds. *Rutter's Child and Adolescent Psychiatry*. 6th ed. Wiley; 2015:419–435.

Galán CA, Tung I, Tabachnick AR, et al. Combating the conspiracy of silence: clinician recommendations for talking about racism-related events with youth of color. *J Am Acad Child Adolesc Psychiatry*. 2022;61(5):586–590.

*Gilliam WS, Mayes LC, Tobon AL. Clinical assessment of infants and toddlers. In: Martin A, Bloch MH, Volkmar FR, eds. *Lewis's Child and Adolescent Psychiatry: A Comprehensive Textbook*. 5th ed. Wolters Kluwer; 2018:291–299.

*Goff PA, Jackson MC, Di Leone BA, Culotta CM, DiTomasso NA. The essence of innocence: consequences of dehumanizing Black children. *J Pers Soc Psychol*. 2014;106(4):526–545.

Kaufman AS, Raiford SE, Coalson DL. *Intelligence testing with the WISC-V*. Wiley; 2016.

Leckman JF, Tayler E. Clinical assessment and diagnostic formulation. In: Tharper A, Pine DS, Leckman JF, Scott S, Snowling MJ, Taylor E, eds. *Rutter's Child and Adolescent Psychiatry*. 6th ed. Wiley; 2015:419–435.

Leibenluft E, Kircanski K. Chronic irritability in youth: a reprise on challenges and opportunities toward meeting unmet clinical needs. *Child Adolesc Psychiatr Clin N Am*. 2021;30(3):667–683.

Loeber R, Green SM, Lahey BB. Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *J Clin Child Psychology*. 1990;19:136–143.

McGuire R, Halligan SL, Meiser-Stedman R, Durbin L, Hiller RM. Differences in the diagnosis and treatment decisions for children in care compared to their peers: An experimental study on post-traumatic stress disorder. *Br J Clin Psychol*. 2022;61(4):1075–1088.

Piqueras JA, Martín-Vivar M, Sandin B, San Luis C, Pineda D. The Revised Child Anxiety and Depression Scale: a systematic review and reliability generalization meta-analysis. *J Affect Disord*. 2017;218:153–169.

Sanislow CA, Morris SE, Pacheco J, Cuthbert BN. The National Institute of Mental Health Research Domain Criteria: an alternative framework to guide psychopathology research. In: Geddes JR, Andreasen NC, Goodwin GM, eds. *The New Oxford Textbook of Psychiatry*. 3rd ed. Oxford University Press; 2020.

Shaffer D, Fisher P, Lucas C. Respondent-based interviews. In: Shaffer D, Lucas CP, Richters JE, eds. *Diagnostic Assessment in Child and Adolescent Psychopathology*. Guilford Press; 1999:3–33.

*Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME. NIMH Diagnostic Interview Schedule for Children, Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *J Am Acad Child Adolesc Psychiatry*. 2000;39(1):28–38.

Sparrow SS, Cicchetti DV, Saulnier CA. *Vineland Adaptive Behavior Scales*. 3rd ed. Pearson; 2006. www.pearsonclinical.com

Tolin DF, Sain KS, Davis E, et al. The DIAMOND-KID: psychometric properties of a structured diagnostic interview for DSM-5 anxiety, mood, and obsessive-compulsive and related disorders in children and adolescents. *Assessment*. 2023;10731911221143994.

Towbin KE. Physical examination and medical investigation. In: Tharper A, Pine DS, Leckman JF, Scott S, Snowling MJ, Taylor E, eds. *Rutter's Child and Adolescent Psychiatry*. 6th ed. Wiley; 2015:449–459.

Townsend L, Kobak K, Kearney C, et al. Development of three web-based computerized versions of the Kiddie Schedule for Affective Disorder and Schizophrenia child psychiatric diagnostic interview: Preliminary validity data. *J Am Acad Child Adolesc Psychiatry*. 2020;59(2):309–325.

*Tsatsanis K, Hart L, Powell KK. Psychological and neuropsychological assessment of children. In: Martin A, Bloch MH, Volkmar FR, eds. *Lewis's Child and Adolescent Psychiatry: A Comprehensive Textbook*. 5th ed. Wolters Kluwer; 2018:321–335.

World Health Organization. Social determinants of health. Accessed January 15, 2023. https://www.who.int/health-topics/social-determinants-of-health

35.2 Psychiatric Assessment of Preschool Children

HELEN LINK EGGER, M.D., AND CAROLY SUE PATAKI, M.D.

The infant/preschool mental health field has made real strides in recent years in demonstrating that clinically significant psychiatric symptoms and disorders are already present and highly impairing in early childhood. Recent epidemiologic studies of children ages 2 to 5 years old have shown that the overall rate of psychiatric disorders in preschoolers is remarkably similar to the overall rates reported for older children and adults. For example, the preschool rate of "serious emotional disturbance," defined as psychiatric disorder(s) that are associated with significant impairment, was 12.1%, 9.1%, and 7.1% in three different community preschool studies, compared with a median rate across community studies of 13.2% in older children and adolescents and 15.6% in adults. The growing recognition of the mental health needs of preschool children reveals a large gap in mental health services. Few clinicians have training in the assessment and treatment of young children. This section provides an overview of the components of comprehensive, empirically based psychiatric assessments of preschool children and their families.

DEVELOPMENTALLY APPROPRIATE CRITERIA FOR PSYCHIATRIC SYMPTOMS AND DISORDERS IN PRESCHOOL CHILDREN

Historically, young children's emotional and behavioral problems have been approached dimensionally, with cut-points set to characterize subsets of children at the extreme of the distribution of (1) normative behaviors and emotions and/or (2) temperament traits. In these models, extremes of problems or temperament traits have typically been regarded as precursors of, or risk factors for, later psychopathology rather than as psychiatric disorders similar to those identified at later stages of life. Although symptom checklists or temperament measures often provide clinically useful information about the young child, they do not have the symptom coverage (particularly for rare but highly concerning symptoms such as suicidality or stereotypies) nor details about the characteristics of symptoms (e.g., frequency, onset, duration) to enable the clinician to make a diagnosis. There have been two approaches to the diagnostic classifications of early childhood psychopathology (1) the *Diagnostic and Statistical Manual of Mental Disorders/International Statistical Classification of Diseases and Related Health Problems* (DSM/ICD) diagnostic classification, either unmodified or modified to be developmentally appropriate for young children, and (2) the Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood (DC:0-3 [1994], DC:0-3R [2005], and DC:0-5 [2016]), an alternative system for classifying mental health disorders in infants, toddlers, and preschoolers. These categorical approaches seek to identify clinically significant syndromes, characterized by severity, pervasiveness, persistence, and impairment, that are early-onset “disorders” rather than simply risk factors for later disorders.

Despite its lack of developmental specificity, the DSM has been a useful starting point for examining how psychopathology presents in early childhood, at least in children 2 years old or older. For example, the diagnostic criteria of the fifth edition of the DSM (DSM-5) and its text revision (DSM-5-TR) work quite well as a means of advancing the current understanding of the presentation and course of autism spectrum disorders despite its lack of developmentally specific criteria. The dual approach of testing the applicability of current diagnostic criteria for identifying autism spectrum disorders in young children while exploring the validity and clinical utility of developmentally specific criteria and/or diagnostic algorithms for use with young children provides an exemplary road map for examining the validity of specific behavioral and emotional disorders in young children. Similar approaches are being taken in studies of early-onset attention-deficit/hyperactivity disorder (ADHD), anxiety disorders, depression, post-traumatic stress disorder, and disruptive behavior disorders in young children. The challenges of applying the DSM criteria to young children are greater than in later childhood because of the limitations of children’s language expression and cognitive development, the rapidity of developmental change during the first 5 years of life, and the greater interdependence of very young children on their caregivers.

There have been two approaches to the development of a mental health classification system for young children: a top-down and a bottom-up approach.

The top-down approach begins with the nosology used for adults and older children and then proposes developmentally appropriate modifications of these criteria. In 2003, the Research Diagnostic Criteria-Preschool Age (RDC-PA) was published with the goal of facilitating further research on the diagnostic validity of psychiatric disorders in preschoolers by defining clearly specified, developmentally appropriate criteria for preschool psychopathology. The RDC-PA explicitly sought to modify the DSM criteria to make them developmentally appropriate rather than to propose a new classification system. The RDC-PA facilitated research on the nosology of disorders in young children.

The alternative approach seeks to start with the young child and use clinical and scientific knowledge about infants, toddlers, and preschoolers to characterize mental health disorders in a separate classification. The first edition of the DC:0-3, published in 1994, reflects the “bottom up” approach. The 2005 revision of the DC:0-3 (DC:0-3R, 2005) and the most recent version, DC:0-5, combined the two approaches, as well as explicitly addressing the crosswalk between the separate early childhood classification and the DSM/ICD. The DC:0-3R and DC:0-5 sought to incorporate research and clinical expertise to operationalize the characteristics of diagnostic criteria and, where possible, explicate differences in presentations in infants, toddlers, and preschoolers. The diagnostic classification includes alternative criteria for DSM-5-TR diagnoses, new diagnostic categories (e.g., regulation disorders of sensory processing), and a multiaxial system.

The 2016 DC:0-5 represents the most significant revision of the early childhood classification. Key changes to this version include: One, the name of the classification was changed to DC:0-5 to reflect a common understanding that “early childhood” spans from infancy through the preschool period and to align the title with the fact that early childhood practitioners have been using the nosology through the preschool period despite the name. Second, to distinguish between transient changes in a young child’s emotions or behaviors or typical individual differences, the DC:0-5 requires child distress or the child’s and/or the family’s functional impairment for every disorder in order to avoid pathologizing nonclinical presentations. Third, despite the fact that the DSM no longer uses a multiaxial system, the DC:0-5 retained the multiaxial system, primarily to emphasize the importance of context for understanding psychopathology in young children. The DC:0-5 axes were substantially revised. All disorders, including relationship disorders, are recorded on Axis I, while the characteristics of the caregiver–child relationship are recorded on Axis II. Axis V was revised to reflect the child's developmental competencies. Lastly, all disorders relevant to the young child are included in the classification (previously, users were referred to the DSM when DMS disorders such as autism spectrum disorder were appropriate for use with young children). The manual does include a "crosswalk" from each DC:0-5 diagnosis to a parallel DSM and ICD diagnostic codes to enable the early childhood nosology to be integrated into mainstream nosologies and to facilitate billing for services.

What System Should a Clinician Use?

At this point, it is quite clear that the DC:0-5 is the current diagnostic nosology for diagnosing psychopathology in infants, toddlers, and preschoolers. In addition to the developmentally sensitive criteria, the DC:0-5 manual includes information about the presentation, prevalence, course, comorbidities, risk factors, and differential diagnoses for each disorder. As noted above, the DC:0-5 manual includes a "crosswalk" from DSM-5 codes to DSM/ICD codes because these codes are often needed to obtain insurance reimbursement for services.

COMPREHENSIVE, EVIDENCE-BASED PSYCHIATRIC ASSESSMENTS OF PRESCHOOL CHILDREN

There are five key components of a comprehensive psychiatric assessment of a preschool child: (1) multiple sessions, (2) multiple informants, (3) a multidisciplinary approach, (4) multiple modes of assessment, and (5) a multiaxial diagnostic formulation. These components arise from the multidisciplinary, developmental, relationship-based, family-centered, and prevention-focused orientations that characterize the infant/preschool mental health field's approach to understanding psychopathology in early childhood.

Multiple Sessions

Despite the real reimbursement constraints that clinicians face in conducting multisession evaluations, these constraints cannot dictate the length of an evaluation. It is simply not possible to conduct an adequate psychiatric assessment of a young child in one session. Although it could be argued that this caveat is true for all of pediatric psychiatry, the rapid developmental, cognitive, emotional, and behavioral changes that characterize the preschool period, as well as the young child's physical and emotional dependency on their primary caregivers, make it imperative that information be gathered about the child across sessions and across time. For example, a tired 3-year-old child seen after a long day at preschool may look quite different when seen in the morning after they have had a good night's sleep. The best practice is for the clinician to see the parent(s) without the child for one of the sessions so as to obtain a detailed history of the child's symptoms and the family history and current family structure and functioning. It is essential to obtain information about current and past parental psychiatric and substance abuse problems and treatments, domestic violence in the home, and marital (or partner) difficulties, but it is not appropriate to discuss these issues in front of the child. Nor is it appropriate for the child to hear negative comments about himself or herself (e.g., "He is a bad kid") or to be privy to a parent's worst fears ("I am afraid that he will end up like my brother who killed himself"). These comments and fears are very important pieces of information about how the parent sees this child. Any parent of a preschooler also knows that it is quite difficult for young children to refrain from interrupting adult conversation long enough for the clinician to obtain the breadth and level of detail needed to understand the child's problems. However, even if the child is seemingly absorbed in play in the corner of the office, they will hear the adult conversation. An evaluation should include phone sessions with the child's teacher or child care provider (written consent to contact the teacher or child care provider should be obtained at the first session). A visit to the child's daycare or preschool to observe the child's functioning with other adults and peers in an out-of-home setting may also be indicated.

Multiple Informants

The clinician must gather information from multiple informants. First and foremost, the child is an important informant. This is as true of a 2-year-old with limited expressive language as of a 5-year-old about to enter kindergarten. Information can be obtained by speaking with and/or playing with the child (using structured or unstructured approaches) and observing the child interacting with parents, with another clinician, and with teachers or childcare providers, or peers. It is not possible to assess or diagnose a young child based on parent reports alone. Moreover, the clinician should make every effort to spend time alone with the child without the parents present (severe separation anxiety is an example of a disorder that might make this very difficult).

Of course, an evaluation cannot be conducted without information from the child's primary caregivers. These may be the child's biologic parents and/or adults who act as parents of the child, such as step-parents, grandmothers, or foster parents. An evaluation should include a three-generation genogram identifying the important parental figures and family members (including psychiatric and substance abuse history for each member). Ideally, if there are two parents in the home, both of them will attend the evaluation sessions and fill out all measures and information separately. Often each parent provides a somewhat different description of the child's symptoms. This does not undermine the validity of these reports but rather reflects the fact that the child shows different aspects of themselves within different relationships. This information is diagnostically important and may be useful in guiding treatment planning. The parents will also provide information about themselves (e.g., current psychiatric symptoms and treatment, parental and marital stress). Information must also be obtained from the child's nonparental caregivers, such as teachers or child care providers, nannies, and babysitters who spend a significant amount of time with the child. As described in more detail in the next section, other clinicians, such as the child's pediatrician, are also critical informants.

Multidisciplinary Assessment

A child psychiatrist conducting a preschool psychiatric assessment usually collaborates closely with other clinicians and care providers, including psychologists, primary care and developmental pediatricians, neurologists, occupational therapists, speech and language therapists, early interventionists, case managers, adult psychiatrists and psychotherapists who treat the disorders from which the parents may be suffering, and marital therapists. The child psychiatrist has an opportunity (if not an obligation) to assume a leadership role among all of these care providers. At the evaluation stage, the child psychiatrist should integrate the assessments and recommendations made by other providers into their assessment and then facilitate treatment planning and coordination of communication among these providers. The child psychiatrist should never provide treatment for a young child, particularly treatment with psychotropic medication, without having conducted a comprehensive assessment themselves.

Multiple Modes of Assessment

Because it is not possible to conduct a standard psychiatric interview with a young child about their emotions and behaviors, preschool mental health assessments must draw on information gathered using different approaches, including adult reports, observational assessments, structured or unstructured play/interactions with the child, and standardized psychopathology, cognitive, motor, language, and other developmental assessments. A physical exam (often conducted by the child's pediatrician) is essential. Laboratory measures, including genetic testing, may be indicated.

When possible, empirically tested measures with demonstrated reliability and validity should be used to assess child and family domains. Types of measures include symptom checklists, such as the Strength and Difficulties Questionnaire, which is valid for children 2 or older, the Child Behavior Checklist 1½-5, the Infant Toddler Social Emotional Assessment for children 12 to 36 months old, DSM-referenced rating scales such as the Early Childhood Inventory-4 (ages 3 to 6 years), and checklist measures of specific symptom clusters such as the ADHD Rating Scale or the Spence Preschool Anxiety Scales. Teacher/Caregiver measures such as the Caregiver-Teacher Report Form, as well as those assessing social skills and peer relationships, such as the Social Skills Rating System, are also useful. Similar types of rating scales are available to assess adult domains, such as psychopathology (e.g., Mood and Feelings Questionnaire), parental stress (e.g., Parenting Stress Index), or domestic violence (Conflict Tactic Scales).

In the future, structured diagnostic parent interviews for assessing preschool psychopathology will be the best way to obtain comprehensive psychiatric, developmental, and environmental information about the child and the family, but these measures are not yet user-friendly enough for use in clinical settings. The exception is the Autism Diagnostic Interview

(ADI). The Preschool Age Psychiatric Assessment (PAPA) is currently the only comprehensive parent psychiatric interview for assessing psychiatric symptoms and disorders in children ages 2 to 5 years with demonstrated test-retest reliability and validity. There is also a version of the PAPA for assessing nonpervasive developmental disorder (PPD) symptoms in children with autism and other PPDs (PAPA-PDD). The ePAPA, a web-based version of the PAPA (and PAPA-PDD) to be administered on a tablet personal computer, which is part of the eMeasures system, is used in research studies but has not yet been developed for use in clinical settings.

Direct “interviews” with the young child about their feelings and experiences are also an essential component of a comprehensive assessment. In most clinical practices, unstructured play is used to learn about the child’s emotions. Structured child interviews, such as the MacArthur Story-Stem Battery (MSSB), have shown some promise for children 4 years old and older. Children tell us what they are feeling and experiencing through their play and their behavior. Unlike older children and adults, young children do not yet have the skills to share verbally how they are feeling or to regulate their emotions. Dysregulated early childhood behaviors such as temper tantrums are as common to be a sign of an emotional disorder as a “behavioral” disorder.

Observational assessments, particularly of parent–child interactions, play an important role in preschool assessments. Inspired by the success of the Autism Diagnostic Observational Schedule—a structured assessment for autism in which the examiner “presses” for PDD symptoms with specific tasks with the child—researchers have developed a similar diagnostic observational measure for assessing disruptive behavior disorders in preschoolers (Disruptive Behavior-Diagnostic Observational Schedule [DB-DOS]). The same group has developed a similarly structured observational measure of emotional disorders for use with preschool children.

There are numerous standardized, developmentally specific measures for assessing the preschooler’s status in cognitive, language, motor, and preacademic domains. Reviews of specific measures in this and the other areas mentioned in this section can be found in the textbooks listed in the reference section.

Multiaxial Assessment

A comprehensive psychiatric assessment in the preschool age group must be truly multiaxial, with each axis reflecting the developmental and relationship-based framework of infant and early childhood mental health. Since DSM-5-TR no longer uses a multiaxial system, coding on the DC:0-5 system provides an opportunity for clinicians to be comprehensive in assessing children’s emotions and behaviors within the context of their relationship with their primary caregivers, their medical history, their history of trauma and adversity, and their developmental competencies across the domains of emotional, social-relational, language-social communication, cognitive, and movement/physical development. In addition to the DC:0-5 manual and Zero to Three’s DC:0-5 training, there are excellent infant/early childhood mental health books, including the *Handbook of Infant Mental Health*, the *Handbook of Preschool Mental Health*, and the *Handbook of Infant, Toddler, and Preschool Mental Health Assessment*.

The child’s symptoms across all axes must be understood within the child’s current level of functioning across multiple developmental domains (e.g., cognitive functioning, expressive and receptive language, gross and fine motor skills, prekindergarten knowledge, and social-emotional skills). These developmental assessments are important for (1) understanding the child’s psychiatric symptoms within their developmental capacities (not simply the child’s chronologic age), (2) identifying developmental lags or deficits that can be effectively addressed (e.g., expressive language problems), and (3) ruling out (or ruling in) the presence of a developmental disorder such as a speech delay. Because children 3 years old and younger (and in some states 5 years old and younger) are eligible for early intervention services for developmental delays, it is critical that child psychiatrists identify these delays and provide appropriate treatment referrals.

Assessment of the context of the child’s symptoms, as well as the child’s environment within and outside of the family, is also vital. The pervasiveness of the child’s symptoms, not only across settings but also across relationships, will provide important information about the nature of the symptoms. For example, if a child has explosive tantrums that occur when the child is at home and at school and occur with the parent and with the child’s daycare provider, the child may well be showing a symptom of an emotional or behavioral disorder. The child who has similar tantrums only when separating from their mother may be experiencing a symptom of an anxiety disorder. The child who only has tantrums when being dropped off for a custody visitation with a divorced parent may be expressing distress that arises from the difficulty of moving back and forth between two parental homes. The child who only has tantrums with their mother but shows good emotional and behavioral regulation with all other adults may be showing signs of a relationship disturbance with the mother. The infrequent tantrums that occur only when a child is exhausted (e.g., after a long car trip traveling home from a visit with Grandma) may very well be developmentally normal. These examples highlight the importance of obtaining detailed examples of the child's behaviors with a focus on the characteristics of the symptoms. Relevant questions include the following: What does the child do or say? How often does it happen? How long does it last? When did it start? What triggers the symptom? When, where, and with whom does the symptom occur? What helps? This descriptive information is not the same as asking the parent, "Why does your child do that?" Often, one does not know "why," but detailed descriptions of specific symptoms provide the information needed to rule in or rule out specific disorders. An important difference between DSM and DC:0-5 is that the DC:0-5 provides a place (Axis II) to record information about the relational context of the child's symptoms. Part A of Axis II characterizes the caregiver's and the infant/young child's contribution to their relationship and their mutual adaptive functioning. Part B examines the caregiving environment and the caregiver and child's adaption to this environment.

Assessment of the child's environment and how stressors impact the child is also essential. In the DC:0-5 systems, psychosocial and environmental stressors and risk factors are listed on Axis IV. DC:0-5 includes a broad (but not inclusive) checklist of known stressors, along with a place to indicate the age of onset, duration, severity, and context of these stressors, which is meant to provide a guide to the assessment of this domain. Axis IV should reflect a careful history of the child's environment, including, to name a few, an understanding of (1) parental/family factors such as a family history of psychiatric and substance abuse disorders and treatment, including types and doses of psychotropic medications and types of psychosocial treatments (a history of postpartum depression should be obtained), family structure and functioning including a careful assessment for physical or emotional abuse or neglect and sexual abuse, parent–partner relationship functioning including an assessment of adult-adult domestic violence and emotional abuse, and sibling relationships; (2) pregnancy and birth factors, including premature birth, low–birth-weight, maternal smoking or use of other substances during pregnancy, neonatal complications, and maternal depression during or following the child's birth; (3) the child's school or day care experiences, including teacher–child ratios, frequency of changes in child care settings and providers, amount of time the child is away from parent(s), age when the child started out-of-home care, peer relationships, and relationships with nonparental child care providers; (4) community factors, including neighborhood safety and proximity of relatives; and (5) socioeconomic factors, including parental employment status, family's social economic status, and absence of health insurance. Major stressors (e.g., parental divorce, death of a parent, physical abuse, hospitalization, or natural disasters such as a house fire or hurricane) known to be potentially traumatic life events should be included. However, chronic stressors or those that may seem "minor" (e.g., change in daycare, moving to a new house) should also be included on Axis IV. There is clear evidence that the accumulation of multiple stressors with varying intensities has as profound an impact as single traumatic events in early childhood, just as they do in later childhood and adolescence. Both acute and cumulative stressors significantly raise the risk that the child will have one or more psychiatric disorders and be impaired.

CLINICAL FORMULATION AND TREATMENT PLANNING

Clinical formulation is an integrative skill that involves the pulling together of multiple observations and sources of information about an individual within a general diagnostic scheme so as to guide decision-making. It is worth emphasizing that a psychiatric nosology classifies disorders, not individuals. When a comprehensive psychiatric evaluation has been conducted with a preschool child and their family, and the DSM and/or DC:0-5 axes have been completed, a clinical formulation about the child and a treatment plan addressing each of the child's challenges areas should be developed. Clearly defined symptoms within the child must be identified, ideally using measures that are sensitive to change, so that the clinician will be able to assess the efficacy of the intervention and the outcome of treatment. Of course, the assessment may also identify targets for treatment beyond the child's mental health symptomatology. For example, a treatment plan might include a referral of the child for speech or occupational therapy or a referral of a parent for treatment of their depression.

The evidence is quite clear that approximately 10% to 15% of preschoolers have severe impairing psychiatric disorders that cause suffering and put them at risk of ongoing emotional and developmental problems. These children and their families deserve effective and safe treatment. The evidence base for psychiatric treatments for preschoolers is emerging. There is every reason to hope that there will be major advances in the current understanding of the etiology and course of early-onset mental health problems, as well as the development and dissemination of effective treatments and preventive interventions. However, children and their families cannot wait for the research to catch up with this need. Clinicians who treat preschoolers must (1) know the current state of evidence on the assessment and treatment of preschool psychopathology, (2) know how to decide on and determine the efficacy of treatments that may not have been studied in young children, and (3) be committed to keeping up with and incorporating new findings into their practice. Clinicians and parents should together weigh the risk of the disorder(s) with the risk of treatments so that parents can make fully informed decisions about how to best care for their child. Single-case experimental designs (e.g., *n*-of-one trial) that seek empirically to determine the efficacy of treatments for preschool patients are the best ways to manage these risks. This is particularly vital when the use of psychotropic medication is being considered. The guidelines for the psychopharmacologic treatment of preschoolers recommend this empirically driven (and cautious) approach to treatment. The infant/preschool mental health field has always had an early intervention and prevention orientation. The hope is that the identification and treatment of preschool children with clinically significant psychiatric symptoms and disorders will not only ease their suffering during early childhood but also will contribute to ameliorating or even preventing the onset of new disorders or progression of these disorders into later childhood and adulthood.

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. 2013.

Angold A, Egger HL. Psychiatric diagnosis in preschool children. In: DelCarmen-Wiggins R, Carter A, eds. *Handbook of Infant, Toddler, and Preschool Mental Health Assessment*. Oxford University Press; 2004.

Carpenter KLH, Sprechmann P, Calderbank R, Sapiro G, Egger HL. Quantifying risk for anxiety disorders in preschool children: a machine learning approach. *PLoS One*. 2016;11(11):e0165524.

Carter AS, Briggs-Gowan MJ, Jones SM, Little TD. The infant-toddler social and emotional assessment (ITSEA): factor structure, reliability, and validity. *J Abnorm Child Psychol*. 2003;31:495–514.

Costello EJ, Egger HL, Angold A. 10-year research update review: the epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. *J Am Acad Child Adolesc Psychiatry*. 2005;44:972–986.

Egger HL, Angold A. Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. *J Child Psychol Psychiatry*. 2006;47:313–337.

Egger HL, Erkanli A, Keeler G, Potts E, Walter BK. The test-retest reliability of the preschool age psychiatric assessment (PAPA). *J Am Acad Child Adolesc Psychiatry*. 2006;45:538–549.

Franz L, Angold A, Copelane W, Costello EJ, Towe-Goodman N, Egger H. Preschool anxiety disorders in pediatric primary care: prevalence and comorbidity. *J Am Acad Child Adolesc Psychiatry*. 2013;52:1294–1303.

Gleason MM, Egger HL, Emslie GJ, et al. Psychopharmacological treatment for very young children: contexts and guidelines. *J Am Acad Child Adolesc Psychiatry*. 2007;46:1532–1572.

Luby J. *Handbook of Preschool Mental Health: Development, Disorders, and Treatment*. Guilford Press; 2006.

Martin SE, Hunt JI, Mernick LR, et al. Temper loss and persistent irritability in preschoolers: implications for diagnosing disruptive mood dysregulation disorder in early childhood. *Child Psychiatry Hum Dev*. 2017;48(3):498–508.

Narrow WE, First MB, Sirovatka PJ, Regier DA. *Age and Gender Considerations in Psychiatric Diagnosis: A Research Agenda for DSM-V*. American Psychiatric Association; 2007.

Sim F, Haig C, O’Dowd J, et al. Development of a triage tool for neurodevelopmental risk in children aged 30 months. *Res Dev Disabil*. 2015;45–46:69–82.

Sveen TH, Berg-Nielsen TS, Lydersen S, Wichstrøm L. Screening for persistent psychopathology in 4-year-old children. *Pediatrics*. 2016;138(4):e20151648.

Task Force on Research Diagnostic Criteria: infancy preschool. Research diagnostic criteria for infants and preschool children: the process and empirical support. *J Am Acad Child Adolesc Psychiatry*. 2003;42:1504–1512.

Towe-Goodman NR, Franz L, Copeland W, Angold A, Egger H. Perceived family impact of preschool anxiety disorders. *J Am Acad Child Adolesc Psychiatry*. 2014;53:437–446.

Tseng WL, Guyer AE, Briggs-Gowan MJ, et al. Behavior and emotion regulation modulation deficits in preschoolers at risk for bipolar disorder. *Depress Anxiety*.

2015;32:325–334.

von Klitzing K, White LO, Otto Y, Fuchs S, Egger HL, Klein AM. Depressive comorbidity in preschool anxiety disorder. *J Child Pychol Psychiatry*. 2014;55:1107–1116.

Wichstrom L, Berg-Nielson TS, Angold A, Egger HL, Solheim E, Sveen TH. Prevalence of psychiatric disorders in preschoolers. *J Child Psychol Psychiatry*. 2012;53:696–705.

Zeanah CH, Carter AS, Cohen J, et al. Diagnostic classification of mental health and developmental disorders of infancy and early childhood DC:0-5: selective reviews from a new nosology for early childhood psychopathology. *Infant Ment Health J*. 2016;37(5):471–475.

Zero to Three: Diagnostic Classification. *0–3R: Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood: Revised Edition*. Zero to Three Press; 2005.

Zero to Three: Diagnostic Classification. *0–5: Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood: Revised Edition*. Zero to Three Press; 2016.

Zucker N, Copeland W, Franz L, et al. Psychological and psychosocial impairment in preschoolers with selective eating. *Pediatrics*. 2015;136:e582–e590.

35.3 Neuropsychological and Psychoeducational Assessment of Children

EVANGELINA BARNARD, PSY.D., AND PATRICIA D. WALSHAW, PH.D.

This section describes the process, application, and interpretation of cognitive assessment of children. Within the umbrella term of cognitive assessment, it is useful to differentiate two common yet different methodologic approaches to child assessment: neuropsychological and psychoeducational assessment. Both neuropsychological and psychoeducational assessments are undertaken as ways of understanding the referral questions of children and adolescents and how the processing of information can affect a wide range of adaptive functions. Although testing procedures in these types of assessments might overlap, these approaches analyze behavior according to two different paradigms. The psychoeducational assessment focuses on assessing a child's intelligence and learning style to inform academic and school-based interventions. Neuropsychological assessment is undertaken in the context of growing knowledge about brain–behavior relationships, and it has the additional possibility of describing the child in terms of neural pathways. These approaches provide ways of conceptualizing how children integrate information (and their thinking, learning, and responding) at different levels. The psychoeducational assessment focuses on understanding the child's functioning at a more behavioral level (i.e., whether a child shows learning difficulties in certain areas) and in descriptive terms. Neuropsychological assessment introduces the additional possibility of understanding such behavior at neural levels and in neurobiologic terms. However, regardless of the theoretical differences among the psychologists undertaking these evaluations, in a practical way, it is the patient's referral question that shapes the assessment process and focuses the interpretation of results.

BASICS OF COGNITIVE ASSESSMENT

Cognitive assessment involves more than testing. Although measurements are useful, testing involves more than scores.

The Testing Process

In addition to testing, assessment procedures include examining past records (medical examinations, prior testing, report cards), interviewing the client and their family (in structured and unstructured formats), obtaining information from home and school (and, sometimes, on-site observations), and obtaining rating scales completed by the child's parents and teachers (regarding developmental, behavioral, emotional, and diagnostic concerns). The diagnostic aspect of the process involves an attempt to determine the psychiatric and educational categories for which the client meets the criteria. Testing is only an aspect of an attempt to establish a broad view of the way a child solves problems in the world, to understand their unique interaction with any diagnostic category, and to provide recommendations for interventions. Once quantitative data from any type of evaluation are gathered, they mean nothing without a qualitative interpretation, as well as a conceptualization and presentation that helps to make sense of the child client to their caretakers. Feedback about results and their relationship to the presenting problems can be provided as the evaluation unfolds. Often this discussion is among the first opportunities the parents have had to understand the patient's predicament in a detailed, sympathetic way and to begin a constructive process. It can be helpful for the family to comprehend the patient's difficulties as a natural outgrowth of problems with understanding and learning, so long as practical interventions and solutions are provided.

Measurements in Testing

Although many techniques might be used to help understand a child client and their referral question, the emphasis here is on standardized testing (based on regularized procedures as well as normative data). The aim of testing is to create a way of comparing one individual to a population of such individuals, as well as the strengths and weaknesses within one individual. The psychologist will select an instrument that is valid (it measures what is intended) and reliable (it measures it consistently).

There are different ways of describing a child's placement compared to others of the same demographic, the most common of which are standard scores and percentiles. This involves converting raw scores to standard scores that can be compared to other scores along what is thought of as a normal distribution with predictable statistical properties. It should be noted that there are different kinds of scores and values within the normal distribution and that they differ by mean and dispersion. (Fig. 35.3–1, which depicts Z, T, and intelligence quotient [IQ] scores, as well as rounded percentiles under the "bell" curve.) While standard scores (such as those with a mean of 100) can be informative, parents often respond well to the concept of a percentile. For example, a student who falls at the 68th percentile performs as well as or better than 68% of the children tested. In addition to percentiles and standard scores, many cognitive tests provide age equivalents or grade equivalents. These measures suggest how the child compares to others of their age or grade.

It should be noted that tests are constantly updated. This is often undertaken to correct perceived shortcomings, adjust for cohort changes, and to make instruments more relevant. Because of this, instruments are referred to in "generic form" in this text without any notation that would identify the edition. However, in actual practice, it is important to know that the clinician has used the most up-to-date version of the test unless there is a good reason not to. For example, some clinicians with specialized practices continue to use old editions (or subtests) of instruments because they find that particular instruments or tasks best meet the needs of their particular population or best answer a type of diagnostic question.

(American Psychological Association guidelines require essentially that the tests used to make interpretations or recommendations are not outdated or obsolete for the “current purpose.”) Clinicians must select from among the available tests the one that seems to best meet the needs of any particular client. Table 35.3–1 notes the editions of tests that are current as the present text goes to press.

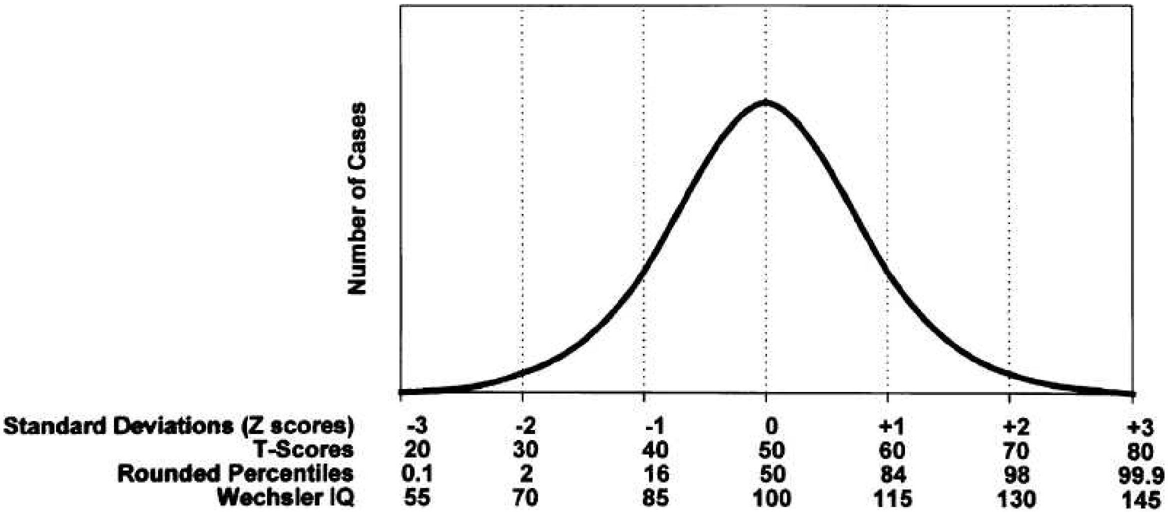

FIGURE 35.3–1. Normal curve and standardized scores.

Cultural and Linguistic Considerations

Cultural and linguistic factors are important considerations in both conducting psychoeducational and neuropsychological evaluations and interpreting findings. Research exploring how an individual’s culture influences their approach to, and performance on, various measures has highlighted discrepancies and raised awareness about the cultural “fairness” of cognitive measures.

Culture provides a framework for ways of thinking, feeling, and behaving (Ardila, 2013). An individual’s cultural background can influence how they approach the assessment, which in turn can impact their perceived performance. For example, while the speed of performance may be valued and understood in Western cultures, other cultures place a much higher value on answering carefully and correctly. Timed measures emphasizing speed over accuracy may therefore produce greatly different results across cultures. Many object naming tests utilize items common to western cultures but rare or unknown in other cultures. In addition to influencing the approach to test measures, culture may also influence how an individual relates to the examiner, their comfort level with the assessment, and their willingness to share details of their medical history and symptoms.

Historically, the majority of assessment measures have been normed on White, non-Hispanic monolingual English-speaking individuals (Casas et al., 2008). While some efforts to improve the diversity of normative data have been undertaken, particularly in adult populations, most assessment norms for children remain relatively homogenous. The impact of this disparity in norms can be seen in test performance across individuals of different ethnic and linguistic backgrounds. For example, one study demonstrated that African American preschool children were significantly more likely to be classified in the “impaired” range on the Expressive Vocabulary Test (EVT) and Peabody Picture Vocabulary Test (PPVT) language tests as opposed to their non-Hispanic White classmates. Another study demonstrated that African American, Hispanic, and Native American children typically perform lower on IQ and achievement tests as compared to their non-Hispanic White counterparts. Indeed, a lawsuit in California in 1979 resulted in a ban on utilizing IQ tests to identify whether African American children should be placed in special education classrooms. Similar disparities in test performance are seen in bilingual and multilingual children as compared to monolingual English-speaking children. These are a few examples; however, they underscore the importance of considering the cultural context of the child being assessed.

If an individual is bilingual or multilingual, ensuring that the assessment is conducted in their primary as well as secondary language(s) is key to gaining a comprehensive understanding of their cognitive abilities. In an ideal scenario, this assessment would be conducted by a trained psychologist who is fluent in the same languages, utilizing measures that have been standardized as well as normed on the population in question. Unfortunately, in many cases, this ideal situation is not feasible due to either lack of appropriate standardized measures, available trained practitioners, lack of norms, or all of the above. If a trained practitioner is not available, the use of a medical interpreter combined with measures normed within this population may be used, although a caveat should be made on the interpretation of the data. Often, appropriate norms are regrettably not available, and assessment with an interpreter utilizing standard norms is conducted with the caveat that the results are more qualitative and may not fully capture the individual’s abilities.

When reviewing an assessment report, it is important to note whether the examiner has identified the child's cultural and ethnic background and the languages in which the child is fluent. Furthermore, as a report consumer, it will be key to identify whether the examiner attempted to utilize the most appropriate norms, acknowledged the impact of culture and ethnicity on the assessment proceedings, and ensured the child was assessed in all proficient languages.

Beyond Scores and Tests

It should be noted that the testing process involves more than scores. Although scores are important, how the patient goes about solving cognitive problems also is carefully observed. The examiner is interested not only in test performance but also in behavioral observations—the client's reaction, types of errors they are making, and behavior during testing. It is important for the psychologist to note how the client arrives at right as well as wrong answers and to explore the client's cognitive strategies on tasks. In general, it is important to note whether the client responds in a deliberate or impulsive way. It also makes a difference to the examiner whether the child says, “This is easy,” as they fail, or laments, “Oh, this is hard,” as they succeed. Both give the examiner a very different sense of how the client will experience school and life.

Although the focus here is on the assessment of broad cognitive functioning and targeted domains, depending on referral issues, the evaluation might also address behavioral, social, and emotional issues using a variety of techniques. It is well established that emotional, behavioral, and social issues can affect cognition and learning and vice versa. The psychologist will use reports from parents and teachers, observations, and test data to examine how these issues are playing a role in the overall picture of the child. These are integrated with other cognitive assessment data to help clarify diagnosis and aid in treatment planning. Specific information about processing problems (how the child perceives their experience and monitors their behavior, and therefore responds to their situation) can help adults to know better how to interact with the child, as well as plan appropriate educational and interpersonal interventions.

Table 35.3–1.
Cognitive Domains and Tests

Domain	Age Range	Description
Intellectual Tests		
Omnibus Intellectual and Ability Tests		
Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition (WPPSI-IV)	2 years, 6 months–7 years, 7 months	These tests consist of several subtests that can be combined in different ways to understand basic intellectual functioning. Many, if not most, cognitive tests (including intellectual, achievement, and domain targeted) are structured in the same way (mean is 100, standard deviation is 15), and so results can be compared across instruments. This is particularly important in identifying learning problems. The Wechsler tests and SB5 yield a full-scale IQ score, as well as overall scores across several indices (e.g., verbal, visual, fluid reasoning, working memory). The KABC scales and their subtests (which minimize verbal instructions and responses) include measures of sequential
Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V)	6–16 years, 11 months	
Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV)	16–90 years, 11 months	
Stanford–Binet Intelligence Scales–Fifth Edition (SB5)	2–85+ years	
Kaufman Assessment Battery for Children–Second Edition (KABC-II)	3–18 years	
Woodcock–Johnson–Fourth Edition Tests of Cognitive Abilities (WJ-IV COG)	2–80+ years	
Differential Ability Scales–Second Edition (DAS-II)	2 years, 6 months–17 years, 11 months	
A Developmental Neuropsychological Assessment–Second Edition (NEPSY-II)	3–16 years, 11 months	

		and simultaneous processing, fluid reasoning and crystallized ability, and long-term retrieval. The WJ-COG scores provide information about intellectual ability, as well as a variety of cognitive and clinical areas. The WJ-COG is co-normed with the WJ-ACH, and the new test of oral language (WJ-OL). The DAS provides a measure of general conceptual ability and a profile of cognitive strengths and weaknesses. While it is not a classic intellectual test and does not provide an IQ measure, the NEPSY provides scores across a variety of cognitive domains (executive functioning/attention, language, sensorimotor, visuospatial, memory/learning, and social perception).
Nonverbal and Language-Free Ability Tests		
Leiter International Performance Scale–	3–75+ years	These tests are often given to assess abilities

Third Edition (Leiter-3)	6–89 years, 11 months	in the absence of language. This can be used for both individuals who have language difficulties and those who do not speak English and where no tests/interpreter are available in that language/culture. Although the nonverbal tests are often seen as language or culture free, strictly speaking, this is not always true because of representational content or language demands. However, some nonverbal tests (such as the TONI-4) depend more on abstract pattern recognition and can make fewer demands on language systems. The WNV provides batteries for younger (4–7:11) and older clients (8–21:11), and claims to be a general ability measure using nonverbal tests. Tests like Raven's Progressive Matrices (which actually come in several forms) are
Test of Nonverbal Intelligence–Fourth Edition (TONI-4)	6–89 years, 11 months	
Comprehensive Test of Nonverbal Intelligence, Second Edition (CTONI-2)	4–21 years, 11 months 5–17+ years	
Wechsler Nonverbal Scale of Ability (WNV)		
Raven's Progressive Matrices		

		based on pattern recognition and thought of as “culturally reduced,” if not unbiased or unloaded. Nonverbal tests can be useful in assessing individuals whose experience falls outside the “cultural norm.”
Quick Intellectual Tests		
Kaufman Brief Intelligence Test–Second Edition (KBIT-2) Wechsler Abbreviated Scale of Intelligence-Second Edition (WASI-II)	4–90 years 6–90 years, 11 months	The KBIT and WASI include verbal and nonverbal scores, but do not include as much information as more comprehensive intellectual tests. However, they provide an estimate of intellectual level in a fraction of the time.
General Achievement Tests		
Wechsler Individual Achievement Test–Fourth Edition (WIAT-4) Woodcock–Johnson Tests of Achievement-Fourth Edition (WJ-IV ACH) Kaufman Test of Educational Achievement–Third Edition (KTEA-3) Wide Range Achievement Test–Fifth Edition (WRAT5) and WRAT-Expanded (WRAT-E)	4–50 years, 11 months 2–80+ years 4–25 years, 11 months 5–94 years 4–24 years	Tests in this domain measure academic performance (knowledge and fluency of academic skills—reading, writing, math). The WRAT5 includes subtests on sentence comprehension, word reading, spelling, and math computation. An expanded version (WRAT-E) provides achievement and

		intellectual information (nonverbal reasoning). The KTEA assesses reading (decoding and comprehension), mathematics, spelling, and written expression. The WIAT and WJ-ACH also systematically address the basic areas that have been identified as relevant for learning disability. The WIAT-4 includes a Dyslexia Index Composite.
Reading Tests		
Gray-Oral Reading Tests–Fifth Edition (GORT-5)	6–23 years, 11 months	The WRMT (which includes a measure of listening comprehension) and GORT (which includes a measure of fluency) involve systematic approaches to assessing different aspects of reading skills. Because of the importance of phonologic awareness in the development of reading skills, tests of phonologic processing (such as the CTOPP) are often included in assessments in which reading problems are identified.
Woodcock Reading Mastery Tests–Third Edition (WRMT-III)	4 years 6 months–79 years, 11 months	
Comprehensive Test of Phonologic Processing–Second Edition (CTOPP-2)	4–24 years, 11 months	

Language Tests		Tests in this domain assess language abilities, either globally or targeting a specific aspect (i.e., expressive, receptive, pragmatic). The PPVT assesses receptive language while the EVT assesses expressive language. The CELF-5 provides a comprehensive assessment of language skills.
PPVT, Fifth Edition (PPVT -5)	2 years, 6–90+ months	
EVT, Third Edition (EVT-3)	2 years, 6–90+ months	
Clinical Evaluation of Language Fundamentals, Fifth Edition (CELF-5)	5–21 years, 11 months	
Attention and Executive Functioning		Tests in this domain assess attention and/or executive functioning abilities. As a test of executive functions, nine stand-alone subtests in the DKEFS assess cognitive flexibility, response inhibition, problem solving, and concept formation. The CPT-3 is a computerized measure of sustained attention, impulsivity, and vigilance. The WCST assesses problem solving, cognitive flexibility, set shifting, and impulsivity.
Delis–Kaplan Executive Function System (D-KEFS)	8–89 years, 11 months	
Conners' Continuous Performance Test, Third Edition (CPT-3)	8+	
Wisconsin Card Sorting Test (WCST)	6 years, 5 months–89 years	
Visuospatial and Visual–Motor Tests		
Bender Visual–Motor	4–85+ years	These tests assess

Gestalt Test–Second Edition (Bender-Gestalt II) Beery Developmental Test of Visual–Motor Integration–Sixth Edition, Revised (VMI-6) Test of Visual Perceptual Skills–Third Edition (TVPS-3) Benton Judgement of Line Orientation (JOLO)	2–99 years, 11 months 4–18 years, 11 months 7–74 years	abilities in visual processing, including visual/perceptual reasoning (*what* an object is or how it is constructed), visual/spatial reasoning (*where* an object is in space), and visual/motor integration. Both the Bender and VMI involve copying geometric figures. Because the Bender allows the student to organize the items on the page and the VMI asks the student to copy each figure in its own space, the two can be used together to assess organizational issues, as well as visual–motor integration. Both the current versions of the Bender and VMI provide ways of assessing perceptual separate from motor skills. The TVPS and Benton are motor free and involve making judgments about visual information—for example, on the TVPS, a client might be asked

		to identify an item from a fragmented presentation while on the Benton a client is asked to identify which lines are in the same spatial orientation.
Auditory Processing Tests		
Test of Auditory Processing Skills–Third Edition (TAPS-3)	4–18 years, 11 months	Tests in this domain assess the ability to process auditory information. The TAPS involves making judgments about auditory information. The TARPS measures the “quality” and “quantity” of the client’s auditory thinking and reasoning.
Test of Auditory Reasoning and Processing Skills (TARPS)	5–14 years	
Memory Tests		
Wide Range Assessment of Memory and Learning–Second Edition (WRAML3)	5–90 years	Memory scales attempt systematically to assess memory skills within different spheres. The WRAML3 core battery consists of verbal, visual, and attention/concentration subtests. The CMS allows for the assessment of attention and working memory, verbal and visual memory, short-delay and long-delay memory, recall and
Children’s Memory Scale (CMS)	5–16 years	
Child and Adolescent Memory Profile (ChAMP)	5–21 years	

		recognition, and learning characteristics. The ChAMP consists of two verbal memory subtests and two visual memory subtests.
Social Cognition and Language		
Test of Problem Solving–3: Elementary (TOPS-3:E)	6–12 years	These tests assess social cognition and reasoning through language as well as indirectly assessing use and understanding of the social pragmatics of language. Elementary and adolescent clients are asked to answer questions about pictures or scenarios to reveal social reasoning.
Test of Problem Solving–2: Adolescent (TOPS-2:A)	12–17 years	
Test of Pragmatic Language–Second Edition (TOPL-2)	6–18 years, 11 months	A test of language used to achieve goals, with information about six areas (setting, audience, topic, purpose, cues, and abstraction).

Note: This table is meant to be a representative list of commonly used tests but is not a comprehensive list of every test available in each domain. Publishers are constantly updating their tests, and these editions will be superseded by newer and better tests. Even now, there may be variations of some instruments that are supplemented, expanded, or newly normed.

PSYCHOEDUCATIONAL AND NEUROPSYCHOLOGICAL ASSESSMENT

Psychoeducational and neuropsychological assessment can be looked at from different perspectives as clinicians attempt to understand referral questions and individual children. Testing instruments can be used in a more “comprehensive” (i.e., measuring a variety of cognitive domains) or more "focused” (i.e., measuring one or two cognitive domains) manner to answer the referral question. Cognitive domains that are explored in testing include intellectual functioning, achievement (academic) functioning, memory, attention, executive functioning, language, visuospatial, and sensory/motor functioning. A common question asked by psychiatrists is what type of evaluation is needed for their specific patient. This is based on the referral question (see Fig. 35.3–2) in order to optimize recommendations for treatment. Often patients are referred for a “comprehensive neuropsychological evaluation” as a kitchen-sink catchall when a focused evaluation may be more helpful in answering the referral question without unnecessary tests, time, and strain on the child and family.

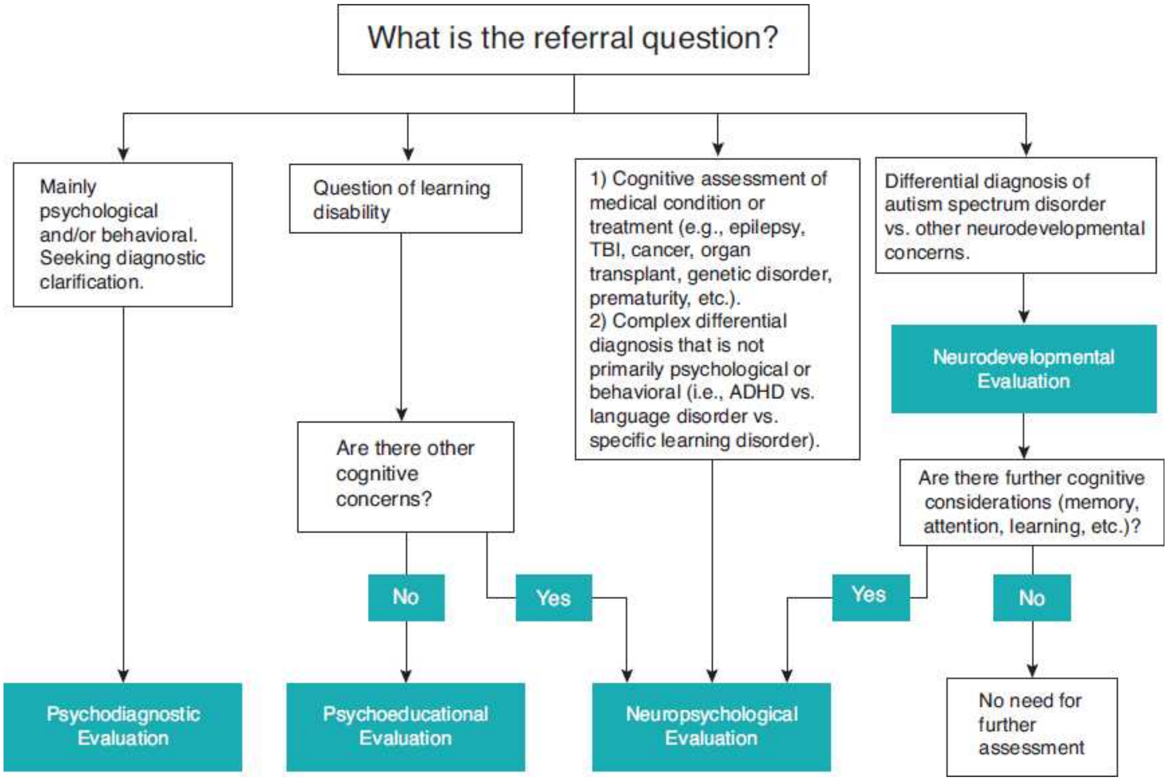
What is the referral question?
Mainly psychological and/or behavioral. Seeking diagnostic clarification.
Question of learning disability
1) Cognitive assessment of medical condition or treatment (e.g., epilepsy, TBI, cancer, organ transplant, genetic disorder, prematurity, etc.).
2) Complex differential diagnosis that is not primarily psychological or behavioral (i.e., ADHD vs. language disorder vs. specific learning disorder).
Differential diagnosis of autism spectrum disorder vs. other neurodevelopmental concerns.
Neurodevelopmental Evaluation
Are there other cognitive concerns?
Are there further cognitive considerations (memory, attention, learning, etc.)?
No
Yes
Yes
No
Psychodiagnostic Evaluation
Psychoeducational Evaluation
Neuropsychological Evaluation
No need for further assessment

FIGURE 35.3–2. Psychological assessment referral algorithm.

Psychoeducational Assessment

Psychoeducational assessment, as described here, is particularly suited for the evaluation of learning issues as they affect school functioning. Psychoeducational assessment involves the combination of intellectual, achievement, and domain-targeted tests to determine the cause of learning problems, as well as, at least in the past, to establish eligibility for services with public schools. These types of evaluations are best when the question is more focused on a specific learning issue (e.g., suspicion of reading disorder, accommodations for learning disorders on standardized tests). Psychoeducational evaluations include measures of academic function (i.e., reading, writing, math, etc.) and fluency and also include basic intellectual function measures for references. Oftentimes assessment of attention deficit hyperactivity disorder (ADHD) is included in this type of evaluation with select measures of attention, and executive function added.

Over the years, the basic definition of learning disability as “unexpected underachievement” has remained, but with new understanding and meaning. In this context, it should be emphasized that there are two different systems for identifying learning problems: currently, Specific Learning Disabilities in school settings and Specific Learning Disorders in clinical settings. For many years, learning disabilities and disorders were essentially interchangeable because of their focus on a severe or substantial (significant) discrepancy (typically one standard deviation [SD] or more) between ability and achievement. However, educational and medical definitions of learning deficits under Individuals with Disabilities Education Act (IDEA) and DSM, respectively, have changed.

In schools, the current IDEA 2004 language forbids states to present a severe discrepancy model to local educational agencies as the only option and indicates they must allow for other possibilities. Underachievement may be identified in several areas (focusing on reading, mathematics, and written expression). Basically, this can be based on failure to respond to research-based intervention as other students have or the identification of a pattern of weaknesses in the context of strengths that indicates a specific learning disability.

In clinics, the new definition provided by the DSM-5 focuses on an absolute academic deficit and stipulates two basic requirements. The first requirement is that the learning difficulties have persisted for at least 6 months “despite the provision of interventions that target those difficulties.” The second requirement is that academic skills (in reading, written expression, or mathematics) should be “substantially and quantifiably below those expected for the individual’s chronological age” and cause significant interference with related activities, “as confirmed by individually administered standard achievement measures and comprehensive clinical assessment.” The new definition also requires two additional factors: that the learning difficulties begin in the school-age years, although they might not become manifest until later, and that the learning difficulties are not better explained by other issues, such as intellectual disability, other disorders, or inadequate instruction. The exact deficit level associated with the learning disorder might be left to the clinician’s judgment, but the threshold is thought to be between 1 and 2.5 SD below the population mean (100), depending on the circumstances of the case. A 1 SD standard would fall at or below 85 (16th percentile); this might be sufficient for diagnosis “with converging evidence.” A 1.5 SD standard would fall at or below 78 (7th percentile); this level is thought to provide “more diagnostic certainty.” Specific learning disorder is an umbrella term, but impairments can be specified in the three academic areas and can involve efficiency as well as accuracy. The severity of the disorder can be specified from mild to severe. The clinical standard for learning disorder might be seen as stricter (if not clearer) than the school standard for learning disability.

For learning problems, educational research has suggested just how complicated the identification of learning difficulties is, while recognizing the need to define them in order to provide services to children, particularly in school. Descriptions and definitions of learning deficits have recognized disparities in different aspects of functioning and tried to parse out factors intrinsic to the child and specific to learning from aspects of their environment and experiences. Procedural definitions of learning disorders under the DSM in the medical community and of learning disabilities under IDEA in the public schools have tried to grapple with this complexity. As a practical matter, both have generally focused on the concept of “unexpected underachievement.”

When it comes to learning disability, what is at stake is the opportunity for a patient to obtain services through school at no cost to the family. Otherwise, it falls to the family to obtain private educational therapy in the community. In schools, students might obtain targeted classroom instruction in areas of weakness, or some intervention through a student study team, or accommodations or modifications in class through informal

504 procedures. However, students are apt to receive more extensive and expensive services under IDEA, through more formal Individualized Education Program (IEP) procedures that identify and make recommendations around disabilities, including specific learning disability.

Neuropsychological Assessment

Neuropsychology is dedicated to the study of brain–behavior relations and has matured into a clinical discipline for the diagnosis and characterization of brain function and dysfunction. This type of assessment is provided by a clinical neuropsychologist with advanced training and expertise in neuropsychology.

Assessment of Functioning. Neuropsychological assessment was originally developed for the assessment of adult patients and was not generally applied to child assessment until later. When it was used, it was typically to assess children's functioning subsequent to acquired injuries or diseases affecting the brain. This functional assessment was important because the effect of trauma on the brain is highly variable across individuals, even when the precise location and size of the lesion are known. In these situations, neuropsychological tests could provide specific functional information that took the child's age and developmental status into account. This continues to be a prominent point of referral for pediatric neuropsychology because it discusses not just the existence of the brain disruption but its *meaning* in terms of the child's ability to function.

Developments in Test Instruments. Advances in pediatric neuropsychology include the introduction of tests specifically designed for use with children. These instruments assess similar behaviors as do their adult counterparts but use paradigms that are more engaging to children and better measure developmental transitions throughout childhood. These instruments are used in clinical assessments but are now also part of many research protocols examining childhood diseases and genetic conditions. Given its precision in measuring behavior, neuropsychological assessment is now involved, not just in the assessment of function after an injury but also in the initial diagnostic processes. Another example of this specificity is the now-routine use of neuropsychological testing in genetic research of developmental disorders of childhood, given the precision that it adds to questions of endophenotype expression.

Applications of Neuropsychological Assessment. The functional and diagnostic assessment of children and adolescents often begins (and ends) with psychoeducational evaluations. There are situations, however, in which the use of educational, intellectual, or psychological testing alone is not able to clarify the diagnosis and determine the most appropriate treatment plan. In those cases, the psychiatrist should consider additional neuropsychological testing. To help treating psychiatrists to know more about when to refer a patient for neuropsychological testing, the following list offers specific information about when this approach to testing might be helpful.

1. *Assessing brain function*: Neuropsychological tests are designed with the expressed intent of assessing brain function. Thus, if patients have a history of traumatic brain injury, neurologic condition (e.g., epilepsy, brain tumor, genetic disorder), or history of medical treatment with known CNS effects (e.g., chemotherapy), neuropsychological assessment would be the testing of choice. In these cases, identification of impacted neuropsychological circuits might contribute to diagnostic clarity and treatment planning (e.g., whether hippocampal vs. frontal circuits are contributing to memory impairment may aid in how to approach learning with a child).
2. *Complex diagnostic picture with clear cognitive issues*: Children presenting with complex diagnostic issues may benefit from neuropsychological evaluation if the psychodiagnostic evaluation does not elucidate a clear picture. For example, a child who presents with symptoms of ADHD and oppositional defiant disorder but who also has a history of language difficulties with speech therapy and poor performance in school may warrant a referral for neuropsychological testing to examine if attentional and oppositional issues are stand-alone or related to ongoing language and/or learning disorder.
3. *Breaking down cognitive functions to better assist with intervention*: Neuropsychological tests are valuable when a systematic assessment of a functional hierarchy might allow for a complex skill to be broken down into its component parts. For example, if a child works slowly on a reading fluency task, is the cause their reading problems or slow processing speed? Many neuropsychological tests employ a functional

hierarchy of demands that allow several basic aspects of a complex skill to be measured independently to reduce confounding variables and more confidently examine the specific behavior initially intended.

In addition, the neuropsychological battery contains many tests that appear to be very simple but are included because of the narrow range of abilities that are required for them to be done well. Examples of these are the examination of sensory skills, such as auditory processing in the right versus left ear, lateralized tests of motor development, such as grip strength or pegboard placing, and the administration of timed, simple tests across several skills to highlight selected low scores that fit into known patterns (such as the language-based weaknesses suggested by compromised rapid color naming alongside weak math automaticity).

4. *Very high or low intellectual level*: Another application of neuropsychological testing occurs in instances where the child is at the higher or lower end of the intellectual range where cognitive testing does not demonstrate a clear score profile. Bright children with learning disorders have many unaffected domains of strength that can be used to compensate on “multiple construct” tests to support the more discrete impairment. Although this will usually result in relatively lower-than-expected scores, in very gifted students, the bottom of the range of such scores may be average to high average, making interpretation difficult. It is important to note, however, that the ability to apply a clever strategy to a test does not mean that the student can likewise use intellect to manage their symptoms in the larger contexts of life (which is the reason the child presented for assessment in the first place). Rather, the child’s clever approach to testing means that there is a greater risk that all scores will cluster together at an above-average point and not demonstrate discrepancies that suggest learning problems. In these cases, students are often not diagnosed at all or until much later, after life’s demands have caused considerable disruption to their academic, social, and/or emotional success.

   Conversely, when a child’s intellectual functioning is very compromised, tests that do not include the functional hierarchies discussed earlier make it more likely that all scores will be low. This is a concern, given the importance of identifying relative cognitive strengths in these children, because these islands of ability can be quite amenable to intervention and so will demonstrate more improvement with treatment than the other, very impaired skills. When these children are young, they should receive intervention for both their areas of profound weakness and their relative strengths, but in adulthood, it will likely be the isolated strengths (which, with specific early intervention, will develop on a higher trajectory than those that are more impaired) that provide the foundations for future participation in aspects of adult life.

5. *Retest to check for treatment effects*: A last benefit of neuropsychological tests that may be important to a psychiatrist is that many of these instruments have been constructed so as to control for "practice effects" (an artificial increase of a score resulting because a test was given twice with too little time between the administrations). When practice effects are controlled, tests can be readministered at short intervals, providing an effective way of assessing treatments such as medication that should show rapid improvement in cognitive abilities. Although this approach needs to be undertaken cautiously (and results interpreted in light of slight improvements seen in nonclinical populations, as are known to occur with some medications), significant changes measured with retesting can add confidence to clinical observations obtained informally or with questionnaires.

   One way of managing practice effects is to have two alternate forms of the same test, which allows them to be readministered after only short intervals. Many subtests from the Delis–Kaplan Executive Function System (D-KEFS) provide these alternate forms, making them a good choice when disorders characterized by executive functioning deficits are part of the referral question. A second way of managing practice effects is with tests that are intentionally boring, such as those measuring sustained attention (e.g., the Connors Continuous Performance Test [CCPT]). When, after a short interval, children with normal attention retake the CCPT, they typically perform *worse* than they did the first time (contrary to what practice effects would suggest). The decrease in scores is attributed to the boring nature of this long, computerized test; the first time the child takes it, there is some excitement about its novelty, but by the second time, the child knows what they are in for from the start, which affects the child's motivation. If a child does significantly better on the second administration, the change is attributed to something other than practice; if done in the course of a medication trial for ADHD, the assumption is that the medicine is helping performance more than the practice is hurting it.

There are several ways that retesting can be used to aid a psychiatrist in quantifying medication effects. For example, when a child who is already on medication for ADHD presents for testing, a 45-minute battery that measures behaviors that are sensitive to the medicine can be administered one morning without medication and a second morning with it. Alternatively, if at the close of testing ADHD is a diagnostic impression and the treating physician thinks a medication trial is appropriate, the child can return for brief reassessment with sensitive tests once the medicine is thought to be at therapeutic levels. Information obtained from retesting can provide data about the benefit of the medication and the appropriateness of the dosing, as well as increasing parents' and patients' confidence about the intervention.

COGNITIVE DOMAINS OF ASSESSMENT

Although psychologists use diverse tests, there is a selection of domains that are measured during the assessment. Table 35.3–1 offers a sampling of tests that psychologists often utilize in neuropsychological or psychoeducational assessments. It should be noted, however, that all of these domains overlap to some degree. The role of the psychologist is to examine all of the data (cognitive, emotional, behavioral, and history) together and look for patterns while isolating areas of dysfunction in order to formulate a conclusion about diagnosis and provide treatment recommendations. A full range of cognitive domains, instruments, and their uses are described in this section.

Intelligence

Intellectual testing is often part of a variety of assessment batteries, including psychoeducational and neuropsychological evaluation, along with more general developmental and clinical evaluations. It is usually undertaken to determine the patient's general intellectual level of functioning. The IQ is valued because of its stability over time in the general population, as well as for the protective aspect of intelligence in a range of potentially disabling situations. It should be noted that, although intellectual level can be measured, it is hard actually to define intelligence.

Definition. Different theorists have had their own take on intelligence. Although, theoretically, Sattler saw intelligence as “multifaceted and hierarchically organized,” practically, there is no agreement about what intelligence is, and this is reflected in the different ways in which it is measured. However, a commonsense definition of intelligence would include the ability to learn from and adapt to the environment as well as the ability to think abstractly and to recognize patterns in diverse situations (on tests and in life). Although intelligence tests yield one IQ score (or several IQ or index scores), they are, in fact, devices for “sampling” many tasks in a variety of verbal and nonverbal areas.

Stability. Although there is some debate, IQ scores tend to be relatively stable starting as young as 5 to 7 years of age. In general, the older the child is when tested and the smaller the interval between test administrations, the greater is the correlation between two IQ scores. Although using an IQ score can be useful as a way of assessing the client’s basic trajectory through life, the prudent practitioner must be aware that there are a number of factors that can affect intellectual functioning and, thus, IQ scores. These factors can include psychiatric illness, lack of motivation, and inattention. In some cases, treatments and interventions implemented to improve cognition and development can improve functioning and, thus, IQ scores.

Protection. Despite conceptual and practical complications, high intelligence is associated with better prognosis in a wide range of psychiatric conditions; lower rates for behavior, conduct, and emotional problems in children; and lower rates of referral for psychiatric problems in adults. In the case of any kind of brain damage (neuronal death), intellectual level accounts for a great deal of variance in predicting overall cognitive and functional outcome, with lower IQs associated with poorer outcomes and higher IQs associated with better outcomes in general.

Assessment. Although IQ is what is obtained with an IQ test, there are a variety of intellectual tests, as well as other ways of calculating intellectual level. There are a number of instruments from which to choose, and the psychologist must make their selection based on the specific characteristics of each test (e.g., normative sample and construction of the instrument) as they relate to the characteristics of the client (e.g., age and referral question). Once the test has been administered, the clinician must make interpretations based on the analysis of overall and subtest scores and their pattern in the context of the diagnostic process. Intellectual tests can be supplemented by a variety of instruments to further characterize any patient's presentation, in order to contribute to their understanding and diagnosis, and to help identify appropriate interventions.

Sometimes, there are misinterpretations of scores at the bottom of the intellectual range. Distinctions must be made between level designations and diagnostic categories. The DSM-5 utilizes the term intellectual disability to describe those individuals who score within the lower end of the intellectual range. The term intellectual disability involves impairments in mental abilities that impact adaptive functioning in three domains (conceptual, social, practical). Severity of the disability is determined by adaptive level, not intellectual score, because it is those impairments that determine the level of supports needed. However, it is important that measurement of intellectual level still be part of any individual's assessment, along with developmental course and adaptive deficits. Under DSM-5, intellectual disability is associated with intellectual levels two SD or more below the population mean, or an IQ score of about 70 or below. In making determinations about low intellectual level, the psychologist must look at the pattern of strengths and weaknesses not only within the instrument, but also within the overall case presentation. For example, some IQ scores can be brought down by specific language disorders or perceptual–motor difficulties.

COMPREHENSIVE INTELLECTUAL TESTS. The two best-known intellectual tests are probably the Wechsler Intelligence Scales and the Stanford–Binet Intelligence Scales (SB). The current editions of both are divided into separate subtests, and the data are analyzed in separate spheres. Each test has an overall mean of 100 (50th percentile) and a SD of 15. Subtests have a mean of 10 and a SD of 3. In the case of the Wechsler tests, three separate instruments (Wechsler Preschool and Primary Scale of Intelligence [WPPSI], Wechsler Intelligence Scale for Children [WISC], and Wechsler Adult Intelligence Scale [WAIS]) are designed for three different age groups over the lifespan. For the SB, one instrument covers a lifetime. All instruments include full-scale IQ scores, as well as (broadly speaking) verbal and nonverbal composite scores. The SB provides an extended score range (at the top and bottom of the scale) that may make it more useful for individuals within the “gifted” and “impaired” ranges. Both the SB and the Wechsler tests have been the “standard” individual intellectual tests for the school-age population. They allow for the assessment of overall functioning and verbal and nonverbal reasoning. These instruments have taken care to relate to achievement tests to facilitate the recognition of learning disabilities according to a “traditional discrepancy model” in which estimated ability is compared to academic achievement. Intellectual testing is commonly a part of psychological evaluation as it provides a way of understanding a patient’s overall level of functioning in school and in life.

OTHER INTELLECTUAL TESTS. Sometimes, specific characteristics of clients, such as age, suggest the use of specific instruments.

1. *Infant and child tests*. Although very young clients can be “tested,” these administrations are often as useful for their observations as their scores. Because of the emphasis on sensorimotor functioning and development of language skills of very young clients, the instruments (such as the Bayley Scales of Infant and Toddler Development) are not as useful for predicting later IQ scores as they are for identifying individuals who are or are not, at risk. The instruments do not become predictive in a more specific way until children are of school age.
2. *Nonverbal and language-free tests.* Although some tests seem especially suited for individuals who have difficulty in hearing or have language issues (such as the Leiter International Performance Scale), the nonverbal tests can have broader applications. Their emphasis is on cognitive problem solving based on visual information. In fact, the Wechsler Nonverbal Scale of Ability claims to be a test of general ability using nonverbal tasks, not a measure of “nonverbal intelligence.” Although these nonverbal tests are useful, it is important to remember that they may not provide the whole picture, and the significance of these tests for the client’s ability and functioning must be carefully assessed.
3. *Quick tests.* The commonly used tests in research and clinical practice are the Kaufman Brief Intelligence Test (KBIT) and the Wechsler Abbreviated Scale of Intelligence (WASI). Both the KBIT and the WASI include verbal and nonverbal tasks to assess crystallized and fluid thinking. The WASI and KBIT2 have been able to provide evidence of good correlations with established Wechsler tests. The available data

suggest that the quick tests, in fact, yield very representative overall IQ scores. Although a clinician might not get all of the processing information included in a full intellectual test, these correlation coefficients suggest that the quick tests can provide a very good estimate of the client's general level of functioning.

General Comment. In discussing intellectual testing, it is important to note what IQ tests are and what they are not. Although IQ scores are commonly used as estimates of ability, they are not measures of "native" or "raw" intelligence. Every client brings a long learning history, including opportunities, to the testing situation. Although some instruments (such as those that emphasize nonverbal functioning) may seem less biased or loaded, it must be acknowledged that probably no intellectual test is either culture free or fair. Nonetheless, because of the way in which tests have been constructed and normed, caution has to be exercised in making conclusions in any case with a diverse cultural/linguistic background or unique medical situation. Although IQ scores are useful as estimates of ability, they are, in fact, measures of functioning. There are a variety of transient factors that can be associated with alterations in IQ scores, including problems with involvement, attention, and disorganization. As a result, the clinician has to be prepared for changes in intellectual level based on changes in medical status. For example, in some cases, clinicians have observed increases in IQ scores subsequent to medical treatment for ADHD, partly due to increased efficiency on the instrument itself and partly due to increased learning over time in the environment and in school. Further, clinicians need to be aware of what IQ scores do and do not measure. They do not measure social intelligence. Nor do they reflect the creativity that artists of different kinds bring to their art and that entertains and informs people daily.

Considerations. Here are a number of factors professionals should keep in mind when they deal with intellectual test results.

1. The intellectual instruments have developed classification systems according to intellectual level. The Wechsler system (based on scores with a mean of 100 and an SD of 15) utilizes the following classifications for standard scores: exceptionally low (≤69), below average (70 to 79), low average (80 to 89), average (90 to 109), high average (110 to 119), above average (120 to 129), and exceptionally

high (≥130). The Stanford–Binet test uses the same classification system for low average to high average scores (80 to 119). However, the classification system has a wider range including scores from 40 to 160, from moderately impaired or delayed to very gifted or highly advanced.

2. The more variability seen on intellectual testing, the less do composite or summary scores have utility. For example, on the WISC or SB, when the verbal and nonverbal scores (or other contributing indices or composites) differ significantly, the full-scale IQ is not informative. Similarly, when subtest scores differ significantly within a sphere, the emphasis should be on specific processing issues rather than overall functioning.
3. Sometimes a practitioner will ask what a patient’s IQ is. It is clear from the multiplicity of the tests and scores available that there is no one IQ. Although IQ is what the client gets on an IQ test, it is prudent to be well aware of which test was administered and which IQ is being quoted. The practitioner should be aware of disparities on different tests or in different areas; this pattern is often important in making decisions for a client.
4. In the context of a psychoeducational assessment, when scores differ significantly on intellectual testing, the full-scale IQ may not be the best estimate of “ability,” and the clinician will make attempts to choose among the possible summary scores on a range of instruments to identify the one that best represents the student’s learning potential.

Further Assessment. The following is a list of patterns that suggest the need for further evaluation.

1. *There are significant discrepancies between verbal and nonverbal scores*. When verbal scores are significantly lower than nonverbal scores, the possibility of a language disorder should be considered. In the case of low verbal scores, a speech and language assessment, including measures of receptive and expressive language, as well as pragmatics (the factors affecting successful communication), might be indicated. It is particularly important to identify language-based learning disorders because of their potential effect on school (and relating) and because there are some very specific treatments that might be helpful. When nonverbal scores are significantly lower than verbal scores, further assessment regarding a variety of possibilities must be considered. More than the verbal tasks, the performance

tasks (which might be timed) are vulnerable to ongoing and transient factors mediated by attention. In the case of particularly large discrepancies, the possibility of nonverbal learning disorder should also be considered. This is a pattern of cognitive assets and deficits that can be associated with peer, as well as academic, problems. The identification of a nonverbal learning disorder can be very helpful because it can lead to a wide variety of interventions, including those around social skills.

2. *The full or composite score is above the 98th or below the 2nd percentile (130 or 70)*. At the intellectual extremes, the examiner may consider whether further assessment regarding educational or adaptive needs might be in order. When a client has a very high intellectual level (e.g., moderately to profoundly gifted), the professional might refer the family to a psychologist specializing in very bright children. In the case of individuals with very low intellectual levels, further assessment can be obtained through the school system or the local public agency serving developmentally delayed clients. Professionals and parents may access the websites http://www.aaidd.org, http://www.thearc.org, and http://www.eparent.com for information about services and products associated with intellectual and developmental disabilities.
3. *The intellectual score shifts downward in a significant way*. Because it is anticipated that intellectual scores will remain relatively stable over time, the mental health professional is advised to look for an explanation, either psychological or medical, for dramatic score changes. This is a time when the psychologist should look at raw, as well as standard, scores. After some exploration, consideration of medical issues, as well as consultation with further medical personnel such as the pediatrician or neurologist, or further psychological professionals, such as a neuropsychologist, might be in order.

Achievement Testing

Achievement tests systematically assess past academic learning. Clinicians understand the results are affected by a wide variety of educational opportunities and impacted by a large variety of comorbid conditions. Once disadvantageous conditions such as sensory disorder, intellectual disability, and lack of appropriate/adequate instruction have been ruled out as primary causes of academic difficulties, the academic shortfall has commonly been attributed to processing problems. These can include everything from discrete cognitive functions, such as phonologic processing, to more general mental constructs, such as executive functioning. In fact, good basic processing, the ability to successfully use a wide variety of cognitive strategies in a wide variety of tasks, has commonly been thought to support intellectual level and academic success.

When achievement testing is undertaken along with intellectual and domain-targeted testing, the overall evaluation is commonly referred to as a psychoeducational assessment. The purpose of the assessment is to identify learning problems and usually to rule out other psychological factors that might be complicating learning. In addition, its purpose has been to establish eligibility for services through the public school system under learning disabilities. Clinicians communicating with school systems will need to be aware of the state and local guidelines for establishing learning disabilities and, for example, how/whether/when the kinds of standard testing described here might be relevant under new identification models. However, practitioners will still value standardized (individual) achievement testing as an objective (normative) view of how the child stands nationally and for the ability to identify academic strengths and weaknesses. Of course, other information will be important—for example, the child’s grades, academic interventions, and behavioral information.

Background. Achievement testing is administered to determine a student’s level of functioning in basic academic areas, including reading, mathematics, and writing. Unlike intellectual testing, achievement testing is not necessarily expected to be stable over time because it measures the child’s success in formal learning and is highly dependent on the home environment and the school curriculum.

Assessment. Psychologists commonly begin an assessment of academic achievement by giving a comprehensive test in order to get a sense of a client’s areas of weakness in contrast to their strengths in reading, math, and writing. These findings are compared to school records. Because reading problems are a relatively common reason for referral and their causes are increasingly understood, where indicated, this comprehensive test is commonly followed by other tests of reading skills that separate out accuracy, fluency, and comprehension.

It needs to be noted that while specific learning disabilities (in schools) and specific learning disorders (in clinics) are commonly thought of in terms of underachievement in a specific academic area, learning difficulties can be more diffuse. They might not impact a particular subject in a significant way, but might impact a range of subjects in a moderate way. Or they might impact overall school performance in a drastic way. Under DSM-5, Specific Learning Disorder might be identified with impairments in reading, written expression, or mathematics. As part of the reconceptualization and reorganization of the diagnostic manual around neurodevelopmental disorders, learning problems associated with processing problems impacting general adaptive and social functioning rather than subject areas might be identified as other specified (or unspecified) neurodevelopmental disorder.

COMPREHENSIVE ACHIEVEMENT TESTS. Each of the tests assesses a range of academic areas so that they can be compared to each other within any individual or to academic achievement according to more external standards (age/grade expectations). The Wechsler Individual Achievement Test (WIAT) and the Woodcock–Johnson Tests of Achievement (WJ-ACH) allow for the systematic assessment of reading (basic skills, fluency, and comprehension), mathematics (calculation, fluency, and reasoning), and writing (spelling and brief to extensive composition), as well as other academic spheres. Both instruments are relevant for identifying specific learning disorders in reading, written expression, or mathematics. The WIAT is a companion test to the WISC. The WJ-ACH is a companion instrument to the Woodcock–Johnson Tests of Cognitive Abilities (WJ-COG), which evaluates processing skills and includes measures of intellectual ability, and the Woodcock–Johnson Tests of Oral Language (WJ-OL), which assesses specific oral language functions. As individual achievement tests, both the WIAT and WJ-ACH are widely used in the community and schools. An additional, comprehensive test is the Kaufman Test of Educational Achievement (KTEA), which is co-normed with the Kaufman Assessment Battery for Children (KABC).

OTHER ACHIEVEMENT TESTS. These and additional achievement tests are also available and can involve general or specific assessments.

1. *General achievement tests.* Some instruments focus on screening, such as the basic Wide Range Achievement Test. However, the fourth edition allows for the assessment of sentence comprehension, as well

as reading, spelling, and math calculations. An expanded version provides even more information. There are other tests that may assess a range of academic areas in more depth.

2. *Specific reading tests.* Because of the importance of reading in school and in life, particular effort has gone into systematically assessing reading skills and the ability to make symbolic associations between sights and sounds, as well as the ability to decode words and understand text. In addition to reading subtests on major achievement instruments, there are specific reading tests, such as the Gray Oral Reading Tests (GORT) and the Woodcock Reading Mastery Tests (WRMT). In addition, because of the importance of phonologic awareness in reading, tests of phonologic strengths and weaknesses, such as the Comprehensive Test of Phonological Processing (CTOPP), are also often given. Since achievement tests assess academic functioning in slightly different ways, psychologists often administer a combination of achievement tests to any one client. For example, reading comprehension might be assessed by asking a student to read a passage and supply a missing word or by asking a student to answer questions about a paragraph. How the student does in different contexts may help to explain the nature of their academic problems.

Considerations. Here are several issues that mental health professionals should keep in mind in assessing achievement results:

1. Although estimates vary greatly, it can be realistic to conservatively estimate that between 15% and 30% of children with ADHD have a learning disability, and between 20% and 30% of children with a learning disability also have ADHD. As a result, when a clinician finds one condition, they should look for the other. Although they can co-occur, ADHD and learning disabilities are separate conditions requiring separate and specific recommendations.
2. Most individuals with a learning disability have problems with reading (between 70% and 85%). Students who have learning difficulties in other academic areas commonly also have reading problems. Isolated learning disabilities in mathematics are much less frequent (perhaps 10% to 15%).
3. Although achievement tests commonly note grade- and age-equivalent scores, as well as standard scores, for a variety of reasons, the standard scores are more reliable than the age or grade equivalents.

The concept of grade equivalent can be misleading. For example, it does not mean that a student functioning at the 6.5 level while in the fourth grade could perform sixth-grade work. The score merely indicates that the typical sixth grader would have received the same score had sixth graders taken the fourth-grade test. This same statement would also be true of age-equivalent scores.

4. Although all schools must provide services to individuals identified as disabled, there are other populations who, despite being at an advantage, may still need help but are not protected by federal laws and regulations. Because giftedness is not a disability, it is not defined and covered under IDEA. There is, however, a federal definition of “gifted and talented” referring to high-performance/achievement capability students who need services or activities “not ordinarily provided,” and the federal government encourages their provision through state enrichment (Gifted and Talented Education [GATE]) programs.

Further Assessment. In practice, when any academic problem is extreme or does not respond to appropriate intervention, further evaluation of the problem is in order. Reading problems, in particular, may need very careful evaluation regarding the basic processes that might underlie the difficulty. A language specialist who deals with reading issues or a neuropsychologist who might shed light on the pathways available or impaired in a student can perform this assessment.

Conclusions

A wide variety of intellectual and achievement tests have been explored as they apply to clarifying cognitive functioning in general and describing competencies in the academic and social spheres. There has been some discussion of learning disabilities and disorders and some attempts to provide information so that psychiatrists might help patients receive services through the school as well as in the community. However, the same processing issues that affect academic skills acquisition can affect behavioral control and social appropriateness. In addition to cognitive and neuropsychological testing, more projective testing might be undertaken to help identify the organizational issues that contribute to school or social problems, along with the more emotional factors that might interact with them. A battery of psychological tests provides a fairly systematic assessment of different kinds of input and output as they are affected by more and less structure.

Descriptions and Assessment of Typical Neuropsychological Domains

Typically, in addition to assessing IQ, academic achievement, and social and emotional functioning, neuropsychologists assess domains of memory, attention, executive functioning, language, visual perception, and sensory–motor development. Tests have been developed to examine specific aspects of these domains in isolation so as to increase diagnostic clarity. Although these domains are discussed as different constructs in this subsection, in truth, they overlap with one another in many different ways. For example, the term *working memory* is often conceptualized as being an aspect of attention, as well as a necessary component of good planning (which is part of executive functioning). It is also a component of memory in that when it is not well developed, and it leads to the phenomenon of "forgetfulness." Many of the measures used to discuss one of these constructs will also be used to discuss others. In addition, neural substrates and pathways supporting their functions are often overlapping or, at least, parallel systems.

Thus, the interdependence of these domains is an assumption of this portion of this section. Each subsection provides an overview of the theory supporting the conceptualization of the domains, followed by a description of the ways in which a neuropsychologist might assess them.

Memory

OVERVIEW. To memorize something, a person must first pay attention to it. Assuming that this attention is intact, the next step is to "make" the memory or *encode* it. Once the memory has been made, however, it must be available on demand to be of use. Knowing the name of France's first president is of no benefit if you cannot *retrieve* it during the history examination.

One metaphor for this two-step memory process is a filing cabinet. *Encoding*, then, is when a person puts information into the "filing drawer." Someone with a true amnesic disorder (such as Alzheimer disease) never gets the information into the drawer. No amount of cueing or reminders later will help the person to recall the information because it never "got into the drawer" in the first place. This type of impairment can be seen in some children, most often those with seizure disorders that adversely affect the temporal lobes. For most children, however, the problem described as "poor memory" is actually a difficulty with retrieval. *Retrieval* is the ability to get information out of the "file drawer" once it has been put in. Poor retrieval is associated with problems of organization (the folders are missing labels) and is more often the issue when children are described as being "forgetful."

To differentiate between encoding and retrieval, children are asked to memorize material and then, 20 to 30 minutes later, recall it. If they are unable to remember it spontaneously, the examiner does not know whether they have not encoded it or are having problems with retrieval. If the child can remember the material with cueing (e.g., "In the story I read you, was the boy's name Johnny or Sam?"), retrieval is implicated. For the child who cannot encode, however, cueing will not help.

ASSESSMENT. In assessing memory, several guidelines should be followed. Both visual and verbal memory tasks should be given. Visual memory tasks (such as learning the location of dots or memorizing faces) are usually aided by the right hemisphere. In most people, verbal memory tasks (such as memorizing a shopping list or a story) are supported by the left hemisphere. In addition, the material to be memorized should include rote tasks (such as word lists), as well as material that is presented in context (such as stories). Some memory tasks assess learning or the child's ability to benefit from several presentations of the material. It is expected that, after three exposures to a picture of dots, the child's memory of it will be stronger than it was after the first exposure. If not, encoding may be implicated. A 20- to 30-minute delay should also be part of the memory assessment, and cues should be available to differentiate between encoding and retrieval difficulties.

Other terms in the neuropsychological literature appear to describe memory but are actually probably better classified as part of the attention system. These include *short-term memory* and *working memory*. These terms are discussed in the following subsection on attention.

Attention

OVERVIEW. The attention literature is large and includes many different conceptualizations. The following illustration demonstrates some elements of good attention.

Suppose you arrive at a lecture hall, open your notebook, and rather than scanning the room indiscriminately, turn your attention to the instructor, who is just beginning to speak (*selective attention*). The lecture is mildly interesting, and you are able to pay attention for the full 20-minute presentation (*sustained attention* or *vigilance*). At the same time that you are listening to the instructor, you are taking handwritten notes incorporating headings and subheadings. It appears that you are able simultaneously to listen, write, and organize rather effortlessly, although you are probably shifting your attention among these competing tasks (*divided attention*). A fire engine goes by the lecture hall, and you look up (*distraction*) but are then able to ignore the dimming noise of the siren (*inhibition*) and continue to listen to the lecture (again, *sustained attention*). Suddenly, the fire alarm rings, and you smell smoke. These distracters capture your full attention (*disengagement* from the lecture), and their importance causes you to change your attention and behavior (*set shifting*) as you hurriedly head for the door.

A breakdown in any one of these areas can lead to a breakdown in attention. Some children are so caught up in their private thoughts that they have difficulty disengaging from their inner musings, not realizing, for instance, that their babysitter is asking them a question. Or, once they are involved in a task, they have trouble switching to something new, causing them to “get stuck” or perseverate.

The attention literature also discusses divided attention. This is the ability to pay attention to two things at the same time, although, in fact, a shifting of attention between two things is probably occurring. Good divided attention is important in multitasking. Some authors use the terms *divided attention* and *working memory* synonymously.

ASSESSMENT. Assessment of attention requires a number of approaches. Children with attention problems exhibit them at home and at school whenever a task becomes less interesting to them. They function better when working one on one with a person or when working on a new activity because it is more stimulating. For this reason, the testing environment may not elicit inattentive behavior (especially on the first day). To assess the child’s attention “in real life” and across settings, attention questionnaires should be completed by both parents and teachers. Many researchers consider this aspect of the assessment of attention to be the most important. Some neuropsychological measures have been found to be sensitive to attention as well. Computerized measures of sustained attention that are designed to be long and boring can capture the loss of attention described here. In addition, specific kinds of performance patterns on these measures have been shown to differentiate different types of attention problems.

Assessment of *verbal* short-term memory might include the repetition of digits or of short sentences. Assessment of *visual* short-term memory can be achieved by having the child point to dots or circles on the page in the same order in which the examiner has just pointed to them. Working memory is usually assessed as the second part of a short-term memory test. It requires that the material that has been stored in short-term memory be manipulated in some way. Verbal working memory can be assessed by having the child repeat digits backward or by doing mathematics in their head. Saying the months of the year backward can also assess verbal working memory (as long as the child is able to give them in their usual order without difficulty). Having the child point to the dots on the page in the reverse order in which they are shown can assess visual–spatial working memory.

Executive Functioning

OVERVIEW. Executive functioning could be considered the mature product of good attention. Although not developed in earnest until children reach adolescence, many aspects of executive functioning begin to appear when children are younger and, thus, can be measured. *Executive functioning* refers to the person’s ability to organize their behaviors to perform a specific goal. Good executive functioning allows a person to identify problems, generate solutions, choose among them, follow through on the chosen strategy, and evaluate its effectiveness as the work progresses. Without good executive functioning, children who are bright have difficulty demonstrating their abilities. Their parents often report school underachievement that cannot be explained by learning problems. The issue is not about “knowledge,” but rather the application of that knowledge to everyday functioning.

ASSESSMENT. Assessment of executive functioning requires several tests, given its many facets. Good attention and working memory, already discussed, are crucial to goal-directed behavior. Inhibition can be tested by giving the child a task in which they must control an automatic response. For example, on tests using the Stroop paradigm, the child is shown a color word printed in ink of a different color (e.g., the word “BLUE” written in red ink). The automatic response is to just read the word “BLUE,” but this task requires that the child, instead, name the color of the ink. After several items, inhibition appears to tire, and children with poor impulse control perform worse than their peers.

Fluency can be assessed by having the child generate category words under a time limit. For example, a child might be asked to name as many kinds of toys as they can in 1 minute. A variant of this task requires the child to create as many designs as they can in a 1-minute period, according to strict guidelines.

Cognitive flexibility is often tested with the Wisconsin Card Sorting Test (WCST), a measure of problem solving. On this test, the child is not told how to solve the puzzles; rather, they must use feedback that their attempts are “right” or “wrong” and are then expected to use this information to generate strategies. During the course of this test, the rules often change without warning, requiring that the child “regroup” and develop a new strategy. This measure generates information about the child’s ability to initially figure out the task, their tendency to perseverate on wrong responses, and their ability to use feedback to generate new responses.

Planning is another aspect of executive functioning. Variants of a “tower” test are often used to assess this ability. On a tower test, the child is shown a picture with colored balls or disks stacked on top of one another on wooden pegs in a specific configuration. The child is told to move the balls or disks on the pegs on an actual model on the table to match the configuration shown in the picture. The child is instructed to move only one ball or disk at a time and to use as few moves as possible. To perform the task well, the child must first “hold back” and not make impulsive moves that may get them “cornered.” The child must also visualize the first few steps of the problem. Thus, both impulse control and visual working memory are required to exhibit good planning on this rather entertaining test.

Language

OVERVIEW. Human language organizes, supports, and communicates knowledge, memories, and ideas. Beyond just allowing people to communicate with others, language organizes both thoughts and emotions, as well as helps people sequence their actions. Although traditionally discussed in terms of left hemisphere functioning, much of the human cortex is involved in various aspects of language. Communication includes both *speech*, the rapid and complex motor movements involved in talking, and *language*, the code used to express thoughts and ideas.

In considering language, perhaps the most common distinction made is between *expressive* and *receptive* language. Expressive language requires the production of language, including articulating clearly, finding the right word, and applying grammar and syntax to one's ideas, in addition to vocal fluency and voice tone (prosody). Receptive language involves the ability to comprehend and remember what is said.

Children with expressive language problems may appear to have little to say and are considered to be shy. In fact, however, their difficulty may be with self-expression. Some children who are very talkative ("fluent") may also have difficulty with finding the word they want or organizing their sentences to make them clear. The paradox of a fluent child with an expressive language disorder may cause their problems to be overlooked.

Receptive language, or the ability to understand what is being said, represents another aspect of the language system. Children with poor receptive language have difficulty in processing information that is spoken to them and may have difficulty learning in the classroom or appear to be inattentive. Sometimes they appear to be oppositional because of their difficulty with understanding (and therefore *doing*) what they are told.

Secondary problems of children with language disorders include difficulties with social interactions and processing of emotions. Language is what humans use to interact and communicate their ideas to one another. When this ability is compromised, children may isolate or try to find less language-intense activities to occupy their time. Emotional problems may ensue from the child's difficulty with using language to express and, therefore, process their inner world.

ASSESSMENT. Assessment of language should include several measures meant to identify the child's specific language profile. Tests should assess all levels of language, including phonemes, single words, simple phrases, complex sentences, and conversation. Measures of both expressive and receptive language should be included. In the assessment of receptive language, children are asked to distinguish between similar sounds and words, remember and repeat word lists and related strings of words, point to a picture that depicts a vocabulary word, and follow increasingly complex directions presented only once. In the assessment of expressive language, children are asked to perform tasks such as listing as many round objects as they can within a time limit, naming a depicted or described item, defining words or concepts, or creating a syntactically complex sentence according to strict guidelines.

In addition, the psychologist might explore *pragmatics*, which is the child's ability to participate in a conversation and use social language. This involves interpreting nonverbal aspects of communication, as well as observing basic social rules, such as turn taking in conversation. Although neuropsychologists often evaluate pragmatics in addition to receptive and expressive language, they also work in concert with speech and language specialists when additional assessment is indicated.

Visuoperceptual Functioning

OVERVIEW. There are several associated constructs in neuropsychology that reflect people's ability to make sense of what they see, to organize it, or to copy it. These abilities are referred to as *visuoperceptual–visuoconstructive abilities*.

Problems with *visuoperception* are distinct from problems with vision. A person with acute eyesight can struggle with perceptual difficulties, such as identifying which of several figures are exactly alike. Some children have difficulty seeing exactly where something is, and these children may have trouble localizing a point in space or judging the direction of a line. *Visuoconstruction* abilities allow a child to join parts to make a whole. These skills require the integration of the motor system with the visual system. Examples include the ability to put together blocks to form a design or to draw three lines to form a triangle.

Problems with visuoperceptual development have academic, as well as social, ramifications. Academic areas, such as mathematics that are less reliant on verbal support are at risk. In addition, concepts such as time and monetary values may not be clearly understood. Students with these difficulties often exhibit a poor sense of direction, and problems with integrating complex visual arrays may lead to feelings of being overwhelmed. They may also have difficulty “reading between the lines,” thereby making comprehension of less tangible reading concepts (such as theme) more elusive.

Social problems are also often seen in students with these delays. Many elements of good social interactions are nonverbal, including the ability to notice and interpret gestures, facial expression, body posture, and tone of voice. Students with visuoperceptual delays may be overreliant on verbal information and may not understand when people are being sarcastic or when something is said in jest.

ASSESSMENT. Assessment of visual processing must address each of the specific elements of this system. The visuoperceptual abilities should be tested using tasks that do not require the child to use their hands to produce the response—for example, activities that require the child to identify designs that match or differ from the target, as well as measures of mental rotation (finding which design one is the same as the target, only rotated). Visuoconstruction tasks add the demand of integrating the hands and eyes in producing the response—for example, having the child draw copies of designs or use blocks to create a replica of a model.

Sensory/Motor Functioning

OVERVIEW. The sensory/motor system is also assessed as part of the neuropsychological examination. Lateralized sensory or motor problems suggest neurologic problems on the opposite side of the brain and are often correlated with cognitive processes localized to the right or left hemisphere. Tasks requiring the perception of visual/auditory fields or specific actions with the right and/or left sides of the body are part of this domain. In addition, the integration of perceptions or movements is also assessed.

Motor assessment is further broken down into the assessment of handedness and tests of large versus fine motor development, as well as the ability to plan motor responses (praxis).

ASSESSMENT. The sensory examination usually includes an assessment of visual fields using clinical methods such as having the child look at the examiner’s nose and then determine whether the assessor is moving the right or left extended hand. Similarly, the assessment of bilateral auditory perception might include the assessor standing behind the child and rubbing their fingers near the child’s right or left ear. Other perceptual tests might assess the ability to name unseen objects placed in the child’s right or left hands. Finger agnosia is tested by touching a finger when the child’s hand is hidden behind a screen and then having them indicate which finger was touched. Integration of perception might include having the child follow directions involving a picture that is shown. Age-based normative data are available for all of these tasks.

Both fine and large motor tests are usually assessed on both the right and left sides of the body. Lateralized fine motor tasks include quickly placing pegs in holes with each hand or squeezing a hand dynamometer with each hand to assess grip strength. Fingertip tapping is one way of testing motor sequencing, as do activities requiring the child to repeat sequences of movements from memory. Handedness is best assessed by having the child do a number of tasks with one hand (e.g., “Show me how you use this spoon,” “Hand me the dime,” and “Throw me the ball”) in random order. Assessment of difficulties with motor planning can be done using pantomime.

Large motor testing involves having the child demonstrate gait while walking forward and backward, running, skipping, walking a straight line, and balancing on one foot. In cases in which the findings of motor screening are significant, the neuropsychologist might refer the child to an occupational or physical therapist for further, more specific, evaluation.

NEUROPSYCHOLOGICAL CONCLUSIONS

Neuropsychological testing is commonly undertaken according to several discrete domains that reflect areas of brain functioning. Typically, these include attention and executive functioning, memory, and language, as well as visuoperceptual and sensory/motor functioning.

Considerations

In considering neuropsychological issues, the following factors should be kept in mind:

1. Although it is always tempting to think in terms of organic versus functional, it needs to be remembered that these are interacting systems, and learning new skills can alter brain structure. This has

been found in the case of taxi drivers in a major European city, as well as in individuals being taught to juggle.

2. It should be noted that the term “attentional problems” can be used diagnostically or descriptively. The fact that clinicians detect problems with attention on a student’s individual intellectual or processing tests (including computerized instruments) does not mean that the student has ADHD. Under DSM-5 and DSM-5-TR, this is a clinical diagnosis that involves specific criteria, is described behaviorally (inattention, hyperactivity, and impulsivity), and is determined by history (several symptoms by age 12 years) and current functioning (several symptoms seen in two settings). Systematic assessment can help to clarify issues around ADHD (and medication), but this will not be done with a single instrument and will require a pattern of responding on a selected group of tools along with interviews, observations, and, usually, rating scales from home and school.
3. It should be noted that learning (and emotional) problems can mimic attention deficit. In assessing ADHD, the clinician will need to rule out artifacts and to consider whether a child stops paying attention because they stop understanding (e.g., what they are reading or hearing) or have primary ADHD. In addition, the clinician will need to consider whether something like ongoing depression or situational anxiety is making it hard for the child to attend.
4. Broad constructs such as executive function can have relevance for a wide variety of diagnostic categories, such as ADHD and ASD. It is possible that about one-third of the adults and children who have ADHD also have executive dysfunction (EDF), which can increase the difficulties facing these clients. As a result, as with ADHD and learning disorder, where there is a large overlap when one sees ADHD, one might look for EDF. The statistical overlap between ASD and EDF is harder to determine because of issues of definition as well as research results. Depending on how it is measured, EDF might be seen as ubiquitous in ASD.
5. After early brain injury, language and motor functioning are the most likely to benefit from “plasticity.” Some research suggests that, with this process of reorganization, other functions (most notably, visuoperceptual abilities) may be “crowded out,” yielding scores that are lower than expected.
6. Interventions for neurologically driven developmental delays have their most profound effect on younger children. Recent studies have shown

that, in children with reading disabilities, bilateral representation of language identified with fMRI before intervention shifted to the left hemisphere by several orders of magnitude in every subject after only 80 hours of reading intervention. These changes in the brain were accompanied by improved reading skills. Thus, the philosophy of delaying intervention until a deficit is fully expressed may keep children from receiving the full benefit that early intervention provides.

7. Risk factors for reading disabilities include family history, early language delays, poor articulation, chronic ear infections, poor early rhyming abilities, inability to recite (not sing) the alphabet by the end of kindergarten, and early brain injury.
8. “Ambidexterity” (consistently using the right hand for some specific tasks and the left hand for other specific tasks) often runs in families in which several members are left handed. In contrast, “ambiguous handedness” (or the use of either hand for the same task; sometimes writing with the right hand, sometimes writing with the left hand) can be a pathognomonic sign suggesting poor cerebral organization for specific behaviors.
9. ADHD more adversely affects abilities typically associated with right hemisphere functioning (such as fine motor skills and visuoperceptual abilities) and affects attention and executive functioning. Psychostimulant medication has been shown to improve functioning in all of these domains in children with ADHD.

DIFFERENT APPROACHES

Psychoeducational, psychodiagnostics, and focused neuropsychological assessment represent different paradigms: The first assesses referral issues in a descriptive way without reference to brain–behavior relationships, and the last assesses them in a functional way from a brain–behavior point of view. In general, when a systematic assessment of brain–behavior relationships is important to understanding referral issues, focused neuropsychological evaluation is indicated. When the referral question centers around learning and school performance, a psychoeducational assessment would be relevant and useful. However, every situation should be decided on a case-by-case basis, in consultation with the parents and involved professionals. It should be noted that developmental, psychoeducational, or psychodiagnostics evaluations and focused neuropsychological assessment are complementary rather than competing approaches. They can be used as alternatives, depending on the needs of the case, or to supplement each other in complicated situations involving testing by one or more professional psychologists. However, regardless of the type of assessment, it is the psychologist's ability to integrate these and other test data with clinical observations, referral issues, and diagnostic possibilities that benefit the patient in a practical way in terms of understanding what is going wrong and making appropriate interventions.

FURTHER READINGS

Ardila A. The impact of culture in neuropsychological test performance. In: Uzzell BP, Ponton M, Ardila A, eds. *International Handbook of Cross-Cultural Neuropsychology.* 1st ed. Psychology Press; 2007, 23–40.

Baron IS. *Neuropsychological Evaluation of the Child: Domains, Methods, and Case Studies*. 2nd ed. Oxford University Press; 2018.

Byrd D, Arentoft A, Scheiner D, Westerveld M, Baron IS. State of multicultural neuropsychological assessment in children: current research issues. *Neuropsychol Rev*. 2008;18:214–222.

Casas R, Calamia M, Tranel D. A screening test of English naming ability in bilingual Spanish/English speakers. *J Clin Exp Neuropsychol*. 2008;30:956–966.

Denckla MB. A theory and model of executive function: a neuropsychological perspective. In: Lyon GR, Krasnegor NA, eds. *Attention, Memory, and Executive Function*. Paul H. Brookes; 1996.

Dennis EL, Thompson PM. Typical and atypical brain development: a review of neuroimaging studies. *Dialogues Clin Neurosci*. 2013;15(3):359–384.

Flanagan DP, Alfonso VC, Mascolo JT, et al. Use of ability tests in the identification of specific learning disabilities within the context of an operational definition. In: Flanagan DP, Harrison PL, eds. *Contemporary Intellectual Assessment: Theories, Tests, and Issues.* The Guilford Press; 2012, pp. 643–669.

Fletcher JM, Lyon RG, Fuchs LS, Barnes MA. *Learning Disabilities: From Identification to Intervention*. 2nd ed. Guilford Press; 2018.

Fuster JM. *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe*. Lippincott-Raven; 1997.

Hale TS. A distributed effects perspective of dimensionally defined psychiatric disorders: and convergent versus core deficits in ADHD. *Front Psychiatry*. 2013;5:62.

Kavale KA, Forness SR. *The Nature of Learning Disabilities: Critical Elements of Diagnosis and Classification*. Lawrence Erlbaum; 1995.

Kenworthy LE, Black DO, Wallace GL, Ahluvalia T, Wagner AE, Sirian LM. Disorganization: the forgotten executive dysfunction in high-functioning autism (HFA) spectrum disorders. *Dev Neuropsychol*. 2005;28(3):809–827.

Kolb B, Whishaw IQ. *Fundamentals of Human Neuropsychology*. 7th ed. Worth; 2015.

Leung RC, Vogan VM, Powell TL, Anagnostou E, Taylor MJ. The role of executive functions in social impairment in autism spectrum disorder. *Child Neuropsychol*.

2016;22(3):336–344.
Lezak MD, Howieson DB, Loring DW, Hannay HJ, Fischer JS. *Neuropsychological Assessment*. 5th ed. Oxford University Press; 2012.
Meltzer L. *Executive Function in Education: From Theory to Practice*. Guilford Press; 2007.
O'Hearn K, Asato M, Ordaz S, Luna B. Neurodevelopmental and executive functioning in autism. *Dev Psychopathol*. 2008;10:1103–1132.
Olson K, Jacobson K. Cross-cultural considerations in pediatric neuropsychology: a review and call to attention*.* *Appl Neuropsychol Child*. 2015;4(3):166–177.
Parke EM, Thaler NS, Etcoff LM, Allen DN. Intellectual profiles in children with ADHD and comorbid learning and motor disorders. *J Atten Disorder*. 2020;24:1227–1236.
Pennington BF. *Diagnosing Learning Disorders: A Neuropsychological Framework*. 2nd ed. The Guilford Press; 2008.
Poldrack RA, Congdon E, Triplett W, et al. A phenome-wide examination of neural and cognitive function. *Sci Data*. 2016;3:160110.
Sattler JM. *Assessment of Children: Cognitive Foundations and Applications.* 6th ed. Sattler; 2020.
Shaywitz S. *Overcoming Dyslexia*. 2nd ed. Vintage; 2020.
Simos PG, Fletcher JM, Bergman E, et al. Dyslexia-specific brain activation profile becomes normal following successful remedial training. *Neurology*. 2002;58:1203–1213.
Sinzig J, Morsch D, Bruning N, Schmidt MH, Lehmkuhl G. Inhibition, flexibility, working memory and planning in autism spectrum disorders with and without comorbid ADHD-symptoms. *Child Adolesc Psychiatry Mental Health*. 2008;2:4.
Stiles J, Moses P, Passarotti A, Dick FK, Buxton R. Exploring developmental change in the neuronal bases of higher cognitive functions: the promise of functional magnetic resonance imaging. *Dev Neuropsychol*. 2003;24:641–668.
Tannock R. Specific Learning Disabilities in DSM-5: are the changes for better or worse? *Int J Res Learn Disabil*. 2013;1(2):2–30.
Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biol Psychiatry*. 2005;57:1336–1346.
Yeates KO, Ris MD, Taylor HG, Pennington BF. *Pediatric Neuropsychology: Research, Theory, and Practice*. 2nd ed. Guilford Press; 2010.

35.4 Psychiatric Emergencies in Children and Adolescents

ALEXANDRA JUNEWICZ, M.D., AND RUTH S. GERSON, M.D.

The past 20 years have seen a dramatic increase in the number of visits to emergency departments (EDs) by children and adolescents in psychiatric crisis, with estimates now reaching 2% to 5% of all visits or over 800,000 visits per year. Between 2007 and 2016, the number of pediatric emergency visits for mental health disorders rose by 60%. Unfortunately, appropriate assessment and treatment are not available to many youths presenting to the ED in a psychiatric crisis. Many present to EDs with limited or no pediatric or mental health clinicians. Nationwide, fewer than 20% of youth presenting to the ED in psychiatric crisis are seen by a mental health professional, and fewer than 50% of youth presenting to the ED with suicidal behavior receive a mental health evaluation, even though those who do not are at increased risk of a future suicide attempt. As the ED is the entry point to psychiatric care for greater than half of youth, improving the care provided to these patients is imperative. This section will underscore the key components of the emergency evaluation of a child or adolescent in a psychiatric crisis and will emphasize the following:

- Understanding the role of developmental, academic, and social context in assessing children in psychiatric crisis
- Differential diagnosis of behavioral disturbances
- Assessment of suicidality, aggression, new-onset psychotic symptoms, and risk-taking behaviors
- Assessment of nonverbal children (including young children and those with autism or intellectual disability [ID])
- Management of agitation in the ED
- Legal issues in emergency child psychiatry, including assessing for child abuse

EPIDEMIOLOGY

The emergency care system has unfortunately been slow to adapt to the increasing demand for child and adolescent psychiatric care. Youth in psychiatric crisis are difficult for EDs to manage. Most children presenting to EDs in psychiatric crisis are seen in general medical or pediatric emergency rooms or, more rarely, in adult psychiatric emergency services (PES). Most EDs lack the specialized staff (child psychiatrists, psychologists, social workers, and nursing staff) necessary to effectively care for youth in psychiatric crisis, who also have a longer length of stay and increased need for inpatient admission or transfer than other youth. EDs also typically lack the physical resources to manage the high volume of child psychiatric patients safely. Agitated or suicidal youth are often placed together in medical examination rooms full of potentially dangerous objects, in busy and noisy holding areas, or on stretchers in trauma bays with security or police supervision. Youths who are seen in the adult PES instead are exposed to adult patients who are severely mentally ill, homeless, or intoxicated. This can be frightening and reinforces the stigma surrounding mental illness and psychiatric treatment. The PES also lacks appropriate staff, as adult psychiatrists and nursing staff trained to work with adults may not know how to approach youth in a developmentally appropriate way.

Youth seeking emergency psychiatric care are frequently clinically acute and clinically complex. Many have previously undiagnosed or undetected psychiatric illness including depression, anxiety, psychosis, post-traumatic stress disorder (PTSD), substance use, autism, and attention deficit hyperactivity disorder (ADHD). Suicidal thoughts and severe aggression are also common precipitants for an ED visit, again often subsequent to a lengthy period of unrecognized or untreated psychiatric symptoms. Suicide is currently the second leading cause of death for 10 to 24 year olds, and as many as 10% of children attempt suicide sometime during their young lives. Studies indicate that 40% of youth who attempt suicide had been seen in the ED in the year before their death, suggesting that an opportunity for life-saving intervention at that ED visit was missed. EDs are also increasingly seeing violent youth, who are often “frequent fliers” in the ED. They frequently have diagnoses of conduct disorder or other disruptive behavior disorders; are involved with the juvenile justice or child welfare systems; and have cognitive impairments that make it difficult for them to function in the community. Many youths present with such severe symptoms that inpatient hospitalization is required. But with the steady decline in access to inpatient child and adolescent psychiatry beds, 23% to 58% of youth in need of psychiatric hospitalization are left boarding in EDs for days, without meaningful treatment, awaiting inpatient admission.

There are significant racial and ethnic disparities among children and adolescents who experience psychiatric crises. While ED visits for mental health complaints have been rising among all youth, minority youth have been disproportionately affected by this trend. Rates of mental health–related ED visits are highest among non-Hispanic Black youth and have been increasing at a greater rate among Hispanic youth. Although suicide rates have traditionally been higher among White than Black individuals, suicide rates have been steadily increasing among Black youth for more than two decades. In recent years, suicide rates have risen for Black youth aged 5 to 11 years while decreasing for their White counterparts. Currently, among children younger than 13 years of age, the suicide rate for Black children is two times higher than the rate for White children. Disparities also exist among other racial, ethnic, and sexual minority groups. For instance, American Indians, Alaska Natives, and lesbian, gay, bisexual, transgender, and questioning or queer (LGBTQ) youth have a higher risk for suicidal ideation and behavior. Despite these data and trends, minority youth area are less likely to have access to mental health treatment and services than nonminority youth. Consequently, minority youth may be more likely to suffer from undetected and untreated psychiatric symptoms, which may worsen to the point of crisis.

Despite the well-documented demand for emergency mental health care, many training programs lack comprehensive instruction on emergency psychiatric assessment of youth. This chapter will outline the key priorities, principles, and components of a comprehensive evaluation and risk assessment for a child or adolescent in psychiatric crisis.

PRIORITIES OF EMERGENCY PSYCHIATRIC EVALUATION FOR CHILDREN AND ADOLESCENTS

Safety

The psychiatric evaluation of a child or adolescent in the ED should have two main priorities: safety in the ED (for patients and staff) and effective clinical and risk assessment to determine whether the child is safe to leave the ED. Ensuring safety in the ED starts with appropriate physical space. Children and teens presenting to the ED with suicidal ideation, aggressive agitation, or other acute psychiatric manifestations are unlikely to be safe in a standard medical examination room. The ED environment, whenever possible, should be modified for safety including removing potentially dangerous items, decreasing overstimulation by avoiding bright lights and unnecessary noise, providing privacy while allowing the patient to be observed continuously, and enhancing patient comfort by giving the patient a comfortable place to lie or sit down as well as food or drink. The child and family should be given clear expectations as to the components of the evaluation, the waiting time, and the interventions and consequences that will be given should the patient become agitated or violent. Adequately trained personnel should be available to help if the patient becomes agitated, and staff should have a clear response plan for calming an agitated or aggressive patient including with medication or restraint if needed (this is discussed further below). These modifications reduce the likelihood of aggression to self or others, and if such aggression does occur, mitigate the risk of serious injury to patient or staff.

Evaluating Medical and Psychosocial Causes of Behavioral Disturbance

The second priority of the ED evaluation, effective clinical and risk assessment, starts with ruling out and treating medical and social/systemic causes of the behavioral disturbance. While children and adolescents generally lack the complex and chronic medical conditions of adult psychiatric patients, medical illness can nonetheless masquerade as a psychiatric illness in subtle and confusing ways. Seizures (particularly temporal lobe seizures and absence seizures), infectious and autoimmune encephalitis, brain tumors, head trauma, and tropical diseases can manifest with hallucinations, mental status change, and behavioral dysregulation. Children may also present with behavioral changes related to undiagnosed chronic diseases such as porphyria, sickle cell anemia, metabolic disorders, and other genetic syndromes. Drug intoxication and other self-poisoning (due to accidental ingestion or intentional acts meant as a suicide attempt or to get high) can also present with psychiatric symptoms. Red flags that are suggestive of a possible medical etiology include the following:

- Clouding of consciousness
- Waxing and waning mental status
- Disorientation
- Memory impairment
- Acute onset of psychosis in a child with no previous psychiatric history
- Abnormal vital signs

When a medical etiology is suspected, even if it is unlikely, an appropriate workup (including neuroimaging when indicated, and cerebrospinal fluid studies if encephalitis is a potential concern) should be completed to ensure that a medical illness is not missed. Even if a youth has already been “medically cleared,” further assessment may be needed as the goal of the medical clearance is not to “rule out” all possible medical etiologies. Youth without suspicion for medical etiology of psychiatric symptoms should receive a focused medical assessment, and testing should be incorporated judiciously. Children who have acute medical illness, who are severely intoxicated, or who have made a suicide attempt by overdose or other method that has caused significant injury, will likely require medical admission prior to psychiatric hospitalization, and should be followed by the psychiatric consult–liaison service during their medical care to avoid unnecessary delay in their psychiatric treatment.

Clinicians should also remember that normal developmental changes, such as puberty, and minor physical ailments such as ear infections, tooth pain, hunger, or fatigue can manifest with or trigger psychiatric or behavioral symptoms, particularly in young children and youth with intellectual or developmental disabilities (Table 35.4–1).

While medical illness can and does masquerade as psychiatric illness, what is even more common in children is psychosocial crisis manifesting as psychiatric disturbance. It is crucial to consider when a child presents with behavioral disturbance or even more classically psychiatric symptoms such as hallucinations or anxiety, that the child may be the “identified patient” in a distressed or pathologic family or social system. Domestic violence and family conflict often manifests with aggressive behavior in a child; reports of auditory hallucinations may be a child’s mimicking the symptoms of a mentally ill parent; or the child may be a victim of physical or sexual abuse or of fabricated or exaggerated symptoms in cases of medical child abuse. Treating the child alone when the source of the symptoms is in the family or social environment not only puts the child at risk for adverse effects from unnecessary medication, but is also psychologically damaging as it teaches the child that they are the problem, not the toxic context in which they find themselves. Alternatively, the child may have true psychiatric illness, but the emergency is not a worsening of symptoms per se but rather that the caretakers are for whatever reason less able to manage those symptoms. A parent may be weakened by illness, divorce, or financial stress, or may lose a key support or structure like an after-school program or grandparent willing to help care for the child; or the child may be removed from the parent and placed in a foster care home where the caretaker is not equipped to care for a child with mental illness. Here, again, the crisis is not psychiatric per se, but rather a breakdown of the caretaking system, which must be shored up if the child is to be discharged safely.

Table 35.4–1.
Common Medical Causes of Psychiatric Disturbance in Children and Adolescents

Fever
Nonconvulsive status epilepticus
Temporal lobe seizures
Absence seizure
Encephalitis
Central nervous system neoplasm or other occult malignancy
Cerebral systemic lupus erythematosus
Childhood confusional migraine
Head trauma or postconcussive syndrome
Porphyria
Sickle cell disease
Substance intoxication
Withdrawal from illicit substances
Anticholinergic delirium
Steroid psychosis or steroid-induced mania
Medication adverse effects such as akathisia from risperidone or irritability from levetiracetam

PRINCIPLES OF EMERGENCY PSYCHIATRIC CARE FOR CHILDREN AND ADOLESCENTS

Comprehensive Assessment of the Child

When a patient is in crisis, treatment must start immediately, so the initial assessment must be therapeutic in itself. Each piece of the assessment is directed toward understanding the nature of the crisis, de-escalating and mitigating the distress of the child and accompanying adults, and monitoring response to therapeutic information. The first step is information gathering. Given the case complexity of youth presenting to the ED in psychiatric crisis, collateral from the parent or guardian and any accompanying adults is a critical source of information about the child's life at home and at school, their strengths, stressors, medical issues and history (including allergies), history of psychiatric treatment (including response and compliance), developmental history, history of substance use, family history, legal history, social history including any prior trauma exposure, and academic history. Parents or accompanying adults will also have a perspective on the current crisis; however, the assessment should not be based exclusively on the parent or guardian's report, even if the child is very young or intellectually disabled. A curious, empathic interview of the child is imperative to understand their perspective, the reasons behind their behaviors, their interpretation of what is going on around them, their strengths and cognitive abilities, and the internalizing symptoms such as depression, anxiety, and trauma-related symptoms that often are missed by parents and other adults (Table 35.4–2).

Table 35.4–2.
Components of a Comprehensive Assessment

Identifying information
Chief complaint
History of present illness
Past psychiatric history
Current medications and medication history
Past medical history
Substance use history
Trauma history
Social history including trauma and child welfare involvement
Developmental and educational history
Family history
Legal history
Mental status examination including cognitive screen
Medical clearance including physical if indicated

Conducting the interview with an aim to truly understand and empathize with the child's perspective can make it very therapeutic for the child. Most children have not had the experience of an adult truly listening to them, striving to understand their experience without judgment or reprimand, and trying to help them. Explaining the purpose of the evaluation and the clinician’s role and decision-making process in a clear, direct manner fosters rapport with children, alleviates their sense of helplessness, and promotes their sense of agency. The child’s privacy should be respected and the confidential nature of the evaluation (and the limits of confidentiality) should be made clear. While a conversational rather than “checklist” approach is preferred, the interview with the child must include questioning about suicidal and homicidal ideation, though with the recognition that children and adolescents often minimize or deny suicidal or homicidal ideation or intent, as they fear getting in trouble or being hospitalized. An empathic, normalizing, and nonjudgmental stance to these questions can increase the likelihood of an honest response and also serves to build rapport, even with hostile adolescents. Questions about hallucinations and delusions are also critical, though the clinician should be careful with phrasing these questions as young children or those with developmental delays may misinterpret the question (or their own internal experience) and falsely answer yes, while teens who are truly psychotic may deny such symptoms for fear of being labeled as “crazy.”

For very young or otherwise minimally verbal or nonverbal children (such as those with autism or ID), the clinician should still attempt to engage the patient to complete a thorough mental status examination. A child with little expressive language may still be able to follow commands, engage in parallel or reciprocal play, and express needs through gestures. The clinician can also observe the child’s fine and gross motor skills through drawing and playing, and can assess frustration tolerance, hyperactivity, distractibility, attachment to the parent or guardian (or evidence of fear of the parent), and anxiety with separation.

A therapeutic stance is equally important for interviewing parents, foster parents, and other caregivers, who are often exhausted from dealing with frightening and disruptive symptoms and behaviors without a good understanding of their cause. Treating child psychiatric patients in the ED often feels like having two patients—the child and the family—and approaching both carefully and empathically is crucial to understand the situation and create buy-in for ongoing treatment. A parent’s anxiety or agitation can be ameliorated by empathizing with how difficult a child’s mental illness can be for a family and acknowledging that the ED evaluation itself is a difficult process. Reducing the parent's anxiety or anger helps them to reflect on and mentalize the child's experience and to consider their own important role in treatment. Interviews with both the parent and the child should be conducted in a quiet, private space free from interruption so that everyone feels safe and respected when discussing their challenges and expressing painful feelings.

In many cases, a child will be brought to the ED by someone other than their parent (e.g., police, emergency medical services [EMS], friends, or child welfare workers). If the parent or guardian cannot be reached, most states allow clinicians in an emergency to initiate evaluation and treatment of the child, though ongoing attempts to reach the parent should be made. Occasionally, an adolescent will insist that the parent not be contacted. While many states allow adolescents to seek mental health services without the parent being involved or notified, these youth should be encouraged to include their parents in the evaluation, and often with support and empathy of the ED staff, a rapprochement can be made between the parent and child.

Often the ED clinician will find that the descriptions of symptoms and behaviors and the explanations for such will vary between the child, the parent, and the other informants. These differences may stem from real differences in the child's symptoms and behaviors across different contexts (e.g., at home or at school). Alternatively, a parent or other adult may not be aware of or attuned to the child's internal state and thus may miss internalizing symptoms and overemphasize externalizing behaviors. The child may have poor insight, may be guarded or minimizing of symptoms, or alternatively may be voicing a major stressor (such as abuse, domestic violence, or school bullying) that the adults are attempting to hide or of which they are unaware. Other sources of information may be needed to obtain a balanced view of the child and family system. A child's teachers, outpatient providers, pediatrician, and social services workers (child welfare workers, other case managers) can provide very helpful perspective on the child's recent symptoms and functioning, his developmental and cognitive level, the family's strengths and vulnerabilities, and the resources and supports available to them in the community. Consent from the parent or guardian should ideally be obtained before contacting any of these individuals, but in potentially life-threatening situations such as suicidal ideation or concern for child abuse, the safety of the patient outweighs the risk of confidentiality, and the clinician can contact these individuals for information without violating the requirements of Health Insurance Portability and Accountability Act (HIPAA).

Trauma-Informed Assessment and Care

The ED clinician's key task is to integrate the child's presentation and collateral information into a clear formulation and effective treatment plan. This must include consideration of trauma, as unfortunately traumatic experiences are highly common in youth presenting to EDs with psychiatric emergencies. High rates of trauma and abuse (over 50%) and child welfare involvement (over 30%) have been documented in this population. ED clinicians should thus utilize a trauma-informed approach to care. Clinicians should inquire about a child's history of trauma with empathy, respect, and calm. Validation, normalization, offering genuine praise, and highlighting resiliency may help a child speak about and process traumatic events. It is important to inquire directly about traumatic experiences but also to recognize that youth exposed to trauma may not disclose it, due to explicit or implicit messages from family members or others about not disclosing, fear of not being believed, or PTSD-related avoidance. Thus "universal trauma precautions" are often recommended.

While it is often difficult to ascertain whether acute psychiatric symptoms displayed by a child in an ED stem from having been traumatized, even if a child's psychiatric symptoms predated traumatic events, it is highly likely that these events exacerbated their symptoms. The ED visit can be an important opportunity for psychoeducation and reframing symptoms in a trauma context with youth and caregivers. Clinicians should remember that their words are powerful, as they can shape the manner in which the child views themself and the way others view the child. Focusing on a child's strengths can be therapeutic and foster agency, resilience, and engagement.

Effective Psychotherapeutic and Psychopharmacologic Management of Agitation

Importantly, the ED visit and interactions with clinicians should not add to a child's trauma history. Key to ensuring that the evaluation is therapeutic is the prevention and effective management of agitation to avoid restraint and seclusion. Children and adolescents experience restraint and seclusion as frightening, and their perception of time can become distorted, such that they perceive themselves as being restrained for much longer than it is in reality. After a restraint, children are angry at and frightened of staff, which will profoundly limit the staff's ability to fully assess and treat them. Restraint can be avoided through proactive engagement to prevent a crisis and by effective verbal and psychopharmacologic de-escalation strategies.

A safe environment is the first step to prevent agitation. Overcrowding, excessive noise, lack of privacy, long wait times, and a tense atmosphere can agitate even a normally quiet child. For a child with impulse control problems, psychosis, trauma history, or autism, such a setting can be catastrophic. A safe ED environment, attention to who is at risk of agitation, vigilance to early signs of distress, and preventive strategies can avert a crisis. When a child is showing early signs of agitation, the ED staff should work as a team to de-escalate the situation. First, the patient should be moved to a safe and quiet area, if they are not already there, separated from other patients and either separated from or rejoined with preferred family members based on the patient's preference. The patient should be offered food, drink, and distractions such as TV or a game. Choices (of food or drink or TV channel) can help a child feel more in control and thus regain emotional control. One member of the staff should approach the patient in a nonthreatening way to listen to their complaint or concern, direct them away from the focus of their anger, and encourage them to verbalize their needs. If the child begins to explain their concern, even if they are still agitated or yelling, staff should empathize by repeating their statement or paraphrasing it back to ensure that the patient's view is understood correctly. As the child makes incremental steps to calm down, for example, by talking to staff or by refraining from pacing, staff should praise these positive steps even if the child is still engaging in other negative behaviors (such as cursing or scowling). Most children in the ED are anxious about the process, their parents' reaction, and the possibility of hospitalization. Others may be paranoid, may have been abused and thus be hyperreactive to perceived threats, or may be hungry or in pain. Staff should work empathically to alleviate distress, while demonstrating calm and clarity that they will work to keep the patient and all others safe and that violence will not be tolerated.

A child's agitation can also stem from a parent's anxiety, anger, or fear. Parents can be anxious in the ED because of concerns about the child's symptoms or safety, fear that the child will be taken away from them (by child protection or by hospitalization), or worry about missing work or affording the cost of the emergency evaluation or ambulance. They may have themselves had negative experiences with psychiatric care. The team must listen to and empathize with the parent, emphasize that the parent is not being blamed for the situation, and assure them that the staff will work with the family to ensure the child's safety. If a parent is agitated, it is most often helpful to separate the parent and the child while staff work to calm the parent, to avoid the child becoming agitated as well.

If the patient continues to be agitated and begins to escalate further, perhaps making threats or beginning to destroy ED property or attempting to elope, oral medications should be offered. The medication should target the underlying cause of the agitation and should be uniquely tailored to the child's individual needs with regards to route of administration, potential side effects, comorbid physical illness, and past experience with PRN medications. Medication should be selected to sufficiently calm, but not sedate, the child. There is no firm evidence or consensus that one medication or even a certain class of medications is preferable to another in managing agitation. Rather, potential adverse effects should be carefully weighed against potential benefits while considering a child's age, size, recent medication administration, medical conditions, and the psychological cause of their agitation. Medications should be chosen carefully, as the medications most commonly used for agitation in adults can have significant adverse effects on children. The potential for harm from medication is significant, and particularly ketamine, barbiturates, and opioid analgesics are strongly not recommended in this population.

Oral medication should always be offered before injectable medication (IM), particularly because many children (even ostensibly tough ones) are afraid of needles and can become more agitated when approached with an injection. Administering an injection will require a physical hold or restraint, with a high potential for injury to staff and children. A child who initially refuses an oral medication may accept the oral medication if offered the choice between oral and IM. This not only allows the patient to avoid the potentially traumatic experience of restraint and IM but also allows the child to make a choice and begin to demonstrate self-control (Table 35.4–3).

If all the above techniques to de-escalate an agitated child have failed, and the child continues to refuse oral medication and engage in violent or dangerous behaviors, an IM may be indicated for acute chemical restraint.

Care should be taken regarding dosing, as most medications are more potent (and thus require lower dose) when administered intramuscularly.

In most cases, when IM is required, a physical hold, restraint, or seclusion is also necessary for safety. The choice of intervention (physical hold, restraint, or seclusion) should be made based on the regulatory standards of the Center for Medicare and Medicaid Services and The Joint Commission on Accreditation of Healthcare Organizations, state and local governmental and oversight agencies, and the policies of the hospital, as well as the individual patient history. For example, for a child who has been sexually abused, restraint may be deeply retraumatizing such that seclusion should be chosen instead; a child who is engaging in self-injury may not be able to be safe in seclusion and thus restraint may be more therapeutic. The physician should ensure that the child is safe, physically comfortable, and medically stable in the restraint or seclusion, and should explain to the child the reason for the restraint or seclusion and that they will be released as soon as they can be safe. Regulation and hospital policy will dictate the allowed duration of restraint or seclusion, but regardless of regulation the child should be removed from restraint or seclusion as soon as possible. Once the child is released, his safe behaviors should be praised, and the physician should debrief with the patient to review what upsets the patient and how such episodes can be avoided. While restraints and seclusion can be traumatizing for children if executed poorly, with appropriate support, explanation, and debriefing, it can be an opportunity for the youth to better express his feelings and understand the roots of his behavioral dyscontrol.

Table 35.4–3. Medications for Agitation

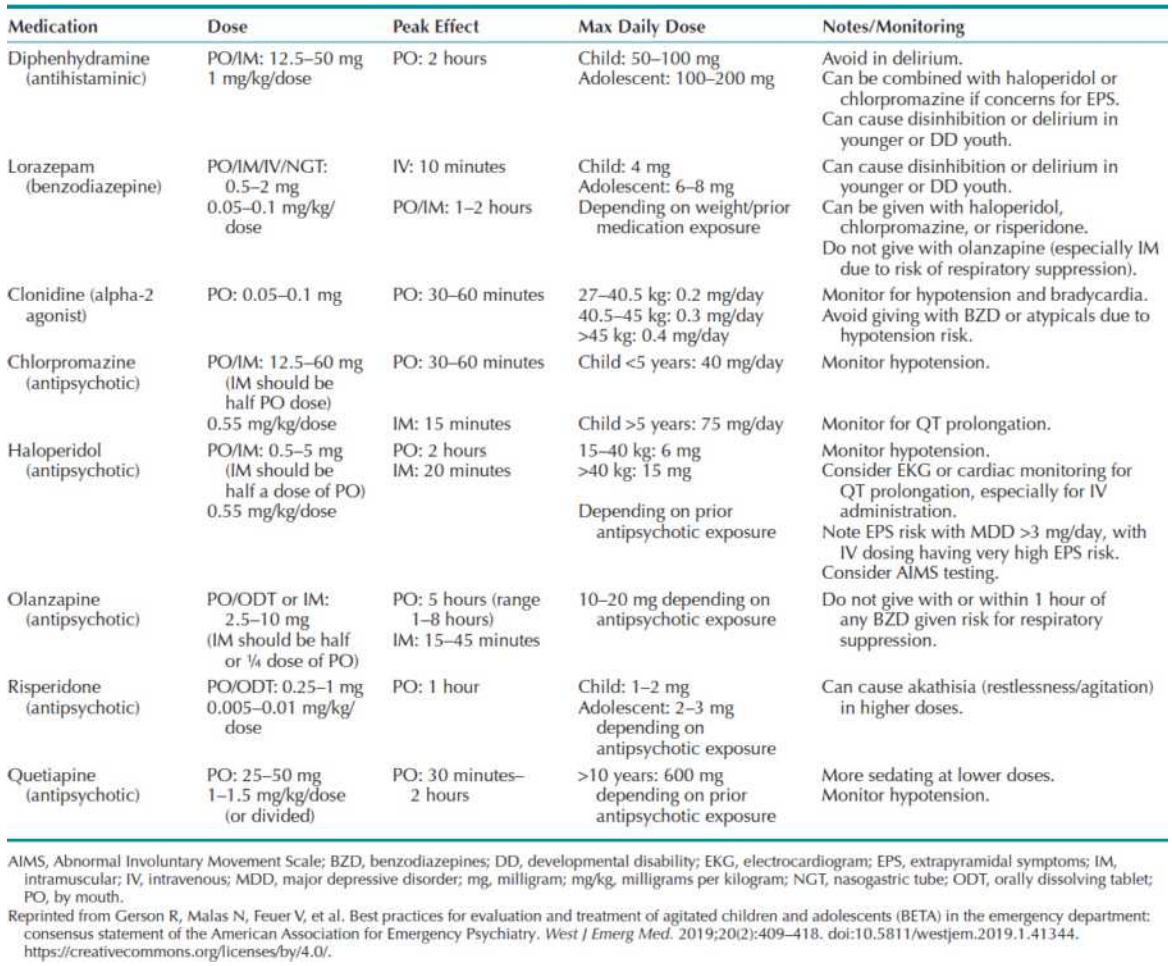

Medication	Dose	Peak Effect	Max Daily Dose	Notes/Monitoring
Diphenhydramine (antihistaminic)	PO/IM: 12.5–50 mg 1 mg/kg/dose	PO: 2 hours	Child: 50–100 mg Adolescent: 100–200 mg	Avoid in delirium. Can be combined with haloperidol or chlorpromazine if concerns for EPS. Can cause disinhibition or delirium in younger or DD youth.
Lorazepam (benzodiazepine)	PO/IM/IV/NGT: 0.5–2 mg 0.05–0.1 mg/kg/dose	IV: 10 minutes PO/IM: 1–2 hours	Child: 4 mg Adolescent: 6–8 mg Depending on weight/prior medication exposure	Can cause disinhibition or delirium in younger or DD youth. Can be given with haloperidol, chlorpromazine, or risperidone. Do not give with olanzapine (especially IM due to risk of respiratory suppression).
Clonidine (alpha-2 agonist)	PO: 0.05–0.1 mg	PO: 30–60 minutes	27–40.5 kg: 0.2 mg/day 40.5–45 kg: 0.3 mg/day >45 kg: 0.4 mg/day	Monitor for hypotension and bradycardia. Avoid giving with BZD or atypicals due to hypotension risk.
Chlorpromazine (antipsychotic)	PO/IM: 12.5–60 mg (IM should be half PO dose) 0.55 mg/kg/dose	PO: 30–60 minutes IM: 15 minutes	Child <5 years: 40 mg/day Child >5 years: 75 mg/day	Monitor hypotension. Monitor for QT prolongation.
Haloperidol (antipsychotic)	PO/IM: 0.5–5 mg (IM should be half a dose of PO) 0.55 mg/kg/dose	PO: 2 hours IM: 20 minutes	15–40 kg: 6 mg >40 kg: 15 mg Depending on prior antipsychotic exposure	Monitor hypotension. Consider EKG or cardiac monitoring for QT prolongation, especially for IV administration. Note EPS risk with MDD >3 mg/day, with IV dosing having very high EPS risk. Consider AIMS testing.
Olanzapine (antipsychotic)	PO/ODT or IM: 2.5–10 mg (IM should be half or ¼ dose of PO)	PO: 5 hours (range 1–8 hours) IM: 15–45 minutes	10–20 mg depending on antipsychotic exposure	Do not give with or within 1 hour of any BZD given risk for respiratory suppression.
Risperidone (antipsychotic)	PO/ODT: 0.25–1 mg 0.005–0.01 mg/kg/dose	PO: 1 hour	Child: 1–2 mg Adolescent: 2–3 mg depending on antipsychotic exposure	Can cause akathisia (restlessness/agitation) in higher doses.
Quetiapine (antipsychotic)	PO: 25–50 mg 1–1.5 mg/kg/dose (or divided)	PO: 30 minutes–2 hours	>10 years: 600 mg depending on prior antipsychotic exposure	More sedating at lower doses. Monitor hypotension.

AIMS, Abnormal Involuntary Movement Scale; BZD, benzodiazepines; DD, developmental disability; EKG, electrocardiogram; EPS, extrapyramidal symptoms; IM, intramuscular; IV, intravenous; MDD, major depressive disorder; mg, milligram; mg/kg, milligrams per kilogram; NGT, nasogastric tube; ODT, orally dissolving tablet; PO, by mouth.

Reprinted from Gerson R, Malas N, Feuer V, et al. Best practices for evaluation and treatment of agitated children and adolescents (BETA) in the emergency department: consensus statement of the American Association for Emergency Psychiatry. *West J Emerg Med.* 2019;20(2):409–418. doi:10.5811/westjem.2019.1.41344. https://creativecommons.org/licenses/by/4.0/.

Choice of Medication in Treating Agitation

Agitation is a symptom, and treatment of symptoms requires a differential diagnosis for their etiology. Potential etiologic factors for agitation include physical pain or distress, anxiety, developmental disability, cognitive impairment, trauma, mania, psychosis, or difficulties with language and communicating needs. Evaluating for medical factors is critical, yet assessment may be limited by the child's agitated behavior; in such cases, clinicians should use their best clinical judgment based on the information available. Interventions should be targeted to the etiologic factors while considering a child's unique needs. Management of agitation specific to certain clinical conditions and situations is discussed below (Fig. 35.4–1).

Delirium. Delirium may stem from physical disease or medication factors and is characterized by an acute change and fluctuations in mental status with inattention, altered awareness, and cognitive deficits. The management of delirium should always seek to identify and address the underlying cause and should begin with nonpharmacologic interventions, such as treating pain or supporting sleep. Anticholinergics, benzodiazepines, and opioid analgesics may precipitate or exacerbate delirium, and thus the use of these agents should be minimized or avoided altogether. If additional medication is required, second-generation neuroleptics such as risperidone, olanzapine, and quetiapine are considered first-line agents. Haloperidol is considered a second-line agent, and if neuroleptics are to be avoided, clonidine may be used. The use of physical restraints should be avoided in medically compromised youth, and risk for falls and medical complications should be monitored in any delirium cases but particularly after psychotropic medication has been administered.

Substance Intoxication or Withdrawal. The approach to agitation in a youth who is intoxicated or in withdrawal should ideally be tailored to the specific substance at hand. However, when intoxication is suspected and the offending substance is unknown, lorazepam is generally preferred for the management of acute agitation. If the patient is suffering from severe agitation or hallucinations, lorazepam may be given together with haloperidol. However, neuroleptics should be used with caution, as they may exacerbate anticholinergic or sympathomimetic toxidromes. Finally, the use of medication is preferable to physical restraints in intoxicated patients, as physical restraints may lead to rhabdomyolysis, a sudden increase in serum potassium, and cardiac arrest.

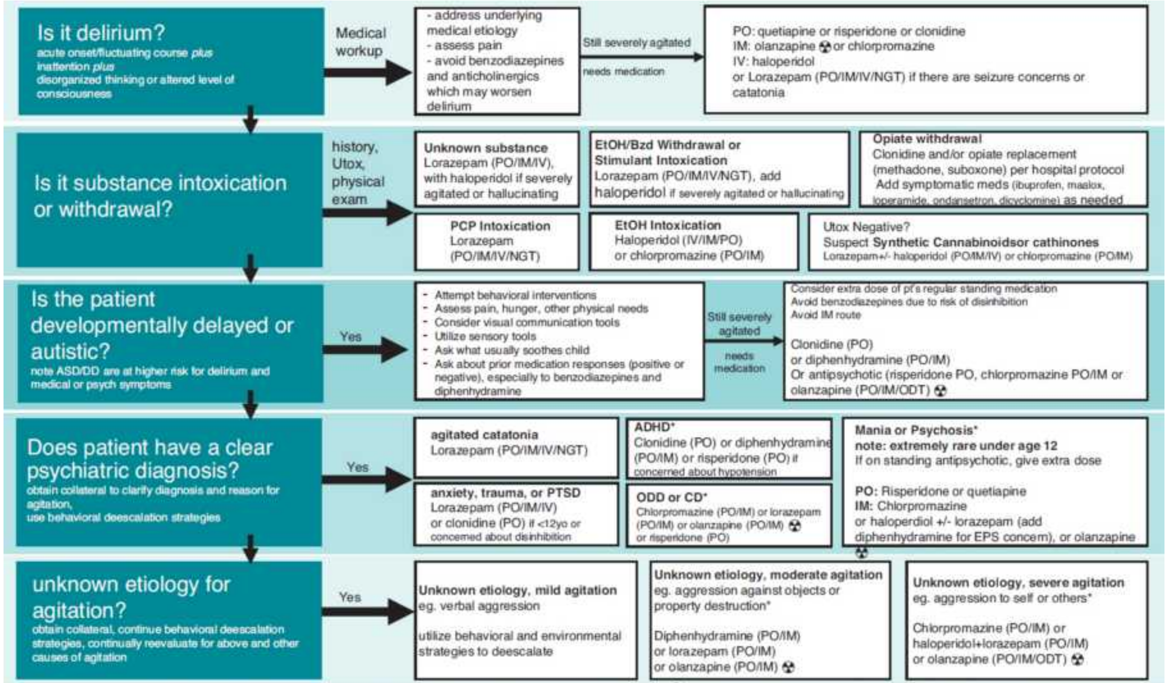
Is it delirium?
acute onset/fluctuating course plus inattention plus disorganized thinking or altered level of consciousness
Medical workup
- address underlying medical etiology
- assess pain
- avoid benzodiazepines and anticholinergics which may worsen delirium
Still severely agitated
needs medication
PO: quetiapine or risperidone or clonidine
IM: olanzapine ☢ or chlorpromazine
IV: haloperidol
or Lorazepam (PO/IM/IV/NGT) if there are seizure concerns or catatonia
Is it substance intoxication or withdrawal?
history, Utox, physical exam
Unknown substance
Lorazepam (PO/IM/IV), with haloperidol if severely agitated or hallucinating
EtOH/Bzd Withdrawal or Stimulant Intoxication
Lorazepam (PO/IM/IV/NGT), add haloperidol if severely agitated or hallucinating
Opiate withdrawal
Clonidine and/or opiate replacement (methadone, suboxone) per hospital protocol
Add symptomatic meds (ibuprofen, maalox, loperamide, ondansetron, dicyclomine) as needed
PCP Intoxication
Lorazepam (PO/IM/IV/NGT)
EtOH Intoxication
Haloperidol (IV/IM/PO)
or chlorpromazine (PO/IM)
Utox Negative?
Suspect Synthetic Cannabinoidsor cathinones
Lorazepam+/- haloperidol (PO/IM/IV) or chlorpromazine (PO/IM)
Is the patient developmentally delayed or autistic?
note ASD/DD are at higher risk for delirium and medical or psych symptoms
Yes
- Attempt behavioral interventions
- Assess pain, hunger, other physical needs
- Consider visual communication tools
- Utilize sensory tools
- Ask what usually soothes child
- Ask about prior medication responses (positive or negative), especially to benzodiazepines and diphenhydramine
Still severely agitated
needs medication
Consider extra dose of pt's regular standing medication
Avoid benzodiazepines due to risk of disinhibition
Avoid IM route
Clonidine (PO)
or diphenhydramine (PO/IM)
Or antipsychotic (risperidone PO, chlorpromazine PO/IM or olanzapine (PO/IM/ODT) ☢
Does patient have a clear psychiatric diagnosis?
obtain collateral to clarify diagnosis and reason for agitation,
use behavioral deescalation strategies
Yes
agitated catatonia
Lorazepam (PO/IM/IV/NGT)
ADHD*
Clonidine (PO) or diphenhydramine (PO/IM) or risperidone (PO) if concerned about hypotension
Mania or Psychosis*
note: extremely rare under age 12
If on standing antipsychotic, give extra dose
PO: Risperidone or quetiapine
IM: Chlorpromazine
or haloperdiol +/- lorazepam (add diphenhydramine for EPS concern), or olanzapine ☢
anxiety, trauma, or PTSD
Lorazepam (PO/IM/IV)
or clonidine (PO) if <12yo or concerned about disinhibition
ODD or CD*
Chlorpromazine (PO/IM) or lorazepam (PO/IM) or olanzapine (PO/IM) ☢ or risperidone (PO)
unknown etiology for agitation?
obtain collateral, continue behavioral deescalation strategies, continually reevaluate for above and other causes of agitation
Yes
Unknown etiology, mild agitation
eg. verbal aggression
utilize behavioral and environmental strategies to deescalate
Unknown etiology, moderate agitation
eg. aggression against objects or property destruction*
Diphenhydramine (PO/IM)
or lorazepam (PO/IM)
or olanzapine (PO/IM) ☢
Unknown etiology, severe agitation
eg. aggression to self or others*
Chlorpromazine (PO/IM) or
haloperidol+lorazepam (PO/IM)
or olanzapine (PO/IM/ODT) ☢

*for these etiologies, in absence of consensus, medication options are listed alphabetically

☢ do not give olanzapine and benzodiazepines within 1h of each other

FIGURE 35.4–1. Agitation management clinical decision flow chart. (Reprinted from Gerson R, Malas N, Feuer V, et al. Best practices for evaluation and treatment of agitated children and adolescents (BETA) in the emergency department: consensus statement of the American Association for Emergency Psychiatry. *West J Emerg Med*. 2019;20(2):409–418. doi: 10.5811/westjem.2019.1.41344. https://creativecommons.org/licenses/by/4.0/.)

Autism Spectrum Disorder or Developmental Delays. Youth with autism spectrum disorder (ASD) or developmental delays may become agitated due to physical distress, sensory overstimulation, or mere changes in their routines. Such youth are also more vulnerable to adverse medication effects and may experience disinhibition or paradoxical agitation with benzodiazepines and diphenhydramine. Thus, behavioral strategies are especially important in addressing agitation in these youth. (Behavioral interventions for youth with ASD and ID are discussed in detail later in this chapter.) If medication is required for acute agitation, an additional dose of the child's regular standing medication is preferable to minimize the risk of adverse effects. Risperidone is a good option, as it is indicated for the treatment of aggressive behavior in ASD. If possible, IM medication and restraints should be avoided in youth with autism or developmental disabilities as such interventions can be highly distressing to these youth due to their sensory sensitivities.

Acute Psychiatric Illness. Youth may become agitated directly due to an underlying psychiatric illness, including anxiety, PTSD, or, more rarely, mania or psychosis. Youth with acute agitation due to anxiety or post-traumatic stress may benefit from diphenhydramine or a low dose of a benzodiazepine. Youth with agitation due to mania or psychosis will likely require haloperidol or another high-potency antipsychotic. However, it is important to remember that youth may become agitated for reasons that are little or completely unrelated to their psychiatric illness. They may be experiencing fear of anxiety, which could be relieved by alpha-2 agonists, diphenhydramine, or benzodiazepines.

Diphenhydramine and lorazepam are often utilized in cases of moderate agitation where the etiology of the agitation is not clear. For severe agitation of unknown etiology, olanzapine, chlorpromazine, or lorazepam combined with haloperidol may be used. These neuroleptics carry more significant risks, including extrapyramidal symptoms (EPS), neuroleptic malignant syndrome, and adverse cardiac effects, such as QT prolongation. Total daily doses should be closely monitored, and “do not exceed” instructions should be clearly written. If there exists significant concern for QT prolongation or other dysrhythmias, cardiac monitoring should be considered. Youth who are antipsychotic naive and are given high-potency neuroleptics for agitation are at increased risk for EPS. Diphenhydramine or benztropine may be utilized as prophylaxis for EPS and acute dystonic reactions, and adding these agents to haloperidol (Haldol) can magnify the sedating effect of the antipsychotic. Olanzapine is more sedating and carries a lower risk of EPS and adverse cardiac effects than other neuroleptics. However, olanzapine carries a higher risk of respiratory suppression if parenterally administered within 1 hour of benzodiazepines, and thus such concomitant administration should be avoided.

Assessment of Developmental Factors in Psychiatric Emergencies

Children are not just little adults, physically or psychiatrically. When faced with a child in crisis, the interpretation and significance of the child’s symptoms will vary greatly depending on the developmental status of the child. The developmental status includes not only the child’s chronologic or physical age but also any developmental delays, cognitive disabilities, genetic or metabolic syndromes, or deficits in speech, language, or learning ability. The ED clinician may not have access to a complete developmental history of the child, but observation of physical signs of genetic or developmental syndromes and of cognitive and motor skills can flag whether a developmental or cognitive delay may be present.

Acute symptoms can have different causes or meaning in children of different age and developmental stage. For example, a young child may present to the ED with “hallucinations” that are age-appropriate imaginary friends. An adolescent with significant ID may have a similar presentation, which would be concerning in a child of that age without ID but may still be normative in the presence of the disability. Sexual behaviors are another important example. Many children are brought to the ED for “sexual acting out.” A 5-year-old child who exposes his genitals to a peer and asks the peer to do the same is engaging in normal, common behavior for a child of that age. A similarly aged child who touches a peer’s or adult’s genitals is still within the normal range of behaviors for that age, though the behavior may reflect poor impulse control or emotional immaturity. If a 14-year-old boy is sent to the ED from school after touching a peer’s genitals without invitation to do so, this could be a predatory behavior, but it is important to assess whether a developmental or cognitive delay is present, as a 14 year old with autism or ID may be closer to the 5 year old’s developmental maturity than his 14-year-old body would suggest. Children with autism and ADHD are particularly likely to have disparities between their physical/cognitive age and their emotional/developmental maturity. Often children present to the ED with concerning behaviors or symptoms that look like psychosis or other acute illness, but a careful developmental history identifies undiagnosed autism or other developmental delay.

Assessment of Contextual Family and Social Factors Contributing to Psychiatric Emergencies

Children and adolescents present to the ED with a range of acute psychiatric complaints, from recent-onset psychosis or mania to longstanding obsessive-compulsive disorder (OCD) or autism. An important question for any child in the ED is why now—what has led the child to be brought today and not yesterday or tomorrow, and what were the precipitating factors that contributed to this? Even when a child presents with an acute or new-onset illness such as psychosis or mania, it did not begin just minutes before their arrival to the ED; rather, someone became aware of the child’s symptoms, or something happened that was the tipping point for the child or family to seek help. Understanding these contextual factors is important to stabilize the crisis. For example, an adolescent may have been abusing opiates for weeks, but today stole money from his mother and became violent toward her. Stabilizing this crisis will require treatment for the child, but also support for his mother if she is to effectively care for him going forward. A young child sent from school for a tantrum may have longstanding ADHD, but the immediate triggers for his tantrum were frustration with schoolwork related to having been transitioned this week from a special education classroom to a mainstream one. His treatment may involve changes in his ADHD medications, but also increased academic supports. A teenage girl with autism is sent to the ED for head-banging and other self-injurious behaviors, but her symptoms are related not only to her autism but also to her caregiver's recent surgery that led to the girl staying with her aunt for 3 days. A safe discharge for her will be more dependent on psychoeducation and support for her aunt than on additional medication for behavioral control.

The key contextual factors to assess are the child's home and family context; educational setting; and social factors. In the home and family sphere, the child's symptoms may be acutely worsened by shifts in the physical environment (e.g., moving into a different apartment or shelter), changes in the family makeup (birth of sibling, remarriage, divorce or death, or a move into foster care), or conflict or violence in the home. These factors may directly impact the child's mental health or may decrease the parent's ability to cope with symptoms, provide structure or supervision, or manage difficult behaviors. A parent may also be suffering from mental illness or substance abuse, or may be abusive, neglectful, or failing to protect the child from abuse or maltreatment. Understanding the child's home and family context can also highlight key strengths and supports for the child that may be useful in treatment and safety planning.

Multiple factors in a child's educational setting can precipitate an ED visit. Many children presenting to the ED are sent from school, often with a mandate for "psychiatric clearance" before the child will be allowed to return. When a child is sent from school, it is important to understand both the general academic context and learning needs of the child, to assess for unmet needs or mismatch, and the immediate triggers for the child being sent to the ED. For example, for the child with ADHD who has a tantrum after being transitioned from special education to mainstream classes, it will be useful to know as well that the tantrum occurred in math, his worst subject, after he was being teased by peers for being a “retard.” This child will have a different treatment plan than a similar boy with ADHD who has a meltdown during a fire drill, when his family recently became homeless after a fire destroyed their apartment building. Each of these boys can be discharged home but are likely to return to the ED with a similar outburst within days or weeks if the contextual factors are not addressed.

Social factors are also important to assess, particularly for adolescents. A child’s social milieu can be a crucial support in a crisis, or a trigger for acute pathology. Self-injurious behavior, substance use, risky sexual behaviors, and suicidality can have strong peer effects among adolescents and preteens. Conflicts within friendships, romantic breakups, bullying, and cyberbullying can precipitate suicidal thoughts and serious violence. Relationship violence and date rape is also unfortunately common among adolescents. Social isolation can be a marker for serious psychiatric illness, be it chronic (in the case of autistic or otherwise “odd” youth) or acute (for depressed, prodromal, or psychotic adolescents).

SPECIFIC PSYCHIATRIC EMERGENCIES IN CHILDREN AND ADOLESCENTS

Suicidality and Self-Harm

Assessment. As noted above, suicide is a leading cause of death in adolescents, and is one of the most common psychiatric presentations of youth to the ED. When a child or adolescent presents to the ED after a suicide attempt, medical stabilization is the foremost step. After this, the ED clinician must determine the patient’s suicide risk and whether inpatient hospitalization (including potentially 1:1 observation for safety) is required. Guidelines vary but generally recommend universal screening of youth presenting with psychiatric concerns, structured risk assessment, safety planning, discussion of means restriction, and referral for treatment.

Suicide screening has been shown to be effective in detecting suicide risk, as well as feasible and acceptable to patients and families. Examples of brief, evidence-based, standardized suicide screening tools for children and adolescents include the Ask Suicide-Screening Questions (ASQ), which consists of four self-report questions, takes 20 seconds to administer, and was developed for use in EDs. Another example is the Tri-Factor Screen for Youth, which consists of 22 self-report questions and takes 3 to 4 minutes to administer. The SAFE-T prompts clinicians to inquire about suicide risk and protective factors and suicidal thoughts, and based on these, suggests a risk level and possible interventions. The Columbia-Suicide Severity Rating Scale (C-SSRS) includes questions on various aspects of suicidal ideation (such as intensity and frequency), preparatory behaviors, and past suicide attempts (including aborted/interrupted attempts, and the lethality and medical consequences of attempts). The C-SSRS has been studied in adolescents with psychiatric emergencies and a pediatric version is available. While depression screening tools, such as the Patient Health Questionnaire-9 (PHQ-9) have been used to identify youth at risk for suicide, these can miss nondepressed youth with suicidal thoughts.

When a youth screens positive for suicide, clinicians should perform further assessment to ascertain the level of suicide risk and determine the level of care needed. A careful assessment of suicidal thinking and behavior is crucial. The ED clinician should ask about the ideation and planning leading up to the attempt, including the anticipated lethality of the attempt, if the child thought someone would discover or rescue them, if a suicide note or other preparation for death was made, and if the child experienced regret or ambivalence after the attempt and immediately called for help, or took steps to ensure no one would find out or be able to help them. The clinician should also assess current active or passive suicidal thoughts, plans, intent, future orientation, and reasons for living, though with recognition that youth may be motivated to minimize their current suicidal thoughts or intent for fear of being hospitalized or upsetting family members. The mental status examination can identify mental states that confer acute risk for suicide, which may be more difficult for a child to minimize or hide; these include hopelessness, preoccupation with death, severe depression, command hallucinations, and feelings of worthlessness. Other risk factors include prior suicide attempts or ideation; substance use; prior psychiatric diagnosis (depression, bipolar disorder, substance use, and conduct disorders confer particular suicide risk); family history of suicide; and access to firearms and other means of self-harm.

It may not be immediately obvious when a child in the ED has made a suicide attempt. Hanging attempts and self-poisonings may have clear suicidal intent, but other youth may present with self-cutting, overdoses of prescription drugs, threats to stab self or jump from high places, attempts to drown in the bathtub, or other acts where the true suicidal intent is unclear. It is important not to dismiss “dramatic” suicidal gestures as related to personality disorder as might be done in adult patients, as in children and adolescents even suicidal gestures and parasuicidal behaviors often demonstrate real risk. If the child engaged in deliberately harmful behavior that could be reasonably expected to end their life with the intent to end their life, the term “suicide attempt” should be used. Moreover, behaviors of unclear suicidal intent can still indicate a very real risk for suicide. Even if the child is repeatedly self-cutting with no current suicidal intent, the fact of engaging in repeated self-injury is a risk factor for later suicide attempts, and even youth who engage in suicidal behavior with ambivalence or for attention can inadvertently hurt themselves (e.g., youth who take overdoses of acetaminophen without realizing its potential lethality). In addition, children might misjudge the dangerousness of their behavior, which therefore may not accurately reflect the severity of their suicidality. A child presenting after a behavior of seemingly “low lethality,” such as an “overdose” using a small number of over-the-counter pills may nevertheless have intended to end their life.

While the determination of a suicide risk level rests on professional clinical judgment, it is important to note that assessing suicide risk or determining a risk level is not the same as “predicting suicide.” Clinicians cannot predict suicide; they can only attempt to assess imminence and severity of risk with the goal of instituting appropriate treatment. There are also no definitive guidelines on weighing suicide risk factors, distinguishing between “chronic” and “imminent” risk, or reassessing risk levels over time. Clinical judgment is imperfect, and while underestimating risk is of concern, a “high-risk” label may also carry unnecessary restrictions that are costly, aversive, and possibly even detrimental, such as separating youth from their families, peers, and communities through hospitalization. It is important to remember that while many youth have risk factors for suicide, few actually complete suicide, and that while suicide risk can be mitigated and reduced, it cannot be completely eliminated.

Management. The first decision in managing a suicidal patient is determining the need for inpatient hospitalization. Inpatient admission may be needed for the following:

- Serious suicide attempt, with or without endorsing persistent or current suicidal thoughts
- Persistent, intense suicidal thoughts or intention
- Suicidal ideation/behavior in the context of severe depression, command auditory hallucinations to self-harm, hopelessness, or self-blame
- Suicidal ideation/behavior in the context of a profound psychosocial stressor, such as recent sexual trauma or disclosure of abuse, severe bullying, relationship loss
- Suicidal ideation/behavior in the context of poor supports (family rejection/stress, no outpatient treatment)

If the clinician is uncertain whether a child is safe for discharge, it is most prudent to hold the child for reassessment of suicidal ideation and for additional collateral history. Discharge with outpatient treatment may be appropriate for youth with a low likelihood of acting on suicidal thinking and no intent or plan to do so, and who have a responsive and supportive family or other caregivers who can intervene if their suicidal thinking or other symptoms worsen.

Before discharge, youth and families should be engaged in safety planning, discussion of means restriction, and commitment to ongoing treatment. These brief interventions have been shown to significantly decrease suicide attempts in the months following an ED visit and can improve continuity of care. Several structured interventions have been developed, and they generally focus on promoting safety, providing follow-up care, and enhancing connection with trusted adults.

Safety planning is a key intervention for children and adolescents presenting to the ED with suicidality, even those not at high risk. Creation of a safety plan involves helping the child identify triggers for suicidal feelings, coping strategies that the child can use if suicidal thoughts present, whom the child can call on for support and what those individuals should do to help the child, and how to access professional help or return to the ED if needed. For youth who struggle to identify triggers, a review of the “chain of events” that led to the current presentation is helpful. The safety plan should include 24/7 crisis hotlines or text lines and should be shared with the parent or guardian. Safety plans can also be useful for self-injurious behavior without suicidal intent by modifying the plan to focus on urges to self-harm rather than on suicidal thoughts. It is also helpful to explore the function of the self-harm, which may include internal reinforcement, such as relief of strong negative emotions or feelings of numbness, or external reinforcement, such as attention from peers or family or release from social situations or conflicts with which the child is otherwise unable to cope. The ED clinician should view self-injurious behavior not as manipulative but as representing an inability to cope or communicate distress and thus demonstrating the need for treatment.

Caregiver engagement is key for safe discharge. The safety plan should be shared with the parent/guardian, and they should be educated on the importance of restricting access to lethal means (and to drugs or alcohol that increase risk for suicidal behavior), increasing supervision, and improving listening and opportunities for positive connection with the youth.

Ensuring access to follow-up care is also critical. The child should be set up with prompt (within a week) outpatient follow-up for reassessment of suicide risk and to work to stabilize the psychological and social factors that led to the suicidal thoughts and behavior. Treatments such as dialectical behavior therapy (DBT) have shown promise and appear to be effective in reducing suicide attempts and self-harm among youth. Unfortunately, these specialized therapies may not be accessible to many youths at risk for suicide; this underscores the importance of the care, counseling, and intervention provided at the ED visit.

Of note, “no suicide” or “no harm” contracts are not recommended. They have not been shown to be effective, may discourage patients from seeking help should suicidal thoughts return, and may give adults a false sense of safety.

Aggression

Assessment. Youth can present to the ED with aggressive behaviors ranging from overblown tantrums to planned, intentional, and even lethal violence. The assessment of such youth starts with managing the child’s behavior in the ED, using the techniques described above to de-escalate and contain the youth’s behavior with verbal interventions, medication, and restraint or seclusion when indicated. The child’s response to these interventions will be telling for both diagnosis and risk assessment. An adolescent who is unable to respond to verbal de-escalation and seems to be disorganized or paranoid may be psychotic or intoxicated. A child who is brought by police or school safety officers cursing and kicking but is able to speak calmly to the clinician once the uniformed officers have left may be, by contrast, an impulsive child with ADHD or PTSD who became agitated due to teasing at school but with no true intent to harm anyone.

Once the initial agitation is addressed, the clinician must attempt to understand the aggressive behavior. Understanding from the child's perspective the triggers for their aggression can make a frightening experience into a therapeutic opportunity. The clinician should assume that there is always a reason for the aggressive behavior, even if the child cannot immediately articulate it. The reason may not be obvious to the clinician or even to the child at first, but empathic and patient questioning (and information from collateral sources) can uncover the triggers, both external (such as teasing or other provocative behavior from peers, being told "no," a disruption of routine, being touched by staff or police) or internal (frustration, anxiety, paranoia, command hallucinations, trauma flashbacks). The clinician should also attempt to ascertain the context of the aggressive behavior, including evidence of planned or premeditated aggression, homicidal intent, weapons involvement, or a history of severe or escalating violence.

Aggressive acts can be reactive (in response to a trigger or perceived insult) or instrumental (planned with intent to harm or to use aggression to get something). Reactive aggression is unplanned, unprofitable, and poorly controlled. Reactive aggression may be triggered by events that appear benign to adults, like a school staff putting a hand on the child's shoulder to calm them, but which feel intrusive or threatening to the child due to their background and history. Children with autism, ID, and learning disabilities may become aggressive when frustrated because they lack the verbal skills to solve the problem and express themselves, or because they do not know any other way to ask for what they want. On the other hand, instrumental aggression is planned, profitable, and self-controlled. It often represents a reaction to narcissistic injury or stems from a belief that aggressive behavior is an appropriate means to gain desired end or exert power.

Management. Once the initial agitation is addressed, the clinician's job is to determine whether the child can be safely discharged. Risk for future violence depends in part on the nature of the violence, and on the presence or absence of underlying psychiatric illness. Evidence of planned or premeditated aggression, homicidal intent, weapons involvement, or a history of severe or escalating violence, all confer an increased risk. Risk may be influenced by parental disciplinary practices, child-rearing style, and history of substance use or criminality. Risk may also be elevated by social factors such as association with gangs or delinquent peers.

Management must be tailored based on the nature of the aggressive behavior, whether it stems from an underlying psychiatric illness, and the level of risk. Youth with a range of psychiatric illness can be prone to reactive aggression, and the treatment for each must be targeted to the illness or symptoms present (Table 35.4–4). Instrumental aggression is less common in mentally ill youth, but can be seen in conduct disorder and substance use.

Psychosocial interventions and behavioral strategies are central and should be considered first-line treatment. When aggression is reactive, the treatment involves identifying triggers, teaching skills to cope with those triggers without aggression, and addressing the core symptoms that predispose the youth to aggressive behaviors. Children with impulsive aggression benefit from parent–child skills training programs and the effective parental use of verbal warnings, prompting, “time outs,” and other behavioral interventions. Treatment of underlying conditions (such as ADHD, autism, or PTSD) is also important.

Table 35.4–4.
Examples of Psychiatric Illnesses in Children That Can Lead to Aggression

Illness/Symptom	Treatment Focus
ADHD	Address impulsivity
ODD	Set clear limits, reinforce positive behaviors
Psychosis	Treat command hallucinations/paranoia that are triggering aggression
Mania	Stabilize mania to address impulsivity/irritability
Autism or intellectual disability	Promote verbal problem solving and other coping behaviors
Conduct disorder	Set clear consequences for antisocial behaviors and reinforce positive behaviors
Substance use	Increased supervision, motivation enhancement techniques to reduce use

ADHD, attention deficit hyperactivity disorder; ODD, oppositional defiant disorder.

When aggression is instrumental, treatment involves setting clear consequences, often in collaboration with parents and law enforcement, while promoting prosocial behavior with positive reinforcements. Youth with instrumental aggression often benefit from interventions such as multisystemic therapy (MST), which addresses family relationships, parent–child communication, parenting skills, school involvement, and peer activities.

If medication is necessary, it should comprise only one aspect of a comprehensive treatment approach and should be used to address the core conditions and symptoms that predispose the youth to aggressive behaviors. While the use of emergency medication may at times be required (as discussed above), the use of PRN medication is not considered standard outpatient treatment for aggression in youth, and should be avoided. If severe aggressive behavior persists despite adequate medication trials to target the underlying psychiatric condition, antipsychotic and/or mood-stabilizing medication may be considered.

Determining whether inpatient hospitalization is indicated depends both on immediacy of risk for further aggression, and on whether the risk for aggression is due to mental illness and will be helped by hospitalization. There are youths who are engaging in criminal behavior not due to psychiatric illness, and for them, inpatient hospitalization may not be helpful; in such cases, clinicians should collaborate with the parents and law enforcement to identify other supports to contain the youth's behavior. If the child's aggression stems from abuse at home or if the family cannot safely supervise the youth, a referral to child protective services may be necessary.

If the aggressive behavior is rooted in psychiatric illness yet there is no imminent risk of harm or significant change in the youth's mental status and the youth's family can contain the behavior, the youth may be safely discharged with outpatient treatment. Prior to discharge, clinicians should engage the youth in creating a safety plan. Similar to safety plans for suicidality (see above), safety plans for aggression should identify triggers and warning signs for aggressive behaviors, coping skills to safely manage situations that may lead to aggression, and supports. Clinicians should also help families establish a crisis or emergency plan and provide 24-hour crisis hotlines in case the youth's aggressive behavior worsens. Active family engagement is critical to the successful management of youth aggressive behavior. Families of aggressive children may by struggling with stigma or feelings of overwhelm, guilt, and shame. Moreover, their child's behavior may have compromised their ability to obtain community support. Providing referrals for support groups, parent advocates, and educational resources will enhance the family's support and facilitate their ability to engage in their child's treatment.

Risky Behaviors and Substance Use

Assessment. Drug use, running away, truancy, risky sexual behaviors including prostitution, gang involvement, and other dangerous behaviors can prompt parents to bring a child to the ED desperate and looking for help. Such behaviors are uncommon in youth and can have serious and even fatal consequences; yet often the child is not immediately "dangerous to self or others." The ED clinician must assess whether the child's behaviors are age-appropriate limit testing, a reflection of underlying psychopathology, or a response to problems in the home or failures of supervision, and whether the risk of the behaviors is sufficient that hospitalization is indicated.

Determining safety is, as always, primary; this may involve assessing whether the youth is currently intoxicated with substances or at risk for withdrawal or whether a rape kit or sexually transmitted disease (STD) testing and prophylaxis are indicated. Some risky behaviors are isolated incidents that the youth will learn from and not repeat. Youth in the ED for such behaviors will usually claim that this is the case, so collateral information is crucial to assess for repeat behaviors, a decline in functioning, or associated psychiatric symptoms. Risky behaviors can be related to a range of psychiatric illnesses, including depression, mania, PTSD, ADHD, and conduct disorder, and can also be seen in intellectually disabled or autistic youth who may not comprehend the dangerousness of their actions. They can also reflect problems in the child's environment, such as a parent's inability to provide sufficient supervision due to illness or work schedule, or the child's reaction to a divorce or other major stressor in the family. In cases of truancy, the behavior may be a response to bullying, insufficient academic supports, or an inappropriate classroom placement for a child with special learning needs. Approaching the patient with an empathic and nonjudgmental stance is key to understanding why they are engaging in these behaviors and assessing their motivation to change.

Management. When risky behaviors put the child at immediate risk for harm, or when they are secondary to underlying psychopathology or problems in the environment, further intervention is needed. For a manic youth having promiscuous unprotected sex and abusing drugs, though this youth may not be suicidal or homicidal, they are endangering themselves and others and will continue to do so until their mania is treated. For an adolescent with ID who is running away and associating with gang members because his mother is medically ill and unable to supervise him, coordination with child protective services and law enforcement may be required to provide more support or place the child in a residential facility with greater supervision. Bolstering the parent's ability to relate to and engage the child will help them to prevent risky behaviors in the future. Often, the parent is so upset, angry, or anxious about the child's behavior that they are unable to effectively relate to the child, and the ED clinician must help the parent to contain their affect and set reasonable expectations. If the child and parent are unable to agree upon basic rules and behavioral expectations, then outpatient treatment has little chance for success and the ED clinician should coordinate with social work, case management, child protective services, and any services currently involved with the child to bolster the family's capacity to keep the child safe.

Psychosis

Assessment. Children and adolescents are often sent to the ED from school or from outpatient providers due to concern for psychosis, particularly for command hallucinations to harm self or others. Yet, chronic psychotic disorders like schizophrenia are extremely rare in children, instead usually developing insidiously in the late adolescent or young adult years. The ED clinician will often be in the position of discerning whether a child who is sent for "hearing voices" or being "internally preoccupied" is a child with early-onset schizophrenia, in need of inpatient hospitalization, or, more likely, that this is a red herring. Young children can report "hearing voices" as a way to express their own thoughts (particularly thoughts that are shameful, uncomfortable, or that might get them in trouble), and older children and adolescents with ID or autism often do the same. Imaginary friends are also normal in children, and again, adolescents with ID whose cognitive age is much lower than their chronologic age may endorse imaginary friends much longer than their neurotypical peers. Children with autism and ID are also often sent to the ED for being "internally preoccupied" or "responding to internal stimuli," but this may be a symptom of their autism, not of psychosis. Often these children have never been diagnosed with autism or ID if the family has not been consistent with medical providers or closely involved with the child's school. Children who have suffered a loss can report hearing or seeing their loved one in the context of grief. Anxiety disorders and OCD can present with thinking that borders on delusional, and post-traumatic flashbacks can look like the child is experiencing hallucinations. Depression and mania can be accompanied by either nihilistic and shameful or grandiose delusions and, at times, hallucinations berating or shaming the patient. An accurate diagnostic assessment is crucial as the treatment will vary greatly depending on the diagnosis and clinical formulation. Collateral information about the child's life, family, functioning, recent stressors, and prior traumas is important, particularly if the child is disorganized or cognitively limited and unable to give coherent answers.

If the clinician determines that the child is truly psychotic, organic causes must be ruled out. A complete physical examination, lab analysis including complete blood count, electrolytes, liver, kidney, and thyroid functioning, full toxicology screening, and brain imaging are indicated, as well as other tests (Lyme, erythrocyte sedimentation rate, rapid plasma reagin, lumbar puncture for cerebrospinal fluid analysis, etc.), depending on the patient's history and associated symptoms.

Management. Patients who are truly psychotic can be dangerous in the ED as they may act erratically and expose themselves or others to danger. These patients often require a 1:1 sitter for observation and close monitoring. Youth with psychosis can rarely be stabilized in a few hours in the ED and generally require inpatient hospitalization for stabilization. If the psychosis is secondary to substance intoxication, the symptoms may clear when the intoxication subsides, but with the new designer drugs such as synthetic cannabinoids (k2, spice) and substituted cathinones (bath salts), substance-induced psychosis is of much longer duration than with more traditional drugs and may require inpatient hospitalization for safety. If the psychotic symptoms are due to severe depression, mania, or OCD, or if the ED clinician lacks the collateral information and time to observe the patient to make a definitive diagnosis, hospitalization may be the best option to keep the child safe and treat the child carefully.

If the reported psychotic symptoms are in fact a manifestation of ID or immaturity, anxiety, grief, trauma, or a normal reaction to stress, the child can likely be discharged with a referral for outpatient treatment and reassurances for the child and family. Families in these cases are often deeply relieved and should be reminded that even though their child is not being diagnosed with schizophrenia, follow-up psychiatric care is still important. Depending on the underlying psychopathology, the ED clinician may also consider starting treatment from the ED (e.g., medication for anxiety or psychoeducation about trauma and flashbacks).

Youth with Autism and Intellectual Disability

Assessment. Children and adolescents with ASD and/or ID, particularly those with limited or minimal verbal communication skills, are often brought to the ED for behavioral symptoms, including aggression, self-injury, inappropriate sexual behaviors, sleep cycle reversal, or running away. These cases can be difficult for the ED clinician because youth with ASD or ID can be medically and psychiatrically complicated, and also they often have a difficult time in the ED due to the unpredictability and unfamiliarity of the ED, the noise, bright lights, and the need for blood draws and physical examinations. Children with ASD or ID often are blocked from admission to inpatient psychiatric units due to their diagnosis, leaving them boarding in the ED for days awaiting an inpatient bed on a specialized unit.

When youth with ASD or ID present to the ED, the first step to keep them safe is to understand what works at home, including preferred toys and caregivers, communication tools (such as picture communication tools and sign language), and soothing techniques. Understanding what typically triggers negative behaviors at home will also help the ED clinician to avoid outbursts in the ED. When a child with ASD or ID is brought to the ED for problematic behaviors, it is important first to understand if the behavior is longstanding or new. Longstanding behaviors such as self-stimulatory behaviors, impulsive aggression, or head banging may be related to the underlying ASD/ID and indicate that the community treatments are insufficient or ineffective and should be re-evaluated. New behaviors are often triggered by changes in the environment (changes in caregiver or home environment, school vacation, changes in school placement) or physical/medical problems (constipation, dental problems, ear infections, seasonal allergies, pain, fatigue, sedation due to polypharmacy, adverse effects from medication such as akathisia, onset of puberty). More rarely, the youth with ASD/ID can have the new onset of a psychiatric disorder such as a mood or anxiety disorder, but environmental changes and medical problems should be ruled out before a new diagnosis is made.

Management. Keeping patients with ASD and ID safe in the ED requires the removal of unnecessary stimuli (noise, light, needle sticks), clear communication (simple, concrete, and written or picture based if possible) of what to expect from the evaluation, attention to physical comfort (pain, fatigue, hunger, constipation), and use of soothing activities or family members when possible. Oftentimes, the parents of ASD and ID youth are the “experts” in knowing which strategies calm and help their child. Family members may also serve as “interpreters,” helping ED staff understand the significance of their child’s actions and behaviors. Involving child life specialists, occupational therapists, speech–language therapists, and other health professionals may help tailor ED care to the youth’s needs. Medication is often used to sedate ASD and ID patients in the ED. However, this is dangerous as they may have paradoxical reactions to medications such as diphenhydramine and benzodiazepines yet have little positive effect from antipsychotics. Behavioral interventions, while more complicated to execute, are much more likely to be effective.

Many youths with ASD/ID can be stabilized quickly if the medical or physical etiology of their distress is discovered and resolved. Effective pain control, dental care, management of insomnia and constipation, and resolution of polypharmacy and medication adverse effects can be accomplished in the ED and provide great relief to patients and families. Often when these problems are resolved, families discover that their child, who had been increasingly aggressive or difficult to manage for a long time, is now calm and comfortable.

As there are few inpatient units that accept youth with ASD or ID (particularly if they are aggressive or engage in self-injurious behaviors), ED clinicians should familiarize themselves with community services available for people with intellectual and developmental disabilities. Often home-based and respite services are available to patients who qualify, though families may not know how to access these services.

Child Abuse or Neglect

Assessment. Unfortunately, it is all too common that youth who present to the ED for psychiatric complaints are found to be the victims of child abuse or neglect. Children with psychiatric and developmental disorders and behavioral disturbances are more likely to be abused and neglected, and abuse and neglect can contribute to the development of a range of psychiatric illnesses. Other risk factors include poverty, single-parent households, large family size, social isolation, parental substance use, and personal history of abuse for the parent. Assessing for abuse or maltreatment in the ED is very difficult, because children rarely admit to abuse, due to shame, guilt about betraying the family if the abuser is a family member, or fear of reprisal or removal from home. Clear evidence of abuse is rare. Children can present with medical symptoms such as unexplained physical injuries, signs or symptoms of sexually transmitted diseases, or blood in their underwear, and then the psychiatric clinician consults the pediatric team for further evaluation for abuse. Children who are being abused or maltreated can also present to the ED with a wide range of psychiatric symptoms.

The ED clinician must ask in a careful, thorough, nonjudgmental way about any abuse or maltreatment in the home, school, or elsewhere. The child should be interviewed separately from the parent and should be reassured that he will be safe and protected no matter what he discloses. It is important to interview the child gently, starting with more subtle queries into the child's home life (such as who lives in the home, who prepares meals, who helps with homework) and then moving to questions about who sleeps in bed with whom, who gets into arguments and how they are resolved, who yells the most at home, what punishments there are for misbehavior, and other contextual questions before asking directly about abuse. Collaboration with the pediatrician, if any physical injuries are present, can also be helpful as these may be a marker for abuse, or the child may require a genital examination if there is a concern for sexual abuse. When child abuse pediatrician consultation services are available, it can be helpful to obtain a specialized interview and physical examination to discover and document evidence of abuse. At the very least, when abuse is suspected, child protective services should be consulted immediately. It is not the ED clinician's job to determine whether abuse has definitely occurred (that is the job of child protective services). But if there is any concern, as a mandated reporter, the ED clinician is obligated to call child protection. Parents should be notified that such a call is being made and why; even if the parent is the perpetrator or failed to prevent abuse, they are often remorseful and looking for help if the clinician is able to empathize with why the abuse has happened (such as a battered mother unable to prevent the abuse of her children, or an overwhelmed parent struggling to discipline a child with significant behavioral problems). The parent may need help as much as the child does, so maintaining an empathic alliance with the parent will give them and the child the best chance for recovery.

Medical child abuse can manifest with exaggeration or fabrication of a child's psychiatric symptoms. This is very rare and difficult to diagnose but can manifest in the ED with hospital shopping, strange recurrences of symptoms despite good treatment, requests for specific medications, tests, or procedures, and refusal to consent for the ED clinician to contact outpatient providers or other key collateral sources. The ED clinician should be cognizant that medical child abuse can present with psychiatric symptoms and should utilize tools for collateral information such as state prescription databases, pharmacy, and insurance records if such an etiology is suspected.

Management. Collaboration with child protective services will help to ensure the child's immediate physical safety, and this may require keeping the child in the hospital until an investigation is done and a safe place is found to which the child can go. The ED clinician can also play an important role in assessing for psychiatric sequelae of maltreatment, providing psychoeducation to the child and family (or foster family), and ensuring that the child receives appropriate psychiatric follow-up care. If the child presents to the ED with behavioral problems, helping the family or caregiver to identify how these behaviors may be triggered by or related to the trauma or abuse can help avoid such behaviors (and repeat trips to the ED) in the future. Family therapy can be a very helpful referral in these cases. Anticipatory guidance can also be helpful for both the child and caregiver to understand what types of symptoms may come up in the future and how to access help if needed. If a child is to be discharged, clinicians should consider the child's environment and circumstances after leaving the ED. Whether the child will be discharged to the custody of child protective services, a foster home, a group home, a residential facility, or returning home with their caregivers will impact their needs for treatment referrals and aftercare. If a child is in a chaotic or unpredictable situation, this should prompt reconsideration of any medications that require frequent monitoring and increased attention to ensure access to follow-up care.

Eating Disorders

Assessment. Eating disorders are a source of significant morbidity and mortality in children and adolescents. When an eating disorder is discovered by a parent or teacher, the adult will often send the child to the ED for medical and psychiatric evaluation due to concern for the medical consequences of eating disorders as well as their association with suicidality and self-harm. The ED clinician may also discover an undiagnosed and unrecognized eating disorder in a child or adolescent who presents to the ED with dizziness, nausea, syncope, suicidal ideation, or self-injury. Younger children and boys may present to the ED with unexplained weight loss, and often no one has asked them about restricting, purging, or excessive exercise because clinicians are unused to seeing eating disorders in prepubertal patients or in males.

Management. Children and adolescents with eating disorders may require psychiatric hospitalization for safety due to suicidality or self-harm, or they may require medical admission for medical stabilization. Collaboration with pediatrics is crucial in all cases to determine the treatment plan and need for hospitalization. Assisting the family to identify resources for support and education is also important, as an effective treatment of eating disorders is a long-term process that requires close family involvement for the greatest success.

Psychosocial Crisis

Assessment. Children are profoundly affected by a parent’s state of mind, mental health, and relationships. Parental mental illness, divorce, or severe conflict between parents can manifest in emotional and behavioral problems in the child. These stressors put the child in a desperate situation, and for a child with underlying psychiatric or behavioral symptoms, these family crises can be catastrophic. These situations can be difficult to assess. The parent is often unable to objectively reflect the stress on the child or even on the parent’s own internal life and struggles. The child may lack the emotional maturity to describe clearly the problems at home, feel guilty, ashamed, protective of the parent, or be afraid of being separated from the family. “Listening between the lines” is important, and when the given history does not make sense, it can be an indication that hidden stressors at home may be at the root of the child’s presenting problem. The child and parents should each be interviewed separately, and the clinician should approach the parent empathically and make it clear that their purpose is to help the child, not to be intrusive, judge the parents, or report them. If parents give deeply discrepant reports or perspectives on the child’s symptoms, blame the other parent or the child, or belittle or berate the other parent or child, this suggests a significant conflict in the family that may be triggering the child’s symptoms or behaviors or making it impossible for the family to effectively manage those symptoms or behaviors. When speaking with the child, vague or inconsistent statements that seem odd or unusual to the interviewer (such as a child who states that the parent is sleeping most of the day and cannot be awoken) may serve as “red flags” and reassurance and specific questioning to the child may help them share the problems at home.

Management. Safety is, as always, the first priority; if a parent is severely mentally ill or if the conflict in the home is so extreme, then child protective services should be notified. The ED clinician may, in extreme cases, also need to refer the parent for their own emergency assessment. In less severe cases, with an empathic and strong rapport with the parents, the ED clinician may be able to help the parent to understand their role in the child’s symptoms and their need for help and treatment. Getting the parent into individual or family treatment and helping the child to voice and cope with the stress at home will reduce the child’s distress and the strain on the family.

LEGAL CONSIDERATIONS AND CONFIDENTIALITY

Clinicians working with youth in emergency psychiatric settings should be familiar with the laws and institutional regulations that govern emergency psychiatric treatment and hospitalization and also those that apply to the evaluation and treatment of minors. Many states allow for emergency psychiatric evaluation and treatment of a minor without a parent’s consent, and several allow adolescents to access outpatient psychiatric care without the parent’s knowledge or consent. If a child requires hospitalization, each state has statutes that dictate a standard for psychiatric hospitalization, generally differentiating between voluntary, involuntary, and emergency or diagnostic admissions. When working with minors, these laws generally dictate that a parent or guardian can hospitalize a child or adolescent under a “voluntary” admission, even without the minor’s assent, and often provide a standard by which the decision to hospitalize a child or adolescent can be made without the parent’s consent. Importantly, separate consent is commonly required in order to treat a hospitalized youth with medication. If parental rights have been surrendered or terminated, a foster care or state child welfare agency may have a court order providing them with the legal authority to provide consent for a child. States also have procedures to bypass consent if a parent or guardian cannot be reached or is refusing to consent to treatment deemed essential to the child’s well-being. Clinicians working in the ED should know these statutes as they apply in their area, as well as the local and institutional policies for consent and hospitalization.

Clinicians should also be familiar with mandated reporting laws for cases of suspected physical or sexual abuse or other maltreatment. If a patient reports serious intent to harm another person, the clinician has a responsibility to warn this individual, which supersedes the patient's right to confidentiality. Navigating the child's confidentiality vis-à-vis their parents can also be complicated. If the child is in imminent danger or a danger to others if the parent is not notified, the clinician should tell the parent, but first should attempt to understand the child's desire for secrecy. Often a child fears the parent will react with reprisal, anger, or distress and may be able to share with the parent with the clinician's support and guidance.

MODELS OF CARE AND INNOVATIONS

In recent years, new developments and innovations have sought to enhance and improve the quality of care for youth with mental health emergencies. Medical EDs have dedicated space with specialized behavioral health staff, and clinical pathways to standardize assessment, intervention, and disposition have been implemented. Hospitals have developed specialized youth psychiatric EDs modeled on adult psychiatric EDs, such as the Comprehensive Psychiatric Emergency Program (CPEP) model. These programs allow for a 3- to 5-day period of extended observation and active treatment in a specialized psychiatric care environment. They seek to stabilize acute psychiatric symptoms and thus avert the need for inpatient psychiatric hospitalization. Telepsychiatry has allowed child and adolescent psychiatrists to provide emergency mental health care to youth in EDs without access to in-person child and adolescent psychiatry staffing. Other service developments offer alternatives to ED visits. Mobile crisis programs, triage phone lines, and psychiatric triage teams provide youth with immediate and effective mental health care in the community. Intensive community-based treatment programs, such as MST or DBT, provide youth and families with a high level of outpatient support and access to clinicians for emergency coaching. These developments have improved the quality and raised the standards of care for youth experiencing mental health emergencies. As the number of mental health–related ED visits among youth continues to rise, investment in training, resources, and new innovations should seek to continue to improve the quality of psychiatric emergency care for children and adolescents.

CONCLUSION

Good ED care for children in psychiatric crisis means an accurate, developmentally informed diagnosis and formulation, stabilization of the crisis, and a treatment plan that helps the child and family to overcome stigma and systems barriers to access the right care. The practical challenges of connecting a patient to appropriate treatment should not be underestimated either. If a patient requires inpatient hospitalization, it can be difficult to find an open bed on a psychiatric unit for children and adolescents, but treatment on an adult unit or "boarding" in pediatrics should be avoided as both can be countertherapeutic. If the child will be waiting in the ED for several days before an appropriate inpatient bed is available, the ED provider should begin treatment (therapy or medication) to the degree that is possible in the ED.

When a patient is safe for discharge, the ED clinician should make every effort to connect the patient to an outpatient provider, not just give the parent a phone number for a provider or for their insurance panel. Connecting to outpatient care can be a very difficult and lengthy process and some parents cannot navigate the system due to insurance issues or the parent's own cognitive or language difficulties. Providing the patient with an actual appointment for follow-up has been shown to decrease future dangerous behaviors and so is crucial for the safe discharge of high-risk patients. ED clinicians should also familiarize themselves with case management services; child welfare services; community supports such as family resource centers; social services within the school, shelter, and family court systems; and respite and crisis services available through the state or city mental health, child welfare, or developmental disabilities services providers. These additional services can be crucial for avoiding a repeat presentation to the ED, particularly if psychosocial or family factors have influenced the current symptoms.

FURTHER READINGS

Asarnow JR, Babeva K, Horstmann E. The emergency department: challenges and opportunities for suicide prevention. *Child Adolesc Psychiatr Clin N Am*. 2017;26(4):771–783.

Bath E, Njoroge WFM. Coloring outside the lines: making Black and Brown lives matter in the prevention of youth suicide. *J Am Acad Child Adolesc Psychiatry*. 2021;60(1):17–21.

Brent DA. Master clinician review: saving holden caulfield: suicide prevention in children and adolescents. *J Am Acad Child Adolesc Psychiatry*. 2019;58(1):25–35.

Bridge JA, Horowitz LM, Fontanella CA, et al. Age-related racial disparity in suicide rates among US youths from 2001 through 2015. *JAMA Pediatr*. 2018;172(7):697–699.

Bridge JA, Marcus SC, Olfson M. Outpatient care of young people after emergency treatment of deliberate self-harm. *J Am Acad Child Adolesc Psychiatry*. 2012;51:213.

Case SD, Case BG, Olfson M, Linakis JG, Laskis EM. Length of stay of pediatric mental health emergency department visits in the United States. *J Am Acad Child Adolesc Psychiatry*. 2011;50:1110.

Chun TH, Katz ER, Duffy SJ, Gerson RS. Challenges of managing pediatric mental health crises in the emergency department. *Child Adol Psych Clin N Am*. 2015;24:21.

Chun TH, Mace SE, Katz ER; American Academy of Pediatrics Committee on Pediatric Emergency Medicine, American College of Emergency Physicians, Pediatric Emergency Medicine Committee. Evaluation and management of children and adolescents with acute mental health or behavioral problems. Part I: common clinical challenges of patients with mental health and/or behavioral emergencies. *Pediatrics*. 2016;138(3).

Chun TH, Mace SE, Katz ER; American Academy OF Pediatrics Committee on Pediatric Emergency Medicine, American College of Emergency Physicians, Pediatric Emergency Medicine Committee. Evaluation and management of children with acute mental health or behavioral problems. Part II: recognition of clinically challenging mental health related conditions presenting with medical or uncertain symptoms. *Pediatrics*. 2016;138(3).

David-Ferdon C, Vivolo-Kantor AM, Dahlberg LL, Marshall KJ, Rainford N, Hall JE. *Youth Violence Prevention Resource for Action: A Compilation of the Best Available Evidence*. National Center for Injury Prevention and Control, Centers for Disease Control and Prevention; 2016.

Dolan MA, Fein JA; Committee on Pediatric Emergency Medicine. Pediatric and adolescent mental health emergencies in the emergency medical service system. *Pediatrics*. 2011;127:e1356–e1366.

Gerson R, Malas N, Feuer V, Sillver GH, Prasasd R, Mroczkowski MM. Best practices for evaluation and treatment of agitated children and adolescents (BETA) in the emergency department: consensus statement of the American Association for Emergency Psychiatry. *West J Emerg Med*. 2019;20(2):409–418.

Grupp-Phelan J, Harman JS, Kelleher K. Trends in mental health and chronic condition visits by children presenting for care at U.S. emergency departments. *Public Health Rep*. 2007;122:55.

Haddad F, Gerson R. *Helping Kids in Crisis: Managing Psychiatric Emergencies in Children and Adolescents*. American Psychiatric Publishing; 2015.

Henderson SW, Gerson R, Phillips B. What is “high risk” and what are we actually supposed to do about it? *J Am Acad Child Adolesc Psychiatry*. 2019;58(6):561–564.

Heppell PJ, Rao S. Social services and behavioral emergencies: trauma-informed evaluation, diagnosis, and disposition. *Child Adolesc Psychiatr Clin N Am*. 2018;27(3):455–465.

Horowitz L, Tipton MV, Pao M. Primary and secondary prevention of youth suicide. *Pediatrics*. 2020;145(Suppl 2):S195–S203.

Jones L, Bellis MA, Wood S, et al. Prevalence and risk of violence against children with disabilities: a systematic review and meta-analysis of observational studies. *Lancet*. 2012;380(9845):899–907.

Kalb LG, Stapp EK, Ballard ED, et al. Trends in psychiatric emergency department visits among youth and young adults in the US. *Pediatrics*. 2019;143(4):e20182192.

King CA, Horwitz A, Czyz E, Lindsay R. Suicide risk screening in healthcare settings: identifying males and females at risk. *J Clin Psychol Med Settings*. 2017;24(1):8–20.

Knapp P, Chait A, Pappadopulos E, Crystal S, Jensen PS; T-MAY Steering Group. Treatment of maladaptive aggression in youth: CERT guidelines I. Engagement, assessment, and management. *Pediatrics*. 2012;129(6):e1562–e1576.

Lo CB, Bridge JA, Shi J, Ludwig L, Stanley RM. Children’s Mental Health Emergency Department visits: 2007–2016. *Pediatrics*. 2020;145(6):e20191536.

McEnany FB, Ojugbele O, Doherty JR, McLaren JL, Leyenaar JK. Pediatric mental health boarding. *Pediatrics*. 2020;146(4):e20201174.

Mohr WK, Mahon MM, Noone MJ. A restraint on restraints: The need to reconsider the use of restrictive interventions. *Arch Psychiatr Nurs*. 1998;12(2):95–106. https://doi.org/10.1016/S0883-9417(98)80059-9

Mroczkowski MM, Havens J. The state of emergency child and adolescent psychiatry: raising the bar. *Child Adolesc Psychiatr Clin N Am*. 2018;27(3):357–365.

Plemmons G, Hall M, Doupnik S, et al. Hospitalization for suicide ideation or attempt: 2008–2015. *Pediatrics*. 2018;141(6):e20172426.

Scotto Rosato N, Correll CU, Pappadopulos E, Chait A, Crystal S, Jensen PS; Treatment of Maladaptive Aggressive in Youth Steering Committee. Treatment of maladaptive aggression in youth: CERT guidelines II. Treatments and ongoing management. *Pediatrics*. 2012;129(6):e1577–e1586.

Torio CM, Encinosa W, Berdahl T, McCormick MC, Simpson LA. Annual report on health care for children and youth in the United States: National estimates of cost, utilization and expenditures for children with mental health conditions. *Acad Pediatr*. 2015;15(1):19–35.

Wharff EA, Ginnis KM, Ross AM. Family-based crisis intervention with suicidal adolescents in the emergency room: a pilot study. *Soc Work*. 2012;57:133.

36

Genetics in Child Psychiatry

GIL D. HOFTMAN, M.D., PH.D., JAMES T. MCCRACKEN, M.D., AND ERIKA L. NURMI, M.D., PH.D.

One of the primary goals of human genetics is to identify DNA sequence variations that impact the onset and progression of human disease. Individual susceptibility for almost all psychiatric conditions is driven, at least in part, by genetic variation. Before the draft of the human genome in 2000 and subsequent completion of the Human Genome Project (HGP) in 2003, the role of genetics in medicine and psychiatry primarily focused on the identification of single-gene mutations that played causative roles in relatively rare Mendelian diseases such as neurofibromatosis, Huntington disease, and tuberous sclerosis. However, classic genetic approaches for Mendelian disease failed when applied to most common human diseases (e.g., heart disease, diabetes, asthma), including neuropsychiatric disorders. These common disorders often have high heritability, which is the proportion of variation in a population trait that can be attributed to inherited genetic factors (see Table 36–1). However, common disorders, despite high heritability, are complex in nature and caused by the hypothesized interaction of multiple genetic and environmental influences. While the HGP provided scientists with the 3 billion bases of primary DNA sequence comprising the human genome, assembling the puzzle pieces of complex genetic disease requires an understanding of human variation, DNA regulation, and other complexities that remains incomplete (see Limitations).

Fueled by continuing rapid technologic advances along with the efforts of large consortia (e.g., Psychiatric Genomics Consortium) and collaborative projects (e.g., ENCODE, PsychENCODE, 1,000 Genomes and International HapMap Projects), the amount of available genomic information has exploded over the past two decades. Identifying susceptibility genes underlying psychiatric illness has proven more vexing in psychiatry than for complex diseases in other areas of medicine (e.g., hypertension, inflammatory bowel disease, diabetes, cancer). One barrier to these efforts is the phenotypic complexity of psychiatric illness and the likelihood that most categories of psychiatric illness, as classified in the DSM-5, are heterogeneous in nature. In addition, most studies aimed at identifying susceptibility genes underlying psychiatric illness have been underpowered and suffered from hypothesis-driven bias. International consortia with growing sample sizes employing hypothesis-free genome-wide approaches have begun to identify multiple risk loci for neuropsychiatric disorders. These studies have highlighted the substantial overlap between categorical DSM diagnoses and implicate common risk pathways for susceptibility spanning traditional diagnostic boundaries. Recent research directions, such as the NIH Research Domain Criteria (RDoC), attempt to characterize psychiatric illness on dimensional and neurobiologic measures that may more accurately reflect underlying biology.

One major challenge remains to leverage the wealth of available data to advance the current understanding of genetic influences in mental illness, to ground these findings in neurobiology across development, and to make them clinically actionable. Such knowledge, once in hand, should provide opportunities for improved diagnosis, treatment, and prevention. Therefore, characterizing the relationship between DNA sequence variation and illness predisposition will likely help identify foundational processes of biologic pathophysiology and provide targets for novel strategies for prevention and treatment. Despite the complexity of childhood- and adolescent-onset neuropsychiatric disorders, clear progress has been and continues to be made, and that will be a key focus of this chapter.

Genetic information is also being applied in other areas to advance medicine and psychiatry. Knowledge of genetic background for treatment selection and the impact of genetic variation on drug response is referred to as *pharmacogenomics* or *pharmacogenetics.* Such fields advance the promise of “precision medicine”—the personalization of treatment based on genetics and other individual factors, which in other areas of medicine has yielded important, widely applied improvements in clinical care. While psychiatry lags behind in providing clinically actionable findings, public and professional enthusiasm for using genetic information to guide prescribing urges the field forward.

With the continued rapid advances in technologic, computational, and statistical methods in genetics, it is likely that genetic information will become even more important in understanding and diagnosing psychopathology, and in directing treatment recommendations. Psychiatrists need to appreciate the nature of the genetic discoveries as they emerge and to understand how to interpret and apply them in their practice; to do so requires basic knowledge of the research methods applied and the strengths and limitations of these findings. An update on current evidence in the genetics of childhood-onset psychopathology forms a needed background. The goal of this chapter is to highlight progress in genetic research on the etiology of pediatric-onset mental illness as it points to future opportunities to inform clinical practice and treatment practices.

THE GENETIC BASIS OF PSYCHIATRIC DISEASE

DNA Variation and Contribution to Disease Risk

In each individual human genome, there are millions of DNA variants, known as *polymorphisms*. Most changes appear to be neutral with respect to human illness but likely are critical for the diversity, robustness, and evolution of the human species. The earliest appreciated disease-associated genetic variations were cytogenetically visible interchromosomal translocations that disrupt or delete genes or regulatory elements, thus contributing to disease. More recently, large *segmental duplications* (SDs) or *copy number variations* (CNVs), involving extensive duplicated, repeated, inverted, or deleted segments of DNA larger than 1 kilobase (KB) in size, are observed commonly in affected and unaffected individuals. When certain CNVs are observed repeatedly in a population of individuals with a disease, the CNVs can point to disease-causing loci in the genome, such as those CNVs producing the Prader–Willi syndrome (15q11–q13 paternally derived deletions), some cases of syndromic autism (15q11.2–q13 maternal duplications), and William syndrome (7q11.23 deletion). Whereas gross chromosomal derangements typically affect multiple genes and produce rare but characteristic syndromes, subtler variations are more likely to be involved in complex disease susceptibility. CNVs are likely to contribute to disease through changes in gene dosage or disruption of genes or regulatory elements in the genome.

Table 36–1.
Definitions of Key Terms Used Throughout the Chapter

ATAC-seq	Identifies regions of the genome that are not histone bound in cell nuclei and are open to gene regulatory processes.
Common variant	A polymorphism with a minor allele frequency >1% (occurs in >1% of the population).
Copy number variant (CNV)	Structural chromosomal variants typically greater than 1 kilobase in size.
DNA Microarray	A small glass slide spotted with short DNA probes. Pieces of sample DNA hybridize to the chip and then are detected by scanning software. Chips can contain approximately 1.8 million different genetic markers.
Genome-wide association study (GWAS)	A test of whether particular alleles at SNP markers spanning the entire genome are associated with disease, traits, or other phenotypes of interest. A GWAS is most often conducted using chip microarrays in many (tens to hundreds of thousands) of individuals.
Heritability	A statistic that measures how much of the variation of a trait can be attributed to variation of genetic factors. It can be estimated by comparing individual phenotypic variation among related individuals in a population, by examining the association between individual phenotype and genotype data, or by modeling summary-level data from GWAS.
Hi-C	A chromosome-conformation capture

	method that can capture genomic regions that are near each other in cell nuclei. Hi-C does this in an all-interactions-to-all-interactions manner.
Liability-Threshold Model	A large number of variables (e.g., genes) is summed to create a risk score; the observed outcome is determined by whether that risk score is larger or smaller than a threshold (e.g., having or not having a SCZ diagnosis).
Linkage Disequilibrium (LD)	The nonrandom association of alleles at different loci due to recombination interacting with selection, mutation, and genetic drift.
Pleiotropy	Genes that affect multiple traits or increase risk for multiple disorders. For example, many of the same genes increase the risk for SCZ, BIP and ASD.
Polygenicity	A trait that is influenced by multiple genes or a disorder that is influenced by many genes of small individual effects. For example, thousands of genes, each with small effect, increase the risk for SCZ.
Polygenic Risk Score (PRS)	An estimate of an individual’s genetic liability to a trait or disease, calculated according to their genotype profile and relevant GWAS data.
Pharmacogenetics	How variation in a single gene influences the response to a single drug.
Pharmacogenomics	How the entire genome influences drug

	responses.
Quantitative Trait Loci (QTL)	Regions of DNA (loci) associated with variation of a particular trait or phenotype. For example, expression QTLs (eQTLs) identify genetic predictors of gene expression.
Rare Variant	Alternative forms of a polymorphism that are present with a minor allele frequency of <1%.
RNA Sequencing (RNA-seq)	Identifies the amount of all RNA molecules in a cell or tissue. Can also capture splicing and isoform-level information.
Single Nucleotide Polymorphism (SNP)	Variation at a single position (single base or nucleotide) in a DNA sequence.
Whole Exome Sequencing (WES)	A type of genome sequencing focused on protein-coding parts of the genome (~3%); good for finding gene-disrupting rare variants.
Whole Genome Sequencing (WGS)	Resequencing of the genome can identify all kinds of genetic variation in concept.

The contribution of CNVs to complex psychiatric illness is becoming increasingly appreciated in recent years. CNVs in aggregate are common (about 12% of the genome) and often polymorphic, with individuals carrying from zero to many copies of a given segment of DNA. As of 2019, there are 6,450 CNVs annotated in the human genome. Small DNA stretches that are deleted or inserted, collectively referred to as *“indels,”* are even more frequent, with more than 450,000 annotated indels in the human genome. CNVs are now understood to represent another genetic mechanism contributing to complex disease risk. For example, the DiGeorge/velocardiofacial syndrome (22q11.2 heterozygous deletion CNV) increases the odds of schizophrenia (SCZ) by about 50-fold compared with the general population. When compared to typically developing controls, total CNV burden is elevated in individuals with autism spectrum disorder (ASD) and SCZ. Although not all CNVs are pathogenic, disease-associated CNVs are larger and more likely to encompass gene-rich regions of the genome.

The simplest and most common type of individual genetic variation is the *single-nucleotide polymorphism* (SNP), widely employed in genetic studies. SNPs are changes of a single DNA base pair (e.g., adenine [A] to guanine [G]) that occur in approximately one per every 1,200 bases throughout the genome. Smaller indels and SNPs can influence disease through more targeted changes in individual genes. SNPs can directly alter protein structure through changes that alter amino acid sequence (nonsynonymous changes) in the gene product or distort start or stop signals. Even synonymous (silent) polymorphisms that do not alter amino acid–coding potential can influence RNA stability, splicing, targeting, editing, and efficacy of translation—many of the effects of these synonymous variants are unknown. Although noncoding changes are often overlooked due to presumed lack of function, they can exert powerful effects on the regulation of gene expression and transcript processing, as illustrated by the noncoding risk alleles described in macular degeneration and Crohn disease. Recent technologic advances have allowed for deeper interrogation of DNA regulatory elements, opening the door to studying and understanding noncoding regions of the human genome. For most mental illness, genome-wide studies have suggested an underlying genetic architecture consisting of both common ancestral alleles, each of which carries a small disease risk, and rare, highly penetrant mutations, that if present, are likely to result in disease.

Finding Risk Genes

Common risk variants are challenging to identify due to their small effect sizes and presence in healthy people, while causality of rare risk variants is difficult to prove due to the low signal-to-noise ratio. In both cases, very large sample sizes are needed, but discrete technologies are best suited to uncover common versus rare variants. The basic strategy behind studies aimed at disease gene discovery is to identify regions of the genome that are shared between affected individuals; the genes that underlie the disease must reside within these shared regions.

Neutral polymorphisms, although not directly contributing to the disease phenotype, can serve as markers for functional risk alleles since polymorphisms are not inherited independently but in large DNA blocks, a phenomenon known as *linkage disequilibrium* (LD). LD can persist for substantial distances across a chromosome and for numerous generations but decay as a function of the rate of recombination between homologous chromosomes. Therefore, it is not always necessary initially for an investigator to genotype a risk allele directly to find association with disease but merely to genotype one of the often many markers “linked” with the risk allele. Although an associated marker may have no direct role in disease etiology, it is present on an ancestral portion of DNA that also contains the disease locus and can thereby pinpoint the genomic location of the disease allele.

Two major shifts in the psychiatric genetics field are responsible for recent success in finding risk genes for mental illness after decades of struggle. First the development of genome-wide, high-throughput technologies enabled affordable, high-resolution data to be generated on large numbers of samples. In parallel, sufficiently large databases began to be collected due to the formation of large, worldwide collaboratives sharing tens to hundreds of thousands of patient samples. Modern successful genome-wide approaches, such as genome-wide association mapping, copy-number variant mapping, polygenic risk analyses, and whole genome or exome sequencing, overcame the limitations of and thus have supplanted prior limited gene-finding strategies like linkage analysis and candidate gene studies.

Genome-Wide Association Studies. In the new millennium, microarray (“gene chip”) technology has made ambitious GWAS possible; chips are available that can affordably assay millions of SNPs simultaneously. Given sufficient sample sizes, GWAS effectively detects common risk variants of small effect. Since only common, known SNPs are included on arrays, however, GWAS unfortunately, fails to capture the contribution of rare or novel variants. Currently, the consensus *P* value for significance in GWAS studies is set at 5×10^{-8}, correcting for the 50,000 predicted false positives that occur when a million tests are performed. While many psychiatric GWAS fail to meet this stringent level of significance, several large consortia have been successful by recruiting tens to hundreds of thousands of participants.

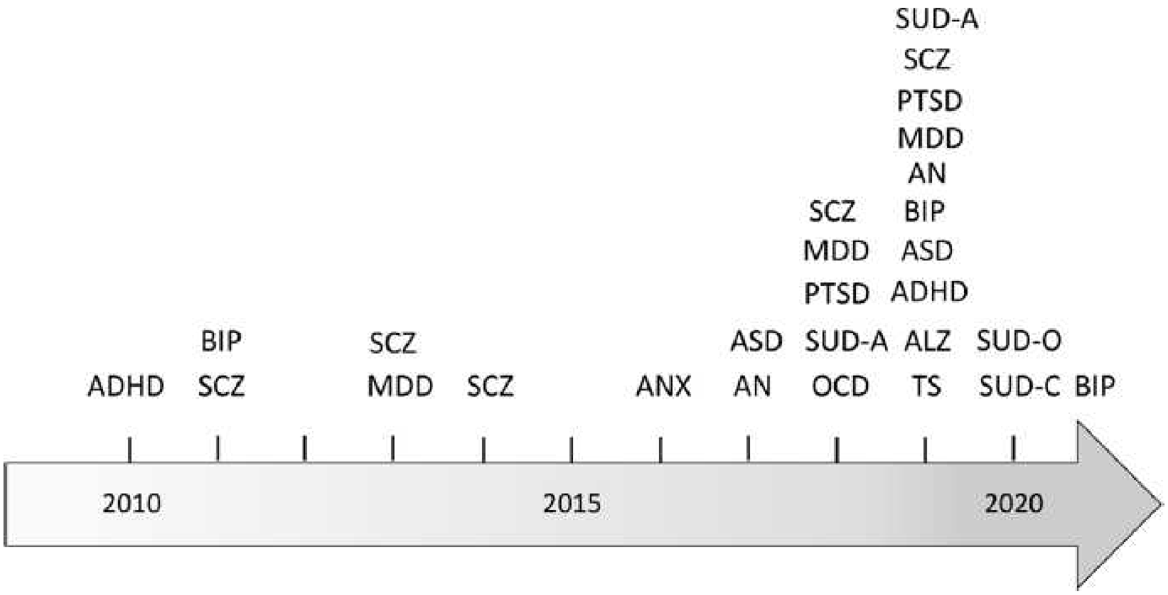

FIGURE 36–1. Timeline of recent key GWAS publications in psychiatric disorders since 2010. ADHD, attention-deficit/hyperactivity disorder; ALZ, Alzheimer disease; AN, anorexia nervosa; ANX, anxiety; ASD, autism spectrum disorder; BIP, bipolar disorder; MDD, major depressive disorder; OCD, obsessive-compulsive disorder; PTSD, posttraumatic stress disorder; SCZ, schizophrenia; SUD-A, alcohol use disorder; SUD-C, cannabis use disorder; SUD-O, opioid use disorder; TS, Tourette syndrome. (Data from Psychiatric Genomics Consortium. 2023. https://pgc.unc.edu/for-researchers/download-results/)

The first overwhelmingly successful GWAS in psychiatry examined SCZ in adults and provided a proof-of-concept that large enough sample sizes are indeed able to identify common risk variants of modest effect in psychiatric disease. Due to the cooperation and sharing between several multisite initiatives, tens of thousands of individuals with SCZ were analyzed in these studies. These initial promising results were further expanded by the Schizophrenia Working Group of the Psychiatric Genetics Consortium (PGC), who conducted a meta-analysis of 49 independent datasets totaling 34,241 cases and 45,604 controls, as well as 1,235 parent–offspring trios. This well-powered meta-analysis produced 128 statistically independent associations between individual SNPs and SCZ diagnosis at 108 loci (including 25 previously reported associations and 83 novel associations). Of interest, at least 75% of these loci contained one or more protein-coding genes, many of which were involved in synaptic plasticity and dopamine (DRD2), glutamate (GR1A1, GRIN2A, GRM3), and calcium (CACNAC1C, CACNA11, CACNB2) signaling pathways.

Since this original success, several other consortia have published similar successes in bipolar disorder (BIP), major depressive disorder (MDD), attention-deficit/hyperactivity disorder (ADHD), and recently ASD, while initial publications for many other disorders such as obsessive-compulsive disorder (OCD), Tourette disorder (TS), anxiety disorders (ANX), and substance use disorders (SUD) show promise with increasing sample sizes (Fig. 36–1). Interestingly, cross-disorder GWAS have revealed loci that span diagnostic borders, demonstrating effects in multiple psychiatric categorizations (e.g., SCZ and BIP, or ASD, ADHD, BIP, MDD, and SCZ; Table 36–2). These cross-phenotype risk genes imply a likely common underlying mechanism for certain aspects, or dimensions, of psychiatric disorders or the presence of resilience genes that confer protection against mental illness.

A substantial leap in pediatric psychiatric genetics occurred in 2019 when the first successful GWAS for ADHD, ASD, and anorexia were published. A GWAS of over 20,000 individuals with ADHD and 35,000 controls identified 12 genome-wide significant loci. Strong genetic overlap of ADHD diagnosis with population-wide cognitive traits supported the evolving conceptualization of categorical mental illness diagnosis lying at the far end of spectrum within the population (Fig. 36–2). A study published the following year showed that genetic risk for ADHD does not vary between adult and pediatric ADHD populations, underscoring the neurodevelopmental nature of the disorder. A similarly sized GWAS in ASD identified five genome-wide significant loci. Interestingly, common variants in ASD identified by GWAS align well with prior whole exome sequencing (WES) studies, highlighting high-confidence risk genes involved in chromatin regulation and neuronal and synaptic development and function. This study also supported the hypothesis that distinct genetic architecture plays a role in different ASD subtypes. In line with the findings of the increased role of large, damaging CNVs in ASD with intellectual disability, the higher functioning ASD subtype was more familial and associated with higher IQ and educational attainment in this large GWAS. GWAS of nearly 17,000 patients with anorexia and 55,000 controls reported 17 genome-wide significant loci with striking face validity, implicating risk genes impacting both psychiatric and metabolic factors. Aside from the above exceptions, the majority of GWAS for pediatric psychiatric disorders have resulted in only suggestive associations, most likely due to modest sample sizes. Other child-focused GWAS, such as those examining more dimensional traits such as “behavioral disinhibition” and “preschool internalizing problems” have again shown several loci approaching genome-wide significance. PGC Working Groups are currently collecting childhood-onset samples to facilitate large GWAS of the magnitude of those successful in adults.

Table 36–2.
Recent Large GWAS, CNV, and WES Studies Across Psychiatric Disorders

	GWAS (Common variants)			CNVs (Copy Number Variants)				WES (Rare Variants)			
Disorder	**Case N**	**Control N**	**N GWAS loci**	**Case N**	**Control N**	**Trios/ quads**	**N CNV loci**	**Case N**	**Control N**	**Trios/ quads N**	**N Genes**
ADHD	20,183	35,191	12	896	2,455	–	1	4,263	5,233	–	0
AN	3,495	10,982	1	–	–	–	–	–	–	–	–
ASD	18,381	27,969	5	–	–	5,574	10	5,556	8,809	6,430	102
BIP	20,352	31,358	17	9,129	63,068	–	1	1,201	5,233	–	0
MDD	135,458	344,901	44	23,979	383,095	–	–	–	–	–	–
OCD	2,688	7,037	0	1,613	1,789	–	1	–	–	–	–
SCZ	40,675	64,643	145	21,094	20,227	–	8	4,264	9,343	1,077	1
TS	4,189	9,488	0	2,434	4,093	–	2	–	–	802	1

GWAS loci are reported from the largest SNP-based studies. Loci enriched for rare copy number variants (CNVs) are reported from studies with >3,000 cases and controls. Individual genes associated with rare coding variants are reported from exome-sequencing studies with >3,000 cases and controls or >500 proband-parent trios.

ADHD, attention-deficit/hyperactivity disorder; AN, anorexia nervosa; ASD, autism spectrum disorder; BIP, bipolar disorder; MDD, major depressive disorder; OCD, obsessive-compulsive disorder; SCZ, schizophrenia; TS, Tourette syndrome.

Adapted from Rees E, Owen MJ. Translating insights from neuropsychiatric genetics and genomics for precision psychiatry. *Genome Med.* 2020;12(1):43. doi:10.1186/s13073-020-00734-5. https://creativecommons.org/licenses/by/4.0/.

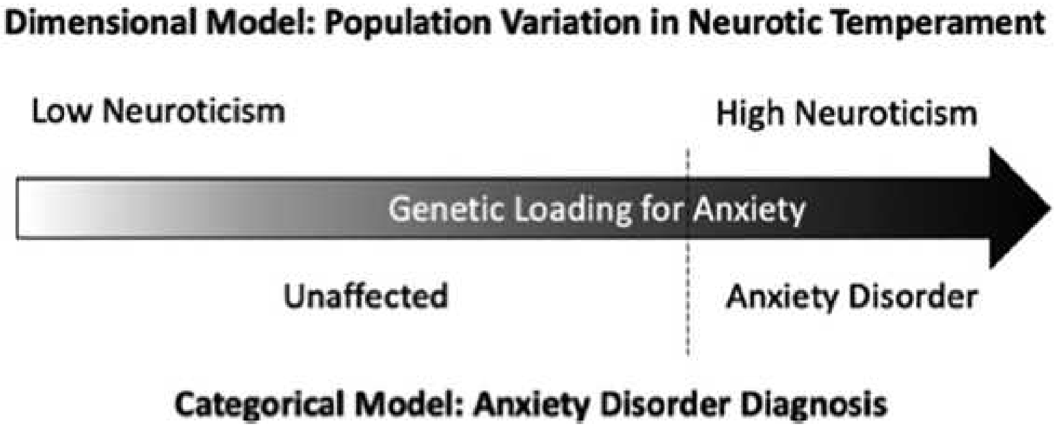
Dimensional Model: Population Variation in Neurotic Temperament
Low Neuroticism
High Neuroticism
Genetic Loading for Anxiety
Unaffected
Anxiety Disorder
Categorical Model: Anxiety Disorder Diagnosis

FIGURE 36–2. Dimensional versus categorical models of mental illness. Genetic loading in the population can be described dimensionally as a neurotic temperament or categorically as an anxiety disorder. Genetic risk for an anxiety disorder exists on a spectrum in the population, from low to high genetic loading. Neurotic temperament is a personality measure that captures anxiety risk dimensionally. Scores on temperament batteries reveal a distribution of individuals ranging from low to high neuroticism. A categorical disease model is overlaid by clinical medicine, where anxiety disorder is defined by impairment in functioning and distress. Individuals in the population are classified as either affected or unaffected. In uniting these two frameworks, increasing genetic risk produces higher neuroticism across the population, and at a defined threshold, anxiety disorder criteria are met.

Polygenic Risk Analysis. In addition to identifying discrete risk variants, GWAS can also inform genetic architecture and heritability. SNP heritability (h^2 SNP), the variance explained by SNPs, can be estimated from GWAS data. Polygenic risk scores (PRS) can be constructed to measure the variance explained collectively by common variation across samples and to test for genetic overlap between traits or shared risk between disorders. A PRS analysis is performed by selecting all of the SNPs meeting a given significance threshold for a trait of interest in a discovery sample and then testing whether this collective set of selected SNPs can also predict a phenotype in an independent replication sample. Using PRS, investigators can assess the genetic relationship between mental disorders and underlying shared risk for psychiatric pathology. However, PRS analyses suffer from Eurocentric ancestry GWAS sample bias at this time. Thus, emphasizing inclusion of samples from populations with non-European genetic ancestry may allow PRS use to be more equitable in the future (see Health Equity in Genetics Research).

Copy Number Variation Analysis. Comparative genomic hybridization (CGH) is a microarray-based technology that can detect CNVs in the human genome rapidly and accurately. *CGH arrays* are glass slides spotted with thousands of probes, from short oligonucleotides to hundred-megabats clone inserts, spanning the genome or focused on a genomic region of interest. Genomic DNA from a test subject and a reference sample are differentially labeled and then cohybridized to the array. Differences in hybridization signals between the test and reference sequences can be detected and decoded to reveal deletions or duplications relative to the reference sample. This method can be used to search for *de novo* (new, noninherited) CNVs in disease populations. CGH has also been used to identify polymorphic CNVs that can be tested for association with disease. Given that local SNPs are frequently in LD with CNVs, it is possible to test association at known CNVs by using nearby SNPs, bypassing the need to measure specific CNV genotypes directly. As an alternative to CGH, GWAS chips can be analyzed for fluorescent signal strength to similarly identify and quantitate CNVs, with increased signal across many adjacent SNPs pinpointing a deletion and increased signal reflecting a duplication.

CNV studies using these microarray-based methods have successfully revealed novel submicroscopic CNVs using advanced computational CNV-predicting algorithms present in populations with intellectual disability, ASD, and SCZ (see Table 36–2). CNVs have been detected in patients with sporadic ASD at 10 times the frequency in controls, and many CNVs are significantly associated with ASD. For this reason, microarray and Fragile X testing is now standard of care for ASD. These methods produce a distinct molecular diagnosis in over 20% of children tested.

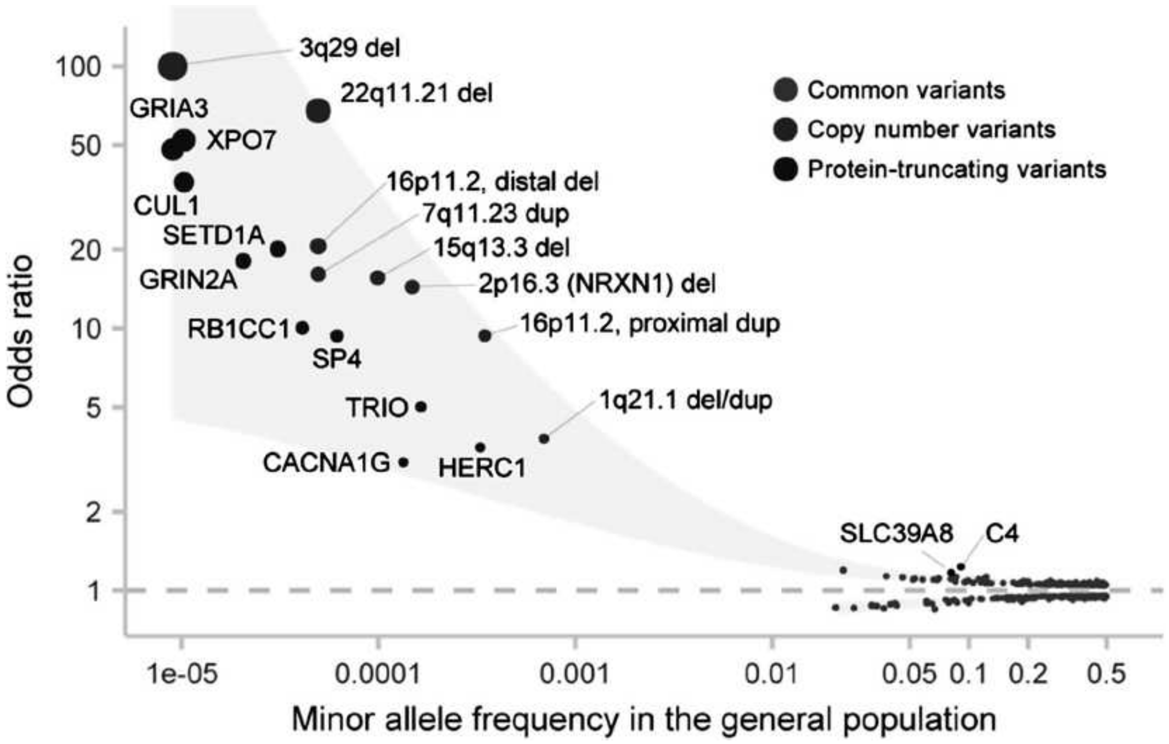

FIGURE 36–3. Inverse relationship between risk variant effect size and frequency. Example plot showing the gradient of differences in odds ratios among common, copy number, and rare variants in SCZ risk. Common variants typically increase the odds for SCZ by one- to twofold, whereas CNVs and rare variants can increase the odds for SCZ by 20- to 100-fold. Examples highlighted in this figure include common variants like complement C4, recently shown to be associated with SCZ with biologic plausibility given its hypothesized role in synaptic pruning; CNVs like 22q11.2 deletion syndrome, long associated with increased risk for SCZ; and rare protein-truncating variants like glutamate and calcium ion channel subunits (GRIA3, GRIN2A, and CACNA1G). (Adapted from Singh T, Neale BM, Daly MJ. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv*. 2020.09.18.20192815. doi: 10.1101/2020.09.18.20192815. https://creativecommons.org/licenses/by/4.0/.)

NV studies have also shed light on biologic processes that are affected in SCZ. Inherited and *de novo* CNVs associated with SCZ more frequently contain genes involved in neurodevelopmental pathways, encoding proteins that function in synaptic architecture, calcium signaling, and glutamatergic neurotransmission. In general, SCZ-associated CNVs have large individual effect sizes but are extremely rare in the population, whereas associated SNPs have small individual effect sizes but are common in the population (Fig. 36–3). The aggregate effect of both CNVs (*de novo* and inherited) and SNPs must be considered in order to fully understand genetic risk for psychiatric disease. In addition, smaller *de novo* CNVs have also been implicated in BIP and ADHD, but at a much lower rate and reduced overall genetic burden compared to ASD and SCZ. Table 36–3 summarizes recurrent CNV loci and their association with psychiatric illness.

Table 36–3.
Summary of Copy Number Variation Associated with Psychiatric Disorders

Locus	Copy Number Change	Disease Association
1q21.1	Deletion/duplication	SCZ, ID, ASD, ADHD
2p16.3	Deletion	SCZ, ID, ASD
3q29	Deletion	SCZ, ID, ASD
7q11.2	Duplication	SCZ, ID, ASD, ADHD, Anxiety
7q36.3	Deletion/duplication	SCZ, ID, ASD
15q11.2	Deletion	SCZ, ID, ASD, ADHD, OCD
15q11–13	Duplication	SCZ, ID, ASD
16p11.2	Deletion/duplication	SCZ, ID, ASD, ADHD, Anxiety, Mood
16p13.11	Duplication	SCZ, ID, ASD, ADHD
17q12	Deletion	SCZ, ID, ASD
22q11.2	Deletion	SCZ, ID, ASD, ADHD, Anxiety, Mood, VCFS
Xq28	Deletion/duplication	ASD

SCZ, schizophrenia; ID, intellectual disability; ASD, autism spectrum disorder; ADHD, attention-deficit/hyperactivity disorder; VCFS, velocardiofacial syndrome.

Genome Sequencing. While GWAS approaches are well suited to examining common genetic variation, genome sequencing studies can identify rare variation leading to disease. In the past decade, the advent of massively parallel sequencing made exome (protein-coding regions of the genome) and whole genome sequencing (WGS) possible. This technology generates millions of overlapping short reads from a DNA template that are assembled and aligned using a reference genome. By comparison to the reference genome, software is used to call genotypes at millions of known and novel variants across the sequenced genome. Since healthy individuals have been shown to carry tens of thousands of novel variants and hundreds of damaging mutations, the signal-to-noise obstacle presented by genome sequencing is staggering. One successful strategy limits the analysis to *de novo* variation by comparing proband to parent genomes. Three independent groups simultaneously published scans of *de novo* variation in ASD in 2012 (see Table 36–2). Collectively, these studies revealed that highly disruptive (nonsense and splice-site) *de novo* mutations in brain-expressed genes are associated with ASD and carry large effects. These *de novo* variants are overwhelmingly paternal in origin (4:1 bias) and positively correlated with paternal age. Multiple independent observations of *de novo* variants within the same gene among unrelated probands and across studies reliably identified several risk genes, including genes involved in sodium transport (*SCN1A, SCN2A*), chromatin remodeling (*CHD8*), and axon guidance (*NTNG1*).

Functional Genomics. Over the past decade, numerous technologic advances have been required to convert the single locus workflow into an integrative functional genomics approach. For example, genome-wide catalogs of functional activity have been generated (e.g., ENCODE project). More recently PsychENCODE has followed suit to generate maps of histone modifications, transcription-factor binding, chromatin assembly, and additional regulatory annotations across brain regions and cell types. For example, just as WES is able to uncover rare variants, other methods like Hi-C and ATAC-seq allow for chromatin capture and interrogation of chromatin state and epigenetic profiling, respectively (Fig. 36–4). These approaches facilitate an understanding of how genetic variation impacts not simply the protein sequence but also the regulation of gene expression. Additionally, overlap between GWAS and functional data allows for more detailed clues about cell type contributions to disease pathogenesis and accelerates the discovery of locus-specific disease mechanisms. There have also been efforts to map quantitative trait loci (QTLs), which are regions of DNA associated with specific traits, across hundreds of individuals and tissues by the Genotype-Tissue Expression (GTEx) consortium. Although QTL approaches can implicate a haplotype in a molecular or cellular phenotype, they cannot identify specific causal variants. Additionally, newer technologies like massively parallel reporter assays (MPRAs) and clustered regularly interspersed short palindromic repeats (CRISPR) genome editing can support functional characterization of specific sequence alterations at scale.

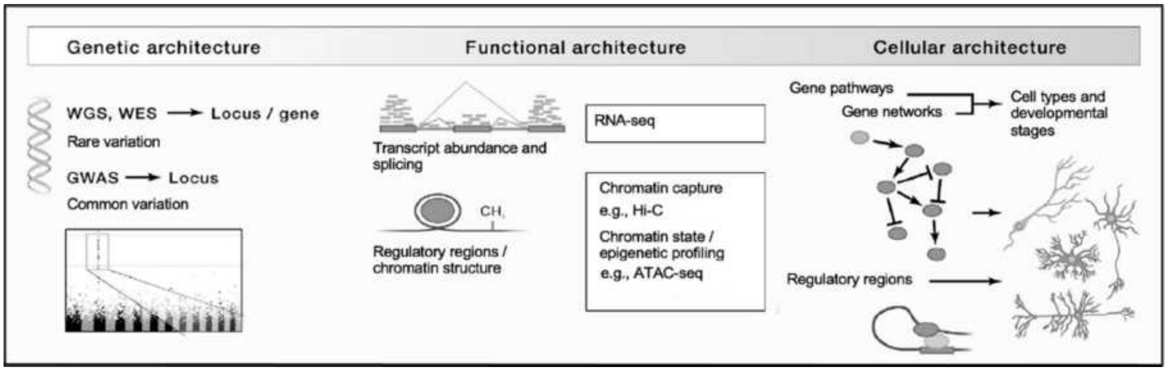
Genetic architecture
Functional architecture
Cellular architecture
WGS, WES → Locus / gene
Rare variation
GWAS → Locus
Common variation
Transcript abundance and splicing
RNA-seq
CH
Regulatory regions / chromatin structure
Chromatin capture e.g., Hi-C
Chromatin state / epigenetic profiling e.g., ATAC-seq
Gene pathways
Gene networks
Cell types and developmental stages
Regulatory regions

FIGURE 36–4. Functional genomics: from genes to neurobiology. From left to right, methods used to elucidate genetic, functional, and cellular architecture underlying psychiatric illness are shown. For genetic architecture, methods include whole genome sequencing (WGS), whole exome sequencing (WES), and GWAS. Examples of functional architecture include RNA-seq and chromatin-directed assays like Hi-C and ATAC-seq. Cellular architecture is probed using gene pathways, gene networks, and regulatory regions in specific tissue types, cell types, and developmental stages. (Adapted from Sullivan PF, Geschwind DH. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell*. 2019;177(1):162–183.)

Pathway Analysis. Individual common SNPs appear to confer only small effects on psychiatric phenotypes, and while they have larger effect sizes, disease-associated CNVs are rare in the population. Therefore, multiple risk alleles in combination are likely required for disease expression (Fig. 36–5). As the first successful studies identifying risk alleles in psychiatric illness have been published, common pathways impacting specific cellular or neurodevelopmental processes have emerged. Many studies now include pathway analyses to add functional relevance and validity to GWAS or sequencing results. Using bioinformatic and gene ontologic approaches (e.g., gprofiler which uses GO Gene Ontology, KEGG, and other databases), investigators can determine whether associated genes belong to particular gene networks or cellular processes above chance levels. Using these approaches, scientists can begin to build gene interaction networks based on a priori knowledge of gene function gleaned from the research literature. These data can inform the development of prediction models, generating testable hypotheses about how a particular risk allele might alter a biologic process to drive disease risk.

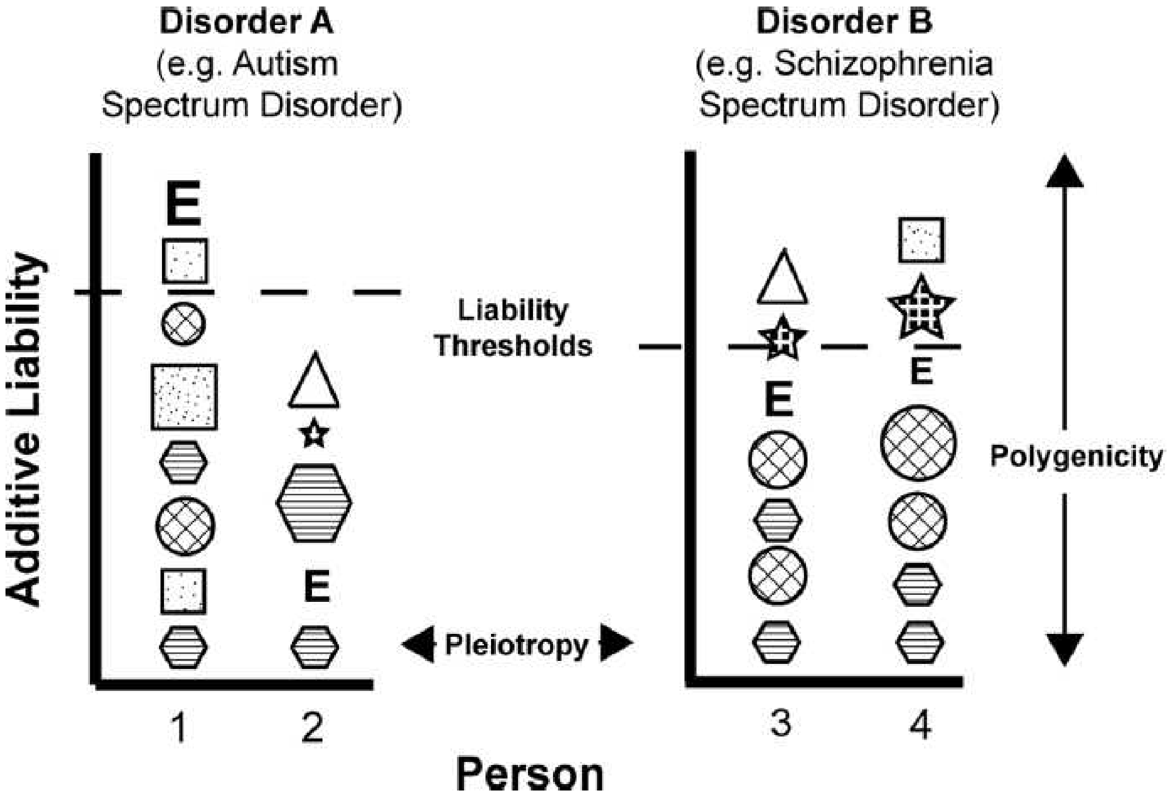
Disorder A
(e.g. Autism Spectrum Disorder)
Disorder B
(e.g. Schizophrenia Spectrum Disorder)
Additive Liability
E
Liability Thresholds
Polygenicity
Pleiotropy
1
2
3
4
Person

FIGURE 36–5. Hypothesized liability-threshold model with a focus on polygenicity and pleiotropy. The left panel represents Disorder A (in this example, ASD), and the right panel represents Disorder B (in this example, schizophrenia spectrum disorder). The left and right panels each have two individuals. Individual #2 does not reach the threshold for Disorder A. Each shape represents a risk gene (common or rare variant) for the disorder. The shapes combine additively, and when they cross a threshold, a person has the disorder. *E* represents environmental impact. Shapes shared across disorders represent pleiotropy. Poly Singh et al. medRxiv 2020 genicity is represented by multiple different shape types within an individual that add up to result in a disorder. Liability thresholds can be unique for different disorders. The size of each shape represents the relative risk contribution of each gene variant. (Adapted from Gandal MJ, Geschwind DH. Polygenicity in psychiatry-like it or not, we have to understand it. *Biol Psychiatry*. 2021;89(1):2–4.)

PHARMACOGENOMICS IN CHILD PSYCHIATRY

Identifying genes that predict treatment response or adverse effects can improve the safety, tolerability, and effectiveness of psychotropic medication. Thus far, most pharmacogenetic studies in adult psychopharmacology have been limited to investigations of single genes or small numbers of genes and frequently test only single variants within these genes for association with drug response. These candidate gene studies have provided inconsistent support for a variety of functional variants largely in neurotransmitter signaling genes. Given the lack of reproducibility and known biases undermining candidate gene studies, these results should be interpreted with caution until confirmed by hypothesis-free genome-wide methods and prospective designs. A notable exception is a large electronic health record survey of drug level and clinical outcome of escitalopram (Lexapro) treatment based on *CYP2C19* genotype. This study strongly supported the clinical validity and utility of incorporating *CYP2C19* genotype into clinical decision making.

While the current gold standard for finding common genetic variation, large-scale GWAS are a particular challenge for detecting pharmacogenomic phenotypes. Well-controlled clinical medication trials are limited in size by feasibility and cost. Population studies are complicated by the inherent heterogeneity of the sample and the use of imprecise self-report or electronic health record data. Combining existing samples also presents considerable obstacles due to differences in outcome measures, study protocols, and diverse populations. No genome-wide significant results emerged from several GWAS conducted for pharmacogenomic predictors of treatment response in adult psychiatry: a meta-analysis of 2,245 individuals of European ancestry with depression (STAR*D, GENDEP, MARS), a depression treatment study of citalopram (Celexa) in 865 adults by the International SSRI Pharmacogenomics Consortium (ISPC), a SCZ treatment trial of 738 subjects receiving multiple antipsychotics (CATIE), and a multiple treatment arm study of 458 patients with bipolar disorder (STEP-BD). GWAS for adverse effects have similarly failed to identify compelling signals in adult samples, including antidepressant treatment-emergent suicidality and metabolic, cardiac, and movement side effects of antipsychotics.

The few pharmacogenomic GWAS in pediatric populations have been more successful thus far. One success was found in a GWAS of antipsychotic-induced weight gain in youth, identifying a risk allele with face validity near the melanocortin 4 (MC4R) gene. When meta-analyzed with three replication cohorts, a genome-wide significant result was reported. GWAS data generated by the OCD Collaborative Genetics

Association Study (OCGAS) was used to examine self-reported positive response or nonresponse to serotonin reuptake inhibitors in 804 subjects with childhood-onset OCD; one SNP near DISP1, a gene involved in neurodevelopment, achieved genome-wide significance. Interestingly, enrichment pathway analysis of suggestive associations highlighted the potential roles of glutamatergic and serotonergic neurotransmitter systems. Several small GWAS of response to methylphenidate and stimulant-induced cardiovascular side effects in children with ADHD, however, failed to produce genome-wide significant results.

CLINICAL CORRELATIONS

The application of genetics in the practice of child psychiatry is in its infancy but is growing rapidly. As described below, genetic testing now has a recommended role in many clinical evaluations and in a growing number of treatment decisions. Given the expected expansion of knowledge in child psychiatric genetics, clinicians need to stay informed, as additional practice implications are sure to follow.

Diagnostic Genetic Testing in Current Psychiatric Practice

Published guidelines from the American Academy of Pediatrics and the American College of Medical Genetics both recommend CGH and GWAS analyses as a routine part of the evaluation of any child with suspected ASD and all children with intellectual disability in an effort to determine molecular diagnoses. In addition, these clinical guidelines include the recommendation for genotyping for the fragile X triplet-repeat mutations and identified mutations in the Rett syndrome *MECP2* gene. The guidelines note that girls with learning disabilities and suspected other developmental delays have been found to possess these mutations in the absence of full syndromic features. Children with macrocephaly should also be tested for mutations in the *PTEN* gene, given its role in growth pathways and association with head circumference. At present, estimates are that routine genetic testing in ASD, as well as intellectual disability, results in positive findings in as many as 20% of individuals.

Some centers have begun routinely obtaining WGS for suspected Mendelian disorders in place of high-density microarray testing for some patients, hoping to identify rare mutations not captured by arrays. Other centers are now collecting WGS data for every patient entering their medical system. With the expansion of clinical testing taking place, more and more patients will present with already available genome sequence data, challenging physicians to identify findings of clinical import. As a result, psychiatrists need to attempt to stay current with the research literature and develop facility with the many searchable genetic databases to aid their interpretations of patient information.

Pharmacogenomic Testing in Current Psychiatric Practice

Pharmacogenomic testing has unquestionably entered the clinical practice of child psychiatry. The U.S. Food and Drug Administration (FDA) has required the inclusion of pharmacogenomic information of clinical importance for over 30 psychotropics commonly used and encountered in mental health treatment. The FDA provides a comprehensive list of these labels on its website. The majority of pharmacogenomic labeling pertains to specific drug safety concerns associated with specific genotypes and their corresponding phenotypes. Arguably the most compelling example is the FDA recommendation for *HLA-B*1502* allele testing prior to prescribing carbamazepine (Tegretol) and oxcarbazepine (Trileptal) in patients of Asian ancestry. The *HLA-B*1502* allele appears to be primarily restricted to Asian populations and is strongly associated with increased risk for severe cutaneous adverse drug effects such as Stevens–Johnson syndrome (SJS). It has been found to be present in greater than 90% of individuals who developed SJS associated with carbamazepine. A majority of other pharmacogenomic considerations relate to genetically mediated drug metabolism differences associated with other safety risks.

Reflecting the growing appreciation of the importance of individual genetic differences in pharmacokinetics and drug distribution, the FDA approved the first pharmacogenetic test, the AmpliChip CYP450, in January 2005. Test results from this chip provide relatively comprehensive testing of well-defined, clinically relevant polymorphisms in two cytochrome P450 genes (*CYP2D6* and *CYP2C19*) that are known to be associated with large differences in the metabolism of many psychotropic medications. Polymorphisms in the genes encoding these enzymes commonly result in poor-metabolizer and ultrarapid-metabolizer phenotypes, producing extreme variability in circulating drug levels. An estimate of the United States’ annual treatment cost of adverse antidepressant reaction in CYP2D6 poor metabolizers was $420 million in 1999; consequently, some authors advocate the economic viability of testing CYP genotypes prior to antidepressant therapy. Current guidelines do not endorse standard genetic testing for antidepressants; however, testing is recommended for prescribing tricyclic antidepressants, as well as specific antidepressants from other classes. These guidelines, available at the Clinical Pharmacogenetics Implementation Consortium (CPIC) website, advise avoidance of tricyclic antidepressants in individuals with ultrametabolizer CYP2D6 status because of potential lack of efficacy and either avoidance or 50% dose reductions and monitoring of therapeutic drug levels in individuals with the “poor metabolizer” phenotype of CYP2D6 or CYP2C19. Similar guidelines are available for selective serotonin reuptake inhibitors (SSRIs). CPIC also published a guideline to assist with atomoxetine dosing in patients with ADHD. Importantly, CPIC review of the effects of CYP2D6 metabolizer status on treatment outcome revealed that a third of normal metabolizers will not achieve therapeutic blood levels of atomoxetine (Strattera) within the FDA-specified dosing range, and therapeutic blood monitoring may be beneficial in optimizing atomoxetine dose. Genotype-tailored treatment is routinely performed in many other areas of medicine (tamoxifen, warfarin, antineoplastics) and will continue to expand in relation to the safe and effective use of psychotropic therapies.

Numerous commercial entities have begun to offer a wider range of genetic “decision support tools (DSTs),” ostensibly to inform psychiatric practice. While several commercial testing providers have published data supporting their products, these studies do not evaluate individual tests but demonstrate that the whole package improves patient care when compared with treatment as usual. Unfortunately, proprietary algorithms are employed that cannot be validated externally. A few positive meta-analyses of the use of DSTs in adult psychiatry have been published but are based on poor-quality primary studies with substantial heterogeneity and mixed results. Problems with study design, posthoc outcome reporting, bias, test validity, and generalizability temper enthusiasm for positive findings. Furthermore, these studies are funded and analyzed, if not entirely conducted, by industry with major financial conflicts of interest. Prior to 2021, no data existed in pediatric samples, and the use of DSTs to guide pharmacotherapy for a variety of child mental health diagnoses was supported only by data from middle-aged women with depression who had failed at least one prior medication trial. The first randomized controlled trial of pharmacogenomic guided treatment compared to treatment as usual in adolescents with depression failed to show superiority. The one significant finding revealed that physicians receiving genetic data were less likely to prescribe an evidence-based treatment for depression in youth, demonstrating a common phenomenon whereby physicians inappropriately follow weak genetic guidance over the empirically driven evidence base. Clinicians should be aware that the supporting evidence for most pharmacogenetic testing is extremely limited and lacks the wider replication required before such tests can be considered “evidence-based.” Current responsible guidelines suggest consideration of *CYP2D6* and *CYP2C19* testing in adults with depression who have failed at least one prior adequate medication trial and HLA genotyping to prevent serious adverse events with carbamazepine and oxcarbazepine use. Results may inform which drugs to avoid or dose more cautiously in a given patient but currently cannot predict whether any medication will be effective. The FDA has issued several warnings against the use of DSTs in psychiatry. Similarly, the American Academy of Child and Adolescent Psychiatry cautions against the use of these products in child psychiatry in an online position statement.

LIMITATIONS OF CURRENT RESEARCH

Even with complete knowledge of the genetic architecture of a psychiatric disorder, deeper and actionable understanding requires a detailed understanding of the functional genomic and cellular architectures. Functional genomics approaches examine how implicated loci impact regulatory processes that influence gene expression and the functional coordination of these genes that influence biologic processes. Cellular architecture approaches target when and where these functional processes are operative within brain regions, layers, cell types (neuronal, glial, progenitor), and developmental stages (pre- and postnatal).

Understanding brain region, layer, and cell type specificity in a psychiatric condition may reveal the selective vulnerability of neuronal subtypes to brain injury and explain how despite exposure to a common insult, disease pathogenesis may reflect disrupted function in a specific subset of brain cells. For example, mutant huntingtin protein (encoded by the *HTT* gene) is widespread in Huntington disease, but striatal degeneration is disproportionately affected. Anoxia disproportionately leads to cell death of hippocampal CA1 pyramidal neurons compared with pyramidal neurons in other hippocampal subregions.

One barrier to identifying genetic susceptibility to psychiatric conditions includes the phenotypic complexity of psychiatric illness. Another is the lack of direct brain tissue accessibility over time within an individual across the course of psychiatric symptom development and progression. Additional key issues include polygenicity (many genes contribute small effects to a psychiatric disorder) of each psychiatric condition and pleiotropy (risk genes are shared across psychiatric disorders) that make causal links between susceptibility genes and the development of a psychiatric condition incredibly challenging (see Fig. 36–3). These numerous promising associations do not necessarily imply causation. Grounding genetic associations with functional data in autopsy-derived brain tissue, stem cells obtained from blood and skin, and live model systems will be critical for understanding their impact and contributions to causal mechanisms responsible for illness pathogenesis.

Ongoing challenges relate to disseminating lessons from high-quality science to the clinic while discouraging the implementation of untested strategies. The current state of the use of genetic data in clinical practice demonstrates this pitfall: (1) genetic testing is recommended in the evaluation of ASD and intellectual disability, but its use is woefully underutilized, and (2) commercial pharmacogenomic testing is widely available and often a primary factor in treatment when its utility is not yet established.

FUTURE DIRECTIONS

Advancing Technology, Computational, and Basic Neuroscience Methods

Emerging technologies and novel data collection and mining methods offer promising avenues forward. Digital technologies provide opportunities for real-time data collection, facilitating the incorporation of objective data collected by wearable devices to study both the momentary and longitudinal course of symptoms and behavioral domains. Electronic health record mining can deliver large databases of clinical information for scientific study. These avenues, in addition to continued subject recruiting across diverse populations, will supply ever-expanding datasets for analysis. Patterns and processes too complex for human appreciation can be detected by computational data-mining approaches like machine learning, leveraging advances in artificial intelligence. Once hypotheses are generated by analysis of big data, functional studies are needed to validate conclusions. Recent molecular tools to examine functional effects include MPRAs that can screen thousands of noncoding genomic regions simultaneously to test for functional effects, CRISPR (clustered regularly interspaced short palindromic repeats) assays that can manipulate genes in a fully developed organism, optogenetic technologies that can activate or suppress neural pathways in vivo, and single-cell multiomic integrative approaches that reveal the complex processes occurring in a single neuron.

The field of epigenetics attempts to understand the impact of experience (e.g., environmental influences) on the regulation and expression of genes. Early findings have already shed light on the mechanisms by which early childhood adversity can lead to risk for adult mental disorders (post-traumatic stress disorder, mood disorders) via methylation of genes involved in the stress–response system. Causal relationships are often a challenge for epigenetic conclusions; however, new approaches, such as Mendelian randomization, have begun to offer solutions. As epigenetic mechanisms are revealed, they are likely to point to novel prevention and treatment strategies that could be applied to youth and potentially alter the trajectory of their development and adaptation. Such exciting prospects should encourage close attention to this important area of psychiatric research.

Pediatric Populations

Genetic studies in pediatric samples are dwarfed by those conducted in adults. Furthermore, age-of-onset is often not obtained in adult samples and represents a missed opportunity. Interestingly, most childhood-onset neuropsychiatric disorders appear to be more heritable than similar adult-onset phenotypes, suggesting that research to identify genetic influences might be more fruitful in pediatric samples. Likewise, studies of genetic influences on treatment and prognosis may be more revealing in younger, treatment-naive subjects. As more adult-onset psychiatric disorders are increasingly being conceptualized as neurodevelopmental in nature, with identifiable prodromal and subclinical childhood manifestations, the study of such early phenotypes may represent genetic subgroups more tractable for etiologic investigations.

Precision Psychiatry

Precision medicine seeks to employ knowledge of individual differences to inform diagnosis, prognosis, and treatment. Over the coming decades, precision psychiatry will likely incorporate not only genetic but also molecular, biochemical, electrophysiologic, imaging, and behavioral biomarkers. Machine learning and other computer-driven approaches can be used to analyze biomarker data and develop data-driven, predictive algorithms that improve with experience. Digital phenotyping with smartphones and wearable sensors might also contribute to precision psychiatry.

As the adoption of clinical genetic testing expands, spurred by future technical and scientific progress, more and more genetic information will become available to clinicians. Many publicly available databases can be accessed by clinicians in their efforts to understand the possible role of genetics at the level of the individual patient, including the guidelines published by the FDA and CPIC, as well as information about specific genetic variants curated by the Pharmacogenomics Knowledge Base (PharmGKB).

Health Equity in Genetics Research

For most diseases, including psychiatric disorders, large-scale GWAS-aggregation efforts have been biased toward information from individuals of European ancestry. While many common risk alleles discovered in one population group are detectable in others, allele frequencies can vary substantially. Therefore, refocusing GWAS and sequencing studies on diverse populations will most likely generate a rich landscape of novel risk alleles.

The use of PRS to improve psychiatric disease outcomes via precision medicine remains one possible and practical area for the clinical translation of genetic risk. However, clinical implementation of PRS raises ethical and scientific challenges since PRS are many-fold more accurate in individuals of European ancestry than others due to Eurocentric GWAS biases. Thus, clinical uses of PRS today would benefit those of European descent at the expense of non-European populations. Critically, new research must prioritize greater diversity in genetic studies with early public dissemination of summary statistic findings to promote health equity and decrease disparities for those who are already most underserved.

Expanded Genetics Training for Mental Health Providers

As a field, psychiatrists should be able to critically appraise the evidence supporting genetic testing and pharmacogenetic/pharmacogenomic testing. Current clinical applications include estimating risk for psychiatric illness from family structure and epidemiologic studies, ordering and interpreting genetic tests for ASD and intellectual disability, and knowing when to consult genetic counselors and medical geneticists. Some anticipated novel clinical applications in the pipeline include estimating psychiatric risk with specific CNV and SNV information, ordering exome sequencing genetic tests for rare variants in SCZ and BIP, and applying genetic risk scores (e.g., PRS) to guide personalized psychiatric medicine. Covering basic principles of genetics, common and rare genetic variants, pharmacogenetics, epigenetics/gene expression, ethical and social issues in genetics, and principles of risk communication during psychiatry residency training may help bridge the education gap and increase confidence and comfort for future psychiatrists to address these issues clinically.

ACKNOWLEDGMENTS

This chapter represents an update and revision of previous chapters in this series, most recently by Meredith Hannan, M.D., Ph.D. from the last edition. We are indebted to Dr. Hannan for her work and retain much of her prose in this current edition.

Further Readings

Besterman AD, Moreno-De-Luca D, Nurnberger JI Jr. 21st-century genetics in psychiatric residency training: how do we get there? *JAMA Psychiatry*. 2019;76(3):231–232.

Bousman CA, Bengesser SA, Aitchison KJ, et al. Review and consensus on pharmacogenomic testing in psychiatry. *Pharmacopsychiatry*. 2021;54(1):5–17.

Brown JT, Bishop JR, Sangkuhl K, et al. Clinical pharmacogenetics implementation consortium guideline for cytochrome P450 (CYP)2D6 genotype and atomoxetine therapy. *Clin Pharmacol Ther*. 2019;106(1):94–102.

Claussnitzer M, Cho JH, Collins R, et al. A brief history of human disease genetics. *Nature*. 2020;577(7789):179–189.

Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell*. 2019;179(7):1469–1482.e1411.

Demontis D, Walters RK, Martin J, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet*. 2019;51(1):63–75.

Gandal MJ, Geschwind DH. Polygenicity in psychiatry-like it or not, we have to understand it. *Biol Psychiatry*. 2021;89(1):2–4.

Goldstein-Piekarski AN, Holt-Gosselin B, O'Hora K, Williams LM. Integrating sleep, neuroimaging, and computational approaches for precision psychiatry. *Neuropsychopharmacology*. 2020;45(1):192–204.

Grove J, Ripke S, Als TD, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431–444.

Hicks JK, Bishop JR, Sangkuhl K, et al. Clinical pharmacogenetics implementation consortium (CPIC) guideline for CYP2D6 and CYP2C19 genotypes and dosing of selective serotonin reuptake inhibitors. *Clin Pharmacol Ther*. 2015;98(2):127–134.

Hicks JK, Sangkuhl K, Swen JJ, et al. Clinical pharmacogenetics implementation consortium guideline (CPIC) for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants: 2016 update. *Clin Pharmacol Ther*. 2017;102(1):37–44.

Jukic MM, Haslemo T, Molden E, Ingelman-Sundberg M. Impact of CYP2C19 genotype on escitalopram exposure and therapeutic failure: a retrospective study based on 2,087 patients. *Am J Psychiatry*. 2018;175(5):463–470.

Lappalainen T, Scott AJ, Brandt M, Hall IM. Genomic analysis in the age of human genome sequencing. *Cell*. 2019;177(1):70–84.

Malhotra AK, Correll CU, Chowdhury NI, et al. Association between common variants near the melanocortin 4 receptor gene and severe antipsychotic drug-induced weight gain. *Arch Gen Psychiatry*. 2012;69(9):904–912.

Manoli DS, State MW. Autism spectrum disorder genetics and the search for pathological mechanisms. *Am J Psychiatry*. 2021;178(1):30–38.

Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584–591.

Nurnberger JI Jr, Austin J, Berrettini WH, et al. What should a psychiatrist know about genetics? Review and recommendations from the residency education committee of the international society of psychiatric genetics. *J Clin Psychiatry*. 2018;80(1):17nr12046.

Phillips EJ, Sukasem C, Whirl-Carrillo M, et al. Clinical pharmacogenetics implementation consortium guideline for HLA genotype and use of carbamazepine and oxcarbazepine: 2017 update. *Clin Pharmacol Ther*. 2018;103(4):574–581.

Qin H, Samuels JF, Wang Y, et al. Whole-genome association analysis of treatment response in obsessive-compulsive disorder. *Mol Psychiatry*. 2016;21(2):270–276.

Rees E, Owen MJ. Translating insights from neuropsychiatric genetics and genomics for precision psychiatry. *Genome Med*. 2020;12(1):43.

Rikani AA, Choudhry Z, Choudhry AM, et al. The mechanism of degeneration of striatal neuronal subtypes in Huntington disease. *Ann Neurosci*. 2014;21(3):112–114.

Rovira P, Demontis D, Sanchez-Mora C, et al. Shared genetic background between children and adults with attention deficit/hyperactivity disorder. *Neuropsychopharmacology*. 2020;45(10):1617–1626.

Savatt JM, Myers SM. Genetic testing in neurodevelopmental disorders. *Front Pediatr.* 2021;9:526779.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421–427.

Schmidt-Kastner R, Freund TF. Selective vulnerability of the hippocampus in brain ischemia. *Neuroscience*. 1991;40(3):599–636.

Singh T, Poterba T, Curtis D, et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*. 2022.

Srivastava S, Love-Nichols JA, Dies KA, et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med Nov* 2019;21(11):2413–2421.

Sullivan PF, Geschwind DH. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell*. 2019;177(1):162–183.

Voort JLV, Orth SS, Shekunov J, et al. A randomized controlled trial of combinatorial pharmacogenetics testing in adolescent depression. *J Am Acad Child Adolesc Psychiatry*. 2021;61:46–55.

Watson HJ, Yilmaz Z, Thornton LM, et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat Genet*. 2019;51(8):1207–1214.

37

Neuroimaging in Psychiatric Disorders of Childhood

FRANK P. MACMASTER, PH.D., KAYLA D. STONE, PH.D., SNEHA CHENJI, PH.D., AND DAVID R. ROSENBERG, M.D.

Over the last three decades, there has been a revolution in our understanding of the living human brain. The advent of brain imaging has allowed for noninvasive assessment of brain structure, function, and chemistry. Until the development of brain imaging, the investigation of the neurobiologic underpinnings of psychiatric illness in children was difficult. Indeed, for many years, the commonly held belief was that psychiatric conditions did not occur in pediatric populations. Brain imaging research has allowed us to move away from this era of skepticism by demonstrating that psychiatric illness does indeed affect children, and these illnesses often have clear neurobiologic underpinnings. The application of brain imaging to describe the biology of the illness can lead to improved diagnosis and innovative treatments. Moreover, the study of brain maturation may make it possible to find critical time periods during which brain abnormalities could lead to neuropsychiatric symptoms. The research findings may also indicate potential "targets" for treatment intervention.

This chapter has three sections. First, core brain imaging methods are described. Second, special considerations (such as sex, gender, and development) are addressed. Third, relevant findings from brain imaging studies of psychiatric disorders of childhood are summarized. Disorders include attention-deficit/hyperactivity disorder (ADHD), mood disorders such as major depressive disorder (MDD) and bipolar I disorder (BP-I), anxiety disorders, obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), Tourette syndrome (TS), childhood-onset schizophrenia (COS), substance use disorders, feeding and eating disorders, autism spectrum disorder (ASD), and specific learning disorders including dyslexia. The purpose of this chapter is to provide an understanding of brain imaging in psychiatric disorders of childhood.

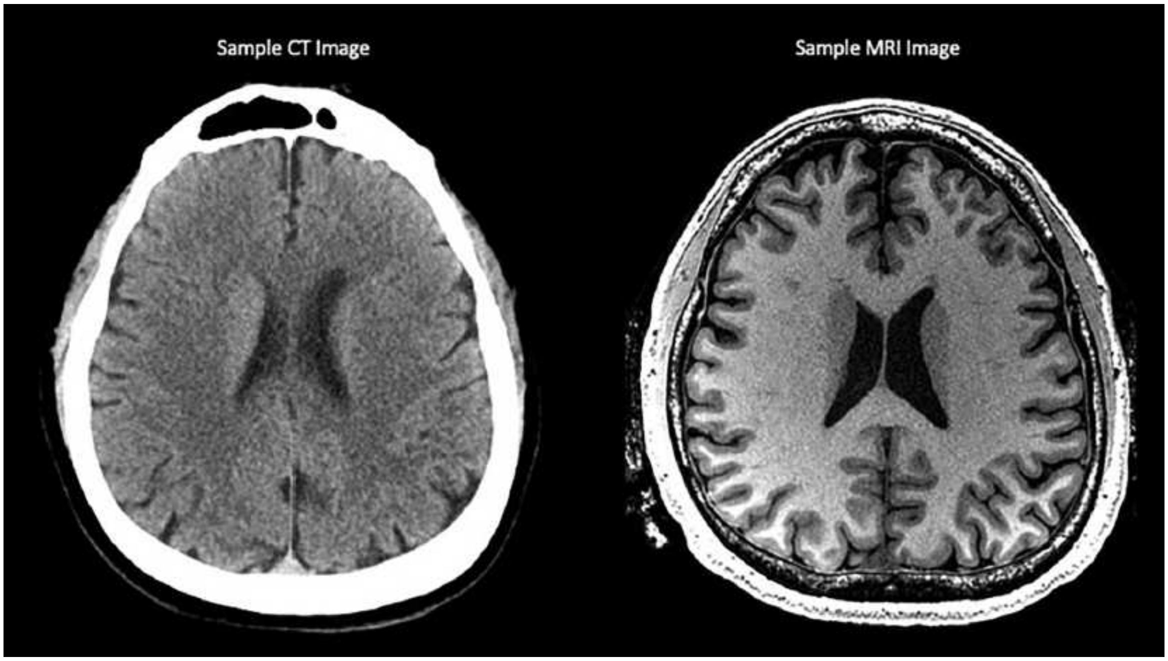

FIGURE 37–1. Sample CT and MRI image of the same person demonstrating differences in contrast and resolution. (Courtesy Frank P. MacMaster, Ph.D.)

TECHNIQUES AND PRACTICAL CONSIDERATIONS

Neuroimaging Techniques

Thirty years ago, the study of the brain was largely limited to postmortem work. However, postmortem studies are not practical for developing an understanding of psychiatric disorders of childhood. Nearly all postmortem brains come from adults. Hence, they are confounded by age, chronic illness, and exposure to treatment. The application of brain imaging techniques has overcome this obstruction to facilitate progress in understanding the developing brain.

Morphology/Structure

Morphologic studies include the examination of the structure of grey and white matter. Computed tomography (CT) was the first imaging method to make routine the direct, in vivo, and noninvasive evaluation of the living human brain. In practice, magnetic resonance imaging (MRI) has replaced quantitative CT in brain research due to its improved resolution and tissue contrast (Fig. 37–1). Early studies in psychiatry began with more of a "radiologic" approach—that is, looking for the presence or absence of visible damage or lesions. An example of this is the link between MRI T2 hyperintensities in left frontal stroke and depression. These lesions can be observed with the naked eye. This approach was not adequate, as most psychiatric illnesses do not involve a visible lesion. Hence, the field focused on more quantitative methods.

A measure should be both valid and reliable. Measurement of biologic structures is challenging, as many parameters change with development and even over shorter periods of time. In the human brain, not all aspects are MRI-visible. This is critical when designing an experiment using MRI as the method. The effect must be of necessary scale to be detectable. However, the technology is continually improving. For instance, MRI techniques have been developed to examine cortical layers (2 to 4 mm thick) using MRI scanners operating at 7T. It may someday be possible to evaluate single cells in vivo using MRI-based technologies. In addition to the ability to be detected and measured, the structure must also be of clinical interest. In other words, it should be critically involved in a specific function associated with the psychiatric disorder in question.

The two principal steps in morphologic MRI studies are data acquisition and quantification. With both of these steps, bias, and error can diminish the validity of the study. Bias can obscure real effects or produce incorrect findings. Error is random variation. Both bias and error can reduce statistical power and result in erroneous findings. There is no single accepted method for reducing bias and error in each of the steps involved in MRI experiments. Furthermore, there is little consensus regarding how to properly perform anatomic quantification, with methods varying considerably. Due to inherent variation in regional brain volumes, it is accepted that large sample sizes are required to demonstrate even moderate effects between groups.

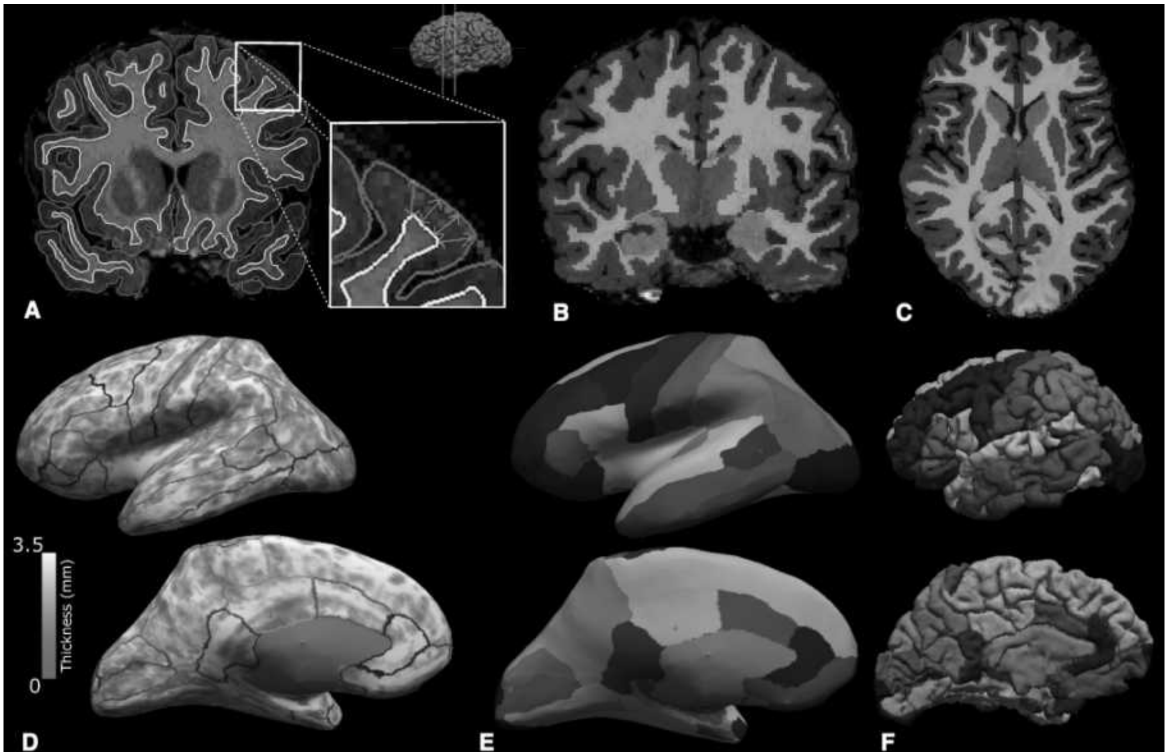

FIGURE 37–2. Cortical thickness and subcortical volume from a pediatric participant. Panels represent outputs from FreeSurfer. **A**: Outlines depict white matter edges (*yellow*) and pial edges (*red*) which enable calculation of cortical thickness. **B, C**: Coronal and axial views of subcortical volumes, respectively. **D**: Heat map representing cortical thickness across various brain regions. **E, F**: Sample parcellations using Desikan–Killany atlas showing inflated brain and gyri, respectively. (See eBook for color image.) (Courtesy Sneha Chenji, Ph.D.)

Grey Matter: Volume, Shape, Thickness, Area, and Gyrification. A “structure” should be distinct from its neighbors. Within the brain structure, there should be a common aspect to or resultant from the parts comprising the structure. This could be cell type, neurotransmitter receptor distribution, or function. For instance, caudate could be a descriptor because it has specific related functions and is contiguous spatially. When trying to understand the brain, you need to know where functionally specific structures are located, what function they perform, and how the distributed processing of that function is organized. The volume of a brain structure is reflective of the size, shape, and density of its parts. It is reasonable to assume that functional changes are reflected physiologically with changes in cellular composition and/or activity. Consequently, by examining the volumetric properties of a structure, one can make inferences about its functionality. Researchers may also study the shape of a brain structure and compare groups and conditions. The quantitative analysis of biomedical shape variation and shape change, while more complicated mathematically, offers useful information on two- and three-dimensional properties of structures that may be relevant to disease states. More recently, the investigation of cortical thickness, surface area, and gyrification (i.e., the pattern and degree of cortical folding) has become increasingly common in the study of pediatric psychiatric populations (Fig. 37–2). These approaches are typically automated and rely on established templates identifying structures and localization of differences.

White Matter Measures: Diffusion Tensor Imaging. Diffusion tensor imaging (DTI) is based on the MR signal resulting from the restricted movement of water in brain tissue. The fundamental architecture of white matter has the axons running in parallel bundles or tracts. This forces water molecules to move along the main bearing of the tract. This oriented dispersal pattern is termed anisotropic diffusion. The location, direction, and anisotropy of the white matter tracts can be measured, yielding metrics that enable investigation of in vivo white matter changes. Fractional anisotropy is a common scalar metric (ranging from 0 to 1) used to measure white matter integrity. It indicates the directional preference of water diffusion. Additional metrics such as axial diffusivity (diffusion rate parallel to the axon bundle) and radial diffusivity (diffusion rate perpendicular to the axon bundle) provide insight into white matter development including myelination and axonal packing (Fig. 37–3). These metrics can be used to measure individual tracts (e.g., superior longitudinal fasciculus) or can be used to measure the structural connectivity pattern of a brain circuit (e.g., using graph theory). DTI measures vary in different brain regions. For example, white matter in the frontal lobes may show lower anisotropy. This is due to the presence of crossing white matter fibers. Thus, careful consideration is required when choosing DTI parameters. Advances in diffusion imaging parameters enable delineation of crossing white matter fibers and provide further insights into intricate tracts such as in the frontal lobe. Myelin water imaging—which quantifies the fraction and relaxation parameters of water trapped in the lipid bilayers—has been applied in neurodegenerative conditions but could offer additional insight into childhood-onset neuropsychiatric disorders.

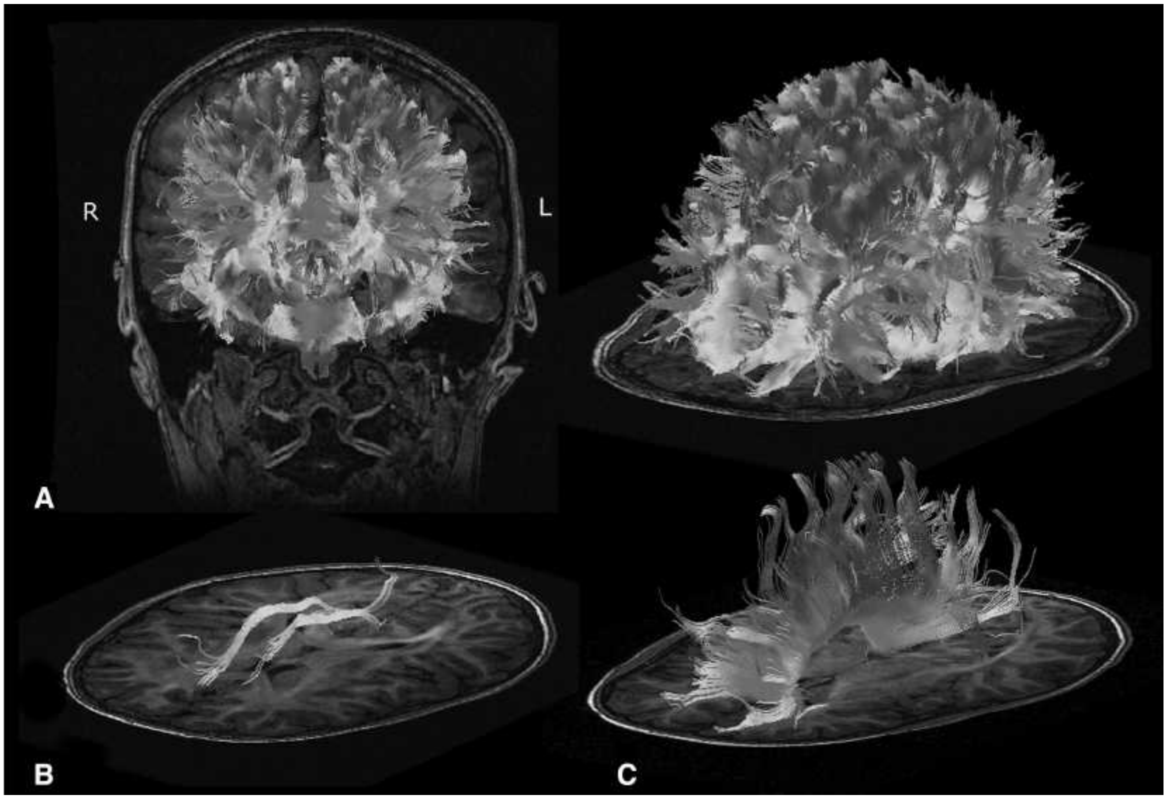

FIGURE 37–3. A: Whole brain tractography using diffusion tensor imaging in a pediatric participant. Images show white matter tracts obtained using ExploreDTI. **B:** Sample cingulate tract. **C**: Corpus callosum obtained using seed-based tractography. (Courtesy Sneha Chenji, Ph.D.)

Chemistry

Brain imaging methods that evaluate brain chemistry may be more sensitive than structural imaging studies for identifying localized abnormalities in psychiatric illness. As indicated above, the radiologic hallmarks of psychiatric illness are often subtle, not typically visible to the naked eye, and require very sensitive methods of detection. Positron emission tomography (PET) and single-photon emission computed tomography (SPECT) methods have been critically informative regarding the pathogenesis and mechanisms of treatment response of several neuropsychiatric conditions in adults. However, these techniques use ligands labeled with radiation, thereby increasing the risk of their use in children and adolescents. In contrast, MR-based techniques are free of ionizing radiation and, as such, have rapidly become the preferred means for examining neuroanatomy, blood flow, and chemistry in children. Novel methods, like 3D PET, do use less radiation and may become an alternative in the future.

Magnetic Resonance Spectroscopy. Magnetic resonance spectroscopy (MRS) measures chemicals in the brain. Although many nuclei with odd-numbered protons or neutrons, including phosphorus (^{31}P), carbon (^{13}C), fluorine (^{19}F), sodium (^{23}Na), and proton (^{1}H) compounds, can be detected using MRS, ^{1}H-MRS is the most commonly used technique, providing measures of the following: N-acetyl aspartate (NAA), a neuronal marker; creatine/phosphocreatine (Cr), involved in energy metabolism; choline compounds (Cho), membrane components; myo-inositol, a secondary messenger involved in signal transduction; and glutamate + glutamine (Glx), and γ-aminobutyric acid (GABA)—all of which occur in relatively high concentrations in the brain (Fig. 37–4). ^{31}P-MRS is the second most common approach in developmental populations. ^{31}P-MRS allows for the study of the products of membrane synthesis (phosphomonoesters [PMEs] like phosphoethanolamine and phosphocholine) and breakdown (phosphodiesters [PDEs] like glycerophosphoethanolamine and glycerophosphocholine), along with substrates of energy metabolism. The choice of MRS versus PET/SPECT measures of brain chemistry involves important trade-offs. First, MRS does not have the power to detect chemicals occurring at lower concentrations, such as monoamines, neuropeptides, opiates, and so on. Second, MRS has relatively poor spatial and temporal resolution as compared to that of PET. However, with data acquisition and spectral editing programs becoming increasingly sophisticated and the increases in magnetic field strength, striking improvements in the spatial and temporal resolution of MRS are being fulfilled.

Function

Functional Magnetic Resonance Imaging. The selection of functional magnetic resonance imaging (fMRI) versus PET/SPECT measures of cerebral blood flow also involves trade-offs. fMRI delivers excellent spatial resolution (1 mm or less) along with potentially superior temporal resolution (about 1 second) compared to the PET/SPECT approaches. Second-to-second temporal resolution is now possible with event-related fMRI. Unlike PET/SPECT, fMRI does not provide an absolute measure of cerebral blood flow. Rather, fMRI detects changes in blood oxygenation-level–dependent (BOLD) signal using a T2*-weighted sequence. It relies on the paramagnetic properties of deoxyhemoglobin to induce local changes in image intensity. Using carefully designed neurobehavioral paradigms during fMRI scans, BOLD signals can be used to generate a pattern of activation (Fig. 37–5). Early techniques involved contrasting task conditions with rest conditions to determine areas of activation. New analytical techniques allow for the quantification of “baseline” or “resting-state” blood flow using fMRI images. Initial techniques involved exploring changes in BOLD signal in a specific a priori region of interest (seed) relative to the whole brain (seed-based correlation analysis). More recent techniques use advanced approaches such as Independent Component Analysis (ICA) and graph theory metrics. These provide a model-free approach (no a priori seed) to explore whole brain networks at rest. The study of so-called resting-state networks offers an innovative method for determining differences in brain function between groups and conditions.

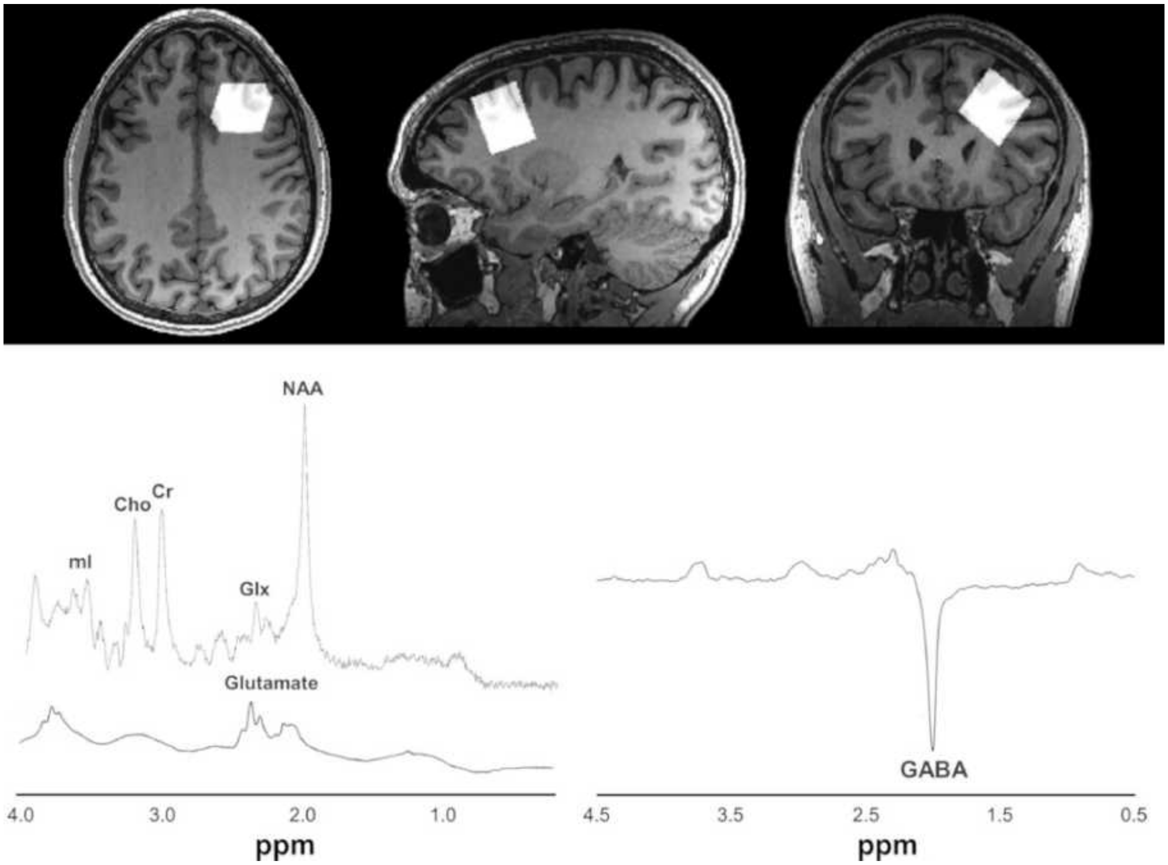

FIGURE 37–4. Sample proton magnetic resonance spectrum from the left dorsolateral prefrontal cortex. Spectra were obtained from a 15.6-mL volume of interest as shown on the magnetic resonance images. At higher field strengths, distinguishing glutamate from Glx is possible. Angulation of voxels to best accommodate anatomy is also possible. Cho, choline; CR, creatine/phosphocreatine; Glx, glutamate/glutamine; mI, myo-inositol; NAA, *N*-acetyl-aspartate; GABA, γ-aminobutyric acid. (Courtesy Frank P. MacMaster, Ph.D.)

MRI-based approaches are a viable tool in basic science studies. As a scientific method, it has served many purposes—mapping developmental trajectories, disease processes, and the effect of treatment. Thanks to its use of nonionizing radiation, MRI allows for repeated measures over time with a high degree of safety. MRI further avoids the physiologic changes inherent in the death process and the effects of the lag between death and fixation processing that are part of postmortem studies. Furthermore, for psychiatric diseases of childhood, like OCD and ADHD, there are virtually no postmortem studies. Hence, the investigation of psychiatric diseases in childhood has benefited from MRI approaches. It should be noted that MRI is obviously of lesser utility in children with braces or claustrophobia, in which case, alternative techniques mentioned above could be considered to measure the developing brain.

Other Functional Techniques. Electroencephalography (EEG) is one of the initial techniques widely used to study brain activity and is often used in studying pediatric neuropsychiatric groups. A standard EEG is a visual representation of the difference in voltage between two different cerebral locations using electrodes placed on the scalp. Essentially, EEG is a measure of the electrical activity generated by neurons. While it lacks the spatial resolution of an fMRI scan, it has excellent temporal resolution. Quantitative EEG can be used as a post hoc analysis approach to measure subtle differences in frequency changes and coherence of activity in brain regions. Recent advances allow for simultaneous recording of EEG and fMRI, thereby allowing for multimodal integration of brain function. Additional techniques to study brain function include the use of functional near-infrared spectroscopy (fNIRS) and magnetoencephalography (MEG). The use of fNIRS is growing in pediatric groups due to advantages such as higher tolerance to motion (compared to fMRI). Additionally, the device is portable, enabling its use both inside and outside of laboratory settings. The latter technique, MEG, relies on the magnetic field generated by electric activity in the brain. It is advantageous over other functional techniques, such as EEG, in identifying the focal point origin of epilepsy.

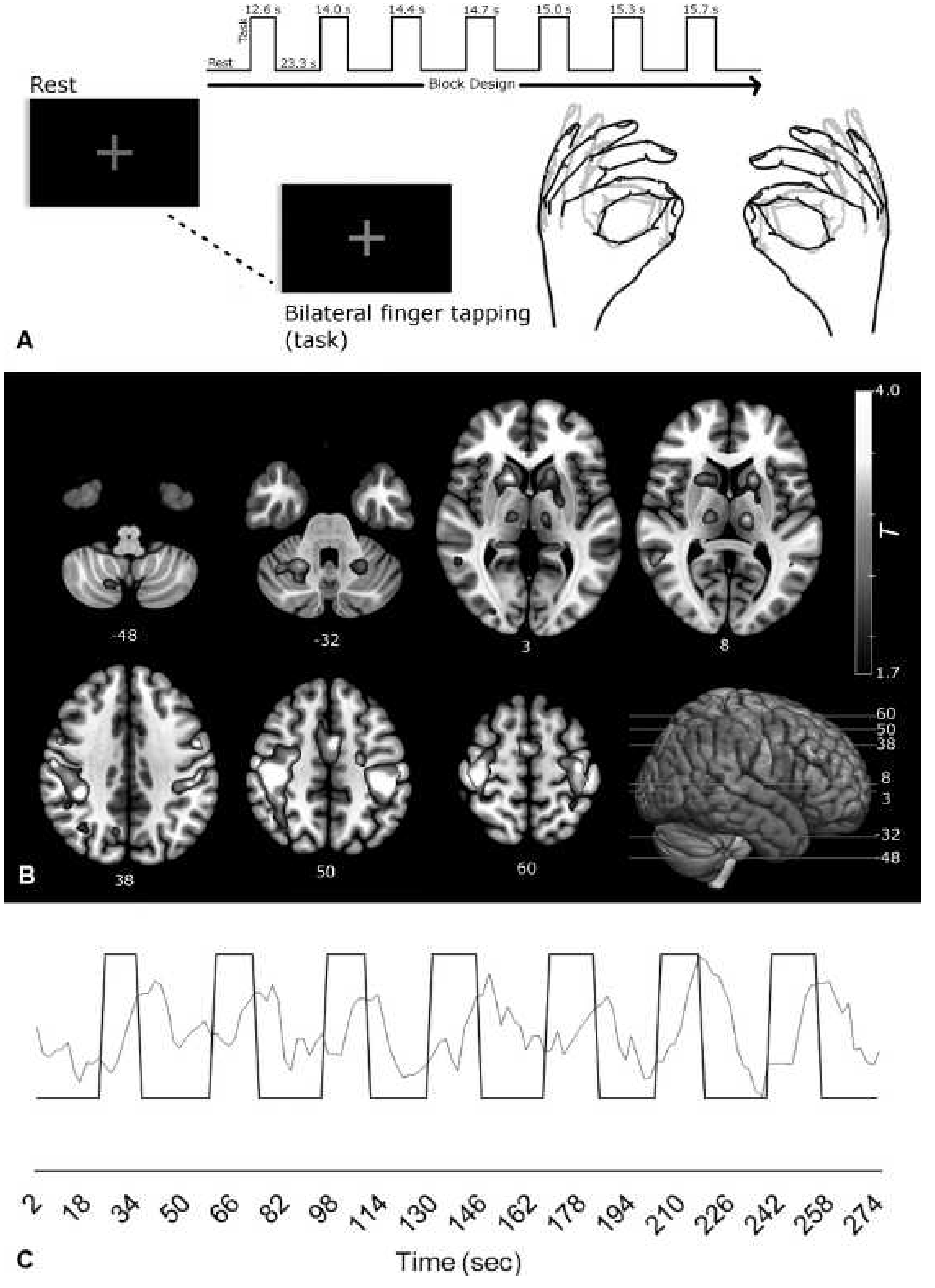
12.6 s
14.0 s
14.4 s
14.7 s
15.0 s
15.3 s
15.7 s
Task
Rest
23.3 s
Block Design
Rest
Bilateral finger tapping (task)
A
4.0
T
1.7
-48
-32
3
8
38
50
60
60
50
38
8
3
-32
-48
B
2
18
34
50
66
82
98
114
130
146
162
178
194
210
226
242
258
274
C
Time (sec)

FIGURE 37–5. **A:** Paradigm for bilateral finger tapping task during functional magnetic resonance imaging (fMRI). **B:** Data demonstrating brain activation during the finger-tapping task. **C:** Average signal intensity from activated regions shows the correspondence with the task paradigm (not tapping/tapping). (Courtesy Sneha Chenji, Ph.D.)

Critical Considerations

When acquiring and interpreting data from MR-based neuroimaging studies of children, a variety of factors merit consideration. A concise overview is presented below.

Technical Factors. With advances in computational techniques, there has been a shift from a priori hypothesis to the use of machine learning algorithms in exploring brain development. It allows the possibility to explore the relationship between several factors (e.g., between brain measures, behavioral measures, and environmental factors). However, the application of such advanced algorithms also requires large sample sizes. While current practices include sample sizes of ~25 participants, there is a growing emphasis to include sample sizes of more than 100 participants. In addition to enabling complex algorithms, large samples are necessary for replication of neuroimaging findings.

Additionally, MRI scanner strength has improved in the last decade, with humans being scanned at an ultra-high field strength of as high as 10.5T. However, there is limited research on such high field strengths in pediatric groups. Currently, most pediatric scans have been completed at 3T, with some studies performed at 7T. Adequate data supporting safety at ultra-high magnetic field strengths are required in pediatric groups. Finally, the capacity of children to remain quiet and still while being scanned is of particular concern to investigators using any imaging technique. Motion artifacts are problematic in studies as they can often render data of such poor quality as to make it unusable. Remaining still can be challenging even for healthy children, let alone those suffering from neuropsychiatric and neurodevelopmental disorders. Children are often naturally curious about the scanning environment. Some children can become quite anxious in the sometimes small, confined space of the magnet. Even with earplugs and headphones, the loud noises inherent in certain specific imaging sequences can be disturbing. Pharmacologic sedation has been used successfully in children who would be otherwise unable to tolerate neuroimaging in clinical settings, but in research studies, such procedures may entail unwanted side effects and alter the neurobiology being studied. For research studies, sedation and its potential risks are more ethically difficult to justify, especially in healthy control participants. Desensitization to the scanning environment with a “mock scanner” (Fig. 37–6) before the actual scan has been found to reduce subjective distress and heart rate, facilitating the successful completion of the study. Unfortunately, “mock scanners” can be costly and require considerable space in already crowded hospitals. Other means of desensitization, such as a careful description of the scanner, videos of its use, and attention to the questions of patients and their parents, may also be helpful in reducing anxiety and habituating children to the scanning environment to facilitate successful MR-based study (Table 37–1). Checking for motion is crucial for quality data analysis. Motion correction can be applied at the level of MRI acquisition using scanner-based tools. However, minor motion can be accounted for using post hoc motion correction techniques during data analysis.

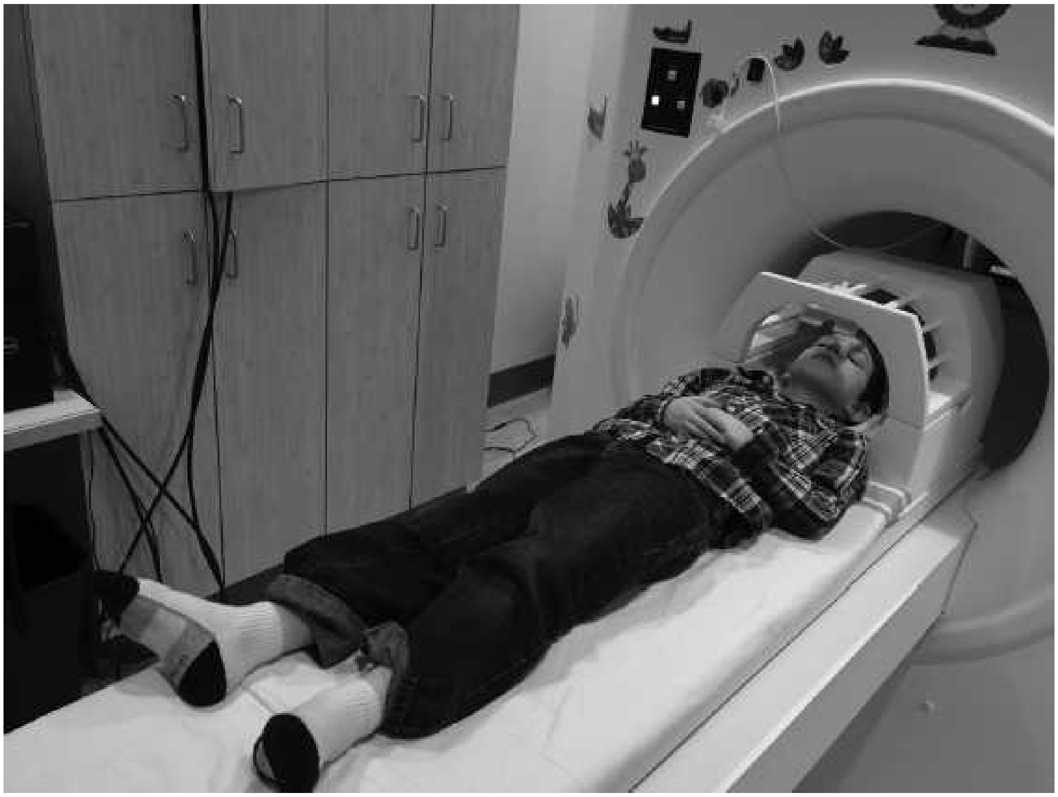

FIGURE 37–6. Child entering the mock scanner, which is designed to replicate the experience of a magnetic resonance scan, complete with the sounds that occur during the protocol to be used. (Courtesy Frank P. MacMaster, Ph.D.)

Neuroanatomical Factors. Smaller body size and brain maturation are key technical considerations in performing MR-based neuroimaging research in children. Children may not fit well into head coils designed for adults, leading to increased magnetic field inhomogeneity and decreased signal-to-noise ratios. Greater variability of the cardiorespiratory cycle in children may also contribute to increased physiologic noise during fMRI. Head motion, which is more common in children, can influence functional connectivity measures as well. However, the brain is surrounded by less bony and muscular tissue in children than in adults, so that MR surface head coils lie closer to the brain regions being imaged, possibly enhancing signal detection.

Table 37–1.
Recommendations for Scanning Young Pediatric Groups

Recommendation	Description
Prescanning training sessions	Provide links to audio, video, or online resources to MRI scanning. For instance, incorporating MRI scanning into a story (e.g., space flight) and having children practice with their parents has been found to increase compliance.
Mock scanner	Scheduling a visit to allow the child to get familiar with MRI scanning using a simulated environment can help reduce anxiety
Play therapy	On the day of scheduled visit, spending time with the child and playing short games to make the child comfortable helps reduce anxiety. Read the story again, and explain MRI scanning in simple terms (e.g., a camera that makes loud noises and takes pictures of the brain) Allow the child to take an MRI-safe toy (e.g., a teddy bear or a fidget toy) into the scanner
Parent in the MRI scanner room	Consider having a parent present in the MRI scanner room to help reduce anxiety. Parent should also be screened for MRI safety prior to entering the room.
Playing a movie	Have a list of children and youth movies that can be projected using an audiovisual device during MRI scan. This reduces the chances of children being bored and fidgeting during an MRI scan.
Giving breaks	Communicate that children can take breaks in between scans. Allow child to wiggle their toes or feet (without moving their head) between scans or allow a bathroom break at a halfway point during the scanning session. This provides flexibility and autonomy to a child during scanning.

	Talking to the child through the MRI microphone and providing a time frame for holding could help (e.g., this next set of pictures is about 5 minutes long, hold nice and still), although children do prefer to watch movies uninterrupted.
Providing incentives	At the end of the scanning session, children could be provided an incentive such as a toy or a certificate for completing their scan. For example, a group of researchers scanning children provided an astronaut certificate for their “space flight.”

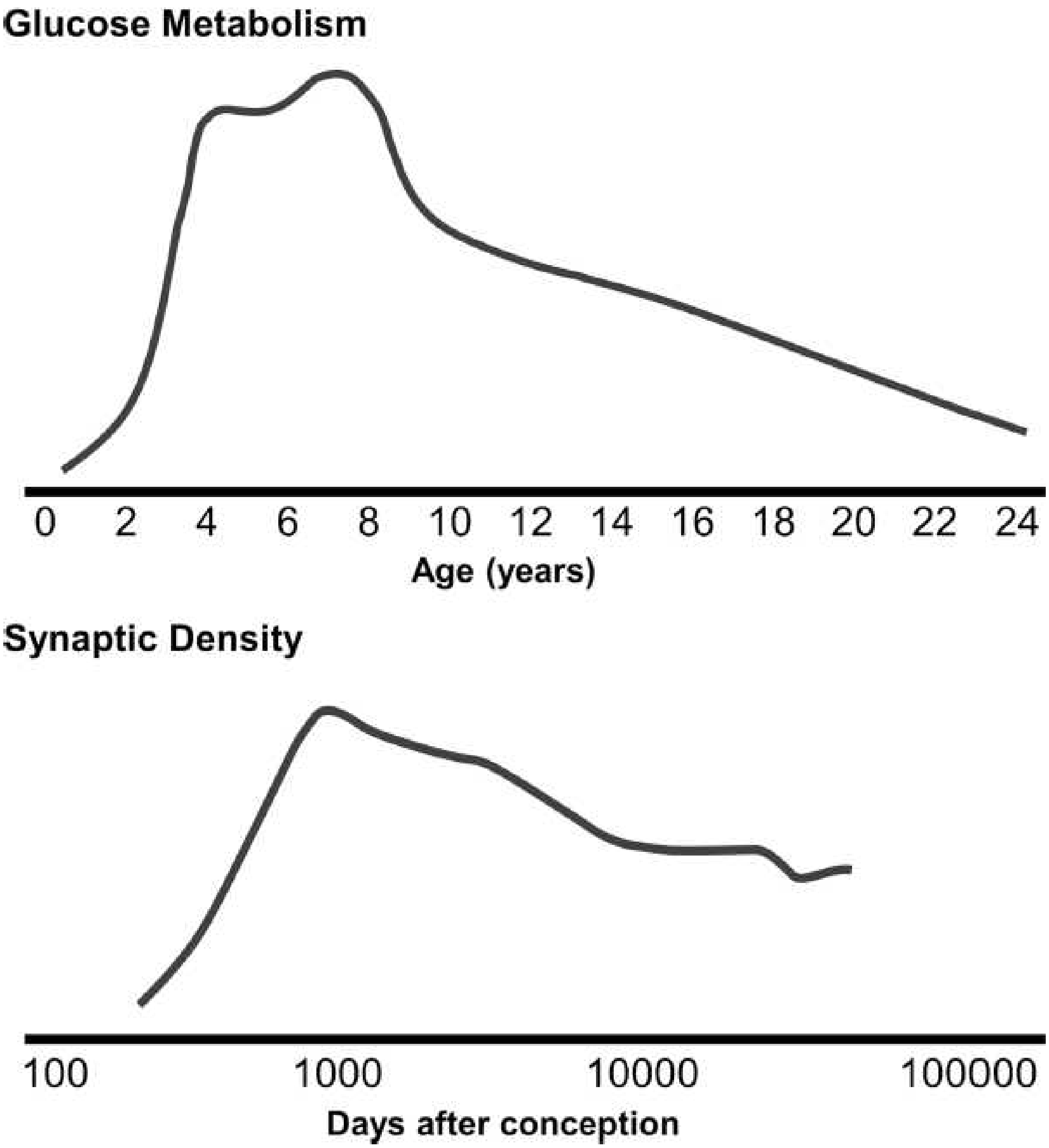

FIGURE 37–7. Schematic adapted from positron emission tomography studies of glucose metabolism (**top**) and synaptic density (**bottom**) as a function of age in children through young adulthood. Note the steep rise through childhood followed by a sharp decline just as puberty and adolescence begins. (Top, data from Chugani HT. Biological basis of emotions: brain systems and brain development. *Pediatrics.* 1998;102(5 Suppl E):1225–1229. Bottom, adapted from Rakic P, Bourgeois JP, Goldman-Rakic PS. Synaptic development of the cerebral cortex: implications for learning, memory, and mental illness. *Prog Brain Res.* 1994;102:227–243. doi:10.1016/S0079-6123(08)60543-9.)

Peak synaptic density within the brain occurs between 4 and 8 years of age (Fig. 37–7). This peak is associated with greater metabolic, and perfusion demands, which may lead to more robust increases in regional cerebral blood flow during neuronal activation and/or increased “noise” in fMRI experiments. Furthermore, the networking between brain regions is immature and may result in a weaker, more diffuse fMRI signal. Structurally, greater ratios of gray to white matter and less cerebrospinal fluid (CSF)—occupied free space around vasculature in children may lead to fundamental differences in the generation of MR signal in children versus adults. These factors must be accounted for when attempting to compare subjects of disparate ages.

The interpretation of fMRI signals in children is further complicated by structural differences because data processing typically involves the “warping” of study subjects’ brains to a standard brain atlas derived mainly from adults. The development of standardized brain atlases for children of different ages offers a solution to this methodologic problem. Still, brain volumes may vary during development even in same-aged children, so that the segmentation of brain into gray matter, white matter, and CSF may be necessary to generate individual descriptors of structural differences for use in statistical correction between study subjects. The use of study-specific templates for fMRI and voxel-based morphometry (VBM) is the most logical course. The volumetric MRI study of normal brain development in children and adolescents jointly sponsored by the National Institutes of Mental Health, Child and Human Development, and Neurologic Disorders and Stroke has recently been initiated and will generate a large normative database specific to pediatric structural maturation. Given the aforementioned variation in the developing brain and its effect on analysis methods, the importance of this resource cannot be overestimated.

Cognitive Factors. Another possible confound in the interpretation of certain fMRI paradigms in children relates to their ability to perform the cognitive tasks under investigation. Across development, children vary considerably in cognitive skill and their capacity to sustain attention and remain “on task.” Potential tasks need to be vetted thoroughly in children and adolescents before robust interpretations of functional data can be made. Cognitive-based fMRI studies comparing children and adults should account for accuracy and reaction time to avoid falsely attributing performance differences to underlying neurobiology. Assessment of effort should also be considered because some children may guess or give up when a cognitive task is too difficult or too boring. In designing a cognitive paradigm for use in an fMRI study of children, tailoring the level of task difficulty to individual capability may help to ensure consistent performance across subjects at different developmental stages. The possibility that children may use different cognitive strategies on a given task can be examined by correlating task performance with neuropsychological measures. The correlation of activation patterns with performance rather than age may help to identify true functional differences based on brain maturity rather than structural- or performance-based differences, which can vary with age and between same-aged subjects during development.

Functional neuroimaging studies of executive functions, including working memory (the ability to respond to new situations on the basis of stored information) and neurobehavioral response inhibition (inhibition of context-inappropriate responses), have demonstrated similar regional activation patterns (e.g., prefrontal cortex [PFC]) in children and adults, with children recruiting larger volumes of activation with greater signal intensity. This suggests a less specific neural network and/or less efficient processing of cognitive stimuli in childhood. Not only may children display different activation strengths in response to stimuli, they may also show regional differences as well. On a task of spatial working memory, adults activated the anterior cingulate, a structure critically involved in arousal and mediating response conflict, whereas children did not. Lower accuracy on the task among children may be related to an incompletely developed anterior cingulate, resulting in a relatively reduced ability to modulate competing responses. In general, the functional connectivity between the anterior cingulate and other cortical and subcortical structures appears to increase with age, as demonstrated in a ^{1}H-MRS study in humans.

Opposite laterality of activation may also distinguish healthy children from healthy adults on tasks of response inhibition and working memory. Children may also recruit subcortical (e.g., basal ganglia, thalamus) and limbic (e.g., amygdala, hippocampus) structures to support incompletely developed PFC in mediating response inhibition. Moreover, whereas both healthy children and adults show altered activation of the amygdala in response to fearful faces as measured by fMRI, children but not adults show activation for neutral faces, as well as fearful faces, further suggesting that more specific neural processing of information develops with age.

Clinical Factors. Recent research has highlighted variations in symptom manifestation of neuropsychiatric conditions in pediatric groups compared to adults. For example, inattention (a common symptom across many neuropsychiatric conditions) may manifest differently in children versus adults. There is a growing emphasis on integrating information at various levels (i.e., from genes to behavior) to better understand differences/similarities in the presentation of mental disorders. This serves as the guiding principle for the Research Domain Criteria (RDoC) Initiative launched by the National Institute of Mental Health (NIMH). The RDoC framework includes three basic assumptions. First, RDoC conceptualizes that psychiatric illness can be addressed as disorders of brain circuits. This is akin to lesions in neurologic conditions such as stroke. However, a brain region does not operate in isolation. The first RDoC assumption thus accounts for the dependencies between different brain regions that serve an underlying behavior (e.g., the role of the cortico-striato-thalamo-cortical circuit instead of just the caudate in ADHD), Second, it assumes that dysfunction in the brain circuits can be identified using clinical neuroscience tools (MRI, fMRI, EEG, etc.). Finally, RDoC assumes that integrating data across scientific fields (e.g., genetic and clinical neuroscience) will help in identifying biosignatures which will support clinical symptoms and management. This allows for a transdiagnostic approach in mental health. For example, the manifestation of reward processing could be studied across mood disorders, addiction, and psychosis to identify underlying genetics, brain structure or function, fear response, and determine appropriate treatment. This will aid in the development of precise and individualistic treatment based on clinical symptom manifestation and the underlying neurobiology, instead of relying on more generic diagnostic labels.

NORMAL DEVELOPMENT

Childhood and adolescence are times of great change and vulnerability. As most psychiatric disorders begin during this period, they can only be understood in the context of brain development. Brain size increases considerably in utero through 2 years of age, during which time synaptogenesis is most active. While total brain volume is comparatively static after 5 years of age, there are two dynamic processes that are ongoing through adolescence into young adulthood—synaptic pruning and myelination. Evidence suggests that increased myelination, connectivity, and changing functional activation patterns indicate that more efficient/specialized connections between brain regions are being made from childhood through adulthood.

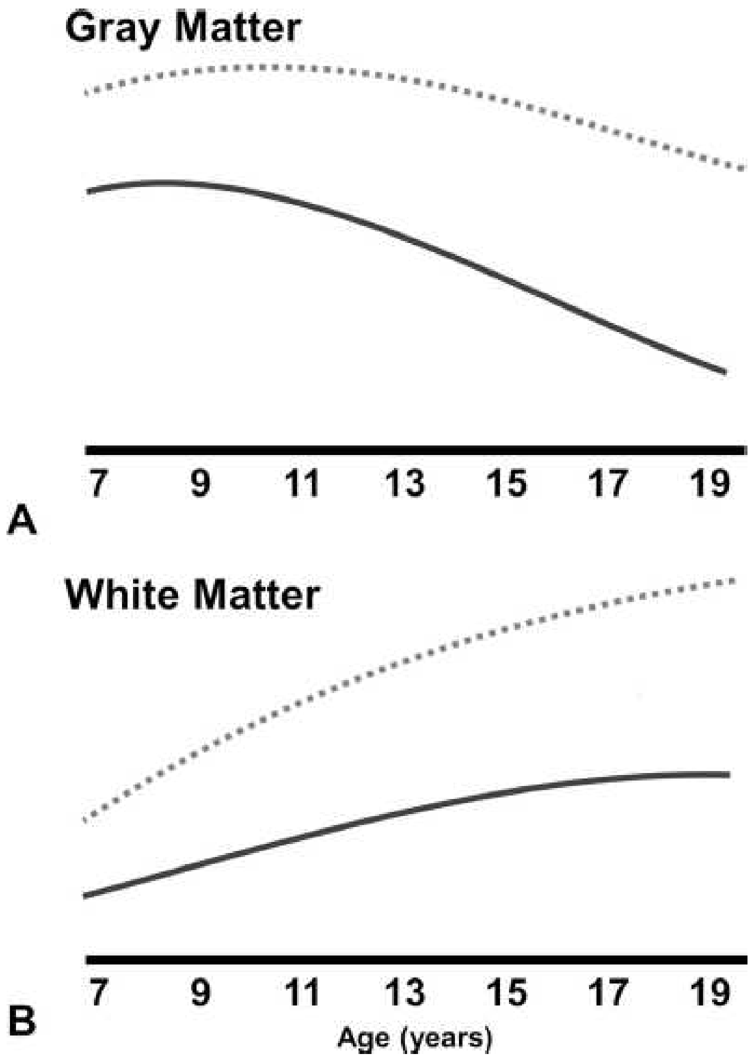

FIGURE 37–8. Schematic showing growth and aging changes in gray and white matter from childhood through adolescence. **A:** Gray matter volume reached a maximum by 9.5 years of age in females (*solid line*) and 10.5 years of age in males (*broken line*) and thereafter declined linearly. This is most pronounced in the frontal and temporal lobes. **B:** White matter volume increased rapidly until 12 to 15 years of age then slowed to eventually plateau. (Adapted from Giedd JN, Raznahan A, Alexander-Bloch A. Child psychiatry branch of the National Institute of Mental Health longitudinal structural magnetic resonance imaging study of human brain development. *Neuropsychopharmacology.* 2015;40(1):43–49. doi:10.1038/npp.2014.236, with permission from Springer Nature.)

Gray Matter

Gray matter volumes peak in midchildhood to early adolescence, earlier in females than males. They then plateau briefly and show a decrease following puberty (Fig. 37–8A). This pattern of gray matter maturation parallels changes in local cerebral metabolic rates for glucose in the cerebral cortex as a function of age (Fig. 37–7). These developmental changes most likely represent an increase in the efficiency of neural processing via the pruning of excessive synaptic connections. It is not clear which cellular changes (i.e., loss of neurons, dendrites, axons, and/or neuropil) are responsible for the reduction of gray matter. The relationship between metrics of brain morphology and age is so strong that cortical thickness, as measured using MRI, can accurately predict an individual's age during childhood and adolescence. When considering the developmental basis of normal and abnormal cognitive/emotional function, it is important to note that it is heterochronous. That means that the timing, degree, and rate of cortical and subcortical gray matter change appear to be regionally specific.

White Matter

The volume of white matter increases roughly linearly in the brain during childhood and adolescence. This is consistent with myelinization during this time (Fig. 37–8B). Growth in cortical white matter appears to be regionally specific. The most dramatic increase in white matter occurs in the dorsolateral PFC (DLPFC) between 5 and 17 years of age. Consistent with the progressive development of executive function through adolescence and into early adulthood, a segmentation analysis of MRI data found that gray matter reduction in the frontal cortex continued into young adulthood after gray matter volumes had stabilized in other cortical regions. The "reduction" of frontal lobe gray matter was thus hypothesized to reflect the effects of an increase in myelination (and hence the relative amount of white matter). This interpretation is supported by the inverse relationship between cortical gray matter and brain growth—presumably due to white matter increase—in the frontal lobes. It has been suggested that myelinization may be as important to brain maturation as synaptic pruning. Indeed, the capacity for attention—a function largely attributable to the frontal lobe—has been shown to increase with larger areas of the anterior cingulate cortex over development, possibly the result of increased myelination. The corpus callosum (CC)—the white matter tract connecting the hemispheres—demonstrates a particularly dynamic growth pattern (Fig. 37–9). Unlike the posterior–anterior progression of brain development described in much of the preclinical literature, CC growth occurs in a rostrocaudal wave between 3 and 15 years of age (see

Fig. 37–9). It may be that CC growth parallels synaptogenesis in associated cortical regions rather than pruning and myelinization. In a large study of 103 healthy participants (5 to 32 years of age), nonlinear development of major white matter pathways was observed. There were significant within-subject changes in the majority of children and early adolescents, with these changes being completed by late adolescence for projection and commissural tracts. In contrast, association tracts demonstrated postadolescent within-subject maturation of both fractional anisotropy and mean diffusivity.

Biologic Sex Differences

Biologic sex also plays a significant role in brain development. In a large longitudinal MRI study of cortical volumes in 145 healthy children and adolescents (4 to 20 years of age), gray matter volume was found to peak and then decline after puberty in the parietal, frontal, and temporal lobes but on different time scales (Table 37–2). In contrast, gray matter in the occipital cortex increased through early adulthood (approximately 20 years of age). With regard to biologic sex differences, the peak for gray matter volume in the frontal and parietal lobes occurs nearly 1 year earlier in girls than boys. This seems to coincide with sex-specific differences in the timing of puberty and suggests the possible influence of gonadal hormones on brain development.

Sex-specific and age-related differences in gray matter volumes have also been noted in the basal ganglia, hippocampus, and amygdala (Table 37–3). Although reduction of basal ganglia volumes has been demonstrated in both sexes through young adulthood, gray matter loss in this region is most dramatic in boys, especially in the caudate between the ages of 6 to 12 years. It is possible that this maturational difference is related to the preponderance of basal ganglia–associated psychiatric disease (e.g., OCD and ADHD) in boys during this critical developmental window. In contrast to the basal ganglia and some areas of the cerebral cortex, the volume of the amygdala and hippocampus increases with age. Volumetric and functional abnormalities of these regions have been associated with depression and anxiety disorders, known to occur more commonly in females, particularly after the onset of puberty. The relatively smaller amygdala and larger hippocampal volumes that emerge in girls than in boys around the time of puberty may be particularly relevant in this regard. The preponderance of estrogen receptors in the amygdala and androgen receptors in the hippocampus suggests that gonadal hormones may influence age-specific changes in these regions and possibly the emergence of mood and anxiety disorders in adolescent girls. The pituitary gland also demonstrates a striking sex difference, with females developing a larger pituitary gland than males, particularly during adolescence. This may be associated with the differential timing of puberty in males and females.

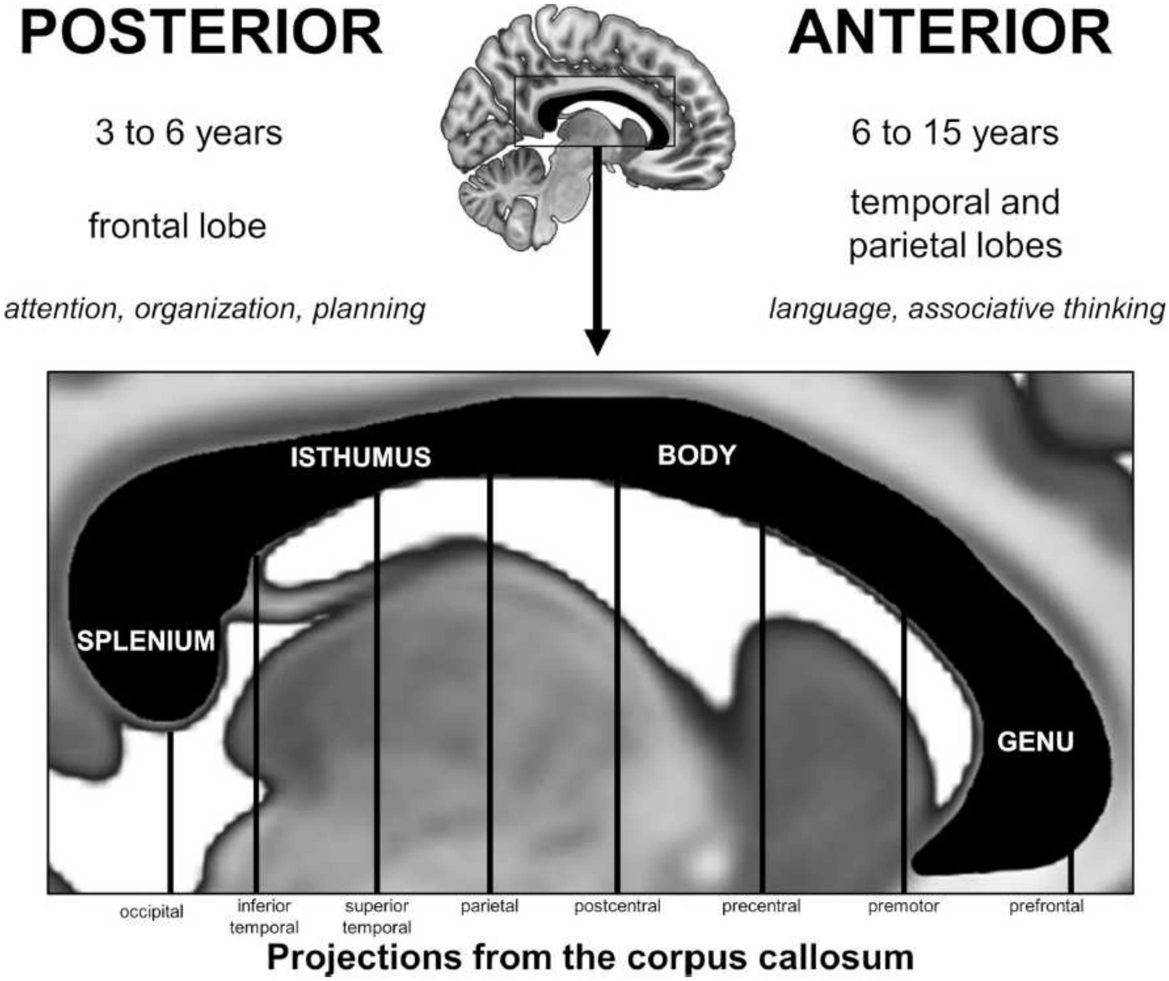
POSTERIOR
3 to 6 years
frontal lobe
attention, organization, planning
ANTERIOR
6 to 15 years
temporal and parietal lobes
language, associative thinking
ISTHUMUS
BODY
SPLENIUM
GENU
occipital
inferior temporal
superior temporal
parietal
postcentral
precentral
premotor
prefrontal
Projections from the corpus callosum

FIGURE 37–9. Schematic showing the corpus callosum, its subdivisions, and projections. (Data from Thompson PM, Giedd HN, Woods RP, MacDonald D, Evans AC, Toga AW. Growth patterns in the developing brain detected by using continuum mechanical tensor maps. *Nature.* 2000;404(6774):190–193.)

SPECIFIC NEUROPSYCHIATRIC DISORDERS

Attention-Deficit/Hyperactivity Disorder

The phenomenology, nosology, and empirically based treatment for ADHD have been well delineated. This makes pediatric ADHD a prime candidate for neurobiologic studies of development. ADHD is also relatively consistent in expression across the lifespan. Therefore, the results of studies of the disorder in children, closer to illness onset, are applicable to adult subjects. Various neuroimaging strategies, including PET, fMRI, volumetric/anatomic MRI, and ^{1}H-MRS, have been used to better elucidate the pathogenesis and mechanisms of treatment response in ADHD.

Table 37–2.
Peak Cortical Gray Matter Volumes by Age and Sex

Cortical Region	Girls (Age in Years)	Boys (Age in Years)
Parietal	10.2	11.8
Frontal	11.0	12.1
Temporal	16.7	16.5

Data from Giedd JN, Raznahan A, Alexander-Bloch A, Schmitt E, Gogtay N, Rapoport JL. Child psychiatry branch of the National Institute of Mental Health longitudinal structural magnetic resonance imaging study of human brain development. *Neuropsychopharmacology.* 2015;40(1):43–49.

Prefrontal Cortex. Core deficits in attention, arousal, and impulse control characterize children with ADHD. The PFC modulates these functions and is the seat of “executive” control. It is not surprising that the PFC has been implicated in the pathophysiology of ADHD. Smaller prefrontal cortical volumes, particularly of the superior and inferior frontal gyri, have been consistently noted in volumetric MRI studies of children with ADHD. However, one study found girls with ADHD to have normal frontal lobe volumes; this finding may be related to the lower prevalence of ADHD in females than males and an underlying sex difference in the pathophysiology of the disorder. Lower global cerebral metabolism in girls with pediatric ADHD than in male patients and normal controls has been noted, further suggesting sex-specific differences in the neurobiology of ADHD. Smaller PFC volumes correlate with impaired response inhibition in children with ADHD. A large-scale study of children with ADHD found that the structural maturation of the PFC lags behind typically developing children by 2 to 3 years. When considering the potential behavioral ramifications of such a lag in development, the implications for the symptoms of ADHD are clear. In addition, using phosphorus magnetic resonance spectroscopy (^{31}P-MRS), lower levels of membrane phospholipid (MPL) precursors were observed in older ADHD children as compared to healthy children. Lower MPL precursor levels suggest a possible underdevelopment of neuronal processes and synapses in ADHD. It was suggested that bottom-up developmental dysfunction, starting in the basal ganglia, might result in ADHD (Fig. 37–10). During a response conflict task that typically involves the anterior cingulate in normal subjects, adult patients with ADHD failed to demonstrate a significant anterior cingulate response. Of interest, both increased and decreased PFC activation during response inhibition have been found in different studies. These discrepant findings may be accounted for by the inverse relationship noted between symptom severity and perfusion in children with ADHD. Specifically, perfusion-based functional neuroimaging studies of "sicker" patients would start from a lower baseline than studies of less symptomatic patients. This may result in a reduction in signal when the PFC is involved.

Table 37–3.
Sex Differences in Brain Volumes

Region	Gender Difference (Age in Years)
Total brain volume	10% larger in boys (4–18)
Cerebellum	8% larger in boys (4–18)
Caudate and putamen	More dramatic decrease in boys (preadolescent)
Amygdala	Larger in girls (postadolescent)
Hippocampus	Larger in boys (postadolescent)

Data from Giedd JN, Raznahan A, Alexander-Bloch A, Schmitt E, Gogtay N, Rapoport JL. Child psychiatry branch of the National Institute of Mental Health longitudinal structural magnetic resonance imaging study of human brain development. *Neuropsychopharmacology.* 2015;40(1):43–49.

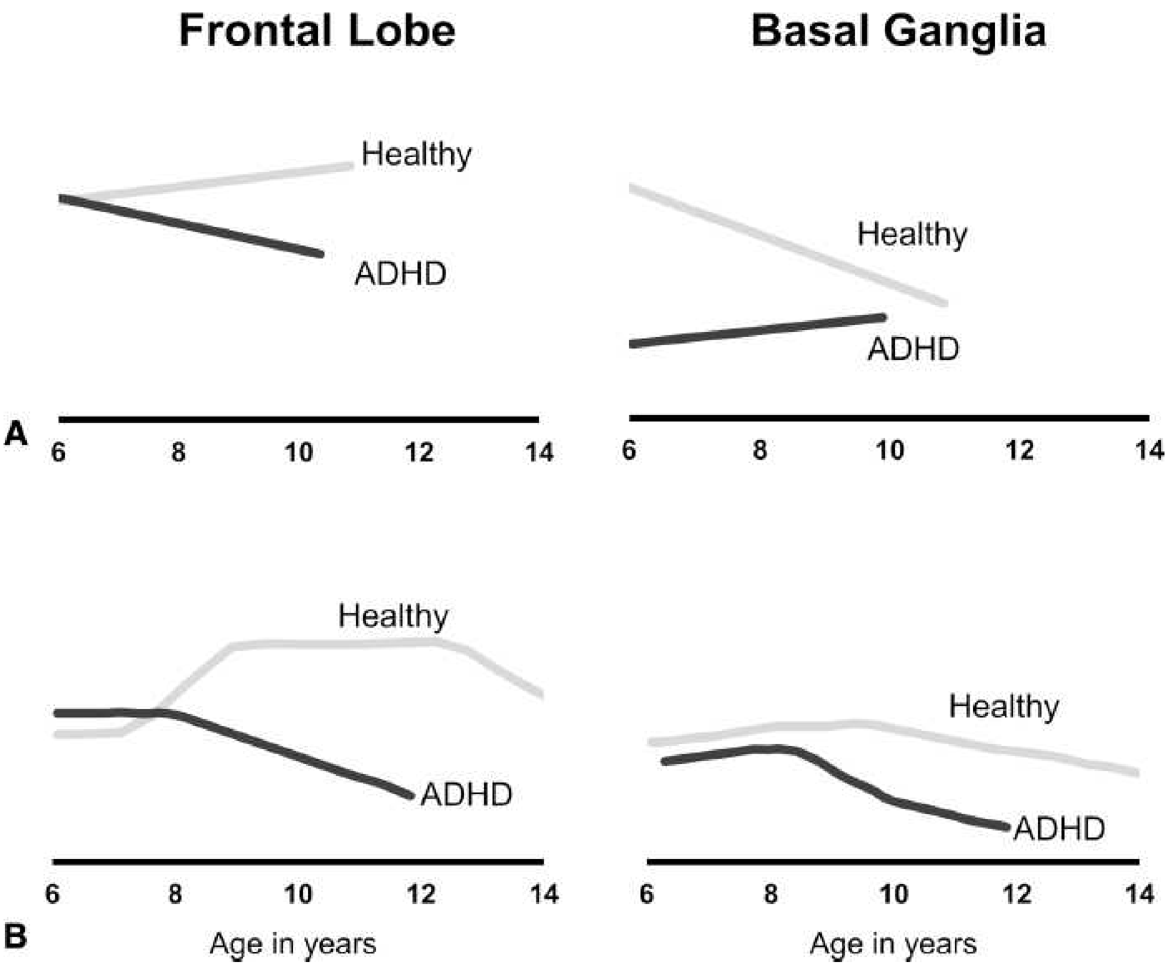

FIGURE 37–10. Schematic showing (**A**) cortical thickness and (**B**) membrane phospholipid (MPL) precursor levels as associated with age in children with attention-deficit/hyperactivity disorder (ADHD) (*dark line*) and typically developing healthy children (*light line*) in the frontal lobe and basal ganglia. (**A:** Left, data from Yang XR, Carrey N, Bernier D, MacMaster FP. Cortical thickness in young treatment-naïve children with ADHD. *J Atten Disord.* 2015;19(11):925–930. Right, adapted from Carrey N, Bernier D, Emms M, et al. Smaller volumes of caudate nuclei in prepubertal children with ADHD: impact of age. *J Psychiatr Res.* 2012;46(8):1066–1072. **B:** reproduced with permission from Stanley JA, Kipp H, Greisenegger E, et al. Evidence of developmental alterations in cortical and subcortical regions of children with attention-deficit/hyperactivity disorder: a multivoxel in vivo phosphorus 31 spectroscopy study. *Arch Gen Psychiatry.* 2008;65(12):1419–1428. Copyright © 2008 American Medical Association. All rights reserved. doi: 10.1001/archgenpsychiatry.2008.503.)

Basal Ganglia. Another aspect of the frontal–striatal loop, the basal ganglia, has been repeatedly demonstrated to be abnormal (with regard to volumes and reversed asymmetry) in pediatric ADHD. The deficits in response inhibition and attention noted in children with ADHD have also been associated with smaller volumes and lower activation of the right and left caudate. The basal ganglia have been implicated in ADHD, TS, and OCD, and may account for the high degree of comorbidity among these conditions. ^{1}H-MRS studies have noted lower levels of the putative neuronal marker NAA and greater levels of Cho, a critical membrane component, in the striatum bilaterally in pediatric ADHD patients. More recently, ^{1}H-MRS studies of pediatric ADHD patients identified greater frontal–striatal Glx levels in medication-free pediatric patients with ADHD 7 to 16 years of age as compared to age- and sex-matched controls. Lower NAA levels suggestive of reduced neuronal viability may result from excess activity of Glx neurons, which can be toxic. Moreover, ADHD children also show reduced striatal GABA. Given evidence that changes in Cho, NAA, Glx, and GABA may be reversible in other neuropsychiatric conditions, longitudinal ^{1}H-MRS studies of these metabolites in ADHD are critical. As mentioned above, ADHD patients have lower PFC and basal ganglia MPL precursor levels than healthy children. In the basal ganglia, these deficits appear earlier, possibly indicating a bottom-up change in developmental trajectory. No significant differences were noted in the superior temporal region.

Cerebellum and Other Regions. Structural neuroimaging studies have demonstrated smaller cerebellar volumes in girls and boys with ADHD. This smaller volume may be localized to the vermis, a brain region involved in maintaining attention. In an fMRI study of motor function in pediatric ADHD patients, greater impairment in task performance was associated with lower cerebellar activation. Furthermore, motor impairment in ADHD patients was also associated with lower temporal lobe activation but greater thalamic activation. Moreover, a large-scale study showed that the temporal pole and fusiform gyrus are also significantly thinner in ADHD children compared to controls. This suggests that complex network-wide dysfunctions might underlie the symptoms of ADHD. This hypothesis is supported by the findings of smaller anterior and poster aspects of the CC in ADHD patients.

Medication Effects on Brain Function in ADHD. Methylphenidate (MPH) is the first-line treatment for ADHD. Studies found that MPH changes frontal–striatal activation in pediatric patients with ADHD, with greater PFC and striatal activation as measured by fMRI demonstrated in

ADHD patients. In addition, MPH was found to improve ADHD patient performance on the response inhibition task. Predrug activation during the task was greater among patients than with controls, suggesting that frontal cortical hyperactivity in ADHD may play a compensatory role that is further enhanced by MPH. In the striatum, predrug activation was lower in pediatric patients with ADHD than in pediatric controls. This deficit appeared to be corrected by MPH. MPH-treated healthy children also improved in their ability to inhibit inappropriate response. However, improved performance in healthy children was associated with a distinct pattern of frontal–striatal activation. Aligned with this, another study showed that MPH normalized fronto-striato-thalamo-parietal dysfunction and upregulated ventral fronto-temporal regions in adolescents with ADHD during a sustained attention task. These differences in MPH-related activation could reflect a fundamental abnormality of frontal–striatal dopamine systems in patients with pediatric ADHD as compared to healthy children. Positive correlation between ADHD symptom severity and increased midbrain accumulation of ^{18}F-dihydroxyphenylalanine (^{18}F-DOPA), a marker of dopamine synthesis, lends further support to the role of dysfunctional dopamine systems in ADHD. There are data suggesting a developmental shift in dysfunction of dopamine from the midbrain in childhood and adolescence to frontal cortex during adulthood in ADHD. Interestingly in children with ADHD, responders to MPH have greater gray matter concentrations compared to nonresponders in the caudate. This is similar to a larger posterior hippocampus predicting better response to antidepressants in adults with depression. The predictive power of imaging for treatment response has yet to be fully explored in child and adolescent psychiatry.

A recent meta-analysis of fMRI studies of the effect of stimulants on brain function in ADHD examined 14 data sets with 212 children with ADHD. Stimulants consistently enhanced right inferior frontal cortex and insula activation, with a smaller effect on putamen activation. This change in activation was present in those receiving MPH alone. The inferior frontal cortex and insula play a role in cognitive control and an increase in activation may represent the positive clinical response found with treatment.

Depressive Disorders and Bipolar and Related Disorders

Biologic studies of depressive disorders and bipolar and related disorders are providing increasingly converging evidence that alterations in medial temporal (e.g., amygdala, hippocampus) and prefrontal–striatal circuits may be involved in the symptomatology of patients with MDD and BP-I (Fig. 37–11). Although the literature is not entirely consistent, it appears that lesions to the left hemisphere are more associated with depression, whereas right-sided lesions tend to be more associated with mania.

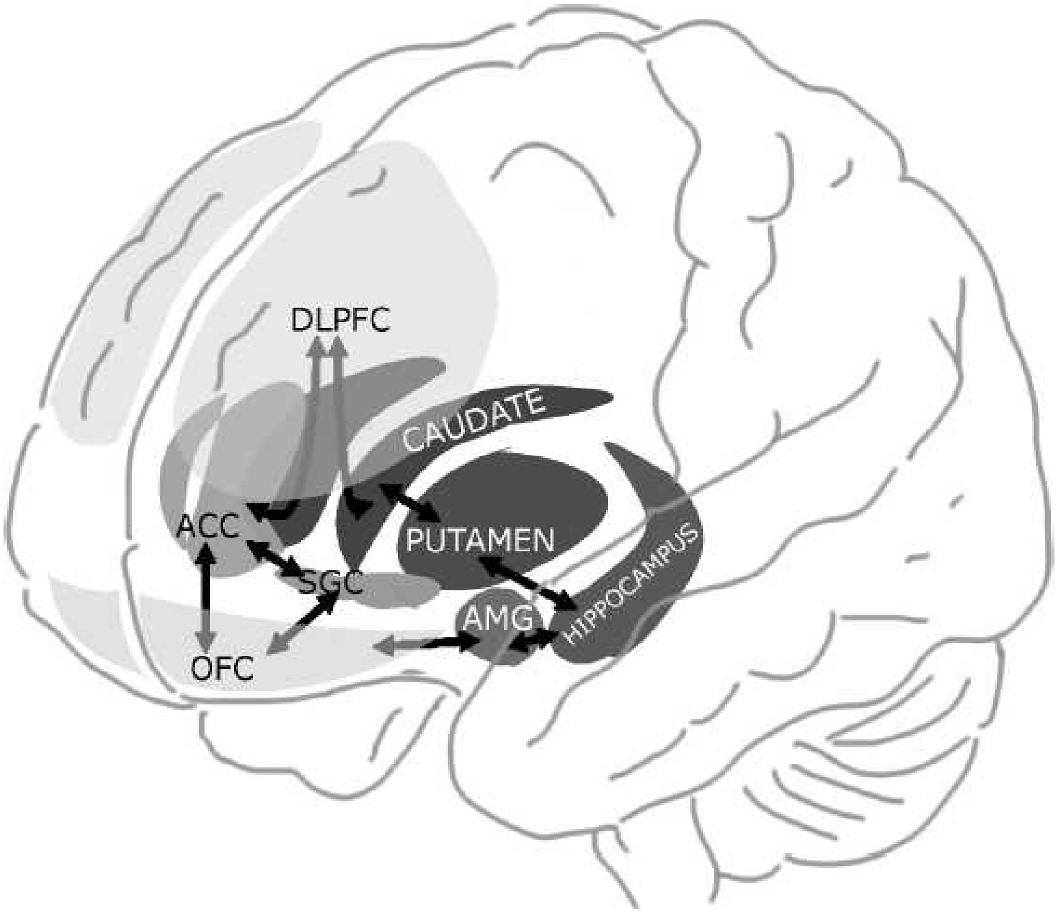

FIGURE 37–11. Brain regions implicated in the pathophysiology of major depressive disorder. ACC, anterior cingulate cortex, AMG, amygdala, DLPFC, dorsolateral prefrontal cortex, SGC, subgenual cingulate cortex. (Courtesy Sneha Chenji, PhD.)

Frontal Lobe and Striatum. In adults and adolescents with MDD, smaller overall frontal lobe volumes have been among the more consistent findings in structural neuroimaging studies. In addition, adolescents with MDD were found to have significantly smaller frontal white matter volumes but larger frontal gray matter volumes than healthy comparison adolescents, even after controlling for age and whole brain.

Cortical thickness in the middle frontal gyrus bilaterally and the left caudal anterior cingulate was found to be greater in adolescents with MDD as compared to healthy control participants. Replicating and extending this finding in adults, thicker caudal anterior cingulate was also found with first-episode MDD. Left middle frontal gyrus was inversely correlated with age in MDD subjects. This may indicate an altered developmental path in youth with MDD. MRI and PET investigation in adults with MDD and BP-I have reported strikingly localized, left-sided smaller volume in the subgenual region of the prefrontal cortex (SGPFC). This smaller SGPFC was associated with lesser cerebral blood flow. These effects were most prominent in MDD and BP-I patients who had at least one first-degree relative with MDD or BP-I (familial MDD or BP-I). In patients with MDD or BP-I without family history of MDD or BP-I (nonfamilial MDD or BP-I), no significant differences were observed in the left or right SGPFC. In an additional sample, a smaller left SGPFC volume was also noted in 30 young women with early-onset MDD (18 to 23 years of age) and 18 middle-aged women with recurrent MDD (24 to 52 years of age) than in age-matched controls (19% difference).

A study of 22 psychotropic-naive pediatric patients with MDD (9 to 17 years) compared to 22 case-matched controls found that patients with nonfamilial MDD had significantly larger left-sided but not right-sided prefrontal cortical volumes than patients with familial MDD and controls. In both the familial and nonfamilial pediatric MDD subgroups, left prefrontal volumetric abnormalities remained significant after controlling for severity of anxiety. This suggests that this region is integral to the pathogenesis of depression in children. Reduced left-sided prefrontal cortical volume was associated with increased severity of depression in familial but not nonfamilial patients with MDD. Longer illness duration also tended to be associated with reduced left prefrontal cortical volume in familial but not nonfamilial MDD patients. These results, coupled with investigations in adults, suggest that left-sided prefrontal cortical volume reduction in familial MDD patients may reflect degeneration of left prefrontal cortical volume, whereas increased left prefrontal cortical volume in psychotropic-naive pediatric patients with nonfamilial MDD may reflect alteration in the developmental maturation of the PFC.

Left PFC has been further implicated in pediatric MDD in two recent ^{1}H-MRS investigations. Increased left orbitofrontal Cho levels were observed in 22 depressed adolescents as compared to 43 control adolescents.

Increased orbital frontal Cho levels were most prominent in adolescent patients with familial MDD as compared to both nonfamilial MDD patients and controls (Dr. Perry Renshaw, personal communication). Increased left but not right DLPFC Cho concentrations were also observed in treatment-naive pediatric patients with MDD as compared to age- and sex-matched controls. Increased Cho concentrations in pediatric MDD patients versus controls persisted even after controlling for severity of anxiety. Mean left dorsolateral prefrontal cortical (DLPFC) Cho concentrations were increased in familial MDD patients as compared to both nonfamilial MDD patients and healthy pediatric controls. Lower Cho concentrations were also observed in the right anterior cingulate. Alterations in Cho as measured by ^{1}H-MRS have been hypothesized to result from abnormalities in signal transduction involving phosphatidylcholine second messenger cascades and could reflect changes in glucose metabolic rates and/or neuroendocrine function. Increased left prefrontal Cho levels may also be consistent with functional neuroimaging studies demonstrating reduced prefrontal cortical metabolism in adults with MDD. Functional neuroimaging studies have demonstrated that reduced prefrontal cortical activation in depressed adolescents and adults normalizes following treatment-induced symptom remission. Because Cho alterations in adult patients with MDD appear to be reversible with treatment, the Cho resonance may represent a potential target for treatment intervention in pediatric MDD, warranting longitudinal ^{1}H-MRS investigations in these patients. Finally, consistent with findings in adult patients with MDD, lower anterior cingulate Glx concentrations have been noted in pediatric MDD patients than in healthy controls. Glx was also associated with severity of illness in the MDD patients. In addition, distinct differences were noted in familial versus nonfamilial MDD patients. Subsequent investigation, using a more robust spectral fitting method, found that the lower anterior cingulate Glx in patients was due to glutamate but not glutamine. In 20 medication-free youth with MDD, lower anterior cingulate GABA was observed as compared to controls, with this difference being more pronounced in those with anhedonia. Recently, a small case series in adolescents with MDD replicated a finding in adults with MDD that showed an increase in DLPFC glutamate in response to repetitive transcranial magnetic stimulation treatment.

In contrast, recent ^{1}H-MRS investigation in 10 medication-free children with BP-I versus 10 healthy children found increased glutamate and glutamine (Glx) levels bilaterally in the frontal lobes and basal ganglia in children with BP-I. Cho levels did not differ in frontal lobes or basal ganglia in patients with BP-I or controls. Increased lipid levels were also found bilaterally in the frontal lobes of children with BP-I. In contrast, ^{1}H-MRS investigation of anterior cingulate cortex in pediatric patients with BP-I versus healthy pediatric controls found no significant differences in Cho, Glx, GABA, or the neuronal marker NAA. Consistent with findings in adults, effective lithium treatment was, however, associated with a reduction of the ^{1}H-MRS neurochemical marker myo-inositol in the anterior cingulate cortex of children with BP-I. A recent VBM study found that subjects with BP-I had reduced gray matter volume in the left DLPFC as compared to healthy controls.

In concert with the frontal lobe, the striatum plays an important role in executive and emotional processing. Evidence from adult imaging studies implicates the striatum in the pathophysiology of depressive disorders. A recent study noted that patients with MDD had significantly smaller right striatum (2.3% smaller) and caudate volumes (3.9% smaller) than healthy subjects. Furthermore, right striatum volumes were inversely correlated with increased severity of depression. Moreover, increased functional connectivity between the left DLPFC and the striatum in depressed adults is associated with better treatment response to noninvasive brain stimulation therapy. These findings provide fresh evidence for abnormalities in the striatum of pediatric MDD and bolster the hypothesis of dysfunction in the cortico-limbic–striatal–pallidal–thalamic circuit in the pathophysiology of MDD and serve as a foundation for the development of better treatment approaches for pediatric MDD patients.

Amygdala, Hippocampus, and Pituitary. Although prefrontal cortical involvement in pediatric MDD may be specific to the depressive state, abnormalities of the amygdala in depressed children and adolescents may be more associated with the severity of anxiety than with the severity of depression. Ratios of left and right amygdala to the hippocampal volume were found to be significantly larger in treatment-naive pediatric patients with MDD than in 23 age- and sex-matched controls. Furthermore, ratios of amygdala to hippocampal volume in pediatric patients with MDD were positively correlated with the severity of anxiety but not with the severity of depression or duration of depressive illness. Amygdala volumes were larger bilaterally in patients with MDD than in controls. In a magnetic resonance spectroscopic imaging ^{1}H-(MRSI) study of the amygdala in adolescents with depression, lower choline-to-creatine (Cho-to-Cr) ratios were noted in the left, but not right, amygdala in patients than in controls. This change in cytosolic choline may be secondary to corticosteroid-neuroendocrine–driven changes, given that increases in cortisol can result in lower Cho concentrations. Right and left hippocampal volumes were smaller in MDD patients. Not only are comorbid anxiety symptoms and disorders extremely common in patients with MDD but also excessive fearfulness in adolescence has been found to predict the development of MDD in early adulthood. Smaller left hippocampal volumes have been noted in pediatric patients with MDD with a familial history of depression than in controls. It may be that patients with a family history of depression are driving the smaller hippocampal volumes noted in previous reports. Smaller right hippocampal volumes have been found in youth with MDD with a normal body mass index (BMI) compared to MDD youth with high BMI and healthy controls. However, smaller left hippocampal volumes were inversely related to increased BMI in the high BMI MDD group. Therefore, variabilities in factors such as familial history and BMI should always be considered. In a closer examination of the hippocampus using ^{1}H-MRSI, reduced left, but not right, NAA concentration was noted in MDD patients and those with treatment-resistant MDD (i.e., unresponsive to at least two medications) as compared to age- and sex-matched controls. This indicates that neuronal viability may be reduced in pediatric MDD.

In adolescents and adults with MDD, at rest, the amygdala shows abnormal connectivity with the hippocampus, parahippocampal gyrus, insula, caudate, and brain stem. Abnormal connections in adolescents with MDD were more prominent in the cognitive control network and default mode network. Much of this connectivity has been inversely correlated with depressive symptoms. Use of emotional facial expressions as fMRI stimuli is common as a probe of affective processing in the amygdala in adults and children. An fMRI study of amygdala response to fearful faces in children with generalized anxiety disorder (GAD) and panic disorder compared to children with MDD and healthy pediatric controls demonstrated altered activation of the amygdala in both patients with MDD and patients with GAD/panic disorder, with the magnitude of signal change in amygdala activation being most associated with the severity of the child’s rating of everyday anxiety. This increase in amygdala arousal to negative stimuli was recently replicated in 32 adolescents with MDD and 23 age- and gender-matched controls.

It should be noted that greater amygdala volume is one of the more consistent findings in adults with BP-I. fMRI studies in adults with BP-I have also demonstrated increased amygdala activation in response to fearful facial affect. More recently, an event-related fMRI study reported that pediatric patients with BP-I demonstrated reduced memory for emotional faces and had increased neural activation in the striatum and anterior cingulate cortex when successfully encoding happy faces, and in the orbitofrontal cortex when successfully encoding angry faces. There were no between-group differences in neural activation during fearful face encoding. Consistent with adult BP-I studies, multiple fMRI studies reported increased amygdala activation in response to fearful faces. Differences across studies of BP-I may have to do with differences in disease state. Longitudinal MRI studies of amygdala volume and function in pediatric patients with MDD are clearly warranted to determine whether increased amygdala volume and altered activation level may serve as potential biomarkers for BP-I.

More recently, larger pituitary gland volumes were noted in adolescents with MDD than in case-matched controls. This was replicated in an independent treatment-naive sample. In the second sample, it was noted that the difference in volume appeared to be driven by nonfamilial MDD patients. Of interest, depressed patients with BP-I demonstrated larger pituitary gland volumes than controls. This was not consistent with a published adult study of BP-I and a pediatric sample. The main difference between these reports is the disease state (depressed vs. mainly euthymic). This reinforces the need for studies of BP-I across the illness course and the need for further studies to differentiate MDD from BP-I.

Obsessive-Compulsive Disorder

There is a growing literature pointing to a dysfunction in ventral prefrontal–striatal–thalamic and medial temporal circuitry in OCD (Fig. 37–12). The phenomenology, nosology, and empirical treatment for childhood OCD have been well established, rendering it an ideal candidate for developmental neurobiologic study. In comparison to depressive and bipolar and related disorders, OCD is more consistent in its expression across the lifespan. This allows investigators to study the illness close to onset, with the results remaining applicable to adult subjects. Studies in pediatric OCD have focused on treatment-naive populations. This minimizes potential confounds such as illness duration and medication exposure. As described in the following sections, evidence is converging that suggests a pharmacologically reversible, Glx-mediated thalamic–cortical–striatal dysfunction in pediatric OCD that may serve as a reliable pathophysiologic and treatment response marker for the illness (Fig. 37–13).

Basal Ganglia. Volumetric neuroimaging studies in pediatric patients with OCD have demonstrated smaller basal ganglia volumes in patients that were associated with OCD symptom severity but not illness duration. In addition, smaller globus pallidus volumes were also noted in treatment-naive pediatric OCD patients than in controls. Recently, however, 26 adolescents with OCD showed greater gray matter volume in the bilateral caudate as compared to healthy controls (*N* = 26). This may reflect fundamental differences in what properties are measured using VBM as compared to the manual tracing technique. It was recently shown, in 29 adolescents with OCD, that those who responded to an allocated treatment (selective serotonin reuptake inhibitor [SSRIs] or cognitive behavioral therapy [CBT]) had larger caudate volumes than nonresponders and healthy controls. In another study, greater caudate Glx concentrations, as measured by ^{1}H-MRS, were also observed in treatment-naive pediatric OCD patients than in age- and sex-matched healthy pediatric controls (Figs. 37–14 and 37–15). No significant differences in Glx concentrations were observed between OCD patients and controls in the occipital cortex, a region not implicated in the pathogenesis of OCD. After 12 weeks of treatment with the SSRI, paroxetine, but not with CBT, caudate Glx concentrations in OCD patients decreased to levels that did not differ significantly from those in controls (Fig. 37–16). Greater pretreatment caudate Glx concentrations predicted a more robust response to paroxetine in OCD patients. More recent investigation suggested that caudate Glx concentrations may persist even after medication discontinuation in pediatric OCD.

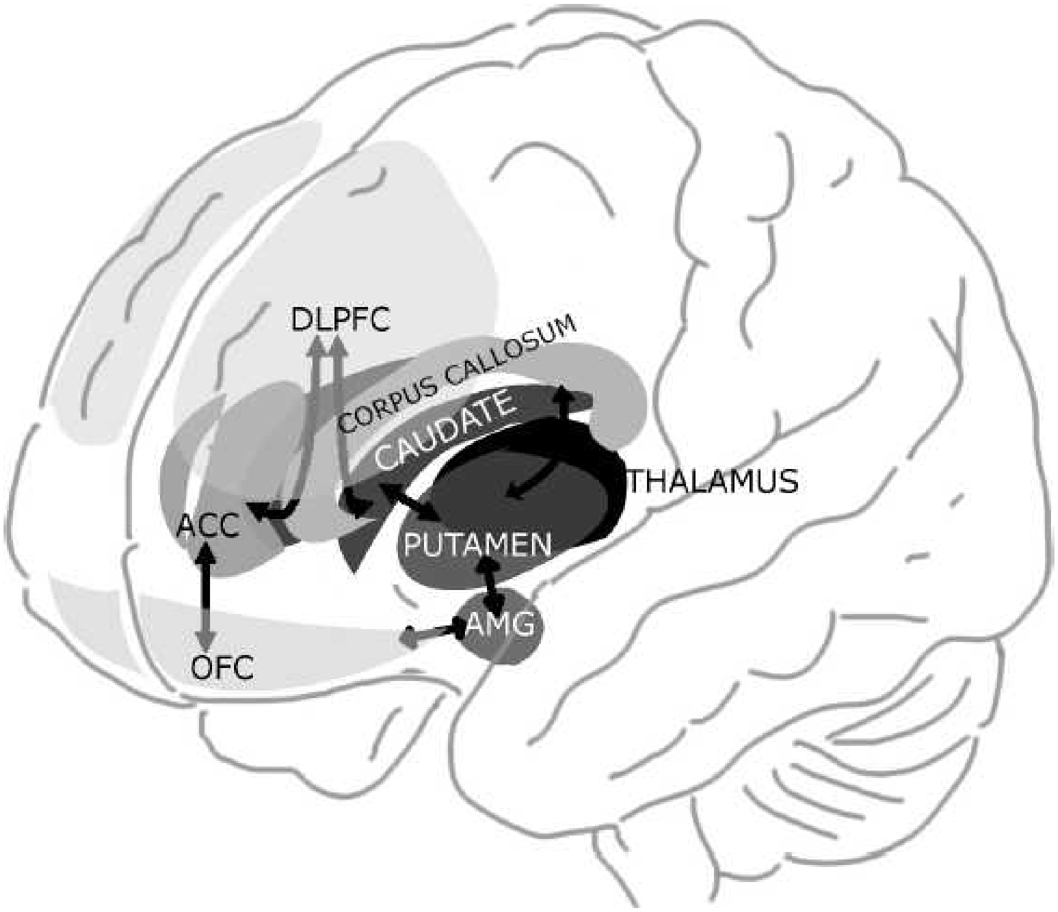

FIGURE 37–12. Brain regions implicated in the pathophysiology of obsessive-compulsive disorder. ACC, anterior cingulate cortex, AMG, amygdala, DLPFC, dorsolateral prefrontal cortex. (Courtesy of Sneha Chenji, Ph.D. Data from Rosenberg DR, MacMillan SN, Moore GJ. Brain anatomy and chemistry may predict treatment response in pediatric obsessive-compulsive disorder. *Int J Neuropsychopharmacol.* 2001;4(2):179–190.)

Prefrontal Cortex. MRI investigation in treatment-naive pediatric patients with OCD as compared to age- and sex-matched healthy controls demonstrated larger anterior cingulate volumes but not DLPFC volume in the OCD patients. These findings are consistent with findings of increased regional cerebral blood flow and metabolic activity in the anterior cingulate cortex in adult and pediatric patients with OCD. Larger anterior cingulate volumes were associated with increased OCD symptom severity but not the duration of illness. Pediatric OCD patients also failed to demonstrate the normal age-related growth in anterior cingulate volume seen in healthy controls (Fig. 37–17). Anterior cingulate volumes were inversely correlated with striatal volumes in the OCD patients. It should be noted that two independent studies noted larger anterior cingulate volumes in pediatric OCD, with the more recent study finding differences in the anterior cingulate gray but not white matter. Lower anterior cingulate Glx was also seen in OCD patients more than in healthy controls. Of interest, the lower anterior cingulate Glx concentration is associated with the G/G genotype of the glutamate receptor gene polymorphism GRIN2B-rs1019385 in OCD patients. A recent fMRI study found greater activation and connectivity of the medial PFC/dorsal anterior cingulate in pediatric OCD patients compared to controls. It was suggested that this might reflect an exaggerated concern for correct performance. Studies of the CC found a larger area and lower signal intensity of the genu of the CC in treatment-naive pediatric OCD patients than in controls. The genu is the region of the CC that connects left and right ventral prefrontal cortical regions, such as the anterior cingulate cortex. This provides further evidence for the involvement of ventral prefrontal cortical regions, such as the anterior cingulate cortex, in the pathogenesis of OCD. In right prefrontal white matter, pediatric OCD patients demonstrated greater NAA and Cho concentrations when compared to healthy control participants. The greater NAA concentration correlated with OCD symptoms as well. Using the manual tracing technique, greater right white matter volume in the orbital PFC has also been noted. Indeed, a recent examination of frontal–striatal–thalamic white matter found greater age-related increases in fractional anisotropy in pediatric OCD patients as compared to controls. In the cingulum bundle, this correlated with OCD symptom severity.

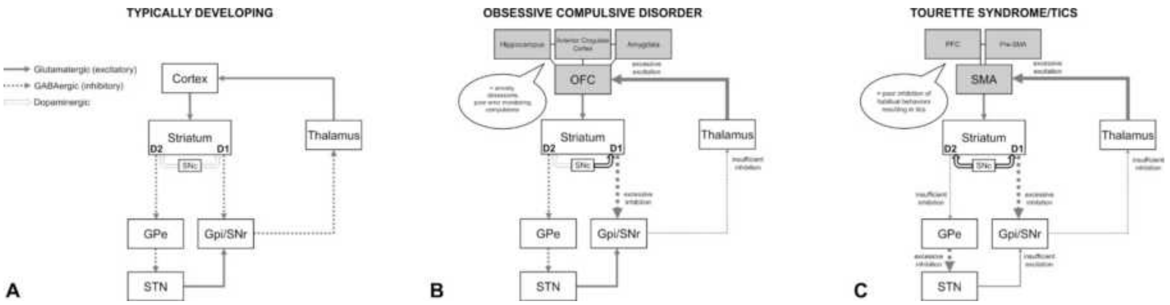
TYPICALLY DEVELOPING
Glutamatergic (excitatory)
GABAergic (inhibitory)
Dopaminergic
Cortex
Striatum
D2
D1
SNc
Thalamus
GPe
Gpi/SNr
STN
A
OBSESSIVE COMPULSIVE DISORDER
Hippocampus
Anterior Cingulate Cortex
Amygdala
OFC
excessive excitation
Striatum
D2
D1
SNc
Thalamus
insufficient inhibition
excessive inhibition
GPe
Gpi/SNr
STN
B
TOURETTE SYNDROME/TICS
PFC
Pre-SMA
+ poor inhibition of habitual behaviors resulting in tics
SMA
excessive excitation
Striatum
D2
D1
SNc
Thalamus
insufficient inhibition
insufficient inhibition
excessive inhibition
GPe
Gpi/SNr
excessive inhibition
insufficient excitation
STN
C

FIGURE 37–13. Schematic diagram showing selected aspects of cortico-striatal pathway loops in **A**: typically developed individuals. **B**: Obsessive compulsive disorder (OCD). **C**: Tourette syndrome (TS). Although some studies have reported normal integrity of striatal dopamine systems, preliminary neuroimaging and postmortem results suggest hyperinnervation of the striatum by dopaminergic neurons, which may contribute to the pathogenesis of tics in TS and compulsions in OCD. Note that in both TS and OCD, there are excessive excitatory projections from the thalamus to cortical areas that mediate the final behavior(s), as noted by the descriptions in the *bubbles* on the left side. These cortical areas are connected to other cortical/subcortical regions involved in the disorders. Excitatory (glutamatergic) projections are indicated by *solid arrows* and inhibitory (GABAergic) projections by *dotted arrows.* Dopaminergic projections are indicated by *outlined arrows*. Thicker arrows indicate increased projections; thinner lines indicate decreased projections. Notably, the details of these pathways across conditions have evolved over time. This figure is intended to provide foundational knowledge on the direct and indirect pathways across typical development and neurodevelopmental conditions. GABA, γ-aminobutyric acid; Gpe, external segment of the globus pallidus; Gpi/SNr, internal segment of the globus pallidus or the substantia nigra pars reticulata; SNc substantia nigra pars compacta; STN, subthalamic nucleus; PFC, prefrontal cortex, Pre-SMA, presupplementary motor area; SMA, supplementary motor area, OFC, orbitofrontal cortex. (A&C, adapted with permission from Peterson BS, Thomas P. Tourette's syndrome: what are we really imaging? In: Ernst M, Rumsey JM. *Functional Neuroimaging in Child Psychiatry*. Cambridge University Press; 2000:242–265. B, data from Lawrence AD, Sahakian BJ, Robbins TW. Cognitive functions and corticostriatal circuits: insights from Huntington's disease. *Trends Cogn Sci*. 1998;2(10):379–388 and Menzies L, Chamberlain SR, Laird AR, Thelen SM, Sahakian BJ, Bullmore ET. Integrating evidence from neuroimaging and neuropsychological studies of obsessive-compulsive disorder: the orbitofronto-striatal model revisited. *Neurosci Biobehav Rev.* 2008;32(3):525–549.)

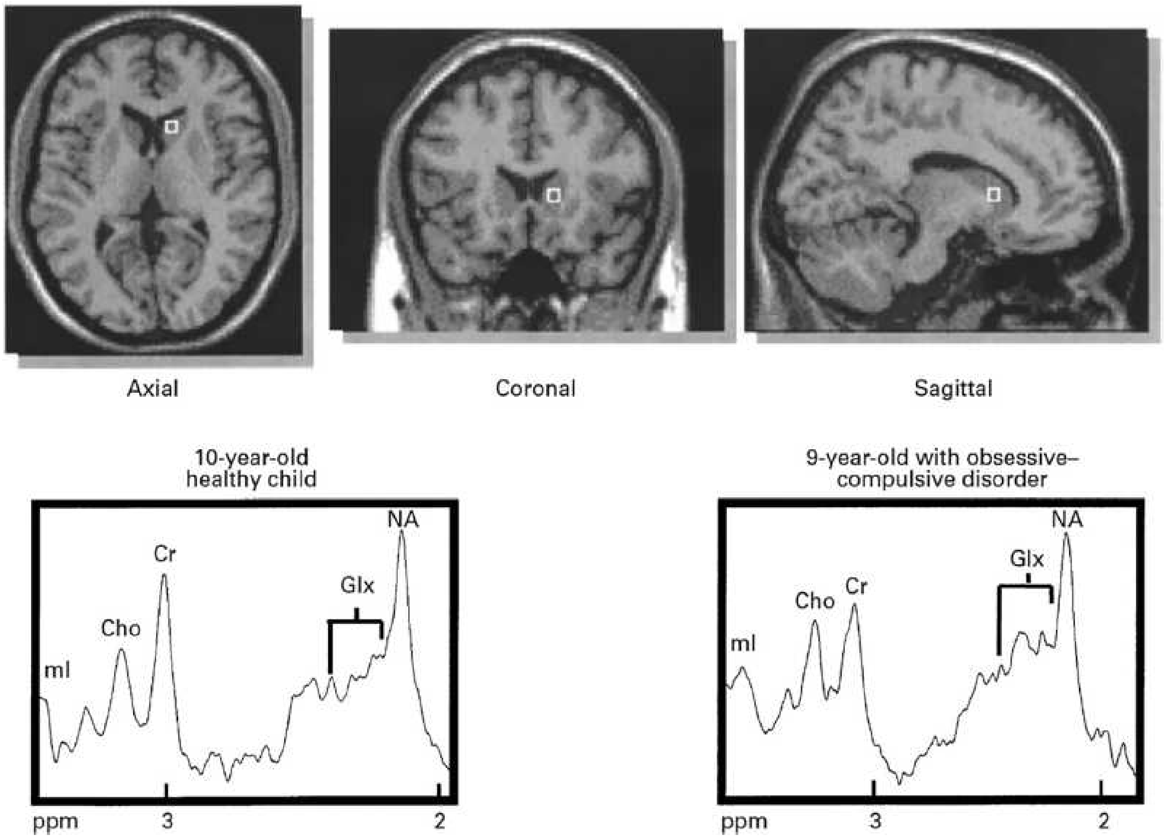
Axial
Coronal
Sagittal
10-year-old healthy child
mI
Cho
Cr
Glx
NA
ppm
3
2
9-year-old with obsessive–compulsive disorder
mI
Cho
Cr
Glx
NA
ppm
3
2

FIGURE 37–14. **Left**: Illustration of voxel placement in left caudate nucleus. **Right**: ^{1}H-Magnetic resonance spectroscopy of a 0.7-mL volume of interest centered in the left caudate in a 10-year-old healthy control and a 9-year-old treatment-naive patient with obsessive-compulsive disorder (OCD) as shown on the T1-weighted magnetic resonance images. Cho, choline; CR, creatine/phosphocreatine; Glx, glutamate/glutamine/γ-aminobutyric acid; mI, myo-inositol; NAA, N-acetyl-aspartate; GABA, γ-aminobutyric acid. (Reprinted with permission from Rosenberg DR, MacMillan SN, Moore GJ. Brain anatomy and chemistry may predict treatment response in paediatric obsessive-compulsive disorder. *Int J Neuropsychopharmacol*. 2001;4(2):179–190. doi: 10.1017/S1461145701002401.)

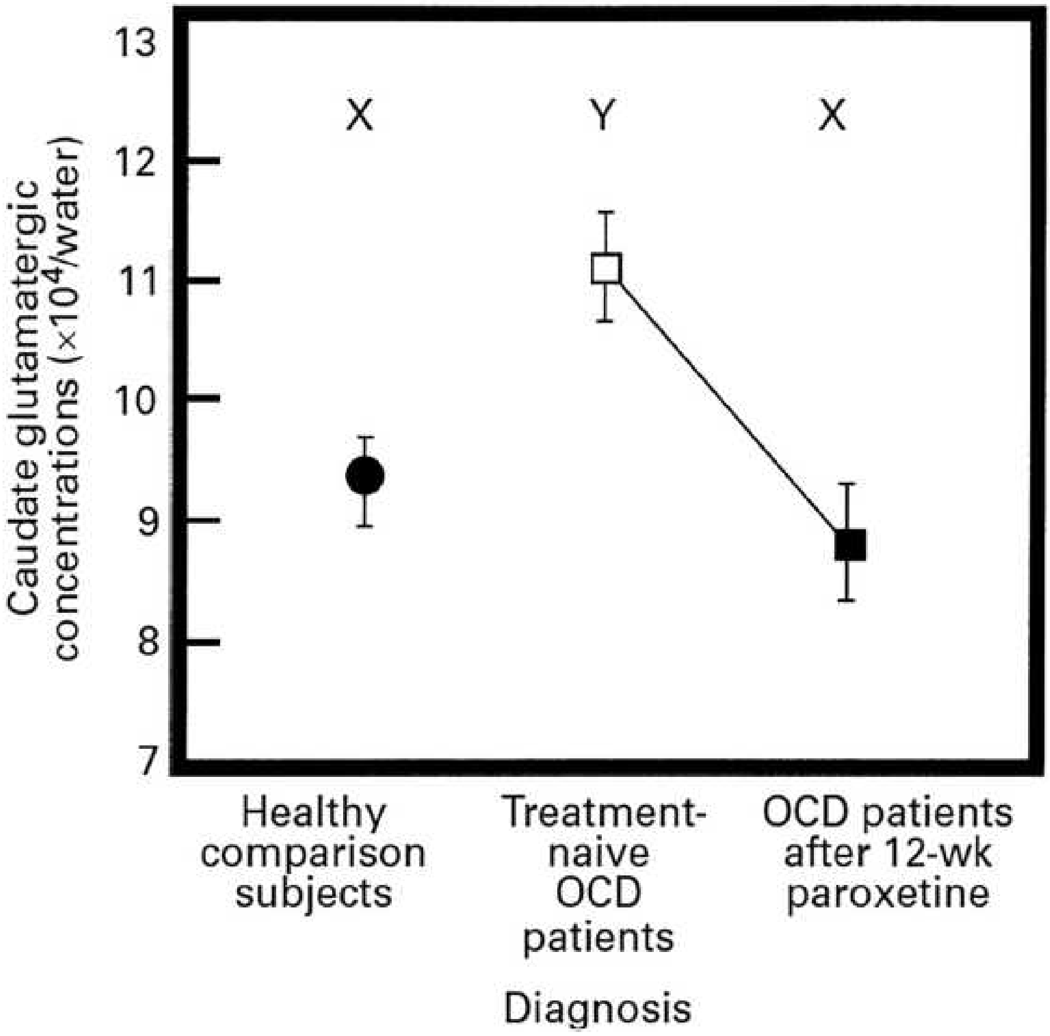

FIGURE 37–15. Caudate glutamatergic concentration by diagnostic and treatment condition. Groups not sharing the same letter are significantly different at $P < .05$. OCD, obsessive-compulsive disorder. (Reprinted with permission from Rosenberg DR, MacMillan SN, Moore GJ. Brain anatomy and chemistry may predict treatment response in paediatric obsessive-compulsive disorder. *Int J Neuropsychopharmacol.* 2001;4(2):179–190. doi:10.1017/S1461145701002401.)

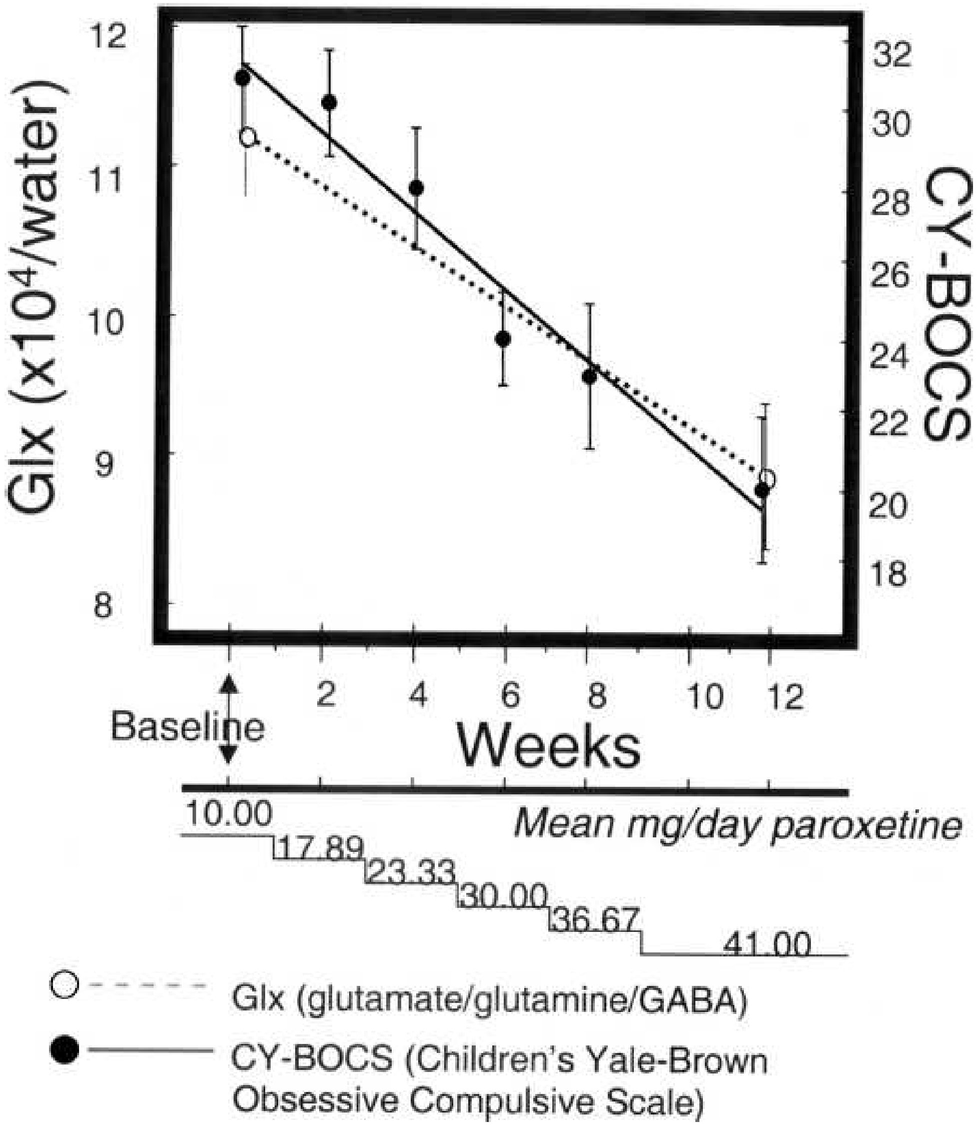

FIGURE 37–16. Decrease in left caudate glutamatergic concentrations associated with reduction in obsessive-compulsive score of the Children's Yale-Brown Obsessive Compulsive Scale. (Reprinted with permission from Rosenberg DR, MacMillan SN, Moore GJ. Brain anatomy and chemistry may predict treatment response in paediatric obsessive-compulsive disorder. *Int J Neuropsychopharmacol*. 2001;4(2):179–190. doi:10.1017/S1461145701002401.)

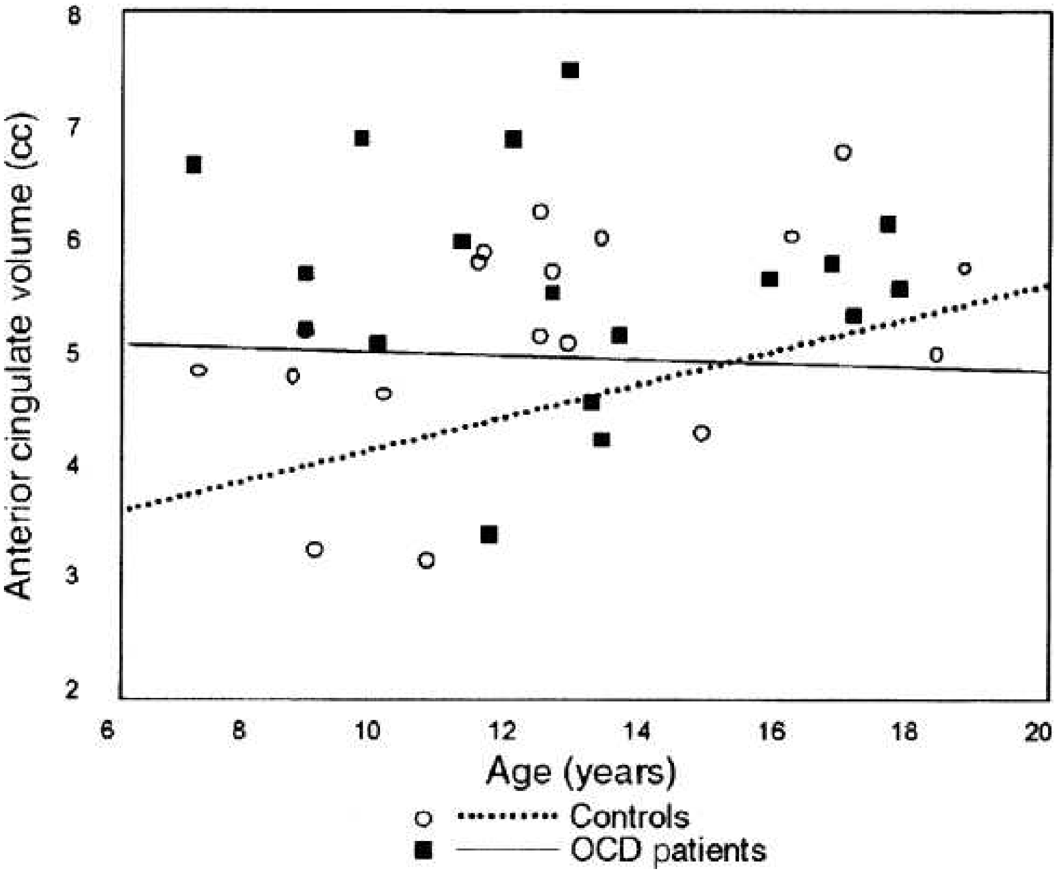

FIGURE 37–17. Anterior cingulate volume as a function of age for pediatric patients with obsessive-compulsive disorder (OCD) and normal controls. Opposite neurodevelopmental trajectories in the patients as compared with the healthy subjects may reflect an absence of early neuronal pruning in children with OCD. (Reprinted with permission from Rosenberg DR, MacMillan SN, Moore GJ. Brain anatomy and chemistry may predict treatment response in paediatric obsessive-compulsive disorder. *Int J Neuropsychopharmacol*. 2001;4(2):179–190. doi:10.1017/S1461145701002401.)

Thalamus. The thalamus acts as the relay station for neuronal input to the cortex and plays a critical role in the circuitry implicated in imaging studies of OCD. In an MRI study of treatment-naive pediatric patients with OCD as compared to age- and sex-matched healthy controls, OCD patients were found to have larger thalamic volumes. After therapy with the SSRI paroxetine, thalamic volumes decreased in size to the point at which they no longer differed significantly from those of controls (Fig. 37–18). This reduction in thalamic volume was correlated robustly with a reduction in OCD symptom severity (Fig. 37–19). Indeed, a larger pretreatment thalamic volume predicted a better response to paroxetine treatment. When subdividing the thalamus into its component nuclei, the changes in volume were most prominent in medial rather than lateral thalamic regions. With CBT, no significant changes in thalamic volume were observed before and after 12 weeks of treatment in treatment-naive pediatric OCD patients. This suggests that the reduction in thalamic volume may be specific to SSRI therapy as opposed to a more general treatment effect. A large-scale study with 2,551 school-aged children showed that children with probable OCD also have larger thalami than controls, suggesting these differences in thalamic volume may inform longitudinal analyses. In a closer examination of the thalamus in OCD, a ^{1}H-MRS study of treatment-naive pediatric OCD patients and age- and sex-matched pediatric controls was conducted. Lower levels of the neuronal marker NAA were found in OCD patients than in healthy controls. After 12 weeks of drug therapy with paroxetine, an increase in NAA in the medial but not lateral thalamus was seen. Thus, as enlarged thalamic volumes in OCD patients appear to be reduced toward normal after SSRI treatment, lower levels of NAA may be conversely increasing. This suggests a potential neurotrophic effect of SSRI treatment. In addition, significantly greater left and right medial thalamic Cho concentrations were observed in OCD patients than in both healthy control subjects and patients with MDD. Cho concentrations did not differ between patients with MDD and control subjects. This suggests that alterations in medial thalamic Cho differentiate patients with OCD from healthy control subjects and patients with MDD.

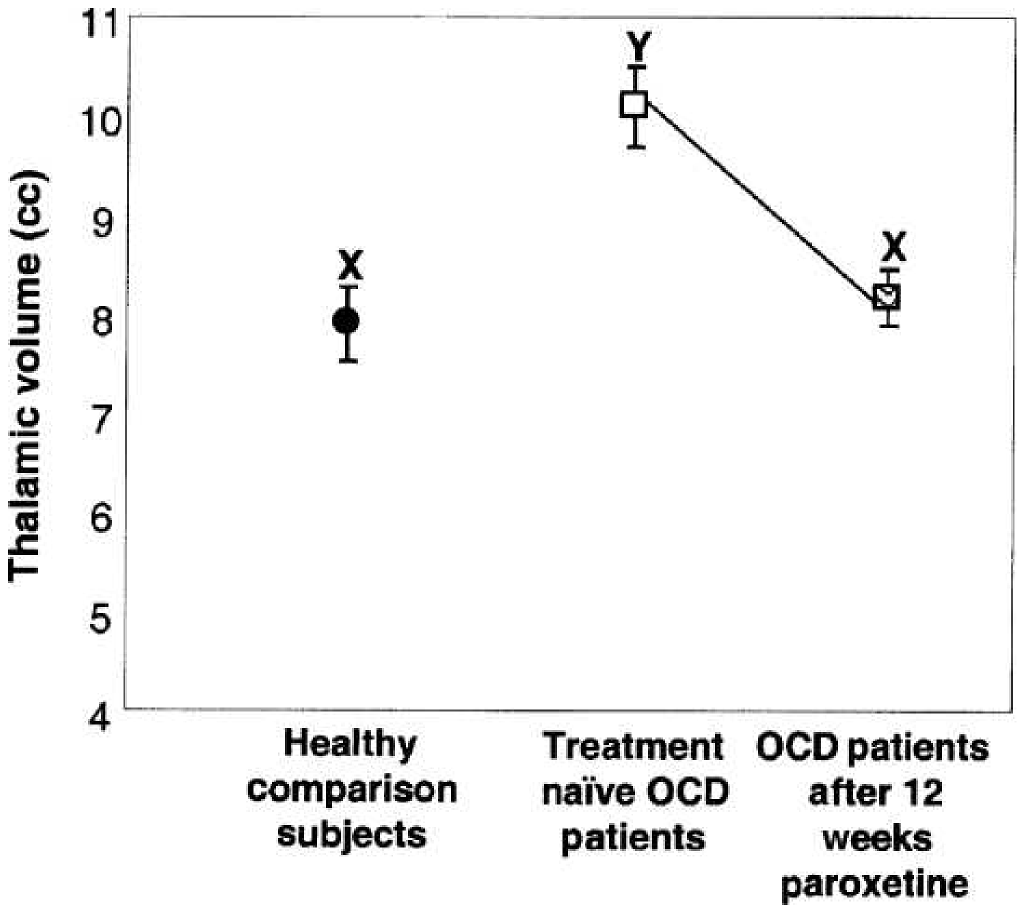

FIGURE 37–18. Data (presented as mean ± standard deviation) for three groups: Healthy controls, pretreatment obsessive-compulsive (OCD) patients, and post-treatment OCD patients. (Reprinted with permission from Rosenberg DR, MacMillan SN, Moore GJ. Brain anatomy and chemistry may predict treatment response in paediatric obsessive-compulsive disorder. *Int J Neuropsychopharmacol*. 2001;4(2):179–190. doi: 10.1017/S1461145701002401.)

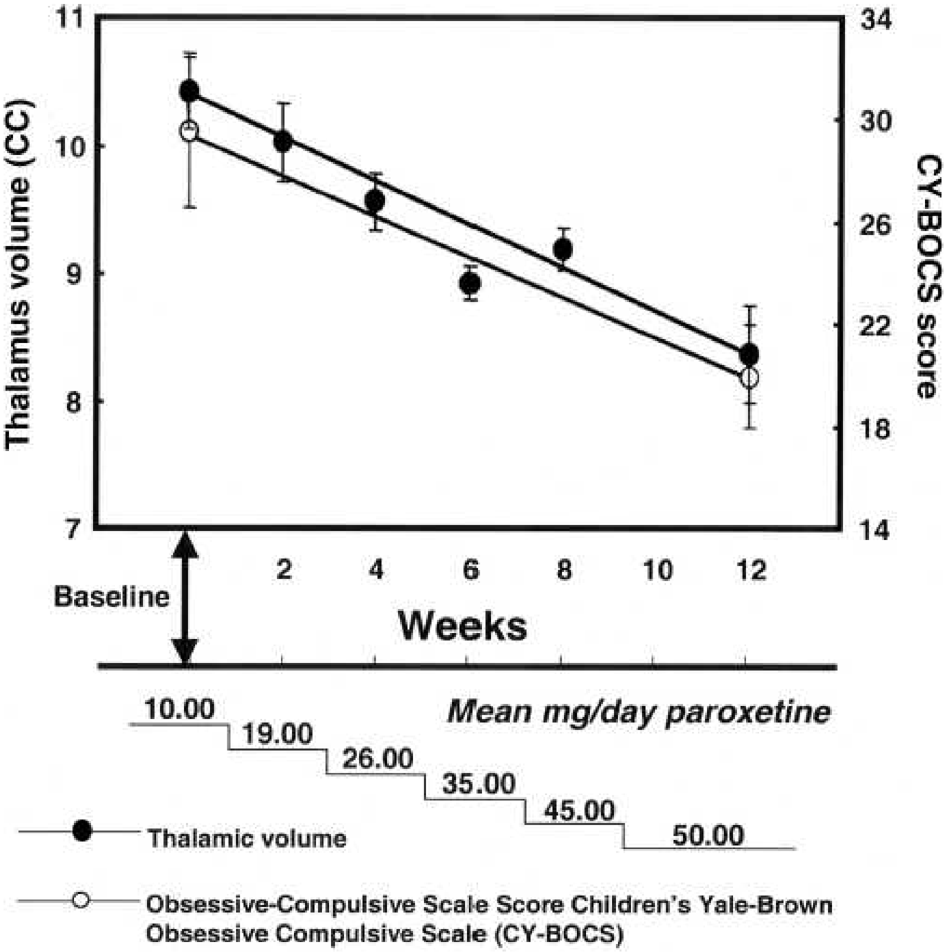

FIGURE 37–19. Decrease in thalamic volume associated with reduction in obsessive-compulsive score of the Children's Yale-Brown Obsessive-Compulsive Scale (CY-BOCS). (Reprinted with permission from Rosenberg DR, MacMillan SN, Moore GJ. Brain anatomy and chemistry may predict treatment response in paediatric obsessive-compulsive disorder. *Int J Neuropsychopharmacol*. 2001;4(2):179–190. doi:10.1017/S1461145701002401.)

Amygdala. Given its role in fear conditioning, the amygdala is believed to be involved in the pathophysiology of OCD. The amygdala is also a target of SSRIs, the mainstay of pharmacotherapy for OCD. In a recent study, amygdala volumes were evaluated for in vivo volumetric changes in 11 pediatric patients with OCD before and after 16 weeks of monotherapy with the SSRI paroxetine. Eleven healthy pediatric control subjects had baseline and follow-up scans, but none of these subjects received medication. OCD patients at baseline demonstrated significant asymmetry of the amygdala (left greater than right) as compared to healthy controls. The volume of the left amygdala decreased significantly following treatment in the OCD patients. This reduction in the volume of the left amygdala in patients correlated significantly with higher paroxetine dose at the time of the follow-up scan and total cumulative paroxetine exposure between the scans. No changes in right or left amygdala volume were noted in healthy controls over time. There are several possible mechanisms that may account for the paroxetine-mediated change in amygdala volume. Disease progression cannot be entirely ruled out, but this is unlikely, given the degree of change noted and the short time in which these changes occurred. The significant inverse relationship between left amygdala reductions and paroxetine treatment lends credence to the idea that the reversal of hypertrophy is due to treatment. Functionally, in most anxiety disorders, an overactive amygdala is often noted. In pediatric OCD, however, amygdala activity is reduced in response to emotional stimuli. Caution is noted due to the small sample sizes involved. Further research is warranted.

Pediatric Autoimmune Neuropsychiatric Disorders Associated With Group A β-Hemolytic Streptococcal Infection

While a smaller basal ganglion appears to characterize patients with early-onset OCD, larger basal ganglia occur in the subset of patients with pediatric autoimmune neuropsychiatric disorders associated with group A β-hemolytic streptococcal (GABHS) infection (PANDAS). In PANDAS, it is hypothesized that GABHS-associated antibodies cross-react with neurons in the basal ganglia. This causes the observed symptoms of OCD, as well as tics/TS and ADHD in susceptible children. This hypothesis is supported by evidence indicating larger basal ganglia volumes in some patients with OCD, tics, and/or ADHD that are positively correlated with the incidence of GABHS infection and the presence of antibody titers. Furthermore, immunotherapy interventions such as plasmapheresis ameliorate OCD symptoms in some PANDAS patients with marked OCD symptoms. Recovery is also associated with a reduction in basal ganglia volumes. Of interest, thalamic volumes do not differ between PANDAS patients and controls. Although additional studies are needed, this suggests possible unique neurodevelopmental subtypes of OCD.

Anxiety Disorders

Compared with OCD, there have been relatively few neuroimaging studies of pediatric anxiety disorders.

Generalized Anxiety Disorder

In children with GAD, regions associated with fear learning and extinction, mentalization, and amygdala regulation (Fig. 37–20) exhibited greater cortical thickness than controls (Fig. 37–21). These areas included the right inferior frontal gyrus and the left inferior and middle temporal gyri. Thinner parietal and occipital areas have been associated with poorer treatment response. Moreover, smaller hippocampal volumes have been shown to correlate with increased anxiety symptom severity. Furthermore, amygdala intrinsic functional connectivity is disrupted in children with GAD, affecting the amygdala’s connections with the medial PFC, insula, and cerebellum. Functionally, greater left medial PFC and right ventrolateral PFC activation was observed in adolescents with GAD in response to a continuous performance task with emotional distractors. Disrupted connectivity arising from the amygdala was also seen. As discussed earlier (see depressive and bipolar and related disorders), the amygdala has extensive interconnections with cortical–limbic circuitry and plays a significant role in the processing of anxiety or fear. Indeed, fMRI studies in a combined sample of children with GAD and/or panic disorder compared to healthy children found exaggerated amygdala activation in response to fearful faces in patients with GAD. Furthermore, amygdala volume was significantly larger in pediatric patients with GAD than in healthy controls. In response to treatment with medication or CBT, right ventrolateral PFC activation related to emotional faces increased. In addition to thicker cortices and larger amygdala volumes, subcortical regions are also affected by GAD. The right putamen was also found to be larger in children with GAD. Interestingly, maltreatment during childhood was associated with larger left thalamic volumes in children with GAD as well. Both the putamen and thalamus have been implicated in other anxiety disorders, but their precise role has yet to be delineated.

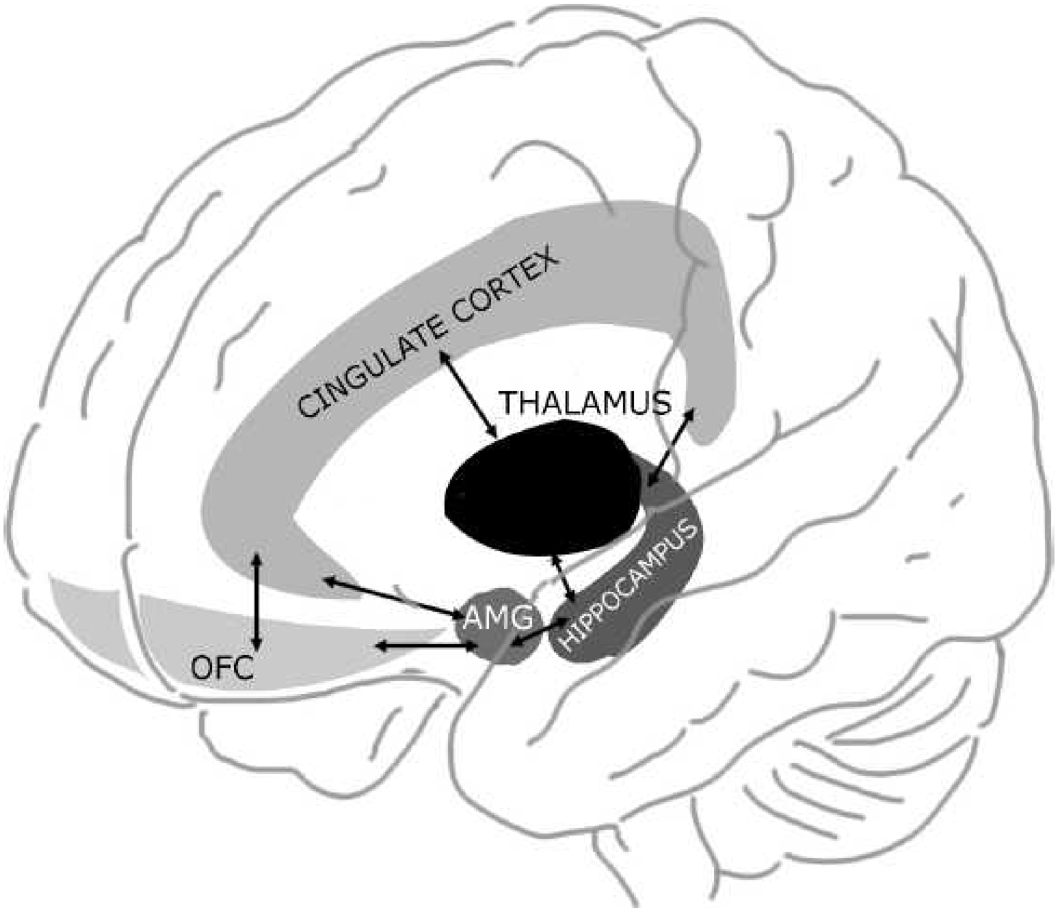

FIGURE 37–20. The thalamus, cingulate, amygdala (AMG), hippocampus, and the orbital frontal cortex (OFC) are activated in the emotion of fear. (Courtesy Sneha Chenji, Ph.D.)

Post-traumatic Stress Disorder

The amygdala has been implicated in some studies of adult patients with PTSD, and recently hypertrophy of the amygdala in children with PTSD has been observed. In addition, smaller hippocampal volumes have been consistently found in adults with chronic PTSD but not in children with PTSD. In children with PTSD, smaller whole-brain, prefrontal cortical, and right temporal lobe volumes are associated with an earlier age of onset and longer duration of abuse. More recently, a ^{1}H-MRS study demonstrated a lower concentration of the neuronal marker NAA in the anterior cingulate cortex of maltreated children with PTSD than in that of controls. Consistent with that finding, a lower concentration of gray matter has been found in youth with PTSD and increased activation in response to a threat in the anterior cingulate cortex. Given the mixed evidence, discrete neurobiologic profiles may be present in pediatric versus adult patients with PTSD. Longitudinal studies to characterize the developmental neurobiology of the illness are needed. Careful assessment for comorbid conditions—for example, anxiety and depression—are also necessary to discern unique neurobiologic pathways among different anxiety disorders and depression.

Tourette Syndrome

TS is often comorbid with ADHD and OCD. Indeed, certain brain circuits are common in the pathogenesis of all three conditions.

Cortex. A recent large-scale volumetric MRI study of children and adults with TS found striking abnormalities in cortical development. In the parieto-occipital region, a greater volume was observed in children than in adults with TS. Age-associated reduction in parieto-occipital volumes occurred predominantly in females. In contrast, relatively larger parieto-occipital volumes were observed in older men with TS, suggesting that sexually dimorphic neurodevelopment may contribute to the greater prevalence of TS in males than in females. Larger DLPFC volumes were found in the younger TS patients but tended to decrease toward adulthood. Because tics often diminish with age, smaller dorsolateral prefrontal cortical volumes in older TS patients were hypothesized to reflect the relatively unusual persistence of tics in the adults participating in the study.

fMRI studies demonstrating increased activation of the PFC during tic suppression in adult TS patients suggest that larger prefrontal cortical volumes observed in children with TS may represent healthy structural compensation to allow for tic suppression. Functional neuroimaging studies in healthy children and adults have implicated the orbital PFC in the processing of inhibitory control, such that the inverse correlation between orbital prefrontal volumes and tic severity in children and adults with TS likely reflects a fundamental deficit in inhibitory control allowing for the “release” of tics. In the motor and premotor cortices, increased volumes and functional activity appear to occur in adults but not children with TS, possibly reflecting the cumulative effect of years of repetitive, abnormal movements. More recently, it was noted that the normative developmental correlates of activity in fronto-striatal circuits that subserve self-regulatory control are disturbed in persons with TS.

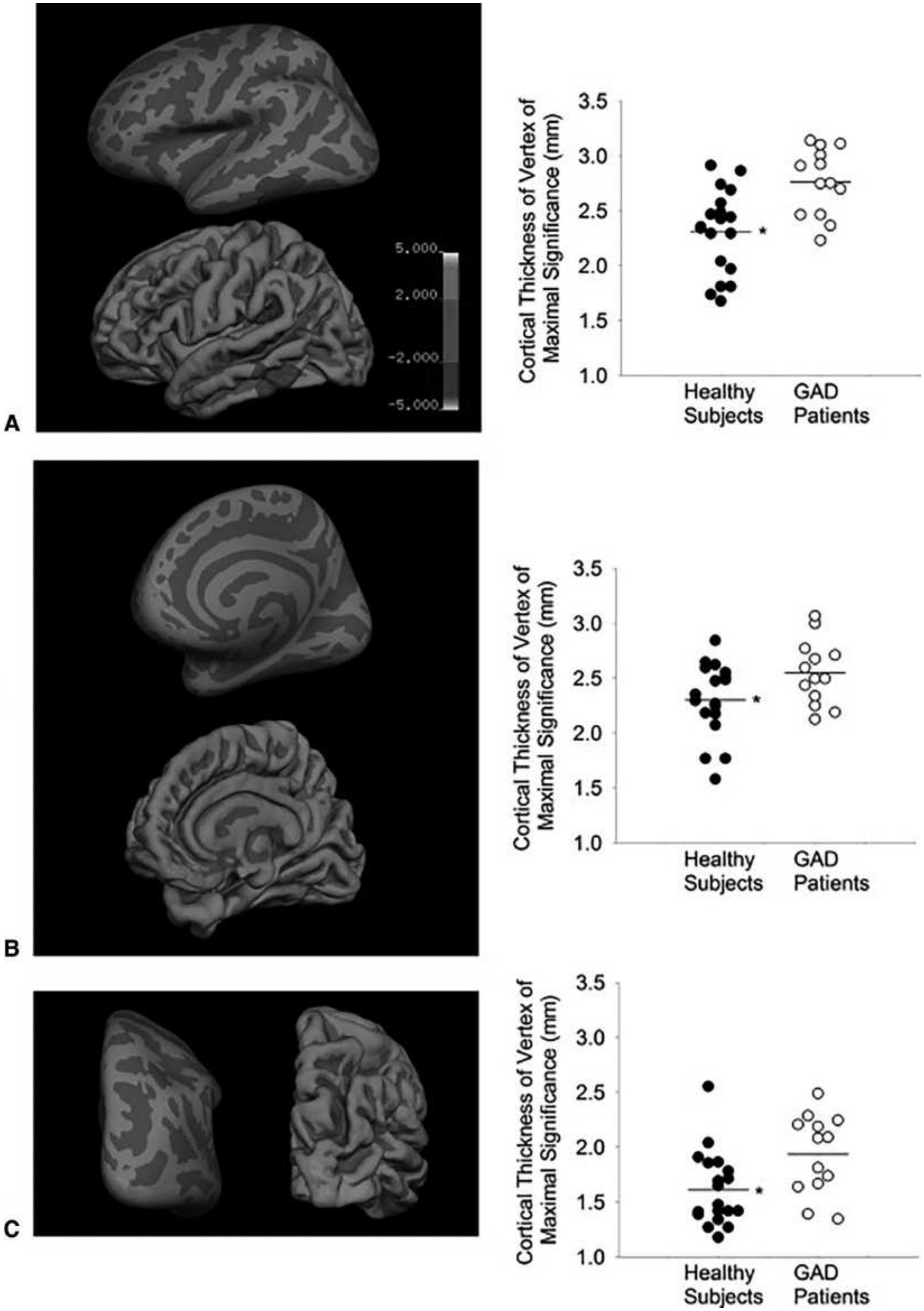

FIGURE 37–21. Cortical thickness in adolescents with generalized anxiety disorder (GAD) as compared to healthy controls. Warm colors (*red to yellow*) indicate greater cortical thickness in adolescents with GAD compared to

healthy comparison subjects, and for each region, both cortical and inflated views are presented. Scatter plots are shown for the maximally significant vertex within each depicted region and healthy patients are represented by (•) while adolescents with GAD are represented by (°). The left temporal findings are shown in panel **A**, while the prefrontal differences and occipital findings are shown in panels **B** and **C**, respectively. Finally, an *asterisk* denotes a statistically significant difference between groups (*P* = .006 for the left temporal differences; *P* = .0004 for the frontal differences, and *P* = .0001 for the occipital differences). (See eBook for color image.) (Reprinted from Strawn JR, John Wegman C, Dominick KC, Swartz MS, Wehry AM, Patino LR, Strakowski SM, Adler CM, Eliassen JC, DelBello MP. Cortical surface anatomy in pediatric patients with generalized anxiety disorder. *J Anxiety Disord.* 2014;28(7):717–723. doi: 10.1016/j.janxdis.2014.07.012.)

Basal Ganglia. Reduced basal ganglia volumes have been observed in both children and adults with TS. Volume reduction has been found to be particularly prominent in the globus pallidus and putamen in patients with TS, implicating motor pathways within cortical–subcortical circuitry. This is consistent with the motor phenomena of tics. In support of the hypothesized role of sex in the neurobiology of TS, abnormal basal ganglia asymmetry has been demonstrated in boys but not girls with TS. Although volumetric neuroimaging studies of identical twins with TS have generated conflicting findings, one twin study found a greater reduction of the caudate in the more severely affected twin, suggesting that the neurobiologic underpinnings of TS expression may be influenced by environmental exposure, as well as by sex and genetic predisposition. Enlargement of thalamic volumes has also been noted in children and adults with TS as compared to healthy comparison subjects. Increased apparent diffusion coefficient (ADC) in cortico-striatal projection pathways, including the left external capsule and left and right subcallosal fasciculus pathway, have also been noted in children with TS. Based on the efficacy of dopamine receptor blockade in the treatment of TS, one recent PET study of children with TS found evidence for excessive dopamine decarboxylase activity in the left caudate and right midbrain, suggesting subcortical dysfunction of the dopamine system in pediatric patients with TS (see Fig. 37–12).

Corpus Callosum. Conflicting results have been observed with regard to the CC in TS in youth, with reports of larger, smaller, and no difference in the area. A study of monozygotic twins discordant for TS using both MRI and DTI methods failed to observe differences with regard to fiber-tracking to the CC. Fractional anisotropy was lower in the posterior aspects of the CC in the affected twin, however, indicating a less-coherent level of organization.

Other Brain Regions. A recent structural MRI investigation evaluated the association between the size of the cavum septi pellucidi (CSP) and a diagnosis of TS. They found a smaller CSP in children with TS than in healthy pediatric controls. These findings were replicated in an independent sample of adult patients with TS as compared to healthy adults. An inverse association between CSP size and ADHD symptom severity was also noted. In addition, the overall volume of the hippocampus and amygdala was found to be significantly larger in the TS group than in controls. A surface analysis found that the larger volumes in the TS group were derived primarily from the head and medial surface of the hippocampus and the dorsal and ventral surfaces of the amygdala. The volumes of these subregions declined with age in the TS group but not in the controls. Hence, the subregions were significantly larger in children with TS but significantly smaller in adults with TS than in controls. Clinically, volumes in these subregions correlated inversely with the severity of tic, OCD, and ADHD symptoms in patients. This suggests that the enlargement of the subregions may have a compensatory effect on tic-related symptoms.

Childhood-Onset Schizophrenia

Although it is a rare condition, COS has recently been the focus of considerable research investigation. It is believed that an enhanced understanding of the developmental brain changes in early-onset schizophrenia might lead to an improved understanding of the neurobiology of adult schizophrenia. Indeed, using brain region data as input, machine-learning methods can correctly classify COS and healthy controls with 73.7% accuracy.

Total Brain Volume, White Matter, and Ventricles. Aside from disrupted hypothalamic–pituitary–adrenal (HPA) axis function in mood disorders, one of the most consistent findings in psychiatry is a reduction in total brain volume and an increase in the volume of the ventricles in adult patients with schizophrenia. In COS, the smaller total brain volume is even more dramatic than in adult patients. This suggests that COS may be a rare, more severe version of the disease. Smaller total brain volume in COS is accompanied by ventriculomegaly (i.e., enlarged ventricles). The smaller amount of brain tissue in early-onset schizophrenia may result from a progressive loss of both gray and white matter in periventricular regions, including the posterior cingulate, caudate, and thalamus (Fig. 37–22).

In young patients with COS, slower growth rates were observed in parietal white matter. However, the CC did not demonstrate any differences from controls with regard to volume and growth trajectory. Furthermore, patients with COS have demonstrated reduction in fractional anisotropy in the cuneus when compared to healthy controls. This reduction was also observed in their healthy siblings, possibly indicating some shared genetic risk. This indicates regional effects on white matter with COS.

PFC, Insula, and Temporal Lobe. Akin to the aforementioned finding in the cuneus, during a weather prediction fMRI task, siblings of COS patients demonstrated less activation of the frontal and striatal regions compared to controls. This indicates that impairments in cortical–striatal function may play a role in the liability of schizophrenia. It has been suggested that changes in frontal deficits are modulated in part by polymorphism of the *catechol-o-methyltransferase* (*COMT*) gene. Lower levels of the ^{1}H-MRS neuronal marker NAA and gray matter volumes in the frontal and anterior cingulate cortex are found in patients with COS and unaffected offspring of patients with schizophrenia at increased risk for developing schizophrenia as compared to controls. This suggests that an exaggerated pruning of neurons may be involved in the pathogenesis of schizophrenia. Controlling for medication exposure in neuroimaging studies is critical because typical neuroleptics have been shown to decrease frontal cortical activation, whereas atypical agents (e.g., risperidone) increase frontal activation. The insula plays a critical role in delineating self from nonself-stimuli. Given this function, it has been implicated in the positive symptoms associated with schizophrenia. In 98 patients with COS contrasted with 100 healthy control participants, smaller right and left insula volumes were noted in patients. Right insula volume correlated inversely with positive symptoms in patients with COS.

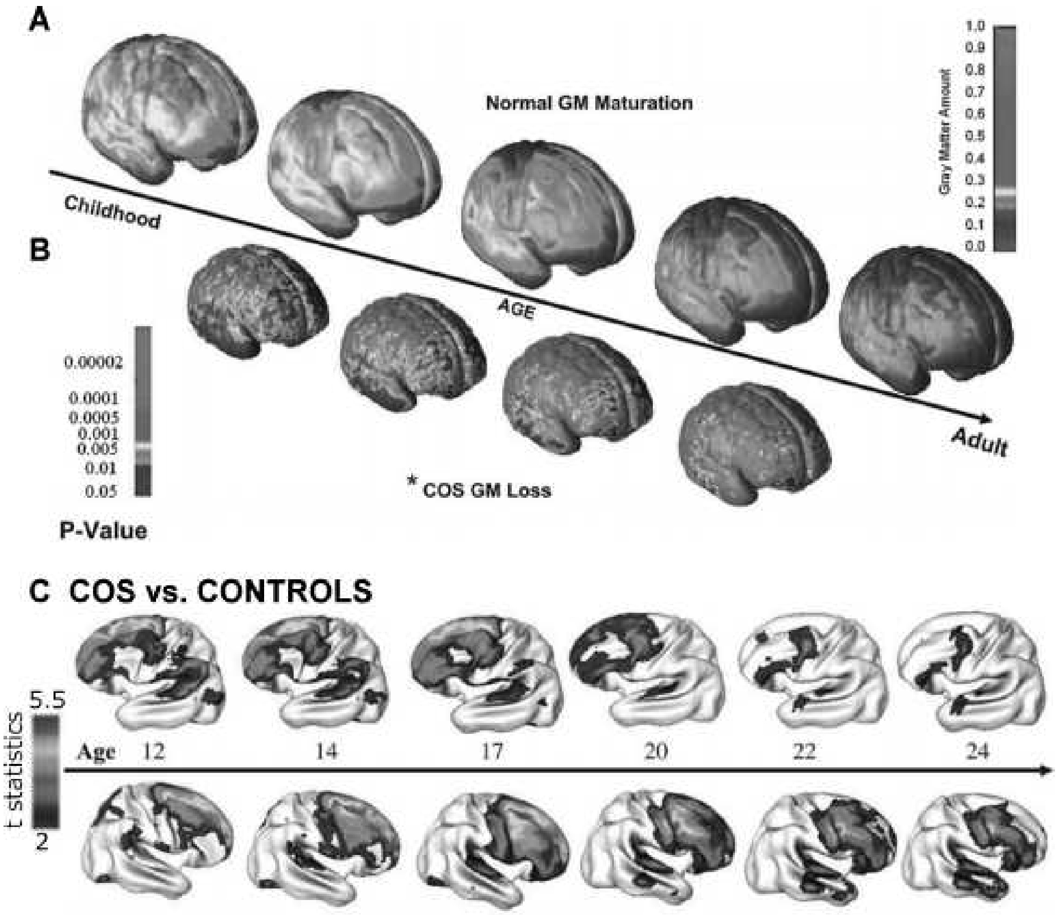
A
Normal GM Maturation
Gray Matter Amount
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Childhood
AGE
Adult
B
0.00002
0.0001
0.0005
0.001
0.005
0.01
0.05
P-Value
* COS GM Loss
C COS vs. CONTROLS
5.5
t statistics
2
Age 12
14
17
20
22
24

FIGURE 37–22. **A:** Right lateral view of normal gray matter (GM) maturation in healthy youth (4–22 years) scanned longitudinally every 2 years. The amount of grey matter change is indicated by the color bar (**upper right** panel; red to pink indicating more GM, blue indicating GM loss). It represents parietotemporal (i.e., back-to-front) progress of cortical maturation. **B:** Right lateral view of cortical maturation in childhood-onset schizophrenia (COS) compared to controls. The color bar (**lower left** panel) indicates *P*-values for reductions in GM amount in COS and suggests global brain changes in COS compared to controls. Cortical maturation in COS appears to follow an exaggeration of normal GM maturation. *COS and healthy controls were aged 16–21 years. **C:** Progression of GM loss in COS youth (12–24 years) compared to healthy controls indicating areas of statistically significant thinning over an average of 3-year intervals (corrected for false discovery rate of *q* = 0.05). (A–C, reprinted from Gogtay N. Cortical brain development in schizophrenia: insights from neuroimaging studies in childhood-onset schizophrenia. *Schizophr Bull.* 2008;34(1):30–36; A, originally adapted from Gogtay N, Giedd JN, Lusk L, et al. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci USA.* 2004;101(21):8174–8179. B, originally adapted from Thompson PM, Vidal C, Giedd JN, et al. Mapping adolescent brain change reveals dynamic wave of accelerated gray matter loss in very early-onset schizophrenia. *Proc*

Natl Acad Sci USA. 2001;98(20):11650–11655. C, originally adapted from Greenstein D, Lerch J, Shaw P, et al. Childhood onset schizophrenia: cortical brain abnormalities as young adults. *J Child Psychol Psychiatry.* 2006;47:1003–1012.)

Consistent with studies of adult patients with schizophrenia, a smaller volume of the temporal lobe has been reported in COS, although a contradictory report exists. Within the temporal lobe, smaller superior temporal gyrus volumes associated with increased symptom severity have been observed in patients with COS as compared to healthy controls. In adult patients with first-episode schizophrenia, volume reduction in the superior temporal gyrus was correlated with the duration of pretreatment illness. Superior temporal gyrus volumes appear to increase toward normal after pharmacotherapy, suggesting that this region may be integral to symptom expression in schizophrenia. In a recent report, it was noted that patients with COS showed significant bilateral deficits (9% to 10%) in total hippocampal volume. This smaller volume in such patients remained consistent from age 9 to 26 years. Of interest, subregional maps showed heterogeneous changes, with a loss of hippocampal volume in both anterior and posterior ends, whereas the body of the hippocampus gained in volume. This suggests that the subunits of the hippocampus are differentially affected in schizophrenia. Indeed, a reduction in the volume of the amygdala–hippocampal formation has also been demonstrated in adults with first-episode schizophrenia, in patients with COS, and in offspring and siblings of patients with schizophrenia. The reduction in amygdala–hippocampal formation in patients with COS and high-risk unaffected relatives suggests that it might play a critical role in the pathogenesis of schizophrenia. Earlier onset of illness in patients with traumatic perinatal histories has also been shown to be associated with a more severe reduction in amygdala–hippocampal volume. Shape analysis revealed a striking inward deformation in the anterior hippocampus in patients with COS. These inward deformations were associated with positive symptoms. More recently, it has been noted that the younger, healthy siblings of patients with COS showed significant gray matter deficits in the left prefrontal and bilateral temporal cortices and smaller deficits in the right prefrontal and inferior parietal cortices compared with controls. It may be that a loss in prefrontal and temporal gray matter in COS is a familial/trait marker of the illness.

Subcortical Regions. Increased caudate volume has been reported in COS, although this might be due to typical neuroleptic exposure. When typical neuroleptics were replaced with clozapine in patients with COS, caudate size decreased to normal within 2 years. However, caudate volume reduction has been reported in neuroleptic-naive adult patients with schizophrenia, so that a primary alteration in caudate volume cannot be ruled out. Recently, a shape analysis of the caudate in COS found an outward displacement in the dorsal aspect of the caudate body. One of the most consistent findings in psychotropic-naive young adults with schizophrenia and COS is decreased volume of the thalamus. Reduced thalamic volumes have also been found in the primary relatives of patients with schizophrenia.

Substance-Related and Addictive Disorders

Alcohol Use Disorder. Given the high prevalence of conditions such as binge drinking among adolescents, the need for research in this area is great. Consistent with adult studies, a neuroimaging study of alcohol use in adolescents found a significant volume reduction in bilateral hippocampi as compared to healthy pediatric controls. Hippocampal volume reduction correlated positively with age of onset and negatively with duration of alcohol use, suggesting that alcohol may interfere with normal growth of the hippocampus. Smaller prefrontal gray and white volumes have been noted in adolescents who abuse alcohol. In youth at high risk for alcoholism, a smaller amygdala has been consistently noted. This is suggestive of a possible abnormality in the processing of fearful stimuli. An fMRI connectivity study found differences in resting state synchrony between the left nucleus accumbens and bilateral inferior frontal gyri and the left postcentral gyrus in youth at risk as compared to healthy controls. The results suggested that youth at risk had less segregation between the nucleus accumbens and frontal executive functions and less integration with reward-related areas like the orbital frontal cortex. Aligned with this, another study with 150 adolescents showed that Alcohol Use Disorder symptomology was associated with activation in other brain areas mediating reward and punishment (i.e., striatum, anterior cingulate cortex). ^{31}P-MRS has been used to probe for metabolic abnormalities of frontal, occipital, and parietal cortex. In children with both a positive family history of alcoholism and a disruptive behavior disorder, decreased PDE levels were observed, suggesting abnormalities in energy metabolism.

These abnormalities were also associated with learning disability. In a longitudinal morphometric study, 40 adolescents were followed over 3 years. During that period, half initiated heavy drinking. At baseline, subjects who later started drinking heavily had smaller left cingulate, pars triangularis, and rostral anterior cingulate volume as compared to those who did not. Over time, reductions in left inferior and middle temporal gyri, left caudate, and brain stem were noted as compared to substance-naive youth. This study suggests that there are preexisting markers in future drinkers and a further volume reduction over time after alcohol abuse was initiated.

Cannabis Use Disorder. Cannabis-using adolescents demonstrate thinner left and right supramarginal, left and right inferior parietal, right pars triangularis, left pars opercularis, left superior frontal, and left superior temporal cortical thickness compared to noncannabis-using subjects. Greater cannabis use was also associated with thicker bilateral superior frontal gyri. In adolescents who are regular cannabis smokers, gray matter volume reduction was observed in the medial temporal cortex, temporal pole, parahippocampal gyrus, insula, and orbitofrontal cortex. These are brain regions rich in cannabinoid CB1 receptors and associated with motivational, emotional, and affective processing. In 17 chronic cannabis-using adolescents as compared to 17 healthy controls, reduction in anterior cingulate NAA, Cr, inositol, and glutamate were observed. Akin to what was observed in childhood-onset schizophrenia, changes in prefrontal and insula cortical thickness were observed in adolescent cannabis users. In a resting state connectivity study, heavy cannabis use in adolescents was also associated with middle frontal gyri activity and inversely with the middle temporal gyrus. During an fMRI monetary incentive delay task, adolescent males who use cannabis showed striatal hyperactivity in anticipation of reward. A systematic review of diffusion-weighted imaging studies of cannabis use during adolescence found abnormal white matter microstructure in neocortical association pathways as well as in projection and thalamic pathways. One of the primary problems with studies of cannabis use in adolescents is the limited ability to quantify and control for exposure; with most studies relying on self-report of use with little data on the strength of cannabis used.

Feeding and Eating Disorders

Several studies have documented decreased global brain volume in children and adolescents with anorexia nervosa (AN), which may be somewhat reversible. However, decreased large-scale brain volume loss is probably the result of malnutrition and is not, therefore, likely to be particularly helpful in elucidating the neurobiologic underpinnings of AN. In adolescents with AN, a lower concentration of frontal NAA has been noted than in controls. Of interest, a lower frontal PDE level was reported as well, indicating a possible decrease in the fraction of MPLs. Furthermore, adult females with AN show a reduction in glutamate in the anterior cingulate cortex, occipital cortex, and putamen. However, spectroscopy studies have yielded conflicting results overall and are limited for other eating disorders. fMRI studies show greater activation in AN, bulimia nervosa, and binge eating disorder patients than in controls when viewing food-related stimuli and/or making choices about food. However, these findings have been mixed for the latter two groups. During working memory tasks, greater activation of temporal and parietal areas was noted in adolescents with AN that resolved with treatment. Functional studies of reward and punishment have also shown that adolescents with AN are especially responsive to losses compared to wins. It has been suggested that altered response to reward in adolescents with AN may be due to disturbed dopaminergic dysfunction, but that has yet to be thoroughly explored. Similarly, a study of 122 adolescent females revealed a positive relationship between brain activity in response to winning money and binge eating behaviors. Studies of bulimia nervosa have also noted reductions in frontal and temporal regions in affected adolescents compared to healthy subjects. Furthermore, fMRI data indicate that adolescents with bulimia nervosa demonstrate abnormal patterns of activation during tasks tapping into regulatory control and conflict resolution—with frontal and striatal regions failing to activate. Individuals with bulimia nervosa and individuals with binge eating disorder also show hypoactivation in fronto-striatal circuits. Finally, adolescents with bulimia nervosa also show increased engagement (compared to controls) in areas that mediate attention and self-referential thoughts when at rest, which might be related to excessive concerns about body shape and weight in this group.

Autism Spectrum Disorders

The number of imaging studies investigating the underlying neurobiology of ASDs has increased eightfold since 2003 (Fig. 37–23). This tremendous progress occurred despite inherent limitations. First, there have been advances in scanning young children (see Table 37–1). Early studies into ASD were hampered by being conducted in adults or adolescents. Use of mock scanners and thoughtful approaches to the scanning procedure have allowed very young children to participate. Given that typical onset of ASD is by 3 years of age, studying mostly older patients places the research at an obvious disadvantage. Second, the majority of neuroimaging studies, especially fMRI studies, have been conducted in high-functioning patients with ASD. Differences in study populations, medication exposure, and methodologic differences in image acquisition and analysis may have also contributed to the lack of reproducibility of findings.

Global Brain Abnormalities. It is believed that alterations in global brain growth are involved in the pathogenesis of ASD. Greater total brain volume has been one of the more consistent findings in patients with ASD. When examining functional connectivity, it appears that in adolescents and adults with ASD, connectivity is reduced. In younger children with ASD, however, intrinsic functional connectivity is increased. These findings have resulted in a call for research on the effect of puberty on the neurobiology of ASD. In a large-scale study (*N* = 717) of children 6 to 10 years of age, those with higher autistic traits (scores on the social responsiveness scale) demonstrated widespread areas of decreased gyrification. Whole-brain serotonin synthesis capacity has been measured in 2- to 15-year-old children with ASD with PET studies using the tracer α-[^{11}C] methyl-L-tryptophan (^{11}C-AMT) for PET. Children with ASD were compared to healthy siblings without ASD and to epileptic children without ASD because using PET to study the brains of healthy pediatric controls is rare. Sedation was required for all of the children with ASD or epilepsy and for many of the siblings without ASD who underwent PET scanning. Disruption in serotonin synthesis capacity was observed in children with ASD compared to their siblings and children with epilepsy. In normal children, whole-brain serotonin synthesis capacity was greater than 200% of adult capacity until age 5 years, and then it declined to adult levels, whereas in children with ASD, serotonin synthesis capacity increased very gradually from age 2 to 15 years to levels 1 to 1.5 times normal adult values.

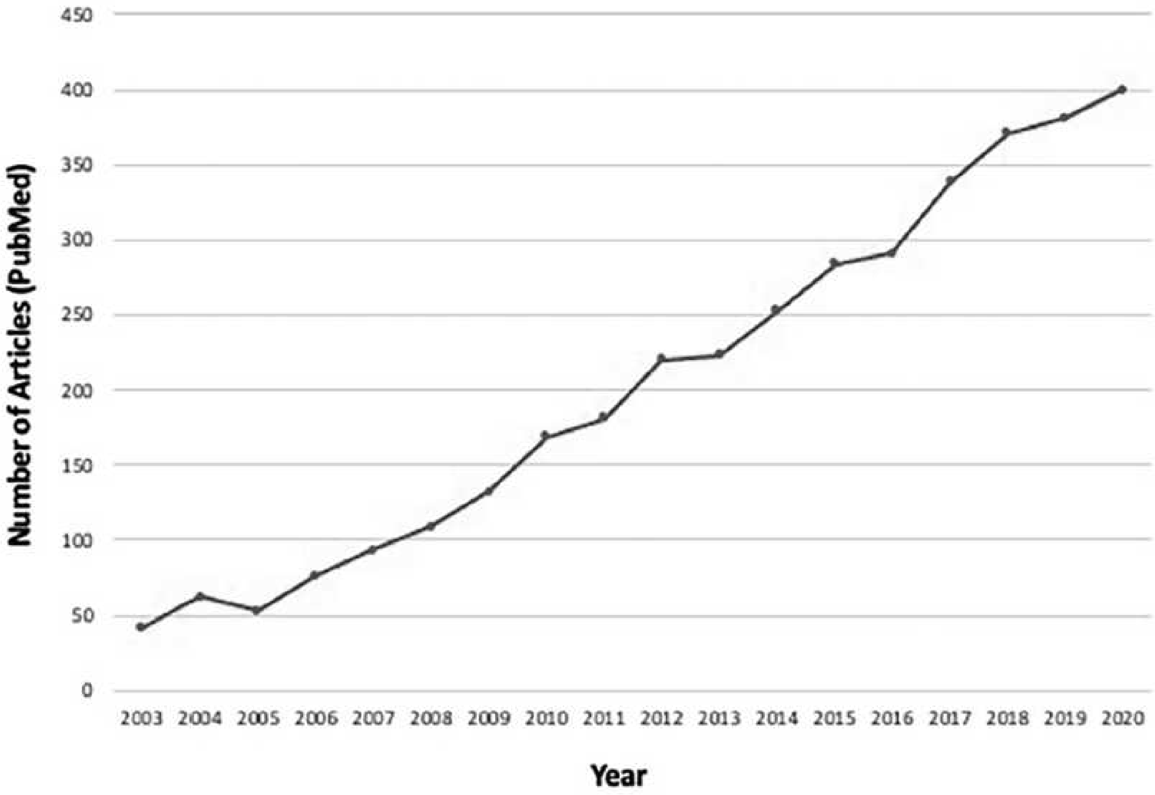

FIGURE 37–23. Number of magnetic resonance imaging articles on autism published each year 2003 through 2020. (Data gathered from PubMed search.)

Table 37–4.
Considerations for Brain Imaging Research

Common Questions	Considerations
Which brain measures should I use?	Think about your research question, are you interested in brain structure, chemistry, or function? Which is more important for your work—the spatial or temporal sensitivity? If you are using brain imaging to track short-term prognosis or treatment effect, brain function, chemistry, or temporal sensitivity is robust compared to brain structure. If you are interested in development or long-term prognosis, any of these measures will be useful. However, keep in mind, the brain changes rapidly during childhood and adolescence in as little as 6 months.
What parameters should I choose?	Consider your time and budget for the type of imaging modality you choose. Check for recent and relevant literature. Reflect on and integrate the parameters they have used—are these parameters feasible for your study, given the budget and time? There are general resources that have been published previously discussing MRI and EEG parameters extensively.
What is the ideal sample size?	This is going to depend on your study design. Identify effect sizes based on recent meta-analyses and use statistical tools to determine sample size, required power or alpha value a priori based on your study design or analysis.

Note that effect sizes for brain imaging are generally low, which means that large sample sizes might be required for statistical significance.

Regional Brain Alterations. In addition to global brain alterations in serotonin synthesis, comparisons of children with ASD who were otherwise healthy without seizures and their siblings without ASD showed asymmetries in serotonin synthesis in frontal cortex, thalamus, and cerebellum. Lower serotonin synthesis was observed in the left frontal cortex and thalamus in children with ASD. Lesser accumulation of ^{18}F-DOPA in the anterior medial PFC was observed in medication-free children with ASD. Lower levels of DLPFC phosphocreatine, as measured by ^{31}P-MRS, in high-functioning adolescent and young adult men with ASD suggest altered metabolism and increased membrane degradation in this region were correlated with neuropsychological and language deficits. The temporal lobe plays a critical role in emotion processing. Hence, it is not surprising that pediatric ASD has been associated with lower metabolic rate, perfusion, and NAA in temporal regions. Volumetric studies of the amygdala and hippocampus have been inconsistent in ASD. In the cerebellum, the dentate nucleus has also been implicated in ASD. The dentate nucleus is involved in the regulation of language and motor output. Abnormal activation of the right dentate has been observed in patients with ASD when performing a language task. In response to sensory stimuli, along with the expected sensory areas, the hippocampus, amygdala, and orbital frontal cortex demonstrate hyperactivity in youth with ASD. Furthermore, greater brain serotonin synthesis has been reported in the right dentate nucleus. Reductions in the neuronal marker NAA in the dentate as measured by ^{1}H-MRS have also been reported. Finally, a large-scale examination of 1,571 individuals with ASD revealed larger lateral ventricle and intracranial volume, increased frontal and decreased temporal cortical thickness, as well as lower subcortical volumes of the pallidum, putamen, amygdala, and nucleus accumbens compared to 1,651 controls. The trajectory of these subcortical volumes was associated with age, peaking during adolescence before decreasing in adulthood.

Specific Learning Disorders

The left hemisphere plays a critical role in the development of reading skills in healthy children. This function appears to be impaired in children with a reading disability, as suggested by lower functional activation in dorsal and ventral regions. Indeed, similar results have been noted in adults with a reading disability, with less functional connectivity between these circuit substrates.

More recently, bilateral prefrontal regions that link brain regions associated with phonologic awareness and rapid naming also demonstrated abnormal functional connectivity in children with dyslexia. The planum temporale is associated with a number of auditory and language-related processes. In children with dyslexia, the planum temporale demonstrates a greater rightward asymmetry. Perhaps to compensate for this dysfunction, children and adults with a reading disability “overactivate” the DLPFC, including the inferior frontal gyrus. Adults with a reading disability also engage the parallel circuitry in the right hemisphere. Right hemisphere activation during reading is normal in children but typically diminishes into adulthood. This suggests a change in the lateralization that characterizes the neural substrates that mediate reading in individuals with persistent reading disability.

FUTURE DIRECTIONS

The use of neuroimaging in children with psychiatric diseases provides an exciting opportunity to explore the developmental abnormalities that lead to mental illness (see Table 37–4). Neuroimaging studies in childhood-onset neuropsychiatric disorders that use biomarkers as both moderator (present at baseline) and mediator (changing with treatment) may potentially result in an enhanced mechanistic understanding of the pathogenesis and maintenance of these illnesses, which could lead to better diagnostic and treatment approaches. One such approach might include the identification of a neurobiologic treatment target for novel treatment tools, such as brain stimulation. Disentangling the impact of comorbidities on disorders and how many disorders seem to converge along key brain regions and circuits is also of great importance. MR-based technologies are particularly useful in youth because they entail no radiation exposure and can be successfully applied longitudinally. Although technologic advances will undoubtedly continue to provide more precisely resolved images, investigators must pay special attention to developing consistency in methodologic approaches across studies, along with larger sample sizes to account for the high variability of brain volume and function during childhood and adolescence. With the advent of large-scale, longitudinal, publicly available neuroimaging datasets, together with continual advancements in technology, ample opportunities lie ahead to unravel the pathophysiology and mechanism of treatment response in various childhood-onset neuropsychiatric disorders.

FURTHER READINGS

Ahmed F, Ras J, Seedat S. Volumetric structural magnetic resonance imaging findings in pediatric posttraumatic stress disorder and obsessive-compulsive disorder: a systematic review. *Front Psychol*. 2012;3:568.

Chugani D. Autism. In: Ernst M, Ramsey JM, eds. *Functional Neuroimaging in Child Psychiatry*. Cambridge University Press; 2000:179.

Chugani HT. Biological basis of emotions: brain systems and brain development. *Pediatrics*. 1998;102(5 Suppl E):1225–1229.

Croarkin PE, Wall CA, McClintock SM, Kozel FA, Husain MM, Sampson SM. The emerging role for repetitive transcranial magnetic stimulation in optimizing the treatment of adolescent depression. *J ECT*. 2010;26(4):323–329.

*Ecker C, Bookheimer SY, Murphy DGM. Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan. *Lancet Neurol*. 2015;14(11):1121–1134.

Eichele H, Plessen KJ. Neural plasticity in functional and anatomical MRI studies of children with Tourette syndrome. *Behav Neurol*. 2013;27(1):33–45.

Ernst M, Rumsey JM. *Functional Neuroimaging in Child Psychiatry*. Cambridge University Press; 2000.

*Giedd JN, Raznahan A, Alexander-Bloch A, Schmitt E, Gogtay N, Rapoport JL. Child psychiatry branch of the National Institute of Mental Health longitudinal structural magnetic resonance imaging study of human brain development. *Neuropsychopharmacology*. 2015;40(1):43–49.

Gilbert AR, Moore GJ, Keshavan MS, et al. Decrease in thalamic volumes of pediatric patients with obsessive-compulsive disorder who are taking paroxetine. *Arch Gen Psychiatry*. 2000;57(5):449–456.

Jacobus J, Tapert SF. Effects of cannabis on the adolescent brain. *Curr Pharm Des*. 2014;20(13):2186–2193.

*Keshavan MS, Giedd J, Lau JYF, Lewis DA, Paus T. Changes in the adolescent brain and the pathophysiology of psychotic disorders. *Lancet Psychiatry*. 2014;1(7):549–558.

*MacMaster FP. Translational neuroimaging research in pediatric obsessive-compulsive disorder. *Dialogues Clin Neurosci*. 2010;12(2):165–174.

MacMaster FP, Rosenberg DR, Kaufmann J. Neurobiology of early onset mood disorders. In: Martin A, Scahill L, Kratochvil CJ, eds. *Pediatric Psychopharmacology*. Oxford University Press; 2010:124–139.

Pavuluri M. Brain biomarkers of treatment for multi-domain dysfunction: pharmacological FMRI studies in pediatric mania. *Neuropsychopharmacology*. 2015;40(1):249–251.

Peterson BS, Thomas P. Tourette's syndrome: What are we really imaging? In: Ernst M, Ramsey JM, eds. *Functional Neuroimaging in Child Psychiatry*. Cambridge University Press; 2000.

Rosenberg DR, Hanna GL. Genetic and imaging strategies in obsessive-compulsive disorder: potential implications for treatment development. *Biol Psychiatry*. 2000;48(12):1210–1222.

Rosenberg DR, Keshavan MS. A.E. Bennett research award. Toward a neurodevelopmental model of obsessive–compulsive disorder. *Biol Psychiatry*. 1998;43(9):623–640.

Rosenberg DR, MacMaster FP, Keshavan MS, Fitzgerald KD, Stewart CM, Moore GJ. Decrease in caudate glutamatergic concentrations in pediatric obsessive-compulsive disorder patients taking paroxetine. *J Am Acad Child Adolesc Psychiatry*. 2000;39(9):1096–1103.

Rosenberg DR, MacMillan SN, Moore GJ. Brain anatomy and chemistry may predict treatment response in paediatric obsessive–compulsive disorder. *Int J Neuropsychopharmacol*. 2001;4(2):179–190.

*Rubia K, Alegria AA, Brinson H. Brain abnormalities in attention-deficit hyperactivity disorder: a review. *Rev Neurol*. 2014;58(Suppl 1):S3–S16.

Sowell ER, Toga AW, Asarnow R. Brain abnormalities observed in childhood-onset schizophrenia: a review of the structural magnetic resonance imaging literature. *Ment Retard Dev Disabil Res Rev*. 2000;6(3):180–185.

Stanley JA, Kipp H, Greisenegger E, et al. Evidence of developmental alterations in cortical and subcortical regions of children with attention-deficit/hyperactivity disorder: a multivoxel in vivo phosphorus 31 spectroscopy study. *Arch Gen Psychiatry*. 2008;65(12):1419–1428.

Strawn JR, John Wegman C, Dominick KC, et al. Cortical surface anatomy in pediatric patients with generalized anxiety disorder. *J Anxiety Disord*. 2014;28(7):717–723.

Thompson PM, Giedd JN, Woods RP, MacDonald D, Evans AC, Toga AW. Growth patterns in the developing brain detected by using continuum mechanical tensor maps. *Nature*. 2000;404(6774):190–193.

Val-Laillet D, Aarts E, Weber B, et al. Neuroimaging and neuromodulation approaches to study eating behavior and prevent and treat eating disorders and obesity. *Neuroimage Clin*. 2015;8:1–31.

Yang XR, Kirton A, Wilkes TC, et al. Glutamate alterations associated with transcranial magnetic stimulation in youth depression: a case series. *J ECT*. 2014;30(3):242–247.

38

Temperament: Risk and Protective Factors for Child Psychiatric Disorders

David C. Rettew, M.D.

Individuals differ in the types and intensity of their emotional and behavioral responses to various environments. Some people are quick to experience anger or anxiety, whereas others are more laid back. Some people seek out new and exciting experiences, whereas others prefer a predictable routine and the comfort of a few familiar friends. Some people are able to delay their impulses and overcome obstacles while working toward a long-term goal, whereas others become quickly distracted or frustrated. These widely recognized human qualities refer to temperament or personality traits that, along with physical and cognitive dimensions, distinguish one person from the next.

Scientists have been observing and studying child temperament and personality development for many years. The bulk of this work has existed beyond the view of most mental health clinicians, under the notion that this area was distinct from clinical psychiatric disorders. Accumulating research, however, has demonstrated strong links between various temperamental traits and nearly all types of psychopathology. The prevailing question today no longer is *if* temperament is associated with psychopathology but *how* temperament interacts developmentally in shaping children toward and away from serious psychiatric disturbance. This chapter offers a brief overview of temperament, including its history, definition, core dimensions, stability, and underlying neurobiology. Following these sections is a summary of the known relations between specific temperamental traits and specific psychiatric disorders. Finally, the chapter concludes with a discussion of possible mechanisms through which temperamental traits operate to increase risk or, conversely, promote wellness and resiliency throughout development.

HISTORICAL ANTECEDENTS

When most students think of temperament, the two names that often arise are child psychiatrists Stella Chess and Alexander Thomas, who brought the concept of temperament and its role in developmental psychopathology into the spotlight more than 50 years ago. While their work remains pivotal, the foundations of temperament theory can be traced back in time much further. The ancient Greek physician Galen postulated the existence of four temperament types, which he called sanguine (extraverted and stable), choleric (extraverted and unstable), melancholic (introverted and stable), and phlegmatic (introverted and unstable), based on the relative concentrations of four "humors" (blood, yellow bile, black bile, and phlegm). Although this early conceptualization of temperament may appear somewhat comical by today's standards, a thought experiment that substitutes the four humors with modern neurotransmitters such as serotonin and norepinephrine reveals how much of this framework remains intact. Indeed, the very word "temperament," derived from the Latin *temperare*, which means "to mix," conveys this association between the functioning of important biologic substrates in the brain and basic human behavioral and emotional predispositions.

Fast forward to the 20th century and the dawn of modern psychiatry: it is worth noting that Sigmund Freud made few qualitative distinctions between the origins of normal personality traits and the factors responsible for psychiatric symptoms, although the nature of those forces were certainly not described as temperament based. As the tenets of behaviorism and learning theory gained prominence, there again were few distinguishing features between the postulated mechanisms underlying personality and those that caused the formation of psychiatric disorders.

It was in this climate during the 1950s that Chess and Thomas began their work on temperament. Based on interviewing parents and observing children across many years, they proposed nine basic dimensions of temperament, as shown in Table 38–1, and three temperament types—easy, slow to warm up, and, most famously, difficult. These basic elements of temperament were hypothesized to be substantially heritable rather than based on learning or unconscious conflicts. It was this heritable quality of temperament that ran against the prevailing wisdom of both the psychoanalytic and the behavioral schools of thought, and the work of Chess and Thomas was initially subjected to intense criticism.

Through the 1980s and continuing today, a medical model of psychiatric illness began to emerge. Under this framework, disorders like depression or panic disorder arise like other medical illnesses such as diabetes or the flu. Genetic variations and environmental pathogens are viewed as important etiologic factors in this model, and the disorders are conceptualized as binary entities that a person does or does not possess. While such an approach acknowledges the influence of more innate dispositional factors, this disease model of psychiatric disorders also divorces an individual's symptoms from their temperament. A panic disorder, for example, is an illness that "comes out of nowhere" rather than being related to a temperamental proclivity toward experiencing fear.

Table 38–1.
Thomas and Chess Classification of Temperament

Nine Temperament Dimensions
Activity level
Rhythmicity
Approach–withdrawal
Adaptability
Response threshold
Response intensity
Mood
Distractibility
Attention span/persistence
Three Temperament Types
Easy
Difficult
Slow to warm up

Table 38–2.
Definition of Temperament

Individual differences in basic reactivity and self-regulation include the following elements:

1. Heritability: temperamental traits are often passed on from one generation to the next through both genetic and environmental mechanisms.
2. Early appearance: the antecedents, if not the full traits, are evident in infancy and toddlerhood.
3. Stability: although there is certainly the potential for change, temperamental traits are moderately stable across many years.

Thus, researchers and clinicians interested in temperament continue to challenge and be challenged by the changing models and theories of psychiatric disease. The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), while specifically acknowledging the dimensionality of most emotional–behavioral symptoms, retained a disease model of psychopathology in its operational structure. The implications of this decision will be discussed later in this chapter after additional background information on temperament is provided.

DEFINITION OF TEMPERAMENT

Despite the long history of temperament investigation, the precise definition of temperament remains elusive. One prominent temperament researcher, Mary Rothbart, described temperament as “constitutionally based individual differences in reactivity and self-regulation.” Although the debate continues, there exists some consensus on key components of a temperamental trait. First, there needs to be a heritable basis to the trait. This quality does not mean that temperament is purely genetic, but rather that there is at least some genetic influence on the phenotypic expression of that trait. Second, a temperament dimension also needs to be relatively stable across time. Similar to the point about genetic influence, stability does not mean inflexibility or, certainly, destiny, but rather that the trait at one time point can predict future time points even many years into the future. Finally, a temperament dimension is expressed early in life, with its rudiments evident as early as infancy. This latter characteristic distinguishes temperament from other important human qualities, such as life satisfaction or generosity, which incorporate higher cognitive functions. These aspects of temperamental traits are summarized in Table 38–2.

Table 38–3.
Current Dimensions of Temperament and Personality

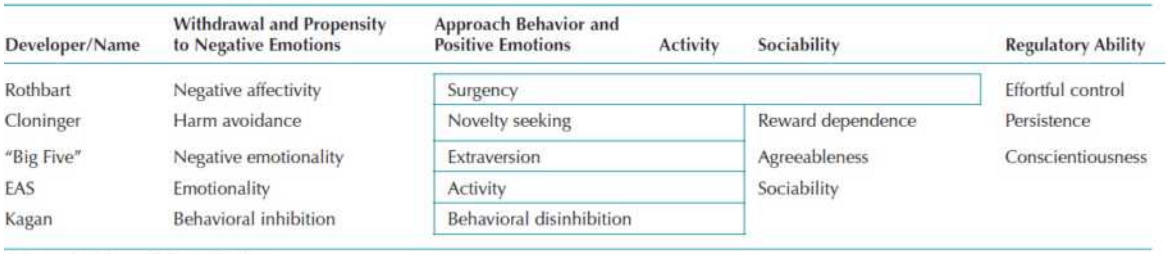

Developer/Name	Withdrawal and Propensity to Negative Emotions	Approach Behavior and Positive Emotions	Activity	Sociability	Regulatory Ability
Rothbart	Negative affectivity	Surgency			Effortful control
Cloninger	Harm avoidance	Novelty seeking		Reward dependence	Persistence
"Big Five"	Negative emotionality	Extraversion		Agreeableness	Conscientiousness
EAS	Emotionality	Activity		Sociability	
Kagan	Behavioral inhibition	Behavioral disinhibition			

EAS, emotionality, activity, sociability.

MAJOR DIMENSIONS

The core dimensions of temperament continue to be a source of discussion. As previously described, Chess and Thomas hypothesized nine core temperament dimensions and three temperamental types, with the latter referring to a constellation of traits that often coexist as a specific profile. Subsequent testing using different samples and more sophisticated statistical modeling have supported many of these dimensions, whereas others were found in need of modification.

Currently, there exist a number of alternate, even competing, frameworks that have received some empirical support for their reliability and validity. Most temperament taxonomies are structured around basic incentive response systems of approach and withdrawal often merged or crossed with emotional valence (propensities toward positive and negative emotions). Across these various taxonomies, however, there exist named dimensions that overlap somewhat but not completely with a corresponding dimension in another theoretical model. The temperament trait of novelty seeking as described by Cloninger is similar but not identical to a dimension referred to by Rothbart and colleagues as extraversion or surgency. Adding additional complexity is a dimension's relative place along a hierarchical organization from lower- to higher-order traits. Activity level, for example, exists as a component of the surgency dimension in Rothbart's framework but functions as a broad higher-order dimension in the Buss and Plomin model. Furthermore, differences of opinion exist between the conceptualization of temperament traits as quantitative factors that are distributed normally along a continuum versus more binary yes/no designations. All of these variations between the different models of temperament that exist today provide a rich environment for further research and discussion but can be challenging when trying to compare findings across studies. These conceptual challenges notwithstanding, several temperament dimensions are fairly consistent across different nomenclatures and constitute the highest level of temperamental traits. These traits are depicted in Table 38–3.

Clearly, there is a tremendous amount of overlap between the core dimensions of temperament and personality, and a full discussion comparing these two domains is beyond the scope of this chapter. Briefly, it is probably fair to say most temperament and personality theorists view temperament as one of the core foundations of personality, with the latter being further developed through additional life experiences, cultural influences, and cognitive abilities. Because these influences act on children and adults alike, however, a clear boundary between a temperament and a personality trait is obscure and somewhat akin to pinpointing exactly when someone is, for example, an adult. At the same time that some prominent personality researchers such as Costa and McCrae argued that their personality dimensions can largely be seen as temperament, child researchers found evidence that the “Big Five” taxonomy of personality is largely applicable in youth samples. These efforts suggest that at least some of the historical distinction between these two constructs was semantic and based on the age of the sample being studied with the term temperament being applied to children and the term personality applied to adults.

Withdrawal and Propensity to Experience Negative Emotions

A common core dimension in temperament found across many different taxonomies refers to a child’s threshold to experience negative emotions such as sadness, fear, or anger, the intensity of that emotion when triggered, and the duration of that emotional experience. This trait has been described using a number of different labels, including negative affectivity, emotionality, and neuroticism. Some researchers have described a similar trait that reflects an underlying sensitivity to signals of punishment and have referred to this dimension as harm avoidance or the behavioral inhibition system (BIS). Some evidence supports greater separation between the tendency to experience anger and frustration and the tendency to experience fear or anxiety. Nonetheless, components of this dimension often include elements such as worry, shyness, irritability, and difficulty being soothed once aroused. This trait also figured heavily in the Chess and Thomas designation of the “difficult” child.

Approach and Positive Emotions

The other temperament dimension that exists throughout various taxonomic structures refers to the degree to which a person actively engages the world and other people. This trait often also encompasses a tendency to experience positive emotions such as joy and excitement. Additional components of this dimension according to the various frameworks include elements of dominance, preference for high-stimulus activities, and approach behavior. Especially when considered in a more social context, this trait is often referred to as extraversion, whereas from a slightly broader perspective, it has been labeled novelty or sensation seeking or the behavioral activating system (BAS).

Emotional Regulation and Constraint

The third dimension common to most temperamental systems relates to emotional and attentional regulation especially in the face of obstacles or in the pursuit of a long-term goal. As a separate dimension, it has been referred to as effortful control, persistence, or, under the “Big Five” framework, conscientiousness. Children with lower levels of this trait are often described as impulsive, distractible, and disorderly. Parallels to more philosophical terms encompassing a person’s degree of voluntary control such as “will” are apparent and distinguish this dimension from behavior that is restrained by anxiety in children with high negative affectivity or low approach.

Other Dimensions

A number of other traits have sometimes been classified as broad-level traits under some frameworks while in others they are a component of one of the three dimensions described above. Three of them are the following.

Social Affiliation. Although aspects of social affiliation can be subsumed under broad conceptualizations of extraversion, other researchers have viewed a person’s tendency toward prosocial behavior as a distinct higher-order trait. Elements of social affiliation are also found in terms such as shyness, although merging these concepts has been seen as problematic by some because it equates the *desire* for social contact with the anxiety or *discomfort* associated with social encounters. As an aspect of temperament, this dimension has been called sociability and reward dependence, among others, whereas under a personality "Big Five" taxonomy it is related to the trait of agreeableness. In some systems, this dimension encompasses elements such as empathy and the desire and ability to share emotional states with others, as well as with dependency and compassion.

Activity Level. A person's level of motor activity is similarly viewed as a component of extraversion in some models, whereas in others, such as that proposed by Buss and Plomin, it is given additional standing in itself. Its ability to be readily observed by others, genetic influence, and emergence as a distinguishing feature very early conform to some of the core requirements of a temperamental trait.

Behavioral Inhibition and Disinhibition. One well-studied temperament trait has been called behavioral inhibition (BI). It was defined by Harvard psychologist Jerome Kagan and colleagues to refer to a subgroup of young children (about 10% to 15% of the population) who consistently responded to new situations and people with reticence and fear. His research group later defined a group of children with the exact opposite response to novelty and referred to them as possessing behavioral disinhibition. Unlike the conceptualization of most temperament dimensions, BI has been studied primarily as a binary trait that a person either does or does not possess. As will be discussed in more detail, children who manifest BI as toddlers have been shown to be at risk for psychiatric disorders later in life, in particular social anxiety disorder.

Temperament Types

As mentioned, an alternative approach to understanding temperament focuses on distinguishing groups of individuals who may possess similar levels across a variety of traits. This emphasis on temperament types should not be considered to be a true substitute to a variable-centered focus but rather a complementary approach, in that these person-centered classifications also rely on specific dimensions to define various classes or types.

The Difficult Child. Probably the most famous temperamental type is the "difficult" child designation of Chess and Thomas, which was distinguished from the "easy" and "slow to warm up" groups (and from children who didn't fit any of the three types). The "difficult" category referred to a cluster of about 10% of young children who possessed the particular combination of low rhythmicity, low response threshold and adaptability, and high response intensity. Although not hypothesized to reflect psychopathology per se, it was this group of children in longitudinal studies who were most likely to have poor outcomes and greater psychiatric disturbance.

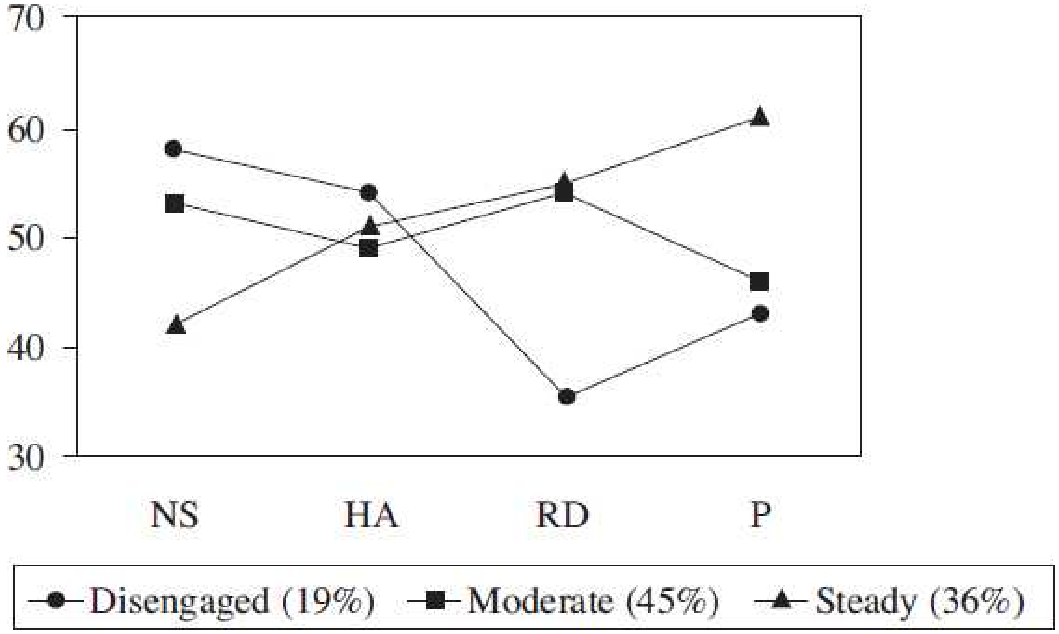

FIGURE 38–1. Temperament profiles as determined through latent profiles analysis. The "moderate" group has average levels across all dimensions, whereas the "steady" group is characterized by low novelty seeking (NS) and high persistence (P). The "disengaged" group has moderately high levels of both novelty seeking and harm avoidance (HA), with low levels of reward dependence (RD). The groups have been found to be related to levels of psychopathology and wellness. (Adapted from Rettew DC, Althoff RR, Dumenci L, Ayer L, Hudziak JJ. Latent profiles of temperament and their relations to psychopathology and wellness. *J Am Acad Child Adolesc Psychiatry.* 2008;47(3):273–281.)

New Approaches to Temperament Types. Statistical procedures such as cluster analyses and latent profile analysis can now be used empirically to identify groups of children who manifest similar levels of traits across several dimensions. These analyses can be performed without preconceived theories of how different traits *should* exist together. Studies that apply these kinds of analyses typically find evidence for a relatively small (often around three to six) number of groups distinguishable by their unique temperamental profile. In one study, latent profile analysis of a group of children of ages up to 18 years revealed the presence of three temperament types (Fig. 38–1). The first, as might be expected, contained people who did not have extreme levels of any dimension and was labeled the “moderate” group. A second group was characterized by low novelty seeking and high persistence and was named the “steady” group. Finally, the least common group contained those who possessed high novelty seeking, high harm avoidance, low persistence, and, most prominently, low reward dependence (low sociability and attachment to others). This last group was labeled the “disengaged” group. As will be discussed, these groups had strong associations with both positive and negative outcomes. Similar efforts by other investigators and using other samples have also tended to find evidence for a relatively small (often around 3 to 6) number of groups distinguishable by their unique temperamental profile.

FEATURES OF TEMPERAMENT

Temperamental traits can be observed in infancy and, according to some, even before birth. They have been observed across a wide number of cultures. Although temperament contributes to an individual’s unique human characteristics, these traits have been widely studied in animals ranging from nonhuman primates to dogs to guppies.

Sex Differences

A review and meta-analysis of sex differences using Cloninger’s temperament framework showed higher levels of reward dependence and harm avoidance in women than in men, whereas novelty seeking and persistence were found to be roughly equal. Of interest, another meta-analysis of child studies using Rothbart’s temperament structure found slightly different results with corresponding dimensions. Specifically, effortful control was found to be higher in girls, surgency was higher in boys, and negative affectivity was roughly equal. These results suggest the possibility that gender differences in temperament may be age-related. Overall, sex differences with many temperament dimensions can be found but are often smaller in magnitude than might be predicted from cultural stereotypes.

Continuity

By definition, temperamental traits are at least moderately stable across time, and most research supports this assertion. This quality does not mean, however, that a person's temperamental profile should be viewed as unchangeable or as someone's "destiny." Studies that measure how strongly a level of a temperament trait at one time point predicts the same trait in the future routinely find modest but statistically significant correlations even decades later. Longitudinal assessments of BI have shown that, while most children do not retain that designation years later, the presence of BI at follow-up can be predicted by initial BI status. In general, increasing stability of a temperament trait has been related to (1) higher baseline age of temperament assessment, (2) shorter time intervals between assessments, and (3) possessing more extreme levels of the trait.

An interesting offshoot of the temperament stability question relates to the possibility that physiologic "footprints" of traits persist even in the absence of observable behavior. In one study, adults who at the age of 2 years were classified as being behaviorally inhibited were scanned using functional magnetic resonance imaging (fMRI). When shown pictures of novel faces, the previously inhibited adults showed higher amygdalar activity than adults who were not inhibited as toddlers, despite the fact that these subjects did not report any increased anxiety. This intriguing result and others like it suggest the possibility that the physiology underlying a temperamental trait can linger despite an outward lack of the trait itself. One mechanism through which this may occur is that other brain areas and circuits, such as those involved in executive functioning, might mature with development and keep a person's behavior more regulated despite enduring internal feelings of discomfort.

Another type of stability, mean level stability, refers to the overall level of personality traits at one age compared to each other. This can be a difficult phenomenon to measure, especially in childhood and adolescence due to the fact that what behaviors constitute a particular trait can change across development. These challenges notwithstanding, there is evidence that across children and adolescents, different traits have different trajectories over time and may even vary between girls and boys. Levels of extraversion, for example, have been shown to dip somewhat in midadolescence before climbing in late adolescence and early adulthood.

NEUROBIOLOGY AND ETIOLOGY

There has been an explosion of research into the neurobiologic underpinnings of temperament and personality. Although a thorough review of this literature is beyond the scope of this chapter, many important findings are particularly relevant for understanding the boundaries and interplay between temperament and psychopathology. For many years, research on the neurobiology of psychiatric disorders existed in parallel with similar efforts in temperament and personality. As these independent efforts began to reveal similar findings, however, there have been increased efforts to assess temperament and psychopathology concurrently.

Genetic Influences

The genetic influence on temperamental traits has been well established through a large number of behavioral and molecular genetic studies. Twin and adoption studies across all ages have generally shown that genetic factors explain about 20% to 60% of the variability in temperament scores within study populations, depending on age and the particular dimension being examined. These heritability coefficients tend to be highest in twin studies when temperament is assessed by a questionnaire completed by a single informant (such as a parent rating both twins) and lower in adoption designs or studies that measure temperament using observational protocols. Furthermore, this genetic influence has often been found to be mainly of the additive type, in which multiple genes exert small, independent effects. Nonadditive genetic influence, such as in genetic dominance or epistasis, have also been found to a lesser degree in some studies. Although most studies usually do not find major sex differences in the *magnitude* of genetic factors, it is not uncommon to find evidence that different genes contribute to particular dimensions between males and females. One somewhat counterintuitive finding that has emerged in some studies is that the heritability of some traits appears to be fairly low in the neonatal and early infancy period—a time in which some might expect to find the largest genetic influence—and then grow over time. These findings lead to the conclusion that environmental factors can play important roles even in infancy, and new genetic and environmental factors come online to contribute to both temperament stability and change across development. These results are reflected in many anecdotal reports of parents who marvel at how different their children's temperaments are from each other despite the children sharing half of their genes and being raised in the same household. Indeed, studies have documented only modest correlations between siblings. Although statistically significant, the moderate level of temperamental differences between siblings suggests that the effects of genetic factors are limited and that individual environmental experiences may play a large role in shaping traits over time.

The question of exactly *which* genes contribute to a particular trait is left for molecular genetic studies that attempt to link temperamental traits to chromosomal regions or known alleles of specific genes. While the validity of older candidate gene studies has more recently been questioned, genome-wide association studies continue to be performed and yield positive results, with some recent efforts showing temperament related to single nucleotide polymorphisms (SNPs) in a number of genes involved in synaptic plasticity, learning, and stress reactivity.

Environmental Factors

As mentioned, the other major source of influence for temperamental traits comes from environmental factors. These experiences have been documented to have stronger influences on temperament than many have assumed. From a behavioral genetic perspective, environmental factors come in two main types. The first are those events that have a similar effect on all members of a family and are called shared or common environmental events. Socioeconomic status, parenting practices, and school quality are all potential shared environmental factors to the degree that their effect on siblings in a family is equal and in the same direction. Nonshared environmental factors, by contrast, are environmental events that are not shared by members of the same family and can contribute to making siblings different from each other. Examples here could include a child's peer group or a powerful positive or adverse experience.

Many of the same twin and adoption studies that have shown evidence for genetic influence also show unequivocal evidence for environmental influence, particularly of the unshared type. At the same time, there has been much less support for robust shared environmental effects. This pattern does not necessarily mean that things such as parenting or poverty are unimportant, but rather that they may not affect people as uniformly as was once hypothesized.

Genetic–Environmental Interplay

The old nature-versus-nurture debates assumed that genetic and environmental effects work through mutually exclusive pathways. A child inherits a particular gene or experiences a certain event, and those factors affect behavior directly. More recently, however, it has become increasingly appreciated that genes and environments have bidirectional influences. Environmental events can control gene expression. Conversely, genetically influenced tendencies can cause certain environments to be more likely to occur (gene–environment correlations), and those environments can, in turn, affect behavior. Finally, the effect of a particular gene can depend on the presence of a particular environment and vice versa (gene–environment interactions). Understanding these complex relations has become a major focus of neuroscience research and has brought with it a new perspective on human development that crosses old boundaries.

Neurotransmitters and Brain Circuitry

Whether behavioral influences are genetic, environmental, or an intricate combination of the two, their ultimate effect is to shape specific brain regions and their connections to other areas. Advances in the knowledge of brain function and its plasticity from environmental factors are rapidly rendering obsolete terms such as “biologic” or “organic” to describe some behaviors but not others. Nevertheless, there is still much to learn about the specific pathways and neurochemical substrates involved in temperament. Many of the specifics remain unknown, and current models and explanations are acknowledged to be vast oversimplifications.

The trait of negative affectivity and related terms such as harm avoidance and BI have often focused on the amygdala and its multiple afferent and efferent pathways. In a study combining genetic and neuroimaging techniques, researchers found significant associations between reduced amygdala/anterior cingulate connectivity and higher levels of harm avoidance. Amygdalar hyperreactivity has also been postulated to be one of the core features underlying the trait of BI.

Ascending pathways from the reticular activating system to the medial septal area and modulated by the orbital frontal cortex have also been related to passive avoidance and behavioral extinction. Related to these functions, dysregulation of the hypothalamic–pituitary–adrenal (HPA) axis has also been found to be associated with this dimension.

Animal studies have implicated ascending serotoninergic pathways from the dorsal raphe to the basal ganglia, substantia nigra, nucleus accumbens, and amygdala. These pathways can be modulated by γ-aminobutyric acid (GABA), as well as by dopaminergic pathways from nigrostriatal regions. Dopamine pathways have frequently been found to play an important role in the approach behavior and spontaneous activity related to the trait of novelty seeking or extraversion. The mutually interacting systems of harm avoidance and novelty seeking may be related to the typical negative correlation between these two traits. Some brain laterality has also been reported in EEG studies, with left frontal activation being related to more positive emotions and approach, and right frontal activation being related to more negative emotions and withdrawal.

Social affiliation and related traits such as reward dependence are hypothesized to be related to noradrenergic pathways from the locus ceruleus, as well as serotoninergic projections from the median raphe, which innervate a number of areas in the cortex, thalamus, and hippocampus. Neuroactive hormones such as oxytocin may also play important roles in social behavior, as illustrated by the marked behavioral differences between two species of voles that correspond to alterations in the distribution of oxytocin receptors in brain reward centers. Dimensions such as persistence or effortful control are thought to be mediated through many areas of the anterior cortex that modify the function of subcortical regions. Many of these pathways also involve dopaminergic transmission. As previously described, the anterior cingulate gyrus is connected both to limbic structures such as the amygdala and to areas of the lateral cortex, thereby providing a link between key emotion processing and executive function regions.

In describing the genes, neurotransmitters, and neuropathways related to temperament, it becomes quickly apparent that the discussion involves all of the “usual suspects” in psychiatric illness. Realization of this repeated finding has prompted systematic efforts to delineate the boundaries between temperamental traits and psychiatric illness in the hope of identifying potential targets of therapeutic intervention.

TEMPERAMENT AND PSYCHOPATHOLOGY

Despite the current conceptualization of psychiatric illness in DSM-5 as binary entities that exist in all-or-none form, there is now overwhelming evidence that almost all symptoms of psychopathology exist on a continuum. This spectrum ranges from a large segment of the population who present with few or no difficulties in a particular area to a gradually diminishing number of individuals with higher degrees of symptomatology and finally to a small number of people with extreme levels. Having both a quantitative and dichotomous perspective on a particular disease is not unique to psychiatry. Although all people have a measurable blood pressure, for example, individuals with a pressure greater than 140/90 mm Hg are considered to have the diagnosis of hypertension.

As the dimensional aspects of psychiatric disorders have become increasingly appreciated and measured, it is not surprising that people have begun to hypothesize that temperament or personality dimensions may underlie this wide distribution of symptom severity. This realization may have been somewhat inevitable, given hypothesized temperament dimensions such as mood, activity level, and attention span, all of which have descriptive similarities to specific psychiatric conditions. Indeed, when the range of psychiatric instruments is extended to include those on the opposite end of the spectrum (e.g., going beyond the distinction between children with attention problems and no attention problems to include those with excellent abilities), the distribution of the scores begins to approximate the bell-shaped curve of many temperament traits. This understanding has led to the speculation that what is currently classified as psychiatric disease may actually represent the extreme tail of normally distributed traits that exist in the entire population.

Table 38–4.
Possible Mechanisms of Temperament/Psychopathology Associations

Model	Description	Example
Spectrum/continuum	Most psychiatric disorders are best conceptualized as existing on a continuum with temperament or personality traits.	Activity level is a temperamental trait, and people with attention-deficit/hyperactivity disorder are at the end of that spectrum.
Risk/vulnerability	Temperament is different from psychopathology but can increase the risk of its occurrence.	People with high levels of novelty seeking are at increased risk for substance abuse.
Common factor	Temperament and psychopathology share common etiologic factors, which are then acted on by other factors that lead some people to develop a disorder.	People with posttraumatic stress disorder share similar genes with temperamentally anxious people but have experienced severe trauma.
Pathoplastic	Temperament and psychopathology are unique but have bidirectional influences on each other.	The long-term prognosis of depression is better for individuals with high effortful control.
Scar	The pathophysiology of psychiatric disorders can alter temperament.	The plaques and tangles associated with Alzheimer disease cause personality changes that are observed prior to other symptoms.

To distinguish “true” psychiatric illness from subthreshold or developmentally expected levels of behavior, clinical teaching has dictated that one of the most important considerations is *impairment*, that is, whether the behaviors, thoughts, or emotions interfere with a person’s daily life and functioning. While this convention is theoretically appealing, research has demonstrated that impairment itself also tends to exist along a continuum. People who are severely depressed, for example, are quite impaired, whereas those who are a little depressed are somewhat impaired, and so on, with no convenient cut-off. Complicating matters further in the field of child psychiatry is the fact that impairment is often not defined by the child but rather by caregivers, who may differ in their definition of impairment based on their own strengths, weaknesses, and expectations.

As psychiatry continues to demonstrate these strong associations between temperament and psychopathology, several mechanistic hypotheses have been proposed to explain this link. These models are summarized in Table 38–4.

The two most investigated potential mechanisms have been described as the *spectrum model* and the *risk or vulnerability model*. The spectrum perspective, as previously mentioned, postulates that psychiatric symptoms and temperament dimensions lie on a common continuum with no clear qualitative distinctions. The difference between a trait and a disorder, therefore, is simply a matter of degree or magnitude. Implicit in this theory is that there also is common underlying neurobiology between temperament and psychopathology. The factors that determine a child’s level of normal developing activity, for example, are the same as those that underlie another child with hyperactivity, albeit perhaps more amplified.

The risk or vulnerability model of temperament/psychopathology relations considers temperamental traits and psychiatric disorders to be qualitatively different entities but with temperament acting as a risk or protective factor for future development. Although it is easy to differentiate conceptually between a risk factor such as sunlight exposure and a disease such as melanoma, the description of something like shyness being a risk factor for social anxiety disorder lacks these salient points of obvious divergence. Proponents of this mechanism, however, point out that there are many examples of normally distributed variables that mask a number of very different underlying processes. Intelligence quotient, for example, often shows a normal distribution in large samples. To assume, however, that there is a common quantitative mechanism underlying this distribution would miss, among other things, the small numbers of individuals whose score is related to a distinct and qualitatively different mechanism, such as the trisomy of chromosome 21 that leads to Down syndrome. In addition, there is also a hybrid of the spectrum and risk models that proposes true phenotypic differences between temperament and psychopathology but some common underlying features, such as shared genetics, that are then acted on by other factors, such as specific adverse or positive experiences to create important points of divergence.

These three models tend to be the most dominant in the literature when it comes to temperament/psychopathology interactions. However, two other models, both of which preserve the uniqueness of pathophysiologic processes, deserve mention. In the pathoplastic model, temperament and psychopathology are distinct but can mutually influence one another. Temperament can increase the chance that a child will suffer a major depression, for example, and that depressive episode can, in turn, exert effects on the child's personality going forward. Finally, the "scar effect," in many ways the reverse of the risk model, proposes that it is the pathophysiologic mechanisms that underlie psychiatric disorders that then modify one's personality. An example of this effect may be Alzheimer disease. Although personality change is common with this condition and often precedes more specific Alzheimer symptoms, it is thought that the early-developing pathology of plaques and tangles affects personality, rather than the other way around.

Scientists have thus turned to modern research designs and tools to help disentangle these complex relations among temperament, personality, and psychopathology. In so doing, there has been a realization that what were previously thought to be clear boundaries might be much more muddled. What appears to be a continuous trait, like intelligence, may develop from a number of small, discrete causes. Conversely, what appears to be a binary yes/no illness might have more embedded gradations when examined closely. In Huntington disease, for example, the number of trinucleotide repeats in the huntingtin gene affects the onset and course of illness. It is also important to keep in mind that the various mechanisms are not mutually exclusive. Within a particular trait–disorder pair, there may not be a single mechanism at work across different individuals and perhaps even within individuals. Some individuals with obsessive-compulsive disorder may have strong temperamental correlates, whereas the symptoms of other individuals might be the result of an autoimmune process.

As is clear from these examples, improved understanding of the way illness is conceptualized has important clinical implications. From a research standpoint, it is vital to ensure that any boundaries, should they exist, between temperamental traits and psychiatric illness be delineated because all results downstream of that categorization are likely to be affected. If children who just barely meet DSM-defined attention-deficit/hyperactivity disorder (ADHD) have a brain physiology that is nearly identical to children who just miss the diagnoses, then using the latter group as controls in an fMRI study, despite their lack of diagnosis, is unlikely to yield fruitful results. Thus, the next section is devoted to describing the current state of knowledge with regard to the existence and mechanism of temperament/psychopathology relations (Table 38–5).

Table 38–5.
Associations Between Temperament Traits and Various Categories of Psychiatric Disorders

	Negative Affectivity	Extraversion	Effortful Control
Attention-deficit/hyperactivity disorder	High	High	Low
Other disruptive behavior disorders	High or low	High	Low
Major depression	High	Low	Low
Anxiety disorders	High	Low	Low
Substance use	High	High	Low
Autism	High	Low	Low
Eating disorders	High	—	High (anorexia) or low (bulimia)

Adapted from Rettew DC. *Child Temperament: New Thinking About the Boundary Between Traits and Illness*. W. W. Norton & Company; 2013.

ADHD and Disruptive Behavior Disorders

ADHD is a prime example of a disorder that, at least phenotypically, appears to be conceptualized as the upward range of a continuum defined by temperamental traits related to activity level and attention span. Indeed, much of the public controversy surrounding this diagnosis is related to the perception that what is being diagnosed as illness is normal temperament, albeit at somewhat higher levels than are typically observed. As might be expected, low levels of effortful control and persistence have been found, in addition to higher levels of novelty seeking and extraversion, especially among children who meet the criteria for the hyperactive-impulsive or combined subtype of ADHD. A study of 8,500 twin pairs found evidence for a continuum model of ADHD with nearly identical heritability estimates (around 0.60) for those with diagnostic versus subthreshold levels of behaviors. Another study subtyped a group of children with ADHD according to temperament and identified three groups that they labeled mild, surgent (with high levels of approach behavior), and irritable (with higher emotionality and poorer regulatory skills). These temperament-based groups, compared to traditional DSM-based presentations, were shown to provide better cohesion with regard to stability over time, functional brain connectivity, and cardiac physiologic parameters.

Other disruptive or externalizing disorders include oppositional defiant disorder (ODD), conduct disorder, and, in adults, antisocial personality disorder. All of these entities are rather heterogeneous, making the task of linking specific temperamental traits with these disorders difficult. Aggression, for example, is often divided into more instrumental and reactive subtypes. In the former, sometimes known as type 1 psychopathy, aggressive acts can be planned in advance and without provocation. These behaviors are often associated with individuals who tend to lack empathy and possess what are termed callous and unemotional traits, now incorporated into the conduct disorder diagnosis in DSM-5. Physiologically, many of these individuals have been shown to have an underaroused autonomic nervous system that craves additional stimulation. In reactive or type 2 psychopathy, individuals quickly perceive threat from others and have difficulty modulating their emotional responses. This latter type is more strongly linked to deficits in emotional regulation, as well as to increased negative affectivity. Oppositionality has similarly been found to break into three dimensions, labeled irritable, headstrong, and hurtful. These dimensions show different comorbidity patterns with regard to other DSM disorders.

Mood and Anxiety Disorders

Often collectively referred to as internalizing disorders, mood and anxiety disorders are some of the most common psychiatric conditions in childhood and adolescence. Although they are considered to be separate categories of illness according to current clinical nomenclature, there is good empirical evidence to support their being closely linked. Consistent with this hypothesis is the finding of similar temperamental correlates common to both mood and anxiety disorders. By far the strongest association is with the dimension referred to as negative affectivity, neuroticism, or harm avoidance. Major depression has also been associated with the combination of high negative affectivity and low approach. Bipolar disorder, especially the purported “broad” phenotype now often diagnosed as disruptive mood dysregulation disorder and characterized by nonepisodic irritability and explosiveness, has been shown to be related to a somewhat unique temperamental combination of high novelty seeking *and* high harm avoidance. This “approach plus avoidance” pattern stands in contrast to the general tendency of novelty seeking and harm avoidance being inversely related to each other.

More generalized and chronic internalizing disorders such as generalized anxiety disorder (GAD), dysthymia, and social anxiety disorder have been hypothesized as the high end of a spectrum of negative affectivity and/or low extraversion. Direct analyses of this hypothesis, however, suggest that such a model is oversimplified. As shown in Figure 38–2, many people who possess even very high levels of harm avoidance do not meet the criteria for GAD, whereas some individuals with moderate levels of harm avoidance do. On the genotype level, behavioral genetic studies indicate that approximately half of the genes that contribute to neuroticism are the same genes that contribute to internalizing disorders, whereas half do not. Thus, despite the obvious content overlap between descriptions of anxious temperament and descriptions of anxiety and depressive disorders, the majority of evidence appears to point to temperament dimensions having more complex relations with these disorders rather than simply being a different point along a continuum, especially of a single temperament trait. The cyclic and episodic nature of other types of mood disorders, such as recurrent major depression or panic disorder, further suggests that elements other than temperament need to be incorporated into the current theoretical models.

Low effortful control or persistence has also been associated with internalizing disorders. Indeed, temperamental dimensions relating to regulatory control appear to be a common correlate to most psychiatric disorders, with other dimensions related more to the specific *type* of psychopathology. As mentioned, the construct of BI developed by Kagan and colleagues likely combines elements of high negative affectivity, low approach, or surgency, and perhaps low effortful control. Longitudinally, however, the strongest association found has been between BI and generalized social anxiety disorder with a meta-analysis revealing that 43% of young BI children meeting the criteria for social anxiety disorder later in development.

Autistic Spectrum Disorder

To many clinicians observing a nonverbal child with stereotypical behaviors, autism would seem to be a condition that is more binary in nature. However, as the name suggests, accumulating evidence has demonstrated that many of the core symptom clusters of autism, such as social cognition, communication, and affiliation, as well as the stereotypical behaviors themselves, exist along a continuum when measured quantitatively. Temperament dimensions such as low sociability or reward dependence, high negative affectivity, and low extraversion have all been associated with autism. At the same time, it is important to remember that the finding of a continuous distribution of phenotypic behaviors does not prove that a common neurobiologic mechanism underlies the entire spectrum. Indeed, studies showing focal areas of cytoarchitectural disorganization in the neocortex of individuals with autism suggest that, at least for some individuals, a “scar” mechanism may be at work in which the pathophysiology of the disorder drives the temperamental differences that are also observed. There is some additional support for the possibility that qualitative differences in brain structure and function exist between individuals who meet the criteria for autism and those who possess some of the same traits (albeit at lower levels) from the neuroimaging literature, although much more work is needed to explore this intriguing hypothesis.

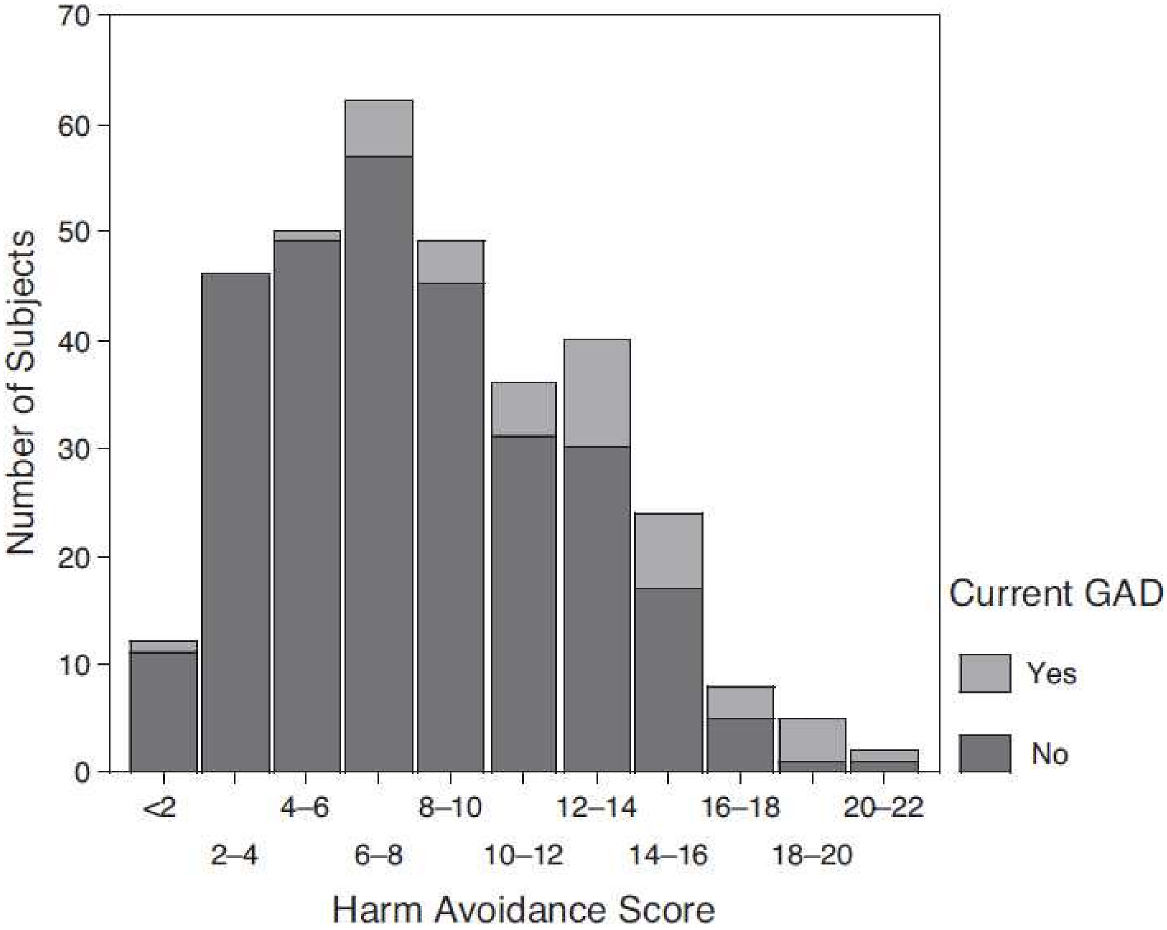

FIGURE 38–2. Distribution of harm avoidance (HA) scores in children and adolescence. Subjects whose results are given also have generalized anxiety disorder (GAD), according to the criteria of the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). The pattern reveals substantial but not complete overlap between harm avoidance and GAD, with a few subjects with high HA without GAD, some individuals with moderate HA with GAD, and very few individuals with low HA with GAD. (Reprinted from Rettew DC, Doyle AC, Kwan M, Stanger C, Hudziak JJ. Exploring the boundary between temperament and generalized anxiety disorder: a receiver operating characteristic analysis. *J Anxiety Disord.* 2006;20(7):931–945. doi:10.1016/j.janxdis.2006.02.002.)

Other Disorders

Some types of disorders are defined by behaviors that are conceptually more distinct from core temperamental traits and thus do not lend themselves as easily to spectrum hypotheses. Discrete actions such as binging and purging in eating disorders, hallucinations in psychotic disorders, and the taking of illicit drugs in substance use disorders may fall into this area. Many of these entities also tend to have onsets in adolescence or later and thus have most of the research directed at personality traits rather than temperament per se.

In substance use disorders, the choice of substances that is used, each with unique physiologic effects, might be partially related to an individual's unique temperamental profiles. Different agents also vary in their legality, which itself could lead to unique associations with particular traits. Alcohol use, for example, is legal for adults but not children and thus may have different temperamental associations by age. One follow-up study of temperament found indications that adult alcohol abuse was associated with being both temperamentally undercontrolled and overcontrolled as a child. Novelty or sensation seeking, especially when defined apart from positive emotionality, has also been associated with increased substance use. Once again, dimensions such as effortful control are related to lower amounts of substance use disorders. An interesting caveat in personality research in substance use acknowledges the shifting cultural perceptions of particular substances over time. Cigarette use, for example, has become viewed as increasingly deviant over the last 50 years across many segments of Western culture. One might expect, consequently, that its use is less associated with traits such as agreeableness over time or compared to cultures in which its acceptance level is greater.

Whereas most individuals with eating disorders regardless of specific diagnosis share a preoccupation with weight and self-image, there also appear to be substantial differences that generally fall along diagnostic lines, particularly between the restricting subtype of anorexia nervosa (AN) and bulimia nervosa (BN), and these may reflect underlying differences in temperament and personality. In particular, low effortful control and persistence has been found with BN but not in AN, where effortful control can be higher than controls. High neuroticism or harm avoidance has also been found for both disorders and parallels the frequent finding of anxiety disorders that can co-occur or precede eating disorders.

Temperamental associations with psychotic disorders have been less studied, although parallels especially with the negative symptoms of schizophrenia would appear to be fertile ground for investigation. High harm avoidance has been found in both patients with schizophrenia and their relatives, whereas high novelty seeking may be associated with more violent behavior. While positive symptoms such as hallucinations might be assumed to be relatively rare in the population and be present in more of an all-or-none form, recent research has suggested that lower-level forms of psychotic thinking are more prevalent than previously thought.

A Case Example

Previous “diathesis–stress” models of psychopathology that include temperament often frame temperamental traits as the diathesis component, which is then acted upon by various stressors, such as parenting behavior, traumatic events, and other factors to produce psychiatric symptoms. While data described in this chapter reveal that such a model is likely overly simplistic when it comes to how temperament is conceptualized, certain aspects of this model can still be useful. One of the principal advances from older conceptualizations of the diathesis–stress model is the full appreciation that the stressors often do not act on the diathesis randomly but rather are probabilistically more likely to occur because of those predispositions. As previously discussed in the section on temperament development, these processes have been called genotype–environment correlations and are readily applicable to the understanding of how risk factors like certain temperamental traits contribute toward the creation of environments such as parental behaviors and peer groups that accentuate those initial predispositions. A case example may illustrate this mechanism best.

Kayla was a child who from birth appeared to be more fearful and easily upset than other infants. She cried frequently and forcefully, and as a toddler, she was more reluctant to leave her mother's side to play with other children and to explore new situations. Her behavior was a challenge to her mother, who herself struggled with significant anxiety and shyness. As a result, her mother was not prone to seek out activities such as play dates with other parents, and when she did, Kayla's behavior made them frequently go badly. Hoping to avoid another embarrassing scene, Kayla's mother became increasingly overprotective in her efforts to shield her daughter from situations that might provoke discomfort. By the time kindergarten started, Kayla had had significantly fewer opportunities to interact with other children and tended to keep to herself. Throughout elementary school, she avoided team sports and other structured social activities, preferring other activities such as learning to play the piano. She did manage to make some lasting friendships, although they tended to be with other children who were also quiet and reserved. Kayla became an easy target of teasing by others, and as an adolescent, she became increasingly self-critical about her demeanor. At the age of 14 years, she saw a psychiatrist for the first time and was diagnosed with social anxiety disorder and major depression.

This illustration demonstrates how insidious the progression can be from temperamental traits to psychiatric disorders and how difficult it can be to ascribe the causation of symptoms as purely genetic or environmental. The case also illustrates the role of other factors in shaping a temperamental trajectory toward more maladaptive outcomes.

OTHER CONCEPTS IN TEMPERAMENT–PSYCHOPATHOLOGY RELATIONS

Parenting Behavior

The fact that some parenting behavior is evoked by children does not mean that parenting is either unimportant or inevitable. Neither does it mean that suboptimal parenting is the fault of the children. Rather, a more transactional view appreciates the bidirectional influences of children and parents that together can create important home environments with important developmental effects. Research has shown, for example, that parents of anxious children are more likely to teach avoidant rather than prosocial solutions to problems and may facilitate the perception of threat in ambiguous situations. As shown in the case example, these parents can become overprotective, which limits the crucial exposure opportunities that can help children to overcome their anxieties. Lack of parental warmth, poor supervision, harsh and inconsistent discipline, and obviously abuse are all important parental behaviors that have been found to be related to various types of psychopathology and serve as critical intermediaries between temperament and psychopathology. An important meta-analysis from 2016 demonstrated the somewhat counterintuitive finding that children with “difficult” temperaments were more sensitive to parenting influences “for better or for worse,” meaning that they benefited more from positive parenting and were more vulnerable to negative parenting than those with other temperament profiles.

Cognitive Factors

Although some cognitively-related attributes, such as attentional skills, are often included under the temperament umbrella, other aspects, such as intelligence, processing speed, and memory, are generally considered outside of the temperament domain. Nevertheless, these cognitive processes likely also play important moderating roles. Children with relatively diminished attention spans but above-average intelligence, for example, may be able to use that intellectual ability to counteract their attention deficits and prevent the clear impairment required for a diagnosis of ADHD. Other cognitive functions, such as the attributional style that a person applies to interpret themselves and the world around them, also likely shape and are shaped by temperament. These types of appraisals and attributions have been found to be important in a variety of both internalizing and externalizing disorders.

Trauma and Adverse Events

Clearly, both single-episode traumas, such as natural disasters and car accidents, and more chronic stressors, such as forms of abuse and poverty, can propel temperamentally at-risk children toward full-blown psychiatric symptoms. At the same time, it is becoming increasingly clear that one of the strongest predictors of more severe symptoms after a stressful event is poorer functioning prior to that event. Thus, a temperamentally aggressive child might be more likely to develop antisocial behaviors in the face of significant stressors, and a more anxious child might develop posttraumatic stress disorder after a single traumatic event.

Peer Groups and Activities

A powerful behavioral influence, especially for older children and adolescents, is a child's peers and the shared experiences that are done together. Similar to parenting behaviors, it is very likely that peer groups have a bidirectional influence on temperament and personality traits. A high-novelty–seeking child, for example, is likely to have similar-minded friends who may push each other to ever riskier and more dangerous activities. These group experiences provide opportunities both to magnify more maladaptive behaviors and to channel temperamental tendencies productively. Research has shown, for example, that adolescents on sports teams are less likely to smoke cigarettes. Given that participation in sports teams appears to have its strongest environmental influence early in life, a possible intervention target emerges for future research.

Temperament and Resiliency

Just as some temperamental traits can mix with the aforementioned variables to generate negative momentum toward psychopathology, other traits can serve as protective factors in the face of adversity and, indeed, may be important not only for a lack of psychopathology but also for true mental health and wellness. The previously mentioned study using latent profiles analysis found that compared to a group of children with moderate levels of most temperament traits, the "steady" group, characterized by high levels of persistence and low levels of novelty seeking, performed better at school by parent and teacher report, engaged in more activities, and functioned better socially. Just like at-risk traits, these dimensions are likely operating with a number of other factors through the same mechanisms of gene–environment correlations and interactions. The Project Competence study, which followed children into adulthood, found robust predictions between child temperament/personality and adult functioning even after controlling for a number of baseline and demographic characteristics.

Goodness of Fit

Although it is easy to slip into the notion that some temperament traits are good and others are bad, it is important to remember that all traits possess both positive and negative qualities. High novelty seeking, for example, generally contains positive elements of being energetic and spontaneous in addition to being prone toward extravagance and impulsivity. Thomas and Chess popularized the notion of goodness of fit decades ago, and it remains a cornerstone of child development theory. Simply put, the idea behind goodness of fit is that particular temperamental traits are not positive or negative per se but become so in an environmental context. A fearful temperament, for example, might be very adaptive in more dangerous settings but becomes less useful when the environment is safer and more predictable.

Research in animals and humans has supported this important hypothesis, although not necessarily at extreme levels. In a study by Rettew and colleagues, certain child temperament traits were related to behavior problems only in the presence of certain parental traits (Fig. 38–3). High child novelty seeking, for example, showed no relation to attention problems in families with a low-novelty–seeking mother but did show a high degree of attention problems when paired with a high-novelty–seeking mother.

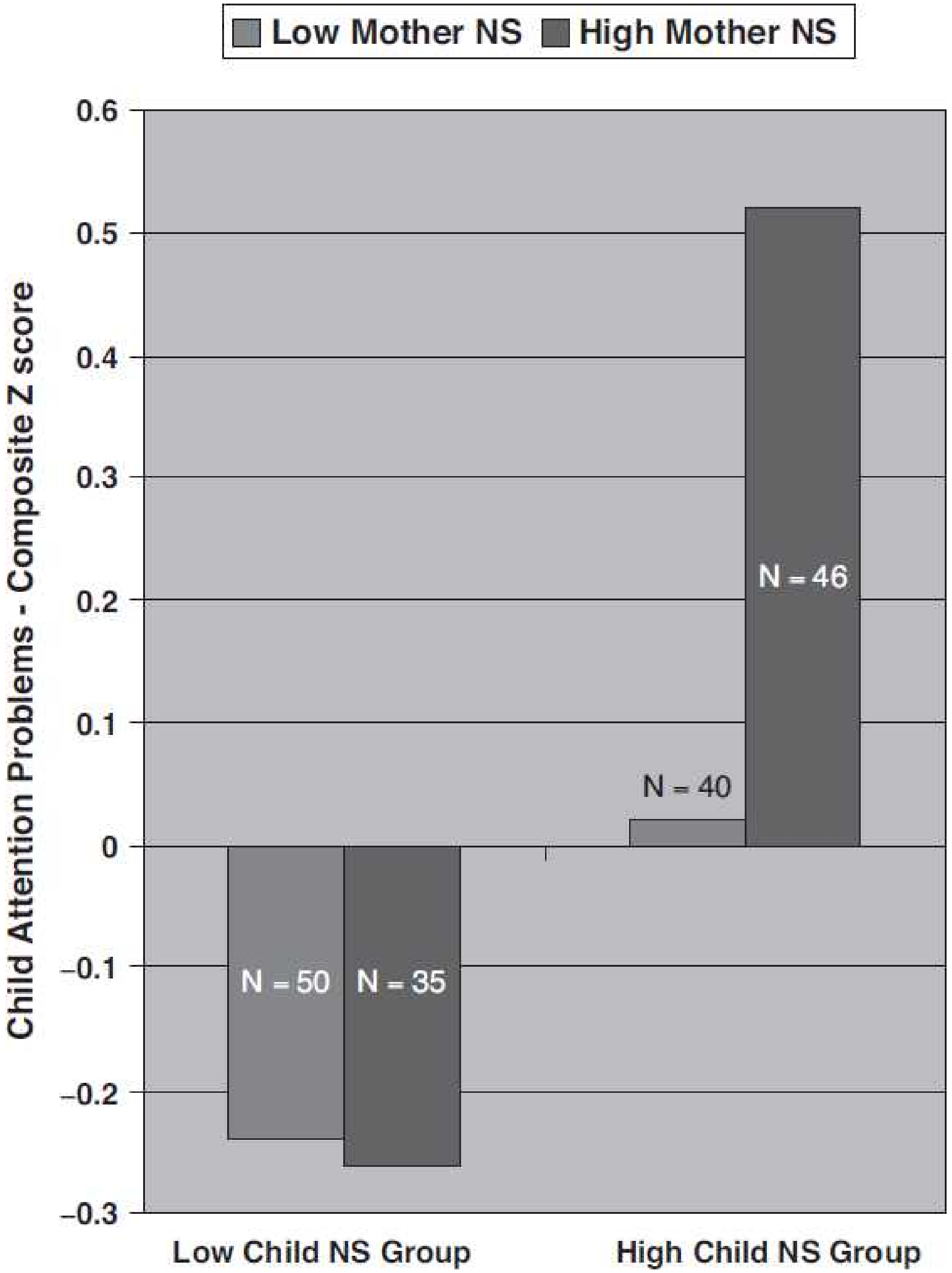

FIGURE 38–3. Relations between child and mother novelty seeking (NS) and attention problems. Children with low NS show few attention problems regardless of maternal NS. In high-NS children, however, a higher degree of attention problems is found in families with a high-NS mother but not in families with a low-NS mother. Results support the construct of goodness of fit in temperament/psychopathology associations. (Reprinted from Rettew DC, Stanger C, McKee L, Doyle A, Hudziak JJ. Interactions between child and

parent temperament and child behavior problems. *Compr Psychiatry*. 2006;47(5):412–420. doi:10.1016/j.comppsych.2005.12.008.)

CLINICAL IMPLICATIONS

It is clear at this point that the benefits of improved understanding regarding precisely how temperamental predispositions push children toward or away from more pathologic outcomes have much more than academic implications and may guide major efforts in early identification and intervention. If negative cycles of genetic–environmental interplay leading to disabling symptoms can be identified, they potentially can be stopped, perhaps even reversed. At the same time, more positive cycles can be cultivated and encouraged. Returning to the case example may be worthwhile.

After diagnosing Kayla with major depression and social anxiety disorder, her psychiatrist recommended cognitive–behavioral therapy to help Kayla overcome her dysfunctional beliefs and provide motivation for her to learn and apply new skills. The clinician also recognized similar symptoms in Kayla's mother and recommended a psychiatric evaluation, which resulted in her being prescribed a selective serotonin reuptake inhibitor (SSRI) and receiving individual treatment. Over time, Kayla learned to identify and challenge many of the negative thoughts that surrounded her behavior. Through the use of incentives, she was able to take small risks in approaching others socially. Her therapist recommended that she use her musical skills to her advantage, such as using her piano playing as a means to get out into public and receive positive social rewards. Her initial gains propelled her to take additional risks, and she joined the school orchestra. Meanwhile, her mother found that the decrease in her own anxiety allowed her to implement the therapist's suggestions more easily. Although Kayla never became a social butterfly, her depressive symptoms resolved, and she graduated from high school with confidence in her abilities and comfort with who she was.

Clinically oriented scientists are working hard to harness the energy that causes temperamentally predisposed tendencies to snowball into full-blown psychiatric syndromes and create pathways that not only will lead out of pathology but also into wellness. Given the complexity of these interactions, it is exceedingly likely that the specific relations between a particular trait and a particular disorder will vary considerably among individuals. Nevertheless, there is a tremendous promise that further study of temperament will not only dramatically alter some of the fundamental definitions of psychiatric disorders but also illuminate mechanisms that can improve the lives of children.

FURTHER READINGS

Borghuis J, Denissen J, Oberski D, et al. Big five personality stability, change, and codevelopment across adolescence and early adulthood. *J Pers Soc Psychol*. 2017;113:641–657.

Brickell I, Burton C, Mota NR, Martin J. Insights into attention-deficit/hyperactivity disorder from recent genetic studies. *Psychol Med*. 2021;51(13):2274–2286.

Buss AH, Plomin R. *Temperament: Early Developing Personality Traits*. Lawrence Erlbaum;1984.

Clauss JA, Blackford JU. Behavioral inhibition and risk for developing social anxiety disorder: a meta-analytic study. *J Am Acad Chi Adolsc Psychiatry*. 2012;51:1066–1075.

Cloninger CR, Svrakic DM, Przybeck TR. A psychobiological model of temperament and character. *Arch Gen Psychiatry*. 1993;50:975–990.

Else-Quest NM, Hyde JS, Goldsmith HH, Van Hulle CA. Gender differences in temperament: a meta-analysis. *Psychol Bull*. 2006;132:33–72.

Gray JA. *The Psychology of Fear and Stress*. Cambridge University Press; 1987.

Hettema JM, Neale MC, Myers JM, Prescott CA, Kendler KS. A population-based twin study of the relationship between neuroticism and internalizing disorders. *Am J Psychiatry*. 2006;163:857–864.

Johnson W, Hicks BM, McGue M, Iacono WG. Most of the girls are alright, but some aren't: personality trajectory groups from ages 14 to 24 and some associations with outcome. *J Pers Soc Psychol*. 2007;93:266–284.

Kagan J. *Galen's Prophecy*. Westview Press; 1994.

Karalunas SL, Fair D, Musser ED, Aykes K, Iyer SP, Nigg JT. Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: Toward biologically based nosologic criteria. *JAMA Psychiatry*. 2014;71:1015–1024.

Lahey BB, Moore TM, Kaczkurkin AN, Zald DH. Hierarchical models of psychopathology: empirical support, implications, and remaining issues. *World Psychiatry*. 2021:20:57–63.

Larrson H, Anckarsater H, Råstam M, Chang Z, Lichtenstein P. Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: A quantitative genetic study of 8,500 twin pairs. *J Child Psychol Psychiat*. 2012;53:73–80.

McCrae RR, Costa PT Jr, Ostendorf F, et al. Nature over nurture: temperament, personality, and life span development. *J Pers Soc Psychol*. 2000;78:173–186.

Oldehinkel AJ, Hartman CA, De Winter AF, Veenstra R, Ormel J. Temperament profiles associated with internalizing and externalizing problems in preadolescence. *Dev Psychopathol*. 2004;16:421–440.

*Rettew DC. *Child Temperament: New Thinking About the Boundary Between Traits and Illness*. W. W. Norton & Company; 2013.

Rettew DC, Althoff RR, Dumenci L, Ayer L, Hudziak JJ. Latent profiles of temperament and their relations to psychopathology and wellness. *J Am Acad Child Adolesc Psychiatry*. 2008;47(3):273–281.

Rettew DC, Doyle AC, Kwan M, Stanger C, Hudziak JJ. Exploring the boundary between temperament and generalized anxiety disorder: a receiver operating characteristic analysis. *J Anxiety Disord*. 2006;20:931–945.

Rettew DC, Stanger C, McKee L, Doyle A, Hudziak JJ. Interactions between child and parent temperament and child behavior problems. *Compr Psychiatry*. 2006;47:412–420.

Roberts BW, DelVecchio WF. The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychol Bull*. 2000;126:3–25.

*Rothbart MK, Ahadi SA, Evans DE. Temperament and personality: origins and outcomes. *J Pers Soc Psychol*. 2000;78:122–135.

*Rutter M, Dunn J, Plomin R, et al. Integrating nature and nurture: implications of person–environment correlations and interactions for developmental psychopathology. *Dev Psychopathol*. 1997;9:335–364.

Shiner R, Caspi A. Personality differences in childhood and adolescence: measurement, development, and consequences. *J Child Psychol Psychiatry*. 2003;44:2–32.

Slagt M, Dubas JS, Deković M, van Aken MAG. Differences in sensitivity to parenting depending on child temperament: a meta-analysis. *Psych Bull*. 2016;142:1068–1110.

Stergiakouli E, Martin J, Hamshere ML, et al. Shared genetic influences between attention-deficit/hyperactivity disorder (ADHD) traits in children and clinical ADHD. *J Am Acad Child Adolesc Psychiatry*. 2015;54:322–327.

Stoner R, Chow ML, Boyle MP, et al. Patches of disorganization in the neocortex of children with autism. *New Eng J Medicine*. 2014;370:1209–1219.

Stringaris A, Goodman R. Three dimensions of oppositionality in youth. *J Child Psychol Psychiatry*. 2009;50:216–223.

Sullivan SA, Kounali D, Cannon M, et al. A population-based cohort study examining the incidence and impact of psychotic experiences from childhood to adulthood, and prediction of psychotic disorder. *Am J Psychiatry*. 2020;177:308–317.

Sylvester DM, Barch DM, Harms MP, et al. Early childhood behavioral inhibition predicts cortical thickness in adulthood. *J Am Acad Child Adolesc Psychiatry*. 2016;55:122–129.

*Thomas A, Chess S. *Temperament and Development*. Bruner/Mazel; 1977.

Whittle S, Yucel M, Fornito A, et al. Neuroanatomical correlates of temperament in early adolescents. *J Am Acad Child Adolesc Psychiatry*. 2008;47:682–693.

Zwir I, Arnedo J, Del-Val C. Uncovering the complex genetics of human temperament. *Mol Psychiatry*. 2020;25:2275–2294.

39

Intellectual Disability

EVAN TANIGUCHI, M.D., SOO-JEONG KIM, M.D., AND KAREN E. TOTH, PH.D.

HISTORICAL OVERVIEW

The field of developmental disabilities has wrestled with how to capture the condition now referred to as “intellectual disability” or “intellectual developmental disorder,” and has worked hard to encourage “people first” language. Individuals should not be defined by their illnesses or conditions—for example, John is not a “leper,” or a “schizophrenic,” or “autistic.” Instead, people *have* illnesses or conditions—for example, John *has* schizophrenia. Sensitivity around labels such as “low functioning” and “high functioning” has also grown. Although these lexical and grammatical sensitivities may seem trivial for some, there is no faster way to lose credibility in some settings than to characterize one’s patient as “low functioning.” Why such heightened sensitivity in this field? While it is well recognized that intellectual gifts have been distributed on a continuum for as long as mankind has walked the earth, the understanding and treatment of persons with intellectual disability (ID) has moved like a pendulum between extremes. At one end, persons with ID have been exalted, and considered “Les enfants du Bon Dieu” (“children of the Good God”). Certain media may convey the message that persons with intellectual or developmental disability are somehow blessed with simpler, more straightforward understandings of basic human truths. And at the other end of the pendulum swing, such persons have also been vilified. The word “cretin” (congenital hypothyroidism), takes its origin from “Christian or Christlike,” even as its modern dictionary definition includes descriptors like “stupid, vulgar, and insensitive.” In the introduction to his text on “mental defectives” over a century ago, Barr (1904), the chief physician at the Pennsylvania training school for the feeble-minded, observed that:

> “The terms idiot and imbecile indiscriminately applied for centuries to a despised and neglected class, have naturally, through the mere force of association, become synonymous with and even expressive of opprobrium and reproach. So closely welded has been this association of word and idea, that even the advance of the nineteenth century with all its revelation of true conditions has failed to sunder them.” (p. 17)

Evidence for the recognition and treatment of ID dates to the earliest medical writings. Hippocrates described microcephaly and craniostenosis, and Galen actively explored the causes of cognitive disability. In the Middle Ages, Avicenna proposed treatments for meningitis and hydrocephalus and even defined levels of intellectual function. Esquirol (1838) is credited as the first medical writer to define the disorder of ID and was careful to distinguish ID from mental illness or disease. He focused on congenital onset in noting that:

> “Idiocy is not a disease, but a condition in which the intellectual faculties are never manifested; or have never been developed sufficiently to enable the idiot to acquire such an amount of knowledge as persons of his own age, and placed in similar circumstances with himself, are capable of receiving. Idiocy commences with life, or at that age which precedes the development of the intellectual and effective faculties, which are from the first what they are doomed to be during the whole period of existence.”

Modern history in the field of ID begins in the late 18th and early 19th centuries. At that time, Jean-Marc Itard attempted a natural experiment to educate Victor, a “wild-child” discovered in the forests of Aveyron, France. Although Itard himself judged his work with Victor a failure, this famous experiment marked the important consideration that persons with disabilities might be educable. Itard's efforts generated interest in educational and other interventions for persons with disabilities. In some ways, Itard and those that followed provided a bridge from beliefs that individuals with ID were subhuman (even suggested by Linnaeus to be a separate species in the genus Homo with names like Juvenis Lupinus Hessensis—a young man found in Hesse among wolves, or Juvenis Hannoverianus—a young man found in Hanover), to a more enlightened appreciation of their humanity and their potential.

Even so, in the preface to his text, Barr (1904) still straddles this divide as he categorically writes off the severely affected, while advocating for the less impaired:

> "In endeavoring to emphasize the utter hopelessness of cure, and also the needless waste of energy in attempting to teach an idiot, I have sought to make clear the possibilities that may be attained in the training of the imbecile, the urgent need of preventing the backward child from degenerating into imbecility, and of safeguarding the absolutely irresponsible amoral imbecile from crime and its penalty."

Howe and Wilbur are credited with introducing the enlightened approach to educating individuals with ID to the United States with the establishment of training schools like those that were appearing in Europe. By the turn of the century, there were 30 such institutional settings in 21 states, both public and private, and some reserved for women, some for individuals with epilepsy. Many of these institutions had residents numbering in the hundreds, and gradually the focus shifted away from education and more toward custodial care. During their peak usage, from 1950 until the late 1960s, institutions were home to approximately 1 per 1,000 Americans. Since the mid-1960s, many fewer persons with ID have been institutionalized in the United States. This move toward "deinstitutionalization" has arisen from many sources. First, the overcrowding and neglect common in many large institutions came to light in various exposés during the late 1960s. Blatt and Kaplan's photo story, "Christmas in Purgatory" (reproduced in Look magazine), Geraldo Rivera's television reports, and Robert F. Kennedy's attacks on the Rome and Willowbrook institutions in New York all shocked the nation.

Deinstitutionalization

Other forces also led to deinstitutionalization. Perhaps the most important of these forces was the philosophy of "normalization," the idea that individuals with ID were entitled to a more normal lifestyle, including a culturally normal rhythm to one's day (school or work, leisure time, sleep), week (weekdays and weekends), and year (vacations, holidays). In 1972, Wolfensberger extended the idea of normalization to the service delivery system itself, calling on all residences, schools, and other services for persons with IDs to be as normative as possible. Parent and professional advocacy groups also fought hard for legislative and legal victories to decrease the size of large institutions.

Over the past 60 years, changes in U.S. residential institutions have been dramatic. Several states have closed their institutions, and the remaining facilities have become much smaller. Typically, individuals remaining in large or even medium-sized institutions have the most severe impairments. Many of these residents also have severe behavioral problems or motor or sensory disabilities in addition to their severe-profound ID. Overall, from 1967 to 1994, the total institutional population—defined as state-operated institutions with 16 or more residents—decreased by almost two-thirds, from 194,650 to 65,818. In California, the number of persons served by the Department of Developmental Services increased 61.6% between January 1992 and December 2001. During this same period, however, the State developmental center population decreased by 44.4%. In many states, Medicaid funding waivers (the ability to access Medicaid dollars in circumstances in which family income criteria or other eligibility criteria are waived) are in place explicitly to support children with ID who meet "institutional level of care" standards. These so-called "deeming waivers," or "Katie Beckett waivers," are meant to ensure that Medicaid support can follow an individual out of an institutional setting if such a move can be made safely and cost-effectively. Katie Beckett was a young woman in Iowa who became ventilator-dependent following a catastrophic CNS infection and whose parents discovered that if they took her home from the skilled nursing facility, she would lose her Medicaid coverage. As the costs associated with institutionalization are invariably in the tens of thousands for any individual (ranging from a low of $104,025 per year in Arkansas to a high of $375,000 in Tennessee in 2009), the cost-effectiveness criterion is rarely a limiting factor, but clinicians may be called upon to attest to the level-of-care criterion for their patients who are seeking support under this waiver.

School. Schooling is the other major service for persons with ID. Following the move to "school" all American children that began in the mid-1800s, teachers and administrators soon discovered that a subset of children was having difficulties in performing school lessons. As a result, classes for "problem children" began in many cities and towns in the late 1800s. In 1894, the first classes specifically for children with ID were formed in Providence, Rhode Island. In rapid succession, special education classes were begun in Springfield, MA (1894), Chicago (1898), Boston (1899), New York (1901), Philadelphia (1901), and Los Angeles (1902).

Local school districts have historically held an ambivalent attitude toward special education classes. Often housed within the least desirable buildings and rooms, these classes were frequently furnished with scarce equipment and materials and had few specially trained teachers. By 1968, Lloyd Dunn declared that most children with ID could be "mainstreamed" in classes with same-age students, questioning the need for segregated special education classes for most children with ID. Dunn's article took note of the Supreme Court's ruling that "separate could not be equal" in educational settings, that research had not shown that children with mild ID learn better in special education versus regular education classes and that educational techniques had advanced sufficiently to allow for the effective schooling of most children with ID alongside other children.

In both education and living settings, then, the post-1970 era witnessed the strong influence of mainstreaming, community living, and normalization. Overall, such movements have proven beneficial as persons with ID increasingly take their rightful place within modern society. Many professionals do, however, question whether normalization has sometimes gone too far. Not every child with intellectual or other disabilities may be able to be schooled optimally with typically developing peers, nor might every adult be able to live independently in the community. Indeed, some families are advocating for specialized schools or residential settings for children with autism even as state agencies fear a return to institutional approaches. These issues continue to be hotly debated among professionals, families, and organizations dedicated to improving the lives of persons with ID.

Research

Like service delivery, the modern history of research also begins in the mid-19th century. In 1838, Esquirol differentiated ID from mental illness and proposed several levels of impairment. Later in the century, various classification systems were proposed. Probably the most well-known was a system of "ethnic classifications" proposed by J. Langdon Down in 1866—hence the term "Mongolism" for persons with Down syndrome (DS), the type of ID he first identified.

In addition to classifying persons with ID, the field of genetics plays an important historic role in IDs research. Unfortunately, early in this century, that role was mainly negative. Spurred by the rediscovery of Mendel's genetic findings, some researchers attempted to show that ID was inherited. Most notorious were Dugdale's (1877) study of the Jukes, and Goddard's (1913) study of the Kallikaks. In both cases, multi-generational families were used to argue that ID was inherited, and that sterilization was necessary for the "preservation of the race." Such eugenic scares led to court decrees and sterilization laws in California and other states during the 1920s.

Also playing an important role in ID research was the new science of psychological testing. Henry Goddard, the research director at the Vineland Training School in New Jersey, was the first to use Binet and Simon's IQ tests in the United States. After testing Vineland residents over a several-year period, Goddard concluded that "the vast majority of feeble-minded children are not changing and are not improving in their intelligence levels," a finding that another leading worker, Walter Fernald, called "the most significant … and the most discouraging that we have ever known."

Since early in this century, research has progressed in both the behavioral and biomedical fields. Behaviorally, recent work has allowed for better diagnoses and classification. Nearly a century ago, psychologists invented tests of motor, nonverbal intelligence, achievement, adaptive behavior, and other skills. Following similar work with typically developing children, psychologists have learned much about the development of cognitive, linguistic, social, and adaptive skills in persons with ID. In addition, many studies now examine the presence of psychiatric disorders in children and adults with ID, and families, schools, group homes, and workshops have all received research attention. Much of this work has occurred from the 1960s until today, a time of strong, often federally supported, progress in ID behavioral research.

But arguably the most important progress has occurred in the biomedical and genetic fields. As early as 1934, the Norwegian physician, Folling, hypothesized that several of his patients with ID were unable to metabolize phenylalanine. Such insights later led to a dietary treatment of this inborn error of metabolism; the phenylalanine-free diet, though not perfect, has significantly limited PKU as a cause of ID in most industrialized countries. In addition, PKU can now be screened at birth through the Guthrie test, after which dietary treatments can be instituted. PKU remains one of the major success stories in ID research even as newer treatments for this condition are explored.

Modern genetics constitutes another success story in ID research. In 1959, Lejeune, Gautier, and Turpin discovered that most cases of DS were due to trisomy of the 21st chromosome. Fragile X syndrome (FXS), the second most common genetic cause of ID, was discovered in 1969, and its cytogenetic diagnosis was refined in the mid-1970s. With the banding techniques of the 1970s and the molecular genetic techniques of the past three decades, exact causes are being described for many of the nearly 1,500 different genetic disorders associated with ID. The genetic—and, increasingly, the behavioral—characteristics of many of these disorders are now known (see below).

Organizations

The field of ID is also a history of organizations. The first such organization was the Association of Medical Officers of American Institutions for Idiotic and Feeble-Minded Persons, which began in 1876. As the first professional group devoted exclusively to persons with ID, the Association of Medical Officers was the forerunner to the American Association on Intellectual and Developmental Disabilities (AAIDD, formerly AAMR), the field’s most influential organization. Through its various journals and committees, AAIDD has been at the forefront of research, policy, and legislative advances for persons with ID. In addition to AAIDD, other groups have played important roles, particularly concerning changes in federal policy toward people with disabilities. The Council for Exceptional Children (CEC) was begun in 1922 and continues to champion the profession of special education and the education of all children with disabilities. The Arc, formerly the National Association for Retarded Citizens (NARC), is the main parent-led advocacy group. The Arc was particularly influential in the passage of the Education for All Handicapped Children Act of 1975 (PL 94–142), the federal law that, for the first time, mandated the “free, appropriate public education” of all children with disabilities throughout the United States. Each of these groups was also influential in the passage of the Americans with Disabilities Act (ADA), federal legislation that took effect in 1992 and outlawed discrimination against persons with disabilities. The National Association for the Dually Diagnosed, now simply NADD, was established in 1983, specifically in recognition of the lack of attention by the professional community to the needs of individuals with ID who also experience psychiatric illness. Over the ensuing years, the organization has grown to be the principal association in North America to advance mental wellness for persons with developmental disabilities through the promotion of excellence in mental health care.

Federal agencies constitute an additional category of influential organizations. Begun during the presidency of John F. Kennedy (who had a sister with ID), the National Institutes of Child Health and Human Development (NICHD) have long supported disability research. Other federal programs, including the nationally distributed network of nearly 70 University Centers for Excellence in Developmental Disabilities Education, Research, and Service (UCEDDs), and NICHD-funded Mental Retardation Research Centers (MRRCs), have also played important roles in research and policy. Taken together, professional, parent, and federal organizations provide an important infrastructure to the ID field.

Before discussing definitional issues, it is worth revisiting terminology. In different countries throughout the world, different terms are used for ID. In Europe, the term “intellectual disabilities” has long been used, and a major international journal is the *Journal of Intellectual Disabilities Research*. In Great Britain, alongside “intellectual disability,” the term “learning disability” (or learning disabilities) is often used—a term with a very different meaning in the United States.

DEFINITION OF INTELLECTUAL DISABILITY

Over the years, the American Psychiatric Association (APA) has adopted definitions of ID from the AAIDD, previously the American Association on Mental Retardation (AAMR). While the definitions in the third and fourth editions of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM; APA) are very similar, DSM-5 (APA, 2013) includes substantive changes to the definition of ID. For the first time in DSM, the term “intellectual disability” replaces “mental retardation,” and the language of DSM-5 is more closely aligned with that of the AAIDD. (The term proposed for ICD-11, to come into effect in 2022, is “intellectual developmental disorders.”) Further, severity levels are now classified by adaptive rather than intellectual function or IQ scores. This section reviews the changes (see Table 39–1 for summary).

Intellectual and Adaptive Functioning

Intellectual Functioning. Standardized intelligence tests that meet appropriate psychometric criteria for reliability and validity are commonly used to identify an ID. These instruments are summarized in Table 39–2, along with each test's age range and cognitive domains.

Both in terms of the testing situation itself and in the choice of an appropriate measure, administering such tests to people with ID is quite challenging. As people with ID have increased risks of comorbid psychiatric or behavioral disorders, examiners need to ensure that difficulties such as hyperactivity or poor frustration tolerance do not impede optimal test performance. Even in persons without behavioral challenges, certain personality characteristics may interfere with testing. Many people with ID, for example, look to others for solutions to difficult problems, and are quick to acquiesce, or become easily discouraged by failure. To minimize these problems, examiners can begin with easier tasks to provide success and incentives or rewards for effort.

For individuals who are nonverbal, there are several measures that can be selected that do not rely on expressive language (see Table 39–2). People with certain genetic syndromes (e.g., Williams syndrome, DS) also present unique test challenges, as many show syndrome-specific profiles of cognitive strength and weakness that are not easily described by a global IQ score (see phenotype section). Persons with ID and co-occurring sensory or motor deficits also require adaptation to routine test procedures. Finally, when testing individuals from some minority groups, examiners need to be sensitive to language and cultural issues, and use a multi-method test approach, especially when diagnosing persons with mild ID.

Adaptive Functioning. Edgar Doll (1935) was the first to develop a formal definition and measure of adaptive behavior. Two decades later, the AAMR (now AAIDD) officially included deficits in adaptive behavior in its definition of ID. Since that time, deficits in adaptive behavior have been formally included in all definitions of ID. However, until DSM-5, IQ score alone was used to classify individuals in terms of *severity* of ID. Table 39–3 illustrates the substantive nature of the changes in severity classification from DSM-IV to DSM-5.

Table 39–1.

A Comparison of DSM-IV and DSM-5 Diagnostic Criteria for Intellectual Disability

DSM-IV Diagnostic Criteria

A. Intellectual Function	B. Adaptive Function	C. Onset
Significantly subaverage intellectual functioning: an IQ of approximately 70 or below on an individually administered intelligence test	Concurrent deficits or impairment in present adaptive functioning (i.e., the person’s effectiveness in meeting the standards expected for their age or cultural group) in at least two of the following skill areas: communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health, and safety	The onset is before age 18 years.

DSM-5 Diagnostic Criteria

A. Intellectual Function	B. Adaptive Function	C. Onset
Deficits in intellectual function, such as reasoning, problem solving, planning, abstract thinking, judgment, academic	Deficits in adaptive functioning that result in failure to meet developmental and socio-cultural standards for	Onset of intellectual and adaptive deficits during the developmental period

learning, and learning from experience, confirmed by both clinical assessment and individualized, standardized intelligence testing	personal independence and social responsibility. Without ongoing support, the adaptive deficits limit functioning in one or more activities of daily life, such as communication, social participation, and independent living across multiple environments, such as home, school, work, and community.

Data from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. American Psychiatric Association; 1994 and American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

Though meanings vary, adaptive behavior is typically viewed as the performance of behaviors required for social and personal sufficiency. Further, adaptive behavior is an inherently developmental and social construct. Adaptive behavior changes as children grow into adolescence and adulthood and demands for social adaptation are also defined by expectations from others—from one’s family, society, and culture. Adaptive skills typically change across various settings. One’s adaptive performance on the job or at school may differ from one’s adaptive functioning with friends or at home. Measurements of adaptive behavior, then, need to have a developmental orientation, to be socially and culturally sensitive, and to represent the many settings in which people live, work, and play.

In contrast to Doll’s time, many instruments now exist that measure adaptive behavior across multiple domains (e.g., community skills, self-care, and home living). Table 39–4 summarizes commonly used measures of adaptive behavior, their age ranges, and the domains assessed. Although most of these measures are administered as interviews with parents or other care providers, some measures are administered directly to persons with ID. Central to all measures of adaptive function is the idea that adaptive behavior is measured by performance as opposed to ability. If people with ID are able to (possess the skills to) perform certain behaviors, but do not regularly do so or do not do so without prompts, reminders, and/or assistance, then they necessarily have compromised adaptive functioning.

Table 39–2.
Measures for Assessing Intellectual Ability

Measure	Age Range (years)	Domains Assessed
Wechsler Preschool and Primary Scale of Intelligence, 4th ed. (WPPSI-IV)	2.6 to 7–3	Verbal IQ, performance IQ, full scale IQ
Wechsler Intelligence Test for Children, 5th ed. (WISC-V)	6 to 16–11	Verbal IQ, performance IQ, full scale IQ
Wechsler Adult Intelligence Scale, 4th ed. (WAIS-IV)	16–90	Verbal IQ, performance IQ, full scale IQ
Stanford–Binet Intelligence Scales, 5th ed.	2 to 85+	Verbal, quantitative, abstract/visual, short-term memory, composite score
Kaufman Assessment Battery for Children, 2nd ed., NU (KABC-II-NU)	3–18	Sequential and simultaneous processing, mental processing composite
Kaufman Adolescent and Adult Intelligence Test (KAIT)	11 to 85+	Crystallized and fluid scales, composite IQ
Differential Ability Scales, 2nd ed. (DAS-II)	2–6 to 17–11	Verbal, nonverbal reasoning, spatial abilities, general conceptual ability
Raven’s Coloured Progressive Matrices (CPM)	5 to older adults	Figural reasoning; group test
Test of Nonverbal Intelligence, 4th ed. (TONI-4)	6–0 to 89–11	Intelligence, aptitude, abstract reasoning, and problem solving

Leiter International Performance Scale, 3rd ed. (Leiter-3)	3 to 75+	Game-like tasks assessing cognitive, attentional, and neuropsychological abilities
Universal Nonverbal Intelligence Test, 2nd ed. (UNIT-2)	5–0 to 21–11	Reasoning, memory, symbolic, and nonsymbolic tasks

Severity of Intellectual Disability

There are remarkably variable strengths and competencies of people with ID. DSM-5 (similar to AAIDD 2002) provides a classification system based on adaptive functioning across three domains: conceptual, social, and practical. Table 39–3 summarizes the classification system in DSM-IV (by IQ score) and the DSM-5 system based on characteristics across the three domains and associated with mild, moderate, severe, and profound ID. As people with ID can also be classified by the etiology of their conditions, later sections detail how adaptive functioning changes across diverse genetic and other etiologic groups.

Table 39–3.

DSM-IV Severity Criteria for Intellectual Disability (APA, 1994)

Severity Criteria				
Mild	**Moderate**	**Severe**	**Profound**	**Unspecified**
IQ level 50–55 to approximately 70	IQ level 35–40 to 50–55	IQ level 20–25 to 35–40	IQ level below 20–25	When there is a strong presumption of ID but the person's intelligence is untestable by standard tests.

DSM-5 Severity Criteria for Intellectual Disability (APA, 2013)

Severity Criteria	Conceptual Domain	Social Domain	Practical Domain
Mild	For preschool children, there may be no obvious conceptual differences. For school-age children and adults, there are difficulties in learning academic skills involving reading, writing, arithmetic, time, or money, with support needed in one or more areas to meet age-related expectations. In adults, abstract thinking, executive function (i.e., planning, strategizing, priority setting, and cognitive flexibility), and short-term memory, as well as functional use of academic skills (e.g., reading, money management), are impaired. There is a somewhat concrete approach to problems and solutions compared with age-mates.	Compared with typically developing age-mates, the individual is immature in social interactions. For example, there may be difficulty in accurately perceiving peers' social cues. Communication, conversation, and language are more concrete or immature than expected for age. There may be difficulties regulating emotion and behavior in age-appropriate fashion; these difficulties are noticed by peers in social situations. There is limited understanding of risk in social situations; social judgment is immature for age, and the person is at risk of being manipulated by others (gullibility).	The individual may function age-appropriately in personal care. Individuals need some support with complex daily living tasks in comparison to peers. In adulthood, supports typically involve grocery shopping, transportation, home and childcare organizing, nutritious food preparation, and banking and money management. Recreational skills resemble those of age-mates, although judgment related to well-being and organization around recreation requires support. In adulthood, competitive employment is often seen in jobs that do not emphasize conceptual skills. Individuals generally need support to make health care–related and legal decisions, and to learn to perform a skilled vocation competently. Support is typically needed to raise a family.
Moderate	All through development, the individual's conceptual skills lag markedly behind those of peers. For preschoolers, language and pre-academic skills develop slowly. For school-age children, progress in reading, writing, mathematics, and understanding of time and money occurs slowly across the school years and is markedly limited compared with that of peers. For adults, academic skill development is typically at an elementary level, and support is required for all use of academic skills in work and personal life. Ongoing assistance on a daily basis is needed to complete conceptual tasks of day-to-day life, and others may take over these responsibilities fully for the individual.	The individual shows marked differences from peers in social and communicative behavior across development. Spoken language is typically a primary tool for social communication but is much less complex than that of peers. Capacity for relationships is evident in ties to family and friends, and the individual may have successful friendships and sometimes romantic relations in adulthood. However, individuals may not perceive or interpret social cues accurately. Social judgment and decision-making abilities are limited, and caretakers must assist with life decisions. Friendships with typically developing peers are often affected by communication or social limitations. Significant social and communicative support is needed in work settings for success.	The individual can care for personal needs involving eating, dressing, elimination, and hygiene as an adult, although an extended period of teaching and time is needed for the individual to become independent in these areas, and reminders may be needed. Similarly, participation in all household tasks can be achieved by adulthood, although extended teaching and ongoing supports are needed for adult-level performance. Independent employment in jobs that require limited conceptual and communication skills can be achieved, but considerable support from coworkers and others is needed to manage social expectations, job complexities, and responsibilities such as scheduling, transportation, health benefits, and money management. A variety of recreational skills can be developed but require supports and learning opportunities over an extended period. Maladaptive behavior is present in a significant minority and causes social problems.
Severe	Attainment of conceptual skills is limited. The individual generally has little understanding of written language or of concepts involving numbers, quantity, time, and money. Caretakers provide extensive supports for problem solving throughout life.	Spoken language is quite limited in vocabulary and grammar. Speech may be single words or phrases and may be supplemented through augmentative means. Communication is focused on the here and now. Language is used for social communication more than explication. Individuals understand simple speech/gestures. Relationships with familiar others are a source of pleasure and help.	The individual requires support for all activities of daily living (meals, dressing, bathing, and elimination), requires supervision at all times, and cannot make responsible decisions regarding well-being of self or others. In adulthood, tasks at home, recreation, and work require ongoing support/assistance. Skill acquisition in all domains involves long-term teaching and support. Maladaptive behavior, including self-injury, is present in a significant minority.
Profound	Conceptual skills generally involve the physical world rather than symbolic processes. The individual may use objects in goal-directed fashion for self-care, work, and recreation. Certain visuospatial skills, such as matching and sorting based on physical characteristics, may be acquired. However, motor and sensory impairments may prevent functional use of objects.	The individual has very limited understanding of symbolic communication in speech or gesture. They may understand some simple instructions or gestures. The individual expresses their own desires and emotions largely through nonverbal, nonsymbolic communication. The individual enjoys relationships with familiar others, and initiates and responds to social interactions through gestural and emotional cues. Sensory and physical impairments may prevent many social activities.	The individual is dependent on others for all aspects of daily physical care, health, and safety, although he/she may participate in some of these activities as well. Individuals without severe physical impairments may assist with daily work tasks at home, like carrying dishes to the table. Simple actions with objects may be the basis of participation in some vocational activities with high levels of support. Recreational activities may involve music, movies, walks, or participating in water activities, all with support. Physical and sensory impairments are frequent barriers to home, recreational, and vocational activities. Maladaptive behavior is present in a significant minority.

Data from the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, DSM-4 (Copyright © 1994). American Psychiatric Association; Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-5-TR (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Table 39–4.
Common Measures of Adaptive Behavior

Measure	Age Range	Domains Tested
Vineland Adaptive Behavior Scales, 3rd ed. (Sparrow, Cicchetti, and Saulnier, 2016)	Birth–90: Interview and Parent/Caregiver Form; 3–21: Teacher Form	Communication: receptive, expressive, written; daily living skills: personal, domestic, community socialization: interpersonal relations, play and leisure, coping skills; motor skills: fine, gross
Adaptive Behavior Assessment System, 3rd ed. (ABAS-3); (Harrison and Oakland, 2015)	Birth to 89–11	Assesses conceptual, social, and practical aspects of adaptive behavior, as well as 11 skill areas within 3 major adaptive domains: Conceptual, Social, and Practical.
Scales of Independent Behavior, Revised (SIB-R) (Bruininks, Woodcock, Weatherman, and Hill, 1996)	Birth to 80+	Motor: fine, gross Social interaction and communication: social interaction, language comprehension, language expression Personal living skills: eating and meal preparation, toileting, dressing, community living skills: time and punctuality, money and value, work
Diagnostic Adaptive	4–21	Focuses on the critical

Behavior Scale (DABS) (AAIDD release 2015)		“cut-off area” for the purpose of ruling in or out a diagnosis of ID or related disability.
Street Survival Skills Questionnaire (SSSQ) (Linkenhoker and McCarron, 1993)	9–40+	Administered to person; basic concepts, functional signs, tools, health

Mild Intellectual Disability. Mild ID constitutes the largest group of people with ID, possibly as many as 85% of all persons with ID. These individuals appear similar to unaffected individuals and often blend into the general population in the years before and after formal schooling. Many achieve academic skills at the sixth-grade level or higher, and some graduate from high school. As adults, many of these individuals hold jobs, marry, and raise families—yet at times they may appear slow or need extra help negotiating life’s problems and tasks. Historically, people with mild ID were thought to show relatively few clear-cut organic causes for their delay. However, in recent years there has been an increase in the number of people with genetic syndromes who function in the mild range of ID. Examples include most people with Prader–Willi syndrome (PWS), as well as some males and most females with FXS .

A more striking finding is that more people with mild ID come from minority groups and low socio-economic (SES) backgrounds than would be expected from their percentages in the general population. This overrepresentation of minority groups has been used to criticize standardized intelligence tests that are used for placement in special education classroom settings, as well as to highlight the importance of environmental-cultural influences on intellectual development and assessment.

Moderate Intellectual Disability. Moderate ID is seen in approximately 10% of the population with ID and includes people with more impaired cognitive and adaptive functioning. People with moderate ID are typically diagnosed in their preschool years, and some show a clear organic cause for their delay. Many persons with DS, the most common chromosomal cause of ID, often function in this range, as do many adolescents and adults with FXS . Most children with moderate ID require special education services, achieving academic skills at the second to third grade level. Need for supportive services continues throughout the lifespan, and with proper supports, many live, work, and thrive in their local communities. In a study by Ross and colleagues, 20% of persons with IQs from 40 to 49 lived independently, whereas 60% were considered partially dependent and 20% totally dependent on others. In a similar way, some individuals in this range are employed in the competitive job market, and need minimal job supervision, whereas others require more extensive supervision on the job, and may work in sheltered workshops or other, more segregated settings.

Severe Intellectual Disability. Severe ID occurs in about 3% to 4% of the population of people with ID. Individuals at this level often have one or more organic causes for their delay, and many show concurrent motor, ambulatory, and neurologic problems, as well as poorly developed communication skills. Most persons with severe ID require close supervision and specialized care throughout their lives. Some individuals learn to perform simple tasks or routines that facilitate their self-care or their ability to perform in a sheltered workshop or pre-workshop type of setting.

Profound Intellectual Disability. Profound ID affects relatively few individuals (1% to 2%) with ID, and involves pervasive deficits in cognitive, motor, and communicative functioning. Impairments in sensory-motor functioning are often seen from early childhood on, and most individuals require extensive training to complete even the most rudimentary aspects of self-care such as eating and toileting. Most people with profound ID have identifiable organic causes for their delay, and most require total supervision and care throughout life. These four levels—mild, moderate, severe, and profound—have long been used to classify severity of ID yet have come under attack because of their focus on deficits instead of competencies. How people with ID are defined and classified—based on strengths, impairments, or some combination of both—has profound implications for studies on the epidemiology of this condition.

PREVALENCE

According to DSM-5, 1% of the population has an ID. Indeed, a meta-analysis of 52 peer-reviewed papers has estimated the global prevalence of ID to be 1,037 per 100,000 population. When stratified by severity, 3 to 4 in 1,000 have severe ID while 5.4 to 10.6 in 1,000 have mild ID (Martin et al., 2018). Rates are higher in males (perhaps due to greater risks of environmental exposures or X-linked conditions), generally low in the early years, peak around ages 10 to 14, and decrease in the late school years and adulthood. However, a recent epidemiologic study based on a national register in Sweden found that the prevalence of ID in older adults is increasing as a result of medical and social advances, and that approximately 0.5% of the population above 55 years of age has ID. Individuals of lower socioeconomic status and of ethnic minority groups show higher-than-expected rates of ID, likely secondary to lack of access to obstetric care, higher rates of environmental toxins, low access to healthy diet options and lower access to education (Thapar et al., 2015). When prevalence rates are adjusted for age, birth order, and economic status the risk associated with minority group status is reduced to eliminated. Genetic variability among ethnic groups is not suspected to be a major factor.

Etiology of Intellectual Disability

As noted by Esquirol in 1845, ID is not a disease in and of itself, but the developmental consequence of a variety of pathogenic processes. ID results from brain dysfunction, generally due to abnormal brain development or brain injury resulting from genetic and environmental causes (Table 39–5). Many genetic causes of ID have been known for decades, while modern genetic techniques have identified hundreds of new causes of ID.

Genetic causes of ID can result in multiple malformation conditions (e.g., DS) or what appears to be “non-syndromic” ID. That said, most genetic causes of ID are associated with characteristic clinical or neurobehavioral features, if the patients are examined thoroughly. Genetic disorders with ID display a wide variety of inheritance patterns and genetic mechanisms including chromosome rearrangements (often de novo), autosomal recessive, autosomal dominant (inherited and de novo), X-linked, triplet repeat expansion, and imprinting. The important role for de novo mutations has only recently been appreciated. Knowledge of these genetic mechanisms can aid with diagnosis and is essential for recurrence risk counseling. More and more, the genetic diagnosis influences treatments to implement (e.g., specific melatonin/propranolol treatment for sleep disruption in patients with Smith–Magenis syndrome) and treatments to avoid (e.g., valproate in ornithine transcarbamylase deficiency). In addition, some disorders are associated with health implications for family members (e.g., fragile X–associated tremor/ataxia syndrome [FXTAS] and premature ovarian failure in adult relatives of individuals with fragile X).

Table 39–5.
Etiologic Classification of Intellectual Disability

Etiology	Examples	Estimated Frequency (%)
A. Prenatal Causes		
1. Genetic disorders		4–28
Chromosomal aberrations	Down syndrome	
Monogenic mutations	Tuberous sclerosis; phenylketonuria and other metabolic disorders; fragile X syndrome	
Multifactorial	“Familial” mental retardation	
Malformation syndromes due to microdeletions	Prader–Willi, Williams, Angelman syndromes	
2. Congenital malformations		7–17
Malformations of the central nervous system	Neural tube defects	
Multiple malformation syndromes	Cornelia de Lange syndrome	
3. Exposure		5–13
Maternal infections	Congenital rubella, HIV	
Teratogens	Fetal alcohol syndrome	
Toxemia/placental insufficiency	Prematurity	
Other	Radiation, trauma	

B. Perinatal Causes		2–10
1. Infections	Meningitis	
2. Delivery problems	Asphyxia	
3. Other	Hyperbilirubinemia	
C. Postnatal Causes		3–12
1. Infections	Encephalitis	
2. Toxins	Lead poisoning	
3. Other postnatal causes	Traumas, brain tumors	
4. Psychosocial problems	Poverty, psychotic illness	
D. Unknown Causes		30–50

With advances in medicine generally and in molecular genetics in particular, new causes of ID, or the genetic etiologies of formerly unspecified syndromes, are identified each year. For example, a national database of genetic disorders, *Online Mendelian Inheritance in Man,* lists >2,500 genetic conditions associated with ID. The most common causes of ID include DS, FXS, prematurity, and fetal alcohol syndrome. Variability in patient cohorts and methods for defining etiologies make it difficult to estimate the proportion of patients accounted for by each cause. For example, persons drawn from institutional settings, where severe and profound ID and visible congenital anomalies are more likely to be found, more commonly have an identifiable chromosomal abnormality than do persons with mild disability residing in community settings.

Understanding the cause of ID in an individual person is essential for providing prognostic information, guiding treatment, and educating families about recurrence risk. For instance, distinguishing between a one-time injury such as hypoxia and a genetic neurodegenerative disorder such as a lysosomal storage disorder makes a huge difference in treatment, emotional impact on the family, how they plan their child’s life, and how they approach having additional children. A specific diagnosis can also reduce family stress by providing an "answer" for why their child has a significant disability and allowing them to connect with other families affected by the same condition. More and more, an accurate diagnosis will guide more appropriate treatments to maximize neurocognitive function, prevent decline, and direct specific medical screenings or treatments that optimize the patient's health and adaptive function.

In the era of rapidly expanding molecular genetic diagnosis, each provider must decide how to manage the explosion in genetic information and the impact on diagnosis and treatment based on their level of interest, educational background, and access to other subspecialists. At a minimum, psychiatric providers need to know when to suspect an underlying genetic condition and how to ensure adequate evaluation and management (through reference materials and/or consultants). For instance, ornithine transcarbamylase deficiency results from a defect in the urea cycle that can cause encephalopathy due to elevated ammonia levels. Patients can present at any age, and the best way to improve behavioral issues in a patient with undiagnosed or inadequately treated ornithine transcarbamylase deficiency, is to recognize the possibility of the condition, diagnose it using biochemical (plasma ammonia level) and genetic testing (*OTC* gene sequencing), and provide dietary and medication treatment to lower ammonia levels and maintain nutrition. Valproate, haloperidol, and other medications should also be avoided, since they have been associated with hyperammonemic crises in these patients. Indeed, the decompensation associated with administration of these medications in an undiagnosed patient can prompt workup and accurate diagnosis of ornithine transcarbamylase deficiency by the vigilant psychiatrist. Genetic diagnosis is now directly influencing treatment choices in a variety of diseases, most prominently cancer, and it is only a matter of time before gene-specific treatments become widespread in psychiatry.

Medical Diagnostic Approach to the Patient With Intellectual Disability

Clinical History and Examination. The elements of a comprehensive clinical evaluation for the patient with ID are summarized in Table 39–6. As for any patient, a good history is essential, and in many cases collateral information is extremely useful. Family members may be able to fill in important aspects of the family pedigree, providing information about relatives ranging from learning problems to physical anomalies, developmental disabilities, and even response to treatment.

The clinical examination should be performed with particular attention to the presence of minor anomalies, aberrant growth, and physical development (see Table 39–6). Increasingly, the examination will look for behavioral symptoms that may comport with a known behavioral phenotype (see later discussion) and help in diagnosis.

Genetic Evaluation. Given the incredible advances in DNA sequencing technologies and other methods for identifying mutations across the entire human genome, clinical genetic testing in patients with ID is very much a moving target. Currently, patients without an obvious syndrome or extrinsic cause should be evaluated for chromosomal abnormalities and fragile X (by chromosomal microarray and fragile X repeat length testing, respectively), since these are the most common causes of “nonsyndromic” ID. Genetic disorders should still be considered in patients who do not follow the expected course for a presumed environmental cause. Unless a specific aneuploidy (such as trisomy 21) or a balanced translocation is suspected, a karyotype is no longer the appropriate test when chromosome microarray is available. Beyond these basic tests, expert clinical genetics consultation may be the most cost-effective next step, rather than extensive laboratory testing and imaging. Genetic consultation can help guide diagnostic testing as well as providing counseling about prognosis, surveillance for complications, and recurrence risk.

When a specific disorder is suspected, a single-gene test may be the best option; however, as genetic testing becomes more widely available, the spectrum of presentations associated with each gene has broadened, so it can be difficult to determine which single-gene test to perform. Fortuitously, DNA sequencing has become much more efficient, so panels of genes, or even most of the coding regions in the entire genome (the “exome”), can now be sequenced in a single test. As the cost of sequencing decreased and human sequence variation becomes better characterized, genome-wide sequencing methods will become first-line genetic tests, despite the challenges with interpreting the significance of many DNA variants. It is important to remember that currently available sequencing tests do not detect repeat expansions or imprinting disorders.

Table 39–6.

Elements of the Clinical Evaluation of the Patient with Intellectual Disability

Clinical History

Prenatal exposures (alcohol, infections, etc.)

Birth history

CNS infection, traumatic brain injury

Developmental regression

Focal neurologic findings, hypertonia, hypotonia, or ataxia

Extra-CNS malformations and medical issues

Visual impairment or hearing loss

Family Pedigree (Three-Generation)

Relatives with learning problems, psychiatric disorders, intellectual disability, abnormal brain imaging findings, neurologic or degenerative disorders

Consanguinity

Family resemblance (reduced between family members with and without some genetic conditions)

Physical Examination

Assessment of major and minor physical anomalies including facial features and skin findings

Growth and physical development

Head circumference compared to norms (congenital versus “acquired” microcephaly/macrocephaly)

Growth trajectory (comparison with earlier measures)

Description of facial features (micrognathia, hypertelorism, thin upper lip)

Complete neurologic examination

Skin examination for café-au-lait and other birthmarks

Documentation of behavioral phenotype

Adjunct Diagnostics

Hearing evaluation

Ophthalmology evaluation of eye structure, eye movements, and visual function

Psychometric assessments

Laboratory Testing (When Indicated)

Creatine kinase (to evaluate for muscular dystrophies)

7-dehydrocholesterol testing if Smith–Lemli–Opitz syndrome is suspected

Metabolic studies for lysosomal, peroxisomal, and mitochondrial disorders

Urine organic acids, plasma amino acids

Congenital disorders of glycosylation testing

Genetic Testing (When Indicated)

Chromosomal microarray

Fragile X testing

DNA sequencing (single gene, gene panel, whole exome)

DNA methylation testing if Prader–Willi or Angelman syndromes are suspected

Imaging

Structural brain magnetic resonance imaging (MRI)

Functional brain MRI (fMRI; still largely research-based)

Diffusion tensor brain MRI (DTI)

Other Studies

Muscle biopsy for suspected mitochondrial disease (or muscular dystrophy)

Neuroimaging. Neuroimaging is clinically indicated for individuals with ID in the presence of microcephaly or macrocephaly, abnormal head circumference, seizures, focal neurologic signs, and/or progressive cognitive impairment or loss of psychomotor skills. MRI is generally superior to CT unless acute hemorrhage or calcification are suspected. MRI has been useful in elucidating a diagnosis in around 5% of people with nonprogressive ID (Decobert et al., 2005). More commonly, in ~50% of patients, subtle and nonspecific structural findings indicating altered neurodevelopment are present such as ventriculomegaly, dysplasia of the corpus callosum, minor abnormalities of white matter signal, cortical gyration, and cerebellar vermis. The American Academies of Neurology and Pediatrics and the American College of Medical Genetics periodically publish updated practice guidelines for the use of brain imaging in the evaluation of patients with ID. Magnetic resonance spectroscopy is lower-yield and less commonly employed, although it can yield a diagnosis when a metabolic disorder is suspected despite otherwise normal genetic and metabolic work-up. Diffusion tensor imaging holds promise for identifying disorders associated with defects in major axon pathways that do not affect brain morphology; however, it is not yet in wide clinical use. The use of MRI technology to study brain structure and regional activation (fMRI) and white matter tracts (diffusion tensor imaging) remains a powerful potential tool in the study of ID and their underlying mechanisms. Information gathered from MRI research studies have shown neuroanatomical and functional effects of specific genetic disorders associated with mental retardation (Gothelf et al., 2005). However, large studies utilizing structural MRI or quantitative methods such as those performed in ADHD or autism have not been performed in ID, and research involving the use of functional MRI (fMRI) in ID is still emerging. Advanced imaging techniques may serve to better elucidate the neurobiology of ID, resulting in improved diagnostic precision and treatment monitoring in the future.

Electroencephalography. Electroencephalographic investigations, such as waking/sleep video-EEG-polygraphic studies may be useful in patients with a clinical history of seizures and those with severe language deficits and/or specific genetic syndromes such as Angelman syndrome. EEGs are often requested in patients with known cortical malformations associated with high seizure risk (e.g., polymicrogyria) when acute behavioral or cognitive changes have occurred.

Metabolic Testing

EVALUATION FOR INBORN ERRORS OF METABOLISM. The Inborn Errors of Metabolism (IEMs) or biochemical genetic conditions are a large and fast-growing group of disorders that are for the most part individually quite rare, but collectively affect 1% to 1.5% of the population. These disorders originate from disruptions to genes involved in biochemical and metabolic pathways, many of which are important in brain development and/or function. In general, symptoms and findings arise from either the accumulation of substances that prove toxic in excess or the inability to synthesize enough of a required product. Well-known IEMs causing ID include phenylketonuria (PKU), Lesch–Nyhan syndrome, and the lysosomal and glycogen storage diseases. It is important for the practicing clinician to retain the IEMs in the differential diagnosis, since many require specialized treatment to prevent morbidity, progression, and mortality, as well as to optimize development. Frequently, treatment involves supplementation with a deficient enzyme or metabolic intermediate with or without dietary restriction. It should be noted that each IEM may have variant ages of onset such as "infantile," "juvenile," or "adult" forms, with concomitantly different mixes of medical and neuropsychiatric symptoms. Later-onset forms may manifest in less medically severe findings and a presentation more dominated by neuropsychiatric symptoms. Specific features in each disorder vary, but the child psychiatrist should consider the IEMs in the differential diagnosis of every individual with ID.

Table 39–7.
Selected Clinical Findings and Laboratory Abnormalities Which Increase Suspicion for Underlying Metabolic Disorder

Growth abnormality
Recurrent, unexplained illness
Exacerbation of symptoms by physiologic stress
Seizures
Ataxia
Loss of psychomotor skills
Hypotonia
“Coarse” appearance
Eye abnormalities (cataracts, ophthalmoplegia, corneal clouding, retinal abnormality)
Recurrent somnolence/coma
Abnormal sexual differentiation
Arachnodactyly
Hepatosplenomegaly
Metabolic/lactic acidosis
Hyperuricemia
Hyperammonemia
Low cholesterol
Structural hair abnormalities
Unexplained deafness
Bone abnormalities (dysostosis; occipital horns; punctate calcifications)
Skin abnormalities (angiokeratoma; “orange-peel” skin, ichthyosis)
Brain imaging abnormalities cortical and/or cerebellar atrophy, evidence of stroke, abnormal white matter signal, lactate peak on MRS

Adapted from Curry CJ, Stevenson RE, Aughton D, et al. Evaluation of mental retardation: recommendations of a consensus conference. *Am J Med Genet.* 1997;72(4):468–477.

While guidelines published by the American Academies of Pediatrics and Neurology do not recommend routine metabolic testing for all persons with ID, several studies suggest that IEMs may account for up to approximately 2.5% of patients after screening by clinicians. More recent studies have shown that even after broad-based metabolic screening, a second line of testing identified IEMs in 14% of patients, many of which were treatable (Engbers et al., 2008). A careful history should be taken to make sure the patient received usual newborn screening tests (particularly in the case of immigrants or where the patient lacked usual perinatal care) and whether the testing followed current recommendations (especially in older individuals). IEMs should be particularly considered in the patient with ID who presents with catatonia, mania, seizures, encephalopathy, and/or visual hallucinations, when symptoms appear exacerbated by physiologic stress and illness, occur with abdominal pain or vomiting, or functional regression (Table 39–7).

The authors recommend the following screening labs for patients with ID when an IEM is suspected: complete blood count, complete metabolic panel (electrolytes, BUN, creatinine, glucose, transaminases), lactate, pyruvate, urine organic acids, plasma amino acids, ammonia, plasma acylcarnitines, and total and free carnitine. Testing during a symptom exacerbation is recommended, since assays may be normal during periods of relative metabolic stability. Referral to a specialist biochemical geneticist is also recommended for evaluation and/or management.

GENETIC INTELLECTUAL DISABILITY SYNDROMES AND BEHAVIORAL PHENOTYPES

Remarkable genetic advances have sparked renewed research interest in the behavioral features of persons with fragile X, Prader–Willi, and other genetic syndromes. In their reviews in the 1970s of behavioral aspects of chromosomal disorders, S. Kessler and R. Moos, for example, discussed only five syndromes. Since that time, the number of disorders for which a “behavioral phenotype” has been suggested has grown considerably. A good working definition of a “behavioral phenotype” is that proposed by Dykens, namely, that of a syndrome in which the probability of the expression of certain behaviors or constellations of behaviors is greater than what would otherwise be expected. Stereotyped hand movements, for example, have been described in many contexts, but a particular form of stereotyped hand wringing is almost invariably seen in Rett syndrome. Hyperphagia and obesity are nonspecific symptoms, but without aggressive dietary intervention, the probability of an individual with PWS becoming obese is nearly 100%.

Table 39–8 provides a representative sample of syndromes for which behavioral phenotypes have been described—some anecdotally, others with controlled studies. Although the study of behavioral phenotypes is only at its beginning stages, this line of work holds much promise for both clinicians and researchers. Identifying syndrome-specific behavioral patterns helps clinicians to zero in on some diagnostic possibilities over others, and these behavioral profiles also guide intervention and treatment. From a research standpoint, phenotypic work advances the understanding of gene–brain–behavior relationships. To illustrate these advantages, the following subsections review behavioral and genetic features for fragile X, Prader–Willi, Down, and Williams syndromes. Although not representative of all syndromes, these examples amply demonstrate the promise of phenotypic research for both intervention and science.

Examples of Behavioral Phenotypes in Intellectual Disability Syndromes

Fragile X Syndrome. FXS, the most common inherited cause of ID, results in a wide range of learning and behavioral problems, with males being more often and more severely affected than females. The fragile X gene (*FMR1*) represents a relatively newly identified type of human disease, caused by an amplification (or excessive repetition) of trinucleotide sequences (CGG) that make up DNA. Greater than 200 triplet repeats, people are fully affected with the syndrome. Numbers of repeats between 55 and 200 are termed “premutations.” One in 148 to 291 females, and one in 290 to 855 males may carry the premutation, with approximately 1 in 4,000 males and 1 in 8,000 females being fully affected with the syndrome.

Many of the behavioral features in people with FXS are a function of the number of CGG repeats and a cascade of subsequent effects. The number of repeats relates to the degree that the *FMR1* gene is methylated and subsequently silenced or blocked from transcription and translation. This, in turn, affects the amount of *FMR1* protein (FMRP) produced, and the range of clinical expression in FXS is associated with the amount of FMRP. Fully affected, fully methylated males have no FMRP, whereas high-functioning males, although rare, have some FMRP via CGG size and/or methylation mosaicism. In affected females, only one X chromosome carries the *FMR1* gene, whereas the other produces some FMRP. Because all females have one X chromosome that is randomly inactivated in all cells, the degree of involvement in females with fragile X is at least in part related to the ratio of cells that have affected X active to silenced (inactive). About half of fully affected females thus show mild to moderate ID, whereas half have average IQ but show learning difficulties and/or social affective impairments. Both males and females with the full syndrome exhibit similar neurocognitive profiles, including deficits in auditory short-term memory, processing of sequential information, and attention, with relative strengths in processing simultaneous information, long-term memory, and distinguishing between one’s own and others’ representations of the world.

Although FXS is a genetic disorder, there is also evidence for environmental contributors to IQ and other specific neurocognitive deficits, including factors relating to the home environment (enrichment opportunities and socioeconomic status). All of these factors taken together—gender, variation in the *FMR1* repeats, and environment—contribute to the significant within-syndrome variability found in FXS.

Fragile X Syndrome, Autism Spectrum Disorder (ASD), and Other Comorbidities. It has been reported that many fully affected males with FXS show symptoms consistent with ASD such as language delay, echolalia, stereotypies, self-injurious behavior (SIB), perseveration, poor eye contact, and tactile defensiveness. Therefore, it is not surprising that early workers wondered whether FXS would reveal a common genetic cause of ASD. Studies to examine the prevalence rate of ASD among individuals with FXS has been somewhat inconsistent, partly due to variable methodology and revised diagnostic criteria for ASD, with prevalence rates ranging from 5% of males with FXS having autistic disorder to half of males and nearly 20% of females with FXS meeting DSM-5 criteria for ASD.

In addition, social and performance anxiety, slow-to-warm up temperament, including shyness, social withdrawal, mutual gaze aversion, hyperarousal, hyperactivity, and attention deficits are common among individuals with FXS. Attention-deficit/hyperactivity disorder (ADHD) symptoms are higher among affected boys relative to control subjects, and girls with FXS have lower prevalence rates of ADHD than boys with FXS. Among adults, problems in attending and in sustaining effort have been found in the neuropsychological profiles of women with expanded *FMR1* repeats, which may contribute to problems in math, abstract reasoning, and planning.

FXS thus involves vulnerabilities toward shyness, gaze aversion, social anxiety, avoidant disorders, schizotypal disorder, ADHD, and ASD. These difficulties vary in severity but are typically found in persons across the IQ spectrum, from those with moderate ID to those with mild learning disabilities.

Individuals with premutation (typically family members of affected individuals) are also at risk for psychiatric and medical issues including social dysfunction, primarily shyness, gaze aversion, and social anxiety and the adult-onset FXTAS. FXTAS is characterized by progressive intention tremor and gait ataxia, with associated dementia and global brain atrophy resulting in significant morbidity and mortality.

Prader–Willi Syndrome. First identified in 1956, PWS affects about 1 in 15,000 births and is best known for its food-related characteristics. Whereas babies with PWS invariably show hypotonia and pronounced feeding-sucking difficulties, young children between

2 and 6 years of age develop hyperphagia and food-seeking behavior such as food foraging and hoarding. Hyperphagia is likely associated with a hypothalamic abnormality resulting in a lack of satiety. Food preoccupations are lifelong, and without prolonged dietary management, affected individuals invariably become obese. Indeed, complications of obesity remain the leading cause of death in this syndrome.

PWS is the first known human disease to show the effects of genomic imprinting, in which genes are modified and expressed differently depending on whether they are inherited from the mother or the father. PWS is caused by the loss of the paternal copy of the PWS “critical region” on chromosome 15q11.2–13, whereas loss of the maternal copy of 15q11.2–13 results in a completely different and more severe developmental disorder, Angelman syndrome. Up to 70% of PWS cases are caused by a paternally derived deletion on the long arm of chromosome 15. Less than 5% of PWS cases are caused by a defect in the imprinting center. The remaining cases are attributed to maternal uniparental disomy (UPD) of chromosome 15, in which both members of the chromosome 15 pair come from the mother. The prevalence of PWS by UPD is reportedly increasing, likely due to the recent trend of advancing maternal age.

Table 39–8.

Phenotypes of Selected Conditions Associated with Intellectual Disability

Disorder	Cause	Inheritance	Prevalence[a]	Recurrence Risk	Testing	Clinical Features	Behavioral Phenotype
Chromosomal							
Cri-du-chat syndrome	5p15.2 deletion	Dominant deletion (usually de novo)	1/20,000–50,000	Very low	Chromosome microarray	Microcephaly, round face with hypertelorism, epicanthal folds, down-slanting palpebral fissures, broad, flat nose, low-set ears, micrognathia, prenatal growth retardation, respiratory and ear infections, congenital heart disease, gastrointestinal abnormalities	Severe ID, infantile "catlike" cry, hyperactivity, stereotypies, self-injury
Down syndrome	Trisomy 21	Sporadic	1/2,500–1/32 births based on maternal age	Low, unless due to translocation	Karyotype	Hypotonia, facial gestalt (upward-slanting palpebral fissures, flat, wide nasal bridge), single palmar crease, short stature, hypothyroidism, congenital heart disease, intestinal atresia	Moderate ID, strength in social development, hyperactivity in childhood, stubborn, verbal > auditory processing, increased risk of depression and Alzheimer-type dementia in middle age
Smith–Magenis syndrome	17p11.2 deletion	Dominant deletion (usually de novo)	1/25,000	Very low	Chromosome microarray	Broad face, flat midface, progressively coarsening facial features, short, broad hands, small toes, and hoarse, deep voice	Severe ID, hyperactivity, severe self-injury including hand biting, head banging, and pulling out finger and toe nails, stereotyped "self-hugging" and "lick and flip" behaviors, attention seeking, aggression, characteristic sleep/circadian rhythm disturbance
22q11 Deletion syndrome (Velocardiofacial/DiGeorge)	22q11 deletion	Dominant deletion (de novo or inherited)	1/4,000	Low if de novo, up to 50% if parent affected	Chromosome microarray or FISH	Congenital heart disease, cleft soft palate, long tapered digits, T-cell deficit, hypocalcemia	Variable ID, learning disability, ADHD, autism, psychosis (schizophrenia 25%)
Williams syndrome	7q11–23 deletion	Dominant deletion (usually de novo)	1/7,500	Low if de novo, up to 50% if parent affected	Chromosome microarray	Broad forehead, depressed nasal bridge, stellate iris pattern, widely spaced teeth, full lips, narrow chin, elastin arteriopathy +/– hypertension, short stature, hypercalcemia, hypothyroidism	Mild to moderate ID, anxiety (specific phobias and general anxiety), emotional lability, hyperactivity, outgoing, sociable, verbal skills > visuospatial skills
Single Gene							
Fragile X syndrome	*FMR1* trinucleotide repeat expansion	X-linked with anticipation	1/4,000 males, 1/8,000 females	Up to 50%	FXS testing for repeat length	Long face, large ears, flat midface, high, arched palate, short stature, macro-orchidism, mitral valve prolapse, joint laxity, strabismus	Variable ID (milder in females), verbal IQ > performance IQ, hyperactivity, inattention, anxiety, stereotypies, gaze aversion, social avoidance, shyness, irritability
Rett syndrome	Altered gene expression due to loss of *MECP2* function	X-linked, 99% de novo	1/8,500 females (rare in males)	Low	*MECP2* gene sequencing	Acquired microcephaly; gait ataxia and apraxia; tremors; seizures	Moderate to severe ID, regression at 6–18 months, stereotyped "wringing" hand movements, fits of screaming and inconsolable crying, autistic features, panic-like attacks, bruxism, episodic apnea and/or hyperpnea
Cornelia de Lange syndrome (Cohesinopathy)	*NIPBL, SMC1A, SMC3, RAD21, HDAC8*		1/40,000–100,000		Panel sequencing vs. exome	Continuous eyebrows, small, upturned nose, anteverted nostrils, thin, down-turned upper lip, microcephaly, short stature, small hands and feet, malformed upper limbs, failure to thrive	Severe ID, self-injury, language delays, avoidance of being held, stereotypic movements
Tuberous sclerosis complex 1 and 2	Loss of *TSC1* or *TSC2* function	Autosomal dominant	1/10,000		*TSC1/TSC2* sequencing	Hamartomas of central nervous system, skin, kidney, heart, epilepsy	Normal IQ to profound ID, social deficits, autism, hyperactivity, impulsivity, aggression, self-injury, sleep disturbance, anxiety, depression
Neurofibromatosis type[a] (RASopathy)	Loss of *NF1* function	Autosomal dominant (de novo or inherited)	1/2,500–4,000	Low if de novo, 50% if parent affected	*NF1* sequencing	Café-au-lait macules, intertriginous freckling, Lisch nodules, cutaneous neurofibromas, optic pathway glioma, short stature, and macrocephaly in 30–45%	10% moderate to profound ID (verbal IQ > performance IQ), 50% speech and language difficulties, hyperactivity, impulsivity, distractibility, anxiety
Imprinting							
Prader–Willi syndrome (PWS)	Lack of paternal-specific imprinting of 15q11.2-q13	Imprinting (90% sporadic)	1/10,000–30,000	Low, except with paternal mutation	PWS/AS methylation testing	Hypotonia, failure to thrive in infancy, obesity, small hands and feet, micro-orchidism, crypto-orchidism, short stature, almond-shaped eyes, fair hair and light skin, flat face, scoliosis, orthopedic problems, prominent forehead, and bitemporal narrowing	Mild to moderate ID, compulsive behavior, hyperphagia, hoarding, impulsivity, borderline to moderate mental retardation, emotional lability, tantrums, excess daytime sleepiness, skin picking, anxiety, aggression
Angelman syndrome (AS)	Loss of *UBE3A* function via mutation or lack of maternal-specific imprinting of 15q11.2-q13	Imprinting (90% sporadic)	1/12,000–24,000	Low, except with maternal mutation	PWS/AS methylation testing and *UBE3A* sequencing	Fair hair and blue eyes (66%); dysmorphic faces including wide, smiling mouth, thin upper lip, and pointed chin; epilepsy (90%) with characteristic electroencephalogram; ataxia; small head circumference; 2–5% microcephalic	Moderate to severe ID, happy disposition, paroxysmal laughter, hand flapping, clapping, profound intellectual disability, sleep disturbance with nighttime waking, possible increased incidence of autistic features, love of water and music (anecdotal)
Metabolic							
Smith–Lemli–Opitz	Loss of *DHCR7* function, defect in cholesterol metabolism	Autosomal recessive	1/20,000–40,000	25%	7-dehydrocholesterol level, *DHCR7* sequencing	Characteristic facial features (narrow forehead, epicanthal folds, ptosis, short mandible, short nose, anteverted nares, low-set ears), cleft palate, cardiac defects, underdeveloped male genitalia, postaxial polydactyly, and 2–3 toes syndactyly	Mild to profound ID, self-injury, hyper-reactivity, rapid body-arching movements, sleep disturbance, autism
Phenylketonuria (prototypical biochemical genetic disorder)	Accumulation of phenylalanine	Autosomal recessive	1/10,000	25%	Newborn screening, plasma amino acids	Symptoms absent neonatally, later development of seizures (25% generalized), fair skin, blue eyes, blond hair, rash	Mild to profound ID, language delay, destructiveness, self-injury, hyperactivity

Mitochondrial	Defects in electron transport chain	Autosomal recessive, autosomal dominant, X-linked or maternal	1/5,000	Up to 25%	Lactate, pyruvate, plasma amino acids, MRS, muscle biopsy	Multisystem involvement (brain, eye, heart, kidney, muscle), decompensation with physiologic stress, regression, seizures, hearing loss; can present in adulthood as psychiatric disease.	Unknown
Lesch–Nyhan syndrome	Loss of *HPRT1* function, accumulation of uric acid	X-linked	1/380,000	50% of sons	Urine urate-Cr ratio, 24-hour urine urate, *HPRT1* sequencing	Hypotonia followed by extrapyramidal movements, spasticity, severe motor disability, uric acid accumulation in kidney (calculi), joints (gout), skin (tophi)	Mild to moderate ID (after normal pre- and perinatal course), severe self-injury/biting behaviors, aggression
Ornithine Trans-carbamylase Deficiency	Loss of *OTC* function and ammonia accumulation	X-linked	1/14,000–77,000	Carrier mothers: 50% of sons; affected fathers: 50% of daughters are carriers	Ammonia level, *OTC* sequencing	Encephalopathy in response to protein ingestion or stress, respiratory alkalosis, altered mental status/delirium	Normal IQ to severe ID (more severe in males), variable age of presentation, ADHD, executive function deficits
Lysosomal storage diseases (diverse group of >50 disorders)	Accumulation of various macro-molecules	Autosomal recessive, X-linked	1/5,000–10,000	25%	Enzymatic and gene sequencing	Coarse facies, corneal clouding or macular cherry-red spot, macroglossia, dysostosis multiplex (skeletal findings), organomegaly (heart, liver, spleen), cardiac arrhythmias	Broad range of ID, gradually progressive
Toxin							
Fetal alcohol syndrome	Alcohol neurotoxicity	NA	1/3,000 in western countries, 1/300 with fetal alcohol effects	NA	Clinical diagnosis	Microcephaly, short stature, midface hypoplasia, short palpebral fissure, thin upper lip, retrognathia in infancy, micrognathia in adolescence, hypoplastic long or smooth philtrum	Mild to moderate ID, irritability, ADHD, memory impairment

*Estimated prevalence in mixed European populations.
ADHD, attention-deficit/hyperactivity disorder; ID, Intellectual Disability; IQ Intelligence Quotient.
As genetic testing has become more widespread, the known clinical presentation for each genetic disorder has expanded beyond the classic presentations described in the early literature. More and more, disorders are being grouped by their shared molecular pathophysiology, e.g., Cornelia de Lange syndrome as a cohesinopathy (defects in the cohesion DNA-binding complex involved in DNA repair, chromosome segregation, and gene regulation) and Neurofibromatosis 1 as a RASopathy (defects in the RAS/MAPK network of genes involved in cell cycle regulation and differentiation).

Although people with PWS invariably obsess about food, a remarkably high proportion of persons also show nonfood obsessions and compulsive behaviors. These nonfood symptoms include skin picking, hoarding, needing to tell, ask, or say things, and having concerns with symmetry, exactness, ordering, arranging, cleanliness, and sameness in daily routine. Often these symptoms are associated with distress or adaptive impairment, suggesting marked risks of obsessive-compulsive disorder in this population. Indeed, obsessive-compulsive disorder is many more times likely in people with PWS than in the general population of people with ID.

In addition, even as compared to others with ID, children and adults with PWS show high rates of temper tantrums, aggression, stubbornness, underactivity, excessive daytime sleepiness (in part, related to central and/or obstructive sleep apnea syndromes) and emotional lability. Coupled with food seeking, these impulsive behaviors often lead people with PWS to need more restrictive levels of care than would be predicted by their mild levels of ID.

Psychopathology and cognitive profiles may be associated with an individual’s genetic variant of PWS. Thus, research has shown behavioral differences between PWS due to paternal deletion and maternal UPD. Cases with deletions may show lower IQ, especially verbal IQ, and more frequent or severe problem behaviors, such as skin picking, hoarding, temper tantrums, overeating, and social withdrawal. Although a dampening of symptom severity is suggested in many UPD cases, there are occasional cases of more severe problems in UPD, primarily autistic-like features and relatively low IQ. In addition, persons with UPD may do better than their counterparts with deletions on verbal tasks but more poorly on visual-spatial tasks, including solving jigsaw puzzles. Indeed, although many persons with PWS are remarkably skilled in solving puzzles, this unusual strength seems to be limited to those with paternal deletions. The mechanisms that explain such differences in maladaptive behavior and cognition across genetic subtypes are unknown. In sum, many people with PWS are at increased risk for obsessive-compulsive, impulse control, and affective disorders. Yet even those who do not meet diagnostic criteria for these psychiatric disorders often show significant maladaptive behaviors that interfere with optimal adaptive functioning.

Down Syndrome. Occurring in approximately 1.2 in 1,000 live births, DS is the most common chromosomal abnormality leading to ID. It is most often the result of nondisjunction of chromosome 21. The syndrome has been the subject of considerable investigation.

More than 80 clinical features characterize the phenotype of DS, including cognitive impairments, hypotonia, facial dysmorphisms, congenital heart disease, and other anomalies. Studies of children with DS have shown distinctive cognitive and linguistic profiles. Notable strengths have been highlighted in visual versus auditory processing. Language impairment can be extensive, with particular difficulty in expressive language, grammar, and pronunciation.

In contrast to the relative overall weakness in the cognitive arena, persons with DS are often described as being particularly socially adept. However, studies of adaptive functioning suggest that these skills are not uniform, with communication scores trailing behind measures of daily living skills and socialization. The stereotype of a “Down personality” as being happy, good-tempered, affectionate, placid, and stubborn has been difficult to verify.

Compared to others with ID, persons with DS appear to suffer less often and less seriously from psychopathology. Adults with DS appear to be somewhat less prone to psychiatric disturbance than controls. This trend seems to extend to children and adolescents as well, in whom rates of psychiatric and behavioral problems are greater than in the general population but appreciably less than in other groups with ID. Commonly noted problems include difficulties with attention, impulsivity, hyperactivity, and aggression. In contrast to these problems, depression seems to be less common among children and adolescents than expected norms, although some youth appear to become increasingly withdrawn in their adolescent years. For many, this shift toward increased withdrawal seems gradual and subtle, and it may be an early harbinger of a later onset of depression or dementia in the adult years. Autism spectrum disorder appears to be relatively rare.

Williams Syndrome. First identified in 1961, Williams syndrome is caused by a microdeletion on chromosome 7 that includes the gene for elastin, a protein that provides strength and elasticity to certain tissues such as the heart, skin, blood vessels, and lungs. The condition affects 1 in 7,500 individuals, and persons with Williams syndrome often show hyperacusis, hypercalcemia, neuromusculoskeletal and renal abnormalities, and characteristic facial features described as elfin-like, cute, and appealing. People with Williams syndrome also typically show cardiovascular disease, especially supravalvular aortic stenosis, and these problems are likely associated with elastin insufficiency.

Apart from these medical issues, recent attention has focused on the interesting, possibly unique, cognitive-linguistic profile shown by many persons with this syndrome. Children with Williams syndrome show relative strengths in verbal short-term memory and language, and that in general verbal and nonverbal abilities develop at different rates for children with Williams syndrome (Jarrold et al., 2001). Indeed, for many years it was thought that these children might even perform at chronologic-age levels on a variety of linguistic tasks. Although age-appropriate performance in language has now been found in only small percentages of children with Williams syndrome, these children nevertheless show relative strengths in language (vocabulary, grammar, and short-term auditory memory), as well as in auditory processing and in some areas of music. In contrast, many children with Williams syndrome perform especially poorly on a variety of visuospatial tasks.

Studies have yet to fully examine the personality or psychiatric features of people with Williams syndrome. Early descriptions of people with Williams syndrome hinted at a “classic” Williams syndrome personality, described as pleasant, unusually friendly, affectionate, loquacious, engaging, and interpersonally sensitive and charming. Such qualities may change over the course of development, with adults being more withdrawn and less overly friendly than children.

Relative to controls, adolescents and adults with Williams syndrome appear more likely to initiate interactions with others, to enjoy social activities, to be kind-spirited and caring, and to empathize with others’ positive feelings or when others are in pain. At the same time, however, they appear to have difficulty in making or keeping friends and risk of “stranger danger.” Indeed, although sociability in Williams syndrome has generally been viewed as a strength, these features also seem to reflect the type of social disinhibition that is characteristic of people who are anxious, impulsive, and overly aroused. In fact, ADHD and anxiety including specific phobia and generalized anxiety are commonly co-occurring in children with Williams syndrome.

Smith–Magenis Syndrome

Bryan was a 19-year-old young man with moderate intellectual disability and a history of psychiatric issues including impulsivity, aggressive behaviors, and a striking history of self-injurious behaviors such as self-hitting, skin-picking, and pulling out fingernails and toenails. He had been given a diagnosis of autism. Bryan had restricted language abilities but got his basic needs met using a small set of phrases and gestures, though his speech was hard to understand and muffled. He lived at home among a devoted family including his parents and elder brother, and participated in a transitional school program, where he was segregated in a classroom for children deemed “lower functioning” with “behavioral needs.” Bryan and his parents visited clinic with the aim of finding a better medication regimen that might be able to support his adaptive functioning, articulating their “dream” that their son might one day live independently or participate more easily in social activities.

At the evaluation, Bryan’s medication regimen consisted of extended-release dexmethylphenidate, trazodone, buspirone, risperidone, and sertraline, with as-needed lorazepam and additional risperidone for agitation and aggression. These medications were not fully controlling his psychiatric symptoms. A careful history was taken, which elicited a history of infantile feeding difficulties, hypotonia and failure to thrive, and a striking pattern of lifelong sleep disturbance. Bryan fell asleep successfully, but woke every night around 2 to 3 a.m., did not go back to sleep, and was sleepy in the afternoons. He was noted to have somewhat heavy and coarse features with brachycephaly and prognathism. Accordingly, first-line genetic testing for intellectual disability consisting of chromosome array and fragile X was sent, revealing a 3.5-Mb interstitial deletion of chromosome 17p11.2, consistent with a diagnosis of Smith–Magenis syndrome (SMS). SMS is a genetic syndrome with a prevalence of 1:15,000–1:25,000, with disruptions to the *RAI1* gene (located within the deletion) accounting for the majority of features. A number of important studies support the presence of distinctive disruption to melatonergic control of the sleep cycle in SMS and the use of melatonin and beta-blockers to normalize sleep, with resultant improvement in behavioral and cognitive function. Bryan was started on melatonin 3 mg in the evening and propranolol 10 mg in the morning. Over 2 months, these were increased to final doses of 5 mg and 20 mg, respectively. His overnight sleep period increased from 4 hours to 8 hours, and by the conclusion of the titration, his self-injury was reduced by 90%, and aggressive behaviors stopped. He began to exhibit new learning, such as starting to search for and complete crossword puzzles, and he displayed an increased interest in social interaction and improved attentiveness. His psychiatrist was able to eliminate the use of risperidone, trazodone, and all as-needed medications. Importantly, standard-of-care recommended medical screenings for SMS were completed, revealing that Bryan was deaf. He received bilateral hearing aids and speech therapy and began to speak more. A year later, Bryan’s adaptive function was significantly improved, and he was able to transition to a supported group living situation with two male peers and began work at a part-time job.

Smith–Lemli–Opitz Syndrome

Brianna was a 13-year-old young woman carrying a diagnosis of intellectual disability and Smith–Lemli–Opitz syndrome (SLOS). SLOS is a congenital syndrome characterized by growth retardation, microcephaly, and multiple malformations, including a distinctive facial appearance. Most affected individuals have moderate to severe intellectual disability, but some have normal development, and it is increasingly thought that the disorder and carriers may be much more common than previously understood. In SLOS, mutations in the *DHCR7* gene result in derangement of cholesterol metabolism with elevated serum 7-dehydrocholesterol (7DHC) and low serum cholesterol. Affected individuals almost always display a variety of structural midline and para-midline brain findings, most commonly abnormalities of the corpus callosum. Sterol levels display a strong correlation with MRI brain findings and the degree of somatic malformations, supporting a link between altered cholesterol/sterol metabolism, deficient brain myelination, and changes in neurotransmitter function that may underlie the neuropsychiatric phenotype.

SLOS is, therefore, a specific etiology resulting in intellectual disability, as well as a disorder that may be simultaneously conceptualized as being a biochemical genetic disorder, multiple congenital anomaly syndrome, and/or brain malformation. As with many metabolic disorders, it is thought that dietary supplementation of the deficient substrate, in this case, cholesterol, may support ongoing brain development. Brianna received cholesterol supplementation since diagnosis and ongoing dietary optimization, as well as support for learning via an Individualized Educational Plan at school. She received low-dose trazodone to support sleep fragmentation; however, as she entered her teenage years shortly after puberty, Brianna began to exhibit increasing challenges with cognitive control and behavioral dysregulation, becoming much more aggressive towards caregivers and peers and with self-injurious behaviors. Simultaneously, she talked of visual and auditory hallucinations, particularly centered around seeing, hearing, and interacting with a person called “Cassie.” These phenomena waxed and waned somewhat on a multiweek or monthly basis, but generally, periods of increased aggressive behaviors coincided with what was interpreted as perceptual or psychotic phenomena.

Olanzapine was started to address aggressive behaviors, self-injury, and presumed psychosis and was titrated to 10 mg twice per day. Brianna’s aggressive behaviors declined somewhat, but the perceptual phenomena appeared unresponsive, and she gained some 40 lb. With concern for treatment resistance and weight gain, she was switched to haloperidol 5 mg twice per day. With the use of this agent, Brianna became more irritable with continuing hallucinations. At this point, the treating child psychiatrist sought consultation with a developmental neuropsychiatrist and biochemical geneticist at a specialist center. Laboratory studies revealed that 7DHC levels were elevated. In accordance with best practice guidelines for SLOS, both trazodone and haloperidol, which are contraindicated in this disorder due to their link with 7DHC elevation, were stopped. Brianna was started on valproic acid, which proved effective in reducing her behaviors and addressing psychosis. Melatonin was prescribed for sleep initiation difficulties with good effect.

DEVELOPMENTAL CONSIDERATIONS IN CHILDREN WITH INTELLECTUAL DISABILITY

Throughout the 20th (and now 21st) century, children with ID have been an ongoing topic of developmental research. Small groups of researchers have examined the development of various Piagetian cognitive domains, language abilities, social skills, and even morality. Besides providing information about basic milestones or natural histories, these studies aid in understanding the underlying processes of development. Researchers have been able to specify the ordering of development (sequences), the ways in which levels in certain areas relate to levels in other areas (cross-domain relations), and the points at which the speed of development (rate) changes. Such analyses are now being applied to children with different etiologies of ID. In addition, developmental analyses have extended beyond studies of children themselves. Many developmental psychologists examine mother–child interactions, family behaviors in families of different-aged children, and the child's interactions with peers, neighbors, and schools. This section examines both child-related and environment-related aspects of development in children with ID.

Child-Related Aspects of Development

Overview. Children with ID traverse the same sequences in their development as do children with typical development, but in the setting of certain forms of ID there are strengths and weaknesses not generally found in typically developing children. Children with ID also show particularly slowed development at certain ages and tasks, which seem related to specific types of ID, although the reasons for such slowings are the subject of ongoing research.

When examining child–environment interactions, it is found that mothers act “the same and different” compared to mothers of typically developing children at similar levels of functioning. Mothers of children with ID are similar in that they shorten their speech, focus on keywords, and in other ways, structure the environment for their child's development. They seem, however, to be much more didactic in their interactions. Families of children with ID also differ widely from one another. When the child with ID shows less maladaptive behavior, and when the parent's style of coping is more problem-focused, the result seems to be reduced levels of parent and family stress.

Sequences. Typically developing children show a specific, possibly universal, ordering to their development. For example, in Piagetian cognitive development, children proceed from sensorimotor, to preoperational, to concrete operational, to formal operational thought. Even within these four larger stages, smaller orderings prevail. Thus, within sensorimotor development, infants proceed in order through Piaget's six substages in each of several subdomains. Do children with ID also follow a “similar sequence” to their development? For the past 50 years, similar sequences have been found for a variety of children over many tasks: in almost 20 Piagetian domains, in symbolic play, and in linguistic grammar and pragmatics. Similar sequences even hold for those children with genetic or other organic causes for their ID. The only possible exceptions include some children with uncontrollable seizures (where accurate testing is difficult) and some children with autism, who may show different orderings due to their particular disabilities on certain social tasks. For almost all children and for a wide variety of behaviors, development for children with ID proceeds along similar sequences to those found in typically developing children.

Cross-Domain Relations. If children with ID proceed in the same developmental ordering as do typically developing children, might they also show the even or flat profiles that groups of typically developing children show from one domain to another? Such a “similar structure” to development was proposed by Zigler in the late 1960s. Contrary to similar sequences—which hold for all children with ID—cross-domain structures may differ based on which group one considers. With a few exceptions, children with cultural-familial ID do show even or near-even performance across various intellectual domains. In contrast, children with different organic forms of ID show specific intellectual strengths and weaknesses. For example, three separate groups show weaknesses in sequential (or bit-by-bit, serial) processing compared to simultaneous (i.e., gestalt, holistic) processing. Such sequential deficits are found in boys with FXS, children with PWS, and children with Smith–Magenis syndrome. Children with DS may have particular weaknesses in expressive versus receptive communication and specific problems in grammatical abilities. Children with Williams syndrome show greater deficits in visuospatial processing skills, and a subset of these children shows heightened abilities in language. Different etiologies of ID, then, show different characteristic intellectual strengths and weaknesses.

Developmental Rates. A further issue concerns a child's rate of development. Children with ID develop at slower rates than typically developing children. As noted in the Prevalence section, nonspecific or mixed groups show reasonably steady rates of development, as shown by stable IQ after the early childhood years. Children with different types of ID, however, may have periods of accelerated or slowed development, and these periods seem to be related to either age or task difficulties. For example, boys with FXS show stable developmental rates until the age of about 9 or 10, at which time development slows. Similarly, in DS, both cross-sectional and small-scale longitudinal studies show that these children slow over the 6-to-11-year period. Throughout this age span, children with DS slow in their development of linguistic grammar, intellectual skills, and adaptive behavior. In both FXS and DS, it remains unknown why such age-related slowings occur, whether all children with these disorders show such slowings, or if slowings are seen across all domains.

A second type of slowing relates to pronounced difficulties in mastering a specific developmental task. These can occur at a variety of ages, whenever children with a particular type of disability reach a specific developmental task. Consider young children with DS (who also show age-related slowing from 6 to 11 years). These children show difficulties in acquiring certain tasks of infant intelligence, even after accounting for their already slower rates of development. Later, toddlers with DS show increasing delays in expressive as opposed to receptive language skills. As a result of both age- and task-related slowings, then, children with DS generally show their highest IQ scores during the first year of life, after which IQ declines.

Contextual Aspects of Development

Mother–Child Interactions. Following research on interactions between typically developing children and their mothers, mother–child interaction studies in ID began during the mid-1970s. This research has shown that mothers of children with ID are both similar and different in their interactions than mothers of typically developing children of the same mental age (MA). Specifically, the two groups of mothers behave similarly on what might be called the structural aspects of behavior. In language, for example, mothers shorten their speech, simplify their vocabularies, raise their pitch, repeat key words, and generally provide the “motherese” typical of mothers of typically developing children. As a result, mothers of typically developing 2-year-olds act almost identically to mothers of children with ID who are at 2 years MA.

However, compared to mothers of typically developing, MA-matched children, mothers of children with ID are oftentimes more didactic. They more often initiate the topic of interaction, take longer speaking turns, talk at the same time as their child, and in other ways control the interaction. As a result, the child is allowed fewer opportunities to initiate and control the conversation. It remains unclear why such stylistic differences occur. The prevailing view holds that mothers of children with ID are worried about their child’s development: in their zeal to teach and instruct their children, these mothers might inadvertently be using a style of interaction that is not the most effective. Although only a few studies exist to date, young children with ID seem to develop faster when their mothers are less intrusive. A recent study examined mother–child and father–child interactions in 89 children with ID ranging from 20 to 47 months (Vilaseca et al., 2019). Results indicated that maternal responsiveness and paternal teaching predicted cognitive development. These studies provide evidence of the importance of paternal involvement and maternal responsiveness for cognitive and language development in children with ID.

Family Reactions. Over the past few decades, family studies in ID have changed enormously, from a predominantly negative to a more balanced perspective. Before the early 1980s, families of children with ID were perceived to be families in crisis. Mothers were examined for their “mourning” reactions, couples for divorce, and mothers, fathers, and unaffected siblings for the presence of depression and other forms of psychopathology. Since the early 1980s, however, family researchers began to change how they conceptualized families of children with ID. From an earlier focus on family pathology, researchers now saw families of children with ID as families facing increased stress. And while children with ID might add stress to the family system, now researchers understood that this stress could result in negative or positive adaptations. This changing view implied a more normative view of these families: just as every family faces a variety of stressors in life (e.g., illness, loss of job, etc.), so too do families of children with ID.

The move from a negative to a stress-coping perspective also has other implications. First, it allows for a change away from a predominantly “between-group” research strategy. Previously, most studies compared families of children with disabilities to families of typically developing, nondisabled children of the same chronologic or mental age. Now the between-group focus has been complemented by within-group studies. Given that families of children with ID vary in their ability to cope, what causes better adjustment in one family and worse adjustment in another? What child, parent, or family characteristics foster better adaptation? To date, most work has examined child characteristics such as the child’s age and degree of impairment. For both variables, findings have been inconsistent. Some studies find it more stressful to parent older children with ID, and some suggest that families may experience more stress when the child begins puberty (11 to 15 years) and again when the child reaches early adulthood (20 to 21 years). However, other studies do not find such relations between increased family stress and the child’s age. Similarly, some studies find that it is more difficult to parent children with more severe levels of disability, whereas other studies do not.

Another characteristic that has been examined is the child’s degree of maladaptive behavior. So far, strong connections have been found between child maladaptive behavior and family stress in families of children with mixed etiologies, with PWS, and with 5p- (cri-du-chat) syndrome. Less clear is why such connections occur. Although most researchers feel that child maladaptive behavior leads to increases in familial stress, increased familial stress may also elicit more child behavioral problems. In studies of parent coping style, parents who more actively, constructively attempt to manage their child do better than those parents who adopt a “palliative” coping style, one that either dwells on or ignores parental emotions. This difference in coping style may help buffer parents and families from the increased stresses of parenting a child with ID.

Psychopathology in Intellectual Disability

Developmental disability is a significant risk factor for psychopathology in general, and, as may be inferred from the description of causes above, this increased risk may derive from biologic vulnerabilities as well as risks that accrue from the environment. Moreover, individuals with IQs below 70 have a two- to fivefold higher rate of psychiatric disorders overall compared with typically developing persons. A host of explanations, highlighted in Table 39–9, have been put forward in efforts to account for this added risk. Included are developmental experiences with which these individuals must contend, including perceived rejection from peers and frustration from parents. The repeated experiences of failure or difficulty achieving what appears to come naturally to typically developing peers, the recapitulation of experiences of failure with each developmental stage, and less well-developed or perhaps less supportive peer groups all take a toll on ego development. In addition to specific biologic vulnerabilities, the common comorbidity of ID with physical illness, for example, epilepsy, may also increase the risk of mental illness.

Table 39–9.
Possible Contributions to Increased Vulnerability to Mental Disorders in Persons with Intellectual Disability

Neuropathologic process(es) responsible for intellectual disability may also cause or increase the risk for mental illness.
There may be an increased likelihood of loss and separation, particularly in out-of-home placements.
Communication deficits may predispose to emotional/behavioral disturbance.
There is vulnerability to exploitation and abuse by others.
There may be inadequate coping skills.
Family stress may be heightened in the presence of a child with developmental disability.
There is a risk of a limited network of social relationships and repertoire of social skills.
There is a risk of reduced opportunities for the development and exercise of recreational and occupational skills.
There is a possible adverse effect of the disability on self-esteem, and possible dysmorphology.

Moreover, the treatments for epilepsy and other medical conditions may carry some “behavioral toxicity” which can increase the likelihood of diagnosed mental illness. Phenobarbital was widely reported to increase the risk of motoric hyperactivity and disinhibition in children and in individuals with developmental disorders, and phenytoin may cause “cognitive toxicity.” While neither of these agents is used as frequently today, similar concerns may obtain for essentially any of the medications that are used for the management of epilepsy or other chronic medical conditions.

In general, individuals with developmental disabilities may experience the full range of mental illnesses as occur in the general population. A meta-analysis of epidemiologic studies of psychiatric disorders in individuals with ID revealed higher rates of psychiatric disorders in children with versus without ID, and higher rates in children and adults with more severe forms of ID versus milder forms (de Ruiter et al., 2007). However, evidence of a greater prevalence of psychiatric disorders in adults with *mild* disability as compared to adults in the general population is less clear. Other studies have shown that, in children, there is evidence that the developmental *course* of psychopathology is similar from ages 6 to 18 for children with and without ID. For instance, in general, internalizing behaviors increase with age and externalizing behaviors decrease with age. Specific risk factors for psychiatric disorders in children with intellectual impairments include both family and child characteristics: moderate versus mild levels of intellectual impairment, adaptive behavior deficits, poor socialization, language impairments, living with one biologic parent, and low family SES.

In 2007, the National Association for the Dually Diagnosed (ID and mental illness), in collaboration with a number of other professional organizations, published a manual to assist clinicians in psychiatric diagnosis with this population. The DM-ID provides a number of explicit examples of adapted diagnostic criteria for each diagnosis and is currently under revision. Additional specific psychiatric disorders as they appear in individuals with ID are reviewed below.

Attention-Deficit/Hyperactivity Disorder. The rates of ADHD in ID are estimated to between 9% and 18%. Although considerable research has been done on the treatment of ADHD, virtually all of these studies have specifically excluded children with ID. Moreover, the diagnosis of ADHD is based upon developmental considerations, namely, hyperactivity, impulsivity, and inattention significantly greater than that expected for a given developmental age, and thus the threshold for diagnosis in persons with severe to profound ID should be elevated.

For persons with ID, the diagnosis of ADHD is qualified as being excessive for an individual’s *mental age.* In the context of profound ID, attention span, distractibility, or “on-task” behavior are predictably quite variable. Individuals given the diagnosis of ADHD in this context should, in comparison to their peers with similar levels of intellectual impairment, exhibit unusually short attention span, excessive psychomotor activity level, remarkable impulsivity, and so on. In many cases, the clinician will encounter a situation in which an individual does not evidence remarkable psychomotor activity or attentional difficulties but may be unusually impulsive. In these situations, one should entertain the diagnosis of an impulse control disorder. Such a diagnosis, for example, might be appropriate for an individual who inexplicably strikes out at a peer in the absence of any identifiable environmental stressor.

Impulse Control and Stereotypic Movement Disorders: Self-Injury and Aggression. Every effort should be made to determine and treat the underlying cause of SIB or aggression, not merely to suppress the behavior. However, in some cases, for example, stereotypic movement disorder with SIB, the symptom is essentially the diagnosis. Aggression and SIB are common in ID and increase with greater severity of cognitive impairment. SIB typically takes the form of chronic, repetitive, and frequently stereotyped behavior causing trauma. It occurs in the context of specific genetic syndromes, for example, self-biting in Lesch–Nyhan syndrome and fingernail pulling in Smith–Magenis syndrome, but also and more commonly in persons with unknown or nonspecific causes of ID. DSM-5 added specifiers to the diagnosis of stereotypic movement disorder for these situations.

In a review of psychiatric consultations in an institutional setting, it was reported that SIB was specifically cited as a reason for referral in 36% of the 251 cases examined. However, SIB as a presenting symptom is generally unhelpful in predicting the ultimate psychiatric diagnosis.

As SIB and aggression are nonspecific symptoms, one must take into account the presence or absence of a variety of factors to arrive at a presumptive diagnosis. Among these factors are the chronicity of the behavior, what function it serves, whether it is invariant in its topography (e.g., hitting only the right ear suggesting an ear infection), whether it is situational, whether it occurs in concert with regression from a previous level of function, and whether there are associated neurovegetative signs that correlate with its onset.

Oppositional Defiant Disorder/Conduct Disorder. The DSM-5 diagnosis of oppositional defiant disorder or conduct disorder also requires comparisons to others of similar MA. Further, both diagnoses assume some degree of deliberateness on the part of patients (e.g., disobedience motivated by spite or resentment), which can be very difficult to discern in nonverbal subjects with profound cognitive deficits. As such, there is a lower limit in terms of severity of ID below which the reliability and utility of these diagnoses becomes suspect and an unspecified disruptive or impulse control disorder is likely to be utilized for diagnostic purposes.

Anxiety Disorders. Specific anxiety disorders, for example, separation anxiety, obsessive-compulsive disorder, panic disorder, generalized anxiety disorder, and so on, rely heavily on an individual’s ability to describe the subjective symptoms of anxiety. Patients who are clearly avoidant, who exhibit autonomic arousal in the face of stimuli that most of their peers would not find aversive, who present with other features of anxiety but who cannot articulate their subjective states, might be given a diagnosis of unspecified anxiety disorder.

In some cases in which individuals engage in behavior that appears compulsive or driven, and for whom such behavior seems ego-alien, the diagnosis of unspecified obsessive-compulsive or related disorder might be considered. Oftentimes these patients engage in self-restraint (securing their extremities in their clothing) or cling to their parents or care providers, seemingly to prevent SIBs. Simple repetitive behaviors or the insistence on sameness that can accompany autism spectrum disorders should not be given a separate OCD diagnosis. However, it may be useful to organize treatment with an additional diagnosis beyond “autism” if, for example, the focus of treatment is a sleep disturbance, impulsive aggression, anxiety, or the like.

Although common, anxiety disorders appear to be underdiagnosed in persons with ID. The wide variability in prevalence rates, from 1% to 25%, is attributed to difficulty in making a diagnosis. Moreover, an individual with ID may not be able to identify subjective anxiety as an underlying cause of distress, and a patient's aggression or agitation may be suggestive of a disorder of impulse control rather than reflective of underlying anxiety. Indeed, common symptoms of anxiety in the population with ID include aggression, agitation, compulsive or repetitive behaviors, self-injury, and insomnia. Panic may be expressed as agitation, screaming, crying, or clinging, which may even appear as delusional or paranoid behavior. Phobias may also occur in this population and may even be more common in persons with developmental disabilities. Experienced clinicians have noted that persons with developmental disabilities are at high risk for abuse, thus placing this population at a greater risk for posttraumatic stress disorder (PTSD). With data suggesting that it may be seen in nearly 8% of the general population, PTSD is an important diagnosis to consider in individuals with ID.

Eating Disorders. Because of the reliance of diagnostic criteria upon subjective experiences, the diagnoses of anorexia nervosa and bulimia are effectively precluded for individuals with severe or profound ID. For example, one would be hard pressed to identify in nonverbal patients such things as classic distortions in body image, or guilt feelings associated with bingeing. Food refusal or self-induced vomiting would have to be viewed as atypical eating disorders if such symptoms were to occur in the absence of other diagnosable disorders (e.g., depression, rumination, etc.). Pica is perhaps the most common eating disorder among persons with ID.

Organic Mental Disorders. One could argue that everyone with ID has some “organic” cerebral dysfunction, and thus any psychiatric illness should be regarded as organic or due to a general medical condition. In his classic study of psychiatric illness in a sample of institutionalized patients with DS in the 1970s, Menolascino appropriately argued that psychiatric nosology did not have to be reinvented to accommodate individuals with a “tissue diagnosis.” Moreover, he reasoned that patients with the so-called “dual diagnoses” of mental illness and DS need to be distinguished from those with DS alone. Thus, the application of the diagnoses of organic mental syndromes and disorders is best approached as if patients do not have ID. The same principle should apply to personality disorders. The diagnosis of “organic” personality disorder is probably best reserved for individuals whose pre-existing personality was altered in a pathologic way by some additional cerebral insult. In essence, this category has more clinical utility if reserved for patients whose ID is acquired after a period of typical development, usually secondary to trauma experienced in childhood or early adolescence.

Psychosis. Patients with developmental disorders are at an increased risk for schizophrenia (approximately 1% prevalence rate in persons with ID), bipolar disorder, and other mental illnesses that may include symptoms of thought disorder and hallucinations. The diagnosis of schizophrenia requires, at least to some extent, that a patient communicate their experience of delusions or hallucinations. Thus, in individuals with profound ID with limited communicative ability, the diagnosis of classic schizophrenia is arguably impossible. Nonetheless, some individuals display presumptive evidence of response to hallucinations (e.g., striking or shouting at empty space, throwing imaginary peers from furniture) or adopt catatonic postures that can appear to be psychotic in origin. In these cases, the diagnosis of psychosis should be considered if these signs exist in the absence of sufficient evidence of a supervening mood disorder.

Mood Disorders. Even in profound ID, the diagnosis of mood disorders is fairly straightforward. In general, a change in mood from baseline is obvious (recent onset lability, tearfulness, mood elevation, irritability). If coupled with changes in interests, activity level, sleep, appetite, or sexual behavior, of sufficient duration and causing sufficient impairment in adaptive function, the diagnoses of mania or of depression can be made in nonverbal patients.

Mood disorders are not uncommon in persons with ID, although it is unclear from studies to date whether or not mood disorders present at higher rates than in the general population. As noted above, learning problems, social skills deficits, and low self-esteem are often associated with developmental disabilities and represent risk factors for the development of mood disorders. There are no striking differences between the expression of mood disorders in persons functioning in the mild and moderate range of ID and their typically developing peers. Differences may emerge among persons with severe to profound disability, but potential equivalents of mood disturbance may be recognizable, including irritability, crying, problems in sleep or appetite regulation, agitation, mood lability, social withdrawal, and isolation. Aggression or SIB may also be seen as behavioral manifestations of dysphoria in persons regardless of developmental level.

Other Disorders. The diagnosis of Tourette syndrome is difficult in persons with profound ID. Frequently, these individuals display stereotyped or other movements, and it can be challenging to discriminate intentional from unintentional movements or sounds, or to distinguish vocal tics from spontaneous, stereotyped, or echolalic vocalizations in individuals frequently incapable of functional speech. The diagnosis of stereotyped movement disorder might also be considered in such circumstances.

Since elimination disorders require an MA of 4 or 5 years in order to be considered, the diagnoses of functional encopresis or enuresis are seldom made in the context of severe ID. In some instances, individuals appear to lose previously acquired skills, for example, urinary continence, but such losses typically do not occur in isolation, suggesting alternative diagnoses (e.g., delirium, depression, etc.).

Somatoform disorders, depersonalization disorders, and sexual disorders are less frequently diagnosed in the context of ID, though certainly not precluded. Sleep disorders ultimately require the subjective input of the patient regarding the adequacy of rest, occurrence of nightmares, and so on. Given the frequent history of abuse reported for people with ID as a group, one should not overlook the possibility of PTSD when sleep disturbance is a presenting problem.

Provisional Diagnoses. In many cases individuals with ID may not clearly fall into a single diagnostic category. Comorbidity is common. Additionally, some individuals may have psychiatric symptoms that significantly interfere with habilitative function, but which do not allow for a clear distinction between certain diagnoses. It may be difficult to distinguish between an impulse control disorder, perhaps characterized by an individual who engages in impulsive aggressive acts, versus an anxiety disorder, suggested by an individual who strikes out in the context of a stressor that would go unnoticed by most people. The clinician should always make a best effort at a working diagnosis and be prepared to make modifications as indicated by data gathered through collateral sources and from increasing familiarity with a particular patient.

Approach to Maladaptive Behavior

As with psychiatry in general, there is little specificity that can be attached to a given symptom. Persons with ID will typically be referred for evaluation because of self-injurious, aggressive, impulsive, or hyperactive behavior. Because of the lack of diagnostic specificity for each of these symptoms, a diagnostic decision tree cannot be constructed. Instead, it is perhaps more useful to ask a series of questions about the expression of a particular behavior. If the behavior is of recent onset, one is more likely to consider an acute medical or psychiatric etiology. If the behavior is highly situational, occurring primarily in the context of the stress of task demands, the likelihood of a psychosis or mood disorder is probably reduced. If attempts are made to avoid the behavior by self-restraint, the inference of some ego-dystonic features may be tenable. Assessing the sum of these and collateral data will lead the clinician to a presumptive diagnosis that will form the basis for a treatment plan.

Diagnostic Instruments, Rating Scales

A number of diagnostic and behavioral rating scales have been developed for the population with ID. While neither diagnostic nor treatment decisions should be based entirely upon the results of a given instrument, the systematic collection of data through such scales is a very useful adjunct to the clinical diagnostic or monitoring process. Among the most common general behavior rating scales are the Aberrant Behavior Checklist (ABC) and the Developmental Behavior Checklist (DBC). The Behavior Problem Inventory (BPI) is particularly useful for self-injurious, aggressive, and stereotyped behaviors. The Diagnostic Assessment of the Severely Handicapped (DASH), the Reiss Screen for Maladaptive Behavior, and more recently the Reiss Scale for Children's Dual Diagnosis are all instruments whose primary utility is to poll for the presence of mental disorders.

TREATMENT

Prevention

Preventive treatment strategies for ID largely focus on specific medical interventions to reduce the prevalence of ID, and treatments that mitigate associated complications, for example, treating associated mental disorders. The merits of primary prevention are obvious, and the successes enjoyed with PKU should continue to provide powerful incentive for the ongoing collaborations of basic scientists and clinicians. Evidence suggests that newborn metabolic screening is extremely successful in reducing the incidence and severity of some ID syndromes (Brosco et al., 2006). The impact of other programs is less clear. For example, although folic acid supplementation appears to significantly reduce the risk of neural tube defects, and may also reduce the risk of autism, compliance with recommendations to increase *dietary* folate appears to be disturbingly low, although compliance with supplementation is as high as 80% (Pathirathna et al., 2020). Further, it appears that the prevalence of trisomy 21 is increasing despite the availability of noninvasive prenatal testing (NIPT); termination rates are unchanged or decreasing as compared to termination rates prior to NIPT. Finally, a recent review of trends in developmental disabilities in the United

States from 2009 to 2017 (Zablotsky et al., 2019) indicated a significant decrease in certain developmental delays but significant increases in ID, ADHD, and autism.

Influence of Etiologic Factors on Treatment

As suggested above, the approach to treatment begins with the diagnostic process. In some cases, the underlying cause of ID may be particularly important in considering treatments. For example, in ID associated with PKU, attempts have been made to minimize or to attenuate hyperactivity and impulsivity by dietary modification. As animal models of this disorder are developed and explored, it appears that dopaminergic abnormalities play a particularly salient role in the expression of maladaptive behaviors. Thus, early and continuous intervention with dopamine agonists plays a unique role in the treatment of behavioral and cognitive disturbance associated with PKU.

Abnormalities in serotonergic function have been reported in the context of DS. This hyposerotonemia led to trials of serotonin replacement with mixed results, but serotonergic drugs may yet hold relevance for persons with DS. In the context of PWS, serotonergic abnormalities are also being reported as responses to serotonergic agents. Taken together, the identification of the underlying cause of ID has become increasingly important in considering biologic treatments. In the same way, the diagnosis of mental disorders or syndromes in persons with ID will guide and influence treatment strategies.

TREATMENT OF MENTAL DISORDERS

Experience over the past decade has made it clear that ID is a multidisciplinary problem, and optimal treatment is multimodal. Biologic, psychological, social, and developmental dimensions should all be considered when designing a treatment plan for an individual with ID. Further, an integrative treatment approach should focus not only on ameliorating symptoms, but on restoring mental well-being, and caring for the family system as well. Typically, a treatment plan will include attention to psychoeducational, psychotherapeutic, and psychopharmacological interventions.

Psychotherapy

One should never assume that a person with ID cannot benefit from psychotherapeutic intervention simply because of their impairment in intellectual functioning. That being said, there are a number of potential barriers to effective psychotherapeutic treatment approaches, including limited verbal ability, cognitive flexibility, psychological mindedness, and self-efficacy. To overcome these barriers, certain modifications in approach are beneficial. It is important, for example, that an active therapeutic stance be employed, as well as concrete and supportive interventions and careful attention to the language abilities and developmental level of the patient in treatment. When these types of alterations are made, many patients with ID are clearly able to benefit. Specific psychotherapeutic approaches that have been shown to be effective include behavioral (in particular, applied behavior analysis [ABA] models), cognitive-behavioral, psychoeducational, and skills training (e.g., coping skills, social skills). Behavioral therapies are demonstrably effective in managing many maladaptive behaviors, particularly aggression and self-injury, in persons with ID. Interventions that should be considered include environmental modifications, caregiver education, and ABA. There is a large empirical evidence base for ABA, and recent attention has been paid to dealing with low–base-rate behaviors, improving generalization, and incorporating choice-making to improve outcomes. Typically, a behavioral assessment begins with a “functional analysis” of behavior or a detailed examination of the variables that reinforce or maintain behaviors. One considers the antecedent events and consequences of a behavior in question and typically must test hypotheses by systematically manipulating key variables. Matson and colleagues have shown that a checklist, called the Questions About Behavioral Function (QABF), can be used successfully to derive clear behavioral functions for most individuals (84%) across target behaviors of self-injury, aggression, and stereotypy. Further, when the results of the functional analysis informed the behavior program, subjects improved significantly more than those whose behavior programs were more generic. While screening instruments may thus be quite useful in beginning to identify reinforcing variables for maladaptive behavior, the development and implementation of an effective behavior program are best conducted by a behavioral psychologist or board-certified behavioral analyst (BCBA). In addition to behavior therapies, there is some evidence for the use of cognitive-behavioral therapy to treat depression in individuals with mild ID. Skills training approaches, such as teaching individuals with ID brief relaxation techniques, may be useful in reducing anxiety even in the context of moderate to severe ID. Parent skills-training approaches have been shown to improve developmental, behavioral, and family outcomes. Supportive groups for parents and siblings (e.g., SibShop) may also be of particular benefit. Family therapy can be particularly useful and provides a setting for education, support, and for consolidation of behavioral treatment and other interventions.

An important area of current research is to examine not only efficacy but also service delivery. A recent review of a range of psychosocial interventions (behavior analytic, cognitive training and rehabilitation, and parent training interventions) delivered by nonspecialist care providers shows that interventions delivered in this way do produce good outcomes compared to no treatment or treatment as usual. This finding is critical given the scarcity of specialists in rural and low-resource areas and countries.

Pharmacotherapy

There is a relative paucity of controlled clinical drug trials that include persons with ID, particularly individuals with the most severe cognitive impairment. As a result, clinicians must generally extrapolate from the general population with regard to pharmacotherapy in persons with ID. While studies remain limited, some emerging evidence suggests that individuals with ID may experience benefits and side effects of certain classes of medications differently than individuals without ID, and individuals with ID may require relatively unique dosing strategies. Additionally, drug–drug interactions should be considered prior to a new medication trial that may affect the availability or effectiveness of concurrent medications. Multiple psychiatric problems are common in individuals with ID and as a result, polypharmacy is common. Target symptoms, along with associated data on baseline severity, frequency, and duration of such behaviors or symptoms (preferably from evidence-based objective scales or measures), should always be clearly identified before starting a medication. Regular monitoring for side effects should occur after initiating medications and adherence to treatment reviewed before increasing dosages. Studies are also making it clear that genetic copy number variants associated with both loss of function and gain of function may be associated with some ID syndromes. As these data inform more personalized or targeted treatment approaches, it is likely that some medications will emerge as being specifically beneficial or specifically contraindicated in certain ID syndromes relative to the general population.

Commonly Cited Target Symptoms and Conditions for Medication Use

AGGRESSION

Antipsychotics. Antipsychotic medications have long been used in persons with ID for indications including aggression, hyperactivity, disruptive behaviors, and self-injury. This is the most widely prescribed class of psychotropic medication for persons with ID. More recently, due to relatively common side effects, they are now being reserved for older children or adults for whom intensive behavioral interventions or other medication options have failed, but they remain among the most widely studied and utilized medications for aggression.

There is evidence to support the use of dopamine antagonists in SIB and aggression both in theory and in practice. Due to the availability of the atypical neuroleptics, use of first-generation antipsychotic medications have dramatically reduced in this population.

Numerous randomized controlled trials and open label studies appear to suggest moderate efficacy of risperidone in reducing aggression in children and adults with ID. While side effects have been generally tolerable, common adverse events can include extrapyramidal side effects including akathisia and parkinsonism, somnolence, headaches, gastrointestinal side effects such as constipation, and weight gain. Other studies have found significant improvement in aggressive behavior with aripiprazole compared with placebo and adverse effects have been generally tolerable. There exists a small number of studies reviewing olanzapine, quetiapine, ziprasidone, and clozapine involving people with ID (Roy et al., 2013). Most studies assessing antipsychotic medications, including those conducted by Research Units on Pediatric Psychopharmacology Autism Network, in ASD include children with comorbid ID, though these studies usually have not stratified efficacy results based on IQ. Notably, a study by Tyrer could not distinguish atypical or traditional neuroleptic antipsychotics from placebo for effects on aggression (Tryer et al., 2008).

Individuals with ID appear to be at greater risk of developing tardive dyskinesia than the general population, with recorded rates ranging from 18% to greater than 30% following chronic use of first-generation antipsychotics. There also appears to be some emerging evidence, along with expert opinion, that children with ASD may be particularly prone to adverse effects with certain psychotropic medications. However, spontaneous abnormal involuntary movements are not uncommon in this population which may confound some of this data. The more widespread use of second-generation antipsychotics in persons with ID have reduced the risk of tardive dyskinesia and extrapyramidal symptoms, as well as cognitive toxicity. Constipation, which is often associated with distress and agitation when unaddressed, should be monitored when starting antipsychotics especially those with anticholinergic properties.

Mood Stabilizers. There is ample evidence to support an antiaggressive effect of lithium salts in the context of ID. The side effects associated with this agent, particularly the cognitive dulling that is occasionally reported by some recipients without ID, should raise the threshold for the use of this agent in the setting of ID. Valproic acid and carbamazepine may help to reduce impulsive aggression, though some of this evidence appears to be extrapolated from studies of the general population. Other research has examined valproic acid and topiramate as adjunctive agents with positive results, but studies remain limited for use of non-lithium mood stabilizing agents in ID.

Selective Serotonin Reuptake Inhibitors (SSRIs). The existing evidence on the use of serotonin reuptake inhibitors for the management of aggressive problems in children and adults with ID is scant. Responses to the SSRIs are varied with some studies reporting favorable results while others show negative effects or worsening symptoms (Sohanpal et al., 2007).

Beta-Adrenergic Antagonists. Small studies and case reports utilizing beta-adrenergic antagonists have also been reported to be of use in the population with ID and developmental disorders. It is not clear, however, whether the mechanism is central or peripheral. A systemic review by Ward found 14 studies utilizing primarily propranolol for “problem behaviors” including aggression in ID, with doses of propranolol up to 340 mg (Ward et al., 2013). However, none of the identified studies were RCTs, and the efficacy of beta-blocker medications remains debatable.

ANXIETY AND MOOD DISTURBANCE

Antidepressants. Special considerations in the use of antidepressant medication in ID include the common medical comorbidities as noted previously. Tricyclic antidepressants (TCAs) in particular (e.g., clomipramine) must be used with the knowledge that the risk of lowering seizure threshold is relevant. Although this risk in the general population is in the order of 1 in 1,000, in individuals with ID this risk may increase to closer to 1 in 5. Cardiac anomalies are common in some ID syndromes, and the anticholinergic side effects of some medications may be particularly significant in persons with DS. Individuals with ID may require lower levels of antidepressant drugs than their normally developing peers, and disinhibition has been described even with typical antidepressant doses of the serotonin reuptake inhibitors. Constipation associated with the anticholinergic effects of some TCAs can cause significant distress which, in people with ID, may be expressed as sleep disorders, loss of appetite, agitation, and aggression.

There are some individuals—particularly a subgroup within autism spectrum disorder—who appear to exhibit extraordinary sensitivity to drugs of this and other classes. This sensitivity, when it occurs, is manifest by disinhibition, and may present as an increase in target symptoms at doses of medication that might otherwise appear quite standard. In this context, individuals may be converted to responders with significant dose reductions, for example, reducing a starting dose of fluoxetine or citalopram from 10 to 2 mg, daily. Moreover, it is increasingly the norm that initiation doses for SSRIs in this population are lower, and the rate of dose increase is slower, than for the general population.

Mood Stabilizers. Some anticonvulsant drugs may improve cyclical mood disorders and impulsive aggression. Valproate has probably been studied most in this regard. Similarly, there is evidence to suggest that in the setting of cyclical mood disturbance, lithium may also be helpful.

Anxiolytics. Although benzodiazepines are commonly prescribed in treatment for anxiety in the general population, there are unique concerns when they are used in the context of developmental disorders, particularly regarding the possibility of increased confusion, cognitive impairment, unsteadiness, and paradoxical excitement. Barron and Sandman reviewed disinhibition associated with benzodiazepines that occurred in 35% to 68% of an institutionalized intellectually disabled population as compared to controls. Nevertheless, alprazolam, clonazepam, and lorazepam are widely used for the treatment of acute anxiety, particularly that associated with procedures. Their use should be considered in the absence of evidence of previous abnormal responses to these agents.

On the other hand, recent findings from animal modeling of haploinsufficiency of the SCN1A gene (Dravet syndrome) by Catterall and colleagues suggest that activation of GABA(A) receptors with low-dose clonazepam improves behavioral and cognitive impairments in mice. There is growing interest in this mechanism in genetic conditions associated with ID.

Buspirone. Buspirone is another serotonergic agent that has been reported to be of benefit in some persons with developmental disorders. Ratey and colleagues reported the use of buspirone in an open trial in persons with diagnosed anxiety disorders manifested by aggressive behaviors and SIBs. A study by Verhoeven and Tuinier also explored buspirone in the treatment of anxiety in persons with ID, suggesting that this agent should be among those considered in treating this population. Moreover, advantages include a relatively benign side-effect profile, specifically the absence of common motor or cognitive toxicities, at doses that are used to treat anxiety.

SELF-INJURIOUS BEHAVIORS (SIBS)

SSRIs. Trials of serotonin reuptake inhibitors in clinical practice are increasingly common among patients with SIB. Favorable results have been reported for fluoxetine, paroxetine, sertraline, and trazodone in this regard but all reports involve small numbers of subjects. While clomipramine has been shown to be useful in a well-controlled study, due to its effect of lowering seizure threshold, clomipramine is generally not a first-line treatment for compulsive SIB in individuals with ID who may have comorbid epilepsy.

Antipsychotics. One retrospective review has suggested that atypical antipsychotics improve aggressive behavior but not SIBs (Ruedrich et al., 2008). Small studies exist for some neuroleptic options (e.g., risperidone, olanzapine) with some encouraging results. Small studies and case reports involving clozapine have also shown some reduction in SIB. However, the literature involving antipsychotics for behavioral problems is generally lacking in outcome measures specific to self-injury. Although suggested by some, there is no convincing evidence that the mechanism of action of antipsychotics in SIB (or aggression) is merely to suppress behavior generally through a nonspecific sedating effect. Such an outcome is clearly undesirable in an individual with pre-existing cognitive impairment, and many of the earliest reports specifically note the absence of sedation.

Mood Stabilizers. There is evidence for the use of Lithium both as a monotherapy and concurrently with other psychotropics in reducing SIB. However, the risks of side effects and monitoring requirements may restrict use of this medication. Less evidence appears to exist for other options including valproate and topiramate though both have been studied with reductions observed in adult patients with ID.

Opioid Antagonists. Some individuals who self-injure appear to have altered pain sensitivity, as was suggested nearly a century ago by Ireland. This observation, coupled with data indicating that opioid antagonists can attenuate stereotypy and self-injury in animal models and data indicating that opioids may modify the function of dopaminergic systems, has fueled interest in this class of drugs. Naltrexone is the opioid antagonist most widely used for SIB, but the literature thus far is mixed. A 2013 Cochrane Review found some evidence of positive results for the indication of SIB in ID, but evidence was not conclusive (Rana et al., 2013). On balance, naltrexone appears to be well tolerated in persons with developmental disorders and IDs, with sedation as the side effect most likely to be observed.

NMDA Antagonists. A growing appreciation exists for the potential therapeutic utility of drugs acting at glutamate receptor subtypes. Glutamatergic and dopaminergic interactions in the neostriatum are the focus of research on the pathogenesis of a host of neuropsychiatric illnesses, including schizophrenia, obsessive-compulsive disorder, SIB, and aggression.

Dextromethorphan, an antitussive agent that also binds to the *N*-methyl-D-aspartate (NMDA) receptor, was reported by Welch and Sovner to have markedly attenuated SIB in a 25-year-old individual with congenital rubella syndrome. The patient was maintained on dextromethorphan for 16 months with sustained benefit. It is surprising and perhaps ominous that no follow-up studies have been reported with this drug in SIB. Lamotrigine is an anticonvulsant drug that also appears to antagonize glutamate by reducing its release. A recent case report suggested that lamotrigine may also be useful in reducing self-injury in the context of a stereotyped movement disorder.

Others. SIB and aggression can also dramatically improve with the withdrawal of phenobarbital. The potential behavioral “toxicity” of this and any drug should not be overlooked in designing treatment strategies for persons with ID. Moreover, medical comorbidity is the rule in individuals with developmental disabilities, and the importance of identifying and treating underlying medical problems or refining that treatment cannot be overstated.

OTHER OBSESSIVE OR COMPULSIVE BEHAVIORS. Of interest, there is some experience with a small number of persons with PWS treated with the anticonvulsant topiramate whose skin-picking behavior improved. The mechanism underlying this effect is unknown, and it might even involve peripheral effects of the drug, for example, on sensation, or perhaps even rate of wound healing. For PWS, obsessive-compulsive disorder occurs more commonly than in the general population, and compulsive skin picking may improve along with other compulsive behaviors with an SSRI.

SLEEP DISTURBANCE. Sleep disturbance is a little-studied but often significant problem for many persons with ID. Jan prescribed melatonin openly to children with moderate to severe ID of mixed etiologies. The children ranged in age from 1 to 11 years. All but two of the children were characterized as dramatic responders based on measures of hours of sleep, nocturnal awakenings, delayed sleep onset, and early-morning awakening. Melatonin was administered as a single 3-mg dose 1 to 2 hours before target bedtime each night. Pillar and colleagues also administered a single 3-mg dose and monitored sleep efficiency, as well as activity via wrist actigraph, in their sample of five children (3.5 to 13 years old) with severe ID. Sleep was significantly improved in the subjects receiving melatonin, and two of the subjects were still receiving melatonin up to 18 months after the completion of the study. This preliminary evidence suggests that melatonin deserves broader consideration for the treatment of children with ID and disturbed circadian rhythm of sleep.

SEIZURE DISORDERS. Crawford and colleagues performed an open-label but randomized, parallel-group, multicenter, add-on study comparing gabapentin with lamotrigine in 109 patients with treatment-resistant focal epilepsy and ID (Crawford et al., 2001). Subjects in both the gabapentin and lamotrigine groups experienced a reduction in seizure frequency by nearly half. Both drugs were equally well tolerated. Gabapentin may have had an edge with respect to some behavioral indices, denoted by the investigators as communication, cooperation, and restlessness, but lamotrigine also was credited with improving challenging behavior. Thus, both medications may hold promise for persons with ID and treatment-resistant epilepsy, and their effects on behavior should be the focus of additional study.

A larger study in which lamotrigine was added to the anticonvulsant regimen of children with ID suggested that those with autism experienced gains that could not readily be attributable to better seizure control alone. A recent double-blind, placebo-controlled trial of lamotrigine in autism, however, did not demonstrate significant results.

ADHD AND DISRUPTIVE BEHAVIORS. Despite reports of paradoxical responses to stimulant medications in persons with ID, with higher-than-expected rates of emergent motor tics and emotional lability, a growing body of literature supports the use of stimulant drugs for the treatment of ADHD in the context of ID. The available literature appears to suggest that overall, the effect size of methylphenidate in children with ID is lower than in children without ID. A higher dose of methylphenidate may be associated with a higher rate of adverse effects. Significant adverse effects include sleep difficulties, poor appetite, weight loss, irritability, social withdrawal, and increased motor activities including tics. A problem awaiting additional study concerns the likelihood of a response to stimulants in relation to the severity of cognitive disability. Data are mixed with respect to whether IQ negatively or positively predicts treatment response.

Atomoxetine is a nonstimulant that is used for similar target symptoms in the general population. There are no studies specifically examining its effectiveness in ID, but growing evidence suggesting that in autism spectrum disorders, atomoxetine may be useful for target symptoms of inattention, hyperactivity, and noncompliance/oppositionality.

An open trial of the use of amantadine in children with various developmental disabilities and disruptive behaviors was quite promising. A subsequent small multisite, controlled trial in children with autism yielded mixed results. Memantine has also shown promise in open trials for agitation in the context of autism.

Considerable evidence is accumulating to support the effectiveness of risperidone in treating severe disruptive behavior in children and adolescents with cognitive disability. Sedation and weight gain were the most common side effects. Similar studies exist for aripiprazole, although both controlled trials were sponsored by the manufacturer. A 2017 Cochrane Review found a lack of evidence to support the use of quetiapine, ziprasidone, or any other atypical antipsychotics besides risperidone for disruptive behavior disorders in children and youth and no evidence for children under 5 years of age although this review was not specific to children with ID.

COGNITION. There is no evidence that medications, nutritional supplements, or other theorized nootropics have demonstrated improved cognition in ID. However, there are a growing number of strategies emerging from animal model and genetic identification studies that may eventually redefine this class of therapeutics for the population with ID. Some early studies suggest possible enhancement of cortical plasticity in animal models for fetal alcohol syndrome. Medications with indications for major neurocognitive disorders, such as acetylcholinesterase inhibitors, do not have evidence for use in ID. The efficacy for various nutritional supplements, though sometimes popular in practice, has not been well established in clinical trials. For example, a randomized controlled trial of antioxidants and/or folinic acid for infants with DS found no evidence of cognitive benefit. Another randomized controlled trial found no evidence of cognitive benefit for a popular synthetic supplement piracetam in children with DS but did find numerous adverse effects. Even as some novel treatments progress to clinical trials, the question as to whether there are particular developmental windows for therapeutic effectiveness remains important to address.

Services and Supports

In contrast to even 20 years ago, individuals with ID can now access a wide variety of services. These services can be found throughout the United States and last throughout the lifespan; see Table 39–10 for a listing of some of these services.

Early intervention and educational services became widespread following the passage of PL 94–142 (the Education for all Handicapped Children Act of 1975); this legislation has recently been extended and modified by IDEA, the Individuals with Disabilities Education Act (IDEA) of 1990, with amendments in 1997 and 2004.

Early Intervention. The IDEA part C provides federal grants to states to provide early intervention services for young children from birth until a child’s third birthday. Such services are provided by different “lead agencies” coordinating in each state and as a result services may vary widely between states. For instance, in some states, birth to 3 (B-3) care may begin with early intervention specialists providing services, such as speech and language therapy, in the child’s home, while in other states B-3 care can also entail center-based services. Once the child is preschool age, most services are provided through the public school district. These services are center-based and include both one-on-one therapies and small group education services. B-3 and preschool services focus on multiple domains of functioning, including cognitive development, language, motor, social, and adaptive functioning.

With the passage of Part H of PL 99–457 (the Education of the Handicapped Amendments of 1986) early intervention services have increasingly emphasized the family's role. Under IDEA, agencies are required to develop an Individualized Family Service Plan (IFSP) for each child and family receiving early intervention services, focusing on family strengths and needs in caring for their young child with ID. An IFSP is similar to an Individualized Education Program (or Plan), called an IEP (see below), in that both contain goals that are regularly monitored and updated as goals are reached, and new goals are written. Unlike an IEP, however (1) Eligibility for early intervention as per Part C of IDEA is determined by state eligibility criteria (which are variably defined) and is often different from eligibility for special education under Part B of IDEA; (2) The IFSP includes both child and family goals and objectives as opposed to an IEP which is focused on the individual student; (3) IFSP includes services to help a family in natural environment settings, not just in daycare/preschool, but at home and in the community; (4) The IFSP includes services to help a family in vivo in the child's natural environments in the home or community as opposed to solely school environments; (5) The IFSP team involves a service coordinator who assists the family whereas students with an IEP have a case manager, usually a special educator at the school.

Table 39–10.
Service Needs and Resources for Families of Children at Different Ages

	Needs	Resources
Ages 0–3 years		
Child	Evaluation: physical, motor, cognitive, linguistic, social-emotional; early intervention services	Multidisciplinary evaluation, which results in an individualized family service plan (IFSP), with child and family receiving either center- or home-based early intervention services for set number of hours per week
Mother	Emotional support, caretaking behaviors	Support groups by disability, region, and etiology; part of early intervention evaluation, intervention, and IFSP
Family	Support, financial assistance, information	Support groups; depending on the problem, state developmental disabilities or insurance payment for some services; hospitals, agencies, groups
Ages 3–21 years		
Child	Evaluation, referral, and individualized education program (IEP)	School system: involves legal process of evaluation and placement (notification, hearings, appeals if necessary); information on “transition” to adult services as child nears

		age 21 years (and school services end).
Family	Information, financial assistance, support	Local and national groups; departments in some states, including respite care, camps, art (Very Special Arts) or athletic activities (Special Olympics), scholarships for adolescents with some disabilities (deafness, blindness)
Older than 21 years		
Offspring	Residential services, work	Run by state developmental disability departments; parents and offspring have major say concerning whether residential or work placements are appropriate.
Family	Support, information, guardianship issues	Continuation of many of the services provided during the school years, particularly for individuals with severe disabilities, provisions for residential and work status after parents can no longer serve as offspring's legal guardians

Reprinted with permission from Hodapp RM. *Development and Disabilities*. Cambridge University Press; 1998:196.

School. From ages 3 to 21, specialized education services are provided through local school districts. The IDEA requires that public school systems serve all children with disabilities and outlines the procedures and legal provisions to do so (see Table 39–11). This process includes an assessment of the child’s academic performance, social/behavioral functioning, and adaptive functioning, and provides annual goals and instructional objectives. The IEP is the result, a legal document that binds the school district to provide the services described in the IEP. Educational services have become a right—not a privilege—for parents of children with intellectual and developmental disabilities.

Table 39–11. Summary of IDEA

Major Provisions	Description
Zero Reject Policy	▶ Schools must educate all children with disabilities (see covered disabilities in Table 39–12) ▶ A state's education agency is responsible for locating, identifying, and evaluating all children, from birth to age 21, who are suspected of having disabilities (**child find system)**
Nondiscriminatory Identification and Evaluation	▶ Schools must use nonbiased, multifactored methods of evaluation to determine whether a child has a disability and whether special education is needed. ▶ **Protection in evaluation procedures:** evaluation procedures must not discriminate on the basis of race, culture, or native language.
Free, Appropriate Public Education (FAPE)	▶ All children with disabilities must receive a free, appropriate public education provided at public expense. ▶ An Individualized education program (IEP) must be created. ▶ An IEP must specify the child's unique educational needs, levels of performance, measurable annual goals and objectives, and services that will be provided to help the child attain those goals.
Least Restrictive	▶ Students with disabilities must be educated with children without

Environment (LRE)	disabilities to the maximum extent appropriate including with aide, assistance or other services. ▶ The IEP must contain a justification when the nature or severity of a child’s disabilities is such that they cannot receive an appropriate education, participate in extracurricular activities, or receives services such as transportation with nondisabled peers.
Due Process Safeguards	▶ Consent from a child’s guardian(s) must be obtained for all subsequent evaluations and placement decisions regarding special education. ▶ Schools must maintain confidentiality for students, preserve records and make records available for guardians on request. ▶ Guardians may obtain an independent evaluation at public expense if there is disagreement in the evaluation. ▶ When there is disagreement on the identification, evaluation, placement, or provision of a free, appropriate public education for the child, the guardians may request mediation or a “due process” hearing.
Parent and Student Participation and Shared	The design and implementation of special education services must be

Decision Making	done in a collaborative manner with the student and their guardians.

An element of IDEA involves the right to a public education “within the least restrictive environment.” This idea of least restrictive environment means that the child will be placed in the same settings as typically developing children insomuch as their disability will allow them to meet IEP goals in that setting. Full “inclusion” conceptualizes *all* education provided alongside typically developing children as every child’s right, while proponents of “least restrictive environment” posit that the environment is determined by the child’s individual educational needs.

Special educators have long advocated “non-categorical” approaches to children with ID. Such approaches do not consider the cause of the child’s disability, and even knowing about the etiology of the disability has been considered harmful, possibly leading to increased stigmatization. However, advocates for “etiology-based” educational interventions suggest that the cause of the child’s ID can influence behavior, and relative strengths (or weaknesses) in language versus other abilities and across cognitive domains. Indeed, some syndrome-specific organizations (e.g., DS, PWS, Williams syndrome) have published best practice guidelines for both medical and psychosocial interventions, though such guidelines for rarer syndromes are lacking.

For example, children with DS show difficulties in linguistic grammar, expressive language, and articulation, even as their abilities in visual short-term memory appear relatively strong. Children with Williams syndrome generally show higher language (verbal) than visuospatial abilities, while children with either Prader–Willi or FXS show stronger abilities in simultaneous (i.e., Gestalt) processing compared to sequential processing. In each instance, educators and test-constructors have devised specific, tailored educational approaches that play to the child’s strengths rather than focusing on improving weaknesses alone. Similar approaches might also be designed and prove effective for children with different ID syndromes. However, while “etiology-based” educational interventions hold promise, such approaches may not prove effective for every child with a particular syndrome and need to be considered in light of the individual child’s strengths and weaknesses.

Adult Services. Federally mandated school services serve students with disabilities from age 3 to age 21. There is a federal mandate requiring youth with disabilities age 16 and older to have a postsecondary transition plan written into their IEP. This plan typically includes teaching vocational and adaptive skills in support of long-term life goals. Students with ID, their parents, and the school team work together to determine the most appropriate long-term goals for each young adult.

If a student with an IEP does not graduate with a high school diploma, they are eligible for the school district’s “transition program.” Each district’s transition program is unique, but most contain some degree of work experience in the community. Once an individual with ID turns 21, service delivery comes under the auspices of each state’s department of developmental disabilities. In contrast to educational services, adult services are not federally mandated. States thus vary widely in their services and how they offer them. Services are often spread out across various locations, and eligibility and access to services can be difficult.

During the adult years, residential and vocational services predominate. Residential services run the gamut from institutional to community-based, and from constant to no supervision. More restrictive placements involve large or small institutional residences, which mainly serve the most severely impaired individuals. Group homes for four to eight residents are probably the most common residential option; these homes typically exist within residential neighborhoods and provide round-the-clock staff to ensure safety and consistent programming. Supervised living arrangements usually consist of apartments that are shared by a few persons with ID; these residents receive visits from program staff, who often instruct residents in cooking, cleaning, and shopping. Unsupervised apartments, where individuals live by themselves or with a roommate under little supervision, mainly succeed with clients who have milder intellectual impairments and fewer behavioral problems.

Vocational services also range from most to least restrictive. Most restrictive are sheltered workshops, in which individuals with ID assemble or test products, working next to others with disabilities. Supported employment involves a job coach or other specially trained individual to work with the client to find a job, to learn how to take public transportation to get to a job, and to acquire the technical and social skills needed to handle the job. These coaches often initially shadow the individual with ID at the workplace and provide support as needed. Competitive employment is the least restrictive job option for adults with ID and involves a job with the same pay levels and benefits that a nondisabled person would have, but also with the same pressures and responsibilities for work performance. As in unsupervised apartment living, competitive employment is an option for the least impacted individuals.

Supports

In addition to services provided by early intervention, schools, and adult service agencies, individuals with ID and their families can benefit from other supports, including short- or longer-term respite care (allowing families a break from full-time care), specialized summer camps, and programs such as Special Olympics. In addition, most states have specialty clinics for behavioral and medical management of certain types of ID, such as autism spectrum disorder.

A wide variety of organized groups are available for parents of children with disabilities and for specific types of disability. Sibling groups and networks are also available, and the federal government fosters service and advocacy through the University Affiliated Programs (UAP’s) in almost every state.

Legal Issues in Intellectual Disability

FORENSIC ISSUES FOR INDIVIDUALS WITH ID. There is some evidence to suggest that individuals with borderline IQ and mild ID are overrepresented within the criminal justice system. Individuals with more severe or profound ID may be "protected" by factors including supervision from caregivers and a reduced range of social activities and community interface. Persons with ID are also more likely to be the victims of crime and other forms of abuse and maltreatment.

Relevant Legislation. Early intervention and educational services became widespread following the passage of PL 94–142 (the Education for all Handicapped Children Act of 1975); this legislation was extended and modified by IDEA, the IDEA of 1990, with amendments in 1997 and 2004. Major provisions of IDEA are outlined in Table 39–11.

ID is considered one of 13 disabilities defined under IDEA (see Table 39–12). Section 504 of the ADA protects anyone with a "physical or mental impairment which substantially limits a major life activity," and who "has a record of such impairment," or is "perceived as having such impairment" from discrimination. Students who do not qualify for an IEP can still receive classroom accommodations under a "504 Plan" according to ADA.

The ADA of 1990 spells out four aims: (1) to provide a clear and comprehensive national mandate for the elimination of discrimination against individuals with disabilities; (2) to provide clear, strong, consistent, enforceable standards addressing discrimination against individuals with disabilities; (3) to ensure that the Federal Government plays a central role in enforcing the standards established in the Act on behalf of individuals with disabilities; (4) and to invoke the sweep of congressional authority, including the power to enforce the 14th Amendment and to regulate commerce, in order to address the major areas of discrimination faced day-to-day by people with disabilities. The ADA represents a major turning point in the assurance of the rights of persons with disabilities who, even 50 years ago, were frequently institutionalized and compulsorily sterilized in misguided efforts to protect society from this population. In contrast, the prevailing paradigms in developed countries are those of community integration and support. Disabilities law over the past three decades ensures that all children with disabilities receive a "free appropriate public education" in the "least restrictive environment," and that reasonable accommodation is made for the employment of adults. Clinicians working with this population have a personal obligation to provide an accommodating environment, and should familiarize themselves with local guardianship, informed consent, and abuse reporting laws to best serve this population.

Table 39–12.
Disabilities Defined by IDEA

Autism
Deafness
Blindness or other visual impairment
Developmental delay
Emotional disturbance
Hearing impairment
Intellectual disability
Multiple disabilities
Orthopedic impairment
Other health impairment
Specific learning disability
Speech or language impairment
Traumatic brain injury

FURTHER READINGS

American Association on Intellectual and Developmental Disabilities. Mental Retardation: Definition, Classification, and Systems of Supports. Washington, DC: Author; 2002. Atkins v. Virginia. 536 U.S. 304 2002.

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. (DSM-5). American Psychiatric Association; 2013.

Brosco JP, Mattingly M, Sanders LM. Impact of specific medical interventions on reducing the prevalence of mental retardation. *Arch Pediatric Adolesc Med*. 2006;160:302–309.

Crawford DC, Acuna JM, Sherman SL. FMR1 and the fragile X syndrome: human genome epidemiology review. *Genet Med*. 2001;3:359–371.

Decobert F, Grabar S, Merzoug V, et al. Unexplained mental retardation: is brain MRI useful? *Pediatr Radiol*. 2005;35(6):587–596.

de Ruiter KP, Dekker MC, Verhulst FC, Koot HM. Developmental course of psychopathology in youths with and without intellectual disabilities. *J Child Psychol Psychiatry*. 2007;48(5):498–507.

Ellis JM, Tan HK, Gilbert RE, et al. Supplementation with antioxidants and folinic acid for children with Down's syndrome: randomised controlled trial. *BMJ (Clinical Research ed.)*. 2008;336(7644):594–597.

Engbers HM, Berger R, van Hasselt P, et al. Yield of additional metabolic studies in neurodevelopmental disorders. *Ann Neurol*. 2008;64(2):212–217.

Esquirol, E. *Mental Maladies; A Treatise on Insanity*. Lea and Blanchard; 1845.

Gothelf D, Furfaro JA, Penniman LC, Glover GH, Reiss AL. The contribution of novel brain imaging techniques to understanding the neurobiology of mental retardation and developmental disabilities. *Ment Retard Dev Disabil Res Rev*. 2005;11:331–339.

Hodapp RM. *Development and Disabilities: Mental, Motor, and Sensory Impairments*. Cambridge University Press; 1998.

Individuals with disabilities education act (IDEA). Individuals with Disabilities Education Act, sites.ed.gov/idea/

Jarrold C, Baddeley AD, Hewes AK, Phillips C. A longitudinal assessment of diverging verbal and non-verbal abilities in the Williams syndrome phenotype. *Cortex*. 2001;37:423–431.

Loy JH, Merry SN, Hetrick SE, Stasiak K. Atypical antipsychotics for disruptive behaviour disorders in children and youths. *Cochrane Database Syst Rev*. 2012;(9):CD008559. Update in: Cochrane Database Syst Rev. 2017 Aug 09;8:CD008559.

Martin A, Bloch MH, Volkmar FR, Dykens E, Hodapp R. Intellectual disability. In *Lewis's Child and Adolescent Psychiatry: A Comprehensive Textbook*. 5th ed. Wolters Kluwer; 2018:434–441.

McQuire C, Hassiotis A, Harrison B, Pilling S. *Pharmacological Interventions for Challenging Behaviour in Children With Intellectual Disabilities: A Systematic Review and Meta-Analysis*. BMC\Park Press; 1976:361.

Pathirathna ML, Wimalasiri KMS, Sekijima K, Sadakata M. Maternal compliance to recommended iron and folic acid supplementation in pregnancy, Sri Lanka: A hospital-based cross-sectional study. *Nutrients*. 2020;12(11):3266.

Rana F, Gormez A, Varghese S. Pharmacological interventions for self-injurious behaviour in adults with intellectual disabilities. *Cochrane Database Syst Rev*. 2013;(4):CD009084.

*Reiss S, Aman MG, eds. *Psychotropic Medication and Developmental Disabilities: The International Consensus Handbook*. OSU Nisonger Center; 1998.

Roy D, Hoffman P, Dudas M, Mendelowitz A. Pharmacologic management of aggression in adults with intellectual disability. *J Intellectual Disability–Diagnosis Treatment*. 2013;1(1):28–43.

Ruedrich SL, Swales TP, Rossvanes C, Diana L, Arkadiev V, Lim K. Atypical antipsychotic medication improves aggression, but not self-injurious behaviour, in adults with intellectual disabilities. *J Intellect Disabil Res*. 2008;52(Pt 2):132–140.

Sedel F, Baumann N, Turpin JC, Lyon-Caen O, Saudubray JM, Cohen D. Psychiatric manifestations revealing inborn errors of metabolism in adolescents and adults. *J Inherit Metab Dis*. 2007;30(5):631–641.

Sohanpal SK, Deb S, Thomas C, Soni R, Lenôtre L, Unwin G. The effectiveness of antidepressant medication in the management of behaviour problems in adults with intellectual disabilities: a systematic review. *J Intellect Disabil Res*. 2007;51(Pt 10):750–765.

Sparrow SS, Balla D, Cicchetti DV. *Vineland Adaptive Behavior Scales*. American Guidance Service; 1984.

Thapar A, Simonoff E. Intellectual disability. In *Rutter's Child and Adolescent Psychiatry*. 6th ed. Wiley-Blackwell; 2015:719–733.

Tyrer P, Oliver-Africano PC, Ahmed Z, et al. Risperidone, haloperidol, and placebo in the treatment of aggressive challenging behavior in patients with intellectual disability: A randomised controlled trial. *Lancet*. 2008;371:57–63.

Vilaseca R, Rivero M, Bersabé RM, et al. Demographic and parental factors associated with developmental outcomes in children with intellectual disabilities. *Front Psychol*. 2019;10:872.

Ward F, Tharian P, Roy M, Deb S, Unwin GL. Efficacy of beta blockers in the management of problem behaviours in people with intellectual disabilities: A systematic review. *Res Dev Disabil*. 2013;34(12):4293–4303.

Zablotsky B, Black LI, Maenner MJ, et al. Prevalence and trends of developmental disabilities among children in the United States: 2009–2017. *Pediatrics*. 2019;144(4):e20190811.

40

Specific Learning Disorder

Oren Boxer, Ph.D.

Humans are not born to read, write, spell, or perform arithmetic calculations: These are skills that must be learned. In contrast to oral language skills, which humans have possessed for over 100,000 years, cultural symbols used to represent specific numbers and speech sounds were invented only a few thousand years ago. Reading was invented approximately 5,400 years ago; symbolic arithmetic is more recent since the Arabic notation was not available even a thousand years ago.

These arbitrary cultural symbols, which are the foundational codes for reading, writing, and arithmetic, are not acquired “naturally” by simple exposure; each human being has to learn these symbols as well as how they map onto specific speech sounds, specific quantities, and numerals. The process of developing these literacy and numeracy skills reorganizes and changes everyone’s brain. This is because the human brain did not evolve to accommodate the recent human invention of symbols, which are far too recent to have exerted any evolutionary pressure on brain evolution. Thus, during the human developmental period, existing functions and networks in the brain are “recycled” for the new uses. For example, learning to read involves the creation and automation of an efficient interface between vision and spoken language (“visual word form area” in the left visual cortex); learning arithmetic involves creating connections with nearby areas of the parietal cortex that code for space and eye movement.

Multiple skills are involved in learning to process one's cultural symbol systems. These include the awareness that spoken language can be segmented into smaller units (words, syllables, phonemes) and mapped onto written, visual symbols (e.g., letters, characters, numerals); the ability to rapidly identify letters, letter clusters, and words (orthographic awareness); the ability to extract meaning from this written language; and the ability to transcode meaning between symbol systems (e.g., to understand that the symbol "3," the quantity of three dots (***), the spoken word "three" (/th/ /r/ /ee/) and the written word "three"; all represent "threeness" (see Table 40–1 for definitions of key terms related to reading and number sense).

For a subgroup of children, at least 6%, the multidimensional developmental task of learning to process and use cultural symbols present major lifelong challenges that are unexpected, given their abilities in other developmental domains. The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) uses the term "Specific Learning Disorder" (SLD) to describe persistent and impairing problems in acquiring and using the cultural symbols that are required for reading, writing, and arithmetic.

Table 40–1.

Definition of Key Terms Used in Relation to Specific Learning Disorder

Term	Definition
Alphabetic principle	The insight that words are made up of sounds and each sound can be signified by a symbol.
Phoneme	Smallest sound units in a word: The smallest unit of speech that distinguishes one word from another (e.g., the word “peach” has three phonemes: /p/ /ea/ /ch/; the phoneme /p/ distinguishes the word “peach” from “teach”).
Phonemic awareness	An explicit awareness that there are discrete speech sounds (phonemes) in speech. It begins to emerge between ages of 2 and 4 years, takes several years to develop, and is crucial to success in reading and spelling. For example, the child must be able to isolate the phoneme /p/ in words such as “pet,” “peach,” and “pear” to fully understand that the grapheme “p” represents this sound and to associate the sound with this letter when attempting to decode a word like “port.”
Phonologic awareness	An explicit awareness that spoken language comprises discrete units, ranging from words and syllables to smaller subsyllabic units of onsets and rimes, and phonemes. It includes phonemic awareness.
Phonologic coding	Translating the letters or spelling patterns of a written word into speech patterns to identify the word and gain access to its meaning.

Phonics	An instructional approach for teaching children the systematic relationship between letters and sounds and how to use that system (alphabetic principle) to read words. In other words, instruction in phonics requires the child to have some level of phonemic awareness.
Onset-rime	Parts of monosyllabic words in spoken language: These units are smaller than syllables but larger than phonemes. An “onset” is the initial consonant sound of a syllable (the onset of “peach” is “p-”). The rime is the part of the syllable that contains the vowel and all that follows it (the rime of “peach” is “-each”).
Orthography	The representation of the sounds of a language by written or printed symbols. Patterns link graphemes (letter) to sounds (phonemes). Orthography describes or defines the set of symbols (e.g., graphemes, diacritics) used and the rules about how to write these symbols.
Orthographic transparency	Deep orthographies are writing systems (such as English) that do not have a one-to-one correspondence between the spoken phoneme and the written grapheme. A “shallow” orthography means that the correspondences between letters and sounds (graphemes/phonemes) in the writing system are close to one-to-one (e.g., Finnish, German).

SIGNIFICANCE OF SLD FOR PSYCHIATRISTS

SLD has high comorbidity rates with other common psychiatric disorders, ranging from attention-deficit/hyperactivity disorder (ADHD) (most common), anxiety, and mood disorders, and suicide, to substance abuse (least common). Thus, clinicians need to consider the presence of SLD when assessing for these other psychiatric disorders. Also, the burgeoning concern about health literacy and health numeracy in the medical literature highlights the need for clinicians to understand the medical implications of SLD. Health literacy is defined as the ability to obtain, process, and understand basic health information and services needed to make appropriate health decisions; health numeracy is defined as the individual-level skills needed to understand and effectively use quantitative health information (e.g., dose, quantity, probability) to guide health behavior and decision making. Basic reading skills (e.g., letter–sound knowledge and word recognition), reading comprehension, and basic numeracy skills are core attributes that underlie health literacy and numeracy. Poor health literacy and numeracy are independent predictors of health disparities, worse health outcomes, and mortality.

Recent surveys of the readability of patient education materials and medical research consent forms typically indicate that many of these materials have higher-than-acceptable levels of readability for the public. For example, the American Academy of Pediatrics reported that at least half of their materials manifest unacceptably high readability levels. Few studies have investigated the readability of patient education materials in psychiatry, which is of particular concern, given the low educational attainment of many patients with mental health illnesses. Low numeracy skills constrain patient choice, reduce medication compliance, impair risk communication, and affect medical outcomes. SLD tends to run in families, so patients and their families may have great difficulty accessing patient education materials and comprehending dosage instructions or nonpharmacologic interventions. Clinicians in the workplace may also have SLD themselves (e.g., almost 2% of medical students declare a diagnosis of SLD in reading upon entry to medical school), which may pose additional challenges in the effective performance of their clinical duties. Thus, an understanding of SLD and knowledge of effective strategies to enhance the clinical performance of clinical trainees and certified clinicians with SLD is of key importance for clinical educators.

To understand SLD, it is necessary to understand how the neuroplasticity of the human brain makes it possible to learn to read, write, and calculate, how the acquisition of literacy and numeracy transforms the brain, and how disturbances in brain development disrupt the normal acquisition of reading, writing, and arithmetic. Two of the most engaging accounts of the history and neuroscience are Maryanne Wolf's Proust *and the Squid* written in *2007,* and Stanislas Dehaene's *Reading in the Brain*, published in 2009. Two additional accounts of the earliest invention of numerals and how the human brain learns to understand and manipulate numbers, are Stanislas Dehaene's book, *The Number Sense*: *How the Mind Creates Mathematics,* written in 2011, and *Finding Zero*, by Amir Acxel, a mathematician's personal account of his journey to find the first "zero"—the keystone of our current number system, which he located on an inscribed stone in an old shed in Cambodia.

DEFINITION

SLD differs from most other mental disorders recognized by DSM-5, in that not only is it a clinical condition recognized and defined in the medical field, but also it is a discrete category of special educational needs (a.k.a., "exceptionalities") recognized and defined in the field of education. One major challenge for clinicians, educators, parents, and researchers alike, is that the purpose of classification differs in the medical/mental health and educational systems, as do their underlying concepts and definitions of SLD. Thus, clinicians who diagnose SLD must understand both the medical and educational definitions, as well as their different implications for accessing school-based supports and intervention services.

Medical/Psychiatric Definition

According to DSM-5, SLD is a Neurodevelopmental Disorder that involves marked and persistent difficulties in learning and using one’s cultural symbol systems (e.g., alphabetic letters, characters, Arabic numerals) that are required for skilled reading, writing, and arithmetic. SLD disrupts the normal pattern of learning these essential academic skills and is not simply a consequence of an intellectual disability, sensory deficits, other mental or neurologic disorders, lack of opportunity of learning or proficiency in the language of instruction, or inadequate educational instruction. Difficulties learning keystone (foundational) academic skills of reading, writing, and arithmetic may also impede learning in other academic subjects (i.e., history, geography, science) but those problems are attributable to difficulties learning the basic underlying academic skills.

Neurodevelopmental disorders are complex and heterogeneous conditions arising from perturbations of the central nervous system, typically during the pre- and perinatal period, which interfere with the development of language and speech, motor skills, attention, behavior, memory, learning, or other neurologic functions. Neurodevelopmental disorders are typically heritable, show early onset, persist through the lifespan, and impair academic, social, occupational, and mental health outcomes. DSM-5 delineates six neurodevelopmental disorders, many of which often co-occur: Intellectual Disabilities, Communication Disorder, Autism Spectrum Disorder, ADHD, SLD, and Motor Disorders. The neurodevelopmental disorder of learning (i.e., SLD) represents the extreme bottom end of normal distributions of culturally acquired skills, rather than a discrete categorical entity.

Like other neurodevelopmental disorders, SLD is understood to be a heterogeneous condition, in which the learning difficulties may be either selective (e.g., occur only with learning one specific aspect of reading such as word decoding) or broad (e.g., occur in learning to decode words, spell, understand what one has just read, as well as in arithmetic calculation), and frequently coexists with other neurodevelopmental (e.g., with Communication Disorders, ADHD) or psychiatric traits or disorders (e.g., with anxiety, anxiety disorders).

Moreover, SLD is understood to be a heritable, chronic condition of childhood that typically persists into adulthood, albeit with changes in the ways the learning difficulties manifest.

Medical practitioners must understand that a medical diagnosis of SLD, using the DSM or ICD classification systems, does not necessarily mandate eligibility for special assistance at school.

Federal/Educational Definition

In the field of education, the term “Learning Disabilities” is used to delineate a specific category of “special learning needs,” that mandates the provision of special education and other services to permit more equitable access to the curriculum and promote their learning.

In the United States, the conceptualization and definition of “Learning Disabilities” has been strongly influenced by the federal definition adopted by the U.S. Office of Education in 1969, operationalized by the U.S. Office of Education initially in 1977 (Individuals with Disabilities Education Act [IDEA]), and reauthorized in 2006 to determine eligibility for special education services that must be provided by all public schools. The 1977 IDEA definition of a specific learning disability refers to:

> A disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations. The term includes such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental dysphasia.

Note that the federal definition of Learning Disabilities includes oral or spoken language impairments, which are classified as Communication Disorders in psychiatric taxonomies and differentiated from SLD. Moreover, the educational definition incorporates notions of presumed etiology (i.e., deficits in underlying psychological processes), whereas psychiatric classification does not.

Identification of specific learning needs in the educational system mandates eligibility for special education services. Due in part to differences in the purpose of identification, the educational system may identify some children as having special educational needs who would not meet the clinical threshold for a diagnosis of SLD in the medical system. The reverse is also true in that some students meet the diagnostic criteria for an SLD in the medical system but are not identified as requiring special education in their school and therefore are not SLD eligible. Children with higher intellectual development are especially at risk for this scenario because they can often access the curriculum and perform at grade level despite underlying cognitive weaknesses which significantly undermine the efficiency of their learning. For example, an intellectually gifted student (typically defined as having an IQ standard score of 130 or greater) would need to perform below grade level (academic achievement near a standard score of 78) to be eligible for special education services. This student would need to perform over three standard deviations lower in one or more academic domains. A student with an average IQ (standard score of 100) would only need to perform 1.5 standard deviations lower to be identified as having special education needs.

One substantive problem with the educational definition is that standards of identification of SLD differ across school boards, postsecondary institutions, and across countries, so that an individual identified as SLD by one school board, or in one region of the country, or in one country, may not be identified in another. Moreover, as both educational laws and psychiatric taxonomies change over time, different groups of individuals are identified as having SLD or Learning Disabilities, depending on the specific set of criteria applied. Thus, the terms Specific Learning Disorder and Learning Disabilities are not synonymous. To date, there is no consensus definition of this condition: Current assessment and diagnostic practices in both medical and educational systems are varied, often opaque and discriminatory, and therefore vulnerable to legal challenges.

HISTORY AND COMPARATIVE NOSOLOGY

History

The term “Learning Disability”, credited to Samuel Kirk in 1962, was first recognized as a formal category in the field of special education in the 1960s, although the first clinical case of a child’s unexplained difficulties learning to read had been described much earlier in the United Kingdom in 1896. In 1963, parents, educators, and policy specialists in the United States, agreed on the term “Learning Disabilities” to refer to a large group of individuals who had unexpected and significant difficulty learning to read, write, or do arithmetic, despite average or above intellectual ability.

Kirk defined learning disability in 1962 as:

> a retardation, disorder, or delayed development in one or more of the processes of speech, language, reading, spelling, writing, or arithmetic resulting from a possible cerebral dysfunction and/or emotional or behavioral disturbances and not from mental retardation, sensory deprivation, or cultural and instructional factors. (Kirk SA. *Educating Exceptional Children.* Houghton Mifflin; 1962:263.)

Early terminology, which focused primarily on difficulties learning to read, included terms such as reading blindness, word blindness, congenital word blindness, strephosymbolia, specific reading disability, specific developmental disability, unexpected reading failure, specific reading retardation, and garden-variety poor readers, and “dyslexia” (from the Greek *dys* and Latin *lexia*). Descriptions of difficulties learning arithmetic emerged later. For instance, Gerstmann, in 1957, described an “isolated disability to perform simple or complex arithmetical operations and an impairment of orientation in the sequence of numbers and their fractions” (Gerstmann syndrome). The term “dyscalculia” (from the Greek *dys* and Latin *calculia*) was used by Cohn in 1968, to refer to a “failure to recognize numbers or manipulate them in an advanced culture,” and later by Kosc as “a disorder of the special abilities for mathematics without a simultaneous defect in general mental abilities.” The first descriptions of *developmental* disorders of written language in the 1960s were based on case studies of isolated impairments of written expression, with other forms of verbal behavior remaining intact. Since these early descriptions involved visual–motor integration, the term *dysgraphia* (literally “difficulty with writing”) was the first diagnostic label associated with disorders of written expression. Although it is sometimes used as a synonym for the disorder of written expression, the term *dysgraphia* is used more correctly to describe difficulties with the mechanics of handwriting. By contrast, SLD in written expression emphasizes difficulties with accurate and fluent generation and composition of written text.

Terminology and conceptualizations of SLD differ widely among professionals from different disciplines or different countries; they also differ between professionals and the lay public. For instance, in the field of research, the term “dyslexia” is typically understood to refer primarily to problems in decoding single words and is associated with phonologic deficits, but lay understanding differs and is broader (e.g., often incorporates concepts such as sensory, perceptual, motor, and emotional difficulties). Also, in Europe, the term “dyscalculia” is used to refer to developmental problems in basic numerical processing, such as automatic or implicit processing of quantities or numbers. By contrast, the term is used rarely in North America, which instead uses terms such as “mathematical disabilities” or “arithmetic difficulties” that implicate a broad array of problems with arithmetic facts, arithmetical procedures, and mathematical reasoning.

Nonetheless, early perspectives of Learning Disabilities shared some commonalities: (a) recognition that impaired academic skills (reading, arithmetic) occur in the context of average or even superior intellectual abilities; and (b) a focus on impairments in specific component skills (e.g., word reading, calculation) rather than viewing an academic domain as a unitary construct. Other important insights were that these disabilities were believed to be congenital, heritable, and manifest more commonly in males than females. However, whether learning disabilities could occur either as an isolated domain-specific deficit or be accompanied by deficits in other academic domains remained controversial, as it does today.

The advent of IQ tests in the early 1900s such as the Stanford-Binet, permitted direct evaluation of the observed difference in a person’s academic skills and intellectual ability. In 1920, the neurologist Samuel Orton postulated a neurologic basis for developmental reading difficulties, suggesting that delayed development of specialization of the left hemisphere for language was potentially causal. However, it was not until the 1970s that an adequate knowledge base was established, which localized the difficulties at the single-word level and pinpointed the cause in most cases as language-based deficits in phonologic processing. The 1940s, known as the era of the “brain-damaged child,” incorporated a broader array of cognitive deficits into the category of Learning Disabilities, including perceptual, perceptual–motor, and attention difficulties. However, it was not until the 1960s that Learning Disabilities emerged as a formal category and the notion of a discrepancy between IQ and achievement prevailed as one of its defining features. The concept was broadened to include perceptual as well as linguistic processing deficits as underlying learning disabilities.

As previously mentioned, in the 1970s in the United States, the conceptualization of Learning Disabilities was consolidated in the federal definition (U.S. Office of Education, 1977) and in the definitions of professional organizations (e.g., National Joint Committee on Learning Disabilities, 1978), and achieved official federal status as an eligible category for direct services (Public Law 94–192). The notion of underlying deficits in cognitive processing as causal factors was instantiated in the conceptualization of Learning Disabilities and the condition included conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental dysphasia (thereby continuing its earlier conceptualizations), but excluded learning problems resulting from visual, hearing, or motor disabilities, mental retardation, emotional disturbance, or environmental, cultural, or economic disadvantage. Thus, this prevailing legal definition restricts the concept of Learning Disabilities to a language-based disorder and does not include nonlanguage-based learning difficulties, such as dyscalculia, as defined by some researchers. In other words, speech and oral language problems are included in the federal category but are coded separately as Communication Disorders in psychiatric (ICD, DSM) taxonomies. Thus, the federal/legal and medical categories of Learning Disabilities differ in the range of learning difficulties they encompass as well as in the imputed boundaries between disabilities/disorders.

Comparative Nosology

The first psychiatric/medical descriptions of SLD in reading or mathematics occurred in DSM-III (1980), with SLD in written expression added in DSM-III-R (1987), under the general category of “Academic Skills Disorder” on axis II. These learning difficulties were classified into three separate disorders: *developmental reading disorder, developmental arithmetic disorder,* and *developmental expressive writing disorder,* respectively. By contrast, DSM-IV-TR–coded learning disorders on axis I, as does DSM-5. Both of these DSM versions dropped the “developmental” descriptor that is used by other clinicians and researchers to distinguish between developmental versus acquired learning problems (e.g., as a result of traumatic brain injury, dementia, stroke). There was no compelling logic for the coding shift from axis II to axis I, but the shift may have served to render these impairing learning disorders more salient for diagnosticians, whose clinical practice focuses primarily on other types of mental health problems. The 2013 revision of DSM (DSM-5) no longer recognizes subtypes of SLD distinguished in previous DSM versions. Instead, it recognizes an umbrella category of SLD and then uses specifiers for clinicians to use to detail more precisely the range of academic skills that are impaired *at the time of assessment*, thereby recognizing the possibility of changes in the manifestation of SLD with development or educational instruction. Specifiers refer to three major academic domains that may be affected (i.e., reading, writing, mathematics) as well as the type of skills impaired. For instance, *SLD in reading* may manifest as difficulties in accurately decoding single words (a.k.a., dyslexia), fluency in reading text accurately, or reading comprehension; *SLD in written expression* may manifest as impairments in spelling words accurately, inaccurate use of grammar and punctuation, or in the clarity or organization of written expression; *SLD in mathematics* may manifest as impairments in basic number sense, memorization or retrieval of arithmetic facts, accurate or fluent calculation, or accurate math reasoning.

The DSM-5 conceptualization of “SLD in reading,” “SLD in written expression,” and “SLD in mathematics”—although comparable with—is not identical to other terms used primarily in the fields of education and in educational and clinical psychology (e.g., “reading disabilities,” “developmental dyslexia,” “dyslexia,” “developmental expressive writing disorder,” “written language disabilities,” “dysgraphia,” “spelling disorder,” “math disabilities,” “specific disorder of arithmetic skills,” “dyscalculia”).

The DSM-5 nosology of SLD also differs from that of ICD-10, which differentiates even more specific subtypes of SLD than does any version of the DSM. For instance, although ICD-10 distinguishes three subtypes that appear to be like those of DSM-IV-TR (“specific reading disorder,” “specific spelling disorder,” “specific disorder of arithmetic disorder,” as well as “mixed disorder of scholastic skills”), they differ in their requirement for the restriction of impairment to that specific academic domain. For instance, the categorical subtype of Specific Reading Disorder excludes individuals who also manifest impairments in arithmetic; Specific Spelling Disorder excludes those who also manifest a reading disorder; individuals who manifest impairments in more than one academic domain are given the diagnosis of Mixed Disorder of Scholastic Skills.

Ongoing concerns about the DSM (and ICD) approach to nosology have spurred a recent initiative by the U.S. National Institute of Mental Health (NIMH) to focus on alterations in the brain and its substrates that might signal the location and source of human mental stresses or woes, specifically the RDoC (Research Domain Criteria Project were created by Cuthbert and Insel). The RDoC framework is a research strategy with a dynamic matrix of elements at its core that includes six major domains of human functioning (social processing, negative valence systems, positive valence systems, arousal and regulatory systems, and cognitive systems). Within each domain are behavioral elements that range from normal to abnormal with the understanding that behaviors are affected by environmental and neurodevelopmental variables.

EPIDEMIOLOGY

SLD is one of the most frequently diagnosed Neurodevelopmental Disorder. A recent epidemiologic study suggested that the lifetime prevalence of SLD is around 10%, with the prevalence being greatest among children with special health care needs (28%) compared to typically developing children (5.4%). However, prevalence rates vary as a function of the definition, cut-offs, and instrumentation used. Moreover, most estimates are based on the prevalence of a single manifestation of SLD (e.g., in reading or mathematics) without considering the co-occurrence of SLD across academic domains. Nonetheless, SLD in the domain of reading (particularly at the word level) is the most common and well-studied manifestation, affecting at least 90% of all individuals identified as having SLD. It is estimated that two of five children in special education have been placed because of reading difficulties. However, children showing a deficit in one academic domain (e.g., reading) are highly likely to show deficits in other domains as well (e.g., spelling, written expression, mathematics), either concurrently or emerging later, as demands for efficient and fluent use of basic academic skills increase.

To date, there are a handful of studies of prevalence of SLD based on DSM-5 criteria. The most recent one attempted to understand the prevalence of SLD, as well as comorbid disorders, and risk factors in elementary school children in Turkey (Senol et al., 2020). The study found that the prevalence rate of SLD was 6.6%, impairment in reading was 4%, in mathematics was 3.6%, and in written expression was 1.8%. Nearly 63% of the children with SLD had one or more comorbid diagnoses, with ADHD being the most common (54.9%). Sex differences were observed as well with a higher prevalence of SLD among males. Another study conducted in a German school population found that SLD across several academic domains (and subdomains, such as word reading, fluency, and comprehension) occurred as frequently as an isolated disorder in any one domain. The rates of multiple-domain SLD were high (particularly for co-occurrence of arithmetic and spelling difficulties) even when applying the stricter criteria of achievement scores of 1.25 to 1.5 standard deviations below the mean for age. No sex differences were observed when considering SLD across multiple academic domains, however, more girls manifest problems in arithmetic and more boys had spelling problems and combined reading and spelling problems compared to girls. A third study, conducted in Brazil, a low- to middle-income country, reported that almost 8% of the representative sample (about 1,600 second- to sixth-grade children from 4 geographic areas) manifest SLD across the academic domains (global impairment), 7.5% in reading, 6% in arithmetic, and about 5% in written expression. Moreover, SLD with global impairment was significantly associated with ADHD; SLD in arithmetic was associated with both ADHD and Anxiety Disorders.

Although more cross-cultural epidemiologic studies of the prevalence of SLD based on full DSM-5 criteria are needed, the practical implications of the existing studies are that assessment for SLD should include all three academic domains as well as the subskills. Clinicians are advised to specify the diagnosis of SLD by providing additional information about the learning domains that are affected at the time of assessment (i.e., implementing Specifiers for SLD listed in DSM-5, as described in the later subsection on diagnosis).

ETIOLOGY

Different levels of explanation must be considered in understanding causal influences on learning academic skills. Converging evidence from cognitive, neuroimaging, twin, family, behavioral, and molecular genetic studies indicates that SLD is a multifactorial neurodevelopmental disorder arising from the complex interplay of biologic and environmental risk factors that shape the development of brain systems underlying academic learning. Reflecting the multiple constellations of variables that shape development and contribute to learning disorders, research in SLD has shifted from a single-deficit model to a more dynamic multiple deficit model (MDM; Pennington 2006). This model proposes that the etiology of behavioral disorders is multifactorial and influenced by multiple risk and protective factors which can be environmental or genetic. The interaction of risk and protective factors alter the development of neural systems that mediate cognitive functions and ultimately produce behavioral symptoms that are identified in the assessment of learning disorders. Over the past 15 years since this model was introduced, it has helped to identify comorbidity between learning disorders and neurodevelopmental disorders (e.g., ADHD), and has also aligned with NIMH RDoC in efforts to identify transdiagnostic risk factors in psychiatric disorders (Insel et al., 2010).

The use of cognitive multiple deficit models in dyslexia and dyscalculia have also proven successful. A recent study, Peterson et al., in 2017, aimed to employ a cognitive multiple deficit model to identify cognitive predictors unique to specific skill domains (word reading, math ability). The study found that a single common cognitive predictor, verbal–conceptual skills, contributed to the overlap between reading and math. It also found that processing speed contributed to the overlap between math and attention. Results of the overall study revealed that cognitive factors accounted for a large majority of the variance in reading and math ability (approximately 80% to 90%).

Like any predictive model, the MDM has some problems. First is that despite the overall theme that multiple deficits ultimately contribute to an SLD, research in dyslexia using the MDM have found that some children (approximately 25%) can be explained by single deficits. Because much of the work using MDM has been cross-sectional it is unclear if the association between cognitive deficits and behavioral symptoms are causal, consequence, bidirectional, or related to other variables. Despite these limitations, it appears that the MDM has great potential to not only help map the etiology of SLD, but also highlight related and effective treatment. There is some evidence that individual cognitive factors are helpful in treatment planning. For example, optimal reading instruction may vary as a function of cognitive characteristics such as vocabulary development. This suggests that further development of the MDM can result in a more detailed understanding of neurodevelopmental disorders and thus a more tailored treatment approach.

Despite growing utility of the MDM in understanding the etiology of SLD, the notion of a single deficit model remains and continues to present in SLD literature and in clinical care. For example, in contrast to the growing evidence that dyslexia can present without a phonologic deficit, some in the assessment and clinical community assert the necessity of a phonologic deficit in the disorder.

Emerging evidence supports the likelihood that multiple and interacting genetic factors, brain differences, cognitive deficits, and environmental stressors underlie SLD, each contributing to the overall outcome.

Genetic Factors

Behavioral Genetics. Family and twin studies indicate that most dimensions of cognition (general intelligence, learning to read, spell, do arithmetic) are moderately heritable: About 50% of variance in a population is attributable to genetic influences, with increased importance with increasing age. For example, heritability estimates of individual differences in reading skills increase from about 20% of variance at age 6 to about 80% at age 12. Increases in heritability estimates with age most likely reflect a reduction of environmental influences once children enter formal education and increasing correlations between genotype and environment as with increasing maturity and independence, individuals tend to select environments that fit their abilities (genotype). A recent meta-analysis, based on 5,330 monozygotic and 7,084 dizygotic twin pairs in unselected samples (i.e., typically developing), estimated heritability of academic achievement at 73% for reading words, 44% for spelling, 49% for reading comprehension, 57% for mathematics, and 645 for language. Moreover, a common genetic factor has been demonstrated among spelling, reading fluency, and reading comprehension, with genetic influences accounting for around 80%

of the covariation between core academic skills, such as reading, spelling, and arithmetic, with many of the same genes affecting the learning of various academic subjects independent of intelligence.

Twin studies involving large samples with diverse abilities (e.g., including typical readers and those with dyslexia) reveal common genetic influences across the range of ability and disability, suggesting that a particular set of genes affects academic learning in both typically developing samples and those with SLD. Also, despite the paucity of behavioral genetic research on SLD in mathematics, family studies indicate that the familial prevalence of dyscalculia is about 10-times higher than expected by chance and that genetic factors contribute to the observed covariance between reading and mathematics deficits.

It is important to note that not all individuals with SLD inherited the condition since twin studies also reveal that both common and unique environmental influences play a role, as discussed below. However, from a clinical perspective, the high heritability estimates for SLD highlight the importance of probing family history in the assessment for SLD and ensuring that parents and patients with SLD can read and comprehend any written reports or information brochures provided to them.

Molecular Genetics. To date, most molecular genetic studies have focused on genes implicated in reading ability and disability, particularly those associated with word reading and spelling difficulties, also known as dyslexia. Several candidate genes have been implicated (e.g., *DYX1C1*, *DCDC2*, *ROBO1*, *KIAA0319*, and the *MRPL19*/*C2ORF3* locus), with some being replicated in many studies. One problem is that most of the candidate risk SNPs identified so far for dyslexia have been in noncoding regions of the genome which are not well understood. Recent studies report genetic interactions between two *DCDC2* risk haplotypes (*READ1* alleles) and the risk haplotype *KIAA0319*, which influence both reading and oral language, as well as intelligence, with both protective and deleterious alleles. Findings from these studies are consistent with the generalist genes hypothesis and underscore the close association between Communication Disorders and SLD, which are conceptualized distinct disorders in DSM and ICD systems.

Molecular genetic studies of SLD in mathematics are sparse and findings inconsistent. For example, two genome-wide association studies failed to identify any significant associations and a third reported a significant genetic component underlying mathematical abilities but did not identify specific risk factors. Another genome-wide association study identified one marker (myosin-18B), that was not only associated with mathematical ability in two independent cohorts selected for dyslexia, but also was associated with brain alterations (reduced depth of right intraparietal sulcus that has been linked with numerical processing) in adults with low mathematical performance. However, this association was not replicated by another group of researchers using independent samples.

Neurologic Substrate

SLD in Reading. Most research has focused on dyslexia. Structural and functional neuroimaging studies have revealed brain regions that are most consistently involved in reading words and which are altered in dyslexia. These regions include inferior frontal, superior and middle temporal, and temporoparietal regions. Experienced readers also involve an area of the left fusiform gyrus, known as the *visual word form area*, which is activated for orthographic (print) processing. This reading network typically develops with age as children gain automaticity in specific skills required for reading.

Meta-analytic studies have revealed patterns of structural and functional differences in children, adolescents, and adults with SLD in reading words (dyslexia). The most common difference is reduced activation in *left* temporal, parietal, and fusiform (visual word form area) regions.

Given the abnormal development of white matter across posterior and anterior language networks related to phonologic skills, it has been proposed that dyslexia results from a white matter disconnection syndrome. This may also explain other deficits seen in dyslexia which rely on a strong connection between posterior auditory processing areas and anterior motor planning areas.

Family risk studies of dyslexia which are designed to compare children at high risk for dyslexia (significant family history of the disorder), without a family history of the disorder have helped to identify developmental precursors. Precursors are often cognitive or neural differences found in children from infancy to preschool who have not yet been exposed to formal reading instruction. Because many of these studies are longitudinal, these children have been followed over time. Once they have been diagnosed with dyslexia the researchers are able to review historical data related to neural and cognitive development and identify what discriminated them from children who were not diagnosed with dyslexia. These neural and cognitive differences were seen before the children were exposed to formal reading instruction, thus identified differences in dyslexia are not simply the result of poor reading development.

Other family risk studies following prereading children at high risk for dyslexia have revealed altered neurophysiologic responses to auditory stimuli (e.g., mismatched negativity) that show potential as an early endophenotype of dyslexia. Some of these biomarkers enhance or even outperform behavioral measures in predicting long-term reading outcomes: But, at present, none of these biomarkers can be used clinically to rule in or rule out the diagnosis of SLD in reading.

SLD in Mathematics. Evidence for the neurologic base of mathematical cognition and SLD in mathematics is sparse in comparison to that for reading and SLD in reading. However, converging evidence from structural and functional imaging studies indicate the *right* intraparietal sulcus as a major (but not sole) locus of brain differences and cognitive deficits in children and adults with SLD in mathematics. Alterations in other interconnected brain regions have also been observed, including posterior regions (bilateral intraparietal sulcus and fusiform gyrus) and prefrontal cortex (premotor cortex and middle frontal gyrus). Also, network analysis reveals multiple dysfunctional white matter circuits, suggesting a neurodevelopmental disconnection condition.

SLD Writing. fMRI studies of written language and spelling provide little evidence of a neurologic underpinning. Many of these studies have small sample sizes and are not exclusive to children. Only a handful of anatomical studies have isolated SLD in writing. Overall, many of the identified brain areas that relate to writing include frontal and parietal regions, as well as coordinated activation with the cerebellum. There does not seem to be a consensus of a neural network that is specific to written language, and many of the studies suggest a considerable overlap with networks related to linguistic processing and motor planning.

Cognitive Factors

The generalist genes hypothesis in learning disorders hypothesizes that shared genetic factors lead to shared cognitive difficulties across the various types of SLD. One domain-general cognitive deficit is executive functioning, which includes the control and coordination of short-term, working- and longer-term memory systems, encoding and retrieval systems, attention, and planning. Working memory, which refers to the ability to temporarily hold and manipulate auditory–verbal and visual–spatial information, while learning new information or problem solving, is believed to be impaired in most individuals with SLD. Thus, individuals with SLD have difficulties suppressing irrelevant information from working memory, as well as simultaneously holding and processing information, which manifests as difficulties updating what they know or have just learned. In other words, the more restricted working memory capacity associated with SLD makes it more difficult for these individuals to suppress irrelevant information and update new knowledge as they learn.

Working memory deficits will impair learning to read new words, which requires the ability to remember letter–sound combinations, their order, as well as the many special rules (e.g., “silent” letters, such as “g” in gnash or height, “e” at the end that changes the vowel sound but itself remains silent, as in “site”). Working memory deficits have an even greater impact on reading comprehension and writing: Reading comprehension requires updating information in working memory to monitor the meaning of each sentence and each paragraph, extract the main ideas, and construct mental images of the emerging situations and events; writing requires the ability to plan, organize, compose, and edit, while keeping in mind the writing goals, content, audience, as well as spelling, vocabulary, and grammar. Poor working memory will also impair mathematical abilities, including the ability to retrieve and hold math facts for use in computation, follow arithmetic procedures in calculation, and in math reasoning. A recent meta-analysis indicates the domain-general nature of working memory deficits across the various manifestations and severity of SLD, but with some indication of more severe verbal working memory deficits in individuals with SLD in reading and more severe deficits in visual–spatial working memory in those with SLD in mathematics.

Slow processing speed is another domain-general factor that has a negative impact on learning. For instance, rapid automatized naming, which refers to the ability to rapidly name familiar alphanumeric (letters, numbers) and nonalphanumeric (colors, objects) stimuli, is impaired in a substantial proportion of individuals with SLD. Although some research has shown distinct rapid naming profiles in students with SLD in reading versus SLD in mathematics (impairment in alphanumeric in the former and in nonalphanumeric stimuli in the latter), a recent meta-analysis concluded that rapid naming of nonalphanumeric stimuli is impaired in both manifestations of SLD, whereas alphanumeric naming may be more specific to SLD in reading. However, rapid naming deficits are not unique to SLD: They are also shared with other neurodevelopmental disorders (e.g., ADHD, Autism Spectrum Disorders, and Communication Disorders).

Language-based factors, although often linked in the older literature exclusively with SLD in reading (particularly its manifestation as “dyslexia”), are associated with all manifestations of SLD. For instance, oral language difficulties interfere with all aspects of reading, writing, and mathematics, although in the latter domain, the impact is primarily on those aspects of mathematics involving verbal processing. Impairment in a specific language-related skill, phonologic processing, is widely recognized as the core cognitive deficit in SLD in word reading. However, some evidence suggests that early phonologic deficits may have deleterious effects subsequently on learning arithmetic. For example, one study reported that over 50% of children with phonologically based reading difficulties but no apparent difficulties in learning basic arithmetic at age 5 years manifest learning difficulties in mathematics as well as continued problems in reading at age 7.

An ongoing controversy in the field of SLD is whether SLD in reading single words (dyslexia) is associated with a unique and potentially diagnostic set of cognitive deficits. However, the conclusions from a large-scale investigation of two population-based samples (U.S. cross-sectional study; international longitudinal study) are that cognitive predictors believed to play a causal role in dyslexia (e.g., phonologic processing, rapid automatized naming) are neither necessary nor sufficient for the diagnosis of SLD in reading. Moreover, the investigators concluded that the relationship between the cognitive measures and reading skill is probabilistic not deterministic or diagnostic. Also, it remains unclear whether cognitive factors are a cause, consequence, or correlate of SLD. Hence, DSM-5 does not include cognitive processing deficits in the diagnostic criteria for SLD.

Environmental Factors

Twin studies indicate that shared and nonshared environmental factors contribute up to about 40% or 50% of academic skill development, and to SLD, but the nature of these environmental factors is not well understood. Candidate risk factors include low socioeconomic status, the prenatal environment, and the postnatal learning environment provided by parents.

Prenatal Environment. Prenatal exposure to toxic substances, such as alcohol, tobacco smoke, lead, and other environmental toxins have been linked to neurodevelopmental outcomes. Prenatal exposure to tobacco smoke remains high despite the overall worldwide reduction in smoking among adults: A recent study reveals that 7.2% of women who gave birth in the United States smoked during pregnancy. Women between the ages of 20 to 24 had the highest rates of smoking at 10.7%. Known risks of maternal smoking and alcohol intake during pregnancy include premature birth, low birth weight, as well as fetal alcohol syndrome for prenatal alcohol exposure. However, findings for the outcomes in terms of subsequent academic achievement in the offspring are inconsistent, perhaps due to methodologic shortcomings of the studies in terms of exposure assessment, measurement of academic outcomes, and inadequate control or consideration of confounding factors. Nonetheless, a recent review of 30 studies concluded that both prenatal exposure to the ingredients in tobacco smoke and alcohol can result in a spectrum of adverse developmental outcomes, ranging in severity from intellectual disorders to more subtle negative impacts on academic achievement, memory, verbal abilities, and poor executive functioning.

More robust evidence comes from recent studies that postdate the above review. For example, a recent large-scale longitudinal study focusing specifically on reading outcomes of prenatal exposure to tobacco smoke found that high prenatal exposure to nicotine tobacco smoke (>17 mg nicotine per day) was negatively associated with reading performance, particularly in decoding single words and spelling, with greatest effects on children with poorest phonologic skills—suggesting a possible gene by environment effect. Another longitudinal epidemiologic study conducted in the United Kingdom examined the health and academic trajectories of children born in the late 1950s—an era in which there was a high rate of maternal smoking during pregnancy (about 30%)—from the prenatal period through to age 16. The research revealed that low birth weight, prenatal exposure to tobacco smoke, and poor school-aged health are negatively related to academic achievement in reading and math, suggesting a cumulative effect of early adversities. Also, a meta-analysis of the academic outcomes associated with children born very preterm (<32 weeks) or with very low birth weight (<1,500 g) reveals negative effects on IQ scores (by an average of 11 IQ points) and poor academic performance, with greater frequency of SLD in mathematics compared to SLD in reading. About two-thirds of children born very prematurely or with very low birth weight have special educational needs, especially for SLD in mathematics.

Family Environment. Parents of children with SLD may well have SLD themselves and so may be less able to engage in "code-focused" or "meaning-focused" activities with their children who have SLD: These activities refer to parent–child engagement in rhyming and counting games, teaching letter names and sounds, letter–sound pairs, teaching numbers and how to count (code-focused); and reading stories aloud in an interactive way or playing board games that involve symbol recognitions and counting (meaning-focused). However, although parents of children with SLD report being more stressed than parents of children without SLD, there is no evidence of poorer family functioning in families in which someone has SLD. Moreover, although the home literacy and numeracy environment, including parental SLD, predict children's growth in literacy and numeracy abilities in preschool and kindergarten years, by school age, genetic influences as well as basic skills knowledge (e.g., letter and number knowledge) tend to predominate.

Gene X Environment Interaction. The complex interplay between genetic and nongenetic influences has garnered relatively little empirical attention, given the general consensus that SLD arises from gene-by-environment interactions. However, some recent studies provide preliminary evidence for such interactions in SLD in single-word reading (a.k.a. dyslexia). For example, one study reports gene X environment effects between adverse prenatal factors (e.g., maternal smoking during pregnancy, low birth weight, socioeconomic status) and a genetic marker associated with dyslexia (*DYX1C1-1259C/G*) in 165 nuclear families in which at least one member had dyslexia. Findings revealed that the interaction of common “C” allele with lower birth weight was associated with lower reading scores, and the interaction between the rare “G” allele and maternal smoking during pregnancy was associated with poorer short-term and working memory scores.

Table 40–2.
Common Risk Factors, Signs, and Symptoms of Specific Learning Disorder

Preschool

- Lack of interest in playing games with language sounds (e.g., repetition, rhyming)
- Trouble learning nursery rhymes
- Problems counting correctly
- Frequent use of baby talk and mispronunciation of words
- Failure to recognize letters in own name
- Trouble remembering names of letters, numbers, or days of the week

Kindergarten

- Unable to recognize and write letters, write own name, or use invented spelling
- Trouble breaking down spoken words into syllables (e.g., “cowboy” into *cow* and *boy*)
- Trouble recognizing words that rhyme (e.g., “cat,” “bat,” “hat”)
- Trouble connecting letters with their sounds (e.g., letter “b” makes the sound /b/)
- Unable to recognize phonemes (e.g., does not know which of the set of words [“dog,” “man,” “car”] starts with same sound as “cat”)
- Unable to count in sequence beyond “4” (e.g., “1,2,3,4…um 7,11”)

Primary Grades (Grades 1–3)

- Receives reports of “not doing well in school”
- Continues to have problems recognizing and manipulating phonemes
- Unable to read common one-syllable words, such as “mat” or “top”
- Problems in connecting sounds and letters (e.g., “big” for “got”)
- Unable to recognize common irregularly spelled words (e.g., “said,” “two”)
- Difficulty in sequencing numbers and letters
- Complains that reading is hard and avoids or refuses to do it
- Difficulty learning simple math facts of 10
- Counting is often inaccurate above about 5 or 6 items

- Unable to identify the relevant numerical information or generate a math equation needed to solve a simple arithmetic word problem

Middle Grades (Grades 4–6)

- Mispronounces or skips parts of long, multisyllable words (e.g., says “conible” for “convertible” and “aminal” for “animal”)
- Confuses words that sound alike (e.g., “tornado” for “volcano”)
- Trouble remembering dates, names, telephone numbers
- Trouble reading small function words (e.g., “that,” “the,” “an,” “in”)

Middle Grades (Grades 4–6) *Continued*

- Gets the first part of a word correct, then guesses (e.g., reads “clover” as “clock”)
- Poor comprehension with or without slow, effortful, and inaccurate reading
- Trouble completing homework or tests on time
- Fear of (refuses) to read aloud, avoids reading
- Terrible spelling; unable to recognize correct spelling from plausible alternatives (e.g., necessary/necessery; friend/frend)
- Continues to rely on finger counting for simple arithmetic
- Difficulties mastering basic math facts (e.g., cannot remember that 3 + 2 = 5)
- Shows high levels of anxiety around any activity requiring basic arithmetic

High School, College, Work

- Slow, effortful reading of single words and connected text
- Trouble pronouncing multisyllable words
- Frequent need to reread material to understand or get the main point
- Trouble making inferences from written text
- Avoids activities that demand reading (reading for pleasure, reading instructions)
- Terrible spelling; the same word is misspelled in different ways within one piece of work
- Excessive problems generating written text: short, immature, not representative of the student’s knowledge or creative ideas

- Avoids writing activities; written expression is very short, immature, and disorganized
- Great difficulties in using arithmetic in functional activities (e.g., calculating tip percentage, calculating the amount of change required after a purchase, monitoring bank account)

General

- First-degree relative (parents, sibling) with current or history of SLD
- Slow learning to talk
- Anxious and embarrassed when asked to read or work out a problem in public (e.g., in class, workplace)
- Avoidance or refusal to read, write or do arithmetic
- Reluctance to go to school

The clinical implications of findings from current research on etiologic factors of SLD include the importance of including in an assessment for SLD the family history of SLD, the individual's prenatal and birth history, and for children with suspected SLD, an assay of their home literacy and numeracy environment. Clinicians are advised to be alert to the increased risk for SLD associated with prematurity, low birth weight, and prenatal exposure to tobacco smoke, alcohol, or other drugs.

CLINICAL FEATURES, ASSOCIATED PROBLEMS, AND DIAGNOSIS

Clinical Features

The learning difficulties manifest as an array of observable and persistent behaviors or symptoms that change across the lifespan, many of which are summarized in Table 40–2. These include (but are not restricted to) poor word recognition and difficulties sounding out words, frequently guessing words, reading slowly and hesitantly; difficulty understanding what is read (may get the basic facts, but not the deeper meaning); marked difficulties spelling words correctly;

poor sentence and paragraph structure in written work, with numerous punctuation and grammatical errors; difficulties understanding numbers and their relative magnitude and relationships; persistent use of primitive counting strategies (e.g., counting on fingers); difficulties recalling number facts and getting lost in the midst of arithmetic computation; and difficulties applying basic arithmetic concepts to solve quantitative problems.

These observable symptoms are not simply the consequence of an intellectual disorder, sensory deficits or neurologic disorders, lack of proficiency in the language of instruction, inadequate educational instruction, or lack of opportunity to engage in full-time formal schooling.

Signs and Symptoms of SLD in Reading. The most common and earliest manifestation of SLD in reading, particularly in the English language, is inaccurate, effortful, and slow reading at the word level (a.k.a., “dyslexia”). Recent evidence, however, has revealed that signs of SLD in reading can be detected even earlier. Differences in oral language development can accurately identify children who are at risk for dyslexia before they are exposed to formal reading instruction. In contrast to the common understanding that phonologic deficits are the strongest predictor of an SLD in reading, phonologic deficits are not a strong predictor of poor reading until kindergarten (5 years old). A 2016 meta-analysis by Snowling and Melby-Lervag found that the strongest predictor of an SLD in reading was poor grammar and speech articulation at ages 2 to 3 years, as well as grammar and vocabulary at 3 to 4 years. By 5 years old, phonologic processing and vocabulary most strongly predict reading delays, while most oral language difficulties have resolved at that point. From kindergarten and most of elementary school, a child with reading deficits will demonstrate a reluctance and resistance to reading aloud; will read hesitantly, slowly, and inaccurately, often with visible effort and discomfort (e.g., sighing, squirming, becoming flushed). They will also make guesses based on the sound of the first letter; they may recognize a few words by sight, but they are unable to “sound out” unfamiliar words. They will struggle to reconstruct the word having sounded out its component parts (e.g., sounds out /b/ /a/ /n/ /k/, yet cannot reconstitute to “bank”); adolescents and adults may read accurately but so slowly and with such effort that reading is not functional for everyday use, and so may avoid reading.

Lack of automaticity in letter–sound mapping not only interferes with reading at the word level, but will also make it difficult for the person to understand what has just been read at the level of the sentence, paragraph, or entire text. However, some individuals with SLD in reading at the word level may have very little difficulty in comprehending connected text (sentence, paragraph, entire text) if the need to decode words is bypassed by having the text read aloud at a reasonable pace (e.g., by using text–speech software).

These observable behaviors are due to an underlying difficulty in learning how the letters (graphemes) and speech sounds (phonemes) of one’s language map onto one another and being able to do so automatically. The general consensus in the scientific community is that the problems with letter–sound conversion arise from an underlying deficit in processing phonemes—the elementary constituents of speech sounds.

Common features of SLD in reading comprehension include reading text aloud with a flat intonation and no modulation to acknowledge punctuation marks (e.g., no pauses at commas or periods; no change in pitch to represent a question mark or exclamation point); may read the text accurately and fluently—often too fast—but is unable to extract and use the inherent meaning (e.g., may not grasp the sequence of events, relationships among events or people, may not be able to draw inferences from text or process the deeper meaning of the text); tends to simply reread the text without using effective strategies before restarting to read (e.g., does not predict, scan for or review unfamiliar words), during reading (e.g., does not make notes or self-question to check one’s understanding; does not create visual images of the scenarios), and after reading (e.g., does not check the accuracy of understanding; does not make a synopsis).

Over time, these challenges in reading comprehension tend to result in a reduction in or avoidance of reading print (e.g., newspapers, novels, nonfiction, lengthy texts regardless of format, such as digital or paper). In turn, restrictions on the amount of reading will have a negative impact on vocabulary growth so that the individual’s range of vocabulary use and understanding is noticeably reduced.

Signs and Symptoms of SLD in Written Expression. One of the most common clinical features is the noticeable and persistent poor spelling for age—often described as “atrocious spelling” (e.g., often adds, omits, or substitutes vowels and consonants; may spell phonetically but incorrectly, such as “pikchr” for “picture”). Sentences are often very short or incomplete and contain grammatical errors, omission of words, little or no use of punctuation, as well as spelling errors. Also, written paragraphs are often short, poorly constructed, and organized (e.g., does use a starting sentence to introduce the topic; includes several topics in one paragraph without a logical order; does not maintain the correct verb tense, incorrect use of pronouns) and their intended meaning is unclear. A common complaint from teachers (and parents) is “procrastination”: The youngster fails to start writing, writes very little, and with great effort and reluctance. In general, the person’s written work does not accurately reflect that individual’s knowledge, creativity, or depth of understanding of an issue: The person’s spoken or visual presentations are often very much better than the written one. Also, essential written activities, such as note-taking, pose major challenges. This is because note-taking requires the physical act of writing or typing notes to occur simultaneously with a complex web of other cognitive processes involved in listening, comprehending, and retaining key information while continuing to process incoming and new information and summarizing key points in a legible and organized format.

Signs and Symptoms of SLD in Mathematics. One common sign that persists across the lifespan is the sole reliance on finger counting for simple addition and subtraction, which is inappropriate for age. Note that typically, children start to switch from finger counting to more efficient calculation strategies, such as retrieval of math facts from memory, by about age 8 or 9 (e.g., see the interview of an accomplished author and Defense correspondent, describe his lifelong struggles with SLD in mathematics, including his use of finger counting under the boardroom table! http://www.mathematicalbrain.com/qt/paul.mov).

Other common features include a poor sense of numbers in terms of their magnitude and relationship; inaccurate counting and calculation; failure to attend to key mathematical symbols (e.g., symbols for addition, subtraction, multiplication, and division); difficulty with identifying the previous or next item in a sequence (e.g., difficulty identifying the name of previous or next day, month, or number, despite being able to rote list those items); difficulty understanding the concept and use of “zero” or decimal points; and “gets lost” in the middle of multidigit calculation and may switch procedures or apply them incorrectly (e.g., in the problem 57−38, may start to subtract 7 from 8 instead of 8 from 7, then either subtracts 3 from 5 without considering the trading procedure, or may switch to adding 5 and 3). Many struggle with understanding the “language” of mathematics, especially in *word problems*, that are emphasized in current curricula across all grade levels. Word problems consist of two interwoven semiotic layers: A narrative about some event or sequence of events in the real world, and a latent web of mathematical relations typically expressed in words, and related by a question involving concepts of quantity such as “more,” “more than,” “less than,” “altogether.” Thus, many individuals with SLD in reading have great difficulty with the language-based demands of word problems. Moreover, older children, adolescents, and adults manifest difficulty understanding and using money, telling the time, using calculators, and remembering numerical PIN codes.

Math anxiety is another common feature that persists across the lifespan. Often described as “a feeling of tension, apprehension, or fear that interferes with math performance,” it may manifest as sweaty palms, a racing pulse, breaking out in a cold sweat, or a choking feeling in the back of the throat, when required to perform mathematics in everyday life (e.g., answer math questions in class; complete math homework, counting out money to buy something, splitting a restaurant bill or calculating the tip, entering specific numbers for cash withdrawal or deposit at an ATM machine).

General Signs and Symptoms of SLD. In school-aged children, the typical presenting problem concerns the difficulty in learning at school ("not doing well at school—and we don't know why"). The puzzling picture for parents and teachers is that the child appears intelligent and often succeeds or even excels in other aspects of the curriculum (e.g., oral discussions, art, drama, sports) that do not depend on the core academic skills.

In addition to the domain-specific features summarized above, SLD imposes a great emotional toll. For example, other common features include embarrassment and low self-esteem, and extreme discomfort or even refusal when asked to read, write, or spell. Children and adolescents may be reluctant or refuse to go to school; may report being exposed to daily criticism, taunting, or bullying; may look anxious, stressed, sad, depressed, or express despair or suicidal ideation.

Associated Problems

SLD is commonly, but not invariably, associated with a developmental history of problems in speech and language, attention, and motor skills. Moreover, SLD commonly, but not invariably, co-occurs with other neurodevelopmental (e.g., Communication Disorders, ADHD) and mental health disorders (e.g., Anxiety and Mood Disorders). These associated problems are clinically important because the diagnosis of SLD cannot be confirmed until the child has started school and been exposed to formal academic instruction, but Communication Disorders in particular are likely to be apparent and diagnosed earlier and so indicate a risk for later SLD. Likewise, a developmental and family history of SLD may indicate a risk for later anxiety and mood disorders in adolescence and adulthood.

Speech and Language Impairments. Oral language constitutes an essential foundation for abilities in written language. Thus, it is not surprising that a developmental history of Communication Disorders (Speech Sound Disorder, Specific Language Impairments (SLI), alone or in combination) in preschool years is a common precursor of SLD, particularly for poor skills in reading words and reading comprehension, spelling, arithmetic fact retrieval, and calculation. Speech sound disorder, defined as problems in producing sounds in one’s language accurately and intelligibly, whereas language impairments refer to problems in developing the structural aspects of language, such as grammar (syntax) and vocabulary (semantics). For example, one longitudinal study of a community-based sample of children identified with pervasive speech/language disorders in kindergarten found an estimated three- to sixfold greater risk for SLD (all manifestations alone or in combination) in young adulthood compared to typically developing youngsters. Moreover, according to a recent population-based study, the cumulative risk of SLD in written language by age 19 is about 64% in youngsters with speech/language impairments compared to 18% in those without.

Attention-Deficit/Hyperactivity Disorder. Epidemiologic and clinical studies indicate that the overlap between SLD and ADHD is greater than expected by chance alone. For example, around 20% to 45% of children with ADHD also meet diagnostic criteria for SLD, but the strongest link appears to be between the inattention dimension of ADHD and SLD, particularly SLD in reading and writing. Inattention in the classroom (even at subclinical levels), as rated by teachers, is a robust predictor of concurrent and subsequent poor academic outcomes, as well as a poor response to evidence-based academic intervention. Importantly, there is little robust evidence that medication treatment for ADHD (including stimulant and nonstimulant medications) has any substantial effect on SLD, despite its positive effects on the behavioral symptoms of ADHD and academic productivity (i.e., completing more arithmetic problems in a specified time frame). By contrast, preliminary evidence suggests that intensive educational intervention may improve (but not normalize) the behavioral problems. The clinical implications of these findings are that assessment for SLD should include assessment for ADHD and vice versa: Youngsters with SLD who have coexisting inattention or ADHD require multimodal intervention, including intensive educational intervention, since medication for ADHD will not address the SLD and intervention for SLD will be less effective if the inattention remains unaddressed.

Emotional Problems, Anxiety and Mood Disorders, Suicide Risk. Individuals with SLD are highly likely to manifest concurrent social, emotional, and mental health problems. For example, they are only too aware of their academic difficulties, leading to low self-esteem and these negative self-perceptions are very difficult to change and contribute to their problems in social and academic functioning. Avoidance of or reluctance to engage in activities requiring academic skills is common in children, adolescents, and adults. Also, recent meta-analyses confirm that the majority of youngsters with SLD (about 70%) have higher rates of anxiety and depressive symptoms compared to non-SLD peers. Episodes of severe anxiety or anxiety disorders (including somatic complaints or panic attacks) are common across the lifespan and accompany both the circumscribed and broader expression of learning difficulties. Moreover, SLD is associated with an increased risk for suicidal ideation and suicide attempts in children, adolescents, and adults. The co-occurrence of internalizing symptoms and disorders is clinically significant, whereas SLD is treated primarily in the field of education; thus, these internalizing symptoms often remain unaddressed. Thus psychiatrists and other mental health practitioners play an important role in identifying, monitoring, and treating concurrent internalizing problems.

Criteria for Diagnosis

There are no biomarkers to date for SLD. Thus, SLD is a clinical diagnosis, which is best made by an experienced clinician, based on a comprehensive assessment with careful clinical synthesis of information from multiple sources, including developmental, medical, educational, and family history; school reports; teacher observations, including response to classroom instruction and other intervention provided at school; as well as direct standardized psychometric assessment of the individual's academic skills—including reading (accuracy, fluency, comprehension), spelling, written expression, and arithmetic (number sense, counting, algorithmic computation). The DSM-5 diagnostic criteria for SLD are presented in Table 40–3.

Table 40–3.

DSM-5 Diagnostic Criteria for Specific Learning Disorder, With ICD-9-CM Codes (ICD-10 Codes Are Shown Parenthetically)

Diagnostic Criteria

A. Difficulties learning and using academic skills, as indicated by the presence of at least one of the following symptoms that have persisted for at least 6 months, despite the provision of intervention that targets those difficulties:
    1. Inaccurate or slow and effortful word reading (e.g., reads single words aloud incorrectly or slowly and hesitantly, frequently guesses words, difficulties sounding out words)
    2. Difficulty understanding the meaning of what is read (e.g., may read text accurately but not understand the sequence, relationships, inferences, or deeper meanings of what is read)
    3. Difficulties with spelling (e.g., may add, omit, or substitute vowels or consonants)
    4. Difficulties with written expression (e.g., makes multiple grammatical or punctuation errors within sentences; employs poor paragraph organization; written expression of ideas lacks clarity)
    5. Difficulties mastering number sense, number facts, or calculation (e.g., poor understanding of numbers, their magnitude, and relationships; counts on fingers to add single-digit numbers instead of recalling the math fact like peers; gets lost in the midst of arithmetic computation and may switch procedures)
    6. Difficulties with mathematical reasoning (e.g., severe difficulty applying mathematical concepts, facts, or procedures to solve quantitative problems)

B. The affected academic skills are substantially and quantifiably below those expected for the individual’s chronologic age and cause significant interference with academic or occupational performance, or activities of daily living, as confirmed by individually administered standardized achievement measures and comprehensive clinical assessment. For individuals aged 17 years and older, a documented history of impairing learning difficulties may be substituted for the standardized assessment

C. The learning difficulties begin during school-age years but may not become fully manifest until the demands for those affected academic skills exceed the individual’s limited capacities (e.g., as in timed tests, reading or writing lengthy complex reports for a tight deadline, excessively heavy academic loads)

D. The learning difficulties are not better accounted for by Intellectual Disabilities, uncorrected visual or auditory acuity, other mental or neurologic disorders, psychosocial adversity, lack of proficiency in the language of academic instruction, or inadequate educational instruction.

Note: The four diagnostic criteria are to be met based on a clinical synthesis of the individual’s history (developmental, medical, family, educational), school reports, and psychoeducational assessment.

Coding Note: Specify all academic domains and subskills that are impaired. When more than one domain is impaired, each one should be coded individually according to the following specifiers.

Specify if:

315.00 (F81.0) With impairment in reading:

Word reading accuracy

Reading rate or fluency

Reading comprehension

Note: *Dyslexia* is an alternative term used to refer to a pattern of learning difficulties characterized by problems with accurate or fluent word recognition, poor decoding, and poor spelling abilities. If dyslexia is used to specify this particular pattern of difficulties, it is important also to specify any additional difficulties that are present, such as difficulties with reading comprehension or math reasoning.

315.2 (F81.81) With impairment in written expression:

Spelling accuracy

Grammar and punctuation accuracy

Clarity or organization of written expression

315.1 (F81.2) With impairment in mathematics:

Number sense
Memorization of arithmetic facts
Accurate or fluent calculation
Accurate math reasoning

Note: *Dyscalculia* is an alternative term used to refer to a pattern of difficulties characterized by problems processing numerical information, learning arithmetic facts, and performing accurate or fluent calculations. If dyscalculia is used to specify this particular pattern of mathematic difficulties, it is important to also specify any additional difficulties that are present, such as difficulties with math reasoning or word reasoning accuracy.

Specify current severity:

Mild: Some difficulties learning skills in one or two academic domains, but may be of mild enough severity that the individual may be able to compensate or function well when provided with appropriate accommodations or support services, especially during the school years.

Moderate: Marked difficulties learning skills in one or more academic domains, so that the individual is unlikely to become proficient without some intervals of intensive and specialized teaching during the school years. Some accommodations or supportive services at least part of the day at school, workplace, or at home may be needed to complete activities accurately and efficiently.

Severe: Severe difficulties learning skills, affecting several academic domains, so that the individual is unlikely to learn those skills without ongoing intensive individualized and specialized teaching for most of the school years. Even with an array of appropriate accommodations or services at home, school, or workplace, the individual may not be able to complete all activities efficiently.

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 76–78 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

This diagnostic category is indicated by using one overarching term, “Specific Learning Disorder,” which incorporates difficulties in learning academic skills, such as reading, writing, and mathematics, which were classified as separate disorders in previous versions of DSM and in ICD-10. The generic term reflects the evidence from behavior–genetic studies showing that learning difficulties in the various academic domains share genetic variance and from clinical and epidemiologic studies that individuals showing a deficit in one academic domain frequently show deficits in other domains as well, although deficits isolated to one academic domain may occur. Using this overarching term, diagnosticians can then specify the details of the areas of learning difficulty and their severity as manifest at the time of assessment.

The manifestation of learning difficulties *at the time of assessment* is coded by adding *Specifiers* to the generic diagnosis of SLD to provide more precise information about the domains that are affected. Specifiers are deficits in reading, deficits in writing, deficits in mathematics, each of which can be further specified by detailed descriptions (e.g., SLD with impairment in reading [currently manifests as problems in word reading accuracy, reading fluency, and comprehension] and impairment in mathematics [currently manifests as problems with memorization of math facts, as well as in accuracy of calculation]). Use of the terms “dyslexia” and “dyscalculia” to refer to problems in learning to read or do arithmetic, respectively, is permitted according to clinician and patient preference, but detailed coding of the manifestation of SLD should still be completed. Severity ratings (mild, moderate, severe), which reflect the dimensional nature of SLD, permit further description by considering the range of learning difficulties and impairments, as well as the likely need for accommodations and support to best promote ongoing learning. DSM-5 specifies a set of four diagnostic criteria, all of which must be met to confirm a clinical diagnosis of SLD.

Criterion A. This criterion lists a range of observable descriptive behaviors or symptoms (Criterion A1 to A6, Table 40–3) which may be observed, probed during the clinical assessment, and confirmed by evidence from the school reports, teacher observations or ratings, previous psychoeducational assessment reports, and current performance scores on standardized achievement tests. At least one of these symptoms must be present. Moreover, Criterion A also requires evidence that the symptoms have persisted for at least 6 months despite the provision of intervention that targets each of the affected academic domains. For children, the term "*persistent*" is confirmed by the evidence of restricted progress in mastering the affected academic skill, over at least 6 months, despite the provision of intervention that targets that affected skill. For older adolescents and adults, persistent difficulties refer to ongoing difficulties with literacy or numeracy skills that first manifest during the school years, as confirmed by school reports, evaluated portfolios of past and current work, or previous assessments.

The term "*intervention*" refers to extra help at school or home, small group or individualized instruction, using any instructional approaches available at the school that focus on the affected skill. DSM-5 does not require the provision of intensive evidence-based (research-based) intervention—as specified by a current educational model in the United States (Response-to-Intervention [RTI]). RTI is a multitiered approach to academic and behavioral intervention designed to provide early, systematic, and appropriately intensive assistance to children who are at risk or already underperforming academically or behaviorally, compared to grade-appropriate standards. To date, the RTI approach, while laudable in its goals, not only lacks a strong evidence base but also is not consistently available within the United States and is rarely available outside of the United States.

Criterion B. This comprises two components: (i) academic achievement that is well below average and (ii) impairment resulting from the learning difficulties. In terms of low academic achievement, the DSM-5 applies the "average-person standard" that has been applied in disability discrimination laws for years. An individual's scores on standardized, individually administered, culturally sensitive academic achievement tests are compared to the population mean for chronologic age, not to IQ or to achievement scores of subsets of the population, such as peers attending university, medical, or law school. Examples of commonly used standard achievement tests are shown in Table 40–4.

DSM-5 does not specify what cut-off scores to use for determining low academic achievement within the diagnostic criterion itself, but provides guidelines within the text. One suggested guideline for below-average academic performance is to require evidence that scores on standardized academic achievement tests fall at least 1.5 standard deviations below the general population mean for age: This cut-off score is equivalent to a standard score >78 or >7th percentile. Standardized tests vary in the number and composition of items relevant to age, as well as in the particular aspect of reading, arithmetic, spelling, and written expression. Moreover, it is important to recall that standardized testing occurs in rather artificial situations (quiet space, minimal noise or other distractions, use of a standardized procedure that often precludes repetition of a question by the examiner), and performance may be influenced by factors other than academic skill level, such as motivation and effort, familiarity with the testing situation, anxiety, hunger, and sleep status, etc. Thus, greater diagnostic certainty is established by the evidence of low scores on more than one test or subtest.

A second guideline suggests that more lenient cut-off scores may be used (1.0 to 1.5 standard deviations below mean for age) when supported by other converging evidence of learning difficulties, such as academic history, other test scores, school reports, teacher observations, etc.

For individuals aged 17 and older, a documented history of long-standing and impairing learning difficulties (based on self-reported history of difficulties learning to read, spell, do arithmetic or written work; previous standardized testing; school reports; previous systematic clinical assessments for SLD) may be substituted for the standardized assessment. The goal of this substitution is to avoid the time and expense needed for unnecessary repeated standardized assessments, based on the evidence that childhood-onset SLD typically persists into adulthood.

Identification of SLD in individuals in postsecondary education poses challenges. DSM-5 does not preclude the possibility that an individual with high IQ, giftedness, or attending postsecondary education may have SLD. Evidence required for greater diagnostic certainty in this situation includes previous or current reports (from school, significant others), self-reported history of marked learning difficulties, a manifestation of self-report of an extraordinary effort required or dependence on the provision of support and accommodations needed to attain/maintain performance; previously evaluated portfolios of academic work; previous clinical assessments indicating SLD; family history of SLD; or current evidence of low academic scores (e.g., below the population mean for age). Extensive testing is rarely required: Research evidence indicates that tests of fluent word reading, reading comprehension, word spelling, number sense, and fluency of arithmetic computation suffice.

In the absence of available culturally appropriate and normed standardized academic achievement tests (as occurs in many low- to middle-income countries or non–English-speaking countries), converging evidence of markedly low academic performance must be derived from careful clinical review of scores on locally available tests, school reports, and clinical assessment. (See Tables 40–1 to 40–4 for a description of commonly used tests.)

Table 40–4.

Examples of Commonly Used Standardized Tests of Reading, Written Expression, Arithmetic, and Related Skills

Name of Standardized Test	Reading Skills
Comprehensive Test of Phonologic Processing (CTOPP); standardized. Ages: 5 years to young adulthood	Basic reading skills (e.g., phonologic awareness, phonologic coding, working memory, rapid naming)
Woodcock Johnson-III Achievement battery	Letter–word identification, word attack, and reading fluency reading comprehension
Test of Word Reading Efficiency (TOWRE); Ages: 5 years to young adulthood	Single-word reading efficiency (e.g., speedy and accurate naming of single words and pseudowords)
Gray Oral Reading Test; Nelson–Denny Reading Test; Gray Silent Reading Test	Reading fluency and comprehension (e.g., rate of reading connected text for meaning)
	Written Expression Skills
Wechsler Individual Achievement Test (WIAT-II) Ages: 5–19 years	Spelling, written expression
Test of Written Language (TOWL; 3rd edition) Ages: 7.5–17.9 years	1. Elicited writing, including vocabulary, spelling, capitalization and punctuation, logical sentences, and sentence combining 2. Spontaneous writing combining
Test of Early Written Language (TEWL; 2nd edition) Ages: 3–11 years	Basic, global, and contextual writing
Test of Written Spelling (TOWS;	Spelling of phonetically regular

4th edition) Ages 6–18 years	and nonregular words
Woodcock–Johnson Psychoeducational Battery Ages: 5 years–adult	Dictation, proofing, writing fluency, writing samples
	Mathematics Skills
Woodcock–Johnson Achievement Battery-III	Math fluency, calculation, quantitative concepts, applied problems
Test of Early Mathematical Abilities—3rd edition Age: 3–0 to 8–11	Numbering skills, number-comparison facility, numeral literacy, mastery of number facts, calculation skills, and understanding of concepts
Test of Mathematical Abilities Age: 8–0 to 18–11	Mathematical symbols and concepts; computation; mathematics in everyday life; word problems; attitude toward math
Teacher Academic Attainment Scale (TAAS)	Subjective rating scale of student academic achievement (7 items)
Math Anxiety Questionnaire	Math anxiety
Abbreviated Math Anxiety Scale	1. Learning anxiety: anxiety related to learning math 2. Testing: anxiety related to being tested in math
Subjective Numeracy Scale	Subjective rating of one's health numeracy
Test of Functional Health Literacy in Adults (s-TOFHLA)	Objective measure of health literacy
Rapid Estimate of	Objective measure of health

Adult/Adolescent Literacy in Medicine (REALM)	literacy

Moreover, Criterion B emphasizes impairment, in requiring that the learning difficulties cause significant impairment in academic, occupational, social functioning, or in activities in daily living that requires academic skills. Thus, low academic scores per se are not sufficient to meet the specifications of Criterion B: Evidence of associated impairment is required. This component is in line with the standard of the U.S. “Americans with Disabilities Act (ADA)” for determining a disability.

Criterion C. This third criterion requires evidence that the symptoms of SLD manifest in the school years—most reliably interpreted as during the elementary (primary) school years. Symptoms of SLD may manifest in preschool years, but a diagnosis cannot be confirmed until the individual has received formal instruction in the core academic skills. However, a diagnosis of SLD can be made for the first time in adolescence or adulthood. This is because the manifest learning difficulties may have been overlooked, not assessed, or attributed solely to concurrent emotional behavioral difficulties. Also, some individuals (particularly those with higher adaptive and intellectual functioning, or in highly supportive environments) may have coped with the demands for the core academic skills during the school years, albeit by putting forth extraordinary effort, or by using compensatory skills, that no longer suffice in later years, when high demands on accurate and fluent academic skills (e.g., reading or writing lengthy complex reports for a tight deadline, excessively heavy academic or occupational loads) start to exceed the individual’s limited capacities and compensatory mechanisms. In such cases, greater diagnostic certainty is established by converging evidence from comprehensive clinical assessment, and the ability to rule out loss of previously acquired academic skills as a result of a major medical event or condition.

Criterion D. The fourth criterion refers to a range of exclusionary conditions. These include intellectual disability, other major mental or neurologic conditions, uncorrected visual or auditory acuity, severe psychosocial adversity, lack of proficiency in the language of instruction (for adults, lack of proficiency in the language of the workplace or social community), lack of or inadequate educational instruction, as may occur for children excluded from education for various reasons (e.g., because they are girls; living in war zones, no schools, no teachers, etc.). These exclusionary criteria are evaluated by means of a comprehensive clinical assessment.

As of 2012, UNICEF reports that about 58 million children are excluded from primary school education and 63 million from lower secondary school education (http://www.uis.unesco.org/Education/Pages/out-of-school-children.aspx). Certainly, it makes no sense to make a diagnosis of SLD in immigrant children who previously have been excluded from education. The issue of excluding from the SLD category those children whose low academic scores may be due to instructional factors (inadequate instruction; a limited opportunity to learn) is important but has been challenging for clinicians to operationalize. One component of Criterion A attempts to address this challenge by requiring evidence that symptoms of learning difficulties have persisted for at least 6 months despite the provision of intervention that targets the affected academic skills. As noted previously, the RTI approach was not adopted by DSM-5 due to insufficient evidence, and lack of universal availability of RTI approaches. However, in many countries, the quality of teacher training has improved over recent years (especially in teaching children to read), affording some assurance that a student’s limited progress is not due to inadequate instruction, so SLD is a possibility.

Identification of SLD in individuals whose first language (mother tongue) differs from the language of instruction is challenging. Greater diagnostic certainty is accorded by evidence of the following: The individual has a history of language delay or disability in the native language; the individual has had difficulty developing literacy or numeracy skills in the native language (assuming adequate instruction in the native language); there is a family history of SLD in parents, siblings, or other close relatives, the individual is not making adequate progress over 6 months compared to other nonnative speakers, despite the provision of targeted intervention.

Does Information from Tests of Intellectual Capacity or Cognitive Processing Assist in the Diagnosis of SLD or Guide Educational Treatment?

Intelligence testing (and IQ profiling using various subtests) with the requirement for a significant discrepancy between intellectual and academic achievement has been the primary and often sole criterion for the diagnosis of SLD in previous versions of the DSM, continues to be used internationally to determine eligibility for special education services and accommodations, and figures strongly in the understanding of SLD by various stakeholder groups (e.g., teachers, psychologists, parents and parent organizations, lobby groups and researchers). Moreover, this “discrepancy” model has become ingrained in clinical training and practice, despite the several decades of research evidence against its validity for diagnosis or utility in guiding educational instruction. For example, discrepancy scores are unreliable, and children with learning difficulties who do and do not meet the IQ-achievement discrepancy do not differ in symptomatology, associated cognitive deficits, or response to intervention. In other words, the purported differences between those with discrepant and nondiscrepant IQ-achievement scores have not proven to be clinically or educationally meaningful. Thus, a measure of intelligence is not required for a diagnosis of SLD (unless there are concerns about the individual’s intellectual ability and adaptive behavior).

Accordingly, based on the available scientific evidence, DSM-5 eliminated the requirement for a significant discrepancy between the IQ score and an achievement score for the diagnosis of SLD. The DSM-5 does not prohibit the inclusion of intellectual assessment in the diagnostic assessment of SLD. Occasionally, information from an IQ test might be helpful in changing a teacher or parent’s overall view of the child and their misunderstanding of the child's intellectual ability. However, when formulating a diagnosis of SLD, clinicians should also keep in mind the DSM-5 changes to the conceptualization and criteria for Intellectual Disability: Specifically, poor adaptive and intellectual functioning, with the latter suggested by IQ score less than 70 ± 5 standard points. This means that individuals with an IQ score of greater than 70 ± 5 standard points and adequate adaptive functioning could potentially—but not necessarily—meet the criteria for SLD since the differential diagnosis of Intellectual Disability can be ruled out.

In recent years, testing of cognitive processes (e.g., phonologic processing, rapid automatized naming, auditory and visual processing, processing speed) and executive functions (e.g., working memory, inhibitory control, cognitive flexibility, sustained attention) has been incorporated increasingly in assessment for SLD, in addition to IQ testing. The goal of such testing is to document patterns of cognitive strengths and weaknesses to inform the development of differentiated and more effective educational instruction and intervention. However, it is unclear whether the relationship between SLD and these processes or executive functions is causal or correlational. There is little or no robust evidence that training these cognitive processes and executive functions will boost academic learning particularly in students with SLD. Also, recent meta-analyses conclude that effect sizes of the magnitude of the relationship between children's cognitive characteristics and response to intervention were small and unable to improve clinical prediction from measures of academic skills per se. Moreover, there is little evidence that cognitive profiles are valuable for planning intervention (i.e., little evidence for Aptitude X Treatment interactions). Thus current recommendations are for assessment and intervention for SLD to focus on educational rather than cognitive factors—a position that is understandably threatening to clinicians and other professionals trained in the psychometric tradition and who might lack expertise in curricular and pedagogical practices. Hence the DSM-5 criterion B does require the use of individually administered standardized tests of academic achievement in reading, writing, and arithmetic, where possible.

DIFFERENTIAL DIAGNOSIS

SLD can be distinguished from normal variation (individual differences) in learning academic skills due to other internal and external factors by the severity of the learning difficulties, the specificity of the difficulties in the presence of intact intellectual abilities, and their persistence in the presence of adequate educational opportunity and receipt of the same instruction as the peer group, targeted home or school-based intervention, competency in the language of instruction (even when different from one's mother tongue).

A diagnosis of SLD is not usually made before the child is about 7 years old because evidence of failure to learn to read, write, or do arithmetic is required. Difficulties in learning foundational academic skills may occur for a variety of reasons (e.g., lack of adequate exposure to print, limited education) and are associated with numerous mental health conditions, including mental retardation, developmental coordination disorder, communication disorders, ADHD, conduct disorder, hearing or visual problems, and prematurity. The most common distinguishing feature of SLD in reading is a persistent impairment in phonologic processing and spelling; that of SLD in mathematics is a persistent impairment in number facts. In a child with visual or hearing deficits, the learning problems must be in excess of the difficulty typically associated with these sensory disorders. In a child with learning difficulties in the presence of low intellectual abilities (i.e., an IQ score of 70 ± 5), it is important to assess adaptive behavior to be able to rule out Intellectual Disabilities.

SLD is distinct from Intellectual Disabilities: Individuals diagnosed with intellectual disability have reduced cognitive capacity that has a global impact on learning and adaptive functioning in everyday life. By contrast, individuals with a diagnosis of SLD have at least average intellectual capacity but exhibit significant difficulties in one or more *academic* areas, while coping well or even excelling in other areas of academic, sporting, or artistic achievement. Moreover, SLD is distinguished from “learning difficulties,” which is a broader term that encompasses difficulties learning as a consequence of a range of causes, including intellectual disabilities, physical or sensory deficits (e.g., hearing or visual impairments), emotional or behavioral difficulties, inadequate environmental experiences, or appropriate educational opportunities.

SLD in Reading, Mathematics, or Written Expression is frequently associated with ADHD, and a concurrent diagnosis is not only permissible but is also essential. An explicit statement should be made concerning the presence of both clinical conditions to ensure that the child can access relevant and effective interventions. Moreover, even if the individual does not meet the full criteria for ADHD but does manifest persistent symptoms of inattention (particularly in the school setting), clinicians should report these problems, given the increased risk conferred for poor response to intervention if inattention is not targeted in the treatment plan.

It might be possible to determine inadequate instruction as the primary explanation for reading problems through detailed interview with parents and teachers. Probes for the number of schools attended, absenteeism (voluntary due to illness or vacation; required due to class or school suspensions and expulsions), and type of reading instruction provided at the school(s) in the primary grades may be informative. However, differential diagnosis may be complicated by parental history or current manifestation of SLD, given that parents with SLD would be less likely to be able to provide the necessary early reading or numerical experiences for the child.

COURSE AND PROGNOSIS

Converging evidence indicates that SLD is persistent and does not remit with increasing age in the absence of effective intervention. SLD tends to be a persistent and chronic condition that remains with the individual his or her entire life, albeit with developmental changes in its manifestation.

Onset, recognition, and diagnosis of SLD usually occurs during the first few years of elementary school when children are expected to learn to read, write, spell, and do basic arithmetic. Precursor signs (e.g., delayed onset of talking; difficulties with counting or rhyming; problems with fine motor skills required for writing) commonly occur in preschool years before the start of formal schooling. Moreover, preschoolers with poor phonologic short-term memory (STM) skills and specific language impairment (SLI) are highly likely to meet diagnostic criteria for SLD in reading at school age. However, referral and formal diagnosis rarely occurs until a couple of years later in second grade, when the children have failed to respond to formal classroom instruction in reading.

SLD in elementary school–aged children (grades 1 to 3) typically manifests first as marked difficulty learning to connect letters with the sounds they make (particularly in English-speaking children), to recognize and read common one-syllable words (e.g., mat, top) or common but irregularly spelled words (e.g., said, two). Reading aloud is slow, inaccurate, and effortful. Some children may also (or instead) show problems in learning to count, remember math facts, and perform simple single-digit arithmetic. Recent evidence suggested that it is possible to estimate an individual child’s risk for SLD in reading at school age from the following factors measured in preschool years (age 3.5 years): Familial history of reading problems, poor grammar and speech articulation, poor letter and object naming speeds, and poor phonologic awareness.

In the middle grades (grades 4 to 6), children with SLD may not only continue to show problems with word reading and basic number sense but also may manifest problems with spelling, writing, and understanding word problems in mathematics. Others may gradually be able to sound out words, but their reading is slow and effortful, they have difficulty understanding what they have just read, and also problems start to manifest in their written work (e.g., poor spelling, poor punctuation, words left out of sentences, short, disconnected sentences, illegible and disorganized writing). Some youngsters continue to count on their fingers since they cannot remember math facts. It is not uncommon, even at this young age, for children to dread going to school or to try to avoid reading, writing, or arithmetic.

By adolescence, some youngsters may show improved word recognition and the ability to sound out words, but reading remains slow and effortful, often with poor comprehension, and so they cannot keep up with the increasing demands for reading complex material. They often need to reread material to get even the main point and have great difficulty making inferences from written text. Written work is markedly impoverished: numerous spelling errors, illegible writing, with poor sentence, and paragraph structure. Those who showed problems in basic number sense at a younger age may use long and inefficient strategies to work out even simple arithmetic, such as “7 × 8” (e.g., draw 8 groups of 7 items, and then counting all of the items). Ongoing dread of and failure in school may result in high-school dropout, anxiety, and depression.

A longitudinal investigation of a 1990 U.S. cohort of children reveals that individuals with SLD in word reading (a.k.a., dyslexia) show an achievement gap in reading that is discernible upon school entry in grade 1 (about age 6) and that persists throughout childhood and adolescence. A more recent study by Shaywitz et al., in 2015, examined 414 students from the Connecticut Longitudinal Study to determine if differences between dyslexic and typical readers in their reading scores and verbal IQ are evident as early as first grade, as well as whether the trajectory of these differences is seen throughout adolescence. The study found that a significant achievement gap was evident as early as first grade, and the gap persisted into middle and high school. It remains unknown whether early reading intervention (i.e., kindergarten, grade 1) will eliminate the observed persistent difference in growth trajectories in reading. Likewise, a longitudinal investigation of a large sample of kindergarten children recruited from 1998 to 1999 and followed through 8th grade revealed that children with SLD in mathematics enter kindergarten with lower skill levels, and the achievement gap persists through grade 8.

Young adults with persistent problems in both reading accuracy and fluency appear to develop compensatory mechanisms to cope with reading, such as greater reliance on memory-based rather than analytic strategies to assist in word identification. By adulthood, many individuals try to hide the fact that they cannot read, spell, or do arithmetic by avoiding both leisure and work-related activities that require these skills, or use alternative approaches to access print (e.g., text-to-speech or speech-to-text software, audiobooks, audiovisual media).

The clinical course of SLD is often complicated further by comorbidity with other mental health problems, such as ADHD or conduct disorder, as well as with low self-esteem, anxiety, depression, and increased risk for suicidal ideation and suicide attempts. The causal relationships between SLD and the various comorbid disorders are unknown and longitudinal epidemiologic studies suggest that the relationships are complex.

On a more positive note, a recent study reported that adults with SLD in reading (a.k.a., “dyslexia”) are overrepresented among entrepreneurs and manifest higher levels of visual problem-solving skills compared to peers with average reading abilities. This subgroup of individuals attributed their success to an admixture of high self-motivation, educational accommodations, and a supportive social network, particularly parents.

Juan P, a 20-year-old Hispanic college student, presented at the primary care clinic in his home city, with complaints of increasing problems sleeping, inability to concentrate or manage his college schedule, failing some courses, and feeling “down” most days. Clinical history indicated a normal birth, slowness in learning to talk, childhood problems with concentration, hyperactivity and angry outbursts, as well as struggles learning to read, at school. Family history was positive for ADHD (older brother) and SLD (father, now deceased, was reported to have had “dyslexia” and dropped out of school early). He had been diagnosed with ADHD, Predominantly Combined Type, at age 9, but did not meet criteria for SLD, since the discrepancy between his intellectual capacity and reading achievement was deemed not to be “substantial” as required by DSM-IV and his local school board.

During the clinical interview, JP reported that since starting college a year ago, he had been unable to remain focused while reading or listening to lectures, felt overwhelmed, and was unable to complete written assignments on time. He complained of feeling restless, agitated, exhausted, and often in a “blue mood.” He denied any substance use or suicidal thoughts. He stopped taking medication for ADHD when he was aged 12 and had not sought registration with Student Disability Services at his college for help with his ADHD because he thought he could get through college by himself. The current psychoeducational assessment confirmed that JP’s reading fluency and comprehension, spelling, and written expression, were substantially and quantifiably below that expected for age and were causing significant impairment.

Diagnostic formulation: *Attention-Deficit/Hyperactivity Disorder*, with predominantly inattentive presentation, of mild to moderate severity; *Specific Learning Disorder*, with impairment in reading (fluency, comprehension) and written expression (spelling, organization of written expression), all currently of moderate severity; depressive symptoms but did not meet criteria for a mood disorder.

TREATMENT

Treatment focuses almost exclusively on academic intervention and rarely includes mental health services even though students with SLD experience higher levels of inattention (and ADHD), anxiety, and depression than their non-SLD peers. Moreover, treatment of SLD is typically carried out in education settings or specialized psychology clinics, and so is not the primary responsibility of psychiatrists or primary care providers.

The primary role of medical health care providers is that of coordinator, which, as a first step, includes the provision of psychoeducation about the disorder and its longer-term implications. Thus, physicians need to be able to explain what SLD is and refer parents (or the adult patient) to organizations, websites, or reading materials that will help (see the References for some useful websites). Also, parents may need help in understanding their child's needs and rights under the IDEA, including possible need for accommodations. Given that SLD tends to run in families, physicians may need to modify how any numeric health information is communicated (prevalence rates, risk ratios, dosage, etc.), as well as to check that the individual or caregivers are able to comprehend any relevant printed or website material.

A second critical step is to establish a monitoring plan with the patient and family, to permit early detection and management of commonly associated mental health problems, such as ADHD, anxiety, and depression, as well as to ascertain whether recommended accommodations have been put in place along with intensive, individualized, explicit instruction. Moreover, the need for transitional planning is essential to help individuals with SLD through major developmental and life transitions (e.g., transition from elementary to secondary school; from secondary school to the workforce or postsecondary education; marriage, parenthood, etc.) and facilitate access to services.

One of the major problems is that the standard instruction in regular classrooms is too little, too general, and nonsystematic. For example, one large-scale study found that children receiving special education over a 1-year period advanced in reading by only 0.04 standard deviations per year, which means that it would take at least 8 years to raise these students' scores from the 5th to the 9th percentile. By contrast, students with SLD in reading who received 1 to 2 hours daily of intense and focused intervention for a total of 35 to 70 hours of instruction showed dramatic gains in performance on standardized measures of both decoding and comprehension, which were maintained for several years after intervention. The instruction programs were very explicit and visible, conducted in small, interactive groups of children matched for reading abilities, which promoted engaged and repeated reading of connected text presented in a way that controlled task difficulty, and included instruction in basic reading abilities (e.g., teaching phoneme-to-grapheme correspondences), as well as the application of metacognitive strategies for decoding unfamiliar words and comprehending connected text. Of note, preliminary evidence suggests that intense and focused instruction may, in fact, alter the aberrant brain activation profiles observed in children with SLD. Thus, clinicians need to be aware that recommendations for placement in special education may not be sufficient to improve reading or arithmetic skills. Rather, they need to advocate for the provision of intense, focused, and explicit instruction in each of the affected academic domains.

Also, given the importance of literacy and numeracy skills in health, clinicians are advised to screen for and monitor health literacy and numeracy in all patients (and parents of pediatric patients), regardless of their referral problem. Low literacy or numeracy cannot be inferred reliably on the basis of the patient's education level, intelligence, or other observable behaviors. Reliable short screening forms can be used to assess health literacy and numeracy in adolescents and adults and to guide health communication.

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Ashkenazi S, Black JM, Abrams DA, Hoeft F, Menon V. Neurobiological underpinnings of math and reading learning disabilities. *J Learn Disabil*. 2013;46(6):549–569.

Bishop DVM. The interface between genetics and psychology: lessons from developmental dyslexia. *Proc Biol Sci*. 2015;282(1806):20143139.

*Butterworth B, Kovas Y. Understanding neurocognitive developmental disorders can improve education for all. *Science*. 2013;340(6130):300–305.

Büber A, Başay Ö, Şenol H. The prevalence and comorbidity rates of specific learning disorder among primary school children in Turkey. *Nord J Psychiatry*. 2020;74(6):453–460.

*Callaway E. Dyscalculia: number games. *Nature*. 2013;493(7431):150–153.

Cho K, Frijters JC, Zhang H, Miller LL, Gruen JR. Prenatal exposure to nicotine and impaired reading performance. *J Pediatr*. 2013;162(4):713–718.

de Zeeuw EL, de Geus EJC, Boomsma DI. Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. *Trends Neurosci Educ*. 2015;4(3):69–76.

Donker M, Kroesbergen E, Slot E, Van Viersen S, De Bree E. Alphanumeric and non-alphanumeric rapid automatized naming in children with reading and/or spelling difficulties and mathematical difficulties. *Learn Indiv Differ*. 2016;47:80–87.

*Ferrer E, Shaywitz BA, Holahan JM, Marchione KE, Michaels R, Shaywitz SE. Achievement gap in reading is present as early as first grade and persists through adolescence. *J Pediatr*. 2015;167(5):1121–1125.

Fortes IS, Paula CS, Oliveira MC, Bordin IA, de Jesus MJ, Rohde LA. A cross-sectional study to assess the prevalence of DSM-5 specific learning disorders in representative school samples from the second to sixth grade in Brazil. *Eur Child Adolesc Psychiatry*. 2016;25(2):195–207.

*Hulme C, Snowling MJ. The interface between spoken and written language: developmental disorders. *Philos Trans R Soc Lond B Biol Sci*. 2013;369(1634):20120395.

Katusic SK, Colligan RC, Weaver AL, Barbaresi WJ. The forgotten learning disability: epidemiology of written-language disorder in a population-based birth cohort (1976–1982), Rochester, Minnesota. *Pediatrics*. 2009;123(5):1306–1313.

*Kaufmann L, von Aster M. The diagnosis and management of dyscalculia. *Dtsch Ärztebl Int*. 2012;109(45):767–777.

Kere J. The molecular genetics and neurobiology of developmental dyslexia as model of a complex phenotype. *Biochem Biophys Res Commun*. 2014;452(2):236–243.

Klassen RM, Tze VM, Hannok W. Internalizing problems of adults with learning disabilities: a meta-analysis. *J Learn Disabil*. 2013;46(4):317–327.

Kucian K, von Aster M. Developmental dyscalculia. *Eur J Pediatr*. 2015;174(1):1–13.

Locke R, Scallan S, Mann R, Alexander G. Clinicians with dyslexia: a systematic review of effects and strategies. *Clin Teach*. 2015;12(6):394–398.

Lyytinen H, Erskine J, Hämäläinen J, Torppa M, Ronimus M. Dyslexia—early identification and prevention: highlights from the Jyväskylä longitudinal study of dyslexia. *Curr Dev Disord Rep*. 2015;2(4):330–338.

Mascheretti S, Bureau A, Battaglia M, et al. An assessment of gene-by-environment interactions in developmental dyslexia-related phenotypes. *Genes Brain Behav*. 2013;12(1):47–55.

Moll K, Kunze S, Neuhoff N, Bruder J, Schulte-Körne G. Specific learning disorder: prevalence and gender differences. *PLoS One*. 2014;9(7):e103537.

Nelson JM, Harwood H. Learning disabilities and anxiety: a meta-analysis. *J Learn Disabil*. 2011;44(1):3–17.

Nelson JM, Harwood HR. A meta-analysis of parent and teacher reports of depression among students with learning disabilities: evidence for the importance of multi-informant assessment. *Psychol Sch*. 2011;48(4):371–384.

Niklas F, Schneider W. Casting the die before the die is cast: the importance of the home numeracy environment for preschool children. *Eur J Psychol Educ*. 2014;29(3):327–345.

Peng P, Fuchs D. A meta-analysis of working memory deficits in children with learning difficulties: is there a difference between verbal domain and numerical domain? *J Learn Disabil*. 2016;49(1):3–20.

Pennington BF, Santerre-Lemmon L, Rosenberg J, et al. Individual prediction of dyslexia by single versus multiple deficit models. *J Abnorm Psychol*. 2012;121(1):212–224.

*Peterson RL, Pennington BF. Developmental dyslexia. *Annu Rev Clin Psychol*. 2015;11:283–307.

Plomin R, Kovas Y. Generalist genes and learning disabilities. *Psychol Bull*. 2005;131(4):592–617.

Polanska K, Jurewicz J, Hanke W. Smoking and alcohol drinking during pregnancy as the risk factors for poor child neurodevelopment—A review of epidemiological studies. *Int J Occup Med Environ Health*. 2015;28(3):419–443.

Rimfeld K, Kovas Y, Dale PS, Plomin R. Pleiotropy across academic subjects at the end of compulsory education. *Sci Rep*. 2015;5:11713.

Sexton CC, Gelhorn HL, Bell JA, Classi PM. The co-occurrence of reading disorder and ADHD: epidemiology, treatment, psychosocial impact, and economic burden. *J Learn Disabil*. 2012;45(6):538–564.

Simms V, Cragg L, Gilmore C, Marlow N, Johnson S. Mathematics difficulties in children born very preterm: current research and future directions. *Arch Dis Child Fetal Neonatal Ed*. 2013;98(5):F457–F463.

Smedberg J, Lupattelli A, Mårdby AC, Nordeng H. Characteristics of women who continue smoking during pregnancy: a cross-sectional study of pregnant women and new mothers in 15 European countries. *BMC Pregnancy Childbirth*. 2014;14(1):213.

Snowling MJ, Melby-Lervåg M. Oral language deficits in familial dyslexia: a meta-analysis and review. *Psychol Bull*. 2016;142(5):498–545.

Stoeckel RE, Colligan RC, Barbaresi WJ, Weaver AL, Killian JM, Katusic SK. Early speech-language impairment and risk for written language disorder: a population-based study. *J Dev Behav Pediatr*. 2013;34(1):38–44.

Swanson E, Vaughn S. Learning disabilities: academic and mental health needs. In: Holt MK, Grills AE, eds. *Critical Issues in School-based Mental Health: Evidence-Based Research, Practice, and Interventions*. Routledge; 2015.

Tannock R. Specific learning disabilities in DSM-5: are the changes for better or worse? *Int J Res Learn Disabil*. 2013;1(2):2–30.

Tannock R. Provision of evidence-based intervention is not part of the DSM-5 diagnostic criteria for specific learning disorder. *Eur Child Adolesc Psychiatry*. 2016;25(2):209–210.

41

Developmental Coordination Disorder

VIJAY VISHWANATH, M.D., PH.D., AND WENDY G. MITCHELL, M.D.

Developmental coordination disorder (DCD) is a neuromotor disability emerging in the early developmental period, which can affect both fine and gross motor skills, interfering with daily activities. Children with DCD have a motor performance that is typically slower, less accurate, and more variable than peers of the same age. DCD is estimated to occur in 5% to 6% of children aged 5 to 11 years, and 1.8% of children 7 years of age exhibit severe DCD. DCD affects boys more often than girls; most recent population studies report ratios of male:female of about 2:1. Children exhibiting DCD often show impairment in gross motor skills such as jumping, hopping, and running, fine motor skills, including handwriting, tying their shoelaces, or using utensils, and balance. Motor coordination disorder is comorbid up to 50% of the time with attention-deficit/hyperactivity disorder (ADHD). Other disorders that may co-occur with DCD are communication disorders and specific learning disorders. Delayed motor milestones may be the first sign of DCD in the absence of intellectual disability. Without intervention, DCD often continues to cause impairment throughout childhood and adolescence, leading to significant consequences physically, educationally, socially, and psychiatrically.

DEFINITION

The essential feature of DCD, according to DSM-5-TR, is impaired motor skills leading to deficits in coordination compared to children of the same chronologic age and intellectual level, interfering with activities of daily living, social activities, and school productivity. Key symptoms are marked delays in achieving motor milestones, such as walking, crawling, and sitting, constantly dropping objects, inability to catch or throw a ball, and poor handwriting. Children with DCD perform motor coordination tasks slower and less accurately than those peers of the same chronologic age and intellectual capacity. Specific manifestations of DCD are detailed in Table 41–1.

HISTORY

In the early 1900s, childhood motor clumsiness was termed *motor deficiency syndrome*. This syndrome was characterized by awkward voluntary actions, excessive tendon reflexes, mild hypertonicity, and neurologic overflow movement. The term *clumsy child syndrome* first appeared in the literature in the 1930s to describe a set of symptoms associated with clumsy motor behaviors of unknown etiology, not associated with a known neurologic disorder. The observation was made that many children with motor coordination difficulties also exhibited ADHD, communication disorders, and specific learning disorders. During the 1940s, remedial programs were established in an attempt to improve academic learning deficits through the treatment of coordination difficulties observed in children.

By the 1960s, *clumsy child syndrome* was officially distinguished from other categories of learning disorders. From this period to the mid-1980s, significant investigations were conducted into perceptual-sensory causes for motor impairments, as well as social and emotional problems associated with such impairments. Children with clumsy behavior were noted to have deficits rather than solely delayed fine and gross motor coordination skills.

COMPARATIVE NOSOLOGY

Many terms have historically been applied to the description of *clumsiness* in a child. Terms used in the past that overlap with a description of DCD include congenital maladroitness, choreiform syndrome, psychomotor syndrome, developmental apraxia, developmental dyspraxia, congenital clumsiness, developmental agnosia, and developmental dyspraxia–dysgnosia. It is only appropriate to make a diagnosis of DCD after any suspicion of a neurologic disorder is ruled out. In the DSM-5-TR, DCD is not diagnosed if symptoms are attributable to an intellectual developmental disorder or a neurologic condition such as cerebral palsy, muscular dystrophy, or a degenerative disorder. Although currently, this syndrome is still occasionally referred to as clumsy child syndrome, DCD and specific developmental disorder of motor function are more appropriate terms.

Table 41–1.
Manifestations of Developmental Coordination Disorder

Gross Motor Manifestations

Preschool age

Delays in reaching motor milestones, such as sitting, crawling, and walking

Balance problems: Falling, getting bruised frequently, and poor toddling

Abnormal gait

Knocking over objects, bumping into things, and destructiveness

Primary-school age

Difficulty with riding bikes, skipping, hopping, running, jumping, and doing somersaults

Awkward or abnormal gait

Older

Poor at sports, throwing, catching, kicking, and hitting a ball

Fine Motor Manifestations

Preschool age

Difficulty learning dressing skills (tying, fastening, zipping, and buttoning)

Difficulty learning feeding skills (handling knife, fork, or spoon)

Primary-school age

Difficulty assembling jigsaw pieces, using scissors, building with blocks, drawing, or tracing

Older

Difficulty with grooming (putting on makeup, blow-drying hair, and doing nails)

Messy or illegible writing

Difficulty using hand tools, sewing, and playing piano

DCD was included as a diagnosis in DSM-III-R and appears in essentially the same form in DSM-5-TR. The tenth edition of the *International Statistical Classification of Diseases and Related*

Health Problems (ICD-10) also includes the disorder, but the slightly more formalized definition requires an intelligence quotient of at least 70 and performance at least two standard deviations below the mean on a standardized test of motor coordination. In 1994, an international consensus meeting held in London, Ontario, agreed on the term DCD to describe children with a heterogeneous set of motor impairments that caused impairment in performing motor tasks compared to same-age peers. ICD-11 updated the disorder in by listing five essential criteria that more closely align with the DSM-5-TR criteria.

Although DSM-5-TR and ICD-10 and 11 view the disorder as a unitary undifferentiated syndrome, several investigators have suggested that there may actually be several distinct subtypes. Supporting that view is the individual variation in symptomatology, only some of which is clearly age-related. One recent attempt was to reclassify DCD proposed subtypes, such as ataxic syndrome (marked by unsteadiness and mild tremors), hypotonic syndrome (characterized by below-average resting muscle tone), tension syndrome (manifested by constant and excessively high levels of muscle tone), dyspraxic syndrome (identified by the inability to link a series of submovements necessary to execute smooth motor movements), manual–graphic syndromes (two separate manual syndromes, one marked by inadequate dexterity but adequate ability to reproduce written symbols and the other characterized by adequate dexterity but inability to reproduce written symbols accurately using fine motor movements), visual-perceptual clumsiness (ranging from ocular-motor problems involving inefficient movement of eye muscles to severe retinal deterioration), and mixed syndromes (a combination of symptoms already presented under the aegis of the previous six classifications). It is always important to keep in mind that significant ataxia, hypotonia, or any asymmetry between left- and right-side muscle function may be an indicator of a neurologic disorder and requires a neurologic assessment.

EPIDEMIOLOGY

The prevalence of DCD has been estimated to be approximately 5% to 6% among school-aged children. DCD has been noted in Europe, Asia, Africa, Australia, and North America. Clinical studies of children with DCD have historically found greater rates in boys than girls, of up to 7:1. More recent population studies have found less disparity, and report rates of approximately 2:1, male to female.

ETIOLOGY

A meta-analysis of performance deficits in DCD by Wilson and colleagues (2013) found that several clusters of deficits in learning motor skills emerged, suggesting that brain networks including parietocerebellar and cerebellar regions are involved. The clusters of motor skills dysfunction include “internal modeling of action” (predictive control of motor skills), rhythmic coordination and timing of motor movements, executive function, and dynamic control of posture and gait. The result of insufficient internal models, due to deficits in neural pathways in DCD leads to a lack of fluency and efficiency in coordinated movement. That is, the sequence of a complex motor movement in children with DCD is not properly or efficiently ordered, leading to awkward motions. Poor ability to coordinate a sequence of motor movements in DCD may be attributable to dysfunction in the corticocerebellar brain region. In terms of executive function deficits, impaired attention, working memory, and inhibition may contribute to poor action planning, and exacerbate overall poor motor coordination. The exact etiology of DCD is not yet clear; however, it is likely to include heritable factors, with potential prenatal and prenatal-developmental contributions. There is much heterogeneity in the expression of DCD. Motor control is dependent on numerous skills, including perception of objects in space, regulation of muscle activity, and balance. Children with DCD appear to be less proficient and less consistent in regulating muscle activity in heterogeneous ways. Some children with DCD are poor in manual skills, while exhibiting proficiency in balance, and for others, the opposite is true.

DCD and ADHD

Epidemiologic surveys have documented the comorbid occurrence of DCD in 30% to 50% of children diagnosed with ADHD. A study of children comorbid for ADHD and DCD in the Netherlands by Fliers et al., in 2007 confirmed previous research demonstrating a consistent increased risk of each of these disorders based on the presence of the other one. This study also investigated the association between the various types of motor coordination problems and inattentive versus hyperactive or impulsive subtypes of ADHD. In this study, data suggested that inattentive rather than hyperactive or impulsive symptoms were related to higher measures of motor coordination symptoms on the Developmental Coordination Disorder Questionnaire (DCD-Q) that was filled out by parents. This finding lends support for hypotheses that deficits in information processing are contributing problems underlying both ADHD and DCD. Although parents report that motor coordination problems often diminish with age, half of adolescents comorbid for ADHD and DCD continue to be impaired by coordination problems. One possible commonality between DCD and ADHD is a deficit in cerebellar activity. The coexistence of DCD and ADHD may reflect a shared heritable disturbance of the dopamine pathway. These two disorders may also share dysfunction in the basal ganglia and the prefrontal cortex.

DCD and Premature Birth

Evidence indicates biologic and cognitive-developmental correlates of DCD. The biologic-organic factors associated with the disorder include various birth and perinatal problems, such as low birth weight, prematurity, hypoxia, and neonatal malnutrition. A study by Marlow and colleagues investigated motor and “executive” cognitive function in 241 children who were born severely prematurely, at less than or equal to 25 complete weeks of gestation. These children were assessed at a median age of 6 years and compared with 160 classmates born at full term. Children underwent assessment of cognitive and neuromotor skills, including the Kaufman Assessment Battery for Children (K-ABC) and the NEPSY (a teacher-based assessment of academic achievement). Motor function was assessed by clinicians using the following items from the Movement Assessment Battery for Children (MABC): Heel-toe walking (15 steps), 1-leg balance (more than 20 seconds), and time to post 12 coins. Results from this study indicate that compared to classmates, children born severely prematurely and in the absence of cerebral palsy exhibited a high prevalence of impairment in gross motor function at early school age, in addition to impaired visual–spatial, perceptual-motor attention, and executive function. In each of the above impairments in motor skills or executive functions, at least half of the deficits were not accounted for by limitation of the overall cognitive score. Thus, motor and executive function disturbances in extremely prematurely born children are not due to compromised intellectual function but are more specifically associated with perceptual deficits. Children who were born extremely prematurely were observed to need more time for tasks compared to full-term classmates, and they were noted to have greater difficulty inhibiting associated movements and maintaining a secure postural basis during activities. As many as 7% of extremely prematurely born children were unable to balance on one leg for 20 seconds (compared to 1% of classmates). This study showed that extremely prematurely born children did not differ significantly from classmates on observed off-task motor behavior (17% vs. 25%), mirror movements (32% vs. 46%), and overflow movements (45% vs. 69%). Some investigators have postulated a continuum of reproductive casualty. There are also suggestions of incomplete cerebral dominance or lateralization, neurochemical abnormalities in the brain, and structural lesions in the parietal lobes.

DCD and Speech-Language Comorbidities

DCD frequently occurs concurrently with speech–language disorders and specific learning disorders, including reading disorders and dyslexia. A study showed that more than half of children with severe dyslexia exhibited motor coordination impairments severe enough to warrant intervention. Reading disorders, dyslexia, and speech and language disorders have been hypothesized to be associated with cerebellar involvement, as has DCD. One longitudinal study of children with language impairments revealed a continuing correlation between peg-moving performance and language performance. Another study found that fine motor skills, response speed, and upper limb speed were associated with language disturbances, whereas gross motor coordination and visual–motor control were not. Furthermore, evidence indicates that articulation problems are related to motor impairment, although the motor impairment is not necessarily dyspraxia.

DIAGNOSIS AND CLINICAL FEATURES

In 2012, The European Academy of Childhood Disability (EACD) published an interdisciplinary clinical guideline for the assessment, diagnosis, and treatment of children with DCD. These guidelines were subsequently revisited in 2019 to publish the international clinical practice recommendations on the definition, diagnosis, assessment, intervention, and psychosocial aspects of DCD. The essential feature of DCD is poor skill in daily motor activities for a given child's age and intellectual level. Impairments may be observed in gross motor activities, such as running, kicking, or catching a ball, as well as in fine motor coordination, resulting in poor ability to write letters neatly, to drink from a cup, or to use a knife and fork well. Children with DCD exhibit a heterogeneous pattern of impairments and thus function more competently in some motor task areas than others.

The current proposed criteria for diagnosis of DCD per the 2019 international clinical practice recommendations by EACD is similar to the DSM-5-TR diagnostic criteria for DCD (Table 41–2). Careful history-taking (preferably from parents, collateral sources other than parents, and the children themselves), problem-oriented clinical observation, and examination are essential to support the identification of the four diagnostic criteria. However, the recommended order of criteria application to arrive at the final accurate diagnosis of DCD involves first considering the age and medical history of the patient (criterion C), then ruling out other medical conditions (criterion D), analyzing the impact on activities and academic achievement (criterion B), and finally, quantification of motor impairment (criterion A). Due to the large variability of motor abilities in normal development, a formal diagnosis of DCD under the age of 5 years is only initiated by motor assessment in case of severe impairment and is usually confirmed by a second motor assessment at least 3 months apart.

Table 41–2.
DSM-5-TR Diagnostic Criteria for Developmental Coordination Disorder

A. The acquisition and execution of coordinated motor skills is substantially below that expected given the individual's chronological age and opportunity for skill learning and use. Difficulties are manifested as clumsiness (e.g., dropping or bumping into objects) as well as slowness and inaccuracy of performance of motor skills (e.g., catching an object, using scissors or cutlery, handwriting, riding a bike, or participating in sports).

B. The motor skills deficit in Criterion A significantly and persistently interferes with activities of daily living appropriate to chronological age (e.g., self-care and self-maintenance) and impacts academic/school productivity, prevocational and vocational activities, leisure, and play.

C. Onset of symptoms is in the early developmental period.

D. The motor skills deficits are not better explained by intellectual developmental disorder (intellectual disability) or visual impairment and are not attributable to a neurological condition affecting movement (e.g., cerebral palsy, muscular dystrophy, degenerative disorder).

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 85, 86 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Assessment Instruments for DCD

The Developmental Coordination Disorder Questionnaire. The DCD-Q is a 17-item parent questionnaire developed in Canada and validated to identify motor coordination problems in children aged 5–15 years by Wilson et al. (2000). For each item, parents rate the degree of coordination in their child compared to others of the same age on a 5-point scale. By adding the scores for each item, a total DCD-Q score can be calculated. In the study validating this instrument for young children, a score of less than or equal to 53 (15th percentile) was used as the criterion for DCD. The 17 items on the DCD-Q are:

1. Throw ball
2. Catch ball
3. Hit ball
4. Jump over
5. Run and stop
6. Plan activity
7. Writing speed
8. Writing legibly
9. Pencil pressure
10. Cutting
11. Team sports
12. Avoids sports
13. Rides bicycle
14. Learns skills
15. Clumsy
16. Awkward
17. Fatigues easily

The Movement Assessment Battery for Children. The MABC-second edition, was introduced by Barnett, Henderson, and Sugden in 2007 and provided an indication of gross and fine motor skills for children 4 years to 16 years of age. The MABC has two parts: a checklist and a performance test. On the MABC checklist, teachers rate the motor performance of children on 48 items reflecting everyday activities in the lives of children from 4 to 12 years. The checklist is divided into 4 sections, with 12 items in each section. On the performance test, there are four age-related item sets, each consisting of eight items that measure manual dexterity, ball skills, and static and dynamic balance. The checklist is a reliable instrument within sections and overall. Since the nature and severity of problems in DCD are very broad, an instrument meant to screen children with DCD must be able to measure a wide variety of motor skills.

Bruininks–Oseretsky Test of Motor Proficiency for Identifying Children with Motor Impairment. The Bruininks–Oseretsky Test of Motor Proficiency (BOTMP) assesses motor skills in children from 4½ to 14½ years old and provides a comprehensive index of motor proficiency as well as individual measures of gross and fine motor skills. In addition to the long form (LF), there is also a short form (SF), which consists of 14 items from the LF that have been validated. A recent study of the BOTMP in the assessment of preschoolers investigated the correlation between the LF and the SF of the instrument. Although the correlation between the two forms was high, short-form total scores were found to be significantly higher than long-form scores, identifying fewer children as having a motor impairment. Therefore, although the LF has been validated for young children, the SF is questionable in its ability to identify motor impairment in very young children.

Additional tests include the Frostig Movement Skills Test Battery, the Slosson Drawing Coordination Test, the Southern California Sensory Integration Test, and the Stott Test of Motor Impairment, which provide a comparative measure of a child’s abilities.

CLINICAL FEATURES OF DCD

Symptoms of DCD may be broken down into the following categories: dyspraxia, synkinesia, hypotonia, hypertonia, tremors, impersistence, and asymmetries. *Dyspraxia* describes a child's difficulty in exhibiting sequenced, coordinated motor movements when presented with an oral request. Movements identified as *synkinesia* are best described as unintentional muscle movements, also termed *motor overflow*. Mirror movements, such as finger twitching on the opposite hand when the child is asked to perform a finger opposition task or facial grimaces when the child is asked to make hand movements, are examples of synkinetic motor actions.

Two categories that describe muscle tone abnormalities are *hypotonia* and *hypertonia*. Hypotonia can be observed in all parts of the body and is characterized by a flaccid or sleepy quality in the child's facial expression and general muscle tone. Children with this condition may exhibit poor posture and obesity. Children with hypotonia often appear to be floppy and weak, but on actual muscle strength testing, they may have normal muscle strength. *Hypertonia* generally describes high levels of muscle tone in a child. For example, a child with hypertonic movement cannot finish drawing a line at a given point, instead overshooting the intended stopping point. Another example is a child who tends to throw a ball too hard and inaccurately at short distances.

The movement category of *tremors* is marked by irregular unsteadiness in motor movements observed in tasks requiring walking, drawing, or both. For example, legibly formed letters might reveal, on closer inspection, rapid tremors in the actual lines themselves. Unsteadiness might be exhibited in the child's gait as high-amplitude fluctuations in leg muscle movement.

The category of *impersistence* refers to the child's inability to sustain and maintain various body postures for reasonable periods of time. Asking a child to stick out their tongue is an example of a task used often to investigate impersistent motor movements. In this example, the clinician would keep track of how long the child could maintain this tongue position.

The final movement category, *asymmetries*, describes motor behaviors that affect only one side of the body. This category includes unusual flexions or muscle weakness, or both, on one side of the body or general muscle weakness in the limbs on one side.

These asymmetrical muscle movements can be observed by requiring a child to move laterally in a rapid manner (“like a basketball player on defense, not crossing your feet”). If the child executes the movement more fluidly on one side than the other, it might indicate an asymmetry. A child who is observed to have asymmetrical motor movements is likely to be referred for a magnetic resonance imaging (MRI) study to rule out an organic cause for this symptom. If the neurologic workup is negative, then this symptom may be considered part of a DCD.

Specific manifestations vary across individuals; thus, there is no typically clumsy child. In some instances, manifestations may be quite specific; one case report describes a child who could sew and do jigsaw puzzles without difficulty but could not write neatly.

The most common way to identify DCD is through the observation of delays in achieving motor milestones (e.g., sitting, standing, crawling, and walking). Gross motor coordination manifestations may be seen in difficulties with activities such as sitting, walking, running, balancing, throwing, and kicking. Fine motor coordination manifestations may be identified through difficulties in performing tasks, such as using utensils, tying laces, buttoning or zipping clothing, drawing, and, especially, printing or writing. During motor tasks, children with DCD may exhibit overflow movements in unsupported limbs, mirror movements, slight abnormalities of reflex, or poor ability to make movements smooth.

Some manifestations of DCD are age dependent. Young children show symptoms, such as delayed development, in early motor milestones and delayed acquisition of self-help skills, such as self-dressing (i.e., tying ties, fastening snaps, zipping zippers, buttoning, and unbuttoning) or self-feeding (handling a knife, fork, or spoon). Preschool-aged children may also have balance problems, awkward gait, and a history of frequently bumping into objects, falling down, and getting bruised. Problems with balance, hopping, jumping, catching and bouncing a ball, and tracing appear to be among the more common symptoms of school-aged children with DCD. A common sign of DCD among school-aged children is avoidance of sports activities because of poor performance or teasing by classmates.

In addition, children with DCD frequently have comorbid communication and learning difficulties. Among the more common of these difficulties are expressive language disorder, mixed receptive–expressive language disorder, phonologic disorder, reading disorder, disorder of written expression, and mathematics disorder.

Secondary complications for children with DCD often involve rejection from peers who are critical of their poor abilities in sports activities and other tasks. Over time, children who are chronically “picked last” for sports teams and who are unable to keep up physically with their peers may become demoralized or depressed. The DSM-5-TR diagnostic criteria for DCD are listed in Table 41–2. The ICD-10 and 11 diagnostic criteria for specific developmental disorder of motor function (ICD-10) and developmental motor coordination disorder (ICD-11) are presented in Table 41–3A and B.

Table 41–3A.
ICD-10 Diagnostic Criteria for Specific Developmental Disorder of Motor Function

A. The score on a standardized test of fine or gross motor coordination is at least two standard deviations below the level expected for the child’s chronological age.
B. The disturbance described in Criterion A significantly interferes with academic achievement or with activities of daily living.
C. There is no diagnosable neurological disorder.
D. Most commonly used exclusion clause. IQ is below 70 on an individually administered standardized test.

From World Health Organization. *The ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic Criteria for Research*. 10th rev. World Health Organization; 1993, with permission. https://apps.who.int/iris/handle/10665/37108.

Differential Diagnosis

Informal observation by parents, teachers, and clinicians of gross motor coordination (e.g., hopping, jumping, and standing on one foot), fine motor coordination (e.g., finger tapping and shoelace tying), and hand–eye coordination (e.g., catching a ball and copying letters) is the most common initial screen for identification of coordination difficulties. However, given the vast normal variation in motor skills, it is often not easy to make a diagnosis of DCD. When coordination difficulties interfere with academic achievement or activities of daily living, however, formal testing and evaluation of these skills are recommended (Table 41–4).

Ruling Out Neurologic Disorders. Children with developmental motor disorders progress in gross and fine motor milestones, although at a slower rate than peers. They do not regress. Thus, a child who previously walked, ran, and climbed at age-appropriate levels but who no longer does these activities well should be evaluated for neurologic disorders. A child whose balance has deteriorated needs further testing to determine the cause of the change. Similarly, a child whose fine motor development was at the age expectation but who stopped using a hand, switched apparent hand dominance, or lost previous abilities should be assessed for other medical conditions. Even a child whose initial motor landmarks were delayed should be carefully investigated if they lose previously established abilities.

Table 41–3B.
ICD-11 Essential (required) Features for Developmental Motor Coordination Disorder

- Significant delay in the acquisition of gross or fine motor skills and impairment in the execution of coordinated motor skills manifesting as clumsiness, slowness, or inaccuracy of motor performance.
- Coordinated motor skills are markedly below that expected given the individual's chronological age and level of intellectual functioning.
- Onset of coordinated motor skill difficulties occurs during the developmental period and is typically apparent from early childhood.
- Coordinated motor skills difficulties cause significant and persistent limitations in functioning (e.g. in activities of daily living, school work, vocation and leisure activities).
- Difficulties with coordinated motor skills are not solely attributable to a Disease of the Nervous System, Disease of the Musculoskeletal System or Connective Tissue, sensory impairment, and not better explained by a Disorder of Intellectual Development.

Reprinted from World Health Organization. ICD-11 for Mortality and Morbidity Statistics, 2023. https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/148247104.

Table 41–4.
Motor Signs and Symptoms Indicating Referral for Comprehensive Neurologic Evaluation

- Hypertonia
- Hypotonia
- Ataxia
- Tremor
- Chorea
- Weakness
- Asymmetry

Family history is important in consideration of neurodegenerative conditions such as spinocerebellar degeneration, Huntington disease, peripheral neuropathies, or other neuromuscular disorders. Further investigation should be pursued even if the positive family history is of onset in adulthood. Many dominantly inherited neurodegenerative conditions show “anticipation,” presenting in subsequent generations at earlier ages, even early childhood.

A physical examination, including a neurologic examination, should be performed. Skin stigmata, for example, may suggest a neurocutaneous syndrome (i.e., multiple café-au-lait spots, axillary or inguinal freckles suggesting neurofibromatosis type I (NF1); hypomelanotic macules, fibrous cephalic plaques, facial angiofibromas suggesting tuberous sclerosis). Excessive diffuse skin pigmentation, particularly in non–sun-exposed areas, is a clue to possible corticotropin excess, seen in boys in the early phases of adrenoleukodystrophy, at a time when the only other complaints may be loss of adept motor skills and behavior problems. Dysmorphic features, even subtle, may suggest an etiology. The presence of scoliosis, a dimple, bulge, or hair tuft over the spine, or limb asymmetries may suggest that the motor findings are due to a structural spinal cord abnormality. Examination of the ocular fundi is sometimes challenging in small children but is particularly important if loss of motor milestones is accompanied by other symptoms such as headache or vomiting.

Screening evaluation of cranial nerve function in young children is most easily assessed by observation of eye movements, facial movements, listening to the quality of speech, and watching the child swallow. For example, a child who cannot easily swallow water from an open cup or who coughs or gags when doing so may need to have a gag reflex formally tested. The extremities should be evaluated for muscle bulk, tone, and symmetry of length and girth. Asymmetric extremity size suggests either a structural brain or spinal cord lesion. In young children, strength is best assessed by observation of the child performing specific activities, such as rising from the floor, climbing onto a stool, reaching overhead for a toy, jumping, and hopping. Diffuse or symmetric muscle atrophy or weakness may suggest a neuromuscular disorder (i.e., muscular dystrophy, congenital myopathy, and neuropathy), malnutrition, or systemic illness. For example, boys with early manifestations of Duchenne muscular dystrophy may appear motor delayed and maladroit well before there are obvious complaints of weakness. Although testing of tendon reflexes may yield somewhat inconsistent results in young children, consistent asymmetries, the complete absence of reflexes, or frank hyperreflexia, clonus, or extensor plantar responses mandate further investigation. Gait should be observed in multiple ways: Watch the child walking, running, tandem walking, skipping, and walking on heels and toes. In addition to watching the gait pattern, the examiner should pay attention to posture and movement of the upper extremities as the child walks or runs. Abnormalities such as asymmetric arm swing or dystonic posturing of one or both hands are sometimes only appreciated with activity. Balance can be assessed both by watching the child perform ordinary activities, such as rising from the floor, walking, or turning, and by more structured tasks, such as standing on one foot. Frank ataxia, as opposed to immaturity and delay in acquiring skills, should mandate additional investigations. Many different genetic conditions present with truncal ataxia, with or without other deficits. The initial presentation may be relatively subtle. Fine motor abilities can be observed both directly (i.e., finger-to-thumb tap, thumb-to-sequential-finger-tap) and by watching the child during age-appropriate activities. For example, watch the child pick up a toy, color, write, use a handheld computer game or tablet-type computer, drink from a cup, or use a spoon. Look for asymmetries, dystonic posturing, or tremor. Look for tremor or dystonic posturing of the outstretched arms.

Poor motor coordination frequently coexists with learning disorders and disruptive behavior disorders, such as ADHD, particularly in children with high levels of impulsivity. A child may meet the criteria for DCD and ADHD. The tendency for children with DCD to be accident-prone, to fall down, to hurt themselves, and to break objects may, at times, be construed as destructiveness. In the school setting, a child’s motor difficulty with holding a pencil and writing legibly may result in great frustration for the child and the teacher. For children whose impaired writing ability and poor production are inadvertently interpreted as a lack of motivation, demoralization is common, and an evaluation for special educational services may be warranted. Children with DCD are at higher risk for poor social skills and are often victimized by bullies.

Jimmy was referred for evaluation to determine his need for adaptive physical education (PE), age 9 years of age. Jimmy was failing PE, was unable to run a mile, and he could not throw or catch the ball, and was generally slow motorically. His classmates laughed because he looked so “funny” while running, and he always seemed to be getting hit with the ball instead of catching it. Jimmy was always picked last for team sports games and was beginning to be bullied by his classmates for his poor physical skills. Jimmy already had an accommodation for more time on tests due to his previously diagnosed ADHD. He was being treated with Vyvanse, which was working well. A developmental history revealed that Jimmy was born prematurely at 31 weeks, and his motor milestones were all delayed, although speech milestones were right on time. Currently, Jimmy was unable to tie his shoes correctly, and he struggled to handwrite letters correctly, but his parents were hoping that he would outgrow those problems. Upon questioning Jimmy’s current motor skills, it became apparent that not only was he having coordination problems but that he was now being rejected by peers and had become quite socially isolated. At home, Jimmy’s parents had accepted the fact that he was just a clumsy boy and noted that Jimmy constantly spilled his drinks and was awkward when he used a fork. His food often fell off a fork or spoon before it reached his mouth, and he had great difficulty using a knife and a fork. A comprehensive assessment of fine and gross motor skills yielded the following results: Jimmy was able to hop, but he could not skip without briefly stopping after each step. He could stand with both feet together but could not stand on tiptoe. Jimmy held a ball bounced to him at chest level, and he could not catch a ball bounced to him on the ground from a distance of 15 ft. Jimmy’s agility and coordination were measured with the Bruininks–Oseretsky Test of Motor Proficiency (BOTMP), which revealed functioning levels commensurate with those of an average 7-year-old child.

Jimmy was referred to a neurologist for a comprehensive evaluation to rule out neurologic disorders. On neurologic examination, Jimmy was found to have no signs of a specific neurologic disorder. Neurologic examination was negative for conditions that could account for his poor motor skills. Based on the history and findings on the BOTMP, Jimmy was given a diagnosis of DCD. Jimmy's symptoms included gross and fine motor impairments, poor balance, and mild hypotonia. Based on an individualized educational plan (IEP) meeting with his school, it became evident that Jimmy required an adaptive PE program for DCD and had other specific learning disorders. In addition to his known ADHD, Jimmy was found to have a writing disability. A treatment plan was developed that included twice weekly with an occupational therapist who used perceptual-motor exercises to improve his fine motor skills, particularly targeting writing and the use of utensils. Jimmy was also enrolled in a treatment program using motor imagery training to improve clumsiness and was also referred to a social skills group to help him deal more effectively with his peers.

Jimmy was relieved to be receiving help, especially for his handwriting and PE class, which he had been close to failing. Over a period of 4 months of treatment, Jimmy showed some improvement in the legibility of his handwriting, although he remained a slow writer. He was responsive to the stimulant medication, and his academic achievement improved as his attention span increased. Jimmy was receiving more praise from his teachers than ever before, and his classmates were teasing him less. As he began to feel better about himself, he began to try to play sports informally with his peers, although not competitively. He was given a modified PE program in school, and he was not required to play on teams, but he practiced throwing and catching a ball and was allowed to walk his mile at a slower time. Jimmy continued to exhibit some impairment in fine motor skills over the year, and he was cooperative with the therapeutic interventions, showed some continual improvement, and demonstrated improved peer relationships.

COURSE AND PROGNOSIS

Children with DCD represent a heterogeneous group with a variety of impairments in gross and fine motor behaviors needed for everyday living as well as in static balance. The profile of specific deficits for a given child is individual and has been shown in recent studies to be modified positively by specific motor skills interventions. The outcome of motor coordination disorder is influenced by the severity of the impairment, the age of initiation and duration of therapeutic interventions, and the presence of comorbid disorders. A recent study of preschool children with DCD at 7 to 8 years of age revealed that those children who had significant impairment at 5 years of age continued to show impairment 1.5 years later. Longer-term studies have shown that motor coordination improves over the long term in at least 50% of children when they reach adolescence. A 2-year follow-up study of children of primary school age who had failed a motor test battery revealed motor performance that, although improved, was still inferior to that of controls. The children with motor impairments also showed poorer educational performance than controls.

A 10-year follow-up of clumsy children and their age-matched, sex-matched peers was done in children over a period of 8 to 9 years. When first identified, the clumsy children were approximately 6 years of age; at follow-up, they ranged from 15 to 17 years of age. Follow-up testing indicated that the clumsy children were still significantly below the controls in most motor skills (e.g., dexterity, ball skills, balance, cutting with scissors, catching, and walking backward) and in global ratings by physical education teachers. The clumsy adolescents showed higher rates of dysgraphesthesia, dysdiadochokinesia, and motor slowness. In addition, the educational achievements of the clumsy adolescents were significantly poorer, and they had significantly lower perceived social competence than their nonclumsy peers. This finding substantiates reports of continuing secondary problems, including peer problems, low self-esteem, and academic struggles in children and adolescents with significant coordination difficulties.

A long-term follow-up of children with DCD and those with ADHD and DCD at 22 years of age suggests that there are psychosocial ramifications into adulthood. Psychosocial adjustment difficulties present in young adults with only DCD and those who were comorbid for ADHD included unemployment, use of substances, continued difficulties with reading, and legal problems.

DCD tends to improve in most individuals over time, and those with more severe impairment at a younger age, as well as those comorbid for additional psychiatric disorders, have the poorest prognoses.

TREATMENT

Interventions for developmental motor coordination can be divided into three major categories: (1) *Task-oriented* approaches such as Neuromotor Task Training (NTT) and Cognitive Orientation to Occupational Performance (CO-OP); (2) *Process-oriented* approaches such as Sensory Integration (SI) and kinesthetic training; and (3) *traditional physical and occupational therapy* such as perceptual-motor training, and a therapist-guided parent or teacher intervention. A meta-analysis of interventions to treat DCD in children by Smits-Engelsman and colleagues (2013) found that the strongest evidence of effective treatment for DCD was for task-oriented approaches and traditional motor training therapies, with process-oriented therapy yielding only weak efficacy.

Task-Oriented Approaches

Task-oriented therapies focus on learning particular motor skills with a focus on the aspects of the movements that pose the greatest difficulties for the child. NTT is a hybrid of practicing specific problematic specific motor skills with the addition of cognitive behavioral strategies to increase motor planning, reduce anxiety, and increase the incentive for trying. CO-OP is another form of task-oriented therapy which directs a child to form a mental model of how a particular set of movements will be accomplished and then review with the therapist how well the mental image fits the actual motor task trials. CO-OP consists predominantly of cognitive behavioral problem-solving approaches and utilizes cognitive behavioral strategies proposed by Donald Meichenbaum.

Computer gaming with virtual physical motor tasks has been found to benefit children with DCD, often without explicit instructions or learning. A study of children with DCD playing a computer game designed in which children need to catch a ball virtually on the screen showed that they used visual cues to help them virtually catch the ball without even realizing it. These children were able to improve their computer scores implicitly by practicing the game and making use of a perceptual cue without having to be specifically instructed on how to improve. This has implications for treatment in that in addition to providing direct instruction to children with motor coordination problems, improvement can also occur with practice through environmental cues that are not necessarily identified.

A newer task-oriented strategy is Nintendo Wii Fit Training (Wii training). The Nintendo Wii is a gaming system that combines motion sensor technology in handheld controls, and a balance board provides both aspects of biofeedback and virtual reality to the player. In this way, an interactive video game can serve as a therapeutic tool for motor skills improvement.

A study of 56 children between 6 years and 10 years with DCD compared NTT (n = 37) to Wii training (n = 19) by Ferguson and colleagues. Children in the NTT group were provided two sessions per week (45 to 60 minutes) for 9 weeks in which they played games such as soccer, tag, and netball under the guidance of therapists who facilitated the learning of motor skills, provided positive feedback, and manipulated the environment when necessary. Children in the Wii group participated in 30 minutes of gaming three times a week for 6 weeks under the supervision and guidance of therapists who encouraged, praised, and motivated the children. Games included sports such as cycling, soccer, skateboarding, and skiing. Games required that the children shift their weight from side to side and jump over virtual objects, and some games required twisting of arm movements. The results found mean motor performance scores in both groups improved during the course of the study. The NTT group showed greater improvement in overall motor performance, functional strength, and cardiorespiratory fitness, while the Wii group showed significant improvement in anaerobic performance. Thus, this study suggests that both NTT and Wii training can provide valuable therapeutic gains.

Overall recommendations for first-line interventions for children with DCD include task-oriented therapies of various types. Younger children can benefit from NTT for problems such as handwriting, that is, writing letters, even before they learn to read. Older children and those with verbal skills may be able to benefit from CO-OP approaches that incorporate more verbal planning. Nintendo Wii and other gaming interventions can be beneficial for children with DCD of all ages.

Process-Oriented Approaches

There has been only a weak suggestion from studies that process-oriented approaches are effective therapies for DCD. However, they are often used in conjunction with other strategies. Process-oriented interventions aim to target neural structures, such as the cerebellum, for improving balance, or sensory processes, such as proprioception, in children with DCD. Interventions aim to enhance and remediate sensory input processing, for example, sensory integration therapy (SIT). SIT approach was introduced in the work of Ayres in the late 1970s, who focused remediation on enhancing sensory information processing in order to achieve more accurate motor execution. Sensory integration programs are often delivered as a related service to children with DCD under the direction of occupational therapists. These programs prescribe specific physical therapies that attempt to modify the motor and sensory functioning of these children. These therapies are based on the notion that sensory integration disorders interfere with body movements and body awareness. Activities are designed to target the tactile, vestibular, and proprioceptive systems. A sensory integration program that addresses disorders affecting the proprioceptive system (bumping into walls and inability to button, skip, or write) might involve using scooter boards to improve balance and body awareness. Multisensory modalities, including visual, auditory, and tactile materials, are utilized in this form of treatment. Motor imagery training for clumsiness in children has received attention recently based on results from controlled investigations of this modality compared to perceptual-motor interventions. One study used a successful motor imagery intervention through CD-ROM, which contained six main training components for children with coordination difficulties. The components included visual imagery exercises involving predictive timing, relaxation, and mental preparation, visual modeling of fundamental motor skills, mental rehearsal of skills from an external perspective, mental rehearsal of skills from an internal perspective, and overt practice. This model of treatment is based on the notion that deficits in a child's ability to internally represent the movement tasks contribute to poor motor coordination. The ability for visual imagery of objects is not typically impaired in children with DCD. The previously mentioned training exercises were designed to improve the timing and accuracy of a given child's ability to visualize motor tasks, as well as to reduce muscle tension while imagining completing the given motor task. Finally, the children were trained to mentally rehearse the task before actually practicing it.

Traditional Physical and Occupational Therapy

Traditional physical and occupational therapies target actual activities of daily function and seek to improve the execution of these motor behaviors. A variety of cognitive-motor approaches have been utilized to improve the planning and execution of motor skills. One cognitive-motor approach to improving functional skills by Sugden and Hendersen has been updated and labeled ecologic intervention, which places emphasis on the actual control of movement and is done in a real-life setting.

Traditional perceptual-motor training consists of a combination of gross motor, fine motor, and perceptual-motor activities related to a child's actual daily motor tasks, such as hopping, skipping, jumping, climbing and moving through hoops, climbing ladders, and jumping on a trampoline, to practice competency. Specific tasks addressing fine motor skills, for example, handwriting training, includes activities to improve the grip of a pencil, wrist rotation, and posture. The results of the previously mentioned study were extremely notable in that the mental imagery treatment was as efficacious as traditional perceptual-motor training exercises in improving a given child's coordination.

An example of functional skill intervention would be those that address impairment in handwriting. It is estimated that between 10% and 30% of children exhibit persistent impairment in handwriting, a skill that remains a critical means of communication even in the face of the pervasive use of computers. Handwriting impairment is associated with academic skills disorders. Handwriting treatment approaches may include perceptual-motor, visual–motor, motor control, individualized interventions or exercises, and supplementary handwriting instruction. Few controlled studies have examined which techniques are most efficacious, and a survey of 50 pediatric occupational therapists found that 90% of them advocated an eclectic approach to interventions for handwriting improvement. A study of school-based occupational therapy intervention for school-aged children with poor handwriting legibility showed efficacy in improved legibility of letters. Children with handwriting difficulties in this study received direct traditional perceptual instruction on writing letters for approximately 30 minutes per week for approximately 7 months. This consisted of a variety of handwriting interventions, such as using vibration or resistance to writing to increase appropriate grip and writing on a chalkboard or a vertical surface to increase arm strength and stability while writing. At the end of this period, children made significant gains in handwriting legibility compared to a control group that did not receive the intervention. Handwriting speed in the children who received the treatment increased from 32 to 37 letters per minute. Although this was a trend toward speed, this did not represent a statistically significant increase in speed. Some students in the treatment group wrote their letters with a greater degree of deliberate care, and although this increased legibility, it may have actually diminished their speed. Overall, the increase in legibility applied to letters more than numbers but did not significantly increase writing speed.

FURTHER READINGS

Adams IL, Lust JM, Wilson PH, Steenbergen B. Compromised motor control in children with DCD: a deficit in the internal model? A systematic review. *Neurosci Biobehav Rev*. 2014;47:225–244.

Blank R, Barnett AL, Cairney J, et al. International clinical practice recommendations on the definition, diagnosis, assessment, intervention, and psychosocial aspects of developmental coordination disorder (long version). *Dev Med Child Neurol*. 2019;61:242–285.

*Blank R, Smits-Engelsman B, Polatajko H, Wilson P; European Academy for Childhood Disability. European Academy for Childhood Disability: recommendations on the definition, diagnosis and intervention of developmental coordination disorder (long version). *Dev Med Child Neurol*. 2012;54(1):54–93.

Caçola P. Physical and mental health of children with developmental coordination disorder. *Front Public Health*. 2016;4:224.

Camden C, Wilson B, Kirby A, Sugden D, Missiuna C. Best practice principles for management of children with developmental coordination disorder (DCD): results of a scoping review. *Child Care Health Dev*. 2015;41(1):147–159.

Dewey D, Kaplan BJ, Crawford SG, Wilson BN. Developmental coordination disorder: associated problems in attention, learning, and psychosocial adjustment. *Hum Mov Sci*. 2002;21(5-6):905–918.

Feder KP, Majnemer A. Handwriting development, competency and intervention. *Dev Med Child Neurol*. 2007;49(4):312–317.

Ferguson GD, Aertssen WFM, Rameckers EAA, Jelsma J, Smits-Engelsman BCM. Physical fitness in children with developmental coordination disorder: measurement matters. *Res Dev Disabil*. 2014;35(5):1087–1097.

*Ferguson GD, Jelsma D, Jelsma J, Smits-Engelsman BCM. The efficacy of two task-oriented interventions for children with Developmental Coordination Disorder: Neuromotor Task Training and Nintendo Wii Fit Training. *Res Dev Disabil*. 2013;34(9):2449–2461.

Fliers E, Rommelse N, Vermeulen SHHM, et al. Motor coordination problems in children and adolescents with ADHD rated by parents and teachers: effects of age and gender. *J Neural Transm (Vienna)*. 2008;115(2):211–220.

Gillberg C, Kadesjo B. Why bother about clumsiness? The implications of having developmental coordination disorder (DCD). *Neural Plast*. 2003;10(1–2):59–68.

Hay JA, Hawes R, Faught BE. Evaluation of a screening instrument for developmental coordination disorder. *J Adolesc Health*. 2004;34(4):308–313.

Miller LT, Polatajko HJ, Missiuna C, Mandich AD, Mcnab JJ. A pilot trial of a cognitive treatment for children with developmental coordination disorder. *Hum Mov Sci*. 2001;20(1–2):183–210.

Noten M, Wilson P, Ruddock S, Steenbergen B. Mild impairments of motor imagery skills in children with DCD. *Res Dev Disabil*. 2014;35(5):1152–1159.

Novak I. Evidence to practice commentary. New evidence in developmental coordination disorder (DCD). *Phys Occup Ther Pediatr*. 2013;33(2):170–173.

Pless M, Carlsson M, Sundelin C, Persson K. Preschool children with developmental coordination disorder: a short-term follow-up of motor status at seven to eight years of age. *Acta Pediatr*. 2002;91:521–528.

Prunty MM, Barnett AL, Wilmul K, Plumb MS. The impact of handwriting difficulties on compositional quality in children with developmental coordination disorder. *Br J Occup Ther*. 2016;79:591–597.

*Rasmussen P, Gillberg C. Natural outcome of ADHD with developmental coordination disorder at age 22 years: a controlled, longitudinal, community-based study. *J Am Acad Child Adolesc Psychiatry*. 2000;39:1424–1431.

Rivard L, Camden C, Pollock N, Missiuna C. Knowledge to practice in developmental coordination disorder: utility of an evidence-based online module for physical therapists. *Phys Occup Ther Pediatr*. 2015;35(2):178–194.

Schoemaker MM, Flapper B, Verheij NP, Wilson BN, Reinders-Messelink HA, de Kloet A. Evaluation of the developmental coordination disorder questionnaire as a screening instrument. *Dev Med Child Neurol*. 2006;48(8):668–673.

*Smits-Engelsman BCM, Blank R, Van Der Kaay AC, et al. Efficacy of interventions to improve motor performance in children with developmental coordination disorder: a combined systematic review and meta-analysis. *Dev Med Child Neurol*. 2013;55(3):229–237.

Sugden DA. Current approaches to intervention in children with developmental coordination disorder. *Dev Med Child Neurol*. 2007;49:467–471.

*Sugden DA, Chambers ME. Intervention in children with developmental coordination disorder: the role of parents and teachers. *Br J Educ Psychol*. 2003;73:545–561.

Venetsanou F, Kambas A, Aggeloussis N, Serbezis V, Taxildaris K. Use of the Bruininks-Oseretsky Test of Motor Proficiency for identifying children with motor impairment. *Dev Med Child Neurol*. 2007;49(11):846–848.

Wade MG, Kazeck M. Developmental coordination disorder and its cause: the road less travelled. *Hum Mov Sci*. 2018;57:489–500.

Williams J, Kashuk SR, Wilson PH, Thorpe G, Egan GF. White matter alterations in adults with probable developmental coordination disorder: an MRI diffusion tensor imaging study. *Neuroreport*. 2017;28(2):87–92.

*Wilson PH, Ruddock S, Smits-Engelsman B, Polatajko H, Blank R. Understanding performance deficits in developmental coordination disorder: a meta-analysis of recent research. *Dev Med Child Neurol*. 2013;55:217–228.

Wilson PH, Thomas PR, Maruff P. Motor imagery training ameliorates motor clumsiness in children. *J Child Neurol*. 2002;17(7):491–498.

Woodward S, Swinth Y. Multisensory approach to handwriting remediation: perceptions of school-based occupational therapists. *Am J Occup Ther*. 2002;56(3):305–312.

*Zwicker JG, Missiuna C, Harris SR, Boyd LA. Developmental coordination disorder: a review and update. *Eur J Paediatr Neurol*. 2012;16(6):573–581.

42

Communication

42.1 Language Disorder

Tuana Kant, M.Sc., Emiko Koyama, Ph.D., Mahliqa Ashraf, H.B.Sc., and James L. Kennedy, M.D., F.R.C.P.C., F.R.S.C.

Communication disorders are among the most common problems exhibited by young children. Nonetheless, communication disorders often go unrecognized, particularly when they are comorbid with other emotional or behavioral problems. Thus, child psychiatrists, psychologists, and other clinicians need to be alert to the potentially damaging impact of communication problems on a child's overall well-being. Proficiency in verbal communication is critical to success in virtually all human activities and is therefore a highly valued skill-set in society and evolution. Individuals suffering from communication disorders may be unable to competently participate in everyday interpersonal, learning, and occupational situations. Limited participation in such key life activities may further restrict growth in communication and related domains.

To develop effective communication, children must master (a) language, the conventional code used to express ideas about the world, and (b) speech, the complex and rapid motor movements that translate such ideas into spoken words. Children need to also learn how to use language and speech appropriately in a given social context. The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) and its text revision (DSM-5-TR) recognize five diagnostic categories for communication disorders. One category concerns a disorder in language (language disorder), two categories relate to disorders in speech (speech sound disorder and childhood-onset fluency disorder), one category is associated with language use (social [pragmatic] communication disorder), and the final category covers communication disorder not otherwise specified.

DEFINITION AND COMPARATIVE NOSOLOGY

Individuals with language disorders show persistent impairments in the expression and reception (production and comprehension) of language. These impairments will manifest across different modalities, including speaking, writing, and sign language. The expressive difficulties may include limitations in vocabulary range, sentence complexity, and coherence in discourse. These expressive problems, however, are of less overall concern than the receptive difficulties. Children with receptive problems may manifest deficits in basic auditory processing skills, such as auditory discrimination, detection of rapid sound changes, association of sounds and symbols, and memory for sound sequences. Additionally, they may be slow to learn definitions, especially those that are uncommon or abstract. They may not understand the questions or comments of others and may misinterpret complex directions and explanations. They may have difficulty following conversations or learning from lectures. Because of comprehension failures, individuals with language disorders may give illogic or inappropriate responses. Furthermore, their comprehension problems may go unrecognized and instead may be misattributed to inattention, stubbornness, or other psychological or behavioral problems. Because of these problems, individuals with a language disorder would often have limitations in academic achievement, occupational success, and other social participation.

The definition of language disorder in DSM-5-TR combines elements of both the expressive and mixed receptive-expressive language disorders of the revised fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR) and the 11th edition of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-11). Both the DSM-IV-TR and the ICD-11 conceptualized the expressive and receptive subtypes as being mutually exclusive language disorders characterized by significant language problems, despite normal levels of nonverbal intelligence. DSM-IV-TR, however, differentiated between the developmental and acquired subtypes, whereas the ICD-11 only recognizes the developmental type, and acquired language problems such as acquired aphasia with epilepsy are listed under separate categories.

EPIDEMIOLOGY

The prevalence estimates of language disorder differ from roughly 3% to 15%. Receptive problems are thought to be somewhat less common than expressive problems. The variation in prevalence estimates is most likely due to differences in language measures used to define the disorder, the age of the children surveyed, and the nature of the sample tested (community or clinic referred).

Language disorders are more common in boys than in girls. Various studies have yielded sex ratios as high as 5:1 and as low as roughly 2:1.

Familial aggregation is characteristic of language disorders. A variety of methods (aggregation, pedigree, twin, adoption, and linkage studies) have provided clear evidence of genetic influences on language disorder. Complementary work also points to environmental input as a key factor that likely contributes to familial aggregation.

Approximately 10% to 15% of children younger than 3 years of age show delayed onset and slow initial progress in language development. Because of the large range of normal variability in the language production skills of young children, it is not yet clear whether an initial delay consistently warrants a diagnosis of language disorder. The majority (50% to 80%) of these children eventually acquire language skills within the normal range, but some show persistent language difficulties into adulthood. Children who continue to exhibit language problems after age 4 are more likely to experience long-term language deficits which may persist into adulthood, although the specific language profiles may change over time.

ETIOLOGY

There are no known causes for language disorder. Various factors have been shown to correlate with language disorder, but none has yet been established as causal. Language disorder most likely has multiple determinants, including various biologic and environmental factors. Different factors or combinations may be implicated in some cases and not others, as suggested by the heterogeneity in the population. At least seven general types of possible causal factors have been investigated: (1) brain abnormalities, (2) genetic factors, (3) environmental influences, (4) neurodevelopmental immaturity, (5) general cognitive factors, (6) specific linguistic factors, and (7) auditory processes. Each of the proposed factors enjoys some empirical support, but none has gained widespread acceptance as a plausible account for all cases of language disorder. Moreover, the factors are not mutually exclusive and may interact in complex ways.

Structural or functional brain abnormalities can now be studied in a sophisticated fashion with advanced neuroimaging techniques. Such abnormalities have long been considered possible causation factors in language disorder, because (1) language characteristics of language disorder resemble those of acquired language disorders that result from known neurologic insults, and (2) children with language disorders show elevated rates of soft neurologic signs (e.g., mixed cerebral dominance, attention problems, or fine motor difficulties). Neuroimaging studies of children with language disorders have demonstrated structural and functional abnormalities, often in brain regions known to be crucial for language processing (e.g., left hemisphere or certain subcortical structures). Nonetheless, many questions remain. Both the language disorder phenotype and the techniques used to examine the brain abnormalities differ greatly among the studies, making it difficult to draw conclusions about how underlying brain abnormalities contribute to the phenotypic language disorder. Furthermore, because of the important role experience holds in brain development, it is not clear whether the observed structural abnormalities are a cause or consequence of a language disorder.

Genetic factors have now been clearly implicated in language disorder, but it is far less clear just what these factors are and how they operate pathologically. Loci on chromosomes 13, 16, and 19 have been tentatively linked to differing aspects of language disorder in several studies. Moreover, 22q11 deletion syndrome and sex chromosomal aneuploidy have also been associated with language disorders. Although these regions are unlikely to be the sole determinants of genetic influences on language, their investigation may yield important insights. For example, some recent linkage studies have suggested that chromosome 16q may be associated with phonologic short-term memory and chromosome 19q with expressive language. Work also continues to refine or broaden the language disorder phenotype to accurately identify possible subtypes and modes of transmission. An interesting possibility suggested by several lines of evidence is that autism and language disorder may share some relevant genetic characteristics. Genetic factors may also be important in understanding the origin of the brain abnormalities noted earlier. However, genetic assessments are not yet part of the initial diagnostic techniques for language disorders. With advancements in the field of genetics, further robustical research is needed to implement genetic etiologic analyses in clinical practice during the tests for children presenting with symptoms of language disorders.

Various environmental factors may also be critical, at least in some cases of language disorder. Children from families of low socioeconomic status (SES) are at elevated risk for language disorder. Low SES, in turn, correlates with other known risk factors, such as single-parent family status, large family size, lower parental education, pregnancy and birth complications, lack of appropriate learning materials in the home, abuse and neglect. Children from low SES families are also at risk for malnutrition, other health problems, increased exposure to known toxins (e.g., lead), and limited educational and social opportunities. The amount and quality of linguistic input available to children have also been shown to vary dramatically across families, with some children hearing as few as 700 utterances a day and others as many as 11,000! Furthermore, and not surprisingly, these variations in input were strongly related to language development as measured by standardized tests. In adverse conditions, supportive extended family and neighbors may act as protective factors assisting the children’s language development. Further work is needed to clarify whether or how various environmental and input factors may interact with one another and, perhaps, with biologic factors to cause language disorder.

Neurodevelopmental immaturity may also underlie language disorder. Proponents argue that language development of children with language disorder generally follows a normal sequence and course but at a slower rate; that is, it is delayed rather than deviant. Delayed maturation of neural structures or slow myelination of neural pathways might also account for associated characteristics that are common in children with language disorder, such as attention and motor problems and other behavioral difficulties. Furthermore, such neurodevelopmental immaturities may occur more often in boys than in girls. A related view is that children with language disorder simply have language skills that fall at the lower end of the normal distribution. When basic yet essential skill sets that allow humans to mingle or operate independently are compromised due to language disorders, the evolutionary advantages humans possess to communicate are decreased, as such, children presenting with a diagnosis would be recognized as disadvantaged.

A variety of general cognitive factors have also been proposed to account for language disorder. Speed of cognitive processing is one such factor. Children with language disorder typically perform slower than their peers without the disorder on a variety of linguistic (e.g., naming or grammatical judgments) and nonlinguistic (e.g., fine motor or visual-spatial) tasks. This slower processing speed might somehow interfere with language learning. Some studies, however, failed to show any significant correlations between processing speed and language achievement. Another potential factor is cognitive flexibility, where children with language disorder may show reduced flexibility in adapting thoughts and behaviors to novel environments. Children with language disorder also show impairments in feedback processing, thus may have reduced feedback-based learning, affecting their overall learning capabilities compared to their peers. Others propose that phonologic working memory, the immediate maintenance and manipulation of verbal information, is a key cognitive factor in language disorder. Children with language disorder show deficits in phonologic working memory, which have been linked to deficits in vocabulary learning and reading proficiency. However, strong causal inferences about the role of phonologic working memory are not yet possible based on the available correlational evidence. Overall, differences in cognitive functions are found in preschoolers with language disorder when compared to preschoolers without, where verbal functions are more varying in children with language disorder.

In contrast to the neurodevelopmental immaturities and general cognitive delay perspectives, that imply that children with language disorder experience an overall delay but nevertheless follow a typical course of language development, some researchers have argued that children with language disorders may show specific lingering deficits in certain aspects of language that a young child with otherwise matched language ability may not show. For example, children with language disorders have particular difficulty in learning certain grammatical features (e.g., verb tense markers or pronouns in English) to the extent that their use of these features is poorer than that of younger children at the same milestone of language development. Much-needed cross-linguistic work in languages with varied grammars is helping theorists hone in on the commonalities and differences in the grammatical features that pose particular problems for children with language disorders who are learning different languages. At this point, however, these linguistic theories are primarily descriptive rather than explanatory.

Auditory problems have also been investigated as possible causes of language disorders. One concern is otitis media, an inflammation of the middle ear that results from ear infections or allergies. Otitis media is often accompanied by mild hearing loss, which is usually temporary if the underlying infection is treated. Repeated episodes of otitis media and associated hearing loss are common in young children during the earliest stages of language development. Although otitis media alone is not sufficient to cause language disorder, it may contribute in combination with other risk factors. A second concern centers on the auditory perception of sounds. Many (but not all) children with language disorders have difficulty perceiving small sound differences and rapid auditory transitions, which are aspects of the acoustic signal thought to be important in understanding speech. It has proven difficult, however, to develop convincing theoretical and empirical links between auditory perceptual skills and the various language difficulties shown by children with language disorder.

DIAGNOSIS AND CLINICAL FEATURES

The hallmark of language disorder is impairment in receptive and expressive language development. The rate of language growth is slower in children with language disorder than children with typical language development. The receptive and expressive language characteristics of affected children may be largely similar to those of younger children who are acquiring language in a typical fashion. Certain language features, such as the grammatical marking of tense, aspect, number, gender, or case may be especially difficult for children with language disorders to acquire. They may have difficulty learning the correct use of auxiliary or modal verbs (e.g., is, were, could), plurals (e.g., cats, geese), verb tense endings (e.g., -ing, -s, -ed), and pronouns (e.g., he, she, they). Children with language disorder may also demonstrate spelling difficulties where they may have decreased sensitivity to silent letters and make errors or omit silent letters while they are spelling.

Comprehension often varies with the complexity of the language. For example, words for familiar, concrete objects, actions, or relations may be comprehended (e.g., *wagon*, *apple*, *walk*, *give*, *in*, *over*), but those with less familiar, abstract meanings may be unknown (e.g., *neuron*, *freedom*, *receive*, *pretend*, *between*, *when*). Similarly, sentences with simple grammatical forms may be understood (e.g., “The boy is smiling.” “The phone rang.”), whereas those with more complex structures may be misinterpreted (e.g., “The car that rammed the bus was towed to the garage.” “Before she finalized the arrangements, she consulted with her lawyer.”). Complex or lengthy instructions, questions, and explanations may not be comprehended. When children with language disorders misconstrue the language and intentions of others, they may inadvertently make inappropriate responses or comments during conversations. Their comprehension failures may also contribute to confusion or inaccuracy in telling stories, relating events, and giving explanations, problems that may be further exacerbated by their expressive language limitations. When their comprehension difficulties are not recognized, children with language disorders may be mistakenly thought to be inattentive, lazy, or uncooperative.

The specific manifestations of language disorder vary with age. Some children with language disorder eventually acquire language skills within the normal range, but the overall prognosis is poorer for children with severe receptive language problems than for children with just expressive language difficulties. The expressive and receptive features adversely affect personal interactions and academic achievement because the spoken language is poorly expressed and understood.

Amy was a sociable, active 5-year-old who was diagnosed with a language disorder. She often played with Lisa, her kindergarten classmate. During pretend play one day, each girl told the story of Little Red Riding Hood to her doll. Lisa's story began: "Little Red Riding Hood was taking a basket of food to her grandmother who was sick. A bad wolf stopped Red Riding Hood in the forest. He tried to get the basket away from her, but she wouldn't give it to him."

By contrast, Amy's story illustrated her marked difficulties in verbal expression: "Riding Hood going to grandma house. Her taking food. Bad wolf in a bed. Riding Hood say, what big ears, grandma? Hear you, dear. What big eyes, grandma? See you, dear. What big mouth, grandma?

Eat you all up!"

Amy's story contained many features characteristic of children with language disorders at her age, including short, incomplete sentences, simple sentence structures, omission of grammatical function words (e.g., is and the), and inflectional endings (e.g., possessive -'s and present tense verb -s), problems in question formation, and incorrect use of pronouns (e.g., her for she).

Nevertheless, Amy performed as well as Lisa did in understanding the details and plot of the Riding Hood tale, as long as she was tested with methods that did not involve verbal responding. Amy also demonstrated adequate comprehension skills in her kindergarten classroom, where she readily followed the teacher's complex, multistep verbal instructions (e.g., "After you write your name in the top left corner of your paper, get your crayons and scissors, put your library books under your chair, and line up at the back of the room.").

Julio was a quiet, sullen 8-year-old whose language problems were no longer obvious in casual, social conversations. His speech now rarely contained the incomplete sentences and grammatical errors that were so evident when he was younger. His expressive problems, however, still surfaced in tasks involving elaborate or abstract uses of language, like those required in much of his third-grade academic work. An example was Julio's explanation of the outcome of a recent science experiment: "The teacher had stuff in some jars. He poured it, and it got pink. The other thing made it white." Although each sentence was grammatical, his explanation as a whole was difficult to follow because key ideas and details were omitted or poorly explained. Julio also showed problems in word finding, that is, in using specific words for the concepts and actions, he was describing. Instead, he relied on vague and nonspecific terms, such as thing, stuff, and got.

In early elementary grades, Julio had kept pace with his classmates in reading, writing, and other academic skills. By third grade, however, the increasing demands for written work began to negatively affect his overall academic standing. His written work was characterized by problems similar to those noted in his oral expression, such as poor organization and lack of specificity. Classmates also began to tease him about his difficulties, and he reacted quite aggressively, sometimes to the point of fighting. Nonetheless, Julio continued to show relatively good comprehension of spoken language, including classroom lectures concerning abstract concepts. He also comprehended sentences that were grammatically and conceptually complex (e.g., "The car the truck hit had hubcaps that were stolen. Had it been possible, she would have notified us by mail or by phone.").

Susan was a pleasant 2-year-old who did not yet use any spoken words. She made her needs known with vocalizations and simple gestures (e.g., showing or pointing) like those typically used by younger children. She seemed to understand the names of a few familiar people and objects (e.g., *mommy*, *daddy*, *cat*, *bottle*, and *cookie*). However, compared to other children her age, she had a small comprehension vocabulary and showed limited understanding of simple verbal directions (e.g., “Get your doll.” “Close your eyes.”). Nonetheless, her hearing was normal, and her motor and play skills were developing as expected for her age. She showed interest in her environment and in the activities of the other children at her daycare.

Min was a shy, reserved 5-year-old who grew up in a bilingual home. Min's parents and older siblings spoke English and Cantonese proficiently. Her grandparents, who lived in the same house, spoke only Cantonese. Min began to understand and speak both languages much later than her older siblings had. Throughout her preschool years, Min continued to develop slowly in comprehension and production. At the start of kindergarten, Min understood fewer English words for objects, actions, and relations than her classmates did. She showed difficulties in following classroom instructions, particularly those that involved words for concepts of time (e.g., *tomorrow*, *before*, or *day*) and space (e.g., *behind*, *next to*, or *under*). It was also hard for Min to match one of several pictures to a syntactically complex sentence that she had heard (e.g., "It was not the train she was waiting for." "Because he had already completed his work, he was not kept after school."). Min occasionally tried to speak with other children. However, these conversations usually broke down because she misinterpreted what others said or could not express her own thoughts clearly. Consequently, her classmates generally ignored her, preferring instead to play with more verbally competent peers. Min's infrequent interactions further limited her opportunities to learn and practice her already weak language skills. Min also showed limited receptive and expressive skills in Cantonese, as revealed by an assessment conducted with the assistance of a Cantonese interpreter. Nonetheless, her nonverbal cognitive and motor skills were within the normal range for her age. She showed no difficulties in solving spatial, numerical, conceptual, or analogic problems, provided they were presented nonverbally.

Fred received a diagnosis of language disorder when he was a preschooler. By 8 years of age, he had also received the comorbid diagnoses of specific learning disorder and attention-deficit/hyperactivity disorder (ADHD). This combination of language, learning, and attention problems made it virtually impossible for Fred to succeed in school. His comprehension and attention difficulties limited his ability to understand and learn important information from classroom instructions, discussions, and lectures. Because the unlearned information was critical to the understanding of future academic lessons, Fred fell further behind his classmates. He was also disadvantaged by the fact that he could read only a few familiar words. This meant that he could not learn other academic information that his classmates acquired by reading textbooks, library books, newspapers, and other written materials. His poor reading ability also limited his opportunities to learn the new vocabulary, complex sentence forms, and sophisticated ideas that other children absorbed from reading. Despite his academic problems, however, Fred continued to show nonverbal intellectual skills within the low normal range, although his scores were somewhat lower than those he had earned as a preschooler.

The disorders of speech, learning, and development may co-occur with language disorder. The two most frequently co-occurring disorders are speech sound disorder and specific learning disorder. Speech sound disorder, which is characterized by difficulty in the correct production of speech sounds, occurs in roughly 40% to 60% of preschoolers with language disorder. Specific learning disorder, especially difficulties in reading, occurs in approximately 50% to 90% of school-aged children with language disorder and most frequently for children with receptive language problems. Difficulties in learning to read may be apparent early or may not surface until later in the learning process when knowledge of sophisticated vocabulary and sentence structures is required to support reading comprehension. Because of this high comorbidity between language disorders and reading problems, some researchers have theorized that language impairment and reading difficulties may be different manifestations of the same underlying problems that manifest themselves at different ages. Following this idea, they suggest that in the evaluation of spelling difficulties in language disorders, phonologic representations, morphologic awareness, and reading skills should all be assessed. Genetic studies also suggest that families with a known risk of language disorders also tend to be at higher risk for reading problems and vice versa. However, other researchers suggest that reading and language problems should be considered separate disorders, as children with language problems manifest problems that go beyond phonologic awareness and include syntax and semantics. Future studies in molecular genetics may help clarify this debate. Similarly, children with language disorder have an increased risk of exhibiting mathematics disorder. A few children with language disorders also experience childhood-onset fluency disorder (stuttering) and developmental co-ordination disorder.

The risk for concomitant psychiatric disorders is high for children with language disorder, especially when receptive language is impaired. Estimates vary, but comorbid psychiatric disorders may affect as much as 70% of school-aged children with language disorder who have severe receptive language problems. The most common comorbid condition is ADHD, occurring in 30% to 60% of cases. Internalizing disorders are also noted but not as often as ADHD. Genetic studies show that the co-occurrence of internalizing disorders with language disorder may be due to common genetic influences. Therefore, children with language disorder may be at a higher risk for internalizing disorders if they already have a genetic predisposition for internalizing problems. Boys are at particular risk for externalizing disorders, where genetic influences may also affect this co-occurrence. Behavior difficulties tend to occur most frequently in children who have concomitant language and reading problems. Girls with language disorders may be withdrawn and quiet, making the detection of language disorders more difficult.

Most children with language disorder do not show frank neurologic symptoms (e.g., tremor, paralysis, or seizures). As a group, however, children with language disorders show elevated rates of mild-to-moderate abnormalities (e.g., in reflexes or fine motor function) on careful neurologic examination. They are also more likely than typically developing children to show mild-to-moderate abnormalities in measures of brain function or structure, such as electroencephalographic (EEG) readings, magnetic resonance imaging (MRI) and functional MRI (fMRI) scans, or diffusion tensor imaging (DTI)-derived fractional anisotropy (FA).

Furthermore, the number of abnormal neurologic signs may be associated with the severity of language difficulties.

With advancing technology, new techniques are being tested to improve the detection and intervention of language disorders. Language disorders may go undiagnosed without proper detection. Automatic speech recognition was tested and showed better accuracy than manual transcribers, providing a promising technique to transcribe sentences and improve the detection of language disorders in school-aged children. Moreover, automated computationally driven techniques in scoring the results from language tasks for children have also been shown to be an effective and accurate method to be used instead of manual entry, which may increase the efficacy and objectivity of measurements. Eye tracking has also shown promising results in detecting early steps of language disorders, where children with language disorder may show slower and atypical verb-based prediction behavior. Overall, researchers suggest that instead of focusing on linguistic measures, analyses of a combination of both linguistic and nonlinguistic features may provide a more accurate detection of a language disorder.

PATHOLOGY AND LABORATORY EXAMINATION

There are no laboratory tests that can provide a diagnosis of language disorder. Neuroimaging techniques, such as MRI and fMRI, sometimes reveal structural or functional brain abnormalities associated with language disorder. There are also new techniques being tested with EEG, where machine learning is being tested to classify infants' neural responses to rhythms in order to identify high-risk children for language disorders. However, such abnormalities have neither been found in all cases nor have their possible etiologic or diagnostic values been determined.

DIFFERENTIAL DIAGNOSIS

Standardized evaluations of expressive and receptive language and nonverbal intellectual functioning are the key elements for differential diagnosis of language disorder. Receptive language testing is particularly important because comprehension deficits may be less obvious than expressive difficulties. Receptive and expressive language performance must fall substantially below the range of normal language performance expected for a child’s age. In addition, the severity of the receptive and expressive language problems must hinder academic performance or social communication. Speech-language pathologists often use observational techniques and analyses of spontaneous language to assess the functional impact of language difficulties and to complement information obtained from standardized language tests. A thorough functional assessment of language abilities may also be substituted when appropriate standardized tests are not available (e.g., if a child speaks a language for which tests have not yet been developed).

A significant discrepancy between receptive language and nonverbal abilities precludes the alternative diagnoses of selective mutism and intellectual disabilities. In selective mutism, receptive language and nonverbal performances usually are similar in level and markedly higher than expressive language skills. In intellectual disabilities, performance is impaired substantially in all three areas of receptive language, expressive language, and nonverbal abilities relative to age-level expectations.

Deficits in receptive language may also be evident in cases of hearing impairment, environmental deprivation, or autism spectrum disorder. A hearing test should be part of the assessment of all cases of suspected language disorder. Collection of a careful history can usually identify severe environmental deprivation. Observations of disturbed social interaction, limited communicative intent, and unusual repetitive behaviors suggest autism spectrum disorder. A concurrent diagnosis of language disorder may be made when receptive and expressive language deficits exceed those expected for the degree of intellectual disabilities, hearing impairment, or environmental deprivation.

Differential diagnosis requires attention to other comorbidity issues. Speech sound disorder is a possible comorbid diagnosis when difficulties in speech sound production accompany language disorder. School-aged children identified with either language or specific learning disorder should be screened for difficulties in the other area because the two disorders frequently co-occur. Similarly, children with psychiatric diagnoses, particularly ADHD, should be screened for language disorder because it is known that a high proportion of children referred for psychiatric assessment shows previously undetected comorbid language disorders. Moreover, children with receptive and expressive language deficits may show signs of reduced autonomy in their play behaviors, as well as reduced eye contact with their parents, which can be monitored even in their prelinguistic skills. Furthermore, sometimes a child shows expressive language that is so difficult to follow and so disorganized that the practitioner questions whether it should be considered disordered thinking as a symptom of schizophrenia.

Illustrative of this concern is the case study presented here with a brief excerpt from a conversation with Max, a highly anxious 9-year-old with language disorder who had just visited the dentist. The stream of language in this case study raises serious concerns about a possible diagnosis of schizophrenia. However, a child with language disorder may produce language this disorganized, especially in situations of heightened anxiety. Diagnosis can usually be clarified by reassessment when the child is less anxious and by careful attention to the presence or absence of comorbid symptoms.

What do you suppose you'd like to do today?

Taking a dentist appointment …

[You're taking a dentist appointment.]

Yeah, he has a little cavity …

[He has a little cavity.]

…and I'm going to take it from him cause he's stuck there he sure doesn't like it but he knows there is too far (mumbled)…

…too far and so he was losing a lot of hair so he needs a new way…sounds like he's bald…so he has everything (mumbling) a guy who's not recep…teaching he will have to it's important like he will or he won't.

Acquired language problems may result from a variety of neurologic insults, including (1) cerebral vascular accidents, such as strokes; (2) infections, such as encephalitis, meningitis, or Zika virus; (3) severe respiratory problems that result in intraventricular hemorrhage or anoxia; (4) near drownings; (5) closed head injuries from falls, motor vehicle accidents, or other blows to the head; (6) open head injuries, such as those from gunshot wounds; (7) toxins, such as lead or household poisons; (8) damage from tumors or radiation treatment; or (9) seizure activity. Landau–Kleffner syndrome (acquired aphasia with epilepsy) is a rare condition marked by an abrupt loss of language skills, which is often associated with the onset of seizures or abnormal EEG activity.

Acquired language problems vary considerably in their language symptoms as a function of the nature and severity of the underlying neurologic condition. Acquired disorders also show distinctive language characteristics as well as some features shared with language disorder. For example, closed head injury, which often occurs in motor vehicle accidents or falls, typically are sparing of basic skills in vocabulary and grammar (after initial recovery stages), whereas more complex skills, particularly in discourse, remain markedly impaired. Comprehension for lengthy or complex discourse may be limited. Likewise, expression in discourse may be disrupted by disorganized, tangential, inappropriate, and illogic comments, even if basic vocabulary and sentence structures are used correctly. Communication effectiveness may be further hindered by accompanying deficits in memory, attention, affect, behavior, and motor skills. In return, these problems and deficits in language skills acquired as a result of brain trauma may create additional obstacles in the rehabilitation process for the injury itself.

Landau–Kleffner syndrome also shows a unique pattern. Children with Landau–Kleffner syndrome seem to be developing normally until around the ages of 3 to 7 years, when they start to have trouble with expressive and receptive language for no apparent reason. The decline in language ability may be sudden or gradual. Hearing loss is usually suspected by parents; however, laboratory tests disconfirm this suspicion. In about 80% of children with Landau–Kleffner syndrome, one or more episodes of epileptic seizures occur, typically during sleep. Behavioral problems such as hyperactivity, aggressiveness, or depression may co-occur.

Elliott, 4 years of age, was a cheerful, talkative preschooler who suddenly experienced a series of seizures for no apparent reason. His language skills, which had been excellent, simultaneously showed a precipitous decline. He no longer understood words, commands, and sentences that had posed no difficulty for him before the seizures. He also lost his expressive skills and began using shorter, less sophisticated sentences, which became increasingly incoherent as he omitted or misused important words. He also developed significant memory and attention problems. Nonetheless, he continued to show nonverbal intellectual skills that were grossly within the normal range. Although Elliott's seizures were eventually controlled with medication, he experienced persistent problems in comprehension and production of language, which never fully resolved.

Course and prognosis are also variable for children with acquired language problems. Some spontaneous recovery generally occurs rapidly within days or weeks of the neurologic insult, but additional improvements may show a protracted course over months or even years. Approximately two-thirds of children with acquired disorders eventually recover language functioning that is broadly within the normal range, although subtle deficits may persist. Those who do not attain normal language functioning show persistent language and cognitive deficits of varying severity. Prognosis depends on many factors, including the severity, extent, and location of brain damage, the specific etiology, the age of onset, the extent of preinjury language development, and the existence of other preinjury disorders.

Language disorders also occur in children from diverse linguistic and cultural backgrounds. A difficult diagnostic challenge is to differentiate language errors that are a normal part of acquiring a new language from those that are indicative of a disorder. A valid assessment usually requires the co-operation of a well-trained informant from the same cultural and linguistic background as the child. The informant assists in collecting and interpreting history information, assessing the child's skills in the other language, and discussing assessment results and recommendations with the family.

COURSE AND PROGNOSIS

Language disorder is characterized by considerable variability in severity, course, and outcome. Individuals who have mild-to-moderate language disorders are less likely to suffer from significant long-term mental health problems than individuals who have severe language disorders. Youngsters with early receptive language delays show higher rates of persisting disorder and less favorable long-term outcomes than those with early expressive-only delays. Some, however, do eventually attain language functioning within the normal range. Better outcomes are generally associated with higher levels of early gestural communication, receptive language, and nonverbal intelligence.

Children with persistent language disorders continue to develop language comprehension and production skills during the preschool and school-aged years. However, the rate of development may be quite slow, causing them to fall further behind peers who are developing language at a normal rate. Overall, the communicative prognosis for children with persistent language disorders continues to be poorer than for those with expressive-only disorders.

Children with language disorders are also at risk for academic problems once they enter school. A negative spiral of poor academic performance, low self-esteem, reduced motivation, and social isolation may stem from the language comprehension difficulties, particularly when they are coupled with attention, reading, or mathematics problems. Fortunately, some progress has been made in using performance on early language-related tasks to predict which kindergarten children may be at risk for later reading difficulties. Moreover, interventions to improve phonologic awareness, a key skill for early reading, have proven successful with children with language disorders, at least in the short term.

Many adolescents and young adults with a history of language disorder continue to show problems in communicative effectiveness, academic performance, and social functioning. These problems may limit vocational choices and social participation. Individuals with language disorder are also at heightened risk of being bullied or victimized in general, including sexual assault or abuse for girls and women. Serious psychiatric problems, such as social anxiety disorder or antisocial personality disorder, may also develop in some individuals. Even though the symptoms of language disorders may improve through development in most children, the internalizing or externalizing behavior problems that co-occur with language disorders may not get better with this improvement.

The societal consequences are significant, as evidenced by the high rate of language difficulties in adults who live in poverty, commit crimes, incur psychiatric illnesses, and achieve limited literacy.

TREATMENT

Experts disagree about the timing of intervention for language disorder. Some advocate a wait-and-see attitude for young children with early expressive delays (late talkers) because most acquire normal language functioning during the preschool years. They argue that intervention should be provided only to those with persistent expressive delays at 4 or 5 years of age. Others suggest that early intervention might prevent or minimize later language, academic, and behavioral difficulties. Unfortunately, few relevant data are available on the efficacy of early- versus late-language intervention.

Direct and mediated interventions have been shown to promote short-term language growth in children with language disorders. In direct interventions, the primary intervention agent is a professional (usually a speech-language pathologist) who works directly with the child to improve language functioning. In mediated interventions, the primary intervention agents are parents, teachers, or paraprofessionals, who have been taught by the speech-language professional to deliver the intervention to the child. Mediators are encouraged to incorporate language facilitation techniques into daily activities and routines, thereby providing regular, naturalistic opportunities for children to use newly learned language features.

Further research on treatment efficacy is needed to address important outstanding issues. For example, studies are required to assess (1) the ways in which interventions might be tailored to address the specific language needs of individual children, (2) the long-term merits of early intervention for children with expressive delays, (3) the extent to which the benefits of intervention achieved in laboratory studies can also be demonstrated in routine community-based interventions, and (4) what the potential harms of early screening or intervention for language delays may be, including the risk involved in false-positive and false-negative results, labeling, parent and child anxiety, and unnecessary evaluation and intervention. One recent study of routine early interventions suggests that systematic efforts may be required to improve their effectiveness.

Children with language disorders are generally recommended for intervention to address their receptive and expressive language needs. Various language intervention techniques and programs have been developed, but some have not been rigorously evaluated to demonstrate their efficacy. The available efficacy literature, however, suggests that language intervention can foster short-term progress toward specific communicative goals, a conclusion that extends to various intervention techniques and programs. With advancements in technology, different technologic intervention techniques have been tested to improve the treatment outcome. For children with co-occurring speech sound and language disorders, electropalatography (EPG), a technique that monitors contacts between the tongue and hard palate, has been shown to be effective in improving regular speech and language therapy outcomes. Another method that is being developed with advancing technology is a speech and language guidance program as an application for smartphones and tablets, where parents and children have an opportunity to benefit from language therapy remotely with lower costs. Unfortunately, there has been limited study thus far on the possible long-term benefits of language intervention in limiting or reversing the negative impacts of language disorder on linguistic, academic, and psychosocial outcomes.

Interventionists have, however, begun to focus on long-term, broad-based outcomes. Preschoolers with language disorders are receiving interventions designed to promote not only oral language outcomes but also social communication and early literacy skills. Kindergartners with language disorders are being taught certain key prereading skills, which may prevent the early reading failure that is linked to subsequent academic and psychosocial problems. More attention is being given to enabling peers, parents, and teachers to learn methods for fostering frequent, natural, and optimal communicative interactions with children who have language disorders. In schools, there are efforts to co-ordinate the various types of interventions that may be required by a child with a language disorder (e.g., language intervention, special reading help, special education support, or behavioral interventions), as well as to train childcare professionals in delivering group interventions which have been effective in increasing children’s phonologic production skills. Even though the parameters for effective dose and intensity of interventions in clinical practice are not yet clearly defined, recent research showed that twice a week may be a more effective method than once a week. Despite some initial promising results from these efforts, much remains to be learned concerning optimal goals, techniques, and programs for language intervention.

Language disorders may also necessitate modifications to other interventions that rely heavily on language as a medium (e.g., psychotherapy or cognitive behavioral modification). Comprehension can be facilitated by simplification of vocabulary, sentence structure, questions, and verbal instructions and by repetition of verbal information as required. Compensation for expressive difficulties can be achieved by using nonverbal response modes to supplement verbal responses. Ambiguous or unclear responses should also be clarified. Assessment procedures that depend heavily on language (e.g., projective tests or diagnostic interviews) may require similar modifications.

Overall, people with language disorders are underserved. The discovery and application of better detection and intervention techniques are necessary to provide better care for this population. Raising awareness, providing detection tests in early stages and in-school settings, and better communication with parents and teachers are some ways that can enhance our current level of knowledge and experience with language disorders.

ACKNOWLEDGMENTS

This chapter was written in memory of the late Dr. Joseph H. Beitchman, who continues to inspire us with his dedication, devotion, and extensive research work in childhood language disorders.

Further Readings

Ahmed ST, Lombardino LJ, Leonard CM. Specific language impairment: definitions, causal mechanisms, and neurobiological factors. *J Med Speech Lang Pathol*. 2001; 9(1):1–15.

Baydar N, Küntay AC, Yagmurlu B, et al. “It takes a village” to support the vocabulary development of children with multiple risk factors. *Dev Psychol*. 2014;50(4):1014–1025.

Beitchman JH, Brownlie EB, Bao L. Age 31 mental health outcomes of childhood language and speech disorders. *J Am Acad Child Adolesc Psychiatry*. 2014;53(10):1102–1110.

*Beitchman JH, Brownlie EB. Language disorders in children and adolescents. In: Wedding D, ed. *Advances in Psychotherapy: Evidence-based Practice*. Hogrefe; 2014.

Bishop DV. How does the brain learn language? Insights from the study of children with and without language impairment. *Dev Med Child Neurol*. 2000;42:133–142.

Bishop DVM, Snowling MJ. Developmental dyslexia and specific language impairment: same or different? *Psychol Bull*. 2004;130(6):858–886.

*Brownlie EB, Bao L, Beitchman J. Childhood language disorder and social anxiety in early adulthood. *J Abnorm Child Psych*. 2016;44:1061–1070.

*Buschmann A, Jooss B, Rupp A, et al. Children with developmental language delay at 24 months of age: results of a diagnostic work-up. *Dev Med Child Neurol*. 2008;50(3):223–229.

Calder SD, Claessen M, Leitão S, Ebbels S. Evaluating two different dose frequencies and cumulative intervention intensities to improve past tense production for early school-aged children with developmental language disorder. *Int J Lang Commun Disord*. 2021;56(6):1278–1295.

Cohen NJ, Vallance DD, Barwick M, et al. The interface between ADHD and language impairment: an examination of language, achievement, and cognitive processing. *J Child Psychol Psychiatry*. 2000;41:353–362.

Costa CH, Molini-Avejonas DR. Development of speech therapy app for parents. *Codas*. 2020;32(5):e20190123.

Dale PS, Price TS, Bishop DVM, Plomin R. Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *J Speech Lang Hear Res*. 2003;46(3):544–560.

Falcaro M, Pickles A, Newbury DF, et al. Genetic and phenotypic effects of phonological short-term memory and grammatical morphology in specific language impairment. *Genes Brain Behav*. 2008;7(4):393–402.

Gale R, Bird J, Wang Y, et al. Automated scoring of tablet-administered expressive language tests. *Front Psychol*. 2021;12:668401.

Gibbon S, Attaheri A, Ní Choisdealbha Á, et al. Machine learning accurately classifies neural responses to rhythmic speech vs. non-speech from 8-week-old infant EEG. *Brain Lang*. 2021;220:104968.

Law J, Rush R, Schoon I, Parsons S. Modeling developmental language difficulties from school entry into adulthood: literacy, mental health, and employment outcomes. *J Speech Lang Hear Res*. 2009;52:1401–1416.

Leonard LB. *Children with Specific Language Impairment*. Bradford; 1998.

*Mayes AK, Reilly S, Morgan AT. Neural correlates of childhood language disorder: a systematic review. *Dev Med Child Neurol*. 2015;57:706–717.

McGregor KK. How we fail children with developmental language disorder. *Lang Speech Hear Serv Sch*. 2020;51(4):981–992.

McGregor KK, Oleson J, Bahnsen A, Duff D. Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *Int J Lang Commun Disord*. 2013;48(3):307–319.

Mohan M, Bajaj G, Deshpande A, Anil MA, Bhat JS. Child, parent, and play – an insight into these dimensions among children with and without receptive expressive language disorder using video-based analysis. *Psychol Res Behav Manag*. 2021;14:971–985.

Paul R. *Language Disorders from Infancy through Adolescence: Assessment & Intervention*. 2nd ed. Mosby; 2001.

Plug MB, van Wijngaarden V, de Wilde H, et al. Clinical characteristics and genetic etiology of children with developmental language disorder. *Front Pediatr*, 2021;9:651995.

Plym J, Lahti-Nuuttila P, Smolander S, Arkkila E, Laasonen M. Structure of cognitive functions in monolingual preschool children with typical development and children with developmental language disorder. *J Speech Lang Hear Res*. 2021;64(8):3140–3158.

Prior M, Bavin E, Cini E, Eadie P, Reilly S. Relationships between language impairment, temperament, behavioural adjustment and maternal factors in a community sample of preschool children. *Int J Lang Commun Disord*. 2011;46(4):489–494.

Robinson RO, Baird G, Robinson G, Simonoff E. Landau-Kleffner syndrome: course and correlates with outcome. *Dev Med Child Neurol*. 2001;43:243–247.

Spinath FM, Price TS, Dale PS, Plomin R. The genetic and environmental origins of language disability and ability. *Child Dev*. 2004;75:445–454.

*Tomblin JB, Zhang X, Buckwalter P, O'Brien M. The stability of primary language disorder: four years after kindergarten diagnosis. *J Speech Lang Hear Res*. 2003;46:1283–1296.

*Toppelberg CO, Shapiro T. Language disorders: a 10-year research update review. *J Am Acad Child Adolesc Psychiatry*. 2000;39(2):143–152.

Toseeb U, Oginni OA, Dale PS. Developmental language disorder and psychopathology: disentangling shared genetic and environmental influences. *J Learn Disabil*. 2022;55(3):185–199.

▲ 42.2 Social (Pragmatic) Communication Disorder

MIRELLE KASS, B.A., YOUNG SHIN KIM, M.D., M.S., M.P.H., PH.D., AND BENNETT L. LEVENTHAL, M.D.

Social (pragmatic) communication disorder (SCD) is a persistent disruption in the communication of social concepts, skills, and intentions. While social communication deficits coexist as a component of multiple disorders, for individuals with SCD, these problems stand alone.

Social pragmatics is derived from the concept of pragmatic language. Language is the full spectrum of expressive and receptive communicative behaviors, including speech, gesture, and other skills. Pragmatics encompasses those behaviors that contribute to implication, relevance, and effectiveness of any form of an individual's communication. In this framework, for effective communication, an individual must follow pragmatic language rules to create and transmit any given message. Similarly, individuals must understand pragmatic rules when attempting to interpret receptive communication.

Language pragmatics relate to how context contributes to understanding and meaning in communications, with each component of communication having its own structure and rules. When speech is used, the rules apply to what is being said and how it is being said. The “what” refers to the choice of words, sentence structure, and connections between different elements of content, such as conjunctions and dependent clauses, as well as the use of different types of sentences (statements, questions, etc.). The “how” of speech applies to the rate, rhythm, prosody, volume, tone, and other characteristics of speech production. Pragmatic rules also apply to nonverbal communication, including gesture, body posture, reciprocity, and facial expression. When the pragmatic rules can be correctly and flexibly applied in context, expressive and receptive language can be effective. When they cannot be so applied, individuals have broad difficulties in social intercourse leading to significant problems in adaptation.

SCD is a diagnostic category first introduced in DSM-5. The main characteristic of SCD is the persistent deficit in the social use of expressive and receptive, nonverbal and verbal communication, as well as difficulty using the linguistic context to make inferences and to resolve ambiguities. Symptoms of SCD include deficits in the understanding and the application of social and language rules for verbal and nonverbal communication. Individuals with SCD have difficulty adapting language to the needs of the listener or situation. Similarly, those with SCD have difficulty following the rules of conversations (such as turn taking), storytelling, humor, sarcasm, and other nuanced speech. As a result, individuals with SCD have difficulty communicating and understanding ambiguous and nonliteral language, as well as integrating linguistic context and prior social knowledge.

The elements of SCD overlap with the “A” criterion of DSM-5-TR autism spectrum disorder (ASD). However, SCD is distinct from ASD due to the absence of the DSM-5-TR ASD “B” criterion, “restricted, repetitive patterns of behavior, interests, or activities….” Although SCD can co-occur with other communication disorders, such as language disorder, the social communication and pragmatic deficits for SCD must exceed what would be expected from structural language deficits alone. The diagnosis of SCD cannot be made along with the diagnosis of ASD.

HISTORY

While social and language pragmatics have long been a clinical concern for many disorders, SCD was not considered a diagnostic category until the creation of DSM-5. Social communication and pragmatic deficits have been documented in the education and speech and language literature for decades, if not longer. Much of the concern about SCD, and other language disorders, focused on children although it is apparent that SCD and other developmental language disorders (DLDs) persist into adulthood.

Previously, semantic–pragmatic deficits characterized children demonstrating problems with verbosity, word finding, word choice, and poor topic maintenance. A similar term introduced in the late 1980s was “semantic–pragmatic disorder,” which was thought to occupy the diagnostic space between ASD and specific language impairment. Both terms reflect the idea that semantic (i.e., the meaning of an expression) and pragmatic (i.e., the use of language in context) difficulties tend to co-occur. In the 1990s, research found that semantic deficits did not necessarily co-occur with pragmatic deficits, and that semantic abilities did not reliably differentiate between typically developing and clinical groups. As a result, the term pragmatic language impairment was coined and appears commonly in the speech and language literature.

While there was much research and clinical progress in understanding etiologic substrates and clinical presentation of autism, as a result of unifying diagnostic criteria between DSM-IV and ICD-10, there was still substantial difficulty distinguishing the subcategories in DSM-IV pervasive developmental disorder (PDD). Research found no etiologic distinctions and a lack of diagnostic reliability with these criteria. In part, this was due to Asperger syndrome becoming a preferred diagnosis for many families because it was associated with much less stigma than the DSM-IV diagnosis of autistic disorder—even today, some still falsely “blame” parenting failure as a cause of autism.

The addition of SCD to DSM-5 is closely associated with the changes made to the diagnostic criteria concerning ASD. In DSM-IV-TR, PDDs included autistic disorder, Asperger syndrome, and PDD—not otherwise specified (PDD-NOS). PDD-NOS was for patients who did not meet the full criteria for either autistic disorder or Asperger syndrome. Children or adults who met one or two of the three major autism criteria showed core deficits in one area, such as social deficits, but did not meet the diagnostic criteria for other symptoms, such as restrictive, stereotypical, and repetitive behaviors, would have been given the diagnosis of PDD-NOS. Research has shown that more than half of the children diagnosed with PDD-NOS were given the diagnosis due to the fewer restrictive and repetitive behaviors they exhibited.

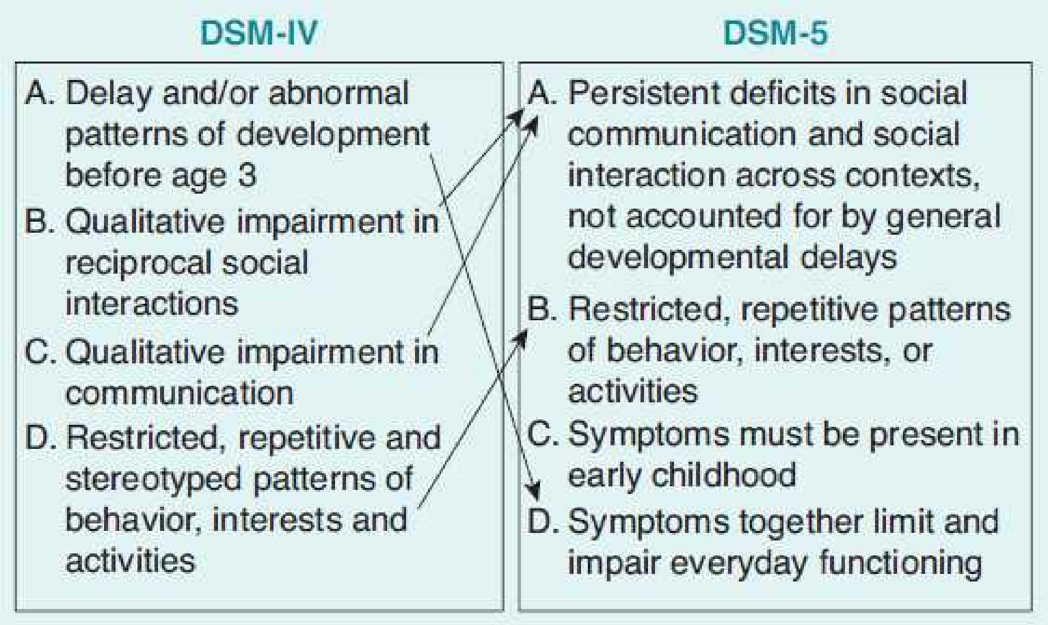
DSM-IV
DSM-5
A. Delay and/or abnormal patterns of development before age 3
B. Qualitative impairment in reciprocal social interactions
C. Qualitative impairment in communication
D. Restricted, repetitive and stereotyped patterns of behavior, interests and activities
A. Persistent deficits in social communication and social interaction across contexts, not accounted for by general developmental delays
B. Restricted, repetitive patterns of behavior, interests, or activities
C. Symptoms must be present in early childhood
D. Symptoms together limit and impair everyday functioning

FIGURE 42.2–1. Social (pragmatic) disorder changes from DSM-IV to DSM-5.

With the advent of DSM-5, PDD was abandoned. In part, this was due to the clarification of diagnostic criteria which were reorganized from three (B to D) to two (A and B) main criteria: (A) persistent deficits in social communication and social interaction, and (B) restricted, repetitive patterns of behavior, interests, or activities. Interestingly, most of the previous PDD diagnostic criteria were rearranged and preserved in DSM-5 ASD (Fig. 42.2–1).

These changes caused considerable stress for some individuals and families as they either did not want to be associated with the term “autism” or feared that they might lose access to services. As will be discussed below, no one was “left out”; however, the change in nosology is still disconcerting for some.

The addition of SCD to DSM-5 was a result of the changes made to the diagnostic criteria to create ASD. In DSM-5, the absence of PDD -NOS and Asperger syndrome meant that there was no longer a place to categorize individuals who only met the DSM-5 ASD “A” criterion (or DSM-IV “D” criterion). Therefore, children or adults who met only the social deficits diagnostic criterion but not the “B” criterion (restrictive, stereotypic, and repetitive behaviors) could not be the new diagnosis: SCD. While there remains a lack of clarity, this diagnosis may allow room for individuals with social interaction deficits and mild or no restrictive and repetitive behaviors. These seeming language disorders remained in the DSM-5 section for neurodevelopmental disorders.

While the diagnosis of SCD may appear to be consistent with the broader concept of pragmatic language impairment, multiple definitions of pragmatic language impairment exist, many of which do not include nonverbal communication. Therefore, further study will be required to determine to what extent SCD is related to the understanding of language pragmatics found in the broader language pragmatics literature.

DIAGNOSTIC FEATURES

SCD is a disruption in the development and use of language pragmatics leading to interference with social functioning (Table 42.2–1).

For this diagnosis, pragmatics refers to the rules that govern social interactions when using both verbal and nonverbal language. It is these rules that allow people to relate to one another reciprocally as well as to create and maintain social transactions. For typically developing children, language pragmatics (initially nonverbal) appear very early and expand rapidly—think of the social smile, visual regard, and pointing as examples. While many think pragmatics are completely intuitive to the typically developing child, it is likely that caregiver reinforcement of these skills increases their quantity, quality, frequency, variability, and utility. In contrast, children with SCD may never exhibit these behaviors, or they are so rare that caregiver reinforcement is not possible. This is rarely a parent/caregiver problem but, rather, a developmental delay or developmental deviation. As a result, children with SCD have early difficulty learning and using even the most basic rules of social communication. While this often begins with delays in nonverbal communication, it generally continues with delays and/or deviations in spoken language, when to listen, how to phrase a question, how to stay on topic, and knowing when the talk is over. As a result of deficits in pragmatics, individuals with SCD may struggle to engage in and/or sustain conversations. Problems may include frequent interruptions, monopolizing conversations, lack of reciprocity, failing to effectively communicate wants and/or needs, or not talking at all. Individuals with SCD may also say things that are inappropriate or unrelated to the conversation at hand. Taken together, these communication problems often make it difficult for individuals with SCD to develop and sustain meaningful social relationships and communicate well with others at school or in the workplace. In addition to the language problems, the communication deficits often lead to frustration (with attendant behavior problems), anxiety, and depression. The symptoms of SCD include the following:

1. Problems using nonverbal and verbal language, such as greeting and sharing information, in ways appropriate for the social context.
2. Impairment in changing communication to match the context or the need of the listener, such as speaking differently in the classroom than on the playground or talking to a child than to an adult.
3. Difficulties following social conventions in conversation and story-telling, such as taking turns in conversations, rephrasing when misunderstood, and knowing how to use verbal and nonverbal signals to regulate the interaction.
4. Difficulty making inferences when something is not explicitly stated, and difficulty understanding the meaning of nonliteral or ambiguous language, such as idioms, humor, and metaphors, and when there are multiple meanings depending on the context.

Table 42.2–1.
DSM-5-TR Criteria for Social (Pragmatic) Communication Disorder

A. Persistent difficulties in the social use of verbal and nonverbal communication as manifested by all of the following:
   1. Deficits in using communication for social purposes, such as greeting and sharing information, in a manner that is appropriate for the social context.
   2. Impairment of the ability to change communication to match context or the needs of the listener, such as speaking differently in a classroom than on a playground, talking differently to a child than to an adult, and avoiding use of overly formal language.
   3. Difficulties following rules for conversation and storytelling, such as taking turns in conversation, rephrasing when misunderstood, and knowing how to use verbal and nonverbal signals to regulate interaction.
   4. Difficulties understanding what is not explicitly stated (e.g., making inferences) and nonliteral or ambiguous meanings of language (e.g., idioms, humor, metaphors, multiple meanings that depend on the context for interpretation).

B. The deficits result in functional limitations in effective communication, social participation, social relationships, academic achievement, or occupational performance, individually or in combination.

C. The onset of the symptoms is in the early developmental period (but deficits may not become fully manifest until social communication demands exceed limited capacities).

D. The symptoms are not attributable to another medical or neurological condition or to low abilities in the domains of word structure and grammar, and are not better explained by autism spectrum disorder, intellectual developmental disorder (intellectual disability), global developmental delay, or another mental disorder.

The communication problems facing individuals with SCD often impair effective communication overall, as well as social participation, academic achievement, and/or occupational performance. The many difficulties with social interactions begin early in life, although the extent of the impairments may not become fully apparent until social communication requirements become more demanding in the preschool or school-age period.

DSM-5-TR criteria require that the symptoms of SCD are not attributable to another psychiatric, other medical disorder, or neurologic conditions, including, but not limited to intellectual disability, global developmental delay, ASD, or any other mental disorders. The deficit in social communication should not be solely attributable to problems in the mechanics of language, as is the case for individuals with language disorder.

Individuals with SCD may show little interest in social interactions. This may appear to be similar to those with ASD. Individuals with SCD may have difficulty adapting language to different listeners (e.g., talking the same way to an adult and to a child) and to different situations (e.g., talking the same way in a classroom as on the playground). They may have trouble making inferences and/or understanding abstractions or implied thoughts or ideas, while they may be readily able to understand that which is stated explicitly. They may not give necessary information when they are speaking to an unfamiliar person, making it difficult to contextualize their utterances. A significant challenge is having difficulty properly gaining the attention of others to make a request or initiate a conversation. People with SCD are often quite literal in their use and understanding of words. This extends to understanding nuanced language and behavior, making inferences with incomplete information, and noticing social subtleties and social cues. As a result, they have difficulty understanding sarcasm, riddles, and humor, and they have trouble telling even the simplest jokes. Some, but not all, individuals with SCD have difficulty understanding nonverbal communication, including gestures, facial expressions, and body posture, in some cases in spite of more than ample vocabulary. It is important to remember that SCD is both an expressive and receptive communication disorder.

Some children with SCD may also have trouble with reading comprehension, despite having reasonably well-developed phonemic awareness and the ability to read well phonetically—they can read it, but they don't understand what they have read. These reading issues may not be noticed until midelementary school or later.

It has been suggested that SCD rarely occurs by itself. Language disorder is the most commonly occurring disorder associated with SCD.

SCD was found to be more common in eighth-grade children with a history of DLD, especially when deficits in pragmatics and structural language occur simultaneously. Having a language disorder of this type during kindergarten was found to increase the risk for SCD during adolescence. However, it was not predictive when “pragmatic and social skills were disproportionately lower than structural language abilities,” according to Ellis Weismer.

Attention-deficit hyperactivity disorder (ADHD) is a common comorbidity associated with SCD. The hyperactivity and inattention problems can aggravate the problems of pragmatic competence. In addition, a lack of social pragmatic skills may lead to children becoming frustrated and having behavior problems.

The following vignettes illustrate SCD in youth with deficits in receptive and expressive nonverbal communication, along with relatively intact structural language.

J is a 6-year-old boy with an age-appropriate vocabulary who often makes comments that are completely out of context. He makes unusual associations with words or phrases and then perseverates on the association. In a similar way, he responds inappropriately to questions from his peers. Taken together, this leads to great difficulty playing with other children. For example, in the midst of playing a game with peers, J said: “What a lovely bouquet.” which he then repeated several times, leaving his peers perplexed. They tried to reengage J, who then just wandered away, repeating the phrase. J’s articulation is clear, and he is able to speak in grammatically correct sentences. He is disappointed and sad because he cannot play with his peers but has little understanding of why.

N is a 17-year-old youth who is often physically bullied by other youth. He is frequently the object of teasing because he is prone to misinterpret the intentions of others and is easily provoked. He has a history of misjudging jokes and even the positive intentions of others. When he speaks, he has virtually no facial expression and very few gestures. He has difficulty initiating and maintaining friendships. He seeks friendships with individuals 4 years older than he but generally does better with children who are much younger. Irrespective of the age group, he has difficulties in social situations. He is aware of this and wants to change so that he can have friends. He plans to follow his therapist's recommendation to avoid the "mean kids" but is not always sure how to tell who they are. He has no repetitive, restricted interests or behaviors, and his cognitive ability is in the "above average" range.

Diagnostic assessment tools for SCD remain quite limited. Studies that screen for SCD and related disorders often use the Children's Communication Checklist (CCC). This 70-item parent-report questionnaire screens for pragmatic abilities and skills within the context of their child's overall language skills. The second edition of this questionnaire, the CCC-2, is one of the few validated tools that measure deficits in pragmatic abilities. Translations of this tool exist for 30 different languages, and normative data for the CCC-2 are available for over 1,400 children from the United States and the United Kingdom.

The Test of Pragmatic Language (TOPL) measures a few characteristics of pragmatic communication, including audience, topic, visual, gestural cues, physical setting, purpose, and abstraction in children between the ages of 6 and 18. The Comprehensive Assessment of Spoken Language measures expressive language skills, such as idiomatic language, nonliteral language, and pragmatic judgment, in children and adolescents between the ages of 3 and 21. However, these instruments are nonideal for diagnostic purposes due to their dichotomous scoring systems, according to Swineford et al. The Pragmatic Rating Scale measures 19 behaviors related to pragmatics, all of which are included in the SCD criteria. This measure includes a 15-minute conversation, derived from the Autism Diagnostic Observation Schedule (ADOS), that is scored from "normal behavior to strikingly abnormal behavior." The Pragmatic Protocol is used for children at a minimum age of 5 and involves a 15-minute observation period in which unstructured conversation occurs. Although the material of these instruments is relevant to SCD, there is a lack of validity and reliability data for either tool.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of SCD can be complex as it must take into consideration the relationship between SCD and psychiatric disorders, especially ASD, intellectual disability, ADHD, and language disorder. Social communication or pragmatic language deficits are common among children referred for psychiatric or psychological assessment. Similarly, comorbidity rates are especially high for children with language disorder, including ADHD, conduct disorder, and traumatic brain injury. Differential diagnosis is especially challenging due to the overlap in symptoms among various psychiatric disorders and the lack of well-validated assessment tools with high levels of reliability. When the DSM-5 was applied to reassess 80 children with DSM-IV-TR PDD diagnoses, 6 children were found to meet SCD criteria, and another 9 children did not meet DSM-5 ASD criteria. Furthermore, six subjects diagnosed with either Asperger disorder, ASD, or PDD-NOS were found to meet DSM-5 criteria for SCD. A study by Mandy et al. (2017) found that their entire sample of 801 children with ASD satisfied criterion B for SCD, and 600 of them also met all four symptom types of criterion A; only 2 of these children did not meet any symptoms of SCD criterion A.

Social communication is one of the two core deficits of ASD. Children with ASD may show adequate structural language abilities yet still have difficulty with pragmatic language, such as pedantic or overly formal speech, verbosity, and reciprocal conversations. For children with ASD, these pragmatic language and social communication difficulties are accompanied by fundamental social deficits, along with restricted interests and repetitive behaviors. The past and/or present symptoms of clinically significant specialized interests and restricted, repetitive behaviors help differentiate ASD from SCD.

Because a diagnosis of ASD can be given to individuals with just a history of fixed interests and repetitive behaviors, individuals who only show significant social and pragmatic language deficits but no present behavioral patterns characteristic of ASD may still be given a diagnosis of ASD. Complicating the differential diagnosis further, stereotyped language, which was previously considered to be a symptom of pragmatic language impairment, is now considered part of the repetitive and restrictive behaviors required for giving a diagnosis of ASD. Finally, some children may show an increase in repetitive behaviors and specialized interests as they grow older, further complicating the differential diagnosis.

When considering a primary language disorder as part of the differential diagnosis, usually children with language disorder show deficits in word knowledge, grammar, and discourse. Discourse includes conversational exchange (reciprocity); narrative deficits in these areas tend to be manifest in social communication problems. However, for children with primary language disorders, social and pragmatic language deficits are thought to be a result of, or secondary to, the structural language impairment. Therefore, a clear relationship between language function and measures of social deficits often supports a diagnosis of language disorder. Other language dysfunction, such as limited vocabulary, prosodic abnormalities, articulation problems, receptive language deficits, and reading problems, increase the likelihood of a language disorder diagnosis.

Some studies have shown group differences in symptom profiles of social communication and pragmatic deficits, and making a differential diagnosis based solely on the basis of social communication, and pragmatic deficits can prove to be challenging. For example, children with Williams syndrome have large vocabularies and are hyperverbal while having fewer problems with stereotyped language, nonverbal communication, social relations, and coherence when compared to children with ASD. An observational study comparing children with SCD and children with language disorder has found that, as a group, children with SCD tend to have more problems with turn taking in conversations, provide no response or inappropriate responses to the communication at hand, and make little use of nonverbal signals when compared to children with language disorder. Much more work is necessary to establish an appropriate level of sensitivity and specificity when assessing social and pragmatic language profiles for the purpose of differential diagnosis, as well as establishing risk factors associated with different clinical populations.

Some have suggested an overlap between SCD symptoms and those in social anxiety disorder. This is due to the apparent uneasiness in social situations in both disorders. However, the time for the onset of symptoms helps differentiate between these two disorders. Children with social anxiety disorder have typically developing social communication skills yet do not utilize them in some settings due to anxiety associated with social interactions. Children with SCD, on the other hand, have early onset of symptoms and usually never had effective social communication.

Although there are no diagnostic tools that specifically screen for SCD according to DSM-5 criteria, some assessment tools have been able to differentiate these diagnoses, according to Yuan and Dollaghan. A study by Morales-Hidalgo et al. identified the ADI-R, the ADOS-2, and a newly developed assessment, the EDUTEA, as three tools that appropriately differentiate ASD/SCD diagnoses from ADHD diagnoses. (The EDUTEA is an 11-item instrument with two factors: social communication impairments and restricted behavior patterns.) High reliability was found for both factors (.93–.95), along with high sensitivity (87.0%) and specificity (91.2%). In addition, the EDUTEA demonstrated sufficient ability to differentiate ASD/SCD diagnoses from ADHD diagnoses, as exemplified by a high sensitivity (83%), specificity (73%), and AUC score (.78).

ETIOLOGY

There is no published systematic research on the etiology of SCD, including examining the heritability and environmental factors associated with SCD using the DSM-5 criteria. The relatively recent labeling of this disorder means that there are few studies examining the familial aggregation and genetic factors associated with SCD. Although the etiology of SCD remains unknown, in the absence of evidence, theories abound.

There is accumulating evidence suggesting that a family history of ASD, language disorder, or specific learning disorder may increase the risk of having social communication and pragmatic language deficits. Many studies are emerging, supporting correlations of impairments in pragmatic language with common gene variants. Duplications at 19p13.12p12 on the short arm of chromosome 19 were identified in a male adult who presented with SCD and intellectual deficits, suggesting that this variant may contribute risk for SCD. In addition, the MET promoter variant rs1858830 C from the receptor tyrosine kinase has been found to be associated with difficulties in social communication in a sample of 748 individuals with ASD from 367 families. However, these genetic variants are not specific to pragmatic language impairments; they are associated with other neurodevelopmental and psychiatric disorders, including ASD, intellectual disorders, and language disorders.

Social inference theory suggests that individuals with SCD, much like individuals with ASD, may have deficits in social cognition, especially in theory of mind (ToM), the ability to make inferences about the mental state of others and use that information to understand and predict their intentions and behaviors. Some have suggested that ToM is an essential precursor for the development of social communication. However, it is equally possible that understanding the mental states of others develops as a result of social communication.

The weak central coherence theory posits that social communication deficits may be associated with a more general processing problem of integrating disparate sources of information into a coherent whole that then allows for specific transactions. This theory suggests that social communication deficits occur due to social communication problems, narrowly focusing on inappropriate or less crucial aspects of the communication while ignoring the bigger picture, which, in turn, makes it difficult for them to use linguistic and social context to derive meaning.

The executive dysfunction theory by Vengas and Davidson suggests that individuals with pragmatic language and communication deficits have a general issue with executive functioning. Executive functions are associated with (but not limited to) activity in the prefrontal cortex, which is associated with high-level cognitive functioning impulse control, behavior monitoring, the flexibility of thinking, abstract thinking, concept formation, problem solving, and planning. The executive dysfunction theory helps explain the great heterogeneity in pragmatic language difficulties experienced by individuals with social and pragmatic difficulties. Executive dysfunction has been implicated in multiple psychiatric disorders, such as ASD and ADHD, each of which is associated with pragmatic deficits.

At this point, there is simply not enough research to support any etiologic substrate of SCD or other social communication or pragmatic language deficits among individuals with SCD.

EPIDEMIOLOGY

There are no reliable prevalence estimates for SCD. It remains unclear as to whether demographic factors such as age, gender, and socioeconomic backgrounds correlate with SCD prevalence estimates. Out of 1,081 children and adolescents seen in a social communication clinic in England, only 8% met SCD criteria, suggesting that the prevalence of this disorder remains very low. A study in the United Kingdom found that 22% of the children attending a language clinic had pragmatic language impairment. However, due to the differences in diagnostic criteria, it is unclear what percentage of these children will also meet the diagnostic criteria for SCD.

Deficits in pragmatic language have been reported without the presence of social deficits or repetitive activities and behaviors. Another estimate showed that over 50% of the children who were given the diagnosis of PDD-NOS had early onset of social communication deficits. In a study of 94 children with PDD-NOS diagnoses, social communication deficits were significantly more present in the children who had comorbid psychiatric diagnoses when compared to those without comorbid disorders. However, none of the above studies used the DSM-5 criteria for SCD; therefore, prevalence is likely to differ from that of SCD.

Table 42.2–2.
Prevalence Estimates: DSM-IV PDD, DSM-5 ASD, DSM-5 SCD

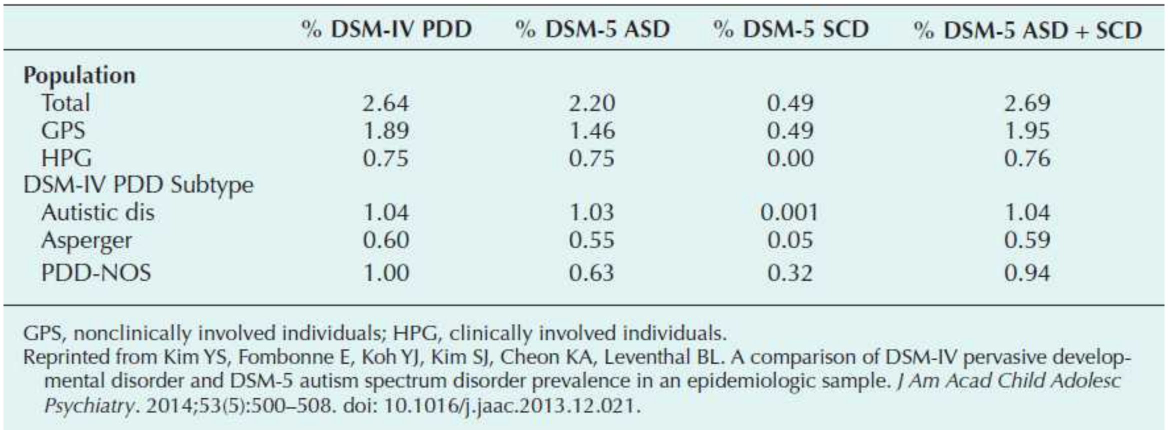

	% DSM-IV PDD	% DSM-5 ASD	% DSM-5 SCD	% DSM-5 ASD + SCD
Population				
Total	2.64	2.20	0.49	2.69
GPS	1.89	1.46	0.49	1.95
HPG	0.75	0.75	0.00	0.76
DSM-IV PDD Subtype				
Autistic dis	1.04	1.03	0.001	1.04
Asperger	0.60	0.55	0.05	0.59
PDD-NOS	1.00	0.63	0.32	0.94

GPS, nonclinically involved individuals; HPG, clinically involved individuals.
Reprinted from Kim YS, Fombonne E, Koh YJ, Kim SJ, Cheon KA, Leventhal BL. A comparison of DSM-IV pervasive developmental disorder and DSM-5 autism spectrum disorder prevalence in an epidemiologic sample. *J Am Acad Child Adolesc Psychiatry*. 2014;53(5):500–508. doi: 10.1016/j.jaac.2013.12.021.

A field trial comparing the prevalence of ASD using the DSM-5 criteria and PDD using the DSM-IV criteria showed a decrease in the prevalence of ASD in one site. Nevertheless, when the prevalence of ASD was combined with the prevalence of SCD, the prevalence was not significantly different from that of DSM-IV PDD. This suggests that the decrease, if any, in the estimated prevalence of ASD may be accounted for by the inclusion of DSM-5 SCD.

In a total population sample from Korea, Kim et al. initially diagnosed the children using DSM-IV criteria. They were subsequently rediagnosed using DSM-5 criteria. They found that the population prevalence of SCD was 0.49%. In addition, they reported that of the 1.00% previously diagnosed with PDD-NOS, 0.32% met the criteria for SCD, while of the 0.60% diagnosed with Asperger syndrome, 0.05 met criteria for SCD (0.55% met ASD criteria) (Table 42.2–2).

DEVELOPMENT AND COURSE

The diagnostic category of SCD is so recent that there are not sufficient longitudinal studies to give a clear picture of the development of SCD.

The diagnosis of SCD is difficult prior to the development of speech as a tool for complex social interactions between the ages of 3 and 5. However, for those children who are speaking by age 2 and those children with limited social reciprocity, SCD should be a part of the differential diagnosis during evaluations. By 4 or 5 years old, most children should have developed sufficient social communication and language skills. Those who have not done so should undergo thorough evaluations with consideration of SCD as a diagnosis, among other NDDs. While most cases are evident early in childhood, some milder forms of SCD may escape detection until adolescence, when the social communication and language demands intensify. Although SCD requires relatively intact structural language and impaired social communication, there is great heterogeneity in the severity and patterns of symptom manifestation among those who are diagnosed with SCD. One study found that 64 out of 66 children who had been previously diagnosed with DSM-IV PDD-NOS had “one distinct symptom pattern, namely social impairments without significant repetitive and stereotyped behaviors.” As a group, children with ASD tend to have more severe social communication and pragmatic language deficits than children with SCD. Children with language disorder, in turn, tend to show milder symptoms of social and pragmatic language deficits than children with SCD.

With no epidemiologic longitudinal follow-up studies, it is difficult to describe the long-term outcome of SCD. Given the heterogeneity of symptoms, it is likely that outcomes will differ greatly. Clinical experience suggests that some children with social and pragmatic skills deficits improve considerably over the course of development while others have persistent social communication skills deficits extending into adulthood.

In general, studies suggest that children with pragmatic language difficulties tend to perform better in academics when compared to children with structural language problems. In addition, children with social communication deficits may also struggle with written expression and reading comprehension. Gilmour et al. found that deficits in pragmatic language skills were frequently observed in children who were persistently disruptive in school settings. There is some evidence suggesting that pragmatic language difficulties in elementary school years may be associated with emotional and behavioral difficulties, higher rates of referral for psychiatric and psychological assessments, and longer duration attending special education.

Social communication problems rarely occur in the absence of other psychiatric disorders. Comorbid disorders, as well as other risk factors, are likely to affect the developmental course and prognosis of SCD.

TREATMENT

Social (pragmatic) deficits rarely occur without other cooccurring behavioral or emotional difficulties. These behavioral and emotional problems may serve as risk factors for further social communication difficulties, and social difficulties, in turn, may accentuate the cooccurring behavioral and emotional problems.

Since SCD is relatively uncommon and only recently recognized as a separate clinical disorder, diagnosis-specific treatments are not readily available. However, this is not unusual and represents the situation for most neurodevelopmental disorders for which most treatments are directed at specific symptoms rather than the diagnosis. As a result, in the current clinical context, the clinician should consider any treatments that are focused on improving social as well as social pragmatic language functions.

One randomized controlled trial examined the social outcomes of children with social communication deficits (both with symptoms of SCD and ASD) who received intensive, individualized, multifaceted interventions as compared to children who received treatment as usual.

The intervention focused on developing (1) social interaction skills and social understanding, (2) verbal and nonverbal pragmatic skills, and (3) language processing skills. The children received up to 20 sessions from specialized speech–language professionals and trained assistants. The children who received the experimental intervention showed significant improvement in blind ratings of communication skills, teacher ratings of learning behaviors, and parental ratings of social and pragmatic skills. There were, however, no significant changes in the standardized assessment measures of social communication, which may be due to the lack of sensitivity of these measures to detect change.

Several treatment options have been proposed for children with social communication and pragmatic language difficulties who also present behavioral/emotional problems. Several behavioral interventions and cognitive behavioral treatment programs have been proposed, involving multiple collaborators, including speech–language therapists, mental health professionals, family members, teachers, and peers. Irrespective of who implements the treatment, significantly greater improvements have been observed in children who begin treatment early, with maximum benefits occurring in children who begin between the ages of 3 to 4.

The Social Communication Intervention Programme (SCIP) is a manualized, theoretically driven intervention for children with difficulties in social communication. In one study of school children with difficulties related to pragmatics and social communication, SCIP was more effective in addressing pragmatic competence as well as modifications in social communication and language skills; however, no improvements in language functioning were observed. A more recent SCIP feasibility study found that all 15 school-aged children made significant progress according to the practitioner ratings; parent report suggests that 1 of the 15 children did not make progress.

Functional communication training is a behavioral technique that has been used to treat problem behaviors associated with communication problems. After a functional analysis is conducted to identify the environmental triggers that evoke the problematic behavior, a targeted behavioral reinforcement intervention is initiated. Usually, the intervention focuses on increasing socially acceptable communicative responses and behaviors in a variety of social contexts.

Cognitive behavioral therapy (CBT) may also be useful, especially in the presence of anxiety and mood symptoms or specific comorbidity that is amenable to CBT.

Didactic training in social skills on its own has yielded few meaningful social changes. This is likely because individuals with SCD do not generally have impaired knowledge of social norms and appropriate behaviors. Interventions targeting underlying deficits, developing strategies to minimize distracting information, or making explicit important verbal and nonverbal cues may be more beneficial.

A social cognitive reinforcement program has been developed for children with SCD. This intervention has three components: executive function training, social cognition training, and communication skills training. It is divided into three steps: introduction (session 1), development (sessions 2 to 7), and conclusion (session 8). Twenty-two children with SCD participated in a comparison study between this intervention and a typical social skills training program. The children who engaged in the social cognitive reinforcement program improved significantly in multiple areas: nonverbal communication, self-assertion, comprehension, and self-control, as demonstrated by notable improvements measured by the Reading the Mind in the Eyes Test, the Children’s Color Trails Test, a picture arrangement task, and a picture completion task. Furthermore, at the 3-month follow-up, significant improvements in executive functioning, pragmatic communication, and social skills were observed in the children who participated in the social cognitive reinforcement program, suggesting that a multicomponent intervention may be the most ideal treatment program for children with SCD.

SUMMARY

SCD is a recent addition to psychiatric nosology. While the clinical presentation is characterized by considerable heterogeneity, SCD represents a fundamental deficit in the ability to use language pragmatics to fully engage in adaptive, reciprocal social interactions. SCD is a relatively rare condition with a prevalence in the range of 0.5%. SCD appears to have an onset in early childhood, and, for many, it likely has a chronic course extending through the lifespan. The etiology of SCD is unknown, and, to date, there are no diagnosis-specific treatments. However, symptom-directed interventions appear to show considerable promise.

FURTHER READINGS

Adams C, Gaile J, Roddam H, Baxendale J, Clitheroe L, Emsley R. Evaluation of a manualised speech and language therapy programme for children with social communication disorder: the SCIP feasibility study. *Pilot Feasibility Stud*. 2020;6:137.

*Adams C, Lockton E, Freed J, et al. The Social Communication Intervention Project: a randomized controlled trial of the effectiveness of speech and language therapy for school-age children who have pragmatic and social communication problems with or without autism spectrum disorder. *Int J Lang Commun Disord*. 2012;47(3):233–244.

Amoretti MC, Lalumera E, Serpico D. The DSM-5 introduction of the social (pragmatic) communication disorder as a new mental disorder: a philosophical review. *Hist Philos Life Sci*. 2021;43(4):108.

Baron-Cohen S, Leslie AM, Frith U. Does the autistic child have a “theory of mind”? *Cognition*. 1985;21(1):37–46.

Ben-Yizhak N, Yirmiya N, Seidman I, Alon R, Lord C, Sigman M. Pragmatic language and school related linguistic abilities in siblings of children with autism. *J Autism Dev Disord*. 2011;41(6):750–760.

Bishop DV; Courtenay Frazier Norbury. Exploring the borderlands of autistic disorder and specific language impairment: a study using standardised diagnostic instruments. *J Child Psychol Psychiatry*. 2002;43(7):917–929.

Campbell DB, Warren D, Sutcliffe JS, Lee EB, Levitt P. Association of MET with social and communication phenotypes in individuals with autism spectrum disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2010;153(2):438–446.

Demily C, Rossi M, Chesnoy-Servanin G, et al. Complex phenotype with social communication disorder caused by mosaic supernumerary ring chromosome 19p. *BMC Med Genet*. 2014;15(1):132.

Ellis Weismer S, Rubenstein E, Wiggins L, Durkin MS. A preliminary epidemiologic study of social (pragmatic) communication disorder relative to autism spectrum disorder and developmental disability without social communication deficits. *J Autism Dev Disord*. 2021;51(8):2686–2696.

Fuller EA, Kaiser AP. The effects of early intervention on social communication outcomes for children with autism spectrum disorder: a meta-analysis. *J Autism Dev Disord*. 2020;50(5):1683–1700.

Gerber S, Brice A, Capone N, Fujiki M, Timler G. Language use in social interactions of school-age children with language impairments: an evidence-based systematic review of treatment. *Lang Speech Hear Serv Sch*. 2012;43(2):235–249.

Gibson J, Adams C, Lockton E, Green J. Social communication disorder outside autism? A diagnostic classification approach to delineating pragmatic language impairment, high functioning autism and specific language impairment. *J Child Psychol Psychiatry*. 2013;54(11):1186–1197.

Gilmour J, Hill B, Place M, Skuse DH. Social communication deficits in conduct disorder: a clinical and community survey. *J Child Psychol Psychiatry*. 2004;45(5):967–978.

Jung HK, Kim JM, Park E. Social cognitive reinforcement program for children with social communication disorders. *Korean J Child Stud*. 2018;39:45–66.

Kaba D, Soykan Aysev A. Evaluation of autism spectrum disorder in early childhood according to the DSM-5 diagnostic criteria. *Turk Psikiyatri Derg*. 2020;31(2):106–112.

Kim YS, Fombonne E, Koh YJ, Kim SJ, Cheon KA, Leventhal BL. A comparison of DSM-IV pervasive developmental disorder and DSM-5 autism spectrum disorder prevalence in an epidemiologic sample. *J Am Acad Child Adolesc Psychiatry*. 2014;53(5):500–508.

Lord C, Bishop SL. Recent advances in autism research as reflected in DSM-5 criteria for autism spectrum disorder. *Annu Rev Clin Psychol*. 2015;11:53–70.

Mandy W, Charman T, Gilmour J, Skuse D. Toward specifying pervasive developmental disorder-not otherwise specified. *Autism Res*. 2011;4:121–131.

Mandy W, Wang A, Lee I, Skuse D. Evaluating social (pragmatic) communication disorder. *J Child Psychol Psychiatry*. 2017;58(10):1166–1175.

*Martin I, McDonald S. Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain Lang*. 2003;85(3):451–466.

Martin J, Hamshere ML, Stergiakouli E, O’Donovan MC, Thapar A. Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. *Biol Psychiatry*. 2014;76(8):664–671.

Miller M, Young GS, Hutman T, Johnson S, Schwichtenberg AJ, Ozonoff S. Early pragmatic language difficulties in siblings of children with autism: implications for DSM-5 social communication disorder? *J Child Psychol Psychiatry*. 2015;56(7):774–781.

Morales-Hidalgo P, Hernández-Martínez C, Voltas N, Canals J. EDUTEA: a DSM-5 teacher screening questionnaire for autism spectrum disorder and social pragmatic communication disorder. *Int J Clin Health Psychol*. 2017;17(3):269–281.

Murphy SM, Faulkner DM, Farley LR. The behaviour of young children with social communication disorders during dyadic interaction with peers. *J Abnorm Child Psychol*. 2014;42(2):277–289.

*Norbury CF. Practitioner review: social (pragmatic) communication disorder conceptualization, evidence and clinical implications. *J Child Psychol Psychiatry*. 2014;55(3):204–216.

*Swineford LB, Thurm A, Baird G, Wetherby AM, Swedo S. Social (pragmatic) communication disorder: a research review of this new DSM-5 diagnostic category. *J Neurodev Disord*. 2014;6(1):41.

Topal Z, Demir Samurcu N, Taskiran S, Tufan AE, Semerci B. Social communication disorder: a narrative review on current insights. *Neuropsychiatr Dis Treat*. 2018;14:2039–2046.

Troia GA. How might pragmatic language skills affect the written expression of students with language learning disabilities? *Top Lang Disord*. 2011;31(1):40–53.

Vanegas SB, Davidson D. Investigating distinct and related contributions of weak central coherence, executive dysfunction, and systemizing theories to the cognitive profiles of children with autism spectrum disorders and typically developing children. *Res Autism Spectr Disord*. 2015;11:77–92.

Velleman SL, Mervis CB. Children with 7q11.23 duplication syndrome: speech, language, cognitive, and behavioral characteristics and their implications for intervention. *Perspect Lang Learn Educ*. 2011;18:108–116.

Yuan H, Dollaghan C. Applying item response theory modeling to identify social (pragmatic) communication disorder. *J Speech Lang Hear Res*. 2020;63(6):1916–1932.

42.3 Speech Sound Disorder

EMIKO KOYAMA, PH.D., MAHLIQA ASHRAF, H.B.SC., TUANA KANT, M.SC., AND JAMES L. KENNEDY, M.D., F.R.C.P.C., F.R.S.C.

Speech sound disorder is the most common communication disorder in childhood. Children with speech sound disorder do not correctly produce the speech sounds that are expected for their age, developmental stage, and dialect. Sounds may be omitted (e.g., saying *hou* for *house*), substituted for other sounds (e.g., saying *yewo* for *yellow*), or distorted because they are produced in an atypical way (e.g., saying /sh/ with too much air escaping from the sides of the mouth). Often, the speech sound errors occur in systematic patterns. For example, consonants that should involve a steady airflow instead may have an interrupted airflow (e.g., *bacuum* for *vacuum*), or consonants that should appear at the end of words may instead be omitted (e.g., *be* for *bed*). In severe cases of speech sound disorder, speech may be completely incomprehensible to others.

Linguistic and motor learning, as well as their integration and coordination, are essential for speech-sound development. Linguistic learning presumably involves the gradual and implicit acquisition of a complex rule system that underlies the representation, perception, and production of speech. Children acquire knowledge about the individual speech sounds that comprise a particular language and their permissible contrasts and combinations. Children also induce the regularities governing syllable and word stress and the intonation patterns (melody) of speech.

Speech sound development also involves considerable motor learning. Speech production entails the quick, accurate, and coordinated actions of the many muscles that support breathing, voicing, and articulation of speech sounds. Major articulators that shape speech sounds include the tongue, lips, upper teeth, the gum behind the upper teeth (alveolar ridge), the hard and soft palates, the uvula, the pharynx, and the glottis (the opening between the vocal cords). Eventually, most adults learn to control the complex coordination of these fine motor activities well enough to enable the production of intelligible speech at rates as high as 250 words per minute.

Children with speech sound disorder generally develop speech sounds in a typical sequence but with a delayed onset and slow rate of growth. Early words generally include consonant sounds designated as the *early eight* (/m/, /b/, /y/, /n/, /w/, /d/, /p/, and /h/), which are also commonly used during the babbling stage from about 6 months of age. As development continues, the *middle eight* consonant sounds (/t/, /ng/, /k/, /g/, /f/, /v/, /ch/, and /j/) are mastered gradually, followed by the *late eight* (/sh/, /th/ [two variants], /s/, /z/, /l/, /r/, and /zh/). Children may not achieve full mastery of the late eight consonant sounds until 7 or 8 years of age. Throughout the course of typical development, children may produce speech sounds incorrectly and inconsistently.

Word formation normally proceeds in a systematic fashion. The first words that children use generally produce consist of simple sequences of consonants (C) and vowels (V) that comprise syllables of the forms CV (e.g., bye), VC (e.g., up), or CVC (e.g., dog) and short repetitive sequences of the form CVCV (e.g., dada). More complex syllable and word shapes are gradually mastered, such as those with multiple, nonrepetitive syllables and consonant cluster sequences, such as /thr-/, /kl-/, /-rst/, or /-nt/. Vowels are generally thought to be acquired early, and most studies suggest that vowel production is relatively accurate by 2 years of age. As such, numerous vowel errors made beyond age 2 years may be early warning signs of impaired speech.

Speech sound errors may also be described in terms of phonologic processes or rules that encompass a pattern of errors. For example, young children often use a phonologic process of final consonant deletion in which they regularly omit consonants at the end of words. This deletion process is often used by young children but may usually disappear from their speech by 3 years of age. Similarly, there are other common processes that gradually disappear during normal development, including reduplication (repetition of a syllable to form a multisyllabic word, e.g., *bahbah* for *bottle*), unstressed syllable deletion (omission of a syllable that is not stressed, e.g., *nana* for *banana*), stopping (use of an interrupted airflow for sounds that should have a continuous airflow, e.g., *tee* for *see*), consonant harmony (e.g., using a sound similar to one that occurs elsewhere in a word, e.g., *tat* for *cat*), and consonant cluster reduction (omitting or simplifying sounds in a consonant cluster, e.g., *top* for *stop*).

These phonologic processes all occur regularly in the speech of young children, but all are usually overcome before 5 years of age.

Even if a child mispronounces words, it may not affect the overall comprehension of the child's utterances. If a child's spoken word or phrase cannot be understood, the child's speech is said to be *unintelligible*. Intelligibility is influenced not only by the produced speech sounds but also by other factors, such as pragmatic context and the complexity of the sentence. In normal speech development, roughly half of children's speech is considered intelligible at about 2 years of age, roughly 75% is intelligible by 3 years of age, and by age 4 years, nearly all of the child's spoken language becomes intelligible. If less than two-thirds of the child's utterances are intelligible by age 4, clinical attention may be warranted.

COMPARATIVE NOSOLOGY

Speech sound disorder in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders, Text Revision* (DSM-5-TR) largely corresponds to phonological disorder in the revised fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR). The designation of *speech sound disorder* is typically reserved for speech sound problems with no known cause, which are thought to reflect difficulties in learning the regularities of phonology (the sound system of a language). Speech sound disorder may also be designated when causes are known (e.g., hearing impairment), but the speech impairment exceeds what is expected given the known cause. The 10th edition of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) provides the category of *specific speech articulation disorder* for speech problems of unknown etiology and designates a different category for speech impairment arising from sensory, structural, or neurologic conditions. The 11th edition of the ICD includes *developmental speech sound disorder*, which has similar criteria to DSM. In the research literature, terms such as (developmental) phonological disorder, (functional) articulation disorder, developmental verbal dyspraxia, and childhood apraxia of speech are commonly used in association with speech sound disorder.

EPIDEMIOLOGY

The prevalence of speech sound disorder is around 5% to 10% in preschoolers, 3% in 6- to 7-year-old children, and 0.5% in 17-year-old children. Estimates, however, vary widely in the research literature, reflecting the diverse criteria and samples that have been used across studies. Slightly more than half show problems only in speech sound production; the remainder also have language difficulties. Overall, speech sound disorder occurs far more frequently than speech impairment with known structural or neurologic origins.

The literature consistently shows that speech sound disorder occurs more often in boys. Estimates for sex ratios are as high as 4:1 and as low as 1.5:1. Speech sound disorder also shows high rates of familial aggregation. Approximately 20% to 40% of children with speech sound disorder have at least one immediate family member who is also affected. There is evidence in favor of genetic contributions to familial aggregation, but environmental factors may also play a role. Familial risk factors such as low maternal education have also been identified as potential risk factors.

ETIOLOGY

The majority of children with speech sound disorder has no known cause. Proposals regarding causation typically focus on many of the same factors considered as possible causes for language disorders, including subtle brain abnormalities, genetic influences, environmental factors, neurodevelopmental immaturity, linguistic factors, and auditory processes. In addition, speech motor control and speech mechanism factors have also been entertained as possible causes.

Speech sound disorder most likely has multiple causal determinants. Although some evidence supports each proposed determinant, no single one is implicated in all cases of speech sound disorder. The idea that brain abnormalities may be responsible for speech sound disorders is supported primarily by indirect evidence, which includes reports of associated soft neurologic signs at higher than normal rates and findings of structural or functional brain abnormalities in children with related disorders, such as language disorder and reading disorder. The evidence in support of genetic factors comes from a variety of methods, including pedigree, twin, adoption, and linkage studies. Recent studies have found susceptibility of speech sound disorder genes in chromosomes 3, 6, and

15, which are also thought to be candidate regions for dyslexia. One of the genes that have been extensively associated with speech problems is the *FOXP2* gene on chromosome 7. Children with changes or mutations on the *FOXP2* gene may experience difficulties producing sequences of speech sounds and may also experience delays in language, social, and motor development. The prevalence of *FOXP2*-related speech problems is estimated to be relatively rare, affecting 1 to 2 in 1,000 people. The advancement of work on genetic influences will probably stimulate corresponding advances in the understanding of environmental determinants. At present, however, no conclusive evidence links specific environmental factors to speech sound disorders, although it is reasonable to assume that such factors contribute in as-yet unspecified ways.

Neurodevelopmental immaturity is suggested by the fact that children with speech sound disorder usually follow the same sequence in speech sound development as children without the disorder, albeit at a slower rate. Additional indirect support comes from findings of soft neurologic signs and subtle deficits in motor skills. Children with speech sound disorders may also have linguistic deficits that interfere with efficient learning of the regularities of phonology (e.g., permissible sounds, combinations, or positions of occurrence) from the available language input. Difficulties with auditory discrimination and perception or temporary hearing loss may also contribute to some cases of speech sound disorder. Children with speech sound disorder also show elevated rates of subtle abnormalities in the structure and function of the speech mechanism, but it is not yet known whether these abnormalities are causes or just correlates of the disorder.

DIAGNOSIS AND CLINICAL FEATURES

A diagnosis of speech sound disorder is given when a person experiences persistent difficulties in speech sound production due to different degrees of impairment in phonologic knowledge and coordination of movements involved in speech. The onset of speech sound problems is during childhood. The speech sound production is impaired for what is expected based on the speaker’s age and dialect and cannot be attributed to medical or neurologic conditions such as cerebral palsy, cleft palate, traumatic brain injury, or hearing impairment. A further criterion is that the speech sound difficulties must interfere with social communication, academic achievement, or occupational success. Speech sound disorder ranges in severity from mild, in which errors occur on one or two sounds and speech intelligibility is hardly affected, to severe, in which errors occur on most sounds and speech intelligibility is significantly impaired.

A delayed onset and slow rate of speech sound development are typical of most children with speech sound disorders. However, the course and sequence of development usually resemble those of typically developing children. Simple syllable and word shapes are gradually elaborated to include more complex shapes. The rate of acquisition may be slow but usually follow the typical sequence, with the early eight consonant sounds usually developing first, followed by the middle eight, and then the late eight consonant sounds. The errors in speech sound production resemble those that occur in normal development, but the relative distribution of omissions, substitutions, and distortions may be somewhat different. Children with speech sound disorder generally use phonologic processes that also occur in normal development but may use these processes well beyond the age at which they normally disappear. Children with speech sound disorder also sometimes use unusual or idiosyncratic phonologic processes that are not observed in normal development.

Ramon was a talkative, likable 3 year old with virtually unintelligible speech, despite his normal hearing and language comprehension skills. Ramon's level of expressive language development could not be determined because his speech was so difficult to understand. The rhythm and melody of his speech, however, suggested that he was trying to produce multiword utterances, as would be expected at his age. Ramon produced only a few vowels (/ee/, /ah/, and /oo/), some early developing consonants (/m/, /n/, /d/, /t/, /p/, /b/, /h/, and /w/), and limited syllable shapes (V, CV, and CVCV). This reduced sound repertoire made many of his spoken words indistinguishable from one another (e.g., he said *bahbah* for *bottle*, *baby*, and *bubble*, and he used *nee* for *knee*, *need*, and *Anita* [his sister]). Moreover, he consistently omitted consonant sounds at the end of words and in consonant cluster sequences (e.g., /tr-/, /st-/, /-nt/, and /-mp/). On occasion, Ramon reacted with frustration and tantrums to his difficulties in making his needs understood.

Kent was a pleasant, cooperative 5-year-old boy who had been diagnosed with a speech sound disorder when he was a preschooler. His hearing and language comprehension skills were within normal limits. However, he showed some mild expressive language problems in the use of certain grammatical features (e.g., pronouns, auxiliary verbs, and past-tense word endings) and in the formulation of complex sentences. He correctly produced all vowel sounds and most of the early-developing consonants but was inconsistent in his attempts to produce later-developing consonants (e.g., /r/, /l/, /s/, /z/, /sh/, /th/, and /ch/). Sometimes, he omitted them; sometimes, he substituted other sounds for them (e.g., /w/ for /r/ or /f/ for /th/); occasionally, he even produced them correctly. Kent had particular problems in correctly producing consonant cluster sequences and multisyllabic words. Cluster sequences had omitted or incorrect sounds (e.g., *blue* might be produced as *bue* or *bwue*, and *hearts* might be said as *hots* or *hars*). Multisyllabic words had syllables omitted (e.g., *efant* for *elephant* and *getti* for *spaghetti*) and sounds mispronounced or even transposed (e.g., *aminal* for *animal* and *lemon* for *melon*). Strangers were unable to understand approximately 80% of Kent's speech. Fortunately, Kent often spoke more slowly and clearly than usual when he was asked to repeat something, as he often was.

Natasha was a shy, reserved 8-year-old girl with a history of significant speech delay. During her preschool and early school years, she overcame many of her earlier speech errors. A few late-developing sounds (/r/, /l/, and /th/), however, continued to pose a challenge for her. Natasha often substituted /f/ or /d/ for /th/ and produced /w/ for /r/ and /l/. Overall, her speech was easily understood despite these minor errors. Nonetheless, she was often reluctant to speak in front of others because of the teasing she received from her classmates about her speech.

Other language, learning, and developmental disorders may co-occur with speech sound disorder. Language disorder, especially deficits in expressive language, frequently co-occurs with speech sound disorder. Children who have both speech sound disorder and language disorder are at increased risk for other problems, such as reading, spelling, and other learning problems, attention deficit hyperactivity disorder (ADHD), and developmental coordination disorder. Children with isolated speech sound disorders are at a relatively small risk for severe academic, behavioral, or psychiatric disorders. If a speech sound disorder persists into adulthood, it may contribute to social anxiety disorder.

PATHOLOGY AND LABORATORY EXAMINATION

There are no laboratory tests that can provide a diagnosis of speech sound disorder. When necessary, specific laboratory tests may help identify and document an underlying neurologic condition that is responsible for apraxia or dysarthria.

DIFFERENTIAL DIAGNOSIS

The most important task in the diagnosis of speech sound disorder is to judge if speech sound production is inappropriate for the age and dialect of the speaker. The judgment may be challenging because the speech of children with speech sound disorder often resembles that of younger children who are developing normally. The same speech errors that may indicate disorder in a 6-year-old child may also indicate typical development in a 4-year-old child. Thus, it is critical to carefully compare a particular child’s pattern of speech sound errors to the normal patterns of speech sound development in children.

One must also carefully consider the dialect spoken by the child. Sound differences that may be quite acceptable in one dialect may suggest disorder in another. For example, the production of *caht* for *cart* is acceptable for speakers from Boston but not for those from most other parts of North America.

One of the challenges facing clinicians is how to diagnose speech sound disorders in bilingual children. Depending on such factors as the child’s first language background and the timing of acquisition of the second language, even typically developing bilingual children may exhibit more frequent and atypical phonologic error patterns compared to their same-age monolingual peers. Research is needed to establish the typical and atypical speech patterns of children acquiring a second language. Meanwhile, an assessment of speech sound production skills in English and the child’s other language, accompanied by a careful history, may lead to a correct diagnosis.

The diagnostic assessment should also include a hearing test and an oral examination. The hearing test identifies temporary or permanent hearing difficulties that may require medical or other interventions to enhance the prognosis for the development of understandable speech. The results from the oral examination may help clarify whether the speech sound production problems are developmental, structural, or neurologic in origin. For example, children with neurologic conditions may be slow or inaccurate in performing various speech and nonspeech movements (e.g., sequencing syllables, as in *putuku*; protruding the tongue; or rounding the lips) or may show signs of muscle weakness or paralysis. Drooling, feeding, or swallowing problems suggestive of dysarthria, may also be observed or reported in the history. During the oral examination, the diagnostician may also identify structural abnormalities in the mouth, such as tumors or submucous clefts of the palate, which may need to be considered in the overall management of the child.

Selective mutism is an anxiety disorder that is characterized by the consistent failure to speak in specific social settings. Children with speech sound disorder may not speak due to embarrassment about their speech impairments. However, children with selective mutism typically exhibit normal speech in settings where they feel “safe,” such as with family or close friends.

Screening for possible co-occurring communication disorders is also important for children diagnosed with speech sound disorders. When an intellectual developmental disorder is suspected, a cognitive assessment may help confirm or rule out this diagnosis. In cases in which speech difficulties are in excess of those expected for a particular degree of intellectual disability, a diagnosis of co-occurring speech sound disorder may be made.

A practitioner should make a referral to a speech–language pathologist for assessment of speech sound production skills when (1) a 2 year old uses little or no speech, (2) a 3 year old is poorly understood by familiar members of the family, (3) a 4 year old is poorly understood by peers and strangers, (4) a child of any age seems to be withdrawing from social interactions as a result of teasing or frustration associated with speech difficulties, (5) a parent or teacher is concerned about a child’s speech, or (6) a child (or adult) experiences a sudden deterioration in speech production skills. The last may be the first warning sign of a developing neurologic condition and, therefore, also warrants referral for neurologic examination.

COURSE AND PROGNOSIS

Diagnosis may occur as early as 2 years of age in severe cases of speech sound disorder, which may be accompanied by co-occurring language involvement. However, diagnosis around 4 years of age is more usual. Speech sound disorders show considerable variability in severity, course, and outcome. In approximately 75% of mild-to-moderate cases, speech sound production skills normalize by 6 years of age, with or without intervention. Thus far, no specific predictors of normalization have been identified among the various speech, hearing, oral, cognitive, linguistic, or psychosocial variables that have been investigated. For children whose speech has not normalized by 6 years of age, speech sound development proceeds in spurts and plateaus until 8 or 9 years of age. If normalized speech has not been achieved by then, residual articulation problems often persist into adulthood unless successfully treated. These residual errors usually involve distortions of later-developing sounds, such as /l/, /r/, /s/, /z/, /sh/, /ch/, and /th/. Specifically, distortions of the sounds /s/, /z/, /sh/, and /ch/ are termed *lisps*. Residual errors often have little effect on speech intelligibility but may influence others to erroneously perceive a speaker as lazy, immature, or of low intelligence.

Children with mild-to-moderate speech sound disorders are at relatively little risk for adverse long-term outcomes. Although speech typically normalizes by 8 to 9 years of age, subtle deficits in the speed and accuracy of articulation during complex speech tasks (e.g., rapid pronunciation of multisyllabic words) may remain detectable into adulthood but have little functional impact. Overall, language, cognitive, academic, psychosocial, and psychiatric outcomes are usually within the normal range for adults with a history of mild-to-moderate speech sound disorder.

Children with severe speech sound disorders that persist to school age are at a greater risk for poor outcomes. They may not develop adequate phonologic awareness skills (perception, identification, and mental manipulation of speech sounds), which, in turn, may lead to problems in learning to read and spell. Parents and teachers should carefully monitor the reading and spelling development of school-aged children with severe, persistent speech sound problems. If required, reading and spelling performance can then be fostered through explicit instruction in phonologic awareness. It is not yet clear whether severe speech sound disorders are associated with an increased risk for adverse psychosocial or psychiatric outcomes.

Children with co-occurring language disorders are, however, known to have elevated risk for adverse long-term outcomes in communicative, academic, psychosocial, or psychiatric functioning. Those with receptive difficulties are at the greatest risk.

TREATMENT

Treatment is typically recommended for children with moderate-to-severe speech sound disorders. A speech–language pathologist may provide the treatment directly to the child or indirectly through paraprofessionals, parents, teachers, peers, or other volunteers who are taught how to provide treatment for the child. Although there are many intervention techniques, styles, and programs that have been advocated, two main approaches to treatment are discernible.

The *traditional approach* is typically used with children who produce substitution or distortion errors for just a few sounds. In this approach, the child participates in activities that highlight the auditory and motor characteristics of a particular sound. The child then practices the production of the sound while the clinician provides feedback and cues concerning the correct placement and movement of the articulators. Treatment is discontinued once the child produces the new sound accurately in spontaneous speech.

A *phonologic approach* is usually chosen for children with patterns of multiple speech sound errors that are described by phonologic processes (e.g., final consonant deletion or consonant cluster reduction). In this approach, the child participates in activities that demonstrate the communicative value of correctly using certain classes of sounds, such as final consonants or consonant clusters. The child then practices the production of various final consonants or clusters in meaningful communicative situations until the targeted groups of sounds are incorporated into the child's phonologic system. Treatment is discontinued once the child uses the new classes of sounds correctly and consistently in everyday speech.

Traditional and phonologic approaches have proven effective in accelerating the rate of speech sound development. However, not all variations of these approaches have yet been subjected to rigorous investigations of their efficacy. Additional research is also needed to compare the relative merits of various techniques for children with different profiles of speech errors, although some work of this nature has begun to appear in the literature. Overall, treatment success depends on whether the new speech skills become established as automatic patterns of production. A key factor in achieving this goal may be the extent to which a child or adolescent is motivated to change existing speech patterns.

In addition to speech sound intervention, children with co-occurring language disorder may need interventions to promote language or fluency. Few studies have systematically evaluated the efficacy of such joint interventions. It is possible that speech production gains may occur as an additional benefit of language intervention that did not focus specifically on speech sound production.

ACKNOWLEDGMENTS

This chapter was written in memory of the late Dr. Joseph H. Beitchman, who continues to inspire us with his dedication, devotion, and extensive research work in childhood language disorders.

Further Readings

Baker E, McLeod S. Evidence-based practice for children with speech sound disorders: part 1 narrative review. *Lang Speech Hear Serv Sch*. 2011;42(2):102–139.

Bauman-Waengler J. *Articulatory and Phonological Impairments: A Clinical Focus*. 2nd ed. Allyn & Bacon; 2004.

Beitchman JH, Brownlie EB, Bao L. Age 31 mental health outcomes of childhood language and speech disorders. *J Am Acad Child Adolesc Psychiatry*. 2014;53(10):1102–1110.

Bowen C. *Children's Speech Sound Disorders*. 2nd ed. Wiley-Blackwell; 2014.

Campbell TF, Dollaghan CA, Rockette HE, et al. Risk factors for speech delay of unknown origin in 3-year-old children. *Child Dev*. 2003;74:346–357.

Dodd B, Bradford A. A comparison of three therapy methods for children with different types of developmental phonological disorder. *Int J Lang Commun Disord*. 2000;35:189–209.

Eadie P, Morgan A, Ukoumunne OC, Ttofari Eecen K, Wake M, Reilly S. Speech sound disorder at 4 years: prevalence, comorbidities, and predictors in a community cohort of children. *Dev Med Child Neurol*. 2015;57:578–584.

Felsenfeld S. Finding susceptibility genes for developmental disorders of speech: the long and winding road. *J Commun Disord*. 2002;35:329–345.

Gierut JA. Treatment efficacy: functional phonological disorders in children. *J Speech Lang Hear Res*. 1998;41:S85–S100.

Hesketh A, Adams C, Nightingale C, Hall R. Phonological awareness therapy and articulatory training approaches for children with phonological disorders: a comparative outcome study. *Int J Lang Commun Disord*. 2000;35:337–354.

Johnson CJ. Getting started in evidence-based practice for childhood speech-language disorders. *Am J Speech Lang Pathol*. 2006;15:20–35.

Johnson CA, Weston AD, Bain BA. An objective and time-efficient method for determining severity of childhood speech delay. *Am J Speech Lang Pathol*. 2004;13:55–65.

Kent RD. Research on speech motor control and its disorders: a review and prospective. *J Commun Disord*. 2000;33(5):391–427.

Law J, Garrett Z, Nye C. Speech and language therapy interventions for children with primary speech and language delay or disorder. *Cochrane Database Syst Rev*. 2003;CD004110.

Lewis BA, Freebairn LA, Hansen A, Taylor HG, Iyengar S, Shriberg LD. Family pedigrees of children with suspected childhood apraxia of speech. *J Commun Disord*. 2004;37:157–175.

MacDermot KD, Bonora E, Sykes N, et al. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am J Hum Genet*. 2005;76:1074–1080.

Morrisette ML, Gierut JA. Lexical organization and phonological change in treatment. *J Speech Lang Hear Res*. 2002;45:143–159.

Munson B, Bjorum EM, Windsor J. Acoustic and perceptual correlates of stress in nonwords produced by children with suspected developmental apraxia of speech and children with phonological disorder. *J Speech Lang Hear Res*. 2003;46:189–202.

Namisivayam AK, Coleman D, O'Dwyer A, van Lieshout P. Speech sound disorders in children: an articulatory phonology perspective. *Front Psychol*. 2020;10:2998

Rice ML, Smith SD, Gayán, J. Convergent genetic linkage and associations to language, speech, and reading measures in families of probands with specific language impairment. *J Neurodev Disord*. 2009;1(4):264–282.

Rvachew S, Chiang PY, Evans N. Characteristics of speech errors produced by children with and without delayed phonological awareness skills. *Lang Speech Hear Serv Sch*. 2007;38(1):60–71.

Rvachew S, Nowak M. The effect of target-selection strategy on phonological learning. *J Speech Lang Hear Res*. 2001;44:610–623.

Schick JH, Kundtz AM, Taylor HG, et al. Evidence of linkage for speech impairment with chromosome 7 markers. *Am Soc Human Gen*. 2001.

Shriberg LD, Austin D, Lewis BA, McSweeny JL, Wilson DL. The speech disorders classification system (SDCS): extensions and lifespan reference data. *J Speech Lang Hear Res*. 1997;40:723–740.

Smith SD. Genes, language development, and language disorders. *Ment Retard Dev Disabil Res Rev*. 2007;13:96–105.

Tyler AA, Lewis KE, Haskill A, Tolbert LC. Efficacy and cross-domain effects of a morphosyntax and a phonology intervention. *Lang Speech Hear Serv Sch*. 2002;33:52–66.

Tyler AA, Lewis KE, Welch CM. Predictors of phonological change following intervention. *Am J Speech Lang Pathol*. 2003;12:289–298.

Verdon S, McLeod S, Wong S. Reconceptualizing practice with multilingual children with speech sound disorders: people, practicalities and policy. *Int J Lang Commun Disord*. 2015;50(1):48–62.

Wren Y, Miller LL, Peters TJ, Emond A, Roulstone S. Prevalence and predictors of persistent speech sound disorder at eight years old: findings from a population cohort study. *J Speech Lang Hear Res*, 2016;59(4):647–673.

Yavas M, Goldstein B. Phonological assessment and treatment of bilingual speakers. *Am J Speech Lang Pathol*. 1998;7:49–60.

42.4 Stuttering

EMIKO KOYAMA, PH.D., MAHLIQA ASHRAF, H.B.SC., JAMES L. KENNEDY, M.D., F.R.C.P.C., F.R.S.C., AND TUANA KANT, M.SC.

Stuttering is a condition in which the flow of speech or fluency is disrupted by involuntary speech–motor events. Stuttering may include a variety of specific dysfluency types, such as sound and syllable repetitions, sound prolongations, dysrhythmic phonations, and complete blocking or unusual pauses between sounds and syllables of words. More severe forms of stuttering are frequently accompanied by a variety of struggle behaviors, commonly referred to as *accessory*, or *secondary*, characteristics of the disorder. These struggle behaviors may be observed at the respiratory, phonatory, or articulatory level of the speech mechanism and may manifest themselves in disordered breathing, glottal fry (abnormal voice quality), lip pursing, and tongue clicking. Additional behaviors, such as eye blinks, facial grimacing, head jerks, and abnormal body movements, may be observed before or during instances of disrupted speech. In addition to the core and secondary physical symptoms of stuttering, the intended communication is often further altered or distorted by linguistic avoidance due to anticipatory sound or word fears. Linguistic avoidance is a commonly observed feature of stuttering, which results in word substitution and circumlocution as the individual approaches the feared sound or word during the communicative attempt. Furthermore, it has been reported that varying levels of both utterance length and temporal complexity function to modulate stuttering frequency in adult stuttering speakers. Similarly, situational avoidance is often characteristic of chronic stuttering, and the type of communicative situation often influences the amount and severity of stuttering. Stuttering patterns are extremely variable, ranging from mild, occasional dysfluencies in speech to severe stuttering that renders communication all but impossible for the affected individual and often results in serious psychological difficulty and social withdrawal. Moreover, the person who stutters may exhibit widely fluctuating periods and severity of dysfluency within the course of a day, depending on the communicative situation, anxiety level, fatigue level, and other associated variables, making stuttering a multifactorial disorder. It is to be noted that stuttering severity is not solely a measure of the frequency of observed stuttering but is usually a function of the communicative hardship it places on the speaker. In other words, an individual employed in an office setting who exhibits a stuttering rate of 15% may feel less handicapped than an airline pilot who is 4% dysfluent, but whose fluent communication is critical. Most recently, treatment efficacy studies of stuttering have begun to utilize measures of functional communication and quality of life to supplement traditional data gathered from counts of stuttering frequency.

HISTORY AND COMPARATIVE NOSOLOGY

Psychiatric syndromes can be defined from the point of view of the observer or the patient. Depressed individuals may not look depressed, but if they report subjective symptoms sufficient to meet criteria, a diagnosis of a depressive disorder can be made. Likewise, stuttering can be defined from the point of view of the listener or the speaker. The speaker’s sense of subjective distress or loss of control of their speech apparatus reflects an approach to nosology that focuses on the speaker’s experience. The *Diagnostic and Statistical Manual of Mental Disorders* (DSM) focuses on the observer’s point of view—that is, on the observable behaviors that are impinging on the patient’s quality of life.

In both DSM-III and DSM-III-R, stuttering was placed in the category of disorders usually first evident in infancy, childhood, or adolescence. In both of these systems, stuttering was coded on axis I, along with clinical psychiatric syndromes. This remains controversial as stuttering is not commonly viewed as a psychiatric disorder among speech–language professionals, and speech–language pathologists have argued that stuttering should not be included in DSM at all.

In DSM-IV-TR, stuttering is categorized as a communication disorder. This is consistent with the speech–language pathology classification, in which all speech and language disorders are kept within the same category. However, it remains to be shown that it belongs in this category; aside from the similarity of causing communicative difficulties, stuttering may pathologically differ from the other communication disorders in this category. The 10th revision of the *International Statistical Classification of Disease and Related Health Problems* (ICD-10) separates stuttering (stammering) from the speech and language disorders and places it in the category of other behavioral and emotional disorders with onset occurring in childhood and adolescence. The diagnostic criteria in ICD-10 are similar to the DSM-IV-TR diagnostic criteria, but a few differences are notable such as the removal of age-dependent criteria in DSM-IV-TR.

DSM-IV-TR requires that the disorder interferes with academic or occupational achievement or social communication for a diagnosis, whereas the ICD-10 diagnosis applies only if it is severe enough to cause notable disturbances in fluency. ICD-10 requires a minimum duration of 3 months of consistent symptomology; DSM-IV-TR has no duration criterion. DSM-IV-TR further specifies that the disturbance must exceed what is appropriate for the individual’s age, and, if a speech–motor or sensory deficit is present, the speech difficulties must exceed those usually associated with these problems. ICD-10 has an additional category for cluttering, which is not accorded separate status in DSM-IV-TR.

In DSM-5 and its text revision (DSM-5-TR), stuttering is rephrased as childhood-onset fluency disorder (stuttering) and categorized within the communication disorders. Stuttering is defined by regular uncontrolled disturbances in the normal fluency and patterning of speech that are inappropriate for the individual’s age and language skills, persist over time, and are characterized by frequent and marked occurrences of one (or more) of the following: phoneme and syllable repetition, with prolonged vowels and consonants, aberrant interjections and pauses between words and syllables, abrupt pauses in speech, linguistic avoidance of words that triggers stuttering, words produced with excessive physical strain, and monosyllabic repetitions (e.g., “I-I-I see him”).

DSM-5 and DSM-5-TR removed the use of interjectional phrases such as “um,” “ah,” and “you know” as part of the diagnostic definition of stuttering. DSM-5-TR also states the disturbance should not be attributable to a speech–motor or sensory deficit, dysfluency associated with a neurologic insult (stroke or hemorrhage), another medical condition (neurogenic stuttering), or better explained by another mental disorder.

The 11th revision of the *International Statistical Classification of Disease and Related Health Problems* (ICD-11) also contains the diagnosis of developmental speech fluency disorder, with similar criteria to the DSM-5-TR.

With onset in early developmental periods, these disturbances facilitate the comorbid development of social anxiety when speaking, which results in limitations to effective communication, social participation, or academic, or occupational performance, individually or in any combination. DSM-5 included the presence of social anxiety or situational avoidance as part of the diagnostic criteria for stuttering.

EPIDEMIOLOGY

Surveys conducted chiefly in the United States and Europe indicate that the prevalence of stuttering in the general population is approximately 1%. Although data regarding incidence vary widely, most specialists would agree that the likelihood of an individual ever exhibiting some form of stuttering is between 3% and 4%. Of those who begin to stutter, the majority, around 80%, stop by adulthood. The estimated percentages of spontaneous recovery in stuttering, again, vary widely because of differing definitions of stuttering and normal nonfluency, age groups studied, and data collection techniques. Several studies have estimated spontaneous recovery to be as high as 80%. Stuttering is more prevalent among boys than girls, with the ratio of boys to girls among U.S. schoolchildren being between 3:1 and 5:1. Current studies have challenged the above estimates of prevalence, incidence, and gender ratios. Among other factors, it has been postulated that most of the risk for stuttering onset is over by age 5. Even in its earliest stages, stuttering has been shown to be both qualitatively as well as quantitatively different from normal dysfluency.

A high familial incidence is indicated by the presence of stuttering among the relatives of approximately 50% of individuals who stutter. In addition, a high degree of concordance of stuttering exists in identical twins. Individuals who stutter appear to have normal developmental histories, except for speech and language. Numerous studies have demonstrated that young stuttering children are somewhat delayed in early language and articulation development; however, studies of older stuttering children appear to demonstrate that language skills are once again age appropriate. With regard to personality and social and emotional adjustment, children who stutter are often regarded by their parents or primary caregivers as nervous and prone to specific fears. They are often described as overly sensitive to their environment. Adults who stutter tend to be somewhat more anxious and emotionally reactive than their nonstuttering peers. For the most part, studies relating intellectual ability to stuttering are inconclusive; however, there is a significantly higher prevalence of stuttering in the developmentally delayed population.

ETIOLOGY

There has been growing agreement among researchers and clinicians that stuttering is a multifactorial disorder. Specifying the level of causation can facilitate understanding the etiology of stuttering.

The first of these levels pertains to underlying causation, which refers to the genetic, neurophysiologic, and psychological factors that may predispose a child to fluency disruption. Research in these areas has been rather cyclic since the 1940s, and various models implicating psychological, learning, and neurophysiologic variables in stuttering have been presented.

One early psychoanalytic theory of stuttering postulated that disrupted speech represents part of the individual’s psychological history, such as disturbed or traumatic early parent–child interaction. Empirical support for this psychoanalytic model is largely lacking, and studies investigating personality factors associated with stuttering have not revealed anything suggesting emotional disturbance or neurosis as a possible cause.

Later psychological models of stuttering have drawn on the principles of classical and operant conditioning to explain stuttering. Several behaviorally oriented researchers offered a model of operant conditioning in which the early dysfluencies of the child were subject to an intermittent schedule of reinforcement, resulting in incipient stuttering. The two-factor theory of stuttering posited that dysfluency resulting from negative emotions becomes associated with speech-related stimuli through a process of pairing, or conditioning, as explained by classical conditioning. The individual’s response to the fluency failure, the repetitions, sound prolongations, and attempts to terminate the instance of disruption are instrumentally conditioned.

Perhaps the best-known psychological theory of stuttering is the diagnosogenic (semantogenic) model as proposed in the 1940s. This theory postulates that stuttering is caused by parents’ diagnoses of the developing child’s normal dysfluencies as stuttering, particularly if the child is also socially sensitive. One of the stronger observations used to support the diagnosogenic theory was that of American Indian tribes of which ones lacking a specific word for stuttering in their tribal language had lower rates of stuttering in their overall population, relative to tribes that did. The underlying fundamentals of this correlation reason that if a child with normal dysfluency is subject to a premature diagnosis of stuttering, they would have a mental block preventing them from realizing their potential. However, there is an understanding that specific vocabulary is not necessary for the sake of communication, which leaves this theory under skepticism as it is still very much possible a child is reprehended and stigmatized for their normal dysfluency.

The large proportion of stuttering individuals who have relatives with a similar speech disorder and the high concordance for stuttering in identical twins have long suggested that biologic inheritance plays some part in stuttering. Current models of genetic transmission that can account for familial patterns of stuttering provide strong evidence for the biologic inheritance of stuttering. Moreover, research utilizing the most current genotyping technologies has confirmed the evidence for a strong genetic factor in stuttering. However, these genetic models point to a complex interaction between heredity and environment. Most recent studies have pointed out several possible specific candidate genes. However, it is too early to declare that specific genes underlying stuttering in the general population have, in fact, been identified. Moreover, at this point in time, it remains unclear as to which specific combinations of motor, linguistic, temperament, or other aspects of speech are being transmitted.

The theory of cerebral dominance, proposed in the late 1920s, was one of the earliest neurophysiologic models used to explain the cause of stuttering. This theory postulated that stuttering resulted from the lack of an appropriately developed hemispheric dominance of speech and language function.

The theory of incomplete or atypical cerebral lateralization for speech has received varying amounts of support, but not until relatively recently did advances in electrophysiologic techniques and brain imaging begin to yield more reliable data on cortical and subcortical processing of speech in stuttering. Much of the earlier research into the nature of these hypothesized brain processes was based largely on the use of behavioral observations and electromyographic measures. Relevant observations included the effects of delayed auditory feedback on the speech fluency of nonstuttering and stuttering speakers, differences in dichotic listening and dual-task interference performance between stuttering and nonstuttering speakers, and articulatory discoordination and oral kinesthetic deficiencies in stuttering individuals. These observations, among others, represented indirect evidence of differences in neural processing between stuttering and nonstuttering individuals.

Most recently, several brain-imaging techniques used to measure regional cerebral blood flow and oxygen consumption have been applied to the investigation of stuttering. Data from these initial studies using single-photon emission computed tomography (SPECT) and positron emission tomography (PET) point to differential activation patterns in stuttering individuals during both oral and silent reading and even immediately after treatment for stuttering. Findings from these studies have reported a general overactivation in several regions of cortical and subcortical motor networks in individuals who stutter. Recent functional imaging studies have reported increased activation in the right frontal-lateral cortex and the medial-anterior cingulate during both silent and oral reading in stuttering individuals. Moreover, the neural activation in stuttering adults is characterized by a generalized bilateral recruitment of cortical and subcortical neural resources involved in both language formulation and speech production, whereas this activation is lateralized for nonstuttering individuals.

Data coming from studies investigating the effects of behavioral treatment seem to suggest that treatment initially results in an increase in recruitment of resources, including cerebellar activity, after an intensive 3-

week treatment period, followed by a significant reduction in the overall amount of activation after 1 year. At the 1-year mark, the overall activation pattern observed in the stuttering speakers is very similar to that observed in the normally fluent speakers. This finding has been interpreted as signaling increased automaticity in the application of learned speech–motor skills.

A recent study attempted to establish whether decreased functional connectivity between speech-related cortical areas was the structural basis for stuttering. The investigation reported that stuttering subjects showed disturbed patterns of activation in the premotor and motor cortices, particularly below the regions that represent the laryngeal and tongue in the left sensorimotor cortex. Further evidence pointing to anomalous anatomical cortical structures in stuttering has been reported in a study using volumetric MRI scans. The study found that anomalous anatomical configurations of perisylvian cortical speech–language areas are more common in adults with stuttering. In addition to several other findings, the authors report that both the right and left planum temporale were significantly larger in the stuttering subjects compared to the controls, and planar asymmetry was reduced, implicating speech and language nonlaterality. The authors point out that the left planum temporale is composed of auditory association cortex, which is important in the higher-order processing of linguistic information and is structurally larger than its homologous counterpart in healthy adults, thus the resulting leftward asymmetry. Researchers have speculated that the left planum temporale may be the neuroanatomical substrate for language, and there is some evidence of a relationship between planar asymmetry and language laterality. Additional structural anomalies have been reported in recent studies examining the neuroanatomical differences in speech-related cortex between stuttering and nonstuttering individuals using voxel-based morphometry. Results revealed significant differences in localized gray and white matter densities of left and right hemisphere regions involved in auditory processing and speech production.

The past decade has seen exponential growth in the number of peer-reviewed articles investigating both functional and structural brain differences in stuttering individuals. While there appears to be much in the way of unclear or conflicting evidence, there have been some consistent findings.

First, several studies have demonstrated that people who stutter demonstrate greater activity in the right hemisphere cortex (gray matter) than nonstuttering individuals. Such overactivation is observed particularly in an area of the right frontal lobe, which is homologous to Broca's area of the left hemisphere. Structural studies have further demonstrated that white matter differences between stuttering and fluent speakers, located primarily in the right primary auditory and frontal cortices, were observed contralateral to the gray matter differences (Fig. 42.4–1). This finding implies that stuttering individuals have decreased functional connectivity between perisylvian areas responsible for hearing speech and producing verbal motor movements, to which the brain tries to overcompensate by contralaterally increasing the formation of white matter tracts. In a recent study of stuttering children, reduced gray matter volumes in the bilateral-inferior frontal gyri and left putamen were observed. These stuttering children also demonstrated increased gray matter volumes in the right ventral premotor and superior-temporal gyri relative to fluent children. These findings are interpreted as reflecting developmental factors resulting in a deficit in speech–motor learning. A number of studies have also reported that stuttering subjects show significantly increased activation in the cerebellum, which is critical for integrating sensory and motor functions early in the acquisition of fine motor skills. In addition, studies have shown decreased corpus callosum tract volume, which communicates information between the right- and left-brain hemispheres, including the finer details of language production, such as tone, humor, irony, and emotions.

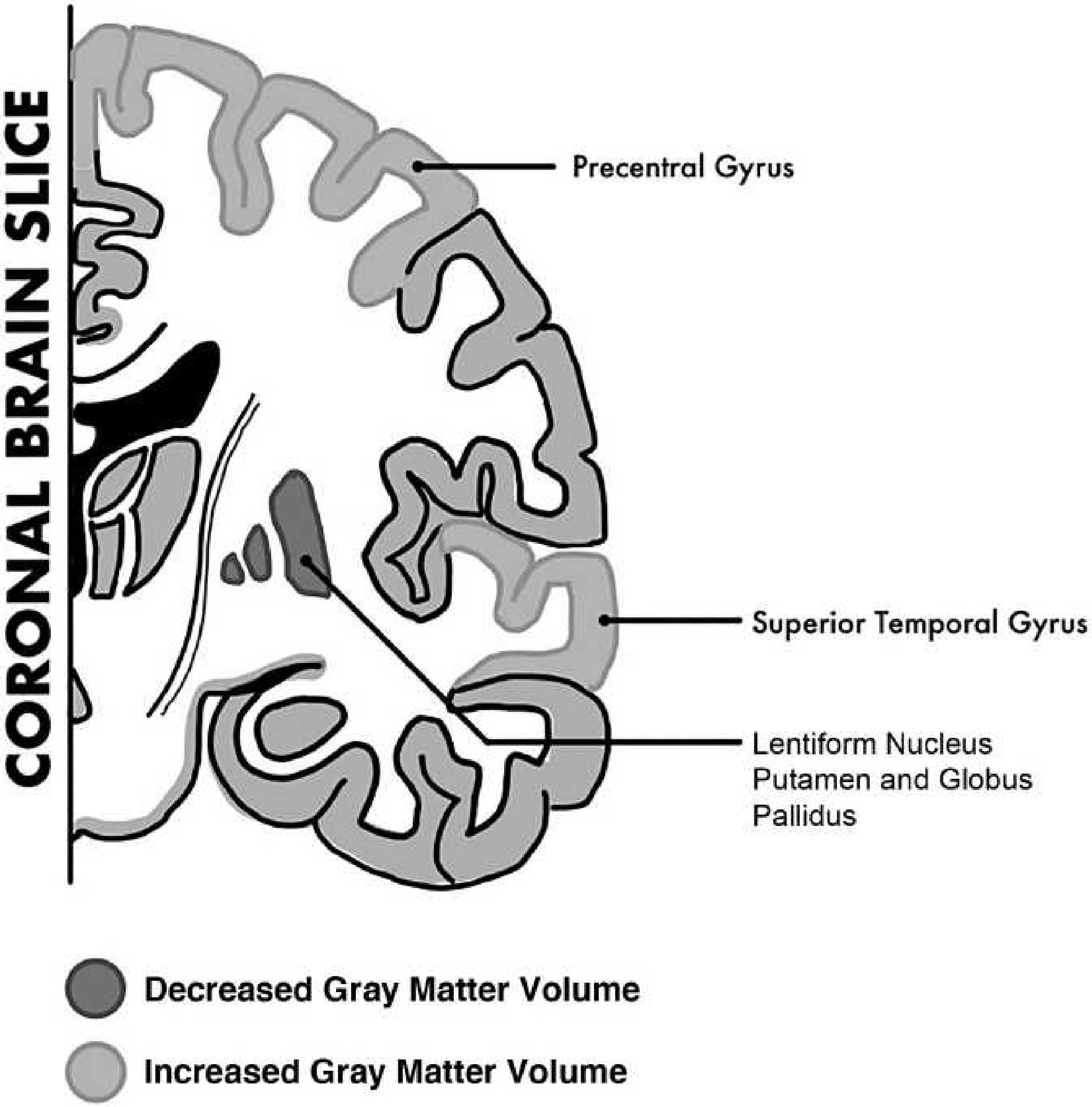
CORONAL BRAIN SLICE
Precentral Gyrus
Superior Temporal Gyrus
Lentiform Nucleus Putamen and Globus Pallidus
Decreased Gray Matter Volume
Increased Gray Matter Volume

FIGURE 42.4–1. Illustrative drawing of the brain anatomy in a coronal slice. The *highlighted regions* show the significant anatomical changes in stuttering children when compared to nonstuttering children.

A second etiologic level of stuttering involves the precipitating causes—that is, when, where, and under what conditions stuttering is most likely to occur. At this level, the concern is with the loci of stuttering and the specific speaking situations impacting fluency. Linguistic factors associated with stuttering have led some theorists to view it as a higher-level cognitive dysfunction. The vast majority of stuttering occurs on the initial sounds of words or syllables. Fluency-facilitating conditions include choral reading, repeated reading of the same material, whispering, and speaking under conditions of altered auditory feedback or masking noise. Based on these and other phenomenologic features of stuttering, new theories have implicated higher-level linguistic processes as the fundamental causal elements for speech disruption in stuttering. These processes involve the planning or adjustments of linguistic plans via internal monitoring loops. One such theory, the covert repair hypothesis, emphasizes the role of internal monitoring in detecting and correcting phonologic errors in speech planning and production. The theory suggests that the unintended side effects of these elaborate models of internal speech monitoring and repair create internally conflicting conditions, which are, in part, responsible for the dysfluency in both normal and stuttering speech. The brain’s executive section favors the abrupt interruption of normal articulation when there are word-finding or grammatical errors.

A third level of causation and an area of active investigation concerns the motor speech and vocal tract dynamics that are characteristic of stuttering. The vocal tract events generating disrupted speech patterns have been the focal point of much of the physiologic research on stuttering since the early 1980s. It is reasoned that increased knowledge of vocal tract events and the component behaviors of stuttering will ultimately have a direct impact on the diagnostic assessment, measurement, and therapeutic modification of these behaviors. Much information is available on the respiratory, laryngeal, and articulatory dynamics of stuttering. Physiologic studies indicating differences between stutterers and nonstutterers are closely scrutinized because they may, in fact, be dealing only with peripheral events surrounding stuttering, not with underlying causation. The above argument notwithstanding, individuals who stutter, as a group, exhibit relatively longer laryngeal, respiratory, and articulatory reaction times. It has been speculated that the speech–motor systems of people who stutter are limited in their skill or ability to prepare and perform the motor actions that are required to implement the various demands imposed by linguistic, cognitive, emotional, and motor aspects of speech. Additional evidence for this argument comes from within the stuttering patient population, which generally shows boys produce articulations with slower reaction times and diminished amplitude than girls. This observation is upheld by statistical evidence that boys are also five times more likely to be diagnosed with stuttering, providing further evidence that such seemingly peripheral observations may be a potential point of the target for both diagnostic measures and therapeutic modifications such as surgical intervention.

The final level of causation involves the environmental factors associated with the onset and development of stuttering. Much information has been collected on the personality traits of young children who stutter and their parents, as well as the psychodynamic, sociologic, and conditioning factors influencing stuttering and their interaction with linguistic development. Considerable inconsistency exists in data obtained from research on the personality of stutterers. There is little (if any) conclusive evidence that stutterers, as a group, exhibit a definable character structure or predictable set of personality traits. This inconsistency is further evidenced in a series of recent articles investigating temperament, emotional regulation, and emotional reactivity in young children who stutter. Some of these studies suggest that greater emotional reactivity experienced by preschool children who stutter, together with their relative inability to regulate their emotions and control their attention, may lead to fluency breakdown. Other studies have been unable to distinguish between the children who stutter and those who do not on these variables. Moreover, there appears to be much overlap in overall adjustment profiles between stutterers and nonstutterers, with lower social adjustment and self-esteem scores characterizing individuals who stutter. However, these traits are more appropriately explained as a product of affect and personality as opposed to the underlying cause of stuttering itself in these individuals.

DIAGNOSIS AND CLINICAL FEATURES

Stuttering is typically characterized by involuntary disruptions in the fluency of verbal expression. These disruptions may be audible or silent and are usually characterized by repetitions or prolongations of sounds or syllables, tense pauses, or the complete cessation or blocking of speech sounds.

Accessory or secondary features frequently accompany instances of stuttering. These secondary behaviors involve both the speech mechanism and related or unrelated body structures. Such behaviors may include eye blinks, facial grimacing, lip and tongue tremors, tics, vocal fry, head jerks, abnormal movements of the limbs, and breathing irregularities. Linguistic and situational avoidance behavior is often observable with confirmed stuttering. The affected speaker will often scan the intended utterance to predict stuttering and then attempt to avoid such disruption by substituting words, revising phrases, circumlocutions, or stopping the entire communicative process. Situational avoidance is often reported, such as speaking in front of a group or on the telephone. Most adult stutterers report specific sounds, words, or situations that increase stuttering. Most of this information comes from studies carried out on chronic adult stutterers. Children who stutter have not been studied extensively, but it is generally accepted by clinicians that stuttering in its earlier forms is more often restricted to simpler types of dysfluency, such as repetitions and prolongations, and less often involves the more severe struggle and avoidance behaviors. Definitions of *stuttering* vary widely and are usually based on listeners' observations of the disrupted speech pattern. However, some authors emphasize that the affected individual's sensations of the stuttering moment are crucial to any complete definition. Many forms of stuttering are so subtle that they are undetectable by the listener and are only perceived by the speaker.

Although stuttering varies markedly in terms of frequency and severity, both within and among individuals who stutter, some fairly predictable events or situations are commonly observed that either facilitate fluency or result in a deterioration of the stuttering (e.g., choral reading, singing, whispering, speaking in the presence of masking noise or under conditions of delayed auditory feedback, speaking when alone). Conditions commonly observed to increase stuttering include talking on the telephone, talking to groups, making introductions and saying one's own name, talking to authority figures, and attempting to communicate during stressful situations such as job interviews or formal presentations.

Although severe forms of stuttering are most often observed in adults, children and adolescents can be similarly affected. Stuttering can have a significant impact on the individual's quality of life, causing significant social, emotional, educational, and vocational problems. Preschool and young school-aged stutterers frequently develop social and school-related anxiety. In fact, school children who stutter have a sixfold increase in their likelihood of having a diagnosable social or general anxiety disorder, as compared to a nonstuttering child. The adolescent who stutters tends to withdraw from normal social interaction. The young adult stutterer often reports feelings of restricted academic and career choices and often forgoes the opportunity to develop specific talents and career paths.

Adam was a pleasant and likable 8-year-old boy in the second grade. He had a younger brother and older sister whose speech developed normally. Adam's speech was also developing normally until around the age of 3½, at which time he began to show numerous repetitions of words, syllables, and sounds. Adam's pediatrician felt that the dysfluent speech was only a phase that Adam would soon outgrow. Adam's speech progressively worsened, and he began to show signs of struggle as he tried to force his words out on increasingly frequent occasions. The family remained convinced that the physician's predictions would, in fact, be borne out and that Adam would eventually grow out of the problem. By the time Adam was in the second grade, his speech was characterized by severe stuttering, taking the form of silent blocks lasting up to 20 seconds. These blocks were associated with generalized upper body tension and breathing abnormality just prior to the instance of disruption. In addition to being visibly frustrated by his inability to communicate at school and with his friends, Adam became withdrawn and socially isolated. He would not participate in class and would often refuse to go to school if he knew that he would be called on during a certain activity. Adam's mother also reported that there had been several instances of bullying and teasing by two of Adam's classmates. On several occasions, Adam returned home from school crying because of the remarks made concerning his speech. At this point, Adam's mother consented to have the school speech–language pathologist perform a full assessment and begin therapy sessions in order to provide Adam with the relief and support that he had been lacking.

The case study of Adam illustrates two commonly observed scenarios with regard to children's early speech development. First, childhood dysfluent speech patterns are often mistakenly interpreted as normal developmental phenomena. Although it is true that many young children will pass through periods of disrupted speech patterns and will spontaneously recover from these, a significant minority, estimated as high as 30%, of these children will continue to exhibit disrupted speech patterns, and if left unattended, they will develop confirmed stuttering. Recent studies point to the importance of early clinical intervention and, at the very least, referral for assessment by a speech–language pathologist to determine whether treatment is indicated. The case study highlights a second frequently observed social impact on children who stutter: teasing and bullying. Recent studies have demonstrated that more than 80% of children who stutter have, at some point, experienced bullying, usually in the form of teasing or name-calling. Children who stutter will often experience feelings of shame and reduced self-esteem as they are identified by their peers as different and vulnerable. They will often be excluded from their peer group activities and become increasingly withdrawn and socially isolated. Because they feel inadequate as communicators, these feelings of inadequacy quite often evolve into generalized feelings of low self-worth and esteem.

Susan is a 28-year-old bilingual speaker of English and Bengali. At the time of her assessment, her speech was characterized by severe amounts of hard glottal attacks, sound prolongations, and laryngeal blocks. Secondary behaviors observed included eye blinks and lip tremor. She reported that she regularly used avoidance strategies such as starters, fillers, and word switching to avoid stuttering in public whenever possible. Susan also reported that she avoided speaking with unfamiliar listeners and speaking in group situations. Susan reported that she had participated in treatment at the ages of 8 and 10 with some success in structured settings, but her gains had been difficult to maintain.

Based on her assessment, it was recommended that she participate in an adult group intensive treatment program focusing on fluency-shaping strategies. Susan was judged to be an excellent candidate for this model of treatment based on the following criteria: First, she demonstrated through her assessment interview and reading sample that she possessed the intellectual capacity to manage the rigorous demands of the intensive nature of the program. Second, she demonstrated through her oral mechanical examination and trial therapy probe that she was capable of managing the physical demands of altering her motor movement patterns. Third, she evidenced the emotional readiness to benefit from group therapy and to address the cognitive restructuring requirements of the program. Fourth, she evidenced a high level of motivation to successfully complete treatment and participate in a long-term maintenance program.

Susan’s treatment program included 3 weeks of daily clinic sessions lasting for close to 5 hours per day. She further dedicated between 2 and 4 hours to the completion of home assignments each evening.

In addition to successfully acquiring the motor skills required for fluent speech, Susan responded well to the cognitive restructuring goals of the program. She took to heart the suggestion that optimal success in treatment cannot be achieved without a genuine acceptance of one’s self as a person who stutters. She gained a level of comfort and openness about stuttering by completing a number of assignments that required her to practice with a variety of partners outside the group. This afforded her the opportunity to discuss her speech condition and therapy program with a number of family members, friends, and coworkers.

Susan is currently in the maintenance phase of her treatment program. She has taken many positive steps to create a lifestyle that supports her goal of long-term fluency maintenance. Susan carefully completes her daily schedule of maintenance practice. When she neglects her practice, she realizes the negative effect on her speech but manages these ups and downs with a positive attitude. She thoughtfully selects short-term goals at each follow-up meeting that will support her long-term goal of successful maintenance. Most recently, she revealed that she had been accepted to a graduate program in speech–language pathology.

DIFFERENTIAL DIAGNOSIS

The ability to produce fluent speech is a skill that develops as the young child grows, and it depends on such factors as motor coordination, timing, linguistic, and cognitive development, as well as emotional development and maturity. Fluent speech production results when the motor skills to produce a desired message occur sequentially and synchronously in a smooth manner. This requires a fully developed system composed of respiratory, phonatory, resonatory, and articulatory components acting interdependently to produce fluent speech. Given that these systems are not fully developed at 3 or 4 years of age, preschool children quite commonly exhibit dysfluency in speech during their communication efforts. The factors that distinguish normal nonfluency from incipient stuttering involve both the type and frequency of the observed disrupted speech patterns. Children who stutter exhibit predominantly part-word repetitions, voiced or unvoiced sound prolongations, and broken or fragmented words, whereas the normally nonfluent child is more likely to demonstrate whole word and phrase repetitions, revisions, and end syllable prolongations. Several researchers have emphasized the degree of speech fragmentation as the distinguishing factor between stuttering and normal nonfluency. The child who is developing normal speech patterns usually does not evidence more than 10% of any type of dysfluency in speech, whereas the stuttering child often exhibits higher amounts of speech disruption. Another distinguishing factor is the presence or absence of struggle during speech attempts. The stuttering child exhibits difficulty in initiating or maintaining proper breath flow or voicing during speech attempts. The normally developing child is more likely to produce easy, effortless speech with continuous, uninterrupted breathing, even if it is intermittently marked by repetitions of whole words or phrases. As such, a caregiver is in a key position to differentiate whether the child's stuttering requires clinical attention or the child is undergoing a phase of dysfluency.

The discussion thus far has centered on developmental stuttering, the disorder of fluency most commonly observed in children and adults. It should be noted that there are other types of speech dysfluencies that may be observed in clinic populations, two of which will be mentioned here.

Acquired neurogenic stuttering is a condition that can appear for the first time in adulthood and is usually associated with a variety of temporary, progressive, and nonprogressive neurologic disorders, such as traumatic brain injuries, strokes, and extrapyramidal diseases. In addition to adult cases of neurogenic stuttering, there have also been reports in which childhood stuttering worsened or reappeared with the onset of the neurologic disorder. Several features of neurogenic stuttering help to distinguish it from developmental stuttering: general adversities from the problem but no apparent anxiety, instances of dysfluency occurring in any part of the utterance, lack of secondary behavioral symptoms, stuttering on small grammatical words and substantive words with approximately equal frequency, and reduced variability of stuttering across different communicative situations.

Cluttering is another disorder characterized by rapid speech, reduced intelligibility, and frequent telescoping of words under the DSM-5. Cluttering is often marked by grammatical and syntactic errors and is not typically associated with any specific sound or word fears or difficulty maintaining breathing levels when producing sentences. Cluttering has often been associated with a central language imbalance, with some authors suggesting that cluttering exists when an individual presents with one or more impairments in each of five broad communicative dimensions reflecting cognitive, linguistic, pragmatic, speech, and motor abilities. Some of the features differentiating cluttering from stuttering are self-monitoring and general awareness. Whereas people with developmental stuttering are usually hypersensitive to the sound of their own speech, the person who clutters is often only vaguely conscious of their aberrant speech pattern and inconsistencies, let alone the existence of their speech disorder. Cluttering patients are often met with poor prognosis, largely as a consequence of the individual’s denial to seek medical help for a condition they are seldom aware of.

COURSE AND PROGNOSIS

The onset of stuttering can occur anytime between the ages of 2 and 7, peaking at approximately age 4. Most of the risk for the onset of stuttering is usually negated by the age of 6. The stuttering fluctuates in severity initially, as every patient has a unique course of development. Initially, days or weeks may go by with little or no speech difficulty, only to be followed by a sharp increase in the amount and type of stuttering behavior, and this may be dependent on environmental triggers. Despite fluctuations in severity, young children who begin to demonstrate signs of true stuttering do not continue to be dysfluent in the same manner they began. As children begin to demonstrate more struggle and fragmentation of words and syllables, the stuttering pattern becomes more complex. If these children are left unattended, their stuttering episodes slowly become more concrete and chronic, proving more difficult, but still possible, to treat with speech therapy. In its most advanced form, speech disruptions are accompanied by secondary behaviors such as those previously described. Complete recovery from stuttering is rare after puberty. As many as 80% of stuttering children experience spontaneous recovery, with girls more likely to become spontaneously fluent than boys. Investigations of prognosis in therapy for the older, confirmed stutterer have focused primarily on the influence of personality factors in treatment; subjects with less favorable speech attitudes were less likely to obtain long-term benefits from behaviorally oriented speech therapy for stuttering.

TREATMENT

The vast majority of treatment techniques for stuttering have been conducted on adults. Speech–language pathologists have traditionally practiced two distinct forms of therapy. The first deals directly with the psychological and cognitive aspects of the problem—that is, the attitudes, feelings, and emotions displayed by the individual who stutters. Treatment procedures use various types of psychotherapy directed at self-acceptance, attitude change, and avoidance or anxiety reduction. These therapy sessions aim to reduce feelings associated with the triggering of a stuttering episode. The second major speech therapy approach involves modification of the stuttering response to fluent-sounding speech by the application of systematic steps and rules of speech mechanics. In this framework, therapy procedures attempt to reconstruct the respiratory, phonatory, and articulatory gestures used in speech production, either by stuttering modification or fluency-shaping techniques. Therapy formats for adults who stutter have been intensified, and several major stuttering treatment programs now offer full-time treatment approaches administered in blocks of several weeks. Current research and clinical efforts are directed at the maintenance of fluent speech in the posttreatment environment. A significant number of individuals have been reported to exhibit either partial regression or even total relapse of stuttering without the continued practice of speech skills after treatment.

Varying definitions of stuttering and of successful treatment have largely hindered measures of treatment efficacy. Current therapies report success rates ranging from 60% to more than 90% when the sole dependent variable is self-perceived fluent-sounding speech immediately after treatment. However, thorough treatment outcome measures for stuttering should include three elements: (1) a significant positive change in speech output is produced, (2) the change is generalized across speaking situations, and (3) the change is maintained over time.

Far fewer studies have been done on the effects of speech therapy on young stuttering children. Clinicians tend to be somewhat divided on the issue of direct therapeutic intervention for the preschooler who stutters. Many advocate an indirect approach, incorporating parent counseling and environmental manipulation to treat the dysfluent child. This approach incorporates such techniques as parental modeling of relaxed, unhurried speech, and reduction of communicative stressors in the home environment. A second approach trains parents of very young children to apply behavioral techniques of verbal contingencies to stuttering and to verbally reinforce stutter-free speech in the child’s everyday speaking environment. If it is determined that these indirect approaches are not sufficient to fully address the stuttering, direct techniques of speech modification or fluency shaping, such as respiratory control and gentle phonatory onsets, are taught in a manner similar to those used with older populations.

Currently, there are no approved medications for stuttering on the market. However, advancements in neuroscience, such as medications that modify the dopamine receptors that reduce stuttering severity, may give patients a unique treatment option, especially for individuals who are unresponsive to traditional treatment methods.

ACKNOWLEDGMENTS

The chapter was written in memory of the late Dr. Joseph H. Beitchman, who continues to inspire us with his dedication, devotion, and extensive research work in childhood language disorders.

Further Readings

Ajdacic-Gross V, Bechtiger L, Rodgers S, et al. Subtypes of stuttering determined by latent class analysis in two Swiss epidemiological surveys. *PLoS One*. 2018;13(8):e0198450.

Ambrose NG, Yairi E, Loucks TM, Seery CH, Throneburg R. Relation of motor, linguistic and temperament factors in epidemiologic subtypes of persistent and recovered stuttering: initial findings. *J Fluency Disord*. 2015;45:12–26.

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Baxter S, Johnson M, Blank L, et al. The state of the art in non-pharmacological interventions for developmental stuttering. Part 1: a systematic review of effectiveness. *Int J Lang Commun Disord*. 2015;50(5):676–718.

Beal DS, Gracco VL, Brettschneider J, Kroll RM, De Nil LF. A voxel-based morphometry (VBM) analysis of regional grey and white matter volume abnormalities within the speech production network of children who stutter. *Cortex*. 2013;49:2151–2161.

Bleek B, Reuter M, Yaruss JS, Cook S, Faber J, Montag C. Relationships between personality characteristics of people who stutter and the impact of stuttering on everyday life. *J Fluency Disord*. 2012;37(4):325–333.

Chang SE, Zhu DC. Neural network connectivity differences in children who stutter. *Brain*. 2013;136(Pt 12):3709–3726.

Chang SE, Zhu DC, Choo AL, Angstadt M. White matter neuroanatomical differences in young children who stutter. *Brain*. 2015;138(Pt 3):694–711.

Choo AL, Smith SA, Li H. Associations between stuttering, comorbid conditions and executive function in children: a population-based study. *BMC Psychol*. 2020;8(1):113.

Cruz C, Amorim H, Beca G, Nunes R. Neurogenic stuttering: a review of the literature. *Rev Neurol*. 2018;66(2):59–64.

Drayna D, Kilshaw J, Kelly J. The sex ratio in familial persistent stuttering. *Am J Hum Genet*. 1999;65(5):1473–1475.

De Nil LF, Beal DS, Lafaille SJ, Kroll RM, Crawley AP, Gracco VL. The effects of simulated stuttering and prolonged speech on the neural activation patterns of stuttering and nonstuttering adults. *Brain Lang*. 2008;107(2):114–123.

*Guitar B, McCauley R. *Treatment of Stuttering: Established and Emerging Approaches*. Lippincott Williams & Wilkins; 2010.

Howell P. Signs of developmental stuttering up to age eight and at 12 plus. *Clin Psychol Rev*. 2007;27(3):287–306.

*Howell P, Au-Yeung J, Pilgrim L. Utterance rate and linguistic properties as determinants of lexical dysfluencies in children who stutter. *J Acoust Soc Am*. 1999;105:481–490.

Iverach L, Jones M, McLellan LF, et al. Prevalence of anxiety disorders among children who stutter. *J Fluency Disord*. 2016;49:13–28.

Kang C, Drayna D. A role for inherited metabolic deficits in persistent developmental stuttering. *Mol Genet Metab*. 2012;107(3):276–280.

*Kroll RM, De Nil LF, Kapur S, Houle S. A positron emission tomography investigation of post-treatment brain activation in stutterers. In: Hulstign W, Peters HMF, eds. *Speech Motor Production and Fluency Disorders. Proceedings of the Third International Conference on Speech Motor Production and Fluency Disorders*. Elsevier; 1997:307–320.

Kroll RM, Scott-Sulsky L. The fluency plus program: an integration of fluency shaping and cognitive restructuring procedures for adolescents and adults who stutter. In: Guitar B, McCauley R, eds. *Treatment of Stuttering: Established and Emerging Interventions*. Lippincott Williams & Wilkins; 2010:448.

Kronfeld-Duenias V, Amir O, Ezrati-Vinacour R, Ben-Shachar M, Civier O. The frontal aslant tract underlies speech fluency in persistent developmental stuttering. *Brain Struct Funct*. 2016;221(1):365–381.

Maguire GA, Nguyen DL, Simonson KC, Kurz TL. The pharmacologic treatment of stuttering and its neuropharmacologic basis. *Front Neurosci*. 2020;14:158.

Namasivayam AK, van Lieshout P. Speech motor skill and stuttering. *J Mot Behav*. 2011;43(6):477–489.

Neef NE, Anwander A, Friederici AD. The neurobiological grounding of persistent stuttering: from structure to function. *Curr Neurol Neurosci Rep*. 2015;15(9):63.

Ntourou K, Conture EG, Walden TA. Emotional reactivity and regulation in preschool-age children who stutter. *J Fluency Disord*. 2013;38(3):260–274.

Onslow M, Packman A. The Lidcombe program of early stuttering intervention. In: Ratner NB, Healey EC, eds. *Stuttering Research and Practice: Bridging the Gap*. Lawrence Erlbaum; 1999:193–210.

Perez HR, Stoeckle JH. Stuttering: clinical and research update. *Can Fam Physician*. 2016;62(6):479–784.

Sommer M, Koch MA, Paulus W, Weiller C, Büchel C. Disconnection of speech-relevant brain areas in persistent developmental stuttering. *Lancet*. 2002;360:380–383.

Walsh B, Mettel KM, Smith A. Speech motor planning and execution deficits in early childhood stuttering. *J Neurodev Disord*. 2015;7(1):27.

Yairi E, Ambrose N. Epidemiology of stuttering: 21st century advances. *J Fluency Disord*. 2013;38(2):66–87.

*Yairi EH, Seery CH. *Stuttering: Foundations and Clinical Applications*. 2nd ed. Pearson; 2015.

42.5 Unspecified Communication Disorder

TIM BRESSMANN, PH.D.

The section on communication disorders in the current fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders, Text Revision* (DSM-5-TR) focuses on developmental speech and language disorders, which is consistent with previous versions. The category unspecified communication disorder is open ended and may capture other notable aspects of a patient's symptomatology as they relate to verbal communication.

This chapter will focus on voice disorders. Beyond the linguistic information that is conveyed through a speaker's speech, the underlying voice quality can add important information about the speaker's emotional and physical status. The patient's voice quality and quantity can be an important clinical sign in the anamnesis of a range of psychiatric conditions, and psychiatric and voice symptoms can sometimes interface in interesting ways. Voice disorders can equally be an outlet or a cause for stress and emotional turmoil. The boundary between structural and functional (psychological) etiologies is often blurred.

The voice is part of the speech production mechanism. Speech is a complex physiologic phenomenon that relies on the coordination of multiple anatomical systems. Airflow from the lungs causes the vocal folds to vibrate periodically, generating a raw voice that is filtered and shaped in the pharynx. The velopharyngeal sphincter functions like a valve that can direct the airflow orally to produce oral speech sounds. If the velum is lowered, speech sounds will be nasal. The movements of the lips, jaw, and tongue create differentiated, linguistically meaningful speech sounds. Accordingly, the subsystems of speech can be broken down as follows: respiration (airflow from the lungs), phonation (sound generation in the larynx), resonance (shaping of the sound quality in the vocal tract), oral–nasal balance (direction of airflow into the oral or nasal direction at the level of the velopharyngeal sphincter), articulation (modulation of the sound stream into consonant and vowel sounds with the tongue, jaw, and lips), and suprasegmental (speech rhythm, loudness, and intonation).

DEFINITION

In the DSM-5-TR, unspecified communication disorders are described as “presentations in which symptoms characteristic of communication disorder that cause clinically significant distress or impairment in social, occupational, or other important areas of functioning predominate but do not meet the full criteria for communication disorder or for any of the disorders in the neurodevelopmental disorders diagnostic class.” The wording of this descriptor is somewhat unfortunate because it implies that unspecified communication disorders do not reach a threshold of criteria to qualify as proper communication disorders. The DSM-5-TR definition goes on to state that the category is applicable in cases where the clinician chooses not to specify the reason that the criteria for another communication disorder are not met or for situations when there is insufficient information to make a more specific diagnosis.

However, the DSM-5 categories are not comprehensive and detailed enough to capture the full range of communication disorders. While this is reasonable because the DSM-5 focuses on psychiatric conditions, speech–language pathologists may feel that this leaves the manual at an uncomfortable halfway point. Many pediatric communication disorders are included with a fair level of detail, while other important and frequent acquired communication disorders such as aphasia (acquired language disorder), dysarthria (neurogenic motor speech disorder), as well as structurally related speech disorders related to head and neck cancer, to name but a few, are omitted in the DSM-5. The clinician may make a diagnosis of an unspecified communication disorder not because of insufficient information but simply because of a lack of a more appropriate diagnostic category.

Examples of disorders that lend themselves to the diagnosis of unspecified communication disorder include voice disorders but also structurally related disorders such as oncologic pathologies (laryngectomy or glossectomy). Because the latter occur only rarely during childhood and are rather self-evidently related to surgical resections, this section is focused on voice disorders.

Voices are as varied, individual, and recognizable as human faces. What is considered acceptable as normal voice quality, loudness, and pitch may vary markedly between languages or in different cultures. The clinician will therefore try to determine whether a client's voice is appropriate (rather than "normal") with regard to the speaker's physical makeup, culture, and language. A voice may be considered disordered when the speaker experiences marked functional limitations, such as a voice that does not project well or gives out. The vocal quality may attract negative attention and may be perceived as unpleasant by the speaker or the environment. Speaking may be uncomfortable or painful.

HISTORY

The beginnings of modern vocology date to the late 19th century. Much of the research during this time was conducted by singing teachers who were interested in the physiology of the anatomical structures they were trying to educate in their students. The early 20th century saw a rapid improvement in laryngeal imaging methods and a more thorough understanding of the vocal mechanism. During this time, researchers also observed that many voice disorders are the result of chronic hyperfunctional misuse and that the voice can be a target of psychosomatic processes. These findings roughly coincided with the development of early psychoanalysis. As a result, functional voice disorders were often interpreted as a manifestation of a suppressed personal conflict, resulting in a conversion dysphonia. In recent decades, voice scientists and speech–language pathologists have begun to differentiate more carefully between true psychogenic dysphonias and functional voice disorders that are related to excessive muscle tension. Patients with muscle tension dysphonia will experience voice problems chiefly because of poor musculoskeletal alignment. Muscle tension dysphonia can often be treated on its own merit so that a psychodynamic interpretation of the disorder is no longer regarded as necessary or helpful in every case.

COMPARATIVE NOSOLOGY

Human communication disorders are as complex and as varied as psychiatric conditions. The DSM-5 continues to recognize the relevance of communication disorders to the psychiatric profession. However, it is neither the aim nor the scope of the DSM-5 to provide the level of detail and differentiation that a speech–language pathologist would bring to the diagnosis and treatment of communication disorders. There are manifold communication disorders that cannot meet the relatively narrow criteria for the types of developmental communication disorders laid out by the DSM-5. Consequently, the DSM-5 category unspecified communication disorder is a diagnostic wild card.

EPIDEMIOLOGY

There are few reliable epidemiologic estimates for the prevalence of voice disorders, and the available figures vary considerably between studies. In part, this is due to uncertainty about the concept of the healthy voice: The acceptability of different vocal qualities and quantities varies considerably in different cultures, languages, and on different socioeconomic levels. Voice quality is an auditory–perceptual impression, so what is considered to be typical or pathologic is very much in the ear of the beholder. The estimated prevalence for pediatric voice disorders ranges from 6% to 23%. The prevalence of adult voice disorders has been estimated between 3% and 10%. The available epidemiologic reports vary in group sizes and compositions as well as in methodology. It is also difficult to estimate how many of the speakers who have been identified as having a pathologic voice quality will require treatment. The need for behavioral, surgical, or pharmaceutical intervention can probably be estimated to be closer to the lower, more conservative end of the prevalence estimates.

The occurrence of voice disorders varies across the lifespan. In children, boys tend to have more voice disorders, probably due to more aggressive vocal behavior and boisterous vocal play. By adolescence, the prevalence is roughly balanced between the sexes. In adults, voice disorders are more frequent among men than women. However, women seek help more often and sooner than men. The highest prevalence of voice disorders in adults is found in the age group of 45 to 64 years. Overall, most voice disorders are benign in nature (vocal nodules, polyps, edemas, and functional dysphonias).

ETIOLOGY

The precise etiology of a voice disorder is often difficult to determine and needs to be evaluated based on the nature of the pathology, the patterns of voice use, the emotional and physical states, and the underlying or accompanying medical conditions of the individual patient. Sudden onsets tend to denote acute processes, which are mostly reversible, and slow progressive onsets tend to denote longer-standing vocal trauma or degeneration that may require more comprehensive behavioral interventions, often in combination with phonosurgery.

Voice disorders are characterized by deterioration in vocal quality and quantity. A lesion on the free margin of the vocal fold results in incomplete vocal fold closure and a breathy, rough, or hoarse voice quality with a decreased loudness and pitch range. A referral to the otolaryngologist is necessary if a persistent change in vocal quality and quantity is observed. The otolaryngologic examination should always comprise a visual inspection of the larynx in order to rule out potential malignant lesions, such as squamous cell carcinoma.

Voice disorders can be congenital or acquired. The relatively rare congenital disorders of the larynx include conditions such as laryngomalacia (softness of the cartilages of the larynx and the epiglottis), subglottal stenosis (laryngeal webbing between the vocal folds), and rare laryngeal clefts. The immediate concern in infants with these pathologies is to keep the airway patent and ensure the patient can breathe.

Most voice disorders are acquired. Acquired voice disorders present as a hen-or-egg problem: It is difficult to determine whether abusive vocal behavior has led to a vocal pathology or whether a pre-existing pathology has triggered a vicious cycle of maladaptive behavior. Consequently, voice therapy and vocal re-education are paramount in all vocal pathologies.

In both children and adolescents, acquired voice disorders can be caused by a number of factors. The most important factor that causes injury to the vocal folds is vocal misuse (vocal hyperfunction) through excessive yelling, singing, talking, aggressive vocal play, or abusive body-focused repetitive behaviors such as excessive throat clearing. Voice disorders related to vocal misuse are often labeled as hyperfunctional voice disorders. Acute and traumatic vocal misuse can lead to lesions in the vocal fold, such as acute inflammatory (usually bilateral) vocal nodules, acute vocal fold polyps, or acute vocal hemorrhage. If these conditions are the consequence of a singular event (e.g., a spectator sports event), they are usually easily reversible by vocal rest and vocal hygiene. Protracted vocal misuse may require more comprehensive behavioral and possibly surgical interventions.

Long-standing voice misuse may result in chronic, fibrotic vocal nodules or chronic, indurated polyps. The delicate and supple superficial layer of the vocal fold (Reinke space) is particularly prone to irritation and swelling, resulting in chronic inflammatory edema of the vocal folds (Reinke edema, chronic laryngitis, polypoid degeneration). Granulomas or contact ulcers on the arytenoid cartilages may be a consequence of low-pitched, strained phonation, sometimes in combination with gastroesophageal reflux. Many of these pathologies will be accompanied by excessive muscle tension, which is often visible and palpable in the extrinsic strap muscles of the larynx. This tension leads to an elevated position of the larynx during phonation. Muscle tension dysphonia can be a symptom that accompanies an organic pathology, or it can be a voice disorder by itself.

Excessive muscle tension is also often found in psychogenic dysphonias. Patients with psychogenic voice disorders can present with severe dysphonia or even complete aphonia (loss of voice), although the larynx is free of any visible pathology. The symptoms of psychogenic dysphonias can be variable: Patients may claim a total loss of voice and present with only whispered or breathy phonation or they may present with an excessively rough and hoarse voice quality, often accompanied by pain and discomfort when speaking. In some psychogenic dysphonias, the ventricular folds (false folds) are adducted together with the true vocal folds. The ventricular folds are situated above the true vocal folds. This mixed glottal–ventricular phonation can be visualized endoscopically and results in rough, low-pitched voice quality (comparable to the singing technique used by Louis Armstrong). The onset of psychogenic dysphonias is mostly sudden, sometimes related to a bout of cold or laryngitis in combination with a high level of life stress or emotional upheaval. According to reports, patients with psychogenic dysphonias are predominantly female, but men and children may also be affected.

A specific psychogenic dysphonia that begins in adolescence (but sometimes extends into adulthood if left untreated) is puberphonia, in which the prepuberty pitch and voice quality are retained by the patient. The result is an age-inappropriate high-pitched voice. It is more obvious in men than in women since the male voice drops about an octave (12 semitones) in puberty, compared with about a terce (3 semitones) in females. A puberphonic voice disorder is called postmutational falsetto in men and juvenile voice in women. It is idiopathic but may be associated with resistance to adolescence and adulthood or a feminine self-identification (in men). It is often possible to elicit a more physiologic pitch from these speakers during nonspeech maneuvers such as coughing, throat clearing, and laughing. The discrepancy between the pitch of the speaking voice and during nonspeech tasks can be an indicator of psychogenic dysphonia or of postmutational falsetto.

The complete refusal to vocalize or speak is called selective mutism. This disorder is predominately observed in younger children. These children first achieve the regular milestones in speech and language development but, at some point, refuse to speak because of emotional or psychosocial distress. Most children with selective mutism limit their interaction to a small number of immediate family members with whom they have close relationships. The incidence of prolonged mutism (i.e., longer than 6 months) has been estimated at less than 1 child in 1,000. Girls are more prone to this disorder than boys. The DSM-5 classifies selective mutism as an anxiety disorder.

A disorder that is difficult to categorize in children as well as in adults is idiopathic paradoxical vocal fold movement. This disorder also goes by vocal cord dysfunction or exercise-induced laryngeal obstruction. Patients with this disorder are still frequently misdiagnosed with either asthma or with a panic disorder. The patients experience spontaneous respiratory distress, often related to strenuous physical activity, in a fashion that is similar to an asthma attack. However, no swelling of the bronchi and the trachea occurs in an episode of idiopathic vocal fold movement. The patient experiences isolated involuntary laryngospasms, resulting in complete closure of the vocal folds. Consequently, asthma medication will not provide the necessary relief. Instead, this disorder responds well to behavioral therapy and specific breathing exercises to open the vocal folds.

Viral infection with the human papillomavirus can affect the respiratory tract in children and adults and lead to multiple lesions. In order to prevent the spreading of papillomatosis into the trachea and lungs, the lesions are surgically removed. Since papilloma lesions tend to recur, multiple rounds of surgery may be required. Although this treatment is effective for viral infection, it may result in severe scarring of the larynx and the vocal folds. An incomplete or irregular vocal fold closure with a hoarse and breathy voice may result. While this condition is initially organic, the disease and its often-lengthy course of treatment can have significant downstream effects on the patient's psychosocial well-being and quality of life.

Neurogenic voice disorders are related to lesions or degenerative processes in the nervous system. Peripheral vocal fold paralysis can be caused by strokes, degenerative disorders, mechanical or surgical trauma, or viral infections. The peripheral paralysis can be uni- or bilateral, and the location of the lesion along the nerve will determine the position of the affected vocal fold and the consequences for breathing, swallowing, and speech. Lesions in the central nervous system have a more generalized impact on the whole speech motor system on all levels: respiration, phonation, resonance, articulation, and prosody. Depending on the localization of the lesion along the central motor pathways, a spastic, hyperkinetic, hypokinetic, ataxic, or pseudobulbar movement disorder will result. The associated motor speech disorders are labeled accordingly (as spastic, hyperkinetic, and so forth dysarthrias) and are thought to roughly reflect the nature of the underlying movement disorder. Apraxia of speech is a motor speech disorder that is characterized by deficits in muscle coordination and sequencing of speech gestures. This also affects the coordination of the laryngeal voicing and devoicing for voiced and unvoiced speech sounds (e.g., to differentiate /f/ and /v/).

Spasmodic dysphonia is an insidious neurogenic voice disorder that is believed to be a focal dystonia (like torticollis or blepharospasm), possibly caused by small lesions in the basal ganglia. In spasmodic dysphonia, phonation is interrupted by adductor (adductor spasmodic dysphonia) or abductor spasms (abductor spasmodic dysphonia). The voice quality in spasmodic dysphonia is unsteady and often effortful, marked by a chaotic pattern of vocal fold adduction and abduction. Visceral voice production during coughing or throat clearing is often unimpaired. However, during conscious, purposeful speech, the spasms become obvious. Spasmodic dysphonia is very rare in children and adolescents, but occasional cases have been reported. Diagnostically, it can be a challenge to differentiate spasmodic dysphonia from other disorders, such as psychogenic dysphonia or idiopathic paradoxical vocal fold motion.

A number of factors may aggravate or precipitate voice disorders, such as insufficient hydration, other medications, and gastroesophageal reflux. The vocal folds can be dehydrated by insufficient intake of fluids, active and passive smoking, or exposure to irritating or toxic chemical fumes in the work or home environment. Severe dehydration makes the vocal folds less flexible and more vulnerable to mechanical trauma caused by loud singing, shouting, aggressive throat clearing, and so forth. Medications can also contribute to vocal fold dehydration. Bronchodilators can influence airflow and air passage. Diuretics, corticosteroids, and decongestants can reduce the fluid level in the vocal fold. Antihistamines, antitussives, and antirefluxiva can reduce pharyngeal and laryngeal secretions. Long-term hormonal therapies (especially with testosterone) can change the vocal fold structure and permanently lower the pitch. A number of psychotropic drugs can cause xerostomia and have dehydrating effects on the vocal folds, including tetracyclic, tricyclic, antipsychotic, and anxiolytic drugs. Usually, the level of oral secretions is a good predictor of the laryngeal secretions and the lubrication of the vocal folds. If marked xerostomia is observed, the patient should be counseled about hydration by a speech–language pathologist before they develop voice problems. Finally, acid reflux from the esophagus can lead to esophagitis (gastroesophageal reflux disease, GERD) with ulcerations or granulomas on the posterior aspect of the vocal folds or the arytenoid cartilages and a resulting voice disorder. Reflux esophagitis with irritation of the larynx may also be the result of self-induced vomiting in bulimic patients. While the most acute symptoms of gastroesophageal reflux are typically observed at night when the patient is lying flat, a subgroup of patients with voice disorders report voice symptoms and pain while awake and upright. A common descriptor for this variety of reflux is laryngopharyngeal reflux.

The perceived subjective impairment resulting from a voice disorder is highly variable across patients. In the extreme, voice disorders may cause anxiety and depression and may impact the patient’s work and social interactions. Anxiety about the restitution of the premorbid voice quality is usually highest among professional voice users, such as singers and actors. Professional voice users are often very emotional about even minor changes in their vocal quality and range. Voice problems in professional voice users may sometimes be related to unrealistic or unhealthy artistic demands on the voice, such as prolonged singing or shouting with a deliberately rough and loud voice. In this sense, the vocal folds of a professional voice user are like the knees of a professional basketball player: Irrespective of talent, the body has individual physiologic limits for every organ. Some rock music singers have seemingly endless vocal stamina and never sustain a vocal injury, even though they misuse their vocal folds every night on stage. Other singers hit the physiologic limits of their larynx a lot sooner, despite a correct singing technique.

A population with unique voice therapy needs is transgender patients who strive to adjust their voice quality and speech. For the transgender patient's satisfaction with the treatment outcomes and quality of life, the voice change can be more relevant than the physical appearance. Therefore, it is important to ensure that the patient is referred to a speech–language pathologist who is competent in voice therapy for this group. In female-to-male transgender patients, male hormone therapy will lead to a thickening of the vocal folds, which will lower the speaking pitch. However, the vocal tract size will still be female, and the patient may have to work with a speech–language pathologist on developing masculine speaking mannerisms. The transition is often more difficult for a male-to-female transgender patient. The patient must learn to raise the speaking pitch and develop more feminine speaking habits. The larger male vocal tract size can make it more difficult for male-to-female transgender patients to project a female vocal image. Transgender patients can differ in their ambitions as well as their abilities to change their vocal gender as perceived by others, so the patient and the speech–language pathologist have to establish the most efficient strategies to achieve a voice quality that will be comfortable for the patient and effective for their social speech communication.

A final etiology that needs to be mentioned in this section is velocardiofacial syndrome (22q11 microdeletion syndrome). This congenital syndrome is related to a microdeletion on the long arm of chromosome 22 and has highly variable symptoms. As a result, patients are sometimes not identified at birth but only later when symptoms become manifest. Many of the patients show various developmental delays, especially with regard to speech and language development. In childhood, attention deficit hyperactivity disorder, anxiety disorders, phobias, or emotion regulation disorders are considerably more frequent than in the general population. In adulthood, schizophrenia and schizoaffective disorders are frequently observed. As part of the syndrome, many patients present with cleft palate, which leads to hypernasality (too much air and sound emitted through the nose during speech). The cleft palate in this population is often isolated so that it does not affect the lip and the alveolar ridge and is not visible in the patient’s face. It may even be submucous (concealed under a lining of mucosa) and hidden from a casual oral inspection. Patients with velocardiofacial syndrome are a prime example of the importance of interdisciplinary treatment of complex disorders. The members of the craniofacial treatment team need to be aware of the patient’s risk for mental health problems during their lifespan. On the other hand, a psychiatrist who notices hypernasal speech in a patient who has not been followed by a craniofacial team should make the appropriate referral.

DIAGNOSIS AND CLINICAL FEATURES

Any unexplained change in voice quality, especially when it involves hoarseness or pain and persists longer than 2 weeks, should be examined by an otolaryngologist. The diagnosis of voice disorders is based on several diagnostic procedures. One important component is the perceptual evaluation of the voice disorder by an experienced otolaryngologist or speech–language pathologist. This evaluation assesses the patient’s voice quality and quantity. Accepted general descriptors of voice quality are breathiness (resulting from incomplete closure or hypoadduction of the vocal folds) and roughness (irregular vocal fold vibration resulting from lesions on the vocal fold or hyperadduction of the vocal folds). Hoarseness is defined as a mixture of roughness and breathiness. Vocal quantity can be described in terms of the pitch and dynamic range of the voice. If the pitch is inappropriate, a speaker’s voice is too high or too low for their age, maturation, and sex. If the dynamic range is inappropriate, a patient may speak too loudly or too softly or may have difficulties raising their voice. Associated perceptual assessments concern the patient’s oral–nasal balance, which may make speech sound hyper- or hyponasal or may add a nasal twang quality. It should be noted that hyper- or hyponasality are not voice disorders per se. Finally, suprasegmental features of speech (prosody, speech melody) are assessed. Pathologic prosodic patterns include monotonous, unmodulated voice versus exaggerated inflections, excessively slow or fast speech, and inappropriate or uneconomical breaks for inhalation during speech. Quantitative instrumental measures of voice quality, voice range, or airflow can supplement the perceptual diagnosis. This is combined with a detailed interview that assesses voice usage and medical history together with health practices, social and occupational circumstances, and psychological well-being. In pediatric patients, developmental and family information need to be considered.

PATHOLOGY AND LABORATORY EXAMINATION

Most lesions of the vocal folds are benign and do not require detailed histologic analysis. However, suspicious lesions should be biopsied and analyzed in order to rule out malignancies. The malignant laryngeal tumors are usually of the squamous cell carcinoma type. Irrespective of the clinical presentation of the voice disorder, the vocal folds should always be examined visually by an experienced otolaryngologist in order to correctly assess or rule out any organic causes of the problem. This can be done with a simple mirror examination, but ideally, a flexible endoscopy or a videostroboscopy should be performed. The videostroboscopic examination allows the physician to slow down the display of the vocal fold motion so that the patterns of vocal fold vibration can be observed. High-speed video kymography allows filming the vocal fold motion in real time, with a temporal resolution of several thousand exposures per second. Although a complete and comprehensive assessment of a patient’s vocal structures can be crucial to determining the correct treatment, otolaryngologists may sometimes contend with only superficial investigations, especially of pediatric clients. Malignancies of the pediatric larynx are extremely rare, while on the other hand, a videostroboscopic examination of a small child can be very tedious. However, this wait-and-see attitude may sometimes delay appropriate treatment. Voice disorders in pediatric and adolescent populations should be taken just as seriously as adult voice disorders.

DIFFERENTIAL DIAGNOSIS

Today, most voice centers use videostroboscopy or flexible endoscopy. With these methods, organic voice disorders can be readily detected. However, the challenge in voice diagnosis is to identify the behavioral components of the voice disorder. In particular, the differential diagnosis between hyperfunctional voice disorders and psychogenic voice disorders can often be difficult for otolaryngologists and speech–language pathologists. As a result, the necessary referral to psychiatric services may be delayed. Psychogenic voice disorders often have a characteristic onset. They come about suddenly, sometimes related to a minor viral infection. The discrepancy between visceral vocalizations, such as yawning, and voice production in speech is often informative. If a normal voice quality can be produced during coughing, throat clearing, humming, yawning, and sighing but not in speech, then it is likely that the patient can be helped back to a normal voice by a short course of voice therapy. After an acceptable voice quality has been restored, the patient should be referred for a psychiatric follow-up. On the other hand, if the patient presents with a total loss of voice but does not cooperate with the therapist’s efforts to restore voice production, a different disorder to consider in pediatric patients is selective mutism.

One of the most difficult diagnoses for an otolaryngologist or a speech–language pathologist is spasmodic dysphonia, which is often diagnosed by exclusion of other applicable pathologies. In particular, psychogenic and neurogenic voice disorders must be ruled out. Since not every clinician is aware of spasmodic dysphonia, some patients still undergo prolonged diagnostic and therapeutic odysseys before the correct diagnosis is finally made. The voice breaks in spasmodic dysphonia may sometimes resemble the tonic blocks of a person who stutters. While the onset of spasmodic dysphonia is in adulthood, stuttering begins during childhood. Another difficult diagnosis is that of idiopathic paradoxical vocal fold motion, which can very much resemble an asthmatic attack. However, the airway constriction in these patients will be restricted to the level of the larynx and will not include the bronchi and the trachea. Finally, patients with velocardiofacial (22q11 microdeletion) syndrome may often present with hypernasality as well as a variety of psychiatric disorders throughout their lifespan. Therefore, it is important to ensure that any psychiatric patient with hypernasality of unclear etiology is referred to a craniofacial team for multidisciplinary assessment and genetic testing.

COURSE AND PROGNOSIS

As voice disorders may be caused by many different conditions, there is a high degree of variability in the long-term prognoses and treatment success. Acute organic processes that are associated with a one-time event of vocal trauma usually have a good prognosis. Examples of such pathologies are acute nodules or polyps. Vocal nodules in children are sometimes not removed by surgery as the vocal fold expands during puberty, and the nodules tend to resolve spontaneously. Organic pathologies that are related to long-standing vocal misuse will require more comprehensive therapeutic interventions that also address the associated abusive behaviors. Examples of such pathologies are chronic nodules, chronic polyps, and Reinke edema. If there is persistent voice misuse combined with aggravating factors such as smoking or alcohol abuse, the long-term prognosis can be poorer. Lesions such as nodules may recur quickly after a surgical resection if the maladaptive behaviors cannot be modified. Voice disorders with a pronounced psychogenic component should be addressed with a combined effort by the speech–language pathologist and the psychiatrist. Although the voice disorder can usually be managed effectively in speech therapy, the underlying emotional or psychosocial factors that have contributed to the voice disorder have to be evaluated by a psychiatrist.

TREATMENT

Intervention in voice disorders is dependent on the nature of the problem. Organic pathologies are often managed surgically with laser surgery or microsurgery, followed by a course of voice therapy. In functional voice disorders, the speech–language pathologist will counsel the patient about the nature of the disorder and commence a course of vocal re-education aimed at modifying or eliminating vocal misuse. If psychological factors contribute to the genesis or the persistence of the voice disorder, a referral to psychiatric services will be necessary.

There are many different approaches to voice therapy. Most of these fall within five main streams of voice therapy: symptomatic, physiologic, psychogenic, hygienic, and holistic. This distinction is largely academic because the boundaries between the different approaches are blurred. Most voice therapists will move between these different schools of thought and pursue a more personal eclectic approach to achieve optimum outcomes for their patients. *Symptomatic* voice therapy uses interventions and voice exercises that are geared toward the voice symptoms of the individual patient. *Physiologic* voice therapy draws on concepts from physical therapy, such as respiration or posture exercises, to achieve the foundations for physiologic voice production. *Psychogenic* approaches to voice therapy draw on the cognitive behavioral approach to behavior modification to change the way in which a patient reacts to life stresses with a change in voice quality. This type of therapy also uses relaxation techniques such as progressive muscle relaxation. *Hygienic* voice therapy is often a supplement rather than a standalone therapy approach. Hygienic voice therapy emphasizes proper hydration and reduction of gastroesophageal reflux. Finally, a number of schools of voice therapy are *holistic* in that they try to achieve goals on multiple levels through a fixed program of exercises. To give just a few examples of such holistic approaches, the *vocal function exercises* treat the vocal muscles like other muscles in the body and provide consist of endurance and stretching exercises to increase the vocal range and stamina. The *accent method* uses a combination of controlled respiration, chanting, and dancing to elicit relaxed phonation. *Resonant voice therapy* aims at maximizing oral resonance and achieving good voice projection through a combined program of vocalizing and increasing self-awareness.

The voice can be an outlet for psychological stress. While everybody has experienced symptoms such as a shaky voice or a dry mouth when feeling excited or threatened, in some patients with psychogenic voice disorders, more significant somatoform symptoms such as throat pain and globus sensation can occur together with a change in voice quality. These patients need help to find their normal voice again and often require a referral to psychiatric or counseling services to address the issues underlying the psychogenic voice disorder.

On the other hand, it is also easily possible to conjure up voice symptoms with the purpose of malingering. Garth was a boy of 8 years of age with typical speech and language. When the family moved to a new neighborhood, Garth initially found it difficult to settle into his new school. It was during this time that he came down with frequent bouts of throat pain, accompanied by coughing and alarmingly rough and hoarse voice quality. When these episodes occurred, he could stay home and rest. Because of the recurrent and worrisome voice symptoms, the family met with the school speech–language pathologist when Garth was going through another bout. The speech–language pathologist made Garth laugh with a joke and tested a few nonspeech vocalizations, such as yawning and sighing. On all of these nonspeech vocalizations, voice quality was regular. During the conversation, it also became clear that there was a pattern where the throat pain and voice symptoms became worse around the middle of the week but then improved toward the weekend when Garth was able to participate in family activities again. The patient was confronted with this pattern and agreed that he might have overstated the severity of the pain and voice symptoms. After a few sessions with a psychotherapist specializing in pediatrics and with the support of his family, Garth transitioned smoothly into his new school environment.

Steven was a 17-year-old teenager who was referred to the speech–language pathologist because of his high-pitched voice. Steven had experienced teasing and negative comments about his voice in the classroom and in other social situations. He was very motivated to change the pitch of his voice. The patient was of normal height and built for his age. An otolaryngologic mirror examination had ruled out the presence of a vocal fold web (patients with a laryngeal web often have a high pitch because the vibrating edge of their vocal folds is shorter). Upon palpation, the larynx was normal in size and had a normal movement range when Steven was asked to swallow (laryngeal elevation) and yawn (laryngeal lowering). An acoustic measurement revealed a fundamental frequency of around 230 Hz, which was twice as high as the appropriate pitch for a male speaker. When the patient was asked to throat clear and to hum, the fundamental frequency dropped to an appropriate range. The speech–language pathologist used gentle vocalization exercises such as chanting in combination with manual circumlaryngeal massage in order to help the patient relax his larynx and lower his fundamental frequency. Steven required only six sessions with the speech–language pathologist to change the pitch of his voice. However, he was apprehensive about using the new voice with his classmates at school. He was worried that the change was quite startling and that he would be ridiculed for seeing a speech–language pathologist. Steven decided that he would spend the Christmas holidays with his family and avoid seeing his friends during these days. When he went back to school in the new year, he told his classmates that he had been sick with flu and that the associated cough had done some damage to his voice box, which had changed his voice.

The patient described here was very motivated to change his voice to an age-appropriate pitch. However, some patients may initially resist the speech–language pathologist's efforts. Speaking with a different fundamental frequency changes a physical activity that is deeply ingrained, and that is as difficult to change as one's posture or gait. Patients may sometimes complain that the new way of phonating feels uncomfortable, hurts their larynx, or even affects their breathing. These initial complaints are usually overcome, and postmutational falsetto should be resolved with a short course of voice therapy. The effect on the patient's appearance in social situations can be stark, so it is important to help the patient develop strategies to gently introduce the environment to the new voice.

In puberty, the male larynx grows, and the fundamental frequency drops by an octave or 12 semitones. By contrast, the female larynx shows less growth, and the fundamental female frequency only drops three to four semitones. This means that female puberphonia (also called juvenile voice) can be less obvious and more difficult to recognize than the postmutational falsetto of male speakers. As a result, female puberphonia may be overlooked and treatment delayed.

FURTHER READINGS

Adler RK, Hirsch S, Mordaunt M. *Voice and Communication Therapy for the Transgender/Transsexual Client: A Comprehensive Clinical Guide*. 3rd ed. Plural; 2012.

*Andrea M, Dias Ó, Andrea M, Figueira ML. Functional voice disorders: the importance of the psychologist in clinical voice assessment. *J Voice*. 2017;31(4):507.e13-507.e22.

Baker J. The role of psychogenic and psychosocial factors in the development of functional voice disorders. *Int J Speech Lang Pathol*. 2008;10(4):210–230.

Boone DR, McFarlane SC, von Berg SL, Zraick RI. *The Voice and Voice Therapy*. 9th ed. Pearson, Allyn & Bacon; 2014.

*Butcher P, Elias A, Cavalli L. *Understanding and Treating Psychogenic Voice Disorder: A CBT Framework*. John Wiley; 2007.

Connor NP, Cohen SB, Theis SM, Thibeault SL, Heatley DG, Bless DM. Attitudes of children with dysphonia. *J Voice*. 2008;22(2):197–209.

D'haeseleer E, Behlau M, Bruneel L, et al. Factors involved in vocal fatigue: a pilot study. *Folia Phoniatr Logop*. 2016;68(3):112–118.

Dietrich M, Verdolini Abbott K, Gartner-Schmidt J, Rosen CA. The frequency of perceived stress, anxiety, and depression in patients with common pathologies affecting voice. *J Voice*. 2008;22(4):472–488.

Gallena SJK, Solomon NP, Johnson AT, Vossoughi J, Tian W. The effect of exercise on respiratory resistance in athletes with and without paradoxical vocal fold motion disorder. *Am J Speech Lang Pathol*. 2015;24:470–479.

Gothelf D, Frisch A, Michaelovsky E, Weizman A, Shprintzen RJ. Velo-cardio-facial syndrome. *J Ment Health Res Intellect Disabil*. 2009;2(2):149–167.

Gundel H, Busch R, Ceballos-Baumann A, Seifert E. Psychiatric comorbidity in patients with spasmodic dysphonia: a controlled study. *J Neurol Neurosurg Psychiatry*. 2007;78(12):1398–1400.

Koufman JA, Block C. Differential diagnosis of paradoxical vocal fold movement. *Am J Speech Lang Pathol*. 2008;17(4):327–334.

Lang C, Nir Z, Gothelf A, et al. The outcome of children with selective mutism following cognitive behavioral intervention: a follow-up study. *Eur J Pediatr*. 2016;175(4):481–487.

Miller T, Deary V, Patterson J. Improving access to psychological therapies in voice disorders: a cognitive behavioural therapy model. *Curr Opin Otolaryngol Head Neck Surg*. 2014;22(3):201–205.

Roy N, Holt KI, Redmond S, Muntz H. Behavioral characteristics of children with vocal fold nodules. *J Voice*. 2007;21(2):157–168.

Roy N, Merrill RM, Gray SD, Smith EM. Voice disorders in the general population: prevalence, risk factors, and occupational impact. *Laryngoscope*. 2005;115(11):1988–1995.

*Stemple JC, Roy N, Klaben B. *Clinical Voice Pathology: Theory and Management*. 6th ed. Plural; 2020.

Verduyckt I, Remacle M, Morsomme D. On the accuracy of adults' auditory perception of normophonic and dysphonic children's personality. *Logoped Phoniatr Vocol*. 2015;40(3):136–147.

Viana AG, Beidel DC, Rabian B. Selective mutism: a review and integration of the last 15 years. *Clin Psychol Rev*. 2009;29(1):57–67.

Whurr R, Lorch M. Review of differential diagnosis and management of spasmodic dysphonia. *Curr Opin Otolaryngol Head Neck Surg*. 2016;24:203–207.

43

Autism Spectrum Disorders

Fred R. Volkmar, M.D., Giacomo Vivanti, Ph.D., Ryan K. C. Yuen, Ph.D., and Kevin Pelphrey, Ph.D.

OVERVIEW

The spectrum of autism is characterized by severe and persistent problems in social interaction and difficulties with change and unusual responses to the environment consistent with the first description of the condition by Leo Kanner in 1943 (Kanner, 1943). In DSM-5, the term for the prototypic disorder is changed to autism spectrum disorder (ASD), reflecting an awareness of the broader range of difficulties observed. This term is meant to correspond (roughly) to the earlier diagnosis of autistic disorder in DSM-IV, although as discussed subsequently, the concept is somewhat narrower (despite the spectrum name). DSM-5 also includes a new communication disorder—social communication disorder—to encompass problems in social (pragmatic language) communication.

DEFINITION

ASD now encompasses, to varying degrees, the various concepts previously recognized in DSM-IV as part of the overarching pervasive developmental disorder category. In DSM-5, the historically used triad of impairments in social interaction, communication, and repetitive and restricted interests/behaviors was collapsed into two domains: social communication and restricted and repetitive behaviors. The definition requires that all four of the social communication criteria be met (see Table 43–1) and at least two of the four restricted interests and repetitive behaviors must be met. Sensory problems are now included in the latter group. These difficulties must be present from early childhood. Given the significant concern that this approach excludes many of the individuals who would have had a diagnosis of some pervasive developmental disorder in DSM-IV, a specific exception is made so that individuals with "well-established" diagnoses of Asperger disorder or pervasive developmental disorder not otherwise specified (PDD-NOS) can keep their label (and thus not be excluded from services). Effectively this solution continues the older DSM-IV system for previously diagnosed cases while creating a new (and somewhat different) one for newly diagnosed cases.

The DSM-5 also provides a series of specifiers, for example, if a known etiologic factor is present. Another change is the inclusion of a descriptor of required levels of support (this is left relatively ambiguous), with a range from levels 1 to 3. Other specifiers can indicate association with intellectual disability and language impairment or catatonia (see Table 43–2).

The changes in DSM-5 have been somewhat controversial. The older DSM-IV system used 12 polythetic criteria grouped in the three traditional areas (impairments in social interaction, communication, and restricted/repetitive interests) and resulted in over 2,000 combinations of criteria that could yield an autism diagnosis while the current approach offers only 12 such possibilities. A substantial body of work now makes it clear that, likely as a result of the way the DSM-5 system was developed, it takes a narrower view of autism. The two groups (in DSM-IV terms) who are at greatest risk of losing a diagnosis are the higher cognitively functioning and the more able younger children. Perhaps 80% of the higher cognitively functioning cases that previously would have had diagnoses of Asperger's or PDD-NOS lose their diagnosis with DSM-5, as do a significant number of more able individuals who previously would have had a diagnosis of autistic disorder.

The more stringent DSM-5 approach basically reflects the fundamental source of the data used, that is, high-quality data from research diagnostic instruments designed to assure consistency in autism diagnosis for cross-site genetic studies. The rationale for the new social communication disorder (a communication disorder diagnosis) is murky and the condition does not simply correspond to the older DSM-IV labels of Asperger's or PDD-NOS. The utility of the various proposed modifiers is debated with concern over their reliability given that clinical presentation can dramatically differ depending on observational context.

HISTORY

The first report of autism as a syndrome was provided by Leo Kanner in 1943. He described 11 children who had what he termed autistic disturbances of affective contact. He felt that this difficulty was innate and viewed this, along with insistence on sameness and difficulties with change, as the two essential features of the condition. Kanner noted the stereotyped motor mannerisms of autism and suggested that these were part of the child's difficulties with change and preference for sameness. Although he did not regard communication skill deficits as essential, he did observe that many of his cases never talked, and those who did have speech that was unusual for echolalia (repetitive use of phrases), unusual voice quality and volume, pronoun problems, and general difficulties in social use of language (pragmatics).

In giving autism a greater historical context, it should be noted that cases of autism had been noted before Kanner's report, for example, possible reports of some "feral" children as well as case reports from state training schools in the United States in the 1800s. His description also gave some false leads relative to confusion of the condition with schizophrenia (the use of the word autism—previously used to describe self-centered thinking in schizophrenia), a lack of apparent associated medical conditions, a tendency for parents to be well educated and successful, and the impression that the children did not have associated intellectual disability, given that on some of the nonverbal parts of IQ tests they did well. Even with these caveats, however, Kanner's phenomenologically based description remains classic.

A year after Kanner's report, Hans Asperger, a Viennese medical student, reported on a group of boys with a condition he termed "autistichen psychopathen" (probably best translated as autistic personality disorder). This group of boys had major social difficulties (hence the word autism as also used by Kanner) but good verbal language abilities, remarkable circumscribed interests that were all encompassing, and motor difficulties. The latter features were an area of difference from Kanner's description. Over time, the validity of the concept he described has been much debated and, although DSM-5 has no longer included it, the previous recognition in DSM-IV led to considerable public awareness and interest and an increase in research. In some ways, one of the most significant aspects of Asperger's work was to make it clear that there was a broader spectrum of autism—that is, consistent with the name change in DSM-5 if not with the definition used. His emphasis on genetic factors (he mentioned that fathers had similar problems) also prefigured the considerable interest in genetics discussed subsequently. Indeed, the recognition of a "broader autism phenotype" (BAP) has been the focus of much interest.

Table 43–1.
Autism Spectrum Disorder

Diagnostic Criteria

A. Persistent deficits in social communication and social interaction across multiple contexts, as manifested by the following, currently or by history (examples are illustrative, not exhaustive, see text):
   1. Deficits in social–emotional reciprocity, ranging, for example, from abnormal social approach and failure of normal back-and-forth conversation; to reduced sharing of interests, emotions, or affect; to failure to initiate or respond to social interactions.
   2. Deficits in nonverbal communicative behaviors used for social interaction, ranging, for example, from poorly integrated verbal and nonverbal communication; to abnormalities in eye contact and body language or deficits in understanding and use of gestures; to a total lack of facial expressions and nonverbal communication.
   3. Deficits in developing, maintaining, and understanding relationships, ranging, for example, from difficulties adjusting behavior to suit various social contexts; to difficulties in sharing imaginative play or in making friends; to absence of interest in peers.

Specify current severity:

Severity is based on social communication impairments and restricted repetitive patterns of behavior (see Table 43–2).

B. Restricted, repetitive patterns of behavior, interests, or activities, as manifested by at least two of the following, currently or by history (examples are illustrative, not exhaustive; see text):
   1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g., simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases).
   2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns or verbal–nonverbal behavior (e.g., extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat food every day).
   3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g., strong attachment to or preoccupation with

unusual objects, excessively circumscribed or perseverative interest).

4. Hyper- or hyporeactivity to sensory input or unusual interests in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement).

Specify current severity:

Severity is based on social communication impairments and restricted, repetitive patterns of behavior (see Table 43–2).

C. Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies in later life).

D. Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning.

E. These disturbances are not better explained by intellectual disability (intellectual developmental disorder) or global developmental delay. Intellectual disability and autism spectrum disorder frequently co-occur; to make comorbid diagnoses of autism spectrum disorder and intellectual disability, social communication should be below that expected for general developmental level.

Note: Individuals with a well-established DSM-IV diagnosis of autistic disorder, Asperger disorder, or pervasive developmental disorder not otherwise specified should be given the diagnosis of autism spectrum disorder. Individuals who have marked deficits in social communication, but whose symptoms do not otherwise meet criteria for autism spectrum disorder, should be evaluated for social (pragmatic) communication disorder.

Specify if:

With or without accompanying intellectual impairment

With or without accompanying language impairment

Associated with a known medical or genetic condition or environmental factor (Coding note: Use additional code to identify the associated medical or genetic condition.)

Associated with another neurodevelopmental, mental, or behavioral disorder (**Coding note:** Use additional code[s] to identify the associated neurodevelopmental, mental, or behavioral disorder[s].)

With catatonia (refer to the criteria for catatonia associated with another mental disorder, for definition) (**Coding note:** Use additional code 293.89 [F06.1] catatonia associated with autism spectrum disorder to indicate the presence of the comorbid catatonia.)

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 56, 57 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

In the years following its first definition, work on autism was limited and hampered by diagnostic confusion. This began to clear as diagnostic guidelines were suggested, as some of the misconceptions of autism were clarified (e.g., that autism could be associated with intellectual deficiency, was seen in all social classes, and was associated with a number of medical conditions, including epilepsy and some genetic conditions). During the 1970s, several lines of evidence converged to show that autism was a strongly genetic and brain-based disorder that responded best to structured teaching. Autism was first recognized officially in 1980 in DSM-III, and research both on biology and treatment expanded dramatically with this recognition.

COMPARATIVE NOSOLOGY

Before DSM-III, autism was regarded as a form of childhood schizophrenia, but during the 1970s, the work of Kolvin, Rutter, and others made it clear that these were actually two very different conditions. This recognition, along with other evidence on the validity of autism, led to its official recognition in DSM-III in 1980. The DSM-III definition was monothetic and heavily influenced by the work of Rutter but unfortunately lacked a developmental orientation (the condition was termed and defined as “infantile autism,” and cases not meeting the strict criteria given, but who once had done so, received a diagnosis of “residual infantile autism”—even though their problems were not particularly residual). In DSM-III-R, a more developmentally polythetic definition of autistic disorder was proposed based, in part, on the work of Lorna Wing. It did a better job in covering the entire range (in age and levels of severity) of the condition but there were some difficulties with overdiagnosis, particularly among the more intellectually impaired. By the time of DSM-IV, 10 years of work had accumulated, and the DSM-IV definition, developed in conjunction with ICD-10, included 12 criteria grouped in three sets. This polythetic definition converged with the ICD-10. Also, in DSM-IV, Asperger disorder was first recognized, along with Rett and childhood disintegrative disorder (CDD); in the two latter groups, a marked deterioration of function was observed. For Rett’s, the potential for confusion with autism was minimal, but it was felt it needed to be included given its distinctiveness and potential for genetic association (and indeed a specific gene was found to be associated with the condition). The unusual onset of CDD (autism that appeared after some 3 to 4 years or so of normal development) appeared to deserve special recognition; the condition itself had a long history antedating that of autism. In addition, in DSM-IV, and consistent with DSM-III and III-R, a category for cases that did not achieve a specific diagnosis but whose troubles resembled those of autism was included—PDD-NOS; such cases would today be recognized as part of the BAP.

Table 43–2.
Severity Levels for Autism Spectrum Disorder

Severity Level	Social Communication	Restricted, Repetitive Behaviors
Level 3		
“Requiring very substantial support”	Severe deficits in verbal and nonverbal social communication skills cause severe impairments in functioning, very limited initiation of social interactions, and minimal response to social overtures from others. For example, a person with few words of intelligible speech who rarely initiates interaction and, when they do, makes unusual approaches to meet needs only and responds to only very direct social approaches.	Inflexibility of behavior, extreme difficulty coping with change, or other restricted/repetitive behaviors markedly interfere with functioning in all spheres. Great distress/difficulty changing focus or action.
Level 2		
“Requiring substantial support”	Marked deficits in verbal and nonverbal social communication skills; social impairments apparent even with supports in place;	Inflexibility of behavior, difficulty coping with change, or other restricted/repetitive behaviors appear frequently enough to be obvious to the casual observer and

	limited initiation of social interactions; and reduced or abnormal responses to social overtures from others. For example, a person who speaks simple sentences, whose interaction is limited to narrow special interests, and who has markedly odd nonverbal communication.	interfere with functioning in a variety of contexts. Distress and/or difficulty changing focus or action.
Level 1 “Requiring support”	Without supports in place, deficits in social communication cause noticeable impairments. Difficulty initiating social interactions, and clear examples of atypical or unsuccessful response to social overtures of others. May appear to have decreased interest in social interactions. For example, a person who is able to speak in full sentences and engages in	Inflexibility of behavior causes significant interference with functioning in one or more contexts. Difficulty switching between activities. Problems of organization and planning hamper independence.

communication but
whose to-and-fro
conversation with
others fails, and
whose attempts to
make friends are odd
and typically
unsuccessful.

The definition of Asperger's in DSM-IV was problematic in several ways, and although a major text revision was undertaken in DSM-IV TR, the criteria themselves were unchanged so that divergent views of the concept remained. With DSM-5, it was argued that the presence of these divergent views argued against its inclusion. Unfortunately, this decision occurred even as data were emerging to suggest important differences from autism as well as the need for diagnostic refinement of the concept.

A major shift in focus of research also occurred with growing interest in the earliest manifestation of diagnosis of autism in infants and young children. This interest reflected the growing body of research on infants at risk, that is, siblings, as well as the awareness that earlier and effective treatments could be instituted and lead to better outcomes.

EPIDEMIOLOGY

Prevalence

Various factors complicate the study of the epidemiology of ASD. These include the broad range of syndrome expression, difficulties in case definition and identification, case evaluation, and methods used in surveys. The first studies of epidemiology suggested that autism was rare with a rate of roughly 1 per 2,000. However, changes in definition and study design and methodology have led to marked changes in this estimate. There are now over 60 epidemiologic studies, and current estimates suggest a rate of about 69 per 10,000 (or 1 in about 145 children). Time trends are difficult to study but it does appear that changes in diagnosis, practice, and educational policy (children with ASD often receive the most intensive services) and vastly increased public awareness account for some portion of this increase. Indeed around the world, increased rates often are associated with greater service availability and public awareness.

Current work on the issue of a possible true increase in the incidence of ASDs is attempting to control for various methodologic factors and may provide better data with which to address this issue. The impact of the DSM-5 criteria on rates remains to be clarified.

Sex Ratio and Gender Differences

A number of studies, using different approaches in both clinical and epidemiologic samples, have repeatedly observed higher rates of ASD in males than in females, on the order of three to five boys for every girl. This ratio has appeared to be more marked among those who are more cognitively able where rates of male predominance are much higher. Several lines of work raise questions about observed gender rates and differences. There is some suggestion that, among females with "classic" autism, severity is increased, suggesting some possibility that boys may be less vulnerable than girls to the cause/causes of autism. At the other end of functioning, there is also a growing body of work suggesting that factors such as diagnosis camouflaging may lead to underdiagnosis in more able females.

Social Class and Culture

In Kanner's original sample, parents were noted to be highly achieving and of high social status, but it is likely that this sample was biased in terms of case ascertainment. Epidemiologic studies have generally failed to reveal such associations. As noted previously, the label of ASD may be more acceptable to better-informed parents and lead to "diagnostic substitution." Within disadvantaged families, ASD may be underdiagnosed, reflecting lack of awareness and services and, potentially, even problems with diagnostic criteria. Similarly, cultural issues in the diagnosis of autism have also become a concern as efforts have been made to expand detection and treatment into developing countries. These issues may reflect basic problems with criteria and/or the ways they are operationalized in assessment instruments. Autism clearly is seen in all cultures and levels of ability, but cultural factors may play an important role in both diagnosis and treatment.

ETIOLOGY

Psychosocial Theories

In Kanner's original case series, most parents were remarkably successful, and this led to the notion that perhaps their success might have come at the cost of a lack of interest in the child (there was little awareness of the factors that did, of course, highly bias referral to the one person in the country who was interested in such children). This idea led, particularly in the 1950s and 1960, to notions like that of the "refrigerator" mother who failed to respond to the child and their emotional needs. As a result, often the child was removed to a presumably more neutral hospital setting and psychotherapy was undertaken for all concerned—with little impact. During the 1970s, a growing body of work began to clearly show the superiority of structured teaching and behavioral methods in intervention for children with ASD. Since the 1980s and the official recognition of autism as a distinct diagnostic category, a large body of research has grown on effective intervention in ASD.

More recent psychological theories have been more sophisticated and attempt to address the atypical cognitive and behavioral profiles of individuals with ASD, including the notions of core deficits in theory of mind, weak "central coherence," and executive dysfunction among others. These influential accounts, generated in the 1990s, shared the idea that a single "primary deficit" or "cognitive style" provides a parsimonious explanation for the heterogeneous manifestations observed in ASD. The theory of mind account posits that ASD symptoms originate from a fundamental inability to attribute mental states to others (e.g., inferring that a classmate is searching for a toy in her bag because she *believes* that the toy is there, even if the toy is on the floor). These difficulties hinder the ability to adjust behavior according to the social context, especially when more inferential reasoning is required to understand others' intentions (e.g., interpreting a sarcastic remark, whereby the literal content of a sentence does not match the speaker's intended meaning) and when the knowledge available to the child (e.g., the toy is on the floor) is different from that of others (e.g., the classmate's belief that the toy is in her bag).

The weak central coherence hypothesis points to a diminished drive to integrate details into coherent wholes as the central feature of ASD. This is reflected in strengths across tasks requiring one to ignore global configurations to focus on specific details (e.g., finding a triangle embedded in a larger meaningful figure) and weaknesses in situations demanding an appreciation of the "big picture" (e.g., disambiguating words that have multiple meanings based on the context, or understanding that a tree decorated with specific lights and ornaments is still a Christmas tree even if the ornaments have different colors compared to the previous year's Christmas tree).

Finally, the executive dysfunction hypothesis explains the difficulties observed in ASD as resulting from executive function impairments, including difficulties in attention shifting, updating and addition/deletion of information in working memory, inhibition of dominant or prepotent responses, abstraction, and processing of multiple cues—an account based on similarities between the behavior of individuals with ASD and that of patients with executive impairments due to frontal lobe lesions.

Although these accounts have generated a wealth of empirical knowledge on ASD, research has failed to lend unequivocal support to these and other theories positing a single cognitive process/dysfunction at the origin of all ASD manifestations. Core limitations across these accounts include unclear support for the specificity, universality, construct validity, explanatory power, and causal role of proposed "primary" deficits. In addition, explanations based on a "static" and universal cognitive style are challenged by research documenting developmental changes and variable outcomes in the ASD population. Accordingly, more recent theoretical approaches to ASD are less "grand" in scope and focus on developmental paths leading from early risk factors to the variable expressions of ASD symptoms.

Biologic Theories

The importance of biologic factors in pathogenesis was clarified as children with autism/ASD were followed over time. It became clear that they were at high risk for the development of seizures both early in life and again in adolescence (the latter being a very unusual time to develop epilepsy). Similarly, it became clear that with minimal or no effective intervention, children often ended up in the intellectually impaired range of development and exhibited various neurobiologic soft signs and biochemical abnormalities, for example, delayed development of hand dominance, persistence of “primitive” (early) childhood reflexes, and high levels of the neurotransmitter serotonin. The study of possible biologic abnormalities in the pathogenesis of autism remains an active area of research and has converged with genetic work, for example, in the study of single-gene models where it is possible to develop animal models most readily.

Genetic Factors

In the late 1970s, twin and then family studies began to demonstrate the central role of genetics in the pathogenesis of ASD. The concordance rate for strictly defined autism in monozygotic twins is about 60%, while the rate for concordance with the BAPs is 80% to 90%. These rates are much higher than dizygotic twin concordance rates. Overall, these and other data suggest that genetic factors account for a more substantial share of the risk for autism than environmental factors, although both sets of factors are seen to contribute substantially to population risk.

Although there has been general agreement on the central role of genetic factors for some decades, the search for specific genes has been a challenge. In recent years, the availability of large data sets, collaborative studies, and advances in genomic technologies have significantly increased knowledge of the genetic architecture of ASD and related conditions.

These studies have produced several important themes. One major area of progress has been via a focus on mutations that are rare in the population, may be new in the individual, and which carry significant biologic effects. These changes lead to single-gene dysfunction as well as larger variations in chromosomal submicroscopic structure copy number variations (CNVs). In addition, there is a substantial body of work that shows considerable evidence that polymorphisms that are frequent in the population (common genetic variants) convey risk, but they only confer a small effect on ASD risk individually. However, the aggregation of these small effects can contribute to a polygenic risk of ASD. The associated variants significantly overlap with those identified in other relevant disorders, such as schizophrenia and major depression. Third, while recent studies have underscored an exceptionally high degree of genetic heterogeneity underlying ASD, so far, the genes that have been discovered appear to converge on a much smaller number of biologic processes. To date, multiple studies have highlighted the role of, for example, chromatin modification, synaptic and signaling proteins, and targets of fragile X mental retardation protein (FMRP) as playing important roles in pathophysiology and underscore the importance of midfetal human cortical development. Fourth, there is emerging evidence that rare, de novo variations in the non–protein-coding regions of the genome carry risk for ASD. They include regions that are responsible for regulating brain development and repetitive regions that can interrupt gene functions. Moreover, it has also been repeatedly shown that the identical genetic variation in different individuals may increase the liability (sometimes sex dependent) for a very broad range of neurodevelopmental disorders, ranging from schizophrenia, to epilepsy, intellectual disability, or specific language impairment. These findings raise important questions about the specificity of genetic risks underlying ASD. Finally, while the relative contribution of genes is thought to be quite large, influences other than variation in the DNA sequence are likely to turn out to be important as well, including epigenetic mechanisms and environmental factors. Indeed, one of the most important outcomes of the successful search for ASD relevant factors will be the light it will shed on nongenetic, potentially modifiable risks.

Other Medical Conditions and ASD

ASD has also been associated with many different conditions which have a strong genetic component. Fragile X syndrome (FXS) is one of the best known of these associations. In this condition, the marker X chromosome is noted to be “fragile” when grown in folic acid–deficient media. This mutation usually involves a triplet repeat of cytosine–guanine–guanine in the regulatory domain of the gene *FMR1*, and this mutation may amplify in subsequent generations. Characteristic physical signs include distinctive facial features, enlarged testes, and associations with both intellectual deficiency and ASD. Behavioral problems are often noteworthy. It appears that perhaps 20% to 30% of boys with fragile X will meet criterial for ASD; conversely, about 1% of boys with ASD are found to carry the fragile X mutation. Understanding the biology and pathogenesis of this condition has potential implications for understanding the etiology of autism.

The advance in genome sequencing has also allowed discovery of novel conditions that are associated with ASD. This “genotype-first” approach involves a performance of comprehensive phenotyping of individuals identified with mutations in a specific gene through large-scale sequencing. One such example is an overgrowth syndrome that is caused by mutations in the *CHD8*—a gene that is highly expressed in brain during the early prenatal development period. *CHD8* codes for an ATP-dependent chromatin-remodeling factor that binds to beta-catenin and regulates its target genes. Physical signs of the condition related to this gene include macrocephaly, distinct facial features, gastrointestinal issues, sleep problems, and cognitive impairment. It is estimated that over 80% of individuals with *CHD8* mutations meet diagnostic criteria for autism spectrum conditions and that mutations in *CHD8* can be found in 0.35% of individuals with idiopathic ASD. The success of identifying novel conditions for ASD using this “genotype-first” approach will help refine the autism spectrum and allow development of more targeted treatments or interventions for ASD.

Pre-, Peri-, and Neonatal Factors

Several studies have shown increased rates of pre-, peri-, and neonatal complications in children with ASD. Much of the difference relates to the observation that something unusual is noted about the child at birth. The association with intellectual disability and these factors may be much stronger than the risk for ASD. This may reflect the operation of genetic as well as perinatal factors. The genetic predisposition to autism may interact with perinatal factors in producing the syndrome.

Other Etiologies

There has been much speculation on the association with a number of factors, although, to date, these mostly consist of anecdotes or case series rather than well-replicated and controlled scientific data. Potential confounds are often unappreciated, and there has been a tendency for case reports to be overemphasized. An early impression of the association of autism with disorders such as congenital rubella had to be modified as these children seemed to lose “autistic-like” features over time. In addition, there is now a greater appreciation of the complexities of diagnosis in what often was a severely disabled population.

Although there is clearly a very strong role for genetic factors in the pathogenesis of autism, there is at least some potential role for environmental factors. However, precise mechanisms remain to be specified, and this is an area of active work.

POSTMORTEM AND NEUROIMAGING STUDIES

Neuroimaging and Brain Development

ASD is a neurodevelopmental condition that emerges early in life. Much has been learned about ASD from postmortem, model systems, and neuroimaging studies over the last few decades. However, of these, magnetic resonance imaging (MRI) studies have provided the greatest contribution to the current understanding of attributes of brain structure and function associated with ASD. This section will review the main themes of this research, focusing on neuroanatomical studies using structural MRI (sMRI) and brain function using task-based (tb-) and resting state (rs-) functional MRI (fMRI).

Brain imaging as a tool and as an approach to understanding brain-based disorders emerged on the research landscape in the 1990s, with the arrival of fMRI to complement sMRI, which became a useful research tool about a decade earlier. Neuroimaging in ASD research, however, did not gain traction until the late 1990s, after which it accelerated rapidly, as it did in all subfields of neuropsychiatry. To date, the dominant research approach has been the case-control design, which considers individuals with ASD as categorically distinct from other groups, including typically developing controls (TDCs). Thus, studies focus on describing how the brain is different in ASD than TDCs, at what ages those differences emerge, and how specific brain features map onto phenotypic elements of ASD, genetic risk variants, and outcome measurements. Recent data, however, demonstrate that ASD is extremely heterogeneous, both in its core features, with severity ranging from the boundary of typical development to profoundly impaired, and with respect to many frequently co-occurring conditions such as intellectual disability, epilepsy, anxiety, and attention deficit hyperactivity disorder (ADHD). This is relevant to understanding the neurobiology of

ASD research literature, because often studies do not replicate upon follow-up and thus cannot provide reliable information. This is inherent both to the heterogeneity that accompanies the current conceptualization of the essence of ASD, as well as to the practice of publishing results based on inadequate sample size, for example, where case and control groups have less than about 20 to 25 individuals each. Inadequate sample size is an acknowledged problem across all of psychological science because it leads to failed replications, undermining the ability to accumulate a stable base of knowledge. Thus, the review of studies herein is confined to research where there have been several studies arriving at similar conclusions, with a bias away from describing studies based on small samples.

Developmental Neuroanatomy

The observation that individuals with ASD tend to have enlarged heads dates back to the origin of the diagnosis itself. Although subsequent studies of head circumference confirmed this observation, it did not receive much attention until the advent of MRI, which provided the excellent soft tissue contrast needed for more precise quantitation of brain morphology. The first MRI study from the middle of the 1990s found total tissue volume of individuals in the ASD group to be about 6% larger than the control group. Subsequent studies suggested that the size increase includes both gray and white matter. Courchesne and colleagues were the first to report that megalocephaly was observable as young as 2 years of age (but was not present at birth). This has been replicated and extended several times. For instance, Hazlett and colleagues showed that while gray matter volume and cortical surface area are enlarged in ASD, cortical thickness is not, which is an important observation, because it constrains the possible biologic mechanisms that could account for these size differences.

The Infant Brain Imaging Study (IBIS) has brought greater precision to the current understanding of the earliest signs of developmental brain differences in ASD. IBIS is a multisite, national study collecting data prospectively on at-risk baby siblings of youth with ASD, with the first brain MRI between 3 and 6 months of age, repeated at regular intervals up until 24 months of age. Initial IBIS results showed that axonal fibers had a blunted development in ASD between 6 and 24 months of age, with results clustering in brain regions implicated by fMRI study as part of the “social brain.” These data suggest that ASD begins to develop neurobiologically before the end of the first year of life in those who go on to a formal diagnosis at older ages (e.g., age 3). This is consistent with emerging data from very careful behavioral testing, which also shows significant deficits in the first year of life for those who go on to develop ASD. Another IBIS study found white matter microstructural abnormalities at 7 months of age in the group who went on to have ASD, in particular in fiber bundles associated with visual orienting and with the ability to disengage attention. These data help pinpoint brain circuits that appear to underlie “sticky attention,” a phenomenon in ASD that causes significant behavioral management issues and undermines learning. Both studies confirm that brain connectivity is an early feature that goes awry in ASD. A third IBIS study has focused more explicitly on how the white matter’s individual axonal fibers are organized in ASD using diffusion tensor imaging (DTI). This study found significant fiber network inefficiencies in the brains of toddlers with ASD at 24 months of age. Thus, a developing theme is that white matter shows aberrant organizational properties in the first year of life, which accentuate with time and subserve core features of ASD.

While most studies have not tried to localize areas of the brain that may be disproportionately enlarged, those that have often find frontal and temporal regions most enlarged. In addition, Munson and colleagues discovered early overgrowth of the amygdala, a structure vital to emotional processing, emotion regulation, and memory, in a sample with an average age of 4 years. Subsequent studies have verified and extended these findings showing associations between degree of overgrowth and severity of clinical symptoms. However, findings are confined to preschool ages, as increases are noted later in childhood and adolescence. Several MRI studies have suggested that the brain’s white matter is disproportionately enlarged in ASD, which is viewed as supportive of more recent conceptualizations of ASD as involving disrupted connectivity. However, not all white matter fiber tracts are enlarged. For example, numerous studies have found decreased cross-sectional width of the corpus callosum. Moreover, there is evidence that later myelinating areas, in particular, contribute the most to the increase in brain volume in ASD relative to typically TDCs (i.e., frontotemporal and frontoparietal regions). Allometric models of cerebral development suggest that doubling the number of neuronal processing units (the gray matter “minicolumn”) supports a fourfold increase in neuronal connectivity, potentially generating a fourfold increase in white matter volume. However, even though white matter enlargement seems a bit disproportionate compared to gray matter enlargement in ASD, it may be that the white matter enlargement is less than would be expected under model assumptions. The implication is that the bigger brains in ASD are actually less well connected than smaller brains among TDCs.

The neurodevelopmental trajectory of ASD was refined by subsequent studies that found that ASD increases in head size were observable only below age 12 years, after which group differences disappeared. This observation has been replicated many times, but the finding is not without opposing results. The single largest and best controlled study to date, albeit on head size (an imperfect proxy for brain size), found significant head circumference enlargement regardless of age. This study is remarkable for its very large sample size (*n* = 1,889 individuals with ASD from young childhood to young adulthood), and for controlling known confounds such as age, body size, and IQ measurements, as well as familial effects on head size, since both parents and an unaffected sibling were also included in the analyses (total *n* = 7,225). In summary, the brain growth and volume literature suggests that brain overgrowth begins in the first few years of life in ASD, and that by adolescence, brain size is much more typical. This abnormal pattern of growth across infancy and childhood is presumed to have very important implications for the functional properties of the brain and to likely perturb functional organization of systems that support typical social–emotional and cognitive skills. This atypical growth pattern also provides clues as to the possible cellular–genetic bases of ASD, especially mechanisms that govern neural numbers.

Consistent with this summary, a postmortem study of brains of those with ASD recently pinpointed a possible mechanism for brain volume differences in ASD. This study found good evidence for early deficits in neural pruning in ASD. While relatively small by MRI standards, the 20 brains analyzed (about half with ASD) represent one of the larger published postmortem studies. Samples came from individuals who died between 2 and 20 years of age. Synaptic pruning after the first year of life is well known as a normal developmental process, which is presumed to yield improved computational efficiencies. While typical controls lost about 40% of their dendritic spines (that part of a neuron that receives signal transmissions) across the age span, those with ASD lost half that much. The investigators also used a mouse model of ASD to study possible causes of this pruning failure. They found that mTOR, a key protein in brain function and development, was hyperactive, which impaired the ability to prune synapses. Overactivity of the mTOR pathway led to blockage of autophagy processes, reduction in pruning, and resulted in mice with ASD-like social behaviors. Significantly, the investigators were able to correct the pruning deficits and ASD-like behaviors in mice through use of an immunosuppressive drug, rapamycin, which inhibits mTOR activity. While this is unlikely to become a standard treatment in humans, owing to its very serious side effects, mTOR is currently being tested in clinical trials for a special subgroup of ASD (those with tuberous sclerosis), where the risk has been judged appropriate. Finally, the investigators also found proteins that reflect deficient autophagy and synaptic pruning malfunctioning in the human postmortem tissue. Collectively, these data are consistent with growing evidence of neural connectivity abnormalities in ASD, especially regional overconnectivity. Given that a great many of the known ASD genetic risk variants interact with the mTOR pathway, deficits in synaptic pruning provide plausible mechanisms contributing to the development and maintenance of ASD. Lack of pruning would also lead to an abundance of synapses, which would create the opportunity for seizures, a common comorbidity of ASD.

Recently, attention has focused on neuroanatomical changes later in adolescence and adulthood. One particularly relevant study matched a very large number of cases and controls (ages 7 to 25 years) on age, gender, and IQ, and found significant, widely distributed reductions in cortical thickness with age in ASD relative to TDCs. These declines were concentrated in the same areas of the brain shown in other studies as exhibiting hypertrophy early in life—frontal and temporal lobe regions, suggesting a unique inverted U-shaped developmental growth trajectory. Prefrontal cortex and the anterior temporal lobe also showed significantly reduced surface area with age (an age-by-group interaction). These results are consistent with several other recent studies that showed accelerating cortical thinning with age, including in the temporal lobe. Collectively, these data raise the concern that there might be undetected declines in cognitive and social–emotional functioning as individuals with ASD move into adulthood.

Functional Neuroimaging

fMRI studies of ASD have painted a complex portrait of differences in brain functioning in ASD. While MRI studies of neuroanatomy focus on measurements of a trait of a person, fMRI study results are influenced both by enduring personal traits (e.g., some people are very good at recognizing faces and others are not) and also mental and emotional states at the time of the assessment. Because the MRI experience can vary between individuals (e.g., stress provoking for some, sleep provoking for others), measurement signal can be influenced by state factors. The signal analyzed from fMRI comes from fluctuations in the magnetic field with different blood oxygen levels that accompany changes in neural activity. While the measurement is in this way indirect, fMRI has advantages over other neuroimaging techniques because of its ability to localize activity very precisely to different parts of the brain, with resolution in most studies of about 3 mm^3 (voxel size). Although this is precise by many standards, it is important to recognize that an average cortical gray matter voxel contains about 630,000 neurons.

fMRI studies can be of different types. Some are task based and have focused hypotheses about regions of the brain (e.g., the amygdala) that are hypo- or hyperactive in ASD. This type of study, based on a specific task (e.g., showing emotionally engaging images vs. neutral images with the participant providing a judgment that is recorded by a button press), has expectations that specific brain “modules” will be differentially engaged and has dominated the literature, particularly the early literature. This approach has been successfully used to test a range of theories about the primary deficits in ASD. The social, communication, and behavioral problems that occur with ASD suggest that the syndrome affects a functionally diverse and widely distributed set of neural systems. At one point or another, disturbance in nearly every neural system in the brain has been proposed to be a fundamental mechanism causing ASD. These theories are typically derived from beliefs about the most salient behavioral and psychological features of the disorder, sometimes with explicit consideration of developmental course. For example, those who emphasize difficulty with complex information processing as the principal characteristic of ASD postulate widespread cortical abnormalities sparing early sensory and emotional processes as the neural basis of ASD. Such cortical abnormalities, when present from early in life, would impair learning and contribute to developmental features of ASD. On the other hand, those who focus on the social reward and saliency deficits and their role in social difficulties often highlight the limbic system and reward system circuitry in the pathogenesis of ASD. Less sensitivity to social reinforcement and drive for social engagement would curtail social development. Theory-driven research is able to articulate specific hypotheses and pursue programmatic research that accumulates knowledge across time. Table 43–3 summarizes the knowledge about brain functioning that has been gained with this type of approach. Replicated studies that support a role of specific brain areas whose functions are atypical in ASD are represented in this table.

Table 43–3 reflects the literature and the strong interest in the field in localizing specific functions related to social ability to discrete locations in the brain. Perhaps the best studied process and brain region involves face perception and fusiform gyrus (FFG). The fusiform has a specific role in person identity from visual images. Numerous studies show that it is often underactive in ASD, paralleling difficulties on psychological tests of face recognition. The fusiform is functionally synchronized with the limbic system, and in particular the amygdala. The amygdala, along with the reward system, is believed to play a role in directing a person's attention to socially relevant things in the environment, such as people and their social interactions. Both the amygdala and the reward pathway have been shown to be hypoactive in ASD, though the amygdala's social responsiveness is confounded by its role in signaling anxiety and fear. There has been considerable speculation regarding the relationship between ASD symptoms and a constellation of brain areas referred to as the mirror neuron system (MNS). MNS has proven responsive to both the perception and implementation of specific body movements. The posterior part of the superior temporal sulcus has been highlighted in many fMRI studies of social perception (especially recognizing nonverbal cues, such as facial expressions, gestures, and eye gaze direction), as well as part of larger systems for representing the intentions of others (i.e., for generating a theory of mind of others). The hub of theory of mind activity, however, is generally localized to the medial prefrontal cortex. Orbital prefrontal cortex represents the value of experiences and of rewards and as such, it plays a key role in emotional learning.

Table 43–3.
Brain Regions of Potential Importance in ASD

Brain Area	Function
Amygdala	Emotional arousal, emotion perception, and emotional learning
Extended amygdala, ventral striatum, and nucleus accumbens	Social reward circuitry
Fusiform gyrus (especially lateral aspect of the right fusiform)	Face (person) recognition
Posterior, superior temporal sulcus	Interpreting nonverbal communications, action observation, and theory of mind
Orbital prefrontal cortex	Emotional learning
Medial prefrontal cortex	Social cognition (interpreting what others might be thinking and feeling)

In the past decade, there has been increased attention given to task-free fMRI studies, most notably what has been termed resting state (rs-) fMRI (rs-fMRI). With this approach, participants are not required to engage in any specific mental activity. Rather fMRI signal is gathered continuously during a rest period, generally of about 5 minutes long and often with more than one rest session. This approach has been most valuable in the context of functional connectivity studies of brain systems, an approach which examines the cross-correlation of disparate regional brain activity fluctuations. In essence, the approach searches for synchronous fMRI activity in order to create networks of regions that appear to be working together in a time-locked fashion. This approach has yielded numerous study findings, but many have not been consistently found across studies, and so confirmation of these findings, therefore, awaits larger studies. Those that are consistent often identify brain system differences associated with ASD that can be supported by lesion studies, or task-based fMRI (e.g., as described in Table 43–3). rs-fMRI is an important component of a major NIH initiative—the Human Connectome Project. This multicenter program of research is collecting imaging data on over 1,500 participants. The Human Connectome and related endeavors will allow connectivity matrices derived from graph theory on resting state and diffusion data from individuals with ASD to be compared to several other groups, including, for example, typical development, Alzheimer, and ADHD. These publicly available resources will generate the large samples needed to make gains in characterizing neural circuits with multimodal imaging in basic cognitive neuroscience and in ASD. This is significant given that other lines of evidence support the general theory of underconnectivity of the brains of individuals with ASD. This approach, along with increased implementation of multimodal imaging studies (where sMRI, fMRI, and other techniques are combined to address hypotheses about the neural architecture of ASD), promises to allow important advances in the understanding of ASD in the coming years.

One of the most relevant components of the Human Connectome Project for understanding ASD involves exploring, in unprecedented detail, how the brain typically grows and connects during childhood and the transition through puberty to adolescence and young adulthood. The Lifespan Human Connectome Project Development (HCP-D) Study is currently enrolling 1,300+ healthy children, adolescents, and young adults (ages 5 to 21) to discover how different parts of a child’s brain are connected and how these connections (the “connectome”) change as the brain develops. Characterizing brain development in healthy children will also allow us to better understand differences that can occur in children with ASD. Similarly, data collection is now ongoing for a large-scale, two-site longitudinal baby connectome project. The study includes

500 participants who will be studied with brain imaging and behavioral phenotyping at four to six visits from birth to age 5 years.

These longitudinal, developmental studies are providing the basis for a developmental perspective (i.e., early and longitudinal study of brain and behavioral development) on the cognitive neuroscience of ASD. Such a developmental perspective is crucial for several reasons. First, very little is known about the neural mechanisms supporting cognition in infants and children or about the changes in brain function, connectivity, and temporal dynamics that underlie normative behavioral development. This paucity is unfortunate because this basic knowledge is proving essential to understanding ASD. The developmental connectome studies will document the changes in brain functional connectivity as a function of the infant/child's age and sex and also detect changes in connectome during critical periods of childhood before, during, and after major developmental transitions. The availability of these data points will facilitate efforts to characterize atypical developmental pathways and will advance development of more effective interventions for ASD. Second, neurobiologic markers of developmental processes may prove useful in early identification of infants at risk for developing ASD. To the extent that the projects elucidate developmental trajectories of the neural circuitry supporting pivotal, early developing social abilities, it can inform the design of more effective programs for identifying and remediating risk for difficulties in these areas. It is generally accepted that earlier educational interventions are more effective for treating a variety of symptoms in ASD. Therefore, early identification of children with difficulties in cognition is pivotal to optimizing individual intervention outcomes. Developmental neurobiologic markers for individual differences in cognitive abilities would then not solely be important for improving early identification but also could offer advantages for earlier interventions. Moreover, it could be that a marker would relate to the severity of specific deficits and might help us to better understand the heterogeneity that is characteristic of ASD. With this information, more targeted treatments could be developed on a child-by-child basis and implemented early in ontogeny, which might then guarantee the most effective course of intervention possible, and thus the assessment of individual treatment needs could be far more efficient and standardized. Finally, fMRI (tb- and/or rs-) might actually provide a means to better quantify treatment effectiveness and reveal whether behavioral improvements correspond to compensatory changes in brain function or normalization of developmental pathways.

Connectome-Based, Individual Differences and Prediction

The human brain is a complex dynamic system, and it is notoriously difficult to build comprehensive theoretical models for even simple neurobiologic and cognitive processes. While there exists a degree of specialization and modularity in the human cortex, the functionality of a specific cognitive process is often largely distributed across brain networks. This makes distilling an experimental result down to a handful of cortical regions challenging and potentially misleading. With these challenges in mind, the advent of technologies in machine learning from computational sciences and graph theory from mathematics have sparked new modeling approaches in cognitive neuroscience.

One particularly influential modeling approach is the connectome-based predictive model (CPM). CPM leverages robust idiosyncrasies of functional connectivity graphs across individuals to make predictions about those individuals. CPM graphs can be constructed from resting-state functional connectivity data as well as functional connectivity data acquired during cognitive tasks. CPM is particularly suited for individual differences research, as the graphs are highly predictive of an individual’s intellectual disability. Irrespective of nuisance artifacts such as head motion or anatomical differences. This will allow the field of ASD neuroimaging research to move beyond case-control studies to fully explore the heterogeneity of ASD. Importantly, CPM predictions remain accurate over multiple scan sessions, and the reliability of the CPM graphs for both rs- and tb-functional connectivity increases monotonically as a function of scan time. Successful analysis frameworks using CPM include predicting an individual’s ID, age, attention, and general intelligence, among other features.

The development of artificial intelligence/machine learning–informed techniques like CPM marks a new era in which brain imaging could evolve into an integral part of a translational research chain. Novel behavioral treatment and pharmacotherapies for ASD may be further developed with the tremendous benefit of directly and more precisely assessing impairment and change in targeted neural circuits.

Neuroimaging-informed, developmentally sensitive biologic markers of underlying latent constructs such as social perception, social motivation, and social learning could be used to tailor an individualized intervention plan from treatment outset, informing decisions related to the selection of a behavioral treatment approach as well as its dose, duration and intensity, and any use of concurrent pharmacotherapy. To illustrate the power of such an approach, we highlight a study that utilized fMRI and a social perception task to predict treatment response in young children with ASD.

Pivotal response treatment (PRT) is an evidence-based treatment for young children with ASD. Extensive research has shown its efficacy in promoting language acquisition and social communication development in young children. While supported by a growing evidence base, PRT does not work for all children (no treatment does), and it is currently difficult for clinicians to predict whether a given child will benefit from PRT (or any other behavioral treatment). Unfortunately, this state of affairs means that a family may spend considerable time and financial resources on treatment that is not effective for their child. Utilizing machine learning and a rigorous cross-validation approach, Yang and colleagues discovered fMRI multivariate pattern analysis (MVPA)-based neurobiomarkers that accurately predict response to PRT at the individual subject level (Fig. 43–1). These neurobiomarkers consist of the pretreatment patterns and levels of activity and connectivity in response to coherent versus scrambled biologic motion in the neural circuits that support social information processing (posterior superior temporal sulcus [pSTS], FFG, amygdala, and inferior parietal lobule [IPL]) and social motivation/reward (orbitofrontal cortex [OFC], insula, putamen, pallidum, and ventral striatum [VES]). In sum, the development of biologic markers of treatment response may pave the way forward for precision medicine approaches in ASD.

Understudied Groups in ASD Neuroimaging

The majority of ASD research has focused on individuals with relatively intact language and cognitive functioning. Those characterized as having “low-functioning ASD,” including minimally verbal individuals, children with regressive trajectories, and those with intellectual disability, have tended to be overlooked. Methodologic and recruitment difficulties have played a role in this trend, but a desire to study a “pure” form of ASD, uncomplicated by potentially confounding factors, has likely contributed as well. This gap in the literature is exacerbated in the domain of human neuroimaging, where a high degree of behavioral compliance is often necessary to obtain usable data. We suggest that one of the next frontiers of brain research in ASD lies not so much in the development of new technologies or statistical methods as in increased attention to understudied subpopulations within the spectrum, particularly those with significant cognitive, language, or adaptive deficits. Neuroimaging approaches are a particularly powerful means to parse underlying differences (or similarities) in the biologic bases of varying ASD phenotypes. Expressions of ASD characterized by a more severe phenotype, including ASD with intellectual disability (ASD + ID), ASD with a history of developmental regression (ASD + R), and minimally verbal ASD (ASD + MV) are understudied generally, and especially in the domain of neuroimaging. ASD + MV may display abnormalities in preattentive auditory perceptual processing and structural deficits in brain structures that support expressive language function. In the future, longitudinal assessments and more detailed phenotyping of language profiles are particularly needed with this population. ASD + ID brain research findings have been highly variable. This may be related, in part, to underlying genetic heterogeneity in ASD + ID. Recent developments in imaging genetics of variants linked to ASD + ID (e.g., FMR1, NRXN1, CNTNAP2) suggest that these variants may shape brain structure in predictable but often opposing ways, indicating the need to take genotype into account in future ASD + ID work. ASD + R neuroimaging findings have focused mostly on electroencephalographic (EEG) abnormalities and epilepsy, but do not provide support for the hypothesis that epileptic encephalopathy contributes to the phenotype in the majority of cases; however, there is some suggestion that further exploration of nonepileptiform abnormalities may be worth investigating in some ASD + R cases. In the few MRI studies of ASD + R, there is some suggestion that brain enlargement may also be a worthwhile avenue of exploration. Investigating more extreme expressions of the phenotype may be a productive direction for future research.

NEUROCHEMISTRY

Dating back to 1961, there are several studies that have reported that approximately one-third of children with ASD have increased peripheral levels of the neurotransmitter serotonin. This has been the focus of much research, but the significance of this finding remains unclear, as the finding is not specific to ASD, and the relation of peripheral levels to central levels of serotonin is unclear.

Another line of work has focused on other neurotransmitters such as dopamine. A hyperdopaminergic functioning of the brain might explain the overactivity and stereotyped movements seen in ASD. This would be consistent with the general observation that administration of stimulants, which increase levels of dopamine, sometimes worsens behavioral functioning in ASD. Studies of cerebrospinal fluid (CSF) of dopamine metabolites and catecholamine metabolites have been inconsistent. It is clear that agents that block dopamine receptors are effective in reducing the stereotyped and hyperactive behaviors of many autistic children.

The endogenous opioids have been investigated, given the possibility that these compounds—enkephalins and endorphins—might lead to social withdrawal and unusual sensitivities to the environment. This was the rationale for use of opioid antagonists such as naltrexone in treating children with ASD. Although it does appear that there may be a modest effect of these agents on the high levels of activity and agitation overall, results have been disappointing. More recently, work has centered on the potential use of agents like oxytocin that may impact attachment and social brain systems more directly.

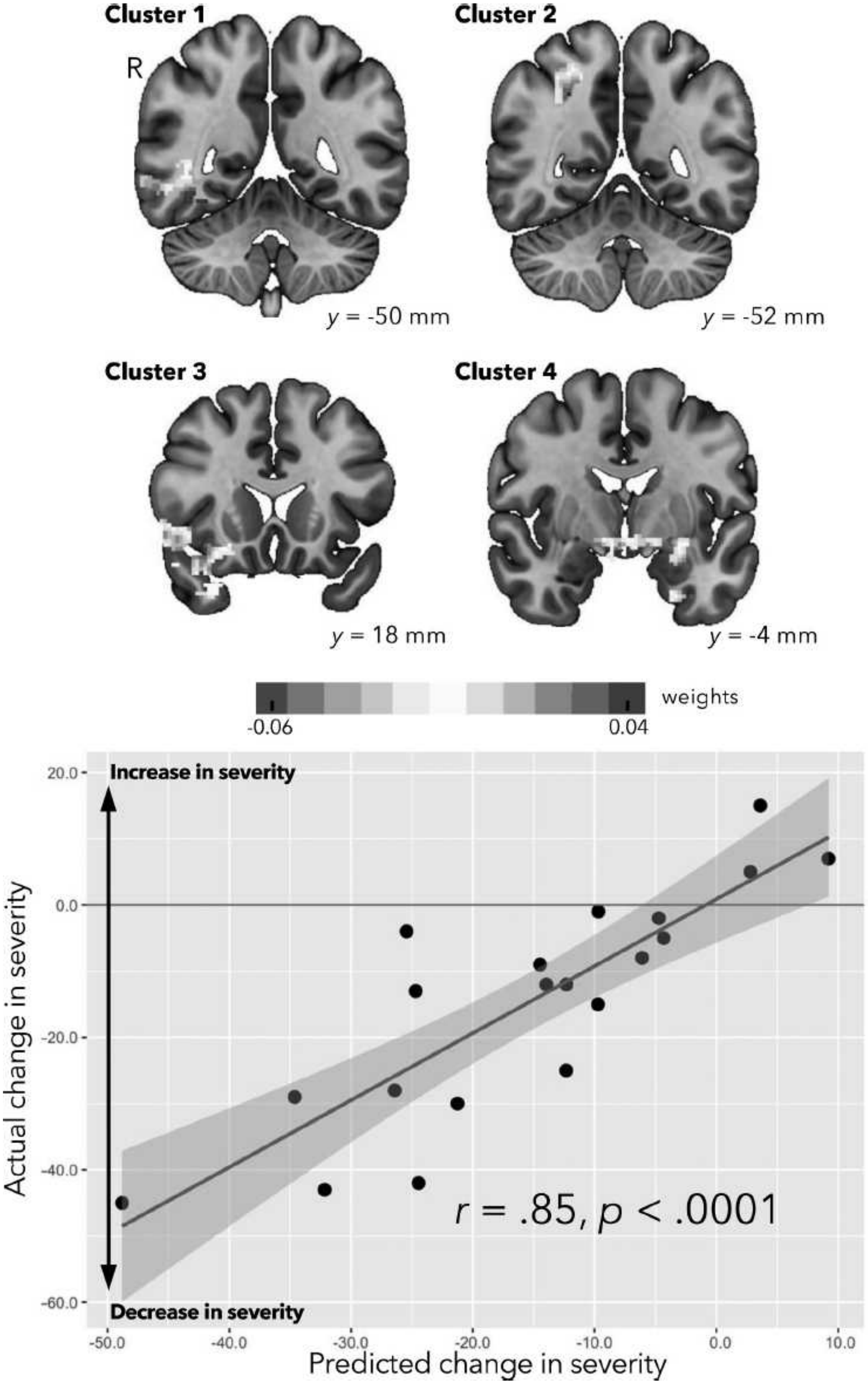
Cluster 1
R
y = -50 mm
Cluster 2
y = -52 mm
Cluster 3
y = 18 mm
Cluster 4
y = -4 mm
-0.06
0.04
weights
Increase in severity
Decrease in severity
20.0
0.0
-20.0
-40.0
-60.0
Actual change in severity
-50.0
-40.0
-30.0
-20.0
-10.0
0.0
10.0
Predicted change in severity
r = .85, p < .0001

FIGURE 43–1. The top panel illustrates four clusters of brain activity identified using functional magnetic resonance imaging (fMRI) multivariate pattern analysis (MVPA). The bottom panel illustrates how well these neurobiomarkers accurately predict each child's response to pivotal response training (PRT), an evidence-based behavioral intervention for core symptoms of autism. (See eBook for color images.) (Reprinted with permission from Yang D, Pelphrey KA, Sukhodolsky DG, et al. Brain responses to biological motion predict treatment outcome in young children with autism. *Transl Psychiatry*. 2016;6(11):e948. doi: 10.1038/tp.2016.213. Licensed under CC BY 4.0.)

IMMUNE THEORIES OF ASD

There is some work on possible immunologic contributions to autism; for example, maternal antibodies that might impact the fetus in utero. There is no evidence that vaccines cause autism. There is some suggestion that possible pre- and perinatal exposure to inflammation and infection in very young children might be a contributing factor. Epidemiologic studies show that gestational exposure to infection is predictor of ASD. Animal models of maternal immune activation (MIA) have provided strong experimental support for maternal immune dysregulation as a contributor to ASD. While hyperactive immune response can disrupt normal brain maturation and cause behavioral abnormalities later in life, the key molecular and cellular pathways responsible for MIA-induced neurodevelopmental disorders remain to be discovered.

DIAGNOSIS AND CLINICAL FEATURES

The DSM-5 diagnostic approach requires three essential social communicative features and at least two restricted/repetitive features thus mixing the older DSM-III monothetic and the more recent DSM-IV polythetic approaches. Deficits are required in the areas of social interaction across settings in area of social–emotional reciprocity, as well as in nonverbal pragmatic communication. Difficulties in all three features are required. In addition, a minimum of two of four restricted and repetitive behaviors must be observed. These include (1) stereotyped motor mannerisms; (2) restricted interests and insistence on sameness, and problems in dealing with change and overreliance on routine; (3) highly restricted interests that are abnormal in focus or intensity; and (4) hypo- or hypersensitivity to sensory input. These difficulties must be present from early childhood and be the source of clinically significant impairment. Various additional codes can be used to denote specific features such as intellectual delay or catatonia. In addition, symptoms cannot be better explained by any intellectual difficulty. Reflecting the concern of individuals losing a diagnosis, those with previously "well-established" DSM-IV diagnoses of autistic disorder, Asperger disorder, or PDD-NOS can retain a diagnosis in DSM-5. For cases in which there are major problems in social communication but who do not meet criteria for autism, the new category of social communication disorder can be considered.

Although DSM-5 gives various specific developmental/behavioral examples, it makes it clear that these are not to be taken as the only examples.

One of the challenges in ASD is the marked range of syndrome expression in ASD over both age and levels of intellectual and communicative ability. For the most cognitively impaired, the person is usually largely or entirely mute and exhibits severe deficits in social interaction. At somewhat higher levels of intellectual ability, there may be more passive interest in social interaction albeit with some spoken language, while the most cognitively able individual with ASD often exhibits eccentric and one-sided social styles. In the latter group, pragmatic language deficits are often striking. This group has been termed "active but odd" and often presents the greatest challenge for the new DSM-5 approach. There is considerable potential for misdiagnosis, particularly for the extremes in age and levels of cognitive functioning. Some older persons slip through the cracks and don't appear for diagnosis until adulthood.

The evaluation of the individual with ASD should include a detailed history, comprehensive medical (including genetic) and developmental examination, and psychological testing. Official practice guidelines for screening, assessment, and treatment are available. Specific tests are noted subsequently.

Age of Onset

Kanner believed initially that autism was congenital, and it does appear that onset is indeed early in life. Figure 43–2 illustrates age of parental concern—usually around 12 and 18 months as language fails to develop. Sometimes, parents worry that the child is deaf even though they would respond very dramatically to some environmental sounds. Occasionally, parents will report that the infant is “too good,” making few demands on them and having limited interest in social interaction. Such a situation is in stark contrast to the noteworthy social engagement of the typically developing child who finds in people and particularly the human face some of the most interesting and salient features in the world. At times, parents may report that the child seemed to develop some language but then stopped or lost verbal abilities or development plateaus without further gains. This type of regression/developmental stagnation may be reported in 20% or so of cases. Rarely, a child develops normally for a few years and then loses abilities and an autistic-like presentation develops. In addition to careful developmental history, use of videos or even photographs may be helpful. Cases with marked regression after a considerable period of normal development are rare and had their own category in DSM-IV. In contrast to more classic autism/ASD cases, DSM-IV type Asperger’s may not be picked up until the child enters preschool and social difficulties with peers become striking. In these cases, although language has developed, social pragmatic language use is a source of great difficulty, and intense special interests are common. Many of these cases no longer receive a DSM-5 diagnosis of ASD (unless they already had a “well-established” diagnosis). Table 43–4 summarizes some of the features important in differentiating autism from other disorders in the first years of life.

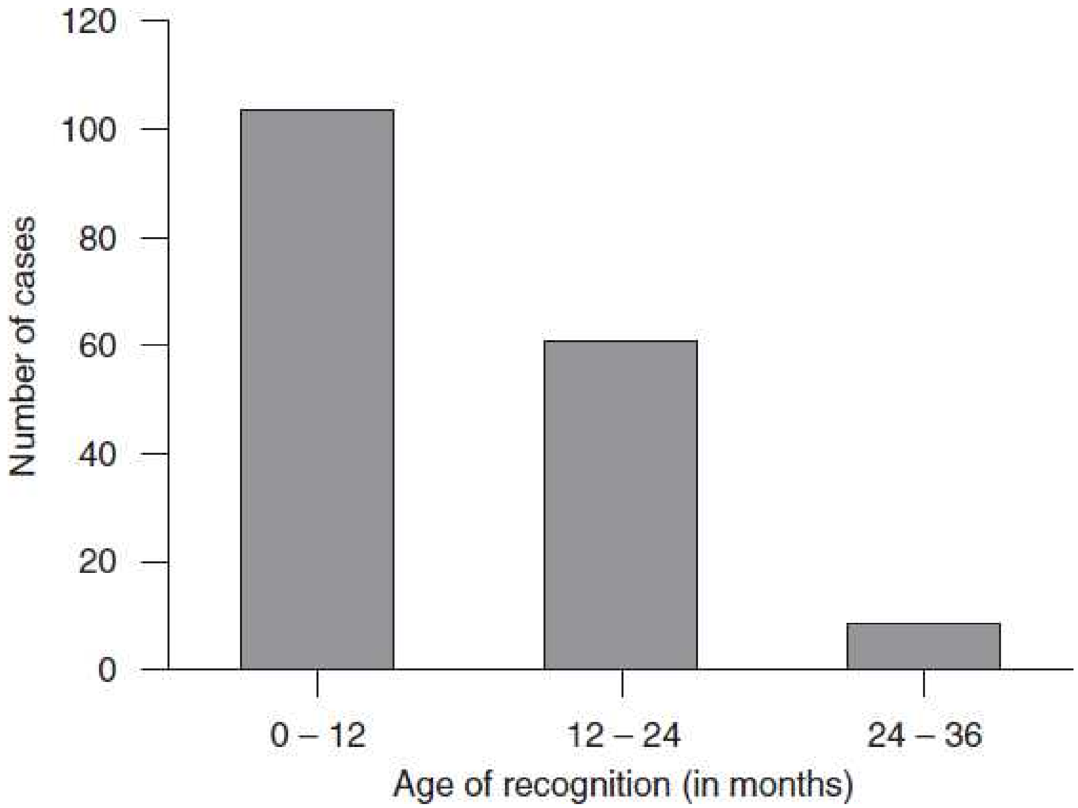

FIGURE 43–2. Age of recognition in autism in a series of 174 consecutive cases.

Impairments in Social Communication

The typically developing infant has a quite striking interest in social interaction and the social environment from the time of birth. This interest is an important foundation for development of a host of other abilities, for example, observational learning, joint attention, selective attention, communication, executive functions, and so forth. In contrast, for the infant/young child with ASD, the face and social interaction are of much less interest. The young child with ASD may have trouble engaging in some of the early social games of infancy. Social interest usually does increase in ASD with intervention, but problems in social interaction remain a hallmark of the condition throughout the life span.

Problems in language and more generally in communication are a hallmark of ASD. Language may fail to develop at all, or when it does, various difficulties often emerge. These include tendencies to echo speech (echolalia); this reflects the child's tendency toward a gestalt learning style (learning things in chunks). Other problems may include difficulties with tone of voice, loudness of speech, problems with pronouns, and reciprocal language. Very idiosyncratic language may develop, that is, specific words or phrases that only the parents understand. Although the syntax and morphology of language can be relatively spared, acquisition of vocabulary and semantic skills may be delayed and social language (pragmatics) difficulties are particularly challenging. As children become older, irony, sarcasm, and humor may be areas of difficulty.

Although prosody is quickly acquired in typically developing infants, those with autism have major difficulties with it, and intonation can be monotonic, loud, and robot like. Perception of social cues is an area of difficulties, for example, the child may not perceive cues from a conversational partner and engage in a one-sided conversation.

Table 43–4.

Symptoms Differentiating Infants and Toddlers with Autism from Typical and Developmentally Delayed Peers

	Social Interaction	Communication	Stereotypical and Repetitive Behaviors
First year	Limited ability to anticipate being picked up	Poor response to name	Excessive mouthing
	Low frequency of looking at people	Infrequent looking at objects held by others	Aversion to social touch
	Little interest in interactive games		
	Little affection toward family members		
	Content to be alone		
Second and third years	Abnormal eye contact	Low frequency of verbal or nonverbal communication	Hand/finger mannerisms
	Limited social referencing	Failure to share interest (e.g., by pointing, giving, showing)	Inappropriate object use
	Limited interest in other children		Repetitive interests/play
	Limited social smile		Unusual sensory behavior (hyper/hyposensitivity to sounds, texture, taste, or visual stimuli)

Low frequency of looking at people	Failure to respond to communicative gestures (showing, pointing, giving)	Poor response to name
Limited sharing of affect/enjoyment	Use of other’s body as tool (takes hand to get object without making eye contact as if hand is agent)	
Limited range of facial expressions	Unusual vocalizations	
Unusual attachment object (hard not soft, the specific object less important than category of object)		
Little interest in interactive games		
Limited functional play		
No pretend play		
Limited motor imitation		

Adapted from Chawarska K, Volkmar FR. Autism in infancy and early childhood. In: Volkmar FR, Klin A, Paul R, Cohen D, eds. *Handbook of Autism and Pervasive Developmental Disorders*. 3rd ed. John Wiley & Sons; 2005:230.

Problems in play are common, particularly with more symbolic–imaginative play. The child with ASD may be more engaged with the taste or smell or feel of play materials or their use for self-stimulatory activity than their intended use.

Restricted Activities and Interests

It is typical for the child with ASD to have problems with change and variation in routine. Attempts to interrupt some activity may lead to catastrophic distress on the part of the child. Parents may note the child's overreliance on routine and their difficulties in dealing with change. Often areas of engagement and interest remain highly restricted, for example, collecting strings and using them for self-stimulation, memorizing numbers, etc. When the younger child with ASD has an attachment object, it often contrasts with that of such objects in the typically developing child in that it may be hard rather than soft and the actual object is less important than the category of objects, for example, any firetruck or train will do. Repetitive and stereotyped movements frequently include toe walking, finger flicking, and body rocking. The child may enjoy spinning objects or may be entranced by repetitive visual displays, for example, of a ceiling fan rotating. Repetitive behaviors, in their more full-blown form, are often among the last to develop in younger children but usually are seen by 3 years of age.

Associated Features

Although initial conceptualizations by Leo Kanner and Hans Asperger highlighted the cognitive potential of individuals with ASD, up to 33% of children on the autism spectrum likely now meet criteria for intellectual disability (IQ ≤ 70), with another 24% in the borderline range (IQ 71 to 85). Intellectual disability appears to be more prevalent in girls than in boys, although it is unclear if this finding reflects a true sex/gender difference or challenges with diagnosing ASD in more cognitively able females.

Performance in IQ tests in individuals with ASD often reflects an uneven cognitive profile, with peaks and valleys at the subtest level. These include relative strengths in visual–spatial and rote memory tasks and relative weakness in verbal comprehension and abstract reasoning, suggesting a “concrete” cognitive style and difficulties going beyond literal information to infer abstract concepts. However, this pattern might be absent or even reversed across different subgroups, including individuals with higher cognitive and language functioning, who might show superior performance in verbal versus perceptual reasoning. Importantly, cognitive tests standardized on typical populations do not always provide an accurate estimate of intelligence in individuals with ASD, particularly for children with lower verbal skills and compliance, requiring the combination of different sources of information to estimate cognitive functioning in this population. Regardless of IQ level, individuals on the autism spectrum often have difficulties translating their cognitive capacity into real-life competencies, as reflected in their adaptive functioning skills lagging behind mental age. A related phenomenon observed in ASD is the presence of “savant” or “splinter” skills—that is, outstanding competencies above the individual’s general level of ability (e.g., normative reading skills in a child with below-average verbal IQ) and/or above the population norm (e.g., calendar calculation skills that are not normally present in the typical population). These exceptional talents, estimated to be present in 10% to 50% of the ASD population depending on operational criteria, are often not capitalized upon to facilitate adaptive functioning (e.g., using exceptional drawing skills to build a successful career as a graphic designer). Fortunately, the appreciation of the unique talents and assets offered by individuals with ASD in the workplace is increasingly appreciated, offering the potential to reverse this trend.

Developmental accounts of ASD have been propelled by prospective research on children at familial risk for autism, whereby younger siblings of a child with autism are followed from birth to examine antecedents of autism symptoms and diagnosis. One prominent model generated by this literature is that early emerging differences in the social domain, including reduced engagement in social play routines, reduced sharing and responsiveness to emotions, and reduced frequency of spontaneous imitation, reflect a fundamental deficit in social motivation or social attention. According to this notion, the difficulties in understanding and relating to the social world in ASD are the downstream outcome of reduced interest in social stimuli during infancy. Evidence in support of this account includes the findings of reduced expressions of pleasure during social interaction, as well as reduced attention toward biologic motion, social scenes, and eye gaze during early development in children with ASD. Early emerging deviations in social attention might constrain engagement in social experiences, thus altering the developmental trajectories of neural specialization and behavioral expertise in the social domain. However, it remains disputed whether this developmental cascade originates from alterations in reflex-like social orienting mechanisms that are present from birth, or later disruptions in the developmental transition from reflexive to volitional attentional social engagement during the first months of life. In addition, the notion of fundamental alterations specific to the social domain has been challenged by findings pointing to widespread network abnormalities that are not specific to the social domain in infants with ASD. Accordingly, it has been proposed that autism symptoms might reflect the downstream consequence of widespread alterations in synaptic processing, possibly disrupting social processing due to the complex and transient nature of social stimuli compared to nonsocial aspects of the environment.

Work using eye-tracking technology has been employed to study aspects of social engagement in a more quantitative way. The study of preferential looking has been extended from adults to very young children. Initial work with adults (see Fig. 43–3) illustrated the marked differences in patterns of looking while viewing social scenes—essentially resulting in difficulties for the person with ASD to fully capture the complexity of the interaction. Other examples include differences in ways biologic motion is perceived. These results highlight the significance of early-onset differences in social information processing on subsequent development.

Currently, no consensus has been established on whether “social-first” or “multiple deficits” models provide the most viable account of the pathways from early alterations to symptom expression in ASD. However, regardless of causal pathways, the notion of early emerging alterations leading to downstream developmental deviations and delays has been influential to early diagnosis and intervention, propelling efforts to identify and pre-emptively address early emerging manifestations of ASD before children are fully symptomatic.

Despite the theoretical advances from “static” to developmental models of ASD capable of informing practice, the establishment of a comprehensive, unitary account of ASD is hindered by several conceptual and methodologic challenges. These include the definitional haze around social processes and mechanisms, with unclear overlap and fuzzy boundaries around analogous constructs (e.g., social reciprocity, social orienting, social reward, social motivation, social cognition), and challenges with tying explanations based on cognitive and behavioral processes to molecular mechanisms of typical and atypical development. Accordingly, research efforts are increasingly focused on the interplay between genetic, neural, cognitive, and behavioral processes underlying the development of ASD.

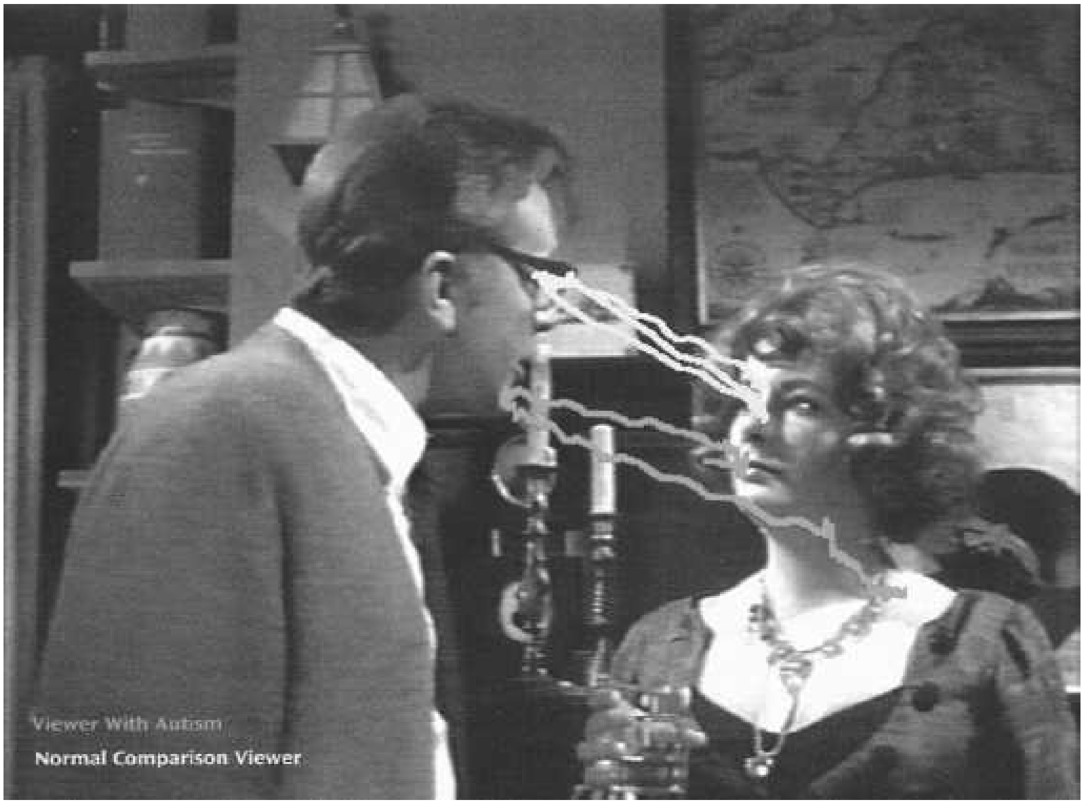

FIGURE 43–3. Eye-tracking data from a typically developing young adult (**top line**) and a young, high-functioning man with autism (**bottom line**). The man with autism focuses almost exclusively on the mouth of the person speaking, avoiding the upper part of the face and not noticing the emotional responses of the listener. (Reprinted with permission from Klin A, Jones W, Schultz R, Volkmar F, Cohen D. Defining and quantifying the social phenotype in autism. *Am J Psychiatry* 2002;159(6):895–908.)

Motor and Sensory Abnormalities. A range of motor abnormalities has been documented in ASD, including delayed walking, abnormalities in gait (e.g., toe walking), posture, balance, and coordination, as well as difficulties in motor imitation and pantomime. In addition, individuals with autism often display motor stereotypies, including repetitive tapping, hand and arm flapping, body rocking, finger flicking, waving in front of the face, rigid tensing of the entire body, as well as atypical patterns of hyper- and/or hypomotor activity. Although these abnormalities are not ASD specific, they are observed more frequently in individuals on the autism spectrum than in other conditions. The developmental course of motor abnormalities in ASD involves early subtle manifestations during the first 2 years of life, followed by increasingly deviant motor trajectories during childhood. Motor abnormalities might decrease in adolescence and adulthood, particularly in more cognitively able individuals, although motor clumsiness and awkward postures or movements might be observed throughout the life course and across the spectrum of ASD severity.

While the nature of motor abnormalities in ASD is poorly understood, these appear to be interrelated with the pathognomonic features of ASD. Differences in motor imitation and motor gestures in infants and toddlers have been found to predict diagnostic status and communication outcomes, reflecting an interplay between motor and social dimensions of development in ASD in the first years of life. In addition, motor and social abnormalities might continue to exert reciprocal effects across the lifespan. For example, school-aged children who engage less frequently or less accurately in motor imitation have fewer opportunities to join peers in coordinated activities that facilitate social bonding (e.g., joint drumming, giving “high fives”). Similarly, teenagers with ASD who engage more frequently in solitary versus social activities are less likely to adopt the postures and movement patterns of those surrounding them, resulting in idiosyncratic motor behaviors. In addition, atypical motor development has been linked to arousal abnormalities, rigidity/inflexibility, developmental dyspraxia, functional and anatomical abnormalities, as well as abnormal connectivity in the brain systems underlying motor planning and execution. Individual differences are substantial, suggesting multifactorial and heterogeneous causal paths.

Another feature that has been frequently documented in ASD is an atypical perception of and response to sensory information. Sensory abnormalities encompass a range of phenomena, including seeking of specific sensory stimulations (e.g., deep pressure), unusual sensory interests (e.g., peering at or smelling objects), as well as overreaction and/or underreaction to sensory inputs. For example, a child with ASD might experience certain sounds as being excessively loud or certain textures as being aversive, resulting in idiosyncratic reactions such as covering the ears in response to a vacuum cleaner, refusing foods with specific textures, and defensively avoiding being touched. Conversely, hyporeactivity might result in children not responding to specific sensory stimulations, such as pain, or their name being called, and seeking sensory inputs in idiosyncratic ways, such as rubbing a piece of fabric in their hands, squeezing objects, or staring at the circular motion of clothes in a washing machine. Hypo- and hypersensitivity might coexist in the same individual in response to different classes of stimuli and at different points in development. Sensory abnormalities have been observed across the spectrum of ASD severity but appear to be more frequent and/or pronounced in more severely impaired individuals. In addition, atypical sensory responsiveness in ASD has been found to be associated with anxiety, although causal paths are unclear. While these phenomena appear to be more prevalent in ASD compared to other conditions, heterogeneity is substantial, pointing to the importance of individualized intervention strategies.

PHYSICAL CHARACTERISTICS

In ASD, higher rates of minor physical anomalies are observed, for example, ear malformations. These may reflect the embryologic period in which older regions of the brain develop. In general, though, children with autism are described as attractive and without dysmorphic features. Interestingly, a significant number of toddlers with autism develop macrocephaly.

PATHOLOGY AND LABORATORY EXAMINATION

Electroencephalography and Seizure Disorders

Various EEG abnormalities are observed even in the absence of clinical seizures (see Fig. 43–4). This is particularly the case when ASD is associated with intellectual deficiency. The observation of EEG differences in ASD for looking at faces as compared to typical controls has emerged as one potential biomarker of ASD.

As noted previously, seizures can have their onset at any time in the developmental period. Individuals who are more cognitively impaired are likely at greatest risk.

Neuroimaging Studies

As noted previously, a large body of research now exists on the social brain using neuroimaging techniques. At present, in the absence of a specific clinical indication, neuroimaging studies are not routinely needed in the usual evaluation of a child with possible ASD.

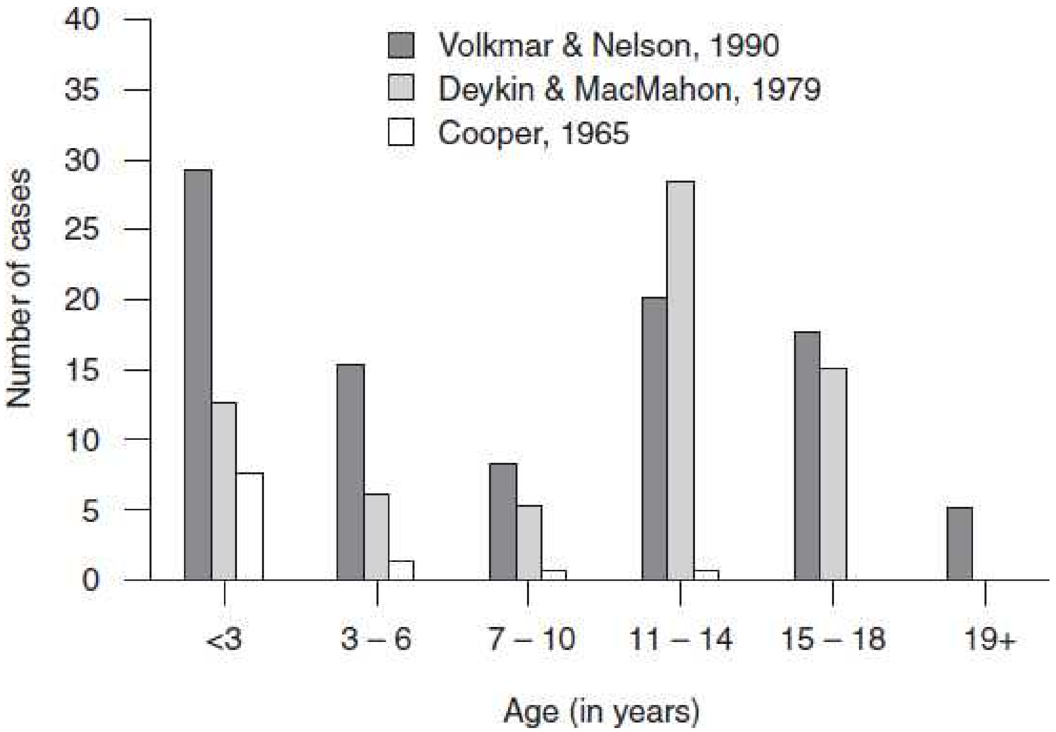

FIGURE 43–4. Rates of first seizure (incidence rate per 1,000) in samples of individuals with autism. (Data from cases with autism from Volkmar FR, Nelson DS. Seizure disorders in autism. *J Am Acad Child Adolesc Psychiatry*. 1990;29:127–129; and from cases with autism/pervasive developmental disorder from Deykin EY, MacMahon B. The incidence of seizures among children with autistic symptoms. *Am J Psychiatry*. 1979;136:1310–1312; as compared to rates in a normative sample from Cooper JE. Epilepsy in longitudinal survey of 5,000 children. *Br Med J*. 1965;1:1020–1022.)

Laboratory Studies

As indicated by history and examination, various laboratory tests may be relevant as part of the assessment. Screening should be done for frequently associated genetic conditions. Indications are summarized in practice guidelines provided in the reference list and these now identify such factors in between 10% and 20% of cases. EEG is indicated when seizures are suspected. Issues of assessment are addressed in various sets of practice guidelines.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of ASD includes other neurodevelopmental disorders, such as intellectual deficiency or language–communication disorders, as well as primary intellectual deficiency. Visual and hearing difficulties should also be evaluated. Over 40 diagnostic instruments and screeners have been developed, but it must be noted that these do not replace good clinical judgment. These can be based on parent, teacher, or clinical report, or on direct examination. Some of these instruments require extensive training. Selected instruments are summarized in Table 43–5. Some young children do not exhibit all features until around age 3.

For the child with mild to moderate intellectual deficiency (but without ASD), social skills are usually consistent with development in areas of cognition and communication. In individuals with severe and profound retardation, the frequency of both autism and autistic-like behaviors increases but present some diagnostic challenges. Sometimes the presence of isolated stereotyped behavior is mistakenly used as a single factor in diagnosis of autism. At the other end of the IQ range, there can be confusion relative to those personality disorders that involve social isolation.

Aids to memory such as videotapes or baby books may help parents recall important historical information. As noted previously, in a small number of cases, the child may seem to develop normally but then either lose skills or plateau as more typical autistic behaviors develop. Other conditions to consider with a history of loss or developmental stagnation include mutism, Rett syndrome, the DSM-IV CDD, language disorder, schizophrenia, and degenerative central nervous system (CNS) disorders. Severe neglect or deprivation, for example, in some institutional rearing settings, can produce some features suggestive of autism, although usually not all features are present and social difficulties tend to dramatically remit if appropriate care is provided. Childhood-onset schizophrenia is rare before age 5 and indeed before puberty. In addition, the more classical features of schizophrenia are present, albeit in a more developmentally appropriate form. In some instances, the diagnosis can only be clarified with certainty over time.

Table 43–5.
Summary of Selected Assessment Instruments for Autism Spectrum Disorder

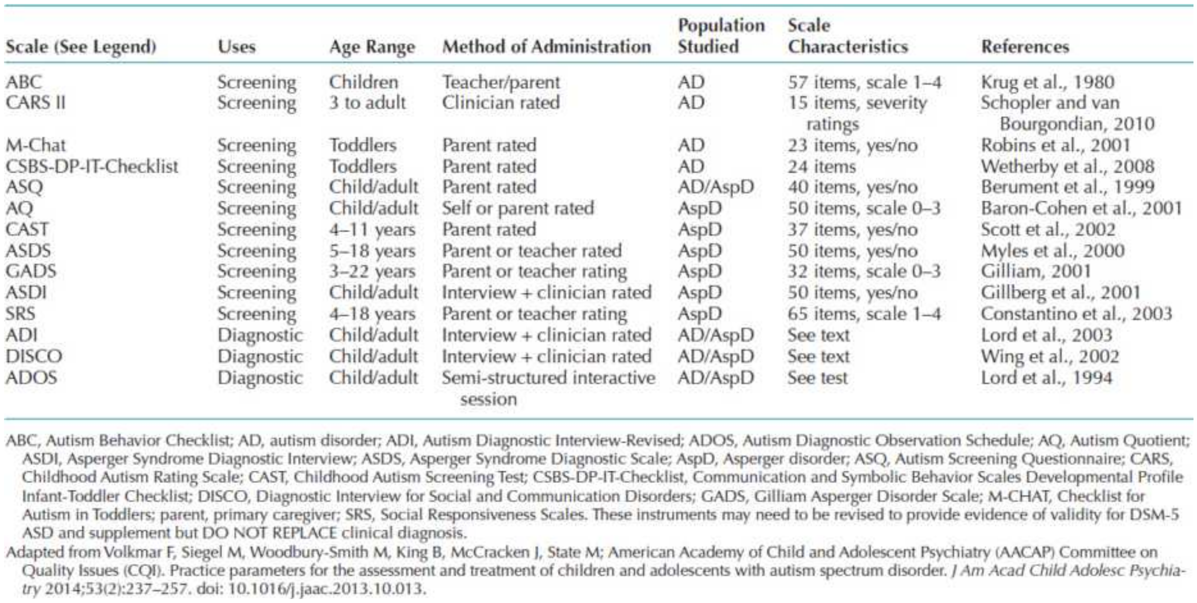

Scale (See Legend)	Uses	Age Range	Method of Administration	Population Studied	Scale Characteristics	References
ABC	Screening	Children	Teacher/parent	AD	57 items, scale 1–4	Krug et al., 1980
CARS II	Screening	3 to adult	Clinician rated	AD	15 items, severity ratings	Schopler and van Bourgondian, 2010
M-Chat	Screening	Toddlers	Parent rated	AD	23 items, yes/no	Robins et al., 2001
CSBS-DP-IT-Checklist	Screening	Toddlers	Parent rated	AD	24 items	Wetherby et al., 2008
ASQ	Screening	Child/adult	Parent rated	AD/AspD	40 items, yes/no	Berument et al., 1999
AQ	Screening	Child/adult	Self or parent rated	AspD	50 items, scale 0–3	Baron-Cohen et al., 2001
CAST	Screening	4–11 years	Parent rated	AspD	37 items, yes/no	Scott et al., 2002
ASDS	Screening	5–18 years	Parent or teacher rated	AspD	50 items, yes/no	Myles et al., 2000
GADS	Screening	3–22 years	Parent or teacher rating	AspD	32 items, scale 0–3	Gilliam, 2001
ASDI	Screening	Child/adult	Interview + clinician rated	AspD	50 items, yes/no	Gillberg et al., 2001
SRS	Screening	4–18 years	Parent or teacher rating	AspD	65 items, scale 1–4	Constantino et al., 2003
ADI	Diagnostic	Child/adult	Interview + clinician rated	AD/AspD	See text	Lord et al., 2003
DISCO	Diagnostic	Child/adult	Interview + clinician rated	AD/AspD	See text	Wing et al., 2002
ADOS	Diagnostic	Child/adult	Semi-structured interactive session	AD/AspD	See test	Lord et al., 1994

ABC, Autism Behavior Checklist; AD, autism disorder; ADI, Autism Diagnostic Interview-Revised; ADOS, Autism Diagnostic Observation Schedule; AQ, Autism Quotient; ASDI, Asperger Syndrome Diagnostic Interview; ASDS, Asperger Syndrome Diagnostic Scale; AspD, Asperger disorder; ASQ, Autism Screening Questionnaire; CARS, Childhood Autism Rating Scale; CAST, Childhood Autism Screening Test; CSBS-DP-IT-Checklist, Communication and Symbolic Behavior Scales Developmental Profile Infant-Toddler Checklist; DISCO, Diagnostic Interview for Social and Communication Disorders; GADS, Gilliam Asperger Disorder Scale; M-CHAT, Checklist for Autism in Toddlers; parent, primary caregiver; SRS, Social Responsiveness Scales. These instruments may need to be revised to provide evidence of validity for DSM-5 ASD and supplement but DO NOT REPLACE clinical diagnosis.

Adapted from Volkmar F, Siegel M, Woodbury-Smith M, King B, McCracken J, State M; American Academy of Child and Adolescent Psychiatry (AACAP) Committee on Quality Issues (CQI). Practice parameters for the assessment and treatment of children and adolescents with autism spectrum disorder. *J Am Acad Child Adolesc Psychiatry* 2014;53(2):237–257. doi: 10.1016/j.jaac.2013.10.013.
]

John was the second of two children born to middle-class parents after normal pregnancy, labor, and delivery. As an infant, John appeared undemanding and relatively placid; motor development proceeded appropriately, but language development was delayed. Although his parents indicated that they were first concerned about his development when he was 18 months old and still not speaking, in retrospect, they noted that, in comparison to their previous child, he had seemed relatively uninterested in social interaction and the social games of infancy. Stranger anxiety had never really developed, and John did not exhibit differential attachment behaviors toward his parents. Their pediatrician initially reassured John's parents that he was a "late talker," but they continued to be concerned. Although John seemed to respond to some unusual sounds, the pediatrician obtained a hearing test when John was 24 months old. Levels of hearing appeared adequate for development of speech, and John was referred for developmental evaluation. At 24 months, motor skills were age appropriate, and John exhibited some nonverbal problem–solving skills close to age level. His language and social development, however, were severely delayed, and he was noted to be resistant to changes in routine and unusually sensitive to aspects of the inanimate environment. His play skills were quite limited, and he used play materials in unusual and idiosyncratic ways. His older sister had a history of some learning difficulties, but the family history was otherwise negative. A comprehensive medical evaluation revealed a normal EEG and computed tomography (CT) scan; genetic screening and chromosome analysis were normal, as well.

John was enrolled in a special education program, in which he gradually began to speak. His speech was characterized by echolalia, extreme literalness, a monotonic voice quality, and pronoun reversal. He rarely used language in interaction and remained quite isolated. By school age, John had developed some evidence of differential attachments to family members; he also had developed a number of self-stimulatory behaviors and engaged in occasional periods of headbanging. Extreme sensitivity to change continued. Intelligence testing revealed marked scatter, with a full-scale IQ in the moderate range of intellectual deficiency. As an adolescent, John's behavioral functioning deteriorated, and he developed a seizure disorder. Now an adult, he lives in a group home and attends a sheltered workshop. He has a rather passive interactional style but exhibits occasional outbursts of aggression and self-abuse.

COURSE AND PROGNOSIS

Typically, ASD is a disability that is lifelong although with a considerable range in adult functional outcome. Over time, fewer and fewer individuals have needed the considerable interpersonal and familial support observed in the first cohorts of cases. It appears that with earlier and more evidence-based intervention, overall outcome is improving. There are some cases of "optimal outcome" where individuals technically no longer meet full diagnostic criteria (although some interpersonal oddity or other difficulties may remain). Some cases, despite early identification and good intervention, do not do well.

More people with ASD are entering the workforce and attending college but may face their own challenges in these settings. More individuals with ASD are now living independently and some are marrying. Initial outcome studies suggested rates of adult independence at around perhaps 5% of cases; this number has dramatically increased, as has the number of individuals who are living semi-independently. Vocational challenges often remain. Associated mental health issues, for example, anxiety and depression, can be a source of additional difficulty for adults. Even adults with more severe associated intellectual disability can make gains, and longer-term goals here include acquisition of important adaptive skills to maximize self-sufficiency to the degree possible.

Factors associated with better long-term outcome include presence of communicative speech by the age of 5 or and nonverbal IQ above 70. Cognitive ability that is overall within the normal range is a positive prognostic sign. Treatment studies of large groups of younger children will typically show a few children who make very dramatic gains in the first years of life. Several studies have now demonstrated observable brain changes in response to treatment using techniques like EEG or fMRI response to social stimuli.

During the elementary school years, gains are usually observed in social compliance and communication, particularly if good evidence-based treatments are provided. Kanner was the first to note that, during adolescence, some children make major gains while a somewhat smaller number deteriorate. Sometimes gains in adolescence are associated with a greater interpersonal awareness on the part of the teenager who has a desire to fit in; this both motivates change but also can be a source of anxiety and depression.

TREATMENT

No treatment has unequivocally proven to reverse the core symptoms of ASD. However, numerous interventions for ASD have shown beneficial effects, including improvements in language, communication, cognitive, social, and adaptive functioning. Effective approaches rely on various applications of behavioral, developmental, and special education principles. Given the heterogeneous nature of the autism spectrum, intervention goals must be individualized, using comprehensive evaluation procedures to tailor objectives to the individual's profile of strengths and needs. Treatment goals must be updated periodically based on progress and consideration of evolving age- and context-specific needs and challenges. The intensity and timing of intervention are important factors in shaping outcomes, with children receiving intervention at a younger age and more intensively (i.e., 15+ hours per week), showing better intervention response. However, individuals with ASD benefit from evidence-based intervention at all ages, including adults who did not receive services during childhood.

Many evidence-based practices in ASD use principles and procedures derived from applied behavior analysis (ABA), a discipline focused on the examination and manipulation of the contingencies that create/maintain specific behaviors (e.g., self-aggression in response to specific demands), with the goal of discouraging maladaptive responses and teaching new skills. In the comprehensive ABA-based approach known as Early Intensive Behavioral Intervention (EIBI), the acquisition of new skills is facilitated through a structured learning environment designed to minimize distractions, with targeted skills broken down into discrete behavioral units that are taught in isolation. Stripped-down directive instructions are repeated until the child’s behavioral response meets a predefined criterion, and correct responses are systematically reinforced, while various error correction procedures are used to discourage incorrect responses.

Other interventions embed ABA principles within naturalistic developmental behavioral approaches. These emphasize the importance of responding to the child’s cues, building learning episodes on the child’s initiative and choices, establishing emotional attunement and positive relationship between child and adult, and creating learning environments that resemble the “real world” where targeted skills would naturally be utilized. In addition, several intervention practices focus on the importance of environmental accommodations to create “autism-friendly” learning experiences, particularly in educational settings and workplaces. These include arranging spaces and using visual supports that capitalize on the visual processing strengths of those with ASD to facilitate understanding of expectations and communication (e.g., using pictures that illustrate what is expected in a specific space, or how to accomplish a specific task). Relatedly, many intervention programs include the use of visual and/or augmentative approaches to facilitate and teach communication. These include the use of visual cues (e.g., giving a partner a picture of an item in exchange for the actual item) or electronically generated sounds (e.g., pushing a button in a device that “speaks electronically” to request an item) that are designed to address difficulties in understanding language or using speech.

Although these intervention approaches are sometimes depicted as mutually exclusive, they are better characterized as falling along a multidimensional continuum from more structured to more naturalistic, from more adult directed to more child directed, from more focused on changing behavior to more focused on accommodating differences, and from more comprehensive to more targeted on specific skills. A growing body of literature has examined outcomes for different intervention packages along this continuum, identifying several elements that are in common across approaches shown to be effective. These include (1) timeliness (the earlier, the better), (2) intensity, (3) use of manualized and clearly operationalized instructional strategies, (4) focus on the core features of ASD, including sensory issues, as well as functional/adaptive skills, (5) assessment procedures to generate and update individualized sets of goals, (6) collection systems to monitor progress and implementation fidelity, (7) involvement of caregivers in establishing treatment goals and delivering intervention, (8) developmental appropriateness and developmental sequencing of goals (i.e., teaching age-appropriate skills and their developmental prerequisites), (9) promoting the individual’s active engagement and communicative intent, (10) incorporating intervention strategies and opportunities in everyday routines, (11) thoughtful environmental arrangements (including minimizing distractions/maximizing focus toward salient stimuli), (12) creating predictable and consistent routines and experiences, with a balance between structure and novelty to support flexibility, (13) incorporating the individual’s special interests when possible, and provision of reinforcements that are meaningful to the individual, (14) use of task analysis to break down complex functional skills into teachable steps, (15) use of scripts, visual supports, and practice opportunities to remind the individual about the steps involved in the targeted skill (e.g., rehearsing the steps involved in buying groceries through visual scripts and supported role-playing), (16) focus on generalization skills (e.g., using the skills acquired in a therapy sessions at home and school), (17) addressing challenging behaviors through the systematic analysis of their function, and teaching more desirable behavioral alternatives, (18) using prompts to facilitate the individual’s response to the instruction, while also fading supports over time to promote autonomy.

Although research has documented the beneficial effects of these practices, there are important gaps that require additional research effort. First, despite the increased number of studies evaluating interventions for younger children, there is limited knowledge on interventions for older individuals, especially those on the most severe end of the spectrum. In addition, most existing literature has tested the efficacy of interventions in the highly resourced context of clinical trials, with more limited knowledge on intervention outcomes when these are transported into community clinical and educational programs.

In addition, although contemporary approaches have shifted away from the use of psychodynamic psychotherapy to address autistic symptoms, cognitive behavioral therapy (CBT) has been shown to effectively address co-occurring anxiety and depression symptoms that are common in children and adults with ASD. Several modifications have shown to facilitate the success of CBT for those with ASD, including using visual and multimodal teaching strategies, a focus on explicit problem-solving behaviors, and incorporating special interests into sessions. There may be an increased risk for suicidal ideation and suicide, and clinicians should be alert to this issue.

Psychopharmacology

The core social difficulty of ASD has, so far, proven resistant to treatment, although work is ongoing with agents such as oxytocin (a hormone involved in attachment). Medication can be of significant benefit for other features such as irritability, overactivity, aggression, and self-injury. Since children and adolescents are often enrolled in special educational programs, there may be opportunities to facilitate independent data collection of value in assessment of drug treatments. Potential benefits and side effects should be considered (e.g., with antipsychotic medications), and baseline laboratory studies (e.g., liver, kidney function, ECG, weight, and so forth) should be obtained. Depending on the agent used, it may be important to closely monitor for side effects, for example, tardive dyskinesia.

Antipsychotics. These agents have been extensively studied and, with relatively low doses, they may produce decreased irritability and agitation. This in turn may foster greater engagement in a treatment program. It is important to monitor for side effects. More recently atypical neuroleptics have been utilized. In one large double-blind, placebo-controlled study, risperidone was shown to quickly have a significant benefit in reducing agitation and other problematic behaviors. This and other atypical neuroleptics have now been approved for use in autism.

Selective Serotonin Reuptake Inhibitors. Interest in this group of agents arose given the awareness of higher peripheral serotonin levels in autism as well as in the increased risk for anxiety and depression. Several studies using placebo control methods have shown benefits for repetitive behaviors and impulsive aggression. Some children may have a paradoxical agitation.

Stimulants. Many school-aged children exhibit significant problems with attention and overactivity that may be responsive to the traditional stimulants and related medications. However, associated side effects, for example, agitation, may be higher than in the general pediatric population.

Other Agents. A number of other agents have been used in ASD. Often data is limited to case reports or open-label studies, and controlled studies are not available. Lithium has occasionally been used, particularly with family history of bipolar illness and clinical presentation suggesting possible bipolar disorder.

Complementary and Alternative Treatments

These are frequently used by parents although, by definition, they lack a strong evidence base. This evidence is based on anecdote and sometimes a case series. One claim was that the gut peptide secretin was suggested to produce marked improvement, but a number of double-blind, controlled studies, did not show benefit over placebo. Parents should be encouraged to pursue proven treatments and avoid those with some risk to the child, either directly due to side effects or more commonly due to lack of participation in effective treatment programs provided in schools. Some resources for parents on these treatments are available (see Volkmar and Wiensser, 2021).

FURTHER READINGS

Asperger H. Autistic psychopathy in childhood. In: Frith U, ed. *Autism and Asperger Syndrome*. Cambridge University Press; 1991:37.

Chawarska K, Macari S, Volkmar FR, Kim SH, Shic F. ASD in infants and toddlers. In: Volkmar FR, Rogers SJ, Paul R, Pelphrey KA, eds. *Handbook of Autism and Pervasive Developmental Disorders*. 4th ed. John Wiley & Sons; 2014.

Chown N. *Understanding and Evaluating Autism Theory*. Jessica Kingsley; 2017.

Hazlett HC, Gu H, Munsell BC, et al; IBIS Network, Clinical Sites, Data Coordinating Center, Image Processing Core, Statistical Analysis. Early brain development in infants at high risk for ASD spectrum disorder. *Nature*, 2017;542(7641):348–351.

Hazlett HC, Poe MD, Gerig G, et al. Early brain overgrowth in autism associated with an increase in cortical surface area before age 2 years. *Arch Gen Psychiatry*. 2011;68(5):467–476.

Heller T. Dementia infantilis. *Zeitschrift Erforschung Behandlung Jugenlichen Schwachsinns*. 1908;2:141–165.

Hermelin B. *Bright Splinters of the Mind: A Personal Story of Research with Autistic Savants*. Jessica Kingsley; 2001.

Herrington JD, Schultz RT. Neuroimaging of developmental disorders. In: Shenton ME, Turetsky BI, eds. *Understanding Neuropsychiatric Disorders: Insights From Neuroimaging*. Cambridge University; 2011.

Howlin P. Adults with autism: changes in understanding since DSM-III. *J Autism Dev Disord*. 2021;51(12):4291–4308.

Huerta M, Bishop SL, Duncan A, Hus V, Lord C. Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of pervasive developmental disorders. *Am J Psychiatry*. 2012;169(10):1056–1064.

Ingersoll B, Wainer A. The Broader Autism Phenotype. In: Volkmar FR, Rogers SJ, Paul R, Pelphrey KA, eds. *Handbook of Autism and Pervasive Developmental Disorders*, 4th ed. John Wiley & Sons, Inc; 2014.

Kanner L. Autistic disturbances of affective contact. *Nerv Child*. 1943;2:217–250.

Klin A, Jones W, Schultz R, Volkmar F, Cohen D. Defining and quantifying the social phenotype in autism. *Am J Psychiatry*. 2002;159:895–908.

Lord C. Follow-up of two-year-olds referred for possible autism. *J Child Psychol Psychiatry*. 1996;36:1065–1076.

Lord C, Corsello C, Grzadzinski R. Diagnostic instruments in autistic spectrum disorders. In: Volkmar FR, Rogers SJ, Paul R, Pelphrey KA, eds. *Handbook of Autism and Pervasive Developmental Disorders*. 4th ed. John Wiley & Sons; 2014.

Lyall K, Schmidt RJ, Hertz-Picciotto I. Environmental factors in the preconception and prenatal periods in relation to risk for ASD. In: Volkmar FR, Rogers SJ, Paul R, Pelphrey KA, eds. *Handbook of Autism and Pervasive Developmental Disorders*. 4th ed. John Wiley & Sons; 2014.

Maenner MJ, Shaw KA, Baio J. Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveillance Summaries*, 2020;69(4):1–12.

McCracken JT, McGough J, Shah B, et al; Research Units on Pediatric Psychopharmacology Autism Network. Risperidone in children with autism and serious behavioral problems. *N Engl J Med*. 2002;347:314–321.

McPartland JC, Pelphrey KA. The implications of social neuroscience for social disability. *J Autism Dev Disord*. 2012;42(6):1256–1262.

National Research Council. *Educating Young Children with Autism*. National Academy Press; 2001

Powers MD, Palmieri MJ, Egan SM, Rohrer JL, Nulty EC, Forte S. Behavioral assessment of individuals with autism: current practice and future directions. In: Volkmar FR, Rogers SJ, Paul R, Pelphrey KA, eds. *Handbook of Autism and Pervasive Developmental Disorders*. 4th ed. John Wiley & Sons; 2014.

Presmanes-Hill A, Zuckerman KE, Fombonne E. Epidemiology of autism spectrum disorders. In: Volkmar FR, Rogers SJ, Paul R, Pelphrey KA, eds. *Handbook of Autism and Pervasive Developmental Disorders*. 4th ed. John Wiley & Sons; 2014.

Robertson, C. E. Baron-Cohen, S. Sensory perception in autism. *Nat Rev Neurosci.* 2017;18(11):671–684.

Rutter M, Bailey A, Bolton P, Le Couteur A. Autism and known medical conditions: myth and substance. *J Child Psychol Psychiatry*. 1994;35(2):311–322.

Rutter M, Thapar A. Genetics of autism spectrum disorders. In: Volkmar FR, Rogers SJ, Paul R, Pelphrey KA. *Handbook of Autism and Pervasive Developmental Disorders*. 4th ed. John Wiley & Sons; 2014.

Schaefer GB, Mendelsohn NJ. Genetics evaluation for the etiologic diagnosis of autism spectrum disorders. [erratum appears in Genet Med. 2008;10(6):464.] *Genet Med*. 2008;10(1):4–12.

Smith IC, Reichow B, Volkmar FR. The effects of DSM-5 criteria on number of individuals diagnosed with autism spectrum disorder: a systematic review. *J Autism Dev Disord*. 2015;45(8):2541–2552.

Tang G, Gudsnuk K, Kuo SH, et al. Loss of mTOR-dependent macroautophagy causes autistic-like synaptic pruning deficits. *Neuron*. 2014;83(5):1131–1143.

Vivanti G, Bottema-Beutel K, Turner-Brown, L. *Clinical Guide to Early Interventions for Children with Autism*. Springer, 2020.

Vivanti G, Messinger D S. Theories of autism and autism treatment from the DSM III through the present and beyond: impact on research and practice. *J Autism Dev Disord*. 2021:51(12):4309–4320.

Volkmar FR, Wiesner LA. *A Practical Guide to Autism*. Wiley; 2021.

Volkmar FR. A brief history of autism. In: Fatemi SH, ed. *The Molecular Basis of Autism*. Springer; 2015:3–14.

Volkmar F, Siegel M, Woodbury-Smith M, King B, McCracken J, State M; American Academy of Child and Adolescent Psychiatry (AACAP) Committee on Quality Issues (CQI). Practice parameters for the assessment and treatment of children and adolescents with autism spectrum disorder. *J Am Acad Child Adolesc Psychiatry*. 2014;53(2):237–257.

44

Attention-Deficit Disorders

44.1 Attention-Deficit/Hyperactivity Disorder

Lily T. Hechtman, M.D., F.R.C.P., A.B.P.N., Margaret D. Weiss, M.D., Ph.D., and Philippe Hwang, M.B., B.Ch., B.A.O., M.Sc.

Attention-deficit hyperactivity disorder (ADHD) is the most common childhood behavioral disorder diagnosed in outpatient settings in the United States. It is a chronic disorder that can cause impairment into adolescence and adulthood. In 2007, Polanczyk and colleagues estimated worldwide prevalence of ADHD in children and adolescents to be 5.29% using a comprehensive meta-analysis including hundreds of articles and more than 100,000 patients. Similarly, a meta-analysis from 2012 by Willicut using full DSM-IV diagnostic criteria found the prevalence of ADHD to be 5.9% to 7.1%. The National Survey of Children's Health (NSCH) showed that 9.4% of U.S. school-aged children had been diagnosed with ADHD by 2016. It has been suggested that the increasing rates of diagnosis and treatment may be due to increased awareness and access to care as opposed to an increase in the number of children in the population meeting criteria for ADHD.

HISTORY

ADHD first was described by George Still in 1902, who wrote about children who were restless, impulsive, and inattentive, with intense affective responses and conduct problems. After the influenza pandemic and the epidemic of encephalitis lethargica from 1919 to 1920, children who survived the flu frequently developed severe behavior problems and are now thought to have suffered organic brain damage. For that reason, the childhood condition was termed "minimal brain damage syndrome," even though brain damage could not be proven. In the early 1960s, the condition was renamed "minimal brain dysfunction." Not only did these terms postulate unproven etiologic mechanisms, but they were also stigmatizing. The ninth revision of the *International Classification of Diseases and Related Health Problems* (ICD-9) and the second edition of the DSM adopted the same descriptive term for the condition—hyperkinetic syndrome of childhood. The 1980 version of the American Psychiatric Association (APA) classificatory system, DSM-III, renamed this diagnosis attention-deficit disorder (ADD) due to research suggesting that the main disability was the difficulty of maintaining sustained attention and problems with impulsivity. Much of the diagnostic criteria for ADHD have remained the same between DSM-IV in 1994, the text revision (DSM-IV-TR) in 2000, and the most recent version (DSM-5) released in May 2013.

DIAGNOSIS AND DEFINITION

The fifth edition of the APA's *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), published in May 2013, describes ADHD under neurodevelopmental disorders as a pattern of inattentive and/or hyperactive–impulsive behavior inconsistent with developmental level which interferes with functioning in social, educational, or work settings. Symptoms are divided into two categories: inattention and hyperactivity/impulsivity. There are five main diagnostic criteria: (1) onset before age 12 years old; (2) duration greater than 6 months; (3) children must have at least six symptoms from the inattention and/or hyperactive/impulsive symptom list, while older adolescents and adults must have at least five; (4) several symptoms must be present in two or more settings and interfere with functioning; and (5) symptoms that do not occur exclusively during the course of schizophrenia or another psychotic disorder and are not better accounted for by another mental disorder, such as depression. DSM-5 has no exclusion criteria for those with autism spectrum disorder (ASD), allowing both autism and ADHD to be diagnosed simultaneously in the same patient. Possible specifiers for the disorder in DSM-5 include combined presentation (criterion met for both inattention and hyperactivity–impulsivity), predominantly inattentive presentation, or predominantly hyperactive/impulsive presentation. Other specifiers describe current severity of the disorder being mild (few to no symptoms in excess of those required), moderate or severe (many symptoms in excess of those required), or to specify if the disorder is in partial remission (fewer than full criteria met in past 6 months, but still results in impairment).

The inattention symptom criteria list includes the following: (a) often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or during other activities; (b) often has difficulty sustaining attention in tasks or play activities; (c) often does not seem to listen when spoken to directly; (d) often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace; (e) often has difficulty organizing tasks and activities; (f) often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort; (g) often loses things necessary for tasks or activities; (h) often easily distracted by extraneous stimuli (or unrelated thoughts for adolescents and adults); and (i) often forgetful in daily activities.

Hyperactivity and impulsivity symptoms criteria include the following: (a) often fidgets with or taps hands or feet or squirms in seat; (b) often leaves seat in situations when remaining seated is expected; (c) often runs about or climbs in situations where it is inappropriate (for older adolescents and adults, feeling restless without the overactivity can apply); (d) often unable to play or engage in leisure activities quietly; (e) often “on the go,” acting as if “driven by a motor"; (f) often talks excessively; (g) often blurts out an answer before a question has been completed; (h) often has difficulty waiting their turn; and (i) often interrupts or intrudes on others.

In the DSM-5, enhanced item descriptors were added to diagnostic criteria to demonstrate symptoms across the lifespan. For example, motor hyperactivity appears far less frequently in affected adolescents and adults, who continue to have attentional and organizational problems, poor anger management, job instability, and problems with social relationships and self-esteem. All 18 of the DSM-IV-TR symptoms of ADHD were retained in the DSM-5, but enhanced elaborators of each were added to better describe difficulties experienced by adolescents and adults. For instance, the inattention criteria "often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort" now includes specific examples of this symptom "(preparing reports, completing forms, reviewing lengthy papers)" to be applied to older adolescents and adults in addition to "(schoolwork or homework)," which is primarily specific to children.

COMPARATIVE NOSOLOGY

Previously, ICD-10 criteria differed from DSM-5 criteria significantly, with fewer children diagnosed with ADHD under ICD classification (called "hyperkinetic disorder" [HD]). In ICD-10, if a diagnosis of generalized anxiety or major depression was present, these disorders would be considered primary and HD would be excluded. For example, investigators in the United Kingdom applied an algorithm to the 579 children with DSM-IV ADHD combined subtype who had been recruited into the National Institute of Mental Health (NIMH) Multimodal Treatment Study of Attention Deficit Hyperactivity Disorder (MTA trial) in order to generate an ICD-10 diagnosis of HD. Only 25% of the MTA sample continued to meet the diagnostic criteria for HD in ICD-10.

However, the ICD-11 classification of ADHD is closer to that found in the DSM-5, removing this exclusion criterion. Therefore, the rate of diagnosis may increase following the adoption of the ICD-11 despite the incidence remaining unchanged. Both diagnoses are now based on similar structures with presentations that are predominantly inattentive, hyperactive, or combined, and that have been present for over 6 months. The DSM-5 focuses on meeting criteria out of lists of symptoms, whereas the ICD-11 uses more generalized symptoms for each presentation, but both specify that these symptoms must be evident in multiple settings. The ICD-11 specifically describes essential characteristics of the disorder without providing a precise age of onset, duration, or number of symptoms necessary for the diagnosis.

ETIOLOGY

ADHD is a complex disorder of multifactorial etiology, thought to have multiple genetic and environmental variables of small effect acting together to produce vulnerability to ADHD. Although the etiology of ADHD yet has to be determined, there is a growing consensus that the condition involves functional and anatomical dysfunction in the brain's cortico-basal ganglia-thalamo-cortical circuitry. These areas support the regulation of attentional resources, the programming of complex motor behaviors, and the learning of responses to reinforcement. Theories involving these areas have been examined in studies involving neurobiologic studies of healthy humans, humans with ADHD, and animal models.

Genetics

Twin, sibling, adoption, and family studies, all suggest a strong genetic component in the development of hyperactivity, inattention, and impulsivity, with the heritability index estimated to range from 0.6 to 0.98. Heritability estimates include both genetic influences, as well as the effects of gene–environment interaction. ADHD-like behaviors have also been reported in rare genetic disorders such as fragile X syndrome, neurofibromatosis type 1, DiGeorge syndrome, tuberous sclerosis, Turner syndrome, Williams syndrome, Angelman syndrome, and Prader–Willi syndrome.

Twin Studies. Twin studies suggest that 76% of the variance in the transmission of ADHD is attributable to genetics, making ADHD one of the most heritable psychiatric disorders. Monozygotic twins have been found to be more concordant for ADHD symptoms of hyperactivity, inattention, and impulsivity than are same-sex dizygotic twins. The concordance rate among monozygotic twins ranges from 59% to 92%, whereas the concordance rate in dizygotic twins ranges from 29% to 42%.

Sibling and Half-Sibling Studies. Pettersson and colleagues reported on family data of 4,408,646 full and maternal half-siblings and genetic data of 125,533 cases and 208,215 controls in the Swedish sibling study. The strength of this study was the large sample size, and that psychiatric status was based on clinical diagnosis using standard diagnostic criteria. There was, however, the assumption that siblings and maternal half-siblings lived in the same household and thus shared the same environment. ADHD was one of the eight psychiatric disorders studied and was found to have a heritability estimate of 80%. An earlier sibling study found that 10 of 19 full-sibling pairs were concordant for ADHD, versus 2 of 22 half-sibling pairs; a significant difference supporting the genetic transmission of ADHD in families.

Adoption Studies. Studies have found that the biologic relatives of children with ADHD are more likely to have ADHD themselves. This relationship is not seen in adoptive relatives of children with ADHD. Adoptive relatives of children with ADHD have a risk of ADHD similar to relatives of control children, while the biologic relatives of children with ADHD are more likely to have ADHD or associated disorders. In a study of international adoptees aged 10 to 15 years, Van den Oord and colleagues estimated that genes accounted for 47% of the variance of inattention scores on the Child Behavior Checklist. Eighteen percent of biologic relatives of adopted-away children with ADHD were found to be at risk for the disorder versus 6% of the child’s adopted relatives.

Family Studies. Family studies have shown that first-degree relatives of children with ADHD are two to eight times more likely to have the disorder. First-degree relatives of children with ADHD have a 20% to 25% risk for ADHD, compared with 4% to 5% for relatives of controls. If a parent has ADHD, 50% of their offspring are likely to have that condition. The increased incidence of ADHD is also extended to second-degree relatives.

Mode of Inheritance. Several hypotheses of ADHD's mode of transmission have been advanced. Omenn explored the possibility of sex-linked transmission because of the male preponderance in the condition. Other theories include a polygenetic inheritance model, including those of Morrison and Stewart, but limitations in sample size made this hard to prove. Rhee and colleagues postulated a polygenetic multiple threshold model after analyzing sex differences in an Australian twin and sibling-pair study. However, differences in fit between genetic models were modest, as is true for comparisons of multifactorial and single-gene inheritance. This led to the suggestion that symptoms of ADHD may be caused by several interacting genes of modest effect. This hypothesis is consistent with ADHD's high population prevalence and high concordance in monozygotic twins, but modest recurrence risk for first-degree relatives. However, diagnostic uncertainty impedes progress in developing genetic models of inheritance.

Molecular Genetic Studies. There has been increasing interest in attempting to identify the specific “risk” genes that may be implicated in the development of ADHD. This search has yielded generally inconsistent results due to complexity of the trait and heterogeneity across studies. Candidate genes have been chosen for study based on theoretical considerations related to their function, and the majority of studies have concentrated on dopamine-related genes given the implication of dopamine dysfunction in ADHD. Candidate genes discovered so far are responsible for less than 5% of genetic variation in ADHD.

The dopamine transporter gene (DAT1) codes for a sodium-dependent dopamine transporter protein that regulates reuptake of dopamine from the synaptic cleft. Cook et al. were the first to show an association between ADHD and DAT1 with 480–base-pair allele. Several studies have indicated that DAT1 might be more strongly associated with the hyperactive/impulsive or combined subtypes of ADHD. Another early gene studied was the dopamine type D2 receptor gene (DRD2). The gene is not specific to ADHD and is not believed to be a primary cause of the disorder. The dopamine-4 receptor seven-repeat allele gene (DRD4) has been linked to novelty seeking and connected to the impulsivity seen in ADHD. DRD4 has also been related to the inattentive subtype of ADHD. The synaptosomal-associated protein 25 gene (SNAP25) codes for a protein involved in neurotransmitter regulation and axonal growth. Researchers have found an association between SNAP25 and sensitivity to methylphenidate (MPH) adverse events in preschoolers with ADHD. Genes that code for dopamine β-hydroxylase (DBH), catechol-O-methyltransferase (COMT), dopamine 5 receptor gene, norepinephrine transporter gene (SLC6A2), serotonin transporter gene (SLC6A4), α-2A adrenergic receptor (ADRA2A), thyroid receptor B gene, androgen receptors, and factors in immune function and regulation have also been reported to correlate with ADHD symptoms. Subsequent studies and meta-analyses have not sufficiently supported the association between ADHD and these genes.

GENOME-WIDE SIGNIFICANT COMMON VARIANTS. Genome-wide association studies (GWAS) scan the entire genome to detect common DNA variants having small etiologic effects. GWAS assay hundreds of thousands or even millions of single nucleotide polymorphisms (SNP). Thus, genome-wide statistical significance must have extremely high “p” values and very large samples. Such a large sample was formed by the consortium of ADHD researchers who carried out a GWAS meta-analysis of 12 studies involving 20,183 people with ADHD and 35,191 children and found 12 independent loci (Demontis et al., 2019) which reached significance. Loci identified include loci pertaining to dopamine regulation, educational attainment, and speech and language disorders. The GWAS analyses also showed that much of ADHD’s heredity is due to the polygenic effects of many common variants each having a very small effect.

GENOME-WIDE ASSOCIATION STUDIES AND COPY NUMBER VARIANTS. Copy number variants (CNV) of the genome often delete or duplicate a large genome segment spanning part or even the entire gene. They clearly affect gene functioning. Such deletions and duplications are overrepresented in ADHD samples and significant effects have been found for small CNV. It has been shown that CNV affect the metabolism of glutamate receptor genes which may be associated with cognitive and clinical impairments of ADHD. Magnetic resonance spectroscopy showed dysregulation of glutamate–glutamine concentration in ADHD patients. The area of genetics in ADHD is evolving and the picture is far from complete. However, the strong heritability of this condition is clear.

Neuroanatomical Aspects

Investigators have postulated the presence of an anterior attention network for promoting multiple aspects of attention, including focusing, executing, sustaining, and shifting functions. The attention networks are widespread through the cortex of the brain and are closely related to the default mode network described by Castellanos and others. MRI studies suggest that there are structural abnormalities in the brains of children with ADHD, including differences in the frontostriatal areas, temporoparietal lobes, cerebellum, basal ganglia, corpus callosum, amygdala, hippocampus, and thalamus. A meta-analysis of voxel-based morphometric (VBM) studies by Nakao and colleagues of children and adults with ADHD found global reductions in gray matter volumes, especially the right lentiform nucleus extending to the caudate nucleus, with slightly larger gray matter volumes in the left posterior cingulate cortex. These areas were demonstrated to normalize with age as well as with stimulant medication treatment. Developmentally, children with ADHD have a 3- to 5-year delay in cortical thickness maturation, particularly in the frontal and temporal brain regions. Diffusion tensor imaging (DTI) studies show white matter tract abnormalities and dysfunctional connectivity in ADHD during rest and while engaging in cognitive tasks.

Recent studies with larger samples by Hoogman et al. found lower surface area values in the frontal, cingulate, and temporal regions, with the largest effect being in total surface area and decreased cortical thickness in the fusiform gyrus and temporal lobes in children with ADHD. Neither surface area nor thickness differences were found in groups of adolescents and adults with ADHD, again suggesting some normalization occurring with age.

In a study using machine learning analysis of MRI data, significant differences were found in children and adults with ADHD compared to controls. However, prediction performance and effect sizes were better for the child versus the adult samples, suggesting greater differences in children versus adults with ADHD.

Deficits in neural activity within frontostriatal and frontoparietal circuits are consistent findings in ADHD. Functional MRI studies using cognitive tasks to target the neural circuits show significantly reduced frontal lobe activity in subjects with ADHD compared to controls. Meta-analysis shows hypoactivation in the frontoparietal and ventral attention networks in children with ADHD, areas that control goal-directed executive processes and decision making. Hart and colleagues found that right dorsolateral prefrontal cortex activation was reduced in medication-naive patients, but normal in long-term stimulant-medicated patients. Electroencephalography (EEG) studies have reported an elevated theta/beta ratio (TBR) in subjects with ADHD versus controls; however, increased theta band activity has also been associated with substance use and bipolar disorder. Recent meta-analysis determined that excess theta and elevated TBR was not a reliable diagnostic measure for ADHD but could have prognostic value for treatment outcome.

Neurotransmitters in ADHD

The nonspecific catecholamine hypothesis of ADHD posits that the psychophysiology of ADHD symptomatology emerges from an imbalance among various neurotransmitters, including norepinephrine, epinephrine, and dopamine. Human studies investigating the catecholamine hypothesis of ADHD produced conflicting results, possibly due to the difficulty of measuring brain levels of these neurotransmitters in peripherally obtained samples.

Dysfunction in brain norepinephrine systems, particularly a lack of inhibition of locus coeruleus neurons, could explain the inattention, cognitive deficits, and higher levels of gross motor activity seen in children with ADHD. Randomized clinical trials (RCTs) have suggested that the norepinephrine reuptake inhibitors (NRIs), tricyclic antidepressants (TCAs), and atomoxetine (ATX) are effective in reducing ADHD symptoms by restoring a more normal ratio of epinephrine and norepinephrine. Molecular genetic studies have targeted genes that code for dopamine receptors, and positron emission tomography (PET) scans in adults with ADHD have provided imaging data that correlate stimulant-driven increases in dopamine concentration. There is weak evidence for serotonin’s role in ADHD. Medications that exclusively affect serotonin function, such as the selective serotonin reuptake inhibitors (SSRIs), have no efficacy in the treatment of children with ADHD. For that reason, serotonin is thought to play a secondary role in the pathology of ADHD.

Recently, the glutamate neurotransmitter system has also been implicated in ADHD. It has been reported that CNV in glutamatergic genes were associated with cognitive and clinical impairments of ADHD. In addition, a pharmacogenomic GWAS study found SNP in glutamate receptor genes (e.g., GRM7) was one of the most significant findings. Thus, the glutamate neurotransmitter system may also have a role in ADHD, and this impact needs to be elucidated.

Environmental Factors

ADHD has been associated with numerous environmental factors, although it is difficult to conclusively determine causality. In utero risk factors that have been associated include prenatal exposure to alcohol, nicotine, recreational drugs, glucocorticoids, poor maternal diet, and maternal stress during pregnancy. Perinatal ADHD risks include prematurity, low birth weight, and obstetric complications. ADHD has been associated with childhood illnesses such as meningitis, encephalitis, cardiac disease, thyroid disease, epilepsy, head trauma and concussions, autoimmune and metabolic disorders, and in survivors of pediatric brain tumors. Psychosocial adversity including poverty, childhood maltreatment, negative parenting, family discord, and bullying have all been associated with ADHD. It has been difficult for investigators working with children affected by adversity to determine whether their ADHD symptoms reflect a response to negative parenting, a harsh environment, a genetically influenced biologic problem, or some interaction among these factors.

Dietary Influences. Dietary factors as a cause and possible treatment of ADHD are controversial. There have been reports of a connection between ADHD and nutritional deficiencies of essential fatty acids (omega-3 and omega-6) as well as mineral deficiencies (e.g., zinc, copper, and iron), but the beneficial effects of supplementation are inconclusive. Some studies have indicated that a small subset of children may have sensitivity to certain food additives, excess refined sugar intake, or nutritional deficiencies (such as fatty acids or iron), causing behavioral effects. There has been some previous support for elimination diets to treat ADHD, such as the Feingold (additive and salicylate-free) diet, oligoantigenic (hypoallergenic) diet, and ketogenic diet. However, in 2011, the U.S. Food and Drug Administration (FDA) determined that existing data do not support a causal link between additives and behavioral disturbance in children.

DIAGNOSIS AND CLINICAL FEATURES

The diagnosis of ADHD requires the integration of data from multiple sources, including clinical interview, direct observation, behavioral rating scales, and the subjective observations of parents and teachers. Most children with ADHD are referred for care because of impairment in academic, family, and/or peer relationship functioning. Symptoms of impulsivity, overactivity, and inattention drive this impairment across the lifespan.

Clinical Features

DSM-5 Diagnostic Criteria. DSM-5 criteria were outlined previously in this chapter. We will here elaborate on key symptoms and their presentation.

Hyperactivity. Hyperactivity describes the inappropriate excessive motor activity observed in ADHD, including restlessness, fidgeting, and appearing to be “driven by a motor.” These behaviors are often first noticed when the child is a toddler but can be normative before age 4. The level of gross motor activity usually decreases with age, however, the inner sense of restlessness may continue into young adult life. Children with ADHD display sustained levels of high activity in different settings such as home, school, and playgroup. Activity increases occur during sleep as well. Hyperactivity may be absent in novel situations or when the child is engaged in an activity they find particularly engrossing, such as playing video games.

Impulsivity. Impulsive actions occur without forethought about consequences and may be associated with a desire for immediate rewards or to avoid delay. Impulsivity might be seen when the child engages in dangerous activities, yells out in class, or interrupts or intrudes on others during games or conversations. Impulsive behavior might also result in trouble with parents, teachers, or other children, including verbal or physical fights. Children with ADHD can demonstrate overly quick and error-prone performance on standardized tasks. Symptoms of impulsivity can persist into adulthood, even after hyperactive symptoms have diminished.

Attentional Difficulties. Inattention includes difficulty sustaining focus, trouble maintaining organization, and being easily distracted by extraneous stimuli. The symptoms of inattention often become prominent in elementary school, when the child is approximately 8 to

9 years old, and symptoms can be lifelong. Attentional difficulties are most often seen in routine settings in which the youth with ADHD must sit and carry out tasks that involve repetition under conditions of low levels of reinforcement and external motivation. Sluggish cognitive tempo (SCT) is a type of attentional difficulty that has been closely associated with the inattentive subtype of ADHD. SCT describes a constellation of symptoms such as daydreaming, staring, tendency toward confusion, mental fogginess, sleepiness, apathy, and physical hypoactivity. There is growing evidence suggesting that SCT is a condition distinct from ADHD; however, SCT is currently not considered a separate disorder or diagnostic subtype.

Associated Factors. Children with ADHD might have areas of impairment that are not listed under the DSM-5 symptom criteria covered exactly by the 18 symptom exemplars of hyperactivity, impulsivity, or inattention. Children and adolescents with ADHD often show difficulty with time management and do not develop an internal sense of pace in planning tasks. This poor sense of time leads to problems in estimating the actual difficulty of waiting in line, planning how much time a task requires, or even knowing when to come home when out playing with other children. Children with ADHD often lack persistence and can become bored playing interactive games with peers. They also find it difficult to delay gratification.

Neuropsychological testing and normalized rating scales such as the Behavior Rating Inventory of Executive Function (BRIEF) have found that individuals with ADHD demonstrate impaired executive functioning, including abnormal response inhibition. On any given measure of executive function, less than half of children with ADHD have been found to be impaired. Executive functioning problems also occur in other psychiatric disorders of childhood, such as depression, and are not specific to ADHD. It has been shown that academic functioning is more strongly affected by an impulsive need to get through tests quickly, a deficit closely linked to poor behavioral inhibition, rather than poor executive functioning.

ADHD is often associated with problems in emotion regulation, resulting in temper outbursts, mood lability, and reactivity. Moods can change dramatically with no obvious connection with what is going on in the environment. Social skills are often significantly impaired in children with ADHD compared to controls. Individuals with ADHD may have problems accurately interpreting nonverbal social cues and have trouble cooperating with other children and following rules in games. Peers identify these individuals as intrusive, bossy, and insensitive to the needs of others. Children with ADHD often overreact and respond to frustration in social situations which can lead to verbal or physical aggression. This is a strong stimulus for peer rejection, which has been shown to be a reliable long-term negative predictor of development, particularly in adolescence.

MEDICAL AND LABORATORY EXAMINATION

Children with ADHD should be given a medical evaluation, including a thorough history and physical examination, to rule out physical disorders that can mimic the symptoms of ADHD. The medical history should cover the prenatal, perinatal, toddler, and preschool phases of development. Inquiries should be made about pregnancy complications, such as maternal illness (eclampsia, diabetes), maternal smoking, alcohol, or illicit drug use. The perinatal period should identify the presence of labor problems, delivery complications, prematurity, jaundice, and low birth weight. Feeding and sleeping routines during the perinatal and postnatal periods need to be described. Developmental milestones, illnesses, injuries, and symptomatology need to be reviewed. A detailed family, medical, and psychiatric history should be completed.

The medical history should include specific questioning about any past heart disease, syncope, dizziness, chest pain, shortness of breath, or palpitations at rest or exertion in the child. In addition, it is recommended to screen for family cardiovascular history including sudden death, early-onset cardiac disease, arrhythmias, or prolonged QT syndrome. It is also helpful to document a premedication baseline for common side effects such as stomachache, headache, insomnia, or rapid heartbeat.

The medical examination can identify physical problems that can increase the risk of serious adverse events should a stimulant medication treatment be initiated, including tachycardia, arrhythmia, or abnormalities in blood pressure. Height and weight at baseline should be recorded. Physical examination of a child with ADHD may reveal nonlocalizing neurologic signs, such as poor coordination, dysrhythmias, and overflow movements. Medical problems that might aid in the differential diagnosis, such as petit mal epilepsy, hearing and vision difficulty, thyroid dysfunction, and hypoglycemia, all need to be ruled out. No blood, genetic, or neuroimaging tests are available to diagnose ADHD.

RATING SCALES

Many standardized behavioral rating scales for ADHD have been developed. Rating scales should not be used independently to diagnose ADHD but can be particularly helpful in establishing the presence of core ADHD symptoms in more than one setting. ADHD-specific scales have been found to have greater than 90% sensitivity and specificity if used in the correct population. Clinical advantages of rating scales include the ability to obtain collateral information from different settings and to identify children who require further treatment because, although they are improved, they are not in remission. There are several commonly used rating scales in the public domain to assess for ADHD symptoms. The National Institute of Children's Health Quality (NICHQ) Vanderbilt Assessment scales for parents and teachers are now in their third edition and widely available on the internet (nichq.org), including a scale specific to the preschool population. The Vanderbilt scales also screen for oppositional disorder and conduct disorder (CD), depression, anxiety, school performance, learning, family, and peer relationships, and participation in activities. The Swanson, Nolan, and Pelham (SNAP) scale includes a 26-item form to look at ADHD and oppositional disorder and a 90-item version that includes a broad screen for other disorders and problems and which has been subject to psychometric validation. More recently, The Strengths and Weakness of ADHD-

Symptoms and Normal Behavior (SWAN) scale was developed to evaluate attention and behavior dimensionally, from below average to far above average. Proprietary scales such as the ADHD RS and Conners must be purchased but have the advantage of being consistent with DSM-5 and including psychometrically validated age and gender scoring paradigms. The Conners scales, now in their fourth edition, include forms specific to preschool, adolescent self-report, and comprehensive screening for a wide range of problems. The Weiss Functional Impairment Rating Scale has a parent and self-report version that evaluates the specific symptom-driven impairments a patient may experience. This is helpful in identifying those areas of functional impairment that may not respond to medication treatment alone and require other targeted interventions.

DIFFERENTIAL DIAGNOSIS

A clinician must make a differential diagnosis when judging whether a patient meets criteria for ADHD. The distinction between differential diagnosis and comorbidity is not well defined in the DSM-5. If the impairing condition is best accounted for by another disorder, that other disorder rules out ADHD. Conditions whose symptoms mimic those of ADHD should be considered and ruled out when possible, with attention to the possibility of either medical or psychiatric comorbid diagnoses.

For example, symptoms of inattention may be present in medical conditions such as hypoglycemia, petit mal epilepsy, substance abuse, sleep apnea, and hypothyroid or hyperthyroid conditions. Disrupted or poor sleep quality may present with tiredness, as well as difficulties with inattention, learning, and impulse regulation. Children undergoing ADHD evaluation should also be screened for clinical signs and symptoms suggestive of sleep apnea mimicking ADHD.

Symptoms of ADHD may also be present in psychiatric illnesses such as anxiety, depression, oppositional defiant disorder (ODD), and early psychosis in adolescents. The symptoms of children with oppositional defiant behavior must be distinguished from the reluctance and resistance that children with ADHD exhibit when they avoid effortful mental tasks. For children with intellectual disability, inattention may be a natural accompaniment of the intellectual limitations, so the diagnosis of ADHD should only be applied if that symptom is excessive for the mental age of the child and the impact of intellectual challenges has been ruled out.

Furthermore, children who experience traumatic events (e.g., serious injury, physical/sexual abuse, neglect, violent acts, unexpected death of a loved one, life-threatening disasters, or parental divorce) can develop a stress response that mimics ADHD symptoms, including disorganization, difficulty concentrating, distractibility, and difficulty sleeping.

Symptoms of impulsivity, gross motor hyperactivity, or inattention may also appear during the use of medications, such as the akathisia that accompanies second-generation antipsychotics or from bronchodilators. Such children should not receive a diagnosis of ADHD, but one of other substance-related disorder.

COURSE AND PROGNOSIS

Parents often notice very high levels of gross motor activity when the child with ADHD is a toddler, just when the child has learned to walk independent of the parents' help. However, the energy, oppositionality, and curiosity of toddlers can be confused with the excessive, almost random motion of older children with ADHD, so that one must be cautious when applying the ADHD diagnosis to a preschooler. Usually, the ADHD diagnosis is first applied in primary school, during grades one to six, when adjustment to the sedentary learning style is compromised. School-aged children with ADHD may experience academic failure and peer rejection. Behavior problems at school and at home are not infrequent, with ODD being a common comorbidity.

During childhood, 65% to 75% of children with ADHD have one or more comorbid conditions, which in addition to ODD, can include conduct disorder (CD), anxiety and mood disorders, tics or Tourette syndrome, learning disorders, and autistic spectrum disorders. Many of these comorbidities continue into adolescence.

In adolescents, the external overactivity lessens but the internal restlessness does not. The adolescent with ADHD continues to have academic and social problems which erode self-esteem and at times promote the gravitation to a negative peer group. Thus, adolescents with ADHD who are untreated, compared to peers with no mental disorder, have a threefold increase in substance use and abuse, more frequent delinquent acts and trouble with the law, and increased rates of car accidents when they begin to drive. Adolescents with ADHD also start sexual intercourse earlier, with a higher number of sexual partners and greater frequency of unprotected sex and corresponding higher rates of sexually transmitted infections and unwanted pregnancy. The rate of depression and anxiety also increases in adolescence because of the many functional problems due to increasing demands and decreasing supports.

Approximately 60% of children with ADHD continue to be impaired well into adult life, with prevalence estimates suggesting that 4% of adults may suffer from ADHD. In adulthood, hyperactivity may be expressed by inner feelings of restlessness, inability to relax and excess talkativeness, fidgeting, and inability to remain seated for long periods of time, for example, in meetings, symposia, or in church. Impulsivity may be expressed in impatience, interrupting others, acting without thinking, and ending or starting ventures on impulse, for example, jobs, relationships, and at times sensation-seeking behavior. Inattention is seen in distractibility, disorganization, chronic lateness, or missing appointments. The distraction and disorganization often result in careless mistakes and forgetfulness. Procrastination is another important problem. Even when things are started, they are often not finished and so deadlines are thus missed.

The inattention and distractibility also make it appear that the adult with ADHD is not interested and not listening to what is being said. This can strain relationships at home and at work. Some adults with ADHD also suffer from labile moods and “short fuses.” These characteristics, coupled with the problems described above, result in more frequent losses of jobs and relationships. The consequences of this are lower job status, lower income, and a sense of underachievement or failure on the part of the adult with ADHD. The inattentiveness and impulsivity also increase the likelihood of driving accidents, and infractions or other types of accidents are also more prevalent. In general, adults with ADHD may have an unhealthy lifestyle, with smoking, alcohol and drug abuse, risky sexual behavior, and chronic sleep problems frequently seen.

In adulthood, comorbidity is the rule, with more than 75% of adults with ADHD having at least one comorbid condition. Anxiety, depression, substance use, and some personality disorders are the most common comorbidities. Gambling and other addictions are also frequently seen.

Prospective follow-up has found that age is associated with a decline in symptoms, but less often with full functional remission. Hyperactive and impulsive symptoms tend to decline at a higher rate than inattention symptoms. It should be noted that even though the male/female ratio in children with ADHD can be 3 to 4:1, in adulthood it is 1:1. It is possible that many girls who tend to be less disruptive are not identified or referred in childhood but have significant impairment as adults and seek help then.

TREATMENT OF ADHD

The efficacy of both medication and psychosocial interventions for the treatment of ADHD has been well established. Treatment with ADHD medications reduces educational underachievement, accidental injury, bone fractures, traumatic brain injury, substance use, cigarette smoking, sexually transmitted infections, teenage pregnancy, depression, suicide, and criminal activity. Nonmedication treatments for ADHD are less effective than medication treatments for ADHD symptoms but are frequently useful to help problems that remain after medication has been optimized. The preferences of families and availability of services largely guide the choice of treatment.

Stimulants

Stimulants are considered first-line agents to treat ADHD symptoms based on decades of proven safety and efficacy. Therapeutic effects of stimulants include a decrease in hyperactivity, impulsivity, and inattention, as well as improvement in associated behaviors. The two groups of stimulant medication that have received U.S. FDA approval for the treatment of youth with ADHD are MPH and amphetamines (AMP). There are approximately 30 stimulants on the market in the United States, many in both generic and branded versions (see Table 44.1–1 for brands, dose, onset, and duration of action). Stimulants are classified with drugs that can be abused or cause dependency under the category of schedule II–controlled substances by the U.S. Drug Enforcement Agency (DEA) (Table 44.1–1). In choosing between preparations, clinicians need to consider the needed duration of action, flexibility of dosing, time of day of impairment, time of day when maximum drug is released, whether the child can swallow pills, availability, and cost. Whenever possible, single-day dosing is helpful in improving compliance and coverage. Adherence and persistence with stimulant treatment have been problematic and can be managed by addressing efficacy and side effects and frequent follow-up. In addition, direct instruction, use of a pill box, and parental supervision, monitoring, or use of digital interventions can help patients to remember to take their medication.

Table 44.1–1.
Methylphenidate Formulations

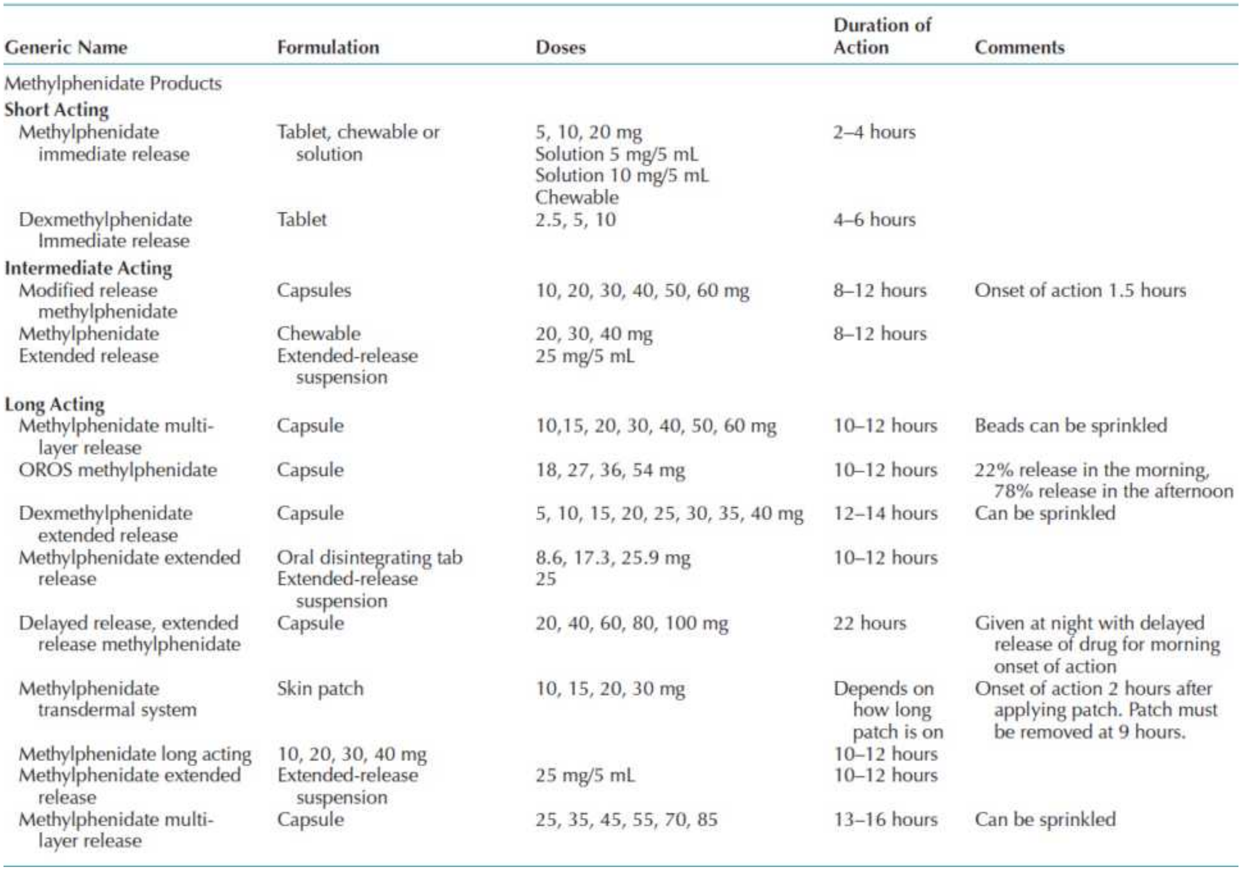

Generic Name	Formulation	Doses	Duration of Action	Comments
Methylphenidate Products				
Short Acting				
Methylphenidate immediate release	Tablet, chewable or solution	5, 10, 20 mg Solution 5 mg/5 mL Solution 10 mg/5 mL Chewable	2–4 hours	
Dexmethylphenidate Immediate release	Tablet	2.5, 5, 10	4–6 hours	
Intermediate Acting				
Modified release methylphenidate	Capsules	10, 20, 30, 40, 50, 60 mg	8–12 hours	Onset of action 1.5 hours
Methylphenidate Extended release	Chewable Extended-release suspension	20, 30, 40 mg 25 mg/5 mL	8–12 hours	
Long Acting				
Methylphenidate multi-layer release	Capsule	10,15, 20, 30, 40, 50, 60 mg	10–12 hours	Beads can be sprinkled
OROS methylphenidate	Capsule	18, 27, 36, 54 mg	10–12 hours	22% release in the morning, 78% release in the afternoon
Dexmethylphenidate extended release	Capsule	5, 10, 15, 20, 25, 30, 35, 40 mg	12–14 hours	Can be sprinkled
Methylphenidate extended release	Oral disintegrating tab Extended-release suspension	8.6, 17.3, 25.9 mg 25	10–12 hours	
Delayed release, extended release methylphenidate	Capsule	20, 40, 60, 80, 100 mg	22 hours	Given at night with delayed release of drug for morning onset of action
Methylphenidate transdermal system	Skin patch	10, 15, 20, 30 mg	Depends on how long patch is on	Onset of action 2 hours after applying patch. Patch must be removed at 9 hours.
Methylphenidate long acting	10, 20, 30, 40 mg		10–12 hours	
Methylphenidate extended release	Extended-release suspension	25 mg/5 mL	10–12 hours	
Methylphenidate multi-layer release	Capsule	25, 35, 45, 55, 70, 85	13–16 hours	Can be sprinkled

Pharmacology and Pharmacokinetics. Psychostimulants increase central nervous system activity in brain regions and increase blood pressure and pulse in the periphery. Although the exact mechanism of action is not completely understood, stimulants increase the concentration of dopamine and norepinephrine in the synaptic cleft. MPH appears to act primarily as a dopamine–norepinephrine reuptake inhibitor, while AMPs block the reuptake of both norepinephrine and dopamine as well as facilitate dopamine release. Some patients respond selectively to either MPH or AMP, with the majority of patients responding to both. AMP products are approximately two times more potent but may have lower tolerability in children. Dextro methylphenidate (d-MPH) is twice as potent as MPH, since it is the d isomer of the racemic mixture. Medication is introduced at a low dose and titrated up to optimize effect and minimize side effects. There are wide differences in interindividual dose–response, so each individual needs to be titrated to their own optimal dose, and dose–response needs to be re-evaluated over time. When switching within the same class of stimulants (MPH or AMPs), direct switching is possible. However, when switching from one stimulant class to the other, decreasing the first class before switching to the second is recommended.

Formulations

FORMULATIONS OF METHYLPHENIDATE. Table 44.1–1 summarizes the commonly used short-acting immediate-release preparations (3 to 6 hours), intermediate acting (8 to 10 hours) and long acting (12 to 16 hours) of MPH and AMP drugs, as well as their availability as a tablet, capsule, suspension, chewable, or patch. These options increase the clinician's ability to provide individualized treatment for patients who need a short onset of action, a particular duration of action, patients who cannot swallow pills, patients who need more medication at the end of the day, or patients who wish to take medication in the evening rather than in the morning. These preparations differ in the pharmacologic mechanism they use to extend drug release. These include the osmotic release oral system (OROS) osmotic pump, use of beads with multiple layers, and drugs that combine immediate- and delayed-release beads. There are also differences between various formulations in the amount of drug released at a particular time of day. Choice between formulations depends on the length of time that a child needs medication and what time of day they are most impaired. There are also differences in time to onset of drug action, which is important in timing the early-morning administration of the drug to assure it is active at the time treatment needs to be effective.

Table 44.1–2. Amphetamine Formulations

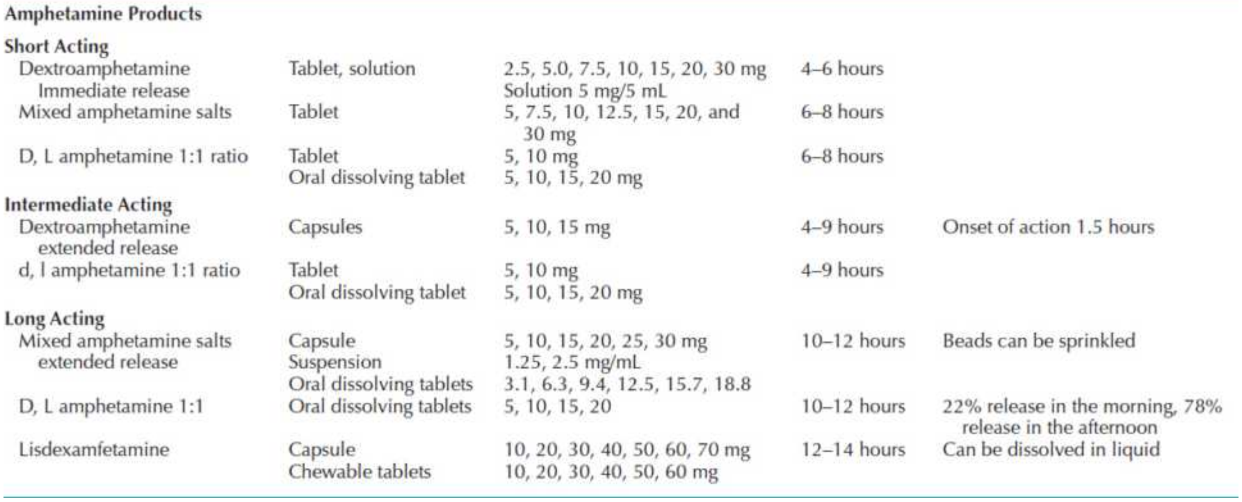

Amphetamine Products				
Short Acting				
Dextroamphetamine Immediate release	Tablet, solution	2.5, 5.0, 7.5, 10, 15, 20, 30 mg Solution 5 mg/5 mL	4–6 hours	
Mixed amphetamine salts	Tablet	5, 7.5, 10, 12.5, 15, 20, and 30 mg	6–8 hours	
D, L amphetamine 1:1 ratio	Tablet Oral dissolving tablet	5, 10 mg 5, 10, 15, 20 mg	6–8 hours	
Intermediate Acting				
Dextroamphetamine extended release	Capsules	5, 10, 15 mg	4–9 hours	Onset of action 1.5 hours
d, l amphetamine 1:1 ratio	Tablet Oral dissolving tablet	5, 10 mg 5, 10, 15, 20 mg	4–9 hours	
Long Acting				
Mixed amphetamine salts extended release	Capsule Suspension Oral dissolving tablets	5, 10, 15, 20, 25, 30 mg 1.25, 2.5 mg/mL 3.1, 6.3, 9.4, 12.5, 15.7, 18.8	10–12 hours	Beads can be sprinkled
D, L amphetamine 1:1	Oral dissolving tablets	5, 10, 15, 20	10–12 hours	22% release in the morning, 78% release in the afternoon
Lisdexamfetamine	Capsule Chewable tablets	10, 20, 30, 40, 50, 60, 70 mg 10, 20, 30, 40, 50, 60 mg	12–14 hours	Can be dissolved in liquid

FORMULATIONS OF AMPHETAMINE. Table 44.1–2 summarizes the commonly used AMP medication for ADHD. Although absorption of MPH products is generally not affected by foods, some long-duration preparations of dextro amphetamine (d-AMP) show decreased concentrations when taken with high-fat meals. Whereas MPH is metabolized in the gut wall and plasma by esterases, the metabolism of d-AMP and mixed amphetamine salts (MAS) takes place in the liver. Variants in the human CES-1 gene coding for the hydrolyzing enzyme have led to interindividual variations in MPH metabolism and plasma level.

AMP product package inserts approved by the FDA warn that urinary excretion of AMP metabolites can be increased by foods of acidic content (vitamin C and orange juice), thus lowering plasma concentration. Conversely, foods that shift urinary pH higher reduce excretion and thus increase the plasma concentration of d-AMP or MAS.

Efficacy. A systematic and rigorous meta-analysis of randomized, double-blind, controlled medication trials in 10,068 children and adolescents demonstrated large effect sizes for both AMPs and MPH on parent ratings and large effect sizes for MPH on teacher ratings. Intercontinental Medical Statistics (IMS) data (2015) indicates that the majority (52%) of children are on MPH products, whereas the majority (76%) of adults are on AMP products. The MTA study found that 77% responded to MPH, and an additional 10% of MPH nonresponders responded to d-AMP. Teacher rating scales showed significant reductions in ADHD symptoms for stimulants compared with placebo, with effect sizes ranging from 0.8 to 1.0 on behavioral measures and an effect size of 0.5 for cognitive measures for those children randomized to active drug. Placebo response rates in controlled studies have ranged from 17% to 40%, and for reasons that are unclear have increased over time. The Preschool ADHD Treatment Study (PATS) determined that preschool-aged children required lower optimal doses of immediate release (IR) MPH, experienced higher rates of adverse events, and had a diminished response as compared to school-aged children. Preschool children also had a slower clearance of the drug.

AMP and MPH both share the ability to reduce gross motor overactivity, increase sustained attention on tasks, and reduce impulsive responses on laboratory measures. Some individuals respond preferentially to MPH, whereas others respond preferentially to AMP. The majority of evidence indicates that MPH and AMP both have large effect sizes.

Cognitive and Behavioral Effects. Greater effects due to stimulants are found to occur in behavioral domains rather than in cognitive areas. Psychostimulant medications have been shown to reduce gross motor overactivity, out-of-seat behavior, calling-out in the classroom, disruptiveness, impulsive behavior, and noncompliance while improving self-perception, academic productivity, peer nominations of the child, and mother–child interaction. Stimulants have been shown to improve vigilance and reaction time, reduce variability, and enhance short-term memory and learning of verbal and nonverbal material in children with ADHD. The MTA Titration Trial showed a linear dose–response curve for behavioral measures between 0.5 and 1.0 mg/kg/day. Recent controlled trials revealed improvements in school-based productivity, grade point average, test scores, and reading attendance in children with ADHD consistently treated with stimulants.

Adverse Effects

SHORT-TERM ADVERSE EFFECTS. Common side effects of the stimulant medications include decreased appetite, weight loss, delayed onset of sleep, headaches, stomach aches, and small increases in blood pressure and pulse. Most of the short-term side effects can often be dealt with by changing the time or dose of medication, change in diet, and sleep hygiene with or without use of melatonin to improve sleep-onset latency. Headaches tend to abate with time and stomach aches are decreased if medication is administered on a full stomach. Stimulants can also be associated with psychiatric side effects such as an increase in dysphoria, anxiety, irritability, social withdrawal, being “too quiet,” and emotional dysregulation.

Parents may complain of a worsening of behavior when medication is wearing off, sometimes called rebound. Rebound presents the sudden return of more severe overactivity, impulsivity, and inattention alongside increased irritability and emotional reactivity when medication is wearing off than was present at baseline off medication. Attempts to demonstrate their presence in full-day, laboratory classrooms have been unsuccessful, so these effects may be generated as the child interacts with task demands in their natural environment. Use of stimulants with a longer duration of action or a small booster dose of short-acting stimulant medication before the onset of rebound may be helpful. If these measures are not effective, a small dose of a short-acting α-blocking agent shortly before the onset of rebound can be useful.

Infrequent side effects include tics or the unmasking of Tourette syndrome. If the tics are functionally impairing, treatment with an α-blocker may be helpful.

Rare side effects include paranoia or other psychotic symptoms, priapism, choreiform movements, and self-directed behavior, such as lip licking, nail biting, and rubbing of sores. These problems are first treated by discontinuing the medication until the adverse event disappears.

Two large recent studies in children and one in adults found that ADHD medications are not associated with an increased risk of severe cardiovascular events. General cardiac side effects of stimulant drugs are an increase in heart rate of approximately 1 to 2 beats/minute and an increase in systolic and diastolic blood pressure of 3 to 4 mm Hg. Stimulants have not been found to induce clinically significant changes in electrocardiographic parameters. The evidence suggests a potential for severe adverse cardiac events in some children with certain forms of congenital heart disease or arrhythmias. Routine electrocardiograms are not recommended prior to initiating stimulants in the absence of concerning physical examination findings or personal or family cardiac history.

Long-Term Adverse Events

EFFECTS ON GROWTH. The use of stimulants in children has been associated with growth suppression. Two large-scale NIMH-supported RCTs have shown that MPH can reduce the rate of height and weight gains in developing children with ADHD. The reductions in linear growth have been approximately 1 cm in height and 2.5 kg in weight during the first year of treatment with stimulants. Net growth loss in adulthood is greater with consistent than inconsistent use of medication, with a maximum loss of about 1.3 inches. Small, single-site studies are discordant concerning whether psychostimulants suppress human growth hormone (HGH) or affect prolactin levels. Clinicians need to regularly monitor the height and weight of children with ADHD treated with psychostimulant medication. Drug holidays during summer vacations increase the probability of growth rebound and thus may mitigate problems of growth deceleration when it occurs.

SUBSTANCE ABUSE. There is no evidence to suggest that the use of stimulants in young people increases the risk for later substance use disorders (SUDs). A meta-analytic study by Wilens and colleagues suggests that stimulant treatment in childhood has a protective effect, bringing the risk down to the level of the general population. Other studies have revealed that stimulant therapy neither increases nor decreases the risk for subsequent SUD in children and adolescents with ADHD. Clinicians should take into consideration if family members are currently abusing prescription medication before prescribing stimulants for ADHD.

Stimulant medication can be misused or diverted for the purpose of academic performance enhancement or euphoric effects, with immediate-release stimulants more often misused compared to extended-release formulations. A systematic review of 21 studies indicated the prevalence of psychostimulant misuse in the year prior to the study ranged from 5% to 9% in grade and high school–aged children and 5% to 35% in college-aged individuals. Close monitoring, pill counts, or switching to long-acting, less easily abused formulations should be tried and, if not successful, nonstimulant medication should be employed by the clinician if misuse or diversion is suspected.

Summary of Stimulant Medication. Continuous treatment of children with ADHD with psychostimulants is both effective and safe over the first 2 years. Long-acting preparations provide the convenience of once-daily dosing, better coverage, and reduced stigma associated with taking medication at school and can decrease the potential for abuse. Short-acting forms are useful for younger children who often require smaller doses than are available in extended-release preparations. Many adverse events can be handled by adjustment of dose and/or the time of administration of the medication. Long-term adverse events such as growth rate suppression concern parents but may be managed by instituting drug holidays to permit growth rebound.

Nonstimulant Medication in the Treatment of Children with ADHD

Whereas psychostimulants constitute the first-line treatment for children with ADHD, other medications need to be used for the treatment of patients who do not respond to or cannot tolerate stimulants. Some nonstimulant medications, such as Atomoxetine (ATX) and long-acting clonidine or guanfacine, have been approved by the FDA for the treatment of ADHD in children and adolescents. Advantages of nonstimulant medications include longer duration of action and the absence of psychostimulant effects such as behavioral rebound or delay of sleep onset. These drugs provide a moderate rather than a large effect size of stimulants but, nonetheless, provide clinical benefit.

Atomoxetine. ATX, a norepinephren reuptake inhibitor (NRI), is an approved alternative to stimulants. ATX may be considered in individuals where stimulants have failed or parents have refused stimulant treatment. In particular, for patients who are unable to tolerate stimulants because of worsening anxiety or tics, ATX is a reasonable alternative since it may ameliorate these comorbid conditions. ATX is well absorbed after oral administration and is minimally affected by food. At steady state, ATX provides 24 hours of coverage.

ATX is metabolized by the cytochrome P450 (CYP) 2D6 hepatic enzyme system. A fraction of the population are poor metabolizers of CYP 2D6–metabolized drugs. Such individuals may experience elevated plasma concentrations and extended plasma half-life; therefore, very slow initial titration may be required for all patients to prevent excessive dosing of these patients. Patients taking other medications metabolized by CYP 2D6, such as fluoxetine, paroxetine, and quinidine, might show signs of competitive inhibition, namely increased plasma levels of both ATX and the competing drug.

In order to mitigate side effects and to optimize the dose for individual patients with variable 2D6 activity, patients are usually started at 0.5 mg/kg and titrated up slowly with increases every 1 to 2 weeks up to 1.2 mg/kg. Patients who are slow metabolizers may require lower doses, just as patients who are ultrafast metabolizers may require higher doses. The time to therapeutic effect is in the range of 1 to 4 weeks, with full response at 6 to 12 weeks. Early signs of response in the first month are predictive of response at 3 months, which can be helpful in encouraging parents to persist with treatment in the absence of the immediate benefits seen with stimulants.

ATX has been shown to reduce ADHD behaviors in children, adolescents, and adults with a medium effect size of 0.5 to 0.7. Common adverse events associated with ATX involve abdominal discomfort, nausea, decrease in appetite, sedation, daytime sleepiness, headache, dizziness, vertigo, irritability, and mood swings. To minimize the risk of stomach aches, ATX is best taken on a full stomach. Minor increases in blood pressure and pulse have been reported and thus should be regularly monitored. Rare adverse events in youth include priapism, disturbances in sexual functioning, increased suicidality, and liver toxicity. ATX has a black box warning of possible suicidal ideation, although more recent studies have questioned if this is actually a concern. Some patients may also experience adverse psychiatric side effects such as new-onset anger or irritability.

α-Adrenergic Agonists. The α-adrenergic agonist agents were initially approved for the management of hypertension in adults. The long-acting α-adrenergic agents guanfacine ER and clonidine ER have been FDA approved as monotherapy and adjunctive treatment of ADHD in children 6 to 17 years old. The immediate-release clonidine and guanfacine have not been approved to treat ADHD, although they are used as second-line or adjunctive agents. Guanfacine has also been found to be effective as an adjunctive agent in improving response to stimulants, particularly in decreasing blowups. The supposed mechanism of action of the α-adrenergic agents involves stimulating α-receptors in the prefrontal cortex leading to improvement of inattention, impulsivity, and hyperactivity. The α-adrenergic agonists are clearly less efficacious than the psychostimulants, with several meta-analyses suggesting a moderate effect size. However, a small multisite, double-blind, RCT showed that children and adolescents comorbid for ADHD and Tourette syndrome benefited from clonidine alone or in combination with MPH, in both cases significantly better than placebo. Guanfacine has proven efficacy in reducing oppositional symptoms in children with comorbid oppositional behaviors and ADHD.

Common side effects of the α-adrenergic agents are sedation, fatigue, somnolence, headache, hypotension, syncope, and bradycardia. Clonidine is very sedating and has been used to treat sleep difficulties of children with ADHD, as well as to offset the delay in sleep onset caused by stimulant medications. Guanfacine is a more selective α-adrenergic agonist and is slightly less sedating than clonidine. The α-adrenergic agonists need to be discontinued very slowly to prevent rebound adrenergic overdrive, with hypertension, agitation, fever, headache, tachycardia, chest pain, sleep disturbance, nausea, and vomiting. Families need to be warned of the potential dangers of sudden discontinuation of this medication, even for holidays or weekends with another parent.

Bupropion. Bupropion is an antidepressant with dopaminergic and noradrenergic properties that has shown modest efficacy in ADHD but is less effective than stimulants or ATX. It is available in extended-release preparations that permit once-daily dosing. Bupropion may be considered as an option for patients presenting with significant depression. Starting dose for young adolescents is 75 mg twice a day to a maximum of 200 to 300 mg/day. The doses can be increased once weekly until improvement is noted. Bupropion usually takes 4 to 6 weeks at a therapeutic dose to show effect. Side effects of bupropion include fatigue, dry mouth, insomnia, headaches, nausea, vomiting, constipation, tremor, and skin rash. Seizures will reliably occur if the adolescent’s dose exceeds 400 mg/day. This medication is FDA approved for the treatment of depression and smoking cessation, but not ADHD, in individuals older than 18 years of age. Bupropion is considered a second-line, off-label treatment for ADHD.

Other Alternative Medications. Anticonvulsants, including carbamazepine, lithium, and SSRIs have not been shown to be effective for the treatment of ADHD. Second-generation antipsychotic medications such as risperidone have been shown to be effective for the treatment of aggression in children with disruptive behavior disorders, especially when combined with parent training and MPH.

Summary of Medication Treatments

Most children with ADHD respond to stimulant medication with good effect on all the core symptoms, making stimulant treatment the first-line medication treatment. If stimulants are ineffective or poorly tolerated, nonstimulant alternatives such as ATX can be used. Monotherapy with α-adrenergic agents is a third choice because they are less effective. Patients with a suboptimal response to stimulants or inability to tolerate an optimally effective dose may benefit from combination treatment with a stimulant and nonstimulant. Clonidine and guanfacine are sedating and have significant cardiovascular side effects. For all these drugs, careful monitoring and slow, cautious discontinuation are essential.

Treatment of ADHD Comorbidities

Tic Disorder. Stimulants alone or in combination with clonidine, guanfacine, or ATX appear to reduce ADHD symptoms in individuals with both ADHD and Tourette syndrome or tic disorders. Although stimulants have not been shown to worsen tics in most patients with tic disorders, they may exacerbate tics in some individual cases. In these instances, decreasing the dose of stimulants, adding α-agonists, or switching to ATX may be alternative strategies. Habit reversal therapy (HRT) and cognitive behavioral intervention for tics are evidence-based therapies that have been shown to be effective for tics.

Seizure Disorder. Children with seizure disorders may also suffer from ADHD. Although there are concerns about psychostimulants, particularly MPH, lowering the seizure threshold, experiencing a first seizure when starting a psychostimulant is extremely rare. Studies of EEG epileptiform activity, seizure rates, or interactions between antiepileptic drugs (AEDs) and psychostimulants have not shown any additional seizure risk. On the contrary, studies have shown that children with ADHD and a seizure disorder that is adequately controlled on antiepileptic medications show reduced ADHD symptoms and no change in seizure frequency.

Aggression, Oppositional Defiant Disorder, and Conduct Disorder. Individuals with ADHD and comorbid aggression, ODD, and/or CD require comprehensive intervention which includes pharmacotherapy and psychosocial treatments. Stimulants have been shown in randomized, controlled trials to improve ADHD, aggression, ODD, and CD symptoms. α-Blockers (guanfacine and clonidine) alone or as adjuncts to stimulants have also been shown to decrease ADHD, aggression, and oppositional symptoms. Stimulants also reduced negative social interactions and covert antisocial behavior (stealing and vandalism but not cheating). However, one cannot solely rely on medication to treat children with ADHD and these comorbid conditions. A comprehensive treatment approach which may include parent training, social skills training, and anger management for the child, as well as an appropriate academic program and support, may also be needed. Adjunctive treatment with α-blockers (guanfacine XR or clonidine) can be helpful in managing extreme explosive outbursts. If these are not effective, replacing these drugs with risperidone or aripiprazole can be useful, although this benefit has to be weighed against the risk of the metabolic effects of the latter drugs. Careful metabolic monitoring is essential. Treating aggression, ODD, and CD early and effectively is critical, as long-term consequences of these disorders include later drug and alcohol abuse and antisocial behavior with involvement of the criminal justice system.

Anxiety Disorder. It had been suggested that children with ADHD and comorbid anxiety may be less responsive to and experience more side effects with stimulant medication. The large multimodal study of ADHD (MTA) did not find this to be the case. However, children with ADHD and anxiety disorder seemed to benefit significantly from psychosocial treatment as well as medication. Thus, for children with ADHD and comorbid anxiety, slow and careful titration of stimulant medication for the ADHD symptoms is recommended. Evidence-based psychosocial treatments such as cognitive behavioral therapy (CBT), parent training, and social skills training combined with an appropriate academic program and academic support is suggested. However, should all these measures prove inadequate, SSRIs (fluoxetine or sertraline) to address the anxiety can be added.

Mood Disorder. In community samples, major depressive disorder (MDD) in youth with ADHD is more than five times higher than in youth without ADHD. Depressive disorders in youth with ADHD usually occur several years after the onset of ADHD and may be due to the cumulative effects of ADHD-related impairments and negative environmental circumstances they need to cope with. Youth with depression and ADHD have greater levels of psychosocial impairment than those with either condition alone. For individuals with both disorders, depressions start earlier, last longer, are more likely to recur, and have higher rates of suicides and hospitalizations than those who have depression alone.

A combination of pharmacotherapy and psychosocial interventions is recommended for patients who have both ADHD and major depression. In general, as per the Texas Children’s Medication Algorithm Project (CMAP), it is suggested that the clinician begin with the most effective medication for the most impairing condition for the patient. Only after the full effect of this medication is determined for both conditions, should a second medication be added to deal with residual symptoms of the second condition. Thus, in some cases of the combination of ADHD and mood disorders, both stimulants and SSRIs may need to be used.

Psychosocial interventions both for the depression, for example, CBT, and the ADHD, for example, organizational, study skills remediation, and appropriate academic and/or vocational programs, may all be needed.

There has been controversy regarding the comorbidity of ADHD and bipolar disorder. Emotional dysregulation and irritability can be an important source of impairment in children with ADHD, and these children may require additional and specialized intervention. DSM-5 introduced the diagnosis of disruptive mood dysregulation disorder (DMDD) to assist clinicians in attempting to characterize children who have chronic and frequent severe temper outbursts, anger, and irritability in multiple settings during latency or adolescence. Classical symptoms of bipolar disorder, such as severe elation of mood, grandiosity, racing thoughts, and hypersexuality, are used to confer the bipolar diagnosis. In these circumstances, the bipolar disorder needs to be stabilized with mood stabilizers and/or antipsychotics before stimulant treatment can be introduced.

Developmental Disorders. ADHD symptoms of inattention and hyperactivity are seen in children with intellectual and developmental delays, Autistic Spectrum Disorders (ASD) or both. Stimulants have been shown to be beneficial in both groups, particularly in those children with IQs greater than 50. MPH has been found to be more effective than placebo in reducing symptoms of ADHD in children with autism, and approximately one-third of children and adolescents with autism are on concomitant stimulant treatment. In the Research Units of Pediatric Psychopharmacology (RUPP) study, 49% of children with autism and ADHD responded to stimulants, with up to 18% unable to tolerate stimulant treatment. A trial of guanfacine found a 45% decrease in behavioral but not cognitive symptoms. ATX has also been found to be effective in the treatment of ADHD in autistic children.

Psychosocial Treatment of Children with ADHD

Psychosocial treatment of children with ADHD refers to nonmedication treatment. This type of treatment includes different modalities such as psychoeducation, academic organization skill teaching and remediation, parent training, behavior modification, CBT, social skills training, and individual therapy. Of these modalities, parent training, intensive behavior modification, social skills training, and academic organizational skills have shown efficacy for children with ADHD in controlled trials. Standardized parent training programs have been evaluated in systematic reviews including parent–child interaction therapy, friendship coaching, and the positive parenting program. Children with ADHD may also have difficulties regulating emotions and anger management and would benefit from nonpharmacologic interventions to address these potential problems.

Behavioral interventions are the preferred choice of treatment for preschool children with ADHD and are recommended as adjuncts to medication in school-aged children and adolescents. Parent training and intensive behavioral interventions are based on principles that will enhance parenting techniques. These include parental psychoeducation about the course, risk factors, and long-term outcomes of ADHD, as well as education on and encouragement of the use of strategies that will address behavior both in and outside of the household. Techniques such as token economy, contingency management, setting small achievable goals, checklists, and closer attention to a child’s behavior especially when the child is behaving appropriately are encouraged. While the effects on behavior and performance in the classroom have been shown, the effects do not tend to generalize to other settings. In practice, behavioral therapy is often carried out in consultation with parents and teachers to establish ideal in-school and home management. However, published trials tend to rely on parent and teacher ratings of improvement but do not have independent evaluators. There is some evidence in recent reviews that, even with blinded assessments, the effects of behavioral therapy persist for improving parenting and decreasing child conduct problems.

Academic interventions can take several forms to aid the child with ADHD. Organizational skills such as time management education are useful and have been shown to be effective. However, for some children, their academic difficulties may have put them behind their peers, and they might also require some level of academic support or remediation. A diagnosis of ADHD may qualify a child for special education or related services, as well as accommodations in a regular classroom. All these avenues should be explored to best support the child’s learning.

Certain alternative interventions including physical activity and mindfulness may be useful supplements in the treatment of ADHD but have not yet been rigorously studied. Other interventions such as dietary interventions, neurofeedback/biofeedback, and mineral and vitamin supplementations are still largely unproven. Recent studies suggest some benefit from fish oil.

Practice guidelines published for child and adolescent psychiatrists strongly suggest a combination of behavioral interventions, such as parent training or classroom management, alongside pharmacologic treatment in the management of ADHD. This is especially important due to the potential deficits along multiple axes in ADHD, and this combined multimodal treatment would benefit most the children who are responding only partially to stimulant medication or who exhibit a comorbid psychiatric disorder whose impairment is not responding to the stimulant medication.

SUMMARY AND CONCLUSIONS

ADHD is a prevalent condition that begins in childhood but often continues into adolescence and adulthood. It is accompanied by significant academic, occupational, social, and emotional impairment. The patients with ADHD often have other comorbid conditions. These comorbidities include oppositional defiant, conduct, learning, and anxiety disorders in childhood and anxiety, depression, and substance use and abuse in adolescence and adulthood.

Treatment must focus on both the ADHD and the various comorbidities. Thus, pharmacotherapy and psychosocial treatment as well as appropriate academic programs and support are needed for optimal outcome. It must be stressed that ADHD is a chronic condition for which ongoing long-term monitoring and treatment is required to optimize functioning.

SPECIFIC CASE OF CHILDHOOD ADHD

Ben is an 8.5-year-old adopted boy who lives with his parents and 11-year-old brother. He attends third-grade regular education in his local public school and received special services including speech, physical, and occupational therapy. Ben had suffered several upper respiratory infections at age 2 years requiring several rounds of antibiotics. Developmentally, some motor delays were reported, with crawling by 1 year and walking by age 2 years. He was speaking one word by his first birthday, but by age 3½ years, his expressive language was behind his peers in fluency. He demonstrated difficulties following multipart directions, even though his hearing tested in the normal range.

Ben has had difficulty with daydreaming and wandering about since kindergarten. His teachers complained that Ben did "not seem to listen," had "poor concentration," and was constantly "out of his seat." Ben had never had close friends and other children seemed to avoid him because he could not stick with an activity for very long and he was always "in their space." Ben's adoptive parents were drained from the nightly battles over homework and discouraged by his lack of focus and "spaciness." His inattention and impulsivity put him in danger and caused injury. He had run into traffic on several occasions, even after being hit by a truck with resulting injuries.

Ben's parents and his third-grade teachers completed the SNAP rating scale. Clinician assessment revealed seven symptoms of inattention. His mother also endorsed all nine symptoms of hyperactivity/impulsivity. These symptoms were moderately to severely impairing across multiple settings, including school, home, peer relations, personal safety, and during athletics.

Ben was started on extended-release MPH at 18 mg daily, which was later increased to 27 mg daily. His teachers reported improvement in attention and hyperactive behavior at school. Homework was no longer such a struggle, and he was noted to play quietly with his toys in the evenings.

ACKNOWLEDGMENTS

I would like to acknowledge the contribution of Dr. Laurence Greenhill, who was a coauthor of this chapter in several previous editions. His contribution to this chapter continues to be felt and is much appreciated.

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Andrade BF, Courtney D, Duda S, et al. A systematic review and evaluation of clinical practice guidelines for children and youth with disruptive behavior: rigor of development and recommendations for use. *Clin Child Fam Psychol Rev*. 2019;22(4):527–548.

Barbaresi WJ, Campbell L, Diekroger EA, et al. Society for developmental and behavioral pediatrics clinical practice guideline for the assessment and treatment of children and adolescents with complex attention-deficit/hyperactivity disorder. *J Dev Behav Pediatr*. 2020;41 Suppl 2S:S35–S57.

Becker SP. Systematic review: assessment of sluggish cognitive tempo over the past decade. *J Am Acad Child Adolesc Psychiatry*. 2021;60:690–709.

Catalá-López F, Hutton B, Núñez-Beltrán A, et al. The pharmacological and non-pharmacological treatment of attention deficit hyperactivity disorder in children and adolescents: a systematic review with network meta-analyses of randomised trials. *PLoS One*. 2017;12(7):e0180355.

Chen Q, Brikell I, Lichtenstein P, et al. Familial aggregation of attention-deficit/hyperactivity disorder. *J Child Psychol Psychiatry*. 2017;58(3):231–239.

Cortese S, Adamo N, Del Giovane C, et al. Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: a systematic review and network meta-analysis. *Lancet Psychiatry*. 2018;5(9):727–738.

Cothros N, Medina A, Pringsheim T. Current pharmacotherapy for tic disorders. *Expert Opin Pharmacother*. 2020;21(5):567–580.

Coughlin CG, Cohen SC, Mulqueen JM, Ferracioli-Oda E, Stuckelman ZD, Bloch MH. Meta-analysis: reduced risk of anxiety with psychostimulant treatment in children with attention-deficit/hyperactivity disorder. *J Child Adolesc Psychopharmacol*. 2015;25(8):611–617.

Daley D, Van der Oord S, Ferrin M, et al; European ADHD Guidelines Group. Behavioral interventions in attention-deficit/hyperactivity disorder: a meta-analysis of randomized controlled trials across multiple outcome domains. *J Am Acad Child Adolesc Psychiatry*. 2014;53(8):835–847, 847.e1–e5.

Danielson, ML, Bitsko RH, Ghandour RM, Holbrook JR, Kogan MD, Blumberg SJ. Prevalence of parent-reported ADHD diagnosis and associated treatment among US children and adolescents, 2016. *J Clin Child Adolesc Psychol*. 2018;47(2):199–212.

Demontis D, Walters RK, Martin J, et al; ADHD Working Group of the Psychiatric Genomics Consortium (PGC), Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium, 23andMe Research Team. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet*. 2019;51(1):63–75.

Faraone SV, Banaschewski T, Coghill D, et al. The World Federation of ADHD International Consensus Statement: 208 evidence-based conclusions about the disorder. *Neurosci Biobehav Rev*. 2021;128:789–818.

Faraone SV, Larsson H. Genetics of attention deficit hyperactivity disorder. *Mol Psychiatry*. 2019;24(4):562–575.

Greenhill L, Kollins S, Abikoff H, et al. Efficacy and safety of immediate-release methylphenidate treatment for preschoolers with ADHD. *J Am Acad of Child Adolesc Psychiatry*. 2006;45(11):1284–1293.

Greenhill LL, Swanson JM, Hechtman L, et al; MTA Cooperative Group. Trajectories of growth associated with long-term stimulant medication in the multimodal treatment study of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*. 2020;59(8): 978–989.

Hoogman M, Muetzel R, Guimaraes JP, et al. Brain imaging of the cortex in adhd: a coordinated analysis of large-scale clinical and population-based samples. *Am J Psychiatry*. 2019;176(7):531–542.

Joshi G, Wilens T, Firmin ES, Hoskova B, Biederman J. Pharmacotherapy of attention deficit/hyperactivity disorder in individuals with autism spectrum disorder: a systematic review of the literature. *J Psychopharmacol*. 2021;35(3):203–210.

Kooij JJS, Bijlenga D, Salerno L, et al. Updated European Consensus Statement on diagnosis and treatment of adult ADHD. *Eur Psychiatry*. 2019;56:14–34.

The MTA Cooperative Group. Moderators and mediators of treatment response for children with attention-deficit/hyperactivity disorder: the multimodal treatment study of children with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry*. 1999;56(12):1088–1096.

Nakao T, Radua J, Rubia K, Mataix-Cols D. Gray matter volume abnormalities in ADHD: voxel-based meta-analysis exploring the effects of age and stimulant medication. *Am J Psychiatry*. 2011;168(11):1154–1163.

Nigg JT, Lewis K, Edinger T, Falk M. Meta-analysis of attention-deficit/hyperactivity disorder or attention-deficit/hyperactivity disorder symptoms, restriction diet, and synthetic food color additives. *J Am Acad Child Adolesc Psychiatry*. 2012;51(1):86–97.

Pereira-Sanchez V, Castellanos FX. Neuroimaging in attention-deficit/hyperactivity disorder. *Curr Opin Psychiatry*. 2021;34(2):105–111.

Pettersson E, Lichtenstein P, Larsson H, et al. Genetic influences on eight psychiatric disorders based on family data of 4 408 646 full and half-siblings, and genetic data of 333 748 cases and controls. *Psychol Med*. 2019;49(7):1166–1173.

Pliszka S; AACAP Work Group on Quality Issue. Practice parameters for the assessment and treatment of childhood attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*. 2007;46(7):894–921.

Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am J Psychiatry*. 2007;164(6):942–948.

Pliszka SR, Greenhill LL, Crismon LM, et al. The Texas Children's Medication Algorithm Project: Report of the Texas Consensus Conference Panel on Medication Treatment of Childhood Attention-Deficit/Hyperactivity Disorder. Part II: Tactics. Attention-Deficit/Hyperactivity Disorder. *J Am Acad Child Adolesc Psychiatry*. 2000;39(7):920–927.

Vitiello B, Lazzaretto D, Yershova K, et al. Pharmacotherapy of the preschool ADHD treatment study (PATS) children growing up. *J Am Acad Child Adolesc Psychiatry*. 2015;57(7):550–556.

Willcutt EG. The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. *Neurotherapeutics*. 2012;9(3):490–499.

Wolraich ML, Hagan JF Jr, Allan C, et al; Subcommittee on Children and Adolescents with Attention-Deficit/Hyperactive Disorder. Clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Pediatrics*. 2019;144:e20192528.

44.2 Adult Manifestations of Attention-Deficit/Hyperactivity Disorder

JAMES J. MCGOUGH, M.D., M.S.

Although attention-deficit/hyperactivity disorder (ADHD) was originally conceptualized as a developmental delay of childhood, it is now recognized as a lifespan neurodevelopmental disorder with widely acknowledged adult manifestations. Longitudinal studies indicate that a majority of school-age children diagnosed with ADHD continue to have clinically meaningful symptoms and related impairments as young adults. Community epidemiologic studies suggest that prevalence rates of ADHD among adults are among the highest of all psychiatric disorders. Adults with ADHD have patterns of psychiatric comorbidity and neurocognitive deficits similar to children but suffer from a much broader array of functional impairments. Both structural and functional brain imaging studies demonstrate distinct abnormalities in groups of adults with ADHD compared to those without, and genetics studies conducted in adults have replicated pediatric findings. Adult ADHD is associated with significant medical, occupational, economic, and social impacts and has emerged as a major public health concern.

DEFINITION

ADHD was classified as a Neurodevelopmental Disorder in the 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) and its text revision (DSM-5-TR), reflecting the disorder's initial emergence in childhood and common co-occurrence with other early-onset brain-based syndromes. In a departure from previous editions, DSM-5 fully acknowledged the persistence of ADHD across the lifespan with modified diagnostic criteria for older adolescents and adults to better reflect developmental changes in symptom sensitivity with increasing age. While criteria in younger patients require evidence of six or more inattentive or hyperactive/impulsive symptoms, this threshold was reduced to five symptoms in either domain for the diagnosis of patients aged 17 years and older. This modification recognized that the population prevalence of ADHD symptoms decreases with increasing age, but adults with as few as four symptoms suffer significant clinical impairments compared with similarly aged peers. As in youth, symptoms must have been present for a minimum of 6 months and exceed what is expected for the patient's expected developmental level. Additionally, there must be some evidence of symptoms prior to age 12 with concomitant functional impairments in multiple life settings, such as home, school, work, social relationships, or personal activities. Current symptoms must be associated with a clinically significant level of functional impairment. Symptoms cannot occur exclusively during an episode of schizophrenia or other psychotic illness and cannot be better attributed to other psychiatric conditions such as mood, anxiety, dissociative, or substance use disorders.

DSM-5 eliminated ADHD subtypes. Instead, *specifiers* describe the overall clinical presentation evidenced in the prior 6 months. Adults with at least five symptoms of inattention and hyperactivity/impulsivity exhibit the combined presentation. Those with five or more inattentive but fewer than five hyperactive/impulsive symptoms exhibit the inattentive presentation. Those with a minimum of five hyperactive/impulsive but fewer than five inattentive symptoms exhibit the hyperactive/impulsive presentation.

Other specifiers describe disorder severity. *Mild ADHD* is specified when the minimal number of symptoms required for diagnosis occur and functional difficulties are at the threshold for clinical significance. *Severe ADHD* is specified when there are many symptoms in excess of the required threshold, or in the presence of marked functional difficulties in one or more settings. Moderate ADHD is specified when the number of symptoms and associated degree of impairment falls between “mild” and “severe.”

DSM-5 further permitted the diagnosis of “other specified ADHD” when an individual exhibits significant impairment due to ADHD symptoms but fails to meet full diagnostic criteria. In this instance, the clinician should specify which criteria are not met, for example, “other specified ADHD with insufficient symptom number.” Clinicians may also assign the diagnosis “unspecified ADHD” if they choose not to specify which criteria are lacking.

HISTORY

For the first half of the 20th century, the condition now recognized as ADHD was variably referred to as minimal brain damage (MBD), minimal brain dysfunction, or hyperkinesis. While these terms conveyed some expectation that the disorder arose from underlying brain pathology, the syndrome was principally viewed as a developmental delay that was most often outgrown during adolescence. This position was maintained in the second edition of the DSM (DSM-II), published in 1968, which placed the diagnostic category *hyperkinetic reaction of childhood* within the grouping “Behavioral Disorders of Childhood and Adolescence.” Unique among editions of the DSM, DSM-II further implied that core features of short attention span, restlessness, and hyperactivity were a reaction to unsupportive environments and/or inadequate parenting.

In the 1970s, concurrent with development of the third edition of the DSM (DSM-III), seminal work by Paul Wender and colleagues at the University of Utah identified a group of adults with persistent childhood ADHD symptoms. Wender asserted that symptoms under consideration for inclusion in the DSM revision were developmentally inappropriate for older patients. He proposed the Wender Utah Criteria for Adult ADHD, empirically derived from extensive experience with older patients, as an alternative to the DSM. Wender Utah Criteria required (1) retrospective diagnosis of childhood ADHD made on the basis of patient interview and, when available, confirmation by a third-party informant; (2) evidence of ongoing impairment in adulthood from hyperactive and impulsive symptoms adulthood; and (3) the presence of additional symptoms felt to be typical of adults with the disorder, including mood lability, irritability, impulsivity, disorganization, and poor stress tolerance. The diagnosis could not be made in patients suffering from severe depression, psychosis, or severe personality disorders.

DSM-III, published in 1980, continued to define attention deficit disorder as a childhood condition. However, with some deference to work by Wender and others, DSM-III allowed the diagnosis of attention deficit disorder, residual type, for adults previously diagnosed as children who had ongoing evidence of some symptoms. This category inherently implied that adults could not meet full diagnostic criteria. Text of the revised third edition (DSM-III-R) allowed that ADHD might continue through adulthood in up to 30% of affected children.

Throughout the 1980s and 1990s, and in the absence of clearly applicable DSM criteria, Wender Utah Criteria were used widely in clinical practice as well as research studies of adult medication response, brain imaging, and genetics. Concurrently, longitudinal follow-up studies of adults previously diagnosed with childhood hyperactivity reported ADHD persistence in 40% to 60% of cases. These converging lines of evidence influenced subsequent DSM revisions and increased recognition of ADHD in adults. Although Wender Utah Criteria had a tremendous impact in promoting initial acknowledgment and study of adult ADHD, their use was largely abandoned as DSM criteria were increasingly applied outside of pediatric age groups. More recently, building on Wender's contributions, there has been renewed recognition of difficulties related to emotional dysregulation in some adults with ADHD, which, while not diagnostic, predict increased risk for comorbidity and functional disability.

In the 1990s, emerging research and several best-selling lay publications spurred widespread public awareness of adult ADHD. These studies and resultant increases in public demand for clinical services catalyzed growing interest in the phenomenology, causes, and potential treatments of adult ADHD. Nonetheless, although the fourth edition (text revision) of the DSM (DSM-IV-TR) provided further leeway for adult diagnosis, it continued to maintain that ADHD symptoms are typically resolved by late adolescence.

Revisions in DSM-5 incorporated more than two decades of research and clinical experience in adults. Enumerated ADHD symptom criteria remained unchanged from those previously identified in field trials of school-age children, but text clarifications were added to reflect developmental changes in symptom presentation that occur with increasing age. The age of onset criterion from previous editions of the DSM was modified from requiring evidence of symptoms and impairment prior to age 7 to requiring some recognition of symptoms alone prior to age 12. In response to studies that demonstrated declining symptom prevalence rates with increasing age, as well as significant functional impairments seen in adults with as few as four ADHD symptoms, DSM-5 required only five symptoms in either the inattentive or hyperactive/impulsive domain for diagnosis in individuals aged 17 and older. Finally, descriptions of typical impairments associated with ADHD were broadened to include the range of difficulties frequently seen in adults, notably problems with work, personal activities, social relationships, and general health maintenance. The text revision of the DSM-5, DSM-5-TR did not many any significant changes to this scheme.

Several birth cohort studies suggested the occurrence of adult-onset ADHD in individuals with no history of childhood symptoms, undermining the conceptualization of the disorder as a chronic neurodevelopmental syndrome emerging before age 12. These studies were largely symptom-based and did not require a demonstration of associated functional impairments. One subsequent investigation found that the failure to detect childhood ADHD was a function of higher IQ and lack of comorbidity. Other studies found that over 95% of supposed adult-onset cases failed to retain the diagnosis after comprehensive examination but instead generally suffered from other psychiatric disorders or functional impairments due to heavy substance use. Current evidence suggests that true adult-onset ADHD occurs rarely or not at all. Further research is necessary to better elucidate this question.

COMPARATIVE NOSOLOGY

The 11th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-11) harmonized with DSM-5-TR in renaming the ICD-10 category of hyperkinetic disorders as ADHD and moving it from the grouping of disruptive behavior to the neurodevelopmental domain. Consistent with DSM-5-TR, ICD-11 required *either* inattentive *or* hyperactive/impulsive symptoms, not both as in previous editions, and allows the use of predominately inattentive, hyperactive/impulsive, or combined presentation specifiers. Unlike DSM-5-TR, ICD-11 does not define a precise age of onset, specific symptom number, or duration. Also, unlike DSM-5-TR, ICD-11 continued to exclude the diagnosis in the presence of autism spectrum disorder. ICD-11 is the principal classification system used outside the United States and generally relied upon within the United States for insurance billing purposes.

EPIDEMIOLOGY

Epidemiologic studies demonstrated that ADHD is the most common behavioral disorder of childhood, with prevalence estimates ranging from 5% to 10% of school-age children. Longitudinal studies suggest the persistence of the disorder from childhood to adulthood in approximately 65% of cases. The World Health Organization estimates the prevalence of adult ADHD to be 2.8%. However, there is a wide range of estimates across studies and countries (e.g., 1.1% in South Korea vs. 5.8% in Brazil). Differences in reported prevalence rates are deemed related to methodologic differences between investigations. These converging reports confirm that significant numbers of adults suffer from ADHD and related impairments. Unlike other common psychiatric conditions, such as mood and anxiety disorders that are typically episodic, adults with ADHD are apt to experience chronic and unremitting symptoms that require ongoing management.

ETIOLOGY

ADHD is best viewed as a neurodevelopmental disorder that arises from the interplay of genetic and environmental factors. Similar to studies in youth, meta-analyses suggest heritability estimates in the range of 70% to 80% in adults, with approximately one-third of this risk explained by polygenic effects of common variants with small individual effects. Adult genome-wide association studies (GWAS) identified a similar set of risk loci as seen in studies of youth, supporting the conceptualization of ADHD as a neurodevelopmental disorder with a significant genetic overlap with ADHD in children. Approximately one-third of adults with ADHD have at least one child with the disorder, and similarly, one-third of children with ADHD have at least one affected parent. Ongoing investigations will further elucidate the genetic, environmental, and gene-by-environment interactions that affect ADHD risk.

Early structural and functional brain imaging studies in groups of adults with ADHD revealed localized dysfunctions in discrete areas of the prefrontal cortex and basal ganglia, similar to findings in children. Longitudinal imaging studies in youth demonstrated maturational delays in cerebral cortex myelination. Some have suggested that cortical maturation is associated with symptom attenuation in adulthood, whereas ongoing subcortical dysfunction predicts symptom persistence. Newer imaging research has shifted from examining distinct anatomical abnormalities to consideration of impaired functional connectivity in brain regions engaged in executive, attention, reward, and default networks. The heterogeneity of imaging findings reflects clinical heterogeneity, but ongoing research might provide insights into the persistence or desistance of childhood ADHD into adulthood

Meta-analyses suggest adult ADHD persistence in 65% of child cases, with some estimated variability due to differing methodologies. Approximately 15% of children demonstrate full diagnostic criteria as adults; an additional 40% to 60% show partial remission. Risk factors for persistence include the severity of childhood ADHD symptoms, the presence of childhood comorbidity, and parental mental health. There is no apparent impact from individual IQ, socioeconomic status, or quality of the parent–child relationship. Comorbid juvenile conduct disorder is the strongest predictor of severe adult outcomes, including increased risk for substance use disorders, criminality, and early mortality. Some evidence suggests that early and sustained medication treatment of childhood ADHD is associated with decreased risk for substance use disorders as young adults.

DIAGNOSIS AND CLINICAL FEATURES

Although parents and teachers generally serve as primary informants in the assessment of pediatric ADHD, adult patients themselves are usually the primary sources of information in their own assessments. Unfortunately, research showed that young adults are frequently poor reporters of ADHD symptoms and associated difficulties, although reliability is much improved when patients reach their late 20s. Studies also revealed that both adults with ADHD and parents of these adult children exhibited poor retrospective recall of problems related to childhood ADHD, often reporting ages of onset up to 4 years later than documented in prospectively collected data. As with children, optimal assessment of adult ADHD depends on the clinical synthesis of all available information, including reports of childhood and current difficulties, symptom reports from parents and significant others, school and occupational reports, and a review of pertinent medical records.

Table 44.2–1.
DSM-5 Symptoms and Adult Correlates

Inattention		Hyperactivity and Impulsivity	
Younger Individuals	**Older Individuals**	**Younger Individuals**	**Older Individuals**
Fail to notice small details, make careless errors	Careless, inaccurate work	Frequently fidget, tap hands or feet, squirm while seated	Often fidgety or restless
Difficulty maintaining attention	Problems staying focused on boring work, mind often wanders, trouble with extensive reading	Frequently get out of seat or have trouble staying seated	Can’t sit long at meetings or other activities where seating is required
Seem not to listen even when spoken to directly	Trouble retaining what people say even when addressed directly	Often run or climb at inappropriate times	Often feel internally restless or unsettled
Trouble finishing work or following through on instructions	Sidetracked easily, problems following directions or finishing tasks	Have trouble playing quietly; often too loud	Often louder than necessary or appropriate
Problems with organization	Can’t prioritize, procrastinate, messy, miss deadlines, and bill payments	Often “on the go,” as if “driven by a motor”	Difficulty unwinding or relaxing; always needing to be busy

Avoid activities that require sustained concentration	Delay and avoid work and other activities that are boring or repetitive	Get in trouble for excessive talking	Talk too much in social settings, don’t tolerate silence
Lose or misplace things frequently	Often misplace things needed at home, work, or other activities	Often blurt out answers before questions are finished	Finish others’ sentences or answers before questions are completed
Distracted easily when interrupted	Easily distracted by surrounding noises, interruptions, or extraneous thoughts	Have difficulties when waiting for turn or standing in line	Frustrated when needing to wait
Frequently forgetful	Fail to remember appointments, deadlines, and other obligations	Frequently interrupt or intrude	Interrupt others’ conversations and activities, socially intrusive

Data from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

Consideration of an ADHD diagnosis should be given to any adult with long-standing difficulties from inattention or distractibility. A single DSM symptom, “Is often distracted by extraneous stimuli,” correctly differentiated 97% of ADHD-affected adults compared with 98% of community controls. One additional symptom, “Avoids tasks requiring sustained mental effort,” correctly identified 89% of ADHD cases when compared with other clinical cases without ADHD.

As with children and adolescents, the diagnosis of ADHD in adults is based on DSM-5-TR criteria. DSM-5-TR acknowledged developmental changes observed in adults by requiring only five inattentive and/or hyperactive/impulsive symptoms for diagnosis, compared with six for younger individuals. ADHD symptoms in DSM-5-TR are identical to DSM-IV and were initially derived from field trials conducted in school-age children. However, text descriptions in DSM-5-TR provided adult correlates for these symptoms and allow greater applicability to older patients, as summarized in Table 44.2–1.

Diagnosis is based principally on information derived from a clinical interview with the adult patient. The self-administered Adult ADHD Self Report Screening Scale (ASRS), developed under the auspices of the World Health Organization and calibrated to DSM-5, is easily administered and has high diagnostic sensitivity and specificity. A related semistructured interview, the Adult ADHD Clinical Diagnostic Scale (ACDS), probes retrospective childhood and prior 6-month adult ADHD symptoms. ADHD assessment requires careful consideration of potentially comorbid psychopathology as well as a broad range of functional impairments. Common co-occurring conditions include lifetime histories of disruptive behavior and antisocial activities, nicotine dependence, substance use disorders, anxiety disorders, depression, risky behavior, and poor health maintenance. Typical impairments include difficulties in academic, occupational, social, and family functioning.

Many clinicians remain highly skeptical of the validity of adult ADHD and presume that most adults seek the diagnosis to obtain various academic or occupational accommodations or scheduled medications. ADHD symptoms are easily feigned. Clinicians should be particularly suspicious about adults claiming to suffer severe symptom levels without reasonable evidence of prior functional impairments. Several investigations compared individuals pretending to have ADHD with known adult cases and found that malingerers had atypically high rating scale scores and greater numbers of self-reported ADHD symptoms. This finding contrasts with the anticipated age-related symptom reduction typically seen in the disorder. Similarly, on computerized attention tasks, those feigning ADHD made many more errors in omission and commission than individuals with historically validated cases, suggesting deliberate attempts to perform poorly. Some have proposed that computerized performance validity tests or symptom validity tests, such as the Exaggeration Index of the Connors Adult ADHD Rating Scale, aid in recognizing malingering, but evidence supporting this approach remains inconclusive. Clinical corroboration of self-reported ADHD symptoms and expected associated impairments with third-party informants, past records, and an overall life course remains the best approach for establishing the adult diagnosis.

Laboratory Studies

There are insufficient data to support the use of neuropsychological tests, computerized tests of attention (CPTs), or brain imaging modalities in the diagnosis of adult ADHD. Neuropsychiatric testing can be useful in differentiating ADHD from specific learning disabilities and other cognitive impairments, particularly with a predominance of inattentive symptoms or a history of school failure. Nonetheless, neuropsychiatric testing has no demonstrated value in assessing ADHD itself. Easily administered self-report measures, such as the Behavior Rating Inventory of Executive Function (BRIEF), can provide evidence of typical ADHD-related cognitive impairments supportive of the diagnosis but fail to identify deficits in half of the cases. Research studies using electroencephalography (EEG), functional magnetic imaging (fMRI), positron emission tomography (PET), and other imaging modalities reveal distinct features in groups of adults with ADHD compared with nonclinical controls. However, these findings have poor predictive value in assigning diagnoses to individual patients, and no evidence supports their utility in clinical settings. As with children and adolescents, adult ADHD remains a clinical diagnosis derived through careful history taking, adherence to clinical criteria, and differential diagnosis of other psychopathology.

Differential Diagnosis

Clinicians should consider a diagnosis of ADHD in any adult patient complaining of pervasive difficulties with inattention, hyperactivity, or impulsivity. Several ADHD symptoms, such as problems with sustained attention, motor restlessness, increased speech, and acting “on the go,” are common in other psychiatric disorders. Clinicians must distinguish ADHD from other psychiatric syndromes while remaining sensitive to a high probability of coexisting psychopathology.

Depressive disorders are differentiated from ADHD in that problems with concentration and motor restlessness occur only during periods of dysphoric mood, anhedonia, or increased irritability. Discrete episodes of somato-vegetative symptoms, such as appetite changes and alterations in sleep patterns, further confirm a depression diagnosis. Patients with both ADHD and depression will report ongoing difficulties with core ADHD symptoms that persist outside of discrete mood episodes. Similarly, patients with bipolar I disorder report racing thoughts, rapid speech, and impulsive behaviors only in the context of sustained periods of elevated or irritable mood. Somewhat more controversial is the overlap of ADHD with hypomania and rapidly cycling bipolar disorder. A clear past history of discrete depressive episodes or strong family history might assist in clarifying the diagnosis, but some patients will meet the criteria for both conditions. Anxiety disorders are distinguished from ADHD by subjective worry that is either generalized and difficult to control or specific to situations or stimuli. Patients with substance use disorders and ADHD should have histories of ADHD symptoms that precede the initiation of substance use. Patients with histories of early school difficulties and childhood conduct disorder are likely to be comorbid for both ADHD and substance use. Adults with ADHD and a childhood history of conduct disorder might also exhibit antisocial personality disorders.

Adults with histories of school difficulties and inattentive symptoms in the absence of hyperactivity or impulsivity should be evaluated for specific learning disabilities or other executive functioning deficits. This generally requires a battery of neurocognitive tests to measure academic aptitude and achievement. Patients with significantly idiosyncratic interests and impaired social functioning might also be diagnosed with an autism spectrum disorder.

Although considerable research has described considerable psychiatric comorbidity in adults diagnosed with ADHD, far less attention has been paid to the prevalence rates of ADHD co-occurring with other adult psychiatric disorders. Evidence suggests up to 16% of adults currently suffering from major depression, 10% with current anxiety disorders, and up to 25% with current substance use disorders meet ADHD criteria. In view of its high overall adult prevalence and associated developmental risk for comorbid psychopathology, all adult psychiatric assessments should include routine diagnostic consideration of ADHD.

A 42-year-old woman was evaluated for complaints of distractibility and general nervousness. The patient had spent the past 16 years at home caring for her children. Her youngest child recently began elementary school, and she obtained part-time work as an assistant manager in a large retail department store. She was excited about the position, particularly the flexible work hours and the opportunity to earn extra income for the family. Nonetheless, she reported particular difficulty remaining focused on her job, had general trouble organizing her time, and was in danger of losing her job due to her failure to meet several deadlines. She was tense, irritable, had trouble relaxing, and described difficulty falling asleep because she was worried about what she needed to do the next day. She complained her "mind goes completely blank" when overstressed. She denied depressed mood, loss of pleasure in her usual activities, change of appetite, or suicidal preoccupations.

The patient reported that she was considered a bright student, particularly in elementary school. She enjoyed school but was generally shy and uncomfortable in unfamiliar social situations. Beginning in middle school and throughout college, she felt increasingly bored, easily distracted, and particularly frustrated with long reading assignments. She graduated with a 3.0-grade point average and subsequently worked for 2 years as a retail clerk. She became pregnant soon after her marriage and left her job to raise her family. She described her marriage as "typical," generally stable and satisfying, with occasional marital tension due to financial stresses and child-rearing. Her middle child, a 12-year-old boy, had been diagnosed with ADHD, which was successfully managed with stimulant therapy and behavioral techniques learned through parent training.

The patient completed an Adult ADHD Self Report Screening Scale (ASRS), which suggested the presence of a high number of symptoms. An ADHD Rating Scale 5 (ADHD-RS-V) completed by her husband reflected a similar symptom pattern. Clinical synthesis of all available information supported an ADHD diagnosis with a retrospectively determined age of onset in middle school, a history of impairment at school, work, and home, and a comorbid diagnosis of generalized anxiety disorder.

The treating clinician outlined several options for medication therapy, specifically an initial stimulant trial with subsequent selective serotonin reuptake inhibitor (SSRI) treatment for residual anxiety versus atomoxetine (Strattera) monotherapy. The patient agreed to a medication trial with the main target symptoms of distractibility, nervousness, and sleep disturbance.

On stable medication, she reported improved focus at work, better general organization, less general anxiety, and improved sleep. She was initially followed with monthly medication visits, which were reduced to visits every 3 months once her clinician deemed her to be stable. After 6 months of medication treatment, the patient was pleased to report that she recently finished reading a novel, her first since college. She resolved her initial difficulties at work and was given a commendable employment evaluation. She was feeling less stressed at home but remained troubled by what she viewed as her inadequacies and inability to manage all her responsibilities. In response, her clinician referred her for a brief course of cognitive-behavioral therapy (CBT) to address residual anxiety symptoms and organizational skills.

COURSE AND PROGNOSIS

The prevalence of ADHD symptoms declines with age. Many children with the combined presentation of ADHD exhibit fewer hyperactive-impulsive symptoms by midadolescence and subsequently exhibit the inattentive presentation as adults. Given the age-related decline of ADHD symptoms in the general population, adults can remain significantly impaired relative to same-age peers with fewer symptoms than are required for diagnosis in childhood. This suggests that reported rates of ADHD remission are more a function of syndrome definition than true syndrome remission. Long-term outcome studies of ADHD-affected children are limited, and there are insufficient data to draw conclusions regarding the developmental course of ADHD from early to middle adulthood. Between 25% and 45% of ADHD children do not suffer ADHD-related impairments as adults. Nonetheless, many adults with ADHD continue to struggle with the consequences of the disorder throughout their lives.

Adults with ADHD have significant rates of co-occurring psychopathology and limitations in adaptive functioning when compared with unaffected peers. They exhibit higher rates of childhood conduct disorder, with a subset of these developing antisocial personality disorders. Significantly higher rates of substance use disorders are consistently reported compared with adults without ADHD. There is no clear evidence that adults with ADHD prefer one type of recreational drug compared to another, but those who develop drug dependency have a more difficult time achieving sobriety than individuals without ADHD. Similarly, adults with ADHD initiate nicotine use at earlier ages, demonstrate higher rates of lifetime nicotine addiction, and have more difficulty maintaining smoking cessation efforts. Adults with ADHD suffer high rates of lifetime mood and anxiety disorders. The differential impact of comorbid psychopathology on long-term ADHD outcomes has not been adequately studied.

ADHD confers a significant risk for lifetime functional impairments. Lower self-esteem, poorer lifetime academic achievement, less successful work histories, and more failures in long-term relationships are evident. Adults with ADHD often make poor choices regarding personal health, including participation in high-risk activities, unprotected sex, and unsafe driving. Conversely, some of these succeed in making lifestyle adjustments that minimize evidence of impairment. Individuals find occupations for which they are well suited, can live alone or find suitable close relationships, or can utilize other talents that compensate for ADHD-related difficulties. Additional research is required to identify predictive factors that mitigate these impairments.

A 26-year-old man was evaluated for severe distractibility after being involved in a minor traffic accident. In early grade school, he had particular difficulties making careless mistakes on tests, forgetting his assignments, and sitting still in class. Although he was considered intelligent, he earned average grades, and he had no college aspirations. After high school, he had several short-term jobs in the construction industry, which were enjoyable because he spent most of his time outdoors. In the past year, he obtained part-time work as a wedding videographer and found himself increasingly interested in making and editing videos.

During his clinical interview, the patient described ongoing difficulties maintaining his attention on uninteresting tasks, easy distractibility, problems with time management, generally feeling restless, and frustration when required to sit for extended periods. He endorsed six inattentive and five hyperactive/impulsive symptoms on a DSM-5-TR ADHD checklist of current symptoms. After a clinical interview, the patient was diagnosed with the combined presentation of adult ADHD.

After discussing the nature of ADHD with his clinician and receiving additional educational materials, the man agreed to a medication trial. A once-daily extended-release preparation of stimulant medication was prescribed in escalating doses over several weeks to determine the optimal dose that maximized symptom reduction and minimized side effects. At his follow-up visit, the patient described robust improvement in his ability to focus and was surprised at his increased motivation to attend to the technical details of video editing. After several months, he enrolled in a community college to pursue formal coursework in video production.

TREATMENT

Good long-term clinical care is contingent on adequate patient education about the nature of the disorder and the effects of treatment. Once a diagnosis is made, clinicians should provide sufficient information to assist adult patients in choosing treatment options that meet individual patient needs. Given the degree of misinformation about ADHD from the Internet and other lay sources, patients are well served when clinicians guide them to appropriate reference materials. Many individuals are relieved to understand that they have a biologically based condition that is shared by a significant number of other adults. National organizations such as Children and Adults with ADD (CHADD, www.chadd.org) or the National Alliance for the Mentally Ill (NAMI, www.nami.org) provide access to local support networks and reliable educational materials.

Pharmacotherapy remains a treatment cornerstone. Medications found to be useful in the treatment of childhood ADHD are generally shown to be effective in adults. FDA-approved medications for adult ADHD include both stimulants and nonstimulants. Of the nonstimulants, atomoxetine has the greatest amount of clinical trial evidence supporting its use in adults. Medications commonly used off-label include bupropion, immediate alpha-2-agonists, tricyclic antidepressants, and the wake-promoting agents modafinil and armodafinil. Stimulants are the most frequently prescribed ADHD medications in adults, and their greatest clinical effects are seen with robust dosing. One noteworthy difficulty for adults is the frequent necessity for multiple daily dosing, even with extended-release formulations, in order to provide adequate symptom relief throughout the day. College students, in particular, have irregular daily class schedules and might not be best served with reliance on morning-administered extended-release stimulants. Well-considered use of immediate-release compounds is necessary at times to meet the clinical needs of these patients.

Many clinicians are reluctant to prescribe stimulants based on concerns about abuse, misuse, and illicit diversion. Misuse and diversion of prescribed stimulants among college students are common. In most cases, these medications are used with the expectation of improving academic performance, although there are no data to support this outcome. Approximately 30% of stimulant misuse among college-age students is for recreational purposes, that is, to cause euphoria. Misuse for academic purposes is highest among white male students attending academically competitive universities, who are in fraternities, are in academic distress, and with prior histories of conduct-disordered behaviors and recreational drug, alcohol, and nicotine use. Several research reports recommend using prodrug and certain extended-release stimulant formulations, deemed less likely to be used for recreational purposes compared with older immediate-release forms if the clinician is concerned about potential misuse. Nonstimulant medications remain an option in patients at high risk for stimulant abuse or misuse. Proper record keeping and thorough documentation of prescriptions written are essential to good practice. When available, clinicians can verify that patients are not obtaining inappropriate amounts of controlled substances by checking prescription records on government-sponsored databases.

As they move toward independent living outside the structure of high school and their parents' homes, transitional-age youth, that is, those ranging in age from 16 to 26 years, have particular clinical needs. During these years, individuals with ADHD have an increased risk for substance use problems, academic underperformance, occupational failures, and relationship difficulties. Treatment adherence and responsible medication use might also be of concern. In working with this age group, clinicians should be particularly mindful of providing adequate counseling regarding responsible sexual activity, alcohol consumption, recreational drug use, and appropriate medication use.

Although many adults have significant lifetime histories of other psychiatric disorders, there have been few treatment studies for comorbid ADHD. Current evidence suggests that clinicians should first initiate therapy for the condition (i.e., ADHD vs. other psychopathology) that is causing the most distress. One example is sequential pharmacotherapy with either a stimulant or SSRI for co-occurring ADHD and anxiety or depression. A particularly vexing problem is whether stimulant medications should be provided to patients with known histories of substance use disorder. It is essential that physicians maintain proper documentation of patients' diagnostic assessments and prescription medication records. Some clinicians are comfortable initiating ADHD pharmacotherapy in the presence of intermittent alcohol or cannabis use but are advised against stimulant pharmacotherapy in those patients with chronic use or the use of other psychoactive agents such as opioids or methamphetamine.

There is growing support for certain nonmedication treatments in adult ADHD. The strongest evidence exists for CBT, which has been administered successfully both in individual and group settings. Well-controlled blinded investigations of omega-3 fatty acid supplementation have also yielded small but significant positive effects on ADHD symptoms. Other nonmedication approaches, such as ADHD coaching, mindfulness training, meditation, yoga, cognitive remediation programs, EEG neurofeedback, and the use of digital technologies such as electronic scheduling and reminder systems, remain popular, despite a lack of sufficient scientific data to support their general use. Individuals are best served when psychosocial interventions are individually tailored to their own specific difficulties. For example, patients who suffer from marital or relationship difficulties might be referred to couples therapy. The emphasis in this treatment should be on the couple's relationship problems, not the individual's ADHD. Parents might benefit from participation in a parenting course, particularly if they themselves have children with ADHD. Insight-oriented psychotherapy might be useful for some patients with significant self-esteem or interpersonal difficulties.

Proper management of adult ADHD is also likely to include appropriate educational planning and consideration of academic and occupational accommodations. Occupational counseling and vocational training can be better options for some individuals than insistence on attending a 4-year college. Adults who attend college should be referred to their school's office for students with disabilities. The Americans with Disabilities Act (ADA) has particular implications for adults with ADHD. Regulations implementing the ADA require that *reasonable* accommodations be made in areas of education and employment for individuals with documented disabilities. Colleges and universities develop their own guidelines for the provision of special accommodations based on Federal mandates. At higher-grade levels and for standardized testing at the college level and beyond, it is generally necessary to conduct formal neuropsychological testing to document ADHD-related learning impediments prior to granting accommodations. In occupational settings, the ADA requires that reasonable accommodations be made for persons with disabilities but does not require that anyone be hired for a position who is not otherwise qualified. It is essential that thresholds for disability be determined in reference to population or societal norms and not speculations about the individual’s potential capabilities. *Reasonable* accommodations should have minimal impact on employer expectations, job performance, and coworkers.

A 19-year-old woman was evaluated at a drug rehabilitation center for methamphetamine dependence. In early grade school, she was extremely hyperactive, reckless, and prone to impulsive judgments. Her father suffered from alcohol dependence and was physically abusive to her and her mother. She began smoking cigarettes at age 11 and using alcohol and marijuana at age 12. At age 13, she was raped by a high school senior whom she had been dating. Two years later, she quit high school and ran away from home. She began smoking crystal methamphetamine the previous year and was court-ordered for drug evaluation after an arrest for prostitution.

A careful history of early school experiences revealed significant problems with distractibility, completing assignments, sitting still in class, and general irritability. On a DSM checklist assessing childhood symptoms, she endorsed six inattentive and eight hyperactive/impulsive symptoms. Currently, she endorsed five inattentive and six hyperactive/impulsive symptoms, general dysphoria, fatigue, anhedonia, and suicidal preoccupation.

The patient enrolled in a long-term, court-funded residential treatment program for drugs and alcohol. After several weeks her mood symptoms diminished, but she continued to suffer from marked inattention, distractibility, and restlessness. After careful consideration, her physician initiated nonstimulant medication treatment for ADHD. At discharge from her residential program, she resolved to stay in outpatient drug treatment and obtain part-time work. At the 1-month follow-up, she described a good response to treatment and had begun studying to obtain a general equivalency diploma (GED). Her physician continued to monitor her on a monthly basis with intermittent, unannounced drug testing to monitor ongoing sobriety.

FUTURE DIRECTIONS

Ongoing advances in neuroscience are apt to increase the acknowledgment of adult ADHD as a biologic syndrome. Changes in clinical practice increasingly reflect a shift from viewing adult ADHD as a unique and rare phenomenon to routine consideration of ADHD both alone and as potentially comorbid with other traditional adult psychopathologies. There will be ongoing interest in developing medication treatments with less perceived potential for abuse and longer durations of effects. Concurrently, there will be sustained efforts to develop effective nonmedication therapies. Although revisions made to ADHD criteria for DSM-5 and DSM-5-TR better reflect the developmental aspects of the behavioral phenotype, ongoing research will increasingly focus on the identification and modification of underlying dysregulation in brain neural networks. Such discoveries would raise the possibility of early identification of risks and treatments that target specific dysregulated systems in individual patients.

FURTHER READINGS

Alexander L, Farrelly N. Attending to adult ADHD: a review of the neurobiology behind adult ADHD. *Irish J Psychol Med*. 2018;35:237–244.

Asherson P, Buitlelaar J, Faraone SV, Rohde LA. Adult attention-deficit hyperactivity disorder: key conceptual issues. *Lancet Psychiatry*. 2016;3:568–578.

Biederman J, DiSalvo M, Woodworth KY, et al. Toward operationalizing deficient emotional self-regulation in newly referred adults with ADHD: a receiver operator characteristic curve analysis. *Eur Psychiatry*. 2020;63:e21.

Brikell I, Kuja-Halkola F, Larsson H. Heritability of attention-deficit hyperactivity disorder in adults. *Am J Med Genet B Neuropsychiatr Genet*. 2015;168(6):406–413.

Caye A, Swanson J, Thapar A, et al. Life span studies of ADHD – conceptual challenges and predictors of persistence and outcome. *Curr Psychiatry Rep*. 2016;18:111.

Cortese S, Coghill D. Twenty years of research on attention-deficit/hyperactivity disorder (ADHD): looking back, looking forward. *Evid Based Ment Health*. 2018;21:173–176.

Faraone SV, Larsson H. Genetics of attention deficit hyperactivity disorder. *Mol Psychiatry*. 2019;24:562–575.

Harrison AG, Armstrong IT. A comparison of the self-report patterns of analog versus real-world malingerers of attention deficit hyperactivity disorder. *J Neural Transm (Vienna).* 2021;128:1065–1077.

Harrison AG, Harrison KA, Armstrong IT. Discriminating malingered attention deficit hyperactivity disorder from genuine symptom reporting using novel personality assessment inventory validity measures. *Appl Neuropsychol Adult.* 2012;29(1):10–22.

Kooij SJ, Bejerot S, Blackwell A, et al. European consensus statement on diagnosis and treatment of adult ADHD: the European Network Adult ADHD. *BMC Psychiatry.* 2010;10:67.

Manos MJ. Psychosocial therapy in the treatment of adults with attention-deficit/hyperactivity disorder. *Postgrad Med*. 2013;125:51–64.

McGough J. Treatment controversies in adult ADHD. *Am J Psychiatry.* 2016;173:960–966.

McGough JJ. *ADHD*. Oxford University Press; 2014.

Mongia M, Hechtman L. Cognitive behavior therapy for adults with attention-deficit/hyperactivity disorder: a review of recent randomized controlled trials. *Curr Psychiatry Rep*. 2012;14:561–567.

Posner J, Polanczyk GV, Sonuga-Barke E. Attention-deficit hyperactivity disorder. *Lancet.* 2020;395:450–462.

Rabiner DL. Stimulant prescription cautions: addressing misuse, diversion and malingering. *Curr Psychiatry Rep*. 2013;15:375.

Reed GM, First MB, Kogan CS, et al. Innovations and changes in the ICD-11 classification of mental, behavioral and neurodevelopmental disorders. *World Psychiatry.* 2019;18:3–19.

Rovira P, Demontis D, Sánchez-Mora C, et al. Shared genetic background between children and adults with attention deficit/hyperactivity disorder. *Neuropsychopharmacology.* 2020;45:1617–1626.

Roy A, Hechtman L, Arnold LE, et al. Childhood factors affecting persistence and desistence of attention-deficit/hyperactivity disorder symptoms in adulthood: results from the MTA. *J Am Acad Child Adolesc Psychiatry*. 2016;55:937–944.

Scholz L, Werle J, Philipsen A, Schulze M, Collonges J, Gensichen J. Effects and feasibility of psychological interventions to reduce inattention symptoms in adults with ADHD. *J Ment Health*. 2023;32(1):307–320.

Sibley MG, Rohde LA, Swanson JM, et al. Late-onset ADHD reconsidered with comprehensive repeated assessments between ages 10 and 25. *Am J Psychiatry.* 2018;
175:140–149.

Ustan B, Adler LA, Rudin C, et al. The World Health Organization adult attention-deficit/hyperactivity disorder self-report screening scale for DSM-5. *JAMA Psychiatry*. 2017;74:520–526.

Wilens TE, Isenberg BM, Kaminski TA, Lyons RM, Quintero J. Attention-deficit/hyperactivity disorder and transitional aged youth. *Curr Psychiatry Rep*. 2018;20:100.

Wilens TE, Morrison NR, Prince J. An update on the pharmacotherapy of attention-deficit/hyperactivity disorder in adults. *Expert Rev Neurother*. 2011;11:1443–1465.

Zhang-James Y, Helminen EC, Liu J, Franke B, Hoogman M, Hoogman M; ENIGMA-ADHD Working Group. Evidence for similar brain structure abnormalities in youth and adult attention-deficit/hyperactivity disorder: a machine learning analysis. *Transl Psychiatry*. 2021;11:82.

45

Disruptive Behavior Disorders in Children and Adolescents

Erin P. Vaughan, M.A., Eva R. Kimonis, Ph.D., M.S., and Paul J. Frick, Ph.D.

Oppositional defiant disorder (ODD) and conduct disorder (CD) are included in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders, Text Revision (DSM-5-TR) in the chapter on disruptive, impulse control, and conduct disorders. The disorders included in this chapter all involve problems with the self-control of emotions or behavior. While many disorders included in the DSM-5-TR also involve problems with emotional and/or behavioral regulation, the disorders listed in this chapter are unique in that the problems in regulation bring the individual into significant conflict with societal norms (CD) or authority figures (ODD) or violate the rights of others (CD). Importantly, the DSM-5-TR and its global public health counterpart, the International Classification of Diseases 11th Revision (ICD-11), recognize the causal heterogeneity involved in the self-control problems of individuals who meet ODD and CD diagnostic criteria. Further, many of the symptoms that define ODD and CD occur to some degree in typically developing individuals, making it important to diagnose these disorders in the context of what is normative for a person's age, gender, and culture, while also considering the degree of impairment that is associated with the symptoms of the disorder.

Considered together, ODD and CD are among the most frequent reasons for children and adolescents to be referred to psychology and psychiatry clinics. This is because these disorders cause significant disruption to those around the child and are associated with a high degree of concurrent impairment. Children and adolescents with ODD or CD are at risk for a variety of other mental health problems (e.g., anxiety, depression, substance use), social problems (e.g., peer rejection, conflict with parents, siblings, and teachers), academic problems (e.g., school suspension and dropout), and legal problems (e.g., arrest for illegal behavior). ODD and CD also increase risk for psychiatric disorders, physical health problems (i.e., cardiovascular disease, serious injury, and sexually transmitted diseases), poor educational and occupational adjustment, and legal problems in adulthood. As a result, ODD and CD are associated with higher rates of public expenditure (e.g., mental health treatment, medical treatment, incarceration) than other psychiatric conditions that are common in children and adolescents.

OPPOSITIONAL DEFIANT DISORDER

Definition

The essential features of ODD can be grouped into three symptom clusters: angry/irritable mood, argumentative/defiant behavior, and vindictiveness. Angry and irritable mood involves frequent temper outbursts, especially when the person does not get their way, and a general pattern of being angry, resentful, and easily annoyed. Argumentative and defiant behaviors are expressed by persistent stubbornness and unwillingness to compromise, give in, or negotiate with adults or peers. Defiance may also include deliberate or persistent testing of limits, usually by ignoring requests or commands, arguing, failing to accept blame, and blaming others for misdeeds. Individuals with ODD often do not regard themselves as oppositional or defiant but justify their behavior as a response to unreasonable demands or circumstances. Symptoms generally begin early in life at home in the child’s interactions with the parents, other primary caregivers, and siblings. As the disorder progresses and increases in severity, angry/irritable and argumentative/defiant symptoms generalize outside of the home toward teachers or other authority figures at school or in the community. While significant symptoms of ODD at home are predictive of current and future impairment, displaying symptoms outside the home is an important marker for the severity of the disorder.

Given that ODD symptoms can be present to some degree in typically developing individuals, they are considered indicative of a disorder when an individual exhibits four of eight possible symptoms over the past 6 months. Behaviors also must be present outside of interactions with siblings. While the required frequency of behaviors is not explicitly stated in symptom definitions (with the exception of vindictiveness, which must occur at least twice in the past 6 months), tentative guidelines are provided to distinguish symptomatic behavior from behavior within normal limits. For younger children (below age 5 years), behaviors should occur most days of the week in order to exceed normal behavior for young children. For individuals over the age of 5 years, behaviors should occur at least once a week. However, there is a degree of flexibility in these guidelines, in order to allow for clinical judgment on what is or is not within the expected range of behaviors for a particular individual based on their developmental level, gender, and culture. Importantly, the behaviors must also be associated with significant distress or impairment within the individual, their social context, or other important areas of functioning.

Disruptive mood dysregulation disorder (DMDD) shares several symptoms in common with ODD, including angry or irritable mood and temper outbursts. In fact, most individuals who meet diagnostic criteria for DMDD also meet criteria for ODD. However, the main differences between ODD and DMDD are in the frequency and intensity of symptoms and the pervasiveness of a negative mood between temper outbursts. That is, individuals with DMDD display more intense mood outbursts and do not have periods of normal mood between angry outbursts (i.e., anger or irritability is present most of the day every day). In individuals who meet diagnostic criteria for both ODD and DMDD, only a diagnosis of DMDD is given.

While ODD includes problems with emotional regulation (e.g., irritability, anger) that increase risk for later emotional disorders, these symptoms are not included in the diagnostic criteria for CD. As a result, when criteria are met for both ODD and CD, both diagnoses can be given. However, symptoms of ODD and CD are highly correlated and are often considered to represent a single dimension of disruptive or externalizing behaviors. In fact, most children who develop CD prior to adolescence begin by showing ODD symptoms, whereas most children with ODD do not go on to develop CD.

History

ODD first emerged as a diagnosis distinct from CD in DSM-III. Previously, the Group for the Advancement of Psychiatry had suggested the concept of childhood oppositional personality in 1966, emphasizing the symptoms of disobedience and pervasive negativism. Early conceptualizations of ODD emphasized the passive-aggressive nature of the oppositional symptoms. With the introduction of DSM-III in 1980, oppositional disorder was included with explicit behavioral criteria. Subsequent editions of the DSM added a defiant label and expanded the ODD symptom list from two of five possible items (DSM-III) to five of nine possible symptoms (DSM-III-R) and finally to four of eight total symptoms (DSM-IV, DSM-IV-TR, DSM-5, and DSM-5-TR). Although DSM-5 and DSM-5-TR include the same eight ODD symptoms as in DSM-IV-TR, they are newly grouped into three categories (i.e., angry/irritable mood, argumentative/defiant behavior, and vindictiveness) on the basis of factor analytic and longitudinal research showing that these symptoms show distinct patterns of covariation and the different symptom clusters uniquely predict different comorbidities and future problems. Specifically, the angry/irritable dimension is associated with greater emotional disorders, the argumentative/defiant dimension is related to attention-deficit/hyperactivity disorder (ADHD), the vindictive symptom is related to callous-unemotional (CU) traits, and all three dimensions are highly related to CD. Thus, DSM-5-TR conceptualizes ODD as a disorder with both emotional and behavioral dimensions. DSM-5 attempted to provide a more lifespan view of all disorders and, in doing so, eliminated the chapter on disorders first evident in infancy, childhood, or adolescence in which ODD was included in previous editions of the DSM, relocating this disorder to the disruptive, impulse control, and conduct disorders chapter.

Despite ODD being included in recent editions of the DSM, spanning many decades, the disorder has been somewhat controversial for two reasons. First, the oppositional and argumentative behaviors that form the criteria for ODD are commonly displayed in normally developing children, especially at certain developmental periods, such as in the toddler and early preschool years (i.e., the terrible twos) and in adolescence. Second, ODD is frequently comorbid with a host of other disorders and, in fact, is rarely displayed without symptoms of other disorders. As a result of these two issues, there were concerns that ODD over-pathologizes a normative behavior pattern and that, unless accompanied by another disorder (e.g., ADHD, CD), the symptoms of ODD are transient and benign and should not be considered as a separate disorder. ODD was ultimately included in the DSM-5, however, based on research showing that its symptoms did predict impairment, even when controlling for comorbid conditions, and clearer guidance was provided on how to differentiate ODD from more normative displays of these behaviors.

Comparative Nosology

The DSM-5-TR and ICD-11 include similar descriptions of ODD. While the 10th edition of the ICD included ODD as a subtype of CD, ODD is classified as a separate disorder in the ICD-11. Similar to DSM-5-TR, ICD-11 defines ODD as a persistent pattern of defiant, disobedient, provocative, or spiteful behavior that is outside the normal range expected based on an individual's age and developmental level. The behaviors must be present outside of interactions with siblings and must cause significant impairment in functioning. The primary difference between the DSM-5-TR and ICD-11 criteria for ODD is their use of specifiers. In ICD-11, ODD is classified with specifiers to designate the presence or absence of chronic irritability-anger and the presence or absence of limited prosocial emotions (LPE). The presence of chronic irritability-anger requires a pervasive angry or irritable mood in addition to frequent temper outbursts, leading the diagnosis of ODD with chronic irritability-anger to strongly resemble DSM-5-TR's diagnosis of DMDD, which is not otherwise represented in ICD-11. The presence of LPE is defined by elevated levels of callous-unemotional (CU) traits. The symptoms for this LPE specifier include a lack of empathy, lack of remorse or guilt, lack of concern over poor performance, limited expression of emotions, and relative indifference to punishment. Although the LPE specifier applies to both ODD and CD diagnoses in ICD-11, it is only used for a diagnosis of CD in DSM-5-TR, excluding the fifth symptom related to indifference to punishment.

Epidemiology

DSM-5-TR cites an average overall prevalence estimate of 3.3% for childhood ODD, which is based on a review of probability estimates from school and community samples worldwide. Interestingly, this overall rate did not differ across countries that varied in their racial and ethnic composition. These prevalence rates were based on studies with children and adolescents, and there have been few studies estimating the prevalence of ODD in adults, although it appears much rarer in this age group. ODD is more prevalent in males than females (1.6:1) in middle childhood, but this gender ratio is narrower in preschool children and adolescents. ODD is also more prevalent among children and adolescents from families of low socioeconomic status.

Etiology

One difficulty in interpreting the research on the causes of ODD is that, while the distinction between ODD and CD is made in clinical settings, most research combines symptoms of both disorders when studying risk factors and causal mechanisms. Further, most children who develop CD prior to adolescence show ODD earlier in development and, over time, start displaying the more severe CD symptoms in addition to ODD symptoms. Thus, the causal models developed to explain the development of childhood-onset CD, which are reviewed in the following section, are also relevant for explaining the onset of ODD. A critical piece of information that is currently unknown is what factors lead some children to develop ODD but not to go on to develop CD, and what factors lead other children to progress on to develop more severe CD symptoms.

What is critical to all theories of ODD is that it is a transactional disorder involving both dispositional characteristics of the child and problems in the child's psychosocial context. That is, ODD is related to factors within the child that make them difficult to socialize. As a result, the child evokes negative responses from others in their environment, in particular caregivers and authority figures, which further increases the child's defiant and negative behavior. Consequently, most causal theories for ODD involve gene-by-environment interactions and dynamic, interactive models involving reciprocal interactions between the child and their caregivers over time.

The degree to which ODD symptoms present within typically developing individuals differs across developmental stages, and several ODD behaviors are highly prevalent in 2- to 4-year-olds and adolescents. However, in addition to considering the child's developmental stage when diagnosing ODD, it is also important to consider the developmental processes that contribute to severe and impairing levels of ODD behaviors. For example, the defiance that is somewhat normative during the third and fourth years of life, as a child develops some degree of independence from caregivers, can lead to impairing behavior if it is met by harsh and inconsistent responses from parents. Similarly, the questioning of authority that often accompanies identity development in adolescence can lead to severe and impairing behaviors, if supported by a deviant peer group or if the adolescent is not appropriately monitored and supervised by parents or other authority figures.

Biologic Factors. Twin studies estimating the genetic and environmental contributions to oppositional defiant behavior find support for additive genetic, unique environmental, and shared environmental influences for both boys and girls. However, these studies generally find support for a genetic vulnerability to oppositional and defiant behaviors that interacts with the child's environment across development to result in the expression of ODD. This research supports the clinical observation that oppositional and disruptive behaviors frequently cluster in siblings and biologic relatives of a child with ODD.

It is likely that genetic predispositions lead to ODD through the child's temperamental characteristics. Temperament is defined as the child's characteristic style of emotional and behavioral responses in a variety of situations and to a variety of environmental stimuli. Specific temperamental styles can be observed early in life and show a modest persistence over time, often developing into personality traits later in life. A cluster of temperament dimensions described as the "difficult temperament" has been shown to be a risk factor for ODD in children across a number of racial, ethnic, and socioeconomic groups. Although many definitions exist, the difficult temperament is generally characterized by emotional hyper-reactivity (especially through frequent and high-intensity responses of negative emotion), poor behavioral control (e.g., overactivity and difficulty sustaining attention), and low adaptability to new situations. This cluster of temperamental characteristics has been shown to predict a wide range of emotional and behavioral problems, and not only ODD. Whereas a difficult temperament alone is a relatively weak predictor of ODD, it becomes increasingly predictive when combined with various psychosocial stressors.

Psychological Factors. One psychosocial factor that is consistently linked with ODD is harsh, coercive, and inconsistent parenting. However, as noted above, this risk factor is usually considered in causal models along with the child's difficult temperament. An influential theory for the development of ODD that integrates both parenting and the child's temperament is Gerald Patterson’s coercion theory. Coercion theory applies learning theory, specifically negative reinforcement, to explain how oppositional child behavior develops. A problematic behavior is negatively reinforced when it allows an individual to escape from a noxious, aversive, or unpleasant situation or task. The negative behavior of one member of a parent–child dyad serves to terminate the ongoing negative behavior of the other, thereby negatively reinforcing the first member’s original negative behavior. In the case of the child with a difficult temperament who expresses high-intensity negative emotion in response to parental demands (e.g., temper tantrums), this often leads the parent to use increasingly harsh and coercive discipline to change the child’s behavior. The child’s improved behavior negatively reinforces the parent’s use of harsh discipline techniques. However, over time, the child learns that if they increases the negative behavior (e.g., longer and more intense temper tantrums), the parent will back down on their discipline attempts. This process reinforces the child’s negative behavior. Thus, the child and parent become entangled in a coercive cycle in which they each use increasingly negative behavior to change the other’s behavior. Over time, the child learns that aversive behavior “works” to solve social problems and may begin to generalize their angry and oppositional interpersonal behavior outside of the home to interactions with teachers and other authority figures at school and in the community, as well as to peers.

Sociologic Factors. Ecologic variables, including parental depression and other mental health problems, family discord, family stress, and poverty, are also associated with increased risk of developing ODD. However, these factors appear to have their influence on the development of ODD through their effects on parental discipline. For example, parents living in impoverished living conditions experience a large number of stressors and daily hassles that make it difficult to engage in positive parenting and to set and enforce consistent limits on their children. Similarly, studies of the relation between maternal depression and ODD find that depression in the mother reduces maternal–child interaction quality, making the parent less involved and responsive to their child’s needs. Maternal depression can also lead to increases in negative parent–child interactions, including irritability or hostility between the mother and child.

Diagnosis and Clinical Features

Diagnostic criteria for ODD are listed in Table 45–1. The diagnosis is not limited to a particular age group, but most commonly emerges between the ages of 4 and 8 years old. While studies of ODD in adults are rare, there is evidence that adults can display this disorder as well. Impairment associated with ODD largely depends on the number of settings in which behaviors are present. Problem behaviors are usually first evident at home in relationships with those most familiar to the child. In more severe cases, ODD symptoms may also occur outside of the home, such as in daycare or school settings, usually in situations where authority figures set limits or place demands on the child. Thus, it is particularly important to investigate behavior in school, with friends, and in other settings to determine the degree of ODD severity. In addition, the child with ODD often does not see their behaviors as a problem and commonly places blame on others for any problems. Thus, the clinical evaluation should include multiple reporters including both parents (where applicable) and teachers, in addition to the child. Finally, given that the behavior associated with ODD can be displayed to some degree in typically developing children and adolescents, it is important that the diagnostic process involves some method for comparing the child’s level of behavior to what is normative for the child’s age, gender, and culture.

Jonah is a 7-year-old boy whose mother reports that he has been difficult to control at home since he was a toddler, with his behavior worsening when his younger sister was born. In addition to impulsive and hyperactive behavior, Jonah is often defiant and argumentative with his parents. Requests to clean his room, get ready for school or for bed, or share his toys with his younger sister are met with angry refusals that escalate into temper tantrums. His mother reports that her attempts to discipline him are unsuccessful, and she displays her frustration with his defiance by shouting and making threats that are not enforced. When Jonah began preschool, these negative behaviors began to be present with his teachers and peers as well. According to his second-grade teacher, he does not follow directions in class, blames other students for his mistakes, argues with adults, and loses his temper on a weekly basis. These behaviors have resulted in several discipline notes being sent home, and he has been kicked out of two afterschool programs due to his difficulty following rules and repeated conflicts with teachers and classmates. In addition to impairment at both home and school, reports from Jonah's parents and teacher on standardized rating scales reveal that his behavior is more severe than would be expected for a boy his age (over the 95th percentile compared to age norms). His performance on cognitive and academic testing indicates abilities within the normative range, and his mother reports an unremarkable developmental and medical history.

Table 45–1.

DSM-5-TR Diagnostic Criteria for Oppositional Defiant Disorder

A. A pattern of angry/irritable mood, argumentative/defiant behavior, or vindictiveness lasting at least 6 months as evidenced by at least four symptoms from any of the following categories, and exhibited during interaction with at least one individual who is not a sibling.

Angry/Irritable Mood

1. Often loses temper.
2. Is often touchy or easily annoyed.
3. Is often angry and resentful.

Argumentative/Defiant Behavior

4. Often argues with authority figures or, for children and adolescents, with adults.
5. Often actively defies or refuses to comply with requests from authority figures or with rules.
6. Often deliberately annoys others.
7. Often blames others for his or her mistakes or misbehavior.

Vindictiveness

8. Has been spiteful or vindictive at least twice within the past 6 months.

Note: The persistence and frequency of these behaviors should be used to distinguish a behavior that is within normal limits from a behavior that is symptomatic. For children younger than 5 years, the behavior should occur on most days for a period of at least 6 months unless otherwise noted (Criterion A8). For individuals 5 years or older, the behavior should occur at least once per week for at least 6 months unless otherwise noted (Criterion A8). While these frequency criteria provide guidance on a minimum level of frequency to define symptoms, other factors should also be considered, such as whether the frequency and intensity of the behaviors are outside a range that is normative for the individual's developmental level, gender, and culture.

B. The disturbance in behavior is associated with distress in the individual or others in his or her immediate social context (e.g., family, peer group, work colleagues), or it impacts negatively on social, educational, occupational, or other important areas of functioning.

C. The behaviors do not occur exclusively during the course of a psychotic, substance use, depressive, or bipolar disorder. Also, the criteria are not met for disruptive mood dysregulation disorder.

Specify current severity:

Mild: Symptoms are confined to only one setting (e.g., at home, at school, at work, with peers).

Moderate: Some symptoms are present in at least two settings.

Severe: Some symptoms are present in three or more settings.

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 522, 523 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Pathology and Laboratory Examination

No laboratory tests are available for the diagnosis of ODD, and assessments are based largely on clinical examination and history. Several validated rating scales are available to assess oppositional and defiant behaviors. These scales document symptom severity relative to same-age and gender peers. Rating scales may assist in treatment planning by identifying specific targets for interventions and by documenting baseline disorder severity and subsequent response to treatment. Rating scales completed by both parents and teachers can aid in determining if the child's behavior is displayed in multiple situations and settings, and with multiple people. However, the cross-situational pervasiveness of behaviors can also be assessed through the clinical interview. Rating scales to measure the degree of functional impairment caused by symptoms are also available, although this too is often assessable through clinical interviews. Finally, observing parent–child interactions in young children can not only aid in determining the severity of a child's behavior but can also document specific parent–child interactional patterns and parenting practices that serve to maintain the ODD behaviors and could be targeted in treatment.

Differential Diagnosis

Children with ODD often have comorbid diagnoses of ADHD and CD. The behaviors associated with ADHD, especially the problems with impulse control, often lead children to have problems regulating their emotions and behaviors and make them difficult to discipline consistently. The diagnosis of ODD is added to the diagnosis of ADHD when the child shows active defiance and noncompliance that cannot be solely attributed to impulsive acts. Over time, the behaviors associated with ODD can develop into the more serious behavior problems associated with CD (e.g., aggression toward people or animals, recurrent destruction of property, covert theft, or deceit). Maintaining the diagnosis of ODD alongside a CD diagnosis may be required to capture the emotional symptoms (e.g., angry/irritable mood) that can be part of ODD but not CD. Defiant and disruptive behaviors may occur after a major stressor such as a traumatic event or the loss of a loved one. If the behavioral disturbance lasts less than 6 months, an adjustment disorder with disturbance in behavior is the more appropriate diagnosis. In individuals with intellectual disability, a diagnosis of ODD is only given if the oppositional behaviors are markedly greater than is commonly observed among individuals of a comparable cognitive level. Irritable, defiant, and oppositional behaviors can be associated with mood and anxiety. However, in most cases of children who meet criteria for ODD, the emotional distress is caused by impairments associated with the child's behavior problems (e.g., disciplinary problems at school, conflict with parents, peers, and teachers). Bipolar disorder presenting in those younger than 18 years can include defiance of authority, oppositionality, and irritability, but the behavior problems are confined to the mood episodes. Also, oppositional behaviors may be a feature of psychotic illness in children and adolescents, but ODD is not diagnosed separately if the ODD symptoms occur only during the course of a psychotic or mood disorder. Refusal to follow the directions of authority figures needs to be distinguished from disorders of hearing and language. Children with hearing loss or receptive language disorders may look oppositional on the basis of a failure to hear or understand spoken commands.

Although DSM-5-TR classifies DMDD as a depressive disorder, DMDD shares several symptoms in common with ODD, including an angry or irritable mood and temper outbursts. The key difference between DMDD and ODD is that, while the angry/irritable mood and temper outburst symptoms of ODD are only required to take place “often” (typically once a week), DMDD requires an angry or irritable mood between temper outbursts that are present most of the day, nearly every day. Temper outbursts in DMDD must also occur three or more times a week on average. Finally, DMDD symptoms must be present for at least 12 months (rather than 6 months for ODD) and in two settings (while ODD symptoms can be present in only one setting for a mild severity). Diagnosis of DMDD must also be made between ages 6 and 18, and symptoms must have emerged before age 10, while no such age requirements are needed for ODD. Despite these differences, research suggests that most individuals with DMDD would also meet the criteria for ODD. As such, DMDD can be conceptualized as a severe variant of ODD, with the primary problems being with emotional control rather than behavioral control. Thus, if criteria for both DMDD and ODD are met in an individual, DMDD is diagnosed as the more severe condition, and ODD is not diagnosed.

Comorbid Conditions

ODD often co-occurs with other psychiatric disorders and, in fact, it is rare for a child to show ODD without other comorbid problems. The most common comorbid diagnosis with ODD is ADHD. Studies have shown that 60% to 80% of children with ODD assessed in clinics may also meet criteria for ADHD, and about 50% of children with ADHD will also meet diagnostic criteria for ODD. Approximately 15% to 20% of clinically referred children and adolescents with ODD have comorbid depression, bipolar disorder, anxiety disorders, or language disorders. Concurrent substance abuse is also more common in individuals with ODD, especially teenagers with comorbid ADHD.

Course and Prognosis

The combination of ADHD and ODD confers risk for more problematic outcomes, including progression to CD. Children with an earlier age of onset of ODD (i.e., preschool or early elementary school) are also three times more likely than those with a later age of onset (i.e., later elementary school years or beginning in adolescence) to develop CD. Approximately 30% of children with ODD will eventually develop CD. Although the diagnosis of ODD is relatively stable over time, about two-thirds of children diagnosed with ODD will no longer meet diagnostic criteria within the following 3 years. Just as ODD confers risk for the eventual development of CD, CD confers risk for the development of antisocial personality disorder (ASPD) in adulthood. Longitudinal studies have shown that about 10% of children with ODD will meet diagnostic criteria for ASPD as adults. These are usually the children with ODD that progress to CD. Children with the angry/irritable mood symptoms of ODD are also at risk for developing anxiety and mood disorders later in life. The development of mood disorders in children with ODD often results from the many impairments associated with their behavior problems.

Treatment

Research suggests that treatment of ODD is critical, not only for reducing the impairments associated with the disorder, but also to prevent the later development of CD, ASPD, and anxiety and depressive disorders. Treatment for ODD tends to be most effective when it is implemented early in childhood (i.e., prior to age 9 years). The most successful of these interventions is parent management training (PMT), also called behavioral parent training, which is based on social learning principles. There are several variants of PMT, such as Parent Management Training Oregon Model (PMTO), Parent–Child Interaction Therapy (PCIT), Helping The Noncompliant Child (HNC), Incredible Years, And Triple P, that have been successfully utilized in clinic, home, school, and online delivery settings and implemented with families in individual and group formats. There is recent evidence to support the effectiveness of various forms of self-administered PMT as well, including interventions provided via books, video, internet, and smartphone applications.

Although many different types of PMT exist, the underlying rationale for all programs is based on the general view that child oppositional problem behaviors are inadvertently learned, developed, and sustained in the home by chronic and maladaptive parent–child interactions. The goals of PMT programs are to teach parents how to develop and implement structured contingency management programs and to use more effective discipline strategies (e.g., increase consistency with and use of a variety of approaches). Roughly half of PMT programs examined in a meta-analysis were found to also integrate a positive parent–child relationship-building component (e.g., to increase parental warmth, sensitivity, and responsiveness). Several of these programs are implemented in two phases. The first phase focuses on improving the quality of the parent–child relationship by teaching parents how to interact with their child in a positive and responsive manner, how to use positive reinforcement to encourage compliant and prosocial behavior, and how to reduce coercive parent–child interactions using selective attention. The second phase focuses on teaching parents how to give clear and effective instructions to their children and to use consistent and noncorporal punishments, such as time-out, to decrease noncompliant, attention-seeking, and other negative behavior. In many PMT programs, therapists model the skills for the parent during treatment sessions, in addition to providing didactic instruction and discussion. Some PMT programs involve real-time coaching of the parent during interactions with the child to practice the skills taught. Most PMT programs also rely on regular home practice assignments and exercises to support learning of parenting skills.

When the child with ODD has comorbid ADHD, which, as noted previously, is quite common, ODD symptoms often respond to psychopharmacological intervention with stimulant medications. However, there is little scientific evidence to suggest that ODD responds directly to medications when it is not accompanied by ADHD.

CONDUCT DISORDER

Definition

The essential characteristic of CD is a repetitive and persistent pattern of behavior in which the basic rights of others or major age-appropriate societal norms or rules are repeatedly violated beginning in childhood or adolescence. These behaviors fall into four main groupings: aggression toward people or animals, destruction of property, deceitfulness or theft, and serious rule violations. To meet criteria for a diagnosis of CD, at least 3 of 15 criterion behaviors must be present during the past 12 months, with at least one behavior present in the past 6 months (see Table 45–2). The disturbance in behavior must also cause clinically significant impairment in social, academic, or occupational functioning. However, the level of severity of CD can vary greatly, from relatively mild lying and stealing to severe aggression involving serious harm to others involving a weapon or forced sexual activity. Severity is classified as mild, moderate, or severe based on the number of symptoms and the degree of harm to others caused by the symptoms. Except for requiring that the criteria for adult ASPD are not present if the individual is aged 18 years or more, there are no exclusionary criteria. Thus, individuals with any other diagnosable disorder, such as post-traumatic stress disorder (PTSD), depression, bipolar disorder, or schizophrenia can be additionally diagnosed with CD, if their behavior over the past year meets criteria.

Table 45–2.

DSM-5-TR Diagnostic Criteria for Conduct Disorder

A. A repetitive and persistent pattern of behavior in which the basic rights of others or major age-appropriate societal norms or rules are violated, as manifested by the presence of at least three of the following 15 criteria in the past 12 months from any of the categories below, with at least one criterion present in the past 6 months.

Aggression to People and Animals

1. Often bullies, threatens, or intimidates others.
2. Often initiates physical fights.
3. Has used a weapon that can cause serious physical harm to others (e.g., a bat, brick, broken bottle, knife, gun).
4. Has been physically cruel to people.
5. Has been physically cruel to animals.
6. Has stolen while confronting a victim (e.g., mugging, purse snatching, extortion, armed robbery).
7. Has forced someone into sexual activity.

Destruction of Property

8. Has deliberately engaged in fire setting with the intention of causing serious damage.
9. Has deliberately destroyed others' property (other than by fire setting).

Deceitfulness or Theft

10. Has broken into someone else's house, building, or car.
11. Often lies to obtain goods or favors or to avoid obligations (i.e., "cons" others).
12. Has stolen items of nontrivial value without confronting a victim (e.g., shoplifting, but without breaking and entering; forgery).

Serious Violations of Rules

13. Often stays out at night despite parental prohibitions, beginning before age 13 years.
14. Has run away from home overnight at least twice while living in the parental or parental surrogate home, or once without returning for a lengthy period.
15. Is often truant from school, beginning before age 13 years.

B. The disturbance in behavior causes clinically significant impairment in social, academic, or occupational functioning.

C. If the individual is age 18 years or older, criteria are not met for antisocial personality disorder.

Specify whether:

Childhood-onset type: Individuals show at least one symptom characteristic of conduct disorder prior to age 10 years.

Adolescent-onset type: Individuals show no symptom characteristic of conduct disorder prior to age 10 years.

Unspecified-onset type: Criteria for a diagnosis of conduct disorder are met, but there is not enough information available to determine whether the onset of the first symptom was before or after age 10 years.

Specify if:

With limited prosocial emotions: To qualify for this specifier, an individual must have displayed at least two of the following characteristics persistently over at least 12 months and in multiple relationships and settings. These characteristics reflect the individual’s typical pattern of interpersonal and emotional functioning over this period and not just occasional occurrences in some situations. Thus, to assess the criteria for the specifier, multiple information sources are necessary. In addition to the individual’s self-report, it is necessary to consider reports by others who have known the individual for extended periods of time (e.g., parents, teachers, coworkers, extended family members, peers).

Lack of remorse or guilt: Does not feel bad or guilty when he or she does something wrong (exclude remorse when expressed only when caught and/or facing punishment). The individual shows a general lack of concern about the negative consequences of his or her actions. For example, the individual is not remorseful after hurting someone or does not care about the consequences of breaking rules.

Callous-lack of empathy: Disregards and is unconcerned about the feelings of others. The individual is described as cold and uncaring. The individual appears more concerned about the effects of his or her actions on himself or herself, rather than their effects on others, even when they result in substantial harm to others.

Unconcerned about performance: Does not show concern about poor/problematic performance at school, at work, or in other important activities. The individual does not put forth the effort necessary to perform well, even when expectations are clear, and typically blames others for his or her poor performance.

Shallow or deficient affect: Does not express feelings or show emotions to others, except in ways that seem shallow, insincere, or superficial (e.g., actions contradict the emotion displayed; can turn emotions "on" or "off" quickly) or when emotional expressions are used for gain (e.g., emotions displayed to manipulate or intimidate others).

Specify current severity:

Mild: Few if any conduct problems in excess of those required to make the diagnosis are present, and conduct problems cause relatively minor harm to others (e.g., lying, truancy, staying out after dark without permission, other rule breaking).

Moderate: The number of conduct problems and the effect on others are intermediate between those specified in "mild" and those in "severe" (e.g., stealing without confronting a victim, vandalism).

Severe: Many conduct problems in excess of those required to make the diagnosis are present, or conduct problems cause considerable harm to others (e.g., forced sex, physical cruelty, use of a weapon, stealing while confronting a victim, breaking and entering).

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 530–532 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Subtypes and Specifiers of Conduct Disorder

As with all disorders in the category of disruptive, impulse control, and conduct disorders, the behaviors associated with CD can be the result of a number of different factors. In an attempt to acknowledge the heterogeneity inherent in the diagnostic criteria for CD, a number of subtypes are recognized in addition to the severity criteria. DSM-5-TR indicates subtypes for CD based on the age of onset of the symptoms and the presence of LPE. Childhood-onset type refers to the onset of at least one CD symptom prior to the age of 10 years. This distinction is based on a long history of research in both mental health and criminology, showing that the age of onset of antisocial behavior has important clinical, etiologic, and prognostic implications. Specifically, research finds that a relatively small group of children (generally boys) with the onset of antisocial behaviors in childhood (as contrasted with onset in adolescence) are more likely to be aggressive and to continue to display problem behaviors, including criminal behavior, into adulthood. Also, compared to the later-onset group, individuals with childhood-onset CD display more neuropsychological deficits, including comorbid sustained attentional deficits, failure to inhibit impulsive responding, chronic learning failure, and academic underachievement.

The second subtype of CD recognized in the DSM-5-TR, the adolescent-onset type, defines individuals for whom all symptoms begin after 10 years of age. The adolescent-onset group typically involves more girls than the childhood-onset subtype and is likely to show aggressive and violent behavior. The adolescent-onset subtype is less likely to show aggressive and violent behavior into adulthood but, importantly, are not completely free of risk for problems as adults. Their antisocial behavior in adolescence can result in continuing problems (e.g., school dropout, substance abuse, and criminal behavior) which can influence their adult adjustment.

It is important to note that while the importance of distinguishing between early and later onset CD has strong research support, the exact age at which to define the subgroups has been debated. Research has used ages ranging from 9 to 14 to make this distinction. Some have also argued that the distinction should be based on pubertal status or made continuously (i.e., the earlier the age of onset, the more severe the CD). The DSM-5-TR also recognizes that pinpointing the age of onset in a person through historical reports can be difficult and, as a result, CD unspecified onset is used when criteria for a diagnosis of CD are satisfied, but the age of onset of the first CD symptoms cannot be reliably determined.

While the distinction between childhood-onset and adolescent-onset CD is included in most recent editions of the DSM, the DSM-5 newly added another specifier to recognize that there can be clinically and etiologically important distinctions within the onset subtypes, particularly within the more severe childhood-onset subtype (DSM-5-TR continued this with identical criteria). Thus, a CD specifier called “with limited prosocial emotions” (LPE), which is defined by the presence of significant and pervasive levels of CU traits, was added to the DSM-5. The LPE specifier is defined by the presence of two of the following four symptoms:

- a lack of remorse or guilt,
- callousness/lack of empathy,
- lack of concern about performance in importance activities, and
- shallow or deficient affect.

At least two of these characteristics must be present in multiple relationships and settings for at least 12 months. While this LPE specifier is new to the DSM-5, many of these symptoms were included in the definition of the undersocialized subtype of CD in the DSM-III. These traits are considered the affective components of conscience in developmental research, and the affective components of psychopathy in research with antisocial adults. The basis for adding the LPE specifier to the DSM-5 came from research showing that CU traits predict more aggression and poorer outcomes, even when controlling for the number of CD symptoms and age of onset. Individuals with CD and co-occurring CU traits typically show a childhood onset to their CD symptoms, which are more severe and aggressive, lead to more harm to their victims, and involve aggression that is more likely to be premeditated (i.e., planned) and instrumental (i.e., for gain) than other youth with early-onset CD. In contrast, children with childhood-onset CD who do not show elevated CU traits tend to show aggression that is confined to reactive acts occurring in context of high emotional arousal and in response to real or perceived provocations from others. In addition, although youth with CD and elevated CU traits do respond positively to many evidence-based treatments for CD, they often begin and end treatment with more severe behavior problems.

In addition to these clinically important differences, there are a number of other differences between children with CD with and without elevated CU traits that have implications for etiologic theories. For example, the conduct problems of children with elevated CU traits tend to show stronger influences from heredity. These children also show a number of deficits in how they recognize and react to emotional stimuli, such as showing poor recognition of fearful face stimuli and less emotional reactivity to signs of fear, sadness, and distress in others. In addition, children with CD and elevated CU traits tend to show reduced responses to cues of punishment and lower fear and anxiety relative to other children with CD. In contrast, individuals with childhood-onset CD without elevated CU traits often show elevated rates of emotional reactivity, especially to others' distress cues and provocation from others, as well as other signs of difficulty in regulating their emotions and behaviors (e.g., high rates of anger and anxiety). Thus, these two subgroups of children with CD show very different emotional and cognitive correlates that are important to consider in etiologic theories and could also be important for tailoring treatment to the individual needs of individuals with CD.

History

Descriptions of youths with conduct, delinquent, and antisocial symptoms antedate the very existence of the field of child and adolescent psychiatry. As early as 1837, Prichard used the term "moral insanity" to describe young adult patients with recurrent patterns of antisocial behavior who were not overtly psychotic. Benjamin Rush, the father of American psychiatry, described cases exhibiting "innate, preternatural moral depravity" and speculated that these patients had a defect in the "moral faculties of the mind." In the 19th century, scientific thought emphasized the inherent traits and physical defects (i.e., phrenology) that appeared to mark the antisocial individual.

Beginning in the early 20th century, a shift in thinking occurred as clinicians, social workers, pediatricians, welfare workers, and academics began to consider the impact of the environment and adverse individual or social conditions on adolescents who commit crimes. Concern over keeping the youthful offender out of jail, combined with hope that advances in the new psychoanalytic and psychodynamic therapies being introduced into America in the first two decades of the new century might undo the psychological damage that early childhood adverse experiences might cause, lead to new optimism in the treatment of young people with antisocial behaviors. This shift in thinking combined with hope for new treatments led to the beginning of the child guidance movement in the United States with the opening of the Juvenile Psychopathic Institute in Chicago in 1909, and the Judge Baker Clinic in Boston in 1917. The initial purpose of these clinics was to advise the courts on the psychology of juveniles under arrest. In 1923, Karl Menninger brought together a group of psychiatrists concerned with youthful offenders and started the American Orthopsychiatric Association. Thus, psychiatry's concerns for youths with conduct problems and antisocial behaviors have been intimately intertwined with shaping psychiatry’s own development as a medical discipline over the past 100 years.

CD was first recognized as a mental health disorder in the DSM-II published in 1968. Continuing a focus on adverse experiences, this early definition of CD in the DSM-II was conceptualized as being a reaction to negative environmental factors, and three distinct reactions were described: runaway reaction, undersocialized aggressive reaction, and group delinquent reaction. Thus, from its very first definition in classification systems for mental disorders, there has been recognition that there are multiple distinct subgroups of individuals with CD who can be differentiated by the behaviors they display and the causes leading to these behaviors.

Over the years, the basic definition of CD as behavior in which the rights of others or major age-appropriate norms are violated has not changed much, but the methods for subtyping individuals within this category have been the major source of change in the diagnostic criteria over time. The DSM-III introduced a different method for subtyping individuals with CD according to whether or not they were (a) aggressive and (b) considered “socialized” (e.g., had lasting friendships, felt guilt/remorse). Relative to the socialized subtype, individuals with the undersocialized subtype of CD showed poorer outcomes in adulthood, and several unique neurophysiologic correlates to their CD, such as low serotonin levels and autonomic irregularities. While the undersocialized subgroup was defined by some characteristics shared in common with the DSM-5-TR’s LPE specifier (e.g., does not feel guilt or remorse), the undersocialized label created considerable confusion as to its core features and most suitable operationalization. Some definitions focused on the child’s ability to form and maintain social relationships, whereas others focused primarily on the context (alone or with a peer group) in which antisocial acts were typically committed. Few definitions focused directly on the LPE themselves.

In an attempt to address this definitional confusion, the next revision of the DSM (i.e., the DSM-III-R) revised how CD was subtyped. The criteria for the undersocialized subtype were changed to focus solely on whether the antisocial acts were committed alone and whether the pattern included aggressive symptoms, both of which were judged to be more easily assessed than the CU features included in the original undersocialized definition. This change was accompanied by relabeling the specifier as the solitary-aggressive subtype. The criteria for the second subtype focused solely on whether the antisocial acts were committed with other antisocial peers, and this subtype was assumed to be primarily nonaggressive in nature, leading to its renaming as the group subtype.

In the DSM-IV, these subtypes were replaced by the childhood-onset and adolescent-onset distinction that was retained in the DSM-5. The focus on onset was intended to capture the distinction between early and later onset of antisocial behavior that had been long studied in research on criminal behavior and proven to be important for predicting different outcomes. Also, given that the childhood-onset subtype was more likely to be aggressive, the distinction based on onset was viewed as adequately capturing the distinction between aggressive and nonaggressive subtypes. Since the publication of the DSM-IV, substantial research supported that CU traits could be assessed reliably in children and adolescents and their presence designated a unique group of children, primarily within the childhood-onset category. An estimated 25% to 35% of children and adolescents with childhood-onset CD show elevated CU traits. However, this minority of children with early-onset CD show a more severe, stable, and difficult-to-treat pattern of behavior, as well as distinct risk factors that have led to etiologic theories for different causal pathways leading to CD in those with and without elevated CU traits.

Comparative Nosology

The ICD-11 description of conduct-dissocial disorder strongly resembles diagnostic criteria for CD in DSM-5-TR, describing a persistent pattern of behavior that violates the basic rights of others or major age-appropriate norms, rules, or laws. While previous editions of the ICD classified CD as either CD confined to the family context, undersocialized CD, socialized CD, or ODD, ICD-11 brought its classification more in line with DSM-5. The ICD-11 definition includes specifiers based on the age of onset (childhood or adolescence) and level of prosocial emotions (limited or typical). One difference between the ICD-11 and DSM-5-TR specifier for LPE is that the ICD-11 includes a fifth symptom defined as “insensitivity to punishment.”

Epidemiology

The worldwide prevalence of CD is estimated to be 2% to 2.5%, with about 10% of individuals affected at some point during childhood or adolescence. There is little evidence that the prevalence of CD differs by world region, and findings of differences between racial and ethnic groups are mixed and not easily teased apart from socioeconomic factors. The prevalence of CD increases from childhood to adolescence, with the median age of onset of CD in the United States being 11.6 years old. Overall, boys show higher prevalence rates of CD than girls at approximately a 2:1 male-to-female ratio. However, like ODD, the sex ratio for CD varies across development, being narrower in early childhood and adolescence than in middle childhood. This narrowing of the sex difference in adolescence is due largely to the finding that girls typically show an adolescent onset to their CD.

Etiology

The vast body of research on antisocial and aggressive behavior generally or on CD specifically has produced several clear conclusions. First, the development of CD typically results from a host of risk factors in both the child's biologic and genetic background, as well as problems in their social environment. Thus, focusing on any single risk factor or even a single class of risk (e.g., biologic risk, familial risk, neighborhood risk) will likely be inadequate to explain how the vast majority of children and adolescents develop CD. Second, research on CD has clearly documented that children can develop CD through many different causal processes, which is reflected in the multiple subtypes of CD that have been proposed, each of which is associated with at least somewhat distinct risk factors.

Table 45–3.
A Summary of Risk Factors for Conduct Disorder

General Risk Factors	Risk Factors Unique to Specific Subtypes		
	CU Traits	Childhood-Onset	Adolescent-Onset
Prenatal exposure to toxins and birth complications	Stronger genetic contribution	Stronger evidence for shared and nonshared environmental influences	Low traditionalism
Socioeconomic disadvantage	Structural and functional deficits in amygdala	Frontal lobe dysfunction	High rebelliousness
Physical and sexual abuse and neglect	Insensitivity to punishment	Low executive functioning/verbal IQ	Poor parental monitoring and supervision
Deviant peer affiliation	Low fear and anxiety	Emotional dysregulation, including anxiety	Few bonds with prosocial institutions
Parental depression and substance use	Low reactivity to cues of distress in others	High reactivity to threat and provocation	
Community and media violence	Deviant social goals	Hostile attributional bias	
	Low parental warmth	Harsh and inconsistent parenting	

While research has historically focused on simply uncovering the multitude of dispositional and environmental factors that can place an individual at risk of showing CD symptoms, more recent work is focused on developing theoretical models that can explain both multiple risk factors and the heterogeneity of causal mechanisms. Further, many of these theories have attempted to embed this research on CD within research on normal development. That is, current theories of CD not only attempt to describe dispositional risk factors that have been empirically linked to CD, but also propose and test how these risk factors can affect important developmental processes in the child that make them more likely to display antisocial and aggressive behavior. These etiologic theories could have important treatment implications because, even if a child shows a historical risk factor that cannot be changed (e.g., prenatal exposure to toxins), if one can document what effects this exposure has had on the developing child, these effects could be the target of treatment.

This developmental psychopathology approach to explaining the causes of the various subtypes of CD, as defined in the DSM-5-TR, has led to theories specifying very different developmental mechanisms underlying adolescent-onset CD, childhood-onset CD with LPE, and childhood-onset CD without LPE. Adolescent-onset CD is viewed as an exaggeration of the normal process of identity formation, which leads to rebellious behavior as an attempt to establish autonomy and results in the rejection of traditional status hierarchies. This view explains why the CD behavior in this group does not onset until adolescence when this need for autonomy becomes more prominent, and why there is greater desistance from antisocial behavior in adulthood for individuals in this group. While most adolescents show some level of rebellious behavior, only a small minority show CD. The exaggeration of the normal process of rebellion can be due to associations with deviant peers who encourage maladaptive forms of rebellion, poor supervision by adults, a lack of connection to prosocial institutions, or simply a rebellious temperament. In contrast, childhood-onset CD is viewed as more of a characterologic disturbance because these behaviors start earlier in development and persist across multiple developmental stages. However, the different correlates to CD in children with versus without elevated CU traits suggest that the type of disturbance may differ for these groups. Children with elevated CU traits show deficits in how they process emotional stimuli that are evident very early in development, which can make it difficult for them to develop empathy, guilt, and other aspects of conscience. In contrast, children in the childhood-onset group without elevated CU traits show elevated levels of emotional reactivity and cognitive deficits (e.g., lower verbal intelligence and deficits in executive functioning) that can make it difficult for them to learn to regulate their emotions and behaviors.

Thus, this theoretical model specifies at least three common causal pathways to CD corresponding to the subtypes recognized by the DSM-5-TR, each involving different developmental mechanisms and somewhat different risk factors. This research is summarized in Table 45–3. In the subsequent sections, discussion of the various types of risk factors that have been associated with CD is organized within this framework.

Biologic Factors

HERITABLE FACTORS. Research has long shown that antisocial behavior, aggression, and conduct problems tend to cluster in families. Familial clustering can reflect shared environmental factors (i.e., factors that all siblings in a family share such as socioeconomic status), genetic factors, or the interaction between genetic vulnerability factors and adverse environmental experiences (i.e., gene-by-environment correlations and interactions). However, there is clear evidence from twin and adoption studies that there is a genetic influence on CD. Adoption studies have shown that biologic children of antisocial parents who are adopted away at birth to nonantisocial adoptive parents show elevated rates of CD. Further, twin studies have consistently shown that about 50% of the variance in CD symptoms are due to genetic influence, with the rest accounted for by shared and nonshared environmental influences. Importantly, there appears to be a different pattern of genetic influence across the subtypes of CD. Specifically, children with CD and elevated CU traits show stronger genetic influences compared to other subtypes.

Recent research has focused on gene-by-environment interactive effects in the development of CD. Several of these studies specifically investigate the gene for the enzyme monoamine oxidase type A (MAOA), which is involved in the metabolism of several central nervous system (CNS) monoamine neurotransmitters that are important in the etiology of many psychiatric disorders. Polymorphisms in the MAOA gene explained variance in the eventual development of CD for boys who experienced sexual or physical abuse. That is, abused boys with a gene variant for low MAOA activity were twice as likely to develop CD as those with a gene variant coding for high MAOA activity. In boys without an abuse history, variations in MAOA were unrelated to the eventual development of CD. This research demonstrates the importance of examining both environmental risk factors and genetic vulnerability factors, along with their interactions, in order to understand the etiology of CD.

PHYSIOLOGIC FACTORS. CD is associated with a number of physiologic factors related to activation of the autonomic nervous system and how it responds to emotional stimuli. The different developmental pathways to CD are particularly important for interpreting this research because the subtypes of CD can have very different and even opposing patterns of physiologic arousal. Specifically, CD in children with elevated CU traits is associated with reduced physiologic arousal to various emotional stimuli, such as signs of emotional distress in others or cues to punishment. These findings have been consistent across studies using very different measures of physiologic arousal, including the child's self-report of arousal, attentional orienting to pictures of emotional stimuli, skin conductance changes in response to visual or auditory emotional stimuli, changes in startle reflex following exposure to emotional stimuli, and paradigms that measure physiologic and behavioral responses to changes in reward and punishment contingencies. These physiologic correlates to CD in those with elevated CU traits are combined with neuroendocrine and neurocognitive differences that can also reflect emotional underarousal to certain stimuli.

Importantly, not only do children with CD without elevated CU traits not show this same consistent pattern of underarousal as is seen in those with elevated CU traits, but they often show the opposite pattern of arousal by being hyper-reactive to emotional stimuli. This includes emotional reactivity to provocation, which could help to explain the reactive aggression that is often displayed by children in the low CU group. From this research, it is apparent why using the DSM-5-TR LPE specifier to designate those with CD and elevated CU traits from those without elevated CU traits is critical when studying psychophysiological correlates to CD. Since some individuals with CD show overarousal and others show underarousal, grouping them together could lead to the conclusion that arousal plays no role in the development of CD since these two opposing patterns could cancel each other out. Further, these patterns of arousal could be critical for understanding how other correlates can influence the development of CD in children from the two subtype groups. The reduced emotional arousal may make children with elevated CU traits less likely to develop empathy and guilt because they do not become aroused to signs of distress in others and therefore are not motivated to avoid hurting others. The lack of response to punishment could also make many forms of discipline less effective for children with elevated CU traits, such that a focus on positive reinforcement becomes more important. In contrast, the overarousal in those without elevated CU traits may make them more likely to evoke reactive and hostile discipline from parents, leading to harsh and coercive interchanges between the parent and child.

NEUROENDOCRINOLOGIC FACTORS. Several studies have examined neuroendocrine risk factors associated with CD subtypes. The hypothalamic–pituitary–adrenal (HPA) axis is a major stress-regulating system in the body, whose final product is the release of cortisol, a stress hormone, into the bloodstream. Cortisol has many effects on biologic arousal systems, preparing the organism to respond to environmental threats. Research has studied basal, diurnal, and stress-reactive levels of cortisol in children and adolescents with CD, with somewhat mixed findings. These inconsistent results may be due to a failure to consider the subtypes of CD. Consistent with other biologic risk factors, there is some support for differences in emotional reactivity by CD subtype, such that those with elevated CU traits tend to show blunted cortisol activity in response to a stressor.

NEUROTRANSMITTER FACTORS. Research identifying specific neurotransmitters in relation to CD has been inconsistent, again possibly due to the failure to consistently consider differences across subtypes. However, three classes of neurotransmitters have been found to be associated with CD. First, there is some support for a relationship between lowered catecholaminergic activity (dopamine and norepinephrine) in the CNS and increased risk for aggressive behavior, ODD, and CD. Second, serotonin is a ubiquitous CNS neurotransmitter that plays an important role in stabilizing information processing in neural systems, resulting in coherent behavioral, emotional, and cognitive outputs. Disturbed serotonin neurotransmission is associated with disinhibited, aggressive, and impulsive behavior in humans, perhaps by increasing sensitivity to cues that elicit irritation and aggression and by diminishing sensitivity to stimuli that signal punishment or consequence for one’s actions. Decreased sensitivity of the serotonin 1A receptor (5-

HT_{1A}) is associated with increased impulsive aggression, whereas sensitivity of the serotonin 1B receptor gene (5-HT_{1B}) is associated with CU traits in boys. Thus, the exact nature of serotonin involvement in CD may differ depending on the subtype. Third, recent studies have found associations between methylation of specific sites on the oxytocin receptor genes and lower levels of circulating oxytocin in children with elevated CU traits. Oxytocin has been associated with a range of biologic and psychological functions, including greater attention to emotional stimuli.

NEUROANATOMIC FACTORS. Structural MRI studies of children and adolescents with CD or aggressive behavior consistently document reductions in grey matter volume across various cortical (i.e., ventrolateral, medial prefrontal, middle/superior temporal, and anterior insular cortices) and subcortical (i.e., amygdala, caudate, and putamen) regions. The most consistent abnormalities in CD are reduced cortical thickness in the ventromedial prefrontal/orbitofrontal cortex, superior temporal cortex, fusiform gyrus, precentral gyrus, and precuneus. These regions are implicated in moral and affective decision-making, face processing, motor functions, and self-referential processing, respectively. Most structural imaging studies have failed to consider the subtypes of CD. However, there is some evidence of a negative correlation between CU traits and cortical thickness in the right superior temporal cortex and lingual/fusiform gyrus, which are both involved in moral decision-making and face processing.

Functional MRI (fMRI) studies of CD have focused on emotion processing, using tasks involving viewing emotional images, facial expressions, or empathy-eliciting stimuli. These studies have reported lower activation in the dorsal and rostral anterior cingulate cortex (ACC), anterior insula, medial prefrontal cortex, ventral striatum, and amygdala in youth with CD. Reduced neural activity in the amygdala, anterior insula, and ACC have been specifically related to elevated CU traits in children with CD. These regions are implicated in fear conditioning and processing distress and pain cues in others. Few studies have directly compared CD children with and without CU traits on fMRI tasks, but those that have done so have found clear differences in brain activation, similar to those found for measures of physiologic arousal. That is, children with CD and elevated CU traits show reduced emotional reactivity in the amygdala in response to distress cues in others, whereas those without elevated CU traits show enhanced arousal in the amygdala.

PRENATAL AND PERINATAL FACTORS. There are a number of prenatal and perinatal risk factors that have been associated with CD, largely through their influence on the CNS of the developing child. Maternal smoking, drinking, drug use, and stress during pregnancy are the most consistently documented prenatal risk factors for CD. Perinatal risk factors for CD include birth complications and malnutrition. Hypoxia during birth can disrupt white matter tracts that have been associated with CD, and malnutrition could contribute to neuronal loss that has been associated with risk for CD.

Psychological Factors

COGNITIVE FACTORS. Some of the biologic factors associated with CD could have effects on the developing child through their influence on the child’s cognitive functioning. Lower intelligence, particularly verbal intelligence, and executive functioning deficits, are some of the most frequently studied cognitive correlates of CD, particularly for childhood-onset CD. These cognitive deficits could explain why some children with CD have difficulty anticipating the consequences of their behavior, as well as why CD is often linked to academic difficulties. Some children and adolescents with CD also fail to appropriately attend to and accurately encode social cues from others in the environment. For example, children with CD often attribute hostile intentions to others, even when the expressed cues are neutral and nonthreatening. Such deficits in social cognition are thought to explain the high rate of reactive aggression displayed by some children with CD.

These cognitive deficits seem to be most strongly related to children with CD who do not show CU traits, while children with elevated CU traits show a different set of cognitive deficits. Relative to other youth with CD, children, and adolescents with elevated CU traits display an insensitivity to punishment cues, which includes responding more poorly to punishment cues after a reward dominant response set is primed, responding more poorly to gradual punishment schedules, underestimating the likelihood that they will be punished for misbehavior, and being less sensitive to potential punishment when peers are present. Youth with CD and elevated CU traits are also more likely to view aggression as an acceptable means for obtaining goals, blame others for their misbehavior, and emphasize the importance of dominance and revenge in resolving social conflicts.

FAMILY FACTORS. Various forms of parental mental health problems have been associated with CD. As noted above, because antisocial behavior tends to run in families, parents of children with CD often also show antisocial and aggressive behavior. Higher rates of parental substance use and depression are also consistently linked to CD. Some of the most common parenting styles that are associated with CD involve low levels of positive parenting (i.e., infrequent use of positive reinforcement, low parental involvement, and low parental warmth and sensitivity) and high levels of negative parenting (i.e., poor monitoring and supervision of the child, inconsistent discipline, and use of coercive and harsh physical discipline). There is evidence that the link between problematic parenting and CD is bidirectional. That is, negative parenting can lead to increases in CD over time, and CD also evokes increasingly negative and decreasingly positive parenting over time.

Relevant to the subtypes of CD, poor monitoring and supervision are particularly important for adolescent-onset CD, possibly by allowing for greater time to affiliate and be influenced by deviant peers. Harsh and inconsistent discipline appears particularly important to the development of childhood-onset CD without elevated CU traits, by leading to harsh and coercive interactions between parent and child and by not providing the child with appropriate skills to regulate their emotions and behavior. Meanwhile, in those with elevated CU traits, parental warmth appears to be particularly important for the development of CD. Warm and responsive parenting is important to the formation of a secure parent–child attachment, which is disturbed in children with elevated CU traits and is foundational to their ability to form close bonds with others and to attend to and decipher others’ emotions.

Childhood physical abuse, sexual abuse, and a history of neglect are associated with an increased risk for the development of a wide variety of mental health problems, including CD. Further, childhood maltreatment could lead to different subtypes of CD through different causal mechanisms. An adolescent who did not have behavior problems as a child may rebel in maladaptive ways to escape abuse in the home, such as dropping out of school, running away from home, and stealing to support themselves. Also, physical abuse can lead a child to develop a hostile attributional bias in which they are hypervigilant to threat cues, making them more likely to act in a hostile, angry, and aggressive manner. Finally, chronic abuse and emotional deprivation can desensitize a child to the emotions of others and lead to a CU interpersonal style.

Sociologic Factors

NEIGHBORHOOD AND CULTURAL FACTORS. A number of extrafamilial risk factors have been associated with CD, including living in neighborhoods where the child might be exposed to high rates of community violence. Other societal factors that have been linked to CD are the easy availability of firearms and high levels of exposure to media violence. Further, poverty, low socioeconomic class, unemployment, poor housing, and overcrowded living conditions have all been linked to CD. There is some indication that the effects of socioeconomic deprivation are mediated through their effects on parenting behaviors. That is, parents living in environments with multiple chronic stressors may find it more difficult to provide consistent, warm, and responsive parenting.

PEER FACTORS. A consistent finding from the research is that children with CD are rejected by their peer group. They tend to have fewer friends, are liked less by their peers, and are more likely to be excluded from activities with peers compared to typically developing children. This rejection by peers is not surprising given that children with CD are often aggressive toward others and bully other children. While children with CD are often the instigator of aggression, they are also more likely to be the victim of aggression from others.

Children with and without elevated CU traits both appear to have problems with peers. Specifically, both groups tend to be disliked by a large number of peers. Although children with elevated CU traits are better able to make friends than other children with CD, their friendships tend to be unstable and unsatisfying, and their peers are more likely to describe them as “mean.”

CD is also related to association with deviant peers. Children with CD often have friends who also have behavior problems and engage in antisocial behavior. This association with deviant peers seems to be related to all subtypes of CD. However, while deviant peer affiliation is often considered to be a primary cause of adolescent-onset CD, it is thought to be a consequence of rejection by prosocial peers for the childhood-onset CD subtype. Children with elevated CU traits are especially likely to commit antisocial acts with peers. They are also likely to be the leaders of their antisocial peer groups and instigators of antisocial acts.

Diagnosis and Clinical Features

The DSM-5-TR diagnostic criteria for CD are provided in Table 45–2. The assessment of CD involves 3 primary goals:

- determining the presence and severity of the primary symptoms of CD, including their pervasiveness and the degree of harm they cause others;
- determining the presence of key indicators of the distinct developmental pathways to CD (i.e., the age of onset of the first CD symptom and the presence of CU traits); and
- assessing common comorbidities and key risks and maintaining factors that should be the target of treatment.

To assess the primary symptoms of CD, information should be obtained from multiple informants, including the child, their parents, and their teacher. The child's teacher appears to be particularly important for assessing conduct problems in children prior to adolescence. In contrast, self-report becomes increasingly important as the child approaches adolescence, and parents and teachers may be less aware of the adolescent's behavior. Finally, parents are often the best source of information on the timing of symptom onset. The use of standardized rating scales is recommended; these scales are often used to document the severity and pervasiveness of CD symptoms, compare the child's behavior to a normative comparison group, and identify treatment targets. The assessment of CU traits should also be conducted by collecting information from multiple sources. Particular care should be taken to determine if the indicators of CU traits are shown by the child in most situations and with most people rather than occasionally or only in certain relationships and settings. It is important for clinicians to screen for the presence of common comorbid psychiatric conditions such as ADHD, depression, bipolar disorder, substance use disorders, and learning disabilities (e.g., reading, speech, and language delays) that frequently accompany a diagnosis of CD.

Zach is a 12-year-old boy whose parents sought a psychological evaluation as a result of his continued pattern of severe behavior problems. From a very young age, he has displayed inattention, hyperactivity, and impulsivity, but his parents are more concerned about his defiance and oppositionality, which later progressed to rule-breaking and physical aggression. For example, he once locked his mother out of the house for several hours and then ran away from home overnight. In addition, he has repeatedly destroyed the toys and property of his two siblings, lied to avoid punishment for his behavior, stolen from extended family members, and threatened violence against his mother. At school, his teachers had repeatedly complained about the danger he posed to other students due to his frequent bullying, fighting, and physical cruelty to others, including an incident in fourth grade when he choked another student. Zach's parents and teachers, along with Zach himself, report that he does not feel guilty or apologize when he does something wrong or hurts someone. Instead, he blames others for his behaviors, such as stating that the classmate whom he choked "deserved" it for making a negative remark to him on the playground. He does not appear to care about the feelings of others and does not display empathy toward others. In addition, his emotions often appear shallow or inappropriate, even in distressing situations such as his grandfather passing away. Zach rarely puts forth effort in school or extracurricular activities but manages to do well on exams and earn A and B grades. At age 7, he received a diagnosis of ADHD and was placed on stimulant medication, which improved his inattention and hyperactivity. However, his oppositionality and conduct problems have only worsened since they began around age 5, and his parents do not know how to control him, given that punishment does not appear to help.

Pathology and Laboratory Examination

No laboratory tests are useful or required for making the diagnosis of CD. All individuals meeting diagnostic criteria for CD merit a complete physical examination to rule out any associated physical trauma associated with a high-risk lifestyle. Particular attention should be paid to the possibility of drug abuse and the risk of sexually transmitted diseases. In girls with CD, early-onset sexual activity may lead to the possibility of early pregnancy.

Due to the association with learning disabilities, psychoeducational testing to assess potential academic difficulties is also necessary.

Differential Diagnosis

Isolated instances of conduct problem behaviors that do not cause substantial daily impairment need to be differentiated from a consistent pattern of rule-breaking and violations of the rights of others that may constitute a disorder. Early-onset CD needs to be differentiated from ADHD, which can be marked by hyperactivity, impulsivity, risk-taking, and problems with affect regulation. CD and ODD share symptoms in common with intermittent explosive disorder (IED) that bring the individual into conflict with adults and authority figures such as teachers and parents. However, IED does not include the nonaggressive features of CD and contains no element of premeditation in order to achieve some contingent goal such as money or social dominance. ODD behaviors do not include overt aggression toward people or animals, destruction of property, or a pattern of theft and deceit. While ODD generally precedes CD in the course of development, many children with ODD do not go on to meet the criteria for CD. When diagnostic criteria are met for both ODD and CD, both diagnoses may be assigned.

Comorbid Conditions

Children and adolescents meeting diagnostic criteria for CD often meet criteria for other psychiatric diagnoses. ADHD frequently co-occurs with CD, particularly childhood-onset CD. Many youths with CD also have internalizing disorders, such as depression and anxiety. These internalizing disorders often onset after the CD, potentially as a result of the impairments caused by the behavior problems, such as conflict with others, disciplinary problems at home and school, or legal problems. Substance use disorders and learning disorders are also common among individuals with CD.

Course and Prognosis

The development and trajectory of CD depend greatly on the different developmental subtypes, and the three common patterns of development are illustrated in Figure 45–1. As noted in the figure, childhood-onset CD is typically preceded by ODD and ADHD in early childhood. For some, CU traits will also begin to emerge in early childhood, which can lead the child to show more severe and varied aggression, which includes both reactive and instrumental forms of aggression. Other children in the childhood-onset pathway who do not have elevated CU traits begin showing more severe behavior problems later in childhood. Their behaviors are less likely to be aggressive but, when aggression does occur, it tends to be limited to reactive forms. Research suggests that between 20% and 40% of children who show childhood-onset CD will be arrested or show ASPD as an adult. In addition, about 20% of those in the childhood-onset pathway with elevated CU traits will show significant levels of adult psychopathic traits.

In contrast, individuals who follow the adolescent-onset trajectory do not start showing significant behavior problems until they approach adolescence, at which point they start showing a range of problems involving symptoms of both ODD and CD. Their behavior problems tend not to be aggressive. However, their antisocial behavior can lead them to have problems at school, leading to school dropouts, as well as problems with the law. Their association with deviant peers can also lead to problems with substance abuse, which can continue into adulthood. In adulthood, individuals in the adolescent-onset group are less likely to have legal problems or ASPD compared to those in the childhood-onset group, but they can still have social, occupational, and mental health consequences of their adolescent behavior.

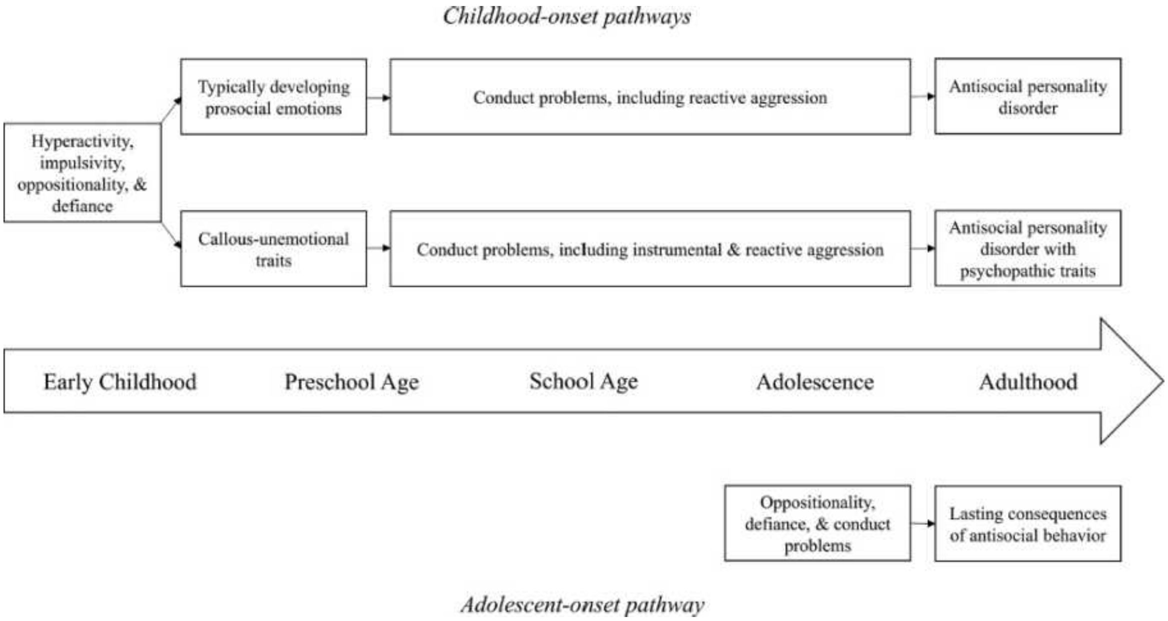
Childhood-onset pathways
Hyperactivity, impulsivity, oppositionality, & defiance
Typically developing prosocial emotions
Conduct problems, including reactive aggression
Antisocial personality disorder
Callous-unemotional traits
Conduct problems, including instrumental & reactive aggression
Antisocial personality disorder with psychopathic traits
Early Childhood
Preschool Age
School Age
Adolescence
Adulthood
Oppositionality, defiance, & conduct problems
Lasting consequences of antisocial behavior
Adolescent-onset pathway

FIGURE 45–1. Three common trajectories of conduct disorder are illustrated over the course of development.

Treatment

Given that most children who develop childhood-onset CD show ADHD and ODD prior to developing CD, successful treatment of these conditions can be considered secondary prevention of CD. Thus, pharmacologic treatment to treat ADHD, such as with CNS stimulants, has also been shown to reduce conduct problems in children with comorbid diagnoses of ADHD and either ODD or CD. Further, the PMT interventions that were described previously in the treatment of ODD are also effective for young children who have begun to show CD. Children with elevated CU traits have typically not responded as well to PMT and other behavioral interventions as other children with CD. Children with CD and elevated CU traits often start and end treatment with more severe behavior problems, despite showing some improvement. These findings have led to attempts to enhance PMT interventions for children with elevated CU traits by targeting their distinct risk factors and needs. These novel treatment targets include: (a) teaching parents how to increase their warm and responsive parenting, (b) emphasizing the use of consistent rewards to encourage positive behavior, including prosocial behavior, over punishment-based parenting strategies such as time-out, and (c) parent–child emotion recognition training. To date, such interventions have shown promising results but await randomized controlled trials to better document their effectiveness.

For older children and adolescents with CD, interventions that are multicomponent and can be tailored to the unique needs of the youth appear most effective. Such interventions recognize that CD is often caused and maintained by multiple factors, and the causal processes may differ across youths with CD. Further, the subtypes of CD can help to guide these comprehensive and individualized approaches to treatment. For example, interventions that focus on enhancing identity development and increasing contact with prosocial peers, such as mentoring programs or programs that provide structured after-school activities, may be effective for youth with adolescent-onset CD. In contrast, interventions that focus on anger control or on reducing harsh and ineffective parenting may be more effective for children in the childhood-onset group who do not exhibit elevated CU traits but often show problems with emotional regulation and come from families that use dysfunctional parenting practices. Finally, interventions that teach children to pay attention to emotional cues in others, motivate them to change their behavior in ways that capitalize on their self-interest using positive reinforcement over punishment, or focus on increasing parental warmth and responsivity may be more effective for children with elevated CU traits.

Several treatment models have been developed that are designed to provide this comprehensive and individualized approach to treatment. Two of these have proven to be quite effective in treating adolescents with CD in randomized controlled trials. Functional family therapy (FFT) represents a unique integration and extension of family systems and behavioral, cognitive, and affective approaches to treatment. FFT focuses on engaging the adolescent and the family in treatment, developing individualized motivations for change, and adapting its treatment to modify structural and systemic family processes that may be maintaining the youth’s problem behaviors. FFT also relies upon a number of proven behavioral principles to treat CD. For example, FFT targets enhancements in communication and mutual problem-solving skills between parents and adolescents, as well as improvements in parental contingency management. Finally, FFT focuses on the generalization of new skills to other settings, such as school or the legal system.

Another comprehensive and individualized treatment model for adolescents with CD is multisystemic therapy (MST). MST was specifically developed as a treatment to address the multiple risk factors that can lead to severe antisocial and delinquent behavior in adolescents. As its name implies, MST tries to intervene in multiple systems, familial and extrafamilial (e.g., maladaptive parenting, deviant peer or gang affiliation, poor school achievement), that can influence the adolescent's behavior. Emphasis is placed on making changes in the adolescent's social ecology that will reduce problem behavior and support these changes after treatment ends. Treatment can range substantially across individuals, depending on an initial needs assessment.

As is the case for PMT, both FFT and MST have been proven to reduce behavior problems in youth with elevated CU traits. However, also consistent with the research on PMT, children with CD and CU traits still often leave treatment with more severe behavior problems than other children with CD. Thus, future work should focus on how to enhance these and other comprehensive approaches to treatment to be more effective for the subgroup of youths with CD and CU traits.

FUTURE DIRECTIONS

There have been a number of major advances in our understanding of the causes of ODD and CD, as well as advances in the treatment of these disorders in recent years. These advances have largely been the result of the fact that research has gone beyond simply documenting the many different risk factors that can play a role in the development or maintenance of these disorders. Instead, recent research is trying to determine the most common ways in which these risk factors can negatively impact important developmental processes, leading to behavior that brings the child into conflict with authority (ODD) or to act in a way that violates the rights of others or major age-appropriate norms (CD). This work has helped to delineate several common developmental pathways to ODD and CD that are now reflected in psychiatric diagnosis.

The importance of this approach to research on causal factors is exemplified by work showing that some children with CD show under-reactive responses to emotional stimuli (i.e., those with elevated CU traits), whereas others show hyper-reactive emotional responses (i.e., those without elevated CU traits). Failure to differentiate these subtypes likely is the reason that previous research on the emotional correlates of CD has been inconsistent. It is possible that future research in molecular genetics and neuroimaging will result in more consistent findings on the neurocognitive correlates of CD once potential differences across subtypes are consistently tested.

This research on developmental pathways to CD has also aided the development of more effective treatments for CD. Treatments no longer focus on changing a single risk factor in the same way for all children and adolescents with CD. Instead, treatments now focus on targeting multiple risk factors and are tailored to the individual needs of the youth with CD. Interventions are beginning to use research on the developmental pathways to CD to guide these tailored approaches to treatment. A promising direction of future treatment research is adapting evidence-based treatments with enhancements specifically designed to be more effective for children with elevated CU traits, a subgroup who heretofore have not responded as well to traditional treatments.

FURTHER READINGS

Blair RJR, Veroude K, Buitelaar JK. Neuro-cognitive system dysfunction and symptom sets: a review of fMRI studies in youth with conduct problems. *Neurosci Biobehav Rev*. 2018;91:69–90.

Canino G, Polanczyk G, Bauermeister JJ, Rohde LA, Frick PJ. Does the prevalence of CD and ODD vary across cultures? *Soc Psychiatry Psychiatr Epidemiol*. 2010;45:695–704.

Caspi A, McClay J, Moffitt TE, et al. Role of genotype in the cycle of violence in maltreated children. *Science*. 2002;297:851–854.

Cecil CAM, Lysenko LJ, Jaffee SR, et al. Environmental risk, Oxytocin Receptor Gene (OXTR) methylation and youth callous-unemotional traits: a 13-year longitudinal study. *Mol Psychiatry*. 2014;19:1071–1077.

Colman I, Murray J, Abbott RA, et al. Outcomes of conduct problems in adolescence: 40 year follow-up of national cohort. *BMJ*. 2009;338:a2981.

Comer JS, Chow C, Chan PT, Cooper-Vince C, Wilson LA. Psychosocial treatment efficacy for disruptive behavior problems in very young children: a meta-analytic examination. *J Am Acad Child Adolesc Psychiatry*. 2013;52:26–36.

Dadds MR, Frick PJ. Toward a transdiagnostic model of common and unique processes leading to major disorders of childhood: the REAL model of attention, responsiveness, and learning. *Behav Res Ther*. 2019;119:103410.

Demmer DH, Hooley M, Sheen J, McGillivray JA, Lum JA. Sex differences in the prevalence of oppositional defiant disorder during middle childhood: a meta-analysis. *J Abnorm Child Psychol*. 2017;45:313–325.

Fairchild G, Hawes DJ, Frick PJ, et al. Conduct disorder. *Nat Rev Dis Primers*. 2019;5:1–25.

Fairchild G, van Goozen SH, Calder AJ, Goodyer IM. Research review: evaluating and reformulating the developmental taxonomic theory of antisocial behavior. *J Child*

Psychol Psychiatry. 2013;54:924–940.

Frick PJ. Developmental pathways to conduct disorder: implications for future directions in research, assessment, and treatment. *J Clin Child Adolesc Psychol*. 2012;41:378–389.

Frick PJ, Nigg JT. Current issues in the diagnosis of attention deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder. *Annu Rev Clin Psychol*. 2012;8:77–107.

Frick PJ, Ray JV, Thornton LC, Kahn RE. Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents? A comprehensive review. *Psychol Bull*. 2014;140:1–57.

Frick PJ, Viding E. Antisocial behavior from a developmental psychopathology perspective. *Dev Psychopathol*. 2009;21:1111–1131.

Granic I, Patterson GR. Toward a comprehensive model of antisocial development: a dynamic systems approach. *Psychol Rev*. 2006;113:101–131.

Hudziak JJ, Derks EM, Althoff RR, Copeland W, Boomsma DI. The genetic and environmental contributions to oppositional defiant behavior: a multi-informant twin study. *J Am Acad Child Adolesc Psychiatry*. 2005;44:907–914.

Kaminski JW, Claussen AH. Evidence base update for psychosocial treatments for disruptive behaviors in children. *J Clin Child Adolesc Psychol*. 2017;46:477–499.

Keenan K, Wakschlag LS, Danis B, et al. Further evidence of the reliability and validity of DSM-IV ODD and CD in preschool children. *J Am Acad Child Adolesc Psychiatry*. 2007;46:457–468.

Kimonis ER, Fleming G, Briggs N, et al. Parent-Child Interaction Therapy adapted for preschoolers with callous-unemotional traits: an open trial pilot study. *J Clin Child Adolesc Psychol*. 2019;48(S1):S347–S361.

Matthys W, Vanderschuren LJ, Schutter DJ. The neurobiology of oppositional defiant disorder and conduct disorder: altered functioning in three mental domains. *Dev Psychopathol*. 2013;25:193–207.

Moffitt TE, Arseneault L, Jaffee SR, et al. Research review: DSM-V conduct disorder: research needs for an evidence base. *J Child Psychol Psychiatry*. 2008;49:3–33.

Noordermeer SDS, Luman M, Oosterlaan J. A systematic review and meta-analysis of neuroimaging in oppositional defiant disorder (ODD) and conduct disorder (CD) taking attention-deficit hyperactivity disorder (ADHD) into account. *Neuropsychol Rev*. 2016;26:44–72.

Odgers CL, Caspi A, Broadbent JM, et al. Prediction of differential adult health burden by conduct problem subtypes in males. *Arch Gen Psychiatry*. 2007;64:476–484.

Pasalich DS, Dadds MR, Hawes DJ, Brennan J. Do callous-unemotional traits moderate the relative importance of parental coercion versus warmth in child conduct problems? An observational study. *J Child Psychol Psychiatry*. 2011;52:1308–1315.

Rowe R, Maughan B, Pickles A, Costello EJ, Angold A. The relationship between DSM-IV oppositional defiant disorder and conduct disorder: findings from the Great Smoky Mountains Study. *J Child Psychol Psychiatry*. 2002;43:365–373.

Thornton LC, Frick PJ, Shulman EP, et al. Callous-unemotional traits and adolescents' role in group crime. *Law Hum Behav*. 2015;39:368–377.

Viding E, Price TS, Jaffee SR, et al. Genetics of callous-unemotional behavior in children. *PLoS One*. 2013;8:e65789.

Viding E, Sebastian CL, Dadds MR, et al. Amygdala response to preattentive masked fear in children with conduct problems: the role of callous-unemotional traits. *Am J Psychiatry*. 2012;169:1109–1116.

Whelan YM, Stringaris A, Maughan B, Barker ED. Developmental continuity of oppositional defiant disorder subdimensions at ages 8, 10, and 13 years and their distinct psychiatric outcomes at age 16 years. *J Am Acad Child Adolesc Psychiatry*. 2013;52:961–969.

Wilkinson S, Waller R, Viding E. Practitioner review: involving young people with callous unemotional traits in treatment–does it work? A systematic review. *J Child Psychol Psychiatry*. 2016;57:552–565.

46

Feeding and Eating Disorders of Infancy and Early Childhood

Irene Chatoor, M.D., and Loredana Lucarelli, Psy.D.

The term *feeding and eating disorders* describes a variety of conditions including food refusal, food selectivity, eating too little, food avoidance, fear of feeding, pica, and rumination in infancy and early childhood. Particularly, the term *feeding* is commonly used to emphasize the dyadic nature of the “feeding relationship” in infants and young children that involves complex interactions between the child and the parent. During a reciprocal feeding process, optimal parent–child interactions depend on sensitive, responsive parenting as well as the child’s ability to clearly communicate hunger and satiety cues. It is estimated that in the general population, 25% to 45% of infants and young children have various feeding and eating problems and that severe disorders associated with poor weight gain occur in 1% to 5% of pediatric hospital admissions and between 4% and 14% in ambulatory care. However, an exact evaluation of the prevalence or incidence rates of feeding and eating disorders, and of particular diagnostic subtypes, has not been published up to now because of the variability of definitions used and the lack of a general diagnostic classification of feeding and disorders shared by clinicians and researchers. Overall, several studies have found that early feeding and eating problems not only cause distress to the young child but affect the whole family and often continue into later childhood and adolescence.

DEFINITION

Criteria for Feeding Disorder of Infancy or Early Childhood in DSM-IV have been criticized for being too restrictive and not capturing many young children with feeding difficulties, and an alternative system in DC: 0–3R (2005) has often been used. This system includes six subtypes of feeding behavior disorder proposed by Chatoor. Avoidant/Restrictive Food Intake Disorder (ARFID) in DSM-5 replaced Feeding Disorder of Infancy or Early Childhood in DSM-IV, and more recently, ARFID was introduced in the 11th Revision of the World Health Organization's International Classification for Diseases (ICD-11). ARFID has expanded the earlier, more limited diagnostic criteria by introducing three main categories of feeding/eating problems based on the three subtypes of Feeding Behavior Disorder in DC: 0–3R, and the text addresses the three remaining feeding disorders from DC: 0–3R. In addition, the ARFID criteria can be applied to individuals of all ages, consistent with the recognition of the continuity of psychopathology across the age range from early childhood to adulthood. ARFID is defined as a persistent disturbance in feeding or eating that can result in severe malnutrition, significant weight loss or a failure to gain weight, growth compromise, and/or a marked interference with psychosocial functioning. ARFID provides a diagnostic label for a heterogeneous group of individuals across the age range who engage in avoidant or restrictive eating behaviors without weight or body image concerns. The new criteria of ARFID in DSM-5 and its text revision (DSM-5-TR) describe three variants of an eating or feeding disturbance that is characterized by, for example, apparent lack of interest in eating or food; avoidance based on the sensory characteristics of food; concern about aversive consequences of eating. These new DSM-5 criteria of ARFID with the accompanying text map directly on three of the DC: 0–3R subtypes. The new subtype, "apparent lack of interest in food or eating," has been described in DC: 0–3R as infantile anorexia (IA); the second variant, “avoidance based on the sensory characteristics of food,” describes children who select or avoid food in relation to the appearance, taste, texture, smell, or temperature of food and has been defined in DC: 0–3R as Sensory Food Aversions; the third variant “concern about aversive consequences of eating” is described in DC: 0–3R as Feeding Disorder Associated with Insults to the Gastrointestinal Tract and has been described by Chatoor as Post-Traumatic Feeding Disorder (2002).

In addition, in the text, there is allowance to diagnose ARFID in infants and children with the remaining three feeding disorders described in DC: 0–3R, those who have difficulty with state regulation, those who may not engage with the caretaker during feeding, or those whose feeding problems are associated with a medical condition when the disturbance of intake is beyond what can be explained by the physical condition.

To delineate feeding disorders from less severe feeding difficulties, DSM-5-TR expanded the criteria of impairment to affect one or more of four areas of health or functioning:

- Significant weight loss (or failure to achieve expected weight gain or faltering growth in children)
- Significant nutritional deficiency
- Dependence on enteral feedings or oral nutritional supplements
- Marked interference with psychosocial functioning

HISTORY AND COMPARATIVE NOSOLOGY

Prior to the inclusion of the diagnostic category of Feeding Disorder of Infancy or Early Childhood in the DSM-IV, there were no nationally defined criteria for the diagnosis of feeding disorders. In the absence of a recognized standard, different authors have used several labels for feeding problems: picky eaters, choosy eaters, selective eaters, problem eaters, food refusal, feeding resistance, food phobia, food aversion, and dysphagia. However, none of these diagnostic labels has been operationalized. The field of feeding disorders has been further complicated by the fact that feeding disorder and failure to thrive (FTT) have been used interchangeably. The term *failure to thrive* refers to inadequate weight gain based on standard growth charts. Historically, FTT has been dichotomized along organic and nonorganic axes. Whereas the former attributes growth failure to a medically diagnosable cause, the latter has commonly been thought to reflect maternal deprivation or neglect. A third category was later added and represents a mixture of organic and environmental etiologic factors of growth failure.

Concerns have been expressed about the usefulness of the label FTT as a “catch-all” diagnosis for all feeding disorders. First, the overlap between FTT and feeding disorders is confusing because not all infants with FTT have feeding disorders, and not all infants with feeding disorders have FTT. Second, Goldbloom pointed out that the term FTT is purely descriptive and does not address the process that caused the infant to grow inadequately. Although many authors agree that FTT should be viewed as a symptom rather than a diagnosis, FTT continues to be used as a diagnostic label, particularly in the pediatric and psychological literature.

As described above, because of the limitations of the diagnostic criteria of Feeding Disorders of Infancy or Early Childhood in DSM-IV, an alternative classification of feeding disorders has been adopted by DC: 0–3R, which served as a template for the new diagnostic category of ARFID in DSM-5 and DSM-5-TR.

The original diagnostic criteria were published by Chatoor in 2002 and were modified with the help of the 2003 Task Force for Research Diagnostic Criteria for Infants and Preschool Children. They were further developed with the help of a work group that was supported by the American Psychiatric Association, and the revised criteria were published in 2007 in *Age and Gender Considerations in Psychiatric Diagnosis.* As explained above, the diagnostic criteria for these six feeding disorder subtypes were also included in the 2005 edition of the *Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood: Revised Edition* (DC: 0–3R).

The six feeding disorder subtypes are as follows: (1) feeding disorder of state regulation, (2) feeding disorder of caregiver–infant reciprocity, (3) IA, (4) sensory food aversions, (5) feeding disorder associated with a concurrent medical condition, and (6) feeding disorder associated with insults to the gastrointestinal tract. The first four feeding disorders tend to start during specific developmental periods and are arranged accordingly, whereas post-traumatic feeding disorder and feeding disorder associated with a concurrent medical condition can occur at various stages of the child's development.

Since recently published research on feeding disorders has frequently used the classification in DC: 0–3R, the authors will discuss the various feeding disorders as listed in DSM-5-TR and will use the descriptions of the feeding disorders in DC: 0–3R as subtitles.

SPECIFIC FEEDING DISORDER SUBTYPES

Avoidant/Restrictive Food Intake Disorder

Subtype: apparent lack of interest in eating or food (DSM-5-TR)
Infantile Anorexia (DC: 0–3R)

Diagnosis and Clinical Features. IA is characterized by poor appetite, lack of interest in eating, and poor growth in an otherwise active and playful child. It was first described by Chatoor in the 80s and was adopted in DC: 0–3R (2005) as one of the subtypes of Feeding Behavior Disorder. It usually begins in the first 3 years of life, most commonly between 9 and 18 months of age, when infants learn to walk, begin to talk, and are transitioned to spoon and self-feeding. The children are active and playful and appear to be interested in everything else but food and eating. They resist going into the highchair; they refuse to open their mouth to be fed and throw the food and feeding utensils. The parents become increasingly worried because of the children's poor food intake and growth, and they often resort to distracting, coaxing, and bribing to get the children to eat. Some parents become so anxious that they try to feed the children whenever they are willing to open their mouth, and some become so desperate that they resort to force-feeding their children. Maternal worry about the child's underweight mediates and moderates the mother–child feeding interactions and perpetuates the feeding disorder. Consequently, there is increasing conflict between the children and their parents, and mealtimes become miserable for the whole family.

Differential Diagnosis. IA needs to be differentiated from other feeding disorders that are characterized by food refusal. ARFID subtype: "concern about aversive consequences," usually has a more sudden onset of the food refusal following a traumatic event to the oropharynx or gastroesophageal tract, such as choking, severe gagging, or vomiting, or the insertion of feeding or endotracheal tubes. The infant with this feeding disorder may refuse all food, only solid food, or only the bottle, depending on what mode of feeding the infant associates with the traumatic experience(s). Infants with ARFID subtype: "avoidance based on the sensory characteristics of food" or sensory food aversions (DC: 0–3R) consistently refuse specific foods but eat well if offered preferred foods.

Etiology and Course. Toddlers with IA have been found to have a higher level of physiological arousal and more difficulty down-regulating their arousal. This may explain why they have so little awareness of hunger because they seem very excited by their play and enjoy talking, but they have difficulty turning away from these pleasurable activities. Studies from different centers have found significant correlations among difficult child temperament, irregular feeding and sleeping patterns, negative and willful behaviors by the toddlers, and mother–child conflict during feeding. At the same time, mothers' insecure attachment to their parents and mothers' drive for thinness and bulimia, anxiety, and depression also correlated significantly with mother-child conflict during feeding. Finally, mother–child conflict during feeding correlated strongly with the child's weight: the more conflict between mother and child, the lower the child's weight. However, correlations need to be used with caution to explain etiology, although the feeding histories of children with IA suggest that the lack of appetite and food refusal by the toddlers evoke severe anxiety and feelings of helplessness in the parents, who then try to engage the children in distractions and do anything to get the children to eat. The more insecure the parents feel, the more they allow the child to control the feeding situation, and the more ineffective the parents become in setting appropriate limits to the child's demanding behaviors. Consequently, parents and children become trapped in maladaptive interactional patterns.

Research further revealed that toddlers with IA, on average, performed in the normal range of cognitive development. However, the Mental Developmental Index (MDI) score of a control group of healthy eaters was 11 points higher than that of the group of toddlers with IA. Further analyses showed that not weight but mother-toddler interactions during feeding and play seemed to contribute significantly to these differences. Mother–toddler conflict and struggle for control during feeding and maternal intrusiveness during play were associated with the lower cognitive performance of the toddlers.

A longitudinal study from Italy by Lucarelli and colleagues of 58 boys and 55 girls with IA who received only minimal intervention at the time of diagnosis revealed some interesting findings. The diagnosis of IA was made by two independent clinicians, and following the publication of DSM-5, the research team clarified that the young patients originally diagnosed with IA and enrolled in this study met the criteria for the ARFID subtype "apparent lack of interest in eating or food." The mothers of the children who were diagnosed with IA were recontacted and invited to take part in a pediatric and psychological assessment at three time points at 5, 7, and 11 years of age. Mothers received some psychoeducation about this specific eating disorder at the time of the diagnosis and during subsequent assessments but did not pursue any specific psychotherapeutic treatment for various reasons: lived or moved to another city, lived in other areas of Italy without appropriate resources, refusal by one or both parents to participate in the study, or precocious termination of treatment. However, the children were followed closely by the pediatricians at the hospital in Rome and in their community, and they were periodically called back for subsequent appointments at the hospital in Rome. About 70% of the children with IA seen at follow-ups from 2 to 11 years of age, showed ongoing eating problems and higher scores in anxiety, moodiness, aggressive behaviors, and somatic complaints when compared to the controls. These longitudinal data also showed significant correlations between continuing maternal eating and psychopathology, especially depression and anxiety, and internalizing problems of the children. At 11 years of age, the girls' emotional/behavioral problems and their mothers' psychopathology and disturbed eating attitudes were more severe than that of the boys and their mothers. These results demonstrated a significant effect of sex–namely that at the age of 11 years, especially the girls and their mothers demonstrated increased eating pathology and associated psychopathological symptoms. Further longitudinal research is necessary to better understand how these children will negotiate adolescence. It raises the question of whether these children will develop a fear of becoming fat and transition to the full picture of anorexia nervosa or other restrictive eating disorders during the critical developmental period of adolescence, which represents a peak period for the onset of eating disorders, especially in girls. Overall, this study on the natural course of IA highlights the importance of early diagnosis and specific interventions for children diagnosed with IA.

A follow-up study of 20 children at 6 months to 2 years at the time of diagnosis, who had completed a specific treatment described later, demonstrated that 17 children had changed considerably. The conflict in the parent–child relationship over the child's food intake was resolved. The children had learned to recognize hunger and increased their food intake, which led to weight gain and improved growth. However, the parents reported that many of the children continued to lose their appetite when they were excited, if they had houseguests, if they had a birthday party, or if they traveled. Once they calmed themselves, they would regain their appetite and usually make up for the lost meals. There were three children who continued to struggle with their parents over their poor eating. In two cases, day care seemed to be too stimulating to allow these children to relax enough for eating. In the third case, the home situation was very overcrowded and disorganized.

Clinical experience suggests that older children who have been untreated and who present for the first time with the diagnosis of IA during school age are usually thin and short for their age. However, their head size and their intellectual development appear to be normal for their chronologic age. This was confirmed in a follow-up study by Chatoor and colleagues of children who were treated as toddlers for IA. It demonstrated that two-thirds of the children had outgrown their feeding disorder whereas one-third did not. Regardless of their malnutrition at the time of diagnosis or at follow-up, there was no correlation between their weight and intellectual functioning. Indeed, one-third of the children functioned in the superior range of intelligence.

To date, few studies have followed children with feeding disorders over time with the exception of the Italian study described earlier. Other longitudinal studies which focused on the development of infants/toddlers with feeding problems have not attempted to differentiate between subtypes of feeding disorders. As a result, it is not known how specific feeding disorders in early childhood are related to eating disorders in adolescence and adulthood. This will be a challenge for future research.

Treatment. This subsection describes a treatment that is based on a transactional model for the development of IA. According to this model, the infant’s difficult temperament characteristics of emotional intensity, stubbornness, lack of hunger cues, irregular feeding, and sleeping patterns evoke conflicts over issues of control, autonomy, and dependency in a vulnerable parent who is anxious and insecure in regard to limit setting. The treatment addresses the infant’s temperament and the parents’ anxiety in regard to limit setting to facilitate internal regulation of eating of the infant.

1. The infant’s temperament is discussed with the parents to help them understand that the infant’s curiosity and emotional intensity

interfere with the infant’s awareness of hunger. Because the infant is very strong-willed and has learned that food refusal is a powerful tool with which to control the parents’ attention, efforts by the parents to distract or entertain the infant to get them to eat may help temporarily but only perpetuate the problem by getting the infant more excited.

2. The parents’ source of anxiety in regard to limit setting is explored. Some parents have been sensitized by previous losses of pregnancies, by infertility problems, or by the harshness and/or emotional unavailability of their parents. Understanding the parents’ sensitivity in regard to limit setting is critical to helping them to be more effective in this area with their infant.
3. After this ground work, the parents are trained with respect to changing their behavior so as to facilitate internal regulation of eating by their infant. To boost the infant’s biologic rhythms of hunger and satiety, the parents are encouraged to feed the infant only at regular 3- to 4-hour intervals and not to offer anything but water in between these regular feedings. They are helped to understand that it is important to encourage the infant’s self-feeding by commenting on the infant’s feeding skills (“you get the spoon in your mouth all by yourself”), but they are not to comment on how much or how little the infant eats. It is helpful to give the infant only a few bites at a time and after a while change to another food to keep the infant interested in eating and engaged with the parent without the parent becoming intrusive or distracting the infant. In addition, this helps the infant to continue eating until full and to learn to regulate their food intake by their internal signals of hunger and satiety. Eating should not be a performance for the parents. There should be no distractions during feedings, and no television or games, so that the infant is kept focused on eating. If the infant engages in inappropriate behaviors during feeding (climbing out of the high chair, throwing food, or feeding utensils), the parents need to learn to use the “time-out” procedure. The time-out is used to accomplish two goals: to help the infant to stop behaviors that interfere with eating and to learn self-calming when upset because they cannot have their way.

This three-step intervention with the parents is best accomplished in double sessions, lasting 2 hours each, with the first two sessions being held within a few weeks of each other. The intensity of this brief intervention facilitates a close therapeutic alliance between the therapist and the parents and enables many parents to make major shifts in their interactions with their infants. Many of these parents need only a few follow-up sessions to deal with some specific questions that may arise. Frequently, after the infant has changed their eating pattern, the parents ask for help concerning the sleeping problems that many of these children have as well. However, there are some families who are more deeply entrenched in maladaptive interactional patterns with their infants, and some infants are so severely malnourished that more intensive interventions or hospitalization may be required. In particular, considering the associations between the severity of IA and risk factors such as high mother-child interactional conflict, difficult child temperament, and emotional and eating dysregulation in both the children and their mothers, the therapeutic approach should have a dual focus on the child and on the parents.

Rachel was 18 months old when she was referred by her pediatrician because of food refusal and malnutrition. The mother looked exhausted as she described how Rachel had never been a big eater from birth but that she stayed on the 25th percentile for weight and height on the growth chart until she was about 10 months of age. Her refusal to be fed by spoon became so strong that her intake of solid food became less and less. She depended primarily on bottles with milk for her nutrition. In spite of all of the parents' efforts to distract Rachel during feeding, to offer different foods, and to feed her whenever she would accept anything (including at night) and their desperate efforts to force food into her mouth, Rachel's food refusal had gotten worse, and she had fallen below the 5th percentile on the growth chart for weight and height. The mother admitted that trying to feed Rachel had taken over her entire life and that she felt exhausted and envious when she watched her friends' children eating without any effort on the part of the parents.

The observation of mother and infant during feeding and play revealed that Rachel was a thin little girl who seemed "all eyes," observing everything in the room, including the pictures on the wall and her image in the mirror. However, she paid little attention to the food. She ate some finger food but refused to open her mouth and arched her back when her mother tried to feed her with spoon. After a few minutes, she appeared bored sitting in the high chair, threw the food, and indicated by grunting that she wanted to be taken out of the chair. Once she had her freedom, she ran around the room entertaining herself by exploring the furniture while her mother sat in her chair looking anxious and defeated.

The treatment addressed Rachel's special temperament characteristics, her intense curiosity and emotional intensity, her strong will, and her poor awareness of hunger, which seemed to have led to her increasing food refusal. The parents were helped to understand that they had a child with special temperament characteristics that required special parenting with special emphasis on limit setting. Specific instructions were discussed on how to handle mealtime, how to deal with Rachel's food refusal, and how to use "time-out" with a young child.

Although the mother struggled initially with how to set limits to Rachel's inappropriate behaviors—for example, stopping her from throwing food and keeping her in the high chair—the parents supported each other in implementing the new feeding instructions and were pleased with the change in Rachel's behavior. Once Rachel realized that her screaming did not change the rules, she stayed in the high chair without protest and became interested in feeding herself with the spoon. Within a few weeks, she seemed more aware of hunger, and her food intake increased accordingly. When seen for follow-up 1 year later, there was no longer conflict over Rachel's eating. Rachel had gained weight and grown in length, which brought her back to the 10th percentile on the growth chart.

In recent years, several studies have explored the effects of cyproheptadine therapy on children with feeding difficulties and poor growth. Although most children seem to tolerate the medication quite well, their appetite and food intake improves and resulting in better weight gain. The initial positive effect often wears off after a few weeks, and the medication is then given 3 weeks on and 1 week off or in shorter intervals of 4 days on and 3 days off in order to maintain the positive effect on appetite stimulation,

Avoidant/Restrictive Food Intake Disorder

Subtype: avoidance based on the sensory characteristics of food (DSM-5-TR)

Sensory Food Aversions (DC: 0–3R)

Diagnosis and Clinical Features. Sensory food aversions are common. However, whereas some children refuse to eat only a few specific foods, others may refuse most foods and take in only a limited diet of a few foods. The diagnosis of a feeding disorder should only be made if the food selectivity results in dietary deficiencies, and/or has led to oral motor and speech delay, and/or in older children is associated with social anxiety around eating and has led to marked interference of psychosocial functioning.

Within this disorder, food refusal is related to the taste, texture, temperature, or smell of particular foods. When the infant is fed specific foods, the infant’s aversive reactions range from grimacing to spitting the food out and wiping the tongue or mouth to gagging and vomiting. After an initial aversive reaction, the infants usually refuse to continue eating that specific food. In addition, infants tend to generalize and refuse to even try foods of similar color, appearance, or smell. Occasionally, sensory food aversions become apparent in the newborn period when the infant is switched from breast milk to formula or from one formula to another. However, most commonly, infants begin to show aversive reactions when introduced to baby foods with different flavors and different textures or in the second year of life, during the introduction of different table foods. After the initial aversive reactions described earlier, some infants show such intense food refusal that they become limited to eating only a small variety of foods. They may refuse to eat whole food categories, such as vegetables, fruits, or meats. In extreme cases, they even refuse to eat a preferred food if it touched a “feared” food on the plate or if the food is not prepared by a specific restaurant or company. Children who refuse to eat vegetables and fruits usually develop vitamin deficiencies, and children who refuse to drink milk or eat meats may develop protein, zinc, and/or iron deficiencies. In addition, children who refuse to eat foods that require more chewing, such as meats, hard vegetables, or fruits, present with delay in oral motor development, which may be associated with articulation difficulties.

In addition to their sensory food aversions, many of these children experience hypersensitivities in other sensory areas as well. Parents frequently report that these children do not like to get their hands “messy,” become distressed when asked to walk on grass or sand, and do not like to wear socks or shoes, certain fabrics, or labels on clothing. Many of these children are also hypersensitive to smells, and some may have difficulty with loud sounds.

Differential Diagnosis. This feeding disorder needs to be differentiated from oppositional food refusal which is common during the toddler years. Toddlers who like to exercise control over their parents may state that they do not want to eat a certain food and demand something else to eat. However, the food choices of oppositional toddlers are inconsistent and vary according to the mood of the toddler, whereas toddlers with sensory food aversions are usually consistent in their food choices, and they become distressed when the parents insist that they eat a certain food to which they are averse. Sensory food aversions also need to be differentiated from ARFID subtype: concern about aversive consequences of eating (DSM-5-TR) or described as post-traumatic feeding disorder by Chatoor, which is characterized by the child’s refusal to eat all solid food or to drink from the bottle, depending on what mode of eating the child associates with the traumatic event(s). However, some children with sensory food aversions may develop a post-traumatic feeding disorder if they have a strong aversive reaction to a food resulting in gagging or vomiting, and/or the parents continue to insist that the child eat a food to which the child is averse. These children may become fearful of eating altogether.

Etiology and Course. There are no empirical studies that have examined the etiology of sensory food aversions in infants and toddlers. However, research with preschool and older children and with adults has related taste sensitivities to the bitter substance propylthiouracil, to strong food preferences, and to the number of taste buds on the individual’s tongue (those with strong taste sensitivities, the “supertasters,” have significantly more taste buds than “nontasters”). A study by Cooke and colleagues that examined food neophobia (the fear to try new foods, which is one of the characteristics of sensory food aversions) in a large sample of twins aged 8 to 11 years estimated that 78% was heritable and another 22% of the variance was explained by nonshared environmental factors. Some studies have documented that parents and their biologic children can share taste sensitivities and various models of heritability—for example, incomplete penetrance, a two-locus model, and multilocus and multiallele models—have been suggested. In 2003, Kim and colleagues reported that specific polymorphisms on the taste receptor protein gene family Tas2r on chromosome 7 differentiated supertasters from nontasters.

In addition, certain aspects of the eating environment have been linked to the development of food preferences. For example, parents with strong taste sensitivities may offer a restricted range of food to their children, and limited exposure to a variety of foods may enhance toddlers' food selectivity. On the other hand, if parents try to coax children or become forceful to get their children to try new foods, the children become oppositional and increasingly anxious and scared to try new foods. This pattern of forcing children to eat certain foods without adequate support to manage their anxiety may be a risk factor in sensory food aversions outcome. In support of this, longitudinal research by Nicholls and colleagues has shown that selective eating is associated with anxiety and obsessive-compulsive symptoms, social limitations, and school difficulties.

A longitudinal study of children's food preferences by Skinner and colleagues, who followed children from 2 to 3 years to 8 years of age, found that the number of foods liked at 8 years was predicted by the number liked at 4 years of age. Children were more likely to accept newly tasted foods between 2 and 4 years of age than during school-age years. School-age children with sensory food aversions may experience social anxiety and try to avoid social activities (parties, sleepovers, summer camp) that involve eating. Parents of children with sensory food aversions often report that they were unable to eat a variety of foods as young children and that they learned to eat new foods primarily during their adolescent or young adult years when they wanted to fit in with their peers.

Treatment. Research by Birch and colleagues indicated that repeated exposures to new foods enhance young infants' acceptance not only of that food but also of similar foods. However, toddlers required more than 10 exposures to the same new foods until acceptance was reached. In addition, it is very difficult to get toddlers to try new foods if they are reluctant to do so. Birch and colleagues demonstrated that coercive techniques, such as threatening children to have them remain seated at the table until they finished eating everything on their plate or depriving them of certain privileges unless they tried the new food, have a significant negative effect. On the other hand, toddlers are very responsive to modeling by their parents. They are more willing to try a new food if they observe their parents eating the food but not offering it to them.

No empirical studies have systematically examined how best to treat infants and young children with sensory food aversions, but clinical experience suggests that if infants or toddlers show strong aversive reactions to a new food, such as by spitting the food out or gagging or vomiting, the parents are best advised to give up on having the child eat this food. On the other hand, if the infant just grimaces, the parents may expose the infant to the new food at another time and pair a small amount of the aversive food with a preferred food. A gradual increase of the aversive food over several meals may allow the infant to get used to the new food.

Modeling eating new foods by the parents and waiting until the toddler asks to try a new food generally is more effective than putting new foods on the toddler’s plate and requesting that the toddler try them. If the parents stay neutral as to whether the toddler likes the requested food, the toddler remains neutral as well and does not appear to become afraid of trying new foods. Once children fear trying new foods, their diet becomes more and more limited, and, by 3 years of age, most young children are not swayed by what their parents eat. Occasionally, preschool children may be willing to try new foods in a preschool setting, but more often they become anxious in social situations and try to avoid eating with others. On the other hand, preschool children are rich in phantasies, and feeding the baby dolls with playfood can help to explore their fears and encourage them to first feed the baby doll and then eat themselves new foods.

John was brought for an assessment by the multidisciplinary feeding disorders team because he refused to eat vegetables, fruits, and meats. His parents had started to give him a multivitamin daily, but they were worried that he was missing other essential nutrients in his limited diet, and they were concerned that his articulation was so poor that nobody outside of the family could understand what he was saying. At the age of 2½ years, his daily diet consisted of Cheerios or pancakes in the morning and peanut butter sandwiches, some cheese (which had to be “American”), pasta, cake, crackers, ice cream, and milk during the rest of the day. His feeding problems started when he was introduced to baby food. He accepted baby food sweetened with fruit but grimaced when given baby food with vegetables and spat it out. He gagged when introduced to baby food with lumps of meat, and he refused to accept any baby food after that. At 9 months of age, John was offered soft table foods, which he ate only if he could finger-feed himself. He had another episode of gagging when introduced to green beans at 15 months of age, and consequently, he refused to touch any green vegetables and fruits. He never tried any meats, and he became very distressed when his parents put any food on his plate that he did not want. When his parents insisted that he should at least try one bite of a new food before he could leave the table, John cried throughout the whole meal and ate nothing, not even his preferred foods. The parents had strong disagreements about how to handle John’s food refusal. The father had similar food dislikes as a child and wanted to leave him alone. The mother was very concerned about John’s health because of his limited diet, and she tried everything—bribing, coaxing, distracting, and threatening—just to get John to try new foods. Both parents experienced mealtime as very stressful and felt that John was getting increasingly rigid about his food likes and dislikes.

The observation of feeding revealed intense conflict between John and his mother regarding the various foods that he refused to eat. The nutritional assessment showed that although John was taking in adequate calories, he was not getting enough zinc and iron from his diet and, without the multivitamin, his diet was deficient in vitamins. The oral motor assessment revealed that John used an immature chewing pattern that was felt to be due to his lack of practice with chewing crunchy and chewy foods. This oral motor delay was felt to account for his articulation problems.

The parents were instructed to obtain vitamins complete with zinc and iron to compensate for the nutrients missing in John's diet. They were also offered speech therapy for John to help with his articulation problems. In addition, the parents were given a few parent training sessions to help them emotionally neutralize mealtime. After the mother's anxiety about John's health was alleviated, she was able to offer John his preferred foods without attempting to get him to eat new foods. The parents were instructed to eat their regular diet in front of John without offering him any of their food and to give him only a small piece of their food if he were to ask for it. They were not to praise him if he liked the food nor become disappointed if he no longer wanted to have it. After she had stopped offering John any new food for about 2 months, John surprised his mother by asking her to have some of her peach. He liked it and asked for more on another day. Gradually, he began to eat soft fruits without the peel, and after several months, he tried chicken nuggets. His diet increased in variety very slowly, and when he entered preschool, he tried a number of foods at school that he had not touched at home. His diet became more complete. However, John was not able to eat the variety of foods the rest of his family enjoyed.

Avoidant/Restrictive Food Intake Disorder

Subtype: concern about aversive consequences of eating (DSM-5-TR)

Feeding Disorder Associated with Insults to the Gastrointestinal Tract (DC: 0–3R)

Post-Traumatic Feeding Disorder (Chatoor et al., 1988)

Clinical Features. This feeding disorder may present with total food refusal, refusal to swallow solid food, or refusal to drink from the bottle, depending on what mode of feeding the child associates with the traumatic event. The onset of the food refusal is frequently rather sudden and follows traumatic experiences that involve the oropharynx or gastrointestinal tract, such as severe gagging, choking, vomiting, insertion of feeding or endotracheal tubes, or force-feeding. It is estimated that post-traumatic feeding disorder may affect 4% of infants with gastroesophageal reflux who do not have a history of esophageal surgery and up to 40% of children who undergo esophageal surgery. Frequently, the parents are not aware that the event was so frightening to the child that it triggered the food refusal because other children who may undergo the same experience will not necessarily develop a post-traumatic feeding disorder. It appears that children who develop a post-traumatic feeding disorder are more prone to anxiety and/or are more sensitive to pain than the average child. Older children who refused to eat solid food after an incident of gagging or choking reported that they were afraid that the food would get stuck in their throat and choke them to death. Infants and young children express this fear through their behavior by crying in anticipation of being fed or when seeing the high chair, the bottle, or the spoon. Some infants or children are brave enough to put the food in their mouths but cannot bring themselves to swallow it. They keep the food in their cheeks and/or spit it out. In severe cases, the infants refuse to feed altogether. These infants are in acute danger of dehydration because the fear of eating appears to override any awareness of hunger or thirst.

Often the food refusal of these children is so intense that the parents react with severe anxiety to their child’s food refusal. They worry about their child’s nutrition. Frequently, they try to coax and distract the child, they offer various types of food, and they try to feed the child day and night without success. However, some very young infants who are afraid to drink from the bottle when they are awake may drink when they are asleep and not aware of what they are doing. However, if they wake up and see the bottle, they usually push it away and cry. In general, the food refusal of these infants poses an acute threat to their health, and they require acute intervention.

Differential Diagnosis. This feeding disorder needs to be differentiated from IA and from food refusal because of sensory food aversions. IA is characterized by an inconsistent pattern of food refusal, depending on the mood of the child. Anorectic children are not afraid to eat and can eat all types of food if they want to eat. Sensory food aversions usually involve foods with a certain taste, texture, temperature, and/or smell. Usually, the food refusal is more selective and not as global as seen in a post-traumatic feeding disorder. However, sometimes children with sensory food aversions develop a post-traumatic feeding disorder if they experience a strong aversive reaction or if they are force-fed aversive foods that trigger gagging and vomiting.

Etiology and Course. The observation of infants with a post-traumatic feeding disorder reveals that their food refusal seems to be caused by fear. However, no empirical studies have been done to explain why only some infants with distressing experiences with the oropharynx react with a post-traumatic feeding disorder. There are also no systematic longitudinal data on the course of this feeding disorder. However, individual case studies indicate that many infants and children get locked into their fears of eating. Some drink milk from the bottle and eat only pureed food until they reach school age, when the social embarrassment of their eating behavior causes the parents to seek help. In extreme cases, when the infants refuse all food, gastrostomy feedings have to be implemented. To survive, some of these children may depend on gastrostomy feedings for years.

A recent study by Lucarelli and colleagues showed that children with a post-traumatic feeding disorder scored in the clinical range of the aggressive behavior on the Child Behavior Checklist, which is mainly characterized by angry moods, temper tantrums, stubbornness, child screams a lot, easily frustrated, and wants a lot of attention. Taking into account the history of traumatic feeding experiences, the emotional dysregulation in these children may further contribute to their ongoing feeding problems. Moreover, this study found that their mothers had significantly higher scores in anxiety symptoms, and some of them received a diagnosis of anxiety disorder. These data suggest reciprocal influences between the dysregulation in emotional functioning of these children and the anxiety symptoms of their mothers. Actually, maternal anxiety may compromise their parenting capacity to moderate the impact of the child’s traumatic experience and consequent emotional dysregulation.

Treatment. Because of the complexity of many of these cases, particularly those that result from trauma inflicted by unavoidable medical procedures (e.g., intubation, suctioning), a multidisciplinary team consisting of a pediatrician or gastroenterologist, nutritionist, occupational therapist, and a psychiatrist or psychologist is best equipped to meet the needs of these infants. In cases of total food refusal, the infant is in acute medical danger of dehydration and malnutrition, and intravenous fluids need to be given. It is important to assess whether the infant might be coaxed into accepting any form of food or whether nasogastric tube feedings need to be instituted. Unfortunately, the insertion of feeding tubes can intensify the fear of feeding, and, if this is the case, the infant is better served by a gastrostomy tube. In any case, the infant’s nutritional requirements need to be addressed prior to the implementation of the psychiatric treatment.

The psychiatric treatment of patients who exhibit total or partial food refusal centers on the desensitization of the infant to the fear of eating. Outpatient treatment can only be carried out if the infant accepts bottle feedings or receives adequate nutrition through a nasogastric or gastrostomy tube. It is best if the child is given the tube feedings at night and during the day there are scheduled oral feedings at 3- to 4-hour intervals with no feedings in between. If the child feeds from the bottle, the bottle feedings should follow the oral feedings in the high chair, and there should be no bottle feedings at other times, to make the child hungry at mealtimes. Behavioral techniques of positive reinforcement, praise and social interactions of food acceptance, and preferred toys or favorite videos for distraction are instituted in the initial management to overcome the child’s fear of eating. Encouragement of self-feeding is frequently helpful for these infants to gain mastery over the anticipatory anxiety of eating. When the infant begins to accept food from the spoon, it is important to proceed slowly when advancing the texture of food to avoid any incident of gagging that may set back the infant. The behavioral manipulation of the infant’s eating frequently leads to external regulation of eating in response to the reinforcers. Once the infant has overcome the fear of eating, it is important to phase out the external reinforcers and to encourage eating according to internal signals of hunger and satiety. The same principles described with regard to facilitating internal regulation of eating in children with IA should also be applied in this final stage of the treatment.

Inpatient treatment becomes necessary if the child is unable to overcome the fear of eating and remains dependent on bottle or tube feedings. Inpatient treatment involves two major components: appetite manipulation and contingency management. For children who have feeding tubes, manipulating calories and ensuring hydration through water or Pedialyte are used to allow the child to become hungry. Positive and negative reinforcements for food acceptance and food refusal are the behavioral interventions most commonly used. However, these techniques need to be adjusted to the individual child and require a skilled team to help the child to overcome the resistance to feeding and to learn to eat in response to hunger and fullness. Counseling and psychological intervention for parents are necessary to help them cope with the child’s traumatic and stressful experiences and support their parenting capacities during the interventions for their child.

Robert was 7 months old when he was referred by a gastroenterologist for an urgent psychiatric evaluation because of severe crying when offered the bottle and refusal to drink unless he was almost fully asleep and seemed unaware of his feeding. The mother reported that Robert was an 8-lb, full-term baby who had a vigorous suck and was feeding well from birth. However, after a few weeks he started to spit up during and after feedings, and by 3 months he developed projectile vomiting. In spite of the vomiting, he gained weight until he was about 4 months of age when he was referred to a gastroenterologist and diagnosed with gastroesophageal reflux. Several medications were tried, but the vomiting continued, and by 6 months of age Robert started to arch himself during feedings as if he were in pain. He would cry and refuse to continue with the feeding. Soon afterward, the mother noticed that he would start to cry at the sight of the bottle, and he refused to accept the bottle totally, as long as he was awake. To stave off starvation, the mother relied on feeding him during the night when he was asleep. His food intake greatly diminished, and he started to lose weight. The mother tried to introduce solid food, but Robert seemed to have difficulty moving the food back to his throat and spat most of it out. At the time Robert was referred, his mother was exhausted and frightened because of Robert’s continued weight loss. A recent examination by the gastroenterologist had shown that the reflux was markedly improved and that Robert’s ongoing feeding difficulties could not be explained by his medical condition.

The observation of mother and infant during feeding revealed that Robert became distressed the moment his mother positioned him for feeding, and he started to cry and arch himself as soon as he saw the bottle. His mother put the bottle away and tried to comfort him, but it took more than 10 minutes until he could settle in her arms. When the mother got the bottle and tried to feed him again, he cried even louder and would not be calmed until the bottle was out of his sight. Because it was obvious that Robert was too frightened to feed from the bottle, the mother tried to feed him in his infant seat with the spoon, but he batted the spoon away and did not want to have anything put in his mouth.

As the first step of treatment, the parents were made aware that infants as young as Robert can make associations between pain, vomiting, or other distress and feeding and consequently become frightened of feedings. The parents were instructed to get Robert on a schedule of regular naps and to feed him from the bottle only when he appeared almost asleep to prevent any further activation of his fears while looking at the bottle. After a successful attempt at the hospital to feed Robert in the high chair with a spoon in his hand and lots of encouragement to put it in his mouth, the mother was able to use another spoon and put some food in Robert's mouth. Consequently, she was encouraged not to focus on the amount of food she was able to get into him, but to look at the feedings as practicing sessions to help Robert overcome his fear of feeding and to develop oral motor skills that would allow him to handle solid food.

In the first 2 weeks, Robert would cry and refuse to continue feeding if he woke up and saw the bottle. However, gradually he seemed to realize that it was safe to drink from the bottle. The mother reported that after 4 weeks, she could even start bottle-feeding when Robert was still awake. Robert also progressed quickly from accepting very little food from the tip of the spoon to eating several spoonfuls. After 3 weeks, Robert was finishing a whole baby jar at one meal. When Robert was 9 months old, the mother called to report that Robert was feeding better and starting to gain weight. At 10 months of age, Robert was taking the bottle when awake, and he had learned to use the spoon and feed himself.

Avoidant/Restrictive Food Intake Disorder (DSM-5-TR)

Subtype: Very young infants may present as being too sleepy, distressed, or agitated to feed

Feeding Disorder of State Regulation (DC: 0–3R)

Clinical Features. This feeding disorder begins in the postnatal period and is characterized by irregular, poor feedings and inadequate food intake by the infant. Infants with this feeding disorder exhibit difficulties in state regulation, which interfere with their ability to feed effectively. They have difficulty reaching and maintaining a state of calm alertness necessary for feeding. Some are too irritable and cry excessively and cannot calm themselves for feeding. Others are too sleepy and cannot wake up or stay awake long enough to feed adequately. The parent may be anxious or depressed or may present with more severe psychopathology. Mother-infant interactions during feeding are characterized by irritability and/or sleepiness of the infant, maternal tension, and poor engagement between mother and infant.

Feeding disorder of state regulation can be seen in infants without any organic problems, but it is frequently associated with prematurity or dysmaturity and gastrointestinal, cardiac, or pulmonary disease.

Differential Diagnosis. This feeding disorder needs to be differentiated from pure organic disorders that can interfere with the regulation of the state of the infant. As described earlier, organic problems can trigger the feeding disorder and should always be addressed first. The diagnosis of feeding disorder of state regulation should be considered if treatment of the organic problem—for example, cardiac or pulmonary disease—cannot resolve the infant's difficulty in state regulation.

Etiology and Course. There are no empirical studies that have explained why some infants have more difficulty in regulation of state than others. However, the foundation for the regulation of feedings, which is so closely linked to the regulation of state, is further developed in the first months of life. Infants with feeding problems during the early months trigger anxiety in their mothers and frequently continue to have difficulty in self-regulation of their food intake during the transition to self-feeding in the second year of life.

In a longitudinal study from Sweden, 70% of infants with feeding difficulties and food refusal in the first year of life were observed to have eating problems during childhood.

Treatment. Treatment needs to be individualized and can be directed toward the infant, toward the mother, and toward the mother-infant interaction. If the infant is too sleepy to feed and difficult to arouse, gentle infant massage to stimulate the baby may help to improve the infant's alertness and facilitate better feeding. If the infant is too irritable and fidgety to feed, reducing stimulation before and during feedings by taking the child to the quietest room in the house and swaddling the baby during feedings may help to settle the infant and facilitate better feedings. However, if the infant tires easily and is unable to take in adequate calories to grow in spite of all efforts to improve the feedings, nasogastric tube feedings may have to be considered to supplement oral feedings. On the other hand, if the mother is exhausted, overly anxious, or depressed, her difficulties have to be addressed to enable her to be more effective with her infant.

Videotaping the feedings and observing the tape together with the mother is frequently helpful to heighten the mother's awareness of the infant's cues and to explore alternative methods that she can use to facilitate the infant's regulation of state and ability to feed.

Thomas was 3 months old when he was referred for a psychiatric evaluation because of his feeding difficulties and poor weight gain since birth. His parents were college educated, and both had pursued their professional careers until Thomas was born. Although Thomas was full-term and weighed 7 lb at birth, he had difficulty drinking from the breast. When he was 4 weeks old, his mother reluctantly switched him to bottle feedings because he was losing weight. Although his intake improved somewhat on bottle feedings, he gained weight very slowly and was still less than 8 lb at 3 months of age. His mother appeared tired and described how Thomas would drink only 1, 2, or 3 oz at a time, wiggle and cry, and refuse to continue with the feeding. After a few hours, however, he might cry again as if he were hungry. However, she could not settle him to feed, and he would continue to cry inconsolably. The mother described that she would attempt to feed him on an average of 10 to 15 times in a 24-hour period, that Thomas would cry a lot during the day and at night, and that everybody in the family was getting very little sleep.

The observation of mother–infant interactions during feeding and play revealed that Thomas was a very alert and wiggly baby who had difficulty settling in his mother’s arms. While drinking from the bottle he would kick his feet and move around with his arms, and soon the nipple of the bottle would slip out of his mouth. This upset him, and he started crying. His mother appeared anxious and tried to restart him by changing his position in various ways, but this only agitated him more. After repeated unsuccessful attempts to continue the feeding, mother and baby appeared exhausted, and the mother gave up.

The assessment revealed that Thomas was a very active and excitable baby who had difficulty keeping calm during feedings. After reviewing the videotape with the mother, the therapist explored ways in which the mother could better facilitate calming during feedings. Using a quiet corner in the house, swaddling Thomas in a blanket, and singing to him before starting the feeding were the most useful suggestions. Thomas stayed calm during feedings, was able to drink larger amounts of milk, and waited longer between feedings. This, in turn, relieved the mother's anxiety and helped both to have calmer interactions.

Avoidant/Restrictive Food Intake Disorder (DSM-5-TR)

Subtype: Infants and young children may not engage with the primary caregiver during feeding.

Feeding Disorder of Caregiver–Infant Reciprocity (DC: 0–3R)

Clinical Features. This feeding disorder is characterized by a lack of engagement between mother and infant, leading to inadequate food intake and growth failure of the infant. The onset of growth failure is usually between 2 and 8 months of age. The infant shows a lack of age-appropriate social responsivity—for example, lack of smiling response, lack of vocal reciprocity, lack of anticipatory reaching out when about to be picked up, and lack of molding and cuddling when held. The mother frequently presents with acute or chronic depression, drug or alcohol abuse, and/or high psychosocial stress, which appear to interfere with her care of the infant. Mother–infant interactions are characterized by a lack of mutual engagement and a lack of pleasure in their relationship. The mother frequently appears detached and noncontingent in her responses to the infant's cues.

Differential Diagnosis. This feeding disorder needs to be differentiated from organic conditions that lead to a lack of weight gain and weakness in the infant. However, the mother and infant usually show better mutual engagement and the infant responds more readily to the examiner.

Etiology and Course. This feeding disorder has been referred to in the literature as "maternal deprivation" or "deprivation syndrome." It has been postulated that lack of emotional nurture and infrequent feedings lead to growth failure and developmental delays in these infants. Characteristically, infants with this feeding disorder become engaged and gain weight when admitted to the hospital and are in the care of a nurturing nurse.

This feeding disorder has also been referred to as “nonorganic failure to thrive,” and in the DSM-III, this disorder was defined as “Reactive Attachment Disorder of Infancy” associated with FTT. However, in the DSM-IV, the same diagnosis was changed to encompass only problems in relatedness without growth failure. However, several studies of infants with the diagnosis of “nonorganic failure to thrive” have demonstrated very high rates of insecure attachment, ranging from 50% to 90% of the infants studied. However, not all infants with insecure attachment are failing to thrive. It appears that only a severe degree of a disturbance in the attachment process of the infant to the caretaker, or what some authors describe as “nonattachment,” results in the severe disturbances in relatedness and growth failure seen in infants with this feeding disorder.

Children with a feeding disorder of caregiver–infant reciprocity are at high risk for delayed cognitive and disturbed emotional development. When such children are followed up to school age, deficits in cognitive performance, behavioral organization, ego control, and ego resiliency and behavioral symptoms have been reported.

Treatment. Depending on the severity of the growth deficiency and the evidence of neglect, various interventions have been suggested. In cases of no evidence of deprivational behavior on the part of the mother, or if the parents have sought medical care in the past and have some support in the extended family or the community, an outpatient approach may be safe. However, if the growth deficiency is more severe, if there is serious hygiene neglect, if the mother is abusing drugs or alcohol, or if the mother lives a chaotic lifestyle, immediate hospitalization is indicated. During hospitalization, it is critical to assign a warm and nurturing nurse to engage the infant during feedings. While the nutritional, emotional, and developmental rehabilitation of the infant takes place, a more in-depth evaluation of the mother and the mother–infant relationship needs to be addressed as well.

Many of the mothers have experienced unsatisfactory relationships with their caretakers when they were growing up, and consequently, they are distrustful and avoidant of professionals. It is important to identify any positive behavior the infant shows toward the mother to see whether the mother has any potential to engage in a mutually satisfying relationship with her infant. In addition, the mother's support system and her ability to become engaged with a therapist need to be explored prior to returning the infant to her care. In some situations of severe neglect and associated abuse, Protective Services need to be involved, and the infant has to be placed in foster care.

Discharge from the hospital is a critical time when all services need to be in place to ensure the follow-through of the treatment plan. The treatment plan should be individualized and may range from home visits by a health care professional to daycare for the infant, to mother-infant psychotherapy, and to family therapy for the parents.

Anna was 8 months old when she was admitted to the hospital through the emergency room because of a respiratory tract infection and severe “failure to thrive.” After she had recovered from the infection and a medical workup for her malnutrition was found to be negative, a psychiatric consultation was requested. After repeated unsuccessful attempts, the psychiatrist finally reached the parents and set up an appointment to obtain more history and to observe the mother and infant during feeding. The parents explained that the father had difficulty obtaining time off from his construction job and that the mother did not have any means of transportation to the hospital. The couple was not married because the father had not gotten his divorce from his first wife, but they lived together with Anna in one room in a house that they shared with nine other members of the father’s extended family.

The mother, who became pregnant with Anna during her last year in high school, described Anna as a quiet baby and denied any feeding problems. However, when the mother temporarily left the room, the father turned to the therapist and complained that the mother did not give enough attention to Anna and that she would talk on the phone or read from a magazine while feeding the baby.

The observation of the mother and infant during feeding revealed that the mother looked sad and preoccupied while she held the baby loosely on her lap. The infant lay in a “surrender position” with her arms raised beside her head, and she looked away while drinking from the bottle. However, when she looked at her mother, the mother looked away as if she could not stand to have eye contact with her infant.

Further evaluation of the mother revealed that she felt very lonely and depressed. She was unable to finish high school because of the pregnancy with Anna. Her family had abandoned her for becoming involved with a married man. She described the father’s family as hostile toward her. The baby’s father worked long hours and was hardly around. The telephone calls the mother made to her girlfriends from high school were the only relief in her unhappy life.

During the hospitalization, Anna gained weight, became more engaged, and rewarded her favorite nurse with frequent smiles. After the social worker had arranged for transportation for the mother, she came regularly for her appointments with the social worker and spent more time with Anna. While nurturing the mother and trying to help her work on her situation with her family and the father's family, the social worker also spent time with mother and infant during play. She would point out how the baby was making progress, how she was trying to get her mother's attention, and how she had that special smile for her mother.

After 4 weeks in the hospital, Anna seemed stronger, and her mother seemed committed to continuing working with the social worker. The hospital staff felt comfortable about discharging Anna. After Anna had returned home with her mother, a visiting nurse came to the home twice a week to check on the mother and infant, and the mother visited weekly with Anna at the hospital to see the social worker. After 6 months, the visiting nurse came less frequently, but the mother and Anna stayed in treatment with the social worker for another year, when both seemed engaged and comfortable with each other.

Avoidant/Restrictive Food Intake Disorder

ARFID can be diagnosed if the feeding disturbance may persist after being triggered by a medical condition and following resolution of the medical condition.

Feeding Disorder Associated with Concurrent Medical Condition (DC: 0–3R)

Clinical Features. Although respiratory problems or cardiac diseases that may tire out the infant during feeding are easily recognized, some medical conditions are not readily diagnosed, and food refusal may be the leading symptom. For example, food allergies can be difficult to diagnose in infants, and “silent” gastroesophageal reflux without vomiting can be overlooked because vomiting is usually the leading symptom that alerts pediatricians to this condition. Typically, infants with reflux can drink 1 to 2 oz until the reflux becomes activated. However, once reflux occurs, the infants experience distress and may arch, cry, and refuse to continue feeding. Some infants can be calmed and resume feeding, but others may become increasingly agitated while their caretakers continue to offer the bottle or food. Infants with respiratory distress may feed for a while and take a few ounces until they tire and stop feeding. In general, the feeding difficulties of these infants lead to inadequate food intake, failure to gain weight, or loss of weight.

Differential Diagnosis. This feeding disorder needs to be differentiated from other feeding disorders characterized by food refusal, such as IA, sensory food aversions, and post-traumatic feeding disorder. What helps in understanding this feeding disorder is the observation of feeding, which reveals that the infant usually initiates feeding without difficulty and that mother and child engage in pleasant reciprocal interactions until the infant experiences distress and stops feeding.

Etiology and Course. FTT of infants and young children was initially believed to be caused by either organic or nonorganic causes (e.g., maternal deprivation). In 1981, Homer and Ludwig described a third category of FTT, which is caused by a combination of various organic and nonorganic problems. Since then, it has been accepted that organic conditions can be complicated by psychological difficulties and lead to severe feeding problems and growth failure. However, although many young infants suffer from gastroesophageal reflux, only less than one-third of them are reported to present with feeding difficulties and growth failure. Clinically, these infants appear particularly alert and sensitive. However, no empirical studies have been done to explore why some infants develop severe reactions to their reflux, whereas others hardly seem bothered by it.

Only a few studies have been reported on the follow-up of these infants. Fortunately, with therapy, up to 80% of children with gastroesophageal reflux improve within 12 months. For those with more serious feeding difficulties and the inability to take in adequate amounts of food for growth, modern technology allows for gastric tube feedings over a long time to safeguard their survival. However, the regulation of oral feedings is often very difficult, and even after the medical condition has been successfully treated, it may take time until the child learns to eat independently.

Treatment. Because of the interaction of organic and psychological factors contributing to the severe feeding difficulties of these children, a collaboration between the pediatrician or pediatric specialist and the psychiatrist is critical. Optimal medical treatment of the child’s illness is necessary before psychological interventions can be successful. Direct observation of infants with their primary caregiver during feeding is most helpful in monitoring how well the medical condition is responding to the treatment. In addition, the observations during feeding can guide the therapist in helping the parents to develop more appropriate strategies that allow infants to calm down instead of escalating their distress during feeding. However, in situations in which the medical illness cannot be adequately treated, and the infant continues to experience distress during feedings, supplemental nutrition through nasogastric or gastrostomy tubes must be considered. Then the therapist must work with the parents to maintain the infant’s oral feeding skills while most of the nutrition is given via tube feedings. The parents need to learn to feed the infant to the point of discomfort but not beyond it. Reviewing the videotape of the infant’s feeding and discussing the infant’s cues with the parents is usually helpful in sorting out how to best feed the infant. In general, these are very difficult cases that require individualized attention by an experienced multidisciplinary team.

Sarah was 6 months old when she was referred by her pediatrician to the Multidisciplinary Feeding Disorders Team because of severe food refusal and poor growth. The mother reported that since birth Sarah occasionally “spat up” a little during and after feedings, but during the last 6 weeks, she had become increasingly difficult to feed. Usually, she took 1 or 2 oz of her formula and then started to arch and push the nipple out from her mouth. If her mother tried to continue feeding Sarah, she cried and became increasingly agitated until her mother picked her up and comforted her by holding her on her mother’s shoulder and walking her. Sometimes after 10 to 20 minutes, her mother could resume feeding Sarah, but often, Sarah started to cry again after suckling from the bottle for a few minutes only. In spite of her mother’s efforts to feed Sarah every 2 hours, Sarah drank only half of the formula that she was supposed to take in for her age, and in recent weeks, she had started to lose weight.

The observation of feeding revealed pleasant mother-infant interactions at the beginning of the feeding, but after a few minutes of suckling, Sarah pushed the nipple of the bottle out of her mouth, arched her back, and started to cry when her mother tried to get her back to feeding. The harder her mother tried, the more agitated Sarah became, and mother and baby looked very distressed and gave up feeding. These feeding interactions raised concerns that Sarah experienced gastroesophageal reflux that caused her pain and interfered with her feeding. Further medical testing confirmed the diagnosis of moderate gastroesophageal reflux, which was somewhat surprising to the pediatrician and the gastroenterologist because Sarah had not shown any vomiting, the most common symptom of reflux.

Sarah was treated with medications for reflux, and her food intake doubled quickly. However, if she drank more than 3 or 4 oz at a time, the symptoms recurred. However, once the mother understood the reason for Sarah’s food refusal, she learned to ease Sarah’s feeding difficulties by thickening the formula and feeding her smaller amounts at more frequent intervals.

Gradually, Sarah tolerated larger amounts of formula at each feeding, and at 15 months of age, she was found to have outgrown the gastroesophageal reflux.

Rumination Disorder

Clinical Features. Rumination has been defined as repetitive regurgitation of gastric contents that are subsequently rechewed and then reswallowed or spit out. The regurgitation occurs very soon after a meal and tends to persist for 1 to 2 hours. The sequence of events leading to regurgitation has been documented by using combined multichannel intraluminal impedance and manometry. These events include an increase in intra-abdominal pressure identified in the stomach, which initiates the reflux, followed by peristaltic fore contractions, which clear the intraesophageal content. Rumination has been described in infants, individuals with developmental disabilities, and in children, adolescents, and adults with psychiatric or neurologic disorders. The major change for rumination disorder in the *Diagnostic and Statistical Manual for Mental Disorders,* Fifth Edition, was its relocation from the section titled “Disorders Usually First Diagnosed in Infancy, Childhood, or Adolescence” to the new section, “Feeding and Eating Disorders.” This change emphasizes that this disorder occurs across the age range, including adulthood. Minor changes have been introduced to the DSM-5 and DSM-5-TR phrasing of diagnostic criteria for rumination disorder to improve clinical utility by ensuring applicability across the age range and removing some ambiguity inherent in the DSM-IV phrasing.

Some infants can be observed to put the thumb or the whole hand in the mouth, suck the tongue rhythmically, or arch the back to initiate regurgitation. However, some infants ruminate when left alone after they have been placed in the crib to sleep. They are frequently found in a puddle of vomitus, which is then falsely attributed to some organic problem. The latter type of rumination is most commonly observed in infants who have received little emotional stimulation and who have learned to stimulate and soothe themselves through rumination. Although initially most of the regurgitated food is vomited, some infants gradually learn to hold more of the food in their mouth, to rechew, and to reswallow it. “Experienced” ruminators are able to bring up the food through tongue movements and reswallow it without losing any of it. Their rumination can only be discovered by observing the movements of their cheeks and the foul oral odor. Social functioning may be significantly impaired due to the recognized social undesirability of this behavior.

Differential Diagnosis. The term *regurgitation* implies that the behavior is volitional, distinguishing it from vomiting or gastroesophageal reflux. The regurgitation is not associated with nausea and is not explained by an illness or medical condition. Individuals may cough, engage in the tongue or abdominal muscle contractions or movements, or place fingers in the mouth to facilitate food being brought back up. Because, in young children, reflux and rumination frequently coexist, it is important to observe the infant in different situations to see the initiation and the rhythmic oral movements of the ruminatory activity. As in pica, the diagnosis of rumination disorder should be given in the context of another mental disorder only if the symptoms require additional clinical attention.

Epidemiology. Rumination is believed to be a rare disorder. It appears to occur more often in males than in females and also in individuals with mental retardation. There are a few reports of children, adolescents, and adults who were diagnosed with rumination disorder, and some had associated eating disorders.

Etiology and Course. Various mechanisms of etiology have been proposed. Several authors have attributed rumination to separations or a disturbed mother-infant relationship—for example, lack of responsiveness or neglect by the mother secondary to stressful family relations or life events. Others have considered rumination as a symptom of gastroesophageal reflux. Several authors have postulated that rumination is a learned behavior that is maintained with special attention to the regurgitation. In more recent years, opiate receptor insensitivity or reduced endorphinergic transmission has been implicated in rumination.

A biopsychosocial model has been proposed by which rumination is seen along a continuum in which an infant may have gastrointestinal pathology and little psychopathology at one end of the spectrum, or the opposite, in which an infant may have no organic pathology and severe psychopathology in the mother-infant relationship at the other end of the spectrum. Frequently, vomiting secondary to gastroesophageal reflux or vomiting associated with an acute illness precedes the beginning of rumination. It appears that at some point the infant learns to initiate vomiting and to rechew the food to achieve relief of tension, to self-soothe, and/or to self-stimulate. It seems that in situations in which the infant fails to elicit or loses either caring attention or tension-relieving responses from the mother, rumination becomes a means of self-regulation. Once the infant has experienced rumination as a means of self-soothing or self-stimulation, the rumination develops into a habit that is difficult to break.

The onset of rumination is frequently in infancy, most commonly in the first year of life. However, rumination may start at any point in life. It is believed that in some individuals the disorder remits spontaneously. However, some infants suffer serious complications, such as electrolyte imbalance, dehydration, and weight loss, and in earlier reports mortality rates as high as 25% have been described.

Treatment. Depending on the theories of the etiology of rumination, different treatments have been suggested. These range from mechanical restraints as described in the early literature and surgical interventions (Nissen fundoplication) to prevent reflux, to behavioral and psychodynamic treatments.

Based on the assumption that rumination is a learned habit that is reinforced by the increased attention that is given to the regurgitation, unlearning by behavioral techniques has been suggested. These treatments include the use of aversive taste stimuli (lemon juice or hot sauce), oral hygiene, differential reinforcement of incompatible behaviors, and electric shock. Proponents of the assumption that rumination is based on an unsatisfactory mother-infant relationship have suggested a psychodynamic approach. A mother substitute who feeds the infant while the mother is helped through psychotherapy has been proposed.

After a thorough evaluation that should include the assessment of the infant and the mother, the therapist needs to individualize the treatment. In severe situations, if the infant is malnourished and continues to lose most of the food through rumination, a jejunal tube may have to be inserted before any psychosocial treatment can begin. Frequently, a combination of psychodynamic and behavioral interventions to enhance the mother–infant relationship in general and to treat the symptom of rumination, in particular, is most effective.

Justin was 9 months old when he was referred by a gastroenterologist for a psychiatric evaluation because of concerns that he continued to vomit because of rumination. Justin was born full-term and had developed nicely until he was about 6 weeks old when he began to vomit increasing amounts of milk during and after feedings. He was diagnosed with gastroesophageal reflux, which was treated with thickened feedings and medication. Justin responded well to the treatment; he stopped vomiting almost completely and gained weight adequately. Because Justin was doing so well, his mother decided to go back to work when Justin was 8 months old. She transitioned his care to a young woman who would come to the house during the mother's working hours. Justin started to vomit soon after his mother left the house. The vomiting seemed to increase from day to day in frequency and in intensity, and after 2 weeks of the mother's return to work, Justin vomited several times daily and was losing weight. He was seen by a gastroenterologist, and during the barium swallow, it was noticed that Justin put his hand in his mouth, which seemed to trigger vomiting. Justin was put back on medication for gastroesophageal reflux, but he continued to vomit with increasing frequency, which led to a psychiatric consultation.

Observation of mother and infant during feeding revealed that as soon as Justin finished feeding, he put his hand in his mouth and vomited. When his mother restricted his hand, Justin moved his tongue back and forth in a rhythmic manner until he vomited again. This happened repeatedly, and Justin continued the rhythmic tongue movements even when he could not bring up any more milk.

Because of his poor nutritional state and moderate dehydration, Justin was admitted to the hospital, and a nasojejunal tube was inserted for feedings. When Justin was awake, a special nurse or the parents played with him and tried to distract him whenever he attempted to put his hand in his mouth or thrust his tongue rhythmically. Justin became increasingly engaged, and his ruminatory activity decreased accordingly. After 1 week in the hospital, small feedings were started, but Justin tried to ruminate again, and the oral feedings had to be stopped. At this point, the mother decided to stop working and take Justin home to continue the treatment there. The mother started small feedings, played with Justin after feedings, and was able to keep him from ruminating. After 4 weeks of slow increments in his feedings, Justin was able to take all of his feedings by mouth without ruminating, and the nasojejunal tube could be removed.

Pica

Clinical Features. Young children with this disorder typically eat plaster, paper, paint, cloth, hair, insects, animal droppings, sand, pebbles, and dirt. Many of the children engage in other oral activities, such as thumb sucking or nail biting, which they seem to use for relief of tension and for self-soothing. In some younger patients, as well as those with neurodevelopmental or learning disorders, some clinicians regard pica as a form of self-soothing behavior. The pica may lead to anemia, diarrhea/constipation, worm infestation, toxoplasmosis, lead poisoning, and malnutrition. In some cases, it may lead to intestinal obstruction as the result of hairball tumors. In addition to the heterogeneity of substances consumed, pica's associated behavioral features can differ considerably among individuals (e.g., its compulsive nature and relationship to emotional arousal or need for oral stimulation).

The major change for pica disorder in the *Diagnostic and Statistical Manual for Mental Disorders*, Fifth Edition, was its relocation from the section titled "Disorders Usually First Diagnosed in Infancy, Childhood, or Adolescence" to the new section "Feeding and Eating Disorders." This change emphasizes that this disorder occurs across the age range, including adulthood. In particular, changes to the phrasing of diagnostic criteria in DSM-5 are intended to guide clinicians in distinguishing eating behaviors that warrant a diagnosis of pica from behaviors that are developmentally normal, culturally supported, or socially normative or that support a diagnosis of a different mental disorder.

Differential Diagnosis. Before the age of 18 to 24 months, mouthing and sometimes eating non-nutritive substances can be relatively common and should only be diagnosed as pica if it is persistent and inappropriate for the developmental level of the child. In DSM-5-TR, 2 years is suggested as the minimum age for a diagnosis of pica; at younger ages, mouthing different kinds of objects is considered developmentally normal. Consumption of substances consistent with local cultural practices or social norms would not warrant a pica diagnosis. Pica may be comorbid with a number of mental disorders, which are more likely to come to clinical attention than pica itself, (e.g., intellectual development disorder, autism spectrum disorder, schizophrenia, and obsessive-compulsive disorder). In addition, pica can co-occur with hair-pulling disorder and skin-picking disorder when hair or skin is ingested. When non-nutritive substances are ingested as a means to suppress appetite in anorexia nervosa, a pica diagnosis is not warranted. Some individuals with ARFID, particularly those whose selective or restricted intake is based on sensory aspects of food, may also present with comorbid pica.

Epidemiology. A survey of a large clinic population with a wide range of ethnic backgrounds revealed that 75% of 12-month-old infants and 15% of 2- to 3-year-old toddlers were reported by their parents to put non-nutritive substances in their mouths. However, the true incidence of pica is not known. Available data suggest that in the United States pica is rare in healthy children older than 2 years. However, a study reported that 33.9% of children in Detroit who were in treatment for sickle cell anemia had pica eating. Pica has also been reported as prevalent in children elsewhere, including in some school-age populations in Africa (as high as 77%). Among individuals with mental retardation, the prevalence of the disorder appears to increase with the severity of the retardation. The prevalence of pica among institutionalized individuals with mental retardation has been estimated to range from 10% to 33%.

Etiology and Course. Organic, psychodynamic, socioeconomic, and cultural factors have been implicated in the etiology of pica. Inadequate intake of iron and calcium has been shown in rats to induce pica and has been considered to lead to abnormal cravings and pica in humans. Low plasma levels of iron and zinc have been found to be associated with pica, and it has been suggested that hypozincemia with low iron levels may be a possible cause rather than a consequence of pica. Psychosocial factors—for example, poverty, maternal neglect and abuse, and disorganized family situations with inadequate nurturing and supervision of the children—have been implicated in the etiology of pica. The cultural acceptance of pica in certain population groups has also been considered to play a role in its etiology. Some authors have proposed a multifactorial model of etiology by which organic, familial, socioeconomic, and cultural factors interact with each other.

In most instances, the disorder is believed to last for several months and then remit. However, it may occasionally continue into adolescence or, less frequently, into adulthood. Several authors have pointed to the seriousness of the developmental effect of the disorder. The younger children showed delays in their speech and psychosocial development. Half of the adolescents were observed to suffer from depression and/or personality disorders, engaged in other forms of disturbed oral activities, such as thumb sucking and nail-biting, and abused tobacco, alcohol, or drugs. More recently, some authors pointed to a strong relationship between pica during early childhood and bulimia nervosa during the adolescent years.

Treatment. A variety of treatments have been proposed. These include physical restraints, aversive and nonaversive behavior therapy, the teaching of children to exchange inedible items for edibles, environmental enrichment, and individual and group therapy for the children and their mothers. It is important to conduct a comprehensive evaluation of the child, the caretakers, and the child’s environment to develop an individualized treatment plan. This should include helping the adults to become aware of the short- and long-term dangers of pica, providing a lead-free and childproof environment, facilitating a more satisfying parent–child relationship, and applying behavioral techniques to extinguish the maladaptive behavior of pica.

Susan was 3 years old when her mother took her to the pediatrician because of abdominal pain and lack of appetite. The mother complained that Susan put everything in her mouth but did not want to eat regular food. The pediatrician observed that Susan looked pale, thin, and withdrawn. She sucked her thumb and quietly looked down while her mother reported that Susan liked to chew on newspapers and put plaster in her mouth.

The medical examination revealed that Susan was anemic and suffered from lead poisoning. She was admitted to the hospital for treatment, and a psychiatric consultation was obtained.

Further exploration of the history and the observation of mother and child during feeding and play revealed that the mother was overwhelmed with having to care for three young children and had little affection for Susan. The mother was unmarried and lived with her three children and five other family members in a three-bedroom apartment in an old housing project. Her 4-year-old son was hyperactive and demanded almost constant supervision. Her 18-month-old infant was an engaging and active little girl, whereas Susan would sit quietly, rock herself, suck her thumb, or chew on newspaper.

The treatment plan included the involvement of social services and protective services to remove any lead paint from the walls in the apartment and to look for better living arrangements for the family. The mother was helped to enroll Susan and her brother in a preschool program that provided them with more structure and stimulation and gave the mother a few hours of relief every day. In addition, Susan was seen with her mother and her younger sister in play therapy to help the mother understand the different temperaments of the two children and to make her more responsive to Susan’s weak attempts to engage her. Once the mother felt more supported and less overwhelmed by her situation, she became more empathic and understanding of Susan. When Susan put something in her mouth, the mother was able to engage her in a play activity rather than scream at her and scold her for whatever she was doing. Over the period of 1 year, the relationship between Susan and her mother gradually improved, and Susan seemed less in need of putting her thumb or inedible things in her mouth.

To conclude our chapter, it is pivotal to underline that a problem arises when the children exhibit symptoms that appear to fit the diagnoses of more than one feeding disorders subtype (i.e., IA and post-traumatic feeding disorder, IA, and sensory food aversions). We recommend that the symptoms should be differentiated as much as possible to understand what the underlying diagnoses may be, because, as pointed out earlier in the treatment targeted to specific diagnostic subtypes, each feeding disorder requires a specific therapeutic approach, and the treatment needs to combine these individual approaches for each disorder to be successful.

FURTHER READINGS

Ammaniti M, Lucarelli L, Cimino S, D'Olimpio F, Chatoor I. Feeding disorders of infancy: a longitudinal study to middle childhood. *Int J Eat Disord*. 2012;45(2):272–280.

Ammaniti M, Lucarelli L, Cimino S, D'Olimpio F, Chatoor I. Maternal psychopathology and child risk factors in infantile anorexia. *Int J Eat Disord*. 2010;43(3):233–240.

Banerjee SP, Bhandari RP, Rosenberg DR. Use of low-dose selective serotonin reuptake inhibitors for severe, refractory choking phobia in childhood. *J Dev Behav Pediatr*. 2005;26(2):123–127.

Benton D. Role of parents in the determination of the food preferences of children and the development of obesity. *Int J Obes Relat Metab Disord*. 2004;28(7):858–869.

*Birch LL. Development of food preferences. *Annu Rev Nutr*. 1999;19:41–62.

Bourne L, Bryant-Waugh R, Cook J, Mandy W. Avoidant/Restrictive food intake disorder: a systematic scoping review of the current literature. *Psychiatry Res*. 2020; 288:112961.

Bryant-Waugh R, Markham L, Kreipe RE, Walsh BT. Feeding and eating disorders in childhood. *Int J Eat Disord*. 2010;43(2):98–111.

*Chatoor I. Diagnostic classification of feeding disorders. In: *Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood: Revised Edition (DC: 0–3R)*. Zero to Three Press; 2005.

Chatoor I. Feeding disorders in infants and toddlers: diagnosis and treatment. *Child Adolesc Psychiatr Clin N Am*. 2002;11(2):163–183.

*Chatoor I, Ammaniti M. A classification of feeding disorders in infancy and early childhood. In: Narrow WE, First MB, Sirovatka PJ, Regier DA, eds. *Age and Gender Considerations in Psychiatric Diagnosis: A Research Agenda for DSM-5*. American Psychiatric Association; 2007.

*Chatoor I, Conley C, Dickson L. Food-refusal after an incident of choking: a posttraumatic eating disorder. *J Am Acad Child Adolesc Psychiatry*. 1988;27(1):105–110.

Chatoor I, Ganiban J, Hirsch R, Borman-Spurrell E, Mrazek DA. Maternal characteristics and toddler temperament in infantile anorexia. *J Am Acad Child Adolesc Psychiatry*. 2000;39(6):743–751.

Chatoor I, Ganiban J., Surles J, Doussard-Roosevelt J. Physiological regulation in infantile anorexia: a pilot study. *J Am Acad Child Adolesc Psychiatry*. 2004; 43(8):1019–1025.

Chatoor I, Surles J, Ganiban J, et al. Failure to thrive and cognitive development in toddlers with infantile anorexia. *Pediatrics*. 2004;113(5):e440–e447.

Chial HJ, Camilleri M, Williams DE, Litzinger K, Perrault J. Rumination syndrome in children and adolescents: diagnosis, treatment, and prognosis. *Pediatrics*. 2003;111(1):158–162.

Cooke LJ, Haworth CM, Wardle J. Genetic and environmental influences on children's food neophobia. *Am J Clin Nutr*. 2007;86(2):428–433.

Green AD, Alioto A, Mousa H, Di Lorenzo C. Severe pediatric rumination syndrome: successful interdisciplinary inpatient management. *J Pediatr Gastroenterol Nutr*. 2011;52(4):414–418.

Gueron-Sela N, Atzaba-Poria N, Meiri G, Yerushalmi B. Maternal worries about child underweight mediate and moderate the relationship between child feeding disorders and mother-child feeding interactions. *J Pediatr Psychol.* 2011;28:1–10.

Hartmann AS, Becker AE, Hampton C, Bryant-Waugh R. Pica and rumination disorder in DSM-5. *Psychiat Ann*. 2012;42:426–430.

Kern L, Starosta K, Adelman BE. Reducing pica by teaching children to exchange inedible items for edibles. *Behav Modif*. 2006;30(2):135–158.

Linscheid TR. Behavioral treatments for pediatric feeding disorders. *Behav Modif*. 2006;30(1):6–23.

Lucarelli L, Cimino S, D'Olimpio F, Ammaniti M. Feeding disorders of early childhood: an empirical study of diagnostic subtypes. *Int J Eat Disord*. 2013;46(2):147–155.

Lucarelli L, Cimino S, Zavaroni D, Ammaniti M. Infantile post-traumatic feeding disorder and infantile anorexia: a clinical evaluation on maternal and child psychopathological risk factors. *Infant Ment Health J*. 2010;32(Suppl):7.

Lucarelli L, Sechi C, Cimino S, Chatoor I. Avoidant/restrictivefood intake disorder: a longitudinal study of malnutrition and psychopathological risk factors from 2 to

11 years of age. *Front Psychol*. 2018;9(1608):1–12.

Lucarelli L, Speranza AM. The relationship between maternal attachment/psychopathology and the severity of infantile anorexia. *Infant Ment Health J*. 2014;35(Suppl):106.

*Marchi M, Cohen P. Early childhood eating behaviors and adolescent eating disorders. *J Am Acad Child Adolesc Psychiatry*. 1990;29(1):112–117.

Nicholls D, Christie D, Randall L, Lask B. Selective eating: symptom, disorder or normal variant. *Clin Child Psychol Psychiatry*. 2001;6:257–270.

San'Anna AM, Hammes PS, Porporino M, et al. Use of cyproheptadine in young children with feeding difficulties and poor growth in a pediatric feeding program. *J Pediatr Gastroenterol Nutr*. 2014;59(5):674–678.

Singhi S, Ravishanker R, Singhi P, Nath R. Low plasma zinc and iron in pica. *Indian J Pediatr*. 2003;70(2):139–143.

Tutuian R, Castell DO. Rumination documented by using combined multichannel intraluminal impedance and manometry. *Clin Gastroenterol Hepatol*. 2004;2(4):340–343.

47

Tic Disorders

Rahil R. Jummani, M.D., Elizabeth A. Deckler, M.D., and Barbara J. Coffey, M.D., M.S.

The best known of the tic disorders is Tourette disorder (TD), also known as Tourette syndrome, named after the French neurologist Georges Gilles de la Tourette, who in 1885, was among the first to recognize and describe the disorder and its comorbid conditions. Since the descriptions of Gilles de la Tourette, a spectrum of tic disorders has come to be recognized and codified in psychiatric nomenclature. No longer is TD viewed as a rare, bizarre psychological entity with primarily disturbing manifestations, such as the shouting of obscenities, screeching, or barking. The comorbid psychiatric disorders that frequently occur in individuals with TD have become more widely recognized in recent years; it is these disorders that may often warrant treatment more than the tics themselves. Although the underlying etiologies of these disorders continue to be elusive, substantial progress has been made in the last few decades using innovative neuroimaging techniques, genomics studies, and clinical science. Tic disorders have emerged as complex, heterogeneous neurodevelopmental disorders that are the manifestation of an intricate interplay of genes and environment. Understanding of the etiology, genetics, neuroanatomy, and epidemiology of tic disorders has been significantly enhanced. Novel treatment approaches have also emerged, providing alternatives to the established pharmacologic management of tics. Specific behavioral interventions have been established as first-line treatments for tic disorders. Deep brain stimulation (DBS), botulinum toxin injection, targeted surgery, and stem cell transplantation are examples of promising new treatments. The future holds tremendous possibilities as new insights into the etiology, pathophysiology, and treatment of tic disorders are rapidly developing. This chapter provides a review and update of the history, comparative nosology, epidemiology, etiology, diagnosis and clinical features, pathology, differential diagnosis, course and prognosis, and clinical treatment of tic disorders, including TD.

DEFINITIONS

Tics are rapid, recurring motor movements (motor tics) or vocalizations (phonic or vocal tics) that are nonrhythmic, involuntary or semivoluntary, and sudden in onset. Tics typically involve one muscle or a group of muscles and are characterized by their anatomical location, number, frequency, duration, and complexity (Table 47–1). They can be classified as *simple* (involving one muscle, muscle group, or sound) or *complex* (slower, more purposeful movements involving coordinated muscle groups or multiple sounds). Simple motor tics cause a brief, jerking movement (*clonic tic*); a briefly sustained abnormal posture (*dystonic tic*); or an isometric contraction (*tonic tic*). Examples of simple motor clonic tics are eye blinking, shoulder shrugging, and head-turning. Dystonic tics include sustained eye closure (blepharospasm), ocular deviations, teeth grinding (bruxism), mouth opening, twisting and tilting of the neck (torticollis), and shoulder rotation. Tonic tics are typically manifested by tensing of abdominal or limb muscles. Complex motor tics may include touching objects, jumping, rotating, or imitating others' gestures (*echopraxia*). Examples of simple vocal tics are throat clearing, coughing, and sniffing. Repeating syllables, words, or one's own phrases (*palilalia*) or repeating others' words or phrases (*echolalia*) are examples of complex vocal tics. Less commonly occurring tics include the involuntary vocalization of obscenities (*coprolalia*) and the involuntary production of obscene gestures (*copropraxia*).

Tics are often emitted in bouts of one or several tics separated by variable periods of no tic activity. Tics change in frequency, anatomical location, number, complexity, and severity over minutes, hours, and days, as well as over longer periods of time. Tics may persist during sleep. Simple motor tics are the most common initial presentation of a tic disorder, and eye blinking is the most common initial tic.

Tics are experienced as irresistible but can be suppressed for varying periods of time. Some patients describe suppressing tics for several hours, for example, while at work or school, and then experiencing a bout of tics. There is usually a reduction in the frequency and severity of tics during concentration on mental or physical tasks such as reading. Stress, excitement, boredom, fatigue, and exposure to heat can exacerbate tics. Tic frequency may also increase during relaxation following a period of stress or sometimes when an individual is actively attempting to suppress tics. Some children, and occasionally adults, are unaware of their tics.

Many individuals with TD characterize their tics as a voluntary response to an uncomfortable feeling that precedes them. These feelings, or premonitory sensations, have been described by many individuals with TD; they can be localized or general, and physical or mental, in nature. These premonitory sensations distinguish tic disorders from other movement disorders. An example is a burning feeling in the eye before an eye blink. However, there are often nonlocalizable and more generalized premonitory phenomena, such as urges, sensations, or anxiety. Many individuals report a “just right phenomenon,” that to relieve the uncomfortable urge they must repeat a particular movement or sound or pattern of movements or sounds, until “it feels good” or “it feels just right.”

Table 47–1. Examples of Tics

Motor		Vocal	
Simple	**Complex**	**Simple**	**Complex**
Eye blinking	Touching objects or self	Throat clearing	Syllables or words
Nose twitching	Hitting or kicking	Coughing	Phrases
Shoulder shrugging	Hand gestures	Sniffing	Swearing
Sustained eye closure	Throwing	Blowing	Repeating one's or others' words
Tensing of abdominal or limb muscles	Squatting or jumping	Barking	Changes in the rate, rhythm, and volume of speech

Complex tics may be difficult to distinguish from compulsive behaviors. This is further complicated by the frequent presentation of obsessive-compulsive disorder (OCD) symptoms with tics. The difference between some complex tics and compulsions is subtle. It is important to determine whether the behavior is being pursued as a strategy to prevent or relieve anxiety rather than simply to satisfy an urge. If anxiety relief or prevention is the case, a compulsion rather than a tic is likely. Studies by Miguel and coworkers demonstrated that more cognitions and autonomic anxiety, and fewer sensory phenomena are reported with repetitive behaviors in individuals with OCD without concomitant tics, whereas more sensory phenomena, such as the "just right phenomenon," and fewer cognitions are associated with repetitive behaviors in individuals with TD.

The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders text revision (DSM-5-TR) and the eleventh revision of the International Statistical Classification of Diseases and Related Health

Problems (ICD-11) provide diagnostic criteria in common use for tic disorders. According to the DSM-5-TR (with the equivalent ICD-11 diagnoses in parentheses), tic disorders include provisional tic disorder (transient motor tics in ICD-11), persistent motor or vocal tic disorder (chronic motor or vocal tic disorder in ICD-11), TD (Tourette syndrome in ICD-11), and other specified tic disorder and unspecified tic disorder (other specified primary tics or tic disorders, primary tics or tic disorders, unspecified, other specified tic disorders, and tic disorders, unspecified in ICD-11).

By DSM-5-TR criteria, a provisional tic disorder is characterized by the presence of single or multiple motor and/or vocal tics for less than 1 year. A persistent motor or vocal tic disorder is diagnosed when either motor or vocal tics, but not both, are present for more than 1 year. For this category of tic disorders, DSM-5-TR provides specifiers to indicate whether motor or vocal tics are present: *With motor tics only* and *With vocal tics only.* TD (Tourette syndrome in ICD-11) is defined by the presence of both multiple motor and one or more vocal tics for more than 1 year.

According to the DSM-5-TR, other specified tic disorder is a tic disorder that does not meet the criteria for any of the other tic disorders, and the clinician specifies the reason criteria are not met. An unspecified tic disorder, on the other hand, is a condition that also does not meet criteria for any of the other tic disorders, but for which the clinician does not elaborate a specific reason. Included are presentations for which information to make a more specific diagnosis is lacking.

In addition to the foregoing criteria that the DSM-5-TR and ICD-11 share for the diagnosis of a specific tic disorder, the DSM-5-TR specifies a few common criteria that must be met to diagnose provisional tic disorder, persistent motor or vocal tic disorder, or TD. The onset of tics must be before the age of 18 years, and a substance or medical condition must not be causally implicated. The DSM-5-TR notes that tics wax and wane in frequency.

HISTORY

In 1825, the French physician Jean-Marc Gaspard Itard provided the first clinical description of Tourette syndrome in a French noblewoman, the Marquise de Dampierre, who displayed body tics, barking sounds, and uncontrollable utterances of obscenities. In his description, Itard also noted the presence of comorbid obsessive thoughts. In 1885, Georges Edouard Albert Brutus Gilles de la Tourette, a French neurologist, described eight patients, in addition to the patient originally reported by Itard, with childhood-onset tics and coexisting emotional symptoms and behavioral problems, such as inattention, hyperactivity, impulsivity, obsessions, and compulsions. Gilles de la Tourette coined the term *coprolalia* for vocal tics that manifested as uncontrollable utterance of obscenities. Furthermore, he and his mentor, Jean-Martin Charcot, viewed the disease as progressive and believed they had identified a hereditary degenerative disorder.

Psychoanalytic views of TD, introduced in the 1920s and 1930s, thrived in the early twentieth century and interpreted tics as expressions of repressed childhood sexual conflict. In the 1940s, Margaret Mahler attributed tic symptoms to infantilization, parental overindulgence, and overprotection, as well as sexual conflict. Serge Lebovici, the leading French child psychoanalyst during the 1950s and 1960s, asserted that neurotic mothers were the cause of tics in their children.

As more biologically based theories of the etiology and pathophysiology of tic disorders emerged, sedatives, stimulants, carbon dioxide inhalants, and other agents came to be used in the treatment of tics, in conjunction with, or as alternatives to, psychoanalytic psychotherapy. In the 1960s and 1970s, as the scientific era of the understanding of Tourette syndrome began with the work of Arthur and Elaine Shapiro, the first randomized clinical trials using neuroleptic medications were published. The perception of Tourette and tic disorders as rare, bizarre, psychological conditions slowly began to change. Advances in the last 50 years have led to the use of the α-adrenergic agonists and second-generation neuroleptics as the primary psychopharmacologic agents for the treatment of tics. Comorbid psychiatric disorders have come to be more widely recognized and treated. Over the last several decades, the Tourette Association of America (TAA) (https://tourette.org) has effectively promoted clinical awareness of the condition and supported research into its causes and treatment. Behavioral interventions have come to the forefront. The quest for a more comprehensive understanding of the underlying etiology and pathophysiology of tic disorders continues.

COMPARATIVE NOSOLOGY

The nosology of the tic disorders has depended, since their recognition, on the presence and classification of their primary symptom—tics. The two predominant current classification systems of tic disorders, the DSM-5-TR and the ICD-11 criteria, are based primarily, as described earlier, on the type of tics present (motor, vocal, or both) and the duration of symptoms (greater or less than 1 year). ICD-11 also distinguishes between primary tics and tic disorders and secondary tics in its categorization. As new information about the etiology and pathogenesis of tic disorders emerges, a new nosology based on an etiologic or pathogenic understanding may replace the current symptom-based taxonomy. Over the years, with newer iterations of the DSM, individual criteria have changed slightly as advances in the understanding of the prevalence and course of tic disorders have emerged. For example, the age of onset of TD has varied from before age 15 years to before age 21 years. The most recent version, DSM-5-TR, has instituted changes to the criteria that reflect the current understanding of the neurodevelopmental and temporal nature of tics, as tic disorders are now contained in the section on Neurodevelopmental Disorders. Tics are now recognized to wax and wane, and restrictions on the time frame of tic-free periods have been removed.

EPIDEMIOLOGY

Provisional tic disorder, which can encompass both motor and vocal symptoms, is the most common of the tic disorders; the prevalence is estimated at up to 20% of school-aged children. A methodologically rigorous study of more than 800 schoolchildren ages 4 through 16 years in the Basque region, however, indicated a prevalence of approximately 6.5% for all tic disorders. The vast majority of tics were mild in severity and duration. The prevalence of persistent tic disorders is between 2% and 4%. Prevalence estimates for TD also vary. As in all population-based studies, results may be influenced by a variety of factors, including criteria used for diagnosis, methods of sampling and inquiry, gender and age distributions in the study population, and size of the sample. Once considered an extremely rare disorder, current lifetime rate estimates vary from 1 to 10 per 1,000 children in European and Asian populations, and a common prevalence figure for individuals meeting criteria for TD is 1 per 200 for the full spectrum of chronic tic disorders. A population-based, epidemiologic study of lifetime prevalence of TD in Israel estimated a point prevalence of 3.1 to 5.5 per 10,000 in 16- and 17-year-olds. Prevalence rates have been estimated to be as high as 4% in one community-based study, and 3% in a survey of middle school youth in the United Kingdom. In a school-based community survey, Kurlan and colleagues reported prevalence rates of tic disorders of 23% in special education classes and as high as 8% for TD, as compared to 19% for tic disorders and about 3% for TD in regular classrooms. The 2007 U.S. National Survey of Children’s Health (NSCH) surveyed parents of children and adolescents aged 6 to 17 years. The estimated prevalence of a lifetime diagnosis of TD by parent report was 3 per 1,000. In a meta-analysis of 13 studies of children, Knight and colleagues reported a prevalence of TD at 0.77% (95% confidence interval, 0.39% to 1.51%).

The average age of onset for TD is age 6 years with peak tic severity between 10 and 12 years of age. Limited and somewhat contradictory information exists on the longitudinal course of the disorder. Recent studies suggest that whereas tics may often persist into adulthood, tic severity often declines significantly in adolescence. Only 20% or fewer of those with TD continue to experience significant functional impairment by age 20 years. In most cases, tics in adults represent the persistence or recurrence of childhood-onset tics, but in extremely rare instances, patients may have their first occurrence of tics during adulthood. Studies in both children and adults suggest that males are at least three to four times more likely than females to manifest TD.

Comorbidity with psychiatric disorders, especially attention-deficit/hyperactivity disorder (ADHD) and OCD, is common in both clinically referred individuals and community samples. In addition, mood disorders and anxiety disorders, such as separation anxiety disorder and generalized anxiety disorder, often co-occur. In the 2007 U.S. National Survey of Children’s Health (NSCH), 79% of those with TD had at least one comorbid mental health or neurodevelopmental condition. ADHD was the most frequent co-occurring condition, and the prevalence of comorbid ADHD with TD was reported to be as high as 60% to 70%. In addition, approximately 50% of patients with TD have prominent obsessive-compulsive symptoms. A more recent estimate of psychiatric comorbidity in those with TD at about 86% is the result of the work of Hirschtritt and collaborators.

Jeffrey was a 5-year-old boy whose parents sought evaluation for recent onset of tics. Jeffrey developed onset of eye-blinking tics at age 4 years and subsequently began to repeatedly push out his hands and legs in front of him, kicking and stomping his feet. These symptoms persisted for about 3 months and remitted. The movements returned about 1 year later, somewhat abruptly, according to his parents. At that time, Jeffrey developed eye blinking and nose twitching. Sometimes, he hopped while walking. He also developed vocal tics around that time, including clicking, throat clearing, sniffing, and coughing. Approximately 1 month prior to the evaluation, Jeffrey developed truncal tics that occasionally began to disrupt his gait.

Jeffrey's parents were also concerned about his problems paying attention in school in the last year. Hyperactive since preschool, he had a very short attention span and was often distracted. He could be impulsive, often unaware that he was doing things that could be dangerous. He often interrupted his parents when they were speaking. He was restless and fidgety in school, and his teachers were concerned that he might disrupt his classmates.

Jeffrey was the product of a pregnancy complicated by intrauterine growth retardation. When the amniotic fluid was found to be at low volume at 36 weeks, delivery was induced. The delivery was uncomplicated, and the birth weight was 5 pounds, 10 ounces. Except for a minor delay in the acquisition of speech, developmental milestones were within normal limits.

Currently in kindergarten, Jeffrey has attended a well-regarded public school since he was 2 years of age. He lived with his parents and 2-year-old sister, had many friends, and enjoyed play dates.

Jeffrey had frequent upper respiratory infections. There were no known allergies and no major medical problems. There was no known family history of TD, tics, or OCD.

MENTAL STATUS EXAMINATION

Jeffrey was a healthy-looking 5-year-old boy, well-groomed and dressed. He was initially very shy, clinging to his mother. He appeared to be quite active and restless. Multiple tics were noted, including nose twitches, finger-snapping, and a cough. Jeffrey was somewhat oppositional when asked to do a drawing.

IMPRESSION

In summary, Jeffrey was a 5-year-old boy with multiple motor and vocal tics for short episodes, including a bout lasting 3 months at age 4 years, and just over 2 months currently, consistent with provisional tic disorder. TD could not be entirely ruled out, but there was no diathesis based on family history. Jeffrey also appeared to meet criteria for ADHD, combined presentation, which may have predisposed him to the development of tics. The ADHD symptoms were causing impairment in his academic functioning.

TREATMENT PLAN

Parent and teacher ratings of ADHD symptoms were recommended at initial evaluation, and at follow-up after treatment was initiated. Treatment of ADHD was recommended, starting with a cautious trial of a methylphenidate-based stimulant. The plan was to add a low dose of an α-adrenergic agonist such as guanfacine or clonidine if tics increased significantly. Consultation and psychoeducation were recommended for Jeffrey’s teacher.

ETIOLOGY AND PATHOPHYSIOLOGY

The etiology and pathophysiology of TD and other tic disorders remain unknown. Considerable data accumulated during the last few decades from neuroanatomical and neurochemical studies of TD point to a diffuse process in the brain involving cortico–striato–thalamic–cortical (CSTC) pathways in the frontal lobes and basal ganglia, including the striatum (caudate nucleus and putamen). Several neurotransmitters and neuromodulators have been implicated, including dopamine, serotonin, glutamate, GABA, and endogenous opioids. Although a specific animal model has not been identified, some parallels may exist with horses with crib biting, which display spontaneous motor tics accompanied by vocalizations, dogs with acral lick syndrome, and mouse models with repetitive movements and stereotypies.

A host of biochemical, imaging, neurophysiologic, and genetic studies support the concept that TD is an inherited neurodevelopmental disorder. Twin studies have reported 89% to 94% concordance for TD. However, environmental factors have also been implicated. In one study of 16 pairs of monozygotic twins, low–birth-weight was a strong predictor of more severe tics. Several gestational and perinatal insults, including hypoxia–ischemia, exposure to androgens, heat, and infectious agents, have been implicated in the pathophysiology of tic disorders. The effect of prenatal ischemic events may be to alter dopamine signaling, heralding the development of tics. Delivery complications, especially forceps deliveries, and fetal exposure to high levels of coffee, cigarettes, or alcohol were associated with male gender and OCD in a sample of TD probands.

Androgen exposure is speculated to play a role in the development of tics, given the higher prevalence of TD in males. Studies in both children and adults suggest that males are at least three to four times more likely than females to manifest TD. X-linked genomic transmission has also been considered, but this appears less likely in the context of frequent male-to-male transmission within families. The role of stress, given the strong association of tics with anxiety, is also compelling. TD patients have been found to have significantly higher levels of cerebrospinal fluid (CSF) corticotrophin-releasing factor (CRF) than normal controls or patients with OCD who do not have a tic disorder, demonstrating a heightened reactivity of the hypothalamic–pituitary–adrenal axis. Increased reactivity of the noradrenergic sympathetic system has also been shown in patients with TD. A dopamine-mediated pathway in the hypothalamus that modulates sweating in response to changes in temperature has also been linked with tic activity. Increases in ambient and core body temperatures have been shown to transiently increase tics in some individuals. A stress diathesis model modulated by the hypothalamic–pituitary–adrenal axis, the associated stress-related neurotransmitters and hormones, and their targets may represent the final common pathway that may explain the variable expression of the tic disorders.

Pediatric Autoimmune Neuropsychiatric Disorders Associated with Streptococcal Infection

A postinfectious mechanism for the development of tics and OCD has been postulated since the late 1800s. A poststreptococcal infection autoimmune etiology has long been well established for Sydenham chorea, the movement disorder that was the model for the development of pediatric autoimmune neuropsychiatric disorders associated with streptococcus (PANDAS) hypothesis. In addition to chorea, some Sydenham’s patients display motor and phonic tics, as well as obsessions, compulsions, hyperactivity, impulsivity, and inattention. In 1998, Swedo and colleagues first described a putative subgroup of 50 children with PANDAS with OCD and/or tic disorders with acute symptom onset and/or exacerbation precipitated by infection with group A β-hemolytic Streptococcus (GABHS). Although GABHS is considered the initial autoimmune response-inciting event in children with PANDAS, viruses, other bacteria, and even noninfectious immunologic responses are thought to be capable of precipitating later symptom exacerbations. Case reports have also implicated Lyme disease and Mycoplasma pneumonia in the development of infection-triggered tics. By molecular mimicry—the postulated autoimmune mechanism—antibodies are formed against the inciting agent and subsequently attack neurons because of some structural similarity.

Specific diagnostic criteria for PANDAS have been described, including the presence of OCD and/or tic symptoms, prepubertal onset, episodic course with sudden, explosive onset and/or exacerbations and remissions, and a temporal relationship of GABHS infection and neuropsychiatric symptoms. A case-control study of 144 children with OCD, TD, or another tic disorder conducted by Mell and colleagues provided the strongest evidence for the involvement of

GABHS in the onset of tic disorders and OCD. Patients with OCD, TD, or another tic disorder were found to be more likely than controls to have had a streptococcal infection in the 3 months prior to the onset of symptoms. Risk was greatest among those with multiple streptococcal infections in the preceding 12 months. Others have found no more than a chance association between GABHS infections and tic symptom exacerbations. A case-control study by Perrin and colleagues found little evidence of increased tics, obsessions, or compulsions after well-documented and treated GABHS infections.

Several groups have reported increased antistreptococcal antibody titers in individuals with OCD and tic disorders, but several other groups have not found this association. A double-blind, placebo-controlled, crossover study of prophylactic oral penicillin to attempt to reduce PANDAS recurrences failed to show differences between drug and placebo. More recently, in two prospective case-control studies of children with PANDAS, Kurlan and colleagues reported that Group A B-hemolytic streptococcus infection was not the only antecedent associated with exacerbations of tics and OCD symptoms in the cases; similarly, Leckman and colleagues found there was no evidence for a temporal association between GABHS infections and tic/OC symptom exacerbations in children who met published PANDAS diagnostic criteria.

Treatments based on a postulated autoimmune mechanism have been nonspecific, and the results have been inconsistent. Immune therapies such as intravenous immunoglobulin (IVIG), plasmapheresis, corticosteroids, and plasma exchange have been explored with inconsistent results. Immunomodulatory treatments remain experimental, and current recommendations are for use only in research protocols.

The link between an infectious etiology and tic disorders/OCD remains a work in progress. Building on previous work, the most recent investigation has shifted its focus on acute, explosive onset OCD symptoms, which Swedo and colleagues have described as pediatric acute onset neuropsychiatric syndrome (PANS). Anxiety, emotional lability, developmental and behavioral regression, deterioration in schoolwork and handwriting, and excessive urination are among the secondary characteristics.

Other new areas of research include the investigation of proinflammatory cytokine elevation, as well as decreased numbers of regulatory T-cells in some TD patients. Clinical implications suggest vigilance in the evaluation of tic or OCD patients with pharyngitis or streptococcus exposure, a low threshold for obtaining streptococcus throat culture, antistreptolysin O (ASO) and anti-DNAse B titers, and vigorous treatment with antibiotics when a streptococcal infection is confirmed. Antibiotic treatment has been shown to be remarkably effective in some unambiguous cases of sudden onset or exacerbation of tics with associated laboratory-documented streptococcal infection.

Neuroimaging Studies

Structural MRI studies in TD demonstrate volume changes in brain regions associated with CSTC circuits. Some studies have shown a loss of the normal basal ganglia asymmetry, as well as decreased basal ganglia volume in TD patients. In ten pairs of monozygotic twins, the right caudate was smaller in the more severely afflicted individuals, supporting evidence for the role of epigenetic mechanisms. A large study of basal ganglia volume in 154 patients with TD and 130 controls, conducted by Peterson and colleagues, found a significant decrease in the volume of the caudate nucleus in both children and adults. Larger dorsal prefrontal regions in children, but significantly smaller volumes in adults, were also reported by this group. Dorsal prefrontal volume was inversely related to tic severity, such that patients with greater prefrontal volumes exhibited decreased tic severity. Larger prefrontal volumes in children with TD may be a manifestation of synaptic plasticity associated with the constant need to suppress tics in social contexts. Adults with persistent tics may represent a subsection of the overall TD population who do not generate a neuroplastic, compensatory response in the prefrontal cortex, leading to increased severity of the tic disorder and its continuity into adulthood.

Larger parieto-occipital region volumes and smaller inferior occipital volumes have also been observed in individuals with TD when compared with controls. Bloch and colleagues found an inverse correlation between caudate volume in childhood and tic severity in late adolescence and early adulthood. Smaller caudate volumes were associated with more significant later tic activity, as well as with OCD symptoms. Another more recent study reported increased volume of the putamen bilaterally in boys with TD.

Decrease in corpus callosum size in children and an increase in corpus callosum size in adults with TD have also been reported. Corpus callosum size correlates inversely with prefrontal volume and is positively related to tic severity. These findings may indicate age-dependent changes in the white matter tracts in the disorder. Alternatively, decreased corpus callosum size, like increased dorsolateral prefrontal volume, may represent a compensatory mechanism children with TD develop to facilitate suppression of tics. Adults with TD tend to show the opposite pattern, exhibiting relatively decreased dorsal prefrontal and increased corpus callosum volumes that may be associated with the persistence of symptoms into adulthood.

Several studies consistently implicate primary cortical dysfunction in TD. Cortical thinning has been documented in frontal and parietal lobes, especially the sensorimotor cortex; the degree of thinning may correlate with tic severity. Increased sensorimotor cortical volume has been correlated with strength of premonitory sensations in a study of adults with TD. Evaluation of subgroups of TD patients has shown a relationship between symptoms and location of cortical thinning. While most have found thinner cortex or decreased volume in several different cortical areas, a few studies have reported overall brain volume decreases in TD subjects.

Larger left thalamic volumes were found in 18 patients with TD compared to 16 control subjects in one study. Increased gray-matter volumes bilaterally in the ventral putamen and decrease in gray matter in the left hippocampal gyrus have also been shown. Peterson and team demonstrated significantly increased overall hippocampus and amygdala volumes in TD patients, with a decline in volume with age in the TD group, but not in healthy controls. In children and adults with TD, these volumes were inversely correlated with the severity of tics, OCD symptoms, and ADHD symptoms. An explanation for these findings is that the subregion plays a compensatory and neuromodulatory role in tic, OCD, and ADHD symptoms. In addition, cerebral hyperintensities—bright patches in brain parenchyma on T_2-weighted MRI that are indicative of neuroinflammatory or neurodegenerative processes—have been detected at increased rates in children and adolescents with TD, OCD, or ADHD in subcortical regions (primarily the basal ganglia and thalamus). Taken together, volumetric studies suggest that cortical and subcortical impairment, including striatal anomalies, and disruption of the temporolimbic pathways of the CSTC circuits are implicated in the pathophysiology of TD.

Functional neuroimaging studies using state-dependent cerebral blood flow, oxygenation, and energy (glucose) metabolism parameters as a measure of brain activity (e.g., functional MRI [fMRI], positron emission tomography [PET], and single photon emission tomography [SPECT]) suggest altered activity (increased or decreased) in various areas of the brain (e.g., frontal and orbital cortex, striatum, putamen) in TD patients compared to controls, or across brain regions within TD patients. An fMRI study by Bohlhalter and colleagues that analyzed activity two seconds before and after tic onset indicated that paralimbic and sensory association areas are critical to tic generation. Other fMRI studies in individuals with TD indicate that greater ability of basal ganglia structures to suppress cortical activity may be related to decreased tic severity, supporting findings from PET and SPECT studies of the role of the basal ganglia in TD. Functional neuroimaging studies suggest that voluntary tic suppression involves deactivation of the putamen and globus pallidus, coupled with partial activation of regions of the prefrontal cortex and caudate nucleus.

Event-related PET scanning has connected increased or decreased activity in many brain areas with the presence of tics, including the prefrontal cortex, premotor and primary motor cortex, anterior cingulate cortex, putamen, and caudate. fMRI scanning has demonstrated increased neural activity in individuals with TD compared to controls throughout various motor areas, including the sensorimotor cortex, putamen, pallidum, and substantia nigra; this increased activity correlates with symptom severity. On the other hand, weaker activity has been demonstrated in the TD patients' motor control pathways, including the caudate and anterior cingulate cortex. More severe tics have been observed in those with progressively less activity in these regions, suggesting a disruption in these areas results in failure to control tics or premonitory urges. Other fMRI studies have also shown age-related abnormalities in frontostriatal and frontoparietal connections, implicating abnormal maturational changes in TD patients, strengthening the understanding that tics are the result of neurodevelopmental deficits.

Abnormal dopamine neurotransmission that contributes to the pathophysiology of TD is supported by increased levels of dopaminergic innervation of the striatum of TD patients in SPECT studies. A study in monozygotic twins concordant for TD but discordant for severity demonstrated abnormalities in D_2 receptors in the caudate area in the more severely afflicted twin. These neuroimaging findings are consistent with results of well-established tic suppression treatments with neuroleptic drugs that block dopamine receptors. Furthermore, TD patients have been shown to release more dopamine in response to amphetamine at dopaminergic synapses than healthy controls and dopamine-releasing drugs may precipitate or exacerbate tics. Midbrain dopaminergic neurons play a central role in motor control and attentional processes through direct connections to the striatum and cortex, respectively. Glutamatergic, GABAergic, serotonergic, cholinergic, noradrenergic, and opioid systems may also be implicated, given that they are present in CSTC circuits and interact with the dopaminergic system.

Taken together, structural and functional neuroanatomical findings support the understanding that TD results from abnormal activity in diverse cortical and subcortical brain regions. Active suppression of tics may require activation of dorsolateral prefrontal circuits that increase overall activity in the caudate nucleus, thereby inhibiting activity in the putamen.

Genetic Findings

Genetic data have been derived historically from family pedigree and twin studies. TD, persistent tic disorders, and OCD cluster in families. Twin studies have shown a high concordance rate for tic disorders among monozygotic (MZ) pairs (ranging from about 50% to 90%), in contrast to a relatively low rate among dizygotic twins (often around 20%). Other studies have shown that first-degree relatives of TD patients have a higher risk of TD, persistent tics, and OCD than normal controls. One large population cohort, multigenerational family study with 4,826 individuals diagnosed with TD or chronic tic disorders in the Swedish National Patient Register, estimated the heritability of tic disorders to be 0.77.

Historically, genetic analyses were thought to be consistent with an autosomal dominant mode (with incomplete penetrance) of inheritance; some investigators argued that a semidominant, semirecessive pattern could also exist. However, older studies were confounded by relatively high rates of bilineal transmission with apparent genetic contributions from both paternal and maternal lines.

More recently, investigators have proposed multiple gene loci in combination with environmental factors. The precise type or severity of tic disorder may be different from one generation to another and may be modified by nongenetic factors. A systematic genome scan of 76 affected sibling pair families with 110 sib pairs conducted by the Tourette Syndrome Association International Consortium for Genetics (TSAICG) demonstrated that 4q, 8p, and four other regions on chromosomes 1, 10, 13, and 19 were regions of interest. Efforts to confirm and extend these findings, however, have led to the identification of new regions and a failure to replicate the original findings.

Specific TD populations have been studied in South Africa, Costa Rica, and French-speaking Canada. Regions near the centromere of chromosome 2, as well as on 6p, 8q, 11q, 14q, 20q, 21q, and X, have been implicated. A linkage analysis study of a large pedigree of multigenerational TD patients in the United Kingdom is suggestive of linkage at loci on chromosomes 5, 10, and 13 in at least one of the families. The largest genetic linkage study, undertaken by the TSAICG, consisting of more than 2,000 individuals, using a whole-genome scan with 390 microsatellite markers, demonstrated strong evidence of linkage for a region on chromosome 2p.

Cytogenetic abnormalities in TD families have also been studied. A TSAICG study reported disruption of the contactin-associated protein 2 gene (CNTNAP2) on chromosome 7, coding for a membrane protein important for signal conduction along myelinated neurons at the nodes of Ranvier, in TD patients. Affected individuals were found to have alterations in chromosomes 2 and 7. Other TD patients have been described with *de novo* duplication of the long arm of chromosome 7, a balanced translocation between chromosomes 6 and 8 with a translocation breakpoint on chromosome 8q, and an 18q21.1–q22.2 inversion. Abelson and colleagues identified a candidate gene, SLITRK1, on chromosome 13q31.1 in some individuals with TD, with expression of the gene product in the basal ganglia and deep layers of cortex, brain regions implicated in the pathophysiology of TD. Ercan-Sencicek and colleagues reported a rare functional mutation in the *HDC* gene encoding l-histidine decarboxylase, the rate-limiting enzyme in histamine biosynthesis, in linkage analysis in a two-generation TD pedigree.

Overall, the current understanding of TD and tic disorders implicates multiple vulnerability genes that likely determine the form and severity of tics in the wide variety of afflicted individuals and their families. Candidate genes have been identified in individuals with TD, including the dopamine receptor genes, the dopamine transporter genes, several noradrenergic genes, and some serotonergic genes. The TSAICG genome-wide association study (GWAS) reported that no genes reached threshold significance. A more recent analysis by Huang and colleagues of a European ancestry sample of 2,434 TD cases and 4,093 ancestry-matched controls for rare copy number variants has led to the identification of two genome-wide significant loci, *NRXN1* deletions, and *CNTN6* duplications, each contributing a significant increase in TD risk.

Willsey and peers have used whole-exome sequencing and *de novo* variant detection techniques to identify four likely risk genes with multiple *de novo* damaging variants in unrelated individuals: *WWC1*, *CELSR3*, *NIPBL*, and *FN1*. They estimate that *de novo* damaging variants in about 400 genes contribute risk in 12% of clinical cases.

DIAGNOSIS AND CLINICAL FEATURES

A comprehensive, detailed history with observations derived from multiple sources is the cornerstone of clinical evaluation of tics and comorbid neurodevelopmental and psychiatric disorders. Neurologic, laboratory, and imaging examinations in patients with tic disorders are usually normal. Since tic disorders are diagnosed based on history, reliable sources of information are essential. This is particularly important because in many patients tics may be suppressed during the initial examination. It is critical to determine the degree to which tics interfere with emotional, social, family, and academic functioning, as treatment strategies are most effective when the severity and impact of tics are taken into consideration. Diagnostic evaluation is closely connected with the first steps of treatment.

Historical inquiry should include detailed medical and developmental information, medication history (including substances of abuse or recreation), educational and occupational data, social and interpersonal history, and a thorough family pedigree covering at least three generations. A careful, descriptive, longitudinal assessment of the tic disorder is important, including onset, course, current severity, presence of premonitory sensations and the capacity for tic suppression, overall burden, and treatments to date. The physical examination should include height, weight, presence or absence of dysmorphic features, posture, gait, reflexes, and a systematic rating of current abnormal movements. The Abnormal Involuntary Movement Scale (AIMS) is a systematic assessment procedure (and rating form) that can be adapted for use with children. Videotaped standardized interviews and evaluation procedures are used in some research studies; these could also be used in the clinical setting. However, the absence of observed tics on initial examination does not preclude the diagnosis of a tic disorder. Unusual histories, co-occurring mental status changes, or the presence of seizures always warrant referrals for thorough neurologic evaluation.

Psychiatric evaluation should include a formal assessment of the behavioral and emotional problems known clinically to co-occur with tic disorders, including ADHD, OCD, anxiety disorders, mood disorders, and manifestations of impaired or dysregulated affect (e.g., aggression). The use of structured or semistructured interviews, such as the Diagnostic Interview Schedule for Children (DISC) or the Children’s Schedule for Affective Disorders and Schizophrenia (K-SADS), can improve classification and evaluation of psychiatric comorbidity. Standardized rating scales developed specifically for the tic disorder population have improved diagnostic reliability in research studies and can also be helpful in clinical care. The Yale Global Tic Severity Scale (Y-GTSS) and the Tourette Syndrome Symptom List (TSSL) rate tics, compulsions, and other associated features. Specific rating scales for OCD (the Children’s Yale-Brown Obsessive-Compulsive Scale [C-YBOCS]) and ADHD (the IOWA Conners parent and teacher rating scales, or the Swanson, Nolan, and Pelham [SNAP] scale) can also be used. Because rating instruments can provide quantifiable baseline data, such as frequency and intensity of tics, these data can also be used to measure treatment response (Table 47–2).

Table 47–2.
Rating Scales

Domain	Type	Reliability and Validity	Sensitive to Change
Tics			
Yale Global Tic Severity Scale (YGTSS)	Clinician	Excellent	Yes
Abnormal Involuntary Movement Scale (AIMS)	Clinician	Excellent	Yes
Tic Symptom Self-Report (TSSR)	Parent/self	Good	Yes
ADHD			
Swanson, Nolan, and Pelham-IV (SNAP-IV)	Parent/teacher	Excellent	Yes
Abbreviated Conners Questionnaire	Parent/teacher	Excellent	Yes
OCD			
Yale-Brown Obsessive-Compulsive Scale (YBOCS) and Children’s Yale-Brown Obsessive-Compulsive Scale (CYBOCS)	Clinician	Excellent	Yes
National Institute of Mental Health Global Obsessive-Compulsive Scale (NIMH-GOCS)	Clinician	Excellent	Yes
Mood and Anxiety Disorders			
Children’s Depression Inventory (CDI)	Clinician	Good	Yes
Children’s Depression Rating Scale (CDRS)	Self	Good	Yes
Multidimensional Anxiety	Self	Good	Yes

Scale for Children (MASC)			
General			
Achenbach Child Behavior Checklist (CBCL)	Parent/teacher	Excellent	No
Clinical Global Impressions Scale (CGI)	Clinician	Excellent	Yes

Auxiliary data from outside sources are essential. Pediatric and medical records may document developmental and medical history, adequacy of medication trials, hospitalizations, and laboratory findings. Review of educational records is advised for children because many TD patients manifest tics or associated problems in school settings. Report cards document academic performance. Direct phone contact with teachers may provide data about attentional functioning, learning challenges, social and emotional competencies, and any behavioral difficulties. Neuropsychological or speech and language testing may be indicated for patients with impairments in school or occupational functioning. Identified areas of strength and weakness are subsequently conveyed to appropriate personnel for inclusion in educational or vocational planning.

PATHOLOGY AND LABORATORY EXAMINATION

In general, laboratory tests are indicated only to rule out other disorders when such conditions are highly implicated in the differential diagnosis of a given patient. For example, specific tests are available for the definitive diagnosis of Huntington disease, Wilson disease, and neuroacanthocytosis.

DIFFERENTIAL DIAGNOSIS

Differential diagnosis of tic disorders can be challenging (Table 47–3). Medical and family histories, physical and neurologic examinations, and baseline laboratory testing enable clinicians to rule out potential causes of *de novo* tics such as infection, trauma, substance use, or toxin ingestion. Diagnostic hallmarks of tic disorders include the waxing-and-waning natural history, onset in childhood or adolescence, repetitive, rapid, nonrhythmic phenomenology, and involuntary features. In the absence of at least some history of simple tics, the diagnosis of a tic disorder may be questioned. Furthermore, the presence of vocal tics is helpful in ruling out other disorders because phonic manifestations are rare in other neurologic conditions. Tics must be distinguished from other types of abnormal movements: myoclonus, tremors, chorea, athetosis, dystonias, akathisia, and compulsive rituals.

Diagnostic considerations among the movement disorders include Sydenham chorea, poststreptococcal neurologic complications such as choreiform, writhing, and truncal movements; Huntington disease, an autosomal dominant disorder presenting with chorea and dementia and with an onset typically in the fourth or fifth decade; Parkinson disease, typically a late-life disorder, characterized by flattened facies, bradykinesia, rigidity, cogwheeling, and “pill-rolling” resting tremor; Wilson disease, an autosomal recessive disorder of copper metabolism that presents with liver abnormalities and neurologic symptoms: early dementia, mood disorders or psychosis, asterixis (a flapping tremor of the hands); Parkinsonism; structural diseases such as hemiballismus associated with subthalamic nucleus impairment; and neuroacanthocytosis, a rare movement disorder marked by progressive muscle weakness and atrophy, progressive cognitive loss, chorea, acanthocytosis (spiked red blood cells), and, sometimes, facial tics. In psychiatric settings, patients who have been exposed to neuroleptics are at risk for neuroleptic-related dyskinesias, including tardive and withdrawal dyskinesias. It is important to remember that dyskinesias may also be observed in tic disorder patients secondary to neuroleptic therapy but may be easily overlooked.

Table 47–3.
Differential Diagnosis of Abnormal Movements

Descriptive Terms	Observable Movement Pattern(s)
Akathisia	Motor restlessness (unpleasant need to move), usually in the lower extremities
Athetosis	Slow, irregular, writhing movements, usually in the hands and fingers but often involving the neck
Ballismus	Intermittent, coarse, large amplitude, jerking, shaking, flinging movements
Chorea	Irregular, spasmodic movements, usually affecting the limbs or face
Compulsion	Repetitive behavior that is more goal-directed and may be undertaken to prevent or reduce anxiety or distress
Dyskinesia	Choreiform or dystonic movement that is involuntary, stereotyped, and not suppressible
Dystonia	Sustained, tonic contraction that progresses to abnormal postures
Myoclonus	Sudden, brief, clonic, shock-like jerks, usually involving the limbs
Periodic movements of sleep	Periodic dorsiflexion of the foot and flexion of the knee, occurring during sleep
Spasm	Stereotypic, prolonged movement involving the same group of muscles
Stereotypy	Repetitive, usually meaningless, gestures, habits, or automatisms
Tic	Stereotyped, rapid, recurring motor movement (motor tic) or vocalization (phonic or vocal tic) that is nonrhythmic, involuntary or semivoluntary, and sudden in onset

Tremor	Unintentional, somewhat rhythmic muscle movement involving oscillations of one or more parts of the body

COURSE AND PROGNOSIS

The typical course of TD is characterized by onset of simple motor tics of the eyes, face, head, or neck by about age 6 or 7 years, followed by a rostral-caudal progression of motor tics over several years to involve the shoulders, trunk, and extremities. Eye blinking is the most common initial tic. Vocal tics typically start at ages 8 to 9 years; more-complex tics often begin later, as do obsessive-compulsive symptoms, at about ages 11 to 12 years. As many as half of clinically referred TD patients may show signs of ADHD, such as hyperactivity, impulsivity, and distractibility, prior to the onset of tics. Tics range from mild to severe and tend to stabilize over time; most children evaluated even in specialty clinics have mild to moderate tics. Some patients have lengthy periods during which most or all manifestations attenuate or remit. Recent studies indicate that tic severity improves significantly in most patients by early to mid adolescence.

Social functioning may be impaired because of bullying or rejection by peers or anxiety about having tics in social situations. Tics may impair academic functioning by their disruptive impact in class or interference with homework completion. Academic underachievement across educational levels is seen in many seeking treatment for TD or chronic tic disorders. Occupational functioning may be impaired by inability to use a computer or complete work tasks. By early adulthood, tics are greatly diminished in most individuals, and some are tic-free. Persistent tic symptoms may cause considerable distress and can lead to social isolation and personality changes. In severe cases, tics may directly interfere with daily activities. For example, violent head and neck tics may cause physical complications, such as compressive cervical myelopathy.

Fernández de la Cruz and peers report that there is a fourfold increase in the risk of suicide in individuals with tic disorders. The strongest predictor of risk is the presence of tics beyond the second decade of life. Psychiatric comorbidity and a history of suicide attempts increase risk as well.

Comorbid Psychiatric Disorders

Hirschtritt and colleagues conducted a large cross-sectional study of TD patients and unaffected family members, estimating a lifetime prevalence of psychiatric comorbidity in those with TD at 85.7%. 57.7% had two or more psychiatric comorbidities. The relationship between TD and frequently observed comorbid psychiatric disorders has not yet been disentangled. Obsessive-compulsive symptoms, developmentally inappropriate motoric hyperactivity and inattention, anxiety, and aggressive dyscontrol have often been described in patients with TD. Motor, vocal, behavioral, cognitive, and emotional dysfunction may represent manifestations of an underlying central core disinhibition.

A majority of clinically referred individuals with TD meet full or subthreshold diagnostic criteria for ADHD, OCD, or both at some time during the course of their illness. Interestingly, these findings are also reported in community samples. ADHD and OCD are more likely to interfere with overall functioning and academic and work performance than the tics. Investigation of functional impairment in youth with TD conducted by Storch and colleagues found that non–tic-related problems caused more dysfunction than the tics per se, most often either ADHD or OCD symptoms.

ADHD is diagnosed by history of an enduring pattern of developmentally inappropriate inattention and/or hyperactivity and impulsive behavior; there is a bidirectional relationship between tics and ADHD. Some investigators have reported that 50% to 75% of TD patients also meet criteria for ADHD, whereas about 10% to 30% of ADHD patients also meet criteria for a tic disorder. Co-occurring TD and ADHD are often associated with academic difficulties, peer rejection, family conflict, and disruptive behaviors such as explosive outbursts, overt aggression, low frustration tolerance, and poor treatment adherence. The nature of the relationship between ADHD and TD is not firmly established, but it has been clearly demonstrated that ADHD predicts greater disability than that associated with tic disorders alone, including an earlier diagnosis of TD, and a much higher rate of anger control problems and oppositional defiant disorder, sleep difficulties, learning disability, mood disorders, social skills deficits, sexually inappropriate behavior, and self-injurious behavior.

There also appears to be a bidirectional relationship between TD and OCD in most patient cohorts. OCD is characterized by obsessions—recurrent, intense, intrusive ideas, thoughts, impulses, or images that cause anxiety or distress, or compulsions—repetitive behaviors or mental acts sought to prevent or ward off anxiety or distress caused by an obsession. 20% to 40% of TD patients meet full criteria for OCD, and up to 90% have been reported to have subthreshold symptoms such as repetitive counting, touching, ordering, arranging, or symmetry needs.

Family studies provide the most compelling evidence for the association of tic disorders with OCD. First-degree relatives of TD probands have higher frequencies of TD, as well as OCD, and first-degree relatives of OCD probands have higher frequencies of OCD and tic disorders than relatives of control subjects. Studies of OCD probands indicate that the earlier the onset of obsessions or compulsions in the proband, the higher the genetic loading in family members for both OCD and tics.

In addition to ADHD and OCD, mood and anxiety disorders have been described in clinically referred and community samples of TD patients, including major depression, bipolar disorder, generalized anxiety disorder, separation anxiety disorder, panic disorder, and simple and social phobias. Some estimates suggest that up to 60% of children with TD will experience major depression at some point, and about 40% will have a non-OCD anxiety disorder. Vermilion and team compared anxiety symptoms in anxious youth with and without tic disorders. The anxiety profile in youth with tics appears to be different from those without tics. Those with tics are more likely to experience separation anxiety and physical symptoms of anxiety than anxious youth without tics.

Coffey and colleagues reported that comorbid mood disorders were the most robust predictors of psychiatric hospitalization and illness severity in a cohort of TD patients. Tic severity was only marginally significant as a predictor of psychiatric hospitalization. In a controlled study of older adolescents with TD, Gorman and colleagues found that, independent of comorbidity with ADHD, TD patients had significantly higher rates of depression and poorer overall outcome than healthy controls.

More recently, the work of Rizzo and peers demonstrated that depression in those with TD is associated with tic severity, the presence of comorbid ADHD, co-occurring anxiety, disruptive behavior, and a family history of depression. Whether mood symptoms are primary (i.e., related to the underlying pathophysiology of TD) or secondary to the demoralization and/or impairment related to having a chronic illness remains to be further clarified.

A substantial minority of youth with TD evaluated in clinical settings may meet criteria for intermittent explosive disorder manifested by rage attacks or oppositional defiant disorder. Budman and colleagues reported that children with TD as well as ADHD and OCD are more likely to experience explosive outbursts, and the number of outbursts is positively correlated with the number of comorbid conditions. Tic type and severity are not linked to the presence of explosive outbursts.

Learning and academic problems are common in children with TD; in a database of 5,450 patients with TD, 1,235 (22.7%) had learning disabilities. Young patients with TD may experience problems with executive functions, such as planning, persistence, organizational skills, and social problem-solving. Patients with TD have been consistently shown to have difficulties with procedural learning, fine motor control, motor inhibition, nonverbal memory, and visual motor integration. In addition, handwriting may be poor. Comorbid ADHD confers a greater risk of learning and academic problems in those with tic disorders.

Many clinically referred children with TD are reported to have difficulty with sensory-perceptual functioning, such as extreme sensitivity to touch or sound. Investigators have hypothesized that dysregulation of central sensory integration and/or sensory gating problems may result in these sensitivities, but more research is needed. Some of these children may have comorbid autism spectrum disorder (ASD), which occurs at higher rates in children with TD than in the general population. Estimates range from 6% to 16%. Emerging research examining other conditions that may frequently co-occur with tic disorders includes stuttering, sleep disturbances, migraine headaches, asthma, and metabolic and cardiovascular disorders.

Mark was a 13-year-old boy referred by his pediatrician for consultation for tics and anxiety. Mark developed the onset of collar chewing at around age 4 or 5 years. Between ages 5 and 6 years, he began to make sniffing and snorting sounds, and at age 6 or 7 years, he developed eye blinking and head and neck movements. Mark was subsequently referred at age 8 years to a pediatric neurologist who diagnosed TD. No treatment was initiated at that time because Mark's tics were not causing any distress or impairment. His parents reported that since age 10 years, the severity of Mark's tics had diminished. He was currently repeating words and making guttural sounds and bird noises at times.

His parents reported that Mark had a long history of anxiety. As a very young child, he was anxious about separation for preschool. He was described as a child who was slow to warm, with a cautious temperament, waiting until he was familiar with a situation to enter a group. Over the years, teachers commented that they would like to see more class participation from him. His parents reported that Mark often preferred to stay at home rather than go out to see friends on play dates.

Mark reported that he did not like to speak in public because he was afraid that what he would say would "not be perfect." Mark described ruminating on "things he should have said or done differently." Mark always worried about performance, about his mother dying, and about bad things happening. Mark had slept with his parents until the age of 10 years because of a history of nightmares.

Mark had concerns about contamination, and he used Purell and washed his hands frequently. Mark hoarded objects and preferred to do things a certain number of times; in his backpack, he kept 20 books, 20 blue pencils, and 20 black pencils. In the past, he used to sharpen pencils excessively to get them to a "just right" point.

Mark was the product of a full-term, uncomplicated pregnancy; developmental milestones were described as entirely within normal limits. Currently in eighth grade, he had attended private school since kindergarten and was an A student. Mark enjoyed sports and spending time with friends.

There was no family history of TD or OCD. Mark's father had been treated for anxiety. A paternal uncle had a long-standing history of eye blinking. Mark had no history of hospitalization, head injury, seizures, allergies, or other major medical problems.

MENTAL STATUS EXAMINATION

Mark was a slender, peripubertal boy in age-appropriate casual clothing. He was very restless. There were multiple motor and vocal tics, including eye blinking, shoulder shrugging, facial grimacing, mouth opening, coughing, and grunting. Mark showed a range of affect and was pleasant and cooperative. Mark reported that his tics did not really cause him any problems, but he worried a lot.

IMPRESSION

In summary, Mark was a 13-year-old peripubertal boy with a long history of multiple motor and vocal tics consistent with TD. Given the contamination concerns, obsessional worries and ruminations, and the tendency to count and arrange, Mark also met the criteria for OCD. In addition, there were symptoms suggestive of social anxiety disorder, given his reluctance to speak up in class, his slow-to-warm-up behavior, and his history of shyness. The excessive worry about performance was suggestive of a generalized anxiety disorder.

TREATMENT PLAN

Treatment of Mark's anxiety was indicated because it appeared to be pervasive and was causing distress. The recommendation was to start cognitive-behavioral therapy to address both his social anxiety and obsessive-compulsive behaviors. Given that Mark was neither in distress nor impaired by his tics, and was probably past his lifetime peak severity, simple monitoring of the tics was recommended. It was likely the tics would decrease once the anxiety was treated.

TREATMENT

Individuals with mild and even some moderate tics may not need treatment if the tics are not causing significant distress or impairment. Simple, active monitoring may be all that is necessary for some individuals with mild tics. Indications for treatment of tics include physical pain or discomfort, interference with social interactions, and impairment in any aspect of educational or occupational functioning. The first step in treatment is proper education of the patient, family members, teachers, and others who interact with the patient. Parents, educators, and physicians must work as partners to advocate for the optimal school environment for children with tic disorders. National and local support groups, such as the Tourette Association of America (https://www.tourette.org), can provide additional information and can serve as a valuable resource for patients and their families.

Nonpharmacologic treatments have been developed that are as successful as pharmacotherapy. In fact, the first-line treatment for mild to moderate tics is cognitive-behavioral intervention. Medications are usually considered only when symptoms are unable to be addressed by psychotherapeutic interventions and/or interfere with academic or occupational performance, peer relationships, social interactions, or activities of daily living. Pharmacotherapy for tic disorders is individualized and tailored specifically to the needs of the patient; the most distressing and/or impairing symptoms are targeted first. The goal of treatment is not to eliminate tics but to relieve tic-related discomfort or distress and to achieve a degree of control of tics that allows the patient to function as normally as possible.

General Guidelines

Behavioral intervention is the cornerstone of effective treatment for tic disorders. Treatment goals should be to reduce symptoms, support adaptive functioning and strengths, and enhance developmental progress. Self-esteem should always be supported, given that most patients with tic disorders will have at least some sense of personal and emotional vulnerability. Early in the diagnostic process, education in phenomenology and the natural history of tic disorders is essential.

Clarification that the patient's symptoms are not voluntary in most situations eases psychological burdens for both patients and families. This is especially important because families observe that tics, at times, can be suppressed. Struggles over whether the symptoms are tics or "negative behavior" are pointless and should be curtailed; a more practical goal involves personal management of, and responsibility for, symptoms and behavior, regardless of their origin.

Containment is another cornerstone of treatment. Even with the use of effective medications, it is rare for tics to remit completely. The use of a "tic room" or a "time out" area provides an opportunity to contain problematic tics or compulsions and to "de-stimulate." Because emotional conflicts and stress frequently increase symptom intensity and frequency, time-limited withdrawal from stressful situations can be beneficial. A balanced, healthy diet and exercise regimen may contribute to stress reduction and overall well-being. Caffeine and excess amounts of sugar should be minimized because they might exacerbate tics in some patients.

Psychoeducational Approaches

Educating patients and their families is essential. Patients and their parents frequently experience relief at the time of diagnosis. There should be ongoing opportunities for discussion of patients' and families' questions and concerns. Referral to the Tourette Association of America and to its local chapters is recommended for all newly diagnosed patients and their families to maintain continuing education and support.

Behavioral Therapies

Behavioral treatment is a primary intervention for many tic disorder patients. Simple behavioral approaches can be used with the majority of TD patients to "de-stimulate" and contain the symptoms. Specific paradigms are indicated for obsessions, compulsions, and tics. Relaxation techniques, such as deep breathing, guided imagery, mindfulness, and the use of relaxation tapes, can be useful for the anxious TD patient.

Habit-reversal therapy or training (HRT) forms the essence of therapeutic intervention in a cognitive behavioral treatment known as comprehensive behavioral intervention for tics (CBIT) in youth and adults. The superiority of CBIT over psychoeducation and supportive therapy alone is well established; CBIT is considered the first-line treatment for mild to moderate tics when intervention is warranted. The primary components of CBIT include awareness training, competing response training, relaxation training, contingency management, social support, and relapse prevention. Awareness training teaches patients to identify the premonitory urges or sensations that typically precede tics, as well as to recognize the situations in which tics are likely to occur. The second component of CBIT, teaching the patient a competing response to the targeted tic, has been developed to prevent the emergence of the tic. Parents are taught to provide appropriate support and encouragement to the child in practicing the technique at home. Tic substitution (substituting a more socially tolerable tic for a less socially tolerable one) can be useful.

Due to barriers to accessibility of behavioral interventions for tics, Singer and co-workers examined home-based parent-administered HRT. Their study found the home-based intervention to produce tic reduction comparable to in-person behavior intervention with a therapist. However, there were more dropouts from the parent-administered therapy group. Various interactive online treatment programs based on CBIT are now available for youth with tic disorders.

Behavioral interventions are also critical for the management of oppositional, defiant, or explosive outbursts, which are not uncommon in clinically referred youth with tic disorders. Scahill and colleagues demonstrated that parent management training is helpful in improving disruptive and explosive behavior problems in children with tic disorders, and Sukhodolsky and team have developed anger control training (ACT) for adolescents with TD and disruptive behavior. In children and adolescents, cognitive-behavioral paradigms are also the cornerstone of the treatment of mood and anxiety disorders, which often co-occur in clinically referred individuals with TD.

Psychotherapy

Individual, group, and family therapies are supportive adjuncts to behavioral interventions and pharmacotherapy. Individual supportive therapy is indicated for patients having difficulty adjusting to the disorder, to their peers or family, or with school or occupational functioning. It is particularly useful when there is evidence of moderate stress or anxiety or another comorbid psychiatric disorder responsive to this form of psychotherapy (e.g., mild depression). It can help to restore or maintain self-esteem and promote mastery and coping. Furthermore, supportive psychotherapeutic interventions may facilitate the normal developmental tasks of friendship development, school adjustment, and day-to-day stress management.

Family therapy is extremely useful in dealing with the many complex interpersonal and family issues that arise for children with TD. At a minimum, both parents should be involved in the initial evaluation. Ideally, family members can be sources of support and nurturance. Ongoing family therapy is indicated when family development (i.e., growth and maturation) has slowed or halted because of the focus on the patient with the tic disorder. It may also help with maladaptive reactions from or to siblings or with specific symptoms that are affecting the entire family.

Group therapy is another important adjunct treatment. Support and education for parents and family members of patients with TD can be found through the Tourette Association of America, which has created a well-organized network of state and community support groups. Informal, unstructured activities can be arranged through local chapters of the Tourette Association of America. Formal, structured support is also possible through social skills groups.

School and Occupational Interventions

Learning problems and classroom difficulties occur commonly in tic disorder patients. Specific developmental disorders and ADHD- or

OCD-related symptoms may interfere with academic performance.

Clinicians should be available to provide consultation, guidance, and education to teachers or employers of patients with tic disorders. Useful special educational interventions include creating a moderate, task-oriented structure in the classroom, preferential seating, one-to-one support, and individualized educational or work plans. Flexibility is a key element in educational and occupational intervention. To promote tic control, there may be a need for optional “time outs” from class or work settings. Time limits may have to be extended or eliminated for exams. Some patients may also benefit from writing aids, such as laptops. Laptop and tablet computers may allow patients with handwriting difficulties to take notes in class without disturbing others. Adaptive, as contrasted to regular physical education, can be particularly helpful for children with TD. This form of physical education can be arranged through special education departments in local schools. Programs can be tailored to each child’s needs, strengths, and challenges.

Specific workplace interventions include structured tasks, organization of tasks into smaller units, flexible time limits, and ample physical space.

Pharmacotherapy

The decision to treat tics and/or comorbid symptoms with medication should be based on a comprehensive evaluation. Patients with symptoms that significantly interfere with adaptation to family, school or work life, peer relationships, or developmental progress should be treated, especially if psychotherapeutic options are either unavailable or not optimally beneficial. Patients who suffer significant emotional distress or inadvertently cause distress to significant others in their lives may also be candidates for treatment, even when symptoms are relatively mild. Most patients with very mild symptoms need only monitoring, education, guidance, and support. Moderate to severe symptoms should usually be treated aggressively.

The clinical workup before initiating pharmacologic treatment should include a complete physical examination and neurologic screening, psychiatric evaluation, vital signs including height, weight, and body mass index (BMI), an Abnormal Involuntary Movement Scale (AIMS) screening, basic laboratory screening (including a complete blood count with differential, urinalysis, and a comprehensive metabolic panel including liver and thyroid function tests, fasting blood glucose, and metabolic screen including cholesterol and triglycerides), and an electrocardiogram (ECG).

Medication should be initiated at the lowest possible dose, and dose titration should be gradual in general. Most maximum doses will be low (compared to doses needed for other indications for these same medications) in TD patients. When possible, a single medication (monotherapy) should be used initially. At times more than one agent will need to be prescribed simultaneously (targeted combined pharmacotherapy) when monotherapy has not been effective or has resulted in limiting adverse effects or when target symptoms include more than one disorder.

Determining an adequate duration for a medication trial for patients with tics can be quite challenging because the natural history of TD is characterized by the waxing and waning of symptoms over time. A 1-month baseline observation period before treatment is initiated is recommended, if feasible. In general, it is best to wait at least several weeks after changing a dose before making any conclusions about a therapeutic response. External stressors must always be considered as potential determinants of increased symptomatology during medication trials. Comorbid conditions, such as ADHD and OCD, which significantly affect the individual, should be targeted aggressively because it is well established that these conditions are often more impairing than tics, and they respond to treatment.

α-Adrenergic Agonists. The α-adrenergic agents are recommended as a first-line pharmacologic treatment for most patients with mild to moderate TD, especially in those with comorbid ADHD, in whom they are more effective. Clonidine (Catapres), an α_2-adrenergic (presynaptic) agonist, has been used for over three decades for the treatment of TD. When given in low doses, it is postulated that presynaptic noradrenergic effects mediate the observed clinical improvement in motor and vocal tics. Clonidine activates presynaptic autoreceptors in the locus coeruleus to reduce norepinephrine release and turnover in the cerebral cortex. In addition, clonidine also reduces the disinhibition, impulsivity, inattention, and hyperactivity often present in young TD patients. Data supporting its use have been mixed, but there have been well-designed studies that demonstrate clinical effectiveness.

Clonidine should generally be started at 0.025 mg daily to twice daily for prepubertal children and increased by 0.025 mg every 5 to 14 days. Prepubertal patients will generally need dosing three or four times daily. Adolescents or adults can be started on 0.05 mg daily and increased by 0.05-mg increments to twice-daily dosing. The total daily dose typically is 0.05 mg to 0.45 mg (i.e., up to 8.0 μg/kg). Common side effects include sedation, bradycardia, orthostatic hypotension, headache, dry mouth, stomachache, midsleep awakening, and irritability. Sedation, the most common and sometimes limiting side effect, usually abates over several weeks; when it does not, dose reduction may be helpful. Hypotensive effects are minimal in this dose range, but blood pressure, pulse, and ECG should be monitored at baseline and follow-up visits. More concerning is the risk of rebound increases in blood pressure, anxiety, and tics if abrupt discontinuation of clonidine occurs. A transdermal form (Catapres TTS-1, TTS-2) is available and may be an alternative when patients require very frequent dosing and/or cannot tolerate the sedative effects. A long-acting oral formulation is also available (Kapvay) and designed to be administered twice daily.

Guanfacine (Tenex), an alternative α_2-adrenergic agonist, may also be efficacious for hyperactivity, impulsivity, and tics. Guanfacine activates postsynaptic prefrontal α-adrenergic cortical receptors. Prepubertal patients are typically started on 0.25 mg daily and increased by 0.25-mg increments every 5 to 14 days given twice daily; older patients and adults are typically started on 0.5 mg daily and increased by 0.5-mg increments to about 3.0 mg daily. Adverse effects, generally less significant and more tolerable than with clonidine, include sedation, fatigue, midsleep awakening/insomnia, irritability, stomach upset, constipation, and hypotension/lightheadedness. A long-acting, once-daily extended-release form (Intuniv) is also available. While previous uncontrolled studies and randomized control trials have supported the clinical use of guanfacine, a recent small, randomized, double-blind, placebo-controlled pilot study conducted by Murphy and colleagues found that there was no difference in efficacy between extended-release guanfacine and placebo.

Neuroleptic Agents. Since the late 1960s, when haloperidol was first introduced as a treatment for TD patients, haloperidol (Haldol) and pimozide (Orap) have been the only Food and Drug Administration (FDA)–labeled, formally approved agents for the treatment of TD. These typical neuroleptic medications block postsynaptic D_2 (dopamine) receptors in the basal ganglia, decreasing dopaminergic input from the substantia nigra and ventral tegmentum and thus reducing tics. These medications are generally effective in tic reduction; however, considerable variation exists in clinicians' and patients' thresholds for neuroleptic use. The decision to use a neuroleptic may be guided by both symptom severity and quality-of-life concerns. Significant adverse effects, including acute dystonic reactions, oculogyric crises, torticollis, Parkinsonism, akathisia, tardive dyskinesia, weight gain, sedation, dry mouth, blurred vision, galactorrhea, gynecomastia, constipation, urinary retention, social phobia, loss of drive/energy/personality, and electrocardiographic changes including tachycardia and QTc prolongation, limit their use to second- or third-line treatments. In general, neuroleptic agents are not recommended for mild tics. Of the typical neuroleptics, haloperidol and pimozide have been the most investigated, with double-blind, placebo-controlled studies supporting their use.

The newer, atypical neuroleptics, or second-generation antipsychotic agents, are recommended as a first-line treatment when a neuroleptic is indicated for moderate to severe tics. These agents, which block both D_2 dopamine and 5-HT_2 serotonin receptors, have the potential advantage of fewer extrapyramidal adverse effects (EPSs) (acute dystonic reactions, oculogyric crises, torticollis, Parkinsonism, akathisia, tardive dyskinesia). Their dual pharmacologic activity appears to protect against the neurologic adverse effects associated with typical neuroleptic agents. Among them, risperidone (Risperdal) has been found to be efficacious for tics. Risperidone is a 5-HT2 receptor antagonist at low doses and a D2 antagonist at high doses. It acts at the α-2, D3, D4, and H1 receptors with moderate to high affinity as well. Multiple trials, including randomized, controlled trials, support risperidone's efficacy. It has been shown to be superior to placebo and equally effective as pimozide and clonidine for tic reduction. Doses are typically in the 1 to 4 mg range. Common adverse effects include appetite increase, weight gain, lipid and glucose metabolism abnormalities, and sedation. Sleep disturbance, social phobia, galactorrhea, gynecomastia, menstrual irregularities, and erectile dysfunction have also been reported.

Aripiprazole (Abilify) is a novel atypical neuroleptic with partial agonist-antagonist effects on the dopamine and serotonin system. It is a partial D2, D3, and D4 receptor agonist as well as a partial 5-HT1A and 5-HT2C receptor agonist. Several published placebo-controlled studies and open-label trials have reported effectiveness and tolerability in TD. A meta-analysis of ten studies with a total of 302 patients conducted by Liu and peers confirms that aripiprazole is safe and effective for the treatment of tics in children and adolescents. The most common adverse effects include weight gain, akathisia, and sedation. Aripiprazole became the most recent and only the third agent to be FDA-approved for the treatment of tics in 2014.

Additional atypical options with some evidence base for efficacy in TD include ziprasidone (Geodon), olanzapine (Zyprexa), and quetiapine (Seroquel). Results of one placebo-controlled trial of ziprasidone indicated that it was as efficacious as risperidone for tic reduction. Weight gain is typically less likely than that with other neuroleptic agents. Ziprasidone carries the added risk of altered cardiac conduction, especially QTc prolongation. An ECG is recommended before initiating treatment, during dose adjustments, on reaching the maintenance dose, and annually thereafter. Other adverse effects include somnolence, sedation, and akathisia.

Evidence for the effectiveness of olanzapine is supported by open-label and case-report studies, but its significant association with weight gain, the highest among the atypical neuroleptics, limits its use. Only two open-label studies demonstrate quetiapine's effectiveness for treating tics. Sedation and weight gain are the most common adverse effects.

Fasting glucose and a lipid panel are recommended before initiating treatment with a second-generation antipsychotic, as well as at the maintenance dose and at regular intervals thereafter. Weight, waist circumference, and BMI should also be monitored. An ECG may be helpful, especially when there is a known individual or family cardiac history.

Neuroleptics should be initiated at the lowest possible dose and increased very gradually. The starting dose for aripiprazole is between 1 to 2 mg per day with maximum doses of 10 to 20 mg daily. Risperidone is typically started at 0.125 to 0.25 mg daily to twice daily and titrated upward over weeks to about 1 to 2 mg daily in prepubertal children; adolescents are typically started on 0.25 mg daily to twice daily. For haloperidol, children are typically started at 0.125 to 0.25 mg daily to twice daily and increased by 0.125- to 0.25-mg increments weekly or biweekly. Adults are usually started on 0.5 mg with weekly or biweekly increases of 0.5 mg. Pimozide is an alternative D_2-blocking agent, which may cause fewer side effects than haloperidol. However, because pimozide also has calcium-channel–blocking properties, normal cardiac function (ascertained from cardiac assessment and an ECG) is a prerequisite to its use. Pimozide has been shown to prolong QTc intervals, and the risk for cardiac conduction abnormalities may increase when pimozide is combined with drugs that inhibit the cytochrome P450 3A4 isozyme. An ECG is also recommended during dose adjustments, on reaching the maintenance dose, and annually thereafter. Approximately half as potent as haloperidol, pimozide should be started at 0.5 mg to 1.0 mg for children and at about 1.0 mg for adolescents or adults. Most patients require less than 10 mg daily.

Common adverse effects of neuroleptics include weight gain, sedation, muscle cramps, and stiffness. Dose reduction or switching to an alternative drug is the best way to address these adverse effects. Fortunately, EPSs do not appear to be common sequelae from neuroleptic use at usual doses in tic disorder patients; nevertheless, clinicians should be alert to the possibility of these reactions and consider appropriate interventions (e.g., dose reduction, anti-Parkinson agents) on a case-by-case basis. EPSs may be less common with the atypical agents than with the typicals. Tardive dyskinesia, although probably not common, has been described in patients with TD and may be difficult to differentiate from tics.

Weight gain is likely with both typical and atypical neuroleptics, but the risk appears to be greater with the use of atypical agents such as olanzapine and risperidone. The possibility of weight gain should be addressed proactively in all youth who are candidates for neuroleptics. Education at the outset of treatment regarding the need for a healthy diet and regular exercise is helpful; at times nutritional consultation may be indicated.

Tiapride, sulpiride, and amisulpride are D_2-receptor–blocking neuroleptics used in tic treatment in Europe but not available in the United States. Tetrabenazine and deutetrabenazine, nonneuroleptic dopamine antagonists that act by inhibiting central vesicular monoamine transporter type 2 (VMAT2), may be useful for the treatment of tics but are currently FDA-approved only for the treatment of chorea in Huntington disease and tardive dyskinesia. There are two open-label studies of tetrabenazine for tics with limited demonstrated effectiveness, and more study is needed. A recent randomized, double-blind, placebo-controlled study of deutetrabenazine in children and adolescents with TD did not demonstrate efficacy for the management of tics. Although results indicated active drug did not separate from placebo, it was well tolerated and was not associated with depressive symptoms (unlike tetrabenazine).

Selective Serotonin Reuptake Inhibitors. Serotonergic agents such as fluoxetine (Prozac), paroxetine (Paxil), fluvoxamine (Luvox), citalopram (Celexa), escitalopram (Lexapro), and sertraline (Zoloft) are first-line agents for pharmacologic treatment of significant OCD, anxiety disorders, and mood symptoms. All the selective serotonin reuptake inhibitors (SSRIs) except citalopram and escitalopram have been studied in placebo-controlled investigations in children with OCD. Although there is no evidence to support the use of SSRIs for tics, their use may be warranted. Since many TD patients have obsessions, compulsions, anxiety, and depression, and neuroleptics and α_2-agonists do not effectively alter these specific symptoms, the SSRIs may be beneficial in some patients. In addition, complex motor tics such as repetitive squatting and touching may be very similar to compulsive rituals. Furthermore, stress and anxiety are known to exacerbate tics, and the use of SSRIs to reduce anxiety may indirectly serve to reduce tics.

Fluoxetine has received the most study in TD, but efficacy is likely relatively comparable among serotonergic agents. The choice typically rests with consideration of the adverse-effect profiles of the agents. Often, higher doses of SSRIs are required for the treatment of OCD symptoms than for anxiety disorders or depression, and many patients with co-occurring OCD and a tic disorder respond less well to SSRIs. The addition of small doses of a neuroleptic or an atypical neuroleptic, which increases the response to the SSRIs, may be required.

Mild OCD symptoms often respond to cognitive behavioral therapy alone, which is equally effective for OCD symptoms with or without comorbid tics. When obsessions or compulsions are of moderate to severe intensity or impair adaptive functioning, SSRIs are potentially indicated in addition to therapy. Fluoxetine is typically started at 2.5 to 5.0 mg daily for children and at 5.0 to 10 mg for adolescents or adults. Dose ranges are not well established for children and adolescents with TD. It appears that most children will respond to 10 to 20 mg daily; adults may require 20 to 80 mg daily. Fluoxetine, in combination with a neuroleptic, can be useful for both tics and obsessive-compulsive symptoms; neuroleptic doses may require reduction or adjustment because of the cytochrome P450 enzyme–inhibiting properties of fluoxetine. Common adverse effects include activation, insomnia, headaches, anorexia, and weight loss.

Sertraline, fluvoxamine, citalopram, and escitalopram are also options. There is also considerable data available on the use of clomipramine (Anafranil), a tricyclic antidepressant, in OCD, but adverse effects may be more problematic. These medications may represent alternatives for patients who experience limiting side effects on fluoxetine. Current recommendations for the treatment of comorbid tics and OCD are to use the appropriate agent indicated for each symptom. For example, an α-agonist or atypical neuroleptic can be combined with an SSRI to target significant tics and OCD symptoms.

Table 47–4.
Pharmacotherapies for Tourette and Tic Disorders

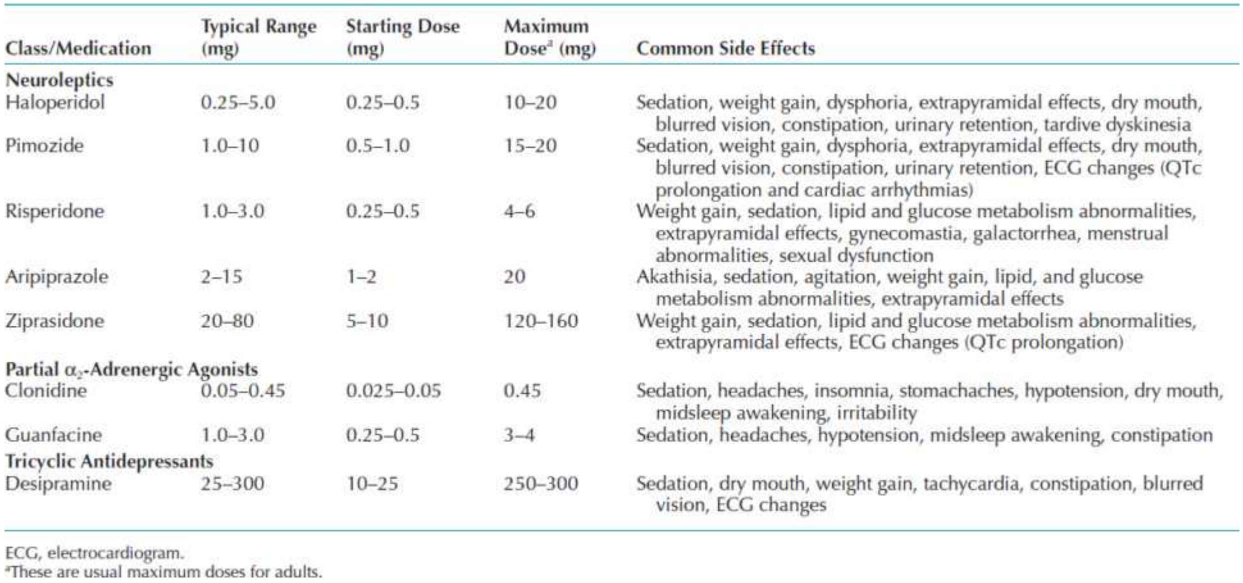

Class/Medication	Typical Range (mg)	Starting Dose (mg)	Maximum Dose[a] (mg)	Common Side Effects
Neuroleptics				
Haloperidol	0.25–5.0	0.25–0.5	10–20	Sedation, weight gain, dysphoria, extrapyramidal effects, dry mouth, blurred vision, constipation, urinary retention, tardive dyskinesia
Pimozide	1.0–10	0.5–1.0	15–20	Sedation, weight gain, dysphoria, extrapyramidal effects, dry mouth, blurred vision, constipation, urinary retention, ECG changes (QTc prolongation and cardiac arrhythmias)
Risperidone	1.0–3.0	0.25–0.5	4–6	Weight gain, sedation, lipid and glucose metabolism abnormalities, extrapyramidal effects, gynecomastia, galactorrhea, menstrual abnormalities, sexual dysfunction
Aripiprazole	2–15	1–2	20	Akathisia, sedation, agitation, weight gain, lipid, and glucose metabolism abnormalities, extrapyramidal effects
Ziprasidone	20–80	5–10	120–160	Weight gain, sedation, lipid and glucose metabolism abnormalities, extrapyramidal effects, ECG changes (QTc prolongation)
Partial α_2-Adrenergic Agonists				
Clonidine	0.05–0.45	0.025–0.05	0.45	Sedation, headaches, insomnia, stomachaches, hypotension, dry mouth, midsleep awakening, irritability
Guanfacine	1.0–3.0	0.25–0.5	3–4	Sedation, headaches, hypotension, midsleep awakening, constipation
Tricyclic Antidepressants				
Desipramine	25–300	10–25	250–300	Sedation, dry mouth, weight gain, tachycardia, constipation, blurred vision, ECG changes

ECG, electrocardiogram.

[a] These are usual maximum doses for adults.

Tricyclic Antidepressants. Some studies of desipramine (Norpramin) have suggested potential efficacy for the treatment of tics as well as ADHD symptoms in patients with comorbid ADHD. Doses are not established but may parallel doses used in ADHD without tics. Careful cardiovascular monitoring must take place, including blood pressure, pulse, and baseline and follow-up ECGs (Table 47–4).

Clomipramine is used for the treatment of OCD when the patient has failed to respond or cannot tolerate two trials of an SSRI. It can be started at 10 to 25 mg daily in children and 25 to 50 mg in adolescents or adults. Total dose ranges are typically 100 to 200 mg daily; blood levels may be obtained, but no therapeutic range has been established. Common adverse effects are those seen typically with other tricyclic antidepressants, including sedation, dry mouth, constipation, blurring of vision, weight gain, tachycardia, and ECG changes.

Benzodiazepines. Benzodiazepines have not been well investigated in tic treatment but may hold some promise (e.g., clonazepam [Klonopin]). Some studies suggest a reduction in tics independent of anxiolytic effects. This effect may also result from a reduction in anxiety that, in turn, could reduce tic frequency. In addition, benzodiazepines can be used to treat comorbid anxiety symptomatology or panic attacks that can also occur in TD patients. Doses are not established but are often quite low (e.g., 0.125 mg daily to twice daily for clonazepam). Adverse effects of particular concern in the juvenile population are sedation, disinhibition, cognitive impairment, and dependence.

Stimulants. Stimulants are first-line agents for the treatment of ADHD. There has been controversy over the role of stimulants in TD and whether their use causes or exacerbates tics; however, recent studies have demonstrated that TD patients with clinically significant ADHD symptoms may be candidates for stimulants. In the first systematic investigation of methylphenidate use by Gadow and colleagues, behavior improved, and tics did not worsen at moderate doses of 0.1 to 0.5 mg/kg methylphenidate. When group data are analyzed, there is no significant increase in tics when stimulants are used in patients with tic disorders; however, individual patients may experience at least a transient increase in tics. Interestingly, in some studies, stimulant use is associated with a decrease in tics.

Stimulants are equally effective in improving ADHD symptoms with or without comorbid tics. A randomized, controlled study of clonidine and methylphenidate in ADHD and chronic tics conducted by the Tourette Syndrome Study Group reported that the combination and each drug alone was more effective than placebo in reducing both tics and ADHD symptoms. An increase in tics did not occur in the methylphenidate group to any extent more significantly than in the placebo group. Additionally, a meta-analysis examining pharmacotherapy for youth with comorbid ADHD and tics demonstrated methylphenidate, α_2 agonists, desipramine, and atomoxetine (Strattera) were effective for ADHD symptoms. Supra-therapeutic doses of dextroamphetamine appeared to worsen tics. There was no short-term evidence that methylphenidate exacerbated tics.

A more recent meta-analysis of 22 randomized, double-blind, placebo-controlled trials involving 2,385 children with diagnosed ADHD conducted by Cohen and colleagues demonstrated that the risk of new-onset or worsening of tics associated with stimulant use (5.7%) was similar to that observed with placebo (6.5%). Age, type or dose of stimulant used, and duration of treatment did not affect risk.

A judicious approach to the treatment of ADHD in the context of a tic disorder is to evaluate whether tics or ADHD symptoms are causing the most distress or impairment. If ADHD is more problematic, it is reasonable to start with a methylphenidate-based stimulant. Stimulants should be initiated in low doses and gradually increased, monitoring for any exacerbation of tics. If tics are more problematic, an alpha agonist is recommended first. Both medications can be safely combined if necessary. Common adverse effects of stimulants include insomnia, appetite suppression, and weight loss.

Nonstimulant ADHD Treatment: Atomoxetine. Atomoxetine, a selective norepinephrine reuptake inhibitor, is approved for treatment of ADHD in children, adolescents, and adults. It has also been studied in youth with ADHD and chronic tics. One study reported that atomoxetine improved both ADHD symptoms and tics in children and adolescents with comorbid ADHD and TD. It is dosed based on body weight and titrated from a daily starting dose of 0.5 mg/kg to 1 to 1.5 mg/kg. Therapeutic response, unlike stimulants, takes a few weeks. Common adverse effects include headache, nausea, fatigue, and stomachaches.

Other Alternatives. Many other medical treatments have been used in TD patients. Immune therapies such as IVIG, plasmapheresis, and corticosteroids have been explored in patients with PANDAS/PANS. These remain experimental treatments under investigation. Opioid antagonists (e.g., naltrexone) may be an alternative for patients with self-injurious behaviors. The dose range for naltrexone may be approximately 25 to 75 mg per day. Nicotine, both in gum and in the transdermal form, has also been studied in TD. Results have been equivocal. In a double-blind, placebo-controlled trial of 61 patients with TD at doses up to 7.5 mg daily, mecamylamine (Inversine), a nicotine antagonist, was found to be no more effective than a placebo. In addition, a randomized, controlled trial of pramipexole, a mixed dopamine agonist with presynaptic autoreceptor effects, failed to show efficacy.

Baclofen, a GABAergic muscle relaxant used to treat spasticity, has been studied in one large open-label and one controlled trial. In the controlled trial, there was no difference between drug and placebo in Yale Global Tic Severity Scale Total Tic score, but baclofen was superior to placebo with regard to tic-related impairment. A placebo-controlled study of omega-3 fatty acids derived from fish oil revealed no improvement in tic measures, but clinically meaningful improvement in tic-related impairment compared to placebo was reported.

The use of botulinum toxin to suppress tics by injection into discrete muscles involved in tic generation has been examined in several studies. Botulinum toxin blocks acetylcholine release at the neuromuscular junction and produces a temporary, reversible diminution of muscle activity, which may last weeks to months for dystonic tics. Placebo-controlled and open-label trials have demonstrated efficacy for isolated tics. Botulinum toxin may be the most useful agent for the treatment of solitary, focal dystonias. There are reports of suppression of ear-wiggling tics and laryngeal tics. Adverse effects include soreness at the injection site and muscle weakness.

Targeted combined pharmacotherapy—the specific combination of two or more medications at one time—has evolved as a treatment option. Ideally, this approach should be used only after adequate trials of monotherapy have failed. However, many patients with TD also have comorbid ADHD or OCD, and often both, and frequently monotherapy addresses one problem but not the other. Judicious use of SSRIs and neuroleptics for TD and comorbid OCD is one example of this approach; in some patients with OCD and tics, this approach may be more effective than monotherapy. Another example is the combination of methylphenidate and clonidine to target tics and ADHD symptoms. In general, for comorbid tic disorders and OCD or ADHD, each condition should be treated with its respective first-line option.

Surgical Treatment. Initial reports on surgical interventions for TD appeared in the 1960s. Stereotaxic neurosurgical techniques have since been used in a few cases of severely impaired adults. Bilateral capsulotomy was associated with at least 50% improvement in both motor and vocal tics in 12 patients who had TD, with no severe adverse effects. In 1999, deep brain stimulation (DBS) was introduced for the treatment of intractable tics. Several patients have since undergone bilateral thalamic stimulation with promising results. However, given the significant medical risks of these procedures, there has been limited experience with use of these techniques in youth, and they have generally been reserved for severe and refractory tics that significantly interfere with function and persist into adulthood.

Transcranial magnetic stimulation involves generating a brief, powerful magnetic field by a small coil positioned over the skull, inducing an electrical current in the brain. This noninvasive approach has been demonstrated to effect long-term changes in cortical excitability, which may be abnormal in tic disorders. It is an exciting potential future treatment for tics. Neural stem cell transplantation is another proposed treatment for tics, and studies are underway.

Jack was a 14-year-old boy referred by his mother for management of his TD. Jack developed the onset of skipping and clapping episodes when he was approximately age 7 years. This was followed by stretching movements of his mouth and nose accompanied by facial grimaces. His mother reported that Jack's tics remitted entirely until just before his 12th birthday when he began to "give the finger" for no apparent reason and began to curse using the "F" word. Children often teased him, and he was in significant distress about this.

Jack was the product of a full-term, uncomplicated pregnancy; developmental milestones were described as within normal limits. Jack's parents divorced when he was 2 years of age. Jack attended private school from prekindergarten through, at the time of referral, the ninth grade. He had difficulty with mathematics. He had a group of friends, including a best friend, and enjoyed music and video games.

Jack's maternal grandmother had OCD, and maternal great-uncles had eye-blinking tics. Jack had no history of head injury or seizures. Electroencephalographic and computed tomography scans were within normal limits. Past medication included adequate trials of clonidine, guanfacine, risperidone, and ziprasidone. Medications at the time of referral included haloperidol 4 mg, which led to excessive sedation, including somnolence in school.

MENTAL STATUS EXAMINATION

Jack was a moderately overweight adolescent boy. Of note were multiple motor tics, including pelvic thrusting, hand movements, outward leg twisting, head and neck thrusting, eye blinking, abnormal inhalation, chin hitting, shoulder licking, and genital scratching. In addition, there was copropraxia in the form of "giving the finger." There was also a change in his vocal inflection intermittently and a tendency to yawn on occasion.

IMPRESSION

In summary, Jack was a 14-year-old adolescent referred for management of his TD. His clinical picture was consistent with moderately severe TD. He had been treated with a variety of tic medications that were ineffective or intolerable.

TREATMENT PLAN

The recommendation was for a trial of aripiprazole. In addition, a trial of Comprehensive Behavioral Intervention for Tics (CBIT) was recommended. Jack and his parents were also counseled that at age 14, it was possible he was currently seeking treatment at his lifetime peak tic severity, and tic symptoms might begin to attenuate over the next few years.

SUMMARY AND FUTURE DIRECTIONS

The French neurologist Georges Gilles de la Tourette in 1885 was among the first to recognize and describe the neurodevelopmental disorder that now bears his name, as well as its psychiatric comorbid conditions. Since the first case descriptions of Gilles de la Tourette, a spectrum of tic disorders has come to be recognized and codified in psychiatric nomenclature. Tic disorders are complex neurodevelopmental disorders with heterogeneous phenotypes characterized by core disinhibition mediated by the CSTC tracts. Tic disorders are genetically determined in most cases. There is a growing appreciation and understanding of the role of psychiatric comorbidity in the diverse clinical presentation, and it is often the neurobehavioral, occupational, and psychosocial problems that co-occur in clinically referred individuals that warrant treatment rather than the tics per se. Treatment strategies should be multimodal in nature and include cognitive-behavioral, individual and family intervention, social and educational monitoring and support, and, where indicated, pharmacotherapy. Future investigation is needed to better understand the phenomenology, developmental psychopathology, and novel approaches to the treatment of tics. Tremendous progress has been made in the last few decades using innovative neuroimaging techniques, genetic studies, and clinical interventions. Understanding of the etiology, genetics, neuroanatomy, and epidemiology of tic disorders has been significantly enhanced. Novel treatment approaches have also emerged, providing alternatives to the pharmacologic management of tics. Targeted behavioral intervention has been proven to be as effective as medication and is now first-line treatment. DBS, botulinum toxin, targeted surgery, and stem cell transplantation are examples of promising new treatments. Complementary and integrative treatments such as music, exercise, and herbal remedies are also being investigated. The future holds tremendous promise as new insights into the etiology, pathophysiology, and treatment of tic disorders are rapidly developing.

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Centers for Disease Control and Prevention (CDC). Prevalence of diagnosed Tourette syndrome in persons aged 6–17 years—United States, 2007. *MMWR Morb Mortal Wkly Rep.* 2009;58(21):581–585.

Coffey BJ. Persistent tics, Tourette syndrome, and psychopathology: where are we now, and where are we going? *J Am Acad Child Adolesc Psychiatry.* 2017;56(4):281–283.

Coffey BJ, Biederman J, Geller DA, et al. Distinguishing illness severity in children and adolescents with Tourette's disorder. *J Am Acad Child Adolesc Psychiatry*. 2000; 39:556–561.

Coffey BJ, Jankovic J, Claassen DO, et al. Efficacy and safety of fixed-dose deutetrabenazine in children and adolescents for tics associated with Tourette syndrome. *JAMA Network Open.* 2021;4(10):1–13.

Cohen SC, Mulqueen JM, Ferracioli-Oda E, et al. Meta-analysis: risk of tics associated with psychostimulant use in randomized, placebo-controlled trials. *J Am Acad Child Adolesc Psychiatry.* 2015:54(9):728–736.

de la Cruz LF, Rydell M, Runeson B, et al. Suicide in Tourette's and chronic tic disorders. *Biol Psychiatry.* 2017;82(2):111–118.

Hirschtritt ME, Lee PC, Pauls DL, et al. Lifetime prevalence, age of risk, and genetic relationships of comorbid psychiatric disorders in Tourette syndrome. *JAMA Psychiatry.* 2015;72(4):325–333.

Houghton DC, Capriotto MR, Conelea CA, et al. Sensory phenomena in Tourette syndrome: their role in symptom formation and treatment. *Curr Dev Disord Rep*. 2014;1:245–251.

Huang AY, Yu D, Davis LK, et al. Rare copy number variants in NRXN1 and CNTN6 increase risk for Tourette syndrome. *Neuron.* 2015;94(6):1101–1111.e7.

Kawikova I, Leckman JF, Kronig H, et al. Decreased number of regulatory T cells suggests impaired immune tolerance in children with Tourette syndrome: a preliminary study. *Biol Psychiatry*. 2007;61:273–278.

Knight T, Steeves T, Day L, et al. Prevalence of tic disorders: a systemic review and meta-analysis. *Pediatr Neurol*. 2012;47:77–90.

Leckman JF, King RA, Bloch MH. Clinical features of Tourette syndrome and tic disorders. *J Obsessive Compuls Relat Disord.* 2014;3(4):372–379.

Leckman JF, King RA, Gilbert DL, et al. Streptococcal upper respiratory tract infections and exacerbations of tic and obsessive-compulsive symptoms: a prospective longitudinal study. *J Am Acad Child Adolesc Psychiatry*. 2011; 50:108–118.e3.

Liu Y, Ni H, Wang C, et al. Effectiveness and tolerability of aripiprazole in children and adolescents with Tourette's disorder: a meta-analysis. *J Child Adolesc Psychopharmacol.* 2016;26(5):436–441.

Mataix-Cols D, Isomura K, Pérez-Vigil A, et al. Familial risks of Tourette syndrome and chronic tic disorders: a population-based cohort study. *JAMA Psychiatry.* 2015;72(8):787–793.

Murphy TK, Fernandez TV, Coffey BJ, et al. Extended-release guanfacine does not show a large effect on tic severity in children with chronic tic disorders. *J Child Adolesc Psychopharmacol*. 2017;27(9):762–770.

Pérez-Vigil A, Fernández de la Cruz L, Brander G, et al. Association of Tourette syndrome and chronic tic disorders with objective indicators of educational attainment: a population-based sibling comparison study. *JAMA Neurol.* 2018:75(9):1098–1105.

Peterson BS, Choi HA, Hao X, et al. Morphologic features of the amygdala and hippocampus in children and adults with Tourette syndrome. *Arch Gen Psychiatry*. 2007;64:1281–1291.

Piacentini J, Woods DW, Scahill L, et al. Behavior therapy for children with Tourette disorder: a randomized controlled trial. *JAMA*. 2010;303:1929–1937.

Quezada J, Coffman KA. Current approaches and new developments in the pharmacological management of Tourette syndrome. *CNS Drugs*. 2018;32(1):33–45.

Rizzo R, Gulisano M, Martino D, Robertson MM. Gilles de la Tourette syndrome, depression, depressive illness, and correlates in a child and adolescent population. *J Child Adolesc Psychopharmacol.* 2017;27(3):243–249.

Robertson MM, Eapen V, Singer HS, et al. Gilles de la Tourette syndrome. *Nat Rev Dis Primers.* 2017;3(16097):1–20.

Scahill L, Specht M, Page C. The prevalence of tic disorders and clinical characteristics in children. *J Obsessive Compuls Relat Disord.* 2014;3(4):394–400.

Singer HS, McDermott S, Ferenc L, Specht M, Mahone EM. Efficacy of parent-delivered, home-based therapy for tics. *Pediatr Neurol.* 2020;106:17–23.

Storch EA, Lack CW, Simons LE, et al. A measure of functional impairment in youth with Tourette's syndrome. *J Pediatr Psychol*. 2007; 32:950–959.

Sukhodolsky DG, Vitulano LA, Carroll DH, et al. Randomized trial of anger control training for adolescents with Tourette's syndrome and disruptive behavior. *J Am Acad Child Adolesc Psychiatry.* 2009;48(4):413–421.

The Tourette Syndrome Association International Consortium for Genetics. Genome scan for Tourette disorder in affected-sibling-pair and multigenerational families. *Am J Hum Genet*. 2007;80:265–272.

Vermilion J, Pedraza C, Augustine EF, et al. Anxiety symptoms differ in youth with and without tic disorders. *Child Psychiatry Hum Dev.* 2021;52:301–310.

Willsey AJ, Fernandez TV, Yu D, et al. De Novo coding variants are strongly associated with Tourette disorder. *Neuron.* 2017;94(3):486–499.e9.

48

Elimination Disorders

Edwin Jens Mikkelsen, M.D.

ENURESIS

The disorder of enuresis has been recognized for centuries. However, substantial progress in the understanding of the disorder did not begin to occur until there were effective treatments. The first effective treatment was the application of the bell-and-pad method of conditioning, which was subsequently followed by pharmacologic treatment in the form of imipramine (Tofranil) and, more recently, desmopressin acetate (DDAVP). Each of these advances has led to etiologic research that might explain the efficacy of the particular intervention. This chapter reviews the accumulated clinical knowledge concerning enuresis, as well as the emerging research that may lead to a greater understanding of the disorder.

Definition

Enuresis is defined as the involuntary or intentional voiding of urine. The severity is determined by the frequency of urination; quantity is not a diagnostic consideration per se. Quantity can become a factor in treatment decisions if a child emits only small quantities of urine; in actual practice, quantity usually does not figure heavily into the treatment plan. Frequency, however, can be important in planning a hierarchy of treatment approaches. The major diagnostic qualifier relates to the age of onset. Finally, the definition precludes a physical cause for the disorder.

The length of time before continence is considered established varies in the literature between 6 months and 1 year. The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) and its text revision (DSM-5-TR) require a chronologic age of 5 years or, alternately, a developmental level equivalent to that age. In practical clinical terms, children who have never maintained continence for more than 1 year are referred to as having *primary enuresis*, whereas children who have achieved continence for 1 year or longer and then resume wetting are referred to as having *secondary enuresis.* The DSM-5-TR requires a frequency of at least two episodes a week for 3 consecutive months. However, it allows an exception for distress or impairment in functions. This means that an individual can receive the diagnosis if they are deemed to experience significant distress and do not meet the frequency criteria.

The disorder is also further defined according to the timing of the episodes throughout the day. Episodes occurring only at night are referred to as *nocturnal*; daytime wetting is labeled *diurnal.* The term "nocturnal and diurnal" refers to a combination of these two subtypes. Most children display only nocturnal enuresis, but some manifest diurnal or nocturnal and diurnal patterns.

History

A scholarly review of the history of enuresis found references dating back about 3,550 years to the Papyrus Ebers. The term derives from the Greek word *enourein*, which means "to void urine." In retrospect, many of the treatment approaches that were used in the past now appear almost sadistic in nature. Although some of these interventions may have been misguided but well-intentioned therapies (as in the use of leeching for other disorders), they may also reflect the commonly held view that the disorder is somehow under the (direct or indirect) control of the child and that punishment would lead to its cessation.

A confluence of somewhat disparate observations contributed to the evolution of the current understanding of enuresis. Although each of these is discussed in more detail in the following sections, a brief reference to them here will provide an overview of the various lines of research and their relative contributions to the present understanding of the disorder. The long-standing observation that enuresis tends to run in families evolved into the observation that a positive family history was a positive prognostic indicator for response to pharmacologic interventions and, ultimately, to sophisticated research with large pedigrees pointing toward specific genetic loci and modes of transmission. The observation that imipramine was an effective treatment for enuresis ushered in the modern era of research because the efficacy of imipramine suggested biologic contributions to the etiology of the disorder. Imipramine has now been largely supplanted by DDAVP as a pharmacologic treatment modality, and related research into its mechanism of action has contributed interesting insights into the pathophysiology of the disorder. Juxtaposed with the research into pharmacologic treatment and organic contributions to the disorder has been the impressive literature, which now dates back several decades, indicating that behavioral treatment in the form of the bell-and-pad method of conditioning is equally effective as the two leading pharmacologic interventions for the disorder. The observation that the wetting episodes usually occur at night also led to decades of research investigating the relation between sleep architecture and the occurrence of enuretic events. Although none of these lines of research has provided a parsimonious explanation, each has provided important contributions to the evolving understanding of the disorder. The accumulating literature with regard to enuresis increasingly suggests that a single unified theory of enuresis that applies to all children will not be identified.

Comparative Nosology

The variety of differing nosologic and diagnostic schemas seen with many disorders does not exist for enuresis. The behavior is concrete and definite: The child wets or does not. That criterion makes it possible to integrate research data from different countries, cultures, and treatment centers with greater reliability than is possible for many other disorders. The efficacy of a given treatment is reported in terms of its effect on the frequency of enuretic events. The primary differences with regard to diagnosis concern the frequency required to make a diagnosis of enuresis as a pathologic state and the period of continence necessary to separate primary from secondary enuresis, with a duration of 6 months occasionally being used instead of 1 year. Only slight differences exist between the eleventh revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-11) and the DSM-5-TR (Tables 48–1 and 48–2).

Table 48–1.
DSM-5-TR Diagnostic Criteria for Enuresis

A. Repeated voiding of urine into bed or clothes, whether involuntary or intentional.

B. The behavior is clinically significant as manifested by either a frequency of at least twice a week for at least 3 consecutive months or the presence of clinically significant distress or impairment in social, academic (occupational), or other important areas of functioning.

C. Chronological age is at least 5 years (or equivalent developmental level).

D. The behavior is not attributable to the physiological effects of a substance (e.g., a diuretic, an antipsychotic medication) or another medical condition (e.g., diabetes, spina bifida, a seizure disorder).

Specify whether:

Nocturnal only: Passage of urine only during nighttime sleep.

Diurnal only: Passage of urine during waking hours.

Nocturnal and diurnal: A combination of the two subtypes above.

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, p. 399 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Epidemiology

The first comprehensive data concerning the incidence of enuresis came from the Isle of Wight Study. They found that 15.2% of boys (7 years of age) were enuretic less than once a week, whereas 6.7% were wetting at least once a week or more. In 7-year-old girls, the frequency was 12.2% for those wetting less than once a week and 3.3% with a frequency of once a week or more. By 9 and 10 years of age, 6.1% of the boys were wetting less often than once a week, and 2.9% were wetting once a week or more. The corresponding figures for 9- and 10-year-old girls were 3.5 % wetting less than once a week and 2.9% wetting once a week or more. The frequency for boys decreased considerably by 14 years of age, at which time 1.9% were wetting less than once a week, and 1.1% were wetting at a frequency of once a week or more. The rates also decreased for girls by 14 years of age, with 1.2% wetting less than once a week and only 0.5% wetting once a week or more.

Table 48–2.
ICD-11 Diagnostic Criteria for Enuresis

- Repeated and persistent voiding of urine into bed or clothes (e.g., several times per week over several months), which may occur during the day or at night.
- The individual has reached a developmental age when urinary continence is ordinarily expected (approximately equivalent to a chronologic age of 5 years).
- The symptoms are not better accounted for by the physiologic effects of a substance or medication or by another medical condition (e.g., a Disease of the Nervous System, a Disease of the Musculoskeletal System or Connective Tissue, or congenital or acquired abnormalities of the urinary tract).

Specifiers:

Nocturnal Enuresis:

Inappropriate voiding of urine occurs only during the night. This is the most common form of enuresis and typically occurs during the first part of the night soon after the individual has gone to sleep.

Diurnal Enuresis:

Inappropriate voiding of urine occurs only during waking hours. This form of enuresis is also referred to as urinary incontinence.

Nocturnal and Diurnal Enuresis:

Inappropriate voiding of urine occurs both during the night and during waking hours.

Reprinted from World Health Organization. ICD-11 for Mortality and Morbidity Statistics, 2023. https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/1157749237.

A New Zealand study involving 8-year-old children found the prevalence of nocturnal enuresis to be 7.4%, with 3.3% primary, and 4.1% labeled secondary. A similar large Scandinavian study involving 7-year-old children reported a prevalence of 9.8%, with the highest prevalence (6.4%) involving children with nocturnal enuresis, 1.8% with diurnal enuresis, and 1.6% with mixed nocturnal and diurnal enuresis. A report from Australia described an incidence of 5.1% for children in the range from 5 to 12 years of age. A review of the published epidemiologic literature concerning enuresis indicates remarkably consistent findings from different countries and cultures, even though the research methods are somewhat variable. A standard consensus format that defines uniform reporting parameters would greatly enhance the epidemiologic literature base.

The epidemiologic literature is consistent in confirming that enuresis is a self-limiting disorder that will eventually spontaneously resolve. Common point prevalence figures are in the range of 7% of boys and 3% of girls at 5 years of age, dropping to 3% of boys and 2% of girls by 10 years of age. Only 1% of boys still wet at age 18 years of age, and still fewer girls wet at this age. A spontaneous remission rate of between 5% and 10% per year after 5 years of age is usually cited. Secondary enuresis may occur at any time, but most commonly, it begins between 5 and 8 years of age.

Etiology

Inclusion in the DSM-5-TR (as well as previous editions) definition of enuresis of children who wet intentionally with those whose enuresis is involuntary can be problematic. The vast majority of enuretic children do not wet intentionally or even on a subconsciously motivated basis. Increasingly, the research is pointing toward causal factors that may involve irregularities in physiologic processes. Clearly, children who wet intentionally are in a different phenomenologic grouping than those who do so on an involuntary basis, even if they do meet the other diagnostic requirements. The most likely explanations for voluntary intentional wetting are an oppositional defiant disorder (ODD) or a psychotic disorder. There are also a small number of children who originally have enuretic events involuntarily and subsequently manifest the behavior on a voluntary learned basis as well.

There is a correlation between enuresis and psychological disturbance that increases with age. This association is even more significant for enuresis, which persists into adolescence. Children living in socially disadvantaged situations and experiencing psychosocial stress have a greater frequency of enuresis than those who are not. The type and range of behavioral disturbance seen in children with enuresis are broad, and no marker can reliably differentiate between behaviorally disturbed and nondisturbed enuretic children. Thus, the associated behavioral disturbances are nonspecific and may represent a coincidental or secondary relationship rather than a causal correlation. Further supporting a secondary correlation are the repeated findings that children with enuresis have significantly more developmental delays and minor neurologic dysfunction than nonenuretic children compared to controls and other children attending a psychiatric clinic. Reports of delayed bone development in children with enuresis have also been linked to a possible general delay of maturation as a common underlying factor. However, new studies refute this finding, and no clear conclusion can be drawn at this time. The most recent research into a general developmental delay contributing to enuresis has used evoked potentials and electromyographic recording of startle eye-blink performance. Separate investigations using these methods have pointed to delayed maturation of brainstem functions.

Another approach to this etiologic question is represented by a large, longitudinal, population-based study in the United Kingdom, which compared the intelligence (as assessed by the Wechsler Intelligence Scale for Children–Third Edition [WISC-III]) of controls with children who still exhibited bedwetting in the age range of 7 years and 6 months to 9 years and 3 months. The authors reported that the bedwetting was associated with lower intelligence quotient (IQ) even after children with an IQ of less than 70 were excluded. However, this finding was primarily related to lower scores on the Performance section of the WISC-III. This observation led the authors to conclude that enuresis may be related to deficits in the maturation of the central nervous system.

One comprehensive effort reviewed the psychodynamic literature concerning enuresis and encopresis to define the concepts conveyed in that literature and then to determine empirically how often those concepts were verified and supported by the analysis of clinical material. The literature review yielded 24 generalizations that were consistently applied to enuretic and encopretic children. In subsequent statistical analysis of the clinical material, only 2 of the 24 generalizations reached statistical significance. Another recent study in this regard used a standardized behavioral rating scale, the Eyberg Child Behavior Inventory (ECBI), which was applied to children with primary nocturnal enuresis (PNE) (N = 92), a nonclinical control sample (N = 92), and a clinical sample (N = 92). The results indicated that the children with PNE had higher scores than the nonclinical control sample but significantly lower scores than the clinical sample. However, the mean score for the children with PNE was actually lower than the clinical cutoff score that had been determined for the ECBI.

Related to the association between behavioral disturbance and PNE are studies that address the frequency of diagnostic comorbidity. The most commonly reported comorbid diagnosis is attention-deficit/hyperactivity disorder (ADHD). In general, studies have reported that the co-occurrence of ADHD and PNE is approximately 30% greater than the frequency that would be expected from chance alone. A large contemporary study found that 32% of children diagnosed with ADHD also met diagnostic criteria for enuresis, whereas only 14% of a control group met the criteria. This finding was statistically significant at the $P < .001$ level. The rates of primary and secondary enuresis were not significantly different in the two groups. This study then investigated the degree of psychopathology and impairment in ADHD children (with and without comorbid enuresis) and found no significant differences. Similar findings were reported for the rates of psychopathology when the control children with enuresis were compared to those without the condition. Thus, the diagnosis of enuresis did not appear to increase the risk of psychopathology for the controls or the degree of impairment in the children who also had a diagnosis of ADHD. However, there was a higher incidence of learning disabilities and impaired academic performance in the control children with enuresis when compared to the control children without enuresis. This difference in the incidence of learning disabilities was not found when the ADHD children with enuresis were compared to those without the comorbid disorder. Thus, the co-occurrence of enuresis did not appear to increase the risk of learning disabilities in children with ADHD.

Another investigation used a family pedigree methodology to study the transmission of enuresis in the relatives of index family members who were diagnosed with PNE and in control probands who were diagnosed with ADHD as compared to those without this diagnosis. The results indicated that the two disorders were transmitted independently of one another.

Enuresis has been noted to occur statistically more frequently in the premorbid histories of children who develop early adolescent bipolar I disorder as compared to controls. An increased incidence of enuresis and other childhood psychiatric disorders has also been found in the history of adults with bipolar disorder and is associated with earlier onset of the disorder.

The association between behavioral disturbance and enuresis is stronger with secondary enuresis. This has been demonstrated by controlled studies and anecdotal reports of an association between the onset of enuresis and the loss of a parent through divorce or death. However, it should be noted that a large study from the Netherlands reported similar rates of psychopathology in children with primary enuresis as compared to those with secondary enuresis. The prevalence of psychodynamic factors in children with secondary enuresis would intuitively make sense because they have, by definition, demonstrated the physiologic competence to maintain continence for a prolonged period of time before losing that ability. The most specific study of this association indicated that exposure to four or more stressful life events in a year and delayed acquisition of nocturnal continence were risk factors for the development of secondary enuresis. An association between the occurrence of traumatic events and the development of secondary enuresis has also been demonstrated.

There has been one innovative attempt to explore the possible relationship between physiologic and psychological factors. Specifically, it was postulated that children with enuresis who did not present with a concomitant behavioral disturbance would be those with dysfunctional or abnormal bladders, whereas those with a behavioral disturbance would be more apt to have normal bladders. The study found the converse to be true: The group with behavioral disturbance had more dysfunctional bladders with lower volumes. The behaviorally disturbed group also had a greater number of developmental delays. A subsequent investigation reported no difference in bladder size among controls, children with PNE, and those who had been diagnosed with PNE but who were now continent. It also was demonstrated by two independent investigations that fluid loading can precipitate enuresis in children who have no prior history of enuretic events. The most recent investigations into this issue have used ultrasound techniques, which investigate both the volume capacity of the bladder and the thickness of the bladder wall in both controls and children with enuresis. This research suggested three subtypes: large-capacity bladder with a thick wall, normal-capacity bladder with normal wall thickness, and low-capacity bladder with a thick wall. The children with enuresis were then assessed with regard to their response to treatment with DDAVP. The results indicated that the children with the two abnormal bladder subtypes were less responsive to treatment than those with normal bladder physiologic characteristics.

The observation that enuresis tends to occur in family members has been known for decades. A large Scandinavian study involving more than 3,000 children found that a child’s risk for developing enuresis was 5.2 times greater if the mother also had the disorder and 7.1 times greater if the father had the disorder, lending further support to a genetic influence. Within the last few years, there have been tremendous advances in the study of the genetic transmission of enuresis, which suggests that genetic contributions are likely to be significant in families with multigenerational histories of enuresis. A Scandinavian group identified 11 families that manifested PNE over three generations in a pattern that was suggestive of an autosomal-dominant inheritance, with a penetrance greater than 90%. Genetic linkage studies with these families indicated markers on chromosome 13q. Similarly designed studies have also implicated chromosomes 4p, 8, 12q, and 22, suggesting genetic heterogeneity. A large multigenerational study that specifically investigated chromosome 22 found that only 39.3% of families had linkage to genetic markers on this chromosome. The genetic modes of transmission that have been discussed include sporadic, autosomal recessive, and autosomal dominant. Linkage studies have also suggested the possibility of mutation in the aquaporin-2 (AQP2) water channel locus in families with a dominant mode of transmission of primary enuresis, in that AQP2 appears to be essential for concentrating urine. Although genetic linkage studies continue to provide important information with regard to the etiology of enuresis, it increasingly appears that the genetic contributions are heterogeneous in nature and vary substantially among pedigrees that contain multiple generations of individuals with enuresis. Thus, it is likely that there is more than one mode of genetic transmission that contributes to the disorder; and different mechanisms may be operating in different family pedigrees.

One particularly important etiologic hypothesis was that urinary tract obstruction frequently caused enuresis. The significance of that belief revolves around the related implications for surgery to correct the obstruction. An extensive review of the relevant literature found no basis for concluding that bladder neck repair or urethral dilation was a reasonable or effective treatment for enuresis. That generalization does not apply to patients with specific anatomical or pathophysiologic findings.

Research into the association between sleep physiology and enuretic events has evolved steadily since the 1970s. The research can be seen as evolving through the following four stages. During the first phase of the research, it was postulated that enuretic events were dream equivalents that occurred in a deep sleep. The second major theory grew out of the observation that enuretic events originated in delta sleep and followed arousal signals. Thus, enuresis was seen as a disorder of arousal, which led to the hypothesis that enuretic children who were not psychologically disturbed did not produce arousal signals, whereas behaviorally disordered children did but failed to respond to them. The third phase of sleep research is represented by the two largest controlled studies, both of which found that when the time of night is also considered, enuretic events occur in all sleep stages in proportion to the amount of time spent sleeping in each stage. A fourth phase of investigation has attempted to couple cystometry with sleep studies in an attempt to identify subtypes. These investigations have yet to produce any definitive results that can be replicated.

A long-debated issue has been whether children with nocturnal enuresis are more difficult to awaken from sleep than other children. Earlier literature had suggested that anecdotal reports from the parents of children with enuresis that they were difficult to arouse stemmed from their lack of experience in attempting to arouse their other children, who did not have enuresis, from a deep sleep. However, studies have suggested that children with enuresis may, in fact, be more difficult to arouse than control children. Recent reports have also found that obstructive sleep apnea may play a role in the development of PNE in some children.

The efficacy of DDAVP as a treatment for enuresis has led to the theory that some children lack the ability to concentrate urine produced during the night and thus cannot reduce urine volume, and manifest enuretic episodes as a result. That hypothesis has led to studies investigating the circadian variation of plasma atrial natriuretic peptide (ANP). Children with enuresis have not consistently been found to differ from controls with regard to ANP. In one study, 14 of 55 children with enuresis had lower ANP concentrations than controls, and 9 of the 14 children had an excellent response to treatment with DDAVP. An investigation into the concentration of vasopressin in plasma and urine in 18 children with PNE and 20 matched controls found the concentration of vasopressin in plasma at 8:00 AM was significantly lower in the enuretic group. Total 24-hour urinary vasopressin excretion was also lower in the enuretic group but not significantly so.

Another hypothesis that has been investigated is that endogenous arginine vasopressin (AVP) levels would be different in children with enuresis who respond to DDAVP as opposed to those who do not. This would suggest the existence of a specific subgroup of children with enuresis based on their endogenous AVP levels and their positive response to DDAVP. A study involving a small number of subjects supported this hypothesis by indicating that AVP levels in the morning differentiated between controls and children with enuresis, as well as between DDAVP responders and nonresponders. However, a large study that also controlled for water intake over 72 hours and that sampled AVP more frequently found that only 14 of the 55 children with enuresis had a significant decrease in AVP when compared to controls. Of these, 9 had a significant positive response to DDAVP. Further complicating this line of research is the finding that AVP is not secreted in a continuous manner over 24 hours but rather is secreted in a pulsatile pattern, suggesting that definitive studies of this variable require frequent sampling. Investigations that have adopted this methodology have yielded conflicting results, in that one reported no differences between DDAVP responders and nonresponders, whereas another study with a small number of subjects that used hourly measurements found that in the 11:00 PM to 4:00 AM time period, children with enuresis (*N* = 9) had significantly lower AVP levels than did controls (*N* = 8).

There have also been related studies into the role of urine osmolality. One hypothesis is that a decreased ability to concentrate urine could contribute to an increase in the volume of urine. In general, these studies indicated that children with enuresis do have lower urinary specific gravities (as compared to controls) but not to a level that reaches statistical significance. It has also been reported that the levels of AVP can be increased in controls and children with enuresis by fluid restriction and, furthermore, that plasma osmolality can affect AVP secretion. In this regard, a comparison of controls, DDAVP nonresponders, and DDAVP responders indicated that fluid deprivation produced an increase in AVP levels in all groups, but the response was less robust in the DDAVP responders than in the other groups.

Another variation on this line of research is that refractory enuresis persisting into adolescence and adulthood may be secondary to a relative insensitivity to AVP, which can be positively affected by DDAVP. These observations contribute to the hypothesis that the primary pathophysiology of enuresis may occur at the receptor level, and this may also explain why some children do not respond to DDAVP.

Further complicating the research into the mechanism of action of DDAVP are two independent investigations that suggested that there is a central nervous system mechanism that may account for its efficacy in some children with enuresis.

Thus, as the foregoing review suggests, what initially appeared to be an intuitive explanation for the mechanism of action of DDAVP has not been supported by a number of studies and a precise mechanism that applies to all children who respond to DDAVP has not been identified.

There have also been investigations that have implicated tubular mechanisms with regard to the urinary excretion of potassium and sodium by children with enuresis. Although this research has primarily focused on the possible mechanism of action of DDAVP, it has also led to a reanalysis of imipramine’s mechanism of action. Imipramine can produce a decrease in the excretion of sodium and potassium, which then produces a reduction in osmolal clearance and a decrease in urinary output. It has also been reported that imipramine can affect the secretion of nocturnal urinary antidiuretic hormone (ADH) and that this may account for its efficacy.

Diagnosis and Clinical Features

The concreteness and simplicity of enuresis make the diagnosis relatively easy. The DSM-5-TR continues the same age criteria that were used in the DSM-IV and DSM-IV-TR by specifying that the diagnosis is not made in a child whose chronologic or mental age is younger than 5 years of age. The stated rationale is that, by this age, continence can be expected to have occurred. As noted by the reference to a mental age of 5 years, children with developmental disabilities may have a greater chronologic age. The wetting must occur at least twice per week for at least 3 consecutive months, or, if less frequent, it must produce significant distress or functional impairment. Physical causes, such as a bladder infection, must be excluded. Qualifiers to the diagnosis indicate whether it is primary or secondary enuresis. As previously noted, *primary enuresis* refers to those children who have never maintained continence for any length of time, whereas *secondary enuresis* refers to those who had achieved continence for at least 6 to 12 months and then resumed wetting. The other qualifiers refer to the timing of the enuretic event during the day. Although most children exhibit only nocturnal enuresis, some have daytime (diurnal) patterns or a combined nocturnal and diurnal pattern.

The inclusion of children who wet intentionally with those who wet involuntarily under the same diagnostic heading can be problematic because that distinction has significant clinical implications. Children who are wetting intentionally almost certainly are doing so as a manifestation of a psychological disturbance. Although there is an association between behavioral disturbance and involuntary enuresis, the nature of the correlation is relatively nonspecific. Thus, although the coexistence of other behavioral problems is not a diagnostic issue per se, it is of clinical importance.

John was a 10-year-old boy who had consistently wet the bed throughout his life. He averaged three to four nocturnal episodes per week, and his longest period of continence was a few weeks. Thus, John warranted a diagnosis of primary enuresis. John's father had been enuretic until 12 years of age, as were his father and a paternal uncle. Accordingly, the family had always assumed that this was a hereditary genetic disorder that John could not control. In keeping with this perception, John's parents had never adopted a punitive approach to the bedwetting. The wetting was handled in a matter-of-fact manner, with John washing his sheets every morning and showering before going to school.

The family sought consultation at this time because John was increasingly being asked to sleep overnight at the homes of friends and also wanted to attend an overnight Boy Scout camp. These concerns led the family to meet with their pediatrician, who suggested a trial of the bell-and-pad method of conditioning. This approach was not effective, in part because the family lived in cramped quarters and the alarm that was meant to awaken John also woke his two brothers who shared the bedroom with him.

At this point, the family's pediatrician suggested that they meet with a child psychiatrist with whom she worked. The child psychiatrist described the clinical literature on DDAVP. However, when John's father indicated that his health insurance did not cover prescriptions and that the family was having difficulty covering their household expenses on his income, which was only moderately above minimum wage, the psychiatrist felt compelled to discuss the relative costs of DDAVP and imipramine, which was less expensive than DDAVP. However, the psychiatrist also indicated that John would require a baseline electrocardiogram (ECG) (which his insurance could cover).

Ultimately, the family decided that John would need medication only for brief, discrete periods (overnights at friends and 1 to 2 weeks of summer camp) because they had worked out an acceptable plan for responding to the wetting on a day-to-day basis. Furthermore, they reasoned that, within approximately 2 years, John would experience a spontaneous remission of his enuresis, given that this was the family pattern. Thus, they decided that, because DDAVP has a rapid onset of action, they would pay out of pocket for the relatively small amounts of DDAVP that would be needed for these discrete periods over the next 2 years. The pediatrician advised the family of the Federal Drug Administration (FDA) alert regarding the risk of hyponatremia and seizures. They decided to use the oral form of DDAVP because the FDA identified the nasal spray as being more likely to produce hyponatremia and indicated that this form of DDAVP was no longer indicated for childhood enuresis. They were also advised to (1) limit John's fluid intake on the nights when DDAVP was administered, (2) discontinue its use if John developed the flu or another illness that could affect his fluid balance, and (3) discontinue the medication and notify their physician if symptoms of nausea, vomiting, or headache developed.

Pathology and Laboratory Examination

Because urinary tract infections can produce enuresis, a urinalysis should be part of every evaluation. Using radiographic procedures with contrast media to detect an anatomical or physiologic cause for the enuresis is more problematic because the procedures are invasive and painful and the diagnostic yield is low. A large study carried out in a pediatric primary care setting found a 3.7% incidence of obstructive lesions in children with enuresis. Others have reported similar findings. The literature cited earlier in the section on etiology concerning the use of noninvasive ultrasound techniques to establish bladder capacity and wall thickness have not been replicated to a degree that would warrant their use on a routine basis. However, they may be of use for children who are refractory to standard treatment.

Differential Diagnosis

As suggested previously, the primary differential diagnosis is a urinary tract infection. This is especially true for girls, who are more prone than boys to develop urinary tract infections. A urinary tract infection should be the first consideration for a girl who has been continent for a considerable period of time and has recently begun wetting. Although the diagnostic yield is smaller with boys, a urinalysis should still be carried out.

Enuresis can result from anatomical malformations or obstructive lesions, but the percentages are relatively low. If the history and interviews with the child suggest that the enuresis is intentional, then it is almost certainly related to an underlying psychological disturbance. The relation between psychological disturbance and involuntary enuresis is less clear. The coexistence of another behavioral disturbance should be noted and attended to clinically. The most common comorbid psychiatric disorder is ADHD.

It has also been documented that enuresis has a negative effect on self-image and that enuretic children have more negative feelings about their disorder as compared to children who have other chronic illnesses.

There have also been case reports linking the onset of enuresis in children to the administration of both the selective serotonin reuptake inhibitor (SSRI) antidepressants and the atypical antipsychotic agent risperidone (Risperdal). Thus, the possibility of enuresis as a side effect of these medications should be considered if a chronologic correlation exists.

Secondary enuresis has also been identified as an important symptom of the abrupt onset of Type 1 Diabetes in children and is related to the polydipsia and polyuria that accompany that disorder.

Course and Prognosis

The natural history of enuresis is significant because it is a self-limited disorder that can spontaneously remit. The fact that the diagnosis is not made until 5 years of age takes into account children who have delays in toilet training that are not outside the accepted range (between 2 and 5 years of age). The prevalence of enuresis is relatively high, between 5 and 7 years of age, and then drops off substantially. The vast majority of enuretic children experience a spontaneous resolution of the problem at some time, and only a few remain enuretic into adulthood. By 14 years of age, only 1.1% of boys wet once a week or more. The peak age for secondary enuresis is between 5 and 8 years of age.

Treatment

The methods for the treatment of enuresis that have been empirically proven to be effective are primarily behavioral and pharmacologic. Psychotherapy may be useful for ameliorating some of the associated behavioral problems that can be seen with enuresis, especially secondary enuresis. A particularly common clinical scenario for secondary enuresis is the development of wetting after the loss of a parent through death or divorce. In these patients, psychotherapy is the primary treatment modality. A review of the efficacy of psychotherapy for primary enuresis found a 20% success rate, which is probably not significantly greater than would be expected by the rate of spontaneous remissions and random chance.

Behavior Therapy. A comprehensive review of several studies determined the success rate for behavioral interventions as 75%. Recent studies have yielded comparable response rates. The primary behavioral intervention is the bell-and-pad method of conditioning. A pad is placed on the bed, with a wire running to a bell. When the child wets, the moisture completes a circuit in the pad, ringing the bell and waking the child. With repeated use, the child learns to awaken before wetting occurs. Several studies have indicated that concomitant behavioral disturbance reduces the likelihood of success of behavioral interventions.

Recent behavioral studies have focused on embellishment of the standard method and comparisons with different methods. For example, no difference was found between an audio alarm and a vibratory alarm. A study that compared the basic bell-and-pad method with methods in which the bell and pad were coupled with more response contingencies for wetting versus not wetting found no statistical difference other than a slightly lower relapse rate with the adjunctive use of reward contingencies. One innovative strategy involved replacing the bell and pad with an alarm clock timed to go off after 2 to 3 hours of sleep when maximum bladder capacity would be expected. The authors reported responses that were equal to results with the standard bell-and-pad method.

The relationship between bladder capacity and response to behavioral treatment has also been investigated. One study found that bladder capacity did not affect outcome with the bell-and-pad method of conditioning, whereas another found that children with smaller bladder capacities tended to do slightly better when retention-control training was linked to the bell-and-pad method. A meticulous investigation into which treatment variables might relate to the efficacy of the bell-and-pad method found that the inability to be awoken by the alarm correlated significantly with failure, as did low functional bladder capacity.

There is some evidence to support the use of biofeedback for children who have small bladder capacities and unstable detrusors and who have been refractory to prior treatment. The newest advance in this behavioral methodology has been the use of an external ultrasonic monitor that is attached to a waistband. The monitor then signals the alarm when the bladder reaches maximal capacity. Clinical research with this device has yielded success rates similar to those with the basic bell and pad and has also reported increases in bladder capacity. The bell-and-pad approach has also been coupled with the use of DDAVP for children who have been refractory to prior treatment owing to family dysfunction or a concomitant behavioral disorder, or both. Although the bell and pad is the most intensively studied method of behavioral intervention, it is not the only type of behavioral treatment. A recent comprehensive review of this subject also noted other behavioral methods, such as evening fluid restriction, night-time toileting, rewards, overlearning, and retention-control training.

Pharmacotherapy. The initial description of imipramine's efficacy for enuresis was followed by more than 40 double-blind studies that confirmed its therapeutic value. This marked the advent of the pharmacologic era of the treatment of enuresis, with imipramine being the dominant pharmacologic entity used for more than two decades. Although imipramine has now been largely replaced by DDAVP, there are still reasons to include a discussion of its use in a comprehensive review of enuresis. Imipramine is still a consideration for children who are refractory to other forms of treatment. There is also a significant cost differential between imipramine and DDAVP, so that imipramine may well be the first pharmacologic choice for those families who do not have any form of insurance and who have limited financial resources.

Later studies with imipramine focused on the relation between response and concentration of imipramine in the blood. Although one study found no connection between positive response and imipramine concentration alone or in combination with its metabolite desipramine (Norpramin), three studies found a significant correlation between positive response and total concentration of imipramine and desipramine. More specifically, optimal response was found at combined concentrations (imipramine plus desipramine) of greater than 60 ng/mL in one study and of greater than 80 ng/mL in another. The most recent study also found that efficacy was related to increasing doses but noted tremendous variation in serum concentrations among individual children receiving the same dose. The adverse effect of dry mouth also correlated with concentration in blood, providing clinicians with a crude index of blood concentration in children who are extremely reluctant to have their blood drawn.

Clinically, when one encounters a history of a lack of response to imipramine, one must ascertain the specific dose used because nonresponse is frequently related to dose: primary care physicians often prescribe only 25 to 50 mg per day. However, low-dose responders do exist, and it is reasonable to start treatment at a daily dose of 25 mg, titrating up by 25 mg every 4 to 7 days. This enables identification of the low-dose responders because the response to imipramine usually occurs within 1 to 2 days. The majority of children exhibit a positive response in the range from 75 to 125 mg. The standard maximal limit for dose is 5 mg/kg body weight, and ECG monitoring is recommended at doses greater than 3.5 mg/kg. A pretreatment baseline ECG is also recommended. Blood concentration values may also be clinically useful if higher doses fail to elicit a response.

There is also a possible risk of overdose by the child with enuresis. Children can engage in magical thinking, and there have been reports of children who believed that, if taking a few pills would stop the wetting for a night, then consuming the whole bottle would make it disappear forever. Thus, parents should be advised to control the medication carefully and to store it in a secure place. Severe overdoses may require treatment with physostigmine (Antilirium), whereas more moderate overdoses may be treated by the symptomatic management of arrhythmias and seizures.

Because almost all children at some point experience spontaneous remission of the enuresis, it does not make sense to prescribe medication for years without checking for a spontaneous remission. One solution to that problem is to taper the imipramine dose at 3-month intervals. Should enuresis reappear as imipramine dose is tapered, it can simply be returned to the optimal level for another 3-month period. Clinically, one often finds that more children remain dry after 3 to 6 months of treatment than can be expected from spontaneous remission alone. In general, treatment with imipramine should be viewed as necessary until spontaneous remission occurs.

Three clinical subtypes of enuretic response to imipramine exist: true nonresponders, true responders who maintain a response over time, and transient responders who respond for a few weeks and then resume wetting. Increasing the dose for transient responders usually recaptures the response but only for another few weeks, and ultimately one reaches a point at which the dose cannot be increased further. However, the initial transient response can be replicated after a medication-free interval; thus, one can use the medication for especially important, time-limited social events, such as summer camp.

As indicated previously, DDAVP has now surpassed imipramine as the first-line pharmacologic approach for enuresis. A thorough review done as early as 1993 revealed 18 randomized, controlled trials, including a total of 689 subjects. Many subjects had been refractory to previous treatment for enuresis. DDAVP reduced enuretic frequency by 10% to 91% in these studies. Positive prognostic factors were age (older than 9 years of age) and lower pretreatment frequency of enuretic events. In most studies, wetting resumed after the medication was discontinued. Follow-up studies indicated that 5.7% of children remained dry after cessation of medication, but many of those cases could have been due to spontaneous remission. The most commonly reported adverse effects were mild abdominal pain, epistaxis, headache, and nasal stuffiness. Several subsequent studies attempted to find factors that might predict or relate to an increased likelihood of a positive response to DDAVP. No one factor appears to predict a positive response, but larger functional bladder capacity and a positive family history of enuresis appear to be positively correlated with treatment success. Factors that are not reliably related to treatment outcome are urine osmolality, plasma osmolality, ADH secretion, urine production, and nocturnal AVP concentration. However, individual studies can be found that link these factors with a positive outcome.

A large multicenter follow-up treatment study of DDAVP from Sweden (*N* = 399) used a 4-week baseline observation phase, a 6-week dose titration period (range from 20 to 40 µg), and a 1-year follow-up period. The study design also included an assessment for the rate of spontaneous remission by means of a 1-week medication-free period every 3 months. Sixty-one percent (*N* = 245) of the children met the requirement of at least a 50% reduction in frequency during the dose titration period that was required to enter the 1-year follow-up phase of the study. The average frequency of enuretic events in the baseline observation period was 5.3 events per week, and this declined to 0.8 nights by the end of the year-long follow-up period. As with other studies, older age was a positive prognostic sign. Of interest was the observation that 77 children experienced a complete remission of their enuresis within the first 6 months of the active treatment phase. There has been one report of an attempt to increase the rate of spontaneous remission by aggressively titrating the dose until cessation of wetting is obtained and then maintaining this dose for 4 to 6 weeks before beginning to decrease the dose by 10 µg per month until it is discontinued. The authors reported that, by using this strategy, complete remission was achieved for 71% of the children. The median length of follow-up was 18 months, and the mean dose of DDAVP was 20 µg.

Studies have also specifically investigated the efficacy of DDAVP for children who were refractory to a conditioning treatment and imipramine. One of these involving 52 children (dose range from 20 to 40 µg intranasally) found complete cessation of wetting in 53%, a partial response in 19%, and minimal or no response in 28%. A long-term follow-up study (mean length of follow-up of 13 months) of children treated with DDAVP found no hematologic or hormonal adverse effects.

Treatment with DDAVP has been compared to the bell-and-pad method of conditioning. The difference in response rate was not statistically significant; 86% improved with the bell and pad, and 70% improved with DDAVP. Another investigation comparing the combined use of the bell-and-pad method and DDAVP with placebo indicated that the combination achieved significantly more dry nights. Similar results were found in a study comparing alarm monotherapy to alarm treatment coupled with DDAVP. The combination proved superior, especially for children with severe, almost nightly, wetting.

The development of oral tablets greatly simplified the administration of DDAVP. The nasal spray was difficult for some children to use, and this could (in some cases) contribute to variable administration. A controlled multicenter study found no difference in response between oral DDAVP and the traditional nasal spray. A dose of 400 µg produced more dry nights than did a 200-µg dose.

A significant dose-effect correlation was found in a placebo-controlled study that used 200-, 400-, and 600-µg oral doses. In a separate follow-up study (7 years), the rate of spontaneous remission at 2 and 7 years was reported as greater than would be expected solely on the natural history of the disorder. As with other treatment modalities, consistent findings in the clinical literature concerning positive responses to DDAVP are older age, larger bladder capacity, and lower pretreatment frequency of enuresis. The newest innovation in the use of DDAVP has been the development of a sublingual desmopressin oral lyophilisate (MELT). A randomized, controlled study that compared this formulation to oral tablets found it to be equally effective. The children also expressed a greater preference for the MELT. This increase in preference was especially pronounced in the younger age group (5 to 11 years of age).

As the use of the DDAVP nasal preparation became more common, there were an increasing number of case reports of hyponatremic seizures with prolonged use. A review of 14 articles that reported data on serum sodium concentration in DDAVP-treated patients and 11 articles that reported individuals who developed altered consciousness or a seizure while taking DDAVP for enuresis concluded that hyponatremia is a potential adverse effect of DDAVP treatment. The authors also noted that excess fluid intake was a contributing factor in 6 of the 11 case reports. They concluded that patients receiving DDAVP for nocturnal enuresis should be advised not to consume more than 8 oz of fluid on nights when DDAVP is administered. The data on hyponatremia also suggest that periodic monitoring of serum sodium concentration would be prudent. The number of reported cases of hyponatremia and seizures, including two deaths, prompted the FDA to issue an alert concerning DDAVP in December 2007. The alert identified 61 postmarketing incidents of hyponatremia-related seizures that were related to the use of DDAVP. Sodium levels obtained at the time of the seizure ranged from 104 to 130 mEq/L. A particular concern was that two of the individuals who experienced seizures with hyponatremia died. The intranasal formulation of DDAVP was implicated in 36 of the cases, and 25 of these were individuals younger than the age of 17 years, most of whom were receiving DDAVP for treatment of nocturnal enuresis. The FDA alert concluded that the intranasal formulation of

DDAVP was no longer indicated for the treatment of PNE. The alert also suggested that treatment with the tablet form should be halted during illnesses that could produce disturbances in fluid dynamics and/or produce electrolyte imbalance. The alert also recommended that fluids should be restricted from 1 hour before to 8 hours after the administration of the tablets. The latter recommendation is consistent with earlier published reports that linked the amount of fluid intake to the risk of acute hyponatremia.

Although imipramine and DDAVP have been the dominant pharmacologic treatment modalities over the last several years, there continue to be reports of success with anticholinergic agents. The majority of these reports relate to oxybutynin, both as a first-line intervention and for individuals who are refractory to other interventions.

General Treatment Considerations. The results of the studies reviewed, in conjunction with the natural history of enuresis, can help the clinician to develop a decision tree tailored to a given child. The first decision, naturally, is whether to treat at all. That decision is affected by the frequency of the wetting, the age of the child, and the amount of psychological distress that the wetting causes the child. Enuresis has a negative effect on the self-esteem of these children, and self-esteem improves after successful treatment. One may choose not to treat a child who is infrequently wetting and who is at an age at which spontaneous remission may be expected.

If one does elect to treat, it makes sense first to try the bell-and-pad method of conditioning because its efficacy has been repeatedly demonstrated, and behavioral intervention is considered less invasive than pharmacologic approaches. A large follow-up study that compared observation only with imipramine, DDAVP, and the bell and pad indicated that, although all of the active treatments were significantly superior to observation alone, the positive response was much more likely to be sustained after treatment cessation with the bell and pad than with either pharmacologic treatment. Despite these apparent advantages, one large population-based investigation revealed that only a relatively small number of enuretic children

(38%) saw a physician, and, of that number, only 3% were treated with the bell-and-pad method of conditioning, whereas more than one-third received pharmacologic treatment. A more geographically restricted survey of primary care physicians yielded significantly different responses, in that 80% of physicians indicated that they recommended the bell-and-pad method of conditioning. The most recent comparative study of the long-term efficacy of DDAVP versus the enuretic alarm involved 142 children who were randomized to either form of treatment for 6 months, with the determination of success rates at 6 and 12 months. Interestingly, only 5.2% of the DDAVP group withdrew after randomization, as compared to 30.7% of those assigned to the enuretic alarm group. The results for those who completed 6 months of treatment (defined as at least 90% in wet nights) indicated that 76.8% of the DDAVP group was deemed to be successful, as compared to 61.8% of the alarm group. The success rates at 12 months were comparable, with 77.8% of the DDAVP group meeting success criteria, as compared to 75% of the alarm group. However, when the large drop-out rate after the assignment to the alarm group was taken into account, the success rates were adjusted to 68.8% for the DDAVP group and 46.2% for the alarm group. Thus, this study essentially confirmed the results of numerous prior studies in that the two treatment modalities were found to be equally effective, but the ability to implement the bell-and-pad alarm system was too difficult for many families to implement successfully.

When the bell-and-pad method of treatment is unsuccessful or not feasible, it makes sense to proceed to pharmacologic treatment, assuming that the enuresis is severe enough to warrant pharmacologic intervention. For decades, imipramine was the first choice for pharmacologic treatment. Although precise figures are not available for its relative use in the treatment of PNE, DDAVP has surpassed imipramine as the most popular pharmacologic treatment.

ENCOPRESIS

Definition

The DSM-5-TR criteria are virtually unchanged from those of the DSM-IV and DSM-IV-TR and define *encopresis* with four related criteria: (1) the repeated inappropriate passage of feces, usually involuntary, (2) occurrence at least once a month for at least 3 months, (3) a chronologic or mental age of 4 years of age, and (4) exclusion of a substance or medical condition as a cause.

For several years, the clinical and research literature concerning encopresis has distinguished between retentive and nonretentive encopresis. The DSM-5-TR recognizes this distinction and lists two subtypes: With constipation and overflow incontinence and without constipation and overflow incontinence. The subtype with constipation and overflow incontinence corresponds to retentive encopresis. The subtype without constipation and overflow incontinence corresponds to nonretentive encopresis.

Some clinicians refer to a *primary type* in which the person has never established fecal continence and a *secondary type* in which the disturbance develops after a period of established fecal continence.

History

The history of encopresis is not as richly detailed as that of enuresis, possibly because encopresis affects fewer children, and the events are less frequent. Unless the feces are particularly fluid, it may also be easier for children to hide the symptoms. A high frequency of encopresis and enuresis was found in children who were separated from their parents in World War II, thus documenting the effect of environmental factors.

The more contemporary history of encopresis is noteworthy for the establishment of a treatment methodology that involves cathartics to re-establish bowel dynamics (coupled with education), which has proven to be quite successful. The innovative use of biofeedback technology showed much initial promise after its introduction in the 1980s, but it now appears that it is not more effective than the older method. Constipation is a common precursor to encopresis, and recent research has also focused on effective treatment for this disorder.

Table 48–3.
DSM-5-TR Diagnostic Criteria for Encopresis

A. Repeated passage of feces into inappropriate places (e.g., clothing, floor), whether involuntary or intentional.
B. At least one such event occurs each month for at least 3 months.
C. Chronological age is at least 4 years (or equivalent developmental level).
D. The behavior is not attributable to the physiological effects of a substance (e.g., laxatives) or another medical condition except through a mechanism involving constipation.

Specify whether:

With constipation and overflow incontinence: There is evidence of constipation on physical examination or by history.

Without constipation and overflow incontinence: There is no evidence of constipation on physical examination or by history.

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, p. 402 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Comparative Nosology

As noted above, the DSM-5-TR has adopted the widely held clinical distinction between retentive and nonretentive encopresis, although the terminology differs slightly. The current DSM-5-TR definition includes voluntary and intentional encopresis, which is obviously a distinction of major clinical importance. This statement appears to minimize the importance of a rather substantial subgroup of children with nonretentive encopresis who are encopretic because they cannot adequately control the anal sphincter or do not recognize the need to defecate in time (Table 48–3).

The ICD-11 diagnostic classification for encopresis essentially follows the DSM-5-TR approach (Table 48–4).

The most comprehensive categorization of encopresis distinguishes three types of patients: (1) intentional—children who do have bowel control but intentionally deposit feces in inappropriate places for psychological reasons; (2) involuntary—those who cannot adequately control the sphincter or lack an awareness of the process, or both; and (3) involuntary—those whose soiling is related to excessive fluid from retentive overflow (75% of this category), diarrhea, or anxiety. Thus, this classification schema incorporates the retentive and nonretentive dichotomy and the distinction between voluntary and involuntary incontinence.

Table 48–4.
ICD-11 Diagnostic Criteria for Encopresis

- Repeated and persistent passage of feces in inappropriate places (e.g., at least once per month over a period of several months).
- The individual who has reached the developmental age when fecal continence is ordinarily expected (approximately equivalent to a chronological age of 4 years).
- Fecal soiling is not better accounted for by the physiological effects of a substance (e.g., excessive use of laxatives) or another medical condition (e.g., aganglionic megacolon, spina bifida, anal stenosis, chronic diarrhea, congenital or acquired abnormalities of the bowel, or gastrointestinal infection).

Specifiers

Encopresis with constipation and overflow incontinence:

- Encopresis with constipation and overflow incontinence is the most common form of fecal soiling, and is characterized by retention and impaction of feces. Stools are typically—but not always—poorly formed (loose or liquid) and leakage may range from occasional to continuous.
- There is often a history of toilet avoidance leading to constipation.

Encopresis without constipation and overflow incontinence:

- Encopresis without constipation and overflow incontinence is not associated with retention and impaction of feces, but rather is characterized by reluctance, resistance, or failure to conform to social norms in defecating in acceptable places in the context of normal physiologic control over defecation.
- Stools are typically of normal consistency and inappropriate defecation is likely to be intermittent.

Reprinted from World Health Organization. ICD-11 for Mortality and Morbidity Statistics, 2023. https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/313374467.

Epidemiology

Two large epidemiologic studies have yielded consistent results. The Isle of Wight Study reported a prevalence of 1.3% for boys and 0.3% for girls (10 to 12 years of age) soiling at least once a month. A large study involving 8,863 children found a 1.5% prevalence rate for boys (7 and 8 years of age). In that study, the ratio of boys to girls was more than 3:1. A significant correlation between encopresis and enuresis has also been found.

Etiology

The dramatic nature of fecal soiling has contributed to psychodynamic speculation. The term *chronic neurotic encopresis* has been used to describe the small number of children whose encopresis is predominantly psychological. The clinical characteristics described in these children are (1) a distant father and neurotic mother, (2) early and often harsh bowel training, and (3) a history of neurologic delay. The one systematic study that attempted to verify the clinical features usually attributed to enuretic and encopretic children in the literature by analyzing actual case material could not substantiate the existence of 20 of the 22 factors developed through the literature review. As with enuresis, the distinction between primary and secondary encopresis relates to the issue of associated psychopathology. A study involving 63 boys with encopresis found that boys with primary encopresis were more likely to have developmental delays and associated enuresis, whereas those with secondary encopresis were more likely to have experienced higher levels of psychosocial stressors and to be diagnosed with conduct disorder. Another large study examined several psychological and physiologic variables in children with chronic constipation as compared to children with encopresis unrelated to constipation. The group with constipation had significantly longer colonic transit times (CTTs), whereas both groups had relatively high frequencies of abnormal defecation dynamics. A subsequent independent investigation of CTT in children with chronic constipation found that a CTT greater than 100 hours was related to a poor outcome at 1-year follow-up.

A large population-based study in the United Kingdom involving 8,242 children (7 to 8 years of age) investigated the association of soiling with a variety of psychological disturbances. The children with encopresis were divided into those who frequently soiled (at least once a week) or occasionally (defined as less than once a week). There was also a large control group. Both of the symptomatic groups had significantly higher rates of behavioral and psychological disturbance than the controls. The rates of disturbance were also higher in the group that frequently soiled than in the occasional-soiling group. However, there was no particular behavioral profile that correlated with encopresis. The authors noted higher rates of attentional, obsessive-compulsive, and oppositional presentations in the symptomatic children than in the controls.

A premorbid history of encopresis has been found to occur more frequently in individuals with prepubertal and early adolescent bipolar disorder phenotype (PEA-BP) than in controls. A recent large, retrospective study (in a military health-system database) involved 32,773 children who were identified as having ADHD. The results indicated that 4.1% of those children had constipation, as opposed to 1.5% of control children without ADHD. The corresponding rates for fecal incontinence were 0.9% in the children with ADHD, as compared to 0.15% of the controls. Both of these results were statistically significant with high levels of confidence, leading the authors to conclude that ADHD was a risk factor for fecal incontinence.

A recent large birth cohort study involving 358 children also found higher rates of both enuresis and encopresis among children with ADHD.

No correlation between encopresis and social class was reported in two independent studies. The issue of neurologic competence as a contribution to encopresis has been investigated. One study found no abnormalities of anal–rectal motor or sensory function but did find that a significant number of encopretic boys had abnormal anal–rectal expulsion dynamics. A sophisticated physiologic study using a triple-lumen catheter attached to a hydraulic manometry infusion system found that age of onset, frequency, and duration of encopresis were all highly correlated with the amount of anal sphincter spasm that occurred when defecation was attempted. The amount of pain reported with bowel movements correlated positively with the frequency of encopresis and inversely with maximum squeeze pressure. An investigation involving the concentrations of gastrointestinal (GI) hormones in controls and children with retentive encopresis indicated that the latter had higher concentrations of pancreatic polypeptide after a meal and less of a motilin response. However, the authors noted that these abnormalities could be the result of, or the cause of, chronic constipation.

An extensive series of studies concerning physiologic factors and the interplay of physiologic and psychological factors has been carried out. The principal physiologic findings were (1) 56% of children with retentive encopresis could not defecate rectal balloons (research technique used to assess sphincter competency), and most had abnormal anal sphincter contractions; (2) only 14% of those who could not defecate the balloons were found to have responded to treatment at 1-year follow-up, compared to a 64% success rate for those who could; (3) at 1-year follow-up, 70% of children who could relax the anal sphincter at the time of initial evaluation were improved, as opposed to 13% of those who could not; and (4) none of the children who initially presented with an abdominal fecal mass was improved at 1-year follow-up, regardless of other factors.

One subset of these studies physiologically compared constipated children to controls during the act of bearing down. All of the control children displayed decreased anal sphincter activity when bearing down, compared to 58% of the constipated children who could defecate a balloon and 7% of constipated children who could not defecate the balloon. The constipated children who could not defecate the balloon were also significantly less likely to respond to laxative treatment. Nonrelaxation of the anal sphincter was found in 75% of encopretic children, as opposed to 13% of a control group in a study that used anal electromyography. This study also reported lower pressures for the encopretic children at rest and on squeezing; this observation has also been described by other investigators.

An attempt was also made to assess the relative effect of physiologic and psychological factors. Specifically, anal–rectal manometric and electromyographic assessments were correlated with behavioral and social competence profiles as related to treatment outcome. Psychological variables were not found to be predictive of outcome, but physiologic variables were.

Diagnosis and Clinical Features

The objective and concrete nature of the encopretic event makes diagnosis relatively easy and uncomplicated from a phenomenologic perspective if the child meets the frequency, duration, and age requirements set forth in the DSM-5-TR. The involuntary versus intentional dichotomy is not a requirement or reason for subclassification according to the DSM-5-TR, but it has important clinical implications. Similarly, the retentive versus nonretentive schema has clinical usefulness. Differentiating between primary and secondary encopresis is important, given that one might expect medical or psychological factors to play a greater role in secondary encopresis.

Henry was an 11-year-old boy with almost-daily encopresis and a number of associated behaviors, including hiding the feces around the house. He resided in a specialized foster care setting, having been removed from his biologic parents at 7 years of age because of physical and sexual abuse. Both parents were involved with substance abuse, and his early history was not well documented. However, the parents did indicate that he had never exhibited sustained bowel continence for several months. Henry had also been enuretic until 6 years of age, but this had resolved to an occasional nocturnal episode every 4 to 6 months.

Henry also qualified for a diagnosis of ODD. Although he had experienced physical and sexual abuse, he did not have flashbacks or other symptoms that would meet the criteria for post-traumatic stress disorder (PTSD). Henry also had ADHD and was being effectively treated with 10 mg of methylphenidate (Ritalin) twice a day.

The foster family resided in an urban area that had access to a nationally recognized children's hospital. The ambulatory care department had a specialized behavioral encopresis program that coupled the bowel training method with a psychoeducational component and psychotherapy. The psychiatric consultant to the specialized foster care program doubted that this program would be successful for Henry because he had so much associated psychopathology, and the feces were often deposited around the house in a symbolic manner. In addition, the encopresis was not of the retentive-overflow type, and the feces were always well-formed. However, because no apparent harm could come from the referral, the consulting child psychiatrist agreed to it. Much to the surprise of the consultant, the several-week outpatient bowel training course, coupled with the psychoeducational component and psychotherapy, resulted in a complete cessation of the encopresis. On one of her visits to the home, Henry proudly showed his case manager a diagram of the functioning of the digestive system that was part of the psychoeducational program. In retrospect, it appeared that, although there were symbolic aspects to Henry's encopretic behavior, the soiling was ego-dystonic, and he was highly motivated to change the behavior, although this motivation could not be prospectively detected by the treatment team because of his oppositional defiant manner of responding to adults.

Pathology and Laboratory Examination

The extensive, detailed physiologic studies cited previously suggest that physiologic factors are important in many children with encopresis, given that they correlate with treatment outcomes at 6-month and 1-year follow-up. However, that does not imply that a physiologic workup that is this thorough is required for every child who presents clinically. The clinician needs to rule out definite physiologic illnesses, such as Hirschsprung disease. Beyond that, clinicians might consider more detailed physiologic investigations, such as those used in the experiments described previously, for children who prove refractory to conventional treatment. That type of specialized testing could also be difficult to obtain outside of an academic research center. A simple plain abdominal roentgenogram may aid in the diagnosis and confirmation of retentive encopresis. In one study, 78% of children who met the diagnostic criteria for encopresis also met roentgenographic criteria, whereas 22% did not.

Children with a condition referred to as *slow transit constipation* may present with intractable encopresis. Although the etiology of this disorder is not fully understood, some of these children appear to have abnormal innervation of the colon. Clinical signs of this disorder include severe constipation within the first year of life and refractory encopresis at an early age. Colonic transit studies are necessary to confirm the diagnosis.

Psychological testing can be useful for those children who exhibit concomitant behavioral problems or for those who are intentionally encopretic. Finally, clinicians should not overlook the importance of a detailed history obtained from the parents and the child that documents the time course, frequency, and circumstances surrounding the encopresis.

Differential Diagnosis

Encopresis can be a symptom of other processes or can be a syndrome in itself, as described previously. The medical illnesses that need to be ruled out include stenosis of the rectum or anus, endocrine abnormalities, smooth muscle disease, and Hirschsprung disease.

The child's psychological profile should be considered because children with intellectual deficits or pervasive developmental delay may have episodes of soiling related to those processes. Children with impulse-control disorders or attention-deficit disorder may, at times, have episodes of soiling related to lack of attention. Encopresis has been reported as one of a number of symptoms that can be seen following sexual abuse. A large retrospective study investigated the hypothesis that the association between encopresis and sexual abuse was of sufficient significance and that encopresis could be classified as an indicator of sexual abuse. The three study groups (4 to 12 years of age) included 466 children with documented sexual abuse, 429 children with externalizing psychiatric disorders, and 641 controls. A history of sexual abuse had been ruled out in the latter two groups. The rates of fecal soiling were similar in both the abuse (10.3%) and psychiatric groups (10.5%). The soiling rate in the control group was much lower (2%). Thus, the authors concluded that the symptom of fecal soiling could not, in and of itself, be used as a predictive indicator of sexual abuse. A thorough history and psychiatric assessment should make it possible to identify those children, as well as those children who soil on an intentional-oppositional basis. Extreme stress may also provoke encopretic episodes in children who are otherwise functioning well.

Course and Prognosis

As with enuresis, continued maturation provides increasing numbers of spontaneous remissions over time. However, the rate of spontaneous remission is not as clearly defined as that for enuresis. That trend should be factored into the assessment of the efficacy of any long-term intervention. The results from relatively simple behavioral interventions involving educational, behavioral, and physiologic components are striking (78%), suggesting that the disorder responds to treatment in most cases.

Treatment

The first clinical approach to encopresis is primarily behavioral, with educational and physiologic components. That method entails initial educational interventions with the child and family around bowel function. The process is meant to correct misconceptions that the family may have and to dissipate anxiety that may develop around the soiling. The physiologic component is instituted next with an initial bowel catharsis, followed by daily doses of laxatives or mineral oil. The behavioral element of the program entails daily timed intervals on the toilet, with success being rewarded. Follow-up studies have reported a 78% success rate for this method. The physiologic research findings concerning abnormal defecation dynamics in children with encopresis led to an interest in biofeedback training as an adjunctive treatment. Although this treatment received positive reports from uncontrolled studies, a review of controlled studies indicated that biofeedback training did not provide additional benefit over conventional treatment. A large, long-term, controlled outcome study that compared conventional treatment with biofeedback training plus conventional treatment indicated that the positive results for both groups were high and quite similar. Specifically, the conventional method yielded an 86% improvement rate, as compared to 87% for the combined group. This suggests that the combination treatment was no better than the conventional treatment alone. However, it was also found that length of time at follow-up was a powerful factor because it significantly correlated with improvement for the group as a whole at the $P < .01$ level. This indicates that the passage of time is an important variable to consider in any investigation of treatment approaches.

A recent systematic meta-analysis of the literature with regard to behavioral and/or cognitive treatments for fecal incontinence concluded that treatment with biofeedback did not improve the results obtained from the traditional treatment.

One element of the aforementioned physiologic studies, namely, the ability to defecate a rectal balloon, has been shown to correlate with treatment outcome. A study involving 139 children with constipation and encopresis and 20 controls found that all of the controls were able to defecate the balloon, as compared to 47% of the encopretic group.

One year after beginning treatment, 51% of the children who could defecate the balloon at baseline were considered to be recovered, but only 34% of those who could not defecate the balloon were improved. Although the group comparisons were considered significant, the balloon test alone could not predict response to treatment on an individual basis.

Psychotherapy may be of use for children who have concomitant behavioral problems and those with psychodynamic factors that appear to be especially strong and are contributing to the disorder. This would also be true for those who began soiling after a stressful or traumatic event. Naturally, psychotherapeutic approaches are of primary importance for children who are intentionally soiling or depositing feces in inappropriate places.

Pharmacologic treatment for encopresis has not been extensively studied or reported. There are six case reports (encompassing 15 individual patients) that have described imipramine as effective for encopresis. Twelve of the 15 patients were boys. The therapeutic effect had relatively rapid onset (as it does with enuresis), occurring within a few days to 2 weeks, usually at doses of 25 to 75 mg. Although the subtype of encopresis is not clearly delineated in all of the case reports, one would anticipate that, if imipramine were to be helpful, it would be in those with nonretentive encopresis. There is one well-controlled study that substantiated the efficacy of the prokinetic-propulsive agent cisapride (Propulsid). This agent is no longer commercially available in the United States, having been removed from the market owing to concerns about side effects.

The relatively high success rate of the combined educational, physiologic, and psychological approach described makes it a reasonable first approach once other diagnoses have been ruled out. Psychotherapeutic and pharmacologic approaches could be considered for children who prove refractory to that form of treatment.

FURTHER READINGS

Baeyens D, Roeyers H, Demeyer I, Verté S, Hoebeke P, Walle JV. Attention-deficit/hyperactivity disorder (ADHD) as a risk factor for persistent nocturnal enuresis in children: a two-year follow-up study. *Acta Paediatr*. 2005;94(11):1619–1625.

Benninga MA, Voskuijl WP, Akkerhius GW, Taminiau JA, Buller HA. Colonic transit times and behaviour profiles in children with defecation disorders. *Arch Dis Child*. 2004;89:13–16.

*Borowitz SM, Cox DJ, Sutphen JL, Kovatchev B. Treatment of childhood encopresis: a randomized trial comparing three treatment protocols. *J Pediatr Gastroenterol Nutr*. 2002;34:378–384.

Brazzelli M, Griffiths PV, Cody JD, Tappin D. Behavioral and cognitive interventions with or without other treatments for the management of faecal incontinence in children. *Cochrane Database Syst Rev*. 2011;2011(12):CD002240.

Butler RJ, Holland P, Gasson S, Norfolk S, Houghton L, Penney M. Exploring potential mechanisms in alarm treatment for primary nocturnal enuresis. *Scand J Urol Nephrol*. 2007;41(5):407–413.

El-Mitwalli A, Bediwy AS, Zaher AA, Belal T, Saleh ABM. Sleep apnea in children with refractory monosymptomatic nocturnal enuresis. *Nat Sci Sleep*. 2014;6:37–42.

Freeman KA, Riley A, Duke DC, Fu R. Systematic review and meta-analysis of behavioral interventions for fecal incontinence with constipation. *J Pediatr Psychol*. 2014;39(8):887–902.

Freitag CM, Rohling D, Seifen S, Pukrop R, von Gontard A. Neurophysiology of nocturnal enuresis: evoked potentials and prepulse inhibition of the startle reflex. *Dev Med Child Neurol*. 2006;48(4):278–284.

Fritz G, Rockney R, Bernet W, et al. Practice parameter for the assessment and treatment of children and adolescents with enuresis. *J Am Acad Child Adolesc Psychiatry*. 2004;43(12):1540–1550.

Glazener CMA, Evans JHC, Peto RE. Alarm interventions for nocturnal enuresis in children. *Cochrane Database Syst Rev*. 2005;(2):CD002911.

Hergüner S, Kilincaslan A, Gorker I, Tüzün U. Serotonin-selective reuptake inhibitor-induced enuresis in three pediatric cases. *J Child Adolesc Psychopharmacol*. 2007;17(3):367–369.

Hergüner S, Mukaddes NM. Risperidone-induced enuresis in two children with autistic disorder. *J Child Adolesc Psychopharmacol*. 2007;17(4):527–530.

Joinson C, Heron J, Butler R, et al. A United Kingdom population-based study of intellectual capacities in children with and without soiling, daytime wetting, and bedwetting. *Pediatrics*. 2007;120(2):e308–e316.

Joinson C, Heron J, Butler U, von Gontard A, Avon longitudinal study of parents and children study team. Psychological differences between children with and without soiling problems. *Pediatrics*. 2006;117(5):1575–1584.

*Joinson C, Heron J, Emond A, Butler R. Psychological problems in children with bedwetting and combined (day and night) wetting: a UK population-based study. *J Pediatr Psychol*. 2007;32(5):605–616.

Klages T, Geller B, Tillman R, Bolhofner K, Zimerman B. Controlled study of encopresis and enuresis in children with a prepubertal and early adolescent bipolar-I disorder phenotype. *J Am Acad Child Adolesc Psychiatry*. 2005;44(10):1050–1057.

Landgraf JM, Abidari J, Cilento BG Jr, Cooper CS, Schulman SL, Ortenberg J. Coping, commitment, and attitude: quantifying the everyday burden of enuresis on children and their families. *Pediatrics*. 2004;113:334–344.

Lottmann H, Froeling F, Alloussi S, El-Radhi AS, Rittig S. A randomized comparison of oral desmopressin in lyophilisate (MELT) and tablet formulations in children and adolescents with primary nocturnal enuresis. *Int J Clin Pract*. 2007;61(9):1454–1460.

McKeown C, Hisle-Gorman E, Eide M, Gorman GH, Nylund CM. Association of constipation and fecal incontinence with attention-deficit/hyperactivity disorder. *Pediatrics*. 2013;132(5):e1210–e1215.

Mellon MW, Natchev BE, Katusic SK, et al. Incidence of enuresis and encopresis among children with attention-deficit/hyperactivity disorder in a population-based birth cohort. *Acad Pediatr*. 2013;13(4):322–327.

Mellon MW, Whiteside SP, Friedrich WN. The relevance of fecal soiling as an indicator of child sexual abuse: a preliminary analysis. *J Dev Behav Pediatr*. 2006;27(1):25–32.

Mikkelsen EJ. Enuresis and encopresis: ten years of progress. *J Am Acad Adolesc Psychiatry*. 2001;40:1146–1158.

Önol FF, Guzel R, Tahra A, Kaya C, Boylu U. Comparison of long-term efficacy of desmopressin lyophilisate and enuretic alarm for monosymptomatic enuresis and assessment of predictive factors for success: a randomized prospective trial. *J Urol*. 2015;193(2):655–661.

Robben JH, Sze M, Knoers NV, Eggert P, Deen P, Müller D. Relief of nocturnal enuresis by desmopressin is kidney and vasopressin type 2 receptor independent. *J Am Soc Nephrol*. 2007;18(5):1534–1539.

Robson WLM, Leung AKC, Norgaard JP. The comparative safety of oral versus intranasal desmopressin for the treatment of children with nocturnal enuresis. *J Urol*. 2007;78(1):24–30.

Schulz-Juergensen S, Rieger M, Schaefer J, Neusuess A, Eggert P. Effect of 1-desamino-8-D-arginine vasopressin on prepulse inhibition of startle supports a central etiology of primary monosymptomatic enuresis. *J Pediatr*. 2007;151(6):571–574.

Touchette E, Petit D, Faquet J, Tremblay RE, Boivin M, Montplaisir JY. Bed-wetting and its association with developmental milestones in early childhood. *Arch Pediatr Adolesc Med*. 2005;159(12):1129–1134.

Van Hoecke E, De Fruyt F, De Clercq B, Hoebeke P, Walle JV. Internalizing and externalizing problem behavior in children with nocturnal and diurnal enuresis: a five-factor model perspective. *J Pediatr Psychol*. 2006;31(5):460–468.

*von Gontard A, Freitag CM, Seifen S, Pukrop R, Rohling D. Neuromotor development in nocturnal enuresis. *Dev Med Child Neurol*. 2006;48(9):744–750.

von Gontard A, Schaumburg H, Hollmann E, Eiberg H, Rittig S. The genetics of enuresis: a review. *J Urol*. 2001;166:2438–2443.

49

Other Disorders of Infancy, Childhood, and Adolescence

49.1 Reactive Attachment Disorder and Disinhibited Social Engagement Disorder

KARIN M. BEST, PH.D., MARY MARGARET GLEASON, M.D., AND CHARLES H. ZEANAH, M.D.

As early as the 13th century, astute observers and lay researchers documented a range of atypical reactions among some children who experience social neglect or physical and social neglect. Observational data and research findings coalesced in the 1970s with the inclusion of reactive attachment disorder (RAD) in infancy and early childhood in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III, 1980). The description of RAD was framed in the observational and theoretical work of John Bowlby.

The DSM-III's revision (DSM-IIIR, 1987) and the next *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV, 1994; DSM-IV-TR, 2000) delineated two subtypes of RAD—an emotionally withdrawn/inhibited pattern and an indiscriminately social/disinhibited pattern. The differentiation was grounded in emerging research that provided evidence suggesting that the subtypes differed in presentation, neurobiology, and developmental course. Research continued to demonstrate that both subtypes emerged only after an infant or very young child experienced significantly inadequate caregiving.

Informed by the seminal research of the English-Romanian Adoption (ERA) study and the Bucharest Early Intervention Project (BEIP), the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) and its text revision (DSM-5-TR) defined the subtypes as two distinct disorders: RAD, which describes children who display the emotionally withdrawn/inhibited pattern of behavior; and disinhibited social engagement disorder (DSED), which describes children who demonstrate the indiscriminately social/disinhibited pattern. These disorders often continue to be categorized as “disorders of attachment,” and the criteria for each reference are early-rearing conditions that limit the development of attachments. However, whether DSED should be considered an attachment disorder remains controversial. DSM-5 classified both Trauma and Stress-Related Disorders, and DSM-5-TR took the same approach. RAD and DSED in DSM-5-TR are also defined in the *International Classification of Disease, eleventh edition* (ICD-11), with similar criteria.

The ERA study and BEIP have contributed to the behavioral delineation of RAD and DSED and the development of “gold standard” methodology and measures for the diagnosis of young and early school-age children. In the past 10 years, research has broadened to include efforts to characterize them in adolescence and in the adult years, with the goal of better understanding their respective sequelae. This research includes a detailed follow-up of the ERA and BEIP samples as well as a retrospective application of childhood diagnostic criteria to high-risk adolescent samples, some of which have extensive childhood health registry data available. The diagnostic precision of studies that apply diagnostic criteria retrospectively is limited, and findings must be interpreted with due caution as exploratory. The efforts include attention to delineating the adolescent behavioral equivalents of RAD and DSED, which both the DSM-5-TR and ICD-11 define with criteria specific to younger children. Additional research also focuses on identifying familial and life experience risk factors for the development of RAD and DSED, exploring child characteristics and skills that may mediate and moderate the long-term sequelae, and identifying alterations in brain structure and activity associated with each diagnosis. This work is central to the development and refinement of evidence-based intervention techniques and programs.

Clinically, the diagnoses of RAD and DSED have been embraced, and there is evidence that both are *overdiagnosed in community settings*. In 2016, the American Academy of Child and Adolescent Psychiatry published practice parameters to guide accurate diagnosis and to discredit dangerous and unproven interventions, such as various holding/restraints, that have resulted in the death of children.

In what follows, we highlight contexts and details relevant to the diagnosis of RAD and DSED. Our goals are to support accurate assessment and diagnosis, provide an update on emerging information about the long-term sequelae of RAD and DSED symptomatology, and provide information regarding care and evidence-informed treatment.

ATTACHMENT THEORY

The normal development of the infant and young child's attachment system, first described by John Bowlby, is the trajectory to which behavior indicating disordered attachment is compared (Table 49.1–1).

Attachment, as described by Bowlby's influential theory, refers to a biobehavioral system that coordinates the balance between the very young child's need, on the one hand, to seek safety by proximity to a caregiver or set of caregivers and, on the other, to explore surroundings, to seek novelty, and to develop mastery. Bowlby argued that human infants are motivated by an attachment behavior control system to seek proximity to a small number of identified caregivers. He inferred that the reason for seeking proximity was to attain an internal feeling of security. This internal sense of security is reinforced by repeated experiences of the caregiver's emotional and physical availability in times of need.

Table 49.1–1.
Developmental Features of Attachment in Early Childhood

	Birth–2 Months	2–7 Months	7–12 Months	12–18 Months	Older than 18 Months
Phase of attachment	Limited discrimination	Discrimination with limited preference	Preferred attachment	Secure base	Goal-oriented partnership
Characteristics	Physical attributes of babyishness attract caregivers, but infant expressions of preference are limited to olfactory and auditory realms and require experimental conditions to be demonstrated.	Infant differentiates among different interactive partners; infant may seem more comfortable with primary caregiver but is willing to interact readily with other social partners.	Clearly expressed preference for a small number of caregiving adults; separation protest and stranger wariness are normative.	Use of the attachment figures as a secure base from which to venture out confidently and to explore the world and a safe haven to return to in times of danger; proximity to caregiver promotes feeling of security in infant.	Co-operation with caregiver despite frequent conflicts; goal is to balance autonomous functioning with reliance on caregiver for help when needed; verbal relatedness is important, including attention to emotional integration and fluency of discourse.

Beginning between 7 and 9 months of age, infants are thought to develop internal representations of their relationships with important caregivers. These representations, or internal working models, form the basis for the intense emotional bond between infants and their primary caregivers and their behavior in later relationships. The normal developmental progression of attachment, captured in attachment-related behaviors, is summarized in Table 49.1–1. The commonly accepted method for classifying patterns of attachment is the strange situation procedure, which includes interaction with an attachment figure and an unfamiliar adult in various combinations. Differences in the balance between attachment and exploratory behavior and how those are expressed are used to classify the infant's attachment to this caregiver into one of four types of attachment: *secure*, in which children express distress directly, seek comfort actively, and obtain soothing readily from attachment figures; *avoidant*, in which children minimize signs of distress and show no comfort seeking efforts despite physiologic arousal; *resistant*, in which children show significant distress to separation but fail to obtain comfort from their attachment figure and cannot be soothed; or *disorganized*, in which children show some mixture of the secure, avoidant or resistant, or anomalous behaviors that preclude classification as another pattern. These patterns of attachment represent risk or protective factors rather than diagnoses.

IMPACT OF EARLY ENVIRONMENT

The importance of an infant's early experience was recognized as far back as the 13th century when the Holy Roman Emperor Frederick II reportedly initiated an experiment to see what language children would develop independently. He ordered that a group of infants be provided simple physical care but not spoken to by their caregivers. According to the Franciscan monk whose chronicle recounted the events, the experiment ended with the premature death of the children, who "could not live without clapping of the hands, and gestures, and gladness of countenance, and blandishments."

The relatively common practice of institutionalizing infants in the United States in the first half of the 18th century unwittingly provided more evidence of the detrimental effects of what has been called "maternal deprivation." Although pediatricians had long argued against the practice of institutionalization, particularly because of the high mortality rates related to failure to thrive, other healthcare providers did not take notice of the effects of institutionalization on social and emotional development until René Spitz (in the 1940s) and James and Joyce Robertson (in the 1950s) conducted more rigorous studies (including control groups). Spitz subsequently proposed a diagnostic entity that he called *anaclitic depression*, which foreshadowed the DSM-III criteria for RAD. Anaclitic depression, which included mood and social symptoms as well as immunologic and growth problems, seemed to follow the prolonged separation of infants from their caregivers. It was most severe when the infants were old enough to have developed a preference for the caregiver from whom they were separated.

In the 1970s, research by Barbara Tizard identified the two sets of behaviors that inform the current DSM-5-TR and ICD-11 classifications of attachment disorders. The children she studied resided in institutions with relatively low caregiver-to-child ratios, mixed age groups, and toys and books to enrich the environment. Their caregivers, however, were explicitly instructed not to form attachments to the children. Eight of the 26 children managed to form secure attachments, 8 were emotionally withdrawn and socially unresponsive (similar to children with RAD), and 10 were indiscriminately social with unfamiliar adults (as seen in children with DSED and DAD).

Recognition of the detrimental clinical consequences of institutional rearing eventually led to a marked decrease in state-run orphanages in the United States and the United Kingdom. Nevertheless, institutional care remains widely used for orphaned, abandoned, and maltreated young children in Asia, Central and South America, and Europe, particularly Eastern Europe.

Since the early 1960s, spurred by C. Henry Kempe's classic 1962 article on the *battered child syndrome*, recognition of and research on the developmental effects of maltreatment on children have grown. The research documents substantial consequences of abuse and serious neglect, including associations between abuse and serious disturbances of attachment.

Thus, descriptions of the phenotypes associated with attachment disorders emerged in early case reports and descriptive studies, but systematic research appeared only in the past four decades. Studies of institutionalized and maltreated infants and young children provided broad support for the phenotypic patterns of disordered attachment, now referred to as RAD and DSED.

EVOLVING NOSOLOGY

The criteria for RAD have evolved markedly since the diagnosis was introduced. The DSM-III description included as central features growth failure and lack of social responsivity. The diagnosis had to start by 8 months of age (the age at which preferred attachment to a restricted set of caregivers is usually just beginning to be evident) and could not result from a diagnosable medical condition. Gross neglect of the infant's physical and emotional needs had to be evident, and the presence of autism or mental retardation precluded the diagnosis. Pertinent behaviors cited in the criteria included poor tone, weak cry, excessive sleep, lack of interest in the environment, and weak rooting and grasping when feeding.

Although failure to thrive is an important problem in infant mental health, it is not equivalent to an attachment disorder. The DSM-III's restrictive age range criterion and the requirement for growth failure significantly restricted the application of the diagnosis. Furthermore, it is rarely possible to rule out an autistic disorder or mental retardation in severely malnourished infants younger than 1 year of age. DSM-III-R dropped the failure to thrive criterion and broadened the age of onset to within the first 5 years of life. DSM-III-R also introduced two subtypes of RAD—inhibited and disinhibited—and these were retained through the DSM-IV versions. DSM-IV criteria continued to highlight the etiologic link between RAD and "pathogenic care" at the hands of the young child's primary caregivers and excluded children whose symptoms might be due to cognitive delay or one of the pervasive developmental disorders.

With the publication of the DSM-5, the subtypes of RAD were defined as separate disorders: RAD and DSED. This separation more closely aligned DSM and ICD nosology. The ICD-10 included Reactive Attachment Disorder of Childhood (RAD), which is similar to RAD in the DSM-5, and Disinhibited Attachment Disorder of Childhood (DAD), which is similar to DSED in the DSM-5. ICD-11 instead mirrored the DSM-5 approach with the diagnoses of RAD and DSED, using similar criteria to the DSM.

A significant body of research examining RAD and DSED criteria has affirmed that they can be reliably identified in high-risk children, and the diagnoses are rare in clinical samples, absent histories of severe neglect. Research also indicates that children with RAD respond to improvements in the caregiving environment. For example, in a prospective randomized controlled trial of foster care as an alternative to institutional care in Romania (BEIP), children with

RAD randomized to foster families demonstrated a reduction in symptomatology within months of placement, with symptom levels decreasing to levels in a community sample of never-institutionalized children. In contrast, children with DSED randomized to foster care continued to exhibit behaviors consistent with the diagnosis after family placement, typically exhibiting a reduction but not an elimination of their signs of DSED. Consistent with this, postplacement studies of Romanian children adopted from institutions by English families (ERA) show that these children do not continue to manifest RAD but do demonstrate persistent signs of DSED. That signs of RAD, which reflects a lack of attachment, diminished when children had the opportunity to form an attachment to foster or adoptive parents, and that signs of DSED may persist in a minority of children even after they form attachments to foster or adoptive parents suggests that the core deficit in DSED may not be the lack of an attachment relationship.

ETIOLOGY

Although research has established that the risk of RAD and DSED increases in the context of grossly neglectful care, the degree of neglect that is necessary for the disorders to emerge is not clear. In fact, even in environments characterized by severe deprivation, most children do not develop signs of RAD or DSED. Of note, however, relatively few children who experience extended institutional care emerge without symptoms of psychopathology. (Only 15% of children who were adopted after at least 6 months of institutional care in Romania did not meet the diagnostic threshold for any disorder at any point through their mid-20s.) More research on vulnerability factors and the critical elements of neglectful care responsible for RAD, DSED, and other disorders is needed.

The longer children live in conditions of deprivation, the more likely they are to develop an attachment disorder. Additionally, for children removed from institutional care, the number of caregiving disruptions or changes is associated with the persistence of RAD up to age 12. A critical age of vulnerability to the disorder may exist, as research has found attachment disorders to be very rare in children adopted before 6 months of age, and signs of DSED have not been reported in children whose deprivation began only after age 2. Children with lower IQs and stunted growth may also have attachment disorders because deprivation compromises both IQ and growth; hence, it is difficult to know whether these problems precede, co-occur, or are merely additional consequences of deprivation. Studies are beginning to identify possible genetic polymorphisms as risk factors for attachment disorders.

Preliminary research suggests that children who have experienced serious deprivation may show differences in brain structure and functioning compared to nondeprived children, and many of these children may show signs of attachment disorders. For example, researchers have found reductions in gray and white matter volumes and reduced electrical activity in higher frequencies, and increased electrical activities in lower frequencies in children who have experienced institutional deprivation. Importantly, reduced white matter volume mediated the effect of institutional rearing on electroencephalogram (EEG) alpha power, meaning that reductions in myelinated circuitry likely account for the reduced EEG power. If children were placed with families before 24 months of age, the differences in electrical activity between deprived and nondeprived children were no longer seen at 8 years of age. These findings have been demonstrated in children with histories of institutional deprivation and are not specific to those with attachment disorders, however. At age 16, the institutionalized children not randomized to foster care had higher relative theta and lower relative alpha power on EEG, a pattern characterized as immature, compared to those who experienced foster care. The brain activity of the fostered children did not differ from that of a community comparison group. Of note, however, the number of placement disruptions (but not the age of placement) among the fostered group was associated with adverse changes in brain activity.

Difficulties with inhibitory control also have been shown in children with indiscriminate social behaviors, implicating ventral frontostriatal circuitry, though findings are not consistent across studies. Children adopted from institutions also showed reduced amygdala discrimination between mothers and strangers on functional magnetic resonance imaging (fMRI), and these reductions were moderately associated with the children showing indiscriminate behaviors. Thus, preliminary research is demonstrating that early deprivation has an impact on brain development, although this relationship has been shown primarily in children with behaviors consistent with a diagnosis of DSED and requires replication.

Studies employing a variety of magnetic resonance imaging (MRI) techniques have variously associated RAD with decreased volume in the left visual cortex and differences in the corpus callosum and projections in thalamic pathways. These studies, however, have been conducted with children diagnosed in the community as early adolescents and may be compromised by diagnostic imprecision.

Recent research suggests that institutional rearing increases the risk of lower executive functioning, externalizing, and “p” (a latent measure of risk for general pathology) at age 12; each measure is further associated with the risk of general pathology at age 16, suggesting that lower executive functioning contributes to challenges over time.

Although research on what causes caregivers to be neglectful is limited, caregiver characteristics, from severe personality disturbance to psychopathology, appear to be important. Research using registry data exploring risk factors for the development of RAD among children who did not experience early institutional placement suggests that prenatal maternal smoking, advanced paternal age, maternal or paternal psychiatric disorder, and single motherhood all increase the risk of the RAD diagnosis. Parental intellectual disability and difficulty meeting a baby’s basic needs, particularly in the context of limited social support, may also increase the risk of neglect, though neglect sufficiently severe to contribute to RAD and DSED is rare, even in families with these challenges. Infants who are cared for by multiple caregivers in succession, as sometimes happens in the foster care system in the United States, are also at increased risk of attachment disorders. The traumatic experience of prolonged separation from a caregiver in early childhood, as documented by

Spitz, the Robertsons, and Bowlby, is enough to cause attachment disturbance (as distinct from RAD and DSED). Repeated, prolonged separations typically cause more severe symptoms, and when these separations are accompanied by neglect, DSED and RAD appear to be more likely.

EPIDEMIOLOGY

Very little data exist on the prevalence of either RAD or DSED in the general population. RAD is rarely diagnosed in children with adequate caregiving environments and requires but does not necessarily follow a history of severe deprivation. DSED appears to be more prevalent, though it also is diagnosed only in a small minority of children who experienced inadequate caregiving.

Among high-risk groups diagnosed retrospectively, such as adolescents in tertiary therapeutic treatment settings, prevalence estimates for RAD and DSED each approach 10%.

DIAGNOSIS AND CLINICAL FEATURES

The diagnosis of RAD or DSED requires that a child have a developmental age of at least 9 months (the cognitive level at which attachment formation is possible) and have experienced a high level of insufficient care, which is presumed causative of each disorder.

RAD is characterized by both inhibition of positive emotional and social engagement *and* the lack of selective or discriminating attachment behaviors directed at an adult caregiver. At times when young children should be motivated to seek closeness or comfort for distress or protection from danger, children with RAD do not seek or readily respond to adults’ efforts to comfort, reassure or protect them. Instead, the children may appear withdrawn and unresponsive or may show distress with no anticipation of comfort. Emotion regulation abnormalities include reduced or absent expression of positive emotions and unexplained displays of irritability, sadness, or anxiety. Among children with the requisite history of severe neglect, operationalized as social neglect/deprivation, repeated changes of primary caregivers that limit opportunities to form stable attachments (e.g., frequent changes in foster care), and/or rearing in group settings that severely limit opportunities to form selective attachments (e.g., institutions with high child-to-caregiver ratios), signs of RAD typically recede within weeks of placement in an adequate caregiving environment. Occasionally, symptoms of RAD persist following placement. Symptom persistence often marks continuing de facto neglect. The de facto neglect occurs because the child's withdrawn and negative behavior is off-putting to adult caregivers. The caregivers fail to override their natural responses to the child's behavior, which obscures the child's needs and avoids interactions.

For diagnosis of RAD, a child must first exhibit symptoms before age 5 and not meet the criteria for autism spectrum disorder (ASD).

DSED is characterized by a lack of stranger wariness, which ordinarily becomes apparent between 7 and 9 months of age. The disinhibition in DSED is social, leading to behaviors in which a child actively approaches and interacts with unfamiliar adults. Although sometimes referred to as indiscriminate friendliness, the behavior is often experienced as intrusive and may include overly familiar verbal or physical behavior that goes beyond culturally sanctioned and age-appropriate social boundaries. Children with DSED often fail to track their caregiver's whereabouts, even in unfamiliar settings, and may even be willing to go off with an unfamiliar adult without hesitation. DSED criteria stipulate that the disordered behaviors include socially disinhibited behaviors beyond the behavioral impulsivity that occurs in attention-deficit/hyperactivity disorder (ADHD; ADHD may co-occur).

Diagnosis of RAD or DSED requires a multisession evaluation, particularly when the child's history does not include multiple placements and the caregiver's behavior has not previously been identified as abusive or neglectful. Ascertaining whether the child's behavior is less atypical in the presence of the clinician or another trusted adult is important to differential diagnosis and may take time. This comparison may be especially helpful in distinguishing highly social children, for example, from those with DSED or somewhat behaviorally inhibited children from those with RAD. The child's relationship with the primary caregiver can be adequately assessed with a combination of a careful history and observations of interactions during free play, structured tasks (that stress the dyad), and a brief separation and reunion. Videotaping the assessment is a useful way to document social responses, and the introduction of a clinically trained person whom neither partner has met is often informative. Children who meet the criteria for this disorder may look quite different, depending on their developmental stage.

An 11-month-old girl was recently placed in foster care. Foster care placement occurred after police, who were called to the home during an episode of violence, found squalid conditions, including narcotic pain pills strewn on the floor. Her parents both had chronic and serious histories of substance use disorders, violence in their relationship, and grossly inadequate care of the child and her older siblings. The girl, with her biologic and foster families, was referred for clinical services by state child protective services to assist with long-term case management.

The foster mother reported that the child looked frightened much of the time, cried sporadically and was not soothed by her, and did not respond to social overture when she first arrived in the foster home. For the initial 2 weeks of placement, the child exhibited little interest in exploring and mostly rocked quietly by herself. The girl's appetite increased over the first few days, and she began to stuff herself to the point that the foster mother, at the direction of the child's pediatrician, limited the food offered at meals and snacks. The foster mother reported it was weeks before the girl smiled. This behavioral pattern was suggestive of RAD.

The clinicians encouraged the foster mother to interact with the child despite the girl's lack of interest and to respond to the girl's distress as if the girl had requested it in order to teach her that she could be comforted when distressed. Gradually, the girl seemed to relax more and began to smile—at first fleetingly but then more frequently, to accept comfort for distress and then to seek comfort for her distress. By one month following placement, the foster mother reported the child seemed to be showing attachment behaviors toward her. The foster mother also described feeling closer to the girl, whom the foster mother had found hard to connect with initially. After this point, the ravenous behavior slowly dissipated. The girl became somewhat clingy with her foster mother and was quick to show distress to even minor changes in routine. Her affect brightened, and she came to demonstrate better emotion regulation.

A 45-month-old boy was referred for services by his adoptive parents because of hyperactivity and disruptive behavior at school. He had joined his parents at 30 months, after living most of his life in a Romanian orphanage, where a rotating staff of caregivers provided his care. When he arrived at his new home, he was below the fifth percentile for height and weight; his growth accelerated, and he quickly approached the 10th percentile. Both his adoptive parents were frustrated by their inability to "reach him." They had initially worried he might have a hearing impairment, but testing and his capacity to engage many adults and children verbally suggested otherwise. He approached and engaged socially with complete strangers and, more than once, had followed strangers willingly. Teachers liked him; they also found him attention-seeking and sometimes disruptive. At times, teachers experienced him as appealing, even charming, but at other times, they characterized his feelings as shallow. He was injured frequently due to seemingly reckless behavior and had an extremely high pain tolerance. He was diagnosed with DSED. The intensive intervention focused on decreasing his self-endangering behavior, enhancing his ability to read emotional expressions and describe social situations, and rewarding him for culturally appropriate social hesitance. With several months of treatment, his reckless behaviors decreased, and his ability to read and describe social situations improved. His tendency to engage with unfamiliar adults persisted at a reduced and more manageable level.

DIFFERENTIAL DIAGNOSIS

Since RAD and DSED differ in clinical presentation, various diagnoses must be evaluated in the differential diagnosis of each. For RAD: ASD, and depressive disorders require evaluation; for DSED: ADHD. Additionally, a child's intellectual ability must be considered; classifying some behaviors as atypical occurs in relation to a child's developmental level.

Depressive Disorders

Children with RAD, and those with depression, both show reduced or absent positive affect and social withdrawal. In a longitudinal study of current or previously institutionalized young children, there was a moderate-to-high association between RAD and signs of depression, but the majority of children who were diagnosed with a depressive disorder did not meet the criteria for RAD. In general, children with depression turn to caregivers for comfort and show behaviors that are indicative of an attachment. The neurovegetative signs of depression and the excessive guilt associated with depression are not characteristic of RAD. Only if a child with depression has a history of significant social neglect and does not seek comfort should the diagnosis of RAD be considered.

Autism Spectrum Disorder

The differential diagnosis of RAD and ASD may be challenging in some cases because quasi-autistic behaviors have been noted in formerly institutionalized, adopted children. Of note, however, the course of quasi-autism varies from that of ASD in that at least partial remission may occur after several years of familial placement.

Children with both ASD and RAD show impaired social behaviors, especially social withdrawal, and minimal reciprocity. However, children with ASD typically also show atypical language development; the language skills of children with RAD may be delayed but are not stereotyped or otherwise deviant. Additionally, children with ASD exhibit restricted and repetitive behaviors; these are not found in children with RAD. Finally, children with ASD demonstrate attachment behaviors with caregivers some of the time and very rarely have the history of social neglect or maltreatment required for the diagnosis of RAD.

Attention-Deficit/Hyperactivity Disorder

Deprivation is known to increase the risk both of inattentive and hyperactive behaviors and of DSED. Children with DSED can also meet the criteria for ADHD, requiring careful assessment of ADHD as an alternative diagnosis as well as a potential comorbid condition. Young children who are hyperactive and impulsive may appear socially disinhibited. The social disinhibition of DSED, however, is not necessarily accompanied by inattention or hyperactivity.

Intellectual Disability (Intellectual Developmental Disorder)

The adverse caregiving environment necessary for the manifestation of RAD or DSED may include environmental deprivation associated with developmental delays. However, in longitudinal studies of children from institutions, RAD and DSED were only moderately associated with cognitive delays. In terms of differential diagnosis, children with intellectual disability typically demonstrate social and emotional development that is relatively consistent with their cognitive functioning. Further, it is well established that children with intellectual disability are capable of forming stable attachments as long as their developmental ages are greater than 9 months.

COURSE AND PROGNOSIS

Research addressing the natural course of RAD and DSED diagnoses and symptomatology is limited but increasing.

Among children in institutions, RAD appears relatively stable; young children who met the criteria for RAD at one point in time continued to show the behaviors at other times. Children who are placed into families typically do not show significant signs of RAD over time, which suggests that a stable caregiving environment ameliorates symptoms of the disorder. In the Bucharest study, symptoms of RAD at age 12 were associated with poorer social functioning. Data from the same study, utilizing variable and person center models, identified four patterns of RAD symptomatology through age 12: children who never evidence symptoms, children who evidence an early and swift decrease in symptoms, children who demonstrate persistent and low-level symptoms, and children who, although demonstrating an initial drop in symptoms, maintain a relatively higher level across time. As one might expect, the persistently high group was primarily populated with children who were not placed with foster families; they also spent significantly more time in institutions and experienced a greater number of placement disruptions. Diagnosis of RAD was associated with internalizing and externalizing symptoms and symptoms of ADHD.

Studies of young children diagnosed with DSED suggest that the diagnosis is moderately stable from year to year. Studies that have followed children into later childhood find that those with over-friendly and attention-seeking behaviors at the age of 4 tend to continue to show these behaviors at the age of eight. At age 12, symptoms of DSED were associated with lower social competence and with the caregiver's viewing the child as a victim (e.g., of bullying, relational victimization, rejection). At the age of 16, these children demonstrated superficial and conflicted relationships with peers. In adulthood, DSED symptoms may persist, with an indication that the level of these symptoms does not interfere with overall adaptation.

As only a fraction of institutionalized children are diagnosed with RAD or DESD, and most symptoms decrease, studies of long-term outcomes frequently characterize outcomes by comparing groups based on length of institutionalization before adoption and/or type of placement following initial institutionalization. Children with shorter-term institutionalized care or foster care placement resemble comparison groups (in-county adoptees or never placed children) on most measures at most ages. In the ERA study, young adults (ages 22 to 25) with longer institutional care prior to adoption had lower educational achievement, higher unemployment, and greater mental health service use in early adulthood. Group differences include increasing levels of self-reported emotional symptoms from the cusp of adolescence through adulthood; the difference in self-reported psychopathology emerged 20 years postplacement. Additionally, longer institutional care is associated with high rates of ADHD (as well as ASD and callous-unemotional traits).

Studies of adolescents in residential placement, diagnosed retrospectively with RAD or DSED, suggest those diagnosed with DSED may differ from the general population, from other adolescents in placement, and from those with RAD in aspects of self-esteem;

however, the findings require follow-up and may be compromised by diagnostic imprecision.

In sum, the expanding research corpus highlights the many persistent risks associated with all but very short-term institutionalized care to the children who experience it. Only a fifth of the Romanian adoptees who spent 6 months or more in institutions scored in the typical range across assessments of psychopathy over a 20 period. Simultaneously, research also clearly documents the potent effect of devoted care and raises hope that appropriate intervention can ameliorate symptoms, mitigate long-term risk and improve adult functioning.

TREATMENT

The first step in the diagnostic evaluation of and treatment planning for children exposed to grossly inadequate care is assessing and securing the child’s safety. Child maltreatment is associated with significant morbidity, and mortality is not uncommon, particularly in children younger than 48 months of age. Children may not have had appropriate medical care, and medical referral is usually indicated. Child protective services involvement may be warranted, and assessment of parental fitness may be a necessary part of the child’s evaluation. In some cases, the placement of the child may be necessary. Unfortunately, the foster care system and the family court system sometimes lead to multiple placements in the first years of life, thus further increasing the risk of RAD, DSED. The clinician may play a crucial role in staying involved with the child who is placed in foster care, providing expert testimony in court and individual or family treatment.

Once the child’s safety is established, full attention can be paid to the diagnosis and psychosocial intervention, which must be tailored to the child and living situation. For children who exhibit RAD and DSED, providing and supporting a primary attachment figure is the goal. Caregivers who foster secure attachments in infants and young children are those who are sensitively responsive, value the child as an individual, and place the needs of the child ahead of their own.

Caregivers with such characteristics are well-suited to assist children who have experienced severe early social deprivation.

In general, research following children with RAD suggests that the symptoms of the disorder resolve with placement in a family that can provide a positive caregiving environment. This research additionally finds that these children may continue to show insecure or disorganized attachments in the Strange Situation Paradigm, suggesting that treatment to improve attachments with foster or adoptive caregivers is important.

For children who exhibit the high levels of indiscriminate behavior seen in DSED, the clinical consensus is to advise adoptive or foster parents who are establishing attachment with the child to limit the child's unnecessary contact with adults outside the family for several months, if possible. The rationale is that children need time to consolidate attachment relationships within their family before they experience other relationships. Although this approach has face validity, it has not been evaluated formally. Research finds that some children diagnosed with DSED continue to show indiscriminate behavior, even after being placed in enhanced caregiving environments and forming secure attachments, an indication that more than enhanced caregiving may be necessary to reduce the signs of DSED significantly. It is important to note that the quality of caregiving for children with DSED is associated with the rate of decline of symptoms, even when they do not fully resolve. It has been suggested that intervention to target social cognitive abnormalities that may underlie the indiscriminate social behaviors shown in DSED would be a possible way to improve these behaviors, but research is needed in this area.

Additional interventions should be formulated to reduce child or adult (caregiver) barriers to attachment. Child barriers include behaviors that provide miscues about their needs, including aggressive and oppositional behavior, emotional dysregulation, and withdrawn behavior. Adult barriers include distractions from the newly developing relationship, such as here-and-now crises, their own problematic relationship histories—including unresolved losses and trauma—and lack of psychological commitment to the child.

Attachment-specific treatments often focus on the adult–child dyad and include the Circle of Security, which weaves together psychoeducation with psychotherapy, and Attachment and Biobehavioral Catch-Up, which is manualized and more prescriptive about which behaviors are coached. Both treatments rely heavily on a review of videotaped parent–child interactions. Other interventions that involve caregiver–child dyadic therapy include child–parent psychotherapy. These approaches typically combine developmental training and guidance with an active attempt to address pressing instrumental issues (e.g., poor housing and inadequate medical care). The complexity of these treatment approaches reflects the complexity of the clinical problem. Long-term interventions are often necessary; family and individual treatment are bolstered by other early intervention programs and ongoing medical care.

CONCLUDING THOUGHTS

In the 40 years since the addition of attachment disorders (as they were originally labeled) to the diagnostic canon, a great deal has been learned about the sequelae of profound deprivation in some children. The diagnoses RAD and DSED have been delineated, instruments and methods of assessment developed and validated, and evidence-informed treatments created. Longitudinal research, including a randomized controlled trial, indicates that placement with caring adults after deprivation effectively lessens or dissipates children’s symptoms of RAD and, to a lesser degree, of DSED. Emerging longitudinal data suggest that patterns and sequelae of these disorders may persist into young adulthood. Additionally, emerging work suggests RAD and DSED may be associated with structural and functional differences in the brain. Much remains to be learned.

Forty years of research continue to indicate that RAD and DSED are rare disorders and that they may be overdiagnosed in community settings when established diagnostic assessment instruments and procedures are not embraced and employed. The challenge in future research is to bridge, differentiate, generalize, and/or extend, as appropriate, findings from rigorous research on RAD and DSED to work on the broader topic of child abuse and neglect, as clinicians are eager for guidance.

FURTHER READINGS

Ainsworth MD, Blehar MS, Waters E, Wall S. *Patterns of Attachment: A Psychological Study of the Strange Situation*. Lawrence Erlbaum; 1985.

Bowlby J. *Attachment and Loss*. 2nd ed. Vol. 1–3. Basic Books; 1982.

Chisholm K. A three-year follow-up of attachment and indiscriminate friendliness in children adopted from Romanian orphanages. *Child Dev*. 1998;69:1092–1106.

Debnath R, Tang A, Zeanah CH, Nelson CA, Fox NA. The long-term effects of institutional rearing, foster care intervention and disruptions in care on brain electrical activity in adolescence. *Dev Sci*. 2020;23(1):e12872.

Fujisawa TX, Shimada K, Takiguchi S, et al. Type and timing of childhood maltreatment and reduced visual cortex volume in children and adolescents with reactive attachment disorder. *Neuroimage Clin*. 2018;20:216–221.

Guyon-Harris KL, Humphreys KL, Degnan K, Fox NA, Nelson CA, Zeanah CH. A prospective longitudinal study of reactive attachment disorder following early institutional care: considering variable- and person-centered approaches. *Attach Hum Dev*. 2019;21(2):95–110.

Guyon-Harris KL, Humphreys KL, Fox NA, Nelson CA, Zeanah CH. Course of disinhibited social engagement disorder from early childhood to early adolescence. *J Am Acad Child Adolesc Psychiatry*. 2018;57(5):329–335.e2.

Guyon-Harris KL, Humphreys KL, Fox NA, Nelson CA, Zeanah CH. Signs of attachment disorders and social functioning among early adolescents with a history of institutional care. *Child Abuse Negl*. 2019;88:96–106.

Guyon-Harris KL, Humphreys KL, Miron D, et al. Disinhibited social engagement disorder in early childhood predicts reduced competence in early adolescence. *J Abnorm Child Psychol*. 2019;47(10):1735–1745.

Hodges J, Tizard B. Social and family relationships of ex-institutional adolescents. *J Child Psychol Psychiatry*. 1989;30:77–97.

Hoffman KT, Marvin RS, Cooper G, Powell B. Changing toddlers' and preschoolers' attachment classifications: the Circle of Security intervention. *J Consult Clin Psychol*. 2006;74:1017–1026.

Jung M, Takiguchi S, Hamamura S, Mizuno Y, Kosaka H, Tomoda A. Thalamic volume is related to increased anterior thalamic radiations in children with reactive attachment disorder. *Cereb Cortex*. 2020;30(7):4238–4245.

Kennedy M, Kreppner J, Knights N, et al. Early severe institutional deprivation is associated with a persistent variant of adult attention-deficit/hyperactivity disorder: clinical presentation, developmental continuities and life circumstances in the English and Romanian Adoptees study. *J Child Psychology Psychiatry*. 2016;57(10):1113–1125.

Lawler JM, Hostinar CE, Mliner S, Gunnar MR. Disinhibited social engagement in postinstitutionalized children: differentiating normal from atypical behavior. *Dev Psychopathol*. 2014;26:451–464.

Lehmann S, Monette S, Egger H, et al. Development and examination of the reactive attachment disorder and disinhibited social engagement disorder assessment interview. *Assessment*. 2020;27(4):749–765.

Makita K, Takiguchi S, Naruse H, et al. White matter changes in children and adolescents with reactive attachment disorder: a diffusion tensor imaging study. *Psychiatry Res Neuroimaging*. 2020;303:111129.

Minnis H. What happens to disinhibited social engagement disorder over time? *J Am Acad Child Adolesc Psychiatry*. 2018;57(5):304–305.

Olsavsky AK, Telzer EH, Shapiro M, et al. Indiscriminate amygdala response to mothers and strangers after early maternal deprivation. *Biol Psychiatry*. 2013;74:853–860.

Rok-Bujko P. Early life trauma – review of clinical and neurobiological studies. *Postęp Psychiatr Neurol*. 2021;30(1):37–44.

Seim AR, Jozefiak T, Wichstrøm L, Lydersen S, Kayed NS. Self-esteem in adolescents with reactive attachment disorder or disinhibited social engagement disorder. *Child Abuse Negl*. 2021;118:105141.

Sonuga-Barke EJS, Kennedy M, Kumsta R, et al. Child-to-adult neurodevelopmental and mental health trajectories after early life deprivation: the young adult follow-up of the longitudinal English and Romanian adoptees study. *Lancet*. 2017;389(10078):1539–1548.

Spitz RA. Anaclitic depression; an inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal Study Child*. 1946;2:313–342.

Tizard B, Hodges J. The effect of early institutional rearing on the development of eight-year-old children. *J Child Psychol Psychiatry*. 1978;19:99–118.

Upadhyaya S, Chudal R, Luntamo T, et al. Parental risk factors among children with reactive attachment disorder referred to specialized services: a nationwide population-based study. *Child Psychiatry Hum Dev*. 2019;50(4):546–556.

Wade M, Zeanah CH, Fox NA, Nelson CA. Global deficits in executive functioning are transdiagnostic mediators between severe childhood neglect and psychopathology in adolescence. *Psychol Med*. 2020;50(10):1687–1694.

Wade M, Zeanah CH, Fox NA, Nelson CA. Social communication deficits following early-life deprivation and relation to psychopathology: a randomized clinical trial of foster care. *J Child Psychol Psychiatry*. 2020;61(12):1360–1369.

Zeanah CH, Chesher T, Boris NW; the AACAP Work Group on Quality Issues. Practice parameters for the assessment and treatment of children and adolescents with reactive attachment disorder of infancy and early childhood and disinhibited social engagement disorder. *J Am Acad Child Adolesc Psychiatry*. 2016;55(11):990–1003.

Zeanah CH, Guyon-Harris KL. Commentary: increasing diagnostic precision in reactive attachment disorder—a commentary on Allen & Schuengel (2019). *Child Adolesc Mental Health*. 2020;25(1):11–12.
Zeanah CH, Humphreys KL, Fox NA, Nelson CA. Alternatives for abandoned children: insights from the Bucharest early intervention project. *Curr Opin Psychol*. 2017;15:182–188.
Zimmermann P, Soares I. Recent contributions for understanding inhibited reactive attachment disorder. *Attach Hum Dev*. 2019;21(2):87–94.

49.2 Stereotypic Movement Disorders in Children

Robert Lloyd Doyle, M.A., D.D.S., M.D.

Stereotypic movements are defined as repetitive, sometimes rhythmic, seemingly volitional, but purposeless movements. They encompass a wide range of behaviors, including the tapping of hands or feet and nail biting, and may occur in typically developing children or those with developmental disorders such as spinning behaviors often seen in children with autism spectrum disorders. A stereotype can be as benign as eye blinking or can be quite dangerous, as in the case of intense head banging resulting in traumatic brain injury. However, the term “stereotypy” may be applied to many types of behaviors, including normal repetitive behaviors of infancy, abnormal repetitive behaviors in autistic individuals, dyskinesias caused by psychotropic medications, and symbolic gestures associated with schizophrenia, to name a few, this section focuses on stereotypic movement disorders of childhood as described in the text revision of the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5-TR) and the eleventh edition of the *International Classification of Diseases and Related Health Problems* (ICD-11).

HISTORY

The first mention of purposeless, repetitive behaviors in the clinical literature dates back 300 years to Nehemiah Grew, who wrote, “We see also Mad people … run upon some one action … without variation.” Possibly the first description of a dopamine-driven stereotypy should be attributed to Samuel Gee, who conducted an experiment in which he injected a large dose of apomorphine into a dog. A few minutes after the animal vomited, it began to pace around the room in a “curiously persistent methodical manner.” In the early 1900s, Emil Kraepelin linked stereotypic behaviors to dementia praecox. Of note, Kraepelin’s careful observations of stereotypic behaviors were made before the development of antipsychotic medications, and so the behaviors were likely related to a pathologic brain state rather than to medication-induced dyskinesia. Several seminal papers published in the 1940s set the stage for the following 50 years of research on the subject.

Landmark research on stereotypic behaviors began with Lauretta Bender and Paul Schilder in the early 1940s. Their contribution differentiated impulsions from obsessions and compulsions based on the premise that impulsions were associated with positive affect, whereas obsessions were notably negative. They also speculated that impulsions could have connections to the primitive motor stereotypic behaviors of infancy and that impulsions might become obsessions and compulsions in adolescence. In 1949, Reginald S. Lourie theorized that the feedback from infantile rhythmic behaviors fostered normal motor development. For instance, infants will assume the four-point crawling position on hands and knees, then rock front to back as if rehearsing the more complex coordinated movement in which only the arm and leg on one side of the body moves forward to provide locomotion. This simple repetitive rocking in place could serve as the basis for a repertoire of more-coordinated movements needed to crawl. Thus, according to this theory, repetitive behaviors occur in a normally developing child, but these recede as the child masters more complex, functional behaviors. These behaviors become stereotyped if they persist beyond the developmental purpose they serve, as seen in persons who are blind or have autism or intellectual disabilities,

ETIOLOGY

Several theories have been proposed to explain the emergence and maintenance of stereotypic behaviors. As just described, stereotypic movements may be the precursor to mastery of more sophisticated motor tasks. This path toward more complex motor behaviors may veer in a pathologic direction based on the interplay among abnormal development, behavioral reinforcement, and psychological conflict. In Down syndrome, for instance, stereotypic behaviors are similar to those seen in children without a diagnosis of mental retardation but differ in that in Down syndrome, they appear at a later-than-expected chronologic age and do not disappear. Therefore, developmental factors could contribute to the manifestation of stereotypic behaviors.

Behavioral factors might maintain the stereotypic behavior through positive or negative reinforcement. Careful behavioral analysis can identify social reinforcers and then aim to eliminate them. From a psychodynamic standpoint, stereotypies are considered instinctual, autoerotic, and self-soothing or gratifying behaviors. The psychodynamic factor implies that a stereotypic behavior, such as nail biting, might be a reaction to intense feelings about one or more conflicts. Biologically based theories have replaced most psychodynamic formulations explaining stereotypic behaviors.

Early observations in animals and humans indicated that environmental factors might modify the expression of stereotypic behaviors. For example, R. J. Gallagher and Gershon Berkson showed that even short periods of normal social interaction reduced the expression of stereotypic behaviors in institutionalized infants with profound intellectual disabilities. Most studies show that substituting alternative activities for stereotypic behaviors effectively reduces their expression.

Even though enriched environments usually diminish stereotypic behaviors they do not eliminate them. The predictable persistence of stereotyped behaviors in childhood autism and other conditions, despite environmental enhancement, indicates that organic factors might play a role. One organic theory proposes that endogenous opioid compounds are released in persons with autism as they engage in stereotypic behaviors. The pleasurable *effect* of the endorphin *release* could reinforce the repetitive maladaptive behaviors. Although studies using the opioid antagonists naloxone (Narcan) and naltrexone (Trexan) resulted in modest improvements in stereotypic behaviors, these have not been a definitive cure for stereotypic behaviors in individuals with autism. Therefore, an organic-based explanation is not entirely satisfactory either. Perhaps an interplay between experiential effects on the brain and physiologic processes within the brain might best account for stereotypic behaviors. Bender and Schilder theorized that a common connection must tie all sorts of stereotypy together; however, studies fail to support the existence of an encompassing stereotypic factor.

Nevertheless, research has pointed to several factors that might act independently or in concert to bring about the expression of a stereotypic movement. These include genetic factors, as seen in nail biting and Lesch–Nyhan syndrome, and neurobiologic factors, as evidenced by the many compounds that can precipitate stereotypic or self-injurious behaviors (SIBs). For example, dopamine supersensitivity due to chronically low levels of dopamine has been described in animal models of SIBs. Of interest, postmortem tissue from patients with Lesch–Nyhan syndrome appears to be dopamine deficient as well. Furthermore, dopamine antagonists have successfully treated nonsyndromal SIBs and at least one case of Lesch–Nyhan syndrome. Dopamine has also been implicated in the emergence or exacerbation of certain repetitive behaviors related to stimulant use. Psychostimulants work by blocking the reuptake of dopamine, thereby causing a relative increase in the availability of dopamine in the synaptic cleft. Stimulant-induced repetitive behaviors include bruxism, skin picking, lip biting, compulsions, nail biting, perseveration, tics, and various other stereotypes. Other neurotransmitters, such as serotonin, opioids, and γ-aminobutyric acid, may be involved as well.

EPIDEMIOLOGY

Many stereotypic behaviors emerge in early childhood and are considered part of normal development. These rhythmic, repetitive behaviors appear and progress in a cephalocaudal direction and can be correlated with chronologic age. An estimated 80% of normal children display rhythmic activities that disappear by the age of 4 years. Although determining the prevalence rate in the general population is difficult, studies estimate that some 3% to 15% of children without disabilities engage in head banging. About 20% of children exhibit repetitive body rocking. If these behaviors continue beyond the normal developmental age range for such behaviors, they might evolve into a stereotypic movement disorder. Consider thumb sucking. Endearing in infants, thumb sucking easily becomes a reason for teasing in the school-age child. In an extreme case, an adolescent boy with profound developmental and intellectual disabilities sucked his thumb until a large portion of soft tissue eroded away and exposed the underlying bone. Although rarely to such a degree, finger, and thumb sucking occurs more often in girls than boys and more in whites than in African Americans.

Bruxism—the habitual grinding of teeth—occurs in approximately 56% of normally developing infants. Even if chrome bruxism wears down enamel, this rarely results in long-term problems in young children because deciduous teeth will be replaced by the permanent dentition. Bruxism in adolescents and adults can cause major dental pathology ranging from periodontal disease to temporomandibular joint dysfunction.

Sean was a burly but affable dental student with severe bruxism. He had no personal or family history of psychiatric illness. He was the eldest child of a large, upper–socioeconomic status family. Sean had always maintained a respectable grade point average, and he continued to do so in dental school. By age 23 years, Sean had worn the cusps of his posterior teeth completely away due to his chronic bruxism, which began in childhood. Although bruxism is often caused by malocclusion between maxillary and mandibular dentitions, this could not explain his persistent problem. Sean's posterior teeth had been worn down to flat occlusal surfaces without any cusps to cause interferences in his bite. His case was so impressive that his professors requested permission to take photographs and make models of his teeth to teach future generations of dental students. Ironically, his bruxism worsened as the stress of dental school caused him to chronically clench his teeth together. His dental health was compromised by broad areas of exposed dentin in places once covered by enamel. Sean was at risk of losing his molar teeth, as well as developing painful temporomandibular joint dysfunction.

The treatment took a multimodal approach. Sean was advised to monitor himself for clenching and grinding during the day. Whenever he became aware of these behaviors, he was instructed to simply separate his upper and lower teeth. To prevent neuromuscular programming, his dentist recommended that he stop chewing gum. These interventions helped during the day; however, nocturnal bruxing continued. Sean was not aware of grinding his teeth in his sleep, but he did notice tenderness around his jaw in the morning. On exam, his dentist noted crepitus—the sound of bone rubbing against bone—with mandibular movement. Because crepitus is a sentinel sign of temporomandibular joint dysfunction, the dentist fitted Sean with a custom night guard that served two purposes: it prevented further eroding of enamel, and it curtailed the chronic temporomandibular joint inflammation. Sean also took ibuprofen to relieve inflammation and pain. His dentist offered a prescription for clonazepam (Klonopin) to help hint deal with the pressures of his professional education. Sean, however, did not want to become addicted to a benzodiazepine, so he chose to try other measures first. Buspirone (BuSpar) was also offered as an option because of its reduced abuse potential. Fortunately, his dental condition stabilized fairly quickly, and he set his sights on having his left tooth structure restored with ceramic crowns.

Nail biting occurs quite commonly in childhood and frequently continues into adulthood. Occasionally, nail biting progresses to the point of causing dermatologic or even dental problems. It does not seem to be associated with the child’s age, level of intelligence, or socioeconomic status, but nail biting seems to have a familial basis. Toenail biting, however, does appear to be much more prevalent in girls. An estimated 50% of children exhibit nail biting. It can be seen as early as 4 years of age, with peak prevalence between ages 10 and 18 years. According to one study, 4.5% of adults continue to bite their nails. Adults with onychophagia may not have comorbid psychiatric illness, and children with nail biting do not appear to be more anxious than peers who do not bite their nails. In most adult cases, nail biting seems to cause distress rather than result from stress.

As mentioned earlier, genetics seems to play a factor in the expression of some stereotypic behaviors. In twin studies, children born to parents who were nail biters were nearly three times as likely to be nail biters as control subjects. The concordance rate for monozygotic twins was 66% versus 34% in dizygotic twins. Onychophagia is the most severe type of nail biting. Monozygotic twins with onychophagia had a 75% concordance rate, whereas dizygotic twins only had an 18% concordance.

Of interest, stereotyped behaviors often occur in blind children. Elise Fazzi and others studied a group of congenitally blind children. In their group of 11 boys and 15 girls with and without developmental disabilities, 73% showed stereotypic behaviors. The most frequently encountered stereotypic behaviors were body rocking (30.7%), repetitive handling of objects (30.7%), and eye pressing along with eye poking (30.7%). These researchers also found that the stereotypic behaviors could be decreased by encouraging appropriate adaptive behaviors and increased by restricting the environment, reducing sensory stimulation, and limiting mobility.

Stereotypic behaviors are more common among children with a history of parental neglect, but they may be present in children who do not have emotional problems. An estimated 20% of individuals with severe intellectual disabilities exhibit stereotypic behaviors. Alfred A. Baumeister and Rex L. Forehand found that two-thirds of a population institutionalized with intellectual disabilities exhibited stereotyped behaviors, and J. R. Dura and colleagues later reported that 34% of nonambulatory individuals with profound retardation who were institutionalized exhibited at least one stereotypic behavior. Despite the need for longitudinal studies, evidence from cross-sectional studies suggests that developmental age and experiential history, rather than chronologic age, are the relevant mediating variables for stereotypic movements. Stereotypic movements frequently peak in adolescence and then diminish thereafter.

The prevalence of SIBs varies from between 2% and 3% for children and adolescents in the community to about 10% to 15% for adults with severe or profound intellectual delays living in residential facilities. In individuals with severe intellectual disabilities living in institutions, stereotyped movements increase with chronologic age; however, the opposite is true of children without disabilities. In individuals with intellectual disabilities, the most common stereotypic movements are body rocking, head rolling, repetitive hand movements, self-hitting, head banging, and self-biting. The prevalence of head banging ranges between 3% and 15% in normally developing children. Approximately three times as many boys as girls exhibit the behavior. Head banging usually emerges between 5 and 11 months of age, but only 5% of these cases persist beyond 3 years of age.

Per the DSM-5, stereotypic movements may occur with or without SIBs. A study by Gal and colleagues monitored SIB and stereotypy in typically developing children as well as those who were intellectually impaired, visually impaired, hearing-impaired, or autistic. In this sample of 90 subjects with at least one SIB, only 3 did not exhibit a co-occurring stereotypy. Interestingly, the rank order of prevalence of SIB and stereotypy was same across all five groups.

Given that the strongest predictor for the presence antiseverity of SIB was other repetitive behavior, an earlier model proposed by Guess and Carr in which SIB was thought to evolve from stereotyped movements makes some sense. SIBS, such as head banging, and stereotypy occur in typically developing children, but the SIB emerges later in development than stereotypy. In individuals with intellectual disabilities, SIB and repetitive behaviors appear temporally related; however, stereotypic behaviors still precede SIB.

Of note, Cornelia de Lange syndrome (mutations including the NIPBL gene in region 5p13, SMC1L1 in the Xp11.2 region, SMC3 gene in region 10q125), Fragile X syndrome (mutation at Xq127.3), Smith–Magenis (3.7 Mb interstitial deletion in chromosome 17p11.2), Cri-du-Chat syndrome (deletion in chromosome 5p15), and Rett syndrome (mutation in the methyl CpG—binding protein 2) are genetic disorders that present with various forms of stereotypic movements and SIB. Since these genetic defects are quite diverse, yet each gives rise to stereotypies with and without self-injury in the same individual, finding treatments for stereotypic behaviors remains a challenge.

Nevertheless, theories for stereotypic movements abound, and research continues to reveal new information regarding the etiology of these behaviors. An array of new research techniques will allow genetic information to be linked with neuroimaging results and facilitate understanding of the phenotypic expression of stereotypic behaviors.

NOSOLOGY

In many cases, the origin of stereotypic movements seems contextual to certain psychiatric illnesses or medical conditions in which they occur. The evolution of the DSM has tried to reflect such insights in its various iterations. Recent versions of the DSM had reserved the diagnosis of stereotypic movement disorder for cases in which the stereotypic behavior is not a core symptom of another psychiatric disorder or the expected manifestation of a medical condition. Therefore, the repetitive facial movements due to either prolonged neuroleptic use or rapid discontinuation of this class of medication would be described as stereotypic movements, but these would not have met DSM-IV-TR criteria for stereotypic movement disorder. Although this continues to be the case in DSM-5, the current criteria require the clinician to specify whether stereotypic movement is associated with a known medical or genetic condition, neurodevelopmental disorder, or environmental factor. For instance, a person with Down syndrome and stereotypic movements would be coded as (307.3) stereotypic movement disorder associated with Down syndrome. In the case of autism spectrum disorder, a comorbid diagnosis of stereotypic movement disorder should be made whenever the stereotypic behaviors cause injury or appear sufficiently severe to become a focus of treatment.

Also, the DSM-5 has placed more emphasis on culture-related issues. Although stereotypic movements, with or without self-injury, manifest in all races and cultures, the attitudes of a particular culture may delay diagnosis. Clinicians need to be aware of cultural context and remain sensitive as they explain the diagnosis to patients and their families who come from a different country or culture. This will help patients accept and maintain appropriate treatment for the condition.

At first, the DSM criteria were somewhat broad. The DSM-III provided criteria for a diagnosis of atypical stereotypic movement disorder for conditions such as rocking, head banging, and repetitive hand movements characterized by quick, rhythmic hand rotations or tentative voluntary movements that usually involved the arms or fingers. The DSM-11 stipulated that such movements differed from tics in that they were voluntary but not spasmodic or distressing. In the DSM-III, atypical stereotypic behaviors were considered often to occur in patients with extreme psychosocial deprivation, intellectual disabilities, and pervasive developmental disorders; however, another mental disorder did not have to be present.

In the revised version of the DSM-III (DSM-III-R), atypical stereotypic movement disorder was renamed stereotypy/habit disorder. The criteria for stereotypy/habit disorder became broader so that intentional, repetitive, nonfunctional behaviors such as nail biting, nose picking, skin picking, and object mouthing were included. The DSM-III-R also specified a severity criterion, which required the behavior to cause injury or markedly interfere with normal activities to be considered a clinically significant "disorder." The severity criterion allowed clinicians to determine whether the stereotypic movement was part of normal development, as is often seen with transient nose picking or a more pathologic picture, in which chronic nose picking causes significant tissue trauma or social stigmatization. A stereotypy or habit described by DSM-III-R criteria could include more intense forms of normal motor patterns seen in childhood, or they might be associated with some type of psychiatric or organic pathology. An additional criterion, however, excluded the diagnosis if the patient met the criteria for pervasive developmental disorder or tic disorder. This exclusion criterion recognized and eliminated the redundancy of making the diagnosis of stereotypy/habit disorder in these disorders, which have repetitive behaviors as core symptoms.

Another name change occurred with the publication of the DSM-IV. In an effort to be compatible with the ICD-10, the DSM-IV renamed the disorder stereotypic movement disorder. Because the intentional nature of a stereotypic movement might be impossible to determine in certain cases—for example, individuals with severe intellectual disabilities with limited ability to communicate, the clinician is no longer obligated to prove that the movement was intentional. The DSM-IV also drew boundaries between stereotypic movement disorders and several other disorders, including pervasive developmental disorder, tic disorder, obsessive-compulsive disorder (OCD), trichotillomania, and intellectual disabilities. As in the DSM-III-R, the diagnosis is not made if the stereotypic behavior is attributable to pervasive developmental disorder or tic disorder. Given the similarity between tics and compulsions, the DSM-IV excluded stereotypic behaviors that were compulsions related to OCD. The DSM-IV also stipulated that the stereotypic behavior could not be due to the physiologic effects of a substance or a general medical condition. A diagnosis of stereotypic movement disorder would not be made for repetitive hair pulling if DSM-IV criteria for trichotillomania were met. In addition, a diagnosis of stereotypic movement disorder with SIB could be made in persons with intellectual disabilities if the SIB was significant enough to become a focus of therapy. Subsequent versions of the DSM—the DSM-IV-R and the DSM-IV-TR—used those criteria established by the DSM-IV.

In general, the DSM-IV-TR and DSM-5 as well as ICD-10 criteria for stereotypic movement disorder describe the diverse and unrelated stereotyped movements that are not part of any other recognized neurologic or psychiatric condition. According to the DSM-5, a stereotypic movement disorder is a repetitive, seemingly driven, and nonfunctional motor behavior. It lists hand shaking or waving, body rocking, head banging, mouthing of objects, self-biting, picking at skin or body orifices, and hitting one's body as examples. The DSM-5 definitions exclude behaviors with a clear purpose or adaptive function, but, as acknowledged in the DSM-IV-TR, the driving force behind a stereotyped movement may be difficult to determine in cases in which the person has limited ability to communicate internal states. What is more, a stereotypic behavior may provide a signal that an undiagnosed neurodevelopmental disorder exists, especially in children younger than 3 years of age.

The DSM-5 specifies that for a repetitive motor behavior to be considered a stereotypic movement disorder, it must interfere with social, academic, or other activities and may result in self-injury. In severe cases, the behavior can cause death, but even less severe cases might result in disabling or disfiguring impairments.

Consider the case of Ruth, a middle-aged woman, who was admitted to the inpatient unit at a university-based teaching hospital due to unrelenting head scratching. Her outpatient psychiatrist had tried the full array of pharmacologic interventions, including combinations of anxiolytic, antidepressant, and antipsychotic medications, but the stereotypic scratching progressed to the point that she required surgical intervention. Ruth had scratched through her scalp and penetrated the skull, exposing the dura mater.

In Ruth's case, a complete DSM-5 diagnosis would be stereotypic movement disorder with self-injurious behavior, severe. Per the DSM-5, the diagnosis is specified as with or without SIB. Note, however, that the disorder should be specified as "with self-injurious behaviors," if injury would occur in cases where protective measures were not present to prevent injury. Ruth obviously displayed severe symptoms, given that continuous monitoring and protective measures were required to prevent serious injury. Moderate symptoms require protective measures and behavioral modification; whereas, mild symptoms can be easily suppressed with sensory stimulus or distraction. Clinicians should also specify whether the stereotypic movement is associated with a known medical or genetic condition, neurodevelopmental disorder, or environmental factor. An additional code identifying the associated medical/genetic condition or neurodevelopmental disorder aids in billing and communicating with third-party payers.

The ICD-10 diagnostic criteria for stereotypic movement disorders appear more succinct. Per the ICD-10, the stereotyped movements must either cause physical injury or markedly interfere with normal activities. The duration of symptoms lasts at least 1 month. In contrast to DSM-5, no other mental or behavioral disorder can be present, except for intellectual disabilities.

DIFFERENTIAL DIAGNOSIS

The outward appearance of a complex tic or a compulsion may closely resemble stereotypic movements, and this may lead to diagnostic difficulties. James W. Bodfish, T. W. Crawford, and colleagues found high rates of stereotyped behavior in adults with comorbid mental retardation and compulsion disorder. Compared with a noncompulsive matched control group, their sample showed an increased prevalence of stereotypy (32%), an increased prevalence of self-injury (40%), and an increased prevalence of stereotypy plus self-injury (37%) in those with compulsive disorder. A review by J. L. Rappoport reported that 20% of patients with OCD also exhibit motor tics, which complicates the diagnosis further. Nevertheless, some simple caveats can help to sort out stereotypic movements from other repetitive movements:

Generally speaking, a stereotypic movement is a rhythmic, repetitive motor behavior. It usually is simpler than other forms of repetitive behaviors. Although a stereotypic behavior appears purposeless, it can also serve to alleviate anxiety or another aversive state. Tics ordinarily have an involuntary quality even though tics may be suppressed for limited periods of time and they are not rhythmic. Compulsions appear more complex and ritualized. A stereotypic movement differs from a complex tic and a repeating compulsion in that it is not performed as a preventative measure, nor is it motivated by a sense of incompletion. In addition, complex tics and compulsive behaviors typically are not rhythmic.

With regard to autism, a study by Fiona Zandt and colleagues suggested that compulsions coexist with stereotypic behaviors in children with autism spectrum disorder. They found that children with OCD had more obsessions and compulsions than children with autism spectrum disorder; however, both the OCD group and the group with autistic stereotyped behaviors exhibited more obsessions and compulsions than a comparison group with a typical developmental trajectory. Similarly, J. M. Ringman and J. Jankovic reported on a series of 12 patients with autism spectrum disorders and stereotypic movements. These patients were referred to a specialized movement disorder clinic for the evaluation of tics. Tics were clearly present in seven of the patients, and six of these patients met the criteria for both Asperger syndrome and Tourette syndrome. Even though the diagnosis of stereotypic movement disorder excludes repetitive movements that may be core symptoms of other DSM-5 disorders, the literature suggests, and this author observes, that stereo behaviors in autistic spectrum disorders may be comorbid with OCD or tic disorder.

The DSM-IV-TR specifies that the stereotypic behavior is not due to the direct physiologic effects of a substance or a general medical condition. This criterion would exclude tardive dyskinesias, as in chronic neuroleptic exposure or skin picking secondary to stimulant use. Although tremors associated with Parkinson syndrome and Parkinson disease are abnormal and stereotypic, neither would fit the current criteria for stereotypic movement disorder because they are due to the physiologic effects of neuroleptic medication in Parkinson syndrome and a medical condition in the case of Parkinson disease.

Intellectual disability is a common feature of numerous genetic disorders. If the stereotypic or SIB occurs in the context of intellectual disability, the behavior no longer must be severe enough to become a focus of the treatment to meet the diagnostic criteria for DSM-5. Repetitive motor behaviors noted in a person with intellectual disabilities would be correctly recorded per the DSM-5 as stereotypic movement disorder associated with intellectual developmental disability. In Lesch–Nyhan syndrome, the degree of intellectual disability does not seem to predict the magnitude of self-mutilation. Some patients with mild mental intellectual disability may display extreme forms of self-injury. Notably, stereotypic movements occur in approximately two-thirds of individuals with intellectual disabilities living in residential facilities.

In addition, the DSM-5 criteria no longer require that the repetitive movement must be present for 4 weeks or more. Typically, stereotypic movements peak in adolescence, but they have no typical age of onset. Stereotypic movements in persons with severe or profound intellectual disabilities may last for years; however, an abrupt onset of stereotypic behaviors may signal the presence of an underlying acute medical condition in nonverbal individuals with severe mental retardation. Clinicians should be mindful of clues that such individuals may be experiencing pain, distress, anxiety, or mood shifts and then promptly treat any identified problems. Stressful events also may precipitate stereotypic behaviors. However, stereotypic movements may emerge and become a focus of treatment in individuals with mental retardation who are institutionalized in nonstimulating environments.

ICD-10 AND 11

The ICD-10 and ICD-11 define stereotypic movements in childhood in a fashion similar to the DSM-5 since one of the goals of DSM-5 aimed to provide compatibility between the two diagnostic systems. As in the DSM, physical injury or marked impairment in normal activities must be present for the diagnosis. This criterion attempts to sort out stereotypic movement disorder from stereotypic movements that are benign or typical of normal development. The duration criterion provides another dimension to the impairment and ensures that the stereotypic movement is more than a passing novelty. In addition, a history of another psychiatric or behavioral disorder should not be present and account for the stereotypic movement.

DIAGNOSIS

No clinical tests are available to make the diagnosis of stereotypic movement disorder. Nevertheless, the history and physical exam may be revealing. Dental and physical exams may provide evidence of tissue trauma that would raise suspicions for stereotypic behaviors, especially the self-injurious type. Anemia may be a sign of self-induced rectal bleeding. The delicate architecture of the eye is particularly vulnerable to repetitive trauma from behaviors such as head banging and eye poking. Sometimes a subtle physical sign clues the clinician to stereotypic behaviors and movements that go unobserved by others.

Rating scales assist in diagnosing and tracking responses to treatment in many psychiatric conditions; however, rating scales for stereotypic behaviors have not been so helpful. Several rating scales have been developed, but none provides optimal psychometric properties across the spectrum of stereotypic behaviors. Even commonly used instruments for tracking drug-induced stereotypic behaviors, such as the Angus Involuntary Movement Scale (AIMS) and the Barnes Scale, focus on involuntary movements rather than the wide range seen with stereotypic movements. Berkson and Davenport developed a stereotypy checklist to track stereotyped behaviors mainly due to social deprivation. Unfortunately, experiments intended to test the reliability and validity of the Berkson and Davenport Stereotypy Checklist have shown variable results.

Recently, video- and computer-assisted technologies have been used to study tardive dyskinesia. These technologies may elucidate the characteristics of other stereotypic behaviors, such as that developed for use in monitoring attention-deficit/hyperactivity disorder (ADHD). These motion-detecting devices connected to sophisticated computer programs can give detailed descriptions of the randomness, as well as the range of movements on a second-by-second basis. This type of technology might prove valuable in research or for clinical applications related to stereotypic behaviors.

Eventually, research may lead to feasible diagnostic tests using genetic and/or neuroimaging techniques to assist in the diagnosis and treatment of stereotypic behaviors. In the meantime, a comprehensive clinical exam remains the standard for making an accurate diagnosis.

INJURIOUS BEHAVIORS

Stereotypic movements can manifest in a variety of injurious behaviors that put patients and caregivers at risk. The ramifications of head banging range from hematomas and head injuries to retinal detachment and blindness. However, even seemingly benign stereotypic behaviors could cause cumulative damage over time. For example, hair loss could be a consequence of chronic head rolling. In addition, parents and caregivers are at risk of accidental bites and bruises as they attempt to protect children and many adults who engage in stereotyped self-injury. Even children who typically do not exhibit self-injury can be hazardous to others. A child being held may surprise a parent or caregiver with spontaneous swift lead bobbing that results in a broken nose or blackened eye. Stereotypic movements, with or without the potential for self-injury, may cause major management problems, which hamper habilitation and stifle social adaptation.

Onychotillomania refers to compulsive picking or tearing at the nails. It can produce ridges and a central depression on the nail from chronic rubbing by the nail or another finger. Most commonly, the thumbnail of the dominant hand produces trauma to the nail plate of the second or third finger. Onycholysis refers to the separation of the nail from the nail bed. Onychotillomania and onycholysis may be manifestations of rare conditions involving self-mutilation like Lesch–Nyhan syndrome and congenital insensitivity to pain.

Appropriate attention to the patient’s history will help to sort out other stereotypic or SIBs. For example, a stereotypic behavior may have been present in a patient with schizophrenia prior to any neuroleptic medication exposure. A careful history of symptom onset can prevent a misdiagnosis of medication-induced stereotypy. Secondary gain in fictitious disorders might be the driving force in some instances. In other cases, the inability to control anger or anxiety may lead to repetitive cutting, as seen with patients with borderline personality disorder.

TREATMENT

A variety of treatment options, from psychotherapy to pharmacotherapy, have been used to treat stereotypic and SIBs. Even surgery and electroconvulsive treatments have been used in refractory cases. Behavioral treatments have been used with some success, and these include positive reward techniques such as differential reinforcement of behaviors other than the targeted maladaptive behavior (Differential Reinforcement of Other Behavior [DRO]) and differential reinforcement of behavior incompatible with the targeted maladaptive behavior (Differential Reinforcement of Incompatible Behavior [DRI]). Rewards may be immediate or delayed, as in token economies, but they must be individualized so that they produce the maximum motivation. Rewards also may lose their novelty over time; therefore, changing rewards may provide a fresh incentive for motivation. Punishment can range from misting water into the face to applying mild electric shocks to the patient. Although these techniques might seem mean-spirited, they should be employed if self-injurious or stereotypic behaviors have a high potential for severe harm. Of course, the clinician should document a thoughtful risk versus benefits analysis and discuss this with the guardian or family before using such methods. Other forms of punishment include positive practice, removal of reinforcement, and restitution. A study by Jonathan M. Miller and others used behavioral modification techniques of habit reversal and differential reinforcement of other behaviors in an attempt to suppress stereotypic behaviors in 12 nonautistic children between the ages of 6 and 14 years. After a mean follow-up time of 12.1 months, motor stereotypies were significantly reduced on the Stereotypy Linear Analog Scale and the Stereotypy Severity Scale. These results correlated with the subject's motivation and the number of sessions attended. The Child Global Assessment Scale also improved, but no correlations were seen with the level of motivation or number of sessions.

Pharmacologic approaches have been a mainstay treatment of stereotypic behaviors. An estimated 50% of persons with intellectual disabilities living in residential or institutional settings are prescribed high doses of antipsychotic medication, but this statistic does not give an accurate account of the percentage of these patients who have a known psychotic illness. About 75% of patients receiving psychotropic medications develop some form of motor dysfunction. In the early days of treatment, movement disorders occurred at even higher rates due to the misconception that motor dysfunction was the definitive sign that a patient had achieved an adequate dose of medication to treat psychosis. Even though antipsychotic agents are used to treat stereotypic movements, they also can cause repetitive, purposeless movements. For instance, akathisia secondary to neuroleptic medication may appear clinically as shifting from foot to foot, crossing, and uncrossing of legs, pacing, hand wringing, or rocking, among other manifestations.

The newer generation of atypical antipsychotic medications does not pose the typical risk for motor dysfunction seen in older drugs. To date, risperidone (Risperdal) and aripiprazole (Abilify) have been approved by the U.S. Food and Drug Administration (FDA) for self-injurious and disruptive behaviors in children with autism. For the double-blinded, placebo-controlled clinical trial of risperidone, the dose ranged from 0.5 to 3.5 mg per day in the 101 children studied. In a placebo-controlled study of chlorpromazine (Thorazine) used in doses ranging from 2 to 3 mg/kg per day, four adolescents and young adults with profound intellectual disabilities exhibited a 35% reduction in stereotypic behaviors. Unfortunately, the improvement in stereotypic behaviors was offset by a cognitive decline. Several double-blinded placebo-controlled studies used haloperidol (Haldol) in daily doses of between 0.5 and 4 mg to treat stereotypic behaviors in youth with autism. In these studies, haloperidol significantly reduced stereotypic behaviors without causing notable cognitive impairment.

With FDA approval of risperidone and aripiprazole for the treatment of disruptive behaviors, including stereotypic SIBs, these medications have been used more frequently in individuals with autism. This increase in these and other neuroleptic medications in the autism population may lead to other types of repetitive movement disorders, such as Parkinson syndrome with a “pill-rolling” tremor or tardive dyskinesia. Anticholinergic agents such as benztropine (Cogentin), diphenhydramine (Benadryl), and others effectively counteract extrapyramidal movements associated with neuroleptic use. Until recently, the only way to address tardive dyskinesia was to immediately stop or taper off the neuroleptic and hope that it would resolve over the course of weeks to months. In many cases, tardive dyskinesia would decrease to a degree but not fully resolve. Although not life-threatening, the prospect of social stigma from permanent tardive dyskinesia had been one of the main limiting factors in the use of this class of medications. In 2017 however, the FDA approved valbenazine and deutetrabenazine for the treatment of tardive dyskinesia. These vesicular monoamine transporter type 2 (VMAT 2) inhibitors are not approved in Canada, but tetrabenazine is. Tetrabenazine was approved by the U. S. FDA in 2008 for chorea movements in Huntington disease. In the randomized 6-week trials of valbenazine, Abnormal Involuntary Movement Scale (AIMS) mean scores were reduced (−3.2) on 80 mg/day, (−1.9) on 40 mg/day, and (−0.1) on placebo at week six compared to baseline. For the 12-week trials of deutetrabenazine, AIMS score reductions with a mean daily dose of 39 mg/day was (−3.0) compared to (−1.6) with placebo in the smaller study of 117 subjects. In a larger trial of this drug, AIMS scores were reduced (−3.3) on 36 mg/day and (−3.2) on 24 mg/day compared to (−1.4) on placebo. The 12 mg/day dose did not show improvement compared to the placebo. Although FDA-approved, these medications can reduce yet not necessarily eliminate symptoms of tardive dyskinesia. Thus, judicious use of antipsychotic medications continues to be the best approach for individuals with autism and stereotypic self-injurious, disruptive behaviors. Otherwise, attempts to eliminate stereotypic movements could lead to the emergence of other types of movement disorders.

A variety of other agents have been tried, but these have fallen short due to limitations of study size or design. Although lithium carbonate (Eskalith) has been studied for SIBs in patients with intellectual disabilities, the results of the few studies are inconsistent. Fenfluramine (Ponderax, Pondimin) reduced stereotypic behaviors and other problem behaviors in individuals with autism; however, the results also varied across studies. In addition, fenfluramine has been associated with cardiovascular and pulmonary problems. Several other serotonergic agents have been tried, including fluoxetine (Prozac), sertraline (Zoloft), citalopram (Celexa), trazodone (Desyrel), and buspirone (Buspar), but methodologic issues related to the size or design of each study limited their usefulness.

Clomipramine (Anafranil) and desipramine (Norpramin) have been studied for their effect on stereotypic behaviors in autism and onychophagia. Gordon and coworkers showed that clomipramine was superior to desipramine, but both were better than placebo in reducing stereotypic behaviors associated with autism. Clomipramine was better than desipramine in treating severe onychophagia, but the dropout rate was greater than that seen in other studies of its use in OCD and trichotillomania. Henrietta L. Leonard and colleagues concluded that clomipramine was not the optimal treatment for severe onychophagia.

An endogenous opioid theory for SIBs has prompted the use of opioid antagonists. The original study of naloxone (Narcan) successfully reduced SIBs in two young adults with profound intellectual disabilities. SIBs with low frequency of occurrence responded best to the intervention at low doses. Subsequently, several other studies showed that individuals responded variably to naloxone or naltrexone. However, these two agents may decrease stereotypic, self-injurious, and hyperactive behaviors in some patients.

Clearly, more controlled studies are needed to evaluate these and other therapies based on behavioral and pharmacologic interventions. Recent technologic advances in genetic analysis, neuroimaging, and movement monitoring will promote this goal.

Although stereotypic behaviors manifest in many ways, diagnostic criteria for a stereotypic movement disorder in the DSM-5-TR and the ICD-11 describe repetitive, rhythmic movements that are not due to another psychiatric disorder, medical condition, or medication effect. Criteria for impairment help to clarify whether a stereotypic movement is an inconsequential habit or part of normal development. The careful clinician should use a thorough history and exam to identify the presence and possible causes of stereotypic behaviors. Once the clinician accurately characterizes the stereotypic movement, behavioral therapy with or without medication intervention may be applied. Although the literature on the treatment of stereotypic behaviors is limited, it does provide a starting point for developing treatment plans that will likely be subject to change as warranted.

FURTHER READINGS

Anderson KE, Stamler D, Davis MD, et al. Deutetrabenazine for the treatment of involuntary movements in patients with tardive dyskinesia (AIM-TD): a double-blind, randomized, placebo-controlled, phase 3 trial. *Lancet Psychiatry*. 2017;4:595–604.

Bhidayasiri R, Jitkritsadakul O, Friedman JH, Fahn S. Updating the recommendations for treatment of tardive syndromes: a systematic review of new evidence and practical treatment algorithm. *J Neurol Sci*. 2018;389:67.

Fernandez HH, Factor SA, Hauser RA, et al. Randomized controlled trial of deutetrabenazine for tardive dyskinesia: the ARM-TD study. *Neurology*. 2017;88:2003–2010.

Gal E, Dyck MJ, Passmore A. The relationship between stereotyped movements and self-injurious behavior in children with developmental or sensory disabilities. *Res Del-Disabil*. 2009;30:342–352.

Gedye A. Anatomy of self-injurious, stereotypic, and aggressive movements: evidence for involuntary explanation. *Clin Psvchol*. 1992;48:76–778.

Guess D, Carr E. Emergence and maintenance of stereotypy and self-injury. *Am J Mem Retard*. 1991;96:299–319.

Hauser RA, Factor SA, Marder SR, et al. KINECT 3: a phase 3 randomized, double-blind, placebo-controlled trial of valbenazine for tardive dyskinesia. *Am J Psychiatry*. 2017;174:476–484.

Hewlett WA. Obsessive-compulsive disorder. In: Ebert MR, Loosen PT, Nurcombe B, eds. *Current Diagnosis and Treatment in Psychiatry NY: Lange Medical Books*. McGraw-Hill; 2000–2351.

Kolmen BK, Feldman HM, Harden BL, Ianosky JE. Naltrexone in young autistic children: replication study and ‘learning measures. *J Am Acad Child Adolesc Psychiatty*. 1997;36:1570–1578.

Lewis MH, Baumeister AA. Stereotyped mannerisms in mentally retarded persons: animal models and theoretical analyses. In: Ellis NR, ed. *International Review at Research on Mental Retardation*. Academic Press; 1982.

Lewis MI-I. Abnormal repetitive behaviours: shared phenomenology and physiology. *J Intellect Disabil Res*. 2012;5(56):427–440.

McCracken JT, McGough J, Shah B, et al. Research Units on Pediatric Psychopharmacology Autism Network. Risperidone in children with autism and serious behavioral problems. *N Engl J Med*. 2002;347(5):314–321.

Miller JM, Singer HS, Bridges DD, Waranch HR. Behavioral therapy for treatment of stereotypic movements in nonautistic children. *J Child Neurol*. 2006;21(2):119–125.

Petty J, Allen D, Oliver C. Relationship among challenging, repetitive, and communicative behaviors in children with severe intellectual disabilities. *Am J Intellect Dev Disabil*. 2009;114:356–368.

Posey D, McDougle CI. The pharmacotherapy of target symptoms associated with autistic disorder and other pervasive developmental disorders. *Harv Rev Psychiatry*. 2000;8(2):45–63.

Raley JJ. *Mental Retardation: Developing Pharmacotherapies*. American Psychiatric Press; 1991.

Richman DM, Lindauer SE. Longitudinal assessment of stereotypic. proto-injurious, and self-injurious behavior exhibited by young children with developmental delays. *Am J Menu Retard*. 2005;110:439–450.

Rifleman JM, Jankovic J. Occurrence of tics in asperger’s syndrome and autistic disorder. *J Child Neurol*. 2000;15(6):394–400.

Sajatovic M, Alexopoulos GS, Burke J, Farahmand K, Siegert S. The effects of valbenazine on tardive dyskinesia in older and younger patients. *Int J Geriatr Psychiatry*. 2020;35:69–79.

Sallustro F, Atwell CW. Body rocking, head banging and head rolling in normal children. *J Pediatr*. 1978;93:704–708.

Sprague RL, Newell KM. *Stereotyped Movements: Brain and Behavior Relationships*. American Psychological Association; 1996;211.

Tanguay PE. Pervasive developmental disorders: a 10-year review. *J Am Acad Child Adolesc Psychiatry*. 2000;39(9):1079–1095.

Thompson T, Gray DB. *Destructive Behavior in Developmental Disabilities*. Sage; 1994.

Tripp G, Luk SL, Schaughency EA, Singh R. DSM-IV and 1CD-10: a comparison of the correlates of ADIID and hyperkinetic disorder. *J Am Acad Child Adolesc Psychiatry*. 1999;38(2):156–164.

Truster H. Prevalence and functions of stereotyped behaviors in nonhandicapped children in residential *care*. *J Abnorm Child Psycho*. 1994;22:79–97.

Vollonar F. *Autism and Pervasive Developmental Disorders*. Cambridge University Press; 1999.

Winchel RM, Stanley M. Self-injurious behavior: a review of the behavior and biology of self-mutilation. *Am J Psychiatry*. 1991;148:306–317.

Zandt F, Prior M, Kyrios M. Repetitive behaviour in children with high functioning autism and obsessive compulsive disorder. *J Autism Dev Disord*. 2007;37(2):251–259.

50

Mood Disorders in Children and Adolescents

50.1 Depressive Disorders and Suicide

KAREN DINEEN WAGNER, M.D., PH.D., TAMI BENTON, M.D., AND CATHARYN A. TURNER, II, M.D., M.ED.

Depressive disorders in children and adolescents have a significant adverse effect on a youth's emotional, social, and cognitive development. Early onset of major depression is associated with more lifetime depressive episodes and suicide attempts, greater psychiatric and medical comorbidity, and poorer quality of life than later age onset of depression. Disruptive mood dysregulation disorder leads to impairments in school functioning and relationships with family and peers. Recognition of these disorders is essential so that children and adolescents may receive effective evidence-based treatments.

DEFINITION

The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5-TR) includes in the category of depressive disorders the following: disruptive mood dysregulation disorder, major depressive disorder, persistent depressive disorder, premenstrual dysphoric disorder, substance/medication-induced depressive disorder, depressive disorder due to another medical condition, other specified depressive disorder, and unspecified depressive disorder. To be in the category of depressive disorders, there cannot be a history of ever having had a manic or hypomanic episode. There are coded specifiers for the depressive episode for severity (mild, moderate, severe, with psychotic features) and course (in partial remission, in full remission). The uncoded specifiers for the current episode include with anxious distress, with mixed features, with melancholic features, with atypical features, with mood-incongruent psychotic features, with catatonia, with peripartum onset, and with seasonal pattern.

Disruptive mood dysregulation disorder in children is characterized by recurrent verbal or behavioral temper outbursts (>3 or more times per week) that are excessive and age inappropriate. In between these episodes, mood is irritable or angry most of the day, nearly every day. These symptoms occur in at least two settings (school, home, or with peers). The duration of symptoms must be more than 12 months, without a 3-month symptom-free period in this time frame. The age of onset of the symptoms must be before 10 years old, and the diagnosis should not be made before age 6 years or after age 18 years.

Criteria for major depressive disorder in children and adolescents are at least one major depressive episode in which there has been at least a 2-week period of either depressed or irritable mood or a loss of interest or pleasure. In addition, at least four of the following symptoms are required: weight loss, weight gain, failure to make expected weight gain, or an increase or decrease in appetite; insomnia or hypersomnia; psychomotor agitation or retardation; fatigue or loss of energy; feelings of worthlessness or excessive or inappropriate guilt; diminished ability to think or concentrate or indecisiveness; and recurrent thoughts of death, recurrent suicidal ideation, suicide attempt, or suicide plan. The symptoms must cause impairment in functioning or cause clinically significant distress.

Persistent depressive disorder in children and adolescents is a chronically depressed or irritable mood for at least 1 year. In addition, there must be at least two of the following symptoms: poor appetite or overeating; insomnia or hypersomnia; low energy or fatigue; low self-esteem; poor concentration or difficulty making decisions; and feelings of hopelessness. During the 1-year period, the youth also has not been without depressive symptoms for more than 2 months at a time.

Premenstrual dysphoric disorder requires at least five symptoms in the final week before the onset of menses which start to improve a few days after the onset of menses and are either minimal or absent in the week following menses. One or more of the following symptoms must be present: affective lability, irritability/anger/interpersonal conflict, depressed mood/hopelessness/self-deprecating thoughts, or anxiety/tension/keyed up or on edge. In addition, at least one of the following symptoms must be present: decreased interest, concentration difficulties, fatigue/low energy, appetite changes, hypersomnia/insomnia, feeling overwhelmed/out of control, and physical symptoms. The symptoms must be severe enough to interfere with functioning or cause clinically significant distress.

A diagnosis of other specified depressive disorder is used when full criteria for a depressive disorder are not met, but the symptoms cause significant distress or impairment. The specific reason for not meeting the full criteria is identified. The diagnosis of unspecified depressive disorder is used when the reasons for not meeting the full criteria are not identified.

HISTORY

Although depression in children and adolescents is now recognized as a common and serious illness, it was not until the 1980s that studies demonstrated that children do develop depressive disorders that can be diagnosed in a standardized manner. It was originally thought that children and adolescents developed a “masked” form of depression with predominantly behavioral symptoms, but subsequent studies have shown that children and adolescents have symptoms similar to those of adults with depression.

Disruptive mood dysregulation disorder is a new diagnosis in DSM-5 depressive disorders to address issues related to the potential overdiagnosis of bipolar disorder in children and adolescents. Children with severe nonepisodic irritability were increasingly diagnosed with bipolar disorder but did not have distinct episodes of mania or hypomania. Disruptive mood dysregulation disorder had its origins in the construct of severe mood dysregulation developed by researchers (Leibenluft and colleagues) at the National Institute of Mental Health.

EPIDEMIOLOGY

The prevalence of major depressive disorder in children is approximately 2%. In adolescence, there is a substantial increase in the prevalence of major depression, with lifetime prevalence rates of 12%. In children, the sex ratio is 1:1, whereas, with puberty, adolescent females are two to three times more likely to have an episode of major depression than adolescent males. Children with a family history of affective disorders are about three times more likely to develop a major depressive disorder than children with a first-degree relative without a lifetime affective disorder. The prevalence rate for persistent depressive disorder in children ranges from 0.6% to 4.6%, and in adolescents, it ranges from 1.6% to 8.0%.

Estimates of the prevalence rate of disruptive mood dysregulation disorder are about 3%. The prevalence is higher in young children (3% to 8%) than in adolescents (0.8%). This disorder is more commonly diagnosed in males compared to females.

ETIOLOGY

Genetic

The most predictive factor for a child or adolescent to develop major depressive disorder is a high family loading for major depressive disorder. There is an increased risk of depression in relatives of depressed children (odds ratio = 3.98 compared to healthy controls) and among offspring of depressed adults (odds ratio = 2.12 compared with healthy controls). Younger age of onset of depression in parents increases the risk of depression in children. Parental major depression is associated with greater severity of illness, longer episode duration, more impairment, and more recurrences in depressed offspring. Twin studies have demonstrated that depression is moderately heritable, with rates of about 31% to 42%.

Heritability of disruptive mood dysregulation disorder has not been specifically determined, although there may be a heritable component to the phasic irritability.

Neurobiology

Neuroendocrine studies have evaluated the hypothalamic–pituitary–adrenal (HPA), hypothalamic growth hormone (HGH), hypothalamic–pituitary–thyroid (HPT), and hypothalamic–pituitary–gonadal (HPG) axes in youth to determine whether there are abnormalities related to depression. HPA-axis abnormalities have been detected in depressed children and adolescents with some evidence of nonsuppression in stimulation studies and higher baseline cortisol values. Growth hormone–axis abnormalities have been found in prepubertal depressed children who have a blunted growth hormone response to stimulation with growth hormone–releasing hormone. There have been inconsistent findings regarding abnormalities in the HPT axis in children and adolescents with depression. The majority of studies of corticotropin-releasing hormone have found no abnormalities in depressed youth.

Abnormalities in the serotonergic system have been examined in depressed youth. Serotonin challenge agents that cause the release of prolactin have shown greater prolactin release in depressed children compared to controls.

Neuroimaging studies of depressed youth have found smaller left subgenual prefrontal cortex in adolescent girls, smaller frontal white matter volumes, larger frontal gray matter volumes, and larger lateral ventricular volumes. Amygdala functioning has been shown to be disrupted in depressed children. Disconnectivity in frontolimbic circuitry has been associated with depression in adolescents. Larger pituitary- and amygdala-to-hippocampal volume ratios have been found in depressed youth compared to healthy controls. Depressed children and adolescents have been found to have smaller left and right amygdala volume than healthy controls. Increased pituitary gland volume has been found in treatment-naive youth with major depressive disorder, particularly in males with nonfamilial major depressive disorder. Magnetic resonance spectroscopic imaging has found an increase in choline in the left dorsolateral prefrontal cortex and orbitofrontal cortex and increased choline/creatinine and choline/*N*-acetylaspartate ratios in the left anterior medial frontal lobe in depressed youth. A functional magnetic resonance imaging (fMRI) monetary task showed disruption in reward system neural circuitry in adolescent depression. Highly irritable youth with disruptive mood dysregulation disorder showed increased frontal striatal activation on fMRI following a frustrating task.

A meta-analysis of depression and inflammation among children and adolescents showed significant associations between depression and c-reactive protein (CPR) and interleukin-6 (IL-6). Sleep abnormalities have been examined in depressed youth, and the findings from polysomnography have been mixed. Some studies found no differences between children with major depressive disorder and controls, whereas other studies have found shorter rapid eye movement (REM) latencies, longer sleep latency, a higher number of REM periods, and increased REM time for depressed children. In studies of depressed adolescents, shorter REM latency, more REM sleep, more REM density, and longer sleep latency have been reported. Subjective reports by youths of sleep disturbance (sleep quality, waking in the middle of the night, morning tiredness) were not confirmed by an electroencephalogram (EEG). Of interest, the depressed children who rated their sleep most poorly had a better sleep by EEG measures than those who rated their sleep more positively. In a study using polysomnography as a measure of the recurrence of depression in youth, those youth who had a later recurrence had decreased sleep efficiency and delayed sleep onset at baseline sleep polysomnography. No clinically significant associations have been shown between disruptive mood dysregulation disorder and findings on polysomnography.

Environmental

Adoption studies have supported the role of environment more than genetic factors in the development of pediatric depression. Family factors, including parental substance abuse, poor marital adjustment, parent–child conflict, and lack of family cohesion, are risk factors for depression in children and adolescents. Life stressors, such as abuse, neglect, death of a parent, bullying, and psychosocial adversities, may contribute to depression in youth. Cigarette smoking, physical illness, and functional impairment have been associated with youth depression. Recurrence of major depression may also be affected by environmental stressors.

There may be gene–environment interactions with regard to pediatric depression. Twin studies of life events have shown that the effect of life events on depression is greater in youth with higher genetic liability. Family conflict and family adversity have also been demonstrated to interact with genetic liability and contribute to depression in youth.

Preliminary findings suggest that environmental factors such as parenting, peer victimization, and adverse life events may have a role in development of irritability and disruptive mood dysregulation disorder.

DIAGNOSIS AND CLINICAL FEATURES

It is important for the clinician to conduct a comprehensive evaluation of a child to ensure the accuracy of the primary diagnosis, as well as of comorbid diagnoses. The clinician should interview the child and parent separately so that each can express their concerns. Teenagers are often reluctant to disclose their feelings or issues related to peers in the presence of their parents. Similarly, parents may be reticent to discuss sensitive topics with their children present. Other sources of information may be obtained from discussion with school personnel, from school records, or from medical records.

Children with a diagnosis of disruptive mood dysregulation disorder are persistently irritable or angry most of the day nearly every day and have severe recurrent (>3 times per week) temper outbursts, either verbal or behavioral. These temper outbursts are developmentally inappropriate and are out of proportion to the situation. To make this diagnosis, the symptoms must be present for at least 12 months (no period >3 months without symptoms in this time frame) and be present in at least two settings (school, home, or with peers). The diagnosis cannot be given to children under the age of 6 years, and the onset of symptoms must be before age 10 years.

The irritability/angry mood state for children with disruptive mood dysregulation is chronic (tonic irritability). Particularly when faced with frustration, these children have marked difficulty regulating their emotions and respond with verbal or behavioral outbursts (phasic irritability). For example, a child may throw a chair in school during one of these outbursts. This persistent irritability and severe temper outbursts lead to significant impairment in school, at home, and in social settings. Children and adolescents with major depression have depressed or irritable mood and/or loss of interest or pleasure, along with at least four of the following symptoms: Weight or appetite change, sleep disturbance, psychomotor retardation or agitation, fatigue or loss of energy, feelings of worthlessness or guilt, diminished concentration, and suicidal ideation, intent, or plan. The symptoms must cause impairment in the child’s functioning with regard to, for example, social or school performance or peer relationships, which is critical to the diagnosis in youth. Impairment helps to distinguish these symptoms from a passing phase of childhood or adolescence.

Youth with major depressive disorder often display irritable mood rather than dysphoria. Typically, they are not aware of the extent of their irritability or its effects on their interaction with other people. Teenagers who have some insight into their irritability might say that everything makes them angry whether it is important or not. Loss of pleasure or interest may be noted by the youth's withdrawal from school or sports activities and from friends. Sleep disturbance is common in depressed youth, particularly difficulty falling asleep. Weight loss or failure to gain weight is more common than significant weight gain. Youth with depression are often fatigued and may nap after school. Diminished concentration may be manifested by declining school grades. A child often describes a feeling of worthlessness as no one liking them, including peers, teachers, and family members. Suicide attempts and psychotic features are more common in depressed adolescents than in depressed children.

Depressed youth may not view themselves as having depression if their mood is irritable rather than depressed. Parents often do not recognize individual symptoms of an adolescent's depression, but they recognize the burden on the family in dealing with an extremely irritable adolescent. Children and adolescents with depression are more likely to be brought for evaluation because of declining school grades, substance abuse, suicide attempt, or a change in behavior.

For a diagnosis of persistent depressive disorder, a child or adolescent must have depressed or irritable mood for at least 1 year. In addition, the youth must have two of the following symptoms: poor appetite or overeating, insomnia or hypersomnia, low energy or fatigue, low self-esteem, poor concentration or difficulty making decisions, and feeling of hopelessness. Youth with persistent depressive disorder have fewer melancholic and neurovegetative symptoms than youth with major depressive disorder. Although the depressive symptoms are milder in persistent depressive disorder than in major depressive disorder, the chronicity of persistent depressive disorder results in significant psychosocial impairment (e.g., relationships with friends and family).

Depressive disorders in youth are often comorbid with other psychiatric conditions. Rates of comorbidity range from 40% to 90% of children and adolescents with a depressive disorder, with up to 50% of depressed youths having two or more comorbid disorders. Anxiety disorders often precede the onset of depression and are a very common comorbid disorder. Major depression and persistent depressive disorder also frequently co-occur in youth. Other common comorbidities especially with disruptive mood dysregulation disorder are attention deficit hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), conduct disorder, and substance use disorder. It is important for clinicians to identify comorbid disorders that are associated with depressive disorders in youth, since these comorbidities may adversely affect the course and treatment outcomes of depressive disorders.

LABORATORY EXAMINATIONS/ASSESSMENT INSTRUMENT

No laboratory findings are diagnostic of depressive disorders in youth. If a child presents with symptoms of depression and hypothyroidism (e.g., dry skin, feeling cold, and memory difficulties), then thyroid function tests should be obtained. If hypothyroidism was found, the diagnosis would be depressive disorder due to another medical condition.

The Patient Health Questionnaire-9 modified for adolescents (PHQ-A) is a widely used self-report screening instrument for depression in youth aged 11 to 17 years. The scale has nine depression symptom items rated on a 4-point scale from 0 (not at all) to 3 (nearly every day). Scores range from 0 to 27. PHQ-A score ≥10 is typically the cut-off threshold for further evaluation to determine a diagnosis of major depression. This scale may also be useful for monitoring treatment outcomes.

The Children’s Depression Rating Scale-Revised (CDRS-R) is a 17-item, clinician-administered instrument that assesses symptoms of depression in children and adolescents. The scale assesses cognitive, somatic, affective, and psychomotor symptoms. The items are rated for severity on a scale of 1 to 7 for some items and 1 to 5 for other items for a total score of 113. Typically, a score of 45 or greater is considered a marker of clinically significant depression. The CDRS-R is useful for obtaining baseline symptomatology prior to treatment and for monitoring treatment outcome.

Given that youth with bipolar disorder may present with a depressive episode, it is important to ask about symptoms of mania or hypomania. Some diagnostic checklists for mania have been shown to discriminate bipolar from nonbipolar disorder in youth of ages 5 to 18 years, including the Mood Disorder Questionnaire-Adolescent (MDQ-A), the Parent General Behavior Inventory 10-item short form (PGBI-SF10), and the Parent Young Mania Rating Scale (P-YMRS). These checklists are completed by the parents about their child's behavior. Adolescent self-report measures about symptoms of mania have demonstrated low validity in identifying bipolar disorder.

DIFFERENTIAL DIAGNOSIS

Depressive symptoms may be related to stressors and fulfill criteria for an adjustment disorder with depressed mood. A child may have experienced the death of a significant person in their life, and the depressive symptoms may be related to bereavement. Depression or irritability may be attributable to substance use or withdrawal, which would be considered a substance-induced depressive disorder.

When a child presents with depressive symptoms, it is important for the clinician to consider bipolar disorder within the differential diagnosis. The clinician should ask about a past or present history of manic or hypomanic symptoms.

There is substantial symptom overlap among various psychiatric disorders, which might complicate the differential diagnosis. Irritability may be found in disruptive mood dysregulation disorder, major depression, persistent depressive disorder, bipolar disorder, post-traumatic stress disorder, and generalized anxiety disorder. Children with ADHD who are failing in school and having difficulty with peer relationships may exhibit symptoms such as depressed mood, low self-esteem, irritability, and behavioral outbursts. Anxiety disorders and ODD may have some symptom features of depression and disruptive mood dysregulation disorder. Children with pervasive developmental disorder may present with depressive symptoms but not fulfill diagnostic criteria for a depressive disorder.

Medical conditions may manifest with symptoms of a depressive episode and should be diagnosed as depressive disorder due to a general medical condition if the mood disorder is attributable to the consequences of the medical condition. Neurologic disorders (e.g., seizure disorder and multiple sclerosis), endocrinologic abnormalities (hypothyroidism, hypopituitarism, Cushing disease), infections (e.g., pneumonia, AIDS, and mononucleosis), autoimmune diseases, and cancer are some medical conditions to consider within the differential diagnosis. Medications such as anti-inflammatory agents, antibiotics, steroids, stimulants, and oral contraceptives may lead to depressive symptoms.

COURSE AND PROGNOSIS

Disruptive mood dysregulation disorder has its onset after the age of 6 years. It has been found that disruptive mood dysregulation disorder at age 6 predicted disruptive mood dysregulation disorder, depressive disorders, and ADHD at age 9 years. Academic difficulties, use of education support services, school suspensions, peer relationship problems, and impaired relationships with parents and siblings are frequent outcomes of this disorder. The symptoms are predominantly limited to childhood. Of children who were diagnosed with disruptive mood dysregulation disorder, only 19% continued to meet criteria for the disorder at 12- and 24-month follow-up. Children with disruptive mood dysregulation disorder are at risk to develop depression and anxiety disorders in adulthood.

The mean length of an episode of major depression in children and adolescents is 8 months. Recovery rates from an index episode of depression are high in youth, even independent of treatment. More than 90% of depressed children and adolescents typically recover from an episode of major depression within 1 to 2 years.

Depression relapse rates, with relapse being defined as a return of symptoms in the index episode, are between 34% and 50%, with the greatest likelihood of relapse in the first 6 months to 1 year after treatment discontinuation. Recurrence rates (onset of a new episode of depression) are high, with rates ranging from 20% to 60% after 2 years of remission and up to 70% at 5-year follow-up.

Risk factors for relapse or recurrence include severity of illness, suicidality, comorbid disorders, residual symptoms, psychosocial stressors, family conflict, family history of recurrent depression, negative cognitive style, low socioeconomic status, early onset of parental mood disorder, baseline social difficulties, and history of sexual abuse.

Youth who present with an episode of major depression are at risk of developing bipolar disorder. Overall estimates are that 20% to 40% of youth with a major depressive episode are at risk of subsequent development of bipolar disorder. In a longitudinal study of prepubertal children diagnosed with major depression, the rate of bipolar I disorder was 33% and that of bipolar spectrum disorders (I, II, not otherwise specified [NOS]) was 48% by young adulthood. Risk factors for switch to bipolar disorder are prepubertal onset of major depression, loaded or multigenerational family history for mood disorders, psychotic features, and hypomania when treated with antidepressant medications.

There is increased recognition of the continuity between depression in youth and depression in adulthood. In a follow-up study of depressed school-aged children in community and clinic samples, 18% continued to have major depression into adulthood. In a clinical sample, adolescent-onset major depressive disorder increased the risk twofold of major depressive disorder in adulthood. In a community sample, 53% of adolescents with a history of major depressive disorder developed a new major depressive disorder in young adulthood.

Early age of onset of depression increases the risk of anxiety disorders, substance abuse, psychiatric comorbidity, suicide attempts, psychiatric and medical hospitalization, impairment in relationships with parents, siblings, and friends, more lifetime depressive episodes, nicotine dependence, increased adulthood body mass index, educational underachievement, unemployment, early parenthood, and worse overall quality of life.

The duration of persistent depressive disorder ranges from 2 to 3 years in clinically referred and community-based youth. After recovery after an episode of persistent depressive disorder, the estimated relapse rate is 45%. Youth with persistent depressive disorder are at increased risk of psychosocial difficulties, substance use disorder, episodes of major depression, and suicidality. The risk of suicide attempts in youths with persistent depressive disorder increases with comorbid conduct disorder or substance use disorder.

Children and adolescents with both major depressive disorder and persistent depressive disorder have a worse course of illness than those with either disorder alone. These youth have more-severe depression, longer depressive episodes, more comorbid disorders, more suicidality, and greater psychosocial impairment.

PHARMACOLOGIC TREATMENT

There are no U.S. Food and Drug Administration (FDA)-approved medications to treat disruptive mood dysregulation disorder. There are limited data on pharmacologic treatment specifically for disruptive mood dysregulation disorder in youth. Medications aimed to treat emotional dysregulation, irritability, and aggression include stimulants, selective serotonin reuptake inhibitors (SSRIs), mood stabilizers, and second-generation antipsychotics.

Citalopram adjunctive to stimulant medication has recently been studied in a double-blind placebo-controlled trial of youth aged 7 to 17 years with severe mood dysregulation disorder (98% met criteria for disruptive mood dysregulation disorder). After 5 weeks of open treatment with methylphenidate to optimize control of ADHD symptoms, youth who continued to meet criteria for severe mood dysregulation disorder were randomized to citalopram plus methylphenidate ($n = 25$) or placebo plus methylphenidate ($n = 28$) for 8 weeks. A significantly higher proportion of response (improvement of irritability) was found for the citalopram plus methylphenidate group (35%) versus the placebo plus methylphenidate group (6%). There was no difference between the groups in measure of functional impairment.

Lithium for treatment of severe mood dysregulation disorder has been evaluated in a 6-week double-blind placebo-controlled trial for youth aged 7 to 17 years. After a 2-week single-blinded placebo run-in, those youth who continued to meet criteria for severe mood dysregulation disorder were randomized to lithium ($n = 14$) or placebo ($n = 11$). There were no significant differences in clinical improvement between the youth randomized to lithium or placebo.

Risperidone was evaluated for treatment of 21 youths aged 7 to 17 years with severe mood dysregulation in an open-label trial. A significant reduction in irritability scores was reported from baseline to endpoint.

There are two SSRIs that have FDA approval for acute and maintenance treatment of major depression in children and adolescents: fluoxetine (≥8 years old) and escitalopram (≥12 years old).

Other SSRIs that have also demonstrated efficacy in controlled studies of depressed children and adolescents are citalopram and sertraline. Some studies of SSRIs did not demonstrate superiority of medication to placebo on the primary outcome measure, including citalopram, paroxetine, and escitalopram. A meta-analysis of SSRI trials in youth showed superiority of SSRIs to placebo, with an average response rate of 60% for antidepressants and 49% for placebo.

Other classes of antidepressants have not shown statistically significant improvement with antidepressant medication compared to placebo, including venlafaxine, nefazodone, desvenlafaxine, vilazodone, mirtazapine, and selegiline transdermal system. In failed studies, duloxetine and fluoxetine were not superior to placebo and desvenlafaxine and fluoxetine were not superior to placebo on primary outcome measures. Tricyclic antidepressants have not been shown to be effective in the treatment of depressed children and adolescents. There is some speculation that the early maturation of the serotonergic system compared to the noradrenergic system may account for the lack of efficacy in noradrenergic antidepressants compared to serotonergic agents in youth. There are no controlled trials of bupropion in the treatment of major depression in children and adolescents.

There has been much discussion about why controlled studies of antidepressants in children and adolescents with major depression have not shown efficacy, that is, statistically significant separation from placebo. One explanation is the high rate of placebo response in these studies—up to 59%. Other explanations are methodologic issues in the antidepressant trials, including number of sites, experience of site personnel, patient recruitment source, inclusion and exclusion criteria, duration of treatment trial, dosing, and outcome measures.

Adverse Events

Common side effects of SSRIs are nausea, abdominal pain, headaches, dry mouth, insomnia, somnolence, and sexual dysfunction. Impulsivity, agitation, irritability, and behavioral activation may occur, particularly with young children. Potential serious side effects include serotonin syndrome, increased bleeding, and amotivational syndrome. Common side effects of venlafaxine are anorexia, abdominal pain, insomnia, somnolence, dizziness, and elevated blood pressure. Common side effects of mirtazapine are increased appetite, weight gain, and somnolence. Common side effects of duloxetine are nausea, headache, abdominal pain, and somnolence. Common side effects of desvenlafaxine are headache, abdominal pain, nausea, and dizziness. Common side effects of vilazodone are nausea, abdominal pain, vomiting, diarrhea, nasopharyngitis, headache, and dizziness.

FDA Warning and Suicidality

In a combined analysis conducted by the FDA of 24 short-term, placebo-controlled trials of antidepressant medications in children and adolescents with major depressive disorder, obsessive-compulsive disorder, or other psychiatric disorders, the risk of suicidality (suicidal thinking and behavior) was 4%—twice the placebo rate of 2%. There were no suicides in any of the clinical trials. The FDA directed manufacturers to add a black box warning to the health professional label of antidepressant medications to describe the increased risk of suicidal thoughts and behavior in children and adolescents being treated with antidepressants, as well as to recommend the need for close monitoring of patients on antidepressant medications.

In a later meta-analysis of 27 trials of pediatric major depression, the risk of suicidality (suicidal ideation and attempts) was 3% in the antidepressant group and 2% in youth who received placebo. The pooled absolute rate of antidepressant response was 61%, and the placebo response was 50%. The number needed to treat (NNT) (number needed to treat to get one positive response to medication) was 10, and the number needed to harm (NNH) (number needed to get one event of suicidality) was 112. Based on this finding, the investigators concluded that the benefits of antidepressants in youths outweigh the potential risk of suicidal ideation or attempts.

There have been recent reports of an increase in suicide rate in youth following the black box warning on antidepressants and a significant decrease in the rates of diagnosis and treatment of pediatric depression.

A medication guide that addresses antidepressant medications, depression, and suicidality has been approved by the FDA. The guide states that depression and other serious mental illnesses are the most important causes of suicidal thoughts and actions and that antidepressant medication may increase suicidal thoughts or actions in some children and teenagers when medication is first started. The patient or family member is encouraged to contact the healthcare provider if there is a change in mood or thoughts.

Informed Consent

Prior to prescribing antidepressant medication to a child or adolescent, an informed consent process should occur with the youth and the parent. Elements of informed consent are the purpose of treatment, description of the treatment process, benefits and risks of the proposed treatment, and alternative treatments. Because there is minimal information about the long-term effects of antidepressant medication in children, parents should be informed that there may be unknown risks of these medications.

Medication Treatment Algorithm

An evidence-based consensus algorithm for the treatment of major depression in children and adolescents was reported in which four stages of medication treatment were identified. It was recommended that if a child failed to respond to treatment in one stage, then the clinician should move to treatment in the next stage. The first stage of treatment is with an SSRI. The second stage of treatment is with an alternative SSRI. The third stage of treatment is selection of a different class of antidepressant medication. If a child fails to respond to these agents, then it was recommended that there be reassessment of the diagnosis and treatment consultation.

In the Treatment of SSRI-Resistant Depression in Adolescents (TORDIA) study, it was found that there was no difference in response rates for those adolescents switched to venlafaxine or another SSRI (48% vs. 47%) after treatment failure with an SSRI. Therefore, if an adolescent does not respond to the initial antidepressant, there is approximately a 50% chance of responding to the second antidepressant.

Lower initial starting doses of antidepressants are typically used in children and adolescents than in adults. However, the dose range tends to be similar to that used with adults. Doses are usually increased by 3 weeks if there is no symptom improvement. If there is no improvement with additional dose adjustment after 8 to 10 weeks, then switching to another treatment is recommended. Antidepressants should be continued for 6 to 12 months after symptom remission. At the time of discontinuation of an antidepressant, the dose should be tapered slowly, usually over a 2- to 3-month period.

PSYCHOTHERAPY

There are limited evidence-based psychosocial treatments for disruptive mood dysregulation in youth.

The feasibility and preliminary efficacy of dialectical behavior therapy for preadolescent children (DBT-C) was assessed in a randomized clinical trial for 43 children aged 7 to 12 years with disruptive mood dysregulation disorder. DBT-C consisted of 32 weekly 90-minute sessions and included a parent-training component. The primary outcome measure was the positive response on the Clinical Global Impression Improvement Scale (improvements in behavior outbursts and angry/irritable mood). The rate of positive response was higher in the DBT-C group compared to the treatment as usual (90.4% vs. 45.5% respectively).

Other psychosocial treatments for symptoms of disruptive mood dysregulation disorder include behavioral therapy, cognitive behavioral therapy (CBT), exposure-based CBT, integrative group therapy (parent–child sessions), and parent management training.

The two most studied psychotherapy modalities for depression in adolescents are CBT and interpersonal psychotherapy (IPT).

Cognitive Behavioral Therapy

CBT focuses on cognitive distortions and behavioral deficits associated with depression in youth. Components of CBT for children and adolescents include psychoeducation and mood monitoring, pleasant activity scheduling and behavioral activation, and cognitive restructuring. Relaxation training and improvement in social and problem-solving skills may be included.

The efficacy of CBT was assessed in adolescents with major depression or dysthymia. The youth were randomly assigned to adolescent group CBT, adolescent group CBT with a separate parent group, or wait-list control. The CBT groups showed higher depression recovery rates (67%) than the wait-list group (48%) and greater reduction in self-reported depression. The addition of a parent group did not affect outcome.

CBT (12 to 16 sessions) was compared to nondirective supportive psychotherapy and systemic behavior therapy in a randomized trial. CBT was significantly superior to the other treatments, with greater reduction in depression, more rapid symptom relief, and a higher remission rate—65%, compared to 39% for nondirective supportive therapy and 38% for systemic behavior family therapy. However, in a 2-year follow-up study, there were no significant differences in outcome among the psychotherapies. Eighty percent of the adolescents had recovered from depression, but 30% had a recurrence of depression.

Computerized CBT in the form of an interactive fantasy game (SPARX) has been developed for adolescent depression. A study comparing SPARX to treatment as usual for adolescents with depressive symptoms showed similar treatment response rates.

Interpersonal Psychotherapy

IPT focuses on the relationship between depressed symptoms and interpersonal relationships. Four areas of focus are loss, interpersonal disputes, role transition, and interpersonal deficits. IPT has been modified for depressed adolescents (IPT-A) and has added a focus on developmental issues such as separation from parents, authority figures, peer pressures, and dyadic relationships. The adolescent and parents are educated about the association between depressive symptoms and events in interpersonal relationships. The adolescent learns skills to improve interpersonal relationships that may help to relieve depression.

The efficacy of IPT-A was assessed in a 12-week study of 48 adolescents with major depression who were randomly assigned to IPT-A or clinical monitoring. IPT-A led to greater decrease in depressive symptoms, improvement in social functioning, and improved problem solving compared to clinical monitoring.

The effectiveness of IPT-A in school-based health clinics was evaluated in depressed adolescents who were randomly assigned to IPT-A or treatment as usual for 16 weeks. The IPT-A training was done with existing staff at the health care clinics in high school. The adolescents who received IPT-A had greater symptom reduction and improvement in overall functioning. The greatest treatment effects occurred in older and more severely depressed adolescents.

There is also a group format (IPT-AG) that allows for peer support for depressed adolescents who have similar interpersonal problems. Adolescents can receive feedback from their peers about their interpersonal interactions and strategies to improve them. The efficacy of IPT-AG was compared to individual IPT-A, and there was similar improvement in depression for both formats of IPT.

In a recent systematic review and meta-analysis of psychotherapy for depression, across age groups, effect sizes in children were 0.35 and in adolescents 0.55.

A meta-analysis of 50 controlled psychotherapy trials for treatment of depression in youth found an overall modest effect size of 0.36 at posttreatment and 0.21 at follow-up (mean 42 weeks after treatment).

Combination Treatment

The Treatment for Adolescents with Depression Study (TADS) was the largest randomized, controlled trial to compare the efficacy of CBT, fluoxetine, CBT plus fluoxetine, and placebo in adolescent outpatients with a diagnosis of major depression. Response rates (much or very much improved) for CBT (43%) were not statistically significant compared to those for placebo (35%). Response rates to fluoxetine plus CBT (71%) and fluoxetine alone (61%) were significantly superior to those to CBT alone and to placebo. Fluoxetine with CBT showed the greatest reduction in suicidal thinking. Although remission rates at the end of 12 weeks of treatment were highest in the combination group (37%) compared to the other treatment groups (fluoxetine, 23%; CBT, 16%; placebo, 17%), these overall remission rates were low. The long-term effectiveness of these treatments was assessed over 36 weeks. It was found that for adolescents with moderate to severe depression, treatment with fluoxetine alone or the combination of fluoxetine with CBT accelerated the improvement of depression compared to CBT alone. The addition of CBT to fluoxetine decreased persistent suicidal ideation and treatment-emergent suicidal events.

A recent systematic review and network meta-analysis was conducted to compare the efficacy of antidepressants, psychotherapies, and their combination for treatment of children and adolescents with major depression. Fluoxetine alone or in combination with CBT was superior to CBT alone and psychotherapy for moderate to severe depression in children and adolescents.

TREATMENT-RESISTANT DEPRESSION

Forty percent of youth with major depression do not respond to treatment with antidepressant medication or psychotherapy.

Ketamine has recently been investigated for treatment-resistant depression in adolescents. In an open-label study, 13 adolescents received intravenous ketamine infusion and 38% met criteria for clinical response. In a double-blind crossover trial that included 17 adolescents with treatment-resistant depression, single-dose intravenous ketamine was shown to be significantly superior to midazolam on depression rating scores. Response rates were 76% for ketamine versus 35% for midazolam. Ketamine was associated with transient dissociative symptoms.

Transcranial magnetic stimulation (TMS) has been investigated in uncontrolled studies which have reported some effectiveness for treatment-resistant depression in adolescents. However, a recent double-blind randomized sham-controlled trial did not demonstrate superiority of 10-Hz TMS compared to sham TMS for adolescents with treatment-resistant depression. Response rates were 41.7% for active TMS and 36.4% for sham treatment.

OTHER TREATMENT COMPONENTS

Treatment of depressive disorders in children and adolescents should also include psychosocial factors, psychoeducation, supportive management, family involvement, and school involvement.

Psychosocial factors such as abuse, family conflict, parental psychopathology, peer difficulties, bullying, and family discord might be exacerbating the symptoms of depression and mood dysregulation and should be addressed as part of the treatment plan.

Psychoeducation involves educating the patient and parents about depression, including causes, symptoms, course, and treatment options, including benefits and risks of treatment. Depression should be explained as an illness, not a character flaw of the child or parent. The parent and child should be informed about the importance of adhering to treatment to ensure an optimal outcome.

Supportive management includes assisting the child with problem solving, improving coping skills, and fostering involvement in treatment.

It is important to involve parents in the treatment process. Parents can provide needed information about the child’s symptoms over the course of treatment. Parents may also benefit from guidance in managing their depressed child at home.

Regarding school involvement, children with depression may have academic difficulties that require school remediation or accommodation in the child’s school schedule or program.

SUICIDAL BEHAVIOR IN CHILDREN AND ADOLESCENTS

Although suicide is not a psychiatric diagnosis, suicidal behaviors are a significant public health problem and a focus of attention. Suicidal behavior is often present in individuals with depression, and the presence of a psychiatric diagnosis of depression is one of the most potent risk factors for suicide. However, not all individuals with depression exhibit suicidal behaviors or die by suicide. Furthermore, treatments targeting depression alone have been insufficient for reducing suicide risk.

Definitions

Suicidal behavior is complex and there is no single etiology, it exists on a continuum from wishing to be dead, to thoughts or plans to die by suicide, to death by suicide. Suicidal thoughts or suicidal ideation can vary from passive thoughts of death (e.g., maybe I would be better off dead, I wish I wouldn't wake up in the morning), to active suicidal ideation (e.g., I would like to kill myself, I wonder if I would die if I jumped off the roof.), to suicidal ideation with a plan and intent (e.g., I plan to kill myself with my hunting rifle when my parents go out tonight. I am going to take a bottle of pills after my parents go to sleep). Suicidal threats are the verbalization or a written account of suicidal ideation. A suicide attempt can be defined as "self-destructive behavior with inferred or explicit intent to die." This definition differentiates suicidal behavior from "nonsuicidal self-injurious behavior (NSSI)," which can be defined as a "direct and deliberate destruction of one's own bodily tissue (e.g. cutting, burning, scratching, self-hitting) in the absence of lethal intent and for reasons not socially sanctioned, and without the intent to die. NSSI is most commonly a maladaptive behavior to address affect. Though NSSI is not suicidal behavior, recent literature suggests that for some adolescents, particularly girls in early to midadolescence, NSSI is an independent risk factor for death by suicide. Suicide is an intentional self-destructive act that results in death. Suicide contagion is the exposure to suicide or suicidal behaviors within a person's family, social group, or through media reporting on high-profile deaths by suicide, which can result in an increase in suicidal behaviors and death by suicide among children and adolescents at risk for death by suicide.

Intentional asphyxiation can resemble suicide by hanging. In intentional asphyxiation, there can be a sexual component—often termed autoerotic asphyxiation, or there can be no sexual component. This has been seen more recently in the rise of children and adolescents playing "the choking game." In both forms of intentional asphyxiation, an individual intentionally restricts carotid circulation to the brain using a neck ligature in order to experience euphoria. In autoerotic asphyxiation, there is also simultaneous masturbation or other sexual play to heighten the intensity of sexual climax. The presence of a ligature, sexual paraphernalia, and evidence of sexual activity support the autoerotic asphyxiation as an accidental death versus suicide. In intentional asphyxiation not related to sexual behavior (e.g., child or adolescent playing the choking game after being introduced to it at school or through online videos), there may be no suicide note and no sexual paraphernalia present. Although intentional asphyxiation may not be intentional suicide, this does not exclude suicidal intent as a child or adolescent who is engaging in intentional asphyxia may know the risk that they might die by engaging in this behavior, and thus this may still be suicide.

Epidemiology

Suicide in children and adolescents is a significant public health concern. Suicide is the second leading cause of death among children and adolescents aged 10 to 19 years. The Youth Risk Behavior Survey (YRBSS) includes national, state, territorial, and freely associated state, tribal government, and local school-based surveys of representative samples of 9th- to 12th-grade students. The YRBSS for 2019, which gathered data on risk behaviors using self-report, found that, in 2019, about 37% of high school students had experienced periods of persistent feelings of sadness or hopelessness, 16% of youth had a suicidal plan, 9% made a suicide attempt, and 3% made an attempt requiring medical attention.

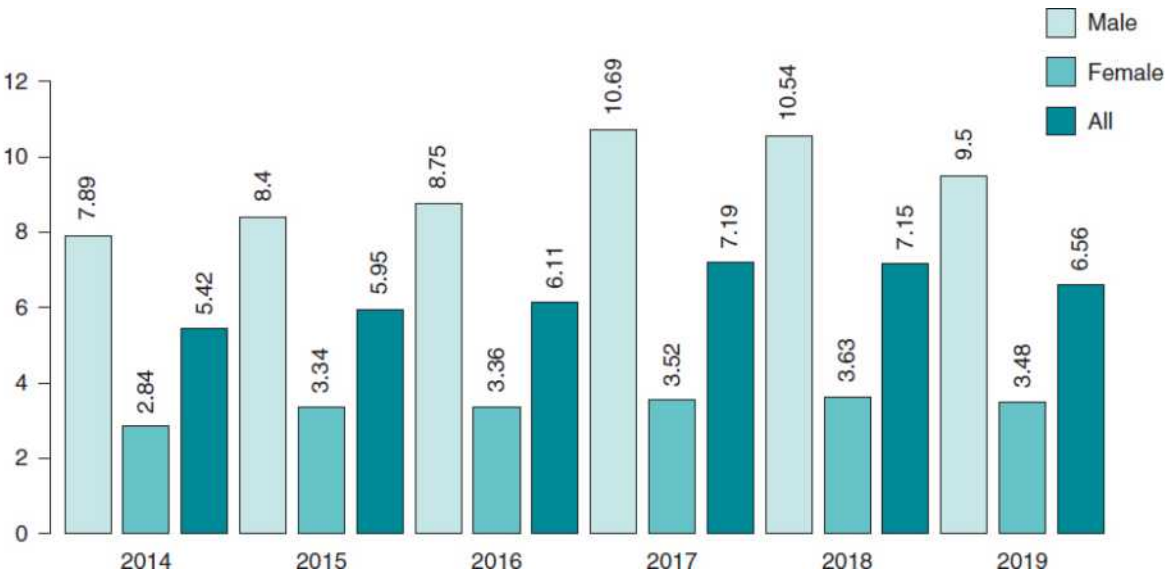

FIGURE 50.1–1. U.S. rates of suicide (per 100,000) in youth aged 10 to 19 years: overall by sex. (Centers for Disease Control and Prevention. Accessed Injury Prevention Control: Data & Statistics [WISQARS]. Updated July 16, 2021. http://www.cdc.gov/injury/wisqars/index.html)

Lifetime rates of suicidal behavior using interview data as opposed to self-report yield lower estimates of lifetime suicidal ideation with plan and attempts, both around 4%. However, these rates may also be skewed as certain populations may not be included in such interviews, and secondary to stigma, not all children or adolescents would volunteer information regarding suicidal behavior.

Although deaths by suicide are much rarer than suicidal behavior, occurring in 10.5/100,000 of adolescents aged 15 to 19 years in 2019, and 2.57/100,000 of adolescents aged 10 to 14 years. Suicide remains the second leading cause of death for children and youth aged 10 to 24 years (see Fig. 50.1–1). There is an unstable death by suicide variable of 0.04/100,000 in children aged 5 to 9 years. This may be in part because suicide deaths in this age range are often not listed by coroners or medical examiners as suicide. Governmental agencies tend to believe that children younger than 10 years old may not appreciate the seriousness of suicide, despite the evidence that children as young as 5 years old do die by suicide. Further, even in older children and adolescents, the cause of death by suicide is often misclassified as unintentional accident. The reasons for this misclassification are myriad: they may include reasons related to social stigma, the age of the child, social or cultural norms, or a coroner's unwillingness and/or inability to accept suicide death in a young person. It is likely that the true rate of death by suicide among children and adolescents is higher than what is reported.

The rates of suicidal ideation, attempts, and deaths by suicide increase with age across childhood and adolescence, peaking in late adolescence. This increase in ideation may be explained by an increased incidence of depression that occurs during adolescence. The increased rate of attempts and suicide deaths in adolescence relative to childhood may also be explained by higher rates of depression, ongoing emotional dysregulation, any substance use, more sophisticated ability to plan and execute a potentially lethal suicide attempt, and easier access to lethal methods. There is limited evidence suggesting that overuse of social media and internet access may lead to increased exposure to suicidal themes, and to an increased risk of suicide secondary to suicide contagion.

Suicidal ideation has previously been thought to occur more frequently in female than male adolescents, though recent literature indicates there is not a true gender disparity in suicidal ideation or thinking about suicide. Suicide attempts do have a gender disparity with adolescent girls attempting suicide with greater frequency than adolescent boys, particularly after puberty. This is a pattern that is also observed with incidence of depression in adolescents post puberty. However, suicide deaths are 1 to 1.5 times as common in males than females aged 10 to 14 years, (with the exception of indigenous people where the rate is 2:1 males to females) and as much as 3 times as common in males than females aged 15 to 19 years. The higher rate of suicide in late adolescent males relative to females may be due to higher rates of substance abuse and conduct disorder and a tendency to use more lethal means of suicide.

The rate of suicide is particularly high in the United States among the indigenous people (Native Americans/Pacific Islanders) with indigenous males dying by suicide at a rate that is three times higher than the overall death-by-suicide rate for children and adolescents (Fig. 50.1–2). There is an elevated risk of suicide among indigenous peoples worldwide and may be related to erosion of traditional cultural values secondary to colonization and attempted genocide, and resulting social ills most notably poverty, unemployment, substance use, and disruption of traditional family structure.

White youth have a higher rate of death by suicide than African American youth, although the rate of suicide among African American boys has climbed over the years to around 75% of the rate in white males. There has been an alarming increase in rates of suicide deaths in younger African American boys (aged 5 to 11 years), while rates in White boys in the same age range have decreased. Suicide rates have been increasing for Black boys aged 10 to 19 years for some time, increasing by 114% between 1980 to 1995. There continues to be an upward trend for Black youth, simultaneously with downward trends for Caucasian youth through 2019. This alarming increase in Black youth suicide prompted Congress to convene an emergency task force to address suicidality and mental health for Black youth in 2019.

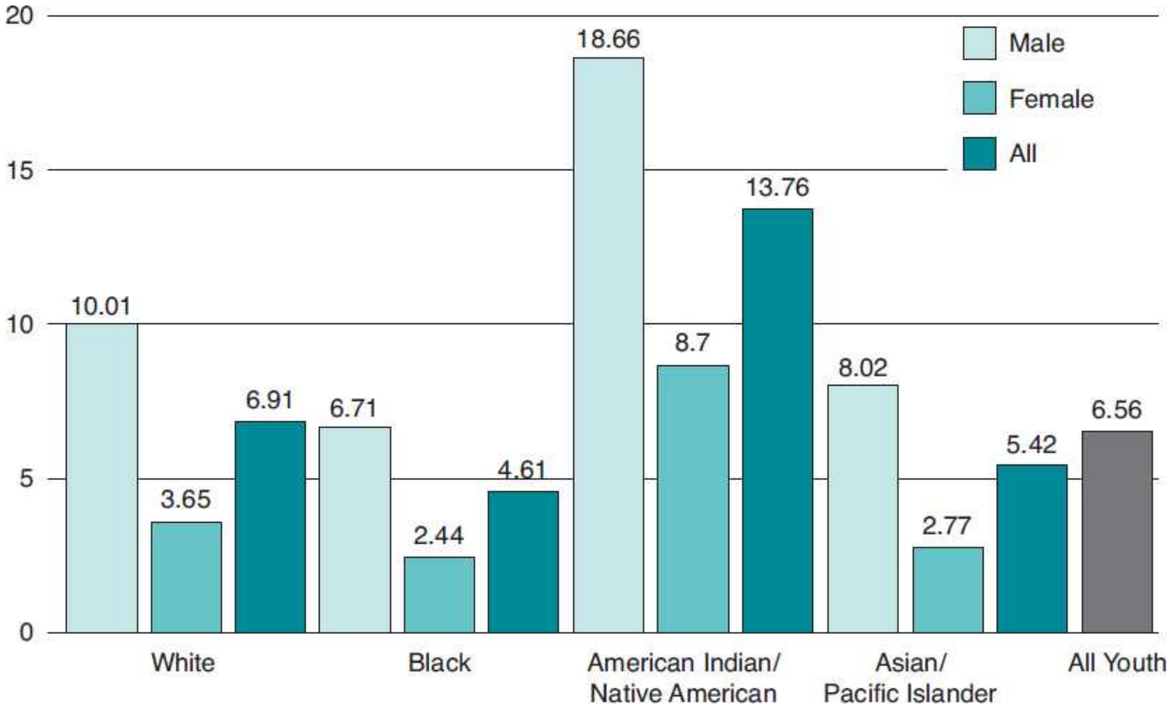

FIGURE 50.1–2. U.S. suicide rate in 10 to 19 year olds (per 100,000) by sex and race (2019). (Centers for Disease Control and Prevention. Accessed Injury Prevention & Control: Data & Statistics [WISQARS]. Updated July 16, 2021. http://www.cdc.gov/injury/wisqars/index)

After initially peaking in the 1990s, the reported suicide rate in adolescents began to decline until 2004, when it increased for the first time in a decade. Since 2007, there has been a gradual increase in the suicide rate both in males and females, with 2018 showing the highest rate of death by suicide in the past 30 years. The decline in the rate during the 1990s had been attributed to several factors, including more restrictive gun control laws and increased prescription of SSRIs though it is difficult to identify causal factors. The use of SSRIs among adolescents has declined since 2004, which is when the FDA initiated a black box warning on antidepressants. Some pharmacoepidemiologic studies have shown a relationship between the decline in use of antidepressants and an increase in the rate of suicide and suicide attempts, but none of these studies are definitive. There are likely other confounders that arose during this period including increased reliance on technology, the use of social media (and media in general), and fewer in-person social interactions as children and adolescents connect with friends through various gaming and online platforms.

The most common method for adolescent suicide attempts is overdose, with the second-most common method being wrist cutting. Although suicide attempts in adolescents are frequently of relatively low intent, ingestion of acetaminophen is associated with a significant risk of liver toxicity and fatality without prompt medical attempt. Much more rarely, adolescents engage in highly lethal suicide attempts, using methods more commonly found in suicide deaths. The most common method in the United States for suicide fatality (in the past 5 years) is by firearm, followed in frequency by hanging (suffocation), drug poisoning, nondrug poisoning, falling or jumping, cutting, and drowning.

Risk Factors, Assessment, and Etiology

Suicide risk is moderated by the interactions between a variety of factors. Most conceptual models of suicide posit that suicide risks result from interactions between predisposing factors (distal, familial/genetic), precipitating factors (proximal, triggering), and developmental factors. The biopsychosocial model for suicide and suicide risks conceptualizes the interplay between genetic, psychological, sociologic, clinical, experiential, and environmental states as causative factors for suicide. While this framework is helpful in understanding the relationships between risk factors, their immediate relevance to suicide risk is imprecise. Any one or all of these factors might be involved at any point in time, and their impacts are moderated by individual traits.

Other relevant risk factors that can help to identify suicide risk are clinical factors that influence the transition from suicidal ideation to attempt, such as prior self-harm events, exposure to suicide by others, depressive disorders, substance use, impulse control disorders, anxiety disorders, eating disorders, and traumatic stress-related disorders. Essentially, any disorder that may decrease restraint, such as impulse control disorders, or increase distress, such as panic, may increase the likelihood of moving from suicidal ideation to suicide attempt, especially if a mood disorder is also present.

Other important risk factors and antecedents to suicide include psychiatric disorders, psychological and biologic factors, familial and environmental factors, chronic physical illnesses, risky health behaviors, and personal, community, or public exposure to suicidal behavior that occurs in media reports.

Psychiatric Disorders. Estimates of the proportion of individuals with a mental illness at the time of their death by suicide or suicide death approximates 90%. These disorders, gathered through retrospective chart reviews, family interviews, and psychological autopsy, may have been undiagnosed prior to their deaths. Mood disorders are strongly associated with suicidal ideation, attempts, and completed suicides. Of adolescents who die by suicide, approximately 60% had a mood disorder at the time of their deaths. Most youths with psychiatric disorders never engage in suicidal behavior or die by suicide, suggesting that the treatment of psychiatric disorders may be a modifiable risk factor for suicidal individuals though not sufficient in itself. The presence of a psychiatric disorder does not in itself predict suicide. However, there are specific features of certain disorders that increase suicide risk, such as bipolar disorder, when an individual is experiencing mixed or rapid cycling states or irritable dysphoric states, or when a depressive disorder is accompanied by suicidal ideation, insomnia, agitation, and irritability. Disorders characterized by greater chronicity, severity, or complicated by comorbidity are more likely to be present in psychiatrically ill adolescents with suicidal ideation, attempts, or completion. Recent data suggests high rates of suicidality among clinically referred mood-disordered youths, with approximately one-third of those with bipolar disorder endorsing one suicide attempt. One study found a 10-year rate of suicide of 7.7% when compared with no suicides among a control group.

Anxiety disorders have also been reported to be associated with suicidal ideation and attempts. Social anxiety, panic disorder, and post-traumatic stress disorder have been most commonly associated and contribute most strongly to suicide risk in the context of an anxiety disorder.

Studies have shown that alcohol misuse increases the risks of suicide attempts and suicide deaths in clinical and nonclinical samples by decreasing restraint from suicidal behaviors for individuals who are contemplating suicide. The presence of a mood disorder further increases the risk for suicidal behaviors. Alcohol misuse may trigger suicidal ideation even in the absence of high levels of depressive symptoms. Although alcohol has been the most studied substance of abuse in suicide, cannabis, and sedative–hypnotic use are more closely associated with suicidal behaviors than other substances. Schizophrenia and eating disorders are also associated with suicide attempts in adolescence, but increased mortality is more likely to occur in adulthood. There is growing evidence describing increasing rates of suicide among youth with autism spectrum disorders.

Medical Conditions. Studies of adolescents with chronic illnesses have found that the overall risk for suicidal ideation and behaviors is greater when compared to adolescents without chronic illnesses. However, the risk for suicidality is dependent upon the specific medical illness, the required treatments, and pre-existing psychiatric conditions. Some disorders, such as those involving the central nervous system, including migraine and epilepsy, inflammatory disorders, such as inflammatory bowel disorders, respiratory disorders such as asthma, endocrine disorders such as diabetes, and chronic pain may have an increased risk of suicidal behavior. Some disorders, such as cancer, may have increased suicide risks at differing times during treatment as a result of remission or relapse status and/or the treatment itself. Factors that appear to contribute to increased risk for depression and suicidal behavior in chronic illness are related to the general experience of illness, including pain and discomfort, disability, disruption of enjoyable activities and achievement, the pathophysiology of the disease itself, for example, inflammation, and effects of treatments used to treat the disorder, for example, steroids.

Biologic Factors. Understanding the neurobiology of suicidal behavior is complicated by the myriad biologic changes in the brain affecting many different neural pathways. However, parsing out the effects of these changes to determine which are specific to suicide and suicidal behaviors or shared with mood disorders and other conditions is not easily achievable. The best-described pathway in suicidal ideation, suicidal behavior, and suicide are the stress response pathway mediated by the hypothalamic-pituitary-adrenal axis (HPA) axis. Disruption of the HPA axis with altered glucocorticoid activity may increase the risk for suicidal behavior. Postmortem brain studies of individuals who died by suicide have shown significant alterations in the prefrontal cortex and hippocampus, areas associated with emotion regulation and executive functioning. These studies also revealed alterations in 5-HT2A binding, decreased activity of brain-derived neurotrophic factor (BDNF) both neuroprotective factors, and dysregulated glutaminergic signaling and gamma-aminobutyric acid (GABA) dysfunction. Several studies have shown an association between inflammation, depression, and suicide generating increasing focus on the microbiome. Neuroimaging studies of depressed people who attempted suicide found alterations of attention and salience networks in that cohort compared with depressed people who did not attempt suicide. In addition, cognitive bias toward suicidal preoccupation, risky decision making, and findings consistent with the characteristics of impulsive aggression, risk taking, and emotion dysregulation have been observed with adolescents who attempted suicide. Neurocognitive testing has demonstrated that adolescents who attempt suicide make impulsive decisions and show a failure to learn from consequences.

Psychological Factors. Personality traits that have been strongly associated with suicidality in adolescents are aggressive–impulsive traits, which are worsened in the presence of depression. Anxiety is weakly linked to suicidality, likely contributing by interacting with individual factors in vulnerable individuals. The aggressive–impulsive behavioral profile has also been found to distinguish those who attempt suicide from psychiatric controls and is more likely to be associated with familial transmission. Also, in studies examining the transmission of suicidal behavior, the transmission of aggressive–impulsive behaviors partially explained the familial aggregation of suicidal behavior. Most adolescent suicide attempts are impulsive, so it is not surprising that impulsivity correlates strongly with suicidality. Other psychological factors include impairments in problem solving, memory bias, and ruminations.

Disordered emotion regulation, characterized as having a lower threshold for dysregulation (prolonged, severe responses to provocation), has been shown to be associated with suicidal behaviors.

Hopelessness is an important factor that has been associated with suicidal behavior, and among depressed youth signals an increased risk for suicide. A negative cognitive style has been associated with suicide attempts independent of depression. There is also evidence that those who attempt suicide are more likely to show a self-identification with death and suicidal behavior and that this self-identification is a robust predictor of future suicidal behaviors.

Familial and Environmental Factors. Although suicide and suicidal behaviors have been known to cluster in families for some time, recent data have begun to clarify these relationships. Parents of suicidal adolescents have higher rates of psychiatric disorders, including depression, substance use, and impulsive aggression. Suicide and attempted suicide have been found to be increased in families in which a parent has attempted or died by suicide even after controlling for parental psychopathology, suggesting that suicidal behaviors may be transmitted independently. Adoption, twin, and national registry studies suggest a heritability of 30% to 50%, while other studies suggest lower rates after controlling for psychopathology. The children of people who attempt suicide show a fivefold greater risk of suicide attempts than population averages, with suicide risk being doubled for individuals who lost a parent to suicide compared with those who lost a parent abruptly to other causes. It is important to note, however, that fewer than 5% of those who die by suicide have a parental history.

While genetic etiologies for suicide and suicidal behaviors have been sought for many years, no single gene or group of genes has been identified as responsible for suicide. Because suicide deaths are rare, genome-wide association studies (GWAS) have been used to investigate suicide attempts, with few studies including suicide as a primary phenotype. While some studies have found associations between genes identified in GWAS and suicidality, these studies have not met reproducibility and significance testing and have produced inconsistent results. Other proposed mechanisms for genetic transmission suggest familial transmission of impulsive aggression and mental disorders that do not include suicidal behaviors. Higher levels of 5-HT receptors have been found in postmortem studies of adults and youth who have died by suicide, suggesting that this may be a marker for suicide death.

The familial transmission of suicide and suicidal behavior is likely explained by familial environmental factors as well as genetic transmission. Environmental factors known to be associated with suicidal behavior are exposure to suicidal behavior within the family. Parent–child conflict increases risk for suicide, especially among younger youth and historically marginalized youth populations. Parental loss, in youth who lose their parent prior to the age of 12, is associated with an increased risk of suicide in offspring.

The most clinically significant familial–environmental risk factor for suicide and suicidal behavior is maltreatment. Physical and sexual abuse and emotional and physical neglect have been strongly associated with suicidal behaviors, even after controlling for parental psychiatric disorders. The strongest associations with suicidality have been reported for sexual abuse. The population-attributable risk of sexual abuse for suicidal behavior has been estimated as slightly less than 20%, suggesting that one in five suicide attempts could be eliminated with the elimination of sexual abuse. In reality, the relationship between abuse and suicidal behavior is more complicated because abuse occurs within the social context of socioeconomic strain, family disruption, parental history of abuse, parental psychopathology—including depression, substance abuse, and suicidal behavior—and low parental monitoring and supervision. The likelihood of suicidal behavior is greater with greater severity of abuse. Studies also show that lesbian, gay, bisexual, transgender, queer (or questioning sexual orientation), and othesr (LGBTQ+) youth are more likely to have experienced sexual abuse than youth who identify as heterosexual.

Youth who are engaged in bullying as victims or perpetrators are at increased risk for suicidal ideation and behaviors. Longitudinal studies suggest that students who were most likely to experience negative outcomes such as suicidal ideation or behaviors showed signs of depression at the time of victimization, suggesting that the impact of bullying may be mediated by the presence of depression. Students at risk of dropping out of school, as characterized by poor academic performance and attendance, are also at high suicidal risk. Youth who live in impoverished communities characterized by violence, trauma, underresourced schools, and lack of community support are at greater risk for depression and suicidal behavior. Youth who have a strong connection to school, even in the face of other risk factors, experience some protection from suicidal risk. A deviant peer group is associated with an increased risk of suicidal behaviors, and the presence of a prosocial peer group is protective for youth, including those who are at high risk.

Youth who experience legal problems are at increased risk for death by suicide. While legal problems commonly overlap with other problem behaviors, there is clear evidence that youth who have contact with the juvenile justice system are at high risk for suicidal behavior and death by suicide. Several studies have demonstrated high rates of suicidal ideation and suicide death in juvenile detention facilities. Studies also show high rates of mental illnesses with some studies finding that 75% of youth who died by suicide while in detention had a history of mental illness, 71% had had a history of prior suicidal ideation, and 45% had prior suicide attempts.

Emerging data for Black and Latina/o/x youth found that perceived racial discrimination is associated with increased suicidal ideation and that positive racial identity is protective against suicidal behaviors in Black and Latina/o/x youth.

Youth who identify as gender nonbinary, as well as LGBTQ+ have consistently been at a higher risk for suicide compared to their non-LGBTQ+ peers. LGBTQ+ identity has been identified as a risk factor for suicidal behavior in youth. This association is especially true for those sexual minority youth who are also racial and ethnic minorities. Suicide risks for these youth are in part mediated by family and peer rejection. LGBTQ+ youth who experience significant family rejection has been found to be nine times more likely to make a suicide attempt compared to those with no rejection. This same population of youth appears to be significantly more likely than other youth to experience harassment and bully victimization, and being bullied due to perceived sexual orientation has been found to have a greater impact on the quality of life for youth than bullying victimization for other reasons. Youth who identify as LGBTQ+ at early ages or with cross-gender appearance or behaviors are more likely to experience bullying. Nonetheless, depression and hopelessness remain stronger predictors of suicidal behavior for sexual minority youth than bullying.

Exposure to Suicidal Behavior. Adolescents are vulnerable to the impacts of suicide contagion that occurs when there has been media coverage of suicide. Studies of media publicity of suicide consistently find that rates of suicide and suicidal behaviors increase, especially among younger adolescents. Research further finds that the increased rates of suicide are proportional to the amount of publicity, the notoriety of the individual who died, details regarding the manner of death, and the extent to which the death was romanticized when compared to presenting the information regarding the death highlighting the relationship between psychiatric illness and the outcome of death by suicide.

Contagion effects can also occur with suicides that occur within the family, community, or among friends. Several studies show increased suicidal thoughts or behaviors following the suicide of a friend or acquaintance. They also found that youth with mental health problems have a greater risk for suicidal behaviors following exposure.

Suicide Risk Assessment

The assessment of suicide risk must consider the spectrum of risky behaviors that impair an adolescent’s health. Suicidal behavior most commonly occurs within the context of risky behaviors such as NSSI, binge drinking, substance use, disordered eating, carrying a weapon, and unprotected intercourse. A plethora of studies have shown that common risk factors for these behaviors include impulsivity and poor problem-solving skills, a weak connection between parent and child and/or a conflictual parent–child relationship, low degree of parental supervision and monitoring, poor connection between child and school, and association with deviant peers.

The assessment must also consider protective factors that are present in the adolescent and the family including family connectedness and support, school connectedness and school safety, intelligence and good social–behavioral and communication skills, and reduced access to lethal means.

Suicide is a rare event, and studies of suicide generally investigate nonfatal suicidal behavior including suicidal ideation and attempts. Research has identified factors that are precipitating and strongly associated with future suicidal behavior. The strongest predictor of future suicidal behavior and suicide death is current and recent past suicidal behaviors. The greater the severity of the previous suicidal ideation (e.g., with plan) and suicidal behavior (e.g., lethality of attempt), the greater the likelihood of another suicide attempt. In one study examining individuals who died by suicide, 40% had previously attempted suicide.

The risk for suicide, suicidal behavior, and suicidal ideation are assessed across five domains: intensity, lethality, precipitants, motivation, and availability of lethal agents.

Intensity and intent: Suicidal ideation is assessed both with respect to intensity (how frequent and how compelling) and severity from passive ideation to active planning. Suicidal intensity refers to the degree to which the individual with suicidal ideation intends to carry out a suicidal act; when the individual has made an attempt, intent refers to the degree to which the individual wished to die. The worst-ever level of suicidal ideation is also predictive of future suicidal attempts and suicide deaths. Suicidal ideation is a strong predictor of future suicidal behavior.

Suicidal intent is assessed using the patient’s self-report and assessment of the actions occurring during the suicidal behaviors including the degree of preparatory behavior, the timing of the event to minimize discovery, and other factors that convey intent. A recent study found NSSI to be a predictor of suicidal behavior associating more strongly with suicide risk than prior suicide attempts. For adolescents who express regret for having survived a suicide attempt, they remain at risk of recurrent and fatal suicidal behavior.

Motivation: Motivation refers to the reason why the individual engaged in suicidal behavior. An understanding of motivation is essential to developing an effective treatment plan. Approximately a third of adolescents who attempt suicide report that they want to escape their intolerable psychological pain and distress, express hostility, or to get the attention and support that was needed.

Precipitant: Parent–child conflict, difficulties in peer and romantic relationships, and disciplinary problems are the most common precipitants for suicidal behavior. In younger children and adolescents, precipitants are most often parent–child conflict, in comparison to older youth who are more likely to report romantic or peer difficulties.

Availability of Lethal Agents. In the United States, approximately 45% of youth suicides involve firearms, although suicide deaths involving firearms are declining, those involving hanging or strangulation are increasing. Studies show an association between the availability of guns, especially handguns, and risk for suicide. Handguns are 2 to 10 times more likely to be found in the homes of suicide victims than in those of psychiatrically matched controls. The only risk factor that differentiated youth who died by suicide and living control youth, all of whom had no clear evidence of psychiatric disorder, is the availability of a loaded gun in the homes of those who died by suicide. Restriction of access to guns has been associated with regional and national declines in firearm-related and overall suicide rates. Restricting the availability of prescription medications and dispensing a limited number of pills/tablets/capsules wrapped in blister packs to prevent easy ingestion of large quantities have been shown to favorably affect rates of fatality by ingestion of prescription drugs. Other prevention methods include placing barriers on bridges to prevent jumping. In countries where poisoning is common with pesticides, using a lock box to contain the poisons has been effective in reducing pesticide ingestions.

Assessment of Imminent Suicidal Risk. Clinicians are frequently asked to assess patients to determine their level of suicide risk in the emergency setting.

Often, the clinician is called upon to assess a patient for imminent risk for suicide, in the emergency or urgent setting—does this person really want to kill themself in the next 24 to 48 hours? When assessing acute risk for suicide, factors that must be assessed urgently are those that increase distress and decrease restraint to determine levels of risk, and availability of lethal means. High levels of distress are found with anxiety and panic, hopelessness, mixed and irritable bipolar states, agitation, psychosis, emotion dysregulation, and insomnia. Insomnia is especially potent as it increases distress and decreases restraint. Factors that decrease restraint include impulsivity, substance use or misuse, a belief that no one cares about the person, or that they are a burden. The most effective interventions target increased restraint and decreased distress, and a plan for safety that will provide resources for recognizing when distress is building and a plan for intervening when risks are increasing, for example, The Safety Planning Intervention.

Treatment

When a child or adolescent is suicidal secondary to an underlying depression, treatment of the depressive episode is important; however, this treatment takes time and a more immediate brief intervention can be used in the interim. In the emergency room setting, targeted safety planning such as the Safety Planning Intervention, which can help patients address suicidal thoughts and feelings and reduce the risk of death by suicide. Safety planning intervention is not the same as a no-suicide contract or a safety contract, rather it is a therapeutic-based approach to helping the suicidal adolescent or child engage in crisis management skills prior to engaging in any sort of self-harm. The Safety Planning Intervention involves identifying antecedents to a suicide attempt, stressors or warning signs that a patient is feeling more dysregulated, coping skills or activities for distraction or avoidance of the negative suicidal thoughts, and creating a list of family friends and trusted adults including providers to whom a child or adolescent can turn before acting on any thoughts of suicide. This type of planning is done in conjunction with the patient, the parent, and a clinician. This particular type of planning works well because it’s not just a relying on a safety contract, which has not proven to be a deterrent to suicide or suicide attempts, rather it is an attempt to help the child or adolescent build skills on which they may rely when they are feeling suicidal, or more importantly when they are feeling negative affect that may lead to suicidal thinking. This is also a plan that is adjustable based on the needs, mindset, and skills that a child or adolescent has developed. Thus, the safety plan can be modified as needed over time. Other brief interventions have included attachment-based family therapy (AFBT) multisystemic therapy (MST), cognitive behavioral therapy for suicide prevention (CBT-SP) especially when paired with parent interaction, and dialectical behavioral therapy (DBT). Both CBT-SP and DBT show evidence for the prevention of future suicide attempts and self-harm in children and adolescents. Several studies suggest that parent involvement or treatment that also addresses substance use can be efficacious in reducing the risk of death by suicide. Although hospitalization often occurs when patients have had a suicide attempt or are at risk for suicide and unable to engage in safety planning, hospitalization in and of itself is not necessarily treatment. It can be the beginning of treatment especially if specific evidence-based therapies to treat depression, underlying anxiety, ADHD, substance use disorders, or other disruptive behavior disorders are incorporated in hospital treatment plans.

While there are no medications specifically targeted to treat suicidality, SSRIs used in the treatment of depression may help prevent suicidal behavior. In children and adolescents with schizophrenia disorders, clozapine has been shown to be protective against death by suicide; however, clozapine is rarely used first line in treating psychotic disorders. In children and adolescents with bipolar disorder, there is evidence that lithium may also have a protective effect; however, with its limited therapeutic range and risk of serious adverse outcomes, including death in overdose, this medication must be administered and monitored closely. In the adult literature, there is evidence for ketamine, or its enantiomer esketamine for treatment-resistant depression, although there are no randomized controlled trials (RCTs) involving children, the literature does suggest in small case-controlled studies and case reports involving children and adolescents that IV ketamine has been used to alleviate depression and suicidal thoughts in selected individuals.

In addition to the brief interventions used in emergency or crisis situations, there are several school-based interventions involving social and emotional learning which show some evidence for suicide prevention. Social and emotional learning with an emphasis on cooperation with others, self-regulation, and maintaining self-control while working toward goals valued by school-aged children, have shown some efficacy in decreasing risk of suicide in youth. Social–emotional learning that focuses on the expression, understanding, and regulation of emotional expression has also shown efficacy in school-aged children and beyond. Examples of these programs include “The Good Behavior Game” and “Promoting Alternative Thinking Strategies program (PATHS).”

For adolescents, there is new research using electronic application (app-based) programs to target suicidal thoughts and behaviors. While these programs or apps are still in the early stages, it may be that targeting adolescents, who spend a significant portion of their day interacting with media through the use of apps, can help reduce suicidal behavior as well as nonsuicidal self-injury. In addition, there is a new computer-assisted technology in the form of the computerized adaptive screen for suicidal youth (CASSY), which although is not a treatment model; it may represent an improved screening for children and adolescents at risk for death by suicide or suicide attempt and therefore allow early intervention to prevent suicide deaths. Given the high rate of suicide among indigenous people, one of the CASSY studies particularly recruited indigenous children and adolescents. There is a pressing need for development of treatment modalities as well as screening modalities to address specific needs of historically marginalized populations, who may not receive the same benefits of current established treatment protocols which often do not include these populations in initial research protocols.

While there is concern about the use of social media and increased suicide rates, there appears to be some literature that suggests moderate social media use can build social bonds which can be protective against suicide. The role of social media in child and adolescent suicide is one that still needs to be more clearly researched and delineated.

Current guidelines regarding the care of the suicidal adolescent include an assessment and psychiatric evaluation to determine the appropriate level of care to safely deliver treatment, the development of a safety plan that can be modified as needed, encouraging family and caregivers to remove or secure lethal means from the home, treatment of any underlying psychiatric disorders, and addressing specific psychosocial issues that may be affecting the child or adolescent. These issues can be associated with adverse childhood experiences, emotional regulation skills, parental mental illness, parental substance use, and parent–child conflict. Ultimately treatment of the suicidal child or adolescent will primarily be that of prevention.

Course and Prognosis

Although suicidal thoughts are strong predictors of suicide attempts, only a third of adolescents who have suicidal thoughts are estimated to act on those thoughts in a suicide attempt. According to a recent review, one factor that most consistently predicts suicide attempts among adolescents with suicidal ideation is related to suicide capability or rather, the degree to which one feels capable of making a suicide attempt. Unfortunately, suicidal behavior is highly recurrent. Approximately 15% to 30% of adolescents will reattempt suicide within 1 year of their original attempt and, in a more recent study, 5% of adolescents made another suicide attempt within 3 months of an initial attempt. Risk for death by suicide is approximately 0.5% to 1% per year. Although factors for adolescents who attempt suicide are multidimensional and heterogeneous, there are factors that discriminate between those who die by suicide compared with those who are treated for suicide ideation or attempts. Factors contributing to death by suicide are male gender, access to a gun, sleep disturbance, mood disorder with comorbid substance abuse or conduct disorder, diagnosis of bipolar disorder, and high suicidal intent. It is important to recognize that suicide is a rare event and that suicide is preventable. Given that suicide is currently the second leading cause of death in adolescents, it is essential that we convey a message of hope when intervening with suicidal adolescents and their families.

FURTHER READINGS

Althoff RR, Crehan ET, He JP, Burstein M, Hudziak JJ, Merikangas KR. Disruptive mood dysregulation disorder at ages 13–18: results from the national comorbidity survey-adolescent supplement. *J Child Adolesc Psychopharmacol*. 2016;26(2):107–113.

Asarnow JR, Mehlum L. Practitioner review: treatment for suicidal and self-harming adolescents—advances in suicide prevention care. *J Child Psychol Psychiatry*. 2019;60(10):1046–1054.

Bridge JA, McBee-Strayer SM, Cannon EA, et al. “Impaired decision making in adolescent suicide attempters. *J Am Acad Child Adolesc Psychiatry*. 2012;51(4):394–403.

Chater AM. Does intentional asphyxiation by strangulation have addictive properties? *Addiction*. 2021;116(4):718–724.

Cipriano A, Cella S, Cotrufo P. Nonsuicidal self-injury: a systematic review. *Front Psychol*. 2017;8:1946.

Colasanto M, Madigan S, Korczak DJ. Depression and inflammation among children and adolescents: a meta-analysis. *J Affect Disord*. 2020;277:940–948.

Croarkin PE, Elmaadawi AZ, Aaronson ST, et al. Left prefrontal transcranial magnetic stimulation for treatment-resistant depression in adolescents: a double-blind, randomized, sham-controlled trial. *Neuropsychopharmacology*. 2021;46(2):462–469.

Cuijpers P, Karyotaki E, Eckshtain D, et al. Psychotherapy for depression across different age groups: a systematic review and meta-analysis. *JAMA Psychiatry*. 2020;77(7):694–702.

Dougherty LR, Smith VC, Bufferd SJ, Kessel EM, Carlson GA, Klein DN. Disruptive mood dysregulation disorder at the age of 6 years and clinical and functional outcomes 3 years later. *Psychol Med*. 2016;46(5):1103–1114.

Dwyer JB, Landeros-Weisenberger A, Johnson JA, et al. Efficacy of intravenous ketamine in adolescent treatment-resistant depression: a randomized midazolam-controlled trial. *Am J Psychiatry*. 2021;178(4):352–362.

Evans DL, Foa EB, Gur RE, et al. Defining youth suicide. In: Evans DL, Foa EB, Gur RE, et al., eds. Treating and Preventing Adolescent Mental Health Disorders: What We Know and What We Don’t Know. A Research Agenda for Improving the Mental Health of Our Youth. Oxford University Press. Oxford Medicine Online. 484–496. Accessed 27 July 2021 https://oxfordmedicine.com/view/10.1093/9780195173642.001.0001/med-9780195173642-chapter-22

Finkelstein Y, Macdonald EM, Hollands S, et al; Canadian Drug Safety and Effectiveness Research Network (CDSERN). Risk of suicide following deliberate self-poisoning.

JAMA Psychiatry. 2015;72(6):570–575.

Gibbons RD, Brown CH, Hur K, Davis J, Mann JJ. Suicidal thoughts and behavior with antidepressant treatment: reanalysis of the randomized placebo-controlled studies of fluoxetine and venlafaxine. *Arch Gen Psychiatry*. 2012;69(6):580–587.

King CA, Brent D, Grupp-Phelan J, et al; Pediatric Emergency Care Applied Research Network. Prospective development and validation of the computerized adaptive screen for suicidal youth. *JAMA Psychiatry*. 2021;78(5):540–549.

Lim KS, Wong CH, McIntyre RS, et al. Global lifetime and 12-month prevalence of suicidal behavior, deliberate self-harm and non-suicidal self-injury in children and adolescents between 1989 and 2018: a meta-analysis. *Int J Environ Res Public Health*. 2019;16(22):4581.

Lindsey MA, Sheftall AH, Xiao Y, Joe S. Trends of suicidal behaviors among high school students in the United States: 1991–2017. *Pediatrics*. 2019;144(5):e20191187.

Lu CY, Zhang F, Lakoma MD, et al. Changes in antidepressant use by young people and suicidal behavior after FDA warnings and media coverage: quasi-experimental study. *BMJ*. 2014;348:g3596.

Lu CY, Penfold RB, Wallace J, Lupton C, Libby AM, Soumerai SB. Increases in suicide deaths among adolescents and young adults following US Food and Drug Administration antidepressant boxed warnings and declines in depression care. *Psychiatr Res Clin Pract*. 2020;2(2):43–52.

Melvin GA, Dudley AL, Gordon MS, Ford S, Taffe J, Tonge BJ. What happens to depressed adolescents? A follow-up study into early adulthood. *J Affect Disord*. 2013;151(1):298–305.

Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self-help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. *BMJ*. 2012;344:e2598.

Ng WYK, Wagner KD. Depression in special populations. *Child Adolesc Psychiatr Clin N Am*. 2019;28(3):xv–xvi.

Nock MK, Green JG, Hwang I, et al. Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: results from the National Comorbidity Survey Replication Adolescent Supplement. *JAMA Psychiatry*. 2013;70(3):300–310.

Ougrin D, Tranah T, Stahl D, Moran P, Asarnow JR. Therapeutic interventions for suicide attempts and self-harm in adolescents: systematic review and meta-analysis. *J Am Acad Child Adolesc Psychiatry*. 2015;54(2):97–107.e2.

Pina AA, Polo AJ, Huey SJ. Evidence-based psychosocial interventions for ethnic minority youth: the 10-year update. *J Clin Child Adolesc Psychol*. 2019;48(2):179–202.

Perepletchikova F, Nathanson D, Axelrod SR, et al. randomized clinical trial of dialectical behavior therapy for preadolescent children with disruptive mood dysregulation disorder: feasibility and outcomes. *J Am Acad Child Adolesc Psychiatry*. 2017;56(10):832–840.

“Ring the Alarm” Congressional Black Caucus Emergency Task Force on Black Youth Suicide and Mental Health. Rep Bonnie Watson-Coleman. 2019. https://nbjc.org/wp-content/uploads/2020/11/Ring-the-Alarm-TASKFORCE-REPORT.pdf

Towbin K, Vidal-Ribas P, Brotman MA, et al. A double-blind randomized placebo-controlled trial of citalopram adjunctive to stimulant medication in youth with chronic severe irritability. *J Am Acad Child Adolesc Psychiatry*. 2020;59(3):350–361.

Turecki G, Gunnell D, Brent DA, et al. Suicide and suicide risk. *Nat Rev Dis Primers*. 2019;5(1):74.

Wagner KD, Pliszka SR. Treatment of childhood and adolescent disorders. In: Schatzberg AF, Nemeroff CB, eds. *Essentials of Clinical Psychopharmacology*. 3rd ed. American Psychiatric Publishing; 2013. 693–750.

Wasserman D, Hoven CW, Wasserman C, et al. School-based suicide prevention programmes: the SEYLE cluster-randomised, controlled trial. *Lancet*. 2015;385(9977):1536–1544.

Zhou X, Teng T, Zhang Y, et al. Comparative efficacy and acceptability of antidepressants, psychotherapies, and their combination for acute treatment of children and adolescents with depressive disorder: a systematic review and network meta-analysis. *Lancet Psychiatry*. 2020;7(7):581–601.

50.2 Early-Onset Bipolar Disorder

GABRIELLE A. CARLSON, M.D., MANPREET KAUR SINGH, M.D., M.S., DANELLA HAFEMAN, M.D., PH.D., AND CAROLY SUE PATAKI, M.D.

Recognition of bipolar disorder in young people has had ups and downs over the past 30 years. Some of the controversy stems from inconsistent use of terms like “childhood,” which might or might not include adolescents. We use early-onset bipolar disorder to refer to onsets before age 18; very early onset to refer to onsets through age 12.

Bipolar disorder itself has a number of meanings, but we refer to the spectrum of bipolar I (at least one lifetime episode of mania with or without depression), bipolar II (at least one lifetime major depression and an episode of hypomania), cyclothymia (hypomanic and subthreshold depressive episodes), and other specified bipolar disorder (usually symptoms of mania or hypomania that don’t meet criteria for full-fledged diagnosis. Finally, onset is a tricky concept. Unless a person’s first episode of bipolar disorder is a clear, uncontroversial episode of mania or hypomania, “onset” is something defined in retrospect. If a person has a manic episode at age 24 and remembers being depressed in sixth grade, the assumption is that their onset was in sixth grade. However, in sixth grade, no one at the time knew a manic episode would occur at age 24, and thus bipolar disorder was starting. Similarly, if a child is grumpy, or volatile, or has sleep problems, and subsequently develops a manic episode, in retrospect, those symptoms are possibly prodromal. They are, however, common symptoms and usually do not lead to bipolar disorder. Even in the context of having a positive history of bipolar I disorder (BP) in a first-degree relative, psychopathology, in general, is much more common than BP or even bipolar spectrum disorder.

The frequency with which very early-onset BP occurs is controversial compared to onsets after puberty when the condition more closely resembles classic adult BP. It is especially challenging to identify a manic episode in a young child, partly due to the distinguishing mania from severe attention deficit hyperactivity disorder (ADHD) with comorbid oppositional defiant disorder. This confusion evolved from a series of diagnostic criteria changes in earlier versions of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), in which *emotion dysregulation* was removed as a core feature of ADHD, and the delineation of an *episode* of mania was not well defined. The subsequent diagnosis of chronically irritable children with severe outbursts led to increased use of neuroleptic medication, prompting the DSM-5 committee to formulate two solutions. First, mania was redefined as: "an abnormal and persistently elevated, expansive or irritable mood" *and* "persistently increased goal-directed activity or energy," which are present that must vary to a *significant degree* and represent an *observable change* from usual behavior. Furthermore, in an effort to provide a diagnosis for youth who were chronically (vs. episodically) irritable, explosive, and equally disabled, but without other symptoms of mania, a new diagnosis, "disruptive mood dysregulation disorder" was added to DSM-5 to identify children whose mood is " … persistently irritable or angry most of the day, nearly every day, and is observable to others" (DSM-5) with frequent, impairing behavioral outbursts.

HISTORY

Recognition of manic depression in child and adolescent psychiatry began with the dissemination of Kraepelin's monograph *Manic Depressive Insanity and Paranoia* after 1921. Child psychiatrists searched for the same condition in their patients but found mostly adolescent depression, which was highly familial. Mania was very rare, even in teens. Brief attempts to broaden the concept of manic depression for younger children were discouraged by a literature review conducted by Anthony and Scott in 1960 that used very stringent criteria to identify manic depression in youth. They found only 3 of 60 cases reported prior to puberty. However, the tension between restricting the diagnosis to classical presentations versus altering the criteria to fit children continues to the present.

In the 1960s, evidence of lithium's efficacy for treating adults with manic depression led to an interest in identifying lithium-responsive clinical profiles in children. Early studies suggested that lithium response was most promising in early-onset cases resembling classic manic depression, whereas there was no support for the use of lithium in treating children with broader behavioral dysregulation. When standard symptom criteria became available in the 1970s for describing mania more precisely in adults, Weinberg and Brumback, two pediatric neurologists, adapted these criteria for use with children. The concept of a "manic-depressive variant syndrome of childhood" was subsequently introduced in 1979 and characterized children between 6 and 16 years of age who were hyperreactive to seemingly minor events and were also hyperactive, had disruptions in interpersonal relationships, and often had a family history of manic depression.

By the 1980s, there was increasing acceptance of the diagnosis of manic depression or bipolar disorder, as it became known to the field, in youth. Research efforts initially focused on describing the phenomenology of pediatric bipolar disorder and established standard assessment techniques. In more recent years, there has been a surge of genetic, high-risk, and treatment studies, as well as investigations of neurobiologic, behavioral, and physiologic markers.

As noted earlier, the diagnosis (vs. the actual condition) of bipolar disorder among youth in the community has fluctuated over the last 25 years. A study of outpatient physician visits in the United States found that juvenile-onset cases of bipolar disorder comprised 42% of outpatient mental health–related visits in 2003, an increase from 6.67% of cases in the early 1990s. Another study, using data from the National Hospital Discharge Survey, revealed increases in the rate of bipolar disorder diagnoses among psychiatrically hospitalized children from 10% in 1996 to 34.11% in 2004, and from 10.24% to 25.86% among psychiatrically hospitalized adolescents during that same period. The increased rate of bipolar disorder appeared to reflect a shift in diagnostic practices such that childhood mood lability and aggression, which previously might have been attributed to conduct problems, ADHD, or impaired family relationships was reinterpreted as bipolar disorder. The consensus currently is that severe irritability is transdiagnostic, and while some of the children may, in fact, have a bipolar spectrum disorder, that is no longer the only diagnostic home for such children.

EPIDEMIOLOGY

Prevalence rates of bipolar disorder vary depending on whether bipolar I, II, or spectrum conditions are included, the age studied, the definitions and methods used to make the diagnosis, and the type of sample interviewed. Taking wide variability into consideration, an updated meta-analysis calculated a pooled rate of 0.6% (95% confidence interval [CI] 0.3% to 1.2%) of bipolar I in adolescents (children younger than age 9 have not been studied in community samples) (Van Meter et al., 2019). The current prevalence of bipolar spectrum disorder is much higher (3.9% (95% CI 2.6% to 5.8%) because of the use of broad bipolar criteria. Lifetime rates are also higher. While tightly defined bipolar spectrum disorder may have an up to 40% rate of conversion to BPI or BPII, manic symptoms alone are less often associated with such a conversion, as seen in both the Longitudinal Assessment of Manic Symptom Study (LAMS), where mania rarely developed and the Oregon Adolescent Depression Study, where it never developed.

ETIOLOGY, MECHANISMS, AND RISK FACTORS

Bipolar disorder is a serious, recurrent, and complex psychiatric condition. Associated conditions include a uniquely high risk of suicide attempts, co-occurring anxiety and attention deficit with hyperactivity, family dysfunction, cognitive inflexibility, and substance use, among others. While mania is the defining feature of BP, episodes of depression often precede the onset of mania and may predominate a bipolar course of illness. Despite advances in accurately diagnosing bipolar disorder in young people, as well as improvements in the pharmacologic management of this disorder, problems related to symptom relapse, medication nonadherence, and co-occurring conditions continue to affect its long-term morbidity and mortality. The adverse impact of bipolar disorder on families and society is clear, but research toward elucidating the pathophysiologic basis of early onset of this disorder has been incremental.

Nevertheless, bipolar disorder is now well established to be brain based, with unique structural, functional, and neurochemical properties in the brain that likely underlie its pathogenesis. Early-onset bipolar disorder is also characterized by dysfunction in the regulation of emotion and cognition. Disturbances in core emotional and cognitive functions aberrant in the disorder provide a basis for understanding the origins of symptom manifestations and the pathophysiology underlying bipolar disorder. In addition, investigating the neural sources of these core features using neuroimaging may be more informative than basing paradigms on nonspecific or transdiagnostic symptoms, particularly when controversies regarding the heterogeneity of symptom presentations in youth inevitably emerge.

Altered interactions among brain networks subserving emotion and cognition must also be developmentally contextualized, with appreciation that key neural features present in adults with bipolar disorder differ from those in youth with bipolar disorder. Traditionally, subcortical gray matter development appeared to precede significant prefrontal development and pruning with a shift in functional dependence from subcortical to prefrontal structures around puberty. Contemporary research suggests there are significant increases in the expression of key presynaptic cell adhesion molecules in the prefrontal cortex prenatally and during the first 3 years of life in individuals at risk for bipolar and other neurodevelopmental disorders. This opens the possibility of a new critical sensitive period for the development of bipolar disorder before puberty when prefrontal–subcortical functional shifts have traditionally been thought to herald the onset of bipolar disorder in adolescence. Early scaffolding of such inborn errors may explain prodromal symptoms that precede mania onset, that can be severe, functionally impairing, and protracted before full threshold symptoms are declared. A family history of bipolar disorder is an independent risk factor for early bipolar development.

Genetic, inflammatory, and environmental stress diatheses have been proposed as etiologic factors for bipolar onset or recurrence. These factors might express themselves phenotypically either through dysregulation of emotion acutely or gradually over time or during a bipolar prodrome. They may also manifest as hyperarousal, hypersensitivity, difficulty with affect decoding or labeling, or as a problem with “repair” of maladaptive emotional responses. With time, chronic illness-related physiologic modifications from emotion dysregulation may incite more stress and inflammatory states, which then may have a direct neurotoxic effect, thereby contributing to the onset and recurrence of mood episodes. Although akin to theories of kindling, it may alternatively be the case that forces that compound episodes are dynamically being counteracted by microlevel forces like remodeling or adaptive neuroplasticity and macrolevel forces like cycles of treatment exposure and adherence. These homeostatic forces may lead to episodic (a feature of illness) rather than continuous expression of emotion dysregulation or extended periods of euthymia (a feature of treatment). Indeed, the cumulative effects of familial vulnerability, inhibited temperament phenotypes, environmental stress, inflammation, and disturbances in emotion regulation may converge in the expression of precursor subthreshold forms of bipolar disorder, persistent hypomania, depression, or mixed mood states, episodic irritability, rapid mood fluctuation, and low social functioning. These pathologic processes may elevate the liability for progression to bipolar disorder and to further neurobiologic sequelae.

BRAIN STRUCTURE AND FUNCTION

Untreated or repeated mood episodes may progress into neurobiologic sequelae that could manifest as structural brain changes that are preceded by neurofunctional impairment. Data demonstrating macroscopic structural and functional changes in selective brain regions in bipolar disorder imply underlying cellular and molecular dysfunction. Aberrant prefrontal–subcortical connectivity may result in failure of normal prefrontal regulation of overreactive subcortical processes associated with emotion dysregulation. This failure may, in turn, result in the appearance of disease-specific anatomic abnormalities that may be unique to youth with bipolar disorder, such as reductions in amygdala volumes, thereby predisposing youth to downstream vulnerabilities in the development of coordinated brain networks important for adaptive emotion processing, including the ventrolateral prefrontal cortex (VLPFC), the anterior cingulate cortex (ACC), thalamus, striatum, hippocampus, and the cerebellar vermis. Abnormal emotion regulation and cognition in bipolar disorder may thus arise from aberrant reciprocal connections between the amygdala and dorsal and ventral prefrontal areas as described by others in mood disorders in general. Comparisons of pediatric and adult bipolar studies of brain function responses to stress add a developmental component to this model. These data suggest that limbic hyperactivity (e.g., in the amygdala) may occur early in childhood and even prior to illness onset, eventually becoming associated with decreased prefrontal activation and control over amygdala activity expected during typical brain development (e.g., in the VLPFC and dorsolateral prefrontal cortex [DLPFC]). As bipolar disorder progresses, prefrontal abnormalities become more important, and neurodegenerative sequelae of the illness develop that appear akin to molecular, cognitive, and clinical manifestations of accelerated aging and dementia. In this context, the unique regulatory role of the prefrontal cortex may make bipolar disorder exclusive to humans and what challenges modeling of the disorder in other animal species.

Pathologic reactions to stress can be detected acutely by clinical symptoms, and over time, by examining brain structural and functional changes, which interrupt healthy development of networks of brain structures that are important for mood regulation. Disrupted relationships among dorsal regulatory areas, including the DLPFC and dorsal anterior cingulate cortex (dACC), and ventral areas, including the amygdala and subgenual anterior cingulate cortex (sgACC), may lead to mood dysregulation or cognitive impairments. Without adequate intervention, children may eventually develop a full manic episode displaying further functional connectivity abnormalities between nodes in mood regulation networks; for example, inverse functional connectivity between amygdala and DLPFC is lost, or task-dependent excessive medial prefrontal cortex (MPFC) activation is observed. With repeated mood episodes into adulthood, further neurodegeneration of prefrontal areas might occur, leading to rapid cycling, treatment resistance, and more severe episodes, causing increased morbidity and mortality. Progression over the lifespan is variable due to individual differences in intrinsic compensatory adaptations and the buffering effects of individual- and system-level interventions. Longitudinal studies that evaluate risk and resilience factors may clarify the episodic nature of bipolar disorder and enhance prediction about onset and recurrence.

Studies in youth at risk for bipolar disorder indicate that most structural and neurochemical differences do not occur prior to the onset of syndromal mania, suggesting the possibility of preventing disease onset and progression if an underlying mechanism can be identified. Further, a relative lack of morphometric abnormalities in symptomatic youth at high risk for bipolar disorder points to the probability that structural effects are a consequence of illness. Studies in healthy offspring of parents with bipolar disorder suggest clues for compensatory or resilience factors that may be important in preventing the onset of illness, including enlargement and increased activation and connectivity in prefrontal regions. Therefore, magnetic resonance imaging (MRI) studies in youth with and at risk for bipolar disorder are identifying brain findings that are now showing demonstrable improvement with early intervention to prevent or ameliorate the natural course of bipolar disorder.

Thus, convergence in network-based abnormalities preceding bipolar onset supports the hypothesis that early neural dysfunction likely underlies the pathophysiology of bipolar disorder. Further, available studies in adults with a prolonged illness have been limited by confounding variables and discontinuous follow-up to make it challenging to elucidate the neurodevelopmental progression of bipolar disorder and its pathogenesis starting from childhood. Many first-generation neuroimaging studies in pediatric bipolar disorder cautiously interpreted early findings with concerns about limitations in small sizes. Although type II error must not be ruled out, and independent replication can strengthen a given finding, indiscriminately large sample sizes as the cost of quality assessments and deep phenotyping are unlikely to facilitate progress in our understanding of early-onset bipolar disorder. To make progress in our field, it is crucial that we find systematic ways of integrating neuroimaging data across multiple modalities, carefully phenotyping youth in a variety of clinical settings, and leveraging multivariate statistical approaches to be able to conduct meaningful next-generation longitudinal neuroimaging studies in at-risk and fully bipolar populations.

ASSESSMENT

The American Academy of Child and Adolescent Psychiatry (AACAP) Practice Parameter advocates (1) screening patients for mania and bipolar depression, both current and past; (2) applying DSM criteria (currently DSM-5); (3) using the designation “bipolar NOS” for children with manic symptoms with a duration of fewer than 4 days (called “other specified bipolar disorder in DSM-5); (4) examining for comorbidities; and (5) using extreme caution when diagnosing young children.

Making an accurate diagnosis of bipolar disorder in children requires systematic and thoughtful evaluation procedures that go well beyond the simple application of DSM-5 criteria for mania and depression. For instance, defining a mood episode in children, especially younger children, is complicated by the fact that a baseline or “usual self,” developmentally, is a moving target. Thus, a previously docile infant who enters the “terrible two’s” may be developmentally normative and yet approximate criteria for mania if this is defined as a persistent mood state different from the child’s “usual self.” Using the above framework, it is also unclear how long to construe a remission if most of an individual’s childhood has been considered an episode. Children stressed by school may have what appears to be mood “episodes” which subside over the summer. In adult studies, in which a typical manic episode lasts for 4 to 8 months, a euthymic period of 8 weeks constitutes remission. Does that mean that a child stressed with “manic symptoms” at school and is better over the summer has had an episode?

Efforts have been made to adapt adult definitions of the cardinal symptoms of mania for young children, taking into account developmental differences. However, children may have difficulties with the concept of elation, or increased self-esteem, which requires an ability to self-assess, a skill absent in young children. Silliness and bravado, which some clinicians construe as childhood manifestations of euphoria and grandiosity, may also be seen in children with ADHD, autism spectrum disorders (ASDs), or children with other social skills deficits. In other words, simply applying criteria using structured interviews without understanding development and context is likely to lead to either the kinds of false positives that artificially inflate rates of bipolar disorder or to cross-site discrepancies in rates of bipolar disorder.

A careful ascertainment of bipolar disorder in youth should start with screening measures completed by multiple informants (i.e., parent, teacher, older child, or adolescent). These are useful in alerting the clinician to current symptoms of mania, depression, and other possible comorbidities. Interviews with caregiver(s) and child and review of other diagnostic information are then required.

Parent rating forms about the child, including the Child Mania Rating Scale, the General Behavior Inventory, and Parent Young Mania Rating

Scale, are good for both screening and to ascertain severity once the diagnosis has been made.

It is most useful to interview caregivers first because they are likely to provide the most reliable historical and contextual information, and this material can be presented within the child interview to address any discrepant responses. Caregiver interviews should include an overview of current symptoms, developmental history, past history, and longitudinal course of depressive and manic (and comorbid) symptoms.

The child interview is critical for eliciting important information regarding manic (or depressive) symptoms the child is experiencing like euphoria, or depression, or trouble concentrating. It is necessary to interview the child to detect the presence of symptoms and features of other conditions that may be confused or co-occur with mania, including ASD, language or thought disorder, psychosis, anxiety, suicidal behavior, physical/sexual abuse, and substance use. Parents may or may not be able to provide information on these symptoms and situations.

Collateral information, such as teacher input, is helpful in determining whether manic-like symptoms are occurring across settings. Although there are the usual discrepancies between parent and teacher reports, home and school observation can be useful. For instance, a child who is quiet and well behaved for 6 hours a day in school and then dysregulated at home is unlikely to be experiencing a manic episode. Outside assessments, including neuropsychological, occupational therapy, psychoeducational, speech and language assessments, and medical records, can also be extremely helpful in identifying developmental, cognitive, and medical factors that may contribute to emotional dysregulation in other conditions as well as those that complicate episodes of mania and depression.

Structured and semistructured diagnostic interviews are used in research settings to elicit symptoms of bipolar disorder and other childhood conditions. However, they may also be useful in clinical settings to confirm a diagnosis. The Schedule for Affective Disorders and Schizophrenia in Children (K-SADS) is the most commonly used interview and has a number of versions, including the K-SADS-PL (present and lifetime version), a version developed at Washington University, the WASH-U-K-SADS, and now a computerized version. In our opinion, the structured interview is best used as a review of symptoms rather than as a history-gathering tool. Clinical judgment is required to determine whether a respondent is under- or overreporting symptoms, misinterpreting developmentally appropriate behaviors, or misunderstanding questions about clinical behaviors. To this end, it is critical that clinicians probe beyond affirmative responses to elicit examples of specific behaviors, as well as the context in which behaviors are occurring.

Several rating scales have been developed to measure the severity of mania once that condition has been diagnosed. One is an interview measure, the Young Mania Rating Scale (Y-MRS), which is used in both adult and youth medication trials.

Family history of mood disorder, bipolar and unipolar, and comorbid psychiatric disorders including oppositional defiant disorder, ADHD, learning disorders, and ASDs should be assessed as part of the parent interview. If there is a positive family history of bipolar disorder, it is important to probe for details regarding clinical presentation, comorbid conditions, and response to treatment in the person who has had bipolar disorder.

Family History

Family history studies estimate that genetic influences explain approximately 60% to 93% of the variance in bipolar disorder. Shared and individual environmental factors account for about one-third and one-fifth of the variance, respectively. Rates of mania in high-risk offspring who have been followed into adulthood occur in 2% to 7%, whereas rates of bipolar spectrum disorder are as high as 20% depending on how broadly the spectrum is defined and how complicated the parent proband is. Mania in high-risk offspring generally starts to emerge in adolescence. There are, however, cross-national differences. Mania onset prior to puberty is not reported in European and Canadian high-risk samples, where it does occur in U.S. samples. The parent probands, however, also differ cross-nationally, with U.S. samples having higher rates of comorbidity, substance abuse, and other psychosocial disadvantages.

The rate of major depression and a wide spectrum of depressive disorders in high-risk children of bipolar parents is higher than the rate of bipolar disorder. That is, major depression, particularly if highly recurrent and remitting, is found more often in families of patients with bipolar disorder and is part of the bipolar diathesis in these families. In fact, depression is the more common mode of onset of bipolar disorder in high-risk samples. Although the estimated heritability of bipolar disorder is very high, bipolar variants can be attributed to gene by environment variables and correlations. For example, environmental risk factors such as adverse childhood experiences (ACEs), including physical and sexual abuse, severe parental neglect, and unstable home environments with poor attachments to primary caregivers, may play a role in the risk of developing the disorders. Exposure to parental mental illness, particularly a parent with bipolar spectrum disorder, or harsh and negative parenting styles, exposure, and poor social support may interact with a genetic vulnerability. Evidence suggests that temperament, specifically high emotionality, may be a risk factor for and of mood disorders in high-risk individuals. Irritability in mothers may also be a risk factor.

Bipolar disorder with other comorbidities in parents is an important complication that further contributes to psychopathology in offspring. In utero exposure to alcohol and illicit drugs has been associated with a higher risk for an earlier onset of bipolar disorder in vulnerable individuals. Family history differs in bipolar individuals who are lithium responsive and lithium nonresponsive. Relatives of lithium-responsive patients have a higher rate of episodic, remitting bipolar and unipolar disorders and not of chronic psychiatric illnesses such as schizophrenia, schizoaffective disorder, or substance use disorders, which are found in relatives of patients with the broader subtype.

Comorbidities

Children and adolescents meeting the criteria for BP nearly always present with co-occurring conditions. Symptoms of mania may overlap with those of concurrent ADHD and oppositional defiant/conduct disorder, depression, and anxiety, precisely the conditions with which co-occurrence is highest. ADHD is the most common comorbid diagnosis among children with bipolar disorder, with published rates of 70% to 95% among preadolescent youth and 50% among adolescents. The frequency of comorbid diagnoses is influenced by how different interviews manage symptom overlap between mania and the comorbid disorder. If the interview or investigator deems that a given symptom counts toward all disorders, rates of comorbid ADHD and ODD will be higher than if symptoms counted toward mania are only those presenting an intensity significantly beyond a usual baseline. By counting any overlapping symptoms toward all diagnoses, the Treatment of Early Mania Study found that 92.8% of the youth with BP had comorbid ADHD in their index episode.

The LAMS study identified ADHD, bipolar spectrum disorders, and comorbidity in a multisite sample of 707 children who were rated by their parents as having elevated symptoms of mania. The LAMS study found that 76.2% of their sample met criteria for ADHD, 22.9% met criteria for bipolar spectrum disorders, 16.5% had both disorders, and 17.5% of the sample had neither disorder. The majority of children who met the criteria for BP (69%) also had ADHD. The high rate of comorbid ADHD in the entire sample was best explained by independent rates of these disorders in youth rather than a hypothesis that ADHD is a risk factor for bipolar disorder in youth.

Rates of conduct disorder and oppositional defiant disorder are reported to be as high as 79% in youth with BP. The combination of externalizing disorders and bipolar disorder appears to run in families and is associated with significant functional morbidity. Comorbid anxiety disorders have been reported in up to 77% of children with narrow phenotype bipolar disorder, with particularly high rates of generalized anxiety disorder, separation anxiety, and specific phobia. There is some evidence that, in a number of cases, anxiety symptoms may intensify during episodes of depression and mania, resolving when mood symptoms are effectively treated. Finally, features of ASDs appear to be quite common among youth with narrow phenotype bipolar disorder, with one study reporting that approximately 20% of their sample met full criteria for a comorbid autism spectrum condition.

As with adults, substance and alcohol abuse are common comorbidities in adolescents with BP. ADHD and conduct disorder are both risk factors for the development of substance use problems. In an epidemiologic sample, young bipolar patients with substance use problems had conduct disorder rates three times higher (52%) than those without substance abuse (14.8%; $p < .01$). By comparison, young subjects without either mania or substance problems had conduct disorder rates of only 7.75%. Substance abuse in bipolar adults may be more related to childhood conduct disorder than uncomplicated bipolar disorder. Nevertheless, substance abuse and bipolar disorder clearly exacerbate each other, and youth with bipolar disorder are at high risk of substance abuse.

For externalizing disorders comorbid with BP, there appears to be a developmental shift over time. Rates are highest in children (greater than

75%), lower in adolescence (approximately 50%, especially if both depression and mania have onset after puberty), and lowest, but still present, in adults, in whom the rate of lifetime ADHD has been reported to be approximately 20%. Similarly, conduct disorder/antisocial personality co-occur in adults at double the rates in those with bipolar onset before age 30 years (33%) versus after age 30 years (16%).

DIFFERENTIAL DIAGNOSIS

The five most common clinical conditions to be distinguished from bipolar disorder are ones with which it is most often comorbid. Severe ADHD and oppositional defiant disorder in young children are the ones most often confused with mania. The reason for the emphasis on episodes, euphoria, and grandiosity in making the diagnosis of mania emerges because ADHD is, by contrast, a chronic (rather than episodic) disorder for which elation and grandiosity are not considered symptoms. The clearer these symptoms and behaviors are, the easier the differential diagnosis. If severe irritability and explosiveness are used as the hallmark of early-onset mania, the distinction from severe ADHD with oppositional defiant disorder is extremely difficult. Disruptive mood dysregulation disorder, which identifies youth with temper outbursts and chronic, severe irritability, cannot be diagnosed in the presence of manic symptoms lasting more than a day. This exclusion was added in order to help separate children with nonepisodic irritability from those with mania. Another difficult diagnostic issue is the development of conduct disorder in older children with ADHD. Children with conduct disorder often are very irritable and oppositional, stay out late in defiance of their parents, and thus, may appear to need less sleep, engage in early sexual behavior and drug use (pleasure seeking without regard for consequences), and are hedonistic in their self-interest, caring little for the rights of others.

A third difficult diagnostic distinction is between an agitated, anxious depression and a mixed episode of mania. This complicated differential diagnosis is not unique to children and adolescents. In adult psychiatry, the trend has been to err on the side of making a diagnosis of bipolar disorder and avoiding medications to treat depression and anxiety in order to diminish the potential risk of exacerbating mania. It is necessary to keep an open mind about this issue, and if treatments for mania are not helpful, a diagnosis of depression and anxiety alone should be reconsidered.

Substance abuse not only may render a teen depressed, irritable, emotionally labile, and euphoric, it may also precipitate a psychosis that may be indistinguishable from mania. In the short term, it is often impossible to determine whether the psychopathology is primary or substance induced.

Finally, especially in adolescence, first-episode schizophrenia poses an important differential diagnosis. For many years, schizophrenia was considered to be the only psychosis that occurred in adolescents. It is now understood that depression with psychosis (often a harbinger of bipolar disorder) and mania with psychosis are equally likely to occur. In the clearest cases, teens developing schizophrenia will have an insidious onset of psychosis, long-standing cognitive difficulties, and prominent and bizarre delusions. However, that may not be the case, and only careful follow-up over several years will allow the true diagnosis to emerge.

COURSE

There have been five areas of investigation into the course and outcome of bipolar disorder: (1) the duration and recurrence rate of manic and depressive episodes, (2) percent time symptomatic over longitudinal follow-up, (3) functional outcome, (4) diagnostic stability (i.e., how likely is a youth to retain the diagnosis of bipolar disorder), and (5) suicidal thoughts and behaviors.

Episode Duration and Recurrence

The median duration of an episode of mania in adults is 3 to 6 months. Episodes that include both polarities (rapid cycling/mixed), which last longer, have durations of about a year. As with adults, at least 80% of children and adolescents with BD do have a period of recovery, defined as 8 weeks with no or minimal symptoms. The duration of the episode (or time to recovery) in youth is variable, though somewhat longer than typically observed in adult studies. In Geller's sample of children with a mean age at intake of 11 years, the index episode lasted 60.2 (±47.5) weeks; later episodes lasted for shorter periods of time (55.2 and 40 weeks, respectively). The Course and Outcome of Bipolar Illness in Youth study in which bipolar youth aged 7 to 17 years were combined from three different sites reported a median episode duration of 52 weeks for BP. Purely adolescent-onset samples fare somewhat better, particularly when focusing only on mania. The median duration of an index episode from one hospitalized sample with mania (mean age 16 years) was 11 weeks, while outpatient teens with mania (average age 14 years) from India had a median episode duration of only 8.7 weeks. The mean episode duration in two other samples was 29 ± 31 weeks and 20 ± 13 weeks (mania/mixed). In older adolescents (mean age 17 years), the median episode duration was 30 weeks. The reasons for these differences in duration are not clear. However, it may relate to the polarity of the episode; as with adults, manic episodes in youth are generally shorter than depressive or rapid cycling episodes. Among those children and adolescents who achieve remission, rates of recurrence ranged from 27% among community-based adolescents at 2-year follow-up to 70% among child and adolescent outpatients at 4-year follow-up.

General risk factors for longer episode duration include nonadherence with pharmacologic treatment, rapid cycling, psychotic symptoms, earlier age of onset, low socioeconomic status, comorbid anxiety, ADHD, and/or disruptive behavior disorders. Alcohol use disorders and exposure to antidepressants have been associated with increased risk of recurrence; involvement in psychotherapy appears to have a protective effect. Notably, males diagnosed with pediatric bipolar disorder appear to have a less severe course with a greater chance of recovery from the initial episode than females. Moreover, compared to children with bipolar I or II disorder, those with *other specified* bipolar disorder endured a more chronic course of illness with persistent subthreshold symptoms and delayed treatment response. Predictors of a more rapid recovery include lithium treatment and living in an intact biologic family. Recently investigators developed a risk calculator to predict mood recurrence in the Course and Outcome of Bipolar Youth (COBY) study using variables known from the literature to increase risk. This risk calculator was able to predict, with good accuracy (area under the receiver operating curve [ROC] curve; AUC >.8), whether an individual would go on to have a recurrence in the next 2 to 5 years. Furthermore, this risk calculator was externally validated in an adult sample (Collaborative Depression Study), showing similar accuracy rates, indicating that a predictive model developed in an adolescent and young adult sample also yields across the lifespan.

Percent Time Symptomatic

Longitudinal studies indicate that youth with BD spend, on average, 60% of the time symptomatic with threshold or subthreshold mood episodes. While the COBY study found that youth spent a larger percentage of that time in subthreshold and depressive episodes, Geller's study found that youth spent 40% of the time manic.

Regardless, percentages at the group level do not describe what is going on for the individual patient. In the COBY study, latent class growth analysis was conducted on available follow-up data (367 youth followed up every 6 months over an average period of 8 years) to find subgroups of mood trajectories. Four distinct classes were identified: (I) "predominantly euthymic" (24% of the sample), (II) "moderately euthymic" (34.6% of the sample), (III) "ill with improving course" (19.1% of the sample), and (IV) "predominantly ill" (22.3% of the sample). Group I was found to be euthymic, on average for 84% of the 8 years; group II was euthymic for 47.3% of the 8-year study period; group III was euthymic for 42.8% of the 8-year period, and group IV was euthymic for 11.5% of the 8-year follow-up period. Factors related to a better course, with longer periods of euthymia, included (1) older age at onset of symptoms and first episode, (2) less severe depression, (3) less severe manic or hypomanic symptoms, (4) less frequent subsyndromal episodes, (5) fewer suicide attempts, (6) lower rates of sexual abuse, and (7) less family history of bipolar disorder or substance use disorders. The youth in the groups with greater periods of euthymia had a less complicated course overall, with lower rates of comorbid ADHD and anxiety disorders and lower rates of all psychiatric disorders in family members. In fact, almost half of the group with the most time euthymic did not experience another episode during the 8-year follow-up. However, although the other half of group I was also euthymic for about 84% of the 8-year study period, they did continue to have new mood episodes during the follow-up period. In a second cohort of 144 adolescents with BD-I/II, Weintraub, Miklowitz, and colleagues conducted a latent class analysis of mood symptoms over 2 years of follow-up, similar to the COBY study. They identified four latent classes, including a "predominantly euthymic" class (29.9% of the sample).

Psychosocial Functioning

Psychosocial impairment often persists beyond mood episodes in BD. In the COBY study, 92.8% experienced impairment in school activities, and

40.5% had difficulties with interpersonal functioning. Psychosocial impairment was more common in those with more severe illness, but 44% of the “predominantly euthymic” class continued to experience significant difficulties. Importantly, impairment in psychosocial function preceded depressive episodes, indicating a possible target for intervention. The Child Global Assessment Scale score in Geller and colleagues’ 8-year follow-up of early-onset mania was 33.5 (SD 10.6) for those with the disorder at age 18 or older. Another 1-year follow-up study reported functional recovery in only 39% of the sample. Across studies, predictors of psychosocial impairment include early BD onset, severe illness course, low socioeconomic status, poor medication adherence, and current comorbidity.

Diagnostic Stability

A comprehensive investigation of the stability of diagnosis of bipolar disorder in adults is critical for both prognosis and treatment. The questions regarding BP in youth are for whom will the cycles be mainly depressive or manic, how often will they reoccur, and for whom will BP completely and forever remit? This is important to keep in mind to determine how long a medication (mood stabilizer or antipsychotic) should be administered during and after the resolution of an episode and whether there are observable predictors of longitudinal course. There are a variety of ways to assess diagnostic stability, including via medical records, national registries, and longitudinal follow-up. Based on a Spanish study of medical records, over 95% of 72 youths (median age 12.96, 37.5% with BD-I) maintained a bipolar disorder diagnosis after a median of 3.86 years of follow-up. In an analysis of the Danish population database, 86% maintained a BD diagnosis at 3 years and 73% at 10 years; those who no longer had a BD diagnosis had changed to schizophrenia. Predictors of diagnostic change to schizophrenia included male sex, history of substance abuse, diagnosis given during hospitalization, and parent with a schizophrenia spectrum disorder. A longitudinal study by Wozniak and colleagues found that 73.1% of youth diagnosed with BD-I continued to meet the criteria for this diagnosis 4 years later (at a mean age of 13.4 years old).

The question of diagnostic stability was recently addressed in the COBY study. Investigators found that 57% of youth with child-onset BP continued to have hypomania/mania in adolescence. They found that approximately ~40% of participants who had a manic or hypomanic episode in adolescence (13 to 17 years old) also had one during young adulthood; when this outcome was limited to manic episodes, the number was ~25%. The most important predictors of diagnostic continuity were age at most recent manic or hypomanic episode during adolescence and family history (of mania and suicide attempts). After accounting for adolescent course, the age of onset and pediatric comorbid disorders were not significant predictors of diagnostic stability. Even in those without syndromal mania or hypomania, most had subthreshold manic episodes.

Interestingly, the issue of diagnostic stability is not unique to pediatric bipolar disorder but also a challenge in adults diagnosed with BD. In a 10-year follow-up of early-episode 15 to 50 years old patients ever diagnosed with first-episode mania with psychosis by Carlson and colleagues, diagnostic consistency occurred in only half the sample. Much of the diagnostic instability was predicated on a change in the course of illness. BPI became clearer over time in 17% (as people with psychotic depression developed mania, for example), but in the rest, even though BPI seemed obvious early, other conditions emerged. The most powerful predictors of diagnostic instability (i.e., not maintaining a BPI diagnosis) were having any childhood psychopathology (OR 2.8, CI 1.4 to 5.6) and specifically having a childhood externalizing disorder (OR 4.7, CI 2.5 to 8.9). The more acute and uncomplicated the diagnosis at onset, the more stable the diagnosis; however, a poor premorbid adjustment in terms of childhood psychopathology and nonacute onset predicted diagnosis change. Prognostic implications were substantial as the consistently diagnosed bipolar group had significantly better functioning on Global Assessment Functioning (GAF) ($p < .001$), spent less time symptomatic over their first 48 months of illness (26% vs. 47%) ($p < .001$) and much less time psychotic (7% vs. 31%) ($p < .001$).

Suicidal Thoughts and Behaviors

Two separate longitudinal studies, including the COBY study, found a 5-year rate of suicide attempts to be around 20% in children and adolescents with bipolar disorder. In a public mental health system, the lifetime rate of suicide attempts in adolescents with a diagnosis of BD-I was 57%. High quality of interpersonal relationships and treatment with lithium were both identified in the COBY study as protective factors against suicide attempts.

Functional outcome outside of symptom status has not yet been addressed in longer-term longitudinal studies. The Child Global Assessment Scale score in Geller and colleagues' 8-year follow-up of early-onset mania was 33.5 (SD 10.6) for those with the disorder at age 18 or older. Another 1-year follow-up study reported functional recovery in only 39% of the sample. ADHD comorbidity and poor medication adherence were associated with poor function. However, it appeared that functional outcome was also associated with syndromal status.

TREATMENT

Our current evidence base for the treatment of mania addresses both medication and psychosocial intervention with considerably more data for the former. Industry-sponsored and investigator-initiated clinical trials have addressed the relative efficacy of atypical antipsychotics with less robust data for anticonvulsants and some recent data supporting lithium. Trials of medication for bipolar depression are much smaller, and to date, only two medications have been shown to be more effective than placebo. Psychosocial interventions are beneficial in addition to medication, not replacements for it. Medication management requires obtaining baseline laboratory tests relevant to the medication used and monitoring side effects as well as monitoring symptoms of mania and depression.

Practice parameters and clinical guidelines, such as the AACAP Practice Parameters, are consistent in advocating the importance of medication.

Pharmacotherapy

As part of the *Best Pharmaceutical Act for Children,* newly developed medications to be used in children or adolescents must be tested for safety and effectiveness in that sample. U.S. Food and Drug Administration (FDA) labeling requirements allow fewer studies to be done if the condition—in this case, mania—is the same condition that exists in adults, such that similar measures to test safety and efficacy are valid in that age group. A consensus conference convened in 2002 established that youth between ages 10 and 17 years can experience the same condition. The same BP disorder in younger children was less well established, hence the cutoff at age 10 years. This is important because it means that the medications and doses used for adults are being studied in older children and adolescents. As a result of this mandate, we have the most information on antipsychotic medications, which will be discussed first.

As of 2021, the following medications which are approved for the treatment of acute mania in adults are also approved for youth: chlorpromazine for children down to age 1 (no studies were ever done, however), olanzapine (ages 13 to 17), for children ages 10 to 17, quetiapine, risperidone, aripiprazole, and asenapine, and lithium is now approved for mania and mixed mania down to the age of 7. There has been a positive randomized clinical trial in youth for ziprasidone, and there is a second trial underway but no FDA approval yet as of this writing.

Olanzapine was the first second-generation antipsychotic to be studied and approved. With doses ranging between 10 and 20 mg, 44.8% of youth on olanzapine showed at least a 50% reduction in symptoms compared to a placebo response rate of 18.5%. This was significant ($p < .001$). The mean decrease from the baseline Y-MRS entry score of 33 was 17.7 points, and the effect size was 0.84.

Other atypical antipsychotic trials for risperidone, aripiprazole, and quetiapine were required to use low- and high-dose alternatives. Both high and low doses of risperidone and aripiprazole were significantly effective as measured by a 50% reduction in Y-MRS entry scores and a decrease of Y-MRS scores of between 14 and 18 points compared to placebo responses of 8 to 10 points. In general, 45% to 65% of subjects improved on active medication versus 18% to 37% on placebo. The absolute treatment benefit (percentage of drug responders minus placebo responders) averaged 19% to 38%, and the number needed to treat (NNT) in successful trials in which drug was better than placebo varied from about 2.6 (high-dose aripiprazole and high-dose risperidone) to around 5 (low-dose aripiprazole and high-dose quetiapine). Side effects were usually greater for the high-dose condition without a correspondingly higher improvement. Asenapine, which is administered twice a day sublingually, was tested at low (2.5 mg), medium (5 mg), and high doses (10 mg). At the endpoint, YMRS responder rates for all 3 doses ranged 42% to 54% compared with PBO (28%). The 10-mg dose had the most robust separation from placebo on the Y-MRS in the clinical trial.

There had been small trials of lithium in children for many years, but none addressing acute mania with a placebo control until relatively recently. The TEAM study, which randomized children to risperidone, lithium, or divalproex, had reported a lithium response rate of 35.6%. There was no placebo arm, but placebo responses in industry-sponsored trials are about 25% to 30%. Thus, lithium would not have beaten the placebo in that study. However, the Collaborative Lithium Trials (COLT), specifically designed to test the efficacy of lithium, found a 4.5-point difference on the Y-MRS scale at week 8 between lithium- and placebo-treated participants (p = .031). Even so, rates of response (≥50% drop in Y-MRS) did not differ significantly between lithium and placebo groups, respectively (37.7% vs. 28.6%, Robb et al., 2014).

Several anticonvulsants have an FDA treatment indication for mania in adults. Data for bipolar youth is less robust. An industry-sponsored trial with divalproex ER requested by the FDA, which was only 4 weeks in duration (the duration of adult trials), was negative; drug and placebo responses were 24% and 23%, respectively. Atypical antipsychotics outperform divalproex in clinical trials. In addition, where it was measured, the time to response for atypicals was shorter than for divalproex. Blood levels were adequate in both the industry-sponsored trial and in the "Treatment of Early Age Mania" study [TEAM]. In the latter study, a poor response likely was due to the >90% rate of comorbid ADHD. As noted below, comorbid ADHD is a predictor of poor treatment response.

Published trials of two other anticonvulsants in youth, topiramate, and oxcarbazepine, have not been found to be significantly better than placebo. There have been no published placebo-controlled trials of carbamazepine.

Correll and colleagues provide a meta-analysis with relative effect sizes of different classes of medication. Atypical antipsychotics for mania/mixed mania show a larger effect size than mood stabilizers (effect size [ES] 0.65, CI 0.53 to 0.78 compared to ES 0.20 CI 0.02 to 0.39), and a 50% reduction in Y-MRS entry score is seen in 50% to 70% of subjects compared to 20% to 37% on placebo. The NNT in successful trials where the drug was better than placebo varied from 3.8 to 5.3.

The presence of ADHD complicates and attenuates treatment response for youth with BP. Attempts to treat both conditions include one study in which mixed amphetamine salts were added to divalproex, decreasing ADHD symptoms in that group. In another study, children with bipolar disorder stabilized on a thymoleptic received varying doses of methylphenidate (MPH) and placebo. ADHD rating scale scores were significantly improved compared to placebo in the MPH-treated youth.

MPH added to aripiprazole did not destabilize youth in a cross-over study (though ADHD was not significantly improved, either). Finally, a small open-label study of atomoxetine added to a mood-stabilized regimen for BP and ADHD found atomoxetine was effective in treating the ADHD, although 2 subjects out of 12 reported worsening mood at some point during the 8-week study.

Given the level of severity and impairment with which most children and teens begin a bipolar treatment study, the improvement, when it occurs, is noteworthy, but it often is not sufficient. Few data are available to predict treatment response. The use of more than one medication has increasingly become acceptable for treating mania/bipolar disorder in both adults and young people. In youth, there are three types of studies: (1) comparing two medications to one—for example, combined divalproex and quetiapine versus quetiapine alone; (2) adding one to another if the first drug does not work—for example, adding risperidone to lithium or divalproex; and (3) starting two together—for example, lithium and divalproex—and discontinuing one.

Although the types of adverse effects of medication are similar in children and adolescents, the frequency and severity may differ. For instance, weight gain with metabolic syndrome, especially in medication-naive patients, is significant. Unlike other antipsychotics, aripiprazole is associated with a reduction in prolactin. However, it also may cause a dose-dependent increase in extrapyramidal symptoms. Risperidone, olanzapine, and quetiapine may increase prolactin and cause more weight gain than aripiprazole. Asenapine is very sedating and also may cause dysgeusia and oral hypoesthesia because of its sublingual administration.

Cases of polycystic ovaries secondary to divalproex may be of particular concern in young people. Similarly, disinhibited behavioral responses to many types of central nervous system medications seem to occur more often in children and perhaps slightly less so in teens. This is of particular relevance to antidepressant use, where misinterpreting the side effect of activation as a manic episode may result in withholding helpful medication and prescribing unnecessary treatments. Indeed, the DSM-5 revision has attempted to prevent that error by stating that the activation is a manic episode if it continues after withdrawal of the medication and beyond the medication’s physiologic effects.

Prevention of Relapse

Children and adolescents with bipolar disorder appear to have higher relapse rates and more episodes over the course of a year than adults. Regarding relapse prevention for mania, there is some suggestion that three medications, aripiprazole, lamotrigine, and lithium, may increase the time to relapse for mania. Beyond these reports, it is prudent to assume that medications that stabilize a child with BP should be continued for at least an entire school year while a child is in school. The AACAP Practice Parameter suggested 12 to 14 months. In cases in which it has taken a long time to stabilize the child and in which there is clear evidence from symptom monitoring that treatment has been effective, it may be wise to continue the treatment until the child has improved function, academically, socially, and within the family. The decision to continue medication treatment or not should be assessed annually. In cases in which treatment response is more tenuous and side effects more troublesome, longer maintenance is more difficult to support.

Treatment of Bipolar Depression

Treatment of bipolar depression in young people has received relatively little attention. Even in adults, only three medications (olanzapine/fluoxetine combined [OFC], quetiapine, and lurasidone) have FDA-approved indications for bipolar depression; two of those, OFC and lurasidone, are approved in youths aged 10 to 17 years. Other studies in young people consist of two open trials—one of lithium monotherapy in hospitalized adolescents with bipolar depression and one using lamotrigine alone or with other medications in outpatient adolescents. Symptoms declined in both studies. In the absence of placebo-controlled studies, and given the high placebo response in depression studies in youth in general, the treatment utility of these medications in youth remains to be demonstrated. Moreover, the adolescent samples studied had considerably more subthreshold manic symptoms (Y-MRS scores were 12 or fewer) than in the adults, whose Y-MRS scores were 2 or fewer. Thus, it is not even possible to extrapolate intelligently from adult data.

Of the placebo-controlled studies, one used flexibly dosed OFC (6/25 to 12/50 mg), which demonstrated superiority to placebo (effect size of 0.46), and the other used lurasidone (mean dose 33.6 mg/day). It had a similar effect size (0.45). Both have been approved by the FDA for the acute treatment of bipolar I depression in patients 10 to 17 years of age. Neither short- nor long-acting quetiapine has demonstrated efficacy compared to placebo in bipolar depression in children or teens.

There is a modest if anecdotal, literature on the efficacy of electroconvulsive therapy (ECT) in mood disorders in teens. The reader is referred to the AACAP's ECT Practice Parameter on its indications and contraindications.

Psychosocial Treatment

Psychosocial interventions (1) target the recognition of future episodes before they progress too far, (2) enhance treatment adherence, and (3) address environmental stressors that act as possible precipitants to further episodes, (4) help parents and children manage mood dysregulation, and (5) improve the quality of family relationships. There are a number of empirically supported psychosocial interventions. These include multifamily and individual family psychoeducation, family-focused therapy for adolescents, child- and family-focused CBT for younger children, dialectical behavioral therapy for adolescents, and collaborative problem solving. Each emphasizes the importance of psychoeducation and destigmatization, as well as increasing parental collaboration and empowerment by reducing parent self-blame for "causing the disorder." Other components include helping parents to distinguish between normative developmental behavior and bipolar symptoms, taking proactive steps to decrease the risk of relapse, developing tools for effectively managing emotional arousal, improving family communication skills, and teaching adaptive problem-solving strategies.

FUTURE DIRECTIONS

Developments in neuroscience and genetics, as well as longer-term treatment and follow-up studies, will undoubtedly help to advance the understanding of this complex and disabling disorder, in particular its relation to other psychiatric disorders such as ADHD, conduct disorder, developmental disorders, and psychotic disorders. It is hoped that such advances can shed light on the etiology and underlying pathophysiology, including the identification of dysfunctional brain circuits that may underlie pediatric mania and progress toward mechanistically informed treatments.

Further Readings

Arnold LE, Demeter C, Mount K, et al. Pediatric bipolar spectrum disorder and ADHD: comparison and comorbidity in the LAMS sample. *Bipolar Disord*. 2011;13:509–521.

Althoff RR, Faraone SV, Rettew DC, Morley CP, Hudziak JJ. Family, twin, adoption, and molecular genetic studies of juvenile bipolar disorder. *Bipolar Disord*. 2005;7:598–609.

Benarous X, Rafiln M, Bodeau N, Dhossche D, Cohen D, Consoli A. Adverse childhood experiences among inpatient youths with severe and early-onset psychiatric disorders: prevalence and clinical correlates. *Child Psychiatry Hum Dev*. 2017;48(2):248–259.

Birmaher B, Axelson D, Goldstein B, et al. Four-year longitudinal course of children and adolescents with bipolar spectrum disorders: the Course and Outcome of Bipolar Youth (COBY) study. *Am J Psychiatry*. 2009;166:795–804.

Birmaher B, Gill MK, Axelson D, et al. Longitudinal trajectories of and associated baseline predictors in youths with bipolar spectrum disorders. *Am J Psychiatry*. 2014;171:990–999.

Birmaher B, Merranko JA, Gill MK, et al. Predicting personalized risk of mood recurrences in youths and young adults with bipolar spectrum disorder. *J Am Acad Child Adolesc Psychiatry*. 2020;59(10):1156–1164.

Carlson GA, Klein DN. How to understand divergent views on bipolar disorder in youth. *Annu Rev Clin Psychol*. 2014;10:529–551.

Cirone C, Secci I, Favole I, et al. What do we know about the long-term course of early onset bipolar disorder? A review of the current evidence. *Brain Sci*. 2021;11(3):341.

Correll CU, Sheridan EM, DelBello MP. Antipsychotic and mood stabilizer efficacy and tolerability in pediatric and adult patients with bipolar I mania: a comparative analysis of acute, randomized, placebo-controlled trials. *Bipolar Disord*. 2010;12(2):116–141.

Danese A, Widom CS. Objective and subjective experiences of child maltreatment and their relationships with psychopathology. *Nat Hum Behav*. 2020;4(8):811–818.

Duffy A, Carlson G, Dubicka B, Hillegers MHJ. Pre-pubertal bipolar disorder: origins and current status of the controversy. *Int J Bipolar Disord*. 2020;8(1):18.

Findling RL, Robb A, McNamara NK, et al. Lithium in the acute treatment of bipolar I disorder: a double-blind, placebo-controlled study. *Pediatrics*. 2015;136:885–894.

Frangou S. Neuroimaging markers of risk, disease expression, and resilience to bipolar disorder. *Curr Psychiatry Rep*. 2019;21:52.

Geller B, Luby JL, Joshi P, et al. A randomized controlled trial of risperidone, lithium or divalproex sodium for initial treatment of bipolar I disorder, manic or mixed manic phase, in children and adolescents. *Arch Gen Psychiatry*. 2012;69:515–528.

Goldstein BI, Birmaher B, Carlson GA, et al. The International Society for Bipolar Disorders Task Force report on pediatric bipolar disorder: Knowledge to date and directions for future research. *Bipolar Disord*. 2017;19(7):524–543.

Hafeman DM, Merranko J, Axelson D, et al. Toward the definition of a bipolar prodrome: dimensional predictors of bipolar spectrum disorders in at-risk youths. *Am J Psychiatry*. 2016;173:695–704.

Kondo DG, Hellem TL, Shi X-F et al. A review of MR spectroscopy studies of pediatric bipolar disorder. *AJNR Am J Neuroradiol* 2014;35:S64–S80.

Leibenluft E. Severe mood dysregulation irritability and the diagnostic boundaries of bipolar disorder in youths. *Am J Psychiatry*. 2011;168:129–142.

Menculini G, Balducci PM, Attademo L, Bernardini F, Moretti P, Tortorella A. Environmental risk factors for bipolar disorders and high-risk states in adolescence: a systematic review. *Medicina (Kaunas)*. 2020;56(12):689.

Pataki C, Carlson GA. The comorbidity of ADHD and bipolar disorder: any less confusion? *Curr Psychiatry Rep*. 2013;15:372–378.

Paterson C, Wang Y, Hyde TM, Weinberger DR, Kleinman JE, Law AJ. Temporal, diagnostic, and tissue-specific regulation of NRG3 isoform expression in human brain development and affective disorders. *Am J Psychiatry*. 2017;174(3):256–265.

Schenkel LS, Towne TL. Errors in identifying emotion in body postures and facial expressions among pediatric patients with bipolar disorder. *J Clin Exp Neuropsychol*. 2020;42:735–746.

Singh MK. *Clinical Handbook for the Diagnosis and Treatment of Pediatric Mood Disorders*. 1st ed. American Psychiatric Association Publishing; 2019. 277–311.

Singh MK, Nimarko AF, Garrett AS, et al. Changes in intrinsic brain connectivity in family-focused therapy versus standard psychoeducation among youth at high risk for bipolar disorder. *J Am Acad Child Adolesc Psychiatry*. 2021;60(4):458–469:S0890856720313344.

Stringaris A, Santosh P, Leibenluft E, Goodman R. Youth meeting symptom and impairment criteria for mania-like episodes lasting less than four days: an epidemiological enquiry. *J Child Psychol Psychiatry*. 2010;51:31–38.

Stringaris A. What we can all learn from the Treatment of Early Age Mania (TEAM) trial. *J Am Acad Child Adolesc Psychiatry*. 2012;51:861–863.

Sun AY, Woods S, Findling RL, Stepanova E. Safety considerations in the psychopharmacology of pediatric bipolar disorder. *Expert Opin Drug Saf*. 2019;18:777–794.

Van Meter A, Moreira ALR, Youngstrom E. Updated meta-analysis of epidemiologic studies of pediatric bipolar disorder. *J Clin Psychiatry*. 2019;80(3):18r12180.

Weintraub MJ, Schneck CD, Axelson DA, Birmaher B, Kowatch RA, Miklowitz DJ. Classifying mood symptom trajectories in adolescents with bipolar disorder. *J Am Acad Child Adolesc Psychiatry*. 2020;59(3):381–390.

Youngstrom E, Meyers O, Youngstrom JK, Calabrese JR, Findling RL. Diagnostic and measurement issues in the assessment of pediatric bipolar disorder: implications for understanding mood disorder across the life cycle. *Dev Psychopathol*. 2006;18:989–1021.

51

Anxiety Disorders in Children

51.1 Obsessive-Compulsive Disorder in Childhood

Adam B. Lewin, Ph.D., A.B.P.P., and John Piacentini, Ph.D., A.B.P.P.

Pediatric obsessive-compulsive disorder (OCD) is a chronic neuropsychiatric condition associated with broad impairments in functioning. Over the past two decades, OCD in childhood has garnered increased attention, given heightened public and professional awareness of substantial upward revision of the prevalence of this disorder in youth. Advances in functional neuroimaging, genetics, neurosurgery, neuropsychology, and psychoneuroimmunology have contributed to a better understanding of etiologic factors contributing to the pathogenesis of OCD in childhood. Concurrently, randomized controlled trials and subsequent meta-analyses from multiple research groups have established individual- or family-based cognitive behavioral therapy (CBT) with exposure and response prevention (E/RP) and selective serotonin reuptake inhibitors (SSRIs), alone or in combination, as first-line treatments for OCD in childhood. Despite promising interventions, OCD remains a chronic and disabling disease which can lead to marked social, academic, and familial dysfunction as well as increased risk for depression, self-harm, substance use, and poorer quality of life in adulthood. For individuals who do not benefit from available treatment, further refinement of treatment approaches remains paramount to diminish suffering and impairment while increasing the likelihood that youth with OCD can maintain relatively normal lives. Because many youth do not receive treatment due to lack of access to child psychiatrists and psychologists trained in evidence-based intervention, improving availability of care is another key goal.

DEFINITIONS

OCD is characterized by the presence of recurrent obsessions or compulsions that are distressing, time consuming, or debilitating. Obsessions are defined as recurrent, persistent, and intrusive thoughts, ideas, impulses, or mental images that are experienced by the individual as distressing and as a product of one's own mind (to differentiate them from thought insertion). These thoughts or ideas frequently generate (or are associated with) feelings of anxiety, tension, disgust, or aversion. Table 51.1–1 lists common obsessional fears. Compulsions are purposeful, repetitive behaviors or mental acts that typically function to neutralize, prevent, or reduce anxiety. Table 51.1–1 also listed common compulsive behaviors. Often, compulsions are performed in response to an obsessive thought or according to some other rigidly applied rules. Although compulsions function to mitigate distress or anxiety or to prevent some dreaded event, they are clearly excessive or are not realistically connected with the triggering stimulus (e.g., tapping three times to prevent illness after someone in the room coughs). While obsessions and compulsions are often ego dystonic and are often recognized by children as senseless or exaggerated, such insight is not required for the diagnosis of OCD in childhood. Otherwise, symptoms of OCD in childhood resemble those seen in adults.

HISTORY

Early examples of childhood-onset OCD, consisting of case descriptions, date back to the late 19th century. In 1895, Sigmund Freud described an 11-year-old child with obsessions. Subsequently, the French physician Pierre Janet provided the first description consistent with OCD in childhood in his 1903 report on a 5-year-old boy. In the first half of the 20th century, descriptions of clinical samples and generalizations emerged. In 1942, L. Berman extracted clinical information on 62 children with possible obsessive-compulsive neurosis out of a case registry of 3,050 cases. In 1955, Louise Despert described more than 60 cases of children with "obsessive-compulsive neurosis," noting a male preponderance, children's recognition that their symptoms are abnormal and unwanted, and the tendency to attempt to conceal their symptoms from treatment professionals and others. Leo Kanner's (1962) textbook of Child Psychiatry contained descriptions of childhood OCD and the consequences of social isolation and family accommodation.

In 1965, Lewis Judd provided descriptive criteria for the diagnosis of OCD in children that closely resembles the current diagnostic criteria, including the requirement of significant impairment in functioning. Reports as recent as 1973 implicated parental withdrawal of attention as a critical causal factor of childhood OCD. However, by the 1980s, the notion that OCD was a neurobiologic disorder with possible childhood onset became more commonplace. Further, findings from the Epidemiologic Catchment Area (ECA) study indicating most adults with OCD report an onset of the disorder during adolescence prompted increased awareness of the disorder in childhood as well as research in this area. Programmatic work in the 1980s under the direction of Judith Rapoport and her group at the National Institute of Mental Health provided the basis for the first systematic studies of epidemiology, phenomenology, and treatment of childhood- and adolescent-onset OCD as well as fundamentally dismissing any notion that the emergence of OCD in youth was unusual or unexpected. The Pediatric Obsessive-Compulsive

Disorder Treatment Study (POTS) emerged as the plenary multimodal treatment trial in 2004.

Table 51.1–1.
Common Symptoms of Pediatric Obsessive-Compulsive Disorder Obsessions

Obsessions
Concerns with dirt, germ exposure, fears of illness, disgust
Fears of harm befalling self or others
Need for symmetry, order, exactness, completeness
Need to save or hoard
Excessive moral, religious, or sexual concerns
Compulsions
Cleaning rituals
Checking
Repeating behaviors
Ordering or arranging
Confessing, reassurance seeking

COMPARATIVE NOSOLOGY

Diagnostic criteria for OCD have remained relatively constant over the past three editions of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM). The fifth revision (DSM-5) published in 2013 defines the following features of OCD: (1) recurrent obsessions or compulsions; (2) obsessions or compulsions cause marked distress, are time consuming (take more than 1 hour per day) or cause significant impairment in functioning; and (3) obsessions and compulsions not due to the presence of another mental disorder, substance, or other medical condition. The DSM-5 further includes specifiers to code degree of insight and to indicate whether the individual has a history of a tic disorder. The 10th revision of the *International Statistical Classification of Diseases and Related*

Health Problems (ICD-10) provides a description of OCD that is similar to the DSM-5.

EPIDEMIOLOGY

Although estimates vary, studies suggest lifetime prevalence for OCD in children between 1% and 3%, indicating that OCD is more common than previously thought and as common in youth as in adults. Sample age may explain some of the variability in these estimates. Older adolescents yield prevalence rates between 0.53% and 4.0%, whereas estimates for school-aged children remain lower. In fact, a large prevalence study from the United Kingdom showed an exponential rise in rates across age groups (5 to 7, 8 to 10, 11, 12, and 13 years). Another recent community sample found a prevalence rate of 3% for “clinical” OCD, defined as symptoms severe enough to cause distress and interference in functioning. The most accepted 6-month prevalence rate is between 0.5% and 1% of the general pediatric population. Of note, rates of OCD in late adolescence make it more common than panic disorder, schizophrenia, and bipolar disorder.

Mean age of onset for pediatric OCD is between ages 6 and 11. Onset age appears bimodal and varies by gender with a male preponderance in cases of prepubertal onset with an initial peak of incidence around puberty and subsequently in young adulthood. The gender distribution becomes roughly equivalent during adolescence. It is currently believed that up to 80% of adult OCD cases have an initial onset during childhood. Children with an early onset of OCD (below age 7 years) are also more likely to be male and to have a family history of OCD in comparison to children with later onsets, suggesting that genetic factors may play a role in development of early-onset cases. Referral bias may account for the underrepresentation of minorities and youngsters from lower socioeconomic strata in clinical samples of OCD youth, as epidemiologic studies with adult samples have not replicated this pattern.

Table 51.1–2.
Comorbid Diagnoses in Pediatric Obsessive-Compulsive Disorder

Diagnosis	Frequency (%)
Anxiety disorders	26–70
Mood disorders	11–62
Tic disorders	11–26
Attention deficit hyperactivity disorder	16–20
Disruptive behavior disorders	9–19
Autism spectrum disorders	3–7

The occurrence of OCD in childhood may be somewhat higher than reported given the secrecy that children often display in reporting embarrassing thoughts and behaviors. In addition, limited insight into the abnormality of their symptoms (and linking OCD symptoms to potential impairment), parent inability to recognize OCD symptoms, and the lack of awareness about the availability of efficacious treatments may contribute to underdiagnosis. Therefore, it is reasonable to assume that many childhood cases of OCD remain undetected.

Comorbidity

In childhood, psychiatric disorders co-occurring with OCD is the norm rather than the exception. For example, in an analysis of 70 clinical cases of youth with OCD, 84% carried at least one additional Axis I psychiatric diagnosis. It is estimated that as many as 50% of pediatric patients with OCD have two or more co-occurring diagnoses (Table 51.1–2). The most common comorbid disorders include other anxiety disorders, depressive disorders, tic disorders including Tourette syndrome, attention deficit hyperactivity disorder (ADHD), and disruptive behavior disorders. Individuals also commonly report sensory abnormalities. Major depressive disorder, while rare in younger children with OCD, increases in prevalence over the course of adolescence and into adulthood.

ETIOLOGY

Neurobiologic Factors

Although the cause of OCD remains unknown, multiple neurochemical, neuroanatomical, and immunologic etiologic theories have been proposed. Clinical descriptions of OCD associated with Sydenham chorea (SC), tic disorders, traumatic brain injury, encephalitis, and Huntington chorea have helped to solidify the broad concept of the disorder as brain based and presumably reflecting impaired striatal functioning. Though the neurocircuitry and neurochemistry involved in OCD symptomology has been better clarified using in vivo brain techniques and evaluation of responses to psychoactive medications, there continues to be a need to further refine etiologic models of the condition.

Neurochemical. Dysfunction of the central serotoninergic neurotransmitter system continues to be a suspect in OCD symptomology. Primary support for this premise is the observed efficacy of SSRIs and altered sensitivity to acute administration of 5-hydroxytryptamine (5-HT) agonists, as well as from results of preliminary studies of 5-HT synthesis. Young patients with OCD have presented with reduced platelet (3H) imipramine binding; conversely, those OCD patients with higher (3H) imipramine binding at baseline differentiated responders versus nonresponders to SSRI treatment.

Research also suggests the involvement of the dopaminergic neurotransmitter system. This may be especially relevant to childhood OCD, given the increased rate of comorbid tic disorders. Clinically, administration of psychostimulants appears to worsen obsessions and compulsions. Conversely, the coadministration of dopamine antagonists with SSRIs may augment the beneficial effects of these agents. However, direct assessment of the dopamine system remains unstudied in children, including a lack of controlled administration of dopamine antagonists. Further, animal studies suggest that excessive glutamatergic output from the cortex to the striatum is associated with repetitive movements. Human studies in children with OCD have reported correlations between striatal glutamate with OCD symptoms and the drop in striatal glutamate with both SSRI medication and exposure-based CBT. In sum, available neurochemical models suggest the possibility of several neurotransmitter abnormalities associated with OCD.

Neurostructural. The pathogenesis of pediatric OCD is characterized by abnormalities in the prefrontal–basal ganglia circuits, specifically frontal–striatal–thalamic–cortical (FSTC) circuit disruption. This premise is supported by imaging studies of OCD, co-occurrence with other disorders with presumed basal ganglia disruption, as well as neuropsychological studies. Volumetric analyses using both magnetic resonance imaging (MRI) and computed tomography (CT) approaches have revealed that treatment-naive children and adults with OCD have smaller basal ganglia segments than adults. For example, reduced globus pallidus volume and gray matter volumetric abnormalities in the anterior cingulated gyrus have been found in pediatric patients with OCD compared to healthy controls. Higher hippocampal volumes and abnormalities in white matter coherence have also been found in child OCD, with greater white matter irregularities linked to increased OCD severity differences compared to healthy controls.

Treatment-naive children appear to also have large thalamic volumes. Thus, it has been hypothesized that children with OCD display abnormalities in normal developmental remodeling of neuronal structures affecting integration, response inhibition, gating/cognitive control, and processing. Positron emission tomography (PET) suggests dysfunction in FSTC circuitry, with studies showing increased glucose metabolism in the orbitofrontal cortex, ventral striatum, right caudate nucleus, and anterior cingulate gyrus, at rest and in response to symptom provocation. Orbitofrontal and caudate metabolic rates are highly correlated, suggesting that interconnecting loops are freed from usual inhibitory influences and reverberate. Disruption of affective/reward processing loops and cognitive control loops have been linked to OCD, and differential development of these systems may be expressed in OCD course and presentation. Recent studies of child OCD suggest that differential development of FSTC circuits and various interconnectivity may explain differential symptom expression (e.g., a finding suggested hypoconnectivity of cognitive control circuity in early-onset cases).

The observations that both medication and behavioral therapies reduce orbitofrontal and caudate blood flow validate the importance of these brain regions in OCD pathogenesis. Paradoxically, lower baseline orbitofrontal glucose metabolism predicts better response to pharmacotherapy. Imaging also suggests the frontal–limbic–basal ganglia system may mediate OCD symptom expression, demonstrating a wider interaction of cognitive and emotional circuits. Specifically, symptom provocation appears to activate paralimbic regions as well as insular, anterior temporal, and amygdalar regions.

Neuroimmunology. Significant attention has focused on the triggering of OCD-like symptoms following group A beta-hemolytic streptococcal infections (GABHS), similar to OCD symptoms observed in SC. These infection-triggered cases of OCD have been coined pediatric autoimmune neuropsychiatric disorders associated with streptococcus (PANDAS) and are believed to represent an autoimmune process related, if not identical to, that associated with rheumatic fever (RF) and SC. (It is noteworthy that patients with RF and SC both display higher frequencies of obsessions and compulsions versus comparison groups, and up to one-third of patients with SC meet research criteria for the diagnosis of OCD.) Immune activation leads to the inflammation of the basal ganglia and disruption of the cortical–striatal–thalamic–cortical (CSTC) function. This acquired basal ganglia dysfunction may result in choreiform movements, tics, obsessions, compulsions, and hyperactivity. PANDAS are characterized by abrupt childhood onset and episodic or saw-toothed course. Relations between the size of the basal ganglia nuclei and OCD symptom severity have been observed using MRI. Similarly, small sample studies of immune-modulating treatments (intravenous immunoglobulin G [IgG] and plasmapheresis) have been associated with improvements in OCD patients with suspected PANDAS. However, these interventions remain experimental, and efforts to establish markers of PANDAS have been inconclusive. In addition, the proportion of child OCD cases potentially attributable to streptococcal infection remains unknown. Additional research into PANDAS is warranted. Although the association between GABHS infection and OCD remains debated, evidence-based psychopharmacologic (e.g., SSRIs) and behavioral (e.g., CBT with E/RP) approaches for treating OCD symptoms in cases with suspected PANDAS should still be strongly considered. Positive titers and exposure to GABHS does not imply an atypical ideological influence; nor does a reduction in OCD symptoms following antibiotics. Formal PANDAS diagnosis should include a sophisticated evaluation of history/course, testing, and exam. Multidisciplinary consultation is recommended prior to experimental, high-risk interventions.

More recently, the term PANS (pediatric acute-onset neuropsychiatric syndrome) was coined to characterize cases with similar clinical presentations as PANDAS but without link or presumed cause with GAS infection. Similar to PANDAS, PANS requires abrupt/dramatic/overnight onset of OCD symptoms, and severe and acute onset of at least two of the following: anxiety, emotional lability, irritability/aggression, developmental regression, academic deterioration, sensory sensitivity, or somatic symptoms (including sleep disturbance, enuresis, or urinary frequency). Similar to PANDAS, diagnosis is not straightforward and even putative PANS caseness does not negate potential benefits of CBT or SSRIs nor necessitate the need for experimental interventions. Diagnosis of PANS/PANDAS is not straightforward and there are no definitive tests; evaluation requires careful multifaceted assessment of medical and psychiatric histories and titers alone are not indicative of diagnosis. As a result, the number of patients with confirmed PANS/PANDAS is significantly lower than the number presenting for treatment or requesting immunologic treatment for these symptoms.

Youth with PANS/PANDAS can benefit from evidence-based psychopharmacology (including antidepressants and antipsychotics) and CBT, but adjustments may be needed, or side-effect profiles may vary.

Family Genetic, Behavioral, and Cognitive Factors

Family Genetic. OCD is a heritable disorder—the familial nature in most cases of OCD is well recognized. Monozygotic twins have a higher OCD concordance rate (0.52) compared to dizygotic (0.21). A meta-analysis of studies exploring familial patterns of psychopathology in relatives of subjects with OCD has documented a fourfold increased frequency of OCD in first-degree relatives of OCD probands (8.2 vs. 2.0 from control families). Rates of OCD among parents of youngsters with OCD range from 17% to 19%. A 2005 review of twin studies suggests 27% to 65% heritability rates for OCD (including nonpediatric onset). In addition, subclinical features of OCD (usually defined as typical symptom constellations without the level of interference required for formal diagnosis) are higher among family members with OCD (e.g., reported in 52% of parents). In fact, relatives of probands with early-onset OCD exhibit even higher rates of OCD and tic symptoms. Analysis of family patterns of OCD finds support for subclinical OCD as an intermediate phenotype that shares similar genetic risk factors.

The relation between OCD and tic disorders, including Tourette syndrome, has emerged from family studies. Tic disorders and some cases of OCD appear to share some susceptibility factors, leading to arguments that OCD may be an alternative expression of the risk factors underlying tic disorders. Some suggest an autosomal dominant model of inheritance with incomplete penetrance. In other words, the phenotypic expression varies in expression from OCD symptoms to tic disorders.

Data from twin and family studies suggest that OCD may be transmitted in families, but specific genes for the disorder have not been identified. Findings from the multisite Obsessive-Compulsive Genetic Association Study (OCGAS), the largest such study to date including 1,065 families/1,406 patients with OCD, failed to identify single nucleotide polymorphisms at genome-wide significance. Investigators posit that larger samples are needed to identify genes associated with OCD. Preliminary findings suggest OCD is polygenetic and likely arises from a large number of diverse genes. Candidate gene studies have yet to produce replicable, significant genes culpable in OCD. The neuronal glutamate transporter gene (*SLC1A1*) has the strongest support, with some findings implicating the serotonin promoter variant (*5-HTTLPR*). Definitive copy number variants associated with OCD have not been identified and, to date, OCD-specific genetic markers remain to be found.

Behavioral Factors. Hobart Mower's two-factor conditioning theory provides a behavioral explanation for OCD symptoms. Via classical conditioning, a neutral stimulus (or event) becomes conditioned to elicit distress because of its association with an unrelated aversive (or feared) stimulus. Fear can be conditioned to both mental events (e.g., blasphemous thoughts) and physical stimuli or behaviors (e.g., a "contaminated" object, items perceived as dangerous, driving, eating). Subsequent to the acquisition of the conditioned fear, rituals develop to reduce or avoid distress. These rituals (or compulsive behaviors) persist and become excessive because they temporarily ameliorate the distress associated with obsessive thoughts via operant conditioning (specifically, a negative reinforcement paradigm). In other words, the escape from/reduction of anxiety associated with performing a compulsive behavior or ritual results in an increased likelihood of the ritual being used in future instances of anxiety.

However, specific traumatic events responsible for the conditioned response are rarely identified. In addition, behavioral theories do not adequately explain the acquisition of obsessive fears. Regardless, these behavioral theories are useful for understanding the maintenance of many OCD symptoms and provide the foundation for behavioral intervention.

Cognitive Factors. The cognitive perspective of the etiology of OCD posits that OCD results from the interpretation of intrusive thoughts rather than the frequency or content of these cognitions. If intrusive thoughts are interpreted in such a way that an individual perceives responsibility for causing or failing to prevent harm, obsessional patterns develop. Attempts to neutralize intrusive thoughts (or obsessions) via motor or cognitive rituals, avoidance, and reassurance-seeking behavior prevent the disconfirmation of the patient's fears and facilitate the proliferation of anxiety. The most common cognitive misinterpretations include thought–action fusion or the tendency to equate thought with action (e.g., thinking something bad is the same as doing it), and exaggerated responsibility for potential untoward consequences of inaction (e.g., "If I don't wash my hands and someone gets sick as a result, then it's my fault."). However, the specificity of these misinterpretations to OCD, especially in children and adolescents, remains unclear.

Similar to behavioral theory, cognitive theory does not adequately explain the etiology of symptoms. It is likely that cognitive misinterpretations and reinforcement theory (see above) maintain and expand OCD symptoms in individuals with biologic propensity for OCD development.

DIAGNOSIS AND CLINICAL FEATURES

Presenting problems among children and adolescents with OCD generally resemble the adult presentation. However, compared to adults, children are more likely to deny impairment due to obsessions and compulsions. Children may also be somewhat less likely to consider their OCD symptoms excessive, perhaps due to lower levels of cognitive awareness. Children do not need to recognize their symptoms as senseless or unrealistic for diagnosis to be assigned; the level of insight varies considerably related to a child's developmental level. Unlike most adults, children might not report symptoms to be ego dystonic. However, children are typically willing to discuss distress from family conflict arising from their rituals (e.g., chronically running late, excessive need for reassurance, time burden on parents, disruption of siblings functioning) or interference in schoolwork resulting in dropping grades. Overall, most children are quite open in describing their symptoms and seek assistance in better managing the illness.

Although comorbidity is the rule rather than the exception in childhood OCD, the content of the obsessions and compulsions cannot be better explained by a co-occurring Axis I disorder. Examples include self-deprecating ruminations associated with depression, obsessive concern with appearance in body dysmorphic disorder, hair pulling in trichotillomania, and fears of harm befalling a loved one in separation anxiety disorder. Similarly, the symptoms also are required not to be substance induced or due to the influence of a medical condition. The range of obsessions and compulsions experienced in childhood is similar to that of adults with OCD, although children are more likely to engage in compulsive reassurance seeking and involve family members in their rituals. Most children with OCD have both obsessions and compulsions; pure obsessional illness is rare in this age group. In fact, most childhood cases of OCD involve multiple obsessions and compulsions (mean number of four current compulsions and four current obsessions in one report). Compulsions in the absence of anxiety-related or otherwise distress-inducing obsessions may occur in (mostly younger) children, who often describe their rituals as being performed in response to an irresistible urge or an otherwise vague sensation (the “just-right” phenomena). The absence of reported obsessions in these cases may be due to underdeveloped awareness or ability to articulate these thoughts or result of a misdiagnosed tic disorder. Specific types of symptoms (e.g., touching/tapping/rubbing) may change with age, but the number of symptoms generally remains constant.

The most common obsessions in childhood include fears of contamination (e.g., dirt, exposure to germs, toxins, contracting an illness), fears of harm befalling self or others (e.g., becoming aggressive, failing to prevent harm), need for symmetry or exactness, hoarding or saving concerns, and excessive moralization or religiosity. Obsessions also may manifest as excessive guilt, self-doubt, or other forbidden thoughts. Common compulsions involve cleaning/washing rituals, checking, repeating behaviors (touching, tapping, rubbing), counting, ordering/arranging, avoidance, doubt/indecision, confessing, seeking reassurance, and slowness. The content and type of obsessive and compulsive symptoms may change over time. OCD is also stress sensitive, and many children experience acute symptom exacerbations during times of psychosocial stress or change (e.g., attending a new school, moving to a new home, death of a family member).

Similar to adult OCD, family members are commonly drawn into the child's OCD symptoms through the provision of reassurance to the child, modification of family routines as a result of the child's symptoms, or even actual participation in the child's compulsive rituals. Such accommodation can negatively impact the family environment and lead to increased levels of family conflict and negative emotional reactions toward the affected child. Accommodation has also been associated with increased feelings of depression and anxiety on the part of the accommodating family member (typically mothers).

Given the typically strange and senseless nature of many OCD symptoms, many children and adolescents will attempt to camouflage their rituals. It is not uncommon for children with OCD to be able to inhibit their rituals for short periods of time to avoid detection (e.g., in class). Not infrequently, teachers, peers, and even parents remain unaware of the child's OCD for months or years, only becoming aware of the problem after the child can no longer control the symptoms or becomes too overwhelmed to cope.

DIFFERENTIAL DIAGNOSIS

Developmentally Normative Rituals and Superstitions

Benign habits, transitory superstitions, or mildly excessive thoughts are not uncommon during childhood. In certain phases of early childhood, rituals or superstitions may be utilized by a child to manage new or stressful situations (first day of preschool, separating from a parent) or achieve a sense of proficiency. For example, a child may require a particular bedtime ritual or demand to wear a particular item of clothing. These normative rituals usually fade by the time of entry into formal schooling and are rarely associated with distress, significant interference with activities, or overlap in content with typical obsessions or compulsions. The absence of functional impairment differentiates mild, age-appropriate superstitions and habits from obsessive-compulsive symptoms.

Other Anxiety Disorders

Despite the presence of preoccupation and worry about issues related to their fears, children with other anxiety disorders typically do not engage in compulsive behaviors to reduce their anxiety. The content of the worry can also be useful in differentiating syndromes. Pediatric patients with generalized anxiety disorder (GAD) present with heightened anxiety relating to their health and safety, the safety of others, and their future. The content of these symptoms, although exaggerated, is grounded in everyday matters (finances, health, grades, safety) as opposed to the unusual or even bizarre content of obsessive worries. Behavioral analogs to compulsions are rare besides the requests for reassurance from others to reduce anxiety. Patients with social anxiety have fears of negative evaluation and extreme scrutiny by others. Again, fears are, by definition, excessive but typically not bizarre. Rituals are rare, although avoidance of feared situations is common. Similarly, the content of fears expressed by children with separation anxiety disorder is, by definition, confined to themes of separation or fears of harm to self or loved ones. Besides avoidance of separation, other behavioral features or rituals are absent from the presentation.

Tic Disorders, Including Tourette Syndrome

Children and adolescents with tic disorders typically display phonic or motor tics that occur frequently and may appear behaviorally similar to compulsive actions, particularly more complex tics. Tics are differentiated from compulsions as usually being avolitional and not preceded by precipitating obsessional thoughts. However, some patients describe their tics as partially volitional in that tic expression is triggered by, and leads to relief from, an uncomfortable premonitory sensation, a functional relationship identical to that between obsessions and compulsions. The content of tics is unlikely to be as purposeful as common compulsions (e.g., washing, arranging), although complex tics may also involve repetition, for example, tapping or touching something a specific number of times. It is noteworthy that there is substantial overlap between tic disorders and OCD in childhood; symptoms of each are not uncommon, and both diagnoses may be warranted.

Autism Spectrum Disorders

The term autism spectrum disorder (ASD) comprises children from a wide range of intellectual (and adaptive) functioning. Pervasive and stereotypic behaviors (including pragmatic language) are characteristic of ASD. In contrast with OCD, perseverations lack an association with anxiety; indeed, ASD perseverations or stereotyped behaviors may contain sought-after or pleasurable topics. Similarly, the stereotyped behavior may be elicited by excitement or frustration or may serve to manage stimulation and is generally not associated with preceding intrusive thoughts or impulses. Obsessions may be described, but they are generally better characterized as idiosyncratic interests accompanied by marked perseveration and no awareness of the disinterest of others (or the excessiveness of their focus). Examples include ordering, arranging, sequencing, and preservative speech. It is not uncommon for parents to report a child appearing “driven” to engage in their “compulsions,” especially younger children. Although OCD can overlap with ASD (6% to 20% of patients with OCD are estimated to have ASD traits), the distinction can usually be made by establishing the developmental context of the behaviors.

Psychotic Disorders

The differential diagnosis of OCD from incipient psychosis in youth can be challenging. The fixed nature of bizarre or unusual obsessions can have delusional qualities and resemble schizophrenic symptomology. Although OCD-related obsessions and compulsions may appear bizarre, insight is usually preserved, and common themes are often evident. Children with OCD are typically aware that symptoms are products of their own mind (i.e., not thought insertion). OCD can usually be distinguished from schizophrenia by the absence of other symptoms, including formal thought disorder and hallucinations, and by the ability to engage in reality testing. In severe cases of OCD, reality testing may be challenged, and the boundary between obsession and delusion may be difficult to distinguish. OCD and psychotic disorders, though rarely, do co-occur.

Illness Anxiety Disorder (formerly Hypochondriasis)

If the distressing thoughts are exclusively related to fears of having a serious medical condition based on a misinterpretation of bodily symptoms, a DSM-5 diagnosis of illness anxiety disorder should be considered instead of OCD. Behaviorally, persistent need for medical evaluation and reassurance is not uncommon. However, if the preoccupation with illness is accompanied by rituals (e.g., washing, avoidance, checking based on the concern about contracting, exacerbating, or spreading illness to others), then a diagnosis of OCD may be appropriate.

COURSE AND PROGNOSIS

OCD, with onset in childhood, is a chronic condition characterized by substantial morbidity. Follow-up studies suggest that 43% to 68% of youth diagnosed with OCD continue to meet diagnostic criteria for the disorder up to 14 years after identification. A recent meta-analysis found that the remission rate may be as high as 62% (52% to 72% within the 95% confidence interval), with shorter duration of income being linked to better outcomes. Worse outcomes at follow-up were predicted by poor initial treatment response, a lifetime history of tic disorder, and a parental Axis I psychiatric disorder. The literature lacks long-term controlled treatment studies, and most follow-up data are from extant treatment trials following only a few months posttreatment. Further, this literature is complicated by multiple confounding variables, including treatment type, comorbidity, age or age of onset, and access to follow-up care. In sum, despite promising treatments for youth with OCD, long-term data are scarce, and there is considerable room for improvement with existing interventions. Early intervention appears to be critically important.

Even with remission following treatment, children with a history of OCD or subclinical OCD symptoms should continue to be monitored for relapse in symptoms. Likewise, relatives (siblings) of patients with OCD should be monitored closely and considered for treatment with the emergence of subclinical symptoms, even in the absence of more severe or impairing symptoms.

TREATMENT

Currently, there are two approaches supported in the research literature for pediatric OCD: CBT with E/RP and pharmacotherapy with SSRIs. According to the American Academy of Child and Adolescent Psychiatry, CBT or CBT with concurrent pharmacotherapy using an SSRI is considered the first-line treatment for pediatric OCD. Multiple controlled trials, meta-analyses, and systematic reviews support these indications and add further support for combined behavioral and pharmacologic approaches in more severe cases. Behavioral approaches are overwhelmingly preferential for parents (75% of parents prefer CBT only compared to CBT + SSRI, with the disparity increasing among parents of children under age 9). Recent data suggest both the robustness and durability of CBT. Overall, treatment methodology is improving for youth with OCD. The goals of treatment are now conceptualized as striving to achieve and maintain remission.

Cognitive Behavioral Therapy with Exposure and Response Prevention

The efficacy of CBT in children has been demonstrated in numerous open trials and at least six controlled trials. Indeed, treatment response rates in the extant CBT trials were quite high, ranging from 57% to 90%. One meta-analysis of 20 studies and 1,293 participants yielded an effect size of 1.21 and relative risk of improvement versus comparison intervention of 3.93. Unlike pharmacotherapies (for which relapse is not uncommon when medication is discontinued), treatment gains from CBT commonly endure after therapy is completed. Strikingly, the Nordic Long-term OCD Treatment Study (NordLOTS) trial found at 3 years post treatment 90% (*N* = 242) of youth were responders (73% in remission). Importantly, CBT with E/RP for OCD is distinguished from other "talk therapies," including role-play–based, supportive, insight-oriented, relaxation, psychoanalytic, and psychodynamic therapies, which have no supported efficacy in the extant literature. Several treatment manuals outlining CBT with E/RP for children and adolescents are available. They consist of exposure (placing the patient in situations that elicit anxiety related to their obsessions), response prevention (deterring the compulsive or ritualistic behaviors that may serve to reduce or avoid anxiety from occurring), and, in some cases, cognitive therapy (training the patient to identify and reframe anxiety-provoking cognitions). In light of the chronic nature of OCD and potential for return of symptoms following completion of acute treatment, cognitive interventions also focus on teaching management, acceptance, and/or tolerance of anxiety and distress. The first step of treatment is psychoeducation about OCD and E/RP. It is important for the child and parents to understand the rationale for exposure to fear and restraint from anxiety-reducing compensatory behaviors. Next, a "fear hierarchy" is developed rank ordering symptoms. Following this, patients are systematically exposed to the feared situations (least feared first) while being instructed not to engage in their ritualistic behavior. Over repeated exposures, the associated anxiety can dissipate through the process of autonomic habituation.

In addition, when the feared consequences of not ritualizing fail to occur, the patient's heightened expectations of harm disappear, reducing anxiety even further. In addition, inhibitory learning techniques, including prolonged exposure and a focus on disproving expectations that feared outcomes will occur (*my obsession will come true if I refrain from my rituals*), may also enhance treatment outcome.

The format and intensity of CBT with E/RP are variable. Studies suggest that weekly outpatient CBT for child OCD is equivalent in efficacy to intensive CBT. It is recommended that treatment of child OCD be conducted in vivo and in the context of the family in order to promote generalization. There is also support for telehealth-based ERP. In addition, CBT targeting OCD may also improve other anxiety and depressive symptoms.

In addition to E/RP, adjunctive cognitive strategies are utilized with pediatric patients. However, younger children and those with impaired cognitive functioning receive fewer and less sophisticated cognitive components to treatment due to developmental and cognitive limitations. A child's age, cognitive functioning, and insight into the nature of their OCD are paramount in determining the direction of treatment, as an introduction of cognitive components of therapy depends on the child's developmental level and insight.

In addition, children's rituals frequently involve parental and family participation. Moreover, E/RP requires significant out-of-session practice for understanding and extinction of OCD symptoms. Many children are unwilling or unable to complete the necessary "homework" from therapy on their own (perhaps due to anxiety, lack of motivation, or distraction). Thus, substantial parental involvement appears warranted, especially for younger children. Family treatment should include significant psychoeducation about OCD as well as mechanisms of E/RP. Parents may benefit from observing techniques presented by a skilled therapist so that they can assist their child with exercises in a supportive, noncoercive approach. In addition, this can help parents to correctly extinguish ritualistic behavior and disengage from OCD rather than scold/react/blame.

Studies suggest family-based CBT is effective for youth with OCD symptoms as young as age 3.

Table 51.1–3.
Medications and Dose Ranges for Pediatric Obsessive-Compulsive Disorder

Medication	Typical Dosage (mg/day)
Fluoxetine (Prozac)	10–80[a]
Sertraline (Zoloft)	50–200[a]
Fluvoxamine (Luvox)	50–300[a]
Paroxetine (Paxil)	10–60[a]
Citalopram (Celexa)	10–40[b]
Escitalopram (Lexapro)	5–20
Clomipramine (Anafranil)	50–200[a]

[a] Mean doses as per the American Academy of Clinical Child and Adolescent Psychiatry 2012 Practice Parameters, based on randomized controlled trials.

[b] FDA has issued some warnings for the use of citalopram in excess of 40 mg.

Pharmacotherapy

The efficacy of pharmacotherapy for OCD in pediatric populations has been demonstrated in several controlled trials with serotonin reuptake inhibitors (SRIs) and SSRIs. Table 51.1–3 lists medications and dose ranges for pediatric OCD.

Clomipramine. Clomipramine, a tricyclic antidepressant, is the most researched SRI in the treatment of pediatric OCD. In a double-blind, 8-week, placebo-controlled study, 60% of pediatric patients showed significant improvement. Patients treated with clomipramine reported a 37% mean reduction in OCD symptoms compared to 8% for the placebo group (with an effect size approaching 1.0). In another 10-week controlled trial, a significant difference was found between clomipramine and placebo, with 75% of pediatric patients showing at least moderate improvement. Other research found that clomipramine was superior to the noradrenergic reuptake-inhibiting tricyclic antidepressant desipramine. This crossover trial found that 64% of patients who initially received clomipramine during their first treatment showed relapse of OCD symptoms during desipramine treatment. Overall, a 2003 meta-analysis of pharmacotherapy trials in children identified clomipramine to be significantly superior to SSRIs in reducing OCD symptoms. Nevertheless, the risk profile, adverse effects, and required electrocardiogram and blood-level monitoring associated with tricyclic antidepressants (e.g., antiadrenergic, anticholinergic, and antihistaminergic adverse effects) are of concern with clomipramine. Thus, clomipramine is relegated to a second- or third-line treatment choice for children and adolescents and should not be used until sufficient trials of CBT with E/RP, two or more SSRIs, and combination CBT with SSRI therapy have been attempted.

Selective Serotonin Reuptake Inhibitors. Several placebo-controlled trials have demonstrated the efficacy of SSRIs. In a 20-week, double-blind, placebo-controlled trial of the SSRI fluoxetine in children and adolescents with OCD, 44% reductions in OCD symptoms were reported. A 13-week controlled trial also demonstrated the efficacy of fluoxetine, with 55% of patients treated with fluoxetine rated as much or very much improved. Another 16-week, placebo-controlled trial of fluoxetine in children reported that 57% of patients demonstrated significant improvement. Several open trials also present favorable findings for the use of fluoxetine for pediatric OCD.

Data also support the use of the SSRI sertraline for the treatment of pediatric OCD. A 12-week, multicenter, randomized, placebo-controlled trial of children and adolescents with OCD found that 42% of patients receiving sertraline were rated as much or very much improved. Further, in a 52-week, open-label extension of the previous study, 71% of children (aged 6 to 12 years) and 61% of adolescents (aged 13 to 18 years) demonstrated 25% decreases in OCD symptoms and were rated as much or very much improved.

More recently, another multisite, randomized, controlled trial found that 21.4% of pediatric patients entered clinical remission following a 12-week course of sertraline.

Finally, controlled trials have found the SSRIs fluvoxamine and paroxetine to be efficacious and well-tolerated treatments for children and adolescents with OCD. A 10-week, multicenter, placebo-controlled trial of fluvoxamine found a 42% response rate. A 10-week, placebo-controlled trial of paroxetine for pediatric patients with OCD found that 61% of patients responded to the medication. In an open-label trial of paroxetine for pediatric OCD (*N* = 335), 71% of patients demonstrated clinical improvement.

Overall, clinically significant reductions in OCD symptomology have been documented in children and adolescents using SSRIs, including fluoxetine, sertraline, fluvoxamine, and paroxetine, with a relatively minimal side-effect profile (as compared to clomipramine). Based on these findings, SSRIs are the consensus first-line medication for pediatric OCD. Although there are no controlled comparisons between these medications in children, research suggests that SSRIs are equally efficacious in children, and the specific choice should be based on the patient’s medical history, concomitant medications, and adverse events. Poor clinical response to one SSRI is not necessarily predictive of failure with other SSRIs, suggesting adequate trials of multiple SSRIs may be indicated before augmentation. In addition, significant clinical response is unlikely within the first few weeks of taking an SSRI; generally, 10 to 12 weeks at adequate dosage is necessary to fully evaluate the efficacy of the medication.

It is noteworthy that on October 15, 2005, the U.S. Food and Drug Administration (FDA) issued a black box warning for antidepressant medications to alert health care providers to an increased risk of suicidality (suicidal thinking and behavior) in children and adolescents being treated with these agents. In light of this warning, increased attention to symptoms of depression among pediatric patients being treated with antidepressant medications appears warranted. This should include a formal assessment of suicidal ideation prior to and throughout the course of pharmacologic treatment. An implication of this warning, combined with strong research support for CBT with E/RP, suggests attempting an adequate trial of the CBT prior to pharmacologic intervention. Notably, a recent large-scale Scandinavian study suggested that for CBT nonresponders, continued CBT was as effective in comparison to SSRI augmentation. Meta-analyses suggest other considerations: Adding CBT to current SSRI treatment may be effective for non- or partial responders. Sertraline and fluoxetine may be superior to fluvoxamine. SSRI activation is more likely for youth with OCD versus SNRI activation, suggesting SNRIs as a second-line psychopharmacologic option for youth who become activated on SSRIs.

It is also noteworthy that the FDA issued an advisory regarding the use of paroxetine in children and adolescents (June 19, 2003), recommending that paroxetine not be used in the treatment of major depressive disorder. Although the FDA did not comment on the use of paroxetine for OCD in pediatric patients, one in every five children and adolescents with OCD has co-occurring depression. Overall, activation, disinhibition, and possible worsening of suicidal thoughts associated with SSRI or SRI exposure may be a special concern in depressed children and certainly demand caution and careful monitoring. Activation may be more common in younger children, especially when on higher SSRI doses. For these reasons and given the availability of other safer and effective pharmacologic options, paroxetine should not be considered a first- or even second-line treatment for OCD in children and adolescents.

Current FDA approval has been granted for four medications for pediatric OCD use: sertraline, fluoxetine, fluvoxamine, and clomipramine, although the minimum approved age varies by the compound. The most frequent SSRI side effects include nausea, insomnia, headache, and asthenia. During clinical trials, the inability to tolerate treatment due to side effects was reported in 15% or less of cases.

Other Treatments

In pediatric cases that are unresponsive to CBT and trials with multiple SSRIs, pharmacologic augmentation may be considered. There is growing support for SSRI augmentation with atypical antipsychotics for refractory OCD in adult patients. For example, an 8-week, double-blind, placebo-controlled trial performed to determine the efficacy of risperidone augmentation in adults with treatment-resistant OCD found a 40% response rate. In addition, two recent meta-analyses of off-label use of neuroleptics (including haloperidol, risperidone, olanzapine, and quetiapine) in adults with OCD suggest that initial findings from open trials, controlled trials, and case studies are promising despite a need for substantial further research, especially in youth. Lithium carbonate and buspirone augmentation have generally not been efficacious augmentation strategies for treatment-refractory OCD in adults but may ameliorate depressive symptoms. Thus, augmentation may also be the case in children with comorbid conditions; for example, a tic disorder may require the addition of an α_2-agonist or neuroleptics. In sum, unlike the adult literature, there is a paucity of empirical data supporting augmentation strategies for pediatric OCD. In addition, significant undesirable side effects should be considered prior to prescribing neuroleptics and benzodiazepines in children.

Kosey is an 11-year-old male of Middle Eastern ancestry from a Southeastern urban community. He lives with his parents and older brother. Kosey originally presented to his advanced practice provider with new-onset obsessive-compulsive symptoms. The parents described their son as a perfectionist. Even as a young child, he would need to know the routine, was highly sensitive to certain sounds and textures, and frequently apologized for seemingly normal behaviors (e.g., dropping a toy, coloring outside of the lines). He would worry about his own performance, health, and safety. He was always questioning adults to ascertain, “Am I doing this right?” or “You’re not angry at me?” In addition to questioning, Kosey would always check over his homework, chores, and personal details (e.g., appearance) very carefully. However, approximately 6 months ago, more unusual and atypical behaviors developed. Kosey would rewrite homework assignments six or seven times, shredding papers if he had to erase the slightest mistake. He would change his clothes several times per day, stating that items “got dirty” when there was no evidence of any need for cleaning. His parents noticed a pile of his previously favorite sports jerseys and expensive athletic shoes stashed in the attic. When questioned, Kosey claimed they were “contaminated” and could not be worn again, despite any indication of soiling. He became late for school and activities, locking himself in his room for extended periods before leaving the house. He was evasive when questioned about this behavior but stated, “There are things I need to do to make sure things will be okay.” He later clarified that he would repeat a sequence of prayers and sayings in his head until “they felt just right,” and he could subsequently go out of the house. He refused to go to school on days he would need to change for physical education, confessing to his parents that he worried he might be “gay” (despite no signs of beginning puberty). Whereas Kosey’s elementary school teachers had tolerated tardiness and late assignments, given his history of being a stellar student, his middle school teachers did not accept late work and reprimanded him for being late. Consequently, his grades were plummeting. At the same time, his extracurricular interests appeared to be diminished, and rituals replaced the time he had spent on age-appropriate activities. Fights were erupting between Kosey and his parents because Kosey's rituals made his mother late for work, and both parents grew tired of their son's increased need for reassurance.

There was no history of OCD in Kosey's immediate family. However, his father had a history of untreated "anxiety," and his mother had suffered from "postpartum panic." His brother also struggles with panic. His uncle had been diagnosed with Tourette syndrome, and his grandmother was described as extremely "compulsive and ritualistic." Kosey's problems had never been mentioned to a health care professional before. For years, his parents felt they were just idiosyncratic aspects of his personality. Moreover, he had been functioning well, and there was no evidence of distress, except when his parents were not available to provide reassurance. Symptoms spiked during the COVID-19 pandemic, with extensive worries about contaminated packaging, air, and risks of infection and death to his family. They presented for treatment, given Kosey's increased distress and impairment. It was recommended by his nurse practitioner that he begin on 50 mg of sertraline (his mother had responded well to sertraline). Kosey's parents wondered if PANDAS was a consideration, given the spike in symptoms when Kosey's grandfather was hospitalized with COVID-19. They were reassured that PANDAS was unlikely a consideration given the long course of symptoms and lack of fluctuating course. They were reluctant to start medication but did so, given the difficulty in finding a clinical psychologist. His dose was gradually increased to 150 mg, but his parents were concerned that he was not responding, despite reading that it can take several weeks or longer for medication to work at therapeutic doses. Consequently, the family was referred to a board-certified child and adolescent psychologist in a neighboring city with expertise in child OCD. Kosey and his family presented for initial consolation with the clinical psychologist, who agreed that CBT with E/RP was indicated, and she reviewed the treatment model with the family. Kosey and his family participated in weekly sessions for about 12 weeks, focusing on exposure to his feared thoughts and prevention of ritualistic behavior in a systematic method. He was highly motivated, and most of the visits were conducted via televideo. Therapy was family based, emphasizing family-based in vivo practice of E/RP, generalization of exercises outside of session, and minimizing disruption to school. A hierarchy of feared symptoms was developed at visit two, and in vivo E/RP commenced at visit three.

After about 3 months of CBT with E/RP, Kosey reported an improved ability to manage his obsessive thoughts and to resist rituals. He reported decreased anxiety as well as urges to ritualize. His parents reported an improved understanding of OCD as well as how to react to his reassurance seeking (and other compulsive behaviors) to facilitate his recovery. The family understood the algorithm for intervening with future obsessions or compulsions as well as the potential need to resume CBT if symptoms resurged. The family chose to remain with the psychologist for a few additional sessions to address family conflict that had developed surrounding the OCD behaviors; at this time, his presentation became consistent with clinical remission. His family desired to discontinue medication but was advised to consult with a child and adolescent psychiatrist to discuss discontinuation versus continuance to best maintain remission.

FURTHER READINGS

AACAP. Practice parameter for the assessment and treatment of children and adolescents with obsessive-compulsive disorder. *J Am Acad Child Adolesc Psychiatry*. 2012;51:98–113.

Browne HA, Hansen SN, Buxbaum JD, et al. Familial clustering of tic disorders and obsessive-compulsive disorder. *JAMA Psychiatry*. 2015;72(4):359–366.

Douglass HM, Moffitt TE, Dar R, McGee R, Silva P. Obsessive-compulsive disorder in a birth cohort of 18-year-olds: prevalence and predictors. *J Am Acad Child Adolesc Psychiatry*. 1995;34:1424.

Freeman J, Sapyta J, Garcia A, et al. Family-based treatment of early childhood obsessive-compulsive disorder: the Pediatric Obsessive-Compulsive Disorder Treatment Study for Young Children (POTS Jr)—a randomized clinical trial. *JAMA Psychiatry*. 2014;71(6):689–698.

Garcia AM, Sapyta JJ, Moore PS, et al. Predictors and moderators of treatment outcome in the Pediatric Obsessive Compulsive Treatment Study (POTS I). *J Am Acad Child Adolesc Psychiatry*. 2010;49(10):1024–1033.

Ipser JC, Stein DJ, Hawkridge S, Hoppe L. Pharmacotherapy for anxiety disorders in children and adolescents. *Cochrane Database Syst Rev*. 2009;3:CD005170.

Kong XZ, Boedhoe PSW, Abe Y, et al; ENIGMA OCD Working Group. Mapping cortical and subcortical asymmetry in obsessive-compulsive disorder: findings from the ENIGMA consortium. *Biol Psychiatry*. 2020;87(12):1022–1034.

Kotapati VP, Khan AM, Dar S, et al. The effectiveness of selective serotonin reuptake inhibitors for treatment of obsessive-compulsive disorder in adolescents and children: a systematic review and meta-analysis. *Front Psychiatry*. 2019;10:523.

Lewin AB. Evidence-based assessment of child obsessive compulsive disorder (ocd): recommendations for clinical practice and treatment research. In: Farrell LJ, Ollendick TH, Muris P, eds. *Innovations in CBT for Childhood Anxiety, OCD, and PTSD: Improving Access and Outcomes*. Cambridge University Press; 2019.

Lewin AB, Wu MS, McGuire JF, Storch EA. Cognitive behavior therapy for obsessive-compulsive and related disorders. *Psychiatr Clin North Am*. 2014;37(3):415–445.

Liu J, Cui Y, Yu L, et al. Long-term outcome of pediatric obsessive-compulsive disorder: a meta-analysis. *J Child Adolesc Psychopharmacol*. 2021;31(2):95–101.

McGuire JF, Orr SP, Essoe JK, McCracken JT, Storch EA, Piacentini J. Extinction learning in childhood anxiety disorders, obsessive compulsive disorder and post-traumatic stress disorder: implications for treatment. *Expert Rev Neurother*. 2016;16(10):1155–1174.

McGuire J, Piacentini J, Lewin AB, Brennan EA, Murphy TK, Storch EA. A meta-analysis of cognitive behavior therapy and medication for child obsessive compulsive disorder: treatment efficacy, response, and remission. *Depress Anxiety*. 2015;32:580–593.

Melin K, Skarphedinsson G, Thomsen PH, et al. Treatment gains are sustainable in pediatric obsessive-compulsive disorder: three-year follow-up from the NordLOTS. *J Am Acad Child Adolesc Psychiatry*. 2020;59(2):244–253.

Mills JA, Strawn JR. Antidepressant tolerability in pediatric anxiety and obsessive-compulsive disorders: a Bayesian hierarchical modeling meta-analysis. *J Am*

Acad Child Adolesc Psychiatry. 2020;59(11):1240–1251.

Murphy TK, Brennan EM, Johnco C, et al. A double-blind randomized placebo-controlled pilot study of azithromycin in youth with acute-onset obsessive-compulsive disorder. *J Child Adolesc Psychopharmacol*. 2017;27(7):640–651.

O’Neill J, Piacentini J, Chang S, et al. Glutamate in pediatric obsessive-compulsive disorder and response to cognitive-behavioral therapy: randomized clinical trial. *Neuropsychopharmacology*. 2017;42:2414–2422.

Pazuniak M, Pekrul SR. Obsessive-compulsive disorder in autism spectrum disorder across the lifespan. *Psychiatr Clin North Am*. 2020;43(4):745–758.

Pediatric OCD Treatment Study (POTS) Team. Cognitive-behavior therapy, sertraline, and their combination for children and adolescents with obsessive-compulsive disorder: the Pediatric OCD Treatment Study (POTS) randomized controlled trial. *JAMA*. 2004;292:1969.

Piacentini J, Wu M, Rozenman M, et al. Knowledge and competency standards for specialized cognitive behavior therapy for pediatric obsessive-compulsive disorder. *Psychiatry Res*. 2021;299*:*113854.

Rozenman M, Piacentini J, O’Neill J, Bergman RL, Chang S, Peris TS. Improvement in anxiety and depression symptoms following cognitive behavior therapy for pediatric obsessive-compulsive disorder. *Psychiatry Res*. 2019;276:115–123.

Stewart SE, Yu D, Scharf JM, et al; North American Brain Expression Consortium, UK Brain Expression Database. Genome-wide association study of obsessive-compulsive disorder. *Mol Psychiatry*. 2013;18(7):788–798.

Swedo SE, Frankovich J, Murphy TK. Overview of treatment of pediatric acute-onset neuropsychiatric syndrome. *J Child Adolesc Psychopharmacol*. 2017;27(7):562–565.

Thienemann M, Park M, Chan A, Frankovich J. Patients with abrupt early-onset OCD due to PANS tolerate lower doses of antidepressants and antipsychotics. *J Psychiatr Res*. 2021;135:270–278.

Torp NC, Dahl K, Skarphedinsson G, et al. Effectiveness of cognitive behavior treatment for pediatric obsessive-compulsive disorder: acute outcomes from the Nordic Long-term OCD Treatment Study (NordLOTS). *Behav Res Ther*. 2015;64:15–23.

van den Heuvel OA, de Wit SJ. Medial frontal hyperactivation in the developing obsessive-compulsive disorder brain: an adaptive response rescued by medication-related reduction of limbic interference? *J Am Acad Child Adolesc Psychiatry*. 2018;57(6):368–369.

Vattimo EFQ, Dos Santos AC, Hoexter MQ, et al. Higher volumes of hippocampal subfields in pediatric obsessive-compulsive disorder. *Psychiatry Res Neuroimaging*. 2021;307:111200.

Zai G, Arnold PD, Richter MA, Hanna GL, Rosenberg D, Kennedy JL. An association of Myelin Oligodendrocyte Glycoprotein (MOG) gene variants with white matter volume in pediatric obsessive-compulsive disorder. *Psychiatry Res Neuroimaging*. 2021;307:111231.

▲ 51.2 Posttraumatic Stress Disorder in Children and Adolescents

JUDITH A. COHEN, M.D.

Two-thirds of American youth experience at least one potentially traumatic event (PTE, e.g., child abuse, domestic or community violence, disaster, accidents, medical trauma, war, or multiple/complex trauma). A third experience multiple traumas, increasing the risk of serious mental and medical health problems. Up to 10% develop full-blown posttraumatic stress disorder (PTSD); many more suffer from subsyndrome PTSD symptoms with equivalent functional impairments. New criteria for PTSD in DSM-5 increase attention to developmental and cultural manifestations of trauma symptoms. Genetic and psychobiologic factors also influence the development of PTSD. Recognition of the frequency with which children develop PTSD symptoms has led to improved strategies for assessing and treating traumatized children.

DEFINITION

PTSD is defined by developing characteristic features after exposure to trauma, that is, events involving a threat to life, safety, or physical integrity, experienced or witnessed directly by the child or affecting someone close to the child (criterion A). The clinical presentation varies, and not all children manifest emotional fear, helplessness, or horror as previously required. Children aged 7 years or older must have one intrusion symptom, such as recurrent trauma-related memories or dreams or intense psychological or physiologic response to trauma reminders (criterion B); one symptom of persistent avoidance of trauma-related stimuli (criterion C); two symptoms of negative alterations in cognitions such as persistent negative trauma-related beliefs or persistent posttrauma negative emotional state (criterion D); and two arousal symptoms (criterion E). The revised criteria include developmentally sensitive diagnostic criteria for children 6 years and younger. These children must have one intrusion symptom (criterion B), one symptom of avoidance or negative alteration in cognitions (criterion C), and two arousal symptoms (criterion D). The duration of symptoms must be more than 1 month and the disturbance must cause clinically significant distress or functional impairment.

HISTORY

Descriptive accounts of children traumatized by the London blitz during World War II, of the kidnapping and subsequent burial of children in their school bus, and of children exposed to a fatal schoolyard sniper attack were published by early child trauma pioneers. Subsequent empirical studies of children exposed to sexual abuse, peer suicide, cancer, burns, and terrorist attacks confirmed that increased levels of trauma exposure and lack of adequate social support strongly predict the emergence of PTSD and other psychological symptoms. The September 11, 2001, terrorist attacks, the Asian tsunami of 2004, and Hurricane Katrina have documented that peritraumatic panic, delayed evacuation, and perceived life threat to self or family members are significant risk factors for later PTSD development. In the past 25 years, a number of empirically rigorous randomized controlled trials (RCTs) have documented the superiority of trauma-focused psychosocial treatments over usual care for pediatric PTSD. Few pharmacologic RCTs have examined the treatment of pediatric PTSD.

COMPARATIVE NOSOLOGY

PTSD was first officially recognized in 1980 in DSM-III. DSM-III-R was the first version to include representative items for childhood PTSD. Importantly for children, this version increased the number of avoidance or numbing symptoms needed to meet diagnostic criteria from one to three. DSM-IV further elucidated childhood variants of symptomatology. The development of an alternative algorithm for diagnosing PTSD in preschool children and growing research led to the new diagnostic criteria in DSM-5 for children younger than 7 years. Research documents that many do not respond with horror, fear, or helplessness supported the removal of this criterion in DSM-5. Other research led to including a new cluster for alterations of mood and cognitions in DSM-5.

EPIDEMIOLOGY

Surveys document PTSD prevalence rates in the general youth population of 8% to 9%, with another 14.1% reaching subclinical levels. More relevant are prevalence rates among trauma-exposed children, as by definition, PTSD is only possible in this population. Nonreferred samples of adolescents exposed to various forms of trauma have documented that 25% to over 90% develop PTSD, depending on the trauma experienced, time since exposure, and the assessment method used. Chronic traumas, such as ongoing child abuse, or traumas that result in disruption of family and the broader social support network, such as war or refugee experiences, tend to result in a greater risk of PTSD than acute traumas, such as natural disasters. Exposure to multiple traumatic events is associated with an increase in PTSD. Female gender also tends to predict a greater risk of developing PTSD as well as other mood and anxiety disorders. Furthermore, the above rates only include children meeting full diagnostic criteria; several studies have indicated that the majority of trauma-exposed children develop PTSD symptoms severe enough to interfere with functioning, even in the absence of a PTSD diagnosis.

ETIOLOGY

PTSD is one of the few DSM disorders that require the presence of an etiologic agent (a serious traumatic event). Exposure factors also predict outcome with severity and frequency of trauma exposure, exposure to interpersonal violence, greater perceived life threat, death of a family member, greater peritraumatic panic, and delayed evacuation (if during a mass trauma experience), all predicting higher risk of developing PTSD. Nevertheless, not all children exposed to such events develop PTSD, so other etiologic factors must also be considered.

Biologic Factors

Female gender is a risk factor for developing PTSD. Children with pre-existing anxiety disorders are at increased risk of developing PTSD following trauma exposure, and a history of depression in first-degree relatives also predicts increased vulnerability to developing PTSD. Prior trauma exposure is a risk factor for developing PTSD upon exposure to a new trauma, with a greater number of traumas predicting a greater risk of the disorder. Recent studies have demonstrated that genetic risk factors interact with other risk and protective factors, such as social support, in determining whether trauma-exposed children will develop PTSD or will be resilient despite traumatic experiences. Numerous psychobiologic abnormalities have been found in children with PTSD, including complex alterations in hypothalamic–pituitary–adrenal (HPA) axis functioning, increased excretion of adrenergic and dopaminergic metabolites, smaller intracranial volume and corpus callosum area, as well as lower intelligence quotient (IQ) and memory deficits. In addition to these alterations, adults with PTSD have evidence of amygdala hyperactivity, medial prefrontal cortex underactivity, and decreased hippocampal volume. However, because of the paucity of prospective studies, it is unclear whether these findings are indicative of etiologic (causal) factors in the development of PTSD or the effects of PTSD on children’s developing brains.

Psychological Factors

Classical and operant conditioning are involved in the development of PTSD. Extreme stressors typically produce negative emotions (terror, anger, sadness) as well as physiologic arousal from the activation of the sympathetic nervous system (the “fight, flight, or freeze” response). Classical conditioning occurs when neutral cues that were present at the time of the trauma become associated or “paired” with the trauma to the extent that they, too, elicit these intense psychological and physiologic reactions. Operant conditioning occurs when the child learns to avoid trauma-related cues in order to reduce the likelihood of experiencing these distressing reactions. As avoidance behaviors are repeatedly reinforced, they may also generalize. New responses can be acquired through the pairing of exposure and inhibitory learning.

Children’s maladaptive cognitions are also relevant to PTSD development. Shame, self-blame, believing that one is undeserving of being loved, protected, or being treated well, or believing that others cannot be trusted are common maladaptive cognitions among children who experience interpersonal trauma. Such children may have difficulty differentiating between what was done to them (the traumatic experiences) and who they are (their identity). This has important implications for how such children’s trauma symptoms are manifested (i.e., in assessing complex PTSD symptoms as described below).

Social Factors

Many pretraumatic social factors, including lower socioeconomic status, economic deprivation, family dysfunction, parental separation or death, lower intelligence, minority racial or ethnic status, and family psychiatric history, have been found to predict a higher risk of developing PTSD. The absence of a strong social support network has been found to increase the risk of developing PTSD. Parental psychopathology has also been implicated as a risk factor in children developing PTSD following exposure to a variety of traumatic events. Thus, there is growing evidence that social factors may have a role in the etiology of this disorder. However, a recent study demonstrated that twins who had been sexually abused were at significantly greater risk of developing a variety of adverse outcomes than their nonabused twins, despite shared genetics and family environment. Social etiologic factors should, therefore, not be overestimated.

DIAGNOSIS AND CLINICAL FEATURES

Exposure to a Traumatic Event

To qualify for a diagnosis of PTSD, the child must have been exposed to a traumatic event that consisted of direct personal experience; witnessing an event in person; learning about the trauma of a close family member or close friend (in the case of death, the event must have been violent or accidental); or experiencing repeated or extreme exposure to aversive details of the traumatic events (but not only through media descriptions). A number of qualifying traumatic events are listed in DSM-5, which specifies that in the case of child abuse, there does not have to be a threat of physical violence or injury (e.g., genital fondling without penetration would qualify).

Re-Experiencing

The child must have at least one persistent intrusion symptom related to the traumatic event. Children may have recurrent, involuntary, and intrusive thoughts, memories, images, or body sensations reminiscent of the traumatic event. In young children, this may take the form of repetitive play in which traumatic themes are present (e.g., sexually abused children may repeatedly re-enact inappropriate touching in their play). Children may have recurrent distressing dreams about the event, or in the case of younger children, there may be frequent frightening dreams without any recognizable content. Children may have dissociative reactions; these may include periods when they act or feel as if the event were reoccurring in the present moment (a flashback). In young children, trauma-specific re-enactment may occur (e.g., sexually abused children may insert objects into their own or other children's private parts or otherwise recreate abusive acts perpetrated against them). Children may experience intense psychological distress or physiologic reactivity when exposed to triggering events, people, or situations that remind them of the original traumatic event (e.g., children who witnessed a drive-by shooting may become agitated or have difficulty breathing when they hear thunder or a backfiring car). These reminders may be highly idiosyncratic to individual children.

The above criteria also apply to diagnose PTSD for children 6 years or younger, but in these children, spontaneous and intrusive memories might not appear distressing. It may not be possible to ascertain that the frightening content of dreams is related to the traumatic event.

Avoidance and Numbing

The child must experience persistent avoidance of stimuli associated with the traumatic event and numbing of general responsiveness, as indicated by one of the following symptoms. Children may avoid or make active efforts to avoid memories, thoughts, or feelings about or closely associated with the traumatic event. They may avoid or make efforts to avoid external reminders (e.g., discussions about the traumatic event or people, places, situations, or activities that remind them of the trauma). For example, after the sudden death of a loved one, children may refuse to say the name of the deceased or insist that others not do so.

In order to diagnose PTSD for children 6 years and younger, one or more symptoms representing *either* persistent avoidance of trauma-related stimuli *or* negative alterations in cognitions and mood associated with the traumatic event must be present.

Negative Alterations in Cognition and Mood

Children may be unable to recall one or more important aspects of the traumatic event (dissociative amnesia). Children may have persistent, overly negative beliefs about themselves (e.g., "No one loves me"); others (e.g., "I can't trust anyone"), or the world (e.g., "Nothing is safe"); or maladaptive cognitions about why the trauma happened (e.g., "I should have known he would hurt me") leading to distorted self- or other-blame. Children may develop negative emotional states such as shame, guilt, anger, sadness, or fear that persist after the traumatic event. Children may have diminished interest or participation in their normal activities, for example, not wanting to play with friends (social withdrawal) or participating less in the classroom. They may feel detached or estranged from others (e.g., feel that no one understands them; affiliate with deviant peers) or may experience diminished capacity to feel emotions, especially warm interpersonal emotions (e.g., children may say, "I don't feel like I love anyone anymore.").

In order to diagnose PTSD for children 6 years and younger, one or more symptoms representing *either* persistent avoidance of trauma-related stimuli *or* negative alterations in cognitions and mood associated with the traumatic event must be present.

Arousal and Reactivity

The child must also experience persistent symptoms of arousal and reactivity that were not present prior to the traumatic event, as evidenced by at least two of the following. Children may have increased irritability and angry outbursts, with little or no provocation; these may be expressed as aggression toward others (and are often misdiagnosed in children as externalizing behavior disorders). They may have reckless or self-destructive behaviors (e.g., self-injury, substance abuse, reckless driving, or unprotected sexual activity). They may exhibit hypervigilance (e.g., children may ask their parents to check several times every night to be sure that all the doors are locked) or exaggerated startle reactions (e.g., jumping whenever the telephone rings or a pencil drops in the classroom). Children may have difficulty concentrating or completing tasks, which may impair academic performance. Children may have trouble falling or staying asleep, or may have restless sleep, whether or not they have trauma-related dreams.

The above criteria also apply to children 6 years and younger, except that these children do not typically exhibit reckless or self-destructive behaviors.

Duration, Types, and Clinical Features

These symptoms must be present for at least 1 month and must be accompanied by significant distress or impairment in an important area of functioning. PTSD with dissociative symptoms should be specified if the child meets the above criteria and experiences persistent or recurrent depersonalization or derealization symptoms that are not attributable to substance abuse. If the onset of symptoms does not occur until more than 6 months after the traumatic event, the diagnosis is PTSD with delayed expression. “Complex PTSD,” “complex trauma,” and developmental trauma disorder have been suggested to represent the complexity of symptoms existing in children experiencing repeated, interpersonal traumas involving attachment disruption. Such children have severe dysregulation of affect, behavior, cognitive, attachment, perceptual, and biologic domains that result in marked functional impairment. It is not clear whether the new DSM-5 PTSD criteria will more adequately describe these complex problems or whether a new diagnosis will be required. The International Classification of Diseases, 11th Edition (ICD-11) conceptualizes this as two distinct disorders, PTSD and Complex PTSD.

PATHOLOGY AND LABORATORY EXAMINATION

Although several alterations in the laboratory, neuropsychologic, and neuroimaging tests have been reported in children with PTSD, there is currently no laboratory or other test that can reliably distinguish children with PTSD from other children on an individual basis.

DIFFERENTIAL DIAGNOSIS

Assessment of pediatric PTSD is complex and beyond the scope of this article. PTSD symptoms overlap and co-occur with several other psychiatric conditions. Acute stress disorder (ASD) occurs and is resolved within the first month after traumatic exposure. Children with less significant trauma-related adjustment symptoms may be diagnosed with an adjustment disorder. Several diagnostic criteria of PTSD overlap with those of anxiety disorders and major depressive disorder. For example, recurrent intrusive thoughts are a hallmark of obsessive-compulsive disorder, while avoidance of social situations, social withdrawal, and tantrums are characteristic of social phobia. The specific content of these thoughts or fears may clarify that they are not trauma-focused. However, because children frequently do not disclose traumatic experiences unless specifically asked, it is important for the assessing clinician to inquire directly about traumatic exposure whenever children present with anxiety-related symptoms.

In a similar manner, irritability, feelings of emptiness, decreased interest in normal activities, sleep disturbance, psychomotor agitation, inappropriate guilt, and decreased concentration are all symptoms shared by PTSD and major depressive episodes. Distinguishing between these two disorders may be particularly difficult in children with avoidance symptoms that prevent them from discussing the traumatic event. Careful history-taking is essential in determining whether the child's symptoms are trauma-related and to what extent the child meets the criteria for either or both diagnoses. Children with traumatic bereavement are particularly likely to meet partial or full criteria for both disorders.

Children and adolescents presenting with substance use disorder may have a past history of trauma that is either covert (never disclosed) or minimized by the child or assessing clinician. Accurately identifying PTSD in such children may be critical to their optimal treatment, as standard substance use interventions do not contain important elements of trauma-focused treatment. If such children are using drugs or alcohol to self-medicate trauma-related symptoms, it is unlikely that their substance use will improve without adequately addressing these symptoms.

Additionally, PTSD may coexist with or be misdiagnosed as behavioral disorders such as attention-deficit/hyperactivity disorder (ADHD) or conduct disorder. Particularly in younger children who are less able to verbalize their anxiety, PTSD-related irritability, sleep difficulties, and difficulty concentrating may present as classic ADHD. Eliciting a history of trauma exposure and evaluating the timing of these symptoms may be important in distinguishing these disorders.

PTSD can mimic medical conditions such as hyperthyroidism, seizure disorder, caffeinism, migraine, asthma, or side effects from prescription or nonprescription medications.

Symptoms of complex PTSD (severe affective, behavioral, relationship, and cognitive dysregulation) may resemble those of bipolar disorder. Particularly if such children are avoidant of disclosing a trauma history, they will likely be diagnosed with bipolar disorder rather than PTSD. Since first-line treatments of these disorders are very different (medications for the former; psychotherapy for the latter), it is crucial to always inquire about children's trauma histories and assess their symptoms from a trauma-informed neurodevelopmental perspective.

Psychotic symptoms may occur in up to 20% of children with PTSD; flashbacks or other intrusive phenomena must be distinguished from true hallucinations or substance-induced disorders.

COURSE AND PROGNOSIS

The natural course of childhood PTSD has not been adequately determined. Based on the available studies, it appears that while PTSD symptoms spontaneously remit in some children, they persist for years or even decades in many others. There is also growing evidence that trauma-focused treatment significantly alters the natural course of pediatric PTSD. Studies have documented continued remission of PTSD 2 years posttreatment in children who received trauma-focused cognitive-behavioral therapy (TF-CBT).

The prognosis for children with untreated PTSD has become a matter of increasing concern since research has documented the numerous psychobiologic abnormalities associated with this disorder, and the severity of some of these (decreased intracranial volume, decreased corpus callosum area, and lowered IQ) is directly correlated with the duration of symptoms. Growing evidence that childhood sexual abuse is predictive of adult suicidality, as well as depression and suicidality in the offspring of such victims, also suggests a poor prognosis for unresolved PTSD in childhood. This emphasizes the importance of identifying symptomatic children and providing effective interventions early in the course of this disorder.

TREATMENT

Children with PTSD receive many treatments, but the effectiveness of most of these interventions has not been evaluated. To date, trauma-focused psychotherapies have the most empirical evidence of efficacy for treating pediatric PTSD. Although several selective serotonin reuptake inhibitors (SSRIs) have been shown to effectively treat adult PTSD symptoms, two RCTs have failed to show superiority to a placebo in traumatized children. Similarly, although open trials of adrenergic agents have had promising results, two RCTs have failed to show the superiority of propranolol to a placebo in treating pediatric PTSD. During the COVID-19 pandemic, treatment for children with PTSD quickly transitioned to being provided almost entirely via telehealth. Two studies have found that when trauma-focused psychotherapy is provided via telehealth, children experience equivalent outcomes with lower attrition than when treatment is provided in typical clinic settings.

Psychotherapeutic Treatments

Cognitive-Behavioral Therapies. Currently, TF-CBTs have the strongest evidence for resolving child PTSD and associated problems. TF-CBT has been found in 23 RCTs to be superior to other active treatments or comparison conditions in decreasing PTSD as well as depressive, anxiety, cognitive, behavioral symptoms, and functional impairment in traumatized children. TF-

CBT has been tested in youth who have experienced sexual trauma, domestic violence, terrorism, disaster, war, traumatic grief, and multiple and complex trauma in multicultural cohorts in the United States, Europe, Africa, Asia, and Australia. TF-CBT is typically provided over 12 to 25 treatment sessions. Its core components are summarized by the acronym PRACTICE: *Psychoeducation* about the child's traumatic experiences and responses; *Parental or caregiver inclusion* in treatment, including a parenting skills component; *Relaxation skills,* which are individualized for each child and parent; *Affective* modulation skills; *Cognitive* coping techniques (connections between thoughts, feelings, and behaviors are taught and practiced); *Trauma narration and processing*, consisting of carefully calibrated efforts to encourage the child to describe increasing details about the traumatic event and to identify and process inaccurate or unhelpful thoughts related to the traumatic event; *In vivo mastery* of trauma reminders, in which children with overgeneralized avoidance and fear develop graduated mastery of these through calibrated exposure to the feared stimuli; *Conjoint* child–parent sessions in which optimal communication is emphasized and practiced; and *Enhancing safety* and future developmental trajectory, in which the child develops and practices effective strategies for maintaining personal safety. A stepped-care TF-CBT model has been found to be noninferior to standard TF-CBT for improving young children's PTSD symptoms and for addressing cost and access issues for some families. Neuroimaging data suggest that changes in brain functioning are correlated with improvements in PTSD symptoms during TF-CBT treatment. A group CBT model for school settings, Cognitive Behavioral Intervention for Trauma in Schools (CBITS), has been tested in schools and found to be effective in reducing PTSD and depressive symptoms. Other CBT-based treatments, such as the UCLA Trauma/Grief Program for Adolescents and a child adaptation of Narrative Exposure Therapy, have also been tested with positive results among refugee or war-exposed children. Cognitive-based therapy for single-incident traumas also has been found to be effective.

Eye movement desensitization and reprocessing (EMDR) utilizes exposure and cognitive reprocessing interventions with bilateral eye movements. EMDR has been adapted for children and tested in two well-designed studies. A recent study comparing EMDR to TF-CBT found that both treatments were equally effective and efficient at improving children's PTSD symptoms, with TF-CBT being superior for improving comorbid psychiatric symptoms.

Psychodynamic and Attachment-Based Therapies. Child–parent psychotherapy (CPP) is an attachment-based treatment of trauma-exposed infants and preschool children and their primary caretakers. CPP has been tested in one RCT for preschoolers exposed to marital violence and found to be superior to case management in improving these young children's PTSD symptoms and in their mothers' PTSD symptoms. CPP includes the following components: Providers begin by providing developmental guidance or psychoeducation about the child's behavior and how it may be understood in light of normal development and changed through simple behavioral interventions. If needed, additional interventions are added to help the parent to understand the child's behaviors and enhance child–parent attunement. These may include behavior management, cognitive reframing, and processing, developing a joint narrative of the child and parent's traumatic experiences, and interpreting the child or parent's behaviors, emotions, and past experiences.

Psychopharmacologic Treatments

RCTs of sertraline (Zoloft), D-cycloserine, and propranolol (Inderal) have failed to show significant differences from placebo in improving pediatric PTSD. While open treatment trials have demonstrated clinical improvement in child PTSD symptoms following the use of adrenergic blocking agents (guanfacine [Tenex], clonidine [Catapres]), a dopamine antagonist (risperidone [Risperdal]), or an opiate (morphine [Roxanol]), none of these medications have been tested in well-designed double-blind, randomized trials.

Pharmacologic treatments thus lack evidence for treating pediatric PTSD. Children who have comorbid psychiatric conditions for which effective pharmacotherapy is established should receive the appropriate psychopharmacologic treatment of those disorders; it is not yet known whether such treatment will also impact favorably on coexisting PTSD symptoms.

Biologic Treatments

Since dysregulation of the HPA axis is generally found in both children and adults with PTSD, the use of a corticotropin-releasing factor (CRF) antagonist has been suggested as a potential intervention for preventing this disorder from developing or in reversing PTSD symptoms once they are present. This intervention has not yet been tested in children or adults.

C was a 15-year-old girl referred to treatment following hospitalization for a suicide attempt. From 6 to 12 years old, C had been sexually abused by her paternal grandfather. C disclosed the abuse to her parents at 12 years old, who were supportive; however, a week later grandfather committed suicide. The following year, C came out about being gay to her parents and close friends, who were supportive. Her parents thought C was doing well until the suicide attempt. In the hospital, C disclosed the following: Since coming out, C experienced worsening physical, sexual, and verbal abuse in school related to her gay identity but kept this from her parents. During the current school year, C was repeatedly sexually assaulted by three male peers who threatened to kill her younger brother if she disclosed the abuse. Guilt (she “froze” during these episodes and did not say “no” or fight back) and negative self-cognitions led to the suicide attempt.

During the initial evaluation, C reported intrusive thoughts and nightmares about sexual abuse and her grandfather’s gory suicide. C tried unsuccessfully to push these images out of her mind. She acknowledged maladaptive cognitions, including that she “deserved to be raped for killing my grandfather” and that she was “a terrible person.” She endorsed sleep disturbance, increased startle, irritability, and sadness. She scored 58 on the UCLA PTSD reaction index (RI), consistent with very severe PTSD, and 20 on the beck depression inventory (BDI), consistent with a major depressive episode. During the assessment and treatment, the therapist also used strategies from the Family Acceptance Project (https://familyproject.sfsu.edu/) to evaluate behaviors that C’s parents demonstrated related to C’s sexual orientation, and how they could expand these in order to further support and advocate for C. These strategies are described in the TF-CBT LGBTQ Implementation Manual (https://www.tfcbt.org/tf-cbt-lgbtq-implementation-manual/).

C and her parents attended 23 sessions of TF-CBT. Twenty of these were individual sessions (half of the session was spent with C and half with parents); three sessions were conjoint child–parent sessions in which the family met together. During the stabilization phase (11 sessions), C and her parents learned about trauma reminders (e.g., the threat of sexual assault would serve as a trauma reminder for C in light of her history of sexual abuse), trauma impact (e.g., fight, flight, freeze, and trauma impact on affect, behavior, biology, cognitions, relationships, and perceptions), as well as a variety of coping skills for trauma reminders. C and her parents reported that this helped them to make sense of C's response to the rapists as a trauma response and was "a huge relief" for C. Although parents and C reported that parents were "supportive" of C's sexual orientation, through further psychoeducation and parenting skills, they became more aware of specific supportive behaviors they could engage in to show this support to C (e.g., encouraging her to have LGBTQ friends over, not allowing others in their family to make verbal slurs about LGBTQ individuals, and actively advocating for her with other family members, at school, and in their faith community, etc.). As her parents used these strategies, C reported that she felt more supported and happier at home. C preferred using visualization and the cognitive triangle (changing thoughts in order to change feelings and behaviors). She carried a notebook with her and practiced changing her thoughts whenever she had a negative feeling. She found that this was very helpful and, when paired with visualization, enabled her to reduce her distress level significantly in most situations. C's parents were very upset at the school for failing to protect C from harassment and physical assaults at school, and with C's agreement, they investigated other schools that C could attend. The therapist worked with parents to become more effective advocates for C in school, their extended family, and the community. C reported that this also contributed significantly to reducing her distress level.

During the trauma narrative phase (7 sessions), C developed a detailed description and cognitively processed her personal trauma experiences. Through this process, her understanding and cognitions about these experiences changed substantially. She identified how her grandfather manipulated her into worrying about her parents' response if she disclosed it and contributed to her self-blame for his suicide. Her new cognitions were that only the perpetrators were responsible for the abuse, her grandfather chose to kill himself rather than face up to prison, and she did not deserve to be hurt for any reason. During this process, C also made the decision that she wanted to go to a different school because "they didn't protect other gay kids and me the way they should have." Her parents actively advocated for this change and helped to find an alternative school. They also prepared a brief response letter to C's narrative stating their regret for not knowing about the abuse and bullying and how they would be more actively supportive in the future.

During the integration and consolidation phase (5 sessions), C shared her narrative with her parents and engaged in safety planning with them. This included how C could protect herself from bullying going forward and how her parents could also advocate for her. C made the transition to her new school successfully and found this to be a more supportive environment. C and her parents addressed healthy sexuality together and in individual sessions, and C reflected on her future hopes and plans. At the end of treatment, C's RI score was 15, and BDI was 10 (both normal).

FUTURE DIRECTIONS

1. Evaluating DSM-5 diagnostic criteria for children ages 7 to 17 years, particularly those with complex trauma, to determine whether the new criteria adequately describe the complex difficulties of these youth or whether an additional diagnosis is needed.
2. Developing and testing effective treatments for children who have PTSD in the presence of comorbid psychiatric disorders.

3. Evaluating the relative efficacy of pharmacologic treatments for pediatric PTSD, alone and in combination with psychotherapy.
4. Developing a biomarker for pediatric PTSD and other effective neurobiologic tools for assessing pediatric PTSD.
5. Developing effective dissemination strategies for training community therapists in evidence-supported treatments for children with PTSD, including making these treatments usual care by embedding evidence-based treatments into routine graduate school training for counseling and social work students.
6. Developing and testing the use of social media applications for screening and treating pediatric PTSD. Given the large number of children who have PTSD, some will never reach mental health providers, but most have access to electronic media.

FURTHER READINGS

Caspi A, McClay J, Moffitt TE, et al. Role of genotype in the cycle of violence in maltreated children. *Science*. 2002;297:851–854.

*Cohen JA, Mannarino AP, Deblinger E. *Trauma-Focused CBT: Treatment Applications*. Guilford; 2012.

*DeBellis MD, Zisk A. The biological effects of childhood trauma. *Child Adolesc Psychiatr Clin N Am*. 2014;23(2):185–222.

Dorsey S, Lucid L, Martin P, et al. Effectiveness of task-shifted trauma-focused cognitive behavioral therapy for children who experienced parental death in Kenya and Tanzania: a randomized clinical trial. *JAMA Psychiatry*. 2020;77:464–473.

Foa EB, Keane TM, Friedman MJ, Cohen JA. Effective treatments for PTSD. *Treatment Guidelines from the International Society for Traumatic Stress Studies*. 2nd ed. Guilford; 2008.

Garrett A, Cohen JA, Zack S, et al. Longitudinal changes in brain function associated with symptom improvement in youth with PTSD. *J Psychiatr Res*. 2019;114:161–169.

*Lieberman AF, Van Horn P. *Child-Parent Psychotherapy with Infants and Young Children*. Guilford; 2008.

Meiser-Stedman R, Smith P, McKinnon A, et al. Cognitive therapy as an early treatment for post-traumatic stress disorder in children and adolescents: a randomized controlled trial addressing preliminary efficacy and mechanisms of action. *J Child Psychol Psychiatry*. 2017;58:623–633.

Salloum A, Wang W, Robst J, et al. Stepped care versus standard trauma-focused cognitive behavioral therapy for young children. *J Child Psychol Psychiatry*.

2016;57:614–622.
Saunders BE, Adams ZW. Epidemiology of traumatic experiences in childhood. *Child Adolesc Psychiatr Clin N Am*. 2014;23(2):167–184.
Scheeringa MS, Zeanah CA, Cohen JA. PTSD in children and adolescents: towards an empirically based algorithm. *Depress Anxiety*. 2011;28(9):770–782.
Stein B, Jaycox LH, Kataoka S, et al. A mental health intervention for schoolchildren exposed to violence: a randomized controlled trial. *JAMA*. 2003;290:603–611.
Stewart RW, Orengo-Aguayo R, Young J, et al. Feasibility and effectiveness of a telehealth service delivery model for treating childhood posttraumatic stress: a community-based, open pilot trial of trauma-focused cognitive behavioral therapy. *J Psychother Integr*. 2020;30(2):274–289.

51.3 Separation Anxiety, Generalized Anxiety, and Social Anxiety

BENJAMIN N. SCHNEIDER, M.D., RUBEN G. MARTINEZ, PH.D., AND TARA S. PERIS, PH.D.

Fear and worries are a common and expected part of healthy child development. For most children, these worries are short lived and limited in the impairment they bring, and they track to developmental status. Among younger children, fears may focus on the dark, monsters, or strangers, whereas for older children and adolescents, fears may focus on friends, school, or decision making. These developmentally appropriate worries wane naturally over time as new experiences and skills overshadow them. Although transient fears and worries are a natural and necessary part of child development, for some youth, they will persist and bring a level of impairment that requires clinical attention.

The three most common pediatric anxiety disorders—separation anxiety, generalized anxiety, and social anxiety—share the same underlying construct of diffuse, future-oriented worry that is often focused on unknown threats. They have significant symptom overlap, are highly comorbid with one another, and show similar familial relationships and continuity with adult anxiety and depressive disorders. For these reasons, the three disorders are often grouped together as they are in this section. The three anxiety disorders are described separately in this chapter to highlight core constructs central to diagnosis; however, treatments are discussed for the disorders as a group, given the significant overlap in approach.

Anxiety disorders comprise the most prevalent class of mental health problems among youth, affecting up to one in three individuals by the time they reach adulthood. Children with anxiety disorders have an early onset and a persistent or recurring course with high levels of short- and long-term impairment. In addition to being comorbid with one another, anxiety disorders are often comorbid with nonanxiety disorders. Fortunately, there are highly effective psychosocial and pharmacologic treatments for these disorders. Current challenges include the identification of treatment methods to optimize outcomes and address co-occurring emotional and behavioral problems and the wide-scale dissemination and implementation of evidence-based treatments.

SEPARATION ANXIETY DISORDER

Definition

Separation anxiety disorder is characterized by excessive and developmentally inappropriate fear of catastrophic separation, avoidance of such separations, and irrational beliefs about the consequences of being apart from a caregiver (Table 51.3–1). Attachment figures are usually parents but may also be siblings, grandparents, or other caregivers or for adults, children or spouses. Whereas developmentally appropriate separation anxiety typically presents around age 6 months and declines between ages 2 and 3 years, children with separation anxiety disorder have either persistent and worsening or new-onset separation anxiety in the school-age years (i.e., ages 6 to 12 years). This section differentiates between developmentally normal separation anxiety in infancy and toddlerhood and separation anxiety disorder in school-aged children.

Comparative Nosology

Separation anxiety disorder appeared initially in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) and was classified as an anxiety disorder of childhood or adolescence. Although it was subsequently categorized in the DSM-IV as a “disorder usually first diagnosed in infancy, childhood, or adolescence,” separation anxiety is currently classified in the DSM-5-TR as an anxiety disorder. Since first introduced, the diagnostic criteria for separation anxiety disorder have been modified in four ways, along with modified wording in DSM-5 (continued in DSM-5-TR) to capture symptom presentation in adulthood. Unlike earlier versions, the DSM-5 does not require onset prior to age 18 and recognizes that some adults report onset after age 18 years. The minimum duration was increased from 2 weeks in DSM-III to 4 weeks in DSM-IV and subsequent versions, and DSM-5 specifies at least 6 months in adults to avoid overdiagnosis of transient fears. Also, evidence of serious distress or impairment was added for diagnosis starting in DSM-IV. Finally, recurrent excessive distress about (1) anticipated or (2) actual separation was collapsed into one symptom in DSM-IV, and thus the symptomatic threshold for separation anxiety disorder was reduced from three of nine symptoms in DSM-III-R to three of eight symptoms in DSM-IV and currently in DSM-5-TR.

Table 51.3–1.

DSM-5-TR Diagnostic Criteria for Separation Anxiety Disorder

A. Developmentally inappropriate and excessive fear or anxiety concerning separation from those to whom the individual is attached, as evidenced by at least three of the following:
   1. Recurrent excessive distress when anticipating or experiencing separation from home or from major attachment figures.
   2. Persistent and excessive worry about losing major attachment figures or about possible harm to them, such as illness, injury, disasters, or death.
   3. Persistent and excessive worry about experiencing an untoward event (e.g., getting lost, being kidnapped, having an accident, becoming ill) that causes separation from a major attachment figure.
   4. Persistent reluctance or refusal to go out, away from home, to school, to work, or elsewhere because of fear of separation.
   5. Persistent and excessive fear of or reluctance about being alone or without major attachment figures at home or in other settings.
   6. Persistent reluctance or refusal to sleep away from home or to go to sleep without being near a major attachment figure.
   7. Repeated nightmares involving the theme of separation.
   8. Repeated complaints of physical symptoms (e.g., headaches, stomachaches, nausea, vomiting) when separation from major attachment figures occurs or is anticipated.

B. The fear, anxiety, or avoidance is persistent, lasting at least 4 weeks in children and adolescents and typically 6 months or more in adults.

C. The disturbance causes clinically significant distress or impairment in social, academic, occupational, or other important areas of functioning.

D. The disturbance is not better explained by another mental disorder, such as refusing to leave home because of excessive resistance to change in autism spectrum disorder; delusions or

hallucinations concerning separation in psychotic disorders; refusal to go outside without a trusted companion in agoraphobia; worries about ill health or other harm befalling significant others in generalized anxiety disorder; or concerns about having an illness in illness anxiety disorder.

Reprinted from the *Diagnostic and Statistical Manual of Mental Disorders*, *Fifth Edition* (Copyright © 2013). American Psychiatric Association. All Rights Reserved.

Epidemiology

Epidemiologic studies show the lifetime prevalence rate of childhood and adolescent separation anxiety disorder is at least 4%, with peak onset between 7 and 9 years of age. Separation anxiety disorder is more common in childhood than in the preschool years or adolescence and is the most common anxiety disorder in children under the age of 12. Although the peak age of onset is in middle childhood, preschool children can be diagnosed with separation anxiety disorder (range from 0.5% to 10.5%) and can be distinguished from children with developmentally normal separation anxiety. Children with separation anxiety symptoms who do not meet diagnostic criteria are very common, with reported rates as high as 50%. Community studies suggest that girls are more likely than boys to meet criteria even in the preschool years, but clinical studies find the disorder to be equally common in girls and boys.

Etiology

Although the exact cause of separation anxiety is unknown, genetic, early temperamental, and family or environmental variables have all been associated with early-onset anxiety disorders. Although genetic factors likely play a significant role, early experiences, many of which are family-based, may play an important role in the onset of childhood anxiety disorders. Methods to sort the various contributions of genetics, temperamental and biologic vulnerabilities, and family or environment are often confounded. However, it is likely that genetically based, temperamental, or biologic vulnerabilities predispose affected children and their parents to interact in ways that either protect children from or predispose children from developing an anxiety disorder. For example, children with separation anxiety raised by a presently anxious parent may differ in outcome from children with separation anxiety raised by parents whose anxiety is well controlled. Although in both situations, the child and the mother share genetic, temperamental, and biologic vulnerabilities, the effect of the child's anxiety on the affected (and treated) parent and the affected (and treated) parent's impact on the child may account for significant differences in both short- and long-term outcomes.

Twin studies provide perhaps the greatest evidence for the role of biologic and environmental factors (shared and nonshared) in the etiology of childhood anxiety disorders. Although historically both genetic and environmental factors have been implicated in the etiology of childhood anxiety disorders, a large well-designed study of 6-year-old twins suggests significant genetic influence for the separation anxiety phenotype. The heritability estimate for separation anxiety was high, 73%, with the remaining variance attributed to nonshared environment, suggesting genetic effects on separation anxiety are substantial and more significant than environmental effects. Another investigation of the role of genetic and shared environment effects in a sample of twins and siblings over time reported that the influence of genetic effects increased and shared environmental effects decreased with age. Finally, newly emerging data suggest that the psychobiologic trait of CO_2 hypersensitivity may increase risk for separation anxiety (and supports a proposed genetic association between separation anxiety and later-onset panic disorder).

Behavioral inhibition (BI) to the unfamiliar, characterized by fear, withdrawal, and avoidance of unfamiliar situations or things, is considered a temperament vulnerability in infants and toddlers that predicts anxiety disorder outcomes in later childhood. BI is usually assessed using observational methods in laboratory settings for children aged 9 months through 6 years. Approximately 10% to 15% of children are thought to display BI and longitudinal outcomes, suggesting young children with BI are more likely than uninhibited children to exhibit clinical levels of multiple anxiety disorders including separation anxiety upon long-term follow-up. In cross-sectional studies, BI is associated with a range of anxiety symptoms and disorders in youth, and it has traditionally been considered a vulnerability that lacks specificity for any childhood anxiety disorder. However, longitudinal data found BI to predict *social anxiety disorder (SAD)* in middle childhood, but not other anxiety disorders.

Parental psychopathology is also highly associated with childhood anxiety disorders. Multiple pathways are operative in linking parent and child well-being. Parents and children may share the same genetic or temperamental vulnerability. Anxious parents are likely to model anxiogenic thoughts and behaviors that are readily adopted by their children. Bidirectional influence may also be operative inasmuch as anxious youth may pull for adults in their environment to maintain proximity and support. Regardless of the mechanism, children of parents with an anxiety disorder are up to five times more likely to be diagnosed with an anxiety disorder. More specifically, parental panic disorder and major depression are associated with an increased risk for separation anxiety disorder in children.

Longitudinal studies examining the role of infant–mother attachment in the development of childhood anxiety often used the Ainsworth Strange Situation to assess attachment and then follow children over time. In multiple studies, children with an anxious or resistant attachment during infancy were more likely to develop anxiety disorders than were securely attached infants. Although infant temperament or maternal psychopathology may affect attachment behavior, cross-sectional assessment of attachment status appears to be the more important factor in determining the presence of anxiety disorders at longer-term outcome.

Several decades of research identify a common set of parenting behaviors found among parents of anxious youth. These include high levels of symptom accommodation, parental intrusiveness, overprotection, and psychologically controlling behavior in the child’s daily activities. Such anxiety-enhancing parenting approaches are thought to promote dependency and reduce the development of autonomy. Controlling and overprotective parenting may contribute to the development of anxiety by preventing the child from having a range of experiences that foster broad and complex coping skills. Furthermore, parental overprotection may communicate to the child that the world is a dangerous place, introducing a chronic and pervasive response bias. Parents may also support anxious behavior by encouraging anxious cognitions and avoidance in their children or children may mimic anxious cognitions and behaviors modeled by their parents.

Diagnosis and Clinical Features

The defining feature of separation anxiety is excessive and persistent worry about harm befalling an attachment figure (e.g., parent getting into car accident, spouse becoming ill or dying) or oneself (e.g., getting lost, being kidnapped, having an accident) upon separation. Symptoms may include avoidant behaviors such as refusal to go to school or work, sleep alone, or be without the attachment figure, in order to minimize separation and to reduce distress. Other common symptoms include nightmares with references to separation, and physical symptoms such as stomachaches or nausea in response to possible or actual separation. DSM-5-TR requires three of eight symptoms, a duration of 4 weeks in children and adolescents, and corresponding functional impairment. The diagnosis can be made in adults who are typically excessively concerned about their spouses or children and symptomatic for at least 6 months.

At home, children with separation anxiety frequently shadow their caregivers. They often are reluctant to be in any room alone and may not want to sleep alone in their bedroom. These children often request a parent to fall asleep with them or find their way to their parent’s bed at night. They are uncomfortable staying with babysitters and may call or text parents frequently. In public, children with separation anxiety stay close to their caregiver and prefer to sit on their lap or be in physical contact. They are reluctant or refuse to attend birthday parties, sleepovers, camps, and extracurricular activities if the attachment figure is not present. Somatic complaints, especially gastrointestinal (e.g., stomachaches, nausea) and cardiovascular (e.g., pounding heartbeat, light-headedness) are commonly reported, particularly before other anxiety-provoking separations. Children may show their separation anxiety distress with tantrums or becoming panicky or oppositional when anticipating separation.

At school, the child may make repeated requests to call their parents to check in or visit the school nurse with physical complaints. When children experience anxiety at school they may become restless and fidgety and be so preoccupied that they are unable to attend to classroom work. Separation anxiety disorder is also a common cause of school refusal and may result in frequent tardiness or high numbers of school absences.

Younger children, in comparison to adolescents, with separation anxiety disorder tend to report a greater number of symptoms and more nightmares with separation themes. Some older children and adolescents with separation anxiety may not present obvious impairment or distress, as their parents and school find ways to accommodate the child's fears to decrease their distress.

As separation anxiety not only affects the child, disruption of family life or parent stress may result from accommodating and managing the child's anxious and avoidant behavior. It can be helpful in the assessment to attempt to quantify the amount of accommodation and to ascertain what the anxiety severity would be if the parents did not offer ongoing accommodation around the child's anxiety.

Posey is a 9-year-old Latina female living with her biologic parents and sister in a large metropolitan region. She presents for care with her mother, who reports that Posey has rarely, if ever slept in her own bed. Posey's difficulty sleeping in her own bed is disrupting family life as evidenced by parents who are losing sleep and coparenting/family conflict that has increased. At a recent slumber party, Posey began to complain of a stomachache, asking her host to call her mother to take her home. She reported being embarrassed that she had to leave her friend's house, but that she "needed" to be with her parents. Upon interview, Posey reports that she worries about the safety of her family ("something bad" happening to her parents) and that when she does not sleep near or with her parents, she has nightmares that someone will "break in and hurt them." She reports associated difficulty sleeping, and a near-constant level of stress when not in the presence of both parents. Both parents describe her as clingy and note that she needs them to stay present at extracurricular activities like soccer and piano.

Differential Diagnosis

Separation anxiety disorder can be distinguished from developmentally normal separation anxiety based on its onset or increasing severity in the school-age years, the duration of symptoms, and the degree of impairment or distress or need for parental accommodation.

Separation anxiety should also be differentiated from other emotional and behavioral disorders. Children with separation anxiety, generalized anxiety, and obsessive-compulsive disorder (OCD), all have concerns about harm. Children with separation anxiety worry about events that lead to being apart from an attachment figure. Children with generalized anxiety are more fearful of the actual event such as a dog bite, lightning striking, or getting sick than the impact on separation. Children with OCD have repetitive intrusive and senseless thoughts of harm befalling them or their loved ones if they do not complete a ritual. Children with separation anxiety may have social avoidance (e.g., parties or school) similar to children with SAD, generalized anxiety disorder (GAD), OCD, or even post-traumatic stress disorder (PTSD), but the rationale for the avoidance will differ for each of the disorders. Children with separation anxiety may be described as panicking upon separation or anticipation of separation, but the panic is distinguished from spontaneous panic attacks that characterize panic disorder.

Finally, children with separation anxiety may exhibit defiance, irritable or labile mood, and explosive behavior when presented with a situation that involves separation. These behaviors or moods should be distinguished from oppositional defiant disorder, which involves more general patterns of argumentativeness and misbehavior, and from bipolar disorder, which has prominent elevation in mood, decreased need for sleep, grandiosity, and increased goal-directed activities.

Course and Prognosis

The course of separation anxiety is not consistent for all children. Multiple rigorous prospective studies suggest that the majority of children with separation anxiety improve with time. Indeed, some appear to spontaneously remit while some see symptoms and impairment gradually wane. For others, separation anxiety symptoms persist over time and transfer to other close relationships. In still other cases, symptoms of separation anxiety resolve but are replaced with symptoms of another anxiety disorder. In some cases of separation anxiety, affected individuals may see symptoms evolve into more adult patterns of anxiety or depression. Prospective, longitudinal studies including both community and high-risk samples have found that those meeting DSM-IV criteria for separation anxiety disorder were at an increased risk of developing subsequent panic disorder with agoraphobia, specific phobia, GAD, OCD, pain disorder, and alcohol dependence, even after controlling for the risk of parental psychopathology. Other meta-analyses indicate an increased risk for all anxiety disorders, but less evidence for elevated risk for depression and substance use.

SOCIAL ANXIETY DISORDER

Definition

SAD, also called social phobia, is characterized by excessive, persistent, and impairing anxiety about potential embarrassment or catastrophic humiliation in social or performance situations. The worry may be restricted to a specific performance situation such as fear of being negatively evaluated during an athletic or musical performance. However, SAD is more commonly generalized such that worry about being scrutinized or embarrassed occurs across a range of activities and situations.

Comparative Nosology

SAD (social phobia) first appeared in DSM-III, with minimal alterations in subsequent editions. One impactful change from DSM-III to DSM-IV was the removal of the diagnosis of avoidant disorder. Avoidant disorder was classified as an anxiety disorder of childhood or adolescence, and characterized by fear and avoidance of strangers resulting in significant impairment in peer relationships. Concern about the overlap between the two disorders and the continuity between childhood avoidant disorder and adult SAD resulted in the removal of avoidant disorder as a diagnostic category in DSM-IV. In DSM-5 (continued in DSM-5-TR), there is no longer a requirement that individuals over age 18 must recognize their fear as excessive and unreasonable; instead, the anxiety must be out of proportion to the actual threat posed by the event after taking into account sociocultural factors. Whereas the 6-month duration was limited in DSM-IV to individuals under age 18, this criterion now applies to all ages. Finally, unlike DSM-IV, there is no longer a specifier for “generalized” but there is a specifier for “performance only” in DSM-5-TR.

Epidemiology

Social anxiety is a common anxiety disorder, with a cumulative incidence of 11%. Prevalence rates have increased over time, likely because of the removal of the avoidant disorder diagnostic category from DSM-III to DSM-IV; that is, youth who were once characterized with avoidant disorder are now diagnosed with SAD, resulting in increased prevalence rates. According to data from the National Comorbidity Survey Replication Adolescent Supplement and the Early Developmental Stages of Psychopathology Study (EDSP), the 12-month prevalence for adolescents is 5.2% to 8.2%. The lifetime prevalence of social anxiety is 9.5% for females and 4.9% for males. Like generalized anxiety, the prevalence of SAD increases with age, possibly reflecting changes in cognitive and emotional development (i.e., the capacity for self-evaluation and self-consciousness).

Age of onset for SAD is during adolescence. The average age of onset in epidemiologic studies is between 10 and 16.6 years and varies depending on the sample studied and the strategy for determining retrospective reports for the age of onset. According to the EDSP which involved a community sample of 3,021 individuals aged 14 to 24 years, the incidence of DSM-IV social anxiety was greatest among those aged 10 to 19 years.

Etiology

Similar to both separation anxiety disorder and GAD, factors that contribute to the development of SAD include genetics, temperament, and family variables. Twin studies are useful in distinguishing between environmental and genetic influences. There are many discrepancies in the literature regarding the contributions of genetic and environmental influences because of differences in measurement (i.e., rating scale vs. psychodiagnostic interviews), informants (i.e., self vs. parent reports), and criterion (i.e., symptoms vs. severity vs. clinical diagnosis). However, the literature is consistent in showing that anxiety disorders are heritable and that both shared and nonshared environmental factors also play a role. Concordance rates for anxiety disorders including SAD are higher for monozygotic twins than dizygotic twins (34% vs. 17%). Familial transmission is also evident by studies showing a high concordance rate of social anxiety among first-degree relatives. Overall, studies investigating genetic transmission find evidence for an inherited propensity toward anxiety in general rather than for a specific type of anxiety.

As discussed earlier, BI is a temperamental construct characterized by increased physiologic reactivity and avoidant behavior in the face of novel or challenging stimuli. Inhibited children have greater sympathetic activation than uninhibited children, evidenced by higher and more stable heart rates and greater heart-rate acceleration under stressful and novel conditions. Many studies support the existence of a specific link between BI and social anxiety. Both longitudinal and retrospective studies show that toddlers with BI assessed by observational or questionnaire methods are more likely than uninhibited children to exhibit social anxiety as adolescents. In a recent prospective longitudinal study, 215 children assessed for BI were assessed for psychiatric outcomes in a 5-year follow-up at a mean age of 9.6 years. Structured diagnostic interviews revealed that BI present at age 4 or 6, but not at age 21 months, predicted social anxiety. BI was not associated with other specific anxiety, mood, or behavior disorders. In another study, adults from a clinical sample provided retrospective reports of childhood inhibition, and those with the highest levels of inhibition also had the highest levels of clinical SAD. It is important to note that less than 50% of toddlers and young children characterized as behaviorally inhibited go on to develop SAD. Similarly, children who are considered shy or have shy temperaments most likely do not develop SAD. Thus, BI is considered a manifestation of a biologic predisposition that serves as a risk factor for anxiety.

Environmental variables such as parent–infant attachment and parenting behaviors also are associated with the expression of SAD. As stated in previous sections, retrospective and observational research shows links between insecure infant attachment and the development of anxiety disorders. One study examined 172 adolescents 16 years after participating in the Ainsworth Strange

Situation test at age 12 months and confirmed that infants with insecure attachments were more likely to meet criteria for anxiety, including SAD. Adult parents with SAD may sponsor social isolation by discouraging opportunities for social engagement. They also provide information that has the potential to heighten anxious arousal.

Diagnosis and Clinical Features

The defining feature of social anxiety is excessive fear about being scrutinized in social or performance situations (Table 51.3–2). Although the same criteria apply to children and adults, some provisions are indicated for children. First, social anxiety cannot be limited to adult settings and must occur in peer contexts. Second, in children, the anxiety evoked by the social situation may involve crying, tantrums, freezing, or shrinking from social situations with unfamiliar people. Common fears related to social anxiety include public speaking, participating in class, initiating or participating in conversations, attending social events, and eating or writing in front of others. Fear of public speaking is considered the most commonly feared social situation; however, test anxiety is commonly feared among adolescents.

Table 51.3–2.

DSM-5-TR Diagnostic Criteria for Social Anxiety Disorder

A. Marked fear or anxiety about one or more social situations in which the individual is exposed to possible scrutiny by others. Examples include social interactions (e.g., having a conversation, meeting unfamiliar people), being observed (e.g., eating or drinking), and performing in front of others (e.g., giving a speech).
Note: In children, the anxiety must occur in peer settings and not just during interactions with adults.
B. The individual fears that he or she will act in a way or show anxiety symptoms that will be negatively evaluated (i.e., will be humiliating or embarrassing; will lead to rejection or offend others).
C. The social situations almost always provoke fear or anxiety.
Note: In children, the fear or anxiety may be expressed by crying, tantrums, freezing, clinging, shrinking, or failing to speak in social situations.
D. The social situations are avoided or endured with intense fear or anxiety.
E. The fear or anxiety is out of proportion to the actual threat posed by the social situation and to the sociocultural context.
F. The fear, anxiety, or avoidance is persistent, typically lasting for 6 months or more.
G. The fear, anxiety, or avoidance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
H. The fear, anxiety, or avoidance is not attributable to the physiological effects of a substance (e.g., a drug of abuse, a medication) or another medical condition.
I. The fear, anxiety, or avoidance is not better explained by the symptoms of another mental disorder, such as panic disorder, body dysmorphic disorder, or autism spectrum disorder.
J. If another medical condition (e.g., Parkinson’s disease, obesity, disfigurement from burns or injury) is present, the fear, anxiety, or avoidance is clearly unrelated or is excessive.

Specify if:

Performance only: If the fear is restricted to speaking or performing in public.

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 229, 230 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Some differences have been noted with respect to the presentation of social anxiety among different age groups. Adolescents endorse anxiety across a broader range of situations and more avoidance compared to children; however, both groups experience similar levels of distress. In addition, adolescents endorse more cognitive symptoms of social anxiety that likely reflect their greater cognitive maturity compared to children. Poor social skills, however, are evident across all youth with SAD.

Infants and toddlers cannot meet DSM-5-TR anxiety diagnosis guidelines primarily due to developmental issues; however, many very young children exhibit higher than normal levels of social fear and associated distress. For example, some children have difficulties playing with new children, refuse to participate in playgroups and classes, play alone at parties or on the playground, and refuse to eat in restaurants or attend public events because of severe distress. Some pilot data reveal that children aged 18 months to 5 years can be diagnosed with a modified version of SAD and are more likely than their nonanxious counterparts to display high levels of shyness, inhibition, and to have anxious parents. Many preschool-aged children meet DSM-5-TR criteria for selective mutism which involves restricted speech in social situations due to high levels of fear and anxiety.

John is a 16-year-old Caucasian male who lives with his biologic mother in a suburban area. He presents for care after experiencing an “anxiety attack” when asked to answer a question in class last week. Upon interview, John reported that he was randomly selected by his teacher to answer a difficult math question, at which point he “locked up,” began breathing heavily, and refused to respond. Upon further interview, he reported that he never raises his hand in class because “if I get the answer right, people will think I’m a nerd, and if I get the answer wrong people will think I’m stupid.” He reports that similar worries (e.g., being seen as weak or stupid) have stopped him from trying out for the football and soccer teams, as well as the debate club. He reports having few friends outside of his childhood friends, and his mother tells you she is worried that he is missing out on the high school experience.

Differential Diagnosis

SAD can be distinguished from other disorders based on the predominant feature that fears center around the potential for embarrassment or receiving negative criticism. Although social anxiety and separation anxiety may both involve avoidance of social situations, the fear with separation anxiety is associated with concern about being apart from a caregiver and is not associated with potential humiliation or embarrassment. The generalized type of social anxiety is differentiated from GAD on the basis of the individual feeling humiliated or embarrassed in public rather than a more global worry. Also, children with GAD may worry about a host of social issues (e.g., friendships, performance, how others view them), but worries are not limited to social or performance situations. Although individuals with SAD may experience panic in the face of challenging situations, the fear and associated somatic and cognitive symptoms occur in response to specific situations and are not unexpected or uncued as they are in panic disorder.

Social anxiety is also distinguished from selective mutism, which is characterized explicitly by a failure to speak in situations where speech is expected and has a much younger age of onset relative to SAD. Selective mutism is considered by some professionals to be a developmental variant of social anxiety.

Finally, it is important to distinguish SAD from developmentally normal fears. For example, it is not uncommon for children to be anxious about speaking in front of the class. However, normative anxiety does not result in significant impairment in functioning or clinically significant distress as with SAD.

Course and Prognosis

SAD that emerges by the early teen years is associated with progressive worsening or persistence of symptoms through early adulthood. Late adolescence is characterized by a waxing and waning course during which symptoms are exacerbated by social stressors. Factors associated with persistence of symptoms include early onset, more anxiety cognitions, severe avoidance and impairment, co-occurring panic, parental social anxiety and depression, BI, and harm avoidance.

Few individuals experience spontaneous remission; data from the prospective longitudinal EDSP study suggest a natural remission rate of 15.1%. Individuals with SAD withdraw from or avoid situations, limiting opportunities for optimal social, academic, and emotional development. Research shows that children with SAD are less socially skilled than their nondisordered counterparts, limiting their capacity to form intimate relationships with friends, family members, or partners. Symptoms and impairment at school may result in academic underachievement, including failure to graduate high school or pursue college. Disabilities in educational and social functioning may lead to demoralization and other emotional problems. Longitudinal studies show, for example, that SAD is consistently linked with increased risk for subsequent depression. Further, individuals with comorbid social anxiety and depression are at increased risk for suicidal ideation and suicide attempts.

Several longitudinal studies find that SAD is associated with the development of substance use problems. For example, one prospective study based on a subsample of 816 individuals from the Oregon Adolescent Depression Project found that SAD is a unique risk factor for the subsequent onset of alcohol and cannabis dependence. Specifically, a lifetime history of SAD at age 17 was associated with 4.5 times greater odds of alcohol dependence and 6.5 times greater odds of cannabis dependence by age 30. Also, results from a different prospective study showed that youth with SAD were significantly more likely 4 years later to smoke cigarettes and have a higher rate of nicotine dependence than a control group.

GENERALIZED ANXIETY DISORDER

Definition

In contrast to separation and social anxiety, which generally have one anxiety-triggering experience (i.e., separation and social humiliation, respectively) or focal fear, GAD is characterized by excessive and uncontrollable worry about a diffuse set of events. These events are often future oriented and center on possible poor outcomes related to school, friendships, health, and world events. In addition, for many youths, generalized anxiety focuses on living up to an internal standard (in school, friendships, extracurriculars) that is difficult to meet. Symptoms of this disorder occur with frequency and are experienced as difficult to stop. In addition, affected youth may report poor concentration, restlessness, and irritability, but other symptoms include fatigue, sleep disturbance, or muscle tension. As with other anxiety disorders, generalized anxiety must cause marked distress, interfere with social, emotional, and educational functioning, and is not simply a reaction to recent stressors.

Comparative Nosology

Prior to DSM-IV, a diagnosis of generalized anxiety was reserved for individuals 18 years or older, and children with multiple fears and worries could be diagnosed with overanxious disorder. Overanxious disorder was characterized by excessive or unrealistic worry for a period of at least 6 months. Due to the poor reliability and validity of this diagnostic category and symptomatic overlap and continuity with adult generalized anxiety, overanxious disorder was incorporated into the DSM-IV GAD diagnosis. Research comparing cases diagnosed by DSM-III-R and DSM-IV criteria found that the change in terminology did not affect the characteristics of cases. Criteria remained unchanged for DSM-5 or DSM-5-TR.

Epidemiology

GAD is a common diagnosis in children, with a reported lifetime prevalence of 9.0%. Community samples of preschool children report a prevalence range of 0.5% to 6.5% for overanxious disorder or GAD. Based on data from two community epidemiology studies, the 3-month prevalence rate for GAD among children aged 9 to 12 is 1.4%. The prevalence for adolescents aged 13 to 18 years ranges between 2.3% and 5.0%. Whereas rates of separation anxiety decline with age, rates of generalized anxiety increase with age, with a mean age of onset of 8 to 9 years, and median age of onset of 12 years. Girls are twice as likely as boys to meet diagnostic criteria.

Pediatric GAD is commonly associated with other anxiety disorders and mood disorders. Data from the National Comorbidity Study—Adolescent Supplement (NCS-A) and Great Smoky Mountains Study (GSMS) found that up to 2/3 of affected youth met criteria for another anxiety disorder, and about half met criteria for a mood disorder. Preschoolers with generalized anxiety have the highest rates of comorbidity compared to preschoolers with other anxiety disorders. Few significant differences in rates and patterns of comorbidity across age and gender have been identified; younger children are more likely than adolescents to have separation anxiety as a comorbid diagnosis, and males are more likely than females to have an externalizing disorder as a co-occurring diagnosis.

Etiology

Similar to the other anxiety disorders, both genetic and environmental factors contribute to the development of GAD. However, heritability estimates are lower for generalized disorder (approximately 30%), suggesting that environmental factors, both shared (e.g., uterine environment in the case of twins, exposure to common parenting in the case of siblings) and nonshared (e.g., personal traumatic experience, peer relationships), may have a more prominent etiologic role. Twin studies show that generalized anxiety may be better explained by additive genetic factors and nonshared environmental factors compared to shared environment factors. Low heritability estimates in twin studies and high familial rates of generalized anxiety suggest that increased rates in families may be more related to environmental factors. For example, parental anxiety, parenting approaches, or exposure to stressful life events may negatively affect childhood anxiety outcomes. Similarly, childhood anxiety may also impact parental or sibling anxiety as the child's anxiety may change the larger home environment.

Anxiety sensitivity is a temperamental or trait-like factor under genetic influence that is considered a risk factor for anxiety in general. It is characterized by a tendency to interpret physiologic symptoms as a sign that something harmful will result. Such factors, while not specific to anxiety, may predispose children in certain environments to become symptomatically more anxious and impaired. Similarly, an attentional bias toward threatening cues and a tendency to interpret ambiguous situations as threatening are associated with children with current anxiety disorder. Although such relationships cannot be considered causal, it is possible that such attentional biases coupled with environmental variables such as parenting influences (e.g., hearing information from a parent about the potential dangers of a situation) may ultimately shape anxiety-related beliefs and behaviors. Anxiety sensitivity has been related specifically to panic in adolescents and otherwise has been linked to anxiety in general.

Diagnosis and Clinical Features

The hallmark feature of GAD is excessive worry that is difficult to control (Table 51.3–3). The threshold for diagnosis differs slightly for adults and children; adults are required to experience three of six associated symptoms, whereas children are required to have only one associated symptom. A diagnosis also requires the child to experience excessive worry more days than not for 6 months, the worry and associated symptoms must interfere with daily functioning, and the anxiety is not triggered by recent stressful events.

Table 51.3–3.
DSM-5-TR Diagnostic Criteria for Generalized Anxiety Disorder

A. Excessive anxiety and worry (apprehensive expectation), occurring more days than not for at least 6 months, about a number of events or activities (such as work or school performance).

B. The individual finds it difficult to control the worry.

C. The anxiety and worry are associated with three (or more) of the following six symptoms (with at least some symptoms having been present for more days than not for the past 6 months):

Note: Only one item is required in children.

1. Restlessness or feeling keyed up or on edge.
2. Being easily fatigued.
3. Difficulty concentrating or mind going blank.
4. Irritability.
5. Muscle tension.
6. Sleep disturbance (difficulty falling or staying asleep, or restless, unsatisfying sleep).

D. The anxiety, worry, or physical symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

E. The disturbance is not attributable to the physiological effects of a substance (e.g., a drug of abuse, a medication) or another medical condition (e.g., hyperthyroidism).

F. The disturbance is not better explained by another mental disorder (e.g., anxiety or worry about having panic attacks in panic disorder, negative evaluation in social anxiety disorder, contamination or other obsessions in obsessive-compulsive disorder, separation from attachment figures in separation anxiety disorder, reminders of traumatic events in posttraumatic stress disorder, gaining weight in anorexia nervosa, physical complaints in somatic symptom disorder, perceived appearance flaws in body dysmorphic disorder, having a serious illness in illness anxiety disorder, or the content of delusional beliefs in schizophrenia or delusional disorder).

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022:250, 251.

The vast majority of pediatric patients with generalized anxiety (75% to 80%) endorse physical symptoms. Common physical symptoms include headaches, tension, restlessness, gastrointestinal distress, and heart palpitations. Cognitive characteristics of generalized anxiety include negative thinking errors such as catastrophizing (i.e., expecting the worst possible outcome) or overestimation of the likelihood that something undesirable will occur. Generalized worries cover a range of everyday issues, such as being on time, upcoming activities, failure of loved ones or friends to meet basic expectations, or whether unexpected events (e.g., inclement weather) will change daily plans or schedules. Children and adolescents with generalized anxiety are often described as perfectionistic and overly sensitive. They are plagued by negative self-image, need for reassurance, and insecurities about competence, performance, and expectations about the future.

Jesse is a 14-year-old Black female who lives with her biologic parents and two older siblings in a suburban area. She presents for evaluation with her biologic father complaining of stomachaches, fatigue, and attention difficulties. Upon further interview, her father reports that Jesse has "always been nervous," and describes numerous domains about which she frequently seeks reassurance, such as whether her family will have enough food to eat and a place to live, the possibility of a natural disaster destroying their home, and whether her grades will get her into a good college. Though her parents provide reassurance related to these worries, she will ask multiple times. She has no history of attention issues and has always been an A student, but she recently has been having difficulty concentrating in class; she reports worrying about many things for most of the day, and that she "can't stop" the worry. As a result, her grades have become Bs and Cs, which she reports has caused even more worry.

Differential Diagnosis

Worry associated with generalized anxiety has multiple targets that include worries regarding everyday life experiences. Therefore, it may be more difficult to differentiate generalized anxiety from normative anxiety experienced by all individuals. However, the overall amount of anxiety, poor anxiety control, the lack of amelioration with appropriate reassurance and support, and longer-term impairment associated with generalized anxiety differentiate it from normal worry.

Generalized anxiety can be differentiated from emotional disorders based on the greater number of anxiety triggers and the specific nature of the worries. Children with generalized anxiety are concerned with past and future experiences, physical health, weather, environmental changes, and unexpected events. They may appear to have similar concerns to children with separation and social anxiety but differ in important ways. For example, children with generalized anxiety may avoid going to school, not because of separation or social concerns but because of academic performance worries. That said, it is not uncommon for younger children with GAD to also be diagnosed with co-occurring separation anxiety and older children with co-occurring SAD. Individuals with generalized anxiety and perfectionism may superficially resemble those with OCD. For example, a child with generalized anxiety and concern about academic performance may routinely check over homework for accuracy or study material repeatedly for the purpose of improving the work product or getting it “perfect.” In contrast, children with OCD know the work product is good or perfect, but experience doubt that their perception is accurate and then recheck for the purpose of rechecking. In addition, individuals with generalized anxiety may have recurrent uncontrollable worries that could be mistaken for obsessions; however, worries of GAD do not appear senseless, intrusive, or unwanted, rather they are described as powerful bona fide concerns and are usually tied to events or situations that would stimulate normal concerns. For example, normal children may worry about bad weather as it happens, GAD children will worry that bad weather may happen on a cloudless day, and children with OCD may worry that if they do not count to three in sets of three then a loved one may die in a weather-related incident.

Ruminations (i.e., recurrent depressive cognitions) are characteristic of depression and tend to regard experiences negatively and hopelessly. For example, repetitive thoughts such as “No one will like me no matter what I do” can be readily distinguished from “I need to be perfect for my teachers and parents” by identifying the depressive (ruminative) quality of the former and the perfectionistic (generalized anxiety) quality of the latter.

Course and Prognosis

Prospective and cross-sectional studies of youth suggest that GAD is an episodic or recurrent disorder and that prognosis is dependent on comorbidity and severity. Individuals with a comorbid diagnosis tend to have higher ratings of severity and functional impairment.

Those with greater severity and psychosocial impairment at baseline are more likely to retain a diagnosis over time.

Like other anxiety disorders, GAD is considered a risk factor for the development of additional psychiatric problems over time. Specifically, generalized anxiety may lead to the subsequent development of anxiety disorders, substance use, depression, or behavior problems. The New York Child Longitudinal Study investigated the associations between anxiety disorders and later psychiatric disorders in 9- to 18-year-old children and adolescents. DSM-III overanxious disorder was related to subsequent major depression, SAD, and generalized anxiety. Whereas children with separation anxiety symptoms were less likely than other children to begin drinking alcohol and did so later than others, those with generalized anxiety symptoms were more likely than other children to begin drinking and did so earlier. Results from the GSMS found that DSM-IV generalized anxiety was associated only with later conduct disorder but did not predict anxiety or other psychiatric problems.

TREATMENT OF ANXIETY DISORDERS

Both cognitive behavioral therapy (CBT) and antidepressant medications are effective for treating youth anxiety disorders (e.g., GAD and SAD). Expert reviews of the literature and meta-analyses repeatedly affirm the evidence base for both interventions, whether administered alone or in combination.

The Child/Adolescent Anxiety Multimodal Study (CAMS) remains the largest double-blind, placebo-controlled trial for child and adolescent anxiety disorders. The CAMS compared the relative efficacy of CBT, sertraline, combination therapy (CBT plus sertraline), and placebo for n = 488 youth with GAD, separation anxiety disorder, and/or SAD. All active treatments were superior to placebo, and combination therapy led to significantly higher response rates than CBT or sertraline (81% vs. 60% vs. 55%, respectively). Remission rates (loss of principal diagnosis, loss of all anxiety diagnoses) ranged from 20% to 68% and followed the same patterns as response rates based on treatment condition. Factors associated with better response and remission rates in the CAMS included the absence of comorbid internalizing disorders, absence of SAD at baseline, lower baseline anxiety symptom severity, lower caregiver strain, nonminority status, and younger age. A recent paper using CAMS data goes further to suggest that youth with mild–moderate anxiety benefit most from CBT monotherapy, whereas those with more severe illnesses may be best served by beginning with combined treatment.

Beyond acute and short-term follow-up of anxious youth, there has been considerable interest in understanding how these youngsters fare over time, particularly when they have had the benefit of evidence-based treatment. To address this question, youth who participated in the CAMS trial were followed as part of a prospective longitudinal study as they entered early adulthood (Child/Adolescent Anxiety Multimodal Treatment Extended Long-term Study; CAMELS). On average, youth were recontacted approximately 6 years following their CAMS treatment and followed for another 4 years. Findings from CAMELS suggest that roughly 20% of youth treated in CAMS remained consistently remitted over time as they entered early adulthood. Roughly half exhibited a fluctuating course marked by intermittent relapse. The remaining 30% were chronically ill, with anxiety disorder evident at every time point. These findings speak to the persistent nature of anxiety and the need to consider models of care that integrate chronic disease approaches and use surveillance and booster sessions to maintain wellness.

Psychotherapy

Over 50 years of evidence support the effectiveness of exposure-based CBT for youth anxiety. Although there is variability across treatment protocols, they are generally linked by their focus on psychoeducation, emotion identification, somatic management techniques, cognitive restructuring, exposure tasks, and relapse prevention. Traditional individual child treatment typically involves 10 to 20 weekly treatment sessions, and numerous studies support the use of additional formats, including group, family, brief intervention, and telehealth administration. In this section, we provide a brief synopsis of the evidence base for CBT before shifting to focus on newer trends in psychosocial treatment research for anxiety.

A 2017 study reviewed 100 anxiety treatment outcome studies in children and adolescents. For symptom reduction, CBT (*d* = 1.19), exposure alone (*d* = 1.05), modeling (*d* = 1.42), CBT with parents (*d* = 1.25), educational interventions (*d* = 1.26), and CBT with medication (*d* = 2.37) garnered the most consistent symptom reductions. The review established individual CBT as the most highly supported treatment for improving functioning, with an average within-group effect size of *d* = 1.76. CBT with parent involvement (*d* = 1.99), CBT with medication (*d* = 2.04), and exposure (*d* = .96) were categorized as having good support and being probably efficacious. These moderate-to-large effect sizes and the durability of treatment effects over time support the conclusion that CBT is an effective treatment for anxiety disorders in children and adolescents. They are bolstered further by findings from a 2019 meta-analysis which included clinical trials from 1963 to 2016 and found a mean posttreatment effect size of Hedges g = .66.

Given the robust evidence supporting the use of CBT, more recent studies have focused on understanding the mechanisms underlying its effects. Research suggests that both cognitive restructuring and exposure practice offer particular benefits during the course of treatment, with the steepest clinical gains observed following the introduction of exposure. Higher doses of exposure practice, more variable distress when practicing, and greater mastery of exposure tasks are all associated with better clinical outcomes. Moreover, contrary to what many may think, exposure practice does not diminish the therapeutic relationship but, indeed, enhances it.

In addition to efforts to optimize CBT by isolating its central ingredients (exposure) and maximizing their effects (via contemporary models of fear learning and extinction), research has also continued to consider factors that affect CBT outcomes. Family dynamics remain central, given their common links to youth anxiety and the identification of well-specified targets (e.g., symptom accommodation, overcontrol). Research consistently suggests that family-focused treatment is efficacious for youth anxiety; however, its additive value beyond individual child treatment remains a subject of some debate. This is likely due to methodologic issues across studies, including if/how putative family mechanisms are specified, how interventions map to specific family targets, and whether measurement approaches allow for tracking of family change and mechanistic effects. Certainly, there is emerging evidence from separate research groups that parent-only interventions may be comparable to child-focused CBT for anxiety.

A few somewhat distinct approaches to treating youth anxiety merit mention. First, parent–child interaction therapy (PCIT), originally developed to treat externalizing problems in young children and to treat various early anxiety problems, has been modified and adapted with success for youth anxiety. This approach targets child symptoms by reshaping parent–child interaction patterns associated with the maintenance of child anxiety symptoms. This goal is achieved through the use of live, unobtrusive parent coaching, delivered through a bug-in-the-ear receiver from a therapist situated behind a one-way mirror. Recently, the first randomized controlled trial (RCT) was conducted to test the efficacy of a 9-week PCIT treatment for separation anxiety in children aged 4 through 8 years. At posttreatment, 73% of children assigned to PCIT no longer met the criteria for a diagnosis of SAD, and results were largely maintained at a 3-month follow-up visit. Preliminary efficacy studies of another PCIT adaptation, the Coaching Approach behavior and Leading by Modeling (CALM) Program, designed to target the full range of anxiety disorders affecting young children (GAD and SAD, in addition to separation anxiety disorder) suggest the promise and feasibility of this approach.

A separate approach has focused on CBT adaptations designed particularly for social phobia, recognizing that this particular condition moderated the outcome in the CAMS trial. In a comparative treatment trial, children and adolescents aged 7 to 17 years were randomly assigned to Social Effectiveness Therapy for Children (SET-C, a behavioral approach), fluoxetine, or placebo. Results revealed both active treatments to be superior to placebo in reducing social distress and behavioral avoidance and increasing general functioning. SET-C, however, was superior to fluoxetine on these outcomes and was the only treatment superior to placebo in terms of improving social skills, decreasing anxiety in specific social interactions, and enhancing ratings of social competence. Although both treatments are considered efficacious, SET-C provides an added benefit by enhancing social skills.

A third innovation comes in the form of efforts to expand access to CBT intervention. Computer-assisted treatments which combine computer-based sessions with online and phone contact with a CBT clinician have been developed. These include protocols such as BRAVE-ONLINE and Camp-Cope-A-Lot. Preliminary trials of these programs, including one RCT, suggest that computer-assisted and in-person CBT produce statistically similar clinical improvements. Given the initial promising findings and the capacity of computer-based CBT to increase treatment accessibility, these programs merit continued investigation. Separate from this approach, telehealth CBT has also gained traction, with preliminary evidence that it produces outcomes comparable to clinic-based care.

Finally, technologic advances are also being integrated into treatment with the goal of enhancing learning and generalization of treatment effects. One such modality uses virtual reality technology to facilitate graded, in vivo exposures. Within this approach, youth are immersed in virtual environments where they interact with clinician-operated avatars to practice socially challenging tasks (e.g., greeting peers, answering questions, and assertiveness skills). The preliminary acceptability and feasibility of this approach have been demonstrated, and further research will assess its efficacy as a stand-alone treatment and adjunct to traditional CBT.

Pharmacology

Antidepressant medications are the mainstay of psychopharmacology of pediatric anxiety disorders. We have grouped the disorders in this psychopharmacology section—as is not uncommon in research studies (see CAMS trial)—as the treatments are mostly overlapping. Where appropriate, we will highlight key differences between disorders.

In a recent meta-analysis of the impact of antidepressant dose and class on treatment response in children with generalized, separation, and/or SADs, antidepressant-related improvement occurred early in the course of treatment, with approximately 50% of the improvement seen at week 12 occurring by the fourth week of treatment.

SSRIs are considered first-line medications for this population due to the favorable treatment response (effect sizes greater than 0.8) and low side-effect profile identified in a number of single-site and multisite RCTs. Agents that have demonstrated efficacy include fluoxetine (Prozac), sertraline (Zoloft), escitalopram (Lexapro), paroxetine (Paxil), duloxetine (Cymbalta), venlafaxine (Effexor), and fluvoxamine (Luvox). There is some data to support the use of tricyclic antidepressants (TCAs), including imipramine (Tofranil), but given the higher potential side effect burden, these are less commonly used.

Despite the demonstrated efficacy of SSRIs for anxiety in RCTs, no SSRIs are approved by the U.S. Food and Drug Administration (FDA) for use with children and adolescents. Published trials specifically targeting generalized anxiety included one effectiveness trial with sertraline and two trials—one positive and one negative—with venlafaxine (Effexor), which were combined and published together. More recently, duloxetine was found to be effective in a large, industry-sponsored, randomized, placebo-controlled trial of children and adolescents 7 to 17 years old with GAD leading to the first FDA indication of any medication for a non-OC, non-PTSD anxiety disorder in children and adolescents.

There have been three studies that have assessed the efficacy of antidepressants in SAD—two industry-sponsored trials and one federally funded study. Venlafaxine extended-release preparation was found to be more effective than placebo in a large sample of children and adolescents (*n* = 293) aged 8 to 17 years. Paroxetine was found to be more effective than placebo in a large sample (*n* = 322) of children and adolescents aged 8 to 17 years. Finally, the only federally funded study that assessed an SSRI, fluoxetine, and found it to be more effective than placebo was the SET-C study described above.

Open-label follow-up studies of SSRIs suggest that youth who benefit acutely continue on a positive trajectory for up to 6 months. Further, since anxiety disorders tend to run a chronic course with associated impairment, longer treatment with medication may be justified to support the consolidation of initial gains and prevent the return of anxiety symptoms.

There is no clear data to determine how long to treat children with medication to maximally reduce the chance for a return of symptoms upon medication discontinuation. Therefore, making the decision when to discontinue medication is often determined pragmatically. Treatment for 6 to 12 months after remission in symptoms may provide extended benefit without undue burden and may offer an opportunity to reduce the chance for a return of symptoms after medication discontinuation. However, longer-term treatment may be necessary if children experience a return of symptoms when medication doses are missed or during known trigger points such as the start of school in the fall or after the winter holidays (separation anxiety), grade reports, or final examinations (generalized anxiety), or group activities at school or camp (social). Such tenuous symptom control suggests that continued treatment, dose increases, or more intensive psychosocial treatments be provided to stabilize the patient before considering medication discontinuation. For patients and families who choose to pursue a discontinuation trial, it is important to consider timing. For example, it may be ill-advised to try coming off a previously helpful medication during a time of likely anxiety triggers or when family or medical support is not available (i.e., start of school for separation anxiety or participation in performance activities for generalized or social anxiety).

Although adults with GAD may be treated acutely with benzodiazepines, there are few data to support the efficacy of this class of medications for youth. For example, a small double-blind, placebo-controlled randomized trial of alprazolam (Xanax) for the treatment of pediatric overanxious and avoidant disorders demonstrated no benefit of medication over placebo. More to the point, benzodiazepines have not been extensively studied for the acute or chronic management of anxiety in children, and while potentially useful at reducing the experience of anxiety, they may lead to both psychological and physical dependence and thus complicate treatment long term. Given the efficacy of SSRIs and other antidepressants for childhood anxiety, benzodiazepines should not be considered as first-line treatment. If used at all, benzodiazepines may be considered for short-term augmentation of antidepressants early in the course of treatment for some highly anxious children.

There is little evidence supporting the use of other medications such as antihistamines, α-adrenergic agonists, or antipsychotics for the treatment of childhood anxiety disorders; however, it is not uncommon to see patients clinically who are treated with such agents to address problems with insomnia related to anxiety. Combining other medications with SSRIs to address anxiety-related insomnia should only occur when antidepressants have been maximally adjusted upward and anxiety and insomnia persist. Then the use of sedating antidepressants is preferred to the use of antihistamines, α-adrenergic agonists, or antipsychotics.

Buspirone (BuSpar), a nonbenzodiazepine, nonantidepressant anxiolytic, is effective in studies of adults with GAD and was evaluated in a large, multisite, industry-sponsored RCT. Although the trial was not published, a report of the negative outcome of the study is available on the FDA website.

Overall, the best evidence supports the use of SSRIs in the treatment of childhood anxiety disorders. The value of the SSRIs probably extends across the class so that medication choice can be based on a number of factors, including evidence of efficacy in high-quality clinical trials, half-life, the risk for drug–drug interactions, any subtle differences in side-effect profiles, and patient, family, and provider preference. SNRIs are also commonly used, bolstered by the FDA indication for duloxetine. The TCAs, which can be effective, have not been consistently able to separate from placebo in larger clinical trials, so they should be considered only as an alternative for those who have not responded to SSRIs or SNRIs. The value of TCAs in attention-deficit/hyperactivity disorder (ADHD) may make them a second-line choice for children with anxiety disorders and ADHD.

Despite ample evidence for the use of SSRIs and SNRIs in pediatric anxiety disorders, there remains tremendous heterogeneity in response. Nearly two in five patients with anxiety disorders fail to respond to these medications. There is growing awareness that pharmacogenetic factors likely account for some of this variability. Drug-metabolizing enzymes change throughout development, and there is data suggesting that SSRI exposure may be significantly different in children than in adults given the same dose. Though the Clinical Pharmacogenetics Implementation Consortium (CPIC) has published a guideline including dosing recommendations for adults, the same is not yet available in pediatric patients. Larger pharmacogenomics studies in pediatric patients are necessary to build evidence-based guidelines for the use of such data in children and adolescents with anxiety.

Safety. The SSRIs as a class share a number of important side effects that include physical symptoms such as stomachache and headache and can also be associated with changes in behavior such as activation/disinhibition, insomnia, apathy, manic reactions, and the potential for suicidal behavior. Activation/disinhibition (and similar or related symptoms such as akathisia) are quite common in the pediatric population, with a 10% to 15% difference between SSRI and placebo groups. Activation often occurs early in the course of treatment—days after the initial dose or first meaningful dose increase (e.g., sertraline 25 or 50 mg; fluoxetine 10 or 20 mg), and goes away with medication discontinuation. It appears to be more common in younger children and those whose anxiety co-occurs with symptoms of ADHD, autism spectrum disorders, or other early developmental psychopathology. The SSRIs impact sleep in two ways: (1) activation can make it more difficult to fall asleep, and (2) SSRIs disrupt sleep architecture; hence users may recover from anxiety, may experience a restful sleep, but may sleep more lightly, and experience middle-of-the-night awakenings. At higher doses of the SSRIs, apathy may occur, leaving patients relatively free of anxiety but unmotivated or feeling disconnected from living.

One area of concern to prescribers of SSRIs in anxious children is the induction of a manic reaction that may herald a new and more severe course of illness. However, manic reactions in the acute phase and large-scale RCTs of the SSRIs are exceedingly rare (<1%). Indeed, rates of manic symptoms are low even in longer-term follow-up studies. The concern regarding a manic switch is probably greatest when the anxiety disorder is complicated by problems with impulse control or aggressive behavior suggestive of bipolar disorder. Differentiating a manic reaction from the activation side effect can be challenging but an important distinction. Misdiagnosing an activation side effect as a manic reaction may shift the prescriber away from the antidepressants to the use of antipsychotics and mood stabilizers, neither of which are particularly effective for anxiety disorders. Manic reactions can be differentiated by their presentation, time course of onset, and response to medication discontinuation. Manic symptoms include elevated and expansive mood, grandiosity, decreased need for sleep, and increased goal-directed activities, and symptoms are distinct from the irritability and impulsivity that are commonly seen with activation. Activation occurs early in the course of treatment; manic reactions occur later. With medication discontinuation, activation symptoms decrease in relationship to blood level, whereas manic symptoms may persist and require medication treatment to reduce symptoms.

Finally, the risk of suicidal behavior emerging or worsening during antidepressant treatment is probably the public’s greatest concern. A recent meta-analysis suggests that the suicidal behavior risk difference between the antidepressant and placebo in a high-quality clinical trial is <1% with numbers needed to harm of >100 and the number needed to treat in high-quality RCTs in depression (NNT = 3–4 in NIMH-funded trials), OCD (NNT = 5), and the triad of anxiety disorders (NNT = 3). These data, the observed reduction in antidepressant use with the advent of a black box warning, and the subsequent spike in youth suicide gave rise to concerns that the black box warning may have had unintended negative consequences. While the black box warning on an antidepressant has not been changed, prescribers are regaining comfort in the use of the antidepressants.

The side-effect profiles of SNRIs are similar to the SSRIs but now include symptoms consistent with norepinephrine reuptake, including nausea, and the potential for initial worsening of anxiety (not activation per se), and for heart rate and blood pressure increases. The side-effect profile of TCAs is well known and includes prominent anticholinergic effects, dry mouth, constipation and urinary retention, and noradrenergic effects on cardiac function, elevated heart rate, and orthostatic blood pressure changes, and the potential for prolongation of important cardiac intervals including the QTc.

FUTURE DIRECTIONS

Issues that warrant further clarification pertaining to the treatment of youth with separation anxiety, generalized anxiety, or social anxiety include the identification of moderators of treatment outcome, identification of critical elements of psychotherapy that contribute to the outcome, questions of treatment sequencing, and whether youth who fail to respond to either medication or CBT as their first treatment will respond to the other treatment, strategies for enhancing outcomes and addressing treatment resistance, and elucidating long-term outcomes, particularly following medication treatment.

FURTHER READINGS

Beidel DC, Turner SM, Sallee FR, Ammerman RT, Crosby LA, Pathak S. SET-C versus fluoxetine in the treatment of childhood social anxiety disorder. *J Am Acad Child Adolesc Psychiatry*. 2007;46:1622–1632.

Biederman J, Petty CR, Hirshfeld-Becker DR, et al. Developmental trajectories of anxiety disorders in offspring at high risk for panic disorder and major depression. *Psychiatry Res*. 2007;153:245–252.

Bittner A, Egger HL, Erkanli A, Costello EJ, Foley DL, Angold A. What do childhood anxiety disorders predict? *J Child Psychol Psychiatry*. 2007;48:1174–

1183.

Buckner JD, Schmidt NB, Lang AR, Small JW, Schlauch RC, Lewinsohn PM. Specificity of social anxiety disorder as a risk factor for alcohol and cannabis dependence. *J Psychiatr Res*. 2008;42:230–239.

Compton SN, March JS, Brent D, Albano AM 5th, Weersing VR, Curry J. Cognitive-behavioral psychotherapy for anxiety and depressive disorders in children and adolescents: an evidence-based medicine review. *J Am Acad Child Adolesc Psychiatry*. 2004;43:930–959.

*Connolly SD, Bernstein GA; Work Group on Quality Issues. Practice parameters for the assessment and treatment of children and adolescents with anxiety disorders. *J Am Acad Child Adolesc Psychiatry*. 2007;46:267–283.

Costello EJ, Egger HL, Angold A. The developmental epidemiology of anxiety disorders: phenomenology, prevalence, and comorbidity. *Child Adolesc Psychiatr Clin N Am*. 2005;14:631–648.

Ginsburg GS, Becker EM, Keeton CP, et al. Naturalistic follow-up of youths treated for pediatric anxiety disorders. *JAMA Psychiatry*. 2014;71:310–318.

Ginsburg GS, Becker-Haimes EM, Keeton C, et al. Results from the child/adolescent anxiety multimodal extended long-term study (CAMELS): primary anxiety outcomes. *J Am Acad Child Adolesc Psychiatry*. 2018;57(7):471–480.

Gleason MM, Egger HL, Emslie GJ, et al. Psychopharmacological treatment for very young children: contexts and guidelines. *J Am Acad Child Adolesc Psychiatry*. 2007;46:1532–1572.

Gosch EA, Flannery-Schroeder E, Mauro CF, Compton SN. Principles of cognitive-behavioral therapy for anxiety disorders in children. *J Cogn Psychother*. 2006;20:247.

Higa-McMillan CK, Francis SE, Rith-Najarian L, Chorpita BF. Evidence base update: 50 years of research on treatment for child and adolescent anxiety. *J Clin Child Adolesc Psychol* 2016;45:91–113.

Hirshfeld-Becker DR, Biederman J, Henin A, et al. Behavioral inhibition in preschool children at risk is a specific predictor of middle childhood social anxiety: a five-year follow-up. *J Dev Behav Pediatr*. 2007;28:225–233.

Kagan ER, Frank HE, Norris LA, et al. Antidepressant use in a 3-to 12-year follow-up of anxious youth: results from the CAMELS trial. *Child Psychiatry Hum Dev*. 2021;52(1):41–48.

Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen HU. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *Int J Methods Psychiatr Res*. 2012;21:169–184.

Manassis K, Lee TC, Bennett K, et al. Types of parental involvement in CBT with anxious youth: a preliminary meta-analysis. *J Consult Clin Psychol*. 2014;82:1163–1172.

*Morris TL, March JS. *Anxiety Disorders in Children and Adolescents*. 2nd ed. Guilford Press; 2004.

Negreiros J, Miller LD. The role of parenting in childhood anxiety: etiological factors and treatment implications. *Clin Psychol (New York)*. 2014;21:3.

Ramsey LB, Bishop JR, Strawn JR. Pharmacogenetics of treating pediatric anxiety and depression. *Pharmacogenomics*. 2019;20(12):867–870.

*Reynolds S, Wilson C, Austin J, Hooper L. Effects of psychotherapy for anxiety in children and adolescents: a meta-analytic review. *Clin Psychol Rev*. 2012;32:251–262.

Rooksby M, Elouafkaoui P, Humphris G, Clarkson J, Freeman R. Internet-assisted delivery of cognitive behavioural therapy (CBT) for childhood anxiety: systematic review and meta-analysis. *J Anxiety Disord*. 2015;29:83–92.

Strawn JR, Mills JA, Sauley BA, Welge JA. The impact of antidepressant dose and class on treatment response in pediatric anxiety disorders: a meta-analysis. *J Am Acad Child Adolesc Psychiatry*. 2018;57:235–244.e2.

Strawn JR, Mills JA, Schroeder H, et al. Escitalopram in adolescents with generalized anxiety disorder. *J Clin Psychiatry*. 2020;81:20m13396.

*Strawn JR, Welge JA, Wehry AM, Keeshin B, Rynn MA. Efficacy and tolerability of antidepressants in pediatric anxiety disorders: a systematic review and meta-analysis. *Depress Anxiety*. 2015;32(3):149–157.

Swan AJ, Kendall PC, Olino T, et al. Results from the child/adolescent anxiety multimodal longitudinal study (CAMELS): functional outcomes. *J Consult Clin Psychol*. 2018;86(9):738.

*Walkup JT, Albano AM, Piacentini J, et al. Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *N Engl J Med*. 2008;359(26):2753–2766.

Warren SL, Umylny P, Aron E, Simmens SJ. Toddler anxiety disorders: a pilot study. *J Am Acad Child Adolesc Psychiatry*. 2006;45:859–866.

Weisz JR, Kuppens S, Ng MY, et al. What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. *Am Psychol*. 2017;72:79.

Weisz JR, Kuppens S, Ng MY, et al. Are psychotherapies for young people growing stronger? Tracking trends over time for youth anxiety, depression, attention-deficit/hyperactivity disorder, and conduct problems. *Perspect Psychol Sci*. 2019;14(2):216–237.

Yap MBH, Pilkington PD, Ryan SM, Jorm AF. Parental factors associated with depression and anxiety in young people: a systematic review and meta-analysis. *J Affect Disord*. 2014;156:8–23.

51.4 Selective Mutism

R. Lindsey Bergman, Ph.D., and Araceli Gonzalez, Ph.D.

Selective mutism (SM) is a childhood behavioral disorder characterized by a persistent failure to speak in one or more specific social situations despite speaking in other situations. Although the syndrome was recognized as long ago as 1877, the disorder (earlier called elective mutism) did not appear in the *Diagnostic and Statistical Manual of Mental Disorders* until its third edition (DSM-III) in 1980. Current conceptualizations of SM emphasize the relationship between considerable social anxiety and the failure or extreme reluctance to speak. In addition, researchers continue to investigate the presence of speech or language anomalies as well as oppositional symptoms among children with SM. Although there has been increased interest related to SM recently, in comparison with many other childhood emotional or behavioral disorders, there is a relative lack of consensus on many issues related to the epidemiology, phenomenology, and treatment of the disorder.

DEFINITION

The DSM-5-TR defines SM as a consistent failure to speak in specific social situations where speaking is expected (e.g., to peers at school), despite speaking regularly in other situations. DSM-5 also specifies that in order for the diagnosis to be assigned, the lack of speech must interfere with achievement or communication, must last at least 1 month (beyond the first month of school), and must not be due to a lack of knowledge or comfort with the language required in a given situation. Finally, DSM-5-TR specifies that the lack of speech cannot be better accounted for by a communication, developmental, or psychotic disorder. Criteria for SM are similar in the *International Statistical Classifications of Diseases and Related Health Problems* (ICD-11).

HISTORY

In the 19th century, the term *aphasia voluntaria* was first noted in the psychiatric literature and was used to describe children who voluntarily chose not to speak in certain situations. Many other terms including voluntary mutism, speech phobia, suppressed speech, functional mutism, and speech avoidance have also been used to describe children who meet the current criteria for SM. The term *elective mutism* was first used in 1934 and appeared in DSM-III in 1980. Some believed that descriptors such as "voluntary" and "elective" that appeared in these early terms suggested a volitional choice to refrain from speaking, and possibly, an oppositional attitude. This attitude was also reflected in the DSM-III definition of elective mutism, which contained a reference to a "refusal" to speak in the definition of SM. In DSM-III-R, elective mutism was renamed SM, perhaps reflecting shifting conceptualizations of the disorder such that attention was more focused on social anxiety as opposed to oppositional or volitional aspects of the nonspeaking behavior.

COMPARATIVE NOSOLOGY

Consistent with the shifting conceptualization described above, SM was reclassified as an anxiety disorder in the current edition of the DSM (DSM-5) and its text revision (DSM-5-TR). In the DSM-IV, SM was included in the section Disorders Usually First Diagnosed in Infancy, Childhood, or Adolescence. In the DSM-5, disorders within that category were reorganized into different categories and SM is now included as an anxiety disorder. This reclassification of SM in the Anxiety Disorders section of the DSM-5 is a logical fit given growing evidence for the phenomenologic similarities between SM and social anxiety (see Etiology section). However, the nature of relationship between SM and social anxiety is complex and requires further research as discussed below. Other than the relocation of the SM diagnosis from Disorders Usually First Diagnosed in Infancy, Childhood, or Adolescence to the Anxiety Disorders section, the diagnostic criteria for SM are virtually identical in the DSM-IV and -5. Interestingly, a diagnosis of SM does not require the presence of fear or anxiety.

The disorder is called SM in the current (11th) edition of the ICD. In the ICD-11, SM appears in the section of Anxiety or fear-related disorders.

EPIDEMIOLOGY

SM prevalence rates vary by study but have been generally below 1%. Few community-based prevalence studies have been published and the majority of this work was completed in European samples. These studies have yielded disparate findings due to different sampling strategies and definitions of SM. A U.K. study identified 0.69% of 4 to 5 year olds as completely mute (i.e., not speaking at all) at school. However, this rate dropped to 0.08% after approximately 8 months of school. Another survey, also in the United Kingdom, identified only 2 of 3,300 (0.06%) 7 year olds as selectively mute at 36 months of age. More recent investigations have focused on the school environment to investigate the prevalence of SM. Two Scandinavian surveys found teacher-reported rates of DSM-IV-TR SM of 0.18% and 1.9%, respectively, and another study with Irish primary school children aged 4 to 12 years also found a prevalence rate of 0.18. One U.S. study surveyed kindergarten, first-, and second-grade classrooms in an urban public school district and reported a DSM-IV-TR SM prevalence rate of 0.71% (16 of 2,256). A study in Israel yielded comparable results with a prevalence rate of 0.76. Another clinic-based study in the United States found a prevalence rate of 0.50% in a primary care clinic. These studies provide evidence that the point prevalence rate for SM is comparable to rates of other childhood psychiatric disorders, including obsessive-compulsive disorder (0.5% to 1%) and major depressive disorder (0.4% to 3%).

Overall prevalence does not seem to vary markedly by sex or race/ethnicity; however, there is evidence that slightly more girls than boys seem to be affected by SM, with an average reported female-to-male ratio of 1.2 to 1. There is some evidence to suggest that this gender difference may become more pronounced among older children. Although it is commonly reported that children with SM show significant symptoms from the age at which they spoke fluently, with the age of onset typically reported from 2 to 5 years old, SM is often not noticed until the time of entry into school. It has also been reported that immigrant and bilingual or language minority children are at higher risk of SM, and the prevalence of the disorder may be three times higher in these populations. However, this required further study, and it should be noted that SM must be distinguished from a nonverbal period that is sometimes observed in children during second-language acquisition.

ETIOLOGY

The conceptualization of SM has undergone a considerable shift in recent decades. At this point in time, evidence suggests that SM results from multiple developmental pathways involving the complex interplay of genetic, neurobiologic, temperament, and behavioral (e.g., avoidance) variables. Previously, it was often suggested that SM was related to a variety of conditions including oppositionality, trauma, or family neuroses. While a variety of psychological and developmental factors likely play a role, there is a general consensus that SM is closely related to anxiety, and social phobia in particular. Children with SM and/or social phobia tend to exhibit similar levels of nonverbal social anxiety, trait social anxiety, and cortisol reactivity to a presentation task. The body of work is consistent with a conceptualization that SM is a developmental variant or precursor of social phobia, rather than an extreme form of SM. Strong support for this position is derived from studies reporting that 69% to over 95% of children with SM also met criteria for social anxiety disorder. Similarly, family history data reveal an elevated rate of social phobia among parents of children with SM, and severity of children's SM symptoms may predict parents' diagnosis of generalized social phobia. Research regarding genetic risk of SM is in its infancy. One recent study examining genetic susceptibility factors for SM in families with social anxiety–related traits found evidence of increased risk of SM in individuals with a genetic variation in *CNTNAP2*, a gene implicated in various forms of language impairment. Notably, *CNTNAP2* has also been implicated in autism spectrum disorder (ASD). While SM and ASD may exhibit several phenotypic similarities (e.g., difficulties with social communication) and there is recent evidence for a potential shared genetic susceptibility, they are distinct conditions (further discussed below in Differential Diagnosis).

Evidence of a relationship between SM and social phobia also come from studies demonstrating several cognitive and behavioral similarities. Recent studies also noted similarities between SM and social anxiety in the content of their reported fears, such as fear of negative evaluation by others. Researchers and clinicians have also noted the resemblance between SM children and children described as shy or “behaviorally inhibited.” In fact, reluctance to speak is often considered to be a sensitive index of behavioral inhibition. Consistent with this is the finding that parents of children with SM rate their children as not responding well to new situations or handling transitions well, both widely considered to be fundamental elements of behavioral inhibition. Since there is evidence that the presence of behaviorally inhibited temperament is associated with increased risk of future anxiety disorders, it is possible that behavioral inhibition is an etiologic factor in the development of SM. By contrast, a small proportion of children with SM do not exhibit social phobia with regard to nonverbal communication and will engage in social interaction using gestures, facial expressions, body language, and other socially engaged behavior. Such cases raise questions about the nature of the relationship between SM and social phobia, suggesting that some individuals may experience anxiety specifically related to expressive language and not to social situations more generally. Some researchers have investigated the notion that the “nonspeaking behavior” of children with SM essentially should be seen as an emotion regulation strategy aimed to reduce anxiety around social interaction. Finally, additional evidence linking SM to social phobia comes from the finding that pharmacologic agents used to treat adult social phobia are similarly effective in the treatment for SM.

As mentioned previously, oppositionality and recent trauma are often cited as etiologic factors related to SM; however, studies utilizing systematic diagnostic procedures do not provide support for these relationships. In comparing children who met criteria for only social phobia to those with both SM and social phobia and those with no disorder, researchers found no significant differences on parental reports of oppositionality. Furthermore, mean scores for ratings of oppositionality were in the normal range for children with SM. Thus, findings suggest that the majority of children with SM do not present with clinically significant oppositional behaviors. However, 29% of children with SM were rated as meeting criteria for oppositional defiant disorder (ODD) by clinicians. Thus, a subset of children with SM may exhibit comorbid ODD.

Somewhat more confusing are conflicting findings related to the prevalence of communication disorders among children with SM and the association between developmental delay and SM. Several studies have examined possible deficits in auditory or speech and language processing in children with SM. In one study, a little over a third of children with SM displayed speech and language disorders, particularly articulation and expressive language disorders. In another study, children with SM were compared to children with an anxiety disorder and normal controls on measures of language, nonverbal working memory, and social anxiety. When compared to children with other anxiety disorders and healthy controls, children with SM had lower scores on language measures and deficits on tasks requiring them to remember verbal information. In another study, researchers found that children with SM did not exhibit deficits in visual memory, but they did show significantly poorer verbal memory than controls. Thus, some children with SM may exhibit meaningful deficits in speech and language processing. While the majority of children with SM do not exhibit speech and language disorders beyond the context of SM, it was recently purported that a subset of children with communication and language difficulties who are also behaviorally inhibited may become nervous to speak in new situations, which may in turn contribute to risk for SM.

Although some children with SM may experience language delay or dysfunction and deficits in speech and language processing, it is noteworthy that among researchers it is not suggested that these problems supersede social anxiety as a major contributing factor toward nonspeaking. Furthermore, although language delay or speech disorder may be more common among children with SM, these problems could be a result, rather than a cause, of the disorder. Finally, speech and language problems associated with SM are likely to be subtle and are exclusionary criteria for the diagnosis of SM when they cannot be ruled out as the cause of the lack of speech. Ultimately, however, it is important to note that many in the field believe that neurodevelopmental and language anomalies occur more frequently among children with SM than among children with other anxiety-related disorders.

DIAGNOSIS AND CLINICAL FEATURES

The primary symptom exhibited by children with SM is the failure to speak in some situations where speaking is expected despite speaking in other situations. The situation in which the child has no difficulty speaking is almost always in the home environment. In fact, some families of children with SM report that the child seems to be “overly” talkative at home. However, in severe cases, there may be some members of the household to which the child with SM does not speak. If a child who is presumed to have SM does not speak at all at home, there is reason to seriously question the diagnosis of SM.

Several studies indicate that the situation in which speaking is most commonly avoided is at school. Within this common pattern (speech at home, failure to speak at school), there is large individual variation in terms of speaking patterns among children with SM. In the school environment, children with SM more commonly fail to speak to adults (e.g., teachers) than to peers, although a lack of speech with peers is not uncommon. However, many children with SM have individually distinct and stable patterns regarding to whom they do, and do not, speak. For instance, there are often particular extended family members with which the child does not speak and others to whom they consistently speak. Some children with SM have little or no difficulty speaking with strangers, whereas others place strangers at the top of their hierarchy of people with whom it is difficult to speak.

The label of SM can be used somewhat broadly to describe children who are consistently reluctant to speak or display low frequency of speech. Although some argue that children who speak inconsistently, reluctantly, or with low volume do not, in fact, have SM, these speaking behaviors are relevant for the diagnosis of SM because children who completely fail to speak in some situations often speak inconsistently, with low frequency or volume, or reluctantly in others. For example, some children with SM may fail to speak at all in some situations and speak normally in others (e.g., home), while in other situations they might speak but not in a "normal" voice but whisper fairly readily. Also fairly common is the child who does not speak at all in some situations, but in other situations speaks in a "disguised" or odd voice that is mechanical, gruff, or baby like.

Given that SM appears to be so closely related to social phobia, it is not surprising that children with SM also have difficulty with nonverbal but socially relevant tasks such as eating in front of others, having their picture taken, and communicating nonverbally. When social anxiety symptoms not related to speech are present and meet criteria for social anxiety disorder, the additional diagnosis should be given. Even among those who are not also diagnosed with social anxiety disorder, children with SM are typically described as shy and self-conscious. However, it is important to note that not all children with SM are shy; although not typical, there are children with SM who present as outgoing, friendly, and eager to interact but only interact nonverbally (e.g., smile for the camera, raise their hand in class but not answer when called upon). This presentation has led some researchers to view SM as a type of specific phobia of speaking.

The following case exemplifies a common presentation of a child with SM.

James is a 5-year-old Chinese American boy who lives with his biologic mother, father, and older sister. James's parents reported that he speaks normally at home but that he started to have difficulty speaking to adults and children at school upon entry to preschool. He continued to have difficulty speaking throughout his preschool year and has ongoing difficulty in kindergarten. At home, his parents report that he is extremely talkative and outgoing when he is around immediate family members and extended family members close to his age. At school, James does not speak to any of the adults but answers classmates with one-word answers and speaks only in a whisper. By his parents' report, James also exhibits extreme social anxiety, to the point of "freezing" in certain situations when attention was focused on him. He often watches other children playing from a distance, does not join in group activities, and does not participate in circle time at school. James's parents also report that he hides behind them when strangers are around and that he dislikes having his picture taken. At the time of his evaluation, James had not received prior treatment for SM or any other emotional or behavioral disorders. James was born in the United States and speaks fluent English and limited Mandarin. His parents speak to him in both English and Mandarin and he mostly responds to them in English. According to his parents, he developed normally with no delays and appears to have above-average intelligence. They also reported that James enjoys riding on his bicycle, singing, and art activities.

During initial evaluation, James hid behind his parents and failed to respond in any way to the intake clinician. James's parents reported that this behavior is typical of him when in a novel situation with new people, but that he might answer with one-word answers softly, communicate nonverbally, and make eye contact with people once he gets more comfortable with them. James's parents reported that he speaks spontaneously and fluently when he is at home with his family. They also reported that he can be quite animated, is a lot of fun, and is very talkative at home. James received diagnoses of SM (moderately severe) and social phobia (moderately severe). Behavioral treatment was recommended.

Behavioral treatment was initiated following the intake evaluation. Initially, the therapist played nonverbal games and engaged in nonverbal activities with James to help him feel comfortable in the session (e.g., coloring, blowing bubbles, hiding pennies, and giving nonverbal clues). Next, James and his mother were asked to create a list of easy, medium, and hard speaking situations and a list of small, medium, and large rewards. These lists then became the basis for assignments (and reinforcement) for speaking tasks that gradually increased in difficulty. In general, sessions included time with James and his mother together to review past and future assignments as well as time with James and the therapist alone.

When treatment began, James did not communicate at all verbally with the therapist. The therapist gradually increased rapport (via unstructured play) while having James try increasingly difficult tasks, such as first whispering to his mother with the therapist in the corner, then nodding yes or no in response to the therapist’s questions, pointing, using a puppet to “talk for him,” whispering to his mother while facing the therapist, and then subsequently responding verbally to the therapist. A successful technique in working with James was having him whisper to his mother while playing games with the therapist and then having his mother move a few inches away each time he was asked to respond to the therapist. As the whisper became more audible, the therapist was able to hear his speech and responded directly to the whisper (without waiting for his mother to tell the therapist what James said). James became more comfortable with the therapist hearing his voice, and he was slowly able to direct his speech toward the therapist. After three sessions, James consistently and spontaneously talked to the therapist. James received stickers for completing each speaking assignment, and after filling up “sticker charts,” he received rewards (i.e., small toy, special activity with parent, or treat from the reward list).

James was also given assignments that involved both his teacher and classmates. These were implemented in a gradual fashion and included waving to the teacher, playing a game with his parent first in an empty classroom then while his teacher was in the classroom, leaving a phone message for the teacher, playing an audiotape of him saying “hello” to the teacher, whispering to his parent loudly enough for his teacher to hear, whispering “hello” to the teacher, and speaking “hello” to the teacher in a regular voice. Assignments also involved classmates and included playing games with his parent and one classmate, asking classmates questions about their favorite things (e.g., color, animal, and food), playing games with classmates without his parent, and play dates at home. After approximately 12 sessions, James spoke a complete sentence to a group of three classmates, spoke to his teacher directly, and answered a question aloud in front of the entire class.

During the last few sessions, which were tapered over several weeks, James’s mother took an increasingly primary role in assigning and following up on speaking assignments. James’s mother had him continue practicing talking assignments with extended family members and unfamiliar adults and children. James entered the first grade and, after only a few days, began regularly speaking to his teacher (both privately and in front of peers) and to most peers in class. Following the completion of therapy, James’s mother continued to monitor his speaking behaviors and to promote speaking in new situations by encouraging (and rewarding) James’s gradual successes with novel people and situations.

DIFFERENTIAL DIAGNOSIS

As discussed above, the current understanding of SM emphasizes that social anxiety underlies most selectively mute behavior. The resulting conceptual and diagnostic overlap between SM and social phobia complicates the clinician’s determination of whether a child with SM also meets the criteria for the presence of comorbid social phobia. One of the most successful strategies for the clinician attempting to make this determination is to direct queries toward the presence of social phobia symptoms that do not involve speaking. For example, the presence of fear and avoidance of eating in public, having pictures taken, using public restrooms, or writing in front of others provide strong evidence for an additional diagnosis of social phobia in a child who meets the criteria for SM. Although some may perceive the child with SM to be oppositional, distinguishing SM from primary ODD is fairly straightforward in many cases. Most importantly, the child with SM is unlikely to engage in significant oppositional behaviors outside of the context of verbal communication. If a child who appears to have SM (failure to speak in certain situations despite speaking in other situations) also exhibits oppositional behaviors in a variety of situations, both diagnoses are given. For children whose failure to speak does not appear to follow a consistent pattern with regard to the situation and targeted individuals, appears primarily motivated by instrumental gains, and occurs in the absence of any other anxiety symptoms, it is possible that ODD rather than SM is the correct diagnosis.

There is a growing body of work examining similarities between SM and ASD. A few recent studies report a high degree of overlap in symptoms of SM and ASD, as reported in a retrospective review of medical records and parent reports on developmental and social communication scales. Additional research is needed to determine the between—group discriminative validity of existing measures of ASD among children with SM. Although children with SM can be misdiagnosed as having autism due to their lack of speech in the clinic setting, this diagnostic error can be avoided with a more thorough evaluation of autism-like features in children with suspected SM, with close consideration of (in)consistency of behaviors across settings. One study examined joint attention (one person’s purposeful focus of attention with another person, often impaired in children with ASD) in children with SM, other anxiety disorders, and a control group. Findings indicated joint attention abilities in children with SM did not differ from the other groups in an unstructured free play condition, although children with SM displayed less joint attention than the other groups during structured tasks. Although this indicates some level of behavioral or psychological overlap between SM and autistic-like features, the context-specific display of behavior in children with SM is critical to distinguishing SM from ASD and may be more reflective of behavioral inhibition than autism-related behavior. Critically, the lack of speech associated with SM is, by definition, limited to certain situations, and evidence indicates that related behaviors that overlap with other diagnostic categories (e.g., poor joint attention) might also be limited to specific situations. However, the cognitive, language, and social communication difficulties of a child with ASD are present across settings. Similarly, the diagnostic criteria requiring that the child speak normally in some situations can often rule out a severe communication disorder.

Several studies have identified subsets or profiles of children with SM who also have mild or moderate levels of oppositionality. Although one may perceive some children with SM to be oppositional, distinguishing SM from primary ODD is fairly straightforward in many cases. Crucially, many children with SM may appear to be oppositional, almost exclusively around situations during which they are reluctant to speak but are being prompted to do so. If a child with a purported diagnosis of SM consistently exhibits oppositional behaviors in a variety of situations, both diagnoses are given. For children whose failure to speak does not appear to follow a consistent pattern with regard to the situation and targeted individuals, appears primarily motivated by instrumental gains, and occurs in the absence of any other anxiety symptoms, it is possible that ODD rather than SM is the correct diagnosis.

The DSM-5 criteria state that a child should not be diagnosed with SM if the failure to speak is due to a lack of knowledge of, or comfort with, the spoken language required. Researchers in the area of bilingual children with SM report that immigrant children and those learning a second language in their new country may be overlooked or misdiagnosed. Thus, the diagnosis of immigrant and language minority children can be difficult, and it is important to understand typical bilingual child development to diagnose SM properly. Many children learning a second language may go through a nonverbal or “silent period” due to the realization that their native language is not understood by classmates and teachers and that their skills in the new language are not sufficient. To diagnose SM in bilingual children and children who are acquiring a second language, it is important to know that bilingual children with SM exhibit a failure to speak in both their native and newly acquired languages in several unfamiliar settings and for an extended period of time (>6 months). Although a child learning a new language may not feel completely comfortable speaking a new language for several months, most children show no progression of speech as they gain mastery over their new second language, whereas a child with SM typically does not.

COURSE AND PROGNOSIS

SM appears to have an insidious onset, with most parents reporting failure to speak in some settings as a problem that was evident from the time their child began talking. However, most typically, the lack of speech outside of the home is not considered to be a problem until school entry. Although SM is generally considered to be a disorder that is difficult to treat successfully, at the same time, it is believed to spontaneously remit over time in a large majority of cases. Based in part on one study that found a large reduction in SM over a 6-month period (from the beginning to the end of the school year), it was believed that the disorder spontaneously remitted over time in a large number of cases and that in these cases, and that SM should be described as transient rather than persistent. This distinction has been refined somewhat based on a school-based study that revealed that although close to 30% of children with SM showed some improvement from the beginning to the end of a school year, as a group, the SM children remained significantly more symptomatic than matched unaffected peers.

In a controlled long-term outcome study (mean follow-up period of 13 years), young adults with SM in childhood were compared to young adults with anxiety disorders in childhood and young adults with no psychiatric diagnosis in childhood. Researchers found that 18% of patients with SM in childhood showed only slight improvement in young adulthood, while the rest of the patients with childhood SM were found to be either markedly improved or totally improved in young adulthood. They reported a complete remission rate of 58%. Unfortunately, these data did not include comparisons related to variation in course with or without treatment. The only controlled treatment study that assessed youngsters 1 year following successful behavioral treatment reported that children maintained or improved upon initial gains when reassessed.

Although it seems intuitively useful to discriminate between transient and persistent SM and to assume that persistent mutism may be related to more severe impairment, this distinction requires further investigation. Similarly, there is no information related to ongoing difficulties that might result from even relatively short-term or transient SM. Given that there is little information regarding lasting consequences related to an early disturbance in verbal communication and that nonspeaking behavior is often conceptualized as a learned behavior that is reinforced considerably over time, it makes sense to target treatment toward younger, recent-onset youngsters. Additional support for the importance of early intervention (vs. delaying treatment to establish if SM will resolve spontaneously) comes from recent findings that age of onset, duration of SM, and age at treatment all predict treatment effect size. Specifically, young age and shorter duration of SM were associated with larger treatment effects.

TREATMENT

Although in the last 20 years, there has been an increase in the amount of research on SM, the majority of published data on the treatment of SM continues to rely on single-case studies and case series of markedly varying quality. In addition, studies reporting SM treatment outcomes tend to rely on small samples or single cases and do not identify diagnostic procedures, assessment or outcome methods, number of treatment sessions, or details of the treatment method. In spite of these limitations, the conviction that behavioral techniques are an essential component of treatment for SM is widespread. Recently, two small randomized controlled trials of behavioral treatment for SM were completed and supported the efficacy of such interventions compared to a waitlist control. Follow-up studies as long as 5 years posttreatment provide support for the durability of findings. In general, these and other published reports in the literature describing treatment efforts provide a solid basis for the development of behavioral treatments that include parent and school involvement in treatment. In addition to behavioral strategies, a small but growing literature provides support for the efficacy of a pharmacotherapeutic intervention (primarily the selective serotonin reuptake inhibitors [SSRIs]). In addition to a growing effort aimed at empirical validation of treatment for SM, recent efforts have also resulted in better instruments for the assessment of specific symptoms of SM. Instruments such as the Selective Mutism Questionnaire, a parent self-report measure of the affected child's lack of speech, and the School Speech Questionnaire, the companion teacher measure, are commonly used in the assessment of symptoms in both research and clinical settings.

Behavioral Treatment

Multiple published reports describe the successful utilization of relatively standard behavioral techniques such as contingency management, stimulus fading, systematic desensitization, negative reinforcement, and shaping for the treatment of SM. A combination of behavioral techniques is probably the most common and successful treatment approach. Behavioral approaches typically utilize a hierarchy or rank-ordered list of situations in which the child has difficulty speaking. Then, the child is guided to systematically engage in speaking-related behaviors (e.g., mouthing speech, making sounds, and whispering) in increasingly more difficult situations. Over repeated successful attempts, the associated anxiety dissipates through the process of autonomic habituation. In addition, when the feared consequences of speaking fail to occur, patients' heightened expectations of harm disappear, reducing anxiety even further. Typically, the child receives positive reinforcement (rewards) following attempts to engage in speaking-related behaviors to shape the child's behavior. There are also several reports of treatments utilizing somewhat unique self-modeling behavioral interventions, including audio and video editing techniques and, most recently, techniques utilizing enjoyable electronic applications on mobile phones and tablets.

A number of considerations serve to complicate the behavioral treatment of SM and to distinguish it from treatment for social phobia, a disorder to which it is closely related. These factors include the child's failure to speak (even to the therapist), the young age of most SM patients, and the need for significant school involvement in almost all cases. The young age of children with SM necessitates considerably more parental involvement in treatment than is typically recommended for other anxiety disorders.

Although not without exception, most children with SM do not initially speak to the treating clinician. Obviously, treatment approaches for SM need to consider the initial communication difficulties that are often present and include strategies for facilitating verbal communication with the therapist. The initial lack of verbal communication with the therapist is assumed to reflect the general anxiety underlying failure to talk in situations outside the home and is addressed accordingly. To promote speaking in the treatment setting, traditional behavioral techniques (shaping, gradual exposure, reinforcement) are first introduced within the session. Recent reports show that the use of mobile apps in treatment is successful at reducing the time it takes children to begin verbalizing during treatment, with many study participants speaking during the first treatment session. The authors posit that the success is due, in part, to the principle of systematic desensitization: that is, the fear associated with speaking is incompatible with the enjoyment associated with the use of the app.

The SM child’s failure to speak represents a somewhat unique challenge for others in the child’s life, particularly parents and school personnel. Although difficulty in differentiating oppositional behavior from anxiety-related avoidance characterizes many child anxiety disorders, when failure to speak is involved, parents and others are particularly likely to attribute avoidance to oppositional motivations and to get frustrated and angry. In contrast, other reactions to nonspeaking include the tendency to accommodate the child’s lack of speech and speak “for” the SM child, thus decreasing the necessity of independent communication. This tendency, along with other ways that nonspeaking may be reinforced, also needs to be addressed in the treatment of the disorder. Recently, proponents of parent–child interaction therapy (PCIT) have combined components of PCIT with more standard behavioral SM treatment to develop a treatment that teaches parents how to encourage and coach their children to increase speech while reducing accommodation relating to avoidance of speech.

Nearly all forms of youth anxiety involve school-related impairment to some extent. However, unlike SM, they do not typically revolve as extensively around the school environment. For this reason, a treatment that integrates school participation as a central component of the intervention is necessary for the effective treatment of SM. Most importantly, close rapport with school personnel that includes regular communication and support must be established to ensure the success of behavioral interventions in the school setting. Furthermore, treatment programs should include teacher education and training in the behavioral methods utilized for treatment. The utilization of teachers in the treatment of SM has been highlighted as a particularly cost-effective treatment approach. In the last few years, treatment programs have been developed that include an intensive group element that simulates the school environments while additionally providing treatment condensed over a few days. This approach also has the advantage of providing treatment over a reduced period of time to children who have to relocate to find clinicians experienced in the treatment of SM.

Other Psychosocial Therapies

Although the most common accounts of successful psychosocial treatment for SM involve behavior therapy, there are descriptions in the literature of successful treatment for SM with play therapy, family therapy, psychodynamic therapy, and group therapy. Although most of these descriptions are flawed by their small sample sizes and failure to adequately describe the diagnostic or treatment intervention processes, it is noteworthy that descriptions of play, family, and group therapy all contain considerable emphasis on the use of behavioral interventions. In general, traditional psychotherapy aimed at addressing underlying intrapsychic conflicts or other interventions (including play therapy, art therapy, supportive therapy, and family therapy) not containing the behavioral elements described above have not been found to be effective for SM.

Speech and Language Therapy

As noted by the DSM-5, the criteria for SM include a caveat that the symptoms cannot be primarily due to a speech or language disorder. However, it is not uncommon for children with SM to also have some degree of speech or language difficulty that may exacerbate speech-related anxiety or self-consciousness. Although there is a lack of research on the efficacy of multimodal approaches, the combination of cognitive behavioral techniques and speech or language therapy can be quite effective for the treatment of children with comorbid SM and speech or language difficulties. In addition, even in cases without the significant contribution of speech or language dysfunction, the small, supportive group setting of auxiliary speech services found in many schools may provide a beneficial environment in which to implement the graded behavioral intervention approach described earlier.

Psychopharmacologic Treatments

Currently, there are no approved U.S. Food and Drug Administration (FDA) medications for the treatment of SM. However, several small studies suggest that SSRIs may be effective treatments for SM in children. This is not surprising, given the central role of anxiety in the phenomenology of the disorder. Results from a single double-blind, placebo-controlled trial of fluoxetine in children with SM indicated significant improvement in parent ratings of mutism and anxiety. Results from open trials and case studies have also supported the efficacy of various SSRI treatments. In addition, a large multisite study of fluvoxamine for the treatment of separation anxiety disorder, social phobia, and generalized anxiety disorder in children included several children with SM who appeared to benefit from this treatment. These results, along with the mounting evidence that SM is closely related to social phobia, have begun to establish SSRIs as an important treatment option for children with SM. However, given concerns about medication use in young children, medication is not the first option for many families. In practice, behavioral therapy, when practical and available, should be considered the first-line intervention strategy. In treatment-resistant cases, combination treatment (behavioral therapy plus medication) appears to be the strategy of choice.

FURTHER READINGS

*Bergman RL, Gonzalez A, Piacentini J, Keller ML. Integrated behavior therapy for selective mutism: a randomized controlled pilot study. *Behav Res Ther*. 2013;51(10):680–689.

*Bergman RL, Keller ML, Piacentini J, Bergman AJ. The development and psychometric properties of the selective mutism questionnaire. *J Clin Child Adolesc Psychol*. 2008;37:456–464.

Bergman RL, Piacentini J, McCracken JT. Prevalence and description of selective mutism in a school-based sample. *J Am Acad Child Adolesc Psychiatry*. 2002;41(8):938–946.

Black B, Uhde TW. Treatment of elective mutism with fluoxetine: a double blind, placebo-controlled study. *J Am Acad Child Adolesc Psychiatry*. 1994;33:1000–1006.

Bunnell BE, Mesa F, Beidel DC. A two-session hierarchy for shaping successive approximations of speech in selective mutism: pilot study of mobile apps and mechanisms of behavior change. *Behav Ther*. 2018;49(6):966–980.

*Carlson JS, Mitchell AD, Segool N. The current state of empirical support for the pharmacological treatment of selective mutism. *School Psychol Quart*. 2008;23:354.

Catchpole R, Young A, Baer S, Salih T. Examining a novel, parent child interaction therapy-informed, behavioral treatment of selective mutism. *J Anxiety Disord*. 2019;66:102112.

Cohan SL, Chavira DA, Stein MB. Practitioner review: psychosocial interventions for children with selective mutism: a critical evaluation of the literature from 1990–2005. *J Child Psychol Psychiatry*. 2006;47(11):1085–1097.

Cornacchio D, Furr JM, Sanchez AL, et al. Intensive group behavioral treatment (IGBT) for children with selective mutism: a preliminary randomized clinical trial. *J Consult Clin Psychol*. 2019;87(8):720–733.

Diliberto RA, Kearney CA. Anxiety and oppositional behavior profiles among youth with selective mutism. *J Commun Disord*. 2016;59:16–23.

Letamendi AM, Chavira DA, Hitchcock CA, Roesch SC, Shipon-Blum E, Stein MB. Selective Mutism Questionnaire: measurement structure and validity. *J Am Acad Child Adolesc Psychiatry*. 2008;47:1197–1204.

Manassis K, Tannock R, Garland EJ, Minde K, McINNES A, Clark S. The sounds of silence: language, cognition, and anxiety in selective mutism. *J Am Acad Child Adolesc Psychiatry*. 2007;46(9):1187–1195.

Milic MI, Carl T, Rapee RM. Similarities and differences between young children with selective mutism and social anxiety disorder. *Behav Res Ther*. 2020;133:103696.

*Muris P, Ollendick TH. Children who are anxious in silence: a review on selective mutism, the new anxiety disorder in DSM-5. *Clin Child Fam Psychol Rev*. 2015;18:151–169.

Muris P, Ollendick TH. Selective mutism and its relations to social anxiety disorder and autism spectrum disorder. *Clin Child Fam Psychol Rev*. 2021;24:294–325.

Nowakowski ME, Tasker SL, Cunningham CE, et al. Joint attention in parent-child dyads involving children with selective mutism. *Child Psychiatry Hum Dev*. 2011;42:78–92.

Oerbeck B, Stein MB, Wentzel-Larsen T, Langsrud Ø, Kristensen H. A randomized controlled trial of a home and school-based intervention for selective mutism—defocused communication and behavioural techniques. *Child Adolesc Ment Health*. 2014;19(3):192–198.

Oerbeck B, Overgaard KR, Stein MB, Pripp AH, Kristensen H. Treatment of selective mutism: a 5-year follow-up study. *Eur Child Adolesc Psychiatry*. 2018;27(8):997–1009.

Piacentini J, Bergman RL. Developmental issues in the cognitive treatment of childhood anxiety disorders. *J Cogn Psychother*. 2001;15:165–182.

Poole KL, Cunningham CE, McHolm AE, Schmidt LA. Distinguishing selective mutism and social anxiety in children: a multi-method study. *Eur Child Adolesc Psychiatry*. 202:30:1059–1069.

Research Unit on Pediatric Psychopharmacology Anxiety Study Group. Fluvoxamine for the treatment of anxiety disorders in children and adolescents. *N Engl J Med*. 2001;344:1279–1285.

*Scott S, Beidel DC. Selective mutism: an update and suggestions for future research. *Curr Psychiatry Rep*. 2011;13(4):251–257.

Schwenck C, Gensthaler A, Vogel F. Anxiety levels in children with selective mutism and social anxiety disorder. *Curr Psychol*. 2019;18:1–8.

Sharkey L, McNicholas F. Selective mutism: a prevalence study of primary school children in the Republic of Ireland. *Ir J Psychol Med*. 2012;29(01):36–40.

Starke A. Effects of anxiety, language skills, and cultural adaptation on the development of selective mutism. *J Commun Disord*. 2018;74:45–60.

Steffenburg H, Steffenburg S, Gillberg C, Billstedt E. Autism spectrum disorder in children with selective mutism. *Neuropsychiatr Dis Treat*. 2018;14:1163–1169.

Stein MB, Yang BZ, Chavira DA, et al. A common genetic variant in the neurexin superfamily member CNTNAP2 is associated with increased risk for selective mutism and social anxiety-related traits. *Biol Psychiatry*. 2011;69:825–831.

Steinhausen HC, Juzi C. Elective mutism: an analysis of 100 cases. *J Am Acad Child Adolesc Psychiatry*. 1996;35:606–614.

Toppelberg CO, Tabors P, Coggins A, Lum K, Burger C. Differential diagnosis of selective mutism in bilingual children. *J Am Acad Child Adolesc Psychiatry*. 2005;44:592–595.

Vogel F, Gensthaler A, Stahl J, Schwenck C. Fears and fear-related cognitions in children with selective mutism. *Eur Child Adolesc Psychiatry*. 2019;28(9):1169–1181.

Yeganeh R, Beidel DC, Turner SM. Selective mutism: more than social anxiety? *Depress Anxiety*. 2006;23:117–123.

52

Early-Onset Psychotic Disorders

Kaitlin Boyle, M.D., Linmarie Sikich, M.D., M.A., and Tara Chandrasekhar, M.D.

Positive psychotic symptoms such as hallucinations and delusions may be present in a number of child psychiatric disorders, including the primary psychotic disorders (schizophrenia and schizoaffective disorders) as well as more common disorders such as depression, bipolar affective disorder, severe anxiety, and posttraumatic stress disorder (PTSD). They can also occur in the context of medical illness such as delirium, or neurologic disorders, including seizure disorder, autoimmune encephalitis, genetic disorders, metabolic abnormalities, intoxication, and medication side effect. In addition, transient psychotic-like experiences, including sleep-related hallucinations and imaginary friends, may occur in young children with a mental age of fewer than 8 years who have no mental health problems. The diagnostic distinctions between these disorders are primarily related to the duration and context of the hallucinations and the presence of specific co-occurring symptoms.

Children and adolescents, as well as their families, who have primary psychotic disorders and affective disorders with psychotic features, are often devastated by these illnesses. These illnesses are frequently characterized by recurrent crises, interruption of normal developmental processes, a deteriorating course, and treatment resistance. Often the care of such youth requires extraordinary monitoring and advocacy by their parents, frequent hospitalizations, and multiple long-term interventions provided by mental health, educational, and social service agencies. Such care often incurs very high societal costs. However, many believe that if effective treatments can be identified and implemented early in the course of illness, this may translate to improved long-term outcomes. In addition, many researchers believe that it may be easier to identify the neurobiologic processes involved in the pathogenesis of schizophrenia and related disorders in affected youth because they have a more severe but clearly developmental form of the disorder that appears more strongly associated with genetic factors than adult-onset schizophrenia.

DEFINITION

Positive psychotic symptoms include hallucinations, delusions, other sensory and cognitive misperceptions, and bizarre or disorganized thinking and behavior including catatonia. Negative psychotic symptoms include flattened affect, apathy, avolition (difficulty making choices), poverty of speech, and poverty of thought content. Early-onset schizophrenia (EOS) is defined as schizophrenia with onset of clear positive symptoms prior to 18 years of age. Onset of schizophrenia prior to age 13 is referred to as very early onset or childhood onset. The diagnoses of psychotic disorders in children and adolescents are generally made using the same criteria as in adults.

The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) and its text revision (DSM-5-TR) requires that youth with schizophrenia, like adults with the disorder, have at least 6 months of positive or negative symptoms and significant social or occupational dysfunction or failure to achieve expected levels of social development. In contrast, The *International Classification of Diseases* 11th edition (ICD-11) requires only a month of symptoms and does not include any impairment criteria. In addition, many researchers believe that marked neurocognitive declines in executive functioning, general cognition, social cognition, and memory are a key feature of schizophrenia, despite not being included in the diagnostic criteria. Indeed these symptoms, not positive symptoms, seem most closely related to long-term functional outcome.

Schizophrenia spectrum disorders also include schizoaffective disorder and schizophreniform disorder. Schizoaffective disorder is defined by the prominence of psychotic symptoms, presence of significant periods of co-occurring major depression or mania, and psychotic symptoms present for at least 2 weeks in the absence of any mood symptoms.

Schizophreniform disorder is defined by its short duration of psychotic symptoms (1 to 6 months) and generally evolves into either schizophrenia or schizoaffective disorder.

Major depression with psychotic features and bipolar affective disorder meets all diagnostic criteria for the respective affective disorder but also include positive psychotic symptoms that are present only during periods of affective disturbance. Key distinguishing features of these disorders and other disorders with psychotic symptoms presenting during childhood are presented in Table 52–1.

HISTORY

Schizophrenia has been reported to occur in children since the diagnosis was introduced by Emil Kraepelin. From the early 1900s to 1980s, the term childhood schizophrenia or prepubertal schizophrenia was used to refer to a range of severe psychiatric disorders presenting in youth including autism, leading to uncertainty about the diagnosis of research subjects during this period. In the 1970s, Israel Kolvin and colleagues conducted seminal work to distinguish between primary psychotic disorders and autistic disorder. Beginning with ICD-9 and DSM-III, autism has been recognized as a distinct disorder and schizophrenia has been diagnosed using identical criteria in youth as adults.

COMPARATIVE NOSOLOGY

The diagnostic criteria for schizophrenia and schizoaffective defined by the ICD-11 are similar to those of DSM-5-TR. ICD-11, published in 2019, only requires 1 month of symptoms to make the diagnosis of schizophrenia versus the 6 months required by DSM-5-TR. In addition, there is no functional impairment criteria as required in the DSM-5-TR diagnostic criteria. Jorge Armenteros and colleagues found a high rate of diagnostic agreement between DSM-III-R, DSM-IV-TR, and ICD-10 in hospitalized psychotic adolescents. Similarly, there is good agreement between the diagnostic criteria for major depression with psychotic features and bipolar affective disorder with psychotic features between ICD-10 and DSM-IV-TR. It is highly likely that the same level of agreement will be found with increased use of DSM-5-TR and ICD-11 criteria.

Table 52–1.

Disorders Presenting with Psychotic Symptoms in Children and Adolescents

Disorder	Distinguishing Features
Early-onset psychosis (EOP)	Any of a number of disorders in which positive psychotic symptoms are a prominent feature.
Early-onset schizophrenia (EOS)	Schizophrenia with onset prior to age 18. ▸ Involves a change from usual functioning, typically within 1–5 years of diagnosis, often has negative and neurocognitive symptoms rather than just positive symptoms. ▸ Affective symptoms may occur but are not present for a substantial part of the illness. ▸ If patient also has an autism spectrum disorder, positive symptoms must include hallucinations or delusions. ▸ Impact on functioning may include failure to achieve developmentally expected social functioning or deterioration in social or occupational functioning. ICD-11 does not require impairment. ▸ DSM-5-TR requires symptoms to be present for 6 months, ICD-11 requires symptoms for only 1 month.
Schizoaffective disorder (SA)	▸ During a substantial portion of the psychotic illness, the affected individual also meets criteria for a major depressive episode, manic episode, or mixed manic/depressive episode. ▸ There are at least 2 weeks during the lifetime duration of the illness in which psychotic symptoms are present in the absence of any affective symptoms.
Schizophreniform disorder	▸ DSM-5-TR diagnosis reflecting schizophrenia criteria for at least 1 month but fewer than 6 months. ▸ Not included in ICD-11. ▸ Generally evolves into schizophrenia or schizoaffective disorders but may less frequently evolve into bipolar disorder.

Early-onset schizophrenia spectrum disorders (EOSS)	▶ Includes EOS, schizoaffective disorder, and schizophreniform disorder with onset prior to age 18 and Very Early Onset Schizophrenia (VEOS) or Childhood Onset Schizophrenia (COS). ▶ At this point, there is no clear evidence that the treatment or prognosis of early-onset schizoaffective disorder is significantly different from EOS.
Major depression with psychotic features	▶ Youth describes clear depressed mood vs. numbness or apathy. ▶ Psychotic symptoms are only present if mood symptoms are present. ▶ Psychotic features are usually mood congruent. ▶ Formal thought disorder is not usually present.
Bipolar disorder with psychotic features	▶ Psychotic symptoms are only present if mood symptoms are present. ▶ There is decreased need for sleep rather than insomnia. ▶ Negative symptoms are absent.
Substance-induced psychosis	▶ Psychotic symptoms are limited to periods of substance intoxication or withdrawal. ▶ Psychotic symptoms did not precede substance use. ▶ Psychotic symptoms generally resolve within 7–30 days of abstinence.
Posttraumatic stress disorder (PTSD)-associated psychosis	▶ Positive psychotic symptoms relate directly to trauma and are not associated with other areas of life or random events. ▶ There are other symptoms of PTSD.
Anxiety-related hallucinations	▶ Positive psychotic symptoms are limited to periods of high anxiety. ▶ There are no negative symptoms or signs of a formal thought disorder.

EPIDEMIOLOGY

There are a limited number of studies reporting on the population prevalence of rare positive psychotic symptoms (≤4) in children and adolescents. A meta-analysis of 12 studies involving a total of 24,103 youths from non–help-seeking samples described a median prevalence of brief psychotic symptoms of 9.8% among youths 11 to 21 years old. Prior studies have suggested that the prevalence of psychotic episodes decreases with age. Importantly, having such symptoms is associated with a threefold increase in one’s current or subsequent risk for having any psychiatric disorder and about a fourfold increase in risk for a schizophrenia spectrum disorder. In contrast, population-based studies of childhood psychiatric studies in the United States, which generally rely on semistructured interviews administered by nonclinicians, do not report any cases of schizophrenia and bipolar disorder. However, one study examining all the youths who had sought mental health care at a regional hospital in Germany serving ∼100,000 children found that only 3 between 6 and 10 years old (0.003%) were diagnosed with schizophrenia. In this same sample, the incidence of schizophrenia increased by about fourfold between 10 and 14 years to 13 of 100,000 (0.013%) and by another fourfold between 14 and 18 years to as much as 0.055%. These results are generally consistent with estimates from the National Institute of Mental Health (NIMH), suggesting the prevalence of childhood-onset schizophrenia is less than 0.04%. Further, at least 5% of individuals with schizophrenia (∼0.05% of the entire population) report onset of illness prior to 14 years of age and up to 40% of those with schizophrenia have been estimated to become ill prior to 18 years of age.

The reports of persistent psychotic symptoms in clinical samples are much higher and increase with the restrictiveness of the clinical setting. In one outpatient pediatric mood and anxiety clinic, 2,031 children aged 5 to 21 years (mean age 14.3 years) were evaluated using a structured diagnostic interview. Nearly 10% had definite or possible psychotic symptoms. Among those with psychotic symptoms, 62% had depressive disorders, 24% had bipolar disorder, and 14% had schizophrenia or schizoaffective disorder. Inpatient samples of children with depression have reported that as many as 60% will have psychotic symptoms. It is clear from the available studies that the presence of persistent psychotic experiences alone does not equate with a diagnosis of a schizophrenia spectrum disorder and instead serves as a marker for a variety of mental health disorders.

Given the greater prevalence of major depression and bipolar affective disorder in youth, it is likely that as many or more cases of pediatric psychosis will be related to affective disorders rather than schizophrenia. It appears that EOS and affective psychoses are more common in individuals with family histories of the respective disorders. EOS appears slightly more common in boys than girls, whereas major depression is much more common in female adolescents than in male adolescents. Both genders appear equally affected by bipolar disorder. A clear relationship with onset of puberty has not been demonstrated.

Clinicians should also be aware of potential racial disparities in the diagnosis of psychotic disorders. Studies have shown that African American patients are diagnosed with schizophrenia at a higher rate than White patients despite epidemiologic data suggesting similar prevalence rates. This highlights a very important issue suggesting potential clinician bias that needs to be recognized and addressed. A misdiagnosis can lead to many issues beyond not receiving the appropriate treatment.

ETIOLOGY

Schizophrenia is a syndrome presenting with varied symptoms and heterogeneous etiologies, almost all of which remain unknown. Schizophrenia is generally thought of as a neurodevelopmental disorder. Some researchers have speculated that multiple disruptions or “hits” to normal neurodevelopmental processes are required for schizophrenia to develop. Such disruptions might result from genetic vulnerabilities; intrauterine adversities as a result of famine, infection, or inflammation; or various environmental exposures later in life including severe psychological stress including trauma and exposure to drugs of abuse, particularly cannabis.

Disruption of normal neurodevelopmental processes, particularly those that involve maturation of the frontal cortex and its connections during the second and third decades of life have been repeatedly demonstrated in histopathologic studies. In adults, alterations in the morphology of specific subclasses of γ-aminobutyric acid (GABA)ergic interneurons and of the volume and synaptic density of pyramidal neurons that receive thalamic afferents have been repeatedly demonstrated. These differences suggest aberrations in the normal developmental processes of synaptic refinement and pruning. Further, subtle differences in neuronal connectivity, especially in the hippocampus and parietal, temporal, and frontal cortices, have repeatedly been identified in both histopathologic and imaging studies. These differences appear to lead to less efficient cortical processing as demonstrated by functional magnetic resonance imaging (fMRI). Given the correlation between these connections and cognitive function, such connectivity alterations may be a potential underlying mechanism of cognitive issues in schizophrenia.

EOS appears to be associated with more profound neurodevelopmental disruptions than adult-onset schizophrenia. Frequently individuals with EOS have social and attentional difficulties during early childhood. Many have evidence of autism spectrum disorder (ASD) or symptoms. Structural imaging studies have repeatedly demonstrated an exaggeration of the normal pattern of cortical volume loss that occurs throughout adolescence, with excessive and progressive reductions being especially prominent in the cingulate, temporal, and frontal cortices. Further, the magnitude of cortical volume loss has been associated with the severity of positive psychotic symptoms. In addition, studies of prodromal individuals, many of whom are adolescents, have shown that regional gray matter volume reductions, especially in these same regions are present before appearance of overt positive symptoms and progress during the first few years of the illness.

The heritability of schizophrenia has also been consistently demonstrated. However, the underlying genetic links are more complicated. The ability to perform genome-wide association studies has allowed for the detection of over 100 common independent loci that are significantly associated with increased risk for schizophrenia. These account for only a small increase in overall risk, leading many to argue that schizophrenia is a polygenic disorder. However, individuals with EOS are significantly more likely than those with adult-onset schizophrenia to have chromosomal abnormalities and DNA copy number variants, which individually pose a greater risk for schizophrenia. A prominent example is the 22q11.2 deletion associated with velocardiofacial syndrome (VCFS). Interestingly, many copy number variants also significantly increase the risk for other neuropsychiatric disorders, including ASD, intellectual disability, attention deficit hyperactivity disorder (ADHD), and epilepsy. Some argue that schizophrenia lies on a spectrum of neurodevelopmental disorders with the disorders listed above. Further, youths with EOS are at least twice as likely as individuals with adult-onset schizophrenia to have a parent with the disorder. Genetic abnormalities suggest the involvement of immunologic and neurodevelopmental functions.

Although multiple environmental factors have been associated with risk for schizophrenia, cannabis use has been of particular interest in recent years given that it may be modifiable. Several studies show a correlation between heavy cannabis use and psychotic experiences as well as higher rates of depression and anxiety. Use during adolescence has been linked with increased risk of psychosis as well as long-term cognitive deficits. Alterations to the endocannabinoid system by exogenous cannabinoids during the critical adolescent period potentially serve as the link between early cannabis use and increased risk for psychosis and cognitive effects.

DIAGNOSIS AND CLINICAL FEATURES

All types of early-onset psychosis are diagnosed using the same criteria as the adult forms of the disorder. However, the characteristics of specific symptoms often show important developmental characteristics that may be confusing for clinicians who infrequently work with pediatric patients. Hallucinations are frequently multimodal, with the affected youth seeing, hearing, and/or feeling the hallucinated being. Children will frequently name the hallucinated being, often using stereotypes such as “an angel,” “the devil,” or “a monster” or will refer to it on the basis of a physical characteristic such as “the red sweater guy” or “the man with no skin.” Youth very seldom complain about their hallucinations if not asked directly. Family members often are not aware of hallucinations but may endorse that the child asks about being called when no one has said anything. Delusions are seldom highly complex or systematic prior to age 16. Instead youth often report vague symptoms like “everyone is out to get me” or “people are laughing at me.” Often reports from collateral sources are needed to determine whether such reports are delusional or reality based. Frequently, youth with EOS present with nonspecific symptoms such as poor attention or increased aggression.

Onset of EOS is often insidious, occurring over a few years. In addition, premorbid symptoms such as being odd and socially isolated or having multiple (but generally mild) developmental delays in cognitive, motor, sensory, and social functioning are common and appear more severe than in adults with adult-onset schizophrenia. Thus, when assessing youth for EOS, it is important to ask about changes in functioning and behavior that have occurred over the past 3 to 4 years, to specifically inquire about hallucinations and delusions using language that the child understands, and to encourage the child to describe their experiences in detail using their own words. One should ascertain whether the child is in control of the phenomena, which is not seen with true hallucinations, or whether they can only moderate their own response to the phenomena. It is also crucial to determine the frequency and persistence of positive symptoms and associated negative symptoms. Finally, often, but not always, EOS is associated with a formal thought disorder.

Clinicians are frequently reluctant to diagnose EOS even when there is clear evidence of the disorder. Although the reasons for this are unclear, unfamiliarity with the developmental variations in symptoms, concerns about overlabeling, likely poor outcome, and stigma are frequently involved. However, failure to make the diagnosis when indicated may reduce access to appropriately intense interventions and relevant psychoeducational resources, increase the likelihood of treatment nonadherence, and result in hopelessness if the youth does not robustly respond to treatment. Further, longer duration of untreated psychosis has been consistently associated with poor prognosis in adult-onset schizophrenia. Notably, a review in 2016 found the mean duration of untreated psychosis to be 3.5× longer in EOS than in adult onset.

Particular attention should be paid to risk for suicide in patients with psychosis. People with schizophrenia have a significantly increased risk for suicide compared to the general population with some studies suggesting this risk is highest during the year after the first hospitalization and may be as high as 10% during the first decade of the illness. Vignette illustrating typical presentation of EOS:

Jack was first brought to a psychiatrist at age 12 when his parents were divorcing. He was obviously upset, was doing less well in school over the past year, and had gone on a hunger strike for the past month because he thought it would make his parents get back together. His parents noted that he had always been somewhat socially awkward, had a limited number of friends, and seldom had extended conversations. He was treated with fluoxetine (Prozac) and psychotherapy. After about 2 months, he was eating normally and seemed less preoccupied with his parents' divorce. However, he would make frequent statements that he needed to protect his younger brother, because he did not trust his parents to take care of the brother. He also said he was trying to find a house where he, his brother, and his father could live together.

A year later, after his mother remarried, he became markedly more irritable and was frequently noncompliant with her requests. His mother reported that there were episodes in which he would suddenly become extremely upset with little or no provocation. For example, she described an episode when they were out walking and he suddenly jumped on her back, saying that she looked like a prostitute. On another occasion, when his stepfather had instructed his brother to wash a glass that had been left out in the living room, Jack became very upset and said he planned to beat his stepfather to death with a golf club while his stepfather was sleeping. Jack's band instructor noted that he frequently seemed preoccupied and sometimes seemed not to know what to do next. Between these episodes, Jack was frequently pleasant and engaged superficially with the family, but had very seldom seen his friends or left the house. He experimented with marijuana by himself three or four times and said it made him feel calmer.

When asked about the incidents with his mother and stepfather, Jack seemed minimally concerned and did not provide an explanation. During his psychiatric evaluation, Jack reported that he often felt extremely angry without really knowing why. He acknowledged some periods of confusion. When asked directly about ever hearing anyone or anything talking to him, he said that his legs often talked to him and told him what to do. He denied significant mood symptoms and described his mood in nondescript ways. He had some poverty of content. At times his speech was disorganized and hard to follow.

Organic workup including MRI and electroencephalogram (EEG) was noncontributory. Family history was significant for a great uncle who had been odd, socially isolated, and had been hospitalized multiple times at the state psychiatric hospital.

Jack was started on an atypical antipsychotic and showed a remarkable improvement in his oppositionality and irritability. He also became more engaged with family members and his grandfather noted he was having conversations with Jack for the first time ever. However, Jack continued to have increasing trouble in school, particularly as he entered high school. Although his positive symptoms were well controlled, he was failing all but one class. It was unclear to what extent his difficulties in school were related to sedation from the medication or cognitive limitations. Cognitive testing showed he was now functioning in the low average to borderline range and had particular difficulty with abstract thought and problem solving. Reading achievement was grade appropriate but he was significantly behind in math. Several meetings with the schoolteachers were held and an individualized education plan was developed. However, he continued to have difficulty. Various medication changes were tried without significant change in his presentation. When the antipsychotic dose was lowered to try to reduce possible sedation, he became more irritable and had thoughts of beating his mother to death while she slept because she had removed his video games for 2 days as a punishment. His school program was changed from a college preparatory program to an occupational preparation program. He felt much less stressed by this but was unable to find and maintain work at a part-time job as required for graduation. He did not seem particularly concerned by this and had no specific plans for the future. His family was considering applying for disability so that he could maintain his insurance after he turned 18 and was trying to figure out how to more effectively structure his time. Eventually, he began treatment with clozapine (Clozaril) given the persistence of negative symptoms and vulnerability to positive symptoms when his antipsychotic dosage was reduced. He did well and has been able to remain out of the hospital since that time.

PATHOLOGY AND LABORATORY EXAMINATION

The physical examination of youth suspected of having EOS should include a careful neurologic examination. If there is a focal neurologic finding or impairment of consciousness, structural MRI is indicated. However, the volumetric, spectroscopic, and diffusion tensor imaging abnormalities consistently reported showing differences in group comparisons within research studies are not of sufficiently great magnitude to be used for diagnostic purposes. If there is any indication of episodic psychosis in the absence of mood symptoms or excessive sedation or confusion following psychotic behaviors, an EEG should be obtained to rule out a seizure disorder. Exposure to illicit substances should be assessed with a urine toxicology screen. Blood work should include screening for thyroid abnormalities, electrolytes, complete blood count, liver function tests, ceruloplasmin, and urinary copper for Wilson disease, heavy metals, and serum organic and amino acids to rule out a metabolic disorder. Antibodies to nuclear proteins and other proteins associated with lupus, as well as those detected with autoimmune encephalitis panels, should be obtained from both plasma and cerebrospinal fluid (CSF). Autoimmune encephalitis should be specifically considered if the onset of psychotic symptoms is very sudden and if neurologic symptoms, including seizures, emerge at about the same time. A paraneoplastic antibody panel or ultrasounds looking for teratomas should also be considered if the autoimmune encephalitis panel is negative and the patient does not respond to standard antipsychotic treatment. Negative antibody tests do not rule out autoimmune encephalitis since it is likely that all relevant antibodies have not yet been discovered. In cases with concomitant neurologic findings, it may be appropriate to undertake a trial of immunomodulatory therapy. Graus criteria for autoimmune encephalitis include acute onset of symptoms, focal neurologic finding(s), new seizures, CSF pleocytosis or MRI suggestive of encephalitis, and reasonable exclusion of alternative causes. It is very important to consider potential medical etiologies as these are managed differently than a primary psychotic disorder. Table 52–2 describes the components of a medical workup.

Table 52–2.
Medical Workup for First-Episode Psychosis

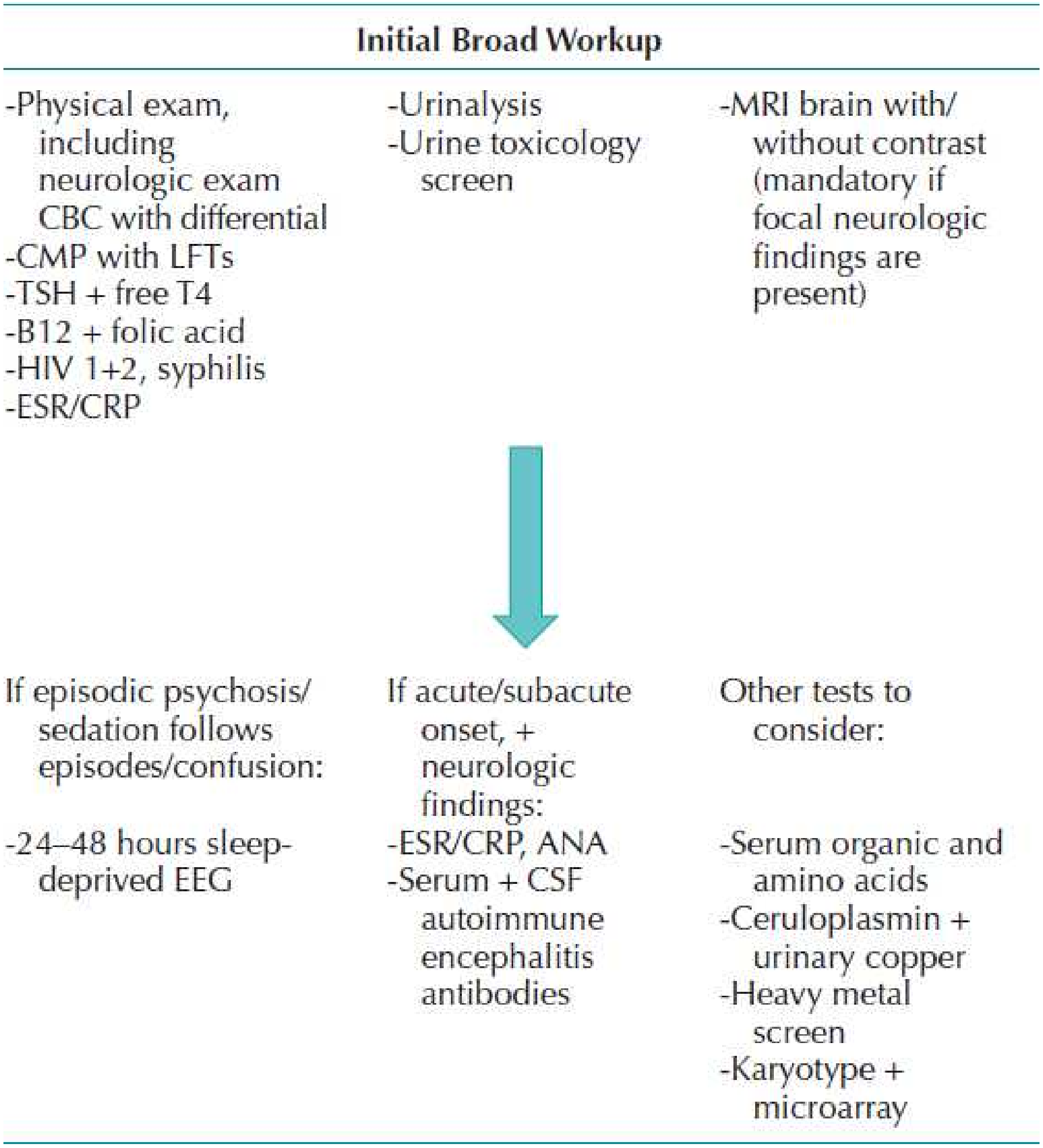

Initial Broad Workup		
-Physical exam, including neurologic exam CBC with differential -CMP with LFTs -TSH + free T4 -B12 + folic acid -HIV 1+2, syphilis -ESR/CRP	-Urinalysis -Urine toxicology screen	-MRI brain with/without contrast (mandatory if focal neurologic findings are present)
If episodic psychosis/sedation follows episodes/confusion:	If acute/subacute onset, + neurologic findings:	Other tests to consider:
-24–48 hours sleep-deprived EEG	-ESR/CRP, ANA -Serum + CSF autoimmune encephalitis antibodies	-Serum organic and amino acids -Ceruloplasmin + urinary copper -Heavy metal screen -Karyotype + microarray

Genetic testing including testing for chromosome 22q11.2 deletion syndrome should also be considered, particularly if the youth has a dysmorphic appearance, any congenital abnormalities or learning disabilities, or other cognitive limitations. Approximately 25% of individuals with 22q11,2 deletion syndrome become psychotic during adolescence, with most experiencing schizophrenia. Since the incidence of VCFS is estimated at 1 in 4,000 live births, as many as 10% of individuals with schizophrenia may have VCFS (Fig. 52–1). Projective psychological testing may be useful in confirming suspected cognitive disorganization.

However, failure to demonstrate psychotic disorganization during projective testing is not sufficient to reject the diagnosis. In addition, neurocognitive evaluation including intelligence quotient (IQ), auditory and visual memory, working memory, problem solving, processing speed, and attention measures may be useful in helping adapt academic expectations for individuals with EOS who often are expected to perform at the same level as prior to onset of their illness despite significant changes in their capabilities.

DIFFERENTIAL DIAGNOSIS

Misdiagnosis of EOS is a major concern. Factors that may lead to misdiagnosis include children's inherent difficulties in describing psychotic symptoms that are outside their normal experience, overinterpretation of transient psychotic symptoms, and the high prevalence of positive psychotic symptoms in child psychiatric disorders other than EOS. However, failure to recognize EOS when present may slow the implementation of appropriate treatments and worsen long-term prognosis.

Evaluation of a child or adolescent who may have psychotic symptoms should always include separate interviews of the child and their guardian with a focus on clarifying the child's developmental trajectory, changes in functioning, internal experiences elicited in a developmentally sensitive way, past exposure to trauma, and environmental stressors. Family history, including information about extended psychiatric hospitalizations and suicides, should also be obtained. The clinician can usually make an accurate diagnosis with attention to co-occurring symptoms, the time course of symptoms, and strict adherence to DSM-5-TR criteria. However, affected youth and their families should be educated that the diagnosis will be re-evaluated over time as additional information becomes available.

Schizoaffective Disorder

Like EOS, schizoaffective disorder is characterized by persistent positive and/or negative psychotic symptoms and functional decline or failure to realize expected developmental progress. However, during a substantial period of the active and residual phases of illness, the individual also experiences prominent mood symptoms that meet the criteria for a major depressive episode, manic episode, or mixed manic episode. Psychotic symptoms must be present in the absence of any mood symptoms for at least 2 weeks during the lifetime duration of the illness. If an individual is in the midst of their only psychotic episode and has had significant mood symptoms throughout, it may not be possible to distinguish between schizoaffective disorder and an affective psychosis. However, the presence of negative symptoms would increase the likelihood of a schizoaffective disorder. EOS with a secondary major depression may be distinguished longitudinally because the depressive symptoms are not present for a substantial period of the illness.

Major Depression with Psychotic Features

Individuals with major depression are generally able to describe their mood vividly rather than simply label it as depressed. In contrast, individuals with EOS may say they are depressed but instead describe a state of apathy and emotional numbing that reflects their experience of negative symptoms. Insomnia, failure to enjoy activities, social withdrawal, lack of motivation, cognitive slowing, agitation, suicidality, and psychomotor retardation may be seen in both disorders. Likewise, weight loss and decreased food intake may be seen in either disorder. However, in EOS, these symptoms are usually the result of delusional beliefs about the food or one’s weight. Although mood-incongruent hallucinations and delusions may occur in major depression, they are not common.

Bipolar Affective Disorder with Psychotic Features

It can be extremely difficult to distinguish a mixed manic state from EOS or schizoaffective disorder because hallucinations, delusions, irritability, and agitation are common in all three disorders. The primary distinguishing factors are the presence of a decreased need for sleep and hypersexuality in bipolar disorder versus negative symptoms in the schizophrenia spectrum disorders. Youth with bipolar disorder, especially those who require hospitalization, appear to have greater numbers of varied psychotic symptoms than adults hospitalized with bipolar disorder.

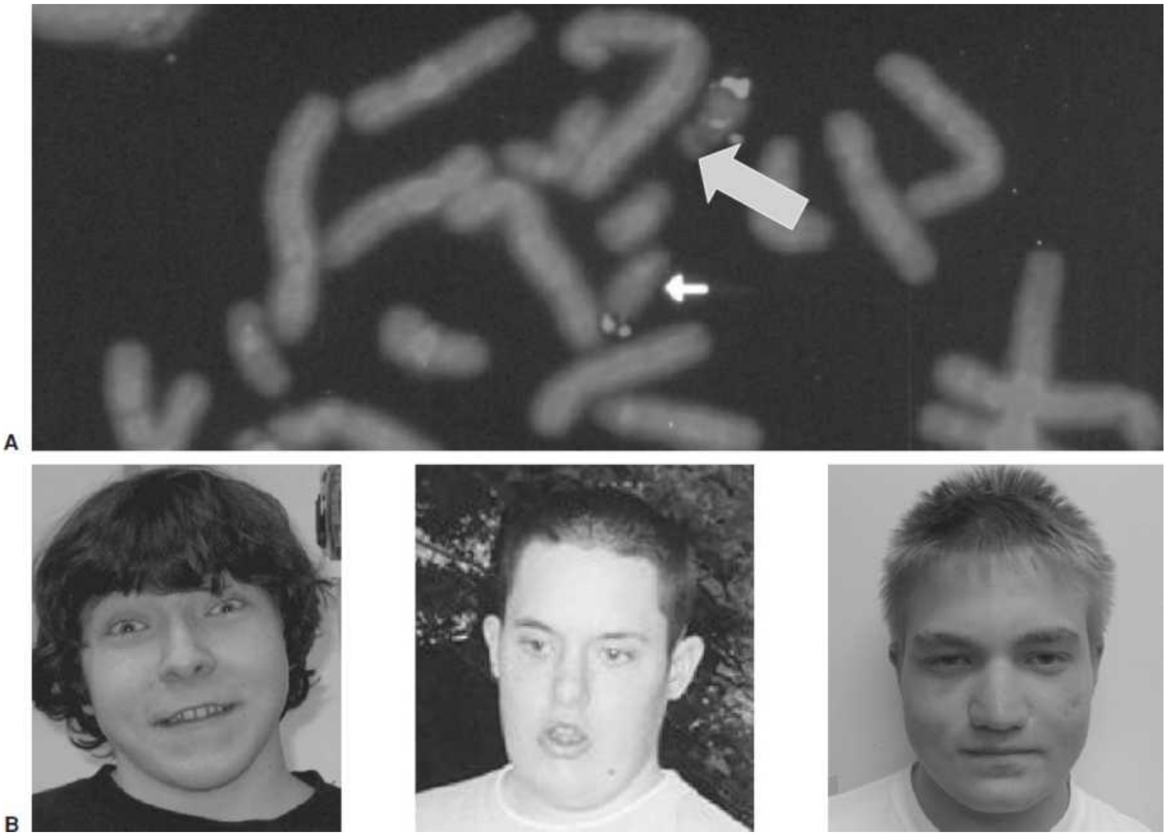

FIGURE 52–1. Features of chromosome 22q11 deletion syndrome (velocardiofacial syndrome [VCFS]). **A:** Demonstrates fluorescent in situ hybridization. One bright probe marks the end of each of the two chromosome 22s. Another probe, which binds to a marker within the typical large deletion in VCFS, is present on the upper chromosome 22 (noted with the larger *gray arrow*) but not on the lower chromosome 22 (noted with a smaller *white arrow*). **B:** Illustrates the facial features of three youths with VCFS. Note the elongated face, tubular nose with a wide nasal bridge, bulbous tip, and small lateral nares. Frequently individuals with the 22q11 deletion have low facial tone, small mouths, learning disabilities or mild mental retardation, and a variety of congenital malformations, especially of the heart and palate. A history of feeding problems in infancy is common.

Substance-Induced Psychosis

In substance-induced psychosis, positive symptoms are limited to periods of substance intoxication or withdrawal and are not better accounted for by a diagnosis of EOS, schizoaffective disorder, or affective disorder. DSM-5-TR notes that such symptoms should resolve within 1 month of discontinuing the substance, although in clinical practice, most substance-induced psychoses in youth resolve much more quickly (e.g., 7 to 10 days). If positive psychotic symptoms have been present prior to the onset of substance use or during periods of sustained abstinence from substances, one of the other psychotic diagnoses is more appropriate. In addition, one is much more likely to see negative symptoms in schizophrenia spectrum disorders than in substance-induced psychosis, although a decline in functioning is common in both. Comorbid substance use is common in older adolescents with EOS. Substance use, particularly cannabis use, appears to be both a precipitating and exacerbating factor. Nonetheless, many youths with EOS appear to self-medicate with drugs of abuse.

Autism Spectrum Disorders

Studies have demonstrated a genetic overlap between ASDs and schizophrenia. Many researchers also agree that the risk of developing psychotic disorders is increased within the ASD population compared to the general population. ASD and EOS are distinguished by their developmental course and the presence of hallucinations and formal thought disorder. Although ASD may not be diagnosed until late childhood, the onset of symptoms prior to 3 years of age is required to make a diagnosis. Such abnormalities may be present in individuals with EOS and schizoaffective disorder, but additional symptoms should present later in life that leads to the psychotic diagnosis. Although individuals with ASD may be concrete and rigid in their thinking and verbal responses, they do not exhibit a formal thought disorder. Further, in order to make the diagnosis of a primary psychotic disorder in an individual with ASD, there must be clear hallucinations or delusions and a change in the level of functioning.

Anxiety Disorders and Posttraumatic Stress Disorder

Transient hallucinations may occur in youth with severe anxiety disorders. However, these youths seldom experience negative symptoms and are less likely to experience delusions or a formal thought disorder. In individuals with PTSD, the psychotic symptoms are generally limited to reliving events that occurred during the trauma and to the perpetrator of the trauma or very similar themes. Hallucinations that are very detailed or specific to certain situations can also be an indicator of a trauma-related phenomenon. Data regarding the long-term course of individuals with these symptoms is lacking.

Organic Disorders

The list of potential organic etiologic agents of EOS is extensive. Entities that must be considered include (1) delirium, (2) seizure disorders, (3) central nervous system lesions (e.g., brain tumors, congenital malformations, and head trauma), (4) metabolic disorders (e.g., endocrinopathies and Wilson disease), (5) autoimmune disorders (e.g., autoimmune encephalitis and lupus), (6) neurodegenerative disorders (e.g., Huntington chorea and lipid storage disorders), (7) infectious diseases (e.g., encephalitis, meningitis, or human immunodeficiency virus [HIV]-related syndromes), and (8) toxic encephalopathies (both substances of abuse and medications). Individuals should be specifically asked about the use of solvents, phencyclidine, hallucinogens, cannabis, stimulants (including street-acquired amphetamines), synthetic or designer drugs, corticosteroids, and anticholinergic agents. Also, it is important to examine for comorbid medical or developmental disorders that will likely impact the type and extent of psychosocial treatments required. For example, there is an increased risk of psychotic disorders associated with 22q11.2 deletion syndrome, which is also associated with significant learning disabilities.

Emily, a 16-year-old girl with no previous psychiatric history, was brought to the emergency room due to a sudden behavioral change and suicidal ideation. A previously cheerful and accommodating girl, for the past 2 weeks, she had been increasingly irritable, moody, and easily frustrated. The day prior to presentation, she began endorsing suicidal ideation and expressing a belief that her thoughts were controlled by an external presence (whom she referred to as "the man"). She also reported that her eyes had been replaced by cameras that allowed her to see ghosts. She was admitted to an inpatient psychiatric unit, where she was agitated, pacing, and continuing to endorse hallucinations. Her physical examination was notable for ataxia and movement disorder (semirepetitive orofacial and limb movements). Routine labs (electrolyte, blood count, liver enzymes, thyroid function, and urine drug screen) were unrevealing. An EEG was grossly normal, and her brain MRI showed no evidence of mass lesion or midline shift. She was treated with risperidone in escalating doses with little benefit.

On hospital day 7, she developed increasing fluctuations in her blood pressure and pulse and then experienced a generalized tonic–clonic seizure. She was transferred to the intensive care unit due to worsening dysautonomia and central hypoventilation. Lumbar puncture results showed protein, glucose, neutrophils, and lymphocytes within the normal range. Though CSF anti-N-methyl-D-aspartate (NMDA) antibodies were negative, serum anti-NMDA receptor antibody panel was positive, increasing the suspicion of anti-NMDA receptor encephalitis as a cause of her symptoms. Imaging for an underlying tumor revealed an ovarian teratoma, which was surgically removed. She was concurrently treated with corticosteroids and experienced a gradual improvement in her psychotic symptoms and remission of her movement disorder. At a follow-up visit 3 months later, she had returned to school part-time. She denied psychotic symptoms but reported residual fatigue and impaired concentration.

COURSE AND PROGNOSIS

Subtle developmental symptoms are frequently present throughout childhood in individuals who have EOS. The onset of overt psychotic symptoms is frequently insidious, particularly in individuals younger than 15 years. In older adolescents, approximately equal proportions of youth have insidious and subacute onset, with rapid onset over the course of several days being rare. Rapid recovery from the initial episode appears rarer than in adult-onset schizophrenia. In contrast, many youths with EOS show improvement but continue to experience significant positive symptoms despite medication adherence for 2 to 5 years. The course tends to be chronic rather than episodic, as is reported in adult-onset schizophrenia. Further, there appears to be a significant decline in cognitive functioning during the 4-year period surrounding the onset of the illness. This cognitive decline appears to reflect both biologic brain processes and an interruption of normal development and learning. A recent study found that 30% of individuals with EOS attempted suicide over the first 10 years of illness, with one in six of those who attempted completing suicide. Suicidality was associated with more depressive symptoms and fewer negative symptoms in the first episode.

Long-term prognosis in EOS also appears worse than in adult-onset schizophrenia. A meta-analysis of 320 studies of outcomes in adult-onset schizophrenia found that about 40% of individuals had good outcomes, whereas about 60% had a chronic disabling course. In contrast, a number of studies have found that very few individuals with EOS (5% to 25%) experience good outcomes and that 75% to 80% have poor outcomes, with about half of those having an extremely poor outcome. Few live independently as adults (7% to 29%), participate in full-time education, or are employed in the open market (17% to 28%). Another study found that fewer (25%) individuals with psychosis complete high school compared to age-matched peers (84%). Treatment resistance may also be more common, although this has not been adequately studied. Figure 52–2 schematically illustrates the differences in the course between EOS and adult-onset schizophrenia.

TREATMENT

The treatment of EOS is multimodal and should include pharmacotherapy, family and patient psychoeducation and psychotherapy, and interventions in the community to support ongoing education and transition to employment when appropriate. The evidence is strongest for pharmacotherapy. A variety of traditional antipsychotics have demonstrated efficacy in reducing psychotic symptoms. A series of studies have compared various atypical antipsychotics to placebo in adolescents with EOS and have demonstrated considerable antipsychotic efficacy, frequently reducing the Positive and Negative Syndrome Scale (PANSS) by about 20%. In these studies, olanzapine (Zyprexa) has been demonstrated to be associated with the greatest acute weight gain and lipid changes. Risperidone (Risperdal) seems to have more moderate acute weight changes, and aripiprazole (Abilify) more limited acute weight changes but a greater frequency of akathisia. Extrapyramidal symptoms appear infrequent but do occur. Only a few comparative trials have been done. These studies have generally found that the various antipsychotics differ in their side-effect profiles but have few differences in efficacy. Specifically, midpotency traditional agents appear to have relatively little impact on weight, lipids, and glucose tolerance but are modestly more likely to cause moderate or severe extrapyramidal symptoms, including akathisia. Significantly, these symptoms are generally well controlled with the addition of anticholinergic medications or propranolol. There is also a strong body of work indicating that individuals with treatment-resistant EOS respond significantly better to clozapine than to haloperidol (Haldol) or olanzapine.

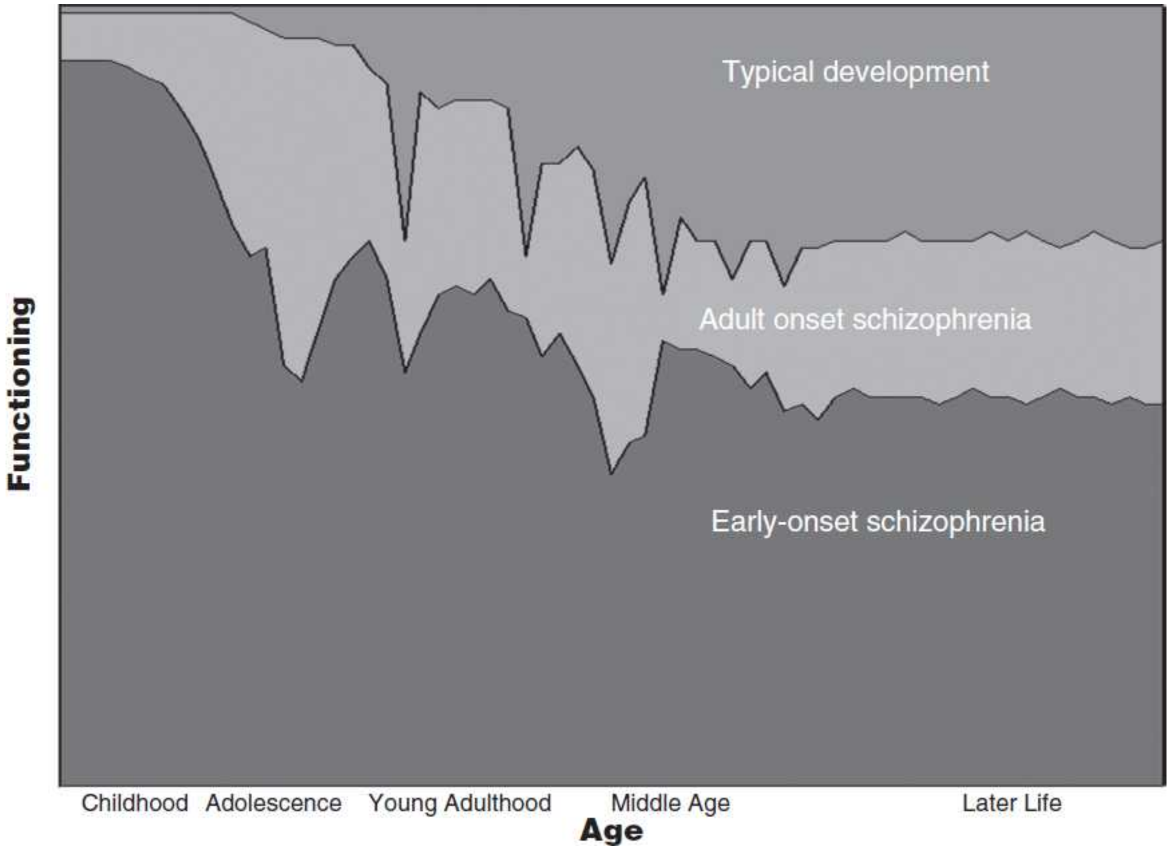

FIGURE 52–2. The comparative course of early-onset schizophrenia (EOS) and adult-onset schizophrenia. Note the lower premorbid functioning of EOS, more insidious onset, and more chronic course with more limited periods of recovery.

There is much debate regarding the role of Electroconvulsive (ECT) in children and adolescents. Different regions of the United States have various rules and restrictions around the use of ECT in this population, often making it difficult for this modality to be considered. The evidence that is available is reassuring that ECT can be as safe and effective in the pediatric population as in adults.

Compared to the literature on adult schizophrenia, few studies of psychosocial interventions have been conducted thus far in EOS. However, both cognitive remediation therapy (CRT) and supported employment interventions show promise in improving the functioning of affected youth. A recent study found that clinically stabilized youth who received CRT experienced significant improvements in verbal memory and executive functions and modest improvement in daily living, adaptive functioning, and reduced family burden compared with controls. Most of these changes were maintained after 3 months. Similarly, a small

Australian study of supported employment (Individual Placement and Support [IPS]) found that youth engaged in IPS gained significantly more jobs (23 vs. 4), earned significantly more money ($2,432 vs. $0), and worked for a longer time period (45 vs. 19 days). An overall positive vocational outcome (return to work or education) was achieved by 85% of the subjects, compared to 6% in the control group. Therapeutic interventions must focus on functional improvement in addition to symptomatic improvement in order to facilitate recovery. Targets of treatment include reduction in self-stigma and caregiver burden and restoration of self-esteem, self-efficacy, hope, and optimism.

Specialized first-episode psychosis programs have been developed to deliver multimodal care to individuals early in their course of illness. Often these programs include interventions incorporating medication management, cognitive behavioral therapy, family interventions, and work and/or school support services. Two-year outcome studies from the NIMH Recovery Aftern an Initial Schizophrenia Episode (RAISE) Early Treatment Program revealed significantly greater improvement in quality of life and symptom improvement versus care as usual.

FUTURE DIRECTIONS

Improving the recognition and timely diagnosis of early-onset psychosis will be important in reducing the disability and personal and societal costs associated with these debilitating illnesses. It will be important to further characterize the long-term course of EOS and other early-onset psychotic disorders, particularly with regard to neurocognitive and biologic outcomes. In addition, it is essential that psychosocial treatments such as cognitive remediation, social skills training, and supported education and employment programs are rigorously evaluated and manualized where appropriate. More effective and safe medication strategies are also required to provide the most effective and safest treatment for these highly vulnerable individuals. This will require both more comparative trials and more studies of long-term outcomes. The development of additional psychopharmacologic treatments, particularly those that are more focused on the neurodevelopmental processes that appear so prominent in EOS, is also essential. Finally, EOS also can serve as a more homogenous model for understanding the pathophysiology of schizophrenia and identifying key neurodevelopmental pathways in which to intervene.

FURTHER READINGS

Armando M, Pontillo M, Vicari S. Psychosocial interventions for very early and early-onset schizophrenia: a review of treatment efficacy. *Curr Opin Psychiatry*. 2015;28(4):312–323.

Birnbaum R, Weinberger DR. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat Rev Neurosci*. 2017;18(12):727–740.

Driver DI, Thomas S, Gogtay N, Rapoport JL. Childhood-onset schizophrenia and early-onset schizophrenia spectrum disorders: an update. *Child Adolesc Psychiatr Clin N Am*. 2020;29(1):71–90.

Gara MA, Vega WA, Arndt S, et al. Influence of patient race and ethnicity on clinical assessment in patients with affective disorders. *Arch Gen Psychiatry*. 2012;69(6):593–600.

Graus F, Titulaer MJ, Balu R, et al. A clinical approach to diagnosis of autoimmune encephalitis. *Lancet Neurol*. 2016;15:391–404.

Healy C, Brannigan R, Dooley N, et al. Childhood and adolescent psychotic experiences and risk of mental disorder: a systematic review and meta-analysis. *Psychol Med*. 2019;49(10):1589–1599.

Hjorthøj C, Stürup AE, McGrath JJ, Nordentoft M. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. *Lancet Psychiatry*. 2017;4(4):295–301.

Jiang L, Xu Y, Zhu XT, Yang Z, Li HJ, Zuo XN. Local-to-remote cortical connectivity in early-and adulthood-onset schizophrenia. *Transl Psychiatry*. 2015;5:e566.

Kane JM, Robinson DG, Schooler NR, et al. Comprehensive versus usual community care for first-episode psychosis: 2-year outcomes from the NIMH RAISE Early Treatment Program. *Am J Psychiatry*. 2016;173(4):362–372.

Killackey E, Alvarez-Jiminez M, Allot K, Bendall S, McGorry P. Community rehabilitation and psychosocial interventions for psychotic disorders in youth. *Child Adolesc Psychiatric Clin N Am*. 2013;22:745–758.

Kumar A, Datta SS, Wright SD, Furtado VA, Russell PS. Atypical antipsychotics for psychosis in adolescents. *Cochrane Database Syst Rev*. 2013;10:CD009582.

Lewandowski KE. Cognitive remediation for the treatment of cognitive dysfunction in the early course of psychosis. *Harv Rev Psychiatry*. 2016;24(2):164–172.

Liu JJ, Norman RMG, Manchanda R, De Luca V. Admixture analysis of age at onset in schizophrenia: evidence of three subgroups in a first-episode sample. *Gen Hosp Psychiatry*. 2013;35(6):664–667.

Lubman DI, Cheetham A, Yücel M. Cannabis and adolescent brain development. *Pharmacol Ther*. 2015;148:1–16.

Puffer CC, Wall CA, Huxsahl JE, Frye MA. A 20 year practice review of electroconvulsive therapy for adolescents. *J Child Adolesc Psychopharmacol*. 2016;26(7):632–636.

Puig O, Penadés R, Baeza I, et al. Cognitive remediation therapy in adolescents with early-onset schizophrenia: a randomized control trial. *J Am Acad Child Adolesc Psychiatry*. 2014;53(8):859–868.

Puig O, Penadés R, Baeza I, et al. Processing speed and executive functions predict real-world everyday living skills in adolescents with early-onset schizophrenia. *Eur Child Adolesc Psychiatry*. 2012;21(6):315–326.

Reichert A, Krelker S, Mehler-Wex C, Warnke A. The psychopathological and psychosocial outcome of early-onset schizophrenia: preliminary data of a 13-year follow-up. *Child Adolesc Psychiatry Ment Health*. 2008;2(1):6.

Remschmidt H, Martin M, Fesichhaker C, et al. Forty-two-years later: the outcome of childhood-onset schizophrenia. *J Neural Transm (Vienna)*. 2007;114(4):505–512.

Rund BR, Barder HE, Evensen J, et al. Neurocognition and duration of psychosis: a 10-year follow-up of first-episode patients. *Schizophr Bull*. 2016;42(1):87–95.

Sanchez-Gistau V, Baeza I, Arango C, et al. Predictors of suicide attempt in early-onset, first-episode psychoses: a longitudinal 24-month follow-up study. *J Clin Psychiatry*. 2013;74(1):59–66.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421–427.

Schneider M, Debbané M, Bassett AS, et al; International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. Psychiatric disorders from childhood to adulthood in 22q11.2 deletion syndrome: results from the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. *Am J Psychiatry*. 2014;171(6):627–639.

Schothorst PF, Emck C, van Engeland H. Characteristics of early psychosis. *Compr Psychiatry*. 2006;47(6):438–442.

Sikich L, Frazier JA, McClellan J, et al. Double-blind comparison of first- and second-generation antipsychotics in early-onset schizophrenia spectrum disorders (TEOSS) study. *Am J Psychiatry*. 2008;165:1420–1431.

Stentebjerg-Olesen M, Pagsberg AK, Fink-Jensen A, Correll CU, Jeppesen P. Clinical characteristics and predictors of outcome of schizophrenia-spectrum psychosis in children and adolescents: a systematic review. *J Child Adolesc Psychopharmacol*. 2016;26(5):410–427.

Tillman R, Geller B, Klages T, Corrigan M, Bohofner K, Zimerman B. Psychotic phenomena in 257 young children and adolescents with bipolar I disorder: Delusions and hallucinations (benign and pathological). *Bipolar Disord*. 2008;10:45–55.

Ventura J, Subotnik KL, Gitlin MJ, et al. Negative symptoms and functioning during the first year after a recent onset of schizophrenia and 8 years later. *Schizophr Res*. 2015;161(2–3):407–413.

Volkow ND, Swanson JM, Evins AE, et al. Effects of cannabis use on human behavior, including cognition, motivation, and psychosis: a review. *JAMA Psychiatry*. 2016;73(3):292–297.

World Health Organization. International *Classification* of *Diseases* for *Mortality* and *Morbidity Statistics*. 11th revision; 2018.

53

Child Psychiatry: Psychiatric Treatment

53.1 Psychodynamic Psychotherapy (Child Psychiatry)

DAVID L. KAYE, M.D., AND MICHAEL SHAPIRO, M.D.

Psychodynamic psychotherapy (PDP), rooted in psychoanalytic theory and used interchangeably with psychoanalytic psychotherapy by most practitioners, has a long and storied tradition in child and adolescent psychiatry both in the United States and internationally. While for many years the backbone of child psychiatric theory and practice, it was under siege for much of the past four decades from different quarters. The ascendancy of neurobiologic approaches and understanding, third party payers' demands for briefer and less expensive treatments, and the public's demand for empirical support all contributed to the diminished influence of psychoanalysis. Despite this, psychoanalysis has remained a vibrant, and still the most comprehensive, framework for understanding psychotherapy and the human mind. In the last two decades increasing empirical support for PDP with adults, children, and adolescents has energized the approach. The groundbreaking review by Shedler, The Efficacy of Psychodynamic Psychotherapy, published in American Psychologists in 2010 was a landmark in establishing the efficacy of PDP with adults. Since then Leichsenring has published a recent systematic review (2015) that concludes that PDP is efficacious for a wide range of adult mental disorders. In 2012, the American Academy of Child and Adolescent Psychiatry published the first ever Practice Parameter for PDP with children. Two volumes of the Child and Adolescent Psychiatric Clinics of North America (January, April 2013) were devoted to PDP. Evidence-based, manualized approaches such as interpersonal therapy (IPT) and mentalization-based therapy (MBT) are based solidly in PDP principles. Further there has been a softening of the boundaries between psychotherapy treatments with inclusion of aspects of different therapeutic approaches now commonplace in child therapy.

Although the place of psychoanalytic psychotherapy within child and adolescent psychiatry has declined from the hegemony it once held, it has remained a primary framework for many practitioners and a psychotherapeutic foundation of many, if not most, residency training programs. In the past, it was often seen as a stand-alone treatment for child psychiatric disorders. In current clinical work, practitioners work pragmatically, flexibly, and in concert with other approaches, including family therapy, behavior therapy, cognitive-behavior therapy, psychopharmacology, and school interventions. Regardless of the approach, the understanding of patients stems from an overarching psychoanalytic framework. Psychoanalytic concepts also continue to be applied broadly to clinical work in residential or other intensive treatment milieus, social services, police, juvenile justice settings, and schools. In looking over his career Leon Eisenberg, a long-time critic of psychoanalysis, observed that:

> “Where I erred was in failing to appreciate the powerful and lasting contribution psychoanalysis made to psychiatry by teaching trainees to listen to patients and to try to understand their distress, rather than merely to classify them by some diagnostic algorithm, or snow them with drugs, or lock them away, or release them to homelessness. Basing care on what is unique to the individual patient is central

to clinical competence, whether in medicine or surgery or psychiatry.” (Excerpt from Eisenberg L. The past 50 years of child and adolescent psychiatry: a personal memoir. *J Am Acad Child Adolesc Psychiatry*. 2001;40(7):743–748. doi: 10.1097/00004583-200107000-00008.)

Far from a single and monolithic set of ideas, psychoanalysis encompasses a wide diversity. Disagreement exists about the nature of the mind, development, psychopathology, and treatment. Despite this, psychodynamically oriented practitioners share the following precepts:

1. The primary emphasis and focus of PDP are the individual’s subjective, inner experience of self and others.
2. A primary value is the search for emotional truth. Seen as elusive and often hidden from ourselves, emotional honesty and truth is sought as an end unto itself. Through this pursuit and recognition, human beings have their best chances for a satisfying life.
3. A primary goal in treatment is self-acceptance and self-understanding through the instrument of a relationship. The primary “tool” of psychoanalytically informed therapy is the relationship between the doctor and the patient.
4. Unconscious mentation is critical to understanding patient’s subjective experiences and relationship patterns. Unconscious affects, motivations, and process (including defense mechanisms) are accepted as critical determinants of normal and abnormal human behavior.
5. Behavior, emotion, and relationship patterns can be understood through the lens of psychic determinism. Human behavior has a logic of its own, happens for psychological “reasons,” and is understandable through following a link of multiple impulses, affects, defenses, and their associated relational experiences. Often these links are unconscious and the work of therapy includes greater awareness of the impact of defenses and warded off emotions and wishes.
6. The child is the father to the man. Early experience in a child’s life is critical in initiating important and enduring patterns in perceiving, thinking, feeling, and behaving that are carried

forward in life via internalized beliefs, working models, assumptions, expectations, and attitudes (i.e., transference).

7. Treatment critically focuses on transference. Focusing on the here-and-now transference reactions of the patient are central to understanding the patient and promoting growth in the psychoanalytic process.
8. Treatment goals are focused on psychological growth broadly, not just symptom amelioration. For example, in child psychotherapy goals often include a return of the child to a healthy developmental pathway, including greater self-acceptance, self-awareness, and self-expression.
9. Emotional barriers (“resistance”) to psychological growth or resolution of symptoms can be expected in the course of treatment.

HISTORY

The first psychoanalytic reference to child cases was Sigmund Freud’s case of “Little Hans” in 1909. Although Freud was skeptical that children could be analyzed, in this case he provided supervision to Hans’ physician-father, who successfully analyzed the 5-year-old boy’s phobia of horses. A few years later Sandor Ferenczi reported his attempt to directly treat a child, but found that the child quickly became bored with the analytic process and wanted to return to his toys to play. It was not until a series of women psychoanalysts entered the scene that it became appreciated that play was the central medium of expression for children, and should be incorporated into work with those with emotional disturbance. Hermine Hug-Hellmuth, a teacher by prior training, was the first psychoanalyst to report the use of play in the treatment of children in 1921. She believed that treatment of children was primarily educational and limited in its goals.

By far, the major figures in the early development of psychoanalytic treatment of children were Anna Freud and Melanie Klein. Klein began to develop a technique for working with children while in Berlin in the early 1920s, and in 1926 moved to London, where she continued her work. She saw children as a cauldron of passionate drives (both aggressive and sexual) with rigid, harsh superegos. Her technique relied on equally powerful and direct interpretations of unconscious impulses. She appraised play as the equivalent of adult-free association and eschewed any role for parents in the treatment. Anna Freud, also a teacher by training, began her psychoanalytic training in the early 1920s and shortly thereafter formed a child psychoanalysis study group in Vienna. She viewed children as immature and undeveloped, necessitating a gentler and educative approach that was careful not to interpret too deeply for fear of overwhelming the child. She recognized work with the parents as difficult but necessary to further the development of the child and also to support treatment with the child. She saw the goals of treatment as being to return the child to their developmental path and not to effectuate a definitive or complete cure. In 1927, Anna Freud wrote the first book on child analysis, *Introduction to the Technique of Child Analysis.* The differences between Klein and Freud led to a confrontational Symposium on Child Analysis in May 1927. These “controversial discussions” brought points of disagreement into sharp focus, and this schism has persisted to the present day, albeit lessening in the recent past. The Kleinian school became dominant in the United Kingdom and South America, while the Freudian school held sway in North America.

Into this fray entered the analysts of the “independent school” in Great Britain in the 1940s, including Donald Winnicott, Wilfred Bion, and others associated with the object relations perspective. Starting in the Kleinian camp, they attempted to bridge the divide with the Freudians. Winnicott contributed enormously to the analytic understanding of child development and treatment, remaining perhaps the most well-recognized influence in the field. His rich body of work included an emphasis on the importance of the early mother–child relationship; the role of play, mirroring, and the establishment of the “holding environment” in healthy development and in analysis; the meaning of the transitional object; and the true versus false self.

In North America an additional perspective was emerging, beginning with Kohut's work in the 1970s on the development of the self. This evolution has continued to the present with major contributions from interpersonal, relational, and intersubjective theorists. Harry Stack Sullivan, Robert Stolorow, Stephen Mitchell, Joseph Lichtenberg, James Fosshage, and Irwin Hoffman among others have been major figures representing these perspectives. All of these authors have written primarily about clinical work with adults, with a few current authors applying the concepts to work with children (e.g., Gary Landreth, Charles Schaeffer). These theorists have also been informed by convergences with infant and attachment researchers such as Daniel Stern, John Bowlby, Mary Main, and Mary Ainsworth. Rather than emphasizing the role of drives, defenses, resistance, conflict, and other traditional concepts these authors assert that the basic building blocks of mental life are relations with others upon which major psychic structures are built. This reflects a shift in the foreground from an intrapsychic focus to a relational field model, from objectivism to constructivism. Consequently there has been a shift in emphasis in treatment from *interpretation* to the *"co-creation of developmentally needed relational experiences"* (Jay Greenberg and Mitchell). Recently there has been a convergence of attachment, self-psychological, and relational perspectives. This is embodied by the work of Peter Fonagy, Mary Target, Efrain Bleiberg, and others who have written about the central importance of *mentalization*, or self-reflective function. This ability to read and understand the feelings, thoughts, and impulses associated with human behaviors is fundamentally dependent upon the establishment of secure attachments. In the past several years this has led to MBT, a manualized therapy approach that has been evaluated empirically primarily in adolescents. Even more recently neuroscience-based understandings have begun to be integrated into PDP with books such as Daniel Siegel's *The Developing Mind* and Delgado, Strawn, and Pedapati's *Contemporary Psychodynamic Psychotherapy for Children and Adolescents: Integrating Intersubjectivity and Neuroscience*.

While clinical work with infants is beyond the scope of this chapter, a great deal of parent–infant psychotherapy has been based on psychodynamic principles. Psychodynamic theory has influenced the field beginning with Selma Fraiberg and early programs developed by Alicia Lieberman, up to present-day evidence-based treatments such as Parent–Child Interaction Therapy.

It should also be noted that the term “psychodynamic” was originally used to describe the ever-changing balance of mental and neurophysiologic energy intrinsic to humans’ inner life. This was contrasted with an approach that emphasized behavior, descriptive phenomenology, or anatomic structure. In this sense psychodynamic and psychoanalytic are not synonymous. In the United States there have been a number of influential writers who would be considered within this broad psychodynamic tradition, even though they were not psychoanalysts. These psychodynamic therapists (e.g., Frederick Allen, Clark Moustakas, Ann Jernberg, John Meeks, Virginia Axline, and more currently Charles Schaefer, Garry Landreth) have also had a major place in the history of child and adolescent psychotherapy and continue to influence current practice.

THEORETICAL ISSUES IN PSYCHODYNAMIC WORK WITH CHILDREN

Beginning from its earliest days child psychodynamic psychotherapists have recognized that there are developmental differences between child and adult patients. While the overarching theoretical precepts of PDP reviewed above hold, there are fundamental differences in work with children and adolescents as compared with adults.

Children Are Not “Little Adults”

Child patients differ in significant ways from adults. These differences are crucial in informing practice. First, children’s limitations in language, cognition, and experience must be taken into account to engage and sustain children in treatment. Because of cognitive and emotional limitations their understanding of the problem or the need for treatment is likely to be incomplete or inaccurate. Children tend to externalize problematic issues and have little ability to understand how therapy might help them now, let alone in the future. Further they have limited capacity for tolerating pain and anxiety, creating another barrier for keeping child patients in treatment. Meeting with a stranger who will meet and talk intimately with the child is certainly extraordinary, and may be confusing and threatening. For all these reasons the child therapist must *actively* engage*,* educate, and work with the child to create the secure space needed to establish a therapeutic alliance.

Another major difference is that children do not come to treatment at their own request. They are brought by their parents, often for reasons that are more of concern to the adults rather than the child. Further, the stigma associated with treatment is often difficult for children (especially adolescents, who are struggling to feel “normal” and “like other kids”). The reasons for treatment must be explored and explained in a way that is understandable and meaningful to the child. The rationale that engages the child may be different from the reasons the parents bring the child in.

A 7-year-old boy was brought for treatment by working-class parents who were concerned about depressive irritability and argumentativeness, which had created great tension in the family. When seen individually the therapist asked the child about his understanding of why they were meeting. The child responded “I don’t know.” The therapist then asked why the child thought the parents had brought him, and he again avoided the question. Later in the interview the therapist returned to the issue while the child played out a puppet scene of different animals fighting. The therapist explored how the animals felt and the child said “sad.” As the play developed the therapist explored whether the child might also be sad. This seemed to resonate and the child nodded his head “yes” and went back to playing with the puppets. As the session closed the therapist suggested that in meeting together they might come to understand “why you feel sad and what we can do to help you feel better. How does that sound to you?” The boy looked up shyly and nodded his approval.

The Child Psychotherapist Must Recognize the Critical and Central Role of the Parents or Caretakers in the Child’s Life

In the early days of child PDP the child therapist would have no contact with the parents, while a social worker would meet alone with the parents. Over the years the importance of establishing and utilizing the relationship with the parents moved more to the foreground. Contemporary psychodynamic psychotherapists endorse that at a minimum the therapist must be able to sustain enough of a relationship with the parent(s) or caretakers to support the treatment of the child. The parents must be engaged sufficiently to bring the child to treatment, pay for the treatment, and withstand their own emotional reactions that may arise as the child changes. Further, a positive alliance with the parents can sustain a child patient through the usual feelings of reluctance to coming to treatment. Ideally the parents can be assisted in fostering sturdier attachment, boundaries, and mentalization. This all requires the psychodynamic child therapist to develop a solid working relationship with the parents while at the same time preserving a sense of confidentiality and trust with the child.

A 10-year-old girl was seen in twice-weekly psychodynamic psychotherapy. As the child left her sessions, her mother each time asked “how is she doing?” The therapist made a separate appointment with the parents to discuss these concerns further, while also exploring how the child felt about this meeting. The girl had concerns about “what they are going to say about me,” but assented with support and understanding from the therapist. After discussion with the parents it was agreed that the therapist would meet with the parents once every 3 weeks to discuss the child’s progress and current functioning, provide support and guidance to the parents, and monitor the parents’ anxiety. Once this plan was initiated, the mother’s anxiety seemed to lessen and she no longer waited expectantly after the sessions with her daughter. The therapist remains attuned to the impact of these meetings on the child patient and discusses with the child as needed.

The Child Patient Is Dependent on the Parents Materially, Physically, and Emotionally

This has profound implications and has led to the conviction, even among the most traditional child psychoanalysts, that therapeutic work with the parents is essential and necessary in working psychodynamically with children. Furthermore the ongoing interactions with caretakers have a continuing effect on the growth and progress of children in therapy. Because a child’s personality is unfolding and shaped strongly by continuing developmental experiences, the ongoing parent–child interactions have a “double-barreled” effect. They not only affect the child’s developing internal world of expectations, assumptions, and beliefs about self and others, but also then in turn impact on the child’s exposure to future developmental experiences. By inhibiting or encouraging exposure to growth-promoting developmental experiences the ongoing parent–child interaction exerts a multiplier effect on the child. This observation strongly influences the overall aim of PDP with children, which, as initially articulated by A. Freud, is to restore the child to the path of normal development. Once back on track, normal development can then take over and provide the experiences needed for internal growth. From a psychological perspective the child’s dependency also impacts the individual treatment directly. Adolescents, especially during the early teen years, are struggling mightily to separate from their parents and establish an independent sense of self (i.e., identity). As a result looking to and depending on another adult, like a therapist, can be fraught with complication and ambivalence. This can interfere with the establishment of a therapeutic relationship even when the adolescent recognizes a need for help.

Children and Adolescents Evoke Stronger Emotional Responses in Psychotherapist’s Attempts to Help Them

As is known by anyone who has spent any time working in child mental health settings children evoke strong feelings in the therapists who work with them. This often results in desires to protect and rescue the child, especially when therapists observe here-and-now pathogenic interactions and traumatic experiences enacted by parents. At other times the child’s behaviors can be exasperating, infuriating, and frightening. All of this results in feelings within the therapist that are generally much higher in intensity than those generated in adult work. This places a premium on the therapists’ recognition and containment of their own emotional responses. While these reactions may be helpful clues in understanding the child’s world, they can make for highly challenging work.

Play Is the Central Medium for Communicating With Younger Children

Since its initial recognition by early psychoanalysts, play has been recognized as the critical in communicating with children in PDP. This is especially true for younger children whose language skills have developed to a limited degree and for whom play is their “natural” language. Older children are increasingly comfortable talking about their own inner states and by 9 or 10 may find playing “baby-ish.” Whether or not to use play materials with an individual child must be determined on a case by case basis. Psychodynamic psychotherapists will encourage play by having a number of toys visible and accessible to the child. This helps to engage the child and conveys that the therapist is willing to meet them on “their turf” and create a safe space for talking about their problems. Much of the “talking” is done “in the displacement” of the play as opposed to the direct discussion that occurs in therapy with older adolescents and adults. Most therapists recommend a limited number of toys that allow for the symbolic expression of a wide array of affects and themes (e.g., doll house, dolls, puppets, animals, blocks). Art materials (drawing, painting, clay) and techniques (e.g., Winnicott’s Squiggle Game), storytelling (e.g., Gardner’s Mutual Storytelling Technique), and role playing are often incorporated into dynamic therapy. Therapists also may utilize structured games (e.g., checkers, chess) or therapeutic games (e.g., Talking, Feeling, Doing Game) in some circumstances.

As Winnicott has pointed out, play exists “in the space between” the inner world and external life, *between* patient and analyst. Play requires active participation from both the patient and the analyst. It is aimless and nonpurposeful; there are no social consequences. As such, play allows for the safe exploration of affects, conflicts, areas of concern, ideals, roles, and wishes. This accords play a self-healing effect in and of itself. Play is a forum for problem solving and experimentation. Play also offers a window into the inner life of the child. This allows for a deepening of the understanding of the child and may be the basis for targeted interventions. As Winnicott has written, when play is not possible then the work of the therapy is to help the child to be able to play.

Transference in Children and Adolescents Is More Complex Than With Adults

Much controversy has surrounded the concept of transference in work with children. Children's internal representational world is still being actively formed and altered by ongoing current interactions with parents and others. As such transference reactions, while organized to a degree, are still in flux. Nevertheless transference reactions occur ubiquitously in PDP as children enact the beliefs, assumptions, and expectations they are developing. These reactions are here-and-now examples of the prototypical interactions that are interfering with their growth and development. Interpretations of transference phenomena primarily address this here-and-now aspect ("perhaps the nervous feeling you have is because of some anger you are feeling") and much less so address genetic roots (i.e., interpretations about the child's reaction being connected to previous experiences with parents are less applicable with younger children). PDP focuses its greatest attention on these transference reactions, seeing this as the most powerful arena to help children gain flexibility and greater freedom in their understanding of themselves and their interactions with others. Psychodynamic work with children is made more complex because the child's transference reactions intersect with the parent's emotional needs and vulnerabilities. This can lead to divided loyalties between the child's desire for growth and the parent's needs. As in all PDPs countertransference reactions require the therapist to discern what belongs to their own psychological dynamics and that which is engendered by the child's behaviors. Countertransference reactions can be invaluable clues to understanding the child's transference reactions and warded off self-experiences.

Barriers to Change Exist Within the Child Patient (i.e., Resistance) and Within the Family System

Resistance is a concept that originates in classical theory to describe the inevitable intrapsychic barriers to growth. Traditionally this is thought to derive from conflicts between the patient's ego, id, and superego. Subsequent theorists have conceptualized this differently. Some have understood these barriers as a striving for growth and attempt to master traumatic experience. Others have seen the terror and dread of changing and giving up familiar ways of seeing oneself and others as central. Still other contemporary theorists account for these impasses as a result of psychic deficits that arise from the absence of needed relational experiences and empathic failures. Resistance in this frame is further impacted by the therapist's ability in the here and now to provide these growth promoting experiences.

While resistance is typically seen as intrapsychic, there are also systemic barriers to change. As the child becomes less symptomatic and grows psychologically the parent's self-concept, parent–child or marital relationship may become challenged, likely due to the child's previous role as the identified patient in the dysfunctional system. This system can continue to create impediments to the child's progress and may require the therapist to attend and respond to the parents' anxiety about change.

A 16-year-old girl was seen in individual psychodynamic psychotherapy for generalized anxiety disorder and recurrent depression. As the teen's symptoms lessened and she developed greater confidence, she began to experiment with desires to move away from the orbit of family expectations and expressed interests outside the parents' expectations. This, in turn, stimulated the mother's criticism about the teen possibly "ending up like her father." This led to increased intra-martial conflict, revealing a longstanding issue in the marriage. The parents were not able to negotiate this in a more satisfactory way, and ultimately, they were referred to marital counseling for themselves. Without this forum for their own growth, the pressure on the child to inhibit their own growth would remain unchecked.

There Is an Increasing Focus on Attachment, Mentalization, and Self-Regulatory Issues versus Oedipal Issues

Classical theorists have highlighted the central importance of oedipal issues in understanding both psychopathology and treatment. Early case descriptions appear to explain clinical phenomena almost entirely through this lens and treatment appears to hinge on adequately addressing through interpretation these oedipal issues consistently; triadic relationship patterns and issues of competition, sharing, sexuality, and rivalry are paramount. In contrast, contemporary theorists, no doubt responding to the major societal changes and vicissitudes of life in the 21st century, have instead placed the primary focus of understanding psychopathology and treatment to early parent–infant interaction, “basic trust,” attachment, separation–individuation concerns, self-reflective process (mentalization) and self-regulation. While not disavowing a place for triadic issues this has prompted a major shift in emphasis to dyadic issues.

A 12-year-old boy is brought for treatment because of multiple fears, languishing academic performance and ambivalent attendance to household responsibilities. Toward the end of the psychotherapy, after the child had made much progress, his father had a relapse of substance addition. While the original symptoms could be understood as a reflection of oedipal issues (e.g., castration fear and guilt over hostile feelings toward a parent, leading to academic inhibition and fearfulness), contemporary theorists would emphasize the boy’s real anxiety over the safety of his father, and the potential for guilt for hostile feelings toward a loved one with a chronic illness.

Interpretation Was Emphasized as the Central Determinant of Therapeutic Action in Classical Theory, But Contemporary Views Place More Emphasis on Attachment Perspectives

While interpretation was seen in classical viewpoints as the central determinant of how psychodynamic psychotherapy works ("therapeutic action"), alternative explanations are now in ascendancy, namely the primary importance of a consistent and caring attachment figure. Classical analysts have emphasized the central role of interpretation and "making the unconscious conscious" in the treatment of children. Interpretations proceed from that closest to consciousness to those mental contents furthest away (i.e., unconscious). This is especially important with children. These interpretations may be directed toward defenses, wishes or intentions, or transference phenomena. Interpretations are traditionally made in a specified order, from the manifest level to what is unconscious: first affect or behavior, then defense, then impulse. In other words the analyst first would interpret the affect or behavior ("You appear to be feeling angry right now"), followed by the defense ("perhaps you are angry because you were disappointed that I did not remember your birthday"), and last the underlying impulse ("perhaps you were really hoping that someone would remember your birthday"). Timing of interpretations must be aligned with the patient's readiness and the state of the therapeutic alliance. Theorists differ in what they see as the primary impulses or fundamental motivators. Classically oriented analysts follow Freud's original drive theories and see sexuality and aggression as primary; contemporary analysts might emphasize self-cohesion, self-regulation, attachment, mastery and assertion as the most powerful organizers of motivation. Contemporary theorists hence would see therapeutic action as stemming from the provision of needed relationship experiences that allow for the development of a cohesive and vital self. Fonagy, Bateman, and others have attempted to bridge these views, attachment theory, and cognitive neuroscience by highlighting the goal of *mentalization.* Mentalization, or the self-reflective capacity, enables the child to perceive and process mental activity in the self and others. This allows the child to develop a model of the inner thoughts and feelings that underlie human behavior. Optimally this capacity develops out of the child's early attachment with parents who relate to the child as an intentional being with underlying thoughts, feelings, and intentions. As a result of these experiences the child is then able to understand other's behaviors and inner states. These theorists posit that this is the centerpiece of psychoanalytic work with children (and adults) and success depends on the degree to which this develops.

PRACTICAL MATTERS

Psychodynamic Psychotherapy With Children and Adolescents Can Take Differing Forms and the Principles Can Be Applied in Work With Individuals, Groups, and Families

Classically, child psychoanalysis involved seeing the child four or more times per week, lasting over a period of years. In this form psychoanalysis is rarely used currently, although there is some research to suggest it may have a place with particularly disturbed children (see Research section). PDP occurring once, or perhaps twice, per week over a period of months to years, is much more common in both private practice and public clinic settings, although it is increasingly difficult because of payment issues and the hectic pace of family life in the United States. The nationalized health systems in the United Kingdom and Canada continue to support such treatments widely.

While typically open-ended in length, there have been recent adaptations of brief, time-limited PDP to work with children. For example, MBT has been applied to work with adolescents, children, and families. Psychoanalytic principles have also been widely applied to adolescent and family group therapies.

As noted previously, psychoanalytically informed principles often are integrated into all treatment efforts with the child. Owen Lewis has delineated such a model of integrated treatment in which psychodynamic principles fundamentally inform the gamut of treatment interventions (e.g., behavior therapy, psychopharmacology, family therapy, and so forth). The more recent review by Malberg and Mayes (The Contemporary Psychodynamic Developmental Perspective) elaborates on this evolution.

Perhaps Even More Than in Work With Adults, the Therapeutic Frame and Setting Are Critical in Working With Children and Adolescents

Child therapy requires a dedicated office that can serve as a concrete reminder of safety, security, and predictability for the child. Using the same office for every session fosters these goals. Some analysts create special cubby holes or boxes for children to keep their particular toys, drawings, and so forth. This also serves to personalize the space for the child. The office should be neither so small that the child feels closed in, nor so large that the child is overwhelmed and overstimulated. An explicit therapeutic contract should be discussed with the child and parents. Children should not be forced to attend therapy although reluctance is not uncommon. Assent (vs. consent) is a more appropriate goal. It is generally recommended that therapy occur at the same time, on the same day of the week. This also strengthens the predictable frame necessary for successful work. All these ingredients contribute to establishing what Winnicott referred to as the “holding environment.”

Confidentiality Is a Particularly Complex Issue With Children and Adolescents

On the one hand, there is a clear need to maintain a sense of safety and security in order for the child to reveal their deepest worries, hopes, thoughts, and desires; on the other hand, therapists may perceive parents or guardians could be instrumental in making effective change if they were aware of such deep worries, thoughts and desires that a child or adolescent may forbid the therapist from sharing. In more serious situations, therapists have a legal and ethical obligation to safeguard the child’s welfare and may be required to break confidentiality. When issues of safety (e.g., suicidality) arise parents must be informed. At other times the involvement of child protective authorities may be needed, which in some situations may end the treatment. In any event the limits and extent of confidentiality should be carefully explained to both the child and parents at the beginning of treatment.

Working With Parents Is Essential in Psychotherapy With Children and Adolescents

As already noted parents are decisive players in PDP with children. Although there are times that the parents need to honor the necessary boundaries of treatment with the child, it is seldom appropriate that the parents are excluded from the treatment. Typically parents are included from the beginning to obtain the history and to begin to establish a working relationship that can sustain the treatment of the child. At the conclusion of the evaluation period it is helpful to explicitly review the assessment and establish a treatment plan with the parents. The analyst must reach concordance with the parents on the nature of this treatment contract from the beginning and may need to re-visit this throughout the course of treatment. Regular meetings with the parents (with or without the child) are routine parts of PDP. These ongoing meetings serve to maintain the working relationship, provide further information about the child, and create a forum for parent guidance or exploration of parental psychological issues that impede the child’s growth. When it becomes apparent that the parent(s) are in need of their own treatment then referral is generally made to another therapist. Providing individual PDP to two members of the same family contemporaneously is discouraged due to the inherent conflict of interest and encroachment on confidentiality that may undercut the alliance with one or both persons.

CLINICAL CONSIDERATIONS

The Aims of Psychoanalytic Psychotherapy Are Ambitious

Psychoanalytic psychotherapy aims to go well beyond the relief of symptoms and/or greater coping or adaptation. With adults, the goals have been to seek a “cure,” broadly improving the quality of life (i.e., “to work and to love”), and inoculating the patient from future stressors. In contrast, A. Freud first articulated the aims of child psychoanalysis as a return to the path of normal development.

Subsequent analysts have emphasized the aims of increased mentalization, self-understanding, self-regulation, personality integration, self-expressiveness, self-esteem, flexibility, interdependence, capacity for intimacy, increased frustration tolerance, and capacity to play. It is assumed that reaching these goals in a “good enough” fashion will result in a diminution of symptoms as well as improved functioning in school, in the community, and with peers and family members.

The Indications for Psychodynamic Psychotherapy Most Likely Include Children and Adolescents With Internalizing Disorders

Indications for PDP have historically focused on the presence of certain strengths or loose categories of problems generally seen as due to intrapsychic conflict. Typically, indications have included those children with neurotic level disorders and sufficiently good internal and external resources such that the gains will be following sustainable termination of treatment. This situation rarely exists in clinical practice. Further, research has indicated that this group of children may benefit as well or better from less intensive treatment efforts involving the school, family, or community resources. At the same time, children with more severe disturbances often subsumed under the category of borderline conditions, may require an intensive effort such as psychoanalysis for effective results to be achieved (see the section on Research and Evaluation). Another way to look at indications is to consider the goals of the treatment. For example, if the goal is to understand one’s feelings, thoughts, and the connections between this inner world and the outer world of behavior (mentalization or self-reflective function), then PDP is likely to be helpful. Because PDP is now generally integrated with other treatment approaches, it is not necessary for this treatment to be all-encompassing but rather to aid in forwarding critical developmental processes.

Some Children With Externalizing Disorders May Be Helped With Family-Focused Psychodynamically Informed Therapies

Regulation-focused Psychotherapy for Children (RFP-C) and Parent–Child Interaction Therapy targeting Emotion Development (PCIT-ED) are two examples of recent manualized psychotherapies targeting the emotional regulation of young children. This is an area of increasing research (see the section on Research and Evaluation), with the best results seen with modalities that include a heavy component of parent/guardian involvement. Historically, externalizing disorders have been considered a contraindication of individual PDP; the key elemental difference perhaps being the recruitment of parents/guardians to assist children in regulating their intense emotional outbursts.

There Are Some Children for Whom Individual Psychodynamic Psychotherapy Is Not Likely to Help or Is Contraindicated

Contraindications may include some children with externalizing problems, particularly when there are “callous-unemotional” features or when there is no comorbid internalizing disorder or an inability to restrain impulses consistently enough to play. Another group of children for whom there is little to justify this treatment would be children with more than mild intellectual disability, classical autism, many autism spectrum disorders, and traumatic brain injury (and other neurologic conditions) when receptive language abilities are substantially compromised. These may be relative contraindications as there may be focal psychodynamic goals even in some of these cases. Again, treatment for many of these individuals will be multimodal, and other treatment modalities may be more critical (i.e., school interventions for children with autism-spectrum disorders).

AN OVERVIEW OF THE PHASES OF TREATMENT

Beginning Treatment: Focusing on the Establishment of a Working Alliance

This requires not only creating a therapeutic ambiance for the child but also establishing a supportive, working relationship with the parents. The initial goal with the child is to create a playful, confidential environment in which they and the therapist can explore the world of the child's thoughts, feelings, and behaviors. In the beginning, some exploration and understanding of why the child is meeting with the therapist is useful, as well as an explicit clarification of the treatment frame and contract ("we will be meeting for 45 minutes every Tuesday at 4:00. This time will just be for us. We'll try really hard to meet every week"). It is also helpful for the therapist to explain to the child that they are "someone who meets with kids to help them with their feelings," "to see if we can understand you," "to help you feel better," "to help you understand how you feel and why you do some of the things that cause you trouble," and so forth. The gist of these comments is to explain the role of the therapist and to convey a steadfastly hopeful attitude to the child (and parents) through what will be inevitable ups and downs. For the playroom to acquire a "specialness" to the child, these messages must be communicated repeatedly. The ground rules also should be explained ("we will talk and play together during the time we are together. Anything is okay as long as no one gets hurt and nothing gets broken."). Of course, once the child begins to express themselves they will see what the therapist is "made of." This is the true acid test for the child. Only over time can the child get the sense that the therapy is not a punishment, that the therapist has hope for the child, the child can feel understood, and problems can be solved. For the most part, in the initial phase, the therapist keeps their comments and interpretations focused on the displacement of the play rather than making direct comments about the child's real-life situation. This initial launching typically takes a few weeks (in the healthiest children) to a number of months to achieve a secure working relationship.

Mid-Phase Treatment

This begins when the child has established a clear working alliance. The child has a familiarity with the frame of treatment, even though emotional upheavals may continue in the sessions. During this phase, the primary emphasis is on the "work" of therapy, that is, the slow and repetitive recognition of feelings, defenses, and their relationship to behaviors. Over this time, the same theme in a play session may occur for months at a time. The therapist continues to create with the child a sense of hope and comfort as increasing recognition and understanding of the child's inner world accrues. It is in this phase that the interpretation of prototypic problematic patterns and transference reactions occurs. The focus on the here-and-now reactions of the child with the therapist is especially fruitful, although much work continues to be done in the play displacement. Gradually in this phase, the child becomes more able to connect this displacement with their own life. Reconstructive statements linking current emotional or behavioral patterns with previous life experiences are generally not emphasized in work with children.

Ending Treatment

Optimally treatment ends when the child is free from symptoms, has made sufficient progress in psychological growth and maturity, and is back on track developmentally. In the real world, termination generally comes about for other reasons. Frequently termination occurs when partial improvement has been made and practical considerations (insurance, finances, transportation, moves, graduation) become paramount. Termination provides an opportunity for the child to focus on issues of separation, loss, and mourning. Paulina Kernberg has elegantly stated:

> "As the child approaches the end of her treatment, the same issues that brought her to therapy once again come to the fore: (1) Do they (the parents) care? (2) Does the therapist care? (3) Do I care for you (the therapist)? (4) Do I care for them (the parents)? (5) Can I care for myself? (6) Will you remember me? (7) Will I remember you?" (From Kernberg PF. *Children With Conduct Disorders*. copyright © 1991. Reprinted by permission of Basic Books, an imprint of Hachette Book Group, Inc.)

In the process of ending, it is important to include the child actively in setting a final date. In general, 6 weeks to 3 months is sufficient time to allow the child and family to concentrate meaningfully on these issues. Setting the date so that it does not coincide with other real-life endings (e.g., graduation, moving dates) permits the child and family to experience the fullest impact of the termination. During the ending phase, children frequently regress, with a recurrence of symptoms and defenses in the face of this stressor. Denial of the importance of the relationship by the child, parent, and therapist is not uncommon and should be understood as a reaction to the impending separation and loss. In the course of ending, the child should have an opportunity to explore and claim the gamut of feelings that inevitably arise: anxiety, guilt, anger, sadness, satisfaction, ambivalence, and so forth. This frees the child from developing defensive patterns to face the inevitable future losses of life.

After 8 months of weekly psychotherapy, an 11-year-old girl's symptoms of perfectionism and severe recurrent abdominal pain had improved sufficiently that the child came in and stated that "I don't have pain! I don't need to come here anymore." She later added that "maybe I need help with boys." This led to a series of discussions about termination, which resulted in the setting of a final date 6 weeks hence. In the first meeting after the termination date was set, she revealed that she was able to comfortably talk with the therapist, unlike with her parents, with whom she felt she could not be open. The next week she became irritated by a comment from the therapist and said, "If I had known you would say that I would have left weeks ago!" She continued to talk about other separations in her life, especially the death of her grandmother. In the final sessions, she alternated between denial and recognition of her emotional reactions to the upcoming termination. She continued to discuss other endings, moves, and friends leaving. In the last session, she said: "Okay. I have something to say. I think I'll miss you. You helped me." To which the therapist responded: "Well, we have worked well together. And I will miss you too." She left the session knowing that she could make future appointments if the need arose.

RESEARCH AND EVALUATION

Claims of lacking an evidence base have been haunting PDP for decades. Because of psychoanalysis' emphasis on the fundamental uniqueness of each individual, single case studies were historically the preferred form of publishable research. These were typically reported as detailed narratives without the systematic or objectified approach typically taken in reports of behavioral or psychopharmacologic treatments. With the ascendancy of randomized controlled trials (RCTs) and meta-analyses, this traditional psychoanalytic approach has been widely seen as lacking.

While there remains persistent criticism within psychoanalytic circles of the overvaluation of the RCT, there has been acceptance of the need for RCTs and meta-analyses involving PDP. RCTs are seen as necessary, but not sufficient, to establish a treatment as "evidence-based." Patient values and preferences, outcome measures broader than symptom change, long-term outcomes, and clinician's observations also are held as important factors as well. Its typical length, ambitious aims, the complexity of operationalizing goals, and the difficulty in manualizing the approach have been serious barriers to applying the RCT model to psychoanalytic psychotherapy.

Despite these concerns, there have been increasing efforts and positive results in evaluating PDP for children and adolescents, as well as adults. A more recent problem is that the evolution of psychotherapy has occurred in such a way that psychodynamic principles are increasingly integrated into more focal or modular treatments. For example, IPT is clearly rooted in psychodynamic principles and has been adapted to adolescents. Many other branded therapies, for example, Parent–Child Interaction Therapy, Mentalization-Based Treatment, Adolescent-Focused Therapy, Attachment-Based Family Therapy, and Regulation-Focused Psychotherapy, among others, have strong psychodynamic foundations. Further, the emphasis on the therapeutic alliance that began with psychoanalysis is now accepted as a key component for virtually all psychotherapy. Many of these treatment approaches have been investigated empirically, although they are not generally included in reviews of the traditional research base for PDP. Much of the literature on PDP has been conducted on small numbers of real-world patients with multiple diagnoses and with broad definitions of PDP. While the largest studies have been observational, there have been a number of RCTs sufficient to generate systematic reviews and meta-analyses. The largest empirical study of psychoanalytic therapy with children has been carried out at the Anna Freud Centre (formerly known as the Hampstead Clinic) by Fonagy, Target, Midgely, and others there. Following extensive searches and rigorous inclusion criteria, they identified over 30 primary studies that involved over 1,200 patients. This represents a significant increase in studies from 10 years ago but still must be considered limited, especially when compared to studies of other psychiatric treatments. Building on previous RCTs, meta-analyses, and systematic reviews, Midgely et al. in 2021 published their seminal “narrative synthesis” of the evidence base for PDP for children and adolescents, including over 120 papers representing over 80 separate distinct studies. The general conclusions are:

1. There is increasing evidence to support the effectiveness of psychoanalytic psychotherapy of children and adolescents, with both the quality and quantity of empirical papers in this field increasing over time.
2. Children with internalizing disorders, primarily depression and emotional disorders, tend to respond most favorably to this treatment modality. The treatment of these disorders has the most high-quality evidence and suggests that brief PDPs are just as effective as other psychosocial interventions, including cognitive behavioral therapy.
3. A number of studies support positive responses to PDP for eating disorders and the effects of trauma in children.
4. There are small but increasing numbers of RCTs focusing on anxiety disorders, and there already exist treatment manuals for PDP for treating children and adolescents with anxiety disorders.
5. There are small but increasing numbers of studies on emerging personality disorders, with several observational studies of the psychodynamic treatment of BPD published in the last 3 years. Mentalization-Based Treatment and Transference-Focused Psychotherapy, both of which were initially developed to treat BPD in adults, have been adapted for adolescents. MBT has shown promise for self-harm as an alternative to DBT.
6. Younger children show larger effect sizes than older children and adolescents, but a sufficient frequency of sessions must be achieved to produce positive outcomes.
7. The psychodynamic treatment of externalizing disorders has received less research attention and, in general, is seen as less effective. Despite this accepted wisdom, some studies show

promising findings, particularly when the child also presents with some emotional difficulties. Young children with disruptive disorders are more likely to respond to interventions that lean heavily on involvement from parents and guardians if they are to respond.

8. There is low quantity but high-quality evidence that PDP may be helpful for adolescents with certain medical or physical conditions, such as diabetes, headache, and treatment adherence.
9. Children and adolescents with conduct disorders or psychotic disorders, receptive language disorders, and intellectual disabilities are also unlikely to respond to this form of treatment. Children with autism generally do not respond well to individual psychotherapy, although there is some evidence that there may be a place in the broader autism-spectrum group for approaches influenced by psychodynamic principles.

A range of methodologic limitations continues to abound, including small sample sizes, lack of or inappropriate control groups, and inadequate outcome measures. More high-quality research is still needed.

Following these findings, recent national guidelines have included PDP as recommended treatment for the two conditions with the most substantial evidence base; the UK NICE guidelines now include PDP as an evidence-based treatment for depression in children and adolescents; in 2015, the AACAP Practice Parameter for Eating Disorders included Adolescent-focused therapy as a recommended treatment in cases of anorexia nervosa when family-based treatment was not available or acceptable to the patient and family. While the empirical data supporting psychoanalytic psychotherapy with children is growing, larger-scale studies for additional targeted populations continue to be needed. At the same time, psychodynamic approaches will continue to be incorporated into other forms of therapy as the field continues to strive to find “what works for whom.”

FUTURE DIRECTIONS

Psychoanalytically informed psychotherapy has a long tradition in child and adolescent psychiatry. Many practitioners over many decades have found this orientation extremely helpful for large numbers of child and adolescent patients. There now exists better empirical support for PDP for children and adolescents with certain conditions, including depression, trauma, eating disorders, anxiety, and personality disorders. More methodologically sound studies are needed and underway. There is still a need to understand those patients or clinical problems for which this technique is especially or uniquely relevant, its mechanism of action, and critical areas of technique. An increasing number of manualized psychotherapeutic modalities have incorporated elements of psychodynamic thinking and practice. This trend is expected to become more prevalent as the field moves to more targeted, individualized therapies.

However, as has been highlighted by Shedler and others, there is a uniqueness to psychodynamics that ultimately may not be able to be isolated in traditional research studies. The supposition that the core elements of psychodynamics can be isolated, pre-packaged, and scripted for randomized control trials so that any therapist picking up a manual will be able to fit that manual to an individual patient will be extremely challenging to test. Evidence-based medicine is inherently designed to eliminate confounding variables and things left to chance, but in doing so, it may also eliminate the individual differences and variations that are co-created by individual patients and therapists. As Shedler has said, it is perhaps these differences and variations, the “unique rhythms,” that are both eliminated in many randomized studies and yet also compose “nearly everything that matters.”

FURTHER READINGS

*Bateman A, Fonagy P. Mentalization based treatment. *Psychoanal Inq*. 2013;33:595–613.

*Bromfield R. *Doing Child and Adolescent Therapy: Adapting Psychodynamic Treatment to Contemporary Practice*. 2nd ed. John Wiley and Sons; 2007.

*Delgado S, Strawn J, Pedapati E. *Psychodynamic Psychotherapy with Children and Adolescents: Integrating Intersubjectivity and Neuroscience*. Springer; 2015.

Di Giuseppe M, Prout TA, Rice T, Hoffman L. Regulation-focused psychotherapy for children (RFP-C): advances in the treatment of ADHD and ODD in childhood and adolescence. *Front Psychol*. 2020;11:572917.

Eisenberg L. The past 50 years of child and adolescent psychiatry: a personal memoir. *J Am Acad Child Adolesc Psychiatry*. 2001;40(7):744.

Fonagy P, Cottrell D, Phillips J, et al. *What Works for Whom? A Critical Review of Treatments for Children and Adolescents*. 2nd ed. Guilford Press; 2014.

Freud A. *The Psychoanalytic Treatment of Children*. Imago Publishing; 1946.

Gatta M, Miscioscia M, Svanellini L, et al. Effectiveness of brief psychodynamic therapy with children and adolescents: an outcome study. *Front Pediatr*. 2019;7:501.

Goodyer IM, Reynolds S, Barrett B, et al. Cognitive behavioural therapy and short-term psychoanalytical psychotherapy versus a brief psychosocial intervention in adolescents with unipolar major depressive disorder (IMPACT): a multicentre, pragmatic, observer-blind, randomised controlled superiority trial. [published correction appears in *Lancet Psychiatry*. 2017;4(8):582]. *Lancet Psychiatry*. 2017;4(2):109–119.

Greenberg JR, Mitchell SA. *Object Relations in Psychoanalytic Theory*. Harvard University Press; 1983.

Hug-Hellmuth H. On the technique of child analysis. *Int J Psychoanal*. 1921;2:287–305.

Kernberg PF, Chazan SE. *Children with Conduct Disorders: A Psychotherapy Manual*. Basic Books; 1991.

*Kernberg P, Ritvo R, Keable H; AACAP Committee on Quality Issues. Practice parameter of psychodynamic psychotherapy with children. *J Am Acad Child Adolesc Psychiatry*. 2012;51(5):541–557.

Klein M. *The Psychoanalysis of Children*. Hogarth Press; 1932.

Leichsenring F, Luyten P, Hilsenroth MJ, et al. Psychodynamic therapy meets evidence-based medicine: a systematic review using updated criteria. *Lancet Psychiatry*. 2015;2:648–660.

Lewis O. Integrated psychodynamic psychotherapy with children. *Child Adolesc Psychiatr Clin North Am*. 1997;6(1):53.

Lichtenberg JD, Lachmann FM, Fosshage JL. *Self and Motivational Systems: Toward a Theory of Psychoanalytic Technique*. Analytic Press; 1992.

Lock J, La Via MC; American Academy of Child and Adolescent Psychiatry (AACAP) Committee on Quality Issues (CQI). Practice parameter for the assessment and treatment of children and adolescents with eating disorders. *J Am Acad Child Adol Psychiatry*. 2015;54(5):412–425.

Luby JL, Barch DM, Whalen D, Tillman R, Freedland KE. A randomized controlled trial of parent-child psychotherapy targeting emotion development for early childhood depression. *Am J Psychiatry*. 2018;175(11):1102–1110.

*Meeks J, Bernet W. *The Fragile Alliance*. 5th ed. Krieger Publishing; 2001.
Méndez J, Sánchez-Hernández Ó, Garber J, Espada JP, Orgilés M. Psychological treatments for depression in adolescents: more than three decades later. *Int J Environ Res Public Health*. 2021;18(9):4600.
Midgley N, Mortimer R, Cirasola A, Batra P, Kennedy E. The evidence-base for psychodynamic psychotherapy with children and adolescents: a narrative synthesis. *Front Psychol*. 2021;12:662671.
Normandin L, Ensink K, Weiner A, Kernberg OF. *Transference-Focused Psychotherapy for Adolescents With Severe Personality Disorders*. American Psychiatric Assoc Pub; 2021.
Preter SE, Shapiro TD, Milrod B. *Child and Adolescent Anxiety Psychodynamic Psychotherapy: A Treatment Manual*. Oxford University Press; 2018.
Sandler J, Kennedy H, Tyson R. *The Technique of Child Analysis: Discussions With Anna Freud*. Hogarth Press; 1980.
Shedler J. The efficacy of psychodynamic psychotherapy. *Am Psychol*. 2010;65(2):98–109.
Shedler J. What is the evidence base for “evidence-based” therapy? *Psychiatr Clin N Am*. 2018;41:319–329.
Siegel D. *The Developing Mind: How Relationships and the Brain Interact to Shape Who We Are*. 3rd ed. Guilford Press; 2020.
Volkert J, Taubner S, Byrne G, Rossouw T, Midgley N. Introduction to mentalization-based approaches for parents, children, youths, and families. *Am J Psychother*. 2022;75(1):4-11.
Weiss S. Vicissitudes of termination: transferences and countertransferences. In: Schmukler AG, ed. *Saying Goodbye: A Casebook of Termination in Child and Adolescent Analysis and Therapy*. The Analytic Press; 1991:265.
*Winnicott DW. *Playing and Reality*. Tavistock Publications; 1971:54.

53.2 Cognitive-Behavioral Psychotherapy for Children and Adolescents

ANNE MARIE ALBANO, PH.D., AND SANDRA S. PIMENTEL, PH.D.

The terms *cognitive-behavioral therapy, cognitive therapy,* and *behavior therapy* are often used interchangeably. *Cognitive-*

behavioral therapy (CBT) is a form of psychotherapy that combines behavior therapy and cognitive therapy into a broad approach composed of empirically based methods built on the fundamentals of social learning theory and experimental psychology. It is the treatment of choice for various mental health problems due to its efficacy. CBT is widely taught in various mental health training programs and is a mandated component of psychiatric residency training in the United States. CBT is a mandated first-line treatment in the United Kingdom for specific mental health problems such as anxiety, eating disorders, and depression.

HISTORY

CBT is best viewed as a multifaceted theoretical approach with a strong allegiance to the experimental method and a resultant set of techniques applied within a CBT formulation. The cognitive-behavioral approach is the convergence of distinct movements in the study and science of human behavior, including classical conditioning, operant conditioning, and the cognitive revolution. In the early 1920s, Ivan Pavlov's work demonstrated the processes of classical (otherwise known as respondent) conditioning and established the stimulus-response learning paradigm. Essentially, classical conditioning is a form of learning in which a neutral stimulus is repeatedly paired with a stimulus that elicits a reflexive response. In his classic experiment, Pavlov paired the neutral stimulus of a bell with food, a stimulus that naturally produces salivation. Eventually, the presentation of the neutral stimulus (bell) by itself elicited the conditioned response of salivation. This process is illustrated in Table 53.2–1. This work greatly influenced John B. Watson, who applied Pavlovian classical conditioning principles to understanding the origins of psychological disorders in humans. Known as the "father of behaviorism," Watson promoted the tenet that psychology should be concerned only with directly observable stimuli and responses and rejected any notion of unconscious and dynamic processes (i.e., mentalistic constructs) underlying behavior, including cognitive processes. Early laboratory research with animals demonstrated the acquisition of the fear response through classical conditioning learning trials.

Watson and colleagues demonstrated that fear reactions are also conditioned in humans who previously did not evidence such a reaction. In their classic experiment, Watson and Rosalie Rayner conditioned a fear reaction to a white rat in a young child named Albert. Initially, Albert played happily with the rat and showed no evidence of fear or avoidance of this animal. Next, the experimenters followed the introduction of the rat with a loud noise made in close proximity to the child, causing him to startle. The pairing of the loud noise (aversive stimulus) and the rat (neutral stimulus) in a sequence of trials resulted in Albert developing fear of the rat (conditioned response). Subsequently, not only did Albert become fearful of the presence of the rat, eventually his fear generalized to other white, furry stimuli, such as cotton balls. Inasmuch as this experiment demonstrated the process of Pavlovian learning principles in the etiology of psychological disorders, Mary Cover Jones demonstrated that one could eliminate human fear conditions in children through a similar learning paradigm that would foreshadow later methods of behavioral treatments for anxiety disorders. With Watson as her dissertation mentor, Jones used an approach involving the techniques of modeling and graduated exposure to treat 3-year-old Peter’s fear of rabbits. Initially, Peter was exposed to observing other children playing happily with a rabbit (i.e., modeling), and, subsequently, while Peter was eating his favorite food, the rabbit was shown to him during a time when he was in a calm state. Gradually, the rabbit was moved closer to Peter until he was able to pet and play with it without experiencing fear (exposure). Although many will recognize Jones’ work as straightforward examples of the basic behavioral procedures of modeling and graduated exposure, her approach to treating Peter was revolutionary at the time. In fact, it was decades before the techniques employed in Peter’s treatment were systematically studied and identified as valuable therapeutic procedures.

Table 53.2–1.
The Process of Classical Conditioning

Before conditioning		
Food (Unconditioned stimulus)	→	Salivation (Unconditioned/reflexive response)
During conditioning		
Bell + Food (Neutral stimulus) + (Unconditioned stimulus)	→	Salivation (Unconditioned response)
After conditioning		
Bell (Conditioned stimulus)	→	Salivation (Conditioned response)

While Pavlov and Watson were experimenting with classical conditioning paradigms, Edward Thorndike of Columbia University was beginning pioneering work in understanding the processes of operant (instrumental) conditioning. Thorndike demonstrated that by changing the consequences of behavior, the occurrence of the behavior is either increased or decreased—that is, providing a reward after the occurrence of a certain behavior increases the probability of that behavior occurring again, whereas withholding the reward, over time, results in the decrease of the behavior. These principles, influencing the later work of Harvard psychologist B.F. Skinner and his students, are widely used today in therapy and educational settings, such as in parent education and training. A common example is that of a young child who throws a temper tantrum in response to being denied a treat. If the parent provides attention to the child through admonishing, coaxing, or giving in and/or providing the treat, the tantrum behavior is likely to occur again at the next instance of the child's being initially denied the treat. Note that the reward in this case is parental attention, regardless of whether it is the positive attention of coaxing or giving in or the negative attention of admonishment. In any case, the child gets the parent's attention. However, if the parent refrains from giving attention and ignores the child's crying and whining, the tantrum will stop and subsequent tantrums are unlikely to occur when the child is denied or frustrated. The child will learn that tantrum behavior does not result in positive outcomes. Skinner articulated fully the theory and principles underlying the process of reinforcement in his seminal tome "Contingencies of Reinforcement: A Theoretical Analysis." This type of learning is central to parent management training (PMT), an empirically based treatment approach, and similar parent training programs. Change in child behavior (decrease in tantrums) is facilitated and enhanced through the simultaneous pairing of positive child behavior with parental attention, so parents are also taught to give attention and praise when their child behaves in acceptable and desired ways (e.g., "I like the way you are playing nicely with your sister"). The procedures of systematically ignoring undesirable behavior while systematically reinforcing desired behavior are known together as *differential reinforcement of other behavior.* Throughout the 1950s and 1960s the principles of operant conditioning were greatly expanded and researched by Skinner and other experimental behaviorists; eventually, this work moved from the laboratory analog studies of learning processes with animals to applied evaluations of the utility of operant procedures in the treatment of adult psychiatric inpatients. For example, Ted Ayllon and Nate Azrin introduced the first comprehensive token economy program at Anna State Hospital in Illinois. Despite initial resistance and skepticism from the hospital staff, the token economy resulted in significant changes in targeted behavior of inpatient adults with severe mental illnesses and disabilities. Token economy systems flourished in psychiatric settings and training schools and were adapted for widespread use in programs serving youth, such as hospitals, residential facilities, and schools.

The emergence of behavior therapy as a distinct field of study began in the 1950s, in part due to advances in the understanding of both classical and operant learning processes, and also because of a growing discontent with traditional psychoanalysis. A major goal of psychoanalytic treatment is for the patient to develop insight into their problems. *Insight* refers to the emotional and intellectual understanding and acceptance of one’s problems and is a crucial step in achieving therapeutic benefit, and gaining insight is associated with uncovering the unconscious conflicts and motivations underlying the problematic behavior. The assumption of unconscious conflicts is a major premise of psychoanalysis that is rejected by the behavioral approach. The focus on achieving insight, along with the often inordinate length of time spent in psychoanalytic therapy, contributed to discontent with psychoanalysis. South African psychiatrist Joseph Wolpe, dissatisfied with his patients’ progress in psychoanalysis, was among the founders of contemporary behavior therapy and moved its applications from experimental settings and institutions to more mainstream psychotherapy contexts. In the 1950s, Wolpe developed procedures of systematic desensitization and assertive training, directly stemming from Cover Jones’ work with children in the 1920s. In each procedure, phobic and anxious patients are taught to replace fear reactions with deep relaxation and proactive communication behaviors, respectively. These procedures provided a more rapid therapeutic response than traditional analysis and proved to be more cost-and time-effective to patients. Very quickly, Wolpe’s procedures were extended to individual and group therapy contexts and adapted for use in children and adults. Worldwide, behavior therapy was developed, advanced, and promoted by pioneers such as Wolpe, Cyril Franks, Hans Eysenck, Stanley Rachman, and Arnold Lazarus. As a field, behavior therapy took shape through the systematic study of behavioral principles and the testing of its methods in single-case and large research trials. In 1966, Franks became the first president of the newly formed Association for Advancement of Behavior Therapy (now the Association for Behavioral and Cognitive Therapies), the flagship professional organization whose missions remains that of promoting the development and dissemination of behavior and CBT therapy.

Radical behaviorists, steeped in the tradition of Watson, eschewed the study of cognition, in that thought processes were viewed as mentalistic constructs not readily observable or verifiable by others. From a purely behavioral perspective, directly addressing inappropriate or abnormal behavior will change any upsetting or inappropriate thoughts and feelings associated with the behavior, and, hence, direct access to thoughts is not necessary. During the 1960s however, the cognitive revolution began with the work of Albert Bandura, Aaron T. Beck, and Albert Ellis. The cognitive therapist takes a different view, espousing that, if one changes dysfunctional thoughts, change in functional behaviors will follow. The cognitive therapist assumes that thoughts are made observable through methods such as direct interviewing, self-report scales, journal or diary entries, and listing of thoughts. At Stanford University, Bandura integrated the principles of classical and operant conditioning with observational learning to formulate the tenets of social learning theory. Through observational learning, the process of observing the actions and reactions of others can effect a change in the behavior of an individual. In essence, observing the actions of others and their resulting consequences, without the individual needing direct experience with a situation, can shape the emotional and behavior repertoire of an individual. Of course, risk and protective variables interact with observational learning to shape behavior such that the temperamentally inhibited child may learn to respond with anxiety and withdrawal when watching the distress of a peer who is teased by the class bully, whereas the child with a more rambunctious disposition may focus on the behavior of the bully and learn to respond in similarly aggressive ways. This theory places critical importance on the role of cognitions in the development and maintenance of behavior, both normal and abnormal, in that cognitions involve thoughts, images, beliefs, expectations, memories, and attributions, among other simple and complex cognitive processes. In the bully example, the inhibited child may think, “oh no, that can happen to me and embarrass me.” This type of thinking leads to anxious withdrawal and fear reactions. Conversely, the more extroverted child may think, “It is fun making this kid cry and I want to do it next,” leading to imitation of the aggressive behavior. The response of the environment to either child’s response then becomes involved in the complexities of this learning paradigm to further influence the observational learning by either rewarding/accepting or punishing/extinguishing the particular response. The process of observational learning is the basis for concern expressed by parents and other concerned parties over violence depicted in television, movies, social media and video games and one of the reasons for parental ratings found in the media and music industry.

The cognitive revolution continued with the work of psychiatrist Aaron T. Beck at the University of Pennsylvania, who observed within the course of his psychoanalytic treatment of patients that certain problems were evident in the thoughts of depressed individuals and that traditional psychoanalysis did not allow for the direct confrontation of these thoughts. Beck's study of depressed patients revealed a certain negative triad of thoughts related to pessimistic views of the self, the world, and the future. Through the Socratic method of questioning the veracity of these thoughts, Beck was able to assist patients with gathering evidence to refute the thought and engage emotional and behavioral resources to counter the depression and increase healthy functioning. Cognitive therapy emerged from this work and became a mainstay in the treatment of a wide range of emotional disorders, including depression, anxiety, personality disorders, and, more recently, psychoses. Concurrent with the development of Beck's approach, psychologist Albert Ellis took note of the irrational beliefs guiding his patients and undermining their ability to proactively manage activities and stressors in everyday life. Ellis developed rational emotive therapy, which seeks to challenge certain core irrational beliefs such as "I must be loved by everyone" and "I must be perfect in my performance/relationships." According to Ellis, irrational beliefs are conditioned cognitive responses that serve to render an individual vulnerable to disappointment and failure and engender a sense of helplessness and eventual psychopathology. Both Beck's and Ellis' therapies integrate behavior principles and processes, such as behavioral exposure and behavior activation, to activate and challenge negative, irrational thinking. From these roots, the cognitive approach, integrated with behavior therapy in forming contemporary CBT, was expanded and adapted to meet the needs of children, adolescents, adults, and families to address a wide range of psychological and behavioral problems. Recent “third wave” influences on the practice of CBT reflect the integration of mindfulness and meditation principles found in dialectical behavioral therapy (DBT) as well as the acceptance and commitment therapy (ACT) focus on values and committed action. The foundation of the third wave approaches rests on the theoretical and empirical foundation of CBT.

THEORETICAL ISSUES

The value of CBT as an evidence-based and front-line treatment is ever-expanding and evolving. Since its inception and integration of the behavioral and cognitive movements, treatment packages such as self-instructional training, problem-solving therapy, harm reduction, behavioral family therapy, and behavioral couples therapy have been advanced. Problem-specific treatments have emerged and include the Coping Cat for childhood anxiety disorders, the Adolescent Coping with Depression Course, Brief Behavioural Activation for Adolescent Depression, CBT of Childhood OCD: It’s Only a False Alarm, The Incredible Years and Parent-Child Interaction Therapy for disruptive behavior in young children, Multisystemic Therapy for serious child and adolescent disruptive behavior and substance abuse, the Maudsley Approach for adolescent eating disorders, and the Unified Protocols for Transdiagnostic Treatment of Emotional Disorders in Children and Adolescents. All CBT protocols are based on the principles of social learning theory and several basic theoretical assumptions of CBT (Table 53.2–2). First, the cognitive-behavioral therapist does not assume that an underlying unconscious process or drive is fueling the behavior of any individual. Rather, the cognitive-behavioral therapist assumes that all behavior, whether adaptive and pathologic, is subject to conditioning principles, although there may be limits on learning due to constitutional constraints (e.g., intellectual capacity, physical limitations) or contextual constraints (e.g., dysfunctional home, depressed community settings, systemic racism or bias). As such, the therapist works with the patient (and with the parents in child cases) to examine samples of behavior across contexts to uncover the antecedents and consequences of behavior. The *ABCs of therapy* refers to the antecedents-behavior-consequence chains that serve to establish and perpetuate behavioral patterns. Rather than assume that an individual is predisposed to behaving in a certain way due to traits such as aggressiveness, introversion, or obsessiveness, the cognitive-behavioral therapist takes a more interactionist approach and assumes that the individual is interacting with the environment in developing response patterns to discrete and specific stimuli that may bear a resemblance to prior similar stimuli or experiences, thus shaping apparently trait-like responses. Through the process of CBT, behavioral antecedents and resultant consequences are identified and addressed within the treatment plan. The cognitive-behavioral therapist, based on empirical findings, applies procedures steeped in the principles of learning to change or modify the patient’s behavior. Through a process of collaborative empiricism, the therapist and patient (including parents or caregivers in the case of dependent youth) specify clearly defined goals for treatment and establish a time frame for achieving subgoals and overall goals. The therapist and patient will then engage in a continual process of evaluating whether goals are achieved and, if not, examining impediments to change, including resistance, throughout the course of therapy. CBT is applied flexibly to meet each individual patient’s needs. That is, techniques are adapted to be developmentally appropriate for a given age or cognitive level and otherwise adapted to meet the needs of an individual patient.

Table 53.2–2.
Assumptions of Cognitive-Behavioral Therapy

- Cognitive-behavioral therapy, relative to traditional psychotherapies, concentrates on behavior (this includes verbal behavior and covert behavior, such as thoughts, beliefs, memories, attributions, and images), rather than on some presumed unconscious underlying cause.
- Cognitive-behavioral therapy assumes that maladaptive behaviors are acquired through learning in the same way that any behavior is learned.
- Psychological principles, especially learning principles, are highly effective in changing and modifying maladaptive behavior.
- Cognitive-behavioral therapy involves setting specific, clearly defined, and observable treatment goals.
- Cognitive-behavioral therapy rejects classic trait theory—that is, the notion that an individual has an inherent predisposition to similar behavior in a wide range of situations.
- The cognitive-behavioral therapist adapts the method of treatment to a patient's problem.
- Cognitive-behavioral therapy is focused on the here and now yet does not neglect the past or history of the patient.
- Cognitive-behavioral therapists place great value on the empirical testing of their assessment and treatment methods.

Contrary to popular belief, CBT does not ignore historical information. The patient's history, including parental and related family members' histories, are essential for understanding the original setting events and learned reactions comprising the etiology of the problem. However, the work of CBT is focused on the present maintaining factors that prevent healthy functioning and allow the persistence of the patient's problem. Thus, CBT is focused on the here and now, with emphasis on understanding the current reinforcement paradigm, allowing persistence of the problem and teaching new skills to promote adaptive behavior. Finally, scientific rigor, whether on a large-group scale or within the single-case approach, is highly valued by the cognitive-behavioral therapist. It is expected that the cognitive-behavioral therapist will be current with the empirical literature and apply the scientific method to their practice of therapy. Thus, within the clinical practice, each patient is regarded as an individual case study in which the therapist and patient again collaborate to test hypotheses of behavioral change and refine their techniques or approach to maximize the potential for therapeutic benefit.

TECHNIQUES

Techniques are not the essence of CBT, nor are they sufficient for the resolution of any clinical problem. A thorough understanding of the empirical bases and theoretical principles of CBT, along with appropriate training and experience, is a prerequisite to the development of competency to formulate a sound treatment plan and deliver CBT to persons in need. In addition, CBT is a very action-oriented approach focused on the here and now. It requires a directive rather than passive stance. Therapists must be ready to teach, model, guide, and reinforce the patient in working toward their goals. Cognitive-behavioral therapists are very active in treatment and often conduct sessions in vivo—that is, in the situations in which the child or adolescent may be experiencing distress. The cognitive-behavioral therapist should expect to make a trip to the school, home, or other settings in which the child has difficulty, as appropriate, with careful planning to maintain confidentiality while promoting therapeutic benefit. Moreover, certain disorders, such as the phobias and obsessive-compulsive disorder (OCD), require the therapist to model the handling or touching of phobic or feared objects such as dogs, insects, and potentially “contaminated” objects such as doorknobs or toilet seats. Cognitive-behavioral therapists take care to devise these in vivo experiences to be maximally beneficial to the child while minimizing risk.

Behavioral Assessment

A classic phrase coined by Barrios and Hartmann in their review of CBT is that “behavioral assessment is the cornerstone of (cognitive) behavior therapy.” In essence, CBT is dependent on a carefully constructed and thorough functional analysis of the patient’s presenting problem. The functional analysis involves the use of appropriate assessment methodologies to uncover the relationships and conditions under which the patient developed and has maintained their problematic behavior (the ABCs defined earlier). Assessment in CBT involves multiple methods, multiple informants, and multiple time periods. Assessment involves multiple methods to capture the most useful and valid information concerning the patient and their functioning across a variety of contexts. Thus, diagnostic interviews, self-report questionnaires, diary forms, and behavioral observation techniques can all be used, given time and cost restraints. The use of multiple methods and multiple informants allows for data to be captured without relying on any one form of assessment or individual that may be subject to biases or limited in scope. For example, although children generally are able to report internal states with accuracy, younger children tend to report what is within their immediate experience, and, as such, even recent events may be distorted or inaccurately represented. In addition, the nature of some disorders may interact with youth report on any particular method of assessment. Adolescents with social phobia in general have difficulty admitting to this form of anxiety and tend to minimize or deny symptoms. In contrast, parents of youth with social phobia are able to identify and report the behavioral avoidance associated with the syndrome. Consider, for example the comparison of responses to questions from a diagnostic interview by a 15-year-old boy with social phobia and that of his parents (Table 53.2–3). Although the youth denies a lack of friendships and minimizes his social anxiety, the parents’ report establishes his isolation due to avoidance of social interaction and activities with peers. In this case, the youth may be unable to admit to the anxiety due to self-presentational demands; however, the parents are readily able to identify the specific behavioral contexts (after-school activities, mobile phone use) in which behavioral avoidance is apparent. This information, taken within the broader context of a diagnostic interview, will be compiled to establish a diagnosis and specify goals for treatment (e.g., engage in social activity outside of home and school; engage with friends on the phone or texting).

Behavioral assessment often involves obtaining samples of behavior through behavioral approach tasks. The behavioral approach task can be individually tailored for each patient or standardized for a particular disorder. Standardized behavioral tests for youth with social phobia involve giving a speech in front of a camera or a small, live audience. Youth with separation anxiety disorder may be asked to stay in the clinician’s office while the parent is sent on an errand. The behavioral approach task allows the clinician to assess the behavioral limits of the patient. If the child refuses to attempt the task, important information about their initial reactivity and ability to manage anxiety is accessed. No child or adolescent is ever forced to attempt a behavioral approach task; rather, they are allowed to avoid or quit the task at any point. The behavioral approach task also provides information on how long the patient can stay in the situation and what methods make the situation better or worse for the patient. For example, children who fear the dark may be better able to sit in a darkened room with the use of a flashlight or with another person with them. These safety signals provide information about what coping strategies the child is using and whether the strategy needs to be refined or removed through treatment. The behavioral approach task also allows for access of specific cognitive, physiologic, and behavioral content by having the patient list their top six thoughts that occurred during the task, taking measures such as heart rate or respiration, and recording the length of time engaged in the task or how close the patient can be to a feared stimulus. Optimally, the real stimulus that is feared (e.g., separation situation, small animal, heights) should be used in a behavioral test for maximal effect. However, certain stimuli or situations are inherently difficult to create in a therapist’s office due to confidentiality (e.g., social situations) or practical matters (e.g., airplane travel, heights), setting a limit on the ecologic validity of the stimulus and patient’s experience. Virtual reality environments, found to be efficacious and of high utility in the assessment and treatment of adult psychiatric conditions, offer a promising medium for creating age- and developmentally appropriate situations that mimic real-world stimuli and scenarios for behavioral tests as well as for use in exposure therapy in youth. Overall, the behavioral approach task provides both qualitative and quantitative information to establish baseline functioning and evaluate the progress of treatment. Sample behavioral approach task situations for anxiety disorders are given in Table 53.2–4. These tasks can also be used in designing exposure situations during treatment.

Table 53.2–3.

Comparison of Adolescent and Parent Reports in the Assessment of Social Anxiety Disorder

Interview Question[a]	Adolescent Report	Parent Report
Do you worry about what other people think of you or that you may do something embarrassing in front of others and then they'll think badly of you? Do you worry about this social anxiety more than other kids your age?	Yes, I worry I'll sound dumb sometimes. But I don't think it's any more than anyone else. I don't think so.	He is always worried about what he says and often says he does not know what to say. He fights with us about this. We try to push him to go out and talk to people, but he's too shy. He won't speak up and he is anxious that others will think he's stupid or awkward. This is different than our other kids who are quite social.
Do you think you have the same number, fewer, or more friends than other kids your age?	About the same as other kids.	Fewer. Actually, we don't think he has any friends that we're aware of and hasn't had since elementary school.
How often do you see your friends outside of school? What kinds of things do you like to do with friends?	Well, not much, maybe once a month. There's not much to do and nothing's really going on. And I have a lot of homework.	He doesn't see anyone outside of school. He just stays in his rooms and plays videogames. He doesn't go to any parties or to school events. We feel like

		we are constantly nagging him to go.
Do you text your friends to do things, or do you wait for them to reach out to you?	I text them back sometimes.	He doesn't seem to call or text anyone. We see our other kids always texting or on social media. Not him. One classmate called to get homework and even that was hard for him.

[a] Questions are modified appropriately when addressed to parent.

The use of multiple informants involves obtaining information from the child, parents, teachers, and, potentially, peers, other relatives, and former caregivers or therapists. The use of multiple informants addresses issues of responder bias while accessing the opinion of individuals who have experience with the child over a sustained period and in multiple contexts. Information gained from parents and teachers, for example, may reveal the occurrence of behaviors that are not likely to present in the therapeutic environment. However, reports from others may be limited by an inability to appreciate the antecedents or consequences of behavior and the inability to understand the child's internal dialogue or ideas. Critical to maximizing the use of the information gathered through behavioral assessment methods and the input of multiple informants is an understanding of role of inequalities related to social determinants of health (e.g., poverty/economic instability, housing and food insecurity, access to healthcare and resources, discrimination/racism). Structural racism and systemic bias continue to result in marginalized youth and families experiencing inequities and disparities leading to unmet or poorly administered mental health care. A comprehensive assessment needs be grounded in a competent and informed evaluation of the cultural and contextual factors affecting the child’s strengths and impairments in functioning. This then allows for a personalized, culturally responsive conceptualization and pragmatic treatment plan. Failure to include and address culture and context in the assessment risks perpetuating the disparities experienced by the youth and family. Finally, assessment occurs throughout the course of CBT. The initial assessment is used to establish a baseline diagnostic picture and establish the limits of functioning and identify behavioral excesses, strengths, and deficits. The goals for treatment are defined through data accessed during the assessment, and specific treatment protocols or techniques are chosen based on the assessment results. Assessment continues throughout the course of treatment to track progress, continue to gather data in specific domains (e.g., daily ratings of anxiety or depression, number of social contacts, diary forms to elicit though patterns and beliefs), and inform the therapist and patient about treatment response.

Table 53.2–4.
Sample Behavioral Approach Tasks for Child and Adolescent Anxiety Disorders and OCD

Social anxiety disorder: Reading aloud in front of the therapist and office staff, initiating a conversation with an unfamiliar person, sending a text to a peer, "liking" a social media post, dropping books in a crowded waiting room, asking for assistance from a store clerk, asking for directions to the restroom.

Selective mutism: Whispering to the parent in front of the therapist, speaking at a low volume/whispering to the therapist, asking a question in a normal voice to the therapist, speaking on the phone with a relative, leaving the therapist a voice mail, asking the teacher a question, calling and inviting a friend to get together, video talking with a friend or teacher.

Separation anxiety disorder: Staying alone in the therapy room for 10 minutes, staying alone in a darkened room for 10 minutes, parents leaving the child at the clinician's office to run an errand, being dropped off at therapy and not knowing parent's whereabouts, parent being late to pick up child at school.

Generalized anxiety disorder: Purposefully making mistakes on a homework assignment, being late for an appointment, reading an upsetting story online or watching an upsetting news clip about crime or world events, being told of unexpected changes to routines or plans.

Animal/insect phobia: Being in the same room with a caged animal, staying in the room with the animal on a leash, touching the animal, allowing the animal to roam free in the room, allowing the animal to interact freely with self.

Injections/doctor phobia: Entering an examination room, looking at the medical instruments, holding a syringe, allowing self to be examined by a physician, allowing needle to rest against skin.

Darkness phobia: Sitting in a dimly lit room with another person, sitting in the dimly room alone for 10 minutes, sitting in the room for an unknown period of time, sitting in complete darkness while holding a flashlight, sitting in complete darkness.

Obsessive-compulsive disorder: Touching "contaminated" objects (e.g., doorknobs, shoe soles), not washing hands, shaking hands with various office staff, saying "forbidden" words or thoughts aloud (e.g., “God is dead”), using “unlucky” or “bad” numbers on purpose.

Table 53.2–5.
Daily Diary for an Adolescent With Major Depressive Disorder During the Early Phases of Treatment

Column A What was happening? (stimulus)	Column B Mood Rating (0 = best, 10 = most depressed ever)	Column C What was I thinking?	Column D What did I do?
Parents were arguing.	8	This is miserable. I hate it here at home. Why don't they stop?	Went to my room and cried. Didn't eat dinner and stayed there all night.
Had an essay due in English class.	7	I can't think of anything to write about. There's nothing interesting to write. I can't do this.	Listened to music and lay in bed. Tried to write but couldn't. Watched some TV and then went to sleep.
Friends posting pictures while at a party I didn't know about.	5	No one tells me when there's something going on. I never get invited. I'm no fun and they know it.	Blocked all those people on social media. Didn't answer any of their texts when they tried to reach me. Felt like cutting myself.

Psychoeducation

A core component of CBT for just about any disorder or problem is psychoeducation for the child or adolescent and parents. Psychoeducation typically begins in the first treatment session and is carried on throughout therapy. Developmentally appropriate information is delivered to the youth and parents concerning the nature of the disorder in question (e.g., addressing the proposed etiologic pathways and maintenance of depression) and the process and procedures of CBT. It is important for patients and parents to learn that certain emotions may be “charged up” or “cooled down,” given changes in the environment (being teased by a bully, getting an A+ on a difficult examination, canceling a picnic due to rain), physical state (e.g., hunger, fatigue), time of day or year (school holidays, final exams) and other internal and external factors. Individuals may be born with a predisposition or vulnerability to experience certain emotions more strongly and readily than others, and over time, by interacting with the wider environment, a person can develop a style of thinking or psychological bias to interpret things in a more emotional manner (the “negative filter”). Although these biases exist, children and parents are further taught that individuals can learn strategies to manage their emotional reactions to a large degree, either with or without the assistance of medications. Emotions are broken down into three components—the physical (e.g., tension, irritability, physical symptoms of anxiety and depression), cognitive (e.g., automatic thoughts, beliefs, cognitive distortions), and behavioral (e.g., actions, aggression, withdrawal, avoidance). The relevant cognitive-behavioral techniques and principles tied to each component are offered as the ways in which the child will learn to recognize the triggers of their emotional reactions and use appropriate coping strategies effectively to manage the emotion and situation. Parents are also given specific information to address any gaps in their understanding and expectations for normative child development.

Table 53.2–6.
Daily Diary With Cognitive Restructuring Component for an Adolescent With Depression

Column E Which dispute handles can I use?	**Column F What is my Rational Response?**	**Column G New mood rating based on evidence and realistic thoughts**	**Column H What is my action plan?**
How often do they argue? Is this the worst thing that could happen? What are they usually like together?	Mom and dad are loving to each other, and all couples disagree and argue at times.	3	Focus on what they are usually like together. Sit with the family for dinner and, if the time is right, let them know I was upset hearing their argument.
How many times have I written an essay? How likely is it that I wont be able to think of anything if I try? Do I have to do this all at once?	I've prepared for class and read all the material, so I do have ideas to use in this essay. I still have 4 days to get this finished and that gives me plenty of time to work on this.	3	I will review my notes from class and make an outline for the essay. After a short break for a snack, I will begin to prepare the content of the essay. If I get this done tonight, I can rest and finish tomorrow.

How important is this particular party to my social life? Am I jumping to conclusions that no one tried to tell me about it? Do I have all the facts?	This is just one party and I’ve been invited before to others. With multitasking school, part time job and my social life, I’m going to miss things here and there.	2	Breathe and use my calming skills, and be proud of myself for not cutting. Reach out to my closest friends and ask if they had a good time. Make plans to get together for lunch and some fun.

Psychoeducation also involves the introduction and training of the child and parents in self-monitoring. Structured diary forms are often used to guide the monitoring of triggers to emotional states (both “positive” and “negative”) and rating of the intensity of emotion (e.g., from 0 to 10). Monitoring sets the stage for later tracking of cognitive and behavioral responses that become the focus of other treatment components. A sample monitoring form for an adolescent with depression is shown in Tables 53.2–5 and 53.2–6.

Somatic Management Techniques

The vegetative and physiologic symptoms accompanying anxiety and depression are addressed through structured relaxation and exercise assignments. Deep diaphragmatic breathing and progressive muscle relaxation training form the mainstay of techniques used to dampen anxious or aggressive arousal and increase a child or adolescent’s ability to control their unhelpful or inappropriate responding. Mindfulness exercises may also be introduced to teach children and adolescents strategies for nonjudgmental awareness of their bodies. Prescribed exercise regimens involving dancing, walking, bicycling, pilates, yoga, playing team sports, and engaging in similar physical activities are geared toward behavioral activation to reduce fatigue and increasing stamina and energy in the depressed adolescent. Nutritional counseling and sleep hygiene methods are also used as indicated to address concomitant problems such as eating and sleep disturbance accompanying the internalizing disorders.

Cognitive Interventions

The goals of cognitive interventions are to build on psychoeducation and provide corrective information concerning the presenting problem while systematically uncovering and refuting the automatic thoughts and schema that drive the youth's emotional distress. Automatic thoughts are discrete, rapid, and often context-specific thoughts that vulnerable individuals treat as fact about themselves, the world, and their future. All individuals have automatic thoughts each day. Typical automatic thoughts for youth with anxiety and depression are listed in Table 53.2–7. Cognitive restructuring involves identifying these thoughts through the use of diary forms, employment of Socratic questioning in the therapy room, and then developing rational responses based on actual evidence for the thoughts to assist with their emergence in everyday life. Apps have also been developed and disseminated to assist youth in monitoring and challenging thoughts in the moment. In effect, patients are taught to treat the thoughts as hypotheses, not facts. The therapist then guides the patient in planning home-based behavioral experiments to test whether the thoughts are true through the process of gathering evidence to either accept or refute them. Dispute handles, also shown in Table 53.2–7, are questions used to examine the evidence for the thought and guide the child or adolescent toward considering alternatives to their negative biases.

Table 53.2–7.
Typical Cognitive Distortions and Dispute Handles for Youth With Anxiety and Depression

Cognitive Distortions

I'm not as smart as other kids.
I don't know what to say to others.
I'm not as attractive as others.
No one wants to hang out with me.
They think I'm a geek because I was nervous giving that oral report.
What if I do something stupid?
Nothing will work out for me.
This is the worst situation anyone could ever be in.
I know I'm going to fail.
If I don't ace this test there's no point in trying anymore.
I should have done better.

Dispute Handles

Am I 100% sure that ____ will happen?
Do I have evidence that ____ will happen?
How many times has ____ happened before?
Is ____ really so important that my whole future depends on its outcome?
What is the likelihood that ____?
Does ____'s opinion reflect that of everyone else?
Do I have a crystal ball?
Am I responsible for this entire conversation?
What is the worst that could happen?
What can I do to cope/handle the situation?

Depending on the age and cognitive developmental level of the child, cognitive interventions can be simplified or rendered more complex. Young children and those who are more concrete in thought processes can be presented with a simple cue to "stop and think" when experiencing strong emotions such as fear or disappointment. A picture of a stop sign or favorite cartoon character holding up a stop sign may suffice to teach the child to resist the negative emotion and engage in proactive coping. Emotive imagery is a technique by which the child's favorite superhero or character provides "guidance" in managing the difficult situation. For example, a 7-year-old boy with separation anxiety disorder might be taught to imagine how Spider-Man's parent was late in picking him up from school. The therapist would coach the child to imagine Spider-Man handling each situation and then have the child imagine himself in each situation, with Spider-Man "standing by" in the image and assisting the child with coping strategies, such as deep breathing and thinking realistic, helpful thoughts. Older children are often provided acronyms to remember a series of steps in cognitive restructuring: Feeling frightened? Expecting bad things to happen? Actions and attitudes to help me cope with this situation. Results on how the situation turned out and how I handled it and reinforce myself for trying. Homework in cognitive restructuring involves the use of additional columns in the diary for (see Tables 53.2–5 and 53.2–6). These forms are reviewed in session, whereby the therapist and patient evaluate each step of the restructuring process and work through plans to overcome challenges and reinforce success experiences.

Exposure/Behavior Change Techniques

In the anxiety disorders, behavioral exposures are used to provide systematic success experiences while managing difficult situations and refining cognitive restructuring skills. Anxiety is partly maintained by avoidance of or escape from challenging situations, and thus the child does not gain experience to refute anxious thoughts and learn that the feared negative outcomes are unlikely to occur or can be managed. A child with separation anxiety disorder tries to avoid, at all costs, the experience of being away from loved ones for fear that something terrible will happen to result in permanent separation. By avoiding separation situations, the child maintains a focus on the catastrophic fear and does not learn that the feared negative outcome is highly unlikely. Moreover, avoidance prevents the child from learning that, when adverse or less-than-desirable or expected events occur, such as a parent returning home late from work, the event is independent of the child's fear and avoidance behavior. Exposure directly addresses the fear and avoidance partly through the process of habituation, by which the anxiety response naturally dissipates during the exposure. Exposures also provide opportunity for inhibitory learning, in which the child can learn that catastrophic feared events did not occur and new beliefs about potential outcomes come to inhibit previous beliefs. The child concurrently gains experience with coping and interacting with the feared stimulus or situation. Exposure is usually conducted in a stepwise manner of introducing increasingly challenging situations. Flooding, which is not recommended for young children or those with more severe anxiety levels, involves having the patient confront a situation that provokes the highest levels of anxiety and stay in the situation until the habituation occurs. Exposure may be imaginal or in vivo but is typically arranged hierarchically for children, from least to most challenging. Role plays with the cognitive-behavioral therapist may be included in preparation for exposures; however, they do not substitute for exposures. During imaginal exposure, the therapist may construct scripts about anxiety-provoking situations and ask the child to listen to the scene as if it were really happening. As the child's subjective anxiety decreases, scenes of greater intensity are presented in succession. In contrast, in vivo exposure involves having the child actually engage in the feared situation. Again, when the child reports tolerable levels of anxiety, the situation may be enhanced to produce a greater challenge to the child and require a new level of cognitive and behavioral coping. In contrast to situational exposures, interoceptive exposure involves inducing the sensations of panic through certain exercises and activities that cause physiologic responding, such as tachypnea, tachycardia, palpitations, diaphoresis, and paresthesias. Again, these exposures are designed to break the fear of internal sensations of anxiety and increase tolerance for normative changes in body functioning. Recent experimental studies indicate that habituation may not be necessary to reduce anxiety, but rather, exposure may assist in the learned tolerance of expected anxiety reactions that would previously have resulted in avoidance. For example, learning to tolerate the discomfort of dental procedures, the arousal accompanying asking someone for a date, and other situations that are inherently challenging to anyone, can result in improved functioning and quality of life, and reduced anxiety overall. In all types of exposures, children are encouraged to use their somatic management skills, as well as cognitive restructuring. Use of safety signals may be allowed for less challenging exposures, but systematic removal may be necessary for increasing exposure potency and testing potential feared outcomes.

Behavioral activation tasks are structured experiences planned for youth with major depression or persistent depressive disorder. The reinforcement theory of depression asserts that low rates of contingent reinforcement contribute to depressed mood and anhedonia and set up a cycle by which withdrawal behavior occurs and, thus, the youth is even less likely to engage in or have access to reinforcing events. For example, withdrawal from team sports, procrastination in homework assignments, and loss of interest in self-care may occur over time in depressed adolescents. By withdrawing from the team, social reinforcement and contact with peers comes to halt. Withdrawal also may have transactional effects on the teen as peers may disengage and taper contact in response, thereby further reinforcing potential distorted beliefs about themselves. Procrastination leads to a sense of dread and failure in school activity, and the drop in self-care compounds the youth’s negative self-image. The result is a systematic decrease is positive reinforcement and corresponding increase in negative feelings about the self. In CBT, the therapist surveys the youth for activities that may bring social reinforcement, success (self-efficacy or accomplishment), or pleasant feelings (e.g., relaxation). Typically, monitoring of mood occurs concurrently with the monitoring of these activities to provide feedback to the youth that engaging in such activity that the youth is likely to attempt, setting up parameters for engaging in the activity, and self-monitoring and processing the outcome of the event. Planned activities increase in type and number as treatment progresses, with the goal that the youth will assume the responsibility of planning these activities and that they will again become a natural and fixed part of their behavioral repertoire.

Parent Involvement in Treatment

Parents and family members are often inadvertently drawn into the cycle of negative affect accompanying the internalizing disorders. Family members may provide reassurance or reinforce escape and avoidance as a way of calming the anxious child or protecting the child from becoming upset and anxious. In depression, parents may take over tasks the adolescent is unable to complete due to anhedonia and vegetative symptoms, such as doing the teenager's homework or other compulsory acts. The intention of the parents is to prevent the child or adolescent from becoming even more anxious or depressed and, thus, to allay the distress experienced by their child. Unfortunately, although well intentioned, these parent behaviors are incorporated into vicious cycles of continued anxiety or withdrawal. In CBT, parents are provided a model of negative affect and education about the proposed etiologic mechanisms and maintenance of these states. The therapist and parents collaborate to examine maintaining variables in the child's environment and family system, to develop plans to change parental behaviors when necessary, and to rearrange contingencies in the immediate environment to foster the child's recovery. For example, in selective mutism, usually another child or family member will speak for the child and reinforce the child's mutism. The parents of adolescents with social phobia order for their adolescent in restaurants, speak to teachers on the adolescent's behalf, and have fewer expectations for their adolescent's social life. Parents of depressed youth may fear worsening the child's depression by challenging or encouraging participation in activities and thus prefer to "leave well enough alone." Parents are often taught to be good CBT "coaches" for their child. In essence, parents are taught to ask questions (use the Socratic method) to guide their child in arriving at a coping/action plan rather than providing an answer for the child. Much like a good coach, parents are taught to provide opportunities to practice the skills learned in CBT while refraining from judging the child. Reinforcing and praising the child's effort at coping, no matter how small, is essential for parents to learn in CBT. Parents are also taught systematic skills to ignore certain child behaviors (e.g., whining, refusal to speak, irritability) and to reinforce with positive attention those behaviors that approximate the goal behaviors. Parents are also trained in the use of contingency management procedures that may involve setting specific goals and contracts, such as "If Jonathan sleeps alone is his bed until 7 a.m., he gets 15 minutes of extra iPad time after school." The goal of contingency management procedures is to rearrange the reinforcement paradigms in the child's environment such that the child receives contingent positive reinforcement for appropriate behavior—in some cases, this may be stickers, small treats, or special privileges—whereas reinforcement is removed or withheld when inappropriate behavior occurs. These procedures also provide structured and, if necessary initially, scripted opportunities for parents to attend to positive behaviors and to effectively communicate this reinforcement. Cognitive-behavioral therapists do not advocate the use of punishment, which has been found to be ineffective in maintaining long-term behavioral change, in addition to being ethically unacceptable.

There are a few guidelines and caveats when planning the level and type of involvement of parents in the child or adolescent's therapy. First, the greater the distress and impairment for the child, the more likely it is that the parents are needed to reinforce the treatment program in the home environment. This is especially indicated in cases involving school avoidance behavior, in which the parents must rearrange contingencies in the home and actively work in cooperation with the school and therapist to promote attendance. Second, externalizing disorders or other complex comorbid conditions tend to fare better in CBT with parent involvement. Third, younger children have been found to respond better in acute treatment when parents are actively incorporated into the treatment plan. Fourth, the developmental goals of adolescence involve forming independent self-identity and behavioral skills; thus, parent involvement should take this and other developmental issues into consideration. Finally, in cases in which parental psychopathology or marital or family discord is present, the involvement of the parents may hinder rather than help the child's progress in treatment. In these cases, involvement of the unaffected parent or minimal parent involvement may be indicated, with referrals given as appropriate for concurrent marital or family therapy, parent management training, and individual parent therapy.

Relapse Prevention

As treatment draws toward a close, the therapist begins the process of transferring the responsibility for continued progress from the therapist to the child. That is, specific strategies and techniques are used to consolidate gains, reinforce coping skills, generalize gains to a wide range of contexts, and bolster the child's sense of self-efficacy. In the initial phase of relapse prevention, treatment sessions are moved to a biweekly and then monthly schedule to allow the child time to practice skills on their own and decrease reliance on the therapist. By identifying upcoming events or changes in the child's life (e.g., starting school, moving to a new neighborhood, facing developmental challenges calling for increased independence), the child is encouraged to develop a coping plan for each event and to plan their self-guided exposures in advance of the event. Hence, if a child is changing to a new school, plans to visit the school and meet the teacher or some peers are encouraged, and cognitive restructuring is completed for any negative thoughts or expectations that may occur. Similarly, booster sessions can be planned for anticipated tough times (e.g., the start of school, before annual examinations) to reinforce coping skills and assist with transitions. Throughout treatment, therapists can work with parents and children to incorporate technology to optimize coping. For example, deep breathing or mindfulness exercises can be recorded on smartphones, or coping reminders can be set for a specific date or time to encourage regular practice or situation-specific coping. In addition, smartphone applications and websites are available that provide age-appropriate CBT support (e.g., Kids Feelings App; Camp Cope-A-Lot). Finally, the child's diaries, workbooks, and journals from treatment become excellent relapse prevention aids for the child to refer to after the end of treatment. Photos taken during exposures or behavioral activation tasks that illustrate the child's accomplishments are pasted into these journals, bookmarked on the child or adolescent's smartphone, or put on posters in the child's room to reinforce gains. Finally, the Coping Cat program for anxious youth advocates the development of "commercials" by which the therapist and child produce a recording that incorporates all of the skills acquired during treatment. The child is the "star" of the recording and is shown as if teaching others how to overcome a problem. Commercials can also be produced in the form of comic strips, short books, poems, and other creative methods that inspire the child. Whatever the form the commercial takes, the child can refer to it for relapse prevention purposes and to reinforce gains. A resource for parents' "Do's and Don'ts" can also be generated for the parent component of relapse prevention. Table 53.2–8 presents a summary of the CBT components reviewed here.

CLINICAL ISSUES

Indications

CBT is indicated for the treatment of a range of disorders of childhood and adolescence, including anxiety and mood disorders, attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder, conduct, and oppositional disorder, habit disorders, OCD, PTSD, substance abuse, eating and feeding disorders, and parent–child discord, in addition to other conditions and behavioral disturbances. CBT is generally accepted as the first-line treatment for a range of disorders in childhood, with a combination treatment of CBT and medication preferred for more intractable or complex cases.

Limitations

Child or adolescent-focused CBT may not be the treatment of choice in situations involving extensive parent–child discord or marital problems. In such cases, as noted previously, the preferred course of treatment would involve family or marital therapy. The level of training and experience of the therapist typically limit the effectiveness of CBT, and, despite the widespread dissemination of CBT manuals and workshops, partially trained or inadequately trained therapists are no substitute for specialists in CBT. Hence, families and referral sources must be fully informed as to the degree of expertise in CBT of the therapist and adjust expectations accordingly. Similarly, the motivation and adherence to the treatment plan demonstrated by the patient and family may limit CBT. The therapeutic relationship, contrary to popular belief, is of critical importance to the cognitive-behavioral therapist. Thus, the cognitive-behavioral therapist will work to establish trust, positive regard, openness, and collaboration with the patient and family, no matter how rapidly a course of treatment is planned, to facilitate treatment progress.

Complications

Incomplete assessment and failure to fully uncover the antecedents and consequences of behavior may result in complications in therapy. It is not unusual for younger children or those who have successfully avoided phobic stimuli to underestimate their true level of fear. Hence, exposure may be too complex or overwhelming for a child, requiring the therapist to break the situation down into smaller, more manageable parts. Therapists who allow a child to escape or prematurely end an exposure session before adequate habituation and processing of the exposure can reinforce the anxiety and escape. Thus, it is essential that enough time is allotted for an exposure session and that the parameters of the exposure be clearly and fully examined with the child or adolescent. CBT does not purport to teach positive thinking but rather to teach optimistic, realistic, and coping-focused thinking. As such, exposure to less-than-ideal situations and outcomes is a part of treatment.

Contradictions

CBT is a flexible set of techniques and approaches that has been effectively applied to a wide range of child and adolescent disorders. There are no known contraindications for the use of CBT in young people.

CASE EXAMPLES

SOCIAL ANXIETY DISORDER

James was a 15-year-old boy in a single-parent, middle-class home and enrolled in the 10th grade of a public school. His mother sought evaluation and treatment for him for a longstanding history of “shyness” and anxiety in social situations. The evaluation revealed social anxiety disorder as the primary disorder in an adolescent who was temperamentally inhibited since early childhood. James had a great fear of making mistakes, looking foolish, and otherwise being the focus of attention when performing academic tasks or simply trying to engage in conversations with peers and teachers. At the same time, James was forthcoming with the evaluator and revealed that he knew he had always been attracted to other boys but did not act on his attraction for fear of rejection due to being gay. Also, he had not told his mother or other close relatives or friends of his sexual orientation. These concerns led him to feel dysphoric more days than not, as “I can’t even be my true self.” James was motivated for treatment and reported that he wished to make friends and feel more comfortable opening up and being himself with other people, and be more confident in social situations. Consequently, James was treated with cognitive-behavioral group treatment for adolescents with a social phobia —a 16-session course of treatment combining psychoeducation, cognitive restructuring, social skills, behavioral exposure, relapse prevention, and four sessions of parent involvement. James was initially quiet in the group; however, as treatment progressed, the other group members encouraged his interactions, validated his feelings, and expressed unconditional and nonjudgmental positive regard for him, resulting in James fully engaging in the therapy. The therapists designed social interaction exposures geared toward his engaging in conversations at school, volunteering to speak in class, making mistakes in front of others, and attending school-related social activities (e.g., clean up in the public park, car wash) where he would need to have conversations with others. A self-identified key goal for James, and among his most challenging homework exposures, however, was talking with his mother about his sexual orientation. The therapist role played various ways this could go—from his mother being accepting and open to her questioning whether he was confusing his feelings with a “phase.” Although he was prepared for a less-than-desired outcome with his mother, in actuality, she expressed relief that he had “finally” come out to her (“A mother knows these things”), and she encouraged their making plans to seek outlets for greater support and connection for him and for her to know how best to support James. He considered the school and peer-based exposures successful and subsequently went on to other social events and increased his class participation and engagement with school-based activities. In this case study, the importance of exposing youth to a range of potential outcomes for a given personal or social situation was crucial to James’s trust in the therapy and therapists and belief in the credibility of treatment. Moreover, he was appropriately prepared through behavioral exposure and practice to handle what could have been an awkward and discouraging situation.

Table 53.2–8.
Summary of the Main Components of Cognitive-Behavioral Therapy

Component	Focus/Goals	Associated Techniques
Psychoeducation	Provide corrective and accurate information about the nature of the disorder or problem in question.	Didactic instruction, self-monitoring (diaries, journals), assigned reading
Somatic management techniques		
Anxiety-related	Target autonomic nervous system arousal and related physiologic symptoms.	Breathing retraining (deep diaphragmatic breathing), relaxation training (progressive muscle/cue-controlled/applied relaxation), meditation, exercise
Depression	Plan behavioral activation tasks involving increase in activity to counteract negative, vegetative symptoms.	Exercise- and employ-related physical, social, mastery, and pleasurable activities
Sleep hygiene	Restructure sleep cycle through scheduling of awake activities and self-soothing activities to promote healthy and restful sleep.	Set tasks and guidelines geared toward decreasing arousal and promoting relaxation prior to bedtime; decrease/extinction of arousal activities such as exercise or

		use of stimulants (caffeine) in close proximity to bedtime.
Cognitive restructuring	Identify maladaptive (automatic thoughts, schema, irrational beliefs) thoughts, beliefs, and images and challenge thoughts with realistic, coping-focused thinking; in children, uncover “unhelpful self-talk” and replace with coping-focused self-talk (“coaching”).	Monitor thought processes through the use of diaries, apps, or questionnaires to reveal thinking styles, identification and labeling of automatic thoughts; teach rational disputation of automatic thoughts using realistic evidence and problem-solving steps.
Exposure/behavior change techniques	Systematic exposure to avoided situation(s) to provide experience with using cognitive restructuring and other deficient skills (e.g., social skills, problem-solving skills, conflict resolution, affect regulation).	Imaginal exposure, in vivo exposure, gradual or massed (more intense) exposure to anxiety-provoking situations; in depression, use of behavioral activation tasks or behavioral experiments to counter negative and self-defeating behaviors and gather evidence to refute depressive ideation.

Reinforcement techniques	The systematic application or withdrawal of stimuli to either increase desired behavior or decrease undesirable behavior.	Patient earns reinforcement through behavioral contracts, level programs, or parent-, therapist-, and/or school-administered contingency management programs; reinforcers can include tangible items (stickers, money), social reinforcers (praise, positive attention, elevation in status), or desired activities (free time, computer/videogame use, extended curfews); time out from positive reinforcement is the removal of the individual from a reinforcing situation for a set period; punishment, which is the application of a noxious or undesired stimulus, does not result in long-term behavioral change and is not
Positive reinforcement	Provide a desirable stimulus (praise, attention, tangible reward) after a desired behavior to increase the probability of the behavior occurring again.	
Negative reinforcement	Remove an undesirable stimulus after the production of a desired behavior to increase the probability of the behavior occurring again.	
Punishment	Apply a noxious stimulus after an undesirable behavior to decrease the probability of it occurring again.	

		recommended for use by child behavior therapists.
Relapse prevention	Focus on consolidating cognitive-behavioral therapy skills and generalizing treatment gains over time; decrease reliance on therapist and others (e.g., parents, spouse) for managing problems.	Fading of sessions (from weekly to biweekly), role reversal (patient acts as therapist for a session), planned booster sessions.

SELECTIVE MUTISM

Anna was a 5-year-old girl from an intact middle-class family referred by her kindergarten teacher due to selective mutism. Anna was described as always having been a shy child who was behaviorally inhibited. Although she would speak and interact with immediate family members, she would withdraw and become quiet around peers and strangers. When she started preschool, she did not speak to her teachers or any school personnel and also began to withdraw from speaking around extended family members at holidays and parties. She also experienced distress about separation from her mother, who would generally respond to her and make requests for her. The school had also allowed her to nod and use nonverbal cues to communicate her preferences, and she was not required to read aloud or speak in front of her peers. She occasionally played and engaged with peers on the playground and was otherwise pleasant and friendly. During the evaluation, it was revealed that Anna enjoyed eating Skittles as a treat. Treatment was designed to provide an incentive for speaking through the delivery of a contingent reinforcer of high value, the Skittles. Hence, Skittles became available only in the school and therapist's office and, temporarily, were not available in her home. The therapist enacted a process of graduated shaping of communication behaviors—first nonverbal and then vocal noises—and trained the teachers to do the same. Skittles were kept in full view of Anna at all times during the initial phase of treatment, and when she was "caught" gazing at the package, the therapist or teacher would prompt her for acknowledgment that she wanted the treat. Pointing, looking, and nodding in their direction resulted in receiving two Skittles. While in session, the therapist would also leave the room while she and her mother were left to play; if the therapist returned and she continued to vocalize or speak, she would receive additional reinforcement via Skittles. Next, the therapist and parent spent time "making noise" with objects (e.g., rattling keys, pressing telephone buttons) and then with their mouths (e.g., animal sounds, silly sounds), and Anna was prompted to join in and make similar sounds to receive the Skittles. Finally, prompts to ask for the Skittles were enacted, first with an “Ssss,” then the full word, and finally the full sentence/question (e.g., “Can I have a Skittle?”). This shaping procedure took three extended-length sessions and collaboration, along with telephone consultation with her school teachers. In addition to treat requests, Anna was increasingly reinforced for additional situations on her fear and avoidance hierarchy: making eye contact, saying hello to neighbors, ordering her own food in a restaurant, purchasing her own Skittles from a drug store, singing in the park, etc. The treat was faded over time—that is, delivered on a variable ratio schedule of every three to eight times that she spoke to promote further speaking and decrease the association with the treat. As a part of their relapse prevention plan, Anna’s parents were reminded to allow or prompt Anna to speak for herself in various pre-identified possible situations.

OBSESSIVE-COMPULSIVE DISORDER

Ayana was a 13-year-old girl with a family history of anxiety and mood disorders. Her parents brought her to treatment due to recurrent obsessions involving contamination and germs and corresponding compulsions by which she had her parents overcook her food, check her food for contaminants, and wash her hands repeatedly to the point of their becoming raw and bleeding. The evaluation revealed a fear that, unless her parents checked her food for bugs or germs, the meal was contaminated, and she would get sick and die. Hence, after cooking it, her parents had to pull apart her food physically and examine it to her satisfaction, which could take upward of 1 hour for each meal. This caused much distress and family discord. In addition, Ayana washed her hands repeatedly after opening a door, using a cell phone, tying her shoelaces or putting on or taking off her shoes, or touching any object that she deemed dirty. First, Ayana, her parents, and the therapist mapped her OCD and constructed a hierarchy based on her compulsions, rituals, and avoided situations (from the least upsetting, checking food prepared by her mother, to the most upsetting, touching something that was wet and slimy and then touching her mouth). The process of exposure and response prevention was introduced in combination with a contingency reinforcement plan for Ayana’s completion of “brave challenges.” Systematically, the therapist had Ayana first engage in imaginal exposure to a scene (e.g., “You watch dad take a bite of his medium rare burger, then you imagine taking a bite of a medium rare burger with the red juice oozing out. You realize it is not super well done, and they did not check it for germs.”) until her anxiety dropped to minimal levels. This usually occurred after 20 to 30 minutes. Next, in vivo exposures were gradually introduced, by which foods were introduced with “contaminants” in them (e.g., putting pieces of uncooked rice into the burger to mimic “grit”), phasing out the parents’ checking of meals as well as her own checking behaviors. As treatment progressed, Ayana learned that her feared consequence of becoming sick was not likely to occur. Washing rituals were addressed by having her touch items with various substances coating them and then touching her face and mouth. At first, she was asked to delay washing her hands over increasing time increments and then to resist for larger blocks of time. Importantly, it was critical to address some of Ayana's mental ritualizing thoughts regarding when she would be next able to wash her hands. Treatment occurred over a 14-session program, with some sessions including extended time to allow for exposure and response prevention to occur. Her parents were taught to assist her with these exposures in the home and refrain from engaging in her rituals or accommodating them (e.g., removal of antibacterial sanitizing gels). They were also provided coaching scripts so as not to engage in disputes over meals. Relapse prevention plans were enacted to expand her range of food choices and situational contexts (cafeterias, food stands, restaurants) for exposure. By the end of treatment, Ayana was eating without the need for checking and with minimal to no anxiety, and her handwashing had decreased to within the normative range of daily washing.

GOALS OF TREATMENT

In the abstract, the overall goals of CBT are to increase an individual's freedom of choice—that is, to free the individual from the confines of a disorder which removes the opportunity to make decisions and choices for one's self out of personal preference and which, instead, determines an individual's choices as a result of fear, depression, anger, or other emotional constraints. The goals of CBT are symptom relief, education about the nature of one's symptoms, vulnerabilities, and patterns of behavior, and the teaching of adaptive, coping-focused, realistic, and optimistic behavior (cognitive and overt) to maximize the person's potential, prepare them to meet the challenges of everyday life and improve their quality of life.

ETHICAL ISSUES

CBT is a safe, tolerable, and effective treatment for a range of behavioral health problems, with increased efficacy linked in some clinical studies to more seasoned and well-supervised therapists. The main risks of CBT may lie within exposure-based treatments, which are associated with temporary emotional discomfort when confronting feared situations. Contrary to the myth that exposure may result in adverse reactions (e.g., unrelenting anxiety, psychosis, breach of therapeutic alliance), any discomfort and boundary issues in conducting exposure can be skillfully and carefully managed by the experienced therapist. When conducting exposure in community (in vivo) settings, such as schools, parks, and playgrounds, therapists need to be mindful of confidentiality concerns and take precautions to minimize compromising the patient’s right to privacy.

RESEARCH AND EVALUATION

CBT is considered the “gold standard” psychosocial treatment approach with more empirical support than any other psychosocial intervention. Efficacy in the treatment of childhood internalizing and externalizing disorders through a number of controlled, randomized trials, including combined trials of medication and CBT. Long-term follow-up of youth treated for anxiety, OCD, or depression suggests maintenance of treatment gains, in addition to further improvement over time. It is incumbent on clinical trial researchers in CBT to test this approach in youth presenting with variations in cognitive and developmental functioning, comorbidity patterns, and various environmental adversities to test the effectiveness of therapy programs in mental health service settings. Further research is needed to evaluate treatment algorithms, discontinuation procedures, and pretreatment factors that predict the maintenance or relapse of treatments over the long term. The critical issue faced by today’s clinical scientists involves moving treatments from the laboratory to the community. This rests on the collaboration and cooperation of clinical scientists, clinicians, patient advocates, and consumers for critical feedback to flow in multiple directions to adapt and adopt evidence-based treatments in community settings.

FURTHER READINGS

Barrios BA, O'Dell SL. Fears and anxieties. In: Mash EJ, Barkley RA, eds. *Treatment of Childhood Disorders*. 2nd ed. Guilford Press; 1998:249.

Bioulac S, de Sevin E, Sagaspe P, et al. What do virtual reality tools bring to child and adolescent psychiatry? *Encephale*. 2018;44(3):280–285.

CATS Consortium. Implementing CBT for traumatized children and adolescents after September 11: lessons learned from the Child and Adolescent Trauma Treatments and Services (CATS) Project. *J Clin Child Adolesc Psychol*. 2007;36(4):581–592.

Cavallo DA, Cooney JL, Duhig AM, et al. Combining cognitive behavioral therapy with contingency management for smoking cessation in adolescent smokers: a preliminary comparison of two different CBT formats. *Am J Addict*. 2007;16(6):468–474.

Chu BC, Pimentel SS. *CBT Treatment Plans and Interventions for Depression and Anxiety Disorders in Youth*. Guilford Press; 2023.

Comer JC, Puliafico AC, Aschenbrand SG, et al. A pilot feasibility evaluation of the CALM Program for anxiety disorders in early childhood. *J Anxiety Disord*. 2012;26(1):40–49.

David D, Cristea I, Hofmann SG. Why cognitive behavioral therapy is the current gold standard of psychotherapy. *Front Psychiatry*. 2018;9:4.

David-Ferdon C, Kaslow NJ. Evidence-based psychosocial treatments for child and adolescent depression. *J Clin Child Adolesc Psychol*. 2008;37(1):62–104.

Freeman JB, Chaote-Summers ML, Moore PS, et al. Cognitive behavioral treatment for young children with obsessive compulsive disorder. *Bio Psychiatry*. 2007;61(3):337–343.

Freeman JB, Garca AM, Coyne L, et al. Early childhood OCD: preliminary findings from a family-based cognitive-behavioral approach. *J Am Acad Child Adolesc Psychiatry*. 2008;47(5):593–602.

Hayward C, Varady S, Albano AM, Thienemann M, Henderson L, Schatzberg AF. Cognitive behavioral group therapy for female socially phobic adolescents: results of a pilot study. *J Am Acad Child Adolesc Psychiatry*. 2000;39:721–726.

Hirshfeld-Becker DR, Masek B, Henin A, et al. Cognitive behavioral intervention with young anxious children. *Harv Rev Psychiatry*. 2008;16(2):113–125.

Jablonka O, Sarubbi A, Rapp AM, Albano AM. Cognitive behavior therapy for the anxiety triad. *Child Adolesc Psychiatr Clin N Am*. 2012;21(3):541–553.

Kazdin AE, Weisz JR. Identifying and developing empirically supported child and adolescent treatments. *J Consult Clin Psych*. 1998;66:19–36.

Kearney CA, Albano AM. *When Children Refuse School: A Therapist's Manual*. Psychological Corporation; 2000.

Kendall PC, ed. *Child and Adolescent Therapy: Cognitive Behavioral Procedures*. 4th ed. Guilford Press; 2011.

Mash EJ, Barkley RA, eds. *Child Psychopathology*. 3rd ed. Guilford Press; 2014.

Pavlov IP. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. Oxford University Press; 1927.

Pediatric OCD Treatment Study (POTS) Team. Cognitive-behavior therapy, sertraline, and their combination for children and adolescents with obsessive-compulsive disorder: the Pediatric OCD Treatment Study (POTS) randomized controlled trial. *JAMA*. 2004;292(16):1969–1976.

Pina AA, Polo AJ, Huey SJ. Evidence-based psychosocial interventions for ethnic minority youth: the 10-year update. *J Clin Child Adolesc Psychol*. 2019;48:2:179–202.

Sanchez AL, Jent J, Aggarwal NK, et al. Person-centered cultural assessment can improve child mental health service engagement and outcomes. *J Clin Child Adolesc Psychol*. 2022;51(1):1–22.

Scheeringa MS, Salloum A, Arnberger RA, Weems CF, Amaya-Jackson L, Cohen JA. Feasibility and effectiveness of cognitive behavioral therapy for posttraumatic stress disorder in preschool children: two case reports. *J Trauma Stress*. 2007;20(4):631–636.

Shooshtary MH, Panaghi L, Moghadam JA. Outcome of cognitive behavioral therapy in adolescents after natural disaster. *J Adolesc Health*. 2008;42(5):466–472.

Silverman WK, Pina AA, Viswesvaran C. Evidence-based psychosocial treatments for phobic and anxiety disorders in children and adolescents. *J Clin Child Adolesc Psychol*. 2008;37:105–130.

Skinner BF. *Contingencies of Reinforcement: A Theoretical Analysis*. Prentice Hall; 1969.

Stallard P. Evidence-based practice in cognitive-behavioural therapy. *Arch Dis Child*. 2022;107:109–113.

Storch EA, Merlo LJ, Larson MJ, et al. Impact of comorbidity on cognitive-behavioral response in pediatric obsessive-compulsive disorder. *J Am Acad Child Adolesc Psychiatry*. 2008;47(5):583–592.

Szigethy E, Kenney E, Carpenter J, et al. Cognitive-behavioral therapy for adolescents with inflammatory bowel disease and subsyndromal depression. *J Am Acad Child Adolesc Psychiatry*. 2007;46(10):1290–1298.

Tsiros MD, Sinn N, Brennan L, et al. Cognitive behavioral therapy improves diet and body composition in overweight and obese adolescents. *Am J Clin Nutr*. 2008;87(5):1134–1140.

Van den Akker EL, Puiman PJ, Groen M, Timman R, Jongejan MY, Trijsburg W. A cognitive behavioral therapy program for overweight children. *J Pediatr*. 2007;15(3):280–283.

Walkup J, Albano AM, Piacentini J, et al. Cognitive-behavioral therapy, sertraline, or a combination in childhood anxiety. *N Engl J Med*. 2008;359(26):2753–2766.

53.3 Group Psychotherapy

MARGO L. THIENEMANN, M.D.

Group therapy is the counseling, guidance, or training of children and adolescents involving interactions of three or more members that has the potential for reciprocal influence. Empirical support for group therapy for children and adolescents is growing. In 1997, Matthew J. Hoag and Gary M. Burlingame published a meta-analysis of the 56 English-language studies of child group therapy that used randomization or matching in their experimental designs. Compared to wait-list or placebo control, group therapies helped to a moderate degree (effect size 0.61, range –0.04 to 2.99). More recent studies have addressed specific populations, as discussed below.

Group interventions for children and adolescents can be categorized by goals, therapeutic processes, duration, and population served and may be referred to as *psychotherapy groups, psychoeducational groups, support groups,* and *self-help groups.* Delivering therapy via groups may be better than individual therapy for addressing specific illnesses or may more effectively address treatment issues such as motivation for change or the need for social reinforcement of learning.

HISTORY

Therapists have been describing group treatments for children for approximately 100 years. Early in the 20th century, adolescents met for psychodrama groups in Vienna. In the 1930s, *activity group therapy* developed. In these groups, children played among therapists who used nonverbal, noninterpretive interactive strategies to effect “ego repair.” After World War I and during the Depression, when delinquent youth needed constructive supervision and resources, and manpower was short, adolescents were treated in groups. After World War II, understaffed military hospitals in the United States and England used groups to deal with psychiatric casualties. The 1960s saw the emergence of encounter groups, the community mental health movement, antiestablishment groups, and increased tolerance in society for talking openly about one's feelings.

Today, groups are used frequently and ubiquitously. Psychotherapy, support, psychoeducational, and self-help groups may be found in all human service facilities, including schools, hospitals, clinics, and on the internet. Conceptual understanding and beginning clinical skills of group therapy are listed among the required core competencies for child and adolescent psychiatrist training, according to the Accreditation Council of Graduate Medical Education (ACGME), though Pollock and Kymissis noted that few training programs include adequate training in group psychotherapy.

Environmental factors have shaped these historical trends. Pollock and Kymissis, discussing the future of adolescent group therapy in 2001, posited that theoretical and clinical developments, training opportunities, scientific research, economics, social change, and public and professional attitudes all influence the type, quality, and availability of group therapy. Psychoanalytic theory informed early groups, which aimed to use group interactions to recapitulate family interactions. Developmental theory, considering the peer group's role in identity formation, pointed to using the group for adolescents' therapy. The development of learning theory and cognitive-behavioral therapy (CBT) has led to problem-focused, time-limited group interventions.

Societal changes and economic factors have increased the appeal of group interventions. According to Miffitt, the current health care climate of high demand and limited resources makes the group therapy format cost-effective and fiscally responsible. Due to the developmental tasks of adolescence, which include identity formation and development of the concept of self through peer relationships, peer approval, and belongingness, group modalities are perhaps especially useful for pre-adolescent and adolescent children (paraphrased from Miffitt, 2014). Today, many group therapies have been researched, and as all evidence-based therapies require treatment manuals, the practitioner will have access to guidance.

Some factors serve as potential obstacles to the use of group psychotherapies. Managed care, because of group therapy's potential cost-effectiveness, initially provided an incentive for increasing the use of outpatient group therapy. Despite increased societal comfort with self-disclosure, therapists and patients may reject group therapy, considering it diluted and less effective. Perhaps, as has been the case for pharmacologic, individual, or other treatment modalities, research will demonstrate that specific group therapies for specific problems administered by specifically trained individuals help, and the interventions may gain adequate support for dissemination.

THEORETICAL ISSUES

Therapeutic Power of the Group

Gathering individuals together for therapy may help them more than working with them alone. Existential psychiatrist Irv Yalom published The Theory and Practice of Group Psychotherapy, in which he described factors that make psychological treatment in groups especially effective. Yalom listed the instillation of hope, universality, imparting of information, altruism, corrective recapitulation of the primary family group, development of socializing techniques, imitative behavior, catharsis, the addressing of existential factors, cohesiveness, and interpersonal learning as factors leading to change. In children and adolescents, the use of the group to address developmental issues may also be an advantage.

Hope. In groups, children can meet other children with similar troubles. As they see peers overcome obstacles and actively addressing problems, they can develop hope that they, too, can help themselves get better. Families may gain hope; a parent of an unmotivated anxious child remarked that seeing other children tackling their anxieties gave her hope that her child would be able to do so. Similarly, in multi-family groups, parents often remark that seeing other children improve behavior-wise gave them hope that their families could also begin to function in a healthier manner.

Universality. Children with psychiatric disorders frequently feel they are “weird,” “crazy,” or “losers.” Getting to know peers with similar issues can lessen the pain and reduce the erosion of self-esteem that comes with having a mental health disorder. Adolescents are in the midst of exploring and consolidating their identities during the time that a number of major psychiatric illnesses begin to manifest themselves. Having the opportunity to interact with others with similar difficulties can help teenagers integrate having a psychiatric illness into a broader sense of self. In a recent pilot study by Straub et al., reviewing the most positive evaluations of their time-limited cognitive-behavioral group for teen outpatients, the most positive elements identified by participants were being with other adolescents and feeling comfortable in the group. By seeing others in the group, teens can come to see a disorder as only one of many aspects of themselves.

Imparting Information. Children learn together at school, on teams, and in families. In group therapy, children can learn about the course and characteristics of their disorder. In addition to presenting live examples, the group format is suited to psychoeducation via leaders’ instruction, educational videos, and structured learning activities. When the child teaches and helps other members, their learning is reinforced. Following the Coping Cat anxiety protocol model, members of one adolescent group treatment for obsessive-compulsive disorder (OCD) ended the group by making a didactic video. The group members delighted in their educational video, which taught about the etiology, symptoms, treatment, and impact of OCD. They watched it repeatedly and shared it with their families.

Altruism. Supporting others in the group is gratifying and a positive coping mechanism. Altruistic actions can boost one’s sense of well-being and self-esteem and help to defend against upsetting feelings. Children and adolescents in group therapy often make thoughtful suggestions to other members. One study of adolescents participating in cognitive-behavioral group therapy reported that most participants tried to help their peers to design behavioral interventions in every session.

Corrective Recapitulation of the Primary Family Group. Through examining psychodynamic processes and addressing treatment resistances, the group format can provide a wealth of opportunities for addressing family conflicts. In a group setting, children develop transference relationships with leaders and peers. Providing a safe, constructive atmosphere in which maladaptive interpersonal patterns are addressed and reworked may provide group members a chance to work through dynamic conflicts originating in the family.

Development of Socializing Techniques. Yalom stated that all therapy groups help patients develop social skills. Social deficits, social dysfunction, and isolation are prominent concerns for children with psychiatric disorders (including those with pervasive developmental disorders), learning disorders, attention-deficit disorder, other externalizing disorders, and social anxiety and depressive disorders. Groups allow patients to practice new skills and receive feedback from professionals and peers in real-time. In one controlled group therapy study of girls without friends, girls in the treatment group did learn to approach peers and begin conversations. Reinforcement and criticism from peers may be the most meaningful feedback for adolescents.

Imitative Behavior. Seeing how others behave greatly influences what children and adolescents do. Children in groups observe other children working on solving problems like theirs. Group leaders often demonstrate strategies, like social problem-solving and conflict resolution, as ways to deal with these. Many group protocols call for role-playing and rehearsal of skills. Sexually abused children practice acting assertively, and impulsive children practice entering an ongoing conversation successfully. In the pilot study by Straub et al. examining a CBT group for depressed teens, participants reported that they felt understood by and comfortable with the group leaders and rated their advice as helpful. Imitating effective behaviors modeled by leaders whom children respect and peers with whom children may identify effectively teaches functional behaviors.

Catharsis. Group therapy can allow one to express one’s distress in the presence of individuals who can truly understand one’s suffering. Although catharsis provides some relief, catharsis alone is insufficient for change. Members of one adolescent trichotillomania treatment group noted that talking at support groups had helped initially but left them feeling dejected after hearing about others’ experiences. They felt more hopeful when they could learn new tools with which to address their behaviors.

Existential Factors. In the trichotillomania group just mentioned, one member openly expressed hopelessness and refused to address her hair-pulling. The other group members challenged her stance, urging her to take responsibility for attempting to change her behavior. Unwilling to take charge at that time, she left the group. Issues of personal responsibility will change, loneliness, and demoralization frequently emerged in the context of group therapy, which provided a rich format in which to discuss these universal existential concerns. Adults in one substance abuse group rated the group’s ability to address existential issues as its most powerful component.

Cohesiveness. The experience of being part of something, being self-revelatory *and* accepted, and working together feels good and can improve children’s motivation to work in therapy. When children and adolescents are secretive about their disorders, they may have a sense they do not belong anywhere. In a group, they can enjoy a sense of closeness when they can be open about their troubles, are accepted, and are understood. A study by McConachie et al.

investigating group therapy for anxiety in children with autism spectrum disorder found that the most helpful strategies were learning alternative coping strategies, sharing experiences, and feeling accepted.

Interpersonal Learning. By working in the here and now of relationships in the group sessions, individuals have an opportunity to address dysfunctional interpersonal difficulties. The group can be a safe place to develop and practice healthy ways of interacting. For example, in a mentalization-based therapy group, teens are often encouraged to discuss when they find other participants' behaviors confusing or distressing so that the behaviors can be understood and improved.

Developmental Factors

Each stage of development has its tasks. Younger children's egocentric focus differs from preadolescents' yearning for same-sex soul mates and from adolescents' empathic, individuating, and affiliative capacities. Understanding the developmental needs and capacities of a set of children or adolescents is necessary for designing and enacting a successful group therapy intervention. A parent-only, group cognitive-behavioral intervention for children with anxiety disorders was used to take advantage of the collaborative working relationship between school-age children and their parents (and of the benefits of parents working together at their common stages of parenting). Aronson delightfully discussed ways in which adolescent developmental issues are addressed in a paper, "Where the Wild Things Are." In it, he likened the journey of Maurice Sendak's Max trying on his new identity—affiliating with peers, individuating, and returning with a new relationship to his family—to psychological processes occurring in adolescent group therapy. Whether the group is a cluster of same-age peers or contains a range of ages, the group interaction can foster developmental stage-appropriate maturation.

Other Factors

Most geographic areas have an inadequate number of psychotherapists, and most individuals have little surplus money for psychotherapy. The group format uses therapist resources in a lower-cost way to provide care that otherwise might not be available due to a lack of resources.

TECHNIQUES

Gains made in symptom resolution and psychosocial development in groups can result from general therapeutic factors and from learning specific information and skills. Before discussing specific diagnoses and problem-focused interventions, this subsection reviews general technical issues.

Selection of Patients

Number. A typical group contains 6 to 10 members—a size large enough to allow for a rich exchange but small enough to prevent members from being neglected. Having enough at the outset ensures against depletion due to dropouts and absences. Sometimes, goals call for a smaller or larger group. In one treatment study, aggressive young boys practiced coping with being teased. Because of the intensity of the intervention and the potential for strong emotions and behavioral acting out, the interventions were done in triads. On the other hand, Meezan and O'Keefe studied a large-group format used to address abusive and neglectful caregivers and their children. Two or more members of at least two generations of each family met with five to seven other families and four therapists. More change (including a reduced incidence of child abuse and improved child behavior) occurred in the families in the intervention group than in a control group using an individual family treatment. The authors believed that families improved more because they met families who had "been there" and other families who motivated them and encouraged them to change. Families felt that compared to feedback from therapists, feedback from other families was easier to take and helped them the most.

Describing an inpatient community group of 18 severely disturbed and conduct-disordered boys between the ages of 11 and 16 years, another author remarked that although maintaining structure and creating a safe environment presents a challenge with a large group of disturbed children, the larger group was powerful and a good choice for reluctant participants and those limited in their capacity to share their thoughts and feelings.

Age and Sex. Participants' age defines the nature of the interactions and the structure of the intervention. Some investigators have included children of different ages and have found that members benefit from the interesting modeling and teaching opportunities provided by the family-like grouping. Most often, groups are composed of children at similar developmental levels (e.g., 8 to 12 years and 12 to 18 years). When patients are of elementary school age, mixed-sex groups are often appropriate, especially for the younger grades. For this age group, games and activities can be fun while they facilitate interactions and learning. As the children age, they can tolerate a format increasingly focused on verbal interaction. Children in preadolescence prefer same-sex peers and can be distracted and inhibited by members of the opposite sex. Adolescents enjoy and can use mixed-sex groups again. Mixing highschool–aged adolescents with younger ones may expose the younger children to topics with which they are not comfortable.

Gathering Group Members

Difficulty gathering a cohort of patients at once can impede the initiation of group work. Wood and coworkers studied a group format for treating adolescents who deliberately harm themselves. Sixty-three adolescents who had intentionally injured themselves were randomly assigned to either group therapy along with routine care or routine care alone. The group format was chosen for both clinical and administrative reasons: because poor peer relationships and impaired problem-solving were thought to be amenable to group intervention and because group therapy was hoped to decrease the use of individual therapy and train a team of therapists for dealing with self-harm. The authors advanced the field by writing a manual describing necessary complex psychotherapeutic techniques. The manualized therapy—a developmental group psychotherapy—included a variety of techniques, including problem-solving, CBT, positive corrective therapeutic relationships, dialectical behavior therapy (DBT), and psychodynamic group therapy. After the assessment, six acute sessions were followed by long-term group therapy, which was open-ended. Group members could join the groups at any time. The themes of acute sessions included relationships, school problems, peer relationships, family problems, anger management, depression and self-harm, hopelessness, and feelings about the future. While the group therapy did not reduce levels of depression or suicidal thinking, participation in the group (the median number of sessions was 8; the range was 0 to 19) did lead to a change in the incidence of deliberate self-harm and the use of other therapy resources. Those who had more group sessions were less likely to harm themselves again, whereas those who had more individual sessions were more likely to repeat self-harm. This study is interesting not only for the clinical outcome but also for a model for running a continuous group, adding newcomers after a six-session lead-in.

Excluding Patients

Certain individuals in group therapy may undermine the process to the extent that, although the individual might benefit from participating, their detriment to the group calls for exclusion. Particularly in adolescent groups, individuals who actively use drugs may disrupt the group by showing up under the influence or reveling in escapades. When this does not serve the group's goals, these individuals should be excluded. Fragile children and adolescents may experience the group as unduly stressful, thus contraindicating their participation. Most group protocols exclude actively suicidal or psychotic individuals, who may monopolize the leader's attention, upset other members, and require the leader to leave the session to appropriately triage the patient. On the other hand, crises and decompensations are part of the territory of treating mentally ill children and adolescents. Witnessing the therapist compassionately and competently handle a troubled peer may be reassuring to a group member. Including an unwilling “captive” member in a group could drag the group down in a mutinous way or, skillfully handled, highlight issues of motivation and personal responsibility for each member.

Investigating the importance of carefully selecting members (considering such factors as level of acting out or ego development), Johnson and Gold sorted 34 latency-aged boys into four groups based on a measured leadership quality score. Groups met weekly for 20 sessions. All group-treated boys improved in popularity ratings and leadership. The authors remarked that the therapist’s ability to match techniques to the characteristics of the group members was more important than group composition.

Parents

Parents spend many more hours with their children than the therapists can and are able to influence their children by teaching, shaping, and modeling. A growing literature supports including parents in group interventions, often in parallel parent groups and sometimes as a sole intervention. A University of Michigan group tested the effects of a child-only, child-plus-mother, and waiting-list intervention for children who had witnessed intimate partner violence. The children in the child-only group participated in a 10-week group therapy aimed at increasing their knowledge about families and family violence, managing their emotions, and managing interpersonal conflict. The separate mothers’ group taught the mothers about the influence of violence on children’s development and worked to improve parenting and disciplinary skills. The child-plus-mother therapy was the most effective in improving externalizing behaviors, with gains continuing after the group ended. When sexually abused children aged 3 to 12 ($n = 617$) were treated using group therapy with a concurrent parent group, it was found that children who participated in the group improved significantly in their symptoms, which included somatic complaints, interpersonal difficulties, psychological state, and behavioral problems. These children were less likely to be revictimized or become sexual offenders, particularly when the parents also attended the group. Parents are an important resource for the reinforcement of skills and psychoeducation taught in groups.

Setting

Group therapies for children typically take place in the clinic or school. The advantages of clinic-based groups include confidentiality and coordination of care. The more available school setting offers the specific advantages of convenience for the children, opportunities for developing friendships with schoolmates who are accessible, and the potential for doing primary prevention work. Some social skill group therapies based in clinics have improved socializing with group members, but results have failed to generalize to school peers. Conducting social skills groups in the schools increases the chance for generalizability of skills. At the time of this writing, most school and clinic work was performed online due to COVID sequestration. Over time, the effectiveness of this distance therapy will be researched and more understood.

Therapist

The effective group psychotherapist understands child development, psychodynamic concepts, characteristics of the child-targeted psychopathology, and facility with the therapeutic intervention. In addition, the therapist needs knowledge of factors particular to groups. The therapist must foster a climate of security, tolerance, and acceptance for differences in feelings and behaviors; be able to control the tensions and anxiety levels in individual group members within acceptable limits; and control group phenomena such as bullying, scapegoating, monopolizing, and instigating. Groups benefit from therapists who are flexible and playful, can regress without becoming a peer, have the ability to provide structure in an authoritative (not authoritarian) manner, and provide a good co-therapy relationship. The group therapist must be aware of and contend with their countertransference to the group and its members. As with all countertransference reactions, the therapists' anxiety and resistance can be used to further understand the group process. By analyzing one's countertransference, one can alert oneself to possible enactments. Group resistance may reflect the therapist's own anxiety.

Co-Therapist Relationship

Having more than one therapist expands the capacity for simultaneously processing multiple members' reactions and allows more efficient administration of the group. Having a co-therapist also provides back-up in the event a leader is absent. The co-therapist relationship has been likened to a marriage or sibling relationship: It takes a cooperative, complementary partnership that must withstand multiple potential stressors. Often, co-therapists are individuals with unequal training and experience. The pair must withstand group members responding to them as a parenting duo and/or mentally assigning them good therapist/bad therapist roles. Throughout the therapy session, co-therapists must negotiate the dips and turns of members' (and their own) emotions in front of the group. Witnessing how co-therapists problem-solve together provides the group's child and adolescent members with an opportunity for modeling and recapitulation of the family. The co-therapist relationship requires a respectful, egalitarian stance in which therapists are open to processing their interaction and their reactions to the group. Unfortunately, due to declining reimbursements, co-therapy is becoming less feasible in community settings.

Tools

Creative group therapy tools range from simple games to video interventions and can facilitate the therapeutic process. One author described a low-tech approach—a variation of pass the parcel, in which a package is passed around the group for each member to remove a layer of wrapping and answer the question written on it.

Heiman and Ettin used a variety of interesting techniques in a therapy group for sexually abused girls. They began with a puppet-making exercise included to promote disclosure and build group cohesion. Girls built puppets in dyads and used them to introduce each other and as a supportive “friend.” Girls named the group in the second session, a task that began their working together as a group and set the stage for them to take ownership and define the goals of the group. Other tools included a board game that asked questions about specifics of abuse, a coloring book about girls’ rights and bodies, a collage project about girls at different ages used to facilitate a discussion of appropriate sexual behavior at different ages, letter writing to perpetrators to express strong feelings, and throwing into the trash drawings and writings symbolizing things they would like to put into the past.

A group for sexually abused boys made a group flag, which required creativity and compromises that led to group solidarity. This same group rolled a “feelings cube,” which directed them to describe a situation causing a certain feeling, act the feeling out, or answer questions posed by group leaders and members. They made and photographed clay figures portraying their abuse. Each of these tools was chosen to help the girls and boys confront and process difficult issues.

Videos can facilitate learning and discussion in groups for children. Prevention films for sexually abused children can be used to teach self-protection skills. Films of social vignettes can provide a take-off point for discussion and role playing for social skills groups. Watching the video stories together is an efficient teaching medium and provides a shared experience from which discussion can flow in a lively way.

Activities can also serve to introduce levity that promotes camaraderie and communication. In DBT groups, games are often used to enforce mindfulness techniques such as observing one’s feelings and radical acceptance. In one activity, members choose a favorite from a bowl of mixed candies. Then they are asked to pass the favorite to the person on their left and mindfully eat the new piece of candy. In a mentalization-based therapy technique, teens are encouraged to “play the doctor” using a stethoscope and dressing in white coats to empower themselves in discussing what they think may be going on in the minds and hearts of other family or group members. Similarly, playing freeze tag where participants “freeze” showing how they look when feeling a certain emotion can help teens understand the concept of the “opaque mind” and illustrates that they can only guess at what goes on in another’s mind. Role play is yet another effective tool to promote perspective taking and communication.

In the era of social media, creating original video recordings can be an exciting developmentally appropriate and available tool. For instance, to consolidate knowledge of information and skills acquired in one anxiety group, members chose to make an instructional video. In one report, therapists introduced a continuously recording video camera to a group to spark group interest when it was waning. They observed that the camera’s eye acted as both an additional group member and as a transference object.

Food, obviously symbolically and physically nurturing, may be included in group therapies for practical and therapeutic reasons. Children coming to the clinic after school may be hungry and feeding them may help to improve their attention. Some treatment protocols use a snack break to allow practice in a less structured socializing for socially phobic or impulsive children. Others use food as reinforcement for specific behaviors. The group therapist must decide whether the distraction caused by food is a reasonable cost for the gained therapeutic benefit.

Length of Intervention

The rate of improvement is greatest in early therapy sessions. With increased emphasis on cost containment and measuring outcome, most recent group therapies are scheduled to last only between 8 and 12 sessions. Therapeutic factors may also call for time-limited work. Formally dividing time into phases like introductory/group forming, working, and consolidation/termination phases may help with focus and momentum. This is not to say that longer interventions are not beneficial. For some children, more sessions are needed before there is significant improvement, still others benefit from more time for reinforcement of learning and perhaps taking on of leadership roles.

Group Rules

For individuals to feel safe in taking the risks necessary for change and for therapists and group members to serve multiple individuals' needs, groups need rules. Developing a set of rules is usually a group activity early in the intervention. Fairly universal rules include those regarding confidentiality, personal safety, attendance, not coming to the group while intoxicated, and respecting others. Deblinger notes that the use of praise, reflective listening, selective attention, positive reinforcement, and negative consequences such as time-out may improve focus in more unruly groups. While, at first glance, prohibiting socializing with other group members outside of sessions may seem unnecessary, it has proven consistently to be an important rule. For example, after members of a structured adolescent cognitive-behavioral group met with each other outside of the group, one member left the group, uncomfortable with feeling pursued by the other and others felt left out. Members of a trichotillomania group requested permission to use instant messaging on the internet to seek support from each other when they had urges to pull. Although most felt that this was helpful, some eventually desisted, feeling that boundaries were impinged on.

Family Involvement

Families may be involved in group therapies for children and adolescents in a variety of ways: they may be included with the children in the session (multifamily group), in parallel parent groups, with families included in parts of each session, certain sessions, or not included at all. Childhood anxiety disorder researchers have examined the role of family involvement in treating anxious children.

One study compared outcomes of group therapy in three conditions: parent-and-child parallel group therapies, a child-only intervention, and a parent-only intervention. Its authors found that, although all groups improved, those children in the parent-and-child parallel group intervention used new coping strategies more actively outside of the group. Barrett and coworkers included parents in group therapy for anxiety disorders because the parents were at increased risk for having anxiety disorders and because the therapists believed that family interactions play a major role in the development of children's anxiety. They randomly assigned 79 children to group CBT, group CBT with family involvement, and wait-list control. The family component encouraged parents to reinforce courageous behavior and ignore excessive complaining and anxious behavior and it taught them communication and problem-solving skills to deal successfully with their children's emotional upsets and anxieties. After treatment and 1 year later, those in the combined-treatment group had a better outcome than those in the child-only condition (95% diagnosis-free vs. 70%).

Similarly, in a recent multi-center randomized controlled trial by Scholten et al., parent involvement in group therapy for children with chronic illnesses did reveal additional positive effects on parent-reported internalizing, child-reported externalizing, information seeking, and social competence, when compared to the child-only arm. Parental involvement also appeared to contribute to durability of improvement.

Process in the Structured Group

Group therapies increasingly use structured and manualized treatment protocols. Critics and skeptics of manualized treatment fear that therapy from a manual will be stiff and "cookie-cutter." Kendall, who developed the first manual for treating childhood anxiety disorders, the Coping Cat, noted that the therapist cannot be a "cookbook" therapist but must have a sound understanding of the theory behind the intervention and adequate flexibility and creativity to use the tools in a way that suits the needs of each patient.

Manuals and outlines of therapy interventions inform and guide but do not control the group therapist. The therapist will lose effectiveness and opportunities to help if they ignore psychodynamic forces and other resistances to treatment. In an article on group therapy for sexual abuse patients, Heiman and Ettin stated that group dynamics and collective process do not disappear even within a highly structured psychoeducational group. If group members' affective reactions are not monitored and the collective dynamics are not acknowledged, then the group curriculum may have little impact. Especially in a group of adolescents, joining is crucial to promote learning. In some cases, the group process can act to disrupt the viability and stability of the group. For example, if a group member decides that a game or activity is silly, it may be difficult to have the rest of the group participate. At times, it may be necessary to pull such an individual out of the group to explore their reasons for disagreement or isolation and encourage rejoining. Group processes can be harnessed to maximize therapeutic impact. By responding to the messages underlying children's behaviors, rather than suppressing them, an opportunity exists to respond and transform the function of children's actions and maximize the efficacy of structured activities.

CLINICAL ISSUES

Indications

Group therapy should be considered if the child or adolescent has a condition for which there is evidence that group therapy is effective. This subsection provides descriptions of empirically supported group therapy interventions for specific pediatric populations. Rather than being an exhaustive review of all reported studies, it presents examples of treatments with support to describe the treatment elements and factors that made group therapy the choice. For some children, group therapy alone may be sufficient, whereas others require concomitant pharmacologic, individual, or family intervention.

Addiction, Substance Abuse. Many interventions for substance abuse are in use, many with little empirical support. Exposure to a credible peer: the type of peer who could be a negative influence under other circumstances, but, when seen in the therapy where each has a goal of not using may be helpful through positive modeling, positive reinforcement, and an alternative peer group.

The mutual self-help groups, Alcoholics Anonymous (AA) and Narcotics Anonymous (NA) may be a cost-effective option for supporting treatment gains for some substance-abusing adolescents; one study reported a decrement in health care costs over 7 years of 4.7% for each session attended! While a number of researched treatment programs have up to moderate effect sizes, the rate of relapse is high. Participation in AA and NA improves abstinence and recovery in a dose-dependent way. Research suggests that general group elements, including meeting others with similar troubles, learning from others' experiences, receiving encouragement, hope and support are powerful contributing factors. In a review, Nash notes that initially, adolescents gain most from participation with same-age peers, with peers improving the likelihood of engagement. Later, adolescents have been found to improve more in groups that also include older members. Research has elucidated particular potential barriers for adolescent participation in 12-step groups: adolescents may not see substance use as a problem due to short duration of use and the feeling that they do not need help. Adolescents' developmental need for autonomy may not mesh with accepting powerlessness against a problem; limited insight and enjoyment of substance use may not align with considering lifelong abstinence. Finding peer role models may be important to buy-in and continuing participation. For instance, Kelly and colleagues followed 160 adolescents at intervals for 8 years (84%) after an in-patient stay for substance abuse. They found that a high initial program participation dropped consistency over the 8 years. Those who attended most were with those with more episodes of alcohol intoxication and those who felt they had a problem with alcohol. Those attending most frequently during the first year had significantly better outcomes over the 8-year duration of the study. Those who perceived that they could use alcohol in moderation attended fewer groups and were consistently less abstinent. The data suggested that in adolescents with substance abuse, attendance at one or two AA meetings per week increased abstinence and that attending at a frequency of three times per week was associated with *complete* abstinence. The authors speculated that the mutual self-help group model succeeds because it may help maintain motivation, provide a new social context of sober peers, enhance coping skills and improve self-efficacy. Nash summarized, "Though not the only model for post-treatment recovery support, research to date suggests that similar to adults, adolescents' involvement in 12-step groups predicts improved AOD use outcomes, and greater participation (i.e., frequency, duration, and extent of involvement) predicts abstinence and SUD remission better than attendance alone."

With the internet being one of the most commonly used methods of communication and entertainment (and, in COVID sequestration times, school) internet "addictions" are emerging among adolescents. There have been growing concerns that online video game play may be addictive and invite people to escape from their problems or daily responsibilities. Snodgrass and coworkers sought to explore some of these concerns using the online game World of Warcraft (WoW), played by groups of people regularly, simultaneously, internationally, and day and night. They also investigated online gamers' motivation, looking at the factors of achievement, social connectivity, and immersion factors that tend to foster problematic play in the online worlds. "Achievement" refers to competitiveness drawing players deeper into game, where they spend more time online than they might wish. "Social connectiveness" refers to getting drawn into communities or associations that also demand more of their time. "Immersive play" refers to players who not only temporarily escape but more persistently or maladaptively avoid real-world problems. Researchers conducted 30 formal interviews and 252 people participated in a web survey. Results showed that subjects' online play could be a source of distress in their lives. They also showed that players lacking high levels of offline success (i.e., good job, high income, having a good social network of family and friends) may attempt to compensate by playing excessively and players *with* high levels of offline success are less likely to overplay. Players who experienced dissociation (losing track of the real world while immersed in the virtual world) were more likely to have problematic play. Playing multiplayer WoW to gain a sense of social connectiveness correlated with such dissociation and problematic play. While WoW group interaction can be problematic when players, who dissociate too fully and lose themselves in the virtual world of WoW, use it as a primary social outlet, thereby risking more problematic behaviors both online and offline.

Du and colleagues used a school-based, multimodal cognitive-behavioral group therapy to treat internet addiction in adolescents. Fifty-six adolescents aged 12 to 17 (with their parents and teachers) were randomly assigned to group therapy or a control group. Group CBT involved 6 to 10 students participating in an eight, 1.5-to-2-hour session group CBT run by two child and adolescent psychiatrists simultaneously with parent group therapy. Teachers attended workshops that included didactics, discussion of cases, and role plays. Groups focused on recognizing and managing one's feelings, fostering healthy family communication, strengthening relationship skills, using the internet in a healthier way, controlling one's impulses, recognizing when behavior is becoming addictive, and coming up with ways to stop addictive behavior. Parents also learned skills to improve emotional awareness and behavior, communication, problem-solving techniques, and managing adolescents with internet addiction. After the intervention, and 6 months later, the intervention group demonstrated better time management skills. As well, the treatment group showed large effect sizes for improvement in anxiety which is important, as anxiety is often associated with internet addiction and emotion and behavior regulation.

Aggression/Conduct Disorder. Children who are aggressive and have conduct disorder have social difficulties, especially in frustrating circumstances. Concerned that parent training for oppositional defiant and conduct disorder did not help all families or generalize to school, Wester-Stratton and coworkers designed a social skills and problem-solving training protocol (Incredible Years: Dinosaur Curriculum) for children between the ages of 4 and 8 years. The investigators used the group as a supervised place to practice social skills. Ninety-nine children with the diagnosis of Oppositional Defiant or Conduct Disorder were randomly assigned to a control condition or a group intervention that included directly taught social skills, problem solving, and anger management. Groups of five or six children met weekly for approximately 6 months. Children in the treatment condition watched videotaped vignettes of children coping with stressful situations. They discussed the vignettes and collaborated to practice a repertoire of acceptable solutions and coping skills. They also discussed feelings they had about troubles caused by their own behaviors. To enhance motivation, attention, and learning and provide a developmentally appropriate milieu, therapists used puppets, cue cards, coloring books, small reinforcements, group activities, and games. The treatment group demonstrated more significant clinical improvement in conduct problems as rated by parents and teachers. Those gains were persisted at 1-year follow-up. The authors noted that many of the children made their first good friends in the group therapy and postulated that intervention might be even more effective if delivered in the subjects' actual peer setting.

In response to a paucity of resources for treating conduct disorder in South Africa and in the hope that group cognitive-behavioral treatment could help children and adolescents with conduct disorder to respond more appropriately to social situations, a group of investigators designed and studied a 12-week intervention program. An occupational therapist and a psychology student were trained and supervised by an experienced adolescent group therapist. The sessions included: (1) an overview of rules, expectations, and rewards; (2) teaching problem solving/decision making; (3) impulse control, response gap thinking; (4) assertiveness training; (5) requesting behavior change; (6) reward session; (7) empathy, perspective taking; (8) sex differences, interactions, relationships; (9)

sex education video, relationships; (10) norms, values, role models; (11) giving compliments; and (12) closure. Parents of treated and control subjects were invited to six sessions of parent behavior management training. The treatment group showed significant improvement in defiance, physical aggression, and delinquent aggression posttreatment and at 6 months, whereas the control group demonstrated improvement only in defiance and a remarkable increase in truancy. The authors noted that this brief intervention using community workers showed promise as a sustainable intervention.

Attention-Deficit/Hyperactivity Disorder. Although the Multimodal Treatment Study demonstrated that stimulant medication is the treatment of choice for uncomplicated attention-deficit/hyperactivity disorder (ADHD), children with greater difficulties may require additional psychosocial interventions to optimize function. Willis and colleagues summarized that socially, children with ADHD are often disruptive and/or developmentally inappropriate, act impulsively and in a self-centered way, and are less sharing, turn-taking, and cooperative than other children. They also demonstrate deficits in social cognition, perspective taking, and social problem solving—all of which may alienate peers, who may see them as immature, intrusive, and/or inept. Moreover, children with ADHD do not realize these things; peer relationships may represent their greatest impairments.

Hinshaw and coworkers designed a CBT intervention for boys with hyperactivity, including peers for practicing social skills. An initial study compared outcomes for a group CBT intervention for boys between the ages of 8 and 13 years treated with methylphenidate or placebo. A second study compared the treatment on a daily basis in the context of an intense summer program with a psychoeducational intervention. In each, the boys were first taught problem-solving strategies: (1) define the goal or problem, (2) generate multiple solutions, (3) consider the outcome of the various solutions, and (4) choose and operationalize a behavior. In groups of 3, the boys first practiced applying these problem-solving strategies to a series of academic and fine-motor tasks; then they applied them to interpersonal problem vignettes. Practicing recognizing signs of anger, rehearsing strategies for dealing with provocation, and then maintaining self-control while experiencing planned provocation in the form of teasing from group peers and group leaders improved significantly in the group intervention and with medication, significantly more than the psychoeducation only subset.

Post hoc analysis suggested that children who had more experience seeing peers model coping behaviors in the group benefited most.

Pfiffner and McBurnett randomized 27 children with ADHD to receive an 8-week group social skills training—with either a parent group and teacher feedback, or wait-list control. A group format was selected so children could practice newly learned social skills and receive social reinforcement from peers. Children reported on their behavior points earned during the week, pooled their points for a group pizza party, role-played specific social skills with immediate peer feedback on performance, and shared fun activities. Evaluation after several months supported the utility of social skills training, with or without parents' and teachers' involvement. The authors noted that strong group reinforcement in a stimulating, developmentally appropriate program led to significant improvements in social skills and problem behavior.

Hoping to improve chronic management of ADHD, a larger (71 subjects in 3 clinics completing) controlled study compared a monthly group therapy intervention with a computerized clinician decision support system, the Child Health Improvement through Computer Automation (CHICA). CHICA promotes clinicians' adherence to clinical care guidelines for ADHD. Up to nine families per group participated in a 60-minute group discussion followed by a 30-minute medication visit, with parents and children in separate, simultaneous sessions, scheduled at the end of the day. The authors randomized treatments to sites, acknowledging the system-level changes necessary for personnel and scheduling. The parent group intervention included psychoeducation about ADHD, educational advocacy, and positive parenting strategies; the child intervention also included psychoeducation and reflection on ADHD’s impact, and what they saw as strengths and challenges for themselves. Compared to those in individual therapy with therapist participation in CHICA, families in the group improved in measured outcomes: they attended more visits over 12 months, experienced better parental emotional health, had fewer challenges related to misbehavior, reported significant improvements in child functioning at home and more time for themselves, other siblings, and routine household activities. Unfortunately, children receiving individual intervention reported a significant drop in mean emotional health. Bauer and colleagues, in this extension of earlier work, concluded that, especially with nearly half of children with ADHD being managed in primary care, group intervention in the familiar primary care clinic may carry less stigma and serve as an effective adjunct to competent medication management, filling a gap between primary and specific mental health care services.

Willis and colleagues, in a recent review of studies of stand-alone social skills training for children with ADHD, while expressing concern about the degree of current studies’ rigor, still suggested that parent training is beneficial. They noted that the most powerful elements included emphasizing to parents that they, as parents, have important roles as their children’s social skill educators and peer relationships coaches. Reviewers also endorsed the concept of simultaneous parallel parent and child sessions and recommended maximizing opportunities for positive peer interactions, even making friends. The review suggested that targeting the child’s natural peer group to increase inclusiveness of the child outside the group setting (such as at school) would help.

Anxiety. Avoidance, social isolation, and difficulty with social skills are significant troubles for most anxious children and adolescents. An opportunity for direct exposure to peers, practicing social interaction, and receiving social support and reinforcement make group therapy a theoretically attractive treatment modality for anxious youth.

In 1998, Barrett reported the results of a CBT group for anxious children and their families. The manual-based treatment focused on providing children and families with skills in managing anxiety and avoidance, problem-solving, and mutual family support. Posttreatment and 12 months later, 65% of children in the child-only group, 85% of those in the group with family involvement, and 25% of the wait-list control group were free of diagnosis. Silverman and coworkers also compared a group intervention for anxious children and their parents to a wait-list control group. They considered the disadvantages of group therapy—dampening of the therapist–child relationship, overly taxed therapist resources, and possible negative modeling—along with possible advantages. Sixty-four percent of the treated subjects versus 13% of the wait-listed subjects were free of their primary diagnosis posttreatment. At 6 months, 79% of the treated group no longer had their primary anxiety diagnosis.

Beidel compared a combined 12-week group and individual treatment for latency-age social anxiety disorder with an active control condition—the Test Busters study skills protocol. The treatment condition consisted of child and parent education about social anxiety, social skills training, peer generalization experiences, and in vivo exposure. In the group social-skills training component, children learned about greetings, starting conversations, maintaining conversations, listening and remembering skills, skills for joining groups, positive assertion, negative assertion, telephone skills, using instruction, modeling, behavioral rehearsal, corrective feedback, and homework. Immediately after the social skills training session, the anxious children joined a group of not-anxious peers for a 90-minute “fun” unstructured group activity (such as bowling or in-line skating). Each week, the children met individually with the therapist to engage in exposure activities, practicing a single task, such as reading aloud in front of others. In this treatment, investigators thoughtfully combined two forms of group intervention and an individual session. The first part allowed for economic use of resources for teaching and an opportunity to practice skills in a structured session. The second part, an unstructured group activity, included non-affected peers and provided an opportunity to expand on and generalize the skills in a more challenging situation. The individual exposure session allowed for customized, individual, symptom-driven intervention. At the end of the group, 67% of those in the treatment condition and 5% in the study skills condition no longer met diagnostic criteria for social anxiety disorder—with gains maintained at 6 months.

A Cochrane Database of Systematic Reviews of CBT for anxiety disorders in children and adolescents in 2015 reviewed 41 studies of adequate quality with a total of 1,806 participants. It concluded that CBT is significantly more effective (60%) than no therapy (waiting list, 16%) in reducing symptoms of anxiety in children and young people. Group, individual, with and without parent therapy formats were indistinguishable in their outcome. CBT was found no more effective than other "active therapies" such as psychoeducation, other supportive interventions, bibliotherapy, and treatment as usual. This suggests that group therapy for childhood anxiety is as effective as other active interventions, and it may offer the advantage of lower cost. It also points to the need for a better understanding of what makes treatment effects more durable.

Bereavement. Universality may be an especially powerful therapeutic factor in groups gathered to deal with loss. Pfeffer and coworkers used manual-based group therapy to assist children after the suicide of a parent or sibling. Families of children between the ages of 6 and 15 years were contacted within 3 weeks of the suicide death of a parent or sibling and were randomly assigned to receive or not receive intervention. Children with psychiatric diagnoses were excluded. The intervention consisted of ten 1.5-hour weekly group sessions of two to five children within 3 years of age of each other, with simultaneous parent groups. The children's curriculum consisted of offering psychoeducation about death and its permanence, identifying feelings of grief, defining what suicide is, and discussing why people commit suicide. Children's groups discussed preventing suicidal urges and presented problem-solving skills. The therapy facilitated expressions of grief and promoted identification with the positive aspects of the deceased. Parents learned about childhood bereavement, learned to help their children express their grief and discuss the suicide, learned to identify morbid reactions in the children, and learned to promote children's healthy functioning. The parents' grieving was also supported. Children in the therapy developed new supportive relationships. They showed greater reductions in anxiety and depression than did the nonintervention children.

Depression. Group therapy is among the most empirically supported therapies for adolescent depression. A large study compared group CBT, group CBT with families, and a wait-list control group and three intensities of follow-up care for depressed adolescents. One hundred and twenty-three adolescents with major depression were randomly assigned to 16 sessions of group CBT, group CBT, plus nine sessions of a parent group or a waitlist. Mixed-gender groups of up to 10 adolescents met for 2 hours twice weekly for 8 weeks. The therapy used a manual and workbooks and taught skills of mood monitoring, improving social skills, increasing pleasant activities, decreasing anxiety, and depressogenic thoughts, and improving communication and conflict resolution. In the parent condition, parents reviewed the adolescent course content and met once per week and twice jointly with the adolescents. Posttreatment, 67% of treated and 48% of wait-listed subjects were diagnosis free. Neither parental involvement nor booster sessions made a significant difference in the measured outcome.

Hypothesizing that treating depressed Latino youth in groups would be successful, researchers recently compared group and individual CBT and interpersonal therapies (IPTs). Authors considered that Puerto Rican cultural value placed on *personalismo* (preference for personal contact in social situations) and *familismo* (placing the interest of the family over the individual) would be served in the group therapy format. Analysis of outcome favored CBT over IPT, which may have resulted from its increased structure and directiveness (consistent with *respeto*) and found results to be similar in the individual and group formats.

Stallard and coworkers designed a clinical and cost-effective study of classroom-based CBT for depression in high-risk adolescents. Of the 5,030 students that were asked to participate, 1,064 adolescents aged 12 to 16 years old were considered high risk and were randomized to a CBT group, the usual school curriculum: Personal, Social, and Health Education (PSHE) delivered by teachers, or an attention control group (the usual school PSHE with two additional facilitators). This study examined whether these preventative interventions for depressive symptoms could be both clinically effective and cost-effective. All three groups showed decreased depression scores with no significant differences between groups. Because it required trained therapists, classroom-based CBT was not shown to be cost-effective compared to the teacher-administered groups. The study did suggest that a preventive school-based group intervention could decrease adolescents' degree of depressive symptoms.

Internet-based interventions may be useful in decreasing social isolation, which may contribute to depression. Van der Zanden et al. studied a treatment that involved groups of six or fewer people aged 16 to 25 in 90-minute chat-room therapy sessions. Participants were randomly assigned to this online cognitive-behavioral group course known as Master Your Mood (Grip op Je Dip) and a wait-list control group. The therapy focused on restructuring thinking patterns by recognizing one's unproductive, unrealistic thoughts and transforming them into more helpful, realistic thoughts. Performance of daily activities and mood monitoring were also done. Reductions in depression and anxiety and an increased sense of mastery were seen after intervention and at 3-month and 6-month follow-ups.

A functional MRI study by Straub and colleagues found alterations in the strengths of brain area connectivity after only 5 sessions of group psychotherapy for 19 depressed adolescents. Compared to controls, after therapy, depressed adolescents demonstrated a change in functional connectivity consistent with their clinical improvement; conductivity changes were consistent with decreased rumination and self-referential processing and increased cognitive control.

Suicidality, Self-Harm, Borderline Personality Disorder Traits. DBT is the first well-established, empirically supported treatment for decreasing repeated suicide attempts and self-harm in youths.

Two independent randomized controlled trials with diverse samples of 250 self-harming youths have demonstrated that DBT is effective in reducing self-harm in adolescents. Recently, McCauley and colleagues in Washington State, California, Pennsylvania, and Texas compared two manualized therapies: Dialectical Behavioral Therapy and Individual and Group Supportive Therapy (IGST), each of which employ individual and group elements, with adolescents at high risk for suicide. DBT aims to teach patients skills and encourages the employment of those skills. The DBT intervention is comprised of weekly multifamily group meetings and daily patient diary cards for teaching DBT skills, weekly individual sessions to reinforce engagement and skill use, up to seven family sessions, weekly therapist team consultation, and 24-hour brief crisis coaching. DBT uses mindfulness and CBT skills to enhance emotion regulation, distress tolerance, and building a life worth living. IGST is a manualized supportive therapy that considers “thwarted belongingness” as a suicide risk factor; IGST emphasizes acceptance, validation, and feelings of connectedness and belonging. IGST contained the same therapeutic intervention types (minus diary cards) as DBT.

173 adolescents who had attempted suicide, had self-harmed more than 3 times, rated high on suicidal ideation, or had more than 3 diagnostic criteria for borderline personality were randomized to DBT or IGST, then treated for 6 months, and followed up for 1 year. The authors postulated that the DBT treatment’s greater family involvement, with all learning coping skills, may have been responsible for its advantage.

Eating Disorders. Outpatient group therapy for eating disorders (EDs) has hardly been examined. Heinicke and colleagues did complete a study of manualized group therapy for girls who identified themselves as having disordered eating, using a self-help manual, My Body, My Life: Body Image Program for Adolescent Girls, in which subjects were asked to read and complete contained activities each session. The intervention consisted of six weekly group sessions facilitated by a trained therapist and postintervention sessions at 2-month and 6-month follow-ups. Seventy-three adolescent girls were randomized to the intervention group or delayed treatment. The 28 girls who completed the intervention showed improvements in body dissatisfaction, disordered eating, and depression postintervention and at 2-month and 6-month follow-ups.

Many individuals with EDs participate in group therapies in the context of residential ED treatment programs, which generally have large effect sizes on ED pathology. Group therapies may include daily process groups, topic-focused groups on ED symptoms, creative arts therapy, nutritional counseling, trauma, and substance use-focused therapy groups, group exposure for food, body image (e.g., mirror exposure), and co-occurring emotional problems. Detangling the effects of group therapy from individual therapy and other aspects of residential treatment seems very difficult. Thompson-Brenner and colleagues worked to implement the Unified Protocol for Transdiagnostic Treatment of Emotional Disorders (UPTTED) at 2 residential and 14 partial hospitalization sites, following 616 patients (mean age 26 +/– 11 years) with anorexia nervosa, bulimia nervosa, and ED not otherwise specified over 3 years. Transdiagnostic approach elements include motivation enhancement, mindfulness and emotion awareness, cognitive appraisal and reappraisal, addressing maladaptive avoidance, developing interoceptive awareness, and emotion exposures. After the implementation of UPTTED, individuals improved significantly in experiential avoidance, anxiety sensitivity, and mindfulness. Groups assessed before and after UPTTED was instituted had large positive effects on depression and ED symptoms at discharge, with postimplementation UTM patients showing larger reductions in ED and depressive symptoms and pre-UTTED patients worsening at 6 months follow-up. Length of stay had a significant positive effect on symptoms, experiential avoidance, anxiety sensitivity, and mindfulness. This study adds to the literature in its effort to examine the effects of residential treatment in a systematic way and to address psychopathology using an accepted transdiagnostic psychotherapy.

Gender Identity Issues. Choosing group therapy to address shame and isolation, Rosenberg used supportive group therapy as a component of treatment for children with gender identity issues. In a case report, she described her regimen of individual plus group treatment for working with 12 children and their parents. All but one child had comorbid symptoms of anxiety, inattention, and social isolation. Each of the children had been taunted and derided. The child group component of treatment was proposed to help self-esteem through the experience of universality. In the parent group, parents expressed relief to talk openly, for the first time, about their children. They shared their experiences and fears that their children might be harmed physically, emotionally, or by sexually transmitted illnesses. They were coached on ways to interact assertively with school personnel, friends, families, and neighbors and on how to access resources. The author reported that the children made significant clinical improvements in comorbid symptoms. The author of this 2002 study noted, “None of these children abandoned their stereotypically cross-gender interests, but they added gender-syntonic interests. All of the children in treatment for several months or more begin to express acceptance of their biologic gender.” A recent World Professional Association for Transgender Health Standards of Care notes that 6% to 27% of prepubertal children who were referred to clinics for assessment of gender dysphoria had dysphoria that persisted into adulthood.

Menvielle and Tuerk also reported the benefits of a group for parents of non–gender-conforming children. In the monthly parent support group, therapists focused on child themes and parent themes. Discussions focused on reducing the polarity of gender-based interests (girl stuff vs. boy stuff), communicating acceptance of the child’s thoughts and feelings about gender and interests, setting limits on behaviors that invite trouble, facilitating peer relationships, dealing with bullying, and discussing the possibility that the child will be homosexual as an adult. Parent-focused themes included helping parents to grieve, developing coping mechanisms for dealing with embarrassing situations, and learning to intervene when necessary in social and school situations. The authors noted that the group helped to destigmatize the nonconformity for parents and that humor helped the group to build cohesiveness and reduce tension, elicited compassion, and provided insights.

Improving Adherence to Medical Regimens in Chronic Illness. Parents of children with chronic illnesses must deal with many demands: attending to their sick children, dealing with insurance and the health care system, continuing to work their jobs, caring for the rest of their families, and struggling with their own emotions. In New Zealand, investigators developed and tested a Strong Parents–Strong Children 6-week group intervention to help parents to manage these stresses and learn practical skills for accessing services. Parents learned to plan time for themselves, address catastrophizing cognitive distortions, and use relaxation techniques. They practiced strategies for obtaining and sharing information from relevant health professionals and maintaining family life and social contacts with the goal of handing off aspects of illness management to the child when developmentally appropriate. Parents who participated in the group perceived themselves as more effective and more able to manage stress and perceived the family as more cohesive, less anxious, and better able to cope than those on the waiting list.

For some adolescents, adherence to life-saving medical regimens is poor. Some preliminary studies suggest that treating adolescents in groups together with parent or “treatment buddy” involvement may help. Viner and colleagues compared a measure of diabetes control (hemoglobin A1c) in adolescents who elected group therapy with that in adolescents who were assessed but declined therapy. The intervention consisted of motivational enhancement therapy, analysis of strategies that help (solution-focused therapy), and a parallel parent group. Those treated in group therapy improved in glucose. A “treatment buddy” model also led to improved antiviral medication regimen adherence in a pilot of a group intervention involving adolescents; nearly all participants increased adherence to medication, as well as other important health behaviors.

Obsessive Compulsive Disorder. Recent years have seen the development of a number of centers offering intensive group-based CBT with family involvement for children with OCD. An intensive outpatient program for children with OCD and social anxiety disorders at McLean engages children in group psychoeducation, unstructured group snack time, and group exposure therapy, where participants work on their particular exposures 4 days per week. Other weekly sessions include exposure therapy in a public setting, a caregiver guidance group, and a multifamily group. Six patients participated per group meeting for a minimum of 4 weeks, after which children and their parents reported that children's functioning had improved at school, at home, and in social environments.

Youth with severe OCD are also treated within residential treatment programs. In one, Rogers Memorial, patients participate in 27 hours of exposure therapy per week alone and with staff, with process groups and school on weekdays, and experiential therapy groups several times per week. An examination of outcomes found that participation in the program (average 78 days) was associated with improvement in OCD and depressive symptoms. Anecdotally, a number of participants have reported that meeting and living with other adolescents with OCD was significantly helpful to them.

Posttraumatic Stress Disorder/Sexual Abuse. Group therapies are widely used to treat individuals who have been traumatized by molestation or other events. In an Indonesian community plagued by poverty and a decade of armed violence between religious groups, Tol and colleagues introduced a 15-session school-based intervention administered in schools by volunteers with no formal mental health training. The manualized CBT protocol included 18 group sessions, with an individual pull-out session to establish a stimulus hierarchy at week 10. Elements included an overview, information gathering, group rules, relaxation training, anger management through problem-solving, examination of self-talk and cognitive restructuring around PTSD, the introduction of a narrative about the trauma, imaginal and in vivo exposure, confrontation of PTSD-induced dysfunctional beliefs, summary, and graduation party. Subjects treated with trauma-processing activities, cooperative play, and creative-expressive elements improved and maintained improvement over 6 months in measurements of posttraumatic stress disorder (PTSD) symptoms and hope in the setting of continued environmental conflict and stress. Girls showed greater improvement than boys. The authors concluded that a classroom-based intervention could be an appropriate approach to target large numbers of children with stress-related symptoms, in which the start dates but not the psychometric assessment dates were staggered. The authors hoped that such time-limited manualized treatment might help in situations of severe natural or human traumas.

British investigators Trowell and coworkers randomly assigned 71 symptomatic, sexually abused girls between the ages of 6 and 14 to participate in either 30 sessions of manualized, focused, weekly individual psychoanalytic therapy or 18 sessions of manualized psychoeducational and psychotherapeutic group psychotherapy (hours spent in therapy were comparable). In the individual psychoanalytic therapy condition, play materials were provided, and topics to be addressed were designated in the manual, but the topics were addressed as they arose, and each was covered. The group topics were prearranged with notebooks, task sheets, and play material. Time was also allowed to address the relationships between the girls and their relationships with the co-therapists, and this was linked to past and current relationships, losses, and disruptions. The outcome was measured posttreatment at 1 and 2 years. Individually treated subjects had significantly greater improvements in PTSD symptoms. Each treatment led to improvement in general functioning and in anxiety and depressive symptoms.

Deblinger and coworkers performed a randomized comparison study of group supportive versus group CBT for young children who had been sexually abused and their mothers. The interventions aimed to improve the adjustment of the children by helping them to learn to communicate their feelings, identify inappropriate touching, and learn abuse response skills. In the children’s support group, the material was presented in a didactic format. The children’s CBT format was interactive, using a workbook, role play, behavioral rehearsal, and joint parent–child exercises. In groups parallel to the children’s, the parents were randomly assigned to cognitive-behavioral group therapy or support groups. In the CBT condition, parents were assisted in coping with their emotional reactions to the abuse and their children, taught about effective communication with their children in general and about sexual abuse, and provided behavior management skills. The support group was run with a manual guiding the therapist with skills of active listening, unconditional positive regard, reflecting feelings and empathy, and providing information about group dynamics. The magnitudes of effect sizes for changes posttreatment and at 3 months for children and mothers who participated in the CBT groups were large on most measures. For the support group subjects, the effect sizes were medium. Parents in CBT expressed greater satisfaction with the therapy. Children in the cognitive-behavioral groups (which used interactive behavioral methods) demonstrated greater knowledge and retention of body safety skills.

In the specific context of child sexual abuse, group psychotherapy might be considered a “treatment of choice” because of its impact on the negative effects of sexual abuse: isolation, stigma, and self-esteem. The group format can foster a sense of belonging among peers and help overcome the element of secrecy, which is a common component of child sexual abuse. Group and individual treatments of sexually abused children and adolescents have now been found to be equally efficacious, finding that the adjusted expense of individual therapy is 64% more than group treatment formats. However, a 2016 Cochrane review of child therapies and adolescent therapies for PTSD with a variety of traumas found that the evidence for psychological therapies versus treatment, as usual, was low overall for longer-term outcome but found that the effectiveness of CBT in decreasing PTSD symptoms was moderate in the short term. The review found no difference in the change of symptom scores between group and individual therapy. As with most psychotherapies, well-designed studies with well-defined populations looking for moderators of change are needed.

Social Skills. Group treatment settings provide obvious opportunities for learning and practicing social skills. Children and adolescents with various disorders may have impaired ability to perceive important aspects of social intercourse, deficient knowledge of what to do to interact effectively, and/or ineffective means of being with others. Social skills have been reported to improve using therapy in groups for children and adolescents with autism spectrum disorder, aggression, or ADHD.

Kolko and coworkers described one model that illustrates the use of many child group therapy components. Targeting social perception, knowledge of social rules, and peer group behavior in hospitalized boys with ADHD and conduct disorders, compared outcomes of three times per week, 5-week group social skills training with social activity group participation. Fifty-six boys were assessed by self-report on assertiveness and loneliness, by staff reports on social problems and social acceptance by peers, and by peers on popularity. The subjects were rated on 15 role-plays for competence and on behavioral observations that detailed social involvement, gaze, physical space, appropriate voice loudness and inflection, using openers and compliments, making assertive requests, responses to provocations, and appropriate play. In each session, group rules were read, the last lesson was reviewed, and questions were asked to assess retention. Skills training was accomplished by reviewing and rating videotaped role plays of situations handled correctly and incorrectly. Incorrectly played situations were correctly role-played, discussed, and then rehearsed in the group. Points earned in the unit’s token economy were recorded in the group. Outside the group, a “Secret Agent” awarded stars when the group members were caught appropriately putting the skill learned into action.

However, the Cochran Database in 2019 reviewed 25 randomized trials of social skills treatment programs involving 2,690 participants. Each study had methodologic issues. Analysis revealed that by teacher report, the main outcome variable, social skill training did not improve social skills, emotional competencies or general behavior compared no social skills training. The authors found no adverse effects of the therapy but did note possible improvement in ADHD in those treated.

A meta-analysis of group social skills interventions for children and adolescents with autism spectrum disorder noted that while many studies have been published, the formats, subjects, measures, and durations varied. The effect sizes of included randomized, controlled studies were small according to parents, observers, and teachers and large by self-report. Gates and colleagues observed that learning about particular skills and actually using them are two different things. They recommended that providers work to provide opportunities for practicing the skills in real-life situations.

The same year the above review was published, a randomized controlled study of a social skills group training was published. Olsson and colleagues recently reported the results of a randomized, controlled multisite (12 sites) study of 296 youth engaged in a 12-week intervention for youth with autism spectrum disorder. The manualized SSGT KONTAKT is described as follows: it “aims to improve social interaction and communication skills, social motivation, awareness of self and others, problem-solving capacities, and self-confidence. The program applies elements of cognitive-behavioral therapy (CBT), computer-based cognitive training, behavioral activation, psychoeducation, observational learning, and parent involvement through various mandatory, recurring, and variable treatment formats. Training sessions focus on understanding social rules and relationships, initiating social overtures, developing conversation skills, identifying and interpreting verbal and nonverbal social signals, managing conflicts, and developing social communication strategies. The teaching formats include individual goal identification, group discussions, social and role play, emotion-processing training, group activities, and tailored homework assignments aiming to facilitate the generalization of skills to everyday life. Elements such as group discussions and role plays are applied more among the adolescents in the variable parts of the training, whereas more playful elements are used in the children's group." Authors reported a small to moderate effect size in autism symptoms and adaptive function in female adolescents compared to standard care, according to parents unblinded to the treatment group. Authors urged that work to improve social skills training is needed, perhaps by adding strategies to improve motivation, focus, and developmental relevancy, along with a longer intervention, is important. The protocol encouraged flexibility within structured sessions. Flexible elements included social interaction games, role play, group discussion, group discussions, pantomime games, and computer-based affect recognition games focused on the recognition of emotions in human faces and eyes. In this pilot study, the youth showing the most improvement included those with higher IQs and higher initial levels of communication difficulties.

A meta-analysis of group social skills interventions for children and adolescents with autism spectrum disorder noted that while many studies have been published, the formats, subjects, measures, and durations varied. The effect sizes of included randomized, controlled studies were small according to parents, observers, and teachers and large by self-report. Gates and colleagues observed that learning about particular skills and actually using them are two different things. They recommended that providers work to provide opportunities for practicing the skills in real-life situations.

Terrorist Attacks. In the 9/11 terrorist attacks, at least 6,000 children lost a parent, and many others developed PTSD. Group psychotherapy served as a place where children could put their experiences into words and play, find support, and begin self-repair. Scheidlinger and Kahn made several observations during their group work with children after 9/11. They used play, drawings, and discussion questions with elementary school–aged children to help them to express their feelings and fears. Play was especially important for traumatized children aged 4 to 8 years, helping them to describe their experiences. Expressing feelings and thoughts in the group setting allowed other group members to share their experiences and relate to others, a significant step toward building trust and unity within the group. The clinicians observed that traumatic events can continue to affect processes of social and mental development, and they emphasized the importance of making effective interventions like group therapy available in events with global impact.

Haen described a two-phase group therapy intervention for children directly affected by 9/11. Many children felt alone and feared for their safety. It was imperative that a safe environment be provided for the children to minimize the sense of isolation they often felt. The first phase of group therapy aimed to develop trust, safety, and security and included work toward psychoeducation and connecting to one’s support system. The second phase was designed to provide an open forum to address and discuss distortions and misconceptions developed by the children after 9/11. When children blamed themselves for the loss of their parent(s), work focused on correcting cognitive distortions. To facilitate the second phase of group therapy, Haen used play, art, and other nonverbal tools to communicate feelings and thoughts. In one example, drama therapy allowed group members to role-play and alter endings to traumatic events. In some cases, they were able to prevent imagined disasters. These tools helped participants to accept and cope with losses and improved the perception of security and safety for young clients.

Limitations of Group Therapy

When nonbehavioral approaches such as nondirective and psychodynamic models are used, meta-analysis found that group therapies demonstrate a smaller effect size. In group therapy, the therapist is constrained in their ability to attend to specific individual troubles and has less flexibility in a crisis. Gathering sufficient group members may prove difficult for individual practitioners.

Training

Running group therapies requires more organization up front: Enlisting co-therapists and patients, locating facilities, establishing a mutually acceptable meeting time, obtaining materials for the participants, coordinating communications, obtaining insurance authorization, and much more. The cost of training therapists to administer empirically supported therapies (which is of course most desirable) can be large and present a barrier to adopting them in clinical practice. Crome and colleagues modeled analysis of costs of adding a CBT intervention using psychologists. Considering direct and indirect costs as well as the opportunity cost of lost income during training, they predicted that at the lower end, reading manuals and attending workshops might cost $9,000, while the upper level reached $55,000; the largest expense was the missed income during training. While the field wants to move toward using the most evidence-supported treatments available, authors expressed concern about scalability.

Complications

A patient’s ability to discuss emotional issues and be vulnerable practicing new skills in a group is dependent on their feeling safe and respected. When confidentiality is breached, when occurrences of disrespect are inadequately processed, or when individuals feel physically unsafe or threatened by interaction outside of the group, the group is counter-therapeutic. Adequate therapist training and supervision, careful screening of patients, and close attention to the group rules and process may guard against these occurrences.

Choice of Treatment

Each group therapy aims to ameliorate symptoms and promote psychosocial growth. The rationale for choosing to provide therapy in a group format must be that group therapy either has evidence to support its choice or has qualities that make it theoretically suited to address the patient’s individual problems.

ETHICAL ISSUES

As implied previously, group therapy, although possibly cost-effective, should not be prescribed with that sole consideration in mind. It is properly used as one tool of many in the armamentarium of psychotherapists when evidence supports its use or, lacking other more empirically supported therapy, theory provides a rationale.

RESEARCH AND EVALUATION

As delineated by the ACGME Program Requirements for Graduate Medical Education in Child and Adolescent Psychiatry ACGME-approved focused revision: June 13, 2020; effective July 1, 2020, those treatments possessing empirical support should be identified, developed, disseminated, and used. A bustling movement is growing to develop well-established treatments. Doubtless, group therapies for children and adolescents will be among those whose use will be supported, but as the reader may surmise, more work needs to be done to establish which interventions, and which elements of interventions, work best for whom in well-designed, multisite, controlled studies.

The author would like to acknowledge Drs, Sonal Rana and Nicole Quiterio for their contribution to this chapter in a previous edition and Abby Miller for her help in manuscript preparation.

FURTHER READINGS

ACGME Program Requirements for Graduate Medical Education in Child and Adolescent Psychiatry ACGME-approved focused revision: June 13, 2020; effective July 1, 2020.

Aronson S. Where the wild things are: the power and challenge of adolescent group work. *Mt Sinai J Med*. 2004;71:174–180.

Bauer N, Sullivan P, Szczepaniak D, et al. Attention Deficit-Hyperactivity Disorder Group visits improve parental emotional health and perceptions of child behavior. *J Dev Behav Pediatr*. 2018;39(6):461–470.

Billow RM. The therapist's anxiety and resistance to group therapy. *Int J Group Psychother*. 2001;51(2):225–242.

Bowman ND, Schultheiss D, Schumann C. "I'm attached, and I'm a good guy/gal!": how character attachment influences pro- and anti-social motivations to play massively multiplayer online role-playing games. *Cyberpsychol Behav Soc Netw*. 2012;15(3):169–174.

Choque-Olsson N, Flygare O, Coco C, et al. Social skills training for children and adolescents with autism spectrum disorder: a randomized controlled trial. *J Am Acad Child Adolesc Psychiatry*. 2017;56:585–592.

Crome E, Shaw J, Baillie A. Costs and returns on training investment for empirically supported psychological interventions. *Aust Health Rev*. 2017;41:82–88.

Du Y, Wenqing J, Vance A. Longer term effect of randomized, controlled group cognitive behavioral therapy for internet addiction in adolescent students in Shanghai. *Aust and New Zea J Psych*. 2010;44:129–134.

Flesichhaker C, Bohme R, Sixt B, Bruck C, Schneider C, Schulz E. Dialectical behavioral therapy for adolescents (DBT-A): a clinical trial for patients with suicidal and self-injurious behavior and borderline symptoms with one-year follow-up. *Child Adol Psych Ment Health*. 2011;5:3.

Freitag CM, Jensen K, Elsuni L, et al. Group-based cognitive behavioural psychotherapy for children and adolescents with ASD: the randomized, multicentre, controlled SOSTA-net trial. *Child Psychol Psychiatry*. 2016;57(5):596–605.

Gates JA, Kang E, Lerner MD. Efficacy of group social skills interventions for youth with autism spectrum disorder: a systematic review and meta-analysis. *Clin Psychol Rev*. 2017;52:164–181.

Gillies D, Maiocchi L, Bhandari AP, Taylor F, Gray C, O'Brien L. Psychological therapies for children and adolescents exposed to trauma. *Cochrane Database Syst Rev*. 2016;10(10):CD012371.

Haen C. Rebuilding security: group therapy with children affected by September 11. *Int J Group Psychother*. 2005;55:391–414.

James AC, James G, Cowdrey FA, Soler A, Choke A. Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane Database Syst Rev*. 2013;(6):CD004690.

Johnson DL, Gold SR. An empirical approach to issues of selection and evaluation in group therapy. *Int J Group Psychother*. 1971;21:456–469.

La Greca AM, Ehrenreich-May J, Mufson L, Chan S. Preventing adolescent social anxiety and depression and reducing peer victimization: intervention development and open trial. *Child Youth Care Forum*. 2016;45(6):905–926.

Leonard RC, Franklin ME, Wetterneck CT, et al. Residential treatment outcomes for adolescents with obsessive-compulsive disorder. *Psychother Res*. 2016;26(6):727–736.

McCauley E, Berk MS, Asarnow, JR, et al. Efficacy of Dialectical Behavior Therapy for adolescents at high risk for suicide: a randomized clinical trial. *JAMA Psychiatry*. 2018;75:777–785.

Nash AJ. The twelve steps and adolescent recovery: a concise review. *Subst Abuse*. 2020;14:1178221820904397.

Pollock KM, Kymissis P. The future of adolescent group therapy: an analysis of historical trends and current momentum. *J Child Adolesc Group Ther*. 2001;11:3–11.

Sperling J, Boger K, Potter M. The impact of intensive treatment for pediatric anxiety and obsessive-compulsive disorder on daily functioning. *Clin Child Psychol Psychiatry*. 2020;25(1):133–140.

Storebø OJ, Andersen ME, Skoog M, et al. Social skills training for attention deficit hyperactivity disorder (ADHD) in children aged 5 to 18 years. *Cochrane Database Syst Rev*. 2019;6(6):CD008223.

Straub J, Metzger CD, Plener PL, Koelch MG, Groen G, Abler B. Successful group psychotherapy of depression in adolescents alters fronto-limbic resting-state connectivity. *J Affect Disord*. 2017;209:135–139.

The World Professional Association for Transgender Standards of Care (SOC) for the Health of Transsexual, Transgender, and Gender Nonconforming People. Version 8.

Thienemann M, Moore P, Tompkins K. A parent-only group intervention for children with anxiety disorders: a multiple-baseline design study. *J Am Acad Child Adolesc Psychiatry*. 2006;45:37–46.

Thompson-Brenner T, Boswell JF, Espel-Huynh H, Brooks G, Lowe MR. Implementation of transdiagnostic treatment for emotional disorders in residential eating disorder programs: A preliminary pre-post evaluation. *Psychother Res*. 2019;29:1045–1061.

Wang LJ, Lu SF, Chong MY, et al. A family-oriented therapy program for youths with substance abuse: long-term outcomes related to relapse and academic or social status. *Neuropsychiatr Dis Treat*. 2016;12:699–706.

Webster-Stratton C, Reid J, Hammond M. Social skills and problem-solving training for children with early-onset conduct problems: who benefits? *J Child Psychol Psychiatry*. 2001;42:943–952.

Willis D, Siceloff ER, Morse, M, Neger E, Flory K. Stand-alone social skills training for youth with ADHD: a systematic review. *Clin Child Fam Psychol Rev*. 2019;22:348–366.

Yalom ID. *Inpatient Group Psychotherapy*. Basic Books; 1983.

53.4 Family Therapy

JOHN SARGENT, M.D., AND NEHA SHARMA, D.O.

Family therapy represents an organized, well-developed, and variable set of interactions with family members that addresses mental health and substance use problems that affect them. It also has been used to assist with recognized relationship problems, including communication difficulties and difficulties occurring within spouse and parent–child relationships. (In this section, “parent” refers to any adult with custody of the child and responsibility for the child’s well-being and development.) Tenets of family therapy can also be used to address conflicts among colleagues and family members including those that are created by implicit bias, racism, and discrimination.

These interventions take advantage of the importance of family relationships to individuals and the willingness of individuals to alter their behavior and thinking to help family members. Family therapy theory and practice are based on a range of ideas and ideologies (Table 53.4–1) and are oriented to respond to a wide range of problems. The therapist’s responsibility is to appreciate the nature of the problem, develop a shared appreciation of the problem with family members, and collaborate with the family to develop the best overall adaptation of family members, including ameliorating the manifestations and effects of the problem.

Table 53.4–1. Family Tasks as Child Development Unfolds

Infancy

- Ensuring parental availability
- Ensuring resources to enable parental caretaking
- Providing support to parents when stressed
- Providing education about infants and parenting
- Encouraging balance between jobs of adult and parent, especially during times of transition

Preschool Children

- Enabling parents to respond to separation/individuation efforts of the child
- Encouraging parental consistency in discipline and limit setting
- Supporting positive parent–child connections based on individual nature of the child and the child's competencies
- Encouraging play, use of fantasy, and creativity in the interactions between parents and child
- Encouraging child's relating ability with other children and with other caretaking adults

School-Age Children

- Ensuring parental consistency in understanding, accepting, and being available for the child
- Developing rule-based behavioral expectations and consequences for the child
- Encouraging the child's competent engagement with school, friendships, and community activities
- Gradually assisting the family in developing a consistent, positive, and hopeful appreciation of the child
- Encouraging the child's independent perception of and participation in family life

Adolescence

- Assisting the parents with adolescent problems that require rules, consequences, and consistency
- Encouraging parental figures in assuming a supervisory role with children, with parents willing to allow for experimentation and learning, although also establishing essential rules

Encouraging the adolescent’s voice and role in family life, as well as their competency and involvement in extrafamilial experiences
Assisting the parents in encouraging self-direction, self-definition, integrity, and values in adolescent children
Assisting parents with refocusing on important aspects of their lives as they separate from their children

Table 53.4–2.
Assumptions of Family Treatment

Family therapy is an integral part of the practice of child and adolescent psychiatry.
Child and adolescent emotional and behavioral difficulties always affect family life and relationships.
Family relationships always affect child development, emotional status, and behavioral capacity.
Adaptation of the child and adolescent with psychopathology to the family is always a goal of treatment and requires change on the part of the family, accommodation by the child when possible, and consideration by treating professionals.
The family can always be a source of understanding, support, and healing for children and adolescents.
The family should agree to, participate in, and assist in directing all mental health interventions at all levels of care.
Family involvement in all treatment should be defined and structured to achieve family goals and facilitate achieving goals for the child with psychiatric difficulties.
Various therapeutic interventions and classes of interventions can be used to achieve these goals.

Family interventions address problems that have neurophysiologic and genetic elements and have developed in relation to unique environmental occurrences. Interactional responses often naturally occur that reinforce the problem. Family members are also emotionally affected by their relatives' difficulties, leading to a further mismatch between family members' needs and others' behavior. A child may be identified as the patient while the family struggles with a poor marital relationship or with parental mental health concerns. Family therapy is much more interested in the solutions to problems and improvement in the overall well-being and functioning of the family than in the etiology of the problem. This leads to a future-oriented focus of treatment and also suggests that assessment is carried out in a manner that is engaging and therapeutic. Often, family members are frustrated and frightened, and often their interactions are unproductive and repetitive, reinforcing and maintaining the problems. Therapy occurs through the interventions of the therapist and the relationships they develop with family members and with the family as a whole. The therapist's behaviors are intentional and carefully chosen and carried out to facilitate engagement, mutual care, support, and a resolution of problems that the family is experiencing. The therapist also appreciates that improving family relationships and interactions often leads to improved overall functioning, which can become mutually reinforcing and self-perpetuating. Family therapy is a treatment approach strongly supporting social justice, equity, and opportunity for all family members.

This section focuses on the use of family assessment, family interventions, and family therapy to address the problems that families have with their children's development (see Table 53.4–1), emotions, cognitions, and behavior. It is based on the premise that the practice of child and adolescent psychiatry always involves interacting with and understanding family interactions and functioning, and always involves engaging the family in the treatment and monitoring process (Table 53.4–2). At the core of family work, there is acceptance of each family's cultural and social values. The section includes case examples demonstrating the range of applications of family therapy with diverse families, as well as the various forms that family treatment may take. Tables reviewing these significant models of family therapy (Table 53.4–3), the theoretical foundations of family therapy (Table 53.4–4), specific treatment approaches for common child mental health problems (Table 53.4–5), and unique family treatment approaches to differences in family composition, family history, and the culture and ethnicity of the family (Tables 53.4–6 and 53.4–7) are included. Family distress brought on by the effects of racism, discrimination, and exclusion is universal target of family treatment.

Table 53.4–3.
Varieties of Family Therapy

Psychodynamic

- Assumptions and Major Concepts
  - Family relationships are based on previous parent–child relationships of each parent.
  - Difficulties arise from developmental arrest, current interactions and projections, and current stressors.
  - Improvements develop through the process of family members gaining insight into problematic relationships from the past.
  - Therapy consists in shared exploration of past relationships and current distortions and role confusions.
  - Therapy also encourages the development of mutual appreciation of the needs and limits of family members.
- Major Theoreticians and Practitioners
  - Nathan Ackerman
  - James Framo
  - Ivan Boszamenyi-Nagy

Experiential/Existential

- Assumptions and Major Concepts
  - The family is a developing organism, and the goal of therapy is to encourage the growth of the family and its members through the process of therapy.
  - This process is best supported by encouraging expressiveness, openness, and communication within the family throughout the therapeutic encounter.
  - The family is responsible for its solutions; the therapist is a facilitator who encourages growth and confronts stagnation.
- Major Theoreticians and Practitioners
  - Carl Whitaker
  - Virginia Satir

Intergenerational/Bowenian

- Assumptions and Major Concepts
  - The family is an emotional relationship system, with the goal of relationships to promote engagement, foster differentiation, and avoid fusion.

In conflictual situations, engaging a third party (triangulation) stabilizes the family system.

Individuals marry partners at the same level of differentiation, leading, at low levels of differentiation in married couples, to severe personal and relationship difficulties.

Therapy is aimed at clarifying current relationships, promoting individuation, resolving current conflictual interactions, and working through intergenerational and introjected difficulties and patterns of behavior.

Major Theoretician and Practitioner

Murray Bowen

Structural

Assumptions and Major Concepts

Focused on current organization of the family as demonstrated through repeated patterns of interaction.

Particularly attentive to how the family's organization promotes or inhibits task performance, especially child rearing.

Analysis of interactions allows the therapist to determine the nature (distance vs. closeness) and the flexibility of family relationships in relation to task performance.

The therapist promotes interactions in sessions to determine family relationships and then intervenes in those interactions to introduce greater flexibility, alternative patterns and structures, and more effective task performance.

Enmeshment (ineffective closeness) and disengagement (excessive distance), as well as rapid fluctuations between extremes, are manifestations of dysfunctional family relationships when associated with significant symptoms in a family member.

This form of family therapy was developed to respond to child and adolescent emotional and behavioral problems and is particularly effective in those situations.

The therapist is both the director of therapeutic action and participant in the process of therapy.

Major Theoretician and Practitioner

- Salvador Minuchin

Systemic/Strategic

- Assumptions and Major Concepts
  - Symptomatic behavior is embedded within the patterns of communication and relationships in a family.
  - Ineffective attempts at solutions to problematic behavior become part of the pattern also and thus part of the problem.
  - The therapist is responsible for developing and prescribing tasks and behaviors that disrupt usual patterns of interaction while strengthening other, more productive patterns.
  - The therapist maintains a stance of neutrality and distance with respect to the family as whole and individual members.
  - Specific methods of questioning designed to elicit information about family relationships (circular questioning) are used; posing hypotheses, reframing, externalizing blame, and the use of positive connotation are common techniques.
- Major Theoreticians and Practitioners
  - Jay Haley
  - Cloé Madanes
  - Mara Selvini-Palazzoli and colleagues (Milan group)

Cognitive-Behavioral

- Assumptions and Major Concepts
  - Concepts of learning theory—conditional contingency, reinforcement, extinction—explain problematic behavior and identify therapeutic interventions.
  - Sessions are entirely problem- and solution-focused in the present.
  - Skills training is emphasized throughout treatment; homework, record keeping, and between-sessions tasks are used regularly.
  - Therapy is aimed at changing both cognitions and behavior and uses new cognitions among family members to direct and evaluate behaviors.

Particularly useful for managing mild to moderate disruptive behaviors in children and adolescents, focusing treatment on parent skills training.

Major Theoreticians and Practitioners

- Gerald Patterson
- James Alexander

Table 53.4–4.
The Theoretical Foundations of Family Therapy

General Systems Theory

The whole is greater than the sum of the parts.

Actions that affect one part of a system affect the whole.

Each part of a system is in a constantly changing relationship with all other parts.

Cybernetics

Interaction among family members is circular and continuous, with each behavior being the effect of one behavior and the cause of another.

Circular causality encourages attention to pattern of interactions and feedback between behaviors and relationships.

Communication Theory

All behavior is communication (both actions and words); behaviors define the nature of the relationships among family members.

Structural Theory

Relationships are structured in a hierarchical, potentially developmentally oriented fashion.

The structure of a family can only be determined by noting the patterns of interaction among family members.

Structures are functional or adequate insofar as they provide for the achievement of family goals.

Social Constructionist Theory

Each individual constructs or makes sense of their experience from their perspective.

Therefore, many different interpretations of a set of circumstances may be simultaneously true.

Families, therefore, must and do develop shared constructions of their situation.

The constructions may become unsatisfactory when problems or changes require interaction with systems (health care, school, extended family, community agency, etc.) that have alternative constructions of events.

Psychodynamic Theory

Early experience is formative in defining who people are.
The capacity for projection and projective identification can seriously inhibit the flexibility, resilience, and growth of relationships.

Attachment Theory

Affect regulation occurs within close relationships.
Secure attachment facilitates social relationships and the ability to know and communicate about mental states.

Over the last 20 years, manualized family therapy treatments have been developed and identified as evidence-based treatments through randomized, controlled trials. These treatment approaches have been developed for specific child and adolescent disorders, such as traumatic stress, substance abuse, anorexia nervosa, and conduct disorder.

PROCESS OF FAMILY THERAPY

Engagement

The key to successful family treatment is successful engagement of the therapist with the family and with family members. This process requires that the therapist possess the skills of connecting with more than one person at the same time and making each person feel understood, comfortable, and cared about. Engagement conveys the therapist’s hope for the family’s successful future and their belief that with effort, patience, mutual care, and concern, changes can occur and that a successful future can be realized. Family therapy in whatever fashion is always a creative act, an act of vision. The therapist’s connection with the family and its members is based on their assertion of the meaning and value of this family. Through the acceptance inherent in engagement, the therapist works to create a context of safety that promotes greater flexibility in the interactions of family members and greater possibility that the situation can be accepted and improved. In addition to warmth, curiosity, hope, and vision, the therapist must be able to hold and act on opposite ideas:

that the family possesses strengths and has difficulties, that the family is doing the best it can, and that the family needs to change, and that the family and its situation are unique and also resemble other families in their connections to each other, their concern for each other, and their wish that every other family member have the best life possible.

The therapist also engages with the family as an entity. By acting in a manner that reflects the therapist's belief in the integrity of the family, family members can begin to relax, see each other differently, and behave in a manner that reflects their interest in membership in their family and their commitment to its well-being. The therapist asks about the family's history, culture, rituals, and beliefs; respects what they hear; and treats the family as an entity in which they believe in and cares about. This helps the family members to come together, work together, and change the ways in which they treat each other. In doing this, family members begin to understand that relationships can improve, conflicts can be dissolved, and mutual understanding and mutual support can develop. Through successful engagement with individual family members and with the family as a whole, the therapist creates a solid foundation for family change and symptomatic improvement.

Safety

Throughout the engagement and assessment process the therapist's first responsibility is to ensure the safety of the child, the family members, and others in the community, as well as their safety. In some families aggressiveness is a part of their interactions, emotions may run high when disappointments, hurts, and conflicts have occurred, and family relationships may feel and at times be unsafe. Thus the therapist must be attentive to signs of the possibility of violence, aggressiveness, abuse, and suicidal behavior. If there is concern about safety, consideration of hospitalization or involvement of human services or the justice system must be considered. The therapist's role is to work with the family to decide together on the level of care necessary to promote the safety of all while also beginning treatment in a manner that ensures the safety and connection of all. The therapist creates a safe space where family members can speak to and listen to each other and where hurts and disappointments can be spoken about, and reconnections can be achieved. The therapist can establish ground rules for sessions and, if concerned, meet with family members individually to ensure a commitment to self-control and mutual respect in treatment and ultimately throughout family life. The quality of the therapist's engagement with the family and their presence, integrity, and his belief in their vision of successful resolution of problems are crucial in leading the family to the changes needed to ensure nonviolent family life. This sense of safety also leads to a space for advocacy when needed for issues of equity and justice.

Assessment

Family therapy is based on an ongoing assessment of the child's problem and family life. It is important for the therapist to recognize that the assessment is part of the therapeutic process. Assessment is completed in an interactive manner that is based on the therapist's interest in the family's strengths and individual capacities and interests, as well as the therapist's professional responsibility to appreciate the family's problems. The therapist, connected to the family in the treatment system, elicits a discussion about the questions the family has about their child and the concerns and issues the family experiences in their life.

Table 53.4–5.

Family-Oriented Treatment Approaches for Specific Clinical Problems

Psychosis

Ensure appropriate diagnostic evaluation, accurate diagnosis, and adequate antipsychotic treatment.

Psychoeducational treatment model is very helpful, providing education about the disorder, outlining family roles in treatment, and providing guidelines about how to respond to crisis and specific problems.

Efforts at reducing blaming and intensity of affective expression within the family are often helpful.

Ensuring family collaboration with the treatment approach, especially concerning communication with professionals, medication, vocational training, and work and placement, is extremely helpful.

Parent support is an important aspect of treatment.

Depression

Ensure that all involved recognize and accept the child’s depressed mood.

The therapeutic orientation should include attention to building attachments and connections, reducing personalized criticism, and building support among family members.

Building individual competence is essential.

Cognitive-behavioral strategies are most effective.

Setting limits to deal with acting out, personal irritability, and poor academic performance is often helpful.

Special attention should be paid to parental psychiatric problems.

Psychopharmacology should be integrated within the therapeutic approach.

Anxiety Disorders and Phobias

Collaboration in treatment reduces the power of the child’s fears and vulnerability.

Cognitive-behavioral approaches with the family are helpful.

Agreement about the expectations for the child, especially regarding fears, is necessary.

Distraction, relaxation, social skills assistance, and focus on building peer relationships all are helpful for the entire family to focus on.

Recognition of temperament and the possibility of recurrence of worries during stressful periods often aid long-term outcome for the family.

Disruptive Disorders

Attention-Deficit/Hyperactivity Disorder (ADHD)

Agreement about diagnosis is essential.

Education concerning the various aspects of ADHD is helpful.

Family involvement in measuring and monitoring the results of behavioral interventions and psychopharmacology is helpful.

Parent support groups can help long-term adaptation.

Specific plans for impulsivity, reactivity, and aggressiveness are helpful.

Independence, self-regulation, self-organization, and personal responsibility are goals of treatment.

Special attention should be directed to developing the strengths, competencies, and self-esteem of the ADHD child and the family as a whole.

Conduct Disorder

Collaboration among parents and all involved professionals and agencies is essential.

Responsibility and accountability for the child are primary goals of treatment.

Parameters of treatment must be explicitly stated.

Behavioral approaches highlighting consistency, expectable consequences, and predictable responses to transgressions are the cornerstone of treatment.

Plans must be made for responses to violence and other serious problems.

Impulsivity in treatment should be avoided; treatment must be consistent, continuous, developmentally oriented, and focused on building successful adaptation of the child, as well as on reduced behavior problems.

Medication is to be used for specific target symptoms, and its use and effectiveness must be supported by all participants in treatment.

Substance Abuse

Agreement about avoiding substance use among all involved adults in necessary.

Family tolerance for affective expression usually needs to be expanded.

The family must also attend to comorbid problems, depression, learning problems, ADHD, and so forth.

The role of consequences and limits for substance use must be expressed clearly.

Parental substance use also must be addressed in treatment.

Family participation in mutual support groups (Narcotics Anonymous, Alcoholics Anonymous) is often very helpful.

Peer problems, legal difficulties, and relapses should be addressed directly with the family.

Eating Disorders

Structural techniques build family cohesiveness and a consistent approach to the disturbed eating behavior.

Simultaneous support for individuation and connection is necessary.

Cognitive-behavioral techniques assist in establishing consistency in family relationships, as well as in responses to eating disorder symptoms.

Family therapy offers an opportunity for dealing with gender issues, as well as concerns associated with control and perfectionism.

The age of the eating-disordered individual defines how treatment is structured; early adolescents respond best to family-oriented treatment; older adolescents and young adults respond better to family therapy approaches that they arrange and manage in conjunction with simultaneous individual therapy.

Abuse (Physical or Sexual)

Safety of the child is paramount.

Parents should collaborate with the therapist and social service agency staff.

A clear role for parents in the treatment process should be outlined, with specific expectations and consequences defined.

Parent training and skills building are helpful, especially if the unique contributions of the child to the family situation are highlighted and the parents have the opportunity to practice responses and receive praise and support when they demonstrate competence.

Honesty, clarity, and directness often reduce defensiveness and evasiveness on the part of family members.

As much as possible, the child should have a developmentally appropriate, active role in treatment.

Encouragement of verbal expression can be helpful in affect regulation and maintenance of consistent responses to behaviors among family members.

Academic or Learning Problems

The child’s learning or academic problem must be defined to the satisfaction of parents and school professionals.

Parents assist with developing questions to be addressed through psychological, psychiatric, and educational evaluations.

Roles and responsibilities of parents, school professionals, mental health professionals, and the child should be clearly identified.

Specific educational and school-based goals should be set; these should be development of academic competency, self-esteem, and greater connection between the child and school.

Steady progress through school marks development and achievement and builds greater maturity.

Other difficulties influencing learning (e.g., depression, anxiety, ADHD, and other disorders) need to be addressed simultaneously.

Table 53.4–6.
Variations in Family Composition

Single-Parent Families

Recognize the strength, responsibility, burden, and stresses of the custodial parent.

Support the independent competence of both the parent and the child.

Encourage the parent's use of alternative caretakers and other modes of personal connection and support.

Encourage the child's exploration of their relationship with the noncustodial parent.

Assist with transitions that may occur with new relationships that the parent might develop.

Divorced Families

Encourage parental collaboration in divorcing and parenting processes.

Assist the independent emotional reactions of each family member to the separation and divorce process, including the recognition that this is a highly painful experience for children.

Encourage the development of effective separate relationships between each parent and each child.

Build strengths of sibling relationships.

Assist the competency and expression of each child.

Assist parents in collaborating in parenting tasks.

Grandparent- or Relative-Headed Families

Recognize the effort and strength of the child-rearing adult.

Encourage the child's emotional adaptation to not living with their biologic parents.

Encourage effective limit setting and adult affiliation with the child in these families.

Emphasize the child's competency and extrafamilial engagement and accomplishment.

Support consistency in involvement of biologic parents to the greatest extent possible.

Adoptive Families

- Recognize that the child’s problems may be associated with preadoptive phenomena.
- Support with sensitivity the parents’ emotional reaction to their child’s difficulties.
- Possible special attention to the child’s background and special attributes.
- Build connections among family members while dealing with the losses and separations associated with adoption.

Families with Same-Sex Parents

- Similar strategy as adoptive families
- Support parents to feel empowered and effective in their parental role.
- Guide parents to differentiate between their personal childhood experience and their child’s experience.
- Build connections in the community such that family members feel welcomed and accepted.

One can view the process of assessment in child-oriented family therapy as the iterative development of an imagined video of the child and family at home, their relationship with the child’s school, and their presence in their community. This understanding appreciates not only the child’s symptoms and abilities, but also the interactions within which those symptoms and strengths are embedded. Thus, symptoms—either those that are spontaneous or those that are reactive—are viewed within the interactions that precede them and those that follow them. One way to understand the patterns in which symptoms are embedded is to encourage interaction among family members in sessions and to be attentive to spontaneous interactions as they occur in sessions. Understanding these patterns provides the therapist with many possibilities for discussion and intervention.

Table 53.4–7.
Family Therapy Approaches: The Effects of Cultural Differences

Cultural Differences and Family Worldview

- Worldview defines the relationship of a family to people, institutions, and objects.
- Cultural experience affects a family's worldview.
- Race, religion, and traditional position within the power structure also affect a family's worldview.
- These factors influence family relationships, family roles, and expectations of parenting, child rearing, and child development.

Examples of Cultural Considerations in Family Treatment

- White families: focus on individuals, use verbal expression, and distinguish between physical and mental factors.
- African-American families: focus on actions and behavior; realize that extended family is important, as are religion and church; also realize that history of oppression and racism affects relationships with institutions.
- Hispanic families: acknowledge bilingual background with intrafamilial differences in acculturation and language use; religion and extended family are important; respect for the family's heritage and background is very helpful in ensuring engagement in treatment.
- Asian-American families: recognize specific national origin; family, privacy, and saving face are central to family life; recognize intrafamilial differences in acculturation and language use.

Attending to Cultural Issues in Family Therapy

- Basic Values
  - Any family is a microculture.
  - The family's culture is necessarily a synthesis and amalgam of cultures.
  - Problems develop when the family experiences conflict internally between the cultures of the extended family or with external cultures, institutions, or systems.

The family is responsible for defining itself and its culture and for developing the most successful adaptation with other cultures.

The therapist's responsibility is to support and facilitate that adaptation, recognizing (1) that there are many ways to be a family and raise children that are healthy and (2) that there may be issues that raise ethical concerns for the therapist, and in those situations further knowledge of the family's culture and consultation or supervision may be helpful.

Interactive Approaches

The therapist must assume a stance of respect for and acceptance of the family as it is.

The therapist should ask about the meaning of requesting mental health assistance for the family.

Honest curiosity expressed in open questioning about how the family operates and how the family's cultural background influences its operation is essential.

The therapist should refrain from judging the family, but rather should ask about how it achieves its goal.

The therapist should avoid behaviors based on assumptions about a particular cultural background.

The therapist should inquire about child-rearing practices in the family and responses to academic, emotional, behavioral, and social problems in the child as they reflect cultural beliefs and cultural expectations.

Clinical Approaches

Specific responses to the child's symptoms and assistance with parenting skill need to be embedded within the therapist's and the family's understanding of cultural beliefs and expectations.

Negotiation of the therapist's role and behavior should be based on the family's sense of their culture's expectations of professional behavior.

The therapist can help the family to understand their cultural background as a method of helping the family to define and understand themselves and clarify their approach to therapy.

Explorations of culture assist the family in assuming direction and responsibility for treatment.

Discussions about respective cultural backgrounds and what those backgrounds mean can assist in conflict resolution and enhance the development of the children.

The therapist can assist the family in making accommodating measures with schools and other institutions based on cultural considerations.

The family and the therapist can also decide together on other questions to be answered through the treatment process. Intellectual concerns, educational abilities, and vocational aptitudes all can be evaluated if questions exist. Diagnostic questions and the use of psychopharmacology also can be discussed in a collaborative fashion with the family. The guiding issues are always: What are your questions? What are your goals? What of these alternatives do you want to do? The key to understanding this process is to appreciate several key principles:

1. Assessment results are impressions and not judgments.
2. All assessments of the family are hypotheses based on limited data and thus are subject to change.
3. Because interactions can change and are largely dependent on context, the therapist holds the belief that if the context changes, the interactions can change.
4. Any intervention represents a question: will this intervention be successful in achieving the goal? If not, a different intervention is required.

The process of developing a shared opinion of both the child and the family, including an appreciation of strengths and difficulties among members of the treatment system, is another opportunity for therapeutic intervention. There may be opportunities to discuss and resolve disagreements between caregivers about the child’s capacities and problems. There may be opportunities, depending on developmental ability, for the child to participate in this discussion. The therapist supports the child's voice and expects the child to participate responsibly. As this process proceeds, the family's ability to negotiate effectively can be supported and reinforced by the therapist.

If other agencies or systems are involved with the family, especially in response to the child's problem, their representatives also will need to participate in the agreements about strengths and problems. This can be difficult, especially if the family has a history of conflict with these systems. The family may believe that a system with which they must be involved, such as the child's school, is treating the child ineffectively. The family will need the therapist's encouragement and potentially the therapist's coaching to be able to come to a collaborative approach with school personnel so that the child's education proceeds successfully. Similar negotiations will need to occur with the child protective system, the juvenile justice system, after-school programs, coaches, extended family, and church leaders, depending on who is involved with the child and what the child's needs, abilities, and problems are. At times the therapist will need to convene meetings of the family and those involved. The therapist, with the family's permission, may need to speak individually with the teacher, probation officer, and so on. These individuals may need to change their impressions of the child or family, and at times they may need to change the way in which they interact with the family. Supporting these new impressions is part of the therapist's responsibility for the success of their interventions. These changes become part of the therapeutic effort and goals for the treatment and become the focus of addressing implicit bias and social injustice when needed.

PROMOTING EFFECTIVE AND REWARDING PARENTING

Ensuring the safe and effective development of children is one of the main responsibilities of parents and caretakers. Caretakers need to recognize and provide structure to ensure an immature child’s ability to grow and explore their world in a way that supports developmental change and newly learned skills. They also need to provide the love and care that encourages the growth of self-confidence and self-esteem while also reinforcing developmental gains. Caretakers must also be observant to note situations in which children may have the maturity to recognize and safely navigate them. Parenting provides the foundation for secure attachment including the availability needed to support the stressed child while also modeling effective emotion regulation. Through parenting children learn to recognize situations they cannot master and to ask for help and support when necessary. Family life changes as children mature and develop. Parenting approaches need to appreciate these changes and be adjusted and modified to respond to changes and growth in the child’s aspirations and abilities. Parenting that included physical control of the child during toddlerhood will need to change to parenting based upon clear and agreed upon rules and consequences for a school aged child and then to more supervision that encourages judgment, impulse control, and the development of prosocial values during adolescence. Not only do parenting approaches need to change and transform in step with developmental maturation caretakers will need to be alert to situations that demonstrate lack of skills and regression and require parenting approaches more appropriate for less mature children.

Family therapy is very helpful in addressing parenting challenges in families. The therapist will need to appreciate the source of these problems which can be a mismatch between parenting approach and the child’s developmental capacity, an inherent difference in the child impacting development and function, a lack of agreement or cooperation between parental figures or external stressors impacting parental ability to recognize or address the child’s needs. Stress related to immigration and acculturation can further magnify these mismatches and differences between the parent and child. During acculturation, the child often gets accustomed to the new culture faster than the parent. This separation in the acculturation process can cause more conflicts in the family, also referred to as acculturative family distancing. Furthermore, the parents' added challenge of imparting culture to the next generation in isolation can make parents more rigid and less open to child's need to integrate two cultures. Thus, engaging and empowering parents to be effective in a culturally sensitive manner is an important target of family therapy. The therapist's ability to engage the parents and the child, appreciate the source of parenting difficulty and ensure the caretakers' trust and motivation often utilize all of the family therapist's skills in connecting, directing, and reinforcing new and effective behaviors in both caretaker and child. An accurate assessment, psychoeducation about children's difference, enactments during family sessions and a relationship of respect and regard with caretakers are important skills in enhancing parenting and building family competency and flexibility which in turn build family connections and affection.

Ensuring a Balance between Support and Limits in Parent–Child Interactions

One of the most challenging aspects of child-focused family therapy is developing an effective balance between, on the one hand, enhancing parent–child warmth, support, and the interactions that reflect secure parent–child attachment and mutual respect, and on the other, working to establish effective rules, supervision, and consequences for misbehavior and rule-breaking. This often requires skill, focused attention, and persistence on the part of the therapist. Because parental warmth forms the foundation of the mutual respect needed for effective discipline, the therapist's first task is to build that warmth. This often occurs at a time when parents feel betrayed, helpless, angry, and victimized. Parents may mistrust the therapist based on experiences of racism and health inequities within the health care system. An open discussion of these experiences and recognition of the impact of discrimination and difference upon the family can be extremely helpful in ensuring the partnership and relationship necessary for family therapy to be effective. At this point, it is the therapist's belief in the parents, in their love for their child, and in the possibility that they might help their child grow up safely and successfully that matters most. The therapist models warmth in their relationship with the parents. Compassionate understanding characterizes these interactions. Interactions that caringly convey the therapist’s belief that the parents care for their child and believe in the child’s future characterize the therapy. This must occur even if the child expresses skepticism or disbelief about the parents caring for them. As trust develops between the therapist and the parents, who know that the therapist believes that they care truly about their children, a wide range of interventions is possible. These include establishing parent interactions with the child aimed at understanding the child’s disagreements with their parents in a nondefensive way and in a dispassionate fashion while establishing expectations for their child’s behavior as well as consistent consequences for failing to meet expectations. In this process, the therapist helps the parents to deal with their disappointments with their child so that consequences are not vindictive and helps the parents appreciate that their child could want their approval and could wish to earn their regard. In this approach, parents understand and support their child’s emotions, disappointment, frustrations, and anger while also holding firm expectations, limits, and consequences. In a sense, parents are feeling one way—frustrated and angry—and acting another way—calm and firm—while allowing the child to act one way—frustrated and angry—and expecting the child to feel another way—accepted and respected even while adhering to parental limits and appreciating parental involvement and concern.

Parents often want advice about establishing discipline and limits, especially in situations in which the child has a disruptive behavior disorder and defies parental rules and limits. This often represents a challenging therapeutic concern. Often in these situations there is a difference in approach between parental figures. Everyone is exhausted, hurt, frustrated, and angry. The parents will frequently claim that consequences do not work. Disrespect and defeat characterize parent–child interactions. In these situations the therapist might choose to meet alone with parents to help them to begin to work together, as well as to develop strategies to reinstitute effective limits. The therapist can help parents to prioritize their expectations and plan to establish one rule at a time. In the process the therapist can also help the parents to develop realistic expectations of their efforts with their child. The parents can receive validation of the experience of frustration and their sense of being betrayed by their child. As long as the parents hold on to their experience of being hurt by their child, it is very unlikely that they will notice small attempts at improvement on their child's part. They also will have difficulty holding to their expectations if the child becomes angry or has a temper outburst when a limit is imposed. In such frustrating situations, guided by the therapist, a parent can gain support from their partner that could result in decreased emotional intensity and improved ability to maintain positive expectations. Another frequent occurrence that can be discussed with the parents alone is the possibility that the child will act in a totally unaffected way when the limit is imposed, as if to say that the parents have no effect on the child. By discussing these possible scenarios, the therapist can help the parents prepare to institute appropriate limits in a planned and controlled manner. In these sessions the parents can practice talking with their child and maintaining firmness even if the child is particularly challenging. The therapist may also choose to meet alone with the child to see whether they have concerns that the therapy might resolve. Defiant, disruptive children and adolescents often want to be trusted and understood, and to have autonomy.

The therapist can help the child to receive parental approval if the child performs acceptably and gains autonomy in relation to their ability to act responsibly. When both parents and child are prepared, a session with all of them together can be held outlining the rules, expectations, and consequences. The child has the opportunity to ask for what they want, and the therapist can help the parents to promise the child the desired autonomy if the child abides by the rules and accepts the consequences. The consequences should be simple, easy to undergo, brief, and have some relationship to the transgression. The key to consequences is that they are meaningful but not onerous. Losing access to a cell phone or computer for one evening is a meaningful but transient consequence. If the parents and child form a committed relationship with the therapist, the therapist can then, with patience and time, bring them back together in an initially tenuous but gradually caring manner. Building understanding and trust, establishing a realistic plan for limits, and noticing small successes provides a pattern of improvement that all can recognize and believe in.

Engaging Families in All Levels of Care

Family members should be actively involved in decisions about treatment programs, levels of care and care delivered, and inception and discontinuation of treatment. This expectation helps the parents or caretakers develop monitoring, supervisory, and organizational skills that they might be lacking. Because the parents possess decision-making power, they must learn to understand their child, make the child’s behavior make sense, and begin the process of accepting the child’s differences. This can accentuate feelings of support and understanding, which may have been undermined by their responses to symptoms and their consequences. This is especially true in the case of families with children with disruptive disorders, in which difficulties in school or with the justice system may have felt hurtful and shaming.

The journey to get out-of-home services often results in parents feeling ineffective and disempowered. The purpose of family intervention is to empower parents to conduct effective parenting. Thus, family involvement is essential for children being treated in treatment programs such as hospitalization, day treatment, or residential treatment. For the home environment to reinforce positive behaviors and not reinforce negative behaviors, it is necessary for program clinicians and staff to be attentive to the interactive sequences of which the symptoms were a part and intervene with the family to alter those sequences so that the home environment is different when the child is discharged or the program ends. If discipline is problematic, parents can be consulted when program rules are broken and consequences are required. Parents can be informed by the child of successful experiences so that their praise is added to that of the staff. These successful experiences can translate into hope for change for parents. Parents can be expected to meet regularly to have a meaningful voice in treatment review, learning to develop realistic expectations of their child based on how the child is able to function rather than based on everyone's wishes about how they might function. It can be difficult at times to share power with parents, but with patience the staff can help the parents to feel empowered throughout the treatment program. If passes home are part of the treatment program, activities can be planned for the entire family that emphasize practice with new interactions at home prior to discharge.

FAMILY THERAPY FOR SPECIFIC PROBLEMS

Attention-Deficit/Hyperactivity Disorder

Families with a child with attention-deficit/hyperactivity disorder (ADHD) need to learn about the disorder, understand appropriate treatment, and support the development and functioning of their child. This requires that parents agree about the presence of the problem, agree to seek the help of a physician, and agree to have their child take medication if prescribed. It is also important for the parents to learn to be consistent with behavioral expectations and to appreciate and reinforce the child's strengths. Often, prior to diagnosis, the child may have experienced failure at school and rejection by peers and marked negative interactions at home. Family therapy after diagnosis of ADHD may be needed to help the family assist the child with these previous difficulties, and also to help the child to be more successful at school and with friends and to feel consistently accepted at home. Comorbidity with learning disability is common, and the family may need assistance with ensuring appropriate education for the child. ADHD often occurs in association with complex trauma in situations of abuse and trauma. Sometimes the concentration problems of the child are a manifestation of anxiety caused by trauma and sometimes the child has both ADHD and Complex PTSD. Therapeutic flexibility and accurate assessment are important aspects of family treatment. Transitions, including moving and changing schools, can be challenging. The family may also need assistance during the child’s adolescence, particularly with impulse control, exposure to drugs and alcohol, and self-management as the child becomes a young adult. A chronic illness framework often is extremely helpful for families with children with ADHD.

Jason was a 7-year-old boy who attended the second grade and who had been noted by his teacher to be extremely active in class and to have significant difficulty concentrating and following class rules. He especially had difficulty during unstructured time in school, both in the cafeteria at lunchtime and on the playground at recess.

When his teacher checked with his previous teachers from kindergarten and first grade, they stated that Jason had difficulties with his attention and activity level in both of those grades. Despite these concerns he was quite bright and had been able successfully to complete his grades and pass to the second grade. Jason's teacher invited his mother to come to school for a meeting after having sent home reports concerning his overactive behavior and his inability to finish his work.

At the meeting with Jason's mother, his teacher spoke about her concerns and suggested that Jason should undergo a psychiatric evaluation and might need medication to help him concentrate and manage his activity. His mother reported that she was also concerned about how active he was at home and how difficult it was to help him get ready for school in the morning or finish his homework in the afternoon.

Jason lived at home with his mother, stepfather, and 4-year-old stepsister. His parents had divorced when Jason was 3 years old, and his mother remarried his current stepfather 1 year prior to Jason turning 7 years old. Jason saw his biologic father on a regular basis and spent alternative weekends at his father's house. His father had also remarried, and Jason had an 18-month-old half-brother who lived with his father and stepmother.

Jason was seen by a child psychiatrist shortly after his mother and teacher met. The child psychiatrist believed that Jason was a child with ADHD and suggested to the mother that a trial of stimulant medication might be helpful for him. The mother was given rating scales to complete after Jason had started on the medication, along with written information about both ADHD and the stimulant that was prescribed for him.

The child psychiatrist also recommended that there be a meeting of all four adults involved in Jason's life—his mother, stepfather, stepmother, and biologic father—so that they could talk together about Jason's difficulties and ensure consistent treatment. The goals of the meeting were to encourage regular use of the stimulant medication and to help everyone to understand the need for clear and focused directions, focused attention, follow-through on homework, and consistent discipline in both households.

The family attended this meeting. The psychiatrist noted that Jason's biologic father was concerned about the use of medication and asked about the risks and side effects of the medicine. Jason's stepmother was concerned that there was no way successfully to help Jason without medication. As the family talked together with the child psychiatrist, asking questions and gathering information, they came to the conclusion that the best approach was to institute a trial of stimulants. The trial was begun, and Jason's teacher reported to the family that he was much more able to focus his attention and manage his behavior in school. He was noted to have less difficulty during unstructured times in school and was also able to do better academically than he had previously.

The child psychiatrist also suggested that the family come back for a second meeting, after Jason had been on medication for 6 weeks, to review everyone's impression of Jason's progress. The child psychiatrist recognized that the biologic father and stepmother were now comfortable using medication for Jason. They agreed to follow through with the medicine and to help Jason manage his behavior better in their home.

Three years later, as Jason continued to do well in school on stimulants, during a routine medication review with the child psychiatrist, Jason's mother reported that his father was getting divorced for a second time. Jason tearfully admitted that he was very upset by this and that he was going to miss his times with his stepmother and younger half-brother.

The child psychiatrist asked to speak to the father, who agreed to come for a session with Jason and talk with him about what was happening in his family. Jason's father agreed to help him have continued phone contact with this stepmother as the father and stepmother proceeded with their plans for divorce.

Six years later, Jason was old enough to obtain a driving permit. He had continued to take stimulants daily and had done well in school. He had grown up without significant emotional and behavior troubles and was eager to learn how to drive. Jason's father and mother had continued to meet with the child psychiatrist every 4 months to review Jason's medication and to check in with how everyone was doing. They were concerned about Jason driving. Jason was sure that he would be able to drive perfectly well even though he was known to be more impulsive and more disorganized than other teens his age. The child psychiatrist helped the family to discuss the privilege of driving. They decided that they would let him drive with one of them in the car and not let him drive after dark for the first several months that he was driving. They decided not to allow him to drive with peers until he had driven successfully for several months. This negotiation was difficult for Jason, but Jason's father and mother both felt very supported by the child psychiatrist. They were comfortable with their decision that Jason could learn to drive as long as he was monitored quite closely and supervised very carefully.

Oppositional Defiant Disorder and Conduct Disorder

The primary goal of family therapy for oppositional defiant disorder and conduct disorder in one or more children is the establishment of a warm and mutually respectful relationship between parents and child, which forms the foundation of effective limits, appropriate supervision, and responsible prosocial behavior by the child. Often the family situation is complicated by any or several of a variety of family risk factors. These include parental stress, parental psychopathology or substance abuse, criminality and incarceration, poverty, marked parental preoccupation with time-consuming work, or preoccupation with other children. Overcrowding, living in a violent community, and domestic violence also occur frequently. Negative coercive cycles often occur in the interactions between parent and child, in which the child is demanding and the parent becomes controlling, leading the child to seek attention through misbehavior and leading the parent to believe that only direct control will make the child behave appropriately. The child fails to learn self-control and self-direction, and the parent becomes gradually and insidiously disempowered and negative toward the child. As the child grows, the parent increasingly becomes hopeless about the child and begins to believe that the child will fail to grow up successfully. This results in parents feeling helpless and re-emphasizes that improvement must come from outside the family. Therefore, the parent hopes that something else will lead the child to develop responsibility and seeks placement of the child in residential treatment or in the juvenile justice system. This perspective localizes the problem in the misbehaving child, leaving the child feeling blamed and even more disconnected from their family.

Two forms of family therapy deal specifically with these issues. Parent–child interaction therapy is specifically designed to treat families with young (preschool) children with oppositional defiant disorder. Multisystemic therapy has demonstrated effectiveness for families with adolescents with delinquency who are involved with the juvenile justice system. Parent–child interaction therapy focuses on building the positive aspects of the relationship between the parent and young child. Through the therapist’s praise, the parent relaxes and becomes more accepting and more openly caring toward the child. The parent receives the therapist’s praise and care and passes that on to the child, leading to closeness, the experience of being understood and cared for, and warmth and attunement in the parent–child relationship. This also leads to improved cooperation between parent and child, greater compliance on the child’s part, and, on the part of the child, much greater interest in the parent’s praise and positive regard. Parent–child interaction therapy has been shown in more than 30 clinical trials to be effective for families with preschool children (between ages of 2 and 6 years) with oppositional defiant disorder. It has also frequently been requested by child protective services to improve parent–child relationships in situations in which mild physical abuse has occurred.

Multisystemic therapy has been shown to be effective in reducing recidivism and delinquency for children and adolescents in treatment trials and through follow-up. Trained therapists with very small caseloads help the parents to manage the stresses in their lives and supervise and set limits on their child. The therapist is available to the family consistently and the treatment is intensive, with three or four contacts with the family weekly over 4 months. The therapist builds family cohesion and encourages the child to identify and earn privileges by being responsive to parental direction. Rules are defined and consequences are specified early in treatment. Parents are expected to be available, observant, and consistent in monitoring their children. For this to occur, other difficulties that the family experiences need to be addressed through the therapy. This treatment is best suited to moderate delinquency and works best early in the course of delinquency.

Another form of family treatment for conduct disorder is parent management training. In this therapy, parents attend sessions without the child and learn and practice techniques of setting limits and imposing consequences. Most important in this treatment is helping the parents plan and practice how they will withstand the child’s demanding and defiant behavior and how they will respond if the child’s behavior becomes worse before it gets better (as is expected). In this treatment the therapist supports a united front among adult authority figures and builds mutual support among these authorities. Parental self-control is another goal of these parent-coaching sessions. The therapist needs to ensure that the parents learn the skills necessary to set effective limits and manage their lives at the same time.

Isaac was a 4½-year-old boy who was brought to a child mental health clinic for evaluation of defiant, challenging, and difficult behavior. He had been referred by his pediatrician after his mother discussed his defiant and difficult behavior with the pediatrician during a regular checkup. The history also revealed that Isaac had been expelled from his preschool and had to be looked after by a neighbor in her house because of aggressive and difficult behavior at school.

When the clinician met with Isaac and his mother, Isaac appeared to be a well-behaved, quiet child who was able to focus his attention and play easily with the interviewer. However, it was noted that his mother was extremely negative about him, and was frustrated, angry, withdrawn, and sad. During the interview the clinician noted that Isaac's mother was exhausted, overwhelmed, and observably sad and irritable.

The more comfortably Isaac played with the clinician in his mother's presence, the more upset and withdrawn she became. She reported that Isaac's father had left before he was born and that she had an older, 8-year-old boy with developmental disability and cerebral palsy. That child was in special education classes. She lived alone with the children. She stated that she was estranged from her mother and that she worked as a cashier in a grocery store. The family lived in a small apartment in an impoverished area of the city, and the mother reported that at times it was difficult to have enough food for her, Isaac, and his brother to eat.

The clinician was struck by the poignancy of Isaac's mother's situation, by her isolation and lack of support, and by her level of exhaustion and sense of being overwhelmed. The therapist was also touched by Isaac's attention to his mother and his tendency to watch her and comfort her when she became sad during the interview. The clinician was attentive to and supportive of the mother, and during the interview she began to calm down and admitted to being depressed, sad, and lonely.

Isaac and his mother arrived for their second session on time and both seemed eager to talk. The mother had noticed that Isaac was a little bit better behaved at home, and she seemed more relaxed and comfortable.

The therapist continued to support the mother and encouraged her to engage with Isaac in a game during the session. The therapist was warm, caring, and praising of the mother's efforts with her son. The therapist also recommended that Isaac and his mother have some regular time to enjoy each other's company during the week and suggested 15 to 20 minutes every day, during which their neighbor could look after Isaac's older brother. As Isaac received more attention from his mother at home, their negative interactions decreased. He seemed to relax and become less quick to anger, both at his babysitter's house and at his mother's house.

The clinician also encouraged the mother to reconnect with her siblings and her mother, who were in the community, and to reach out to them for support and understanding. The mother noticed that she was more relaxed and less stressed and that she was more able to experience enjoyable times with her sister. She also began to communicate in a more open way with her mother.

As the sisters became more involved with Isaac's family, there were more family gatherings and more opportunities for the mother to experience some relief from her caretaking obligations. Over a period of 3 months of weekly sessions and 2 months of biweekly sessions, Isaac's mother became much more relaxed, was able to obtain babysitters for her children, and got more involved in her older son's special education programs. She became more assertive about obtaining a raise at work and socialized with her colleagues. She spent more time with her family and her neighbors. She also resumed attending church, and Isaac became involved in the church-sponsored preschool kindergarten group.

The therapist stayed in touch with Isaac and his mother and heard from her that Isaac had successfully begun kindergarten during the next academic year. His behavior was much improved and his curiosity, desire to play and learn, and attentiveness to adults had helped him to learn and grow in kindergarten.

Joseph was referred to the intensive in-home treatment team after his discharge from the juvenile detention facility in his community. He was 15½ years old and had been placed in the detention facility after his arrest for vandalism and breaking curfew. The vandalism involved breaking windows in his school, and well as painting graffiti on several businesses in his downtown neighborhood. Joseph lived with his mother, stepfather, 12- and 19-year-old brothers, and 17-year-old sister.

When the in-home treatment therapist went to meet Joseph and his family, the mother was at home with the younger brother. Joseph was not at home. He had told his mother that he needed to visit his friends after school and that he would be home in time for the appointment. He was late and had not informed his mother.

The history the therapist obtained was that Joseph had been a very good student and a fine boy until the age of 9, when his grandfather suddenly died and his mother moved the family into the grandmother’s house to look after her. Joseph had been close to his grandfather and had reacted to his death with sadness and withdrawal from the family.

His mother noted that he began to act up at that time and said that he had started hanging around the “wrong kids” at the age of 11 and had begun truancy when in middle school. Joseph had regularly received reprimands from the school administration, mostly for fighting, defying teachers’ authority, and misbehavior during recess or lunch period. Joseph’s mother said that at this point she believed that Joseph was outside of her control and she had hoped that 8 weeks in the juvenile detention facility would help him to learn to manage his behavior. Because he had not come home for the therapy appointment, she was worried that perhaps things were just as bad as they were prior to his arrest and stay in the detention center.

The therapist noted that the family lived in a cramped apartment in a rundown building that needed maintenance. When the therapist asked Joseph’s mother about her husband, she stated that he also was supposed to have come home for the therapy session. She said that he often stopped at the bar on his way home to hang out with his friends and drink. Joseph's mother was increasingly worried about her 12-year-old son, who seemed to be following in Joseph's footsteps, and reported that her 17-year-old daughter would likely be moving out of the home with her boyfriend when she had her baby in approximately 3 months.

Joseph's 19-year-old brother was in and out of the house, and his mother was never sure whether she could expect him home on a specific evening. She thought that Joseph's older brother was having difficulties with substance use, and she reported that there were times when she smoked marijuana or drank just to get away from her concerns and feel better for a little while. The therapist asked Joseph's mother about the family's economic situation and learned that the stepfather had a regular job as an auto mechanic but was not paid well and that the mother worked part time in a department store as a cashier.

The family had Children's Health Insurance for the children. This managed Medicaid program would provide financial support for the intensive home-based treatment program. During the first meeting, the therapist worked to understand the family's situation and recognized Joseph's mother's affection for her children and the dismay that she was experiencing because of their difficulties. The therapist spent 90 minutes getting to know Joseph's mother and her youngest son. During this first meeting, Joseph's sister came into the house during their conversation and told her mother that she had seen Joseph hanging out on the street with friends whom he used to get in trouble with before he was in the detention center.

The therapist proceeded with the assessment and commented on the strengths of the family, including their strong connections with each other, and on the hopes of the family for the futures of the younger brother and the sister. The therapist developed a deep respect for Joseph's mother and closed the first session by saying that she would come back the next day and meet with the mother again. The therapist wrote a message to Joseph, saying that she had been there and that she had heard about him and knew that she would very much enjoy meeting him and invited him to be present next evening for her visit.

Over the next 6 weeks, the therapist visited Joseph and his family in their home three to four times a week. She invited Joseph's mother to call any time if she had concerns, and she helped Joseph's mother to think about her priorities for their treatment process. Joseph's mother said that her goals included helping Joseph stay out of detention and resume school; repairing her relationship with the children's stepfather and encouraging him to take care of his chronic alcohol use; improving the family's financial situation by obtaining a full-time job; and deciding with the stepfather whether to continue to allow Joseph's older brother to come and go in the household.

Mother and stepfather set up rules and consequences for Joseph with three expectations: (1) that he would attend school all day, each school day, (2) that he would attend an afternoon program at the local Boys and Girls Club, where he could receive tutoring to complete his homework, as well as engage in basketball and video games with other young people at the center, and (3) that Joseph would return home after the Boys and Girls Club for dinner and that he would have access to his cell phone and his video game set in the evening if he followed through with the plan.

Joseph initially strongly resisted the rules and expectations, telling his mother that "she had no right telling him what to do," that he was grown, and that "he could make up his own mind and make his own decisions." The therapist helped Joseph's mother reiterate to her son that she would be happy for him to have greater freedom if he demonstrated greater responsibility and was able to comply with the house rules. The therapist also visited Joseph's school and talked to his counselor and teachers and encouraged them to be warm, engaging, and supportive of Joseph to help him know that they wanted him to be in school every day and that they wanted to get to know him and learn about him. The therapist helped Joseph to write a biography of his grandfather, which Joseph researched by talking with other community leaders and neighbors who had known him. The therapist also encouraged Joseph to resume his interest in basketball and baseball and to learn as much as he could about the history of his family and the history of his family's role in his community.

As adults who expressed interest in him and his ideas surrounded Joseph, he began significantly to improve his mood, decrease his level of anger, and increase his hopes for the future and his interest in school and his after-school program. The intensive treatment program continued to work with Joseph and his family for 3 months and encouraged the adults in Joseph's life to express more responsibility and interest in him and to offer him more opportunities to be a part of his school and a part of his neighborhood. The therapist also encouraged the mother and stepfather to resume attending their church, which the family did, and the therapist talked to the minister about being particularly welcoming for Joseph, his older sister, and his younger brother. The therapist also talked with Joseph's stepfather about his alcohol use. The stepfather was receptive to decreasing his alcohol use and being home more in the evenings and paying more attention to his family and his wife.

Together with consistent, nearly daily contact with the family, involvement with several important adults, and child-serving institutions in Joseph's community, the therapist, by creatively using Joseph's interest in his family and in his grandfather and identifying Joseph's interest in telling the stories of his family, was able to improve Joseph's school attendance, sense of himself, and the family's overall sense of well-being in the 12 weeks of contact. When the program ceased, Joseph had several adults involved in his life, his mother had a full-time job that felt rewarding, her relationship with her daughter who was about to have her first child had significantly improved, and Joseph's younger brother had resumed his academic progress and his athletic activities.

A family therapist at the local mental health center was able to meet Joseph and his family for the last 3 weeks of the intensive program once a week at their clinic. That therapist continued to meet with Joseph and his family over the next several months on a weekly or bimonthly basis.

At the close of treatment, 1 year after Joseph had been discharged from the detention center, he was at home consistently, was attending school, had caught up academically, and had resumed regular activities with peers who were interested in sports and academic success. Joseph's mother and stepfather had decided to remain married and had improved their relationship.

Family Therapy for Depression

The core of treating depression in children is the consistent building of competence and connections for the child. The problem in accomplishing these goals is that the depressed child has problems with concentration, which directly interferes with competence, especially at school, and marked irritability, which greatly interferes with successful relationships with family, peers, and important adults. Furthermore, the depression creates a marked lack of motivation, which leads the child to not believe in or pursue treatment. Substance abuse is often a comorbid problem, which may leave the child feeling better in the short term but with a significant long-term problem in addition to the symptoms of depression. The longer depression persists, the more likely it is that such a child will develop significant and persistent secondary problems, including academic failure, legal difficulties, and further alienation from parents, who may also be experiencing stress.

In the face of these difficulties, family treatment offers several advantages. The parents can be asked to ensure that the child participate in treatment, even if the child would rather not, as long as the child's problems persist. Suicidal ideation, plans, and attempts can occur at this point and require specific focused attention by the therapist and the parents. Parental warmth and encouragement can be achieved by linking the child’s problems to the experience of depression. Depression can be defined as a phenomenon that has taken over their child, and which is painful and hurtful while draining the child’s energy and motivation. The therapist can reinforce and support the parents caring for and commitment to their child. As the parents express their care for the child, the therapist can build the child’s trust in the parents.

The therapist can then encourage the parents to expect the child to participate in school, both in academics and in activities while enlisting school personnel to engage with the child and provide remedial support if needed. The therapist encourages parents to be available and attentive to the child’s accomplishments and supportive of the child’s emotional experience. Therapeutic support for building on the child’s interests, strengths, and abilities is also important, as are efforts to build relationships with prosocial peers while also limiting substance use. Prescription of selective serotonin reuptake inhibitors (SSRIs) while monitoring suicidality can promote this process for the child and family if needed.

Lydia was a 14-year-old girl who was brought to the emergency room of her local hospital after ingesting an overdose of 15 acetaminophen tablets. Lydia had taken the overdose after an extremely difficult day in which she was teased by friends at school, was criticized by her teacher for not having completed her homework, and then engaged in an argument with her mother about her wish to go out with friends that evening. Lydia challenged her mother and threatened to leave the house. Her mother was worried that Lydia might be meeting a teenage boy and got very angry at Lydia and called her a demeaning name. During the argument Lydia rushed into the bathroom and found the bottle of acetaminophen tablets and took the overdose. She then became quite frightened and told her mother that she had taken the pills. Her mother called an ambulance, and Lydia was brought to the emergency room.

Lydia maintained consciousness and talked openly with the clinician in the emergency room about her increasing distress over the last 12 months. She had become increasingly involved with peers who were in trouble, and her performance in school had declined. She also reported that she had stopped participating in her usual activities at school, including quitting the cheerleading squad. She had been sexually active with a 16-year-old boy on one occasion and reported that the boy had stopped going out with her after they had had sex, and he had talked about the experience in an embarrassing way with other young people at her school. She had previously been a good student and popular in her class but had recently become more irritable, more isolated, and less interested in academics and her friends.

Lydia's mother reported that her daughter had changed from a happy, upbeat teenager to a surly, irritable child who was very difficult to communicate with. Lydia lived with her mother and her three younger brothers, who were 9, 7, and 5 years old, respectively. Their father had left the family to work in another community and was sending money home to the mother. He was away for several months at a time doing construction work and was only in contact with the family by phone once a week. Lydia had been very close to her father, and his recent occupational absence had been distressing for her. She had felt isolated and disconnected from her family.

Lydia's relationship with her mother had always been ambivalent; she had been very helpful but not very emotionally close to her mother and had become more quick to anger and argumentative with her mother over the last several months.

Lydia was admitted to the adolescent medical unit for monitoring because of the possibility of toxicity from the acetaminophen. While in the hospital, she and her mother were seen by a child psychiatry consultant, who invited Lydia to talk openly with her mother about why she had made the suicide attempt and what had happened that day that had so upset her. The consultant also encouraged Lydia to talk about missing her father and her recent sadness. Her mother, while being somewhat defensive, was supported by the clinician to listen to her daughter's difficulties and concerns.

Lydia was able to be open with her mother about the problems she was having in school, the difficulties that she was having with peers, and the fact that school suddenly had become very difficult for her. The clinician held daily sessions in the medical hospital over the next 3 days with Lydia and her mother to enhance the closeness and supportiveness of their relationship. He encouraged Lydia's mother to listen to her daughter in a nonjudgmental and nonreactive fashion. As the mother felt the support of the clinician and his encouragement to be clear with Lydia about her concern about the impulsive suicide attempt, her mother was also able to be more understanding and available to her daughter during these meetings. The clinician also arranged a phone session with Lydia, her mother, and her father.

After 3 days of medical hospitalization, Lydia was able to apologize to her mother for taking the overdose and to ask her for enhanced involvement and understanding at home. The family was then referred to an outpatient clinician at the local mental health center, where they were seen 3 days after discharge. Lydia reported improved mood and increased attention from her mother at home and had returned to school. The new therapist asked permission to contact Lydia's teachers to arrange more support at school. She strongly praised the enhanced closeness between Lydia and her mother.

Lydia was not initially placed on psychopharmacology. She had noted improved mood at home and had managed her behavior with less irritability and impulsivity in the initial week following her hospitalization. Future sessions were held weekly and were divided, with half the session for Lydia individually and half of the time with Lydia and her mother. The therapist focused with Lydia on her life at school, her schoolwork, and her peer relationships. The meetings between Lydia and her mother led to enhanced understanding, improved communication, and greater mutual support. Regular phone calls involved Lydia's father.

With the enhanced family support and Lydia's increasing competence as a teenager, her mood improved dramatically, and her relationship with her mother grew closer. The telephone contact between Lydia and her father re-engaged them and restored the warmth they had shared.

After 3 months of outpatient treatment, Lydia was doing well and came for sessions once every 2 weeks. The therapist noted that Lydia's mother was sad and missed her husband and offered the mother the opportunity to come for individual meetings to talk about this and how she could carry on at home while her husband was away.

Traumatic Stress in Children

Trauma-focused cognitive-behavioral therapy for children has been identified as an evidence-based treatment for traumatic stress in children. It is important to recognize that this treatment has an essential family component. Family education, emotional support for the parent who may be distancing themselves due to the child's traumatic experience, and family involvement in supporting the child while developing and sharing the trauma narrative are essential elements in this treatment. This often requires that parents accept the reality that they could not protect their child from harm or that one or both of them might actually have perpetrated the harm. Helping the parents acknowledge their feelings and the realities of their child's experience of exposure to or injury by trauma can often be quite challenging. The therapist has two goals: to support and care for the parents and to assist the child in understanding and dealing with what has happened to them. This can be horrific and is often frightening or infuriating for the therapist. Therapists in this situation often need an opportunity to discuss cases with colleagues or with a supervisor. Attention to the safety of all participants provides the foundation for engagement, and engagement builds the trust needed to share the painful, shameful, and sad aspects of trauma and its aftermath. These realities are present regardless of the source and nature of the trauma. Focusing on a successful resumption of family life founded on the belief that the trauma has led to closer relationships and greater resilience is often healing in itself. The recognition of family resiliency is then coupled with acceptance of the unpleasant emotions that are normal responses to trauma among them all. A step-by-step therapeutic approach, including the family, is extremely helpful. This approach leads from providing safety to ensuring that basic needs are met. It continues with education for understanding, behavioral competence, emotional support, and narrative creation and validation. The trauma becomes a remembered but not excessively arousing experience. As with family treatment of other child problems, the treatment can lead not only to the resolution of traumatic stress but also to an enhanced awareness of family strength among all members.

Monica was 6 years old and was referred by her pediatrician to a child psychiatrist for evaluation after she had been involved in a serious motor vehicle accident. One month prior to the referral, Monica had been riding in the car with her mother. She had been seated in a booster chair in the back seat. Their car was hit by another car that was driving very fast. Only Monica and her mother had been in the car, and Monica's mother was seriously injured in the accident. When the ambulance came to pick Monica up, they noticed that she was crying for her mother and that she was shaking in the back seat, although she did not seem to have been physically injured in the accident.

Monica's mother had lost consciousness in the front seat and needed to be taken to the emergency room in a helicopter. The police had removed Monica from the back seat, and one of the policemen held her on his lap while she was crying out for her mother. They found the mother's pocketbook and called Monica's father at work. As Monica's mother was being flown to the emergency room in the helicopter, Monica's father came to the scene of the accident and picked up his daughter.

Monica had previously been a healthy girl attending first grade. She had two younger sisters, ages 4 and 2 years, respectively. Her father was a lawyer in a firm in the community, and her mother was a part-time bookkeeper for a local physician. The family had significant resources, and the children attended either private school or private daycare.

Monica was slow to warm up and approached new situations with fearfulness and reticence. Now, 3 months into the school year, she had established friends and had a very positive relationship with her teacher. When her father brought her home and got a neighbor to stay with his three daughters while he went to the hospital, everyone noted that Monica was extremely upset.

During the mother's hospitalization, which lasted for 2 weeks, she required abdominal surgery to remove her spleen and was found to have a fractured tibia and fibula in one leg and required a cast and crutches. Monica had been able to talk with her mother by phone several times in the hospital beginning on the fourth day after the accident, and near the end of her mother's hospitalization, Monica had been able to visit her in the waiting area of the hospital.

When Monica's mother came home, she learned that Monica had been crying every morning on the way to school and also while at school. Monica also had become increasingly withdrawn at home from her younger sisters and at school from her peers and from school activities. Because of this withdrawal and problems with separation, Monica's pediatrician recommended that she and her family see a child psychiatrist.

The child psychiatrist obtained the history just described and noted that Monica's mother was also having nightmares and flashbacks of the accident and was increasingly hypervigilant. She had become frightened to drive the car. Monica's father was very concerned about his wife and about Monica and attended the initial session with them. He stated that he knew that his wife and his oldest daughter were both quiet and shy and were very much alike and very close to each other.

The child psychiatrist determined that Monica's father was deeply invested in his marriage and in wanting to help Monica and his wife recover from the physical and psychological effects of the automobile accident. During the conversation Monica's mother stated that she felt extremely guilty for having the accident and was frightened both about what had happened to her and also the possibility that the accident could have hurt her daughter. She began crying in the session and it was obvious that her distress deeply affected both Monica and her father.

The child psychiatrist talked with the family about the fact that they had a normal reaction to an extraordinarily frightening and traumatic event. He also explained that the trauma for Monica was somewhat different from the trauma for her mother. Monica's trauma involved not only the accident but also her fear of her mother's well-being and her experience of being separated from her mother at the scene of the accident and during her mother's hospitalization.

The psychiatrist contrasted that with the mother's fears from her injury and treatment, her worries about her daughter, and her guilt from the car accident. The child psychiatrist also spoke about the strengths he recognized in the family: the father's intense connection to the family, his dedication to being helpful to his wife and his daughter, and their impressive closeness and love for each other. As he said this, the family relaxed. The therapist suggested realignment of the family's energy now that the threat to survival was over. He suggested that the family begin to orient themselves to resume their lives as a family while also recognizing their fragile emotional state.

The child psychiatrist recommended a family-based course of therapy over the next 8 to 10 weeks in which the mother, father, and daughter would work on reconnecting with their future lives while integrating their experience of trauma, injury, and anxiety into their past. The psychiatrist proposed homework for the family; he suggested that the father spend time alone with Monica, talking with her about the fact that her mother was better now and that it was her job to resume being a big sister to her little sisters, a student at school, and a regular 6-year-old girl. The therapist also suggested that the mother and father talk together to help the mother to accept the accident and also to appreciate that because of her appropriate attention to her child's safety, her child was in the booster seat in the back seat and was not injured. They could discuss the fact that the mother had taken very good care of Monica and that Monica was having trouble because she continued to be frightened that her mother was not all right.

Over several sessions, Monica's mother was able increasingly to reassure her daughter that she was okay and encouraged her daughter to re-engage with her school and friends and enjoyment of activities and fun as a child. She also gradually reassumed her role as an effective and caring mother for her three daughters. The father stayed actively involved with all of the children and maintained his belief in his wife's ability to recover and resume her role as his wife and their children's mother.

It became apparent to the therapist and Monica’s parents that they had too rigidly adopted complementary roles as breadwinner and caretaker and that some readjustment of their partnership would need to be addressed in the therapy and recovery process. Monica’s mother was grateful to her husband for being more flexible and responsive to her and their children’s needs. He continued to be invested in his career, but with less intensity and less single-mindedness.

After 4 weeks of treatment, Monica was increasingly acting as she had prior to the accident, as a successful student, a confident big sister, and someone who could more easily separate from her mother and the family. She had fewer episodes of tearfulness and was able to sleep through the night.

After 8 weeks, Monica’s teacher reported to the parents that she had, in fact, resumed the previous functioning and had caught up academically and appeared to all to have recovered. Her mother was less frightened and was sleeping well, and had resumed work and driving. She and Monica could discuss the accident openly without crying.

Treatment of Anorexia and Substance Abuse

Family therapy–based treatment models have demonstrated effectiveness for both anorexia nervosa and adolescent substance abuse. While anorexia was formerly thought of as a disorder of middle and upper class families, it has more recently been recognized in children and adolescents of all classes and ethnicities. In both of these problems the child is involved in an engaging, self-reinforcing, and dangerous behavior. This behavior often is pursued in an extreme fashion and ultimately leads to failure to achieve other developmental skills. The lack of these skills also often renders the child anxious and further narrows the child’s orientation to their life. There may be a genetic predisposition to either anorexia or substance abuse. Despite this, the child will need to learn to manage the impulse to diet or the impulse to use substances in a dangerous manner. In the majority of these situations, the parents will bring the child for treatment and will be concerned about their inability to lead their child out of their self-destructing behavior patterns. Two evidence-based family treatments have been developed for these problems, and their effectiveness has been documented through randomized, controlled treatment trials. These treatments are multidimensional family therapy for adolescent substance abuse and family-based treatment for anorexia. Both of these treatments assign responsibility to the parents for reducing the intensity of the symptomatic behaviors. In family-based treatment for anorexia, the parents are given the responsibility of helping their child gain weight with the therapist's support. As the parents take on this responsibility, the therapist helps them to be successful in attaining the child's attention and the child's compliance. The child gains greater support and autonomy as they participate in treatment. The therapist also supports the child's interest in and competence with activities and with peers. If treatment is initiated early in the course of the disorder, it can be effective and relapse is rare. Family treatment for substance abuse in adolescents is similar. However, often school failure, legal difficulties, a history of poor supervision, and the influence of negative peers complicate the clinical situation. The treatment manual for multidimensional family therapy identifies ways of addressing these challenges while simultaneously providing the supervision, support, and limits necessary to reduce substance use. Once again, the keys to treatment effectiveness are opportunities to behave in prosocial ways with intense family support for competency and success associated with consistent, unwavering expectations and reasonable but not excessive consequences.

Joan was 14 years old and lived with her parents and 12-year-old sister. She was referred to a child psychiatrist by her pediatrician, who had recently evaluated her for significant weight loss and dramatically reduced food intake.

Joan's parents were middle class, and both worked long hours. They had become concerned about her dieting and weight loss but had been unable to have it cease. When Joan and her mother met with the child psychiatrist, Joan stated very clearly that her weight was a major concern to her, that she still felt that she needed to diet and lose weight and that under no circumstances was she going to eat what the pediatrician and nutritionist had recommended. Joan's father was a busy executive who traveled twice monthly and who had grown increasingly distant from his daughters over the last year. Joan's mother stated her concern that her husband was not as involved with his daughters as he should be and that now his sole interaction with Joan concerned her eating. He yelled at her, trying to get her to eat. Joan's mother had thought that her daughter was experiencing anorexia nervosa, and the therapist concurred. The therapist strongly suggested that Joan's mother and father come to a session later that week to discuss a plan to help Joan with her anorexia. The therapist received a phone call from the father stating that he needed to be out of town on business, but the therapist firmly stated that he should postpone his trip and attend the session with his wife and daughter.

The therapist's first efforts in the meeting were to emphasize to both parents the need for them to work together and to assist Joan with rules that would encourage her to follow through with the food recommendations of her pediatrician and her nutritionist.

As the therapist helped Joan's parents plan how they would expect Joan to eat the required amount and gain weight as suggested by her pediatrician, the therapist also suggested that Joan finish her meals before she left the table and pursued other activities or completed her homework. Joan was upset and frightened at the prospect that she would gain weight. The therapist helped the parents to be understanding but firm and to reiterate their expectation of consistent weight gain. The parents were also advised to be explicit about caloric expectations and meal planning.

As Joan gained weight, she gained more freedom to choose which foods the family had in the house, but she was still required to eat the required amount of calories. Her rate of weight gain was consistently 1 lb per week, and she went from 80% of her ideal body weight to 95%. As Joan began gaining weight, the therapist noted increasing distress in her parent's relationship. The therapist suggested private couple sessions for the parents to discuss their relationship. Individual sessions were held with Joan to assist her with assertiveness, peer relationships, and ways of decreasing her perfectionism with regard to her schoolwork.

Joan participated in her individual sessions eagerly and was soon noted to be spending more time with friends and less time preoccupied with food, dieting, and weight. When she reached 100% of her ideal body weight, her menses resumed. The therapist was also meeting regularly with Joan's parents. Joan's mother expressed her concern that her husband was more involved in work than with the family and that this was not all right with her. She became assertive with him, and they began to negotiate a different approach to family and work. They also were advised to spend some time alone together. Joan's father responded well to the couple sessions, and her parent's marriage improved and became more satisfying for both of them.

Over the 8 months that therapy continued, Joan resumed normal adolescence, family relationships improved significantly, and Joan's father became more engaged with his daughters.

Stephanie was a 15-year-old girl who was referred to the substance abuse treatment program because her parents were concerned about her alcohol and marijuana use. She lived with her stepmother and biologic father. Prior to this school year, she had been a successful student and had played in her high school band.

Stephanie's biologic mother had significant trouble with alcohol and drugs when Stephanie was a preschooler. After failed attempts at treatment, Stephanie's father separated from her mother and gained full custody of his daughter. Stephanie knew that her mother was in another state and believed that she was still having trouble with alcohol use. She had not had contact with her mother for 12 months prior to her referral.

Stephanie's father had remarried when Stephanie was 6 years old, and Stephanie had two younger half-siblings, a 7-year-old brother and a 4-year-old sister. This year Stephanie had increasingly felt disconnected from her stepmother and her younger half-siblings. Her relationship with her father had begun to deteriorate as she became more defiant of rules and more involved with peers. She was home less frequently and was less involved with family activities. Stephanie's stepmother had been called by her school when she had been found somnolent and disoriented at school.

Stephanie's stepmother and father took her to the emergency room, where she was found to be intoxicated with alcohol. She said she had gotten the alcohol when she and some friends had gone to a friend's house during lunch period. Stephanie's father and stepmother had been concerned that Stephanie might be using substances frequently and in a problematic way. Over the last several months, she had been coming in late after curfew, staying over at friends' houses frequently, and had been less involved with her previous friends and more involved with a group of friends that the family did not know very well. Stephanie had also dropped out of the band and stopped playing her instrument. Her academic performance had decreased significantly.

Stephanie agreed that she was using drugs and alcohol but steadfastly stated that she did not have trouble with them and that she was only doing what her friends did. When Stephanie, her father, and her stepmother met with the family therapist at the drug treatment center, it became apparent to all that Stephanie's problems were quite significant and were seriously jeopardizing her physical safety as well as her developmental progress.

In the first session Stephanie tearfully admitted to involvement with unprotected sex, especially when she was intoxicated. With guilt and shame she admitted that she did not want to be doing what she was doing but had found it very difficult to refuse substances or affection and physical intimacy when she was with peers.

The family therapist suggested to Stephanie's father that she was scared and lost rather than defiant and delinquent. Stephanie's father, with some difficulty, acknowledged that this might be true. The family met with the therapist three times a week at the program while Stephanie participated in the adolescent after-school drug treatment groups. The therapist helped Stephanie's parents to institute a curfew and ensure adult supervision after dinner and on weekends. Stephanie reluctantly, but in fact with relief, accepted these rules and adhered.

Regular drug testing was part of Stephanie's program, and over the first month of treatment, she was consistently free of evidence of drug and alcohol use. The therapist also strongly recommended that Stephanie attend the adolescent 12-step program meetings with the teens in the treatment group. The therapist also spoke with Stephanie's counselor and teacher at school. They were advised to strongly encourage Stephanie to rejoin the band, resume her musical practice, and participate in tutoring and additional homework to improve her academic standing. Stephanie responded to this with some initial reluctance, but ultimately with the support and encouragement of her teachers, school counselor, therapist, and band leader, she was able, after a rocky period of a couple of weeks, to be more involved in school and more involved with the friends she had distanced herself from after she had gotten involved with her substance-abusing friends.

During this time, Stephanie was also seen by a child psychiatrist. The child psychiatrist reported that Stephanie was sad, but he did not believe that she warranted antidepressant medication. The child psychiatrist told Stephanie, her father, and her stepmother that he was very pleased with the progress that they were making in the intensive substance abuse treatment program.

Stephanie’s mood continued to brighten. She was encouraged to have evening and weekend times with her father and stepmother. Her stepmother was advised to spend time talking with Stephanie about her friends and her activities. She was encouraged to avoid disciplining Stephanie but to inform her father about any concerns. Stephanie responded to this with improved connections with both her father and her stepmother. She also spent more time with her half-siblings.

Three months into treatment, at the time that Stephanie’s discharge from the intensive treatment program was planned, her urine tested positive for alcohol. Her parents also discovered that she had come home after curfew, apparently intoxicated after attending a party without parental permission.

The therapist helped the family to recognize that Stephanie had had a difficult weekend and continued to have difficulty resisting the attraction to alcohol and the encouragement of peers to use substances. The therapist worked intensively for the next 2 weeks to reconnect Stephanie with her father and stepmother and to re-establish more-intensive supervision for Stephanie. The therapist encouraged the father and stepmother to maintain hope for Stephanie’s success in the future. Stephanie asked her parents to be understanding and forgiving of her episode of alcohol abuse. Her father and stepmother realized that they could not yet expect Stephanie to be able to make mature decisions when exposed to substances and that they would need to provide intensive supervision.

The therapist also spoke with Stephanie's father and stepmother about the difficulties they were having as a blended family and worked with them to understand the challenges of that experience. They worked on building one family of five rather than two families, one with four members and another with three members. At the same time, the therapist spoke with Stephanie about her experience of abandonment and hurt in her relationship with her mother and about her worries about her mother's well-being. Stephanie talked about her fear that she would grow up to be like her mother. As Stephanie held these conversations with her therapist, she was able to be increasingly mature and thoughtful about herself and her previous experience of neglect and abandonment. She also was encouraged to speak about this with her father in a session.

The involvement of the entire family in treatment, the appreciation of the family's history and experiences, Stephanie's involvement in a teen substance use program, and the active involvement of her school had all been elements in treatment. Individual therapist support to help her to understand and accept her difficult relationship with her mother was also essential. All of this helped the family to bond together more openly and to be increasingly successful in the developmental challenges that they were facing.

Family Treatment for Divorcing and Blended Families

Family transitions occurring as a result of parental separation and divorce and remarriage are disruptive to children. If parents are able to be congenial with each other and to amicably resolve issues related to the divorce (including financial issues, grieving the end of the marriage, and child care and custody issues), the children, though they still grieve the end of their family and accommodate the changes to family life with sadness, will usually do so without significant relationship problems and persistent behavior problems.

However, if significant resentment, anger, and conflicts continue between the former partners, the effect on the children can be significant and persistent. Problems with academic performance and mood and behavior problems can develop. The conflicting parents often do not manage these difficulties successfully, and family treatment can be necessary. The first step in this treatment process is to identify and support the concern that each parent has for the children and for resolving the children’s problems. Parents may be motivated by different concerns—for example, one by a child’s academic difficulties and the other by the child’s sadness. In any event, bringing together antagonistic parents is an important element in treatment. The family therapist should work extremely hard to help the parents to understand that their children need both of them and need them to work together in parenting for the children’s well-being. This process may be difficult and time-consuming, but it provides the best foundation for the children’s successful outcome. As the therapist helps the biologic parents work together, they can also help the children express their concerns with each parent individually in separate therapy sessions.

Several aspects of post-divorce life can be problematic for children. Often a parent can be very involved in their life, including work and relationship building. This can lead the children to be unsupervised and to further engage in problematic behaviors. To address this, both parents (if possible) will need to supervise the children and reiterate their belief in the children’s strengths and capacity to be successful.

When stepparents are involved in problematic post-divorce situations, matters can worsen. Stepparents may decide that enhanced discipline is needed and may attempt to act to deliver that. As the stepparent works to impose discipline, a child often becomes indignant and defiant. The more forceful and demanding the stepparent becomes, the more defiant the child becomes. Family therapy can reorganize this situation and lead both biologic parents to take a central role in the child’s life and success and address challenging behaviors. Family therapy can also help stepparents to take a peripheral but very important role as supporter and ally for both the child and biologic parent. The stepparent can express concerns and worries about the child, but rather than to the child directly, these concerns can be expressed to the biologic parent, who can act on them in concert (if possible) with the biologic parental partner. The therapist can also encourage the stepparent to develop a warm, accepting, and caring relationship with the child based on mutual interest. This decreases the child's antagonism toward the stepparent and may significantly improve the mood of the family while also enhancing trust and attachment between the child and the biologic parent.

As time passes, the child can adapt to the changing family and maintain a supportive and engaged relationship with both biologic parents and accept stepparents and step-siblings. The therapist can also help the new partners build their relationship while they manage successful relationships with all children and with former spouses.

Kevin was an 11-year-old boy who lived with his mother, maternal grandmother, 17-year-old sister, and 14-year-old high-functioning autistic brother. Kevin was referred to see a child psychiatrist because his mother was concerned about his excessive worries. The whole family had been living with their maternal grandmother since Kevin's parents separated and divorced when he was a year old. Post-divorce, the parents had a long history of custody battles. At the time of the initial visit, the parents had shared legal custody, but the mother had physical custody. Their divorce agreement specified that if there is a disagreement regarding medical and academic needs and decisions, the mother's decision supersedes the father's preference. The father's consent to provide psychiatric treatment was received, though he did not see a reason for an evaluation. Kevin and his brother would stay with their father every other weekend and have dinner with him at least once a week.

Kevin's mother wanted him to be engaged in treatment because she believed that he suffered during the custody battles. He always had difficulty academically, but his performance deteriorated when his sister refused visitation with their father 3 years ago. When he was anxious, he would be short of breath and would pace around the house and scream. This would occur specifically when he was struggling with his homework and when he was supposed to visit his father. This behavior would take place every 3 to 4 days. Nightmares would wake him up shaking in the middle of the night.

His father was invited to participate in the evaluation and treatment. Initially, his involvement was limited, but when he realized that some of Kevin's difficulties were related to his upcoming wedding, he became more engaged.

Kevin and his father were encouraged to engage in conversations to communicate his difficulties with "his growing family." Kevin was able to feel safe in therapy to inform his father of his grievances, including his concerns about his stepmother taking over his mother's role. He worried about his mother being alone while also grieving the permanency of their divorce. With the support and involvement of his father's fiancé, his mother and his father were able to hear his concerns, to encourage him to speak-up, and to compromise when necessary. They were able to validate the loss of his original family but also to reinforce that they all had a special relationship with him individually. He was reassured that these relationships were going to remain unchanged and that one relationship could not take the place of the other. The family members were able to discuss their expectations of him in both households. They agreed to keep the rules and these expectations consistent so that it would be less confusing for Kevin. As all relationships improved, Kevin's anxiety and difficulty in school were resolved.

Family Therapy for Families from Diverse Backgrounds

Family therapists, and all providers, should be sensitive to the family's cultural norms in order to provide the best care. When discussing diversity, one often thinks of ethnicity and race exclusively. Concerns about diversity must also include differences in socioeconomic status, sexual orientation, and nontraditional family units (i.e., same-sex parents). Each family has its own culture and values that define its norms and guiding principles. Culture is the foundation that dictates one's sense of how to be, what is "normal," right and wrong, and reactions to others. The degree of assimilation to the culture of the local area (community, region, or nation) plays an important role in the family's goals for the treatment. A family's culture also influences the degree of mental health stigma and directly impacts the level of trust in mental health care. Family therapy may be tailored to increase engagement with the family to overcome distrust, mental health stigma, and past experiences of being misunderstood, misperceived, or discriminated against. Being curious about the family's history of hardships, trauma, losses, and experiences of discrimination and exclusion early in treatment helps the therapist understand the family.

As the therapist engages with a family of a different culture, they will need to be aware of how the family meets three challenges that families from diverse background experience. The first challenge is to assimilate into the culture of the host country without losing their traditional values. The second challenge is to balance passing on their cultural traditions to the next generation with supporting the youth's success in the western culture. The third challenge is to manage how different family members adapt to the mainstream culture at a different pace which can result in acculturative family distancing (AFD). AFD can increase family conflicts while decreasing family cohesion through miscommunications and misunderstandings among family members.

Family therapy addresses where each family member is in their acculturation process. The therapist will also need to appreciate each member's expectations from treatment, the caretakers' parenting approach, the strengths of parent–child relationships, and parent–child conflict. The therapist can navigate AFD by highlighting the love and affection between parent and child and by creating an atmosphere of mutual acceptance and compromise, a both/and approach that is based on connection and shared commitment to family life and the development of the child's independence. By elaborating on the narratives of the parents and the child, the therapist increases mutual understanding, further fostering trust and collaboration. The therapist's task is to honor and respect how the family understands their differences. By recognizing and elaborating on the different ways family members express their love, the therapist can open the door for acceptance and growth.

Success with diverse families is defined by the family's ability to promote the value of success in America as well as their traditional cultural values, traditions, and rituals while also dealing with the presenting concerns.

Ana is a 17-year-old Mexican-American girl who was sent to the ER from her school. At school, her English teacher had raised concerns because of an essay that she submitted for an assignment. The assignment was to share reflections on a difficult moment. Ana shared her challenges at home and her struggle with living in two worlds that sometimes have opposing expectations. She also shared that she felt trapped and wished there was a way to escape her situation. Subsequently, the teacher reached out to the guidance counselor, who spoke to Ana in depth. When probed, Ana admitted to having thoughts about walking onto a high-traffic highway. She was unclear whether she wanted to die, but she was clear that she found her situation to be unbearable. The guidance counselor also felt that Ana had a flat affect, a depressed mood, and was socially isolated. The teacher had also endorsed her declining grades and decline in the quality of her class and homework. Following the school protocol, the guidance counselor called the parents, shared her concerns, and informed them that Ana could not return to school unless a psychiatrist had deemed her safe to return.

This information was surprising to the parents. As suggested by the pediatrician and the guidance counselor, they took their daughter to the nearest hospital. After prolonged waiting in the emergency room, they spoke to the psychiatrist. Ana shared her story with significant shame and guilt. She felt that she was burdening her parents, who worked hard to take care of her and her siblings. She also knew that her father had to request time off from work to be with her. Though the psychiatrist was thoughtful, the parents felt attacked, as if they were too demanding and uncaring about their daughter. They also felt a bit ashamed that they were not aware of the degree of stress on their daughter. Once the psychiatrist understood the family's concerns, she referred the family for an outpatient family therapy visit with a child psychiatrist at the same institution. She also informed the outpatient psychiatrist of the parent's experience in the emergency room and at the school.

At the first visit, it became clear that though the mother spoke English, a Spanish interpreter would be preferred so she could share her thoughts more freely. It was also apparent that the parents felt forced to engage in this service and were not happy about sharing their family stories with a stranger. Ana was palpably aware of these concerns and appeared shut down. The child psychiatrist understood these challenges and tried to gain trust by asking about their goal for the meeting. Mother and father shared that they want Ana to be safe, be less stressed but also acknowledge that the family is the most important. For them, family included her siblings but also grandparents, aunts, uncles, cousins, and sometimes even family friends. Ana's goal was to feel normal among her friends at school and be accepted by her parents. She felt that too often, her actions were divided between "an American" or "a Mexican" way of living.

Through the second, third, and fourth visits, the psychiatrist was humbly curious about the parent's experience in Mexico, their journey to America, and their hopes for their children. Ana's father came to America in his 20s in search of opportunities. He struggled with poverty and the loss of family support upon coming to America. He worked hard at a retail store, a restaurant, and as a truck driver to support his family and to make his parents proud of him. Throughout his time in America, he had been discriminated against and had to deal with prejudice. But, with hard work and dedication, he was able to bring his other siblings to America. Together, they were able to open a small business and navigate the ups and downs of the economic challenges. He married his wife when he was visiting his family in Mexico. Her transition to America was very isolating since he worked long hours, she did not have a Visa to work, and she did not speak English. Though they had been in America for over 20 years, they had adhered to their cultural values to retain their "way of life." They spent a significant amount of time with their own family members (including Ana’s uncles) as a way to be protective of their culture.

The family therapist learned that Ana’s parents valued family time, gender roles, and importance of cultural involvement as they had learned prior to immigration. Ana’s parent’s expectation of “family first”—taking care of brothers’, helping with chores, and being involved with the Mexican American community—conflicted with her own desires to be more like her peers. She enjoyed American pop music and wished that her family would be accepting of her new white American boyfriend. They worried that Ana would drift away from them as she was more comfortable American culture. Furthermore, Ana would often be playing the role of translator for her mother and explaining the American cultural norms to her. This situated her in a unique position of supporting her parents to navigate American culture but, also increased the burden on her as well. She felt that they didn’t understand her hopes to feel “normal” and not be isolated from and different among her peers. She also felt that the more she tried to fit- in among her peers the more isolated, disconnected, and she was in her family.

Over time, the psychiatrist was able to appreciate the parents’ worries about losing their traditions and language (as Ana preferred to speak in English rather than Spanish.) He also recognized that the parents were completely dedicated to the well-being of their daughter and her future success. With the psychiatrist’s support, parents were able to appreciate Ana’s burden of straddling two cultures. They began to be more lenient with their expectations and allowed Anna to spend more time with her friends and boyfriend. Though they had still not met the boyfriend, they were more curious about him. They were accepting that Ana would only flourish if they embraced both cultures. The psychiatrist spent a significant amount of time processing their grief over losing some traditions while supporting their unique ways of retaining Mexican cultural values.

Ana was also encouraged to appreciate her parents’ challenging journey to America in search of more opportunities and privileges for her and her siblings. She was able to respect their dilemma in holding on to their Mexican roots while trying to raise children with Mexican values and American values. Ana was proud of her parents for creating their own path to success and opportunities with limited support. She was inspired by their resilience, their dedication, and their love. This also resulted in her feeling less isolated now that her parents were accompanying her in navigating both American and Mexican cultures.

The child psychiatrist continued to support open communication between Ana and her parents. Ana reported improved mood at home and was less irritable with her parents. She also improved her grades and her participation in school. Therapy sessions were now taking place less frequently but continued meetings between Ana and her parents led to enhanced understanding, improved communication, and greater support for all to manage a bicultural lifestyle.

Family Therapy for Families with Children with Developmental or Medical Challenges

Families with children with developmental delay, autistic spectrum disorders, or chronic medical illness face unique challenges. Family therapy can help those families who are having difficulty managing emotions, behavior, and illness treatment or achieving the optimal development of the affected child or siblings. These situations often are extraordinarily sad and difficult to deal with. There can be financial challenges, career advancement can be negatively affected, illness or handicap care routines can be onerous, and opportunities for parental intimacy and family enjoyment can be threatened. Parental burden and stress can be extreme, and at times the expected positive effects of medical treatment may not occur. Marital relationships may be threatened, and siblings may not receive adequate attention and supervision. Parental isolation may also seriously jeopardize family well-being. Family therapy can address these difficulties by offering the parents, the affected child, and siblings a safe environment in which to discuss reactions to the child’s problems and to collaborate on making the situation the best that it can be. Parental cooperation and parental mutual understanding are goals. All those involved can be guided by the therapist to learn about the child’s condition, current and potential treatments, and prognosis. The parents, the child, and siblings can be encouraged to ask questions and to be assertive with the medical care team. The family and child can be encouraged to recognize the areas of the child’s life that are not affected by the illness or handicap and the child’s strengths and abilities. The therapist can also recognize the courage and care of all in dealing with the challenges of the child’s condition. As the family recognizes their commitment to the affected child, they can come to see themselves as heroic and resilient, and the child can note their courage and competency regardless of the illness or handicap. The therapist can help the family work with the school and teachers and with extended family members to understand and support the affected child while also helping the child engage teachers, neighbors, peers, and extended family in an assertive, positive, and confident manner. Finally, the child’s active participation in their care can be noted and recognized as a sign of competence and mastery that will help the child throughout life. The family therapist can also support the family if the child’s condition fails to improve or deteriorates. The therapist elicits and ratifies legitimate responses of sadness and powerlessness and accepts their powerlessness over the illness while staying involved with and attuned to family members’ sadness and grief.

Dave is a 9-year-old boy who lives with his parents and 1-year-old brother. He had been undergoing treated for high-functioning autism spectrum disorder, attention-deficit/hyperactivity disorder, and anxiety disorder for 2 years. His treatment included a low-dose atypical antipsychotic, a low-dose selective serotonin reuptake inhibitor, a stimulant, and an alpha agonist. Additionally, he was receiving individual therapy every 2 weeks to increase his coping skills. The family had been referred for family therapy but, due to both parent's full-time jobs and other commitments, they had not engaged in that treatment. The family self-referred to a different child psychiatrist because they felt that Dave's treatment had become stalled.

The family's chief complaint was Dave's explosive behavior around limit setting at home and social difficulties with his peers. His father struggled to follow through with consequences, whereas his mother was always able to do it. This inconsistency had resulted in confusion for Dave and also increased conflict between the parents. Dave's mother felt that she was not being supported by her husband. Dave's father felt that his wife was very strict with Dave and would attempt to compensate for this strictness. He felt that Dave's developmental delay made school so difficult for him that he deserved to have a less demanding environment at home. As the parents talked with the therapist they realized the need to negotiate expectations and consequences for Dave. They also recognized the need to support each other more and redevelop the warmth in their relationship. Although Dave remained difficult and challenging, he gradually improved and became much less explosive.

FUTURE DIRECTIONS

Family therapy has evolved from a therapy developed in opposition to common psychiatric care to a basic orientation to the treatment of child mental health challenges. All situations require the assessment of the involvement and education of family members. Some situations require the specific and monitored engagement of family members in the treatment plan. In many instances, treatment of the child requires the involvement of the family in the manner outlined in the manual of an evidence-based model of family intervention. Knowledge of these treatments and skills implementing them are necessary elements in child mental health care. For many child mental health problems, the problem leads to intuitive family interactions that are counterproductive and often worsen the situation. Family therapy stops these negative cycles and can introduce successful, self-reinforcing, positive family interactions that promote family life and optimal child development. These interventions can help the child's situation to improve, and the family's experience of warmth, care, and effectiveness can emerge and persist.

FURTHER READINGS

Baldwin S, Christian S, Berkeljon A, Shadish W. The effects of family therapies for adolescent delinquency and substance abuse: a meta-analysis. *J Marital Fam Ther*. 2012;38(1):281–304.

Barkley R. *Defiant Children. A Clinician's Manual for Assessment and Parent Training*. 2nd ed. Guilford; 1997.

Boyd-Franklin N. *Black Families in Therapy: Understanding the African American Experience*. 2nd ed. Guilford; 2006.

*Combrinck-Graham L. *Children in Family Contexts: Perspectives on Treatment*. 2nd ed. Guilford; 2006.

Coolhart D, Shipman DL. Working toward family attunement: family therapy with transgender and gender-nonconforming children and adolescents. *Psychiatr Clin North Am*. 2017;40(1):113–125.

*Diamond G, Josephson A. Family-based treatment research: a 10-year update. *J Am Acad Child Adolesc Psychiatry*. 2005;44:872–887.

Diamond G, Rosson J, Levy S. Attachment-based family therapy: a review of the empirical support. *Fam Process*. 2016;55(3):595–610.

Fisher CA, Skocic S, Rutherford KA, Hetrick SE. Family therapy approaches for anorexia nervosa. *Cochrane Database Syst Rev*. 2019;5(5):CD004780.

Handen B, Johnson C, Butter EM, et al. Use of a direct observational measure in a trial of risperidone and parent training in children and pervasive development disorders. *J Dev Phys Disabil*. 2013;25(3):355–371.

Henggeler SW, Schaeffer CM. Multisystemic therapy: clinical overview, outcomes and implementation research. *Family Proces*. 2016;55(3):514–528.

Henggeler S, Sheidow A. Empirically supported family-based treatments for conduct disorder and delinquency in adolescents. *J Marital Fam Ther*. 2012;38(1):30–58.

Jewell T, Blessitt E, Stewart C, Simic M, Eisler I. Family therapy for child and adolescent eating disorders: a critical review. *Fam Process*. 2016;55:577–594.

Josephson AM; AACAP Work Group on Quality Issues. Practice parameter for the assessment of the family. *J Am Acad Child Adolesc Psychiatry.* 2007;46:922–937.

Kazdin H, Weisz J. *Evidence-Based Psychotherapies for Children and Adolescents*. Guilford; 2003.

Lewis AJ. Attachment-based family therapy for adolescent substance use: a move to the level of systems. *Front Psychiatry*. 2020;10:948.

Liddle HA. Multidimensional family therapy: evidence base for transdiagnostic treatment outcomes, change mechanisms, and implementation in community settings. *Fam Process*. 2016;55:558–578.

Lock J, Le Grange D, Agras W, Moye A, Bryson S, Jo B. Randomized clinical trial comparing family-based treatment to adolescent-focused individual therapy for adolescents with anorexia nervosa. *Arch Gen Psychiatry*. 2010;67(10):1025–1032.

*Madsen W. *Collaborative Therapy with Multi-Stressed Families: From Old Problems to New Futures*. Guilford; 2003.

McFarlane WR. Family interventions for schizophrenia and the psychoses: a review. *Fam Process*. 2016;55:460–482.

*McGoldrick M, Giordamo J, Garcia-Preto N. *Ethnicity and Family Therapy*. 3rd ed. Guilford; 2005.

McGoldrick M, Hardy K. *Re-visioning Family Therapy: Race, Culture and Gender in Clinical Practice*. 2nd ed. Guilford; 2008.

Miklowitz DJ, Chung B. Family-focused therapy for bipolar disorder: reflections on 30 years of research. *Fam Process*. 2016;55(3):483–499.

Miklowitz DJ, Schneck CD, George EL, et al. Pharmacotherapy and family-focused treatment for adolescents with bipolar I and II disorders: a 2-year randomized trial. *Am J Psychiatry*. 2014;171(6):658–667.

Minuchin P, Colapinto J, Minuchim S. *Working with Families of the Poor*. 2nd ed. Guilford; 2006.

Minuchin S, Lee W, Simon G. *Mastering Family Therapy: Journeys of Growth and Transformation*. 2nd ed. Wiley; 2006.

*Nichols M, Schwartz R. *Family Therapy: Concepts and Methods*. 8th ed. Allyn & Bacon; 2007.

Pineda J, Dadds M. Family intervention for adolescents with suicidal behavior: a randomized controlled trial and mediation analysis. *J Am Acad Child Psychiatry*. 2013;52(8):851–862.

Robbins MS, Alexander JF, Turner CW, Hollimon A. Evolution of functional family therapy as an evidence-based practice for adolescents with disruptive behavior problems. *Fam Process*. 2016;55(3):543–557.

Sargent J. Family therapy in child and adolescent psychiatry. *Child Adolesc Psychiatr Clin North Am*. 1997;6:151.

Sharma N, Sargent J. Diverse families and family treatment. In: Parekh R, Al-Mateen CS, Lisotto MJ, Carter RD, eds. *Cultural Psychiatry with Children, Adolescents, and Families*. American Psychiatric Association Publishing; 2021:151–161.

Sharma N, Sargent J. Overview of the evidence base for family interventions in child psychiatry. *Child Adolesc Psychiatr Clin North Am*. 2015;24(3):471–485.

Sholevar C, Schwoeri L. *Textbook of Family and Couples Therapy: Clinical Applications*. American Psychiatric Publishing; 2003.

Spain D, Sin J, Paliokosta E, et al. Family therapy for autism spectrum disorders. *Cochrane Database Syst Rev*. 2017;5(5):CD011894.

Waldron H, Brody J. Functional family therapy for adolescent substance use disorders. In: Weisz JR, Kazdin AE, eds. *Evidence-Based Psychotherapies for Children and Adolescents*. 2nd ed. Guilford Press; 2010:401–415.

*Walsh F. *Normal Family Processes: Growing Diversity and Complexity*. 3rd ed. Guilford; 2006.

*White M. *Maps of Narrative Practice*. W. W. Norton; 2007.

53.5 Pediatric Psychopharmacology

Jeffrey R. Strawn, M.D., and Laura B. Ramsey, Ph.D.

Psychotropic medications play a critical role in treating many psychiatric disorders in children and adolescents. However, using these medications effectively and appropriately requires an understanding of pediatric pharmacokinetics and pharmacodynamics, in addition to the nuances of response that are revealed by numerous double-blind, placebo-controlled trials. In this chapter, we review the developmental pharmacology that is relevant to psychotropic medication use in pediatric populations and the evidence for multiple psychotropics across internalizing and externalizing disorders.

DEVELOPMENTAL PHARMACOLOGY

Pharmacokinetics and pharmacodynamics in children and adolescents differ considerably from adults, and these differences give rise to differences in dosing, differences, response, and differences in tolerability. Importantly, pharmacokinetic and pharmacodynamic factors interact (Fig. 53.5–1) and are mutually influenced by development. These pharmacokinetic factors and pharmacodynamic factors subtend differences in side effects (e.g., extrapyramidal symptoms), dosing (e.g., lithium), and efficacy (e.g., benzodiazepines) between children, adolescents, and adults.

Pediatric Pharmacokinetics

In general, metabolic capacity at birth is less than in adults and increases to levels seen in adults by approximately 2 years of age. However, near pubarche, metabolism may be greater than what is commonly observed in adults. In addition to developmental differences in hepatic metabolism, renal function, and changes in fat distribution during childhood and adolescence affect the pharmacokinetics of many medications in youths.

Also, children and adolescents have greater total body water and relatively less adipose tissue compared to adults. This affects the distribution of many medications and the accumulation of lipophilic medications (and their metabolites).

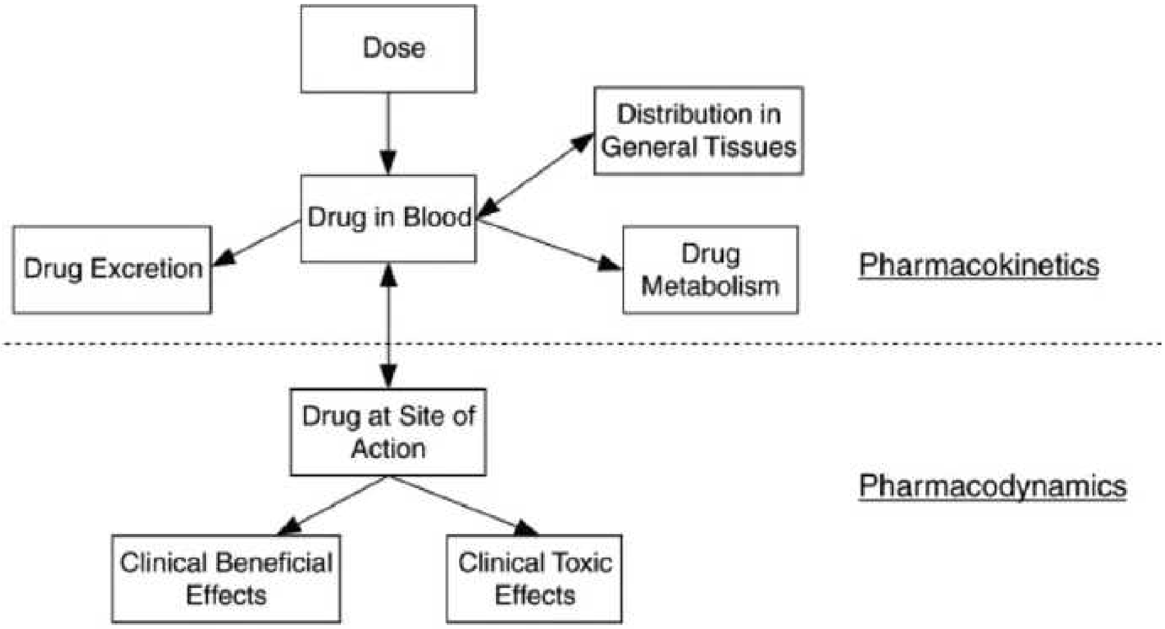
Dose
Distribution in General Tissues
Drug in Blood
Drug Excretion
Drug Metabolism
Pharmacokinetics
Drug at Site of Action
Pharmacodynamics
Clinical Beneficial Effects
Clinical Toxic Effects

FIGURE 53.5–1. Schematic representation of the components that comprise the pharmacokinetic-pharmacodynamic interface. (Reprinted with permission from Capparelli EV. Clinical pharmacokinetics in infants and children. In: Aranda JV, van den Anker JN. *Yaffe and Aranda's Neonatal and Pediatric Pharmacology: Therapeutic Principles in Practice*. 5th ed. Wolters Kluwer; 2020.)

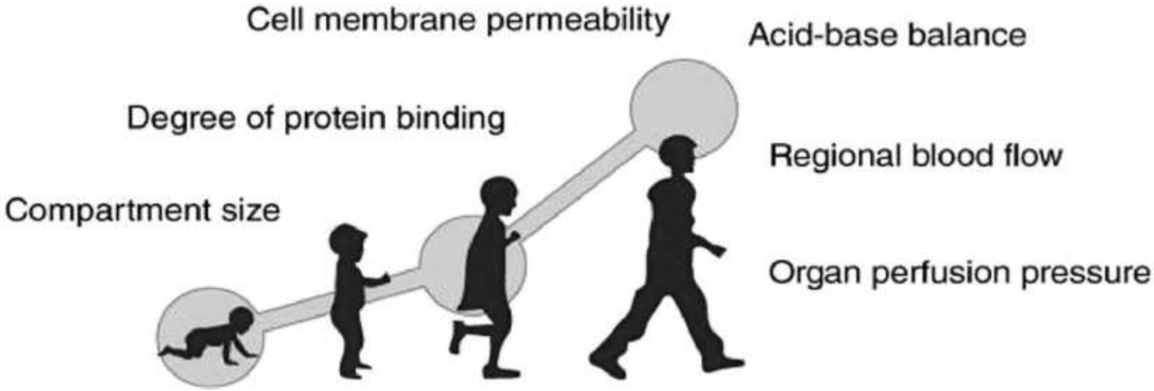

FIGURE 53.5–2. Factors affecting the rate and extent of drug distribution. (Reprinted with permission from Stancil SL, Chapron A, Abdel-Rahman AM. Drug absorption, distribution, metabolism, excretion, and transporters in newborns and children. In: Aranda JV, van den Anker JN. *Yaffe and Aranda's Neonatal and Pediatric Pharmacology: Therapeutic Principles in Practice*. 5th ed. Wolters Kluwer; 2020.)

Absorption and Bioavailability. Absorption is influenced by the route of administration of a drug and, for orally administered medications, is influenced by gastric pH and the presence of food in the stomach. Bioavailability is the fraction of the drug that reaches the systemic circulation and is also influenced by concurrently administered medications and first-pass metabolism.

Drug Distribution. Drugs are distributed throughout the body through the bloodstream, and distribution of psychotropic drugs to the site of action is affected by protein binding and transport across the blood-brain barrier. The volume of distribution (V_D) reflects the degree to which it is distributed in body tissue (rather than the plasma). Larger V_D indicates a greater amount of tissue distribution. Smaller VD suggests that the drug stays primarily in the plasma. Drugs with high lipid solubility (e.g., some benzodiazepines), low rates of ionization, or low plasma protein binding have higher volumes of distribution compared to drugs which are more polar, more ionized or those that have more plasma protein binding and differs across development (Fig. 53.5–2). Additionally, age-dependent changes in body composition can also influence distribution.

Metabolism. Many psychotropic medications are metabolized by cytochrome P450 enzymes CYP3A4, CYP2C19, or CYP2D6 in the liver. The activity of CYP2D6 is fairly constant from 1 year of age through adulthood, but the activity of CYP2C19 and CYP3A4 may be increased in children compared to adults. Drug–drug interactions can occur through inhibition or induction of these enzymes. In addition, in younger children, hepatic clearance of drugs may be increased as liver blood flow is increased compared with adults. This is largely related to the larger ratio of liver to total body mass in the children compared to adults and adolescents (Fig. 53.5–3) and, as a result, the first-pass effect of some drugs may be increased.

Excretion. Most drugs and their metabolites are excreted through the kidney or liver. Renal clearance is substantially higher in pediatric populations compared to adults and decreases from ages 2 to 20 years.

Pharmacogenetics in Children and Adolescents

Variants in the genes encoding the metabolizing enzymes (e.g., CYP2C19 and CYP2D6) influence the enzyme activity, so that patients can be poor metabolizers (with little to no activity), intermediate metabolizers, normal metabolizers, rapid metabolizers, or ultrarapid metabolizers (with more than normal activity). Dosing recommendations based on these metabolizer groups are available from the Clinical Pharmacogenetics Implementation Consortium, the Dutch Pharmacogenetics Working Group, and some are included on the FDA labels. The CYP2C19 and CYP2D6 enzyme activity affects the systemic concentration, which can be normalized with dose adjustments, using lower doses in poor metabolizers and higher doses in ultrarapid metabolizers. Individual genes should be used to adjust medication dosing rather than using combinatorial pharmacogenetic testing for medication selection.

ROLE OF PHARMACOTHERAPY IN CHILDREN AND ADOLESCENTS

While psychotropic medications often play an integral part in the treatment of children and adolescents with psychiatric disorders, they are usually used in combination with other modalities. Education about diagnosis, course of illness, prognosis, and available treatment options is important. Additionally, depending on the individual, a treatment plan may warrant individual, group, and/or family therapy; educational interventions; parent training; substance use disorder treatment; or any of several additional interventions.

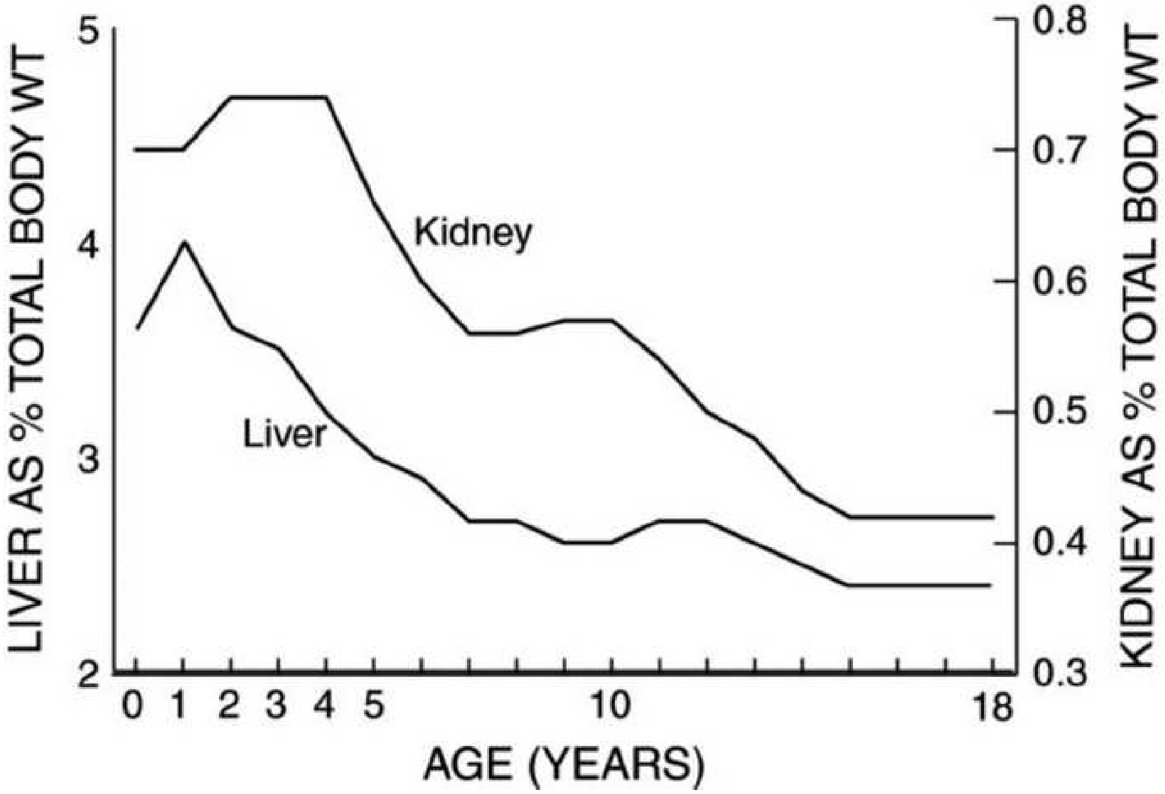

FIGURE 53.5–3. Change in relative liver and kidney mass expressed as a percentage of body weight (WT) from infancy to young adulthood. (Reprinted with permission from Kauffman RE. Drug action and therapy in the infant and child. In: *Yaffe SJ, Aranda JV. Neonatal and Pediatric Pharmacology: Therapeutic Principles in Practice*. 4th ed. Wolters Kluwer; 2011.)

With all disorders and symptoms encountered in children and adolescents, thorough and developmentally sensitive assessment is critical prior to and throughout psychopharmacologic treatment. Such treatment plans are developed from comprehensive and accurate diagnostic evaluations, consideration of comorbidity, subsyndromal symptoms, family factors, personality/temperament, and learning issues (including learning disorders and learning weaknesses). Failure to address comorbidities or other factors may limit the benefits of pharmacotherapy or any therapeutic intervention.

Additionally, rating scales can be used to inform screening and diagnosis of some psychiatric disorders, although these are not a substitute for full clinical assessment by trained clinicians. Standardized measures generally provide normative data that allows a clinician to compare a particular patient's symptoms to those of children of similar age. In each section, common symptom rating scales will be discussed in addition to the evidence supporting their use.

ANXIETY DISORDERS

Selective Serotonin Reuptake Inhibitors in Anxiety Disorders

Multiple meta-analyses suggest that selective serotonin reuptake inhibitors (SSRIs) are superior to other medication classes in pediatric anxiety disorders, but patients are more like to discontinue SSRIs because of adverse events compared to SNRIs and SSRIs are more likely to produce activation in children and adolescents with anxiety disorders compared to SNRIs. To date, sertraline, fluoxetine, escitalopram, fluvoxamine, and paroxetine have been evaluated in pediatric anxiety disorders. These individual clinical trials are reviewed in detail elsewhere.

In pediatric patients with anxiety disorders, the probability of SSRI-related improvement increases over time although response is, overall, logarithmic—70% of the improvement, observed at week 8, occurred in the first 4 weeks of treatment. This trajectory may not be unique SSRIs in pediatric anxiety disorders as SSRIs appear to produce similar, early improvement in youth with major depressive disorder (MDD) and OCD. Additionally, compared to SNRIs, SSRIs produce faster and greater improvement. Only 40% of the treatment response observed for SSRIs occurs in patients receiving SNRIs by the eighth week of treatment and the divergence in SSRI–SNRI trajectory emerges near the fourth week of treatment.

SNRIs in Pediatric Anxiety Disorders

Several studies have examined venlafaxine in youth with generalized (κ = 2, pooled analysis) and social anxiety disorders (κ = 1) as well as duloxetine in generalized anxiety disorder (κ = 1) and atomoxetine in pediatric patients with ADHD and comorbid generalized, separation and/or social anxiety disorders. These studies have generally suggested benefit compared to placebo. However, meta-analyses of SSRIs and SNRIs in youth with anxiety disorders suggest that SSRIs produce faster, and greater improvements compared to SNRIs.

Other Agents in Pediatric Anxiety Disorders

While SSRIs are the most commonly prescribed initial medications for children and adolescents with anxiety disorders in the United States and approximately half of these youth continue treatment for at least 6 months. However, nearly one-third of anxious who are treated with medication in the United States begin treatment with a non-SSRI medication.

Benzodiazepines in Pediatric Anxiety Disorders. Despite evidence for the use of benzodiazepines in the acute treatment of anxiety in youth, studies of chronic use of benzodiazepines in youths with generalized, separation, and social anxiety disorders have failed to show benefit.

α_2-Agonists in Pediatric Anxiety Disorders. The α_{2A}-agonist, guanfacine has been evaluated in youth with generalized, separation, and/or social anxiety disorder. Extended-release guanfacine was well tolerated although somnolence and fatigue were more frequent in GXR-treated patients compared to those receiving placebo. Differences in response were observed between placebo and extended-release guanfacine, although statistically significant differences were not observed for the dimensional measure of anxiety.

5-HT$_{1A}$ Modulators in Pediatric Anxiety Disorders. Buspirone was evaluated in one flexibly dosed (n = 227) and one fixed-dose (n = 341) trial in children and adolescents with generalized anxiety disorder. Buspirone was not statistically superior to placebo in any of these trials. Discontinuation related to an adverse event occurred more commonly in patients treated with buspirone relative to those treated with placebo.

Tricyclics in Pediatric Anxiety Disorders. Two randomized, placebo-controlled trials evaluated imipramine in pediatric anxiety disorders. In the first study, imipramine was superior to placebo in the study's primary outcome measure (response rate: 81% vs. 47%, $p < .05$). With the exception of dry mouth, no side effects were significantly more common compared to the rates observed in patients receiving placebo. However, a second study of imipramine in children with SAD failed to replicate the earlier study's findings. Imipramine was associated with behavioral activation (e.g., "angry outbursts," irritability, etc.) and dry mouth was commonly observed.

Tricyclics (i.e., tricyclic antidepressants, TCAs), have fallen out of favor since the introduction of SSRIs and SNRIs due to comparatively poorer tolerability, including anticholinergic side effects, the need for electrocardiographic monitoring as well as their narrow therapeutic index that results in an increased risk of lethality in TCA overdoses.

SSRIs + Psychotherapy in Pediatric Anxiety Disorders. Few studies have examined CBT in comparison to an SSRI or in combination with an SSRI in pediatric anxiety disorders. CAMS (*n* = 488) found CBT + sertraline to be more effective than CBT or sertraline monotherapy. Remission rates for CBT + sertraline were superior to CBT (66% vs. 35%) after 12 weeks of treatment. Yet, youth in CBT continued to improve relative to COMB such that by 24 weeks, the gap in remission rates narrowed (65% vs. 45%) with no significant difference by week 36 (67% vs. 58%). Interpretation of the effects of CBT in CAMS has always been complicated by design decisions that may have diminished the effects of CBT: (1) CBT was shortened to conform to the ethically acceptable duration of pill placebo exposure; (2) sessions focused on exposure tasks; the key ingredient in CBT were reduced. (3) At the end of CAMS' 12-week acute phase, regardless of remission status, CBT participants entered maintenance phase with a reduced visit schedule, and no initiation of any new CBT treatment potentially limiting restricting potential gains from weeks 12 to 24 and beyond. (4) CAMS CBT had no parent involvement until well into treatment, and even then, it was limited, an approach that is at odds with current practice. In youth with social anxiety disorder, Social Effectiveness Therapy for Children (SET-C) + fluoxetine and SET-C both improved anxiety symptoms relative to placebo.

Select Key Studies

The largest study of an SSRI in pediatric patients with anxiety, the Child/Adolescent Anxiety Multimodal Study (CAMS), compared sertraline (*n* = 133), cognitive-behavioral therapy (CBT) (*n* = 139), combination therapy (CBT + sertraline) (*n* = 140), and placebo (*n* = 76). CAMS included youth with separation, generalized, and/or social anxiety disorders. In CAMS, sertraline was initiated at 25 mg/day and titrated to 200 mg/day with a mean dose of 134 ± 60 mg per day in the combination-therapy group and 146 ± 61 mg per day in the sertraline group. More than half of sertraline patients (55%) exhibited improvement based on Clinical

Global Impression-Improvement (CGI-I) scores while 81% of youth treated with cognitive-based therapy (CBT) + sertraline met response criteria for ($p < .001$) (compared to 60% in patients receiving CBT ($p < .001$). All active treatments were statistically superior to placebo. While this is the largest randomized controlled trial of any psychopharmacologic or psychotherapeutic combination (or their combination) in pediatric patients with anxiety disorders, there was no CBT+ placebo treatment. Some have suggested that the “absence of such a group prevented the investigators from determining whether the addition of sertraline to cognitive behavioral therapy resulted in more improvement than each treatment given separately because of an additive effect of two active treatments or because of the placebo effect of adding a pill to cognitive behavioral therapy.” In this regard, those receiving CBT in conjunction with a tablet knew they were receiving active medication, thus accentuating the expectation of treatment success in patients receiving sertraline + CBT.

PHARMACOTHERAPY FOR MAJOR DEPRESSIVE DISORDER

Multiple RCTs support the efficacy of fluoxetine and escitalopram in youths with MDD—and both are FDA approved for MDD in pediatric patients. There is also support for the use of medication in conjunction with CBT for initial outcome and to reduce the risk of relapse. In an RCT of 46 adolescents who responded to 12 weeks of fluoxetine treatment, relapse-prevention–focused CBT was associated with lower relapse rates compared to medication alone. In fact, the booster effect of CBT substantially reduced the risk of relapse eightfold. Beyond fluoxetine and escitalopram, several other SSRIs have demonstrated efficacy in youth with MDD. However, interpretation of the clinical trial literature for SSRIs and SNRIs in youth with MDD is complicated by several factors. In recent meta-analyses of these medications in youth with MDD, funding source differentiated medication-placebo differences with smaller medication-placebo differences being observed in industry-funded versus federally funded trials. Thus, the impact of study sponsorship on trial outcome supports the assertion that industry-funded trials with high placebo response rates and small drug-placebo differences are “failed trials” and should not be described as “negative trials” or used to determine public health estimates of antidepressant efficacy in children and adolescents with depression.

Despite the efficacy of the SSRIs, no controlled pediatric trial data suggest benefit in pediatric patients with MDD. Monoamine oxidase inhibitors have very limited data in pediatric use, although one large, double-blind, placebo-controlled trial of transdermal selegeline recently failed to show benefit in youth with MDD, although it was compromised by a very high placebo response rate.

Treatment resistant MDD, in adolescents is prevalent and associated with a substantial psychosocial burden and mortality. In approaching medication-resistant MDD in adolescents, it may be useful to quantify the degree of treatment resistance which may help to identify treatment-responsive phenotypes, and this can be done using the Antidepressant Treatment Record (ATR), which has been evaluated in adolescents. Aside from the Treatment of SSRI-resistant depression in Adolescents (TORDIA) study described below and two recent trials, few have systematically evaluated treatments for this population. In a large study of 10-Hz transcranial magnetic stimulation (TMS) for adolescents with TRD (n = 103), daily treatments over 6 weeks failed to demonstrate superiority to a sham intervention in change in the Hamilton Depression Rating Scale score. Dwyer and colleagues examined a single intravenous infusion of either ketamine (0.5 mg/kg over 40 minutes) or midazolam (0.045 mg/kg over 40 minutes), and the alternative compound 2 weeks later in 17 adolescents and found ketamine to be well tolerated and associated with short-term (2-week) efficacy in reducing depressive symptoms compared with midazolam.

Suicidality

In 2004, an FDA advisory Committee reviewed results of a meta-analysis of 24 controlled clinical trials of 9 antidepressants, which included approximately 4,400 pediatric patients. While there were no completed suicides, the cumulative risk for suicidality (suicidal thinking or behavior) when adverse events were simply collected by spontaneous report was approximately 4% on active medication versus approximately 2% on placebo. This resulted in the FDA issuing a boxed warning that any antidepressant use in children and adolescents and young adults was accompanied by an increased risk for suicidal thinking and behavior.

Subsequent meta-analyses have failed to replicate the original analyses from 2004, even as newer data were incorporated. However, there may be specific increased risks for some SSRIs. Among SSRIs, a meta-analysis found differences in suicidality, with sertraline potentially having lower rates of treatment-emergent suicidality while paroxetine had higher rates of suicidality compared to placebo. That medications within the SSRI class are differentially associated with both higher and lower rates of suicidality compared to placebo with class-wise comparisons (e.g., SSRI vs. placebo) finding no significant differences call into question the boxed warning applied to all antidepressants regardless of indication and specific medication (US Food and Drug Administration. Relationship between psychotropic drugs and pediatric suicidality: review and evaluation of clinical data. Available at: http://www.fda.gov/ohrms/dockets/ac/04/briefing/2004-4065b1-10-TAB08-Hammads-Review.pdf). However, unlike recent network meta-analyses of SSRIs and SNRIs in youth, the analyses that gave rise to the boxed warning on SSRIs and SNRIs as well as earlier analyses of suicidality utilized older meta-analytic techniques could not compare individual medications.

Select Key Studies

Treatment for Adolescents with Depression Study (TADS). In TADS, adolescents with MDD (*n* = 439) were randomized to (1) fluoxetine monotherapy (10 to 40 mg/day), (2) CBT monotherapy, (3) CBT + fluoxetine (10 to 40 mg/day), or (4) placebo. Following 12 weeks of treatment, patients who received fluoxetine + CBT were significantly more improved compared to patients who had received either CBT or fluoxetine monotherapy (Fig. 53.5–4**)**. Additionally, fluoxetine was superior to CBT and all active treatments were superior to placebo.

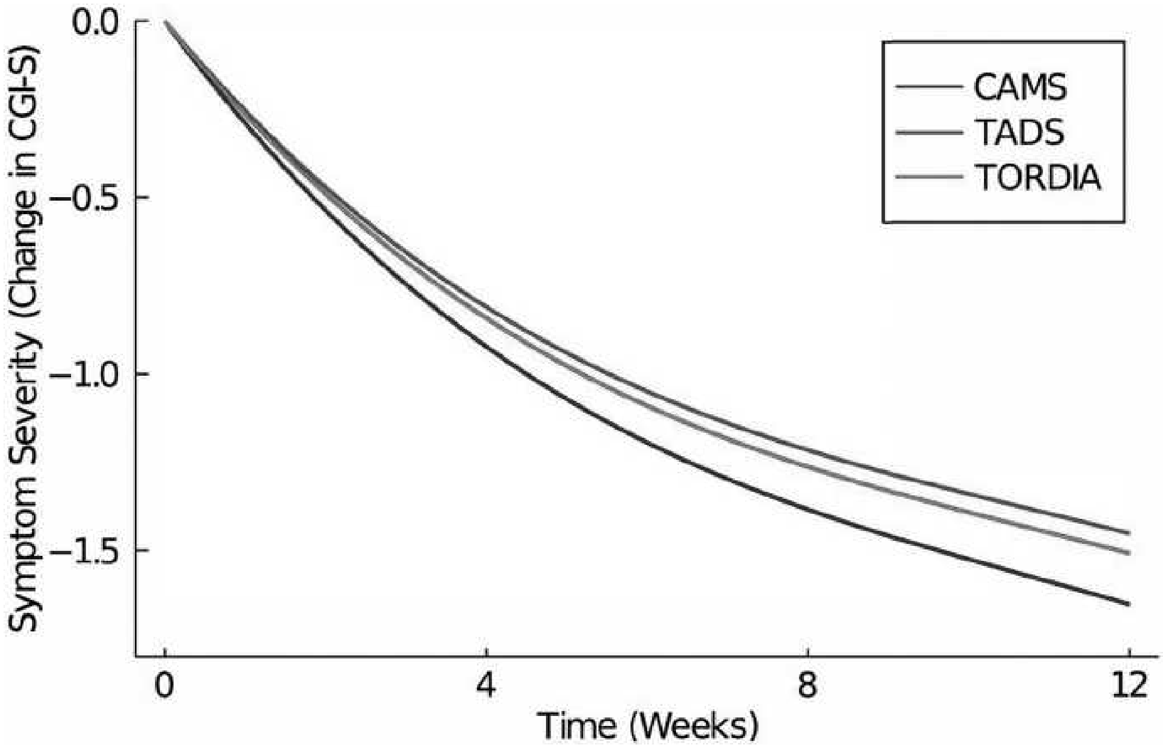

FIGURE 53.5–4. Time course of SSRI response in NIH-sponsored trials of anxiety disorders (Child/Adolescent Anxiety Multimodal Study [CAMS]), depression (Treatment of Adolescent Depression Study [TADS]), and treatment-resistant depression (Treatment of SSRI-Resistant Depression in Adolescents [TORDIA] Study) in children and adolescents. (See eBook for color images.)

Treatment of SSRI-Resistant Depression in Adolescents (TORDIA) Study. In TORDIA, adolescents with MDD aged 12 to 18 (*n* = 334) who did not respond to an adequate dose of an SSRI, for at least 8 weeks, were randomized to one of four treatments over 12 weeks (March et al., 2004). Importantly, these patients had, on average, a more chronic course than those enrolled in TADS. In TORDIA, patients could switch to (1) a different SSRI (paroxetine, citalopram, or fluoxetine); (2) a different SSRI + CBT; (3) venlafaxine ER; or (4) venlafaxine ER + CBT. CBT + medication switch (either to another SSRI or venlafaxine ER) was associated with a marked decrease in depressive symptoms compared to simply changing medication. The two arms with CBT plus medication demonstrated a higher response rate (55%) than a medication switch alone (41%). In the original analysis, no differences in final outcome were detected between switching to venlafaxine ER relative to another SSRI, but patients who were switched to a second SSRI had less suicidal ideation than those treated with venlafaxine and fewer adverse events. Additionally, when tolerability and efficacy of the SSRIs were considered together, adolescents treated with citalopram and fluoxetine fared better than those treated with paroxetine.

Recent reanalysis of the TORDIA data, using contemporary statistical approaches, found that in patients not receiving CBT, SSRIs produced greater and faster improvement in depressive symptoms compared to venlafaxine ER. However, for CBT-treated patients, SSRIs and venlafaxine produced similar improvements in anxiety and depression.

Clinical Implications and the Role of Pharmacotherapy. For children and adolescents who present with a more complex MDD, or have a limited response to supportive psychotherapy, a more specific psychotherapy and/or an SSRI is indicated. Severe MDD and/or chronic recurring MDD will generally require pharmacotherapy, either alone or in combination with psychotherapy. For youths with SSRI-resistant depression, switching to another SSRI represents a reasonable approach and one RCT of adolescents with treatment-resistant depression, suggests that IV ketamine is superior to active placebo in the very short term (i.e., 2 weeks). However, in RCTs of adolescents with treatment-resistant depression, rTMS was not superior to sham.

PHARMACOTHERAPIES FOR BIPOLAR DISORDER

Bipolar disorder in children and adolescents is associated with significant long-term morbidity and disability. Owing to the severity and persistence of bipolar disorder into adulthood, safe and effective long-term treatments are needed.

Mixed Dopamine Serotonin Receptor Antagonists

Placebo-controlled studies suggest that multiple mixed dopamine, serotonin receptor antagonists are efficacy in pediatric patients with acute mania and six agents have received FDA approval for pediatric mania (lithium, risperidone, aripiprazole, quetiapine, olanzapine, and asenapine). In general, antimanic effects are observed within the first 2 weeks of treatment and mood stabilizers may be used as second-line treatments or in combination with mixed dopamine receptor antagonists. While monotherapy is generally preferred, combination pharmacotherapy may be required—such as a mixed dopamine receptor antagonist and mood stabilizer, or an antimanic agent with treatment for comorbid ADHD, anxiety, and/or OCD. Following initial response, most youth require ongoing treatment with medication to prevent relapse. Current recommendations suggest the regimen needed to stabilize the acute mania should be maintained for 12 to 24 months.

Mixed dopamine serotonin receptor antagonists, in youth, have been associated with changes in prolactin, fasting glucose, fasting total cholesterol, necessitating close monitoring (at least bi-yearly to yearly). Some have advocated the use of metformin to decrease weight gain and metabolic risk in youth treated with mixed dopamine serotonin receptor antagonists. Additionally, long-chain omega-3 fatty acids may decrease the risk of hypertriglyceridemia associated with some mixed dopamine receptor antagonists in youth.

Lithium

Based on two NIH-sponsored studies, the Collaborative Lithium Trials (CoLT), lithium is FDA approved for the acute (short-term) treatment of youths 7 years of age and older with manic and mixed episodes associated with bipolar I disorder and for maintenance treatment in pediatric patients ages 7 to 17 years. In the first CoLT study, children and adolescents were randomized to lithium ($n = 53$) or placebo ($n = 28$) for up to 8 weeks. Greater improvements in manic symptoms and response were observed in lithium-treated youth compared to those who receive placebo. Additionally, in the second CoLT study, responders were randomized to continue lithium or to be cross titrated to placebo for up to 28 weeks. The primary outcome measure was relative risk of study discontinuation for any reason. Significantly greater maintenance of response was observed in patients who continued lithium. In youth, the half-life of lithium is significantly shorter than in adults, largely related to increased renal clearance. In youth, pediatric trials have generally used doses of 25 to 30 mg/kg/day and average serum concentrations have been 1.05 mEq/L. Regarding the kinetics of lithium in youth, a population pharmacokinetic study found suggested a maintenance dose of 25 mg/kg/day (divided BID), although this study underscored the importance of ongoing therapeutic drug monitoring. Finally, the tolerability of lithium has been like that observed in adults with gastrointestinal side effects and elevations in TSH being common.

Anticonvulsants

While the anticonvulsants likely have primary roles in the treatment of adults with bipolar disorder, these agents have produced very mixed results in youths. In this regard, improvements in manic symptoms are significantly greater for mixed dopamine serotonin receptor antagonists (approximately 0.65) compared to mood stabilizers (approximately 0.24); however, anticonvulsants are associated with less metabolic side effects in youth compared to mixed dopamine serotonin receptor antagonists. Further, their teratogenic and metabolic effects have limited their usefulness in youth.

Depression Associated with Bipolar Disorder

For youth with depression associated with bipolar I disorder, two medications have received FDA approval, olanzapine-fluoxetine combination (10 to 17 years of age) and lurasidone (10 to 17 years of age). Double-blind placebo-controlled trials of quetiapine and extended-release quetiapine have failed to demonstrate superiority to placebo in youth adolescents with bipolar depression. Open-label studies suggest that lamotrigine and lithium are associated with improvement in depressive symptoms in youth with bipolar I disorder.

Clinical Implications and the Role of Pharmacotherapy. Pharmacotherapy is the primary treatment for mania in youth. However, adjunctive psychosocial treatment may be useful to address disruptive symptoms and their sequela, which often impact family and community. These include interventions that focus on coping and emotional regulation, family therapy (including multi-family psychoeducation and family-focused treatment), CBT and there is emerging evidence that dialectical behavioral therapy may also be helpful. Due to the difficulty of stabilizing active bipolar and frequent comorbid symptoms, mixed dopamine-serotonin receptor antagonists are recommended as first-line treatments, although it is often necessary to employ combined pharmacotherapeutic strategies in these youth. As one example, multiple studies have shown that stimulants can be used in youth with stable bipolar disorder to treat comorbid ADHD.

OBSESSIVE-COMPULSIVE DISORDER (OCD)

OCD affects between 1% and 2% of children and adolescents and is characterized by impairing rituals and obsessions. Pediatric OCD often continues into adulthood. Psychotherapy is recommended as first-line treatment for OCD in youth, but several pharmacotherapies have also demonstrated efficacy.

Pharmacotherapy for OCD

Several SSRIs have demonstrated efficacy and been approved by the FDA for the treatment of OCD in children and adolescents. These medications include fluvoxamine, fluoxetine, clomipramine, and sertraline. Based on the currently available evidence, the American Academy of Child and Adolescent Psychiatry recommends CBT for youth with mild to moderately severe OCD; patients with more severe symptoms of OCD, the combination of an SSRI and CBT may be more effective than CBT alone. Regarding the efficacy of SSRIs for the treatment of OCD in youth, several meta-analyses suggest benefit although the effect sizes in these studies are smaller relative to the effect sizes of these medications for patients with non-OCD pediatric anxiety disorders. The choice of antidepressant may be important; findings in one meta-analysis suggest that in pediatric patients with OCD, the tricyclic antidepressant, clomipramine—which is FDA approved for the treatment of OCD in youth —may be associated with larger effect sizes. Nonetheless, SSRIs are considered first-line psychopharmacologic interventions for youth with OCD and three SSRIs are approved for this indication in youth: fluoxetine, fluvoxamine, and sertraline.

Selected Studies

THE PEDIATRIC OCD TREATMENT STUDY (POTS). POTS randomizing 112 youths with OCD (aged 7 to 17 years) to CBT, sertraline, the combination of the CBT + sertraline and pill placebo. All active treatments were superior to placebo in reducing OCD symptoms, although remission rates for combined treatment was 53.6%, for CBT alone 39.3%, for sertraline alone 21.4%, and for placebo only 3.6%. In POTS (II), 124 youths with OCD (aged 7 to 17 years) were assigned to one of three treatment strategies, all of which included SSRIs. The positive response rates for medication management only (30%) and medication management plus instructions in CBT (34%) were significantly lower than combined medication management plus CBT (69%). These findings indicate that truncated versions of CBT are not efficacious comparatively, and that efforts should focus on making the full CBT protocol available more widely in community settings.

Clinical Implications and the Role of Pharmacotherapy. SSRIs are effective in reducing symptoms of OCD in pediatric patients and are well tolerated. However, in light of the POTS I/II findings, recommendations would be for children and adolescents with OCD to begin treatment with either CBT alone or in combination with an SSRI.

POSTTRAUMATIC STRESS DISORDER

Exposure to traumatic events, including neglect, domestic violence, as well as physical and sexual abuse, is not uncommon for many children and adolescents. A significant number of traumatized children and adolescents develop posttraumatic stress disorder (PTSD), a condition intrinsically linked to the experience of extraordinary stress and characterized by intrusive hyperarousal and avoidant/numbing symptoms. Though the prevalence varies by report, 10% of physically abused adolescents may meet criteria for PTSD. Other samples of youth exposed to urban violence suggest that PTSD rates may be as high as 24% to 34% while following a single motor vehicle accident, the prevalence of PTSD in children at 6 months is approximately 12%. Finally, pediatric patients with PTSD or clinically significant posttraumatic stress symptoms often have comorbid with depression, bipolar disorder, substance use disorders, and externalizing behavioral problems. Additionally, preschool age children with posttraumatic stress symptoms or PTSD may experience dysfunctional attachment, aggressiveness, and impulsivity.

Pharmacotherapy for PTSD

When considered, medications should be utilized adjunctively with evidence-based trauma-focused therapies. Additionally, as sleep difficulties and nightmares are common among youth with PTSD, addressing sleep is an appropriate first step when considering adjunctive pharmacotherapy.

SSRIs. Clinical trials comparing adjunctive antidepressant treatment with placebo in pediatric patients with PTSD are limited and, in general, do not suggest benefit of these medications, although, SSRIs may be considered in youth with co-occurring MDD or anxiety disorders in addition to PTSD.

Antiadrenergic Agents. Medications that centrally modulate noradrenergic tone are supported by clinical research as potential treatments for PTSD. The α_1-antagonist, prazosin has been studies as a treatment for dyssomnias, including nightmares and frequent nighttime awakenings in youth with PTSD. Case reports and a retrospective case series of 34 youth ages 5 to 17 years of age suggest that prazosin, an α_1-adrenergic antagonist, may be effective and well tolerated in treating sleep disturbances in youth with most youth responding to <5 mg qHS. Prazosin should be initiated at 1 mg 30 minutes before bedtime due to concern for first-dose hypotension and may be increased by 1 mg every 3 to 4 days for the first 2 weeks, starting lower and titrating up more slowly in younger children and/or in children and adolescents experiencing nausea, dizziness, or other side effects related to changes in blood pressure. Additionally, an uncontrolled trial suggested that guanfacine may reduce intrusive and hyperarousal symptoms in children with PTSD symptoms.

Mixed Dopamine Serotonin Receptor Antagonists. Limited, uncontrolled trials of second-generation antipsychotic medications in youth with PTSD symptoms have found mixed results and high rates of treatment-emergent adverse events.

Clinical Implications and the Role of Pharmacotherapy. Current evidence for the treatment of PTSD in youth suggests that trauma-focused psychotherapies be considered first-line treatment for children and adolescents with PTSD, with SSRIs and antiadrenergics, considered potential treatment options for comorbidities and for dyssomnias.

SCHIZOPHRENIA

While the onset of schizophrenia is rare prior to 13 years of age, numbers steadily increase throughout the adolescent years, and by 18 years of age

12% to 33% of patients who will ultimately develop schizophrenia will have symptoms.

Pharmacotherapy for Schizophrenia

Aripiprazole, lurasidone, olanzapine, quetiapine, risperidone, and paliperidone have received FDA approval for the treatment of adolescents aged 13 to 17 years. The mixed dopamine serotonin receptor antagonists are considered first-line medications for adolescents with schizophrenia. Studies involving clozapine in treatment-refractory early-onset schizophrenia-spectrum disorders have demonstrated superiority of clozapine compared to haloperidol and olanzapine, with regard to reduction in negative symptoms and to a certain degree with overall response. Finally, based on the results of one RCT, the use of omega-3 fatty acids may delay the onset and/or reduce the ultimate expression of psychotic symptoms in high-risk children.

Clinical Implications and the Role of Pharmacotherapy

Psychopharmacology plays a critical role in the treatment of schizophrenia, with the atypical antipsychotics considered among the first-line treatments. Fish oil may be useful in adolescents at high risk to develop fulminant psychotic symptoms. Psychosocial interventions are considered adjunctive to the medication but are strongly recommended.

EXTERNALIZING DISORDERS

Attention-Deficit/Hyperactivity Disorder (ADHD)

ADHD affects 6% to 9% of children and adolescents worldwide and persists into adulthood in roughly 50% of patients. Stimulant medications represent the first-line pharmacotherapy for ADHD, and their efficacy and tolerability are supported by multiple randomized controlled trials and meta-analyses. In recent decades the utility of stimulant medications has been boosted by the development of long-acting/delayed/extended-release preparations. These newer formulations have allowed patients to sustain treatment effects throughout the day and lower potential abuse liability. Similarly, there have been successes in the development of novel medications and devices for the treatment of ADHD.

Pharmacotherapies for ADHD. Stimulant medication—including both amphetamine- and methylphenidate-based agents are highly efficacious in treating youth with ADHD. In general, 65% to 75% of stimulant-treated youth response and response rates are similar for methylphenidate and amphetamine in immediate-release as well as extended-release preparations.

STIMULANTS. For ADHD core symptoms, amphetamines and methylphenidate are associated with improvement relative to a placebo, and this efficacy has been demonstrated in multiple meta-analyses. While network meta-analyses have failed to detect differences in efficacy between one amphetamine derivative over another or between amphetamines and methylphenidate, their tolerability may differ considerably, and there may also be differences in response at the individual level.

Immediate-release stimulant medications must be taken BID or TID and are both rapidly absorbed and metabolized which results in their therapeutic effects wearing off within several hours. To address these limitations, sustained-release formulations were developed, although in some cases “acute tolerance due to a lack of a steep absorption phase and a flat (i.e., zero-order) drug delivery profile” potentially degraded response. Given this, modified extended-release formulations have been develop which include varying amounts of immediate- and extended-release stimulant and which produce ascending (e.g., first-order) delivery profiles (Table 53.5–1). Longer-acting formulations are considered the first-line stimulant formulation and enhance adherence. However, it is not uncommon to need a “booster” dose of a stimulant in the afternoon with many extended-release proportions to extend the duration of coverage. Dosing is important with stimulant medication and in general, adequate trials of methylphenidate consist of >0.8 mg/kg/day while 0.5 mg/kg is often considered an adequate trial for mixed amphetamine salt-based preparations.

Table 53.5–1.
Common Stimulant Medications and Formulations

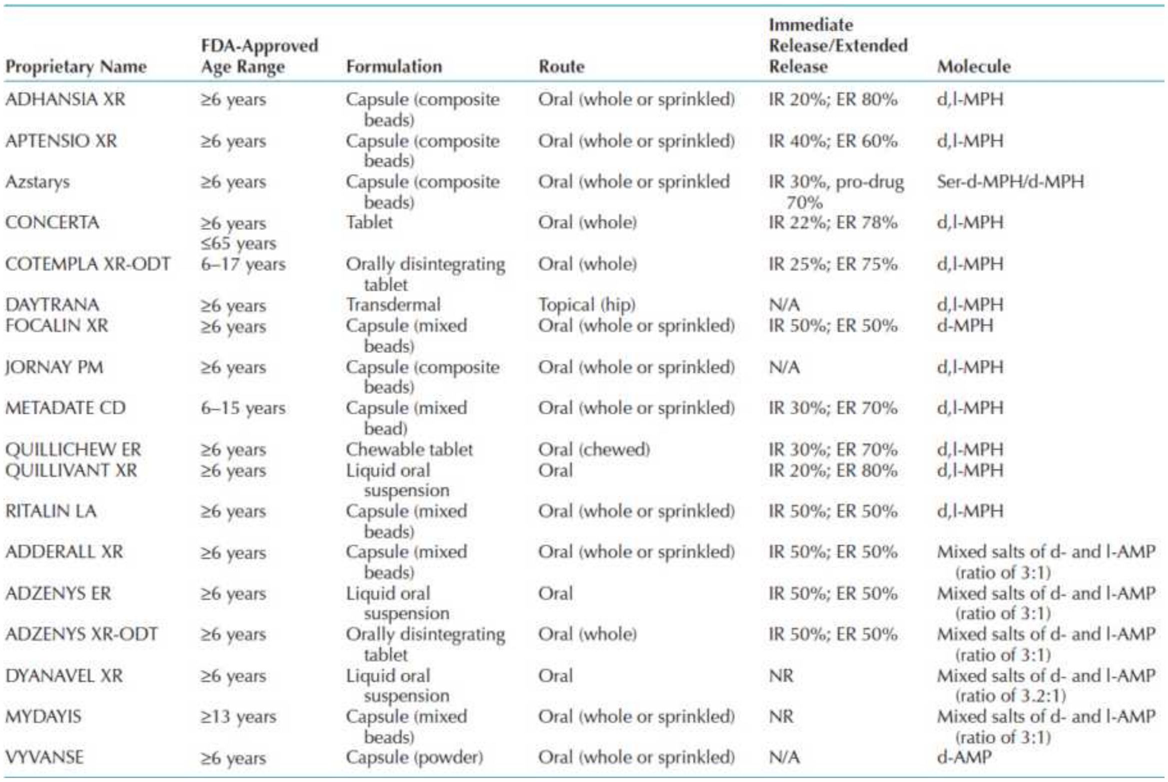

Proprietary Name	FDA-Approved Age Range	Formulation	Route	Immediate Release/Extended Release	Molecule
ADHANSIA XR	≥6 years	Capsule (composite beads)	Oral (whole or sprinkled)	IR 20%; ER 80%	d,l-MPH
APTENSIO XR	≥6 years	Capsule (composite beads)	Oral (whole or sprinkled)	IR 40%; ER 60%	d,l-MPH
Azstarys	≥6 years	Capsule (composite beads)	Oral (whole or sprinkled	IR 30%, pro-drug 70%	Ser-d-MPH/d-MPH
CONCERTA	≥6 years ≤65 years	Tablet	Oral (whole)	IR 22%; ER 78%	d,l-MPH
COTEMPLA XR-ODT	6–17 years	Orally disintegrating tablet	Oral (whole)	IR 25%; ER 75%	d,l-MPH
DAYTRANA	≥6 years	Transdermal	Topical (hip)	N/A	d,l-MPH
FOCALIN XR	≥6 years	Capsule (mixed beads)	Oral (whole or sprinkled)	IR 50%; ER 50%	d-MPH
JORNAY PM	≥6 years	Capsule (composite beads)	Oral (whole or sprinkled)	N/A	d,l-MPH
METADATE CD	6–15 years	Capsule (mixed bead)	Oral (whole or sprinkled)	IR 30%; ER 70%	d,l-MPH
QUILLICHEW ER	≥6 years	Chewable tablet	Oral (chewed)	IR 30%; ER 70%	d,l-MPH
QUILLIVANT XR	≥6 years	Liquid oral suspension	Oral	IR 20%; ER 80%	d,l-MPH
RITALIN LA	≥6 years	Capsule (mixed beads)	Oral (whole or sprinkled)	IR 50%; ER 50%	d,l-MPH
ADDERALL XR	≥6 years	Capsule (mixed beads)	Oral (whole or sprinkled)	IR 50%; ER 50%	Mixed salts of d- and l-AMP (ratio of 3:1)
ADZENYS ER	≥6 years	Liquid oral suspension	Oral	IR 50%; ER 50%	Mixed salts of d- and l-AMP (ratio of 3:1)
ADZENYS XR-ODT	≥6 years	Orally disintegrating tablet	Oral (whole)	IR 50%; ER 50%	Mixed salts of d- and l-AMP (ratio of 3:1)
DYANAVEL XR	≥6 years	Liquid oral suspension	Oral	NR	Mixed salts of d- and l-AMP (ratio of 3.2:1)
MYDAYIS	≥13 years	Capsule (mixed beads)	Oral (whole or sprinkled)	NR	Mixed salts of d- and l-AMP (ratio of 3:1)
VYVANSE	≥6 years	Capsule (powder)	Oral (whole or sprinkled)	N/A	d-AMP

While stimulants are generally well tolerated in youth with ADHD, side effects, including decreased appetite, sleep difficulties, rebound hyperactivity or irritability, headache, and gastrointestinal symptoms, can become problematic. A meta-analysis of appetite loss associated with stimulant medications in pediatric patients with ADHD suggests that nearly half of methylphenidate-treated patients will experience appetite loss (compared to approximately 15% of placebo-treated patients); however, “serious” appetite loss occurs in 10.5% of stimulant-treated youth (compared to approximately 2% of placebo-treated youth). Stimulant-induced appetite decreases appear to be dose related. Several strategies have been proposed for the management of stimulant-induced appetite loss, including careful monitoring of appetite, weight, and body mass index, administration of medication after breakfast, and the use of calorically dense snacks (or fourth meals) later in the day or early evening.

There has been considerable debate about the management of cardiovascular side effects of stimulant medications, and, at times, conflicting recommendations have been provided. Stimulants have been associated with elevations in mean blood pressure (<5 mm Hg) and heart rate (<10 beats/min). A subset of individuals (5% to 10%) may have an even greater increase in heart rate or blood pressure. Although there are limited prospective data related to the relationship between cardiac risk factors and outcomes in stimulant-treated youth, prudent recommendations from the American Heart Association and the American Academy of Pediatrics recommend obtaining a thorough history to identify high-risk youth with potential cardiac abnormalities or pre-existing cardiac disease prior to initiation of stimulant pharmacotherapy. Routine screening electrocardiography is *not* indicated in the absence of other cardiac risk factors.

NOREPINEPHRINE REUPTAKE INHIBITORS (NRIS). The NRI atomoxetine has demonstrated efficacy in multiple controlled trials, although its efficacy is less than that of the stimulants. One meta-analysis reported atomoxetine's effect size as 0.62, while the stimulants were from 0.91 (immediate release) to 0.95 (sustained release). However, atomoxetine exposure varies substantially (30-fold) across individuals, and part of this variability is explained by CYP2D6 activity. Dosing recommendations are available for atomoxetine when the CYP2D6 genotype/phenotype is known.

The second NRI to be approved for children and adolescents with ADHD is viloxazine which also has *in vitro*, antagonistic activity at 5-HT2B and agonistic activity at 5-HT2C receptors (and has high receptor occupancy at clinical doses).

α-AGONISTS. Multiple RCTs support the efficacy of extended-release preparations of clonidine and guanfacine as monotherapy in children and adolescents with ADHD. Typically, extended-release clonidine is administered BID, and the extended-release guanfacine is given either qHS or qAM with equal efficacy throughout the day. Extended-release α-agonists are FDA approved as monotherapy as well as in combination with stimulants. Moreover, in youth who have a partial response to stimulants, α-agonists may boost ADHD response. Immediate-release formulations of clonidine and guanfacine are also used off-label in the treatment of youth with ADHD.

Selected Studies

MULTIMODAL TREATMENT STUDY OF CHILDREN WITH ADHD (MTA). In MTA, children were randomized to one of four groups: medication treatment alone (with a structured algorithm-based approach), psychosocial treatment alone, algorithm-driven medication management, psychosocial treatment, or “community treatment.” The treatment algorithm involved monthly appointments with a physician during which the stimulant was titrated based on parent and teacher ratings. Treatment-related improvements were observed across all groups (at 14 months relative to baseline). However, algorithm-driven medication management was superior to intensive behavioral treatment alone and community treatment. It is noteworthy that the community treatment group had more limited physician follow-up, and, in this group, the average daily stimulant dose was significantly lower than in the algorithmic medication management group. Finally, approximately a quarter of youth who were randomized to behavioral management treatment required pharmacologic treatment secondary to a lack of effectiveness of the behavioral treatment. This finding highlights the role of pharmacotherapy in the management of many pediatric patients with ADHD.

PRESCHOOL ADHD TREATMENT STUDY (PATS). The Preschool ADHD Treatment Study (PATS) was also a large federally funded 14-month study designed to examine medication efficacy for younger children. All children and their parents first participated in a pre-trial, 10-week behavioral therapy, and training course. Only those children who still demonstrated significantly impairing ADHD symptoms were then randomized to receive immediate-release methylphenidate or placebo. Methylphenidate was significantly more effective than placebo at treating preschoolers with ADHD. However, at lower weight-adjusted, smaller effect sizes compared with school-aged children. In this regard, the mean optimal dose of methylphenidate was 14 mg/day (0.7 mg/kg/day), which is lower than the mean optimal dose of 1 mg/kg/day in the MTA study. In terms of tolerability, spontaneously reported moderate to severe adverse effects were present in approximately one in three preschoolers, including emotional outbursts, initial insomnia, repetitive behaviors/thoughts, decreased appetite, and irritability. Moreover, one in ten children discontinued treatment because of intolerable methylphenidate adverse effects (vs. <1% of school-age children in the MTA study).

Clinical Implications and the Role of Pharmacotherapy. Pharmacotherapy is a fundamental component of ADHD treatment in children and adolescents and is superior to behavioral therapy alone. Stimulant medications are among the most effective pharmacotherapies, although they vary considerably in release properties and duration of action. Additionally, in patients who have an incomplete or partial response to stimulants, adjunctive atomoxetine, guanfacine, and clonidine may boost response.

AUTISM SPECTRUM DISORDERS (ASD)

Pharmacotherapy for Autistic Spectrum Disorders (ASD)

ASDs, including autistic disorder, are characterized by impaired socialization, communication, and behavioral problems, frequently presenting with anxiety, irritability, aggression, and inattention/impulsivity-related symptoms. Psychosocial interventions include social skills training and applied behavioral analysis for youth with ASDs, while the role of psychopharmacologic interventions is frequently focused on target symptoms (e.g., irritability, disruptive and aggressive behaviors, impulsivity). Risperidone and aripiprazole have been approved by the FDA for the treatment of irritability associated with autistic disorder in children and adolescents.

Mixed Dopamine Serotonin Receptor Antagonists. Both risperidone and aripiprazole are FDA approved for the treatment of irritability associated with an autism spectrum disorder. In addition, quetiapine, olanzapine, ziprasidone, and paliperidone are commonly used clinically to manage irritability and aggressive behavior in youths with ASD; however. Meta-analyses of pediatric patients with autism (higher functioning) demonstrated improvement in behavioral symptoms associated with mixed dopamine serotonin receptor antagonists but confirmed the significant metabolic liability (e.g., weight gain) associated with this class of medications.

Additionally, first-generation antipsychotic medications (i.e., dopamine receptor antagonists) have been used to manage irritability and aggression in children with ASD. However, comparative data suggest that the mixed dopamine serotonin receptor antagonists are potentially more effective and may have fewer extrapyramidal side effects. More recently, there have been several observational studies examining long-acting injectable forms of the mixed dopamine SSRIs, and these may be particularly helpful when children struggle with swallowing tablets and adherence.

Clinical Implications and the Role of Pharmacotherapy

SSRIs are modestly effective in improving global functioning and specifically improve repetitive behaviors, anxiety, irritability, and hyperactivity in children with ASD. Mixed dopamine serotonin receptor antagonists are commonly used for affective dysregulation, including irritability, aggression, and self-injurious behaviors. The mixed dopamine serotonin receptor antagonists may improve both the affective symptoms as well as—in some samples—ritualistic behaviors. For ADHD, as well as symptoms of inattention and distractability, stimulants are commonly used; however, the degree of stimulant-related improvement for these target symptoms is less than that which is associated with idiopathic ADHD.

SLEEP-RELATED DISORDERS

Insomnia is common in the general pediatric population and is certainly more common in youth with psychiatric disorders due to the disorders themselves as well as a sequela of the medications prescribed to treat the primary disorder. Pharmacologic treatment may be indicated if children do not show significant improvement following behavioral intervention. However, excluding other primary causes of dyssomnia (e.g., obstructive sleep apnea, iron deficiency anemia) is critical prior to pursuing pharmacotherapy, as pharmacotherapy could mask an underlying disorder.

Insomnia

Insomnia is common in the general pediatric population and is certainly more common in youth with psychiatric disorders due to the disorders themselves as well as a sequela of the medications prescribed to treat the primary disorder. Pharmacologic treatment may be indicated if children do not show significant improvement following behavioral intervention. However, excluding other primary causes of dyssomnia (e.g., obstructive sleep apnea, iron deficiency anemia) is critical prior to pursuing

pharmacotherapy, as pharmacotherapy could mask an underlying disorder.

Pharmacotherapy for Insomnia. There are no FDA-approved sleep agents for children. After aggressive sleep hygiene and education, controlled trials have demonstrated improved sleep latency outcomes with melatonin. Data regarding other agents used chronically are largely derived from expert opinion or case series, including benzodiazepines, clonidine, mirtazapine, imipramine, trazodone, and, less frequently, sedating atypical antipsychotics.

The primary regulator of the sleep–wake cycle is melatonin, an endogenous hormone produced by the pineal gland in response to changes in retinal light perception. Exogenous melatonin supplementation may be the preferred initial pharmacotherapy for sleep-onset insomnia due to its chronobiotic properties. In clinical studies of children and adolescents, both immediate- and extended-release melatonin reduced sleep onset latency and increased total sleep duration in pediatric patients, although the increase in total duration of sleep is greater with extended-release preparations. Additionally, melatonin may have a particularly important role in circadian rhythm sleep disorders. In this regard, low-dose melatonin (0.5 mg), when timed relative to the endogenous dim light melatonin onset, is more effective in shifting the sleep phase than higher doses, suggesting that timing may have a greater impact than dosage. Data regarding melatonin administration suggest that the optimal timing for melatonin administration is 4 to 6 hours prior to a child's preferred bedtime, and doses of 0.5 to 1 mg have been effective when given in this window. Variation across studies has contributed to a lack of consensus regarding pediatric melatonin dosing. For example, 0.05 mg/kg may be a minimal effective dose when given 1 to 2 hours prior to bedtime. Finally, studies of selective GABAergic modulators and benzodiazepines have not produced positive results in prospective trials of youth with insomnia.

Clinical Implications and the Role of Pharmacotherapy. Pharmacotherapy should be considered only after a thorough assessment of the etiology of the sleep disruption (e.g., sleep hygiene, potential iatrogenic causes, sleep apnea, insomnia as a sequela of a psychiatric disorder). Evidence-based pharmacotherapy for insomnia is largely limited to melatonin-related agents in the pediatric age range. For pharmacologically induced sleep disruption, consider alternative preparations, classes of medication, or medication discontinuation.

Enuresis

While a specific etiology for enuresis is not commonly identified, a careful assessment is necessary for consideration of potential urologic, developmental, psychosocial, and/or sleep-related causes. If the history, physical examination, and urinalysis are unremarkable and do not suggest a specific etiology, supportive approaches combined with conditioning are considered the first-line treatment. Supportive interventions include education as well as recommendations against punishing the child for enuretic episodes. Behavioral conditioning methods with portable battery-operated alarms have been repeatedly demonstrated to be highly effective.

Pharmacotherapy for Enuresis. In addition to imipramine, two medications have demonstrated efficacy in the treatment of enuresis, desmopressin (DDAVP), and oxybutynin. As a synthetic analog of vasopressin, desmopressin decreases urine production. It is generally given orally or intranasally, dosed initially at 0.2 mg at bedtime, and titrated up to 0.4 to 0.6 mg as needed. Success rates of 10% to 65% have been reported. If successful, a course of treatment is recommended for 3 to 6 months. Additionally, oxybutynin, in combination with imipramine, has demonstrated superiority relative to placebo and imipramine.

Imipramine is FDA approved for enuresis and has been used for decades despite an unknown mechanism of action, with several studies demonstrating 40% to 60% effectiveness. It is given in a single bedtime dose of 1 to 2.5 mg/kg/day. Additionally, RCTs suggest that amitriptyline and desipramine have efficacy in decreasing pediatric enuresis, although they lack FDA approval for this indication. When effective, TCAs should be continued for 4 to 6 months. However, given the risk of arrhythmias with tricyclic antidepressants, an ECG can be considered prior to treatment and periodically during treatment.

Clinical Implications and the Role of Pharmacotherapy. If supportive approaches and behavioral conditioning with a urine-responsive alarm (i.e., “bell and pad”) are ineffective, medication may be considered as a treatment option. Additionally, medication may also be useful intermittently to allow youth to attend overnight activities such as camp or sleepovers without embarrassment.

Tic DISORDERS

Tic disorders are common in youth, occurring in 15% of the general population, and are more common in youth with other psychiatric disorders. In general, when tic disorders do not affect function, pharmacologic intervention is not recommended. However, when tics interfere with function, as is frequently the case with Tourette syndrome, pharmacotherapy may be indicated. Currently, there are three medications with FDA approval for tic suppression in Tourette's, and these include haloperidol (age >12 years), pimozide (>12 years of age), and aripiprazole (>6 years of age). However, secondary to the risks associated with dopamine antagonists and mixed dopamine serotonin receptor antagonists, α_2-agonists (e.g., clonidine and guanfacine) are generally used as the first-line pharmacologic interventions. Additional trials are underway with VMAT2 inhibitors (e.g., tetrabenazine), which deplete vesicular dopamine and have been used in other hyperkinetic disorders.

Psychotropic medications play a critical role in treating many psychiatric disorders in children and adolescents. However, using these medications effectively and appropriately requires an understanding of pediatric pharmacokinetics and pharmacodynamics, in addition to the nuances of response that are revealed by numerous double-blind, placebo-controlled trials. In this chapter, we review the developmental pharmacology that is relevant to psychotropic medication use in pediatric populations and the evidence for multiple psychotropics across internalizing and externalizing disorders.

CONCLUSIONS

There is substantial evidence for pharmacologic treatments for psychiatric disorders in children and adolescents. Pharmacokinetics and pharmacodynamics in children and adolescents differ considerably from adults, and these differences give rise to differences in dosing, differences in response, and differences in tolerability. As such, the safe and effective use of pharmacotherapy requires an understanding of pediatric pharmacokinetics and pharmacodynamics in addition to the nuances of response that are revealed by numerous double-blind, placebo-controlled trials.

FURTHER READINGS

Anagnostou E, Aman MG, Handen BL, et al. Metformin for treatment of overweight induced by atypical antipsychotic medication in young people with autism spectrum disorder: a randomized clinical trial. *JAMA Psychiatry*. 2016;73:928–937.

Arns M, Kooij JJS, Coogan AN. Review: identification and management of circadian rhythm sleep disorders as a transdiagnostic feature in child and adolescent psychiatry. *J Am Acad Child Adolesc Psychiatry*. 2021;60(9):1085–1095.

Batchelor HK, Marriott JF. Paediatric pharmacokinetics: key considerations. *Br J Clin Pharmacol*. 2015;79(3):395–404.

Caldwell PHY, Sureshkumar P, Wong WCF. Tricyclic and related drugs for nocturnal enuresis in children. *Cochrane Database Syst Rev*. 2016;2016(1):CD002117.

Correll CU, Sheridan EM, DelBello MP. Antipsychotic and mood stabilizer efficacy and tolerability in pediatric and adult patients with bipolar I mania: a comparative analysis of acute, randomized, placebo-controlled trials. *Bipolar Disord*. 2010;12(2):116–141.

Cortese S, Adamo N, Del Giovane C, et al. Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: a systematic review and network meta-analysis. *Lancet Psychiatry*. 2018;5(9):727–738.

Cortese S, Holtmann M, Banaschewski T, et al. Practitioner review: current best practice in the management of adverse events during treatment with ADHD medications in children and adolescents. *J Child Psychol Psychiatry*. 2013;54(3):227–246.

Findling RL, McNamara NK, Pavuluri M, et al. Lithium for the maintenance treatment of bipolar I disorder: a double-blind, placebo-controlled discontinuation study. *J Am Acad Child Adolesc Psychiatry*. 2019;58(2):287–296.e4.

Findling RL, Robb A, McNamara NK, et al. Lithium in the acute treatment of bipolar i disorder: a double-blind, placebo-controlled study. *Pediatrics*. 2015;136:885–894.

Fristad MA, MacPherson HA. Evidence-based psychosocial treatments for child and adolescent bipolar spectrum disorders. *J Clin Child Adolesc Psychol*. 2014;43:339–355.

Hammerness PG, Perrin JM, Shelley-Abrahamson R, Wilens TE. Cardiovascular risk of stimulant treatment in pediatric attention-deficit/hyperactivity disorder: update and clinical recommendations. *J Am Acad Child Adolesc Psychiatry*. 2011;50:978–990.

Kollins S, Greenhill L, Swanson J, et al. Rationale, design, and methods of the Preschool ADHD Treatment Study (PATS). *J Am Acad Child Adolesc Psychiatry*. 2006;45:1275–1283.

March J, Silva S, Petrycki S, et al; Treatment for Adolescents With Depression Study (TADS) Team. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *JAMA*. 2004;292:807–820.

Mills JA, Strawn JR. Antidepressant tolerability in pediatric anxiety and obsessive-compulsive disorders: a bayesian hierarchical modeling meta-analysis. *J Am Acad*

Child Adolesc Psychiatry. 2020;59(11):1240–1251.

Mossman SA, Mills JA, Walkup JT, Strawn JR. The impact of failed antidepressant trials on outcomes in children and adolescents with anxiety and depression: a systematic review and meta-analysis. *J Child Adolesc Psychopharmacol*. 2021;31(4):259–267.

Ramsey LB, Namerow LB, Bishop JR, et al. Thoughtful clinical use of pharmacogenetics in child and adolescent psychopharmacology. *J Am Acad Child Adolesc Psychiatry*. 2021;31(4):259–267.

Sakolsky DJ, Perel JM, Emslie GJ, et al. Antidepressant exposure as a predictor of clinical outcomes in the Treatment of Resistant Depression in Adolescents (TORDIA) study. *J Clin Psychopharmacol*. 2011;31(1):92–97.

Strawn JR, Lu L, Peris TS, Levine A, Walkup JT. Research review: pediatric anxiety disorders – what have we learnt in the last 10 years? *J Child Psychol Psychiatry*. 2021;62(2):114–139.

Strawn JR, Mills JA, Croarkin PE. Switching selective serotonin reuptake inhibitors in adolescents with selective serotonin reuptake inhibitor-resistant major depressive disorder: balancing tolerability and efficacy. *J Child Adolesc Psychopharmacol*. 2019;29(4):250–255.

Strawn JR, Mills JA, Sauley BA, Welge JA. The impact of antidepressant dose and class on treatment response in pediatric anxiety disorders: a meta-analysis. *J Am Acad Child Adolesc Psychiatry*. 2018;57(4):235–244.e2.

The Research Unit on Pediatric Psychopharmacology Anxiety Study Group. Fluvoxamine for the treatment of anxiety disorders in children and adolescents. *N Engl J Med*. 2001;344:1279–1285.

Varigonda AL, Jakubovski E, Bloch MH. Systematic review and meta-analysis: Early treatment responses of selective serotonin reuptake inhibitors and clomipramine in pediatric obsessive-compulsive disorder. *J Am Acad Child Adolesc Psychiatry*. 2016;55:851–859.e2.

Walkup JT, Albano AM, Piacentini J, et al. Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *N Engl J Med*. 2008;359:2753–2766.

Woolf AD, Erdman AR, Nelson LS, et al. Tricyclic antidepressant poisoning: an evidence-based consensus guideline for out-of-hospital management. *Clin Toxicol (Phila)*. 2007;45:203–233.

53.6 Inpatient Psychiatric, Partial Hospitalization, and Residential Treatment for Children and Adolescents

CYNTHIA WILSON, M.D., AND ANDRÉS MARTIN, M.D., PH.D.

The demand for psychiatric inpatient beds rose significantly after the influenza epidemic of the 1920s. This was due to the need for long-term custody of children with mental and behavioral problems post encephalitis. Against this background, it is interesting to consider what may be the long-term effects of the COVID-19 pandemic on mental health care, particularly for children and adolescents. The number of inpatient beds peaked in the 1950s, prior to the deinstitutionalization movement of the 1960s. Deinstitutionalization was propagated by poor conditions in some institutions, as well as changes in treatment approaches and the advent of psychopharmacologic interventions.

The number of inpatient and residential pediatric beds has continued to decrease since then, especially starting in the 1980s. A nationally representative study of community hospitals conducted by Case et al. in 2007 showed that the inpatient lengths of stay continued to decrease through the end of the 20th century, going from a median of 12 days in 1992 to 5 days in 2000, a reduction of 63%. Meagher et al. looked at one inpatient setting in Massachusetts, which also demonstrated a significant decline in the length of stay, ranging from a mean of 45 days in 1991 and 11 days in 2008. Though the mean length of stay is quite different in these studies, they both show a steady decline in the amount of time children spent in the hospital. The latter study also showed a significantly increased number of psychotropic medications at admission and discharge, lower global assessment of functioning at intake, and an increased number of prior hospitalizations between patients admitted in 2008 versus 1991. There has been a concurrent increase in the number of psychiatric inpatient admissions for children and adolescents, as noted in an analysis of National Hospital Discharge Survey data between 1996 and 2007 by Blader et al. Possible explanations may be that outpatient, community-based, and systems-of-care approaches that work for many children and adolescents may not be available in the amount necessary to meet current demands. Alternatively, children and adolescents requiring inpatient hospitalization may be sicker today than those admitted in the 1990s.

The effects of institutions closing their doors, a shift to managed care and other third-party funders, the privatization of Medicaid, and changing paradigms of effective psychiatric treatment options combined to result in a shift in the way in which inpatient psychiatric hospitalization is conceptualized and used, particularly in the United States. Emphasis has shifted to acute stabilization, diagnostic assessment, implementation of multidisciplinary treatment interventions, and return to the community and its available supports. The focus of inpatient psychiatric hospitalization has shifted more toward assessment and stabilization and less long-term treatment, although there is recognition that for a minority of children, inpatient and highly structured care settings remain the preferred treatment approach.

REASONS FOR HOSPITALIZATION AND TYPICAL REFERRAL SOURCES

Although there are significantly fewer inpatient beds since the onset of deinstitutionalization in the 1960s, inpatient psychiatric treatment continues to be necessary and essential at times. The most common reasons for psychiatric admission in children and adolescents include attempted suicide, suicidal ideation, homicidal ideation, aggression, psychosis, and an inability to care for activities of daily living due to psychiatric pathology. Inpatient hospitalization is mostly deemed necessary when these symptoms cannot be safely managed in the community. It is often the last resort and is only used if there are concerns about safety or grave disability.

Referrals to child and adolescent inpatient beds can come from a variety of sources, but overwhelmingly from emergency departments. Often inpatient units are affiliated with a hospital or academic institution that has an attached psychiatric emergency room. However, they could also be referred from other hospitals, outpatient psychiatrists, community mental health clinics, or juvenile detention centers. There are additionally free-standing psychiatric hospitals that are not affiliated with a hospital system. A larger hospital system may be better suited to address the needs of medically challenging or complicated children or to pursue complex medical workups.

To complicate matters, in the United States, as in many countries, there is a shortage of qualified child and adolescent psychiatrists. There is additionally a national shortage of sufficient outpatient services to keep children out of the hospital. As a temporary measure, some hospitals have adapted to “board” children on general pediatric wards or in the emergency department while they await a psychiatric bed or until they are stabilized. Children who do not meet strict inpatient admission criteria may be referred to crisis stabilization programs that provide short-term treatment while longer-term outpatient services are arranged.

LEVELS OF CARE

The level of care appropriate for a particular patient is dependent on the current symptom profile, acuity, and availability of outpatient treatment options. Inpatient hospitalization is considered the highest level of care and provides 24/7 observation and medical care. The majority of inpatient units are locked facilities for safety purposes and therefore considered the most restrictive level of psychiatric care. Open or unlocked units typically exclude patients at acute risk of harming themselves. Partial hospital programs (PHPs) and intensive outpatient programs (IOPs) are used when the child is safe enough to remain at home yet has sufficient symptoms to require structured therapeutic day supports. Residential programs provide care for children and adolescents around the clock, typically over a longer period of time (Table 53.6–1). In addition, there are a variety of outpatient services available for children and adolescents, as discussed in other chapters.

Table 53.6–1.
Differences among Levels of Care

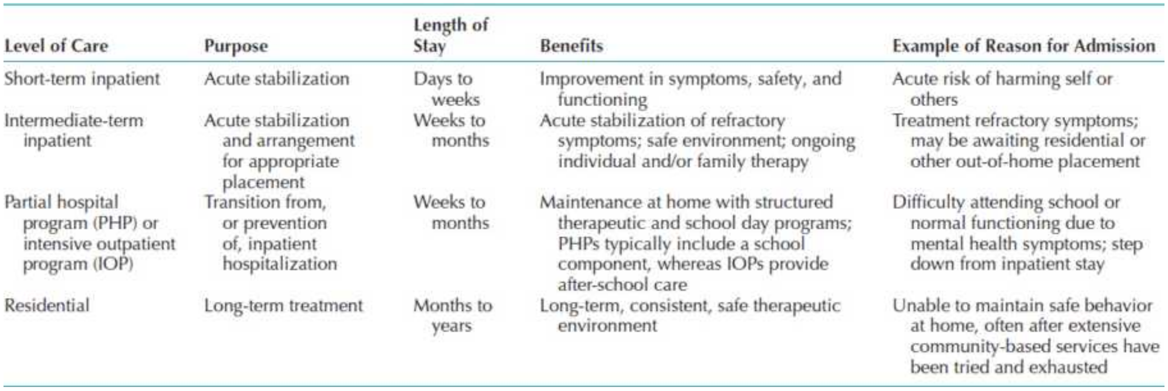

Level of Care	Purpose	Length of Stay	Benefits	Example of Reason for Admission
Short-term inpatient	Acute stabilization	Days to weeks	Improvement in symptoms, safety, and functioning	Acute risk of harming self or others
Intermediate-term inpatient	Acute stabilization and arrangement for appropriate placement	Weeks to months	Acute stabilization of refractory symptoms; safe environment; ongoing individual and/or family therapy	Treatment refractory symptoms; may be awaiting residential or other out-of-home placement
Partial hospital program (PHP) or intensive outpatient program (IOP)	Transition from, or prevention of, inpatient hospitalization	Weeks to months	Maintenance at home with structured therapeutic and school day programs; PHPs typically include a school component, whereas IOPs provide after-school care	Difficulty attending school or normal functioning due to mental health symptoms; step down from inpatient stay
Residential	Long-term treatment	Months to years	Long-term, consistent, safe therapeutic environment	Unable to maintain safe behavior at home, often after extensive community-based services have been tried and exhausted

Inpatient Psychiatric Care

As discussed, the inpatient level of care for children and adolescents is often a last resort and is only used if there are concerns about safety or grave disability. Insurance companies will typically deny payment for inpatient admission if certain stringent criteria are not met. Typical admission criteria include (1) a psychiatric condition that places the child or adolescent at imminent risk of harm to self or others; (2) a psychiatric illness that poses significant medical complications such as anorexia or catatonia; or (3) loss of judgment or impulse control that results in dangerous, threatening, or disorganized behavior leading to grave disability or impairment.

Inpatient psychiatric care for children and adolescents is used for stabilization, most often with the goal of returning to the community with enhanced supports. The first step in this process is to conduct a thorough psychiatric assessment to inform the inpatient treatment plan as well as to make recommendations for treatment following the hospitalization, commonly referred to as aftercare planning. In order to best do so, it is necessary to consider multiple domains of the child's life, including the individual patient, the family, and the system of care. Given the developmental complexities of childhood and adolescence, all of this has to be done with special consideration of a developmental framework. This assesses the successful completion of prevailing or stalled developmental tasks with the eventual goal of a healthy transition to the next childhood developmental phase and, eventually, adulthood. These tasks are different for children, adolescents, and young adults and can have some individual variability.

In considering the individual issues related to the child or adolescent, it is most helpful to consider a biopsychosocial formulation. Approximately 75% of serious psychiatric disorders have an onset prior to the age of 25, and the clinician must be astute to how these disorders can present in childhood and adolescence as opposed to the typical adult presentation. On an individual biologic level, it is necessary to consider hereditary factors, substance use, intrauterine exposure to toxins, any medical conditions that may be contributing, substance use, and the role of puberty. Medication management is often a key component of inpatient hospitalization, although it should not be the sole component. Children and adolescents, at times, are suffering from a psychiatric illness to the extent that medications and interventions typically reserved for adults may be necessary, such as the use of lithium, clozapine, or interventional psychiatry procedures. Electroconvulsive therapy (ECT) and ketamine treatments can be considered for refractory depression in children and adolescents. ECT is sometimes also used for the treatment of refractory psychosis or catatonia. It is necessary on the individual level to have a thorough assessment and formulation to determine the best overall treatment plan. Medication may not always be indicated, and the treatment plan will also likely include various psychological therapeutic techniques. It is helpful to consider from a psychological perspective the various cognitive processes and coping skills of the child and adolescent to drive the psychological treatment both inside and outside of the hospital. For example, a child who is very anxious and displays frequent catastrophic thinking may benefit from the use of cognitive behavioral therapy (CBT). Likewise, a child who is self-harming and engaging in impulsive behaviors may benefit more from dialectical behavioral therapy (DBT).

It is crucial to consider the family system of the child and the adolescent. It is important to keep in mind that the family system may not be two biologic parents and can be complex. Examples of possible family constellations could include one biologic parent, stepparents, extended biologic family, adoptive parents, or child protective services. It is important to discuss with the child and adolescent whom they consider to be their family and support system. Impaired family functioning may be contributing to the psychiatric illness of the child, and likewise, the mental illness of the child may be impacting family functioning. Frequently, family conflict precipitates the crisis leading to psychiatric hospitalization for children and adolescents. Family meetings focused on targeted interventions can help to defuse the acute crisis as well as foster better family functioning when the child or adolescent returns home. It can be a safe space for children and adolescents, as well as parents, to have difficult conversations and then independently process without immediately having to return home together. It is additionally common for children and adolescents to reveal previously undisclosed information during a psychiatric hospitalization, such as a sexual assault or their use of substances. Parents can be both devastated and yet appreciative of learning this information. Such disclosures to parents can lead to a better understanding of their child's mental health struggles. Parents should be strongly encouraged to discuss with their child or adolescent any new rules or guidelines that will be implemented at home, such as restrictions around the use of social media or increased levels of supervision to avoid an acute emotional crisis once the child is discharged and returns home. Such an emotional crisis could lead to readmission to an inpatient unit. Family strengths should also be highlighted and praised. Additional concepts to consider within the family domain are expressed emotion, which is known to exacerbate symptoms of mental illness, and intergenerational trauma. Often, families could benefit from ongoing family therapy outside of the hospital, and this should be included as a recommendation if indicated. This may be in the form of family therapy with an individual therapist or an in-home therapy team, such as intensive in-home child and adolescent psychiatric services (IICAPS). There are additionally in-home models of care focused on problematic substance use and behavioral concerns: multisystemic therapy (MST) and multidimensional family therapy (MDFT).

There is truth in the African proverb and later book title by Hillary Clinton “it takes a village to raise a child.” Children and adolescents are raised in and by the community and the various entities involved in their lives. The needs of emotionally disturbed children cannot be addressed by the mental health system alone. In addition to the mental health care system, it is also important to consider the various other elements of the system of care influencing the social, emotional, and physical health of children and adolescents. This often includes many community agencies such as faith-based organizations, recreational activities, job and skills training, and mentoring. The school system plays an integral role in the life of both children and adolescents. Section 504 of the Rehabilitation Act of 1973 ensures the federal civil rights of those with disabilities, including children and adolescents with mental health diagnoses that may be impacting their academic performance. Under this act, these youth are entitled to a school plan or an individualized educational plan (IEP) until the age of 18. Parents are often unaware of this legal right and, subsequently, need assistance in requesting and giving recommendations for appropriate accommodations. Schools often welcome assessments and recommendations for appropriate accommodations provided by inpatient treating psychiatrists. For many children and adolescents, the system of care can also include government agencies such as child protective services and the juvenile justice system. It is important to consider in a formulation how these interlocking systems may impact the youth’s mental health. It is also important to make recommendations to bolster the mental health of children through increased in-school support, mentors, prosocial activities, improved friendships, and, when necessary, involvement in child protective services.

Milieu Management. The schedule in a child and adolescent inpatient unit is often similar to a typical school day schedule. The therapeutic day includes schoolwork, recreation groups, group therapy, supervised free time, and time for individual and family meetings. One also has to consider the developmental levels of the majority of the patients and how to make the milieu programming developmentally appropriate as well as useful. Among the unique features of inpatient, partial hospital, and residential care is the role of milieu-based, group-wide activities. As part of these activities, children benefit and learn from (and are routinely challenged by) the presence and demands of their peers and staff members. Their interactions and reactions in groups can provide helpful information for a biopsychosocial assessment as well as for appropriate aftercare planning.

The physical layout of a psychiatric unit is generally different from that of a medical inpatient unit. Most often, there is a multipurpose common room, similar to a living room, which can be used for a variety of therapeutic and recreational activities. Some units have a secure outside playground, providing another important outlet for physical expression and exercise. Bedrooms typically have beds for one to three children, and time spent in the community room is encouraged. Inpatient units are designed with attention to safety considerations to minimize the possible risk of harm to oneself or anyone else.

The management of a number of children or adolescents in one enclosed space can be challenging at times. Given their mental health concerns, special attention is often needed to minimize aggression as well as self-harming or suicidal behaviors. Aggression is one of the most common reasons for inpatient hospitalization in children. As such, inpatient units are charged with safe ways of de-escalation and stabilization of aggression and violence. Seclusion into contained spaces, physical holds, and mechanical restraints have been some historical means of dealing with acute violence and the imminent risk of injury to self and others. Following reports of deaths due to asphyxiation during restraints and physical holds in the early 1990s, there have been concerted and federally supported efforts to reduce the use of such restrictive interventions. As a result of these efforts, the use of restraints has decreased dramatically in many settings, although they continue to occur regularly in others.

Every effort should be made for de-escalation and proactive treatment planning. Physical interventions should only be used when all others have failed, and there is an imminent risk of harm to a patient or staff. Following any episode of seclusion or restraint, there should be a debriefing with all staff involved to determine any possible warning signs as well as contributing factors and how this could be prevented in the future. Questions about race and culture and whether or not this contributed to the restraint process should be routinely considered: several entities have reported racial disparities in the utilization of seclusion and restraints. It is clear that this needs to be taken into serious consideration as well as further studied to develop a better understanding of systemic racism in child and adolescent units and the unique challenges of and ways to engage minority youth. There is evidence that behavior modification in an inpatient unit can make a large difference in terms of decreasing inpatient aggression. There is additionally evidence for the use of collaborative problem solving as a milieu management strategy and in the reduction of the need for restraints and seclusion.

A number of additional challenges in the management of the milieu can arise. One such challenging behavior is attempted romantic interludes between peers. Although romantic interest is often developmentally appropriate, it is not conducive to the clinical work of an inpatient setting, and physical contact between peers can be problematic. Another potential behavior is the use of hate speech, either between peers or toward staff. This is best addressed preemptively and with clearly communicated reminders that this will not be tolerated. If such words are used, it should be addressed by staff immediately, and follow-up is needed with the patients involved, as well as with the treatment team, staff members, and particularly with the community of patients, who need to be ensured of their safety and of the antiracist principles governing the inpatient setting.

A further consideration in child and adolescent units is policies around transgender and gender-fluid youth. Children and adolescents should all be asked their preferred names and pronouns, and these preferences are adhered to by staff and in introductions to peers. This can, at times, cause conflict with parents who may not be accepting of the nonbinary or fluid status of their child or adolescent. In addition, questions can arise about rooming assignments, and a clear policy should be put into place that protects confidentiality and is gender affirming.

Multidisciplinary Approach. Inpatient child and adolescent psychiatric units are multidisciplinary by nature. They often include a variety of disciplines, including nurses, milieu counselors, recreational therapists, occupational therapists, social workers, chaplains, dieticians, teachers, pharmacists, psychologists, psychiatrists, and other licensed independent practitioners. In academic settings, there may additionally be residents, fellows, medical students, and other learners. Each staff person's assessment and observations of the child are important and valued. Inpatient mental health services for children require especially high staff-to-patient ratios and 24-hour per-day observation of the child's behavior. Nurses and milieu counselors are specially positioned to observe unit behaviors and peer interactions. Multidisciplinary team rounds typically occur daily, during which representatives from each discipline come together to discuss each patient's progress and treatment planning. Handover of information between shifts provides staff with the framework to improve communication across different staff members and disciplines and is critical to the smooth functioning of an inpatient unit.

Legal Considerations. There are many legal considerations in the inpatient mental health treatment of children and adolescents. Even within the United States, there is significant variation by state regarding at which ages parents may legally hospitalize their children against their will, children are able to voluntarily admit or discharge themselves, children may provide their own consent for outpatient mental health treatment, consent for psychotropic medications, and children and adolescents may seek independent treatment for sexual health. Therefore, it is crucial to know the laws specific to each country, state, or territory. There may be times when a child or adolescent may be hospitalized either against their or their parent's wishes due to safety concerns. These processes also differ by state but typically involve evaluation by one or more physicians and an avenue for a hearing with a judge. There are additional considerations of privacy and how much information can and should be legally shared with parents. Confidentiality and its limits should be openly discussed with children and adolescents at the beginning of their admission. Children and adolescents may additionally disclose that they have been the victims of abusive or assaultive behavior in the community, which would need to be reported to child protective services. Laws regarding what is reportable and what is considered abuse also vary by locality and are essential knowledge for all mental health providers specific to where they practice. It is not uncommon for parents, especially those who are divorced or separated, to disagree on the treatment plan for the child or adolescent, and it is then necessary to know who has legal decision-making abilities. It is often helpful to ask for a divorce decree or custody documents when having to ascertain these distinctions.

Treatment and Discharge Planning. Treating the child in the least restrictive environment should be a guiding principle of any level of care. Short inpatient stays provide a safe, containing environment for a child who is acutely at risk of harm to self or others. At the time of admission, outpatient support services necessary for discharge should start to be explored. Inpatient care should not be viewed as an endpoint but rather as a site for stabilization and a launching point for longer-term community treatment options or residential placement. Because one of the rules of inpatient hospitalization is to protect children from harming themselves or others due to dangerousness or aggression, it follows that at discharge, a child no longer poses such an acute risk or is no longer gravely disabled due to mental illness. There may be times at which suicidal ideation, self-harm, or aggression are chronic, and therefore the goal may be to minimize rather than altogether eliminate this risk and to enhance safety planning.

The inpatient admission assessment often includes the input of multiple disciplines, as well as parents and other collateral sources. Information regarding the child’s past treatments may be evaluated by obtaining prior records or speaking with the child’s outpatient providers and parents. Pediatricians can be a very valuable resource as they have often known the child and family for many years and have unique insight into both developments as well as family dynamics. Observation of the child and adolescent in groups and interacting with peers is a particularly helpful aspect of inpatient hospitalization and can provide clues to the social and emotional strengths and challenges that they experience in the community. Inpatient treatment interventions are individualized and should focus on maximizing strengths as well as addressing any vulnerabilities. Several different modalities of treatment are potentially implemented during the inpatient stay, including psychoeducation, individual therapy, psychopharmacology, and family work. Collateral information from families and outpatient providers aids in understanding the progression of the child's decompensation and resultant admission. It additionally provides information regarding previously tried medication as well as therapy techniques. Details of the dosing of medication, duration of treatment, adherence, and reason for discontinuation are all important to obtain. Typically, the child will be returned to the prior providers, and therefore, collaboration about treatment is essential for continuity after discharge.

Communication of facts about the treatment and time to effect is crucial. Detailed conversations with children and parents about possible side effects and expectations can strongly enhance engagement and adherence. And although in many localities, children and adolescents may not be legally required to provide consent, it is helpful for them to understand the risks and benefits of medication and provide assent. Families and children often have anxiety about taking psychotropic medications, which can be driven by misinformation and misconceptions. Open communication between the treatment team and the family about these beliefs can increase understanding and long-term adherence. It is also important to keep in mind that many children and adolescents are medication naive and to be mindful that their first experience with psychopharmacology in the inpatient setting may determine their perception of psychiatric medications as a whole, as well as their lifelong views on psychiatric treatment. Appropriate lab tests should be obtained prior to starting medications. In addition, the risks and benefits must be weighed for each treatment, and the reality of adherence to the treatment must be considered. For families in which medication adherence is problematic, the use of visiting nurse services may be a consideration. A large and arguably most important part of an inpatient stay is implementing support services in the community. This may be a variety of different types of outpatient treatment options and may include recommendations for education or further testing by the child's school.

Transitional Age Youth. With advances in research on brain development over the past two decades, the traditional view of adolescence ending at age 18 has been debunked, and the term transitional age youth has started to be used for adolescents between age 15 and young adults up to age 25. There has been an increasing voice that this age group deserves special attention and has a unique set of developmental challenges. Some hospitals have started to implement transitional-age youth units for young adults and older adolescents. It is important to consider developmentally the best ways to aid in transitioning children and adolescents successfully into young adulthood. The groundwork for this should begin as early as possible and can be an important topic of milieu programming, as well as family work.

Partial hospital and Intensive Outpatient Programs

The structure of PHPs and IOPs is loosely based on inpatient programs. Programs can include a school component (typically called a PHP) or maybe an after-school program (often referred to as an IOP). There can also be differences in the number of days a week and hours a day of programming. Most programs have a staff structure similar to an inpatient unit, and many PHPs and IOPs are an extension of an inpatient program, with many of the children "stepping down" into this less restrictive level of care after their initial inpatient stabilization. The daily routine of a PHP differs from an inpatient unit in that the child wakes up and gets ready at home, and is then transported to the facility. The child then goes home in the late afternoon to return the following day again. It can be particularly helpful that the child can implement learned coping skills in the evenings and then be able to return to the program the following day to process any conflicts or challenges in the home environment. Children and adolescents are typically admitted to PHP/IOP programs due to difficulties at school or home due to mental health problems that have not been responsive to less intense outpatient treatment efforts or as a step down from an inpatient unit. Typically, these children are not currently at acute risk of harming themselves or others. For such patients, once-weekly visits to their outpatient clinicians may not suffice, given the severity of their symptoms, but inpatient admission may not be warranted either. PHPs and IOPs often last for several weeks to months. Assessment and treatment are similar to those in inpatient programs, including contact with outside providers and school personnel, as well as a biopsychosocial developmentally informed assessment. Discharge from a PHP or IOP program depends on symptom improvement and preparation of the child or family to be able to continue care at home with adequate outpatient services. A step down from PHP/IOP may include traditional weekly or biweekly outpatient therapy with medication support by a psychiatrist or include an in-home therapy service such as IICAPS, MST, or MDFT.

Residential Treatment Centers

“Residential” can have a variety of meanings, from residential placement of a child with no other home option to a voluntary placement sought by families to provide 24-hour care, housing youth in the juvenile justice system, or providing therapeutic treatment with the goal of returning to the community. The history of identifying residential stays as treatment began with the establishment of the American Association of Children’s Residential Centers in 1954. Even at that time, there was variation in the definition of residential treatment, though most centered around a psychodynamic orientation, requiring isolation of the child from the environment as a necessary form of treatment. This trend continued through the 1960s and 1970s until the beginning of deinstitutionalization. In the 1980s, with the advent of managed care, intensive outpatient treatment became more economically efficient as compared to residential treatment. Residential treatment also declined due to improvements in psychopharmacology and advocacy for community-based, wraparound, family-centered outpatient care, partly due to the more current philosophy that children are better assessed in the context of the environment in which maladaptive behaviors occur.

Despite best efforts, some children still require residential placement, whether temporary or longer term, although there are increasingly fewer placement options. Children in need of residential treatment tend to have either severe psychiatric and/or developmental disorders, preventing them from being managed through community and family supports alone, or have such circumstances that prevent them from living either at home or in therapeutic foster care. Residential treatment center staffing differs from inpatient and PHP/IOP settings. For example, the psychiatrist on staff often serves in a consultant role, thereby lessening the psychiatrist’s involvement in treatment. There are also fewer nurses per child, but the number of day-to-day residential staff is increased. Lengths of stay have decreased significantly since the 1950s, with further evidence since the 2000s of 3- to 6-month lengths of stay being efficacious. The use of cognitive behavioral and dialectical behavioral approaches has generally replaced psychodynamic approaches. Reduction in seclusion and restraint has also occurred in the residential setting. Residential programming focuses on individual coping skills, relationship building, and basic living skills, including self-care. There is a routine schedule as in inpatient units but with more individualized free time. Parent training and family therapy are integral aspects of residential treatment and prepare the child and family for the child's return home.

CHALLENGES RELATED TO THE COVID-19 PANDEMIC

There have recently been unique challenges in all levels of care for the child and adolescent psychiatric services. As it has with most things, the COVID-19 pandemic has resulted in major modifications to many policies and procedures. For example, in many inpatient units, it was necessary for all children to be tested for COVID-19, as well as to wear masks, implement physical distancing, and increase emphasis on hand hygiene. In addition, on many inpatient units, visitation was either canceled or significantly reduced, including for parents. This led to family meetings often occurring via electronic platforms such as Zoom. However, there have also been some unexpected advantages to these changes. For example, kids who typically do not receive visitors are not comparing themselves as much to their peers who have regular visitors. Further, family meetings between divorced or separated parents can, at times, be more comfortable and productive via Zoom, and family meetings are more convenient and less volatile when parents do not have to be physically present. The same is true for residential treatment settings. Many PHPs and IOPs have transitioned from in-person to entirely electronic platforms. The majority of outpatient services have also transitioned to electronic platforms, with many providers not taking new patients, thus making inpatient and PHP/IOP discharge planning more difficult due to even greater limitations on available outpatient services. It is too soon at the time of this writing to know how clinical care will shape up post pandemic. But it is certain that the “good old days” will not be back in the same way: synchronous means of electronic communication such as Zoom, tired as we may be of them by now, have opened new venues for care and eased much of the logistic burden of in-person meetings and visits, particularly during working hours. The pandemic, dreadful as it has been, will leave some positive changes in its overdue wake.

GLOBAL PERSPECTIVE

There is significant worldwide variability in terms of the availability of inpatient, PHP/IOP, and residential treatment centers for children and adolescents. There are additional variabilities in the criteria for admission, as well as available community treatment options. This chapter was written from the perspective of services in the United States, and even as the same guiding principles may apply, there are country-, cultural- and region-specific variations we have not examined in this chapter. However, it is true worldwide that children and adolescents suffer from psychiatric disorders, and providing safe and adequate mental health care needs to remain a worldwide priority.

FUTURE DIRECTIONS

Advances in psychiatric treatment options, the availability of a range of community services, and a general shift in philosophy have gradually changed the conceptualization of psychiatric inpatient hospitalizations for children and adolescents from one of longer-term treatment to one of multidisciplinary diagnostic needs assessment and acute stabilization. Although there have been many changes over the past decades, the basic dynamics of staff, patients, and family remain unchanged. Adequate assessment of a child or adolescent cannot occur without a thorough assessment of the individual, family, and system of care through a biopsychosocial and developmentally informed lens. It is also necessary to remember that this is often the first encounter with psychiatric services for patients and families and that these early experiences can set the stage for lifelong views of psychiatric treatment and adherence. PHPs and in-home services are valuable services and contribute to the capability for children to have shorter inpatient stays and home reunification after inpatient or residential stays.

ACKNOWLEDGMENTS

This chapter represents an update and revision of previous chapters in this series, most recently by Dana Kober, MD from the last edition. We are indebted to Dr. Kober for her work and retain much of her prose in this current edition.

Further Readings

Blader JC. Acute inpatient care for psychiatric disorders in the United States, 1996 through 2007. *Arch Gen Psychiatry*. 2011;68:1276–1283.

Blader JC. Pharmacotherapy and postdischarge outcomes of child inpatients admitted for aggressive behavior. *J Clin Psychopharmacol*. 2006;26:419–425.

Carlson GA, Chua J, Pan K, et al. Behavior modification is associated with reduced psychotropic medication use in children with aggression in inpatient treatment: a retrospective cohort study. *J Am Acad Child Adolesc Psychiatry*. 2020;59:632–641.

Case BG, Olfson M, Marcus SC, Siegel C. Trends in the inpatient mental health treatment of children and adolescents in US community hospitals between 1990 and 2000. *Arch Gen Psychiatry*. 2007;64:89–96.

Curie C. SAMHSA's commitment to eliminating the use of seclusion and restraint. *Psychiatr Serv*. 2005;56:1139–1140.

Darwish A, Salmon G, Ahuja A, Steed L. The community intensive therapy team: development and philosophy of a new service. *Clin Child Psychol Psychiatry*. 2006;11:591–605.

Geller JL, Biebel K. The premature demise of public child and adolescent inpatient psychiatric beds: part I: overview and current conditions. *Psychiatr Q*. 2006;77:251–271.

Geller JL, Biebel K. The premature demise of public child and adolescent inpatient psychiatric beds: part II: challenges and implications. *Psychiatr Q*. 2006;77:273–291.

Greene RW, Ablon JS, Hassuk B, Regan KM, Martin A. Innovations: child and adolescent psychiatry: use of collaborative problem solving to reduce seclusion and restraint in child and adolescent inpatient units. *Psychiatr Serv*. 2006;57:610–612.

Henggeler SW, Rowland MD, Randall J, et al. Home-based multisystemic therapy as an alternative to the hospitalization of youths in psychiatric crisis: clinical outcomes. *J Am Acad Child Adolesc Psychiatry*. 1999;38(11):1331–1339.

Henggeler SW, Schoewald SK, Rowland MD, Cunningham PB. *Serious Emotional Disturbance in Children and Adolescents: Multisystemic Therapy*. Guilford; 2002.

LeBel J, Stromberg N, Duckworth K, et al. Child and adolescent inpatient restraint reduction: a state initiative to promote strength-based care. *J Am Acad Child Adolesc Psychiatry*. 2004;43:37–45.

Leichtman M. Residential treatment of children and adolescents: past, present, and future. *Am J Orthopsychiatry*. 2006;76:285–294.

Martin A, Krieg H, Esposito F, Stubbe D, Cardona L. Restraint and seclusion reduction through collaborative problem solving: a five-year prospective inpatient study. *Psychiatr Serv*. 2008;59(12):1406–1412.

Meagher SM, Rajan A, Wyshak G, Goldstein J. Changing trends in inpatient care for psychiatrically hospitalized youth: 1991–2008. *Psychiatr Q*. 2013;84:159–168.

Skehan B, Davis M. Aligning mental health treatments with the developmental stage and needs of late adolescents and young adults. *Child Adolesc Psychiatr Clin N Am*. 2017;26:177–190.

Stroul B, Friedman R. *A System of Care for Severely Emotionally Disturbed Children and Youth*. Georgetown University Child Development Center, National Technical Assistance Center for Child Mental Health; 1986.

Van Deusen T, Wilson C, Kim HJ, Qayyum Z, Millard H, Parke S. Transitional age youth with serious mental illness: high acuity patients requiring developmentally informed care in the inpatient hospital setting. *J Adolesc Psychiatry*. 2018;8(3).

Wickizer TM, Lessler D, Boyd-Wickizer J. Effects of health care cost-containment programs on patterns of care and readmissions among children and adolescents. *Am J Public Health*. 1999;89(9):1353–1358.

Wilens T, Rosenbaum J. Transitional aged youth: a new frontier in child and adolescent psychiatry. *J Am Acad Child Adolesc Psychiatry*. 2013;52:887–890.

Woolston JL, Adnopoz JA, Berkowitz SJ. *IICAPS: A Home-Based Psychiatric Treatment for Children and Adolescents*. Yale University Press; 2007.

53.7 Community-Based Treatment

Andrés J. Pumariega, M.D.

Community-based treatment for children and adolescents has undergone significant transformation over the past 30 years. The traditional community mental health model has largely yielded to the community-systems-of-care philosophy and model. This model is based on an individualized, intensive, and flexible approach to service delivery for children and families within the context of their homes and communities as an alternative to treatment in out-of-home settings. This model also focuses on systems issues that affect care, including interagency coordination, service access and utilization, financing, and clinical and cost-effectiveness of mental health and support services provided to children and their families. Psychiatry, which was central in the traditional model, has only recently re-engaged itself as a participant in this new model. However, as the community systems of care model has been scaled up at regional and statewide levels, inadequate fidelity and resource allocation continue to result in a lack of access to community-based services and excessive utilization of residential services, many of lower quality and therapeutic value. More recently, in response to the shortages in child psychiatrists and requests by pediatric primary care physicians for greater access to child mental health services, collaborative and integrated behavioral health models have begun to emerge to form more effective alliances with this important sector of care. Psychiatrists are meeting the challenge of reaffirming their unique value in these models and in integrating their developing clinical and scientific knowledge and skills base.

The increasing need for and utilization of child mental health services in the United States has spurred their rapid growth. Recent studies suggest that approximately 20% of children and adolescents have a diagnosable mental, emotional, or behavioral disorder, with 5% having a serious mental illness and emotional disturbance. A number of morbidities have been associated with childhood emotional disturbance and mental illness, including suicide, homicide, substance abuse, child abuse, teenage pregnancy, school dropout, youth crime, and associated institutionalization and incarceration. However, less than 1% of children in the United States receive mental health treatment in a hospital or residential setting, with another 5% receiving treatment in outpatient or community-based settings; greater than two-thirds of children in need still receive insufficient or no mental health services whatsoever. There is evidence that the recent COVID-19 pandemic has significantly increased levels of mental health needs and rates of morbidity, as well as rates of mental health utilization among children and youth, aggravating these gaps in services even more.

In contrast, funds dedicated to the financing of public mental health services for children have grown significantly since the 1970s. Medicaid, the public insurance program for the poor and disabled, funds a significant proportion of child mental health services in the United States. In January 2020, over 32 million children were enrolled in Medicaid, and another 6 million were enrolled in Children’s Health Insurance Program (CHIP). Mental health and substance abuse services are more than 45% of the over 600 billion dollars of overall Medicaid costs. The federal government contributes from 50% to 75% of costs, depending upon the state. The total public costs of behavioral health disorders among children and youth come to $247 billion annually, including services costs, lost productivity, and the cost of criminal activity. During the 2009 to 2011 period, an average annual total of $10.9 billion (in 2011 dollars) was spent on the treatment of mental health disorders among children aged 5 to 17 years, with 44.0% of expenditures for prescription medicines ($4.8 billion) and 34.9% were for ambulatory visits ($3.8 billion). The mean annual expenditure per child aged 5 to 17 years for the treatment of mental health disorders (among those with expenses for mental health) averaged $2,192 during 2009 to 2011, with mean expenses per child for ambulatory visits of $999 and $1,246 for prescription medications.

The main cost-containment approach to reduce these expenditures has been to control inpatient expenditures, which has more recently led to a major shortage of child and adolescent acute inpatient beds in the United States. A disproportionate level of these expenditures is still dedicated to residential treatment, which is viewed with increasing skepticism, given their relatively high cost, as well as limited documentation of outcome. In addition, there has been increasing documentation of poor practices and abuses in residential treatment as a result of underfunding and decreasing standards and quality of care. As a result of the growth in expenditures, the resources available to fund community-based child mental health and human services are increasingly limited.

BIPOC (Black, Indigenous, and people of color) children and their families have been traditionally served by public community-based mental health and human services. Children from these populations experience higher levels of stressors, such as poverty, discrimination, immigration, acculturation stress, and exposure to violence and trauma, and they are likely to have higher levels of need for services. It has become clearer that these populations suffer from major disparities in access to care as well as access to quality and effective care, particularly African American and Latino youth. A growing awareness and recent reckoning with the long-term impact of racism and discrimination in the United States has pointed to inherent structural racism in societal institutions, including mental health, health, education, and social services, which are now recognized as resulting in such disparities and the lack of cultural competence of such services. The cost of effectively serving BIPOC children and adolescents is in stark contrast to the much higher cost of psychosocial morbidities of racial/ethnic disparities, including lost human potential and the costs of welfare dependency and institutionalization. The United States continues to have the highest youth incarceration rates in the world, the highest levels of out-of-home or community residential placement for youth, and a high rate of children in state custody. Due to lack of funding for effective community-based mental health services, health and human service agencies (schools, social welfare agencies, child protective agencies, juvenile justice, and public health) have shouldered the increasing burden of psychosocial morbidity experienced by these children and youth. These agencies typically address pieces of the service system puzzle, with little to no coordination with other agencies serving the same youth resulting in increasing care fragmentation.

These trends have increased pressures on public child mental health and social service agencies to demonstrate improved clinical and cost-effectiveness, increasingly turning to managed care approaches to finance and organize mental health and social services. The majority of children covered by Medicaid are under managed care plans, including for behavioral health services. However, most managed care approaches were developed with private sector populations in mind, relying on a priori benefit restrictions based on actuarial data on middle-class populations with adequate social supports, who are traditionally lower users of services. When applied to public child mental health services, these approaches deprive children at high risk of effective intervention and preventive services. They also contribute to fragmentation of care and burden shifting to other child-serving agencies and systems, significantly increasing morbidity and acute need. Some Medicaid entities have successfully implemented community-systems-of-care approaches and integrated them with managed care methods, but many have had difficulty integrating them. Over the last 20 years, there has been an exponential growth of class action lawsuits against state Medicaid, educational, child welfare, and juvenile justice agencies failing to meet federally required service mandates, particularly around their high representation of children with significant mental health needs. The most egregious of such lawsuits involve the practice of parents having to give up custody of their children to obtain necessary mental health services.

HISTORY

From its earliest origins, children's mental health services have emphasized a community orientation. In the United States, these services began in the 1890s in response to the perceived need for counseling juvenile offenders rather than incarcerating them with adults. At that time, much like today, the United States was undergoing rapid sociocultural changes due to immigration, industrialization, and urbanization, and these social strains contributed to marked increases in juvenile crime and status offenses. The earliest juvenile courts in Chicago and Boston established clinics that comprised the first child mental health services in the nation. Their success led the Commonwealth Foundation to commission a study in the 1920s that recommended the development of child guidance clinics throughout the United States staffed with interdisciplinary teams of professionals that could serve children and their families. They provided low-cost services oriented to the needs of the child and the family, with treatment modalities evolving to include individual psychodynamic psychotherapy, family therapy, crisis intervention, and even day treatment programs. They served as models for the community mental health centers that, with some alterations, emerged in the 1960s, and many have survived to this day.

The "medicalization" of psychiatry, starting in the late 1970s, served to move child and adolescent psychiatry toward the more hospital-based, tertiary care model. This left the child guidance clinics, and the community mental health centers that followed them, without significant psychiatric input. Many of the children previously receiving psychiatric services in these clinics were then served in free-standing psychiatric hospitals, residential facilities, or detention facilities, often in the custody of child welfare or juvenile justice agencies.

The modern era of community-based systems of care for children was ushered in by the publication of Jane Knitzer's groundbreaking book *Unclaimed Children*, which exposed the aforementioned consequences of neglecting the provision of community-based mental health services for children and their families. Her advocacy and that of others led to the development of the Child and Adolescent Service System Program (CASSP), which assisted all 50 states in the development of an infrastructure for publicly funded, community-based services. The CASSP initiative was supported by the conceptual work of Stroul and Friedman, who coined the term "community-based system of care for seriously emotionally disturbed children" and enunciated the principles behind such systems of care. Stroul and Friedman's work spurred the development of many innovative community-based treatment modalities, as well as a number of model demonstration programs in different parts of the United States that embody the use of these modalities within the context of an organized interagency system of care. In the early 1990s, the Robert Woods Johnson Foundation established eight pilot demonstration community-systems-of-care programs in different parts of the country, which demonstrated the viability of the systems-of-care approach and demonstrated cost savings, as well as less restrictive levels of care. Starting in 1994, the Center for Mental Health Services (CMHS) of the Substance Abuse and Mental Health Services Administration established the Comprehensive Community Mental Health Services for Children and their Families Program. This program has funded 173 community-systems-of-care sites throughout the nation in widely diverse communities. The *Surgeon General's Report on Mental Health*, published in 1999, focused its recommendations for children's mental health around the systems-of-care model and its benefits for service system reform and individualized care. More recently, given the focus on school violence and the threat of terrorism in the United States, there has been additional focus by the federal government on these two aspects of community children's mental health. The Safe Schools Health Students program has funded 240 schools and communities to develop school-based preventive and clinical programs oriented at decreasing school violence and substance abuse, and the National Child Traumatic Stress Initiative funded many sites to adapt, implement, and evaluate promising intervention approaches addressing child traumatic stress in communities. Additional federal funding for children's mental health services was allocated in the 2010s in response to the continued occurrence of school shootings in various communities, as a result of the Affordable Care Act, and more recently in response to the mental health impact of the COVID pandemic in 2020 and 2021.

THEORETICAL ISSUES

The CASSP initiative set forth the principles inherent in community-based systems of care. The key principles include access to a comprehensive array of services, treatment individualized to the child's needs, treatment in the least restrictive environment possible (with full utilization of the resources of the family and the community), full participation of families as partners in services planning and delivery, interagency coordination, the use of case management for services coordination, no ejection or rejection from services due to lack of "treatability" or "cooperation" with interventions, early identification and intervention, smooth transition of youth into the adult service system, effective advocacy efforts, and nondiscriminating, culturally sensitive services.

The wraparound approach is also critical for systems-of-care interventions and services. This approach is based on the principle of individualization of services and proposes that an individualized service plan be "wrapped around" the child, including clinical, support, and nontraditional interventions. A wraparound approach is greatly facilitated by the use of noncategorical and flexible funding for services, which enables a case manager and family to purchase whatever services the child and family needs at that time (including living needs, educational, social support, and even staff security services). Wraparound plans can build a level of service intensity rivaling that of inpatient or residential settings without removing a child from their home, especially using systematic wraparound protocols that have measured fidelity.

Cultural competence is an integral element of community-based systems of care. It implies that practitioners in systems of care develop the necessary attitudes, skill, and knowledge base to serve minority and culturally diverse children and families in their communities, as well as having service systems that develop policies and practices to remove barriers for access to services and make these more responsive to the needs and values of diverse communities. The community-based systems-of-care approach is congruent with the cultural values of ethnic minority populations, which emphasize strong extended family involvement in the life and upbringing of children and the use of natural community resources first in dealing with the emotional and physical problems of family members. These factors have been shown to be protective from some of the morbidities associated with emotional disturbance, such as substance abuse and suicidality. National guidelines for cultural competence for managed Medicaid services have been published by the CMHS, and the Practice Parameter adopted by the American Academy of Child and Adolescent Psychiatry (AACAP) operationalizes elements of cultural competence at both the clinical and systems levels. Building on this work, more recent efforts by various professional organizations and researchers on children’s mental health have focused on the impact of structural racism on racial/ethnic disparities, such as the rising prevalence of psychiatric morbidities such as suicide in BIPOC children and the barriers to access to effective mental health and social services.

Another important principle inherent in this approach is the prioritization of services to “seriously emotionally disturbed children.” This designation includes the presence of an Axis I diagnosis under the criteria of the *Diagnostic and Statistical Manual of Mental Disorders* of the American Psychiatric Association but places equal emphasis on the child’s inability to function in at least one of their domains (e.g., school, home, socially with peers). This reflects the difficulty many children and their families have in accessing necessary services when the diagnosis was not "severe enough," particularly when the child was diagnosed with a disruptive disorder (although many children with disruptive behavioral disorders have comorbid serious mental illness when evaluated systematically). These problems are more pronounced in the diagnostic assessment of ethnic minority youth with serious emotional disturbance (SED), with serious underestimation of comorbid affective and substance abuse disorders by clinicians. Other studies indicate that the level of care received by children is only partially accounted for by their clinical diagnosis, whereas the level of function and psychosocial stressors are stronger predictors. It is also important to note that racial/ethnic disparities around diagnosis result in underidentification of BIPOC children and youth with significant mental health needs.

The movement to managed care has resulted in the application of community-based systems-of-care principles to these approaches. In fact, these principles embody a more clinically based managed-care philosophy. That is, active care management becomes the primary modality for utilization management rather than arbitrary benefits restriction. In combination with the principles of care in the least restrictive environment, systems of care result in children actually receiving higher-quality services with fewer disruptions in their lives and development at a lower cost than with more restrictive, higher levels of care. The AACAP has developed guidelines for states using managed care for children's behavioral health services funded by Medicaid, integrating community-based systems-of-care principles into their contracts with private managed care providers or into their policies and procedures for public managed services. It also has developed (in conjunction with the American Association of Community Psychiatry) an evidence-based level of care determination tool, the Child and Adolescent Level of Care Utilization System/Child and Adolescent Service Intensity Instrument (CALOCUS-CASII). It uses the level of function and contextual variables as the main variables for level-of-care determination, with recommended levels of care based on treatment intensity using systems-of-care and wraparound approaches. The CALOCUS-CASII has demonstrated significant interrater reliability and validity with mental health, child welfare, and juvenile justice populations and has contributed to significant systems-of-care reform in many states, including the state of Hawaii from 2003 to the present.

More recently, the federal Affordable Care Act that has promoted health care reform has promoted the establishment of Medical Homes, where individuals (including children and youth) receive their primary care and also the coordination of specialty care. This has facilitated a movement in both general and child-and-adolescent psychiatry toward collaborative and integrated behavioral health models, to enhance collaboration with pediatric primary care physicians in Medical Homes, and to enhance access to entry-level community child psychiatric services. The full implementation of these models calls for systematic screening of behavioral health needs served by Medical Homes and a stepped model of care, where either (a) the primary care physician directly delivers psychiatric care in collaboration with a care manager, (b) the primary care physician refers the child for stabilization by the child mental health service or specialist and receives the child back once stabilized, or (c) the primary care physician and child mental health service collaborate longitudinally in the child's care. The primary care physician is trained and/or supported by child psychiatric and mental health consultation in screening, assessment, and entry-level behavioral health services as delivered within the primary care setting. There is both ready access to child psychiatric consultation (direct and indirect) and case management support to enable the primary care physician to deliver these services within the medical home and assist with specialty referrals and continuity tracking of the child's course of treatment and outcomes. In integrated models, there is embedded case management and therapist support in the Medical Home, as well as access to regular child psychiatric consultation on the panel of patients served by the Medical Home.

TECHNIQUES

A community-based treatment approach to services begins with access to services facilitated through multiple portals of entry, including direct referral or referral through child service agencies (such as the school or juvenile justice, child welfare, or public health agencies). A triage system should direct children and families to services and providers appropriate to their special needs. A full interdisciplinary team, including a child and adolescent psychiatrist, should be involved in the development of triage procedures and protocols. Comprehensive mental health assessments should be made by licensed providers who are credentialed and trained to evaluate children and families in consultation with a child and adolescent psychiatrist. Such assessments should include clinical/diagnostic, functional, developmental, family, and cultural/community evaluations of the needs *and* strengths of the child and family. Whenever possible, standardized measures should be used that are reliable, valid, clinically useful, culturally competent to the population being served, and useful for programmatic evaluation.

Children and families who require time-limited, less intensive services should have an initial care plan developed by a licensed mental health professional in consultation with a child and adolescent psychiatrist. The initial care plan for children requiring more extensive treatment should be developed by an interdisciplinary team, including all relevant child mental health and child service professionals. The parent(s), other relevant family members, and the child, if appropriate, should be members of this team. The care plan should identify problem areas and strengths that are relevant to the child and family's functioning, as well as community-based interventions to address these needs. The patient and their family should be equal participants with the clinical team in the development and implementation of the plan of care. In some programs, special "child and family teams" that are composed primarily of the child, family, and other natural supports in the community develop the outlines of care plans, and these are then negotiated by case managers with interdisciplinary teams of professionals and agencies. Plans of care should be individualized to the needs of the child and family, including attention to cultural issues, and should follow the child continuously throughout all levels of care, being amended as needed. The plan of care should promote continuous coordination and integration of all applicable services and agencies that have responsibilities for providing services to the child and family (such as primary health care, juvenile justice, schools, and social services).

Treatment and other services should be specific to the unique needs of the child and their family and use approaches that maintain or reintegrate the child into the family and community. Child and adolescent psychiatrists, psychologists, social workers, psychiatric nurses, and other members of the team should participate in the development of service protocols to ensure that the special developmental, physical, mental, and emotional needs of the child are addressed. The system should have as many diverse therapeutic modalities with children and families as possible, particularly those modalities that have demonstrated effectiveness with seriously emotionally disturbed children. These include Medical Homes, traditional and intensive outpatient services, nonresidential community-based interventions (including home-based, day treatment, and school-based services), flexible support services, community-based residential services (such as therapeutic foster homes and group homes), and brief, acute inpatient services when needed. These interventions are aided by the increasing effectiveness of cognitive, behavioral, family-based, and psychopharmacologic therapies, which should be integrated into all community-based interventions and levels of care.

Case management services are essential in community-based treatment, including assessment of problem areas, measurement of functioning level, determination of needs, linkage with necessary resources, care coordination, advocacy, and monitoring of services provided and their outcomes. The child and family should be assigned a case manager who has the necessary skills to ensure seamless coordination of care across all levels, different providers, and the full-time course of illness/disturbance. The frequency and intensity of case management services are proportional to the clinical and psychosocial needs of the child and family. The patient/family should be participants in the case management process, with shared responsibility with case managers for using and coordinating services. Case management protocols should facilitate the coordination of multiagency, multisystem interventions, integrating services from the various providers and agencies (including primary care and health providers), and the development of wraparound plans that are truly individualized for the child and family’s needs. These protocols should include access to flexible noncategorical funds to enable the purchase of flexible services needed for an effective wraparound plan.

ROLES AND FUNCTIONS OF PSYCHIATRISTS

Given the complexity of these new systems of care, the skills and roles that psychiatrists must display go far beyond the circumscribed clinical roles to which they are usually limited. These roles include serving as a front-line clinician, clinical consultant and collaborator to other professionals, clinical team leader, administrative leader in delivery organization/system, quality assurance/improvement consultant,

consultant to interagency teams, and outcome evaluator/researcher in systems of care. Child and adolescent psychiatrists, given their broad biopsychosocial and developmental perspectives, should be best able to integrate and coordinate community-based treatment delivered by multiple professionals with diverse skills. However, given the emphasis on tertiary care models in the training of many child and adolescent psychiatrists, they often have not developed all the skills needed to serve as effective members and leaders of these new systems. Three publications from AACAP, *Guidelines for Training Child and Adolescent Psychiatrists in Community-Based Systems of Care*, the *Practice Parameters for Child Mental Health in Systems of Care*, and *Best Principles for Integration of Child Psychiatry into the Pediatric Health Home* provide guidance on the skills, knowledge, attitudes, training experiences, and practices that psychiatrists need to function effectively within community-based systems of care.

CLINICAL ISSUES

Indications

Community-based treatment approaches are indicated for the treatment of the majority of children with moderate to serious emotional disturbance or mental illness, which is often associated with comorbidities and multiple psychosocial stressors and needs. It is most indicated for the prevention of out-of-home placement of such children and for the treatment of children served by multiple agencies with limited family resources. It can be contraindicated when the child’s family is too fragmented or inaccessible. It often has to be modified with more intensive and restrictive treatment elements when the child faces imminent danger from whatever source (suicide, homicide, abuse, victimization, severe substance abuse), dangerous symptomatology, or serious decrease in function (such as acute psychotic symptoms or delirium). However, community-based treatment can be delivered in these instances as long as staff and physical security and safety needs are addressed through more intensive wraparound services (such as one-to-one staffing or on-site medical services). These services can be de-escalated when such symptoms or conditions are resolved (usually after brief stabilization) or the family is able to function again to an extent to which it can safely and effectively participate in the child’s care.

Limitations

The community-based treatment approach can suffer from limitations when there is a lack of appropriately trained child mental health professionals who are conversant in evidence-based psychotherapies with children and youth and when there is a lack of an adequate and accessible continuum of care with sufficient options for an appropriate level of care intensity to meet the child and family's clinical and functional needs. In addition, a lack of adequate services coordination and integration can lead to significant care fragmentation. This can occur when psychiatric and nonpsychiatric services are segmented, and there is a lack of timely care coordination services, leaving children for inadequate periods of time in inappropriate levels of care intensity (either too low when they are developing more acute needs or too high after no longer needed).

Complications

The misapplication of the community-based treatment approach can lead to children and youth, and their families being underserved. This can result from a lack of access to necessary clinical and support services, lack of access to more intensive services, excessive pharmacologic treatment to compensate for lack of evidence-based psychotherapy services, and lack of timely care management services.

Contraindications

Community-based treatment services are contraindicated for children who suffer from acute symptomatology. These include acute suicidality with imminent danger to self, psychosis or mania leading to risk to self or others or inability to be adequately protected/cared for, or substance abuse that leads to risk to self or others. In addition, the child who is at risk of victimization in their home or community environment requires out-of-home placement. In all these instances, the community-based treatment approach calls for continued coordination with services in the home community and all possible efforts to return the child back to their family or to a normalized community setting with adequate support.

Shaniquah was a 14-year-old African American youth who lived with her Caucasian adoptive mother and adoptive siblings in a mid-sized Southern town. She presented with significant isolation and withdrawal, lack of motivation, irritability, and problems with aggression and fighting with her adoptive mother and siblings at home and with peers and teachers at school. She was expelled from school after the latest of such fights, in spite of her good academic standing, and this resulted in her making suicidal statements with thoughts of means and intent to die. She was taken by the local mobile crisis services to the local child and adolescent psychiatric unit, where she had been admitted twice before over the past year for suicidality and aggression and there was diagnosed with disruptive mood dysregulation disorder and attention deficit hyperactivity disorder, the latter established since age 7 and for which she had received outpatient services from her pediatrician and the local community mental health center. Her adoptive mother felt unable to care for her at home and had been asking for residential services, but the inpatient team referred them to the new Child and Adolescent Action Team (CAAT), an intensive community-based system of care program being established in different areas of her state. Within the team, she had a masters level therapist who worked with her and her family, a female mentor who worked to connect with her informally and motivate her to engage in prosocial activities, a case manager who worked to coordinate care with the school and identify community resources and activities she would engage with, and a child and adolescent psychiatrist. During her intake and child psychiatric evaluation, it was established that she also suffered symptoms of post-traumatic stress disorder (PTSD), and she began to reveal a past history of sexual abuse by her adoptive mother’s former boyfriend, exacerbated by conflicts with her adoptive mother’s current boyfriend as well as triggers by male peers at school. The therapist and child psychiatrist worked closely to address Shaniquah’s PTSD symptoms with both trauma-focused cognitive behavioral therapy (CBT) as well as appropriate pharmacotherapy. The therapist also identified and began to address significant conflicts between the adoptive mother and Shaniquah, with the former minimizing the impact of trauma on Shaniquah as well as the cultural divide between them (seeking to treat her “the same” as her other stepsiblings of different races/ethnicities). This involved much intensive family therapy, which at times was difficult for both but was quite productive. The CAAT Team female mentor also established a close relationship with Shaniquah, serving both as a role model as well as helping to engage her in prosocial activities of interest that validated her Black identity. The case manager on the team worked from early on to improve communication with Shaniquah's school, helping school staff recognize the impact that male peers had on her and engaging her school counselor and teachers in providing more academic and emotional support as well as support for appropriate social activities. This multiprong approach gradually helped the adoptive mother to better support Shaniquah at home and abandon plans for residential placement, instead now supporting her aspirations to enter a special magnet program at school. Shaniquah herself learned to use psychotherapy and mindfulness techniques to reduce her reactivity, developed a stronger bond with her adoptive mother, and re-engaged in academic and social pursuits.

Goals of Treatment

The goals of community-based treatment usually include the following: (1) to restore the child's function within the family, at school, with peers, and in the community; (2) to maintain or reintegrate the child within the family and community and avert separation or disruption of the community tenure; (3) to individualize interventions to the unique clinical and psychosocial needs of the child and family; and (4) maximally to use the strengths and resources of the family and the community on behalf of the child.

ETHICAL ISSUES

Ethical issues can arise in community-based treatment when the child experiences serious symptomatology in the face of family disorganization or risk of harm within the family and home. The ethical conflict arises from the inherent tension between the child's and family's needs for safety and stability and the goals of retaining the child within their family or community and helping both achieve change within their natural environment, avoiding restrictive care and disconnection from the child's relationships and natural supports. A home in which there is a high level of conflict or abuse can be more restrictive for a child than some out-of-home programs due to the degree to which the child's function and activities are constrained. Also, the child may require a more structured or restrictive environment than the family can provide, such as when the child is experiencing suicidal, violent, or severe runaway behaviors. On the other hand, residential placement has been known to result only in temporary respite, with relapse of the child's emotional and behavioral symptoms once they return home since the structure, dynamics, and resources of the home environment have not changed during their absence. Interdisciplinary teams are usually well equipped to face such ethical dilemmas, given their representation of multiple stakeholders; chief among these are the voices of the family and the child and their advocates.

Another inherent clinical ethical dilemma arises when considering whether to place a child on a pharmacologic regimen versus intensive community-based or psychosocial interventions for disruptive or dangerous behaviors or symptoms. In these instances, the tension is between the biases toward less “medicalized” psychosocial interventions over pharmacologic interventions. The short- and long-term risks versus benefits (and evidence base) of any psychopharmacologic intervention for a child should be carefully weighed. However, many community-based interventions can be adjunctively aided by targeted pharmacologic interventions and vice versa, which can further aid in the reduction of restrictiveness, reduce needed doses of medications, and facilitate better integration into the community. The role of the psychiatrist in addressing these ethical dilemmas is critical.

The utilization of limited community-based treatments exclusively to reduce costs without considering the needs of the child and/or family can be another serious ethical breach in the pursuit of community-based treatment. Unfortunately, the financial incentives to provide less intensive care inherent in most managed care approaches can contribute to such dilemmas, potentially endangering the child if they place the child in greater danger or risk. Community-based approaches need to be pursued within organizational and financial structures that do not have arbitrary benefit limits, allow for wraparound approaches, and use on-site case management to avoid such situations.

RESEARCH AND EVALUATION

Research in child mental health services has particularly focused on the effectiveness of community-based treatment. Research in this area has included the evaluation of community-based residential approaches, nonresidential intensive community-based approaches, and the clinical, functional, and cost outcomes of community-based systems of care, as well as factors affecting the need and utilization of mental health services in different populations of children.

Community-Based Interventions

The Surgeon General reports on mental health and on youth violence point to research evidence supporting the effectiveness of a number of community-based interventions for children and youth, such as intensive case management, therapeutic foster care, partial hospitalization, and intensive in-home wraparound interventions (particularly multisystemic therapy [MST]). Other community-based interventions that show promise include school-based interventions, mentoring programs, family support, and education programs, wilderness programs, crisis mobile outreach teams, partial hospitalization, time-limited hospitalization with coordinated community services, and family support services. Studies involving these modalities have demonstrated significantly better outcomes than traditional outpatient or residential services. These include reduced levels of externalizing and internalizing symptoms, improved family functioning, reduced utilization of more restrictive services, diversion from juvenile justice involvement or detention, and improved cost-effectiveness. In addition, a number of in-home preventive parent–child interventions with infants and young children have demonstrated long-term effectiveness in reducing child abuse, as well as adverse behavioral and academic outcomes (such as the Elmira Project by Olds and coworkers in upstate New York and the child abuse prevention projects in New Orleans by Boris Zeanah and colleagues). On the other hand, there is a lack of research evidence or actual lack of effectiveness for a number of more restrictive interventions, such as inpatient hospitalization in isolation of community services, long-term residential treatment, therapeutic group homes, and boot camps.

The wraparound model of care embodies this approach toward multimodal community-based interventions. This model emphasizes aggressive outreach, the use of least restrictive treatment options, and individualized, flexible, and unconditional child- and family-centered services for children imminently at risk or already in out-of-home placements. A number of studies of wraparound programs in different communities across the nation with diverse populations of at-risk children and families have reported positive outcomes in terms of reduction of externalizing behavioral problems, level of function, reduction in out-of-home placement, improved family management skills and function, and consumer/family satisfaction. MST combines home-based, wraparound, and cognitive behavioral interventions within a systemic context. It has been evaluated with youth at risk of detention/incarceration and at risk of psychiatric or substance abuse hospitalization, with significant results in reducing out-of-home placement, externalizing problem behaviors, rates of recidivism, and costs of treatment. A more recent community-based intervention—cognitive behavioral intervention for traumatic stress—has demonstrated effectiveness in addressing traumatic stress among children of diverse cultural backgrounds in school settings.

Community Systems Evaluations

Outcome data on early system-of-care demonstration projects in Vermont, New York, Ventura County, California, and the eight Robert Woods Johnson pilot sites demonstrated reduced hospitalization rates and days, decreased out-of-home placement, and less restrictive placements, significantly lower incidence of negative behaviors placing the child at risk for out-of-home placement, and significantly lower rates of overall problem behaviors. The Robert Woods Johnson pilot sites also demonstrated significantly lower costs using a managed subcapitation approach for funding flexible services for children with SEDs. Studies of the Civilian Health and Medical Program of the Uniformed Services (CHAMPUS)-funded Fort Bragg Demonstration Project and the Stark County CMHS system-of-care site compared their outcomes to those of usual service sites. These programs achieved significantly lower restrictiveness of care, higher family and consumer satisfaction, increased access to services, and greater funding spent on less restrictive services. However, clinical and functional outcomes and costs were not significantly different overall. Critics of these studies point to their short-term nature, and more recent studies have supported that significant clinical and functional outcomes from system-of-care programs are found on longitudinal evaluation.

The Substance Abuse and Mental Health Services Administration (SAMHSA) has sustained a major initiative, the Comprehensive

Community Mental Health Services for Children and their Families, to promote the development and expansion of community-based systems of care. This initiative, also known as the Children's Mental Health Initiative (CMHI), was authorized by Public Law 102–321 to provide funds to public entities for the purpose of delivering comprehensive community mental health services to children and youth who have an SED and their families. From 1993 to 2015, it funded 173 demonstration grants and cooperative agreements in states, territories, counties, and federally recognized tribal entities to create local systems of care that serve children and youth with serious mental health conditions and their families. It has included both a centrally designed evaluation, with required systematic data collection on all children served, and locally administered evaluations. Two main designs have been used: a longitudinal design with multiple assessment baselines at program entry and every 6 months until 36 months (and after program exit in newer sites), and two sets of comparison studies of system of care demonstration and control community sites. Data collected from grantees show that children and youth who received system of care services experienced positive behavioral and emotional health (as measured by the Child Behavior Check List (CBCL) total, internalizing, and externalizing scales), significant reductions in suicidal ideation and attempts, and improved educational outcomes such as reductions in expulsions and improvement in attendance. The average number of psychiatric hospital days was reduced from 0.98 to 0.63 days at 6 to 12 months in grantee sites, and total costs per child involved in juvenile justice went from $ 2,204 at intake to $ 936 between 6 and 12 months, achieving cost savings for all children served when all child service sector services are included. Grantees also made overall gains in applying systems of care principles to transform service delivery systems. Building on the success of these systems of care outcomes, SAMHSA began awarding 1-year *Expansion Planning grants* in Fiscal Year (FY) 2011 and 4-year Expansion and Sustainability grants in FY 2012 in order to stimulate the wide-scale adoption of the systems of care approach that is part of SAMHSA's Theory of Change. As of FY 2015, SAMHSA has awarded a total of 50 1-year *Expansion Planning grants* and 53 4-year Expansion and Sustainability grants.

In addition, some of the current and past CMHS-funded sites have begun to evaluate the added effect of the integration of evidence-based psychotherapeutic interventions within the system-of-care model. Some system-of-care programs have begun to integrate evidence-based manualized therapies (such as parent–child interactive therapy) and community-based interventions (such as MST), demonstrating enhanced clinical outcomes. The Hawaii children's system of care has pursued the most ambitious integration of evidence-based interventions, creating a menu and an algorithm for the application of such interventions, with significantly improved functional and clinical outcomes and decreased level of care intensity of services over time. However, there are major challenges in integrating evidence-based interventions into a community system of care, including maintenance of fidelity, the need for added training resources, attitudinal changes by clinicians, and the overall complexity of the enterprise.

Services for Underserved and BIPOC Children and Youth

Services research literature with underserved minority youth has been sparse and initially focused on documenting racial differences in services provided. African American children tend to remain in foster care for longer periods of time and to have more foster care placements than White children. Studies have shown culturally diverse children to be underrepresented in mental health institutions and overrepresented in child welfare and juvenile justice settings and placements compared to nonminority youth, even when they are equally psychiatrically impaired. Ethnic and racial differences in the diagnoses of culturally diverse adolescents have been identified by some investigators, including overdiagnosis of conduct disorder and psychoses and underdiagnosis of affective, personality, and substance abuse disorders. Significantly lower rates for overall services utilization persist over a number of years, and specifically for the treatment of depression among African Americans and Latinos have been documented. Some studies have found a significant relationship between race and the service needs of racial and ethnic minorities. Studies have also shown higher retention in the treatment of minority youth if the clinician is of a similar ethnic/racial background. These findings have been replicated over the years in spite of the development of much evidence-based psychotherapy and community treatment programs that are effective in serving BIPOC children and families. Significant among them, the SAMHSA Comprehensive Community Mental Health Services for Children and their Families initiative has demonstrated not only significantly improved access by

BIPOC children and families but also equivalent outcomes that are correlated to cultural competence program measures.

Collaborative and Integrated Behavioral Health in Primary Care

Most of the research evaluations so far for collaborative and integrated behavioral health models have focused on their application to adult populations. There are a number of pilot collaborative care programs that have been developed, which have demonstrated significant successes. For example, Massachusetts has successfully demonstrated the Massachusetts Child Psychiatry Access Project (MCPAP), a program that provides timely telephonic psychiatric and clinical guidance to primary care providers (PCPs) treating children with mental health problems. The program allows enrolled primary care physicians to get assistance for any child in their care. The program has enrolled most primary care practices, representing an estimated 95% of all youth in the state, and has high rates of PCP participation, with over 5,000 patients served from 2008 to 2009. Primary care physicians who use the program have reported significantly higher ratings of their ability to serve children with mental health problems. Similar statewide programs have been developed in New York State and Washington State, with similar results. Studies have begun to demonstrate the effectiveness of collaborative and integrated care models. For example, Kolko et al. studied children and caregivers who participated from eight pediatric practices and were cluster-randomized to Doctors Office Collaborative Care (DOCC; n = 160) or enhanced usual care (EUC; n = 161). In DOCC, a care manager delivered a personalized, evidence-based intervention. DOCC was associated with significantly higher rates of treatment initiation (99.4% vs. 54.2%) and completion (76.6% vs. 11.6%), improvement in behavioral problems and parental stress ($p < .05$ to $p < .001$), remission in behavior and internalizing problems, treatment response, consumer satisfaction, and greater perceived practice change and efficacy. Grimes et al. examined mental health service use outcomes for children receiving integrated care via a collaborative practice model (CPM) and examined the impact on access disparities. They studied 228 children referred by pediatricians, 32 referred to a CPM intervention, and 196 youths to usual care (control group). Holding all else constant, children receiving the CPM intervention had four times higher odds of accessing psychiatric evaluations than children in the usual care control group (adjusted odds ratio [AOR] = 4.16, *p* < .01), with seven times odds for engagement in follow-up for CPM youth over those in the control group (AOR = 7.54, *p* < .01).

FUTURE RESEARCH NEEDS

There are still many areas of significant deficit in the evidence base for children's mental health services and systems of care. Areas in which further research is needed include the effectiveness of culturally competent services and culturally specific interventions, the effectiveness of school mental health services, the effectiveness of mental health services integrated within the juvenile justice and child welfare systems, the effect of different cost containment and reimbursement mechanisms, the effectiveness of a systems-of-care approach in combination with evidence-based interventions (both psychological and pharmacologic), and the integration of child mental health and primary care services. The application of population health and mental health services research findings in service system policy and planning is the ultimate objective and value of such research.

FURTHER READINGS

Abram KM, Paskar LD, Washburn JJ, Teplin LA, Zwecker NA, Azores-Gococo NM; Office of Juvenile Justice and Delinquency Prevention; US Dept of Justice and United States of America. Perceived barriers to mental health services among detained youth. OJJDP Juvenile Justice Bulletin. 2015. Accessed August 9, 2023. https://ojjdp.ojp.gov/sites/g/files/xyckuh176/files/pubs/248522.pdf

Alegría M, Green J, McLaughlin K, Loder S. Disparities in child and adolescent mental health and mental health services in the U.S. A William T. Grant Foundation Inequality Paper. 2015.

Asarnow JR, Rozenman M, Wiblin J, Zeltzer L. Integrated medical-behavioral care compared with usual primary care for child and adolescent behavioral health: a meta-analysis. *JAMA Pediatr*. 2015;169(10):929–937.

Bookman-Frazee L, Garland AF, Taylor R, Zoffness R. Therapist's attitudes toward psychotherapeutic strategies in community-based psychotherapy with children with disruptive behavior problems. *Adm Policy Ment Health*. 2009;36:1–12.

Bruns EJ, Pullmann MD, Sather A, Brinson RD, Ramey M. Effectiveness of wraparound versus case management for children and adolescents: results of a randomized study. *Adm Policy Ment Health*. 2015;42(3):309–322.

*Burns BJ, Hoagwood K. *Community Treatment for Youth: Evidence-Based Interventions for Severe Emotional and Behavioral Disorders*. Oxford University Press; 2002.

*Chorpita BF, Becker KD, Daleiden EL. Understanding the common elements of evidence-based practice: misconceptions and clinical examples. *J Am Acad Child*

Adolesc Psychiatry. 2007;46:647–652.

Cohen R, Ventura AB. The status of care for at risk youth prior to the 1990’s. In: *Transforming Children’s Mental Health Policy into Practice: Lessons from Virginia and Other States’ Experiences Creating and Sustaining Comprehensive Systems of Care*. Lexington Books; 2017:3–16.

Daleiden EL, Chorpita BF, Donkervoet C, Arensdorf AM, Brogan M. Getting better at getting them better: health outcomes and evidence-based practice within a system of care. *J Am Acad Child Adolesc Psychiatry*. 2006;45:749–756.

FAIR Health White Paper. *The Impact of COVID-19 on Pediatric Mental Health: A Study of Private Healthcare Claims*. FAIR Health; 2021.

Franco E, Soler RE, McBride M. Introducing and evaluating parent-child interaction therapy in a system of care. *Child Adolesc Psychiatr Clin N Am*. 2005;14:351–366, x.

Grimes KE, Creedon TB, Webster CR, Coffey SM, Hagan GN, Chow CM. Enhanced child psychiatry access and engagement via integrated care: a collaborative practice model with pediatrics. *Psychiatr Serv*. 2018;69:986–992.

Holt W. *The Massachusetts Child Psychiatry Access Project: Supporting Mental Health Treatment in Primary Care. Case Study 32: High-Performing Health Care Organization. Case Study 32*. Commonwealth Fund; 2010:18.

Kataoka SH, Stein BD, Jaycox LH, et al. A school-based mental health program for traumatized Latino immigrant children. *J Am Acad Child Adolesc Psychiatry*. 2003;42:311–318.

Kolko DJ, Campo J, Kilbourne AM, Hart J, Sakolsky D, Wisniewski S. Collaborative care outcomes for pediatric behavioral health problems: a cluster randomized trial. *Pediatrics*. 2014;133(4):e981–e992.

McMahon T, Hamlett N, Olezeki C, Van Deusen T, Harris N, Flanagan D. Children, adolescents, and young adults in the publically funded system of care. In: Jacobs SC, Steiner JL, eds. *Yale Textbook of Public Psychiatry*. Oxford Press; 2016;133–154.

Pumariega AJ. Residential treatment for youth: introduction and a cautionary tale. *Am J Orthopsychiatry*. 2006;76:281–284.

Pumariega AJ, Atkins DL, Rogers K, et al. Mental health and incarcerated youth. II: service utilization in incarcerated youth. *J Child Family Stud*. 1999;8:205–215.

Pumariega AJ, French W, Millsaps U, Moser M, Wade P. Service intensity/level of care determination in a child welfare population. *J Child Fam Stud*. 2019;28:1502–1511.

*Pumariega AJ, Rothe E, Mian A, et al; American Academy of Child and Adolescent Psychiatry (AACAP) Committee on Quality Issues (CQI). Practice parameter for cultural competence in child and adolescent psychiatric practice. *J Am Acad Child Adolesc Psychiatry*. 2013;52(10):1101–1115.

Rowland MD, Halliday-Boykins CA, Henggeler SW, et al. A randomized trial of multisystemic therapy with Hawaii’s Felix class youths. *J Emot Behav Disord*. 2005;13:13–23.

Stambaugh LF, Mustillo SA, Burns BJ, et al. Outcomes from wraparound and multisystemic therapy in a center for mental health services system-of-care demonstration site. *J Emot Behav Disord*. 2007;15:143–155.

*Stroul BA, Friedman RM. *A System of Care for Severely Emotionally Disturbed Children and Youth*. Georgetown University Child Development Center/CASSP Technical Assistance Center; 1986.

U.S. Department of Health and Human Services. *Mental Health: A Report of the Surgeon General*. Center for Mental Health Services/Substance Abuse and Mental Health Services Administration/U.S. Department of Health and Human Services; 1999.

U.S. Department of Health and Human Services. *Culture, Race, and Ethnicity: A Supplement to Mental Health: A Report of the Surgeon General*. Center for Mental Health Services/Substance Abuse and Mental Health Services Administration/U.S. Department of Health and Human Services; 2001.

*U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration Center for Mental Health Services, Child, Adolescent and Family Branch. *The Comprehensive Community Mental Health Services for Children with Serious Emotional Disturbances Program.* Report to Congress, 2015. Publication No. PEP16-CMHI2015.

Vaughan T, Pumariega AJ, Klaehn R. Systems of care under legal mandates. In: Pumariega AJ, Winters NC, eds. *Handbook of Child and Adolescent Systems of Care: The New Community Psychiatry*. Jossey Bass; 2003:414–431.

Winters NC, Chenven M, Pumariega AJ. Managed care and systems of care. In: Pumariega AJ, Winters NC, eds. *Handbook of Child and Adolescent Systems of Care: The New Community Psychiatry*. Jossey Bass; 2003:380.

*Winters NC, Pumariega AJ. Work group on community-based systems of care, committee on community psychiatry, work group on quality issues. Practice parameters for child mental health in systems of care. *J Am Acad Child Adolesc Psychiatry*. 2007;46:284–299.

53.8 The Treatment of Adolescents

Khadijah Booth Watkins, M.D., M.P.H., Steven C. Schlozman, M.D., and Eugene V. Beresin, M.D., M.A.

The treatment of adolescents must begin with a definition and developmental understanding of adolescence. Although most developmental theorists view adolescence as roughly coinciding with the onset of puberty, it is important to note that the criteria for both the onset and the conclusion of adolescent development are complex and multidetermined. Biologic, psychological, social, cognitive, and cultural paradigms, all play a role in the conceptualization of what exactly constitutes adolescent development.

From a biologic perspective, adolescence begins with the onset of puberty. It is important to note that puberty is not limited to the development of secondary sexual characteristics. In addition, neurodevelopment research suggests that during puberty, the brain undergoes a number of important changes. These changes are most likely necessary for the emerging adolescent to move away from what Piaget characterized as a relatively rule-bound and concrete understanding of the world toward an increasing capacity for abstract thought. Neuron migration, changes in myelination, pruning of gray matter, and increases in white matter, all lead to changes in frontal lobe activity and have been implicated as playing major roles in long-observed cognitive changes. However, there is also growing evidence for incomplete frontal lobe myelination leading to decreased capacities for complicated problem-solving activities in the presence of hotly affective environments. Thus, despite substantial development of brain regions responsible for executive function, the relatively poor myelination of these regions explains some of the classically observed lapses in judgment and some of the impulsivity that characterizes adolescent behavior. In this fashion, Piaget's observations have been validated and also challenged by a growing understanding of adolescent neurobiology. During the ages of 14 to 26, cerebral cells die, a process known as apoptosis. Also, during this time period, myelination of neurons from lower centers, such as the amygdala, cingulate gyrus, hippocampus to the cortex, and the orbitofrontal cortex, takes place. This important neurobiologic task helps the developing adolescent move from their behavior being driven by impulse and emotion, to behavior which is based more on complex, rational decision making, reflection, and problem-solving processes.

Psychologically, adolescence has typically been viewed as a period of separation and individuation coupled with an increasing desire to form a coherent and consistent sense of self. These tasks, roughly characterized as identity formation by theorists such as Erik Erikson, Anna Freud, and Peter Blos, are central to the experience of the adolescent as they continue along the developmental trajectory. According to Erikson's epigenetic model, adolescence is determined by the quest for identity formation at the risk of role diffusion. The onset of important advances such as the development of sexual and romantic attractions, the formation of political and ethical opinions, and the willingness to differ in opinion from both peers and adult caregivers are all key elements of these psychological changes.

In addition, G. Noam and others have proposed that during early adolescence, group identity and cohesion plays a much more important role than individual identity formation. Noting that belonging to a group and conforming to the norms of that group are important aspects of the early adolescent years, these scholars theorize that adolescents first experiment with separation and individuation by embracing the rules and regulations of their peer group. In “Eriksonian terms,” this developmental challenge could be coined *inclusion at the risk of exclusion.* To this end, as they mature, healthy adolescents move away from all of the rules of any one peer group and instead choose those aspects of many different groups that best suit their developing sense of self. The normal early-adolescent emphasis on popularity thus grows to also include the later emphasis on being comfortable that one can maintain a coherent identity without relying on the restrictions and rules of any single peer group.

Social norms and pressures contribute to adolescent development in many ways. Ideals of beauty, power, success, masculinity, and femininity are, to a large extent, socially determined. In modern industrial Western society, the adolescent’s individual and group identities are significantly influenced by mobility and the increasing role of media as a means of information, entertainment, and communication. New technologies such as cellular telephones, broadband connections, and the internet have expanded the adolescent’s circle of relationships and generated instant access to groups of individuals, movies, music, news broadcasts, streaming videos, and advertisements. Although group norms and role models previously were confined to relationships in real time, adolescents today are barraged with detailed and graphic information about the intimate lives of celebrities and other public figures in the news. Current events, often dramatic and traumatic events, are instantly available to the adolescent. The rapid transmission of information plays a major role in both defining the world for the adolescent and providing a constant stream of data through which the adolescent may become immersed and influenced in powerful ways.

Although the media is no substitute for life experience in real time, the power of the media to create a social milieu with figures for identification, imitation, and role modeling has grown exponentially in recent years. For example, studies have demonstrated that the media has a powerful effect on adolescent body image and self-esteem and has complex manifestations across various ethnic, racial, and cultural groups. Also, in the context of increased media use and exposure, there has been a dramatic increase in loneliness in Gen Z and millennials. The BBC, Pew, and Cigna studies, all suggest consideration be given to the increase in social media use as a significant factor, albeit, not the sole source of the upward trend in loneliness. In the Cigna report from January 2020, Gen Z was found to have the highest loneliness score. Also in the Cigna report as well as other studies, it was shown that loneliness is linked to anxiety and depression and it was found that there were significant reductions in loneliness, depression, and anxiety over time with decrease social media usage. All of these factors have complicated the tasks of separation and identity formation in many ways that are only beginning to be understood, and these social changes have led to worries regarding addictive internet use, which seems to start most commonly during adolescence. These issues are addressed later in this section.

Finally, as for all developmental stages, adolescence exists within a specific cultural milieu. Although virtually every culture embraces rituals intended to recognize both the onset of adolescence and the transition to adulthood, the conceptualization of what characterizes adolescence varies in terms of cultural norms. What one culture might consider a normal development is often somewhat startling taken out of cultural context. As the world becomes increasingly multicultural, these considerations are especially important. For example, the Asian youth who begins to question his parents' values is often viewed very differently in Western versus Eastern contexts. Although someone with a Western outlook might view this as a healthy and normal emotional development, the youth's parents might see it as particularly pathologic and dangerous. Cultural differences such as this are obviously generalizations, but keeping these potential differences in mind is extremely important to any clinical consideration of adolescent populations.

CONTROVERSIES IN CONCEPTUALIZING NORMALITY AND PATHOLOGY IN ADOLESCENCE

The period of adolescents has been redefined and conceptualized over time to include young adults, many of whom have not yet assumed adult roles, up to age 24 years. There is a subset of adolescents that "fail to launch." They experience significant challenges with transitioning from high school to college: they have work and career problems, family conflict, challenges in friendships and romantic relationships, and limited independence. Also contributing to the prolongation of adolescence is increase in the number of youths seeking post–high school education, which is now required for success resulting in a delayed earning potential. With that, coupled with the lack of job opportunities and the increased costs of education, the ability to live and thrive independently has been delayed.

The adolescent's quests for identity and separation have led both clinicians and popular culture to view adolescence as a normally tumultuous period. Although most adolescents navigate the transition from childhood to adulthood without serious incident, there are times of upheaval that are experienced as real crises. These periods may be obvious sources of concern to significant figures in the teenager's life, such as parents, siblings, teachers, or other adults, or they may be concealed and go unnoticed. For most adolescents, the movement from complete dependence on family to independence is at once exhilarating and terrifying, and at other times, it is filled with profound despair. Highly charged affect is quite typical of adolescents and is often intense and labile. In the separation process, the adolescent must in one way or another reject their parents and at other times seek parental approval and nurturing—even babying. It is not unusual, for example, to see an adolescent girl's bed covered with stuffed animals, and at the same time to overhear her talking about moving out of the house or fantasizing about sexual intrigues.

Although the normal adolescent is capable of understanding the consequences of their actions, teenagers typically live "in the moment" and are occasionally blind to risky behavior. Because normal myelination proceeds in a caudal-to-rostral pattern, emotionally charged moments may not enjoy efficient use of the more sophisticated regions of the brain until individuals are in their mid-20s. These neurobiologic processes are compounded by cultural norms that create a sense of grandiosity and invulnerability. However, at other times, the adolescent's omnipotence evaporates as the teenager fears a multitude of potential dangers, from physical illness to kidnapping. Finally, the quest for identity carries with it substantial self-consciousness, characterized by fear of public exposure and shame. These fears may result in reluctance to share feelings or discuss issues that are potentially humiliating.

In addition, all adolescent behavior is influenced by personal traits, such as temperament, and by the environment, including relationships with peers, siblings, and parents and the community in which the teenager lives. Consequently, there is a great deal of heterogeneity in adolescent behavior. To further complicate matters, developmental advances are never linear. Progress is punctuated by periods of regression to earlier stages. It takes considerable clinical acumen or a great deal of familiarity with a particular teenager to understand what is normal for one adolescent or pathologic for another.

Some researchers feel that viewing adolescence as a tumultuous period creates a setting in which adolescent tribulations and “out-of-control” behavior are viewed as normative and therefore not deserving of clinical attention. Researchers such as Offer have instead suggested that whereas most adolescents are in fact well adjusted and not psychologically troubled, the tendency to view adolescence as “normally difficult” potentially risks missing the population of youth who are legitimately suffering. In fact, research suggests that approximately 20% of the adolescent population in Western settings is seriously in need of clinical intervention. Some of this attention is needed for managing particularly charged crisis points in normal development for youth at risk, whereas other attention is required to treat the onset of many psychiatric disorders that achieve greater prevalence during adolescence. Among these are mood and anxiety disorders, schizophrenia, eating disorders, and substance use disorders. The suicide rate increases tremendously throughout adolescence: Data from the Centers for Disease Control and Prevention (CDC) show that the suicide is the second leading cause of death in youth aged 10 to 24 years. In addition, data from the Youth Risk Behavior Survey (YRBS) published by the CDC show that 18.8% of teens “seriously considered” suicide during the 12-month period during 2019. Prevalence estimates were highest among females (24.1%); White non-Hispanic students (19.1%); students who reported having sex with persons of the same sex or with both sexes (54.2%); and students who identified as lesbian, gay, or bisexual (46.8%).

Although the overall rates of suicide had decreased compared to the rates in 2018, there is growing concern about the increase in the rates of suicide among Black adolescents. Thus, the psychiatric treatment of adolescents has significant public health repercussions.

Much of these challenges are driven by the paradoxical lack of careful thought among adolescents, who nevertheless clearly have the capacity to embrace rational and abstract problem-solving approaches. In short, often the overall judgment and risk assessment of adolescent thinking is markedly different from that of adults. For example, whereas adults would rarely accept a challenge to drive recklessly on the freeway, such a scenario is not unheard of in late adolescent populations in spite of their growing capacity for sound judgment and abstract reasoning. Some researchers have suggested that the affectively charged environment of the adolescent leads to an increased willingness to take risks, sometimes called a “hot cognition” in the cognitive developmental literature. Again, this is best understood in the setting of the incomplete neurodevelopment outlined earlier. Thus, whereas adolescents might appear adultlike in their thinking during calmer moments, affectively volatile environments can pose a particular risk to their well-being. These observations serve to accentuate the need to view adolescent behavior as potentially in need of clinical attention.

Another area of controversy involves the interplay of media and popular culture in affecting adolescent behavior. Many researchers have suggested that violent or disturbing media increases the likelihood of psychosocial problems within adolescent populations. These studies have led to policy statements by a number of professional organizations, including the American Academy of Pediatrics and the American Psychological Association, in which suggestions for curtailing media impact are made clear. Other researchers have noted the potential use of the media in both understanding adolescent thinking and in allowing adolescents to explore difficult issues through displacement and projection. From a clinical perspective, therefore, the issue of media effects on adolescence is extremely important, and any discussion of the psychiatric treatment of adolescence must take into account this key controversy.

Among the major concerns surrounding digital media use in adolescence is the ways cell phones and social media are handled: Nearly 95% of all teens own a smartphone and 45% are online on a near-constant basis. Fifty-one percent log onto social media more than once a day and 97% use social media such as TikTok, Twitter, Snapchat, YouTube, Instagram, or Facebook. There are considerable dangers of social media use in adolescents.

Though most teens think peers are nice, 88% witnessed others being cruel or mean. Considerable attention has been given to the serious impact of cyberbullying among adolescents. The 2019 cyberbullying data from the Cyberbullying Research Center found that approximately 37% of the students reported experiencing cyberbullying in their lifetimes and 15% admitted to cyberbullying others at some point in their lifetime. Female students, LGBTQ students, and students not sure of their sexual identity were found to have the highest prevalence.

Cyberbullying increases dramatically from grade school through high school and may have significant psychiatric consequences, including depression, suicidal thoughts and behavior, anxiety and post-traumatic stress disorder, social isolation, low self-esteem, poor academic performance, and alcohol and substance use. Furthermore, bullying has been shown to have lifelong psychiatric consequences, particularly anxiety and mood disorders.

Many teens reference their sexual behavior, substance use, and/or violence through digital media. Sexting is a particularly worrisome problem in teens. Sexting is the use of text messaging or other means to post sexually explicit photos. Up to 20% of teens are found to engage in sexting, and while studies vary widely, probably 15% have sent an image, and as many as 27% have received images.

Nonconsensual sexting is extremely common with approximately 1 in 8 youths forwarding explicit images or messages without the permission of the original sender, and about 1 in 12 youths have a sext forwarded without their consent. A 2018 meta-analysis found that one in 5 youths under 18 are exposed to unwanted sexual material online.

The reasons for sexting are highly variable but usually involve the desire to maintain or seek a relationship, peer pressure, boredom, joking around, or showing off. While most teens feel the recipient would want to receive the photo and keep it to themselves, forwarding photos is a significant problem. Most teens are unaware that digital media affords recipients making a photo “go viral” and expand to a very wide audience. The consequences of such exposure may be profound and serious. Many teens who engage in sexting are deeply shamed, anxious, and depressed and few are aware that there could be potential legal problems including a risk of being charged with disseminating child pornography and placement on a sex offenders registry if convicted.

Teen drivers have a higher rate of fatal accidents which can be attributed to their immaturity, lack of skills, and lack of experience and distracted driving can make it difficult to react during a potential crash. Distracted driving is defined as any activity that takes the driver’s attention away from focusing on the road and driving which can include talking to peer passengers, talking or texting on a cell phone, changing the radio, eating, or applying makeup.

Specifically, the impulsive use of the cell phone while driving, which occurs all too often, can lead to injury or death. In fact, motor vehicle accidents are a leading cause of death and nonfatal injury among U.S. adolescents.

The data from the 2019 CDC’s Youth Risk Behavior Surveillance System (YRBSS) survey showed that 39% of high school students who drove in the past 30 days texted or emailed while driving on at least one of those days. Also of note, students who engaged in distracted driving were also more likely to not always wear a seat belt, they were more likely to ride with a driver who had been drinking alcohol and more likely to drive after drinking alcohol.

There can be serious repercussions of social media such as exposure to harmful or inappropriate content, exposure to dangerous people, the oversharing of personal information, privacy concerns including data collection and identity theft, and interference with sleep. Of note, one such danger of social media is the suicide contagion or copycat phenomenon. There have been numerous studies of copycat suicides that have been well documented and the copycat phenomenon is a significant issue for adolescents, replicating celebrity or fictional portrayals of suicide. It’s imperative that parents are aware of what they have seen or read and discuss their reactions. When teens hear of celebrities, or even fictional characters in the media dying by suicide, there is a 2-week window in which they are vulnerable to attempt suicide. The CDC and American Foundation for Suicide Prevention set guidelines for journalists on how to report on deaths by suicide that does not sensationalize, report details of the means, and instead, focus on the survivors. Mention of the method of suicide in reporting has been linked to suicide contagion, or “copycat suicide.”

While there may be dangerous consequences of social media, there are potential benefits. For many teens, the immediate contact affords nurturing social relationships and connection, which is particularly true for teens with high-functioning autism spectrum disorder and shy teens.

Connection and social relationships are key components to building resilience.

Teens are able to stay connected to friends, meet new friends that share similar interest, share artwork and music, and express themselves.

Social media can foster separation from parents and increased autonomy, helping shy or anxious teenagers develop social interactions, and providing forums for discussing moral, social, or religious issues in concert with the adolescent’s increased cognitive capacity. However, given the potential serious risks, it is imperative that clinicians know the media use of their patients. Media literacy is a critical and uncharted aspect of the care of adolescents, and one that should not be neglected. Media literacy on the part of parents, pediatricians, and allied health professionals will serve an integral role in providing the important and necessary supervision as well as guidance for the adolescent.

There continues to be concern regarding addiction as it relates to the internet and gaming. The World Health Organization, in their 11th edition in 2019, included gaming disorder, to be considered an “addiction,” much like gambling addiction. The American Psychiatric Association, however, does not currently include internet gaming addiction in the DSM-5, although it has been noted that these behaviors are worthy of future study in the context of internet gaming disorder.

Proponents of these concerns note that many people become enamored of internet activity during adolescence at the expense of normal functioning. These adolescents are said to ignore more normal social behavior and spend increasing time engaged in internet activity. Clinicians who are more skeptical of the concept of internet addiction point to the lack of good prospective studies, confounding issues such as different uses of the internet among those who display worrisome internet behavior, and poor understanding of molecular mechanisms such as dopamine reward systems that have been postulated in other addictive problems. In addition, what is considered socially normal increasingly involves the internet. Thus, fundamental questions, such as how much time on the internet is too much or whether activities such as internet gaming and gambling hold the same risks as internet pornography, need to be better explored.

Finally, in considering media effects on adolescents, one must consider the ongoing debate regarding massively increased use of video game activity among teens. Because many of these games adhere to increasingly realistic and complex narrative themes, there is concern that especially violent video games will desensitize adolescents to aggressive behavior. In addition, there is also the risk of internet safety and engaging with strangers. However, there are investigations that suggest that video games offer opportunities for virtual displacement of aggression that might otherwise be acted out in the “real world” as well as opportunities to connect with peers and escape into fantasy as ways of coping. As with research concerning internet use, data regarding video games are quite mixed, and much study is needed for a better understanding of the nuances of these activities.

WHEN DOES AN ADOLESCENT REQUIRE TREATMENT?

Psychiatric treatment is indicated for any individual when there is a disturbance in thought, affect, or behavior that affects normal functioning. For adolescents, this would include impairments in a teenager’s biologic functions, such as eating, sleeping, and other daily personal routines, as well as problems with relationships with family and peers, problems in school, and problems in occupational and recreational activities. Although it is true, as noted earlier, that normal adolescence involves a certain degree of emotional and/or behavioral volatility, most adolescents overall have a rather smooth course, punctuated by crisis points that usually resolve quickly.

The major issue for parents, caretakers, and clinicians is making the discrimination between a normative developmental crisis that will eventually self-correct and a behavioral change that indicates a deeper problem worthy of clinical attention. This distinction may not be easy for a number of reasons. First, there are often discrepancies between the way an adolescent looks and how they feel. For example, an adolescent with a depressive disorder who is contemplating suicide may appear no different than before the episode, somewhat quiet and contemplative, or mildly irritable. On the other hand, teenagers with profound insecurity and social phobia may present as arrogant and seemingly fearless. Second, although many adolescents are capable of effective communication, there is often a reluctance to share deeply troubling thoughts or feelings with friends, parents, or other adults. This may be due to developmental lags in expression, but more commonly, it is due to the deep self-consciousness and feelings of vulnerability that are typical of normal adolescence. Consequently, many adolescents are misunderstood by parents and peers, which only compounds a sense of aloneness and personal distress. Third, there may be cultural taboos, stigma, or a sense of betraying family expectations that obstruct a teenager from expressing emotional or psychological problems. Despite these obstacles, there are important questions clinicians should keep in mind in determining whether an adolescent is in need of treatment:

- Has there been a developmental lapse, for example, a falloff from the trajectory with respect to functioning in school, home, friends, and recreational activities?
- What is the level of emotional distress?
- What is the meaning of the problem to the adolescent? Although a problem may seem minor for an adult, it may carry tremendous weight for the adolescent.
- How vulnerable is the adolescent to stress? Is there a range of effective coping skills, or is the adolescent limited in managing stress?
- What supports does the adolescent have with other adults at home or in the community?
- Does there appear to be a recent onset of a psychiatric disorder? Is there a past history of psychiatric disorder and/or treatment? Is there a positive family history of such disorders?
- How impulsive is the adolescent?
- Is there family psychopathology, such as mood disorders, substance use disorders, or severe conflict?
- Are there major changes in the family, such as divorce, unemployment, and so on, that might seriously disturb the adolescent?
- Are there concerns about physical or sexual abuse or neglect?
- Have there been recent losses in the family or in peer relationships, such as loss of a friend, romantic relationship, failure to make an athletic team, and so on? How vulnerable is the adolescent to loss?
- How capable is the adolescent in managing affect, such as anger or sadness?
- Are there other acute or chronic medical conditions in the adolescent or in the family?
- Are there cognitive limitations, such as mental retardation?
- Have there been any major media events, such as suicides, school shootings, terrorism, or copycat phenomena, that have had a strong effect on the adolescent?
- Has there been a significant trauma to the adolescent outside of the family, for example, date rape, sexual harassment, bullying, scapegoating, and so on?

Answering these questions requires gathering data from multiple sources, including the adolescent, family members, teachers, pediatrician, medical records, or other members of the community. The research literature demonstrates that in children and adolescents, there is a poor correlation between self-report and reports from others regarding behavioral symptoms. Hence, data collection must include more than one informant.

MODALITY OF PSYCHIATRIC TREATMENT

The psychiatric treatment of adolescents can occur on a number of fronts. Treatment may focus on individual or group settings and may encompass interventions from pharmacologic, psychosocial, and environmental perspectives. Thus, the treatment of adolescents encompasses considerations of individual need, family and community resources, clinical impressions, and biologic history. Furthermore, treatment often brings together multiple aspects of the adolescent’s world. Legal, academic, medical, and psychological issues, all play potential roles in any clinical intervention.

Biologic Interventions

Biologic treatments primarily involve medication interventions. Agents such as antidepressants, conventional mood stabilizers, typical and atypical antipsychotics, anxiolytics, and psychostimulants, all are used in adolescent populations.

There are other biologic treatments used in the treatment of adolescents such as electroconvulsive therapy (ECT), transcranial magnetic stimulation (TMS), and ketamine treatment; either intravenous infusion or nasal spray.

Adolescent depression can result in significant disability and is potentially lethal. All three modalities are used in the treatment of refractory depression. ECT is currently the most effective modality for major depression that fails to respond to other treatments. Ketamine is also used in the treatment of refractory depression, and if a teen responds to ketamine, it can rapidly reduce suicidality and relieve other serious symptoms of depression. TMS, while shown to be effective in treating adults who did not respond to pharmacologic interventions, it has not demonstrated the same efficacy in adolescents.

Obviously, choosing a pharmacologic regimen is based on clinical and diagnostic impression, as well as the family’s and the adolescent’s willingness to embrace medication treatment. Furthermore, as for all medication treatments, both the patient and the family must have a sound understanding of the risks, benefits, and potential limitations of this particular modality as well as potential adverse effects, side effects, and alternatives to the proposed treatment. Finally, any consideration of medication treatment in the child population should include an understanding that randomized and controlled trials of specific medications are not always available and that many of the most commonly prescribed medications have not been approved by the U.S. Food and Drug Administration (FDA) for use in children. Nonetheless, the last 10 years have seen a dramatic increase in clinical investigation into the efficacy of pediatric psychopharmacology, and clinicians should understand that keeping abreast of these rapid developments is paramount in pursuing medication options for adolescents. In addition, the FDA black box warnings associated with the increased risk for suicide in antidepressant use in children and adolescents need to be taken into consideration when prescribing these agents. Families must be told of the FDA warnings and helped to understand the risks and benefits of these medications despite these concerns. The American Academy of Child and Adolescent Psychiatry website has instructions for how to discuss these complex issues with parents and teens. The risk posed by having these warnings includes the documented decline in the overall willingness to diagnose and treat psychiatric conditions in young people. Finally, as with pharmacologic treatment of adults, treatment of adolescents may be most efficacious when medications are combined with psychosocial therapies.

A 15-year-old boy was noted by his parents to have a difficult time leaving his computer in the morning. Although the problem was initially relatively well managed with gentle reminders that he was going to be late to school, over time, the boy showed a constant tendency to check his standing on an internet social networking site. This behavior resulted in frequent school tardiness, eventual school refusal, and increasingly poor hygiene. He was referred by his pediatrician for treatment with a psychiatrist, with whom he endorsed concerns that if he did not constantly update his website, his “social score” would decrease and that this decrease would translate into worsening social standing at school. He did not enjoy the constant maintenance of his website but felt powerless to stop and was ashamed of his behavior. He was started on fluoxetine and received a course of cognitive behavioral therapy (CBT) for obsessive-compulsive disorder. Over the next few months, his behavior gradually improved, with decreased anxiety and greater social engagement.

Group Therapy

Group therapy is a useful modality in the treatment of adolescents. It’s a form of psychotherapy with multiple patients led by one or more therapists. Typically groups are comprised of teens of similar ages and with different personalities. Group therapy uses this dynamic and interaction between peers to increase understanding of mental illness and improve social skills. Group therapy facilitates communication among peers in a way in which their feelings and experiences are normalized. As a result of the interaction and engagement, teens participating in group therapy are able to learn about themselves and the impact their behavior has on others and are able to receive feedback from peers. There are many different types of group therapies (e.g., psychodynamic, social skills, substance abuse, multifamily, parent support, etc.), and group therapy is often incorporated into the various therapy modalities such as CBT and DBT.

Psychosocial Interventions

Individual and group interventions include psychoanalytic therapy, psychodynamic therapy, CBT, and interpersonal therapy (IPT) approaches. All of these modalities, for both individual and group treatments, have been found to be effective for the treatment of a wide range of adolescent problems, including adolescent depression, anxiety disorders, substance abuse, and suicidality.

Psychoanalytic and Psychodynamic Therapies. Psychoanalytic and psychodynamic treatments seek to help the adolescent to come to terms with and thus better understand unconscious conflicts and maladaptive defense mechanisms. In addition, they may be useful in facilitating more adaptive coping skills. These goals are accomplished by a close relationship between therapist and patient and the use of unconscious processes such as transference, countertransference, and resistance. The best candidates for psychodynamic and psychoanalytic psychotherapy are adolescents who have adequate cognitive and language functions to use insight, self-reflection, and the ability to communicate effectively with the therapist. They also must have sufficient ego strength to tolerate clarification, confrontation, and interpretation. Because psychodynamic therapy may precipitate some regression and will, at times, be stressful in uncovering conflicts, the adolescent should be able to tolerate affect and have sufficient impulse control to thwart acting out. Commonly used psychodynamic therapies include classical psychoanalytic psychotherapy, supportive psychotherapy, and psychodynamic psychotherapies based on ego psychology, object relations theory, and self-psychology. As in the treatment of most adolescent disorders, psychodynamic psychotherapy is often conducted in conjunction with other modalities, such as pharmacotherapy, family therapy, and/or other psychosocial therapies. Although these forms of psychotherapeutic treatments are probably the most commonly applied interventions in the treatment of adolescents, the open-ended nature of these modalities has made randomized trials difficult. Many clinicians, however, feel strongly that skilled psychodynamic treatment is immensely beneficial to suffering adolescents.

A 14-year-old girl gymnast began increasing her daily exercise and restricting her diet. She had never achieved menses due to her low weight. She became fixated on the size of her thighs and belly and lost so much weight that her coach and pediatrician did not allow her to participate in athletics. She became increasingly terrified of getting fat, and she secretly exercised any chance she could. She was a perfectionist in academics, as well as in gymnastics. Because of excessive weight loss, she was hospitalized, and the diagnosis of anorexia nervosa, restricting type, was established without comorbid disorders. Following hospitalization with modest weight gain, an outpatient plan was initiated, including regular meetings with her pediatrician for monitoring of vital signs, weekly meetings with a nutritionist, weekly family therapy, and individual psychodynamic psychotherapy. In her psychotherapy, over the course of 2 years, she was able to understand that her perfectionism was a defense against low self-esteem, that she was very angry with her mother, who lived vicariously through her gymnastics and did not validate who she really was, and that she was, as an only child, the focus in her family, holding her parents' marriage together. Over time, she was able to understand the function of her anorexia as a way to avoid her maturation as a woman, as a dysfunctional means of developing some kind of identity, and as a passive-aggressive attack on her parents. With time, her weight increased, she was able to resume her athletics, and she developed close friends.

New research has focused on the process of mentalization and the theory of mind in the psychodynamic treatment of adolescents. Roughly speaking, mentalization can be considered the process by which individuals come to understand the thinking and emotional states of others. In a sense, the recursive thinking that becomes possible in adolescence is directly tied to this therapeutic skill. Thus, one attempts to build on the adolescent's capacity for empathy as their brain matures to the point at which empathy is more readily accessible.

Cognitive Behavioral Therapy. CBT has gained increasing popularity in recent years, and there are a number of studies suggesting its efficacy for problems such as adolescent depression, eating disorders, anxiety disorders, suicidality, and obsessive-compulsive disorder. CBT aims to help the adolescent to recognize cognitive distortions that lead to faulty interpretations and maladaptive behavior. CBT draws the connection between thoughts, feelings, behaviors, and body sensations and teaches that changing one can alter others. Techniques such as relaxation and contingency reinforcement are used to help the teenager to tolerate affectively charged environments. Treatment is often manual driven, with a specific number of sessions and homework assignments between appointments. This form of treatment does not focus on aspects of unconscious motivations and conflicts. Instead, it focuses on correcting distorted cognitions by modifying information processing.

A 13-year-old boy was noted by his parents to be extremely worried about contracting germs from his surroundings. He insisted on washing his hands more than 25 times a day and would often spend more than 1 hour a day making sure that his body and hands were clean and “free of contamination.” After evaluation, the diagnosis of obsessive-compulsive disorder was made, and the boy and his parents opted for CBT. The therapist started by helping the boy to rank his fears on a scale of 1 to 10, with 10 assigned to objects that he felt were most germ laden and 1 assigned to those objects that were of little or no concern. The boy was then taught both relaxation techniques and how to engage in thought interruption, and he used these techniques in a stepwise fashion to overcome first the more benign elements on his list, then the most severe. Eventually, he was able to modify his behavior by combining his understanding of what worried him most with his capacity to master and eventually overcome his fears.

Behavioral Therapy. Behavioral therapy has been used for years in the treatment of adolescent anxiety disorders, such as simple phobias, panic disorder, separation anxiety, and school refusal, as well as disorders of conduct and oppositionality. This modality is based on principles of operant conditioning and reinforcement and often uses important techniques such as systematic desensitization and contingency management. Behavioral therapy differs from CBT by focusing on behavior modification and not on issues such as faulty cognitions and distortions.

Interpersonal Therapy. IPT was first developed in 1984 by Gerald Klerman and colleagues in an effort to provide a short, psychodynamically informed treatment for depressed adults. IPT is based on the central premise that depression occurs in the setting of interpersonal conflict and that therapy should focus in a time-limited fashion on specific interpersonal issues as a means of ameliorating depressed mood. Like CBT, IPT is manual driven and often includes homework between sessions. Although it has not been studied in the treatment of adolescents as much as CBT, IPT has been shown to be effective in the treatment of nonsuicidal, nonpsychotic adolescent patients.

Dialectical Behavioral Therapy. Dialectical behavioral therapy (DBT) was developed by Marcia Linehan, given the realization that patients with affective dysregulation, suicidal, and other self-destructive behavior did not respond adequately to CBT or psychodynamic psychotherapy. Her model involves replacing ineffective, maladaptive behavior in response to internal or external stimuli with skillful, adaptive behavior. In DBT, maladaptive responses are viewed as learned responses to painful affect. DBT for adolescents was developed by Drs. Alec Miller, Jill Rathus, and Marcia Linehan. It involves learning emotional regulation skills, developing strategies for distress tolerance, enhancing interpersonal communication, and walking the middle path, and it combines these modalities with mindfulness techniques. DBT teaches that two opposites can be true at the same time, creating a new way of viewing the situation.

Typically, DBT involves the utilization of an individual therapist and participation in a multifamily skills-based group, requiring that parents meet in groups, sometimes with their adolescents. In the groups, parents will learn the skills to increase desirable behaviors and decrease undesirable behaviors of their teens, improve the quality of their relationships with their teens, effectively problem solve and negotiate, and manage unhelpful thoughts and emotions in response to teens.

One aspect of this form of therapy is the availability of parent skills coaching and adolescent skills coaching calls to help the adolescent “in the moment” to learn to utilize skills for mastery in real time and to help the parent support their teen using skills at the moment they are needed.

DBT has been shown to be effective for adolescents with borderline personality disorder, suicidal ideation and comorbid depression, bipolar disorder, eating disorders, and adolescent with impulsive and aggressive disorders.

Mindful Meditation. Mindful meditation has become increasingly used as a treatment for a number of psychiatric and medical conditions in adolescents. While more research is needed, there is evidence for its efficacy in treating anxiety, depression, and pain syndromes and its use in modulating affective dysregulation, impulsivity, and aggression. It is often a standard component of DBT for adolescents. Research has demonstrated that meditative techniques are helpful for general well-being and stress management.

Grouping psychosocial interventions into categories such as psychodynamic therapy, CBT, IPT, and DBT formats potentially creates the false impression that practitioners practice distinct forms of each modality. Although most clinicians are most comfortable focusing on one or two of these therapeutic formats, almost all who treat children and adolescents use particular aspects of multiple treatments. Psychodynamic therapies can be aided by contingency management, and effective CBT is made possible with a strong and, in some cases, unconsciously motivated relationship. Indeed, Ablon and colleagues showed that the best indicator of success in all forms of treatment is less related to the therapeutic modality use than to universally important features, such as the trust that the patient places in the clinician and the relationship that the clinician has with the patient.

Environmental and Systemic Treatments

In addition to psychotherapeutic and biologic interventions, the treatment of adolescents often includes therapeutic involvement in the adolescent’s environment. *Family therapies*, for example, aim to help the adolescent to negotiate their changing role in the family dynamic. Because adolescence is characterized by a movement away from the family of origin, many families become potentially distressed and volatile as the adolescent asserts their opinions within the family environment. Family interventions help to clarify family responsibilities and responses as these changes take place.

Similarly, *school interventions* seek to help an adolescent's teachers and school to better understand the specific dynamics between an adolescent's behavior and the school's response to that behavior. School consultants can make huge changes in a teenager's ability to continue along the social and academic trajectory by helping teachers to better understand the adolescent's behavior, by recommending and, in some cases, providing outpatient evaluation, and by suggesting behavioral techniques within the classroom. Sometimes relatively small but psychologically informed interventions such as simply changing classrooms can be immensely beneficial for a troubled teenager.

Finally, there is a growing appreciation of the importance of bringing interventions to the environments in which adolescents are most troubled. *Multisystemic therapy (MST)* has gained increasing popularity for the treatment of conduct disorders among adolescent populations. MST is based on the premise that conduct-disordered teens need interventions in schools, at home, in the office, and in the neighborhood. These interventions must be integrated and coordinated, and current studies document the efficacy of this multifaceted approach in reducing adolescent transgressions of conduct.

All forms of adolescent treatment are complicated by issues particular to adolescent development. Because much of the adolescent patient population is younger than 18 years old, parents must legally and ethically provide informed consent before allowing their children to be treated either with medications or psychotherapy. However, given the emphasis that healthy adolescents place on separation and autonomy, it is clinically paramount that this population understands the risks and benefits of psychiatric interventions and be willing participants in their treatment. Only in extremely rare circumstances, such as severe psychotic states or in the presence of dangerous suicidality, is it clinically useful to prescribe psychiatric medications and other treatments against an adolescent's will. In addition, adolescents may feel differently about their treatment when they are given the opportunity to discuss their feelings without their parents present. For example, side effects such as sexual dysfunction characteristic of many psychotropic agents need to be discussed with appropriate patients in the absence of their parents. In general, clinicians need to treat adolescents as much as possible like the adults they are striving to become while at the same time keeping in mind all of the legal and ethical considerations that are paramount in the treatment of any underage patient.

Hospital and Residential Treatment

Most adolescents can be treated in the outpatient setting. However, at times, inpatient, residential, or day hospital care might be required. In recent years, largely due to third-party payers and the advent of managed care, access to these more restrictive environments has been quite limited. The primary indication for inpatient settings is typically danger to self or others. Adolescents might be admitted to short-term residential or day hospital programs, many of which are in unlocked settings if there is an acute deterioration in behavior with an inability to function in the home and school environments. The most common diagnostic categories requiring hospitalization include severe depression, suicidal behavior, bipolar disorder, obsessive-compulsive disorder, psychosis, eating disorders, substance abuse, post-traumatic stress disorder, and conduct disorder. Inpatient, residential, and day hospitalizations offer the adolescent a therapeutic milieu for observation and multimodal treatment. The adolescent and family may use these services for medication management, behavioral containment, and the acquisition of better problem-solving skills. In many health care systems, a continuum of services has been established, with initial inpatient care leading to step-down units, such as unlocked residential or day hospital programs that offer a less restrictive environment and gradual re-entry into the community. A primary function of hospital-based care is to develop a comprehensive after-care program with coordination between outpatient services and utilization of school-based programs and other community services.

RESILIENCE AND WELL-BEING

Resilience can be defined as the process of adapting well and maintaining personal and social stability in the face of adversity, such as trauma, tragedy, threats, family and relationship problems, serious health problems, financial stressors, or other significant sources of stress. While there are many components to resiliency, the key elements are connection, awareness, wellness, healthy thinking, and meaning.

Resilience, while it may appear to come naturally for some, resilience is learned and must be cultivated and mastered. It is a set of skills that allow flexibility, innovation, and protection. It's an important skill to master as without it, one is easily overwhelmed and distressed and resorts to maladaptive coping skills and strategies.

Well-being integrates mental and physical health and is a dimension of wellness, which has several dimensions, including emotional well-being (coping effectively with life and creating satisfying relationships) and physical well-being (recognizing the need for physical activity, healthy foods, and sleep). According to the CDC, well-being is considered to be the presence of positive emotions and moods, the absence of negative emotions, satisfaction with life, fulfillment, and positive functioning. Higher levels of well-being are associated with decreased risk of disease, illness, and injury; better immune functioning; speedier recovery; and increased longevity. Adolescent well-being depends on both the adolescent and their environment, with family, school, and romantic relationships being decisive environmental factors.

EFFECTIVE ADOLESCENT THERAPY

The treatment of adolescent patients is quite different from the treatment of latency-age or adult patients. Adolescents can seem both extremely childish and surprisingly mature. In addition, the same adolescent population may undergo rapid and unnerving developmental changes in a relatively short amount of time. These issues substantially complicate the clinician's task. Forming an effective alliance with an adolescent patient requires special attention to already important clinical considerations such as confidentiality and safety. For example, the patient who admits to using illegal drugs may not want their parents to know but may acknowledge a desire for help in understanding this growing drug problem. Although patient safety is always the most important clinical consideration, clinicians must view each situation on a case-by-case basis. Often, adolescents come to their treaters for guidance on issues that they are not able to discuss at length with their parents. A friendly, honest, flexible, and available clinician is absolutely essential for the adolescent to feel comfortable in coming forward with important and pressing issues.

Clinicians also need to be curious about their adolescent patients' worlds. Adolescents will express themselves wonderfully in displacement, finding voices in the music they listen to, in the television they watch, and in the stories and movies they embrace. Although it is neither reasonable nor possible for a clinician to stay abreast of the pantheon of popular culture, inviting adolescent patients to discuss aspects of their culture that excite them is a valuable means of building trust and of gaining insight.

FUTURE DIRECTIONS

The psychiatric treatment of adolescents is, by definition, a complex endeavor. Because adolescent behavior manifests as a function of biologic, developmental, psychological, and cultural phenomena, the clinician must be prepared to take into account all aspects of the adolescent's world to effect change. We cannot underestimate the impact of social events and political movements on adolescent development, such as climate change, gun control, sexual assault and harassment, discrimination, racism, injustice, and social justice reform. We need to consider the long-term impacts of COVID-19 as well as the impact of environmental and social events on adolescent development.

Furthermore, no clear hierarchy exists among the various determinants of adolescent difficulties. The same behavioral problems among different adolescents can have strikingly different determinants. One irritable teen might have a biologic predisposition to depression, whereas another might suffer primarily from a chaotic home life. A tendency to address only one aspect of an adolescent's presentation risks oversimplifying the extent to which adolescents suffer and misses a valuable opportunity to be maximally effective. Indeed, it is this very complexity that makes the treatment of adolescents so rewarding and fulfilling.

FURTHER READINGS

Anderson M, Jiang J. Teens, social media and technology. 2018. Retrieved September 20, 2018 from the Pew Research Center website. pewinternet.org/2018/05/31/teens-social-media-technology-2018

Bruce LD, Wu JS, Lustig SL, Russell DW, Nemecek DA. Loneliness in the United States: a 2018 national panel survey of demographic, structural, cognitive, and behavioral characteristics. *Am J Health Promot*. 2019;33(8):1123–1133.

Centers for Disease Control and Prevention. [2021] Youth Risk Behavior Survey Data. Available at: www.cdc.gov/yrbs.

Cuffe S, Hall W. Suicide and SSRI medications in children and adolescents: an update. American Academy of Child & Adolescent Psychiatry. 2007.

http://www.aacap.org/AACAP/Medical_Students_and_Residents/Mentorship_Matters/DevelopMentor/Suicide_and_SSRI_Medications_in_Children_and_Adolescents_An_Update.aspx

Danneel S, Nelemans S, Spithoven A, et al. Internalizing problems in adolescence: linking loneliness, social anxiety symptoms, and depressive symptoms over time. *J Abnorm Child Psychol*. 2019;47:1691–1705.

Erikson EH. *Identity and the Life Cycle*. W. W. Norton; 1979.

Fair DA, Cohen AL, Dosenbach NUF, et al. The maturing architect of the brain's default network. *Proc Natl Acad Sci U S A*. 2008;105:4028–4032.

Freud A. *Normality and Pathology in Childhood: Assessments of Development*. International University Press; 1965.

Goel V, Dolan RJ. Reciprocal neural response within lateral and ventral medial prefrontal cortex during hot and cold reasoning. *Neuroimage*. 2003;20;2314–2321.

Gogtay N, Giedd JN, Lusk L, et al. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci U S A*. 2004;101:8174–8179.

Gorrindo T, Fishel A, Beresin EV. Understanding technology use throughout development: what Erik Erikson would say about toddler tweets and Facebook friends. *Focus*. 2012;10(3):282–292.

Groves S, Backer HS, van den Bosch W, Miller A. Dialectical behavior therapy with adolescents. *Child Adolesc Ment Health*. 2012;17(2):65–75.

Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: a meta-analytic review. *J Consult Clin Psychol*. 2010;78(2):169–183.

Hunt M, Marx R, Lipson C, Young J. No More FOMO: Limiting social media decreases loneliness and depression. *J Soc Clin Psychol*. 2018;37:10:751–768.

Ivey-Stephenson AZ, Demissie Z, Crosby AE, et al. Suicidal ideation and behaviors among high school students - youth risk behavior survey, United States, 2019. *MMWR Suppl*. 2020;69(1):47–55.

Jellinek MS. Impact of media. *J Am Acad Child Adolesc Psychiatry*. 2001;40:1124.

Klerman GL, Weissman MM, Rounsaville BJ, Chevron ES. *Interpersonal Psychotherapy of Depression*. Basic Books; 1984.

Ko CH, Yen JY, Yen CF, Lin HC, Yang MJ. Factors predictive for incidence and remission of internet addiction in young adolescents: a prospective study. *Cyberpsychol Behav*. 2007;10(4):545–551.

*Kutner L, Olson C. *Grand Theft Childhood: The Surprising Truth about Violent Video Games and What Parents Can Do*. Simon and Schuster; 2008.

Madigan S, Ly A, Rash CL, Van Ouytsel J, Temple JR. Prevalence of multiple forms of sexting behavior among youth: a systematic review and meta-analysis. *JAMA Pediatr*. 2018;172(4):327–335.

Markey PM, Ferguson CJ. Internet gaming addiction: disorder or moral panic? *Am J Psychiatry*. 2017;174(3):195–196.

Mesoudi A. The cultural dynamics of copycat suicide. *PLoS One*. 2009;4(9):e7252.

National Center for Statistics and Analysis. *Teen Distracted Driver Data*. (Report No. DOT HS 812 931). National Highway Traffic Safety Administration; 2021. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812931

*O'Malley F. Contemporary issues in the psychiatric residential treatment of disturbed adolescents. *Child Adolesc Psychiatr Clin N Am*. 2004;13:255–266.

Pataki C, Bostic JQ, Schlozman S. The functional assessment of media in child and adolescent psychiatric treatment. *Child Adolesc Psychiatr Clin N Am*. 2005;14(3):555–570, x.

Patchin JW, Hinduja S. The nature and extent of sexting among a national sample of middle and high school students in the U.S. *Arch Sex Behav*. 2019;48(8):2333–2343.

Reinecke MA, Ryan NE, DuBois DL. Cognitive-behavioral therapy of depression and depressive symptoms during adolescence: a review and meta-analysis. *J Am Acad Child Adolesc Psychiatry*. 1998;37:26–34.

Stubbe DE, Weiss G. Psychosocial interventions. Individual psychotherapy with the child, and family interventions. *Child Adolesc Psychiatr Clin N Am*. 2000;9:663–670, ix.

Sudak HS, Sudak DM The media and suicide. *Acad Psychiatry*. 2005;29(5):495–499. https://doi.org/10.1176/appi.ap.29.5.495

*Wiseman CV, Sunday SR, Becker AE. Impact of the media on adolescent body image. *Child Adolesc Psychiatr Clin N Am*. 2005;14(3):453–471, viii–ix.

World Health Organization. *Preventing Suicide: A Resource for Media Professionals.* Update 2017. World Health Organization; 2017. (WHO/MSD/MER/17.5). Licence: CC BY-NC-SA 3.0 IGO.

Yang AC, Tsai SJ, Yang CH, et al. Suicide and media reporting: a longitudinal and spatial analysis. *Soc Psychiatry Psychiatr Epidemiol*. 2013;48;427–435.

54

Child Psychiatry: Special Areas of Interest

54.1 Adoption and Foster Care

Dale Aaron Peeples, M.D., and Sandra B. Sexson, M.D.

Adoption and foster care are social institutions that provide substitute care for children and adolescents who are unable to be cared for by their biologic parents in their own homes. Foster care represents a temporary solution to out-of-home care, whereas adoption is aimed at a permanent assumption of parental responsibilities by parents other than the biologic parent(s). Recent data suggest that approximately 424,000 children are in formal foster care in the United States in a given year, with more than 673,000 being served by child welfare systems each year and close to 251,000 entering the system, and slightly fewer exiting the system annually. All of these numbers reflect a slight increase followed by a plateau over the past decade. Informal fostering also occurs outside of the public system, likely doubling the number of children and adolescents affected. Statistics for adoption are more opaque because adoptions are only formally tracked in the public child welfare system, not in the private system. It is estimated that approximately 2.5% to 3.5% of the U.S. population is adopted. Child welfare data suggest that more than 66,000 children are adopted annually from the child welfare system. Almost half of the adoptions occur through various private arrangements (stepparents, nonbiologic parents who conceived via reproductive assistance, private agencies, or Indigenous Tribes). Public adoptions have grown over the past decade, while international adoptions have seen a significant decrease due to changes in international relationships and shifts in source countries for adoptees. Child welfare statistics show a growing number of children in foster care awaiting adoption over the past decade with most recent estimates at 122,000.

The staggering statistics regarding foster care and adoption demand that psychiatrists and other mental health professionals address the complex needs of this vulnerable population of children and adolescents in the United States. Children in foster care frequently experience multiple placements, and many of those who are reunified with their families of origin are subsequently placed into care again. Children who are adopted are less likely to have another disruption in placement. Currently, parental neglect and the inability to provide adequate care are the most common reasons that children are removed from their homes. Fifty years ago, causes for separation typically revolved around incapacitating parental illness or death or unstable financial situations. Today, the reasons for parental incapacitation are heavily influenced by parental alcohol and drug abuse and all the iterations of poverty. Other, less common, reasons include physical abuse, abandonment, and sexual abuse. Although many children show resilience in the face of all of these stresses, the risk for mental and physical health problems in children in foster care and, to a lesser and somewhat more controversial extent, children in adoptive settings is much greater than that of the general population. Although it appears that many children enter placement situations with vulnerabilities in these areas, the consensus among professionals is that multiple disruptions in placements over time add to this vulnerability, increasing the risk for serious and prolonged mental health problems. This section reviews the history of adoption and foster care in the United States; identifies trends in the management of out-of-home placement, both foster and adoption, over recent years; and reviews the limited literature on outcomes in an effort to inform the practicing psychiatrist on how best to approach the care of this very vulnerable population.

FOSTER CARE

History of Foster Care

History is replete with examples of out-of-home care for children. For centuries, parents have abandoned children because of unwillingness or inability to care for them. History suggests that over the centuries, religious entities took over this responsibility, typically forcing children into child labor situations. Children from cities, seen as dangerous, were sent to farm families as a source of cheap labor, not only leading to mistreatment but also removing them from their families of origin and community systems permanently because of the distances involved. Group care for these children was developed in the 19th century when orphanages were found to be less costly for the state. Pejorative attitudes continued to paint them as coming from bad families with high risk for bad outcomes. Family reunification remained a possibility although, at the beginning of the 19th century, it required mothers to pay a fine and reimburse the state for costs incurred by the state for the child's care. By the early 20th century, such temporary out-of-home placement was common, leading to the first White House Conference on Care of Dependent Children, which formally endorsed foster care over the institutional model.

Although most of the law on foster care comes at the state level, federal law sets the tone and direction for the states. In a chronologic review of federal legislation on child welfare, several themes emerge, including funding programs to improve child safety and well-being, permanency of placement, and improving outcomes of youth in the child welfare system. The first federal funding of children's services came about with Title IV of the Social Security Act of 1935, which helped set national standards for child welfare. It provided funds to help support children in single-parent homes but excluded minority families until the Civil Rights Movement. The Child Abuse Prevention and Treatment Act (passed in 1974 and updated in 2010) provides funds to states to prevent, assess, investigate, prosecute, and treat child maltreatment, as well as providing funds for both data collection and child welfare research. Focus on the emotional and developmental needs of children increased with the Adoption Assistance and Child Welfare Act of 1980. This act was aimed at preventing unnecessary separation of children from their families by providing more services for the prevention of removal, as well as efforts toward reunification. It also sought to address the problem of the unstable and continuous cycle of children being in multiple foster homes over extended periods of time. It established mandated 6-month reviews of each child in foster care. Permanency planning was emphasized, with reunification as the primary goal and adoption as the alternative. Over the years, this act was expanded to provide adoption support for special needs children. Under Title XIX of the Social Security Act, foster care children generally receive health care through Medicaid. The Multiethnic Placement Act of 1994 also took steps to improve permanency, particularly for minority children. The law increased efforts to recruit a diverse representation of foster and adoptive families to reflect the backgrounds of those in the child welfare system. It also eliminated discrimination based on race, color, or national origin.

Many children who were removed from their homes were eventually returned to families of origin (up to 75% in earlier studies, with around 57% as recent as 2016). A significant number of those reunited subsequently return to foster placement. The Adoption and Safe Families Act of 1997 sought to provide safeguards to protect children from neglect and abuse while promoting timely permanency plans and holding the child welfare system more accountable. Concerns that prolonged efforts toward reunification had been unsuccessful led to more stringent guidelines regarding permanency planning. States were required to move toward freeing a child for adoption when a child had been in foster care for 15 of the prior 22 months, with permanency hearings within 12 months after the child went into foster care. Kinship care in the form of permanent placement with a family member became a more popular option in lieu of reunification. The limitations on the aggregate time in foster care also discouraged the prior practice of moving children from foster care back to the family of origin and then back into foster care again over and over, delaying decisions on other, more permanent options.

More recently child welfare legislation has attempted to address those foster children who “age out” of the system, that is, those who remain in the foster care system until transition into adulthood. Initial steps began in 1986 with the addition of an amendment to Title IV-E of the Social Security Act—the Independent Living Initiative. This provided states with funds for preparation for independent living as youth aged out of foster care. This was replaced in 1999 by the Foster Care Independence Act which increased funding to states to help with transition to independent living and provided the option for states to extend Medicaid coverage to the age of 20. All youth aging out of foster care are eligible for Medicaid coverage until the age of 26 because of the Affordable Care Act of 2010. In addition to targeting access to health care, the federal government has also recognized the need of coordinating other services as children age out of foster care. The Fostering Connections to Success and Increasing Adoptions Act of 2008 requires the development of a transition plan at least 90 days prior to a youth aging out of foster care. Ideally, this will help address housing issues, health insurance, education, mentoring, and employment services, as young adults exiting foster care have little or no “family-type” support. Each year, more than 20,000 adolescents “age out” of foster care with limited connections or support systems. In 2018, the Family First Prevention Services Act was signed into law, and continued a trend of working toward maintaining families, and minimizing restrictive placements. It allowed states to use Title IV funds to offer treatment services to the families of youth at risk of entering foster care and continued to develop kinship care and reunification programs. Standards for residential treatment programs servicing foster children were increased, independent living services were extended to the age of 23, and educational vouchers to the age of 26.

Who Is in Foster Care?

In the United States, there are currently more than 423,000 children in foster care. Although child abuse and neglect occur across all racial and ethnic groups, as a percentage, children of color are over-represented in the foster care system compared to the general population. Multiracial children account for 8% of the foster care population while making up 3% of the population. Black and Indigenous children have a relative risk twice that of the general population. Not only are there more Black and Indigenous children in foster care proportionately, but those in foster care are more likely to remain there for longer periods of time and are less likely to experience permanent reunification or adoption.

There are some positive trends seen over the past decade regarding ethnicity and foster care. Although Black children remain over-represented in the foster care system, they have experienced the greatest reduction as a group of children receiving services, by 25%. Otherwise, the total number of children by ethnicity entering foster care has remained relatively stable. The one exception to this is multiracial children, which increased by a third during that time.

Another group of over-represented youth in foster care is sexual and gender minorities. Often, issues around sexuality and gender identity fuel conflict in the home, resulting in high numbers of runaway and homeless youth who identify as a sexual or gender minority. Overall, they are more likely to experience placement instability, group home placement, and age out of the foster care system. Qualitative studies suggest that these youth frequently experience discrimination in the foster system with care providers rejecting their identity, and caseworkers often assuming reunification may be impossible due to the biologic family’s initial rejection of the youth.

Over the past two decades, the number of unaccompanied minors crossing into the United States from the southern border has dramatically grown, with unstable circumstances in Central America, and U.S. immigration policy both influencing this change. Historically, about 6,700 unaccompanied children entered the United States annually, but over the past decade, those numbers climbed to over 72,000 in 2019. This drastic surge strained resources available to these children, leading to extended delays in detainment facilities. In 2018, this problem was exacerbated when a “zero tolerance” immigration policy temporarily went into effect, separated children from their parents, and reclassified them as unaccompanied minors. Most of these children are males over the age of 14, primarily fleeing Guatemala, El Salvador, and Honduras. They often face desperate circumstances such as violent criminal gangs, poverty, and natural disasters. At a bare minimum, unaccompanied refugee minors experience the stress of adjusting to a new culture and separation from family. In addition, many have been exposed to war, rape, torture, and severe deprivation prompting their immigration to the United States. Subsequently, post-traumatic stress disorder (PTSD), depression, and anxiety are common issues among refugee children. In 2002, Congress transferred the care of unaccompanied children from the Immigration and Naturalization service to the Department of Health and Human Services (HSS) to move away from an adult-oriented detention model to a child welfare-based model. Unaccompanied children are individuals under the age of 18 who have no lawful immigration status and who are without a parent or legal guardian. Initially, these youth are apprehended by Immigration and Customs Enforcement. Those that have an asylum claim are managed by the Office of Refugee Resettlement in HSS, who then assist in arranging placement. Temporary shelters are located in 23 states and house youth while they await a more permanent placement, either sponsorship by a relative, community foster care, or a group home. The average length of stay in a shelter grew from 34 days in 2016 to 93 days in 2018. Benefits for these children as they enter foster care are the same as those for children born in the United States. State child welfare systems monitor their care while in the child welfare system. Additionally, foster parents receive training in cultural sensitivity to help with the children's adjustment to life in a new country.

There are 574 federally recognized Indigenous tribes with a varying capacity for management of their own child welfare programs. Consequently, some Indigenous children are managed by state protective services. In the state system, they are over-represented relative to their portion of the population and are disproportionally placed into group home and institutional settings. Estimates about the number of children served by tribal organizations are unavailable, and therefore not included in national statistics. In 1978, the Indian Child Welfare Act was passed which reaffirmed and empowered Indigenous tribes to manage child welfare issues. It was put forth in response to the high rates of Indigenous children in foster care, and the fact that they were often placed in boarding schools and group homes away from their reservation, community, language, and culture. Specific plans for services vary from tribe to tribe. In addition to the challenge of operating many different models, services to children on reservations are also challenged by the rural nature of the setting, limited funding, and frequent turnover in local government, which can impact policy. Federal services are available to help Indigenous communities implement culturally sensitive practice model improvement projects.

Children of all ages are in foster care. Estimates in 2019 suggested that 7% of those in care are less than 1 year of age, 35% are between 1 and 5 years of age, 23% are between 6 and 10 years of age, 21% are between 11 and 15 years of age, 11% are between 16 and 18 years of age, and close to 2% are older than 19 years of age. Many of the older children entered foster care younger than 1 year of age. Infants stay in care longer and are less likely to be reunited with their parents. Males continue to represent slightly over half of children in foster care at 52%.

As a group, children entering foster care comprise a high-risk population for both physical and mental health problems. Studies suggest that one-third of children in foster care have at least one chronic medical problem or disability and over 80% will have mental health problems during their lifetime. Estimates suggest that as many as 80% of children in foster care were exposed to drugs or alcohol in utero and 40% had low–birth-weight either from prematurity or in-utero growth failure; both factors that increase the risk for many medical problems and developmental and mental health problems. It is not surprising that studies suggest that from 50% to 90% (compared to 10% to 20% in the general population and 35% or so in children living below the poverty level) of children who are placed in foster care have significant mental health problems necessitating interventions. In addition, children involved in the child welfare system are at much higher risk for mental health problems than their socioeconomically matched counterparts. Of interest, those entering the foster care system are more likely to get mental health care than those who remain in their homes, regardless of clinical need, leading some to suggest that foster care is a gateway to mental health services for many children. Significant nonclinical predictors of the use of mental health services for children in foster care include older age and being White. In addition, children who have been sexually abused are more likely to access mental health services, as are children living in nonrelative foster care placements versus those in kinship care. Those placed for reasons of physical neglect are less likely to receive mental health services.

There is a growing body of literature illustrating the physiologic changes children in foster care may experience due to early exposure to abuse and neglect. Neurodevelopmental impairment also results from early neglect. In addition to general growth delays, head circumference is decreased in children coming from institutional placements. The Bucharest Early Intervention Project (BEIP) is a longitudinal randomized-controlled trial following children exposed to institutional care and comparing their growth and development to those who were able to find placement in foster care and a community control sample. In this study, when children were followed up at multiple points (ages 5, 8, 12, and 16), electroencephalography patterns consistent with attention deficit/hyperactivity disorder (ADHD) were present more often in institutionalized children than their peers, reflecting immature cortical development. The presence of consistent changes at follow-up illustrates the potential long-term impact brought on by early neglect. In other adoption studies, both magnetic resonance imaging and diffusion tensor imaging reveal decreased total brain volume in both white and gray matter. For children who found placement in foster care as opposed to an ongoing institutional setting, there was an increase in white matter volume suggesting some neural plasticity and the potential for some catch-up once living arrangements improve. With respect to specific brain regions, studies are mixed regarding both the amygdala and the hippocampus, with some finding increased volume and others reporting decreased volume.

Despite similar experiences in terms of neglect and deprivation, outcomes for children do vary, suggesting that some individuals might inherit resistances to these stressors, while others might be more vulnerable to negative environmental effects. Polymorphisms in genetic sequences may offer some explanation for these different outcomes. Variation in the serotonin reuptake transporter gene (5HTT) is associated with moderating the influence of early adverse experiences and the development of anxiety and depression. Studies in adopted children comparing groups adopted before and after the age of 6 months found that variation in the 5HTT allele did impact the development of emotional problems in children adopted after the age of 6 months. In the English and Romanian Adoptees study, variation in dopamine reuptake receptors (DAT/DRD4) was associated with ongoing ADHD symptoms at the ages of 11 and 15, suggesting that children with the high-risk haplotype did not outgrow ADHD symptoms as rapidly as their peers. In the BEIP study, allele variation in the dopamine reuptake catechol-O-methyltransferase gene resulted in significantly lower rate of depressive symptoms in children with one of the variations.

Reviews of the literature consistently demonstrate that minority children, particularly Black children, are less likely to receive services than their White counterparts, despite both having high behavior dysfunction scores on screening instruments. This disparity suggests that caregiver, child welfare, and court officials’ attitudes regarding mental health service needs may be culturally biased, influencing which children in need get mental health services. Psychotropic medication prescribing practice in the foster care population is another area of major concern, with an over-reliance on medication in place of psychosocial interventions. Given that the foster care population is very much at risk of mental health problems, it may not be surprising that this population receives psychotropic medications at rates of two to eight times that of their peers. The concern is whether evidence-based psychosocial interventions, which are often difficult to manage in unstable living environments or in homes with multiple foster children, are being employed when indicated. Certainly, foster children are at higher risk for prenatal exposure to substances, malnutrition, and family history of mental illness, all of which could necessitate a pharmacologic intervention. However, psychological factors such as grief, adjusting to a new environment, and attachment issues may manifest as internalizing and externalizing symptoms causing a psychiatrist to rush to medicate when a psychosocial intervention is in order. This is a particular risk when the history is frequently obtained from sources that only know a portion of the child's history.

Reasons Youth Enter Foster Care

The reasons that children and adolescents require placement outside their families of origin are complex, involving problems with the parents, child factors that contribute to failure of the nuclear family, and societal or environmental conditions. Biologic parents' problems may place their children at risk of neglect or abuse. Substance abuse, mental illness, intellectual disability, domestic violence, and lack of social support systems (low socioeconomic status and being a single parent) all negatively affect the parents' ability to provide adequate care for their children. The child may also bring factors, such as temperament, into the situation that increases the risk of parental failure. Children with difficult temperaments and those with the inability to self-regulate effectively make parenting more difficult. However, even the child with an easy temperament may be at increased risk of neglect in a chaotic and demanding family. Children with cognitive and physical disabilities also may be more difficult to parent. Children who have temper tantrums and oppositional behavior bring more stress to the already vulnerable parent–child interaction. The more needs a child has, the more difficult it is for parents to provide care that is adequate. Finally, adverse environmental conditions may contribute to the failure of capable parenting. Examples that increase stress in families include poverty, unemployment, decreased or absent social supports, and poor nutritional resources. Effective interventions to prevent the need for foster placement, as well as interventions aimed at providing substitute care, must address all of these components if they are to be successful. Too frequently, the child is moved to foster care, but many of the issues involved in the ongoing lack of fit for the child in the family are not aggressively addressed.

Not all children who encounter child protective services (CPS) end up in foster care. After a report of neglect or abuse is received, the caseworker determines if there is enough information to open a case. If there appears to be a safety risk, a case opens, and an investigation commences to determine if the allegations are “substantiated” or “unsubstantiated.” Depending on these findings, a number of options are available. A case could be closed, the family could accept voluntary services, or a referral could be made to the court. For those families that go to court and are separated, a permanency plan is developed focusing on the child’s eventual exit from the foster system. Sometimes a guardian ad litem is appointed by the court to represent the best interests of the child. If the termination of parental rights is considered, the court seeks clear and convincing evidence that the parent is unfit.

The prior description of CPS procedure focuses on formal steps with court supervision. However, families can accept voluntary recommendations which remove them from court oversight. Typically, this scenario comes about when a family voluntarily accepts to move a child to a kinship placement at the beginning of a CPS investigation. States are not required to track this type of arrangement, but estimations suggest this “hidden” or voluntary kinship foster system is similar in size to the formal foster care system. Criticisms of this approach are that the procedure is coercive toward families (with the threat that they will lose their child if they do not accept the safety plan), that the process lacks judicial oversight, that a clear reunification plan is often lacking, and that kinship care providers don’t receive funds and services they would qualify for if CPS went through the formal placement process. Benefits include minimal disruption to the child, flexible visitation with parents, and reduced demand for state resources. Increasingly, additional funding is allotted to support these types of arrangements, but it still remains less than a family would receive if the kinship placement went through the court.

Outcomes

Outcomes for children and adolescents placed in foster care are hard to assess, but The National Youth in Transition Database has monitored youth transitioning out of foster care at 17, 19, and 21. The survey looks at financial self-sufficiency, education, social support, homelessness, and high-risk behaviors (substance use, incarceration, pregnancy). By the age of 21, 86% have part-time or full-time employment. Sixty-nine percent have achieved a high school diploma or GED, with 7% having obtained further vocational training or an advanced degree, and 24% report they are still engaged in some form of education. At 17, 94% report having a supportive adult in their life, but by 21 this number drops to 87%. High-risk behaviors are elevated at all points of the survey. At age 17, 27% report a history of substance use treatment, and 11% report treatment in the past 2 years at age 21. Thirty-three percent note incarceration by the age of 17, with 20% having experienced incarceration in the past 2 years at 21. Childbirth numbers gradually increase with age, and 22% report having had a child by 21. Homelessness was experienced by 27% of 21-year-olds. In this data, four subgroups of young adults transitioning out of foster care can be identified. Forty-five percent appear resilient, having obtained educational skills and employment, while avoiding most negative outcomes. The second largest group is “parents on public assistance,” individuals who typically made gains in education, but had children early and subsequently have limited workforce involvement and are more likely to receive public assistance. Unsurprisingly, this group is three-quarters female. The “troubled and struggling” group comprises about 18%, and has high rates of homelessness, educational and employment deficits, and higher rates of substance use and incarceration. The last group is the incarcerated (15%) who have the highest rate of arrest and substance use, high rates of homelessness, and are over three-quarters male. It can be argued that youth who remain in foster care at 17 may not be representative of all youth who encountered the foster care system, such as those who quickly reunited with family. Still, these statistics show the significant challenges faced by these young adults and the need they have for continued social support during this transitional phase of life. The average young adult is not financially independent until the age of 26, these young adults could broadly benefit from ongoing mental health and substance use services, post-high school educational support, housing and job access, and childcare programs.

Outcomes during foster care look at issues related to the disruption of care, given that improving placement stability is perceived as a major goal in foster care placement. Over a 12-month period, as many as 50% of children in placement experience a disruption, resulting in movement to another home or a more restrictive environment. Child characteristics thought to increase the risk of placement disruption include older age, general health problems, a mental health diagnosis, family history of mental health issues, sexual and gender minority identification, and presence of behavior problems, especially those related to conduct disorders. Studies assessing the impact of gender are mixed, but recent reviews suggest females may now be at greater risk of disruption. Children removed for physical and sexual abuse are also more likely to have trouble with a placement compared to those removed for neglect. In addition, the experience of stress in the foster family in 6 months prior to the child’s placement has been associated with increased rates of disruption, while foster parents that are trained to deal with special needs are more likely to continue caring for challenging children. Emerging data that links the number of behavior problems reported by the foster parent per day with increased disruption of placement suggest the need to focus interventions on reducing the child's behavior problems and increasing the foster parents' ability to deal with them. Children in kinship placements seem less likely to be disruptive than those in nonkinship foster care. Reasons for this are unclear. Typically, there are fewer children in kinship placements, which may affect a family's ability to manage a given child's behavioral problems. In addition, attachments in kinship placements may be easier to foster. Disruptions appear to be less likely when foster parents are warm and child-centered and when foster children are open to fostering. There are well-established, evidence-based interventions focused primarily on teaching parents to intervene appropriately with the children in their care. However, few of these have been implemented with foster parents on a wide-scale basis. Policy and systems also play a role in placement stability for children. There are times when the state may mandate a move for a child to reunite siblings, switch to a long-term placement, or merely try to deal with the reality of funding issues. When children have a high level of turnover in caseworkers, they are at greater risk of changing placements. Subsidies for guardianship and adoption placement, as well as hiring graduate-level social work staff, are associated with improved outcomes in placement. Without a doubt, multiple disruptions lead to increased problems for the child and for the caregivers.

Since the 1980s, child welfare systems have focused more on reunification with the family of origin as a permanent placement option. However, over time, some leaders in child welfare have questioned whether this goal is always appropriate because many children who are reunited go back into foster care. Lower rates of reunification occur when the child is very young, an adolescent, or Black. When the child is removed because of neglect there is a lower likelihood for successful reunification. Family variables that decrease the rates of reunification include poverty and failure to visit when children are in foster care. Longer time in foster care, as well as kinship placement, also lowers the reunification rates. Still, over the years, child welfare has assumed that reunification is the best outcome for children in foster care based on the need for continued contact with the biologic family and the fact that extended periods in foster care usually lead to less stability. However, recent data has accumulated suggesting that reunification may not always be the best option. Reunited children in several studies have had more behavioral problems, engaged in more risky behaviors, and had poorer life-course outcomes. Better outcomes for those who were not reunited included higher scores on intelligence testing, increased general well-being, and fewer problems with repeated criminal arrests. Some of the major parental issues that lead to foster placement—specifically substance abuse—are not easily remedied in the parent, limiting the likelihood of successful reunification. Intensive interventions to support families with mental illness, substance abuse, homelessness, and severe poverty are not readily available in most settings. Reunification in the right situation is still a laudable goal, but child welfare agencies should openly confront the pros and cons in each situation in an attempt to make a permanent decision in a timely manner that is best for the individual child.

Since the 1990s, kinship care has become more popular as a formalized placement alternative to traditional foster care options. Federal laws such as the Personal Responsibility and Work Opportunity Reconciliation Act of 1996, the Adoption and Safe Families Act of 1997, and the Fostering Connections to Success and Increasing Adoptions Act of 2008 have given preference for child placement to kinship arrangements. Most states have options for reimbursing kinship care, although frequently at lower levels than nonkinship foster care. Theoretical advantages to kinship care include the preservation of the child’s sense of family and cultural identity; increased access to relatives other than the kinship provider; greater stability in placement; maintenance of religious, ethnic, and racial values; and, potentially, the existence of previously established relationships with the children that may facilitate attachment relationships. Kinship care provides a way for the child’s family to be preserved even when reunification with the family of origin is not possible. Children in kinship placements as compared to those in foster care also show greater mental health stability and fewer behavioral issues. However, there are numerous concerns about kinship care. As a group, kinship caregivers are more likely to be single, older, with more health concerns, less education, higher rate of unemployment, and from lower socioeconomic status. Frequently, kinship caregivers are grandmothers who parented the foster child's parent, which causes some concern among child welfare agencies. Child welfare workers report having more difficulty supervising kinship placements, citing that kinship caregivers frequently do not adhere to parental visitation restrictions. Some of these negative differences may be alleviated with more equitable resources and support services and increased supervision commensurate with that provided to nonkinship foster parents. Continued research on kinship care is necessary to provide child welfare agencies with reliable data on the long-term outcomes of children reared in either kinship or nonkinship foster placements to inform future decisions.

Additionally, it is being recognized that the majority of kinship arrangements are informal, meaning the court and the public child welfare system have not helped broker the arrangement between the birth parent and the caregiver. Approximately 1.76 million children are in this type of living arrangement, six times as many when compared to families that have formalized the agreement. Unfortunately, this creates obstacles for both the caregiver and the child. Without proof of legal rights, obtaining services ranging from health care to school enrollment may be impossible without assistance. Grandparents find themselves most commonly serving in the role of the informal kinship caregiver. Often they are reluctant to seek out legal action to terminate birthparents' legal rights. They retain hope that the social circumstances (often substance use) that led to the current living arrangement will correct itself, and the child will be able to live with the birth parent once more. Although this arrangement is often prompted out of compassion for the biologic parent, it limits a child's access to social services, and places a child at risk in terms of permanency of placement, as they can be removed from the home at the whim of the birth parent. Policies vary from state to state, but informal kinship caregivers have options beyond adoption that allow them to seek services for children. Unfortunately, these generally require some out-of-pocket expense for legal fees, which may be an obstacle to some families. Seeking guardianship will allow a caregiver to access services for a child, but in some states, it would also require the termination of parental rights. Legal custody, as opposed to guardianship, is appointed by the family court instead of probate court, but otherwise, quite similar in that it often requires the proposed custodian to provide proof of the birthparents' inadequacy, and in some instances, parental rights will be terminated. Some states offer a *de facto* custody option. This is also managed by the family court and requires that the care provider prove that they provided for the child's needs for a period of time (often 6 months to a year) without the involvement of the birth parent. Biologic parental rights are not terminated, and no concerns about the birthparents' inability to raise the child need to be discussed in court. However, once the ruling is made the birth parent is not allowed to remove the child from the de facto custodian without court approval.

Treatment Foster Care

Treatment foster care (TFC) is a well-researched, evidence-based approach that refers to interventions focusing on at-risk children who are unlikely to otherwise be managed in the community and would thus find themselves in an institutional placement such as a mental health facility, juvenile corrections, or a secure residential facility. These are children who have experienced significant trauma, abuse, and neglect which have begun to manifest as major mental illness or budding antisocial behavior. TFC interventions are similar to a number of older services, such as "specialized foster care," "wraparound" foster care, and multidimensional TFC. All try to help at-risk children by developing strong community links, therapeutic training for foster parents, providing treatment directly in the home, limiting the number of children in the home, providing crisis intervention services, and treating the foster parent as a professional as well as providing additional payment to the foster family. Children show an improvement in social skills, a reduction in problem behaviors, and an increase in placement permanency. Such interventions may help reduce antisocial behavior, and subsequent incarcerations for youth in foster care. Additionally, children receiving such services have also shown improvements in school attendance, homework, and employment.

Mental Health Problems

Children in foster care struggle with many psychological issues because the reasons that they enter foster care have almost universally potential negative effects on psychological development and emotional adjustment. The events leading to the need for placement are adverse, often involving abuse, neglect, exposure to violence, and inconsistent parenting. Despite the fact that the foster placement may be a safer environment, it carries with it the loss of familiar surroundings and familiar caretakers, caretakers to whom the child usually has some attachment despite the abusive or neglectful relationship. In addition, foster placements typically are not planned but occur abruptly with little opportunity for preparation. How children handle these major stresses depends on the reasons for placement, the age and/or developmental level of the child, the old and new home environments, and the support systems available for the transitions. Each separation emphasizes the rejection and feelings of abandonment, compromising the ability of the child to develop effective emotional attachments. Younger children are particularly vulnerable to attachment problems, which may affect their ability to relate effectively in the future. Their inability to trust their caregivers to provide them with consistent predictable care leads them to mistrust future relationships and exhibit behaviors of avoidance, anger, and/or oppositionality. These behaviors in turn make it hard for even the most nurturing of adults to continue to offer the child much-needed nurturance. Attachment to foster parents may also be complicated by complex and ambivalent feelings of being unfaithful to the biologic parents when children open up to a relationship with a new caregiver. Typically, children removed from their birth parents blame themselves, perceive themselves as unlovable, and feel guilty about the separation. When children cannot trust their caretakers, their psychological development is complicated by the use of primitive defense mechanisms, and they are at risk of developing pathologic narcissism, which may lead to an inability to express empathy. Sometimes children have experienced care in their families of origin or in deprived placements that are so pathologic that they may develop reactive attachment disorder or disinhibited social engagement disorder, which requires immediate and intensive intervention to avoid catastrophic outcomes.

Children who have experienced pathologic care during their younger years also have difficulties regulating their behaviors, leading to problems with modulation of their emotions, aggression, and even attention and activity levels. Self-regulation is learned through a protective process with a predictable caretaker, an experience many foster children lack. In addition to the pathologic care, there may be other risk factors for regulatory problems, including prematurity, low birth weight, and exposure to drugs and/or alcohol in utero. The world of the abused and/or neglected child is one of extreme unpredictability, which leads to high levels of anxiety and fear. They are unable to sleep, a problem that contributes even further to a pervasive inability to regulate behavior.

Not only have young children in foster care experienced loss of their primary caretakers, who, in many cases, were unpredictable and pathologic in their caregiving, but many have also been traumatized overtly by physical and sexual abuse, as well as exposure to violence. Some may meet full criteria for acute or PTSDs, but many will have subthreshold symptoms that need to be addressed. The stress systems in the brain may be activated in such a way as to keep the child in a chronic state of arousal, an issue that more recent research suggests may impair the child’s ability to manage stress in the future. Early and intensive interventions may remediate these problems at a neurobiologic level.

Children in foster care also have much higher risks for developmental disabilities than children who are not in foster care.

Risk factors that lead to foster care may affect development. These include such negative influences as direct effects of maltreatment (brain injury, spinal cord injury, burns), as well as prenatal exposure to toxins such as alcohol and other drugs. Parental incapacities may have genetic components that affect the child as well, including intellectual impairments and language delays. Developmental disabilities also increase the risk that children will go into foster care and stay in out-of-home care longer, that foster placement will be less stable, and that other problems either within the child or the families involved will be exacerbated.

What about the Foster Parents?

Foster parents are asked to provide a nurturing and accepting environment for children who have had numerous negative and unpredictable experiences with adults in their environment, who come to a “temporary” relationship with ambivalence and difficulties with trust and attachment, and who then frequently test in numerous ways the foster parents’ acceptance of them within their families. Foster families have to deal with frequent, occasionally less than helpful, intrusions into their lives by caseworkers and other supervisors of foster care while sometimes getting little if any support for this challenging and difficult task. Furthermore, they have to do this knowing that their attachment to the child is likely to be disrupted, often under situations that they do not support, such as sending a child back to parents who have a history of being abusive. Foster parents experience grief and loss as children enter and leave their care. They experience rejection and opposition from children under major stressful conditions. Although most placement agencies provide some education to prepare parents for the challenges of foster care, ongoing training specific to the individual child’s needs and support systems to help the foster parent deal with their sense of loss and grief while continuing to provide nurturing support may improve outcomes for both the foster families and the children placed in their care. Professionals working with foster children and foster parents must address these important issues to facilitate optimum outcomes. The development of rituals for entering and leaving families and careful documentation through scrapbooks and photo albums may provide continuity in the life of the child moving across placements, as well as a historical record for the foster family.

The Psychiatrist's Role

The child and adolescent psychiatrist, along with the entire team of mental health professionals, is in a unique position to advocate for and provide effective interventions for the highly vulnerable population of foster children in the United States. In 2015, the American Academy of Child and Adolescent Psychiatry issued an updated guide outlining best practices in caring for children in foster care. Recent trends have led to increased use of psychotropic medications for children in foster care, relative to other youth with Medicaid coverage. Due to a combination of legitimate concern for safety issues surrounding these medications and state budgetary concerns, increasing attention is paid to prescribing practices of physicians treating children in child welfare services. Subsequently, before considering a medication, it is imperative to conduct a comprehensive mental health assessment (including, whenever possible, reaching out to biologic parents for history), and to identify any medical or developmental issues. Immediate intensive interventions are needed that address the breadth of the problem—not just prescribing medications, but also offering intensive parent training to foster parents specific to the child they must parent and implementing other psychosocial treatments along with psychopharmacologic treatments that are evidence-based for the disorders identified. Unfortunately, evidence-based nonpharmacologic interventions are frequently unavailable or poorly reimbursed, including intensive parent training. Child and adolescent psychiatrists often struggle with little continuity of information about the child's functioning, therefore limiting appropriate decision making as well as few available resources for consistency in evidence-based psychosocial interventions.

When medication is indicated, educating all parties involved in the child's welfare is essential, including obtaining assent from the child. Estimates of the number of children in foster care using these medications are two to eight times that of age-matched peers. When it is determined that a child will benefit from a pharmacologic intervention, consider a conservative approach. Avoid starting multiple medications or making multiple changes at once. Follow-up with the patient in a timely manner when changes are made—ideally 2 weeks after an initial visit. If a medication does not appear beneficial, confirm that the dose and duration are appropriate, and then remove the medication by a gradual taper if it truly poses no benefit. Monitor for side effects, including checking vital signs and weight regularly. When possible, use FDA-approved options and discuss "off-label" use with all parties when such an approach is necessary. When polypharmacy is necessary, clearly explain the role of all medications to the child and the caregivers. Try to avoid using medications of the same therapeutic class together for extended periods, try to avoid using doses above the maximum suggested recommendation, and try to avoid the "as needed" use of medications. In preschool-aged children, be very judicious about medication use. The Texas Department of Family and Protective Services and The University of Texas at Austin College of Pharmacy developed a comprehensive guide for prescribing psychotropic medications to foster children, which clinicians can use for further guidance.

It is necessary to keep the best interests of the individual child always at the forefront in assessment and intervention strategies and advocate for the child for permanency planning that considers the child's attachments, predictability, nurturance, support, and safety. Insisting that foster parents and families are involved in the treatment is also mandatory, avoiding the all-to-frequent custom of having a caseworker (or even a "transport" person with no information) brings the child to the psychiatric appointment without the involvement of the caregiver. Custody assignments should foster the child's sense of belonging within a system and should avoid disruption of strong psychological ties unless physical or emotional well-being is compromised. Psychiatrists are uniquely qualified to evaluate these attachments and support systems and recommend and/or implement treatment interventions that are evidence-based. In addition, as care providers, psychiatrists must advocate for the availability of effective services, which, while never readily available, may become even more limited as an effect of the expanding influence of Medicaid and managed care. Almost all children in the foster care system have Medicaid coverage. The number of psychiatrists accepting Medicaid continues to drop, and presently, only about a third of practitioners are willing to take on new patients with this coverage. In the context of the national shortage of child and adolescent psychiatrists, this misfit between payer source and provider means many foster children do not have access to care.

Looking to the Future

Although progress in understanding the problems and needs of children whose families of origin cannot successfully care for them continues, the field of foster care faces unanswered questions. Maltreatment and neglect have established negative effects on the growing child. Multiple placements with failure to provide consistent care over time lead to negative outcomes. Nonetheless, many questions as to how best to intervene are often controversial and, more often, difficult to implement (even when evidence indicates an appropriate intervention) because of limited resources. Children and adolescents in foster care must have physical and mental health assessments and interventions beyond medication management from the very beginning of care. Foster parents need to be involved in these interventions. Agencies need to provide more education and individualized support to foster families who are providing care for these challenging youth. Further research is needed to assess the viability of family preservation, early determination of the need for out-of-home permanency planning, and the advantages and disadvantages of kinship care over nonkinship care. As services shift away from institutional and group home settings, interventions are needed for youth with severe behavioral issues and developmental needs. Implementing more organized support for adolescents who "age out" of foster care is a continuing challenge.

ADOPTION

History of Adoption

Adoption is a practice through which children whose parents are unable to provide for their care are supplied parents, often providing the adoptive parent(s) much-desired children. The practice is ancient and documented throughout history. Prior editions of this volume have summarized the history from ancient Babylonian times through today's practices. It was not until the 1850s that adoption was first recognized legally in the United States. Previously, informal adoptions often resulted in indentured servitude and child labor. In 1851, Massachusetts passed a law on adoption that integrated concepts that are still important. The law required judicial approval, biologic parent or guardian consent, and the determination that the adoptive family was able to care for the child. Under these new laws, children were accorded legal membership in the family, although sometimes without access to the inheritance. Children's needs became an established priority.

Historically, adoption records, when available, were not confidential. Typically, the birth certificate was not changed. In the early 20th century, social workers implemented the concept of *closed adoption*, establishing secrecy among the triad of the adoptee, birth family, and adoptive family. Social workers tried to match adoptees with adoptive parents based on race, religion, and physical and intellectual attributions. These efforts at secrecy were felt to protect the privacy of all involved, enhance permanency, and help the child avoid the social stigma of illegitimacy. Sometimes adoptees were not even told that they were adopted. Obviously, many children were then, and still are, considered unadoptable. The practice of secrecy is controversial. Historically, concerns for all parties in the adoption process were made to argue in favor of secrecy. First, among these was the idea that the adopted child might experience confusion about their identity development and who their “real parents” were. Second, the adoptive parents might feel that they did not have the right to act as the child’s “full parents.” Finally, there was concern that the birth mother might not be able to move on from the loss of giving her child up for adoption due to the ongoing contact, thus prolonging her grieving process. Initial studies went on to refute such concerns. Additionally, as the face of adoption shifted in the United States, openness became unavoidable in many cases. Children adopted later in life through foster care are obviously aware of the family of origin. Similarly, children cared for by extended kin are often encouraged to maintain a relationship with their biologic parents.

Open adoptions, in which records are open and adoptive children may actually have contact with biologic parents, have become more common. The degree of openness varies from family to family, ranging from a one-time meeting between the biologic mother and the adoptive parents to ongoing regular contact with the biologic family. Although research is limited, recent studies eliciting feedback from adoptive parents have suggested that contact between the family of origin and the parents undertaking the adoption results in a higher comfort level with the adoption process for the adopting family, as well as instilling greater empathy in them toward the birth family. Adoptive parents have indicated that they are able to control boundaries with birth families and generally note that they also feel they have a good level over control of communication. As this contact improved empathy toward the family of origin, it also may lead to more positive perceptions about the adopted child’s behavior. Having an agreed-upon and predictable structure for future contact is also seen as a positive to families. As contact continues, adoptive parents note fewer concerns that the biologic parents may seek reunification, as trust between the two parties grows. Contact also reduces fears that adopted children have about their biologic parents. They report less stress and worry about the safety and well-being of their family of origin. Similarly, it is felt that some connection and openness can help them to better deal with mental health issues that the biologic parent may be experiencing. This preliminary data are limited because the responses are almost totally from the adoptive parents, studies are small and nonstandardized, generally recruitment of subjects is through convenience samples, and most are focused on the adoption of White infants. Still, the data offer some newer insights into the management of openness in adoption.

With societal changes, such as increased use of birth control, legalization of abortion, and the decreased stigma of single parenthood, the latter third of the 20th century saw fewer infants available for adoption, and both older children in foster care and children from other countries became potential adoptees for families in the United States. The goal for adoption of older children shifted from historical beginnings, when the benefit was for service to the adoptive family, to the psychological benefit of the child. Children from other countries began to be adopted in the United States after World War II and again after the Korean War, raising the controversy over transracial or transcultural adoptions, one that is not resolved. Recently international adoptions have been in global decline due to shifts in cultural attitudes and regulatory processes, but the United States remains the largest destination country for international adoptions. Previously Russia and Guatemala were a source for many international adoptions, but now China, Columbia, India, South Korea, Bulgaria, and Haiti have the most children placed in the United States annually. In many cases, both with international adoptions and to a lesser extent with adoptions from foster care, the physical appearances of the child and the parent are obviously different, leading to the designation of “visible” adoptions.

Epidemiology of Adoption

Adoption is a well-established institution involving huge numbers of children and adults over the years. Although relevant statistics have not been routinely compiled, estimates suggest that between 2.5 and 3.5% of children in the United States are adopted, with more than 2% in nonrelative adoptive placements and another 1.5% in relative adoptions, including stepparents, who account for a significant number of “relative” adoptions. Foster care adoptions, have gradually increased with the push toward permanency planning, and now represent about 40% of adoptions. Approximately 120,000 to 130,000 children are adopted each year, but this number has varied widely, with only about 50,000 annually in the 1940s and highs of 175,000 in the late 1960s and early 1970s.

Children are adopted in a number of different ways. Infants may be relinquished by their biologic parents and adopted through private agencies or through attorneys, physicians, or clergy. These children are placed in adoptive homes either from the hospital or after a few weeks in “private” foster care. More and more of these adoptions are “open,” allowing initial contact with biologic parents, particularly mothers, and in some cases, ongoing contact between the adoptive and biologic parent(s). Although statistics on this group are not being collected, it is estimated that approximately 50,000 children per year are adopted in this manner. A second group comprises international adoption—the practice of adopting children from outside the United States. This group continued to increase over much of the last two decades, often averaging close to 20,000 adoptions annually, until recent declines due to geopolitics. In 2019, there were just under 3,000 international adoptions. These adoptions typically are arranged through agencies, and a total average of over $30,000 and take anywhere from 3 months to 3 years to complete depending on the source country. The average age of adoption has gradually increased, and it is now relatively rare to have an international adoption completed under 1 year of age. A little under half of the international adoptions are completed before a child is five. These children may have had less than optimal care in orphanages prior to adoption. They were removed from their countries of origin and had little or no information about their biologic families. Finally, children may be adopted from the foster care system. As many as 66,000 children or adolescents are adopted annually through the child welfare system. However, over a quarter of children in foster care are waiting on adoption. These children typically have been in temporary placements, sometimes for a number of years. These adoptees are more likely to be from a racial or ethnic minority, and more than half are 6 years of age or older and may have experienced abuse and neglect, as well as multiple foster care placements.

Legislation continues to head in a progressive direction, gradually removing some of the barriers to adoption and opening the door for nontraditional adoptions, including single-parent, sexual minority, and transracial possibilities. The 2015 Supreme Court ruling of Obergefell v. Hodges recognized the right of same-sex couples to marry. One factor necessitating this equal protection was the protection of family well-being, and subsequent rulings have extended this decision to protect the rights of same-sex couples to adopt. Some states have attempted to prevent sexual and gender minorities from adopting by contracting private religious organizations to assist with adoption placement, who then use decline services on the grounds of “religious freedom.” Gay and lesbian couples are more likely to adopt through the foster care system than heterosexual couples, and overall adoption rates are higher in states that have policies in place that are friendly to sexual minorities. Consequently, these families do a great service to many children remaining in foster care and are generally also more willing to adopt transracial children. As openly gay families having the ability to adopt children is a relatively new phenomenon, longitudinal studies are just beginning to provide some outcome data. In early childhood, there appears to be no difference in internalizing or externalizing behavior by parent and teacher reports in children raised in gay, lesbian, or heterosexual families. Further, there are no differences in gender role behavior in these children. Play preferences in young children appear less gender stereotyped in children reared by homosexual parents, particularly by lesbian mothers. Studies have also found that reduced parental stress is associated with better child adjustment. Consequently, state and federal policies that minimize discrimination against gay and lesbian parents can benefit their children as well. The topic of adoption by transgender parents is a relatively new issue, and there are no specific legal barriers to this group, but in reality, they often face discriminatory attitudes, and many lawyers advise disclosing gender status early to avoid last-minute disruptions to the adoption process.

Approximately, 40% of all adoptions are transracial or transcultural, meaning that the race, ethnic, or cultural origin of the child is different from that of one or both of the adoptive parents. This is most common in international adoptions, where 84% are transracial, but also seen in private adoptions (21%) and domestic adoptions (28%). Transracial adoptions in the United States initially involved only overseas adoptions, but in the 1960s, some transracial domestic adoptions began to occur. Parents were initially encouraged to de-emphasize the importance of racial differences to facilitate the child's integration into the new family. More recently, parents are encouraged to discuss race and focus on the positives of cultural diversity, as well as acknowledge to their children that they are living in a racialized world to help validate any experiences with microaggression or overt racism that may be directed at the child. Recently, many studies have begun to focus on two issues—how to promote the development of a racial/ethnic/cultural identity for a child and how to promote a child's awareness and ability to cope with racism. Contact with members of the child's racial community is helpful but infrequently occurs in transracial adoption. Having such connections may help children cope better with experiences of racism. Also of importance in the United States are the experiences of Black children adopted into White families compared to Black children adopted into Black families. Researchers have described more identity issues in these populations, with Black children reared in Black families identifying themselves as Black much more frequently than those reared in transracial settings. Findings are complicated because multiracial children were included in the "Black" population. Although highly contentious, both politically and socially, review articles conclude that transracial adoption leads to a majority of satisfactory outcomes, similar to other samples of adopted children after infancy.

Obviously, adoptees enter the adoptive family with different life experiences. Those who are adopted as newborns experience family life and development much like children reared in biologic families. However, the majority of others come from preadoption experiences of severe deprivation, traumatic life experiences, and multiple disruptions in caregivers before adoption.

Developmental Issues in Adoptive Families

Although most adopted children seem to progress along typical developmental trajectories, often even demonstrating developmental catch-up after adoption, there are some developmental challenges encountered in the adoptive family that are peculiar to the realities of the situation. Numerous authors suggest that there is an element of loss, embodied in the process of relinquishment, in the psychological development of most adoptees, and that loss, particularly loss of the biologic child that the adoptive parents were not able to have, may be an important psychological issue for the adoptive parents as well. These losses may affect the attachments of the child to the parents and vice versa, attachments that are necessary for typical development to progress. Benoit, Nickman, and Rosenfeld grouped these losses into three categories: *Overt* losses of environmental stability, including relationships; *covert* losses, which embody how the children think about the reasons for adoption and the effects that being “given up” may have on developing self-esteem; and finally *status* loss, which arises from feeling stigmatized by the fact of the adoption, a loss experienced generally but more commonly in transracial or international adoptees. These elements of loss may contribute to difficulties that the child may have in establishing strong attachments to the adoptive family. The parents may also have ambivalence regarding the adopted child, sometimes having failed to mourn the loss of their own fertility. This ambivalence, along with extended-family attitudes that may be unsupportive of adoption, can impair the adoptive parents’ ability to attach wholeheartedly to the adopted child. Delays in the finalization of the adoption may also cause adopted parents to have some reservations regarding attachment for fear of further disappointment and loss. Thus, one of the first major developmental tasks for the adoptive family is the facilitation of successful attachment. Attachment difficulties are more complicated when adoptions occur after multiple disruptions in care over extended periods of time or after prolonged periods of institutional care. In these cases, immediate psychological interventions are imperative.

Adoptive parents have the additional responsibility of telling their adoptive children about the adoption and what they might know about the family of origin. Although in the past, secrecy was frequently attempted; most professionals today recommend early disclosure and ongoing processing of the fact of adoption throughout the developmental years. Many feel that early disclosure, even before children can fully comprehend the information, is recommended. Addressing the reality of the adoption early and in an integrated manner allows children to develop healthy identities as adoptees as they progress developmentally, both cognitively and emotionally, over time and keeps the topic open as something easily talked about within the family. This may become very important when adopted children begin to wonder about their families of origin and begin to ask questions that the adoptive parents must answer as honestly and openly as possible. Anxieties about rejection may spill over into the adoptive family, causing children to worry about being removed from their adoptive homes. These issues are often heightened during adolescence when children are normally struggling with issues of identity and separation from the nuclear family. Increased curiosity about families of origin may be difficult for adoptive families to address. Adolescents' typical desire to know more about their biologic families, even to meet them, may feel like rejection to the adoptive families. When there is adolescent turmoil within families, adolescents may idealize biologic families as a way to exert independence within their adoptive families. It is unclear how more "open" adoptions, particularly those that foster open relationships with biologic families, may affect these issues common in closed adoptions.

Over the developmental spectrum, adoptive parents may find parenting challenges not experienced by biologic parents. Frequently the adoptive parents enter the relationship with unresolved grief regarding their infertility, making them potentially less emotionally available to adopted children. Because of genetic differences as well as early experiences, the goodness of fit temperamentally and psychologically may not be optimal, making it more difficult for the adoptive parents to understand and respond empathically to the adopted children. Any problems the children have, any differences they may manifest, often tend to be viewed through the eyes of the adoption rather than as being normal variants within human development. Support for the adoption from extended families tends to decrease the effectiveness of some of these challenges.

Risk for Mental Health Problems

It has long been established that adopted children are over-represented in both outpatient and inpatient mental health settings. Some of these findings predated the recent increase in the number of children adopted from settings in which they had been neglected or abused prior to adoption. Some of this over-representation may result from the evidence that suggests that adoptees are referred for problems of less severity than nonadoptees, leading researchers to hypothesize that adoptive families were more likely to seek interventions than nonadoptive families. Consistently, adoptees demonstrate more risk, however low, for externalizing behaviors, including substance abuse, conduct symptoms, aggression, and possible suicidal behavior, behaviors that are more distressing to parents and the community and more likely to lead to mental health referrals. Nonetheless, analyses controlling for the severity of the mental health problems demonstrated an increased risk in adoptees over than in nonadoptees. Researchers postulate that the reasons for this over-representation in clinical populations are complex and may represent an actual increase in risk for problems in this population along with a referral bias among parents who, for a multitude of reasons, including higher educational and socioeconomic levels, may be more willing to access mental health interventions for their children. Many of the studies suggesting increased risk fail to differentiate clinical from nonclinical samples, the age of children when adopted, and the type of comparison group. Some suggest that the over-representation statistics in clinical populations may be colored by a small group of very serious cases. Large nonclinical studies have yet to be completed, but the ones that have been done continue to show risk factors for the adopted population, albeit at lower levels. Drug use, antisocial behavior, and school adjustment are compromised, although adopted adolescents have fewer problems with withdrawn behavior and have increased prosocial behavior compared to nonadoptees.

Over the years, it has been assumed that the older the child at the time of adoption, the higher the risk for mental health problems. Studies in the 1990s suggested, however, that although those children adopted as infants statistically demonstrated fewer behavioral and psychiatric problems, about 25% of them did have problems. Having a nonadopted sibling seemed to be a risk factor in these cases. In addition, studies have shown that deprivation in early childhood and later adoption resulted in the highest level of maladjustment. Yet, among these children, more than one-fourth of them demonstrated no serious problems as adolescents. Other studies have suggested few problems in those children adopted as infants, and all tend to show increased difficulties in adjustment relative to the age of adoption, with infants showing the least problems and children adopted after the age of 10 years showing the most problems when compared to matched controls.

Comparison groups in adoption outcome research have been problematic as well. If the comparison group is that of nonadopted children from intact families, adopted children show more problems with adjustment. However, adopted children seem to have better adjustment than those raised in foster care or institutions and even those returning from temporary care to families where there is a history of and potentially ongoing impaired parenting. The complex nature of evaluating outcomes demands addressing these issues, identifying risk factors that children bring to adoption, as well as risks within the adoptive family, and developing early interventions to help adoptive families and adoptees address these risk factors early and intensively so as to improve outcomes across the population.

Adopted children are frequently the source for research evaluating genetic versus environmental contributions to the development of mental illness. Researchers looking at schizophrenia have demonstrated that adoptees at high genetic risk for schizophrenia may be protected in optimal adoptive environments, whereas those who experience disordered parenting situations are more likely to develop schizophrenia, suggesting complex interactions between genetic and environmental contributions to the manifestation of the disorder.

International Adoptions

The practice of international adoptions experienced a gradual decline over the past 15 years. A recent high was in 2004 when 23,000 children were brought to the United States, while 2019 saw under 3,000 adoptions finalized. Geopolitics plays a strong role in the viability of international adoption. The Hague Convention on the Protection of Children and Co-operation in Respect of Intercountry Adoption (Hague Convention) was established in 1993 with the goal of setting international expectations for the regulation of cross-border adoptions. The Hague Convention set forth international standards for adoption, which increased the amount of oversight some countries had in place. The goal was to put safeguards in place so families have informed consent about the adoption process and, whenever possible, to keep children connected to their community. The United States signed on in 1994, but full implementation waited until 2008. There is disagreement over how much The Hague Convention directly reduced international adoption relative to changes in specific countries' adoption policies. For instance, Russia was a major source of international adoptions until nationalistic pressure led to a ban on U.S. adoptions in 2013. China's growing middle class and reversal of the one-child policy have dramatically reduced the number of children available for international adoption and shifted gender ratios toward parity. Families adopting from overseas also need to be aware that the adoption does not automatically grant citizenship. Although Congress has taken steps to streamline the process, it is estimated that over 15,000 individuals in the United States were adopted internationally but never had their citizenship established.

The serious deprivation in the early lives of many of these children has focused attention on the importance of early childhood experiences. Developmental problems, particularly motor delays, as well as a broad spectrum of medical problems, are commonly identified in children adopted internationally, particularly in those reared in orphanages. Approximately 12% of all adoptees have at least one chronic disability, twice the rate among all children in the same age range. Overall rates are similar among international and domestic adoptees. However, international adoptees who immigrated to the United States later in life and consequently experienced longer periods of deprivation prior to adoption have higher rates of impairment. The country of origin also influences disability status, as rates from children of Russian or Eastern European origin are much higher than those children hailing from China and Korea. This is likely due to a greater reliance on institutional care in Eastern Europe as well as alcohol consumption rates that are twice as high in Russia as they are in China. All international adoptees should be given a full medical and developmental evaluation within the first two weeks after arrival in the United States, with interventions instituted immediately for infections, lack of immunizations, and early intervention for growth and developmental delays. Repeated meta-analyses suggest that international adoptees have increased mental health problems, particularly externalizing behaviors. Preadoption conditions appear to be more predictive than the age of adoption. The longer the child is in a deprived early environment, the greater the risk for attachment difficulties. Interventions that foster attachment in these high-risk children should be implemented. Evidence suggests that children who are able to establish secure attachments with their adoptive mothers attain better developmental recovery, both cognitive and social, in middle childhood. Most studies show good outcomes for internationally adopted children. Factors that may increase the risk for developing mental health problems include male gender, older age at placement, preadoption care in an orphanage, and placement with a single parent or in a white-collar family. The story of international adoption over the last several decades highlights the effect of extreme early deprivation and the possible recovery in developmentally supportive families. It has become apparent that adoptions after 24 months of age in these severely emotionally deprived children complicate recovery. Longer deprivation leads to more significant developmental delays and less complete recovery. In addition, recovery, when it occurs, typically does so within the first 2 years after placement. The final chapter on long-term outcomes is yet to be written, but the evolving literature suggests that children severely deprived for long periods of time may continue to have problems with oppositional behaviors, attachment difficulties, odd or bizarre behaviors, including social behaviors, and language delays. It is obvious that adoption improves the outcomes in many of these severely deprived children, many of whom are beginning to tell stories of the confusing move from their countries of origin, being placed frequently in families who do not speak a language they have ever heard, along with the abrupt discontinuation of any exposure to their language of origin. Those who have age-appropriate competence in their language of origin typically have better outcomes. Examination of the typical developmental process through which these young children adjust to English as a second language warns clinicians against diagnosing selective mutism in these children who are adjusting to a world totally dominated by a new language. Nonetheless, retrospectively these international adoptees seem to balance these confusing circumstances and benefit greatly from the more empathic and developmentally appropriate care in their adoptive families.

The Role of the Psychiatrist

Many child and adolescent psychiatrists work intensively in outpatient, inpatient, and residential treatment settings. Some are involved early in the adoption of high-risk children through consultation with schools, child welfare agencies, and juvenile court settings. Although there is little research in the area, case reports identify the need for empathy for both the adoptive parents and the adoptee. Psychiatrists working with adoptive families need to explore their countertransference issues related to adoption, including issues related to abandonment and rescue, which may impair their abilities to work with the dyadic issues in a balanced, empathic, and objective manner. It is imperative that psychiatrists educate themselves regarding the complicated research regarding outcomes to avoid presuming that adopted adolescents will have problems, a practice that may lead to self-fulfilling prophecies. Embarking on a psychotherapeutic process with an adoptee must fully include the parents to increase open communication within the family. At times, adoptive parents harbor fears regarding the child's heredity, which may lead to the expectation of negative outcomes. Although without empirical evidence, group interventions for both adopted children and for their families may have the potential as effective treatment methodologies. Issues that commonly arise in psychotherapy with adoptees involve problems with self-esteem and fears of additional losses or rejection. However, recent studies have demonstrated the development of good self-esteem in most adoptees. Issues related to curiosity about biologic families are common concerns, sometimes leading to the opening of birth records without the consent of the birth parent. Another area of increasing focus relates to the common concern about treatment for attachment problems, which may be too frequently presumed in adoptive situations of early deprivation, causing clinicians to miss other treatable diagnoses. Controversial techniques such as holding therapy and rebirthing pose a significant physical risk of asphyxiation and should be discouraged. Newer interventions to foster attachment involving group educational programs on sensitive parenting coupled with focused feedback on the interactions between the parent and the child led to better infant outcomes, as well as enhanced parent–child attachments. Although the reality of adoption must enter into the psychotherapeutic equation, psychiatrists are cautioned to integrate issues related to adoption with an overall individualized assessment of the child in the family, avoiding the tendency to ascribe all psychopathology to issues related to adoption. Psychiatrists are obviously more likely to work with adoptive situations in which the children and families are not doing well. It is important to remember that, whereas adoptive status may influence the situation and negative experiences prior to the adoption exert great effect on the outcomes, each family needs to be evaluated thoroughly, and targeted interventions must be implemented in an effort to preserve the family because adoption disruptions in older adoptions are higher compared to adoptions in infancy. Data are difficult to come by regarding disruptions, but overall rates are around 10%.

Another consideration regarding treatment is that adolescence is the most challenging time for adopted children for both behavioral and emotional problems, regardless of the stage in life at which the child was adopted. This is partially due to the universal challenges of adolescence, but heightened due to the loss and trauma many adopted children experienced early in life. These experiences resurface during adolescence and can catch the adoptive parent off guard. Many adoptive parents envisioned a “happily ever after” scenario and believed that their love would be enough to help a child process issues related to grief, identity, and early childhood stressors. Consequently, adoptive families are four times as likely to seek counseling services and seven times as likely to seek residential treatment when compared to birth families. Problems may become so severe that a youth is at risk of re-entering foster care or the adoption is legally dissolved. From the limited data available, these outcomes are rare with approximately 10% of children who are adopted through the child welfare system finding themselves placed back into foster care. Of that 10%, approximately one-third may have the adoption terminated. Another 10% have extended disruptions to their living situation such as running away, living with another family member, or having an extended placement in a residential treatment facility or a juvenile corrections facility. As federal policy only began to encourage adoption from foster care in the 1980s and 1990s, postadoption services to families with special needs children are relatively new. Services are becoming more widespread with the recognized need to avoid disruption in placement. However, they tend to vary significantly from state to state and it can prove a challenge to locate all services for families in need. Encouraging families to seek out services that improve their knowledge and skills is vital, as is reminding them that this is a sign of family strength, not parental inadequacy, as so many may believe. Education is where intervention begins and a discussion of the adopted child in light of the impact that early abuse and deprivation can cause will provide a greater understanding of the child, increase empathy, and help them look for ways to provide support and build on a child's strengths to promote resilience. It is also helpful to discuss realistic expectations of a child and to encourage parents to take care of their own well-being to promote better interactions with the child.

Psychiatrists may also work with the biologic parents and families of children who have been given up for adoption or whose children have been removed because of abuse and neglect. These persons often have pre-existing psychopathology, which might become more complicated because of the removal of their children. In addition, for those whose children are in foster care and for whom reunification is a goal, intensive psychiatric interventions, and support may be the most significant influence on the ultimate outcome.

Future Directions

It is clear that for many children and families, adoption meets a special need and facilitates a positive psychological outcome. For a minority, psychological adjustment is more challenging. Identifying, more specifically, the risk factors for less-than-optimum outcomes and developing evidence-based interventions to address critical issues like attachment, externalizing behaviors, and increasing empathic responses of adoptive families to high-risk children is the challenge for the future. Few current interventions take into account the complex nature of repeated trauma, attachment, and identity issues faced by adopted children. Policies that facilitate adoptions earlier when there is little hope for family reunification are imperative to interrupt the prolongation of temporary placements, which increases the risk of compromised adoption outcomes. Additional work is also needed to provide adoptive families with postadoption support to address the needs of the child as they are identified. Egregious policies that remain law in some states, which require adoptive parents to formally "abandon" children in order to obtain residential treatment, call for immediate correction. More needs to be learned about birth parents and optimal management of open adoptions or late reunions, which are becoming more frequent but about which very little is known. Finally, continued efforts are needed consistently to keep the best interests of the child at the forefront in all decisions regarding permanency planning, whether it is related to reunification, long-term foster care, or adoption.

FURTHER READINGS

American Academy of Child and Adolescent Psychiatry. Appropriate use of psychiatric medication for youth in foster care policy statement. 2018;1–2.

Askeland KG, Hysing M, La Greca AM, Aarø LE, Tell GS, Sivertsen B. Mental health in internationally adopted adolescents: a meta-analysis. *J Am Acad Child Adolesc Psychiatry*. 2017;56(3):203–213.

Carnochan S, Moore M, Austin MJ. Achieving placement stability. *J Evidence Based Social Work*. 2013;10(3):235–253.

Council on Foster Care, Adoption, and Kinship Care and Committee on Early Childhood. Health care of youth aging out of foster care. *Pediatrics*. 2012;130(6):1170–1173.

Szilagyi MA, Rosen DS, Rubin D, Zlotnik S; Council On Foster Care, Adoption, and Kinship Care; Committee on Adolescence; Council on Early Childhood. Health care issues for children and adolescents in foster care and kinship care. *Pediatrics*. 2015;136(4):e1142–e1166.

Grotevant HD, McDermott JM. Adoption: biological and social processes linked to adaptation. *Ann Rev Psychology*. 2014;65:235–265.

Gupta-Kagan J. America's hidden foster care system. *Stanford Law Review*. 2020;841–913.

Hernandez P, Lee J. Outcomes of young adults aging out of foster care: a latent class analysis. *Child Welfare*. 2020; 98(3):145–166.

Kandel WA. Unaccompanied alien children: an overview. Congressional Research Service (CRS Report No. R43599). 2019.

Lee T, Fouras G, Brown R, American Academy of Child and Adolescent Psychiatry (AACAP) Committee on Quality Issues (CQI). Practice parameter for the assessment and management of youth involved with the child welfare system. *J Am Acad Child Adolesc Psychiatry*. 2015;54(6):502–517.

Macdonald G, Turner W. Treatment foster care for improving outcomes in children and young people. *Cochrane Database Syst Rev*. 2008;1–76.

McCormick A, Schmidt K, Terrazas S. LGBTQ youth in the child welfare system: an overview of research, practice, and policy. *J Public Child Welf*. 2016;11(1):27–39.

National Working Group on Foster Care and Education. Fostering success in education: national factsheet on the educational outcomes of children in foster care. 2018;1–34.

National Youth in Transition Database. Report to Congress: Executive Summary. 2019;1–20.

Nickman SL, Rosenfeld AA, Fine P, et al. Children in adoptive families: overview and update. *J Am Acad Child Adolesc Psychiatry*. 2005;44(10):987–995.

Sahota PC. Kinship care for children who are American Indian/Alaska Native: state of the evidence. *Child Welfare*. 2019;97(2):63–78.

Sheridan M, Drury S, McLaughlin K, Almas A. Early institutionalization: neurobiological consequences and genetic modifiers. *Neuropsychol Rev*. 2010;20(4):414–429.

Simkiss DE, Stallard N, Thorogood M. A systematic literature review of the risk factors associated with children entering public care. *Child Care Health Dev*. 2013;39(5):628–642.

Texas Department of Family and Protective Services, The University of Texas at Austin College of Pharmacy, Rutgers University-Center for Education and Research on Mental Health Therapeutics. Psychotropic medication utilization parameters for children and youth in foster care (5th Version). 2016.

U.S. Department of Health and Human Services, Adoption and Foster Care Analysis and Reporting System (AFCARS). The AFCARS report preliminary FY 2019 estimates as of June 2020. No. 27. 2020;1–6.

U.S. Department of Health and Human Services, Administration on Children, Youth and Families, Children's Bureau. Trends in foster care and adoption: FFY 2010-FFY 2019. (Based on data submitted by States as of June 23, 2020). 2020;1–4.

U.S. Department of Health & Human Services, Administration on Children, Youth and Families, Children's Bureau. Trends in U.S. adoptions: 2008–2012. *Child Welfare Information Gateway*. 2016;1–8.

U.S. Department of Health & Human Services, Administration on Children, Youth and Families, Children's Bureau. Child welfare outcomes 2016: report to congress executive summary. 2019;1–64.

U.S. Department of Health and Human Services, Administration on Children, Youth and Families, Office of Refugee Resettlement. Unaccompanied children (UC) program [fact sheet]. 2021;1–2.

U.S. Department of State, Bureau of Consular Affairs. FY 2019 Annual report on intercountry adoption. 2020;1–10.

Van Wichelen S. Moving children through private international law: institutions and the enactment of ethics. *Law Soc Rev*. 2019;53(3):671–705.

Winokur M, Holtan A, Batchelder KE. Kinship care for the safety, permanency, and well-being of children removed from the home for maltreatment (review). *Cochrane Database Syst Rev*. 2014;1:1–239.

Wolfgram SM. Openness in adoption: what we know so far—a critical review of the literature. *Soc Work*. 2008;53(2):133–142.

▲ 54.2 Child Maltreatment

HEATHER WILLIAMS, M.D., AND BRADLEY FREEMAN, M.D.

Child maltreatment is a grossly underrecognized problem that is pervasive across all socioeconomic classes, races, ethnic groups, and religions. Although the federal government and every state have programs to investigate allegations of abuse, prevent abuse, and educate caregivers, the cases that are reported and substantiated are just the tip of the iceberg of the true prevalence of child maltreatment. For centuries, the recognition of child maltreatment has been compromised by pervasive denial and systemic constraints related to the recognition, reporting, investigation, and intervention of abuse and neglect. The physical and mental health impacts of child maltreatment can be severe and long-lasting, with the potential for the psychological effects to endure long after any physical or sexual injury has healed.

A diverse group of professionals—including child and adolescent psychiatrists, psychologists, social workers, law enforcement, school personnel, pediatricians, dentists, emergency department workers, and other mental health professionals, are mandated to report suspected child maltreatment. Additional professionals such as pathologists, radiologists, and attorneys, for example, often participate in the identification, evaluation, and substantiation of child maltreatment. There are many roles for child and adolescent psychiatrists and psychologists, social workers, and other mental health professionals in these cases. The above clinicians may provide treatment to children who experience significant symptoms due to trauma, such as behavioral and emotional problems because of abuse. Mental health professionals may evaluate and treat perpetrators of abuse, many of whom may have been victims of abuse themselves earlier in their lives. Clinicians may provide evaluations of children and adolescents suspected of having been maltreated for forensic or legal purposes. A forensic evaluation may assist the court in determining specific details and the chronology of maltreatment in a child or adolescent. Child psychiatrists may be asked to evaluate the credibility of children's and adolescents' reports of maltreatment and the competency of the youth to testify. In civil lawsuits pertaining to child maltreatment, psychiatrists and other mental health professionals may testify about the nature, extent, and cause of the child's psychological injuries. Finally, mental health professionals may address the problem of child maltreatment on the level of public policy by providing education to legislators, judicial officers, and attorneys about the psychiatric aspects of abuse and the developmental needs of children.

DEFINITIONS

DSM-5-TR

The somewhat terse classification system provided by the Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR) lists *child physical abuse*, *child sexual abuse*, *child neglect, and child psychological abuse*. The four conditions appear in the chapter "Other Conditions That May Be a Focus of Clinical Attention" and in the section "Abuse and Neglect." These categories should be used when the focus of clinical attention is the severe mistreatment of one individual by another. The coding system of DSM-5-TR allows the evaluator to distinguish each type of maltreatment: confirmed abuse from suspected abuse; an initial encounter from a subsequent encounter; parental abuse from nonparental abuse; whether the abuse previously occurred during the person's childhood; and whether the mental health services are for the perpetrator of parental or nonparental abuse.

Federal Law

The Child Abuse Prevention and Treatment Act (CAPTA) was passed in 1974 and has been amended several times, most recently in January 2019, by the Victims of Child Abuse Act Reauthorization Act of 2018. In federal law, *child abuse and neglect* mean, as a minimum, any recent act or failure to act on the part of a parent or caretaker which results in death, serious physical or emotional harm, sexual abuse or exploitation, or an act or failure to act which presents an imminent risk of serious harm. In federal law, *sexual abuse* means the employment, use, persuasion, inducement, enticement, or coercion of any child to engage in or to assist any other person to engage in any sexually explicit conduct or simulation of such conduct for the purpose of producing a visual depiction of such conduct. CAPTA further expands the legal definition of child sexual abuse to include the following terms: rape (and in cases of caretaker or familial relationships, statutory rape), molestation, incest, and prostitution. CAPTA was also most recently amended to include other forms of sexual exploitation of children and victims of sex trafficking.

State Law

A large mass of legal definitions and statutes exist at the state level. The legal definitions of terms related to the maltreatment of children vary from one jurisdiction to another, so clinicians should be aware of their state statutes, with a special focus on mandatory reporting laws, corporal punishment laws, and legal age of consent.

Medical Diagnoses

Child maltreatment is a unique area of medicine where the medical diagnosis and legal findings overlap. For this reason, the diagnoses listed below are informed by legal definitions and statutes. However, the diagnosis of child maltreatment in the clinical setting is independent of the investigative process dictated by these laws and is determined based on the medical provider's level of concern for maltreatment. There are also variations in terminology used between the medical and legal fields.

Child Neglect

The most prevalent form of child maltreatment, *child neglect,* is a failure to provide or supervise and can be committed through acts of commission or omission. Neglect can fall into several categories, including the following:

- Physical neglect, such as abandonment
- Nutritional neglect
- Medical neglect
- Inadequate supervision
- Emotional neglect
- Educational neglect

Neglect may also be chronic, where an individual incident of neglect may not appear harmful, but multiple incidents over time create a negative impact on the child.

Child Physical Abuse

As described above, the minimum definition of *child physical abuse* involves an act on the part of a child's caretaker, which results in death, serious physical harm, or imminent risk of serious harm. In the clinical setting, child physical abuse may present with a wide range of symptoms, from subtle bruising on a nonmobile infant to abusive head trauma resulting in long-term disability or death. Determining the level of concern for child physical abuse in the clinical setting depends on several factors, including the constellation of symptoms and injuries, the history provided to explain these injuries, and the developmental capabilities of the child. If there is no history provided to explain the injuries, if the history provided is implausible to explain them, if the child is developmentally incapable of causing the injuries, or if abuse is disclosed, witnessed, or confessed this should raise concerns for child physical abuse, warranting a report to local child protective services and consideration of additional medical evaluation.

Corporal punishment, or physical discipline, includes acts such as spanking or paddling. Each state differs in its laws regarding the use of corporal punishment in the home and at school. Regardless of state law regarding the use of corporal punishment, if the discipline resulted in serious physical harm or imminent risk of physical harm, it should be considered child physical abuse and reported as such.

Abusive head trauma is a specific type of child physical abuse with potentially long-lasting neurocognitive and mental health impacts related to the intracranial injuries and abuse. Medical providers are strongly discouraged from using the term “shaken baby syndrome,” as this inappropriately focuses on one specific possible mechanism of injury instead of the abuse itself.

Child Psychological Abuse

Child psychological abuse is a pattern of intentional verbal or behavioral actions or lack of actions that conveys to a child that they are worthless, flawed, unloved, unwanted, or endangered. Child psychological abuse is divided into the following categories: spurning, rejection or degradation, terrorizing, exploiting or corrupting, denying emotional responsiveness, and isolation. The severity of psychological abuse depends on (1) whether the perpetrator intends to inflict harm on the child and (2) whether the abusive behaviors are likely to cause harm to the child. In contentious divorces, one of the parents may indoctrinate the child to fear or dislike the other parent, thus causing parental alienation between the child and the rejected parent. Depending on the circumstances, parental alienation may be identified by one or more of these DSM-5 terms: child psychological abuse, parent–child relational problem, or child affected by parental relationship distress.

Child Sexual Abuse

Child sexual abuse is defined by the American Academy of Pediatrics as occurring when a child is engaged in sexual activities that they cannot comprehend, for which they are developmentally unprepared and cannot give consent, and/or that violate the law or social taboos of society. Recent definitions and laws have expanded to include children who are victims of human trafficking and internet crimes against children, such as child pornography. State laws vary greatly regarding age of consent, definitions of terms such as penetration, and timing for evidence collection related to an acute sexual assault. Inappropriate sexual contact with a child can include kissing, fondling, digital–genital, oro–genital, genital–genital, and anogenital contact. The vast majority of child sexual abuse is perpetrated in the home by a close household family member.

Children commonly engage in sexualized behaviors with themselves or with other children, described as “sexual play” that is spontaneous, intermittent, mutual, and noncoercive. Normal sexualized behaviors typically include acts such as self-stimulation and voyeurism. It is important to recognize that these behaviors may be distressing to caregivers but are not an indication that child sexual abuse has occurred. Reassurance and education regarding these behaviors are frequently all that is needed when they present. Sexualized behaviors become problematic when they cross socially unacceptable boundaries, become coercive or aggressive, or represent an early onset of adult sexual activity. When the behaviors become problematic, further interventions to extinguish the behaviors and evaluation for possible child maltreatment may be indicated.

HISTORY OF CHILD MALTREATMENT

Historical and literary sources have documented child maltreatment for thousands of years in cultures from all parts of the world. Although some forms of abuse were considered deviant or prohibited in particular cultures, most often the abusive practices were accepted or even encouraged by the societies in which they occurred. The long-term trend in European and American history has been to move from blatant, socially acceptable child abuse to strong societal disapproval.

Since the latter half of the 20th century, child abuse became the subject of widespread clinical and governmental concern. Abusive practices were actively identified, rather than ignored and denied. Modern recognition of child abuse was signaled in 1946 by a pediatrician and radiologist, John Caffey, who noticed a syndrome of children with multiple skeletal injuries and chronic subdural hematomas. Until the 1960s, physical abuse of children was considered rare, partly because physical discipline of children was generally more acceptable and partly because of societal denial concerning violence toward children. In 1962, Henry Kempe and his colleagues brought child maltreatment to the attention of physicians and legislators in a historic article, “The Battered-Child Syndrome.” Within 5 years, child abuse reporting laws were adopted by every state.

It took a separate societal realization during the 1970s to acknowledge the extent of sexual abuse. Child sexual abuse perpetrated by their caregivers, previously referred to as incest, was known to occur, but most people believed that it was unusual and happened primarily among deviant families. It is now known that sexual abuse perpetrated by family members and other forms of sexual abuse are not rare. The term “incest” is no longer preferred in the clinical domain as it does not necessarily pertain to children and it suggests a degree of consent.

In 1974, the federal government passed the CAPTA, which established uniform national standards for the identification and management of child abuse cases. Child abuse legislation, originally directed toward case discovery and perpetrator punishment, now focuses on prevention, rehabilitation, and family reunification. Individual states independently define maltreatment, formulate investigative procedures, collect data, and organize service delivery systems.

Research in the field of child maltreatment has continued to grow, and there are now teams of maltreatment providers in hospitals across the United States and internationally whose purpose is to assess children with concerns for maltreatment and collaborate with community partners to ensure the safety of these children. Maltreatment providers with specialized certification and training exist at most major academic children's hospitals and offer a range of services including consultations and assessments of patients presenting with concerns for child physical abuse, child sexual abuse, and neglect. They typically work closely with their social work colleagues to liaise with local investigative agencies and provide expert testimony in court when needed.

COMPARATIVE NOSOLOGY

The first edition of the DSM (DSM-I) (1952) and the second edition of the DSM (DSM-II) (1968) made no mention of child maltreatment. The Group for the Advancement of Psychiatry (1974) mentioned child maltreatment among the pathogenic factors of childhood mental disorders: "The deleterious effects … of open rejection or neglect, occasionally involving willful injury or inadequate feeding, are well documented." However, the Group for the Advancement of Psychiatry did not include child maltreatment in their proposed system of classification.

When the third edition of the DSM (DSM-III) (1980) was published, the American Psychiatric Association Task Force on Nomenclature and Statistics introduced a chapter called "V Codes for Conditions Not Attributable to a Mental Disorder That Are a Focus of Attention or Treatment." The V Codes, which were modeled on the practice in the ninth edition, clinical modification of the *International Classification of Diseases* (ICD-9-CM), did not include specific categories for child maltreatment. However, it was intended that child abuse would be coded as a *parent–child problem*, the category used "when a focus of attention or treatment is a parent–child problem that is apparently not due to a mental disorder of the individual (parent or child) who is being evaluated." The DSM-III specifically stated that "an example (of parent–child problem) is child abuse not attributable to a mental disorder of the parent." The emphasis on *parent–child* *problem* was apparently on identifying perpetrators of abuse, not the child victims of abuse.

The revised third edition of the DSM (DSM-III-R) (1987) provided a similar definition for *parent–child problem*: “This category can be used for either a parent or a child when the focus of attention or treatment is a parent–child problem that is apparently not due to a mental disorder of the person who is being evaluated.” However, the example involving child abuse was replaced by “conflict between a mentally healthy adolescent and her parents about her choice of friends.”

In the fourth edition of the DSM (DSM-IV) (1994), the title of the chapter changed to “Other Conditions That May Be a Focus of Clinical Attention.” In addition, a specific section was introduced, “Problems Related to Abuse or Neglect,” which included terms related to child maltreatment. The editors noted that this category was “included because of the clinical and public health significance of these conditions.” For the first time, it was possible to diagnose clearly *physical abuse of child*, *sexual abuse of child*, and *neglect of child*. Furthermore, the DSM-IV made it possible to distinguish whether the person being evaluated was the perpetrator of abuse or the victim of abuse. The terminology regarding child maltreatment in the DSM-IV-TR (2000) was the same as that in the DSM-IV.

The fifth edition of the DSM (DSM-5) (2013) and the text revision (DSM-5-TR) (2022) include child maltreatment in the same chapter as in the previous edition, “Other Conditions That May Be a Focus of Clinical Attention.” A new section was included titled “Abuse and Neglect.” This section encompasses abuse of all persons, young and old. Diagnosing abuse with the current DSM nomenclature has become more specific. The child can be diagnosed as being a victim of *child physical abuse, child sexual abuse, child neglect*, and *child psychological abuse*. It remains possible to differentiate whether the clinical service was for the victim or the perpetrator of abuse. For adults, the provider can indicate if the abuse or neglect is from the individual’s childhood history.

Additionally, the diagnoses of parent–child relational problems and child affected by parental relationship distress are also included in the same chapter but under the section titled “Relational Problems.” That section includes the language, “In the extreme, these close relationships can be associated with maltreatment or neglect, which has significant medical and psychological consequences for the affected individual.” The description of parent–child relational problem in DSM-5 indicates the child is affected in either the behavioral, cognitive, or affective domain.

The tenth edition of the ICD (ICD-10, Version 2007) has a slightly different way of organizing child maltreatment diagnoses. The ICD-10 has a chapter titled “Injury, Poisoning and Certain Other Consequences of External Causes” and a section titled “Other and Unspecified Effects of External Causes.” Within that section is a heading for “Maltreatment Syndrome,” which includes the following diagnoses: neglect or abandonment; physical abuse; sexual abuse; psychological abuse; other maltreatment syndromes; and maltreatment syndrome, unspecified. The ICD-10-CM (Clinical Modification) includes codes that enhance the specificity of the abuse or neglect diagnosis. For example, the abuse can be coded as either *suspected* or *confirmed* and the type of encounter can be denoted as either *initial* or *subsequent*.

The eleventh edition of the ICD (ICD-11, Version 2019) has a chapter entitled “External causes of morbidity or mortality.” Maltreatment is a subsection in this chapter and encompasses physical, sexual, psychological, specified, and unspecified categories. This version of the ICD no longer includes the section “Maltreatment Syndrome.” There are extension codes that are provided to add detail to defined categories. Abandonment and neglect are included in the list of codes. Under the “Factors influencing health status or contact with health services” chapter, there is coding for personal history of physical, sexual, and psychological abuse, and neglect in addition to unspecified and specified history of maltreatment.

EPIDEMIOLOGY

Each year, the Children's Bureau, an agency within the Department of Health and Human Services, collects data on child maltreatment. The results are published in an annual document called *Child Maltreatment*. The agency estimated that, in 2018, approximately 4.3 million reports with concern for abuse or neglect were made to child protective services involving 7.8 million children. Of those reports, 678,000 reports were substantiated within each individual state's statutes defining abuse and neglect, 1.3 million received postresponse services, and over 200,000 children were placed into foster care. The substantiated cases were distributed as follows: 61% neglect, 11% physical abuse, 7% sexual abuse, and 15.5% two or more types of abuse and neglect. The Children's Bureau estimated 1,770 children died as a result of maltreatment in 2018, over an 11% increase from the 2014 estimate. Approximately 47% of these deaths were children younger than 1 year of age and boys (2.87 per 100,000) had a higher fatality rates than girls (2.19 per 100,000).

The data were analyzed for patterns of maltreatment by the sex, age, and race or ethnicity of victims. With regard to the age distribution of victims, the age group from 0 to 1 years of age had the highest victimization rate (26.7 per 1,000 children) with the rate of victimization decreasing with increasing age of the victim. The majority of identified child victims were either White (44.5%), Hispanic (22.6%), or African-American (20.6%). American-Indian or Alaska Native children had the highest rate at 15.2 per 1,000 children, and African-American children had the second higher rate at 14.0 per 1,000.

All of these figures are approximations because the actual amount of abuse is unclear. The reporting of abuse and the annual victimization rate increased during the 1980s and reached a high in 1993 (when the victimization rate was 15.2 per 1,000 children in the population). This increase is thought to have resulted from greater public awareness and willingness to report child abuse, improvement in data collection techniques by individual states, and local economic conditions that placed a larger number of families under stress. The victimization rate fell after 1993 and reached a low of 9.2 per 1,000 children in 2012. Much as it is difficult to definitively state why the rate increased in the 80s and 90s, it is also difficult to state why the rates have decreased in recent years. This decrease may be the result of a true decrease in child maltreatment related to prevention programs but may also reflect differences in reporting, data collection, and state laws. Regardless of the reason for fluctuations in the number of reports per year, these numbers likely reflect a significant underestimate of the true extent of maltreatment.

RISK FACTORS

Characteristics of a child that may be risk factors for maltreatment include age less than 4 years old, prematurity, emotional/behavioral difficulties, physical disabilities, and unplanned or unwanted pregnancy. Risk factors associated with caregiver and family characteristics include substance abuse, parental mental illness, lack of understanding of child's needs or developmental capabilities, caregiver history of child abuse or neglect, and young parental age. Additional family characteristics that are often associated with child maltreatment are low caregiver educational achievement, low caregiver socioeconomic status, single parenthood, large number of children in the home, nonbiologic caregivers, and domestic violence in the home. Community and societal risk factors include community violence, concentrated neighborhood disadvantages, poverty, and social isolation. These factors may be difficult to identify and assess. It is also important to be aware of the unconscious bias identification of these risk factors may create, as their presence does not mean abuse or neglect had to have occurred nor does their absence mean abuse or neglect did not occur. These factors must be considered in the overall context of the individual case. Community and societal factors that may be protective for a child are a strong extended family and social support system.

DIAGNOSIS AND CLINICAL FEATURES

Abused children manifest a variety of emotional, behavioral, and somatic reactions. These psychological symptoms are neither specific nor pathognomonic: the same symptoms may occur without a history of abuse. The psychological symptoms manifested by abused children and the behaviors of abusive parents can be organized into clinical patterns. Although it may be helpful to note whether a particular case falls into one of these patterns, that in itself is not diagnostic of child abuse.

Physically Abused Children

Abused children display behaviors that should arouse the suspicions of the health professional. For example, these children may be unusually fearful, docile, distrustful, and guarded. However, they may be disruptive and aggressive. They may be wary of physical contact and show no expectation of being comforted by adults, they may be on the alert for danger and continually size up the environment, and they may be afraid to go home.

The literature regarding the psychological consequences of physical abuse and neglect indicates a wide range of effects: affect dysregulation, insecure and atypical attachment patterns, impaired peer relationships involving increased aggression or social withdrawal, and academic underachievement. Physically abused children exhibit a range of psychopathology, including depression, conduct disorder, attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), dissociation, and post-traumatic stress disorder (PTSD).

Carol, 4 years of age, had a change in her behavior at preschool approximately 3 months after the birth of her sister. Her teacher saw Carol push other children and hit a classmate with a wooden block, causing a laceration of the child's lip. When Carol's teacher took her aside to talk about her behavior, she noticed what seemed to be belt marks on Carol's abdomen and forehead. The teacher reported possible child abuse to protective services. In addition, the family was referred for psychiatric evaluation.

Carol's baby sister was colicky and slept only for short periods of time throughout the day and night. She stopped crying only when her mother held her. Her mother, therefore, had little time for Carol, and Carol's father took over her care on evenings after daycare and on weekends. He began to drink more than usual and became increasingly irritable. The parents argued over the mother's attention to the infant and the requirement that the father takes care of Carol. Carol, who was a bright, curious, and talkative child, constantly asked questions and often asked to carry the baby. When refused, she would lie on the floor and have a tantrum. She also began to have difficulty falling asleep and awoke repeatedly during the night. Carol's father was unable to cope with her requests for attention and often told her to "shut up" and slapped her when she continued her demands. On many occasions, he responded to her tantrums or repeated questions by hitting her with his belt.

While protective services monitored the situation, Carol and her parents began a family therapy program that included parenting training and behavioral therapy for Carol, which was coordinated with the preschool. Carol's father attended Alcoholics Anonymous (AA) meetings and stopped drinking. He was able to control his anger at his daughter. Six months later, Carol's aggressive behavior ceased. She was doing well with peers, was sleeping through the night, and stopped having temper tantrums.

Sexually Abused Children

A variety of symptoms, behavioral changes, and diagnoses sometimes occur in sexually abused children: anxiety symptoms, dissociative reactions and hysterical symptoms, depression, disturbances in sexual behaviors, and somatic complaints.

Anxiety Symptoms. Anxiety symptoms include fearfulness, phobias, insomnia, nightmares that directly portray the abuse, somatic complaints, and PTSD.

Dissociative Reactions and Hysterical Symptoms. The child may exhibit periods of amnesia, daydreaming, trance-like states, hysterical seizures, and symptoms of dissociative identity disorder.

Depression. Depression may be manifested by low self-esteem and suicidal and self-mutilative behaviors.

Disturbances in Sexual Behaviors. Sexualized behaviors in children toward other children or adults are particularly suggestive of sexual abuse. Behaviors exhibited by young children such as repeated masturbating, imitating intercourse, or inserting objects into the vagina or anus are suggestive of sexual abuse. Sexually abused children may display sexually aggressive behavior toward others. Other sexual behaviors are less specific, such as showing genitals to other children. A young child who experienced sexual abuse may manifest age-inappropriate sexual knowledge. In contrast to these overly sexualized behaviors, a child who has been sexually abused may avoid sexual stimuli as they grow and develop.

Somatic Complaints. Somatic signs suggestive of possible sexual abuse include enuresis, encopresis, anal and vaginal itching, and nonspecific somatic complaints such as frequent headaches, and stomachaches.

The above symptoms are not pathognomonic of maltreatment. Nonabused children with anxiety disorders, and other psychiatric disorders may also exhibit some of the symptoms and behaviors. For example, normal, nonabused children commonly exhibit sexual behaviors such as masturbating, displaying their genitals, and trying to look at people who are undressing.

Approximately one-third of identified sexually abused children may not exhibit observable symptoms. Most adults who were abused as children have no obvious significant abuse-related symptoms. However, the following factors tend to be associated with more severe symptoms in the victims of sexual abuse: greater frequency and duration of abuse, sexual abuse that involved force or penetration, and sexual abuse perpetrated by the child's father or stepfather. Other factors associated with poorer prognosis are the child's perception of being less believed, family dysfunction, and lack of maternal support. In addition, multiple investigatory interviews appear to increase symptoms.

Child Sexual Abuse Perpetrators

Child sexual abuse that occurs over a period of time may be characterized by a particular pattern or sequence of steps. Victims of sexual abuse recount a gradual progression of boundary violations by the perpetrator, starting with tiny invasions and escalating to serious, overwhelming intrusions. Healthy, self-confident children may rebuff the intrusions directly (via temper tantrums and verbal disagreements) or indirectly (through silence and distancing maneuvers), or by adopting a strategy that causes the offender to refrain.

Sexual abuse that occurs over a period of time typically involves a close household contact of the child and may evolve through phases. There may be a period of engagement or "grooming" where the perpetrator gains the child's trust or the trust of their caregivers in order to gain access to the child prior to engaging in inappropriate sexual contact. The sexual contact may be ongoing and escalate over time. Sexual abuse includes any form of inappropriate sexual contact with a child, including but not limited to kissing, digital–genital contact, oro–genital contact, anogenital contact, and genital–genital contact. The child may not disclose the abuse for an extended period of time due to fear, shame, or not feeling like they will be believed or supported, and the perpetrator may engage in threatening ways to keep the inappropriate sexual contact a secret. Sometimes the sexual abuse is discovered accidentally without a disclosure. If the child does not feel safe, supported, or believed, it is not uncommon for the victim to recant, deny, or minimize the abuse.

Extrafamilial perpetrators of sexual abuse against children usually have regular access to the child. Children may be repeatedly abused by trusted adults, such as teachers, counselors, family friends, and clergy. In this scenario, the pedophilic perpetrator grooms the child over a period of time. They gain the friendship of children through enjoyable activities and gifts, introduce sexual activities that may seem innocent and even pleasurable, and progress to more intrusive activities.

A financially comfortable couple lived in a pleasant, clean house in a nice neighborhood, but they had no friends. Their four teenagers never had visitors. One day, the oldest girl, 17 years of age, went to the police and told them she had a baby at home and her father was the baby's father. The girl said her father had been having sexual relations with her for more than 4 years and he was now doing the same with her younger sisters. The mother admitted knowing about the situation for years, but she had not reported it to the authorities for fear of losing her husband.

Long-Term Consequences of Maltreatment

Many adults who were abused as children have no easily identifiable significant abuse-related symptoms. However, epidemiologic studies —such as the Adverse Childhood Experiences (ACE) Study—have shown a link between childhood adversity and adult psychopathology in diverse populations. It is likely this vulnerability is mediated by persistent physical consequences of the maltreatment on the child's developing brain.

Neurobiologic Aspects of Child Maltreatment

Genetics. Research has described the interaction between genetics and environmental factors (referred to as G × E interactions). Although G × E interactions have been studied for nearly a century in biology and general medicine, the mapping of the human genome has made it possible to study complex interactions among specific genes and between genes and the environment. For example, a G × E interaction appears to play an important role in the neurobiologic response to childhood maltreatment. In several studies, maltreated male children with a genotype conferring high levels of monoamine oxidase A (MAOA) expression were less likely to develop antisocial behavior as adults. In other studies examining the role of genetics in modulating the response to stress, a functional polymorphism in the promoter region of the serotonin transporter gene (known as 5-HTTLPR and *SLC6A4*, respectively) was found to affect the likelihood of developing depression and suicidality in response to stressful experiences. Individuals with one or two copies of the "short allele" were more likely to develop depressive symptoms than individuals with two "long alleles."

Neuroimaging. Childhood maltreatment exposure increases the person's risk of major depression, substance abuse, anxiety disorders, and PTSD, and maltreated individuals with those disorders tend to have an earlier age of onset, greater symptom severity, more comorbidity, a greater risk of suicide, and poorer treatment response than nonmaltreated individuals with the same conditions. It is noteworthy that imaging findings associated with these disorders—such as reduced hippocampal volume and amygdala hyper-reactivity—are observed more consistently in individuals with a history of child maltreatment. Trauma victims who develop PTSD have changes in brain activation that are consistent with the individual's experienced symptoms. For example, individuals with PTSD who report reliving their traumatic events in the form of flashbacks have reduced activation in the medial prefrontal cortex and anterior cingulate cortex (observed on functional magnetic resonance imaging [fMRI]), which is consistent with less regulatory control over affective arousal states. However, individuals with PTSD who experience dissociation when exposed to trauma memories have increased activation in the medial prefrontal cortex and anterior cingulate cortex, which is consistent with robust inhibition of affective arousal states.

Chronic physical abuse and repeated sexual abuse cause changes in the child's developing brain that persist into adulthood. Chronic severe abuse can lead to a syndrome known as complex PTSD or disorder of extreme stress not otherwise specified (DESNOS). This disorder, which was evaluated but rejected by the Post-Traumatic and Dissociative Disorders Sub-Work Group for DSM-5, is characterized by alterations in the following: ability to modulate emotions; attention or consciousness; somatization; self-perception; perception of the perpetrator; relations with others; and systems of meaning. The core of this disorder is a breakdown in the ability to regulate internal states. Trauma exposure changes the way in which children perceive environmental stimuli, especially their perception of threats, and how they organize and process information. Specifically, individuals with a history of trauma are more vulnerable to reacting with subcortically initiated responses that are irrelevant or even harmful in the present situation.

EVALUATION PROCESS

The evaluation of a child or an adolescent who may have been physically or sexually abused depends on its circumstances and context. Practitioners must consider whether they are conducting a forensic evaluation, which has legal implications and may ultimately be used in court, or a clinical evaluation, which is done for a therapeutic purpose. A forensic evaluation emphasizes collecting accurate and complete data to determine the level of concern for abuse. The data collected in a forensic evaluation must be preserved in an audio and video-recorded format. The results of the forensic evaluation are organized into a report that is read by attorneys, a judge, and others. However, the emphasis in a therapeutic evaluation is to assess psychological strengths and weaknesses, make a clinical diagnosis, develop a treatment plan, and lay the foundation for continuing psychotherapy. The clinician may also be interested in determining what happened to the child, but it is not so essential to distinguish facts from fantasies. Compared with the forensic evaluation, the psychotherapist does not need to keep such detailed records and ordinarily does not prepare a report for court.

In addition to distinguishing a forensic examination from a therapy meeting, a number of factors may affect the evaluation of a child who was abused or may have been abused: whether one is a pediatrician in an emergency department or a child psychiatrist in an office, whether a parent or another person is suspected of the abuse, the severity of the abuse and the victim's relationship to the perpetrator, whether physical signs of abuse are obvious or absent, the age and gender of the child, and the degree of anxiety, defensiveness, anger, or mental disorganization the child exhibits. Often, the examiner must be creative and persistent.

From the psychiatric perspective, the interview is usually the primary source of information and the physical examination is secondary. In practice, children who present with concerns about child maltreatment typically receive a physical exam in close proximity in time to the history being obtained by medical providers or social workers.

Parent Interviews

The evaluator obtains a history from the parents (separately, in most cases) and other pertinent informants, as well as from the child, if verbal. The emphasis of the interview depends on the circumstances.

Suspected Physical Abuse. The examiner should obtain a history from caregivers regarding the cause of the patient's injuries and pertinent past medical history. It is always a consideration that the caregivers are not telling the truth, so it is essential to document the history that is provided by the caregivers and compare it to the objective clinical findings of the patient to determine a level of concern for maltreatment

When the child is brought to the emergency room, a detailed and spontaneous account of the injury should be obtained promptly from parents or other caregivers. The interviewer should allow the caregiver to explain, expound, derail, or detour the storyline. Caregivers often provide histories that are implausible for the injuries noted on the child, such as multiple broken bones from rolling off of the couch, or provide no history at all to explain the injuries, such as stating that the patient went to sleep without injuries and woke-up covered in bruises.

A 30-day-old girl was transferred from a rural hospital to a university medical center because of supposed near-sudden infant death syndrome (SIDS). The child was unresponsive and required mechanical ventilation. A brain scan revealed bilateral subdural hematomas, subarachnoid hemorrhage, and hemorrhage in the parenchyma of the brain. A skeletal survey showed two posterior rib fractures. An ophthalmologist observed extensive retinal hemorrhages. After the child was admitted to the Pediatric Intensive Care Unit, the child abuse consultant interviewed the parents separately. The mother, 28 years of age, said that she had recently started a new job. The baby had been perfectly fine when she had left her in the care of her live-in boyfriend, the child's biologic father. The father, 24 years of age, said that when he checked on the baby, he found her to be not breathing, blue, and unresponsive. He ran to report this to a neighbor and then called 911. The child abuse consultant told the father the child must have been injured in some way and asked whether he had any explanation for this injury. The father said, "I shook the baby after I found her not breathing." The consultant's assessment was that the patient's injuries were most consistent with child physical abuse and notified child protective services and the local police department so they could initiate and coordinate their investigation.

Suspected Sexual Abuse. As with physical abuse cases, caregivers involved in child sexual abuse cases may not be telling the truth. This is especially true when there are concerns for child sexual abuse in the home, and the adult caregiver does not believe or support the patient or the caregiver is the possible perpetrator of the abuse. Parents may also fabricate concerns for child sexual abuse during custody disputes. It is important to remember that it is rare for a child to fabricate a report of sexual abuse.

The examiner should determine how the allegation originally arose and what subsequent statements were made; determine the emotional tone of the first disclosure (e.g., whether the disclosure arose in the context of a high level of suspicion of abuse); determine the sequence of previous examinations, the techniques used, and what was reported; try to determine whether the previous interviews were likely to have distorted the child's recollections; review, if possible, transcripts, audiotapes, and videotapes of earlier interviews; seek a history of overstimulation, prior abuse, or other traumas; and consider other stressors that could account for the child's symptoms. The examiner should also ask about exposure to other possible male and female perpetrators.

In Either Case. Whether physical or sexual abuse is involved, a pertinent psychosocial history should be collected and organized, including the following:

1. Symptoms and behavioral changes that sometimes occur in abused children.
2. Confounding variables, such as psychiatric disorders or cognitive impairment, may need to be considered.
3. Family attitude toward discipline, sex, and modesty.
4. Developmental history from birth through periods of possible trauma to the present.
5. Family history, such as earlier abuse of the parents, substance abuse by the parents, spouse abuse, and psychiatric disorder in the parents.
6. Underlying motivation and possible psychopathology of adults involved.

Collateral Information

The evaluator should consider requesting collateral information from the following, after obtaining authorizations: protective services, school personnel, other caregivers (e.g., babysitters), other family members (e.g., siblings), the pediatrician, and police reports.

Child Interview

Several structured and semistructured interview protocols have been introduced that were designed to maximize the amount of accurate information and minimize mistaken or false information provided by children. These approaches include the cognitive interview, which encourages witnesses to search their memories in various ways, such as recalling events forward and then backward. The Step-Wise Interview is a funnel approach that starts with open-ended questions and, if necessary, moves to more specific questions. The interview protocol developed at the National Institute of Child Health and Human Development (NICHD) includes a series of phases and uses detailed interview scripts. The Cornerhouse Forensic Interview Protocol is another semistructured, nondirective interview method for alleged victims of child sexual abuse.

Although these protocols may be particularly important in a forensic context, experienced clinicians endorse flexibility and consistent good-hearted behavior by the interviewer. As when seeing any patient, the evaluator must size up the situation and use techniques that are likely to help the youngster become comfortable and communicative. One victim might need a favorite object (e.g., a teddy bear or a toy truck); another might need to have a particular person included in the interview. Some children are comfortable talking; others prefer to draw pictures. An unrelated joke, a shared cookie, or the picture on the evaluator’s wall may lead to a disclosure of abuse. Important comments might be made while chatting during the break time instead of during the structured interviews.

Interview Process. The interviewer of the child who may have been abused should remember the following principles:

1. Audio and video record the interview.
2. Use a minimum number of interviews (perhaps two or three), because multiple interviews may encourage confabulation.
3. Avoid repetitive questions, either–or questions, and multiple questions, and try to avoid leading and suggestive questions.
4. Use restatement, that is, repeat the child’s account back to the child (which allows the interviewer to see whether the child is

consistent and ensures that the interviewer understands the child's report).

5. Conduct the examination without the parent present (if the child is young, consider having a family member in the room).
6. Use an examination technique that is appropriate to the child's age and developmental level.
7. Determine the child's terms for body parts and sexual acts; do not educate or provide/introduce new terms.

Interview Content. The interview should not take the form of an interrogation. The interviewer should note the child's affect while discussing these topics and be tactful in helping the child manage anxiety. Young children may not be able to report all of the relevant information. The examiner should explore the following:

1. Whether the child was told to report or not to report anything.
2. Who was the alleged perpetrator.
3. What the alleged perpetrator did.
4. Location of the incident.
5. When it started and when it ended.
6. Number of times the abuse occurred.
7. How the child was initially engaged, and how the abuse progressed over time.
8. How the alleged perpetrator induced the child to maintain secrecy.
9. Whether the child is aware of specific injuries or physical symptoms associated with the abuse.
10. Whether photography or videotaping took place.

The Investigative Interview Protocol. The usual clinical interview may need some modification for evaluating a child who may have been abused. The Investigative Interview Protocol, which was developed at the NICHD, consists of the following components: introduction, rapport building, training in episodic memory, transition to substantive issues, investigating the incidents, eliciting information that has not been mentioned by the child, collecting information about the disclosure, closing the interview, and concluding with a neutral topic.

INTRODUCTION. The interviewer introduces himself or herself and explains in general terms the purpose of the meeting. The interviewer explains the importance of telling the truth. The child can say, “I don’t understand your question” or “I don’t know.” The child can correct the interviewer if the interviewer makes a mistake.

RAPPORT BUILDING. The interviewer encourages the child to talk about the child’s activities and interests but avoids focusing on television, videos, and fantasy.

TRAINING IN EPISODIC MEMORY. Asking the child to describe two specific past events assesses the child’s memory and models the form of the interview for the child. Using nonleading, open-ended questions, the interviewer asks the child to tell everything that happened on two specific days. The interviewer trains the child to give a free narrative account, which is the pattern through the rest of the interview.

MAKING THE TRANSITION TO SUBSTANTIVE ISSUES. The interviewer asks the child to focus on the reason for the interview by making a nonleading statement such as, “I understand that something may have happened to you. Tell me everything that happened from the beginning to the end.”

INVESTIGATING THE INCIDENTS. The interviewer starts with open-ended questions, such as “Tell me everything about what happened.” If the child describes a specific incident, elicit additional information in a nonleading manner by saying, for example, “You mentioned (a particular activity). Tell me everything about that.” After posing open-ended questions, the interviewer proceeds with focused questions that relate to information already mentioned by the child. For example, “You mentioned a friend was there. What was her name? What was she doing”? This process of investigating the incidents is repeated, which are typically “the last time,” “the first time,” and “another time that you remember well.”

ELICITING INFORMATION THAT WAS NOT MENTIONED BY THE CHILD. The interviewer may proceed with direct questions only if they have tried other approaches and forensically important information is missing. For example, “When you told me about the time in the kitchen, you mentioned the man took off his pants. Did something happen to your clothes? Tell me all about that.” The interviewer may need to formulate a direct question by making use of outside information. For example, “I heard you told your mother that something happened at school. Tell me everything about that.”

OBTAINING INFORMATION ABOUT THE DISCLOSURE. The interviewer wants to know how the child came to disclose the alleged abuse. Using nonleading questions, the interviewer may say, “Tell me what happened after (the last incident)” and “Does anybody else know what happened”?

CONCLUDING THE INTERVIEW. At the end of the interview, the interviewer says, “You have told me lots of things today, and I want to thank you for helping me.” The interviewer may also say, “Is there anything else you think I should know”? and “Are there any questions you want to ask me”? The interviewer should not compliment the child with comments such as, “You were very brave to talk about those things that happened.” The interviewer may conclude with a discussion of a neutral topic, such as, “What are you going to do today after you leave here”?

Extended Forensic Interview. In typical child abuse investigations, a single-session forensic interview model is widely recommended by Child Advocacy Centers across the United States. Interview protocols, which try to define the ideal single-session interview, assume the child is a willing and able participant who can provide a significant account of the alleged abuse. If the single interview results in an equivocal account of abuse, the interview is flawed, or the details need further clarification, another interview following the same protocol may be required. These single-session interviews may not suffice for children who have difficulty relating to a stranger about emotional, traumatic information. Such children may benefit from a multisession interview process.

An extended forensic interview (EFI) model was investigated by the National Child Advocacy Center to help determine if such a protocol would be superior to the single-session model for certain children. It was investigated for use in children who either (1) did not disclose but there was other compelling evidence for abuse such as medical findings, (2) the child was unable to provide a full disclosure during the single interview, and (3) the allegations remain unresolved (founded or unfounded) after the single interview. It was determined through field trials at multiple sites that a six-session EFI was the most efficient model. The results from the field testing showed the EFI had utility and had an equal rate of court outcomes as the single interview model. Thus, for single-session forensic interviews that may have initially been unfounded, the EFI helped to resolve the ambiguity surrounding the allegation in almost two-thirds of cases. Like other interviews, the EFI should be recorded.

PHYSICAL EXAMINATION OF SUSPECTED CHILD MALTREATMENT

A comprehensive medical examination should be performed promptly by a medical professional, preferably one with special expertise in child abuse. The assessment should include a head-to-toe examination of all areas of the body including the anogenital region. Any findings should be objectively documented in the medical record and photo-documented when possible. The medical assessment should also include a comprehensive review of systems, past medical history, birth history, and family history to determine concern for underlying medical or inherited conditions that could contribute to the child's clinical presentation.

Just as there is no single psychological characteristic that is diagnostic of abuse, no single physical injury is diagnostic of abuse when considered in isolation. A medical provider must consider the overall constellation of injuries in the context of the history provided to explain them and the developmental capabilities of the patient. The following is a list of injuries that suggest child physical abuse:

- Any injury to a preambulatory infant, including bruising, oropharyngeal injuries, fractures, intracranial injuries, and abdominal injuries
- Multiorgan system injuries
- Injuries in different stages of healing
- Patterned injuries
- Injuries to areas of the body typically protected from accidental injuries, such as the nonbony locations of the torso, ears, face, neck, and upper arms
- Significant injuries that are unexplained

These injuries may range in severity from minor cutaneous injuries to life-threatening intracranial injuries. The younger the patient, the more at risk they are for severe or fatal injuries. Physical abuse also tends to be repetitive and escalate over time, so it is essential to intervene on behalf of the patient even when the injuries appear to be minor to prevent risk of future harm. These seemingly minor injuries noted on infants are called sentinel injuries, and a thorough medical evaluation is needed promptly to assess for occult injuries or those injuries that are not apparent on physical exam.

A 3-month-old female presents with a lingual frenulum tear. The patient's father reports that he was feeding the patient a bottle earlier that morning and noticed the blood coming from her mouth. He denies knowledge of any accidental injuries that could explain the torn frenulum. The patient is discharged home without further evaluation or report to child protective services. The patient returns 2 weeks later in full cardiopulmonary arrest. Head computed tomography (CT) reveals large bilateral subdural hemorrhages and extensive cerebral edema. A full nonaccidental trauma evaluation is completed at that time revealing multiple fractures in different stages of healing and extensive retinal hemorrhages. The patient dies 2 days later as a result of the intracranial injuries. A report to child protective services and law enforcement is made to investigate the case and to ensure the safety of the other children in the home. Upon interview with investigators, the patient's father admits that the initial oral injury was the result of force-feeding when he became frustrated with the patient refusing the bottle. He also confesses that over the following 2 weeks, he had started shaking the baby out of frustration, including on the morning of presentation when the patient presented in cardiopulmonary arrest.

Physical Findings of Sexual Abuse

The vast majority of victims of child sexual abuse will have no signs of anogenital injury, especially in the nonacute setting, which represents the largest portion of sexual abuse exams due to the well-known delay in reporting sexual abuse and resiliency of the anogenital region. It is of the utmost importance for medical providers to recognize an absence of injury does not invalidate the child's report of abuse. A number of anogenital exam findings have been described in child abuse literature. These physical findings are carefully categorized by Adams et al. with regard to their association with sexual abuse. Findings strongly suggestive of sexual abuse include certain signs of acute anogenital trauma and residual (healing) injuries. Other signs of sexual abuse include sexually transmitted infections (STI) (e.g., Neisseria gonorrhea, chlamydia trachomatis, syphilis, human immunodeficiency virus [HIV]), semen identified in forensic specimens from a child’s body, and pregnancy.

SPECIAL TESTS AND LABORATORY EXAMINATION

Child Physical Abuse Evaluation

It is essential that medical providers have a standard approach to the evaluation of suspected child physical abuse in order to identify and treat all injuries as promptly as possible and to reduce provider biases regarding which patients should undergo the evaluation. The American Academy of Pediatrics provides evidence-based recommendations regarding this evaluation based on the patient’s age and clinical presentation. All young infants and older patients with concern for neurotrauma should receive a head CT to assess for intracranial injury. Advanced neuroimaging via MRI of the brain and spine and dilated fundoscopic exam to assess for retinal hemorrhages are indicated for patients whose initial head imaging or exam is concerning for intracranial injury. All children under the age of 2 years or young children with developmental disabilities or neurologic compromise should receive a skeletal survey to assess for occult skeletal trauma. All children up to the age of 5 years should have labs obtained to assess for occult abdominal trauma with a follow-up abdominal CT for elevations in these labs or those children with concerns on physical exam for abdominal trauma. A follow-up skeletal survey approximately 2 weeks after the initial study is typically recommended to assess for occult skeletal trauma not identified in the initial study, and siblings of children with concern for child physical abuse should undergo an assessment for injuries due to the inherent risks associated with their sibling’s injury.

Child Sexual Abuse Evaluation

State laws vary regarding the timing of evidence collection following acute sexual assault and range from 24 hours to 120 hours with special consideration to the age of victim and their disclosures. The acute sexual assault evaluation should be performed by a trained medical professional, including sexual assault nurse examiners (SANE nurses). The acute sexual assault evaluation includes a thorough head-to-toe examination of the victim and anogenital exam with photo-documentation, typically assisted by medical equipment such as a colposcope. Evidence collection kits vary depending on the state, and they typically include several swabs designated for specific areas of the body. Chain of custody must be maintained during and following the evidence collection process. Depending on the age of the victim and type of contact disclosed by the victim, STI, HIV, and pregnancy testing and prophylaxis may be indicated. In prepubertal patients, the anogenital exam should be external. Anesthesia should only be used if there is concern for internal injuries that cannot be visualized on external exam and may need surgical intervention. If a patient is noncooperative with an exam, the patient should be allowed to refuse the exam with accommodations made to collect as much evidence and information as possible to the extent of the individual patient’s comfort level, as the intent is not to retraumatize a victim with the process intended to protect them. When a patient presents with concerns for sexual abuse outside of the acute timeframe established by their state laws, a forensic medical exam should be performed as promptly as possible and includes a head-to-toe exam, including anogenital exam with photo documentation, and any appropriate STI testing indicated by disclosures of sexual contact made by the victim. Most victims will have no findings on exam. If an injury or STI is identified, follow-up exam and testing may be indicated to assess for healing and cure of disease. Should the patient need HIV postexposure prophylaxis, follow-up with infectious disease experts is commonly indicated, as the regimen is prolonged and monitoring for side effects and compliance is needed.

DIFFERENTIAL DIAGNOSIS

Physical Findings

Pediatricians who examine children who may have been abused must consider a wide range of congenital, infectious, toxic, and traumatic conditions in their differential diagnoses. Most importantly, accidental trauma must be considered as a cause for the patient's clinical presentation, emphasizing the importance of a thorough history to determine if a plausible accidental mechanism of injury has been provided to explain the patient's presentation. For bony injuries, congenital and metabolic bone diseases must be considered, such as rickets and osteogenesis imperfecta. For intracranial bleeds and bruising, underlying coagulopathies, such as von Willebrand and hemophilia, must be considered. Testing for rare underlying genetic conditions such as GA1 (Glutaric Aciduria Type 1) may also be considered. These exceedingly rare conditions are far less likely to explain the patient's clinical presentation than trauma but are important to exclude, especially when considering the legal implications of the diagnosis of child physical abuse and abusive head trauma. Medical providers must also recognize mimics of child maltreatment. Most commonly, cutaneous findings such as slate grey nevi, formerly known as Mongolian spots, may be confused for bruising, or conditions such as staph scalded skin syndrome may be confused for inflicted thermal burns. This emphasizes the importance of a thorough exam, medical history, and broad differential diagnosis when evaluating patients with concerns about child maltreatment.

Psychiatric Considerations

The psychiatric evaluation of youngsters who may have been abused may involve assessing the patient's credibility. Although children generally tell the truth when they talk about abuse, sometimes children make false denials (saying that they were not abused, when actually they were) or false allegations (saying they were abused, when actually they were not).

Possible Explanations of Denials of Abuse. A false denial or retraction may occur for several reasons. The child may have been pressured by the perpetrator or family members to recant the allegation. The pressure may consist of bribery, mockery, or threats of injury. The child may be protecting a parent or other family member, even without external coercion (i.e., the child takes on this responsibility through role reversal). The child might be frightened or distressed by the investigation and may decide to withdraw participation. For instance, an interviewer could induce a false denial by asking overly challenging questions. A child may be inhibited by shame or guilt; the child may mistakenly assume responsibility for what happened. Finally, the child may have accommodated the abuse, consciously or unconsciously, instead of objecting to it.

Possible Explanations of Allegations of Abuse. A false allegation of abuse, although rare, may occur for several reasons.

ALLEGATIONS ORIGINATING WITH ADULTS. Sometimes a false allegation arises in the mind of a parent or another adult and is imposed on the child.

Parental Misinterpretation and Suggestion. The parent may have misinterpreted an innocent remark, a neutral piece of behavior, or a benign physical condition as evidence of abuse and may have induced the child to endorse this interpretation. This happens in child custody disputes, as well as in other settings.

Parental Delusion. The parent and child may share a *folie à deux*, or the child may simply give in and agree with a delusional parent.

Parental Indoctrination. A parent may have fabricated the story and may have induced the child to collude in presenting it to the authorities.

Interviewer Suggestion. Previous interviewers may have asked leading or suggestive questions. An interviewer who believes abuse occurred may unwittingly shape a child’s responses until the child validates the interviewer’s assumptions.

Group Contagion. In epidemic hysteria, people modify what they have heard to meet their emotional needs. Thus, rumors may become more convincing as they are retold.

UNCONSCIOUS OR NONPURPOSEFUL MENTAL MECHANISMS IN THE CHILD

Fantasy. A younger child may confuse fantasy with reality.

Delusion. Although rare, older children and adolescents may experience delusions about sexual activities in the context of a psychotic illness.

Misinterpretation. Sometimes children misunderstand what happened and later report it inaccurately.

Miscommunication. The child may misunderstand an adult’s question, and the adult may later misinterpret or take the child’s statement out of context.

Confabulation. Children fill gaps in their memory with whatever information makes sense to them and others at the time.

CONSCIOUS OR PURPOSEFUL MENTAL MECHANISMS IN THE CHILD

Pathologic Lying. Also called *pseudologia fantastica*, this refers to children who tell stories with such zeal that they may become convinced of their truth.

Innocent Lying. Children may make false statements because that seems to be the best way to handle the situation. Developmentally, this is more likely to happen with younger children.

Deliberate Lying. Children may choose to avoid or distort the truth for some personal advantage. This happens more with older children and adolescents.

PERPETRATOR SUBSTITUTION. The child may have been sexually abused and may exhibit symptoms consistent with abuse but may identify the wrong person as the perpetrator, resulting in a false allegation. The child may do this to protect the actual offender, or the child may displace the memories and accompanying affects onto another individual.

The parents of a 9-year-old boy were divorced, and he spent the summers visiting his father. When he returned from visitation, the mother noticed injuries: A healing cut on the boy's right ear and a collection of superficial, vertical scratches from his umbilicus to his pubis. She had the boy disrobe and took digital images of his abdomen. Several days later, the mother took the boy and the images to the boy's regular psychotherapy appointment. She told the therapist that she was concerned the boy had been physically and sexually abused. The therapist interviewed the boy individually, who explained in a matter-of-fact manner his ear was cut accidentally when he had his hair cut and his abdomen was scratched when he went body surfing at a beach. The therapist concluded the boy had not been abused and reassured the mother. The therapist did not notify child protective services.

SPECIAL ISSUES

Neonatal Abstinence Syndrome

Neonatal abstinence syndrome (NAS) is another example of the potential maltreatment of a fetus. Newborns who have been exposed to maternal drug use can undergo withdrawal after their birth. This can be due to illicit substance use by the mother or iatrogenic addiction (e.g., maternal treatment of a pain condition with opioid medication). The newborn's symptoms are reflected by the type of substance exposure. Symptoms of NAS encompass high-pitched crying, jitteriness, tremors, convulsions, sweating, fever, mottling, excessive sucking/rooting, poor feeding, vomiting, and diarrhea. A toxicology workup can be done in addition to thorough history taking, although the mother may be reluctant to disclose substance use. It is estimated that 3% of the 4.1 million women of childbearing age who abuse drugs will continue to use drugs during their pregnancy. From 2000 to 2009, the rate of infants born with NAS rose from 1.2 to 3.4 per 1,000 hospital births per year. Maternal substance use has both medical and developmental concerns for the newborn as well as legal, health, and economic impact for the mother.

Medical Child Abuse

Medical child abuse, or factitious disorder imposed on another, was formerly known as Munchausen Syndrome by Proxy. This occurs when a child receives unnecessary and harmful or potentially harmful medical care at the instigation of a caretaker. The perpetrator is usually the mother and often has a narcissistic, histrionic, or borderline personality disorder and thrives on the center-stage position they assume. Medical child abuse is perpetrated by fabrication (falsifying or lying about symptoms) or induction (an act that induces or causes illness, such as poisoning). Frequently, this form of maltreatment has been ongoing for an extended period of time before it is recognized by medical providers. In families in which medical child abuse is found, more than one child may be involved consecutively or simultaneously. In extreme cases, children can be killed, seriously injured, terrorized, and experience severe adverse effects from medications or other treatments.

A 12-year-old boy was referred for psychiatric evaluation because of oppositional behavior. The psychiatrist interviewed the youngster and his mother and learned they had an enmeshed relationship and their affect toward each other ranged from loving to embittered. An older brother had finished school and worked full-time but still lived at home. The patient's medical history was unusual, in that the mother claimed the boy had a long history of fainting spells that defied diagnosis. The mother, but no medical professional, had witnessed these spells. The mother was a secretary in a medical practice consisting of neurologists and neurosurgeons, and she had induced her employers to repeatedly work up the child with physical examinations, laboratory testing, and electroencephalograms (EEGs). In fact, one of the neurosurgeons did burr holes to place EEG electrodes directly on the surface of the child's brain. No neurologic diagnosis was made, and factitious disorder by proxy was suspected. The child's neurologist said that no further testing would be done and encouraged psychiatric treatment. The psychiatrist helped the patient and his mother achieve age-appropriate psychological separation, and there were no more fainting spells.

Sex Trafficking

Human trafficking is referred to as modern-day slavery. When specifically addressing sex trafficking of minors, it is called commercial sexual exploitation of children (CSEC). CSEC includes a large number of crimes against minors, including prostitution, survival sex, internet-based exploitation, and child pornography. While trafficking may involve the physical transfer and movement of the child, these crimes also frequently occur in or close to the home. LGBTQ youth are especially vulnerable to human trafficking related to survival sex. It is estimated that almost 300,000 American youth are at risk of becoming victims of the commercial sex trade. The average age at which children enter the trade is between 12 and 14 years for girls and 11 and 13 years for boys.

The use of violence, drugs, coercion, and other methods of manipulation are commonplace among the traffickers. If a youth in the sex trade becomes addicted to drugs, the trafficker can use the addiction as a means of control. The National Center for Missing and Exploited Children created the Child Sex Trafficking Team in 2001 due to the need for specialized services in this area. Training of police officers, detectives, and emergency room staff can help to identify youth involved in trafficking.

Cultural Variations

Parenting practices in United States society vary widely. They are partly determined by the cultural heritage of the family, and the cultural context of the alleged act should be considered in evaluating suspected neglect and abuse, including assessing whether the discipline of a child is abusive. Some unusual bruises result from cultural healing practices. Coining or cupping—techniques used in many European and Asian countries—are used to draw toxins from the skin surface and to remove feverish mucus from the system. In coining, a coin is heated in hot oil and massaged against the skin. In cupping, a heated glass cup is used to create suction against the skin. Skin scraping and moxibustion may also mimic physical abuse.

Female Genital Mutilation

Commonly called *female circumcision*, female genital mutilation is performed on an estimated 2 million girls worldwide each year. It is practiced across diverse socioeconomic classes and different ethnic, cultural, and religious groups. Commonly, girls are circumcised between 4 and 10 years of age, but the procedure may be performed on infants, may be postponed until just before marriage, or may be done after the birth of the first child. In some cultures, it is considered part of a ceremonial induction into adult society. Female circumcision, in the mildest form, consists of clitoridectomy—the anatomical equivalent of penile amputation. In its most severe form, total infibulation involves the removal of the clitoris and labia minora plus an incision of the labia majora to create raw surfaces that are then stitched together. This practice has been widely criticized, including opposition by the World Health Organization (WHO) and other major healthcare groups. In the United States, female circumcision is generally considered child abuse. Efforts have been made to educate immigrant communities on the health risks and legal liabilities of the practice.

Male circumcision is not usually considered child abuse, but it does have its critics. Although this simple procedure is one of the most common operations performed worldwide, complications may result, such as hemorrhage and infection. Advocates of male circumcision cite research that it is somewhat protective against HIV infection, other sexually transmitted diseases (STDs), and penile cancer.

Religious Practices

Certain religious practices may be construed as forms of abuse if they are likely to harm children. For example, adherents of Christian Science believe illness should be treated with prayer instead of medical care. As a result, children may sustain chronic illness, disability, or death. This is an area of flux: some state laws and some courts have allowed parental discretion to withhold conventional medical treatment for legitimate religious reasons, whereas others have not.

Teamwork

Child maltreatment cases require a multidisciplinary team (MDT) of medical providers, mental health professionals, law enforcement, and child protective services. Team participation provides consultation, professional support, collaboration, and coordination for continued safety, health, hope, and positive social interactions within the grim realities of case discovery.

In cases of suspected child abuse and neglect, the physician should make an assessment based on a history, a physical examination, and indicated medical studies; should request appropriate surgical, medical, and psychiatric consultations; should diagnose the suspected maltreatment; and should report these concerns promptly to their local child protective services and investigative agencies. Child protective services will determine the safety plan for the patient, which can range from periodic monitoring to out-of-home foster care placement. Members of the MDT may also refer the patient to appropriate mental health services and other medical or developmental therapies that are needed. These activities, taken together, should be a coordinated team effort.

REPORTING LAWS AND PRACTICES

In every state, the District of Columbia, and all areas controlled by the federal government (e.g., military bases), physicians are required to report child abuse to the police or the appropriate child protection agency. The typical statute says a physician who has “reason to believe” or “reason to suspect” child abuse is required to report it, and as long as the physician is “acting in good faith,” they are protected from criminal or civil litigation related to their report. In most states, other designated individuals, such as nurses, psychologists, school officials, police officers, district attorneys, and providers of child daycare and foster care, are also required to report child abuse.

Under-Reporting

A significant factor in reporting errors is the failure of clinicians and other mandated reporters to consider child maltreatment in the differential diagnosis of child injuries. Frequently, mandated reporters fail to pursue the diagnosis of child maltreatment because of diverse cultural and socioeconomic factors. Lack of experience or education in dealing with the issues and the judicial quagmire that result when reporting many child maltreatment cases are also important deterrents. Under-reporting of abuse also occurs when well-meaning but misguided therapists attempt to monitor and treat their abusive clients without involving the proper authorities. Some therapists think they are above the law; others simply have no idea what the law says about reporting child abuse. The most obvious negative consequence of under-reporting is the abused child and siblings are left in dangerous circumstances and are vulnerable to further abuse.

TREATMENT

Child

The first part of the treatment of child abuse and neglect is ensuring the child's safety and well-being. The child may need to be removed from an abusive or neglectful family to ensure protection but emotionally may feel even more vulnerable in an unfamiliar setting. Because of the high risk of psychiatric symptoms in abused and neglected children, a psychiatric evaluation is in order. Along with providing specific treatments for any mental disorders present, the therapist may have to deal with the immediate situation and the long-term implications of the abuse or neglect. Psychotherapeutic issues to be addressed with an abused child include dealing with the child's fears, anxieties, and self-esteem; building a trusting relationship with the therapist in which the child is not exploited or betrayed; and gaining a perspective, over time, of the factors that contributed to the child's victimization. These issues may be addressed in individual or group psychotherapy or both.

The intensity and duration of treatment depend on the severity of the symptoms. If an abused child has no post-trauma symptoms, it probably is adequate to provide a few counseling sessions that include psychoeducation and anticipatory strategies for recognizing problems in the future. For a child who has been minimally assaulted (e.g., a relatively innocuous single episode of fondling by a stranger), it may be adaptive for the child to forget the incident and move on.

Abused children who have mild to moderate psychological symptoms have been treated with a form of cognitive-behavioral treatment that has four components: psychoeducation, anxiety management, exposure, and cognitive therapy. *Psychoeducation* includes teaching the child and the caregivers about the nature of maltreatment, offenders, and strategies to avoid future abuse. *Anxiety management* includes teaching children how to use relaxation and other coping strategies to reduce fearful responses to memories and reminders of abuse. *Exposure* refers to talking, drawing, or writing about the abuse experiences. It is usually helpful to have the child share the painful, ugly details of the abusive experiences with the therapist. This may take time and may require therapeutic creativity and flexibility. Although some children may be able to relate the bad things that happened through ordinary conversation, other children may need a variety of play, artistic, or other projective techniques. Some children avoid painful memories through conscious (suppression) and unconscious (repression) mechanisms. *Cognitive therapy* techniques are used to help the child replace cognitive distortions about the event or negative feelings about himself or herself.

Abused children and adolescents with very severe symptoms (such as complex PTSD or DESNOS) require extensive treatment. These victims of abuse have been treated with a multimodal approach, including specific psychotherapies such as dialectical behavioral treatment (DBT), psychotropic medication, and perhaps hospitalization or residential treatment.

Parents

Developing an intervention plan for an abused or neglected child requires assessing the parents' behavior and functioning in several respects. The psychopathology and the prognosis for achieving adequate parenting skills for the abusive parent and the one who allowed the abuse to occur should be assessed. The evaluator should determine whether the parent's dysfunction is confined to this child or involves other children and whether the dysfunction is short-term or long-term (reflecting a lifelong pattern). The evaluator should assess the parent's willingness to participate in the intervention plan and the availability of personnel and physical resources to implement the various intervention strategies. One must consider the risk of additional physical or sexual abuse if the child remains in the home.

On the basis of the information obtained, several options can be selected to improve the parent's functioning: (1) eliminating or diminishing the social or environmental stresses; (2) lessening the adverse psychological effects of the social factors on the parents; (3) reducing the demands on the parent to a level that is within their capacity through day care utilization or provision of a housekeeper or babysitter; (4) providing emotional support, encouragement, sympathy, stimulation, instruction in parental care, and training in planning for, assessing, and meeting the needs of the child (supportive casework); and (5) resolving or diminishing the parent's inner psychic conflicts (psychotherapy). Some clinics provide group counseling for nonoffending parents.

Intrafamilial Child Sexual Abuse

The first step in the treatment of child sexual abuse perpetrated by a family member is its disclosure, which means overcoming the denial and collusion of the family members and fear and blame felt by the victim. The perpetrator may leave the home, or the child may be removed by protective services, so the inappropriate sexual contact is not likely to recur. In those circumstances in which it is possible for the family to reunite or cases where the perpetrator never left the home, family therapy is useful to re-establish the group as a functioning unit and to develop healthier role definitions for each member. Specific individuals may need therapy regarding unique aspects of their involvement. If the child victim does not feel heard/validated, they can develop a defensive posture and become reclusive, distrusting, and feel unloved in addition to a host of other negative psychological sequelae. The nonoffending caregiver may also develop a sense of guilt that erodes their self-confidence and interferes with family relationships. They may feel unworthy of being a parent which could result in further parenting deficits or overcompensation for perceived faults. When the participants experience severe psychopathology, treatment must be directed toward the underlying illness. While the perpetrators of abuse are learning to develop internal restraints and appropriate ways to gratify their needs, the external control provided by therapy helps to prevent further abusive behavior. At times, legal agencies are involved to help enforce external controls.

PREVENTION

Primary Prevention

Measures that reduce the occurrence of child abuse—*primary prevention*—can be classified on the basis of their audience or intended recipient. *Universal interventions targeted to the general public* include such strategies as a poster campaign admonishing parents not to shake their babies, required videos about the dangers of shaking babies in newborn nurseries, warnings about consuming alcohol and drugs during pregnancy, and raising public awareness through the media, civic associations, church groups, and parent–teacher organizations. To some extent, children can be taught to avoid or resist being abused. However, ultimately, the adults in a community must create safety and security for their children. *Selective interventions are targeted to populations at high risk of child abuse*. In general, for example, there should be less child abuse in a community with fewer single parents (use of birth control and greater involvement of fathers in families) and greater financial security (local jobs and social support systems). In-home programs aimed at primary prevention of child maltreatment have proven effective, but they are expensive and time intensive. These programs typically involve in-home intervention throughout pregnancy and the first 2 to 3 years after birth to provide education regarding child development and childcare to at-risk mothers. Governmental regulations may be helpful, such as the licensing of daycare centers and regulations concerning the operation of residential programs for children. Finally, one can sometimes *intervene with specific persons or families that are at high risk*. Counseling and parenting classes may deter parents from being abusive, although this has not been substantiated. Treatment of parents who abuse substances and services for families with a history of spouse abuse may be helpful. Individuals can help friends and relatives who appear stressed and who are struggling with their parenting responsibilities.

Secondary Prevention

Early detection and prompt intervention, when child abuse occurs, comprise secondary prevention. Education of the medical profession, members of allied health fields, and all who come in contact with children aid early detection. Emergency department personnel, pediatricians, teachers, and counselors have many opportunities to identify abused children. It is hard to know how best to use the limited resources available to child protection agencies. Some critics think too much effort is put into assessing the millions of referred cases, with relatively little attention given to interventions. For instance, an agency might put many hours into collecting information, might arrive at a decision about whether a case is founded, and might engage in expensive legal measures but might offer little or no aid to help a marginal family function better. In general, child abuse and neglect prevention and treatment programs should try (1) to prevent the separation of parents and children if possible, (2) to prevent the placement of children in institutions, (3) to help the parents attain self-care status, and (4) to help the family become self-sufficient. As a last resort, to prevent further abuse and neglect, children may have to be removed from families who are unwilling or unable to profit from the treatment program.

FUTURE DIRECTIONS

When “The Battered Child Syndrome” was published in 1962, the concept of child maltreatment had a rather narrow scope, that is, child physical abuse. During the half-century since that landmark event, the understanding of child maltreatment has grown to include medical, educational, and other forms of neglect; child sexual abuse and the exploitation of children in child pornography; verbal or psychological abuse; observing family violence; the neglect of children in orphanages; and the abuse of adolescents in institutional settings such as residential treatment centers. In the future, the scope of child maltreatment will continue to enlarge and there will be descriptive research regarding less common forms of child abuse, such as medical child abuse; parental alienation, in which a child of divorced parents loses their relationship with one of the parents; and abusive corporal punishment in schools. The phenomenon and consequences of psychological abuse will be studied more thoroughly. The sequelae of child maltreatment have been studied through retrospective and limited prospective research. The study of psychological abuse will be of particular interest to mental health professionals, especially since both *child psychological abuse* and *spouse or partner psychological abuse* were added to DSM-5. The definitions of these mental conditions refer to "nonaccidental verbal or symbolic acts … that result, or have reasonable potential to result, in significant psychological harm." The research project that addressed ACEs took child psychological abuse into consideration, in that two of the ACEs directly reflected child psychological abuse: verbal abuse by a parent or other adult in the household and feeling threatened with physical injury. The ACE researchers found that an emotionally abusive family environment accentuated negative outcomes later in life. Authors from many countries have noted that causing parental alienation in a child—inducing the child to fear or hate a parent, resulting in psychological parentectomy—is a particularly pernicious contemporary form of child psychological abuse.

In the future, there will be systematic studies from an international perspective of child maltreatment, providing information on, for example, child soldiers in Africa, child suicide bombers in the Middle East, child labor in Asia, the trafficking of children and adolescents for prostitution, and the deaths of large numbers of children due to lack of food and medical care.

A major focus of research in the next few years will pertain to resilience. This research will address the question of why some maltreated children manifest very serious psychological and behavioral symptoms, whereas other abused children are free of symptoms both in the short and the long term. At least part of the answer to this question will involve genetic predispositions, specific life experiences, and the interaction between genetic and environmental factors.

In the more distant future, there will be a greater understanding of the chain of events by which trauma damages young victims, starting with specific genetic and other congenital vulnerabilities and predispositions that interact with maltreatment and other forms of adverse childhood experiences. These predispositions and experiences lead to changes in cognitive and affective pathways in the brain, which will be observed through neuroimaging. Alterations in the brain will be shown to cause changes in behavioral and emotional functioning, which result in dysfunctional relationships and additional symptoms and disorders. Although society's goal is the primary prevention of child maltreatment, that objective will not be fully achieved. The research outlined here will inform the secondary prevention of the consequences of child abuse, as well as the biopsychosocial treatment of abused children.

FURTHER READINGS

Adams JA, Farst KJ, Kellogg ND. Interpretation of medical findings in suspected child sexual abuse: an update for 2018. *J Pediatr Adolesc Gynecol*. 2018;31(3):225–231.

*American Academy of Child and Adolescent Psychiatry. Practice parameters for the forensic evaluation of children and adolescents who may have been physically or sexually abused. *J Am Acad Child Adolesc Psychiatry*. 1997;36:423–442.

American Professional Society on the Abuse of Children (APSAC). *Forensic Interviewing in Cases of Suspected Child Abuse*. American Professional Society on the Abuse of Children; 2012.

Anderson J, Ellefson J, Lashley J, et al. The corner house forensic interview protocol: RATAC. *T M Cooley J Prac & Clinial L*. 2010;12:193–331.

Bernet W. False statements and the differential diagnosis of abuse allegations. *J Am Acad Child Adolesc Psychiatry*. 1993;32:903–910.

Brassard MR, Hart SN, Hardy DB. *Psychological and Emotional Abuse of Children.* Springer; 1991.

Byrd AL, Manuck SB. MAOA, childhood maltreatment, and antisocial behavior: meta-analysis of a gene-environment interaction. *Biol Psychiatry*. 2014;75(1):9–17.

Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, Division of Violence Prevention. Child Abuse and Neglect: Risk and Protective Factors. 2021.

Deblinger E, Mannarino AP, Cohen JA, Steer RA. A follow-up study of a multi-site, randomized, controlled trial for children with sexual abuse–related PTSD symptoms. *J Am Acad Child Adolesc Psychiatry*. 2006;45:1474–1484.

Edwards VJ, Holden GW, Felitti VJ, Anda RF. Relationship between multiple forms of childhood maltreatment and adult mental health in community respondents: results from the adverse childhood experiences study. *Am J Psychiatry*. 2003;160:1453–1460.

Finkelhor D. Prevention of sexual abuse through educational programs directed toward children. *Pediatrics*. 2007;120:640–645.

Flaherty EG, Stirling J; American Academy of Pediatrics, Committee on Child Abuse and Neglect. Clinical report—The pediatrician's role in child maltreatment prevention. *Pediatrics*. 2010;126:833–841.

Friedrich WN, Fisher JL, Dittner CA, et al. Child sexual behavior inventory: normative, psychiatric, and sexual abuse comparisons. *Child Maltreat*. 2001;6:37–49.

Gershoff ET. *Report on Physical Punishment in the United States: What Research Tells Us About Its Effects on Children*. Center for Effective Discipline; 2008.

Green AH. Factors contributing to false allegations of child sexual abuse in custody disputes. *Child Youth Serv*. 1991;2:177–189.

Grossi LM, Lee AF, Schuler A, Ryan JL, Prentky RA. Sexualized behaviors in cohorts of children in the child welfare system. *Child Abuse Negl*. 2016;52:49–61.

*Helfer ME, Kempe RS, Krugman RD. *The Battered Child*. 5th ed. University of Chicago Press; 1999.

*Jenny C. *Child Abuse and Neglect: Diagnosis, Treatment and Evidence*. Saunders; 2010.

Kellogg N; American Academy of Pediatrics, Committee on Child Abuse and Neglect. The evaluation of sexual abuse in children. *Pediatrics*. 2005;116:506–512.

*Lamb ME, Orbach Y, Hershkowitz I, Esplin PW, Horowitz D. A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: a review of research using the NICHD Investigative Interview Protocol. *Child Abuse Negl*. 2007;31:1201–1231.

McElvaney R, Greene S, Hogan D. To tell or not to tell? Factors influencing young people's informal disclosures of child sexual abuse. *J Interpers Violence*. 2014;29(5):928–947.

Narang SK, Fingarson A., Lukefahr J; Council on Child Abuse and Neglect. Abusive head trauma in infants and children. *Pediatrics*. 2020;145(4):e20200203.

Nemeroff CB, Seligman F. The pervasive and persistent neurobiological and clinical aftermath of child abuse and neglect. *J Clin Psychiatry*. 2013;74(10):999–1001.

Poole DA, Lamb ME. *Investigative Interviews of Children: A Guide for Helping Professionals*. American Psychological Association; 1998.

*Roesler TA, Jenny C. *Medical Child Abuse: Beyond Munchausen Syndrome by Proxy*. American Academy of Pediatrics; 2009.

Sege RD, Siegel BS; Council on Child Abuse and Neglect and Committee on Psychological Aspects of Child and Family Health. Effective discipline to raise healthy children. *Pediatrics*. 2018;142(6):e20183112.

Sheets LK, Leach ME, Koszewski IJ, Lessmeier AM, Nugent M, Simpson P. Sentinel injuries in infants evaluated for child physical abuse. *Pediatrics*. 2013;131(4):701–707.

U.S. Department of Health & Human Services, Administration for Children and Families, Children’s Bureau. Child Maltreat 2018. 2020.

U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention. Confronting Commercial Sexual Exploitation and Sex Trafficking of Minors in the United States. 2013.

van der Kolk BA, Roth S, Pelcovitz D, Sunday S, Spinazzola J. Disorders of extreme stress: the empirical foundation of a complex adaptation to trauma. *J Trauma Stress*. 2005;18(5):389–399.

Winters GM, Jeglic EL, Kaylor LE. Validation of the sexual grooming model of child sexual abusers. *J Child Sex Abus*. 2020;29(7):855–875.

54.3 Children's Reaction to Illness and Hospitalization

SUSAN BECKWITT TURKEL, M.D., JULIENNE JACOBSON, M.D., AND MARYLAND PAO, M.D.

HISTORY

Hospitalization of young children was a rare event before the mid-19th century, as most children were cared for at home, and hospitals were considered unsafe for them. When young children were admitted to early general hospitals, their mothers typically remained with them to assist in providing their care. This was routine and met with little opposition. Some hospitals would only admit children on the condition their mothers were present. From the 1850s through the 1880s, hospitals encouraged regular visits from parents and siblings. These visits were viewed as important for the well-being of the sick children. Outbreaks of infectious diseases might temporarily restrict visits, but they were reinstated as soon as possible. Any restrictions on visits would be lifted for seriously ill or dying children. The goal was to create, at least in part, a "homeliness" where children could feel safe and comfortable.

The transformation of the hospital into a headquarters for scientific medicine started in the period before the First World War and had profound implications on medical care, research, and education. The exclusion of parents from the hospital began in the late 19th century. Hospitals gradually restricted parental access and removed toys, pictures, and pets from the wards. These changes reflected changing ideas about the spread of infectious diseases, the growing professionalism of nursing, the development of the specialty of pediatrics, and funding changes for healthcare. Visitors, including parents and siblings, were restricted or excluded from the children's wards because of the belief that the emotional upset the visits created would be detrimental to the child's recovery. Fear of infection led to banning personal contact between parents and their children, and glass panels were provided so parents could see their children on the ward. The practice of enabling mothers to remain with their hospitalized children never completely disappeared, but mainstream practitioners during this period were unconvinced of the importance of having the mother in the hospital with her sick child.

Nursing was highly regimented during this period, and children and parents often perceived the staff as uncaring. Pediatric doctors and nurses cared about their patients' emotional well-being but were taught to be disciplined, professional, and scientific and not to show affection or emotion. Consequently, hospitalized children were traumatized by their illness, by separation from their parents and families, and by the lack of warmth in the care they received.

Little changed until after the Second World War. By that time, the prevailing practices were in conflict with emerging psychoanalytic and developmental theories of the social and emotional growth of children and new thinking about the child's need for parental interaction and emotional care. John Bowlby, deputy director of the Tavistock Clinic in England, established a separation research unit in 1948, which led to his formulation of attachment theory. Bowlby warned that maternal deprivation or separation or denial of a warm, intimate, continuous relationship with the mother, or mother substitute could have far-reaching effects on the child's character and mental development. Other psychoanalytic studies used hospitalization as an example of the first early separation from home. Psychoanalysts were aware of the lack of integration between the practice of child guidance, psychoanalysis, and pediatrics. Observation of the profound emotional reactions of children to physical illness and the likelihood that mothers might be unable to make objective observations when they were emotionally upset by the stresses of their child's condition was noted. The need to prevent the negative psychological impact of hospitalization on children was soon accepted by most psychiatrists and some pediatricians, but this did not yet change hospital practice.

James Robertson, a student working with John Bowlby, translated Bowlby's theories into film, starting with "A 2-year-old Goes to Hospital" in 1953. This film emphasized the detrimental effects of an admission that separated a young child from his family, friends, and mother's care. It demonstrated the phases of protest, despair, and detachment (or denial) through which small children pass when isolated from their mothers for a length of time and illustrated how children who had been left in the hospital exhibited greater behavioral problems when they returned home. His film was widely shown to both professional and lay groups and was instrumental in accelerating change in hospital practice.

Following the efforts of Robertson and others, public outcry led to formation of a British parliamentary committee, chaired by Sir Harry Platt, which investigated the hospital care of children. The committee found that children were expected to conform to ward routines, lie quietly in their cots, and were not allowed to play; their parents were allowed to visit only during restricted visiting hours. The committee's report, "The Welfare of Sick Children in Hospital", published in Britain in 1959, led to the revolution in pediatric hospital care in Britain and other western countries. The Platt Report recommended unrestricted visiting and accommodation for parents, especially for children less than 5 years old. The committee sought to minimize potential damage caused by the separation of children and mothers, and to encourage appropriate stimulation for the child in the hospital.

The recommendations of the Platt report took time to be implemented. Gradually the views of pediatricians and pediatric nurses changed from those prevailing since before the Second World War and became less discordant with the views of child psychiatrists. Pediatricians and child psychiatrists now agree about the significant impact of separation, illness, and hospitalization on children and their families.

The length of hospitalization and the reasons children and adolescents require hospitalization had changed considerably. Children used to be admitted to the hospital for much longer periods than they are today. Changes in morbidity and treatment modalities, along with alternatives to inpatient care such as day surgery, home intravenous antibiotics, ventilators, and advances to maintain technology-dependent children at home, have led to shorter and less frequent hospitalizations. The average length of stay has been reduced by almost half and same-day discharges have almost tripled in the past 25 years. The leading causes of hospitalization recently have been asthma, gastrointestinal disorders, pneumonia, and seizures; inpatient surgeries have declined as outpatient procedures have increased. Hospitalizations lasting more than 7 days now are likely to be for psychiatric reasons, chemotherapy, and treatment of malignancies, or to address complications of procedures or prior treatment.

As survival rates have increased, at least 90% of children with chronic conditions now live to young adulthood. Inpatient hospitalization for children with more common conditions has shifted to outpatient care, leaving proportionally more complicated, chronically ill children in inpatient settings.

Hospitalized children now are significantly medically complex, often with comorbid psychiatric disorders. This small proportion of children accounts for most health care expenditures: 20% of children with complex chronic illness and children with special health care needs account for 80% of costs. The prototypical medically complex child has a multisystem disorder, is probably technologically dependent, has frequent inpatient admissions, require multiple medications, multiple subspecialists, and optimal care coordination across inpatient and outpatient settings. Care is usually provided at tertiary care centers. These children have three times longer lengths of stay, over 10 times greater billing charges, significantly higher gastrostomy and tracheostomy placements, intensive resource use, and 15 times inpatient mortality. Seriously and chronically ill children are "growing up in the hospital," and the lives of these children and their families may differ considerably from their peers.

CONTEMPORARY PEDIATRIC HOSPITALS

Contemporary pediatric hospitals and wards are staffed and structured differently than adult units, and children's hospitals typically include features not found in most adult settings. The physical layout of a children's hospital or pediatric ward is usually designed to facilitate the child's primary caregiver staying with the child. Sleeping arrangements, showers, laundry, and other facilities are made available to increase individual patient and parent comfort and facilitate extended stays with children. Private rooms or conference rooms exist to allow doctors and parents to confer in privacy.

Accommodations are now made to encourage visits even from large or extended families, placing the experiences of children and their families at the heart of pediatric healthcare. To decrease stress in waiting areas, toys and activities are available in brightly colored, child-friendly spaces. In an effort to provide maximum continuity of care, every child may be assigned a primary nurse who will continue to be the child's primary nurse in subsequent hospitalizations. While most large medical centers have hotels nearby for family members, children's hospitals often provide additional settings such as Ronald MacDonald House, where entire families may stay in small, subsidized apartments. Social workers, parent advocates, groups for parents and for siblings, and educational groups provide support to families. Disease-specific summer camps for patients and siblings and other recreational group activities for sick children are characteristically managed through children's hospitals or disease-specific charitable organizations. Parents, initially focused on potentially life-threatening physical concerns, often express deep gratitude for these services later and are grateful for additional support for their child's psychosocial development. Psychosocial standards have been developed for children with cancer and their families to ensure many of the services described below are integrated routinely into care.

Child life programs have become standard at children's hospitals. These developmental specialists are available in most pediatric inpatient units and in some outpatient pediatric settings to help minimize disruption in a child's development by hospitalization, surgery, or other potentially frightening or distressing medical experiences. Child life specialists provide recreational activities, often in designated places where children can engage in therapeutic play. Child life specialists help provide information and preparation, and can rehearse useful coping strategies for medical procedures with children. Preoperative surgical preparation programs have been shown to be effective in reducing anxiety and improve children's abilities to cope with the stress of a surgical procedure more effectively. The goal is to help children understand their feelings about their experiences in the hospital and facilitate better adjustment.

Emphasis has been placed on minimizing physically painful experiences. Regular assessment of pain in children using age-appropriate measures has been widely implemented, and the use of nonpharmacologic and pharmacologic techniques adapted for young patients are more standardized and better utilized. These changes have been important in further decreasing the stress of hospitalization for children and adolescents.

Education is a major part of each child's life, and hospital schools and teachers are present in most pediatric settings. Some children's hospitals have arrangements with the local school system to provide teachers, and others employ their own. Either way, collaboration with the child's school is important in helping maintain academic progress as much as possible.

Principles of behavior management are typically incorporated into the culture of pediatric units, where staff can play an important role in modeling and encouraging age-appropriate behavioral management by parents. The absence of limits and consequences for misbehavior can be harmful for children, leading to acting out and difficulties in regulation of emotion and behavior. Although mental health professionals have limited opportunities to monitor behavior directly and cannot substitute for parents and staff on the unit, frequent consultation with psychologists, social workers, and child psychiatrists can be helpful to the care team, to patients, and to families.

Helping children, adolescents, and their families adapt to hospitalization should involve the child, parents, family, and hospital staff. In current family-focused models of care, a child's care is planned around the whole family with parents considered integral to the treatment process. Family-centered care has proven to be valuable in helping families and children cope with the stress, fear, and altered parenting roles that accompany a child's illness and hospitalization. While the creation of a nurturing and developmentally sensitive hospital environment is primarily the responsibility of staff, the assistance of family members is essential in establishing structure and maintaining stability for sick children. "Anticipatory guidance" about parenting struggles in and out of the hospital can be helpful. Parents of children with life-threatening illnesses often feel guilty and need support to maintain usual parenting practices. Parents of chronically ill children should be included in discussions with hospital staff about realistic rules; this leads to increased adherence to daily schedules and consistent behavioral expectations.

Inpatient pediatric care requires constant review by nursing and medical staff as well as by patients and families to remain relevant and effective. While past initiatives have created the current environment of pediatric inpatient care, hospital staff should remain open to adopting additional practices as further studies identify better alternatives.

COLLABORATIVE CARE, INTEGRATED CARE, AND THE MEDICAL HOME

Psychiatric participation in pediatric care has been extended from the hospital to the outpatient setting, and pediatric primary care has become an important site for providing mental health services. Approximately one in four U.S. children suffer from a chronic health problem, and one in five from a psychiatric disorder. Integrating psychiatric into pediatric care can improve overall medical outcomes, increase healthcare utilization, and decrease healthcare costs for medically ill children and adolescents by shortening hospital length of stay, reducing rates of readmission, and improving adherence. Integration of child psychiatric services in pediatric primary care assists in addressing structural and attitudinal barriers, shortages of behavioral health specialists, and geographical difficulties in locating and participating in psychiatric care.

Outpatient integrated care systems are intended to expand psychiatric access. They range from care coordination and collaborative care to colocation to embedded child psychiatric services to fully integrated medical-psychiatric practices, which can make it much easier for families by providing early psychiatric referral, increasing access to care, and improving their engagement in mental health services.

The medical home concept originated in the American Academy of Pediatrics (AAP) in 1968 to provide for children with special health care needs. The AAP defines the medical home model of care as comprehensive, high-quality care that is accessible, family-oriented, coordinated, continuous, compassionate, and culturally effective, with the goal of coordinating medical and mental health care. About half of U.S. children receive health care in a medical home setting.

The integration of behavioral health interventions into pediatric medical care has been a national priority. The 2010 Affordable Care Act promoted the medical home as an efficient way of providing accessible, continuous, comprehensive, family-centered, coordinated care to children and adolescents. School-based health centers are another way a medical home can be structured to provide the same services in settings that are easiest for students and families to access. Most models of integrated and collaborative care have relied on state-level initiatives or additional funding sources to establish them, usually in large public or academic health care systems, which may be difficult to sustain. The use of telehealth technology can extend service delivery and has become an essential tool in increasing the number and variety of services.

The wrap-around service delivery model refers to coordination between multiple agencies providing mental health, child welfare, juvenile justice, developmental support, addiction, and primary health care required for children and adolescents with complex needs.

IMPACT OF ILLNESS AND HOSPITALIZATION ON CHILDREN

Sick children are not a homogeneous group. They come to the hospital having very different life experiences, irrespective of their medical condition. The presence of an acute, life-threatening, or chronic physical illness by itself is insufficient to produce psychiatric problems, but it may make the child more vulnerable to emotional difficulties and may exacerbate any pre-existing psychiatric illness or behavior problems. How a child will react to illness and hospitalization is more a reflection of their age and understanding of their situation, their parents' response, and the nature of the illness, and its treatment rather than the greater severity of the medical condition itself.

Preschool children are hospitalized more than older children, and those younger than 7 years are at greater risk for more frequent and potentially lasting emotional and behavioral problems after hospitalization than older children. The risk of later aggressive and hyperactive behaviors can usually be predicted by the mother's emotional state and the child's premorbid behavior. Baseline internalizing or externalizing behaviors may be the best predictor of increased behavior problems 3 to 6 months after hospitalization, and medically ill children with greater premorbid emotional and behavioral problems have more psychiatric disorders following their illness. When illness is experienced as an insurmountable obstacle or the cause of rejection by loved ones, the child may react with extreme emotional distress, and suicidal ideation and behavior may occur. Overall, 20% of children with chronic conditions have behavioral and emotional symptoms, double the rate of the general population.

Conditions that are less visible, have a remitting-relapsing course, or uncertain prognosis may be associated with greater emotional problems. Males with chronic conditions appear to have more emotional sequelae than females. Children with brain damage have poorer overall adaptation to illness. Adopted children and their parents also seem to have more problems with hospitalization. On the other hand, the personality characteristics of sociability, flexibility, and physical attractiveness appear to be protective, and may mitigate against more severe emotional problems. Children from families with clear boundaries, the capacity to balance competing family needs, the ability to maintain flexibility, and the tendency to ascribe positive meanings to life events, fare best.

Chronic and serious physical illnesses involve recurrent separations and frequently upsetting events, which can be distressing to both parents and children. Surgery, medical procedures, adverse medication side effects, and diagnostic tests can be frightening, confusing, unscheduled, disruptive to sleep, and painful. The burdens of chronic illness are pervasive and may produce chronically stressful situations, resulting in maladaptive behavior, emotional symptoms, and added financial burdens, all contributing to conflict, guilt, and anxiety in the family.

Medical illness and hospitalization can pose serious threats to the developmental process and impact the physical and emotional environment, disciplinary context, schooling, family life, and neurobiology of the chronically ill child, and thus profoundly affect cognitive, emotional, social, and sexual developmental trajectories. Invasive procedures; separation from family, friends, and school, lost opportunities to develop interests, and a diminished sense of self-mastery further complicate development. Environmental restrictions, such as isolation for infection control and limited activities available in the hospital, lead to fewer opportunities to discover personal strengths and weaknesses. Chronically ill children often have difficulty finding something to be “good at” and have trouble feeling confident and competent. Delayed growth and puberty are common in chronically ill children and adolescents, especially when malnutrition, prolonged corticosteroid treatment, or chronic inflammatory processes are prominent. These growth delays typically have additional negative psychological implications.

One week after discharge from the hospital, a child often demonstrates increased anxiety about separation, may experience sleep problems, and may be more aggressive. These behavioral changes should quickly resolve as the child adjusts to the return home, but this may not always happen. Repeated admissions of children coming from dysfunctional or disadvantaged families are very likely to be associated with emotional and behavioral problems following hospitalization, but all children may be at risk. The most difficult period of adjustment to a hospitalization may be near the beginning, when new and painful experiences and separation are most distressing. Prolonged periods of hospitalization bring the risk of developing maladaptive behaviors. In a meta-analysis of studies of children’s behavior after hospitalization, children hospitalized for 2 to 3 days showed more behavioral problems than those hospitalized for longer or shorter periods, and overall, briefer hospitalizations of older children were the least likely to have detrimental effects.

Post-traumatic stress disorder (PTSD) has been described following critical illness and surgery. Serious accidents and hospitalizations following acute physical trauma may be associated with PTSD, while the risk of PTSD in children and parents after simple planned surgery is almost nonexistent. Children and parents after newly diagnosed cancer and children and parents following solid organ transplantation may also be at risk for PTSD. Parental distress and characteristics of the family environment are more correlated with the development of PTSD symptoms in sick children than family composition, course of hospitalization, or the extent of a child’s accidental injuries. In contrast, commitment, help, support, and the open expression of feelings in the family can facilitate a child’s better adjustment to newly diagnosed cancer and associated intensive medical treatment. Sometimes the treatment of life-threatening medical illness may be experienced as repeated trauma, which might not impact during or immediately after treatment, but rather appear as long-term effects on affect modulation and interpersonal relationships.

Attempts have been made to mitigate the negative consequences of hospitalization, not only in the way hospitals are structured and function but also in the provision of preventative educational or behavioral interventions. In a meta-analysis of a variety of interventions, psychoeducation was generally not effective with younger children, and any benefits from the experimental interventions disappeared by 4 weeks after discharge. Other programs which focus on increasing parents’ knowledge and understanding of the range of behaviors and emotions that young children typically display during and after hospitalization and that support direct parent participation in their child's physical and emotional care are more effective and help reduce the mother's emotional stress, depression, and anxiety, reduce the child's negative behavioral symptoms and increase the child's adaptability up to 12 months after hospitalization.

When psychiatric and behavioral symptoms are severe or persistent, a thorough psychiatric assessment, based on multiple brief examinations of the child and information gathered from parents, other family members, hospital staff, and teachers, is needed. A child's biologic vulnerability to depression and anxiety may be inferred from a history of previous psychiatric symptoms or psychiatric treatment, or from a family history of a mood, anxiety disorder, or other psychiatric disorder. Psychiatric syndromes of adjustment disorder, major depression, anxiety, delirium, and psychosis can be diagnosed in medically ill children, and require appropriate intervention.

Behavioral and cognitive approaches, such as distraction, guided imagery, and relaxation for reducing procedural distress, are well-established in the pediatric hospital setting. Supportive individual and family therapy and cognitive behavioral and family therapy may be useful in reducing anxiety and in providing patients and families with adaptive strategies to use in coping with illness and its treatment. Psychopharmacotherapy may be indicated based on the severity and duration of psychiatric symptoms.

IMPACT OF CHILD'S ILLNESS ON FAMILIES

The family of a seriously or chronically ill child is as affected by the illness and frequent hospitalizations as is the child, and the social impact of the illness is felt throughout the family system. Family systems are dynamic and can be sources of risk or protection for a child with a serious medical problem. Chronic conditions present the entire family with an additional set of tasks, responsibilities, and worries. There is personal impact on a parent, financial impact on the family, and implications for family and social relationships. Parents must deal with the day-to-day management of the illness, with the increased financial burden associated with medical bills and time lost from work, and with the uncertainty about treatment and their child's prognosis. This can place additional strain on a marriage and can significantly affect the parents' relationships with other children. Relatives, teachers, and friends may be available to provide support, but they also may be negatively impacted by the child's illness. Social activities usually become limited, decreasing opportunities for interpersonal support.

When illness presents early in the child's life, parents face additional challenges and feel uncertain and helpless. They must make difficult treatment choices, confront the risk of premature death, and come to terms with their lost expectations, fearing the loss of the child's functional abilities and independence.

Family cohesion, open communication, and the ability to adapt to conflict are indicative of healthy family function and psychological adjustment in families with chronically ill children. Nine aspects of a resilient family process have been identified: (1) balancing the chronic condition with other family needs, (2) maintaining clear family boundaries, (3) developing communication competence, (4) attributing positive meanings to the situation, (5) maintaining family flexibility, (6) maintaining a commitment to the family as a unit, (7) engaging in active coping efforts, (8) maintaining social integration, and (9) developing collaborative relationships with professionals.

Parental temperament and coping skills strongly influence how their child copes. Children take their cues for coping and adaptation to illness primarily from their parents, especially their mothers. Parents who are able to isolate their own fears and anxieties during crises are better able to enhance their child's understanding of the illness and facilitate self-care and independence in their child. Parental anxiety can have a negative effect, and children with very anxious mothers often have less distress during procedures when their mothers are not present.

Parents tend to remain at the hospital during the initial diagnosis and early therapeutic interventions, and usually return when symptoms worsen or complications arise during hospitalization. Once the first crisis passes, the mother is most likely to take on the day-to-day care of the child. She is often the parent who remains in the hospital with the child, accompanies the child to physician visits, provides support for procedures, and acts as advocate for the child with health care team. Research on fathers' reactions is consistent with studies on mothers, although fathers are less well-studied. Fathers traditionally take on additional financial burdens associated with a child's illness and endure separation from their spouse and ill child during hospitalizations. Mothers also work and lose pay when off work with their sick child, adding to the economic impact on the family. Men often seek to be more involved in caring for their sick child. Fathers give direct care to the sick child, take turns staying in the hospital, and take on the primary care for the other children remaining at home. Although fathers may not be as constant a parental presence at the hospital as mothers, treatment teams must remember to include fathers who also struggle to adjust and cope with challenging issues of their child's condition and care.

Parents may respond differently to different conditions. For example, with traumatic brain injury in their children, fathers are more likely to use denial and report greater injury-related distress, while mothers are more likely to demonstrate acceptance and emotion-focused coping. Mothers may also be better able to adjust over time, while fathers' adjustment and marital satisfaction may not improve.

Parents need to deal with feelings of helplessness due to their inability to protect their child from harm. Differences in coping style may contribute to emerging family dysfunction. While at risk for maladjustment, there is considerable individual variation, and many do well. Some parents experience marital distress and impaired sexual relationships, but other parents of children with cancer report their marriages are satisfactory and their partners are supportive, and higher divorce rates have not been found. While marital discord is associated with increased psychological problems in chronically ill children, a mutually supportive marriage can benefit the family's ability to cope with stress. Feeling uncertain about the outcome is to be expected, but it can lead to a fragile sense of optimism and hope. Parental hope may be a coping strategy, and parents may avoid information and people that threaten their hope. Open, empathetic communication and a sense of connectedness strengthen the relationship, while physical and emotional distance is the most common challenge.

Siblings of a seriously, chronically ill child are in a very difficult position. They often are asked to take increasingly adult roles to "help out" while receiving less attention and support from their parents, adapting to personal and family changes while dealing with their own strong emotions. The sibling may love and worry about their sick brother or sister and, at the same time, resent them, leading to feelings of guilt. Sibling rivalry and jealousy increase when the focus of both parents and the family's resources are directed toward the ill child with less time and attention for the healthy siblings. Siblings of chronically or severely ill children often feel isolated, and their family life is interrupted. They are more likely to develop depression, anxiety, and a negative self-image and have fewer quality peer activities. Younger siblings may be more resilient, and life-threatening illnesses may exacerbate the siblings' risk.

Sibling adjustment typically parallels the family's adjustment and finding ways to help the mother feel confident in managing her child's illness is integral to her ability to provide the optimal home environment for both the sick child and the healthy siblings. Parents of sick children rate the siblings' quality of life better than the siblings themselves do, so sibling adjustment needs to be assessed individually.

Siblings may be asked to donate hematopoietic stem cells for transplantation. Donor siblings often feel they "had no choice" about donating. Both donors and recipients may have post-traumatic stress symptoms following the transplant, and nondonor siblings may report more anxiety and lower self-esteem. Teachers have observed donors to have both more adaptive coping and more school problems and attention-seeking behaviors than nondonor siblings. Donors themselves report the psychological aspects of donating outweigh any physical consequences, whether their sibling lives or dies. If the recipient dies, the donor sibling frequently feels guilty, and may use spirituality and religion as primary coping strategies. Siblings want to be included in decisions, be provided information, be given choices, have help sharing their feelings, be provided opportunities to meet with peers, and experience a supportive hospital environment when they visit.

Siblings of a dying child benefit from preparation and support as the time of anticipated death nears. They should be included in discussions of what changes to expect and estimates of survival, and again be asked about their own fears and experiences. Increased anxiety and persistent feelings of regret related to a sibling's death have been described in surviving siblings who were not present during the illness or at the time of death. Bereaved siblings who received support from clinicians and were present and involved at the time of their sibling's illness and death often report positive memories, and most perceived physicians and staff as warm, kind, and honest. Surviving siblings who were involved in the process describe their sibling's strong willpower, good mood, and stamina in the face of adversity, enhanced family cohesion and strengthening of family bonds, and feelings of personal growth. A minority of surviving siblings describe negative experiences, and report increased awareness of vulnerability, feeling emptiness and loneliness, unacknowledged and personally stagnant, and see increased family conflict. Families who actively seek information and services, work to solve problems, express their feelings, and are able to balance personal, family, and illness needs show better adaptation to a child's illness and hospitalization.

Maintaining supportive relationships with people in the community can also be an important protective factor. Informal support networks provide emotional and informational support and sometimes tangible aid. Illness-specific resources and formal support groups should be made available to families. When coping difficulties are recognized or psychological symptoms develop, mental health services may be needed and referrals made.

When families are given a new devastating diagnosis, such as childhood cancer, optimally, the family views the illness as a challenge that is controllable and changeable, making long-term difficulties in psychological adjustment less likely. Alternatively, the family may respond as though the event is a matter of chance beyond their control, and exhibit denial, detachment, intellectualization, or substance abuse instead.

Health care providers are very influential and play a powerful role in the lives of families with a sick child. Parents look to them for essential medical information and guidance. Pediatricians need to be aware that anything said by physicians, whether well-reasoned or not, can have an enormous impact on parents and children, and effects may be immediate or long-lasting. When families of pediatric cancer patients were surveyed about information they received from medical teams about what to expect from the effects of cancer and its treatment on a child's psychological health, 86% of parents and 71% of siblings reported they did not receive enough or any information. They felt unable to ask professionals about current and future psychosocial issues and related topics. Half the parents and a quarter of the siblings later wished they had communicated their feelings more to other members of the family, noting limited communication in their own family. Reports of family communication differ and family members may have different requirements, necessitating assessment of the need for support and information for each member.

Medical communication delivered in a clear, thoughtful, and sensitive manner can promote competent family functioning and resilience, despite the presence of a life-threatening illness in a child. However, when health information is not thought through or clearly given, it may be misconstrued, misinterpreted, or misunderstood, and may lead to a constellation of abnormal behaviors and altered parent-child interactions that are referred to as the vulnerable child syndrome (VCS), first described by Green and Solnit in 1964.

The cardinal features of VCS include:

1. pathologic separation difficulties, such as poor sleep in infants, recurrent unexplained abdominal pain, or school phobia in older children
2. infantilization with overprotective, overly indulgent, and overly solicitous parent, and an overly dependent, disobedient, irritable, argumentative, and uncooperative child
3. bodily over concerns, when the child has many physical complaints and the parent (usually the mother) has difficulty differentiating between the child's discomfort and their own distress, and cannot distinguish mild, self-limited symptoms from potentially serious illness
4. Separation anxiety, delaying starting school, and school underachievement.

Predisposing factors to VCS include a history of infertility, miscarriage, illness during pregnancy; prematurity; a life-threatening illness; or a child with a chronic medical illness, especially if there was hospitalization during the first year of life. VCS is linked to more frequent emergency room visits and inpatient hospitalizations. Depending on how their parents perceive the event, even older children seen in an emergency department for minor head trauma or hospitalized for gastroenteritis or asthma may be at risk for VCS.

The prevalence of VCS in a community sample was estimated at 10% in 1995 but may be higher today with improved survival of premature children and children with serious chronic illnesses. The degree of parental perception of vulnerability does not necessarily correlate with the severity of neonatal problems, birth weight, or current health. Rather, higher levels of maternal anxiety at neonatal discharge better predict parental perception of child's vulnerability and are associated with worse developmental outcomes. Affected parents often have difficulty with setting age-appropriate limits, and demonstrate permissiveness alternating with excessive attempts to control their child's activities. They are also likely to seek higher levels of medical service and have more dissatisfaction with the medical care provided.

A child's perception of their own vulnerability does not necessarily correlate with their parents' perception. When children rate themselves as vulnerable, most parents agree; when parents rate their children as vulnerable, only about half the children agree.

Effective communication is at the heart of the preventing and managing VCS. If the child stays healthy and communication between parents and pediatricians is effective, VCS does not necessarily persist into school age. Persistent parental fear that a child remains vulnerable can be elicited by asking parents, "has your child ever been seriously ill"? VCS should be suspected when parents then respond by relating traumatic accounts of their child's perceived near-death experiences. The physician can decrease the risk of unintentionally exacerbating parental anxiety and any belief in their child's vulnerability when discussing diagnosis, prognosis, and anticipated treatment with a parent with excessive anxiety or guilt by being aware of the parents' and their own state of mind. The parents' frightening experience should be acknowledged, parents provided a more accurate assessment of their child's health, and informed when time-limited medical problems are resolved and will not recur. The importance of encouraging appropriate autonomy in the child should be emphasized with discussion of potential adverse effects of excessive parental concern on a child's development and behavior. Reassurance and regular follow-up help to distinguish normal development and behavior from problems needing psychiatric consultation.

A CHILD'S RESPONSE TO ILLNESS AND HOSPITALIZATION AT DIFFERENT AGES

A child's behavioral response and concept of illness and its causes follow a sequence of developmental stages related to their perceptual ability and cognitive development (Table 54.3–1). This section is limited to a description of a child's response to illness and hospitalization at different ages. For a complete description of child development, see the chapters on Child Development and Adolescent Development.

Infants (Birth to 18 Months)

Serious illness and hospitalization separate infants from their parents and set them in a strange, frightening, unpredictable environment. This can provoke anxiety and confusion putting them at risk for disruption in the development of basic trust. Infants younger than 7 months rarely protest separation from their mothers and seem to accept substitute caretakers in the hospital. They may then appear briefly but dramatically upset upon return home. Infants older than 7 months protest initially, appear uncooperative and subdued in the hospital, and then require a readjustment period upon return home. The more infants are provided with contact and comfort from parents when distressed by illness or hospitalization, the better basic trust can be protected. Infants subjected to unexpected medical procedures may have difficulty forming attachments and become anxious. Repeated exposure to pain and the absence of social stimulation in neonates may affect later pain sensitivity, brain volume, neuroendocrine functioning, behavior, and cognition. If parents are unable to hold and soothe their babies because of the nature of an injury or medical care, the babies are at risk for apathy and withdrawal, the attachment process may be disrupted, and separation anxiety may develop. Infants may then refuse to interact with parents, or become clingy, demanding, and difficult to soothe.

It is important to maintain consistent nursing staff and infant routines, preferably with parents or familiar caregivers in the baby's room. If they cannot stay, attempts should focus on maintaining regular times for parental visits. Parents should be encouraged not to leave only after infants are asleep as this adds to an infant's anxiety and distrust. Normal stranger and separation anxiety may become exacerbated by frequent staff changes, so attention to providing familiar affectionate staff on a regular schedule may be helpful. Both parents and babies benefit from staff support, and parents may need staff to model appropriate interactions and interventions with their baby.

Table 54.3–1.
Summary of Impact of Illness and Hospitalization

Age	Child's Response	Possible Effects of Hospitalization	Possible Effects of Chronic Illness
Infancy 0–18 months	▸ Unable to understand ▸ Fear abandonment ▸ Perceive pain	▸ Anxiety and confusion with separation	▸ Decrease access to the environment ▸ Parents' guilt, anger, and grief may interfere with attachment ▸ Difficult to establish trust ▸ Greater helplessness
Toddler 18 months to 3 years	▸ Fear abandonment ▸ Little capacity for reasoning or understanding experience	▸ Regression ▸ Negativism ▸ Restricted activities ▸ Important to maintain routines ▸ Transitional objects help	▸ Delay motor and language development ▸ Parental reluctance to set limits ▸ Problems with bladder and bowel function
Preschool 3–5 years	▸ Fear mutilation ▸ Magical thinking ▸ Attribute illness and pain to punishment for bad behavior	▸ Fear abandonment ▸ Fear mutilation ▸ Hospital play beneficial	▸ Parental overprotection ▸ Regression ▸ Limited initiative ▸ Compromised social function
School Age 6–12 years	▸ Illness is punishment for bad behavior or due to own ideas of cause ▸ Understand simple bodily structure and function ▸ Understand the purpose of treatment	▸ Fear pain ▸ Benefit from clear expectations and concrete explanations ▸ Important maintain roles and rules	▸ Alienation from peers ▸ Fewer social interactions ▸ Limited activities ▸ Parental overprotection
Adolescent 13–18 years	▸ Abstract thinking ▸ Understand body processes and implications of illness ▸ Understand cause and effect ▸ Hypothetical reasoning ▸ Comprehend mind–body issues	▸ Fear alienation from peers ▸ Value privacy ▸ Appearance important ▸ Want to participate in treatment decisions	▸ Concerned about appearing different from peers ▸ Concerned about medication side effects ▸ Show risk-taking behavior ▸ Treatment nonadherence ▸ Immaturity and overly dependent on parents

Toddlers (18 Months to 3 Years)

Toddlers' primary developmental goal is to establish early autonomy and separation-individuation from their parents. These tasks require a secure base. They are difficult to accomplish without having first established basic trust, which may not have occurred effectively if the toddler was ill or hospitalized as an infant. Toddlers fear abandonment when separated from their parents or primary caregivers. Like infants, they benefit from maintenance of routines and stable relationships, which support early independence and efforts at developing competency. Transitional objects may be particularly helpful in this age group, so bringing toys, blankets, or other familiar objects from home to the hospital or doctor's office may help calm an anxious toddler.

The requirements of medical care can be infantilizing and carry high risk for regression, increased dependence, and negativism. It is easy for parents to become overly protective or permissive during a toddler's illness and behavior problems may result. If these parenting patterns persist after recovery, tantrums, and oppositional behavior increase, which leads to parental anger and threat of rejection, followed by guilt and more permissiveness, and a vicious cycle of behavior problems can develop. Providing anticipatory guidance to parents during a toddler's hospitalization and follow-up behavioral management and support after discharge can be helpful.

Preschool Age (3 to 5 Years)

Preschool-age children seem to be more upset by hospitalization than older children or infants, but they are the most likely to benefit and respond positively when their environment is supportive. When the child is prepared for hospitalization, parental visits are consistent, and parents participate in their child's care during hospitalization, the potential emotionally traumatic effects of hospitalization need not occur.

The preschool age is the period of "magical thinking" when children typically attribute illness and pain to punishment, and hospitalization and separation to rejection. Limited understanding and concrete thinking contribute to fears and misconceptions, leading to feelings of guilt and distress. Children this age have a very rudimentary concept of the body and its parts. They fear mutilation, and all parts of the body are considered equally essential. This limited understanding may cause procedures such as phlebotomy and surgery to be especially terrifying. Children are often unable to articulate their fears. They may not understand adult attempts to describe bodily structures and functions, illness effects, or treatment plans. Children of this age often benefit from preparation for procedures through medical play. When the material is presented in a manner congruent with their cognitive function, comprehension is improved, and fear is reduced.

School Age Children (6 to 12 Years)

School-age children are typically active and focused. Developmentally, they are working to become secure in their own abilities and accomplishments. During this period, they acquire academic and social skills and begin to internalize parental and societal values. School-age children, like younger children, often believe that their illness is a punishment for misbehavior. Their strong respect for authority and rules often leads school-age children to feel illness or injury is the consequence of not following rules properly. Because their understanding of anatomy and physiology is more complete, children this age have less fear of mutilation than younger children. However, because their thinking remains concrete, they cannot comprehend hypothetical possibilities or complicated explanations. Adult attempts to explain illness, injury, or anticipated medical procedures may be confusing, especially when explanations are complex or abstract. Medical explanations can be easily misunderstood or misinterpreted, leading to confusion and increased fear, which the child may be reluctant or unable to disclose. Drawings, photographs, and clear explanations at an appropriate cognitive level will better help children understand and master their experiences. It is important to minimize physical pain as much as possible. Difficulty coping with pain, and anxiety about pain, may compromise a child's ability to understand and participate in care. Potentially painful or frightening experiences may be less stressful with adequate preparation and explanation. School-age children should be encouraged to express their fears, ask questions, and talk about their feelings to engage them in the process and help increase their sense of mastery of their illness.

It is important to maintain routines and relationships for school-age children during illness or hospitalization. Children should be supported to meet their responsibilities for school, maintain connections with peers, and participate in as many of their usual activities as possible. These older children have an increasing awareness of the personal significance and implications of their illness. The ability to adapt to new, challenging experiences increases with age and expanding cognitive function. Coping with illness and hospitalization may increase children's resilience and teach skills useful in other settings.

Adolescents (13 to 18 Years)

Adolescence is a complex process of transition from childhood to adulthood and is a challenging time for most. Adolescent's ability to understand their illness, their body and its functions, and the causes, nature, and treatment of their illness expands as cognitive skills increase. They are more skeptical of simplistic explanations or superficial reassurance. As formal operations develop in individuals of normal intelligence, adolescents become able to hypothesize cause and effect, understand hypothetical suppositions and multifactorial explanations, and participate in decision-making and assent/consent for treatment.

Adolescents are usually aware of the implications of their illness. They may be concerned about its effects on their relationships, fertility and sexuality, and vocational possibilities. They fear the loss of independence and control, and of their sense of physical well-being and integrity. Younger adolescents especially do not want to be different from their peers. They are exquisitely sensitive to changes in their appearance and dread permanent disfigurement. They understand the permanence of death and fear it. Issues of puberty, autonomy, personal identity, sexuality, education, and vocational choice are challenging to adolescents. They are more difficult for an adolescent who is also coping with an acute or chronic illness. Illness can be associated with medical setbacks, impaired physical or mental abilities, forced dependence, and a perceived poor prognosis. When an illness is severe or disfiguring, an adolescent's identity, self-image, and ego development can be negatively affected. Identity formation is especially challenging in the hospital setting, and adolescents may come to rely on their illness or disability as their identity, rather than setting goals or aspirations beyond the hospital environment. Biologic, psychologic, and social development are highly interrelated. Delays in one area, such as body image or sexual development, lead to consequences in other domains. Delayed puberty in boys may lead to low self-esteem and apparent immaturity, which in turn can foster dependence on the family and eventual reduced educational qualifications and increased underemployment or unemployment.

Adolescents with a chronic illness are generally well-adjusted and perceive their illness as less severe than their physicians assume they do. Many who live with chronic illness or disabilities are well-adjusted socially and emotionally. When adolescents perceive their illness to be severe or untreatable, they are more likely to have symptoms of depression and low self-esteem. Those adolescents with the greatest uncertainty of their outcome often have more difficulty accepting and discussing their condition and may avoid disclosure to peers. In a study of chronically ill adolescents with a variety of disorders compared to matched controls, there were no differences in levels of anxiety or self-esteem between the two groups, and both held a similarly positive outlook on life. This finding may be related to the ability of adolescents to deny the worst aspects and consequences of their illness in order to function effectively, or it may be related to their experience coping with multiple minor stresses resulting in greater resilience and more successful coping strategies than their healthy peers. Adherence to treatment, rules, and authority may be problematic for some medically ill adolescents, and they may be more likely to demonstrate risky behaviors than their healthy peers. Medically ill adolescents require the same anticipatory guidance and close supportive follow-up as any adolescent at risk.

It is important that adolescents participate in their own care and treatment decisions, although in most cases, diagnosis and treatment information is still directed to the parents. Adolescents highly value their privacy and confidentiality. Hospitalized adolescents prefer to have medical information provided directly to them by health care professionals, but often still chose to have their parents present for support. Adolescents would like to have discussions of diagnosis and prognosis as soon as possible after admission, even when their condition may be terminal. Adolescents often recognize the imminence of their death and want to discuss their feelings. They may request information that the parents are reluctant to disclose to them and may be ready to discontinue treatment when their parents are not. In these situations, it is important that the health care team is ready to support both the adolescent and parents and to provide help to bridge their differences. Advance care planning is encouraged by the Institute of Medicine and the AAP, and developmentally appropriate tools for discussing end-of-life care are available for adolescents and young adults.

RESPONSE TO SPECIFIC ILLNESSES

Asthma

Asthma, a reversible reactive airway disease triggered by immunologic and nonimmunologic factors, is the most prevalent chronic illness of childhood, affecting more than 1 in 20 school-aged children. While asthma can present at any age, 75% of patients present in the first 3 years of life. The sensation of chest tightness, respiratory distress, and anxiety may be particularly frightening to young children, while adolescents with new-onset asthma seem to adapt more easily than anticipated. There is a high comorbidity in asthma with anxiety and depressive disorders, although this risk may be less when asthma is well-controlled. Anxiety plays a greater role when a child's symptoms are mild, and it is more difficult to quantify symptoms when asthma is severe. Trouble breathing typically triggers anxiety, and proinflammatory cytokine release is found in both asthma and depression. Maximizing the child's ability to perceive symptoms and comply with treatment can reduce the morbidity and mortality of asthma.

Rheumatologic Disorders

Rheumatologic disorders in pediatric patients are typically more severe clinically than the same disorders in adults and may have significant psychiatric and medical consequences. These conditions are treated with potent immunosuppressive medications, steroids, and immune modulators that themselves may be associated with psychiatric side effects. Psychiatric symptoms may also be the result of the autoimmune process or related underlying vasculitis, or the psychiatric disorder may be unrelated and coincident. In determining the long-term effects of psychiatric complications of pediatric rheumatologic disorders, parental functioning may play a role, and parental distress and maternal depression have been associated with greater behavior problems in the child.

Juvenile idiopathic arthritis (JIA) can lead to social isolation, chronic pain, and skeletal deformities, which may lead to depression, anxiety, low self-esteem, and lower quality of life. In studies of children and adolescents with JIA, those with less severe physical disabilities were often more affected psychologically, and more emotional and social problems and pain are noted in children whose parents feel stressed, pessimistic, and guilty.

Systemic lupus erythematosus (SLE) in children and adolescents is associated with anxiety, depression, and cognitive dysfunction. Pain, fatigue, and depression may be related to the effects of the chronic illness, needed medications, or SLE itself. Central nervous system involvement, with both neurologic and psychiatric symptoms, occurs in more than half of children and adolescents with SLE, usually early in the course of the disease. Neurocognitive impairment is common, and complaints of memory and concentration problems are frequent even when not documented on neuropsychologic evaluation.

Cystic Fibrosis

Cystic fibrosis (CF) is a lethal inherited disease caused by mutations in the CF transmembrane conductance regulator (*CFTR*) gene. Survival of patients with CF has been considerably extended by improved therapy of pulmonary infections, improved nutritional management, and an organized system of centralized care with physicians experienced in caring for the complications of CF. The goal of treatment has been slowly shifting from symptomatic management to targeting the underlying CFTR defect to halt disease progression and prevent the occurrence of CF complications.

The psychological and social functioning of patients with CF is similar to that of healthy peers until their disease becomes severe and their physical and social activities become increasingly limited. Patients with CF are at increased risk for psychiatric problems, such as depression, and their ability to cope with their disease and comply with treatment is important for their quality of life. The psychological and social problems associated with CF may be intensified during adolescence, when more problems with independence, sexuality, relationships, and noncompliance emerge. It is difficult to develop independence when required daily care is complex and may require assistance, and it is difficult to develop self-esteem when chronic productive cough, physical limitations, and peer relationships are disrupted by recurrent hospitalizations. Boys and girls seem to respond differently: girls seem to have a better body image and to be more distressed coughing in public, and for boys, short stature and pubertal delay may have a greater impact on their quality of life than the issue of shortened survival.

CF is a common indication for lung transplantation. Most transplants are living donor lobar transplants and are usually performed in adolescence. The mean survival is greater than ten years post-transplant, which can allow recipients to develop long-term goals and plan for independence.

Inflammatory Bowel Disease

Inflammatory bowel disease (IBD) is also often associated with depression and anxiety. Children with a good knowledge of their illness, an internal locus of control, a healthy family climate, and an open social network do well, and can maintain good mental health functioning despite severe illness. Disease severity appears to be a significant predictor of depressive symptoms in adolescents. The relationship between social support and adjustment observed with other conditions is also noted in children with IBD; those with high levels of social support from both family and peers show significantly better adjustment than those with limited support from either source. On the other hand, parental depression is associated with greater depressive severity in children more than adolescents with IBD. Cognitive behavior therapy and pharmacotherapy, with fluoxetine the most studied, have been shown to be helpful in treating depression in children and adolescents with IBD.

Insulin Dependent Diabetes Mellitus

Type 1 insulin dependent diabetes mellitus (IDDM) affects many aspects of everyday living and requires a high degree of personal responsibility to manage effectively. Poor diabetic control in childhood is associated with a greater risk for long-term complications and may lead to cognitive difficulties, making careful management more critical. Depression and anxiety are common in children and adolescents with IDDM, and risk is higher during the first year after diagnosis in those with pre-existing anxiety and in those whose mothers have psychiatric problems. Adolescence and associated physiologic changes of puberty can lead to increased insulin resistance; at the same time, peer relationships become more complex, and parental supervision may be withdrawn, all leading to greater problems with compliance and glycemic control. There appears to be a direct relationship between the diabetic adolescent's emotional and physical well-being. Adolescent juvenile diabetics, who are considered well-controlled medically, also tend to be well-adjusted at home and with peers, are appropriately independent, and have good self-esteem and an internal sense of control.

Diabetes is associated with an increased risk for eating disorders, which are associated with poorer glycemic control and an increased risk of diabetic comorbidities. Intentional omission of insulin to lose weight or as a form of suicidal behavior may occur with Type 1. Similarly, some patients with Type 2 will intentionally omit oral hypoglycemic drugs to lose weight. Binge eating disorders are also common. Patients tend to deny their manipulation of diet or insulin, but poor glycemic control, missed clinical appointments, and recurrent episodes of diabetic ketoacidosis or hypoglycemia should lead to suspicion of an eating disorder.

Heart Disease

Depression is common in patients with congenital heart disease and can exacerbate the physical consequences of the disorder. Children with congenital heart disease may have more medical fear and fear of the unknown, more physiologic anxiety, more depression, and more delinquent behavior than the general pediatric population. Those with cyanotic malformations may be at higher risk, which can be further exacerbated by maternal anxiety. Pain and separation from parents are more upsetting to younger children, while older children and adolescents are more focused on their physical status and the long-term implications of their condition. Accommodating to complete heart block and the implantation of a cardiac pacemaker often produces emotional distress; those children and adolescents without psychiatric symptoms who undergo the procedure demonstrate intellectualization, identification with medical personnel, and denial during the medical emergency. Problems in adjustment that occur following cardiac transplantation reflect the patient’s previous level of emotional adjustment more than a specific response to the procedure.

Ventricular assist devices (VADs) have become widely used in the management of end-stage heart failure in children, particularly in cardiomyopathy or myocarditis patients. Although overall outcomes are favorable, medical complications are common, especially for children with complex congenital heart defects. VADs are most useful as a bridge to recovery or eventual transplant.

Surgical Procedures

Some surgical procedures may be more disturbing than others, and children usually have more emotional problems after genitourinary surgery than after tonsillectomy. Although early psychoanalytic studies on the effect of tonsillectomy concluded that this operation was a stressful experience for every child, especially those with a history of previous trauma, currently the majority of children do well. Despite mild adverse behavioral changes in some younger children, more than 90% undergo the procedure without negative consequences.

Hemophilia

Hemophilia is an example of an inherited X-linked coagulation defect that serves as an illustration of individual and family response to a chronic and potentially lethal disorder. Most boys and their families adapt well and are able to maintain age-appropriate functions at home and school despite their illnesses. The few who show poor adjustment tend to have overprotective, overanxious, or guilt-ridden mothers, and the sons often respond either with social withdrawal and fear or with inappropriate recklessness.

Cancer

Cancer has been the second leading cause of death for children ages 5 to 14 years and the fourth leading cause of death in adolescents ages 15 to 19 years. Progress in the treatment of pediatric malignancies has led to an increase in the number of children, adolescents, and young adults surviving free of neoplastic disease, but this increased survival has come at a price. Survivors may suffer endocrinopathies, growth abnormalities, intellectual decline, neurocognitive and psychosocial dysfunction, and increased rates of anxiety, depression, dementia, and personality and behavioral disturbances. They also have an increased actuarial risk of developing a second malignant neoplasm. The prevalence rate of psychiatric disorders in children and adolescents with cancer is mixed, ranging from rare to as many as half of the patients with anxiety disorders, PTSD, and depression. Psychological interventions have shown promise in decreasing distress associated with diagnosis and treatment but have not shown an effect on long-term clinical outcomes. Survivors of childhood malignancies may have impaired quality of life, compromised social functioning and emotional well-being, including suicidal ideation, and diminished academic and occupational success as adults.

Covid-19

A pandemic associated with a new coronavirus began in China in December 2019 and quickly spread through person-to-person contact to multiple countries by January 2020. The virus was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the Coronavirus Study Group, and its related disease named coronavirus disease 2019, or COVID-19, by the World Health Organization. COVID-19 was declared a pandemic by the World Health Organization on March 11, 2020. Transmission is by droplets, spread mainly by coughing and sneezing by symptomatic and asymptomatic individuals.

Children have mainly become infected in family clusters and have had relatively milder illnesses and better outcomes than adults with the original genotype of the virus. The incubation period of COVID-19 ranges from 2 to 14 days, with most patients developing symptoms 3 to 7 days after exposure; children are usually symptomatic 1 to 2 days sooner than adults. Cough, fever, and headache are the most common presenting symptoms. Whether COVID-19 is clinically suspected or patients are asymptomatic, the diagnosis must be confirmed by a positive polymerase chain reaction test. Respiratory involvement is apparent in about 90% of pediatric patients. Gastrointestinal symptoms of abdominal pain, nausea, vomiting, and diarrhea are occasionally seen. Nonspecific skin findings can range from petechiae to papulovesicular rashes, diffuse urticaria, or chilblain-like lesions. Lymphopenia elevated procalcitonin, and elevated C-reactive protein are noted in laboratory studies. Pre-existing cardiac, hematologic, neurologic, and oncologic illnesses are reported most often in hospitalized children with severe COVID-19. SARS-CoV-2 is neurotropic, directly attacks neurons, is associated with a variety of neurologic symptoms, and precipitates an abnormal immune response in the CNS.

Multisystem inflammatory syndrome in children (MIS-C) in pediatric patients is most often seen in children 6 to 12 years old who present with fever and vomiting. MIS-C patients may become critically ill, requiring intensive care, mechanical ventilation, ionotropic support, and occasionally extracorporeal membrane oxygenation. Gastrointestinal (92%), cardiovascular (80%), hematologic (76%), and respiratory (70%) systems are most often involved in MIS-C.

Neurologic involvement is common in children with MIS-C and is usually transient. Patients with neurologic involvement are more likely to have underlying neurologic disorders, including seizures, neuromuscular disorders, autism, or developmental delay. Approximately one in four patients with neurologic involvement with MIS-C present with altered awareness or confusion of delirium. About 10% of MIS-C patients can develop a persistent disability or life-threatening neurologic conditions. Severe encephalopathy, CNS infection, encephalitis, aseptic meningitis, central demyelination, Guillain–Barré syndrome, acute fulminant cerebral edema, seizures, altered mental status, coma, demyelinating disorders, infarction, cerebellar ataxia, peripheral neuropathy, intracranial hypertension, lethargy, and drowsiness are reported.

Most survivors of COVID-19 can get back to normal soon after their initial illness. But in some, symptoms persist far longer, even after an initially mild illness, which has complicated planning to reopen schools. Patients, mostly girls, are reported to experience fatigue, dyspnea, palpitations, chest pain, muscle weakness, dizziness, difficulty concentrating, gastrointestinal problems, and headaches, symptoms similar to acute illness but without an active infection. Symptoms may last 6 to 8 months and can be very debilitating.

With the onset of the pandemic in 2020, the world essentially went into lockdown in an effort to slow the spread of the virus. Emergency nationwide school closures led to an immediate massive effort by schools and teachers to create online courses and deliver them as quickly as possible via the internet. School closures have been severely disruptive to children and adolescents, resulting in adverse academic and psychological effects. Loneliness from enforced social isolation and lack of usual peer interactions has increased the risk of depression and anxiety during and after the period of isolation, and the length of loneliness could be a predictor of future mental health problems.

Infants and toddlers appear to be more affected by the stresses COVID-19 places on their parents, and prolonged parental stress has been associated with increased behavior problems in the child. Younger children are more likely to become fearful, intolerant of uncertainty, more clingy, and dependent on their parents, while older children become more inattentive and worried about COVID-19 infection, school examinations, and other academic consequences. Anxiety, depression, irritability, mood swings, inattention, and sleep disturbance are fairly common, and depression and anxiety are particularly noted in adolescent cohorts.

Those children with pre-existing chronic medical and psychiatric conditions are at higher risk for regression, tantrums, and conflict with parents and siblings, making it more difficult for parents to handle their behavior and manage any special needs when they are home for long periods instead of in school. Mental health concerns are highly prevalent among youth with physical health problems, which worsened during the COVID-19 pandemic. Pre-existing autistic spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), and disruptive behavior disorders worsened. When speech, physical, and occupational therapies became limited, skills declined due to disruption in ongoing chronic care and supportive services. Children and youth who were previously in residential centers were sent back to their dysfunctional families as centers closed, exacerbating family problems. Children living in poverty suffered the consequences of malnutrition, parental unemployment, and homelessness. Physical exercise, eating habits, and lifestyle change when students are out of school. Becoming more sedentary and less physically active with prolonged school closures and home confinement meant more time in front of screens, more irregular sleep, and increased intake of high-caloric “comfort food,” resulting in weight gain and loss of cardiovascular fitness. Isolation in adolescence can increase the risk of poor adult health and depression.

Many families lost members to COVID-19, and some children live with a single parent or have parents who worked on the front lines, potentially adding to family distress. Physical abuse increased and the organizations that work with children with abuse or neglect became more restricted and less available. As the pandemic continued, risk increased for parental mental illness, domestic violence, and child maltreatment. More domestic violence and physical and/or sexual abuse at home increases the risk for suicide in children and adolescents, especially in girls, and suicidal ideation has been reported in 20% of junior high students, with 6% reporting suicidal plans, and 5% making suicide attempts.

TRANSITION FROM PEDIATRIC TO ADULT MEDICAL CARE

As children with chronic and severe medical problems are living longer into adulthood, the previously unanticipated problem of transitioning a pediatric patient to adult care has become important to address. Adolescence describes the time from childhood to adulthood when the teen must change from pediatrician to internist and move from shared responsibility with parents to individual responsibility for their own health. At some time, adolescence ends, and adulthood begins, even for patients with special health care needs. The transition from pediatric to adult-oriented care is a process that should take place in stages over the adolescent years. Six core elements of transition have been designated: (1) Transition policy, (2) Transition tracking and monitoring, (3) Transition readiness, (4) Transition planning, (5) Transfer of care, and (6) Completion of transfer. These six elements apply as well to transition from adolescence to adulthood in the medical home.

Adolescence can be a difficult time, especially for those with chronic illness who face unique challenges in developmental tasks. Effective transfer to adult care involves ongoing coordination and good communication with the patient, family, and adult physicians so that the anxieties of all, including the pediatrician, are adequately addressed.

Transition is not the same as transfer of care. Transition is a multifaceted process that addresses the psychosocial, educational, and vocational needs of adolescents as they move from child-oriented to adult-oriented care. It is predicated on effective communication between pediatrician and internist, and between the patient, parents, and providers. Transition implies a guided educational and therapeutic process rather than an administrative event.

There are two different approaches for transfer of adolescents with chronic medical problems to adult medical care. Patients may be directly transferred to internists. This is typical for patients with pediatric-onset of adult disease. Alternatively, patients may be transferred to a specialized center focusing on pediatric disorders in adults, and will be seen in pediatric clinics with separate areas designated for young adults.

Pediatric care is typically family-centered. The pediatrician looks to the parent for information and to provide instructions. In contrast, adults speak for themselves directly to their physicians. It is useful for the adolescent preparing for transition to begin talking to the pediatrician directly, without their parents' assistance, to prepare for their increased responsibility for their own care.

The time of transition of care is a period of vulnerability and potential risk, best accomplished during a period of relative stability. Many young adults regard health care as a low priority compared with education, employment, housing, relationships, and recreation. Young adults who have survived with pediatric care and parental support are at risk of deterioration after they move on to adult providers who are impatient with noncompliance and unfamiliar with the management of rare and often complex chronic childhood conditions. Those with less severe and visible conditions may have greater problems moving from pediatric to adult care, lacking preparation or only minimal or last-minute planning, resulting in poor communication and limited coordination for the effective transfer of care.

It is important that pediatric relationships are appropriately terminated as part of the transition process. The patient's level of developmental maturity and ownership of their own care is the most prominent determinant of transition outcomes. In order for the process of transition to be effective, the patient needs to have demonstrated adequate self-management skills, long-term social support, self-reliance, autonomy, and independence. Nonadherence with treatment is particularly prevalent in adolescents. Effective physician management requires attention to psychological and social issues as well as medical factors.

A positive sense of the future is associated with a greater sense of control in adolescents, supporting their emotional well-being and growth and increasing the likelihood of a more successful transition (Table 54.3–2). When accomplished effectively, transition can provide the young adult with a sense of the future and of future responsibilities that continued pediatric care fails to reinforce. A carefully planned transition to adult health care should improve self-reliance, enhance autonomy and independence, and support young people in attaining their maximum potential and having meaningful adult lives.

Table 54.3–2.
Recommended Approach to Transition in Adolescence

Period	Task
Early Adolescence Ages 10–13 years 1. Transition Policy	Introduce concept of separate and confidential visits with MD. Straight forward reward-oriented approach. Begin discussion of transfer of care. Need for privacy begins.
Mid Adolescence Ages 14–16 years 2. Track progress 3. Monitor transition readiness 4. Assess skills	Begin discussion of healthy living and risky behaviors in nonjudgmental approach. Expect increased introspection. Begin visits to new setting and meeting new providers.
Late Adolescence Ages 17–21 years 5. Planning and transfer of care 6. Integration into an adult medical care setting	Increased self-responsibility. Transition to adult providers. Future-oriented approach. Progression of personal identity and code of ethics. Physician seen alone. Evolving role of social supports.

For those who are mentally ill, the transfer of their psychiatric care from child-adolescent to adult psychiatrists can be more daunting than medical transition and a potentially more precarious time if preparation is inadequate. The MILESTONE Project (Managing the link and strengthening transition from child to adult mental health care) is intended to address the significant socioeconomic and societal challenges related to the transition process for psychiatric care, which is less structured than the process for medical care. Successful mental health transition requires effective self-management, medication adherence, attendance at medical appointments, understanding of the illness, and coping with mental health crises when they occur.

SOCIAL MEDIA

Any website that allows social interaction is considered a social media site. This includes social networking sites such as Facebook, Twitter, Instagram, and TikTok; gaming sites and virtual worlds such as Neopets, FreeRealms, Minecraft, Webosaurs and video sites such as YouTube, Instagram, TikTok, and various blogs (these examples will almost certainly be dated by publication). These sites offer today's youth a portal for entertainment and communication, and some adolescents log onto their favorite social media sites as often as 10 or more times a day. Just over half of children in the United States have a smartphone by age 11, and 84% of teenagers have their own phones. A significant part of this generation's social and emotional development is now occurring while on the internet and cell phones. Students use the internet to do homework, look up information, collaborate on group projects, and attend school online, especially since the COVID-19 pandemic. Internet and social media sites provide useful applications and allow the development of user-generated content, blogs, patient experience journals, and video and photo-sharing sites. Children and adolescents, particularly those with chronic conditions, seek out social media and other online sources to acquire health information, connect with others affected by similar conditions, and play a more active role in their healthcare decisions. Many patients and families can develop large support networks by sharing and disseminating information about their child's illness on specialized sites, including hospital-sponsored sites such as CaringBridge or CarePages. Internet support groups and interactive peer video games can decrease some of the isolation patients, and families feel with rare or unknown conditions or with necessary hospital medical isolation practices. Applications that monitor and track symptoms (e.g., adherence, mood, nutritional intake, chemotherapy side effects, blood cell count) may be useful, and applications that facilitate delivery of stress management and relaxation techniques (e.g., biofeedback, breathing exercises, mindfulness, imagery) are available but less widely disseminated. These tools are helpful but are not a substitute for clear in-person communication, clinical psychological and psychiatric assessment, and psychotherapy.

Practitioners, patients, and families need to consider threats to confidentiality and privacy and professional and therapeutic boundaries when communicating via social media. The risks to adolescents when using social media can be more than most adults appreciate. Cyberbullying and online harassment are common, which can have profound consequences and precipitate aggressive and suicidal behavior in response. Sexting, or sending sexually explicit messages, photographs, or images via cell phone, computer, or other digital devices, can lead to school suspension for perpetrators and emotional distress for victims. "Facebook Depression" develops when teens and preteens spend excessive time on social media, risking social isolation and involvement with sites that may promote substance abuse or unsafe sexual or other behaviors. The Children's Online Privacy Protection Act (COPPA) prohibits sites from collecting information on children younger than 13 years without parental permission, and sites should mirror COPPA regulations, setting 13 as the minimum age to sign onto sites such as Facebook and Instagram. Parental supervision of social media, online sites, and other digital activities is essential, especially as there is minimal legal protection for minors between 13 and 17 years from the hazards of cell phones and computers.

FURTHER READINGS

Arruda-Colli MNF, Bedoya SZ, Muriel A, Pelletier W, Wiener L. In good times and in bad: what strengthens or challenges a parental relationship during a child's cancer trajectory? *J Psychosoc Oncol.* 2018;36:635–648.

Bailey-Pearce O, Stedmon J, Dallos R, Davis G. Fathers' experiences of their child's life-limiting condition: an attachment narrative perspective. *Clin Child Psychol Psychiatry*. 2018;23:381–397.

Becker JE, Smith JR, Hazen EP. Pediatric consultation-liaison psychiatry: an update and review. *Psychosomatics*. 2020;61:467–480.

Belay ED, Abrams J, Oster ME, et al. Trends in geographic and temporal distribution of US children with multisystem inflammatory syndrome during the COVID-19 pandemic. *JAMA Pediatr*. 2021;175(8):837–845.

Chassiakos YLR, Radesky J, Christakis D, Moreno MA, Cross C; Council on Communication and Media. Children and adolescents and digital media. *Pediatrics*. 2016;138(5):e20162593.

Cobham VE, Hickling A, Kimball H, Thomas HJ, Scott JG, Middeldorp CM. Systematic review: anxiety in children and adolescents with chronic medical conditions. *J Am Acad Child Adol Psychiatr*. 2020;59:595–618.

Cohodes EM, McCauley S, Gee DG. Parental buffering of stress in the time of COVID-19: family-level factors may moderate the association between pandemic-related stress and youth symptomatology. *Res Child Adolesc Psychopathol*. 2021;49(7):935–948.

Costello CR, McNeil DE, Binder RL. Adolescents and social media: privacy, brain development, and the law. *J Am Acad Psychiatr Law*. 2016;44:313–321.

Dinleyici M, Carman KB, Özdemir C, et al. Quality-of-life evaluation of healthy siblings of children with chronic illness. *Balkan Med J*. 2020;37:34–42.

Eilertsen MB, Lövgren M, Wallin AE, Kreicbergs U. Cancer-bereaved siblings' positive and negative memories and experiences of illness and death: a nationwide follow-up. *Palliat Support Care*. 2018;16:406–413.

Fegert JM, Vitiello B, Plener PL, Clemens V. Challenges and burden of the Coronavirus 2019 (COVID-19) pandemic for child and adolescent mental health: a narrative review to highlight clinical and research needs in the acute phase and the long return to normality. *Child Adolesc Psychiatry Ment Health*. 2020;14:20.

Fleary SA. Association between the medical home model of care and demographic characteristics in US children. *Clin Pediatr.* 2019;58:270–281.

Green M, Solnit AA. Reactions to the threatened loss of a child: a vulnerable child syndrome. *Pediatrics*. 1964;34:58–66.

Hart LC, Maslow G. The medical transition from pediatric to adult-oriented care: considerations for child and adolescent psychiatrists. *Child Adolesc Psychiatr Clin N Am*. 2018;27:125–132.

Kieckhefer GM, Trahms CM, Churchill SS, Kratz L, Uding N, Villareale N. A randomized clinical trial of the building on family strengths program: an education program for parents of children with chronic health conditions. *Matern Child Health J*. 2014;18:563–574.

Kim EG, Lee Y-M, Riesche L. Factors affecting depression in high school students with chronic illness: a nationwide cross-sectional study in South Korea. *Arch Psychiatr Nurs*. 2020;34:164–168.

Kolko DJ, Perrin E. The integration of behavioral health interventions in children's health care: services, science, and suggestions. *J Clin Child Adolesc Psychol*. 2014;43:216–228.

LaRovere KL, Riggs BJ, Poussaint TY, et al. Randolph AG for the Overcoming COVID-19 Investigators. Neurologic involvement in children and adolescents hospitalized in the United States for COVID-19 or Multisystem Inflammatory Syndrome. *JAMA Neurol*. 2021;78(5):536–547.

Lövgren M, Udo C, Alvariza A, Kreicbergs U. Much is left unspoken: self-reports from families in pediatric oncology. *Pediatr Blood Cancer*. 2020;67:e28735.

Ludvigsson JF. Case report and systematic review suggest that children may experience similar long-term effects as adults after clinical COVID-19. *Acta Paediatr*. 2021;110(3):914–921.

McKune SL, Acosta D, Diaz N, et al. Psychosocial health of school-aged children during the initial COVID-19 safer-at-home school mandates in Florida: a cross-sectional study. *BMC Public Health*. 2021;21:603.

Nearchou F, Flinn C, Niland R, Subramamiam SS, Henessy E. Exploring the impact of COVID-19 on mental health outcomes in children and adolescents: a systematic review. *Int J Environ Res Public Health*. 2020;17:8479.

Nelville AR, Moothathamby N, Naganathan M, Huynh E, Moola FJ. "A Place to call our own": the impact of camp experiences on the psychosocial wellbeing of children and youth affected by cancer – a narrative review. *Complem Ther in Clin Pract*. 2019;36:18–28.

Perrin EC. Children in hospitals. *J Dev Behav Pediatr*. 1993;14:50–52.

Sadof MD, Nazarian BL. Caring for children who have special health-care needs: a practical guide for the primary care practitioner. *Pediatr Rev*. 2007;28;e36–e42.

Santosh P, Adams L, Fiori F, et al. Protocol for the development and validation procedure of the managing the link and strengthening transition from child to adult mental health care (MILESTONE) suite of measures. *BMC Pediatrics*. 2020;20:167.

Simon TD, Berry J, Feudtner C, et al. Children with complex chronic conditions in inpatient hospital settings in the United States. *Pediatrics*. 2010;126:647–655.

Weiner L, Battles H, Zadeh S, Pelletier W, Arunda-Colli MNF, Muriel AC. The perceived influence of childhood cancer on the parents' relationship. *Psychooncol*. 2017;26:2109–2117.

White PH, Cooley WC, Transitions Clinical Report Authoring Group, American Academy of Pediatrics, American Academy of Family Physicians, American College of Physicians. Supporting the health care transition from adolescence to adulthood in the medical home. *Pediatrics*. 2018;143:e20182587.

Winters NC, Metz WP. The wraparound approach in systems of care. *Psychiatr Clin North Am*. 2009;32:135–151.

54.4 Psychiatric Sequel of HIV and AIDS

MARK DEANTONIO, M.D.

In June 1981, the first cases of what was later called acquired immunodeficiency syndrome (AIDS) in the United States were reported in *Morbidity and Mortality Weekly Report*. Since that time, this syndrome has become one of the greatest public health challenges. In 2019, an estimated 38.0 million people were living with human immunodeficiency virus (HIV) worldwide. HIV continues disproportionately to affect certain geographic regions (e.g., sub-Saharan Africa and the Caribbean) and subpopulations (e.g., women in sub-Saharan Africa, men who have sex with men, injection-drug users, and sex workers). There have been significant changes in the understanding, epidemiology, and treatment of this disease in children and adolescents, modifying psychiatric sequelae. In the United States, the prevention of vertical transmission and the surveillance of blood make it rare for a child or adolescent younger than 15 years of age to become infected with HIV. However, adolescent sexual behavior and drug use, especially when combined with psychiatric disturbance, put individuals at increased risk of HIV infection in adolescence and young adulthood. Psychiatrists in the United States might never encounter an HIV-infected child but have a great opportunity to decrease the risk of infection when psychiatrically treating adolescents who are at risk. Gaining an understanding of the psychiatric sequelae of HIV and AIDS and the dynamics of adolescent risk behavior will allow the physician to craft effective interventions to prevent transmission in youth.

EPIDEMIOLOGY

In the United States at the year-end of 2018, approximately 1.2 million people were living with HIV, including 1,012,100 with diagnosed HIV and 161,800 with undiagnosed HIV. The number of people with HIV (diagnosed or undiagnosed) in the United States has steadily increased as the number of new infections per year has outpaced the number of people dying from HIV. Key 2018 HIV prevalence data in the United States include 58% of persons with HIV had male-to-male sex as their transmission category, Blacks/African-Americans accounted for approximately 41% of all people with HIV, and 58% of persons with HIV were 45 years of age or older. Among the 37,515 persons who were diagnosed with HIV in 2018 in the United States, 66% acquired HIV via male-to-male sex, 43% were Black/African American, and 46% resided in the South at the time of the diagnosis. The HIV incidence (estimated new HIV infections) in the United States has declined slightly between 2014 and 2018, with an estimated 36,400 new infections in 2018. However, the highest number of persons newly diagnosed with HIV in 2018 was between 20 and 35 years of age. Using back-calculation methods, the CDC estimated the proportion of persons with HIV in the United States who are unaware of their HIV status decreased from 25% in 2003 to 13.8% in 2018; however, among persons aged 13 to 24 years with HIV, 45% are undiagnosed. Persons who are aware of their HIV status can benefit from HIV treatment and are less likely to transmit the virus to others. The annual number of deaths for persons diagnosed with HIV remained relatively steady between 2014 and 2018. Most of these deaths are due to non-AIDS causes, particularly liver disease, cardiovascular disease, and non-AIDS malignancies. Among persons with diagnosed HIV who died in 2018 in the United States, Blacks/African Americans had the highest number of deaths and highest death rate. Globally, an estimated 38 million people are living with HIV, and 25.6 million (67%) reside in sub-Saharan Africa. Globally, in 2019, an estimated 1.7 million new HIV infections occurred, and an estimated 6,90,000 persons with HIV died. In 2019, an estimated 25.4 million persons with HIV globally were receiving antiretroviral therapy.

The changing United States epidemiology shows relative control of HIV infection in children under 13 but a significant rise in rates of individuals in their 20s. Sexual behavior, specifically male-to-male sexual contact, is the primary cause of infection, with injection drug use and heterosexual contact decreasing as a cause. The behaviors that cause infection in individuals in their 20s often begin in adolescence. These trends highlight the importance of interventions to change sexual behavior in adolescence.

There is another way in which HIV infection affects children and adolescents: When their parents or family members are infected. The number of U.S. youth affected by HIV in a family member is growing because the majority of infected women are mothers; 60% have children younger than 18 years of age, and 45% live with their children. Living with chronically ill parents, often with comorbid substance abuse, has significant consequences for the affected family members, with potential psychiatric and social sequelae.

ETIOLOGY

Psychiatric sequelae of HIV infection in children and adolescents are due to multiple and interacting etiologies. These include direct and indirect effects of HIV infection on the brain, stresses of living in families often affected by substance abuse, psychiatric disorders, chronic illness, and death. Adolescents who contract HIV infection have a high likelihood of pre-existing psychiatric diagnosis, cognitive deficits, histories of sexual or physical abuse, ongoing substance abuse, homelessness, and chaotic family dynamics. The addition of HIV infection to these adolescents adds profound stress and challenge to an already overwhelming existence.

HIV is transmitted through blood-to-blood or semen-to-blood contact. The virus can also be transmitted by penile–vaginal intercourse, especially if other sexually transmitted diseases (STDs) coexist. Receptive anal intercourse and the use of contaminated syringes are the most common causes of HIV infection in the United States. HIV-infected women can transmit the virus during pregnancy, delivery, or through breast milk. It also can be transmitted through untested blood products. There are four identified types of human retroviruses, which belong to two distinct groups: the transforming human T cell lymphotropic virus (HTLV) types I and II, and cytopathic HIV-1, which is more common in the United States, and HIV-2. HIV-1 is further divided into groups, classes or subtypes, and circulating recombinant forms, all of which indicate geographic origin.

Testing to diagnose HIV infection has evolved to provide quicker, more accurate, and more cost-effective results. Current testing approaches directly detect the virus or demonstrate antibodies to the virus or a component, which appears 2 to 12 weeks after acute infection. A definitive diagnosis among infants younger than 15 months old cannot be made because the antibody crosses the maternal–fetal blood barrier, so all infants of HIV-infected mothers will have maternal HIV antibodies, whether infected or not. Once an individual is diagnosed as infected, the level of plasma HIV RNA (viral load) and the absolute number of peripheral CD4$^+$ T cells are used to monitor clinical status.

In 1985, when HIV testing first became available, the main goal of such testing was to protect the blood supply. At that time, professional opinion was divided regarding the value of HIV testing for individuals and whether HIV testing should be encouraged because no consensus existed regarding whether a positive test predicted transmission of the HIV virus to sex partners or from mother to infant. No effective treatment existed. In 1987, the implications of positive HIV serology became evident. HIV testing and counseling were recommended for all persons in high-risk groups and all persons seeking treatment for STDs. In 1995, after perinatal transmission of HIV was demonstrated to be substantially reduced by the administration of zidovudine to HIV-infected pregnant women and their newborns, it was recommended that all pregnant women be counseled and offered testing for HIV. In 2001, the CDC recommended that all pregnant women should have HIV screening as a part of routine prenatal care. Although the woman needed to be aware of what they were being tested for, it was also felt that the previously recommended pretest counseling was a barrier to testing and should be simplified. At this time, the CDC recommended that HIV testing be offered routinely to all patients in high-HIV-prevalence healthcare settings, despite risk factors. In low-prevalence settings, targeted HIV testing on the basis of risk screening continues to be recommended.

In September 2006, the CDC again revised recommendations for HIV testing of adults, adolescents, and pregnant women. Due to the changes in demographic of the HIV/AIDS epidemic in the United States, increasing proportions of infected persons are younger than 20 years of age, women, members of racial or ethnic minority populations, persons who reside outside metropolitan areas, and heterosexual men and women who frequently are unaware that they are at risk of HIV infection. As a result, the effectiveness of using risk-based testing to identify HIV-infected persons has diminished. In addition, prevention strategies that incorporate universal HIV screening have been highly effective (blood donors and pregnant women). Perinatal transmission rates have been reduced to less than 2% due to prophylactic administration of antiretroviral drugs, scheduled cesarean delivery, and avoidance of breastfeeding. Studies document that the majority of adults who are aware of their HIV infections substantially reduce sexual behaviors that might transmit HIV after they become aware that they are infected. Therefore, to increase the diagnosis of HIV infection, destigmatize the testing process, link clinical care with prevention, and ensure immediate access to clinical care for persons with newly identified HIV infection, the CDC now recommends the adoption of routine HIV testing in all health care settings for individuals from the age of 13 to 64 years. The testing should be undertaken only with the patient’s knowledge and understanding that HIV testing is planned. Patients can refuse testing. Persons likely to be at high risk of HIV infection—injection drug users and their sex partners, sex workers, partners of HIV-infected persons, and persons who have had or whose sex partners have had more than one sex partner since their most recent HIV test—all should be tested annually. This significant change in testing recommendations reflects the ever-changing nature and better understanding of HIV/AIDS.

Pre-exposure prophylaxis, or PrEP, is a new tool in HIV prevention. It is a prevention option for individuals who are at high risk of getting HIV. It is a pill with two antiviral medications (tenofovir and emtricitabine), which are taken daily. PrEP has been shown to be safe and efficacious for adults. Recent studies have now documented the safety and efficacy of adolescents using PrEP. However, these studies also highlight the challenge of using PrEP in adolescents, which is that over time, adherence to treatment declines. New treatments of injectable pre-exposure prophylaxis medications, which could address the adherence issue, have not been studied yet in adolescence.

Disclosure of a child's or adolescent's HIV status brings up complex issues requiring careful assessment of an individual's circumstance. Knowledge of one's HIV status is associated with emotional distress and increased rates of depression. The level of distress depends on how the disclosure occurs, under what circumstances, and the reaction of the people involved. Telling a child that they have HIV appears to be influenced by several factors, including older child age, higher child cognitive functioning, recent major life events, and higher family expressiveness. Many clinicians and caregivers are ambivalent about disclosing to children because of fears that the children will inappropriately disclose to the community at large and face stigmatization as a consequence. Current pediatric practice guidelines encourage disclosure by adolescence, but neurocognitive deterioration may complicate timing. Clearly, it is critical that HIV status be disclosed to adolescents before they become sexually active. Simply telling adolescents that they are HIV positive may not reduce risky sexual behavior or disclosure of that status to their sexual partners. This potential lack of disclosure underscores the importance of emphasizing safer sex practices, including condom use, for all adolescents regardless of HIV status.

Neurodevelopmental Sequelae of HIV Infection

HIV is found primarily in microglia and macrophages and not directly in neurons. The virus is thought to have neurotoxic components and induce a variety of inflammatory factors that cause neuronal cell death. The clinical presentation of HIV neurologic involvement can vary from mild neurocognitive impairment to static encephalopathy, progressive encephalopathy, and dementia. There has been more progress in understanding the effects of HIV on the brain in adults than in children, owing to the dramatic decrease in childhood infections. Studies of infected patients 13 to 18 years of age are limited. In the earliest asymptomatic stages, abnormalities may show up in cognitive abilities evaluated during neuropsychiatric testing but not necessarily as impairments in functioning. In children and adolescents, those infected are also likely to have cognitive deficits due to in-utero drug and alcohol exposure, attention-deficit disorder, and learning disabilities, all of which increase maternal HIV infection risk and complicate neuropsychiatric testing. The clinical presentation of HIV central nervous system (CNS) involvement may vary with the age of infection and mode of transmission. In children or adolescents infected through blood transfusion, encephalopathy generally presents later and follows a more indolent course. By contrast, in almost one-fourth of vertically HIV-infected children, encephalopathy can present and contribute significant morality and mortality as an AIDS-defining illness even when the child is not severely immunocompromised. Early progressive encephalopathy is usually seen in young, untreated infants with high-viral loads and rapid clinical deterioration. Vertically HIV-infected children who are labeled "slow progressors"—those with lower viral loads and higher CD4 counts—often exhibit no cognitive deficits throughout childhood and adolescence.

The most common computed tomographic findings associated with encephalopathy are basal ganglia calcifications, brain atrophy with enlarged ventricles, and enlarged cortical sulci. HIV-infected children have a higher incidence of ischemic strokes and cerebral artery aneurysms, which can rupture. Magnetic resonance imaging (MRI) has been recommended for infected children who present with a lack of neurocognitive progression or dysfunction even though they are otherwise asymptomatic.

In adolescents with advanced AIDS, a gradual progression into dementia is seen. It appears as a subcortical dementia process with motor impairment, cognitive disturbance, and a predominance of apathy over depression. Psychotic symptoms may be present. In addition, opportunistic CNS infections such as toxoplasmosis, progressive multifocal leukoencephalopathy, and CNS lymphoma can complicate the CNS clinical picture. There is no specific treatment for HIV involvement in the CNS. When progressive encephalopathy is present in children and adolescents, increased doses of antiviral agents are used, but there have not been studies documenting efficacy in children and adolescents.

CLINICAL FEATURES

Despite more than 30 years of studying HIV/AIDS in children and adolescents, there have been few studies of psychiatric diagnoses in these individuals. Although there is clearly defined neurologic pathology due to HIV infection, psychiatric illness, specifically due to HIV infection, has not been quantified. Instead, psychiatric diagnoses are related to inheritable psychiatric disorders and reactions to having a chronic illness in the context of often chaotic and traumatic environments.

Early studies of emotional and behavioral disorders in perinatally HIV-infected children showed high rates of anxiety, depression, adjustment problems, and attention-deficit/hyperactivity disorder (ADHD). However, results among studies were inconsistent and failed to use appropriate control groups. Studies of behavioral disorders in HIV-infected children that included uninfected control groups of similar backgrounds generally reported high rates of emotional and behavioral disorders in both groups. One controlled study found elevated rates of ADHD in a sample of prenatally drug-exposed children, with no difference between HIV-infected and uninfected children. Another controlled study found elevated rates of internalizing and externalizing symptoms in HIV-infected children and demographically matched control. However, rates of symptomatology were actually higher in the control group. The Woman and Infant Transmission Study Group Study compared HIV-infected children with uninfected siblings born to HIV-positive mothers. Findings confirmed earlier data, revealing high rates of hyperactivity and impulsivity in both the infected and uninfected children. The Pediatric AIDS Clinic Trials Group reported a higher frequency of psychiatric hospitalization among perinatally infected children than in the general pediatric population and HIV-uninfected members of the cohort. The majority of patients were admitted because of depression, and a minority were due to behavioral disorders. Studies of vertically infected adolescents and psychiatric illness are limited to case reports. The group whose AIDS progresses significantly in adolescence exhibits a dementia process, which may include psychosis and agitation. The group of vertically infected adolescents who are labeled “slow progressors” due to their mild degree of immune suppression appear to present with depression, anxiety, and chronic adjustment disorders related to chronic illness.

Adolescents who become infected due to sexual or injection substance abuse have recently been a focus of study through the Adolescent Trials Network for HIV/AIDS Interventions. There is evidence that having psychiatric illness in adolescence, especially when coupled with substance abuse and cognitive and learning difficulties, increases the high-risk behaviors that can lead to HIV infection. It has been documented that a significant number of HIV-infected adolescents have depressive disorders. Conduct disorder and substance abuse are also present to a significant degree. The majority of these disorders predated their HIV infection. There is evidence in the adult and adolescent literature that having depression while being HIV-infected can alter the function of killer lymphocytes. This causes an increase in activated CD8 T lymphocytes and can increase the viral load. Treatment with selective serotonin reuptake inhibitors (SSRIs) and a decrease in depressive symptoms reverse these changes.

TREATMENT

Psychiatrists play an important role in the treatment of children and adolescents who are HIV-infected. HIV infection has changed over the years to a chronic illness in the United States, which requires multiple layers of care in a variety of settings over many years. Although pediatricians have expertise in managing developmental issues in chronic illness in youth, psychiatrists have a system-oriented approach arising from their work with patients who depend on the support systems provided by family, school, and community. Parents of HIV-infected children and adolescents have a high rate of psychiatric illness, substance abuse, and chaotic and dysfunctional family interactions. Psychiatry brings expertise to the treatment team in crafting interventions to promote family stability and family support of the patient’s medical treatment. Finally, a significant number of children and adolescents who are HIV-infected have a pre-existing psychiatric illness that requires psychiatric expertise to manage. In children and adolescents receiving antiretroviral treatment, lack of almost perfect adherence to daily regimes continues to be the greatest barrier to sustaining a clinically effective response. Lack of adherence can also lead to the development of HIV strains resistant to the available antiretrovirals. Although adherence issues are common with all chronic illnesses in children and adolescents, it is more of an issue in HIV-infected children and adolescents living in situations in which parents may be physically or mentally ill or with children and adolescents in a foster care or residential setting with inadequate supervision. The psychiatrist has a critical role in working with the treatment team, the child or adolescent, and caregivers to problem-solve and intervene around issues of adherence.

The HIV epidemic adds to the challenges of the psychiatric treatment of children and adolescents in inpatient and residential settings. The American Academy of Child and Adolescent Psychiatry has the following recommendations concerning HIV and inpatient psychiatric treatment:

- HIV infection or risk of HIV infection is not a reason to deny a child admission to a psychiatric hospital.
- All patients admitted to a psychiatric hospital should be treated as if they were HIV-infected. Even if all patients were tested on admission for antibodies to HIV, one is still not sure of infection status due to the potential latency period of antibody response documenting infection.
- Universal precautions should be employed for any secretions, including blood from any psychiatric inpatient. There is no need for a change in the management of an HIV-infected child if universal precautions are used for all inpatients.
- HIV-positive or potentially positive patients do not require individual rooms or toilet facilities. HIV-infected or potentially infected patients do not need to be isolated in regard to the milieu or peer groups. Such patients should participate in all aspects of the inpatient treatment program.
- The Academy agrees with the 2006 CDC recommendation of the adoption of routine HIV testing in all healthcare settings from the age of 13 to 64 years. The testing should be undertaken only with the patient's knowledge and understanding that HIV testing is planned. Patients can refuse to test. Persons likely to be at high risk of HIV infection, including injection drug users and their partners, sex workers, partners of HIV-infected persons, and persons who themselves or whose sex partners have had more than one sex partner since their most recent HIV test, all should be tested annually. The child and adolescent psychiatrist should balance the adolescent's request for privacy with the family's wishes. There are specific laws regarding the confidentiality of HIV antibody testing results and who authorizes consent in minors. These laws vary from state to state. Prior to testing, ensure the laws are understood.
- Results of patients' HIV antibody tests will be maintained in a confidential manner. A patient's HIV status should be shared only with those staff who need to know the status to appropriately care for the child while an inpatient. Sharing of HIV status should, of course, be in compliance with applicable state and federal law.

Through the Adolescent Trials Network for HIV/AIDS Interventions, there have finally been published controlled trials of treatment. Interactions between psychotropic and antiretroviral medications are common in HIV-infected children and adolescents. Most medications used in the treatment of psychiatric disorders are metabolized by the cytochrome P450 (CYP) enzymes, especially the subgroups 3A4 and 2D6. Of the antiretrovirals, the protease inhibitors tend to pose the greatest risk of changes in drug levels that are significant in the HIV patient. Ritonavir is the most potent of the protease inhibitors, with the greatest effect on the inhibition of 3A4 and, to a lesser extent, 2D6. Enfuvirtide, a fusion inhibitor, is not an inhibitor of CYP enzymes and does not alter the metabolism of CYP 3A4, 2D6, 2C19, or 2EI substrates. Therefore, it has no significant effect on blood levels of psychiatric medications or other antiretrovirals. Efavirenz has been shown in vivo to induce CYP 3A4. Other compounds that are substrates of CYP 3A4 may have decreased plasma concentrations when coadministered with efavirenz. In vitro studies have demonstrated that efavirenz inhibits 2C9, 2C19, and 3A4 isozymes. Coadministration of efavirenz with drugs primarily metabolized by these isozymes may result in altered plasma concentrations of the coadministered drug. Drugs that induce CYP 3A4 activity would be expected to increase the clearance of efavirenz. Some antiretroviral drugs have been reported to cause psychiatric symptoms in HIV-infected youth, including mania with zidovudine and attention-deficit problems with efavirenz. It is important for psychiatrists to remain in contact with the other medical personnel providing treatment for the HIV-infected child or adolescent, maintain updated lists of all medication, and frequently consult updated web-based information sites regarding drug-drug interactions.

In terms of the strategy of psychopharmacologic treatment, thoughtfulness is primary. The concern a physician has is that the psychiatric symptoms present are really due to the "organicity" of HIV infection and so should only be treated by maximizing antiretroviral treatment. Although this point of view needs to be considered, one can make psychiatric diagnoses and use psychopharmacology despite a possible underlying organicity caused by HIV. The most common disorders presenting in this population are

ADHD, generalized anxiety disorder, major depressive disorder, bipolar affective disorder, psychotic disorder not otherwise specified, and schizophrenia. One should use the same medications with the same dose titration as one would with an uninfected child or adolescent. The exception would be to start and maintain lower doses of medication in patients with clear demented or encephalopathy. Stimulants, in general, have few drug–drug interactions, and lithium has none because it is excreted renally. In adults, it has been reported that the HIV virus can target the destruction of cells in the basal ganglia, which can sensitize an individual to extrapyramidal side effects when they use an antipsychotic medication. This observation has not been reported in children and adolescents, but because children and adolescents overall are more sensitive to extrapyramidal side effects, this should be of concern. Therefore, one should first use atypical antipsychotics when treating children who have psychosis and adolescents who are HIV-infected.

Recent studies have documented effective treatment for depression in adolescents and young adults with the combination of cognitive behavioral therapy and administration of SSRI. Family and group interventions are an important adjunct to the successful treatment of HIV-infected youth. For infected adolescents, prevention programs emphasize reducing transmission acts, enhancing the quality of life, increasing self-care behaviors, improving medical adherence, coping with medication side effects, dealing with appropriate disclosure of HIV status, and interacting with health care providers. Preventive interventions for youth living with HIV using therapeutic approaches based on cognitive behavioral principles have been studied. One intervention included two primary components: (1) increasing positive health behaviors, particularly around medication adherence, and (2) enhancing altruistic motivation to reduce transmission acts. The intervention emphasized coping with being HIV positive, adopting healthy routines, partaking in health care decisions, reducing substance use and risky sexual behavior by identifying risk triggers, and increasing self-efficacy to use condoms and negotiate safe sex. The intervention sought to build skills and improve emotion regulation. These are similar to the strategies used to enhance psychological functioning for youth in psychiatric care. Adolescents who received the intervention reported more lifestyle changes, improved coping, fewer sexual partners, and a lower percentage of unprotected sexual acts than did youth who did not receive the intervention. Additional sessions emphasized increasing life satisfaction and emotional strength. The additional sessions reduced emotional distress and psychiatric symptoms among the infected youth, including somatization and anxiety. Programs have also been developed to help the adult caregivers of children and adolescents with HIV infection. These programs teach effective parenting skills, parental monitoring of medication adherence, and increasing social support among family members and the community. Some of the more innovative interventions also address the caregivers' mental health and quality of life.

Despite potential long-term intergenerational consequences of HIV illness, few services are available to assist children and adolescents of HIV-infected family members. The few programs with published data have yielded positive effects for affected adolescents. These programs address coping with stigma, dealing with illness-related tasks, dealing with parental disclosure, increasing maternal well-being and social support, developing effective parenting skills, decreasing victimization, custody planning, maintaining positive family routines, and handling emotional reactions to AIDS. Two– and 4-year follow-up studies of intervention effects have been both promising and disappointing. Two years after the intervention, treated adolescents reported significantly fewer conduct problems, less emotional distress, lower levels of family-related stress, and higher self-esteem than adolescents in the control group. Outcomes were comparable for youths who were bereaved and nonbereaved by parental AIDS and across ethnic and gender subgroups. Unfortunately, the positive effects had eroded by the 4-year follow-up, underscoring the need for sustained support and treatment for these families. HIV prevention programs for affected children and adolescents must address intergenerational imitation of HIV risk behaviors and substance abuse and help them to cope with HIV-related psychosocial challenges such as bereavement. Unfortunately, psychosocial interventions for the affected have not been broadly implemented or prioritized.

CONCLUSION

Child and adolescent psychiatry, with its developmental and systems approach, has a unique position from which to manage the mental health sequelae of HIV infection in children and adolescents. Although this epidemic has created challenges, it has also given opportunities. The epidemic has forced a reevaluation of child and adolescent sexual development and a more comprehensive understanding of the development of sexual orientation. It has prioritized the need for the assessment of sexual and drug-taking behavior as part of an adolescent psychiatric assessment. Finally, the AIDS epidemic demands that psychiatrists constantly monitor high-risk behaviors in adolescent patients and implement strategies to reduce these behaviors. Such interventions will improve the quality of life and prognosis of adolescent psychiatric patients regardless of HIV infection status.

Additional Information: Author [Mark DeAntonio] died December 31, 2021.

FURTHER READINGS

Allen E, Gordon A, Krakower D, Hsu K. HIV preexposure prophylaxis for adolescents and young adults. *Curr Opin Pediatr*. 2017;29(4):399–406.

Bland S. Expanding Access to Pre-Exposure Prophylaxis (PrEP) for Adolescents and Young Adults: Models for Addressing Consent, Confidentiality, and Payment Barriers. amFAR | March 17, 2019.

Brown LK, Kennard BD, Emslie GJ, et al. Effective Treatment of Depressive Disorders in Medical Clinics for Adolescents and Young Adults Living with HIV: A Controlled Trial. *J Acquir Immune Defic Syndr*. 2016;71(1):38–46.

Bucek A, Leu CS, Benson S, et al. Psychiatric Disorders, Antiretroviral Medication Adherence and Viremia in a Cohort of Perinatally HIV-Infected Adolescents and Young Adults. *Pediatr Infect Dis J*. 2018;37(7):673–677.

Centers for Disease Control and Prevention. Diagnoses of HIV infection in the United States and dependent areas, 2018 (Updated). *HIV Surveillance Report*. 2020;31:1–119.

Centers for Disease Control and Prevention. Estimates HIV Incidence and Prevalence in the United States, 2014–2018. *HIV Surveillance Supplemental Report*. 2020;25(No. 1):1–77.

Evangeli M. Mental health and substance use in HIV-infected adolescents. *Curr Opin HIV AIDS*. 2018;13(3):204–211.

Harris LL, Chernoff MC, Nichols SL, et al. Prospective memory in youth with perinatally-acquired HIV infection. *Child Neuropsychology*. 2018;24(7):938–958.

Harrison S, Li X. Toward an enhanced understanding of psychological resilience for HIV youth populations. *AIDS Care*. 2018;30(sup4):1–4.

Hermetet-Lindsay KD, Correia KF, Williams PL, et al. Contributions of Disease Severity, Psychosocial Factors, and Cognition to Behavioral Functioning in US Youth Perinatally Exposed to HIV. *AIDS Behav*. 2017;21(9):2703–2715.

Hosek S, Celum C, Wilson CM, Kapogiannis B, Delany-Moretlwe S, Bekker LG. Preventing HIV among adolescents with oral PrEP: observations and challenges in the United States and South Africa. *J Int AIDS Soc*. 2016;19(7 Suppl 6):21107.

Hosek S, Henry-Reid L. PrEP and Adolescents: The Role of Providers in Ending the AIDS Epidemic. *Pediatrics*. 2020;145(1):e20191743.

Hosek SG, Landovitz RJ, Kapogiannis B, et al. Safety and feasibility of antiretroviral preexposure prophylaxis for adolescent men who have sex with men aged 15 to 17 years in the United States. *JAMA Pediatrics*. 2017;171(11):1063–1071.

Lewis-de Los Angeles CP, Alpert KI, Williams PL, et al. Deformed subcortical structures are related to past HIV disease severity in youth with perinatally acquired HIV infection. *J Pediatric Infect Dis Soc*. 2016;5(suppl 1):S6–S14.

Li X, Harrison SE, Fairchild AJ, Chi P, Zhao J, Zhao G. A randomized controlled trial of a resilience-based intervention on psychosocial well-being of children affected by HIV/AIDS: Effects at 6- and 12-month follow-up. *Soc Sci Med*. 2017;190:256–264.

Macapagal K, Kraus A, Korpak AK, Jozsa K, Moskowitz DA. PrEP awareness, uptake, barriers, and correlates among adolescents assigned male at birth who have sex with males in the U.S. *Arch Sex Behav*. 2020;49(1):113–124.

Malee KM, Chernoff MC, Sirois PA, et al. Impact of perinatally acquired HIV disease upon longitudinal changes in memory and executive functioning. *J Acquir Immune Defic Syndr*. 2017;75(4):455–464.

Mellins CA, Brackis-Cott E, Dolezal C, Abrams EJ. Psychiatric disorders in youth with perinatally acquired human immunodeficiency virus infection. *Pediatr Infect Dis J*. 2006;25(5):432–437.

Mpango RS, Kinyanda E, Rukundo GZ, Levin J, Gadow KD, Patel V. Prevalence and correlates for ADHD and relation with social and academic functioning among children and adolescents with HIV/AIDS in Uganda. *BMC Psychiatry*. 2017;17(1):336.

Mpango RS, Rukundo GZ, Muyingo SK, Gadow KD, Patel V, Kinyanda E. Prevalence, correlates for early neurological disorders and association with functioning among children and adolescents with HIV/AIDS in Uganda. *BMC Psychiatry*. 2019;19(1):34.

National HIV Curriculum “Epidemiology” of HIV. https://www.edu/go/screening-diagnostic/epidemiology/cove-concept/all; 2020.

Nguyen N, Choi CJ, Robbins R, Korich R, et al. Psychiatric trajectories across adolescence in perinatally HIV-exposed youth: the role of HIV infection and associations with viral load. *AIDS*. 2020;34(8):1205–1215.

Nichols SL, Chernoff MC, Malee KM, et al. Executive functioning in children and adolescents with perinatal HIV infection and perinatal HIV exposure. *J Pediatric Infect Dis Soc*. 2016;5(suppl 1):S15–S23.

Phillips NJ, Hoare J, Stein DJ, Myer L, Zar HJ, Thomas KGF. HIV-associated cognitive disorders in perinatally infected children and adolescents: a novel composite cognitive domains score. *AIDS Care*. 2018;30(sup1):8–16.

Robbins RN, Zimmerman R, Korich R, et al. Longitudinal trajectories of neurocognitive test performance among individuals with perinatal HIV-infection and -exposure: adolescence through young adulthood. *AIDS Care*. 2020;32(1):21–29.

Skeen S, Sherr L, Croome N, et al. Interventions to improve psychosocial well-being for children affected by HIV and AIDS: a systematic review. *Vulnerable Child Youth Stud*. 2017;12(2):91–116.

Smith R, Huo Y, Tassiopoulos K, et al. Mental health diagnoses, symptoms, and service utilization in US youth with perinatal HIV Infection or HIV Exposure. *AIDS Patient Care STDS*. 2019;33(1):1–13.

Tanner MR, Miele P, Carter W, et al. Preexposure prophylaxis for prevention of HIV acquisition among adolescents: clinical considerations, 2020. *MMWR Recomm Rep*. 2020;69(3):1–12.

Van Den Hof M, Ter Haar AM, Scherpbier HJ, et al. Neurocognitive development in perinatally human immunodeficiency virus-infected adolescents on long-term treatment, compared to healthy matched controls: a longitudinal study. *Clin Infect Dis*. 2020;70(7):1364–1371.

Whiteley L, Brown LK, Mena L, Craker L, Arnold T. Enhancing health among youth living with HIV using an iPhone game. *AIDS Care*. 2018;30(sup4):21–33.

54.5 Adolescent Substance Use Disorders

Sean R. Lenoue, M.D., and Paula D. Riggs, B.A., M.A., M.D.

Adolescent substance use disorders (SUDs) are often multifaceted in their etiology and treatment approach. In the United States, the economic impact of substance misuse approaches $450 billion annually. The overall negative impact of substance use on the lives of youth, individuals, and our communities is incalculable. Given the increasing number of states that have legalized marijuana for medical and recreational use, adolescents are exposed to and use higher potency marijuana products. The opioid epidemic continues to devastate the United States, and total drug overdose deaths peaked at the highest recorded level in 2020. Research advances have increased our understanding of the neurobiology of addiction and behavior change, as well as genetic and environmental contributions to adolescent-onset SUD. These advances have informed the development of more effective substance prevention and treatment interventions. Clinicians working in pediatric mental health settings can play an important role in the prevention, identification of risk factors, and early intervention for youth who are at risk for developing SUD. This chapter focuses on specific aspects of the etiology, psychoeducation, prevention, screening, common comorbidities, and treatment relevant to adolescent-onset SUD.

SUDs and other psychiatric disorders commonly co-occur in part due to similar antecedent environmental risk factors, and both are influenced by adversities during critical developmental periods. Generally speaking, most childhood-onset psychiatric disorders increase the risk of developing an adolescent-onset SUD which complicates the clinical management of preexisting psychiatric disorders and increases the risk of developing new-onset mental health problems. Approximately 80% of people struggling with substance use issues begin using substances as teenagers. The majority of adolescents in the United States experiment with drugs or alcohol before graduating from high school, and the estimated risk of developing a SUD by the age of 18 years are 23.8%. Adolescent-onset SUDs are more likely to persist into adulthood and increase the risk of mood and anxiety disorders as well as antisocial personality disorder. Half of all psychiatric disorders begin before the age of 15, and three-quarters have onset by age 24. Thus, both SUD and non-SUD psychiatric disorders are largely pediatric-onset, and there is a bidirectional relationship between the two. Most childhood-onset psychiatric disorders (particularly attention-deficit/hyperactivity disorder [ADHD], trauma, and oppositionality) increase the risk for adolescent-onset SUDs. Adverse childhood experiences (ACEs), including maltreatment and trauma, also increase the risk of developing mental health problems and adolescent-onset SUD (Douglas et al., 2010; Felitti et al., 1998). The onset of SUD during adolescence complicates the clinical management of preexisting psychiatric disorders and increases the risk of developing new psychiatric symptoms and medical problems. Repeated substance use during adolescence increases the risk of addiction to other drugs tried later and interferes with the rapid brain development that occurs throughout adolescence into young adulthood. After initiation, repeated substance alters the brain in ways that contribute to further drug use. Initially, substances of abuse increase firing in the dopaminergic pathways of the ventral striatum, producing a subjective “high” that promotes further drug use. However, prolonged, repeated use of the drug gradually diminishes dopaminergic neurotransmission, producing less intense “highs” and greater postuse dysphoria. At this stage, the neural circuitry in the dorsal striatum, rather than the ventral striatum, drives the preoccupation with drugs, decision making, and habitual use, making it much more difficult to consider or achieve abstinence. Thus, prevention, early identification, and intervention are crucial to efforts to prevent the progression to more chronic and severe substance and mental health problems as adults.

Table 54.5–1.
Psychoeducation Resources for Clinicians, Youth, and Families

- Preventing Drug Use among Children and Adolescents (National Institute on Drug Abuse; National Ins Department of Health and Human Services)
  - https://www.drugabuse.gov/sites/default/files/preventingdruguse_2.pdf
- Treatment Considerations for Youth and Young Adults with Serious Emotional Disturbances and Serio Co-occurring Substance Use
  - https://store.samhsa.gov/sites/default/files/SAMHSA_Digital_Download/Treatment%20for%20Yout Adults%20with%20Mood%20Disorders%20and%20other%20Serious%20Emotional%20Disturban occurring%20Substance%20Use.pdf
- NIDA Facts for different substance use disorders (written for patients and parents in English and Span
  - https://www.drugabuse.gov/publications/drugfacts/understanding-drug-use-addiction
  - https://www.drugabuse.gov/drug-topics
- SAMHSA guide to treatment resources
  - https://www.samhsa.gov/resource-search/ebp
- NIDA interactive games
  - https://teens.drugabuse.gov/teens/games/drug-facts-challenge

At-risk children can often be identified as young as 3 to 5 years of age, often presenting with impulsivity, aggression, hyperactivity, inattention, poor persistence, and emotional lability. This constellation of symptoms, characterized as “behavioral dysregulation,” is highly heritable. Such children often continue to have poor self-control and disruptive behavioral disorders in later childhood and are at high risk for developing SUDs during adolescence. Psychiatrists and other behavioral health clinicians play an important role in the early identification and treatment of childhood trauma/adversities, childhood-onset psychiatric disorders, behavioral and learning/academic problems, and other risk factors to reduce the risk of adolescent-onset SUD. Behavioral health clinicians also play an important role in providing youth and families with research-based psychoeducation and prevention of alcohol, nicotine/tobacco, marijuana, and other SUDs. Excellent psychoeducation resources are found on the NIDA website, and we list some examples in Table 54.5–1.

PREVALENCE OF SUBSTANCE USE

The Monitoring the Future (MTF) survey annually reports substance use trends in a nationally representative sample of more than 11,000 8th-, 10th-, and 12th-grade students from 112 schools. Key takeaways from the most recent MTF report are as follows:

- Vaping: Remains consistent for nicotine and marijuana vaping, elevated overall since 2017 for 8th, 10th, and 12th graders.
- Marijuana: Remains consistent for 8th, 10th, and 12th graders over the past decade, with little change in rates of daily- and past-year use.
- Alcohol: Although adolescent alcohol use and binge drinking have declined over the past decade, alcohol continues to be the most commonly abused substance by adolescents, with more than 50% of high school seniors and more than 40% of 10th graders reporting alcohol use in the past year; 17% and 10%, respectively reporting binge drinking in the past 2 weeks.
- Illicit/Other Drugs: continued decline in adolescent illicit substance use, including heroin and nonmedical use of prescription opioids, with the exception of a slight upward trend in amphetamine, inhalant, and cough medicine misuse among 8th graders between 2019 and 2020.

Nicotine/Cigarette Smoking

While cigarette smoking has reached an all-time low among adolescents, the prevalence of nicotine vaping has risen dramatically in the past 5 years. Research shows that nicotine vaping quadruples the risk of cigarette smoking. Thus it is not yet known whether the increase in nicotine vaping will reverse or slow the downward trend in cigarette smoking in the future (Johnston et al., 2021). The past-year prevalence of nicotine vaping among high school students nearly doubled between 2017 and 2020. The 2020 MTF survey indicates that nearly 35% of high school seniors and 31% of 10th-grade students reporting nicotine vaping in the past year; 8.6% and 5.6%, respectively, reported daily or near-daily nicotine vaping. Recent studies indicate that adolescents who vape nicotine or use e-cigs are three and a half times more likely to use marijuana compared to non-vape/e-cigarette users.

Marijuana

The dramatic increase in nicotine and marijuana vaping (e.g., e-cig; vape devices) since 2017 continues to be an area of significant concern for substance use in youth. Nicotine, cannabis, and stand-alone flavoring are the three primary “vape” products often sought by adolescents. The increased availability and popularity of e-cig and vaping devices and expanding legalized cannabis environment have likely contributed to the rising trend among adolescents, co-use/vaping of nicotine, and marijuana. Results from a recent prospective study of more than 3,000 high school students indicated that nearly 60% of youth who reported vaping were co-using (vaping) both nicotine and marijuana. Only about 23% and 17% were vaping only marijuana and nicotine, respectively. Similar trends have been reported in Colorado and other states that have legalized cannabis for recreational use. Flavorings, preservatives, and carrier components (e.g., vitamin E acetate) of vape products place users at risk for a variety of possible health consequences such as bronchiolitis obliterans or “popcorn lung” and e-cigarette or vaping product use-associated lung injury (EVALI).

Although still early, cannabis legalization has not (yet) been associated with a significant increase in the prevalence of marijuana use among adolescents. However, cannabis legalization has contributed to a dramatic rise in the potency of cannabis products. In commercial cannabis markets, smoked “flower” contains nearly 25% tetrahydrocannabinol (THC) and concentrates (e.g., “dabs,” “wax,” “shatter”) ranging between 70% and 90% THC. The market share of concentrates has risen sharply over the past 5 years and currently represents about one-quarter of cannabis products sold in commercial markets. This has likely contributed to the increasing use of concentrates among teens, especially in marijuana-legalized states. Compared to states that have not legalized marijuana, legalized states report higher rates of adolescent and adult first-time marijuana users and lower perceived risks associated with cannabis use. Colorado and other “early adopter” legalized states have also reported a significant increase in marijuana-related suicides, hospitalizations; emergency department admissions, and poison center calls postlegalization.

Ample research shows that cannabis is addictive; 1 out of 6 adolescents and 1 out of 11 adults who use/experiment will progress to addiction/moderate–severe cannabis use disorder (CUD). Higher THC concentrations are associated with greater addictive potential (abuse liability) and increased psychosis risk, particularly in adolescents. Regular cannabis use in individuals less than 17 years of age increases the risk of psychosis, negatively impacts motivation, drive, and self-efficacy, and is associated with persistent neurocognitive deficits and reductions in adult intelligence quotient (IQ) that may not be fully reversible even with abstinence. There is also a dose-relationship between the frequency of marijuana use before age 17 and poorer academic achievement, greater risk of high school dropout, as well as unemployment or underemployment in adulthood. There is also growing evidence that adolescent cannabis use increases the risk of developing depression, anxiety disorders, and addiction to other drugs in young adulthood. Increasing access to high-potency marijuana through legalization and relaxed regulatory policies, greater social acceptability, and lowered perceived risk are factors that likely contribute to continued and greater interest in marijuana by youth and greater initiation of first-time users. Thus, it is increasingly important that psychiatrists and other clinicians working in pediatric mental health settings have adequate training and research-based knowledge of the risk factors and risks associated with the use of cannabis and other substances during adolescent development.

CLINICAL EVALUATION

Behavioral health clinicians can play an important role in identifying youth who are likely to develop or have substance problems, as well as motivating and encouraging behavior change. This section provides clinicians with information and clinical tools to assist with fulfilling this potential, including discussion of valid substance screening measures, comprehensive clinical and diagnostic evaluation, brief evidence-based interventions, research-based psychoeducation resources for youth and families, and "warm handoff" procedures for referring adolescents to subspecialty substance treatment providers and coordinating care.

Screening

Screening, brief intervention, referral, and treatment (SBIRT) initiatives and programs have been developed and adopted in many states to encourage primary care providers to utilize standardized screening procedures and assessment instruments. Although SBIRT protocols were initially developed for implementation in pediatric primary care settings, over the past decade, SBIRT has been adapted and implemented in a broad range of clinical settings, including outpatient mental health settings. Several brief substance screening and assessment tools have been developed to assist clinicians in identifying high-risk youth and/or those who have started using or abusing substances, including the Brief Screener for Tobacco, Alcohol, and Other Drugs (BSTAD) and Screening to Brief Intervention (S2BI). Both are valid measures designed for self- or clinician administration. The S2BI uses a single stem question, "How many times in the past year have you used"? with four response options (e.g., "never," "once or twice," "monthly," "weekly or more") for nicotine/tobacco, alcohol, marijuana, illicit drugs, inhalants, synthetic drugs (K2, spice), nonmedical use of prescription medications. Both the BSTAD and S2BI may be self- or clinician-administered. In either case, it is important for clinicians to review or administer, interpret, and respond to substance screening results in the context of a confidential clinical evaluation with an adolescent alone (without a parent). If the adolescent reports "no use" → positive reinforcement; "once or twice in the past year" → brief advice using MI; "monthly" → likely to meet DSM-5 criteria for mild SUD → brief intervention or referral; "weekly or more" → likely to meet DSM-5 diagnostic criteria for moderate–severe SUD → referral to subspecialty substance treatment may be clinically indicated if DSM-5 diagnostic criteria for SUD are confirmed.

Brief interventions have been shown to have a modest effect size on reducing substance use and/or preventing progression to more serious SUD when delivered by SBIRT-trained clinicians. Several no-cost or low-cost online resources for training in brief intervention and motivational enhancement approaches can be found on the SAMHSA website (https://store.samhsa.gov/product/TIP-34-Brief-Interventions-and-Brief-Therapies-for-Substance-Abuse/SMA12-3952).

For youth who do not respond to brief interventions, progress to more serious substance involvement, or meet diagnostic criteria for moderate–severe SUD, referral to subspecialty substance treatment may be clinically indicated. It is important to develop "warm handoff" procedures to increase the likelihood that referred youth will engage in substance treatment. "Warm handoff' and coordinated care procedures may be collaboratively developed by partnering with an adolescent substance treatment program or provider offering evidence-based substance treatment. Adolescent engagement in substance treatment may be enhanced by first referring the teen for a brief, confidential expert consultation/evaluation with the subspecialty substance treatment provider as opposed to "referral to treatment." A brief initial consultation/evaluation provides an opportunity to assess the adolescent's "readiness for change" and use motivational enhancement approaches to respectfully address any questions/concerns and further explore and enhance motivation for change. An example of this approach can be heard in the Charting Pediatrics podcast: *When SUD Come to Pediatrics* (S4:E16) (https://www.childrenscolorado.org/health-professionals/professional-resources/charting-pediatrics-podcast/substance-use-disorder/).

Prevention

The Institute of Medicine (IOM) recognizes three categories of prevention interventions: (I) Universal prevention targets the population as a whole); (II) Selective prevention targets high-risk groups of individuals; and (III) Indicated prevention targets individuals who manifest risk behaviors or early signs and symptoms of the illness. According to the most recent Cochrane review, only the following three interventions were determined to have sufficient evidence to be deemed efficacious or effective: (I) The Unplugged Program (directed toward 12 to 14 year olds); (II) Life Skills Training Program (directed toward 13 to 17 year olds) and (III) Good Behavior Game (directed toward 6 to 7 year olds). All are universal, school-based prevention interventions that are generally adopted school-wide and delivered by teachers trained in the intervention and delivered in the classroom setting. The Cochrane review did not identify any selective or indicated interventions or non–school-based/multicomponent community-based prevention interventions meeting rigorous criteria for effectiveness. However, recent studies not included in the Cochrane review but meeting similarly rigorous criteria support the efficacy of at least one selective school-based prevention intervention known as the Project Toward No Drug Abuse (directed toward 14 to 17 year olds), and one indicated school-based prevention intervention known as the Teen Marijuana Check-Up (TMCU). The National Institute on Drug Abuse (www.drugabuse.gov) has developed a more detailed description of evidence-based substance prevention interventions and programs as well as prevention principles that can be adapted for clinical practice (https://www.drugabuse.gov/publications/preventing-drug-use-among-children-adolescents/prevention-principles) included in the psychoeducation resources section of this chapter.

Treatment

As noted earlier, one of the best treatment strategies is prevention and early identification of risk factors for SUDs. Given mental health disorders and SUDs can often occur together and affect one another, treatment for co-occurring mental health disorders and SUDs needs to be multifaceted as well. There are a host of evidence-based practices to aid providers in most effectively treating their patients. The recently published “Treatment Considerations for Youth and Young Adults with Serious Emotional Disturbances/Serious Mental Illnesses and Co-occurring Substance Use” is an excellent resource for community providers and experts who have clients/patients with co-occurring disorders. This resource guide also includes principles and practical tips that can be adapted and applied in developing coordinated care models. Psychosocial interventions and psychotherapy remain the cornerstone of treatment for co-occurring MH and SUDs. When indicated, medication treatment strategies may also be appropriate. The combined use of therapy and evidence-based medications for the treatment of MH/SUDs may be indicated for moderate-to-severe conditions.

Most psychosocial substance treatment interventions address drug-taking in general. Whereas treatment for withdrawal syndromes or other medications used in substance treatment is generally substance-specific or used to treat co-occurring psychiatric disorders (e.g., ADHD, mood disorders). Compared to numerous studies in adults, relatively few randomized controlled trials have evaluated the safety and efficacy of medications for substance abuse (Table 54.5–2) or co-occurring psychiatric disorders (Table 54.5–3) in adolescents.

Evidence-based psychosocial treatments for adolescents with SUD supported by at least one randomized controlled trial include motivational enhancement therapy (MET)/cognitive behavioral therapy (CBT) (both group and individual format), and family-based interventions including multisystemic therapy (MST), functional family therapy (FFT), multidimensional family therapy (MDFT), brief strategic family therapy (BSFT), adolescent community reinforcement approach (ACRA), MET/CBT, and MET/CBT + contingency management (CM). Each of these treatment modalities has demonstrated efficacy based on the results of at least one randomized controlled trial. Overall, these interventions show comparable but modest treatment effect sizes on reductions in substance use but relatively low rates of abstinence, ranging from <20% (most family-based and group MET/CBT interventions) to 30% (individual MET/CBT). At least two randomized controlled trials have reported that approximately 50% achieve abstinence during the treatment when CM/abstinence-based incentives are added to individual MET/CBT compared to 30% abstinence with MET/CBT alone. This is consistent with the results of numerous studies in adults. Moreover, MET/CBT has most consistently been associated with sustained or increasing post-treatment effect size in both adolescents and adults with SUD, and there is considerable consensus that MET/CBT + CM is currently the most effective treatment for CUDs.

Table 54.5–2.
Medications Used in the Treatment of Substance Use Disorders (SUDs) in Adolescents

Medication	Randomized Controlled Trials	Targeted SUD	Main Outcomes
N-Acetylcysteine (NAC)	Gray et al., 2010	CANNABIS USE DISORDER	Significantly greater reductions in cannabis cravings and cannabis use.
Buprenorphine	Woody et al., 2008	OPIATE USE DISORDER	Randomized 152 opioid-dependent adolescents and young adults in a 12-week study. Randomized participants either received a 2-week detoxification with buprenorphine-naloxone (Bup-Nal) or 9 weeks of Bup-Nal maintenance followed by a 3-week taper. Those maintained on Bup-Nal were significantly more likely to remain in the study at 12 weeks. 70% vs. 21% and had significantly less injection drug use and use of opioids or marijuana. However, the 69- and 212-month follow-up groups were not significantly different in reported opioid use, cannabis use, or injection drug use. Thus, 3 months Bup-Nal did not alter the long-term course, suggesting that even longer-term maintenance may be indicated; additional research is needed. This medication is approved in the United States for the treatment of opioid dependence in those 16 years and older.
Bupropion Nicotine Replacement Therapy (NRT)	Moolchan et al., 2005 Muramoto et al., 2007	TOBACCO USE DISORDER (NICOTINE DEPENDENCE)	Bupropion and NRT alone and in combination have been shown to be well tolerated with a relatively good safety profile in adolescents (Moolchan et al., 2005; Muramoto et al., 2007). There is evidence of efficacy in adolescents for NRT delivered as a patch (Moolchan et al., 2005) and for bupropion at 300 mg/day

dose in at least the short term (Muramoto et al., 2007).

Table 54.5–3.

Medications Treating Co-Occurring Psychiatric Disorders in Adolescents with Substance Use Disorders

Medication	RCT	Targeted Disorder	Main Outcomes
Fluoxetine	Riggs et al., 2007	MDD	Adolescents (*n* = 123) aged 13–19 years were randomized to fluoxetine (20 mg daily dose) or a matching placebo. Participants in both groups received weekly individual MET/CBT as outpatient substance treatment throughout the 16-week trial. Results showed that compared to placebo, subjects treated with fluoxetine had significantly greater reductions in depression severity based on CDRS-R scores (primary depression outcome measure) and higher rates of depression remission (70% vs. 52%). Higher than expected depression remission in both groups suggests that the individual MET/CBT (targeting SUD) received by participants in both groups may have contributed to depression treatment response. Significant reductions in substance use were seen in both groups, but no group difference. Regardless of treatment assignment, participants whose depressions remitted by the end of treatment showed significantly greater reductions in substance use compared to nonremitters, whose substance use did not significantly decline from baseline levels of use despite comparable treatment compliance and completion.
Lithium carbonate	Geller et al., 1998	Bipolar	Lithium treatment was efficacious for both bipolar disorder as well as comorbid substance use disorder in adolescence. Longer-term follow-up study needed.

Pemoline	Riggs et al., 2004	Attention-deficit/hyperactivity disorder (ADHD)	Adolescents (*n* = 69) with ADHD and SUD were recruited from the community and randomized to pemoline titrated to a once daily 37.5-mg dose or matching placebo. Reductions in ADHD symptom severity were significantly greater in the pemoline treatment group compared to placebo on one (Clinician Global Impression- Improvement [CGI-I]) ratings but not both primary outcome measures (ADHD DSM-IV symptom checklist scores, adolescent report). Substance use did not decrease significantly in either group in the absence of specific psychosocial treatment for SUD.
Osmotic-release methylphenidate (OROS-MPH)	Riggs et al., 2011	ADHD	Adolescents (*n* = 303) with ADHD and at least one non-tobacco SUD were randomized to OROS-MPH (titrated to 72 mg daily dose) or a matching placebo. Participants in both groups participated in weekly individual motivational enhancement therapy (MET)/cognitive behavioral therapy (CBT) targeting SUD throughout the 16-week medication trial. Results showed significant reductions in ADHD symptom severity (ADHD-RS scores; primary outcome measure) and days of past 28-day substance use but no difference between groups. OROS-MPH was superior to placebo on some secondary outcome measures: (1) greater improvement in problem-solving and coping skills related to their drug use and (2) greater number of negative urine drug screens. Results suggest that adolescents in

			substance treatment (even if not yet abstinent) may benefit from concurrent treatment of co-occurring ADHD with OROSMPH (which showed a relatively good safety profile and low abuse/diversion liability with regular monitoring in this setting).
Atomoxetine	Thurstone et al., 2010	ADHD	Adolescents (*n* = 70) adolescents referred to outpatient substance treatment and meeting diagnostic criteria for ADHD, and at least one non-tobacco SUD were randomized to atomoxetine or a matching placebo. Participants in both groups concurrently received individual MET/CBT targeting SUD throughout the trial. Results showed significant reductions in ADHD symptom severity (ADHD-RS) and substance use but no difference between groups on primary ADHD or substance outcome measures. Adverse events were consistent with those reported for atomoxetine in non–substance-involved adolescents with ADHD despite nonabstinence in most participants.

An important limitation of existing substance treatment interventions is that none specifically address or integrate, or concurrently treat SUD and co-occurring psychiatric disorders, even though the vast majority of adolescents referred for substance treatment have psychiatric comorbidity. In addition to longstanding systemic and economic barriers, relatively few studies have focused on the development and testing of integrated treatment approaches. Only five randomized controlled trials, enrolling at least 60 subjects (≥30 subjects per group), have evaluated the safety and efficacy of medications for a co-occurring psychiatric disorder in adolescents with SUD. These studies are cited with brief summaries of the main outcomes in Table 54.5–3.

SUMMARY, DISCUSSION, AND CONCLUSION

Over the past decade, the opioid crisis, expanding legalized cannabis environment, and the COVID-19 pandemic underscore the critical role of pediatric primary care providers and mental health clinicians as “front line” first responders with regard to substance abuse prevention, screening, early intervention, and treatment. These events have brought an even sharper focus on the critical shortage of child and adolescent psychiatrists and behavioral health clinicians and the even greater shortage of clinicians with dual training and experience in treating co-occurring psychiatric disorders and SUDs. Thus, in addition to usual updates on trends in adolescent substance use and current research, this chapter includes a number of clinical tools, psychoeducation, and training resources to further bolster the substance-related clinical skills of clinicians working in these settings. Over the next decade, we can anticipate continued momentum toward greater integration of behavioral healthcare services in schools and primary care medical settings and a continued burgeoning of behavioral health innovations to enhance scalability and access to evidence-based mental health and substance prevention and treatment interventions (e.g., AI-informed mobile apps, online or web-based interventions) and clinical training opportunities to address the clinical workforce shortage. It will be important for today's clinicians to stay abreast and informed of trends driving the evolution of the field and clinical practice.

FURTHER READINGS

Allen JG, Flanigan SS, LeBlanc M, et al. Flavoring chemicals in e-cigarettes: diacetyl, 2,3-pentanedione, and acetoin in a sample of 51 products, including fruit-, candy-, and cocktail-flavored e-cigarettes. *Environ Health Perspect*. 2016;124:733–739.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. *DSM-5*. 5th ed. American Psychiatric Association; 2013.

Bukstein OG, Bernet W, Arnold V, et al. Practice parameters for the assessment and treatment of children and adolescents with substance use disorders. *J Am Acad Child Adolesc Psychiatry*. 2005;44:609–621.

Centers for Disease Control and Prevention (CDC). 2020a. Overdose deaths accelerating during COVID-19. Accessed March 12, 2021. https://www.cdc.gov/media/releases/2020/p1218-overdose-deaths-covid-19.html#:~:text=Over%2081%2C000%20drug%20overdose%20deaths,Control%20and%20Prevention%20(CDC)

Centers for Disease Control and Prevention (CDC). 2020b. Outbreak of lung injury associated with the use of E-cigarette, or vaping, products. Accessed April 20, 2021. https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html

Chadwick B, Miller ML, Hurd YL. Cannabis use during adolescent development: susceptibility to psychiatric illness. *Front Psychiatry*. 2013;4:129.

Dalwani MS, Tregellas JR, Andrews-Hanna JR, et al. Default mode network activity in male adolescents with conduct and substance use disorder. *Drug Alcohol Depend*. 2014;134:242–250.

D'Onofrio G, Degutis LC. Integrating Project ASSERT: a screening, intervention, and referral to treatment program for unhealthy alcohol and drug use into an urban emergency department. *Acad Emerg Med*. 2010;8:903–911.

Douglas KR, Chan G, Gelernter J, et al. Adverse childhood events as risk factors for substance dependence: partial mediation by mood and anxiety disorders. *Addict Behav*. 2010;35(1):7–13.

Felitti VJ, Anda RF, Nordenberg D, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults – The adverse childhood experiences (ACE) Study. *Am J Prev Med*. 1998;14:245–258.

Geller B, Cooper TB, Sun KAI, et al. Double-blind and placebo-controlled study of lithium for adolescent bipolar disorders with secondary substance dependency. *J Am Acad Child Adolesc Psychiatry*. 1998;37(2):171–178.

*Johnston LD, Miech RA, O'Malley PM, Bachman JG, Schulenberg JE, Patrick ME. *Monitoring the Future National Survey Results on Drug use 1975–2020: Overview, key Findings on Adolescent Drug use*. Institute for Social Research, University of Michigan; 2021.

Koob GF, Volkow ND. Neurocircuitry of addiction. *Neuropsychopharmacology*. 2010;35(1):217–238.

Lac A, Luk JW. Testing the amotivational syndrome: marijuana use longitudinally predicts lower self-efficacy even after controlling for demographics, personality, and alcohol and cigarette use. *Prev Sci*. 2018;19(2):117–126.

*Lanza HI, Barrington-Trimis JL, McConnell R, et al. Trajectories of nicotine and cannabis vaping and polyuse from adolescence to young adulthood. *JAMA Network Open*. 2020;3(10):e2019181.

Meier MH, Caspi A, Ambler A, et al. Persistent cannabis users show neuropsychological decline from childhood to midlife. *Proc Natl Acad Sci USA*. 2012;109:E2657–E2664.

Merikangas KR, He JP, Burstein M, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the national comorbidity survey replication–adolescent supplement (NCS-A). *J Am Acad Child Adolesc Psychiatry*. 2010;49(10):980–989.

Moolchan ET, et al. Safety and efficacy of the nicotine patch and gum for the treatment of adolescent tobacco addiction. *Pediatrics*. 2005;115.4:e407–e414.

Muramoto ML, Leischow SJ, Sherrill D, Matthews E, Strayer LJ. Randomized, double-blind, placebo-controlled trial of 2 dosages of sustained-release bupropion for adolescent smoking cessation. *Arch Pediatr Adolesc Med*. 2007;161(11):1068–1074.

*National Institute on Drug Abuse (NIDA). 2020c, June 2. Principles of Adolescent Substance Use Disorder Treatment. Accessed May 9, 2021. https://archives.nida.nih.gov/sites/default/files/podat-guide-adolescents-508.pdf

Norman AL, Pulido C, Squeglia LM, Spading AD, Paulus MP, Tapert SF. Neural activation during inhibition predicts initiation of substance use in adolescence. *Drug Alcohol Depend*. 2011;119(3):216–223.

Riggs PD, Hall SK, Mikulich-Gilbertson SK, Lohman M, Kayser A. A randomized controlled trial of pemoline for attention-deficit/hyperactivity disorder in substance-abusing adolescents. *J Am Acad Child Adolesc Psychiatry*. 2004;43(4):420–429.

Riggs PD, Mikulich-Gilbertson SK, Davies RD, Lohman M, Klein C, Stover SK. A randomized controlled trial of fluoxetine and cognitive behavioral therapy in adolescents with major depression, behavior problems, and substance use disorders. *Arch Pediatr Adolesc Med*. 2007;161:1026–1034.

*Riggs PD, Winhusen T, Davies RD, et al. Randomized controlled trial of osmotic-release methylphenidate with cognitive-behavioral therapy in adolescents with attention-deficit/hyperactivity disorder and substance use disorders. *J Am Acad Child Adolesc Psychiatry*. 2011;50(9):903–914.

Rocky Mountain High Intensity Drug Trafficking Area (RMHIDTA) program. The legalization of marijuana in Colorado: the impact: Volume 6, September 2019. *Mo Med*. 2019;116(6):450.

*Substance Abuse and Mental Health Services Administration (SAMHSA). Treatment Considerations for Youth and Young Adults with Serious Emotional Disturbances/Serious Mental Illnesses and Co-occurring Substance Use. Publication No. PEP20-06-02-001. National Mental Health and Substance Use Policy Laboratory, Substance Abuse and Mental Health Services Administration. 2021.

Thurstone C, Riggs PD, Salomonsen-Sautel S, Mikulich-Gilbertson SK. Randomized, controlled trial of atomoxetine for attention-deficit/hyperactivity disorder in adolescents with substance use disorder. *J Am Acad Child Adolesc Psychiatry*. 2010;49(6):573–582.

Volkow ND, Baler RD, Compton WM, Weiss SRB. Adverse health effects of marijuana use. *N Engl J Med*. 2014;370:2219–2227.

Walker D, Stephens R, Roffman R. Randomized controlled trial of motivational enhancement therapy with non-treatment-seeking adolescent cannabis users: a further test of the teen marijuana check-up. *Psychol Addict Behav*. 2011;25(3):474–484.

Walley SC, Wilson KM, Winickoff JP, Groner J. A public health crisis: electronic cigarettes, vape, and JUUL. *Pediatrics*. 2019;143(6):e20182741.

Woody GE, Poole SA, Subramaniam G, et al. Extended vs short-term buprenorphine-naloxone for treatment of opioid-addicted youth: a randomized trial. *JAMA*. 2008;300(17):2003–2011.

54.6 Forensic Child and Adolescent Psychiatry

PETER ASH, M.D., AND ELISSA P. BENEDEK, M.D.

Forensic psychiatry is a subspecialty of psychiatry that focuses on the interface between psychiatry and the law. It may involve consulting to courts, evaluation of individuals involved in the legal system, working in corrections, having involvement in legal issues related to practice, teaching, and performing research in forensic psychiatry. Psychiatrists seeking board certification in forensic psychiatry are now required to complete a 1-year fellowship in an accredited forensic psychiatry training program. No such specialized training is available for child and adolescent forensic psychiatry, although, increasingly, trainees are getting more exposure to the area in residencies, and many forensic fellowships offer some additional experience in child and adolescent forensic psychiatry. Continued practice in this specialized area may involve certification, recertification, and maintenance of certification.

This chapter will focus on conducting psychiatric evaluations intended to provide evidence to be used in litigation. Forensic evaluations differ from clinical evaluations in that there is no doctor–patient relationship with an implicit duty of care, and confidentiality is limited inasmuch as most reports are being prepared for use in court. Forensic evaluations also differ in that evaluations are focused on answering a specific question. Often the question will have a specific legal test, which may vary from jurisdiction to jurisdiction. For example, while many jurisdictions use a “best interests of the child” test in determining custody in divorce proceedings, in other jurisdictions this test has a specific statutory definition that lists a variety of factors to be considered. It is important for evaluators to be clear on the relevant test in their jurisdictions. The forensic examiner approaches each case with a degree of skepticism beyond that used in most clinical settings. Malingering and secondary gain need to be considered in litigants and criminal defendants. To strive for objectivity, the forensic examiner often seeks out a wide base of collateral information which may include legal, medical, psychiatric, and academic records, interviews with other family members, depositions, police reports, and arrest records. As noted, confidentiality and patient advocacy do not exist in forensic evaluations; thus, guiding ethical principles are not the traditional *primum non nocere*—first do no harm—and beneficence. In fact, were the forensic psychiatrist striving to act in the examinee’s best interests, it would skew objectivity. Appelbaum suggests that, in forensic psychiatry, primacy be given to the value of justice based on the ethical principles of truth telling and respect for persons. This primacy applies regardless of whether the forensic psychiatrist has been retained by an attorney for one side or is court appointed.

Dr. A was very concerned and sympathetic as she listened to Becky, a 12-year-old girl, discussing the emotional and physical abuse she claimed she had endured while visiting her father during unsupervised visitation. Later, while interviewing Becky's mother, Dr. A listened to a verbatim account of the abuse. It was only upon later reflection that Dr. A recognized that Becky had been coached to parrot a word-for-word version of her mother's story and that, in her sympathy for this allegedly traumatized young girl, she had lost objectivity.

GUIDELINES

Traditionally, child and adolescent forensic psychiatry had focused on issues of child abuse and neglect, custody, and delinquency. Beginning in the early 1980s, many child and adolescent psychiatrists were drawn into civil suits as expert witnesses in cases alleging child sexual abuse. Some experienced role confusion, in which they were trying to serve as both therapist and expert witness, seemingly unaware of this inherent conflict. Added difficulties were created by the lack of clear guidelines as to how one actually assessed children for suspected sexual abuse and the paucity of good research data regarding the long-term effects of child sexual abuse.

The American Academy of Child and Adolescent Psychiatry (AACAP) has issued practice parameters for forensic evaluations which include welcome principles of practice. Anyone undertaking forensic examinations of children or adolescents should become familiar with the American Academy of Psychiatry and the Law's *Ethical Guidelines of Forensic Psychiatry.*

Civil litigation involving children may involve other types of trauma, malpractice, and class action suits against institutions for neglect or abuse. As more youths enter the juvenile justice system, there is a growing need for forensic psychiatrists who are willing to perform evaluations for the court or work within juvenile corrections providing evaluation and treatment for detainees and committed youths. There is a high prevalence of psychiatric disorders in this population, and many youths remain undiagnosed and untreated.

TREATING CLINICIANS TAKING ON A FORENSIC ROLE

Not uncommonly, a treating clinician may be asked to become involved in litigation involving a child or adolescent patient. For example, divorcing parents who have sought treatment for their child to help the child deal with the stress of divorce may ask the treating clinician to provide the court with an opinion as to whom the child should live with, or a child who is in treatment for the effects of an injury or trauma may be suing the person who allegedly caused the injury and the treater is being asked for opinions on the severity of mental injury and prognosis. Treating clinicians who provide opinions to the court are placed in a dual role situation, that of treater and of forensic evaluator, and these roles may conflict (e.g., the confidentiality of the treatment may be lost, which could impair future treatment if the patient becomes less willing to share their thoughts). Requests for the treater to enter a forensic role may be informal or may come as a subpoena. The fact that the parents typically control the release of information, rather than the child patient, adds to the complexity.

Requests for clinicians to take on a forensic role should be dealt with cautiously. In most situations, it is preferable to have the child patient evaluated by a separate forensic evaluator to keep the treatment and forensic roles separate. This is particularly important if the clinician has not formed an opinion on the legal issue and would need to conduct a forensic inquiry which otherwise would not be part of the treatment.

Dr. J was treating 7-year-old Stacy for anxiety arising out of her divorcing parents contesting her custody. Stacy's mother's attorney asked that Dr. J provide an opinion as to how much time Stacy should spend with her father. Dr. J responded that she had not evaluated Stacy with that in mind, did not have an opinion on that issue, and recommended the parents seek a court-ordered evaluation to obtain recommendations for what custodial arrangement would be in Stacy's best interest.

Treating clinicians should also clarify whether they are being asked to be a fact witness or an expert witness. Expert witnesses can testify as to their opinions; fact witnesses cannot. That a clinician wrote in the chart that the patient has major depressive disorder is a fact; whether the patient has major depressive disorder is an opinion. Treating clinicians can be required to answer questions about their treatment, assuming there is an appropriate release, but in most jurisdictions, a clinician cannot be required to provide expert opinions without the clinician's voluntary agreement. Expert witnesses are usually paid for their time; fact witnesses are not. When a treating clinician offers expert opinions, a cross-examining attorney is likely to try to paint the expert as biased in favor of the patient. A treating clinician who is not experienced in forensic issues or practice may wish to obtain consultation as to how to respond to a request to act in a forensic role.

FORENSIC EVALUATIONS

Referrals

Most referrals for forensic evaluations come from attorneys, and some come directly from the court or from child protection agencies or parents. At the outset, it is important to clarify whom the attorney represents, whether the case is civil or criminal, and what the legal issue is. A useful practice is to ask the referring party to put these questions in writing in a cover letter. This also enables the attorneys to think more carefully about exactly what they want to learn from the forensic examination and refreshes the memory of the examiner doing the forensic examination. Conflicts of interest need to be ruled out in taking a case. These might include having treated a relative of the examinee, being on the faculty of a facility being sued, having social ties with an involved party, or having a personal or work relationship with a physician defendant. Forensic psychiatrists need to consider whether they have sufficient expertise in the area in question. A lack of expertise in the topic being litigated does not necessarily preclude involvement as long as one notifies attorneys of this fact. Some attorneys may be willing to provide background materials on issues not often encountered, such as toxic buildings.

Ms. Smith, an attorney, called Dr. W and informed him that she was representing a hospital that was being sued by the parents of a youth who died by suicide in an adolescent inpatient psychiatric unit. She had heard of Dr. W's fine reputation, both as a therapist of adolescents and as an expert witness, and she hoped that he would take the case. Dr. W immediately recognized a conflict of interest in that he did some of his training at this facility and believed that might compromise his objectivity. Furthermore, the defendant's psychiatrist was someone with whom he had a collegial relationship. Finally, inasmuch as his practice was entirely outpatient, he informed Ms. Smith that he did not feel qualified to testify on issues involving the standard of care in an inpatient unit. Dr. W declined to take the case. Dr. J, however, was an unpaid adjunct professor at the same university, had not practiced inpatient psychiatry for years but was not at all reluctant to become involved in the case and did not see a conflict of interest. Needless to say, Dr. J was cross-examined on just these issues.

Other factors to consider when deciding whether to accept a case are the extent of work anticipated; whether the referring attorney is asking just for record review or wants an evaluation as well; whether a written report is expected; the timeline of the case; and whether it is feasible to meet the case's discovery date—the time when attorneys must provide the other side with their expert's name and opinions. Fees should be discussed and a retainer may be obtained up front. It is always unethical to accept a case on a contingency fee, in that it would compromise objectivity. Plaintiff's lawyers, in contrast, act as advocates for their clients, and so they often work on contingency fees in civil suits.

Informed Consent

Unless a court order requires the evaluee's participation, prior to continuing the interview, informed consent should be obtained. Since, in most cases, minors are deemed legally incompetent to provide informed consent, consent is usually obtained from parents or guardians. However, respect for persons suggests that minors should be given information about the evaluation in a manner suitable for their developmental level, which includes explaining in age-appropriate language who has retained the expert, the purpose of the evaluation, that the evaluator is not there as a treating doctor, the lack of confidentiality, and that the expert may be called to testify. It is important to assist the juvenile in understanding the information presented. Assent to the evaluation should be requested. If the minor does not assent, the clinician needs to consider carefully whether it is appropriate to proceed with the evaluation.

Dorothy, an intellectually disabled 12 year old, said that she understood that the evaluation would not be confidential but then could not explain what confidential meant. The evaluator then made a further effort to explain the concept in simpler language and asked her again to explain it back.

In some cases in which an examination is court-ordered, informed consent need not be obtained, but the juvenile should still be informed of the nature of the evaluation. If a juvenile refuses to cooperate with a court-ordered examination, they should be informed of the possible legal ramifications of failure to comply with a court order. It is prudent to clarify that the forensic examination is not a treatment relationship, and examinees may need to be reminded of that in the course of the evaluation, particularly if they persist in viewing the examiner as being there to help them. If informed consent is required, a written release should be obtained for dissemination of the written report, and the examinee should be made aware of the limited confidentiality of the evaluation.

Written Report

The written report becomes the backbone of any forensic evaluation. Care needs to be taken to ensure that it addresses the relevant legal issues and contains the foundations for the expert's opinions. A thorough, well-written report often leads to an out-of-court settlement, avoiding the cost and trauma of going to trial. If the case proceeds to trial, the report becomes the basis of the expert's testimony. Reports should be free of any inflammatory or prejudicial statements, and parties should be treated with respect. A good check on one's possible bias is to ask, "How might my report have differed if written for the other side?" The expert must be aware that many people besides the judge may read the report, including plaintiffs, defendants, and their parents.

Dr. G wrote in a report that the youngster "had the brain of a banana." Needless to say, this was an inappropriate comment.

Dan, an adolescent whose parents were involved in a contentious child custody dispute, was concerned about the possible information that might be contained in the expert's report of his evaluation. He waited by the rural mailbox daily in an attempt to read the report before his parents and was successful in intercepting it. The examiner had made mention of his latent homosexuality. Dan was distressed about this report with just cause.

CUSTODY ISSUES

Child Custody Disputes Between Parents

Custody conflicts most typically arise between divorcing or divorced parents but may also arise between parents who never married and sometimes between parents and others seeking custody or visitation, such as grandparents and surrogate parents. Since the early 1900s, custody decisions have been guided by the concept of the child's best interests, but states vary in how these are defined, with some states specifying particular factors that need to be addressed, so forensic evaluators need to be aware of the specific test in their jurisdiction. In child custody disputes, the child is not a named party in the dispute, and the expert has an advocacy role in speaking for the child's best interests. For this reason, most experts in the field recommend that the evaluator in a custody dispute not be retained by one of the parties but instead be either court-appointed or stipulated to by both parties to avoid any bias that might arise from being retained by one side or the problem of one parent distrusting the evaluator because of suspected bias.

Unlike most forensic evaluations, which focus on an individual, child custody evaluations require assessments of multiple people and their interactions and typically involve individual interviews of each parent, the children, and observations of the child(ren) interacting with each parent and others who live in the proposed custodial environment, in addition to interviews of others who may be involved or have useful collateral information. Unilateral parental evaluations in divorce situations are of very limited utility in that the examiner has only second-hand, and often biased, information about the parent not seen. Furthermore, the examiner cannot make comparisons between the parents if only one parent has been seen.

Child Abuse and Neglect Cases

Forensic evaluations may be requested at any step in the course of litigation involving child abuse and neglect, from the determination of whether abuse and neglect occurred through the assessment of whether the child should be removed from the home to the question of whether a child who has been removed should be returned or whether parental rights should be terminated. All states require physicians to report suspected child abuse and neglect to the state or county protective services agency, and the agency then conducts an investigation. The investigation often includes a forensic interview of the child. While psychiatrists sometimes are asked to conduct these forensic interviews, the trend over the past 30 years has been to have initial forensic interviews conducted by trained personnel in designated centers rather than refer initial evaluations to child and adolescent psychiatrists. In later litigation, either in the juvenile court or in adult criminal court, psychiatrists may be retained to review the initial forensic interview or to reinterview the suspected child victim. Once abuse or neglect is established, a forensic interview may be sought to assist in planning appropriate interventions. If a child is removed from the home, particularly if the removal is occasioned because of neglect by a seriously mentally ill parent, after the parent receives treatment, an evaluation may be sought as to whether the parent has regained sufficient parenting capacity to be reunited with the child.

Ms. Smith was seen for an evaluation as to whether her parental rights should be terminated with her 5-year-old child after her 1-year-old son was murdered by her live-in boyfriend. Ms. Smith did not believe that her boyfriend had murdered her child as he had not yet been convicted. Her history was replete with past physical and emotional abuse by her boyfriend and others. She was compliant with all evaluation appointments and court-ordered psychiatric counseling until such time as her boyfriend was convicted. Even then, she believed him—she protested his innocence and became less compliant with therapeutic sessions. The evaluator expressed concern about the safety of the 5-year-old child.

Assessments of parental capacity must take into consideration the child's developmental needs, the fit between parent and child, and parental strengths and ability to protect, advocate, discipline, and provide emotional and cognitive support. In addition, the psychiatrist needs to assess the parents' cognitive capacities and social supports, as well as their histories and special issues, such as prior sexual abuse, substance abuse, domestic violence, and mental illness.

Making Recommendations

Recommendations to the court should address the legal issues at hand, be feasible, and be based on the child's developmental needs. Forensic examiners need to remember that the court is the ultimate decision maker and not be offended if their recommendations are not followed. Sometimes this may occur because the court has become aware of new evidence or because of legal issues. However, a report that fails to demonstrate the basis for the expert's reasoning and opinions may also lead to recommendations not being followed.

CIVIL LITIGATION

Torts

A *tort* is a claim of wrong done to another person for which a legal remedy, most often a monetary award for the plaintiff, is sought. The plaintiff must demonstrate by a preponderance of the evidence that there has been a breach of duty owed by the defendant to the plaintiff, as a result of which the plaintiff suffered an injury. Torts may be deliberate, as when an individual sets out to harm someone, or unintentional, as when the act results from negligence (as in malpractice and personal injury cases). The plaintiff may seek compensatory damages for pain and suffering, special damages for medical or psychiatric care, and, in intentional torts, punitive damages. While only a minority of personal injury cases do not settle and actually proceed to trial, in contrast to family law cases, for those cases that do proceed to trial, a jury will typically hear the case, and a long period often elapses before the case goes to trial. The statute of limitations in each state limits the period during which an action may be brought and usually runs from the time when an injury is discovered. However, with juveniles, in most states, the statute does not begin to run until the child reaches the age of majority.

Susan, a 10-year-old girl, was severely bitten by a neighbor's dog and sustained a cosmetic injury to her lower lip. Her father instituted a suit against the neighbor in the role of her best friend. A forensic evaluation revealed that Susan had little concern about her injury. She was doing well in school, had many friends, and had future plans to become a teacher. Her concerns centered on her mentally ill mother, who had been institutionalized and then ran away from the family. Susan wondered about her health and her whereabouts. Susan's father, by contrast, obsessed about her injury.

Post-traumatic Stress Disorder

A common claim is that post-traumatic stress disorder (PTSD) results from the negligent or intentional behavior on the part of a defendant. The forensic psychiatrist may be asked to examine the plaintiff by either side. Sometimes, the forensic psychiatrist's role is limited to record review and critiquing the forensic report of the other side. Evaluating a child or adolescent for alleged PTSD and commenting on issues of causality and future damages is a complex and challenging task. Very young children may have difficulty verbalizing the subjective symptoms of PTSD, and it may manifest itself differently in children than adults. Re-enactment of a trauma may be common, but flashbacks are less likely than with adults. Parents may not be the best observers of their children's behavior and may veer toward minimizing or exaggerating psychopathology. Heightened arousal in PTSD may be confused with attention deficit hyperactivity disorder (ADHD) or coexist with it. Children, especially those living in inner cities or war zones or those exposed to domestic violence, may experience cumulative and chronic PTSD that clouds issues of causality. Prior trauma may sensitize them, thereby heightening their vulnerability to subsequent trauma and the development of PTSD. Age at the time of trauma may affect which symptoms develop but not necessarily symptom severity. The course of PTSD is often affected by parental responses to trauma—hence the need to explore family dynamics and other environmental mediating factors. In addition, the forensic psychiatrist should be aware of studies on the longitudinal course of PTSD in childhood.

Key Points

Key points in the forensic assessment of children involve getting thorough developmental and family histories, including whether the child has experienced prior victimization or witnessed violence or sexuality and, if so, how they have dealt with it. In addition, it is important to note from other sources the exact onset of symptoms. For example, pediatric records may indicate that a child's sleep disturbance and bed-wetting preceded a trauma, contrary to the history given by parents. The forensic psychiatrist should be aware of the salient literature on memory in children and the suggestibility of child witnesses. The child who has been through multiple interrogations may feel that they have not been believed and start to embellish their responses or try to please the interviewer by giving the desired responses. Leading questions and either–or questions that give forced choices can be hazardous.

Questions should be put to the child in an unbiased way that encourages the use of narrative recall. It is important to elicit the child's recall of traumatic events both verbally and through the use of spontaneous play and drawings, which may reveal traumatic themes. The child may evidence behavioral signs of trauma within the interview, such as heightened startle response or, in the case of the sexually abused child, undue familiarity. Although young children are not likely to malinger, statements made by interviewers or parents can readily influence them. Source misattribution, in which children confuse something they have read or been told about with something that actually happened to them, may also lead to erroneous allegations. In research settings, children who were repeatedly urged to think about nonevents eventually came to believe they really happened.

Sylvia, a 4-year-old nursery school child, claimed that she had been repeatedly sexually abused by her nursery schoolteacher. In addition, she reported she had accompanied her classmates to a secret location in a nearby park where there were sacrifices of animals that they were forced to observe. After a transcript of the initial evaluator's interview of Sylvia was obtained, it was possible to conclusively opine that she had been asked leading questions repeatedly and then complimented when she gave the desired answers. In addition, she had been told that other classmates had told the interviewer about their horrific experiences with the nursery schoolteacher.

Pitfalls

Common pitfalls in the forensic evaluation of damages in civil litigation include attributing all of a child's symptoms to the trauma being litigated, failure to get an adequate prior history, a bias that leads to selective interviewing, the failure to pursue disconfirming data, asking numerous leading questions, and the need to give attorneys the opinions they seek. Some examiners immediately equate trauma with PTSD and fail to consider other diagnoses or even whether the trauma in question meets the threshold of trauma necessary to meet the diagnostic criteria of the latest diagnostic manual (DSM-5-TR). Similar problems may be encountered with plaintiffs alleging trauma-related depression or other psychiatric disorders.

Dr. A was called by a defense attorney to review discovery material in a case that alleged permanent harm and suffering in 6-year-old Tony, who was alleged to have been sexually abused at age 3 years in his daycare center. Dr. G, the forensic expert for the plaintiff, had evaluated the child and performed psychological testing on him and concluded that the boy's conduct problems were all related to the alleged abuse, which the child had difficulty recalling. However, his early history of the boy was cursory, and he had little information about the mother, who was a single parent and did not review medical records. In her thorough review of discovery material, Dr. A learned that Tony had witnessed extensive domestic violence and his mother's rape, shown signs of hyperactivity since age 2 years, and exhibited much anxiety related to his mother's safety and several separations from her at times when she was unable to care for him owing to depression. Tony also had delayed language development. Dr. G, at the time of his deposition, was asked why he had not asked about these matters. He said he considered the mother's personal life a private matter and did not see its relevance to the litigation. Dr. A, when deposed, pointed out that many other factors besides the alleged abuse might account for Tony's behavioral problems.

JUVENILE JUSTICE CASES

Overview

The U.S. juvenile justice system, which is more than 100 years old, is in a state of rapid change. The founding premise that wayward youths could be rehabilitated through education, special services, treatment, and redirection has been challenged as more hard-core and violent youths enter a juvenile justice system that lacks sufficient dispositions and funding. Until fairly recently, there have been little data on effective treatments and a paucity of adequate dispositions and funding for this population. In response to the upsurge of adolescent violence, which peaked in the early 1990s, out of a concern for public safety, many states passed mandatory waiver statutes which typically specified that any juvenile over a certain age who was charged with certain specified crimes would automatically be waived to adult jurisdiction. Other forms of waiver to adult court include judicial waiver, in which a juvenile is waived after a hearing, and discretionary waiver, in which the district attorney, by statute, is allowed discretion to prosecute certain classes of juvenile defendants in either juvenile or adult court. Juvenile crime has actually decreased dramatically over the last three decades, and in response, there is a trend to loosen some of the broad waiver statutes. Studies suggest that adolescent defendants waived to adult jurisdiction are more likely to recidivate once released than youths who committed comparable crimes and remained in the juvenile justice system.

The Juvenile Court

The juvenile court remains a civil court in which judges have wide discretion, procedures remain more informal than in adult court, and records may remain sealed. However, in actuality, there are many loopholes and exceptions to the concept of the sealed juvenile record. If a juvenile is waived to the adult court, their records go along. Juvenile records may also be released to insurers, the military, employers, and for purposes of licensing and national security.

It was not until the 1960s that juveniles were afforded clear due process rights. In a seminal 1967 case, *In re Gault,* Justice Fortas charged that “in most juvenile courts, the child receives the worst of both worlds: That he gets neither the protections accorded to adults nor the solicitous care and regenerative treatment postulated for children.” As a result of *Gault* and subsequent cases, juveniles were afforded most of the rights afforded to defendants in adult criminal court, with the notable exceptions of the right to jury trials or transcripts of proceedings. In 1970, *In re Winship* held that the standard of evidence in juvenile hearings must be beyond a reasonable doubt, the same standard used in the adult criminal court.

Preadjudicatory Examinations

Courts often request psychiatrists to perform preadjudication examinations of juveniles. These are designed to help the judge to make appropriate dispositions. These examinations are similar to diagnostic evaluations but go further in that they also consider mitigating factors, risk factors for continued delinquency, and the risk the youth may pose to the community. In addition, an assessment of amenability to rehabilitation and recommendations as to what sort of treatment the youth needs may have a bearing on the type of facility to which the youth is committed.

Waiver Evaluations

Waiver, or *bindover,* as it is known in some jurisdictions, refers to the transfer of a juvenile from the juvenile court to the adult court, in which proceedings are criminal, not civil; and public, not closed; and sentences may be harsher and extend into adulthood. The are no juvenile courts in the federal system, so minors charged with federal crimes have a waiver hearing to determine whether they will be prosecuted as adults in federal district court or transferred to state juvenile court jurisdiction. Most adolescent defendants who now enter the adult criminal system do so as the result of mandatory waivers or as a result of prosecutorial discretion. With judicial waivers, evaluations are used for determining which court should have jurisdiction over the youth, not for disposition as in a preadjudicatory examination. Thus, the focus is on helping the judge weigh factors that have a bearing on the decision of whether to waive a youth. In addition to developmental and psychosocial history and diagnostic issues, the examiner considers whether the youth has benefited from any interventions, be they therapeutic or within the juvenile justice system, thus far. Amenability to treatment that can be provided within the juvenile system and the ability to contain the youth within the juvenile justice system are also considerations. Special needs that the adult correctional system might not be able to meet need to be identified and consideration also may be given to the age, size, and maturity of the youth. The risk of future violence must be assessed, and short-term prognoses may be made based on known risk factors.

Competency to Stand Trial Examinations

Competency is a term that may be used in reference to many legal issues, such as competency to stand trial, waive *Miranda* rights, plead guilty, and waive the right to counsel. Competency to stand trial in some jurisdictions is called competency to proceed, to more clearly cover issues that arise in a criminal case prior to trial, such as the decision to accept a plea bargain. The court may also request an evaluation in regard to competency to stand trial if waiver seems likely. An adolescent must be competent to stand trial in adult court. Until the mid-1990s, very few states required that a youth be competent to stand trial in juvenile court under the theory that many youths were immature and so incompetent and that the role of the juvenile court was to promote rehabilitation. Since the mid-1990s, an increasing number of states have required that youths be competent to stand trial in juvenile court, as well. Competency examinations focus on the individual's current ability to understand the charges against them and their severity and potential consequences, understand the trial process, be able to work with their attorney in developing a defense, and be able to participate in courtroom proceedings. While in many respects, the evaluation of competency to stand trial of a youth is similar to that of an adult, issues of developmental immaturity are much more likely to render a youth incompetent. Juvenile defendants who are found incompetent to stand trial pose a challenge for the courts in that a major component of restoration is simply to let the minor mature. However, prosecuting and punishing a youth several years after offending who is now in a different developmental stage from how they were when the crime was committed is problematic.

Criminal Responsibility Assessments

In contrast to competency evaluations, which assess the present state of mind, assessments regarding criminal responsibility attempt to reconstruct the defendant's state of mind at the time of the crime. Retrospective assessments of the state of mind are always more difficult and rely on not just the defendant's version of events but also a great deal of discovery material. *Insanity* is a legal term that is used as a defense to relieve defendants of responsibility for the crime. States have different legal definitions of insanity, and several states have no insanity defense. Legal definitions of insanity typically include a cognitive prong that gets at whether the defendant, at the time of the crime, as a result of mental disease or defect, lacked substantial capacity to appreciate the wrongfulness of their conduct. *Substantial* implies serious mental disorders and does not cover Axis II diagnoses or, usually, PTSD. The exact wording of the insanity defense varies from state to state. Some states have an additional volitional prong that refers to the inability of the defendant to restrain themselves as a result of an abnormal state of mind. In addition, the law requires that the defendant possessed a guilty state of mind (*mens rea*)—that is, intention at the time of the crime. Mental abnormalities may affect the ability to form intent or lead to diminished responsibility for the act. Thus, a murder charge might be reduced to manslaughter if the defendant was not capable of premeditation. A few states now permit a guilty but mentally ill plea, which results in the same sentence as if found guilty but supposedly promotes more treatment while incarcerated. While youth waived to adult criminal court can potentially make use of their state's insanity defense, only a minority of states allow the use of an insanity defense in juvenile court.

Sentencing Evaluations

In juvenile court, if the court has already obtained a preadjudicatory evaluation covering the youth's developmental status, treatment needs, and amenability to rehabilitation, that report will often be used in sentencing. Courts or attorneys may utilize similar reports in the sentencing phase if a prior evaluation has not been done or if a supplemental evaluation is deemed useful.

For youth tried as adults, forensic evaluations are increasingly utilized in sentencing decisions. For serious crimes, a series of U.S. Supreme Court cases have changed the landscape with regard to the sentencing of adolescents. In 2005, in *Roper v Simmons,* the Court held that “evolving standards of decency that mark the progress of a maturing society” had changed to the extent that the death penalty for youth under 18 years of age was “cruel and unusual punishment” and thus unconstitutional. The Court cited three factors regarding adolescents that made them less culpable as a class and thus not deserving of death: (1) that adolescents were developmentally immature and thus more impulsive and more prone to respond to peer pressure, (2) that their behavior was often affected by adverse environmental circumstances over which they had no control (such as where they lived and the circumstances of their families), and (3) that their personalities were still in flux and so their actions could not be seen as reflecting enduring character traits. Since *Roper*, there has been a burgeoning of research on the neurologic development of the adolescent brain demonstrating that while those areas of the brain involved in emotion, particularly the limbic system, mature early in adolescence, other areas, especially the prefrontal cortex, that are involved in restraining impulsive behavior and considering consequences, continue to develop into the early 20s. This developmental sequence helps explain why much adolescent antisocial behavior is time limited to adolescence and is used to support the idea that adolescents are both less culpable than adults and are also more likely to be either rehabilitated from or develop out of criminal behavior. Using analyses along these principles, the Supreme Court held that a sentence of life without parole (LWOP) was unconstitutional for all nonhomicide offenders (*Graham v Florida*, 2010), that mandatory LWOP was unconstitutional and could only be imposed for murder following a hearing (*Miller v Alabama*, 2012), that LWOP sentences should be rare and reserved for the most incorrigible offenders (*Montgomery v Louisiana*, 2016), but that an LWOP sentence could be imposed without a judicial finding that an adolescent was incorrigible (*Jones v Mississippi*, 2021). Some states, by either statute or court holdings, have adopted further specific criteria that need to be considered in considering sentencing adolescents to LWOP. Although these court holdings pertain directly to factors to be considered when sentencing adolescents convicted of murder, these factors of reduced culpability of adolescents and the likelihood of their being rehabilitated by their mid-20s have been applied in sentencing hearings of adolescents facing lesser charges and in sentencing hearings of young adults who are slightly older than 18 years of age but who arguably, in many developmental respects, resemble adolescents under age 18, and so provide a useful framework for structuring opinions about mitigating sentencing in a wide variety of adolescent and early adult cases.

Blended Sentences

Some courts, such as New Mexico, have started using criminal sanctions within the juvenile court, and others, such as Michigan, are experimenting with blended sentences by which the youth, on reaching the age of majority, is automatically transferred from a juvenile correctional facility to an adult one. In Minnesota, judges may impose extended juvenile dispositions with stayed criminal sanctions. Courts also use mandatory minimum sentences for certain crimes. A very promising development has been that of youth drug courts for first-time offenders charged with drug-related offenses. Close court supervision is combined with outpatient drug treatment and a team approach, which appears to be reducing recidivism.

Interventions

The last decade has seen many changes in the juvenile justice system. There has been much more attention to the prevalence of mental illness in this population and the implementation of evidence-based treatments for conduct disorders, substance abuse problems, and comorbid psychiatric disorders. A number of professional organizations have been active in developing guidelines for assessment and treatment in juvenile detention facilities. However, many facilities and programs continue to rely on outdated interventions in which juveniles are treated in groups with their deviant peers. Numerous studies attest to the lack of efficacy of many programs for this population, and research has documented the negative effect that delinquent youth have on one another. Innovative community-based treatment programs, such as intensive probation, are proving effective in helping youth on probation cut down on reoffending. Another effective approach is multisystemic therapy, which was designed to target and influence the known correlates of antisocial behavior in youth. This may occur with youth living either at home or in foster placement. It uses intensive outpatient services that focus on problems within the child and their family and environment and promotes prosocial behavior. It is a cost-effective treatment that provides an alternative to out-of-home placement and has been shown to be effective with delinquents. Collaborative problem solving is another intervention that is being used effectively in many juvenile facilities. Therapeutic foster care also has been shown to cut down on recidivism in some studies. A number of school-based programs now have good evidence of efficacy in reducing violence. In addition, diversion programs and juvenile mental health and drug courts are showing promise in keeping youth out of correctional facilities.

TELEHEALTH FORENSIC EVALUATIONS

Prior to the onset of the COVID-19 pandemic in 2020, telehealth forensic evaluations of juveniles were fairly rare, although utilizing telehealth for providing treatment to juveniles in juvenile correctional settings located in rural areas where there were few local child and adolescent psychiatrists was not uncommon. Following the onset of the COVID-19 pandemic, much outpatient treatment of children and adolescents moved to virtual platforms, many courts moved to virtual hearings, and many correctional institutions restricted the entry of nonemployees, so telehealth interactions became much more common. Clinicians and courts became familiar and more accepting of telehealth operations, and since many schools moved to online schooling, virtual interactions became more familiar to youth as well. As a result, telehealth forensic evaluations of youth became more common, despite very limited data regarding their comparability with in-person evaluations.

Telehealth evaluations often raise privacy issues that differ from in-person evaluations. If the youth is in an institution, the institution may require a monitor to be in the room with the youth, in part out of a concern that the equipment is utilized properly, which may limit what the juvenile evaluee is willing to discuss. If the youth is at home, there is potentially a concern that an adult may be in the room, out of sight of the camera, and affect the child’s responses in ways unknown by the evaluator, either by coaching or simply out of a need to please the adult.

Video techniques generally do not markedly affect interviewing an adolescent who is cooperative and does not have privacy concerns, although it is more difficult to clearly observe nonverbal reactions. Evaluations of prepubertal youth, which, if conducted in person, would include the clinician using toys or drawings with the child, require alternative approaches, the effects of which need to be considered in assessing the reliability of the interview. Formal psychological tests may also require in-person interactions, although several studies of remote administration of particular measures that involve reading questions to adults have shown similar results to in-person administration. Telehealth forensic evaluations are an evolving area of practice but are being increasingly accepted by evaluators, attorneys, and courts.

FUTURE DIRECTIONS

This is an exciting time to be working in the field of child and adolescent forensic psychiatry because expanding research allows for opinions that are based on research data in addition to specialized knowledge of child and adolescent development, which will aid the courts in their deliberations. This subspecialty is gaining recognition and a body of literature along with clinical guidelines. Child and adolescent psychiatrists also have an important role in conducting research in the areas of prevention, intervention, and understanding the effect of trauma on the developing brain.

The juvenile justice system is in need of further reform. Many programs are inordinately expensive, and their efficacy has not been documented. In fact, in some programs, interventions have led to more reoffending or violent behavior. In many jurisdictions, juvenile offenders receive poor legal representation due to excessively large caseloads assigned to juvenile defenders and inadequate compensation rates. The recognition by the courts of the decreased culpability of adolescents, and the rapidly expanding knowledge of the underlying neurobiology of the developing brain, gives hope that public policy will continue to move in the direction of taking account of developmental information in rehabilitating adolescent offenders. Standards for determining child custody also need to be reviewed in light of changing family structures and the vagueness of “the best interest of the child” standard. Finally, it is hoped that forensic psychiatry fellowships will offer more flexible and diverse programs for those trainees wanting to specialize in child and adolescent forensic psychiatry.

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Schetky for her contribution to previous versions of this chapter.

Further Readings

*American Academy of Psychiatry and the Law. Ethics Guidelines for the Practice of Forensic Psychiatry. Accessed January 6, 2016. http://www.aapl.org/ethics.htm

Appelbaum PS. A theory of ethics for forensic psychiatry. *J Am Acad Psychiatry Law*. 1997;25:233–247.

Ash P. But he knew it was wrong: evaluating adolescent culpability. *J Am Acad Psychiatry Law*. 2012;40:21–32.

*Benedek EP, Ash P, Scott CL. *Principles and Practice of Child and Adolescent Forensic Mental Health*. American Psychiatric Publishing; 2010.

Brodsky SL, Gutheil TG. *The Expert Witness: More Maxims and Guidelines for Testifying in Court*. 2nd ed. American Psychological Association; 2016.

Candilis P, Weinstock R, Martinez R. *Forensic Ethics and the Expert Witness*. Springer; 2007.

Ceci SJ, Kulkofsky S, Klemfuss JZ, Sweeney CD, Bruck M. Unwarranted assumptions about children’s testimonial accuracy. *Ann Rev Clin Psychology*. 2007;3:311–328.

Cohen JA, Bukstein O, Walter H, et al; AACAP Work Group On Quality Issues. Practice parameter for the assessment and treatment of children and adolescents with posttraumatic stress disorder. *J Am Acad Child Adolesc Psychiatry*. 2010;49:414–430.

Dodge KA, Bierman KL, Coie JD, et al; Conduct Problems Prevention Research Group. Impact of early intervention on psychopathology, crime, and well-being at age 25. *Am J Psychiatry*. 2015;172:59–70.

Drogin EY. Forensic mental telehealth assessment (FMTA) in the context of COVID-19. *Int J Law Psychiatry. Epub* 2020;71:101595–101595.

Drozd L, Saini M, Olesen N. *Parenting Plan Evaluations: Applied Research for the Family Court*. 2nd ed. Oxford University Press; 2016.

Elkington KS, Teplin LA, Abram KM, Jakubowski JA, Dulcan MK, Welty LJ. Psychiatric disorders and violence: a study of delinquent youth after detention. *J Am Acad Child Adolesc Psychiatry*. 2015;54:302–312.

Everson MD, Sandoval JM, Berson N, Crowson M, Robinson H. Reliability of professional judgments in forensic child sexual abuse evaluations: unsettled or unsettling science? *J Child Sex Abus*. 2012;21:72–90.

Federal Bureau of Investigation. The School Shooter: A Threat Assessment Perspective. 2002. https://www.fbi.gov/file-repository/stats-services-publications-school-shooter-school-shooter/view

Graham v Florida, 560 U.S. 48 (2010).

Grisso T. *Evaluating Juveniles’ Adjudicative Competency: A Guide for Clinical Practice*. Professional Resource Press; 2005.

Grisso T, Vincent GM, Seagrave D. *Handbook of Mental Health Screening and Assessment for Juvenile Justice*. Guilford Press; 2005.

In reGault, 387 U. S. 1; 1967.

In reWinship, 397 U.S. 358; 1970.

Jones v Mississippi, 141 S.Ct. 1307; 2021.

Kraus LJ, Thomas CR, Bukstein OG, et al. Practice parameter for child and adolescent forensic evaluations. *J Am Acad Child Adolesc Psychiatry*. 2011;50:1299–1312.

**Miller v Alabama*, 567 U.S. 460; 2012.

Montgomery v Louisiana, 577 U.S. 190; 2016.

Office of Juvenile Justice and Delinquency Prevention, Evidence-based Programs. Accessed July 23, 2023. https://ojjdp.ojp.gov/evidence-based-programs

Righthand S, Kerr BB, Drach K. *Child Maltreatment Risk Assessments: An Evaluation Guide*. Routledge; 2011.

Roper v Simmons, 543 U.S. 551; 2005.

Steinberg L, Cauffman E, Monahan K. *Psychosocial Maturity and Desistance from Crime in a Sample of Serious Juvenile Offenders (NCJ 248391)*. OJJDP Juvenile Justice Bulletin, Issue. U. S. Department of Justice, Office of Justice Programs; 2015.

Steinberg L. Adolescent brain science and juvenile justice policymaking. *Psychol Public Policy Law*. 2017;23(4):410–420.

Strasburger L, Gutheil T, Brodsky A. On wearing two hats: role conflict in serving as both psychotherapist and expert witness. *Am J Psychiatry*. 1997;154:448–456.

54.7 Ethical Issues in Child and Adolescent Psychiatry

ADRIAN N. SONDHEIMER, M.D., AND TAKAHIRO SODA, M.D., PH.D.

Ongoing scientific developments and discoveries continue to advance and enhance the practice of medicine in areas relevant to the care of children and adolescents. The medical profession is no stranger to such continued accrual of new knowledge and to the potential benefits and risks associated with these advances, especially those that, at times, suggest quantum leaps ahead. As the entire globe currently confronts the COVID-19 pandemic, the development and distribution of countering vaccines are starkly illustrative of both possible outcomes. In past centuries the pace of change in the "knowledge base" was slower, but societies have always needed medical practitioners, designated "healers," whose responsibility it was to diagnoses ailments, provide treatments, care for the sick, give advice to the afflicted and their families, and develop new understandings of maladies. Since time immemorial, society's healers have encountered and considered conflicting theoretical stances, goals, and approaches. Steady advances in knowledge in the pursuit of providing optimal care have continuously led to comparisons of consequent risks versus benefits, and practitioners often had to confront the reality that proposed new solutions to problems can induce harmful effects.

These age-old conflicts among and between competing views and goals in the pursuit of "doing good" are ethical ones, and physicians—today's "healers"— exemplify their professional roles when giving these conflicts their due consideration. Situations in which such conflicts, or ethical dilemmas, arise include the degree to which physicians impose their wills on a patient when the patient's wishes oppose the physician's; the potential for physicians' personal aggrandizement or hubris to influence medical decision making; and the recommendation of treatment options that would benefit an individual patient but might conflict with the needs or desires of the family, community, and/or society.

The field of ethics is a discipline that focuses on moral principles or values which guide attitudes and behaviors while examining notions of right and wrong in relation to these human activities. Acting in an ethical and professional manner has been a fundamental tenet of medical practice throughout its history and remains an essential aspect of medical practice. The ethics of medicine has always intended to concern itself with the pursuit of patients' best interests and the societal good while respecting the individual patient's personal rights. Codifications of musings about these considerations serve to advise, admonish, or assist people to behave in desirable ways, and these written compilations reflect the philosophical worldview and religious, spiritual, political, cultural, and economic circumstances of their societies.

CODES OF MEDICAL ETHICS—NATIONAL AND INTERNATIONAL

Medical ethics is the application of the ethical discipline to medicine. The individuals designated to treat the somatic and psychological ills of the community's members—today's physicians—are commonly accorded privileges and respect for their specialized knowledge and abilities. To promote this trust and protect the public, the healers and their communities have produced formalized written codes, or guidelines, governing the behaviors and conduct expected of physicians. These codes of medical ethics focus substantially on the treatment obligations of the healers to the individuals residing within these communities and on behaviors expected of them as professionals.

Several themes and exhortations are commonly found in medical codes of ethics that date back many hundreds of years and emerge in myriad cultures. These include the primacy of respect for human life; physician responsibility for the patient; emphasis on virtue and duties; provision of help; avoidance of harm; equality of care for rich and poor; setting of fees; methods used to establish diagnoses, with the expected employment of clinical observation and reasoning; use of appropriate therapies; confidential nature of the patient–physician communications; prohibition of physician–patient sexual relationships; and provisions for punishment or sanctions in the event of poor technical performance, code violations, or both.

In the United States, the formation of the American Medical Association (AMA) in 1847 was accompanied by the creation of a code of ethics. A society organized in 1844 by 13 medical administrator superintendents of “American Institutions for the Insane” became the foundation for the creation of the American Psychiatric Association (APA) in 1921. In 1973 the APA produced its first code of ethics, which was annotated and revised in subsequent years.

Considerations of children’s needs were largely absent in the formative years of these professional organizations. The dawn of the 20th century saw the beginnings of professional reflections concerning children’s health needs as separate from those of adults. The American Academy of Pediatrics was founded in 1930. In 2007 that body adopted, for the first time, a document focusing on the mission, values, and professionalism distinctive to its practitioners. APA members, independent psychiatrists, pediatricians, psychoanalysts, and other medical practitioners with a primary interest in children’s mental health founded the American Academy of Child (subsequently, and Adolescent) Psychiatry (AAC[A]P) in 1953. The AACAP introduced its first code of ethics in 1980; it was expanded, with accompanying annotations, in 1995, revised in 2007, and substantially rewritten in 2009. In 2014 the code was further revised and enhanced by the inclusion of discussions related to social media, expansion of research matters to include scholarly activities, and a focus on explicit natures of potential boundary crossings. The AACAP code covers ethical issues relevant to clinical, administrative, and scholarly practice in child and adolescent psychiatry. Its narrative is organized around principles and major specified areas of concern inclusive of Developmental Perspective, Promoting the Welfare of Children and Adolescents (Beneficence), Minimizing Harmful Efforts (Nonmaleficence), Assent and Consent (Autonomy), Confidentiality (Autonomy/Fidelity), Third Party Influence (Fidelity), Scholarly and Research Activities, Advocacy and Equity (Justice), Professional Rewards, and Legal Considerations.

This code uniquely introduces a developmental perspective, thus covering an essential aspect of caring for children, adolescents, and their families. It affirms that abiding by these ethical principles is considered an essential aspect of practicing as a competent and capable child and adolescent psychiatrist. Further, continued learning and maintenance of the ability to identify, analyze, and make decisions about ethical dilemmas is considered a key skill set for all medical students, residents, and physicians. Ethical child and adolescent psychiatrists (CAPs) should possess the knowledge of ethical principles and their use but, even more fundamentally, understand that this knowledge and skill set is an essential aspect of their professional demeanors and should be reflected in their behaviors. The code advises that a key aspect of the ethical practice of CAPs is a developed awareness of one's personal beliefs and the impact of these beliefs on the clinical and professional choices made by the practitioner.

Ethical thought is universalizable, that is, ethical principles and reasoning apply in all situations, nationally and internationally. However, when ethical reasoning is utilized, attention must always be paid to situational context and culture. Thus, while emphases on, for example, competence, honesty, avoidance of exploitation, confidentiality, boundaries, and protection of research participants transcend individual nation-states, their examination, and potential resolutions are partially contingent on geographic location. Many ethical and related codes have been produced by international bodies, including those of the World Medical Association (originated 1949), the World Psychiatric Association (originated 1996), and the United Nations Convention on the Rights of the Child (1989). This last document established the rights of children to survival, development, protection, and participation in society and, by implication, their health care. In 2006, the International Association for Child and Adolescent Psychiatry and Allied Professions (IACAPAP) produced the document Ethics in Child and Adolescent Mental Health, which created principles addressing the rights of children and behavioral guidelines for the professionals who treat them worldwide. More recently, IACAPAP's Declarations of Durban (2014) and Prague (2018) reinforced and elaborated upon these emphases on the rights of children worldwide to mental health and well-being.

However, while children and families in developed countries can struggle to access services, the resources available to developed versus resource-poor countries, as well as the natures of their governance and cultures, serve as the greatest differentiators in mental health approaches to children living in dissimilar locales. In the former, for example, the practitioner might be concerned with the ethics surrounding recommendation of brand versus generic medications in comparison to a practitioner whose pharmacologic armamentarium often is empty.

Similarly, a locale's culture may determine whether a child's protests are welcomed as a sign of developing autonomy versus being viewed as a violation of older adults' decision making for the group, the culturally sanctioned norm. At worst, displaced refugees, homelessness, migration, sex-trafficking, and the production of orphans (secondary to war, capture, and disease) account for huge numbers of children in predominantly resource-poor countries. A paucity of mental health professionals in these locations further ensures that little attention would be paid to children's psychiatric needs. The ethical imperative to improve children's life circumstances and to provide appropriate care, however, remains unaffected by the surround.

MEDICINE AND MENTAL DISTURBANCE

From a historical perspective, within the context of medically influenced writings, mental illness was first formally addressed by the Greek physician Hippocrates and, several centuries later, by Galen in Rome. The ancient Romans dealt with the definition of mental disturbance, the nature of caregivers, guardianship, approaches to care, and legal competence. Thus, Roman society and numerous others that followed, by recommending approaches that they considered to be correct and beneficial medical conduct, were instrumental in contributing to the creation of modern codes of behaviors expected of healers. Focusing on medical illness, deliberately inclusive of those who suffered from mental disturbance, the codes and laws forged by the ancient Romans, reconsidered and expanded in subsequent eras, explicitly addressed ethical conflicts that daily confront medical practitioners, without separating the somatic and psychological spheres.

CHILDREN AND ETHICAL CONSIDERATIONS

When and how did children come into the picture? Some historians have arguably posited that prior to the 17th century, in the western world, children generally were not viewed as developmentally distinct from adults. Pediatrics was first established as a unique specialty in the late 1800s. Over the last several centuries, children and adolescents have been recognized as distinct groups engaged in ongoing growth, as well as knowledge and skill acquisition, in emotional, physical, intellectual, language, moral, and interpersonal spheres. In the last decades, increasing recognition and support have accrued to the perspective that children have inalienable rights to development and participation in society, as described in the UN Convention on the Rights of the Child. The substages of child and adolescent development imply that children's maturation in developmental spheres correlates with chronologic age and, therefore, that expectations of children's capacities for comprehension, reasoning, and the sound making of decisions will vary with the individual child's degree of attainment. From an ethical perspective, insofar as these capacities remain immature during development, children generally are not granted absolute decision-making rights; rather, these responsibilities remain with their guardians. When the goals of the guardians and their children conflict, however, problems may ensue. Furthermore, children are commonly cared for by individuals or agencies concerned with their welfare, such as schools, individual health practitioners and institutional providers, governmental welfare and protective services, and juvenile justice systems. Information exchange among these parties is necessary for the provision of care, but the transmission of data among them can occur without a child's knowledge, leaving affected children vulnerable to the potential misuse of confidential information. Thus, the factors of continuous developmental maturation, the potential for children–guardian conflicts, and numerous avenues of and needs for information transmission can frequently lead to ethical dilemmas concerning children.

Ethical Dilemmas and Reasoning

Practitioners might recognize the presence of a child-related ethical dilemma when experiencing internal conflict about potential divergent approaches to the provision of clinical care. For example, CAPs consulting to a pediatric clinic, with 15 minutes available to see a 6-year-old distressed patient, must consider whether to spend this limited time with the patient and guardian or with the guardian alone. If psychiatrists choose the latter, they can anticipate obtaining a coherent history without needing to "manage" the individual patient, an approach that could rapidly and succinctly provide the needed medical input. On the other hand, direct observation and interaction with the child might prove crucial to an adequate database. How to provide the best care—an ethical imperative?

Ethically problematic situations like this one are commonly ambiguous —they involve gray areas. In the face of such a conflict, given the pressure of time and clinical demands, harried clinicians might not register the ethical dilemmas related to the principles of *beneficence* (promotion of well-being) and *nonmaleficence* (avoidance of causing harm). Therefore, it is most advisable for clinicians to resist an immediate response, consider the alternatives, engage in a reasoning process—largely a risk–benefit analysis—and ultimately make their choice based on a resolution that is least damaging and ideally most helpful to the patient.

Ethical Principles and Forensics

Medical ethics is a field of inquiry based on principles that are considered universal, inclusive of the obligations to pursue *beneficence, nonmaleficence, autonomy* (promotion of individual freedom of action and choice), *fidelity* (faithfulness, loyalty, and respect for the rights of others), and *justice* (treatment of individuals with fairness and equality). Parenthetically, within this last category, consideration is also warranted for the concept of *social justice*, that is, the treatment of groups of individuals with fairness and equality. These principles often serve as the foundations for derivations of legal precepts and administrative decisions. For example, the legal concepts of the individual’s right to privacy and informed consent for medical procedures largely derive from the principles of autonomy and fidelity. Additional concerns include the possibly conflicting rights of individuals and society, as in the case of proposed quarantine or the mandated use of masks for contagious infectious disease, which pits colliding concepts of individual autonomy against those of communal social justice. The pre-eminent ethical principle of “do no harm” tends to trump others, but even that moral standard can be open to debate when applied to specific situations. Moral dilemmas in medical practice often involve the operation of several ethical principles simultaneously, and the principles may conflict and compete.

The consideration of multiple principles in a reasoning process, however, allows for leeway in decision making. The law, by contrast, although generally based on ethical deliberations, defines acceptable choices and behaviors more narrowly. Not uncommonly, the discussion of ethical considerations segues into an examination of legal ones, conflating the two distinct disciplines. At times, the legal response can be at odds with powerful ethical arguments. For example, in custody disputes that include significantly involved nonbiologically related foster parents, certain jurisdictions adhere to laws that mandate awarding custody to the closest biologic relative, even to one not involved in the child’s current care. Such compulsory rulings might conflict with evidence of the establishment of close and long-standing emotional ties between children and their nonrelative foster parents. Similarly, enforcement of separation of children from parents or handcuffing of nonviolent youngsters for purposes of transport may be lawful but entirely unethical. In such instances, implementation of the law would likely cause harmful, in contrast to beneficial, effects.

Although several distinct ethical concerns may exist simultaneously, over time one approach to the dilemmas usually takes priority as the most beneficial, or least harmful. The child and adolescent psychiatrist clinician, administrator, researcher, or teacher, dealing with a problematic case, must strive to arrive at the least damaging resolution, commonly using a risk–benefit analysis. Awareness of relevant laws is necessary for this reasoning process, but the law is not necessarily the deciding factor.

CHILD-RELATED CONCERNS

Areas of ethical concern common to child and adolescent psychiatry practice include confidentiality, parent–child conflicts in desires and rights, assent to or refusal of treatment or research, physician responsibility for continuation or termination of care, child advocacy and protection, agency, boundaries, and professional responsibilities. Dilemmas arise in many contexts, including inpatient and outpatient facilities, private offices, and institutional settings; in the course of obtaining histories and providing treatments; abuse and neglect; divorce and custody; and forensic settings (e.g., court). Participation in administrative and research/scholarly activities often raises ethical issues. Conflicts arise owing to increased physician awareness of, and sensitivity to, matters of culture, race, religion, and diversity, changes in health care delivery and economics, heightened understandings of the scientific bases underlying disease etiology and expression, the influence of various institutions involved in health care, and the ever-increasing presence of electronic methods of communication and implicit reasoning, used both professionally and socially. Consequently, new dilemmas, or old dilemmas in new garb, confront child and adolescent psychiatrist practitioners, administrators, researchers, and educators on a daily basis.

Multiple Agency

Angela, a pregnant 15-year-old, was informed that her 16-week-old fetus was anencephalic. Obstetrics personnel urged termination of the pregnancy, but Angela refused to abort. One day later, she became acutely suicidal. In the emergency department, a psychiatric evaluation found her imminently dangerous to self. Angela refused to agree to psychiatric hospitalization and an involuntary commitment process was begun. Subsequently, the receiving psychiatric hospital refused to admit Angela without a prior termination of pregnancy. Angela's history indicated that the state child protective services agency, which had functioned as Angela's legal guardian for 5 years, had 1 year earlier removed Angela's then 7-month-old child from her care after a finding of neglect.

CAPs are obliged to recommend and pursue optimal care for their patients. In this case, doing so would require a guarantee of the patient's safety, and, therefore, in Angela's case, hospitalization is indicated. The practitioner, however, must also take into account other interested parties, such as, in this case, obstetric and pediatric personnel who strongly argue for termination of the pregnancy to spare anticipated future suffering of the patient and fetus. Other positions requiring consideration include those of the receiving psychiatric hospital that refuses to admit Angela without prior termination of the pregnancy; Angela's foster parent, who refuses to countenance abortion as an option for a child living in her house; the child protective services agency, Angela's guardian, which refuses to take a decision-making role, arguing that the patient is competent to make decisions; and the patient, who refuses to assent to both psychiatric hospitalization and termination of the pregnancy. While Angela's right to make the decision about her hospitalization varies depending on the state in which she lives, she retains the right to decide about her pregnancy throughout the entire United States. The child and adolescent psychiatrist has the obligation to advocate for Angela in order to ensure that her best interests are protected and she does not experience harm. Thus, working with Angela to better understand her views and helping her to make good decisions while utilizing appropriate strategies to manage her distress, is essential.

CAPs commonly face issues of multiple agencies—the need to be responsive to a case's interested parties, who often are pursuing conflicting agendas. As above, it is fundamental that the child and adolescent psychiatrist protect and advocate for the individual patient. That is, CAPs must act as agents for the minor patient while not permitting themselves to be swayed by the goals of the other parties, their own personal biases, the sources of income that pay for the physicians' efforts, or their allegiances to fellow professionals. CAPs have the obligation to act as the agents for their minor patients, despite their engagement in the care of these patients by parents or guardians with whom the child and adolescent psychiatrist may spend more time than with the child and who often are paying the physician's fees.

> "I want you to treat my daughter. She's been misbehaving at home and at school. This has nothing to do with my recent separation from her mom and the fact that my new girlfriend is living in our home. I'm not going to say another word about this—concentrate your efforts on my daughter"!
>
> "I am the only normal one in my family. You've heard what happens at home—my parents are crazy. Who behaves like this? I'll tell you, people who have problems. They should be in treatment, not me."

Who is the patient? When dealing with children and adolescents, it is often useful to use a family systems perspective that focuses on the dynamics between and among the various parties in the nuclear household. This stance can be viewed as a beneficial approach for all the involved parties because it permits CAPs to function as agents for both the family as a unit and its individual members. When the guardians refuse such an approach, however, practitioners face at least two ethical dilemmas. CAPs need to provide optimal care to the child while not alienating the parents and must avoid inappropriately labeling the child with psychopathology when the primary disturbance may reside with the parents. Such an approach requires clinical dexterity and due regard for the autonomy of all the interested parties while maintaining as primary the needs of the child.

EFFECTS OF CULTURE AND RELIGION

As the intermingling of differing, multiple cultures continues to be the ever-growing norm in the United States, CAPs are frequently called on to deal with conflicting notions of expectable behaviors and deviations from them.

> “I brought my 13-year-old daughter to the emergency department because she’s hitting me. She’s sinful, and she’s crazy. We’ve always gone to church together, 5 days a week, and now she refuses to go. She has to go because otherwise, the Devil will take hold of her, as He did to me in my teens when I had her. Her hitting me, her mother, is just the beginning of His hold on her, and through her, the Devil will infect my younger children.”
>
> “I don’t know why my parents are bothering you with this. They want me to be more observant of our rituals, participate more in our culture, and continue with my education. Both of my parents didn’t finish high school and most people my age back home are working at a job. I’m trying to do what they want, but now I found a job and they have no reason to be mad that I’m quitting school.”

Varied culturally and religiously defined misbehaviors are found within an array of ethnicities and denominations. CAPs must rely on the profession’s scientific knowledge base to evaluate case data in order to arrive at psychiatric diagnoses rather than using the belief systems and mores preferred by a culture or religion. However, CAPs should also be aware of, acknowledge, and have respect for the importance of cultural and religious perspectives to patients and their families. Treatment plans and dispositions should be approached in a similar manner, using resources peculiar to the particular belief system to the extent possible, without jeopardizing the care of the child. In addition, CAPs should note that separation-individuation phenomena manifest and provoke consternation in many, perhaps all, cultures. In the interest of providing good care, CAPs are ethically bound to pay attention to both culture-specific and developmentally normative behaviors. Not infrequently, education concerning normal child and adolescent development, conveyed in a caring manner, is much appreciated by the parent living in an otherwise self-contained culture.

Countertransference

The five abovementioned vignettes likely engendered, in the evaluating and treating psychiatrists, affective responses toward the protagonists. *Countertransference*, the term colloquially used to describe these phenomena, is closely related to CAPs’ affective responses to the raised issues. What are the psychiatrists’ attitudes toward abortion, teenagers who engage in sexual activity and become parents, parents reluctant to acknowledge the effect of their behaviors on their children, assaultive youths, individuals for whom religious and cultural beliefs and mores are paramount, and youngsters striving for autonomy? It is important for CAPs to engage in self-inquiry and determine the bases for their countertransferential responses to these factors so as to ensure that these phenomena do not negatively affect optimal care for the children (and adults) under their care. Overidentification by CAPs with the child or adult is dangerous, as are rescue fantasies that focus on children who seem to live with little in the way of adequate adult protection, advocates, or both. CAPs in earlier stages of their training may tend to ally themselves with their patients, whereas older ones, many of whom have become or are parents, may tend to express greater empathy with the guardians. Ultimately, CAPs must focus on the needs of the child and avoid choices that are more reflective of the individual psychiatrist's intrapsychic dynamics.

PROVISION OF CARE

From a clinical perspective, ethical considerations should begin with the first patient contact and extend past the final meetings.

Assent and Consent

> "Hello, doctor. I would like to make an appointment for my 9-year-old boy who probably has ADHD (attention-deficit/hyperactivity disorder), and I know he has lots of behavior problems. By the way, I'm not telling him that we're coming to see the doctor. He gets so upset, and I figured you could explain it better once we get there."

Questions that should follow immediately on such a seemingly innocent and sensible inquiry about an appointment ought to address whether the involved child agrees to a medical procedure (i.e., the psychiatric evaluation and interview). Will the child assent to this procedure, is he averse to it, and/or is he being coerced by the parent to attend? To what extent should children's objections to the initiation of an evaluation, or to treatment that might be recommended at its conclusion, be respected—that is, to what extent is children's *autonomy* considered? Have the children been reasonably well apprised by the guardians of the reasons for the proposed evaluation, and, if not, by the practitioners in the course of this first meeting—in other words, has the assent been informed?

Insofar as children ordinarily are legally considered minors until the age of 18 years, children are generally unable to give legal *consent* to proposed medical procedures, but they may *assent* to them. It is the children's guardians who provide consent. Further complications can arise when disagreements exist between a children's legal guardians regarding proposed care, as may occur in the context of divorced parents. Often one of the parties is legally designated the guardian ordinarily responsible for obtaining medical care for their children. In any case, in emergencies, the urgent need for the provision of care becomes the physician's paramount consideration, irrespective of parental guardianship. In situations short of emergencies, however, lack of consent or opposition by one of the guardians raises ethical and legal dilemmas. The parental disagreement might be exacerbated by psychiatrists' actions favoring the stance of one of the parents, which could negatively affect or harm the child by aggravating the parental conflict.

Standing in contrast, psychiatrists must also consider the children's rights to an evaluation and their need for care.

> "I'm pregnant, my parents don't know, I'm not sure who the father is, and I'm not sure what to do about it. I'm 15 years old, I can think for myself and make my own decisions, I have the money to pay your fee, and I want to discuss this with you without my parents knowing about it."
>
> "I don't see why you want to talk to my parents. Yes, I live with them, but they really don't know anything about me, and I'll be leaving soon when I graduate from high school."

These adolescents are not simply assenting to treatment but are requesting care from psychiatrists. A 15-year-old pregnant girl, in some states' jurisdictions, is automatically considered an adult or "emancipated minor" by dint of her pregnant state, whereas in others, she remains legally a minor. "Mature" or "emancipated" minors can legally access care for themselves. Ordinarily, however, care only can be provided, except in emergency situations, with the consent of an adult. At what age does the adolescent develop the cognitive and emotional maturity to be able to give consent? Is that decision based on chronologic age (e.g., attainment of age 18 years), or is it situation or case-dependent? Given that adolescence is a stage of rapid physical and cognitive maturation and change, this fluidity in growth renders ethical questions regarding autonomous decision making more complex and less certain than when dealing with similar issues in younger children. As adolescents advance toward cognitive maturity, respect for their desires and opinions should progressively be factored into professional decisions in the interest of respecting the youths' entitlement to *autonomy*.

Confidentiality

It is the physician's responsibility to protect, and not convey to others, information about a patient unless the patient has given permission for this transmission. The patient has a right to expect adherence by the physician to this obligation, which serves as a basis for a trusting physician–patient relationship. The principle of confidentiality is often subject to limitations in working with children in that guardians, patients, and agencies want and require information about child patients, and children often are unable to comprehend the risks and benefits involved in the waiver of their rights to confidentiality. In addition, electronic medical records (EMR) and other electronic methods of communication can mean that, without adequate built-in safeguards, multiple individuals and institutions may potentially access confidential patient information.

"It's simple, doc. You've known my kid for a while. She's seriously disabled, she's 10 years old, and we need the state benefits to which she's entitled. Just fill out this 10-page form—it asks for all kinds of information about her—nooooo problem"!

"Doc, can you give me a copy of her chart? The school needs the records to start their evaluation process, and it'll be faster if I just deliver them the information."

A signed release from guardians permitting the transmission of information about children might satisfy legal concerns, but ethical issues remain that warrant exploration. To what extent do children's disabilities affect their capacity to provide assent? Would state/school involvement foster a child's development or hinder it? Is all, some, or none of the requested information pertinent to the child's condition? Is the information requested legally protected, or might it be accessible or transmitted to parties in the future without the patient's or guardian's knowledge? Psychiatrists must evaluate these issues by simultaneously playing the role of child advocates and attempting to protect children from current and possible future misuse of the information that other practitioners or agencies might subsequently provide. Such questions should be addressed with both guardians and, to the extent possible, patients, because families need to be aware of the ways in which information supplied to third parties might subsequently be used. They also should be educated about the potential risks of information disclosure when various forms of electronic communication, such as e-mail, social media, and cell phone texting, are employed.

"About those messages I've sent you about my daughter. You know, asking whether she talks to you and tells you the truth or she lies? Please don't tell her about them or about any of the info I've given you about her."

Adolescent patient to psychiatrist:

> "I don't really talk on the phone. Can you text me? I'll definitely get any messages then. And that way, I can let you know what's happening; it would be great to talk between sessions."

Commonly, parents or guardians are eager to know details concerning information the child and adolescent psychiatrist has obtained in the course of treatment of their minor. The use of a family therapy approach at the onset of treatment, with an explicit message emphasizing the goal of judicious exchanges of information while simultaneously acknowledging privacy needs and rights, can circumvent many confidentiality concerns. Informing family members at that time that information requested or supplied outside of treatment sessions would likely be shared with all obviates otherwise sequestered communications. As well, it is warranted to notify all participants at the outset that potentially self-harmful or externally directed aggressive ideations and acts do not fall under the confidentiality obligation. In forensic contexts, minors and their guardians need to be informed from the outset of the lack of confidentiality concerning personal information (i.e., that the information included in forensic reports will be made available to all interested parties and may become part of the public record). Ultimately, practitioners should always exercise tact and restraint in the provision of information to all concerned parties, all the more so when transmitted information is likely to be reviewed by nonmental health professionals.

Separately, the advent of commonly used digital media and electronic communications, while positively facilitating information exchange of all kinds, does not in any way eliminate the need for alertness to the existence of ethical considerations and legal concerns in these spheres. The considerations that govern face-to-face care apply equally to electronic communications with patients and their families, despite their apparent ease of use. Thus, potentials for boundary violations, confidentiality and privacy transgressions, negligence, and defamatory messages are abundant but sometimes not obvious. Further, the seductive simplicity of seeking and finding information about patients and their families, easily obtainable via the Internet, could result in ethical violations when performed without the patients' and families' knowledge and, similarly, if not necessary for patient care. All of these concerns and behaviors involve potential violations of patients' rights to expect respect by physicians for their *autonomy* and avoidance of breaches of *fidelity*.

Fidelity and Mutual Trust

Fidelity, *honesty*, and *integrity* are three cardinal qualities expected of physicians that contribute to the measure of a physician's professionalism. Patients are entitled to expect physicians to deal with them openly, stay with them and not abandon them through the course of their illnesses or disturbances, and make attention to their needs primary before the physician might focus on possible self-interests. These expectations, however, may come under pressure when a physician feels threatened or hemmed in.

> "You've seen my 16-year-old son for a month. He's angry and unhappy, fighting with me and failing school. I think he's using drugs. I want you to tell his pediatrician to order a tox[icology] screen and have it performed without his knowledge when he sees the doctor next week. My son would refuse if asked directly. If it turns out negative, he won't know I suspected him of drug use; if positive, I'll keep the results to myself and decide later what I want to do about it."

The patient's guardian requests of the psychiatrist to deal with the patient in a nonstraightforward manner to obtain important information. This request is ethically problematic because the creation of mutual trust between physician and patient is a fundamental tenet for the establishment of a sound treatment relationship. A physician–patient relationship in which neither side could expect honest exchange would be doomed from the outset, even if the knowledge sought in a clandestine manner were significant. The means to ends should be discussed in a candid exchange, even if some of the desired ends are, in the short run, thus rendered more difficult to achieve. Short of emergencies, there is no place for a paternalistic approach that would justify withholding vital information from a patient. It is crucial to the establishment and maintenance of successful therapeutic relationships that all family members understand this principle.

CAPs and others involved in the care of children are periodically faced with behaviors that adults experience as repugnant, threatening, or both.

An inpatient pediatric attending addresses the consulting child and adolescent psychiatrist:

> "The nursing staff on my unit have been telling my patient, a 15-year-old boy with hepatic failure, that they wish he would die. He developed severe hepatitis as a child and became a "darling" of the unit during his stays there while receiving and then rejecting two liver transplants. He is back here now being evaluated for a third transplant, but 2

> weeks ago, he was charged with sexual abuse of a child, and the staff has completely turned on him."

CAPs tend to have considerably more experience with victims of abuse than with (alleged) perpetrators. Physicians' efforts, therefore, are mostly devoted to the repair of those individuals who have experienced an emotional injury. Children who are accused of doing (or who do) bad things, however, also require care. Nondiscrimination is a fundamental ethical imperative. The provision of the physicians' care is dependent on a patient's needs and not on the whims, preferences, or biases of the practitioners. Thus, it is incumbent on CAPs to act as advocates for the abusing, malfeasant, and nasty child as much as for the victim deserving of sympathy. In other words, despite possibly harboring negative feelings toward patients, physicians must maintain *fidelity* to them and continue loyally to provide disagreeable children with professional service.

Despite such exhortations, situations arise in which CAPs' negative feelings toward children may interfere with the provision of good care, in particular when CAPs feel threatened by the children.

A 15-year-old male teenager in a detention facility, adjudicated for murder, calmly says to the examining child and adolescent psychiatrist:

> "I know where you live. I don't like you. You stay in my face, and I'll kill you. That's not a threat—that's a promise."

Psychiatrists, like all human beings, are entitled to self-preservation and considerations of self-interest. To the extent possible, CAPs should avoid allowing these concerns to jeopardize the health of their patients. Termination of care may be appropriate in the event child, and adolescent psychiatrists believe themselves to be in danger, but only if provisions are made beforehand for care to be provided by other practitioners.

Boundaries

Physician behavior toward patients should always be initiated and maintained with the patient's best interests in mind. Medicine's professional ethos emphasizes strict physician adherence to the value of integrity. In the context of the physician–patient relationship, the concept of boundaries refers to limits within the relationship that ought not to be breached. Maintenance of such boundaries helps preserve the integrity of the relationship. Although some physician behaviors clearly violate the physician–patient boundary and are impermissible (e.g., sexual relations between physician and patient; physician use of confidential patient-derived information for the physician's gain), other behaviors reside in gray areas. Physicians who divulge personal information to patients, or overlook timely payment for services from patients reporting financial difficulties, may be engaging in good, compassionate practice or enticing and exploitative behaviors that service the physicians' needs, or both.

"I saw you eating dinner at a good restaurant a few weeks ago, so I figured you liked it. You've helped our family so much that I wanted to say thank you with a gift card to that place. Don't worry; it isn't a big deal."

"Can you come to my graduation? I couldn't have made it without you."

"Can you lend me a couple of dollars? I'm really hungry and don't have enough money for the potato chips."

"This case is a great example of the subject I'll be presenting, but I need to send in my talk tomorrow. There's not enough time to find the patient and ask her permission to use her information. I'll change the details a bit, so no one will figure out who she is."

Working with children and adolescents can lend itself to the potential for boundary infringements. CAPs who passively enjoy the spectacle of a child's oppositional or adolescent's seductive behaviors and therefore neglect to set appropriate and timely limits, agree to treat children despite current socialization with the family, accept significant presents, prolong children's care owing to an attraction to a parent, or discuss with colleagues or write for publication about a patient for the purpose of self-aggrandizement are, in all these instances, potentially infringing on or violating patient–physician boundaries.

Patients justifiably expect not to be exploited by their physicians and to feel secure in the treatment setting. Despite best intentions to engage in self-monitoring, physicians' positive countertransferences toward patients could result in behaviors that end in a special or favored status for particular patients. In those instances, it is highly recommended that physicians recognize when they indulge in these behaviors and their potential for leading toward more frank violations. Rapid consultation with respected colleagues is the preferred antidote to strengthen and secure physicians' self-control and provide patients with sound and unimpaired care.

CURRENT AND FUTURE ISSUES

Although the ethical principles of beneficence, nonmaleficence, autonomy, fidelity, and justice remain immutable, their consideration must be applied to constantly changing contexts that include advances in scientific knowledge, innovations in means and modalities of communication, changes in financial structures, shifts in societal priorities, newly created treatment approaches, and novel research foci with patients. It is fascinating to note how rapidly the foci of CAP investigation and discussion change. Thus, in the past decade, prominent areas of concentration have included, among others, treatments of psychiatric prodromes; health insurance and the funding of psychiatric healthcare; the significant influence of the pharmacologic industry; and the potential for improper neuroenhancement. While their visibility and attendant concerns continue, newer issues are achieving ever-greater prominence. For example, racial and economic disparities, equity, and diversity; impacts of social media; alternative sexual lifestyles; and medical marijuana prescription/legalization have come to the fore, while the application of advances in psychiatric genetic knowledge and, separately, emerging technologies and artificial intelligence, are in the process of significantly transforming aspects of CAP practice. Thus, change is a constant, as it presents novel challenges. These challenges, in turn, call for the application of unchanging ethical analysis to the freshly minted ethical dilemmas they pose.

The following subsections focus on 12 relevant areas.

Information Technologies and Data Science

Substantial advances in electronic data collection, storage, retrieval, communication, and reporting have revolutionized medical work. These technologic advances raise novel ethical dilemmas for child and adolescent psychiatry practice. The EMR increases the possibility of the transmission of confidential information concerning minors, without a guardian’s consent, over the time period that the information is electronically stored, that is, theoretically forever. The child, owing to age, lacks control over personal information transmission. That same individual later becomes an adult. At that time, potential employers’, health insurers’, and public agencies’ electronically facilitated access to the past information, despite possible legal protections, could negatively affect the former child, owing both to the stigma that might continue to attach to mental disturbance and the effect of pre-existent illness on employer hiring practices and employee acquisition of health benefits. Will the adult have control over personal health data entered into information banks when that individual was a minor? Who owns this information, who is empowered to make these decisions, and what standards are to be used?

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 was established to create transaction standards and privacy rules overseeing the electronic submission of personal medical information. Although the Act, which became effective in 2003, provided federally mandated privacy protections for this information, loopholes exist, not to mention illegal cybersecurity breaches. In the interest of advocacy for patients, it has become the CAP's professional responsibility to alert children's guardians, as well as the adolescent who may have arrived at a cognitive maturity enabling comprehension of the implications of these privacy concerns.

Simultaneously, electronic systems are increasingly used to provide treatment, as in telemedicine, and many websites complement traditional therapies. In-place security protections of these systems vary. Thus, patients' medical information could be accessed by strangers. As well, individuals are often naïve to the potentially negative implications of entering or posting information on social media, email, and texts, and to an even more recent development, that of allowing access to information by third-party computer or phone applications. Children and adolescents, in particular, due to their developmental immaturity, often lack the capacity to truly understand the immediate and long-term implications of sharing personal data via these mechanisms. Use of these systems by adult professionals equally creates potentially significant complications for the providers as private citizens as well as in their roles as health care personnel. The challenges of keeping patient information secure and maintaining therapeutic relationship boundaries while simultaneously providing effective interventions may frequently be complicated. CAPs must carefully configure their professional and, separately, personal social media accounts to ensure that patients are not "friended" or are not acknowledged as patients, as inadvertent disclosures could easily lead, in varieties of ways, to HIPAA violations. Thus, these digital communication methods can be effective and useful, professionally and personally, but their use requires great caution, vigilance, and self-monitoring.

Combined datasets derived from myriad information sources previously deemed irrelevant, too hard to integrate, or too unreliable, are now used to make associations with numerous psychiatric conditions, which in turn enhance the possibilities for risk prediction. These combined sources include prior medical and pharmaceutical records, neuroimaging data, and genomic information. The datasets are used for primary prevention and continuity of care, secondary prevention for patients with current and ongoing psychiatric diagnoses, and tertiary prevention via highlighting individuals at high risk for expressions of severe disturbance, particularly suicide. The electronic health record system is no longer a passive network waiting for stored data to be used following patient consent. Rather, it has become a dynamic tool that readily alerts physicians to elements of prior medical histories that patients may not currently disclose, and it shapes physician practice by alerting to needs for diagnostic workups or follow-up care, and, in some cases, it initiates “hard stops” on physician entry of further data into the record until required actions of benefit to the patient are performed.

A child presents to a CAP’s office after a referral from a local school. The school principal describes him as “fidgety” and “never listening to his teachers.” The mother notes that “he is always running around, can’t sit still at home, and unless his favorite TV show is on, he never pays attention.” Earlier in the week, he took a book away from a classmate. He recalls that he “just wanted to look at it.” The CAP considers prescribing stimulant medication after diagnosing the boy with ADHD. However, when typing the information into the EMR, a popup appears, preventing further data entry until the CAP gives evidence of diagnostically considering the presence of oppositional defiant and conduct disorders.

These EHR-based technologies are intended to improve patient care by, for example, alerting physicians to order lab tests that could otherwise be overlooked, such as lipid panels when neuroleptic medications are prescribed. However, they can also be programmed to recommend less universally accepted directions, such as requiring EKGs in advance of stimulant prescriptions or considerations of unlikely diagnoses. The “hard stop” can prove problematic because it negatively impacts physician autonomy–a possible contribution to the phenomenon of physician “burnout.” Relatedly, suspicions have been raised that such required orders might be inserted to optimize reimbursement of hospitals or physicians rather than instituted in the best interests of the patient.

Advances in artificial information (AI) technologies have led to the development of machine learning-based prediction tools. They purportedly aid in the identification of individuals at increased risk of suicide, which would likely influence subsequent physician behavior. Similarly, physicians, fed AI-based alerts, are likely to incorporate these decision support tools into their practices, often by intervening sooner and more aggressively in the management of a patient under observation, or treatment, for psychosis. The nature of prior data input presents ethical obstacles to AI's proper use. Thus, for example, data distorted by systemic racist perceptions regarding the indications for and incidence of such diagnoses as ADHD and conduct disorders in children, and the lack of transparency regarding the sources of the data input, present caveats to wholehearted acceptance of the algorithmic outcomes and/or the degrees the professional's behaviors might be influenced by electronic direction. Thus, datasets inputted with good intentions, that is, to increase surveillance of individuals at risk for psychiatric disorders, simultaneously infringe on patient autonomy and rights to privacy. In addition, the flag, or alert, AI may place on a patient may bias future clinical diagnoses. Unfortunately, few clinicians are likely to possess knowledge of possible inadequacies in data input. Instead, they likely would pay close heed to AI's indications and recommendations, akin to conforming to a supervisor's suggestions during training. Therefore, CAPs should be given instruction on the production methods of AI data and its potential limitations, both ethical and informational.

The following is an anticipated likely scenario akin to current analogous presentations: A recently developed application, marketed to parents, advertises its ability to derive “a likeliness to commit suicide score.” It advises installation of the app on their child's cellphone. By monitoring the child's activity, in combination with information obtained from the youth's medical records, the app would alert the parents when it anticipates “the child might harm themselves within 2 months.”

If an algorithm predicts a child to have a substantial risk for suicide within the near-term future, should that individual be brought to a CAP for an evaluation? Might legislation implement and demand such a procedure? Might the psychiatrist's decisions be impacted by knowledge of the algorithm's recommendation? Might the CAP endorse hospitalization despite its inherent limitations on patient autonomy? Who would the psychiatrist trust-the algorithm or his/her personal judgment?

Considerations of potential legal liability arise in the event of CAP disagreement with an electronically derived conclusion or recommendation. AI is the future, and it is here, presenting novel ethical challenges to child and adolescent practitioners.

Genomics; Prodromal Treatments Revisited

More than 20 years ago, the Human Genome Project successfully decoded the DNA sequence of a human individual's 23 paired cellular chromosomes. Since then, numerous discoveries have occurred regarding the role genetics plays in human disorders. These findings have been paralleled by a significant decrease in the time and costs necessary to investigate an individual's DNA. Whereas decoding the first genome required 13 years of work and over 1 billion dollars, an entire human genome can now be sequenced for 1,000 dollars in a single day. For less than $100, single-nucleotide-polymorphism (SNP)-based microarrays can detect various structural abnormalities, including copy number variants (CNVs), and also be used to assess common variations associated with numerous human disorders, with the results analyzed and returned to clinicians, or directly to consumers, in a matter of days. Such DNA-based tests investigating specific genetic disorders have profound implications for illness prognostication, treatment, and counseling, as well as their ontology. This especially holds true with regard to psychiatric disorders, as many of their genetic architectures demonstrate substantial overlap and rare structural variants appear to play significant roles in the manifestation of multiple neuropsychiatric conditions. Current Diagnostic and Statistical Manual (DSM) and International Classification of Diseases (ICD) diagnoses primarily rely on gross observational criteria, not on etiology. Genetic- or biology-based recategorization would likely considerably alter a portion of the current diagnostic categories and boundaries.

As well, studies are revealing differences in treatment responses when patient groupings are based on genetic findings, for example, treatment resistance in schizophrenia with duplications of 15q11.2-q13.1, in contrast to DSM-5-TR-based diagnoses. Recent cross-DSM-disorder genomic analyses have found significant overlaps in genetic risks for DSM-distinct disorders (e.g., schizophrenia and bipolar disorder; obsessive-compulsive disorder [OCD] and tic disorders).

The recommendations by such professional organizations as the AACAP, American Association of Pediatrics (AAP), and American Society for Human Genetics (ASHG) for routine genetic evaluations of patients with suspected autism spectrum (ASD) and intellectual deficiency disorders (IDD) is a recent consequence. Similarly, numerous studies have examined pharmacogenetic assessments to prognosticate optimal pharmacologic treatment of children. For example, FDA labels now recommend dose adjustments for individuals with specific pharmacogenetic propensities that include cytochrome p450 (CYP) and human leukocyte antigen (HLA) polymorphisms. However, the medical field has yet to achieve consensus on the utility of these tests; some professional organizations currently recommend their limited use (e.g., International Society for Psychiatric Genetics; Clinical Pharmacogenetic Implementation Consortium), while others oppose them (e.g., AACAP).

Given these conflicting stances, CAPs are struggling with the integration of genetic information into clinical practice. They must ensure that patients receive appropriate genetic testing when warranted and recommended (*beneficence*), provide counseling for or against testing when immediate benefits do not exist, and interpret tests that they may neither have ordered nor would have recommended prior to their deployment by others (e.g., pharmacogenetic analyses, polygenic risk scores).

From another perspective, these genetic discoveries are again raising possibilities of providing preventive treatments as beneficial interventions. Thus, preventive treatments based on genetic findings might be applied with currently available psychotropic medications or with yet-to-be-developed gene therapies. Informed *assent* and *consent,* and the weighing of the numerous risks versus benefits of these postulated interventions, would need to be mandated as ethical requirements.

Studies in recent decades premised that the appearance of specific behavioral signs and symptoms in childhood or adolescence, coupled with a positive family history of severe disorders, might be pathognomonic or prodromal for the later development of full-blown psychiatric disorders such as schizophrenia. Such theoretical prodromic features were not posited as definitively predictive; they did postulate possible vulnerabilities. Further, these studies addressed whether treatment, with neuroleptic medications or with enhanced psychosocial interventions, might lead to changes in outcomes, such as the prevented development of psychiatric disorders and/or disease progression.

Aside from the complicated matters of informed consent and assent, dilemmas also included the accuracy of disease identification and the huge intrinsic difficulties in monitoring outcomes. It became very clear that it is nigh to impossible to determine whether preventive pharmacologic and/or psychosocial treatment could forestall the later expression of an illness if that disorder was not destined to be expressed to begin with. These concerns, coupled with worries about the potential for unnecessary and harmful medication effects, resulted in the discontinuation of these programs at the experimenting institutions.

Currently, preimplantation genetic testing is available for neuropsychiatric disorders such as Fragile X, Huntington, and Alzheimer disorders; prenatal genetic screening can detect rare chromosomal abnormalities with large effects on brain function, predicting increased risk for ASD/IDD and schizophrenia; and direct to consumer testing is available for genetic variants associated with such neurodegenerative disorders as Alzheimer and Parkinson disorders as well as polygenic risk scores for schizophrenia. Related studies suggest that eliminating embryos with high-risk scores for schizophrenia would lower the prevalence of the disorder to 20% of its current levels.

Clearly, advances in molecular genetics have enabled researchers to risk-stratify individual youths based on polygenic risk scores, narrowing the number of individuals that might undergo future preventive treatments. However, genetic and epigenetic interventions specifically designed to target large-effect genetic variations are ethically challenged due to their rarity, thus making experimental validation difficult. As well, possible indications for early intervention, perhaps in utero or even preimplantation, are not currently countenanced as, to date, no timepoints have been established at which treatments could be delivered to achieve avoidance of the disorders, while in utero treatments give rise to concerns about potentials for damage to both embryo and mother.

All of these developments, and their associated caveats, again raise the specter of possible renewed interest in preventive treatments. Of inherent and great importance, they raise huge ethical concerns surrounding impacts on reproductive freedom (*autonomy*), financial costs to and considerations by society (*justice*), and potential harms to the developing fetus and child (*nonmaleficence*). Clearly, CAPs must be alert to the latent issues raised by enormous leaps in medical knowledge and their potential applications while working in tandem with medical geneticists and counselors.

A related but separate ethical concern involves CAPs' medical record notes regarding patients' genetic information, which requires strong adherence to safeguarding data (*confidentiality).* Breaches could lead to patient stigmatization and employer or hiring agency discrimination. Similarly, though legal guards exist against genetics-based discrimination for health insurance in the United States, those protections do not hold for life insurance coverage. CAPs also must be cognizant of broad societal discrimination against "atypical" individuals as well as parents of children with genetic syndromes.

Conflict, Violence, and Advocacy—Individual, Group, National, and International

As a country, the United States engages in behaviors that cry out for responses in opposition from CAPs, such as the recent separation of children from their migrant parents at its Southwestern border. Throughout the globe, both at intranation and nation-state levels, the world is chronically beset by adult-conducted conflict, often violent. Consequently, children and adolescents suffer. They are captured, kidnapped, abused, separated from their parents, and witnesses to the killings of family members and others. Subsequently, they become homeless, refugees, migrants, and vulnerable to exploitation. These circumstances persist despite United Nations commission declarations, nobly protecting children's rights, supported by most of its members. As well, in relatively more settled countries, children are often witnesses and subject to domestic abuse.

CAPs are ethically bound to act as advocates for traumatized children. Clearly, the professional, in the one-to-one encounter, is expected to evaluate, make recommendations, and provide treatments that might help alleviate or mitigate the child's trauma. In the face of negative injunctions affecting large groups of children, be they legislative or administrative, CAPs, individually and as professional groups, should feel encouraged to make their voices heard in opposition while simultaneously recommending more useful approaches. Likewise, in the realms of intranation or state-endorsed violence, CAPs can be knowledgeable about and endorse interventions that attempt to mitigate the after-effects on children and, on rare occasions, by choice, participate in that work.

Children lack the autonomy and power that cognitive maturity, access to employment, and voting rights provide. In the United States, their guardians, insurers, or both, inadvertently or deliberately, may make decisions that are not in the children's best interests. Therefore, proper treatment and placement venues, possible needs for evaluation and treatment, the foci and choices of treatments and the need for their continuation, the value of extended physician–patient relationships, and recommendations for adjunctive investigations and care are all matters that may require advocacy by the child and adolescent psychiatrist for a child.

A 10-year-old boy with a history of assaultive behaviors was brought to a hospital emergency department after having attempted to stab a peer with a knife 1 hour before. The patient was still agitated, expressing rage at this peer, and possessed negligible insight and responsibility for his behavior. Nevertheless, the boy's father insisted on the discharge of the patient to his supervision in order that they keep their long-established plans to fly the following day to a distant city. Independent of the family plan, the patient did not want to be hospitalized. The father assured the evaluating child and adolescent psychiatrist that he would immediately obtain psychiatric care for his son in the new location. The patient's insurer expressed ambivalence about hospitalization, especially after learning of the father's wishes, but reluctantly agreed to cover the costs in the event the child and adolescent psychiatrist chose this course of care.

The vignette's guardian's concerns are legitimate, and his intentions may be well-meaning. The insurer's concerns are legitimate as well, although considerably influenced by financial considerations. Owing to safety considerations, however, the child and adolescent psychiatrist must opt for psychiatric hospitalization of the child. The practitioner is obligated to advocate for the child's immediate well-being, even in the face of opposition from the guardian, child, and insurer. *Advocacy* for the individual child implies choosing the least damaging or most beneficial course for the child in dealings with parents, guardians, agencies, insurers, and the patients themselves. Individuals, agencies, or insurers may attempt to dissuade CAPs from their chosen course because the interests of the payment providers may conflict with the immediate needs of the child. As the child's potential advocates of last resort, however, CAPs should provide the interested parties with the reasons underlying their positions as well as maintain the medically indicated stance, despite possible resistance or antagonism.

Race, Ethnicity, Socioeconomic Status

Dr. Smith, one of three CAPs in the city, is the only CAP that accepts public insurance. He enters a clinic's waiting room and calls for Robert, referred by his school, for an evaluation following "aggressive behaviors" directed toward a peer. Immediately following the call, Robert's parents approach the clinic receptionist and loudly request a different doctor, "one who knows what it's like to be of our race."

The recognition of posited systemic bias, inequities in mental health systems, and both explicit and implicit bias held by children, adolescents, and their parents, as well as practicing CAPs, requires consideration and address. Studies indicate that mortalities related to specific medical conditions decrease when racial minorities are treated by physicians of the same race. Studies suggest that adult psychotherapy patients may prefer and have a better impression of race-matched therapists; nothing is known about race-matching in child and adolescent psychiatry. What are the clinician's obligations in the event of such requests? Is the patient's or family's autonomy to be respected? Does the answer change if the requesting child or family is Caucasian rather than a child or family of color, or vice versa? If, upon investigation, the request for a race-matched provider arose due to traumatic experiences a child or family experienced at the hands of a perpetrator of the psychiatrist's race, might this fact indicate a different response to avoid the imposition of harm than a request arising from perceptions of provider inferiority due to the latter's ethnicity, race or religious affiliation? At a minimum, the underlying motivations behind such requests deserve address.

Significant disparities exist in the diagnostic rates of childhood-onset neuropsychiatric disorders; one explanation hints at the posited presence of systemic bias. For example, a child has an increased likelihood of being diagnosed with IDD and ADHD, relative to Caucasian children, when that youngster belongs to a specified minority. Conversely, the rates of diagnosed ASD are significantly higher in Caucasians compared to those of minority populations. The concentration of CAPs and other ASD treatment providers in affluent, primarily Caucasian neighborhoods may be factors contributing to these findings.

Unjustifiable uses of violence in the name of authority against Black men and women, the recent and current bans and barriers on entry into the United States from countries whose citizens are predominantly Muslim and/or black- or brown-skinned, and inaccurate attribution of coronavirus disease 2019 (COVID-19) spread to Asian peoples leading to physical assaults on them, all highlight the significant impacts race, ethnicity, identity, and religious affiliation have on the probability that children and adolescents within these and other minorities experience related trauma, directly or indirectly. The ethical principle of beneficence, directed toward helping the individual child, adolescent, adult, or family, might sometimes be placed in opposition to those of justice and equity, as, for example, in the race-blind distribution of resources held up as society's goals and ideals. Yet the emphasis on the latter can involve a denial of the importance to kinship groups of their intermingled traumatic history, skin color, cultural norms, and religious beliefs. Ultimately, the possible presence of implicit or unconscious bias on a professional indicates the benefit to be derived from acknowledging its influence, if any, in order to counteract its effect, thus avoiding causation of harm and leading to helpful decision choices for the individual and family. As well, when CAPs become aware of such biases operating within larger groups in the professionals' spheres of influence, they are urged to identify them and advocate in those settings for the best interests of affected children and youth.

The COVID-19 pandemic has exacerbated and brought to light systemic inequalities and their effects on children and adolescents. In the United States, roughly one-third of the labor force was deemed "essential." Despite restrictions on travel and stay-at-home orders, the vast majority of these workers, other than those laid off, continued to work. Doing so contributed to disproportionate rates of infection, morbidity, and mortality secondary to COVID-19 infection in African- and Latin-American populations. The impact of the death or severe illness of a parent, or of similarly significant support, on the children in these households is obvious. In addition, African and Latin Americans are disproportionately employed in low-wage occupations and are less likely to possess health insurance, which translates into further financial burdens on these already poorer communities and families. With regard to gender, women have been more likely to lose or quit employment than men, as they have been disproportionately tasked with assuming supervision of their children in response to closed schools and home-bound learning. In addition, the COVID pandemic's heavy toll on the service and hospitality sector, in which women are overrepresented, has further negatively impacted children. The lack of infrastructure development in locales populated predominantly by communities of color, as well as those of rural communities, highlighted by the lack of stable high-speed internet availability, leads to difficulties in remote learning as well as access to telehealth. These disproportionate impacts on educational quality and mental health access, affecting children both directly and indirectly, are equity issues about which CAPs should have awareness as they interfere with the provision of beneficial care and can serve as spurs for involvement in advocacy efforts by professionals.

Telepsychiatry

Telepsychiatry was originally conceived in the 1960s. Recently, however, its use has become much more widespread with the increased availability of videoconferencing technology. Increased access to high-speed internet, ownership of personal computers and cell phones with video capabilities, and software meeting the technical demands for patient confidentiality have all contributed to the rapid spread of telepsychiatry for both patients and mental health practitioners.

Telepsychiatry's use skyrocketed, consequent to the recent advent of the coronavirus 2019 disease pandemic, as many ambulatory clinics and doctor's offices closed to mitigate the risks of viral transmission. Numerous practices successfully adapted by surmounting previous barriers which included clinicians' lack of familiarity with the modality, required regulation-compliant electronic health records, and videoconferencing platforms. Systemic barriers were suspended or removed, such as prior exclusion from health insurance coverage for telepsychiatric services. However, problems remain. Telepsychiatric care remains unavailable for those individuals without stable high-speed internet or telephone access. Privacy can be difficult to obtain for individuals living in communal or family settings. Language barriers are compounded as many telepsychiatric systems lack abilities to provide needed support for non-English speaking individuals.

Ultimately, telepsychiatry presents a challenge to traditional psychiatric practice, which has embraced the face-to-face interpersonal encounter as near-sacred. Digital views and exchanges are several steps removed from that experience. Can the differences between the two approaches be adequately and accurately measured, and by what criteria? Would case outcomes differ or be the same? Or are the differences measurably ineffable?

Is care provided via telepsychiatry equal to that provided in person? A few short-term studies performed on adults have indicated that psychopharmacology appointments conducted virtually can achieve outcomes similar to those provided in-person. Studies with children treated for ADHD indicate that direct management and indirect collaborative care models, in which primary care physicians receive consultation from CAPs via telehealth, can achieve outcomes comparable to those achieved in-person. A couple of psychotherapeutic treatments provided digitally, specifically, CBT and ABA, result in target symptom, mood, and satisfaction scale outcomes equivalent to those produced during in-person encounters. These approaches utilize manualized forms of therapy or rely heavily on work performed by patients and caregivers without CAP presence; the CAP-delivered portions are strictly psychoeducational. Psychotherapeutic modalities conducted virtually, such as play, psychodynamically based, and interpersonal treatments, may be less efficacious than ones conducted in person. These have yet to undergo substantive telehealth delivery model research trials in the child-age population. Allied therapies, such as physical and occupational, have demonstrated limits on efficacies when performed remotely. The provision of telepsychiatric care is certainly better than none, but promoting convenience and decreased costs as justifiable reasons for offering possibly less efficacious treatments raises significant ethical concerns. Many aspects of care cannot be provided virtually, or are more difficult to perform, such as monitoring of laboratory values, vital signs, and physical examinations. Increased collaboration among CAPs, primary care practitioners, traveling nurses, pharmacies, and laboratories may help overcome some deficits. However, that degree of optimal and timely collaboration can be difficult to arrange.

Anecdotally, the lack of direct mutual interpersonal contact among children and adolescents due to school closings and education at home in response to the COVID pandemic has led to a great deal of distress among thusly affected youths. Is this phenomenon measurable by significant outcome parameters or simply to be viewed as a necessary response to the times? Similarly, do telepsychiatry and its digital distancing herald the future dominant form of care provision, or will there be a reversion to in-person face-to-face encounters? Which approach is most beneficent and least harmful, and how are these comparisons to be measured? In the meantime, while the questions are asked and studies performed, CAPs must provide good care to the best of their abilities without compromising its quality.

Social Media

A technologic communication and media revolution has taken place worldwide. With the establishment of digital means of transmission via ether and cloud, 67% of the world's citizens currently access mobile phones, 59% the Internet, and 50% social media, enabling them to receive news and instantly exchange information with each other. While censorship of these communications is exercised to different extents in varied locales, in the "free world," almost none exists. In the United States, 75% of children and adolescents have unfettered access to one or more communication media such as Facebook, Instagram, and Twitter (platforms that may last into the indefinite future or evanesce), and the nature and substance of their communications are limited only by the youths' degree of exercised self-restraint. As adolescence (and beyond) as a developmental stage is underscored by the ever-increasing importance of the peer group, decreased enforcement of parental controls, and, not uncommonly, beginnings of interpersonal romantic relationships, adolescents constantly employ social media to these ends. As is true of all maturational developments, some expressions are valuable and constructive, while others are not. The potential risks are considerable. Behaviors such as texting, canceling, ghosting, cyberbullying, hate-messaging, and the encouragement of self-destructive behaviors occur in youngsters with steady frequency, leading to despairing responses by recipients and, at times, precipitating embarrassment or worse in the perpetrators. CAPs need to be aware of these potential pitfalls and harms associated with social media use by their often naïve young patients and have a moral responsibility to alert them to their proper and less risk-laden uses.

On the other hand, the emergence of these platforms as a primary means of social discourse has positively enabled their use as media messengers for alerting others to the experience of personal distress. Mental health clinicians may become involved in the management of clinical crises following the appearance of alarming posts made by a child or adolescent. Numerous social media platforms have programmed monitoring capabilities that can alert local crisis intervention, law enforcement, or emergency medical services personnel to the postings and location of individuals whose pronouncements indicate they may be at risk of harming themselves or others.

Marijuana

During the past decade, the use of marijuana (cannabis), a mind-altering substance, has undergone public rehabilitation. Thus, three of the world's countries have legalized its recreational use, many more have decriminalized its possession, and the majority of Westernized countries generally overlook citizen possession of small amounts. Twenty-one of the United States have legalized its recreational use and possession, and in many remaining states, the trend seems to move in that direction. Simultaneously, state government-approved medical research in the United States examining the potential use of cannabinoids for treatment purposes proceeds apace. Preliminary findings suggest the potential for beneficial use of cannabinoids as antinauseants (e.g., during chemotherapy) and as treatments for chronic pain, neuropathies, pediatric epilepsy, obstructive sleep apnea, and dystonias. Other studies have focused on the possible utility of cannabinoids in the treatment of psychiatric disorders, including schizophrenia, social anxiety, and opioid use.

In the United States, decriminalization and the possible legalization of marijuana applies to adults. However, substances that become legally available to older individuals commonly trickle down to adolescents. Statistics indicate that 13% of adolescents used marijuana in 2018 and 33% of youths aged 18 to 25 years did likewise. For comparison purposes, 36% of youngsters aged 18 to 20 used alcohol, also a mind-altering substance that is legally off-limits to this and younger age groups.

While the occasional use of marijuana usually proves harmless, 6% of adolescents use the substance daily, leaving them habitually mind-altered or "stoned." Such heavy use is associated with chronically diminished or poor academic functioning, slowed cognitions, driving while intoxicated, possibly altered brain structures, potentially exacerbated anxiety and/or depressive disorders, and postulated vulnerability to the precipitation of psychosis. Perhaps most alarming is the association between chronic marijuana use in adolescents that results in 17% of those individuals developing a substance use disorder as adults.

While decriminalization may be indicated for a relatively milder mind-altering substance used on occasion by most, so is vigilance against its misuse. CAPs will commonly encounter teens who occasionally use marijuana and alcohol. It is important for the CAPs to monitor the nature and degree of use carefully in order that therapeutic management and interventions prove beneficial, neither punitive nor permissive. The CAP remains ethically required to resort to an evidence base to support intervention choices rather than personal bias.

Sexual Orientation and Gender Identity

Two significant sexual revolutions took place during the last half-century. The early 1960s witnessed the development of the birth control pill and, subsequently, other pregnancy-prevention devices, leading to a considerable increase in premarital sexual activity, including among adolescents. The very late 1960s and years following saw the beginnings of the era of “coming out,” that is, the public emergence of alternative sexual lifestyles beginning with homosexual self-expression by both genders, resulting in the orientation’s decriminalization, and subsequently involving self-proclaimed identifications as bisexual, transgender, and queer (LGBTQ). Organized psychiatry, reflecting changing societal mores and scientific research indicating the absence of intrinsic psychopathology, moved from its position of classifying homosexuality as a disorder to removing it from the DSM in 1973. The recent decades have witnessed continued vociferous demands of equal rights for, and nondiscrimination against, LGBTQ populations, ultimately resulting in the legitimization of same-sex adoptive parents and foster parenting as well as aspirants successfully attaining political office.

Development of sexual orientation and formation of gender identity solidify during childhood and adolescence, and the public expression of alternative orientations and identities in the adult world has led to adolescent, and sometimes earlier, self-declarations of alternative sexual and gender self-perceptions. Parents of LGBTQ children may respond to their offspring's declarations in a variety of ways, ranging from antagonistic and belittling to supportive and championing of their children's rights. As is true of all notable social change, significant backlash often occurs. The increased visibility and numbers of self-identifying young LGBTQ individuals fueled the development of "conversion therapies," arising prominently in conservative religious communities and countries designed to promote these individuals' adoptions of heterosexual orientations and lifestyles. Organized psychiatry has since documented negative and potentially dangerous outcomes of these therapies, labeling their use as unethical in 1998. CAPs treating LGBTQ youths are obligated to provide them with the care that supports the children's mental well-being (*beneficence*), does not cause harm (*nonmaleficence*), and respects the individual's right to self-assertion (*autonomy*). Simultaneously, while acting as the child's advocate, the professional should pay heed to adult caretakers who respond with distress to their child's inclinations by supportively aiding the parents' adjustment to the pertinent realities. At the same time, the CAP should never adopt a stance that might damage the child's comfortable and self-accepting expression of alternative sexual orientation or gender identity.

Pharmaceutical Industry

For years concerns have been voiced about the complicated relationships between medicine and the pharmaceutical industry. Pressure has been brought to make public traditionally private information detailing the financial relationships between individual physicians and pharmaceutical companies. Drug companies are now required to make all data derived from their studies publicly available, including negative results. Data has been collected on the number of children and adolescents prescribed multiple psychotropic medications, often with unclear indications. Medical schools imposed restrictions on interactions and exchanges between industry representatives and physician trainees and practitioners. Professional organizations instituted more stringent policies governing their dealings with the pharmaceutical industry. These actions represent continuing angst concerning the ethical dimensions of industry–physician and industry–psychiatrist relationships, in turn stimulating ongoing advocacy for transparency in their partnerships.

The industry, “Pharma,” produces the medications prescribed by physicians. It also funds the bulk of child and adolescent psychiatrist research, directly and indirectly, funds the work of academic researchers, and underwrites and provides financial support for professional journals, conferences, and lectures. The industry engages in these practices, in largest part, because the component companies are for-profit concerns, beholden to their shareholders, and their primary goal is to maximize the purchase of their products. The results of Pharma’s largesse to child and adolescent psychiatrist professionals can and does result in increased discovery and dissemination of new knowledge, expanded evidence-based understandings, and improved patient care. However, given the industry’s desire to subtly as well as overtly influence practice, practitioners need to remain intelligent consumers of the professional literature in order to judge the limitations of research results and their interpretations.

CAPs might unwittingly be induced to prescribe more expensive, in contrast to less expensive, medications or brands in contrast to equally effective generic drugs, thus unnecessarily increasing costs for their patient’s families. The influence of Pharma on clinical practice can lead to an increased emphasis on drug studies in professional journals, at the possible expense of studies devoted to nonpharmacologic components of the child and adolescent psychiatrist’s treatment arsenal. These possibilities, coupled with the current healthcare indemnification environment, could lead to excessive child and adolescent psychiatrist dependence on drug prescription and, on occasion, its excessive use. In addition, the expansion of criteria to include youngsters in diagnostic categories from which they were formerly essentially excluded, as in the case of bipolar disorder, or the creation of new categories, for example, disruptive mood dysregulation disorder, might be perceived by some observers as a consequence of the influence of Pharma as easily as it would be viewed by others as real and considerable psychiatric advances.

Thus, the relationships between Pharma and the profession of child and adolescent psychiatry, as well as with its individual practitioners, raise potential conflicts of interest. Can these conflicts be managed, or should they be eliminated, by sundering all but indirect contacts and contributions between the industry and the profession? The ethical dilemmas are many and profound, and the choices made by individual practitioners and professional organizations will significantly affect future patient care, research funding, basic scientific knowledge, and professional education.

Funding and Provision of Patient Care

In the United States, healthcare costs have risen steadily for decades for individuals, employers, and the government. At the same time, many have voiced concerns about the quality of provided care. Various administrative interventions have been implemented in an attempt to improve care as well as to contain costs. The Affordable Care Act, enacted in 2010, aimed to improve access to affordable health insurance for more Americans, enhance the quality of health care, regulate the health insurance industry, and decrease health care spending. Recently the Act survived several years of political attacks. Currently, it appears likely to significantly expand its number of covered individuals and families. In addition, the Act's creation and expansion unintentionally highlighted the shortage of CAPs and other mental health providers in most areas of the country. An earlier reform enacted during the 1980s colloquially termed managed care, had a notable effect on child and adolescent psychiatry practice. Its primary goal was cost containment, which often complicated and compromised the delivery of effective psychiatric care for children and adolescents. Managed care's innovations included gatekeeper requirements, prior approval for treatment–even in emergencies, and limitations on treatment length and access to providers. Its legacy continues to limit care options, both for practitioners and families. Then and now, many of the institutions providing insurance payments for mental health care are for-profit companies with obligations to their shareholders, raising possible conflicts of interest that negatively affect their customers' and patients' health.

Sixteen-year-old Julie was psychiatrically hospitalized for expressing suicidal thoughts after having been physically assaulted by her father. Julie's parents had divorced 5 years previously, and Julie since lived with each parent at different times. Six months before the hospitalization Julie became pregnant, was "kicked out" of her mother's home, and obtained an abortion while living with her father. The recent assault was the father's response to Julie's demand to spend time "chilling" with her drug-dealing boyfriend. During the admission Julie exhibited symptoms of a major depressive disorder, although her suicidal ideations had diminished in frequency and intensity. The hospital utilization review informed the attending child and adolescent psychiatrist that the patient's insurance coverage provided a maximum 5-day stay and demanded that the patient be discharged on the 5th day. The parents stipulated "impossible" expectations of Julie in the event that she might come to live with them; Julie demanded both discharge from the hospital and the right to live with her boyfriend; and a state child protective services investigation found no need for its further involvement in the case.

One of the raisons d'être of the current versions of managed care is the pursuit of the ethical principle of *distributive justice*. In the case of health care provision, that translates into the funding of basic care for the largest number of people by controlling costs expended on the care of individual patients. Furthermore, the fact that many private health insurance companies are for-profit concerns is an obvious additional consideration when examining the motivations contributing to financially related administrative stances and decisions. Nevertheless, the child and adolescent psychiatrist, despite awareness of these "macro" concerns, continues to be bound by the professional obligation to provide beneficial care for each individual patient, independent of funding. Although, in the previous case, Julie's hospitalization coverage is about to expire (in theory, so that there will be coverage available for other individuals in need), Julie continues to require care—for her depression, protection, and sustenance. The attending child and adolescent psychiatrist remains ethically and legally responsible for attempting to make the best or least damaging decisions regarding Julie's needs for care in the immediate future. If the child and adolescent psychiatrist deemed that further hospital stay was indicated, it would be incumbent on the child and adolescent psychiatrist to request additional funding from the insurance company, appeal the company’s refusal of funding in writing, or arrange for hospitalization elsewhere. If comfortable with the discharge option but not with the current dispositional choices, the child and adolescent psychiatrist would be obligated to pursue the state child protective services agency to re-examine the case, given that it is this agency’s responsibility to provide alternative placement possibilities. Although funding mechanisms may change, the child and adolescent psychiatrist’s ethical constraint to provide beneficial care is immutable.

> Dr. Green, a child and adolescent psychiatrist, had been providing a combined treatment approach of psychotropic medication and psychotherapy to Ben, an 11-year-old boy who had developed post-traumatic stress disorder (PTSD) 6 months earlier after exposure to a stressful event. Ben’s mother informed Dr. Green that her health insurance benefits, obtained via her employment, had changed. The new insurer limited coverage for CAPs to brief medication checks, and benefit payments for psychotherapy were available exclusively for nonmedical mental health personnel. Dr. Green believed that the psychotherapists recommended by the insurer had inadequate training to handle Ben’s case.

While frustrating and time consuming, CAPs have an obligation to advocate with insurance companies in their attempts to obtain what they consider optimal care for their patients. Pediatricians share similar concerns about their psychiatrically affected patients. In response, two recent developments have emerged. Some pediatric practices have embedded mental health practitioners, including CAPs, in their practices. A second approach, often state government-sponsored, involves CAP electronic supervision of pediatricians who themselves directly supply mental health care. It is notable that in some parts of the United States, most commonly in large, urban areas, some CAPs have managed the dilemmas and conflicts arising from dealings with insurance companies by choosing not to participate with them and, instead, engaging in fee-for-service practices. While this approach does allow these psychiatrists to provide treatments as they deem appropriate, it raises major ethical issues related to access to care and clear considerations of social injustices. *Split treatment*, defined as the child and adolescent psychiatrist providing psychiatric evaluations and medication management while nonmedical mental health practitioners provide verbal therapies, has become the norm in numerous healthcare settings. It is unclear if cost savings for families and companies result; it should allow for greater CAP availability to a larger number of patients. However, as a consequence, many psychiatrists have far less knowledge of their patients than do their peers who provide combined treatments, while they yet retain an ethical responsibility for their patients' optimal care. In addition, these CAPs often have little contact with the nonmedical mental health practitioner providing verbal treatment. Reduced knowledge of patients and self- and externally-induced pressures to treat more patients in less time raise the possibilities of poorer attention to patients' needs despite the ethical constraints to provide beneficial care, not do harm, and make the patient's interests the primary concern.

Administration

CAPs serve as administrators in a variety of institutional settings and wear multiple hats. Common to child and adolescent psychiatrist administrators are responsibilities for the provision of clinical mental health services to child and adolescent populations, responses to ever-increasing budgetary pressures, and, in that context, maintenance of staff morale and cohesion. Common administrative maneuvers during times of budgetary constraints include attempts at financial savings via personnel layoffs and cutbacks, reduction of services to patients and their families, and shifting of clinical priorities and foci. For example, the child and adolescent psychiatrist administrator may consider whether funds formerly intended for adolescent inpatient care should be shifted to child day-hospital services or vice versa; whether a child outpatient clinic should be maintained or scuttled if expenses exceed income; and whether a child and adolescent psychiatry employee should be assigned an increase in caseload without a corresponding increase in salary or time availability. These considerations, maneuvers, and decisions cause distress to the administrators, patients, and professional staff.

The child and adolescent psychiatrist administrator is constrained by the ethical injunctions to not do harm and to maintain fidelity to patients. Fiscal realities also necessitate considerations of *distributive justice*, translated into the attempt to provide reasonable care to the greatest number. Considerations of these fiscal concerns can come at the expense of patients currently receiving treatment. The administrator must seek to ensure that care remains available to them—if necessary, via referral.

Administrators are also responsible for the behaviors of those who report to them.

Mrs. Smith complained to the child and adolescent psychiatrist director of a mental health clinic: “My 8-year-old has a diagnosis of oppositional defiant disorder and, with him present, Dr. Jones accused me at length of being the cause of his troubles. Since then, my son doesn’t listen to anything that I tell him to do. You better get rid of that Dr. Jones.”

The mental health administrator has the reasonable expectation that child and adolescent psychiatry personnel will perform competently and professionally, which includes according to patients and their guardians appropriate respect. Administrators must also investigate patients’ complaints. When allegations have foundation, the administrator must determine whether a professional’s misbehavior or poorly chosen words reflect a momentary aberration or are one in a series of inappropriate actions. Psychiatric work with children and adolescents has the potential for stimulating poorly restrained countertransferential responses that arise primarily from a professional’s unresolved conflicts, possibly stemming from the physician’s childhood, in contrast to behaviors reflecting an objective response to the clinical presentation. It is the child and adolescent psychiatrist administrator’s responsibility to ensure appropriate professional behavior among the caregivers under their supervision. The administrator must also distinguish between a single untoward incident that may be responsive to personal counseling in contrast to the consistent expression of ingrained unprofessional behaviors that would require a significantly stronger response. In both situations, the administrator must endeavor to ensure, as much as possible, that the behaviors do not recur and that professionalism is maintained.

Research, Neuroethics, and Psychopharmacology

Arguably, the major recent advances in child and adolescent psychiatrist knowledge and treatment have resulted from quantum leaps in knowledge of neural function, particularly at the brain receptor and molecular levels.

This knowledge has correlated with an expansion in the knowledge base of neuropsychopharmacology.

As is true of all treatment techniques, the decision of whether to intervene pharmacologically requires a risk–benefit analysis. The potential dangers of acute or chronic side effects must be weighed against the long-term effects of unhindered chronic psychopathology. Knowledge concerning the short- and long-term effects of psychotropic medications on possibly vulnerable developing brains remains a concern. Although some negative effects have correlated with medication treatments (e.g., aspirin with Reye syndrome and central nervous system stimulants with possible growth retardation), studies have not, by and large, followed psychotropic-taking children into adulthood. Many psychiatric disorders first manifesting in childhood or adolescence (e.g., autism spectrum, bipolar, Tourette's, and OCD) are likely to remain manifest as chronic conditions in the ensuing decades, and the long-term effects of the prescribed psychotropic medications are uncertain.

Psychotropic medications commonly have been prescribed by physicians for "off-label" use. For example, many medications prescribed by CAPs for child psychiatric disorders (e.g., schizophrenia, bipolar disorder) only recently received limited FDA approval and thus became legitimized as "on-label." Thus, some medications that target severe behaviorally and cognitively disruptive childhood disorders, which include antipsychotic, mood-stabilizing, antidepressant, and cardiovascular agents, might not have FDA approval for their prescription to children. In addition, prescriptions of these medications for newly defined psychiatric conditions, such as disruptive mood dysregulation disorder, have as yet to obtain FDA approval for this purpose, and their use in these circumstances would again be termed "off-label." Although considered clinical judgment can result in the conclusion that these medications, in fact, might be indicated and beneficial, a rigorous evidence base for that opinion could remain weak or nonexistent. Because many childhood behavioral disorders contain elements of various disease entities without fully satisfying established diagnostic criteria, pharmacologic management is often attempted based on coupling the overt behaviors with hints or suggestions of other signs or symptoms. Also, practitioners attempt to treat with medications because adequate and appropriate psychotherapeutic and environmental interventions often are not accessible, available, or sufficient. These diagnostic uncertainties and absences of full ranges of care, coupled with inadequate evidence bases, always pose ethical conflicts between the desire to provide good care, for example, by using what are thought to be indicated psychopharmacologic agents—and the possible iatrogenic precipitation of harm.

The lack of knowledge concerning long-term outcomes indicates the continued need for research to expand the child psychiatric scientific database. Political jockeying persists around a mandate to study the effects in children of medications prescribed to adults. While such a directive was originally imposed, that is, the "Pediatric Rule," it was, in turn, struck down by court order; the pharmaceutical industry has since been encouraged to expand drug trials to include children via the use of incentives. In any event, until an increase occurs in the number of drug trials focusing on children and their longer-term outcomes, prescribing new medications to youngsters without advanced knowledge of potential ill effects could put them at risk. Thus, informed consent is crucial in these circumstances. A description of the treatment protocol, risks and benefits that might accrue consequent to the provision or withholding of the treatment, and avoidance of coercion with respect to patient participation, should be expected by the prospective participants in what are essentially experiments whose predictable outcome is indeterminate. Although consent to a child's participation in a research protocol must come from fully informed guardians, the child should be advised of all the pertinent factors to the extent of their capacity to understand and noncoerced assent obtained for participation to occur.

The phenomenal increases in knowledge of molecular biology have widened the spectra of innovative drug treatments, not only for diseases of the brain but also for enhancement of what until now has been considered cognitive or affective functioning, or both, within that broad range called "normal." Stimulant medications, for example, are known to be effective in increasing attention and concentration not only for children diagnosed with ADHD but also in normal populations. Might drugs be used in the future to enhance the sociability of the nondisabled shy younger child or, conversely, reduce the boisterousness of the mildly annoying exuberant child? One needs only to look at controversies regarding steroid augmentation of athletic performance, or neuroenhancers intended to forestall or slow cognitive declines, to comprehend that similar improvement or magnification of unimpaired brain function abilities lie around the corner. These and related concerns are subject to frequent musings by neuroethicists.

The number of psychotropic medications prescribed for children, inside and external to the United States, has seen a significant increase over the last few decades. Potential explanations include gains in professional knowledge, studies indicating medication efficacy, more medication choices among chemical classes of psychotropic medications, access to medical and psychiatric care by larger numbers of children, large sums devoted to advertising of pharmaceuticals targeting physicians, and child and adolescent psychiatrist professional education that emphasizes "biologic" interventions at the expense of time devoted to the verbal therapies. More recently, publicly aired concerns about this general increase, highlighting polypharmacy in the child and adolescent public sector, and concerns about purported relationships between the administration of selective serotonin reuptake inhibitors (SSRIs) to depressed youngsters and subsequently increased suicidality, all occurring in the context of a pharmaceutical industry having significant interests in the marketing of its products, resulted in some curtailing of medication prescription by CAPs. Of note, it appears that the above-cited drop in SSRI prescriptions was followed by a significantly measurable increase in suicidal behaviors in the target population. That example serves as a fine illustration of the ethical dilemmas CAPs face as they weigh the risks and benefits of therapeutic choices, grapple with an enlarging psychotropic armamentarium, consider ever-broadening clinical presentations for which medication interventions are suggested, and experience increasing economic pressures and inducements to write prescriptions. In the end, it is incumbent on CAPs to weigh evidence objectively, distance themselves from personal considerations and biases, and make choices that foster the best interests of the child.

TEACHING ETHICS AND REPORTING VIOLATIONS

The American Council of Graduate Medical Education mandates the teaching of ethics in child and adolescent psychiatry training programs in order for future CAPs to be aware of the potential emergence of ethical conflicts during their careers. Topics deserving didactic exposure during training commonly include matters generic to all physicians, such as confidentiality, boundaries, professional responsibilities, and relationships with healthcare providers, companies, and industries, as well as topics focused more specifically on child and adolescent psychiatry practice. The latter include the topics of child advocacy and protection, informed consent and assent, refusal of treatment, child-guardian conflicting rights, agency, compensation sources, autonomy and paternalism, and research-related risks and benefits. In addition, attention should be paid to pitfalls surrounding communications via digital transmissions and practitioners' ethical responsibility to be aware of federal, state, and local laws governing medical and child and adolescent psychiatry practice. Clearly, ethical concerns regarding children arise from the youngsters' needs for adult protection and nurturance, given their physical and cognitive immaturity. Related training in and knowledge of psychodynamics, the origins of motivations, intrafamilial relationships, and systems theory would augment, and can easily be incorporated into, the child and adolescent psychiatrist trainee's ethics education. Further, child and adolescent psychiatrist trainees should be exposed to the contents of the AACAP *Code of Ethics.* The document, using a principled framework, addresses the needs and rights of children and their guardians, as well as the child and adolescent psychiatrist's responsibilities to both. On occasion, trainees and/or experienced practitioners may experience uncertainty regarding an ethical dilemma concerning a child, and referral to the *Code* may help guide the response.

The AACAP Ethics Committee also serves as a potential resource for advice concerning ethical dilemmas. Inquiries to the committee receive suggestions for approaches to resolution after member deliberations. When faced with an ethical dilemma, the potential value derived from consultation with colleagues, and from the committee in particular, cannot be overstressed. In recent years, inquiries presented to the AACAP ethics committee have included concerns about administrative orders to withhold information from families; "double booking" of patient appointments; provision of simultaneous care to multiple family members; transitional age youth access to firearms; inadequate research supervision; noninclusive and/or biased custodial evaluations; the use of agencies and credit bureaus to collect fees; potential breaches of patient confidentiality; poor or potentially harmful patient care; organizationally blocked access to professional consultation coupled with inadequate supervision; video monitoring on child and adolescent psychiatrist inpatient units; disclosure of case details in, and obtaining patient consents for, publications;

abandonment of children by professionals in the context of threatening parents; potential supply of sexual paraphernalia to patients; practice of child psychiatry without formal training; the impacts of television and other media on children; requests of families for funding of research; treatment by a child and adolescent psychiatrist parent of that psychiatrist's child; reporting alleged abuse/neglect to state child protection agencies; and potential adoption of former patients. In addition, the AACAP Ethics Committee has developed a range of educational materials, disseminated in various fora and maintained on the AACAP ethics website, related to child and adolescent psychiatric practice. To those ends, in the recent past, the committee focused on ethical concerns arising from the impacts of systemic racism, immigration bans, family separation, and the COVID pandemic, highlighting the latter's influence on the development and evolution of telepsychiatric practice. Finally, it remains of note that, in 2010, the AACAP implemented a mandate that continues to require of all members the acquisition of biannual ethics education in order to maintain their status in the organization.

Periodically, fulfilling a different function, on rare occasions, the committee receives complaints concerning practitioners alleged to have engaged in unethical practices in matters concerning children and adolescents. When further investigation is indicated, complainants are referred to appropriate organizations and agencies, such as the APA and AMA district branches and state medical licensing boards. Investigations by these entities could result in "no findings" or recommendations for sanctions, which may include reprimand, suspension, or loss of licensure. Abrogation of licensure by a state medical licensing board, consequent to findings of an ethics violation, automatically results in loss of membership in the AACAP, the guild representing the profession's ethos.

The roles of healers in society have a long history, and communities' designated medical personnel have always been faced with ethical dilemmas. With continued growth in relevant scientific knowledge and related technologic achievements, the dilemmas have grown more complex and complicated. The basic understandings, however, remain the same and undiluted, that is, the challenge of doing good and not engaging in harmful behaviors, equally for all patients. CAPs have the additional burden of caring for the immature youngster, as well as maintaining good relations with their guardians and caretakers. Whether the child and adolescent psychiatrist is wrestling with age-old conflicts surrounding privacy and confidentiality or dealing with novel developments in genetics and transmission technologies, ethical responsibilities represent a never-ending challenge for all practitioners and for the subspecialty as a profession. In the face of unceasing change, therefore, the teaching of ethics, discussions of current dilemmas, revisions of ethical codes, and monitoring of practitioners and practices are a must for the maintenance of the profession's integrity.

FURTHER READINGS

American Academy of Child Psychiatry. *Code of Ethics*. American Academy of Child Psychiatry; adopted in 2009, revised in 2014.

American Psychiatric Association. *The Principles of Medical Ethics with Annotations Especially Applicable to Psychiatry*. American Psychiatric Association; 2013.

Anzia DJ, LaPuma J. An annotated bibliography of psychiatric medical ethics. *Acad Psychiatry*. 1991;15:1–17.

Applebaum PS. Ethical issues in psychiatric genetics. *J Psychiatr Pract*. 2004;10(6):343–351.

Bailey ZD, Krieger N, Agenor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. *Lancet*. 2017;389:1453–1463.

Beauchamp T, Childress J. *Principles of Biomedical Ethics*. 8th ed. Oxford University Press; 2019.

Bloch S, Green SA, eds. *Psychiatric Ethics*. 5th ed. Oxford University Press; 2021.

Boydell K, Hodgins M, Pignatiello A, Teshima J, Edwards H, Willis D. Using technology to deliver mental health services to children and youth: a scoping review. *J Can Acad Child Adolesc Psychiatry.* 2014;23:87–99.

Charland L. A historical perspective. In: Bloch S, Green S, eds. *Psychiatric Ethics*, 5th ed. Oxford University Press; 2021.

Chassiakos Y, Radesky J, Christakis D, Moreno MA, Cross C; Council on Communications and Media. Children and adolescents and digital media. *Pediatrics*. 2016;138(5):e20162593.

DeCamp M, Sulmasy LS. Ethical and professionalism implications of physician employment and health care business practices: a policy paper from the American College of Physicians. *Ann Intern Med*. 2021;174:844–851.

DeJong SM. *Blogs and Tweets, Texting and Friending: Social Media and Online Professionalism in Health Care*. Elsevier Publishing; 2013.

Dingle AD, Stuber ML. Ethics education. *Child Adolesc Psychiatr Clin N Am*. 2008;17(1):187–207.

Enzer N. Ethics in child psychiatry—an overview. In: Schetky DP, Benedek EH, eds. *Emerging Issues in Child Psychiatry and the Law*. Brunner/Mazel; 1985.

Grigsby TM, Hoffmann LM, Moss MJ. Marijuana use and potential implications of marijuana legalization. *Pediatr Rev*. 2020;41(2):61–72.

Haque O, De Freitas J, Bursztajn H, et al. *The Ethics of Pharmaceutical Industry Influence in Medicine*. UNESCO Chair in Bioethics, Ministry of Education; 2013.

Lew C. What do new neuroscience discoveries in children mean for their open future? In: Illes J, Hossain S, eds. *Anticipating the Future*. Oxford University Press; 2017.

Namerow LB, Walker SA, Loftus M, Bishop JR, Ruaño G, Malik S. Pharmacogenomics: an update for child and adolescent psychiatry. *Curr Psychiatry Rep*. 2020;22:26.

Ptakowski KK. Advocating for children and adolescents with mental illnesses. *Child Adolesc Psychiatr Clin N Am*. 2010;19(1):131–138.

Rodgers SM. Transitional age lesbian, gay, bisexual, transgender, and questioning youth: issues of diversity, integrated identities, and mental health. *Child Adolesc Psychiatr Clin N Am*. 2017;26:297–309.

Ross EL, Zuromski KL, Reis BY, Nock MK, Kessler RC, Smoller JW. Accuracy requirements for cost-effective suicide risk prediction among primary care patients in the US. *JAMA Psychiatry*. 2021;78:642–650.

Sabin JE, Skimming K. A framework of ethics for telepsychiatry practice. *Int Rev Psychiatry*. 2015;6:490–495.

Schetky D. Boundaries in child and adolescent psychiatry. *Child Adolesc Psychiatr Clin N Am*. 1995;4(4):769–778.

Sondheimer A. Ethics and international child and adolescent psychiatry. In: Rey J, ed. *IACAPAP e-Textbook of Child and Adolescent Mental Health Care*. International Association of Child and Adolescent Psychiatry and Allied Professions; 2012.

Sondheimer A. Ethics and risk management in administrative child and adolescent psychiatry. *Child Adolesc Psychiatr Clin N Am*. 2010;19(1):115–129.

Sondheimer A, Jensen P, Karnik N. Ethics and child and adolescent psychiatry. In: Bloch S, Green SA, eds. *Psychiatric Ethics*. 5th ed. Oxford University Press; 2021.

Sondheimer A, Klykylo W. The ethics committees of the American Academy of Child and Adolescent Psychiatry and the American Psychiatric Association: history, process, education, and advocacy. *Child Adolesc Psychiatr Clin N Am*. 2008;17(1):225–236.

Stoll J, Sadler JZ, Trachsel M. The ethical use of telepsychiatry in the Covid-19 pandemic. *Front Psychiatry*. 2020;11:665.

Trent M, Dooley D, Douge J. The impacts of racism on child and adolescent health. *Pediatrics*. 2019;144(2):e20191765.

UN Convention on the Rights of the Child. Adopted November 20, 1989. https://www.ohchr.org/en/ohchr_homepage. Accessed December 1, 2023.

UN Study on Violence Against Children. 2003. https://violenceagainstchildren.un.org/content/un-study-violence-against-children. Accessed December 1, 2023.

Vitiello B. Ethical considerations in psychopharmacological research involving children and adolescents. *Psychopharmacology (Berl)*. 2003;171:86–91.

Walter G, Rey JM, Thomas CR, Martin A. Conflict of interest. In: Martin A, Scahill L, Kratochvil CJ, eds. *Pediatric Psychopharmacology Principles and Practices*. 2nd ed. Oxford University Press; 2011:738.

54.8 School Consultation and Collaborating With the Educational System

Sheryl Kataoka, M.D., M.S.H.S., Ilaina Blum, M.A., M.F.T., Edward Castelo, M.D., and Roya Ijadi-Maghsoodi, M.D., M.S.H.P.M.

No setting other than a child's home can influence child development more than schools. In addition to the academic curriculum taught throughout a child's education, schools can be instrumental in teaching students social–emotional learning (SEL) skills such as communicating with peers and adults, problem solving, and recognizing and regulating emotions, as stated clearly by Taylor et al. Wang et al. reported that the environment or climate of a school also plays an important role in shaping students' ability to develop cognitively, psychologically, socially, and behaviorally. Thus, whether a psychiatrist is providing treatment for a youth or consulting with a school or district, having knowledge about the school system and how to coordinate with a patient's educational setting can be essential to supporting students' mental health and well-being.

EDUCATIONAL OUTCOMES AND MENTAL HEALTH

Educational outcomes, such as achievement, school attendance, discipline, and high school (HS) graduation, are associated with youth mental health and well-being, as reported by Henry et al. One indicator of negative school performance is chronic school absenteeism, defined as missing at least 10% of school days per year, whether excused or unexcused absences. Students may frequently be absent from school for several reasons, including having a chronic illness, involvement in the juvenile justice system, housing instability, feeling unsafe at school due to harassment or bullying, teen pregnancy, or not finding value in attending school. Risk factors for chronic school absenteeism include children who have experienced maltreatment or other types of trauma, who have a higher number of adverse childhood experiences (ACEs), and who have a disability or special health care need. Allison et al. found that chronic school absenteeism can influence functioning into adulthood, affecting employment, use of public assistance, and justice involvement. For example, Porche and colleagues found that childhood trauma, mediated by psychiatric disorders, was related to school dropout. Others have also found relationships between school attendance and psychiatric problems such as attention deficit hyperactivity disorder (ADHD), anxiety and depressive disorders, and developmental disabilities. Thus, psychiatric treatment can impact both the mental health of youth and their school performance, with early intervention having the potential to prevent a negative developmental trajectory.

Overview

In this section, we first provide a broad description of the school system with which psychiatrists may interface when working with children. We will present the historical context and relevant policies that have shaped the U.S. educational system, racial and ethnic disparities that exist in schools, common types of schools and the roles of school personnel, and the range of resources, support, and programming that can be found on campuses.

Next, we discuss possible roles a psychiatrist may have with schools, including (1) collaborating with schools to support the treatment plan of their patients, (2) practicing at a school-based health center (SBHC) or through telepsychiatry as part of a school mental health team, (3) consulting to a classroom, school, or district to provide system level supports and mental health education to staff and parents, and (4) consulting to the education system in a time of crisis.

SCHOOL SYSTEMS

Common Terms Used in the Educational System

When discussing educational supports, specifically special education services and programming, psychiatrists should be aware of the “language” of schools. Below is a list of commonly used terms, although differences exist by region (see Table 54.8–1). Refer to the local school district or state’s Department of Education for specific information.

Special Education, according to the Individuals with Disabilities Education Act (IDEA), has six foundational principles: (1) free and appropriate public education (FAPE), (2) appropriate evaluation, (3) Individualized Education Program (IEP), (4) least restrictive environment (LRE), (5) parent and student participation in decision making, and (6) procedural safeguards to ensure that the rights of children with disabilities and their parents are protected and that they have access to the information needed to effectively participate in the process as described by Zirkel (2015).

Table 54.8–1.
Common Education Terms

Common Education Terms	Description
LEA: Local Educational Agency	Public authority tasked with administrative control of schools in a region (i.e., county education office, district)
General/Regular Education	Education for typically developing children based on annually reviewed state standards
Special Education	Specially designed instruction under IDEA to meet needs of child with a disability; can include gifted instruction
IDEA: The Individuals with Disabilities Education Act	Federal law: free appropriate public education for children with disabilities, for special education and related services
FAPE: Free and Appropriate Public Education	In IDEA and Section 504, describes education that's at no cost to parents, meets state standards, and meets student's needs
LRE: Least Restrictive Environment	In IDEA, appropriate educational environment, meets special educational needs, and maximizes being with nondisabled peers
IEP: Individualized Education Program	Written statement for each student with a disability that describes how disability affects education, goals, progress, and services
Accommodations	Changes in how student learns material or demonstrates mastery. Does NOT change what is taught or tested
Modifications	Changes to content and performance requirements. DOES include changes to what is taught or tested

Table 54.8–2.

Examples of Key U.S. Federal Laws Involving the Educational System, by Year

Year	Policy/Legislation	Significance
1965	ESEA: Elementary and Secondary Education Act	Part of President Johnson's "War on Poverty." For equal educational opportunity for all students (i.e., those living in poverty, learning English, disabled). Provided grants to low-income districts, funds for special education centers, scholarships for low-income college students, grants to LEAs to improve the quality of primary/secondary education
1973	Section 504, Rehabilitation Act	Protects the rights of individuals with disabilities, including in public schools (enforced by the Office of Civil Rights) and entitles children with disabilities to a "free and appropriate education."
1974	FERPA: Family Educational Rights and Privacy Act	A federal law that protects the privacy of student education records applies to all schools receiving federal funds. Gives parents certain rights to child's education records until child is age 18 (then right goes to child). *See the section on HIPAA vs. FERPA later in this section.*
1975, 2004	IDEA: Individuals with Disabilities Education Act (Formerly Education for All Handicapped Children Act)	Mandates free and appropriate public education to eligible children with disabilities in the United States, by providing them with special education and related services in the least restrictive environment.
1987	McKinney-Vento Homeless Assistance	In addition to other programs, this act includes the Education for

	Act	Homeless Children and Youth Program
2001	NCLB: No Child Left Behind Act	Contained standards-based education reforms, established high standards and measurable goals to improve outcomes of students defined by states; federal funding was dependent on states implementing assessments.
2015	ESSA: Every Student Succeeds Act	Reauthorization of the ESEA, replacing NCLB. Retained annual standardized testing requirements but allows LEAs to apply for subgrants for local accountability plans. Includes the reauthorization of the McKinney–Vento Homeless Assistance Act, which requires schools to identify students experiencing homelessness and collaborate with other service providers and emphasizes school stability.

Educational History and Policies Influencing Mental Health in Schools

As Flaherty and Osher (2003) described, mental health clinicians have collaborated with school personnel for over 100 years. From the inception of school mental health during the Progressive Era, beginning in the 1890s, schools started expanding to include more students and introducing services such as “mental hygiene” in response to academic and behavioral challenges in the educational setting. Since the latter half of the 20th century, U.S. schools have been shaped by the Civil Rights era and federal legislation that initiated a focus on addressing equity in education for those students with disabilities and those living in underresourced neighborhoods. Table 54.8–2 highlights some of the key federal education laws that have and continue to shape practice.

Disparities in the Educational System

Despite federal legislation that has called for equity in public education for over 50 years, there remain significant racial and ethnic disparities for multiple educational outcomes, affecting children across their lifespan as reported by Carter et al. (2017). For example, according to the U.S. Department of Education, Black preschool (PS) children experience disproportionate PS suspensions at rates over 2.5 times that of their enrolled population in the United States, with similar out-of-school discipline rates continuing into their elementary and secondary education. Native American youth are suspended at four times their enrolled population during K-12 schooling and expelled from school at six times their population. Native American and Black youth are also disproportionately referred to and arrested by law enforcement for school-related offenses at least twice the rates of their enrolled populations. In 2019, the U.S. Commission on Civil Rights reviewed racial and ethnic disproportionality of school discipline for students with disabilities. The report concluded that despite disproportionate discipline for students of color, these students did not commit more disciplinable offenses than their White peers, although received harsher and longer punishments than did White students. They summarized research showing that out-of-school discipline is associated with poor classroom engagement, greater school dropout, and juvenile justice involvement. The Commission recommended that federal nondiscrimination laws needed to be enforced, and schools should be provided resources and training for alternatives to out-of-school discipline. Psychiatrists can play an important role in supporting schools in these efforts, including consulting to educational systems on implementing trauma-informed practices, and on improving school climate, including student support, engagement, and school connectedness, as reported by Ijadi-Maghsoodi et al. (2017) and Kataoka et al. (2018).

Common Educational Settings

Psychiatrists may find that their youth patients attend a wide array of educational settings, each with varying resources and regulations (Table 54.8–3). The strengths and limitations of a school setting also need to be considered, in light of the strengths and needs of each individual child. The following is a description of some of the more common types of educational settings. For more information, see the Institute of Education Sciences Fast Facts (2021).

Table 54.8–3.
Educational Environment for All Students With Disabilities Under IDEA, 6–21 Years Old, 2018

Educational Environment	Proportion of Students With Disabilities
Regular school, with <40% time in general education	13.1%
Regular school, with 40–79% time in general education	18.0%
Regular school, with >79% time in general education	64%
Separate school for students with disabilities	2.7%
Separate residential facility	0.2%
Parentally placed regular private school	1.4%
Homebound/hospital placement	0.4%
Correctional facility	0.2%

Data from U.S. Department of Education. Digest of education statistics. https://nces.ed.gov/fastfacts/display.asp?id=59

Early Education Settings. High-quality early care and education centers (i.e., PSs) that serve children from birth to pre-kindergarten can play an important role in social–emotional and behavioral development as well as readiness for academic learning. One well-studied program for young children, the HighScope Perry Preschool Project, found that attending this PS resulted in greater positive long-term outcomes than for similar peers that did not attend this PS, such as fewer teen pregnancies and greater HS graduation, home ownership, and higher earning jobs. Head Start and Early Head Start are federal efforts to provide enrichment from birth to 5 years old for low-income families, which has also been found to have positive long-term outcomes for children. However, PSs have not been universally available to all children nor of universally high quality. The Institute of Medicine and National Research Council’s report, Transforming the Workforce for Children Birth Through Age 8: A Unifying Foundation (2015), outlines the system changes needed to ensure high-quality PS settings, such as core competencies in practices that support the growth and development of young children and engage diverse families.

Public Schools. All students in the United States have access to free education from kindergarten through 12th grade and are required to participate in education as defined by state laws (by attending a public, private, or some homeschooling programs). Public schools are funded primarily by state and local governments, with the remaining funding from the federal government. Public schools must comply with state and federal regulations including IDEA and the Every Student Succeeds Act (ESSA) (see earlier Policy section for details). In 2016, 90.5% of U.S. students attended a public school, with 4.6% of all students attending a public charter school.

Charter Schools. Charter schools were first established in 1992 in Minnesota, as an alternative to public schools. A charter school is a public school that operates under a legislative contract, or *charter*, with an authorizer or sponsor. Sponsors can include state education agencies, special boards, or even school districts. Charter schools are given autonomy and flexibility in curriculum, structure of the school day or year, and budget management but must meet the accountability standards outlined in the school's charter. If a charter school is not meeting these standards, the sponsor can revoke the charter. These schools are free and open to the public and if seats are not available, admission is by lottery.

Private Schools. The majority of students who attend private school, attend a religious (most frequently Catholic) school. In 2016, of all students in the United States, 7.6% attended a religious private school and 1.9% attended a nonsectarian private school. The Department of Education does not have jurisdiction over private elementary and secondary schools that are parentally placed (as opposed to publicly placed nonpublic schools, described below). Unless a private school receives federal aid, the federal government does not have control over such things as the school's curriculum or personnel and does not accredit these schools. See additional information in the later section entitled, *Private Schools and Special Education.*

Publicly Placed Students in Nonpublic Schools. Students with a disability may be placed in a private school—at no cost to their parent or guardian—to receive the special education and related services that they require in order to access their education, under IDEA. Students in these schools have an IEP that determines that the nonpublic school setting is the LRE for them to access the curriculum.

Homeschooling. According to the National Center for Education Statistics, homeschool students are students who receive their instruction at home instead of at a private or public school (excludes those enrolled in public or private schools for at least 25 hours per week or schooled at home due to a temporary illness). Regulation of home schools is state dependent as there is no federal law that governs or mentions homeschooling.

Alternative Programs/Schools. There is a subset of students in public, general education schools who might be considered “at risk” due to having a higher probability of failing academically or dropping out of school. Often, these students face life circumstances, inequities, and structural risk factors, that can compromise their ability to successfully complete school, such as homelessness, teenage pregnancy, domestic violence, transiency, incarceration, or serious health issues. For students with these challenges, who do not qualify for special education services, yet struggle to remain in their general education program and who want to stay in school and earn a diploma, public schools may offer alternative programs or schools. According to the National Center for Educational Statistics, placement in alternative schools and programs is usually based on recommendations from school personnel or administrators or requested by the parent or student.

Most alternative programs or schools are part of the public HS (though in some areas it extends to middle school [MS] as well). Alternative programs can include a separate classroom or school within a general education campus or an alternative school on a separate campus. Some alternative programs are specially designed for teenage mothers. These alternative programs vary widely by state and Local Educational Agency (LEA) according to Addis et al.

Roles of School Personnel

When psychiatrists work with child and adolescent patients, it is often indicated and beneficial to communicate with school personnel and become familiar with the culture of the school and community. This can help maximize understanding of the student's experience and inform the treatment plan. Below is a brief description of various school personnel who might be involved in a student's school programming:

Teachers are the primary school staff who typically are the most involved with students. According to the U.S. Department of Education in 2017, elementary school (ES) teachers have approximately 23 students for 6 hours each day; MS and HS teachers may have approximately 20 to 25 (or more) students per class, often teach four or more classes per day, and usually teach specific subjects.

Aides or paraprofessionals ("paras") often are assigned to work with a specific student and work under the classroom teacher. Aides can be in a general education or special education class and may accompany the student to activities outside of the classroom (e.g., lunch, recess).

Special education teachers are credentialed to provide alternative instruction to smaller groups of students with learning disabilities, emotional disorders, autism, speech/language deficits, developmental delays, and/or other handicapping conditions that are interfering with learning. For students to receive instruction from special education teachers, they must have an IEP.

Resource Specialist Program (RSP) teachers mostly work with students who have mild or moderate disabilities in either a push-in (i.e., in the general education class) or pull-out (i.e., in a separate, nongeneral education class, such as a resource room) approach. Some resource teachers administer the achievement testing for the IEP assessments.

School mental health counselors are often the primary mental health support within a school. They often provide short-term therapy and crisis support to students on campus. Students do not necessarily need an IEP to meet with a school mental health counselor.

School academic/guidance counselors often primarily assist students in academic or vocational planning and class selection. They may intervene with issues that arise at school but do not typically provide specialized or mental health interventions.

School psychologists conduct educational assessments and help construct IEPs. They may also provide individual and group therapy to students who qualify for counseling services through an IEP, and they often work with several schools in the district, so may not be assigned to one school full time.

School adjustment counselors and/or social workers have been added to the school staff in many locations, with their primary role being to provide psychosocial interventions to students and sometimes their families. Students do not necessarily need an IEP to meet with a school adjustment counselor or school social worker.

School nurses and nurse practitioners address acute health care needs of students, provide health education, and administer medications to students. School nurses usually travel between several schools and may be involved with storing medications taken during the school day.

Occupational therapists (OTs) work with students individually or in small groups to address challenges or deficits in areas such as fine and large motor skills (including writing and using tools) as well as sensory integration and/or sensitivities. They may be involved with developing alternative strategies in the classroom and/or individualized interventions. The OT usually is involved with a student as part of a recommended IEP service.

Speech–language therapists/pathologists (SLP) meet with students who qualify for such services through an IEP and have speech (e.g., articulation), language, and/or communication and social skills difficulties. They meet individually and/or in small groups.

School resource officers (SROs) are employed by the local police department, sheriff's agency, or school system and work closely with school administrators for the purpose of creating a safer environment. The responsibilities of SROs are similar to other police officers (i.e., have the ability to make arrests, respond to calls for service, and document incidents).

School administrators (i.e., principals, vice principals) coordinate between and within the school and the community by managing all services (teaching, discipline, custodial) within their school building and reporting to the *superintendent.*

Superintendents guide educational activities among all the schools within a school district and report to an elected school board in public schools, or an appointed board in private, parochial, or charter schools.

SCHOOL RESOURCES AND PROGRAMMING

In addition to becoming familiar with the types of personnel in a school system, a psychiatrist should also understand the range of resources and programming that may be available at a school. Schools are an ideal setting to disseminate effective health promotion and prevention interventions, provide early detection of mental health conditions, and link students and families to needed services. One overall framework for considering the integration of all supports on a campus is the multitiered systems of support (MTSS), which describes supports generally available to all students on a school campus. Two specialized programs for eligible students include a 504 Plan and an IEP.

If there is suspicion, concern, or evidence that a student is struggling in school, there are various avenues that can be taken to provide supports for that student.

1. A student can receive *tiered interventions* within the general education school program.
2. A student can formally receive accommodations/modifications/supports through a *504 Plan* within the general school program.
3. A student can be assessed by the public school to determine if the student qualifies for special education services through an *IEP.*

Multitiered Systems of Support

Response to Intervention (RTI) is an MTSS used by teachers and other school staff as a set of steps to help a child improve in an area of need by providing targeted educational supports within the general education setting according to Stoiber and Gettinger (2016). Interventions in a multitiered system gradually increase in intensity from Tier 1 to Tier 3 (though some schools include additional tiers), and a student may receive tiered interventions in one, two, or all three tiers until their academic or behavioral needs are met, or until the student needs an IEP evaluation. Progress is tracked within each tier and if the student does not improve within one tier, they will move to the next tier. As the student exhibits sufficient progress, the student may move back to a previous tier. Maintaining the student in the general education setting, while providing the requisite supports, is often a central goal of the tiered approach. This tiered approach, as detailed by Stephan (2015), applies to the whole child, in supporting instruction as well as the social, emotional, and behavioral development of the student (Fig. 54.8–1).

Tier 1 is usually a whole-class approach, typically applied to over 80% of the school population. From an instructional perspective, a classroom teacher may use a research-based curriculum. Similarly, from a social–emotional perspective, students may receive health promotion and universal prevention programming in the classroom as part of Tier 1 supports, such as SEL, which teaches students how to manage emotions and develop supportive relationships.

Tier 2 usually involves more focused instruction of a specific skill, utilizes a different instructional method and at a higher frequency than that provided in Tier 1, and is often provided in smaller group settings than in Tier 1, which allows for relatively more attention and guidance. Approximately 15% of general education students access Tier 2 supports, which can include selective prevention interventions in group modalities.

For Tier 3 services, students may receive highly individualized instruction or may work in very small groups multiple times per week. The student will continue to spend most of the day in a general education classroom. Poor progress following an adequate trial of Tier 3 intervention may result in consideration of an evaluation for special education services through an IEP. Approximately 5% of students will receive Tier 3 services. For social–emotional needs, this could be individual or family therapy.

Psychiatrists can help schools identify and implement appropriate programs depending on the needs and resources of the school system. Several repositories of evidence-based and promising practices have been developed to assist practitioners and agencies (including school districts and schools) select scientifically supported interventions based on designated criteria (e.g., age, setting, and target outcome). See Table 54.8–4 for evidence-based social, emotional, and behavioral interventions by tier of support and presented by PS, ES, MS, and HS.

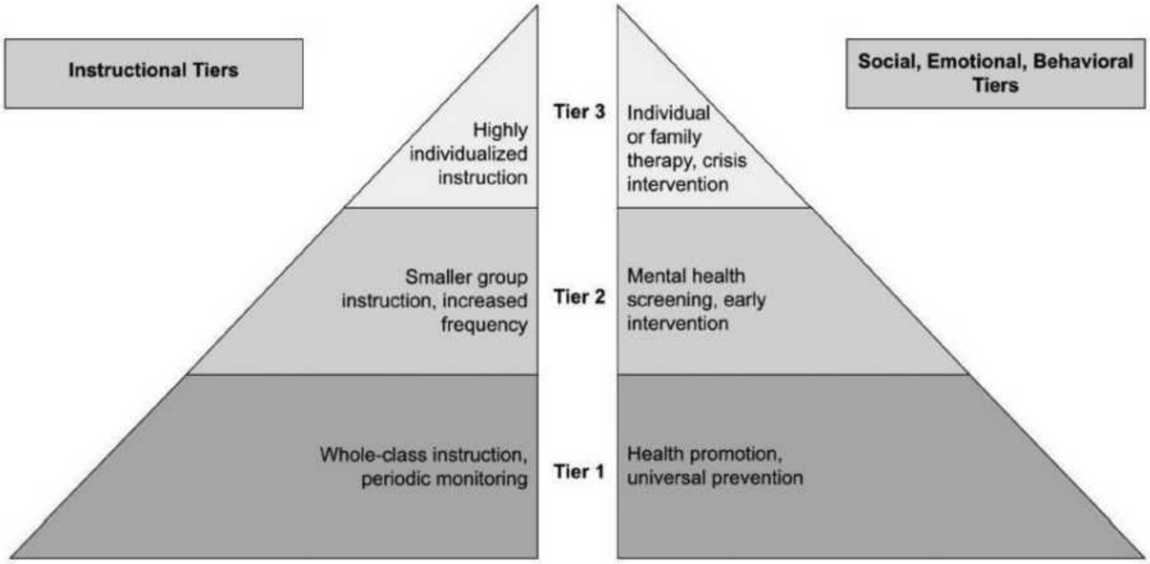
Instructional Tiers
Social, Emotional, Behavioral Tiers
Tier 3
Highly individualized instruction
Individual or family therapy, crisis intervention
Tier 2
Smaller group instruction, increased frequency
Mental health screening, early intervention
Tier 1
Whole-class instruction, periodic monitoring
Health promotion, universal prevention

FIGURE 54.8–1. Tiered systems of instructional and social, emotional, and behavioral supports.

Table 54.8–4. Examples of School Social, Emotional, and Behavioral Interventions by Tier of Support

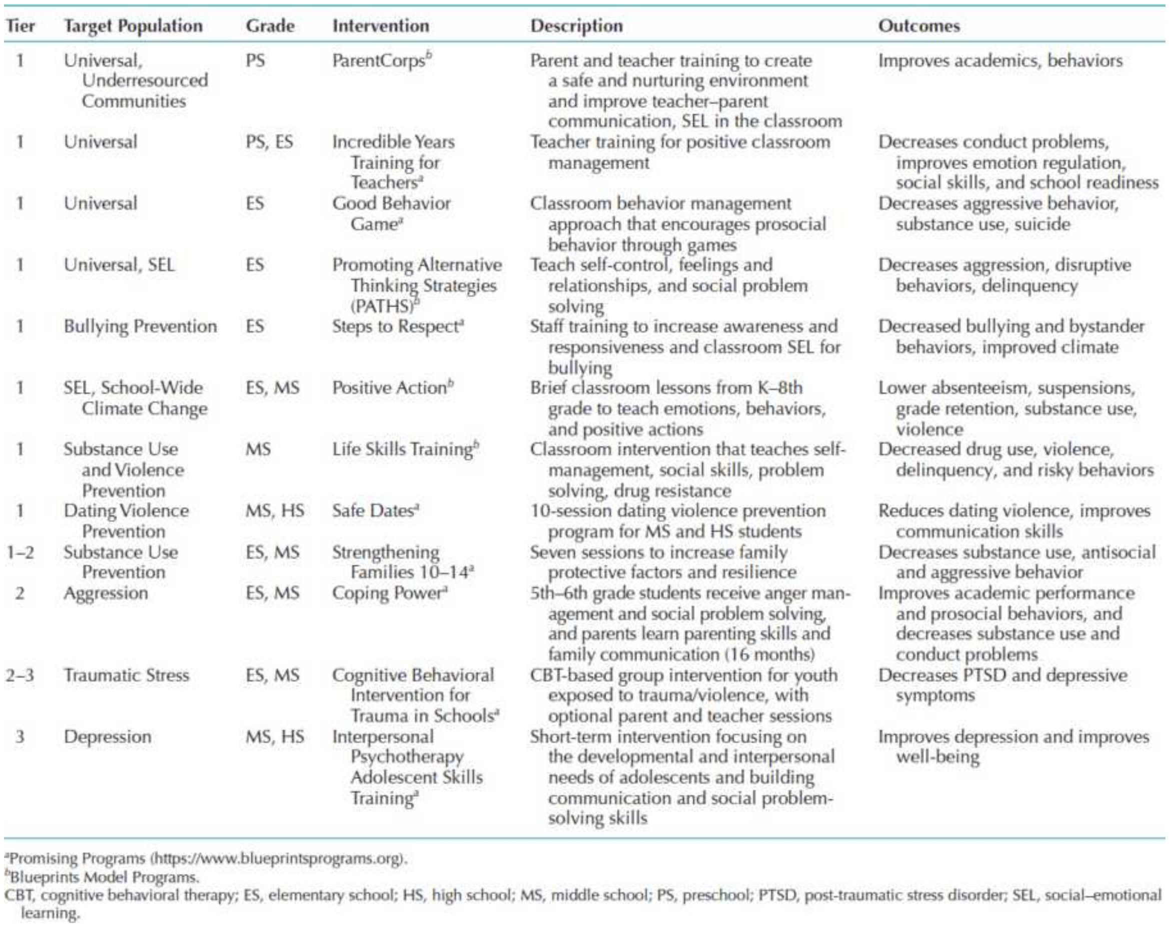

Tier	Target Population	Grade	Intervention	Description	Outcomes
1	Universal, Underresourced Communities	PS	ParentCorps[b]	Parent and teacher training to create a safe and nurturing environment and improve teacher–parent communication, SEL in the classroom	Improves academics, behaviors
1	Universal	PS, ES	Incredible Years Training for Teachers[a]	Teacher training for positive classroom management	Decreases conduct problems, improves emotion regulation, social skills, and school readiness
1	Universal	ES	Good Behavior Game[a]	Classroom behavior management approach that encourages prosocial behavior through games	Decreases aggressive behavior, substance use, suicide
1	Universal, SEL	ES	Promoting Alternative Thinking Strategies (PATHS)[b]	Teach self-control, feelings and relationships, and social problem solving	Decreases aggression, disruptive behaviors, delinquency
1	Bullying Prevention	ES	Steps to Respect[a]	Staff training to increase awareness and responsiveness and classroom SEL for bullying	Decreased bullying and bystander behaviors, improved climate
1	SEL, School-Wide Climate Change	ES, MS	Positive Action[b]	Brief classroom lessons from K–8th grade to teach emotions, behaviors, and positive actions	Lower absenteeism, suspensions, grade retention, substance use, violence
1	Substance Use and Violence Prevention	MS	Life Skills Training[b]	Classroom intervention that teaches self-management, social skills, problem solving, drug resistance	Decreased drug use, violence, delinquency, and risky behaviors
1	Dating Violence Prevention	MS, HS	Safe Dates[a]	10-session dating violence prevention program for MS and HS students	Reduces dating violence, improves communication skills
1–2	Substance Use Prevention	ES, MS	Strengthening Families 10–14[a]	Seven sessions to increase family protective factors and resilience	Decreases substance use, antisocial and aggressive behavior
2	Aggression	ES, MS	Coping Power[a]	5th–6th grade students receive anger management and social problem solving, and parents learn parenting skills and family communication (16 months)	Improves academic performance and prosocial behaviors, and decreases substance use and conduct problems
2–3	Traumatic Stress	ES, MS	Cognitive Behavioral Intervention for Trauma in Schools[a]	CBT-based group intervention for youth exposed to trauma/violence, with optional parent and teacher sessions	Decreases PTSD and depressive symptoms
3	Depression	MS, HS	Interpersonal Psychotherapy Adolescent Skills Training[a]	Short-term intervention focusing on the developmental and interpersonal needs of adolescents and building communication and social problem-solving skills	Improves depression and improves well-being

[a] Promising Programs (https://www.blueprintsprograms.org).
[b] Blueprints Model Programs.
CBT, cognitive behavioral therapy; ES, elementary school; HS, high school; MS, middle school; PS, preschool; PTSD, post-traumatic stress disorder; SEL, social–emotional learning.

504 Plan

If a student needs more formalized supports within the general education setting than can be provided within the tiered model but is not yet considered for special education services as part of an IEP, the school may develop a 504 Plan. This can follow a trial of tiered support or be developed and implemented on its own without a student being tried in any other support systems. A 504 Plan refers to section 504 of The Rehabilitation Act of 1973 that addresses protections for students with disabilities.

Section 504 states: “No otherwise qualified individual with a disability in the United States… shall, solely by reason of her or his disability, be excluded from the participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance….”

A 504 Plan formalizes specific supports, accommodations, and/or modifications that the school will implement for a student based on the student’s identified “impairment.”

Students qualify for a 504 Plan if:

- They have a physical or mental impairment.
- The impairment substantially limits at least one major life activity.

Physical or mental impairment: Some examples of physical or mental impairments are physical disabilities, health conditions, mental health disorders, and learning disabilities. Depending on the severity and duration of the limitation, *short-term impairments* (such as a leg fracture) and episodic impairments such as asthma may qualify a student for 504 accommodations.

Examples of major life activities: Some examples of limits in major life activities that would be included in a 504 Plan qualification include deficits in one or more of the following: caring for oneself, communicating, eating, learning, reading, seeing, sleeping, speaking, or walking.

504 Plans identify adjustments (i.e., supports, accommodations, or even modifications) that a student with an impairment needs in order to sustain participation in and access to the general education classroom. A 504 Plan formally delineates ways the school will help the student access the curriculum and benefit from instruction. Because the supports, adjustments, accommodations, and/or modifications developed and written in a 504 Plan are usually intended to be implemented within the general education setting, general education teachers are typically the primary, or sole, school staff responsible to implement or monitor the accommodations or supports documented within a 504 Plan.

Craig is an 8-year old boy in second grade. He had been asked to leave several private preschools due to significant motor activity and marked irritability, such as becoming upset with teachers and peers over trivial things. After entering kindergarten at a public school, parents continued to get complaints from teachers about Craig's difficulties with inattention, sitting still, tracking oral instructions, and initiating work. His irritability with peers also persisted. Therefore, his kindergarten and first-grade teachers discussed his need for a possible diagnostic assessment. A child psychiatrist evaluated Craig and diagnosed him with ADHD, combined type.

At the end of first grade, the school held a 504 Plan meeting to discuss *accommodations*. Given a "disabling condition" of ADHD, the school determined he met criteria for a 504 Plan and recommended accommodations including preferential seating, frequent checks for understanding, movement breaks, concise and neutral redirection (i.e., nonverbal redirection, such as tapping on his desk), and developing a recess plan (i.e., generating a list of activities with teachers and peers he would try to play with; clarifying when to go to the yard teacher).

There are unlimited types of supports that can be included in a 504 Plan. *Accommodations* are changes in how a student might learn material or be required to demonstrate skill mastery (i.e., be tested), but what is being measured is not changed. Accommodations do not change what is being taught and/or what the student is being tested on. Students receiving accommodations are typically subject to the same grading scale as students who are not receiving accommodations as reported by Harrison et al. (2013).

Modifications are changes made to the content and performance requirements for a student, usually based on or in response to an identified area of deficit, weakness, or disability. This can include a change in what is being assessed, assigned, or taught as compared to the general education curriculum. A distinguishing factor between accommodations and modifications is that modifications refer to the process of altering what a student learns or receives on a test (e.g., a sixth-grade student who might need fourth-grade–level spelling words) whereas accommodations refer to supports that help a student access the general education curriculum (e.g., a sixth-grade student who is not required to write spelling sentences for all 20 weekly spelling words). Modifications are more intensive changes to the difficulty level and/or the quantity of material to be learned. Modifications are usually required for students whose impairments necessitate more intense adjustments. 504 Plans can be revised as indicated and should be updated as a student progresses from one grade level to the next and especially at major grade/school changes (i.e., elementary to MS; MS to HS). See Table 54.8–5 for examples of specific types of accommodations and modifications.

504 Plans and Postsecondary Programming. Unlike services provided through an IEP, which discontinue once a student formally exits HS (i.e., via graduation, taking and passing the GED or HS proficiency examination) or reaches 22 years old, *accommodations provided as part of a 504 Plan can be carried on past HS.* Students entering post-HS educational programs (e.g., colleges and universities) can continue to receive 504 Plan support since most such educational programs receive federal funding. Public and private postsecondary institutions are required to provide accommodations to students with disabilities under either the Americans with Disabilities Act (ADA) Title II or Section 504. According to the U.S. Department of Education Office of Civil Rights (OCR),

> OCR enforces Section 504 in programs and activities that receive Federal financial assistance from [the Department of Education] ED. Recipients of this Federal financial assistance include public school districts, institutions of higher education, and other state and local education agencies. The regulations implementing Section 504 in the context of educational institutions appear at 34 C.F.R. Part 104.

Students who received 504 Plan supports in HS may continue to need such supports when they transition to college or university and are encouraged to contact the center for students with disabilities at their postsecondary school to explore these options and arrange for the necessary supports. Psychiatrists working with students in post-HS educational programs should be aware that their patients may be eligible for continued 504 plan supports.

Table 54.8–5.
Accommodations and Modifications

Categories	Application
Accommodations	
Presentation	▸ Allow students to audiotape lessons. ▸ Allow student to use alternatives to test books (including audio records). ▸ Reduce amount of work per page. ▸ Enlarge print size. ▸ Provide adjusted paper for writing and math (e.g., wide-lined notebook paper; enlarged graph paper). ▸ Use multimodal communication especially for instructions (e.g., pairing auditory and visual information). ▸ Read all instructions aloud to student. ▸ Frequent checks for understanding.
Response	▸ Allow alternative response form (i.e., orally vs. in writing). ▸ Only call on student when student raises their hand. ▸ Eliminate participation grade/use alternative means for grading participation. ▸ Use of a scribe for written responses. ▸ Extended time allowances for responses. ▸ Opportunities to complete work or take tests in a separate location. ▸ Allow student to use a dictionary and/or grade spelling separately. ▸ Do not downgrade for handwriting. ▸ Allowances to use a calculator and/or printed math facts sheet (e.g., multiplication table).
Timing/Scheduling	▸ Extended time allowances for in-class work, homework, projects, and/or tests. ▸ Opportunities to take breaks, including movement breaks. ▸ Extended time allowances during passing periods.

	▸ Allowances to take tests over multiple days/periods.
Setting	▸ Allowances to take tests in a separate location (i.e., quieter setting). ▸ Preferential seating to maximize learning (e.g., closest to the source of instruction; at back of room to allow for movement). ▸ Being provided with special lighting or acoustics. ▸ Nonpunitive use of sensory tools (e.g., chair band, seat cushion, foot rest).
Modifications	
Assignment	▸ Spelling tests using alternative words to be tested on. ▸ Reduce assignments in amount or modify content significantly. ▸ Provide alternative material or assessments that differ from standard grade-level type or amount.
Curriculum	▸ Teach adjusted content and use material that differs from standard level. ▸ Adjust grading requirements and standards as expected of same-grade peers. ▸ Provide alternative projects with adjusted grade-level materials.

Individualized Education Plan

If a student needs more specialized services and programming than what can be provided via a 504 Plan, then a student may be involved in the IEP process (see Table 54.8–6 for comparisons between 504 Plans and IEPs). The IEP process can begin following either of the two “lower-level” supports mentioned above (i.e., MTSS/RTI or 504 Plan supports) or the IEP process can be initiated without the student having received tiered or 504 supports.

The IEP refers to the individualized program that is developed only after a student has been approved to be assessed, has completed the IEP assessment, has been determined to meet eligibility for receipt of special education services under the designated eligibility categories by the IEP team, and the IEP team has agreed upon the recommended goals, services, and programming. In these regards, the IEP itself is the “end product” that is the result of a specific, standardized, and structured process.

Table 54.8–6.
Comparison Between 504 Plans and IEPs

Description	504 Plan	IEP
Who qualifies?	1. Has a physical or mental impairment 2. Impairment limits at least one major life activity	1. Meets eligibility criteria for at least one of the Special Education categories 2. Exhibits difficulties that have existed over an extended period of time and, to a marked degree, adversely affect a student’s educational functioning
Who typically implements services?	General education staff	Specialized staff (e.g., special education teacher, school psychologist, speech therapist, etc.)
Does school receive funding to implement recommended changes?	No funding	Federal/state/local funding for services
Are protections available postsecondary school?	Yes	No

An important aspect of the IEP process is the *timeline*. Although federal law states that assessments must be conducted within 60 days of receipt of parental/guardian consent for the evaluation (i.e., signed assessment plan form), *the exact timelines can vary from state to state*. See Figure 54.8–2 for the steps involved in the IEP process.

IEP Assessments. A psychiatrist should be aware of the types of tools and information included as part of the IEP assessment: review of records, administration of any indicated testing based on the suspected area(s) of disability, observations of the student, and/or consulting with outside providers, such as the psychiatrist. Most initial IEP assessments include formal, standardized tests, especially when the suspected area of disability relates to a student's academic achievement, speech and/or language functioning, processing, and/or social, emotional, and behavioral functioning. Each school and/or district may administer and use different tests, rating scales, or assessment tools. Table 54.8–7 describes some examples of the categories and types of assessments commonly used, though not the only tests used, for IEP assessments.

General Components of an IEP Meeting. At an initial (or triennial) IEP meeting, two main areas are covered: (1) review of current assessments to discuss "present levels of performance (PLOP)" and (2) determination of eligibility.

IEP Meetings Requirements:

1. Must be held annually, at minimum, to collect data, update goals, and make any changes.
2. Can be reconvened at any time in response to parent request or team agreement.
3. Must be convened when there is a request for a change in services/supports (either an increase, decrease, or discontinuation of services).
4. Must be convened if a student with an IEP is being considered for disciplinary action (i.e., “manifestation determination hearing”).
5. Must be convened when a student transfers to a new school/school district in order to accurately document the change in location and provide any appropriate updates based on the services and programming at the new school.

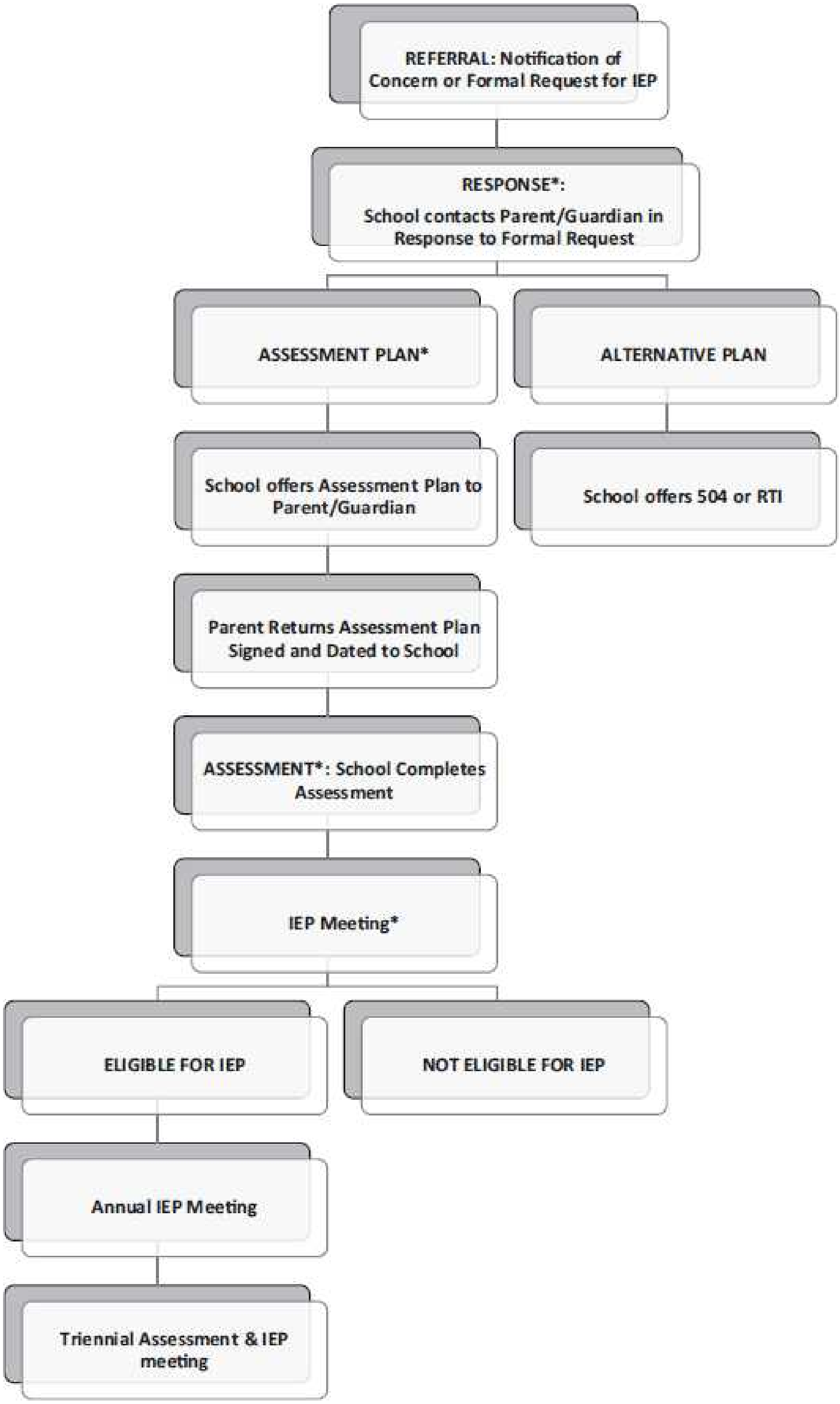
REFERRAL: Notification of Concern or Formal Request for IEP
RESPONSE*:
School contacts Parent/Guardian in Response to Formal Request
ASSESSMENT PLAN*
ALTERNATIVE PLAN
School offers Assessment Plan to Parent/Guardian
School offers 504 or RTI
Parent Returns Assessment Plan Signed and Dated to School
ASSESSMENT*: School Completes Assessment
IEP Meeting*
ELIGIBLE FOR IEP
NOT ELIGIBLE FOR IEP
Annual IEP Meeting
Triennial Assessment & IEP meeting

FIGURE 54.8–2. Steps in the initial IEP process.

If the student is found to meet eligibility criteria, the IEP team meeting will then develop recommendations for the special education programs and services that the student needs; this is known as the team's offer of Free and Appropriate Public Education (FAPE). The IEP team can determine (1) goals and objectives based on individual needs; (2) recommended services and programming and/or placement based on the goals; and (3) accommodations, modifications, and/or supports based on the recommended placement. The IEP team will also decide on the percentage of time that the student is to receive instruction within and outside of the general education setting.

Table 54.8–7.
Examples of Commonly Used Assessment Tools by Category

Category	Instrument
Intellectual or Cognitive Tests	
Cognitive Tests for Children <6 years old	
	Bayley Scales of Infant and Toddler Development-Third Edition (Bayley-III)
	Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition (WPPSI-IV)
Cognitive Tests for Children >5 years old	
	Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V) (6–16 years old)
	Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV)
	Stanford-Binet Intelligence Scale-Fourth Edition
	Woodcock-Johnson IV Test of Cognitive Abilities
	Differential Ability Scales-II (DAS-II)
Educational Achievement Tests	
General Achievement Tests	
	Wide Range Achievement Test-5th Edition (WRAT-5)
	Woodcock-Johnson Test of Achievement-Fourth Edition (W-JIV)
	Wechsler Individual Achievement Test-Third Edition (WIAT-III)
	Kaufman Test of Educational Achievement-Third Edition (KTEA-III)
Tests of Specific Areas of Achievement	
	Gray Oral Reading Test-Fifth Edition (GORT-5)
	Test of Written Language 4 (TOWL 4)
	Woodcock Reading Mastery Tests-Revised-Normative Update (WRMT-R/NU)
	Dynamic Indicators of Basic Early Literacy Skills (DIBELS)
	KeyMath 3 Diagnostic Assessment

Speech and Language Tests

Comprehensive Speech/Language Tests

- Clinical Evaluation of Language Fundamentals-Fifth Edition (CELF-5)
- Peabody Picture Vocabulary Test-Revised (PPVT-3)
- Expressive One-Word Picture Vocabulary Test (EOWPVT)
- Goldman-Fristoe Test of Articulation
- Test of Language Competence (TLC)
- Test of Problem Solving (TOPS)
- Test of Language Development (TOLD-3)

Auditory Processing Screening Assessments

- Test of Auditory Processing (TAPS) (often given by psychologists)
- SCAN-3:A Tests of Auditory Processing Disorders in Adolescents and Adults

Processing Tests

- Beery–Buktenica Developmental Test of Visual-Motor Integration (VMI) (includes VMI, Visual Perceptual, and Motor Coordination subtests)
- Bender Visual Motor Gestalt Test
- Test of Visual Perceptual Skills (nonmotor)-Third Edition (TVPS-3)

Adaptive Functioning

- Vineland Adaptive Behavior Scales (VBAS)

Social/Emotional Measures/Rating Scales

- Behavior Assessment System for Children (BASC-2)
- Child Behavior Checklist (CBCL)
- Children’s Depression Inventory (CDI)
- Beck Depression Inventory (BDI)
- Screen for Child Anxiety Related Disorders (SCARED)

	UCLA PTSD Reaction Index for DSM-5
Behavior Rating Scales	Conners’ Rating Scales Behavior Rating Inventory of Executive Function (BRIEF)

If the student is found to *not meet eligibility criteria* for any of the special education categories, then the IEP team may discuss alternative options (i.e., tiered interventions; development of a 504 Plan). If a parent disagrees with the IEP team’s recommendations, then the parent can discuss that during the meeting. If a resolution cannot be achieved at the meeting, then the parent can take steps to pursue other options (e.g., mediation, complaint, due process hearing).

Parents and guardians are considered to be key members of the IEP team. At or prior to the IEP meeting and in writing, the school is required to inform parents and guardians of their rights under IDEA and state law. One key right is the right to active participation in the IEP meeting and process. Other key rights of parents and guardians include, but are not limited to, the following (categorized as procedural safeguards):

Table 54.8–8.
IEP Eligibility Categories

Eligible Category	**Description**	**Examples of Disorders/Conditions**
Specific Learning Disability (SLD)	Disorder in one or more of the basic psychological processes involved in understanding or in using language (ability to listen, think, speak, read, write, spell, or do mathematical calculations)	▸ Dyslexia ▸ Dyscalculia ▸ Dysgraphia ▸ Brain injury ▸ Developmental aphasia ▸ Perceptual disabilities
Emotional Disturbance (ED)/Emotional Disability (ED)	One or more of the following characteristics over a long period of time and to a marked degree that adversely affects one's educational performance: A. An inability to learn that cannot be explained by intellectual, sensory, or health factors. B. An inability to build or maintain satisfactory interpersonal relationships with peers and teachers. C. Inappropriate types of behavior or feelings under	▸ Affective/mood disorders ▸ Anxiety disorders ▸ Psychotic disorders ▸ Certain impulse control disorders (including trichotillomania)

	normal circumstances. D. A general pervasive mood of unhappiness or depression. E. A tendency to develop physical symptoms or fears associated with personal or school problems.	
Other Health Impairment (OHI)	Having limited strength, vitality, or alertness, including a heightened alertness to environmental stimuli	▶ Attention deficit hyperactivity disorder ▶ Tourette syndrome ▶ Asthma ▶ Diabetes ▶ Epilepsy ▶ Sickle cell anemia
Autism (AUT)	Exhibits diagnostic impairments consistent with the diagnosis of AUT. Educational performance affected by engagement in repetitive activities and stereotyped movements, resistance to environmental change or change in daily routines, and unusual responses to sensory experiences.	▶ Autism spectrum disorder
Speech/Language	Meets one or more of	▶ Apraxia of speech

Impairment (SLI)	the following criteria: A. Articulation disorder. B. Abnormal voice C. Fluency disorders D. Language disorder	▶ Dysarthria ▶ Aphasia ▶ Stuttering ▶ Phonologic disorder ▶ Muteness ▶ Selective mutism
Intellectual Disability (ID)	Subaverage general intellectual functioning, existing concurrently with deficits in adaptive behavior and manifested during the developmental period	Mild, IQ = 50–70 Moderate, IQ = 35–49 Severe, IQ = 20–34 Profound, IQ < 20
Multiple Disabilities	Combination of disorders causing severe educational needs that they cannot be accommodated in special education programs solely for one of the impairments	Combination of conditions
Deaf–Blindness	Requires severe communication and other developmental and educational needs	
Deafness/Hard of Hearing (DHH)	Severe impairment in processing linguistic information through hearing	
Hearing Impairment	Impairment in hearing (permanent or fluctuating) adversely affects educational	

	performance, but that is not included under the definition of deafness	
Orthopedic Impairment (OI)	Severe OIs that include impairments caused by a congenital anomaly and impairments from other causes	▶ Cerebral palsy ▶ Amputations ▶ Fractures
Traumatic Brain Injury (TBI)	An acquired injury to the brain resulting in total or partial functional disability or psychosocial impairment, or both	
Visual Impairment (VI)	Impairment in vision that, even with correction, adversely affects a child's educational performance	▶ Partial sight ▶ Blindness

1. Giving informed written consent for their child to be evaluated and to seek an independent evaluation if they feel compelled to do so.
2. Access to their child's educational records.
3. Dispute resolution through steps that can include state complaint, mediation, a resolution session, and/or a due process hearing.
4. Explanation of these rights in writing, and in their native language when possible.

Eligibility categories. There are usually 13 special education eligibility categories, with some districts using slightly different terms or titles, but in general, the categories are listed in Table 54.8–8 below. Siperstein et al. (2011) reported that students who qualify for emotional disturbance (ED) represent one of the main categories of youth with psychiatric illnesses, a group who has been shown to struggle academically over time, despite receiving services.

Key Parts of an IEP. Although all districts are legally required to include certain sections in their IEPs, all districts have their own IEP format. Therefore, IEPs from different school districts vary in how they are organized. Below is a list of some of the key sections found in most, if not all, IEPs and those that may be useful for psychiatrists to review.

1. *Identifying information.* This can include school of attendance, school of residence (if different), pertinent dates such as the date of the current IEP, date of initial entry into special education, and date(s) of subsequent annual meetings and estimated assessments (i.e., triennial assessment).
2. *PLOP.* A description of the student's current level of functioning and/or test results is included.
3. *Goals*. A list of goals with (a) target areas to be addressed by special education services, based on eligibility and identified areas of need, and (b) the person responsible for addressing the intervention (e.g., special or general education teacher, specialist, counselor).
4. *Eligibility/disability category*. Some districts will indicate a "P" for a primary handicapping condition and "S" for a secondary handicapping condition. Some districts only allow one eligibility. See Table 54.8–8 for eligibility categories.
5. *Recommended services* typically include *instructional settings* such as general education classroom, RSP, Special Day Class (SDC), Specially Designed Instruction (SDI), Specialized Academic Instruction (SAI); "push in" (inclusive education with a specialist that assists the student in a general education setting) and "pull out" (student pulled out of the general education class to receive specialized services). The *location of services* can also vary (on a general education campus or specialized campus such as a nonpublic school).
6. *Additional services and related services* include speech and language therapy, occupational therapy, adaptive physical education, assistive technology, and transportation. *Behavioral Intervention Services* can begin with a functional behavioral assessment (FBA) to develop a positive behavioral intervention plan (strategies and supports to address the target behavior). Also potentially included is *Extended*

School Year (ESY) Services, specialized instruction or related services during extended breaks (such as during summer breaks).

7. *Percentage of time* is also recorded, which designates the percent of time within and outside the general education setting. School districts are required to provide a FAPE within the LRE. Therefore, the IEP team may initially recommend what appears to be a “minimal” amount or lower level of intensity of services before trying more intensive services. However, there is no mandate that an IEP team must start with a lower level of services before considering or offering more intensive services.
8. *Other sections***.** Statewide assessments includes a description of any adjustments to statewide testing that the student may receive.

Craig is now a 9-year-old and in third grade. He has continued to struggle in school despite his 504 Plan accommodations he had in second grade. Last year, he continued to be in and out of his seat, knocking items off his desk, and disrupting the whole class, especially when he would talk out of turn. He continues to have challenges with following instructions, and at home, he struggles with sitting down and completing his homework, even with fewer items per page. As a result, he lags behind in reading and writing, but he is doing better at recess with a set of his own preselected activities and awareness to speak to a yard teacher when upset.

Given his continued disruptions in class and parental concerns regarding his academic achievement, his parents submit a letter to his principal, requesting a formal assessment for an IEP. The IEP team met following the school’s assessment and determined that he meets a primary eligibility of other health impairment (OHI) due to his ADHD and secondary eligibility of specific learning disability (SLD) in reading. The IEP team recommended he receive SAI as a pullout in a special education class for 300 minutes per week, adjustments to reading materials across subjects (such as allowances for math word problems to be read to him), and continuing those accommodations listed in his 504 Plan (which were incorporated into his IEP).

Private Schools and “Special Education”

Some parents choose to enroll their child in a private school, called a Parentally Placed Private (PPP) school enrollment. Private schools are not required to provide “FAPE,” special education services, or an IEP. However, no school can discriminate against a child with a disability for reasons related to that disability. As such, all schools are responsible to provide supports to ensure that a child with a disability can have equal access to educational opportunities at their campus (e.g., providing ramps for students in wheelchairs). In addition, if they choose, some private schools may develop and work with the LEA to develop a plan for specialized services within the private school. This is known as an Instructional Services Plan or an Individual Services Plan (ISP), which tends to be less comprehensive than an IEP. For example, private schools often agree to give accommodations (e.g., extra time on tests, preferential seating). If a student who is attending a PPP school qualifies for special education services through the IEP process, the services offered are very likely to be offered only at the local public school.

ROLE OF PSYCHIATRISTS IN SCHOOLS

Having a thorough understanding of the school system, the personnel, and resources potentially available to students is an important first step for psychiatrists working with schools. This next section will describe several ways that psychiatrists can apply this knowledge in working with the school system. A *psychiatrist in practice* may play a collaborative role, working with the school as part of a patient’s treatment plan, such as coordinating medication treatment with a school nurse, supporting a teacher with specific behavioral strategies, or as an advocate in supporting a student in accessing needed special education services. *A school psychiatrist* may be an employee of the educational system or of an outside organization such as a community agency or university, who evaluates and treats students referred by the school system such as through a SBHC. A *school consultation* model involves a psychiatrist who serves as a consultant to a classroom, school, or district, and who may provide staff- and administrator-level recommendations for system improvement. Finally, a psychiatrist may be called upon to assist a school during a time of crisis, such as during a disaster or sudden death.

The Role of a Psychiatrist in Coordinating Care With Schools

One important role that a psychiatrist may play with a school system is in collaboration as part of patient care. For example, a student's outpatient psychiatrist may work directly with school staff (following parent or guardian consent) to discuss the impact of their diagnosis on daily functioning including school functioning, help establish and clarify expectations based on a student's PLOP, obtain information from the school to inform treatment plans, and promote effective communication among parent/guardians, school staff, and the clinician.

When working with schools, it is important to understand that schools follow a federal law that regulates how data can be disclosed, the Family Educational Rights and Privacy Act of 1974, (FERPA). Limits of confidentiality must be considered within both the health care system (the Health Insurance Portability and Accountability Act [HIPAA]) and educational system (FERPA). See Table 54.8–9 for comparison between HIPAA and FERPA. After obtaining consent and discussing confidentiality parameters, information such as patient diagnoses, assessment results, pertinent information from clinical interventions, and how medications may impact student school functioning can be shared between school personnel and clinicians.

Psychiatrists may communicate with the school to set up school observations to aid in clinical data collection for their client's treatment plan or provide documentation to the school delineating patient diagnoses. Bidirectional communication is crucial in ensuring positive outcomes for students. Identifying the key school personnel who can support a student in the school setting can be a vital part of the collaborative process and treatment plan. The following is a list of important considerations for the psychiatrist when collaborating with school staff:

- When possible and indicated, convey your understanding that the school personnel are the "experts" in that setting.
- Direct communication (i.e., over the phone, virtual or in-person visit, email) is often an ideal way to maximize exchange of specific ideas and information.
- Be aware that teachers may likely have notable time limitations to meet during the school day, so be gracious and aware of this time limitation.
- Note that sometimes the teachers have already had difficult interactions with the student and/or parent. Thus, be mindful of

entering into potentially tense dynamics or conflicts.

- Ideally, one goal from this communication is to *join* with the teacher (two professionals working collaboratively) and to *clarify* (i.e., get specific information about how the student is doing, what challenges the teacher sees as primary difficulties, what the teacher perceives are contributing to that challenge).
- Ask specific yet *open-ended* questions as much as possible (minimize general yes/no or good/bad questions). Examples of key questions to consider:
    - How is the student doing with independent work?
    - What are the student's social interactions like?
    - How often does the student ask for help?
    - How does the student respond to redirection?
    - How does the student do during less structured times?
    - What has been tried that has helped? (and, if so, how long has it helped?)
    - What has been tried that has not helped?
    - How does the student do with homework, classwork, and tests?
    - Does the student need the information to be repeated?
    - Does the student seem to have difficulty adequately explaining their answers?
    - Does the student seem to give up and say things like, "It doesn't matter"?
    - Are there any stressors (including peer relationships, family, housing, food insecurity, racial stress, immigration, economic, trauma, or loss) you are aware of that may be impacting the student?
    - What are the student's relationships or connections like with staff at the school?
- As informed, psychologically minded, and experienced school personnel might be, they are not medical professionals. A psychiatrist needs to describe *how the patient fits a specific diagnosis* rather than solely stating the diagnosis, and abbreviations should be avoided in verbal and written communication.

Table 54.8–9.
Descriptions of HIPPA and FERPA

	HIPPA	FERPA
Law	Health Insurance Portability and Accountability Act of 1996	Family Educational Rights and Privacy Act of 1974
Source of the Law	U.S. Department of Health and Human Services	U.S. Department of Education
Federal or State Law	Federal	Federal
What Entities Does It Cover	Businesses that focus on health care information (health care providers, health plans, and health care data clearinghouses)	Schools and postsecondary institutions (public and private schools)
What Information Does It Cover	Personally identifiable information about a patient stored by an organization (except that stored by a school)	All information in a student's records (student's grades, academic report cards, transcripts, class schedules, disciplinary records, personal and contact information, family information, and medical care given to a student at school).
Who Does It Cover	All patients of an organization.	Covers both current and previous students of the school.
What it Requires	Requires that covered entities apply appropriate administrative,	Schools give legal guardians control over a student's information until the

	technical, and physical safeguards to protect the privacy of individually identifiable health information (protected health information or PHI), in any form. The only individuals with access should be the patient, those providing care, and people the patient allows to access their PHI, such as family members.	student is 18 years old or starts their postsecondary education. These rights transfer to the student when the student reaches the age of 18 or attends a school beyond the high school level.

Communicating Through a School Letter. A psychiatrist may need to write a letter to their patient's school to best transmit critical psychiatric information, especially to convey recommendations for the school setting. Such letters can also ensure that pertinent data needed to substantiate a recommendation for special education services are provided, without unnecessarily compromising patient or family privacy by overincluding information not immediately relevant to schools. Directives from a nonschool professional for inclusion of a student based on a particular special education category or placement in a specific school may not be seen favorably by school personnel; thus, it is more effective to describe how a child's behavior is consistent with specific special education eligibility criteria and how the student's functional school needs are consistent with the types of services and/or programming required and recommended. In addition, the letter should highlight any prior history of difficulties in school and prior school interventions, including if the patient has received formal special education services through an IEP, supports under a 504 Plan, and/or informal interventions within a general education program. The letter should summarize any history of suspected or identified learning issues, any interventions provided within the school or at home, effect of interventions, and attendance problems. A letter that documents the student's level of physical activity and any medical conditions is helpful, as this information is required for the IEP. Provide specific information regarding the patient's RTIs, specifying any current problems or stressors, especially those that may affect the patient's functioning at school. Provide specific examples.

When describing the clinical diagnoses and medications, avoid using abbreviations. It may be helpful to acknowledge previous diagnoses which have been given and to explain the rationale for the current diagnoses, as previous IEPs may include information that is inconsistent with the current clinical assessment. Also indicate if mediation will need to be administered at school or not. Finally, for treatment and educational recommendations, specify recommendations for patient's further or continued psychiatric treatment. A description of the patient's behaviors as consistent with special education eligibility criteria (see Eligibility Categories for specific descriptors) can be helpful. Try to describe the child's current or predicted school functioning in ways that school professionals can recognize (1) whether the child can function all or most of the time in a general education classroom with support throughout the week; (2) whether the child will require intensive specialized instruction and teacher supervision throughout most of the day; or (3) whether it appears that the child will require sustained special instruction and therapeutic support both in the classroom and through the entire school day in a setting where trained personnel are available to deal with crises.

Direct Service Models

Direct service models can involve a psychiatrist providing clinical care for students in the school system or in partnership between community organizations and schools to support an array of mental health services in the school or linked to a community agency. SBHCs and comprehensive school mental health systems (CSMHSs) are two models that reflect such models of direct service, providing mental health care to students via a collaboration with community providers. Utilizing telemental health is another avenue by which psychiatric consultation services can be delivered in school settings.

Many schools have mental health supports less comprehensive than in SBHCs or CSMHPs yet provide important components of student mental health. Examples include early identification screening programs, web-based mental health resources, and on-site or video consultations with off-site mental health clinicians. The model adopted by a school often depends on resources, state law, and student and school needs and priorities.

School-Based Health Centers. SBHCs can improve health equity, often serving underresourced communities to improve access to primary care, mental health, dental care, and other services, with the ultimate goal of supporting health, wellness, and learning. Often located on or near a school campus, the SBHCs help students and families address access to care barriers such as transportation, cost, and parent time off from work. As of 2016 to 2017, the School-Based Health Alliance reported that there were 2,584 SBHCS in the United States, a number that has doubled over the past two decades, according to Love et al. SBHCs integrate into schools, building an interdisciplinary team of providers who deliver prevention and integrated care that effectively meets the complex medical, behavioral, and social health needs of children and adolescents. SBHCs are found across elementary, middle, and HS settings. Research shows that SBHCs significantly improve children and adolescents' access to and utilization of medical, mental health, and substance use services, particularly among underresourced populations.

Comprehensive School Mental Health Programs. CSMHPs provide a full array of tiered services, including universal mental health promotion activities, selective prevention, and indicated early intervention services such as clinical assessment and treatment. CSMHPs rely on collaborative partnerships between school systems and community programs so that school-employed mental health professionals are meaningfully supported by community mental health programs. Increasingly, community mental health staff partner with school-employed mental health staff and educators to provide this more comprehensive integrated model. From a national sample of 1,064 school districts, around 50% reported contractual/formal agreements with community-based agencies to augment service provision. Community mental health program partners may include community mental health centers, hospitals, and universities. The shared goals of these partnerships emphasize improving student behavior, addressing barriers to learning, increasing access to mental health care, and improving student attendance and academic achievement as reported by Hoover et al. (2021).

Telemental Health in Schools. Telemental health for children and adolescents is widely accepted as an effective delivery model of mental health services in a variety of settings, including schools. Improving technology and increased exposure to telehealth allows students, teachers, and administrators greater access to psychiatric consultation. Medication management, evaluation for support services, assistance with diagnostic clarification, and integrative collaboration with school staff can all be completed by utilizing remote services. Several school telemental health programs demonstrated increased efficiency, greater capacity for volume, and increased access to care for children in underserved urban and rural community settings. Several issues to consider include privacy, access to appropriate technology, and the ability to engage students and families. Preliminary studies by Mayworm et al. show the promise of telemental health as a psychiatric consultation delivery model.

School Consultation

One role that a psychiatrist can have with schools is as a school consultant, in response to an identified need or question from the school system (i.e., initiated from a teacher, principal, district leader, etc.). Although the psychiatrist brings a valuable wealth of expertise about mental health, child development, and the interface of the social determinants of health, the consultant should address the school question with an approach of mutual "joining" with the system and school community.

Bostic and Bagnell describe a School Consultation Framework with five components to consider when serving as a consultant in the school:

1. *Who is the consultee, and what are the limits of confidentiality?* The consultant must establish procedures to clarify who is appropriate to contact or to evaluate. The consultant may be asked by the school administrator to observe a classroom and consult to staff about questions on how best to work with particular students. When the consultant is asked to meet directly with a student, *the consultant must obtain parental permission before interviewing any student.* Clarifying confidentiality parameters should occur with staff, students, and parents. This includes delineating how and where information is shared (i.e., if clinical information is written and placed in student or personnel records).

2. *What is the consultation question, and what are the expectations of the consultee?*
   Consultants help others define problems and furnish solutions. Though a consultant may ask for specific interventions for a student, the consultee may actually expect or wish a consultant to speak with administrators or recommend removing the student from the consultee’s classroom. Consultants should focus on the objective of the question to clarify both the consultee’s explicit request and implicit desires. Defining the consultation questions and expectations in collaboration with the consultee is crucial for the success of the consultation, so both agree on the problem and realistic goals for any intervention.
3. *How is the larger system experiencing this problem?*
   A problem will affect school staff members and administrators differently, and each individual may have different expectations for the consultation. The consultant should identify the key stakeholders in a consultation question and answer, including who requested the consultant, who knows about the consultation, who consented to it, and who will be informed about the consultant’s findings. Clearly communicating and defining the goals and expectations of the various individuals and entities may increase buy-in of the proposed solution.
4. *What legal or ethical factors should be considered?*
   A variety of legal and ethical factors may impact the identified consultation and question and proposed solutions. Specific laws governing the education system may yield unique opportunities for students, but families and schools may not consider their applicability in a particular case. If the consultation question involves a specific student, especially if a direct interview or evaluation will be completed by a psychiatrist, written consent signed by a parent or guardian is generally required outlining permission for the release of medical information to the school. Parents may decline the psychiatric consultation offered by the school or deny consent for a psychiatrist to consult with school staff about their child. Structural racism, fears about immigration status, concerns about child protective services involvement, and stigma about receiving psychiatric services are all potential barriers to parents obtaining psychiatric treatment. Exploring and understanding parental concerns and fears, stigma, and expectations; providing psychoeducation; and discussing confidentiality limits may promote acceptance of the consultation. However, if parents decline direct contact between a consultant and

their child, a consultant may still address teacher questions without seeing the child. Parents cannot prevent a clinician from observing a student and consulting directly with school staff.

In cases of suspected abuse, the consultant may help guide or assist staff in expressing concerns to an appropriate agency. In acutely dangerous situations such as those in which suicide or homicide arise, the consultant may aid the consultee in facilitating emergent treatment or warning those at risk of being harmed. With any proposed intervention, legal and ethical ramifications must be considered. To increase the likelihood of progress in a difficult situation, consultants should promote clear and collaborative efforts of the student, family, and school.

5. *How does the consultant understand this problem from a biopsychosocial perspective?*

   Using the biopsychosocial model to conceptualize a consultation question can help to expand potential solutions in the school setting. For example, when asked to evaluate a student's "problem behaviors," a psychiatrist needs to consider the biologic and psychological risk and protective factors, ACEs, the family and community's culture and beliefs, peers and school environment and other social stressors, and social and structural determinants of health. Thinking about these various influences may provide a variety of targets for intervening in a particular problem, including at the classroom and teacher level, with the family, in addition to intervening with the student. Given racial and ethnic disparities in the education system, as discussed earlier, the consultant's biopsychosocial perspective is essential.

Consulting to Schools in a Time of Crisis

Whether responding after a student's sudden death by suicide, preventing violence on campus, or supporting a school during a broader community crisis such as the COVID-19 pandemic or the wildfires in the western states, a psychiatrist's consultation can be valuable in supporting the school community before, during, and after recovery from a crisis. Crises involving a school can occur unexpectedly and suddenly, disrupt the entire school community, and interfere with learning. For over 40 years, crisis teams at the regional and school level have disseminated preparedness and crisis management strategies. Funding from the U.S. Department of Education through such mechanisms as Project SERVE (School

Emergency Response to Violence) and Readiness and Emergency Management for K-12 Schools (REMS) grants have supported schools and districts in response to school crises. As mentioned earlier, schoolwide prevention efforts, such as violence prevention programs and early referrals to needed mental health services through schools, can help mitigate the effects of school crises (see Table 54.8–4 for examples of evidence-based school prevention programs). Vona and colleagues described a multitiered system of a support framework for crisis intervention, with strategies before, during, and after a crisis involving a school.

Responding After a Death by Suicide. Given that death by suicide is a major cause of death for young people, schools can play a critical role in disseminating universal suicide prevention strategies. One example of an effective intervention is the Signs of Suicide (SOS) program, which teaches students about the warning SOS risk and how to seek assistance from trained school staff. Promising results of the SOS program indicate improvement in students' attitudes about intervening with peers and themselves when suicidal intent is present and a decreased likelihood of attempting suicide for those receiving the intervention.

Although rare, one of the most devastating incidents a school community can experience is the death by suicide of a student or staff person. In preparation, schools can develop preplanned, postvention procedures that would be implemented following a death by suicide. The main purpose of postvention is to disseminate resources for those emotionally impacted, identify those who may need more intensive services, provide coping strategies, and facilitate a return to the routine of school. Prevention is particularly geared toward those who may be at risk for suicide themselves and in preventing a contagion effect.

Responding to a Threat of Violence. Although mass school shootings such as those at Sandy Hook Elementary School or at Stoneman Douglas High School are rare events, psychiatrists may be called upon to conduct a threat assessment of the potential "risk" that a student may act violently toward others. Weisbrot outlines several key points to understand regarding threat assessment, such as (1) there are often warning signs and others who were made aware of plans of violence, (2) "pushing out" students through out-of-school discipline may not eliminate the threat to a school and can lead to disengagement of a student who may be in need of supports, and (3) social media posts and writings that threaten harm and violence need to be taken seriously and assessed.

One threat assessment protocol for schools, the Virginia Student Threat Assessment Guidelines (VSTAG), has been found to decrease rates of peer aggression, improve school climate, and decrease out-of-school suspensions detailed by Cornell et al. The VSTAG is a structured assessment technique that rates 19 risk factors for violence as low, medium, or high. When evaluating schools implementing threat assessment techniques, Cornell found that students at “high” risk by formal threat assessment were 12.5 times more likely to attempt violence than those categorized as not at high risk. Males, those in upper ES, those receiving special education services, and disproportionately Black students were more likely to be *referred* for a threat assessment. However, when the structured threat assessment was used, the authors report that no racial or ethnic disparities in out-of-school suspensions nor in determinations of threats deemed “serious” were found, thus demonstrating racial bias in referral. Threats that were more likely to be rated as serious included those involving battery, threats to kill, those involving weapons, and those targeting school administrators.

Responding to Disasters and Other Community-Wide Trauma. One final example of a key consulting role that psychiatrists can have with schools is in supporting efforts by a school or district during and following a natural or man-made disaster or a major catastrophic public health situation such as the COVID-19 pandemic. Community-wide traumatic events such as these can impact the education, health, and well-being of students but often challenge the school staff and administrators in similar ways. In particular, these types of events can reverberate across multiple social determinants of health, exacerbate existing inequities, and disproportionately affect underresourced communities of color. Ijadi-Maghsoodi and colleagues described a unique public–public partnership between a public university and a large, urban county system that had begun creating a countywide infrastructure for web-based trauma and resilience-informed resources prior to the pandemic. This partnership, including psychiatrists and other mental health professionals, collaborated with systems of care such as local school communities in response to the impact of the pandemic, including creating online training for teachers and other school personnel addressing teacher well-being, supporting students with anxiety, psychological first aid for teachers, and the effects of racism in schools (see learn.wellbeing4la.org/ website for details).

In summary, there are many ways that psychiatrists can interface with schools, reinforcing the educational mission of schools and supporting the development of the whole child, starting with health promotion and prevention. Table 54.8–10 includes a list of website resources that may be useful for psychiatrists as they provide school consultation, school mental health services, and coordinating care with schools for their patients.

Table 54.8–10.
Website Resources

Category	Description	Site
Blueprints	Registry of evidence-based interventions for healthy youth development	blueprintsprograms.org/
National Center for School Mental Health	National training, practice, research, and policy information	csmh.umaryland.edu
School Threat Assessment	Trainings and resources for threat assessment in schools	schoolta.com/
Social-Emotional Learning	Research, practice, and policy regarding social–emotional learning	casel.org
What Works Clearinghouse	Reviews the existing research on different *programs, products, practices,* and *policies* in education	ies.ed.gov/ncee/wwc/
Wrightslaw	Special education law and advocacy	wrightslaw.com

ACKNOWLEDGMENTS

The authors would like to acknowledge the previous authors, Jeff Q. Bostic, Sharon Hoover Stephan, and Alexa Bagnell, and we would like to thank Kungeun Lee and Steven Forness for their assistance with this manuscript.

Further Readings

Addis S, Greer K, Dunlap L. Effective strategies for alternative school improvement: a practice guide by the national dropout prevention center. *National Dropout Prevention Center*. Successful Practices Network; 2020.

Allison MA, Attisha E; Council on School Health. The link between school attendance and good health. *Pediatrics*. 2019;143:e20183648.

Bostic JQ, Bagnell A. School psychiatric consultation: an organizing framework and empowering techniques. school consultation/intervention. *Child Adolesc Psychiatr Clin N Am*. 2001;10:11–12.

Carter PL, Skiba R, Arredondo MI, Pollock M. You can't fix what you don't look at: acknowledging race in addressing racial discipline disparities. *Urban Educ*. 2017;52:207–235.

Cornell D, Maeng JL, Burnette AG, et al. Student threat assessment as a standard school safety practice: results from a statewide implementation study. *Sch Psychol Q*. 2018;33:213–222.

Daly BP, Sander MA, Nicholls EG, Medhanie A, Berk EV, Johnson J. Three-year longitudinal study of school behavior and academic outcomes: results from a comprehensive expanded school mental health program. *Adv Sch Ment Health Promot*. 2014;7:24–41.

Flaherty LT, Osher D. History of school-based mental health services in the United States. In: Weist MD, Evans SW, Lever NA, eds. *Handbook of School Mental Health: Advancing Practice and Research*. Springer; 2003:11–22.

Harrison JR, Bunford N, Evans SW, Owens JS. Educational accommodations for students with behavioral challenges: a systematic review of the literature. *Rev Educ Res*. 2013;83:551–597.

Henry KL, Knight KE, Thornberry TP. School disengagement as a predictor of dropout, delinquency, and problem substance use during adolescence and early adulthood. *J Youth Adolesc*. 2012;41:156–166.

Hoover S, Bostic J. Schools as a vital component of the child and adolescent mental health system. *Psychiatr Serv*. 2021;72:37–48.

Ijadi-Maghsoodi R, Harrison D, Kelman A, et al. Leveraging a public-public partnership in Los Angeles County to address COVID-19 for children, youth, and families in underresourced communities. *Psychol Trauma*. 2020;12:457–460.

Ijadi-Maghsoodi R, Marlotte L, Garcia E, et al. Adapting and implementing a school-based resilience-building curriculum among low-income racial and ethnic minority students. *Contemp Sch Psychol*. 2017;3:223–239.

Institute of Medicine and National Research Council. *Transforming the Workforce for Children Birth Through Age 8: A Unifying Foundation*. National Academies Press; 2015.

Joshi SV, Hartley SN, Kessler M, Barstead M. School-based suicide prevention: process, and role of trusted adults and peers. *Child Adolesc Psychiatr Clin N Am*. 2015;24:353–370.

Kataoka SH, Vona P, Acuna A, et al. Applying a trauma informed school systems approach: examples from school community-academic partnerships. *Ethn Dis*. 2018;28:417–426.

Love HE, Schlitt J, Soleimanpour S, Panchal N, Behr C. Twenty years of school-based health care growth and expansion. *Health Aff (Millwood)*. 2019;38:755–764.

Mayworm AM, Lever N, Gloff N, Cox J, Willis K, Hoover SA. School-based telepsychiatry in an urban setting: efficiency and satisfaction with care. *Telemed J E Health*. 2020;26:446–454.

Miller DN, Mazza JJ. School-based suicide prevention, intervention, and postvention. In: Flett GL, Saklofske DH, Leschied AW, eds. *Handbook of School-Based Mental Health Promotion: An Evidence-Informed Framework for Implementation*. Springer; 2018:261–277.

Porche MV, Fortuna LR, Lin J, Alegria M. Childhood trauma and psychiatric disorders as correlates of school dropout in a national sample of young adults. *Child Dev*. 2011;82:982–998.

Schilling EA, Aseltine RH Jr, James A. The SOS suicide prevention program: further evidence of efficacy and effectiveness. *Prev Sci*. 2016;17:157–166.

Siperstein GN, Wiley AL, Forness SR. School context and the academic and behavioral progress of students with emotional disturbance. *Behav Disord*. 2011;36:172–184.

Stephan SH, Sugai G, Lever N, Connors E. Strategies for integrating mental health into schools via a multitiered system of support. *Child Adolesc Psychiatr Clin N Am*. 2015;24:211–231.

Stoiber KC, Gettinger M. Multi-tiered systems of support and evidence-based practices. In: Jimerson SR, Burns MK, VanDerHeyden AM, eds. *Handbook of Response to Intervention: The Science and Practice of Multi-Tiered Systems of Support*. 2nd ed. Springer; 2015:121–141.

Taylor RD, Oberle E, Durlak JA, Weissberg RP. Promoting positive youth development through school-based social and emotional learning interventions: a meta-analysis of follow-up effects. *Child Dev*. 2017;88:1156–1171.

U.S. Commission on Civil Rights. Beyond suspensions: examining school discipline policies and connections to the school-to-prison pipeline for students of color with disabilities. U.S. Commission on Civil Rights; 2019.

U.S. Department of Education Student Privacy Policy Office, The family educational rights and privacy act: guidance for eligible students. Student Privacy Policy Office, U.S. Department of Education; 2020.

Vona P, Jaycox LH, Kataoka SH, Stein BD, Wong M. Supporting students following school crises: from the acute aftermath through recovery. In: Harrison JR, Schultz BK, Evans SW, eds. *School Mental Health Services for Adolescents*. Oxford University Press; 2017:78–103.

Wang MT, Degol JL. School climate: a review of the construct, measurement, and impact on student outcomes. *Educ Psychol Rev*. 2016;28:315–352.

Weisbrot DM. “The need to see and respond”: the role of the child and adolescent psychiatrist in school threat assessment. *J Am Acad Child Adolesc Psychiatry*. 2020;59:20–26.

Zirkel PA. Special education law: illustrative basics and nuances of key IDEA components. *Teach Educ Spec Educ*. 2015;38:263–275.

54.9 Prevention of Psychiatric Disorders in Children and Adolescents

William R. Beardslee, M.D.

Over the last few decades, there has been a significant expansion of empirical research supporting the efficacy and effectiveness of programs to prevent the onset of mental health and behavioral problems in children and youth. The Institute of Medicine issued reports in 1994 and 2009 on the prevention of mental health disorders, including one specifically on the prevention of mental, emotional, and behavioral disorders in young people. The latter chronicled an expanding body of efficacy and effectiveness trials of preventive interventions, highlighting a deeper understanding of risk and protective factors for mental illness, as well as advances in neuroscientific understanding that underpins preventive intervention.

Since then, a number of relevant new reports from the National Academy of Sciences have been released. For example, the NAS published 2016 a report on parenting and young children ages 0 to 8, which describes programs for families with children in a developmental period that is crucial for lifelong prevention, and the NAS also released a 2019 report on addressing child poverty, which is an essential risk factor to address in the prevention of childhood mental disorders. Most notably, the NAS published in 2019, “Fostering Healthy Mental, Emotional, and Behavioral Development in Children and Youth: A National Agenda,” which emphasizes not only the continued strong research supporting preventive interventions since the 2009 prevention report was released but also the importance of widely disseminating and implementing prevention programs so many more children, youth, and families may benefit. The strong empirical support for preventive interventions, as detailed in the Institute of Medicine (IOM) and National Academies of Science (NAS) reports, suggests that it is now time to implement evidence-based programs to decrease the incidence of behavioral, emotional, and mental disorders in children and youth. Practical strategies for correcting problematic developmental trajectories have been identified, and the salient characteristics of target populations have been defined. Rigorous preventive interventions have been implemented across diverse populations and have been shown to be effective. Finally, informative longitudinal studies have followed children into adulthood and demonstrated differential outcomes and the important influences of risk and protective factors at multiple levels.

There are other reasons why the implementation of preventive interventions ought to be pursued with urgency. Prevention Science, at its core, identifies risk factors for the development of future pathology and seeks to reroute that trajectory by increasing protective factors, including adaptive coping skills. Risk factors for child and youth mental health disorders can occur on many levels, ranging from the highly individual, such as genetic predisposition or cognitive style, to broad social and environmental contexts, such as living in a community without adequate access to resources. Some of the most salient risk factors for childhood mental illness, including lack of social connections, having a parent with active mental health challenges, living in poverty, and exposure to chronic stress, are exacerbated by the global COVID-19 pandemic, which at the time of writing, is still devastating most countries around the world. The psychological fallout of this era-defining public health crisis has yet to be seen but is anticipated to be great. Programs to prevent mental health disorders in youth are more important than ever.

Coinciding with the COVID-19 pandemic in the United States was the murder of a Black man, George Floyd, whose highly publicized death was the catalyst for widespread acknowledgment of the problem of racial injustice in this country. Racial inequality is notably intertwined with poor health outcomes, as has been apparent in the way communities of color have been disproportionately affected by COVID-19. Racial disparities and other forms of inequality are important risk factors for mental health disorders and must be addressed to prevent mental health disorders in children and youth.

The study of the prevention of psychiatric disorders offers an unusual opportunity to study mechanisms by which disorder is either expressed or prevented and, as such, is an important part of overall research on development. Preventive intervention research also highlights the influence of broader forces on children's development, such as neighborhood characteristics, inequality, and lack of access to resources.

There has been some reluctance by psychiatrists and other mental health professionals to engage in advocacy for and practice of preventive interventions. This is perhaps because prevention requires a major paradigm shift away from a traditional concentration on diagnosis and treatment of those who are seriously ill and what can be done for the child or adolescent immediately. Prevention asks a different question: what does the child, the family, and even the broader community, need 1, 3, or 5 years down the line in order to ensure the best possible trajectories of healthy development? Although some psychiatrists have been reluctant to embrace developments in preventive intervention, psychiatrists and other mental health professionals are in key positions to identify those children who are at the highest risk for the development of psychiatric disorders and institute appropriate interventions. It is important that the same quality of treatment and prevention services be available to all children and families. This requires being able to implement and disseminate evidence-based prevention and treatment interventions widely.

In addition, as the field of prevention has increased in its understanding of the role of genomic variations in the onset of illness and of the biologic, as well as the psychosocial, context of risk and protective factors for illness, preventive interventions are increasingly conceptualized as including both biologic and environmental components. The full integration of genomics in mental health preventive interventions has been slow because many of the investigators and clinicians who have been involved with the interdisciplinary development of preventive interventions are not familiar with molecular genetics.

In fact, focusing on prevention requires mental health professionals to recognize that mental health and physical health are inseparable and, hence, preventive interventions in one realm often exert a positive effect on another. Successful preventive interventions are inherently interdisciplinary, drawing on findings and insights from psychology, sociology, medicine, adult and child psychiatry, social work, developmental neuroscience, and genetics. Mental, emotional, and behavioral disorders are developmental. That is, they have their onsets at different points across the lifespan, requiring a series of interventions that can be positioned at different points across the lifespan.

The development of a new domain of science can be facilitated by the introduction of new knowledge from related fields that require a major reassessment of previous beliefs. This reassessment can subsequently lead to an integration of these new insights into the traditional core knowledge base of the science.

Every medical specialty is being challenged to consider how genomic variations influence the onset of illnesses as well as the selection of treatments that can be tailored to be maximally effective for each patient. This new orientation provides an opportunity to begin to practice “individualized, personalized medicine.” This approach to diagnostic evaluation focuses on biologic vulnerabilities that can be identified in each individual patient. Consequently, preventive interventions, in the same way as therapeutic interventions, will increasingly be conceptualized to include both biologic and environmental components. Consideration of molecular genetics is essential in this regard.

This chapter focuses on preventive interventions for children, youth, and families. The current understanding of prevention research is guided by several key principles. First, it is essential to anchor medical practice in the recognition that access to care, including prevention, is a basic human right and a basic civil right. There is increasing recognition that the practice of medicine involves moral principles, including addressing inequities and disparities. Second, because improving the health of a population is an essential goal, preventive interventions must be delivered widely to all who would benefit. Third, health disparities remain a growing and profound problem, and preventive interventions can help address such disparities. Fourth, awareness of cultural and linguistic competence and the need for cultural humility must guide the adaptation and delivery of preventive interventions. Finally, given the relative maturity of prevention science and emphasis on implementation in recent authoritative national reports, a focus on prevention is timely and necessary.

DEFINITIONS

In an attempt to clarify the research evidence and advance the science of prevention, the 1994 IOM Committee on the Prevention of Mental Disorders recommended the adoption of an alternative definitional taxonomy for mental health interventions: prevention, treatment, and maintenance. The use of the term “preventive intervention” in this classification system was restricted to interventions provided prior to a patient receiving a psychiatric diagnosis. The term “treatment” was reserved for interventions targeted at patients with a diagnosis. The term “maintenance” referred to the care provided to patients with chronic illnesses who had not had remission as a consequence of their primary treatment.

The IOM Committee recommended that preventive interventions, which were previously referred to as primary prevention, be subdivided into three categories based on their targeted population. These categories are universal, selective, and indicated preventive interventions, as described below.

Universal preventive interventions are targeted to the general public or all of a subset thereof (for example, all 6-year-olds in Massachusetts). By definition, universal interventions are provided to individuals who have all levels of risk, including no risk factors at all. A common example of a universal preventive intervention is the fluoridation of a water supply to improve the dental health of a community. Selective preventive interventions are targeted to individuals whose risk of developing mental disorders has been demonstrated to be increased. These identified risk factors may be associated with an imminent onset or with risk at a later point in development. Biologic markers that indicate a predisposition for a mental disorder but are detected in individuals that are free of symptoms represent an example of medical criteria that can be used to initiate a selective preventive intervention. Indicated preventive interventions are targeted to high-risk individuals identified as having detectable signs or symptoms associated with the ultimate development of a mental disorder. Symptoms or signals, which are early warning signs of the future development of a psychiatric illness, may arise years before a patient develops a sufficient number of symptoms to meet the full diagnostic criteria that are required to establish the occurrence of the psychiatric disorder. Indicated preventive interventions can be initiated early in the natural course of a mental disorder when the first signs and symptoms appear.

Sufficient evidence has accumulated to include mental health promotion as a fourth major category of preventive intervention (see Fig. 54.9–1). Mental health promotion aims to enhance an individual’s ability to achieve age-appropriate developmental tasks or developmental competence and to have a positive sense of mastery and well-being, as well as to strengthen an individual’s ability to cope with adversity.

The term risk factor is used in both prevention and treatment, but the implications of risk for these two concepts are different. In prevention, a risk factor must predate the onset of the illness and can be the characteristic that is used to identify a target population. The reduction of risk factors that are both causal and malleable is a goal of preventive interventions. Once the diagnosis of an illness has been made, risk factors that are subsequently identified may be used to guide treatment decisions and provide insight into prognosis. One well-studied example of a risk factor is parental mental illness. Another is undergoing severe loss, such as the death of a family member.

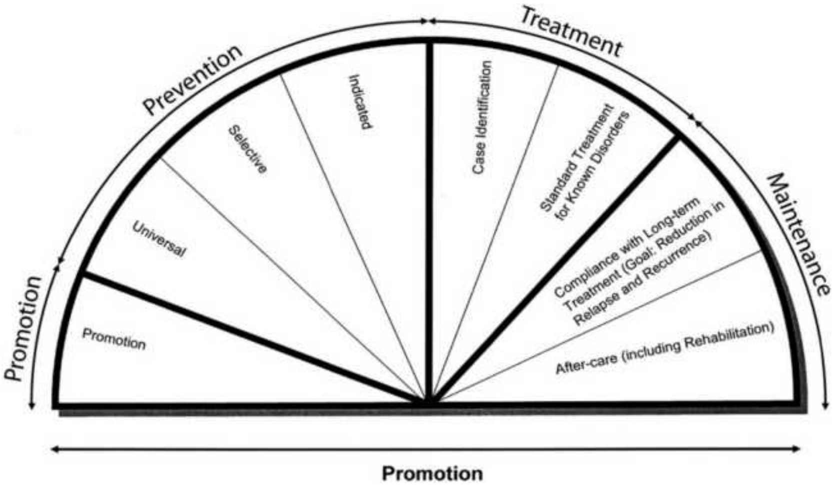
Prevention
Treatment
Maintenance
Promotion
Universal
Selective
Indicated
Case Identification
Standard Treatment for Known Disorders
Compliance with Long-term Treatment (Goal: Reduction in Relapse and Recurrence)
After-care (including Rehabilitation)
Promotion
Promotion

FIGURE 54.9–1. The mental health intervention spectrum for mental disorders. (Reprinted from Mrazek PJ, Haggerty RJ. *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research*. National Academy Press; 1994.)

Protective factors decrease the risk of the onset of mental illness. An individual with one or more protective factors is at a lower risk for the development of a specific psychiatric illness than an individual without the protective factor. Most protective factors are nonspecific and have been perceived to be universally beneficial. Common examples of protective factors include a supportive family environment, adequate family financial resources, and access to high-quality education.

Increasingly, research has highlighted the ways in which biologic and environmental risk and protective factors interact. The study of epigenetics has emphasized the important ways in which environmental influences can profoundly affect gene expression across generations. Exploration of gene–environment correlations and gene–environment interactions offers important ways to study mechanisms.

EXAMPLES OF RESEARCH ON THE PREVENTION OF MENTAL HEALTH PROBLEMS IN YOUTH

Figure 54.9–2 presents a schematic representation of the different interventions that are likely to be effective across the span of childhood. Early in life, interventions for high-risk families (e.g., home visitation, center-based care) have been shown to be valuable. As children enter school, both academic enrichment activities and social skills training are valuable, as are various kinds of parent training. As youngsters enter adolescence, there is evidence that some programs can cut the incidence of major depression, and others can substantially reduce substance abuse.

Across the bottom of the diagram, three types of factors that occur across the span of childhood and adolescence are emphasized: the importance of addressing families in adversity; the importance of interventions that strengthen the fabric of schools and communities; and policy interventions such as those that provide ready access to preventive and treatment services, and that support families. A central theme in the 2009 IOM report is that preventive interventions can reduce or minimize toxic biologic and psychological processes, richly reinforce self-regulatory prosocial behavior, teach prosocial skills and values, and foster acceptance. What follows is a discussion of a selection of preventive interventions that are particularly relevant to psychiatric practice and which illustrate general points about prevention approaches. Rather than trying to comprehensively review all prevention approaches, the focus is on preventive interventions that are particularly relevant to psychiatric practice and that illustrate more general points about prevention approaches. First is a review of community-level interventions for child maltreatment prevention, from which useful lessons about community-level risk factors and prevention can be drawn. Next, this chapter reviews interventions for families that have experienced specific types of adversity. This is followed by a discussion of interventions to prevent externalizing behaviors, depression, psychosis, and substance abuse.

COMMUNITY-LEVEL PREVENTIVE INTERVENTIONS FOR CHILD MALTREATMENT

Maltreatment of children is a significant public health issue and a potent risk factor for later psychiatric illness, including post-traumatic stress disorder, substance misuse, depression, and suicide attempts. Factors associated with child maltreatment can be conceptualized at the individual, family, and community levels. Due to structural inequality in American society, some family-level risk factors, including the experience of food insecurity, unemployment, and poverty, are apparent at the community level, as well. Communities with high levels of unemployment and high rates of poverty are associated with higher rates of child maltreatment, and collective efficacy and social network size in these communities are associated with lower rates of child maltreatment. Reducing inequality and increasing support within a community are, therefore, ideal targets to prevent child maltreatment at the community level. Addressing these risk factors may have additional benefits in reducing mental health problems in children, as poverty, child maltreatment, and other related manifestations of disadvantage are significantly linked to child mental health. Efforts to improve poverty, food insecurity, and unemployment are particularly important in the wake of the COVID-19 pandemic, which has exacerbated existing problems in these areas and weakened protective factors, such as social cohesion and social network ties within a community.

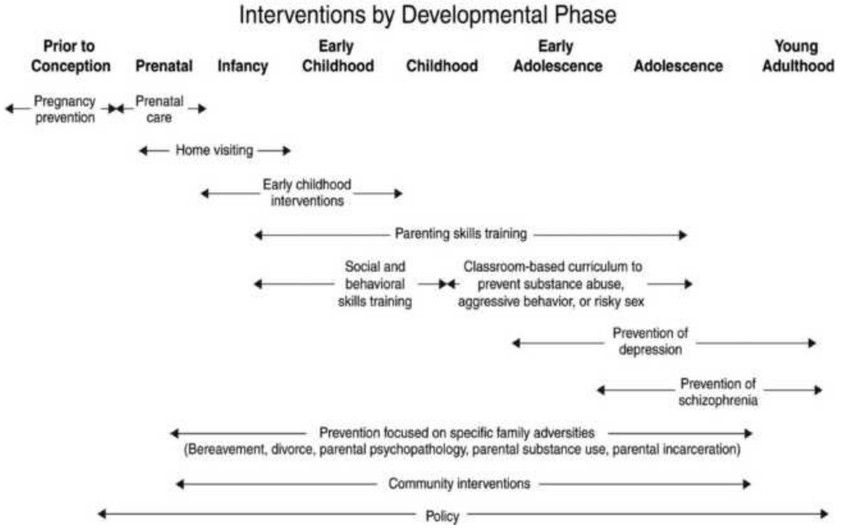
Interventions by Developmental Phase
Prior to Conception
Prenatal
Infancy
Early Childhood
Childhood
Early Adolescence
Adolescence
Young Adulthood
Pregnancy prevention
Prenatal care
Home visiting
Early childhood interventions
Parenting skills training
Social and behavioral skills training
Classroom-based curriculum to prevent substance abuse, aggressive behavior, or risky sex
Prevention of depression
Prevention of schizophrenia
Prevention focused on specific family adversities
(Bereavement, divorce, parental psychopathology, parental substance use, parental incarceration)
Community interventions
Policy

FIGURE 54.9–2. Interventions and their targeted developmental stages.

An example of a community-wide prevention program is the Strong Communities for Children initiative. This program utilized outreach workers to mobilize volunteers and community organizations to form their own local action plan to "keep kids safe" and used resources acquired through the outreach campaign to help build social support networks for parents and provide direct assistance to in-need families. This program was successful in both high- and low-resource communities, with increased social support, collective efficacy, parental efficacy, and child safety at home, and decreased parental stress and child maltreatment, compared to communities that did not implement the Strong Communities program. This program is notable for changing social norms, a formidable feat, and for creating lasting positive change with a small starting budget and a team of a few paid workers. A replication study in Israel is currently underway.

Preventive Interventions Targeting Family Adversity

Examples of effective parenting programs in risk situations are the effective programs designed and implemented by Wolchik, Sandler, and colleagues at the University of Arizona. These programs are the New Beginnings Program (NBP) for families facing divorce and the Family Bereavement Program (FBP) for families facing bereavement.

The NBP is a group-delivered program designed to improve parenting quality and decrease child exposure to parental conflict, two factors associated with child outcomes after divorce. The program has been studied in two efficacy and one effectiveness trial (the former of which tested the program in a research setting, and the latter of which tested the program while implemented in real community contexts).

In the efficacy trials, the program was first found to be associated with improved parenting and reduced child mental health problems post-test, compared to a waitlist condition. The second efficacy trial used a randomized, controlled clinical trial design to compare the NBP parenting sessions alone, the parenting sessions with an additional child-delivered coping program, and a literature control group that provided mothers with books about child adjustment to divorce. The researchers found no significant differences between the mother-only and mother-and-child interventions, but both conditions of the NBP were associated with significant gains in parenting and reductions in child externalizing problems post-test. The results of 6- and 15-year follow-up studies revealed that the young adults who received either version of the NBP intervention were less likely to have mental health problems, substance use issues, high-risk sexual behavior, or involvement in the criminal justice system and were more likely to experience improvements in self-esteem and grade point average. The NBP was then tested in a randomized effectiveness trial, where the program was implemented through family courts in four countries. Strengthened parenting and reduced child mental health problems were associated with the NBP compared to the active control post-test and after 10 months. Non-Hispanic White families experienced benefits with the program, but this was not true, on average, for Hispanic families, emphasizing a need for culturally appropriate adaptation.

For families struggling with the loss of a parent, Sandler and colleagues developed the FBP. With a similar structure to the NBP, FBP provides caregivers and children with 12 group sessions, four of which include shared activities for both the caregiver and the child or adolescent. The control group received three books discussing child, adolescent, and adult grief over the course of 3 months. A 6-year follow-up indicated that effective parenting, operationally defined as caregiver warmth, acceptance, and consistent discipline, had increased in families receiving the FBP.

The results of the NBP and FBP programs demonstrate that relatively brief interventions have the potential for long-term sustained effects in families that have experienced a major stressor, such as divorce or the death of a parent. More generally, interventions that are highly specific and increase the capacity of parents to parent effectively have a strong knowledge base.

Preventions Targeting Youth Externalizing Behaviors

Webster-Stratton originally developed a treatment intervention for young children with conduct disorder and their families. This intervention, The Incredible Years Parenting Series, includes therapist-led group discussion that emphasizes parent training. Parents view standardized videotapes that present parents and their children engaging in both desirable and problematic interactions. Findings indicate that this program improved parent and child interactions and significantly reduced the children's conduct problems and increased cooperation, compared to a control group. This treatment was later modified to become an indicated preventive intervention for children with early symptoms of conduct disorder but who did not meet diagnostic criteria. More recently, this intervention was modified further to allow it to be used as a selective intervention for children enrolled in Head Start programs who were all considered to be at high risk but were not symptomatic at the time of the intervention. Teachers in the active intervention group used more positive classroom management techniques, and their students had fewer conduct problems and more emotion regulation than control students. Since then, this program has been widely disseminated.

Poduska, Kellam, and colleagues document the long-term effects of a universal classroom-based intervention with children during their first- and second-grade years. Their intervention, the Good Behavior Game (GBG), is directed at improving the classroom management skills of teachers, socializing children to become better students, and reducing aggressive and disruptive behavior. A randomized, controlled trial was carried out in 40 first-grade classrooms in 18 elementary schools in Baltimore, Maryland. Significant positive results were found, although these outcomes varied by gender, baseline level of risk, and cohort. Generally, the children with the highest level of risk from the fall of the first grade through the seventh grade benefited more than other children who were assessed to be at lower risk. The most aggressive boys in first grade who received the GBG were the children who demonstrated the greatest reduction in aggressive behavior over the course of first through seventh grades. These boys were less likely to develop antisocial personality disorder as an adult and had a 50% reduction in lifetime illicit drug use. These results highlight the benefit of early behavioral interventions.

Preventive Interventions Targeting Depressive Disorders

Quite fruitful work has been done in the area of the prevention of depression. There are both specific risk factors for depression, such as having a parent with depression, a depressogenic cognitive style, or a prior depressive episode, and nonspecific risk factors, such as poverty, exposure to violence, and social isolation. Research suggests that children who have either specific or nonspecific risk factors are at increased risk for the disorder and that interventions targeting these risk factors may prevent the onset of depressive symptoms and disorders. Recent meta-analyses support the value of preventive intervention approaches targeting depression and their efficacy in preventing episodes of major depression in both adults and youth.

An example of a strong depression prevention program is the Coping with Depression Course (CWD), developed by Lewinsohn in the mid-1970s. After finding much success in the reduction of depressive symptoms in adults who received this prevention and treatment course, Clarke and colleagues adapted the CWD for use with an indicated sample of adolescents and demonstrated the prevention of unipolar depressive disorder. Later, Clarke and colleagues used the CWD course with adolescents with subsyndromal depressive symptoms and parents experiencing depression. Only 9.3% of these adolescents went on to develop a major depressive disorder, compared to 28% of adolescents who were at similar risk but did not receive preventive intervention. Garber and colleagues undertook a four-site replication of this study, again using adolescents at double risk because their parents had experienced major depression and they themselves either had prior depression and/or symptoms of depression. This study demonstrated a strong and sustained effect of the prevention program. At a 33-month follow-up, adolescents in the group cognitive-behavioral prevention program had significantly fewer onsets of depressive episodes compared to those in the comparison group; however, for adolescents whose parents were currently depressed at the time of enrollment, the average rate of onset did not differ statistically between the intervention and nonintervention groups. Brent and colleagues found at a 6-year follow-up, adjusting for current parental depression at enrollment, adolescents in the cognitive behavioral prevention program still had a lower incidence of depression. The effect of the program was strongest against a new onset of depression in the first nine months after enrollment in the program, but gains were maintained through several years of follow-up.

A second example of a preventive intervention targeting youth depression was developed by Compas and colleagues, who explored preventive interventions for children with depressed parents using a family group cognitive-behavioral (FGCB) strategy. Children between the ages of 9 and 15 who had a parent with a history of major depressive disorder participated, along with their parents, in either the 12-session FGCB intervention or received three sets of written materials about depression in the mail every 8 weeks. Eighteen and 24-month follow-ups demonstrated that the FGCB preventive intervention significantly reduced the onset of the children’s depressive episodes for up to 2 years after the intervention ended compared to the control group. Additionally, this preventive intervention reduced the children’s internalizing and externalizing symptoms. Compas and colleagues found a mediating role of parental behavior, specifically parental guilt induction. Similar to Sandler and colleagues’ FBP, it was found that a reduction in negative parental behaviors improved family functioning; One year after the end of the intervention, reductions in parental guilt induction were associated with significantly lower internalizing symptoms and lower levels of negative cognitive style in the youth participating in the FGCB program.

An illustration of a preventive intervention that reduces the risk associated with depression in parents is the Preventive Intervention Project, which targets relatively healthy children aged between 8 and 15 years who have a biologic parent with a clinical mood disorder. Two manual-based programs were tested. The first intervention consisted of a two-session group lecture format for the depressed parents;

the second intervention used a clinician-facilitated format consisting of 6 to 11 sessions, including separate meetings with parents and children, as well as a family meeting. In both formats, psychoeducational information about mood disorders was provided, and parents were helped to better support the initiatives of their children. In the clinician-facilitated format, efforts were made to link information about mood disorders to the depressive illness experiences of the parents in the intervention. At the 4.5-year follow-up, parents who received either the lecture format or the clinician-facilitated intervention reported improvements in their behaviors and attitudes regarding their children, and the internalizing symptom scores of the children who had received either of the interventions improved when they were assessed at follow-up. Moreover, there was a significant relationship between the number of positive changes that the parents reported in their behaviors and attitudes regarding their children and the improved understanding of the course and treatment of mood disorders that was achieved by their children. This family-based approach is particularly applicable to implementation in a clinical setting, which could be accomplished by adding a relatively small number of additional sessions to the standard treatment of those adult patients with mood disorders who have children in the therapeutic range. Beardslee's family-based depression prevention approach has been adapted successfully for use in different ethnic and cultural groups, both in the United States and around the world.

Another area with very promising prevention strategies is that of intervention for mothers with young children. One example of this class of intervention, Silverstein and colleagues used a problem-solving educational intervention with mothers in Head Start with depressive symptomatology. In a randomized clinical trial, mothers with lower-level baseline depressive symptoms who received the educational program had fewer depressive symptom elevations than mothers with lower-level baseline depressive symptoms who did not receive the program. No effect was found, however, for mothers whose level of depressive symptoms was higher at baseline.

Prevention of Early-Onset Psychosis

Several investigators have focused on children and youth at high risk for the development of schizophrenia and the prodromal symptoms of psychosis. The goal of much of this work has been to identify risk factors or biologic markers that could inform studies designed to prevent the onset of an initial episode of psychosis or to optimally treat a patient at the very earliest point in the initial onset of the illness. Although these intervention strategies all have been multidimensional, one critical aspect of these interventions is the prudent use of antipsychotic medications. For example, McGorry and colleagues gave young first-episode psychosis patients a low dose of risperidone therapy to treat their early symptoms; results revealed that risperidone was very effective in reducing psychotic symptoms and improved the individual's quality of life, treatment experience, and adherence.

In the United States, with efforts to reduce the duration of untreated psychosis (DUP) between early symptoms to engagement in treatment, Srihari, McGlashan, and colleagues have developed the Specialized Treatment Early in Psychosis: Early Detection (STEP-ED) program, an adaptation of McGlashan and Johannessen's Early Treatment and Intervention in Psychosis (TIPS) study. The TIPS Early Detection in Psychosis Study, conducted in Norway and Denmark, provided the general public, schools, and Early Detection teams with information about prodromal, first-episode schizophrenia symptoms. This program shortened the DUP by an average of 1.5 years. The STEP-ED researchers have been conducting a trial of the TIPS-adapted model in the United States since 2014 and hope to improve presentation, engagement, and early outcomes through the reduction of DUP. Another way to prevent the onset of psychosis, other than reducing the DUP, is to detect symptoms early and intervene to prevent psychosis. McFarlane and colleagues created and implemented the Family-aided Assertive Community Treatment (FACT) program and conducted a 2-year follow-up study. Adolescents were assigned to one of three groups, clinically low-risk (CLR), clinically high-risk (CHR), or very early first-episode psychosis (EFEP), based on their score on the Positive Symptom Scale of the Prodromal Symptom Scale. Individuals designated to the CHR and EFEP groups received FACT, which incorporates a variety of interventions, including psychoeducational multifamily group (PMFG) therapy, elements of assertive community treatment, supported education and employment, and psychotropic medication. Those in the CLR group received community care, which consisted of monthly monitoring through phone assessments. Results demonstrated that 2 years after treatment, FACT was effective in reducing symptoms, whether positive, negative, disorganized, or general symptoms. Participants in the FACT condition, who were at risk or experiencing very early psychosis, also had improved Global Assessment of Functioning scores, work and school participation, and global outcomes. The findings suggest that preventive interventions can be effective in preventing the onset or progression of psychosis and that family interventions are extremely important in the treatment and prevention of psychotic symptoms. Given consistent evidence that demonstrates that subsequent episodes of psychosis are associated with a less optimistic prognosis and probable neurodegenerative cerebral changes, the development of preventive interventions for schizophrenia and bipolar disorder is a high public health priority. The identification of children who are at risk as a result of carrying susceptibility genes for these disorders is a research strategy that has not yet been systematically implemented.

Prevention of Substance Abuse Disorders

Many researchers have developed preventive intervention programs for substance abuse in children and in adolescents. Botvin and Eng developed a universal prevention program, life skills training (LST), which has been shown to be quite successful in reducing substance abuse in middle and high school students. During 30 lessons over the course of 3 years, students participating in LST were taught skills, information, and attitudes to resist temptations to use substances in adolescence. In dozens of efficacy trials, LST has reduced cigarette smoking, alcohol consumption, methamphetamine usage, and marijuana use by as much as 80% compared to control groups. Similar results have been found with minority group students. Some LST trials have discovered that elementary school programs have positive long-term effects. The LST intervention’s effects have been maintained up to 12 years after the completion of the program. A booster program designed for LST has demonstrated maintenance of and even bolstered the original intervention’s effects.

Likewise, Kumpfer and colleagues developed the Strengthening Families Program: for parents of youth 10 to 14 (SFP). This 14-week, family-focused substance abuse prevention program has demonstrated significantly greater reduction rates of substance use in high school students compared to school-based alcohol prevention programs. SFP also has generated many significant positive effects on family functioning and parent-child relationships. SFP has been successfully adapted for many cultural groups, such as Hispanic, African American, Native American, and Asian Pacific Islander families, and has been widely adopted internationally.

Now that children and adolescents have greater access to mobile devices and computers, Internet-based prevention programs have become increasingly popular. In a meta-analysis by Champion and colleagues, seven school-based substance abuse prevention programs delivered through either the Internet or CD-ROMs, including Consider This, ASPIRE, CLIMATE, Head On, and several other programs, were assessed for their effectiveness in preventing alcohol, tobacco, and cannabis use in adolescents. Five of the seven programs were found to reduce drug and alcohol consumption in the follow-ups, which were completed between 6 and 34 months after the cessation of the program. Although the reviewers discuss how the effect sizes were modest, the reduction of substance use after the implementation of Internet and computer prevention programs provides evidence for how these new methods for disseminating interventions have the potential to be very effective and can be used by individuals in a variety of urban and rural settings.

ISSUES IN THE PREVENTION OF MENTAL ILLNESS

Barriers to Providing Clinical Preventive Interventions

Despite the very strong scientific evidence about the benefits of preventive interventions, there are relatively few psychiatrists who are involved in the administration of evidence-based preventive interventions. There appear to be three major barriers to the adoption of these interventions. First, there is no clinical infrastructure for delivering preventive and early intervention services. The provision of services to large numbers of children requires the coordination of multiple agencies, disciplines, and service delivery systems. Without a well-developed and accessible clinical infrastructure, most families are unable to receive preventive services even where treatment services exist. A second barrier to the provision of early clinical intervention is that there are insufficient resources to pay for these services. The lack of reimbursement through either public payment plans such as Medicaid or by private insurers prevents most psychiatrists and other mental health clinicians from providing preventive services. A third barrier to initiating early interventions is the absence of an academic curriculum designed to train general and child psychiatrists in the assessment of children who demonstrate very early symptoms of psychiatric disorders. Information and training in available evidence-based, early-intervention, manual-based protocols are not routinely provided to psychiatry residents.

There is a significant and growing public health awareness of the disparity between the recognition of the need for both prevention and early intervention services and the absence of a plan for developing an efficient and effective systematic delivery of such services. An immediate goal should be the development of new clinical initiatives that would begin to provide appropriate services to the highest-risk populations.

Economic Issues

Ideally, psychiatric administrators should take responsibility for the management and funding of preventive interventions. However, to achieve this goal, new mechanisms to reimburse the provision of preventive services within the mental health field must be created. One issue will be a revision of diagnostic and billing codes to define specific evidence-based prevention services and to ensure payment for the delivery of these services. The standard medical reference for procedures that will be paid for by insurance companies is the Current Procedural Terminology (CPT) codes of the American Medical Association. CPT codes have been designed to define virtually all clinical procedures that have been accepted as legitimate services. Unfortunately, payment for many of the current CPT codes that could be used to support services designed to prevent psychiatric illness has not been classified as traditional psychiatric interventions that are covered by medical insurance. Furthermore, even if insurance policies were to change and begin to make payments for preventive interventions that were authorized, many of the available CPT codes that are related to prevention-related services have been assigned very low relative value units (RVUs). When a CPT code is given a low RVU assignment, payment for the service will be quite low. Although a common compendium of CPT codes is established for all medical subspecialties, virtually no mental health specialists are able to use available codes for risk reduction and anticipatory guidance because of the difficulties that exist in receiving adequate payment. This needs to change, and appropriate provision for payment needs to be made.

Another administrative problem is that payment for psychiatric services has traditionally required that a psychiatric diagnosis be identified for the patient prior to the clinician receiving payment. By definition, children who are receiving preventive services do not have a psychiatric diagnosis that would be included in the 10th or 11th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10 or ICD-11) at the time of the provision of a preventive intervention. Instead, these children may have problems that could be classified using a “V code.” Unfortunately, many insurance companies make the argument that the services provided to deal with problems that are categorized using V codes are not medically necessary. Therefore, these services are not eligible for medical insurance, and these payers deny payment to any psychiatric patients that have not been assigned a traditional Axis I diagnosis. Newer payment approaches that emphasize payment for covered lives offer considerable promise for funding prevention.

Training Issues

The next generation of psychiatric clinicians must be better trained in the science of the prevention of mental disorders. Much needs to be done to achieve this goal. For evidence-based prevention programs to be implemented on a national scale, it is absolutely necessary that clinicians be appropriately trained to provide these specified interventions. Supervised training experiences to provide trainees with appropriate skills will have to be incorporated into the traditional residency curriculum as one of the “core competencies” that must be developed during residency training. The ability to identify children at risk and deliver evidence-based preventive interventions must become an academic expectation of training. However, for this goal to be achieved, the importance of preventive interventions will need to become more explicitly stated in the residency training requirements that are available to provide guidance to the design of clinical training programs. Ultimately, board certification and recertification must require the demonstration of clinically relevant applications of preventive interventions as part of the core knowledge base that is examined.

Research Issues

Continuing a strong research program in prevention is essential. As noted, it provides one of the best ways of understanding the mechanisms by which either disorder is expressed, or the balance of risk and protective factors is altered, and disorders are prevented. Research investigators should be encouraged to assess multiple risk factors and multiple outcomes crossing the domains of mental health, substance abuse, physical health, and learning disorders. Priority should be given to sequential preventive interventions that can strategically target high-risk children prenatally, in infancy, in early childhood, in elementary school, during the teenage years, and finally, in young adulthood. In addition, research efforts must consider and track high-risk youth and the different interventions they need at different stages (e.g., high-quality child care, daycare, and early education followed by access to high-quality schools). Finally, preventive intervention studies must aim to assess the mechanisms by which such interventions work. Greater attention to genetics, epigenetics, and gene–environment interactions is warranted as this work goes forward.

The following case example illustrates how newer pharmacogenomic approaches offer great promise. George was 16 years old when he started experiencing depressive symptoms after his girlfriend decided to end their relationship. George had a family history of bipolar disorder, and by using pharmacogenomic testing, a clinician familiar with preventive approaches to mental illness was able to treat him safely.

When George presented to the clinician, he reported sleeping for more than 10 hours a night, rarely leaving his home, and having little interaction with his friends. His maternal uncle had struggled with bipolar disorder for many years. Although his parents had no history of mania or hypomania, both his mother and his father had been successfully treated with selective serotonin reuptake inhibitors for episodes of major depressive disorder.

A series of eight outpatient cognitive-behavior therapy sessions targeting his depressive symptoms proved to be minimally effective. Subsequently, George was placed on a dose of 20 mg of fluoxetine per day while being followed closely for the possible emergence of suicidal ideation. Within 2 weeks, he reported that he needed to sleep for only 1 or 2 hours each night and was experiencing marked surges of energy.

His psychiatrist became concerned that George was becoming hypomanic and was at risk of developing bipolar disorder. Prior to selecting another medication to manage these new symptoms, he ordered pharmacogenomic testing of a panel of genes with the goal of minimizing additional adverse reactions. George was determined to have a normal genotype of his cytochrome P450 2C19 gene, indicating that he could effectively metabolize 2C19 substrate medications. However, his cytochrome P450 2D6 and 2C9 genotypes documented that he was a poor metabolizer of both 2D6 and 2C9 substrate medications. In addition, his serotonin transporter genotype was associated with an increased risk for the onset of depression, and his GRIK3 genotype indicated that he was at an increased risk for the development of bipolar disorder.

George was started on a low dose of lithium, and the fluoxetine that he had been receiving was discontinued. After 2 months of outpatient treatment on lithium, he complained of having less energy and had become very discouraged about his ability to complete his planned applications to attend college. Given the results of his pharmacogenomic genotype testing, he was placed on a dose of 10 mg of escitalopram per day in addition to the lithium. After 4 weeks, he reported feeling much better and had not experienced a recurrence of hypomanic symptoms. He was subsequently able to complete his college applications without further difficulty.

George was maintained on lithium and escitalopram throughout his first year at college. During the summer before the beginning of his second year away at school, both medications were discontinued, and he remained euthymic.

Although fluoxetine is an effective and economical treatment for major depressive disorder in adolescents, fluoxetine is primarily a 2D6 substrate medication. Of note, the cytochrome P450 2C9 enzyme plays a secondary role in the metabolism of fluoxetine. Patients with both of these pathways compromised are at very high risk of developing side effects while taking standard doses of fluoxetine.

The clinical observation that manic activation may occur in depressed adolescents when treated with selective serotonin reuptake inhibitors has been frequently reported. However, clinicians have concluded that the benefits of treating many depressed adolescent patients with antidepressant medications have outweighed the risks of activating bipolar symptoms in the relatively few adolescent patients who develop manic symptoms when treated with these medications. Fortunately, it is now possible to identify genetic risk factors for activation and offer alternative medications. Managing the treatment of depression in a manner that minimizes activation can be conceptualized as an example of an indicated preventive intervention designed to prevent the onset of a first full episode of bipolar disorder. More generally, this case illustrates a way of thinking that is likely to be more useful in both treatment and prevention going forward.

LOOKING FORWARD TO THE FUTURE

There is a debate in the prevention field regarding the cost–benefit ratio for universal interventions. Some universal prevention studies have been shown to provide the most benefits to children who are at the highest risk. In contrast, many children who are at minimal risk do not benefit at all. This pattern of response supports the view that early screening and interventions targeted to help populations with specific high risks, including early prodromal symptoms, could be considered for some disorders. Integrating the new, rapidly evolving personalized medicine approaches will become increasingly important in the future, as well as understanding how to effectively disseminate widely those prevention approaches that are strongly supported empirically.

As there are potent nonspecific risk factors for multiple disorders (risk factors including poverty and exposure to violence), more attention also needs to be devoted to addressing these public health drivers. The IOM recognized this in emphasizing the role of poverty, stating in their 2009 report they were "persuaded that the future mental health of the nation depends crucially on how, collectively, the costly legacy of poverty is dealt with."

It is now possible to implement evidence-based selective and indicated preventive interventions to minimize risk factors associated with the onset of mental disorders. Broader implementation of these interventions for children and adolescents at risk for depression, conduct disorder, and psychosis will require continued political commitment. Given the anticipated psychological aftermath of the COVID-19 pandemic and the growing awareness of the terrible inequities in our society, the importance of implementing preventive interventions is greater than ever.

ACKNOWLEDGMENTS

This chapter is dedicated to the extraordinary work and life of Dr. David Mrazek, who passed on May 6, 2013, and his collaborator and wife, Dr. Patricia Mrazek. They were coauthors with me on earlier editions of this chapter. Ms. Anne Diehl, MPH, a freelance researcher, contributed a great deal to this chapter by reviewing the literature, discussing key issues, and assisting with the editing and submission process.

FURTHER READINGS

Allen D, Coombes L, Foxcroft DR. Cultural accommodation of the Strengthening Families Programme 10–14: UK Phase I study. *Health Educ Res*. 2007;22(4):547–560.

Beardslee WR. Master clinician review: parental depression and family health and wellness: what clinicians can do and reflections on opportunities for the future. *J Am Acad Child Adolesc Psychiatry*. 2019;58(8):759–767.

Beardslee WR, Brent DA, Weersing VR, et al. Prevention of depression in at-risk adolescents: longer-term effects. *JAMA Psychiatry*. 2013;70(11):1161–1170.

Beardslee WR, Solantaus TS, Morgan BS, Gladstone TR, Kowalenko NM. Preventive interventions for children of parents with depression: international perspectives. *Med J Aust*. 2013;199(3 Suppl):S23–S25.

Beardslee WR, Wright EJ, Gladstone TR, Forbes P. Long-term effects from a randomized trial of two public health preventive interventions for parental depression. *J Fam Psychol*. 2007;21(4):703–713.

Berwick DM. The moral determinants of health. *JAMA*. 2020;324(3):225–226.

Botvin GJ, Griffin KW. Life skills training: preventing substance misuse by enhancing individual and social competence. *New Dir Youth Dev*. 2014;141:57–65.

Brent DA, Brunwasser SM, Hollon SD, et al. Effect of a cognitive-behavioral prevention program on depression 6 years after implementation among at-risk adolescents: a randomized clinical trial. *JAMA Psychiatry*. 2015;72(11):1110–1118.

Champion KE, Newton NC, Barrett EL, Teesson M. A systematic review of school-based alcohol and other drug prevention programs facilitated by computers or the internet. *Drug Alcohol Rev*. 2013;32(2):115–123.

Clarke GN, Hornbrook M, Lynch F, et al. A randomized trial of a group cognitive intervention for preventing depression in adolescent offspring of depressed parents. *Arch Gen Psychiatry*. 2001;58(12):1127–1134.

Compas BE, Forehand R, Thigpen JC, et al. Family group cognitive-behavioral preventive intervention for families of depressed parents: 18- and 24-month outcomes. *J Consult Clin Psychol*. 2011;79(4):488–499.

Cuijpers P, van Straten A, Smit F, et al. Preventing the onset of depressive disorders: a meta-analytic review of psychological interventions. *Am J Psychiatry*. 2008;165(10):1272–1280.

Garber J, Clarke GN, Weersing VR, et al. Prevention of depression in at-risk adolescents: a randomized controlled trial. *JAMA*. 2009;301(21):2215–2224.

McDonell JR, Ben-Arieh A, Melton GB. Strong communities for children: results of a multi-year community-based initiative to protect children from harm. *Child Abuse Negl*. 2015;41:79–96.

McFarlane WR, Levin B, Travis L, et al. Clinical and functional outcomes after 2 years in the early detection and intervention for the prevention of psychosis multisite effectiveness trial. *Schizophr Bull*. 2015;41(1):30–43.

McGlashan TH, Johannessen JO. Early detection and intervention with schizophrenia: rationale. *Schizophr Bull*. 1996;22(2):201–222.

McGorry PD, Cocks J, Power P, et al. Very low-dose risperidone in first-episode psychosis: a safe and effective way to initiate treatment. *Schizophr Res Treatment*. 2011;2011:631690.

Merry S, McDowell H, Hetrick S, Bir J, Muller N. Psychological and/or educational interventions for the prevention of depression in children and adolescents. *Cochrane Database Syst Rev*. 2004;1:CD003380.

Molnar BE, Scoglio AA, Beardslee WR. Community-level prevention of childhood maltreatment: next steps in a world with COVID-19. *Int J Child Maltreat*. 2021;3(4):467–481.

Mrazek PJ, Haggerty RJ. *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research*. National Academies Press; 1994.

National Academies of Sciences, Engineering, and Medicine. *A Roadmap to Reducing Child Poverty*. The National Academies Press; 2019.

National Academies of Sciences, Engineering, and Medicine. *Fostering Healthy Mental, Emotional, and Behavioral Development in Children and Youth: A National Agenda*. The National Academies Press; 2019.

National Academies of Sciences, Engineering, and Medicine. *Parenting Matters: Supporting Parents of Children Ages 0–8*. The National Academies Press; 2016.

National Research Council, Institute of Medicine. The etiology of depression. In: *Depression in Parents, Parenting, and Children: Opportunities to Improve Identification, Treatment, and Prevention*. National Academies Press; 2009.

O’Connell ME, Boat T, Warner KE, eds. *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*. National Academies Press; 2009.

Poduska JM, Kellam SG, Wang W, et al. Impact of the Good Behavior Game, a universal classroom-based behavior intervention, on young adult service use for problems with emotions, behavior, or drugs or alcohol. *Drug Alcohol Depend*. 2008;95(Suppl 1):S29–S44.

Sandler I, Wolchik S, Mazza G, et al. Randomized effectiveness trial of the New Beginnings Program for divorced families with children and adolescents. *J Clin Child Adolesc*. 2020;49(1):60–78.

Silverstein M, Diaz-Linhart Y, Cabral H, et al. Efficacy of a maternal depression prevention strategy in head start: a randomized clinical trial. *JAMA Psychiatry*. 2017;74(8):781–789.

Srihari VH, Tek C, Pollard J, et al. Reducing the duration of untreated psychosis and its impact in the US: the STEP-ED study. *BMC Psychiatry*. 2014;14:335.

Webster-Stratton C. Mothers’ and fathers’ perceptions of child deviance: roles of parent and child behaviors and parent adjustment. *J Consult Clin Psychol*. 1988;56(6):909–915.

Wolchik SA, Sandler IN, Tein JY, et al. Fifteen-year follow-up of a randomized trial of a preventive intervention for divorced families: effects on mental health and substance use outcomes in young adulthood. *J Consult Clin Psychol*. 2013;81(4):660–673.

54.10 Child Mental Health Services Research

BONNIE T. ZIMA, M.D., M.P.H., AND REGINA BUSSING, M.D., M.S.H.S.

Child mental health services research integrates advances in child psychiatric care and research methods to examine unmet needs for mental health services, service use across care sectors, and quality of care along the care continuum (e.g., prevention, early detection, assessment, treatment, follow-up). To meet this challenge, health services research uses a compendium of quantitative and qualitative research methodologies that bring together expertise from multiple medical specialties and disciplines (i.e., social sciences, biostatistics, economics, health policy, computer science), community partners, and consumers. For child mental health services research, the breadth of expertise also includes the perspectives of primary caregivers, teachers, social workers, pediatric primary care providers, probation officers, and child advocates.

Child mental health services research addresses a wide variety of questions to generate the evidence needed to improve care in real-world treatment settings. Early stages of child mental health services research typically describe access to care and its predictors, the extent of unmet need for child mental health services, and the prevalence of broad indicators of quality of care (e.g., continuity of care, filled medication prescriptions). Findings at this stage are typically a wake-up call alerting others of the potential magnitude of a problem in child mental health care, such as prescription rates of atypical antipsychotic medication among very young children. Some studies may also explore how mutable factors at several levels (i.e., system, community, family, child) independently predict who receives care while adjusting for the clinical need to identify target areas for reducing disparities in access to or quality of care. Examples of questions driving such research are: Among children with a psychiatric disorder and impairment, who is more likely to access care? What are the potential mechanisms underlying disparities in child mental health care?

In addition, child mental health services research focuses on the delivery of care. In real-world treatment settings, implementation of an evidence-based practice (i.e., uptake, engagement, continuity, sustainability) likely varies, and predictors of implementation stages can identify barriers and facilitators for more large-scale use. An example research question is, how should evidence-based practices be modified for use in community-based settings to improve their effectiveness? Following decades of research on integrated care models, mostly for adult depression, there also continues to be a need for research to guide how to integrate, deliver, and sustain the provision of mental health care in pediatric primary care. With advances in technology-enabled care (e.g., telehealth, app-supported evidence-based therapies, feedback from real-time sensor data, clinical decision support tools), research is needed to adapt these advances to improve access to child mental health care, quality of care, and equity.

With the rise in the number of quality measures, child mental health services research also examines the gap between recommended and usual care and their predictors to identify disparities in care. Improving the quality of child mental health care is of public health significance because child psychiatric disorders are prevalent, disabling, and costly. The use of established efficacious treatments in community settings is typically low, and the underuse of recommended care may negatively impact a child’s developmental trajectory. Examples of questions common to quality of care research are: What proportion of children receive care that would pass a national quality measure? What organizational, provider, family, or child characteristics are predictive of acceptable care? Is adherence to national quality measures for child mental health care associated with improved outcomes? And if so, for whom and at what cost? To date, the majority of research examining adherence to quality measures is based on administrative data, such as claims or hospital discharge at the individual level or aggregate data at the health plan level, which have their limitations.

Nevertheless, significant progress in data coordination of electronic health record (EHR) data across large healthcare systems using common data elements has been made, providing new opportunities to leverage very large datasets to address questions relevant to child mental health services research. The rapidly evolving field of data science, defined by the National Institute of Health as “the interdisciplinary field of inquiry in which quantitative and analytical approaches, processes and systems are developed and used to extract knowledge and insights from increasingly large and/or complex sets of data,” provides the methods to integrate EHR-derived data with other relevant data sources, addressing biologic and lifestyle domains. The contents of EHRs may include structured data like vital signs, medication timing and dosing, procedures, and diagnoses, which is compatible with computing algorithms. EHRs also contain unstructured text and images that will require new analytical tools like natural language processing tools to become analyzable. Other data science sources are biologic, such a genomic, proteomic, or metabolomic data, and lifestyle data, such as can be derived from mobile devices and self-tracking tools. Although challenges to coordinate EHR data across different healthcare systems are prominent, advances in data science will fundamentally contribute to the next generation of mental health services research.

In addition to the enriched scope and methods afforded by expanding EHR and data science, child mental health services research also intersects with the rapidly maturing field of implementation science. Implementation science has been defined as the scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine practices, to improve the quality and effectiveness of health services. The emphasis is placed on fitting mental health interventions in the real-world service systems and examines variables at various levels, including patients (e.g., depression remission rates), providers (rates and quality of evidence-based depression treatment), organizations (e.g., collaborative care designs, EPIC workflows supporting evidence-based treatment protocols), and policy (e.g., mental health insurance coverage policies). And finally, the disparate general and mental health outcomes associated with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the COVID-19 pandemic have highlighted the urgency of advancing disparities research from documenting disparities and their predictors to the development of population-level interventions that can be rapidly implemented in real-world treatment settings. Child mental health services researchers will be called upon to continue the core tradition of integrating expertise from these specialties and disciplines with input from community partners, families, and youth to provide the scientific undergirding to improve access to equitable, high-quality child and adolescent mental health care.

As an introduction to child mental health services research, this section briefly describes the historical context that shaped the field, including landmark national reports, the establishment of key professional organizations, research studies, and health policies. Access to care is described as a separate component of quality, following Avedis Donabedian's conceptualization of quality, which argues that accessibility is a separate yet influential component of quality. This is followed by an overview of a well-established model of quality of care with examples of quality of care research and national quality measures, an introduction to a three-tiered conceptual model to organize advances in measurement-based care (MBC), and areas for future research. To supplement, the reader is encouraged to read cross-referenced sections and chapters to stimulate further inquiry about the prevalence of child psychiatric disorders, healthcare disparities, dissemination, and implementation science, MBC, and technologic advances in mental healthcare delivery, health information, and computational methods.

HISTORICAL CONTEXT

Table 54.10–1 summarizes the historical context of child mental health services research by landmark events and their significance. The first half of the twentieth century was characterized by increasing federal initiatives that raised public awareness of the need to improve the quality of life for children, such as a child's right to preventive health care, safe working conditions, and a free public education, particularly for those identified as seriously emotionally disturbed or handicapped. After the midcentury mark, child psychiatry was identified as a relatively new medical subspecialty, with its first national professional organization established in 1953 and board certification in 1959. During the next three decades, landmark epidemiologic studies provided some of the first prevalence estimates of child mental health problems in community-based populations and their correlates. These studies were significant because the findings raised awareness of the substantial proportion of children who met criteria for a psychiatric disorder and the need to improve early detection and access to child mental health services.

During the 1980s, the pediatric community identified psychosocial conditions, such as poverty, family dysfunction, and child exposure to traumatic events, as part of the new morbidity of child health that included learning and behavioral disorders. The conceptualization of child health was further broadened in the following decade to include attention deficit/hyperactivity disorder (ADHD) within the definition of chronic conditions among children with special healthcare needs. Within this same time period, a philosophy for a system of care for children was operationalized. The two core values were that care should be child-centered and community-based. These values were later echoed a decade later in the guiding principles underlying the pediatric medical home. Remarkably, these principles share similar dimensions of "quality" as identified in the "articles of faith" underlying good medical care by R. I. Lee and L. W. Jones more than 50 years earlier. Nevertheless, sustained articulation of these values and principles is important because it provides a relatively consistent framework from which programs serving children with mental health problems can plan and implement new models of care for children with mental health problems. Table 54.10–2 summarizes the guiding principles of these two conceptual models.

During the 1990s, federal initiatives increasingly focused on improving child mental health care and promoting basic science, clinical, and health services research for childhood psychiatric disorders. In preparation for a national epidemiologic study to estimate the prevalence of childhood mental disorders and service use, the Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA) Study was conducted. The MECA Study was a multisite, community-based study that further refined standardized assessment methods of psychiatric diagnosis, impairment, mental health service use, and risk factors and also provided prevalence estimates of childhood psychiatric disorders and mental health service use in four diverse geographic areas.

In response to increasing adaptation of system of care principles to the delivery of community-based services for children and their families, a large-scale effectiveness study was conducted, namely, the Fort Bragg Demonstration Project. Provision of a continuum of care was found to provide more services and be more costly but did not produce better clinical outcomes. These thought-provoking findings encouraged the field to conceptually expand outcomes for children beyond clinical status (e.g., to system-, community-, family-, and parent-level outcomes) and to integrate the influence of varying service use (e.g., intensity, persistence, type) in the determination of effectiveness.

Table 54.10–1.
Historical Context for Child Mental Health Services Research: Landmark Events and their Significance

	Event(s)	Significance
1909	White House Conference on Dependent Children	Raised public awareness of poor living conditions and high cost of homes for emotionally disturbed children. National recommendations for improved child programs stimulated the development of child welfare and foster care system.
1912	Federal Children's Bureau created	Establishment of the oldest federal agency for children. Located in the U.S. Department of Health and Human Services' Administration for Children and Families, the agency is responsible for assisting states in the delivery of child welfare services; providing grants to states, Native American tribes, and communities to operate a range of child welfare services (e.g., child protective services, family preservation and support, foster care, adoption, and independent living); and making major investments in staff training, technology, and innovative programs for children.
1919	The Child Welfare League of America chartered	Establishment of the oldest and largest national nonprofit organization dedicated to developing and promoting policies and programs to protect U.S. children and strengthen U.S. families.
1920	The Dawson Report published in England	Considered to be one of the founding documents in the history of epidemiology by introducing the concept of primary care. Lord Dawson believed health care in a given district should be based on "primary health centers," which he termed "institutions equipped for services of curative and preventive medicine to be conducted by the general practitioners of that district." Difficult cases requiring special treatment would be referred to "secondary health centers." In addition, "secondary services" would be equipped for the care of infectious diseases and chronic ailments such as epilepsy or mental illness. "Teaching hospitals," with medical schools, would treat the most difficult illnesses and provide physician training.
1922	Pediatric Section of AMA approves the Sheppard–Towner Act while the AMA House of Delegates condemns it.	The Sheppard–Towner Act supported federal funding to states for health services for mothers and children. The AMA opposed it, calling it "an imported socialist scheme." Following this disagreement, AMA leadership promptly ruled that no section could independently adopt a resolution. This ruling was described as, "sterilizing the Pediatric Section [that] made the American Academy of Pediatrics inevitable".
1930	White House Conference on Child Health and Protection	Children's Charter was established, outlining a set of 19 principles in the areas of education, child

		labor, vocational training, recreation, family welfare, health, growth and development, disabled children, and related areas, including the Bill of Rights for the Handicapped Child.
1930	American Academy of Pediatrics established	On June 23–24, 1930, at Detroit's Harper Hospital, 35 pediatricians drew up the constitution and by-laws for AAP. The purposes of the AAP were: 1. To create reciprocal and friendly relations with all professional and lay organizations that are interested in the health and protection of children. 2. To foster and encourage pediatric investigation, both clinically and in the laboratory, by individuals and groups.
1933	Roger I. Lee, Lewis Webster Jones, and Barbara Jones, *The Fundamentals of Good Medical Care*, a publication of the Committee on the Costs of Medical Care	Defined quality of care as follows: "Good medical care is the kind of medicine practiced and taught by the recognized leaders of the medical profession at a given time or period of social, cultural, and professional development in a community or population group."Identified eight articles of faith underlining the concept of good care: 1. Good medical care is limited to the practice of rational medicine based on the medical sciences. 2. Good medical care emphasizes prevention. 3. Good medical care requires intelligent cooperation between the lay public and the practitioners of scientific medicine. 4. Good medical care treats the individual as a whole. 5. Good medical care maintains a close and continuing personal relationship between physician and patient. 6. Good medical care is coordinated with social welfare work. 7. Good medical care coordinates all types of medical services. 8. Good medical care implies the application of all the necessary services of modern scientific medicine to the needs of all the people.
1935	L. Kanner, *Child Psychiatry*	First English-language textbook on child psychiatry.
1938	Fair Labor Standards Act signed into law by President Franklin Delano Roosevelt	Prohibited interstate shipment of goods made in firms that employ children younger than age 16 years or children younger than age 18 years in hazardous occupations.
1946	National Mental Health Act	Authorized the Surgeon General to improve the mental health of U.S. citizens through research into the causes, diagnosis, and treatment of psychiatric disorders.
1949	National Institute of Mental Health (NIMH) is established	One of 27 institutes and centers that make up the National Institutes of Health. It is the largest

		research organization dedicated to mental health in the world.
1950	Mid-Century Conference on Children and Youth held	Prepared the Pledge to Children, which summarizes commitments by adults to children to promote education, strengthen family life, safeguard civil equality, improve workplace conditions, enhance democratic citizenship and institutions, and advocate for the development of the child’s individual talents and potential.
1953	American Academy of Child and Adolescent Psychiatry established	The leading national professional medical association, currently composed of more than 6,500 child and adolescent psychiatrists and other interested physicians dedicated to treating and improving the quality of life for children, adolescents, and families affected by these disorders.
1955	Mental Health Study Act	Authorized NIMH to study and make recommendations on mental health and mental illness in the United States. The act also authorized the creation of the Joint Commission on Mental Illness and Health.
1958	Community-based epidemiologic survey of mental health problems among those living in a northeastern U.S. community	First wide-scale epidemiologic study of problem behaviors among a U.S. community-based sample. Using systematic assessment methods for problem behaviors, investigators identified high rates of fear and worries among children aged 6–12 years in a general population sample, suggesting that symptoms are in part developmental. Established the feasibility of child self-report of emotional and behavioral symptoms. Used multiple-informant data and reported discrepancies in mother and child reports for some symptoms.
1959	Board certification in child and adolescent psychiatry offered for the first time by the American Board of Psychiatry and Neurology	National application of standards for clinical knowledge and skills in child and adolescent psychiatry.
1961	Kerr White and colleagues, “The Ecology of Medical Care,” published in *New England Journal of Medicine*	Seminal article that promoted the development of health services research. The authors “observed that traditional indexes of public health—including mortality and morbidity—do not adequately describe the actions of individual patients and doctors.” They argued that doctors should combine research of pathologic processes with exploration of the use of medical care resources, their relative availability in different contexts, and the patient’s decision-making processes. In their view, “the patient may be a more relevant primary unit of

		observation than the disease, the visit, or the admission.” Findings from this study examining the relationship between U.S. and U.K. populations and the experience of illness and different levels of medical treatment suggested that health care delivery and the training of physicians did not bear any logical relationship to the actual experience of illness in a given population. These data suggested that greater attention should be devoted to primary, continuing medical care as opposed to episodes of hospitalization or consultation of specialists. Furthermore, medical education and research should take greater account of the environment of health care and consider more systematically how health care resources should be deployed.
1964–1975	Isle of Wight studies	Extensive longitudinal cohort study over 5 years of intellectual and educational retardation, psychiatric disorder, and physical handicap of the entire population (estimated *N* = 3,500) of children aged 9–11 years. Screening for neuroepileptic disorders among entire population of 5- to 15-year-olds. Survey of 10-year-old children to directly compare to children of the same age range living in an inner borough of London. Additional studies examining the mental health profile of young children aged 5–7 years. Research methods included: 1. Multiple-informant data (child, parent, teacher, school records, psychiatric records, and pediatric records). 2. Two-stage screening strategy. 3. Comprehensive battery of assessments. 4. Identification of psychiatric diagnoses using social impairment as a criterion for severity and symptom pattern as a criterion of type. 5. Longitudinal design to study the age of onset, identification of risk factors, and description of course of illness over time. 6. Causal hypotheses of risk factors (e.g., organic brain dysfunction, family discord, reading difficulties). 7. Standardized neurodevelopmental assessments used to relate neurologic findings to psychiatric disorders. 8. Examination of differences in participants and nonparticipants.
1966	Avedis Donabedian, “Evaluating the Quality of Medical Care,” published in *Milbank Memorial Quarterly*	Proposed conceptual model for the evaluation of quality of care that remains widely used in health services research.

		Identified three main domains for examining quality: (1) structure, (2) process (technical, interpersonal), and (3) outcomes.
1969	Joint Commission on Mental Health for Children	Raised public awareness of high unmet need for child mental health services and overuse of mental health care delivered in restrictive treatment settings. Recommended establishment of a child advocacy system to coordinate federal, state, and local actions, development of community services focused on prevention and remediation, and delivery of child services in settings resembling normal living conditions. Identified need for research examining the effectiveness of child mental health services.
1974	Large epidemiologic study of 2,034 youth ages 6–18 years in New York	Included a sample of 1,000 youth in the welfare sector and 1,034 randomly selected youth. Prevalence estimates of psychiatric disorders were higher among children in the welfare sector than in the general population (21.0% vs. 17.5%). Documented high unmet need for services, such that <50% of children with serious impairment were referred for care.
1975	Education for All Handicapped Children Act passed (Public Law 94–142)	Federal law whose purpose is to ensure that all children with disabilities have access to a free appropriate public education that emphasizes special education and related services designed to meet their unique needs; to ensure that the rights of children with disabilities and their parents or guardians are protected; to assist states and localities to provide for the education of all children with disabilities; and to assess and ensure the effectiveness of efforts to educate children with disabilities.
1977	President’s Commission on Mental Illness and Health	Identified need for studies of the incidence and prevalence of mental disorders in the United States. Recommended an increase in the provision of child mental health services and the capacity to provide a continuum of care to children.
1979	NIMH contracts to develop standardized assessment methods of children and adolescents in general population samples	Stimulated a wave of epidemiologic studies examining the reliability and validity of checklists, rating scales, and diagnostic interviews in geographically diverse populations in the 1980s to 1990s. Promoted the development of assessment tools for use in community-based populations, such as the Child Behavior Checklist, Clinical Global Assessment Scale, Child and Adolescent Psychiatric Assessment, Diagnostic Interview for

		Children, and Diagnostic Interview Schedule for Children.
1980	NIMH Epidemiologic Catchment Area Study	First large-scale, multisite epidemiologic study of psychiatric disorders among 20,291 adults in U.S. households and institutions. Stimulated further development of the Diagnostic Interview Schedule and provided prevalence estimates of treated and untreated psychiatric disorders in the community.
1981	Select Panel for the Promotion of Child Health	Acknowledges significant improvement in services to children with disabilities (including serious emotional disturbances) but identifies wide variation in the availability of child-related services. Developed national recommendations to improve the delivery of services to children, including greater access to care, expansion of services to include prevention and early detection and treatment of developmental problems, improved quality of care in residential treatment settings, enhanced community support for child programs, and greater involvement of families in the delivery of mental health services.
1982	Children's Defense Fund report, *Unclaimed Children*	Raised public awareness of high unmet need for child mental health services in the United States. Identified children with severe emotional and behavioral problems as "unclaimed" by the public agencies responsible for serving them.
1984	Child and Adolescent Service System Program (CASSP) established	NIMH-funded program to support states in the development of interagency efforts to improve the system of care for children with mental health problems. Provided technical assistance to community-based programs in the adaptation of systems-of-care values and principles to the delivery of services in children with severe emotional and behavioral problems.
1986	Beth A. Stroul and Robert M. Friedman, *A System of Care for Severely Emotionally Disturbed Children and Youth*	Pivotal document from the CASSP Technical Assistance Center, Georgetown University Child Development Center, in Washington, D.C., that articulated the core values and guiding principles for the system-of-care philosophy for children.
1989	Agency for Health Care Policy and Research established (later renamed Agency for Healthcare Research and Quality)	The health services research arm of the U.S. Department of Health and Human Services (HHS). Supports research in the following areas: 1. Quality improvement and patient safety 2. Outcomes and effectiveness of care 3. Clinical practice and technology assessment 4. Health care organization and delivery systems 5. Primary care

		6. Health care costs and sources of payment
1989	System of Care for Severely Emotionally Disturbed Children and Youth	Introduced a philosophy underlying a system of care for children with severe emotional and behavioral problems. Identified core values for system of care for children: Care should be child centered and community based. Outlined 10 guiding principles for system of care.
1989–1995	Fort Bragg Demonstration Project	Large-scale effectiveness study of providing a continuum of services to children and their families with four components: A longitudinal cohort study comparing outcomes and costs of providing a model continuum of care vs. traditional mental health services. Provision of a continuum of care was related to the following: 1. Improved access to care. 2. Higher adolescent and parent satisfaction with services. 3. Greater costs. 4. Reduced rates of hospitalization or residential treatment. Continuum of care did *not* produce better clinical outcomes than traditional mental health services.
1989	Institute of Medicine report on Research on Children and Adolescents with Mental, Behavioral, and Developmental Disorders	Comprehensive review by national experts of the dimensions, causes, determinants, and interventions for childhood psychiatric disorders, as well as child mental health research infrastructure. Raised public awareness of an estimated 12% of U.S. children having a diagnosable mental illness, advances in the assessment and treatment of childhood disorders, promising research directions, and resource needs for research.
1989	Agency for Health Care Policy and Research created.	Supported by the Omnibus Budget Reconciliation Act of 1989. Mission included “to support research designed to improve the quality, safety, efficiency, and effectiveness of health care for all Americans.” Activities included large multidisciplinary, multi-institutional projects that focused on patient results in certain medical conditions, development of practice guidelines, and participating in the development of a health care reform plan.
1990	National Plan for Research on Child and Adolescent Mental Disorders	Comprehensive 5-year plan of how NIMH intended to stimulate new basic and clinical research, promote research training, and provide leadership and coordination to stimulate advocacy and research dissemination in child mental health.

1990	The 1990s declared by the U.S. Congress the “Decade of the Brain”	Provided national public endorsement for federal, state, and community initiatives to support a wide range of mental health research, as well as initiatives to improve the quality of care for the mentally ill.
1990	The National Committee for Quality Assurance is created.	A large, nonprofit organization dedicated to improving health care quality through measurement, transparency, and accountability. Products include the Healthcare Effectiveness Data and Information Set (HEDIS), a set of quality measures used to report on quality of care at the health plan level.
1991–1996	Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA) Study	Pilot studies followed by multisite study of 1,285 child–caregiver pairs in four geographic areas using household interviews. Stimulated the development of standardized assessment methods of psychiatric diagnosis, impairment, mental health service use, risk factors, and administrative infrastructure for large-scale epidemiologic study. Produced prevalence estimates of psychiatric disorders with and without impairment criteria and mental health service use.
1992	Passage of the Comprehensive Community Mental Health Services Program for Children and Adolescents with Severe Emotional Disturbances	Authorized federal funding of more than $100 million/year for implementation of systems of care in 38 different communities. Supported the development of community-based continuum-of-care programs for children, including assessment, case management, outpatient treatment, day treatment, in-home services, respite care, therapeutic foster care, and group homes for children with severe emotional disturbances.
1992	American Academy of Pediatrics, Ad Hoc Task Force on Definition of the Medical Home	The guiding principles for the medical care of infants, children, and adolescents should be accessible, continuous, comprehensive, family-centered, coordinated, and compassionate. Comprehensive care encompasses the following services: 1. Provision of preventive care. 2. Ambulatory and inpatient care for acute illness 24 hours a day, 7 days a week. 3. Provision of care over an extended period of time to enhance continuity. 4. Identification of the need for subspecialty consultation and referral and knowing from whom and where these can be obtained. 5. Interaction with school and community agencies to be certain that special health needs of the individual child are addressed. 6. Maintenance of a central record and data base containing all pertinent medical information

		about hospitalizations. This record should be accessible, but confidentiality must be assured.
1997–1999	Development of practice parameters by the American Academy of Child and Adolescent Psychiatry	Twenty practice parameters were developed for the assessment and treatment of specific psychiatric disorders, as well as child mental health problems (e.g., suicide, sexual abuse).
1998	National Institutes of Health Consensus Development Conference on Diagnosis and Treatment of Attention Deficit Hyperactivity Disorder	Review the scientific evidence, open public discussion, and closed-panel deliberations by a 13-member multidisciplinary national expert panel. Findings from this conference support the following: 1. The validity of the diagnosis of attention-deficit/hyperactivity disorder (ADHD). 2. The substantial effect of ADHD on children, their families, and society. 3. Effectiveness of psychosocial and medication treatments. 4. Need for additional research on the long-term effects of stimulant medication treatment. 5. Need to improve early detection, access to care, and quality of evaluation and treatment. 6. Need for further basic science and clinical and intervention research.
1999	White House Conference on Mental Health	First major public introduction to mental illness in the United States and scientific advances in the assessment and treatment of psychiatric disorders. Called for a national antistigma campaign for persons with mental illness.
1999	Surgeon General's Call to Action to Prevent Suicide	Developed a national plan to increase awareness and prevent suicide in the United States.
1999	Surgeon General's Report on Mental Health	First Surgeon General's report dedicated to mental health and mental illness. Communicated three main messages: (1) Mental health is fundamental to health; (2) mental disorders are real health conditions; and (3) the efficacy of mental health treatments is well documented, and a range of treatments exists for most mental disorders. Included a chapter on children and mental health.
1999	Multimodal Treatment Study of Children with ADHD (MTA study)	First large-scale, multisite, randomized clinical trial of treatment of a childhood mental disorder. Compared the effectiveness of intensive behavioral treatment, tightly managed medication treatment, combined behavioral and medication treatment, and standard community care for ADHD. A randomly selected sample of 579 children ages 7.0–9.9 years recruited from six university-based clinic referral sources. No oversampling to include sufficient number of girls or children from minority race/ethnicity backgrounds to test treatment differences.

		The following results were found when using core ADHD symptom reduction as the outcome: 1. Medication management was superior to behavioral treatment and routine community care. 2. Combined treatment did not yield significantly more benefit than medication management alone.
1999	The Agency for Health Care Policy and Research is renamed the Agency for Healthcare Research and Quality	The agency was reauthorized under the Healthcare Research and Quality Act of 1999. Its mission is to produce evidence to make health care safer, higher quality, more accessible, equitable, and affordable. Development of practice guidelines was removed from the new mission, and it was made explicit that the agency is not a regulatory body.
1999	National Quality Forum created.	This nonprofit, membership-based organization was created in response to recommendations from the President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry. The main activities are: (1) convene working groups to foster quality improvement in both public- and private-sectors; (2) endorses consensus standards for performance measurement; (3) ensure that consistent, high-quality performance information is publicly available; and (4) seek real-time feedback to ensure measures are meaningful and accurate.
2000	White House Meeting on Child Mental Health	Major public meeting specifically addressing the need to improve the diagnosis and treatment of children with emotional and behavioral disorders.
2000	U.S. Surgeon General's Conference on Children's Mental Health, Developing a National Action Agenda	Developed recommendations for a National Action Agenda on Children's Mental Health based on feedback from a broad cross-section of more than 300 child mental health stakeholders
2000	National Institute of Mental Health and the Food and Drug Administration (FDA) Workshop on Psychopharmacology for Young Children, Clinical Needs and Research Opportunities	A nationally sponsored meeting of researchers, practitioners, ethicists, industry staff, and family and patient representatives, whose mission was to examine the current state of knowledge regarding psychopharmacology for young children and discuss a variety of topics relevant to research in this age group, including safety, efficacy, investigational methods, and ethical aspects. The meeting was in response to the increasing use of psychotropic medication among preschool-age children, for which the efficacy and safety of psychotropic medication have not been systematically evaluated. Recommendations included additional research in developmental psychopathology to clarify

		diagnosis and provide sensitive and specific methods for clinical trials.
2000	National Research Council and Institute of Medicine, *From Neurons to Neighborhoods: The Science of Early Childhood Development*	A 17-member committee of national experts convened to update scientific knowledge about the nature of early child development; disentangle knowledge from erroneous popular beliefs; and discuss the implications of this knowledge base for early childhood policy, practice, professional development, and research. With 11 national recommendations, this monograph advocated for improving the national capacity to meet the high unmet need for mental health services among preschool children and conducting research examining the effectiveness of early intervention programs for children.
2000	Clinical practice guidelines for ADHD developed by the American Academy of Pediatrics	National recommendations for the evaluation and diagnosis of ADHD, as well as treatment of the school-age child with ADHD in the primary care sector.
2000	Agency for Health Care Policy and Research evidence report, *Treatment of Attention-Deficit/Hyperactivity Disorder*	As part of the Evidence-Based Practice Program, a multidisciplinary research team conducted a systematic review of the scientific evidence for the treatment of ADHD to develop a base from which clinical guidelines, performance measures, and other quality improvement tools could be developed.
2000–2002	Development of quality indicators for childhood psychiatric disorders by the American Psychiatric Association Task Force on Quality Indicators for Children	Conceptualized the quality of care for children in five main domains: Prevention, access, quality/process/appropriateness, satisfaction/perceptions of care, and outcomes/effectiveness. Developed multiple recommendations for quality care and corresponding sample indicators.
2001	Surgeon General's Conference on Children's Mental Health, *A National Action Agenda*	Made the following recommendations: 1. Promote public awareness of children's mental health issues and reduce the stigma associated with mental illness. 2. Continue to develop, disseminate, and implement scientifically proven prevention and treatment services in the field of children's mental health. 3. Improve the assessment and recognition of mental health needs in children. 4. Eliminate racial/ethnic and socioeconomic disparities in access to mental health care services. 5. Improve the infrastructure for children's mental health services, including support for scientifically proven interventions across professions.

		6. Increase access to and coordination of quality mental health care services. 7. Train front-line providers to recognize and manage mental health issues and educate mental health care providers about scientifically proven prevention and treatment services. 8. Monitor the access to and coordination of quality mental health care services.
2001	National Advisory Mental Health Council's Workgroup on Child and Adolescent Mental Health Intervention Development and Deployment, *Blueprint for Change: Research on Child and Adolescent Mental Health*	A 21-member work group of national experts reviewed major research findings on child and adolescent psychiatric disorders over the prior decade and identified three major research priority areas: 1. Basic science and the development of new interventions. 2. Intervention development, moving from efficacy to effectiveness. 3. Intervention deployment, moving from effectiveness to dissemination.
2001	Institute of Medicine's Committee on Quality of Health Care in America, *Crossing the Quality Chasm*	Proposed six aims for improving the quality of care for the twenty-first century: Health care should be the following: 1. Safe: Avoiding injuries to patients from the care that is intended to help them. 2. Effective: Providing services based on scientific knowledge to all who could benefit and refraining from providing service to those not likely to benefit (avoiding underuse and overuse, respectively). 3. Patient-centered: Providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions. 4. Timely: Reducing waits and sometimes harmful delays for both those who receive and those who give care. 5. Efficient: Avoiding waste, including waste of equipment, supplies, ideas, and energy. 6. Equitable: Providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status.
2002	President's New Freedom Commission on Mental Health	National commission, whose mission was to conduct a comprehensive study of the U.S. mental health service delivery system, including both the private and public sector providers, and advise the president on methods of improving the system. The main goals were the following: 1. Review the current quality and effectiveness of public and private providers and federal, state,

		and local government involvement in the delivery of services to individuals with serious mental illnesses and children with serious emotional disturbances and identify unmet needs and barriers to services. 2. Identify innovative mental health treatments, services, and technologies that are demonstratively effective and can be widely replicated in different settings. 3. Formulate policy options that could be implemented by public and private providers and federal, state, and local governments to integrate the use of effective treatments and services, improve coordination among service providers, and improve community integration for adults with serious mental illnesses and children with serious emotional disturbances.
2002	Agency for Healthcare Research and Quality, *Systems to Rate the Strength of Scientific Evidence*	Identified 19 generic systems that fully address key quality domains for a particular type of study and 7 systems that address all three quality domains grading the strength of a body of evidence. These systems may be useful to the development of clinical practice guidelines or other health-related policy advice.
2002	International Consensus Statement on ADHD	A consensus statement developed by a consortium of international scientists that supports the validity of ADHD and its adverse effect on the lives of children and their families using scientific evidence. The statement was developed as a reference to the status of the scientific findings about ADHD to dispel periodic public perception that ADHD is not a real medical condition.
2003	President’s New Freedom Commission on Mental Health: Transforming Mental Health Care in America	Recommendations for Transformed Mental Health System: 1. Americans understand that mental health is essential to overall health. 2. Mental health care is consumer and family driven. 3. Disparities in mental health services are eliminated. 4. Early mental health screening, assessment, and referral to services are common practices. 5. Excellent mental health care is delivered, and research is accelerated. 6. Technology is used to access mental health care and information.
2004	U.S. Food and Drug Administration (FDA) black box warning on pediatric antidepressants	The FDA directed manufacturers of all antidepressant medication to add a boxed warning that reported (1) antidepressants increase the risk of suicidal thinking and behavior in children and adolescents with major depression and other psychiatric disorders, (2)

		anyone considering the use of an antidepressant in a child or adolescent for any clinical use must balance the risk of increased suicidality with the clinical need, (3) patients who are started on therapy should be observed closely for clinical worsening, suicidality, or unusual changes in behavior, and (4) disclose whether the particular drug is approved for a pediatric indication(s) and, if so, which one(s).
2005	Caring for California Initiative	First statewide study of the quality of care for children with ADHD, major depression, and conduct disorder in publicly funded outpatient specialty mental health programs.
2006	Institute of Medicine's (IOM) Committee on Quality of Health Care in America, *Improving the Quality of Health Care for Mental and Substance Use Conditions*	Recommended that the prior IOM health care system redesign strategy was applicable to mental and substance use health care but required some adaptation to address greater obstacles to patient-centered care, a weaker quality measurement and improvement infrastructure, more separated care-delivery arrangements, less use of information technology, a more educationally diverse workforce with a greater use of solo practice, and a differently structured marketplace.
2006	National Advisory Mental Health Council's Workgroup on Services and Clinical Epidemiology Research, *The Road Ahead: Research Partnerships to Transform Services*	Six issues driving recommendations shaping the future research agenda to transform mental health services are (1) partnerships in research, (2) quality of care, (3) fairness, (4) recovery, (5) communication, and (6) ongoing evaluation.
2007	Treatment for Adolescents with Depression Study (TADS) 36-week outcomes	Among 327 patients from 13 academic and community sites randomized to combination therapy, fluoxetine therapy, and cognitive behavior, combined treatment appears to be superior to monotherapy using symptom reduction and overall clinical improvement rating as outcomes.
2007	National Stigma Study-Children	First large-scale, nationally representative survey of public knowledge, attitudes, and beliefs on children's mental health reported substantial rates of stigma (35–57%) and negative perceptions of medication treatment (52–85%).
2008	Mental Health Parity and Addiction Equity Act	A federal law that generally prevents group health plans and health insurance issuers that provide mental health or substance use disorder benefits from imposing less favorable benefit limitations on those benefits than on medical/surgical benefits.
2009	The Children's Health Insurance Program Reauthorization Act	Mandated the identification, refinement, and development of child health care measures for voluntary use in Medicaid and Children's Health

		Insurance Programs. Called for submission of initial core set of 24 quality measures to Department of Health and Human Services by 2010. Funded the 5-year Pediatric Quality Measures Program. Required annual review of the national Core Set starting in 2013.
2010	Patient Protection and Affordable Care Act	Landmark legislation for health care reform. Mandated development of a National Quality Strategy and earmarked support for the Patient-Centered Outcomes Research Institute.
2018	Bipartisan Budget Act	Extended 4-year funding to the Children's Health Insurance Program. Requires all states to report adherence to quality measures in the national Child Core Set starting in 2024.
2019	Wit v. United Behavioral Health	A federal court in Northern California found that one of the largest managed behavioral health care companies rejected the insurance claims of tens of thousands of people seeking mental health and substance use disorder treatment based on defective medical review criteria. This case found evidence of a separate and unequal system of medical care review that is inconsistent with the treatment of physical health conditions, such as cancer or heart disease.
2020	The Coronavirus Aid, Relief, and Economic Security (CARES) Act	A $2 trillion stimulus package which includes $425 million of appropriations to the Substance Abuse and Mental Health Services Administration to respond to the pandemic, $250 million going to new funding for Certified Community Behavioral Health Clinic Expansion grants, $100 million for emergency response activities, and $50 million for suicide prevention.

Scientific advances in the assessment and treatment of ADHD also laid the foundation for conducting the first large-scale, randomized clinical trial of two treatment modalities. The Multimodal Treatment Study of Children with ADHD, or MTA study, examined the effectiveness of highly standardized psychosocial, medication, and combined treatment protocols for ADHD. Based on the use of reduction of ADHD symptoms as the main outcome at 14 months, medication management was found to be superior to intensive psychosocial treatment and usual community care, and combined psychosocial treatment was found to be of no significantly greater benefit than medication management alone. Findings from the MTA study, however, should not be interpreted as conclusive that psychosocial interventions are inferior to medication treatment. Instead, alternative analyses from this rich data set have illuminated key methodologic issues, such as the selection of outcome measures, the use of composite versus individual outcomes as indicators of effectiveness, and the effect of moderators (e.g., sociodemographic factors, clinical severity, parental education, and income) and mediators (e.g., treatment adherence) in the assessment of an intervention's effectiveness. Nevertheless, treatment effects were found not to be sustained. At 36 months, the treatment groups did not significantly differ on any outcome measure. In addition, based on saliva assays for methylphenidate as an objective indicator of medication adherence, only 53.5% of children were consistently adherent to stimulant medication treatment during the initial 14-month treatment period. Further, the findings are not generalizable to community-based populations.

In the early twenty-first century, a report from the Institute of Medicine's (IOM) Committee on Quality of Health Care in America raised national awareness that the "American health care delivery system was in need of fundamental change." The wide difference between the care that should be received and the care that was actually being received was described as the quality chasm. A follow-up report found that the IOM's recommendations to transform the health care system were generally applicable to mental and substance use disorders but also identified distinctive characteristics of health care for mental and substance use conditions. Harold Pincus and colleagues noted that "the [Committee's] fundamental conclusion [is] that coordination and collaboration with the rest of medicine are essential (and vice versa), but it is also critical to ensure access to the depth and expertise provided by psychiatrist and other mental health and substance use specialists."

Table 54.10–2.
Core Values of Children's System of Care and Pediatric Medical Home

Children's System of Care	Pediatric Medical Home
Mental Health Care	Medical Care
Access to a comprehensive array of services	Accessible
Smooth transitions to the adult service system	Continuous
Early identification and intervention	Comprehensive
Full participation of families and surrogate families	Family-centered
Integrated and coordinated care across child-serving agencies Inclusion of case management	Coordinated
Individualized services Least restrictive, most normative environment Protection of child rights and effective advocacy efforts Equitable and culturally sensitive service delivery	Compassionate

Within this same time period, national professional organizations developed practice parameters and treatment guidelines that enabled child mental health services researchers to develop quality indicators based on evidence of effectiveness and clinical consensus. Using documented adherence to quality indicators in the medical record to estimate quality of care, from 1998 to 1999, the Caring for California Initiative (CCI) was the first statewide quality-of-care study for children with common psychiatric disorders in publicly funded outpatient mental health programs. The CCI found that on average less than one-half (48.5%) of medical records passed all eligibility quality domains at the probable acceptable care level, a rate similar to general health care among U.S. adult and child populations. Medication safety monitoring was particularly weak, such that roughly one-fourth (28.5%) of the records of children receiving psychotropic medication treatment had documented monitoring of at least one clinically indicated vital sign or laboratory study. National concerns regarding the safety of psychotropic medication were also raised during this time period. In 2004, the U.S. Food and Drug Administration (FDA) issued a black box warning on all antidepressant medication to alert health care providers of an increased risk of suicidality in children and adolescents. Further, high national rates of stigma related to medication treatment raised awareness of the need for further examination of the mechanisms underlying engagement, uptake, and adherence to recommended care for children with psychiatric disorders and their families.

Although improving the quality of child mental health care has been a long-standing national priority, two federal health policies in the early 21st century legislatively mandated the development and use of quality measures for child health and mental health. In 2009, the Children's Health Insurance Program Reauthorization Act (CHIPRA) was signed. This law mandated the identification, refinement, and development of child health care measures for voluntary use in Medicaid and the Children's Health Insurance Program (CHIP). It called for submission of initial core set of 24 quality measures to the Department of Health and Human Services by 2010, with annual updates starting in 2013. This legislation also funded a 4-year Pediatric Quality Measures Program (2010–2014). Within this program, seven centers of excellence were funded to develop new measures and refine existing ones for “potential core enhancements” as well as ten demonstration programs to develop, test and evaluate, and/or report on adherence to quality measures.

During the following year, the landmark Patient Protection and Affordable Care Act was signed in 2010. Within this complex legislation, the development of the National Quality Strategy was mandated. This strategy to transform health care had three aims: (1) better care; (2) healthy people and healthy communities; and (3) affordable care. The six priority areas were safety, person or family engaged as partners in care, effective communication, and coordination of care, effective prevention and treatment, work with communities to promote wide use of best practices, and make quality of care affordable. In 2022, the National Quality Strategy was substantially revised to prioritize a person-centric approach from birth to end of life as individuals journey across the continuum of care. This legislation also funded the Patient-Centered Outcomes Research Institute (PCORI) to support research that informs better choices among alternative treatment strategies, supports a patient-centered orientation, and directs attention to individual and system differences that may influence strategies and outcomes. As stipulated by the 2018 Bipartisan Budget Act, which included a 4-year funding extension to CHIP, state reporting of quality measures from the Child Core Set will be mandatory starting in 2024. In response, criteria for addition and removal of core set measures have been revised to require a large proportion of states have the capacity to report adherence rates on a national quality measure.

ACCESS TO CARE

Significant gaps continue to exist between clinical needs and mental health service use for children with psychiatric disorders in the community. A number of conceptual models were developed to explain differential access to care. Of note, these models attempted to identify factors that can be changed through quality improvement interventions (referred to as *mutable factors*) to improve the likelihood of contact with services among children meriting assessment and treatment. A brief review of historically relevant access models is provided, because their frameworks continue to inform current quality improvement efforts.

Conceptual Models of Access to Care

Donabedian’s early model of the medical care process acknowledges the complex interactions between patient and provider behaviors, as well as the influence of contextual factors, such as the characteristics of the settings (e.g., formal and informal treatment settings and social, economic, and cultural factors) that take place in the process of seeking and receiving care. From the patient’s perspective, service use and, ultimately, health outcomes are initiated by the recognition of need. This step is followed by the decision to look for care and the processes of seeking care, assuming the sick role, and maintaining and terminating care. On seeking care, provider behaviors also become influential and include diagnosis, therapeutic decision making, and potential collaboration with other providers during these processes.

Subsequent conceptual models further specify factors that are influential to the access to care. Ronald Andersen’s Behavioral Model of Health Services Use, for example, proposes that predisposing (e.g., sociodemographic factors such as gender), enabling (e.g., availability of insurance or treatment resources), and need (e.g., illness severity) factors form the basis of the decision-making process. Alternatively, the Health Belief Model suggests that individual psychological attributes, such as attitudes, beliefs, and perceived benefits of treatment, represent the major determinants of seeking care. Other models of access, such as Bernice Pescosolido’s Family Network-Based Model of Access to Child Mental Health Services, emphasize the role of social networks in this decision-making process.

Overall, help-seeking pathway models propose a series of links between the initial recognition of a mental health problem and the eventual use of mental health services, framing these steps within the context of an “illness career” and network experiences thought to influence career progression as well as adherence to professional treatment suggestions. Few studies, however, have implemented study designs complex enough to analyze network influences, temporal sequencing, and combined multisector care patterns for youth. Studies addressing access to care have been most commonly based on the Andersen model, using cross-sectional designs to identify relevant predictors.

Stigma as Barrier to Access

The 1999 Surgeon General's Report on Mental Health had the foresight to identify stigma as a significant barrier to access and continuity of mental health care. Youth with mental disorders face significant stigma in various forms, including public (expressed as prejudice and discrimination), courtesy (stigma suffered by family members or associates of individuals with mental disorders), and self-stigma (stigmatized individual's acceptance of negative views by others). Successful recovery from mental illness requires awareness of stigma and its potential adverse consequences and effective means to combat it.

To examine factors that shape prejudice and discrimination toward children with mental disorders, the National Stigma Study-Children (NSS-C) used vignettes depicting several stigmatizing scenarios, including ADHD, depression, asthma, or "normal troubles." Results indicate that scenarios of ADHD and depression and those depicting adolescents elicited the highest rejection responses. Respondents who thought a child suffered from a mental illness rather than from asthma or "normal" troubles indicated that they would seek social distance from the affected child, apparently resulting from perceptions that a child with mental illness is dangerous. In a comparison of child and adult depression vignettes, more respondents evaluated childhood depression as serious (83%) than adult depression (51%) and endorsed treatment of all types, including coerced care, for children with depression.

However, high stigma surrounding childhood disorders poses unique access barriers for youth and their families and merits further study to identify ways to combat these harmful perceptions. The NSS-C data were collected as part of the General Social Survey in 2002, and despite this historical limitation, still represent relevant generalizable insight on stigma perceptions relevant to youth. There has been some increased willingness to disclose mental health problems to friends and family in recent years, and public media disclosures by celebrities of their mental health struggles, yet, there is little indication that overall stigma perceptions surrounding mental disorders have lessened, considering that public perceptions studied across 22 years changed toward increasing perceptions of potential dangerousness and more support for coercive treatments. Earlier general societal antistigma programs relied on three pillars: protest, education, and contact. Subsequent formal evaluation of these traditional programs challenged the effectiveness of the hypothesized change agents and newer antistigma interventions consider new paradigms, including inclusion of patients, family, and friends in intervention development. For example, a recently developed college program seeks to create/influence new generations of medical, social, and political leaders during their formative years of college, by fostering safe and stigma-free zones on campus. The program uses a toolbox approach developed in partnership with multiple stakeholders, including the students, based on intervention principles such as outreaching early, an "by students for students" approach, and using "bait and flip" efforts to attract campus-wide engagement. Early results show encouraging reductions in stigmatizing behaviors and prejudice, as well as improved campus mental health culture perceptions using this intervention approach.

QUALITY OF CARE

The Institute of Medicine defined quality as "The degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge." The overarching goal is to improve the alignment between clinical needs, delivery of recommended care, and health policies to sustain quality care. In lay terms, the goal is to deliver the right care, at the right time, at the right place, every time. This simple but powerful statement should also include "and for everyone," to underscore the imperative to reduce disparities in child mental health care that may be confounded by structural racism and social determinants of health.

Models of Quality of Care

Components of quality of care are often conceptualized as structure, process, and outcomes, following Donabedian's model of quality of care. Structural factors include system and clinic organizational factors, as well as provider, child, and family characteristics. Examples are health plan benefit structure, care sector, clinic volume, provider caseload, provider training and discipline, and parent education, as well as child insurance coverage, clinical profile, and sociodemographic characteristics. Processes of care are dichotomized into technical and interpersonal aspects of care. A common technical care process is medication treatment, which may, for example, be assessed using quality indicators or criteria pertaining to medication evaluation, use, and monitoring. For children, variables within the domain of interpersonal style may include the assessment of aspects of parent-provider communication and shared decision making. For example, how well a child and/or parent believes their concerns are listened to or how much involvement the primary caregiver had in the decision process of making the diagnosis, developing a treatment plan, and participating in the care of their child over time. Outcomes may also include child-, parent-, family-, community-, and system-level characteristics. Outcomes for children, parents, and families include symptom reduction, improved functioning, child and parent satisfaction, reduced parent strain, and improved family cohesion. Examples of community- and system-level outcomes are improved school attendance and reduced costs, respectively.

Building on the Donabedian quality of care framework, McGlynn and colleagues proposed a more transactional approach to examining quality of care for mental health. The ordering of the components of quality of care is reversed, such that outcomes influence both care processes and structural factors and processes of care directly affect change in structural characteristics. They also suggested that in examining quality of care for mental health problems, an analytical strategy be used that “simultaneously examines the outcomes, the processes associated with producing those outcomes, and the structural context in which the process was delivered.”

However, with either of these conceptual models, interpretation of data examining the relationship between care processes and clinical outcomes should be done carefully. A study sample may underrepresent children from race or ethnic minority backgrounds, reducing statistical power to examine disparities. A child’s clinical severity may drive more intensive service use (e.g., repeat emergency department visits) while a child’s psychosocial complexity (e.g., poverty, parental unemployment) may hinder access or continuity of outpatient care. Missing data is also a common problem. In the community, clinic follow-up is often poor. If the data source only includes children who remain in care and have a standardized measure of a clinical outcome over time (e.g., EHR), it is difficult to determine if a child who drops out of care no longer clinically needs care or not.

Quality of Care for Child Mental Health

Psychotropic Medication Prescription Trends. Research examining national trends in psychotropic medication use, based on secondary databases (i.e., pharmacy claims), and how rates vary by sociodemographic and clinical characteristics provides important monitoring of potential overuse of psychotropic medication among children. Among commercially insured young children (2 to 7 years) from 2007 to 2017, annual antipsychotic use was 0.27% in 2007, peaked at 0.29% in 2009, and statistically significantly declined to 0.17% by 2017. Antipsychotic use was higher in boys than in girls, and the most common diagnoses were pervasive developmental disorder, conduct or disruptive behavior disorder, and ADHD. Although use of antipsychotic medication declined from 2009 to 2017, a majority of use remained off-label and for conditions lacking effectiveness and safety data. In addition, metabolic monitoring is important for children taking antipsychotic medication, given the risk for increased body mass index, impaired glucose metabolism, and hyperlipidemia. Among children enrolled in Medicaid in two states, 61% of children saw more than one provider type and children seeing primary care and specialty mental health provider with prescribing privileges were significantly more likely to receive metabolic monitoring compared to children seen only by a primary care provider. Together, these findings support the need for improved medication safety monitoring and quality improvement interventions, such as shared care arrangements, for children receiving antipsychotic medication treatment.

Gaps between Recommended and Usual Care. Findings from studies examining care processes for child psychiatric disorders suggest that gaps exist between recommended and usual care. Medical record data from a statewide sample of children receiving care for common psychiatric disorders in publicly funded outpatient mental health programs documented that adherence to quality of care indicators for clinical assessment was high (78% to 95%) but adherence to quality indicators related to service linkage, parental involvement, use of evidence-based psychosocial treatment, and patient protection were moderate to poor (8.0% to 74.1%). Medication safety monitoring was particularly weak, such that roughly one-fourth (28.5%) of the records of children receiving psychotropic medication treatment documented the monitoring of at least one clinically indicated vital sign or laboratory study. Using child-level data from multiple data sources, quality of care for ADHD in primary care and specialty mental health clinics in a large managed care Medicaid program was examined using a longitudinal cohort study design. Continuity of care was poor, and in primary care children received on average only one to two follow-up visits a year. In the specialty mental health clinics, less than one-third of children received any stimulant medication, and in both sectors, stimulant medication refill prescription persistence was poor (31% to 49%). With few exceptions, ADHD diagnosis, impairment, academic achievement, parent distress, parent-reported treatment satisfaction, perceived benefit, and improved family functioning did not differ between children who remained in care and those that received no care.

Table 54.10–3.
National Child Behavioral Health Quality Measures in 2021 Child Core Set

NQF #	Steward	Measure Description	Data Source(s)
0108	NCQA	Follow-up care for children (6–12 years) prescribed ADHD medication Percentage of children newly prescribed ADHD medication who had at least three follow-up care visits within a 10-month period, one of which was within 30 days of when the first ADHD medication was dispensed. Two rates are reported Initiation phase: Percentage of children ages 6 to 12 as of the index prescription start date (IPSD) with an ambulatory prescription dispensed for ADHD medication, who had one follow-up visit with practitioner with prescribing authority during the 30-day initiation phase Continuation and maintenance (C&M) phase: Percentage of children ages 6 to 12 as of the IPSD with an ambulatory prescription dispensed for ADHD medication who remained on the medication for at least 210 days and who, in addition to the visit in the initiation phase, had at least two follow-up visits with a practitioner within 270 days (9 months) after the initiation phase ended	Administrative EHR
0576	NCQA	Follow-up after hospitalization for mental illness Percentage of discharges for children ages 6 to 17 who were hospitalized for treatment of selected mental illness or intentional self-harm diagnoses and who had a follow-up visit with a mental health practitioner. Two rates are reported: Percentage of discharges for which the child received follow-up within 30 days after discharge Percentage of discharges for which the child received follow-up within 7 days after discharge	Administrative
2801	NCQA	Use of first-line psychosocial care for children and adolescents on antipsychotics Percentage of children and adolescents ages 1 to 17 who had a new prescription for an antipsychotic medication and had documentation of	Administrative

		psychosocial care as first-line treatment	
n/a	NCQA	Use of multiple concurrent antipsychotics in children and adolescents Percentage of children and adolescents ages 1 to 17 who were treated with antipsychotic medications and who were on two or more concurrent antipsychotic medications for at least 90 consecutive days during the measurement year	Administrative

NQF, National Quality Forum; NCQA, National Committee on Quality Assurance; ADHD, Attention Deficit/Hyperactivity Disorder; EHR, electronic health record; n/a, not applicable.

National Quality Measures. In response to national recommendations to improve the quality of mental health care, there has been a proliferation of quality measures, predominantly focused on adults. Nevertheless, the 2021 Child Core Set of quality measures for beneficiaries in Medicaid and CHIP, includes four measures related to child mental health. The measures monitor follow-up care for children prescribed ADHD medication, timeliness of follow-up outpatient care(at 7 and 30 days) after hospitalization for a mental illness, metabolic monitoring for children receiving antipsychotic medication, and use of first-line psychosocial care for children on antipsychotic medication. The operational definitions for the measures are summarized in Table 54.10–3. Across measures, clinical validity is scarce, state Medicaid agency adherence rates vary widely, and some states have challenges accessing or linking the data required to publicly report measure adherence rates. Starting in 2024, Medicaid and CHIP agencies will be required to report on all Child Core Set measures and the behavioral health measures in the Adult Core Set, stimulating greater urgency to address limitations in existing data sources and methods.

Future research is needed to examine whether adherence to national child mental health quality measures is associated with clinical outcomes that are meaningful, which likely vary by stakeholder perspective (e.g., satisfaction, guideline-concordant care, cost, value). When examining quality of care, it is imperative to examine variation in quality by sociodemographic and clinical characteristics to identify disparities as well as to explore target mechanisms driving inequities. However to advance this line of research, greater public investment is needed in data infrastructure, such as improved data coordination across large electronic health care record data bases using common data elements, better clinician decision support tools, and greater integration of standardized clinical outcome measures using patient portals to reduce selection bias toward having outcome data from only patients who remain in care.

Measurement-based Care. MBC is conceptualized as a driver for quality improvement. However, given the breadth of MBC types, it is useful to conceptualize these by categorizing them by population levels which are individual (Tier 1), clinical aggregate (Tier 2), and national levels (Tier 3). At the individual level (Tier 1), the main purpose of MBC is to inform clinical decision-making. MBC is conceptualized as routinely collecting patient-reported outcome measures, such as standardized symptom rating scales. The target mechanisms for change include improving early detection of treatment resistance and facilitating patient–provider communication. During psychotherapy, routine outcome monitoring may be especially effective to identify patients who are not clinically improving and improve fit between evidence-based therapies to clinical needs.

At the clinical aggregate level (Tier 2), the main purpose of MBC is to monitor performance. The proposed mechanism of change is that feedback on variation in clinical outcomes by provider, clinic, hospital, or agency will identify target areas for quality improvement and guide resource allocation decisions. Tier 2 is conceptualized as the clinical aggregate level to encompass common methodologic challenges in program evaluation at the patient, provider, clinician, clinic, hospital, or agency (county or

State) levels. Within this continuum, priority is placed on methods that measure provider adherence to recommended care processes and clinical outcomes.

At the national level (Tier 3), the main purpose of MBC is to monitor change in adherence rates to national quality measures. For example, adherence to quality measures in the Child Core Set is used to assess the quality of care provided by states to Medicaid and CHIP beneficiaries. State agency reporting on a standardized set of measures is envisioned to facilitate tracking performance over time. The proposed mechanism of change is public reporting on quality measure adherence rates will drive quality improvement by identifying target areas that state Medicaid agencies can be held accountable to address. Incremental increases in measure adherence rates over time are interpreted as improvement in quality of care.

AREAS FOR FUTURE RESEARCH

Despite substantial progress, there remain far more questions to be addressed which require development of innovative study methods and improved data infrastructure. Although there has been rapid expansion of telehealth and technology-supported treatments, more rigorous, large-scale research is needed to examine the impact of these technologic advances on the delivery of care, quality of care, and disparities in care. Access to the internet, for example, is not equitable and there are concerns that the digital divide may exacerbate disparities in care, especially among our most vulnerable children. In addition, the centralization of EHR data within care networks using common data models provides an exciting opportunity to leverage these databases to examine child mental health care quality and health disparities. This line of research will likely also identify limitations to stimulate further refinement of data coordination methods and predictive analytics for use in child mental health services research. With these advances, lessons learned from landmark events described in the historical context of this field should be heeded and the shared values of quality care should remain critical underpinnings for future research.

Further Readings

American Academy of Pediatrics. The medical home. *Pediatrics*. 1992;90:774.

Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav*. 1995;36:1.

Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. *BMC Psychol*. 2015;3:32.

Bennett TD, Callahan TJ, Feinstein JA, et al. Data science for child health. *J Pediatr*. 2019;208:12–22.

Bickman L. Implications of a children's mental health managed care demonstration evaluation. *J Ment Health Adm*. 1996;23:107.

Bushnell GA, Crystal S, Olfson M. Trends in antipsychotic medication use in young privately insured children. *J Am Acad Child Adolesc Psychiatry*. 2021;60(7):877–886.

CMS National Quality Strategy. April 2023. Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.cms.gov/files/document/cms-national-quality-strategy-handout.pdf. Accessed September 26, 2023.

Donabedian A. Promoting quality through evaluating the process of patient care. *Med Care*. 1968;6:181.

Edgcomb JB, Zima B. Machine learning, natural language processing, and the electronic health record: innovations in mental health services research. *Psych Services*. 2019;70:346–349.

Goldman ML, Druss BG, Horvitz-Lennon M, et al. Mental health policy in the era of COVID-19. *Psych Services*. 2020;71:1158–1162.

Institute of Medicine Committee on Quality Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. National Academy Press; 2001.

Jensen PS, Arnold LE, Swanson JM, et al. 3-Year follow-up of the NIMH MTA Study. *J Am Acad Child Adolesc Psychiatry*. 2007;46(8);989–1002.

McGlynn EA, Norquist GS, Wells KB, Sullivan F, Liberman R. Quality of care research in mental health: responding to the challenge. *Inquiry*. 1988;25:157.

Monteith S, Glenn T, Geddes J, Bauer M. Big data are coming to psychiatry: a general introduction. *Int J Bipolar Disord*. 2015;3:1–11.

Pappadopulos E, Jensen PS, Chait AR, et al. Medication adherence in the MTA: saliva methylphenidate samples versus parent report and medication effect of concomitant behavioral treatment. *J Am Acad Child Adolesc Psychiatry*. 2009;48(5):501–510.

Pescosolido BA, Manago B, Monahan J. Evolving public views on the likelihood of violence from people with mental illness: stigma and its consequences. *Health Aff (Millwood)*. 2019;38:1735–1743.

Pescosolido BA, Martin JK. The stigma complex. *Annu Rev Sociol*. 2015;41:87–116.

Pescosolido BA, Perry BL, Krendl AC. Empowering the next generation to end stigma by starting the conversation: bring change to mind and the college toolbox project. *J Am Acad Child Adolesc Psychiatry*. 2020;59(4):519–530.

Pescosolido BA, Perry BL, Martin JK, McLeod JD, Jensen PS. Stigmatizing attitudes and beliefs about treatment and psychiatric medication for children with mental illness. *Psychiatr Serv*. 2007;58:613.

Rapport F, Clay-Williams R, Churruca K, Shih P, Hogden A, Braithwaite J. The struggle of translating science into action: foundational concepts of implementation science. *J Eval Clin Pract*. 2018;24(1):117–126.

Shenkman E, Thompson L, Bussing R, et al. Provider specialty and receipt of metabolic monitoring for children taking antipsychotics. *Pediatrics*. 2021;147(1):e20200658.

Stroul BA, Friedman RM. *A System of Care for Severely Emotionally Disturbed Children and Youth*. CASSP Technical Assistance Center, Georgetown University Child Development Center; 1986.

Turan JM, Elafros MA, Logie CH, et al. Challenges and opportunities in examining and addressing intersectional stigma and health. *BMC Med*. 2019;17:7.

Wensing M, Sales A, Armstrong R, Wilson P. Implementation science in times of Covid-19. *Implement Sci*. 2020;15:42.

Yamaguchi S. Evaluation of co-productive, student-led, and college-based anti-stigma intervention. *J Am Acad Child Adolesc Psychiatry*. 2020;59:506–507.

Zima BT. Measurement-based data to monitor quality: why specification at the population level matter? *Child Adolesc Psychiatric Clinics N Am*. 2020;29(4):703–731.

Zima BT. U.S. child behavioral health quality measures: advancing a national research agenda. *J Abnorm Child Psychol*. 2020;48(6):745–756.

Zima BT, Bussing R, Tang L, et al. Quality of care for childhood ADHD in a large managed care Medicaid program. *J Am Acad Child Adolesc Psychiatry*. 2010;49(12):1225–1237.

Zima BT, Hurlburt MS, Knapp P, Ladd H, Tang L. Quality of publicly-funded outpatient specialty mental health care for common childhood psychiatric disorders in California. *J Am Acad Child Adolesc Psychiatry*. 2005;44:130.

54.11 Impact of Terrorism and School Shootings on Children

NAYLA M. KHOURY, M.D., M.P.H., AND WANDA P. FREMONT, M.D.

Over the last few decades, terrorist attacks, threats, and their aftermath have been increasingly frequent in countries throughout the world. The effects of terrorist activities have a significant clinical impact on children and their families. Additionally, school shootings, a specific form of domestic terrorism, have been increasing in recent years. A growing amount of research documents the effects of these exposures, especially data that helps to differentiate reactions to terror from those associated with other types of violence. In addition, the indirect effects of terrorism, including exposure to media and television coverage of violent images, have been noted to be of increasing concern. To develop a proactive and strategic response to these reactions, clinicians, educators, and policymakers must understand the psychological effects of terrorism, including school shootings. Many of the effects of terrorism-induced trauma on children are similar to the effects of other human-caused and natural disasters. Therefore it is essential for the clinician to have an understanding of children's reactions to trauma. The important contributing influences on children's reactions to trauma include risk and protective factors, developmental factors, neurobiologic consequences, and family and community strengths and vulnerabilities. Recently, with the increase in terrorist bombings and access to media coverage worldwide, more studies have examined the unique effects of terrorism-induced trauma on children. This research covers a range of violent acts that are classified as terrorism. These include isolated events in countries not at war (the Oklahoma City bombing and September 11 terrorist attacks) and repeated terrorist attacks in areas of political conflict (terrorist activities in Northern Ireland, state terrorism in Guatemala, the Scud missile attacks during the Persian Gulf War, and bombings and kidnappings in Israel). Earlier studies completed in the 1970s and 1980s were primarily descriptive and anecdotal in nature. More recent studies have been more rigorous and focus on protective and risk factors that affect children's reactions, as well as systems of care that support children in the aftermath of these events. They include longitudinal studies that assess the long-term consequences of terrorist activities and school shootings. Diagnostic implications, clinical interventions, and treatment strategies have developed as a result of the increasing knowledge obtained about the effect of terrorism on children.

DEFINITION OF TERRORISM

The U.S. Defense Department defines terrorism as “the calculated use of violence or threat of violence to inculcate fear: Intended to coerce or to intimidate governments and societies in the pursuit of goals that are generally political, religious, or ideological.” The goal of terrorism is not only to cause visible disaster but also to inflict psychological fear and intimidation at any time—during periods of peace or conflict. Unlike family or community violence or the trauma resulting from war, terrorist activities are not confined to a specific geographical area or time. Terrorist events occur suddenly without any forewarning and frequently result in severe trauma. The threat persists indefinitely. New terrorist threats and actions may exacerbate underlying anxieties. Terrorism capitalizes on media coverage. Powerful visual images create strong emotional responses, including fear, panic, despair, and rage, even in individuals who live far from the traumatic events. The events are given well-recognizable labels (e.g., “the blood bus,” “the explosive refrigerator in Israel,” “the Chowchilla bus kidnappings,” “the Scud missile attacks,” “the 9/11 terrorist attacks”).

Mass school shooting incidents are another form of domestic terrorism that occurs worldwide and elicits similar media coverage and strong emotional responses. From the 1999 Columbine High School massacre to the 2018 Parkland high school shooting, these events cause panic and public outcry for increased preventive efforts and legislative reform. Although these types of events occur more in the United States than in any other country, they can happen anywhere. Examples include the 2007 attack in Finland at Jokela Secondary School and the 2008 attack at Seinajoki University of Applied Sciences in Kauhajoki.

Mass school shootings have led to increasing youth homicide rates over the past two decades. The Centers for Disease Control reports that between 1994 and 2016, multiple victim school–associated youth homicide incident rates increased and involved a disproportionate number of youth homicide victims. An overwhelming majority of the associated mortality is due to firearm-related injuries.

While the definition of mass shootings varies depending on the number of injuries/fatalities (typically three to four fatal or nonfatal injuries), they are differentiated from violence related to gangs, family members, or interpersonal violence and are more unpredictable and seemingly random. Similar to other forms of terrorism, they involve unrelated persons in a public space occupied by multiple unrelated persons. Less common than community violence or single-victim incident violence, they have far-reaching consequences for victims, families, and communities.

CHILDHOOD REACTIONS TO TERRORISM-INDUCED TRAUMA

Although many of the reactions of children to terrorism-induced trauma are similar to children’s reactions to other forms of human-caused violence and natural disasters, terrorism-induced trauma results in unique stressors and reactions in children. The clinician must have a good understanding of the variables that affect children’s responses to trauma and the variety of responses that children manifest. It is, therefore, important to review the literature on childhood reactions to trauma and their clinical manifestations. Children vary in their reactions to traumatic events. Many demonstrate resilience and will be unaffected by these events in the long term. Some suffer from anxiety and bad memories that dissipate with time and emotional support. Other children may be more severely traumatized and experience long-term problems that significantly interfere with their functioning. The reader is encouraged to refer to the section in this text on posttraumatic stress disorder (PTSD) in children and adolescents.

Several factors contribute to a child’s differential response to trauma. These include the child’s age and level of psychological maturity, the severity and level of exposure to trauma, the duration of exposure to violence, predisposing risk factors, protective factors and resiliency, the influence of parents’ responses to disaster, family protective factors, the degree of social and community support (including educational, political, and religious influences), and the extent and context of displacement and relocation.

Predisposing risk factors include previous traumatic exposures. Research has shown that other types of trauma, including childhood maltreatment and community violence exposure, influence outcomes. Childhood adversities, including maltreatment and community violence, are known to increase vulnerability to subsequent victimization and psychiatric disorders, including PTSD. While issues of interpersonal and community violence are beyond the scope of this chapter, there are overlaps in documented potential mental health outcomes.

Trauma-related symptoms are documented as potential outcomes for youth regardless of the type of trauma. Children's emotional reactions may develop immediately after the trauma or may occur later. *Acute stress disorder* (ASD) is the most common psychiatric disorder following a traumatic event. *PTSD* is a reaction that may occur after a period of at least 1 month following the event. Children or adolescents suffering from ASD or PTSD may re-experience the trauma by having nightmares or recurring flashbacks of the trauma. They may become fearful, distractible, anxious, or depressed. Their play and study may be affected in addition to their pattern of sleep. In young children, repetitive play may occur in which themes or aspects of the trauma are expressed. Their dreams may be frightening but without any recognizable content. Trauma-specific re-enactment may occur. Adolescents and children may feel numb and withdraw from the external world or avoid situations that arouse recollection of the trauma. Symptoms of increased arousal may occur that include hypervigilance, sleep difficulties, irritability, difficulty concentrating, and outbursts of anger.

In addition to ASD and PTSD, children may develop comorbid disorders, including anxiety disorders (e.g., generalized anxiety disorder, social phobia, specific phobias, panic disorder, agoraphobia, obsessive-compulsive disorder), depression, and substance abuse-related disorders. Children may present with partial or variable symptoms of these disorders.

Children who are indirectly exposed to a traumatic event may also manifest symptoms of anxiety, including trauma-specific fear, fear of being alone, the habit of clinging to others, and event-specific fears.

Children's responses to terrorism-induced trauma include many of the symptoms and reactions to nonterrorist traumatic events. However, terrorism-induced trauma results in unique stressors and reactions in children. Terrorism, by its nature, is likely to cause psychological stress. Because terrorists can strike at any time and at any location, the threat of terrorism is omnipresent and enduring. Mass school shootings can cast doubt on the safety of a school for children and their families. The unpredictable, indefinite threat of terrorist events, the profound effect on adults and communities, and the effect of extensive terrorist-related media coverage exacerbate underlying anxieties and contribute to a continuous state of stress and anxiety. The resulting anticipatory fear and anxiety may be severe.

Other types of violence, such as childhood maltreatment and community and neighborhood violence, are associated with externalizing symptoms in some studies. This has not been found in most studies on youth following terrorism trauma. Instead, mental health outcomes for youth exposed to terrorism or mass school shootings includes increased risk for trauma-related symptoms, anxiety, depression, and functional impairment.

RISK AND PROTECTIVE FACTORS FOR CHILDREN EXPOSED TO TERRORISM

Several risk and protective factors influence children's reactions to terrorist events. These include factors relating to the history of the terrorist event and individual, family, and community strengths and vulnerabilities.

History of the Terrorist Event

The level of exposure to terrorist actions, either directly or indirectly, affects a child's response. The degree of exposure to terrorist actions is related to the prevalence of PTSD. The more severe the traumatic event, the greater the risk of the child developing posttraumatic symptoms. Children who experience loss directly are more symptomatic. Suffering physical injury or witnessing the death or physical injury of others is associated with higher rates of PTSD and comorbid depression and anxiety. The degree of personal loss (i.e., the relationship of the child to the victim) is also correlated with the severity of PTSD symptoms. Knowing an injured or deceased person increases the risk of symptom development. Children who lose a parent manifest the highest rates of PTSD. In addition to the level of trauma, the duration of exposure predicts the risk for the development of psychiatric problems in children.

Individual Factors

Individual factors and vulnerabilities that affect a child's differential response to terrorist events are related to developmental factors. These include a child's age and level of psychological maturity. Additional individual factors influencing responses include neurobiologic responses, predisposing risk factors, protective factors, and effects of resiliency.

Childhood Development Influences on Terrorism Reactions. How children understand and conceptualize traumatic events resulting from terrorist activities is greatly influenced by their developmental level with respect to cognition, emotional, and social factors.

PRESCHOOL-AGE CHILDREN. Preschool-age children between the ages of 3 and 5 years express their concerns for safety in terms of separation from parents and other primary caregivers. At this age, they are not able to cognitively understand the finality of death and do not fully understand the actual events that are occurring. Preschoolers rely on cues about danger from caretakers. They are more comforted than older children by reassuring and supportive parental behaviors. They may manifest their anxiety by having difficulties in separating from their parents. Regressive behaviors, including thumb sucking, enuresis, and fear of the dark, are not uncommon in young children. In addition, children aged 5 years and younger may manifest symptoms of fear, anxiety, sleep problems, and aggressive behaviors.

LATENCY-AGE CHILDREN. Latency-aged children, 6 to 11 years, are prone to derive a sense of safety from clear-cut rules. At younger ages, they have difficulty abstracting and thus tend to polarize complex situations into “right” and “wrong” or “good” and “bad.” As they mature, they develop an increasing sense of empathy and altruism. Children in this age group recognize the finality of death and are capable of fearing for others. Latency-age children are extremely influenced by indirect effects of violence and trauma. Their reactions to terrorist events may include depression, PTSD, and anxiety. Symptoms of anxiety include irrational fears, sleep problems, nightmares, somatic complaints (headaches, stomachaches), irritability, angry outbursts, and school avoidance. Signs of depression may include sadness, withdrawal, feelings of hopelessness, and problems with attention. These problems may result in academic difficulties.

ADOLESCENTS. Adolescents aged 12 to 18 years are likely to be vulnerable to idealistic notions and an unrealistic sense of immunity to harm and omnipotence. This may influence their judgment regarding the degree of danger in their environment. When exposed to an extreme act of violence, such as a terrorist attack, young adolescents may feel overly responsible for protecting their parents from danger. Adolescent responses to terrorist events are more similar to adult responses. They may exhibit symptoms of acute or PTSD, including intrusive thoughts, hypervigilance, and emotional numbing. Additional symptoms may include nightmares, sleep disturbances, avoidance, and withdrawal. They are at increased risk of developing depression, substance abuse problems, and peer problems. Shame, humiliation, self-blame, and guilt may result from a sense of powerlessness.

Neurobiologic Consequences of Trauma in Children. Research on chronic stress, such as child abuse and neglect, and acute stress, such as disaster, war, or medical trauma, supports the hypothesis that trauma produces long-lasting effects on brain functioning. The human brain undergoes significant development during childhood and adolescence. Trauma may have a particularly devastating effect on the developing brain and may cause permanent changes in neurotransmitter systems that are stimulated by stress. Early childhood exposure to traumatic events influences the neurobiologic response to stress in the developing brain in such a way that it predisposes affected individuals to mood and anxiety disorders later in life. An important mediator of the stress response is corticotropin-releasing factor (CRF). Stress causes the hypersecretion of CRF. CFR is a key mediator of the stress response.

Dysregulation of the CRF system in the pituitary gland and brain causes an exaggerated glucocorticoid response. This response contributes to the symptomatology of traumatic syndromes, such as an increased startle response and symptoms of anxiety and mood disorders. Excessive stress hormones also affect myelination and synaptogenesis in the developing brain. This is associated with structural alterations in the corpus callosum, hippocampus, amygdala, ventromedial prefrontal cortex, and cortex. Elevated glucocorticoids have also been shown to affect the immune system causing altered glucocorticoid sensitivity in target immune cells, shifts in immune cell distribution, elevated proinflammatory cytokines, and a decrease in regulatory T cells. These changes result in impaired immunity and inflammation responses. Trauma has been associated with epigenetic changes through alternations in methylation resulting in changes in gene expression. Gene variation is also thought to contribute to the biology of differential risk for developing PTSD. The level of genetic influence on the variability of PTSD risk in the population has been shown to be 5% to 20%, with some variability for sex. New research is emerging on the biology of human resilience, including biologic and genetic protective factors that buffer the adverse neurobiologic responses to stress. In addition, emerging research suggests psychosocial factors such as positive relationships can provide a healthy buffer in the face of childhood adversity. More research is needed to understand whether the neurobiologic changes are reversible or permanent. An understanding of the neurobiologic consequences of trauma is important to determine treatment.

Predisposing Risk Factors. All children who are exposed to traumatic events experience a threatened sense of safety and potential for personal danger. However, children with predisposing risk factors are at greater risk of developing symptoms of anxiety and depression. Risk factors include exposure to past traumatic events during childhood, childhood anxiety, childhood conduct problems and antisocial behaviors, learning disabilities, and a history of psychiatric disorders. Gender also plays a role, with female gender being associated with increased risk for poor mental health outcomes following a traumatic event in the community and in certain forms of terrorism such as school shootings. Certain communities, including under-represented minorities, are more vulnerable to the effects of trauma, given high rates of childhood adversities and historical and ongoing structural violence. Children who have survived the death of a parent or another significant person are more vulnerable to the effects of trauma. In addition, each child reacts differently, depending on temperament and sensitivity and whether they tend to internalize or externalize experiences and emotions.

Protective Factors and Resilience. Many children who are exposed to trauma do not develop pervasive chronic posttraumatic symptoms. Factors associated with resilience among children after exposure to traumatic events are related to both psychological, biologic, and community support characteristics. Resilience is related to the variability in psychological interpretation of physiologic arousal and relies on the presence of consistent caregivers. Initial biologic responses to a fearful event include an increase in cortisol level in the brain and an increase in heart rate. Recovery from a traumatic event includes the ability to respond to the event in a supportive environment without extreme avoidance and in the absence of persistent physiologic arousal. The development of a perpetual state of fear, with both emotional reaction and abnormal physiologic response, leads to posttraumatic symptoms. Factors related to resilience include a child's characteristics and resources (persistence, motivation, adaptability, optimism, willingness to approach novel events, self-esteem, intelligence, social skills), developmental stage, family characteristics (support, open communication, consistency), and community support. Understanding risk and protective factors that contribute to a child's response to trauma is important to develop strategic interventions. Early intervention facilitates the process of recovery in vulnerable individuals.

Family Factors and Vulnerabilities

Children's response to trauma and their ability to cope with these disasters are strongly influenced by their parent's response to the violent events. There is a positive correlation between children's and parents' symptomatology. Parental involvement in mediating stress reactions in children exposed to terrorist attacks is a protective factor and helps to reduce a child's vulnerability to the trauma. Parents who underestimate their children's reactions may unintentionally hinder their children's process of mastering the trauma. Terrorist incidents affect adults profoundly. Therefore parents, adult caretakers, and teachers may not be able to provide the support and reassurance needed to help avoid potential long-term emotional harm to their children. Increased levels of distress and less adequate coping skills have been noted in children whose parents responded to traumatic events with negative emotions. Positive coping responses in children are associated with parents who responded with positive emotional reactions. Family factors that have been shown to increase resiliency and buffer stress responses in children include a stable, secure emotional relationship with at least one parent, a parental model of constructive coping mechanisms, and physical proximity of children to parents.

Community Factors

Children's responses to terrorism are significantly influenced by community members' responses. While community violence can be an important childhood adversity that increases vulnerability for youth, community support and relationships within the community can be critical protective factors. Strong social support and positive community influences increase resilience. Community educational, political, and religious support fosters moral development in children. The presence of a caring adult who is able to provide emotional support, encourage self-esteem, and promote competence has a moderating influence on acute and long-term mental health outcomes. Community value systems, beliefs, and ideology contribute to resiliency by giving meaning to dangerous events, allowing children to identify with cultural values, and enabling children and adults to function under extreme conditions.

CHILDREN'S RESPONSES TO MEDIA EXPOSURE TO TERRORIST EVENTS AND SCHOOL SHOOTINGS

Media coverage of terrorist activities has a significant effect on children's emotional responses. The media plays an important role in providing information to the public, especially during times of disaster. The images portrayed are powerful and intensely emotional. They are frequently repeated over and over again. Terrorist organizations depend on the media dissemination of their activities to increase responses of fear, thus creating an environment of persistent threat. Similarly, school shooters often rely on and seek the attention of media. Similar to suicides, media contagion effect has been well documented in the aftermath of a mass school shooting. Studies have shown that one-third or more children watch the news. Extensive research has examined the influence of violence and aggression on children and their families. Recently, with the increase in terrorism and access to media throughout the world, more studies have examined the effects of children's reactions to media coverage of terrorism. The results research, covering a range of violent acts, has shown that remote exposure, such as viewing terrorism on television, has a significant clinical effect on children's responses.

Because media exposure is so extensive and the effect on children's reactions is significant, it is important to provide a summary of relevant factors pertaining to children's exposure, interpretation, and constructive and destructive effects of television and the media.

Several studies have examined the extent of children's exposure to the media, which has dramatically increased over the last decades. On average, children ages 8 to 12 in the United States spend 4 to 6 hours a day watching or using screens, and teenagers are reported to spend up to 9 hours per day. With the rise of social media and personal electronic devices among youth, it is now possible to get access to news including terrorist events and school shootings nearly 24 hours per day. News coverage in these mediums may be more stimulating, distressing, and unfiltered.

Childhood Developmental Influences on Media Responses

Children's understanding and interpretation of television and the media are related to their cognitive and psychological developmental levels. Children are less able than adults to regulate their emotions after watching traumatizing images.

Preschool-Age Children. Preschool-age children are most affected by visual images and emotional sounds. They are more upset by images of dead bodies or bloodied survivors and crying victims or witnesses than falling buildings or announcer commentaries. They are not able to understand the extent of the tragedy. Preschoolers cannot distinguish between live pictures and replays. If events are shown over and over again, they believe that the catastrophe is repeatedly occurring in real life. They are likely to believe that the terrorism is happening at the same time they are watching the coverage, and they believe that it is occurring in close proximity to their home.

Latency-Age Children. Latency-age children (ages 6 to 11 years) have a better understanding of the event but will interpret it concretely. They have difficulty understanding the effect of the terrorist threats and adult responses to these threats. They are more sensitive to injured children or those whose parents were killed. School-age children are more likely to focus on their safety and their family's safety.

Adolescents. Adolescents are more capable of abstraction than younger children. They are equally affected by visual images; however, they are also concerned about the effect of threats on the future and community responses. They are able to understand more complex philosophical and existential issues such as social justice, politics, and theology. Their personal struggles with identity, meaning, and place in society may be compromised seriously or distorted by the perception of the world and its human interactions as dangerous, untrustworthy, and immoral. This worldview may have a negative effect on their sense of the future.

Although there has been extensive research about children's exposure to the entertainment media, there is far less scientific research about children's exposure to the news, especially exposure to terrorist media coverage. Media coverage of terrorist events has both positive and negative consequences.

Constructive Effects

Several constructive effects of the media have been emphasized. The media plays a vital role in disseminating information quickly to local, national, and international regions. Media coverage facilitates family and community cohesion. It facilitates the dissemination of educational information, and it can serve as a vehicle for discussion in older children and adolescents. The media, the Internet, and the telephone may serve as an extensive support system. Television stations are able to broadcast advice on how to help children to cope with stress and anxiety related to disastrous events. Data obtained after the September 11, 2001, attacks noted that sustained advertising through multiple media outlets was effective in informing individuals about counseling resources and encouraging them to seek mental health services.

Destructive Effects

The increased incidence of negative effects and stress-related symptoms in children exposed to media coverage of terrorist events has been examined recently in studies of the aftermath of the *Challenger* disaster, the Oklahoma City bombing, the Scud missile attacks and the Gulf War, the World Trade Center attacks, and the conflicts between Palestinians and Israelis. Repeated media exposure is especially detrimental for young children because of their cognitive immaturity. The effect of repeated coverage may promote the trauma and increase the intensity and duration of the effects. The media may have a negative effect on parents and caregivers, which directly affects the coping abilities of their children. More research is needed to measure the effects of media coverage of terrorist events and to determine whether they have detrimental or positive effects on mental health outcomes.

LITERATURE REVIEW OF TERRORIST EVENTS

Early research on the effects of terrorism indicated that children exposed to violence are greatly affected by the experience. These studies examined children's responses to terrorist activities in Northern Ireland, state terrorism in Guatemala, and terrorist attacks in Israel. Children manifested many psychological symptoms, including anxiety, depression, anger, feelings of hopelessness with regard to the future, difficulties with trust, anticipatory fears, and an altered sense of security. More recent studies have used statistically valid structured assessments to document children's reactions and symptoms. Research has been conducted in the aftermath of the 1993 World Trade Center bombing, the 1995 bombing in Oklahoma City, the 1991 Gulf War and its effects on Palestinian and Kuwaiti children, Scud missile attacks, and terrorist activities in Israel in the last two decades, the 1998 American Embassy bombing in Kenya, and the September 11, 2001, terrorist attacks. These studies emphasized the importance of a thorough assessment of children's responses and the influence of the media on children's reactions to a terrorist event. The rates of PTSD in children exposed to terrorist activities range from 28% to 50%. Other psychiatric problems have been identified, including anxiety, separation problems, depression, mood changes, sleep difficulties, regressive symptoms, and behavioral problems.

The importance of the influence of risk and protective factors on responses to terrorism has been identified. Research has increasingly focused on factors (personal, family, and community) that contribute to the persistence of psychological problems and factors that foster resilience. The results of these studies have contributed to the design of prevention and intervention strategies to help children and their families to cope and recover from the effects of terrorist attacks.

The 1991 Persian Gulf War and Scud Missile Attacks

Repeated terrorist attacks, including the Iraqi Scud missile attacks and terrorist activities on civilian sites in Israel, occurred during the war in the Persian Gulf in 1991. Further data measuring children's affective reactions and coping skills to terrorist events were gathered during studies conducted by several researchers during and after the attacks. The results of the research supported earlier data that correlated the degree of exposure to trauma to the prevalence of PTSD. Children who directly experienced loss due to terrorist attacks were more likely to develop symptoms of PTSD. The degree of personal loss was also correlated with the number of trauma symptoms. Knowing an injured or deceased person increases the risk of symptom development. The studies in Israel also focused on risk and protective factors and the long-term consequences of terrorist activities. Several studies addressed the importance of parental involvement in mediating stress reactions in children. Earlier research identifying the protective effects of family relationships was conducted on Cambodian adolescent refugees who survived 4 years under the Pol Pot regime from 1976 to 1979, during which 1 to 2 million people were exterminated. Adolescents who were able to re-establish contact with a family member had fewer psychiatric symptoms than adolescents who lived alone or in a foster home. The natural families were successful in maintaining traditional child-rearing practices. Clear adult–child boundaries and high expectations for conformity and responsibility were found in these families. Naomi Bat-Zion and Rachel Levy-Shiff noted that parents' responses were a central mediating factor for Israeli children's reactions to Scud missile attacks. Negative emotional expressions by parents were associated with increased levels of distress in children, whereas positive emotional manifestations were associated with increased coping efforts.

Laor and colleagues studied the long-term consequences on Israeli children of the Scud missile attacks. Children of mothers with poor psychological functioning showed more stress symptomatology in families who were displaced. The association between children's and mother's symptoms was stronger among younger children (ages 3 to 4 years). Unhealthy family interactions—either disengaged or enmeshed styles—exacerbated children's stress. These children manifested more symptoms, and their capacity to recover from the trauma was jeopardized. Conversely, children with adequate family cohesion manifested less stress in reaction to trauma and were better able to recover from the initial effect of the trauma.

Correlation between media exposure to terrorist activities and posttraumatic stress symptoms was verified in several studies. The effect of the media on children living in a war zone was also studied in Palestinian and Kuwaiti children during the Gulf War. Palestinian children who were exposed to traumatic events indirectly by the media had a higher incidence of anticipatory anxiety and cognitive expressions of distress than children who were exposed directly, who in turn manifested more symptoms of PTSD.

1993 World Trade Center Bombing

The first major terrorist attack in the United States occurred on February 26, 1993, when international terrorists bombed the World Trade Center in New York City. Six people were killed, and thousands of persons were trapped inside the building. One study compared the reactions of children who were present in the building at the time of the explosion to a community sample of children who were indirectly exposed (children whose exposure was limited to interpersonal relationships with persons directly affected and media coverage of the event). Children who were directly exposed to the attack were more symptomatic. Children in the comparison group had mild symptoms, which did not interfere significantly with their functioning. A marked discrepancy was noted between the parents' report of PTSD symptoms and fear in their children and the children's direct report of their fears. In a follow-up study 9 months later, parents of both groups noted a decrease in PTSD symptoms and fear in their children. However, the children did not report any significant decrease. The results of this study emphasize the importance of evaluating children directly in addition to evaluating their parents.

1995 Oklahoma City Bombing

On April 19, 1995, the Alfred P. Murrah Federal Building, a U.S. government office complex in downtown Oklahoma City, Oklahoma, was bombed. Until the September 11, 2001, attacks, it was the deadliest act of terrorism on U.S. soil. A total of 168 people were killed, including 19 children, and more than 800 people were injured. Until then, few studies had examined children's reactions to terrorist events. Several studies were conducted in the aftermath of the attack, most notably those of Betty Pfefferbaum and colleagues. The significance of the effect of media exposure on terrorist attacks was identified in these studies. In one study, preschool-age children who attended day care in a facility near the Federal Building were examined. Although many of these children had minor injuries, none of them died or experienced serious injuries. Parents tended to underestimate symptoms in their children. All children met the exposure criterion for PTSD, although none of their parents reported diagnostic criteria for the disorder. Staff reports of children indicated that they met the exposure and re-experiencing criteria (posttraumatic play, discussion with peers). Only three children met the criteria based on parent report. The importance of direct observation of young children after exposure to trauma is essential for an accurate assessment of their reactions. The results of these studies also supported the literature documenting developmental differences in responses to trauma. Children were more likely to exhibit regressive behaviors and functional problems than the numbing and hyperarousal criteria. None of the preschool children reported symptoms of avoidance, restricted range of affect, or sense of foreshortened future that are typical of older children and adolescents.

Research on the Oklahoma City bombing also addressed the effects of media exposure to terrorist events and indirect exposure on children's responses. Lenore Terr and colleagues first studied these effects of distant trauma after the Space Shuttle *Challenger* explosion in 1986. She examined the responses of children who had watched the explosion on television rather than directly witnessed the disaster. Terr and her colleagues defined distant trauma as "the reaction (memory, thinking, symptoms) to a disastrous event, observed at the time of the event, but from a remote, and safe distance." Children's responses included symptom patterns similar to PTSD. Children's reactions to distant trauma also included trauma-specific fear, fear of being alone, and clinging behaviors. Children who lived on the East Coast were more symptomatic than children who lived on the West Coast. East Coast children had a significant increase in dreaming, drawing, and behavioral enactment, trauma-specific fears (e.g., death and dying, explosions, fires, airplanes), and clinging behaviors. Children throughout the United States had diminished expectations of the future. Latency-age children were more affected than adolescents. Most symptoms diminished within 1 year; however, for children with partial PTSD, symptoms (i.e., posttraumatic play for young children and hopeless attitudes about the future for adolescents) often persisted. Terr and colleagues defined a spectrum of trauma-related conditions that resulted from remote exposure to violence, including viewing media coverage of disaster events. Because one of the goals of terrorists is to inflict fear in the broader community, it is important to examine the effects of less direct and distant forms of exposure.

Pfefferbaum and colleagues assessed more than 3,000 middle and high school students 7 weeks after the Oklahoma City bombing. Most of the children heard or felt the blast, more than 40% knew someone who was killed or injured, and approximately two-thirds reported extensive television viewing of the bombing. Seven weeks after the event, one-third of the children reported continuing worry about themselves or their families. Female gender, acute reactions of fear and arousal, interpersonal exposure, and bomb-related television viewing were associated with an increase in posttraumatic stress symptomatology. Children who experienced personal losses were at greater risk of watching significantly more terrorist-related television coverage than were children without direct losses. This further exacerbated the traumatic experience. Children who witnessed the bombings on television reported trauma-related stress for more than 2 years. In addition to broadcast media, print media (i.e., newspaper and magazine) exposure was also examined. Print media exposure correlated more strongly with enduring posttraumatic stress than exposure to broadcast media. Most of the children reported mild posttraumatic stress reactions and minimal effect on functioning. The clinical significance of the results is unclear, and further research is needed to determine whether they are at risk for future stress reactions.

1998 American Embassy Bombing in Nairobi, Kenya

More than 200 individuals were killed and thousands were injured on August 8, 1998, when a terrorist bomb destroyed the American Embassy in Nairobi, Kenya. Pfefferbaum and colleagues extended their studies of children who were indirectly exposed to terrorism, exploring issues that were not examined in the Oklahoma City studies, including exposure to other traumatic events. More than 500 indirectly exposed children in a convenience sample of Nairobi schoolchildren were assessed 8 to 14 months after the embassy bombing. Most of the children in the study lived in a community in which trauma was common and had experienced other crimes of violence and the death, illness, or injury of family or friends. Children who were at greater risk of developing symptoms of posttraumatic stress to the bombing incident had experienced a higher number of previous traumatic events, more severe posttraumatic stress reactions to past events, and stronger acute reactions. Little or no impairment of children’s functioning was noted. Similar to the Oklahoma City school samples, the children in this study experienced little direct exposure to the terrorist actions, and posttraumatic stress reactions and impaired functioning were relatively low. Future studies are needed to explore the significance of risk and protective factors, as well as a more comprehensive set of outcomes, including long-term longitudinal data.

September 11, 2001, Attacks

The deadliest terrorist events were the attacks on September 11, 2001. There were 2,973 fatalities, not including the 19 highjackers—246 on the four airplanes, 2,602 in New York City at the World Trade Center, and 125 at the Pentagon. These events significantly affected U.S. and world history, governmental policies, and the lives of individuals in the United States and other countries. Indeed, “pre-9/11” and “post-9/11” have become commonly used terms worldwide. The terrorist attacks and their aftermath, including repeated media coverage of the events, have had a powerful effect on children and their families.

July 22, 2011, Utøya Youth Camp Attack

On July 22, 2011, a lone man disguised as a police officer opened gunfire, killing 69 people at a summer camp for adolescents on the Utøya islet in Norway and shooting and injuring 66 other youth. The Utøya mass shooting is the most severe terrorist attack during peacetime in Norwegian history.

Multiple studies were conducted with the adolescent survivors and demonstrated functional impairment one year following the attack and risk for medical sequelae up to two years later. In a study by Stene, Schultz, and Dyb, 1 year following the attack, 61% of survivors had impaired academic performance and 29% impaired school well-being, both of which were associated with symptoms of posttraumatic stress, anxiety, depression, and somatic symptoms. Stensland and colleagues conducted a case-control study to examine whether exposure to terrorism increased the risk of migraines. They found increased risk for both migraine and tension-type headache. A variety of sleep disturbances were found in 2-year follow-up in a study done by Gronli and colleagues of 42 survivors when compared with matched controls; these disturbances were more prevalent in a subgroup of survivors with psychiatric disorders. Strom and colleagues demonstrated that school performance was also affected 1 year later but noted that grades improved in the last year of high school, indicating possible recovery with improved school absenteeism. They note that these students reported high satisfaction with school support and conclude that these results suggest the importance of providing in-school support for trauma-exposed students and helping them remain in school.

April 2013 Boston Marathon Bombings

On April 15, 2013, two pressure cooker bombs exploded during the Boston Marathon killing three people and injuring an estimated 264 others. A manhunt for the suspects followed. The subsequent investigation determined that the bombings were an act of terrorism.

Comer and colleagues studied the psychological impact of exposure to this event on children. He examined the rates of PTSD in children who witnessed the event and the impact of media exposure. Children who witnessed the bombings were six-times more likely to develop symptoms of PTSD. Two to six months after the attacks surveys from parents, living within 25 miles of the bombing and manhunt, showed that 11% of the parents identified PTSD symptoms in their children. This rate is similar to what was found among 9/11 New York City children 6 months after the attacks. Watching the manhunt on television was even more strongly associated with problems such as conduct issues or anxiety than exposure to the bombings themselves. Most children watched 1.5 hours of coverage on the day of the bombings, with 20% watching greater than 3 hours. About one-third of parents surveyed tried to restrict children's access to the media. Recognizing the potential psychological impact of media exposure on children, researchers advised that, "if there is a child who shouldn't watch an R-rated movie they should not be watching news with the same amount of violence and aggressive behavior."

School Shootings

The occurrence of school shootings has increased significantly during the past two decades. However, there have been few studies regarding the psychological effects on students who witnessed the shootings. Of those studies that were conducted, numerous challenges, including methodologic and design issues, attrition, under-represented minority youth, and variation in definitions and measures, contributed to difficulties in interpreting data and making conclusions. Research examining youth responses to school shootings was done in the aftermath of shootings at Jokela, Finland (2007), Virginia Tech (2007), Kauhajoki, Finland (2008), Northern Illinois University (2008), Seattle Pacific University (2014), Thurston High School, Springfield, Oregon, 1998.

A review of this research indicates that most young people demonstrated no long-term dysfunction. A minority of youth experienced severe and chronic symptoms. Clinically significant reactions were determined by proximity to the trauma, peritraumatic dissociation, posttraumatic emotional regulation difficulties, social support, and coping styles. Short-term responses included acute stress responses and emotional distress. Longer-term responses included PTSD, anger, and social withdrawal.

Risk factors for developing significant symptoms included proximity to trauma, history of previous trauma, female gender, fewer psychosocial resources, and media exposure to the event. Protective factors included social support seeking, resilience, gratitude, flexible coping styles, and secure attachment styles.

Children's Reactions. Research on children's reactions to these attacks corroborates the results of previous research studies. The degree of exposure, proximity to the disaster area, association with parental reactions, the experience of distant trauma (including media coverage), and history of pre-existing trauma were associated with increased stress symptoms. Children's reactions included symptoms of posttraumatic stress, depression, panic attacks, anxiety, separation anxiety, and agoraphobia. The largest study of children's responses was commissioned by the New York City Board of Education and was supported by the U.S. Department of Education through Project SERV (School Emergency Response to Violence). The study, conducted in collaboration with the New York City Department of Health and Mental Hygiene, surveyed 8,266 randomly selected children in grades four to twelve in the New York City metropolitan area 6 months after the September 11 attacks. Christine Hoven and colleagues surveyed children who attended schools in different areas of proximity to Ground Zero. Those children closest to the World Trade Center were most likely to have seen or heard the planes hitting the towers, seen the flames, been exposed to the dust and smoke, and struggled with fleeing to safety. Children farther away from the towers were more indirectly exposed to the attacks, including interpersonal exposure through direct victimization of a family member or friend or through extensive media exposure.

Seventy percent of all children had either direct or indirect exposure. The incidence of PTSD in this group increased from 2% to 10.5% following September 11. Children who lived outside of New York City showed an equal prevalence of symptoms of PTSD and other psychiatric problems (depression, panic attacks, anxiety, conduct problems) as children living within the city. The diagnosis of separation anxiety doubled, and that of agoraphobia tripled. Children who were exposed to trauma before the attacks were more likely to experience PTSD and psychiatric problems. More than one-half of the total sample of children who exhibited symptoms of PTSD had experienced previous traumatic events. This high frequency of previous trauma exposure emphasizes the need to address and identify prior trauma experiences when conducting needs assessments and intervention strategies following a terrorist event. Hoven and colleagues noted that although some of the children in the sample did not meet the criteria for a DSM psychiatric disorder, they still suffered from disorder functional impairment in the aftermath of the terrorist events, which they suggest was mediated by grief/bereavement reactions. They emphasized the importance of identifying and providing specialized treatment for grief reactions in traumatically bereaved youth.

Parental stress reactions were also noted to have a significant effect on children’s responses. The results of several studies noted a correlation between parents’ responses and the frequency and severity of posttraumatic stress symptoms in children. Gerry Fairbrother and colleagues assessed predictors of posttraumatic stress reactions in children using a parent report administered by a New York City telephone survey. Parents reported that almost 20% of their children experienced severe or very severe posttraumatic stress reaction, and approximately two-thirds of children had moderate posttraumatic symptoms. Parental reactions were associated with the severity of posttraumatic reactions in children. Mark Schuster and colleagues surveyed 560 adults throughout the United States 3 to 5 days after September 11 using random-digit dialing. Parents with substantial stress reactions were more likely to report that their children had symptoms. Thirty-five percent of adults surveyed reported that their children had one or more stress symptoms. Forty-seven percent of adults noted that their children were worried about their own safety or that of a loved one. This data continue to support the importance of evaluating parents’ reactions to a disaster when assessing children’s response and needs for treatment.

Further studies after the September 11 attacks assessed children’s psychological and behavioral responses by surveying their parents, either by telephone or Web-based interviews and assessments. William Schlenger and colleagues assessed psychological responses 1 to 2 months after the September 11 attacks using a Web-based epidemiologic survey based on the Posttraumatic Stress Disorder Checklist and the Brief Symptom Inventory. Adult reports indicated that children who were distressed by the events manifested symptoms of sleep disturbances, irritability, and fear of separating from their parents. Another study examined behavioral problems in New York City children by comparing three cross-sectional telephone surveys of adults. Two surveys were conducted after the September 11 attacks. The first survey was conducted 4 and 5 months and the second between 6 and 9 months after the attacks. The results of these surveys were compared to a representative survey conducted before September 11. The results of the study indicated that behavioral problems decreased in the first few months after the attacks but then returned to the predisaster levels after 6 to 9 months. These results support findings in an earlier study of children’s behaviors before and after Hurricane Andrew.

Children’s Reactions to Media Coverage. The effect of extensive terrorist-related media coverage on children’s stress reactions was noted in several different studies after the September 11 attacks. Media coverage was extensive, and repeated coverage of the graphic images of planes striking the World Trade Center caused considerable concern about the added emotional effects of viewing the traumatic events. Results of a national study conducted a few days after September 11 concluded that children age 5 to 19 years watched 3 hours of television news on the day of the attacks. Only 8% of children did not watch any coverage of the event. Parents who reported that their children were upset restricted their children’s television viewing. In children whose viewing of television was not restricted, the number of hours of television viewing correlated with the number of stress symptoms. A study commissioned by the New York City Board of Education 6 months after the attacks noted that 62% of children spent much of their time and 33% of children spent some of their time learning about attacks on television. Approximately 75% of children increased the amount of time they spent reading newspapers and magazines, and more than 30% obtained information from the Internet. A total of 10.5% (75,000) of students manifested symptoms of PTSD 6 months after September 11, including children who were not affected directly by the event. Among these children, the prevalence of stress was higher among children who spent more time watching television coverage of the events than children who spent less time. The author noted, however, that these results did not determine a direct causality between television viewing and PTSD rates. Exposure to the images on television may have exacerbated children's stress responses; however, anxious children may have been more likely to seek information from the media than nondistressed children.

The effects of television viewing of the September 11 attacks also were examined in adults. Specific disaster-related television images were associated with PTSD and depression among persons who were exposed directly to the disaster. The number of hours of television coverage viewed and an index of the content of that coverage was associated with PTSD symptoms.

Studies in several other geographical areas in the United States support the effects of media exposure. The indirect exposure of September 11 via television, the Internet, and printed media was examined in a study of elementary schoolchildren in the southeastern United States by Saylor and colleagues. PTSD symptoms were reported in children who saw reports on the Internet (versus television and printed material), saw images of death or injury, or feared that a loved one might have died in the attacks. There was no measurable benefit to seeing "positive" or "heroic" images. Older children and boys were noted to have greater media exposure and more trauma-specific PTSD symptoms.

The correlation between television viewing of the September 11 attacks and an increased sense of insecurity and stress symptoms was also demonstrated in a study of elementary school-age children in Washington, D.C. A few studies examined reactions of children who lived on the West Coast. One month after the attacks on the World Trade Center, adolescents in California experienced heightened perceptions of vulnerability to death, which extended beyond the terrorist attacks and generalized to unrelated risks. After exposure to terrorist attacks, even from across the country, adolescents internalized a perception of the world as more risky and dangerous. Children in an ongoing longitudinal study who were assessed after the 9/11 attacks demonstrated posttraumatic stress symptoms and fears at levels comparable to children who directly experienced disasters.

Following mass school shootings, similar findings have been documented in a recent review that suggests media exposure may increase risk of psychiatric disturbance and that vulnerable youth may seek more media exposure in a reciprocal fashion.

Counseling Interventions. Few studies examined the frequency of counseling interventions following the September 11 terrorist attacks. In a telephone survey 1 to 2 months after the September 11 attacks, more than 100 parents were asked to describe the experiences and reactions of their children, most of whom were in school or day care at the time of the disaster. More than 20% of the parents studied reported that they received counseling. Parental posttraumatic stress, male gender, and having more than one sibling living in the household were factors that were correlated with receiving counseling. Two additional phone surveys of 423 and 112 parents, respectively, reported that 10% and 22% of symptomatic children received mental health services. These services were more likely to have been accessed by children whose parents reported having current symptoms of PTSD. Gerry Fairbrother noted that only 27% of children who had been identified in his first study to have severe or very severe stress reactions received any mental health treatment 4 to 5 months after the September 11 terrorist attacks.

In another survey, military health system beneficiaries in the Washington, D.C., area after the September 11 attack on the Pentagon were studied using the Electronic Surveillance System of the Early Notification of Community-Based Epidemics (ESSENCE) database. The number of visits to behavioral health clinics for anxiety disorders and acute stress reactions in children increased after the attack.

Much literature on mass school shootings has focused on a number of primary preventive efforts; some include gun reform laws, school safety initiatives, and increased mental health access for all youth through increased integrated services, such as mental health in school-based or primary care settings.

Primary prevention efforts post-trauma also target identifying youth in need of selective or more indicated services. Secondary or selective prevention efforts include supporting wellness, coping, and resilience, normalizing reactions, and target exposed youth who may be at increased risk. School-based interventions are useful for primary and secondary efforts, given their central role in the lives of most youth, and as such can increase ease and access to treatment. Other community sites such as religious institutions, after-school and youth programs, and even shelters and refugee camps have been effectively used for delivering interventions. Tertiary prevention or indicated services are reserved for those with persistent clinical symptoms and functional impairment in the aftermath of a terrorist attack or mass shooting and involve more specialized mental health services.

Project Liberty. In the aftermath of the September 11 attacks, the largest federal health program in U.S. history, Project Liberty, was created to provide mental health services to approximately 1.5 million people. It was funded through grants totaling $155 million from the Federal Emergency Management Agency (FEMA). The New York State Office of Mental Health (NYSOMH) administered and designed the program and collaborated with almost 200 local agencies to provide service delivery to New York City and 10 surrounding New York counties. The overall goal of Project Liberty was to provide free and anonymous community-based disaster services to help individuals to recover from the psychological distress that they experienced as a result of the attacks. Services provided included disaster-related outreach, individual and group counseling, psychoeducation, and referral services for both adults and children. The services were developed based on the assumption that although most persons' distress reactions were disturbing to them, they constituted normal reactions to a traumatic event. Most services were planned to be of short duration, with the expectation that reactions would dissipate over a short period of time. Interventions helped people to identify their trauma responses and understand them as normal reactions to a traumatic event and to reconnect people with their pre-existing social supports. Populations identified to need the greatest support included victims' families, survivors and their families, displaced individuals, emergency and recovery workers, children, older adults, persons with limited financial and social support resources, certain cultural and ethnic groups, and persons with a history of psychiatric problems. Project Liberty provided face-to-face counseling and educational and outreach services to an estimated 1.2 million individuals in the New York City metropolitan area between the time services began a few weeks after September 11 through December 31, 2003, when services were terminated. An important component of Project Liberty was to provide assurance/quality improvement activities to assess the quality of the delivery and effect of services. Data on demographic variables, symptom clusters, variables predictive of further need versus recovery, and long-term outcomes were gathered. In addition, data were collected to monitor how well the services provided adhere to the treatment manual and guidelines.

As part of the quality assurance program, counselors completed encounter logs that included checklists for reactions to the September 11 attacks. These results were used to help guide program planning and responses to service recipients. Nancy Covell and colleagues collected data on service utilization and reaction patterns among children who received Project Liberty counseling services. Nine percent of Project Liberty community-based service recipients who received counseling between September 2001 and December 2003 were children. Census data for the 15 boroughs and counties serviced by the project indicated that children constituted 25% of the population. During a later phase of the project, funded by Project Liberty and organized through the New York City Department of Education, an additional 36,337 children received counseling services between January 2003 and December 2004. With these additional services, approximately 20% of Project Liberty's total services for individual (including family) counseling were provided to children. The most commonly reported reactions for children's responses paralleled responses recorded for adults. Reactions included sadness, fear, anxiety, concentration difficulties, anger, irritability, intrusive thoughts and images, and difficulty sleeping. Children's responses differed according to their age. Elementary-school–age children were more likely to display signs of withdrawal and isolation, anxiety, and difficulties with concentration and memory. Older children (ages 12 to 17 years) were more likely to exhibit symptoms similar to adults, including avoidance and "numbing" reactions, in addition to substance abuse.

CLINICAL INTERVENTIONS AND TREATMENT

The effectiveness and comparative advantages of specific treatment modalities and interventions for children exposed to terrorism have not been extensively studied. Only recently have researchers developed rigorously designed studies to examine the effectiveness of treatments in children exposed to nonterrorism-induced trauma (e.g., sexual abuse, single-incident traumas, natural disasters, community violence, and war-induced trauma). The results of this research are promising and have significant implications for intervention strategies and treatment of children exposed to terrorism. These include community, school-based, individual, and family interventions and treatments. Essential components used in all treatment approaches include addressing protective and risk factors. Emphasis is placed on providing assistance to parents and caregivers and including them in treatment. Psychoeducation, cognitive restructuring, exposure, and coping skills management are frequently used treatment strategies. Guidelines and suggestions specifically related to media coverage of terrorist events are also crucial interventions to prevent the exacerbation of terrorist-induced trauma. Resources for information on commonly used assessment tools following a terrorist attack are provided in Table 54.11–1. Resources for information on children and terrorism are provided in Table 54.11–2. Guidelines for parents, health care providers, and educators are listed in Table 54.11–3, and guidelines related to media coverage are outlined in Table 54.11–4.

Community-Based Interventions

Based on their research into the effects of the 1995 Oklahoma City bombings, Robin Gurwitch and colleagues identified essential responses to community-wide acts of terrorism that included early community-based intervention; clinical needs assessment to identify children at risk; multimodal, trauma-loss–focused treatment programs; and program evaluation of treatment efficacy. Early community-based interventions focus on safety and protection. Restoration of rest and sleep, emotional support by parents and caregivers in the community, and stress-related symptom reduction are emphasized. Community-wide screening is critical to identifying children at risk. Screening may take place in schools, primary care settings, or neighborhood centers. Although there is no consensus on the most effective method of assessment, numerous structured interviews and self-reports are available. Screening identifies children at risk and children in need of acute, trauma-related services. An essential component of the screening process is distinguishing normal, developmentally appropriate reactions to trauma from abnormal reactions. Screening information may be used to develop specialized treatment programs based on the unique needs of the child, family, or community. Screening results also serve to provide pretreatment data for evaluating outcome studies of treatment effectiveness. Triage and referral, based on knowledge of psychopathology and risk factors, are essential components of the intervention. Active involvement of parents, caregivers, health care providers, educators, and community leaders is emphasized.

Table 54.11–1.
Commonly Used Assessment Instruments

Posttraumatic Stress
Child PTSD Symptom Scale (CPSS)
Diagnostic Interview Schedule for Children/Diagnostic Predictive Scale (DISC/DPS)
Impact of Event Scale Revised (IES-R)
University of California at Los Angeles Post-Traumatic Stress Disorder Reaction Index (UCLA PTSD RI)

Anxiety
Child Behavior Checklist (CBCL)
Diagnostic Interview Schedule for Children/Diagnostic Predictive Scale (DISC/DPS)
Revised Children's Manifest Anxiety Scale (RCMAS)
State-Trait Anxiety Inventory for Children (STAIC)

Depression
Center for Epidemiological Depression Scale (CES-D)
Child Behavior Checklist (CBCL)
Children's Depression Inventory (CDI)
Diagnostic Interview Schedule for Children/Diagnostic Predictive Scale (DISC/DPS)

Behavior
Behavior Assessment Scale for Children (BASC)
Child Behavior Checklist (CBCL)
Diagnostic Interview Schedule for Children/Diagnostic Predictive Scale (DISC/DPS)
Strengths and Difficulties Questionnaire (SDQ)

From Pfefferbaum B, North CS. Child Disaster Mental Health Services: a review of the systems of care, assessment approaches, and evidence base for intervention. *Curr Psychiatry Rep* 2016;18:5. doi: 10.1007/s11920-015-0647-0.

Community and school interventions, family therapy, adult therapy, and group therapy have been used to treat traumatized children and families, help them with coping skills, and address issues related to trauma and loss, but no systematized studies have shown their effectiveness. Psychoeducational group meetings (i.e., parents or community meetings) reduce symptoms of individual, family, and community fear and arousal levels.

Community resilience has emerged as a way to conceptualize how communities can adapt and thrive post-trauma. One approach to assessing a community's needs and relative strengths include the Communities Advancing Resilience Toolkit (CART), which was developed by the Terrorism and Disaster Center (TDC), in Missouri and at the University of Oklahoma Health Sciences Center (OUHSC) in Oklahoma. Theory-based and evidence-informed, the toolkit consists of instruments for collection and assessment to support intervention implementation. The CART Assessment Survey is one such instrument and can be used to obtain baseline information about a community, identify strengths and challenges, and examine change over time after a disaster or intervention. This toolkit is noted to be multilayered and may be effective even in low-resource settings such as refugee camps.

School-Based Interventions

School-based interventions to treat PTSD symptoms have received empirical support in the literature. Annette La Greca and colleagues developed a school-based intervention manual to help children cope after Hurricane Andrew. Their manual has been distributed to schools internationally and has been used in other disasters, including the Oklahoma City bombing. Robin Gurwitch and Anne Messenbaugh expanded the work of La Greca and colleagues after Hurricane Andrew and the work of Storm and colleagues after the Australian bushfires. They developed a manualized treatment program for use following traumatic community events, including acts of terrorism. The program, Healing After Trauma Skills (HATS), emphasizes the active involvement of parents and caregivers and includes educational materials on basic safety skills and psychological stress responses. It uses treatment exercises to address symptoms and behavioral difficulties associated with trauma and loss. Similar manuals and materials have been developed in the aftermath of the September 11 terrorist attacks. Among the materials that have been widely distributed internationally are *Helping America Cope: A Guide to Help Parents and Children Cope with the September 11th Terrorist Attacks* by La Greca and colleagues and the American Red Cross publication *Facing Fear: Helping Young People Deal with Terrorism and Tragic Events*. The preliminary results of the interventions and manuals developed to aid children after terrorist attacks are promising. However, outcome data have not been examined, and further research is needed to evaluate their effectiveness.

Table 54.11–2.
Mental Health Resources on Children and Terrorism

American Academy of Child and Adolescent Psychiatry: http://www.aacap.org

American Academy of Pediatrics: http://www.aap.org

American Psychological Association: http://www.apa.org

American Red Cross. *Terrorism. Preparing for the Unexpected.* https://www.redcross.org/content/dam/redcross/atg/PDF_s/Preparedness___Disaster_Recovery/Disas

National Advisory Committee on Children and Terrorism (NACCT). https://health.mo.gov/emergencies/pediatrictoolkit/CommunityPlanningforChildren/NationalAdvisoryCo

Center for the Study of Traumatic Stress: http://www.centerforthestudyoftraumaticstress.org

Federal Emergency Management Agency (FEMA. *Emergency Planning with Children.* https://www.fema. communities/children

National Academy of Sciences. *Preparing for the Psychological Consequences of Terrorism—A Public H* https://www.ncbi.nlm.nih.gov/books/NBK221643/

International Society for Traumatic Stress Studies (ISTSS): http://www.istss.org

National Child Traumatic Stress Network, *Psychological First Aid Field Operations Guide*: https://omh.ny.gov/omhweb/disaster_resources/pandemic_influenza/doctors_nurses/physchological_fi

National Institute of Mental Health, U.S. Department of Defense, Justice, and Veterans Affairs, and the A *Mass Violence: Evidence-Based Early Psychological Intervention for Victims/Survivor of Mass Violenc* http://www.nimh.nih.gov/research/massviolence.pdf

National Mental Health Association: http://www.nmha.org

National Mental Health Information Center, *Psychosocial Issues for Children and Families in Disasters: A* http://mentalhealth.samhsa.gov/_scripts/printpage.aspx?FromPage=http%3A/mentalheatlh.samhsa.go

New York University Child Study Center: http://www.aboutourkids.org

Substance Abuse and Mental Health Services Administration (SAMHSA). Mental Health Response to Ma Manual https://store.samhsa.gov/product/Mental-Health-Response-to-Mass-Violence-and-Terrorism-A-

U.S. Department of Education. Practical Information on Crisis Planning Brochure. https://www2.ed.gov/a

U.S. Department of Health and Human Services, *Terrorism and Other Public Health Emergencies: A Ref* https://www.calhospitalprepare.org/sites/main/files/file-attachments/hhsmedisreferenceguidefinal.pdf

Increasing attention has been paid to social-emotional competence as a focus of primary prevention efforts and the basis for multiple studied interventions that schools can offer at-risk youth, given its role as a protective factor against mental health outcomes for youth exposed to trauma. Most studies of SEC programs include nonreplicated pilot studies with mixed but promising results.

Selective interventions have more evidence. A recent review by Yohannan and Carlson found school-based CBT and ERASE-Stress interventions feasible and with high rates of fidelity and noted that school-based CBT had high rates of acceptability. Enhancing Resiliency Amongst Students Experiencing Stress program uses a combination of CBT, bioenergetic exercises, meditation, art therapy, and narrative therapy and has been studied in a variety of settings post disasters, war, and terror-related traumatic events.

Cognitive Behavioral Intervention for Trauma in schools (CBITS) was designed and studied by Jaycox and colleagues to be delivered in inner-school settings by school staff in order to facilitate access and ease of treatment; it has been used as a selective intervention for at-risk youth and has been studied with youth from diverse backgrounds. CBITS is a 10-week small group treatment program with one to three individual sessions dedicated to creating a trauma narrative. The key elements of CBITS include core CBT techniques: psychoeducation regarding the effects of trauma, relaxation, cognitive restructuring, graduated in vivo exposure, and social problem-solving. Not specifically designed for post-terrorism trauma, CBITS has been effectively culturally adapted in various communities, such as with Latino and Asian immigrants, African Americans in low SES settings, and American Indian youth.

Individual and Family Interventions

Children and families who manifest significant psychiatric impairment and dysfunction need more intensive health treatment. A variety of treatment modalities have been utilized and studied in traumatized children. Cognitive-based techniques, including trauma-focused CBT, have been the most rigorously studied treatment for traumatized youth in general, although other treatment modalities have an evidence base, such as eye movement desensitization and reprocessing, prolonged exposure, and narrative exposure. Most reviews comparing interventions do not find a difference between active groups, although there remains little evidence for debriefing as an early intervention.

Regardless of technique, interventions should be developmentally appropriate, promote safety and social support, build resilience, and include the family when appropriate, particularly given the association between parents' reactions and the severity of the child's posttraumatic symptoms. Families have a role to play in boosting resilience and are included in most evidence-based trauma treatment for youth. Therapeutic modalities are frontline with medications utilized only as adjunctive or for comorbid disorders.

Most interventions for youth within three months of a traumatizing event come from studies of war-affected youth or following a natural disaster rather than a terrorist or school shooting incident. One exception is a study of hypnosis-assisted therapy conducted by Lesmana and colleagues on children exposed to a terrorist bomb attack in Bali in 2002. This study built upon hypnosis as a technique utilized for trauma treatment in various settings and added a spiritual component for cultural relevance. PTSD symptoms were significantly decreased in 77% of the group compared with 24% of the control group.

Table 54.11–3.
Guidelines for Parents, Health Care Providers, and Educators: Terrorism and Trauma

Helping children and adolescents overcome psychological problems in the aftermath of terrorist events is an important challenge for families and communities. Proactive and preventative methods to help children to cope with loss and anxiety may help to facilitate positive posttraumatic adjustment and allow for healthier outcomes in children's growth and development. Several important general guidelines and suggestions are recommended for parents, health care providers, and educators to help children in the aftermath of terrorist attacks.

Adults should focus on positive ways of understanding and coping with their responses to terrorist events. Since September 11, 2001, many resources have been made available to help parents to cope with the terrorism-related fears and anxieties. The National Mental Health Association, Centers for Disease Control and Prevention, and American Academy of Child and Adolescent Psychiatry are among some of the agencies that have published extensive information.

Any parent who is overwhelmed by the effects of the disaster should seek help from primary care providers, educators, mental health agencies, and clergy. If parents are not able to manage their responses to the trauma, they will not be effective in reassuring their children and helping them to discuss their concerns.

Children cannot be sheltered entirely from knowing about or reacting to the disturbing event. Their fears of not knowing or understanding what has happened may be more disturbing than the truth. It is therefore important for parents to present information to children according to their developmental level. Children should be encouraged to express their concerns, and their feelings should be validated. Listening is more important than talking. All questions should be treated with respect and not ignored or dismissed. Responses should be made with calm reassurance and empathy. If a child's question causes anxiety or discomfort, the adult should not remain silent, because the child might interpret the silence as a sign of danger. It is permissible to tell the child that the question cannot be answered immediately but will be responded to after the adult has had time to think about it.

Younger children may act out their feelings rather than verbalize them. When possible, children should be encouraged to express their feelings in words. Answers to their questions should be honest and brief. More information than is requested should not be provided.

Older children may be hesitant to talk about their feelings. Discussing terrorist activities with them, without overwhelming them, and encouraging questions or correcting misperceptions are helpful.

Some children's and adolescents' responses may be confusing. Their defenses may be covering up their fears and anxieties. For example, they may make comments such as "We should not be upset, nothing has happened to us." Just as adults should be aware of their own defenses and coping styles, they should be sensitive to children's defenses and ways of coping.

If a child chooses not to talk about the events, the decision should be respected. Children may not be ready to share their experiences immediately; they may wait several days or weeks. Adults should make sure that children know that they are available to listen and discuss their feelings whenever the children need them.

Adults need to reassure children that they are safe and will be protected. Maintaining familiar routines helps children to know what to expect and is comforting to them. Reassuring children that they are loved and that they will be cared for will give them a sense of security. However, they should not be offered false assurance. Children should be taught to manage their fears and anger constructively by showing them how to handle their emotions in healthy ways. They can be made to feel more in control by allowing them to make decisions (meals, what to wear, what movie to see, etc.)

Children may be encouraged to draw pictures or write about their feelings. An important way for children to work through their worries is play. At times, it may be helpful for an adult to redirect the play into caring or helpful themes (such as pretending to make meals for emergency helpers or building hospitals). Focus and attention should be given to those who are helping in the community and nation at that time. Children may feel empowered by participating in events to help those who have suffered. Offering children opportunities to channel their feelings into positive actions allows for healthier outcomes.

Children should be given a reasonable sense of hope for the future. An example may be giving children historical references to how a country overcame previous threats to become a strong nation. Other examples include sharing past problems that children have had with classmates or family members and how they resolved their difficulties, or by giving examples of conflict and resolution within the community.

Children should be educated about racism and stereotyping. They should be told that discrimination and prejudice are unacceptable. Children learn from adults, who must be careful to monitor their own morals and values in front of their children.

Most children will develop healthy coping skills if given the support and reassurance from adults. Some children need more help, especially if they are at greater risk or have experienced the loss of a loved one. Children who may need more intensive help should be referred to a mental health professional. Children's prolonged responses to anxiety and stress may interfere with their ability to function at home, at school, or with peers. The signs of stress may be similar to those seen in adults: Avoidance (refusing to go places that remind them of a traumatic event), emotional numbing (diminished response or lack of feeling), re-experiencing trauma (through nightmares or daytime recollections), and hyperarousal (tendency to be easily startled, sleep disturbances). Children's responses also may differ from those of adults and may include disorganized or agitated behavior rather than fear and helplessness, constant repetitive play of themes of trauma that preoccupy and distract them from their daily activities, and frightening dreams that are not related to terrorist activities. In these cases, children should receive support and help from a mental health professional to prevent these symptoms from continuing to cause problems in the future as children grow.

Trauma-focused cognitive-behavioral therapy (CBT) has been the most rigorously studied treatment for traumatized children in general and has been studied for use in children as young as 3. Randomized, controlled studies have shown the effectiveness of CBT in targeting symptoms of PTSD and anxiety in youth following different types of abuse. Cognitive restructuring (reprocessing the traumatic event and identifying traumatic triggers), relaxation training, anger management skills, and grief management are important methods of treating PTSD after disasters. However, the effectiveness of CBT for treating children exposed to chronic trauma, past history of trauma exposure, and serious family dysfunction has not been studied.

Interventions such as ARC (attachment, regulation, competency) and ICT (integrative treatment of complex trauma) have been developed for individuals with early, developmental, and complex trauma.

Minimal data are available on the effectiveness of pharmacotherapy for treating symptoms of PTSD in children. Although other selective serotonin reuptake inhibitors (SSRIs) have not been approved by the FDA, they have been shown to have therapeutic effects on symptoms of depression in children. Because symptoms of anxiety and depression are common in children following trauma, treatment with SSRIs has been considered a treatment option. However, caution must be used because of concern about the potential increase in suicidal ideation when these medications are used in children. Close monitoring for adverse side effects of suicidal thoughts is necessary when prescribing SSRIs to young patients. Given known hyperarousal symptoms following traumatic events, medications that target this system may be helpful adjuncts. Commonly used medications may include alpha agonists to help down-regulate the hyperarousal system and target anxiety and attention difficulties or to help with sleep difficulties.

Table 54.11–4.
Guidelines Related to Media Coverage

Because the media plays a major role in informing the public about terrorism and may be a causative agent in symptom formation, it is critical for parents and families to have guidance about the use and misuse of media in traumatic times. Several media-related guidelines have been identified to help prevent traumatizing children specifically from media coverage of terrorist events. Parents, educators, health care providers, politicians, and journalists must increase their awareness about the potentially harmful effects of media exposure on children. Parents should closely monitor disaster-related television viewing by children.

Children's exposure should be limited. Adults would also benefit from limiting their own exposure to reduce their stress reactions so that they and their young children are better able to cope. General recommendations on limiting screen time: young children 18 months or younger screen time only for video chatting along with an adult; 18–24 months, watching educational programming with a caregiver; 2–5 years of age, limiting children to one hour per weekday and 3 hours per weekend for noneducational screen time; children ages 6 and older, encouraging healthy habits that limit screens for children. Recommendations for mitigating the negative effects of news coverage include: watching news with child when appropriate, asking child what questions they have, provide reassurance regarding his/her safety in simple words, look for signs that news may have triggered fears or anxieties.

Vulnerable persons who have risk factors, such as young children and parents with previous traumatic experiences, should be especially cautious. The use of media should take into consideration a child's developmental level. Young children should not watch adult-oriented shows. Older elementary school-aged children may tolerate being exposed to news "briefs" in which simple concise statements of the facts are made (e.g., reports that occur on major networks between television shows) but may not be ready to watch news shows that spend 30 minutes with interviews, photos, and more detailed discussion of devastation.

Parents and caregivers are advised to watch the coverage with children when they are permitted to watch television. This supervision provides adults with the opportunity to observe children's signs of distress and be available to answer questions and discuss relevant topics.

School staff also must be aware of the importance of monitoring media exposure and should consider policies related to viewing coverage of traumatic events.

Constructive uses of the media should also be taken into consideration. The media may be helpful for parents by providing them access to valuable information to help them when dealing with children in the aftermath of terrorist attacks. Media reports can be used to help reassure children that many adults are working to maintain a safe and protected environment for them Adults may choose to offer taped news reports or specific types of media discussion that will help a child feel secure in the knowledge that adults are taking charge and responding in ways that are protective. Films about resilience, survival, family and community consolidation, and triumph over adversity may play a constructive role by giving children a reasonable sense of hope for the future. Long after an event, children may watch shows with their parents to help them understand terrorism, for example, "Through a Child's Eyes; September 11, 2001," which is an Emmy Award-winning HBO children's special that deals with young school-age children's responses to the September 11 attacks.

Certain limitations exist when applying trauma-specific interventions to children who have experienced ongoing violence, whether in the form of complex trauma, community violence, or the continual threat of terrorism. Crisis intervention alone does not prevent the long-lasting effects on children exposed to terrorism. Compounded effects of multiple traumas have not been addressed. The unpredictability of terrorist attacks, the indefinite nature of violent threats, the lack of specific geographical boundaries, and the effects of media coverage result in unique stressors and pose specific challenges for the treatment of terrorism-induced trauma in children.

FAMILY TRAUMATIC GRIEF

Ben was a 7-year-old boy whose maternal uncle, a firefighter, died in the September 11, 2001, World Trade Center attacks. His uncle did not return to his company on that day. Ben was first seen in treatment 1 month after September 11. He was at school when the attacks occurred. His father picked him up. After they returned home, Ben, his parents, and his 5-year-old sister watched the news and the repeated images of the falling towers. Ben knew that his uncle was a firefighter, but he did not know whether he had responded to the attacks. Ben's mother was very distraught as she tried desperately to find out what had happened to her brother. Ben's father stayed home from work during the first week after September 11 and assumed most of the household chores and child care, which his mother had been primarily responsible for prior to the attacks. Ben was very worried about his mother. He was aware that the family was concerned about his uncle, and he asked frequent questions about him. His parents had been ambivalent about what to tell their children. Initially, when they did not know for certain whether Ben's uncle had died, they were hesitant to accept the possibility of his death and maintained hope that he was alive. They told their children that they were hopeful that he would be found. Both parents made a conscious effort to make sure that their children were not exposed to television coverage. However, after the children were in bed, they spent several hours watching the news coverage. They talked to their family and friends by phone and tried to protect their children by talking in another room while the children were playing. After Ben's parents were notified of his uncle's death, they struggled with their feelings of shock and grief. They did not tell their children about their uncle's death for several days. By continuing to "not know," Ben's parents may have reacted to their terror and loss by using protective defenses of shock and disbelief. Ben and his sister continued to ask questions: "Was Uncle Gary one of the firefighters who died trying to rescue the people in the tower?" The children's direct questioning challenged the adults' defenses of denial. Ben's mother was irritated by his frequent questions and then felt guilty for being angry with him. Ben, in turn, became upset, whiny, and oppositional. He also became very clingy and insisted on being with his parents, following them to each room of the house. He did not want to go to the bathroom by himself. At night, Ben would join his parents in their bedroom, which was very uncharacteristic of him. Ben's parents eventually told their children that their uncle had died. Ben continued to ask questions about his uncle's death. "How did he die? Did he burn to death?" His mother, in a state of traumatic bereavement, did not know how to answer his questions. Ben's father tried to reassure his son that his uncle was now in heaven.

When school resumed the following week, Ben refused to go. He complained of "belly aches" in the morning. When his parents tried to help him to get ready for school and reassure him, he started crying. He refused to get dressed for school and had temper tantrums. Ben's mother took him to school. He went willingly with her, but when it was time for him to leave her and enter the classroom, he started crying and refused to let go of her. She struggled to see him in so much anguish, and after speaking with the school counselor, she decided to take Ben home for the day. The school counselor noted that several of the children had not yet returned to school. Ben had no prior history of school problems and had always enjoyed attending school.

Ben continued to have difficulties separating from his mother and continued to refuse to go to school. He no longer asked questions about his uncle, and his parents avoided talking about him in front of their children. Ben continued to complain of stomachaches. His parents became increasingly concerned and made an appointment for him to see his pediatrician. They noticed that Ben's physical complaints were more frequent on the days he was expected to go to school, but they wanted to make sure that he did not have an underlying medical problem. Ben's pediatrician completed a full medical workup and did not find any underlying organic problems that could be contributing to his complaints of abdominal pain. She reassured Ben's parents that no further medical tests were needed. She explained to them that she believed Ben's current difficulties were related to the recent trauma that the family had experienced and referred them to a child therapist.

The child therapist met with Ben's family initially. Ben's parents focused on their worries about their son. They felt that his problems were related to the recent traumatic events and the death of his uncle. They acknowledged that they felt overwhelmed and were grieving the loss of their uncle. Ben said that he was worried about his parents, especially his mother. “She pretends she is OK, but I don’t believe her.” The therapist then met with Ben alone. Ben said he was worried about his mother. He was also afraid that another attack would occur. The therapist commented that this was a scary time, and a lot of children were frightened, sad, and mad. Ben was quiet and did not answer. While they were talking, Ben drew a picture of his uncle’s fire truck. The therapist asked Ben about his uncle. He said that he liked to spend time with his uncle, who had “awesome baseball cards.” Ben and his uncle’s family would get together every Sunday. Ben commented that the Sunday before his uncle died, he went to a friend’s birthday party while his family visited his uncle’s family. He was very sad that his uncle died, but he also felt “bad” that he chose to go to his friend’s party instead of spending time with his uncle. The therapist assured Ben that his uncle knew that Ben loved him. She explained to Ben that he did not cause his uncle’s death. Ben responded, “I know, but I felt like I jinxed him.”

When the therapist met with Ben’s parents alone, she asked them how they were coping. Ben’s mother noted that in addition to mourning the death of her brother, she was also very frightened that another attack would occur. She had frequent nightmares and was afraid to leave the house. She was afraid for her husband’s safety. He worked long hours and commuted to his job in Manhattan by subway. Her main supports were her sister, who lived out of state. She had a few close friends, but their families were also affected by the September 11 attacks, and she did not want to be a burden to them by talking about her fears and loss. Ben’s father was also worried about his wife. He noted that he coped with stress “by being a workaholic.” Both parents continued to watch the television coverage of the events, usually at night after their children had gone to bed. They stated that they tried to avoid talking about the terrorist attacks in front of their children because they “did not want to upset their children further.”

The therapist continued to meet with Ben and his family. Family sessions were held, in addition to individual sessions with Ben and sessions with his parents. Family sessions focused on encouraging the family to communicate more directly with each other and to express their feelings of grief and anxiety. The therapist helped to normalize Ben’s parents’ reactions by explaining to them that their feelings of fear were normal reactions to an extreme situation. She encouraged them to use their supports, including their friends and family. She helped them to develop coping strategies. Ben’s mother had stopped exercising after the attacks, an activity she had previously enjoyed. Ben’s father was able to reduce his work hours and spend more time with his family during the crisis. Both parents were advised to decrease the amount of time they spent watching television. The therapist explained to them how their reactions and defenses affected their children. Ben’s parents were given guidance regarding how to talk to their children. They were encouraged to listen to their children, and they were taught how to understand their children’s feelings and behaviors. In particular, the therapist helped Ben’s parents to understand how his difficulties with separation and school refusal were related to their responses and anxieties, especially his mother’s worries. The therapist helped Ben’s parents to understand his experience and how he was feeling guilty that he may have caused his uncle’s death by not visiting him. The importance of returning to a regular routine and ordinary daily life was emphasized. This included making sure that Ben’s parents were consistent in helping him to attend school every day. To facilitate his return to school, the therapist spoke with his school counselor to collaborate on treatment. It was agreed that Ben’s mother would drop him off at school and the school counselor would meet him at the school door. Ben would meet with the counselor every day during the first week. He would also be allowed to call his mother during the first week. The therapist and school counselor agreed to collaborate and communicate regularly to make appropriate adjustments to the treatment plan as needed. In addition, once Ben returned to school, he would participate in a peer group that the school counselors had developed to help children with similar experiences and losses.

The therapist also met with Ben individually. She helped to reassure him that the adults in his life, including his parents and teacher, would do their best to make sure that he and the other children were safe. Ben’s therapist helped to normalize his feelings, explaining to him that other children also experienced feelings of sadness, anger, fear, and confusion. She facilitated the symbolic expression of his feelings by supporting and joining him in play and in art projects. She listened to him when he expressed his feelings verbally and clarified his feelings by helping him to find words to name his emotions. She avoided probing questions and followed Ben's lead. Ben's attachment and identification with his uncle were actively facilitated by helping him remember and talk about his uncle.

Lastly, the therapist offered guidance to Ben and his family regarding what symptoms to monitor over time and the recommendation around when to seek more specialized trauma-focused treatment if symptoms persist or become functionally impairing.

FUTURE DIRECTIONS

The continuing frequency of terrorist events around the world and their effect on children and families underscore the urgency for an effective public health approach to providing for the needs of children and their families. Mental health organizations and providers must be educated and prepared to help children in the aftermath of terrorist activities. Similarly, the rise in school mass shootings underscores the need for a stepped-care approach. Terrorism is a community mental health issue because terrorism is a form of psychological warfare whose target is to intimidate and cause fear. Several mental health activists have made recommendations to help communities organize and plan appropriate interventions and treatments. Collaboration among community, educational, and mental health care organizations is essential. In addition, collaborative efforts between mental health organizations nationally and internationally have been initiated and continue to grow and develop. In the United States, the mental health approach to disaster has developed over several decades, primarily based on experiences with natural disasters. Following the 1995 Oklahoma City bombings, many modifications to the federal approach were made, including more effective crisis intervention methods and better triage and referral services, to address the unique effects of violent acts of terrorism. The effect of the events of September 11, 2001, further emphasized the difficulties in identifying the extent of services needed, the scope and depth of personnel and funding necessary to provide services, and the anticipated duration of treatment. The lack of sufficient data and comprehensive information prevented adequate funding by federal mental health agencies to address mental health needs of children and their families. Obtaining money and resources will continue to be a significant challenge. Advocacy for program development and funding to increase community services for children affected by terrorism is needed.

In the general community, trained mental health providers who have experience working with children are needed to identify children at risk and provide community-wide services and more intensive treatment interventions. Standards for mental health providers who work with children and families exposed to terrorist events should be established. Within school settings, there has been increasing effort to train school staff to provide primary prevention efforts of increasing social-emotional health and boosting resiliency. In certain communities, schools have been the focus of selective interventions to identify youth at increased risk following a traumatic event such as a school shooting and in some cases, to provide selected treatment. In all settings where youth reside, more collaboration is needed to establish referral networks to specialized mental health services when needed and to offer a continuum of care to youth in need.

The correlation between media exposure to terrorist events and traumatic stress symptoms in children is well documented. Collaboration between community agencies and the media is necessary. Public policy limiting media coverage is a controversial issue but needs to be addressed.

The research on childhood reactions to terrorism-induced trauma highlights the need for a systematic public health approach that provides community-based interventions, screening of children at risk, triage and referral services, community-based trauma and loss programs, and multimodal outcome studies. During the last decade, policy and procedures for the conduct of organized interdisciplinary research have received more attention. These advances resulted in more cooperation between researchers to address the issue of how ethically to obtain research data during times of crisis. Planning to incorporate research infrastructures into disaster preparedness plans is necessary. Further research on terrorism and its unique effects on children and families should be encouraged. More research on the effects of early interventions and treatment is needed to provide reliable mental health response planning. Further research identifying children who are at risk for developing PTSD and related symptoms will allow for more efficient triage services. Studies on longitudinal effects and the long-term effect on development in children exposed to terrorist violence will provide information useful for initial and later treatment interventions. More reliable and objective measures are needed to quantify parents' reactions and the effects of social supports. Methods for evaluating a child's intrinsic resiliency will help to differentiate coping models learned from parents. Assessment of the capacity of the public health system to respond to terrorist events will help to determine optimal response recommendations for the future. Efforts should be focused on increasing community services for children affected by terrorism and encouraging further research on terrorism and its unique effects on children and families. Responses from mental health providers and organizations, educators, spiritual leaders, and community organizations will be better able to help children to cope and will minimize long-term psychological problems to enhance healthier outcomes and growth.

FURTHER READINGS

Allwood MA, Bell-Dolan D, Husain SA. Children's trauma and adjustment reactions to violent and nonviolent war experiences. *J Am Acad Child Adolesc Psychiatry*. 2002;41:450–457.

Bat-Zion N, Levy-Shif R. Children in war: stress and coping reactions under the threat of scud missile attacks and the effect of proximity. In: Leavitt L, Fox N, eds. *The Psychological Effects of War and Violence on Children*. Lawrence Erlbaum; 1993.

Coates SW, Rosenthal JL, Schechter DS. *September 11: Trauma and Human Bonds*. Analytic Press; 2003.

Comer JS, Dantowitz A, Chou T, et al. Adjustment among area youth after the Boston Marathon bombing and subsequent manhunt. *Pediatrics*. 2014;134(1):7–14.

Covell NH, Allen G, Essock S, Pease EA, Felton CJ. Service utilization and event reaction patterns among children who received project liberty counseling services. *Psychiatr Serv*. 2006;57:1277–1282.

Danieli Y, Dingman RL. *On the Ground After September 11: Mental Health Responses and Practical Knowledge Gained*. Haworth Press; 2005.

*Debellis M, Zisk A. The biological effects of childhood trauma. *Child Adolesc Psychiatr Clin N Am*. 2014;23(2):185–222.

Fairbrother G, Stuber S, Galea S, Pfefferbaum B, Fleishman AR. Unmet need for counseling services in children in New York City after the September 11th attacks on the world trade center: implications for pediatricians. *Pediatrics*. 2004;113:1367–1374.

Fallahi CR, Lesik SA. The effects of vicarious exposure to the recent massacre at Virginia Tech. *Psychol Trauma*. 2009;1(3):220–230.

Ford-Jones A, Nieman P, Psychosocial Paediatrics Committee. Impact of media use on children and youth. *Paediatr Child Health*. 2003;8(5):301–306.

Fremont WP, Pataki C, Beresin E. The impact of terrorism on children and adolescents: Terror in the skies, terror on television. *Child Adolesc Psychiatr Clin N Am*. 2005;14:429–451.

Gurwitch RH, Messenbaugh A. *Healing After Trauma Skills: A Manual for Professionals, Teachers, and Families Working With Children After Trauma/Disaster*. Children's Medical Research Foundation; 2001.

Hoven CW, Duarte CS, Lucas CP, Wu P, Mandall DJ. Psychopathology among New York City public school children 6 months after September 11. *Arch Gen Psychiatry*. 2005;62:545–552.

La Greca AM, Silverman WK, Vernberg EM, Roberts MC. *Helping Children Cope With Disasters and Terrorism*. American Psychological Association; 2002.

Laor N, Wolmer L, Cohen D. Mother's functioning and children's symptoms 5 years after a Scud missile attack. *Am J Psychiatry*. 2001;158:1020–1026.

Leavitt L, Fox N, eds. *The Psychological Effects of War and Violence on Children*. Lawrence Erlbaum; 1993.

Lesmana CBJ, Suryani LK, Jensen GD, Tiliopoulos N. A spiritual-hypnosis assisted treatment of children with PTSD after the 2002 Bali terrorist attack. *Am J Clin Hypnosis*. 2009;52(1):23–34.

Lowe SR, Galea S. The mental health consequences of mass shootings. *Trauma Violence Abuse*. 2017;18(1):62–82.

Neria Y, Gross R, Marshall R. *9/11: Mental Health in the Wake of Terrorist Attacks*. Cambridge University Press; 2006.

*Newman E, Pfefferbaum B, Kirlic N, Tett R, Nelson S, Liles B. Meta-analytic review of psychological interventions for children survivors of natural and man-made disasters. *Curr Psychiatr Rep*. 2014;16(9):462.

*Pfefferbaum B, North CS. Child disaster mental health services: a review of the systems of care, assessment approaches, and evidence base for intervention. *Curr Psychiatry Rep*. 2016;18(1):5.

Pfefferbaum B, Seale T, McDonald N, Brandt E, Rainwater S. Posttraumatic stress two years after the Oklahoma City bombing in youths geographically distant from the explosion. *Psychiatry*. 2000;63:358–370.

Schlenger WE, Caddell JM, Ebert L, Jordan BK, Rourke KM. Psychological reactions to terrorists attacks. Findings from the national study of Americans' reactions to September 11. *JAMA*. 2002;288:581–588.

Schuster MA, Stein BD, Jaycox LH, Coliins RL, Marshall GN. A national survey of stress reactions after the September 11, 2001 terrorist attacks. *N Engl J Med*. 2001;345:1507–1512.

*Stene LE, Schultz JH, Dyb G. Returning to school after a terror attack: a longitudinal study of school functioning and health in terror-exposed youth. *Eur Child Adolesc Psychiatry*. 2019;28(3):319–328.

Strøm IF, Schultz JH, Wentzel-Larsen T, Dyb G. School performance after experiencing trauma: a longitudinal study of school functioning in survivors of the Utøya shootings in 2011. *Eur J Psychotraumatol*. 2016;7(1):31359.

Teicher MH, Samson JA, Polcari A, McGreenery CE. Sticks, stones, and hurtful words: relative effects of various forms of childhood maltreatment. *Am J Psychiatry*. 2006;163(6):993–1000.

Thabet AM, Abed Y, Vostanis P. Emotional problems in Palestinian children living in a war zone: a cross-sectional study. *Lancet*. 2002;359:1801–1804.

Yehuda R. Risk and resilience in posttraumatic stress disorder. *J Clin Psychiatry*. 2004;65(Suppl 1):29–36.

*Yohannan J, Carlson JS. A systematic review of school-based interventions and their outcomes for youth exposed to traumatic events. *Psychology Schools*. 2019;56(3):447–464.

54.12 Impact on Parents of Raising a Child with Psychiatric Illness and/or Developmental Disability

ALICE R. MAO, M.D., MATTHEW N. BRAMS, M.D., KATHRYN FORNERIS, D.O., M.P.H., AND PIETER JOOST VAN WATTUM, M.D., M.A.

The birth of a child is generally anticipated with joy and many complex emotions. Parents have conscious and unconscious expectations about the child's looks, temperament, personality, and other characteristics based on past experience. Expectations develop about the child's developmental course and future, including education, career, and generational continuity. Parents also have expectations for themselves in their new roles, which are shaped by past experiences and other important role models. One of the important tasks of parenthood is the adaptation and modification of these fantasies and hopes to match the reality of the actual child and their abilities.

The task of raising a child requires fostering autonomy while providing appropriate limit-setting at each developmental stage. The child has an increasing need for separation and individuation as they move into adolescence and then into adulthood, which must be combined with nurturing parental structure and support. Ultimately, the goal is for the child to become an adult who is able to successfully form relationships, support themself, and take on a socioculturally defined mature family and societal role. Unfortunately, if a child is not developing normally or is suffering from a psychiatric disorder that is affecting multiple domains of functioning, their ability to achieve these goals is compromised and caregiver burden is increased.

When a child is diagnosed with a psychiatric or developmental disorder, parents typically undergo a complicated adjustment and grieving process. This may include anger or sadness that their child is not developing normally, coping with the narcissistic injury that the “imperfect” child is a negative reflection of their parental competency. In addition, parents may irrationally blame themselves for causing a disorder to manifest in their child. Disorders with heritable causative factors may lead to parental self-blame and/or blame from nonaffected family members. The contrast between the imagined parenting experience of raising a normally developing child and the reality of caring for a child with a psychiatric and/or developmental disorder may lead to intense disappointment and increased caregiver burden.

This section reviews the effect on parents and caregivers of raising a child with serious psychiatric disorders, specifically bipolar disorder (BPD), attention-deficit/hyperactivity disorder (ADHD), schizophrenia, and autism/developmental disorders. Unique stressors faced by parents raising children with these disorders are summarized. Psychological responses to these stressors and research on specific interventions developed to help parents and family members cope effectively are reviewed. Individual psychotherapeutic work with parents of psychiatrically ill children may be enhanced by understanding the life-changing consequences of dealing with a chronic mental illness or developmental disability. Importantly, the parent’s ability to implement complex and multimodal treatment interventions for their child is affected by their responses to their situation. Early intervention for parents after their child receives a psychiatric diagnosis may improve the long-term outcome for the child.

IMPACT ON PARENTS OF RAISING A CHILD WITH AN AUTISM SPECTRUM DISORDER

In the mid-1940s, Leo Kanner and Hans Asperger independently reported on disorders in childhood that were characterized by repetitive or stereotyped behaviors, language abnormalities, and difficulties in social interaction. Initial reports included observations of similarities between parents and their affected children. Asperger suggested possible genetic factors; however, prevailing theories attributed the children's symptoms to their parents' behaviors. Currently, autism spectrum disorders (ASDs) are clearly recognized as biologically based neurodevelopmental disorders.

Unique Stressors

Psychiatrists caring for parents of people with ASDs will be better prepared if they are aware of the multiple issues that affect families. This subsection points out the unique stressors associated with raising a child diagnosed with ASD; highlights issues related to the life cycle of the family; summarizes findings from studies of parents coping with raising a child with ASD; discusses the psychiatric disorders common in these parents, including the "broad autism phenotype" (BAP); and suggests some approaches to therapeutic work with this population.

Parenting a child with special needs, especially when accompanied by an intellectual disability (ID), is stressful. Stressors uniquely impacting the parents of children with ASD include diagnostic uncertainties, the prolonged duration of the diagnostic process, the extreme variability in the child's developmental course, the effect of the child's atypical social communication despite the child's typically normal physical appearance, and the burden of creating and organizing the child's treatment program.

Diagnosis of ASD is often delayed when primary care providers do not screen for or recognize signs of developmental disorders. Referrals to subspecialists may be time-consuming and lack definitive results. Siklos and Kerns reported that parents saw an average of 4.5 professionals and waited almost 3 years to obtain a diagnosis. More than 50% of parents were dissatisfied with the diagnostic process. Geographic variations in the prevalence of ASD, as well as rates of children receiving school-based special education following diagnosis of ASD, are well documented, suggesting that state policies, community resources, and community-level demographic factors impact the diagnosis and treatment of ASD. In 2005, Mandell and Palmer reported varying prevalence rates of ASD among states based on education-related spending, the number of school-based health centers, and the number of pediatricians in the state. In 2020, Safer-Lichtenstein, Hamilton and McIntyre found that state-level demographic factors, including household income, average adult educational attainment, and political leanings, also impacted the rates of ASD diagnosis and treatment. Obtaining a timely diagnosis and accessing subsequent care are important for parents. Suboptimal disciplining techniques used prior to the ASD diagnosis being made can result in caregiver loss of confidence and self-efficacy.

The spectrum of autism disorders is very broad; therefore, predictions about a specific child’s prognosis are imprecise. Irregular patterns of skill development are common. Developing appropriate expectations for a child and predicting future development is challenging for most clinicians because of wide variations in developmental progression.

Of the behaviors associated with the diagnosis of ASD, abnormalities in social communication appear to be a particular source of distress for families. A study performed by Safronoff and Farbotko in 2002 found that autistic children’s difficulty with verbal and nonverbal communication results in parents feeling unable to meet their child’s wants or needs. The inability to understand her child and their behavior is reported significantly more frequently by mothers of autistic children than control mothers.

Lack of consistent behavior across time, people, and environment is frustrating and confusing for parents. Children with more obvious impairments, such as those affected with Down syndrome, are excused and tolerated by others, but the typically normal physical appearance of a child with autism leads others to expect appropriate behavior. Fear of their child’s unpredictable, disruptive, and unusual behavior, as well as others’ responses to it, leads parents to avoid public situations, which in turn leads to restriction of family activities, social isolation, and increased burden in completing everyday household chores, such as grocery shopping.

Typical recommended interventions for children with ASD are obtained from several different systems—public education, ancillary health care therapists, hospital-based, center-based, and community-private practitioners. The family is typically left in charge of coordinating care with a complicated network of professional providers who have differing or conflicting opinions, recommendations, physical locations, and financial/insurance requirements. The lack of curative interventions from traditional care providers, the lack of empirically supported therapies, the nature of ASD symptoms, and the chronicity of these disorders have led to the emergence of unconventional treatments by alternative practitioners with poorly supported claims of cure or rapid improvement, further complicating an already convoluted process of establishing appropriate care for the affected child. Thus, the family is confronted with the challenge of obtaining intensive treatment for their child during a critical time frame with little professional guidance.

Lifespan Issues

Each developmental stage presents the family with unique and specific issues. Early childhood is characterized by the family's realization of an abnormal developmental trajectory, the diagnostic process, and issues of definitive diagnosis. Parental guilt, mourning the loss of raising a normally developing child, the unexpected negative responses of extended family members to the child's unusual behaviors, and the possible additional diagnosis of ID are all issues that color this period. The child's irregular sleep patterns, hyperactive behavior, lack of effective communication, peculiar eating and elimination habits, as well as safety issues, and the lack of ability to obtain appropriate child care, are practical concerns facing most parents. Accepting the diagnosis, finding community services, and planning for the future are tasks for the family. Families report that education about the disorder and information about resources were the most helpful interventions they received during this period. Other useful resources were learning and behavior assistance, speech therapy, and family support.

With a good school program, middle childhood may be a relatively easier life phase. Behavioral problems often improve; families have adjusted their expectations of the child and have accepted the delays in development, abnormal social interaction and communication. Developing an educational program that includes the necessary special education services, offered in the setting that is most conducive to the child's learning, is the goal of collaboration with the school system. The focus may be on the development of adaptive behaviors in the areas of self-care, domestic chores, and social behaviors. However, parents may grieve anew as they recognize the child's inability to participate in many social activities because of their anxiety, language, or developmental differences. Sibling issues can occur at any time; however, they are common during this time period. The affected child behaves oddly, is not a typical playmate, and requires much of the parents' time and energy. Parents often underestimate siblings' worries and overestimate siblings' abilities to understand the implications of an ASD diagnosis.

By adolescence, almost all families report worries about their child's future. Public education services end, and adult services may be uncertain. Behavioral concerns include aggression, increased sexuality, and an increased sense of independence without the skills necessary for self-management. Higher-functioning children may become aware of their differences from peers. Nonaffected siblings may be leaving home, creating a higher burden of care for parents who have relied on their help. The aging of grandparents may also lead to less support and greater family strain.

The most pressing issues for adulthood are appropriate living arrangements and vocational opportunities. The move toward deinstitutionalization has eliminated the option of residential placement for most families; thus, many adults with autism continue to live with their families. There are limited opportunities for appropriate social experiences and employment. Young adults with comorbid ID may find services through this diagnosis. Individuals with higher cognitive skills often continue to lack job-related skills and remain un- or underemployed. Seltzer et al. found that 57% of parents aged 50 or older live with their child with a developmental disorder versus 16% of normally developed children. Research performed by Anderson et al. further explored rates at which young adults with ASD, specifically, live with their parents or guardians, noting that individuals with ASD between the ages of 21 and 25 years old were more likely to live with a parent or guardian and for a longer duration than their peers with intellectual disabilities, learning disabilities, or emotional disturbances.

Guardianship and long-term estate planning are worrisome issues for parents. The divorce rate has been found to be higher in parents of autistic children when compared with parents raising normally developed children (25% vs. 13%). Distinct from typically developing children, where the divorce rate is high until approximately age of 8, the likelihood of divorce remains high through early adulthood for parents of children with autistic spectrum disorder. The financial strain for therapeutic interventions cannot be overlooked; In 2014, Buescher, Cidav, Knapp, and Mandell estimated the cost of caring for an individual with ASD across the lifespan at $1.4 million to $2.4 million USD, with costs increasing for individuals diagnosed with concurrent ID.

Psychological Responses to Raising a Child with an ASD

Stress. Many published studies have documented that parent of children with ASD report more stress than parents of children with other developmental disorders or parents of typically developing children. A variety of factors have been found to cause and/or mediate this finding, the most important of which is the gender of the parent. The preponderance of studies examining gender differences in parental response soundly endorse the observation that mothers report significantly more stress than fathers in general and within given family units. Explanations for this finding include those mothers shoulder the greater burden of child care, even if they work outside the home; perceived parenting competence contributes more to mothers' self-esteem than to that of fathers; and fathers use more effective coping strategies than mothers.

Most studies support a positive correlation between the amount of stress mothers report and the behavioral problems of the child. These behaviors include oppositionality, noncompliance, aggression, irritability, and unpredictability. None of these behaviors are core features of ASD, but they often accompany these disorders. Neither the severity of the ASD per se nor the level of adaptive functioning were found to correlate with reported stress or poor maternal outcome. However, these were not universal findings, as other authors have found that autism severity or the number of services the child received, which was interpreted as a measure of autism severity, did correlate with levels of maternal stress.

Several studies have aimed at understanding factors other than the child's autism that affected the well-being of individual family members. Factors such as the partner's depression and anxiety were positive predictors of stress for both mothers and fathers. Attributes of the affected child were predictors for maternal but not paternal stress. Factors that have been found to decrease parental stress include positive coping styles, belief in the efficacy of treatment, and social support. Parent education and the acquisition of behavioral management skills have also been found to be beneficial. Greater marital quality is predictive of lower levels of stress.

Coping. Coping describes "a process by which an individual responds to threats of stress. Coping strategies are generally divided into two categories; (1) emotion-focused strategies, which aim to reduce or mask feelings of distress, including engaging in denial, hoping for miracles, fantasizing, using food or drugs, distancing, and avoiding; and (2) problem-focused strategies, which aim to solve or change the problem, including confrontive coping, fighting for what is wanted, expressing anger and other feelings, taking chances, and engaging in positive reappraisal. Several studies have examined the consequences of particular coping styles for parents of ASD children. Emotion-focused styles have been found to correlate with increased depression, isolation, and marital problems. Conversely, problem-focused coping was associated with lower levels of depression. Pottie found that escape, blaming, withdrawal, and helplessness lead to decreased positive mood, whereas problem-focused strategies, including social support, positive reframing, emotional regulation, and compromise coping, resulted in increased positive mood. Moreover, coping style more robustly predicted levels of depression than locus of control and social support, although these factors were also contributory.

Research on coping over time suggests that, in general, young adults are more likely to use problem-focused strategies, while middle-aged adults are more likely to use emotion-focused methods. Gray found evidence for this in his longitudinal study of coping with children with ASD. Religion and faith became much greater in relative importance as a coping strategy; attachment to the child and greater appreciation for their good qualities also emerged as more common strategies. In general, the use of an array of coping strategies is beneficial; however, Gray found that parents used fewer strategies over time. Studies show mixed findings regarding increased effectiveness of coping as parents age. Fitzgerald reported that advanced age of affected ASD children was associated with better maternal mental health, implying adaptation on the part of the mothers. However, Smith and coworkers found high levels of depression in parents of both toddlers and adolescents.

A rare focus on fathers revealed that fathers of children with ASD or Down syndrome reported more use of wish-fulfilling fantasy and information seeking than control fathers. Different styles of paternal coping have been hypothesized as the explanation for decreased levels of stress reported by fathers compared to mothers. Studies show that fathers use avoidance and engagement in outside activities more frequently as coping strategies than mothers.

Tunali and Powers proposed a model of cognitive strategies that would facilitate coping for mothers of children with autism. These strategies included greater emphasis on being a “good mother” rather than on career development; greater frequency of recreation with family members rather than individual recreation; less emphasis on individual leisure; less emphasis on others' opinions of child-rearing practices; greater emphasis on parenting and partner support than on intimacy and shared leisure in marital relationships; and higher tolerance of ambiguity.

Quality of Life. Several authors have used the conceptual framework of "quality of life" (QOL) to understand the impact of rearing a child with an ASD. The World Health Organization defines QOL as "individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards, and concerns." QOL is a multidimensional concept that encompasses physical health, mental health, relationships, personal beliefs, and values in relationship to the environment. A significant body of research spanning cultures and countries indicates poorer QOL in parents of children with ASD when compared to parents of typically developing children and parents of children with other disabilities. These differences in QOL could be related to several factors, including stressors related to raising a child with autism, decreased social support for parents of children with ASD, and genetic and environmental factors affecting the parents of children with ASD, themselves, that may impact their perception of QOL.

On the World Health Organization Quality of Life-BREF, parents of children with pervasive development disorder (PDD) were compared to parents of children with mental retardation (MR), cerebral palsy (CP), and typical development. Parents of children with PDD showed a significant impairment in QOL as compared to the other groups, whereas few differences were noted between the MR, CP, and typically developing groups. Within the PDD group, mothers had lower QOL than fathers. These differences were posited to be caused by the stress of raising children with severe developmental disorders, the presence of the BAP in the parents or other family members, and parental psychiatric disorders.

Analysis of data from the 2003 National Survey of Children's Health (NSCH) revealed significantly poorer QOL in families of children with ASD when compared to families of children with ADHD or typical controls, most notably in the areas of burden of care and isolation. In a population-based study of psychological functioning and coping, mothers of children with ASD rated their overall health as comparable to mothers of nonaffected children, but they reported poorer mental health. In a study of health-related QOL in parents of children with high-functioning autism (HFA) or Asperger syndrome, mothers had lower physical health scores than fathers of affected children and control mothers.

Psychiatric Disorders and the Broad Autism Phenotype

As data were gathered supporting the role of genetics in autism, the idea that the genetic liability for ASD may also be expressed in nonautistic relatives of autistic probands with qualitatively similar but milder behavioral and cognitive features of autism gained momentum. These phenotypic features, including social impairments, communication impairments, and rigidity, constitute the concept of the “broad autism phenotype” (BAP), which is supported by cotwin studies and family studies of siblings and parents. Social difficulties include lack of interest in social interaction and lack of close, confiding relationships. In one study, 37% of parents of children with ASD reported no friendships, compared with only 4% of parents with Down syndrome. Communication problems are pragmatic, referring to deficits in the social aspects of language. Clear-cut obsessional and repetitive behaviors occur infrequently; however, rigidity, characterized by resistance to and difficulty in accepting change, is more common. These differences in language, anxiety, rigidity and communication may cause parents to feel inadequate to help in the behavioral treatment of their children who experience similar impairments.

Bolton reported that the BAP was seen in 20% of siblings of those with autism versus 3% of siblings in the control group. Similarly, Sasson et al. (2013) reported that higher rates of all BAP traits were found in parents of children with ASD, with the prevalence of each trait ranging from 14% to 23% versus 5% to 9% for control parents. Additionally, parents of children with ASD were more likely to present with multiple traits of BAP when compared to control parents. Ingersoll et al., 2011 found parental BAP was associated with less social support, increased stress and higher rates of depression. Ingrams et al. found that parents expressing BAP used maladaptive coping strategies.

Studies that focused on the stress of raising a child with autism have found increased rates of several psychiatric disorders in parents of affected children, most robustly major depressive disorder (MDD) and social phobia. Piven and Palmer reported MDD in 33% and social phobia in 14.6% of proband parents in comparison to 11.7% and 3.3%, respectively, in Down syndrome parent controls. Smalley and coworkers found a lifetime prevalence of major mood disorders in 40.3% of parents and 32.4% of siblings of children with autism, compared to 19.2% of parents with children with tuberous sclerosis or seizures.

Results of a large meta-analysis performed by Schnabel et al. (2019) found that 31% of parents of children with ASD had depressive disorders, compared to the World Health Organization’s estimated 4.4% of the general population throughout the world. Rates of anxiety disorders in parents of children with ASD (33%) were also significantly more common than the estimated global prevalence (3.6%). With regard to MDD, the increased risk does not appear to be due to raising a child with ASD, given that the majority of depressed parents had their first episode prior to the birth of their affected children. Moreover, these psychiatric disorders have not been found to be associated with the presence of BAP. Potential explanatory models for these findings have involved overlap of the risk alleles that are common to both autism and recurrent mood or anxiety disorders or a result of assortative mating of parents with BAP with spouses with major depression or social phobia. Other mood disorders and substance use disorders have also been reported to occur more frequently in parents of children with ASD in comparison to general-population parents or parents of children with other developmental disorders.

Summary

Parenting a child with an ASD is a stressful, lifelong process that significantly affects all individuals involved. Psychiatrists providing care to these parents need to be aware of the nature and magnitude of this undertaking. Parents should be encouraged to become active in their child's therapies, join available support services, and use positive coping strategies. Monitoring for the development of depression and anxiety disorders and conducting an evaluation for the broader autism phenotype are also indicated. Although specific data are not available for ASD, cultural factors color parents' beliefs about their children's mental health conditions and need to be considered by mental health practitioners who treat the parents.

IMPACT ON PARENTS OF RAISING A CHILD WITH BIPOLAR DISORDER

There has been increasing awareness of the tremendous burden that is placed on parents and caregivers who are raising a child with BPD. Past theories have implied that the dysfunctional family system is causing the illness in the child, it is the child's illness that negatively impacts the family instead. There is growing evidence to support interventions for family members and caregivers of patients with BPD because decreasing stress may improve long-term outcomes for both the family and the child.

Unique Stressors

Research has shown that children, adolescents, and adults who have BPD have severe global impairments in multiple areas of functioning. They are at higher risk for academic failure, poor interpersonal relationships, substance abuse, psychosocial and occupational difficulties, legal problems, suicide attempts, and neuropsychological deficits. In many cases, the parents of children with BPD may also be struggling with their own mood disorders. The child's unpredictable course may create additional stressors that exacerbate parental illness, leading to impaired parenting. Emotional dysregulation in parent and child may increase conflict and compound family stress. Not only is an episode of illness stressful for each of the family members, but the fear of relapse is also present even when the illness is stabilized.

The diagnosis of pediatric-onset BPD is imprecise. However, there is agreement that the presentation of BPD in children and adolescents differs substantially from the presentation in adults. In contrast to the discrete episodes of manic or depressive symptoms observed in adults, children usually exhibit ultradian cycling, irritability, explosive temper outbursts, and decreased frequency of interepisode recovery. Due to the unpredictable nature of the illness presentation, accompanying affective instability, and often violent or aggressive behaviors, the caregiver role can be demanding, distressing, and harmful to health and QOL. Frequently, the diagnosis of BPD is delayed because the child's behavior is initially labeled as oppositional or defiant; therefore, appropriate pharmacotherapy and psychotherapy interventions are also delayed.

Lifespan Issues

As their child begins exhibiting BPD symptoms, parents are generally bewildered by the child's mood swings, disruptive behaviors, and general inability to conform to their peers. Most often, the children appear developmentally normal and have normal intellectual functioning. Therefore, they do not come to the attention of school staff or teachers until they exhibit disruptive behaviors or receive failing grades. Practitioners are usually reluctant to diagnose a child with BPD. These children have been described as "diagnostically homeless." In other words, they often exhibit symptoms that can fit the criteria for multiple other disorders, such as attention-deficit/hyperactivity disorder, mood disorder not otherwise specified, conduct disorder, and intermittent explosive disorder. When the diagnosis of BPD is finally made, parents have described experiencing an array of emotions ranging from relief to depression.

Adolescents with BPD are even more challenging to parents because their desire for autonomy, coupled with impaired judgment, can lead to harmful consequences. Parents of adolescents with BPD struggle to control their children’s promiscuous behavior, dangerous driving habits, substance abuse, and oppositional behaviors. Frequently, the adolescents are not compliant with pharmacotherapy because the medications have side effects such as weight gain or sedation. Some adolescents perceive that the medications alter their personality and make them more socially inhibited. They frequently take little responsibility for the consequences of their actions and tend to perceive their parents as being overly controlling rather than appropriately protective. Parents struggle with guilt, self-blame, and feelings of helplessness because they can no longer physically restrain the adolescents if they choose to leave the home or refuse to take their medications.

Parents of adults with BPD face the greatest obstacles because they have limited legal recourse to protect or help their children if treatment is refused. Because adults have euthymic episodes with longer periods of interepisode recovery, they may cycle into depression or mania without parents being aware of the polarity switch. Consequently, reinitiating treatment may be delayed. During acute episodes, adults with BPD frequently alienate parents, spouses, or other caregivers with their aggressive, abusive, or impulsive behavior. Parents may be overly taxed by the responsibility of emotionally and often financially supporting their adult child.

Caregiver Burden

The effects on caregivers of living with an ill relative have been defined as family burden. The burden of caring for a psychiatrically ill family member has been divided into two categories:

1. *Objective burden* is any disruption to family life that is verifiable and observable. Examples of objective burden include high rates of separation and divorce in marriages with a psychiatrically ill spouse; restricted social and leisure activities, often related to fears of stigma; and financial difficulties.
2. *Subjective burden* involves the personal feelings of carrying such a burden, such as feeling distressed, unhappy, or upset. The types of subjective burden include guilt if the relative feels

> responsible for contributing to the illness; rejection toward the patient; anger that the illness has spoiled the relative's life; and grief contrasting the present with memories of the person before the illness. Subjective burden experienced by relatives has been linked to the development of affective symptoms.

Review of Research on Caregiver Burden for Bipolar Disorder. There is limited data on the degree of stress experienced by caregivers of children with pediatric BPD. Some information about the consequences of caring for a family member with BPD can be obtained from studies of caregivers of adult bipolar patients. Factors affecting family burden of adult BPD include prolonged illness and high levels of dysfunction among patients, as well as caregivers' illness beliefs and coping style. Presence of suicidal ideation in BPD patients is also a predictor of their caregivers' worsening health and development of depressive symptoms over time.

The Family Experience Study, a longitudinal study of the primary caregivers for patients with BPD, examined the cost and consequence of caring for a relative or friend with BPD. Primary caregivers were evaluated within 1 month after patients entered the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD), using measures of burden, coping, physical health, mental health, resource utilization, and cost. At study conclusion, more than half of the caregivers experienced moderate or high burden levels, which were related to patient problem behaviors, interference with caregivers' ability to fulfill other role obligations, or disruption of household routine. Compared to less burdened caregivers, caregivers experiencing higher burden levels reported more physical health problems, depressive symptoms, smoking and drinking, health service use, and limited social support. They also reported having to provide more financial support to their bipolar relative. Limited data has also shown a differentiation between burdens experienced by female caregivers and those experienced by male caregivers. While female caregivers suffer more from relationship difficulties with the BPD patients, male caregivers suffer more from limited autonomy in their personal lives and also from uncertainty regarding the patient's capacity and fluctuating bipolar symptoms.

Most adult patients with BPD spend more time in periods of remission or with subthreshold symptoms than in an acute episode. However, most studies of caregiver stress are completed when patients are acutely ill, which may overestimate stress. In one study, investigators focused on stress in caregivers of euthymic adult bipolar patients aged 18 to 65 during periods of subthreshold symptomatology. Caregivers showed a moderate level of subjective burden on the Social Behavior Assessment Schedule. The variable most highly related to caregiver distress was found to be the patient's behavior (hyperactivity, irritability, sadness, and withdrawal). Parents were most worried about their children's abilities to perform at work, in school, and in social relationships. It is also important to note whether caregivers are more affected by BPD patients' depressive episodes or by their manic episodes. A study on 500 caregivers reported that only BPD patients' depressive episodes, but not their manic episodes, correlate with more subjective and objective caregiver burden.

Subjective and Objective Burden. Factors that increase caregiver's subjective burden are poorer patient social and occupational functioning, a relapse during the previous 2 years, and a history of rapid cycling. Rapid cycling is defined as having four or more episodes of mania or depression per year and is found in a significant fraction of patients with bipolar disorder. It has been associated with greater disability and worse prognosis, as evidenced by higher use of intensive psychiatric interventions.

Risk factors associated with increased objective burden levels include caregiver responsibility for medication administration, violence from the patient, caregiver social isolation, rejection by the patient during periods of illness, and patient irritability leading to more frequent arguments and conflicts. Some studies have suggested that the effect of clinical variables can be mediated by caregiver coping styles and their level of understanding of the patient's illness and symptoms.

Raising a child with BPD can be complicated if the parent is also suffering from a mood disorder. Research has shown that the offspring of parents with BPD have an increased risk for inheriting mood disorders. A meta-analysis of 17 studies of offspring of parents with BPD found that 52% of offspring had challenges with the interpersonal, cognitive, and behavioral manifestations of BPD. Children of bipolar parents also have an elevated risk of major depression, as well as other psychiatric conditions such as generalized anxiety disorder, oppositional defiant disorder (ODD), and obsessive-compulsive disorder. To further complicate matters, parents with BPD are even less capable of coping with the burden of raising a child with BPD.

Interventions

Family interventions for bipolar disorder are generally psychoeducational, with the goals of teaching family members the signs and symptoms of BPD, developing early intervention strategies for new episodes, and providing consistent medication regimens. The psychoeducational approach recognizes that lack of information about the disorder, compounded by uncertainties about the future, can increase caregivers' stress. Thus, in addition to providing information, clinicians help patients and their families emotionally cope with the illness, its prognosis, and expected treatments.

Family-focused therapy (FFT), which is usually provided in combination with pharmacotherapy, involves both BPD patients and their caregivers. Patients and caregivers are jointly trained in psychoeducation, problem-solving skills, and communication enhancement. Eight randomized trials have supported that FFT decreases mood symptoms, promotes faster recovery from mood episodes, and reduces recurrences.

Although there are some randomized, controlled trials that highlight the positive benefit for the patient of family-focused treatment in addition to pharmacotherapy in BPD, few trials using contemporary methodologies have analyzed the specific effects of caregiver-focused interventions. A randomized, control trial published in 2008 studied the impact that interventions for caregivers had on relapse rates in the euthymic bipolar patients they cared for. Caregivers in the experimental group were provided twelve 90-minute sessions that focused on knowledge of BPD and training in coping skills. The patients did not attend the groups, but were assessed monthly during both the intervention and at the 12-month follow-up. The patients whose caregivers attended the intervention group experienced fewer recurrences than those in the control group. In addition, the intervention had different effects depending on the type of episode analyzed: manic, hypomanic, depressive, or mixed. Significant differences were found in the number of subjects who experienced a hypomanic/manic relapse and the time to such an episode. The differences in depressive and mixed episodes did not reach statistical significance.

Relapse prevention may improve by facilitating early identification of the first signs of an episode, leading to earlier intervention. Additional benefits include dispelling caregivers' misconceptions about the illness and issues such as caregiver criticism or overinvolvement. Some studies have shown that family psychoeducation also develops a mutual understanding with the patient. It also increases positive interactional behaviors and reduces expressed emotion. High expressed emotion with negative affective style and family burden has been related to poor illness outcomes.

A 2002 pilot study of manual-driven, multifamily psychoeducation group (MFPG) treatment for families of children with mood disorders represents a rare study focused on pediatric subjects. It addressed two questions: Do families of children with BPD and families of children with MDD or dysthymic disorder (DD) differ at treatment entry, and do they benefit equally from intervention? Conclusions were that BPD families entered treatment with more-impaired children and more extensive treatment histories than families of children with MDD or DD; however, all families benefited from intervention. A qualitative analysis revealed that families in the study gained emotional support, improved coping skills and decision making as it related to their child's illness, shifted their attitudes with regard to the illness, and experienced improved interaction with their affected children. Clinical impressions suggested that interventions combining families of children with BPD and depressive spectrum illnesses were clinically feasible and potentially beneficial. The authors suggested that intervention modalities geared toward families of children with BPD may also include families of children with other mood disorders, given that both groups may benefit from these types of interventions.

A 2010 randomized controlled trial studied the effectiveness of Family-Focused Treatment-Health Promoting Intervention (FFT-HPI). FFT-HPI consisted of 12 to 15 cognitive-behavioral sessions that taught the caregiver how to manage the patient's illness, set goals for self-care, and reduce depression, strain, and health risk behaviors while caring for the patient. Compared to the control, a series of educational videotapes, FFT-HPI reduced depressive symptoms and health risk behaviors in the caregiver. When caregivers were better adjusted, patients also displayed fewer depressive symptoms. This finding suggested that caregiver-oriented therapy can improve patient outcomes, even when the patient is not part of the therapy.

Conclusions

Although most treatment research has focused on pharmacotherapy for pediatric BPDs, a growing literature suggests that psychosocial interventions are also important to provide families with an understanding of symptoms, course, and treatment of BPDs. Interventions to teach youth and parents methods for coping with symptoms (e.g., problem-solving, communication, emotion regulation, cognitive-behavioral skills) and prevent relapse have been shown to be helpful to parents. Identifying and addressing stressful aspects of raising an individual with BPD could reduce the emotional burden and potential negative impacts on the caregiver's mental health and QOL.

Gaining support from other parents who are raising children with BPD may reduce guilt, self-blame, and inappropriate feelings of responsibility for their child's condition. Moreover, providing support for the parents may ultimately improve the child's outcome. With support and education, parents are better able to implement multimodal treatment interventions to improve adaptive and functional outcomes for their child. In cases of multigenerational BPD, treatment of the parent's mood disorder can also have a positive effect on the child's symptoms, which can bolster parental perceptions of self-competence. In conclusion, psychosocial interventions that involve families, psychoeducation, and skill building may offer added benefits to pharmacotherapy interventions.

IMPACT ON PARENTS OF RAISING A CHILD WITH ATTENTION-DEFICIT/HYPERACTIVITY DISORDER

Parenting a child with ADHD has challenges similar to those faced by parents of children with other developmental and/or behavioral disorders. These challenges include increased parental stress, marital discord, and maternal depression. Advocacy issues and long-term care concerns, although present, are generally less urgent. Nonetheless, rearing children with ADHD involves separate and unique aspects of care that are influenced by both child-derived and parent-derived characteristics. Complicating this scenario is the likely interaction among these characteristics.

Effect of Child Characteristics on Parenting Stress

Several authors report increased parenting stress in mothers of children with ADHD. Child characteristics of ADHD that specifically correlate with parental stress are the severity of ADHD symptoms, aggression, oppositional behavior, and, to a lesser degree, the presence or absence of medical illness in the child. This highlights the effect of comorbid and associated symptoms on the level of stress within the family of the ADHD patient.

Unlike most children with more severe disabilities, children with ADHD, especially of a more moderate level of severity, are expected to master the developmental challenges of typical childhood, adolescence, and adulthood. From kindergarten through middle school, they are expected to relate to their peers, join extracurricular peer group activities, participate in sports, and succeed in mainstream education. In adolescence, they drive cars, date, work, and attend school. They are expected to enter early adulthood with the typical goals of learning a trade, going to college, obtaining employment, and establishing long-term relationships and friendships. However, due to inattention, impulsivity, and/or hyperactivity, ADHD patients do not always meet these goals. Studies show that children with ADHD, especially those with a hyperactive/impulsive component, are at much higher risk for accidental injuries and trauma, emergency room visits, motor vehicle accidents, grade retention, high school dropout, drug use, unplanned pregnancies, and sexually transmitted diseases.

Furthermore, data has also shown that individuals diagnosed with ADHD between 5 and 12 years of age are more likely to engage in criminal activities as young adults compared to non-ADHD individuals. While individuals with the inattentive subtype are more likely to be involved in activities that require planning (e.g., selling drugs and burglary), individuals with impulsive symptoms have a higher rate of crime involvement and are more likely to be arrested and convicted. In addition, more than half of these children were fired from one or more jobs by the age of 21 years. These consequences of ADHD create stress within the family and can lead to a higher incidence of parental divorce or marital discord and child abuse.

Effect of Parent Characteristics on Parenting Stress

Parenting stress is also correlated with specific attributes of the parents who are rearing children with developmental disabilities. For example, in parents raising children with autism, maternal depression and the perception of one’s share of the caretaking burden have been linked to increased stress. Similarly, in Anastopoulos’ study of mothers of children with ADHD, parenting stress, as measured by the Parenting Stress Index (PSI), correlated highly with maternal psychopathology (e.g., depression, bipolar disorder) and lack of employment outside of the home. Seipp and coworkers reported that during observed mother–son interactions with boys with ADHD, mothers with greater overt hostility had less positive parenting behavior, which was associated with greater oppositionality in the son. Moreover, mothers with sons with ADHD and comorbid ODD were less responsive to their child's needs and more overreactive than control mothers. This finding may be more directly related to specific maternal or child characteristics or the interplay between the two.

Rearing a child with ADHD is greatly complicated if parents also have ADHD and struggle with similar cognitive and behavioral manifestations. An estimated 40% to 60% of adults with ADHD are rearing an affected child. Because the prevalence of ADHD in adults is estimated at around 4.5%, there are millions of parents with ADHD rearing children with ADHD. This situation creates a multitude of additional management issues, increasing the family's stress level. Parental ADHD, especially if untreated or poorly managed, may contribute to parental neglect, inadequate supervision, inconsistent and inappropriate discipline, and increased parent–child conflict. Therefore, the parent without ADHD may be unfairly burdened with the responsibility of managing the child with ADHD. Furthermore, maternal depression and early maternal negative parenting techniques are predictors of future conduct problems in children with ADHD. Conversely, early positive parenting techniques, including positive feedback, praise, and optimistic affect, are associated with fewer conduct issues.

Mrs. M, a 34-year-old homemaker, had been married for 10 years and had two children. Both her 8-year-old son and her husband were diagnosed with ADHD, although only her son was medicated. Mrs. M's son continued to have significant impulsive aggression directed toward his younger sibling, serious academic struggles, and poor peer relationships secondary to "immature behavior." Her husband had poor job performance, with deficient organizational skills and failure to attend important meetings due to "forgetfulness." He was reluctant to take medications because of concerns about side effects and "drug dependence." Mrs. M's family of high achievers compared her unfavorably to her more accomplished sister and her family. Moreover, her family had a history of depression but worries about the stigma of psychiatric illness preventing her from seeking treatment. She has been stressed because she and her husband have discussed divorce. She had no sources of emotional support and felt overwhelmed.

Clinical Implications

The assessment of ADHD patients who are parents needs to include consideration of the patient's ability to meet the demands of parenthood, especially in the presence of a childhood psychiatric disorder. The consequences of rearing a child with ADHD should not be underestimated, especially given the frequency with which the disorder also affects one of the child's parents. Studies have shown that the degree of parental dysfunction correlates with the development and severity of the child's ADHD-derived disorders, such as conduct disorder and ODD. Furthermore, poor parental skills can also lead to poorer self-control in ADHD children, which is a catalyst for more disruptive behaviors.

IMPACT ON FAMILIES OF RAISING A CHILD DIAGNOSED WITH SCHIZOPHRENIA

The National Institute of Mental Health estimated that in 2002, 1.1% of the population suffered from schizophrenia. The estimated yearly cost of schizophrenia in the United States was 62.7 billion dollars in 2002. Globally, the cost is estimated to be between 1.5% and 3% of national health care budgets. This subsection discusses the burden borne by parents and siblings of a child with schizophrenia. Whereas the other disorders discussed in this section are usually observed earlier in the child's life, schizophrenia typically develops between the mid-teens to early 30s, with rare cases of earlier onset after a significant period during which the child appeared to be on track to become a healthy, independent adult.

Early psychodynamic theories of schizophrenia focused on the role of mothers of patients with schizophrenia. They were often called "schizogonic" or "schizophrenogenic," and their lack of maternal warmth was, without any sound scientific evidence, blamed for their child's development of schizophrenia. Only in the mid-1960s was this concept countered, and since then, schizophrenia research has been approached from a biopsychosocial perspective based on the notion that the expression of the disorder is due to a genetic predisposition combined with exposure to environmental and individual stressors. This puts the role of the parent and family in a different light.

Caregiver Burden

To better assess the issues with which families of schizophrenic patients grapple, Hoenig and Hamilton introduced the concept of "family burden," which was defined earlier in this section. The Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study found four factors that caregivers identified as most burdensome in their experience with their ill family member: (1) patient problem behavior, (2) the patient's impairment in activities of daily living, (3) lack of patient helpfulness, and (4) resource demands and disruptions in the caregivers' personal routine. Other factors related to the level of burden were social and professional support, age, and education level. Family members who were able to share their experience with others and who received financial support experienced a lower level of burden; younger and less-educated caregivers experienced more burden than older and more-educated caregivers.

Parents and siblings of patients with schizophrenia often grieve the loss of the child or sibling they knew before the onset of the illness. They may also experience feelings of embarrassment and stigma related to their family member’s illness, which, in turn, may lead to isolation. Aging parents frequently worry about what will happen to their children when they can no longer care for them. In a study of 157 mothers of children with serious mental illness, only 11% reported definite plans for their offspring’s future residence; 66% had completed financial plans; and 75% hoped a family member would assume care, but only 25% thought that this would actually occur. Older parents need professional help with planning for the future, in particular estate planning, to enable case management, and ongoing services after the parent’s death.

Siblings play a unique role in the life of the patient with schizophrenia and may have specific needs. They have additional levels of burden, such as guilt about being healthy, fear of developing schizophrenia, and fear of passing on “bad genes” to their offspring. Several coping mechanisms have been identified in siblings, including normalization, caregiving, and grief. Smith and coworkers identified several factors that predict whether siblings will be involved in the future care of the ill sibling. Pull factors, which predict involvement, are early experiences in socialization, strong sibling relationships, and personal caregiver gains, such as developing personal strengths while caring for the sibling. Push factors, which predict disengagement, are geographic distance and a belief that the ill sibling has control of their symptoms.

Cultural Aspects of Burden

Across cultures, families of patients with schizophrenia often suffer from the stigma that accompanies the diagnosis of a psychotic disorder. Several studies have indicated that black caregivers, on average, experience less distress than Hispanic and white caregivers: A seemingly surprising finding in light of the greater likelihood of additional stressful conditions such as prejudice, discrimination, and lower income that African-Americans face. In addition, African-American family members tend to be more involved in caregiving activities, which are also associated with higher levels of distress. Potential protective factors, such as religious involvement and social support, do not appear to differ according to race and ethnicity. Studies indicate that the QOL for caregivers who live with a chronically ill family member is lower than those living alone or with healthy family members, and the financial burden is significant. QOL for siblings is better than for parents and children of schizophrenic patients, but across cultures, women tend to experience more distress than men because they tend to provide more care for ill children or siblings.

Treatment and Family Burden

A multidimensional approach is required to assess and understand the burden experienced by families of schizophrenic patients. The Schizophrenia Patient Outcomes Research Team (PORT) recommended that, in addition to pharmacotherapy, patients with ongoing family contact should be offered family psychosocial intervention lasting at least 9 months, emphasizing education about the illness, family support, crisis intervention, and problem-solving skills training. This may relieve the burden experienced by the family, who in turn is better able to take care of the ill relative. Several support models exist, including single- and multiple-family group models. Although models vary, effective programs share the following characteristics and assumptions:

1. Schizophrenia is an illness.
2. Program leadership is provided by mental health professionals.
3. Participating family members are not seen as patients.
4. Family therapy is not included to avoid the suggestion that the family is to blame for the relative’s illness.

5. The focus remains on patient outcome, including prevention of relapse and improvement of functioning and well-being of the patient.

Families participating in such programs can reduce relapse rates by up to 50%, which is comparable to pharmacotherapeutic effect sizes. However, according to PORT, participation in such programs is limited. Although 75% of patients had contact with family members, only 31% of families had received information, advice, or treatment. Barriers to access included patient factors (concerns about family burden and confidentiality), family factors (lack of awareness regarding benefits, time constraints, reluctance to become more involved), and provider factors (staff resources, time and training, reimbursement issues, lack of awareness of benefits). Community-based programs, such as the National Alliance for the Mentally Ill (NAMI), Family-to-Family, and Journey of Hope programs, are run by volunteers from patients' families and focus on family needs, advocacy, and system reform. Participation in these programs leads to improved overall well-being for family members, greater facility in navigating the mental health care system, and increased understanding of schizophrenia. Shorter duration and lower costs may make them more accessible to family members.

Conclusions

Schizophrenia is a chronic illness that often causes an enormous burden for the patient's family. However, during the last decade, research has increasingly shown that support for patients and their families through psychoeducation and support programs can have a significant effect on the perceived burden by both families and the patient. It can lead to improved outcomes for the patient, with fewer episodes of relapse and better QOL, whereas families are more able to understand the diagnosis and its consequences for their child or sibling and to take on the role of caregiver. In addition, these programs reduce the financial burden of schizophrenia on the health care system. However, many factors hamper the implementation of these programs; it is the role of the treatment provider to offer the best possible and evidence-based care to patients and their families.

SUMMARY

This section has described the consequences of having a psychiatrically ill child on their parents by considering four diverse disorders that usually first appear in childhood or young adulthood. Although not all-inclusive, these disorders were chosen to illustrate themes generally experienced by any family so affected. Common to all of these disorders are the grief and disappointment that parents experience when they realize that their child is not typically developing. Self-blame is frequent, particularly for inheritable disorders. Factors that shape parental response include the age of onset of the disorder, chronicity of the disorder, the extent to which the disorder affects the acquisition of basic developmental milestones and functioning, the presence of aggression, and the stigma associated with the specific disorder. The presence of the disorder or an associated condition in a parent, as well as in the child, greatly affects the outcome for all members of the family. The lack of appropriate services, especially when the youngster reaches adulthood, is a prevalent dilemma for parents. Psychiatric practitioners need to consider the profound effect that raising a child with a serious psychiatric disorder has on a parent who is their patient.

FURTHER READINGS

Aghebati A, Gharraee B, Shoshtari MH, Gohari MR. Triple p-positive parenting program for mothers of ADHD children. *Iran J Psychiatry Behav Sci*. 2014;8(1):59–65.

Allik H, Larsson JO, Smedje H. Health-related quality of life in parents of school-age children with Asperger syndrome or high-functioning autism. *Health Qual Life Outcomes*. 2006;4:1.

Bauer R, Gottfriedsen GU, Binder H, et al. Burden of caregivers of patients with bipolar affective disorders. *Am J Orthopsychiatry*. 2011;81(1):139–148.

Chessick CA, Perlick DA, Miklowitz DJ, et al. STED-BD Family Experience Collaborative Study Group. Suicidal ideation and depressive symptoms among bipolar patients as predictors of the health and well-being of caregivers. *Bipolar Disord*. 2009;11(8):876–84.

Cohen IL, Tsiouris JA. Maternal recurrent mood disorders and high-functioning autism. *J Autism Dev Disord*. 2006;36:1077–1088.

Giallo R, Gavidia-Payne S. Child, parent and family factors as predictors of adjustment for siblings of children with a disability. *J Intellect Disabil Res*. 2006;50:937–948.

*Gray DE. Coping over time: the parents of children with autism. *J Intellect Disabil Res*. 2006;50:970–976.

Hastings RP, Kovshoff H, Ward NJ, degli Espinosa F, Brown T, Remington B. Systems analysis of stress and positive perceptions in mothers and fathers of pre-school children with autism. *J Autism Dev Disord*. 2005;35:635–644.

Herring S, Gray K, Taffe J, Tonge B, Sweeney D, Einfeld S. Behaviour and emotional problems in toddlers with pervasive developmental disorders and developmental delay: associations with parental mental health and family functioning. *J Intellect Disabil Res*. 2006;50:874–882.

Kersh J, Hedvat TT, Hauser-Cram P, Warfield ME. The contribution of marital quality to the well-being of parents of children with developmental disabilities. *J Intellect Disabil Res*. 2006;50:883–893.

*Lecavalier L, Leone S, Wiltz J. The impact of behaviour problems on caregiver stress in young people with autism spectrum disorders. *J Intellect Disabil Res*. 2006;50:172–183.

Mandell DS, Novak M. The role of culture in families' treatment decisions for children with autism spectrum disorders. *Ment Retard Dev Disabil Res Rev*. 2005;11:110–115.

*Marcus L, Kunce LJ, Schopler E. Working with families. In: Volkmar F, Paul R, Klin A, Cohen D, eds. *Handbook of Autism and Pervasive Developmental Disorders*. 3rd ed. Vol. 2. Wiley; 2005:1055.

Miklowitz DJ, Chung B. Family-focused therapy for bipolar disorder: reflecting on 30 years of research. *Fam Process*. 2016;55(3):483–499.

Modesto-Lowe V, Danforth JS, Brooks D. ADHD: does parenting style matter? *Clin Pediatr (Phila)*. 2008;47(9):865–872.

Montes G, Halterman JS. Psychological functioning and coping among mothers of children with autism: a population-based study. *Pediatrics*. 2007;119:e1040–e1046.

Mugno D, Ruta L, D'Arrigo VG, Mazzone L. Impairment of quality of life in parents of children and adolescents with pervasive developmental disorder. *Health Qual Life Outcomes*. 2007;5:22.

*Murray C, Johnston C. Parenting in mothers with and without attention-deficit/hyperactivity disorder. *J Abnorm Psychol*. 2006;115:52–61.

Ostacher MJ, Nierenberg AA, Iosifescu DV, et al.; STEP-BD Family Experience Collaborative Study Group. Correlates of subjective and objective burden among caregivers of patients with bipolar disorder. *Acta Psychiatr Scand*. 2008;118:49–56.

*Perlick DA, Rosenheck RA, Kaczynski R, Swartz MS, Canive JM, Lieberman JA. Components and correlates of family burden in schizophrenia. *Psychiatr Serv*. 2006;57:1117–1125.

*Perlick DA, Rosenheck RA, Miklowitz DJ, et al.; STEP-BD Family Experience Collaborative Study Group. Prevalence and correlates of burden among caregivers of patients with bipolar disorder enrolled in the Systematic Treatment Enhancement Program for Bipolar Disorder. *Bipolar Disord*. 2007;9:262–273.

Reinares M, Vieta E, Colom F, et al. What really matters to bipolar patients' caregivers: sources of family burden. *J Affect Disord*. 2006;94:157–163.

Schieve LA, Blumberg SJ, Rice C, Visser SN, Boyle C. The relationship between autism and parenting stress. *Pediatrics*. 2007;119(Suppl 1):S114–S121.

Schwichtenberg A, Poehlmann J. Applied behaviour analysis: does intervention intensity relate to family stressors and maternal well-being? *J Intellect Disabil Res*. 2007;51:598–605.

Seipp CM, Johnston C. Mother-son interactions in families of boys with attention-deficit/hyperactivity disorder with and without oppositional behavior. *J Abnorm Child Psychol*. 2005;33:87–98.

Shu BC, Lung FW. The effect of support group on the mental health and quality of life for mothers with autistic children. *J Intellect Disabil Res*. 2005;49:47–53.

Siklos S, Kerns KA. Assessing the diagnostic experiences of a small sample of parents of children with autism spectrum disorders. *Res Dev Disabil*. 2007;28:9–22.

Singh MK, DelBello MP, Stanford KE, et al. Psychopathology in children of bipolar parents. *J Affect Disord*. 2007;102:131–136.

Smith MJ, Greenberg JS, Mailick Seltzer M. Siblings of adults with schizophrenia: Expectations about future caregiving roles. *Am J Orthopsychiatry*. 2007;77:29–37.

Smith LE, Seltzer MM, Tager-Flusberg H, Greenberg JS, Carter AS. A comparative analysis of well-being and coping among mothers of toddlers and mothers of adolescents with ASD. *J Autism Dev Disord*. 2008;38(5):876–889.

Tonge B, Brereton A, Kiomall M, Mackinnon A, King N, Rinehart N. Effects on parental mental health of an education and skills training program for parents of young children with autism: A randomized controlled trial. *J Am Acad Child Adolesc Psychiatry*. 2006;45:561–569.

54.13 Pediatric Sleep Disorders

Although precise estimates of the frequency of sleep disorders in children and adolescents are somewhat variable among studies, the understanding of the epidemiology of sleep pathology and its relationship to comorbid psychiatric diagnoses has grown substantially in recent years. Research in cognition and brain development over the last 20 years also has significantly advanced the understanding of the genetic and neurobiologic bases of sleep disorders in children and adolescents. With an increasing understanding of sleep disorders, treatment interventions have also become more refined and useful.

This section reviews the research and clinical methods used to study sleep, new findings on the neurodevelopmental and cognitive effects of sleep loss and fragmentation, and the clinical description and treatment of sleep disorders in children and adolescents and their interface with psychiatric disorders.

SLEEP EVALUATION METHODS

One of the primary tools for studying sleep is *polysomnography* (PSG), which incorporates a variety of measures of physiologic changes experienced during sleep. The American Thoracic Society developed a set of recommendations for the interpretation of pediatric PSG that has become the standard for sleep labs nationally. With the aid of portable technology, PSG is increasingly being performed at home, although the methods and results are not quite as sophisticated as when performed in the lab. While PSG remains the gold standard for sleep studies, other less complicated and less expensive techniques are available for studying a variety of sleep disorders in children and adolescents.

Actigraphy was developed in the early 1970s and has come into increasing use in both research studies and clinical practice. Actigraphy allows for the study of sleep–wake patterns and circadian rhythms via the assessment of body movements. The device is typically worn on the wrist and can easily be adapted for home use. With the advent of personal fitness trackers, manufacturers have increasingly been designing better algorithms for gathering data on sleep; to date, however, these devices have shown high variability and many inaccuracies.

A number of survey instruments exist for detecting problematic sleep in children and adolescents, including self-report questionnaires (such as the Sleep Disturbance Scale for Children, the Child Sleep Questionnaire, the Child and Family Sleep History Questionnaire, and the Sleep Disorders Inventory for Children), sleep diaries, and parent report forms. Sleep logs or diaries provide information on bedtime, sleep latency, wake time, and the number of nocturnal awakenings, in addition to exercise and the use and timing of caffeine, alcohol, and drug use. Sleep logs are typically completed for a 2-week period by parents and caregivers or adolescents themselves. Although subjective, sleep logs provide a wealth of useful information for treating insomnia and sleep phase disorders and are available readily in many forms on the internet. Finally, the Epworth Sleepiness Scale, commonly used with adults as a subjective measure of the propensity to fall asleep in certain situations and to monitor treatment efforts, has recently been modified for use in children and adolescents.

NEUROCOGNITIVE EFFECTS OF SLEEP DISRUPTION

Most of the research examining neurocognitive sequelae of sleep loss and fragmentation in children has focused on the effects of sleep restriction and sleep-disordered breathing (SDB) on daytime performance in children and adolescents. As little as one night of acute sleep restriction in children can result in significant attention deficit. A study conducted in a sample of 7- and 8-year-old healthy children subjected to 1 hour of sleep restriction for 1 week demonstrated impaired attention and concentration, along with significant reorganization in brain cortical electroencephalographic

(EEG) response as measured by multiarray P300 (cognitive) evoked potentials. By contrast, one additional hour of sleep in children aged 8 to 12 years over 1 week resulted in parent and child reports of improved emotional functioning and objective improvements in memory, attention, and math fluency. Similarly, in children 7 to 11 years old, one additional hour in bed, averaging 27 minutes of additional sleep, demonstrated improvements in teachers' blinded ratings of emotional liability and impulsivity.

Obstructive sleep apnea syndrome (OSAS), which is characterized by sleep fragmentation due to multiple arousals and disruption in ventilation following episodes of complete or partial airway obstruction during sleep, has been studied with respect to its effect on neurocognitive functions in children. Numerous studies report academic deficits and learning problems in children with OSAS that appear to be reversible after treatment of the condition. Hyperactive, inattentive, and disruptive behaviors resembling those found in attention-deficit hyperactivity disorder (ADHD) have been increasingly reported in children with snoring and OSAS. Similar behavioral changes are seen in some children with periodic limb movement disorder (PLMD) and restless legs syndrome (RLS).

In adults, sleep deprivation adversely affects functions mediated by the prefrontal cortex, particularly attention and working memory, upon which most cognition depends. Deduction and critical thinking (logic-based tasks) show little change in accuracy with sleep deprivation but do reflect cognitive slowing. Divergent skills involved in decision making, however, are greatly affected by sleep loss and involve trouble assimilating changing information, updating strategies based upon new information, innovation, risk assessment, and maintaining interest, insight, communication, and memory. Unfortunately, a lack of sleep, in addition to increasing levels of responsibility, lead to somewhat of a perfect storm, contributing to the high-risk behavior so commonly observed among adolescents.

A number of studies have indicated reduced intelligence scores among children with SDB compared to normal controls, including significantly impaired general conceptual ability. Lower academic achievement has also been observed in children with OSAS compared to children without reported histories of SDB when matched for age, race, gender, socioeconomic status, and school attended.

The increasing reliance upon technology, such as computers and smartphones, represents perhaps the greatest novel threat to our children's sleep. The United Nations currently estimates that more of the world's population has access to a smartphone than a working toilet. New research has found that excessive screen time contributes to childhood obesity and is associated with aggressive thoughts and behavior, cognitive impairment, and delayed literacy. In addition to the physiologic and psychological arousal from screens that contributes to impaired sleep, the bright lights from screens can block the natural evening flow of melatonin, further contributing to insomnia.

SLEEP DISORDERS IN CHILDREN

It is estimated that approximately 25% of children will suffer from a sleep problem at some point during childhood. The types of problems affecting children vary greatly in their frequency and severity, ranging from the more common bedtime resistance and anxiety to the occasional primary sleep disorders, such as narcolepsy and OSAS. Regardless of the primary complaint, parents frequently report sleep difficulties among their children, and these reports are remarkably consistent with most researchers identifying up to 50% of preschool children, roughly 30% of school-aged children, and approximately 40% of adolescents as having sleep difficulties. Among a representative U.S. sample of 13- to 16-year-olds, the 30-day prevalence of insomnia was 9.4%, and the lifetime prevalence was 10.7%; further, 88% of the adolescents with a lifetime history of insomnia reported current sleep difficulties, further indicating the chronic nature of insomnia among youth. Much higher rates of sleep disorders are generally reported among those with neurodevelopmental and psychiatric disorders. Among children with intellectual disability, for example, 30% to 80% have sleep problems, while 50% to 90% of children with autism spectrum disorders (ASDs) are affected. Despite a high overall prevalence, however, sleep problems are not often discussed during visits with primary care practitioners and are therefore left undiagnosed and untreated.

Classification of Sleep Disorders

The *International Classification of Sleep Disorders*, Third Edition (ICSD-3), and the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (DSM-5), and its text revision (DSM-5-TR) are the primary classification systems for sleep disorders. While the ICSD-3 and the DSM-5-TR cover much of the same ground, the ICSD-3 classification is more comprehensive and precise. The DSM-5-TR specifies 12 main categories of sleep disorders (see Table 54.13–1).

Sleep-Disordered Breathing. The term SDB encapsulates a variety of difficulties, including primary snoring, upper airway resistance syndrome (UARS), and OSAS, and it occurs relatively frequently in otherwise healthy children. Although accurate figures are difficult to obtain, it is generally estimated that 2% to 5% of all children may be affected by OSAS, and 7% to 12% of children may suffer from primary snoring.

Whereas adults with OSAS most often present with *excessive daytime somnolence* (EDS), children rarely exhibit this symptom and instead tend to present with externalizing behaviors, such as hyperactivity, aggression, social withdrawal, and learning difficulties. In one study, fewer than 15% of children with OSAS exhibited EDS at all. Diagnosis of OSAS further requires a consistent clinical history and physical examination. To augment these two diagnostic measures and reveal the extent of the effect on sleep, children should also undergo overnight PSG.

Children with OSAS present with loud snoring, sweating, agitated sleep, and a variety of related sleep problems (e.g., persistent enuresis, night terrors, somnambulism, bruxism, etc.). EDS, somnambulism, bruxism, and anxiety are also frequently noted in adolescents with OSAS. Physical examination often demonstrates craniofacial abnormalities and adenotonsillar hypertrophy, and PSG indicates frequent apneas/hypopneas and oxygen desaturations. Outcomes of OSAS can include significant health problems, such as failure to thrive, hypertension, pulmonary hypertension with or without cor pulmonale, and mental health problems.

Table 54.13–1.
DSM-5-TR Classification of Sleep Disorders

- Insomnia Disorder
  - Hypersomnolence Disorder
  - Narcolepsy
  - Breathing-Related Sleep Disorders
    - Obstructive Sleep Apnea
    - Central Sleep Apnea
    - Sleep-Related Hypoventilation
  - Circadian Rhythm sleep-Wake Disorders
    - Delayed Sleep Phase Type
    - Advanced Sleep Phase Type
    - Irregular Sleep-Wake Type
    - Non-24-Hour Sleep-Wake Type
    - Shift Work Type
  - Parasomnias
    - Non-Rapid Eye Movement Sleep Arousal Disorder
    - Nightmare Disorder
    - Rapid Eye Movement Sleep Behavior Disorder
    - Restless Legs Syndrome
- Substance/Medication-Induced Sleep Disorder
- Other Specified Insomnia Disorder
- Unspecified Insomnia Disorder
- Other Specified Hypersomnolence Disorder
- Unspecified Hypersomnolence Disorder
- Other Specified Sleep-Wake Disorder
- Unspecified Sleep-Wake Disorder

Data from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. text rev. American Psychiatric Association; 2022.

The most common treatment for pediatric SDB is adenotonsillectomy, which cures the problem in approximately 85% of affected children. A continuous positive airway pressure (CPAP) mask is often used for children who fail surgical treatment or for whom surgery is not indicated.

Insomnia. *Pediatric insomnia* has been much less studied and is more poorly understood than adult insomnia. A number of efforts have been made to define pediatric insomnia, estimate its prevalence among different pediatric populations, and develop clinical guidelines for behavioral and pharmacologic treatment.

The prevalence of pediatric insomnia is estimated at about 1% to 6% among general pediatric populations, with a much higher prevalence in children with neurodevelopmental and chronic medical and psychiatric conditions. When bedtime resistance and disruptive nighttime awakenings are included, the prevalence of sleep-disrupted behavior approaches 25% to 50% in preschool-aged children. Psychiatric symptoms have been reported in almost 50% of children with persistent insomnia. It has been proposed that persistent insomnia may also represent an early sign of emotional distress in susceptible children with poorly preserved sleep homeostasis.

The ICSD-3 describes two forms of behavioral insomnia of childhood as unique diagnostic entities to emphasize sleep difficulties that result from inappropriate sleep associations or inadequate parental limit setting. *Limit-setting type* is due to parents having difficulties establishing and enforcing bedtimes, while *sleep-onset association type* is characterized by reliance on maladaptive and inappropriate sleep associations such as rocking, watching television, feeding, and falling asleep in the parents' bed. The child is usually unable to fall asleep in the absence of these conditions at both bedtime and following nocturnal arousals. Behavioral insomnia generally leads to sleep fragmentation and increased nighttime awakenings, insufficient sleep duration, later bedtimes, and daytime sleepiness.

Good sleep hygiene and behavioral interventions, such as extinction or bedtime fading, are the first recommended treatments for pediatric insomnia. Children who do not respond to behavioral interventions may be considered for the pharmacologic management of their insomnia. There are no U.S. Food and Drug Administration (FDA)-approved medications for the treatment of insomnia in children. Nonetheless, a large study of over 1,000 child and adolescent psychiatrists found that 96% recommended at least one prescription medication, and 88% advised an over-the-counter medication for sleep in a typical month. Despite the lack of FDA approval, however, limited empirical evidence supports the use of melatonin and clonidine for childhood insomnia. Meanwhile, benzodiazepines, guanfacine, pyrimidine derivatives (e.g., zaleplon and zolpidem), sedating antidepressants (e.g., trazodone and mirtazapine), sedating antipsychotics (e.g., quetiapine), and sedating antihistamines (e.g., diphenhydramine, doxylamine, and hydroxyzine) are often employed to improve symptoms of insomnia in children and adolescents but have virtually no solid evidence to support their use.

Coincident with the increasing use of herbal medications and dietary supplements in the United States, a number of studies have been conducted over the last decade that demonstrates that melatonin is an effective, safe, and well-tolerated agent, especially in cases of sleep-initiation insomnia caused by circadian factors. Several placebo-controlled studies of melatonin in normal, healthy children with idiopathic sleep-initiation insomnia have shown that melatonin administered at bedtime reduces sleep-onset latency time and increases total sleep time (TST).

Little information is available on the use of the nonbenzodiazepine hypnotic agents zaleplon and ramelteon, but studies have found both zolpidem and eszopiclone ineffective for ADHD-related insomnia. In addition, hypnogogic hallucinations and paradoxical agitation have been demonstrated when adult doses of these agents have been prescribed to children and adolescents. Clearly, more safety trials and pharmacokinetic studies are needed prior to recommending these drugs for use in pediatric insomnia.

Cognitive behavior therapy for insomnia (CBT-I) is the most effective treatment in adults for decreasing sleep latency and the frequency and duration of nighttime awakenings, improving sleep quality, and increasing TST. CBT-I incorporates a number of evidence-based psychological and behavioral strategies, including stimulus control (e.g., removing clocks from the bedroom, maintaining a quiet sleep environment, etc.), sleep restriction (e.g., limiting the amount of time in bed until the patient is sleeping 90% of the time they are in bed), arousal reduction (e.g., relaxation techniques such as belly breathing and biofeedback), addressing cognitive distortions associated with insomnia (e.g., catastrophizing such as “if I don’t fall asleep soon, I’ll be a complete wreck tomorrow”), circadian rhythm maintenance (e.g., going to bed and awakening at the same time each day), and sleep education and hygiene. CBT-I is well accepted by patients, and 70% to 80% achieve a therapeutic response, while 40% achieve clinical remission. Numerous randomized studies of adults demonstrate that CBT-I is even more effective for insomnia than sleep medications and that it is equally effective whether delivered one on one, in a group setting, or via the internet.

Although most studies of CBT-I have focused on adults, recent research demonstrates its efficacy among pediatric populations as well. Reductions in sleep latency and wake after sleep onset, along with increased sleep efficiency among both children and adolescents, have been demonstrated to last for at least 6 months posttreatment. In addition, the first randomized controlled trial of CBT-I in adolescents found a medium to large effect size.

While sleep hygiene is typically recommended by practitioners, sleep hygiene alone is not sufficiently effective for the treatment of pediatric sleep difficulties. Discontinuing exposure to blue light via screens for an hour or two before sleep, limiting caffeine, alcohol, and drugs (such as marijuana and nicotine), in addition to creating a comfortable sleep environment with limited light and noise while being judicious about the timing of exercise, are, however, valid sleep hygiene strategies that can augment the utility of CBT-I.

Narcolepsy. Major strides in the understanding of the etiology of narcolepsy have been taken in recent years as remarkable discoveries have identified a new neuropeptide, hypocretin or orexin, which is linked to animal models of narcolepsy. The exact prevalence of childhood narcolepsy is unknown. The youngest patient on record was diagnosed at age 12 months. In retrospective studies, about one-third of adults with narcolepsy indicate onset prior to 15 years of age and roughly 15% prior to 10 years of age.

Children with narcolepsy are commonly misdiagnosed with other neurologic and/or psychiatric disorders, such as epilepsy, ADHD, mood disorders, and psychotic disorders. Excessive daytime sleepiness is the most common presentation of narcolepsy in school-aged children, although according to some reports, 50% to 70% of childhood narcolepsy patients present with symptoms of cataplexy. Nighttime sleep is frequently disturbed in patients with narcolepsy, and periodic limb movements during sleep are also common. The diagnosis of narcolepsy requires sleep laboratory assessment, including nocturnal PSG and a multiple sleep latency test.

Effective treatment of narcolepsy in children and adolescents demands establishing a regular sleep schedule, providing extensive education on sleep hygiene, and implementing scheduled daytime naps if possible. Excessive daytime sleepiness in children and adolescents with narcolepsy is commonly treated with psychostimulants such as methylphenidate and dextroamphetamine. Modafinil and armodafinil are alertness-promoting medications that can provide long-lasting benefits and have been shown to be effective in pediatric patients with narcolepsy.

Tricyclic antidepressants such as clomipramine, protriptyline, and imipramine have long been known to be effective for symptoms of cataplexy. Serotonin-specific reuptake inhibitors and mixed-action antidepressants have also been prescribed to patients with cataplexy. Sodium oxybate (Xyrem) and calcium, magnesium, potassium, and sodium oxybate in combination (Xywav) have been recently approved by the FDA for pediatric use and are effective for the treatment of excessive daytime sleepiness and cataplexy associated with narcolepsy.

Due to the complexity of narcolepsy symptoms and their effect on child development, a variety of treatment modalities, including individual and family therapy, school advocacy, and academic interventions, should be used to optimize symptom control.

Parasomnias. *Parasomnias* are disruptive physical acts that occur during specific sleep stages or during the transition between sleep stages and include non-REM types (somnambulism and sleep terrors), REM sleep behavior disorder, somniloquy, bruxism, nightmares, and RLS. Although the incidence of general psychopathology in children with non-REM sleep parasomnias is quite low, the overall prevalence of having at least one parasomnia by age 13 years has been reported to be as high as 78%. The events, although rarely requiring clinical intervention, are often very troubling for parents.

Somnambulism, or sleepwalking, is relatively common in childhood, with annual prevalence rates approaching 17%. Although somnambulism rarely results in harm to the subject, it is possible for an episode of somnambulism to become a confusional arousal, in which the likelihood of accidental violence (to oneself or others) is much greater. The prevalence of confusional arousal is less well established, although a 4% incidence was noted in a longitudinal study of 212 children of ages 6 to 16 years selected at random in Stockholm.

Sleep terrors are somewhat less common than somnambulism and are difficult to quantify because of the self-report nature of the data and the fact that parents often confuse them with nightmares. Annual incidence rates have been reported as high as 6% and prevalence rates as high as 17% among children aged 13 years and younger. Most non-REM parasomnias are first noticed by parents of affected children. Although a large percentage of children have at least one incidence of somnambulism, far fewer present with recurrent and disruptive somnambulism. Sleep terrors follow a similar trend, although they are less common than sleepwalking. Numerous studies have demonstrated a connection between anxiety levels and parasomnias in children, showing that increased anxiety correlates with an increased prevalence of sleepwalking and sleep terrors.

Although the treatment of non-REM parasomnias has received minimal formal study, there are numerous measures that have been shown to be effective. In addition to safety measures, the use of scheduled awakenings and naps, and in severe cases, clonazepam or tricyclic antidepressants, may be effective. In addition, a formal sleep study may be helpful in revealing SDB as a comorbid disorder for which treatment may reduce the apparent severity of the parasomnia.

Sleep talking, or *somniloquy*, is considered the most common parasomnia, with a reported prevalence of greater than 50% in children between the ages of 3 and 13 years. Somniloquy is often comorbid with sleepwalking and sleep terrors, suggesting a common pathophysiology among the three disorders. Nightmares are common among children and adolescents and are only considered abnormal when they are recurrent and cause excessive distress. Although sometimes associated with post-traumatic stress disorder (PTSD), up to 70% of children report that the content of their nightmares is influenced by media seen on television or at the movies. Effective treatment for nightmares generally includes CBT and rescripting but can also include hypnotherapy and medication. REM sleep behavior disorder, in which subjects become aroused during REM sleep and “act out” their dreams due to loss of body paralysis, is not generally observed in children and is more closely associated with brain damage and dementia.

RLS is characterized by an irresistible urge to move the legs, commonly associated with discomfort and pain, which is partially relieved by movement. In the adult population, RLS has an estimated prevalence of 10% to 15%. PLMD, no longer classified by DSM-5-TR, is characterized by episodic limb movements during sleep, generally involving the leg along with the extension of the ankle and toe.

Only recently has the presence of RLS in children and adolescents been elucidated, with studies generally demonstrating a prevalence rate of RLS among the general pediatric population to average around 2%. There is considerable interest in the relationship between RLS, PLMD, and ADHD. The prevalence of RLS in patients with ADHD is estimated to be between 10.5% and 44%, emphasizing the need to consider both diagnoses and the possible interrelationship between them in children and adolescents.

Prior to the diagnosis of RLS, a child often presents with sleep disturbances and daytime impairments, including extreme fatigue, irritability, and hyperactivity, and the discomfort in the extremities that accompanies the disorder is often confused with growing pains. Other qualitative signs suggestive of RLS include an urge to move the legs with a worsening sensation when sitting or lying down, the alleviation of discomfort through movement, and worsening symptoms in the evening. Several studies have demonstrated an autosomal-dominant inheritance for early-onset RLS. Family history, therefore, is an important indicating factor in diagnosing RLS in children and adolescents.

Most treatments for RLS/PLMD include a pharmacologic component, although proper sleep hygiene is also very important. Children and adolescents with RLS/PLMD should reduce caffeine intake, eliminate alcohol and tobacco use, avoid stimulating activities prior to bedtime, and minimize sleep deprivation. Correlation among low ferritin levels, iron deficiency, and RLS symptoms has been identified in both children and adults. Providing ferrous sulfate supplementation for those children who are deficient has been demonstrated to reduce PLMD symptoms. Other medications have also demonstrated efficacy in reducing symptoms. One randomized, double-blind, placebo-controlled study of children with RLS aged 7 to 12 years found a combination of carbidopa/L-DOPA to be safe and effective. In adults with RLS, dopaminergic medications, including levodopa, pramipexole, and ropinirole, are most often used. Gabapentin has also been shown to improve RLS symptoms in children. Although these medications are prescribed to children and can be effective in reducing symptoms, none has been approved by the FDA for the treatment of RLS/PLMD in children. As a result, the medications should be administered with great caution and intensive monitoring of potential side effects.

Circadian Rhythm Sleep–Wake Disorders. Circadian rhythm sleep–wake disorders occur due to a misalignment between the internal circadian rhythm and the sleep–wake schedule. Over 10% of children are thought to be affected by circadian rhythm sleep problems. Delayed sleep phase type occurs in as many as 5% to 10% of adolescents, while advanced sleep phase type is more common in the elderly and occasionally in preschool children.

A delay in the normal sleep phase is typical in the peripubertal period, leading to later bedtimes. Extensive homework and extracurricular obligations coupled with early school start times frequently result in excessive daytime sleepiness, academic difficulties, irritability, mood instability, attention and memory impairment, and family conflict. Recognizing the detrimental effects of delayed sleep phase on adolescents, the American Academy of Pediatrics has published a policy statement urging a delay in the school start time so that high school students can achieve adequate sleep.

Treatment of delayed sleep phase typically involves sleep education and hygiene, along with gradual advancement of the sleep phase. Bright light morning therapy (5,000 to 10,000 lux), along with low-dose melatonin (approximately 500 mcg) about 4 to 5 hours before the desired bedtime, have both been found helpful in advancing the sleep phase after several weeks. Chronotherapy, or the gradual delay in bedtime by 3 hours each night until the desired bedtime is achieved, can also be effective, although the schedule is very difficult to maintain.

Nocturnal Enuresis. Although no longer classified by DSM-5-TR as a sleep–wake disorder, *nocturnal enuresis* is a frequent occurrence, affecting approximately 30% of 4-year-olds, 10% of 6-year-olds, 5% of 10-year-olds, 3% of 12-year-olds, and 1% of individuals aged 15 years or older. *Primary nocturnal enuresis* (never consistently dry at night) is likely to have a multifactorial etiology and may be due to difficulties with bladder musculature stability, central nervous system arousability, pontine reflex function, internal sphincter tone, functional bladder capacity, nocturnal urine production, and a maturational delay in antidiuretic hormone (ADH) secretion. *Secondary nocturnal enuresis* (previously dry for at least 6 months), by contrast, is typically caused by urinary tract infections, diabetes mellitus, and psychological factors.

The first step in treatment is to educate the family and patient regarding proper sleep hygiene. All caffeine and alcohol intake should be discontinued because both decrease ADH secretion, and late-night fluid intake should be strictly limited. On occasion, the introduction of an afternoon nap will be beneficial by virtue of decreasing the quantity of nighttime deep sleep, thereby making it easier for the patient to awaken to the internal cues of bladder fullness. Finally, parents sometimes find that briefly awakening the child after 2 or 3 hours of sleep and sitting them on the toilet will decrease the frequency of enuresis without significant impairment in sleep efficacy. Should these educational strategies be ineffective, both behavioral and pharmacologic treatments exist for nocturnal enuresis.

Behavioral treatment most commonly involves the use of an enuretic alarm, which has the highest cure rate and lowest relapse rate. Other behavioral interventions include bladder training to increase the number of micturitions during the day or to enlarge bladder capacity. The use of reward systems, such as star charts, has also been shown to be useful. Finally, cognitive therapy (e.g., explaining the enuretic process and maintaining a daily diary), motivational therapy, pelvic floor muscle training, and biofeedback also have some reported utility.

The two medications most commonly accepted as an appropriate pharmacologic treatment for pediatric nocturnal enuresis are desmopressin acetate and imipramine. The use of amitriptyline and anticholinergic agents, such as oxybutynin and tolterodine, however, has also been described.

Sleep Disorders in Children with Psychiatric Disorders

Sleep problems often occur in the context of mental illness, either as primary concerns (e.g., primary insomnia) or secondary concerns (e.g., as a consequence of depression or anxiety). All sleep problems identified early in childhood have been later correlated with anxiety, depression, attentional difficulties, family or parental distress, continued sleep problems, aggression, delinquency, and social problems. Some of the more common psychiatric disorders among children and adolescents and their effect on sleep (and vice versa) will be reviewed here.

Attention-Deficit Hyperactivity Disorder. Numerous studies based on parent reports have demonstrated the prevalence of sleep disorders among children with ADHD to be higher than that for healthy controls, controls with other psychiatric illnesses, and healthy sibling controls by up to fivefold. Actigraphy has shown children with ADHD to have greater variation in sleep onset time, wake time, and single-night sleep duration than controls, and in at least one study, there were significantly more bedtime struggles and longer total sleep duration in the ADHD group. Although children with ADHD often have difficulty settling down to sleep and may sleep more total hours than controls, restricting the amount of sleep a child receives in an experimental setting can also result in symptoms of ADHD, including inattention, hyperactivity, impulsivity, and poor cognitive performance.

ADHD is also associated with snoring during sleep. Habitual snoring has been reported to be three times more common in children with ADHD than in control children drawn from child psychiatry and pediatric clinics. In addition, children with ADHD have been found to have a greater frequency of PLMD while maintaining normal sleep architecture. Children with ADHD also suffer from SDB at a greater rate than expected. SDB has been associated with poor memory performance on standardized psychometric tests and lower academic achievement in children, whereas in adults, deficits in attention, concentration, memory, and verbal and nonverbal intelligence have been noted.

Children with ADHD may also suffer circadian rhythm abnormalities. Among children with ADHD, there is greater inconsistency or instability in time to sleep onset, wake time, and sleep duration than in controls. Furthermore, a high prevalence of ADHD has been shown among children with chronic sleep-onset insomnia, and sleep-onset insomnia in children with ADHD is associated with later dim-light melatonin release. The use of melatonin near bedtime often improves sleep-onset delay among children with ADHD.

Fragmented or disrupted sleep, rather than lessened TST, may be the true culprit behind the daytime behavioral changes and disruptions seen in some children. To that end, recommendations to families with a disruptive or hyperactive child should include an emphasis on basic sleep hygiene. Many children with ADHD take medication regularly, which can further complicate an understanding of their sleep cycle. Much investigation, however, has suggested that children with ADHD still demonstrate more sleep disturbances than controls regardless of medication use, particularly with regard to sleep-onset latency, sleep efficiency, and TST.

Opinions vary greatly on how to manage insomnia due to either ADHD or the stimulant medications that are used to treat it. Typical recommendations include lowering the dose of the stimulant, changing to a shorter-acting preparation, adding a low dose of stimulant (if the insomnia is believed to be secondary to hyperactive “rebound”), or changing to an entirely different stimulant or treatment, such as atomoxetine. If adjunctive pharmacotherapy is used, typical medications include melatonin, clonidine, antihistamines (e.g., diphenhydramine, cyproheptadine), and antidepressants (e.g., trazodone, mirtazapine).

Autism Spectrum Disorders. Children with ASDs are also susceptible to sleep difficulties, with prevalence estimates ranging as high as 90%. Younger children and those with more severe cognitive or intellectual disabilities tend to demonstrate increased sleep problems, which often persist over time. The most frequently reported sleep problems among those with ASD include difficulty falling asleep, restless sleep, frequent awakenings with difficulty falling back to sleep, early-morning awakenings, irregular sleep–wake patterns, shortened duration of sleep, dyssomnias, and parasomnias.

Numerous hypotheses have been put forth to explain the reported sleep difficulties among children with ASD, including failure to recognize environmental and social cues, poorly developed circadian rhythms secondary to social deficits, altered melatonin production, and abnormalities in the hypothalamic–pituitary–adrenal axis and serotonin systems. Many treatments have been proposed for children with ASD with comorbid sleep difficulties.

Various behavioral treatments have been used with varying success in the ASD population. Most favorably, the use of proper sleep hygiene techniques, stimulus control, graduated extinction, scheduled awakenings, and daytime nap restriction have all shown utility among children with autism and intellectual disability. Light therapy, in combination with chronotherapy, may also be useful in children with autism who have circadian rhythm abnormalities. Pharmacologic treatments also have proven to be useful for many, but not all, children with ASD.

Various prescribed medications have reported utility in the treatment of sleep difficulties in the ASD population, including antihistamines, sedating antidepressants, α-2 agonists, benzodiazepines, and antipsychotics. Melatonin has also demonstrated utility in treating children with neurodevelopmental disabilities and autism.

Mood Disorders. Sleep complaints are among the most prevalent symptoms of major depression in children and perhaps the most prevalent symptom among adolescents. Insomnia has been reported to affect nearly 90% of depressed adolescents, whereas *hypersomnia* affects a considerably smaller proportion, perhaps around 25%. Some adolescents—around 10%—continue to experience insomnia even after the remission of depression. The use of actigraphy in diagnosing sleep problems among depressed children and adolescents highlights the compromised sleep quality and abnormal circadian rhythms these individuals experience while depressed. PSG studies in children with depression often demonstrate decreased REM latencies, increased total REM, and increased sleep-onset time. A recent meta-analysis of PSG characteristics among children and adolescents with major depression demonstrated, in fact, that increased sleep-onset latency was collectively the strongest PSG marker of depression in youth.

A number of large prospective population-based and cross-sectional studies have been conducted in the past decade among adolescents, demonstrating a strong association between sleep problems and suicidality. A close relationship between sleep problems and suicide attempts has also been shown in adolescents who have already engaged in nonsuicidal self-injury. A curvilinear dose–response was observed such that those adolescents with reported TST of 8 to 9 hours per night had the lowest risk for suicidal ideation and attempts. Insufficient sleep duration has emerged as one of the most salient risk factors for suicidality, illustrating the importance of sleep screening as part of a suicide assessment.

Many children who are or will be later diagnosed with bipolar disorder present with sleep-related symptoms. When depressed, these children may experience exhaustion, distractibility, irritability, and impulsivity at levels significantly greater than at baseline. In contrast, children experiencing mania may present with unusual amounts of energy and a decreased need for sleep. Such extremes of sleep behavior warrant clinical assessment.

Seasonal affective disorder (SAD) has a reported prevalence of 3% to 4% in children and adolescents, for whom more than 20% of their parents report that their children are sleepier and generally more tired during winter months. Studies of children with a diagnosis of SAD have found that they did not differ from controls in the level of nocturnal activity but did experience significantly lower diurnal activity levels and had less robust circadian rhythms, with a reduction in amplitude.

There have been a small number of studies examining the effects of pharmacologic and behavioral interventions in depressed children and adolescents suffering from sleep problems. A combination of improved sleep hygiene and cognitive behavioral therapy (CBT) has been shown to be effective in managing the insomnia and awakenings associated with depression. In those cases that require further treatment, the most effective drug therapy appears to be a combination of an antidepressant with the short-term use of a sedative or hypnotic. As noted previously, however, some children and adolescents experience a paradoxical reaction to these medications, which severely limits their utility.

Anxiety Disorders. Although it is difficult to establish causality between early sleep problems and the later onset of anxiety and depression, it is clear that anxiety and sleep problems are intimately connected in childhood and that the presence of a sleep problem at age 4 years is significantly correlated with the development of depression and anxiety by age 15 years. Fears at nighttime are common among children, with up to 75% of those aged 4 to 12 years reporting such fears, which generally take the form of animals, fictitious characters (e.g., witches, monsters), being kidnapped, or being teased by peers. Although no specific numbers are available, for the majority of children with DSM-5-TR anxiety disorders, sleep problems typically follow. In addition to the aforementioned concerns, anxiety is also believed to predispose children to parasomnias and nightmares.

The effects of anxiety on sleep are profound, and the clinical presentation depends largely on the specific disorder. Individuals with *generalized anxiety disorder* (GAD) have a wide range of sleep disturbances, which can include insomnia and hypersomnia. These GAD sleep symptoms may be due to central nervous system hypervigilance and hyperarousal such that the individual is unable to relax sufficiently to fall asleep and maintain good sleep quality. Patients with panic disorder tend to present with insomnia, restless and fragmented sleep, and nocturnal panic attacks. These symptoms often result in trouble initiating and maintaining sleep, increased sleep latency, and reduced sleep efficiency.

Children and adolescents with PTSD may experience significant sleep difficulties. Individuals with PTSD often have trouble initiating sleep and experience significant nighttime arousals and disruptive nightmares. Because children with PTSD may have flashbacks to their traumatic experiences, bedtime can become a stressful time in which they experience increased anxiety and are unable to relax sufficiently for restful sleep. The accompanying nightmares, which are a characteristic symptom of PTSD, and increased limb movements, which are associated with nightmares, lead to disrupted REM sleep.

Some research suggests that early trauma may exert a sustained impact on sleep persisting into adulthood. Bader and colleagues demonstrated that childhood trauma was the strongest predictor of delayed sleep-onset latency, poor sleep efficiency, and high nocturnal activity during adult years, even after controlling for stress and depression. Similar data has been derived from studying sleep in females who experienced sexual abuse as children, suggesting that traumatic exposures, more so than a diagnosis of PTSD *per se*, are closely associated with persistent sleep disturbances.

Cognitive and behavioral techniques, systematic desensitization, and imagery rehearsal therapy have demonstrated efficacy in decreasing nightmares and are integral parts of treating the sleep disturbances accompanying obsessive-compulsive disorder (OCD), PTSD, and GAD. CBT, along with systematic desensitization, has been demonstrated to improve the sleep disturbances associated with OCD. In cases that are particularly severe, serotonin-specific reuptake inhibitors such as fluoxetine, sertraline, and paroxetine further aid in the reduction of symptoms. Children and adolescents with PTSD generally have a complex constellation of symptoms, and the primary psychiatric diagnosis must be treated first before the sleep disturbances will resolve. CBT and play or expressive therapies, in which the children process the traumatic events triggering the PTSD, have been shown to be very effective, and antidepressants and anxiolytics may hasten the recovery process in combination with other therapies.

FURTHER READINGS

Alfano CA, Reynolds KC, Scott N, Dahl RE, Mellman TA. Polysomnographic sleep pattern of non-medicated, non-depressed children with generalized anxiety disorder. *J Affect Disord*. 2013;147:379–384.

*American Academy of Pediatrics. School start times for adolescents. *Pediatrics*. 2014;134(3):642–649.

Augustinavicius JLS, Zanjani A, Zakzanis KK, Shapiro CM. Polysomnographic features of early onset depression: a meta-analysis. *J Affect Disord*.

2014;158:11–18.

Bader K, Schäfer V, Schenkel M, Nissen L, Schwander J. Adverse childhood experiences associated with sleep in primary insomnia. *J Sleep Res*. 2007;16(3):285–296.

Becker SP, Kapadia DK, Fershtman CEM, Sciberras E. Evening circadian preference is associated with sleep problems and daytime sleepiness in adolescents with ADHD. *J Sleep Res*. 2020;29(1):e12936.

Blake MJ, Sheeber LB, Youssef GJ, Raniti MB, Allen NB. Systematic review and meta-analysis of adolescent cognitive-behavioral sleep interventions. *Clin Child Fam Psychol Rev*. 2017;20(3):227–249.

Buckley AW, Hirtz D, Oskoui M, et al. Practice guideline: treatment for insomnia and disrupted sleep behavior in children and adolescents with autism spectrum disorder. *Neurology*. 2020;94(9):392–404.

*Chervin RD, Archbold KH, Dillon JE, et al. Inattention, hyperactivity, and symptoms of sleep-disordered breathing. *Pediatrics*. 2002;109(3):449.

Chiu HY, Lee HC, Chen PY, Lai YF, Tu YK. Associations between sleep duration and suicidality in adolescents: a systematic review and dose-response meta-analysis. *Sleep Med Rev*. 2018;42:119–126.

Cortese S, Brown TE, Corkum P, et al. Assessment and management of sleep problems in youth with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*. 2013;52(8):784–796.

Cortesi F, Giannotti F, Ivanenko A, Johnson K. Sleep in children with autistic spectrum disorder. *Sleep Med*. 2010;11:659.

de Bruin EJ, Bögels SM, Oort FJ, Meijer AM. Efficacy of cognitive behavioral therapy for insomnia in adolescents: a randomized controlled trial with internet therapy, group therapy and a waiting list condition. *Sleep*. 2015;38(12):1913–1926.

England SJ, Picchietti DL, Couvadelli BV, et al. L-dopa improves restless legs syndrome and periodic limb movements in sleep but not attention-deficit/hyperactivity disorder in a double-blind trial in children. *Sleep Med*. 2011;12(5):471.

Gershon A, Singh MK. Sleep in adolescents with bipolar I disorder: stability and relation to symptom change. *J Clin Child Adolesc Psychol*. 2017;46(2):247–257.

*Gozal D. Sleep-disordered breathing and school performance in children. *Pediatrics*. 1998;102(3 Pt 1):616.

Hutton JS, Dudley J, Horowitz-Kraus T, DeWitt T, Holland SK. Associations between screen-based media use and brain white matter integrity in preschool-aged children. *JAMA Pediatr*. 2020;174(1):e193869.

Ivanenko A, Crabtree VM, Gozal D. Sleep and depression in children and adolescents. *Sleep Med Rev*. 2005;9:115.

Jones S, Castelnovo A, Riedner B, et al. Sleep and emotion processing in paediatric posttraumatic stress disorder: a pilot investigation. *J Sleep Res*. 2021;30(4):e13261.

Lee J, Finkelstein J. Consumer sleep tracking devices: a critical review. *Stud Health Technol Inform*. 2015;210:458–460.

Ma ZR, Shi LJ, Deng MH. Efficacy of cognitive behavioral therapy in children and adolescents with insomnia: a systematic review and meta-analysis. *Braz J Med Biol Res*. 2018;51(6):e7070.

Malow B, Adkins KW, McGrew SG, et al. Melatonin for sleep in children with autism: a controlled trial examining dose, tolerability, and outcomes. *J Autism Dev Disord*. 2012;42(8):1729–1737.

*Mindell JA, Emslie G, Blumer J, et al. Pharmacological management of insomnia in children and adolescents: consensus statement. *Pediatrics*. 2006;117(6):e1223–e1232.

Morin CM, Vallières A, Guay B, et al. Cognitive behavioral therapy, singly and combined with medication, for persistent insomnia: a randomized controlled trial. *JAMA*. 2009;301(19):2005–2015.

*O'Brien LM. Cognitive and behavioral consequences of obstructive sleep apnea. In: Sheldon SH, Ferber R, Kryger MH, Gozal D, eds. *Principles and Practice of Pediatric Sleep Medicine*. 2nd ed. Elsevier; 2014:231–238.

Paine S, Gradisar M. A randomised controlled trial of cognitive-behaviour therapy for behavioural insomnia of childhood in school-aged children. *Behav Res Ther*. 2011;49(6–7):379–388.

Plazzi G, Ruoff C, Lecendreux M, et al. Treatment of paediatric narcolepsy with sodium oxybate: a double-blind, placebo-controlled, randomised-withdrawal multicentre study and open-label investigation. *Lancet Child Adolesc Health.* 2018;2(7):483–494.

Rosen G. Circadian rhythm disorders in children. In: Barkoukis TJ, Matheson JK, Ferber R, Doghramji K, eds. *Therapy in Sleep Medicine*. Elsevier Inc; 2012:465–474.

Soehner AM, Bertocci MA, Levenson JC, et al. Longitudinal associations between sleep patterns and psychiatric symptom severity in high-risk and community comparison youth. *J Am Acad Child Adolesc Psychiatry*. 2019;58(6):608–617.

Thorpy MJ, Bogan RK. Update on the pharmacologic management of narcolepsy: mechanisms of action and clinical implications. *Sleep Med*. 2020;68:97–109.

54.14 Internet Gaming Disorder in Youth

MICHAEL FELDMEIER, M.D., MICHAEL OGATA, M.D., CAROLY SUE PATAKI, M.D., AND ROXY SZEFTEL, M.D.

The first commercially available video and computer games for children and adolescents emerged in the 1950s and were played by one or two players sitting together at a single computer. In the 1970s, home and handheld video game consoles became available, including the popular Atari Home Pong and the Magnavox Odyssey 2. By the 1990s, the internet gaming industry emerged with an explosion of online games able to be played by multiple players across the globe and available as a commercial network.

Internet gaming disorder (IGD) was proposed in the United States in 2013 and included in the Diagnostic and Statistical Manual (DSM-5) as a “condition for further study.” Clinicians conceptualized it as a potentially impairing psychiatric disorder that might merit inclusion in future revisions of the DSM. At about the same time, the World Health Organization (WHO) began consideration of game addiction as an important public health issue. In 2018, the WHO proposed game addiction as gaming disorder (GD) which was included in the 11th revision of the International Classification of Diseases (ICD-11).

In the United States, given the paucity of evidence-based clinical studies on GD and the lack of consensus on diagnostic criteria, clinicians and researchers decided to delay its inclusion in the DSM-5.

PREVALENCE

Due to variability in screening measures as well as different uses of the diagnostic criteria of IGDs in the ICD-11 and the DSM-5, the prevalence of this disorder in youth and adults remains difficult to determine. Barriers to acceptance of a unified screening tool for IGD include a lack of consensus on symptom identification, variation in clinician view of how much gaming is impairing, and how cultural and national social norms may play a role in identifying IGD as a disorder. One analysis of four survey results conducted in the United States, Canada, United Kingdom, and Germany showed clinically significant impairment due to internet gaming behaviors or disorders with a prevalence between 0.3% and 1% in adults. Most existing gaming rating tools focus on screening for impairment in relationships, school, or work due to internet gaming addiction and gaming behaviors.

DIAGNOSTIC CRITERIA: GAMING DISORDER (ICD-11) AND INTERNET GAMING DISORDER (DSM-5 CONDITION FOR FURTHER STUDY)

The DSM-5 (Internet Gaming Disorder—condition for further study) and ICD-11 (Gaming Disorder) list similar but not identical diagnostic criteria for these disorders. Similarities include a pattern of repetitive or persistent gaming behavior that impairs functioning in multiple domains (e.g., educational, occupational, interpersonal) for a period of at least 12 months (Tables 54.14–1 and 54.14–2).

Table 54.14–1.

DSM-5: Diagnostic Criteria for Internet Gaming Disorder (proposed as a condition for further study)

Persistent and recurrent use of the Internet to engage in games, often with other players, leading to clinically significant impairment or distress as indicated by **FIVE** *(or more)* of the following in a **12-month** period:

1. **Preoccupation with internet games.** *(The individual thinks about previous gaming activity or anticipates playing the next game; Internet gaming becomes the dominant activity in daily life.)*
2. **Withdrawal symptoms when Internet gaming is taken away.** *(These symptoms are typically described as irritability, anxiety, or sadness, but there are no physical signs of pharmacological withdrawal.)*
3. **Tolerance**—the need to spend increasing amounts of time engaged in Internet games.
4. **Unsuccessful attempts to control the participation in Internet games**.
5. **Loss of interests in previous hobbies and entertainment as a result of, and with the exception of, Internet games**.
6. **Continued excessive use of Internet games despite knowledge of psychosocial problems**.
7. **Has deceived family members, therapists, or others regarding the amount of Internet gaming**.
8. **Use of Internet games to escape or relieve a negative mood** *(e.g., feelings of helplessness, guilt, anxiety)*.
9. **Has jeopardized or lost a significant relationship, job, or educational or career opportunity because of participation in Internet games**.

Note: Only **non-gambling Internet games** are included in this disorder. Use of the Internet for required activities in a business or profession is not included; nor is the disorder intended to include other recreational or social Internet use. Similarly, sexual Internet sites are excluded.

Specify current severity:

Internet gaming disorder can be mild, moderate, or severe depending on the degree of disruption of normal activities.

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, pp. 913, 914 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Table 54.14–2.
ICD-11: Diagnostic Criteria for Gaming Disorder

- **Description**
  - Gaming disorder is characterized by a pattern of persistent or recurrent gaming behaviour (‘digital gaming’ or ‘video-gaming’), which may be online (i.e., over the internet) or offline, manifested by:
    1. impaired control over gaming (e.g., onset, frequency, intensity, duration, termination, context)
    2. increasing priority given to gaming to the extent that gaming takes precedence over other life interests and daily activities; and
    3. continuation or escalation of gaming despite the occurrence of negative consequences.
  - The pattern of gaming behaviour may be continuous or episodic and recurrent.
  - The pattern of gaming behaviour results in marked distress or significant impairment in personal, family, social, educational, occupational, or other important areas of functioning.
  - The gaming behaviour and other features are normally evident over a period of at least 12 months for a diagnosis to be assigned, although the required duration may be shortened if all diagnostic requirements are met, and symptoms are severe.
- Exclusions
  - Hazardous gaming (QE22)
  - Bipolar type I disorder (6A60)
  - Bipolar type II disorder (6A61)

Reprinted from World Health Organization. ICD-11 for Mortality and Morbidity Statistics, 2023. https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/1448597234.

BENEFITS AND RISKS OF ONLINE VIDEO GAME PLAY IN YOUTH

Benefits

Most children, adolescents, and adults who play online games identify distinct benefits to (moderate) video gameplay. Most individuals can identify the negative health outcomes of substance use, for example, but negative health and mental health outcomes due to video gameplay are more difficult to determine and differ based on one's psychosocial situation and needs. Most video game players may never experience clinically significant negative outcomes from their video game use, but those that do can experience very significant impacts on their lives. Benefits of internet gaming for youth include promoting problem-solving, cooperation, and connection with peers all over the world.

Educational—Inherent to all video games is a process of learning the rules and objectives of a game. Typically, this is done either through a tutorial preceding the core of gameplay or concepts that are gradually introduced as a player progresses through the initial sections of a game. Game designers must balance a difficulty curve while maintaining engagement–a game should not be so easy that it becomes boring a predictable, but it should not be so difficult to frustrate a player into quitting the game. It is our observation that through video gameplay, like "real-life play," children can acquire prosocial skills such as frustration tolerance, winning and losing gracefully, cooperating with others, and in some cases, delayed gratification (such as when completing a game despite challenges arising).

Social—There are palpable benefits of internet gaming for youth seeking social connections to peers. Especially for youth who are isolated or lonely, the medium of video games may provide a need that is otherwise unmet. Online video gaming may be perceived as a safe space that allows for peer-to-peer self-disclosure and social exploration.

Risks

Avoidance—In some cases, children may rely on video gameplay as a way of avoiding social and academic challenges. Children may delve into gaming instead of schoolwork or in-person social interactions with peers to shield themselves from emotional stress.

Anger—Some gaming environments can lead to hostile environments or intimidation of younger children who are playing with players who are older or who bully less experienced players. Some children, as well as adults, may express their anger during gaming through verbal altercations or through private messaging. There is no data to support those violent video games lead to violent in-person behaviors in youth.

Financial—Children and adolescents have anecdotally been reported to utilize parental credit cards to make large in-game purchases.

BIOLOGIC CORRELATES OF INTERNET GAMING DISORDER

Proposed Neurobiology of Internet Gaming Disorder

IGD shares neurobiologic alterations that are typical for other addictions, such as:

1. Activation in brain regions associated with reward as evident from cue exposure and craving studies and neurotransmitter systems studies that indicate an involvement of dopamine-mediated reward mechanisms
2. Reduced activity in impulse control areas and impaired decision making
3. Reduced functional connectivity in brain networks that are involved in cognitive control, executive function, motivation, and reward

Structural changes, mainly reduction in gray-matter volume and white-matter density, have been found in groups of individuals with IGD.

Executive control networks in ADHD may increase the susceptibility to developing IGD.

COMORBIDITY

Several comorbid psychiatric disorders have been identified to be associated with youth who present with problematic or impairing video game use. The strongest associations are found with anxiety, depression, and ADHD.

Major Depressive Disorder

Depressive symptoms are frequently associated with youth and adults who present with impairment from internet game use. Depressive and IGD symptoms are hypothesized to be bidirectionally related, consistent with findings in other addictive disorders. Depressive symptoms have been shown to be more prevalent in those with IGD compared to healthy controls. Conversely, a reduction of depressive symptoms has been reported in individuals after remission of IGD**.**

A 17-year-old male presents to a community psychiatric outpatient clinic with worsening performance in school, especially in subjects that he does not enjoy, and withdrawal from friends and family activities. He spends most of his time in his room with the door closed. He performs best in math, his favorite subject, but he ignores class assignments, and he rarely turns in homework. His parents report that prior to becoming a teenager, he was friendly and outgoing and did well in school without studying. His parents believe that he did so well in school as a child because of his high native intelligence. His parents are concerned that he may have attention-deficit/hyperactivity disorder (ADHD) because to complete homework now, he requires them to sit right next to him and repeatedly verbally redirect him. He is aware of his poor performance, and his goals still include going to college and possibly graduate school to focus on mathematics. Recently he has started to have crying spells at night and reports self-deprecating thoughts, which make it difficult to sleep. During an interview with him, he reports that he plays online games and watches YouTube videos about gaming for many hours during the day and late into the night, which is news to his parents. Psychiatric evaluation does not reveal a diagnosis of ADHD, but he meets the criteria for major depressive disorder (MDD). He and his parents agreed to a trial of antidepressant medication, and after a month of treatment with sertraline, his symptoms of depression improved. However, his academic performance is largely unchanged. He discloses when asked that when he sits down at his computer with the intent of completing his homework assignments after working for about 15 minutes, he plans to take a short break which usually extends to several hours. Once he becomes involved in gaming, he is unable to stop because he wants to continue to make gains in his game, and he loses all interest in completing his homework for the day. He agrees to a family session in which he reveals how many hours per day he spends gaming and his inability to control it. His parents are surprised, but they are pleased that he is willing to discuss his gaming in the session. His psychiatrist recommends that his parents enroll him in a 2-hour after-school homework program so that he can focus on homework without the option and temptation of playing online games and that he begins weekly therapy to help him monitor and control his cravings for games.

Attention-Deficit/Hyperactivity Disorder

ADHD is commonly comorbid with problematic internet gaming and maybe a mediating factor for IGD. It has been reported that features of ADHD in undiagnosed individuals with IGD are a significant predictor of comorbidity of these two disorders. Neurologic reports studies have suggested overlap between the pathways involved in ADHD and IGD. Compared to pure-IGD, when ADHD is comorbid, recovery of IGD is less likely, and there is a higher likelihood of recurrence of IGD.

A 9-year-old boy with a history of well-controlled ADHD and intermittent explosive disorder presents to the psychiatric outpatient clinic due to recurrent severe tantrums when his parents limit his time on the computer and insist that his online game be turned off. His parents are concerned about his significantly increasing cravings for online gaming and his out-of-control reactions when limits to his gaming are set. His reactions to his gaming being curtailed have escalated from screaming and crying to hitting and kicking his parents, to throwing household objects to recently physically attacking his mother. The boy feels sad and guilty after his tantrums and often apologizes to his parents, yet the same thing happens the next day. His parents report that there is a noticeably more intense pattern of tantrums occurring after longer and more frequent gameplay sessions, especially when he is playing online with friends and classmates. His parents are unsure of how to limit his online gaming while still supporting his social engagement with peers, especially since online gaming has become the main avenue for social interaction in the context of the COVID-19 pandemic. His parents helped to create a behavioral token economy that provides rewards for social nonvideo game engagement with friends and allows him to earn reasonable online gaming time. Consistent time limits were necessary and implemented by his parents, and after struggles and resistance for a few weeks, the boy was able to accept the limits, enjoy other online activities with peers that do not involve video gaming, and his reactions to limits on his gaming time lessened without escalating to physical aggression.

Anxiety Disorders

Anxiety has been frequently reported as comorbid with IGD. When anxiety is present with IGD, the risk of developing depression is increased. Correlations have also been reported between IGD, generalized anxiety disorder, and social anxiety disorder.

A 13-year-old boy presents at a psychiatric outpatient clinic with resistance to engage in online school and refusal to give presentations by Zoom. His parents brought him to the clinic for evaluation to find out if his online gaming is interfering with his school participation. He has a history of discomfort with meeting new people, speaking in public, worrying more than his peers about what others think of him, and potential natural disasters that may occur. He has had one episode of a panic attack in school after he gave a wrong answer in front of his class, requiring his parents to come and pick him up. He has been diagnosed with generalized anxiety disorder and, in the past, has demonstrated some improvement with cognitive behavioral therapy and fluoxetine medications. He has never had significant depressive symptoms or suicidal thoughts. Currently, he is not on any medication and is not in therapy. His performance in school has been consistently good, usually getting As and Bs. In his spare time, he plays a popular online game and then takes game footage and edits it in a creative way, and uploads those videos to YouTube. He has a following of several hundred subscribers. His parents have been wary of this activity and set firm limits and monitor what he is uploading. His parents emphasize that they have set a family rule that he is not to show his face or swear and that he blocks hostile viewers (trolls) who leave demeaning comments on his videos. He spends about four hours per day online gaming, but he continues to finish his homework, go to sleep at a reasonable time, and participate in family activities. Even though he spends up to four hours on weekdays and up to six hours on weekends either playing games or editing gaming footage, he is accomplishing the tasks required for his age. During the psychiatric evaluation, he reveals that he agrees with his parents' limits on his online gaming, and he feels that he can control it without distress. He discloses that over this past year of Zoom school, his fear of speaking in front of the class has escalated, and especially his anxiety when doing online presentations has made school problematic. In a feedback session with him and his parents, his parents are reassured that his online gaming is not contributing to his trouble with school and that his performance anxiety pertaining to public speaking in the classroom is the source of the anxiety for him. He and his parents agree to a trial of propranolol which he takes on days that he must give a Zoom presentation in school.

TREATMENT OF INTERNET GAMING DISORDER

Evidence-based, gold-standard treatments for IGD have not yet been demonstrated. There is anecdotal evidence for efficacy with modalities including parent training, motivational interviewing, group therapy, and cognitive-behavioral therapy. Currently, several residential therapeutic treatment programs for IGD are in operation in the United States and abroad, which offer “digital detox,” however, long-term outcomes have yet to be verified through evidence-based or randomized clinical trials. Pharmacologic treatment for IGD has been reported only anecdotally, including a case report of efficacy with bupropion. There is no evidence currently to recommend the use of psychopharmacologic agents for IGD other than in the treatment of comorbid psychiatric conditions.

Impact of Treating Comorbid Disorders on IGD

There is currently no evidence-based data to demonstrate the impact of treating comorbid disorders on IGD. It has been reported anecdotally, however, that behavioral interventions in youth with comorbid depressive symptoms and IGD symptoms may be effective in reducing the severity of both disorders. IGD is believed to be a behavioral addiction that is not affected by the concurrent effects of substance use or its treatment.

CLINICAL CONSIDERATIONS

Parent Recommendations on Gaming Limit Setting in Children and Adolescents

Age-appropriate game choice and boundaries are important considerations for parents and an essential skill for clinicians to be able to identify signs of problematic gameplay and provide guidance for parents. Clinician consensus suggests that among other risk factors for developing IGD in youth is many hours of video gameplay per day. Consequently, it is beneficial for clinicians to screen for excessive use in their child and adolescent patients and counsel parents to set age-appropriate playtime limits for gaming. While there is no known specific time-dependent exposure to online gaming hours that would lead to a given child being susceptible to IGD, clinical observation and monitoring of a child’s overall functioning is important. Parents' familiarity with and understanding of expectations of schoolwork and other balanced tasks of childhood (exercise, socializing with peers, extracurricular activities, enrichment) at different developmental ages will help them to determine how much gaming and screen time would be appropriate for their child's circumstances.

Clinical Recommendations

As the potential for online gaming impairment may rise along with increased accessibility of video and the current ubiquitous nature of online gaming in the lives of youths, screening for IGD is a worthwhile part of age-appropriate screening questions during mental health evaluations and during clinical follow-up sessions. Video gameplay in youths, especially problematic use, may be under-reported by youth and often unknown to parents, which makes it even more important for clinicians to ask youth as well as parents for detailed information on their child's online gaming habits. It is suggested that clinician screening questions for youth begin with open-ended, plain-language questions such as "Tell me about what gaming means to you," "What happens when you find yourself gaming too much"?, and "What does gaming get in the way of for you"?

A social media screening questionnaire created by the authors. Michael Feldmeier pertaining to online use in youth, The Social Media Initial Questionnaire-revised (SMIQ-r), can be helpful in systematically evaluating a child's online activities, including video game use. The SMIQ-r is included in this book in Chapter 55.15, "Social Media in Youth." The SMIQ-r may aid clinicians in elucidating a more nuanced picture of a child's gaming internet use experience.

Many families find it helpful to view video game use as a privilege for their children and use video time as a reward to motivate and inspire their children to accomplish their necessary tasks (homework, chores, and fitness). It is important that children and adolescents have a variety of in-person as well as online activities to engage in, including physical fitness activities. The American Academy of Pediatricians lists recommended time limits based on age (see https://services.aap.org/en/patient-care/media-and-children/)

Video game abstinence is not usually recommended since it may promote avoidance of issues rather than help a child learn self-regulation. Gaming is often a coping strategy for many children. In moderation, gaming can represent a beneficial activity allowing children an opportunity for emotional processing, social play, learning, and a reprieve from life stressors. Excessive gaming, however, can become a negative behavior that may block opportunities for emotional development. It is important for clinicians to understand that although video game use is ubiquitous in today's youth, addiction to video gameplay leading to impairment is rare.

Public Policy Impact on Clinical Sequelae of IGD

Emerging worldwide public health awareness of GDs and the impact of problematic gaming behaviors in youth leading to impaired functioning has made it an important public health concern for further investigation. The following questions come to mind: "What can be done to reduce GD prevalence"? and "Who should take responsibility"? Currently, in the United States, the Entertainment Software Ratings Board provides oversight within the video game industry regarding transparency of the content within video games. Outside of age restrictions on online gambling platforms, however, there is no oversight pertaining to the content within the game design used to attract players to initiate and continue the game. As gaming has become more widespread, increased data sets are accessible to gaming companies to use for market research to expand the popularity of a given game and increase players' overall playtime. Clinicians can be helpful to their child and adolescent patients once they understand their patients' video game use and its impact on their lives.

FUTURE DIRECTIONS

Randomized controlled studies investigating online gaming in children and adolescents are lacking in the United States. Available international data by Na Rim Kim and colleagues and Cheng-Fang Yen and colleagues (see References) provide trends about international gaming in adult and adolescent patients, but generalizability and study power are varied given the inconsistencies in self-report tools used in these studies. An internationally recognized screening tool would be instrumental in allowing research to progress. A published review by Daniel L King (see References) compared 32 different forms with no clear clinical tool standing out among the rest in assessing for GD/IGD.

FURTHER READINGS

Andreassen CS, Billieux J, Griffiths MD, et al. The relationship between addictive uses of social media and video games and symptoms of psychiatric disorders: a large-scale cross-sectional study. *Psychol Addict Behav*. 2016;30(2):252–262.

Bonnaire C, Liddle HA, Har A, Nielsen P, Phan O. Why and how to include parents in the treatment of adolescents presenting Internet gaming disorder? *J Behav Addict*. 2019;8(2):201–212.

Gentile DA, Bailey K, Bavelier D, et al. Internet gaming disorder in children and adolescents. *Pediatrics*. 2017;140(Supplement 2):S81–S85.

Greenfield DN. Treatment considerations in internet and video game addiction: a qualitative discussion. *Child Adolesc Psychiatr Clin N Am*. 2018;27(2):327–344.

Kim NR, Hwang SS, Choi J, et al. Characteristics and psychiatric symptoms of internet gaming disorder among adults using self-reported DSM-5 criteria. *Psychiatry Investig*. 2016;13(1):58.

Kim M, Kim D, Bae S, Han DH, Jeong B. Aberrant structural network of comorbid attention deficit/hyperactivity disorder is associated with addiction severity in internet gaming disorder. *Neuroimage Clin*. 2020;27:102263.

King DL, Chamberlain SR, Carragher N, et al. Screening and assessment tools for gaming disorder: a comprehensive systematic review. *Clin Psychol Rev*. 2020;77:101831.

Ko C, Yen J, Chen S, Wang P, Chen C, Yen C. Evaluation of the diagnostic criteria of Internet gaming disorder in the DSM-5 among young adults in Taiwan. *J Psychiatr Res*. 2014;53:103–110.

Kuss DJ, Griffiths MD. Online gaming addiction in children and adolescents: a review of empirical research. *J Behav Addict*. 2012;1(1):3–22.

Liu Lu, Yao YW, Li CR, et al. The comorbidity between internet gaming disorder and depression: interrelationship and neural mechanisms. *Front Psychiatry.* 2018;9:154.

Pataki C, Bostic JG, Tabor-Furmark KS, et al. Functional assessment of social media in child and adolescent psychiatry. In: Beresin EV, Olsen CK, eds. *Child and Adolescent Psychiatry and the Media.* Elsevier; 2019:93.

Przybylski AK, Weinstein N, Murayama K. Internet gaming disorder: investigating the clinical relevance of a new phenomenon. *Am J Psychiatry*. 2017;174(3):230–236.

Razjouyan K, Khademi M, Dorandish ZY, Davari-Ashtiani R. An investigation into the frequency of addiction to video games in children with attention-deficit hyperactivity disorder. *J Family Med Prim Care*. 2020;9(2):669–672.

Rho M, Lee H, Lee T, et al. Risk factors for internet gaming disorder: psychological factors and internet gaming characteristics. *Int J Environ Res Public Health*. 2017;15(1):40.

Rumpf HJ, Achab S, Billieux J, et al. Including gaming disorder in the ICD-11: the need to do so from a clinical and public health perspective. *J Behav Addict*. 2018;7(3):556–561.

Wang CY, Wu YC, Su CH, Lin PC, Ko CH, Yen JY. Association between Internet gaming disorder and generalized anxiety disorder. *J Behav Addict*. 2017;6(4):564–571.

Weinstein A, Leroux M. Neurobiological mechanisms underlying internet gaming disorder. *Dialogues Clin Neurosci*. 2020;22(2):113–126.

Zajac K, Ginley MK, Chang R. Treatments of internet gaming disorder: a systematic review of the evidence. *Expert Rev Neurother*. 2020;20(1):85–93.

▲ 54.15 Social Media and Youth

Christina Guest, M.D., Yeun J. Lim, M.D., Michael Feldmeier, M.D., Roxy Szeftel, M.D., and Caroly Sue Pataki, M.D.

Social media is defined as “electronic communication including websites for social networking and blogging through which users create online communities to share information, ideas, personal messages and other content such as pictures and videos.” Social media, as we know it, originated in the 1990s when groups began forming on the newly accessible internet. This was a novel way for many people, near and far, to share thoughts, social networking, support, and ideas. This was the start of online communities.

Then, social media evolved to chat room formats, such as AOL and Yahoo messaging, in the late 1990s. It later transitioned into blogging websites, online gaming sites, and dating websites. The first social media site, Six Degrees, was created in 1997, allowing users to create profiles and make online friends. MySpace and LinkedIn became prominent in the early 2000s, and LinkedIn became the first job- and career-based social media platform. YouTube emerged in 2005, creating an innovative way for people to communicate from great distances. In 2006, Facebook and Twitter emerged and continue to be among the most popular social media sites. A whole host of additional sites, including Reddit, Tumblr, Foursquare, Snapchat, and Instagram, have emerged to fit specific social networking needs in various subcultures of youth and adults.

The proliferation of social media and media technologies has made it a mainstay of daily life for today’s children and adolescents. Millennials, born between 1977 and 1995, were the first generation of social media natives who grew up with the transition and development of many social media sites and increased access to daily internet technology. Generation Z, born between 1996 and 2015, have incorporated social media as one of their primary forms of communication and social development. These youth have grown up with social media technologies which have seamlessly been integrated into their development, communication styles, and social schemas.

According to the Pew research report of 2021, at least 60% of today's U.S. youth have interacted with a smartphone or other media technology before the age of 5 years, and over 90% of teens in the United States have access to a smartphone device. A primary purpose of social media in youth is to increase their sense of social connectedness and belonging with peers, friends, and family and increase social interactions while also being used for educational purposes and, of course, entertainment. Among children and adolescents, the most popular social media platform is YouTube, as reported by up to 80% of parents interviewed by Pew researchers in 2021.

According to the 2021 Pew Research survey, after YouTube, U.S. youth most often use Instagram, WhatsApp, and TikTok. Of the 95% of U.S. adolescents who report having a smartphone, 45% endorse being "constantly online" and with the primary reason to "just pass time" (57%). Male and female adolescents equally report the use of smartphones and social media to connect with others and to learn new things. Female adolescents, however, report that they are more likely to use smartphones to avoid in-person social interactions compared with males.

In younger school-aged children, according to the Pew Report in 2021, YouTube, TikTok, and Snapchat are the most popular. Young children are often exposed to digital media by the age of 1 year. Some parents report that they rely on social media to calm or "babysit" their children when they are busy with other tasks. Among school-aged children, the vast majority of them use their parents' devices, making it easier for parents to monitor and control their social media use compared to adolescents.

BENEFITS OF SOCIAL MEDIA USE FOR YOUTH

The ubiquitous nature of social media and media technology in the daily lives of children, adolescents, and adults leads to examining and recognizing its benefits. Today's children and adolescents have social media and media technology integrated into their lives from very early on, often within the first year of life. It is therefore understandable that youth would integrate social media and media technologies into their social schemas, that is, a natural way that they communicate with peers and play online recreational games. With the explosion of applications for cell phones and computers and the easy accessibility of the internet, social media and media technology are here to stay and have identified themselves as beneficial tools for many youths. Social media and media technology are now used regularly by youth for increased social connectedness, including for emotional support and building of social circles and support systems, as well as for increased access to resources, such as mental health resources and education. However, the benefits of social media and media technologies continue to seem endless, and the aforementioned benefits are just a small subset of some applicable benefits for youth with mental illness.

Social Media to Increase Social Connectedness in Youth

With the seamless integration of social media into youths' lives, it is without surprise that it is used as a tool for constant communication with their social circles, including family and friends. Applications such as Facebook, WhatsApp, Instagram, TikTok, and YouTube are just a small subset of the numerous social media and media technology programs that youth use to stay connected. Having social connectedness without needing to be physically present with others can increase a sense of community and friendships for some youth. In addition, social media and media technology enable youth to connect to peers and other youth in far-away neighborhoods, cities, and even countries, enabling youth to find more friends and peer connections.

Noah, a 13-year-old male with social anxiety disorder and attention-deficit hyperactivity disorder (ADHD), presents to the outpatient child and adolescent psychiatry clinic with his mother for a follow-up visit. Since kindergarten, he had been extremely shy, quiet, and worried about being judged by other kids, especially in a large group. He has only one close friend at school. His mother was conflicted when he asked for a smartphone for his 13th birthday, as she worried that having a smartphone may steer him away from social interactions further. During this visit, the mother shared that he had been using TikTok and Instagram with his mother's permission and supervision. Noah sheepishly tells the psychiatrist that he had made several new online friends on TikTok and Instagram. They were all following and commenting on other teenagers' accounts that post funny skits about their favorite Japanese Manga characters. He is often sharing funny videos and memes with them. He no longer feels alone anymore because he realizes there are others with the same interests in many different cities and states. He is excited that soon they will meet online and play video games together.

Social Media for Youth Seeking Support

A frequent scenario among youth experiencing emotional distress without adequate support at home or who don't have clinical support, caregivers, or peers in person in place is to seek support through social media. This modality may be more desirable, especially for those youth who are not in danger of harming themselves or others and feel embarrassed or ashamed to ask for help in person. A youth's increased ability to seek support online from peers may help them feel less alone and, at times, help them to access useful resources and online education about their distress and available interventions. These can lead to positive outcomes for youth. An example of youth using social media for support can even take the form of posting photos or content related to feeling lonely and depressed and receiving empathic responses. Youth who are lonely or isolated often lack other support. Thus having the social media platform as a potential for a new sense of community and support that otherwise does not exist for them in person can be beneficial. See case report.

An adolescent female, Lena is a 15-year-old teen with a history of depression who has been feeling isolated and tearful. She turns to social media for support by posting pictures of herself looking wistful, lonely, and depressed on Instagram. Soon, she discovered that there were many other adolescents on Instagram who began to follow her with similar struggles who felt lonely, isolated, and unable to access camaraderie in person. Although Lena feels isolated because she does not have an in-person support system of peers, through her Instagram posts, she is able to share her feelings online and gain a sense of support and community, resulting in feeling less alone and her spirit lifted. Although Lena continues to feel lonely, the positive reactions and encouragement she receives online feel supportive, she is not in danger of harm, and she finds this modality of support very helpful.

Social Media for Youth to Seeking Mental Health Information

Social media is often used by youth to seek information on mental health and potentially to learn about mental health treatments. Youth from Generation Z (born 1996 to 2015) exhibit a strong preference for social media and media technology communication to learn about mental health and potentially receive online mental health treatments. This preference appears to be especially prominent in at-risk youth with emotional vulnerabilities or a history of mental health problems. Clinicians who can use social media platforms to connect with at-risk youth may increase the likelihood that these youths will access mental health interventions. Some youth report a sense of reassurance and security from online support, such as online support groups. An increasing number, albeit still a small percent, of mental health professionals are joining online support groups to provide information and guidance for youth in distress with the goal of increasing youth emotional well-being.

A 17-year-old female, Ricka, joined a clinic WhatsApp chat group for support, information, and potential linkage for treatment. According to the clinician who led the group, Ricka often posted questions about sexual abuse. The social worker was able to respond to such posts with a private message back and an offer to help Ricka find an in-person evaluation in her own city. Through these confidential and private communications, Ricka felt connected to the social worker and developed enough trust to accept her guidance and seek mental health treatment in her own city.

Surveys indicate for some youths, online modes of communication with mental health providers offer less shame, increased youth engagement, as well as an increased sense of control over their sense of emotional stability.

In addition, social media and media technology are now available to some youth around the globe, increasing access to mental health treatment for those who otherwise might not have access. For example, an African study shows an increase in retention, education, and support for youth in human immunodeficiency virus (HIV) treatment programs within Africa by having a social media platform for all to join. This allowed for support, education, and increased access to treatment guidelines and recommendations to a larger group than otherwise would be able to join due to limitations on travel or inconvenience that in-person support groups and education would require. Similarly, digital interventions have been utilized for the treatment and prevention of substance use disorders in youth, another topic that often has stigma associated with it, like HIV, potentially limiting those who would reach out for help. By integrating advertisements within social media directed at adolescents, researchers were able to imbed a substance use and addiction scale, along with digitized adaptations of motivational interviewing to help increase acceptance of substance use treatment and recovery programming. The anecdotal successes of remote engagement for youth, teachers, and clinical providers suggest that online computer-based educational settings and clinical modalities of mental health can be effective and, in some cases, increase access to education and mental health care.

RISKS OF SOCIAL MEDIA USE FOR YOUTH

Parents and caregivers may or may not be aware of what their children are viewing on social media and how much time they are spending on it, especially if they use mobile phones when caregivers are not home or at night in their rooms. Youth of all ages who use social media may be exposed to violence and risk-taking behaviors, videos of stunts or challenges that may inspire unsafe behavior, sexual content, negative stereotypes, substance use, cyberbullies and predators, advertising, misleading or inaccurate information, identity theft, or hacking.

While adequate screen time may increase social connectedness in youth, excessive screen time may lead to impairment in everyday life, has been linked in some youth to sleep problems, lower grades, less time with family and friends, lack of physical activity, weight problems, mood problems, poor self-image, and a fear of missing out (FoMO). Surveys of youth suggest that excessive social media may be contributing to increasing mental distress among youth. Some researchers have hypothesized that problematic social media use occurs with greater frequency in youth with a diminished ability to regulate impulses, display stress or anxiety when social media is restricted, and demonstrate constant thoughts about social media. The potential risks of social media are also associated with increased time spent on social media.

Data from a survey of adolescents from 29 countries showed that adolescents' intense social media use was, in some cases, positively associated with well-being and, in other cases, negatively associated with a reported sense of well-being. Factors associated with a positive or negative response to intense social media use include the prevalence of social media in the country and the level of vulnerability of the youth. In countries with a lower prevalence of intense social media use, those youth with intense use reported more frequent psychological complaints, lower life satisfaction, and lower levels of family support than youth with less social media use. In countries with a higher prevalence of intense social media use, however, intense social media use was predominantly positively correlated to family support and life satisfaction.

It is important to consider that intense use of social media is not as universally correlated with negative well-being in youth. Rather, negative outcomes of intense social media use are based on youth vulnerability due to the contributions of many possible factors, including lack of family support, poor school achievement, and prior mental health challenges. Thus, risks to a sense of well-being in youth with intense social media use may arise not simply from the time spent on social media but perhaps due to many factors, including an inability to monitor and mediate their own use and a preoccupation with social media use.

Bidirectional Association between Social Media Usage and Depressive Symptoms

Social media use appears to be increased in youth with depressive symptoms, loneliness, and high levels of stress. Loneliness reported by youth is also a predictor of more intense usage, and in some cases, it may help to alleviate the depressed mood. Depressive symptoms, and sleep problems, have been shown to be a predictor of excessive social media use.

Negative impacts of social media on depressive symptoms are endorsed more frequently in adolescents with high levels of social isolation and low levels of in-person interaction. In contrast, youth with higher levels of in-person socialization appear to be protected against the negative consequences of excessive social media use and report negative outcomes less often.

Depressive symptoms such as rumination and brooding appear to exacerbate the negative effects of social media, according to surveys in youth. Surveys of college students report that those who tend to compare themselves with others more often also report higher rates of negative affect, such as feelings of low self-esteem after browsing the Facebook profile of an acquaintance, relative to students who infrequently compare themselves to peers.

A 2021 systematic review of the literature suggests that passive social media usage (browsing others' content or comments) may aggravate depressive symptoms, such as loss of interest or low mood. High social media use appears to be predictive of depressive symptoms and low offline social support from both family and peers. Addictive attitude toward Facebook is found to be positively correlated with depression, regardless of age (age range 10 to 18) and gender.

Social Media Impact on Youth Body Image and Risk of Self-Harm

It is well known that many social media platforms showcase images of “ideal” and idealized notions of beauty and body image. Surveys of youth and adults indicate that frequent and prolonged social media use is associated with body image concerns and disordered eating in both females and males. This effect is more pronounced in youth who tend to frequently compare their appearance to others in person. Social media is also often used by youth to join others with similar concerns or beliefs either for support or to become a persuasive group for a specific belief or cause. On social media, eating disorders are common themes. Some online groups that attract mainly adolescent girls who are interested in extreme weight loss and are seeking guidance regarding the best ways to starvation diets provide this information. Online groups such as these provide videos and coach users on how to pathologically lose weight. Although there is some censorship by these sites to monitor the content provided, there are also groups that offer opposing viewpoints, for example, officially aimed to raise awareness of the negative consequences of anorexia.

Ashley, a 16-year-old female with no psychiatric history, went to an appointment to see her pediatrician with her mother. Her mother says that ever since Ashley started to spend 3 to 4 hours a day scrolling through fashion influencers' Instagram posts, she has been trying to skip breakfast and dinner often. Ashley does acknowledge that she has felt like she is "fat" compared to the fashion influencers who often show off their thin bodies in tight clothes or swimsuits. The pediatrician suggested that Ashley cut down on social media and non–school-related electronic use time to 1 hour a day. Her mother started using a timer for social media use, and Ashley agreed to unfollow the influencers who posted photos of their bodies often. In a couple of months, she started to eat breakfast and dinner again.

RISKS OF INTENSIVE SOCIAL MEDIA USE IN YOUTH

Associations Between Social Media Use and Anxiety

Some social media features, such as the ability to receive or post online approval through the number of "likes," have been found to be associated with excessive social comparison, rumination, and increased anxiety symptoms. Youth with pre-existing traits of "perfectionism" may be more vulnerable to reacting to other users' "likes" and "dislikes" of their posts. Social anxiety is a topic that is discussed frequently online by youth in the

United States. Online socializing can be attractive for youth with fear of being negatively judged by others during real-life social interactions. Some youth feel that by communicating and posting online, they will avoid being judged negatively. When youth with fears of in-person social interactions receive symbols of "like" to their online posts, it may reinforce their online choices and thereby replace opportunities to engage in social activities in person that promote socioemotional development. The more that youth share about their lives online, the more they are at risk of being afraid of missing updates and an urge to constantly check their profiles for feedback. This pervasive anxiety and urge to be constantly in fear of not being included is known as the "fear of missing out" (FoMO). This FoMO in youth is marked by worry that other people might be enjoying fun or satisfying experiences without them, pushing them to be constantly on social media to check. High levels of FoMO seem to promote a sense of addiction to online social media platforms.

Intensive Social Media Use and Cyberbullying

Cyberbullying commonly involves cruel messages or postings made online as well as sharing of unflattering or inappropriate pictures online without permission. A study of 181,000 adolescents in Europe and Canada found that youth who engage in social media all day long and with diminished ability to regulate their social media impulses are at increased risk of being cyberbullying perpetrators. Girls and young women tend to spend more time on social media than boys and thus are exposed to the risk of cyberbullying and its negative impacts on mental health than their male counterparts. Youth who have been victims of cyberbullying are at risk for long-standing negative emotional impacts. In some cases, the odds of suicidal ideation, plans, and attempts are higher in youth who have experienced cyberbullying.

Poor Sleep With Excessive Social Media Use

Excessive smartphone use in youth on a regular basis is frequently associated with less sleep at night and compromised mood, school achievement, and relationships with peers and family. Poor sleep may mediate the relationship between nighttime smartphone use and subsequent depressed mood, according to a longitudinal study.

CLINICAL CONSIDERATIONS

Schools, family, and clinical settings are all potential arenas for the detection of adolescents with problematic social media use, as well as for providing guidance and initiating interventions aimed at monitoring and reducing it.

Parent and caregiver media literacy can help families monitor social media and set helpful goals and limits for youth. Youth who are nurtured by positive parental influence and a supportive school environment are more likely to be open to limits placed on their social media usage and recognize its challenges.

Encouragingly, youth are increasingly recognizing the negative impact of social media on their lives and starting to take steps to mitigate this. According to a recent poll, 54% of teens felt they spend too much time on their cell phones, and about half reported cutting back on the time they spend on them.

A Clinical Tool for Clinicians to Explore Social Media Use in Youth

Clinicians may benefit from a structured instrument to explore the positive and negative aspects of a youth's social media use. One targeted tool: **Social Media Initial Questionnaire-revised (SMIQ-r)**, created and revised by the authors, is a self-report questionnaire that can be offered to youth at clinic visits or electronically by using an online link. See Figure 54.15–1.

Social Media Use in Psychiatric Clinic Youth Before and During the COVID-19 Pandemic

When the authors initially offered the SMIQ-r to clinically assess social media use in a Los Angeles County, California, youth psychiatric clinic outpatient population in 2019, youth reported the following:

- More than half reported that social media interferes with homework
- More than half reported that social media interferes with sleep
- More than one-third reported that social media interferes with school
- Most reported that social media interferes with overall functioning

The COVID-19 pandemic has accelerated the consideration of utilizing digital and social media platforms in mental health evaluation and treatment in the recent period. During the COVID-19 pandemic, clinics, schools, and social activities were forced to discontinue in-person interactions leading to gaps in usual social connectedness, education, and treatment for youth, caregivers, and clinicians. Although most youth and adults expressed concern, in some ways, digital and remote platforms increased the ease of connectedness with those who are far away and not able to be reached in person.

During COVID-19, the following addendum questions were added to the SMIQ-r:

- Since school went online, when do you begin your online school? Earlier, later, about the same time.
- Since school went online, when do you go to sleep? Earlier, later, about the same time.
- Since school went online, when do you wake up? Earlier, later, about the same time.
- Since school went online, how many hours do you sleep at night? Less, more, about the same.
- Since school went online, how hard is your schoolwork? Easier, harder, about the same.
- In what ways is schoolwork different than before?
- Since school went online, I am feeling: more anxiety, less anxiety, about the same amount of anxiety.
- Since school went online, I am feeling: more sad, less sad, no change in my mood
- I prefer to have appointments with my psychiatrist: online (with video or phone), or in person.
- What is the biggest change in your life since the COVID-19 pandemic? And did it make your life harder or easier?
- In what ways did the change from the coronavirus make your life harder or easier?

In comparison to before the pandemic, reassessing social media and media technology use in outpatient psychiatric clinic youth from the same clinic during the pandemic, some differences were observed. Compared to the SMIQ-r data in 2019, youth endorsed an even stronger preference for in-person friends, in-person school, and in-person psychiatrist appointments. There was a trend for increased isolation and depression in youth reports during the COVID-19 pandemic. This suggests that an inability to see friends and/or go to school in person was associated with increased levels of anxiety symptoms and sadness in youth.

Electronic SMIQ-r Use

SMIQ-r electronic survey format that most youth filled out on their smartphones appeared to be user-friendly and more efficient than the in-person paper version and reinforces the appeal of electronic and digital formats for patient engagement and as a communication method for youth going forward (Table 54.15–1).

Johnny is an 8-year-old boy being seen in an outpatient mental health clinic for routine medication management of his ADHD. He has been asking his mother about using the family iPad to go on YouTube and "watch videos like my friends." Johnny's mother is concerned about allowing Johnny to use the iPad and social media and asks the clinician for advice. Johnny's mother is reminded of the importance of setting good boundaries with social media use, modeling good social media behaviors herself, and discussing social media safety concerns. Her questions were answered, and boundaries were discussed with Johnny as part of the conversation to encourage collaboration in creating social media family rules. Both Johnny and his mother felt comfortable with the plan for his social media use by the end of the appointment.

CONCLUSIONS

There are many benefits of social media for youth, including enhancing their connectedness and ability to communicate with other youth around the globe, enabling them to seek support for important personal stressors, and seeking education and resources on limitless topics. There are some potential associated risk factors that may arise with excessive social media use: exposure to inappropriate content for a youth's developmental age, and engaging with strangers who have harmful intent. It is critical to note that most youth continue to spend the same amount of time in person with friends while engaging in reasonable or even high levels of social media and technology use compared to youth who use little social media. Overall, with some guidance and review of the risks and benefits of social media use and media technology, clinicians can work with youth and their families to have a safe and beneficial social media experience as an integrated part of a youth’s development.

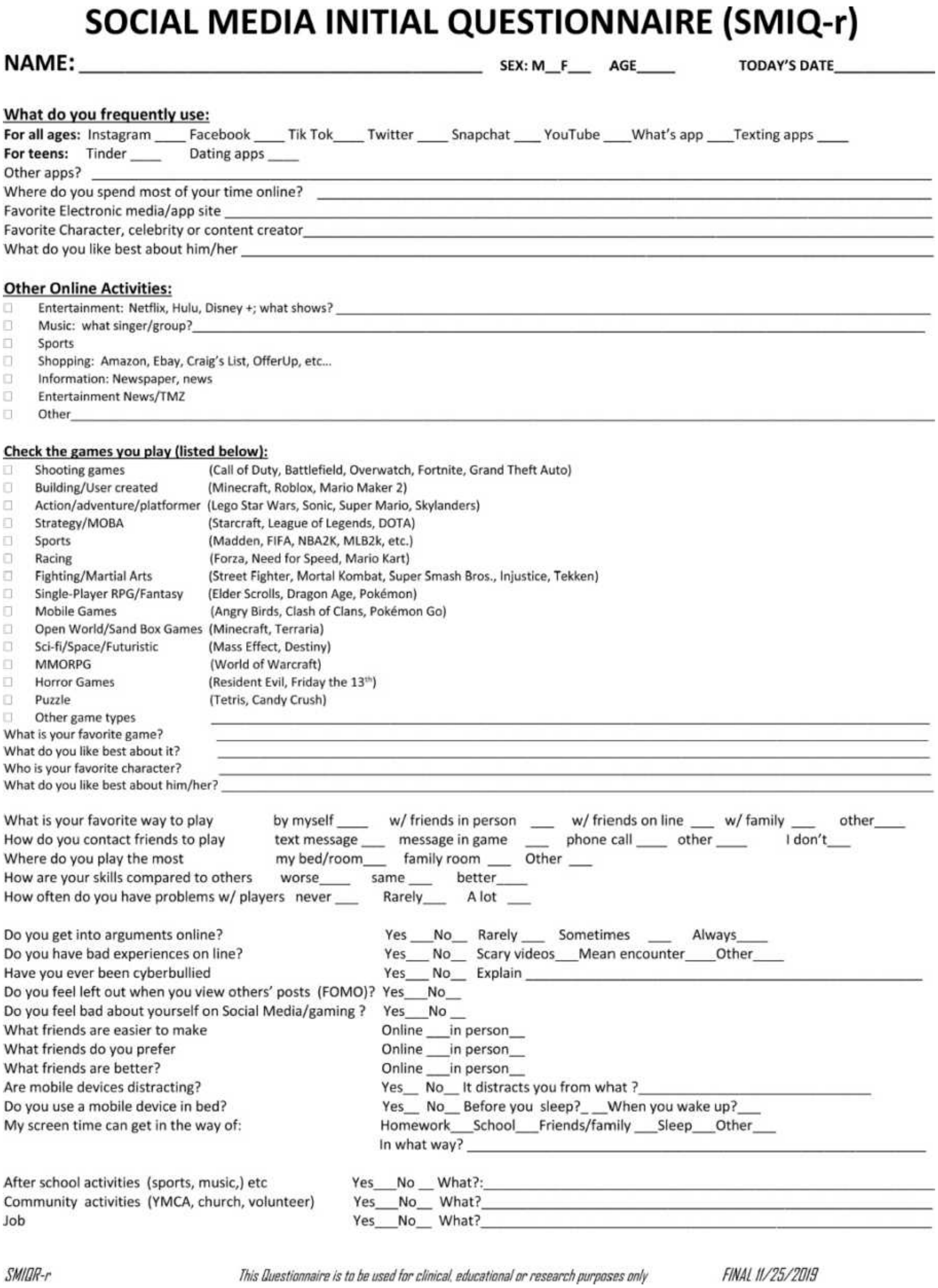

SOCIAL MEDIA INITIAL QUESTIONNAIRE (SMIQ-r)

NAME: ________________________________ **SEX: M__F___ AGE_____ TODAY'S DATE___________**

What do you frequently use:

For all ages: Instagram ____ Facebook ____ Tik Tok____ Twitter ____ Snapchat ____ YouTube ____What's app ____Texting apps ____
For teens: Tinder ____ Dating apps ____
Other apps? ______________________________
Where do you spend most of your time online? ______________________________
Favorite Electronic media/app site ______________________________
Favorite Character, celebrity or content creator______________________________
What do you like best about him/her ______________________________

Other Online Activities:

- ☐ Entertainment: Netflix, Hulu, Disney +; what shows? ______________________________
- ☐ Music: what singer/group?______________________________
- ☐ Sports
- ☐ Shopping: Amazon, Ebay, Craig's List, OfferUp, etc…
- ☐ Information: Newspaper, news
- ☐ Entertainment News/TMZ
- ☐ Other______________________________

Check the games you play (listed below):

☐ Shooting games	(Call of Duty, Battlefield, Overwatch, Fortnite, Grand Theft Auto)
☐ Building/User created	(Minecraft, Roblox, Mario Maker 2)
☐ Action/adventure/platformer	(Lego Star Wars, Sonic, Super Mario, Skylanders)
☐ Strategy/MOBA	(Starcraft, League of Legends, DOTA)
☐ Sports	(Madden, FIFA, NBA2K, MLB2k, etc.)
☐ Racing	(Forza, Need for Speed, Mario Kart)
☐ Fighting/Martial Arts	(Street Fighter, Mortal Kombat, Super Smash Bros., Injustice, Tekken)
☐ Single-Player RPG/Fantasy	(Elder Scrolls, Dragon Age, Pokémon)
☐ Mobile Games	(Angry Birds, Clash of Clans, Pokémon Go)
☐ Open World/Sand Box Games	(Minecraft, Terraria)
☐ Sci-fi/Space/Futuristic	(Mass Effect, Destiny)
☐ MMORPG	(World of Warcraft)
☐ Horror Games	(Resident Evil, Friday the 13th)
☐ Puzzle	(Tetris, Candy Crush)
☐ Other game types	______________________________
What is your favorite game?	______________________________
What do you like best about it?	______________________________
Who is your favorite character?	______________________________
What do you like best about him/her?	______________________________

What is your favorite way to play	by myself ____ w/ friends in person ___ w/ friends on line ___ w/ family ___ other____
How do you contact friends to play	text message ___ message in game ___ phone call ____ other ____ I don't___
Where do you play the most	my bed/room___ family room ___ Other ___
How are your skills compared to others	worse____ same ___ better____
How often do you have problems w/ players	never ___ Rarely___ A lot ___

Do you get into arguments online?	Yes ___No__ Rarely ___ Sometimes ___ Always____
Do you have bad experiences on line?	Yes___ No__ Scary videos___Mean encounter____Other____
Have you ever been cyberbullied	Yes___ No__ Explain ______________________________
Do you feel left out when you view others' posts (FOMO)?	Yes___No__
Do you feel bad about yourself on Social Media/gaming ?	Yes___No __
What friends are easier to make	Online ___in person__
What friends do you prefer	Online ___in person__
What friends are better?	Online ___in person__
Are mobile devices distracting?	Yes__ No__ It distracts you from what ?______________________
Do you use a mobile device in bed?	Yes__ No__ Before you sleep?_ __When you wake up?___
My screen time can get in the way of:	Homework___School___Friends/family ___Sleep___Other___ In what way? ______________________________

After school activities (sports, music,) etc	Yes___No __ What?:______________________________
Community activities (YMCA, church, volunteer)	Yes___No__ What?______________________________
Job	Yes___No__ What?______________________________

SMIQR-r *This Questionnaire is to be used for clinical, educational or research purposes only* *FINAL 11/25/2019*

FIGURE 54.15–1. Social Media Initial Questionnaire (SMIQ-r). (Reprinted with the permission of Michael Feldmeier, MD. https://levelupmentalhealth.com/resources/.)

Table 54.15–1.
Clinician Social Media Guidance for Youth, Parents, and Families

1. **Encourage discussion of safety rules**
    - The most widely used social media sites such as Instagram, TikTok, Facebook, and YouTube, all require account holders to be 13 years or older
    - Review privacy settings on all sites and apps with children to make sure they’re keeping photos and personal information private
    - Utilize parental controls that limit access to inappropriate content
    - Do not give out personal information online
    - Tell a parent or other trusted adult if they or their friends are being bullied or disrespected online
    - For younger teens, do not befriend, chat with someone without a parent’s permission
    - Parents may friend or follow their child’s social media accounts with an agreement between the parent and child
2. **Discuss known risks of social media use in an open, nonjudgmental way**
3. **Encourage caregivers to proactively set boundaries for social media in youth**
    - Limit the amount of time that children and adolescents spend on social media so that it does not interfere with functioning
    - Do not use any electronic devices 1–2 hours before bed, and do not have any smartphones in the bedroom overnight
4. **Remind caregivers that children model social media use on caregiver behavior**
    - Caregivers who follow similar rules with their social media and smartphone use as they expect out of their children will model appropriate use

RESOURCES FOR FAMILIES

The American Academy of Pediatrics provides Family Media Use Plan (healthychildren.org/mediauseplan).

Common Sense Media Family Media Toolkit (https://www.commonsensemedia.org/AAPtoolkit).

FURTHER READING

Alonzo R, Hussain J, Stranges S, Anderson KK. Interplay between social media use, sleep quality, and mental health in youth: a systematic review. *Sleep Med Rev*. 2021;56:101414.

Anderson M, Jiang J. Teens, social media and technology 2018. *Pew Research Center*. 2018. https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/

Auxier B, Anderson M. Social media use in 2021. *Pew Research Center*. 2021. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/

Boer M, van den Eijnden RJJM, Boniel-Nissim M, et al. Adolescents' intense and problematic social media use and their well-being in 29 countries. *J Adolesc Health*. 2020;66(6S):S89–S99.

Cauberghe V, Van Wesenbeeck I, De Jans S, Hudders L, Ponnet K. How adolescents use social media to cope with feelings of loneliness and anxiety during COVID-19 lockdown. *Cyberpsychol Behav Soc Netw*. 2021;24(4):250–257.

Dulli L, Ridgeway K, Packer C, et al. A social media-based support group for youth living with HIV in Nigeria (SMART Connections): randomized controlled trial. *J Med Internet Res*. 2020;22(6):e18343.

Hendricks D. Complete history of social media: then and now. *Small Business Trends*. 2021. https://smallbiztrends.com/2013/05/the-complete-history-of-social-media-infographic.html

Kreski N, Platt J, Rutherford C, et al. Social media use and depressive symptoms among United States adolescents. *J Adolesc Health*. 2021;68(3):572–579.

Olsen K, Beresin EV, Schlozman SC. Media literacy for clinicians and parents. In: Beresin EV, Olsen CK, eds. *Child and Adolescent Psychiatry and the Media*. Elsevier; 2019:75.

Pataki C, Bostic JG, Tabor-Furmark KS, et al. Functional assessment of social media in child and adolescent psychiatry. In: Beresin EV, Olsen CK, eds. *Child and Adolescent Psychiatry and the Media*. Elsevier; 2019:93.

Pew Research Center. April 2021. "Social Media Use in 2021." https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/

Schaeffer K. Most U.S. teens who use cellphones do it to pass time, connect with others, learn new things. *Pew Research Center*. 2019. https://www.pewresearch.org/fact-tank/2019/08/23/most-u-s-teens-who-use-cellphones-do-it-to-pass-time-connect-with-others-learn-new-things/

Social media initial questionnaire-revised. https://levelupmentalhealth.com/resources/

54.16 Gay, Bisexual, Transgender, Queer/Questioning, Intersex, and Asexual (LGBTQIA) Youth

NATALIA RAMOS, M.D., M.P.H., SERENA CHANG, M.D., B.A., AND MIKIKO THELWELL, M.D., M.P.H.

Today's youth are increasingly comfortable with gender and sexuality in the context of greater societal awareness and acceptance of diverse gender identities and sexual orientations. Relatedly, there has been a notable increase in child and adolescent patients seeking gender- and sexuality-related care and support. In order to provide effective evidence-based care to their gender-diverse and sexual minority patients, psychiatrists can increase their knowledge and comfort with key terminology and concepts and learn to implement trauma-informed LGBTQIA-affirming care models, as discussed in this chapter.

In this chapter, we utilize the term *LGBTQIA*, which stands for Lesbian, Gay, Bisexual, Transgender, Queer/Questioning, Intersex, and Asexual, to refer to gender and sexual minority patients except where data refer to specific subgroups therein. Some of these specific terms represent gender identities, while some represent sexual orientations. It is also worth emphasizing that different LGBTQIA patients use different terms to describe themselves and their identities, and the acronym used by the authors encapsulates a limited number of identities. For example, many youths identify with nonbinary, genderfluid, and/or genderqueer identities. A best practice for the clinician is always to ask the youth what term(s) best describe them and what terms they would like the clinician to use.

SEXUAL ORIENTATION

Many Western cultures take a heteronormative approach to socialization—in other words, many youth are taught that heterosexuality is one's expected sexuality. Parental expectations and assignations of gender begin even prior to conception and birth and can shape parenting. Hence, LGBTQIA youth often identify a sense of not belonging, a challenge to personal development during adolescence, when norms and fitting in seem to be the social task at hand. As a result, numerous large national studies have found notable mental health disparities between LGB youth and their heterosexual peers, including increased rates of depression, anxiety, suicidality, and substance use (Kann et al., 2016). It has also been noted that family support and acceptance are strong mitigating factors in reducing these risks for LGB youth.

Table 54.16–1.
Sexual Orientation Terminology

Heterosexual (“Straight”)	**Attracted to the “opposite” sex/gender**
Lesbian	Female-identified attracted to female-identified
Gay	Male-identified attracted to male-identified
Bisexual	Attracted to “both” sexes/genders
Pansexual	Attracted to all sexes/genders
Queer	Denotes identities outside of traditional cisgender heteronormativity. Historically used as a derogatory term but also feels supportive to some.
Allosexual	Having the capacity to be attracted to others in a sexual way
Asexual (“ace”)	Not attracted to anyone in a sexual way
Aromantic (“aro”)	Not attracted to anyone in an emotional/romantic way

An individual’s sexual orientation describes to whom they are physically, romantically, or emotionally attracted (see Table 54.16–1). Sexual orientation is independent of one’s sex or one’s gender. Sexual orientation has three main components:

1. Identity: how one sees themself
2. Attraction: to what gender(s) is one attracted emotionally, romantically, or physically
3. Behavior: in what acts does one engage and with whom

The Diagnostic and Statistical Manual of Mental Disorders (DSM) has a long history of pathologizing LGBTQIA identities, leading to a mistrust of the field of psychiatry and an avoidance of care. Homosexuality was first included in *the Diagnostic and Statistical Manual of Mental Disorders (DSM), First Edition (DSM-I)* (1952) and the original DSM-II (1968) as a “sociopathic personality disturbance.” It was later recategorized as a “sexual orientation disturbance” (as opposed to a mental disorder) in a

1974 edition after years of protests. The diagnosis was changed again to "ego-dystonic homosexuality" in the DSM-III (1980) before being revised yet again to remove the word homosexuality (substituted with "sexual disorder not otherwise specified") in the DSM-III-R (1987).

SEX AND GENDER

Although often conflated with each other, sex and gender are distinct concepts. An individual's *sex* is typically *assigned at birth* based on visualized genitalia and, less often, on other physical characteristics and chromosomes. *Sex assigned at birth* is typically described as female, male, or intersex. Of note, intersex is an umbrella term that describes differences in sexual development (DSD), or a range of traits and conditions in which individuals have naturally occurring variations in chromosomes, gonads, or genitals. Differences in sex development are discussed in greater detail later in this chapter.

An individual's *gender* refers to both one's internal identity and the expression of their gender within their social context (see Table 54.16–2).

Nonbinary Identities

Among Western cultures, much of the population is raised using the gender binary as the social norm. Thus, the concept of gender existing on a continuum or a spectrum can be surprising. In fact, many cultures and societies embrace diverse gender identities outside of Western colonization. In the modern day, many children and adolescents embrace nonbinary identities. In other words, many youths express gender identities that fall outside of the typical "masculine" or "feminine" descriptors. In the United States, several states legally recognize nonbinary identities.

Table 54.16–2.
Gender Terminology

Gender binary	The division of gender into two separate and “opposite” categories (i.e., male and female)
Gender identity	How one thinks about their own gender (i.e., masculine, feminine, both, neither)
Gender expression	The way one dresses, behaves, and “expresses” oneself
Gender roles	The norms that one’s culture creates around male and female identities
Cisgender	One’s gender identity coincides with the sex assigned at birth
Transgender	One’s gender identity differs from their sex assigned at birth. Umbrella term that can encompass other descriptors such as agender, bigender, gender expansive, gender diverse, gender nonbinary, genderqueer, and genderfluid.
Nonbinary	Describes one’s identity as neither fully male nor female.
Transition	The process of changing one’s gender expression to match one’s gender identity.

Gender Dysphoria

Similar to its approach to pathologizing homosexuality, the DSM’s approach to transgender identities has been stigmatizing and marginalizing to many. The DSM-III (1980) listed “transsexualism” for the first time. This term was replaced by “gender identity disorder,” categorized under Sexual Dysfunctions and Paraphilic Disorders, in the DSM-IV (1994). In the DSM-5 (2013), “gender identity disorder” was replaced by “gender dysphoria” (APA, 2023).

The Diagnostic and Statistical Manual of Mental Disorders, fifth edition, text revision (DSM-5-TR) (2022) defines gender dysphoria in children as “a marked incongruence between one’s experienced/expressed gender and assigned gender, lasting at least 6 months, as manifested by at least six of the following [criteria]” (see Table 54.16–3). The first criterion, “a strong desire to be of the other gender or an insistence that one is the other gender," is required for the diagnosis. Note that the criteria for gender dysphoria in adults and adolescents varies slightly. Similar to the diagnosis in adolescents and adults, "the condition is associated with clinically significant distress or impairment in social, school, or other important areas of functioning."

Treatment for gender dysphoria consists of gender-affirming care, which may include therapy or psychiatric treatment for gender-linked distress and co-occurring conditions. Youth and adults with gender dysphoria experience higher rates of co-occurring anxiety, depression, suicidality, trauma, and substance use disorders. Research has not shown correlations between gender dysphoria and rates of schizophrenia or bipolar disorder diagnoses. Recent research also points to a strong correlation between autism spectrum disorder (ASD) incidence among gender dysphoric youth.

Gender Transitions

Gender diversity in children and adolescents has become highly politicized in the United States. Anti-LGBTQIA legislation and lawsuits have focused particularly on bathroom access, youth sports team participation, and access to gender-affirming health care. However, studies show that transgender youth who are supported in their identities experience lower rates of negative mental health outcomes than those who are not. In the TransYouth Study, investigators found that prepubescent youth who socially transitioned—meaning they lived as their experienced gender, with parental support—demonstrated the same rates of depression as matched cisgender peers. Accordingly, many mainstream medical organizations, including the American Psychiatric Association, the American Psychological Association, the American Academy of Child and Adolescent Psychiatry, the American College of Gynecology, and the Endocrine Society, have published statements and guidelines supporting gender-affirming care.

Table 54.16–3.
DSM-5 Criteria for Gender Dysphoria in Children

The DSM-5-TR defines gender dysphoria in children as a marked incongruence between one's experienced/expressed gender and assigned gender, lasting at least 6 months, as manifested by at least six of the following (one of which must be the first criterion):

- A strong desire to be of the other gender or an insistence that one is the other gender (or some alternative gender different from one's assigned gender)
- In boys (assigned gender), a strong preference for cross-dressing or simulating female attire; or in girls (assigned gender), a strong preference for wearing only typical masculine clothing and a strong resistance to the wearing of typical feminine clothing
- A strong preference for cross-gender roles in make-believe play or fantasy play
- A strong preference for the toys, games, or activities stereotypically used or engaged in by the other gender
- A strong preference for playmates of the other gender
- In boys (assigned gender), a strong rejection of typically masculine toys, games, and activities and a strong avoidance of rough-and-tumble play; or in girls (assigned gender), a strong rejection of typically feminine toys, games, and activities
- A strong dislike of one's sexual anatomy
- A strong desire for the physical sex characteristics that match one's experienced gender

Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision, DSM-V-TR, p. 512 (Copyright © 2022). American Psychiatric Association. All Rights Reserved.

Many, and not all, youth with gender dysphoria will wish to transition, which is the process of changing one's gender expression to match one's gender identity. Transitioning *may* include various social, legal, and medical changes (see Table 54.16–4).

Table 54.16–4.
Gender Transitions

Social	Changes in name and pronouns, appearance, dress, grooming (e.g., clothes, hair, make-up); nonmedical changes; may include “coming out” to family, school, peers, or others
Legal	Legal name or gender marker change, resulting in the reissuing of government documents (e.g., birth certificate, driver’s license, passport)
Medical	Gender-affirming procedures with medical doctors; may include: puberty blockers (i.e., Gonadotropin-releasing hormone agonists, like leuprolide), hormone therapies (e.g., estradiol or testosterone), testosterone blockers (e.g., spironolactone), and gender-affirming surgical procedures such as: ▸ For transmasculine/nonbinary patients assigned female at birth: chest masculinization (bilateral mastectomy with chest reconstruction), hysterectomy, phalloplasty, metoidioplasty ▸ For transfeminine/nonbinary patients assigned male at birth: breast augmentation, orchiectomy, vaginoplasty, electrolysis, tracheal shave, facial feminization surgery

There is no single protocol for a gender transition; every gender-diverse individual has unique needs and goals. For example, some people opt solely for social or legal transition. Others may wish to engage in any number of medical transition options with their doctors. Of note, many gender-diverse individuals face legal, social, and financial barriers that prevent access to medical-affirming care. Furthermore, the above medical procedures have developmental prerequisites and may not be readily accessible to minors. The Standards of Care for the Health of Transgender and Gender Diverse People, Version 8 (SOC8) published by the World Professional Association for Transgender Health (WPATH) provides specific guidance around medical transitions and the treatment of gender dysphoria (see https://www.wpath.org/publications/soc for details).

Mental health clinicians have historically been pushed into a “gatekeeper” role in the medical transition process, as standards of care often require patients to pursue psychological evaluation and obtain surgical readiness evaluation letters from psychiatrists or therapists. Insurance carriers or surgeons mandate these evaluations in some cases. This process can create a complicated relationship between mental health clinicians and their gender diverse patients. For example, patients may feel pressure to deny mental health symptoms or pursue care to be accepted or “approved” for gender-affirming medical care. Like many individuals who experience social stressors, gender-diverse individuals may benefit from therapeutic or psychiatric treatment. The best clinical practices allow the clinician to support the youth and family before, during, and after the transition process. Clinicians can help youth target mental health symptoms, manage stressors, seek support, and better understand their goals while providing developmental guidance and skills to caregivers to improve youth well-being.

Differences of Sex Development (DSD) and Intersex Youth

While the literature on the specific mental health needs of intersex youth remains limited, the first national survey of intersex adults in the United States revealed a high prevalence of lifetime anxiety disorders, post-traumatic stress disorder (PTSD), and depressive disorders. Intersex youth face unique stressors and related mental health symptoms across developmental stages due to the medicalization and pathologization of their bodies from a young age. Some intersex traits present clearly at birth, such as those involving external genitalia, while others manifest during puberty.

Intersex youth experience intense stressors with lifelong consequences, including medically unnecessary surgeries, a lack of support around gender identity development, and the loss of autonomy and agency. Parents, meanwhile, may contend with a sense of urgency around sex assignment, including pressure to consent to genital surgeries without understanding the potential long-term effects on their child and regret about irreversible decisions. Often the justification for an early surgical procedure is to improve long-term mental health by helping the child live a “normal” life—a justification made without evidence and typically by clinicians who do not specialize in mental health. Additionally, early medical interventions reinforce sociocultural norms around male and female gender stereotypes, often before the development of the youth's own gender identity.

Overall, there is a great need for increased mental health research, programs, and developmentally appropriate guidelines in caring for youth with DSDs. Mental health professionals can support youth and families in gender identity exploration, navigating challenging health care decisions, and recovery from traumatic stress. Additionally, new parents may experience intense and conflicting emotions in deciding on their child's health care plans. Child and adolescent clinicians can also advocate using psychologically sound evidence to determine care for intersex youth at all developmental stages and seek inclusion in pediatric teams who manage DSDs. Finally, they can assist youth and families in identifying community resources and supports that promote mental well-being.

YOUTH TRAUMA AND RESILIENCE

LGBTQIA youth are more likely than cisgender heterosexual youth to experience traumatic stress due to discrimination, harassment, marginalization, and victimization across social systems, including family, school, and health care settings. Examples of common social stressors include family rejection and neglect, harsh discipline and bullying in schools, and anti-LGBTQIA legislation. Youth who experience multiple forms of psychosocial stress are especially susceptible to trauma sequelae across their lifespan. The stage of development during which the stressors occur may also influence their long-term impact. Youth with chronic or pervasive social stressors, particularly unsupportive, rejecting, or dangerous home environments, are at a higher risk for unhealthy behaviors. Consequently, they experience higher rates of the leading causes of morbidity and mortality in adulthood, including ischemic heart disease, cancer, lung disease, and liver disease, than peers without adverse childhood experiences.

Mental Health Inequities Among LGBTQIA Youth

Despite an increase in the number of youths identifying as LGBTQIA, data continue to demonstrate clear inequities in mental health outcomes among LGBTQIA youth compared to their cisgender heterosexual peers. LGBTQIA youth report disproportionate rates of depression, anxiety, suicidal ideation, and suicide attempts. Nearly two-thirds of LGBTQIA youth report symptoms of generalized anxiety disorder, and more than half report symptoms of major depressive disorder. For many LGBTQIA youth, mental health challenges emerge earlier in the life course and persist over time.

Clinicians can understand these disproportionate mental health outcomes by examining the interlocking systems of marginalization and discrimination that compound to produce a variety of unique mental health challenges for LGBTQIA youth living in cis-sexist heterosexist social environments. Intersectional and minority stress approaches, which have strong empirical support in LGBTQIA health disparities research, explain how structural inequities like a lack of access to stable housing and interpersonal anti-LGBTQIA discrimination lead to increased rates of mental health symptoms among LGBTQIA youth. In addition to the negative outcomes stemming from the external stressors themselves, experiences of anti-LGBTQIA stereotypes and messaging influence internalized psychological processes, such as the perception of the external world as threatening, concealment of identity, and poor self-image. Clinically, providers may observe symptoms related to anxiety, depression, and somatization, in addition to more externalizing symptoms like impulsivity and substance use.

CLINICAL APPROACHES FOR LGBTQIA YOUTH AND THEIR FAMILIES

Intersectional Care for LGBTQIA Youth

In order to provide effective care, clinicians must thoughtfully explore and assess each youth’s unique experiences of stigmatization and support, as LGBTQIA youth do not share a universal experience by any means. For example, asexual adolescents report significantly higher rates of internalized LGBTQIA-phobia than adolescents with other sexual minority identities while experiencing less interpersonal prejudice and engaging in fewer risky behaviors. Notably, LGBTQIA youth of color experience profound inequities across school, juvenile justice, and social welfare settings due to compounding marginalization based on both race/ethnicity *and* gender identity or sexual orientation. Gender-diverse youth of color report more school victimization based on gender identity *and* race/ethnicity than cisgender peers of color, while multiracial Black

LGBTQIA students experience more victimization than LGBTQIA students who identify only as Black.

Clinicians must thoughtfully examine connections between observed symptoms and identity-linked social stressors. Like cisgender heterosexual youth, LGBTQIA youth experience an array of mental health challenges, which can be additionally compounded by their experiences of marginalization and social rejection. Existing evidence-based treatments can be tailored to a youth's unique stressors and strengths. For example, cognitive behavioral therapies (CBT) can address LGBTQIA identity-related stressors and promote adaptive coping strategies. Modified CBT has been shown to reduce depression symptoms, alcohol use, and human immunodeficiency virus (HIV) risk behaviors among gay and bisexual young men. Trauma-focused cognitive-behavioral therapy (TF-CBT) has been formally adapted to LGBTQIA youth and their families. Additional interventions, such as the Family Acceptance Project, target caregivers' accepting and rejecting behaviors and family caregiver–child communication to improve outcomes among LGBTQIA youth and can be adapted to diverse families.

Addressing Risk Factors and Building Resilience

In addition to experiencing more social stressors, LGBTQIA youth are also less likely than cisgender heterosexual peers to have access to key protective factors that mitigate traumatic stress, including family support, stable housing, and social support networks. In working with LGBTQIA youth, clinicians can focus on family, school, and community support that strongly impact the psychosocial well-being of LGBTQIA youth.

Family support is a key factor in the mental well-being of LGBTQIA youth, while family rejecting behaviors detrimentally impact youth mental health. Accepting and supportive parental behaviors related to gender identity and sexual orientation reduces the risk of depression and risk behaviors among LGBTQIA youth, buffering the effects of identity-related stressors. Among prepubescent transgender youth, those who socially transition with parental support demonstrate rates of depression and anxiety comparable to cisgender peers. In working with families, the best practice is to align with caregivers around the shared goal of supporting their child's wellbeing rather than challenging deeply held beliefs.

Experiences of acceptance and support in the school setting also greatly impact LGBTQIA youths' mental health, as do bullying and discrimination. School connectedness and school safety are associated with better mental health and decreased risk behaviors. Unfortunately, the vast majority of LGBTQIA youth experience verbal harassment and discrimination in school settings. Clinicians can help youth recover from traumatic stress, manage co-occurring conditions, build coping strategies, and locate safe spaces and supports. Caregivers can also advocate in the school setting for inclusion policies and support for their youth and help youth connect to affirming activities in their community.

Table 54.16–5.
Key Trauma-Informed LGBTQIA-Affirming Practices in Clinical Settings

- Collect accurate data on sexual orientation and gender identity, including names and pronouns used
- Use the youth’s name and pronouns consistently
- Screen for trauma symptoms and experiences of abuse, discrimination, marginalization, and bullying
- Utilize open-ended questions to discuss sexual orientation and gender identity/expression
- Ensure confidentiality and privacy pertaining to identity
- Engage the youth in collaborative decision-making when feasible
- Avoid gender stereotypes and comments based on appearance
- Make the clinical care space more supportive (e.g., access to gender-inclusive bathrooms, open-ended intake forms)
- Refer to Standards of Care and Continuing Education to stay up-to-date with LGBTQIA care practices

Trauma-Informed Affirming Care for LGBTQIA Youth

By implementing trauma-informed affirming care, clinicians can play key roles in recognizing and targeting psychological symptoms that stem from social stressors and marginalization experienced by LGBTQIA youth. Research consistently shows that social support and positive coping strategies can mitigate stress experienced by LGBTQIA youth. Clinicians who practice from a trauma-informed perspective recognize and support the youth’s strengths and contextualize symptoms and challenges stemming from lived experiences. Additionally, in practicing affirming care, clinicians fundamentally presume that gender and sexual orientation: (1)

exist on a natural spectrum where no gender identity or expression or sexual orientation is pathologic; (2) can be fixed or fluid and, in regards to gender, binary (male/female) or nonbinary; and (3) may evolve or change over the life course for some individuals. LGBTQIA-affirming care also recognizes that gender and sexual orientation vary across cultures and throughout history. Clinicians can follow specific practices to deliver trauma-informed affirming care to LGBTQIA youth (Table 54.16–5).

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Suegee Tamar-Mattis, DO, for graciously offering to consult and review our content related to caring for intersex youth and youth with DSD.

FURTHER READINGS

American Psychiatric Association. *Diagnostic And Statistical Manual Of Mental Disorders.* 5th ed. American Psychiatric Publishing; 2013.

American Psychiatric Association. (2022). *Gender Dysphoria*. https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/DSM-5-TR/APA-DSM5TR-GenderDysphoria.pdf

American Psychiatric Association. (2023). *Gender Dysphoria Diagnosis*. https://www.psychiatry.org/psychiatrists/diversity/education/transgender-and-gender-nonconforming-patients/gender-dysphoria-diagnosis

American Psychological Association. Guidelines for psychological practice with transgender and gender nonconforming people. *Am Psychol*. 2015;70(9):832–864.

Becerra-Culqui TA, Liu Y, Nash R, et al. Mental health of transgender and gender nonconforming youth compared with their peers. *Pediatrics*. 2018;141(5):e20173845.

Bockting WO, Miner MH, Romine RES, Hamilton A, Coleman E. Stigma, mental health, and resilience in an online sample of the US transgender population. *Am J Public Health*. 2013;103(5):943–951.

Callens N, Kreukels BPC, van de Grift TC. Young voices: sexual health and transition care needs in adolescents with intersex/differences of sex development—a pilot study. *J Pediatr Adolesc Gynecol*. 2021;34(2):176–189.e2.

Centers for Disease Control and Prevention (2023). *The Youth Risk Behavior Survey Data Summary & Trends Report: 2011–2021*. U.S. Department of Health and Human Services. https://www.cdc.gov/healthyyouth/data/yrbs/yrbs_data_summary_and_trends.htm

Cohen JA, Mannarino AP, Wilson K, et al. *Trauma-Focused Cognitive Behavioral Therapy LGBTQ Implementation Manual*. Allegheny Health Network; 2018.

Coleman E, Radix AE, Bouman WP, et al. Standards of care for the health of transgender diverse people, version 8. *Int J Trangend Health*. 2022;23(Suppl 1):S1–S259.

de Vries ALC, Noens ILJ, Cohen-Kettenis PT, van Berckelaer-Onnes IA, Doreleijers TA. Autism spectrum disorders in gender dysphoric children and adolescents. *J Autism Dev Disord*. 2010;40(8):930–936.

Drescher J. Out of DSM: depathologizing homosexuality. *Behav Sci (Basel)*. 2015;5(4):565–575.

Eisenberg ME, Resnick MD. Suicidality among gay, lesbian and bisexual youth: the role of protective factors. *J Adolesc Health*. 2006;39(5):662–668.

Ethier KA, Harper CR, Dittus PJ. School environment is related to lower health and safety risks among sexual minority middle and high school students. *J Adolesc Health*. 2018;62:143–148.

Felitti VJ, Anda RF, Nordenberg D, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. *Am J Prev Med*. 1998;14(4):245–258.

Grant JM, Mottet LA, Ranis J. National transgender discrimination survey report health and health care: findings of a study by the national center for transgender equality and the national gay and lesbian task force. 2011.

Institute of Medicine (US) Committee on Lesbian, Gay, Bisexual, and Transgender Health Issues and Research Gaps and Opportunities. *The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding*. National Academies Press (US); 2011.

Kann L, Olsen EO, McManus T, et al. Sexual identity, sex of sexual contacts, and health-related behaviors among students in grades 9–12—United States and selected sites, 2015. *MMWR Surveill Summ*. 2016;65(SS-9):1–202.

Keo-Meier C, Ehrensaft D, eds. *The Gender Affirmative Model: An Interdisciplinary Approach to Supporting Transgender and Gender Expansive Children*. American Psychological Association; 2018.

Kidd KM, Sequeira GM, Douglas C, et al. Prevalence of gender-diverse youth in an urban school district. *Pediatrics*. 2021;147(6):e2020049823.

Kosciw JG, Clark CM, Truong NL, et al. *The 2019 National School Climate Survey: The Experiences Of Lesbian, Gay, Bisexual And Transgender Youth In Our Nation's Schools*. GLSEN; 2020.

Lothwell LE, Libby N, Adelson SL. Mental health care for LGBT youths. *Focus*. 2020;18(3):268–276.

McEwen BS. Neurobiological and systemic effects of chronic stress. *Chronic Stress (Thousand Oaks)*. 2017;1:2470547017692328.

McInroy LB, Beaujolais B, Leung VWY, et al. Comparing asexual and non-asexual sexual minority adolescents and young adults: stressors, suicidality and mental and behavioural health risk outcomes. *Psychol Sex*. 2020:1–17.

Meyer IH. Prejudice as stress: conceptual and measurement problems. *Am J Public Health*. 2003;93(2):262–265.

Olson KR, Durwood L, DeMeules M, McLaughlin KA. Mental health of transgender children who are supported in their identities. *Pediatrics*. 2016;137(3):e20153223.

Ramos N, Barnert E. Bath E. Addressing the mental health needs of LGBTQ youth in the juvenile justice system. *J Am Acad Child Adolesc Psychiatry*. 2021;S0890-8567(21):00419-6.

Rhoades H, Rusow JA, Bond D, Lanteigne A, Fulginiti A, Goldbach JT. Homelessness, mental health and suicidality among LGBTQ youth accessing crisis services. *Child Psychiatry Hum Dev*. 2018;49(4):643–651.

Romero AP, Goldberg SK, Vasquez LA. *LGBT People and Housing Affordability, Discrimination, and Homelessness*. The Williams Institute; 2020. https://williamsinstitute.law.ucla.edu/publications/lgbt-housing-instability

Rosenwohl-Mack A, Tamar-Mattis S, Baratz AB, et al. A national study on the physical and mental health of intersex adults in the U.S. *PLoS One*. 2020;15(10):e0240088.

Ryan C, Diaz R. *Family Acceptance Project: Intervention Guidelines and Strategies*. Family Acceptance Project; 2011.

Ryan C, Russell ST, Huebner DM, Diaz R, Sanchez J. Family acceptance in adolescence and the health of LGBT young adults. *J Child Adolesc Psychiatr Nurs*. 2010;23(4):205–213.

Strang JF, Meagher H, Kenworthy L, et al. Initial clinical guidelines for co-occurring autism spectrum disorder and gender dysphoria or incongruence in adolescents. *J Clin Child Adolesc Psychol*. 2018;47(1):105–115.

Strang JF, van der Miesen AIR, Fischbach AL, Wolff M, Harris MC, Klomp SE. Common intersection of autism and gender diversity in youth: clinical perspectives and practices. *Child Adolesc Psychiatr Clin N Am*. 2023;32(4):747–760.

Tamar-Mattis A, Baratz A, Dalke KB, et al. Emotionally and cognitively informed consent for clinical care for differences of sex development. *Psychol Sex*. 2014;5(1):44–55.

Terada S. Suicidal ideation among patients with gender identity disorder. *J Psychiatr Res*. 2011;190(1):159–162.

The Trevor Project. National Survey on LGBTQ Youth Mental Health. 2021. https://www.TheTrevorProject.org/survey-2021

Truong NL, Zongrone AD, Kosciw JG. *Erasure and Resilience: The Experiences of LGBTQ Students of Color, Black LGBTQ Youth in US Schools: A Report from GLSEN and The National Black Justice Coalition*. GLSEN; 2020.

54.17 Racism and Discrimination's Impact on the Health and Well-Being of Developing Children and Adolescents

Wanjikũ F.M. Njoroge, M.D., Ewurama Sackey, M.D., Nana Asabere, M.D., and Eraka Bath, M.D.

Researchers have always been interested in understanding the way the social construction of race has impacted health. While several well-known and foundational researchers were investigating the impact of "race" on the health and well-being of children, adolescents, and adults, it was not until 2002 that the Institute of Medicine (IOM)'s Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care produced a seminal document titled "Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care." This report definitively highlighted the decades-long impact that unequal health care access and treatment had on the health, lives, and well-being of generations of people of color (BIPoC) in the United States. Specifically, the report highlighted the ways in which communities of color have been historically marginalized by health care providers and the health care system or, worse, brutalized with experimentation (e.g., Sims on Black women of reproductive age, the Tuskegee experiment on Black men, women, and children in Alabama, forced sterilization of Native American and Black women, and even more recently immigrant women in 2020). The IOM's 2002 nuanced report not only outlined the harm done to communities of color but also gave specific recommendations on how to mitigate the injustices that had occurred in and by the health care system and practitioners. However, in spite of the cogent and clearly outlined recommendations, these systems, structures, and institutions continue to operate as created, with the resultant disparate health care outcomes still evident in every aspect of medicine, including psychiatry.

Around the same period, Dr. Camara Jones, former president of the American Public Health Association (APHA), launched a National Campaign Against Racism, one of the largest professional organizations to do so. While at the Center for Disease Control and Prevention (CDC), she led the development and integration of a standardized assessment tool on racism within the Behavioral Risk Factor Surveillance System. Almost two decades later, during the midst of the global COVID-19 pandemic and after the murder of George Floyd launched uprisings around the world for racial justice, the CDC officially recognized the ongoing impact that racism has had on the health of historically marginalized BIPoC in the United States in that race and ethnicity *still* remain significant predictors of access, quality, and type of care that patients receive. Secondary to these pervasive health inequities, in the spring of 2021, the CDC director declared in the midst of the global COVID-19 pandemic, "Instead, the pandemic illuminated inequities that have existed for generations and revealed for all of America a known, but often unaddressed, epidemic impacting public health: racism. What we know is this: racism is a serious public health threat that directly affects the well-being of millions of Americans. As a result, it affects the health of our entire nation. Racism is not just the discrimination against one group based on the color of their skin or their race or ethnicity, but the structural barriers that impact racial and ethnic groups differently to influence where a person lives, where they work, where their children play, and where they worship and gather in community. These social determinants of health have life-long negative effects on the mental and physical health of individuals in communities of color."

DEFINING TERMINOLOGY

As we begin this discussion, it is important to understand the historical significance of the classification of peoples into "races." While this particular terminology has been used in medicine specifically to categorize the risk of disease, surveillance methods, prognostic predictions, and even specific treatment modalities, questions have long existed about the scientific validity of this type of framing. It is important to recognize the myth of race(s) and recognize it as a social construct developed to justify and explain structural and interpersonal/personally mediated racism that has no scientific merit. Carl Linnaeus's 1735 taxonomy of humans created the underpinnings of scientific racism that persist to this day, reflecting Eurocentric classifications of groups by the myth of "race." Because of the longstanding use of "race" as a proxy to attribute differences in health and well-being, health disparities have long been framed using race as the construct rather than racism. "Race" has been confounded with social determinants, such that phenotypic appearance has been designated as a key factor in illness and health rather than understanding how the context in which people live, work, and play influences their experiences and their health and well-being, which are in fact not determined by "biologic race" but rather secondary to structural systems and barriers put into place by the social construct of race. But more simply, these systems and structures have determined access and quality of care and, ultimately, the health of people rather than personal choices and decisions.

This chapter specifically interrogates the ways in which race, ethnicity, racism, and discrimination have impacted children and adolescents' developmental trajectories, behavioral health, and well-being. We highlight ways to prevent and disrupt the various forms of racism and discrimination embedded in the health care system, other systems in which children find themselves, as well as the practice of medicine, specifically child and adolescent psychiatry, thus ensuring all children and adolescents receive equitable psychiatric assessment, treatment, and care. However, any thorough discussion of race, ethnicity, racism, and discrimination must first begin with diagnosing the problem and this requires critical examination of the language and nomenclature that have been historically used to buttress and support different ideologies and beliefs. Thus, to ensure shared understanding throughout the chapter a definition of terms that are frequently used when discussing health equity will first be discussed. Throughout the chapter, Black has been used to describe African American/African peoples from the African diaspora, Latinx has been used to describe Latino/Latina/Latine/Hispanic peoples, Asian American and Pacific Islander (AAPI) has been used to describe Asian American/Asian/Native Hawai'ian/Pacific Islanders, and Native American has been used to describe Indigenous, First Nations, Alaskan Natives, Native tribes—federally recognized, and not.

An important undergird of the American health system, and thus the practice of psychiatry in the United States is the existence of systemic, structural, or institutional racism. All of these terms have been used to explain the systems, structures, and institutions that were put into place during and after slavery, particularly during the Jim Crow era to ensure an inequitable and unjust system in the United States. With respect to psychiatry, this has resulted in not only disparate treatment for Black, Indigenous, and BIPoC, but also defined the ways in which we diagnose psychiatric illnesses which cascades to potentiate inequitable treatment and care. An understanding of the multiple forms of racism is important in being able to both dismantle and correctly diagnose the problems. To that end, it is helpful to understand that most theorists divide racism into multiple buckets including larger systems or structures down to the more nuanced and personalized forms of racism that we see operating between and within individuals (see Table 54.17–1). Understanding these differing forms of racism as operating at all times is also important when understanding the ways in which racism impacts the physical and mental health of individuals, particularly children and youth. When discussing systemic or structural racism, the link to the deliberate construction of neighborhoods, schools, and other foundational systems is incredibly important in understanding how these systems continue to function and negatively impact the lives of BIPoC Americans. While the government purposefully constructed these inequitable systems, they have been maintained and reinforced via other actors. The result has impacted the health of generations of children and families of color.

Table 54.17–1.
Levels of Racism

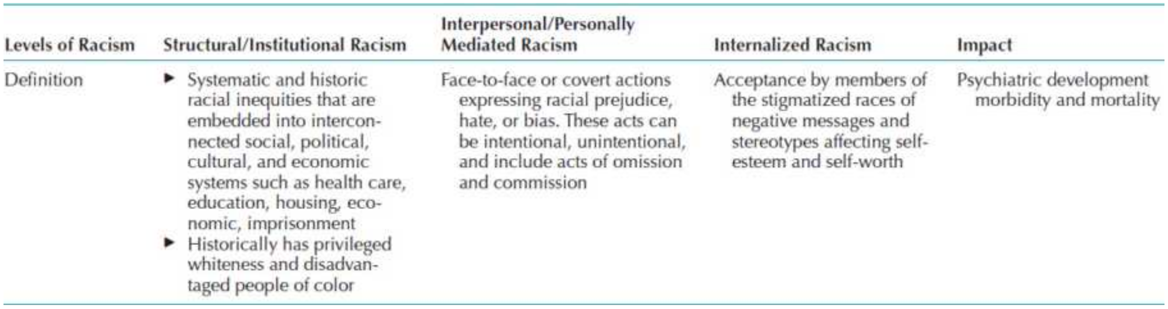

Levels of Racism	Structural/Institutional Racism	Interpersonal/Personally Mediated Racism	Internalized Racism	Impact
Definition	▸ Systematic and historic racial inequities that are embedded into interconnected social, political, cultural, and economic systems such as health care, education, housing, economic, imprisonment ▸ Historically has privileged whiteness and disadvantaged people of color	Face-to-face or covert actions expressing racial prejudice, hate, or bias. These acts can be intentional, unintentional, and include acts of omission and commission	Acceptance by members of the stigmatized races of negative messages and stereotypes affecting self-esteem and self-worth	Psychiatric development morbidity and mortality

In 2019, the American Academy of Pediatrics (AAP) published a policy statement which at the time was revolutionary simply stating, “Racism has a profound impact on the health status of children, adolescents, emerging adults, and their families.” In this chapter, we will focus not only on systemic, structural, and institutional forms of racism and the impacts on Black, Indigenous, and children of color, but we will also highlight how other forms of racism including interpersonal/personally mediated and internalized forms of racism continue to have deleterious impacts on developing children and adolescents. Understanding the multiplicative ways in which developing children differ in their cognitive, regulatory, and emotional developmental trajectories, the impacts of racism on children will be presented along specific developmental ages and stages: early childhood, preschool age, school age, adolescence, and transitional age youth. This chapter will nest the experiences of these developing children and adolescents within the larger supports and systems in which they live and grow including their parents and primary caregivers, schools and childcare facilities, neighborhoods, and communities (see Fig. 54.17–1). Finally, this chapter will speak to the role of the child and adolescent psychiatrist as well as future research, policy, and clinical directions, highlighting culturally appropriate prevention and intervention models to ensure equitable care for all children and youth with behavioral health challenges.

CONTEXTUALIZING RACE-BASED TRAUMATIC STRESS AS AN ADVERSE CHILDHOOD EXPERIENCE

Racial Stress and Trauma

In 2009, the American Psychiatric Association (APA) Presidential Task Force on Traumatic Stress Disorder and Trauma in Children and Adolescents first recognized discrimination as a contributing factor to stress in the lives of youth of color. *Racial Trauma* is defined as "dangerous or frightening race-based events, stressors, or discrimination." Most of the time racial trauma is not a major, isolated episode but results from an accumulation of sometimes daily, covert acts of racial discrimination. Children and adolescents may experience direct racial trauma through structural inequities (neighborhood environments, lower quality schools, health care, family resources, and the like) or via interpersonal/personally mediated racist experiences. Racism and discrimination have been long considered not only a social determinant of health but also increasingly considered as an adverse childhood experience (ACE) initially described by Felliti and colleagues identifying certain experiences that occurred in childhood as having direct impact on adult health. Following this seminal work, the impact of racism and discrimination is being acknowledged on health surveys designed to identify ACEs; an example would be the National Survey for Children's Health [NSCH], Philadelphia ACE Survey. For example, due to racism in the adult carceral system, a significant proportion of Black and Latinx youth have parents who are incarcerated. Youth with incarcerated parents are at increased risk of poverty, criminal activity, drug use, and emotional distress.

In addition, youth of color are also vulnerable to direct, indirect, or vicarious racialized trauma such as personally experiencing police harassment or witnessing police killings of unarmed Black or Latinx citizens. Resulting symptoms of racial stress and trauma include anxiety, depression, hopelessness, and cognitive distortions about self-esteem and safety. Racial trauma has a detrimental effect on the physiologic, emotional, social, and psychological development of BIPoC children and adolescents. Ongoing racial discrimination and incidents of racial trauma can lead to a range of behavioral health disorders, including post-traumatic symptoms such as hypervigilance and avoidance, and can lead to the development of a post-traumatic stress disorder (PTSD) diagnosis, as well as generalized anxiety disorder (GAD), and/or major depressive disorder (MDD) diagnoses.

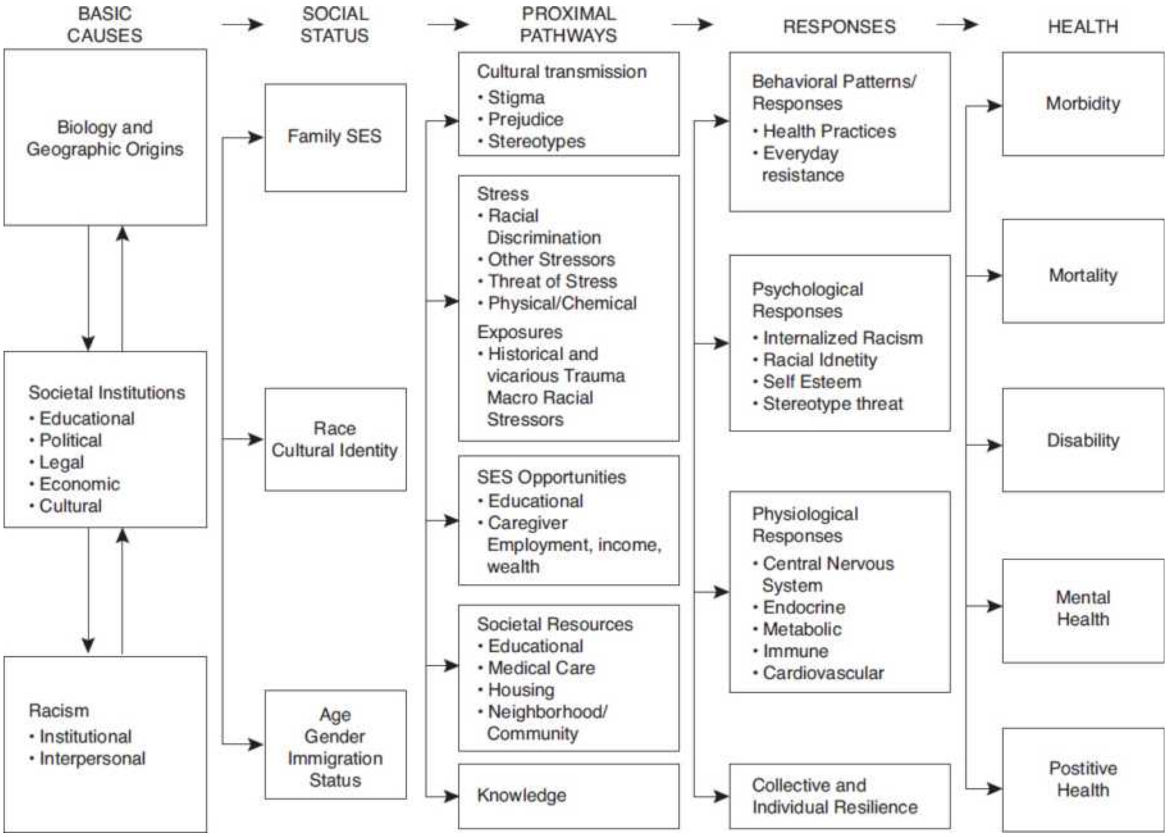
BASIC CAUSES
SOCIAL STATUS
PROXIMAL PATHWAYS
RESPONSES
HEALTH
Biology and Geographic Origins
Societal Institutions
• Educational
• Political
• Legal
• Economic
• Cultural
Racism
• Institutional
• Interpersonal
Family SES
Race Cultural Identity
Age Gender Immigration Status
Cultural transmission
• Stigma
• Prejudice
• Stereotypes
Stress
• Racial Discrimination
• Other Stressors
• Threat of Stress
• Physical/Chemical
Exposures
• Historical and vicarious Trauma Macro Racial Stressors
SES Opportunities
• Educational
• Caregiver Employment, income, wealth
Societal Resources
• Educational
• Medical Care
• Housing
• Neighborhood/ Community
Knowledge
Behavioral Patterns/ Responses
• Health Practices
• Everyday resistance
Psychological Responses
• Internalized Racism
• Racial Idnetity
• Self Esteem
• Stereotype threat
Physiological Responses
• Central Nervous System
• Endocrine
• Metabolic
• Immune
• Cardiovascular
Collective and Individual Resilience
Morbidity
Mortality
Disability
Mental Health
Postitive Health

FIGURE 54.17–1. Pediatric framework for the study of racism and health. (Adapted with permission from Williams DR, Mohammed SA. Racism and Health 1: pathways and scientific evidence. *Am Behav Sci*. 2013;57(8):1152–1173. doi: 10.1177/0002764213487340.)

EMERGING RESEARCH/NEUROBIOLOGY/NEUROPLASTICITY

The health and development of children and adolescents is shaped by a dynamic process of adaptation (at the molecular, cellular, organ system, and behavioral levels) that begins at conception and extends throughout the life course. Importantly, the underlying plasticity that influences that adaptive capacity (known to decrease with age) can lead to a wide range of outcomes that are influenced by the interactions among individual genetic predispositions and the environments in which pre- and postnatal development unfolds. Crucially, as children’s brains develop, the cortical lining thins. Cortical thinning occurs in a typical pattern that begins in primary sensory cortices, then extends to the motor cortices, and finally ends in the association cortices occurring late in adolescence or in early adulthood/transitional age youth. Importantly, cortical thinning reflects myelination, and myelination reduces plasticity, thus early cortical thinning may shorten windows of developmental plasticity. Cortical thinning has also been linked to synaptic pruning, which has the potential to limit the plasticity at later developmental windows. Recent examination of cortical changes throughout development has shown that in adolescence, there are cortical thickness differences between individuals from lower and higher socioeconomic backgrounds across much of the brain. When further articulating when these changes first begin, data has shown that in early childhood (ages 4 to 7), the cortex is thinner (reflecting that of an older child) in individuals from lower socioeconomic backgrounds, suggesting that cortical thickness differences may be influenced by a variety of environmental circumstances, and likewise may also be preventable with appropriate evidence-based interventions. The theory that stress accelerates aging is well established, and the theory that early life stress accelerates maturation has been gaining ground.

STRESS, CORTISOL LEVELS, HYPOTHALAMIC–PITUITARY–ADRENAL AXIS

Racial discrimination and the chronic stress associated with racism and poverty have direct effects at the microcellular level and can lead to physiologic dysfunction. Racism increasing cortisol levels reflects the phenomenon of stress neurobiology. With increased stressors, there is a related increase in sustained release of stress hormones, which is critically important not only for all adults but also for pregnant women and their developing fetuses. We now know that not only does early exposure to maternal responses to stress shape the developing fetuses’ later stress responsivity, but also early postnatal exposure can have similar adverse effects. We are increasingly learning more about the mechanisms involved leading to these alterations such that now agreement exists as highlighted in a technical report issued by the AAP to increase the depth of surveillance. Unceasing stress hormones directly impact cortical structures with neuronal shrinkage of specific areas including the amygdala, hippocampus, and prefrontal cortex. Reductions in these areas can lead to specific behavioral challenges including emotional and behavioral dysregulation, as well as later executive function deficits. With greater elucidation of the science around early neurodevelopment, much is now known about the ways in which these unrelenting, pervasive amounts of stress impact these early developing systems with resultant increasing levels of earlier detection of mental health and behavioral challenges leading to increasingly complicated psychiatric presentations later in childhood.

In the past 20 years, this level of sustained exposure to specific stress hormones (cortisol, corticotropin-releasing hormone [CRH], adrenaline, norepinephrine) has often been labeled “toxic stress” secondary to the harm it produces on children’s developmental trajectories. This intimate intertwining of systemic racist environments impacting developing brains speaks to the need to not only understand the mechanisms of action involved, but also to better understand ways to mitigate these interactions in order to prevent, intervene, or buttress children of color who are disproportionately bearing the brunt of these disparate endemic conditions. All of this evidence speaks to the importance and understanding that early experiences matter greatly, as these very early experiences can alter neuronal development leading to lifelong consequences. With growing evidence of the impact that early, repeated, and pervasive stress leads to later physiologic and psychological changes, understanding racism as an important early adversity that can be mitigated is critical.

IMPACT OF RACISM AND DISCRIMINATION ON DEVELOPMENT

The health and development of young children are shaped by a process of ongoing, interactive adaptation from the cellular through to cortical systems beginning at conception and extending throughout the life course. The underlying plasticity that influences adaptive capacity (which decreases with age) can lead to a wide range of positive or negative outcomes influenced by the inextricable interactions among individual genetic/biologic predispositions and the environments in which pre- and postnatal development unfolds. Disparities in physical and mental health impose substantial human and economic costs on societies—and the relation between early adversity and lifelong well-being presents a rich scientific framework for fresh thinking about promoting health and preventing disease broadly, augmented by a deeper focus on how racism exacerbates inequities more specifically.

With increasing longitudinal research, we are beginning to understand the impact that racism and discrimination specifically have on child neurodevelopment at different ages. Secondary to the differing forms of racism as described earlier in the chapter, as well as the differential impacts on developmental stages, this section unpacks the ways in which these specific forms of racism and discrimination impact children and adolescents across five developmental stages: infancy and toddlerhood, preschool age, school age, adolescence, and transitional age youth.

Infant and Toddler

There is increasing evidence that environmental exposures including, but not limited to—intrauterine conditions, poverty, environmental toxicants, psychosocial stressors, maternal ACEs, racism, and traumatic life events —confer significant neurodevelopmental risk prenatally. Furthermore, there is increased evidence that maternal stress–related aberrations in utero are associated with a child's later risk for neuropsychiatric disorders and stress responsivity, as discussed earlier in the chapter. When thinking about the impact of racism on early childhood, it is critically important to understand that pregnant women's experiences of racism and discrimination impact not only their health and well-being but also that of the growing fetus. Crucially, evidence suggests that early environmental influences including racism impact children's outcomes including their behavioral health. Importantly, the multiplicity of factors has a worsening impact on early childhood development.

Data reflects that very young children experience both institutional/structural racism, as well as interpersonal or personally mediated racism (indirectly) through their mothers' or primary caregiver's experiences, with evidence identifying that racism created inequities for Black and brown mothers and children even before birth. Strikingly, Black babies have been shown to have the highest incidence of ACEs—more than 1.5 times the national average of two or more. Public health researcher, Geronimus developed the analytic framework of weathering, which essentially describes the ways in which the health of Black people is harmed by early deterioration secondary to exposure to structural/institutional racism. Her studies have demonstrated that racial discriminatory experiences lead to negative impact on telomeres. The length of telomeres, the end caps of chromosomes that shorten with cellular division, has been used as a marker of epigenetic age. Individuals who have experienced more stress have shorter telomeres. Geronimus has shown that exposure to stressors, such as racism and discrimination, can increase allostatic load, leading to faster aging or “weathering.” There is mounting evidence that disparities in the pace of aging begin early in life: children from racially and socioeconomically disadvantaged groups mature faster than their socially advantaged and non-Latinx White peers.

Preschool Age

Recent research has reflected the impact of structural racism and discrimination on early maturation noted by early molar development in preschool-aged children. There is mounting evidence that disparities in the pace of aging begin early in life with children from racially and socioeconomically disadvantaged groups, Black, and Latinx children, all maturing faster than their socially advantaged and non-Latinx White peers. Along with these early signs of early maturation as a consequence of racism and discrimination, further evidence reflects that as children age, they become increasingly aware of racism, specifically interpersonal or personally mediated racism, and even begin to struggle with internalized racism. Recent data has shown that Black and non-Latinx White children show pro-White biases by age 4, and that children have preferential play partners and this can be based on skin color or gender. But, importantly for young children, these biases are based solely on phenotypic differences or skin color NOT racist stereotypes. What becomes increasingly clear is that young children have clear in-group preferences. But, as these children aged, they focused on individual characteristics, such as shared interests, to determine in-group and out-group membership.

By age 5, Black children attribute social exclusion to their identified race. Less is known about how young children process interpersonal racism and social exclusion, or about individual differences in their sensitivity to these experiences. Kenneth and Mamie Clark discussed these issues in the 1940s with the seminal doll study showing the impact of structural racism in the form of school racial segregation and its impact on Black preschoolers in the South. Importantly, this data was so compelling that it was used over a decade later in 1954 when Thurgood Marshall and colleagues argued the landmark desegregation case of Brown *v* Board of Education citing the research as evidence of the direct harm segregation was having on young Black children.

As described earlier in this chapter, race is a social construct and very young children initially identify differences in skin color, however as they age they learn that there is a value attached to different skin colors. Theory and research has suggested that implicit intergroup attitudes emerge early in life at levels that are comparable to that of adults.

School Age

School-aged children have a different understanding and experience of racial dynamics. Research has demonstrated that non-Latinx White children can exhibit anti-Black bias by the age of 6. Research studies have also demonstrated that non-Latinx White children are more likely to choose other non-Latinx White children as friends over Black, Latinx, Asian, and Native American children. This can cause racial bullying at school and the exclusion from academic, school-related (sports, arts, groups), and social activities. This results in racial trauma and negative associations with school, learning, and education for non–White-identifying students. Moreover, research proves that children who are the only Black children in their class or school are more likely to experience more racist incidents while having less support from teachers, administration, and their peers at school. At 7 years old, most children have developed racial permanency, the understanding that their skin color is unchangeable, and in-group bias. By 9 years old, most children have an understanding of the status that their racial and/or cultural group holds in the broader society.

Children of color are often able to notice when preferential or oppressive racial dynamics play out at school, in their community, and in society at large. School-aged children are also able to recognize that racial violence in the form of police killings of citizens, xenophobia against immigrants, and race-based attacks are based on White supremacy. Dr. Alvin Poussaint hypothesized that Black people are not able to express appropriate rage in response to the racial violence that they experience which results in repressed anger and internalized psychological self-hatred. Elementary school children of color can internalize racism and racial stereotypes, leading them to unconsciously perceive people as having negative or positive characteristics based on race. This can further result in children of color internalizing White supremacy and believing in White superiority. Academic and social rejection in school by teachers and peers may lead students to avoid school for fear of traumatic school experiences which then affects students' academic participation and performance in school. In addition, children of color are often aware that their non-Latinx White peers are often given preferential treatment and access to additional resources. Accordingly, despite racial permanency, children of color can unconsciously desire to attain Eurocentric physical features to receive equal treatment at school and in broader society.

10 to 12 Years Old

Research has shown that Black children as young as 10 years old are routinely perceived to be older and less innocent than non-Latinx White children of the same age. This stereotyping leads to differential treatment, with Black children being labeled as aggressive, culpable, and needing to be watched and feared. In addition, in Black boys (aged 10 to 15 years), experiences with perceived interpersonal and institutional racism were related to increased internalizing symptoms such as depression and anxiety, and increased externalizing symptoms, such as anger or aggression. School-aged children of color are also able to identify that racial discrimination threatens safety and security. Direct and indirect experiences of race-based trauma can cause typical trauma symptoms such as fear, helplessness, and worry about the physical safety of children and their family members based on race. Race-based trauma can also affect caregivers' emotional availability, attachment, and ability to help children manage their psychological responses to racial stress and trauma. Racial trauma in school-aged children is associated with distractibility, problems in school, phobias, and other anxiety symptoms.

Dr. Kenneth B. Clark hypothesized that racism creates an unconscious hypervigilance in its victims. Accordingly, as school-aged children age, they are able to identify differential treatment based on race, but may not have the ability to recognize that their problems with mood, behavior, and self-image may be secondary to racism. Consequently, children tend to avoid settings in which they are made to feel "othered." A loss of safety and sense of belonging can result in decreased self-esteem and children losing a sense of hope, optimism, and belief in a future filled with endless possibilities. Research has shown that by age 12, there is increased pressure for youth to conform to racial in-group beliefs as evidenced by the tendency to shun nonconforming in-group members.

Early Adolescence

Adolescence is a crucial developmental stage in the life cycle that encompasses numerous physical and psychosocial changes. During this phase, youth become more attuned to their social standing within peer groups and increasingly separate from their family. In addition, adolescence is characterized by an exploration of one's sense of self and personal identity. For minoritized youth, this normative process becomes exceedingly more complex when confronted with increased exposure to institutional, interpersonal, and internalized forms of racism which threaten central parts of identity and may have deleterious effects on developmental pathways and adolescent mental and physical health.

Several studies have reported higher rates of internalizing disorders among Black, Native American, and Asian American adolescents compared to non-Latinx White adolescents. In a longitudinal, school-based epidemiologic study of adolescent depression encompassing a diverse range of adolescents in grades 7 through 12, minoritized youth were more likely to report higher baseline levels of depression. Secondary analyses identified several factors that predicted a higher risk for the onset of a depressive episode, including Black/Latinx ethnicity, female gender, and low-income socioeconomic status (SES). Latinx children and adolescents reported the highest levels of depressive symptoms of all ethnic groups across three separate waves of data collection. Asian Americans reported similarly high levels, followed by Black adolescents. Non-Latinx White adolescents had the lowest depressive scores over time, whereas depression scores for all other ethnic groups converged, despite initial differences.

Black youth are disproportionately impacted by traumatic events (i.e., physical, sexual abuse, witnessing violence). One study demonstrated that 65% of Black youth report traumatic experiences compared to 30% of their peers from other races. In addition, Black youth were more likely to report adverse emotional and behavioral sequelae to trauma exposure, including poor mental health, substance abuse, decreased well-being, and risky sexual behavior. The disproportionate rates of trauma and the associated outcomes observed among African Americans were thought to be moderated by unique race-based toxic stressors.

Late Adolescence/Transition Age Youth

The transition to adulthood has become a critical window in the life course, directing some young people on trajectories of increasing advantage and solidification and others on pathways of increasing disadvantage. In a seminal study of 422 patients independently observed over 5 years in a psychiatric emergency room (ER), it was found that after adjusting for psychotic disorders, the severity of disturbance, dangerousness, psychiatric history, use of restraints, time spent in ER, and other factors, compared to other patients, Black patients received, on average, more antipsychotic doses. In addition, data reflected that clinicians spent less time evaluating Black patients than non-Latinx White patients.

RACISM IN SYSTEMS OF CARE

Children and adolescents spend the majority of their time interfacing with different systems of care including schools, early intervention (EI) programming, child welfare, juvenile justice, and services for those with developmental disorders. However, data reflects that these systems, much like other systems in which children spend their time, may also be impacted by structural/institutional racism and/or interpersonal/personally mediated racism and discrimination. This section explores challenges that BIPoC children and their caregivers may face vis-a-vis experiences of racism and discrimination both within and outside of these systems which reflect long-standing, structurally, and racialized inequities that may adversely impact their developmental trajectories. Furthermore, when thinking of barriers or access to care, institutional/structural racism again plays a role in that the creation of redlining or residential segregation in the 1930s in over 200 cities across the United States has left neighborhoods in the 21st century bereft of certain services and infrastructure, thus placing patients farther away from quality care, as well as barriers around the costs of affordable care.

We posit that antiracist, culturally and trauma-informed approaches that focus on systemic inequities can create more equitable frameworks benefitting the children and adolescents they serve.

Racism in Early Care and Early Intervention

A greater understanding of the mechanisms of early stressors, including racism and discrimination on neurodevelopmental pathways, and how these stressors cause impairment of these critically developing systems has led to more work substantiating the importance and efficacy of early care and EI programming. The EI system was created to ensure that high-quality preventive and EI programming is accessible to very young children in order to prevent adverse outcomes. Several EI programs have shown efficacy in longitudinal assessments of a variety of various health, behavior, development, and educational outcomes. For children participating in high-quality EI programs (e.g., Perry Preschool Program, Carolina Abecedarian Project ([BC], Early Headstart, and the like) in the first 5 years of life, the impact of these programs is notable.

As more is known about the criticality of the first 5 years of life and neurodevelopment, understanding the benefits and challenges inherent in early childhood intervention, and prevention systems are incredibly important. Evidence shows that implicit racial bias and exclusionary discipline against children begin as early as preschool. Dr. Walter Gilliam's work in preschools highlights how teachers' biases and stigmatized beliefs directly impact young children's (particularly young Black boys with an expulsion rate of 3.6 times more than their non-Latinx White age-matched peers) ability to remain in early educational programming. Understanding how to buttress healthy caregiving to ensure that even when children are living in inequitable systems, institutional and federally funded supports exist to level the playing field.

Racism in Educational System

Children of color experience structural/institutional and interpersonal/personally mediated racism on a daily basis. A large portion of children's time is spent in school; a system purposed to facilitate healthy socioemotional development and equip children with knowledge while serving as a safe, nurturing, and affirming environment. Unfortunately, schools have often served the opposite role for children of color who experience repeated, daily structural and interpersonal racial discrimination in school settings from as early as preschool. Structurally, Black, Latinx, and Native American children are more likely to be educated in school systems with less experienced teachers, less advanced courses, and lower resources than their non-Latinx White peers. Multiple studies have shown that children of color perform better academically in school when they share the same cultural background as their teachers, yet youth of color are less likely to have a teacher, counselor, or principal of the same cultural background compared to their non-Latinx White peers. In addition, due to racial bias, Black and Latinx children score lower on standardized testing than their non-Latinx White peers. Subsequently, these factors impact students' engagement and investment in school, compelling many children of color to mentally and physically "check out" of school at an early age, which impacts their educational attainment, income, and status as adults.

In controlled preschool classroom simulations, teachers were more likely to have increased surveillance of Black children and expect misbehavior from them, especially Black boys, even when no behavioral issues were present. Studies have demonstrated that continual correction of students by teachers in front of peers has a negative effect on the peers' perceptions of students. Black and Latinx children are viewed as more aggressive than non-Latinx White students. This bias has resulted in Black preschoolers being suspended at a rate of more than three times their White peers. Black girls are six times more likely to be suspended from school than non-Latinx White girls. At school, similar behaviors in Black and White children, such as inattention, talking back to a teacher, or defiance, are often interpreted differently and can result in additional attention and resources for non-Latinx White students versus punishment for Black students.

In addition to disproportionate punishment, academic enrichment, leadership opportunities, and extracurricular opportunities are structured by racism and bias. Black students are less likely to be recommended for gifted education programs than their non-Latinx White peers, because teachers have been shown to have lower expectations for Black students. Moreover, children of color who are not academically challenged but may experience attentional and behavioral issues in school, are more susceptible to racially biased exclusionary disciplinary practices. Multiple studies have shown that Black students are disproportionately punished with suspension and expulsion. Factors influencing suspension include being Black, being male, and being perceived as looking older than their classmates. Black and Latinx youth are also less likely than their non-Latinx White peers to receive a referral by school professionals for an attention deficit hyperactivity disorder (ADHD) evaluation.

Exposure to racism in schools can render them unsafe spaces for children of color. This, in addition to other institutional forms of racism (family economics, neighborhood environment, nutrition), contributes to educational inequalities. By high school, Black, Latinx, and Native American youth were more likely to have higher rates of absenteeism and lower graduation rates than their non-Latinx White counterparts. This further exacerbates racial disparities as advanced educational attainment is a major predictor in life outcomes such as health and economics.

Racism in Special Education

As stated earlier in the chapter, secondary to the impacts of structural racism, youth of color are more likely to have experienced a variety of adverse outcomes, including low birth weight, exposure to lead and environmental pollutants, and living in unsafe neighborhoods, all of these are structural risk factors for developing disabilities. In addition, significant bias and structural racism has been shown in the rates of disability identification, evaluation, categorization, and access to and utilization of services rendered for Black and Latinx children in the special education system. Children diagnosed with one or more disabilities, such as autism spectrum disorder (ASD), emotional disturbance, and hearing impairment, qualify for special education. Black children are more likely to be diagnosed with emotional disturbance or intellectual disabilities; they are underrepresented in ASD and speech and language delay categories. Missed diagnoses mean that instead of receiving accurate diagnoses and appropriate services, Black youth with disabilities are more likely to be suspended than their non-Latinx White peers with disabilities.

Overall, Black and Latinx children are less likely to be enrolled in special education compared to children who are identified as non-Latinx White. However, when placed in special education programs, youth of color are more likely to be placed in the most restrictive educational settings. This means that they are more likely to spend 100% of their educational time in segregated, special education settings and are less likely to have access to general education curricula, teaching, and resources. Black and Latinx children in the high school special education system are more likely to be placed in tracks where they earn high school certificates instead of traditional high school diplomas, which prevents them from considering postsecondary education as a possibility. These exclusionary practices for Black and Latinx students in special education have significant repercussions because they restrict Black and Latinx youth to a future of limited job opportunities, income/economics, health, and life trajectories.

Racism in Juvenile Justice System

Importantly, challenges that children of color face in schools are also linked to their involvement in the juvenile justice system, a system that reflects larger societal structural racism. Students who experience exclusionary and punitive disciplinary measures in school are more likely to drop out of school and more likely to become involved in the juvenile justice system. Black, Latinx, and Native American youth are disproportionately represented in the juvenile justice system. According to 2021 data, youth of color comprise 49.8% of the U.S. population but make up 62% of the youth charged in the juvenile justice system. Racially minoritized teens have higher arrest rates for nonviolent, low-level offenses, and for noncriminal misbehaviors such as truancy and curfew violations. Research has also shown that race is a strong predictor of youth commitment to facilities where Black youth were more likely to be sent to boot camp or wilderness programs, while their non-Latinx White peers were committed to therapeutic programs such as drug/alcohol diversion programs and mental health facilities. In addition, Black children are five times more likely to be detained in juvenile justice facilities and are more likely to be sent from juvenile to adult court than their non-Latinx White peers.

During incarceration, youth experience solitary confinement, abuse, possible inadequate access to mental health care or education. Evidence has shown that involvement in the justice system “can lead to lower family functioning, impaired social networks, decreased high school graduation, decreased employment rates, increase in violence and victimization, and poorer mental health outcomes.” Subsequently, youth involved in the juvenile justice system are more likely to continue involvement in the justice system throughout adulthood.

Child Welfare

The child welfare system is another system characterized by inequity, racial disparity, and racial disproportionality across multiple levels. Oversurveillance of and overinvestigation of families of color reflect systemic, implicit, and explicit biases that are racialized and extend across every stage of decision making in the child welfare system, from reporting, investigating, intervention, and the placement process. This includes the types of services to which families are referred, are accessible, the degree of placement instability, and engagement with caseworkers. All of these reflect disparities where Black families have disparate contact and outcomes. It is documented that biased decision making contributes to the fact that caseworkers are more likely to substantiate claims of abuse and neglect for Black and Indigenous families resulting in disproportionate separation of Black and Indigenous youth from their families. From entry to placement or exit out of the child welfare system, Black children experience more harm, longer lengths of stay, protracted reunification periods, and slower rates of exit compared to other children.

Specifically, the mental health implications of structural racism within child welfare are significant. Researchers have shown how racism impacts mental health access for youth of color in the child welfare system and that despite having similar presentations in symptoms and diagnoses, youth of color have a higher likelihood than non-Latinx White youth to receive psychotropic medication and behavioral health diagnoses. Solutions proposed by researchers, advocates, and policy members include the development of culturally responsive practices, recruiting and retaining foster families of color, and understanding and addressing bias at various decision points.

Developmental Disabilities

Another system in which disparities exist for children of color is within developmental disabilities services. Proper and timely identification, access to treatment, and care are all challenges for BIPoC families. Multiple studies have identified that there are significant racial/ethnic disparities in the identification of ASD. Historically, there were lower rates of identification in Black and Latinx children compared to their peers. Recently, the CDC has shown an improvement in the identification of ASD in Black children. However, lower identification rates for Latinx children still persist. Black and Latinx children are still more likely to receive ASD evaluations at later ages and stages of severity than their non-Latinx White peers. Studies indicate that this delay in assessment and diagnosis occurs years after Black parents have voiced concern for their child’s development to teachers and pediatricians. Black parents are also more likely to experience difficulties getting ASD evaluations for their children and, on average, visit multiple providers before finally receiving a diagnosis of ASD. This delay in diagnosis denies children of color access to EI (which is related to longitudinal outcomes, including morbidity and mortality), services, and accommodations that are essential in their developmental trajectories and quality of life. Moreover, these children ultimately experience worse health outcomes.

Even when diagnosed with developmental disabilities, children of color are more likely to receive certain diagnoses that reflect racist stereotypes. Black students are twice as likely to be identified as having emotional disturbance and intellectual disabilities as their peers. Native American students are twice as likely to be identified as having specific learning disabilities and four times as likely to be identified as having developmental delays. In addition, children of color with disabilities are more likely to be placed in more restrictive learning environments than their non-Latinx White peers. After diagnosis, research has shown that Black and non-English speaking families receive fewer services for their children with developmental disabilities than non-Latinx White families who have children with developmental disabilities. These differences have also been identified at the state level, in that studies have shown that regional centers in California, which provide services for children with developmental disabilities, spend less funding on Black and Asian children than on non-Latinx White children.

RACIAL INEQUITIES IN MENTAL HEALTH TREATMENT AND OUTCOMES

The data reflects how BIPoC patients not only have less access to care, but when they do receive care, it is consistently poorer quality than that received by their non-Latinx White peers. Racism and racial bias are woven into the foundation of the American practice of psychiatry and mental/behavioral health care. In the 18th century, Dr. Benjamin Rush created the concept of *negritude*—the belief that having black skin was a disease only curable by a patient becoming White. In the 19th century, Dr. Samuel Cartwright invented the concept of *drapetomania,* also known as runaway slave syndrome, which believed that enslaved Africans who attempted to escape slavery had a mental disorder that could only be treated with physical punishment or amputations; he also invented *dysaesthesia aethiopica,* the theory that it was a mental illness for enslaved Africans to refuse to do manual labor. Mirroring the sociopolitical climate of American society, psychiatric institutions intentionally participated in racial segregation. Even when psychiatric institutions were “integrated,” they provided different standards of care and treatment for Black and White patients, resulting in significant inequities in mental health care. During the 1960s and 1970s, there was a shift in the definition and racial demographic of schizophrenia. More Black men were diagnosed with schizophrenia, and the diagnostic criteria began to include characteristics such as aggression, violence, agitation, and stereotypical terms often attributed to Black men. Per psychiatrist and historian Dr. Jonathan Metzl, these racially coded words were added to the definition of the disorder in order to discredit the mental fitness of many Black people who were involved in the Civil Rights Movement. This racialization and criminalization of mental illness also resulted in mental health care and treatment transferring from psychiatric institutions to the carceral system, which disproportionately imprisons BIPoC. To this day, psychiatrists are more likely to diagnose Black patients with schizophrenia and prescribe them higher doses of antipsychotic medications.

Adults of color are also more likely to have worse mental health outcomes than their non-Latinx White counterparts, which is of particular importance when recognizing that the health of the parents or primary caregivers directly influences the healthy development of their child. Child mental health prognosis is directly dependent on the mental wellness of their adult caregivers. In addition, inequities in mental health care extend to children of color. Psychiatrists’ racial bias and differences in access to mental health care have resulted in mental health disparities along color lines. Research has shown that implicit and explicit personally mediated racism can affect mental health care delivery and outcomes.

With respect to psychiatric and behavioral health care, recent literature reviews have indicated that ethnic/racial minority children with mental health difficulties are underserved relative to their non-Latinx White counterparts for prevention, access, quality treatments, and outcomes of care. Mental health care systems perpetuate racism and racial bias through the underdiagnosis and misdiagnosis of common pediatric psychiatric disorders and, moreover, the limited exploration and identification of the manifestations of racial trauma in children. Children and adolescents of color have experienced disparities in mental health care access, diagnostic variability, and treatment availability, resulting in differences in outcomes. Due to structural racism, Black and Latinx youth are more likely to live in areas with a shortage of health providers and, most likely, child psychiatrists. Systemic racism influences treatment seeking and the experience of mental health care for BIPoC children and youth. Studies from the United States cite experiences of racism, discrimination, and culturally incompetent care as barriers to help-seeking behaviors and retention in treatment. Studies in Canada demonstrated that youth are less likely to access outpatient levels of care voluntarily and more likely to initiate treatment in crisis settings or through the criminal justice system. When minoritized youth ultimately access mental health treatment, they are more likely to express dissatisfaction with their care due to experiences of race-based discrimination. In addition, attitudes toward help seeking are further complicated by cultural background (i.e., first- vs. second-generation immigrants) and views of mental health treatment as a westernized construct.

According to the 2006 to 2012 Medical Expenditure Panel Surveys, Black and Latinx children had fewer visits to psychiatrists and other mental health professionals and less substance use care than their non-Latinx White peers, even when controlled for mental health impairment, income, and health insurance. This may also show lower identification of mental illness in minority youth and decreased referral rates to psychiatrists from pediatricians and educators. Psychiatric and behavioral problems in youth of color are more likely to result in increased school punishment and incarceration than increased referrals for mental health care. Black adolescents diagnosed with MDD received mental health treatment at a much lower rate than non-Latinx White adolescents. Black and Latinx adolescents, even when severely impaired, were less likely than non-Latinx White adolescents to receive medication or therapy for mood and anxiety disorders. Once care was initiated, youth of color had fewer follow-up mental health appointments than their non-Latinx White peers.

A large portion of all diagnostic processes allows for clinician subjective evaluation, which leaves room for unconscious racial bias. ADHD is prevalent in children of all races and genders. However, non-Latinx White boys are more likely to be diagnosed with ADHD than any other group. By kindergarten, Black children are 70% less likely to receive the diagnosis of ADHD compared to non-Latinx White children, and by high school, non-Latinx White children are almost twice as likely to receive the diagnosis of

ADHD as Black children. Further studies have shown that Black and Latinx youth are more likely to receive the diagnosis of oppositional defiant disorder (ODD) and conduct disorder (CD) than the diagnosis of ADHD and are more likely to be diagnosed with ODD and CD than their non-Latinx White peers. These diagnoses make youth more vulnerable to adverse punitive disciplinary action in school, poor school performance, and involvement in the juvenile justice system. Failing to diagnose youth of color with ADHD can significantly alter one's life trajectory; it denies youth`s essential medications, therapy, and support services, further perpetuating or exacerbating professional and socioeconomic inequalities. Compounding overreporting of ODD and CD and underreporting depression, anxiety, and ADHD in youth of color is the fact that racial trauma is likely underreported in youth of color due to clinician's lack of awareness, bias, and discomfort surrounding racial topics and the narrow and exclusive scope of the PTSD criteria.

Disparities in mental health diagnosis and treatment have resulted in worsened mental health outcomes for children and adolescents of color. Research from Leventhal et al. showed that self-reported concern about increasing societal discrimination was associated with a higher frequency of substance use, a greater number of substances used, and higher odds of depression and ADHD symptoms. The suicide rate for Black youth has increased from 2.55 per 100,000 in 2007 to 4.82 per 100,000 in 2017, which is the fastest rate increase for any racial/ethnic group. Compared to their non-Latinx White peers, Black youth under 13 years are twice as likely to die by suicide. Research investigating the impact of perceived racial discrimination among children and adolescents found that exposure to perceived discrimination predicted poorer mental health care outcomes.

THE ROLE OF THE CHILD AND ADOLESCENT PSYCHIATRIST

Child and adolescent psychiatrists are poised to leverage both developmental and structural approaches to understand and respond to structural racism and the resultant adverse mental health outcomes in BIPoC children and youth. By acknowledging the role of racism in child and adolescent mental health, psychiatrists and other mental health professionals can proactively engage in systems-level interventions to identify, mitigate, and prevent the adverse effects that stem from structural, systemic, interpersonal/personally mediated, and internalized racism, thus improving the health and well-being of all children, adolescents, emerging adults, and their families. By using innovative strategies, including having an antiracist approach to clinical practice, engaging not only the family but also community partners, and building resilience through supporting racial–ethnic socialization and racial pride, child and adolescent psychiatrists can improve the equitable care of diverse children.

Intentionally antiracist practices, trauma-informed care, and culturally informed care are important factors in ensuring the healthy and optimal development of BIPoC children and youth. Psychiatric treatment of minoritized and marginalized youth should fundamentally be strength based, emphasizing creating safe spaces that foster trust, autonomy, and self-advocacy. Understanding the neighborhoods, schools, and communities in which children live and grow is an important part of child psychiatric practice. Optimizing clinical care includes understanding the lives and experiences of diverse children and adolescents and their families, ensuring that prevention and intervention programs interdigitate with real life. Screening for mental health and behavioral issues should always be nuanced and thoughtful, taking into account differential experiences as well as taking into account diverse cultural practices and beliefs.

Building Resilience

Importantly, many children and adolescents are able to thrive even in the face of racist systems and structures that directly impact their lives, neighborhoods, communities, and, ultimately, their development. As important as it is to describe how racism and discrimination can negatively impact developing children, equally important is recognizing that families have support systems that help to mitigate the impacts of racism and discrimination.

As has been noted, several studies have identified racism as impacting youths’ psychosocial functioning and emotional well-being, particularly around anxiety, depression, and academic functioning. Literature that examines protective and resilience-promoting factors highlights the importance of racial–ethnic socialization and racial pride in decreasing negative mental health outcomes. As described by Stevenson, Hughes, and colleagues, racial/ethnic socialization (R/ES) represents a diverse set of individual coping skills across various psychosocial domains to mitigate experiences of racial-related tension and stress in interpersonal relationships. R/ES, which includes affirming messages passed on between youth and their caregivers around racial pride and identity, has been shown to be particularly important for the academic and socioemotional well-being of Black children. Racial ethnic connectedness is also a form of cultural affirmation and pride that can positively impact academic achievement and decrease conduct and emotional problems.

Research and Antiracism

Despite a considerable increase in the number of studies on racism and health in recent years, relatively few have focused on children, adolescents, and transition age youth. Furthermore, much of the work has focused on the experiences of Black and, to a lesser extent, Latinx children and adolescents, excluding the experiences of Asian Americans, Hawaiian, Pacific Islanders, and Native/Indigenous children and adolescents. The sparse evidence from studies that draw on longitudinal data draws a correlation between racial trauma and poor health but provides few insights about the sensitivity of early exposure, the importance of duration of exposure, clear mechanisms of action, or the latency period between exposure to racism and discrimination and the onset of illness; as a result, the causal pathways through early childhood and later adolescence are not fully understood. What is known is the impact of perceived and observed experiences of discrimination and racism on child and family health outcomes can be profound. Future research efforts should (1) focus on nuanced tools to measure perceived racism and its impact on all BIPoC children and adolescents at different developmental ages and (2) explicitly focus on BIPoC youth as they are underrepresented in research, thereby ensuring that prevention and intervention strategies that are created will be as effective in diverse communities.

By ensuring that the research undertaken recruits and includes diverse families and youth, we will ensure more equitable care, access to treatment, and more holistic preventive strategies. Secondary to the increasing recognition of the dearth of inclusive research on children and youth of color, federal and state agencies are creating legislation to ensure these limitations cease in future funded research mechanisms. Allied organizations are also changing the language they use with youth and their families, enhancing their guidelines to identify and address racism and discrimination and its impact on developing children and youth. Both the AAP and the American Academy of Child and Adolescent Psychiatry (AACAP) have created resources for clinicians, children, and families that highlight the importance of addressing issues of racism and discrimination often and early, recognizing the detriment they have on child and adolescent mental health and well-being.

As this chapter comes to a close, the salient points highlighted throughout the sections include the growing evidence of the significant ways in which multiple forms of racism and discrimination can adversely impact developing children and adolescents. Also highlighted are the burgeoning evidence-based modalities and system changes that, once generalized, will ensure equitable care and treatment for all children. Increasingly, the role of the child and adolescent psychiatrist is to be able to understand emerging principles of neuroscience as they relate to child development and then integrate this understanding with information about the social determinants of mental health. A comprehensive appreciation of child development must include the impacts of racism and discrimination on child health. This chapter has taken a developmental perspective to articulate the myriad ways in which racism and discrimination harm children and youth while also ensuring that the science around prevention, intervention, and discussion of systems and practices to prevent, intervene, mitigate, and ultimately treat if necessary are also clearly understood.

FURTHER READINGS

Assari S, Moazen-Zadeh E, Caldwell CH, Zimmerman MA. Racial discrimination during adolescence predicts mental health deterioration in adulthood: gender differences among blacks. *Public Health Front*. 2017;5:104.

Baron AS, Banaji MR. The development of implicit attitudes: evidence of race evaluations from ages 6 and 10 and adulthood. *Psychol Sci*. 2006;17(1):53–58.

Berry OO, Londoño Tobón A, Njoroge WFM. Social determinants of health: the impact of racism on early childhood mental health. *Curr Psychiatry Rep*. 2021;23(5):23.

Broder-Fingert S, Mateo CM, Zuckerman KE. Structural racism and autism. *Pediatrics*. 2020;146(3):e2020015420.

Brown E. Yale study suggests racial bias among preschool teachers. The Washington Post 2016. https://www.washingtonpost.com/news/education/wp/2016/09/27/yale-study-suggests-racial-bias-among-preschool-teachers/

Camerota M, Willoughby MT. Prenatal risk predicts preschooler executive function: a cascade model. *Child Dev*. 2020;91(3):e682–e700.

Carter RT, Kirkinis K, Johnson VE. Relationships between trauma symptoms and race-based traumatic stress. *Traumatology*. 2020;26(1):11–18.

Center for Disease Control and Prevention. Media statement from CDC Director Rochelle P. Walensky, MD, MPH, on Racism and Health, 2021. https://www.cdc.gov/media/releases/2021/s0408-racism-health.html

Center for Disease Control and Prevention. Spotlight on closing the racial and ethnic gaps in the identification of autism spectrum disorder. Community Report, 2020. https://www.cdc.gov/ncbddd/autism/addm-community-report/spotlight-on-closing-racial-gaps.html

Fadus MC, Valadez EA, Bryant BE, et al. Racial disparities in elementary school disciplinary actions: findings from the ABCD study. *J Am Acad Child Adolesc Psychiatry*. 2021;60(8):998–1009.

Fadus MC, Ginsburg KR, Sobowale K, et al. Unconscious bias and the diagnosis of disruptive behavior disorders and ADHD in African American and Hispanic youth. *Acad Psychiatry*. 2020;44:95–102.

Felitti VJ, Anda RF, Nordenberg D, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: the adverse childhood experiences (ACE) study. *Am J Prev Med*. 1998;14(4):245–258.

Fredrickson GM, Camarill A. *Racism: A Short History*. Princeton University; 2002.

Geronimus AT. The weathering hypothesis and the health of African-American women and infants: evidence and speculations. *Ethn Dis*. 1992;2(3):207–221.

Geronimus AT, Hicken M, Keene D, John B. “Weathering” and age patterns of allostatic load scores among Blacks and Whites in the United States. *Am J Pub Health*. 2006;96(5):826–833.

Graignic-Philippe R, Dayan J, Chokron S, Jacquet AY, Tordjman S. Effects of prenatal stress on fetal and child development: a critical literature review. *Neurosci Biobehav Rev*. 2014;43:137–162.

Hughes D, Rodriguez J, Smith EP, Johnson DJ, Stevenson HC, Spicer P. Parents’ ethnic-racial socialization practices: a review of research and directions for future study. *Dev Psychol*. 2006:42:747–770.

Jones CP. Levels of racism: a theoretic framework and a gardener’s tale. *Am J Public Health*. 2000;90(8):1212–1215.

Lee J, Bell Z, Ackerman-Brimberg M. Implicit bias in the child welfare, education, and mental health systems. *NCYL*. 2015;11–12.

Love B, Hayes-Greene D. The groundwater approach: building a practical understanding of structural racism. *Racial Equity Institute*. 2018.

MacKinnon N, Kingsbury M, Mahedy L, Jonathan E, Ian C. The association between prenatal stress and externalizing symptoms in childhood: evidence from the Avon longitudinal study of parents and children. *Biol Psychiatry*. 2018;83(2):100–108.

Mandell DS, Wiggins LD, Carpenter LA, et al. Racial/ethnic disparities in the identification of children with autism spectrum disorders. *Am J Public Health*. 2009;99(3):493–498.

Marrast L, Himmelstein DU, Woolhandler S. Racial and ethnic disparities in mental health care for children and young adults: a national study. *Int J Health Serv*. 2016;46(4):810–824.

Masten AS. Resilience theory and research on children and families: past, present, and promise. *J Fam Theory Rev*. 2018;10(1):12–31.

McDermott CL, Hilton K, Park AT, et al. Early life stress is associated with earlier emergence of permanent molars. *Proc Natl Acad Sci U S A*. 2021;118(24):e2105304118.

McDonald SW, Madigan S, Racine N, Benzies K, Tomfohr L, Tough S. Maternal adverse childhood experiences, mental health, and child behavior at age 3: the all our families community cohort study. *Prev Med*. 2019;118:286–294.

Njoroge W, Benton T, Lewis ML, Njoroge NM. What are infants learning about race? A look at a sample of infants from multiple racial groups. *Infant Ment Health J*. 2009;30(5):549–567.

Pachter LM, Coll CG. Racism and child health: a review of the literature and future directions. *J Dev Behav Pediatr*. 2009;30(3):255–263.

Paradies Y, Truong M, Priest N. A systematic review of the extent and measurement of healthcare provider racism. *J Gen Intern Med*. 2014;29(2):364–387.

Parker CB. Teachers more likely to label black students as troublemakers, Stanford research shows. Stanford News 2015. https://news.stanford.edu/2015/04/15/discipline-black-students-041515/

Priest N, Doery K, Truong M, et al. Updated systematic review and meta-analysis of studies examining the relationship between reported racism and health and well-being for children and youth: a protocol. *BMJ Open*. 2021;11(6):e043722.

Saleem FT, Anderson RE, Williams M. Addressing the “myth” of racial trauma: developmental and ecological considerations for youth of color. *Clin Child Fam Psychol Rev*. 2020;23:1–14.

Sevon MA, Levi-Nielson S, Tobin RM. Addressing racism and implicit bias part 1: a response to the framework for effective discipline. *Communique*. 2021;49(5):10–12.

Shonkoff JP, Garner AS; Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood, Adoption, and Dependent Care, Section on Developmental and Behavioral Pediatrics. The lifelong effects of early childhood adversity and toxic stress. *Pediatrics*. 2012;129(1):e232–e246.

Smedley BD, Stith AY, Nelson AR, et al. *Institute of Medicine (US) Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care*. National Academies Press; 2003.

Stevenson HC. Racial socialization in African American families: the art of balancing intolerance and survival. *Fam J*. 1994;2:190–198.

Thompson T Jr, Massat CR. Experiences of violence, post-traumatic stress, academic achievement and behavior problems of urban African-American children. *Child Adolesc Social Work J*. 2005;22(5–6):367–393.

Trent M, Dooley DG, Dougé J; Section on Adolescent Health, Council on Community Pediatrics, Committee on Adolescence. The impact of racism on child and adolescent health. *Pediatrics*. 2019;144(2):e20191765.

Van den Bergh BRH, van den Heuvel MI, Lahti M, et al. Prenatal developmental origins of behavior and mental health: The influence of maternal stress in pregnancy. *Neurosci Biobehav Rev*. 2020;117:26–64.

Wamsley L. CDC director declares racism a ‘serious public threat.’ NPR, 2021. https://www.npr.org/2021/04/08/985524494/cdc-director-declares-racism-a-serious-public-health-threat

Weir K. Inequality at school. *Monit Psychol*. 2016;47(10).

Wilkinsom A, Laurore J, Maxfield E, et al. Zero to three: the state of babies yearbook. *Policy Brief*. 2021.

Williams DR, Lawrence JA, Davis BA. Racism and health: evidence and needed research. *Annu Rev Public Health*. 2019;40:105–125.

Williams DR, Mohammed SA. Racism and Health 1: pathways and scientific evidence. *Am Behav Sci*. 2013;57(8):1152–1173.

54.18 Forensic Evaluations of Child and Adolescent Migrants

Olivia Shadid, M.D., and Shawn Singh Sidhu, M.D.

Forensic psychological evaluations of child and adolescent asylum seekers are a rewarding, meaningful, and effective way for health care providers to engage with a unique population while providing a valuable service. The presence of an objective and independent medical evaluation may increase the likelihood of receiving asylum from 37.5% to 89%. Asylum hearings can have life or death significance for migrants seeking humanitarian relief. An official report from Human Rights Watch documents that hundreds of migrants have been murdered shortly after deportation from the United States to their country of origin, while others have reported being tortured, raped, or sexually trafficked after deportation to their country of origin.

MENTAL HEALTH PREVALENCE IN ASYLUM-SEEKING POPULATIONS

According to the United Nations High Commissioner for Refugees (UNHCR), 82.4 million people were forcibly displaced from their homes in 2020, 42% of whom were children. Premigration (armed conflict, disasters, abuse, and neglect), migration (treacherous conditions, trafficking, sleep deprivation), and postmigration (detention, family separation, racism and immigration stress, resettlement stress) trauma result in elevated rates of mental health diagnoses in migrants. Children show high rates of developmental regression and attachment difficulties in detention, while parents show high rates of anxiety, psychosis, and suicidal ideation. Both groups show alarmingly high rates of depression, anxiety, and post-traumatic stress disorder (PTSD).

OVERVIEW OF THE EVALUATION PROCESS

While there can be considerable variation in the asylum evaluation process, most cases follow a typical approach. Initially, an immigration attorney requests a mental health evaluation for their client. Once connected, the attorney assists in arranging an evaluation for their client. This evaluation may take place in the community, at the evaluator’s office, within a detention center or other forensic center, or through a telehealth evaluation. As in clinical interviews, it is helpful to collect collateral from caregivers, and, with very young children, caregivers may be present with the client throughout most of the interview. However, youth should always be given the opportunity to be interviewed without the caregiver present. The length of evaluations can vary considerably depending on the complexity of the case. Following the assessment, the evaluator will submit their written and signed document, or legal affidavit, to the attorney, who can then use this document in court to advocate for their client.

Evaluators

A broad range of mental health and medical professionals can conduct mental health evaluations for asylum-seeking immigrants. Trainees and professionals without an independent license may still conduct these evaluations if they have a co-signer who is independently licensed.

Preparing for the Psychological Evaluation

Prior to the interview, it is important to review any initial documentation, such as the client’s Declaration or Statement of Facts, I-589 Application, Credible Fear Interview, medical records, or any other documents provided by the attorney for accuracy and pertinent details. The evaluator should work with the attorney to discern whether the evaluator is a good fit for the evaluation based on their identifiable characteristics.

Conducting the Psychological Evaluation

Context, Introductions, and Confidentiality. Given the whirlwind of procedural meetings and interactions that asylum seekers may have with multiple individuals, it is of the utmost importance that the evaluator introduces themselves to the client and their caregiver as someone working with their attorney, clearly state their profession, and explain their role as an evaluator to the client. The evaluator should also introduce and explain the role of anyone else at the interview, such as the certified medical interpreter or if a trainee is present. Once the client and their caregiver are comfortable with information about the evaluator, notify the client that the evaluation may take several hours, that the client is in control of the interview, and that they can take breaks or stop at any time. Communicate also that this interview is completely voluntary, and their autonomy will be respected. Notify the client and caregiver that they should feel free to express how they are feeling throughout the interview and to ask questions at any time. To further elucidate the nature of the evaluation, it is useful to state clearly that while the evaluator is a mental health professional, they will not be providing the client with specific treatment plans after the evaluation or treating any mental health problems going forward. The client should also be informed about the confidentiality of the affidavit. The affidavit will be visible to legal authorities and representatives within the court setting, and thus its contents will not be confidential within the legal system. However, the contents of the affidavit are not to be shared in any capacity outside of the hearing with the federal immigration judge.

Interviewing the Client. It is common for asylum seekers to exhibit a broad range of emotions while telling their story and answering very detailed and specific questions about their traumas. The evaluator must carefully gauge the client’s readiness to continue in these circumstances. Above all, it is important that the evaluator take a trauma-informed approach to the interview to avoid potential retraumatization—which consists of attending to psychological and physical safety, transparency, collaboration, empowerment, and understanding cultural, gender, and historical issues—given the high prevalence of trauma among asylum seekers. Evaluators should also take a stance of cultural humility, reflecting and remaining cognizant of potential interviewer biases, privilege, and power dynamics in the interview. It is also essential to allow the client time to feel, process, and ground while discussing traumatic memories. Younger children, in particular, may benefit from the evaluation being conducted over multiple interviews to avoid overwhelming or risk dissociation or retraumatization. Evaluators can make empathic statements and give clients a sense of control if they experience symptoms, with necessary adjustments for telephonic and telehealth interviews. For young children, it can be helpful to start the evaluation with drawing, asking about hobbies, or playing a short game to establish rapport. Throughout the interview, offer breaks for the child to check in with their caregiver.

Developmental level and culture can play a significant role in how a client engages with an evaluator and, ultimately, how a client testifies. The evaluator should not hesitate to ask the client openly about their affect and note this in the affidavit. Child and adolescent clients are much more likely to display an incongruent affect when discussing traumatic or emotionally difficult topics, and even more so if the child has a history of developmental delays. Similarly, a client's culture or socioeconomic status can largely shape how they present for the interview. For example, a client from a culture in which the expression of negative affect is considered rude may display incongruent responses, such as smiling when recalling trauma. Such changes in affect and the topics precipitating them should be noted in the mental status examination. Understanding this dynamic more fully at the outset will help to inform the evaluator's interview and write-up.

Contextual History. While there are many ways to effectively conduct a psychological evaluation, one logical starting point is contextual history, particularly when interview time is limited. This is an opportunity for the client to express in their own words what life was like in their country of origin for individuals who shared the same identifiable characteristics as them. For example, if a family reports they were tortured and persecuted on the basis of belonging to a minority political group, it would be helpful for the court to understand how the predominant political group typically treated political opposition. Was the client's family the only family who was treated unfairly, or was this common among their associates as well? Personal stories and detailed examples help to illustrate the circumstances for the immigration judge and frame the narrative for the remainder of the evaluation. Starting with a contextual history also helps the youth ease into the interview and helps establish rapport with the evaluator.

Focused Trauma History. Given the importance of trauma in the psychological evaluations of asylum seekers and the link between trauma and claims for humanitarian relief, the evaluator should obtain a focused trauma history during the evaluation. This is one of the most critical sections of the affidavit, and high-quality reports have detailed trauma histories. If a client has endured multiple traumas, it can be helpful to ask initially about the traumas that impacted them the most, potentially including traumas during migration and following migration. While qualitative details can be helpful, quantitative details with specific numbers, dates, and times can sometimes cause a client's case to lose credibility erroneously. However, qualitative details can be important in describing how the client was harmed and demonstrating the degree of torture or damage inflicted on the client. This type of detail could include the type of trauma, who the attackers were, the instruments used, the type of injuries, how the client was injured, the type and duration of the pain, and whether or not there was any bleeding, bruising, burning, broken bones, or other forms of physical damage. In the case of sexual trauma, it can be important to ask potentially uncomfortable questions in a gentle manner, such as whether there was bleeding or lacerations, the type of any penetration, any resultant difficulty urinating or defecating, or persistent scarring. Occurrence and symptoms of a potential traumatic brain injury are also necessary to include, such as blows to the head, bleeding and/or bruising to the head or face or neck, whiplash, loss of consciousness, dizziness, unsteadiness, gait abnormalities, decreased reaction time, and short- or long-term headaches or memory difficulties. It is helpful to ask about short- or long-term impairments in function.

It is essential to note whether a client or client's family member's life was threatened or if they feared for their life during an event. Also important is to ask about what was said by perpetrators during the torture, as this ties the trauma into the client's legal claim (e.g., if the perpetrators of the torture mention political parties, and the client is claiming asylum on the basis of political persecution, such statements would be highly relevant to the client's case because they convey a motive for the torture).

In line with a trauma-informed approach, it is helpful for the evaluator to normalize and validate any potentially distressing emotions or physiologic symptoms that may come up during the interview. Evaluators should strive to maintain a supportive and nonjudgmental approach to the interview and avoid asking questions in a manner that may seem like a forced exposure or risk retraumatization. Evaluators should also explain why questions—particularly sensitive ones—are asked.

Social History. There are several elements of the social history that could impact the client's testimony and case. For example, the level of education can have a significant impact on testimony. If a client has little formal education and the judge is asking complex questions, the likelihood of a reliable or consistent answer is low, given the client's likely level of skill in rhetoric and vocabulary. The client should be asked about their school, occupational, and social goals and dreams following resettlement in the United States, as these small personal details can help paint a complete and holistic picture of the client.

Substance Use History. The implications of substance use in humanitarian relief cases are multiple and complex. Most evaluators will include at least a basic overview of the most commonly used substances in their report. If the client was either forced to use or had access to substances as a minor, this could be useful to note, as it could aid in characterizing their social and developmental history. It is important to note if the client was under the influence when the trauma occurred, if the substance was forced upon them, and if the substance was an attempt to self-medicate a significant mental health condition.

Past Mental Health History. A fairly standard past mental health history is warranted in a forensic psychological evaluation for asylum seekers. This should include prior counseling, inpatient hospitalizations, substance use treatment, suicidal and/or homicidal ideation, self-harm, auditory or visual hallucinations or delusions, a history of violence, and psychotropic medication trials. It is relevant to include any treatment that the individual received while in detention or in the community while awaiting their immigration hearing. If the client reports acute psychiatric symptoms, such as suicidal ideation with a plan and intent, or if the client is observed to be experiencing hallucinations and delusions to the extent that it impacts their safety, this should be reported to the attorney, caregiver, and/or detention center immediately. If the client presents with subacute symptoms, such as passive suicidal ideation, the inability to perform basic activities of daily living, or severe weight loss, the attorney, caregiver, and relevant authorities should be made aware that these symptoms are to be followed closely.

Past Medical History. A basic medical history should be included in the psychological evaluation. Developmental issues should be listed. A history of traumatic brain injury or concussion is especially important in this section, as is the documentation of any significant injury or impairments in functioning from past torture or persecution. Seizures must always be ruled out in a mental health evaluation. Current medications should be noted. If the individual required treatment and this was neglected, this should be noted.

Physical Examination. While a full physical examination is contraindicated as part of a mental health evaluation, some mental health evaluators choose to document scars in easily visible areas and the mechanism by which the patient reports that such injuries were obtained. For example, if a client was struck on the forehead and has a scar, the evaluator can document a "*hypopigmented lesion on upper right temporal legion, which the client states is from being struck with the butt of a rifle*." Similarly, regarding functional impairment, the evaluator could document that "*the client reports that they are unable to sit for long periods of time after being struck in the knee with the butt of a rifle, which could impact testimony*." Scars in less visible areas of the body can be described verbally by the client, and this verbal description can be included in the report.

Targeted Diagnostic Screening. A thorough diagnostic mental health screening is arguably the most important portion of the entire psychological interview of asylum seekers. The presence of a mental health diagnosis can be a way of demonstrating objective harm experienced by the client at the hands of the alleged perpetrators. At the same time, a mental health diagnosis can also explain why the client's presentation in court might vary based on their experiences.

While a thorough diagnostic interview based on criteria outlined in the Diagnostic and Statistical Manual–5 Text Revision (DSM-5-TR) remains the gold standard, providers may use diagnostic scales with which they are familiar to add an objective measure to the report. However, note that most scales and screening tools have not been studied in migrant populations. Limiting reported diagnoses to those that are most impairing removes the impression that a client may be overendorsing symptoms to the judge.

When screening for mental health disorders, it is important to proceed criterion by criterion and provide additional detail whenever possible. For example, stating "*Client screens positive for PTSD*" provides far less detail and evidence than "*Criterion B1: client endorses recurrent distressing dreams of the event in which they will dream that they are being physically attacked and awaken in a cold sweat, with the feeling that they were being restrained again and not remember where they are*."

The most common diagnoses in youth asylum seekers are PTSD, major depressive disorder, and anxiety disorders. Anxiety secondary to PTSD, an other specified trauma- and stressor-related disorder such as traumatic stress, or an adjustment disorder should be ruled out before diagnosing an additional anxiety disorder. Less common diagnoses include a psychotic disorder, bipolar disorder, cognitive disorder, or a developmental disorder/intellectual disability. Unaccompanied minors separated from both parents have been shown to have higher scores of PTSD and depression compared to youth asylum seekers who are not separated from their parents.

Aside from directly experiencing trauma, which would also include threats of harm or death, PTSD also can be diagnosed if a client saw a trauma happening to someone else or found out about a trauma happening to a loved one. For children, recurrence symptoms may be more likely to manifest in play, such as imagining facing monsters or acting out the trauma repeatedly, emotional dysregulation, or disrupted attachment.

Assessing psychosis in the context of an asylum evaluation can be complex and requires astute clinical skills. Psychotic symptoms most often result from trauma, PTSD, and/or mood disorders. The presence of multiple simultaneous voices, negative symptoms, disorganized speech, disorganized thought, or disorganized behavior is concerning for a primary psychotic disorder, such as schizophrenia.

The Mental Status Examination

The mental status examination is an important component of the assessment wherein the clinician documents objective findings of the client's presentation, including the client's demeanor, evidence of tearfulness or overwhelming emotion, incongruent or constricted affect, and need for breaks, thus providing insight into the patient’s emotional and cognitive state.

If the client is wearing an item of jewelry or has a tattoo that aligns them with certain social groups, these might be noted. Gender presentation may be important to note, particularly if the client was persecuted based on gender or intersectional identity.

Behavior is important to describe, particularly gestures of distress such as when a client holds their head in their hands, fidgets, looks at the ground, takes deep breaths and sighs, requires multiple breaks, clenches their fists, sweats, talks in either a loud or soft manner, the nature of eye contact, and attempts to avoid discussing distressing content or trauma. Affect may shift throughout the interview, and such alterations may be instrumental in acquiring a more nuanced, dynamic formulation of the client. When assessing cognition, such as attention, memory, orientation, insight, and judgment, it is important to factor in developmental age, educational history, and culture.

The following is a completely fictional sample mental status examination.

Alejandra is a 15-year-old transgender girl appearing younger than her stated age. She is feminine presenting. She wears a rosary necklace which she clutches during the interview, particularly during times of stress. She is well groomed, wearing pink lip gloss, and casually dressed. Alejandra is highly cooperative and engaged with the interview. She has good eye contact but will lower her eyes when talking about her traumas. She was observed being tearful throughout several portions of the interview, ranging from brief tearfulness to outright sobbing. During some portions of the interview, she laid her head down to rest, while at other times, she stared blankly off into the distance when discussing difficult elements. Her speech is very soft and slow, almost a whisper. Her motor examination is significant for psychomotor retardation or slowing. Her mood is, "I feel like I lost a piece of soul that I never got back after I was attacked." Affect is tearful, dysphoric, and anxious. Thought process is linear and logical. Thought content is without any hallucinations, delusions, suicidal ideation, or homicidal ideation. Insight and judgment appear good for age and developmental level.

GENERAL TIPS FOR WRITING AN EFFECTIVE PSYCHOLOGICAL AFFIDAVIT FOR ASYLUM SEEKERS

In the majority of cases, the contents of the affidavit are ultimately the only information from the mental health evaluation that will be available for use in the client's case; thus, what is written is crucial. Psychiatric terminology may be employed when making an official diagnosis or explaining medical phenomena, but consider also explaining such terms in lay language, as most judges and attorneys are not familiar with the technical language of psychiatry and medicine. Words such as "torture," "persecution," and "competency" should also be avoided, as these are legal determinations. If collateral information is obtained, such as from a caregiver, sources of information should be noted in the affidavit.

A few disclaimers can be helpful to note in the affidavit. For example, given the limitation in time, the focused trauma history may not be a comprehensive list of every trauma experienced by the client. This is helpful in the event that the client shares different traumas with immigration authorities than they do with the psychological evaluator. It is also helpful to note that quotations are paraphrased and not exact quotes if the client had used slightly different quotations elsewhere.

Assessment

The assessment and discussion sections of the affidavit are most likely to be of direct benefit and utility to the attorney and the client. The purpose of the assessment is to state whether the evaluator found the client's reported history to be consistent with a history of traumatization. An assessment of whether the psychological findings the client reported are typical reactions to extreme stress may also be documented in this section. The evaluator may also use the qualities of the interview in their assessment of consistency. For example, the evaluator may find that the emotional responses the client exhibited during the interview and the ways in which they described vivid details of certain traumatic events are consistent with the experiences the client reported. This section also offers an opportunity to note whether the evaluator feels the client endorses symptoms indiscriminately or if they endorse some, rather than all, symptoms when they are queried.

Diagnosis

Following the assessment, the evaluator should include whether the client meets the criteria for any diagnoses described in the DSM-5-TR. Although not all youth will meet the full criteria for a disorder, such as PTSD, they often exhibit symptoms of a disorder, such as several symptoms of PTSD, which should be detailed and explained clearly. Youth may describe symptoms of a disorder in the past, but symptoms may be currently in remission, especially if they are living in a supportive, safe environment without ongoing exposure to the trauma. Diagnoses should be clearly labeled with the accompanying ICD-10 F-Codes (e.g., "*Post-Traumatic Stress Disorder–F 43.10*").

Discussion

In the discussion section, the evaluator can offer the attorney and judge information on the client's ability to testify, particularly noting findings or framings that may lead to impairment in this ability. Aspects that can affect the ability to testify range from the effects of psychiatric or physical symptoms to cognitive or educational status to cultural factors to qualities of resilience.

If the client is found to have suffered from trauma, the evaluator can discuss how impaired memory is not only a symptom of trauma disorders but also consistent with neurobiologic understandings of the neurotoxic effects of chronic stress, via cortisol elevation, on the hippocampus, the brain's memory center. Similarly, impaired concentration, memory, sleep, and cognition are also associated with many mental health diagnoses. Multiple cognitive domains, including memory and executive functioning, may also be affected by any traumatic brain injuries suffered by the client. If there is a concern for baseline neurocognitive deficits, such as in a client with an intellectual disability, this should also be noted as a possible impairment in the ability to testify.

Clients suffering from trauma may also have difficulty testifying should they experience "intrusion symptoms" such as flashbacks, distressing memories, or dissociation while testifying. This risk can be mitigated by acknowledging triggers of intrusion symptoms and minimizing their presence in the courtroom. Clients may also have physical conditions that may impair their ability to testify secondary to pain or other difficulties.

For children, the evaluator may discuss why a child may not recall particular events or symptoms or avoid talking about events or symptoms by discussing the impact of developmental stage, trauma, language, and mental health diagnoses on memory and recall.

For children and for any client with developmental or intellectual disabilities, it can be useful to describe what developmental stage they are in, noting the competencies they have attained and those that they have not yet likely mastered. If a client, whether child or adult, has not completed high school, the client may have difficulty with both vocabulary and rhetorical skills possessed by those who have completed higher years of formal schooling.

An evaluator may discuss how the affect displayed by the client during testimony may appear incongruent with the topics they are discussing. In addition, the evaluator may discuss the mental health impacts of detention and family separation on a client. Finally, the evaluator may cite literature that explains why parents may remember or experience certain events differently than their children.

Conclusion and Signature

The affidavit should conclude by restating the mental health diagnoses and explaining that there are highly effective, evidence-based treatments available for the client within the United States for said diagnoses. If the client is in need of urgent treatment, that should also expressly be stated here. If the evaluator has knowledge of mental health care availability in the client's country of origin, they can note whether the availability of such treatments may be limited in the client's country of origin, which could lead to psychiatric decompensation. Finally, the evaluator will write that they swear by penalty of perjury that all information contained in the affidavit ("herein") is truthful to the knowledge of the evaluator and sign and date the affidavit for submission.

Follow-Up

Once drafted, the affidavit should be sent to the client's attorney for review. The attorney may ask clarifying questions or comment on potential issues the affidavit may pose that are unique to the courts in which they practice. Following this revision process, the evaluator will finalize, sign, and date the affidavit. In the majority of cases, this will conclude the asylum evaluation process.

CONCLUSION

Conducting forensic psychological evaluations can be a meaningful and rewarding experience for health care providers and can significantly improve the likelihood of receiving asylum or other forms of humanitarian relief for children and families fleeing torture and persecution. Prior to evaluating the client, the clinician should carefully review all legal documents and arrange for a safe interview space and high-quality interpretation. The evaluator should bring a trauma-informed, culturally sensitive, and developmental lens to the interview and conduct the interview empathically by conveying that the client is in control, meanwhile obtaining a detailed trauma history, contextual history, and mental health review of systems. The mental health status examination should be used as a tool to convey the experience of being in the room with the client.

The evaluation should conclude with a thorough and nuanced assessment and discussion, in which the diagnosis, consistency with trauma, and factors affecting testimony are described completely for legal authorities. The affidavit concludes when the evaluator states that they swear the aforementioned is truthful to the best of the examiner's knowledge and that the document is signed, dated, and submitted.

Finally, as they are at risk for vicarious trauma, evaluators should monitor their own countertransferences and responses to hearing, engaging with, and carrying the stories of often severely traumatized children and adolescents. Evaluators may find solace and integration by debriefing with colleagues and by consciously noting not only the traumatic experiences of clients but also their remarkable resilience despite their traumas.

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Publishing; 2013.

Asylum Denial Rates Continue. Transactional records access clearinghouse. newhouse school of public communications and the whitman school of management. Syracuse University. October 28, 2020. Accessed December 27, 2020. https://trac.syr.edu/immigration/reports/630/#:~:text=During%20FY%202020%2C%2073.7%20percent,26.3%20percent%20of%20the%20time

Baranowski KA, Moses MH, Sundri J. Supporting asylum seekers: clinician experiences of documenting human rights violations through forensic psychological evaluation. *J Trauma Stress*. 2018;31:391–400.

Cleveland J, Rousseau C. Psychiatric symptoms associated with brief detention of adult asylum seekers in Canada. *Can J Psychiatry*. 2013;58(7):409–416.

Deported to danger: United States deportation policies expose salvadorans to death and abuse. An official report by human rights watch. February 2020. Accessed December 27, 2020. https://www.hrw.org/sites/default/files/report_pdf/elsalvador0220_web_0.pdf

Derluyn I, Mels C, Broekaert E. Mental health problems in separated refugee adolescents. *J Adolesc Heal*. 2009;44(3):291–297.

Gartland MG, Ijadi-Maghsoodi R, Giri M, et al. Forensic medical evaluation of children seeking asylum: a guide for pediatricians. *Pediatr Ann*. 2020;49(5):e215–e221.

Habbach H, Hampton K, Mishori R. You will never see your child again: the persistent psychological effects of family separation. An official report by physicians for human rights. 2020. Accessed May 04, 2020. https://phr.org/wp-content/uploads/2020/02/PHR-Report-2020-Family-Separation-Full-Report.pdf

Iacopino V. Medical evaluations of asylum seekers. *Virtual Mentor*. 2004;6(9).

Katz CL, Springer EK, Green A, Ruchman S, Popil E, Justin G. *Best Practices in Conducting Forensic Telephonic Asylum Evaluations*. Mount Sinai Health Systems; 2019.

Kien C, Sommer I, Faustmann A, et al. Prevalence of mental disorders in young refugees and asylum seekers in European countries: a systematic review. *Eur Child Adolesc Psychiatry*. 2019;28(10):1295–1310.

Lustig SL, Kureshi S, Delucchi KL, Iacopino V, Morse SC. Asylum grant rates following medical evaluations of maltreatment among political asylum applicants in the United States. *J Immigr Minor Health*. 2008;10(1):7–15.

Mohlenhoff BS, Chao LL, Buckley ST, Weiner MW, Neylan TC. Are hippocampal size differences in posttraumatic stress disorder mediated by sleep pathology? *Alzheimers Dement*. 2014;10(Suppl 3):S146–S154.

Peeler KR, Hampton K, Lucero J, Ijadi-Maghsoodi R. Sleep deprivation of detained children: another reason to end child detention. *Health Hum Rights*. 2020;22(1):317.

Shadid O, Sidhu SS. The mental health effects of migrant family separation. *J Am Acad Child Adolesc Psychiatry*. 2021;60(9):1052–1055.

Sidhu SS, Vasireddy R. The detention of migrant families. *J Am Acad Child Adolesc Psychiatry*. 2020;59(6):681–683.

Tervalon M, Murray-García J. Cultural humility versus cultural competence: a critical distinction in defining physician training outcomes in multicultural education. *J Health Care Poor Underserved*. 1998;9(2):117–125.

Turrini G, Purgato M, Ballette F, Nosè M, Ostuzzi G, Barbui C. Common mental disorders in asylum seekers and refugees: umbrella review of prevalence and intervention studies. *Int J Ment Health Syst*. 2017;11:51.

United States Department of Health and Human Services. SAMHSA's concept of trauma and guidance for a trauma-informed approach. 2014:1–27. https://store.samhsa.gov/system/files/sma14-4884.pdf

von Werthern M, Robjant K, Chui Z, et al. The impact of immigration detention on mental health: a systematic review. *BMC Psychiatry*. 2018;18(1):382.

Yeager KA, Bauer-Wu S. Cultural humility: essential foundation for clinical researchers. *Appl Nurs Res*. 2013;26(4):251–256.

55

Geriatric Psychiatry

55.1 Overview

55.1a Introduction to Geriatric Psychiatry

Dilip V. Jeste, M.D., Dylan J. Jester, Ph.D., M.P.H., and Ellen E. Lee, M.D.

The population of the United States is aging rapidly, with both an increased number and an increased proportion of older adults. The number of Americans aged 65 and over has increased from 3 million (4% of the total population) in 1900 to 54 million (16% of the population) in 2019. Yet, this historical rise will continue to be dwarfed by that which started in the year 2011 when the first baby boomers (those born between the years 1946 and 1964) reached the age of 65—the traditional definition of "old age." The number of people over 65 will rise to 81 million in 2040 and 95 million (24% of the population) by 2060 (Fig. 55.1a–1).

The average life span in the United States has increased from 47 years in 1900 to 79 years today. The more optimistic estimates project an increase in life expectancy at birth from the current figures of 76 years for men and 81 years for women to 86 years for men and 92 years for women by 2050. Men and women who are 65 years old today are expected to live an average of 16 and 19 years, respectively. By 2050, the average life span at age 65 would have increased to 25 and 30 years, respectively. Thus, reaching age 65 would not indicate imminent death but rather many more years of life. One of the unwelcome scenarios that could torpedo such an age wave would be a continuation of the current epidemic of obesity in younger and middle-aged individuals leading to a rising incidence of diabetes and other metabolic and vascular diseases that may shorten life spans.

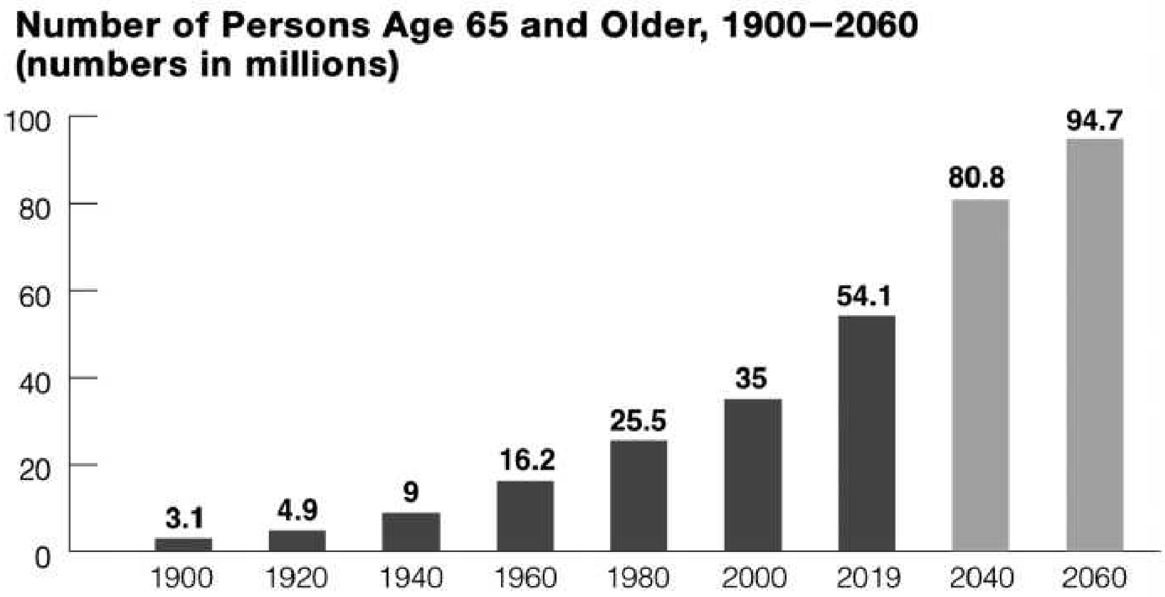

FIGURE 55.1a–1. Number of Americans 65+ (in millions) (U.S. Census Bureau, 2019). (From U.S. Census Bureau. 2020 Profile of Older Americans; 2021. https://acl.gov/sites/default/files/Aging%20and%20Disability%20in%20America/2020ProfileOlderAmericans.Final_.pdf)

The fastest-growing segment of the population in the United States is that of individuals over the age of 85. The 85-and-over age group is expected to increase from about 6 million today to more than 18 million in 2050. Also striking is the growth of the centenarian population. While there were about 5,000 centenarians in the United States in 1900, today, there are around 53,000 people in that age group in the United States or 1.73 centenarians per 10,000 people in the total population. By 2050, that number will have gone up to about 835,000. A particularly notable fact is that whereas most of the centenarians in the United States today are White, nearly half of those in 2050 will belong to ethnic minority groups.

The aging of the population is not restricted to the United States but is a global phenomenon. The world is getting older at a pace that has taken demographers by surprise and presented governments with economic, environmental, and social challenges. Whereas there were about 605 million people over age 60 in the year 2000, that number will reach 2 billion in the year 2050. Two-thirds of all the people in the entire history of the world who have reached the age of 65 are alive today. The number of centenarians in the world will increase from approximately 316,000 at the present time to an incredible 2.2 million by 2050. To put the latter number in perspective, presently, the entire population of countries such as Kuwait, Latvia, Macedonia, and the United Arab Emirates is roughly 2 million people each.

The rates of aging of the population vary widely by region. In some developed countries today, the proportion of older persons is close to 20%—that is, one in five. During the next several decades, that proportion will reach one in four and, in a few countries, as high as one in two (i.e., 50%). Highly developed countries have more stable populations and will experience proportionately less growth in their older populations than developing countries that are undergoing a rapid increase in the number of aged individuals.

We may need to revise our definitions of middle age and old age as well as some of our traditional thinking about chronologic versus biologic aging as well as psychological aging ("felt age" that indicates how old one feels). It has been suggested that middle age be defined as age between 51 and 70 years, "younger old" between 71 and 90 years, and "older old" over age 90. With increasing longevity, the timetable for brain aging may have changed as well.

Table 55.1a–1.
Population by Age and Sex: 2019 (Numbers in thousands. Civilian noninstitutionalized population.[a])

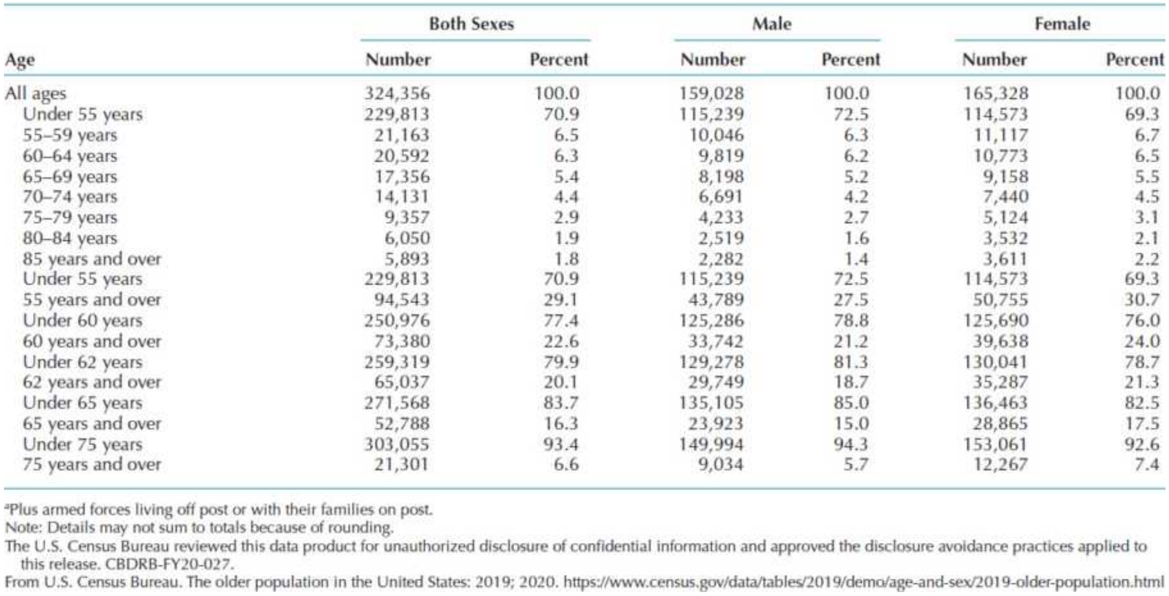

Age	Both Sexes		Male		Female	
	Number	Percent	Number	Percent	Number	Percent
All ages	324,356	100.0	159,028	100.0	165,328	100.0
Under 55 years	229,813	70.9	115,239	72.5	114,573	69.3
55–59 years	21,163	6.5	10,046	6.3	11,117	6.7
60–64 years	20,592	6.3	9,819	6.2	10,773	6.5
65–69 years	17,356	5.4	8,198	5.2	9,158	5.5
70–74 years	14,131	4.4	6,691	4.2	7,440	4.5
75–79 years	9,357	2.9	4,233	2.7	5,124	3.1
80–84 years	6,050	1.9	2,519	1.6	3,532	2.1
85 years and over	5,893	1.8	2,282	1.4	3,611	2.2
Under 55 years	229,813	70.9	115,239	72.5	114,573	69.3
55 years and over	94,543	29.1	43,789	27.5	50,755	30.7
Under 60 years	250,976	77.4	125,286	78.8	125,690	76.0
60 years and over	73,380	22.6	33,742	21.2	39,638	24.0
Under 62 years	259,319	79.9	129,278	81.3	130,041	78.7
62 years and over	65,037	20.1	29,749	18.7	35,287	21.3
Under 65 years	271,568	83.7	135,105	85.0	136,463	82.5
65 years and over	52,788	16.3	23,923	15.0	28,865	17.5
Under 75 years	303,055	93.4	149,994	94.3	153,061	92.6
75 years and over	21,301	6.6	9,034	5.7	12,267	7.4

[a] Plus armed forces living off post or with their families on post.
Note: Details may not sum to totals because of rounding.
The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. CBDRB-FY20-027.
From U.S. Census Bureau. The older population in the United States: 2019; 2020. https://www.census.gov/data/tables/2019/demo/age-and-sex/2019-older-population.html

The growth of our older population during the 20th century is attributable to several factors. During the first half of that century, the main factor responsible for the demographic change was a progressive and dramatic decrease in perinatal and infant mortality due to enhanced sanitation efforts, the advent of antibiotics and vaccines, improved nutrition, and better prenatal and postnatal care. In the latter half of the last century, the greater longevity resulted primarily from an increase in average life span—that is, adults lived longer as a result of declines in mortality from pneumonia, heart disease, and other chronic diseases.

Aging is associated with increased heterogeneity. Tables 55.1a–1 to 55.1a–6 show numbers of people over age 55 in the United States by age and gender, marital status, educational attainment, earnings of full-time workers, types of household, and monetary income. The wide range of diversity in socioeconomic status in older adults is remarkable. Factors such as education, marital status, and income have a potential bidirectional relationship with a number of mental illnesses. For example, adults with schizophrenia are likely to have lower levels of education, be unmarried, and have lower incomes than the general population. These variables could be considered potential risk factors for the illness; it is more likely, however, that these factors are secondary to the development of a serious mental illness. In either case, they are of considerable importance in the management of persons with schizophrenia.

WHAT IS NORMAL AGING?

"Normal" aging, which reflects typical or average (rather than illness-free) aging, must be differentiated from disease. Some bodily functions decline with age, but health problems are not inevitable. Most diseases of old age, such as heart disease, cancer, and dementia, result from pathologic processes rather than from normal aging. Disabilities resulting from diseases can usually be treated and sometimes prevented. For example, most hearing problems can be at least partially corrected with hearing aids, while the risk of heart disease can be reduced by healthy habits with respect to diet and exercise. The same applies to mental health and many mental illnesses in older people.

One point to stress is that aging is associated with more, and not less, diversity. It is a common misconception that older adults look, think, and behave alike. Indeed, there are more differences among older people than among younger adults. Even in the same person, different organs "age" at different rates. It has been suggested that we should abandon the notion of chronologic age ("how old are you?"), and instead, consider the stages of aging that a given individual exhibits in different areas of functioning. Furthermore, just as a child may be at different stages of development mentally and physically, an older adult may have divergent aging of their brain, liver, kidneys, and heart. As a matter of fact, not even all the parts of the brain age at the same rate in a person. Thus, an aged person may have relatively intact memory while having deficits in motor function.

We must recognize that although we need age cut-offs for defining old age for societal purposes (e.g., the qualifying age for starting Medicare or for Social Security), such cut-offs are artificial and arbitrary. No magical change occurs overnight on the 65th or any other birthday. Also, with increasing longevity, the age cut-offs used currently will need to change.

There is often a tendency to confuse the effects of old age with those of its common, but not inevitable, associations. Thus, medical morbidity increases with aging, but many physical diseases can be managed better, and some (e.g., nutritional deficiencies) can be prevented. The diverse treatment of various physical illnesses by different specialists resulting in polypharmacy has an adverse influence on the treatment of mental disorders in older adults. Similarly, psychosocial stressors such as financial hardship and lack of appropriate community services impact the management of older people with psychiatric symptoms.

Table 55.1a–2.
Marital Status of the Population 55 Years and Over by Sex and Age: 2019 (Numbers in thousands. Civilian noninstitutionalized population.[a])

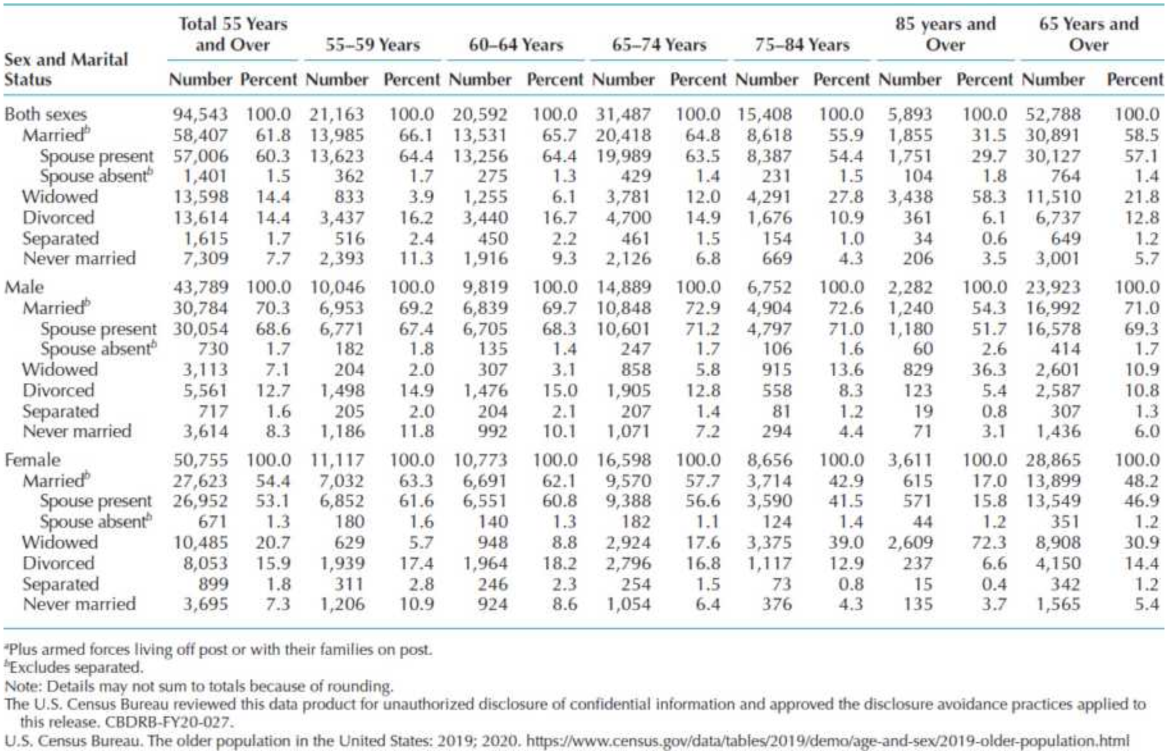

Sex and Marital Status	Total 55 Years and Over		55–59 Years		60–64 Years		65–74 Years		75–84 Years		85 years and Over		65 Years and Over	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Both sexes	94,543	100.0	21,163	100.0	20,592	100.0	31,487	100.0	15,408	100.0	5,893	100.0	52,788	100.0
Married[b]	58,407	61.8	13,985	66.1	13,531	65.7	20,418	64.8	8,618	55.9	1,855	31.5	30,891	58.5
Spouse present	57,006	60.3	13,623	64.4	13,256	64.4	19,989	63.5	8,387	54.4	1,751	29.7	30,127	57.1
Spouse absent[b]	1,401	1.5	362	1.7	275	1.3	429	1.4	231	1.5	104	1.8	764	1.4
Widowed	13,598	14.4	833	3.9	1,255	6.1	3,781	12.0	4,291	27.8	3,438	58.3	11,510	21.8
Divorced	13,614	14.4	3,437	16.2	3,440	16.7	4,700	14.9	1,676	10.9	361	6.1	6,737	12.8
Separated	1,615	1.7	516	2.4	450	2.2	461	1.5	154	1.0	34	0.6	649	1.2
Never married	7,309	7.7	2,393	11.3	1,916	9.3	2,126	6.8	669	4.3	206	3.5	3,001	5.7
Male	43,789	100.0	10,046	100.0	9,819	100.0	14,889	100.0	6,752	100.0	2,282	100.0	23,923	100.0
Married[b]	30,784	70.3	6,953	69.2	6,839	69.7	10,848	72.9	4,904	72.6	1,240	54.3	16,992	71.0
Spouse present	30,054	68.6	6,771	67.4	6,705	68.3	10,601	71.2	4,797	71.0	1,180	51.7	16,578	69.3
Spouse absent[b]	730	1.7	182	1.8	135	1.4	247	1.7	106	1.6	60	2.6	414	1.7
Widowed	3,113	7.1	204	2.0	307	3.1	858	5.8	915	13.6	829	36.3	2,601	10.9
Divorced	5,561	12.7	1,498	14.9	1,476	15.0	1,905	12.8	558	8.3	123	5.4	2,587	10.8
Separated	717	1.6	205	2.0	204	2.1	207	1.4	81	1.2	19	0.8	307	1.3
Never married	3,614	8.3	1,186	11.8	992	10.1	1,071	7.2	294	4.4	71	3.1	1,436	6.0
Female	50,755	100.0	11,117	100.0	10,773	100.0	16,598	100.0	8,656	100.0	3,611	100.0	28,865	100.0
Married[b]	27,623	54.4	7,032	63.3	6,691	62.1	9,570	57.7	3,714	42.9	615	17.0	13,899	48.2
Spouse present	26,952	53.1	6,852	61.6	6,551	60.8	9,388	56.6	3,590	41.5	571	15.8	13,549	46.9
Spouse absent[b]	671	1.3	180	1.6	140	1.3	182	1.1	124	1.4	44	1.2	351	1.2
Widowed	10,485	20.7	629	5.7	948	8.8	2,924	17.6	3,375	39.0	2,609	72.3	8,908	30.9
Divorced	8,053	15.9	1,939	17.4	1,964	18.2	2,796	16.8	1,117	12.9	237	6.6	4,150	14.4
Separated	899	1.8	311	2.8	246	2.3	254	1.5	73	0.8	15	0.4	342	1.2
Never married	3,695	7.3	1,206	10.9	924	8.6	1,054	6.4	376	4.3	135	3.7	1,565	5.4

[a] Plus armed forces living off post or with their families on post.

[b] Excludes separated.

Note: Details may not sum to totals because of rounding.

The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. CBDRB-FY20-027.

U.S. Census Bureau. The older population in the United States: 2019; 2020. https://www.census.gov/data/tables/2019/demo/age-and-sex/2019-older-population.html

DEMOGRAPHICS OF OLDER PEOPLE WITH MENTAL ILLNESSES

Aging and aging-associated disorders (e.g., metabolic and vascular diseases) may influence the expression and management of mental illnesses. Yet, psychiatric epidemiologic studies have tended to exclude older adults. The common notion that psychiatric disorders are rare among older individuals is just plain wrong. This misconception, prevalent even among nongeriatric mental health experts, is based on studies such as the Epidemiologic Catchment Area (ECA) study conducted by the National Institute of Mental Health in the early 1980s. The ECA study was one of the largest epidemiologic investigations of psychiatric disorders using, what was at that time, state-of-the-art methodology, including then-new DSM-III criteria. The ECA study reported that the prevalence of any psychiatric disorder (excluding dementia), either over the previous year or over the entire lifetime of the individuals, was considerably lower among older compared to younger adults. This decrease in the prevalence of mental illness with aging was thought to be due, in part, to higher mortality associated with serious mental illnesses. Yet, an important explanation for this ECA finding could be that psychiatric disorders were underdiagnosed among older adults for a variety of reasons, indicating differences due to a "cohort effect" (i.e., different rates of an illness in people from different generations).

According to the ECA study, 13% of people over age 65 would meet the criteria for psychiatric disorders other than dementia. This is, however, likely to be a serious underestimate because of factors such as misattribution of psychiatric symptoms (by patients, caregivers, and raters) to other causes such as cognitive impairment, physical disorders, or "normal aging"; a lack of age-appropriate diagnostic criteria for major psychiatric disorders including substance abuse or dependence; and underreporting of psychological symptoms due to forgetfulness and social or gender-based stigma. The "real" prevalence of psychiatric disorders other than dementia in older adults is probably at least 25% higher—that is, over 16%. In addition, over 10% of older adults have dementia, usually of the Alzheimer type; neuropsychiatric behavioral disturbances are common in patients with dementia, with at least a third of them (i.e., about 3% of the total older adult population) exhibiting psychotic and/or depressive symptoms. Thus, a total of nearly 20% of persons over age 65 would have diagnosable psychopathological symptoms.

Figure 55.1a–2 shows the prevalence of mental illnesses among younger and older (>50 years) adults. While the overall prevalence of mental illness among young adults has increased in recent years, it has remained relatively stable among older adults. Mood disorders, anxiety disorders, substance use disorders, and personality disorders tend to decrease in prevalence as age increases. However, there may be sex differences between the octogenarian and nonagenarian years of life. Older men who survive to age 85 years or older have a higher prevalence of mood disorders (e.g., major depression, dysthymia), posttraumatic stress disorder, and other anxiety disorders (e.g., panic disorder, specific phobias, generalized anxiety disorder) than older men aged 65 to 74 and 75 to 84.

Table 55.1a–3. Educational Attainment of the Population 55 Years and Over by Sex and Age: 2019 (Numbers in thousands. Civilian noninstitutionalized population.[a])

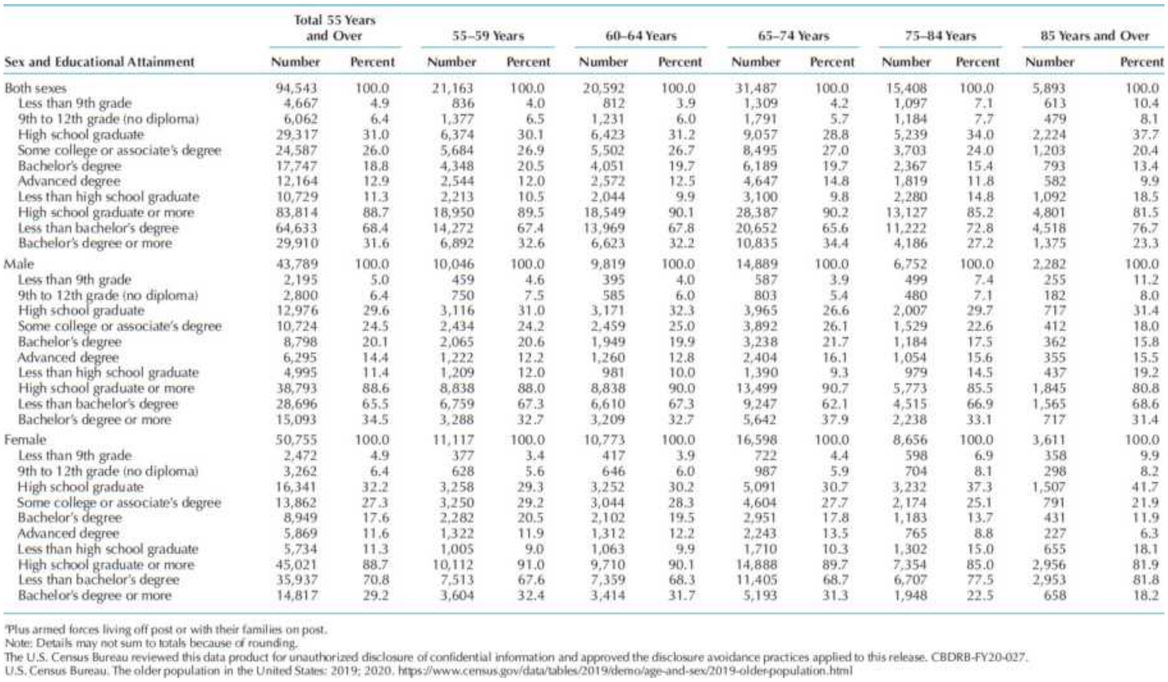

	Total 55 Years and Over		55–59 Years		60–64 Years		65–74 Years		75–84 Years		85 Years and Over	
Sex and Educational Attainment	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Both sexes	94,543	100.0	21,163	100.0	20,592	100.0	31,487	100.0	15,408	100.0	5,893	100.0
Less than 9th grade	4,667	4.9	836	4.0	812	3.9	1,309	4.2	1,097	7.1	613	10.4
9th to 12th grade (no diploma)	6,062	6.4	1,377	6.5	1,231	6.0	1,791	5.7	1,184	7.7	479	8.1
High school graduate	29,317	31.0	6,374	30.1	6,423	31.2	9,057	28.8	5,239	34.0	2,224	37.7
Some college or associate's degree	24,587	26.0	5,684	26.9	5,502	26.7	8,495	27.0	3,703	24.0	1,203	20.4
Bachelor's degree	17,747	18.8	4,348	20.5	4,051	19.7	6,189	19.7	2,367	15.4	793	13.4
Advanced degree	12,164	12.9	2,544	12.0	2,572	12.5	4,647	14.8	1,819	11.8	582	9.9
Less than high school graduate	10,729	11.3	2,213	10.5	2,044	9.9	3,100	9.8	2,280	14.8	1,092	18.5
High school graduate or more	83,814	88.7	18,950	89.5	18,549	90.1	28,387	90.2	13,127	85.2	4,801	81.5
Less than bachelor's degree	64,633	68.4	14,272	67.4	13,969	67.8	20,652	65.6	11,222	72.8	4,518	76.7
Bachelor's degree or more	29,910	31.6	6,892	32.6	6,623	32.2	10,835	34.4	4,186	27.2	1,375	23.3
Male	43,789	100.0	10,046	100.0	9,819	100.0	14,889	100.0	6,752	100.0	2,282	100.0
Less than 9th grade	2,195	5.0	459	4.6	395	4.0	587	3.9	499	7.4	255	11.2
9th to 12th grade (no diploma)	2,800	6.4	750	7.5	585	6.0	803	5.4	480	7.1	182	8.0
High school graduate	12,976	29.6	3,116	31.0	3,171	32.3	3,965	26.6	2,007	29.7	717	31.4
Some college or associate's degree	10,724	24.5	2,434	24.2	2,459	25.0	3,892	26.1	1,529	22.6	412	18.0
Bachelor's degree	8,798	20.1	2,065	20.6	1,949	19.9	3,238	21.7	1,184	17.5	362	15.8
Advanced degree	6,295	14.4	1,222	12.2	1,260	12.8	2,404	16.1	1,054	15.6	355	15.5
Less than high school graduate	4,995	11.4	1,209	12.0	981	10.0	1,390	9.3	979	14.5	437	19.2
High school graduate or more	38,793	88.6	8,838	88.0	8,838	90.0	13,499	90.7	5,773	85.5	1,845	80.8
Less than bachelor's degree	28,696	65.5	6,759	67.3	6,610	67.3	9,247	62.1	4,515	66.9	1,565	68.6
Bachelor's degree or more	15,093	34.5	3,288	32.7	3,209	32.7	5,642	37.9	2,238	33.1	717	31.4
Female	50,755	100.0	11,117	100.0	10,773	100.0	16,598	100.0	8,656	100.0	3,611	100.0
Less than 9th grade	2,472	4.9	377	3.4	417	3.9	722	4.4	598	6.9	358	9.9
9th to 12th grade (no diploma)	3,262	6.4	628	5.6	646	6.0	987	5.9	704	8.1	298	8.2
High school graduate	16,341	32.2	3,258	29.3	3,252	30.2	5,091	30.7	3,232	37.3	1,507	41.7
Some college or associate's degree	13,862	27.3	3,250	29.2	3,044	28.3	4,604	27.7	2,174	25.1	791	21.9
Bachelor's degree	8,949	17.6	2,282	20.5	2,102	19.5	2,951	17.8	1,183	13.7	431	11.9
Advanced degree	5,869	11.6	1,322	11.9	1,312	12.2	2,243	13.5	765	8.8	227	6.3
Less than high school graduate	5,734	11.3	1,005	9.0	1,063	9.9	1,710	10.3	1,302	15.0	655	18.1
High school graduate or more	45,021	88.7	10,112	91.0	9,710	90.1	14,888	89.7	7,354	85.0	2,956	81.9
Less than bachelor's degree	35,937	70.8	7,513	67.6	7,359	68.3	11,405	68.7	6,707	77.5	2,953	81.8
Bachelor's degree or more	14,817	29.2	3,604	32.4	3,414	31.7	5,193	31.3	1,948	22.5	658	18.2

[a] Plus armed forces living off post or with their families on post.

Note: Details may not sum to totals because of rounding.

The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. CBDRB-FY20-027.

U.S. Census Bureau. The older population in the United States: 2019; 2020. https://www.census.gov/data/tables/2019/demo/age-and-sex/2019-older-population.html

Table 55.1a–4.
Earnings of Full-Time, Year-Round Workers 55 Years and Over by Sex and Age: 2018[a] (Numbers in thousands. Civilian noninstitutionalized population.[b])

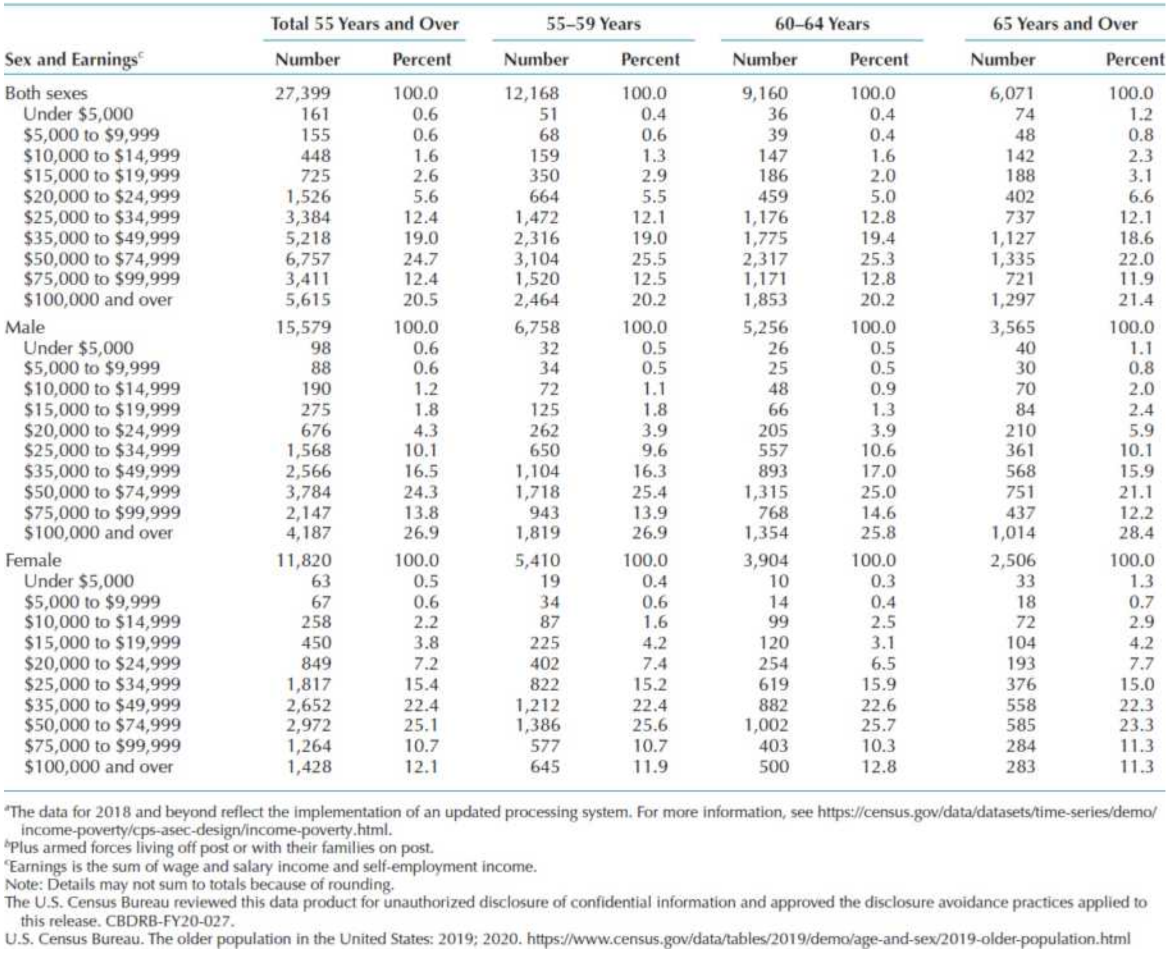

	Total 55 Years and Over		55–59 Years		60–64 Years		65 Years and Over	
Sex and Earnings[c]	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Both sexes	27,399	100.0	12,168	100.0	9,160	100.0	6,071	100.0
Under $5,000	161	0.6	51	0.4	36	0.4	74	1.2
$5,000 to $9,999	155	0.6	68	0.6	39	0.4	48	0.8
$10,000 to $14,999	448	1.6	159	1.3	147	1.6	142	2.3
$15,000 to $19,999	725	2.6	350	2.9	186	2.0	188	3.1
$20,000 to $24,999	1,526	5.6	664	5.5	459	5.0	402	6.6
$25,000 to $34,999	3,384	12.4	1,472	12.1	1,176	12.8	737	12.1
$35,000 to $49,999	5,218	19.0	2,316	19.0	1,775	19.4	1,127	18.6
$50,000 to $74,999	6,757	24.7	3,104	25.5	2,317	25.3	1,335	22.0
$75,000 to $99,999	3,411	12.4	1,520	12.5	1,171	12.8	721	11.9
$100,000 and over	5,615	20.5	2,464	20.2	1,853	20.2	1,297	21.4
Male	15,579	100.0	6,758	100.0	5,256	100.0	3,565	100.0
Under $5,000	98	0.6	32	0.5	26	0.5	40	1.1
$5,000 to $9,999	88	0.6	34	0.5	25	0.5	30	0.8
$10,000 to $14,999	190	1.2	72	1.1	48	0.9	70	2.0
$15,000 to $19,999	275	1.8	125	1.8	66	1.3	84	2.4
$20,000 to $24,999	676	4.3	262	3.9	205	3.9	210	5.9
$25,000 to $34,999	1,568	10.1	650	9.6	557	10.6	361	10.1
$35,000 to $49,999	2,566	16.5	1,104	16.3	893	17.0	568	15.9
$50,000 to $74,999	3,784	24.3	1,718	25.4	1,315	25.0	751	21.1
$75,000 to $99,999	2,147	13.8	943	13.9	768	14.6	437	12.2
$100,000 and over	4,187	26.9	1,819	26.9	1,354	25.8	1,014	28.4
Female	11,820	100.0	5,410	100.0	3,904	100.0	2,506	100.0
Under $5,000	63	0.5	19	0.4	10	0.3	33	1.3
$5,000 to $9,999	67	0.6	34	0.6	14	0.4	18	0.7
$10,000 to $14,999	258	2.2	87	1.6	99	2.5	72	2.9
$15,000 to $19,999	450	3.8	225	4.2	120	3.1	104	4.2
$20,000 to $24,999	849	7.2	402	7.4	254	6.5	193	7.7
$25,000 to $34,999	1,817	15.4	822	15.2	619	15.9	376	15.0
$35,000 to $49,999	2,652	22.4	1,212	22.4	882	22.6	558	22.3
$50,000 to $74,999	2,972	25.1	1,386	25.6	1,002	25.7	585	23.3
$75,000 to $99,999	1,264	10.7	577	10.7	403	10.3	284	11.3
$100,000 and over	1,428	12.1	645	11.9	500	12.8	283	11.3

[a] The data for 2018 and beyond reflect the implementation of an updated processing system. For more information, see https://census.gov/data/datasets/time-series/demo/income-poverty/cps-asec-design/income-poverty.html.

[b] Plus armed forces living off post or with their families on post.

[c] Earnings is the sum of wage and salary income and self-employment income.

Note: Details may not sum to totals because of rounding.

The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. CBDRB-FY20-027.

U.S. Census Bureau. The older population in the United States: 2019; 2020. https://www.census.gov/data/tables/2019/demo/age-and-sex/2019-older-population.html

Furthermore, psychiatric symptoms that do not meet the traditional criteria for psychiatric disorders (that had been developed for use in younger adults) but are nonetheless quite disabling are much more common in older adults than the ECA study suggests. For example, a diagnosis of major depression requires the presence of at least five symptoms from a list of nine symptoms included in the DSM-5. If a person has four of those symptoms, they would not meet the criteria for major depression but may still be quite disabled. Such subsyndromal depression is much more common than major depression in older adults, yet the patient may not receive the necessary therapeutic attention because the patient was not diagnosed with major depression.

There is expected to be a disproportionately greater increase in the numbers of mentally ill older adults than that of the overall older population. There are several reasons for this demographic change. Excess mortality (from suicide, nonintentional injuries, or physical comorbidity) in early life among seriously mentally ill patients such as those with schizophrenia, bipolar disorder, unipolar depression, and substance dependence is likely an important reason for a lower prevalence of nondementia psychiatric disorders in older compared with younger adults. With an improved overall standard of living, better treatment for physical disorders, and more effective treatments for psychiatric illnesses, the high mortality in mentally ill young adults should begin to decrease, resulting in a greater number of those who will reach old age. Unfortunately, this scenario has not materialized yet. As a matter of fact, the longevity gap between people with schizophrenia (and other serious mental illnesses) and the general population has increased in recent decades. The likely reason is that persons with mental illnesses have not been able to benefit from the health and medical advances that have led to increased longevity in the community at large. Also, with the general increase in the numbers of people older than 65 years, there will be more older adults in the community at risk of developing late-onset psychiatric disorders. Furthermore, aging baby boomers are expected to have a higher risk of depression, anxiety disorders, and substance use disorders than those born before World War II. For all these reasons, it is estimated that the number of older mentally ill adults in the population will increase by at least 10% from 2000 to 2030—that is, it will be about 22%. Thus, the numbers of psychiatrically ill older adults will rise by 275%—from 4 million in 1970 to 15 million in 2030, while comparable numbers for those aged 30 to 44 years will increase by only 67%—from over 9 million to 15 million. These estimates are educated crude guesses; nonetheless, they provide some understanding of the magnitude of the challenge facing psychiatry.

Table 55.1a–5.

Households by Type and Age of Householder 55 Years and Over: 2019 (Numbers in thousands. Civilian noninstitutionalized population.[a])

Household Type	Total 55 Years and Over		55–59 Years		60–64 Years		65–74 Years		75–84 Years		85 Years and Over		65 Years and Over	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Total households	58,328	100.0	12,099	100.0	12,073	100.0	19,681	100.0	10,222	100.0	4,253	100.0	34,156	100.0
Family households[b]	34,095	58.5	8,032	66.4	7,630	63.2	11,880	60.4	5,206	50.9	1,347	31.7	18,432	54.0
Married-couple	27,928	47.9	6,389	52.8	6,328	52.4	10,096	51.3	4,247	41.6	867	20.4	15,210	44.5
Male householder[c]	1,590	2.7	451	3.7	399	3.3	443	2.3	190	1.9	107	2.5	740	2.2
Female householder[c]	4,576	7.8	1,191	9.8	903	7.5	1,341	6.8	768	7.5	373	8.8	2,482	7.3
Nonfamily households	24,233	41.5	4,067	33.6	4,443	36.8	7,801	39.6	5,016	49.1	2,906	68.3	15,723	46.0
Male householder	9,659	16.6	1,999	16.5	2,071	17.2	3,238	16.5	1,524	14.9	826	19.4	5,588	16.4
Female householder	14,575	25.0	2,068	17.1	2,372	19.6	4,562	23.2	3,493	34.2	2,080	48.9	10,135	29.7

[a] Plus armed forces living off post or with their families on post.
[b] Households in which at least one member is related to the person who owns or rents the occupied housing unit (householder).
[c] No spouse present.
Note: Details may not sum to totals because of rounding.
The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. CBDRB-FY20-027.
U.S. Census Bureau. The older population in the United States: 2019. 2020. https://www.census.gov/data/tables/2019/demo/age-and-sex/2019-older-population.html

Table 55.1a–6.

Total Money Income of Households by Type and Age of Householder 55 Years and Over: 2018[a] (Numbers in thousands. Civilian noninstitutionalized population.[b])

Household Type and Money Income[c]	Total 55 Years and Over		55–59 Years		60–64 Years		65 Years and Over	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Total households	58,328	100.0	12,099	100.0	12,073	100.0	34,156	100.0
Under $5,000	2,097	3.6	442	3.7	451	3.7	1,203	3.5
$5,000–$9,999	1,937	3.3	310	2.6	416	3.4	1,211	3.5
$10,000–$14,999	3,612	6.2	523	4.3	615	5.1	2,475	7.2
$15,000–$19,999	3,696	6.3	432	3.6	556	4.6	2,709	7.9
$20,000–$24,999	3,206	5.5	424	3.5	479	4.0	2,303	6.7
$25,000–$34,999	5,968	10.2	783	6.5	1,024	8.5	4,162	12.2
$35,000–$49,999	7,262	12.5	1,210	10.0	1,263	10.5	4,789	14.0
$50,000–$74,999	9,646	16.5	2,005	16.6	1,988	16.5	5,653	16.6
$75,000–$99,999	6,233	10.7	1,545	12.8	1,423	11.8	3,265	9.6
$100,000 and over	14,671	25.2	4,426	36.6	3,858	32.0	6,387	18.7
Family households[d]	34,095	100.0	8,032	100.0	7,630	100.0	18,432	100.0
Under $5,000	652	1.9	130	1.6	157	2.1	366	2.0
$5,000–$9,999	415	1.2	87	1.1	71	0.9	257	1.4
$10,000–$14,999	698	2.0	146	1.8	170	2.2	382	2.1
$15,000–$19,999	853	2.5	166	2.1	155	2.0	531	2.9
$20,000–$24,999	1,127	3.3	166	2.1	165	2.2	796	4.3
$25,000–$34,999	2,708	7.9	392	4.9	446	5.8	1,870	10.1
$35,000–$49,999	4,054	11.9	674	8.4	711	9.3	2,669	14.5
$50,000–$74,999	6,319	18.5	1,263	15.7	1,319	17.3	3,736	20.3
$75,000–$99,999	4,816	14.1	1,184	14.7	1,083	14.2	2,549	13.8
$100,000 and over	12,454	36.5	3,824	47.6	3,354	43.9	5,276	28.6
Nonfamily households	24,233	100.0	4,067	100.0	4,443	100.0	15,723	100.0
Under $5,000	1,445	6.0	312	7.7	295	6.6	838	5.3
$5,000–$9,999	1,521	6.3	223	5.5	344	7.8	954	6.1
$10,000–$14,999	2,915	12.0	377	9.3	445	10.0	2,093	13.3
$15,000–$19,999	2,844	11.7	266	6.5	400	9.0	2,178	13.8
$20,000–$24,999	2,080	8.6	258	6.4	314	7.1	1,507	9.6
$25,000–$34,999	3,260	13.5	391	9.6	578	13.0	2,291	14.6
$35,000–$49,999	3,208	13.2	536	13.2	552	12.4	2,120	13.5
$50,000–$74,999	3,327	13.7	741	18.2	669	15.1	1,917	12.2
$75,000–$99,999	1,417	5.8	361	8.9	341	7.7	716	4.6
$100,000 and over	2,217	9.1	602	14.8	504	11.4	1,110	7.1

[a] The data for 2018 and beyond reflect the implementation of an updated processing system. For more information, see https://census.gov/data/datasets/time-series/demo/income-poverty/cps-asec-design/income-poverty.html.
[b] Plus armed forces living off post or with their families on post.
[c] Total money income is the sum of wages and salaries, net income from self-employment, and income other than earnings.
[d] Households in which at least one member is related to the person who owns or rents the occupied housing unit (householder).
Note: Details may not sum to totals because of rounding.
The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. CBDRB-FY20-027.
U.S. Census Bureau. The older population in the United States: 2019; 2020. https://www.census.gov/data/tables/2019/demo/age-and-sex/2019-older-population.html

It is also to be noted that not only will the numbers of older adults with mental illness increase dramatically over the next few decades, but also the types of patients we see will change to some extent. For several disorders, there will be higher proportions of patients at the two ends of the severity spectrum. For example, expected advances in the ability to detect Alzheimer disease (AD) in earlier stages and to decrease the rate of decline via the use of effective neuroprotective agents may result in greater numbers of patients who are more mildly affected and who will present different management problems from patients with advanced dementia. Issues regarding employment, financial planning, driving, home care, and daycare will become more prominent. At the same time, reduced mortality from severe forms of physical illness (e.g., strokes, heart disease) will lead to increased numbers of severely demented patients with comorbid medical conditions living longer, raising different clinical, research, and ethical issues.

GERIATRIC PSYCHIATRY AS A SUBSPECIALTY

There are unique diagnostic and therapeutic considerations in the management of older adults with mental illnesses. Just as children are not small adults, older adults are not merely chronologically more aged people. There are important bio-psycho-social differences that have an impact on the diagnosis, course, and treatment of mental (and physical) illnesses in different age groups. Because of age-related changes in the functioning of the liver, kidneys and brain, older people may only tolerate lower doses of most medications and are at a greater risk of developing a number of side effects. They also have special needs for psychosocial treatments (e.g., they are less likely to be living with their families and more likely to have financial difficulties). Additionally, they may have major physical illnesses and disabilities, visual and hearing impairments, and even less access to health care than younger adults (e.g., due to transportation issues and financial limitations). As a result, fewer than half of older Americans with mental illness receive the needed mental health services. They face the dual stigma of mental illness and aging. These problems are further compounded by therapeutic pessimism (and even nihilism) fostered by rampant societal as well as professional ageism.

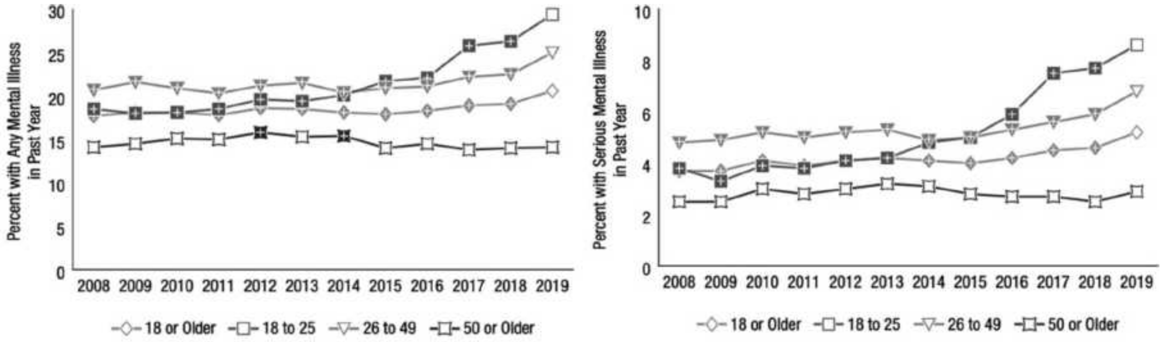

+ Difference between this estimate and the 2019 estimate is statistically significant at the .05 level. + Difference between this estimate and the 2019 estimate is statistically significant at the .05 level.

FIGURE 55.1a–2. Prevalence of psychiatric disorders over time by age. Note: Any mental illness (AMI) is defined as a mental, behavioral, or emotional disorder. AMI can vary in impact, ranging from no impairment to mild, moderate, and even severe impairment (e.g., individuals with serious mental illness as defined below). Serious mental illness (SMI) is defined as a mental, behavioral, or emotional disorder resulting in serious functional impairment, which substantially interferes with or limits one or more major life activities. The burden of mental illnesses is particularly concentrated among those who experience disability due to SMI. (From Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: Results from the 2019 National Survey on Drug Use and Health [HHS Publication No. PEP20-07-01-001]. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2020. https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR1PDFW090120.pdf)

Old-age psychiatry has had a long and distinguished tradition in the United Kingdom and other countries in Europe. In recent years, it has started to come of age in the United States too. Geriatric psychiatry became the second subspecialty of psychiatry (after child psychiatry) recognized by the U.S. Accreditation Council for Graduate Medical Education and by the American Board of Psychiatry and Neurology. There are several geriatric psychiatry organizations, such as the American Association for Geriatric Psychiatry and the International Psychogeriatric Association. There are also several peer-reviewed journals devoted to geriatric psychiatry, including Aging & Mental Health, the American Journal of Geriatric Psychiatry, the International Psychogeriatrics, the International Journal of Geriatric Psychiatry, the Journal of Geriatric Psychiatry and Neurology, Psychogeriatrics, and Psychology & Aging. In basic sciences, there is a growing interest in the molecular biology of aging and the neuroplasticity of aging.

Table 55.1a–7.
Overview of the Education and Training of Professionals in Geriatrics in the United States

Discipline	Progress
Geriatricians	Geriatricians are doctors that specialize in older adults. As of 2016, the number of geriatricians with up-to-date certifications had remained flat at 7,000 for the past 10 years. Approximately 300 geriatricians are trained each year, and many fellowship positions are not filled.
Geriatric Psychiatrists	There were less than 2,000 certified geriatric psychiatrists, and many fellowship positions remain unfilled each year.
Gerontology Advanced Practice Registered Nurses (APRN, NP, and CNS)	In 2013, the certification in gerontology for NPs and CNS was phased out and combined with adult primary and acute care certifications. In turn, the number of adult-gerontology certified APRNs significantly increased to 12,000.
Registered Nurses Certified in Gerontology	From 2013 to 2014, the number of RNs certified in gerontology increased by 1.7%. In 2014, 7,874 RNs were certified out of 3.1 million RNs nationwide.
Gerontologic Social Workers	There is no certification in gerontology for social workers. In 2009, 1,318 individuals in a Master of Social Work program specialized in aging.
Certified Geriatric Pharmacists	Between 2010 and 2015, the number of pharmacists certified in geriatrics increased from 1,210 to 2,158 (a 78% increase).
Geropsychology	In 2010, the American Psychological Association recognized professional geropsychology as a special area of psychology and started accrediting a geropsychology postdoctoral program. In 2014, the American Board of

	Professional Psychology added geropsychology as a boarded specialty.
Geriatric Physical Therapy	Between 2010 and 2015, the number of physical therapists certified in geriatrics increased from 1,006 to 1,936 (a 92% increase). The number of residency programs expanded from 6 to 13.
Gerontology-Certified Occupational Therapists	As of 2015, there were only 18 occupational therapists certified in gerontology.

Reprinted from Substance Abuse and Mental Health Services Administration. Older adults living with serious mental illness: the state of the behavioral health workforce. 2019. https://store.samhsa.gov/sites/default/files/d7/priv/pep19-olderadults-smi.pdf; originally adapted from Warshaw GA, Bragg EJ. The essential components of quality geriatric care. ***Generations***. 2016;40(1):28–37.

Unfortunately, the increasing need for a geriatric mental health workforce is being met with a diminishing availability of experts in the field. There is a severe shortage of workforce trained to provide the necessary mental and physical health care for the growing numbers of older adults (Table 55.1a–7). Despite an urgent need for a trained workforce with expertise in geriatric mental disorders, the number of American Board-certified geriatric psychiatrists, who have maintained their subspecialty certification, has dropped from over 2,200 a decade ago to fewer than 1,800 today. This number is projected to decrease further to 1,650 by the year 2030, amounting to less than one geriatric psychiatrist for every 6,000 older adults with mental health and substance use disorders. Out of more than 40,000 psychiatrists in the United States, only about 5,000 list “geriatric or gerontologic psychiatry” as one of their three primary areas of interest.

The Health Resources and Services Administration (HRSA) and the Substance Abuse and Mental Health Services Administration (SAMHSA) have reported that we have passed the “tipping point” of an inevitable shortfall in geriatrics-trained providers and highlighted the urgency of expanding and strengthening the geriatrics health care workforce. Yet, little is being done to address this challenge. Already geriatric psychiatric services are in high demand, and specialists are in short supply. The trainee pool is a critical issue. Currently, there are over 100 geriatric psychiatry fellowship training slots available nationwide, but the number of slots filled has declined from more than 80% in 2000 to under 50% today. Psychiatry residents, often struggling with large student loans, are drawn away from geriatric psychiatric fellowship, at least partly because of financial concerns. Rates of Medicare reimbursement are decreasing, and the reimbursement system as a whole makes psychiatric care difficult to access for older patients. The dual stigma against aging and mental illness also makes geriatric psychiatry less attractive to many trainees.

COMMON MENTAL ILLNESSES IN OLDER PEOPLE

A brief description of a few of the mental disorders of late life is provided below. A detailed description of each illness, its causes, and treatments are provided in subsequent chapters of the book.

Alzheimer Disease (AD) and Other Dementias

AD is the most common form of dementia. It is a progressive brain disorder that results in cognitive deficits such as memory loss, impaired thinking, and abnormal behavior. The number of older persons with AD will increase substantially over the next three decades for several reasons. The prevalence of dementia doubles every 5 years after the age of 60. Thus, 1% of the people aged 60 to 64 years are affected; 2% of those aged 65 to 69, 4% of those between 70 and 74, 8% of those 75 to 79 years old, 16% of those aged 80 to 84, and 30% to 45% of those age 85 years and older. These figures provide a reasonable guide to the number of patients with AD and other dementias anticipated over the next three decades. From a different perspective, one of the “hottest” areas of research is developing drugs that delay the progression of brain degeneration in patients with dementia. AD and other dementias are seriously under-recognized at present. In various studies, investigators have found that primary care practitioners recognize only a minority of the AD patients identified by researchers using a structured diagnostic interview. We can expect greater awareness of the need for early detection of AD among professionals as well as among the lay public. Currently, many researchers are working on developing biologic markers to detect early AD. Reliable and valid biologic markers will diminish the present dependence on clinical (largely subjective) judgment in diagnosing early AD and enhance the rate of detection of AD even before obvious dementia has set in. Furthermore, as new therapies for the treatment of very early AD (even before dementia is obvious) become available in the coming years, improved recognition of patients before they manifest the disease clinically will increase the number of identified cases in the population.

Interestingly, there are also factors that could serve to reduce the incidence of AD in future cohorts of aged individuals. One such factor is education. The prevalence of AD is lower among those with more years of formal education or greater “cognitive reserve.” There will be greater numbers of older baby boomers with a high school diploma compared to the current cohort of older adults. Education also has a variety of secondary effects that may reduce the frequency of dementia. For example, head injury has been identified in several studies as a risk factor for AD, and head trauma is less common among those with more education who are more likely to wear safety belts, use protective headgear in contact sports, and avoid driving while intoxicated. Such factors associated with a reduced risk of AD point to the way in which we may be able to lower the incidence of AD by proactive societal actions.

Neuropsychiatric behavioral disturbances such as agitation, psychosis, and depression are common in patients with AD and are frequent causes of institutionalization. About a third of the patients with AD have depressive symptoms, whereas the prevalence of psychosis is about 50%. Agitation is even more frequent, especially in later stages. A number of intervention studies of pharmacologic as well as psychosocial treatments have been performed for these neuropsychiatric symptoms, with variable results.

Dementias other than AD, such as Lewy body dementia, frontotemporal dementia, and vascular dementia, are also receiving increasing attention, mainly in the areas of brain imaging and neuropathology. Again, behavioral disturbances and psychotic symptoms are common phenomena in these patients.

Anxiety Disorders

Anxiety disorders are the most common disorders among older adults, with the exception of dementia. There are a number of differences between younger versus older adults in terms of the psychopathology of these disorders. For example, older phobic patients tend to have a fear of situations or inanimate stimuli, such as heights or lightning, whereas younger individuals are more likely to have a fear of animals. Because of physical and psychosocial changes associated with aging, it might be difficult to distinguish between phobias or agoraphobia and nonpathologic avoidance behaviors in older patients, such as avoiding leaving home due to arthritis or fear of crime. Older adults might be more likely than younger patients to manifest a mixture of anxiety and depression. Such comorbidity could be a result of the cooccurrence of two common conditions or could reflect one being an epiphenomenon or a complication of the other. A syndrome of "mixed anxiety– depressive disorder" might have both heuristic and clinical utility for older patients.

Depression

Numerous studies have shown that depression goes undiagnosed in sizable proportions of older adults. This is due, in part, to frequent denial of depression with its associated social stigma and attributing the symptoms of depression to "normal" aging or to physical illnesses. In the past three decades, some SSRIs have become nearly household names, leading to more recognition and acceptance of depression in the public at large. This trend will continue, resulting in reduced social stigma associated with depression. Another important reason for the underdiagnosis of depression in the primary care setting, where most older persons are seen, is the inadequacy of geriatric mental health training at the medical school level. With improvements in medical education, this situation will change in the years to come, resulting in greater detection of depression in older adults.

Major depression manifests somewhat differently in older compared to young adults. Physical symptoms such as a change in appetite and body weight, constipation, and sexual dysfunction are more common in older adults; on the other hand, the physical symptoms in older adults could be a result of an undiagnosed physical illness (e.g., pancreatic cancer). While the point prevalence of subsyndromal depression in younger adults is comparable to or lower than that of major depression, the prevalence of subsyndromal depression (18%) is three times that of major depression (6%) in people aged 65 years or older. Thus, the predominant form of depression in older adults appears to be subsyndromal depression. Notably, subsyndromal depression is also associated with significant increases in psychosocial impairment in major domains of everyday functioning.

A number of large epidemiologic investigations in several different countries have shown that the risk of depression, as well as depression-related suicides, alcoholism, and substance abuse, is significantly higher in the baby boomers than in those born before World War II. The reasons for this type of "cohort effect" are not known but could be related to several different factors: Genetic factors such as new mutations among recent generations. A number of environmental factors may also contribute to greater stress and increased risk of depression among the baby boomers; these include changing family structure with nuclear families, higher divorce rates, changing roles for working women, greater urbanization, and a decrease in "social connectedness" due to geographic mobility and job changes. Detection of depression in older age has likely increased due to improved psychometric tests, the development of the DSM, and reduced societal stigma around seeking treatment for mental illness. Finally, depression is known to be much more common among those with physical illnesses than in physically healthy people. This observation leads to the development of preventive interventions in individuals with physical disorders such as heart disease and cancer.

Schizophrenia and Other Psychotic Disorders

The century-old concept of schizophrenia, based on Kraepelin's definition of dementia praecox, is that this is a disorder with onset in early life—that is, adolescence and early adulthood ("praecox") and progressive deterioration ("dementia"). Yet studies in older patients show that schizophrenia can have onset in middle age (late-onset schizophrenia). Late-onset schizophrenia is more common in women and is usually of the paranoid type but may present initially as brief psychotic disorder or delusional disorder. In terms of its course in late life, schizophrenia is typically not associated with a dementing course. Indeed, a small minority of patients with schizophrenia have a sustained remission of their illness in old age.

Psychotic symptoms may manifest for the first time after age 65. This is generally not schizophrenia but is best labeled as very late-onset schizophrenia-like psychosis.

Substance Use Disorders

As better treatments become available for substance use disorders (including alcoholism), the focus on the war on drugs during the coming decades will shift from punishment to treatment. Studies using currently available diagnostic criteria for substance abuse have underestimated the prevalence of drug abuse in older adults because these criteria were developed and validated only in young and middle-aged samples. For example, DSM-5 criteria for substance use disorders include the development of tolerance to the effects of that substance (e.g., a given amount of a stimulant produces less stimulation than it used to), resulting in a need for increased consumption over time. Yet, liver and kidney damage common in old age would produce higher concentrations of that substance in the blood, while age-associated brain changes would serve to enhance the effects of the substance on the brain. Hence, there is likely to be decreased tolerance to that substance in an older individual (i.e., a given amount produces more effect than it used to), leading to a need for decreased consumption. This older person, although still dependent on the substance, may not meet the "official" criteria for diagnosis. One DSM-5 criterion for substance use disorders is a report of being repeatedly unable to carry out major obligations at work, school, or home, such as absences or poor work performance, suspensions or expulsions from school, and neglect of children. Such adverse consequences are much less likely to occur in older adults who often live alone and are unemployed or retired, resulting in the underestimation of substance use disorders.

It is generally believed that aging is associated with reductions in and changing patterns of drug use (i.e., reductions in illicit drug use). One theory posits that factors associated with "maturation" associated with aging processes, increased numbers of individuals who have been dependent on illicit drugs would become abstinent as they age. This is, however, not entirely supported by empirical data, which suggest that there may be increasing numbers of older addicts and that persons who have been dependent for more than 5 years do not become abstinent as they age. An alternative theory posits that those with the most severe substance use behaviors would die prematurely and, therefore, never reach old age. In addition to increased numbers of aging "life-long" drug abusers, we are also likely to see greater numbers of individuals who, after decades of abstinence, return to drug or alcohol problems. Psychosocial stressors, including isolation, various types of losses, loneliness, and the onset of illness, have been related to late-onset drug use.

There is greater use and acceptance of psychotropic medications among the baby boomers than among those born before World War II. In contrast to younger substance abusers who most often abuse illicit drugs, substance abuse problems in older adults may result from misuse or abuse of over-the-counter (OTC) and/or prescription drugs. Older individuals use prescription drugs three times as frequently as the general population, and the use of OTC medications by this group is even more extensive. The estimated expenditure on prescription drugs by older adults in the United States is $15 billion annually, a fourfold greater per capita expenditure on medications compared to younger individuals. Psychoactive drug use (most often sedative-hypnotic, anti-anxiety, and analgesic) is particularly problematic, including long-acting drugs not recommended for older adults and especially those at risk of falls and inappropriately high doses of benzodiazepines.

Comorbid physical, as well as psychiatric disorders appear to be important predictors of prescription drug use. As greater numbers of people with chronic physical and psychiatric disorders survive into older age, there will be an increased prevalence of substance use disorders in the older population. Moreover, the deceptive marketing and overprescription of narcotics (e.g., Dilaudid, fentanyl, methadone, Oxycontin, oxycodone) for chronic pain have led to a drastic increase in their use and abuse in older adults.

The jury is still out whether the legalization of marijuana for medical use or adult recreational use in a growing number of states may negatively affect the health of older adults. Marijuana is the most commonly abused drug in the United States and around the world, but it is likely that by legalizing marijuana, the number of marijuana users would increase.

POSITIVE SIDE OF AGING

Contrary to the concept that aging is synonymous with progressive and inevitable deterioration of the brain and mind, a number of recent studies across the world show that psychological well-being and happiness increase with age, especially after middle age. Some studies have found that subjective feeling of well-being (a sense of contentedness) is high at the beginning of our adult life, then declines progressively until it hits rock bottom in middle age, producing the so-called mid-life crisis, but then starts to rise again. Other investigations report a progressive rise in well-being from age 20 to 80+. Interestingly, the level of happiness in the early 80s is similar to that in the early 20s. The increase in happiness after middle age occurs even as physical, cognitive, and fertility functions decline. This finding has been reported by researchers from several different countries and cultures. Notably, perfect physical health is neither necessary nor sufficient to feel happy as we grow older. Research shows a paradox of aging: physical health and some aspects of cognitive function (for instance, memory for names and faces) decline with age, while happiness, mental health, and management of interpersonal relationships improve.

What are the possible explanations for this counterintuitive result? Suggested factors include changes in attitude such as greater acceptance of one's physical limitations, contentedness with past accomplishments in life, reduced preoccupation with peer pressure, a more realistic appraisal of one's own strengths and limitations, and increased emotional stability. Intriguingly, a recent study, based on the reports by the individual caretakers of chimpanzees and orangutans in zoos and other protected areas, found an increase in these great apes' well-being after middle age. Thus, better well-being in older age might be rooted, in part, in biologic changes in the brain.

Several aspects of aging can be aided significantly by improved health care, the greater societal value placed on older adults, and a more positive conceptualization of old age. Most psychiatrists today would not agree with Freud's assertion in 1924 that near or above the age of 50, the elasticity of mental processes is lacking, and, therefore, older individuals are "no longer educable." The so-called problems of aging should become opportunities for a better understanding and improvement of the entire life process.

The aging of the population is generally viewed with concern and alarm about its impact on social security, healthcare (including Medicare), housing, jobs for younger generations, etc. Yet, this is also a grand opportunity to make positive use of the growing older adult population so that the new era, dubbed inappropriately by some as "Silver Tsunami," can be labeled the "Golden Wave." The most important advance in neuroscience in the past two decades has been the demonstration of neuroplasticity of aging—the fact that brain growth and development continue into old age. Furthermore, such a positive outcome is related not so much to the genes we inherit from our parents but to our behavior, attitude, and environment. With appropriate physical activities, targeted nutrition, and psychosocial stimulation along with necessary healthcare and social support, successful aging need not be a dream but can become a widespread reality. As the quality of life of older adults improves, they become more productive and contribute more to the rest of society, and their healthcare costs go down. The seniors who learn and teach can be wonderful and wise mentors for younger generations. Older citizens can lead efforts at social entrepreneurship in a collaborative manner.

One of the positive traits that have received increasing attention in recent years is wisdom. Wisdom is comprised of several specific components—***viz***., prosocial behaviors (empathy and compassion), emotional regulation, self-reflection, acceptance of diverse perspectives, decisiveness, and possibly, spirituality. Research has also shown a neurobiologic basis of wisdom, mainly involving the prefrontal cortex and limbic striatum. A number of studies have found higher levels of several components of wisdom in older compared to younger adults. Investigations have also demonstrated that intergenerational activities are useful for both generations. A meta-analysis of 57 randomized controlled trials reported that about half of these trials resulted in significant improvement in a component of wisdom, such as empathy and compassion, emotional regulation, and spirituality, with moderate to large effect sizes. Importantly, wisdom is associated with greater well-being, happiness, resilience, and better health. Recent research has shown an opposite relationship between loneliness and wisdom, especially its compassion component. Whereas loneliness is associated with worse physical, mental, and cognitive functioning, the reverse is true for wisdom and compassion. Thus, wisdom-enhancing interventions can potentially curtail loneliness in older adults and thus contribute to better health and, perhaps, greater longevity.

SUCCESSFUL COGNITIVE AND EMOTIONAL AGING

Research on aging has typically focused on disease and disability. However, during the past two decades, this focus has slowly begun to expand to incorporate studies of “successful or healthy aging.” Understanding why some people seem to adapt to challenges they encounter in late life when others cannot facilitate the development of interventions and policies geared toward older populations. Yet, there are many times as many studies of aging-related disorders as there are on successful aging. One problem in research on successful aging is that it involves multiple components (e.g., physical health, psychological well-being) that extend beyond longevity, and as a result, there are numerous criteria for defining it but no consensus. In 1987, Rowe and Kahn proposed that successful aging involved three components: (1) absence/low probability of disease and disease-related disability, (2) high cognitive and physical functioning, and (3) active engagement with life. According to this model, however, many cognitively intact and socially active older adults would not be categorized as aging successfully because of the presence of chronic physical diseases and disabilities. Also, such criteria fail to take into account the perspectives of older adults themselves about what defines and determines successful aging. Whereas researchers tend to focus on physical health as a critical factor to successful aging, older adults tend to stress the role of psychological factors such as adaptation and acceptance, along with emotional well-being. One focus group study indicated that older adults viewed successful aging as involving a foundation of health and security, along with attitudes and adaptive behaviors that can compensate for difficulties with that foundation and enable engagement in meaningful activity and stimulation. It is particularly noteworthy that a number of aspects of successful aging identified by older adults in this study (e.g., attitude, healthcare, financial security) are potentially modifiable.

Successful aging is a complex and multidetermined trait and state, with numerous genes and environmental factors making small contributions to the overall presentation of the phenotype. No single factor is either necessary or sufficient to determine the outcome in a specific individual. Cognitive preservation may, however, be a prerequisite of independent living, positive adaptation, mastery, and growth, all of which seem to be the major determinants of “success.”

IMPLICATIONS FOR FUTURE PROGRESS

The impending crisis in geriatric mental health with the projections of dramatic increases in the numbers of mentally ill older people as the baby boomers reach old age should be a cause of great concern. However, the good news is that we can do something about this issue. We should begin by recognizing the fact that the expected increases in the ranks of older mentally ill adults are not necessarily a bad thing. It is better for us to diagnose illnesses that are treatable rather than keeping them from being detected in older adults. Advances in the pharmacologic and psychosocial management of psychiatric disorders over the past two decades have already improved the treatment of several serious mental illnesses in older people. Previously unimaginable advances are taking place in the understanding of brain function in normal aging as well as in various neuropsychiatric disorders. New lines of investigations are pointing to possible ways of not just managing but even possibly curing and preventing several diseases for which there was little hope in the past. Brain stem cell implants, gene therapy, and vaccines for AD (despite the initial disappointing results) are among a host of therapies that are moving from the realm of science fiction to reality in a matter of years instead of decades. Early diagnosis of disorders such as depression or AD is important in order to start treatments that will prevent or at least slow their progress. Similarly, it is better that young adults with schizophrenia who currently have a shorter life span than the general population live into old age. Hence, an increase in the numbers of older adults with mental illnesses will be, in part, a welcome development.

At the same time, some of the increased prevalence of mental illness in older people may be preventable. For example, a proportion of cases of late-onset depression or substance use disorders could be prevented by appropriately treating individuals at high risk of developing these disorders. Thus, support groups for bereaved persons could lower the risk of developing major depression. Similarly, persons with stroke or myocardial infarction are at a high risk of depression, and the incidence of poststroke and postmyocardial infarction depression could be reduced with the use of suitable antidepressant treatments and/or psychosocial interventions following strokes or heart attacks. There are now emerging neurobiologic data suggesting that physical exercise may be as beneficial for the brain as for the body. Research also needs to be conducted to determine if engaging in specific types of mental activities may lead to the delayed onset of dementia.

The less developed regions of the world will undergo proportionately larger demographic shifts in the decades ahead than the more developed ones that have had a stable population base for many years. Yet, the less developed nations also have fragile economies and will have limited opportunities to increase health services and treatment for aged individuals. This global situation places a premium on developing diagnostic technologies and treatments that are sufficiently inexpensive and can be widely applied.

Reducing the social stigma associated with aging is as critical as reducing the social stigma associated with mental illness. Mental illnesses are brain disorders and, as such, should receive the same consideration in terms of insurance coverage.

We can change the age-old stereotype that late life is a gray and dark period of life and try to make it a “golden” one for our senior citizens. There are several aspects of aging that can be aided significantly by improved healthcare. If the societal value placed on the older population becomes more positive and the scientific openness to conceptualizing old age increases, it will become clear that the so-called problems of aging are actually opportunities for a better understanding and improvement of the whole life process. To quote Alex Comfort, “No pill or regime known, or likely, could transform the latter years of life as fully as could a change in our vision of age and a militancy in attaining that change.” The same applies to our attitude toward mental health and mental illness in old age.

FURTHER READINGS

Bartels SJ, Naslund JA. The underside of the silver tsunami—older adults and mental health care. ***N Engl J Med***. 2013;368(6):493–496.

Cao Q, Tan CC, Xu W, et al. The prevalence of dementia: a systematic review and meta-analysis. ***J Alzheimers Dis***. 2020;73(3):1157–1166.

Edmonds EC, Martin AS, Palmer BW, Eyler LT, Rana BK, Jeste DV. Positive mental health in schizophrenia and healthy comparison groups: relationships with overall health and biomarkers. ***Aging Ment Health***. 2018;22(3):354–362.

Flaherty E, Bartels SJ. Addressing the community-based geriatric healthcare workforce shortage by leveraging the potential of interprofessional teams. ***J Am Geriatr Soc***. 2019;67(S2):S400–S408.

Fredericks D, Norton JC, Atchison C, Schoenhaus R, Pill MW. Parkinson’s disease and Parkinson’s disease psychosis: a perspective on the challenges, treatments, and economic burden. ***Am J Manag Care***. 2017;23(5 Suppl):S83–S92.

Grennan G, Balasubramani PP, Alim F, et al. Cognitive and neural correlates of loneliness and wisdom during emotional bias. ***Cereb Cortex***, 2021;31(7):3311–3322.

Howard R, Rabins PV, Seeman MV, Jeste DV; International Late-Onset Schizophrenia Group. Late-onset schizophrenia and very-late-onset schizophrenia-like psychosis: an international consensus. ***Am J Psychiatry***. 2000;157:172–178.

Institute of Medicine. Retooling for an Aging America: Building the Health Care Workforce. 2008.

Institute of Medicine. The mental health and substance use workforce for older adults. In: ***Whose Hands?*** National Academies Press (US); 2012.

Jeste DV, Alexopoulos GS, Bartels SJ, et al. Consensus statement on the upcoming crisis in geriatric mental health: research agenda for the next two decades. ***Arch Gen Psychiatry***. 1999;56:848–853.

Jeste DV, Lee EE, Palmer BW, Treichler EBH. Moving from humanities to sciences: a new model of wisdom fortified by sciences of neurobiology, medicine, and evolution. ***Psychol Inq***. 2020;31:(2):134–143.

Jeste DV, Savla GN, Thompson WK, et al. Association between older age and more successful aging: critical role of resilience and depression. ***Am J Psychiatry***. 2013;170(2):188–196.

Jeste DV, Wolkowitz OM, Palmer BW. Divergent trajectories of physical, cognitive, and psychosocial aging in schizophrenia. ***Schizophr Bull***. 2011;37:451–455.

Kok RM, Reynolds CF. Management of depression in older adults: a review. ***JAMA***. 2017;317(20):2114–2122.

Kuerbis A. Substance use among older adults: an update on prevalence, etiology, assessment, and intervention. ***Gerontology***. 2020;66(3):249–258.

Lee EE, Govind T, Ramsey M, et al. Compassion toward others and self-compassion predict mental and physical well-being: a 5-year longitudinal study of 1,090 community-dwelling adults across the lifespan. ***Transl Psychiatry***. 2021;11(1):397.

Lee EE, Liu J, Tu X, Palmer BW, Eyler LT, Jeste DV. A widening longevity gap between people with schizophrenia and general population: a literature review and call for action. special issue. ***Schizophr Res***. 2018;196:9–13.

McKeith IG, Boeve BF, Dickson DW, et al Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB Consortium. ***Neurology***. 2017;89(1):88–100.

Nguyen TT, Jeste DV. Ageism: the brain strikes back. ***Cerebrum***. 2021;2021:cer-05-21.

Reus VI, Fochtmann LJ, Eyler AE, et al. The American Psychiatric Association practice guideline on the use of antipsychotics to treat agitation or psychosis in patients with dementia. ***Am J Psychiatry***. 2016;173(5):543–546.

Reynolds K, Pietrzak RH, El-Gabalawy R, Mackenzie CS, Sareen J. Prevalence of psychiatric disorders in US older adults: findings from a nationally representative survey. ***World Psychiatry***. 2015;14(1):74–81.

Rock D, Golshan S, Kraemer HC, Depp CA. Association between older age and more successful aging: Critical role of resilience and depression. ***Am J Psychiatry***. 2013;170:188–196.

Rowe JW, Kahn RL. Human aging: usual and successful. ***Science***. 1987;237:143–149.

Sajatovic M, Strejilevich SA, Gildengers AG, et al. A report on older-age bipolar disorder from the International Society for Bipolar Disorders Task Force. ***Bipolar Disord***. 2015;17(7):689–704.

Substance Abuse and Mental Health Services Administration. ***Key Substance use and Mental Health Indicators in the United States: Results from the 2019 National Survey on Drug Use and Health*** (HHS Publication No. PEP20-07-01-001). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2020.

https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR1PDFW090120.pdf

Substance Abuse and Mental Health Services Administration. Older adults living with serious mental illness: The state of the behavioral health workforce. 2019. https://store.samhsa.gov/sites/default/files/d7/priv/pep19-olderadults-smi.pdf

Thomas ML, Kaufmann CN, Palmer BW. Paradoxical trend for improvement in mental health with aging: a community-based study of 1,546 adults aged 21–100 years. ***J Clin Psychiatry***. 2016;77(8):e1019–e1025.

Wuthrich VM, Meuldijk D, Jagiello T, Robles AG, Jones MP, Cuijpers P. Efficacy and effectiveness of psychological interventions on co-occurring mood and anxiety disorders in older adults: A systematic review and meta-analysis. ***Int J Geriatr Psychiatry***. 2021;36(6):858–872.

55.2 Assessment

55.2a Psychiatric Assessment of the Older Patient

Davangere P. Devanand, M.D., and Andrea Delgado-Peraza, M.D.

In the older adults, there is an interplay between biologic, psychological, and social issues in the presentation of mental illness. Therefore, a thorough evaluation, including an assessment of the patient's internal and external environment, is necessary for accurate diagnosis and treatment. In neuropsychiatric disorders, localization of signs and symptoms to specific brain regions should also be attempted to improve understanding of the patient's condition and facilitate clinical management.

The advent of a variety of brain imaging techniques, genetic testing, and other developments have helped enhance diagnostic assessment, but these approaches are not adequate substitutes for comprehensive history taking and clinical assessment of older patients. Structured and semi-structured interviews and rating scales can be used to supplement the diagnostic interview process.

Input is often required from the family and significant others. Collateral information becomes essential in most cases as some older adult patients can minimize symptoms, partly because of fear of losing independence or becoming a burden to their families. Another factor can be shame about developing psychiatric problems, especially in those who never previously suffered from mental disorders, which is related to the broader issue of stigmatization of mental illness. When evaluating a patient for possible dementia diagnosis, it is critical to involve a reliable informant as patients with dementia may have difficulty providing history due to language problems, memory loss, or anosognosia (lack of insight).

In older adult patients, the fear of aging and its effects are common. They often state that since they are getting old, their symptoms, for example, depression or cognitive impairment, are normal and to be expected. In such a patient, it is useful to point out that among people in their own age group, only a small minority has their specific condition and that most people function well. The evaluation should aim at obtaining a comprehensive assessment of the patient's mental and physical state. The historical background and values of the older adult patient within a social class and culture, and occupational background influence the perception and presentation of illness and disease.

HISTORY TAKING

The principles in gathering history are similar to those in adults generally: establish rapport, make the patient comfortable, and maintain privacy and confidentiality. Older people have lived a long time, have undergone many stresses, are resilient, and expect to be treated with respect. The older adult patient expects direct inquiry and examination. Hand contact and reaching out and touching the patient are often useful in establishing rapport. A medical approach is often expected and respected by the older patient. This is important even in the age of telehealth that has been amplified by the COVID pandemic.

The psychiatrist needs to take an active role and involve the family and significant others whenever feasible. The psychiatrist should clearly explain the nature and purpose of the evaluation and inquire about the patient's complaints and the patient's and family's goals for the evaluation. The patient's limitations may require modification in the interview technique. In a patient with hearing impairment, the psychiatrist may need to move close to the patient's good ear and speak slowly, loudly, and clearly. If necessary, conducting the interview by writing questions for the patient to read could be done to obtain the essential information. Cognitively impaired patients may find it very stressful to perform cognitive tasks in the mental status examination and may refuse to complete the tests. It is important to explain the nature and purpose of the cognitive tests before administering them, and frequent breaks between tests can help the patient complete the cognitive assessment.

Details about medical illnesses and medications taken are important in older adult patients. Information from other physicians, hospitals, and therapists should be obtained whenever needed to supplement the evaluation. Getting a complete medication list is imperative as older adult patients tend to have polypharmacy, and occasionally cognitive symptoms can improve when inappropriate medications are discontinued. If medication nonadherence occurs, simple techniques like introducing pillboxes are often useful. In severely impaired patients, particularly patients with dementia, information from caregivers and professionals involved in social support and related services is important. Most older adult patients first see their primary care physician for their psychiatric symptoms, and this is the setting where the majority of them are treated. Therefore, primary care physicians may refer their patients to a psychiatrist if diagnosis and management become difficult.

The history should include the following components: identifying data (name, age, sex, education, marital status), chief complaint, history of present illness including onset, duration and symptom course over time, past psychiatric history, medical history and medications are taken, alcohol and substance use history, work history and living situation, and family history of psychiatric, neurologic and other medical disorders.

The initial evaluation session may be long in an older patient, especially if extensive information from collateral sources needs to be obtained, which may be necessary to confirm the information provided by the patient and evaluate reliability. Further, a second visit may be necessary to provide enough time to gather all the additional information from the results of all the tests and procedures conducted with the goal of making an informed diagnosis. In a patient with cognitive decline, laboratory tests to identify potential causes (e.g., syphilis, vitamin B12 deficiency, folate deficiency, hypothyroidism, medication toxicity) of cognitive impairment, and neuropsychological testing and structural/functional brain imaging procedures (CT or MRI, SPECT or PET) may need to be ordered at the first visit. In these cases, the second visit becomes an informing visit at which the patient and family receive an opinion about the diagnosis and likely prognosis. Some diagnostically puzzling cases may need more visits and follow-ups to help clarify the patient’s condition. In particular, the psychiatrist needs to keep an open mind about the possibility that the initial diagnosis may need to be revised based on the patient’s clinical course and response to treatment, for example, change in cognitive status after initiation or withdrawal of medications that have an adverse impact on cognition, including benzodiazepines and medications with anticholinergic properties.

History of Present Illness

The patient should be encouraged to describe the history of the current illness in an open-ended interview. The chief complaint might provide additional information as to whether that patient came of their own volition or was forced to come by a family member, with more structured questions following this initial inquiry. The onset and course of symptoms should be evaluated in relation to other changes, including loss of loved ones, retirement, loneliness or isolation, relation to daily activities, medication usage, response to treatment if applicable, and intercurrent medical illnesses. Many older adult patients consider insomnia and fatigue to be a normal part of the aging process, and these impairments may require explicit inquiry. The time course of development of the present illness is very important in older adult patients. In a patient with cognitive decline, a history of gradual deterioration suggests dementia, episodic cognitive deficits with confusion suggest delirium, focal neurologic symptoms or signs suggest cerebrovascular disease, and a short, downhill course may indicate rapidly progressive dementia like Creutzfeldt–Jakob disease.

Ascertaining age and age at onset of symptoms is critical in the older adult patient. There is considerable evidence that patients with late-onset (onset after 50 to 60 years of age) major depression are less likely to have a positive family history of affective disorder and more likely to have evidence of cerebrovascular disease and cerebral atrophy than older adult patients with early-onset depressive illness. Late-onset schizophrenia differs in some respects from early-onset schizophrenia, with a greater preponderance of females, paranoid subtype, and some differences in neuropsychological test profiles. Patients with early-onset Alzheimer disease (onset in the 40s or 50s) typically have presenilin I or presenilin II gene mutations, whereas late-onset Alzheimer’s patients usually do not have these gene mutations though a large proportion may have the apolipoprotein E e4 genotype that increases risk and some have a positive family history of Alzheimer disease.

Based on the clinical presentation, the inquiry should focus on pertinent symptoms to make a diagnosis of major depression or psychosis or anxiety disorder or other primary psychiatric disorder. Potentially serious symptoms that should be reviewed with the patient include depressed mood, suicidal thoughts, lack of interest, agitation/aggression and impulsivity, delusions, hallucinations, and an assessment of functioning, including ADLs and social functioning.

Insomnia is common in the older adults. During the aging process, there is a small, gradual decline in the number of hours slept per night. Therefore, any change in sleep pattern that occurs over a short period of time should alert the psychiatrist to the presence of an underlying psychiatric disorder, most commonly a depressive illness. Some patients may downplay insomnia, anorexia, and lack of interest as normal features of the aging process, but this constellation of symptoms should trigger a thorough evaluation for the presence of a depressive disorder. Insomnia, apathy, and anhedonia also occur in early dementia, and the presence of these symptoms should lead to a systematic assessment of cognitive abilities. A gradual decrease in libido and sexual functioning occurs in some older adults, but a rapid deterioration in sexual interest should trigger an investigation into the underlying cause. Diabetes, hypertension, and the iatrogenic effects of medication are common causes of sexual dysfunction, but so is depression. Somatization is typical of older adult depressed patients, but there are cultural differences in the degree to which this occurs.

Older adults face a variety of stresses: medical illnesses, physical disability, sensory impairments in vision and hearing, loss of family and friends, loss of social supports, changes in relationships with children and significant others, alterations in living situation, and loss of income. One or more of these factors may be related to the onset of the clinical syndrome. Although many psychiatric disorders are biologically driven, psychosocial precipitants play an important role in causing or exacerbating psychiatric disorders. The death of a spouse is extremely stressful, as is the death of an adult child. Patients with major depression perceive the significantly greater negative emotional impact of the same life events than patients with dysthymic disorder and healthy controls, particularly for interpersonal conflicts.

In the older adults, there is a high rate of delirium after admission to a hospital. This may be related to the lower “cognitive reserve” of older compared to young adults. If there is a suspicion of delirium, potential risk factors should be evaluated, including fever, urinary tract infection, urinary retention, fecal impaction, changes in medications or dehydration.

The following symptoms need to be asked of patients with cognitive decline: difficulties in memory or thinking or concentration, forgetting of recent events, forgetting where things are located, forgetting to take medications or keep appointments, problems in finding the right word to say, getting lost in familiar places, making mistakes with commonly used gadgets that include cell phones and kitchen equipment, and difficulty in managing finances like balancing the checkbook or paying for items in a store. These symptoms may be the first indicators of dementia. However, the isolated symptom of forgetting names is common during the normal aging process and by itself is not indicative of dementia.

A specific inquiry should also be made about pain, visual and hearing impairment, and possible drug toxicity. For patients in institutional settings, interpersonal interactions with other residents are important.

Psychiatric History

Evaluation of psychiatric history should include the course of illness and types of treatment, that is, psychotherapy, medications, and ECT, and history of response or lack of response to these treatments. Treatments that worked well before in the patient, or worked well in family members, should be given priority.

History of depression, anxiety disorders, and psychosis/schizophrenia, and treatments received in prior and current episodes, should be obtained. Personality disorders are less common in older adults depressed patients compared to young adults with depression. Obsessive-compulsive and avoidant personality subtypes are more common than borderline, histrionic, narcissistic and antisocial personality disorders in older adults. Leading questions should be asked about alcohol or substance abuse, history of head trauma, infections like meningitis or encephalitis, stroke, seizure disorder, and other neurologic disorders such as restless leg syndrome or REM sleep behaviors.

Family Interview and History

In the cognitively impaired patient, it is advisable to see the patient alone first as older adults abuse is more prevalent than is reported to the authorities, then the informant (usually a family member) and/or the patient together with the informant. Studies show that 1 in 10 Americans above 60 years old are abused by a family member, which translates to 5 million elders abused per year in the United States. When interviewing both the patient and the family member, assessing the interaction based on verbal and nonverbal communication might provide additional information. The informant interview provides an independent assessment of the patient's environment, for example, the cognitively impaired patient may have allowed the home to become disorganized or left the stove on with adverse consequences. In severe dementia, interviewing the informant (or informants) both separately and in a joint interview with the patient is often necessary. In these patients, impairment in memory and other cognitive functions may be so severe that nearly all the history needs to be obtained from the informant. Even at the stage of mild cognitive impairment, the informant's report of the patient's function has been shown to be a stronger predictor of dementia on follow-up than the patient's self-report of functional ability.

Patients may minimize symptoms in the presence of other people but may be more revealing about symptoms like depressed mood and suicidal ideation when interviewed alone. Paranoid ideation may not be revealed when family members are also present in the room. A joint interview with the patient and family (or another informant) may become necessary if there are conflicts about the patient's habits and behavior that need to be resolved, for example, the patient wants to continue driving after developing dementia. However, the family member's report may not always be accurate. For example, in patients with dementia, there is consistent evidence that the informant's report that the patient is depressed often reflects the informant's own depression rather than the patient's mood state. Studies show that caregivers of older adult patients with dementia are more prone to develop financial, emotional, and physical difficulties and receive treatment with psychotropic medications compared to caregivers of older adult patients without dementia; depression affects 20% to 40% of caregivers.

A genogram is helpful in delineating the familial relationships and the important psychiatric, neurologic, and other medical illnesses in different generations. An inquiry should be made about a family history of psychiatric hospitalization, unipolar depression or bipolar disorder, schizophrenia, anxiety disorders, alcohol or substance abuse, suicide, Alzheimer disease, stroke, epilepsy, Creutzfeldt–Jakob disease, Huntington disease, Parkinson disease, and other neurologic disorders. Interviewing more than one family member, and preferably more than one generation of family members, improves accuracy regarding family history.

In older adult patients with functional impairment, family and other caregiver support need to be assessed. The time spent by family members, services provided by family members, perception of family support, and tolerance by family members of changes in behavior are elements that need to be assessed and considered in the development of a treatment plan. Services that may be needed include transportation, nursing services, physical therapy, continuous supervision, homemaker, and household services, meal preparation, administrative, legal, and protective services, financial assistance, and coordination of the delivery of services. These issues may be of greater concern to the patient and family than the clinical symptoms and usually require coordination with nurses, case managers and social workers in the health support system for the older patient.

The patient and family usually find the transition to assisted living or nursing home to be very difficult. Guilt, frustration, anger, helplessness, isolation, depression, and demoralization on the part of the caregiver can reduce caregiver effectiveness and adversely impact the patient. This can lead to an accident like falling and fracturing a hip, inappropriate nursing home placement, or, in some instances, elder abuse. It is important to inquire how the caregiver is coping with the stress of having to take care of the patient, and an appropriate referral for the caregiver to seek evaluation and treatment may become necessary.

Personal History

Details should be obtained about the patient’s marital status, living situation, educational background, work history, and the nature of relationships that are important to the patient. Premorbid personality does carry into old age, but its impact on illness in the older adults varies considerably. Personality itself may change with aging. This change is believed to be environmentally determined, but it could also be based on changes in neurobiology, for example, catecholaminergic neurons in the locus coeruleus decay markedly with aging and this change may be related to the relatively low rate of late-onset anxiety disorders in the older adult population. Religion is often important to the older adult patient and needs to be included in the history as it is a protective factor when suicidal thoughts or hopelessness are present.

The interviewer may be embarrassed to ask about the sexual history of the older adult patient, but many older adults do have symptoms of sexual dysfunction and want to talk about it. Inquiry about sexual orientation and sexual dysfunction should cover any history of sexually transmitted diseases, erectile dysfunction (including impotence) in men, and anorgasmia in men and women. Sexual dysfunction can be a side effect of medication, particularly SSRIs and related antidepressants. Diabetes and hypertension are also common culprits.

Medical History

The older adult patient typically suffers from several medical illnesses and takes a large number of medications. Patients may not keep track of all the medicines they are taking, and their list often excludes medications obtained over-the-counter or alternative medications sold in health food stores. Therefore, it is important to inquire about all these categories of medications, and it is a good idea for the psychiatrist to ask the patient to bring all the current medications in a bag to the visit. Benzodiazepines, steroids, and anticholinergics can lead to cognitive impairment and even delirium, L-dopa and dopamine agonists prescribed for Parkinson disease can lead to psychosis, and the use of some antihypertensives and steroids has been associated with depression.

Medical history should cover all major illnesses, surgery, hospitalizations, trauma, and treatments. A variety of medical disorders can first present with disturbances in mood, cognition, or psychotic features, including infections, electrolyte and metabolic imbalances, cardiovascular and cerebrovascular disease. Depression can lead to weight loss and malnutrition. Dietary history is important in this regard; the older adult living alone may not obtain adequate nutrition and develop vitamin and other deficiencies.

Alcohol and Substance Abuse

Alcohol abuse is not uncommon in the older adults, and explicit inquiry is usually necessary to elicit this history. Even when alcohol consumption has been constant in amount for many decades, it can lead to cognitive decline in susceptible patients as they grow older because of decreased cognitive reserve due to neuronal loss in the brain. Neuropsychiatric complications of alcohol abuse and dependence, now subsumed under alcohol use disorder in DSM-5, include varying degrees of amnesia, Korsakoff syndrome, malnutrition, and intention tremor and ataxia due to cerebellar damage. Other substance use disorders are less common in older adult patients, but some do use marijuana and tobacco regularly. Even with the advent of new hypnotics, benzodiazepine use remains fairly widespread. This is of particular concern in the older adults because benzodiazepines can contribute to excessive sedation, unsteady gait and falls, and cognitive deficits. Patients with chronic pain, particularly back pain, may be dependent on narcotics.

MEDICAL EXAMINATION

If a medical examination has not been conducted recently by the patient's primary care physician, the psychiatrist needs to complete this component or arrange to have it done by another physician. The neurologic examination is particularly important and needs to include evaluation of sensorium, cognition, motor strength and function, gait, sensation, reflexes, extrapyramidal signs, vision, and hearing. Neurologic deficits, endocrine disorders, cardiovascular disease, occult malignancy, and infections are potential causes of cognitive impairment, agitation, depression, and occasionally psychosis. Urinary tract infection, which can occur without fever, is a common underlying cause of cognitive decline and delirium, particularly in the nursing home.

Mental Status Examination

The mental status examination is essential. Serial mental status exams may be necessary, especially if mental functioning fluctuates over time.

Observation of Behavior. The first part of the mental status examination is a general observation of the patient's appearance, personal hygiene, motor activity, any unusual behaviors, speech, attitude toward the examiner, and the presence of hearing and visual aids. Poor hygiene suggests dementia, but commonly the patient with dementia may look well-groomed because of the caregiver's efforts.

In older patients, paranoid delusional thoughts occur not only in schizophrenia and delusional disorder but also in dementia, particularly Alzheimer disease, and during the course of delirium. Apathetic, withdrawn behavior is common in both depression and dementia. Disturbances of motor activity are frequent in older adult patients, and bradykinesia with slow gait is common in the eighth to tenth decades of life. A slow shuffling gait with mask-like facies, resting tremor and rigidity indicate Parkinson disease or extrapyramidal side effects of antipsychotic medication. Older adult patients are particularly prone to the neurologic side effects of antipsychotics. Lip-smacking and involuntary Bucco facial movements suggest tardive dsykinesia in patients with prior antipsychotic exposure, though this has become less common since the advent of second-generation antipsychotics. Impulsive, disinhibited behavior with minimal to mild cognitive impairment may be the first signs of frontotemporal dementia, though increasing apathy is, in fact, more common than disinhibited behavior in frontotemporal dementia.

Language and Thought Disorder. Speech and language are important to assess in older adults. The quality, quantity, and fluency of speech need to be assessed. Speech can be normally responsive, rapid, pressured, loud or emotional, slow, monotonous, lacking spontaneity, or slurred. Aphasia or slurred speech occurs in stroke and in a variety of neurologic disorders. Mutism occurs in the late stages of primary progressive aphasia, dementia from a variety of causes, and amyotrophic lateral sclerosis. Repetitive speech with perseveration on the same topic is common in patients with dementia and tends to be secondary to the loss of recent memory, which is a hallmark of the illness.

The patient should be able to understand and answer questions appropriately and think in abstract terms. The paucity of speech with long latency is characteristic of major depression, particularly the melancholic subtype. Tangentiality and looseness of associations are characteristic of schizophrenia. The patient with Alzheimer disease may display repetition, perseveration, and fragmented ideas. In contrast, patients with subcortical dementias like Huntington or Parkinson disease usually do not show these patterns of thought process.

Thought content abnormalities include preoccupations or obsessions, including preoccupation with health problems in the older adult depressed patient. Delusions can be persecutory, grandiose, somatic, erotic, guilty, or nihilistic. Hypochondriacal delusions are common in delusional depression, for example, bowels are not working or have been destroyed. Delusional depression in the older adults is associated with a high suicide risk. Bizarre, fixed delusions are common in schizophrenia, but they are rare in dementia where transient persecutory ideas are more common. Distinguishing fact from fiction can be difficult in the older adult patient, who may have had personal effects stolen and may not have a persecutory delusion as the evaluation may suggest.

Perception. Perceptual disorders include hallucinations and illusions. Hallucinations can be simple or complex and organized. Hallucinations can occur in dementia and in patients with hearing or visual deficits. In these patients, it is important to distinguish an illusion, that is, perceptual distortion of a real sensory stimulus, from a hallucination that may indicate more severe psychopathology. Memory loss and disorientation help to identify the problem as being related to dementia. A number of other perceptual abnormalities occur in a variety of neurologic disorders involving damage to specific brain regions. These syndromes include micropsia, which occurs when objects look smaller than their real size, macropsia, which occurs when objects look larger than their real size, visual agnosia, which is the inability to recognize visible objects, prosopagnosia, which is the inability to recognize familiar faces, astereognosis, which is the inability to recognize objects by touch, and autotopagnosia, which is the inability to recognize parts of one’s own body.

Mood and Affect. Mood is the subjective and affect the objective emotional state. Many older adult patients describe their mood accurately as depressed or anxious, but some tend to deny and suppress it. For example, an older adult woman who is crying may say that she is not depressed. Somatization is common in older patients, particularly gastrointestinal symptoms like constipation. Some patients are not sure if they are depressed because they have begun to lose interest in activities but do not experience a clear-cut depressed mood. Many of these patients have dysthymia or minor depression. Depression is common in a variety of other neuropsychiatric syndromes, including dementia, stroke, and Parkinson disease. Mania is not common in the older adults but can occur as a primary late-onset mania or secondary to dementia or stroke, or other neurologic insults. Labile affect is believed to be typical of vascular dementia, but it remains unclear if labile affect is more common in vascular compared to other forms of dementia. Pseudo-bulbar affect is uncontrollable laughing or crying in the absence of the corresponding subjective mood experience and occurs in neurologic disorders like amyotrophic lateral sclerosis and frontotemporal dementia, and less commonly in Alzheimer disease. Medication side effects or toxicity need to be considered as possible etiologies for depression. Of note, late-onset anxiety disorders, including panic disorder, are relatively rare.

Suicidal and Homicidal Ideation. An inquiry should be made about suicidal and homicidal ideation. Suicidal ideation should be probed by first asking whether the patient thinks that life is not worth living. A positive response should lead to follow-up questions about whether there is a suicide plan or history of a suicide attempt. If this line of inquiry indicates suicidal risk, further questions should focus on whether a gun or other means to commit suicide are easily available, and the appropriate clinical intervention with the patient and family members should then be made. Studies consistently show that older adult males, particularly white males, living alone are the demographic group with the highest risk of committing suicide and that they often use guns as the method of suicide. Of note, the majority of patients who commit suicide have visited their primary care physician a few days to weeks before they act. Patients with a new recent diagnosis of dementia may have an increased risk of suicide.

Sleep. Approximately half of older adults complain of sleep difficulties. Careful assessment of sleep habits, including factors such as alcohol and caffeine intake that can disrupt sleep, is important. Other factors contributing to sleep impairment include age-related changes in sleep architecture and circadian rhythms, sleep disorders including sleep apnea, dementia, depression, other medical illness, and polypharmacy. The impact of treatment, or withdrawal of treatment, for example, benzodiazepines, needs to be assessed in the context of these factors that can underlie sleep disorders.

Sensorium and Cognition. Evaluation of sensorium and cognition is critical and essential in older adult patients. Whenever there is suspicion of cognitive impairment by history, it is important to administer a screening test like the Mini Mental State Examination (MMSE) or Montreal Cognitive Assessment (MOCA). Both are adversely affected by low educational attainment, particularly the MOCA, and this fact should be taken into account during the assessment.

At the start of this portion of the exam, the patient should be told that some of the questions are easy and some more difficult and that people score across a wide range on these tests. This instruction is important because patients with significant cognitive deficits may not agree to complete the test if they feel that they are taking an examination and that their deficits are being judged critically. Skills that can be tested in this component of the evaluation can be assessed more thoroughly by detailed neuropsychological testing, including tests of attention and concentration, learning, recall, recognition after a delay (memory), comprehension, verbal fluency, repetition, naming (language), visuospatial abilities, calculation, praxis, and the executive functions of initiation, perseveration, response control, synthesis and integration, and changing mental sets (frontal lobe functions).

A patient needs to be able to focus and be attentive in order to complete the tests. If a patient is distractible and unable to concentrate, the subsequent components of the test may not be valid. Educational achievement and the patient’s primary language are important variables to consider in the interpretation of cognitive tests. A patient with low education or a patient for whom English is not the primary language is likely to score poorly on some components of the cognitive assessment, but memory is affected less by these factors. Therefore, a patient with low education who scores 25/30 on the MMSE but recalls 3/3 objects at 5 minutes is less likely to be developing dementia than a patient with a post-graduate degree who scores 28/30 on the MMSE but recalls only 1/3 objects at 5 minutes.

Cognitive Testing. A brief evaluation of cognitive ability can be performed with the MMSE. A score of less than 24 out of 30 is considered indicative of a cognitive deficit. A variety of conditions, including major depression, can be associated with low MMSE scores. In patients with mild cognitive impairment, deficits in recall of three objects after a 5-minute delay, and impairment in the orientation items, suggest an increased likelihood of developing dementia. However, the MMSE by itself is not sufficient to make a diagnosis of dementia, especially in patients with mild impairment. The MMSE has limitations in evaluating executive function. The Clock drawing test can be used instead to evaluate executive function. In the early stages of dementia, attention is preserved, but short-term memory is impaired. In contrast, attention is impaired in depression, but short-term memory is relatively intact, that is, what is registered in the brain is generally retained. Delirium is characterized by disorientation, which also occurs in the mid to late stages of most dementing disorders. The rapidity and duration of disorientation are useful indicators of delirium versus dementia in the differential diagnosis. If the results of the cognitive screening leave the diagnosis in doubt, which is often the case, a referral for detailed neuropsychological testing is usually indicated. Neuropsychological testing is useful in making the diagnosis of dementia in its early stages, but its utility in differentiating subtypes of dementia, for example, Alzheimer's versus vascular, or in diagnosing depression in patients with dementia, is not as well established.

Insight. Insight refers to the ability of patients to understand their own symptoms and their potential causes. Insight is impaired in most forms of moderate to severe mental illness and can disappear in severe dementia. Judgment can be assessed during the course of the interview by inquiring about the patient's opinion and proposed course of action in incidents that they describe in their own history or in a hypothetical incident. Judgment is compromised in psychotic disorders and dementia.

DAILY FUNCTIONING

Assessment of daily functioning is essential in the evaluation of the older patient, unlike the situation in most young adults. Activities of daily life (ADL) can be classified into basic ADL (e.g., eating, dressing, toileting), instrumental ADL (cooking, shopping, laundry) and more complex social and cognitive functions requiring greater mental effort, for example, handling finances and doing taxes. Basic and instrumental ADLs need to be assessed in older adult patients. The Pfeffer functional activities scale is widely used in large-scale studies of patients with mild cognitive impairment and Alzheimer disease. Other scales for ADL and instrumental ADL include the Katz Index of Independence in Activities of Daily Living and Lawton-Brody Instrumental Activities of Daily Living. The Clinical Dementia Rating (CDR) incorporates cognitive, functional, and social-cognitive items, and is widely used as an overall indicator of the level of cognitive impairment and dementia severity.

NEUROLOGIC EVALUATION

Psychiatric, cognitive, and neurologic examinations overlap considerably in older adults. While patients with manifest neurologic disorders need a consultation with a neurologist, the geriatric psychiatrist needs to be able to perform the neurologic examination because many patients present with syndromes that overlap between the two specialties. In patients who have cognitive disorders or focal neurologic symptoms or signs, a neurologic examination is essential. All components of the neurologic examination are important, but in the older adult patient presenting to a psychiatrist, the areas of frontal lobe function, basal ganglia function (examination for extrapyramidal signs), motor and sensory function, reflexes, coordination, and evaluation of vision and hearing are essential. A neurologic evaluation could help determine parkinsonism's symptoms that characterize Parkinson disease and Lewy body dementia.

RATING SCALES

Although rating scales are used primarily in research studies, they can serve several possible functions in clinical practice: address specific clinical problems and document change, teach, communicate with colleagues, and strengthen the chart or electronic medical record for medicolegal and insurance purposes. The SCID can be administered to make comprehensive DSM diagnoses, but it requires special training and takes too long to be recommended for routine clinical practice. Some of the widely used scales to assess specific domains of psychopathology and cognitive and daily functioning will be summarized here.

Depression

The Hamilton Rating Scale for Depression (HRSD), Montgomery-Asberg Scale (MADRS), Geriatric Depression Scale, and Center for Epidemiological Studies-Depression Scale (CES-D) have been used in a variety of research studies. The Hamilton Rating Scale for Depression and the Montgomery Asberg Depression Rating Scale are particularly useful in patients with severe depression. Another widely used instrument is the Beck Depression Rating Scale, a self-report instrument. Epidemiologic studies have typically used the self-report CES-D. The Geriatric Depression Scale, which can be self-rated or observer-rated, is quick to administer and can be quite useful. To assess depression in patients with dementia, the Cornell Scale for Depression in Dementia has been developed.

To assess cognitive impairment, a plethora of rating instruments are available. The most widely used instrument is the MMSE, and the MOCA is comparable in its utility. Education and cultural background can affect performance on these tests, which is one reason why the score obtained should be evaluated in the context of the entire clinical picture. MOCA is available in multiple versions, including a version for blind people and in different languages. Of note, the MMSE is copyrighted, and using the MOCA requires first taking a 1-hour training course.

To assess psychosis, the Brief Psychiatric Rating Scale (BPRS) is commonly used in young adults and can also be used in older adult patients. To address the unique features and methods of assessment of psychosis in dementia, a variety of scales have been developed. These include the Neuropsychiatric Inventory (NPI), the Columbia University Scale for Psychopathology in Alzheimer's Disease (CUSPAD), CERAD Behavior Rating Scale, and the Behavioral Pathology in Alzheimer's disease (BEHAVE-AD) rating scale. The Cohen-Mansfield agitation inventory is available for use in nursing home patients. Currently, the NPI, which uses a time-efficient, decision-tree approach, is the most widely used scale to evaluate the effects of pharmacologic treatment of psychosis and agitation in patients with dementia.

The standardized Confusion Assessment Method (CAM) can be used to assess for delirium. To assess the neurologic side effects of antipsychotic medications, the Unified Parkinson's Disease Evaluation Scale and the AIMS scale can be used.

BIOMARKERS FOR A DEMENTIA DIAGNOSIS

A biomarker is a measurable parameter that reflects a physiologic or pathologic process. It should be reliable, reproducible, and measurable, preferably in a noninvasive way. Alzheimer dementia continues to be the most common type of neurocognitive disorder, accounting for the majority of patients with dementia in the United States. Structural imaging with MRI shows decreased hippocampal and entorhinal cortex volume in Alzheimer disease, but these measures show only moderate accuracy in differential diagnosis and estimation of prognosis. The brains of Alzheimer disease patients are characterized neuropathologically by the aggregation of extracellular β-amyloid (Aβ) peptides into plaques and intracellular hyperphosphorylated tau into neurofibrillary tangles. Studies show that AD biomarkers, particularly PET imaging showing increased amyloid in the brain, can be detected several decades before the clinical onset of cognitive impairment. Imaging using PET for tau biomarkers may also indicate abnormality during the preclinical stage of AD, though perhaps at a later stage than amyloid. Amyloid and tau biomarkers can also be measured in the CSF with decreased amyloid AB42 and AB42/40 ratio, with an increase in tau and phospho-tau being indicative of AD. These biomarkers can be utilized in clinical practice together with neuropsychological testing and neuroimaging to increase diagnostic precision and earlier diagnosis. One limitation of amyloid and tau biomarkers is lack of specificity, for example, similar abnormalities can occur in Lewy body dementia. Another approach is to use FDG-PET imaging that can reveal distinct regions of low metabolism and hypoperfusion and could differentiate between AD and frontotemporal dementia. Some limitations to CSF or PET/SPECT biomarkers are that they are invasive, time-consuming, and expensive, and each biomarker by itself does not have greater than 75% diagnostic accuracy.

New studies have proposed that AD biomarkers can be measured in the plasma, which would be less invasive, cost-effective, and easily accessible. These include Aβ42/ Aβ40 ratio, NfL, T-tau, and more recently, P-tau181 and possibly P-tau217. Other noninvasive biomarkers include olfactory identification deficits that occur because the brain regions subserving olfactory neuronal projections to limbic regions are affected early in the course of Alzheimer disease.

COMPETENCE

The psychiatrist may be asked about the patient’s level of competence to give consent for medical treatment, power of attorney, sign or change a will, or manage personal finances. Competency is a legal term, and the criteria for judging it have never been precisely defined. Competency needs to be judged in a specific framework and context, that is, a patient may be competent to execute one task but not another and may be competent at one time but may change shortly thereafter. There are some specific measures to evaluate financial abilities, but testamentary capacity remains a judgment call. Both cognitive abilities and general functioning need to be evaluated before making this judgment.

SYNDROMES SPECIFIC TO GERIATRIC PATIENTS

There are several syndromes common in geriatric patients that do not fit easily into the current diagnostic nomenclature. These syndromes, cut across medical and psychiatric diagnoses, are widely recognized and need to be incorporated into the evaluation. Apathy is common, and it can be difficult to separate from depression. Frailty is a syndrome that is associated with increased morbidity and mortality, and gait and grip strength are important metrics in the evaluation of frailty. Technology-based assessments that can monitor activity levels and sleep to inform diagnosis are now used mainly in research studies, but over time they are likely to be included in the standard clinical assessment of patients. Similarly, telemedicine can be extended to telepsychiatry, particularly given the aging of the population and the need to provide psychiatric assessment and treatment to the homebound older adults.

SUMMARY

The evaluation concludes with a summary of the key findings and differential diagnosis, which leads to an overall diagnosis. An etiologic formulation using the biopsychosocial model is the next step. Finally, these conclusions are built into the treatment plan and management strategies for the patient. Holistic management through an integrated healthcare model is warranted most of the time for older adult patients. Improving the social support system, engaging that patient in social and cognitive activities, and encouraging a healthy diet and exercise routine, especially for patients with hypertension, diabetes and obesity, should also be part of the treatment plan.

FURTHER READINGS

Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell scale for depression in dementia. ***Biol Psychiatry***. 1988;23:271–284.

Ashton NJ, Hye A, Rajkumar AP, et al. An update on blood-based biomarkers for non-Alzheimer neurodegenerative disorders. ***Nat Rev Neurol***. 2020;16(5):265–284.

Brown PJ, Devanand DP, Liu X, Caccappolo E. Functional impairment in elderly patients with mild cognitive impairment and mild Alzheimer disease. ***Arch Gen Psychiatry***.

2011;68:617–626.

Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric inventory: Comprehensive Assessment of Psychopathology in Dementia. ***Neurology***. 1994;44:2308–2314.

DeMers S, Dinsio K, Carlson W. Psychiatric care of the older adult: an overview for primary care. ***Med Clin North Am***. 2014;98(5):1145–1168.

Devanand DP, Miller L, Richards M, et al. The Columbia University Scale for Psychopathology in Alzheimer's disease. ***Arch Neurol***. 1992;49:371–376.

Folstein M, Folstein S, McHugh P. "Mini-mental state": a practical method of grading the cognitive state of patients for clinicians. ***J Psychiatr Res***. 1975;12:189–198.

Glover J, Srinivasan S. Assessment of the person with late-life depression. ***Psychiatr Clin North Am***. 2013;36(4):545–560.

Karimi S, Dharia SP, Flora DS, Slattum PW. Anticholinergic burden: clinical implications for seniors and strategies for clinicians. ***Consult Pharm***. 2012;27(8):564–582.

König A1, Aalten P, Verhey F, et al. A review of current information and communication technologies: can they be used to assess apathy? ***Int J Geriatr Psychiatry***. 2014;29(4):345–358.

Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. ***Gerontologist***. 1969;9:179–186.

Lesser RP, Fahn S, Snider SR, Cote LJ, Isgreen WP, Barrett RE. Analysis for the clinical problems in parkinsonism and the complications of long-term levodopa therapy. ***Neurology***. 1979;29:1253–1260.

Lin JS, O'Connor E, Rossom RC, Perdue LA, Eckstrom E. Screening for cognitive impairment in older adults: A systematic review for the U.S. Preventive Services Task Force. ***Ann Intern Med***. 2013;159(9):601–612. Erratum in Ann Intern Med. 2014 Jan 7;160(1):72.

Mackenzie TB, Robiner WN, Knopman DS. Differences between patient and family assessments of depression in Alzheimer's disease. ***Am J Psychiatry***. 1989;146: 1174–1178.

McKeith IG, Galasko D, Kosaka K, et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. ***Neurology***. 1996;47(5):1113–1124.

Morris JC, Blennow K, Froelich L, et al. Harmonized diagnostic criteria for Alzheimer's disease: recommendations. ***J Intern Med***. 2014;275(3):204–213.

Neikrug AB1, Ancoli-Israel S. Sleep disturbances in nursing homes. ***J Nutr Health Aging***. 2010;14(3):207–211.

Ramos-Ríos R1, Mateos R, Lojo D, Conn DK, Patterson T. Telepsychogeriatrics: a new horizon in the care of mental health problems in the elderly. ***Int Psychogeriatr***. 2012;24(11):1708–1724.

Scheltens P, Blennow K, Breteler MM, et al. Alzheimer's disease. ***Lancet***. 2016;388(10043):505–517.

55.2b Complementary and Integrative Geriatric Psychiatry

SARAH A. NGUYEN, M.D., HANADI AJAM OUGHLI, M.D., HARRIS A. EYRE, M.B.B.S., PH.D., AND HELEN LAVRETSKY, M.D., M.S.

Although the terms “alternative,” “complementary,” and “integrative” are often used interchangeably, the National Center for Complementary and Integrative Health (NCCIH) defines these terms as follows: “Alternative” medicine refers to a set of practices of medicine (e.g., traditional, oriental, mind–body, etc.) without sufficient evidence base that is used in place of conventional medicine, whereas “complementary” medicine refers to nonmainstream practices that are used together with conventional medicine. “Integrative” medicine refers to the use of conventional and complementary approaches in a coordinated way targeting specific disorders or underlying neurobiologic processes.

Complementary and integrative medicine (CIM) emphasizes a holistic, patient-focused approach to health care and wellness. CIM targets the whole person rather than one organ system, and historically, these practices are oriented toward a harmonious rebalancing of mind, body, and spirit. CIM approaches incorporate the belief that well-being is a state of balance, encompassing the spiritual, physical, and mental/emotional functioning achieved by leading a balanced, healthy lifestyle. This is achieved by proper nutrition, exercise, sleep habits, and the ability to regulate stress response through meditation or other mind–body practices. The NCCIH identified three categories for complementary and integrative approaches: (1) mind and body practices (massage therapy, meditation, yoga, acupuncture, chiropractic/osteopathic manipulation, hypnotherapy, tai chi, qi gong, healing touch, and relaxation exercises); (2) natural products (herbal medicines, botanicals, vitamins, minerals, probiotics, and other dietary supplements); (3) other complementary approaches (indigenous healing practices, Chinese medicine, Ayurvedic medicine, homeopathy, and naturopathy). In general, CIM practices aim to achieve a greater balance and well-being, making them especially valuable strategies for disease prevention. Because CIM is often considered to be a natural and safer option compared to conventional medicine in addressing common health conditions, these interventions are rapidly gaining popularity in the United States; the top practices include mind–body practices like yoga and meditation and the use of prayer for health and multivitamin supplementation. The baby boomer cohort and subsequent birth cohort are particularly high utilizers of CIM.

Mind–body practices refer to practices that emphasize mind–body connection awareness in order to achieve well-being, stress reduction, or reduction in stress-related disorders. These can be mindful exercise practices rooted in ancient traditions (e.g., yoga, tai chi, and qigong) or meditation-based practices (e.g., mindfulness meditation). The shared feature of mindful activities is putting attention and raising awareness of bodily sensations (e.g., breathing, muscle relaxation) and the ability for self-regulation by rebalancing autonomic nervous system functioning to offset the deleterious effects of chronic stress.

Other integrative interventions are based on improving lifestyle using diet, exercise, and sleep regulation. Dietary interventions address the effect of dietary choices on health and may include Western, traditional, and Mediterranean dietary patterns, as well as caloric restriction, among others. Natural products and supplements include herbs, vitamins, minerals, natural supplements, and probiotics. Body-based practices may include spinal manipulation (e.g., chiropractic, osteopathic medicine), aromatherapy, and massage therapy. Whole medical systems include traditional Chinese medicine (including acupuncture), Ayurvedic medicine, or homeopathy. These CIM interventions are largely based on therapies that have been utilized in many civilizations and religions for hundreds and thousands of years. In the more recent decades, empirical science has begun to explore the evidence behind these therapies. For a graphical representation of the CIM subtypes, refer to Figure 55.2b–1.

There are a number of major reasons why geriatric mental health providers need to recognize the importance of these interventions. (1) ***CIM therapy use is high and rising***. Research suggests the 12-month prevalence of any CIM usage (excluding prayer) in the United States is around 35% to 50%; baby boomers (adults born from 1946 to 1964) report significantly higher rates of use than the Silent Generation (born from 1925 to 1945) for chronic conditions. (2) ***The global population is aging***. The number of older persons (aged ≥60 years) globally is expected to more than double, from 841 million people in 2013 to more than 2 billion in 2050. (3) ***CIM therapies are increasingly cited in clinical guidelines***. It is now commonplace to note the use of therapies like conventional exercise in the management of psychiatric conditions. (4) ***CIM therapies are a source for innovative interventions and have a growing, quality empirical***

research basis. Empirical research into these therapies is exploring their clinical efficacy in trials and the neurobiologic mechanism underlying their effects. (5) ***CIM therapies can help to lower the utilization of conventional medicines***. CIM therapies may be used in replacement of conventional medicines for the management of mild illnesses (e.g., relaxation techniques for mild anxiety instead of benzodiazepines which are known to have adverse effects on older adult populations). (6) ***CIM therapies can have significant side effects and CIM–drug interactions***. With the high use in geriatric populations, this is, therefore, an important reason why clinicians need to better understand CIM therapies.

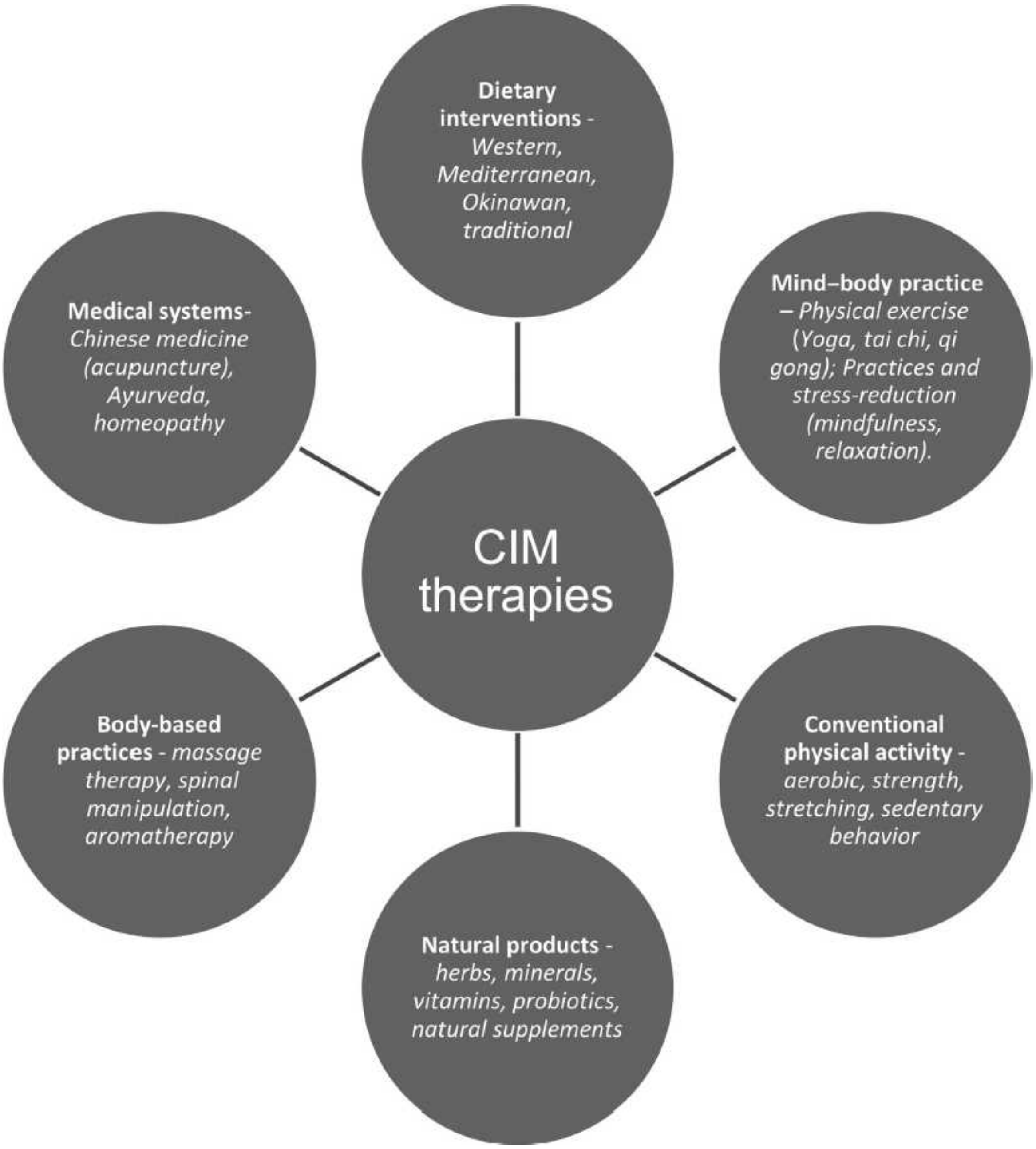
Dietary interventions - Western, Mediterranean, Okinawan, traditional
Mind–body practice – Physical exercise (Yoga, tai chi, qi gong); Practices and stress-reduction (mindfulness, relaxation).
Medical systems- Chinese medicine (acupuncture), Ayurveda, homeopathy
CIM therapies
Body-based practices - massage therapy, spinal manipulation, aromatherapy
Conventional physical activity - aerobic, strength, stretching, sedentary behavior
Natural products - herbs, minerals, vitamins, probiotics, natural supplements

FIGURE 55.2b–1. Types of complementary, alternative, and integrative therapies.

This chapter aims to provide an overview of the importance of understanding CIM therapies in geriatric psychiatry.

THE EPIDEMIOLOGY OF COMPLEMENTARY AND INTEGRATIVE MEDICINE USE

To appreciate the importance of CIM therapies in geriatric psychiatry, it is important to explore the epidemiology of and drivers behind usage. This section will explore these topics in general, aging and psychiatric populations.

CIM Use in Geriatric Populations

CIM usage in general adult populations appears to be high. However, there is variability within studies. In a joint survey conducted by the American Association of Retired People (AARP) and the NCCIH, the use of CIM was reported to have increased rapidly, exceeding a prevalence of 53% among those aged 50 years and above. Previously, studies had shown that older adults were more likely to use prayer for health and spiritual practices than younger adults and were less likely to use other forms of CIM. However, a recent study comparing the baby boomers cohort (1946–1964) to the pre-boomer cohort (those born prior to 1946) demonstrated that a higher proportion of baby boomers used CIM (27.7%) than pre-boomers (16.4%). This data, therefore, suggest that given baby boomers are already using 30% more CIM than their older counterparts, with the aging baby boomer population across the world and associated increases in chronic and painful conditions, CIM therapies will only continue to increase. By 2030, more than 6 in 10 baby boomers will be managing at least one chronic disease. National estimates of CIM use were highest in East Asian countries such as Japan, South Korea, and Malaysia. Studies consistently show that "average" CIM users are more likely to be female, better educated, middle aged, and report poorer health status than nonusers.

The use of CIMs appears to be influenced by a complex mix of economic and sociocultural factors. First, it appears CIM use is tied to the movement for self-care. Therefore, CIM use may provide people with a better sense of control and responsibility over their illness. CIM therapies such as tai chi and yoga appear to improve patients' quality of life (QoL). CIMs can be used in addition to, or in replacement of, pre-existing pain medications. CIM therapies can offer a less harmful way to manage chronic pain by avoiding unnecessary complexity, polypharmacy, and side effects. In low-resource environments where access to biomedical services is poor, there is evidence of a pervasive reliance on traditional healers, even for severe diseases. Delayed medical care—more likely in low-resource environments—is associated with CIM use in the United

States. In high-resource settings where biomedical services are more readily available, a significant amount of CIM therapies are used for health promotion and illness prevention. CIM therapies are often used as adjuncts to conventional therapies for severe diseases like cancer and to self-manage chronic health complaints like lower back pain. It appears CIM usage is driven more so by values and beliefs rather than dissatisfaction with biomedicine. The influence of increasing marketing of CIM therapies is anecdotal but must be considered.

Persons with Mental Illnesses

Persons with mental illnesses appear to use higher numbers of CIMs than the general population. In national samples of general adult populations, persons with depression and anxiety symptoms use CIM therapies more than persons without such disorders. One study indicated that over half of such persons reported CIM use within the prior year. The presence of psychiatric illness also increases the odds of visiting a CIM practitioner. The use of CIM therapies for anxiety and depression was often in conjunction with traditional treatments, raising concerns about how these treatments interact with conventional psychiatric treatments. Using the 2002, 2007, 2012, and 2017 National Health Interview Surveys, there has been a robust increase in the use of yoga, tai chi, and qigong from 2002 to 2017 (5.8% to 14.5%, respectively) by the general population typically for stress, anxiety, and depression. Relatively less information exists specifically about CIM use among older adults with mental illnesses. Therefore, clinicians treating older adults need to be informed of the efficacy and potential side effects of various CIM therapies to better assist their patients in making health care decisions.

Mind–Body Therapies

Mind–body therapies (MBT) are varied and are also rising in popularity, according to the 2017 National Health Survey, with a 5% to 10% increased use of yoga, meditation, and chiropractic care. A growing body of research evidence supports the use of MBTs (including meditation, yoga, tai chi, and qigong) as minimally invasive, cost-effective approaches for the management of stress, late-life mood, and cognitive disorders. In contrast to pharmacologic approaches, MBTs aim to teach patients life-long skills that may continue to confer benefits long after formal training has ended.

MBTs can be best divided into physical and nonphysical exercise-type practices. Mind–body physical exercise practice includes yoga, tai chi, and qigong. Mind–body nonphysical exercise practice includes relaxation, mindfulness, meditation, and acceptance therapies. Mind–body physical practices are often defined as “physical activity executed with a profound inwardly directed contemplative focus.” Mindful physical activity contains the following key elements: (1) a noncompetitive, nonjudgmental meditative component; (2) mental focus on muscular movement and proprioceptive awareness combined with a low to moderate level of muscular activity; (3) centered breathing; (4) a focus on anatomic alignment (i.e., spine, trunk, and pelvis) and proper physical form; and (5) energy-centric awareness of the individual flow of intrinsic energy, vital life force, qi, etc. Mind–body nonphysical practice does not emphasize the physical activity component of the above. More recently, MBTs such as mindfulness yoga have taken a combined approach, integrating yoga practices with interventions from mindfulness-based stress reduction (MBSR) programs. In cases of mind–body practices, individuals go through training to develop life-long skills for self-regulation in the face of adversity and stress and health-promoting practices.

Mind–Body Physical Practices. Mindful exercise interventions have shown promise in addressing depressive symptoms in older adults. A more recent review including 1,282 older adults aged 62 to 83 years with depressive symptoms, frailty, or chronic medical illnesses showed that qigong exercise resulted in significantly improved physical ability compared with active control or usual care (standardized mean difference [SMD] = 1.00 and 1.20, respectively). Although qigong exercise showed favorable effects on depression, balance, and functioning, the overall effects did not reach statistical significance. However, no significant adverse events were reported, suggesting that the qigong exercise may be an option for older adults to improve physical ability, functional ability, and balance and lessen depression and anxiety.

The practice of yoga typically benefits from instruction by expert instructors and requires participant commitment to multiple weekly sessions and ongoing maintenance practice for maximal benefit. Yoga is commonly used in combination with other treatments for depression, anxiety, and stress-related disorders. A prior review of published randomized controlled trials (RCTs) of yoga for depression in adults revealed that all trials found benefits. However, the trial methodologies have generally been weak with a lack of blinding, short duration of the intervention, variable outcome measures, limited information about subjects, randomization procedures, compliance, and dropout rates. Comparative studies of yoga have likewise been limited to older adults. In an RCT of 98 older adults, aged 65 to 92 years, participants were randomly assigned to chair yoga, chair exercise, and control groups and assessed preintervention, postintervention, and 1-month follow-up. Yoga participants improved more than both exercise and control participants with regard to depression, anxiety, well-being, anger, general self-efficacy, and self-efficacy for daily living. Another study of 69 older adults in India did compare the impact of yoga to Ayurveda or a wait-list control condition on sleep and depressive symptoms. Participants in the yoga group practiced physical postures, relaxation techniques, regulated breathing, and devotional songs and attended lectures for more than 7 hours a week during the course of the 6-month trial. The practice of yoga significantly impacted the quality of sleep and level of depressive symptoms when compared to the two control conditions, neither of which demonstrated significant effects. In particular, depressive symptoms, as measured by the short form of the ***Geriatric Depression Scale (GDS)***, decreased in the yoga group from a baseline average of 10.6 to 8.1 by 3 months and 6.7 by 6 months. The average time to fall asleep decreased in the yoga group by 10 minutes, while the total number of hours slept increased by 60 minutes and resulted in a greater feeling of being rested after 6 months. A recent study compared the effectiveness of laughter yoga, group exercise therapy, and control in decreasing depression in older adult women (60 to 80 years). In this study, 70 depressed women were chosen if their Geriatric Depression Score was >10. This study went on for 10 sessions and found a significant improvement in depression scores with both yoga and group exercise therapy as compared with control. While the evidence supporting the use of these therapies in the treatment of late-life mental illnesses is not strong, there are some interesting early results that require further high-quality research.

A recent meta-analysis critically evaluated the effects of tai chi on individuals aged 60 years and older with and without cognitive impairment. In this study, 20 eligible studies with a total of 2,553 participants were identified that met inclusion criteria for the systematic review; 11 of the 20 eligible studies were RCTs, 1 was a prospective nonrandomized controlled study, 4 were prospective noncontrolled observational studies, and 4 were cross-sectional studies. The overall quality of RCTs was modest, with 3 of 11 trials categorized as high risk of bias. Meta-analyses of outcomes related to executive function in RCTs of cognitively healthy adults indicated a large effect size (ES) when tai chi participants were compared with nonintervention controls (g = 0.90; p = .04) and a moderate ES when compared with exercise controls (g = 0.51; p = .003). Meta-analyses of outcomes related to global cognitive function in RCTs of cognitively impaired adults, ranging from mild cognitive impairment (MCI) to dementia, showed smaller but statistically significant effects when tai chi was compared with nonintervention controls (g = 0.35; p = .004) and other active interventions (g = 0.30; p = .002).

A recent review of 13 RCTs examined the impact of yoga on cognition and mental health. Yoga-based interventions have had some positive evidence on attention, executive functions, and memory among cognitive variables. These studies are limited by the small number and the need for more validated modules of yoga intervention with longer-term follow-up and standard study methodology in future trials. In addition, the impact of yoga on depression was comparable to active control among older adults.

For patients with mild to moderate Parkinson disease (PD), a 2016 meta-analysis concluded that mind–body exercises, including yoga, dance, and tai chi, had immediate moderate-to-large beneficial associations with motor symptoms, postural instability, and functional mobility. A more recent RCT of 138 participants with mild-to-moderate PD also showed that a mindfulness yoga program was found to be as effective as stretching and resistance training exercise (SRTE) on psychological distress, physical health, spiritual well-being, and health-related QoL.

RECOMMENDATION AND SUMMARY. We recommend the use of mind–body therapies as a stand-alone or adjunctive treatment in the management of geriatric disorders based on individual preference and the severity of the disease. The majority of studies to date have focused on younger and middle-aged adults, but the overall impact of MBTs has been positive despite the proposed mechanisms by which MBTs improve mood, sleep, and well-being in geriatric populations being less understood. Further high-quality data is needed to better inform these recommendations oriented toward promoting well-being and resilience in this vulnerable population.

Mind–Body Nonphysical Practices and Stress-Reduction Techniques. Existing evidence suggests that, on average, both meditative movement (e.g., yoga) and multicomponent mindfulness-based interventions can be as effective as other active treatments for disorders such as major depression, anxiety, and substance use disorder. Results of studies with healthy adults suggest that multicomponent interventions such as MBSR have the potential to increase empathy, self-control, self-compassion, relationship quality, and spirituality, as well as decrease rumination. Research has also begun to evaluate the neurobiologic mechanisms by which meditative therapies enhance resilience to mental health disorders, and several promising mechanistic domains (neural, hormonal, immune, cellular, and cardiovascular) have been identified.

Relaxation techniques, such as breathing exercises, guided imagery, and progressive muscle relaxation, are designed to produce the body's natural relaxation response. These relaxation therapies are shown to lower stress levels in older adult subjects and so are recommended for clinical use. Two of the most commonly available and widely studied forms in research are MBSR and mindfulness-based cognitive therapy (MBCT). These mindfulness-based treatment protocols are usually brief (typically eight in-person individual or group sessions), multicomponent interventions incorporating both education and practice.

An RCT including 103 adults aged 65 years or older with an anxiety or depressive disorder (diagnosed according to DSM-IV criteria) and subjective neurocognitive difficulties were recruited and randomly assigned in groups of 5 to 8 to MBSR or a health education control condition matched for time, attention, and credibility. The primary outcomes were memory (assessed by immediate and delayed paragraph and list recall) and cognitive control (Delis–Kaplan Executive Function System Verbal Fluency Test and Color Word Interference Test). Other outcomes included clinical symptoms (worry, depression, anxiety, and global improvement). The mindfulness group experienced greater improvement on a memory composite score ($\boldsymbol{p}$ = .046), measures of worry ($\boldsymbol{p}$ = .042), and depression ($\boldsymbol{p}$ = .049) at posttreatment and on worry ($\boldsymbol{p}$ = .02), depression ($\boldsymbol{p}$ = .002), and anxiety ($\boldsymbol{p}$ = .002) at follow-up. However, the groups did not differ on changes in cognitive control. Participants receiving MBSR were more likely to be rated as much or very much improved as rated by the Clinical Global Impressions-Improvement scale (47% vs. 27%, ***p*** = .03). Cortisol levels decreased to a greater extent in the mindfulness group, but only among those participants with high baseline cortisol.

In another study, **13 participants with PD** experienced depression, anxiety, stress, or difficulty coping with PD. Six completed the full, 8-week MBSR course. The MBSR group reported significant improvements in levels of depression, anxiety, and stress at weeks 8 and 16, as measured by the Depression Anxiety and Stress Scale, short version (DASS-21). This small study supports previous preliminary findings that mindfulness-based interventions could benefit people with PD experiencing nonmotor symptoms. Further research using larger sample sizes, a control group, and a longer follow-up period is required.

Another RCT examining the effects of MBSR on chronic insomnia demonstrated that the MBSR program could be a beneficial treatment for chronic insomnia in adults aged 75 years and older. Sixty participants were included, ages 75 and older, over with chronic insomnia. They were randomly assigned to the 8-week MBSR group or the wait-list control group. Assessments using the Pittsburgh Sleep Quality Index (PSQI), Self-rating Anxiety Sale (SAS), and GDS were taken at baseline and posttreatment. There was a significant time × group interaction for the PSQI global score (***p*** = .006); the MBSR group had a decrease in the PSQI global score (Cohen’s d = 1.12), while the control group did not (Cohen’s d = –0.06). Among the sleep PSQI components, there was a significant time × group interaction for daytime dysfunction (***p*** = .048; Cohen’s d of the MBSR group was 0.76, while Cohen’s d of the control group was -0.04). There was no significant time × group interaction for the anxiety SAS score (***p*** = .116). However, for the depression GDS, there was a significant time × group interaction (***p*** = .039; the Cohen’s d value for the MBSR group was 1.20, and it was 0.12 for the control group).

Another randomized pilot trial with MBSR aimed to test the safety and feasibility of MBSR in older adults with MCI. The study found that adults with MCI can safely participate and adhere to an MBSR program. The qualitative interviews revealed that most enjoyed the program and described improved mindfulness skills, well-being, interpersonal skills, acceptance/awareness of MCI, and decreased stress reactivity. Most data suggest a trend toward improvement for measures of cognition and well-being, but a more formal, large-scale RCT is needed.

RECOMMENDATIONS AND SUMMARY. The evidence is growing for the efficacy and underlying mechanisms of complementary and integrative interventions. Given the low cost and relative safety, we recommend mind–body therapies as useful stand-alone or adjunctive therapies for the management of mild–moderate neuropsychiatric and cognitive symptoms (e.g., psychological distress, anxiety, depression, pain, insomnia, and MCI) and physical health (e.g., improving balance, strength, coordination). Care must be taken to create age- and physical- and mental-function–appropriate prescriptions. Our recommendations are based on existing evidence and clinical experience, as well as growing patient preferences for the use of integrative medicine.

Conventional Physical Activity

Conventional physical activity is one of the most promising therapy for the treatment and prevention of age-related cognitive decline and dementia. Importantly, there are a variety of subtypes of physical activity, from aerobic activity to resistance, stretching, and toning. Currently, the American College of Sports Medicine recommends that exercise programs for older adults include both aerobic and nonaerobic physical activities, such as resistance training, balance training, and stretching, for optimal general health. See Table 55.2b–1 for the latest clinical guidelines on physical activity for older adults.

Table 55.2b–1.
Clinical Suggestions for Geriatric Populations

If 65 years or older, generally fit, and have no limiting health conditions

- At least 150 minutes of moderate-intensity aerobic activity (i.e., brisk walking) every week and muscle-strengthening activities on 2 or more days a week that work all major muscle groups (i.e., legs, hips, back, abdomen, chest, shoulders, and arms).

OR

- 75 minutes of vigorous-intensity aerobic activity (i.e., jogging or running) every week and muscle-strengthening activities on 2 or more days a week that work all major muscle groups.

OR

- An equivalent mix of moderate and vigorous-intensity aerobic activity and muscle-strengthening activities on 2 or more days a week that work all major muscle groups.

Other comments

- If there is a limiting condition, it is recommended the health professional and patient consult a qualified exercise physiologist.
- 10 minutes at a time is fine for physical activity.
- The following activities count as muscle-strengthening activities: yoga, heavy gardening.

Adapted from U.S. Department of Health and Human Services. ***Physical Activity Guidelines for Americans***. 2nd ed. U.S. Department of Health and Human Services; 2018. https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf

Conventional physical exercise has received considerable attention for its potential role in the treatment of depression, based initially on the results of observational studies. A large (***n*** = 6,828) cross-sectional study of men and women aged 20 to 88 years found that aerobic fitness and physical activity levels, as measured by a maximal treadmill test and self-report, respectively, were significantly related to the degree of depressive symptoms and emotional well-being. Interestingly, these relationships occurred in a graded, dose-dependent fashion, with progressively higher levels of fitness associated with increasingly lower levels of depressive symptoms and higher degrees of self-reported well-being. For physical activity levels, a similar dose–response relationship was observed;

however, the greatest benefit was seen at between 11 and 19 miles per week of walking, jogging, or running.

The effects of aerobic exercise in depressed adults aged 50 years and older were examined and compared to either monotherapy with sertraline as a standard antidepressant or a combination of both approaches. The study found that after 16 weeks, all three groups demonstrated equivalent significant reductions in depressive symptoms as measured by scores on the ***Hamilton Depression Scale*** and the ***Beck Depression Inventory***. Furthermore, a 10-month follow-up study showed that participants in the exercise group who had improved to the point of remission during the initial study showed lower rates of relapse than those taking antidepressants and independent continuation of exercising related to a reduced probability of being depressed at the end of the trial. Another study of exercise for older adults found that 10 weeks of weight-bearing exercise classes provided as an adjunctive treatment significantly lessened depressive symptoms in a group of older adults who had not responded to antidepressant therapy alone when compared to a health education course. A systematic review examining the role of physical exercise in the treatment of depression in older adults was conducted in 2009. This study noted 11 RCTs with a total of 641 participants aged >60 years. The study noted short-term (0 to 3 months) positive outcomes were found in nine studies, although there was variation in the type, intensity, and duration of exercise. The efficacy of exercise in the medium and long term (3 to 12 months and >12 months, respectively) was less clear. Issues were noted, with only five studies showing appropriate allocation of concealment, five studies showing intention to treat analysis, and blinding occurred in seven studies. The effectiveness of a number of lifestyle interventions promoting increases in conventional physical exercise has been recently reviewed. Overall, interventions were relatively successful in initiating increased levels of physical activity in the short run; however, maintenance of these gains was diminished over time, and there was no convincing evidence that behavioral reinforcement strategies were beneficial. Another review of exercise interventions found that while many older adults increase physical activity levels in response to a wide variety of interventions, the amount of increase is seldom enough to convincingly impact health outcomes in a positive manner.

In a recent randomized controlled study, 28 participants with MCI and 28 participants with Alzheimer disease (AD) were randomized to a control group and exercise group and evaluated before and 3 months after the intervention on physical and cognitive abilities. MCI is a well-recognized risk factor for dementia and represents a critical window of opportunity for intervening and altering the trajectory of cognitive decline. The intervention included a 12-week multimodal physical exercise program on global cognition, executive function, and mobility in older adult people with MCI or AD. Results in this small study suggest that the 12-week multimodal physical exercise program contributed to improvements in the mobility and executive function of older adult individuals with MCI but not of those with AD. Currently, several larger-scale RCTs are underway, but this suggests that physical exercise should be recommended to those in the early stages of neurocognitive disorder for more preventative measures.

Another RCT examined the efficacy of resistance and aerobic training in the improvement of cognitive functions in subjects with MCI. The study occurred over 6 months and involved 86 community-dwelling women aged 70 to 80 years. Physical activity protocols included twice-weekly resistance training, twice-weekly aerobic training, or twice-weekly balance and toning (BAT) training (i.e., control). Resistance training improved selective attention/conflict resolution and associated memory compared with BAT. In contrast, aerobic physical activity improved general balance and mobility and cardiovascular capacity. The study also found aerobic physical activity improved verbal memory, and both resistance and aerobic physical activity improved spatial memory. A recent stringent meta-analysis has examined the efficacy of exercise on cognition in older adults with MCI. MCI was diagnosed on documented criteria or via the Mini-Mental State Examination (MMSE). Fourteen RCTs with 1,695 participants aged 65 to 95 years were utilized. These studies had a duration of 6 to 52 weeks. Overall, 42% of effect sizes (ESs) were potentially clinically relevant (ES > 0.20), with only 8% of cognitive outcomes statistically significant. The meta-analysis revealed negligible but significant effects of exercise on verbal fluency (ES: 0.17 [0.04, 0.30]). No significant benefit was found for additional executive measures, memory, or information processing. The authors critically appraised RCT methods and concluded on a moderate quality, with the majority of trial samples being too small for sufficient power. There is clearly some effect for exercise at moderate to high levels of exertion. In addition, nowadays, there is emerging evidence for the effect of low levels of physical activity or “sedentary behaviors” like sitting. Physical inactivity may increase the risk of AD by 82%. Approximately 13% of AD cases worldwide may be attributable to sedentary behavior. A 25% reduction in sedentary behavior could potentially prevent more than 1 million AD cases globally. Sedentary behaviors may contribute to the risk of AD and dementia by an increased risk of cardiometabolic risk factors—diabetes, hypertension, obesity—diseases associated with an increased risk of dementia.

Our own studies of yoga and memory training in older adults with MCI and tai chi in older adults with depression have demonstrated changes in mood, psychological resilience, brain connectivity, and memory and executive functioning, leading to improved QoL.

Recommendations and Summary. Our clinical recommendation is physical activity should be encouraged for all members of society and particularly older adults. Evidence points to exercise as having a significant potential to lessen symptoms in depressed older adults and, furthermore, to prevent relapse of depression and/or development of depressive symptoms in older adults experiencing acute stressors associated with aging. We recommend physical aerobic and mindful exercise as a stand-alone therapy in the treatment of mild depression and as an adjunct in moderate depression. Recommendations for engagement are given in Table 55.2b–1. More work is required to understand if physical activity recommendations or dementia prevention are the same as generic “healthy aging” recommendations (as in Table 55.2b–1). While shorter-term interventions are useful in the research setting, care must be taken to ensure a sustained engagement in physical activity. Care must also be taken to create age- and physical-function–appropriate prescriptions, given the risks of adverse health effects (e.g., falls, cardiovascular events).

Diet and Use of Herbal and Nutritional Supplements

Several lines of investigation have pursued the role of diet and nutrition in the management of mental disorders. These include targeting increased oxidative stress and inflammation; reduced essential nutrients in psychotic and mood disorders; and gut microbiome dysfunction, all of which have been linked to mental disorders. Over half of all U.S. adults take some form of nutritional supplement. Older adults aged 70 years and older suffering from disorders such as stroke, cancer, obesity, breathing problems, and arthritis are more likely to use herbal and nutritional supplements. Despite the popular belief that natural supplements are safer, their use can be associated with important adverse effects and supplement/herb–drug interactions, which are only minimally regulated by the Food and Drug Administration (FDA). In addition, excessive consumption of certain supplements may be harmful, underreported in consumer reports or media, and underappreciated by patients and health care providers. This next section will discuss the limited evidence available for herbal and nutritional supplements.

Diet. The emerging field of nutritional psychiatry focuses on the consequences and associations between what you eat, how you feel, and, ultimately, how you behave. The focus shifts from single nutrients or supplements to a comprehensive, cohesive, and scientifically rigorous evidence base to support a shift in thinking about the role of diet and nutrition in mental health. In particular, the Mediterranean diet gained attention after a large, multicenter, randomized controlled PREDIMED study showed that adherence to a Mediterranean diet could be a primary prevention strategy for cardiovascular disease. The Mediterranean diet consists of plentiful use of olive oil; high consumption of fruit, vegetables, legumes, cereals, and nuts; regular but moderate intake of wine (especially red wine) with meals; moderate consumption of fish, seafood, fermented dairy products (yogurt and cheese), poultry and eggs; and limited consumption of red and processed meats and sweets. Two RCTs found significant improvement in depression with Mediterranean diet interventions in adults compared to social support controls, with the dietary support group demonstrating significantly greater improvement between baseline and 12-week Montgomery–Åsberg Depression Rating Scale (MADRS) scores compared to social support alone. There was also a noted greater reduction in depression (p = .03) and improved mental health QoL scores (p = .04) at 3 months with sustained effects at 6 months follow-up with Mediterranean and fish oil supplementation compared to social support alone. Another observational cohort study examined the relationship between the Mediterranean-DASH diet on depression in older adults, demonstrating adherence to the Mediterranean-DASH diet was associated with lower rates of depression while Western diets had a positive association with depression. Beyond mood effects, the Mediterranean diet has also been shown to be beneficial for cognition. One of the first RCTs examining dietary patterns on global cognitive function demonstrated nutritional intervention with the Mediterranean diet supplemented with either extra virgin olive oil (EVOO), or mixed nuts is associated with improved global cognition. It is likely that a combination of emphasizing foods that contain multiple beneficial nutrients, as well as a reduction in saturated and unsaturated fats, has a stronger, combined effect on mental disorders than single supplementation alone. Despite the popularity of supplements and herbals, the data has been less consistent than the stronger, reproducible evidence of certain dietary patterns in supporting geriatric mental health.

Natural Products. Natural products may include herbs, minerals, natural supplements, vitamins, and probiotics. A range of these products has been explored with regard to the treatment and prevention of cognitive impairment. There are too many of these products to thoroughly review in this chapter, so only the main products, with a larger body of research and/or epidemiologic usage, are reviewed here with further reading recommended.

ST. JOHN WORT. St. John Wort (SJW) (***Hypericum perforatum***) is a wildflower that has been used for medicinal purposes for thousands of years. Various standardized extracts of this natural dietary supplement have been studied for the treatment of depression. There are around 40 published clinical trials comparing SJW with various antidepressants, including tricyclic antidepressants (TCAs) and selective serotonin reuptake inhibitors (SSRIs), as well as many systematic reviews and meta-analyses. A Cochrane review by Linde et al. examined 5,489 patients with 18 trials comparing SJW with placebo and 17 trials comparing SJW with standard antidepressants. Overall, the results were heterogeneous, but SJW was found to be superior in efficacy to placebo and equivalent to standard antidepressants. Another meta-analysis led by Ng et al. studied 3,808 patients and 27 clinical trials comparing SJW with SSRIs for the treatment of mild to moderate depression. SJW demonstrated comparable response (pooled risk ratio (RR) 0.983, 95% confidence interval [CI] 0.924–1.042, $p < .001$) and remission (pooled RR 1.1013, 95% CI 0.892–1.134, $p < .001$) rates with a significantly lower discontinuation/dropout rate (pooled OR 0.587, 95% CI 0.478–0.697, $p < .001$) rate. In addition, the pooled SMD from baseline scores on the

Hamilton Depression Rating Scale (HAM-D) supported the efficacy of SJW in ameliorating depressive symptoms. In older adults, the available evidence for the use of SJW as an antidepressant is limited. A small RCT comparing SJW (800 mg extract LoHyp-57) and fluoxetine (20 mg) in moderately depressed older adults (60 to 80 years old) found equivalent efficacy for these agents in reducing depression over 6 weeks. Therefore, there is a strong evidence basis for SJW in treating the general adult population suffering from depression.

SJW contains more than 150 chemicals, some of which have psychotropic properties. Hypericin and hyperforin are two compounds that are best understood, and SJW preparations tend to be standardized to one of these compounds. SJW's mechanism of action appears to be multifactorial and includes interaction with the hypothalamus–pituitary–adrenal (HPA) axis leading to decreased cortisol levels. In addition, SJW has a slight monoamine oxidase inhibitor activity, which needs to be taken into consideration when combined with SSRIs as it could result in serotonin syndrome. Other mechanisms include direct serotonergic activity, decreased serotonin receptor density, and synaptic neurotransmitter reuptake.

Classic doses of SJW range from 300 to 1,800 mg/day, which are usually divided into three times daily basis. The standard therapeutic dose is typically 900 mg daily. SJW tends to be well tolerated. Side effects commonly include dry mouth, constipation, dizziness, phototoxicity, and cycling to mania in bipolar disorder patients. SJW is also an inducer of the cytochrome P-450 3A4 liver enzyme system, which leads to decreased activity of multiple medications such as warfarin, methadone, benzodiazepines, zolpidem, olanzapine, clopidogrel, cyclosporine, oral contraceptives, theophylline, digoxin, indinavir, Camptosar, irinotecan, phenprocoumon, and verapamil. Thus, extreme caution is needed for patients with HIV who are treated with protease inhibitors, cancer patients who are receiving chemotherapy, and transplant patients who are taking immunosuppressive drugs while using SJW.

Thus, the evidence concludes that SJW can be used as a natural remedy for the treatment of mild to moderate depression. For an older adult who would rather not take conventional western medications, SJW can be considered an alternative natural treatment. However, all patients using SJW should be informed about potential side effects and be monitored for drug–drug interactions.

S-ADENOSYL-L-METHIONINE. S-adenosyl-L-methionine (SAMe) is a naturally occurring compound normally synthesized in the body formed from the amino acid L-methionine as part of a multistep metabolic pathway involving folic acid and vitamin B_{12}. SAMe is thought to play a role in depression since it is a necessary cofactor for the synthesis of serotonin, norepinephrine, and dopamine. In addition, cerebrospinal fluid levels of SAMe have been shown to be decreased in depressed patients, and levels rise in response to treatment.

The literature so far supports the efficacy of parenteral and oral SAMe preparations as compared to placebo. In addition, SAMe was deemed to be equivalent to TCAs in terms of efficacy. Over 50 clinical trials using SAMe for depressive disorders, including 19 RCTs, 17 open-label trials, and 21 comparison trials with other antidepressants, exist. Among the trials, the oral doses of SAMe ranged from 400 to 1,600 mg/day, with most trials showing benefits with dosing between 800 and 1,600 mg/day. However, dosages up to 3,600 mg/day have been used with success and safely in older populations, supporting the notion that higher dosages may be necessary for clinical benefit.

A meta-analysis by Hardy et al. comparing SAMe to placebo showed a significant improvement of approximately six points in the score of the HAMD. Comparisons with other antidepressants showed no differences in risk ratios and ESs, which supports equivalence between SAMe and other agents. However, there was some concern regarding publication bias, small samples, heterogeneity of the doses used, and delivery systems, and thus, the authors were cautious regarding their interpretations of the results.

A recent systematic review by Sharma et al. determined that relatively few new studies of SAMe have existed since 2002, thereby noting that the meta-analysis by Hardy et al. remains highly relevant. Furthermore, they found promising but limited evidence of efficacy and safety to support the use of SAMe as monotherapy and as augmentation for other antidepressants.

SAMe can be beneficial for depression. However, the data are lacking for the geriatric population. It is generally well tolerated, yet important safety concerns such as increased anxiety, mania, and hypomania have been reported in patients with bipolar disorder. Patients with bipolar disorder should not take SAMe without concomitant administration of mood stabilizers coupled with clinical monitoring. Other side effects may include insomnia, anorexia, dry mouth, sweating, dizziness, and nervousness.

SAMe has a faster onset of action than standard antidepressants, with some patients exhibiting improvement in depressive symptoms within 2 weeks. It could be combined with antidepressants, thereby accelerating the onset of action in many cases. SAMe is a particularly good choice in patients with methylene tetrahydrofolate reductase (MTHFR) deficiency, as it would bypass the steps in folate metabolism that are dependent on this enzyme. An ideal candidate for a trial of SAMe could include a mildly symptomatic patient who would not have serious effects if there was a delay in treatment or a person who has failed multiple trials of conventional antidepressants and/or is unable to tolerate side effects.

VALERIAN. Valerian (***Valeriana officinalis***) is an herbal sedative and mild hypnotic that has been used for over a thousand years worldwide. Its tranquilizing effect is due to its active compounds, including valepotriates and sesquiterpenes, which have gamma-aminobutyric acid (GABA)-ergic activity and function similarly to barbiturates and benzodiazepines. Valerian also appears to promote natural sleep after a few weeks of regular use instead of treating insomnia acutely.

There are over 40 published controlled trials of valerian that include both healthy and symptomatic individuals. Some studies have shown that valerian is comparable in efficacy to that of benzodiazepines and is associated with lesser side effects and no tolerance. There are a few meta-analyses and systematic reviews that are not supportive of its efficacy. In addition, one 4-week pilot RCT involving 36 patients with generalized anxiety disorder (GAD) treated with valerian, diazepam, and placebo showed no significant differences between the valerian and placebo groups or valerian and diazepam groups in HAM-A total scores. The inconsistent findings in these trials could be due to the valerian’s powerful smell that could lead to unmasking. However, this may be overcome by combining the inactive compounds of the valerian into placebos, which is more recently practiced.

The recommended dose of valerian is 450 to 600 mg before bedtime, with little benefit from higher doses. Overall, it is very well tolerated. The most common side effects include gastrointestinal (GI) side effects and headaches. No increased sedation or hangover effect in the morning after its use. Rare toxic reactions include blurry vision, dystonias, hepatotoxicity, delirium, and withdrawal. It has been noted to be safe for older adults. In conclusion, valerian is a safe and effective hypnotic that decreases sleep latency and improves sleep quality and can be used in older adults.

GINKGO BILOBA SUPPLEMENTS. Ginkgo biloba leaf extract is among the most widely sold herbal dietary supplements in the United States. Its purported biologic effects include scavenging free radicals, lowering oxidative stress, reducing neural damage, reducing platelet aggregation, and anti-inflammatory, antitumor, and antiaging activities. Ginkgo's primary indication is for enhancing cognitive function, including memory and abstract thinking, in individuals with organic brain diseases such as Alzheimer dementia. No standards or guidelines regulate the constituent components of Ginkgo biloba leaf extract, nor are exposure limits imposed. Over 30 placebo-controlled, double-blind RCTs of ***Ginkgo biloba*** in patients with various types of dementia have yielded contradictory results. Of the studies that revealed cognitive improvement, the effect was minor and did not last more than 6 months. One RCT demonstrated the efficacy of once-daily–dosed Ginkgo biloba (240-mg extract EGb 761) in outpatients aged 50 years or older with mild to moderate AD or vascular dementia (VaD) over 24 weeks. Both types of dementia showed improvement in cognition and neuropsychiatric symptoms compared to placebo. In addition, a recent meta-analysis by Hashiguchi et al. examined 9 out of 13 studies lasting 12 to 52 weeks with doses >120 mg/day in 2,381 participants. In seven out of the nine studies, Ginkgo overtook placebo. The advantage was preserved when examining its efficacy in dementia types, including AD, VaD, and the combination of AD and VaD. Doses of 240 mg/day appeared to be most effective with good tolerability. Ginkgo has been compared to cholinesterase inhibitors and was noted to be better tolerated, albeit less effective. The evidence does not suggest Ginkgo biloba has any clear preventative effects on dementia. In a trial of AD prevention in older adults aged 72 to 96 years, 120 mg of ***Ginkgo biloba*** twice a day was not effective in reducing the overall incidence of dementia in older adults with normal cognition or MCI. Therefore, ***Ginkgo biloba*** will not prevent the development of AD but may have a minor impact on cognitive decline in those already afflicted with dementia.

As with other natural products, Ginkgo biloba comes with risks. It is recommended to use caution when using ***Ginkgo biloba*** in those who are at increased risk of hemorrhage as it may potentiate bleeding by inhibiting platelet-activating factor (PAF). Nevertheless, a recent meta-analysis of 18 trials including patients with dementia and healthy volunteers, diabetes mellitus, and peripheral artery disease did not find an added risk of bleeding based on hemostatic outcomes. Side effects also include mild GI upset, headache, irritability, dizziness, or allergic reactions. The short-term use of Ginkgo biloba is acceptable under some conditions, dependent on clinician–patient discussions, with the potential risks being considered.

SILEXAN. Silexan is a potent anxiolytic composed of lavender oil. Its active ingredients are linalool and linalyl acetate which comprise 71% of the oil, while the remaining 29% is made up of trace amounts of over a hundred other compounds from the lavender plant. Its mechanism of action is multifactorial and includes potent inhibition of voltage-dependent calcium channels. In addition, it is an N-methyl-D-aspartate (NMDA) antagonist, a serotonin-1A agonist, and it has GABA-ergic effects. A recent randomized, double-blind controlled trial compared two doses of silexan 80 mg or 160 mg/day to paroxetine 20 mg/day and placebo in the treatment of GAD. The 10-week study showed that both doses of silexan worked significantly better than paroxetine and placebo ($p < .01$) on the Hamilton Anxiety Rating Scale as well as on secondary measures of depression, sleep quality, and overall QoL. The effect was dose dependent, with an ES of 0.4 and 0.5 for the 80 mg and 160 mg doses, respectively. A Cochrane-style meta-analysis found a medium-size effect of 0.67 for anxiety and a large-size effect (0.87) in studies that exclusively enrolled participants with GAD.

Silexan is generally well tolerated and has no known withdrawal effects, addictive potential, or known drug interactions. The most common side effect was burping. It is considered safe for consumption by the FDA and National Institutes of Health (NIH). However, several ingredients have weak estrogen-like properties leading to rare cases of breast development in adolescent boys and early breast development in young girls. Thus, its use in children under the age of 18 years should be avoided. There are currently no known studies looking at the anxiolytic effects of silexan in older adults.

OMEGA-3 POLYUNSATURATED FATTY ACID SUPPLEMENTS. Omega-3 fatty acids are a family of polyunsaturated lipids that include eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), both derived from oily fish such as salmon, mackerel, herring, and sardines, as well as alpha-linolenic acid (ALA) derived from vegetable sources such as flaxseed oil. Omega-3 fatty acids are essential for brain development and are a key component of neuronal membranes in the brain. They promote cognitive function due to their antioxidant and anti-inflammatory effects. They also have antidepressant properties and have previously been used for the treatment of major depression as adjunct therapy as well as monotherapy.

The evidence regarding omega-3 fatty acid supplementation is difficult to interpret due to the heterogeneity among studies, specifically regarding the omega-3 preparations, doses, and study design. Numerous meta-analyses examined their efficacy in the treatment of dementia. A recent Cochrane review looked at the effects of omega-3 fatty acids on the treatment of cognitive dysfunction. Six hundred and thirty-two participants with mild to moderate AD were examined over 6, 12, and 18 months. However, there was no evidence of an increased benefit from omega-3 fatty acids supplementation on cognitive function when measured at 6 months with the MMSE or the Alzheimer’s Disease Assessment Scale-Cognitive subscale. Another Cochrane database systematic review explored the effects of omega-3 fatty acids on the prevention of dementia and cognitive decline in cognitively healthy older people. Information was available from three RCTs, including 3,536 participants in total. This meta-analysis found no evidence to support a preventative effect following 24 or 40 months of intervention. Other RCTs suggest that selected patients with an MMSE score >27 were more likely to identify a positive effect of omega-3 fatty acids supplementation.

In terms of depression, a meta-analysis showed that omega-3 fatty acids supplementation with ≥60% of EPA was beneficial, with evidence supporting a role for their use. The data supporting omega-3 fatty acids supplementation in geriatric depression is less extensive. There are currently no meta-analyses available for the geriatric population; however, some RCTs looking at the antidepressant effect of omega-3 fatty acids exist. A small RCT of clinically depressed females aged 66 to 95 years living in a nursing home found that omega-3 fatty acids supplementation improved depression and QoL. These findings are consistent with previous studies on the general adult population.

Typical doses range from 1 to 2 g/day with the use of preparations ≥60% EPA in omega-3 fatty acids combinations. Potential side effects include mild nausea and a fishy aftertaste. Previous concerns regarding the increased risk of bleeding associated with omega-3 fatty acids were disproved. However, caution is recommended in patients taking anticoagulants. There have been some cases of cycling to mania in bipolar patients who were depressed, and therefore, individuals with bipolar disorder should be on concomitant mood stabilizers. More placebo-controlled trials with large samples in older adults with depression are needed to further confirm their effectiveness.

VITAMIN SUPPLEMENTS. Antioxidant compounds such as vitamins A, C, and E have a significant role in the regulation of oxidative stress, a pathway associated with neurodegeneration and cognitive impairment. However, the use of vitamin E supplementation has not been found to have a protective effect against progression from MCI to dementia or cognitive function. Likewise, vitamin B is linked to cognitive impairment during the aging process. It is considered a standard procedure to check a vitamin B_{12} level when considering a diagnosis of dementia because it is seen as a reversible cause of the cognitive decline. Moreover, various measures of vitamin B_{12}, including methylmalonate, homocysteine, and cystathionine, are associated with episodic memory deficits, slowed perceptual speed, and decreased total brain volume in adults aged 65 years and older. Interestingly, data have related elevated homocysteine levels to amyloid and glutamate toxicity and cognitive decline in animal studies. In addition, vitamin B_3 (niacin) deficiency, also referred to as pellagra, leads to the classic triad of dementia, dermatitis, and diarrhea. Therefore, it is logical that B vitamin supplementation should benefit cognition.

Table 55.2b–2.
Other Natural Products and Supplements Suggested for the Treatment of Cognitive Impairment

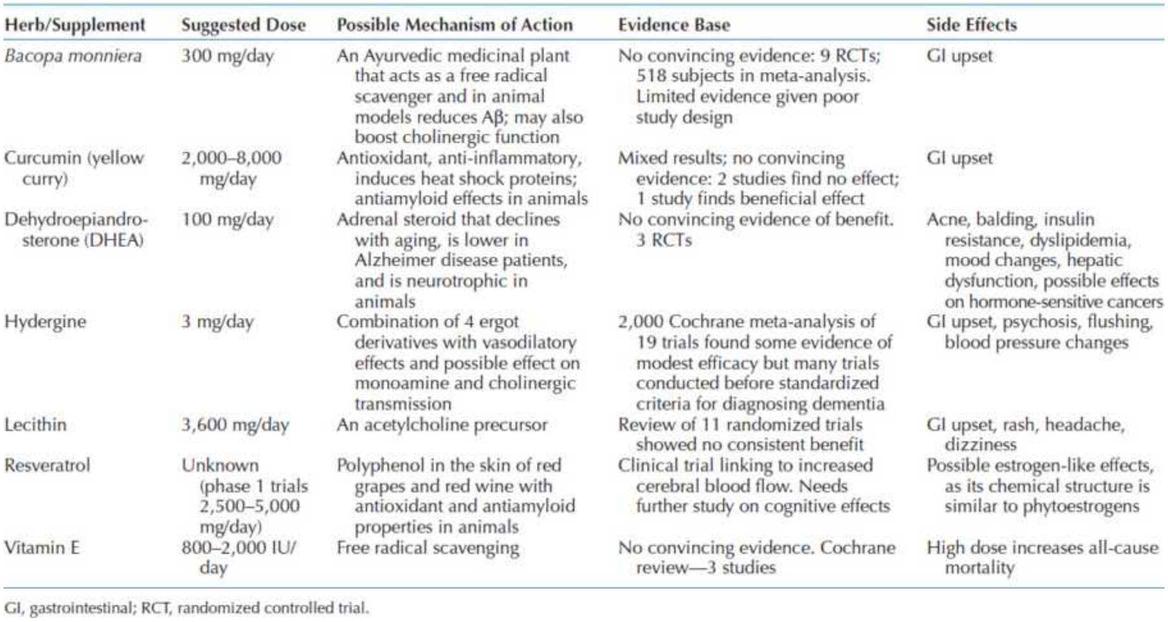

Herb/Supplement	Suggested Dose	Possible Mechanism of Action	Evidence Base	Side Effects
Bacopa monniera	300 mg/day	An Ayurvedic medicinal plant that acts as a free radical scavenger and in animal models reduces Aβ; may also boost cholinergic function	No convincing evidence: 9 RCTs; 518 subjects in meta-analysis. Limited evidence given poor study design	GI upset
Curcumin (yellow curry)	2,000–8,000 mg/day	Antioxidant, anti-inflammatory, induces heat shock proteins; antiamyloid effects in animals	Mixed results; no convincing evidence: 2 studies find no effect; 1 study finds beneficial effect	GI upset
Dehydroepiandrosterone (DHEA)	100 mg/day	Adrenal steroid that declines with aging, is lower in Alzheimer disease patients, and is neurotrophic in animals	No convincing evidence of benefit. 3 RCTs	Acne, balding, insulin resistance, dyslipidemia, mood changes, hepatic dysfunction, possible effects on hormone-sensitive cancers
Hydergine	3 mg/day	Combination of 4 ergot derivatives with vasodilatory effects and possible effect on monoamine and cholinergic transmission	2,000 Cochrane meta-analysis of 19 trials found some evidence of modest efficacy but many trials conducted before standardized criteria for diagnosing dementia	GI upset, psychosis, flushing, blood pressure changes
Lecithin	3,600 mg/day	An acetylcholine precursor	Review of 11 randomized trials showed no consistent benefit	GI upset, rash, headache, dizziness
Resveratrol	Unknown (phase 1 trials 2,500–5,000 mg/day)	Polyphenol in the skin of red grapes and red wine with antioxidant and antiamyloid properties in animals	Clinical trial linking to increased cerebral blood flow. Needs further study on cognitive effects	Possible estrogen-like effects, as its chemical structure is similar to phytoestrogens
Vitamin E	800–2,000 IU/day	Free radical scavenging	No convincing evidence. Cochrane review—3 studies	High dose increases all-cause mortality

GI, gastrointestinal; RCT, randomized controlled trial.

The evidence, however, does not support a robust response in cognition with B vitamins. A multicenter RCT of adults over 50 years old with mild to moderate AD and normal vitamin B_{12} and B_9 levels showed no benefit in cognition with supplementation. Another RCT in China failed to show an improvement with supplementation in cognition with older participants, age 60 years or older, with either AD or VaD and normal vitamin levels. However, the evidence appears to be more promising in individuals without cognitive impairment. A recent RCT of older adults, aged 60 to 74 years, without dementia and normal vitamin B_{12} and B_9 levels showed a benefit in immediate and delayed recall with supplementation. The evidence does not support vitamin B supplementation in treating dementia when levels are normal. However, some data support supplementation in improving cognition prior to the development of dementia. In addition, such use of supplemental vitamins may come at a risk. The Iowa Women’s Health Study demonstrated an increase in mortality with the use of various dietary supplements, including vitamins B_6 and B_9. This study supports the fact that it is important for patients and clinicians to acknowledge that being natural is not always safer than “conventional” medicine.

For further details on other utilized CIM therapies for the treatment and prevention of age-related cognitive decline, refer to Table 55.2b–2.

Table 55.2b–3.
Other Natural Products and Supplements Suggested for the Treatment of Late-life Depression

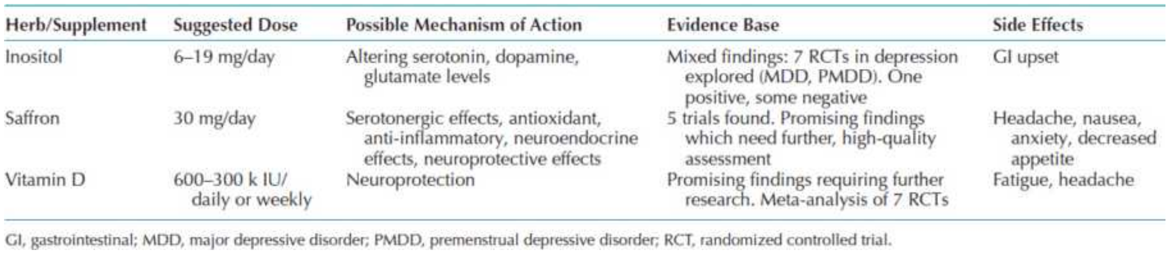

Herb/Supplement	Suggested Dose	Possible Mechanism of Action	Evidence Base	Side Effects
Inositol	6–19 mg/day	Altering serotonin, dopamine, glutamate levels	Mixed findings: 7 RCTs in depression explored (MDD, PMDD). One positive, some negative	GI upset
Saffron	30 mg/day	Serotonergic effects, antioxidant, anti-inflammatory, neuroendocrine effects, neuroprotective effects	5 trials found. Promising findings which need further, high-quality assessment	Headache, nausea, anxiety, decreased appetite
Vitamin D	600–300 k IU/ daily or weekly	Neuroprotection	Promising findings requiring further research. Meta-analysis of 7 RCTs	Fatigue, headache

GI, gastrointestinal; MDD, major depressive disorder; PMDD, premenstrual depressive disorder; RCT, randomized controlled trial.

For other natural products and supplements which may be used for the treatment of late-life depression, see Table 55.2b–3.

Potential Side Effects and CIM–Drug Interactions. As eluded to above, when encountering natural products in a clinical environment, it is important to consider potential side effects and CIM–drug interactions. Table 55.2b–4 outlines an array of clinical findings and potential pitfalls of natural products. It is important to note that CIM use may exacerbate polypharmacy in the older adults, which is a risk factor for drug interactions, medication errors, and hospitalization.

Other Therapies

Acupuncture. Acupuncture, a component of oriental traditional medicine (OTM), is a technique in which practitioners stimulate specific points on the body—most often by inserting thin needles through the skin. Several reports of positive effects of acupuncture for cognitive and psychiatric symptoms of VaD have emerged from China, but a 2007 Cochrane review found the methodologies of most of these studies to be highly problematic, leaving no convincing evidence of positive benefit. Acupuncture, however, has been further examined for the treatment of depression and was found to produce some advantages. Prior reviews of RCTs using acupuncture to treat depression have found some evidence for benefit; however, the majority of included studies have been of relatively low quality, with placebo acupuncture treatment at times yielding similar results to treatment acupuncture. The Cochrane database review of studies examining acupuncture for depression found insufficient evidence to determine the efficacy of acupuncture compared to standard antidepressants, sham acupuncture, or wait-list controls. Another pilot study examining acupuncture as an augmentation strategy in late-life depression (LLD) showed significant overall improvement in HAM-D rating scales, well-being, and activity from baseline to the end of treatment. This suggests that acupuncture may benefit older adults if they are willing to participate actively. More clinical studies investigating the use of acupuncture are needed to further establish its efficacy.

Table 55.2b–4.
Potential CAIM–Drug Interactions Between Natural Products and Prescription Medications

Herb/Supplement	Prescription Drug(s)	Possible Interaction
Chamomile	Sedatives (or alcohol)	Oversedation
Garlic extract	Protease inhibitors	Decreased efficacy against HIV
Ginkgo biloba	Anticoagulants	Excessive bleeding
	Omeprazole (CYP2C19 substrates)	Decreased drug levels (2C19 induction)
	Antipsychotic	Priapism
Ginseng	Anticoagulants	Excessive bleeding
	Monoamine oxidase inhibitors	Mania, hypertension, headache
	Sulfonylureas, insulin	Hypoglycemia
	Stimulants	Insomnia, anxiety, mania
Kava	Sedatives (or alcohol)	Oversedation
	Levodopa	Increased Parkinsonism
Omega-3 fatty acids	Warfarin	Excessive bleeding
St. John wort	Protease inhibitors	Decreased efficacy against HIV
	Cyclosporine	Organ rejection
	Serotonergic antidepressants	Serotonin syndrome
	Oral contraceptives	Ineffective contraception
	Tetracycline	Photosensitivity
Valerian	Sedatives (or alcohol)	Oversedation
	CYP3A4 substrates	Increased drug levels (3A4 inhibition)
Vitamin E	Anticoagulants	Excessive bleeding

Aromatherapy. Aromatherapy, a therapy related to massage, has found some interesting results in small-scale studies. Aromatherapy is the use of essential oils extracted from plants for health purposes. Trials of aromatherapy have, at times, combined the application of the essential oils with a massage, while at other times, the patients simply inhaled the aromas of the oils. One study reported the combination of aromatherapy (with lavender oil, commonly proposed to have calming effects) and massage tended to yield less agitation than either condition alone. Another small study in China was positive for simple inhalation of lavender oil versus control oil in reducing agitated behaviors. The most methodologically sound study of aromatherapy used lemon balm oil (***Melissa officinalis***) in a double-blind, placebo-controlled trial for managing agitation in severe dementia and found significant benefits for aromatherapy. Thus, aromatherapy can be used for agitation in patients with dementia.

Body-Based Practices. Body-based practices may include massage therapy and spinal manipulation. The authors are unaware of any strong evidence supporting these therapies in the prevention and treatment of cognitive disorders. However, they may have merits in stress reduction, pain management, and treatment of anxiety disorders. Massage therapy includes many different techniques in which practitioners manually manipulate the soft tissues of the body. It has found some use in dementia care populations.

Medical Systems

Chinese or Oriental Traditional Medicine. Acupuncture is a part of OTM—it is a technique in which practitioners stimulate specific points on the body—most often by inserting thin needles through the skin. Several reports of positive effects of acupuncture for cognitive and psychiatric symptoms of VaD have emerged from China, but a 2007 Cochrane review found the methodologies of most of these studies to be highly problematic, leaving no convincing evidence of positive benefit.

Ayurvedic Medicine. Ayurveda, which translates as “the science of life” and has been described as “knowledge of how to live,” is a complete natural health care system that began in India more than 5,000 years ago and has been used within the context of mental health and aging for memory enhancement and anxiety reduction. Promoting immune system function through anti-inflammatory remedies, Ayurveda is widely used contemporarily in India as a system of primary health care. With the emphasis on developing an individualized regimen for patients that includes dietary, meditation, herbal, and other components, interest in Ayurveda is growing worldwide to treat a number of maladies, including depression. Through yoga and/or transcendental meditation techniques, Ayurveda promotes lifestyle change and education about how to release stress and tension. Furthermore, Ayurveda is geared toward treating chronic disorders associated with the aging process, and promising preliminary results related to the use of Ayurveda have been found in studies of depression, anxiety, sleep disorders, hypertension, diabetes mellitus, PD, and AD. Of concern regarding the quality of certain Ayurvedic products, in 2008, an NCCIH-funded study examined the content of 193 Ayurvedic products purchased over the internet and manufactured in either the United States or India. The researchers found that 21% of the products contained levels of lead, mercury, and/or arsenic that exceeded the standards for acceptable daily intake.

PERSPECTIVES ON THE DIGITIZATION OF SUITABLE THERAPIES

As we are living through the Fourth Industrial Revolution, where technology trends are accelerating, it is important to consider the role of the digitization of relevant therapies. Therapies can offer digitizable physical activity, meditation, and other forms of patient-empowerment/education tools. While some geriatric patients may shy away from using e-technologies, many are becoming high utilizers of technologies. There are a burgeoning number of research labs focused on technology innovations for geriatric patients, and a large amount of private sector investment capital is moving into this field.

Notable examples of digital therapies include the following. Researchers in Europe recently assessed the feasibility of delivering a Lifestyle-Integrated Functional Exercise program and evaluated the delivery of the intervention by use of digital technology (eLiFE) to prevent functional decline in 61- to 70-year-old adults. They conducted a multicentre, feasibility RCT run in three countries (Norway, Germany, and the Netherlands). Out of 7,500 potential participants, 926 seniors (12%) were screened and 180 participants randomized to eLiFE (***n*** = 61), aLiFE (***n*** = 59), and control group (***n*** = 60). eLiFE participants used an application on smartphones and smartwatches, while aLiFE participants used traditional paper-based versions of the same lifestyle-integrated exercise intervention. At 6 months, 87% of participants completed the ***posttest***, and 77% completed the final assessment at 12 months. Participants were willing to be part of the program, with good compliance and adherence. Despite small errors during start-up in the technologic component, intervention delivery by use of technology appeared acceptable. Another group recently explored the effectiveness of a neurobic exercise program on memory and depression among older adults residing in old age homes. The neurobic exercise program was found to be effective in reducing depression in this population.

A commitment to responsible innovation of these technologies is paramount. For example, most dementia health apps do not have a privacy policy, so bolstered safeguards and improved communication about privacy protection are needed to facilitate consumer safety and trust in the apps. Responsible innovation is an increasingly prominent initiative. A 2020 Organisation for Economic Co-operation and Development recommendation on responsible innovation in neurotechnology proposed the first international standard in this domain. This recommendation “aims to guide governments and innovators to anticipate and address the ethical, legal and social challenges raised by novel neurotechnologies while promoting innovation in the field.” It articulates “the importance of (1) high-level values such as stewardship, trust, safety and privacy in this technological context, (2) building the capacity of key institutions like foresight, oversight and advice bodies, and (3) processes of societal deliberation, inclusive innovation, and collaboration.”

FUTURE DIRECTIONS

Gaining a better understanding of the clinical effect of CIMs, as well as their safety, is clearly critical for the management of geriatric psychiatry conditions. At present, there is insufficient data on the prevalence, effectiveness, efficacy, safety, and health economic benefits of most CIM treatments. This means patients, health providers, and other stakeholders are unable to access reliable and rigorous evidence on CIMs. When considering future directions for this field, it is important to consider the most recent NCCIH strategic plan for 2021 to 2025 and the ongoing emphasis on integrative health research to bring conventional and complementary approaches together in a safe, coordinated way to improve clinical care for patients, health promotion, disease prevention, and ultimately, restoring health rather than just treating disease. The NCCIH's updated plan for 2021 to 2025 has five main objectives: (1) advance basic and mechanistic research relevant to dietary, psychological, and/or physical approaches; (2) advance research on whole person health and on the integration of complementary and conventional care; (3) foster research on health promotion and restoration, resilience, disease prevention, and symptom management; (4) enhance the complementary and integrative health research workforce; (5) disseminate objective evidence-based information on complementary and integrative health interventions. Ultimately, this expansion of the definition of integrative health to include whole person health, or helping individuals improve their health in multiple interconnected domains: social, psychological, and physiologic, including connections between organs and systems, serves as a way to bring conventional and complementary approaches together in a safe, coordinated way.

CONCLUSION

Despite the increasing use of CIM by individuals and a growing body of research examining the efficacy of CIM for treating mental disorders, research and scientific support for its efficacy for older adults remains limited for most practices. Overall, there is stronger and more consistent evidence about the benefits of dietary interventions and mind–body therapies for mood and cognitive disorders, while the evidence for most herbal and nutritional supplements remains mixed and inconclusive. The noninvasive nature of mind–body therapies and Mediterranean diet modifications also make these CIM interventions more favorable over supplementation use due to the variability in formulations available, as well as potential drug–drug interactions. However, studies have consistently faced limitations, such as small sample size, heterogeneity of outcome measures, lack of an active control group, and absence of long-term follow-up, particularly in older adults. Additional research is needed to determine predictors and biomarkers of response that would allow the identification of individuals who are more likely to respond to CIM. Implementing CIM interventions in the communities will represent the next phase of CIM research as preventive, well-being–oriented interventions.

Neuropsychiatric disorders in older adults, such as mood, anxiety, sleep, and cognitive disorders, are among the most common reasons for older adults to use CIM therapies. The use of CIM therapies in geriatric psychiatry is promising, with a growing body of empirical research. There is a wide range of research quality within this field, making firm conclusions difficult. More high-quality research is needed to establish efficacy and safety. When promoting the safety of various natural products, clinicians can offer important information to those patients who are less comfortable with "conventional" medicine and prefer alternative "natural" therapies, such as advising on drug–CIM interactions and directing them toward evidence-based indications. It is clear that CIM use is significant in older adult populations and will likely increase, so empirical research is essential.

FURTHER READINGS

Abbott R, Lavretsky H. Tai Chi and Qigong for the treatment and prevention of mental disorders. ***Psychiatr Clin North Am***. 2013;36(1):109–119.

Birtwell K, Dubrow-Marshall L, Dubrow-Marshall R, Duerden T, Dunn A. A mixed methods evaluation of a mindfulness-based stress reduction course for people with Parkinson's disease. ***Complement Ther Clin Pract***. 2017;29:220–228.

Bonura KB, Tenenbaum G. Effects of yoga on psychological health in older adults. ***J Phys Act Health***. 2014;11(7):1334–1341.

Bunchorntavakul C, Reddy KR. Review article: herbal and dietary supplement hepatotoxicity. ***Aliment Pharmacol Ther***. 2013;37(1):3–17.

Chang PS, Knobf T, Oh B, Funk M. Physical and psychological health outcomes of qigong exercise in older adults: a systematic review and meta-analysis. ***Am J Chin Med***. 2019;47(2):301–322.

Chobe S, Chobe M, Metri K, Patra SK, Nagaratna R. Impact of yoga on cognition and mental health among elderly: a systematic review. ***Complement Ther Med.*** 2020;52:102421.

Estruch R, Ros E, Salas-Salvado J, et al. Primary prevention of cardiovascular disease with a Mediterranean diet. ***N Engl J Med.*** 2013;368:1279–1290.

Eyre HA, Siddarth P, Acevedo B, et al. A randomized controlled trial of Kundalini yoga in mild cognitive impairment. ***Int Psychogeriatr***. 2017;29(4):557–567.

Gates N, Singh MAF, Sachdev PS, Valenzuela M. The effect of exercise training on cognitive function in older adults with mild cognitive impairment: a meta-analysis of randomized controlled trials. ***Am J Geriatr Psychiatry***. 2013;21(11):1086–1097.

Hashiguchi M, Ohta Y, Shimizu M, Maruyama J, Mochizuki M. Meta-analysis of the efficacy and safety of Ginkgo biloba extract for the treatment of dementia. ***J Pharm Health Care Sci***. 2015;1:14.

Jacka FN, O'Neil A, Itsiopoulos C, et al. A randomised, controlled trial of dietary improvement for adults with major depression (the 'SMILES' trial). ***BMC Med.***

2017;15:23.

Kasper S, Müller WE, Volz HP, Möller HJ, Koch E, Dienel A. Silexan in anxiety disorders: clinical data and pharmacological background. ***World J Biol Psychiatry***. 2018;19(6):412–420.

Kwok JYY, Choi KC, Chan HYL. Effects of mind-body exercises on the physiological and psychosocial well-being of individuals with Parkinson's disease: a systematic review and meta-analysis. ***Complement Ther Med***. 2016;29:121–131.

Kwok JYY, Kwan JCY, Auyeung M, et al. Effects of mindfulness yoga vs stretching and resistance training exercises on anxiety and depression for people with parkinson disease: a randomized clinical trial. ***JAMA Neurol***. 2019;76(7):755–763.

Loughrey DG, Lavecchia S, Brennan S, Lawlor BA, Kelly ME. The impact of the mediterranean diet on the cognitive functioning of healthy older adults: a systematic review and meta-analysis. ***Adv Nutr.*** 2017;8:571–586.

Mischoulon D. Popular herbal and natural remedies used in psychiatry. ***Focus (Am Psychiatr Publ).*** 2018;16(1):2–11.

Ng QX, Venkatanarayanan N, Ho CYX. Clinical use of ***Hypericum perforatum*** (St John's wort) in depression: a meta-analysis. ***J Affect Disord.*** 2017;210:211–221.

Nguyen SA, Lavretsky H. Emerging complementary and integrative therapies for geriatric mental health. ***Curr Treat Options Psychiatry.*** 2020;7:447–470.

OECD. Recommendation of the council on responsible innovation in neurotechnology. 2020. OECD/LEGAL/0457. Accessed April 12, 2021. https://legalinstruments.oecd.org/api/print?ids=658&Lang=en

Papakostas GI, Mischoulon D, Shyu I, Alpert JE, Fava M. S-adenosyl methionine (SAMe) augmentation of serotonin reuptake inhibitors for antidepressant nonresponders with major depressive disorder: a double-blind, randomized clinical trial. ***Am J Psychiatry***. 2010;167(8):942–948.

Parletta N, Zarnowiecki D, Cho J, et al. A mediterranean-style dietary intervention supplemented with fish oil improves diet quality and mental health in people with depression: a randomized controlled trial (HELFIMED). ***Nutr Neurosci.*** 2019;22(7):474–487.

Raj D, Santhi S, Sapharina GJS. Effectiveness of neurobic exercise program on memory and depression among elderly residing at old age home. ***J Complement Integr Med***. 2020;17(4).

Rashrash M, Schommer JC, Brown LM. Prevalence and predictors of herbal medicine use among adults in the United States. ***J Patient Exp.*** 2017;4(3):108–113.

Rosenfeld L, Torous J, Vahia IV. Data security and privacy in apps for dementia: an analysis of existing privacy policies. ***Am J Geriatr Psychiatry***. 2017;25:873–877.

Sharma A, Gerbarg P, Bottiglieri T, et al. S-Adenosylmethionine (SAMe) for neuropsychiatric disorders: a clinician-oriented review of research. ***J Clin Psychiatry.*** 2017;78(6):e656–e667.

Siddiqui MJ, Min CS, Verma RK, Jamshed SQ. Role of complementary and alternative medicine in geriatric care: a mini review. ***Pharmacogen Rev.*** 2014;8(16):81–87.

Silva FO, Ferreira JV, Plácido J, et al. Three months of multimodal training contributes to mobility and executive function in elderly individuals with mild cognitive impairment, but not in those with Alzheimer's disease: a randomized controlled trial. ***Maturitas.*** 2019;126:28–33.

Smith CA, Armour M, Lee MS, Wang LQ, Hay PJ. Acupuncture for depression. ***Cochrane Database Syst Rev.*** 2018;3(3):CD004046.

Strauss C, Cavanagh K, Oliver A, Pettman D. Mindfulness-based interventions for people diagnosed with a current episode of an anxiety or depressive disorder: a meta-analysis of randomised controlled trials. ***PloS One.*** 2014;9(4):e96110.

Taraldsen K, Mikolaizak AS, Maier AB, et al. Protocol for the PreventIT feasibility randomised controlled trial of a lifestyle-integrated exercise intervention in young older adults. ***BMJ Open***. 2019;9(3):e023526.

Varteresian T, Lavretsky H. Natural products and supplements for geriatric depression and cognitive disorders: an evaluation of the research. ***Curr Psychiatry Rep***. 2014;16(8):456.

White CM. Dietary supplements pose real dangers to patients. ***Ann Pharmacother.*** 2020;54(8):815–819.

Williams J, Graham C. Acupuncture for older adults with depression-a pilot study to assess acceptability and feasibility. ***Int J Geriatr Psychiatry***. 2006;21(6):599–600.

Yang YP, Wang CJ, Wang JJ. Effect of aromatherapy massage on agitation and depressive mood in individuals with dementia. ***J Gerontol Nurs***. 2016;42(9):38–46.

Zhang JX, Liu XH, Xie XH, et al. Mindfulness-based stress reduction for chronic insomnia in adults older than 75 years: a randomized, controlled, single-blind clinical trial. ***Explore (NY).*** 2015;11(3):180–185.

55.2c The Aging Brain

ROBERT J. BOLAND, M.D., AND IQBAL AHMED, M.D., F.R.C.PSYCH.

Aging is an essential factor in health and disease. Over the adult life span, the body undergoes changes that generally result in loss of functional capacity. Fundamental processes underlying aging drive these changes, and lifestyle factors and diseases can influence them. The brain is not unique in experiencing changes associated with aging that tend to decrease its capabilities. These changes occur at different levels, including clinical evaluation, risk, and protective factors, biomarkers such as brain imaging, and underlying mechanisms and pathophysiology. Although the current understanding of the mechanisms that underlie brain aging is incomplete, there has been much progress in identifying the contributions of aging and disease processes to changes in cognition, mood, and behavior in older adults. Many myths about aging have arisen, and many of us tend to view old age as a time of physical and mental decline and infirmity—however, recent research on successful aging challenges this bias. We have made much progress at defining the cognitive strengths and abilities preserved in healthy brain aging and teasing out contributions of diseases to loss of ability. By delineating the clinical features, biomarkers, and pathophysiology of conditions that are not part of healthy brain aging and impair brain function, such as Alzheimer disease (AD) and vascular pathology, we can, hopefully, develop specific treatments that target these conditions age-associated diseases disease processes. In addition, factors that may limit age-related impairment in cognition and other aspects of brain function and result in greater resilience in the face of aging are also receiving attention; these include cognitive stimulation, social and physical activity, and nutrition. This chapter provides an overview of concepts and challenges in studying and measuring brain aging and its effects on brain anatomy. First, we will review tools and approaches to measure cognitive aging and neuroimaging studies that have revolutionized how brain aging is studied. Next, we will describe the nature and impact of common age-associated brain pathologies and diseases. We will also consider the burgeoning area of identifying factors that may modulate the effects of brain aging, and we will consider what may mediate those effects. Finally, we will summarize the approaches that may help to promote healthier brain function during aging.

CONCEPTS AND CHALLENGES IN STUDYING BRAIN AGING

How are metrics for brain aging defined and established? Table 55.2c–1 lists some essential questions and issues that are critical to this measurement. First, is it helpful to define approximate stages in brain aging based on transitions at specific ages, even though the underlying processes are likely to be continuous and cumulative? Although it is easy to identify age by time, we do not have biologic markers of the aging processes. By default, “milestones” used to define aging are thus measured by chronologic age and may impose artificial divisions and definitions. For example, many aging studies set a cutoff of 65 or older as older adults. Even disease definitions, such as dementia, have considered onset before the age of 65 as presenile and after 65 as senile. This arbitrary cutoff at 65 was selected based on the age for retirement benefits, established in 1889. In developed countries, the average life expectancy now exceeds 75, and many people live into their ninth and tenth decades. In addition, in the United States and Europe, we see a steady increase in centenarians. So, the concepts of young-old, aged 65 to 75, and oldest-old, aged 85 and older, have arisen. However, no agreed-upon measures or markers other than the passage of time distinguish groups of people in their eighth, ninth, and tenth decades. Studies have not yet shown clearly whether the age-associated decline of brain abilities accelerates at particular ages. Indirect support for acceleration comes from epidemiologic studies of age-associated syndromes such as dementia, which show an exponential increase in incidence after 65. Therefore, the inadvertent inclusion of some subjects with preclinical dementia influences many “normal aging” studies with a likely outcome that cognitive decline will appear to accelerate in very old age.

Another question about staging is how to develop benchmarks for “normal” brain structure and performance against which to measure effects related to aging and disease. One approach is to collect normative data for groups of apparently healthy individuals of varying ages, that is, to age adjust. This approach may have clinical utility, for example, to define laboratory test cutoffs or identify suboptimal performance as a likely indicator of disease. Clinical, pathologic, and imaging research studies across the life span suggest that the human brain matures, establishes connectivity, and achieves optimal function by the time people reach their mid-to-late 20s. There is some measurable decline beginning as early as the 30s. Should brain structure and performance at age 30 be used as a benchmark to calibrate the magnitude of changes associated with many years of aging? This cutoff may help estimate relative changes in the number of neurons in particular brain regions or changes in connections between neurons and may also shed light on how human cognitive performance changes. Many studies have followed people aged 60 or older to derive data on brain structure or performance and compare subjects in later decades to people in their 60s. However, this may underestimate the magnitude of changes across the adult age span. This underestimate may particularly affect pathology studies, where studies across the age span have indicated that a disorder such as AD expresses itself clinically only after decades of slowly emerging pathology.

Table 55.2c–1.
Important Questions Regarding Brain Aging

1. Should benchmarks or reference ages be used to calibrate the extent of age-related changes? Are there biomarkers of chronologic age?
2. Is brain aging a steadily progressive process or does it accelerate?
3. How can “normal” aging be distinguished from disease-related changes?
4. Are some brain pathologic changes inevitable?
5. What factors are associated with more or less severe age-associated brain changes?
6. Are some brain regions more susceptible to age-related changes than others?
7. Do plasticity and cognitive reserve attenuate the loss of function associated with brain aging?

Secular changes that may exaggerate differences between younger and older individuals may also inflate cross-sectional comparisons across the age span. For example, levels of education, nutritional habits, and the widespread use of medications such as statins and antihypertensives have changed markedly over the past 50 years. These changes may result in differences between cohorts born decades apart when assessed through cognitive testing during life and even in studies of postmortem brain tissue. Longitudinal data relating measures of clinical performance of the brain to the underlying structure are becoming available from imaging studies, but we usually measure serial imaging over a few years rather than decades. However, secular changes do not inevitably result in better brain health. For example, inactivity, obesity, and increasing rates of type 2 diabetes, as well as effects of air pollution, climate change among adults in Westernized societies, as well as global problems such as pandemics, may counteract improvements in education and cardiovascular management.

Almost any physiologic index shows an increase in its variability in association with aging. This increase holds for many aspects of cognitive and motor performance in older adults. The reasons for increased variability are not well understood, but both genetic and environmental factors may contribute. Identifying factors that contribute to decline in brain performance associated with aging, particularly those that may be modifiable or preventable, may lead to interventions aimed at forestalling or attenuating the degree of age-associated decline that falls within what we consider normal.

Are there boundaries between brain aging and disease? Can disorders like AD be demarcated from normal or optimal brain aging? Aging is associated with an increased prevalence of lesions such as senile plaques (SP) and neurofibrillary tangles, vascular brain changes, and neuronal loss. Even though the clinical syndrome of progressive dementia associated with AD is different from the lesser degrees of cognitive change associated with aging, the key pathologic features of AD are present to some extent in most individuals aged 90 or older. If Alzheimer pathology is an inevitable feature of aging, it is possible that extending life span will lead to the inevitable development of clinical AD. Even if AD pathology is an unavoidable part of the aging process, the pathology builds up slowly, over a decade or longer, and the task of understanding factors that may lead to earlier or later age of onset of clinical dementia and the quest to delay the expression of clinical dementia for as long as possible remains extremely important. Because of the high frequency of AD lesions in older adults, it is difficult to carry out studies of brain aging in which the definition of normal requires the exclusion of any degree of AD pathology. Nevertheless, it may be possible to dissect the impact of aging from that of AD. In this quest, biomarkers for AD pathology have been used in studies to illustrate differences in longitudinal cognitive performance and magnetic resonance imaging (MRI) indices of functional brain connectivity in people who test positive for amyloid and tau biomarkers compared to those who test negative.

Studying Brain Tissue: Neuroanatomy, Genomics, and Imaging

Neurons vary significantly in structure and connectivity throughout the brain. An important question is whether aging selectively affects certain neurons or specific brain regions or whether changes are relatively uniform and widespread. To provide a systematic answer to this question would require an enormous amount of work obtaining quantitative measures or neuronal numbers and axonal and dendritic structure throughout the brain from a series of tissue samples across a wide age span. Studies of neurons during aging have typically targeted brain regions whose connections or neurotransmitter circuitry relate to specific brain processes, such as hippocampal neurons with memory.

An issue that has troubled the field for years is the extent of neuron loss with aging. A series of studies by Harold Brody sparked initial interest in this question. Brody’s work, supported by other investigators, suggested that the superior-frontal and superior-temporal cortices lost up to 50% of neurons between the ages of 18 and 95, and 20% to 30% losses occurred in the precentral gyrus and visual cortex. Later, investigators found that brain tissue processing caused shrinkage, and the initial studies likely overestimated the density of neurons in younger brains. After appropriate adjustment and the introduction of automated counting methods, investigators found markedly lower estimates of neuron loss. However, the number of neurons is only one of many conceivable measures of neuronal well-being. Recent studies have evaluated the dendritic arbor or synaptic density as indices of neuronal connectivity and have assessed the strength of neurotransmitter circuits by measuring levels of neurotransmitters or their receptors in the brain. These approaches, assessing structure, connectivity, and biochemistry, provide a more nuanced picture of age-associated changes.

Technology advances enable comprehensive protein and gene expression studies in discrete brain regions and even in single neurons. Large-scale omic analysis is now feasible for interrogating the mechanisms of brain aging. In particular, transcriptional analysis using arrays to measure relative mRNA expression, followed by bioinformatic approaches, allows identification and analysis of signaling pathways. For example, the BRAINEAC project (www.braineac.org) analyzed mRNA expression in 12 brain regions in brains from over 130 people without neurodegenerative disorders covering the adult age span.

In recent years, neuroimaging studies have brought another dimension to the study of brain aging. It is possible to obtain morphologic, functional, and metabolic information about the brain in living people with excellent precision and reliability. The resolution of MRI is sufficient to obtain overall and regional quantitative measures related to the brain. Multimodal brain imaging studies, combining measures of brain structure, connectivity, blood flow, and chemistry in the same individual, are being brought into play to help to understand structure–function relationships.

Thresholds of Pathology, Clinical Disorders, and Cognitive Reserve

Investigators have attempted to define thresholds to clarify the relationship between the pathology and the clinical expression of brain disorders. For example, some have tried to define the burden of pathology or cell loss typically present when a clinical phenotype manifests. For example, studies of postmortem brain tissue of patients with Parkinson disease have suggested that motor symptoms do not appear until more than 40% of dopaminergic neurons in the substantia nigra are lost. This threshold exceeds the extent of neuron loss found in this brain region in studies of normal aging. Impaired reserve and plasticity may play a role in whether age-associated changes in the brain may manifest as disease or not. Subthreshold changes related to aging may increase vulnerability to pathology. For example, an individual with mild Alzheimer pathology may maintain intact cognitive function with aging due to sufficient brain reserve. However, the addition of one or more small strokes may be enough to impair cognition beyond a tipping point, resulting in clinically recognizable dementia. Similarly, age-related decreases in the strength of norepinephrine or dopamine, and possibly serotonergic, circuits may increase the vulnerability for developing depression, and particular stresses may precipitate it even though they would not have done so at a younger age. Finally, brain aging, with decrease in strength of cholinergic circuits, may increase someone's vulnerability to develop adverse cognitive side effects from medications and develop delirium in the setting of systemic illness.

Cognitive reserve is, in part, a resilience that allows preserved performance in the face of damage, and it may protect against the onset of a clinical disease phenotype. Dr. Yaakov Stern has proposed two types of reserve: brain reserve refers to differences in brain structure that may result in tolerance of pathology, and cognitive reserve describes differences in task performance between individuals. Brain reserve may depend on the intrinsic richness of connections between brain structures or possibly on the total number of mature neurons that remain at the end of brain development. Factors such as education, carrying out cognitively demanding activities, or physical exercise may assist in the maintenance of these connections during aging. Neuroplasticity is an essential component of reserve. Neuroplasticity classically refers to the brain's ability to adapt and strengthen connections or potentially to form new connections between neurons and circuits. Studies of recovery after a stroke or ablative brain surgery support this possibility. However, aging generally leads to less effective recovery after injury, and the brain is no exception. Classically, we considered neurons fully differentiated postmitotic cells, making them particularly vulnerable to cumulative damage sustained over the years. Although neurons do not divide, they can form new synapses and carry out remodeling. This limited view of the regenerative potential of neurons has changed; exciting research findings over the last two decades have challenged the doctrine that mammals do not actively replace neurons. Neural progenitor cells had been identified in birds and rodents and play roles in brain plasticity.

A prominent example is neurogenesis which accompanies the acquisition of new songs by birds. In the 1990s, evidence emerged that similar progenitor cells exist in primate brains, including humans. New neurons can be generated throughout adult life in brain regions such as the dentate gyrus of the hippocampus, as progenitor cells differentiate in its subgranular zone and in the subventricular zone of the lateral ventricle. Although we are still trying to define the roles of these progenitor cells in health and disease, their potential to proliferate, differentiate, and migrate expands the repertoire of compensatory changes within the brain. In rodent and primate models, brain injury, for example, due to stroke, leads to the stimulation of neurogenesis. Recent studies have suggested that depression may be associated with deficient neurogenesis.

Table 55.2c–2.
Mechanisms Implicated in Brain Aging

1. Oxidative stress—damage to DNA, proteins, lipids
2. Mitochondrial damage
3. Protein misfolding and aggregation
4. Decreased endocrine support (e.g., estrogen and testosterone)
5. Decreased neurotrophic factor support
6. Impaired axonal transport
7. Impaired neurogenesis

MECHANISMS IMPLICATED IN BRAIN AGING

The mechanisms and processes that lead to age-related changes throughout the body and ultimately limit life span are not well understood. Aging has been characterized by one of its leading scientists, Dr. Leonard Hayflick, as the effect of an increase in entropy on molecules such as proteins, lipids, and deoxyribonucleic acid (DNA), resulting in the accumulation of structural changes that eventually lead to impaired physiologic function at the level of molecules, cells, and ultimately the organism. Many mechanisms can contribute to the age-related damage that may impair brain function or reduce brain reserve, as summarized in Table 55.2c–2.

Oxidative stress is particularly relevant to the brain, whose long-lived neurons use oxygen more extensively than any other organ. Although there are many antioxidant defenses and repair mechanisms, oxidative damage gradually affects the structure of DNA, ribonucleic acid (RNA), lipid membranes, and proteins. These changes may become irreversible, resulting in impaired function of the affected molecules and organelles. Mitochondrial damage associated with aging leads to less production of ATP and more production and accumulation of reactive oxygen species (ROS) may impair energy metabolism. During aging, damaged mitochondria produce less ATP, and more ROS accumulate. The ROS cause oxidative stress that accelerates aging and possible neurodegeneration; an example is the damage in substantia nigra. Proteins that undergo oxidative changes such as nitrosylation or carbonylation may be prone to cross-linking and to aggregation. In addition, they impose a load on the degradative machinery of neurons, which may be unable to remove these abnormal proteins effectively or break them down completely. Diabetes, which is increasingly prevalent with aging, may promote oxidative stress through mechanisms such as the glycation of proteins and the generation of advanced glycation end-products. Caloric restriction has been widely studied in animals as a method that may extend the life span. Decreasing oxidative damage and improving insulin signaling are two potential mechanisms that may explain these effects. However, whether this is effective for life span extension in humans is unknown, and we have not discovered whether it benefits brain health.

Misfolding of proteins, with the adoption of stable insoluble amyloid configurations, is an essential theme in many age-associated diseases. For example, aggregates of amyloid-β protein (Aβ) and the microtubule-associated protein tau form SP and neurofibrillary tangles are the hallmark lesions of AD. In families with early-onset AD, mutations in genes related to Aβ (amyloid protein precursor [APP] and presenilin 1 and 2) are associated with a relative or absolute increase in the production of longer forms of Aβ or produce mutant forms of Aβ that are more prone to aggregate. Evidence to date does not support the overproduction of Aβ in sporadic AD or with aging. Only a small change in the expression levels of specific proteins may be needed to initiate pathology, or aging may predispose to amyloid deposition in the brain through other mechanisms. Impaired degradation or clearance of Aβ is, in part, responsible for age-associated plaque formation. Factors that modulate the misfolding, aggregation, and clearance of proteins (proteostasis) during aging continue to receive much research attention.

Another age-related factor may be the lack of trophic support of certain types of cells or their connections. Estrogen deficiency after menopause or androgen decline with aging may impair the function of neurons that respond to these hormones. Delivery of neurotrophic factors may be necessary for synaptic plasticity and specific populations of neurons. For example, in aged rodents and primates, the administration of nerve growth factor (NGF) to the brain can enhance cholinergic neurotransmitter circuits and improve performance on cognitive tasks related to memory. Axonal transport is required for neurons to convey proteins and other components needed to maintain the structure and function of their lengthy processes. Impaired axonal transport is one factor in the decrease of growth factors in the aging brain.

Inflammation is a mechanism under exquisite regulatory control throughout the life span. Peripheral inflammation is most likely one mechanism related to aging in general. For example, several cytokines and inflammatory biomarkers, such as C-reactive protein, are increased in serum with aging, and various autoantibodies increase during aging. An increase in microglial activation and other innate immunity mechanisms in the brain may damage or lack appropriate neuronal or synaptic support.

Although there is no obvious selection or survival advantage for genetic factors that influence brain aging, investigators have identified several genes that may impact brain aging. As an extreme case, rare genetic causes of premature aging syndromes, such as progeria, result in hallmarks of premature aging, such as cataract development, skin changes, increased risk of cancer, and immune alterations during childhood. However, these syndromes are not clearly associated with accelerated brain aging or decline. In addition, there are rare genetic mutations associated with early-onset familial AD, inherited in an autosomal dominant pattern. However, these genes do not appear to have sequence changes that alter the risk or age at the onset of late-life AD. The most common variant associated with the risk of late-onset AD is the gene for apolipoprotein E (APOE), which codes for a lipoprotein involved in cholesterol transport within the brain. There are three APOE alleles, e2, e3, and e4. People who carry one copy of the e4 allele have a two- to threefold increased risk of developing AD, whereas those with two copies have a four- to eightfold increased risk. The e2 allele slightly decreases the risk of developing AD and is overrepresented in people who live into their 90s or 100s.

The ability to generate new neurons in certain adult brain regions also decreases with aging in animal studies. Although direct data are lacking from studies of human brain tissue, a decrease in the production of neural progenitor cells may play a role in brain aging. Another study developed an innovative magnetic resonance spectroscopic index of neural progenitor cells and could detect a specific signal in cultured cells and animals. Furthermore, a small group of human volunteers spanning the age range from 20 to 70 showed attenuation of this signal with age in the hippocampus, an area known to harbor progenitor cells. At present, this represents the best available evidence for the age-related decline of neural progenitor cells in humans. That said, most of the research, including the effects of aging, derives from rodent studies.

STRUCTURAL CHANGES ASSOCIATED WITH BRAIN AGING

Brain atrophy accompanies aging, but the extent to which this represents a loss of neurons is controversial. The belief that thousands of neurons are lost every day derives from studies with methodologic flaws. More recent studies indicate that neuronal cell loss is less conspicuous and may be selective for specific brain regions.

Methodologic issues are essential in brain aging studies. By design, aging studies of human brains are cross-sectional. Secular effects, for example, generational changes in nutrition, general health, and comorbidity, can complicate comparisons in brains of individuals in their 30s and 40s with those in their 70s and 80s. Height and gender influence brain weight and size. Counts of neurons should attempt to adjust for these factors. Also, brain region sizes and neuron count vary as much as two- to threefold across populations. To detect significant changes in the face of high variability requires many samples or large effect sizes. Ideally, the brains used for normal aging studies should be from people formally evaluated close to their time of death. As mentioned above, including brains with AD pathology is inevitable if an aging study includes individuals in their 80s and 90s. However, there are staging systems of AD pathology, allowing for the exclusion of people with a substantial lesion burden.

Postmortem tissue preparation is critical—early studies that claimed that extensive neuron loss accompanied aging were flawed because of shrinkage of brain tissue during its preparation. Technical approaches to imaging and counting neurons have improved in recent years, including automated counting of neurons based on profiles and the availability of cell-specific stains. Stereologic approaches to sample neurons in a relatively unbiased manner have become routine. Comparative studies of brain tissue from aging primates and rodents have the advantage of minimizing postmortem changes and optimizing brain tissue preparation. These provide valuable data that complement direct studies on human tissue.

Total brain weight reaches a maximum in young adults around the age of 20, followed by a progressive decline thought to begin around the ages of 30 to 40 and reaches its nadir after 85 (Table 55.2c–3). By that age, on average, brain weight has decreased by 11% compared to the maximum. Investigators have found similar findings of an age-associated decline in studies of chimpanzee brains. Neuroimaging studies have provided both cross-sectional and longitudinal data showing that brain volume decreases with age, accompanied by an increase in the volume of the ventricles.

Quantitative MRI studies have shown that brain volume decreases across the adult age span. Some studies have suggested that brain volume loss may accelerate with aging, with about 0.1% to 0.2% of brain volume lost per year between 30 and 50, increasing to about 0.3% to 0.5% per year after 70. Preclinical diseases related to aging, including vascular changes, amyloid, and tau deposition, may contribute to this apparent acceleration. Although the whole brain, including gray and white matter, shows volume loss, regions such as the hippocampus and frontal and parietal cortices are affected to a greater degree than the occipital cortex.

What explains the decline in brain weight or volume? Neuron loss, as discussed, has been overestimated in the past. However, after Hege Gundersen and colleagues introduced stereologic methods (including the dissector and its modification, the optical dissector), studies by Bente Pakkenberg and Gundersen reported a less substantial loss of neurons across the adult life span, representing about 9.5%. Other studies have questioned whether there was substantial loss of gray matter, attributing age-associated declines to loss of white matter.

Table 55.2c–3.
Structural Changes Associated With Brain Aging

1. Decline of brain weight
2. Neuron loss
3. Neuronal atrophy
4. Synaptic loss
5. Pruning of dendritic trees
6. Nucleolar atrophy
7. White matter changes
8. Gliosis

Neuron atrophy, rather than loss, could contribute to the age-related decline in brain weight. For example, a classic study by Dr. Robert Terry and colleagues used a computer-assisted method to estimate neuron numbers and size in the superior-temporal, medial-frontal, and inferior-parietal cortical regions from ages 24 to 100. They found that the number of large neurons decreased in all of these regions, accompanied by a corresponding increase in the number of small neurons. In addition, investigators have documented the atrophy of pyramidal cells. In particular, pyramidal neurons in the hippocampus and cerebral neocortex show atrophy. This atrophy is associated with shrinkage in the nucleolus or loss of cell body RNA.

Regional and Microstructural Changes

Investigators are re-examining whether particular regions are vulnerable to neuron loss with aging. The hippocampus and its subregions, such as the entorhinal cortex, CA1, and CA2, have been singled out due to their role in memory and learning. The use of high-resolution MRI, for example, at 4 or 7 Tesla strength, allows for acquiring regional anatomical information in great detail. For example, in a study of healthy individuals spanning the age range from 20 to 85, the CA1 subregion of the hippocampus showed the most significant degree of volume loss with age, whereas the entorhinal cortex, CA2, CA3/4, and dentate gyrus showed much less change. How does this compare with pathology studies? Teresa Gomez-Isla and colleagues applied stereologic techniques to the entorhinal cortex and found no appreciable loss of neurons in cognitively healthy individuals in their 70s and 80s; by contrast, there was a substantial degree of neuronal loss in AD, even when clinically very mild. When they similarly studied the superior temporal cortex, an area representative of the association cortex, once again, there was no significant loss of neurons with aging. However, stereologic studies of the CA1 subfields have been inconsistent. Some investigators have found decreases in neurons in CA1 and the subiculum with age, but others failed to find these effects. The reasons for these discrepancies are unclear. Using similar stereologic methods, studies of monkeys have shown no appreciable loss of neurons with age in the entorhinal cortex, CA1, CA2, or subiculum. However, despite the intact number of neurons, decreases in *N*-methyl D-aspartate (NMDA) receptors were evident in aged monkey brains. This finding raises the possibility that changes short of neuronal death may occur during aging and may provide a substrate that could explain the cognitive decline.

Changes in Synapses and Circuits

Neurons function by exchanging information across circuits and networks. The shape and branching of dendrites vary markedly between different neuron types and within individual neurons of a particular class. Dendrites have protrusions called spines, which contain excitatory synapses. These vary in density across the dendrite length and increase in number with successive branches of dendrites. Even in the absence of overt neuronal loss, changes in dendritic branches or spines could impair neuronal circuit function. A consistent finding in anatomical studies is age-related, prominent in pyramidal neurons in areas such as the prefrontal, precentral, and superior-temporal cortices. Studies using fluorescent dyes have revealed a marked loss of dendritic spines at all levels of dendrites in aged monkeys. Dendritic changes may link to the levels of neurotransmitters and strength of certain types of neurochemical circuits. For example, the glutamate R2 and NMDAR1 receptors show decreased expression in monkeys associated with aging, in areas such as the prefrontal cortex for glutamate R2 and projections from the superior-temporal cortex for NMDAR1. Detailed studies of postsynaptic non-NMDA glutamate receptors in monkeys have suggested that these remain functionally unchanged in primates in areas such as the prefrontal cortex. However, there appears to be an increase in the level of inhibitory synaptic input to pyramidal cells in this area. The increased release of γ-aminobutyric acid (GABA) from presynaptic neurons could explain this. In the absence of overt neuron loss, these changes in morphology and biochemistry could form a substrate for declining performance on cognitive tasks that assess memory, learning, and cognitive processing speed. Of interest, circuits that are highly vulnerable to degeneration in AD may also display synaptic changes similar to those described above in association with aging.

Neuroimaging and electrophysiologic methods (e.g., EEG and magnetoencephalography [MEG]) can help develop functional brain maps and identify overall and regional brain connectivity. Functional MRI (fMRI) evaluates cerebral responsiveness to activation in human subjects. This approach takes advantage of the BOLD (Blood-oxygen-level-dependent) signal, which is related to a decrease in deoxyhemoglobin concentration and an increase in cerebral blood flow that accompanies local neuronal activation. Such studies can explore the activation of brain circuitry in response to cognitive tasks. Many fMRI studies evaluate neural activation of regions and circuits in the brain with aging. For example, the extent of hippocampal activation in response to episodic memory tasks decreases in cognitively intact subjects aged 70 or older compared to younger adults. Resting-state fMRI takes advantage of correlated regional blood flow changes at short time scales to investigate and map the degree of large-scale coordination between brain systems. A default network of functionally interacting anterior and posterior brain regions is associated with internally directed mental states and is vital in forming a substrate for memory acquisition and planning. Early studies exploring aging, for example, those by Buckner and colleagues reported a sizeable age-related decrease in functional connectivity between these systems that correlated with worse cognitive performance, particularly prominent after 70. The most likely structural basis for this functional decline was the disruption of white matter integrity. The ability to image preclinical amyloid pathology by amyloid PET scanning has led to a reappraisal that connectivity changes in cognitively intact older adult individuals are more significant in those with positive amyloid scans than without.

In summary, the level of understanding of structural changes in the brain that accompany aging remains descriptive. It will be a great challenge to define the mechanisms that underlie these changes. Brain imaging provides tools to study the effects of risk factors that may explain some of the variability of brain changes in later life. They also may help to interpret the effects of interventions that may decrease the degree of age-associated decline or enhance the function of brain networks or connectivity, as discussed below.

LESIONS IN THE AGING BRAIN

Various pathologic markers are visible on microscopic examination of brain tissue from older adult individuals (Table 55.2c–4). Markers such as corpora amylacea, lipofuscin, or neuromelanin, may represent cellular responses to aging with uncertain clinical consequences because they do not correlate with clinical disorders or with a decline in function. Others such as amyloid deposits and neurofibrillary tangles are markers of AD when they exceed a threshold. In older people who lack defined disorders such as dementia, the overall burden of these lesions may likely contribute to cognitive decline attributed to aging. For example, circumscribed distribution of tau aggregates is a common part of aging in the absence of significant amyloid deposits, and some experts have proposed the acronym primary age-related tauopathy (PART). PART may be asymptomatic or associated with mild cognitive changes.

Table 55.2c–4.
Lesions That Accumulate in the Brain With Age

1. Lipofuscin
2. Corpora amylacea
3. Neuromelanin
4. Amyloid deposits, including senile plaques
5. Vascular amyloid
6. Neurofibrillary tangles and argyrophilic grains
7. Lewy bodies
8. Aggregates of TDP43 and hippocampal sclerosis
9. Vascular (ischemic) lesions

Lipofuscin is the most common age-related marker seen in tissues throughout the body. Postmitotic cells such as neurons accumulate lipofuscin more extensively than mitotic cells. Lipofuscin is particularly evident in neurons in the inferior olive, dorsal root ganglia, anterior horn cells, the dentate nucleus of the cerebellum, and large neurons in the cerebral cortex. Lipofuscin consists of protein and lipid, mostly oxidized and contained within lysosomal membranes. Metals, particularly iron, are also present. Lipofuscin is evident as intracellular autofluorescence or can be seen using lipid stains or via electron microscopy. Lipofuscin likely represents the remnants of incomplete mitochondrial breakdown by lysosomes. The accumulation of small amounts of lipofuscin may allow cells to sequester aged mitochondria of cross-linked proteins. However, when lipofuscin occupies large cell areas, it is less clear that this is adaptive and likely adversely affects cellular processes.

Corpora amylacea are round, cytoplasmic glycoprotein inclusions most commonly seen in astrocytes, which are found in the brain after age 40. They may arise due to the effects of oxidative stress and iron imbalance within astrocytes. However, it is uncertain whether they represent a cellular defense that sequesters iron or oxidized components of mitochondria or whether their presence may impair the ability of astrocytes to function optimally. The number of corpora amylacea is increased markedly in the brain in AD.

Neuromelanin may be found within neurons as early as 3 years of age and increases throughout life. It is particularly abundant in the substantia nigra and the locus ceruleus. The function of neuromelanin is not known. However, it contains melanins, derivatives of dopamine, and lipofuscin. Neuromelanin likely forms after the uptake of extracellular dopamine and its oxidation and may represent a means to sequester potentially damaging dopamine byproducts. In addition, Neuromelanin avidly binds iron, which may be protective, but could act as a source of oxidative stress in association with aging.

Protein misfolding is a prominent feature in many body organs, accompanying aging and associated with the disease. In the brain, aggregates of Aβ form extracellular SP and deposit in the walls of meningeal and intraparenchymal arterioles. Tau forms aggregates with the ultrastructure of paired helical filaments. Its aggregates are in dystrophic neurites, argyrophilic grains, and neurofibrillary tangles.

Both plaques and tangles are associated with aging, and AD patients have them in greater abundance. Because there is no unique morphology of plaques or tangles that characterize AD but is absent from cognitively intact older adult individuals, pathologic criteria used to diagnose AD are quantitative. Both plaques and tangles are necessary for a diagnosis of AD, and the distribution and burden of tangles, as well as neuronal loss, more strongly correlate with cognitive abilities and dementia than does amyloid. There are several sets of pathologic criteria for AD. These have been updated in light of increasing knowledge of the anatomical distribution of tau lesions. National Institute on Aging and the Reagan Institute (NIA-Reagan) criteria combine the presence of abundant neuritic plaques with the anatomical extent of tau pathology (limited to entorhinal cortex, spread to the hippocampus, or spread to neocortex) to assign a low, intermediate, or high probability that AD is present. These were recently modified to also score the extent and nature of Aβ pathology as a “phase” of brain amyloidosis.

Studies of individuals with intact cognitive performance close to the time of death have reported an age-associated increase in the number of plaques and tangles. About one-third of cognitively normal subjects in their 80s meet NIA-Reagan criteria for an intermediate probability of AD. In people without dementia, tangles are usually limited to the hippocampal formation. The spread of these lesions to involve association areas in the temporal neocortex, and more widely, is more clearly linked to the development of clinical evidence of cognitive decline and dementia. In very older adult individuals with AD, there is also a lower tangle burden compared to younger individuals with AD. This finding suggests that brain reserve is diminished in association with aging so that only a small increase in tangles is necessary to lead to the clinical expression of dementia. The question of whether AD pathology contributes to age-associated changes in general cognitive performance is not clear. Autopsy studies have suggested that people in their 50s may have some degree of amyloid deposition and tangle formation. However, tau pathology does not occur through the brain but instead is usually restricted to hippocampal subregions and the entorhinal cortex and may represent PART. Age-related declines in cognitive speed and executive function correlate weakly with plaques and tangles.

Two other lesions occur variably in association with brain aging. With the loss of neurons in the CA1 and sometimes other hippocampal regions accompanied by gliosis, we generally do not consider hippocampal sclerosis a part of “normal” aging. However, its prevalence increases with age, being in about 15% to 20% of brains of people over age 90. When present, hippocampal sclerosis contributes to cognitive loss and dementia and is usually associated with aggregates of a protein called TDP43. Another protein, α-synuclein, forms aggregates within neurons, including structures, called Lewy bodies, which are perinuclear intraneuronal inclusions. These are the identifying pathology of Parkinson disease. They occur in subcortical areas such as the locus ceruleus and substantia nigra, and cortical and limbic regions in small numbers and exist in up to 20% of elders examined at autopsy. Like hippocampal sclerosis, α-synuclein aggregates are more likely to be present in the brains of people with cognitive impairment and are more likely markers of preclinical disease processes than adaptive changes of brain aging.

Age-associated vascular changes in the brain have received increasing focus. The many faces of vascular pathology include cerebral infarcts, lacunar infarcts, microangiopathy, and amyloid angiopathy. Some of these lesions, in particular infarcts, are not inevitable accompaniments of aging but are related to vascular risk factors such as hypertension and diabetes. Cerebral amyloid angiopathy (CAA), with deposition of Aβ in the walls of leptomeningeal and medium-sized and small arterioles, is present in at least 30% of older adult individuals and can occur in the absence of the diagnosis of AD. Vessels laden with amyloid may rupture, resulting in lobar hemorrhage, which may cause neurologic deficits. MRI studies indicate that microhemorrhages, visible on gradient echo MRI sequences, may occur in association with CAA, and if extensive, are associated with impaired cognition. The widespread use of computed tomography (CT) imaging and MRI in studies of aging and cognition has led to an appreciation of white matter changes. On CT, these appear as a periventricular halo of hypodensity. On MRI, they are easiest to detect on T_2-weighted images. They have various appearances, including periventricular hyperintensity, capping of the ventricular horns, and hyperintense spots of varying size. The hyperintense spots may coalesce to form larger cotton–wool-like confluent areas in the fiber tract projections of the centrum semiovale. Initially referred to as leukoaraiosis, many attributed these changes to incomplete ischemias affecting the white matter tracts. Capping of the ventricles is not related to ischemia, but age-related disruption of the ependymal layer increased accumulation of interstitial fluid and subependymal gliosis. Neuropathologic findings vary and often include pallor and loss of myelin. Cognitive testing in older adult individuals with white matter changes reveals decreased performance on timed tests and tasks dependent on executive function. White matter changes are more extensive on MRI in people with vascular risk factors such as diabetes and hypertension. These vascular factors may worsen the severity of age-dependent changes in the structure of long projecting tracts that extend through the white matter.

Detailed microscopic studies of blood vessel morphology have raised further questions about vascular factors in age-associated cognitive decline. For example, a decrease in the diameter of capillaries in hippocampal subregions, independent of AD-related lesions, raises the possibility that microvascular changes could contribute to neuron loss or damage. These changes might alter the available cerebral blood flow to neurons in the hippocampus, resulting in oxidative stress, damage, and eventually neuronal death. Intervention to treat hypertension and diabetes in middle and later life could help forestall some of the cognitive decline attributed to aging, although clinical trials of antihypertensive treatment have had mixed results on cognitive outcomes to date.

Another vascular measure is cerebral blood flow, which can be measured using positron emission tomography (PET). More recently, investigators have used the magnetic resonance technique of arterial spin labeling (ASL) to measure cerebral blood flow. This method magnetically tags water protons in the carotid and vertebral arteries at the base of the brain, the major conduits of blood supply to the brain. As these tagged protons enter the brain, they alter the local magnetization state in proportion to blood flow. There have been relatively few large-scale studies of cerebral blood flow with aging. In general, total cerebral blood flow decreases with age, with studies reporting about 0.5% to 0.6%/year changes. Recent studies that have examined regional cerebral blood flow have identified heterogeneous changes: decreases with age seem to occur mainly in limbic and tertiary association cortices, with little change in regions such as the primary visual cortex. Vascular risk factors may adversely affect global and regional cerebral blood flow.

NEUROCHEMICAL CHANGES IN AGING

Many neurotransmitters and neuropeptides in the brain change with age. Two of these, dopamine and acetylcholine, will be highlighted because of their importance in cognitive and affective function and their potential importance in neuropsychiatric disorders. Much of the decline in neurotransmitter concentrations is due to presynaptic changes, often secondary to loss or atrophy of subcortical projecting neurons, for example, the pigmented cells of the locus ceruleus and substantia nigra (which produce norepinephrine and dopamine, respectively) or the nucleus basalis (which produces acetylcholine). Aging may also affect postsynaptic processes that affect transmitter function, including a decreased density of postsynaptic receptors and possible decreased efficiency of signal transduction.

Populations of dopaminergic neurons show marked changes during aging. In monkeys, neurons in the substantia nigra staining positive for tyrosine hydroxylase decreases by as much as 40% to 50%. This decrease may be associated with increased α-synuclein content in individual neurons, a potential factor related to decreased strength of dopaminergic transmission in the nigrostriatal system. These changes may account for slowing movement and possibly slowing cognitive processes in association with aging. Anatomical studies in primates show that dopamine innervation from the midbrain projects to extensive areas throughout the cortex. There are two broad projections, ventral tier neurons that project predominantly to the striatum and the dorsal tier, whose cells express calbindin and project to limbic, cortical, and selected striatal areas. More detailed anatomical studies may be required to evaluate the specific contribution of these different types of cells with aging. A decline in the strength of dopamine transmission in the mesolimbic and mesocortical pathways may contribute to increased vulnerability to disorders such as depression in aging.

PET imaging can obtain quantitative data for indices of dopaminergic neurotransmission. Studies find age-associated declines in dopamine transporter levels, D_2 receptors, $5HT_2$ receptors, and $5HT_{1a}$ receptors. However, the patterns of decline differ. For example, dopamine transporters and D_2 receptors show a linear decline with age, but serotonin $5HT_2$ receptors decline rapidly during the fifth decade of life and then more slowly.

Cholinergic neurons in the nucleus basalis of Meynert and the septal nuclei project widely to the hippocampus and neocortex. They function to activate brain regions involving sustained attention, long-term potentiation, and memory formation. In animal studies, lesions to cholinergic nuclei result in deficits in memory abilities. Aged rodents, monkeys, and humans show a decrease in the levels of markers of the cholinergic system with age. In AD, there is a more dramatic decrease which parallels the severity of dementia. Although the mechanisms of these age-related declines are unknown, diminished neurotrophic support is one possible factor. Cholinergic neurons respond to NGF, and decreased retrograde transport of NGF from targets to the cell body may atrophy the cholinergic cell bodies. Treatment of aged animals with grafted cells genetically engineered to secrete NGF resulted in improved cognitive test performance and increased cholinergic markers. However, human studies of gene therapy with NGF in AD have been disappointing. It may be that waiting until observable symptoms of dementia emerge is too late a time for gene therapy, and focusing on detecting pathologic processes at an earlier stage will improve outcomes.

CHANGES IN COGNITIVE AND MOTOR ABILITIES ASSOCIATED WITH AGING

Much research has documented changes in cognitive abilities, particularly memory, accompanying aging (Table 55.2c–5). On tests of episodic memory, both the acquisition of new verbal or visual information and recall after a delay show decline with age. These are evident even when comparing the performance of 40-year-olds to that of 20-year-olds. On average, subjects at age 80 perform approximately 1.0 to 1.5 standard deviations below the mean for young adults. Whereas performance appears to decrease throughout the adult age span, the decline may be a little steeper after age 60. Other aspects of cognition, including working memory, perceptual speed, spatial visualization, and inductive reasoning, show a similar age-related decline. Age-related effects on different types of cognitive tests are correlated, suggesting that widespread brain changes underlie them. One such factor is cognitive slowing; if we remove time limits from specific cognitive tests and the scoring criterion is accuracy, older adult individuals improve their performance. Similarly, multitasking and cognitive control are more difficult for older persons. As demonstrated by performance on timed tests, retrieval of semantic knowledge, including words and names, is impaired in aging. However, vocabulary, semantic memory, and cumulative knowledge tend to increase with age and remain preserved in old age. Decline in social cognition has been noted in the ability of the older adult to correctly identify another person's state of mind. This appears to be related to cortical and white matter volume loss in the prefrontal, temporal, and insular regions. The ability to perceive emotion remains more intact.

Table 55.2c–5.
Brain Performance Measures That Worsen With Aging

1. Cognition
    - Mental speed
    - Fluid reasoning
    - Working memory
    - Retrieval
    - Episodic memory
    - Free recall worse than recognition
2. Movement
    - Gait slowing
    - Simple reaction time slowing
    - Upward gaze
    - Balance changes (vestibular, sensory, motor, and brain)

Many aspects of movement show evidence of slowing with aging. One of the more noticeable features is the slowing of gait and tendency to stoop, typical in the oldest. Older adult individuals also show a decline in balance: difficulty standing on one leg or performing a tandem walk usually begins after age 70. The decline is most likely multifactorial, with contributions from declines in nerve and muscle function, changes within the spinal cord, sensory systems (vision and vestibular function), and brain changes involving structures such as the basal ganglia. In addition, simple reaction time and movement time are both slowed with aging, resulting from changes to white matter tracts and neuronal networks that initiate and coordinate rapid movements.

AGE-ASSOCIATED BRAIN DISEASES

Several disorders show a predilection for the aged brain (Table 55.2c–6). Neurodegenerative diseases rank high on the list; AD is the brain disorder most intimately related to aging. Both the incidence and the prevalence of AD show an exponential increase after the age of 60. Parkinson disease has a prevalence of approximately 1% to 2% in the general population. The prevalence of Parkinson disease increases with aging and may be as high as 4% to 5% in individuals aged 85 or older. However, the frequency of Lewy bodies—the signature pathology of Parkinson disease—in the substantia nigra in series of autopsies from people aged 60 to 90 is higher, typically around 5% to 8%. This finding is consistent with the concept that reserve may protect against the clinical expression of Parkinson disease. Amyotrophic lateral sclerosis (ALS) is much less common than AD or Parkinson disease. Its onset is more common among people aged 60 or older than in younger persons without a causative gene mutation.

Table 55.2c–6.
Age-Associated Brain Diseases

1. Alzheimer disease
2. Vascular dementia
3. Normal pressure hydrocephalus
4. Parkinson disease
5. Amyotrophic lateral sclerosis
6. Depression
7. Seizures
8. Sleep disorders
9. Essential tremor

The incidence of stroke increases with age. This increase includes clinically apparent strokes and evidence of stroke on neuroimaging in the absence of a history of stroke (silent infarcts). In population-based studies, silent infarcts are much more common than symptomatic strokes in older adults, and the overall prevalence of infarcts on MRI may be as high as 20% in older adults. In addition, the prevalence of MRI-defined infarcts is associated with aging and with cardiovascular risk factors such as hypertension.

BRAIN AGING IN THE OLDEST-OLD AND CENTENARIANS

In 2016, there were 82,000 Americans aged 100 years or older. This number will likely increase dramatically during this century. Living to be 100 or older is a milestone worthy of celebration if it includes good physical and cognitive function. Several studies have tried to determine the prevalence of dementia in centenarians. Centenarians typically have difficulty performing daily activities, often due to physical or sensory impairment. Because of marked differences in methods of cognitive assessment and incomplete sample ascertainment, estimates vary widely. However, longitudinal data from the 90+ study of nonagenarians living in Leisure World in Southern California identified a marked acceleration in the risk of dementia after the age of 90 with a prevalence of at least 50% by age 100.

In the oldest-old, correlations between pathologic lesions of AD and cognitive status are weaker than those in younger individuals. For example, the Oregon Brain Aging Study followed people aged 85 or older without cardiovascular or major age-associated diseases. Subjects who died without having shown cognitive decline or dementia had scanty SP and neurofibrillary tangles or NFTs, although there was a correlation between the increasing burden of these lesions and impaired performance on cognitive testing. The Leisure World 90+ study includes cognitive testing, clinical evaluation for the presence or absence of dementia, and a high rate of brain examination at autopsy. Whereas counts of SP weakly distinguished people with dementia from those without, the most significant discriminating features for dementia were reduction of presynaptic biochemical markers, overall tangle burden, and the presence of hippocampal sclerosis. Synaptic density may represent a dynamic and adaptive marker initially related to established connectivity during development, later remodeled according to use and plasticity, and then undermined by age- and disease-related processes. A study of more than 300 Dutch centenarians from the

100+ study between 2013 and 2019 showed no decline across investigated cognitive domains for up to 4 years, with the exception of a slight decline in memory function. Cognitive performance was associated with better physical health and cognitive reserve (e.g., higher education). While centenarian brains revealed varying loads of postmortem neuropathologic hallmarks of AD, this was not associated with cognitive performance or rate of decline.

Table 55.2c–7.
Interventions Proposed to Help Maintain Brain Abilities With Aging

1. Control hypertension
2. Treat diabetes and vascular risk factors
3. Regular physical activity
4. Mental activity, e.g., cognitively demanding pastimes
5. Social activity, e.g., friend, family, and community interactions
6. Diet: similar components to a heart-healthy diet, e.g., a diet rich in antioxidants and low in saturated fat, or a Mediterranean-type diet

PROSPECTS FOR HEALTHY BRAIN AGING

In the absence of effective treatment that slows the progression of AD, it is essential to consider how to prevent AD and preserve cognitive abilities during aging. Population-based studies of aging and dementia have suggested that there may be modifiable factors associated with the risk of dementia and or age-associated cognitive decline. These include aspects related to lifestyle and diet as well as medical interventions (Table 55.2c–7). These types of interventions are being evaluated in randomized clinical trials beginning in mid or later life. Early results from one of these studies conducted in Finland are encouraging, suggesting that cognitive performance may improve due to multicomponent intervention in people who have an unhealthy lifestyle.

Converging evidence suggests that physical activity may modify brain aging. In animal studies, maintained physical activity or exercise is linked to decreased markers of oxidative damage in the brain with aging, increased neurogenesis, increased brain perfusion, and lower expression of amyloid pathology in transgenic mice used as models of plaque deposition and AD. Epidemiologic studies have shown that regular physical activity is associated with a decreased risk of developing AD. Vigorous aerobic exercise may not be necessary, as evidence suggests that walking may be associated with risk reduction. Self-reported higher levels of exercise correlate with better cognitive test scores in older adult individuals. There have been many controlled trials of exercise intervention in older adult subjects; a meta-analysis of 18 randomized trials found that aerobic exercise training had beneficial effects for healthy older adults who did not exercise regularly. Effects were more robust for tests of executive function and reaction time than for other cognitive processes and were statistically significant only for interventions that used exercise sessions that were at least 30 minutes long. We need further research to confirm the findings of intervention studies, establish what types of exercise are beneficial, and determine how long and frequent each exercise session should be. In addition to defining a minimal amount of exercise that may show a positive cognitive effect, studies need to identify feasible exercise regimens for most older adult individuals. Along the way, it will be essential to clarify how exercise benefits the brain. Some have suggested that exercise’s benefit is related to its influence on oxidative stress by improving the balance between prooxidants and antioxidants. However, much needs to be done in this area to clarify this.

Sleep may be an essential factor as well. Patients with dementia often have altered circadian rhythms, and several studies show an association between the circadian pattern and cognitive decline. Disordered sleep appears to be an independent risk factor for AD. The direction of the cause and effect is not clear—the relationship seems to be bidirectional. However, some investigators have suggested mechanisms for how circadian dysfunction can cause disease progression. For example, many molecules relevant to memory consolidation, plasticity, and AD are under circadian control. Many clinical studies show that improving sleep can improve cognition. However, these are primarily in younger individuals without dementia, and whether interventions to improve sleep can treat or prevent dementia is an ongoing question worthy of study.

Cognitive stimulation is associated with reduced deposition of amyloid in transgenic mouse models related to AD. It uses such methods as environmental enrichment. Several human studies have found that engagement in cognitively demanding pursuits throughout adult life correlates with better cognitive performance and decreased risk of incident AD. For example, the Rush Memory and Aging Project enrolled over 1,200 people in Chicago, with an average age of 80, who were determined to be free of dementia and followed and reevaluated this cohort annually. They questioned subjects about cognitively active pursuits (such as reading books, attending music or plays, playing card games) and their frequencies. Those subjects who had a moderate or high level of cognitive activity had a lower rate of age-associated cognitive decline, which remained significant even after controlling for age, race, education, depressive symptoms, and chronic medical conditions.

Further analyses have shown that subjects with higher levels of cognitively demanding pursuits had a 2.6-fold lower risk of developing dementia than subjects with a low level. Other studies have reported similar findings, which have given rise to the “use it or lose it” hypothesis. Some fMRI studies support this, showing modifications in specific brain networks, such as attention, default mode, and visual networks. These associations do not necessarily a cause-effect relationship.

Many factors relate to higher levels of cognitive activity in later life, including general health, physical function and mobility, and differences in diet and lifestyle, that may be part of a package that contributes to brain function in aging. There are many intervention studies in this area, recent ones including computerized cognitive training. The Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE) study was a large-scale intervention that enrolled over 2,800 subjects aged 65 or older. Participants were randomly assigned to receive specific training modules for memory, reasoning, speed of processing, or no intervention. A subset of subjects received booster training at 11 or 35 months after enrollment. Cognitive training improved cognitive performance on tasks specific to the training. This improvement did not, however, generalize into other areas. The effect persisted for the 5 years of the study. Reasoning training resulted in slightly less decline in self-reported instrumental activities of daily living; otherwise, there was no functional benefit associated with training. This rigorous study highlights a key feature of cognitive intervention, training to task. Computer-assisted training appears to influence the specific training area, and it has been challenging to show broader improvement in other cognitive abilities or an impact of training on functional abilities. Unless such evidence emerges, there is no apparent reason to recommend computer games and interventions to improve brain health and function in aging.

Nutrition and brain aging is another area of controversy. It is challenging to randomize people to different types of diet for long-term studies, so research efforts in this area have used dietary questionnaires, short-term interventions, and supplements with vitamins or other nutrients. Many components of diet link to cognitive performance with aging. Some examples of factors that may be protective include eating a Mediterranean or heart-healthy diet or consuming foods with a relatively high content of naturally occurring vitamins with antioxidant activities (such as vitamins E and C). A diet developed at the Rush University School of Medicine which combines the DASH (Dietary Approaches to Stop Hypertension) diet and the Mediterranean diet, called the MIND diet (called the Mediterranean-DASH Intervention for Neurodegenerative Delay) appears to be promising in reducing the risk for developing AD. The Rush team found that this diet lowered Alzheimer’s risk by about 35% for those following it moderately well and 53% for those who were rigorously adhering to it. Subsequent studies have not confirmed the neuroprotective benefits of this diet.

Although switching to a healthy diet may have dramatic benefits, studies relying on dietary supplements are less promising.

Supplemental antioxidant vitamins have shown some promise in animal models but have not been promising in humans for improving cognitive benefits or reducing the risk of dementia. Homocysteine levels sometimes correlate with the risk of AD and stroke in several studies. In the United States, fortifying grain products with B-group vitamins has lowered the population’s homocysteine levels. However, intervention studies of lowering homocysteine in people with AD had no benefits on cognition. Omega-3 fatty acids are essential nutrients for brain function and growth during development. Clinical trials that have used omega-3 supplementation have had no benefits in AD and have not shown apparent effects in altering cognitive decline associated with aging.

Cardiovascular risk factors, in particular hypertension and diabetes, are associated with worse cognitive function with aging. For example, in the ACTIVE study discussed above, participants with hypertension had a faster decline in logical reasoning tasks, and diabetes was associated with a faster decline on a test of cognitive processing speed. Population-based studies have found that midlife hypertension is a consistent risk factor for accelerated age-associated cognitive decline, stroke, and incident AD. However, although clinical trials of treatment of hypertension in older adult individuals have shown clear benefits for reducing heart attack and stroke, they have had less consistent benefits on slowing cognitive decline. Problems such as dropout during long follow-up periods and the difficulty of maintaining an untreated comparison group limit these studies. As obesity has increased in adult populations in Western societies, attention focuses on whether metabolic syndrome is associated with cognitive performance in older adults. Studies have found that components of the metabolic syndrome, such as diabetes and abdominal girth, appear to be related to poorer performance on a variety of cognitive measures in cross-sectional studies in older adults. Much of this risk appears to be due to hyperglycemia. However, inflammation may also contribute to the risk. Animal trials suggest that such diabetes treatments as metformin can improve memory impairment. Clinical trials are currently examining whether interventions to maintain stringent control of type 2 diabetes benefit cognitive performance, with preliminary data suggesting that higher insulin doses in diabetic patients can have a protective effect.

Even in the absence of evidence from rigorous intervention studies, it seems reasonable to recommend maintaining physical and cognitive activity, controlling cardiovascular risk factors, and eating a heart-healthy diet to preserve general health with aging, with the bonus that there may be positive spin-offs regarding cognitive abilities. The Institute of Medicine included these lifestyle modifications in their recommendations. They also mentioned avoiding delirium and being cautious about using prescription medications with potential cognitive side effects. It also highlighted the essential planning roles that families, communities, and governments need to develop the environment to support healthy cognitive aging. More recently the Lancet Commission consisting of an international group of experts in dementia proposed prevention interventions incorporating potentially reversible risk factors from different phases of the life-span and not just old age, we are able to propose a novel life-course model of risk, from which population attributable fractions (PAFs) have been derived to show the possible effect on future incidence of successful elimination of the most potent factors. Using this approach, they calculated that more than a third of dementia cases might theoretically be preventable.

Our ability to describe and quantify changes in the brain associated with aging has burgeoned in the past few decades, with parallel advances in studying aging in animal models. Advances in brain imaging continue to have a massive impact on human studies of aging and cognition. Although there is significant progress at a descriptive level, researchers are still handicapped by not understanding which mechanisms are most important in driving the aging process and how they interact to result in biochemical and structural changes in the brain. Because aging is associated with a decline in brain abilities as well as an increased vulnerability to disorders such as dementia and cerebrovascular-associated cognitive impairment, continued research into mechanisms that lead to damage as well as those that support resilience and plasticity within the brain could pay an enormous dividend by improving health and vitality as society continues to benefit from gains in longevity.

ACKNOWLEDGMENTS

This chapter represents an update and revision of previous chapters in this series, most recently by Douglas R. Galasko, M.D., from the last edition. The authors are indebted to Dr. Galasko for his work and retains much of his prose in this current edition.

Further Readings

*Andrews-Hanna JR, Snyder AZ, Vincent JL, et al. Disruption of large-scale brain systems in advanced aging. *Neuron*. 2007;56:924–935.

Beker N, Ganz A, Hulsman M, et al. Association of cognitive function trajectories in centenarians with postmortem neuropathology, physical health, and other risk factors for cognitive decline. *JAMA Netw Open*. 2021;4(1):e2031654

Biagi L, Abbruzzese A, Bianchi MC, Alsop DC, Del Guerra A, Tosetti M. Age dependence of cerebral perfusion assessed by magnetic resonance continuous arterial spin labeling. *J Magn Reson Imaging*. 2007;25:696–702.

Brier MR, Thomas JB, Snyder AZ, et al. Unrecognized preclinical Alzheimer disease confounds rs-fcMRI studies of normal aging. *Neurology*. 2014;83:1613–1619.

Calvert JF Jr, Hollander-Rodriguez J, Kaye J, Leahy M. Dementia-free survival among centenarians: an evidence-based review. *J Gerontol A Biol Sci Med Sci*. 2006;61:951–956.

Colcombe S, Kramer AF. Fitness effects on the cognitive function of older adults: a meta-analytic study. *Psychol Sci*. 2003;14:125–130.

Corrada MM, Brookmeyer R, Paganini-Hill A, Berlau D, Kawas CH. Dementia incidence continues to increase with age in the oldest old: the 90+ study. *Ann Neurol*. 2010;67:114–121.

Crary JF, Trojanowski JQ, Schneider JA, et al. Primary age-related tauopathy (PART): a common pathology associated with human aging. *Acta Neuropathol*. 2014;128:755–766.

*Dickstein DL, Kabaso D, Rocher AB, Luebke JI, Wearne SL, Hof PR. Changes in the structural complexity of the aged brain. *Aging Cell*. 2007;6:275–284.

Drachman DA. Aging of the brain, entropy, and Alzheimer disease. *Neurology*. 2006;67:1340–1352.

Gorelick PB, Scuteri A, Black SE, et al; American Heart Association Stroke Council, Council on Epidemiology and Prevention, Council on Cardiovascular Nursing, Council on Cardiovascular Radiology and Intervention, and Council on

Cardiovascular Surgery and Anesthesia. Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42:2672–2713.

Green MS, Kaye JA, Ball MJ. The Oregon brain aging study: neuropathology accompanying healthy aging in the oldest old. *Neurology*. 2000;54:105–113.

Harmell AL, Jeste D, Depp C. Strategies for successful aging: a research update. *Curr Psychiatry Rep*. 2014;16:476.

Hayflick L. Biological aging is no longer an unsolved problem. *Ann N Y Acad Sci*. 2007;1100:1–13.

Iadecola C. The pathobiology of vascular dementia. *Neuron*. 2013;80:844–866.

*Imhof A, Kövari E, von Gunten A, et al. Morphological substrates of cognitive decline in nonagenarians and centenarians: a new paradigm? *J Neurol Sci*. 2007;257:72–79.

Jack CR Jr, Wiste HJ, Weigand SD, et al. Age, sex, and APOE ε4 effects on memory, brain structure, and β-amyloid across the adult life span. *JAMA Neurol*. 2015;72:511–519.

Karr JE, Areshenkoff CN, Rast P, Garcia-Barrera MA. An empirical comparison of the therapeutic benefits of physical exercise and cognitive training on the executive functions of older adults: a meta-analysis of controlled trials. *Neuropsychology*. 2014;28:829–845.

Klempin F, Kempermann G. Adult hippocampal neurogenesis and aging. *Eur Arch Psychiatry Clin Neurosci*. 2007;257:271–280.

Li W, Sun N, Kondracki A, Sun W. Sex, sleep duration, and the association of cognition: findings from the China Health and Retirement Longitudinal Study. *Int J Environ Res Public Health*. 2021;18:10140.

Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the *Lancet* Commission. *Lancet*. 2020;396(10248):413–446.

Love MN, Ruff G, Geldmacher D. Social cognition in older adults: a review of neuropsychology, neurobiology, and functional connectivity. *Med Clin Rev*. 2015;1(1):16.

Martin M, Clare L, Altgassen AM, Cameron MH, Zehnder F. Cognition-based interventions for healthy older people and people with mild cognitive impairment. *Cochrane Database Syst Rev*. 2011;(1):CD006220.

Morrison JH, Baxter MG. The ageing cortical synapse: hallmarks and implications for cognitive decline. *Nat Rev Neurosci*. 2012;13:240–250.

Ngandu T, Lehtisalo J, Solomon A, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *Lancet*. 2015;385:2255–2263.

*Pakkenberg B, Gundersen HJ. Neocortical neuron number in humans: effect of sex and age. *J Comp Neurol*. 1997;384:312–320.

Parambi DGT, Alharbi KS, Kumar R, et al. Gene therapy approach with an emphasis on growth factors: theoretical and clinical outcomes in neurodegenerative diseases. *Mol Neurobiol*. 2022;59(1):191–233.

Rahimi J, Kovacs GG. Prevalence of mixed pathologies in the aging brain. *Alzheimers Res Ther*. 2014;6:82.

Robinson JL, Molina-Porcel L, Corrada MM, et al. Perforant path synaptic loss correlates with cognitive impairment and Alzheimer's disease in the oldest-old. *Brain*. 2014;137(Pt 9):2578–2587.

Salthouse TA. When does age-related cognitive decline begin? *Neurobiol Aging*. 2009;30:507–514.

Sritawan N, Suwannakot K, Naewla S, et al. Effect of metformin treatment on memory and hippocampal neurogenesis decline correlated with oxidative stress induced by methotrexate in rats. *Biomed Pharmacother*. 2021;144:112280.

Stephan BC, Matthews FE, McKeith IG, Bond J, Brayne C; Medical Research Council Cognitive Function and Aging Study. Early cognitive change in the general population: how do different definitions work? *J Am Geriatr Soc*. 2007;55:1534–1540.

*Stern Y. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol*. 2012;11:1006–1012.

Tan, X, Gao L, Cai X, et al. Vitamin D3 alleviates cognitive impairment through regulating inflammatory stress in db/db mice. *Food Sci Nutr.* 2021;9:4803–4814.

van den Brink AC, Brouwer-Brolsma EM, Berendsen AAM, van de Rest O. The Mediterranean, Dietary Approaches to Stop Hypertension (DASH), and Mediterranean-DASH Intervention for Neurodegenerative Delay (MIND) Diets are associated with less cognitive decline and a lower risk of Alzheimer's disease-a review. *Adv Nutr*. 2019;10(6):1040–1065.

Wilson RS, Bennett DA, Bienias JL, Mendes de Leon CF, Morris MC, Evans DA. Cognitive activity and cognitive decline in a biracial community population. *Neurology*. 2003;61:812–816.

55.2d Psychological Changes With Normal Aging

Tina L. Mayes, Ph.D., Maya A. Reiter, M.S., Ph.D., and Jennifer J. Dunkin, Ph.D.

Advances in medicine and improvements in living conditions over the past century have resulted in a greater and greater number of persons reaching old age. This "silver tsunami" of the population is expected to reach its peak over the next 50 years as the baby boom cohort reaches senior citizen status. Projections estimate by 2050, 2 billion people will be older than 65 years. The oldest-old group of older adult individuals—those 85 years of age and older—is expected to grow at the fastest rate. Although on the surface, these demographic trends reflect favorably on society's attempts to eradicate disease, reduce infant mortality, and improve standards of living, recent debates have focused on the "hidden costs" of these changes. For example, aging is associated with increased risks of, among other factors, physical illness, impaired function, personal losses, and cognitive impairment. If these risks translate into significant impairment for the majority of older people, then there will be increased costs to society as a greater and greater number of people reach old age. Unfortunately, current health care systems focus on ways to manage existing diseases rather than targeting healthy aging. If there is not a shift in health care, the world may not be adequately prepared to address the large number of older adults with serious disabilities in the coming years.

From both public health and quality-of-life perspectives, it is important to determine how to maximize the health and well-being of the aged population. It is with this in mind that a number of large-scale longitudinal studies have focused on the concept of "successful aging" to determine what characteristics are present in individuals who appear to reach old age with favorable outcomes. The results of such studies have been important in shaping public policy as well as changing societal views on the nature of the aging process itself. For example, although there are clearly unique life challenges that face older adults, it may be surprising that the majority of research studies suggest that most older adult people report high levels of life satisfaction, cope well with the challenges of late life, and demonstrate considerable resilience in the face of functional and personal losses. Nevertheless, recent studies suggest that even successfully aging individuals are prone to their own negative aging self-stereotypes and that these self-stereotypes cause distress and appear to be related to poorer functional health. Thus, the challenges as a society are to increase the number of older adult persons reaching old age successfully and to decrease negative views of the aging process.

DESIGN AND SAMPLING ISSUES

Design Issues

Most studies in the field of aging are cross-sectional in design, meaning that groups of individuals of different ages are sampled at one point in time. Although meaningful conclusions about differences between age groups can be drawn from cross-sectional studies, it is inappropriate to draw inferences about change over time from such studies because individuals are not studied prospectively, and there is considerable potential for mistaking cohort differences for longitudinal change. For example, there are vast differences between the average 30 year old and the average 70 year old besides age, including the average number of years of education, exposure to historical events and experiences such as the Civil Rights Movement and Vietnam War, comfort with revealing personal information, and ability to navigate technology. Thus, what may resemble change over time may instead be baseline cohort differences. In addition, cross-sectional designs tend to overestimate the relationships between variables that change with time. Another significant issue involves an overutilized of datasets in the aging literature known as the *tragedy of the commons*. A recent review article points out that researchers frequently use the same dataset to publish multiple papers, thereby creating an inherent sampling bias. Because of these issues, it is likely that different trajectories of change will be necessary for each individual age cohort and some scholars call for restrictions, both self-regulation by researchers and through regulatory systems, to guard against the overuse of datasets in aging research.

Sampling Issues

Until recently, most studies of psychological issues in aging have used predominantly Caucasian, well-educated, healthy, and financially comfortable samples of older adults. Studies have primarily used samples of subjects in the so-called young-old age group (65 to 75 years of age) or, less frequently, the old-old age group (75 to 85 years of age), but rarely the oldest-old group (85 years of age and older). Furthermore, early longitudinal studies of personality development and adjustment used predominantly cisgender-male heterosexual samples. The result is that comparatively little is known about minority older adults, older adult women, LGBTQIA* older adults, inner-city and lower-income older adults, and the oldest-old. Psychological aspects of normal aging in persons with developmental disabilities such as autism spectrum disorder have also received much less attention, which is problematic in light of far higher prevalence rates of mental health as well as neurocognitive disorders in this population. Similarly, many psychometric and neuropsychological tests that measure cognitive functioning do not have adequate normative data for the oldest age groups. Attrition rates are also an important phenomenon: Studies rarely compare individuals who drop out of studies with those who remain in the study. This may overestimate the amount of stability and underestimate the amount of decline in psychological functions, as those who decline in functioning tend to have higher attrition rates than those who do not decline. In terms of cognitive aging, studies have varied considerably in terms of whether individuals with medical conditions known to affect brain function were included in the sample. Although such individuals may be “typical” of their age cohort, their inclusion in studies makes it difficult to determine which cognitive changes can be expected as a function of the aging process itself and which changes may be a function of epiphenomena associated with aging, such as medical conditions.

MENTAL HEALTH AND PERSONALITY

Mental Health Over the Life Span

With the exception of dementing disorders and delirium, most mental disorders are no more common in older adults as compared to younger age groups. Epidemiologic studies have found that prevalence estimates of common disorders such as major depression, anxiety disorders, substance abuse, and psychotic disorders are generally lower for those over the age of 65 versus younger age cohorts, a finding recently replicated by newly revised prevalence estimates of mental disorders using the NIMH Epidemiologic Catchment Area Program data, as well as the National Comorbidity Survey. Furthermore, longitudinal studies that have followed individuals for up to 25 years have found no evidence that those rates of emotional distress increase with age. However, certain specific mental health disorders may be either more prevalent or more severe in adults over the age of 65. For example, the progression of hoarding disorder symptoms, although beginning in adolescence, has been shown by evidence to peak in severity in older adults. Sleep quality and duration, which negatively impact emotional well-being, cognitive functioning, and mental health, have also been shown to decline in aging. Large-scale studies have shown that adults over the age of 65 have higher rates of insomnia symptoms (difficulty initiating sleep, maintaining sleep, and higher rates of early morning awakenings), with women showing higher vulnerability to difficulties with sleep during aging than men. This could mean that assessing sleep patterns and addressing sleep difficulties may be an important aspect of medical care in older adults.

Age differences in the experience and expression of both positive and negative emotions have also been studied to determine if older adults report more negative emotional experiences than younger adults. In a study using a series of large ethnically diverse samples, older adults reported experiencing fewer negative emotions such as sadness, anger, and fear but more positive emotions such as happiness than those in the younger age cohorts. In another study, even among older persons with physical illness, more positive than negative emotions were observed, lending further support to the notion that even advanced age and illness do not necessarily lead to negativity in outlook. In addition, older adults reported a greater sense of emotional control over negative internal emotional states than younger subjects. To explain their data, the authors proposed an *emotional control* model, suggesting that aging is associated with an increased ability to inhibit negative emotional states and maintain positive emotional states.

A study found interesting racial differences between older adult Blacks and Whites on measures of psychological distress. The study included over 1,000 high-functioning older adults between the ages of 70 and 79 and found that Blacks were less distressed than Whites, particularly at lower educational levels. The finding remained even after controlling for social support, socioeconomic status, and functional status. The authors suggest that older adult Blacks may be afforded higher status and stay more engaged within their communities than older adult Whites. If so, then the place of older Blacks within their communities could serve as a model for other ethnic or racial groups.

Stage Theories of Personality Development

Early personality theorists proposed that development was completed by the end of childhood or adolescence. One of the first developmental theorists to propose that personality continues to develop and grow over the course of the life span was Erik Erikson. Erikson believed that development proceeded through a series of *psychosocial stages*, each with its own conflict that is resolved by the individual with a greater or lesser degree of success. Erikson termed the crisis of the last epoch of life *integrity versus despair* and believed that successful resolution of this crisis involved a process of life review and achieving a sense of peace and wisdom through coming to terms with how one's life was lived. For example, Erikson proposed that a successful resolution of this crisis would be characterized by a sense of having lived one's life well, whereas a less successful resolution would be characterized by feeling that life was too short, that one did not choose wisely, and a sense of bitterness that one will not have another chance to live life over.

Erikson's theory is descriptive, and few studies have attempted to empirically validate it. However, a recent longitudinal study of approximately 500 subjects from two age cohorts found not only that the earlier age cohort scored significantly higher on a measure of integrity than the later age cohort but also that scores for both age cohorts on this measure had declined significantly by the final time of testing. These data suggest that the conflict of integrity versus despair may have a more favorable outcome in earlier age cohorts versus later age cohorts, raising the possibility that changing societal values have negatively impacted the struggle for integrity. Although more empirical data are needed, Erikson's theory remains one of the most eloquent descriptions of the potential of late life.

Personality Over the Life Span: Stability or Change?

Although Erikson and other stage theorists focused on unique developmental tasks and stages central to each phase of life, other studies have focused on defining core personality traits within the individual and determining their course over the life span. Studies have consistently revealed five basic personality traits: extraversion, neuroticism, agreeableness, openness to experience, and conscientiousness. Further, these personality traits generally remain remarkably stable over the life span. A number of large, well-designed longitudinal studies that have followed individuals over periods ranging from 10 to 50 years have found strong evidence for stability, although with a few notable exceptions. For example, some studies have found that neuroticism decreases as individuals move from young and middle adulthood to old age, after which it remains stable. Other studies have found slight decreases in extraversion as individuals move into the oldest-old age category, possibly linked to deterioration in health and sensory loss, particularly hearing.

Personality factors may even contribute to mortality in old age. For example, recent studies have found that high levels of extraversion and low levels of neuroticism are associated with a reduced risk of mortality in individuals 65 years of age and older. One explanation for this finding may be that personality traits mediate physical, social, and cognitive activity levels that lead to less healthy and supportive lifestyles.

ADJUSTMENT AND COPING IN OLDER ADULTS

Individuals can expect to experience periods of change, disruption, and even crisis throughout their life span. There is no doubt that older adults experience significant challenges as they enter and negotiate the last phase of life, including bereavement, loss of socially valued roles such as that of workers, and chronic and accumulating health problems. How age affects the process of managing change and discontinuity has been an important research focus, with some authors proposing that people become less effective as they age due to depletion in resources, whereas others have hypothesized that people become better at managing stressful life events over time, secondary to a process of learning more effective strategies. These questions have been investigated from a wide range of theoretical perspectives.

Age Differences in Coping Strategies and Coping Effectiveness

“Coping” is a general term that refers to strategies that individuals use to manage stressful life situations that involve perceived or actual threats. It includes strategies designed to reduce emotional distress, such as distraction, passive avoidance, and positive reappraisal, and strategies designed to solve the problem, such as direct action, confrontation, and information seeking. Cross-sectional studies of age differences in coping strategies have had somewhat mixed results, with some studies finding no differences between age groups in their use of coping strategies and other studies finding that older adults preferentially use strategies designed to reduce emotional distress, whereas younger adults use more problem-focused coping, although the effect sizes are relatively small. This finding has often led to the conclusion that older adults are more passive in their coping style, yet when the type of stressor is controlled, these differences are often eliminated or reduced. In addition, older adults are often attempting to cope with situations that are less changeable than those faced by younger individuals, and therefore, the use of coping techniques that attempt to reduce emotional distress rather than solve the problem may be a more effective approach. In this vein, a study found that older adults who modify their goals and accommodate and adapt to losses and stressors rather than attempting to exert control or deny losses have a better outcome.

Coping effectiveness across the life span has been another area of interest for gerontologists. The effectiveness of a coping strategy is defined by its outcome, usually measured by the level of distress, depression, or health symptomatology. Clearly, if individuals become less effective at coping with the demands of life as they age, this would have considerable implications for clinical practice and the kinds of interventions that might be used with older adults. However, studies of coping effectiveness have generally not revealed age differences; that is, the old appear to cope as well as the young, and their outcome, in terms of depression and other stress-related symptomatology, is just as favorable. For example, in a study investigating age differences in coping across a broad range of stressors using a sample of over 2,000 men ranging in age from the late 40s to over 90, there were no significant age differences in the reporting of negative emotional states in response to stressors or perceived efficacy in coping.

Role Transitions

Role transitions play an important part in the aging process, and there is no domain where it is more important to understand issues of coping and adaptation than during transitional phases of life. There are many transitions experienced by older adults, but older adults are especially likely to experience two major role transitions: retirement, or the transition from worker to nonworker, and bereavement, or the transition from spouse to widow. These major transitions present the individual with the challenge of having to redefine identity and create new roles to replace those that have been lost. Indeed, there is evidence that older adults who are successful in creating new roles for themselves experience a greater sense of well-being than those who are less successful at this process.

Retirement. Retirement is an age-graded event in which individuals in society are expected to retire at a particular age, regardless of health status or personal preference. It is almost entirely a phenomenon of Western, industrialized, and postindustrialized societies and, prior to the late 19th century, was largely a personal decision. The transition and adjustment to retirement have been characterized by social gerontologists as a *psychosocial crisis*, and the image of the successful executive deteriorating or dying soon after retirement has been popularized in the press. However, recent studies have suggested that most recent retirees rank retirement quite low in terms of its impact on a list of psychosocial stressors, with 17% of men and 11% of women reporting dissatisfaction with retirement. Nevertheless, there is some heterogeneity among retirees, and the factors that predict a poor adjustment to retirement include an involuntary or unexpected retirement, low socioeconomic status, health problems, a competitive and highly achievement-oriented personality (the so-called Type A behavior pattern), inability to set future goals, and general difficulty adjusting to change. Further, despite early data suggesting that there is an increased risk of death in the first 2 years after retirement, studies controlling for health status at retirement have failed to reveal any associations between retirement and mortality. In addition, when preretirement mental and physical health are controlled for, there are no clear associations between retirement and subsequent physical or mental health consequences. Thus, it is important to consider retirement in the context of an individual's life and other changes that they may be experiencing rather than in isolation.

Few studies of retirement have specifically investigated ethnic differences in adjustment. For many ethnic minority groups, retirement represents a significant loss of income because many are economically disadvantaged and earn lower wages relative to their Caucasian counterparts. Available data suggest that many continue to work past retirement age, even in the face of disability and health problems. A recent study suggests that middle-aged African American males can expect to spend, on average, 2.5 fewer years retired than Caucasian males and, therefore, can expect to spend more time in the workforce with fewer years of retirement and leisure in their old age.

Spousal Bereavement. Demographic data suggest that 51% of women and 14% of men over the age of 65 will be widowed at least once. Bereaved spouses consistently rank spousal loss as among the most stressful of all life experiences. Early work on bereavement and older adults hypothesized that older adults might be less able to cope with the loss of a spouse than younger adults due to declining physical capabilities and subsequent decrements in the ability to manage stress. Subsequent research has revealed age differences in coping with bereavement but in the opposite direction as predicted: Younger bereaved spouses report greater symptomatology on measures of physical and mental health following bereavement than older spouses. Several factors might account for this difference, including cohort effects, with younger cohorts having a greater comfort level in acknowledging symptoms or whether the death was expected or unexpected, with unexpected deaths causing greater symptomatology and occurring with greater frequency in younger age groups.

Other data are consistent with the idea that, as a group, older adults appear to have a more favorable outcome than expected following the death of a spouse. A longitudinal study comparing over 120 older bereaved spouses with age-matched nonbereaved controls found that depressive symptoms peak within the first few months after a death but decline significantly within a year, and after 2 or 3 years of follow-up, although bereaved groups still report significantly higher levels of depressive symptoms than age-matched controls, these differences are very small. Even during the first few months following bereavement, few older adult widows (between 5% and 17%) score in the severely depressed range on depression rating scales. Interestingly, although older adult bereaved spouses report that loneliness and performing activities of daily living are factors that complicate adjustment, many older spouses also report positive feelings following a death, such as relief, in contrast to younger samples. There are nevertheless important exceptions;

studies suggest that older adult survivors of spouses who committed suicide, who have low social support, and who report greater psychopathology at the time of death are all at a higher risk of poorer long-term adjustment. In addition, older adult women who are highly dependent on their spouses for assistance appear to be at a higher risk for adverse outcomes and may require more professional intervention.

Recent data reveal an interesting relationship between bereavement and an excess risk of mortality in older adults. Most studies have revealed that actually younger bereaved spouses face a greater risk of mortality than older recent widows. In addition, older adult White widows appear to be at a greater increased risk of mortality than older adult Black widows, who may not face increased risk at all.

Resilience of the Aging Self

The above review of coping and adaptation in aging suggests that older adults appear to generally cope well with adversities and challenges despite, in many cases, the more chronic and accumulating nature of the stressors that they face. The concept of *resilience* has been offered as an explanatory factor for this phenomenon and is defined as the process of both maintaining adaptive behavior and positive emotions when presented with stress as well as recovery from adversity. Resilience may involve a process of *selective optimization* or an accommodative process of rescaling goals to match one’s limitations and the demands of the environment and investing fewer resources available in personally salient activities that maintain self-esteem. For example, a study found that the psychological impact of physical disability was less for those elders over 80 than that in those younger than 80, suggesting that a process of scaling down expectations with increasing age may occur and may be adaptive.

Maintaining a strong sense of self-worth in the face of accumulating losses may be an important part of resilience in aging. For example, a recent study found that self-esteem actually declined very little over time in a sample of elders with low socioeconomic status and low levels of education, despite experiences of personally meaningful losses. Rather, these older adults appeared to engage in cognitive strategies to protect their sense of self and well-being. In another study, even individuals over the age of 100 who experienced multiple losses in many domains report similar levels of happiness to middle-aged and young-old individuals, and attitudes such as optimism and self-efficacy are more predictive of happiness than objective resources available.

SOCIAL FUNCTIONING IN LATE LIFE

Research over several decades has consistently validated the importance of social support in “buffering” the effects of stress: Given equivalent stressors, those individuals with strong social ties appear to have better outcomes than those who are isolated or perceive that support is unavailable or unhelpful from those in their network. Yet, older adults appear to have reduced social networks, and the frequency of their social contacts is reduced as compared to those of younger adults, and this may be especially true for the oldest-old, whose social networks have been shown to be one-half the size of those of the young-old. This finding was previously thought to be evidence for the *disengagement* hypothesis of aging, that is, that aging involves a process of withdrawal from society and others and the development of a more inward focus. However, for many older adults, the reduction in social contact is not voluntary and may be more related to disability and reduced mobility than preference. The disengagement hypothesis has been recently reframed as *socioemotional selectivity*, based on the finding that although overall social contacts do decrease in late life, it is only contacted with acquaintances and other peripheral individuals that decreases; social contacts with very close friends and family remain stable, and older adults’ satisfaction with these close relationships remains high. It may be that as individuals age and resources become fewer, those resources are adaptively invested in maintaining close relationships as opposed to maintaining acquaintanceships. For older adults with a *limited future time perspective* or a greater awareness of mortality, there is even more motivation to maintain emotionally meaningful relationships. However, this phenomenon may be at least partially dependent on the external circumstances of older people's lives. For example, in a study of over 500 older adult individuals over the age of 70, the size of social networks and the feelings of connectedness to that social milieu were related to the presence and availability of the nuclear family. Older adults with nuclear family members tended to have social networks that were both larger as well as more emotionally close than older adults without a nuclear family. However, those without a nuclear family reported greater feelings of closeness to friends than those with a nuclear family, suggesting that older adults can adapt to the absence of family by creating satisfying relationships with others.

The social structure of the oldest-old has recently been studied, and data suggest that this group may be particularly vulnerable to social isolation due to multiple losses. In a recent study, most individuals over the age of 85 had at least one surviving relative, but approximately one-quarter had little or no contact with that relative. Most social contacts were derived from children, and for those individuals without children, social contact, and support were not provided from more distant relatives, except within the African American community, where other relatives and friends were available to provide help and support. For the oldest-old, then, the family, especially adult children, appears to play a crucial role in providing social support.

Sexuality in Adults Over the Age of 65

Sexuality is an important aspect of personal life, well-being, personality, and identity. As a group, adults over the age of 65 continue to engage in sexual behavior. That is to say that they experience sexual desire, have sexual thoughts, and are sexually active (both with themselves and with partners). However, as in other life stages, there is a great deal of individual variability in sexuality and sexual behavior in adults over the age of 65. Sexuality bidirectionally interacts with life satisfaction and mental as well as physiologic health in older adults. Indeed, research suggests that sexually active older adults view their health more positively and that a decline in sexuality in older adults is associated with poorer well-being. Although recent trends in sex positivity have somewhat dismantled the widespread myth that older adults stop enjoying and expressing their sexuality, opposite but equally problematic biases may be emerging, such as the misconception that physical sexual activity is a ubiquitously *essential* part of effective aging. Although sexual health is a highly relevant health care domain for older adults, sexuality concerns may be overlooked far too often in medical settings. Research suggests that adults over the age of 65 can experience more barriers to accessing medical care related to sexuality concerns than younger adults do. This could be related to heightened sex-related stigmas about older adults (affecting both patients and providers). Specifically, research suggests that providers inquire about sexuality-related concerns less frequently in older adults as compared to patients in other life stages. Given that it can be easier for patients to discuss concerns related to sexuality when the provider raises the topic, mental and medical health care workers can improve age-sensitive care by incorporating sexuality-related questions into their routine practice with older adults.

Numerous aspects of normal aging that differ tremendously across individuals can impact sexuality and sexuality-related concerns in adults over the age of 65. These include but are not limited to medical conditions (and effects of treatments for these conditions), life stressors, loss of a spouse and spousal illness, substance use, sexual dysfunction (in individuals with all types of genitalia), and normal endocrinologic changes (which may differ by biologic sex). While age-associated changes may be related in general to a reduction in physical sexual activity, this reduction is not necessarily related to decreased intimacy between romantic partners. Notably, evidence suggests that individuals with major neurocognitive disorders may experience multiple sexuality-related impacts, including decreased sexual responsiveness to their partners as well as increased sexually inappropriate behaviors. Spouses who experience a role transition into providing care for their partners report a reduction of sexual intimacy with their partners.

Intersectionalities between sex assigned at birth, gender identity, sexual orientation, culture, and religious background, as well as distinct cohort-related experiences, may influence the expression of sexuality in adults over the age of 65. Attitudes about sexuality and gender-related roles and expectations about sex can vary across individuals and cohorts, necessitating sensitivity on the part of clinicians in discussing sexuality-related topics with patients. There are many specific sexuality-related concerns that impact sexuality in LGBTQIA* older adults. For example, LGBTQIA* adults over the age of 80 experience increased victimization related to gender identity and sexual orientation and have higher internalized stigma compared to adults in the 50- to 64- and 65- to 79-year-old age range. In addition, living arrangements may disproportionally impact the expression of sexuality in LGBTQIA* older adults, as fear of victimization and bias related to sexual orientation may hinder sexual behavior while living with family members or even in assisted living communities or nursing homes. Although LGBTQIA* older adults may be more predisposed to such barriers to expressing their sexuality, many LGBTQIA* individuals are thriving, maintain highly satisfying sex lives, and relate extremely positively to their sexual orientation and sexuality.

COGNITIVE AGING

The study of cognition and aging has been an extraordinarily productive area of research over the past several decades. Nevertheless, the existing literature is fraught with difficulties in interpretation, confusion, or disagreement about the inclusion, or exclusion of subjects with medical problems and multiple medications and the inadvertent inclusion of subjects with early dementia, which tends to overestimate the effects of age by lowering the overall mean of older adult groups. Furthermore, until magnetic resonance imaging (MRI) scanning was widely used, studies unknowingly included subjects with subcortical ischemic disease, confounding the effects of age with disease. In addition, design issues are important; cross-sectional studies, in which individuals of different age groups are compared at one point in time, almost always show broader and sharper age effects than longitudinal studies due to possible confounding of age effects with cohort effects.

Summary of Age Effects on the Major Cognitive Domains

The broad domain of cognition has been divided into subdomains that include attentional processes, language functions, visuoperceptual abilities, learning and memory, processing speed, and executive functions.

Attentional Processes. Attentional processes are a complex group of behaviors by which individuals select information for cognitive processing while ignoring competing information, concentrate on information over a sustained period of time, divide attentional focus between two or more aspects of a task at a time, and shift attentional focus from one set of stimuli to another and back again. Basic attention span is the amount of information that can be apprehended at one time and demonstrates little or no relationship to age. Similarly, sustained concentration or vigilance is not affected by the aging process. However, complex attentional tasks such as divided attention that place more demand on cognitive capabilities show significant declines with advancing age.

Language Functions and Verbal Abilities. Language functions remain remarkably stable with age, with few exceptions. Vocabulary and expressive word knowledge may even increase over the life span. Verbal expression and verbal reasoning skills may demonstrate subtle changes with age, as some studies have shown that spontaneous language may become less precise and more repetitious over time, although most of these conclusions were drawn from cross-sectional studies and may be confounded with cohort effects. Finally, the ability to spontaneously retrieve names of objects from the lexicon appears to show reliable but small decreases with age, although these effects are heavily influenced by educational achievement.

Visuoperceptual and Visuoconstructive Abilities. Visuoperceptual abilities are those functions that involve the recognition and appraisal of nonverbal relationships, specifically those involving shapes, patterns, and objects with spatial relationships to one another. Visuoconstructive abilities are those involving the reproduction of nonverbal material, such as the direct copy of simple and complex designs. Visuoperceptual abilities demonstrate relatively consistent age effects across studies, with older adults generally performing as well as younger adults on tests of simple object and pattern recognition but performing more poorly than younger adults on measures of visual closure and spatial orientation. Similarly, in terms of visuoconstructive functions, older adults perform about as well as younger adults when copying simple two-dimensional designs but more poorly than younger adults when copying complex three-dimensional designs and when planning and strategy are necessary for good performance. It should be noted that age effects are the largest on visuoconstructive tasks that are timed, and removing the speed component of the task often considerably reduces but does not remove the age difference.

Learning and Memory. Effects of age on learning and memory are among the most well-documented and well-studied effects in the neuropsychological literature. However, “memory” is an extremely complex set of functions and systems, and the aging process affects some components of memory to a greater degree than others. Sensory memory, or the extremely brief sensory record of a stimulus, primary memory, exemplified by the ability to recall a string of digits long enough to repeat them, and tertiary memory, or memories from childhood, show no significant age effects or mild and subtle effects that can be detected in the laboratory. Research studies have consistently shown that secondary memory, exemplified by the ability to learn information and recall it after a time delay, is most strongly affected by age. Nevertheless, the age-related effects on secondary memory are relatively small: Studies suggest that performance declines at a rate of 10% per decade of life, beginning in midlife. However, once older adults learn information, it is retained well; studies suggest that older adults retain information at rates only slightly below that of younger adults. In addition, available data suggest that spontaneous recall demonstrates considerably stronger age effects than recognition memory, which is the ability to recognize information that has been previously presented.

Processing Speed. Speed of information processing and pure motor speed demonstrate the most reliable age differences and decrements with age, whether measured cross-sectionally or longitudinally. Indeed, reduced processing speed in aging may be a crucial component of many age-related effects on cognition, such that in a recent study, age effects on a number of neuropsychological tests were considerably diminished after controlling for processing speed, even on tests where the speed of performance was not measured. In addition, studies suggest that speed of processing is an important mediator of memory performance, such that some authors have proposed a speed of processing model to account for age-related decrements in memory and other cognitive processes. Evidence suggests that slower processing speed generally leads to slower reading, but typically their relative ability to contextualize what they read remains intact.

Executive Functions. The term “executive functions” refers to superordinate cognitive skills involved in abstract reasoning, problem solving, planning, response inhibition, the ability to shift cognitive set, and a working memory construct that is defined as the ability to simultaneously hold and process information. Early studies suggested that executive functions decline with age. However, studies that use carefully selected samples and exclude subjects with histories of chronic medical conditions known to affect central nervous system (CNS) functioning or documented evidence of subcortical white or gray matter lesions on MRI scanning reveal very few age differences in executive tasks. Because of the relationship of performance on tests of executive functioning to the presence of medical conditions, especially diseases of the cardiovascular system, these studies suggest that some of these age effects may be preventable.

Models of Cognitive Aging

Taken together, studies suggest that there is a profile of cognitive changes with age that includes significantly reduced information processing speed and pure motor speed and mild decrements in spontaneous recall, executive skills, complex attentional processes, and complex visuoperceptual and visuoconstructive abilities. It should be emphasized that although age-related changes in cognition are statistically significant, most older adult persons still perform within normal limits on cognitive testing. The extent to which these observed changes are related to the process of aging itself, the effects of particular medical conditions on the brain and CNS functioning, and cohort differences in education and experience are not well understood at present, although a recent longitudinal study with a 10-year follow-up interval revealed that significant decline in cognitive functions (excluding psychomotor speed) was apparent only in individuals suffering from poor health. A number of models have been proposed to account for these changes, and these models have included both psychosocial and biologic interpretations. First, it is unlikely that all age-related effects on cognition are secondary to cohort differences because cohort-sequential studies, in which individuals from different age cohorts are followed prospectively, reveal clear longitudinal decline across age cohorts. Unfortunately, studies specifically addressing this important question are few and have only investigated intellectual abilities and not functions such as memory or other neuropsychological variables. It is likely that cohort effects exert the strongest influence on those abilities most dependent on formal education, but this hypothesis has not been formally tested. It is also unlikely that simple disuse accounts for significant aspects of age-related cognitive decline.

The Cattell–Horn model of intellectual changes over the life span proposes that age affects different cognitive processes differently. The model proposes that general intelligence can be subdivided into *crystallized intelligence*, referring to abilities, and information learned through exposure to education and life experience, and *fluid intelligence*, referring to abilities involved in novel problem solving. The model proposes that crystallized intelligence remains stable or increases over the course of the life span and may even increase into old age due to continuous exposure to information. In contrast, fluid intelligence, thought to be dependent on the integrity of brain function, was hypothesized to decline with advanced age. Available data have generally supported this model, with studies showing that there is little evidence of a decline in measures of crystallized abilities such as word knowledge or general fund of information, whereas measures of fluid abilities such as novel problem solving, as discussed above, generally reveal some decline. However, the model, as originally proposed, is limited in that it only discusses changes in intellectual functions, whereas other cognitive functions, such as learning and memory, are not included. Furthermore, the same limitations of studies of executive functions apply to the fluid intelligence concept.

The processing speed hypothesis hypothesizes that many of the observed age-related changes in cognition are, in fact, secondary to reductions in the speed of processing. The model proposes that the speed at which information can be processed limits both the quantity and the quality of cognitive operations. Slowing affects not only how much information is processed within a given period but may also dilute and weaken the associations between neural networks, leading to more impoverished cognitive processes. The model has received considerable empirical support, with most studies finding that a general speed factor is an important mediator of other cognitive processes, such as memory.

Neuropsychological models of age-related decline have attempted to integrate findings from cognitive testing with findings from brain imaging and neuropathologic studies to suggest possible etiologies of cognitive changes. The most promising of these is a model variously termed the frontal deficit hypothesis or the executive function hypothesis, based on the finding that the pattern of cognitive performance in normal aging, although less pronounced, is similar to that seen in many dementing processes affecting primarily subcortical–frontal circuitry. In addition, neuropathologic data suggest that frontal–subcortical structures are the most vulnerable to aging, with many studies revealing that neuronal loss and degeneration of dendritic arborization are most prominent in the frontal cortex and in some subcortical areas. Further, there are definite changes in the vasculature of the brain with age that result in ischemic white and subcortical gray matter disease that predominantly affects cognitive processes mediated by frontal–subcortical circuits. Finally, one study found that age-related changes in the dopamine system, particularly declines in D_2 receptor availability in some basal ganglia structures, were related to age-related declines in performance on both motor and executive systems tasks. However, most studies have found that executive functions appear to decline at the same rate as other cognitive functions and are not among the first age-related cognitive changes to be observed. This casts some doubt on their central role as mediators of age-related cognitive decline.

Finally, some authors have proposed a “common cause” model, suggesting that decrements in sensory functions, including visual and auditory acuity, may explain age-related changes in cognitive processes. As with the processing speed hypothesis, empirical support for this idea is generally strong, with most studies finding that sensory function is an important predictor of age differences in a number of cognitive processes. Nevertheless, it is clear that an as-yet-unnamed variable or group of variables remains unexplained in accounting for variance in cognitive aging because most studies find that age-related variance remains even after controlling for speed and sensory abilities.

It is important to note that the line between a “normal” process of cognitive aging and a pathologic process is not well drawn at present. For example, there is considerable disagreement in the field about the inclusion or exclusion of older adult subjects with medical conditions known to affect brain function or individuals with white matter lesions in studies of normal aging and cognition. Because these conditions affect such a high proportion of functional older adult persons, their exclusion may lead to erroneously high estimates of normal expectations of performance and the subsequent classification of many individuals as impaired. In addition, many older individuals without known cerebrovascular risk factors or other medical conditions have periventricular white matter lesions on MRI scanning, further confusing the line between disease processes and normal aging. At present, until this demarcation is better defined, the study of “typical” aging, which attempts to characterize the majority of older adult people, as well as the study of “optimal” aging, which attempts to characterize the ideal outcome of the aging process as well as characterize changes that are a function of aging itself and not epiphenomena of other conditions, are both important pursuits, as long as the research questions are appropriate to the sample being used.

Finally, it may be that older adults’ performance on cognitive testing may be affected by their own ageist stereotypes. A recent study found that when, prior to cognitive testing, older adults were “primed” with positive images of aging, their performance improved on measures of memory, as well as measures of memory self-efficacy, in contrast to older adults who were primed with negative images of aging.

SUCCESSFUL AGING

In recent years, there has been a paradigm shift in gerontology, from a view of aging as an inevitable deterioration to one that includes the potential for development and growth. In this section, a number of studies were reviewed that, taken together, support the notion that the process of aging can be managed well by the majority of older adults. In this vein, the construct of *successful aging* was created. In the process, however, there has been considerable debate about how to define successful aging. It has become clear that if one defines successful aging as the absence of deficits in physical, cognitive, and social functioning, then few older adults would be so classified. A recent study examined both objective and subjective indicators of successful aging and found that only 10% of their sample could be classified as “successfully aged” based on objective criteria that included quantitative assessments of physical, cognitive, and social functioning and measures of well-being. However, when successful aging was defined as a subjective sense of having adapted well to the demands of life or being self-rated, many more individuals were characterized as having aged successfully. In a recent review of data-based studies, the proportion of successfully aged individuals ranged from 0.4% to 95%, thus underscoring this ongoing debate. In general, older adults rate social engagement and positive thinking as more important than physical functioning in their conceptualizations of successful aging. It is likely that any definition should include both objective and subjective definitions of success.

Studies have found that high education, absence of alcohol abuse and cigarette smoking, absence of depression, maintaining an appropriate weight, regular exercise, and social support are all significant predictors of successful aging. Social class, however, does not appear to be an important predictor. Intrinsic psychological factors such as resilience, spirituality, and optimism have also been associated with successful aging. As more is learned about successful aging, both in terms of its psychosocial correlates and biologic correlates, it may be that this knowledge can be translated into useful interventions that can improve the overall number of older adults reaching old age successfully.

IMPLICATIONS FOR CLINICAL PRACTICE

One of the major points of this section is that, with the exception of cognitive processes, age in and of itself is rarely an important predictor of outcome. Studies of mental health over the life span, coping and adjustment, and personality often fail to reveal any significant age effects when other factors are controlled. In addition, research has revealed an often surprising degree of resilience in older adults. Therefore, the clinician working with older adults should not expect that age itself is a risk factor for poor outcomes. Unfortunately, this notion has probably been partially responsible for the underidentification of depression in many older adult cohorts; many clinicians used to believe that depression was a “normal” response to the aging process. Similarly, in aging adults with developmental disabilities such as autism, there may also be an underidentification of mental health concerns due to a focus on the disability and not on an individual’s particular circumstances. Regardless of disability, clearly, the life themes that older adults struggle with are different from those of younger adults, and the clinician should be prepared to deal with issues of loss of family and friends, loss of independence that leads to increasing dependency, and the struggle to come to terms with the past. In addition, the concept of resilience may be useful in psychotherapeutic work with older adults by using their own strengths to address the challenges associated with aging. Clinicians can encourage patients to invest resources selectively in activities that they find especially gratifying. Moreover, clinicians can enlist the help of family members and encourage basic lifestyle choices (i.e., diet, quitting smoking, and exercise) to further support a better quality of life.

The important exception to some of these conclusions may be individuals in the oldest-old category or those over the age of 85. Research has shown that the oldest-old may be at a higher risk of social isolation than younger age groups, especially those without living children. Clinicians should be alert to the unique life situations faced by individuals in this age group, who may be especially vulnerable not only to isolation but also to increasing frailty and the cumulative deleterious effects of chronic medical conditions. Interventions that both provide social support as well as encourage alternative sources of social support, such as senior centers, may be especially helpful for individuals in this age group without living family.

In terms of cognitive processes, there are clear decrements in some but not all cognitive processes over the life span. However, with the exception of processing speed, these changes are generally of small magnitude. The result is that the typical changes in cognition should not significantly impact the clinical situation, either the psychotherapeutic encounter or the psychiatric consultation. Some older adults may require more repetition and explanation of material, especially material that is novel and highly abstract, but this can easily be managed with modifications of psychotherapeutic approaches. The pattern of cognitive decline seen in normal aging discussed in this section should alert the clinician to changes that deviate from this pattern that might be cause for concern, especially such things as difficulties with recalling information even after cueing, problems with basic attentional processes, problems with constructing basic two-dimensional designs, or language impairment. At the same time, clinicians need to be alert to the effects of negative age-related stereotypes and their effects on both performance and self-concept.

Finally, it is worth noting that because research findings are based on the central or average tendencies of groups of individuals, the conclusions drawn often mask underlying heterogeneity. This may be especially true with older adults, as research has shown that the variability of almost every construct studied, such as personality, cognition, health status, and income, increases over the life span. It is likely that this increased heterogeneity is a complex interactive process of increasingly individualized expressions of genetic predispositions in response to accumulating demands from the environment over the life span. Thus, the clinician working with older adults may have less precise guidelines or standards for normative expectations for older adult patients because older adults can present with a much wider range of behavior, personality, and life experience than is found in younger adults.

FURTHER READINGS

Aldwin CM, Sutton KJ, Chiara G, Spiro A 3rd. Age differences in stress, coping, and appraisal: findings from the Normative Aging Study. *J Gerontol B Psychol Sci Soc Sci*. 1996;51(4):P179–P188.

Berg AI, Johansson B. Personality change in the oldest-old: is it a matter of compromised health and functioning? *J Pers*. 2014;82(1):25–31.

Carr D, House JS, Kessler RC, Nesse RM, Sonnega J, Wortman C. Marital quality and psychological adjustment to widowhood among older adults: a longitudinal analysis. *J Gerontol B Psychol Sci Soc Sci*. 2000;55(4):S197–S207.

Cath DC, Nizar K, Boomsma D, Mathews CA. Age-specific prevalence of hoarding and obsessive compulsive disorder: a population-based study. *Am J Geriatr Psychiatry*. 2017;25(3):245–255.

Cheimariou S, Farmer TA, Gordon JK. The effects of age and verbal ability on word predictability in reading. *Psychol Aging*. 2021;36(4):531–542.

Depp CA, Jeste DV. Definitions and predictors of successful aging: a comprehensive review of larger quantitative studies. *Am J Geriatr Psychiatry*. 2006;14(1):6–20.

Erikson E. *Childhood and Society*. Norton; 1950.

Ferraro K, Wilmoth JM. Gerontology: *Perspectives and Issues*. 4th ed. Springer Publishing Company; 2013.

Folkman S. Coping across the life span: theoretical issues. In: Cummings EM, Greene AL, Karraker KH, eds. *Life-Span Developmental Psychology: Perspectives on Stress and Coping*. Lawrence Erlbaum Associates; 1991:3.

Gross JJ, Carstensen LL, Pasupathi M, Tsai J, Skorpen CG, Hsu AY. Emotion and aging: experience, expression, and control. *Psychol Aging*. 1997;12(4):590–599.

Gulia KK, Kumar VM. Sleep disorders in the elderly: a growing challenge. *Psychogeriatrics*. 2018;18(3):155–165.

Irving P. The *Upside of Aging: How Long Life is Changing the World of Health, Work, Innovation, Policy, and Purpose*. Wiley Publishing; 2014.

Jopp D, Rott C. Adaptation in very old age: exploring the role of resources, beliefs, and attitudes for centenarians' happiness. *Psychol Aging*. 2006;21(2):266–280.

Kocevska D, Lysen TS, Dotinga A, et al. Sleep characteristics across the lifespan in 1.1 million people from the netherlands, united kingdom and united states: a systematic review and meta-analysis. *Nat Hum Behav*. 2021;5(1):113–122.

Kok AAL, Aartsen MJ, Deej DJH, Huisman M. Socioeconomic inequalities in a 16-year longitudinal measurement of successful ageing. *J Epidemiol Community Health*. 2016;70(11):1106–1113.

Kolodziejczak K, Rosada A, Drewelies J, et al. Sexual activity, sexual thoughts, and intimacy among older adults: links with physical health and psychosocial resources for successful aging. *Psychol Aging*. 2019;34(3):389–404.

Kubzansky LD, Berkman LF, Seeman TE. Social conditions and distress in elderly persons: findings from the MacArthur Studies of Successful Aging. *J Gerontol B Psychol Sci Soc Sci*. 2000;55(4):P238–P246.

Lang FR. Regulation of social relationships in later adulthood. *J Gerontol B Psychol Sci Soc Sci*. 2001;56(6):P321–P326.

Levy BR. Mind matters: cognitive and physical effects of aging self-stereotypes. *J Gerontol B Psychol Sci Soc Sci*. 2003;58(4):P203–P211.

Lund DA. Conclusions about bereavement in later life and implications for interventions and future research. In: Lund DA, ed. *Older Bereaved Spouses:*

Research with Practical Applications. Hemisphere Publishing Corporation; 1989:3.
MacLeod MD, Saunders J. Episodic memory and age-related deficits in inhibitory effectiveness. *Exp Aging Res*. 2017;43(1):34–54.
Mehta M, Whyte E, Lenze E, et al. Depressive symptoms in late life: associations with apathy, resilience and disability vary between young-old and old-old. *Int J Geriatr Psychiatry*. 2008;23(3):238–243.
Mejía ST, Ryan LH, Gonzalez R, Smith J. Successful aging as the intersection of individual resources, age, environment, and experiences of well-being in daily activities. *J Gerontol B Psychol Sci Soc Sci*. 2017;72(2):279–289.
Mroczek DK, Weston SJ, Graham EK, Willroth EC. Data overuse in aging research: emerging issues and potential solutions. *Psychol Aging*. 2022;37(1):141–147.
Narrow WE, Rae DS, Robins LN, Regier DA. Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. *Arch Gen Psychiatry*. 2002;59(2):115–123.
Schaie KW. A lifespan developmental perspective of psychological aging. In: Laidlaw K, Knight B, eds. *Handbook of Emotional Disorders in Later Life: Assessment and Treatment*. Oxford University Press; 2008:1.
Sinkovic M, Towler L. Sexual aging: a systematic review of qualitative research on the sexuality and sexual health of older adults. *Qual Health Res*. 2019;29(9):1239–1254.
Thumala Dockendorff DC. Healthy ways of coping with losses related to the aging process. *Educ Gerontol*. 2014;40(5):363–384.
Vliet ECV, Manly J, Tang MX, Marder K, Bell K, Stern Y. The neuropsychological profiles of mild Alzheimer's disease and questionable dementia as compared to age-related cognitive decline. *J Int Neuropsychol Soc*. 2003;9(5):720–732.
Wilson RS, Krueger KR, Gu L, Bienias JL, Mendes de Leon CF, Evans DA. Neuroticism, extraversion, and mortality in a defined population of older persons. *Psychosom Med*. 2005;67(6):841–845.

55.2e Neuropsychological Evaluation

Gauri Savla, Ph.D., and Barton W. Palmer, Ph.D.

As medical technology advances rapidly, people are living longer and are active well into traditional definitions of "old age." Nevertheless, this particular demographic is also at the greatest risk for cognitive deficits related to cerebrovascular, systemic, and neuropsychiatric conditions, as well as long-term use of certain medications. Comprehensive clinical care of older adults, therefore, should almost invariably include some consideration of neuropsychological functioning. Cognitive decline is a core component of neurodegenerative conditions such as Alzheimer disease, dementia with Lewy bodies (DLB), Parkinson disease, as well as cardiovascular conditions such as strokes. Chronic metabolic illnesses (e.g., metabolic syndrome, diabetes) and long-standing autoimmune diseases (e.g., multiple sclerosis) may also be associated with cognitive impairments. Some infectious diseases may also affect cognitive functioning. For example, in the wake of the COVID-19 pandemic, there have been numerous reports of “brain fog” among those experiencing “long-haul” symptoms.

Psychiatric disorders with a typically earlier onset that often extend into later life, such as primary psychotic disorders or severe mood disorders, are also often (but not always) associated with mild to moderate deficits in neuropsychological functioning. These deficits tend to remain more or less stable over the course of the illness (i.e., there are rarely sharp declines in later life without the co-occurrence of a new-onset neurodegenerative illness). An important point to consider is that cognitive deficits, rather than symptoms of primary psychopathology, in schizophrenia and bipolar disorder have a much stronger predictive relationship to the level of adaptive functioning.

Consequently, an evaluation of neuropsychological functioning can be a critical component of effective diagnosis and long-term treatment and care planning in geriatric psychiatry settings.

WHAT IS A NEUROPSYCHOLOGIST?

Neuropsychology as a field of scientific inquiry is not a profession or degree unto itself but falls within the broader field of clinical psychology. The term “clinical neuropsychologist” refers to a licensed clinical psychologist who, in addition to the broader training required for a doctorate and licensure in clinical psychology, has also completed additional training in structural and functional neuroanatomy, neuropathology, and neurobehavioral syndromes. A neuropsychologist, therefore, is an expert in scientifically grounded brain–behavior relationships.

WHAT IS A NEUROPSYCHOLOGICAL EVALUATION?

Standardized neuropsychological tests that measure cognitive abilities and deficits form the core of neuropsychological evaluation. However, neuropsychological evaluation involves much more than administrating and scoring tests. A valid and clinically useful interpretation of test performance requires the neuropsychologist to consider the examinee’s performance within the broader context of their presenting concern as well as prior medical history, psychosocial history and functioning. Thus, in addition to the standardized assessment of neurocognitive abilities, comprehensive clinical neuropsychological evaluations generally include a thorough review of available medical records as well as a clinical interview of the patient and family or other collateral information sources. Personality and emotional functioning are also generally assessed via clinical interviews, behavioral observations, clinical rating scales, and/or standardized psychological tests.

HISTORICAL DEVELOPMENT

The emergence of neuropsychology as a clinical discipline can be traced back to the 1930s and 1940s, when several pioneers of the field, such as Ward Halstead and Arthur Benton, began their studies of brain–behavior relationships using psychometric methods. Many of the tests developed or adapted by Halstead (and the extended battery of tests subsequently expanded by Ralph Reitan), as well as those by Benton, remain core components of contemporary neuropsychological testing. There were also early efforts in the psychometric testing of intelligence and other cognitive functions, as exemplified by the work of Alfred Binet, and the Army aptitude testing programs that emerged during World War I. (These are, by the necessity of space, only selected examples of a more complex history of psychometric testing.) Another key contributor to the development of clinical neuropsychology was David Wechsler. In 1939, Wechsler published the first version of his Bellevue Intelligence Examination, followed in 1945 by its companion, the Wechsler Memory Scale. Many of the subtests in the Wechsler batteries were adapted from earlier sources. As collective batteries and with appropriate revisions and updates in the ensuing decades, the Wechsler scales have remained among the most widely used measures of cognitive functioning in both clinical and nonclinical populations.

The development of modern clinical neuropsychology was fostered by profound advances in theory and empirical knowledge of brain systems and processes, brain–behavior relationships, and the effects of various neuropathologic conditions on everyday functioning. The nature and quality of standardized normative data for valid interpretation of neuropsychological test results have also improved significantly. These advances have occurred in close collaboration with the growth in related fields, such as cognitive neuroscience, behavioral neurology, computer and information science, as well as profound advances in imaging brain structure and function.

COMMON USES OF NEUROPSYCHOLOGICAL EVALUATIONS

Neuropsychological assessment can be invaluable in geriatric mental health care settings; it can aid in the distinction between normal versus pathologic cognitive and functional changes, differential diagnosis among neuropsychiatric and neuropathologic conditions, and determination of long-term functional needs for treatment and care planning.

There have been remarkable advances in clinical brain neuroimaging over the past three decades, which can be a valuable supplement to, but not replace, the findings from neuropsychological assessments. One of the most common concerns with older patients is distinguishing normal age-related changes from early Alzheimer disease or other neurodegenerative conditions. A second frequent presenting concern is determining whether an individual's reports of memory or other cognitive deficits are due to depression or primary neurocognitive dysfunction. Clinical neuroimaging can be helpful in identifying gross changes in brain structure, such as reduction in hippocampal volume. In May 2020, the U.S. FDA approved flortaucipir (AV-1451) to image the presence of tau neurofibrillary tangles in the brain with positron emission tomography (PET) scans. These new and emerging technologies are an important part of neuropsychiatric care and help evaluate structural changes. However, the neuropsychological assessment remains key to the examination of patterns of spared and diminished cognitive functions that are critical to everyday functioning.

As an example of the use of neuropsychological evaluation, consider a patient presenting with a report of "poor memory." The term "memory" has a variety of meanings in colloquial use, such as remembering to perform a task ("prospective memory"), finding misplaced objects (which may reflect attention or organizational skills rather than memory per se), and remembering recent events or newly acquired information ("episodic memory"). Even when the concern is narrowed down to episodic memory, it is important to distinguish between a range of underlying memory-related tasks or processes, including difficulties with initial learning or encoding information or difficulties efficiently retrieving information, and actual loss of the memory trace. Results from standardized memory tests are invaluable in making such distinctions. For instance, if a patient shows poor performance on initial learning or recall trials of a test, but after a delay, they are still able to recall most of whatever amount of information they learned, then their deficit is likely one of learning or encoding rather than forgetting. Comparison of cued recall, performance on a recognition format, or both to free recall permits identification of patients with inefficient retrieval rather than actual forgetting. These types of distinctions are key in the differential diagnostic process. Rapid forgetting, that is, the inability to retain recently learned information, is relatively uncommon with normal aging and even in patients with depression but is quite common among patients with cortical dementias. One would not diagnose a patient based on their performance on an isolated test score (most healthy adults will manifest one or more scores in the impaired range on some measure when given a large test battery), but the presence of rapid forgetting would be an important piece of evidence more consistent with the presence of cortical dementia than with normal aging or depression.

Even when a patient's diagnosis is known, neuropsychological assessment can be valuable in the documentation of the trajectory and rate of cognitive change. For instance, there is considerable variability between patients with neurodegenerative conditions in terms of the rate of cognitive decline. Repeated evaluation with a targeted battery can be useful in determining how fast a particular person is likely to reach a point requiring a more structured or intensive level of care. Conversely, for a patient who has experienced a stroke or other acute brain trauma, repeated neuropsychological testing can be useful in documenting the rate, trajectory, and degree of recovery of cognitive functioning.

Neuropsychological evaluation in geriatric mental health care can also be useful in treatment plans involving the remediation of deficits in everyday functioning. For example, it may be known that a particular patient's "forgetfulness" is causing difficulties in everyday functioning (such as missed appointments, unpaid bills, etc.). In such instances, a key goal of a neuropsychological evaluation is to identify areas of spared cognitive functioning, in addition to documenting the type and degree of any cognitive deficits. With such information, the specific cognitive processes underlying any apparent deficits in independent functioning can be identified, and information about spared abilities can be used to arrange the environment in a way that maximally draws upon the individual's strengths and reduces the potential deleterious functional effects of impaired abilities.

ASSESSMENT DOMAINS AND COMMONLY USED MEASURES

The length and extent of cognitive testing performed by a neuropsychologist frequently vary depending on the nature of the presenting concern and the general context or purpose of the evaluation. Neuropsychological assessments range from screening procedures that can be completed in 30 minutes or less, whereas a more comprehensive evaluation generally requires several hours to complete. The latter type of testing is particularly valuable in identifying cognitive strengths and weaknesses across a range of more specific neurocognitive abilities. Some of the common constructs evaluated as well as examples of tests frequently used to assess these constructs are described below.

Cognitive Screening

The majority of brief cognitive screening measures were designed specifically to stage the severity of Alzheimer disease or other dementias, so they may be insensitive to other functionally relevant forms of cognitive impairment. They also tend to be less helpful in identifying areas of clear cognitive strengths.

One of the most widely used dementia screening measures is the Folstein Mini-Mental State Examination (MMSE). The MMSE has a potential range of 0 to 30, with scores below 24 generally considered “impaired.” Advantages of the MMSE include ease of administration and scoring; it typically takes less than 10 minutes to complete, can be reliably administered and scored by trained clinic staff, and can be tolerated even by very low-functioning patients. In addition, it is so widely used that most clinicians have a tangible sense of how scores at a given range are likely to correspond with actual functioning. MMSE scores can thus serve as a useful means of quickly communicating the severity of a patient’s overall deficits. Disadvantages of the MMSE include its sensitivity to education (highly educated patients with mild Alzheimer disease may perform very well on the MMSE, whereas healthy individuals with low education may earn scores in the “impaired” range) and its relative insensitivity to mild cognitive deficits and deficits associated with noncortical dementias. Of note, the MMSE is a copyrighted test that requires purchase for legal use.

Another brief screening instrument for cognitive performance is the Montreal Cognitive Assessment (MoCA). Similar to the MMSE, the MoCA is brief (less than 10 minutes), easy to administer and score, and yields total scores ranging from 0 to 30, with lower scores representing worse cognitive impairment. It was originally designed to screen for mild cognitive impairment (MCI), rather than cortical dementia *per se*, and thus may be more sensitive to subtle impairments. Another advantage of the MoCA includes availability in multiple languages in addition to English. The test itself is free, but, except for those users in certain exempt categories, the copyright holder now requires users to complete a 1-hour for fee training/certification module (see MoCA website for current details: https://mocacognition.com/).

The Mattis Dementia Rating Scale-Second Edition (DRS-2) is another widely used scale of dementia screening; it generates a total score (maximum possible points = 144), indicating overall cognitive functioning, as well as five subscale scores in the domains of attention, initiation/perseveration, construction, conceptualization, and memory. It takes from 15 to 20 minutes (for a cognitively intact individual) to approximately 25 to 30 minutes (for an impaired individual) to administer. Like the MMSE, it is prone to age and education effects, but available age- and education-corrected norms facilitate the interpretation of DRS-2 scores. The DRS-2 is sensitive to mild dementia, although the ideal “cutoff” score seems to vary slightly among studies; one cross-validated study applied a cutoff of 129 of 144 and yielded a 98% sensitivity and 97% specificity while distinguishing patients with Alzheimer disease from healthy comparison subjects. Overall, DRS-2 scores, as well as subscale scores, have been shown to be good predictors of functional outcomes (basic and instrumental activities of daily living [ADLs and IADLs]) among patients with Alzheimer disease. The cognitive subscale of the Alzheimer’s Disease Assessment Scale (ADAS-Cog) is another widely used screen for cognition among those with Alzheimer disease.

An intermediate option between a brief screening instrument and a lengthy comprehensive neuropsychological battery is to use an instrument such as the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). The RBANS takes approximately 30 minutes to administer and assesses a number of domains of interest. Computerized cognitive screens, such as the MicroCog, were developed to reduce time and administration burden and are available for use.

A screening measure is useful for a quick and cursory evaluation of cognitive functioning but is limited in the extent to which it can assess the breadth and depth of neuropsychological abilities. A comprehensive neuropsychological test battery is the best way to measure more nuanced aspects of cognition across multiple domains of functioning.

Crystallized Knowledge

Crystallized knowledge refers to information and facts such as those typically learned in formal school settings. Tests that measure crystallized knowledge include the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) Vocabulary subtest, in which the examinee is asked to define a common word, and the WAIS-IV Information subtest, in which the examinee is asked to answer “common knowledge” questions, and the Wide Range Achievement Test-Fifth Edition Reading subtest that measures one’s ability to recognize or read isolated words. Crystallized knowledge tends to be resistant to most forms of neurocognitive injury and may be relatively stable (or even improve) over the life span; word reading and vocabulary scores are, therefore, often used, together with educational history, to estimate levels of premorbid cognitive functioning.

Receptive and Expressive Language

Deficits in language functions may reflect specific or localized lesions (such as a stroke in the left frontal lobe) or diffuse neuropathology such as that occurs with Alzheimer disease. Common receptive language measures include such tasks as reading comprehension or comprehension of increasingly complex aurally presented instructions, such as measured by the Token Test or the Complex Ideational subtest from the Boston Diagnostic Aphasia Examination. The Boston Naming Test is the most widely used measure of expressive language functions in assessments of adults and is particularly useful with patients reporting word-finding difficulties. The task involves presenting the examinee with line drawings of objects one at a time and asking them to identify each one in succession. If the examinee cannot identify the item within 20 seconds, then they are given a cue to see if that aids in responding. Confrontational naming is often one of the first abilities to deteriorate in patients with cortical dementia, such as in early Alzheimer disease. Patients with subcortical dementias may have difficulty retrieving the names of the objects, but they tend to do better with cueing.

Another widely used set of tasks that are sometimes categorized under expressive language is the letter and category fluency (word generation) tasks. For instance, in the “FAS” test, the examinee is given three 1-minute trials to name as many words as they can think of that begin with the letters “F,” “A,” and then “S,” respectively. Category fluency involves having the examinee rapidly name words from a given category, such as types of animals. Because word knowledge tends to be organized by meaning (semantics) rather than by the initial letters, healthy individuals generally find the category fluency task considerably easier than the letter fluency task. However, with some neurodegenerative conditions, such as Alzheimer disease, the semantic links begin to deteriorate, so the typical advantage of category fluency over letter fluency performance decreases. This is another example of how considering the pattern of performance across tasks can be useful in the differential diagnostic process.

Attention and Vigilance

The construct of attention comprises a range of specific functions that include consciousness (attentional, arousal, and temporal orientation), simple attention (the ability to repeat recently presented stimuli, most commonly, strings of digits), selective attention (distinguishing target stimuli or information from irrelevant stimuli), sustained attention (ability to stay mentally "on track"), and vigilance (the ability to wait/look out for a particular stimulus or event over time). Some aspects of working memory, best described as a "mental workbench," also fall under the larger domain of attention (with the more complex aspects constituting executive functions). This function of working memory allows for a limited amount of information to be actively held for short periods of time for further processing and requires active rehearsal or allocation of attentional resources for it to be retained; a classic example is the holding of a telephone number in one's working memory so long as is needed to dial the telephone, and the number may be retained via active rehearsal over the delay period.

Attentional functions are among the first to be affected by a variety of neurologic and psychiatric conditions. Attentional decline may therefore be readily observed in older clinical populations. Alterations in consciousness are common in older adult patients with delirium, for example, which in turn may be caused by multiple etiologies (e.g., substance use, prescription medication, urinary tract infections). Other aspects of attention may be moderately to severely impaired in patients with dementia, especially subcortical dementias such as those associated with Parkinson disease. Dementia with Lewy bodies is also associated with impairments in both consciousness and arousal and simple and sustained attention. Mild to moderate impairments in attention are present among patients with Alzheimer disease, but they tend to be able to perform basic attention tasks fairly well even in more advanced stages of the disease. Fluctuating attention, especially within a single attention-oriented task, is often associated with geriatric depression or anxiety.

Orientation and simple attentional tasks are included in the MMSE, MoCA, and DRS-2. The MMSE and MoCA include the serial sevens subtraction task, while the DRS-2 includes some cancellation tasks.

A common measure of simple attention and working memory is the WAIS-IV Digit Span subtest, in which the examinee is asked to repeat strings of digits of increasing length and is then asked to repeat strings of digits in reverse order. Cancellation tasks are common measures of sustained attention or vigilance; for example, in the Digit Vigilance Test from the expanded Halstead–Reitan Battery (HRB), the examinee is asked to rapidly identify and cross out every time a target number appears in a much larger array of numbers. The various versions of the Continuous Performance Test (CPT), such as the “identical pairs” version, in which the examinee is shown a series of numbers rapidly flashed on a computer screen and is required to make a response after any number is shown twice in a row, is also an example of such a task.

Learning and Memory

Memory complaints are prevalent among older individuals, and there is consistent evidence that some aspects of memory, specifically episodic declarative memory, deteriorate with advanced age. The function of an objective memory evaluation is to determine whether memory decline is beyond age-appropriate memory decline. Therefore, normative data on standardized memory tests become even more important to interpret the target person’s test results accurately. Also, given that memory components of episodic memory tests are typically preceded by learning new material (e.g., word lists, stories, geometric designs), it is also important to rule out the possibility of a learning disorder (which may have remained previously undiagnosed or undetected among older patients). The presence of a learning disorder is determined by performances on achievement tests (e.g., reading, spelling, simple arithmetic) and the individual’s educational or occupational history (repetition of a grade, difficulties in school, special education programs, extra tutoring, and behavioral problems in school being common red flags). If a learning disorder exists, then memory test results must be interpreted with caution, and education-corrected norms will give skewed results.

Referral questions for neuropsychological evaluation among geriatric patients typically include a request for “memory assessment,” often because it is the primary cognitive complaint of patients, caregivers, or both. There are several reasons why older individuals may experience impairment in episodic learning and memory, such as dementia, depression, physical conditions, such as pain, and certain medication. Memory performance can also be hindered by primary deficits in other cognitive domains, such as attention or executive functioning. Procedural memory (e.g., driving a car or making dinner) tends to fade at a slower rate and may only be seen in moderate to severe stages of dementia. Finally, autobiographical memory tends to remain intact for the most part, except in patients with very advanced stages of dementia.

Neuropsychological evaluation typically includes measures of simple registration (e.g., asking the patient to repeat three or five words, as in the MMSE and MoCA, respectively) and both auditory and visual episodic learning and memory. Tests of episodic learning and memory typically begin with a series of learning trials followed by short- and long-delay recall trials. The time between the learning and delay trials is filled with unrelated tasks (typically, another neuropsychological test or “distraction stimuli”) so that the patient cannot rehearse the test stimuli. Delayed recall trials may include free recall, cued recall, recognition paradigms, or a combination of these. Cued recall and recognition trials help identify retrieval versus encoding deficits; in the former, the information is encoded but can only be retrieved with help (e.g., cues or recognition), and in the latter, the information is never encoded in the first place (so neither cues nor recognition will help).

The most commonly used tests of auditory learning and memory include the Hopkins Verbal Learning Test-Revised, Rey Auditory Verbal Learning Test, and the California Verbal Learning Test-Third Edition (CVLT-3), all of which are list-learning tasks, and Logical Memory (from the Wechsler Memory Scale-Fourth Edition [WMS-IV]) or the Story Memory Test in which the stimuli are stories or short passages. Visual learning is often measured with the Brief Visuospatial Memory Test-Revised, Visual Reproductions (from the

WMS-IV), and the Rey–Osterrieth Complex Figure Test. These tests require the examinee to learn and remember geometric figures by copying them or, via immediate recall, recalling them by drawing as many figures as possible from memory. These tests, therefore, also have a visuoconstructional element in addition to a visual memory component. The Visual Reproduction test also has recognition components.

Two classic distinctions made between memory test score profiles among older individuals are those of cortical versus subcortical dementia and dementia versus late-life depression. Cortical profiles of word list-learning tasks seen most often among patients with Alzheimer disease include a flat learning slope across the learning trials, rapid forgetting (i.e., little to no words in short- and long-delayed recall), high recency effect (i.e., most words recalled from the last third of the list), high number of intrusions (i.e., words not on the list, similar to confabulations seen in informal assessments of patients with Alzheimer disease), high number of repetitions, and no benefit from cueing or recognition formats. Patients with subcortical dementias, such as Parkinson disease, also have flat learning slopes and difficulty recalling any words during free recall trials but are unlikely to have intrusions or repetitions and do better with cues or on the recognition trial. The performance of patients with depression is marked by overall slowness in responding and lower output of words during free recall. They will almost never have intrusions or repetitions, and they tend to recall words from all over the list. They tend to do significantly better with cueing and recognition, often achieving scores within the normal range for their age.

Visuospatial Cognition

The assessment of visuospatial cognition or perceptual organizational skills is an important component of both screening and comprehensive neuropsychological evaluations. Broadly, these functions consist of perceptual ability (the ability to discern forms or patterns), visuospatial judgment (the location of objects in space), and constructional ability (the manipulation of objects in three-dimensional space). Intact visual acuity is an obvious precondition to all subdomains of visuospatial cognition, and adequate use of hands is necessary for the constructional tasks. Visual field defects and hemiparesis can also cause impairments in visuospatial cognition. Some of the most common neurologic conditions affecting visuospatial cognition among older adults are vascular incidents (lesions causing contralateral hemispatial neglect), Balint syndrome (associated with bilateral lesions in the posterior parietal cortex), and dementias caused by Lewy bodies, Parkinson disease, and Huntington disease. Patients with Alzheimer disease also have impairments in visual cognition, but they are less severe compared with those among individuals with other dementias. Incipient Alzheimer disease is associated with difficulties with constructional ability, but other visuospatial abilities may only get severely impaired in the advanced stages of the disease.

Visuospatial abilities are commonly measured with subtests of the WAIS-IV such as Picture Completion (which is one of the WAIS-IV supplemental subtests), in which the examinee must identify the important missing element in a series of pictured scenes, and Matrix Reasoning, a visual reasoning task wherein the examinee must identify the missing element in a matrix by discerning the rule underlying the relationship between each of the several subparts. The Hooper Visual Organization Test is another widely used measure, wherein the examinee is shown pictures of objects that have been cut and randomly arranged in pieces; they must mentally rotate the various pieces and put them together, and then name the object. Constructional skills are measured by graphomotor tasks, such as the Rey–Osterrieth Complex Figure Test, in which the examinee is presented with a complex geometric figure, which they must copy (the order in which the figure is copied is noted by the examiner), or, most commonly, the Block Design subtest (from the WAIS-IV), in which the examinee is required to rapidly match a two-dimensional design by combining three-dimensional red-and-white blocks.

Executive Functions and Working Memory

There is wide consensus among neuropsychologists that intact executive functions are central to optimal daily functioning, yet, the precise definition of the construct remains elusive. The term “executive functions” is used to describe an array of subdomains, including abstraction and concept formation, planning, response inhibition, cognitive flexibility, and some aspects of working memory. The products of all these executive functions typically involve novel problem solving, initiating and maintaining goal-directed behavior, or disengaging from such behaviors when appropriate to the goals or in response to environmental cues. Executive functions were also previously called “frontal cortical functions” because they were thought to be mediated by the frontal cortex. However, there is now overwhelming evidence that these functions are not limited to the frontal cortex but that they involve several frontal–subcortical circuits as well. Therefore, lesions in frontal as well as subcortical brain structures can disrupt any of the executive functions.

People with Alzheimer disease often show signs of executive dysfunction even in the early stages of the disease. People with frontotemporal dementia and those with any of the several forms of subcortical dementias (such as those due to Parkinson disease) have significant impairments in cognitive flexibility, abstract thinking and conceptualization, and fluency. Executive functions may also be more sensitive to low or inconsistent effort; therefore, patients with depression may do poorly on some or all executive functioning measures.

The Wisconsin Card Sorting Test (WCST) is one of the most commonly used measures of executive functioning, specifically, concept formation and cognitive flexibility. The task involves asking the examinee to discern an unspoken rule that defines correct versus incorrect matches and requires the examinee to change strategies when appropriate. The Delis–Kaplan Executive Function System (D-KEFS) is a battery of nine executive functioning tests and has normative data for adults up to 89 years. The D-KEFS includes the Trail Making Test, Verbal Fluency Test, Design Fluency Test, Sorting, Tower Test, Color–Word Interference, the Word–Context Test, the Proverb Test, and Twenty Questions. These tests, some of them modified from popular, existing tests (e.g., the Trail Making Test is modified from the Trail Making Test, Parts A and B, a.k.a., the Partington Pathways Test), measure a broad range of executive functions and are useful to include in a comprehensive battery for use with older adults.

Mental Processing Speed/Psychomotor Speed

The term “mental processing speed” (or information processing speed) refers to the speed with which one thinks and executes a task. As the tasks used to assess this construct generally include a motor response component, it is also commonly labeled “psychomotor speed,” reflecting the fact that test performance is determined not only by the speed with which one thinks but also by the integration of those thoughts with the motor response. Whether construed as psychomotor speed or pure mental processing speed, tests in this domain tend to be among the most sensitive to virtually any form of neurocognitive injury or dysfunction, including that caused by psychiatric conditions such as depression. Indeed, the steady, if gradual, stowing of processing and psychomotor speed is the most ubiquitous and consistent change seen with normal aging, beginning at least by the early part of the fourth decade of life. When examining an older adult patient for the possible presence of pathologic motor disorders, it is important to consider their psychomotor speed relative to that expected from an age-comparable normative sample.

The Digit Symbol (a.k.a., “Coding”) test is a prototypical example of processing/psychomotor speed tests and is among the most widely used tests to assess this domain. The Digit Symbol test consists of a stimulus key with the numbers 1 through 9, each paired with a unique geometric figure. Beneath this stimulus key is a large matrix of numbers with spaces below each number. The patient’s task is to rapidly fill in the symbol paired with each number. The stimuli used for this task appeared in largely unaltered form since the early 1900s. That is, the stimuli used in the Army Mental Test battery during World War I were almost identical to those used in the adult versions of Wechsler intelligence test batteries, up through the 1997 edition of the Wechsler Adult Intelligence Scale (WAIS-III). The Digit Symbol subtest stimuli were substantially redesigned for the WAIS-IV, published in 2008, but the sensitivity of the task to age-related or other psychomotor slowing remains the same.

Another widely used measure of psychomotor speed is the Trail Making Test Part A (originally published as Partington Pathways Test in the 1930s), which requires the patient to connect a series of circled numbers in order as fast as possible, and measures scanning, visuospatial sequencing, and motor speed. The D-KEFS includes a number of interesting modifications of the Trail Making Test, including a Motor Speed subtest that may prove helpful in distinguishing pure motor slowing from slow mental processing.

OTHER ASPECTS OF COMPREHENSIVE NEUROPSYCHOLOGICAL EVALUATION

Assessment of Mood and Anxiety

Emotional functioning is extremely important to monitor as part of any assessment with older adults. Depression is often undetected among older adults, although increasing awareness of the frequency of psychiatric problems in this population has led to more routine assessments of depression, anxiety, and suicidality in psychiatric, neurologic, and general medical treatment settings. Some useful and quick measures of depression are the Geriatric Depression Scale, Beck Depression Inventory-Second Edition (BDI-II), and Patient Health Questionnaire-Ninth Edition (PHQ-9). The former is particularly helpful for use with older patients and is validated for use with patients with Alzheimer disease. Occasionally, if the patient has been able to tolerate a lengthy neuropsychological battery and may have potential emotional problems, they may be given the Minnesota Multiphasic Personality Inventory-Third Edition (MMPI-3). The MMPI-3 can give a detailed profile of the patient’s emotional functioning as well as further highlight targets for intervention.

Activities of Daily Living/Everyday Functioning

It has been common practice in neurology and psychiatry to informally assess older patients' basic ADLs and IADLs. Basic ADLs include personal functions such as feeding oneself, showering, dressing, and using the toilet, among other things. Intake interviews as part of neuropsychological assessments tend to focus more on evaluating IADLs, such as cooking, driving, managing finances, and so forth (i.e., skills that are more complex than basic ADLs yet necessary for independent everyday functioning).

Diversity

One of the challenges in neuropsychological assessment of older adults has been and remains the fact that the majority of people in the normative samples were non-Latine White, heteronormative, and native English–speaking persons. There have been efforts to develop normative data for African Americans and Latine individuals, and there have been more recent attention to the effects of culture and cultural bias on the validity of standard neuropsychological tests. Yet, much work remains to be done in ensuring culturally appropriate and validated tests and test interpretation for individuals from diverse groups. Even the assumption that a particular test developed in one cultural context measures the same cognitive construct across other cultures needs to be considered with caution. A key distinction is between differences in test performance versus differences in the level of cognitive function. There are known effects of not just years of education but the quality of education that affect test performance without clearly representing differences in the level of cognitive functioning. Caution is also needed when testing a person in English when their first language was something other than English. There is some research suggesting that bilingualism confers some advantages in the resilience of some neurocognitive domains, but it may affect performance (not skill) on verbally focused tasks administered in English. For the purposes of readers of this overview, the key is to stay aware of the importance of considering and addressing diversity issues in the interpretation and conclusions drawn from neuropsychologist test scores.

Telehealth Assessment

One of the byproducts of the social distancing requirements during the COVID-19 pandemic has been increased interest and reliance on remote administration of neuropsychological tests, such as via telehealth. Although there are active efforts to develop such tests, and some standard tests by necessity have been adapted for virtual administration, interpretation of scores under these conditions requires caution. At present, most normative data for neuropsychological tests were developed with 1:1, in-person assessments. There are small-scale studies that provide some preliminary evidence that equivalent performance can be obtained, but the generalizability of that normative data to a telehealth format remains less certain. Standardization on the receiving end is also complicated by the lack of uniformity in the size and nature of the device in the examinees' possession (smartphone, tablet, desktop, and monitor size) and other practical concerns that are closely related to socioeconomic positions, such as accessibility to high-speed internet and its reliability of functioning. An often-overlooked obstacle to valid assessment is the lack of privacy in home settings, where the examinee may be distracted by ambient noise and other family members and pets that can further confound results. With older adults, additional concerns arise from less comfort or familiarity with the requisite technology and visual and hearing impairments that can complicate the efficient use of remote technology. Thus, any conclusions derived from neuropsychological assessments currently administered via telehealth must be provisional and accompanied by the documentation of any of the aforementioned confounding variables.

ISSUES IN THE ASSESSMENT OF OLDER ADULTS

Sensory Impairment

Hearing impairments tend to be fairly common among older adults in general and can considerably impede verbally administered neuropsychological tests. The potential examinee's hearing should be at least informally assessed to determine whether it is adequate for the verbal neuropsychological tests. If hearing is deemed too impaired, they can still be assessed with largely visual tests (assuming they have adequate visual acuity) if it is clear that the person understands the test instructions. Visual acuity may be informally assessed, for example, using the Rosenbaum Pocket Screener. Visually impaired patients can be assessed with a variety of tests, including all verbal memory tests, some attention tests, all verbal D-KEFS tests, and the Tactual Performance Test, which is a measure of visuospatial functioning.

Interpretation of Test Results in the Context of Normal Aging

Normal aging is itself associated with some changes in cognitive functioning, including improvements (at least during the first seven decades of life) in crystallized verbal knowledge, slowing of mental processing speed, psychomotor speed, and subtle declines in some aspects of memory and executive functions. A key advantage of standardized testing is that interpretation is not solely reliant on the absolute level or content of responses (or pathognomonic signs) but also on reference to the level of performance expected from a neurologically healthy individual. The latter is determined by comparing the examinee's test scores to demographically appropriate norms. The most widely used neuropsychological tests have normative data developed and collected using standardized administration and scoring procedures. That being noted, the above-expressed caveats concerning the application of normative data to diverse cultural groups remain a concern.

FURTHER READINGS

Bilder RM, Postal KS, Barisa M, et al. InterOrganizational practice committee recommendations/guidance for teleneuropsychology (TeleNP) in response to the COVID-19 pandemic. *Clin Neuropsychol*. 2020;34(7–8):1314–1334.

Boake C. From the Binet-Simon to the Wechsler-Bellevue: tracing the history of intelligence testing. *J Clin Exp Neuropsychol*. 2002;24(3):383–405.

Celik S, Kokje E, Meyer P, Frölich L, Teichmann B. Does bilingualism influence neuropsychological test performance in older adults? A systematic review. *Appl Neuropsychol Adult*. 2022;29(4):855–873.

Díaz-Orueta U, Blanco-Campal A, Burke T. Rapid review of cognitive screening instruments in MCI: proposal for a process-based approach modification of overlapping tasks in select widely used instruments. *Int Psychogeriatr*. 2018;30(5):663–672.

Fujii D. *Conducting a Culturally Informed Neuropsychological Evaluation*. American Psychological Association; 2016.

Geddes MR, O'Connell ME, Fisk JD, Gauthier S, Camicioli R, Ismail Z; Alzheimer Society of Canada Task Force on Dementia Care Best Practices for COVID-19. Remote cognitive and behavioral assessment: Report of the Alzheimer Society of Canada Task Force on dementia care best practices for COVID-19. *Alzheimers Dement (Amst)*. 2020;12(1):e12111.

Grant I, Adams KM. *Neuropsychological Assessment of Neuropsychiatric and Neuromedical Disorders*. Oxford University Press; 2009.

Kochhann R, Wilson MA. Introduction: Special issue on neuropsychology of aging. *Psychology & Neuroscience*. 2019;12(2):141–143.

Kolb B, Whishaw IQ. *Fundamentals of Human Neuropsychology*. 8th ed. Worth Publishers; 2021.

Lanca M. Integration of neuropsychology in primary care. *Arch Clin Neuropsychol*. 2018;33(3):269–279.

Lezak MD, Howieson DB, Bigler ED, Tranel D. *Neuropsychological Assessment*. 5th ed. Oxford University Press; 2012.

McPherson S, Koltai D. *A Practical Guide to Geriatric Neuropsychology*. Oxford University Press; 2018.

Morris RG, Kopelman MD. The neuropsychological assessment of dementia. In: Crawford JR, Parker DM, McKinlay WW, eds. *A Handbook of Neuropsychological Assessment*. Routledge; 2018:295–322.

Palmer BW. The expanding role of neuropsychology in geriatric psychiatry. *Am J Geriatr Psychiatry*. 2004;12(4):338–341.

Palmer BW, Dawes SE. Cognitive aging: from basic skills to scripts and schemata. In: Jeste DV, Depp CA, eds. *Successful Cognitive and Emotional Aging*. American Psychiatric Publishing, Inc; 2010:37–54.

Pavol MA. Inpatient neuropsychological assessment in older adults. In: Ravdin LD, Katzen HL, eds. *Handbook on the Neuropsychology of Aging and Dementia*. 2nd ed. Springer International Publishing; 2019:89–103.

Pinto TCC, Machado L, Bulgacov TM, et al. Is the Montreal Cognitive Assessment (MoCA) screening superior to the Mini-Mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's Disease (AD) in the elderly? *Int Psychogeriatr*. 2019;31(4):491–504.

Ravdin LD, Katzen HL. *Handbook on the Neuropsychology of Aging and Dementia*. Springer International Publishing; 2019.

Salmon DP, Bondi MW. Neuropsychological assessment of dementia. *Annu Rev Psychol*. 2009;60:257–282.

Salthouse TA. Trajectories of normal cognitive aging. *Psychol Aging*. 2019;34(1):17–24.

Savla GN, Moore DJ, Palmer BW. Neuropsychological assessment. In: Birren JE, ed. *Encyclopedia of Gerontology: Age, Aging, and the Aged.* 2nd ed. Elsevier Academic Press Ltd; 2007:99–107.

Savla GN, Twamley EW, Thompson WK, Delis DC, Jeste DV, Palmer BW. Evaluation of specific executive functioning skills and the processes underlying executive control in schizophrenia. *J Int Neuropsychol Soc*. 2011;17(1):14–23.

Sherman EMS, Tan JE, Hrabok M. *A Compendium of Neuropsychological Tests: Administration, Tests, Norms, and Commentary*. 4th ed. Oxford University Press; 2020.

Sohlberg MKM, Mateer CA. *Cognitive Rehabilitation: An Integrative Neuropsychological Approach*. Guilford; 2001.

Tuokko HA, Smart CM. *Neuropsychology of Cognitive Decline: A Developmental Approach to Assessment and Intervention*. Guilford; 2018.

55.2f Genetics of Late-Life Neurodegenerative Disorders

SEETHALAKSHMI RAMANATHAN, M.D., M.P.A., AND STEPHEN J. GLATT, PH.D.

Genetic studies of any complex condition, including late-life neurodegenerative disorders, attempt to identify the elemental biologic underpinnings of these disorders to facilitate their early identification, early intervention, and, ultimately, prevention. In addition, the identification of genes that influence the development and/or progression of these illnesses may foster medicinal chemistry applications that target faulty DNA sequences or the altered RNA transcripts or proteins produced by such genes. Such discoveries could usher in the long-awaited era of “personalized or precision medicine,” in which the treatment of any given late-life neurodegenerative disorder can be tailored to individual patients based on their genetic and biologic background.

The search for genes influencing late-life neurodegenerative disorders falls under the broader umbrella of genetic epidemiology, which has been defined as a science that deals with the etiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations. Genetic epidemiologists examine the transmission of illness within families, or the distribution of disease in populations, with the goal of finding genetic and environmental causes of illness. Genetic epidemiologic research on a particular disorder such as Alzheimer disease (AD) tends to follow a series of questions in a logical progression. This sequence proceeds as follows: First, we ask, “Is the phenotype familial?” or “Does it run in families?” which can be established by family studies. Second, “What is the relative magnitude of genetic and environmental contributions to the phenotype?” which can be established by twin studies. This, importantly, will reveal the maximum amount of variance in a disease that may be attributable to heritable factors (including genes themselves, but also gene–environment interactions). Third, “How is the phenotype transmitted from generation to generation?” can be addressed by segregation analyses. Fourth, “If genes mediate this transmission, where are they located?” is a question that may be answered by linkage studies of families or extended pedigrees. Fifth and finally, with a heritable component established, we ask, “What are the specific genes and variants in those genes that influence risk for the phenotype?” which can be addressed by association studies, including genome-wide association studies (GWASs). These are difficult questions to answer for any trait, but particularly so for phenotypes as complex as late-life neurodegenerative disorders, which can often be confused or misconstrued with other medical, psychiatric, or substance use disorders. Fortunately, the wide variety, specificity, and power of available methods have allowed genetic researchers to begin to find meaningful answers to some of these questions for each late-life neurodegenerative disorder.

GENETIC METHODS AND PRINCIPLES

Family Studies

The first question that must be asked and answered when attempting to delineate the genetic and environmental components of a complex late-life neurodegenerative disorder is, “Does the phenotype run in families?” or “Is this phenotype familial?” This question can be answered through the use of family studies. The basic design of the family study begins with the ascertainment of a group of subjects who are affected by the disorder (cases) and a comparable group of comparison subjects who do not have the disorder. Next, the biologic relatives of these index subjects, or “probands,” are ascertained and evaluated for the presence of the illness or, in some cases, subthreshold forms of the illness. The rate of the disorder or associated “spectrum conditions” among family members of affected probands is then compared to the rate of the disorder among family members of comparison probands to determine the familial risk or relative risk. If a late-life neurodegenerative disorder has a genetic etiology, then biologic relatives of cases should have a higher likelihood than relatives of comparison subjects of carrying the gene or genes that influenced illness in their relative, and thus they should be at greater risk for the illness themselves. In addition, the risk to relatives of cases should be correlated with their degree of relationship to the proband or the number of genes they share in common. First-degree relatives, such as parents, siblings, and children, share 50% of their genes, on average, with the proband. Thus, first-degree relatives of cases should be at greater risk for the disorder than second-degree relatives (grandparents, uncles, aunts, nephews, nieces, and half-siblings), who share only 25% of their genes with the proband.

Twin Studies

Once a disorder has been established as familial, it becomes necessary to determine if that pattern is attributable to the inheritance of genes or to shared familial and other environmental factors. It is also important to quantify the contribution that genes make relative to that made by environmental factors, as this may encourage or discourage future molecular genetic studies; this may also influence the decisions made by individuals seeking genetic counseling and impact the usefulness of family information in identifying high-risk individuals to target for early intervention and prevention protocols. These questions can be answered by twin studies. In twin-study designs, the ratio of concordant:discordant monozygotic (MZ), or identical, twin pairs is compared to the ratio of concordant:discordant dizygotic (DZ), or fraternal, twin pairs. If the concordance for a late-life neurodegenerative disorder is higher among MZ twin pairs than DZ twin pairs, this is a good indication that there is a genetic contribution to the disorder; if MZ and DZ twin pairs have approximately equal concordance rates, environmental factors are more strongly implicated. Concordance rates among twin pairs are then used to estimate the "heritability" of a disorder. Heritability refers to the ratio of genetic to phenotypic variances or the proportion of variance in risk for a late-life neurodegenerative disorder that is accounted for by variability in genetic factors. A heritability of 1.0 (or 100%) indicates that all variability in the phenotype is due to genetic factors alone. In contrast, a heritability of zero attributes all phenotypic variation to environmental factors.

Segregation and Linkage Studies

Knowing that genetic factors are involved in the etiology of a late-life neurodegenerative disorder—and to what degree—is essential for designing optimal molecular genetic studies to reveal the chromosomal location of the responsible or contributory genes. To determine the mode of inheritance of disease, segregation analysis is used to test the goodness of fit of various genetic models onto observed family and pedigree data. To identify regions of chromosomes that have a high likelihood of harboring risk genes for late-life neurodegenerative disorders, linkage analysis is an appropriate strategy under certain conditions, that is, when the disease segregates strongly through families, and evidence suggests a Mendelian mechanism of transmission. Families are ascertained for linkage analysis through a proband affected with the disorder of interest; then, each individual in the family is genotyped at a series of DNA markers (not necessarily in genes) spaced evenly throughout the genome, and the co-segregation of these DNA markers with the illness is tracked in each pedigree. Evidence for co-segregation at each marker locus is summed across pedigrees to derive an index of the likelihood of the obtained patterns of marker-phenotype co-segregation, given the sampled pedigree structures. This approach has been successful in homing in on causal or major genes for some late-life neurodegenerative disorders. Unfortunately, most forms of late-life neurodegenerative disease are more complex in their etiology, involving the combined and interactive effects of numerous genes and environmental factors, and in such a scenario, linkage analysis is less effective in elucidating contributory genes.

Association Studies

Once regions of certain chromosomes have been implicated from linkage analysis as harboring a risk gene for a disorder (or, increasingly, even in the absence of such linkage information), the next step is to identify what specific gene is segregating through families or populations to give rise to the observed pattern of disease. A gene can be selected for such an analysis subsequent to linkage analysis as a means to follow up on evidence for increased genetic similarity at a locus among affected individuals in a family (i.e., a “positional candidate gene” approach). Alternatively, specific genes can be examined in the absence of linkage information if there is some compelling physiologic reason to suspect that the gene influences risk for a given disorder (i.e., a “functional candidate gene” approach). For example, dopamine-system genes, such as receptors and transporters, are commonly examined as functional candidates for Parkinson disease (PD) based on the neural structures involved in its pathology and treatment. In contrast to linkage analysis, which uses random DNA markers as proxies for nearby risk genes, genetic association analysis is the leading method for determining if a particular gene variant has a direct effect on risk for an illness or is in very close physical proximity to and tight linkage disequilibrium (LD) with such a gene.

The familiar Mendelian model of single-gene transmission of a trait is only one of the many transmission mechanisms that can be tested in genetic association models. Mendelian diseases have recognizable inheritance patterns (autosomal dominant, autosomal recessive, and X linked) and generally are caused by mutations in a single gene within a particular family. In contrast to Mendelian diseases, complex diseases display no clear inheritance pattern, and most forms of late-life neurodegenerative disease fall into this category (rare Mendelian forms do exist, but these typically have early onset). Complex diseases are multifactorial in their etiology and are caused by an interacting web of genetic and environmental risk factors. The genetic risk factors for such complex diseases are generally referred to as *risk* or *susceptibility genes*, in which genetic variants that are common in the population act to increase the risk of developing the disease. Unlike mutations in Mendelian genes, susceptibility genes are neither necessary nor sufficient to cause disease.

Genome-Wide Association Studies

The completion of the human genome project and advancements in the genotyping of single nucleotide polymorphisms (SNPs) laid the foundation for a more broad examination of the human genome *via* the GWAS approach. GWAS, as the name suggests, examines the entire genome for genetic associations with specific traits. This broad approach accords GWAS more flexibility in exploring complex diseases with multifactorial polygenic etiologies, including late-onset Alzheimer disease (LOAD). GWAS techniques examine numerous markers distributed throughout the genome in order to understand the contribution to disease of various types of SNPs, including LD-based markers that are highly correlated with the causal allele, markers for missense mutations associated with Mendelian diseases, and convenience-based markers that are chosen based on logistical conveniences such as cost, heterozygosity rate, and genotyping success rate.

Clearly, such a broad examination has advantages as well as disadvantages. One advantage that GWAS has over other methods is its unbiased nature; however, the examination of the entire genome requires large samples to establish relationships with adequate power due to the inflation of type-I error rate that can occur when performing millions of statistical tests. Another unfortunate constraint of most current GWAS methods is that, unlike candidate genes which are often typed on functional polymorphisms, the DNA markers included in GWASs are often intergenic or otherwise do not constitute any recognized functional motifs; thus, the biologic or clinical significance of a given disease-associated variant may not be immediately recognizable. Researchers have been following and attempting to replicate the extensive list of SNPs identified by GWASs and are now focusing not only on the most strongly associated individual SNPs from meta-analyses of like efforts but on the collective power of large numbers of contributory SNPs in a so-called genetic risk score. Furthermore, the functional outcomes associated with reliably implicated loci and pathways are objects of intensive study.

MAJOR OR MILD NEUROCOGNITIVE DISORDER DUE TO ALZHEIMER’S DISEASE

AD is a progressive neurodegenerative disorder characterized by an insidious onset and progressive loss of memory and other cognitive functions. AD is the most common form of dementia in older adults. Two broad subtypes of AD have been identified: an early-onset subtype, defined as occurring before age 65 years, which accounts for no more than 5% of cases, and a late-onset subtype, defined as occurring on or after age 65 years, which accounts for 95% or more of cases.

Early-onset AD has been characterized as a familial, autosomal-dominant genetic disease with causative mutations in several genes. Early linkage studies of early-onset AD revealed linkage to a locus on chromosome 21q, which was later identified as the gene for amyloid precursor protein (*APP*) which encodes the beta-amyloid protein found in amyloid plaques. However, since linkage to chromosome 21 was not found in most families with early-onset AD, other potential loci were investigated. This led to the discovery of other gene mutations in presenilin 1 (*PSEN1*) on chromosome 14 and in presenilin 2 (*PSEN2*), based on homology to *PSEN1*, on chromosome 1. The net effect of the mutations in these genes is overproduction of the more pathogenic form of beta-amyloid; that is, more beta-amyloid 42 relative to beta-amyloid 40.

The less complicated genetic nature of early-onset AD paved the way for a greater understanding of the genetically complex LOAD, which is sporadic or familial with no clear Mendelian segregation patterns in families. AD has been estimated to be 60% to 80% heritable, with residual influence from nonshared environmental factors. *APOE* is the gene that has been the most reproducibly and consistently identified as a risk factor for the disease, with just a single copy of the *APOE ε4* allele (actually a haplotype of alleles at two physically close loci in the same gene) increasing the risk of AD by approximately 300%. Recent studies have also identified that *APOE* affects the age of developing dementia in a gene-dosage–dependent manner. On the other hand, the *APOE ε2* allele has been shown to reduce the risk for AD relative to the ε*3* allele. *APOE ε4*, however, does not account fully for the genetic variance in AD.

No less than 40 candidate genes have now been implicated by GWAS as reliable risk factors for LOAD; however, the effects of each variant are minimal, and there is no genetic polymorphism that is either necessary or sufficient to cause LOAD. Associated genes are strongly expressed in immune-related tissues and cell types (spleen, liver, and microglia). Gene-set analyses indicate biologic mechanisms involved in lipid-related processes and degradation of

APP. Pathway analysis implicates immunity, lipid metabolism, tau binding proteins, and APP metabolism, showing that genetic variants affecting APP and Aβ processing are associated not only with early-onset autosomal dominant AD but also with LOAD.

Beyond *APOE* ε4 and the several dozen identified genome-wide–significant common SNPs, there may be thousands of additional genetic polymorphisms that make minor nonessential contributions to the overall risk for LOAD. A polygenic risk score (PRS) can be quickly and easily derived from these myriad variants to aggregate their collective effects on risk. The scoring algorithm sums the weighted effect of numerous SNPs to identify a single "genetic load" variable that maximally discriminates cases from unaffected comparison subjects. The best PRS thus far assembled captures most of the SNP heritability for LOAD. Extensive research has indicated that the PRS can boost the accuracy in LOAD diagnosis beyond the performance of *APOE* alone and can capture LOAD phenotypic variability that cannot be explained by *APOE* status alone. Thus, from a diagnostic point of view, having access to a patient's genotypes at *APOE* and at all other reliable risk genes provides the most complete picture of the genetic architecture underlying LOAD.

MAJOR OR MILD NEUROCOGNITIVE DISORDER DUE TO PARKINSON DISEASE

PD is a neurodegenerative disease characterized clinically by a progressively worsening movement disorder and dementia in 20% to 30% of cases late in the course of the disease. Similar to AD, PD largely follows a complex pattern of inheritance with only 5% to 10% aggregating in families and showing an autosomal-dominant or -recessive inheritance. It is, therefore, not surprising that early family studies failed to find much support for familial aggregation of the disorder, causing many to suspect that the etiology of PD was largely (if not exclusively) driven by environmental factors. However, later studies ultimately supported a genetic contribution to the etiology of the disorder. The largest family study has been the Mayo Clinic study, which established that biologic relatives of a patient with PD are at an approximately 70% elevated risk of themselves developing the disorder as compared to relatives of unaffected individuals. It appears that genes may more strongly contribute to the risk for early-onset PD (less than 66 years of age). Relatives of such early-onset PD patients are at a 2.6-fold increased risk of developing PD as compared to relatives of patients with onset of PD later than 66 years, who were not at an increased risk for the disorder. Segregation analyses further suggest that early-onset PD may be driven by an autosomal-recessive mechanism.

Late-onset PD (like AD), on the other hand, has a more complex etiology involving environmental factors and not excluding some major genes of additive or recessive effects. Linkage studies of PD have identified certain distinct loci as possibly harboring risk genes for the disorder in chromosomes 1, 2, 4, 5, and X. Numerous functional candidate genes have also been reliably implicated in the etiology of PD by association analysis. GWASs have further added to this list by identifying unsuspected candidate genes for PD, as well as other late-life neurodegenerative disorders. Genes encoding alpha-synuclein (*SNCA*, which accumulates in Lewy bodies), *MAPT* (which is a constituent of tangles in AD), and leucine-rich repeat kinase 2 (*LRRK2*) have also been repeatedly associated with PD. A recent meta-analysis of GWASs found that, together, *MAPT* and *SNCA* contribute to up to 25.6% of the population's attributable risk for PD. Mutations in parkin (*PARK2*), *PTEN-induced kinase 1* (*PINK1*), and *oncogene DJ1* have been associated with autosomal-recessive parkinsonism, which is early onset and responsive to DOPA. However, considering the significant clinical role of the dopaminergic system in PD, other dopaminergic system candidate genes, such as *DRD1* and *DRD2,* which code for dopamine receptors, are not similarly implicated. Further, genes involved in other neurodegenerative disorders, such as *APOE,* have not been implicated. Another meta-analysis of genome-wide association data of 13,708 cases with PD and 9,522 comparison subjects identified 26 loci, 7 of which were novel candidate genes (*SIPA, 1L2, INPP5F,* *MIR4697, GCH1, VPS13C*, and *DDRGK1*). The genes at these 26 risk loci mediate diverse biologic processes, but collectively they are enriched with genes that bind nucleotides and genes that are involved in dopamine biosynthesis. Finally, the largest and most recent meta-analysis of late-onset PD, involving over 50,000 cases and 1.4 million control subjects, identified 90 independent genome-wide significant risk signals across 78 genomic regions, 37 of which were novel discoveries. These 90 variants explained 16% to 36% of the heritable risk of PD. Similar to GWAS in AD, these new findings are triggering new studies aimed at understanding the various functional aspects of the implicated genes.

MAJOR OR MILD NEUROCOGNITIVE DISORDER WITH LEWY BODIES (DEMENTIA WITH LEWY BODIES)

DLB is characterized by cognitive decline, parkinsonian signs and symptoms, fluctuations in cognition and attention, rapid eye-movement (REM) sleep disturbance, and visual hallucinations. Individuals with DLB can also have repeated falls, autonomic dysfunction, a heightened sensitivity to neuroleptics, disturbances in smell, systematized delusions, apathy, anxiety, and depression. Furthermore, the revised criteria also include consideration of biomarkers, such as changes in imaging studies and EEG.

DLB has a great deal of clinical overlap with both AD and PD. DLB differs from AD in that the onset of hallucinations is much earlier in DLB. Similar to PD, individuals with DLB demonstrate parkinsonian signs; however, these motor manifestations are markedly more and manifest earlier (at least 1 year before the onset of cognitive impairment) in PD as compared to DLB. DLB has been found to be largely sporadic and has been noted to be complex in nature, resulting from genetic and environmental interactions. Considering this sporadic nature, most genetic studies of DLB have been limited to family studies. Family studies have noted that DLB can be autosomal dominant or recessive. Reflecting the overlap in clinical signs, DLB shares considerable genetic overlap with AD and PD. For example, *APOE4*, synuclein genes (*SNCA* and *SNCB*), and *PSEN1* genes have been found to be risk factors in DLB. A recent genome-sequencing effort and GWAS of a large cohort of DLB cases and controls identified five independent risk loci, with mutations particularly enriched in the gene *GBA*. Genetic risk scores (PRSs) and implicated pathways were strongly shared between DLB, AD, and PD.

MAJOR OR MILD FRONTOTEMPORAL NEUROCOGNITIVE DISORDER (PICK DISEASE AND OTHER FRONTOTEMPORAL DEMENTIAS)

Frontotemporal lobar degeneration is a progressive dementia characterized clinically by prominent behavioral, executive function, and language deficits. Frontotemporal lobar degeneration is categorized into three main clinical syndromes, including frontotemporal dementia, semantic primary progressive aphasia, and nonfluent primary progressive aphasia.

Frontotemporal dementia rarely has, if at all, been evaluated within the context of family, twin, or segregation studies; thus, it is difficult to estimate the familial risk, heritability, and mode of inheritance of this group of disorders. However, within the context of several linkage studies, it has been shown that frontotemporal dementia may be inherited in an autosomal-dominant manner with age-dependent penetrance (i.e., risk increases with age). One of the most routinely identified loci linked to frontotemporal dementia has been chromosome 17q21. Approximately 10% to 20% of familial cases are accounted for by mutations in the *MAPT* and *GRN* genes, which are located in this region. In addition to these, *C9ORF72*, genetically linked to chromosome 9p, has been recently identified as having a mutation that may singly account for the most number of cases (11.7%) of familial frontotemporal dementia (FTD). The association with 9p has been found by at least two GWASs, thus giving credence to the importance of 9p in FTD. While *MAPT* is associated with increased deposition of phosphorylated tau, *GRN* and *C9ORF72* have been associated with deposition of the transactive response DNA binding protein of 43 kDa (*TDP-43*). GWAS has also identified associations with immune system processes and possibly lysosomal and autophagy pathways in FTD through 6p21.3, encompassing the *HLA* locus (immune system), and 11q14, encompassing *RAB38*/*CTSC* (transcripts of which are related to lysosomal biology). Like LBD, frontotemporal dementia also shares genes in common with AD and PD, as evidenced by the meta-analytic results of several GWASs, with associations converging at SNPs within the *HLA*, *MAPT*, and *APOE* genes.

MAJOR OR MILD VASCULAR NEUROCOGNITIVE DISORDER (VASCULAR DEMENTIA)

Vascular dementia (VaD) is the second most common form of dementia in older adults. VaD is caused by the effects of cerebrovascular disease and events (i.e., “stroke”) that initiate a cascade leading to central nervous system (CNS) damage and, ultimately, a dementia syndrome in some individuals.

While familial forms of VaD are uncommon (accounting for less than 10% of cases of VaD), cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) and hereditary cerebral hemorrhage with the amyloidosis-Dutch type (HCHWA-D) are among the better studied genetic syndromes of cerebrovascular disease and dementia. CADASIL is a syndrome of subcortical small vessel disease with lacunar strokes, migraine, and dementia. The discovery of linkage to chromosome 19q12 in two unrelated families and, subsequently, mutations in the *NOTCH3* gene explain the cause of most cases of CADASIL. *NOTCH3* is expressed in vascular smooth muscle cells and is important in cell proliferation and differentiation. FCAA, or familial cerebral amyloid angiopathy, is another heritable form of VaD, which is characterized by the deposition of amyloid on the vessel walls. A number of distinct forms of FCAA have been identified. Hereditary cerebral hemorrhage with amyloidosis-Dutch type, HCHWA-D, is an autosomal-dominant syndrome primarily of hemorrhagic strokes and dementia. Since extensive deposits of beta-amyloid protein in brain vasculature had been found in HCHWA-D, linkage to the *APP* gene was tested for and was found in two unrelated families with HCHWA-D. A point mutation in the *APP* gene was found in two patients, providing evidence that *APP* gene mutations can cause HCHWA-D. Italian and Flemish types of HCHWA also are caused by point mutations in the *APP* gene and exhibit autosomal-dominant transmission. The Icelandic type of HCHWA, also with an autosomal-dominant transmission, is caused by a point mutation in the cysteine protease inhibitor cystatin C gene, while the British type of FCAA has been associated with mutations in the *ITM2B* gene (also called *BRI2*).

The genetic underpinnings of most cases of VaD, however, remain largely undetermined. This deficiency stems largely from the heterogeneity of the illness. VaD and AD frequently occur together as “mixed dementia.” *APOE* gene variants are known to affect risk for cerebrovascular disease, and presenilins have been shown to interact with *NOTCH3*—suggesting mechanistic overlap, if not the possibility of additional genetic overlap between AD and VaD, which may be driving the comorbidity. Studies examining candidate genes have noted inconsistent associations with genes coding for proteins involved in lipid metabolism, inflammation, and the renin–angiotensin system. A complex disease such as VaD warrants exploration into epistatic interactions. Studies have identified that epistatic interactions may act as a stronger risk factor for VaD. For example, the combined genotype of tumor necrosis factor (*TNF*α T-1031C) and heat shock 70-kDa protein 1A (*HSP70-1* A-110C) showed a much stronger odds ratio as compared to the individual genotypes. Similar synergistic relationships have been noted between transforming growth factor beta 1 (*TGF-β1*) and angiotensinogen (*AGT*) genes and between family with sequence similarity 134 member B (*FAM134B* [rs10041159]) and tumor necrosis factor receptor superfamily member 19 (*TNFRSF19* [rs9317882]). A recent GWAS also noted an association with a variant near the androgen receptor (*AR*) gene on the X chromosome. While the mechanistic role of this association is unclear, this finding replicates past reports of associations with CAG repeat lengths on AR and VaD.

MAJOR OR MILD NEUROCOGNITIVE DISORDER DUE TO PRION DISEASE (CREUTZFELDT–JAKOB DISEASE AND OTHER TRANSMISSIBLE SPONGIFORM ENCEPHALOPATHIES)

Creutzfeldt–Jakob disease (CJD) is one of several neurodegenerative diseases characterized by progressive dementia and ataxia, called *transmissible spongiform encephalopathies* (TSEs). TSEs are rare diseases that include CJD, kuru, Gerstmann–Sträussler syndrome, and fatal familial insomnia. The sporadic form of CJD is the most common TSE, with one to two incident cases per million people worldwide. The TSEs are unique in that they are inherited, sporadic, or acquired by infection. An abnormal isoform of the prion protein has been implicated as the causal agent in the TSEs. The prion protein exists in one of two states: PrP^{C}, for the normal cellular form of the protein, and PrP^{SC}, for the abnormal isoform isolated from the brains of individuals with TSE. The genetic epidemiologic study of this class of late-life neurodegenerative disorders has largely bypassed behavioral genetic approaches such as family and twin studies; thus, it has never been firmly established that these disorders have a familial or genetic basis. Data from Finland showed that 6 of 30 cases observed nationwide from 1975 to 1984 were familial cases, and all belonged to the same extended kindred. Further evaluation of this pedigree identified autosomal-dominant transmission of an early-onset long-duration form of the illness.

Despite the dearth of behavioral genetic data on CJD and related disorders in general, molecular genetic methods have often been used in the study of these disorders, largely due to their ease of application. Much of the focus of earlier studies had been on the *PRNP* gene, which encodes the prion precursor protein. In fact, most of the affected individuals in the Finnish family described above also carried a nonsynonymous mutation in codon 178 of *PRNP*, whereas none of the 10 healthy controls carried the same variant. Since then, *PRNP* (specifically an amino acid polymorphism at codon 129 coding for methionine or valine) has been confirmed in a number of studies as a reliable risk factor for both variant or vCJD and sporadic or sCJD. Over the last 5 years, a number of other risk genes have been identified, although *PRNP* still remains the strongest universal factor. Variations in associations have been noted across populations for sCJD; SNPs at the *ZBTB38–RASA2* locus (a zinc finger transcriptional activator that binds methylated DNA) have been associated with CJD in U.K. populations, while an SNP in the *CHN2* gene has been found to be associated with vCJD in a German population. Some of the other associations include myotubularin-related protein 7 gene (*MTMR7)* and neuronal PAS (per-ARNT-sim) domain-containing protein 2 gene (*NPAS2*), which have been identified to increase risk independently, and β-site APP cleaving enzyme 1 (*BACE1)* which influences the risk of sCJD by interacting with *PRNP* gene.

FUTURE CLINICAL DIRECTIONS

The DSM-5 introduced a new category of neurocognitive disorder—mild neurocognitive disorder (or mNCD). The introduction of mNCD in DSM-5 helped formalize a uniform clinical definition which, in turn, has furthered our understanding of factors that drive the progression to dementia. Understanding the genetics of mNCD becomes critical for the purposes of early intervention and prevention and also furthering personalized medicine. Complementing the DSM-5 introduction of mNCD, the U.S. National Institute on Aging (NIA) has offered further recommendations to provide a standardized language for researchers. These recommendations focus on the biologic aspect of the neurodegenerative disease and position symptoms as a consequence of the underlying disease process. This introduction is aimed at two critical elements. The first is the delineation of a disease stage prior to the onset of mNCD, called cognitively unimpaired (CU). Currently, considering that the diagnosis is driven by biomarkers and brain pathology, it has been difficult to identify individuals belonging to this category. However, as more clinical biomarkers, such as loss of smell, become defined, it can potentially lead to primary prevention efforts. A second new area the proposed research definition adds is the recognition and identification of newer biologic factors responsible for the dementia syndrome, such as limbic-predominant age-related TDP-43 encephalopathy or LATE, which shares mechanisms with both AD and FTD and argyrophilic grain disease (AGD), which has been identified as a cause for tau hyperphosphorylation, providing more information about the underlying disease process.

Identification of the CU population, in addition to the increased recognition of the role of resilience in symptomatic dementia, opens up an exciting new frontier in the prevention of dementia. Approximately 30% of individuals with the neuropathology of AD do not develop clinical symptoms; of the remaining 70%, the degree of severity of symptoms varies, highlighting a potential cognitive or neural basis for resilience. Recent studies have identified a number of candidate genes modulating the progression of cognitive decline in individuals with AD. As an example, the *BDNF* Val66Met SNP was associated with a faster rate of cognitive decline, while the *PLXNB1* gene was associated with slower cognitive decline. A more recent GWAS of around 5,108 individuals identified resilience genes located on chromosome 8, upstream of *ATP8B1*, which is involved in the bile acid metabolism pathways.

FUTURE RESEARCH DIRECTIONS

Over the last decade, significant strides have been made in understanding the genetics of late-life neurodegenerative disorders. Despite the initial successes, association studies remain plagued by limitations, including their propensity to produce false-positive results and their limited breadth in terms of identifying causal variants rather than markers. Regarding the former, genes identified as associated with a late-life neurodegenerative disorder in an initial study often overestimate the true effect size and subsequently fall victim to the winner's curse, wherein the same magnitude of an effect cannot be replicated. Regarding the latter, association methods remain focused on the physical localization of association signals rather than on true functional or causal mutations. Thus GWAS can implicate regions of the genome that may harbor contributory loci, but much work remains to fine-map the associated region to identify the true source of its association signal. The result of widespread GWAS application has been the elucidation of a more complex profile of gene involvement. The discovery of the complexity led to the identification of newer and advanced methods, including imputation, which relies on the knowledge that certain gene sequences are usually coinherited. GWAS often identifies a large number of loci, some of which are noncoding. This has led to the identification of expression quantitative trait loci (eQTLs). eQTLs are regions in the genome that regulate nearby genes. In their attempts to integrate functionality with observations, researchers are also beginning to look into one of the limitations of GWASs—the tendency to filter out "near misses." Preliminary work in this area has led to the identification of an "inflammatory subtype" of PD, which may have greater involvement of genes associated with the immune system.

The multifactorial polygenic etiologies and heterogeneity of late-life neurodegenerative disorders have proved to be a huge challenge in the discovery of their underlying genetic bases. The multifactorial polygenic nature of these diseases dictates that the "signals" obtained in linkage and association studies will be numerous and of very low intensity, while the heterogeneity of these disorders increases the "noise" against which these already faint signals must be detected. Some of the most effective ways of combating these complexities are through the maximization of power, both by increasing sample size and by studying homogeneous groups of affected individuals. To identify the numerous genes with small effects on risk for late-life neurodegenerative disorders, various data-pooling strategies, such as continued combination of study-level data by meta-analysis or direct combination of primary data by mega-analysis, can be particularly effective. However, to overcome the obstacles introduced by heterogeneity, delineating and studying smaller, more homogeneous subgroups of affected individuals may have the greatest beneficial effects on power. When used in tandem, the practices of pooling and splitting will yield maximal power for genetic studies and hasten the discovery of the full compendium of risk genes for each heritable late-life neurodegenerative disorder.

Other promising avenues for elucidating risk genes include the analysis of gene–gene interactions as well as the aggregation of risk genes into PRSs, which, in aggregate, account for a larger proportion of variance in susceptibility than any single variant in isolation. As an example, a genetic risk score based on *ABCA7*, *BIN1*, *CD2AP*, *CLU*, *CR1*, *MS4A4E*, *MS4A6A*, and *PICALM* was noted to predict the rate of progression (a twofold increase in rate) to AD from mNCD. But, perhaps, one of the most interesting shifts in paradigm is the recommendation to examine neurodegenerative disorders from the bottom-up; that is, identifying common biologic pathways in neurodegenerative disorders that become individualized, leading to unique clinical and pathologic characteristics. Recent GWASs have emphasized the biologic pathways that have subserved shared mechanisms between AD and PD (as well as LBD and FTD), such as apoptosis, autophagy, mitochondrial dysfunction, the ubiquitin–proteasome system, oxidative DNA damage, inflammation, lipid metabolism, and vascular changes.

CLINICAL IMPLICATIONS

Since the first draft of the human genome was produced in 2001, scientists have touted the promise of genetic analysis for identifying risk factors for common diseases and the development of personalized (or, more recently, “precision”) medicine. Yet, the vast potential of genetic studies to change the clinical practice of psychiatry is just beginning to be tapped. Recently, commercial genetic tests for (especially Mendelian forms of) late-life neurodegenerative disorders have become available to individuals with a strong family history of illness (e.g., *APP*, *PSEN1*, and *PSEN2* for AD, *MAPT* for frontotemporal dementia and Pick disease, *PRNP* for CJD, and *PARK*s *1*, *2*, *6*, *7*, and *8* for PD). However, while individuals in the general population can be tested for the possession of particular genotypes to determine their cumulative risk of developing a particular disorder, these commercially available tests do not yield accurate or actionable estimates of risk. Similarly, over the past decade, progress has been made in understanding the factors that contribute to conversion to AD. However, despite these advances, the leads that have been generated from molecular genetic studies are simply not yet understood well enough.

Once the many and diverse genetic risk factors for these conditions become established and widely recognized, they can serve many purposes for early intervention and prevention efforts. For example, understanding risk factors involved in conversion to dementia could speed the initiation of appropriate treatments, which consequently may promote better prognoses through early intervention efforts. Needless to say, the most ideal situation would be preventing the development of cognitive impairment and, perhaps, identification of individuals at high risk of developing cognitive impairment through the development of biomarkers. The operationalization of the definition of early cognitive impairment will clearly advance the investigation into early identification and prevention. Knowledge of genetic risk factors will only help so much in the management of late-life neurodegenerative disorders; in order to fully realize the promise of genomics, linkage, GWAS, and sequencing studies must be followed by functional and mechanistic studies, enabling us to identify the biology and pathophysiology underlying these diseases. Understanding the functional biologic consequences of possessing certain risk-conferring genetic variants will allow for the development of more precise and effective therapeutics. All of this information may someday be harnessed to identify the earliest biologic changes associated with a trajectory toward a late-life neurodegenerative disorder, and this in turn, may help identify specific molecular targets around which the next wave of pharmacotherapies can be designed. Furthermore, the delineation of numerous distinct etiologies for what are now considered unitary disorders (e.g., PD) may allow, in the future, for these disorders to become recognized as distinct and treated as such (e.g., *PARK1*-positive PD, *PARK2*-positive PD, etc.). Recent successes in identifying specific risk genes for these disorders have been encouraging. Replicating and refining these results, and translating them into meaningful clinical applications signals exciting times ahead for research on the genetics of late-life neurodegenerative disorders.

FURTHER READINGS

Adams HHH, de Bruijn RFAG, Hofman A, et al. Genetic risk of neurodegenerative diseases is associated with mild cognitive impairment and conversion to dementia. *Alzheimers Dement*. 2015;11(11):1277–1285.

Baizabal-Carvallo JF, Jankovic J. Parkinsonism, movement disorders and genetics in frontotemporal dementia. *Nat Rev Neurol*. 2016;12(3):175–185.

Beecham GW, Hamilton K, Naj AC, et al; Alzheimer's Disease Genetics Consortium (ADGC). Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet*. 2014;10(9):e1004606.

Bennett DA, Yu L, Yang J, Srivastava GP, Aubin C, De Jager PL. Epigenomics of Alzheimer's disease. *Transl Res*. 2015;165(1):200–220.

Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet*. 2007;39(1):17–23.

Chia R, Sabir MS, Bandres-Ciga S, et al. Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. *Nat Genet*. 2021;53(3):294–303.

Corder EH, Saunders AM, Risch NJ, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet*. 1994;7(2):180–184.

Croes EA, Alizadeh BZ, Bertoli-Avella AM, et al. Polymorphisms in the prion protein gene and in the doppel gene increase susceptibility for Creutzfeldt-Jakob disease. *Eur J Hum Genet*. 2004;12(5):389–394.

Dumitrescu L, Mahoney ER, Mukherjee S, et al. Genetic variants and functional pathways associated with resilience to Alzheimer's disease. *Brain*. 2020;143(8):2561–2575.

Ferrer I, Santpere G, van Leeuwen FW. Argyrophilic grain disease. *Brain*. 2008;131(Pt 6):1416–1432.

Ferrari R, Hernandez DG, Nalls MA, et al. Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol*. 2014;13(7):686–699.

Ferrari R, Wang Y, Vandrovcova J, et al; International FTD-Genomics Consortium (IFGC), International Parkinson's Disease Genomics Consortium (IPDGC), International Genomics of Alzheimer's Project (IGAP). Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. *J Neurol Neurosurg Psychiatry*. 2017;88(2):152–164.

Guerreiro R, Gibbons E, Tábuas-Pereira M, Kun-Rodrigues C, Santo GC, Bras J. Genetic architecture of common non-Alzheimer's disease dementias. *Neurobiol Dis*. 2020;142:104946.

Hohman TJ, McLaren DG, Mormino EC, Gifford KA, Libon DJ, Jefferson AL; Alzheimer's Disease Neuroimaging Initiative. Asymptomatic Alzheimer disease: defining resilience. *Neurology*. 2016;87(23):2443–2450.

International Parkinson Disease Genomics Consortium; Nalls MA, Plagnol V, Hernandez DG, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*. 2011;377(9766):641–649.

James BD, Bennett DA. Causes and patterns of dementia: an update in the era of redefining Alzheimer's disease. *Annu Rev Public Health*. 2019;40:65–84.

Joutel A, Corpechot C, Ducros A, et al. Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature*. 1996;383(6602):707–710.

Leblanc GG, Meschia JF, Stuss DT, Hachinski V. Genetics of vascular cognitive impairment: the opportunity and the challenges. *Stroke*. 2006;37(1):248–255.

Lee C, Kim Y. Complex genetic susceptibility to vascular dementia and an evidence for its underlying genetic factors associated with memory and associative learning. *Gene*. 2013;516(1):152–157.

Levy-Lahad E, Wasco W, Poorkaj P, et al. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science*. 1995;269(5226):973–977.

Lloyd SE, Mead S, Collinge J. Genetics of prion diseases. *Curr Opin Genet Dev*. 2013;23(3):345–351.

Meeus B, Theuns J, Van Broeckhoven C. The genetics of dementia with Lewy bodies: what are we missing? *Arch Neurol*. 2012;69(9):1113–1118.

Morton NE. *Outline of Genetic Epidemiology*. Karger; 1982.

Nalls MA, Blauwendraat C, Vallerga CL, et al; 23andMe Research Team, System Genomics of Parkinson's Disease Consortium, International Parkinson's Disease Genomics Consortium. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2019;18(12):1091–1102.

Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain*. 2019;142(6):1503–1527.

Nelson PT, Wang WX, Partch AB, et al. Reassessment of risk genotypes (GRN, TMEM106B, and ABCC9 variants) associated with hippocampal sclerosis of aging pathology. *J Neuropathol Exp Neurol*. 2015;74(1):75–84.

Saunders AM, Strittmatter WJ, Schmechel D, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology*. 1993;43(8):1467–1472.

Schrijvers EMC, Schürmann B, Koudstaal PJ, et al. Genome-wide association study of vascular dementia. *Stroke*. 2012;43(2):315–319.

Tan EK. The role of common genetic risk variants in Parkinson disease. *Clin Genet*. 2007;72(5):387–393.

Tan EK, Khajavi M, Thornby JI, Nagamitsu S, Jankovic J, Ashizawa T. Variability and validity of polymorphism association studies in Parkinson's disease. *Neurology*. 2000;55(4):533–558.

55.3 Psychiatric Disorders of Late Life

55.3a Assessment of Functioning

David J. Moore, Ph.D., Maulika Kohli, M.S., and Pariya L. Fazeli, Ph.D.

Older adults are the fastest-growing segment of the population in the United States. By 2030, the number of Americans aged 65 and older will more than double to 77 million; roughly 1 in 5 Americans will be over 65. Given that older people represent such a large part of the American population, it is increasingly important that clinicians and researchers in the field of psychiatry be prepared to work with and care for older adults. Older adults, like any other segment of the population, will require care for physical and mental health problems, and they may be at increased risk for impairment in both of these areas. A decline in everyday functioning may be associated with mental and physical health problems. Clinicians are increasingly asked to make decisions regarding patients' capacity to live independently, as well as to determine their capacity to conduct other daily tasks such as managing health care and finances. This chapter defines and reviews domains of everyday functioning and discusses factors (e.g., neurocognitive impairment) that are likely to interfere with daily functioning in older adults. Additionally, this chapter discusses various assessment methods and specific clinical populations, and we make recommendations for future instrument development. The authors hope that this chapter will serve as a resource for clinicians and spur additional needed research on everyday functioning difficulties among older adults.

WHAT IS EVERYDAY FUNCTIONING?

Before discussing the various assessment modalities and other factors related to everyday functioning, a few terms must be defined. Quite simply, *everyday functioning* refers to all of the activities in which people engage in their daily lives. By this definition, *everyday functioning* is different for every person. The nature of one's daily activities and behaviors depends on numerous factors, including culture, socioeconomic status (SES), and the presence or absence of medical or psychiatric conditions. Despite the wide range of possible daily activities, certain basic behaviors are nearly universal. They range from basic *Activities of Daily Living (ADLs)* such as eating and grooming to more complex behaviors, often termed *Instrumental Activities of Daily Living (IADLs)*, such as managing finances and medications. Some scientists would confine the assessment of functioning to only these physical behaviors, but we take a broader view. For instance, the authors believe that social skills and quality of life (QoL), neither of which is based on physical ability, fit nicely under the umbrella of everyday functioning.

Even though many useful everyday functioning instruments exist, so far, researchers have not reached a consensus on the instrument or instruments that would represent the "gold standard." Several factors may have contributed to this. First, the assessment of everyday functioning in older adults is important to a wide variety of disciplines, such as geriatrics, gerontology, occupational therapy, psychiatry, psychology, nursing, general practice, neurology, and psychiatry, each of which has its own particular focus. This diversity of focus is apparent in the literature as well. Because of the multitude of disciplines conducting research on this topic, different terms are often used when referring to very similar constructs. As a result, no unified search term for "everyday functioning" is available in library reference databases (e.g., Medical Subject Headings, or MeSH) in contrast to such terms as "neuropsychology" or "cognition." The most widely used terms for everyday functioning are *ADL*, which refers to very basic skills needed to survive (basic ADLs or BADLs), and *IADL*, which covers higher-order tasks that represent a level of independence that may or may not be present among individuals with intact ADLs. These terms are discussed and defined in more detail below.

Many individuals experience declines in their ability to perform everyday tasks as they age. Some of these declines may be considered normal, whereas other impairments may result from a disease process. Declines in daily functioning can result from numerous factors, including physical, medical, cognitive, and psychiatric factors. Daily functioning ability is so important that it is included as a diagnostic criterion in the determination of many psychiatric and cognitive disorders. For instance, a diagnosis of major neurocognitive disorder due to *Alzheimer Disease* (AD), according to the fifth edition of the *Diagnostic and Statistical Manual for Mental Disorders* (DSM-5), requires cognitive deficits that “interfere with the independence of everyday activities.” Due to the typically late onset, many individuals who are diagnosed with AD are no longer working. The specific means used to determine “independence of everyday activities” are typically left to the discretion of the clinician making the diagnosis. Along these same lines, a diagnosis of mild neurocognitive disorder due to AD—thought to be a transition between normal cognitive functioning and dementia—requires that “cognitive deficits do not interfere with the capacity of independence in everyday activities.” Again, the methods for determining whether daily living skills are intact are often left to the clinician and based upon historical clinical notes when available. Some standardized methods for assessing everyday functioning in mild neurocognitive disorder are available. For example, the Functional Activities Questionnaire (FAQ) is a widely used 10-item (e.g., preparing a meal, monitoring family occasions or personal medications) measure of performance of daily activities, which assesses difficulty ranging from no difficulty to dependent. This brief measure is appropriate for clinical settings and can be completed by the person or an informant. However, differences exist clinically and in research settings for how “independence” is defined. Conservative operationalizations require responses of “no difficulty” or “difficulty but still performing by self,” whereas more liberal definitions allow responses of “performing with assistance.”

DOMAINS OF EVERYDAY FUNCTIONING ASSESSMENT

Daily functioning tasks lie on a continuum that ranges from tasks that are necessary to function independently (e.g., the ability to get oneself to the toilet and use it) to tasks that are optional but common (e.g., driving a vehicle in the United States) and promote living independently. Tasks falling in between these extremes, such as managing one’s finances, are not critical to living independently, but the inability to perform them requires strong environmental or personal support such as a structured living facility or a committed family member or caregiver. Although we prefer not to classify daily functioning skills by category, several terms are used in the literature. These terms should be familiar to clinicians so that they are able to communicate effectively with other members of a multidisciplinary clinical team.

Activities of Daily Living

As discussed previously, some basic daily functions seem to cut across cultures. Termed ADLs, these are personal care behaviors and basic physical skills necessary in daily life. They include walking, grooming, bathing, and dressing.

Table 55.3a–1.
Typical Everyday Functioning Domains by Activities of Daily Living (ADL), Instrumental Activities of Daily Living (IADL), and Others

ADLs	IADLs	Beyond IADLs
Physically based ADLs Transferring ▶ Chair, bed, wheelchair ▶ Toilet ▶ Tub/shower Walking ▶ Given distance ▶ Climbing stairs ▶ Uneven surfaces **General ADLs** Grooming/hygiene Bathing Toileting Continence ▶ Bladder ▶ Bowel Dressing Eating/feeding	Financial management Medication management Cooking Laundry Cleaning Communication ▶ Written ▶ Spoken ▶ Reading ▶ Telephone Shopping Housekeeping Gardening Driving Hobbies Safety	Social Functioning Driving Quality of Life Multitasking

Some ADLs are more physically based than others (Table 55.3a–1). They are sometimes simply referred to as physical functions. The assessment of physical functioning is often left to physiatrists, physical therapists, rehabilitation specialists, and general practice physicians. Often, these practitioners will use the *Functional Independence Measure* (FIM) (later described in the Clinician Rating of Everyday Functioning section) or a similar measure to describe the level of physical dependence. Perhaps the most basic form of everyday functioning assessment is determining whether an older individual has the physical capabilities to perform basic motor tasks. Motor-based assessments such as timed gait and grip strength may provide additional information in this area. For example, the Timed Up and Go Test (TUG) tests the time it takes to rise from an armchair, walk 3 m, turn, walk back, and sit down, and has been shown to be well correlated with scores on scales of balance, gait speeds, and ADL, and is predictive of falls. Another measure of physical functions is the Short Physical Performance Battery (SPPB), which evaluates balance, gait, strength, and endurance by examining the ability to stand with the feet together in side-by-side, semitandem, and tandem positions; time to walk 8 ft; and time to rise from a chair and sit back down five times. This test has been reported to predict mortality, disability, and nursing home admissions.

Other ADLs are required for autonomous functioning at home (e.g., grooming, bathing, dressing). People incapable of conducting these basic activities require significant assistance and are likely to live in a structured setting (e.g., a nursing home) or receive assistance in completing these tasks from a full-time caregiver. A list of common ADLs can be found in Table 55.3a–1.

Instrumental ADLs

Assessments of IADLs were developed to assess certain higher-level abilities that are required for autonomous functioning within society. Examples of IADLs include managing medications, balancing a checkbook, and preparing a meal (see Table 55.3a–1). Assessments of IADLs are targeted at older people who live independently and have little problem with basic ADLs but may need some assistance with higher-order tasks. Most assessments of everyday functioning either focus on ADLs or IADLs, and assessment approaches are discussed below.

Beyond IADLs

There are several abilities that go beyond IADLs but can nevertheless be important for higher levels of daily functioning. These include social functioning, QoL, and driving.

Social Functioning. The ability to function socially can be important for many aspects of an older adult's life. For instance, social skills are essential to maintaining friendships and remaining actively engaged with other people. Social skills are also critical in many everyday situations, such as articulating an error on a recent bill or discussing health care choices with a physician. Assessments of social skills may be particularly important in conditions known to impact these abilities, including dementia and schizophrenia.

Driving. Many older adults continue to drive in later life. Although they typically drive fewer miles per year than younger drivers, they are at greater risk for accidents than all other age groups, with the exception of teenagers. The risk for accidents seems to increase most noticeably after age 75. After adjusting for miles driven per year, those over age 75 have the greatest number of fatal crashes of any group.

Given that older adults are at significant risk for driving-related accidents and even death, it is important for clinicians to understand what factors put these individuals at risk so that they can appropriately intervene. Equally important is to avoid stigmatizing older adults who are still capable of safe driving. Decisions regarding the advisability of continued driving can be difficult since driving often gives older people a sense of independence and allows them to maintain meaningful social relationships that may be otherwise hampered. The final decision regarding continued driving typically rests with each state's Department of Motor Vehicles (DMV). While many states have adopted special provisions for older drivers, such as no mail renewal after a certain age, very few states require individuals to take an exam or a road test once a certain age is reached. Some states require clinicians to report patients diagnosed with dementia to the DMV (e.g., California). The DMV then evaluates the driver to determine if they may retain driving privileges.

Neuropsychological abilities play a large role in determining whether people of any age should be allowed to continue driving; however, given that dementia is more prevalent in older adults, it is especially important in this population. Cognitive domains that are more strongly associated with impaired driving performance include memory, executive functioning, and processing speed. Depression, which is common among older adults, can also contribute to decreased responsiveness and poor concentration and lead to an increased risk for driving accidents. Some evidence exists that cognitive training may help improve processing speed among older adults.

Quality of Life. QoL, while not in itself a functional ability, is perhaps the most important outcome affected by daily functioning. Certainly, improved QoL is an assumed treatment goal with any medical or psychological intervention. QoL can be viewed from two perspectives. The global perspective considers all aspects of a person’s life; these assessments are typically both subjective and multidimensional. Alternatively, health-related QoL (HRQoL) focuses exclusively on the effects of health, illness, and treatment. HRQoL can be assessed generically or with reference to a specific disease. The most commonly used generic HRQoL instrument is the 36-Item Short-Form Health Survey. Although several HRQoL instruments are used with older adults, the authors are unaware of any specifically designed for this population. Physical functioning is often a part of HRQoL questionnaires and may unknowingly bias these instruments against older adults. A sampling of QoL and HRQoL instruments is shown in Table 55.3a–2. By nature, these instruments are primarily self-report, although some collateral-report measures are available. An expert panel of the National Institute on Aging recommended that older adults, particularly those with multiple chronic conditions, complete a brief but comprehensive (i.e., assesses physical symptoms and mental health symptoms) composite screening measures such as the SF-8/SF-36 or the Patient Reported Outcomes Measurement Information System 29-item Health Profile (PROMIS-29) in clinical settings. This recommendation was made to identify domains where additional follow-up measures may be warranted (e.g., depression, daily functioning).

Table 55.3a–2.
Commonly Used Quality of Life (QoL) and Health-Related Quality-of-Life (HR-QoL) Measures for Older Adults

Quality of Life

Comprehensive Assessment and Referral Evaluation (CARE)
EASY-Care
Functional Assessment Inventory (FAI)
IOWA Multidimensional Functional Assessment Questionnaire (OMFAQ)
Perceived Well-being Scale (PWB)
Philadelphia Geriatric Centre Multilevel Assessment Instrument (PGCMAI)
Quality-of-Life Profile: Seniors Version (QoLPSV)

Health-Related Quality of Life (HR-QoL)

36-Item Short-Form Health Survey (SF-36)
12-Item Short-Form Health Survey (SF-12)
8-Item Short-Form Health Survey (SF-8)
EuroQoL Five-Dimension Questionnaire (EQ-5D)
Feeling Thermometer Medical Outcomes Study (MOS)-derived questions
Quality of Well-Being (QWB) Scale
Sickness Impact Profile (SIP)
Patient Reported Outcomes Measurement Information System 29-item Health Profile (PROMIS-29)

FACTORS INFLUENCING EVERYDAY FUNCTIONING

Given the complexity of most everyday functioning tasks, many factors may significantly affect performance. These include, but are not limited to, cognition, mood, physical health, culture and ethnicity, and the physical environment. Research on the relative impact of each of these factors indicates that each factor has shown varying levels of influence. A discussion of each of these factors and the evidence supporting its role in everyday functioning difficulties is provided below.

Cognition and Everyday Functioning

Of all the factors pertinent to daily functioning, the strongest evidence exists for cognitive impairment as a significant contributor to the decline in many everyday functioning abilities. However, this literature discussing the relationship between cognition and everyday functioning is not as well developed and complete as one might expect. A recent comprehensive review of the cognitive correlates of functional status found that (1) the proportion of variance accounted for by cognition was modest, (2) some cognition domains had a stronger relationship to functional status than others, (3) cognitive impairments in executive control may be particularly strong correlates of impaired medication management, multitasking, and financial decision-making, and (4) global cognitive screening instruments may be relatively strong predictors of functional status. Historically neuropsychological tests have undergone a more rigorous test-development process than have everyday functioning assessments, and consequently, their reliability and validity are better established.

Many of the early researchers on functional abilities (e.g., Lawton, Brody) hypothesized that daily functioning skills would be lost in the opposite order from that in which they were acquired. Indeed, more complex daily tasks seem to be more susceptible to cognitive impairments than basic tasks.

The role of neurocognitive impairment for daily functioning in older adults is not the same for all disorders. For instance, some individuals may have cognitive difficulties that have long affected their daily functioning (e.g., an older person with schizophrenia), whereas others may have experienced recent and sudden changes (e.g., an older person who recently suffered a stroke). Moreover, the profiles of neuropsychological impairment vary greatly among the disorders found in older adults, and these distinct cognitive profiles may lead to different declines in daily functioning abilities. For instance, memory impairments offer a unique set of difficulties for everyday functioning as compared to motor impairments. A person with declarative memory impairment may have difficulty remembering to take his medications, whereas a person with motor difficulties may remember to take his medications but have difficulty opening the pill bottle.

Mood and Everyday Functioning

Depressed mood and other psychiatric difficulties appear to contribute less to everyday functioning impairment than cognitive difficulties. Obviously, people who are severely depressed or experiencing florid psychotic symptoms may experience significant impairment of daily functioning; however, people who are only mildly to moderately impacted by psychiatric symptoms appear to suffer a less resultant functional disability than do people with comparable impairments in cognition. This issue is discussed further in the “Clinical Populations and Everyday Functioning” section. It is important to note that psychiatric disturbances may have a greater impact on reporting everyday functioning problems on questionnaires rather than performance-based tasks.

Physical Disability and Everyday Functioning

People who experience physical disabilities typically use appropriate assistive devices and compensatory strategies, allowing them to adapt better to able-bodied environments. For example, people recently confined to a wheelchair may need special accommodations installed in their dwellings. On the whole, there are more available and established compensatory strategies to improve ADL functioning in people with physical disabilities compared to cognitive disabilities, although people with a physical disability may still require assistance with ADLs. Several methods for assessing physical functioning are found in the “Assessment of Everyday Functioning” section.

Culture and Ethnicity in Everyday Functioning

Ethnicity and culture can play a substantial role in the types of everyday activities in which people engage as well as in the support that they receive in accomplishing those activities, in both normative and pathologic settings. Certainly, individualistic cultures differ from collectivist cultures in the amount of support that each offers. For example, Latinos and many Asian cultures have, comparatively speaking, a collectivist approach to functional impairment in family members. Households are frequently multigenerational, encompassing parents, grandparents, and other older relatives, as well as the householders and their children. Hence, families in these cultures often provide greater support to older family members in carrying out daily tasks than do people from families with individualistic cultural backgrounds (e.g., Caucasian Americans). Latino and Asian families, by virtue of their family-oriented support system, may be less likely to seek treatment for an older family member who is experiencing decline and may prefer to handle the matter within the family. Another consideration is the shame associated in some cultures (e.g., Japanese Americans) with acknowledging mental health difficulties in oneself or a family member. In such cases, people may try to convey mental or cognitive difficulties by presenting with physical complaints (e.g., referring to a decline in cognitive abilities as a headache).

Given such cultural considerations, using everyday functioning assessment instruments designed primarily for English-speaking Caucasian people with individuals of other cultural backgrounds or who do not speak English as a first language can be problematic and lack ecologic validity. People from different cultures may perform poorly on assessments of a daily functioning task for cultural reasons unrelated to functional capacity. For example, in one study, Black and Latino subjects over 55 years of age who had recently been prescribed an antidepressant attributed intentional nonadherence to concern about side effects and stigma; they also attributed less importance to antidepressants than other medications. Other considerations include that within different cultures and households across socioeconomic backgrounds, certain everyday tasks may not be performed at all (e.g., online banking), and thus tasks assessing such domains would not reflect actual abilities. In some cases, everyday functioning instruments have been translated and culturally modified for use with other, non–English-speaking populations, for example, Chinese older adults, as shown in Table 55.3a–3; however, there is clearly a need for additional translations, adaptations, and culturally appropriate modifications.

Environment and Everyday Functioning

An older person’s physical environment also can negatively affect every day functioning abilities. People with limited financial resources are often living in physical environments that exacerbate their everyday functioning difficulties. For instance, people with limited ability to climb stairs may live in apartments on the upper floors of buildings without elevators, or they may not be able to afford to have bars installed over the toilet to assist with transferring. Likewise, a person of lower SES may not be able to afford to hire a health care assistant (e.g., for help with managing medications).

ASSESSMENT OF EVERYDAY FUNCTIONING

Standardized assessment instruments and procedures have been developed to evaluate everyday functioning. Functional assessments typically aim to determine the level of ability in a given domain of daily functioning. Depending on the individual being evaluated, an assessment may focus on basic ADLs (e.g., bathing or grooming), physically based ADLs like mobility (e.g., ability to walk without assistance), or IADLs (e.g., cooking, financial management). Table 55.3a–1 shows some of the domains typically assessed.

Table 55.3a–3.
Multidomain Performance-Based Batteries of Everyday Functioning

Battery Name	Brief Description	Site	Scoring	Admin. Time (Minutes)	Notes
Dementia (Primarily Alzheimer Disease–AD, some dementia NOS)					
Activities of Daily Living Situational Test (AST) (Skurla, 1988)	4 subscales; 38 individual items. Item examples include making instant coffee, and finding and calling a phone number.	Clinic	Items scored from 0 to 4. Total score ranges from 0 to 152.	15–20 minutes	
Cognitive Performance Test (CPT) (Burns, 1994)	6 tasks; ratings of specific behaviors are used to place individuals into 1 of 6 levels of disability. Task examples include getting dressed, and preparing a piece of toast.	Clinic	Tasks scored from 1 to 6. Total score ranges from 6 to 33.	45 minutes	Requires training and special equipment. Available for a specific subpopulation (older adult Israelis with and without dementia).
Direct Assessment of Functional Abilities (DAFA) (Karagiozis, 1998)	10 domains; Designed to measure Pfeffer Functional Activities Questionnaire (PFAQ) IADL items. Examples include writing a check and making a sandwich.	Clinic	Items are scored from 0 to 3.	45–90 minutes	The DAFA has been translated and adapted for use with Hong Kong Chinese older adults.
Direct Assessment of Functional Status (DAFS) (Loewenstein, 1989)	7 subscales; 53 individual items. Item examples include balancing checkbook and writing out a check.	Clinic	Items scored from 0 to 2. Total score ranges from 0 to 106.	25–40 minutes	DAFS has also been used with other geriatric samples, e.g., older schizophrenia patients.
Functional Performance Measure (FPM) (Carswell, 1995)	3 subscales; 8 subsections; total of 40 items. Participants are videotaped doing tasks and then the videotapes are rated. Item examples include washing and drying hands, and putting on and taking off sweater.	Clinic	Each subsection scored from 0 to 50. Total score ranges from 0 to 400.	<20 minutes	
Refined ADL Assessment Scale (RADL) (Tappen, 1994)	5 domains; 14 individual items. Item examples include washing hands and brushing teeth.	Clinic	Tasks are broken into component parts and scored from 1 to 6.	5–60 minutes for eating subtask Total >90 minutes	
Structured Assessment of Independent Living Skills (SAILS) (Mahurin, 1991)	10 domains; 50 individual items. Item examples include zipping a jacket, and using a key.	Clinic	Items scored from 0 to 3. Total score ranges from 0 to 150. Both time and accuracy contribute to the score.	60 minutes	An abbreviated version of the SAILS is available. Eating and money subscales validated for older Mexican Americans.
Texas Functional Living Scale (TFLS) (Cullum, 2001)	5 subscales; 21 individual items. Item examples include making change and remembering to take medications.	Clinic	Items scored from 0 to 5. Total score ranges from 0 to 52.	15–20 minutes	
Bay Area Functional Performance Evaluation (BaFPE) (Houston, 1989)	2 general domains; 5 tasks in the first domain, 7 items in the second. Item examples include filling out a bank deposit slip and drawing a person.	Clinic	Rated from 1 to 4.	>45 minutes	This information is based on the revised BaFPE.
Older Psychiatric Patients (Primarily schizophrenia, some mixed psychiatric groups)					
Kohlman Evaluation of Living Skills (KELS) (Thomson, 1992)	5 subscales; 17 areas assessed. Item examples include budgeting money and recognition of dangerous situations.	Clinic	Item scores range from 0 to 1; ½ points are possible.	20–30 minutes	Test sold by publisher. KELS is a combination of self-report, observation, and performance-based tests.
Performance Assessment of Self-Care Skills (PASS) (Rogers, 1984; Holm, 1999)	In the home version, participants are observed conducting various tasks primarily with their own tools and props. Example items include sweeping and mailing bills.	Home	Most recent version is scored on a 4-point ordinal scale ranging from 0 to 3. Higher scores indicate better performance.	>90 minutes	Contains both physical, ADL, and IADL sections.
UCSD Performance-based Skills Assessment (UPSA) (Patterson, 2001)	5 subscales; item examples include shopping for items in mock grocery store and planning a recreational activity.	Clinic	Each subscale transformed to 0–20 score. Total score ranges from 0 to 100.	30 minutes	

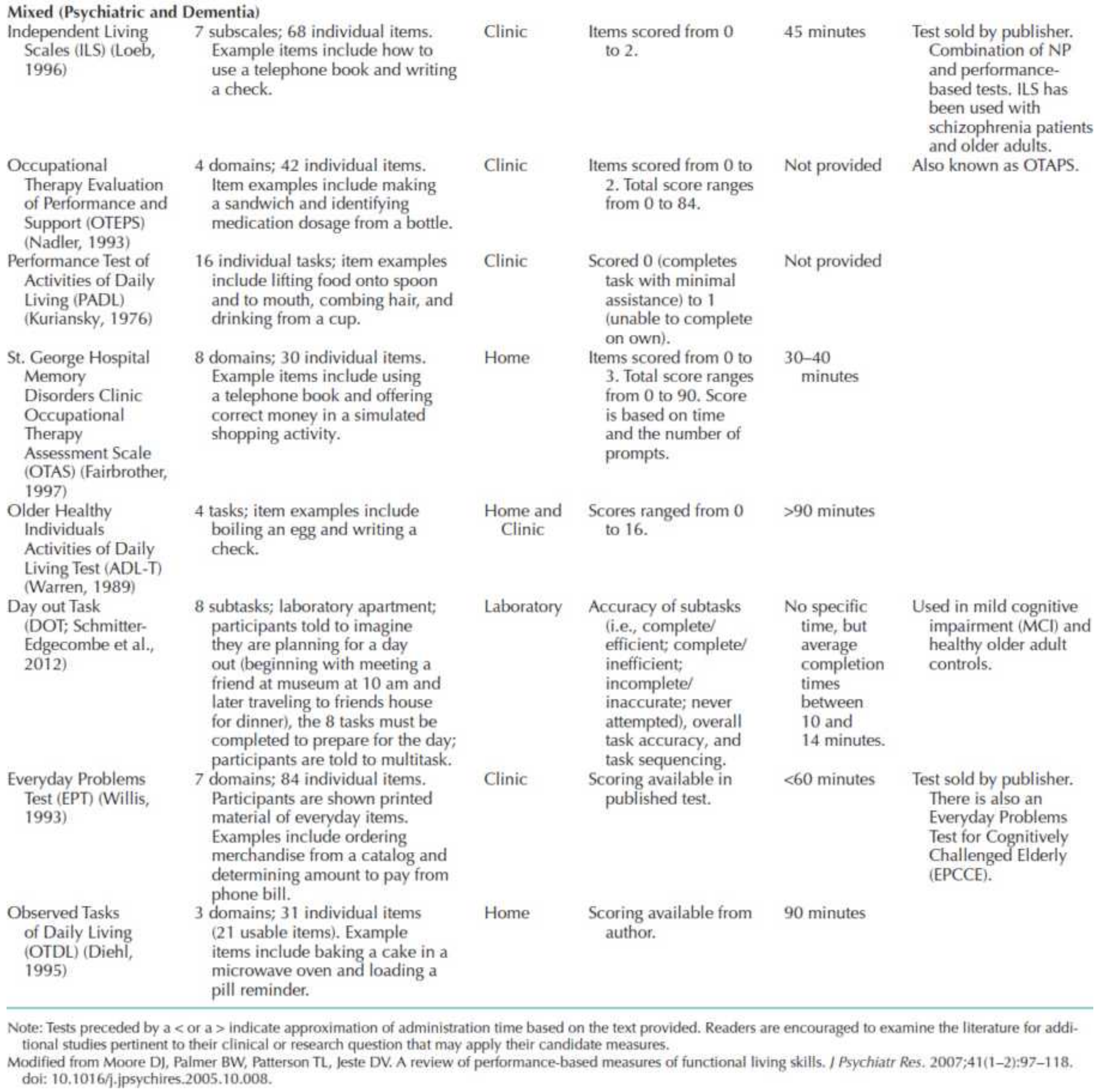

Mixed (Psychiatric and Dementia)					
Independent Living Scales (ILS) (Loeb, 1996)	7 subscales; 68 individual items. Example items include how to use a telephone book and writing a check.	Clinic	Items scored from 0 to 2.	45 minutes	Test sold by publisher. Combination of NP and performance-based tests. ILS has been used with schizophrenia patients and older adults.
Occupational Therapy Evaluation of Performance and Support (OTEPS) (Nadler, 1993)	4 domains; 42 individual items. Item examples include making a sandwich and identifying medication dosage from a bottle.	Clinic	Items scored from 0 to 2. Total score ranges from 0 to 84.	Not provided	Also known as OTAPS.
Performance Test of Activities of Daily Living (PADL) (Kuriansky, 1976)	16 individual tasks; item examples include lifting food onto spoon and to mouth, combing hair, and drinking from a cup.	Clinic	Scored 0 (completes task with minimal assistance) to 1 (unable to complete on own).	Not provided	
St. George Hospital Memory Disorders Clinic Occupational Therapy Assessment Scale (OTAS) (Fairbrother, 1997)	8 domains; 30 individual items. Example items include using a telephone book and offering correct money in a simulated shopping activity.	Home	Items scored from 0 to 3. Total score ranges from 0 to 90. Score is based on time and the number of prompts.	30–40 minutes	
Older Healthy Individuals Activities of Daily Living Test (ADL-T) (Warren, 1989)	4 tasks; item examples include boiling an egg and writing a check.	Home and Clinic	Scores ranged from 0 to 16.	>90 minutes	
Day out Task (DOT; Schmitter-Edgecombe et al., 2012)	8 subtasks; laboratory apartment; participants told to imagine they are planning for a day out (beginning with meeting a friend at museum at 10 am and later traveling to friends house for dinner), the 8 tasks must be completed to prepare for the day; participants are told to multitask.	Laboratory	Accuracy of subtasks (i.e., complete/efficient; complete/inefficient; incomplete/inaccurate; never attempted), overall task accuracy, and task sequencing.	No specific time, but average completion times between 10 and 14 minutes.	Used in mild cognitive impairment (MCI) and healthy older adult controls.
Everyday Problems Test (EPT) (Willis, 1993)	7 domains; 84 individual items. Participants are shown printed material of everyday items. Examples include ordering merchandise from a catalog and determining amount to pay from phone bill.	Clinic	Scoring available in published test.	<60 minutes	Test sold by publisher. There is also an Everyday Problems Test for Cognitively Challenged Elderly (EPCCE).
Observed Tasks of Daily Living (OTDL) (Diehl, 1995)	3 domains; 31 individual items (21 usable items). Example items include baking a cake in a microwave oven and loading a pill reminder.	Home	Scoring available from author.	90 minutes	

Note: Tests preceded by a < or a > indicate approximation of administration time based on the text provided. Readers are encouraged to examine the literature for additional studies pertinent to their clinical or research question that may apply their candidate measures.

Modified from Moore DJ, Palmer BW, Patterson TL, Jeste DV. A review of performance-based measures of functional living skills. *J Psychiatr Res*. 2007;41(1–2):97–118. doi: 10.1016/j.jpsychires.2005.10.008.

Some assessments focus on capacity (what the person can do), while others focus on performance (what the person actually does). Glass describes three “tenses” of functioning: (1) a *hypothetical* tense representing the capacity of the person, (2) an *experimental* tense defined as what the person is capable of doing, and (3) *enacted* tense which represents what the patient actually does. Furthermore, Glass describes common errors in classifying older adults’ functional status. The first occurs when people provide a low self-report of functioning but actually function at a higher level; these people are “overachievers,” and their reports as compared to behavior represent a discrepancy between the hypothetical and the enacted tense. The opposite situation also occurs:

some people report high functioning but engage in few activities that would qualify for that designation. These people can be termed “underachievers” and also show a discrepancy between the hypothetical and the enacted tense.

When interpreting functional assessment data, the clinician or researcher must consider whether the person being assessed has ever done a given task before. For example, performing a financial assessment with a person who has never managed their own finances has little to no meaning.

Table 55.3a–4.
Strengths and Weaknesses of Various Functional Assessment Modalities

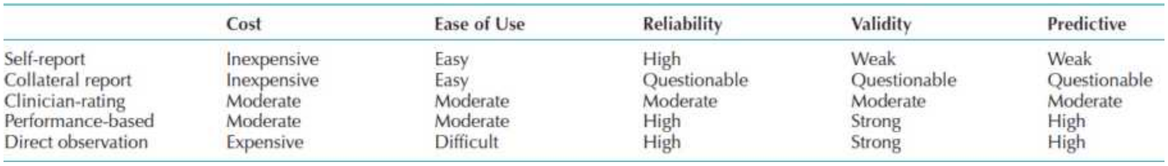

	Cost	Ease of Use	Reliability	Validity	Predictive
Self-report	Inexpensive	Easy	High	Weak	Weak
Collateral report	Inexpensive	Easy	Questionable	Questionable	Questionable
Clinician-rating	Moderate	Moderate	Moderate	Moderate	Moderate
Performance-based	Moderate	Moderate	High	Strong	High
Direct observation	Expensive	Difficult	High	Strong	High

There are many reasons for conducting a functional assessment. In many cases, functional assessments are used to get a “snapshot” of an individual’s current level of functioning. Functional assessments are rarely done as part of routine care; rather, they are often conducted following some acute event (e.g., stroke) or referral complaint (e.g., the patient has difficulty managing medications). An evaluation in this context serves as a baseline for future evaluations and, ideally, would be sensitive to improvement or worsening of abilities over time. Functional strengths and weaknesses can be identified, and interventions can aim to improve areas of disability and utilize areas of intact functioning to compensate for deficits. In the optimal situation, improvements would be observed following a treatment intervention to improve the targeted area of functional impairment. It is important to place particular emphasis on those abilities that are critical for functioning within a given environment. For instance, medication management may not be important for an individual living in a structured living facility that administers medications, but bathing, grooming, and the ability to transfer remain important. In this way, a functional assessment can be used as a tool for discharge planning and estimating the level of functional assistance that a particular individual may need.

Often a person receiving a functional assessment has a significant preexisting cognitive or physical disability, but people with more subtle difficulties may be even more appropriate candidates for a functional assessment. For example, a person with AD is likely to have many functional impairments, and decisions regarding housing placement and the level of care may have been determined prior to the administration of a functional assessment. On the other hand, a person who is having mild cognitive difficulties and who evidenced impairments on tests of arithmetic and memory may be an ideal candidate for a financial capacity assessment (and subsequent treatment) if they are still managing personal finances. The authors argue that functional assessments should be used at earlier stages of clinical care to facilitate gradual life changes (e.g., enlisting a family member to help with finances) rather than wait for a radical upheaval of all aspects of daily life at once.

Everyday functioning assessments are not just for individual-level clinical purposes. These assessments can be used with representative samples of clinical populations to draw general demographic conclusions. For instance, research using the Social Skills Performance Assessment (SSPA) has shown that older schizophrenia patients have difficulties with social skills. As a result, Cognitive-Behavioral Social Skills Training was developed and has targeted these problems among people with schizophrenia and improved the functioning of this group.

There are several different methods for assessing everyday functioning. The most common are clinical interviews, self- or collateral reports, clinician ratings of behavior, laboratory performance-based measures, and direct observation. Several reviews in the literature provide additional information about the everyday functioning assessment instruments designed for older adults.

Each assessment method has inherent strengths and weaknesses (Table 55.3a–4). For example, some types of instruments are easily administered, but may not be reliable (e.g., self-report), whereas others may provide more reliable data but are not feasible to administer routinely (e.g., direct observation). Choosing the type of functional assessment to use is largely driven by the referral question, available resources, and the person’s particular needs. In the sections below, common functional assessment methods are reviewed, and some of the instruments frequently used with each type are discussed. Of course, the gold standard consists of a multimodal approach.

Clinical Interview of Everyday Functioning

Though time is often at a premium when psychiatrists and other physicians conduct clinical interviews, a detailed clinical review of a patient's daily functioning can be extremely helpful in establishing current levels of functioning as well as for treatment planning. Understandably, a thorough assessment of everyday functioning cannot be conducted at every clinical visit; however, psychiatrists can make referrals to other disciplines (e.g., physical and occupational therapy, clinical psychology, neuropsychology) that may be able to spend more time both conducting a clinical interview as well as other assessment modalities described below.

Certain elements are essential to a thorough clinical interview designed to assess a person's daily functioning. First, it is important to get a sense of the person's daily activities. Open-ended questions such as, "Can you describe a typical day for me?" are often helpful. This puts the onus on the patient to elaborate on their daily activities. Queries should be made about difficulties with ADLs and IADLs. All assessments should include questions about other important domains such as social functioning, driving, and safety. Table 55.3a–5 lists some of the critical elements of an everyday functioning interview. Similarly, follow-up questions pertinent to any daily functioning domain of concern should be phrased in ways that give the patient the opportunity to elaborate on any possible concerns or difficulties. It is sometimes helpful to disarm patients by saying things like, "All people have difficulties managing their medications. Can you tell me about the last time that you had difficulties managing yours?"

Table 55.3a–5.
Topics to Cover in Clinical Interview of Everyday Functioning

For All Older Adults

- Current daily activities?
- Difficulties with:
  - Physical demands (e.g., climbing stairs, walking)
  - ADLs (e.g., bathing, dressing, feeding)?
  - IADLs (e.g., cooking, medication/financial management, driving)?
  - Social skills?
  - Safety (e.g., two-story house for frail person)?
  - Decision making?

Additional questions for persons with difficulties

- Experience with daily activities of difficulty?
- Onset (e.g., gradual/discrete)?
- Progression (e.g., continue to decline)?
- Decline associated with critical event (medical, change in medication, grief)?
- Availability, reliability, and commitment of support declines in daily functioning?
- Concrete plan for dealing with daily functioning difficulties?

It is also important to elicit patients’ histories with particular daily tasks. For instance, some older individuals may have little experience cooking and preparing meals, while other older individuals may have little experience managing finances. If a patient has recently lost a significant other or a support system that used to perform a particular task, poor performance may not represent a decline in daily functioning ability but rather a lack of experience. Some assessment performances, however, may represent declines, and it is useful to pose questions such as the following: Was the onset of these difficulties gradual or sudden? Has there been a continued decline? Was the decline associated with any other recent change (e.g., medical problem, change in medications)?

Among patients exhibiting daily functioning difficulties, it is important to find out whether they have support for daily tasks. Clinicians should assess the availability, commitment, and feasibility of the proposed support system. Where possible, clinicians should get consent from the patient to follow up with the identified support person or network to determine if the person or people identified are truly capable of providing support in performing daily tasks.

Self-Report of Everyday Functioning

Self-report is the most widely used modality for assessing problems with daily functioning. Self-reports are limited by the knowledge and objectivity of the respondent. Although self-report is subject to inaccuracies, particularly overestimation of functional abilities, it has several advantages, including cost-effectiveness, ease of administration, and portability. Furthermore, in people without significant psychopathology (e.g., depression, personality disorders) or health problems (e.g., chronic pain, fatigue), self-endorsement of functional difficulties is generally held to equate to actual difficulties; however, the opposite does not always prove to be true. That is, people often fail to report or minimize difficulties that they may be having. Among older people with mild neuropsychological impairment or dementia, self-reports of functional status may have little to no utility, especially among patients with amnestic syndromes. Even among older individuals without cognitive problems, overestimation of abilities is common.

Another disadvantage of most self-report measures is that they are not particularly sensitive to change over time. Individuals tend to hold fairly constant views of themselves regardless of recent personal events and tend not to report changes in daily functioning abilities until they become drastic. Self-report measures appear to elicit more accurate data and correlate more robustly with performance-based measures when queries are specific and pertinent to the individual’s difficulties (e.g., about stair climbing) rather than abstract (e.g., about mobility in general).

Self-report data concerning functional abilities can be collected in several different ways, including structured, semistructured, or unstructured interviews. Interviews allow clinicians to hear, in the patient’s own words, the difficulties that they are having. Clinicians can then ask follow-up questions that help specify the particular difficulty or difficulties. This is the approach most commonly used by general practitioners. Where indicated, problem areas can be addressed with more formal assessments.

Table 55.3a–6.
Self-Report Measures of Everyday Functioning

Self-Report Measures of Functional Status
Lawton–Brody IADL
Katz ADL
Functional Status Questionnaire
Health Assessment Questionnaire
World Health Organization Questionnaire
Short-Form Health Survey (SF-36)
Sickness Impact Profile (SIP)
Functional Disability Index Score (FDIS)
Older American Resources and Services (OARS)
Patients Assessment of Own Functioning Inventory (PAOFI)
Rapid Disability Rating Scale–2 (RDRS-2)
Questionnaire on Clinical and Preclinical Disability (QCPD)
Unmet Needs Indices (UNI)
Identification of Seniors at Risk (ISAR)
Instrumental Activities of Daily Living Compensation (IADL-C) Scale
Everyday Competence Questionnaire (ECQ)

Another way of assessing functional status is through the use of standardized self-report questionnaires. These instruments vary in their levels of reported reliability and validity. The outcome is generally a score that represents the level of functional impairment. Some of the most widely used instruments of self-reported daily functioning are the Katz ADL scale, the Lawton–Brody IADL scale, and the FAQ. Additional self-report assessment instruments are listed in Table 55.3a–6.

Katz ADL Index. The Katz ADL Index uses a three-point Likert-type scale to assess the ability to bathe, dress, toilet, transfer, feed oneself, and maintain bowel and bladder continence. It was designed for use primarily with older hospitalized patients who function at lower levels, and hence it is unlikely to identify functioning difficulties among most community-dwelling outpatients. Another limitation is the scale's hierarchical structure. Functional difficulties are assumed to accrue in this order: bathing, dressing, toileting, transferring, continence, and feeding.

Research has shown that alternative hierarchies of functions may apply equally as well.

Lawton–Brody Instrumental ADLs. The Lawton–Brody IADL scale is another structured questionnaire that focuses on higher-level domains needed for independent functioning at home and in the community. The scale covers eight domains: telephone use, shopping, food preparation, housekeeping, laundry, transportation, medication management, and financial management. This scale is often used in conjunction with the Katz scale to provide indicators of both basic ADLs and IADLs. Although the questionnaire is often used in self-report mode, the clinician or caregiver can also complete it.

Functional Activities Questionnaire. Another self-report instrument of everyday functioning is the FAQ. The FAQ is a 10-item questionnaire that is widely used in conjunction with neuropsychological testing. This questionnaire also has an informant form that can be administered in addition to the self-report form. Respondents rate the patient's ability to carry out IADLs, such as preparing balanced meals and managing personal finances, on a scale ranging from no difficulty to dependent. Ratings account for whether the patient has ever done the activity and their current abilities (e.g., never did [the activity] but could do now = normal). The main focus is on IADLs that require higher cognitive abilities that are observed in the earlier stages of dementia and not more basic ADLs that typically decline later in the dementing process.

Collateral Report of Everyday Functioning

Reports from caregivers, family members, or other people who are close to the individual with everyday functioning difficulties can be extremely valuable. In situations where the collateral reporter knows the patient well, is capable of providing an accurate report, and is able to maintain objectivity, collateral reports can be one of the best ways of gathering information regarding a patient's functional status. Accurate collateral reporters provide the same types of information as a self-report assessment, but they do so from the perspective of a direct observer. Unlike direct clinical observers, collateral reporters are typically part of the individual's daily routine, which obviates some of the problems common to traditional direct observation (see below). Unfortunately, collateral reports also have limitations. Close family members or friends may have reasons for distorting their reports. For example, they may minimize the level of functional disability in an effort to protect the patient from unpleasant but inevitable transitions, or they may overemphasize the level of impairment in order to obtain medical or psychosocial assistance for the patient more quickly. The collateral reporters may also have cognitive or psychiatric impairments of their own that limit their ability to render accurate judgments. Finally, in many cases, older patients are socially isolated such that a knowledgeable informant is unavailable. Where collateral reports are available, the clinician must decide how likely they are to be accurate. Many of the self-report measures listed in Table 55.3a–6 also have collateral-report versions. Additionally, the Physical Self-Maintenance Scale (PSMS) was specifically designed for use by a person with significant knowledge of the skills of the patient, and the Specific Levels of Functioning (SLOF) was designed for caretakers (detailed below).

Specific Levels of Functioning. Another measure is the SLOF, which is a 43-item caretaker report of a patient's behavior and functioning in the following domains: physical functioning (e.g., vision, hearing), personal care (e.g., toileting, eating, grooming), interpersonal relationships (e.g., initiating, accepting, and maintaining social contacts; effectively communicating), social acceptability (e.g., verbal and physical abuse, repetitive behaviors), activities (e.g., shopping, using telephone, paying bills, use of leisure time, use of public transportation), and work skills (e.g., employable skills, level of supervision, punctuality). The SLOF rates the amount of assistance that the patient requires to perform real-world tasks (personal care, activities), and it rates levels of physical functioning and the frequency of certain behaviors (interpersonal relationships, social acceptability, work skills). An important aspect of collateral reports, such as the SLOF, is how well the caretaker or informant knows the patient. Performance-based measures of everyday functioning discussed below have been shown to predict real-world outcomes as determined by the SLOF.

Ecologic Momentary Assessment. Assessing everyday functioning in real-time has several potential advantages. Ambulatory approaches such as ecologic momentary assessment (EMA) and experience sampling methods are growing in research settings to capture everyday phenomena. Such approaches involve sending individuals surveys via electronic methods (e.g., text messaging, email) on a set schedule, often several times a day for several days. Such an approach allows for a more nuanced and ecologically valid assessment of the performance of daily activities. Potential impeding factors can also be assessed using this method (e.g., cognitive impairments). Other advantages of EMA approaches are that they are less sensitive to recall bias (which may be particularly poor among those with memory impairments), as well as greater sensitivity of these methods to detect changes over time. Though EMA and other approaches offer potential advantages, implementing them in a clinical setting has yet to be realized. With the continued rise of telemedicine in clinical settings, the outlook of EMA for evaluating daily functioning is promising.

Clinician Rating of Everyday Functioning

Another common functional assessment approach is the clinician rating. Such an assessment is typically conducted in a clinic or hospital and can be administered by a physician, nurse, occupational or physical therapist, or a specifically trained rater. In contrast to self- or collateral reports, the clinician rating has the advantage that the clinician is generally familiar with specific daily functioning difficulties and the particular behaviors that may put individuals at risk for poor daily functioning. Many clinician-rated instruments have reported good interrater reliability, although it is likely that experienced clinicians will produce more accurate ratings and classifications. These instruments can measure both actual behavior and capacity. For instance, some clinician instruments measure what they see patients actually doing, and others ask patients to conduct a given behavior (capacity). A disadvantage of clinician ratings is that they are typically conducted outside of the patient's habitual environment (i.e., clinic or hospital) rather than in the patient's home or some other everyday setting. Therefore, they may not generalize to the patient's everyday setting (particularly if the patient still lives independently).

Barthel Index. One of the most widely used clinician-rated scales is the Barthel Index. To complete this and other similar instruments (see Table 55.3a–5), the clinician rates several different ADLs on a Likert-type scale that ranges from dependent to independent. The Barthel Index assesses more basic functions such as feeding, bathing, grooming, dressing, bowel movements, bladder control, toilet use, transfers from bed to chair and back, mobility, and stair use. Items on the index mostly range from 0 to 10 (some from 0 to 5) and have behavioral anchors for midpoints (e.g., for the transfers item, a rating of 5 indicates major help from 1 or 2 people). The Barthel Index and the Katz ADL (which assesses bathing, dressing, toileting, transfers, continence, and feeding) are most appropriate for individuals who have significant physical or cognitive impairments.

Functional Independence Measure. The FIM is a clinician-rated assessment that is widely used within rehabilitation settings for the determination of functional dependence/independence. It is important to note that the FIM rates observed performance, not capacity; this limits the measure's usefulness in inpatient or other settings where the clinician has frequent observation of the patient. Each of the FIM's 18 items is rated on a 7-point Likert-like scale that ranges from total dependence to functional independence. The scale requires trained raters (with at least 80% accuracy), and certification is provided for a fee by the Uniform Data System for Medical Rehabilitation (UDS). The FIM features both a motor and a cognitive component. The items covered are listed in Table 55.3a–7.

The Functional Assessment Measure (FAM) was created to supplement the FIM and includes 12 items related to cognitive, behavioral, communication, and community functioning. These items are also listed in Table 55.3a–7. The FIM + FAM comprises a total of 30 items and is often used to evaluate change over the course of an inpatient stay. Both the FIM and the FAM were designed primarily for people with brain injury, but they are also widely used in other settings that provide treatment for older adults.

Table 55.3a–7.
Everyday Functioning Skills Assessed by the Functional Independence Measure and the Functional Assessment Measure

FIM	FAM
Eating	Swallowing
Grooming	Car transfer
Bathing	Community access
Dressing upper body	Reading
Dressing lower body	Writing
Toileting	Speech intelligibility
Bladder management	Emotional status
Bowel management	Adjustment to limitations
Bed, chair, wheelchair transfer	Employability
Toilet transfer	Orientation
Tub and shower transfer	Attention
Walking/wheelchair locomotion	Safety judgment
Stairs	
Comprehension	
Expression	
Social interaction	
Problem solving	
Memory	

Performance-Based Measures of Everyday Functioning

Performance-based measures of everyday functioning use role play and mock settings to simulate traditional everyday functioning tasks. These tasks are administered under consistent conditions and use standardized instructions. For instance, rather than ask a patient or a collateral informant about the patient’s cooking skills and safety (which is limited by the insight, candor, and objectivity of the patient or the informant) or observe the patient at home and watch them prepare a meal (which may be too time- or labor-intensive), a performance-based measure will have the patient prepare a meal using a mock kitchen based in a laboratory or hospital. Performance-based measures assess capacity or what a person “can do” and represent the enacted tense of the Glass model under standardized, optimal conditions rather than in the patient’s habitual environment. This distinction is important because in some disorders, what impairs performance is not a lack of capacity but rather a failure to translate functional capacity into self-initiated functional performance.

It should be noted that the distinction between clinician ratings and performance-based measures of functioning is fundamental. Performance-based measures typically use a mock setting and are scored similarly to a neuropsychological test, with little ambiguity in criteria. By contrast, clinician ratings occur in less-standardized settings (hospital or clinic), are generally geared toward more basic ADLs, and typically require a more experienced or highly trained rater to score appropriately.

A variety of performance-based measures have appeared over the past 30 years in discipline-specific journals (e.g., occupational therapy, clinical and educational psychology, psychiatry, neurology). The authors conducted a review of these measures. Table 55.3a–8, adapted from the authors’ review, shows the functional domains assessed by commonly used performance-based measures. The authors present measures that assess multiple domains as well as some that focus on a specific domain. Tables 55.3a–3 and 55.3.a–9 briefly describe these measures and provide information on scoring and approximate assessment times (more details in Moore et al., 2007). In addition to these tables, which primarily show assessments of traditional ADLs and IADLs, Table 55.3a–10 shows some common performance-based measures of physical functioning and simple motor skills. As there is no specific gold standard measure, a broad representation is presented in the tables.

Performance-based measures of everyday functioning, like all assessment tools, have strengths and weaknesses. The strengths are many. Foremost are the standardized methods of administration, which leave little ambiguity in the scoring. As a result, interexaminer reliability tends to be high. Additionally, these assessments focus on real-life skills in a safe setting (e.g., cooking tasks can be controlled) and are face valid. Furthermore, they do not rely on the patient’s insight or ability to answer questions on a self-report questionnaire which might be influenced by education or reading level, and they do not depend on the knowledge or objectivity of a collateral or proxy reporter. These instruments are good targets for intervention studies designed to improve everyday functioning and have been implemented in this role (e.g., the UCSD Performance-based Skills Assessment, UPSA). The UPSA is a real-world outcome measure for the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) trial, which is designed to assess the effectiveness of antipsychotic medications in improving cognition. The consensus appears to be that performance-based measures have more strengths than weaknesses and offer benefits over other assessment approaches that make them useful for clinical treatment studies.

However, performance-based measures do have weaknesses. One is that the test environment is contrived (e.g., it is not the patient's own kitchen but a mock one). This decreases the patient's ability to follow behavioral scripts or schemas that allow them to successfully complete a task at home but not elsewhere. Performance-based assessments may achieve high reliability at the cost of validity. Perhaps the greatest weakness of performance-based measures of functional capacity, however, is that they measure capacity, not actual performance. For instance, a person may very well have the capacity to complete the task but may not have the initiative or supportive environment in order to complete the task on his or her own. Finally, most of the current performance-based measures are designed to only assess one given functional domain at a time, whereas in daily life, a person is often completing many functional tasks simultaneously.

While there are far more single- and/or multiple-domain tests that evaluate each separately, there have been some multitasking tests developed. For example, the Everyday Multitasking Test (MT) is a laboratory-based multitasking qualitative assessment in which individuals are required to complete four separate tasks within a 12-minute time limit. The four tasks include cooking a meal, advanced finances (e.g., paying bills), medication management (e.g., pill dispensing), and telephone communication. Completion of these unstructured tasks required the integration of multiple cognitive systems, which is reflective of everyday functioning environments.

Virtual Reality Performance-Based Measures of Everyday Functioning

While virtual reality (VR)-based assessment has been used to assess driving ability, there are several VR, performance-based everyday functioning assessments that have been developed recently that show promise for geriatric everyday functioning assessment. These VR tasks also assess multitasking aspects of everyday functioning and may be particularly sensitive to executive dysfunction accompanying age-related neurodegenerative disorders involving frontal pathology (e.g., planning, task switching).

Table 55.3a–8.
Domains Assessed by Several Performance-Based Batteries of Everyday Functioning

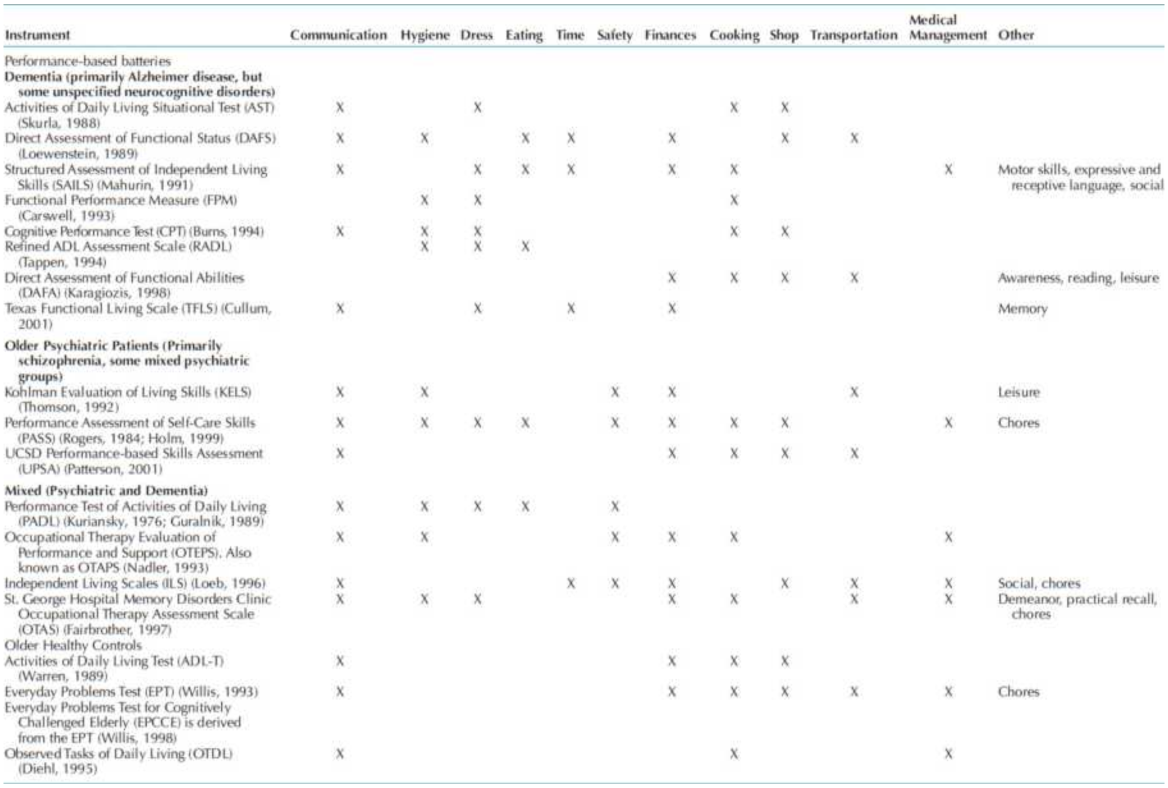

Instrument	Communication	Hygiene	Dress	Eating	Time	Safety	Finances	Cooking	Shop	Transportation	Medical Management	Other
Performance-based batteries												
Dementia (primarily Alzheimer disease, but some unspecified neurocognitive disorders)												
Activities of Daily Living Situational Test (AST) (Skurla, 1988)	X		X					X	X			
Direct Assessment of Functional Status (DAFS) (Loewenstein, 1989)	X	X		X	X		X		X	X		
Structured Assessment of Independent Living Skills (SAILS) (Mahurin, 1991)	X		X	X	X		X	X			X	Motor skills, expressive and receptive language, social
Functional Performance Measure (FPM) (Carswell, 1993)		X	X					X				
Cognitive Performance Test (CPT) (Burns, 1994)	X	X	X					X	X			
Refined ADL Assessment Scale (RADL) (Tappen, 1994)		X	X	X								
Direct Assessment of Functional Abilities (DAFA) (Karagiozis, 1998)							X	X	X	X		Awareness, reading, leisure
Texas Functional Living Scale (TFLS) (Cullum, 2001)	X		X		X		X					Memory
Older Psychiatric Patients (Primarily schizophrenia, some mixed psychiatric groups)												
Kohlman Evaluation of Living Skills (KELS) (Thomson, 1992)	X	X				X	X			X		Leisure
Performance Assessment of Self-Care Skills (PASS) (Rogers, 1984; Holm, 1999)	X	X	X	X		X	X	X	X		X	Chores
UCSD Performance-based Skills Assessment (UPSA) (Patterson, 2001)	X						X	X	X	X		
Mixed (Psychiatric and Dementia)												
Performance Test of Activities of Daily Living (PADL) (Kuriansky, 1976; Guralnik, 1989)	X	X	X	X		X						
Occupational Therapy Evaluation of Performance and Support (OTEPS). Also known as OTAPS (Nadler, 1993)	X	X				X	X	X			X	
Independent Living Scales (ILS) (Loeb, 1996)	X				X	X	X		X	X	X	Social, chores
St. George Hospital Memory Disorders Clinic Occupational Therapy Assessment Scale (OTAS) (Fairbrother, 1997)	X	X	X				X	X		X	X	Demeanor, practical recall, chores
Older Healthy Controls												
Activities of Daily Living Test (ADL-T) (Warren, 1989)	X						X	X	X			
Everyday Problems Test (EPT) (Willis, 1993)	X						X	X	X	X	X	Chores
Everyday Problems Test for Cognitively Challenged Elderly (EPCCE) is derived from the EPT (Willis, 1998)												
Observed Tasks of Daily Living (OTDL) (Diehl, 1995)	X							X			X	

Table 55.3a–9.
Single Domain Performance-Based Batteries of Everyday Functioning

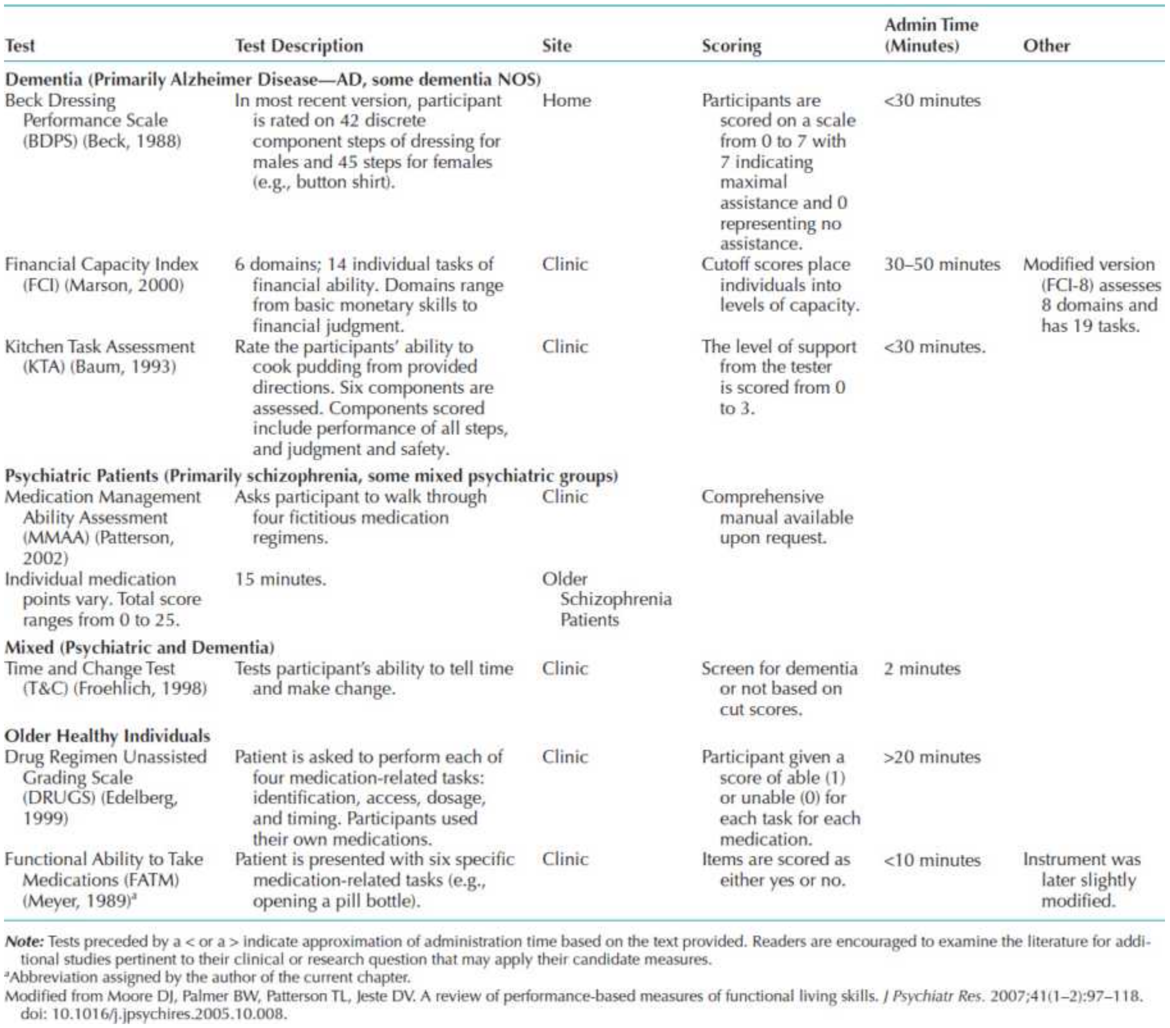

Test	Test Description	Site	Scoring	Admin Time (Minutes)	Other
Dementia (Primarily Alzheimer Disease—AD, some dementia NOS)					
Beck Dressing Performance Scale (BDPS) (Beck, 1988)	In most recent version, participant is rated on 42 discrete component steps of dressing for males and 45 steps for females (e.g., button shirt).	Home	Participants are scored on a scale from 0 to 7 with 7 indicating maximal assistance and 0 representing no assistance.	<30 minutes	
Financial Capacity Index (FCI) (Marson, 2000)	6 domains; 14 individual tasks of financial ability. Domains range from basic monetary skills to financial judgment.	Clinic	Cutoff scores place individuals into levels of capacity.	30–50 minutes	Modified version (FCI-8) assesses 8 domains and has 19 tasks.
Kitchen Task Assessment (KTA) (Baum, 1993)	Rate the participants' ability to cook pudding from provided directions. Six components are assessed. Components scored include performance of all steps, and judgment and safety.	Clinic	The level of support from the tester is scored from 0 to 3.	<30 minutes.	
Psychiatric Patients (Primarily schizophrenia, some mixed psychiatric groups)					
Medication Management Ability Assessment (MMAA) (Patterson, 2002)	Asks participant to walk through four fictitious medication regimens.	Clinic	Comprehensive manual available upon request.		
Individual medication points vary. Total score ranges from 0 to 25.	15 minutes.	Older Schizophrenia Patients			
Mixed (Psychiatric and Dementia)					
Time and Change Test (T&C) (Froehlich, 1998)	Tests participant's ability to tell time and make change.	Clinic	Screen for dementia or not based on cut scores.	2 minutes	
Older Healthy Individuals					
Drug Regimen Unassisted Grading Scale (DRUGS) (Edelberg, 1999)	Patient is asked to perform each of four medication-related tasks: identification, access, dosage, and timing. Participants used their own medications.	Clinic	Participant given a score of able (1) or unable (0) for each task for each medication.	>20 minutes	
Functional Ability to Take Medications (FATM) (Meyer, 1989)[a]	Patient is presented with six specific medication-related tasks (e.g., opening a pill bottle).	Clinic	Items are scored as either yes or no.	<10 minutes	Instrument was later slightly modified.

Note: Tests preceded by a < or a > indicate approximation of administration time based on the text provided. Readers are encouraged to examine the literature for additional studies pertinent to their clinical or research question that may apply their candidate measures.

[a] Abbreviation assigned by the author of the current chapter.

Modified from Moore DJ, Palmer BW, Patterson TL, Jeste DV. A review of performance-based measures of functional living skills. *J Psychiatr Res*. 2007;41(1–2):97–118. doi: 10.1016/j.jpsychires.2005.10.008.

The VR multiple errands test (VMET) was adapted from a non-VR version of the MET due to the need for a more feasible approach to assess multitasking and executive abilities in the context of everyday functioning. Several versions have been established, but in general, participants complete a set of diverse subtasks within a simulated shopping mall and/or office environment and include a planning component, tapping into the role of metacognition and metacognitive strategies in everyday task performance.

Table 55.3a–10.
Performance-Based Tests of Physical Functioning

Physical Performance Test (PPT)
Established Populations for Epidemiologic Studies of the Elderly (EPESE)
Timed Manual Performance Test (TMP)

Another measure is the Virtual Kitchen Challenge (VKC), a noninvasive VR meal preparation (breakfast and lunch) task that takes place in a virtual kitchen. Similar to the VMET, the VKC measures executive abilities in an everyday setting. Unlike the VMET test, tasks must be performed in a specific sequence, however, with little explicit rule constraints because the goal of this test was to explore whether and how often subjects participate in behavior that is clearly not goal-directed when there is no explicit instruction or “rules” directing against such behaviors.

Table 55.3a–11 provides a list and description of several recently developed VR assessments. Although many of these tests were not necessarily designed for older adult populations, they may prove useful in healthy and pathologic aging. The advantages of VR-based assessments are: (1) increased safety, (2) improved standardization and objectivity in terms of stimulus delivery, measurement/scoring, and error reduction, (3) ability to detect subtle deficits, and (4) portability which makes them ideal for clinical settings and has potential for remote delivery. Many of these tasks are new and are therefore still in the development and validation stages; however, they do show promise for geriatric assessment in future research and practice (Table 55.3a–11 lists the populations that have been used to date). It is important for these VR assessments that participants and patients are given a chance to familiarize themselves with the computerized interface, especially when they have no prior computer experience.

Table 55.3a–11.
Virtual Reality (VR) Performance-Based Everyday Functioning Assessments

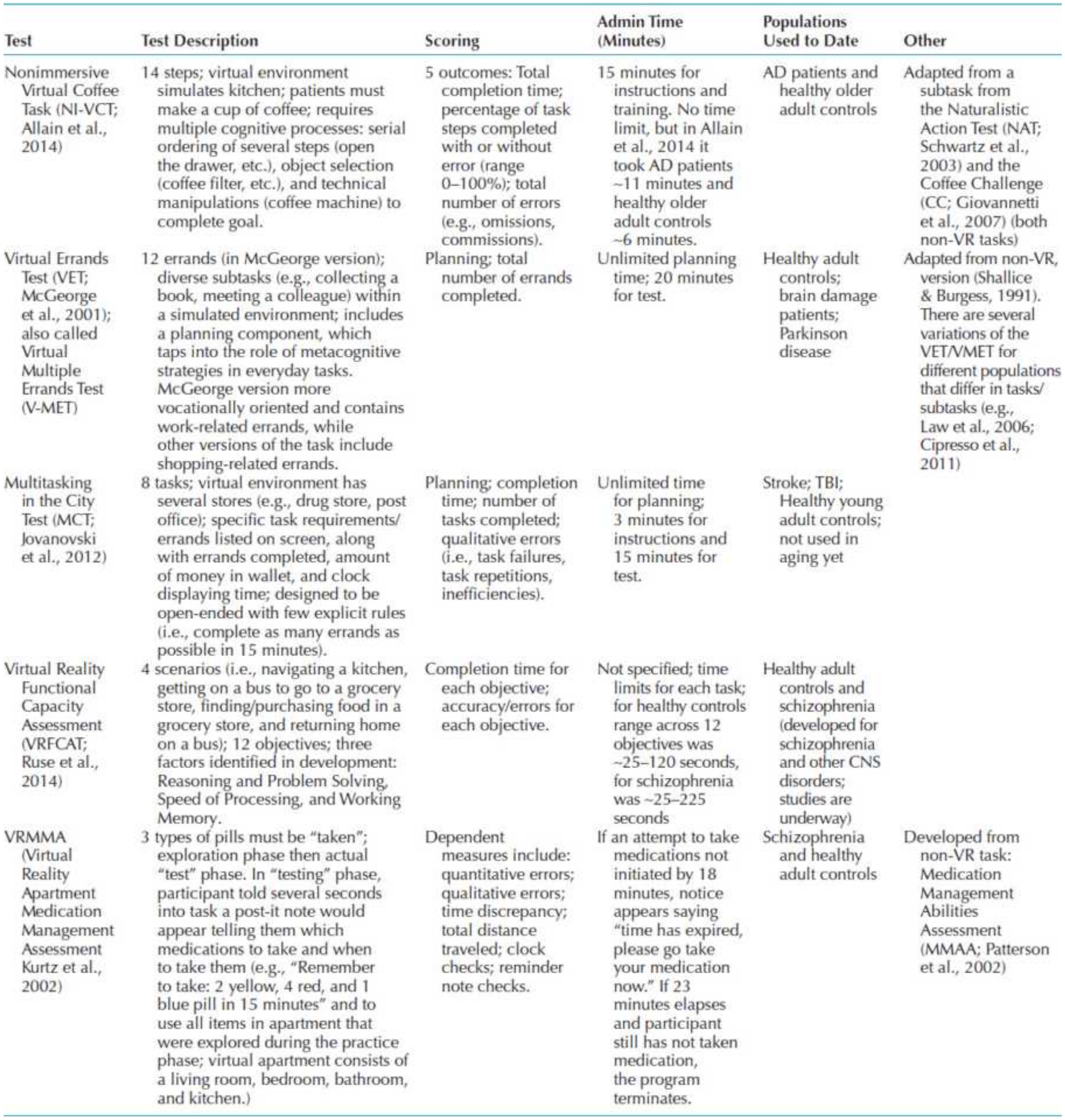

Test	Test Description	Scoring	Admin Time (Minutes)	Populations Used to Date	Other
Nonimmersive Virtual Coffee Task (NI-VCT; Allain et al., 2014)	14 steps; virtual environment simulates kitchen; patients must make a cup of coffee; requires multiple cognitive processes: serial ordering of several steps (open the drawer, etc.), object selection (coffee filter, etc.), and technical manipulations (coffee machine) to complete goal.	5 outcomes: Total completion time; percentage of task steps completed with or without error (range 0–100%); total number of errors (e.g., omissions, commissions).	15 minutes for instructions and training. No time limit, but in Allain et al., 2014 it took AD patients ~11 minutes and healthy older adult controls ~6 minutes.	AD patients and healthy older adult controls	Adapted from a subtask from the Naturalistic Action Test (NAT; Schwartz et al., 2003) and the Coffee Challenge (CC; Giovannetti et al., 2007) (both non-VR tasks)
Virtual Errands Test (VET; McGeorge et al., 2001); also called Virtual Multiple Errands Test (V-MET)	12 errands (in McGeorge version); diverse subtasks (e.g., collecting a book, meeting a colleague) within a simulated environment; includes a planning component, which taps into the role of metacognitive strategies in everyday tasks. McGeorge version more vocationally oriented and contains work-related errands, while other versions of the task include shopping-related errands.	Planning; total number of errands completed.	Unlimited planning time; 20 minutes for test.	Healthy adult controls; brain damage patients; Parkinson disease	Adapted from non-VR, version (Shallice & Burgess, 1991). There are several variations of the VET/VMET for different populations that differ in tasks/ subtasks (e.g., Law et al., 2006; Cipresso et al., 2011)
Multitasking in the City Test (MCT; Jovanovski et al., 2012)	8 tasks; virtual environment has several stores (e.g., drug store, post office); specific task requirements/ errands listed on screen, along with errands completed, amount of money in wallet, and clock displaying time; designed to be open-ended with few explicit rules (i.e., complete as many errands as possible in 15 minutes).	Planning; completion time; number of tasks completed; qualitative errors (i.e., task failures, task repetitions, inefficiencies).	Unlimited time for planning; 3 minutes for instructions and 15 minutes for test.	Stroke; TBI; Healthy young adult controls; not used in aging yet	
Virtual Reality Functional Capacity Assessment (VRFCAT; Ruse et al., 2014)	4 scenarios (i.e., navigating a kitchen, getting on a bus to go to a grocery store, finding/purchasing food in a grocery store, and returning home on a bus); 12 objectives; three factors identified in development: Reasoning and Problem Solving, Speed of Processing, and Working Memory.	Completion time for each objective; accuracy/errors for each objective.	Not specified; time limits for each task; for healthy controls range across 12 objectives was ~25–120 seconds, for schizophrenia was ~25–225 seconds	Healthy adult controls and schizophrenia (developed for schizophrenia and other CNS disorders; studies are underway)	
VRMMA (Virtual Reality Apartment Medication Management Assessment Kurtz et al., 2002)	3 types of pills must be "taken"; exploration phase then actual "test" phase. In "testing" phase, participant told several seconds into task a post-it note would appear telling them which medications to take and when to take them (e.g., "Remember to take: 2 yellow, 4 red, and 1 blue pill in 15 minutes" and to use all items in apartment that were explored during the practice phase; virtual apartment consists of a living room, bedroom, bathroom, and kitchen.)	Dependent measures include: quantitative errors; qualitative errors; time discrepancy; total distance traveled; clock checks; reminder note checks.	If an attempt to take medications not initiated by 18 minutes, notice appears saying "time has expired, please go take your medication now." If 23 minutes elapses and participant still has not taken medication, the program terminates.	Schizophrenia and healthy adult controls	Developed from non-VR task: Medication Management Abilities Assessment (MMAA; Patterson et al., 2002)

Direct Observation of Everyday Functioning

Some measures employ direct observation, which may be more valid than subjective reports. Direct observation has typically been used in inpatient settings where clinicians can easily record behaviors. Although many consider direct observation to be the "gold standard" of functional assessments, it is, in fact, more of a theoretical technique than a practical tool. That is, the most accurate representation of functional ability would consist of observing someone in her natural environment as she conducts her daily tasks without knowing that she is being observed. In most cases, direct observation of this type is unethical and unrealistic. Even where the patient's consent has been obtained for direct observation, it is labor intensive and provides data only about the behaviors engaged during the observation period. Furthermore, a person who knows he is being observed will frequently alter his behavior, sometimes radically.

CLINICAL POPULATIONS AND EVERYDAY FUNCTIONING

The following text discusses selected domains of everyday functioning that are of concern among specific clinical populations of older adults or that have received substantial research (e.g., functional impairments in older people with schizophrenia). This list of clinical populations is in no way exhaustive, but it gives a sense of the ways in which functional impairment can differ from group to group.

Normal Aging

Cognitive declines associated with normal aging seldom reach a level sufficient to impact daily functioning. Among older people without dementia or some other condition known to affect cognition, the neuropsychological domains that show the greatest decline with aging are processing speed and psychomotor speed. Declines in these domains are unlikely to significantly influence basic ADLs such as grooming, bathing, or using a telephone. Normal declines in these domains may, however, have some effect on speeded IADLs such as driving. Driving relies heavily on intact processing and psychomotor speed, and as discussed above, driving is a domain that warrants particular attention from clinicians when assessing older adults.

Dementias and Neurologic Conditions

The cognitive profiles of individuals with dementia can be quite varied. AD, the most common of the dementias, often produces difficulties with memory (particularly with delayed recall) as well as language difficulties (e.g., anomia). Difficulties with memory can lead to problems with many IADLs, such as financial and medication management. Tasks such as cooking and meal preparation are highly dependent on the order in which they are conducted, and a person with memory impairment may forget a critical step or unknowingly repeat a task.

This case presents an example of a thorough assessment of neuropsychological, mood, and daily functioning in a patient suspected of having AD.

DEMOGRAPHICS

Mrs. Smith is a 71-year-old, right-handed Caucasian woman.

EDUCATIONAL AND VOCATIONAL HISTORY

The patient completed 14 years of education. After completing 2 years of college, she accepted a position as a secretary in a law office and remained in that position until she retired at the age of 55.

BEHAVIORAL OBSERVATIONS

The patient was alert and oriented to the day of the week, year, and city but was incorrect by 2 days on the exact date. She knew the month, and she could name the president, governor, mayor, and the name of the building she was in. The patient was pleasant to work with. Her mood was positive, and her affect was congruent with her mood. She mentioned that her children were worried about her memory lately, particularly her daughter. She interacted with the examiner in an appropriate manner. She took a few breaks to stretch her legs. Her frustration tolerance was good. She showed very good attention and low distractibility. Gross visual acuity with glasses was intact, and her hearing was adequate. She was able to understand test instructions with minimal repetition. Overall, the patient seemed to put forth a good effort on all of the tests. The results are considered to be a valid estimate of her current cognitive and everyday functioning.

GENERAL COGNITIVE FUNCTIONING

Mattis DRS-2 out of a total possible score of 144:

Total raw score:	127	Scaled score: 6
Attention raw score:	36	Scaled score: 11
Initiation/perseveration raw score:	37	Scaled score: 11
Construction raw score:	6	Scaled score: 10
Conceptualization raw score:	32	Scaled score: 7
Memory raw score:	16	Scaled score: 2

VISUOSPATIAL FUNCTIONING

Performance on the Clock Drawing Test was normal, at the 50th percentile, consistent with her performance on the DRS construction score.

PSYCHOMOTOR SPEED

Trail Making Test Part A and Trail Making Test Part B were severely impaired, at the <1st percentile.

EVERYDAY FUNCTIONING ASSESSMENT

The patient was administered several measures of everyday functioning. Specifically, the patient was administered portions of the performance-based Independent Living Scale (ILS): Managing Money; the self-report Minimum Data Set (MDS) questionnaire that queries the frequency that one needs help on and the difficulty experienced when performing 20 IADLs; the Short Physical Performance Battery (SPPB); and an on-road driving assessment. On the ILS: Managing Money, her performance fell within the mildly impaired range when compared to matched peers. For example, the patient inaccurately calculated how much change she should get back if she purchased something for $1.19 and gave the clerk $5.00 (her answer was $4.79). When presented with mock check materials and specific directions for completing the check, she correctly filed out the check, except for writing the wrong date. She was able to calculate the balance in her checking account. On the self-report MDS, she reported that she only needs help “some of the time” with cooking and experienced no difficulty with any of the activities. On the SPPB, the patient scored 8 points total of 12 possible points, indicating relatively intact physical functioning: 3 out of 4 points on the balance tests (missed 1 point on holding the tandem stand for less than 10 seconds); 2 out of 4 points on the gait speed tests; and 3 out of 4 points on the chair stands. Finally, on the on-road drive, the patient’s discrete maneuvers were rated unsatisfactory within the majority of the eight domains evaluated but fully intact in the remaining domains.

CORROBORATING REPORT FROM MRS. SMITH’S DAUGHTER

The patient’s daughter stated that her mother had developed memory problems over the last 12 months. She stated that Mrs. Smith is able to remember to take her daily medications with no issues but has trouble remembering appointments and birthdays. Her daughter also reports that Mrs. Smith does require some assistance with managing finances and does not drive alone. Her daughter mentions that since Mrs. Smith has started being more diligent about using a pocket calendar to keep track of events and appointments, her memory functioning is better.

MOOD ASSESSMENT

On the Center for Epidemiologic Studies Depression Scale (CES-D) (a self-report measure of depressive symptomatology), the patient endorsed two items, indicating that she was not experiencing significant clinical depressive symptoms.

SUMMARY AND CONCLUSIONS

Results of the current evaluation indicate that the patient's overall level of cognitive functioning was overall in the moderately impaired range for her age (DRS-2 Total score = 127); however, her memory and executive functioning were severely impaired, fitting the profile of amnestic Mild Cognitive Impairment (MCI) due to AD. The patient is overall independent in most IADLs. However, her memory impairments are having at least a mild effect on some domains, such as managing finances and keeping track of appointments, which is consistent with her daughter's reporting. The etiology of her cognitive deficits is unknown; however vascular disease does not appear to be a primary contributor, as she has no medical comorbidities, vascular or otherwise (e.g., no depressive symptomology). It was recommended that she receive assistance with her complex financial and medical decisions. She could continue to manage her own medication administration with some supervision and aid (e.g., weekly pill organizer). Her driving, while not impaired at this stage, should be closely monitored. She was advised not to drive alone and only drive in familiar areas that are proximally closer to her home. Finally, her family was cautioned that the observed difficulties may continue to progress, and plans should be made now for this possibility.

The use of memory aids such as calendars, electronic reminders, and checklists may help patients to circumvent their problems for a brief period; however, as dementia progresses, cognitive difficulties tend to worsen, and compensatory strategies may be less effective. Many people in the later stages of AD become completely dependent on all their ADLs. People in the early stages of AD can appear to have few daily functioning difficulties. They may be particularly capable of maintaining appropriate over-learned social interactions (e.g., introducing themselves to a new person); however, if pressed for details about newly obtained information, they may have difficulty recalling.

Other neurodegenerative conditions can present interesting challenges for everyday functioning assessments. For example, people with Lewy body dementia can have profound impairments of visuospatial skills. They are likely to have great difficulty with higher-order tasks dependent on intact visuospatial functioning, such as driving, whereas they may have less difficulty with social skills. People with Parkinson disease may have a slightly different set of daily functioning difficulties. Parkinson disease causes particular difficulties with the initiation of motor movements and, more generally, with motor skills. Initiation problems may hinder a person from starting activities that they are cognitively capable of planning and executing. Difficulties with fine motor skills may prevent the accomplishment of basic tasks like opening medication bottles.

Psychiatric Conditions

Older people suffer from the same types of psychiatric conditions as younger people, although in some cases (e.g., depression), the incidence may increase in older adulthood. Accordingly, assessment instruments need to be designed to capture daily functioning difficulties in people with a serious mental illness. This group manifests particular deficits that may be important to evaluate. The following text reviews the difficulties with everyday functioning among these individuals, as well as some of the instruments used to assess those difficulties.

Depression. Depression, neuropsychological dysfunction, and difficulties with everyday functioning are all common in older age. Studies of older people show consistent evidence linking depression to neuropsychological dysfunction, specifically, impairments in executive function, which may relate to everyday functioning difficulties. Investigators have suggested a number of sociodemographic and clinical factors that are associated with worse cognitive performance in depressed patients, such as age, education, and symptom severity. A meta-analysis found that greater depressive symptom severity was related to poorer executive function, processing speed, and episodic memory. Given that neuropsychological dysfunction and difficulties with daily functioning can be more common among older adults, the presence of depression can exacerbate these difficulties.

Particularly among older individuals, the presence of depressive symptoms is associated with a decline in the performance of ADLs. Thirty percent to 60% of individuals who have been hospitalized for depression experience a drop in their ability to perform ADLs, including bathing, dressing, toileting, transferring, and eating. One study of 398 recently hospitalized older individuals found that nearly 40% of older patients experienced persistent depressive symptoms from hospitalization up to 3 months post discharge. This study also found that patients with mild and severe persistent depressive symptoms had a significantly higher risk of functional decline (measured by the Katz-6 ADL index score), falls, and mortality compared to patients with minimal symptoms. Patients with more depressive symptoms were distinguished from patients with minimal symptoms by factors such as more medical comorbidities, older age, and cognitive impairment. Other studies have confirmed that depressive symptoms are strongly associated with poor functional performance, and depression scores for older individuals were significantly higher for those who reported experiencing ADL decline. Among a community-based sample of older adults (aged 65 and older), neuropsychological impairment and depression appear to be two of the strongest predictors of ADL/IADL difficulties, along with age, pain difficulties, taking more than four medications, having a chronic condition, poor mental health and QoL, and BMI. Unintentional nonadherence to psychotropic medications has been shown to be associated with greater cognitive impairment. Among adults with a primary major depressive episode and severe depressive symptoms (mean Beck Depression Inventory = 34.3 [SD = 11.0]), cognitive impairment was strongly associated with impairment in IADLs (e.g., medication taking and finance handling), whereas more severe depressive symptomatology and older age were strongly associated with impairments in basic ADLs. However, patients’ satisfaction in role functioning and relationships improved with age.

Schizophrenia. Daily functioning difficulties are common among people with schizophrenia. As a result, there are several performance-based instruments to capture everyday functioning difficulties common among older people with schizophrenia that have been developed. These instruments include the UPSA, including newer computerized versions; the Medication Management Adherence Assessment (MMAA); and the SSPA. Performance on these everyday functioning measures is strongly correlated with global neuropsychological functioning. A recent study showed that among people with a schizophrenia diagnosis, working memory, sustained attention, and verbal fluency were predictors of functional capacity (UPSA-B scores), and negative symptoms were the strongest predictors of social skills performance (SSPA scores). In another study, cognitive performance in the neuropsychologically normal range was not associated with everyday functioning; however, cognitive impairment predicted functional impairments. Furthermore, negative symptoms, even lower levels, were correlated with worse social functioning. These results suggest that neuropsychological abilities and negative symptoms have differential effects on everyday functioning. These results suggest that neuropsychological abilities and social skills have differential effects on general everyday functioning skills. Consistently, neuropsychological impairment has been shown to be a stronger correlate of functional status than psychiatric symptoms among people with schizophrenia. Negative symptoms of schizophrenia (e.g., flat affect) tend to be correlated with neuropsychological impairment; however, positive symptoms are rarely associated with neuropsychological or daily functioning impairments.

Recent research indicates that scores on the UPSA are associated with “real-world” outcomes, thereby making the UPSA a potentially useful tool to aid in clinical decision making. For example, a recent study found that UPSA scores of 75 or above were predictive of patients’ ability to reside independently (e.g., alone in a home or apartment). In this same study, the UPSA was significantly better at predicting residential independence than the Mattis Dementia Rating Scale (DRS), a commonly used instrument that assesses global cognitive functioning. In another study, higher scores on the UPSA were associated with greater involvement in community responsibilities, such as working for pay, engaging in volunteer work, and attending school, among Hispanic/Latino patients with psychosis. In contrast, the DRS was not significantly associated with greater engagement in community responsibilities. These two studies may suggest that performance-based assessments of daily functioning are more proximal to community functioning than global cognitive functioning.

Recently Hospitalized Older Individuals

Unfortunately, many older people never receive an assessment of their everyday functioning abilities until they are hospitalized. Because information about how the person was functioning prior to hospitalization may be sparse, it is often difficult to determine how much change in daily functioning resulted from the hospital stay and related events. The hospitalization may result directly from a recent decline in functional abilities or from some unrelated cause (e.g., a need for hip replacement or coronary bypass surgery). Regardless of the reason, ADLs among older adult patients are more likely to deteriorate and less likely to improve after hospitalization. Specifically, it has been shown that 30% to 60% of older individuals who have had a recent hospital stay acquire at least one new dependency in daily functioning. A recent study showed that the variables most often associated with a decline in ADL among older patients who have been hospitalized are age, pain, medication use, and body mass index; and the factors highly associated with a decline in IADL are being separated or divorced, living with others (nonspouse), and self-rated memory. Few screening instruments have been developed to measure the risk of functional decline among older adult individuals who have been recently hospitalized; however, a recent review of the literature identified three useful screening instruments: Hospital Admission Risk Profile (HARP), Identification of Seniors at Risk (ISAR), and Care Complexity Prediction Instrument (COMPRI).

People Living With HIV

Neurocognitive impairment commonly occurs in HIV infection, particularly among older individuals living with HIV. Neurocognitive problems may stem from the disease itself, HIV medication side effects, comorbidities associated with HIV (e.g., stroke, diabetes) or a combination of factors. The severity of the HIV-associated neurocognitive disorders (HAND) varies from asymptomatic forms that do not interfere with daily functioning to more moderate and severe forms that interfere moderately or significantly with everyday functioning. Specifically, neurocognitive impairment in HIV has been associated with driving, medication management abilities, and other IADLs. Many of the daily functioning assessments used with older HIV-infected people, such as UPSA and Lawton & Brody ADL Scale, were originally developed in non–HIV-infected populations. While translating these assessments to the older HIV-positive population may be appropriate in terms of measuring functioning, the etiology of the everyday functioning impairment should be considered in the context of the unique risk factors in this heterogeneous population.

DEMOGRAPHICS

Mr. Anderson is a 62-year-old, right-handed African American male.

EDUCATIONAL AND VOCATIONAL HISTORY

The patient completed 13 years of education and received his high school diploma. He denied having any problems learning to read or write and never skipped or repeated a grade. He worked as a cook in a restaurant for approximately 30 years.

BEHAVIORAL OBSERVATIONS

The patient was alert and oriented to the day of the week, date, year, and city. The patient was diagnosed with major depressive disorder in his early 20s and was diagnosed with HIV infection at age 50. He has been worried for some time that he may have some neurocognitive difficulties. He complains of slowed ability to process information, difficulty learning new strategies, and difficulty planning and initiating new activities. During the assessment, the patient was apathetic and appeared somewhat lethargic, but was pleasant and cooperative. Gross visual acuity and hearing were intact. Overall, the patient seemed to put forth very good effort on all of the tests. The results are considered to be a valid estimate of his current cognitive and everyday functioning.

GENERAL COGNITIVE FUNCTIONING

The patient was administered a comprehensive neuropsychological test battery. Using T-scores adjusted for age, education, sex, and ethnicity, Mr. Anderson was found to be impaired in the domains of learning and executive functioning. His performance was within normal limits on verbal functioning, speed of information processing, recall, motor, and attention/working memory. His global neuropsychological performance is mildly impaired (impairment ≥2 domains).

EVERYDAY FUNCTIONING ASSESSMENT

The patient was administered the UCSD Performance-based Skills Assessment (UPSA), which assesses the overall ability to manage everyday functional tasks, including the ability to plan and organize events, manage finances, communicate with others, arrange transportation needs, and conduct household chores. Overall, his performance fell within the moderately impaired range (overall score =

63). His UPSA score suggests some reduced capacity in maintaining his independent living situation.

CORROBORATING REPORT FROM MR. ANDERSON'S THERAPIST

The patient's therapist stated that the patient had experienced some mild difficulty performing day-to-day tasks, including managing his own medications and finances. Specifically, the therapist stated that on occasion, the patient has reported that he sometimes has difficulty organizing and managing his medications. The therapist has increasingly become concerned with the patient's neurocognitive and daily functioning difficulties. The therapist indicated that these neurocognitive and functional impairments have been stable over the past several years and do not appear to be worsening.

MOOD ASSESSMENT

The patient scored 11 on the Hamilton Depression scale, which is indicative of mild symptoms of depression. The patient denied suicidal ideation.

SUMMARY AND CONCLUSIONS

Results of the current evaluation indicate that the patient's overall level of cognitive functioning was mildly impaired as compared to other individuals of his demographic background. The patient also demonstrated moderate impairments on a measure of everyday functional abilities (UPSA). The patient's therapist corroborated many of the patient's daily functioning difficulties. These test results suggest mild to moderate impairments in cognitive and everyday functioning. The longitudinal course of these impairments has been stable. That is, cognitive and everyday functional abilities appear to have remained stable over the past several years, which is again consistent with published literature on this topic. Based on his current level of cognitive functioning and deficits in functional skills, the patient was recommended to seek additional assistance in instrumental activities in daily living activities. It was also recommended that the patient receive functional skills training classes, which may help improve his ability to perform essential everyday skills (e.g., manage finances, and communicate with others). Particularly important is his ability to self-administer his medications. Providers should consider additional training in managing medications as well as methods to reduce errors in administering medications (e.g., blister packs, electronic reminders, pill organizers).

RECOMMENDATIONS FOR FUTURE INSTRUMENT DEVELOPMENT

Assessments of daily functioning continue to evolve. Below, the authors offer recommendations for future development.

Everyday functioning is multifaceted; however, most assessment instruments are designed to test a given domain in isolation. For instance, the MMAA assesses medication management abilities independent of other abilities. Medication management in “real life” occurs over a much longer time span and is intertwined with other daily tasks. The ideal assessment of daily functions would test multiple domains of functioning concurrently. Some multitasking tests of this nature have been developed, but additional assessments of this type are clearly warranted.

Another desirable component of any new assessment instrument is that they are adaptive. That is, instruments should have the flexibility to adjust to the level of mastery of a particular individual. If an item was too difficult, the individual could be tested with easier items or vice-versa. A great deal of research has been done on adaptive testing, particularly with computer-administered, standardized achievement tests. Hopefully, this technology can be applied to the assessment of functional capacity.

VR technology may offer a way of creating more realistic mock environments in which multiple aspects of functioning could be tested. Assessments could even be designed to recreate a patient’s unique habitual environment, and many assessments have already been developed and are currently being tested. Some argue that older adults are too resistant to digital technologies for such technologies to be useful to this population, but the degree of resistance may be exaggerated. Furthermore, performing VR tasks may not require high levels of technical knowledge. These techniques may be useful both for assessments and interventions. These technologies have already been implemented for the treatment of certain disorders (e.g., PTSD).

Very few functional-assessment instruments provide demographically adjusted normative data, with the most important variable being age. This constitutes a significant limitation as compared to other types of assessment, such as neuropsychological tests, for which the assessor typically has an indication of where a given score falls on a continuum. Ideally, all functional assessments would have normative information. In the absence of normative data, an evaluator who is more experienced is likely to have a better sense of what a given score means in the context of other objectively gathered data. The authors understand that normative studies are not easily funded; however, the authors believe that future studies of this type should be a research priority.

In addition to demographically corrected normative data, there are significant advantages to the conorming of different assessment modalities together. If evidence could be obtained of how certain instruments work with one another, then better insight into a given patient's true difficulties would be available. For example, if a self-report, collateral report, and performance-based measure were concurrently administered to a large sample of older adults with varying diagnoses and ages, the results might help us to understand the distinct strengths and weaknesses of the various assessment modalities. It might be found that people who score below a given level of self- and collateral-reported functioning would derive no benefit from receiving a performance-based assessment. On the other hand, in people who report a relatively intact level of functioning, there might still be difficulties with performance-based assessments due to inaccuracies in self- or collateral reports. The combinations are many, but ideally, it would be helpful to know how the most widely used instruments worked together.

Finally, the discussion of norming and conorming is moot if the instruments in question have no validity for predicting actual functioning. Predictive validity is difficult to obtain without a widely recognized gold standard of assessment for everyday functioning abilities; however, the authors believe that using a multimodal assessment strategy allows for an accurate representation of daily functioning. Until reliable and valid instruments with strong predictive validity are established, there will be difficulty in establishing the true usefulness of everyday functioning assessments.

SUMMARY

Older adults are at significant risk for difficulties with everyday functioning. These difficulties can range from impairments in basic behaviors critical for independent functioning, such as toileting, to optional daily tasks, such as driving. Several factors may influence daily functioning, and cognitive impairment appears to be a strong predictor of daily functioning difficulties. Numerous everyday functioning assessment methodologies are available and include clinical interviews, self- and collateral reports, clinician ratings, performance-based measures, and direct observation. Performance-based measures may offer significant advantages such as improved reliability and a better estimate of functional capacity as compared to other assessment approaches such as self-report. More research is clearly needed to develop quick, reliable, and valid assessments of everyday functioning for older adults; the field is currently burgeoning, and improved assessments will continue to be developed in the near future.

ACKNOWLEDGMENTS

The authors would like to thank Virginia Wadley, Ph.D., for contributing to one of the case presentations in this chapter, and Thomas L. Patterson, Ph.D., for contributing to a prior version of this chapter.

Further Readings

Advinha AM, Lopes MJ, de Oliveira-Martins S. Assessment of the elderly's functional ability to manage their medication: a systematic literature review. *Int J Clin Pharm*. 2017;39(1):1–15.

Becattini-Oliveira AC, Dutra DDF, Spenciere de Oliveira Campos B, de Araujo VC, Charchat-Fichman H. A systematic review of a functional assessment tool: UCSD performance-based skill assessment (UPSA). *Psychiatry Res*. 2018;267:12–18.

Carswell A, Spiegel R. Functional assessment. In: Wilcock GK, Bucks RS, Rockwood K, eds. *Diagnosis and Management of Dementia*. Oxford University Press; 1999.

Chan ML, Eng CW, Farias ST, Gilsanz P, Whitmer RA. Cognitive and physical functioning as predictors of everyday functioning: neuropsychology/Early detection of cognitive decline with neuropsychological tests. *Alzheimer's & Dementia*. 2020;16:e044184.

Cipriani G, Danti S, Picchi L, Nuti A, Fiorino MD. Daily functioning and dementia. *Dement Neuropsychol*. 2020;14:93–102.

Connolly D, Garvey J, McKee G. Factors associated with ADL/IADL disability in community dwelling older adults in the Irish longitudinal study on ageing (TILDA). *Disabil Rehabil*. 2017;39(8):809–816.

Cornelis E, Gorus E, Van Weverbergh K, Beyer I, De Vriendt P. Convergent and concurrent validity of a report- versus performance-based evaluation of everyday functioning in the diagnosis of cognitive disorders in a geriatric population. *Int Psychogeriatr*. 2018;30(12):1837–1848.

Giovannetti T, Mis R, Hackett K, Simone SM, Ungrady MB. The goal-control model: an integrated neuropsychological framework to explain impaired performance of everyday activities. *Neuropsychology*. 2021;35(1):3–18.

Giovannetti T, Yamaguchi T, Roll E, et al. The Virtual Kitchen Challenge: preliminary data from a novel virtual reality test of mild difficulties in everyday functioning. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*. 2019;26(6):823–841.

Glass TA. Conjugating the "tenses" of function: discordance among hypothetical, experimental, and enacted function in older adults. *Gerontologist*. 1998;38(1):101–112.

Harvey PD, Strassnig MT, Silberstein J. Prediction of disability in schizophrenia: symptoms, cognition, and self-assessment. *J Exp Psychopathol*. 2019;10(3):2043808719865693.

Kizony R, Zeilig G, Krasovsky T, et al. Using virtual reality simulation to study navigation in a complex environment as a functional-cognitive task; A pilot study. *J Vestib Res*. 2017;27(1):39–47.

Laverick R, Haddow L, Daskalopoulou M, et al; Cognitive Impairment in People with HIV in the European Region (CIPHER) Study Group. Self-reported decline in everyday function, cognitive symptoms, and cognitive function in people with HIV. *J Acquir Immune Defic Syndr*. 2017;76(3):e74–e83.

Mahmood Z, Burton CZ, Vella L, Twamley EW. Neuropsychological predictors of performance-based measures of functional capacity and social skills in individuals with severe mental illness. *J Psychiatr Res*. 2018;102:201–206.

Papakostas GI. Cognitive symptoms in patients with major depressive disorder and their implications for clinical practice. *J Clin Psychiatry*. 2014;75(1):8–14.

Pashmdarfard M, Azad A. Assessment tools to evaluate activities of daily living (ADL) and instrumental activities of daily living (IADL) in older adults: a systematic review. *Med J Islam Repub Iran*. 2020;34:33.

Patterson TL, Moscona S, McKibbin CL, Davidson K, Jeste DV. Social skills performance assessment among older patients with schizophrenia. *Schizophr Res*. 2001;48(2–3):351–360.

Reichardt LA, van Seben R, Aarden JJ, et al; Hospital-ADL Study Group. Trajectories of cognitive-affective depressive symptoms in acutely hospitalized older adults: The hospital-ADL study. *J Psychosom Res*. 2019;120:66–73.

Romero-Ayuso D, Castillero-Perea Á, González P, et al. Assessment of cognitive instrumental activities of daily living: a systematic review. *Disabil Rehabil*. 2021;43(10):1342–1358.

Rycroft SS, Giovannetti T, Divers R, Hulswit J. Sensitive performance-based assessment of everyday action in older and younger adults. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*. 2018;25(2):259–276.

Schmitter-Edgecombe M, Sumida C, Cook DJ. Bridging the gap between performance-based assessment and self-reported everyday functioning: an ecological momentary assessment approach. *Clin Neuropsychol*. 2020;34(4):678–699.

Shirazi TN, Summers AC, Smith BR, et al. Concordance between self-report and performance-based measures of everyday functioning in HIV-associated neurocognitive disorders. *AIDS Behav*. 2017;21(7):2124–2134.

Stickel AM, Tarraf W, Wu B, et al. Cognition and daily functioning: results from the Hispanic Community Health Study/Study of Latinos (SOL) and Study of Latinos-

Investigation of Neurocognitive Aging (SOL-INCA). *J Alzheimers Dis*. 2020;77(3):1267–1278.

Strassnig M, Bowie C, Pinkham AE, et al. Which levels of cognitive impairments and negative symptoms are related to functional deficits in schizophrenia? *J Psychiatr Res*. 2018;104:124–129.

Sumida CA, Vo TT, Van Etten EJ, Schmitter-Edgecombe M. Medication management performance and associated cognitive correlates in healthy older adults and older adults with aMCI. *Arch Clin Neuropsychol*. 2019;34(3):290–300.

Szabo S, Merikle E, Lozano-Ortega G, Powell L, Macek T, Cline S. Assessing the relationship between performance on the University of California Performance Skills Assessment (UPSA) and outcomes in schizophrenia: a systematic review and evidence synthesis. *Schizophr Res Treatment*. 2018;2018:9075174.

Tomioka K, Shima M, Saeki K. Mental component of health-related quality of life is an independent predictor of incident functional disability among community-dwelling older people: a prospective cohort study. *Qual Life Res*. 2021;30(7):1853–1862.

Ward M, McGarrigle CA, Kenny RA. More than health: quality of life trajectories among older adults-findings from The Irish Longitudinal Study of Ageing (TILDA). *Qual Life Res*. 2019;28(2):429–439.

Wesson J, Clemson L, Brodaty H, Reppermund S. Estimating functional cognition in older adults using observational assessments of task performance in complex everyday activities: a systematic review and evaluation of measurement properties. *Neurosci Biobehav Rev*. 2016;68:335–360.

Woodbridge R, Sullivan MP, Harding E, et al. Use of the physical environment to support everyday activities for people with dementia: a systematic review. *Dementia (London)*. 2018;17(5):533–572.

55.3b Psychiatric Problems in the Medically Ill Geriatric Patient

Anna Borisovskaya, M.D., Beth Ann Labardi, M.D., Matthew Alan Schreiber, M.D., Ph.D., and Marcella Pascualy, M.D.

Since the beginning of the 20th century, a dramatic increase in life expectancy has produced an aged population of unprecedented size and longevity. According to demographic forecasts, one out of five people will be over the age of 65 in the year 2060, up from one in seven in 2012. In the United States, a 65-year-old person will typically remain independent until they reach the age of 75 and will live an additional 6.5 years once they become dependent on others.

People over the age of 65 account for twice the average number of prescription drugs, one and a half times the number of physician contacts, and three times the number of hospital-bed days. Overall, the aging population accounts for more than one-third of the total personal health expenditures in the United States. Among the top 5% of highest health care spenders who collectively make up one-half of total spending, adults over age 65 account for 40%.

Over half of the population that is 65 and older report having at least one chronic disability, and a third report having a chronic condition that is severe enough to limit their activities. For the psychiatrist, medical comorbidities and disabilities, which aggregate with age, are the most important factors distinguishing older patients from their younger counterparts.

Some of the most common disorders in the aged population include degenerative arthritis and cardiovascular disorders, mainly hypertension and heart disease. In addition, geriatric syndromes such as falls, vision and hearing loss, urinary incontinence, and impaired cognition disproportionately impact health care costs and quality of life. One recent study found that about 65% of adults over age 65 have two or more chronic diseases, and this increases to 82% in individuals over the age of 85.

Estimates of the prevalence of psychiatric disorders among the medically ill and disabled older adult vary widely across clinical settings, and reliable data on comorbidities is still sparse. For example, 30% of older adults in primary care have at least one active psychiatric diagnosis: 10% to 15% have clinically significant depressive syndromes such as major depression, dysthymia, or minor depression; 5% to 10% have dementia, and 2% to 5% have alcohol abuse or dependence or psychoses. The lowest prevalence rates for psychiatric symptoms and disorders are found among medically healthy persons living independently in the community, and prevalence increases with the severity of medical illnesses and extent of functional dependency (Fig. 55.3b–1). The relationship between functional health and psychiatric comorbidity has been most clearly established for depression, but emerging data suggest similar trends for anxiety disorders, sleep disorders, and psychoses. In addition to impairing day-to-day functioning and quality of life, depression in medically ill older adults has been associated with decreased self-care and adherence to treatment regimens for chronic medical conditions, worse medical outcomes, increased health care costs, and mortality from suicide and medical causes.

The profound changes in the demographics of chronic disease have resulted in an increased need for the psychiatric care of patients with comorbid medical and mental disorders. New models of practice, based on effective collaboration between psychiatrists and primary care providers or medical specialists, are required to address the needs of older patients. Clinical care of psychiatrically disturbed medical patients poses several special challenges for the psychiatrist. Clinical syndromes may vary in presentation from the classical disorders for which treatment guidelines are well established to situations where common disorders such as depression and anxiety are masked by medical comorbidities and functional disability in place of overt expression of emotional symptoms. Thus, one needs a working familiarity with the common medical and functional problems of the older adult and a full range of skills in psychotherapeutic and somatic treatment modalities. Despite these inherent challenges, most psychiatrically disturbed older persons can be substantially helped when approached within a systematic framework.

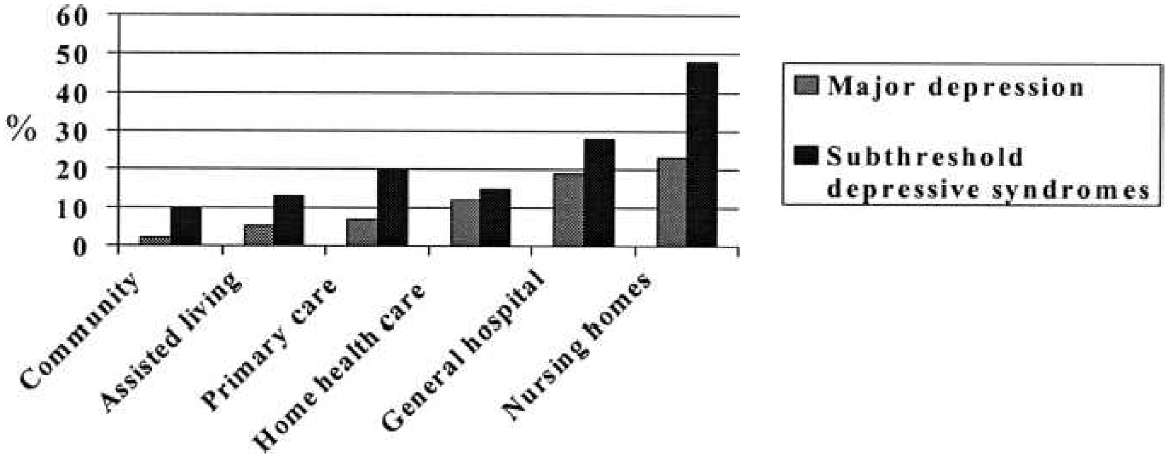

FIGURE 55.3b–1. Prevalence of depression in older adults.

APPROACH TO THE PATIENT

Mastery of a few central concepts helps integrate the range of clinical data essential in the successful evaluation and management of psychiatrically disturbed, medically ill geriatric patients (Table 55.3b–1). The adaptive reserve of all vital organs, including the brain, tends to decline with aging, even in the absence of manifesting disease. This loss of physiologic resilience is generally detectable in clinical settings only after a sudden or severe deterioration in the face of acute illness or psychosocial stress.

Age-related loss of adaptive reserve underlies the downward spiral and multiple organ failures often seen in sick older adults. It is responsible for the narrowed margin of safety observed for many drugs, changes in the presentation of disease, and increases in morbidity of medical and psychiatric disorders. For example, acute myocardial infarction in an octogenarian may present as painless delirium, while a mild chronic anxiety disorder may become fulminant panic in a patient with rectal bleeding who, ultimately, may not have a serious medical problem. Such complex patterns of causation are typical of psychiatric symptoms in the very old and in persons with multimorbidity—the most common chronic medical condition—used to define the presence of several comorbid medical conditions. Sometimes causal factors emerge clearly only after a successful intervention, or not at all.

Table 55.3b–1.
Key Concepts in Psychiatric Care of the Medically Ill Older Adult

Decreased organ reserve alters response to illness, treatment, and social stressors.
Comorbid medical illness affects presentation and treatment of psychopathology.
Nondrug treatments are preferable when effective.
Combining treatment modalities optimizes gains.
Comfort, function, and safety are the major goals of treatment.
Coordinated care is better care.
Family and caregivers are essential to patients' recovery and must be involved in the treatment.

A 78-year-old man with bipolar disorder since age 40 became manic while residing in a nursing home. His situation had been complicated by chronic alcoholism and mild dementia. He had no history of renal disease, and his serum creatinine was normal. After an initial period of manic grandiosity, when he was often seen near a local tavern, he soon deteriorated to crawling on his hands and knees, growling and barking like a dog, and striking out at all who approached him. His lithium level was 1.2 mEq/L with a maintenance dose of 750 mg/day. After a reduction in lithium dose to achieve a level of 0.6 mEq/L, the temporary addition of low-dose olanzapine, and close attention to adequate fluid intake and preventing access to alcohol, both the manic syndrome and the superimposed delirium cleared within 3 weeks without the need for psychiatric hospitalization.

Medical causes should be considered whenever an older adult patient develops a new psychiatric symptom or disorder. When psychotropic medications are used, clinicians are advised to carefully watch for the emergence of side effects, allow sufficient time for medications to reach effective plasma and receptor site concentrations, and seek optimal therapeutic outcomes. Particularly with newer-generation psychotropic medications, the differences between “geriatric” and “adult” doses are minimal; careful titration of medications from a low dose to a usual adult dose may be necessary to achieve the desired therapeutic benefit. Close collaboration with other clinicians and caregivers is essential for maximal safety and effectiveness at every stage of treatment. The greatest overall therapeutic gain, at the smallest cost in treatment-related adverse events, is usually achieved by combining psychotherapeutic, social, environmental, and somatic interventions. A rehabilitation model that specifies progress toward the attainment of specific goals in stepwise increments is a useful way to frame the ongoing process of care that is often most appropriate for chronically ill older adults.

An 83-year-old woman with long-standing diabetes and ischemic and hypertensive cardiovascular disease had become bedfast over several years in a nursing home. Two months before psychiatric consultation was requested, she developed a psychotic illness characterized by delusions of being beaten during the night, aggressive outbursts, and pervasive despondency, which led to disruptive nocturnal phone calls to her daughter and to the local fire station. Initial clinical assessment revealed moderately severe dementia in a personable but angry and frightened woman, eager for relief. Collateral history confirmed progressive cognitive deterioration over the course of her chronic medical illness, on a background of long-standing dysthymia. Trials of several antipsychotic and antidepressant medications had exacerbated her symptoms. A review of her medications showed that 4 of the 12 drugs she was taking on a regular basis were not essential for her medical management. A head computed tomography (CT) scan showed mild cerebral cortical atrophy, moderate periventricular white matter hyperintensities, and numerous lacunar infarcts. The onset of her psychosis was eventually associated with the initiation of aggressive treatment of arterial leg ulcers, including round-the-clock debridement dressings, oral antibiotics, and narcotic analgesics. Control of her diabetes had deteriorated, and her insulin dose had been increased. Nursing staff members had provided excellent physical care but were resentful of her racist attitudes and physical aggressiveness, which exacerbated her loneliness and paranoia. Her treatment plan was as follows: discontinuation of night-time debridements (allowing elimination of narcotics for pain), simplification of her drug regimen, preventing nocturnal hypoglycemia, installation of a night light in her room, problem-solving discussions with staff, weekly psychotherapy, and wheelchair walks, which represented her first opportunity to go outdoors for months. Her distressed phone calls ended soon after. Her mood, sleep, and interactions with staff improved gradually, and her delusions and racist remarks disappeared. Foci of psychotherapy were the discovery of traumas she experienced in Cossack Russia, psychoeducation about her delirium, and discussions of her desire to die rather than live in a deteriorated state. She developed a firm therapeutic alliance and regained the dignified and compassionate manner that characterized her response to adversity in the past.

PSYCHIATRIC SYNDROMES IN MEDICALLY ILL PATIENTS

General Principles

Systemic medical illness can mimic or cause almost any psychiatric syndrome, often through more than one pathophysiologic mechanism. They can also coexist with psychiatric illnesses that were present prior to the onset of such medical problems. The Diagnostic and Statistical Manual of Mental Disorders, fifth edition, text edition (DSM-5-TR) further simplified the classification of psychiatric disorders due to medical diseases, dividing them into cognitive disorders, psychotic disorders, depressive and bipolar disorders, anxiety disorders, obsessive-compulsive disorders, catatonic disorders, and personality changes. Changes from DSM-IV-TR include the addition of obsessive-compulsive and related disorders due to another medical condition, removal of amnestic disorder and sleep disorder due to a medical condition, and specifying depressive or bipolar illness due to another medical condition. Distinguishing a “medical” psychiatric syndrome from a primary psychiatric disorder is often difficult. For example, depressive symptoms in a hospitalized patient may be indicative of an adjustment disorder, demoralization, side effects of medications, delirium, withdrawal from substances, consequences of another medical illness, or major depressive disorder. At the same time, psychiatric illness may arise after serious medical illnesses such as sepsis, acute respiratory distress, myocardial infarction, and cerebrovascular accidents. Clues to the secondary nature of a psychiatric syndrome include partial expression, first onset in old age or a change in clinical features with recurrence of an earlier-onset disorder, lack of family history, a recent change in health or medical treatment, and concurrent cognitive processing deficits that point to impaired brain function.

Psychiatric disorders not caused directly by the effects of a medical illness interact with chronic medical conditions in complex ways. This is best understood for depression in the setting of chronic disease: Depression can amplify somatic symptoms, leading to excess functional impairment and overestimation of disease severity; it decreases motivation and brings about despair about the efficacy of treatment, leading to poor treatment adherence, poor self-care, and finally, poor outcomes. In such cases, consideration should be given to the diagnosis introduced in DSM-5, Psychological Factors Affecting Other Medical Conditions, denoting psychological or behavioral factors that affect a medical condition adversely. Comorbid medical illness, in turn, can delay or prevent the recognition and treatment of psychiatric disorders that might be readily treated in primary care.

Medications as a Cause of Psychiatric Symptoms

Altered pharmacokinetics and polypharmacy place older adult patients at increased risk of developing adverse drug reactions. Many drugs that are used in the older adult tend to have a low therapeutic index increasing the likelihood of toxicity (e.g., anticoagulants, hypoglycemics, antineoplastics, diuretics). Most drugs, in sufficient doses, can affect central nervous system function in a vulnerable host. Drugs affecting the brain, either as their primary site of therapeutic action (e.g., all psychotropic drugs, narcotic analgesics, anti-Parkinson agents, and many antihypertensive agents) or as a secondary site not essential to their therapeutic effect (e.g., digitalis [Lanoxin], cimetidine [Tagamet], glucocorticoids, and nonsteroidal anti-inflammatory agents, such as aspirin and indomethacin), can cause psychiatric symptoms in older adult patients. As an example, corticosteroids used for the treatment of inflammatory states such as rheumatoid arthritis may lead to depression or irritability, while thyroid replacement and even antidepressants can cause anxiety. It’s sobering to realize that adverse drug reactions are common and preventable in the older adult and that the very medications meant to cure the illness can, in fact, increase the burden of disease. Thorough knowledge of the side effects of any prescribed medication, careful monitoring for these side effects, and reduction in polypharmacy are paramount when any new medication appears on an older adult person’s medication list.

Clues to a possible drug-induced psychiatric syndrome include the presence of associated systemic or neurologic signs and symptoms, such as recent loss of appetite, reduced sense of well-being, blurred vision, tremor, slurred speech, dizziness, gait instability, or somnolence. A patient’s complaint of feeling “drugged,” while occasionally indicative of a heightened focus on somatic symptoms, should raise suspicion of drug toxicity.

Identifying a drug as the primary cause of a change in mental state can be accomplished in several ways. Some drugs produce classical symptoms in overdose, such as nausea and distorted color vision in digitalis overdose or tinnitus and dizziness with aspirin. In other instances, a neuropsychiatric syndrome results from pathophysiologic mechanisms common to a number of drugs with very different primary indications but similar pharmacologic properties. An example is the central anticholinergic syndrome, characterized in its most extreme form by tachycardia, pupillary dilation, visual hallucinations, and delirium in patients receiving medications with known high affinity for muscarinic receptors, such as tricyclic antidepressants (TCAs), but also caused by pharmacodynamic interactions of a vulnerable brain with multiple drugs, each having weak but additive individual antimuscarinic actions.

Blood levels may not be helpful in determining whether a medication is causing a neuropsychiatric problem, except for the relatively few drugs readily measured in serum or plasma and for which therapeutic ranges are well established. However, medication toxicity may occur in older adults without elevated plasma levels, and improvement, when a suspect drug is decreased or discontinued is the most crucial test of its role in pathogenesis. A detailed history, a careful physical exam, and knowledge of drug–drug interactions are most conducive to making an accurate diagnosis of toxicity.

An 87-year-old woman came to a psychiatrist with mild depression associated with a recent cross-country move and hepatic cirrhosis of unknown etiology with mild renal insufficiency. Treatment was initiated with fluoxetine 20 mg/day and supportive psychotherapy. Over the ensuing 9 months, she was hospitalized several times with intractable vomiting for which no cause was found. When she again presented with the same picture and now ascites and severe demoralization, a fluoxetine level was obtained. The combined concentration of fluoxetine + norfluoxetine was within the expected range at 138 ng/mL (normal range of 100 to 930 ng/mL). Discontinuation of fluoxetine eliminated nausea, vomiting, and demoralization within a week.

Drug and Disease Interactions

Psychiatric symptoms arising from the interactive effects of medications and underlying disease may be severe enough to lead to psychiatric hospitalization. In a series of nearly 16,000 consecutive adult psychiatric admissions, about 1% was attributable to adverse drug reactions. Older adult patients were most often affected, particularly those with neurologic diseases or those receiving multiple medications. Nearly 30% of older adult persons develop delirium after hospitalization for acute medical illness. Particularly at risk are those with cognitive impairment at the time of admission, the very old (>80 years), patients with fractures or symptomatic infection, and those receiving neuroleptics or narcotics. People with dementia may be at higher risk of toxicity of some medications due to the combined risks of weight loss, increased permeability of the blood–brain barrier, and reduced acetylcholine in the brain, among other physiologic changes, which means paying even closer attention to prescribing medications in this group. Early detection of mental disorders due to drug–disease interactions depends on familiarity with the common medical illnesses of older people, their potential impact on brain function, the pharmacologic properties of drugs used in medical management and their special potential for interactive effects. Table 55.3b–2 shows examples of some drug–drug and drug–disease interactions that may be encountered in clinical practice.

Table 55.3b–2.

Examples of Drug–Drug/Drug–Disease Interactions in the Older Adult

Drug	Indication	Adverse Effect
Cimetidine	Gastroesophageal reflux	Delirium
		Elevated drug levels (e.g., TCAs, benzodiazepines, neuroleptics, propranolol, theophylline)
Metoclopramide	Gastroparesis	Parkinsonism, inappropriate prescription of levodopa
TCAs	Depression, anxiety, insomnia, chronic pain	Confusion
Constipation		
		Tachycardia, cardiac conduction defect, hypotension
		Altered effect of antihypertensive agents (↑ effect of diuretics, ↓ effect of α_2 agonists, e.g., clonidine)
SSRIs	Depression, anxiety, obsessive-compulsive symptoms	Apathy
		Insomnia
		Restlessness, agitation, extrapyramidal symptoms, akathisia
		Central bradycardia
		↑ Potential for diverse drug–drug interactions
		Destabilization of Coumadin dosing

		Gastrointestinal distress, diarrhea
NSAIDs	Arthritis	Cognitive impairment ↑ Serum lithium levels ↑ Blood pressure, ↓ hypertension control
Neuroleptics/atypical antipsychotics	Dementia with psychosis	Death ↑ Cognitive impairment Extrapyramidal symptoms Falls Aspiration pneumonia
Ginkgo biloba	Impaired memory, arterial insufficiency	↓ Clotting time Restlessness (high doses)
Glucocorticoids	Inflammation, COPD	Depression, mania, psychosis, cognitive impairment (especially at high doses), insomnia

COPD, chronic obstructive pulmonary disease; NSAIDs, nonsteroidal anti-inflammatory drugs; SSRIs, selective serotonin reuptake inhibitors; TCAs, tricyclic antidepressants.

Differential Diagnosis

Differential diagnosis of psychiatric syndromes in the medically ill older patient begins with a mental status examination that incorporates a careful assessment of several cognitive dimensions, including attention and orientation (both impaired in delirium), memory and learning, and executive control function, the ability to understand, organize, and carry out goal-directed cognitive and motor sequences. Thorough collateral information and an evaluation of the patient’s past and current medical conditions, emphasizing the disease of vital organs and painful conditions, are essential. Complete medication history and a focused physical examination are important, including, at a minimum, a review of vital signs and a screening neurologic examination to check for abnormal gait or movements, focal areas of weakness, or abnormal deep tendon reflexes. The judicious use of appropriate laboratory tests may clarify the possibilities in complex situations (Table 55.3b–3). Finally, the gathered data can be synthesized into a list of possible differential diagnoses and probable causes. For example, in a patient with blood pressure in the low-normal range, pitting ankle edema, apathy, weakness, and depressed mood, who has been taking a high-potency diuretic for many years, an electrocardiogram (ECG) and echocardiogram may be needed to determine whether congestive heart failure (CHF) or the diuretic is the most salient contributor to the depressive syndrome, whether it is safe to try discontinuing the diuretic, and what special precautions are required in prescribing an antidepressant. In doubtful cases where the differential diagnosis includes a primary psychiatric disorder not caused by physical effects of illness, the electroencephalogram (EEG) may be helpful. The EEG in patients with psychiatric syndromes due to metabolic, toxic, or circulatory causes typically shows diffuse slowing, although it cannot implicate specific diseases or drugs, and the interpretation may be confounded when an underlying dementing illness is present. An increase in dominant frequency accompanying recovery lends weight to the formulation.

Table 55.3b–3.
Practical Evaluation of Psychiatrically Disturbed Medically Ill Older Patients

Establish rapport quickly.

Define the reason for the consultation or visit.

Learn and use a few rapid screening instruments:

- Cognition—e.g., Mini-Cog
- Depression—e.g., Patient Health Questionnaire-9 (PHQ-9)
- Neuropsychiatric symptoms—e.g., Neuropsychiatric Inventory (short version)
- Functional ability—e.g., Functional Activities Assessment

Expect more than one psychiatric diagnosis (e.g., delirium plus chronic depression).

Make a complete problem list:

- Medical
- Psychiatric
- Social
- Functional

Identify all drugs in current use:

- Prescription drugs
- Over-the-counter medications
- Vitamins, minerals, botanicals, and “natural” remedies
- Alcohol and tobacco
- Include those used infrequently
- Ask patients or caregivers to bring all medications from home in a paper bag for your inspection

Interview capable collateral sources:

- Family or other primary caregivers
- Primary physician
- Staff of residential facilities

For acute exacerbations of chronic problems, ask “why now?”

If problem unexplained, search for:

- Undiagnosed medical illness
- Drug effect or interaction
- Occult injury
- Undetected social problem (abandonment, abuse)

Conduct a screening physical examination: vital signs, gait, neurologic signs.

Obtain appropriate corroborating laboratory studies; consider broad-spectrum tests to screen for suspected disease, e.g., complete blood count, blood glucose, serum electrolytes, blood urea nitrogen and creatinine, liver enzymes, thyroid-stimulating hormone and free thyroxine, vitamin B_{12} and folate, syphilis serology if indicated, urinalysis, electrocardiogram, chest radiograph, electroencephalography for delirium, erythrocyte sedimentation rate for occult systemic disorders, hip radiographs for a delirious patient who will not walk, neuroimaging (head computed tomography for screening, or magnetic resonance imaging when microvascular disease is a possibility; single photon emission computed tomography for more precise diagnosis).

When a drug is believed to cause or contribute to a psychiatric disorder, all drugs should be sorted into those with a high potential for causing psychiatric symptoms (such as anticholinergic and sedating drugs), those clearly required for the management of serious medical disease, and those that may be modified or stopped on a trial basis. Familiarity with drug side effects, drug–drug interactions, and the pathways for their elimination may assist in the detection of otherwise occult medication toxicity syndromes. Close adherence to the “rule of maximal simplification” of medication regimens can produce clinical improvement in up to half of all psychiatrically disturbed medically ill older persons.

A 73-year-old man with severe chronic obstructive lung disease (COPD) came for the treatment of anxiety with insomnia and severe daytime fatigue. He had been treated with fluoxetine for several years after initial psychiatric evaluation, to which clonazepam had been added by his pulmonologist in doses up to 2 mg/day without relief. Shortness of breath and activity-induced panic attacks were his most pervasive complaints. Physical examination and imaging disclosed hypertension (190/120), persistent tachycardia, and signs of emphysema but no indication of CHF. A walk with oximetry led to oxygen desaturation from 93% to 83% after 3 minutes. His medication list included two antihypertensive agents, only one of which he took since the other was a diuretic that prevented him from leaving home unless assured of constant proximity to a bathroom. He was also on three scheduled inhaled medications and albuterol when needed (PRN). It turned out that he was using his PRN albuterol at least 12 times daily, but his technique was ineffective in achieving drug deposition in smaller airways. Pulmonary function tests showed no improvement with bronchodilator therapy. Diagnoses of activity-dependent oxygen debt and albuterol toxicity were made, manifested by treatment-resistant anxiety and hypertension, tachycardia likely to endanger his heart if uncorrected, and severe exercise intolerance. Reduction in albuterol use to four times daily, the addition of a second antihypertensive drug without diuretic activity, and prescription of continuous oxygen therapy with judicious oral morphine significantly improved his shortness of breath and activity tolerance and reduced his blood pressure and tachycardia to safer levels.

Older adult patients who present with previously undiagnosed cognitive impairment warrant aggressive assessment. The most prevalent dementia syndrome among geriatric patients is Alzheimer disease, while dementia with Lewy bodies is a close second. Other common causes of cognitive impairment in older patients include vascular dementia, frontotemporal dementia, dementia due to Parkinson disease, Korsakoff syndrome, and potentially reversible cognitive impairment (if caused, e.g., by depression, drugs, B12 deficiency, and others) (see chapter on Dementias for additional discussion). A meta-analysis of rates of reversible dementia found that 9% of cases had a partially or completely reversible etiology but that a total of only 0.6% of cases reversed with appropriate clinical management. These figures likely reflect general improvements in geriatric medical care.

Completing an evaluation of a patient with dementia often requires two or more visits, particularly for frail older adult patients, as well as a thorough collateral history gathering. Medical illness can interfere with the adequate diagnosis of dementia or unnecessarily delayed antidementia pharmacotherapy. Also often overlooked in the management of demented, medically ill patients is the opportunity to improve cognition by eliminating medications with mild but cumulative central nervous system effects that worsen both cognition and function.

Approach to Treatment: Overall Recommendations

When treating an older adult with medical comorbidities, some overarching principles are worth calling to mind. Where appropriate, involve family and other supports. Knowing the extent and reliability of a patient’s social support system can aid in predicting what additional resources may be needed and how likely a given treatment is to be successful. Likewise, economic status can shape treatment and its delivery, especially in older adults managing multiple medical conditions, medications, Medicare copays, and hospital bills. The living environment should be safe and adequately accessible to needed medical resources and should be addressed directly in a multidisciplinary treatment plan if it is not. Do not neglect the patient’s preferences in treatment, but continually assess for medical decision-making capacity.

Prescribing medications in the older adult poses a particular challenge for reasons such as polypharmacy, pharmacokinetic (altered volume of distribution, decreased renal and hepatic clearance) and pharmacodynamic (increased sensitivity to certain classes of drugs) changes that occur with aging, reduced functional reserve that limits compensatory homeostatic mechanisms, and underlying medical and neurologic diseases that impact drug metabolism and tolerance. These variables should be viewed as part of a constantly evolving process that will require continued attention to establish the most appropriate medication regimen for an older adult individual.

Polypharmacy is a significant concern in this population. A survey of community-dwelling older adults (aged 65 years and older) reported that 39.0% of the overall group surveyed used five or more medications and that the average amount of prescription medication use in this population has doubled over the past three decades, with the most pronounced increases being in males over age 80. In addition, regular use of over-the-counter medications and herbal and dietary supplements are also on the rise in the older adult. An explicit-based guide such as the American Geriatrics Society Beers Criteria or the Screening Tool of Older People's Prescriptions (STOPP)/Screening Tool to Alert to Right Treatment (START) can be helpful in the clinician's efforts to evaluate medication regimens and reduce polypharmacy.

These are the general guidelines that we recommend when prescribing medications for an older adult patient:

- Always keep patients' goals of care, current health, and cognitive status in mind.
- Before adding a medication, carefully review the current regimen and attempt to simplify the dosing schedule, adjust, and discontinue any unnecessary medications.
- Aim to make only one medication change at a time so that the effects of changes can be determined and distinguished.
- Expect that some psychiatric medications will take longer to achieve a therapeutic effect in older, medically frail adults. Be patient and use a slow pace of titration.
- Focus on potential drug–drug interactions and keep the anticholinergic load to a minimum.
- Be equally concerned about overprescribing and underprescribing.
- Whenever possible, prescribe an affordable and streamlined regimen.
- Educate the patient and caregivers about side effects and expected improvements.
- Periodically review effectiveness, need for dose adjustment, or continued use. Ideally, use structured means of monitoring effectiveness. Standardized scales (patient self-report, caregiver report, or provider-administered) are invaluable feedback devices, and there are many available to the clinician spanning a range of disorders and symptoms.
- Be vigilant of treatment adherence.
- Do not neglect cognition, as it may be not only a primary target of treatment but a factor complicating the treatment of medical and psychiatric conditions.

PSYCHIATRIC SYMPTOMS ASSOCIATED WITH SPECIFIC MEDICAL DISORDERS

Vascular Disease

Vascular disease, once thought to be the principal cause of dementia and psychoses of later life, is again being increasingly recognized as a cause of late-onset psychopathology as well as a factor exacerbating pre-existing disorders such as recurrent major depression, bipolar disorder, and Alzheimer disease. Ischemic and hemorrhagic strokes cause a variety of disturbances of behavior and emotionality related both to the effects of the lesion on brain function and to psychological problems in adaptation to disability. Major or minor depressions are common, occurring in roughly one-third to a half of all patients following an acute cerebrovascular accident. The most recent review of available evidence shows a significant association between stroke severity and risk of poststroke depression, but not the stroke type or mechanism. Though lesion location has been extensively investigated as a risk factor for poststroke depression, the findings have been inconsistent. Left frontal or left basal ganglia lesions are convincingly correlated with depression within 2 months of a first clinical stroke, but it is still difficult to draw clear conclusions beyond this. The persistence of milder depressive syndromes after the acute phase has been associated with other factors, including functional impairment, cognitive impairment, and inadequate social support. Several studies have shown the presence and severity of depression to predict the severity of impairment in activities of daily living as well as mortality in the year(s) after a cerebrovascular accident. The link between mood and vascular health is also notable for its bidirectional relationship; depression is an independent risk factor for stroke.

The importance of cerebrovascular disease in geriatric psychiatry extends well beyond poststroke depression. The vascular depression hypothesis supposes that cerebrovascular disease may be etiologically related to geriatric depressive syndromes. The clinical presentation of this entity is characterized by cognitive deficits, psychomotor retardation, lack of insight, and disability disproportional to depression severity. Response to antidepressants is generalized as suboptimal in these cases. Though the most common neuropsychiatric outcomes of cerebrovascular disease are depression, anxiety, fatigue, and apathy, other forms of psychiatric or behavioral morbidity may occur as well. Emotional lability, personality changes, psychosis, and mania are less common but may cause significant distress. There is often no clear relation to the location of the brain lesion, and management is especially challenging given the lack of an evidence base.

Cerebrovascular disease can also cause psychomotor slowing, impaired executive function, and significant functional dependency in patients who may or may not be clinically depressed. Common associated clinical features include executive dysfunction, encoding-based memory deficits, and, often, impaired gait and urinary control. Neuroimaging studies, especially those using magnetic resonance imaging, in affected patients typically demonstrate ischemic changes in subcortical and deep white matter and/or lacunar infarcts in the thalamus, the basal ganglia, and sometimes the upper brainstem. Interference with the function of frontal-subcortical circuits appears to be a common underlying anatomical feature of lesions that cause depression and executive dysfunction and can be seen as a partial disconnection syndrome. Many patients with subcortical vascular disease have minimal or no discernible mood change, showing instead only cognitive, motivational, and subtle neurologic signs. A diagnosis of vascular neurocognitive disorder may be appropriate in such cases, but minor depression may also be diagnosed when apathy and disengagement are functionally impairing. Vascular brain diseases can be caused by intrinsic disease of small and/or large vessels supplying the brain due to atherosclerotic narrowing, occlusive plaque emboli, medial (muscular) hypertrophy, or vascular amyloidosis, but cardiogenic causes must also be considered. Cognitive impairment resulting from the larger lesions of ischemic cortical strokes appears to be approximately as common as that resulting from lacunar strokes. This may be due to the lacunar infarct's role in unmasking impairment in preexisting diffuse cerebral small vessel disease. Among cardiogenic etiologies, several pathogenic mechanisms are recognized: hypertension, recurrent or severe hypotensive episodes, CHF, recurrent arrhythmias (especially atrial fibrillation), and especially coronary artery bypass grafting. Approximately 40% of heart failure patients will have cognitive impairment. A similar percentage of those who have had coronary artery bypass graft surgery for ischemic heart disease have cognitive impairment at 5 years postoperatively. Its causes are still the subject of active research, but emerging consensus favors variations in pre- and postoperative control of vascular risk factors as one important determinant. Patients with modifiable but poorly controlled risk factors are more likely to have strokes and myocardial infarctions in the year following surgery. The heart–brain connection is further underscored by evidence that hemodynamic and serum cardiac markers (such as natriuretic peptides and troponin) are related to worse cognition and a higher risk of dementia. All physicians caring for such patients, including psychiatrists, need to be alert to the potential to improve long-term outcomes with aggressive management of vascular risk factors.

In fact, identifying and modifying risk factors for further vascular insults should be a key component in the psychiatric treatment plan of patients suspected of having vascular psychopathology. Relevant workup includes fasting lipids, homocysteine, vitamin B12, and folate levels; a careful history and ECG to identify cardiac arrhythmias or, if already recognized, coordination of care and education of other physicians about the possible cognitive consequences, monitoring serial blood pressure measurements to identify hypertension and hypotensive episodes. Once identified, appropriate therapeutic steps can be taken to reduce risks, such as the prescription of folate/vitamin B6/vitamin B12 combinations, antihypertensives, antithrombotic agents, or lipid-lowering drugs. Depression related either to an acute cerebrovascular accident or to significant cerebrovascular disease is responsive to pharmacologic treatment with antidepressants, but the best agent and duration of therapy are unclear. Suboptimal response or concurrent apathy may call for adjunctive therapies, particularly stimulants such as methylphenidate or amphetamines and psychosocial measures to provide time and activity structure. Selective serotonin reuptake inhibitors (SSRIs) are considered first-line because of the available evidence and safety profile. It should be noted, however, that there is an unresolved but active debate about possible higher mortality in patients treated with SSRIs, with some studies finding higher and others lower mortality rates in treated stroke survivors. Treatment of stroke survivors with antidepressants has been shown to significantly lower the incidence of poststroke depression and may enhance recovery independent of poststroke depression or their effect on mood. Poststroke response to antidepressant treatment results in improvements in activities of daily living and cognitive function and decreased disability. Brief psychosocial interventions focused on care management, psychoeducation, and family support can be beneficial.

When executive dysfunction or gait impairment is significant, extra caution is needed in choosing and titrating psychotropic medications. A broad treatment strategy is often needed for optimal results: securing structure and support to compensate for defective executive function; planned, supervised exercise to improve and maintain gait and balance; and management of frequently associated urinary incontinence. It is worth remembering that anticholinergic medications, frequently prescribed for incontinence due to overactive bladder, can impair alertness and general cognitive abilities and counteract the benefits of concurrent cholinesterase inhibitors. When the therapeutic benefit of anticholinergics used to treat urinary incontinence is questionable, they should be stopped.

Heart disease is now a well-recognized cause of depression and anxiety in patients with myocardial infarction or heart failure. Medical and psychological factors both play a part, and as with cerebrovascular disease, a bidirectional causational relationship exists. Depression is an independent risk factor for coronary heart disease and myocardial infarction, and untreated or inadequately treated major depression is associated with adverse outcomes in individuals with known coronary heart disease and those who have undergone coronary bypass grafting. Depression is thought to impact cardiovascular disease via impaired heart rate variability, systemic chronic inflammation, hypothalamic–pituitary–adrenal axis dysfunction, and endothelial dysfunction. Effective psychiatric treatment often requires both psychotherapeutic and pharmacologic management. The safety and efficacy of antidepressant drugs in patients with chronic heart disease have been evaluated in several studies, and there is evidence that antidepressants may increase all-cause mortality, likely via mechanisms associated with known adverse effects. SSRIs are usually the drugs of first choice. These drugs occasionally cause bradycardia by a central medullary mechanism but lack primary cardiotoxicity and are generally considered the safest and most effective for patients with heart disease. Patients who fail to respond to these agents and do not have serious cardiac conduction defects or severe chronic CHF can generally be safely treated with other newer antidepressants such as bupropion, mirtazapine, or serotonin-norepinephrine reuptake inhibitors (SNRIs) under close blood pressure monitoring. When selecting an antidepressant medication, it is important to remember that many cardiac patients take multiple medications, and potential drug–drug interactions should be minimized. While electroconvulsive therapy (ECT) is an undisputed myocardial stressor and should generally be avoided in the acute postmyocardial infarction period, treatment can usually be given without major complications if anesthesia and treatment schedules are carefully individualized and managed.

A 72-year-old man had a history of recurrent major depression since early adulthood, as well as coronary heart disease. Depressive episodes were typically characterized by low mood, anergia, amotivation, and increased appetite and had a marked seasonal variation. With treatment with a selective serotonin reuptake inhibitor and seasonal phototherapy, he suffered no major episodes in a 3-year period of relative stability. In the month of January, he suffered a cardiac arrest and had a prolonged course of cardiopulmonary resuscitation complicated by difficult intubation in the field. During the hospitalization that followed, he was diagnosed with a diffuse anoxic brain injury, primarily affecting the subcortical watershed areas of the right hemisphere greater than the left. After his acute rehabilitation, he was left with residual left-sided weakness and new psychiatric complaints. He described his mood as “blah,” but not frankly depressed, could not find anything that interested him, and moved and spoke very slowly, even considering his physical limitations. He expressed worry about becoming like his father, who had Alzheimer disease because he could not follow complex instructions or adequately attend to challenging tasks. His wife took over the management of his medication as well as the bills, and his driving privileges were revoked. In addition, to close follow-up with his cardiologist and continued physical and occupational therapy, the man saw his psychiatrist, reporting that his previously effective antidepressant medication was no longer helping. In consultation with his cardiologist, low-dose methylphenidate was added to his medication regimen. At his psychiatry appointment 6 weeks later, his affect was notably brighter and more reactive, his speech and movements were less labored, and he had been able to enjoy dabbling in his old hobby of collecting military memorabilia.

Parkinson, Alzheimer, and Other Neurologic Diseases

Depression occurs with disproportionate frequency in Parkinson disease, affecting up to half of all patients at some point in the disease process. This depression is believed to be caused, in large part, by the neurodegenerative process itself; its presence is associated with evidence of damage to noradrenergic and serotonergic systems and is thought to identify a group of Parkinson patients with a more rapid course of the disease. Nevertheless, antidepressant drugs or other treatments can be effective. SSRIs have the most evidence supporting a benefit and are generally considered first line, given their relative safety and tolerability. TCAs have some evidence of efficacy but are less commonly used due to concerns about adverse effects in older patients and those with Parkinson disease specifically. SNRIs have been studied less, but from available evidence, they appear to be effective. Selegiline and safinamide, monoamine oxidase inhibitors, and anti-Parkinson agents, may be a useful addition to the treatment of depression in Parkinson patients. There have been several publications with positive outcomes in the treatment of depression in Parkinson disease with the use of dopamine agonists. However, concerns about methodology, clinical significance, and potential adverse effects outweigh the evidence, and routine use of these agents is not advised for the treatment of depressive symptoms. There is a paucity of data on the efficacy and tolerability of other standard antidepressant agents when used for Parkinson disease–induced depression. One systematic review suggests a good response to treatment with ECT but cautions about the propensity of adverse effects in this population, especially as the disease becomes more advanced. Importantly, cognitive behavioral therapy (CBT) has also proven effective.

Psychotic symptoms are common in Parkinson disease, with an estimated prevalence of around 60%, with visual hallucinations being the most common. Psychosis is more common as the disease progresses and is associated with comorbidities, including cognitive impairment, mood disorders, and sleep disorders. The first step in addressing psychotic symptoms is to rule out any concurrent processes that could cause worsened cognition or delirium (e.g., infections; anticholinergic medications). These syndromes may respond to lowering anti-Parkinson drug doses. If not, or if motor disability worsens to an unacceptable degree, atypical antipsychotic drugs with less impact on nigrostriatal dopaminergic pathways can be helpful. Pimavanserin [Nuplazid], a 5HTa inverse agonist approved by the Food and Drug Administration (FDA) to treat Parkinson disease psychosis, has been found to be effective and well tolerated. It is unique in its status as an antipsychotic without dopamine-receptor–blocking properties, and given this different mechanism of action, the time to response is significantly longer—on the order of 4 to 6 weeks. Clozapine (Clozaril) and quetiapine (Seroquel) are the most extensively studied antipsychotics, and though clozapine is consistently efficacious, quetiapine is usually prescribed first because of the risk of agranulocytosis and the requirement for frequent blood monitoring with clozapine. Positive response to clozapine can occur at low doses and in timeframes as short as one week. Motor deterioration in a significant percentage of patients has been observed with olanzapine (Zyprexa), risperidone (Risperdal), ziprasidone (Geodon), and aripiprazole (Abilify). Cholinesterase inhibitors may also be beneficial in reducing hallucinations, though the data supporting this assertion is sparse.

Major neurocognitive disorder due to Parkinson disease is more commonly seen in those with more advanced disease and more advanced age. The affected cognitive domains include attention and visuospatial functions, with language functions being relatively preserved. Behavioral and neuropsychiatric symptoms, including those discussed above, are nearly always present. The currently available evidence supports the use of cholinesterase inhibitors in these patients, with a positive impact on global assessment, cognitive function, behavioral disturbance, and activities of daily living rating scales. The rivastigmine (Exelon) transdermal patch is often better tolerated than oral formulations, though exacerbation of tremor remains a concern. Response to memantine (Namenda) in this disorder has been disappointing.

In Alzheimer disease, depression affects up to one-third of patients, especially in milder stages, and may be the major presenting problem. Despite their widespread use, antidepressants lack clear evidence for this patient population. They are also associated with significant adverse effects, including falls, hyponatremia, cardiotoxicity, and increased bleeding risk. Trials of antidepressants may be best reserved for those with historical responses to antidepressants or those with moderate or severe depression who have not responded to behavioral interventions. As always, risks and benefits should be carefully weighed. Apathy is even more common than depression in major neurocognitive disorder due to Alzheimer disease, but similarly, antidepressants and other commonly used agents haven’t proven to be effective. Methylphenidate may be beneficial, but the quality of current evidence is low. With the progression of dementia, classical presentations of depression become less frequent, while agitation, psychosis, and aggressive behavior with or without dysphoria become more common. Mixed behavioral syndromes are the rule in middle and later-stage dementia. Assaultive or threatening behavior demands thorough medical and socioenvironmental assessment to identify precipitating and aggravating factors, but pharmacologic interventions are often required. Second-generation antipsychotics are commonly used off-label in these cases, but their use has been controversial since black box labeling by the FDA. Recent reviews conclude that the use of these agents confers modest benefits for short-term treatment of aggression and psychosis as well as a small but significant benefit on caregiver burden. This is offset, however, by the substantial risk of serious adverse events, including 1.6 to 1.7 times increased mortality risk. Trials of valproate for these dementia-related behaviors have been disappointing. Existing evidence for the use of SSRIs for the treatment of agitation associated with Alzheimer disease is promising but limited. Further high-quality research is also needed on the use of cognitive enhancers for behavioral and psychological symptoms of dementia. Delirium induced by drugs or acute medical illness is a risk at all stages and in all types of dementias, as brain compensatory mechanisms are impaired.

A 75-year-old male Vietnam veteran with moderate major neurocognitive disorder due to Alzheimer disease was brought for psychiatric treatment because of increased episodes of tearfulness and speaking about the loss of life and other traumas witnessed in Vietnam. Upon interview, his family reported that this was a relatively new change for him and that he hadn't previously focused so much on these troubling memories. They also reported that the television in the home was often playing war movies or historical documentaries, as this was the patient's stated preference. The family was instructed to attempt to engage him in activities he used to enjoy or direct him toward different types of television programming. As a result of this, family and caregivers spent more time interacting with the man as well. His mood seemed to improve. A year later, the man returned with his caregiver and family, now with aggressive behaviors of grabbing and hitting as well as yelling out, especially around times of care provision. The patient was assessed for any physical cause of the discomfort, but none could be found. The caregiver received additional information on communicating with individuals with Alzheimer disease and providing distractions and rewards around these difficult times. The patient's behavior improved only minimally with these interventions, and a low dose of quetiapine was started. The behaviors improved, and the psychiatrist was able to gradually taper off the medication after 3 months.

Post-traumatic and postanoxic encephalopathies, hydrocephalus, central nervous system tumors and infections, multisystem atrophy, Shy–Drager syndrome, Huntington disease, demyelinating diseases such as multiple sclerosis, idiopathic inflammatory diseases such as systemic lupus erythematosus and temporal arteritis, and occupational neurodegenerations such as chronic organic solvent encephalopathies may all be associated with depressive, manic, delirious, and psychotic syndromes. Antidepressant trials are generally warranted for the treatment of depressive disorders associated with these diseases, though drug intolerance may be problematic and require the use of innovative and combination treatment strategies. The management of other psychiatric complications is less well established.

Chronic Respiratory, Metabolic, and Endocrine Disorders

Chronic obstructive respiratory disease (COPD) is the most prevalent of the chronic respiratory diseases in the second half of life. Cigarette smoking and, in the developing world, exposure to biomass fuels are significant risk factors. COPD is strongly associated with depression, anxiety, and, in advanced cases, a special form of panic disorder precipitated by exertion. Over 40% of patients with COPD suffer from depressive symptoms. Smoking is a shared risk factor in depression and COPD, and the rates of depression are even higher in COPD patients who are actively smoking. Untreated depression in COPD worsens overall outcomes with more COPD exacerbations per year, contributes to persistent smoking, and has been associated with increased risk of death, poorer social functioning, increased physical disability, poor adherence to treatment, and health care overutilization.

Psychological complications are common and include feelings of hopelessness and helplessness that contribute to the impaired treatment and rehabilitation adherence, anxiety-related dyspnea and secondary panic symptoms fostering further inactivity, fear of sudden death by asphyxiation, intense self-blame in patients whose illness is related to smoking and social isolation due to shame, low mood, and lack of motivation. For milder dysphoric states that do not meet the criteria for a depressive diagnosis, multidisciplinary rehabilitation and counseling approaches improve mood in many patients. In such cases, the efficacy of antidepressants in COPD has not been demonstrated either for alleviation of depression or improvement of dyspnea. However, available evidence suggests that, in patients with clinical depression and the absence of left ventricular disease or hypercapnia, antidepressant pharmacotherapy is beneficial for depression, anxiety, and somatic symptoms. A risk and benefit analysis regarding the use of antidepressants needs to be considered in each patient. Comprehensive nonpharmacologic interventions for anxious, depressed patients with COPD provide individualized education about the disease, its management, and adherence to recommended treatment, teach patients to modify their symptoms with appropriate techniques such as breathing retraining and offer supervised physical training. The use of behavioral activation, CBT, and collaborative care approaches should also be considered. Patients with pulmonary disabilities and those who require home oxygen therapy may benefit from self-help groups.

Deterioration in pulmonary or cardiopulmonary status in patients with COPD can lead to acute and chronic encephalopathies, sustained hypoxemia and hypercapnia, or to the central nervous system effects of drugs used in management (oral or inhaled bronchodilators, high-dose prednisone, cough or dyspnea suppressants, or benzodiazepines). Psychotic states may occur, usually related to the acute impairment of brain function by severe exacerbation of illness. Prior to the establishment of guidelines for oxygen therapy in COPD, neurocognitive impairment was thought to be quite widespread, but recent studies questioned the prevalence of cognitive disorders due to COPD prior to severe and end-stage disease. Patients with lung cancer may develop psychiatric complications similar to those occurring in COPD; in addition, paraneoplastic psychiatric syndromes with psychotic, cognitive, and/or affective symptoms have been reported.

Chronic renal and hepatic disease can lead to well-known delirium states due to their detrimental effects on nutrition, metabolism, clearance of toxic metabolites, and drug disposition. In addition to these transient alterations of mental function, the prevalence of other psychiatric disorders in these conditions approximates those for chronic disease in general. Pharmacologic treatment considerations include the effects of disease on half-life and plasma drug levels and adjusting prescribing accordingly.

In renal disease, depression is associated with clinical progression regardless of dialysis status and must be carefully distinguished from somatic symptoms of the underlying condition. These patients can benefit from assertive antidepressant treatment, taking into consideration drug metabolism concerns carefully. Although suicide risk in advanced renal disease appears to be less elevated than previously thought, both depression and renal disease can have significant effects on quality of life. Dialysis termination raises complex end-of-life issues where nephrologists, palliative care specialists, and psychiatrists can work closely together with patients. Hepatic disease, in general, increases the risk for both acute and chronic cognitive alterations; hepatitis C specifically is frequently comorbid with neuropsychiatric conditions. Fortunately, the advent of direct-acting antivirals has provided for curative treatment, likely greatly reducing both direct and indirect psychiatric morbidity from hepatitis C. These newer agents lead to much fewer psychiatric side effects than prior, less efficacious treatments such as interferon, and evidence suggests patients with psychiatric comorbidities can safely be treated with these medications.

Diabetes is implicated in both depression and cognitive impairment. Depression rates are at least twice those of the general population and can be associated with poor glycemic control, lessened adherence to medications, and reduced self-care, all of which can, in turn, contribute to increased diabetic complications and elevated mortality rates. In short-term studies, treatment for depression improves both glucose management and quality of life. In older patients, type 2 diabetes is predominant and has been consistently associated with an elevated risk of dementia; recent guidelines for the treatment of diabetes in older adults recommend cognitive screening at baseline (without specific recommendations on the method) and regularly thereafter, especially for those at higher risk. The pathogenetic source of diabetic neurocognitive risk is not yet understood, but it has been suggested that both glucose control and insulin itself play a critical role in neuronal health. It remains unclear to what extent optimal glucose control can modify these risks, with longitudinal studies in this area an active area of exploration. Ongoing treatment of comorbid depression is very likely to improve long-term care of these patients, and as with other major chronic diseases in older patients, requires heightened vigilance for depression in the setting of somatic symptoms.

Hormonal changes have a complex and still unclear role in mood in older adults with multimorbidity. For example, even subtle thyroid abnormalities can play a role in mimicking or worsening psychiatric syndromes. Another area of interest and continuing investigation relates to changes in sex hormone levels which evolve substantially with age and can be deranged with illness. Testosterone repletion can be helpful in older men with depression, but only if hypogonadism (as reflected by low hormone levels) is present. Testosterone supplementation in men with normal hormone levels remains controversial and is likely to pose cardiovascular and prostate risks. Conversely, androgen deprivation therapy for prostate cancer poses a clear risk of depression. In women, the relationship between postmenopause sex hormones, cognition, and mood also appears to be highly complex, and hormone replacement likewise carries substantial risks versus uncertain benefits. For both older women and men, there is much to learn about the interrelationship of sex hormones and psychiatric conditions, both in health and disease.

Chronic Pain and Musculoskeletal Disorders

In older patients, the prevalence of chronic pain ranges between 36% and 80%, depending on the setting. Patients suffering from pain often restrict activity to prevent further deterioration and to avoid exacerbation of pain, which leads to deconditioning, isolation, and worsening disability. Over half of older adults with depressive disorders have chronic pain, with depression more likely to be present in patients with multisite pain and higher pain intensity and frequency. Older patients who have both chronic pain and depression are more likely to experience insomnia and suicidal ideation; they have a higher level of impairment and poorer quality of life. Hospitalized older patients with chronic pain, particularly from osteoarthritis, are more likely to exhibit severe frailty. Chronic pain is also a risk factor for premature death and accelerated cognitive decline, making it an important target for the prevention of these outcomes.

The causes of chronic pain in older adults include musculoskeletal disorders, neuropathies, and pain related to cancer, with musculoskeletal disorders being the predominant cause. These include fractures, osteoarthritis, rheumatoid arthritis, microcrystal disorders, myositis, vasculitis, and other arthropathies. Aging musculoskeletal tissues are less resilient due to the loss of ligament elasticity, redistribution of fat, osteoporosis, and subsequent decrease of bone strength. Sarcopenia, the loss of skeletal muscle mass, is a nearly universal aging phenomenon, contributing to disability and functional loss in older patients. The older adults are also at higher risk for infections that may affect the joints due to the comorbid predisposing disorders (peripheral vascular disease, diabetes mellitus), immunosuppression, and surgical interventions that are prevalent in this population. Falls are both a risk factor for incurring injuries that may result in chronic pain and a consequence of chronic pain. Insomnia is often comorbid with chronic lower back pain in older adults and may be a contributor to the severity of pain and associated disability.

Sixty percent to 70% of all joint diseases in the older adult is osteoarthritis, often comorbid with regional pain syndromes. A common complaint in ambulatory patients over the age of 75 is chronic low back pain, while knee osteoarthritis remains the most common cause of disability in older persons. Rheumatoid arthritis is to blame for 10% of all joint pain complaints, with patients who are taking biologic disease-modifying drugs being at the highest risk for depression, suicidal ideation, and poor quality of life. The number of older adults with arthritis or joint symptoms is projected to be around 41.1 million by the year 2030.

Chronic pain is undertreated in older adults for several reasons, among them poor access to care, patients' concern about the meaning of pain and the cost of its diagnosis and treatment, and the patients' and providers' belief that pain is concomitant with aging, although there's evidence that a significant portion of older patients with chronic pain hope that physicians will pay more attention to their pain complaints. Physicians may fail to assess the patients' pain or believe in it when it is reported. There's also a persistent fear that prescribing opioid medications, often necessary for the control of pain when it doesn't respond to other agents, will result in addiction. A recent survey of hospitalized geriatric patients found a prevalence of pain at around 67.3%, but only 49% of these cases were treated; 75% of patients felt their pain medication was of low or no efficacy. Patients with cognitive impairment, particularly those with severe dementia and inability to communicate, are at risk for undertreatment of pain due to failure to identify pain and use appropriate tools for measuring pain.

Treatment of chronic pain in older adults is a complex undertaking, requiring a multidisciplinary approach. A comprehensive assessment should include a thorough history of pain, past medical history, history of substance abuse, physical and mental status examination, a thorough review of medications, and an examination of how pain affects the patient's quality of life and function. Self-report of pain is the gold standard for evaluation of pain in any patient, although this may be challenging in cognitively impaired, noncommunicative patients.

Pharmacologic treatment should be individualized for each patient; options for the treatment of pain exist, but each one has its pros and cons. It should be kept in mind that older adults are underrepresented in clinical trials testing interventions for chronic pain, and the impacts of comorbid medical conditions such as polypharmacy and frailty on the management of medications commonly used for pain are largely unknown. While opioids are an effective way to treat noncancer and possibly neuropathic pain, their duration of action in older adults is prolonged, and long-acting medications are particularly concerning. The consensus statement by the expert panel from the International Forum on Pain Medicine recommends that doses of opioids should be lower, and a longer time interval should be used between doses. Further, consideration of any concomitant medications affecting respiratory function and pulmonary conditions should occur to avoid problems with hypoventilation and respiratory depression. The use of opioids in patients with a history of substance abuse should be carefully monitored. Further, it should be remembered that opioids may cause depression, the risk of which increases with the duration of exposure to opioids. Risk factors for opioid misuse, abuse, and the likelihood of overdose or opioid-induced respiratory depression should be considered in each individual case. The Risk Index for Overdose or Serious Opioid-induced Respiratory Depression (RIOSORD) has been validated in veteran and nonveteran users of prescription opioids and is one of the tools that may be used to evaluate the risk of adverse events with these medications.

TCAs, gabapentin, and pregabalin can manage neuropathic pain, but they are sedating. TCAs are also anticholinergic with risks of cognitive impairment, urinary retention, constipation, visual changes, and cardiac conduction disorders, among others. Some nonsteroidal anti-inflammatory drugs (NSAIDs) may cause gastrointestinal bleeding and discomfort.

Nonpharmacologic treatments should be utilized whenever possible: These include biofeedback, physical therapy, nutritional counseling, chiropractic, acupuncture, aquatic therapy, weight control, corticosteroid joint injections, and surgery if indicated. Psychoeducation, psychotherapy (CBT is proven effective in a variety of chronic pain disorders), and psychiatric medications such as antidepressants should be recommended if depression and/or anxiety symptoms are comorbid with pain. Due to increasing interest regarding the assessment and treatment of pain in older patients, recent years have seen publications of excellent guidelines. The authors recommend “Guidance on the management of pain in older people,” published in *Age Ageing.* 2013;42:i1–i57, by the British Geriatrics Society.

A high index of suspicion regarding the presence of untreated pain should be applied when working with patients with dementia. A change in their behavior, including aggression and agitation, could be related to pain. Other pain manifestations may include changes in facial expression, verbalizations, and body movements, as well as the onset of crying, irritability, and increased confusion. Using scales developed specifically for the evaluation of pain in patients with cognitive impairment will improve outcomes. An example of such a scale is the Pain Assessment in Advanced Dementia (PAINAD), applicable in long-term care and acute geriatric care settings. Investigation of possible causes of pain should be undertaken with the assistance of collateral information, full medical history, and physical examination. Finally, a trial of an analgesic resulting in improvement in behavior may confirm the diagnosis of pain. For further guidance on addressing pain in patients with dementia, we recommend "The importance of pain management in older people with dementia," published in *Br Med Bull.* 2014;111(1):139–148, by Corbett et al.

An 81-year-old man with previously diagnosed dementia related to Alzheimer disease sustained a fall resulting in a right elbow fracture. After the orthopedic surgery, he had a difficult postoperative course complicated by severe confusion and a disrupted sleep–wake cycle. A geriatrician who examined the patient found a confused, thin, older adult man with a marked level of rigidity and cogwheeling in the left arm (the right arm could not be examined due to pain). He was disoriented, and his level of consciousness fluctuated markedly during the interview. He had difficulty focusing on the content of questions and was unable to explain why he was in the hospital. The patient's son reported that his father's memory and executive functioning have been declining for over 4 years, and the course of this disease was marked by visual hallucinations, usually of animals like cats, which were distressing for the patient and for his family. Rigidity and falls became a problem 2 years after the onset of cognitive impairment. After the surgery, the patient was described as markedly below his recent cognitive baseline. The geriatrician diagnosed the patient with delirium related to the recent surgery and the number of sedating medications on board, including morphine, fentanyl, zolpidem, and lorazepam. Discontinuation of the above medications and initiation of scheduled acetaminophen and tramadol for pain led to significant improvement in the patient's mental status and eventual discharge from the hospital to a rehabilitation facility. The geriatrician subsequently reexamined the patient and diagnosed him with dementia with Lewy bodies rather than Alzheimer disease, recommending avoidance of long-acting opioids and sedating medications, as well as initiation of rivastigmine patch to address cognitive impairment. The patient also worked with a physical therapist to improve balance and coordination, with the goal of fall prevention.

Neuropsychiatric Sequelae of COVID-19 Syndrome

Novel coronavirus SARS-CoV-2 emerged at the end of 2019 as a worldwide epidemic causing COVID-19. Though the primary manifestations of COVID-19 are respiratory, with the most common complaints including fever and cough, it can impact other organs, including the brain, through the direct neuroinvasion by coronavirus and through the dysregulated immune response resulting in excessive inflammation and cytokine network dysregulation. Additional mechanisms by which neuropsychiatric sequelae occur may be related to the side effects of drugs used to treat COVID-19, such as steroids, and coagulopathy related to increased prothrombin time. Neuropsychiatric symptoms of COVID-19 include headaches, myalgias, anosmia, ageusia, and asthenia, as well as more serious complications such as encephalopathy, encephalitis, meningitis, stroke, seizures, and neuropathies. Fatal outcomes related to the neuropsychiatric sequelae of COVID-19 have been documented, and in the COVID-19 survivors, the inflammatory process in the brain could trigger long-term mechanisms leading to neurodegenerative and neurologic disorders emerging down the line. Older adult people with medical comorbidities are more vulnerable to developing a fatal illness.

COVID-19 syndrome among hospitalized patients is associated with several psychological factors, including fear, loneliness, boredom, anger, and depression. Over the years 2020 and 2021, public health measures such as quarantine and isolation had their own mental health consequences, as people in the general population exhibit higher levels of stress, anxiety, and poor sleep quality. A month after the outbreak, a survey of people from Wuhan (where COVID-19 was first reported) and the surrounding cities in China showed that symptoms of post-traumatic stress disorder (PTSD) were present in 7% of the respondents. In a study of Indian adults (with India having been a pandemic hotspot), sleep difficulties, paranoia about having the COVID-19 infection, and distress were reported by 12.5%, 37.8%, and 36.4% respondents, with 80% of respondents expressing a need for mental health services. There have also been reports of vulnerable individuals engaging in suicidal behavior out of fear of COVID-19. Those with pre-existing mental health illnesses were more likely to experience anxiety, depression, distress, and insomnia. Additionally, patients with psychiatric disorders may be more vulnerable due to being on psychotropic medications with such side effects as metabolic syndrome, comorbid medical problems, and smoking habits.

In the past, a significant percentage of patients infected with other types of coronavirus (e.g., SARS-CoV-1) have been diagnosed with PTSD, depression, pain disorder, panic disorder, and obsessive-compulsive disorder in the months postinfection, signaling that the burden of COVID-19 will likely emerge over the years following the pandemic. SARS-COV-2 infection has already been implicated in the onset of psychosis, mood disorders, PTSD, and suicide.

Attention must be given to monitoring older adult people with acute COVID-19 illness for neuropsychiatric symptoms, as delirium and depression may be underrecognized and undertreated. If psychiatric medications are continued or initiated for those patients who are receiving antiviral medications, drugs that do not act on the cytochrome P450 should be utilized preferentially.

Those with pre-existing psychiatric conditions and neurodegenerative disorders such as dementia may be at the highest risk for experiencing the neuropsychiatric complications of COVID-19. Individuals who are impacted by the COVID-19 pandemic in other ways, such as social isolation, are likely to experience less distress with support from their family and social networks. As older adult people living alone are particularly at risk for loneliness during quarantines, efforts to decrease their isolation should be actively undertaken. Given the risk of long-term neuropsychiatric sequelae of COVID-19 emerging over the years following the acute illness and the well-known risk factors for long-term cognitive and functional impairments following a critical illness, it is paramount to monitor the survivors' neuropsychiatric status throughout their life course, both for their benefit as well as to establish a framework for understanding the course of COVID-19 disease. Research about the impact of this pandemic on mental health, particularly in the older adult, is still lacking, an issue that must be addressed in the years to come.

Finally, adaptations for delivery systems for mental health care should be implemented to ensure continued access to care through the increased use of telehealth, creation of virtual mutual support meetings, provision of online resources, and creation of prevention programs for relatives of people who died from COVID-19 as had been done in some countries. Early experience with increased use of telehealth during the pandemic shows benefits with this model of care, such as reduced logistical barriers to care and decreased no-show rates. In order to ensure availability of telemental health for all who need it for years to come, lifting reimbursement restrictions, lifting of the Ryan Haight Online Pharmacy Consumer Protection Act restriction on prescribing controlled substances via telemental health, suspension of restrictions on licensure requirements to practice telemental health across state lines are necessary, along with any other systematic changes that may be required.

COVID-19 has been a great challenge for the world to overcome, and it also presents a unique opportunity to redesign our health care system with an eye on quality and mitigation of health care disparities.

SPECIAL ISSUES IN THE CARE OF OLDER ADULT PATIENTS

Impaired Vision and Hearing

The importance of sensory impairments in patients who develop psychiatric disorders in later life justifies special attention to these issues during evaluation and treatment. Degenerative diseases of the eye and ear impair responsiveness to the interpersonal and social environment, increase feelings of vulnerability among older persons, and cause hallucinations in the affected sensory modality in some individuals. The most common vision problems in older adults include spatial contrast sensitivity, and scotopic dysfunction, which are important, respectively, for pattern recognition and adaptation to darkness, as well as slow visual processing. Visual hallucinations are common in later stages of Alzheimer disease as neuropsychiatric symptoms mount, in dementia with Lewy bodies, and in Parkinson disease, contributing to increased caregiver burden. Visual hallucinations in the absence of psychiatric disease (Charles Bonnet syndrome) occur in 10% to 50% of patients with low vision, a potentially distressing experience. In addition, deafness originating in childhood disease but exacerbated by presbycusis appears to be a risk factor for late-onset paranoid illness, and hearing and visual loss both contribute to the onset and persistence of depression.

Longitudinal population studies demonstrated that the presence of either visual or sensory impairment places the older adult at risk for cognitive decline, while those with both impairments may be at even higher risk for faster cognitive decline after the age of 50. In turn, those suffering from comorbid sensory and cognitive impairment are vulnerable to increased risks of depression, death, lower autonomy, and hospitalization. They also contribute to higher health care spending.

Although the U.S. Preventive Services Task Force did not recommend routine screening for visual impairment in asymptomatic older adults, there may be a reason to perform routine screening in those who are frail and thus may be at even higher risk of being adversely affected by sensory impairments. Treatment of common causes of eye disease in the aging population, improvement in corrective lenses, cataract surgery, appropriate prescription of hearing aids or portable magnifiers, and attention to environmental lighting and sound may enhance the efficacy of specific interventions for psychiatric disorders in sensory-impaired patients. Research opportunities on how to improve the care of the aging population with sensory impairments abound, and contributions in this field may further assist clinicians with the care of persons with comorbid sensory and psychiatric disorders.

Falling

Falling and fall-related injuries occur in 30% of community-dwelling older adults and in more than 50% of those living in nursing homes each year. Increasingly, it is recognized that cognitively impaired older adults have an even higher prevalence of falls and associated morbidity. According to recent data from the Centers for Disease Control and Prevention, fall death rates among older adults increased about 30% from 2009 to 2018, with the fastest growing rate among adults aged 85 and older. Falls are the main contributor to injury, related disability, and premature death in older people. Complications of falling include hip fractures (25 per 1,000 women over age 85), other fractures, lingering painful soft tissue injuries, and head injuries with behavioral complications, which are widely feared by older persons as harbingers of death. In older patients with cancer, falls are also associated with an increased risk of chemotherapy toxicity and poorer survival. Risk factors for falling include conditions that cause postural instability, general muscular deconditioning, gait disorders, painful or deforming foot conditions, visual, auditory, vestibular, proprioceptive, and central neurologic impairments, decrease in cognition and speed of processing information, depressive symptoms, environmental hazards, inadequate ambulatory supports, urinary incontinence, and all classes of psychotropic medications. In hospitalized patients, the presence of mental status deficits is one of the strongest risk factors for falls. Overall, the use of medications with sedative effects is the single most powerful predictor of falling in the older adult, all other things being equal. These include drugs used for the management of insomnia, anxiety, depression, and agitation (prescribed and over the counter), alcohol, and drugs used in the treatment of medical diseases that sedate by themselves or enhance the sedative effects of other drugs. The presence of multiple risk factors geometrically increases hazard.

Assessment of these risk factors and remediation whenever possible may prevent falls, with multifactorial interventions being more effective, given that falls often result from the interaction of multiple factors listed above. The goals of prevention include reduction of falls and associated risk for injury and maintaining the highest possible level of independence; just as important is education aimed at recognition of fall risks and their prevention for health care providers. Given the close relationship between gait, cognition, and falls, strategies that improve attention and executive function can help with deficits in gait control and fall prevention in older cognitively impaired adults. Falls are common in older patients with insomnia and sleep apnea, and treating sleep disorders with nonpharmacologic approaches may help improve attention, daytime vigilance, gait, and balance control, thereby preventing falls. Screening for falls and risk of falls is strongly recommended, given their potentially devastating consequences and the availability of useful interventions, such as exercise and physical therapy.

Psychiatric complications of falls include acute cognitive disturbance due to head injury and a common but rarely recognized phobic anxiety state (fear of further falls, which leads to limiting participation in activities of daily living, further deconditioning and curtailment of opportunities for social exchange, and sometimes permanent retirement to a wheelchair). Persistent fall phobias may become less common with aggressive rehabilitation after hip fracture and the dissemination of exercise programs for frail older adults.

Urinary Incontinence and Urinary Tract Infections

Urinary incontinence affects up to one-third of older adults living in the community and can cause significant suffering, decreased quality of life, and functional impairment, usually worsening as the problem becomes more severe. Notably, urgency urinary incontinence causes a higher degree of disability than stress urinary incontinence. Routine screening for urinary incontinence is essential, as older patients are less likely to mention urinary incontinence to their physicians and suffer silently, limiting their trips away from home and their opportunities for social activities. Guidelines for screening for urinary incontinence have been published by several professional societies.

There are many causes of urinary incontinence, including urinary tract infections, altered mental status, restricted mobility, polypharmacy and specifically the use of diuretics, and chronic anatomical changes. A growing body of research has found associations between anxiety, depression, and all types of urinary incontinence in women; the use of antidepressants is independently associated with urinary incontinence as well.

Urinary tract infections are one of the most common infectious diseases in this age group and are associated with multi-infarct dementia in older women. Conversely, patients with dementia are at higher risk for urinary tract infections. Patients with acute or untreated chronic urinary symptoms should be referred to their primary care physician, gynecologist, or urologist for a thorough workup.

Frailty

Frailty is a clinical syndrome comprised of several physical symptoms, including weight loss, exhaustion, sedentary behavior, slow gait, and muscle strength. Those who fulfill 3 of these 5 criteria are considered to have physical frailty. Those who have 1 to 2 of these criteria are considered to have prephysical frailty. Other symptoms of frailty include weakness, inactivity, balance and gait abnormalities, and decreased food intake. However, frailty is also a multidimensional syndrome that encompasses physical, social, and cognitive dimensions. It is thought of as an age-related state of vulnerability related to reduced reserve and capacity of the organism to withstand stress and maintain homeostasis, a syndrome characterized by the decline of multiple physiologic systems. Many medical problems can lead to frailty, including chronic infections and inflammatory conditions, cancer, CHF, and thyroid abnormalities. Some common psychiatric problems may also manifest as frailty, among them depression, psychosis, grief, and dementia. Frailty is associated with, and predictive of, physical disability, falls, hospitalizations, malnutrition, and increased mortality. The prevalence of this syndrome in the older adult is between 14% and 24%, depending on the definition of frailty, and is generally higher among women and blacks.

Recent data indicate that cognitive function is significantly poorer in patients who are frail; moreover, a higher level of physical frailty is associated with a faster rate of global cognitive decline. Impaired cognition and frailty may be independent syndromes, but they often overlap and share common risk factors such as hormonal changes, nutrition, cardiovascular risks, mental health and, most importantly, increasing age. The systemic changes related to chronic inflammatory conditions, poor nutrition, and declining function of endocrine and immune systems affect neurons in the hippocampi and microglia, placing the brain at risk for cognitive impairment, delirium, and dementia. A recent systematic review and meta-analysis found that frailty was a significant predictor of Alzheimer disease, vascular dementia, and all dementia among community-dwelling older people. The interconnectedness between cognitive and physical aspects of decline related to frailty led to the consideration of a concept of cognitive frailty, which consists of the following criteria: (1) presence of physical frailty and cognitive impairment, (2) exclusion of concurrent Alzheimer disease dementia or other forms of dementia. This definition is still being refined, and further research is needed to develop successful preventive strategies for this syndrome.

An association of significance also exists between depression and frailty. A recent study of men older than 75 years of age found that those with current symptoms of major depressive disorder had a fourfold increase in mortality, mostly due to comorbidity with frailty. Another cross-sectional study of older adult men found that those with frailty were much more likely to also suffer from anxiety and depression. Among hospitalized patients, frailty is often comorbid with delirium, conferring a poorer prognosis and decreased survival than either condition alone.

The best approach to the treatment of frailty and associated mental health conditions starts with a comprehensive geriatric assessment, whether conducted during an acute admission or on an outpatient basis. Encouraging studies of using exercise to improve endurance and increase strength and balance in frail patients have been conducted, but there is little known at this time on the potential value of nutritional changes or pharmacologic agents. Interventions of most benefit are likely to be multidisciplinary and multidimensional, addressing the whole person rather than a single disease or organ system at a time and paying attention to the mental health problems of an aging, frail patient.

Overreliance on Drugs for Managing Mental Disorders

Judiciously prescribed medications are likely to remain the mainstay of treatment for older, medically ill patients. However, potentially inappropriate medication (PIM) use is more likely in older patients leading to increased adverse drug reactions and decreased efficacy. In a comprehensive survey of PIM medication lists in the literature, psychoactive medications were strikingly commonly cited, especially TCAs and benzodiazepines; antihistamines, which are often used as sleep aids and for anxiety even in medically ill late-life adults, were also commonly cited as inappropriate in older patients. The only medications almost universally represented across all surveyed PIM lists were diazepam, chlordiazepoxide, and amitriptyline (along with the nonpsychotropic indomethacin).

Considering the risks that can be present for older patients treated with medications, whether overtly cited as PIM or with polypharmacy more generally, psychiatrists and other physicians tend to underutilize psychotherapeutic and behavioral interventions in the care of sick or frail older patients. Care for psychiatric illness can be greatly enhanced by a psychotherapeutic orientation. While access to formal office-based individual psychotherapy is often a limited resource, there is accumulating evidence that brief therapeutic interventions, which can be delivered in the primary care setting where chronically medically ill patients are most often cared for, can provide considerable benefit and are easier to access for these patients. As mental health care becomes more available in an integrated care setting, psychotherapeutic interventions may prove highly efficacious and cost-effective as a complement to, and at times an alternative to, long-term medication use. Even more encouraging from a public health perspective, emerging evidence indicates that older patients can benefit from the delivery of psychological care via telehealth modalities. This has the potential to help greatly improve access to mental health care for older patients who commonly face social, financial, and medical barriers to travel for in-person care.

In turn, increased use of nonpharmacologic treatments could potentially be of benefit in reducing the risks of adverse effects from medication overuse and polypharmacy. Brief structured psychotherapies can be delivered individually, with key family members, or in groups in both specialty mental health and primary care settings. More broadly, the psychiatrist's use of psychotherapeutic techniques from diverse orientations can include the capacity to develop emotional connections with patients and empathic integration of the needs of the patients, their families, and professional caregivers into a comprehensive and workable treatment plan. In this way, elements of psychotherapy can make substantial contributions both to medical management and to patients' comfort and capacity to persevere in the face of demoralization from a serious and ongoing illness.

A HEALTH SYSTEMS APPROACH TO THE CARE OF THE MEDICALLY ILL GERIATRIC PATIENT

Older adults are less likely than younger patients to seek and receive specialty mental health services and, when referred by their primary care providers, are less likely than younger patients to follow through on such referrals. Older adults, their family members, and their physicians often attribute enduring but treatable psychiatric symptoms to "normal aging," stressful life events, grief, social isolation, losses, financial difficulties, care-taking responsibilities, dementia, or medical illness, resulting in hopelessness and therapeutic nihilism in patients and providers. Older adults may also fear stigmatization associated with mental disorders or psychiatric treatments. As a result, the majority of this subset of the population receives care for mental and emotional problems exclusively in the general medical sector, and many who could benefit from the treatment of psychiatric disorders do not receive it. It is worth noting that Caucasian men over the age of 85 are at the highest risk for suicide in the United States, primarily due to their use of firearms during suicide attempts, and that many older adults visit primary care providers in the weeks leading up to the suicide attempt. Primary care visits should be an opportunity to intervene; however, there is evidence that depression in older adults is often unrecognized in this setting, which is particularly problematic for men and members of ethnic minority groups. Recent data indicate that the accuracy of diagnosis of depression in primary care usually improves with repeated visits, and it is recommended that those patients suspected to have depression should be reassessed rather than have one-off visits; the use of rating scales can also aid the diagnosis.

Even if depression is diagnosed, only a minority of patients receive successful treatment despite robust evidence for the efficacy of treatments such as antidepressant medications and psychotherapy in old age. In recent years, the rate of antidepressant prescribing in primary care has increased substantially, but many depressed older adults are started on an antidepressant with little follow-up. As a result, many stop taking medication prematurely, stay on inadequate doses for too long or do not change medication or receive psychotherapy if an initial trial of an antidepressant fails.

Despite most practice guidelines recommending that primary care providers should play a central role in the detection of cognitive impairment and dementia, there is a concerning lack of diagnosis of dementia in this setting, with rates of underdiagnosis ranging between 50% and 80%. Thus, patients with declining cognitive ability, usually in great need of treatment and advice on planning for the future, are not able to take advantage of the interventions available to them. The lack of diagnosis is partly due to a lack of recommendations for systematic screening of all at-risk patients and partly due to a lack of comfort on the part of primary care physicians when it comes to mental health in geriatric patients since their training does not emphasize geriatrics or mental health. Further, when dementia is diagnosed, the diagnosis is often undisclosed to patients and their caregivers, and treatment is not offered —largely due to providers' lack of confidence and comfort with the pharmacologic options available.

Older adults may have difficulty accessing specialty mental health care, especially when it is provided only in off-site specialty clinics. This problem is only compounded by the dearth of geriatric psychiatrists: by the year 2030, there will be only one such specialist per 20,195 older Americans. Some improvements in access have been made in recent years, such as the Medicare Improvement for Patients and Providers Act of 2008, reducing the copayment for outpatient mental health services to 20%, down from 50%, but barriers in private insurance plans regarding psychiatric medications and mental health services such as psychotherapy still exist.

Efforts to improve the care of common mental disorders in primary care have included training of primary care providers, development of treatment guidelines for depression or dementia, and the incorporation of routine screening into primary care practice as recommended for depression by the National Preventive Services Task Force. Attention must be paid to medical education—encouraging physicians to specialize in geriatric medicine and psychiatry, making an effort to promote geriatric education in medical schools and residencies, and even changing the structure of medical textbooks that are often organized in chapters covering a single-disease entity, rather than teaching management of patients with multimorbidity. Clinical guidelines should be redesigned to address multimorbidity. All these measures are necessary, but they should be accompanied by promoting effective models of care.

The present overarching organization of primary care is designed for short visits meant to take care of one problem at a time—and geriatric patients with mental health problems such as depression or dementia, especially those with several comorbid medical conditions or frailty, fall so far outside of such a simplistic paradigm that there is no schema in which to tackle their complex presentations. Collaborative models of care have been demonstrated to substantially improve clinical outcomes. In these approaches, mental health specialists collaborate with primary care providers to manage individuals with comorbid medical and mental disorders within the primary care setting. Such interventions have included systematic case finding using simple screening instruments such as the Patient Health Questionnaire-9 (PHQ-9), patient and family education and activation, clinician education and support, systematic assessment and treatment matching, proactive outcomes tracking, and integration of primary care and specialty mental health services. These models may use a geriatric psychiatry care manager with special expertise in the recognition and treatment of common mental disorders or a mental health specialist such as a psychologist or psychiatrist working with a primary care clinic to provide consultation and treatment on site. On-site services can reduce barriers to access and communication between primary and specialty providers, maintain the comfort of patients accustomed to receiving their care in a single setting, and result in better care for both medical and psychiatric problems. In a similar fashion, geriatric psychiatrists have also developed successful models for integrated care provided on-site in nursing homes. Psychiatrists and other mental health specialists working in these general medical settings can also help refer patients to further specialty mental health care as needed or link them to helpful community services. Innovative programs built to provide psychosocial skills, integrated illness self-management, collaborative care, and behavioral health homes are becoming the current evidence-based approaches to providing the best care for older adults with serious mental illness (Bartels).

This kind of collaborative care may be especially important for older adults with multimorbidity, the prevalence of which is increasing in adults aged 45 and over, with nearly 70% of Medicare beneficiaries having two or more chronic conditions such as diabetes, arthritis, and heart disease, and corresponding increases in health care costs. Depression is more likely to be present in people with multimorbidity, compounding the negative consequences of chronic physical conditions and lending importance to the need for addressing both medical and mental disorders. Multimorbidity requires a comprehensive treatment approach that addresses all problems present in an older adult patient seamlessly, accounting for interactions between treatments and conditions and the patient's own preferences. The National Quality Forum, in its publication of the Multiple Chronic Conditions Measurement Framework, proposed focusing on patient-centered quality measures such as optimizing function, ascertaining patient-important outcomes, and avoiding inappropriate care. There is evidence that organizational interventions, just as patient-oriented interventions, hold promise in addressing multimorbidity.

Centers for Medicare and Medicaid Services and private health insurance plans should redesign payment from focusing on specific diseases and procedures to value-based quality services promising relief for individuals with multiple medical and mental health problems. In this regard, accountable care organizations (ACOs) and patient-centered medical homes (PCMHs) hold promise. Through their emphasis on high-quality primary care, ACOs and PCMHs may help transform the failing health care system.

The PCMH is a model of health care delivery of a “medical home” where a primary care provider will design and coordinate comprehensive, evidence-based, cost-effective care for the patient. All aspects of care, including end-of-life care and mental health care, are coordinated and provided through the PCMH. An essential component of the success of the PCMH is a trusting relationship between the primary care provider and the patient. While this is a new and emerging model of health care, emerging data encourages optimism—we can expect that the PCMH will increase patient satisfaction and decrease health care costs. Following the paradigm of providing quality coordinated care, the last few years have seen the emergence of the ACO, defined by the Centers for Medicare and Medicaid Services as “an organization of health care providers that agrees to be accountable for the quality, cost, and overall care of Medicare beneficiaries enrolled in the traditional fee-for-service program who are assigned to it.” Similar to the PCMH, the ACO has a strong basis in primary care. The ACO is also driven by intertwining the quality and cost of care and by magnifying outcome improvements. As new PCMHs and ACOs continue to emerge, and already existing organizations continue to improve, it should be remembered that providing care for a whole patient, geriatric or not, means attending to their psychosocial concerns, including mental health, substance abuse, living and family situation, and caregivers if applicable (for patients with physical and mental disabilities, dementia, etc.). Financial incentives that reward addressing behavioral health care on par with medical problems, and reward for improved outcomes rather than fee-for-service approaches, must be implemented.

Finally, to provide the best possible care for patients with psychiatric and medical comorbidities during their hospital stay, collaborative interdisciplinary care is necessary and is best accomplished by the creation of an inpatient psychiatric consultation–liaison service. Such hospital-based service can provide effective early management of affective and cognitive disorders, help with detection and management of delirium, perform risk assessments, provide advocacy, and address psychosocial and ethical issues.

FURTHER READINGS

American Geriatrics Society 2015 Beers Criteria Update Expect Panel. American Geriatrics Society 2015 updated Beers Criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc*. 2015;63:2227–2246.

Ardoino I, Franchi C, Nobili A, Mannucci PM, Corli O; REPOSI Investigators. Pain and frailty in hospitalized older adults. *Pain Ther*. 2020;9:727–740.

Bartels SJ, DiMilia PR, Fortuna KL, Naslund JA. Integrated care for older adults with serious mental illness and medical comorbidity: evidence-based models and future research directions. *Clin Geriatr Med*. 2020;36:341–352.

Borisovskaya A, Bryson WC, Buchholz J, Samii A, Borson S. Electroconvulsive therapy for depression in Parkinson’s disease: systematic review of evidence and recommendations. *Neurodegener Dis Manag*. 2016;6(2):161–176.

Cai W, Mueller C, Li YJ, Shen WD, Stewart R. Post-stroke depression and risk of stroke recurrence and mortality: a systematic review and meta-analysis. *Ageing Res Rev*.

2019;50:102–109.

Cannon JA, Moffitt P, Perez-Moreno AC, et al. Cognitive impairment and heart failure: systematic review and meta-analysis. *J Card Fail*. 2017;23(6):464-475.

Charlesworth CJ, Smit E, Lee DSH, Alramadhan F, Odden MC. Polypharmacy among adults aged 65 years and older in the United States: 1988–2010. *J Gerontol A Biol Sci Med Sci*. 2015;70:989–995.

Domenichiello AF, Ramsden CE. The silent epidemic of chronic pain in older adults. *Prog Neuropsychopharmacol Biol Psychiatry*. 2019;93:284–290.

Felde G, Engeland A, Hunskaar S. Urinary incontinence associated with anxiety and depression: the impact of psychotropic drugs in a cross-sectional study from the Norwegian HUNT study. *BMC Psychiatry*. 2020;20:521.

Freid VM, Bernstein AB, Bush MA. Multiple chronic conditions among adults aged 45 and over: trends over the past 10 years. *NCHS Data Brief*. 2012;(100):1–8.

Friedman JH. Pharmacological interventions for psychosis in Parkinson's disease patients. *Expert Opin Pharmacother*. 2018;19(5):499–505.

Gentry MT, Lapid MI, Rummans TA. Geriatric telepsychiatry: systematic review and policy considerations. *Am J Geriatr Psychiatry*. 2019;27(2):109–127.

Greaves D, Psaltis PJ, Ross TJ, et al. Cognitive outcomes following coronary artery bypass grafting: a systematic review and meta-analysis of 91,829 patients. *Int J Cardiol*. 2019;289:43–49.

Hackett ML, Anderson CS, House A, Xia J. Interventions for treating depression after stroke. *Cochrane Database Syst Rev*. 2008;4:CD003437.

Halen NV, Cukor D, Constantiner M, Kimmel PL. Depression and mortality in end-stage renal disease. *Curr Psychiatry Rep*. 2012;14(1):36–44.

Hanagasi HA, Tufekcioglu Z, Emre M. Dementia in Parkinson's disease. *J Neurol Sci*. 2017;374:26–31.

Hossain MM, Tasnim S, Sultana A, et al. Epidemiology of mental health problems in COVID-19: a review. *F1000Res*. 2020;9:636.

Johnson JM, Nachtigall LB, Stern TA. The effect of testosterone levels on mood in men: a review. *Psychosomatics*. 2013;54(6):509–514.

Khow KSF, Visvanathan R. Falls in the aging population. *Clin Geriatr Med*. 2017;33:357–368.

Kojima G, Taniguchi Y, Iliffe S, Walters K. Frailty as a predictor of Alzheimer disease, vascular dementia, and all dementia among community-dwelling older people: a systematic review and meta-analysis. *J Am Med Dir Assoc*. 2016;17:881–888.

Maharani A, Dawes P, Nazroo J, Tampubolon G, Pendleton N; Sense-Cog WP1 Group. Visual and hearing impairments are associated with cognitive decline in older people. *Age Ageing*. 2018;47:575–581.

Mayeda ER, Whitmer RA, Yaffe K. Diabetes and cognition. *Clin Geriatr Med*. 2015;31(1):101–115.

Motter FR, Paniz ÉV, Paniz VMV, Fritzen JS. Potentially inappropriate medication in the elderly: a systematic review of validated explicit criteria. *Eur J Clin Pharmacol*. 2018;74:679–700.

Nagata T, Shinagawa S, Nakajima S, Noda Y, Mimura M. Pharmacological management of behavioral disturbances in patients with Alzheimer's disease. *Expert Opin Pharmacother*. 2020;21(9):1093–1102.

O'Mahony D, O'Sullivan D, Byrne S, O'Connor MN, Ryan C, Gallagher P. STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. *Age Ageing*. 2015;44:213–218.

Orsini A, Corsi M, Santangelo A, et al. Challenges and management of neurological and psychiatric manifestations in SARS-CoV-2 (COVID-19) patients. *Neurol Sci*. 2020;41(9):2353–2366.

Paolucci S. Advances in antidepressants for treating post-stroke depression. *Expert Opin Pharmacother*. 2017;18(10):1011–1017.

Reeve E, Trenaman SC, Rockwood K, Hilmer SN. Pharmacokinetic and pharmacodynamic alterations in older people with dementia. *Expert Opin Drug Metab Toxicol*. 2017;13(6):651–668.

Ryan M, Eatmon CV, Slevin JT. Drug treatment strategies for depression in Parkinson disease. *Expert Opin Pharmacother*. 2019;20(11):1351–1363.

Sugimoto T, Sakrai T, Ono R, et al. Epidemiological and clinical significance of cognitive frailty: a mini review. *Ageing Res Rev*. 2018;44:1–7.

Whitson HE, Cronin-Golomb A, Cruickshanks KJ, et al. American Geriatrics Society and National Institute on Aging bench-to-bedside conference: sensory impairment and cognitive decline in older adults. *J Am Geriatr Soc*. 2018;66(11):2052–2058.

Xu X, Mishra GD, Jones M. Evidence on multimorbidity from definition to intervention: an overview of systematic reviews. *Ageing Res Rev*. 2017;37:53–68.

Zhang W, Low LF, Schwenk M, Mills N, Gwynn JD, Clemson L. Review of gait, cognition, and fall risks with implications for fall prevention in older adults with dementia. *Dement Geriatr Cogn Disord*. 2019;48:17–29.

Zhang Y, Chen Y, Ma L. Depression and cardiovascular disease in elderly: current understanding. *J Clin Neurosci*. 2018;47:1–5.

55.3c Sleep Disorders in Older Adults

Jana R. Cooke, M.D., and Sonia Ancoli-Israel, Ph.D.

Up to half of older adults report chronic difficulties with their sleep. A number of changes occur in sleep architecture and sleep patterns with normal aging. However, problems with sleep are not an inevitable part of aging and are associated with significant morbidity. Sleep disturbances in older adults are generally related to specific sleep disorders, comorbid conditions, or both.

SLEEP AND AGING

Older adults report several subjective sleep complaints, including difficulty falling asleep, waking up often during the night, waking up too early in the morning, and needing to nap during the daytime or evening. Studies using polysomnography (PSG) to objectively measure sleep have confirmed that older people spend more time in bed, have more awakenings at night (decreased sleep efficiency), and take longer to fall asleep compared with their younger counterparts. The proportion of time spent in the lighter stages of sleep (stages 1 and 2) is greater, and the amount of slow-wave (stage 3) and rapid eye movement (REM) sleep is decreased in older adults compared to young adults. Almost all of these changes in sleep architecture, however, occur between early and middle adulthood. A meta-analysis that included 65 studies and 3,577 subjects ranging from 5 to 102 years in age suggested that sleep efficiency (the percentage of time in bed spent sleeping) alone continues to decrease after the age of 60 years, whereas the other sleep parameters remain relatively stable. Older adults are, on average, sleepier than their younger counterparts during the daytime when tested objectively, suggesting that they are less likely to obtain an adequate amount of restorative sleep at night.

CONSEQUENCES OF POOR SLEEP IN OLDER ADULTS

Insufficient sleep is associated with significant morbidity and increased mortality. Patients with difficulty sleeping report decreased physical health status, mental well-being, and quality of life compared to those without sleep difficulties. In older adults, subjectively and objectively assessed sleep problems are associated with an increased risk of falls, along with ambulation, balance, and visual difficulties—even after controlling for medication use. Chronic sleep difficulties at any age can lead to deficits in attention, response times, short-term memory, and performance level. In older adults, subjective reports of too little sleep (less than 4.5 hours per night), objectively measured falling to sleep difficulty, and fragmented sleep have all been associated with poor cognition. Accordingly, poor sleep is also associated with an increased risk for long-term care placement. Both reported sleep difficulties and objectively measured sleep disturbances are associated with increased mortality. In a longitudinal study of 185 healthy older adults, lower sleep efficiency (less than 80%) increased the risk of all-cause mortality almost twofold. In another large prospective study, women over the age of 69 who reported sleeping 9 to 10 hours in a 24-hour period (including napping) were almost 60% more likely to die from cardiovascular or other noncancer causes compared to women who slept between 7 and 8 hours.

SLEEP–WAKE DISORDERS IN OLDER ADULTS

Contributors to sleep disturbances in older adults can include comorbid medical or psychiatric illnesses, medications, circadian rhythm disturbances, psychosocial changes, and/or certain specific sleep–wake disorders.

Comorbid Medical and Psychiatric Illnesses

Complaints of sleep difficulties are strongly correlated with complaints about health and depression. In the 2003 National Sleep Foundation survey of adults aged 65 years and older, those with more medical conditions reported significantly more sleep complaints. Approximately 67% of older adults have multiple comorbidities, including cardiovascular disease, pulmonary disease, cancer, diabetes, and arthritis, and as the number of medical conditions increases, the likelihood of having sleep difficulties increases. The physical discomfort and emotional distress from chronic medical diseases contribute to an increased number of nighttime awakenings and daytime sleepiness which subsequently can have a negative impact on these medical illnesses. Many chronic medical conditions are also associated with an increased prevalence of sleep disorders, including insomnia, sleep apnea, and restless legs syndrome (RLS).

Depression and anxiety both commonly occur in older adults, and both mental health conditions coexist with insomnia. Older adults are more likely to experience major life stressors such as the death of a spouse, medical illness/injury, functional decline, and lack of social contact, all of which can increase their risk for developing depression and anxiety. In a large cross-sectional survey, insomnia was present in 65% of respondents with major depression, 61% of those with panic disorder, and 44% of those with generalized anxiety disorder. Longitudinal studies have also found that insomnia may increase the risk of depression in older adults.

Table 55.3c–1.
Four Cardinal Sleep Rules

1. Do not spend too much time in bed
2. Keep a consistent waketime
3. Get out of bed if you can’t sleep
4. Don’t get into bed unless you are sleepy

Reprinted with permission from Ancoli-Israel S. *All I Want is a Good Night’s Sleep*. Mosby; 1996.

Medications

Medications used to treat chronic medical and psychiatric conditions can contribute to or even cause sleep disruptions. Nearly 9 of 10 older adults chronically take prescription medications, with more than one-third of them taking more than five medications daily. Medications known to disturb sleep include beta blockers, bronchodilators, corticosteroids, decongestants, antidepressants, and diuretics. Clinicians should advise patients to take sedating medications prior to bedtime and stimulating medications and diuretics earlier in the day to avoid further sleep disruptions.

INSOMNIA DISORDER

Insomnia is one of the most common medical and sleep complaints, and it is frequently comorbid with most medical and psychiatric disorders. The diagnosis requires three main features: persistent sleep difficulty, adequate sleep opportunity, and associated daytime impairment. For insomnia to be considered chronic, the symptoms must occur at least three times per week for at least 3 months. Patients may complain of difficulty falling and/or staying asleep or early morning wakening. All other etiologies (e.g., primary sleep disorder, comorbid medical or psychiatric conditions, medications) must be ruled out before making this diagnosis.

Treatment of Insomnia

Cognitive behavioral therapy for insomnia (CBT-I) and pharmacotherapy are the main treatment options for chronic insomnia. A mounting body of research has found that in older adults, the risk of sleep medication side effects (including impaired balance, dizziness, grogginess, and confusion) and the potential risk of long-term effects on cognition is significant. In contrast, CBT-I is safe, the effects are long lasting, and it is endorsed as a first-line treatment for chronic insomnia by multiple societies and guideline panels.

Table 55.3c–2.
Sleep Hygiene

1. Restrict naps to 30 minutes in the late morning or early afternoon
2. Exercise regularly
3. Spend time outside each day
4. Eat a light snack (i.e., milk, bread) before bed
5. Avoid caffeine, tobacco, and alcohol after lunch
6. Limit liquids in the evening
7. Keep the bedroom dark and cool
8. Avoid electronics, especially blue light, 2 hours prior to bedtime

Behavioral Treatments. CBT-I has been shown to have excellent short- and long-term effectiveness in both younger and older adults. This multicomponent approach addresses common thoughts and maladaptive behaviors that interfere with sleep. A typical program consists of 6 to 10 weekly sessions administered by a specially trained clinician. Behavioral components include establishing a strict bedtime/wake time, reduction in time in bed, and sleep hygiene. Cognitive restructuring involves changing unrealistic beliefs or irrational fears about sleep and/or the loss of sleep. Finally, patients are educated about sleep behaviors that promote sleep and those practices that may be interfering with their sleep (Tables 55.3c–1 and 55.3c–2).

The use of CBT-I has been limited by a lack of trained clinicians and the requirement of 6 to 10 sessions for therapy. An abbreviated form of CBT-I called brief behavioral therapy for insomnia (BBTI) was created to overcome these impediments. A course of BBTI consists of four sessions, three of which last 30 minutes or less, some by phone, and can be administered by nurses within the primary care setting. BBTI is exclusively focused on behavior with the core components of stimulus control and sleep restriction. Several randomized trials have shown BBTI to be effective in older adults.

Pharmacologic Treatments. In some cases, it may be appropriate to prescribe medications to treat insomnia (Table 55.3c–3). Outcome studies have demonstrated that CBT-I, in combination with medication, is superior for the treatment of insomnia when compared to medication alone. With this approach, sleep aids provide short-term relief for insomnia, while CBT-I (or BBTI) provides longer-lasting benefits.

In 2019, The American Geriatrics Society Beers Criteria for potentially inappropriate medication use in older adults issued strong recommendations for the avoidance of most medications used for sleep, including benzodiazepines, antidepressants, nonbenzodiazepines/benzodiazepine receptor agonists, and first-generation antihistamines.

Table 55.3c–3.
U.S. Food and Drug Administration–Approved Non-benzodiazepines for Insomnia

Agent	Indication	Required Time in Bed (hours)	[a]Medication Appears in Beers Criteria Recommendations
Eszopiclone (Lunesta)	Sleep initiation/maintenance	8+	Yes
Zaleplon (Sonata)	Sleep initiation	4	Yes
Zolpidem (Ambien)	Sleep initiation	7–8	Yes
Zolpidem MR (Ambien CR)	Sleep initiation/maintenance	7–8	Yes
Zolpidem SL (Edluar)	Sleep initiation	7–8	Yes
Zolpidem oral spray (Zolpimist)	Sleep initiation	7–8	Yes
Zolpidem SL (Intermezzo)	Sleep initiation	4	Yes
Ramelteon (Rozerem)[b]	Sleep initiation	—	No
Doxepin (Silenor)	Sleep maintenance	—	Yes
Suvorexant (Belsomra)	Sleep initiation/maintenance	7	No
Lemborexant (Dayvigo)	Sleep initiation/maintenance	7	No

[a] See text for specific Beer’s Criteria cautions regarding potentially inappropriate use of sleep medications in older adults.

[b] Not DEA controlled.

The selective short-acting nonbenzodiazepines (type 1 gamma-aminobutyric acid [GABA] benzodiazepine receptor agonists), including eszopiclone, zaleplon, and zolpidem, have been shown to be effective in the short-term management of insomnia in older adults. Despite an overall improved side-effect profile, safety concerns remain, including increased risk for falls, delirium, and complex sleep-related behaviors. The 2019 Beers Criteria strongly recommends avoidance of this class of medications in older adults.

Although in the past, benzodiazepines were commonly prescribed for insomnia in older adults, there are significant risks associated with this class of medications, including an increased risk of falls, cognitive impairment, and driving accidents. In 2019, the Beers Criteria also strongly recommended the avoidance of benzodiazepine use in older adults.

A small number of industry-sponsored studies suggest that low doses of the tricyclic antidepressant (TCA) doxepin (≤6 mg/day) may be effective and safe for the promotion of sleep maintenance in older adults. The 2019 Beers Criteria strongly recommended avoiding doses >6 mg/day in older adults due to concerns about its significant anticholinergic and sedating properties and its tendency to induce orthostatic hypotension. A strong recommendation to avoid the use of doxepin at any dose in older adults who are at risk for delirium or who have neurocognitive disorders was also included.

Ramelteon, a melatonin receptor, appears to be safe in older adults, although its effectiveness is limited. Orexin receptor antagonists (suvorexant, lemborexant) are the newest class of FDA-approved sleep medications. These are approved for use in older adults with insomnia, and suvorexant, in particular, is approved for use in patients with Alzheimer disease (AD). But because both meds are relatively new, the experience with them is limited.

A variety of other classes of medications, including antihistamines, antidepressants, antipsychotics, and anticonvulsants, are prescribed for insomnia. As per the 2005 NIH State of the Science Conference on the Manifestations and Management of Chronic Insomnia in Adults, there is no systematic evidence to support this practice, and the risks largely outweigh the benefits in older adults.

CIRCADIAN RHYTHM SLEEP–WAKE DISORDERS

In humans, the sleep–wake cycle is controlled by an endogenous clock operating over an approximate 24-hour period called the circadian rhythm. This rhythm is synchronized by external time cues (“zeitgebers”), such as the light–dark cycle, which serves as the major external time signal in humans. Circadian rhythm sleep–wake disorders (CRSDs) may develop when dyssynchrony between the endogenous clock and external environmental zeitgebers occurs.

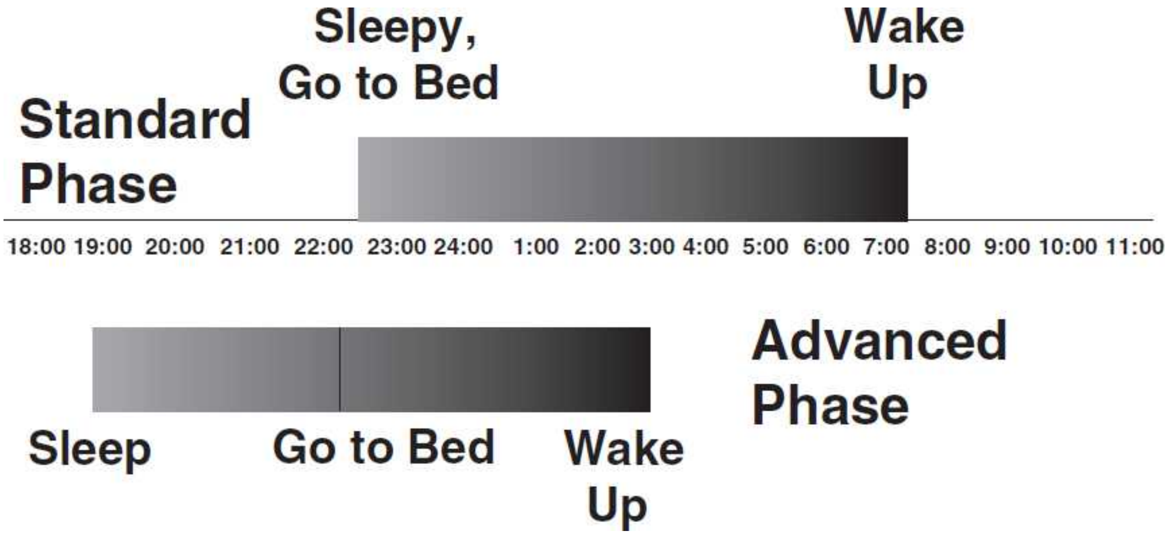

FIGURE 55.3c–1. Standard phase of sleep versus advanced phase of sleep.

Several factors associated with aging can contribute to the development of CRSDs. The circadian pacemaker, located in the suprachiasmatic nucleus, degenerates with age, which may result in weaker rhythms. The nocturnal secretion of endogenous melatonin, which plays an important role in the sleep–wake cycle, gradually decreases with age, resulting in disrupted circadian rhythms. In addition, environmental cues needed to synchronize the circadian rhythm of the sleep–wake cycle may be absent or weak in older patients. For example, one study showed that daily bright light exposure averaged about 60 minutes for healthy older adults, 30 minutes for patients with AD living at home, and 0 minutes for nursing home patients. Studies of sleep in institutionalized patients have found that this low level of light exposure is associated with sleep fragmentation and CRSDs. Gradual decreases in the amplitude of the circadian rhythm are also associated with aging and may lead to an increase in the frequency of nighttime awakenings and the severity of daytime sleepiness.

The most common CRSD experienced by older adults is advanced sleep phase syndrome (ASPS), which is characterized by sleep and wake times that are several hours earlier than societal norms. Individuals with ASPS will typically get sleepy in the early evening, timed with the drop in body temperature, and wake in the early morning (Fig. 55.3c–1). Although sleep itself may be normal, delays in sleep onset due to obligations or societal customs may result in insufficient sleep time (as these individuals experience an early-morning awakening regardless of what time they go to bed) and sleepiness during the day. The prevalence of ASPS is about 1% in middle-aged and older adults and increases with age. Men and women are equally affected. Careful and detailed sleep history and sleep diary are required to make this diagnosis. Activity monitoring with wrist actigraphy, if available, can be helpful as well.

Treatment of Advanced Sleep Phase Syndrome

For older adults with ASPS, treatment can be simply to adopt an earlier bedtime which coincides with the onset of sleepiness, to provide for enough sleep. For many patients, however, this may not be desirable due to interference with lifestyle. Treatment of CRSDs focuses on strengthening and synchronizing the sleep–wake cycle. Bright light is the strongest cue for circadian entrainment and is, therefore, the most common and the most effective treatment for circadian rhythm shifts. For older patients with ASPS, including those who reside in nursing homes, evening light exposure has been shown to strengthen the sleep–wake cycle and delay circadian rhythms. Patients with ASPS should be advised to avoid bright light during the morning and should be educated on sleep hygiene practices, including adherence to a strict sleep schedule. Early-evening artificial bright light via a bright light box has also been shown to improve sleep continuity in both healthy and institutionalized older patients.

There is no evidence to support morning melatonin therapy to delay the sleep phase in ASPS, especially given its sedating properties.

OBSTRUCTIVE SLEEP APNEA–HYPOPNEA

Obstructive sleep apnea (OSA)–hypopnea syndrome describes a spectrum of intermittent respiratory events occurring during sleep, ranging from primary snoring to increased respiratory effort without an increase in airflow, which results in arousal from sleep (flow-limitation events, typically called upper airway resistance), to substantial decreases in airflow (hypopneas), to complete cessation of airflow (apneas), or some combination of these events.

The prevalence of OSA increases or stabilizes with age, per most longitudinal and cross-sectional studies. In a large series of randomly selected community-dwelling older adults 65 to 95 years of age, researchers reported that 81% of the study subjects had an apnea–hypopnea index (AHI) of 5 or greater, with prevalence rates of 62% for an AHI of 10 or greater, 44% for an AHI of 20 or greater, and 24% for an AHI of 40 or greater. In a longitudinal study, the prevalence only changed with associated changes in body mass index. The Sleep Heart Health Study, which reported prevalence rates of OSA by 10-year age groups, found increasing rates of OSA with increasing age for those subjects with an AHI of 15 or greater.

Institutionalized older adult patients appear to have even higher rates of OSA than those older adult patients who live independently. Patients with dementias involving degeneration in brainstem areas responsible for regulating respiration and other autonomic functions relevant to sleep maintenance may have a higher risk of developing OSA. Institutionalized older adult patients with more severe dementia have been reported to have more severe OSA than those older patients with milder or no dementia. This study also reported that patients with more severe OSA had lower cognition scores during testing, suggesting that more severe OSA was associated with more severe dementia.

Risk factors for OSA in older adults are similar to those found in the younger population, including increasing age, gender, and obesity. Other conditions that increase the risk of developing OSA include the use of sedating medications, alcohol consumption, family history/genetic factors, menopausal status, race, smoking, and craniofacial morphology.

Snoring and excessive daytime sleepiness (EDS) are the predominant symptoms of OSA in older adults. Less common presentations include insomnia, particularly sleep maintenance complaints, nocturnal confusion, nocturia, and daytime cognitive impairment, such as difficulty with concentration and short-term memory loss. Unintentional dozing is a common manifestation of untreated OSA in older patients. As with younger adults, excessive sleepiness can lead to social, occupational, and cognitive difficulties, which is of particular relevance in patients with underlying cognitive impairment at baseline.

The presence of OSA is associated with significant morbidity, including cardiovascular disease, depression, and cognitive dysfunction. The presence of untreated OSA has been linked to an increased risk of developing dementia and mild cognitive impairment. Older patients with severe OSA (defined as an AHI of 30 or greater) have consistently been found to have impairments in attention, concentration, memory consolidation, immediate and delayed recall, executive tasks, planning and sequential thinking, visuomotor skills, and manual dexterity. Even older adults with milder OSA may experience similar deficits. In a substudy of the Study of Osteopathic Fractures, those older women (mean age 82) without dementia with OSA (AHI of 15 or greater) were found to have an increased risk of developing cognitive impairment. In middle-aged and older adults, research across a variety of studies suggests that OSA is associated with an increase in AD biomarkers (amyloid-beta and tau) as measured in cerebrospinal fluid, by positron emission tomography and in blood serum.

OSA is an independent risk factor for the development of systemic hypertension; it is one of the only known causes of hypertension. The exact relationship between OSA and other cardiovascular diseases (including myocardial infarction, pulmonary hypertension, stroke, arrhythmias, and congestive heart failure) in older adults remains unclear, although results from a few recent studies have provided some insight. The Sleep Heart Health Study found a positive association between the severity of OSA and the risk of developing cardiovascular disease, including coronary artery disease, congestive heart failure, and stroke. OSA is highly prevalent in patients who have suffered an ischemic stroke. The American Heart Association and American Stroke Association now recommend screening patients who have experienced an ischemic stroke or transient ischemic attack for sleep apnea. One recent study found an OSA prevalence of nearly 62% in patients with a stroke/transient ischemic attack history.

As with younger adults, clinicians should use a stepwise approach in evaluating older adult patients with suspected OSA. A complete sleep history focused on symptoms of OSA, including EDS, unintentional napping, snoring, trouble with maintaining sleep, and symptoms of other sleep disorders (e.g., RLS, insomnia), should be obtained, preferably in the presence of a bed partner or caregiver. The patient's medical history should be reviewed, with attention paid to comorbid medical conditions, medications, the use of alcohol and/or tobacco products, and evidence of cognitive impairment. If these are suggestive of OSA, an overnight sleep recording should be obtained for diagnostic purposes. Both unattended ("home") and attended (in the laboratory) sleep studies can be employed to make the diagnosis of OSA.

Treatment of Obstructive Sleep Apnea–Hypopnea Syndrome

Continuous positive airway pressure (CPAP) is the gold standard and most effective treatment of OSA and has been shown to improve cognition functioning, cardiovascular disease outcomes, and EDS in older adults. CPAP compliance can be an issue for any patient with OSA, regardless of age. Older adults may need additional education and assistance with mask fit and comfort features of CPAP initially. Studies support the idea that patients with mild and moderate AD can also be compliant with CPAP treatment, as the only factor associated with poor compliance was the presence of depression—not age, the severity of dementia, or the severity of OSA. There is also preliminary evidence to suggest that CPAP treatment may normalize AD biomarkers.

Alternative treatments include the use of an oral appliance, upper airway surgery, and a hypoglossal nerve stimulator ("pacemaker"), which is surgically implanted in the chest wall. Clinicians should counsel all patients on weight loss, smoking cessation, and abstinence from the consumption of alcohol. Sedating medications, especially those known to result in respiratory depression, including benzodiazepines, muscle relaxers, and opioids, should be avoided in older adults with OSA.

Most experts agree that the treatment of OSA in older adults should be guided by the patient's symptoms and the severity of the OSA. If an older adult has comorbid conditions such as hypertension, depression, cognitive dysfunction, or cardiovascular disease or if the OSA is severe, treatment should be considered. Age alone should not be a determining factor in treatment decisions.

RESTLESS LEGS SYNDROME/PERIODIC LIMB MOVEMENTS IN SLEEP

RLS is a common sleep-related movement disorder characterized by an urge to move the legs or leg dysesthesia (“a creepy crawling” or “restless” sensation) that occurs while the person is in a relaxed awake, typically inactive state, particularly in the evening, and is transiently relieved by movement. During sleep, many patients with RLS have characteristic limb movements, called periodic limb movements in sleep (PLMS), which may or may not be associated with sleep disruption. When the PLMS results in sleep disturbance, the diagnosis is periodic limb movement disorder (PLMD), which can result in EDS.

The pathophysiology of RLS remains incompletely understood, although both central and peripheral nervous system abnormalities have been identified. Iron stores and alterations in dopaminergic systems have both been implicated in RLS.

Epidemiologic studies reveal the prevalence of RLS in older adults ranges from 9% to 20% and increases with age. Studies have also found that RLS occurs twice as often in older women than in older men. Comorbid conditions associated with RLS include iron deficiency, end-stage renal disease, neuropathy, fibromyalgia, diabetes, Parkinson disease, and multiple sclerosis.

Like RLS, the prevalence of PLMS increases with age, with a reported prevalence of up to 45% in community-dwelling adults older than the age of 65 years. It is equally prevalent in older men and women. Despite its high prevalence in older adults, the severity of PLMS does not appear to progress with age.

The diagnosis of RLS is a clinical one, based on history alone—PSG has no role in the diagnosis. Table 55.3c–4 lists the four essential questions that clinicians should ask. In order to make the diagnosis of RLS, the answer to all four questions needs to be “yes.” Questionnaires such as the Cambridge-Hopkins diagnostic questionnaire for RLS (CH-RLSq) and the RLS Diagnostic Index (RLS-DI) can also be helpful. Other conditions that may be mistaken for RLS, such as peripheral neuropathy, myalgia, leg cramps, positional discomfort, arthritis, pruritus, and vascular insufficiency, need to be ruled out, although many of these can coexist with RLS. In patients with significant cognitive impairment, signs of discomfort in the legs, such as rubbing the legs, which are present or worse during inactivity and which are lessened by activity, may be observed.

Table 55.3c–4.
Four Essential Questions for Restless Legs Syndrome Diagnosis[a]

1. Do you experience the urge to move your legs (or other parts of your body), typically accompanied or caused by an uncomfortable and/or unpleasant sensation in the affected body part?
2. Does the urge to move or the uncomfortable sensation begin and/or worsen during periods of rest or inactivity?
3. Is the urge to move or uncomfortable sensation partially or totally relieved by movement?
4. Is the urge to move or the uncomfortable sensation worse in the evening or at night compared to the daytime, or does it only occur in the evening or at night?

[a] All four questions should be answered in the affirmative for the diagnosis of restless legs syndrome (RLS).

In contrast, the diagnosis of PLMS requires PSG and is seen in more than 80% of patients with RLS who undergo PSG. PLMS is defined as periodic episodes of repetitive limb movements during sleep, typically separated by 20 to 40 seconds. PLMD is diagnosed when the clinical history of EDS with subjective sleep complaints is combined with PSG showing PLMS, excluding other sleep disorders.

Treatment of Restless Legs Syndrome and Periodic Limb Movement Disorder

Medications that suppress the leg movements and/or the related arousals are the currently recommended treatments for both RLS and PLMD. Nonergot dopamine agonists (ropinirole, pramipexole, and rotigotine transdermal) and alpha-2-delta calcium channel ligands (gabapentin enacarbil, pregabalin, and gabapentin) are both considered first-line treatment for RLS. Augmentation is a common complication of long-term dopaminergic therapy and is characterized by worsening RLS symptoms (earlier onset of symptoms, the spread of symptoms to the arms and torso) with increasing doses of the medication. If augmentation occurs, reducing the dose, splitting the dose into daytime/nighttime, and/or discontinuation of the medication is recommended.

In addition, supplemental iron is recommended in patients with RLS whose serum ferritin level is ≤75 mcg/L. A number of medications can exacerbate RLS, including antidepressants, neuroleptic agents, and sedating antihistamines (including those found in the nonprescription sleep aids and allergy relief medications), and should be discontinued if possible. Sleep deprivation, either from insufficient sleep time and/or due to comorbid sleep disorders, can also aggravate RLS and should be reviewed and further evaluated if indicated.

RAPID EYE MOVEMENT SLEEP BEHAVIOR DISORDER

REM sleep behavior disorder (RBD) is a parasomnia characterized by dream-enacting behavior with a reported prevalence of 2% in adults over 60 years of age and 6% in those over 70 years of age. It is more commonly diagnosed in men (9:1 male-to-female ratio), although this may represent a referral bias. The development of RBD is a harbinger of alpha-synuclein neurodegeneration. After a prolonged interval (average 10 years), most older patients with RBD will show signs and symptoms of Parkinson disease or another related disorder such as multisystem atrophy or dementia with Lewy bodies. In addition to synucleinopathies, RBD occurs in nonsynucleinopathic dementias/neurogenerative diseases, with some brainstem lesions, and with certain medications (antidepressants).

RBD is characterized by elaborate, complex, and purposeful motor activities and vocalization associated with dreams (“acting-out dreams”) due to the loss of REM sleep atonia. Patients may punch, kick, yell, or jump/run out of bed, and these actions correlate with dream imagery. The episodes typically occur during the second half of the sleep period, when REM sleep predominates. Sleep behaviors can place the patient and/or bed partner at risk for injuries, which have been reported to occur in 32% to 64% of cases.

The diagnosis of RBD requires a sleep history, in the presence of the bed partner if possible, and an overnight polysomnogram with video recording to confirm the relationship between REM sleep and complex motor behaviors. Even in the absence of abnormal behavior during the PSG recording, REM sleep atonia is characteristically present.

Treatment of Rapid Eye Movement Sleep Behavior Disorder

The primary goal of RBD treatment is to create a safe sleeping environment. Heavy or sharp items should be removed from the bed's vicinity, and doors/windows should be locked. Measures to avoid being injured by falling out of bed can be used (i.e., bedrails, a low-profile bed, or a mattress on the floor).

Melatonin is the preferred first-line therapy for patients with RBD. By an unclear mechanism, melatonin in high doses (6 to 18 mg) augments REM sleep atonia and improves RBD symptoms. Melatonin is generally well tolerated. Low-dose clonazepam (0.5 to 1 mg) is also an effective treatment for RBD, although its side effects may limit its utility, especially in older adults with neurodegenerative disorders. Maintaining good sleep hygiene and educating the patient and their bed partner are also important aspects of RBD treatment.

SLEEP IN NEUROCOGNITIVE DISORDERS

Recent studies support a possible connection between sleep disturbances, especially OSA, fragmented sleep, abnormal duration of sleep, and the development of cognitive impairment, suggesting that the sleep–wake cycle may play an important role in cognitive aging. Patients with neurocognitive disorders, both those living independently and those institutionalized, experience disrupted sleep which may be due, in part, to neurodegeneration in areas of the brain responsible for sleep regulation. It has been reported that 19% to 44% of community-dwelling patients with dementia complain of sleep disturbances, and the prevalence is estimated to be even higher in patients with dementia who are institutionalized. Disrupted sleep commonly results in napping, both intentional and unintentional, and can be manifested as abnormal nocturnal behaviors, including wandering and confusion, which are significant contributors to caregiver stress and may result in the need for patient institutionalization.

Patients with neurocognitive disorders can have primary sleep disorders such as OSA, PLMS, RLS, or RBD or may have disturbed sleep related to pain, medications, circadian rhythm changes, or depression. Therefore, disturbed sleep related to a neurocognitive disorder itself should be a diagnosis of exclusion.

As with all older adults, treatment of specific sleep disorders in older patients with neurocognitive disorders should be targeted for that specific sleep disorder. Patients with OSA should be treated with CPAP if appropriate, whereas patients with circadian rhythm disturbances should be treated with bright-light therapy. Given the high risk-to-benefit ratio of most medications in this patient population, nonpharmacologic interventions such as regular physical activity and restricting time in bed are preferred. Medications, including sedative–hypnotics, antihistamines, and antipsychotics, all have significant side effects, including residual daytime sleepiness (the so-called “hangover effect”) and motor and/or cognitive functioning impairment. Further, the efficacy and safety of sleep medications have not been adequately studied in patients with neurocognitive disorders; there is little evidence to support their use.

SLEEP IN NURSING HOMES

Institutionalized older adults experience more fragmented and disturbed sleep than those living in the community. Nursing home patients have poorer sleep quality, more fragmentation of sleep, more advanced sleep–wake cycles, and more daytime sleepiness and are prescribed sedative–hypnotics more frequently than older adults living in the community or in assisted living facilities. During any 1 hour of a 24-hour day, institutionalized patients with major neurocognitive disorders may not spend a single hour completely awake or completely asleep.

Excessive nighttime noise and inadequate ambient light exposure both contribute to sleep disruption in nursing home patients. In a study of institutionalized patients, most were found to be exposed to less than 10 minutes of bright light per day. In contrast, patients with more light exposure had fewer sleep disruptions. In addition, nursing home patients may spend large portions of their day in bed, and inactivity is known to disrupt circadian rhythms.

Institutionalized patients should have a careful sleep evaluation, including an assessment for the presence of primary sleep and circadian rhythm disorders. A thorough review of medications should be performed, and medical and psychiatric history should be taken. Reduction in nighttime light and noise increases in daytime light exposure and improvements in sleep hygiene should be implemented. The use of sleep medications should, in general, be avoided.

CONCLUSION

For many older adults, aging is associated with significant changes in sleep with a variety of potential contributors, including medications, comorbid conditions, circadian rhythm disruption, and specific sleep–wake disorders. As with younger adults, a thorough sleep history should be taken, and often an overnight sleep recording should be performed. Treatment should address the primary sleep problem and can result in significant improvement in both quality of life and daytime functioning.

FURTHER READINGS

Allen RP, Picchietti DL, Garcia-Borreguero D, et al; International Restless Legs Syndrome Study Group. Restless legs syndrome/Willis-Ekbom disease diagnostic criteria: updated International Restless Legs Syndrome Study Group (IRLSSG) consensus criteria–history, rationale, description, and significance. *Sleep Med*. 2014;15(8):860–873.

American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society updated Beers Criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc*. 2012;60(4):616–631.

Aurora RN, Zak RS, Maganti RK, et al; Standards of Practice Committee, American Academy of Sleep Medicine. Best practice guide for the treatment of REM sleep behavior disorder (RBD). *J Clin Sleep Med*. 2010;6(1):85–95.

Blackwell T, Yaffe K, Ancoli-Israel S, et al; Study of Osteoporotic Fractures Group. Poor sleep is associated with impaired cognitive function in older women: the study of osteoporotic fractures. *J Gerontol A Biol Sci Med Sci*. 2006;61(4):405–410.

Boot BP, Boeve BF, Roberts RO, et al. Probable rapid eye movement sleep behavior disorder increases risk for mild cognitive impairment and Parkinson disease: a population-based study. *Ann Neurol*. 2012;71(1):49–56.

Buysse DJ, Germain A, Moul DE, et al. Efficacy of brief behavioral treatment for chronic insomnia in older adults. *Arch Intern Med*. 2011;171(10):887–895.

Cross NE, Carrier J, Postuma RB, et al. Association between insomnia disorder and cognitive function in middle-aged and older adults: a cross-sectional analysis of the Canadian Longitudinal Study on Aging. *Sleep*. 2019;42(8);zsz114.

da Silva AA, de Mello RG, Schaan CW, Fuchs FD, Redline S, Fuchs SC. Sleep duration and mortality in the elderly, a systematic review with meta-analysis. *BMJ Open*. 2016;6:e008119.

Dam TT, Ewing S, Ancoli-Israel S, Ensrud K, Redline S, Stone K; Osteoporotic Fractures in Men Research Group. Association between sleep and physical function in older men: the osteoporotic fractures in men sleep study. *J Am Geriatr Soc*. 2008;56(9):1665–1673.

Hornyak M, Trenkwalder C. Restless legs syndrome and periodic limb movement disorder in the elderly. *J Psychosom Res*. 2004;56(5):543–548.

Kripke DF, Garfinkel L, Wingard DL, Klauber MR, Marler MR. Mortality associated with sleep duration and insomnia. *Arch Gen Psychiatry*. 2002;59(2):131–136.

Low DV, Wu MN, Spira AP. Sleep duration and cognition in a nationally representative sample of U.S. older adults. *Am J Geriatr Psychiatry*. 2019;27(12):1386–1396.

McCleery J, Cohen DA, Sharpley AL. Pharmacotherapies for sleep disturbances in Alzheimer's disease. *Cochrane Database Syst Rev*. 2014;(3):CD009178.

McGrane IR, Leung JG, St. Louis EK, Boeve BF. Melatonin therapy for REM sleep behavior disorder: a critical review of evidence. *Sleep Med*. 2015;16(1):19–26.

McSorley VE, Yu SB, Lauderdale DS. Associations of sleep characteristics with cognitive function and decline among older adults. *Am J Epidemiol*. 2019;188(6):1066–1075.

Morin CM, Bootzin RR, Buysse DJ, Edinger JD, Espie CA, Lichstein KL. Psychological and behavioral treatment of insomnia: update of the recent evidence (1998–2004). *Sleep*. 2006;29(11):1398–1414.

Mullins AE, Kam K, Parekh A, Bubu OM, Osorio RS, Varga AW. Obstructive sleep apnea and its treatment in aging: effects on Alzheimer's disease biomarkers, cognition, brain structure and neurophysiology. *Neurobiol Dis.* 2020;145:105054.

NIH State-of-the-science conference statements on *manifestations and management of chronic insomnia in adults*. *NIH Consens State Sci Statements.* 2005;22(2):1–30.

Neikrug AB, Liu L, Avanzino JA, et al. Continuous positive airway pressure improves sleep and daytime sleepiness in patients with Parkinson disease and sleep apnea. *Sleep*. 2014;37(1):177–185.

Nishihata Y, Takata Y, Usui Y, et al. Continuous positive airway pressure treatment improves cardiovascular outcomes in elderly patients with cardiovascular disease and obstructive sleep apnea. *Heart Vessels*. 2015;30(1):61–69.

Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep*. 2004;27(7):1255–1273.

Peever J, Luppi PH, Montplaisir J. Breakdown in REM sleep circuitry underlies REM sleep behavior disorder. *Trends Neurosci*. 2014;37(5):279–288.

Russell T, Duntley S. Sleep disordered breathing in the elderly. *Am J Med*. 2011;124(12):1123–1126.

Stone KL, Ewing SK, Ancoli-Israel S, et al. Self-reported sleep and nap habits and risk of mortality in a large cohort of older women. *J Am Geriatr Soc*. 2009;57(4):604–611.

Tranah GJ, Yaffe K, Nievergelt CM, et al; Osteoporotic Fractures in Men Study (MrOS) Research Group. APOE-4 and slow wave sleep in older adults. *PLoS One*. 2018;13(1):e0191281.

Troxel WM, Conrad TS, Germain A, Buysse DJ. Predictors of treatment response to brief behavioral treatment of insomnia (BBTI) in older adults. *J Clin Sleep Med*. 2013;9(12):1281–1289.

Wennberg AM, Canham SL, Smith MT, Spira AP. Optimizing sleep in older adults: treating insomnia. *Maturitas*. 2013;76(3):247–252.

Yaffe K, Falvey CM, Hoang T. Connections between sleep and cognition in older adults. *Lancet Neurol*. 2014;13(10):1017–1028.

Yaffe K, Laffan AM, Harrison SL, et al. Sleep-disordered breathing, hypoxia, and risk of mild cognitive impairment and dementia in older women. *JAMA*. 2011;306(6):613–619.

Zee PC, Attarian H, Videnovic A. Circadian rhythm abnormalities. *Continuum (Minneap Minn)*. 2013;19(1 Sleep Disorders):132–147.

55.3d Anxiety Disorders in Older Adults

HELMET T. KARIM, PH.D., AND CARMEN ANDREESCU, M.D.

In DSM-5, the chapter on anxiety disorders has been revised to include separation anxiety disorder and selective mutism—but given the lack of research in late life, we will not review these, although symptoms suggesting separation anxiety are frequently encountered in Alzheimer disease (AD) and related dementias. No longer included are obsessive-compulsive disorder (OCD) or post-traumatic stress disorder (PTSD), and acute stress disorder (which have separate chapters in DSM-5). The term “anxiety disorders” used in this chapter nonetheless is used to refer to all the DSM-5 diagnoses within anxiety-, trauma-, and stressor-related and obsessive-compulsive and related disorders. When we refer to these disorders in aggregate (i.e., when there are no specific distinctions between them that we are highlighting), we will use the term “anxiety disorders” otherwise, we will specify the distinctions.

There has been increasing evidence that anxiety in late life has greater clinical consequences and is more prevalent than previously thought. Although much less is still known about late-life anxiety than about other common late-life psychiatric conditions such as depression or dementia, current studies show that in community samples, anxiety is more common than depression in older adults. In addition, anxiety disorders in later life can greatly impact quality of life, increase disability levels, and impact recovery from illness as well as having reciprocal associations with several disorders including cardiovascular disease and dementia. When compared to the body of research of anxiety in younger adults or children, limited data suggest that older adults present with the same anxiety disorders as do younger adults, although several additional anxiety syndromes may be particularly relevant to older adults. Many anxiety disorders are thought to “carry over” from adulthood to late life, though there are some studies that show some anxiety disorders may appear de novo in late life. Virtually all epidemiologic research is consistent in finding that the prevalence of anxiety, trauma- and stress-related, and obsessive-compulsive and related disorders is lower among older adults than among younger adults, particularly with respect to panic disorder and social anxiety disorder (social phobia). Rates vary considerably by study, however, leading to contradictory conclusions: either anxiety, trauma- and stress-related, and obsessive-compulsive and related disorders in older adults are extremely rare or they constitute the most common class of geriatric psychiatric conditions, with substantially higher prevalence rates than mood disorders.

Diagnostic issues are complicated by medical illness and medications, which can cause or contribute to anxiety-like symptoms. Clinicians may also attribute symptoms to realistic, age-appropriate health problems, worries, or fears, but severe worry or anxiety can, at times, be dismissed as part of the normal aging process, which may hinder detection. Furthermore, in some disorders, like dementia, anxiety may present as agitation, which complicates diagnosis and treatment.

Although some of these factors may lead investigators to believe that anxiety disorders are uncommon in older adults, it is more likely that anxiety is underdiagnosed in the presence of medical conditions or depression. Clinicians may conclude that anxiety or avoidance is an inevitable or adaptive response to aging and, therefore, not pathologic. In many cases, this conclusion may be inaccurate. Differences in the presentation of certain anxiety disorders in older adults may also muddy the diagnostic waters. Older adults may also express anxiety in ways that are different from younger adults—in particular, assessing avoidance or worry may be difficult as older adults may regard symptoms associated with avoidance or worry as signs of weakness. Older adults may also experience more impairment with even subthreshold levels of anxiety. Finally, there is insufficient research on differences in presentation or etiology between late-onset anxiety disorders and disorders originating earlier in life that have persisted until or recurred in old age. Additional research on these topics is necessary to clarify the many unresolved issues surrounding anxiety in older adults.

DEFINITION, HISTORY, AND COMPARATIVE NOSOLOGY

The definition, history, and comparative nosology of anxiety, trauma- and stressor-related, and obsessive-compulsive and related disorder diagnoses, as defined by the various versions of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) and the *International Statistical Classification of Diseases and Related Health Problems* (ICD). To date, no geriatric-specific anxiety disorders have been included in the diagnostic nomenclature. In general, the DSM excludes anxiety disorders that are “due to the direct physiological effects of a general medical condition.”

EPIDEMIOLOGY

Prevalence

The prevalence of anxiety disorders in older adults is controversial. Investigations have found rates ranging from 0.7% to 19% in community samples and 1% to 28% in medical settings. A recent review of epidemiologic studies in Europe found a range of 4% to 14%. The prevalence of clinically relevant symptoms is much higher, ranging from 15% to 52% in community samples and 15% to 56% in clinical samples. By comparison, another recent review found prevalence rates of major depression ranging from 0.9% to 9% among community-dwelling older adults. Thus, it appears likely that, as in younger adults, anxiety disorders as a class are more common than major depression in older adults—though many of these estimates for anxiety include PTSD and OCD.

Table 55.3d–1.
Prevalence Estimates of Anxiety Disorders in Older Adults

Generalized anxiety disorder	0.7–9%
Specific phobia	3–5%
Social anxiety disorder (social phobia)	0.2–2%
Panic disorder	0.1–1%
Agoraphobia	0.4%
Obsessive-compulsive disorder	0.1–2%
Hoarding disorder	14%
Post-traumatic stress disorder	8.7%

It is generally accepted now that prevalence of anxiety disorders likely decreases among older adults, and sex differences, with women reporting higher rates than men, appear to attenuate among the oldest old. The exception, however, is generalized anxiety disorder (GAD) which maintains similar prevalence rates in younger and older adults. One study found that women were twice as likely to present with GAD or agoraphobia as men, independent of age, but did not find such an association with other anxiety disorders in community-dwelling older adults. There appears to be a lower comorbidity of anxiety disorders and alcohol abuse among older adults than among younger adults, although this may be influenced by relatively low rates of substance use disorders among older adults. Of note, the rates of prescription drug misuse have increased among older adults including opioid and benzodiazepine abuse. Most studies conclude that GAD and phobias are the most common anxiety disorders among older adults, with panic disorder and social anxiety disorder (social phobia) much less common than in younger age groups. Estimated prevalence rates for specific anxiety disorders are listed in Table 55.3d–1.

Incidence

Incidence studies of late-life anxiety are relatively rare, so most data on the age of onset come from retrospective reporting. The incidence of agoraphobia in women may increase in late life, whereas specific phobia, OCD in men, and panic disorder are typically continuations from earlier life or the result of another psychiatric or medical illness. While it is thought that most anxiety disorders are chronic and usually occur earlier in life, GAD and agoraphobia may have onset later in life. In one study, 75% of all anxiety disorders have an onset before age 21 and 95% before age 51. However, approximately half of GAD cases arise for the first time in later life after age 50. Those with an earlier onset GAD (compared to later onset) tend to have a more severe course and pathologic worry, while later onset is often characterized by more medical comorbidity and disability. One study found that agoraphobia in late life was frequently triggered by physical illness or trauma.

Prevalence Within Special Groups

Anxiety disorders are common in diseases of older adults. Rates as high as 30% have been reported in patients with cardiac conditions such as coronary heart disease and atrial fibrillation. In chronic obstructive pulmonary disease patients, prevalence rates of panic disorder range from 8% to 32%, with GAD rates ranging from 10% to 16%. Approximately 40% of patients with Parkinson disease report anxiety symptoms. Vestibular diseases have a similar comorbidity with anxiety disorders

(37% to 42%). In irritable bowel syndrome, 20% of patients have comorbid anxiety. GAD may occur in 15% to 21% of patients with neurologic conditions such as stroke and AD. GAD is thought to have reciprocal relationships with coronary heart disease and dementia, which may worsen overall outcomes, especially if they are left undiagnosed and untreated.

Although rates of anxiety disorders in special populations are typically higher, this is not the case for studies in nursing home patients, among whom the reported rate is approximately 5%. A review of prevalence data on anxiety disorders among caregivers for individuals with dementia found rates ranging from 4% to 16%. In a review and meta-analysis of studies involving widows, prevalence rates were 12% for PTSD, 10% for panic disorder, and 31% for GAD. By comparison, the prevalence of major depression was 22%.

Some data suggest that anxiety may be more common in older individuals from minority groups, although a recent study of older Black individuals found 6.6% reporting an anxiety disorder (vs. 2.4% for major depression), consistent with other research. The distribution of disorders, however, was different from what has been reported in other samples. For example, the most common anxiety disorders in this sample were PTSD (2.9%) and social anxiety disorder (social phobia) (2.1%). Another study examining race and sex differences found highest rates of GAD among older Black women and the lowest rates of GAD among older Black men. One study found that while PTSD prevalence did not differ between Black and White individuals, older White individuals who had experienced nonphysical trauma were more likely to report hyperarousal than older Black individuals. The oldest old and those in the South were less likely to report anxiety disorders. Almost one-fourth (24%) of older Puerto Rican primary care patients met criteria for an anxiety disorder, primarily specific phobia (12%), GAD (11%), or panic attacks (11%). Hispanic individuals often express distress via somatic rather than cognitive complaints, which may explain some of this discrepancy. Across all ethnic groups, phobias appear to be associated with lower income.

Although older veterans report rates of PTSD that are substantially lower than the rates found in younger veterans (e.g., 6% vs. 19% in one study), rates of PTSD are high among aging former prisoners of war (POWs) (e.g., 20%). The prevalence of PTSD appears to be highest among Holocaust survivors (e.g., 45%).

Comorbidity With Mood Disorders

As is the case with younger adults, comorbidity of anxiety and mood disorders is very common. Data from the Epidemiologic Catchment Area survey and two Dutch studies found that 21% to 37% of community-dwelling patients with an anxiety disorder (as defined by DSM-IV) experienced comorbid depression. Although these rates suggest that the disorders frequently overlap, it appears that most DSM-IV anxiety disorders in older adults are unaccompanied by a mood disorder. One study found that 74% of older adults with anxiety did not have a mood disorder, while depression without comorbid anxiety is much less common in late life. Some epidemiologic evidence in younger adults suggests that anxiety disorders precede mood disorders and may partially explain this prevalence. Some research indicates that this comorbidity may increase in certain conditions or populations (e.g., higher comorbidity in residential care facilities). Among nursing home patients, 42% of those with a DSM-IV anxiety disorder experienced comorbid major depression.

ETIOLOGY AND RISK FACTORS

Although much less research has been conducted on anxiety disorders than on depression in older adults, it appears that risk factors are similar for both conditions. Female sex, cognitive impairment, chronic health conditions, poor self-rated health, recent adverse life events, poverty, history of mental illness in parents, lower education, being a caregiver, social isolation, bereavement, adverse events in childhood, and functional limitations are among the most important factors increasing vulnerability. Many of these risk factors are shared in younger and older adults. Psychological risk factors for both anxiety disorders and depression include personality traits (i.e., neuroticism), dysfunctional coping, and other forms of psychopathology. Marital status (unmarried) appears to be associated with depression but not anxiety, being childless, having lower income, and experiencing traumatic events are more typically associated with anxiety than with depression. Older age appears to be somewhat protective for anxiety disorders. However, one may have to keep in mind that anxiety in old age may be often atypical and several symptoms do not fall under the current DSM categories. Also, of interest, one study that examined depression and anxiety disorders simultaneously found that although risk factors were similar for both classes of conditions, odds ratios associated with virtually all risks were higher for anxiety disorders than for depression.

In nursing home patients, anxiety disorders appear to be related to younger age, lower levels of social support, better cognitive performance, depression, and history of stroke. Comorbid anxiety and depressive disorders are associated with the presence of pain, loneliness, and perceived lack of adequate care.

With respect to specific anxiety disorders, female sex, marital status (divorced or widowed), and comorbid psychiatric illness (specifically major depression and social anxiety disorder [social phobia]) are risk factors for agoraphobia in older adults. Late-onset agoraphobia is often associated with medical illness or traumatic events such as victimization or falls. Panic disorder is more strongly associated with disability than with health problems. Phobic disorders are associated with lower income and early parental loss. Risk factors for PTSD include female sex, neuroticism, low self-efficacy, and childhood trauma, and there are some reports of new-onset or worsening PTSD symptoms with cognitive decline. There is some evidence that re-experiencing symptoms in PTSD declines with age. In a cross-sectional study, only 10% of individuals over age 65 reported re-experiencing symptoms compared to 21.5% in those aged 18 to 24 years.

DIAGNOSIS AND CLINICAL FEATURES

While many clinical features of anxiety in younger adults can be encountered in older adults, older adults also experience atypical anxiety syndromes (see next section).

Similar to what has been reported in younger samples, severe and difficult to control worry, sleep disturbance, fatigue, and muscle tension appear to be the cardinal features of late-life GAD. There are some differences, for instance in the nature of worries (e.g., more health related in late life). It is also widely accepted that it is more difficult to assess anxiety symptoms in late life as they often manifest as somatic symptoms and comorbid with medical conditions.

Older adults with panic disorder report fewer and less severe panic symptoms, lower levels of arousal, and less depression than younger individuals. Late-onset panic disorder appears to be associated with lower levels of cognitive and somatic distress during panic attacks. OCD may differ somewhat between older and younger patients, with older persons reporting more contamination and religious obsessions but fewer rituals involving symmetry or counting. Older individuals reported lower severity of obsessions overall (compared to younger individuals). There is strong evidence that compulsive hoarding increases with age and even more so in dementia.

Older adults appear to develop PTSD and other anxiety syndromes after exposure to traumatic events less frequently than younger adults. Older PTSD patients report more social dysfunction than younger patients, and in older adults, the external locus of control appears to be associated with more avoidance, social dysfunction, and depression. Older PTSD patients may present more somatic than psychological symptoms and may report less severe PTSD symptomatology.

Anxiety Syndromes of Particular Relevance to Older Adults

Agoraphobia, fear of falling, hoarding, and complicated grief merit special mention because they may be anxiety syndromes with particular relevance to older adults. Each of these disorders has a higher prevalence in older adults compared to younger adults.

Agoraphobia. As shown in Table 55.3d–1, prevalence rates of agoraphobia vary widely in older adults, suggesting a lack of diagnostic consensus. Distinguishing between adaptive and excessive avoidance is challenging, and patients typically believe that their behavior is reasonable, given age-associated medical conditions and physical limitations. Because panic disorder is rare in older adults, agoraphobia without history of panic disorder (DSM-IV) is the most common symptom presentation. In DSM-5, where panic disorder and agoraphobia have been diagnostically uncoupled, the diagnosis of agoraphobia would apply.

In one study of older adults that used conservative criteria that defined agoraphobia as avoidance due to a panic attack rather than panic-like symptoms, the prevalence of comorbid panic disorder was still much lower (8%) than the prevalence of major depression (25%) or social phobia (20%). This suggests that older adults may limit their activity (i.e., develop agoraphobia) more often than do younger adults in response to panic attacks that do not meet criteria for panic disorder. Only 11% of late-onset patients with agoraphobia had histories of panic attacks in another investigation with looser criteria for agoraphobia; instead, they developed their fear in response to medical illness, physical injury, or traumatic events.

Across the lifespan, clinical features of agoraphobia are fairly consistent; however, age may impact the type of situation feared. The most common situations feared in older adults include being in line, being in shops, and being in open spaces. Older individuals may fear or avoid these situations due to thoughts that help may not be available in the event of embarrassing symptoms (e.g., fear of falling, fear of incontinence).

Fear of Falling. Although under DSM-5 diagnostic criteria fear of falling can only be classified as a specific phobia, other type, if the patient perceives it to be excessive or unreasonable, it is a common and disabling anxiety syndrome in older adults. As many as 25% of community-dwelling older adults report fear of falling. Although individuals with a past fall are more likely to develop fear of falling than those without such history, most individuals reporting fear do not have a history of falls. Typical screening tools are often not well suited for identifying fear of falling so it is commonly underdiagnosed and left untreated.

Fear of falling has potentially serious consequences, including, paradoxically, increased risk of falls. Approximately half of those with fear of falling restrict their activities due to fear, and the risk of falling is highest among those who do, even among those who have never experienced a fall. Fear of falling may lead to maladaptive changes in gait parameters, further increasing the risk of future falls, and may interfere with rehabilitation following a fall. For example, in one investigation of patients recovering from hip fracture surgery, fear of falling predicted poor functional recovery, although pain and depression did not. Older patients are often more likely to receive an opioid or benzodiazepine prescription and are subsequently more likely to experience side effects including cognitive impairment, dizziness, impaired balance, and higher likelihood of falls.

As is true with other anxiety syndromes, fear of falling is frequently, although not always, comorbid with other anxiety disorders and depression. In one sample of individuals with high fear of falling and a recent fall, 36% had depression and 19% had an anxiety disorder, most commonly agoraphobia without panic, and anxiety was a strong predictor of fear of falling even after controlling for factors such as dizziness, visual impairment, number of medications, and depressive disorders and symptoms.

Hoarding. Once considered a subtype of OCD, hoarding disorder is now considered a separable syndrome and is a new diagnostic category in DSM-5. Nationally representative epidemiologic studies of hoarding disorder with DSM-5 criteria are not available, but smaller epidemiologic surveys estimate a 1.5% to 6% prevalence of hoarding behaviors (not necessarily at the level of a hoarding disorder diagnosis) in the general population. Hoarding symptoms may be nearly three times more prevalent in older adults compared with younger adults. Older adults may also show more severe hoarding symptoms than do middle-aged adults and is even more prevalent in patients with dementia.

Older adults who are hoarders are typically female, never married, and living alone. Hoarding is typically a chronic condition with an early onset and progressive course. For those with a late onset of symptoms, hoarding likely begins secondary to a disease process occurring in later life, such as dementia. Pest infestations, fire hazards, and falling are risk factors for older adults with hoarding behaviors. People with hoarding disorders often have poor physical health and a greater number of comorbid medical conditions compared to those without hoarding disorders and several factors may explain these associations including impulsivity, executive dysfunction, and emotion regulation though there is limited research.

Coronavirus Pandemic and Anxiety in Late Life. During the coronavirus disease (COVID) pandemic, older adults experienced greater adverse health effects (e.g., more complications, higher mortality), difficulties adapting to a new status quo (e.g., telemedicine, telecommunication), and greater isolation and potentially more frequent grief. There was some concern that older adults may subsequently experience greater mental health conditions, especially anxiety including agoraphobia. However, compared to younger adults, several studies demonstrated lower rates of anxiety: in August 2020, the Center for Disease Control reported lower rates of anxiety (6.2%) in older adults (>65 years) compared to younger adults (18 to 24 years, 49.1%) or midlife (45 to 64 years, 16.1%). A study from Spain replicated these findings, showing lower rates of anxiety, depression, and PTSD. However, this may be due to differences in coping, emotion regulation, and expressed anxiety, as one study showed that in a study of 7,127 older adults, 12.3% reported worsened anxiety after lockdown. Apprehension regarding COVID was also much higher compared to anticipatory apprehension related to other disorders, like influenza, prompting the term coronaphobia which may be driven by uncertainty regarding the future and the virus, significant changes in daily life due to the virus, low optimism (especially early on), poor health-related outcomes, and misinformation on social media, news sources, and other online sources. Finally, COVID survivors at times experience additional symptoms for months during the convalescent period (long COVID syndrome), including fatigue and pain, but also depression and anxiety. Thus, one study found that 3 months after recovery, 6.5% of COVID survivors experienced anxiety compared to 1.6% in a comparison group who did not have COVID. Further impact associated with social changes during the pandemic as well as with long COVID syndrome in older adults will require additional research.

PATHOLOGY AND LABORATORY EXAMINATION

Most research on the biology of late-life anxiety has focused on PTSD and, more recently, GAD. Studies in PTSD, GAD, and other late-life anxiety disorders have examined cognitive performance, hypothalamic–pituitary–adrenal (HPA) axis function, and other markers of autonomic arousal, neuroimaging, and genetics. These are summarized in Figure 55.3d–1.

Post-Traumatic Stress Disorder

With respect to neuropsychological functioning, older persons with PTSD display poorer explicit memory, visuospatial functioning, attention, inhibitory control, information processing speed, and working memory than those with trauma exposure but no PTSD and those without trauma exposure. Among those with PTSD, improvement in recall may be associated with improvement in PTSD symptoms. It is hypothesized that lack of cognitive resources to manage trauma may result in engagement of more maladaptive strategies for coping with symptoms.

Other research shows evidence of HPA axis dysregulation associated with late-life PTSD. Individuals with PTSD show flatter circadian salivary cortisol patterns and lower urinary cortisol excretion than do those without PTSD, and better explicit memory performance is associated with lower salivary cortisol levels. Urinary cortisol excretion and glucocorticoid receptor–mediated responsiveness appear to be associated with age of first trauma exposure, and cortisol level may be associated with severity of childhood trauma. Moreover, cortisol administration may enhance working memory performance in older veterans with PTSD. Glucocorticoid receptor, plasma adrenocorticotropic hormone (ACTH), and cortisol suppression in response to dexamethasone administration are greater in older PTSD patients. Older PTSD patients may have higher levels of dehydroepiandrosterone (DHEA) and DHEA sulfate ester (DHEAS) and a lower cortisol/DHEA ratio. The cortisol/DHEA ratio appears to be predicted by the severity of childhood trauma and current PTSD symptoms, and both variables are predicted by symptom improvement and coping. A recent meta-analysis, however, revealed mixed findings on the association between PTSD and diurnal cortisol where some of these studies suggested that cortisol levels correlated more with trauma exposure in general rather than PTSD.

Previous studies have also identified associations between PTSD and proinflammatory factors, neurotrophic factors, and neurotransmitters. Much of this evidence is not widely studied in late-life PTSD, but given that anxiety disorders are thought to be lifetime disorders, they may still be relevant in late life. Childhood maltreatment and exposure to trauma are associated with higher levels of interleukin-6 (IL-6), IL-1β, and tumor necrosis factor-α (TNF-α), and exposure to trauma is additionally associated with higher C-reactive protein (CRP). As reviewed throughout, these can promote several downstream effects, including those related to dementia promoting amyloid and tau accumulation as well as worsening vascular risk factors. In addition, midlife individuals with PTSD have lowered peripheral brain-derived neurotrophic factor (BDNF) levels, and while its relationship with brain function is unclear, it is known that BDNF gene expression in the hippocampus, amygdala, and prefrontal cortex is necessary for the formation of emotional memories, including those related to fear extinction. In midlife, individuals with PTSD may have altered peripheral serotonin and norepinephrine levels, but these findings are mixed. There are very limited studies that investigate these mechanisms in late life, but these are urgently needed to better understand PTSD in older adults, given the cohorts of aging war veterans.

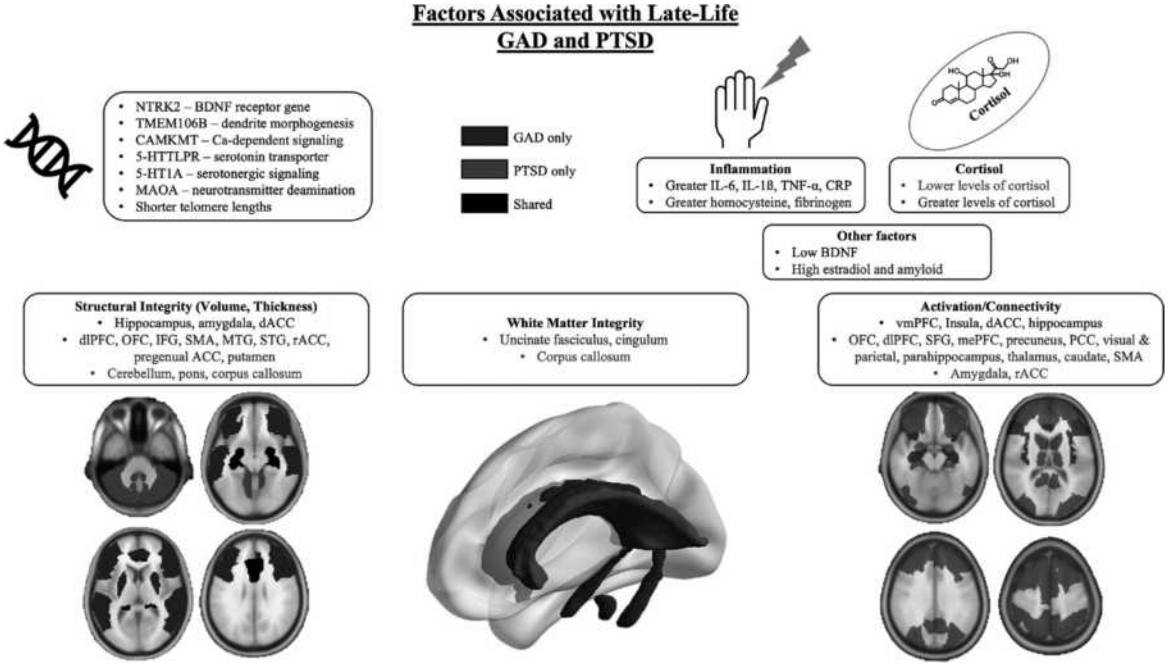
Factors Associated with Late-Life
GAD and PTSD
NTRK2 – BDNF receptor gene
TMEM106B – dendrite morphogenesis
CAMKMT – Ca-dependent signaling
5-HTTLPR – serotonin transporter
5-HT1A – serotonergic signaling
MAOA – neurotransmitter deamination
Shorter telomere lengths
GAD only
PTSD only
Shared
Inflammation
Greater IL-6, IL-1ß, TNF-α, CRP
Greater homocysteine, fibrinogen
HO
OH
Cortisol
Cortisol
Lower levels of cortisol
Greater levels of cortisol
Other factors
Low BDNF
High estradiol and amyloid
Structural Integrity (Volume, Thickness)
Hippocampus, amygdala, dACC
dlPFC, OFC, IFG, SMA, MTG, STG, rACC, pregenual ACC, putamen
Cerebellum, pons, corpus callosum
White Matter Integrity
Uncinate fasciculus, cingulum
Corpus callosum
Activation/Connectivity
vmPFC, Insula, dACC, hippocampus
OFC, dlPFC, SFG, mePFC, precuneus, PCC, visual & parietal, parahippocampus, thalamus, caudate, SMA
Amygdala, rACC

FIGURE 55.3d–1. Factors associated with late-life PTSD and GAD. This figure summarizes findings on genetics, inflammation, cortisol, BDNF, estradiol, amyloid, structural gray matter volume, diffusion-weighted imaging, and fMRI. Shared findings are in *black*, while GAD only is in *blue*, and PTSD only is in *red*. (See eBook for color image.)

In terms of brain structure, although one investigation found no differences in hippocampal volumes in Holocaust survivors relative to controls, among those with PTSD, smaller left hippocampal volumes were found in those who had no prior history of trauma before the index trauma. These findings suggest that smaller hippocampal volume may be a risk factor for the development of PTSD. In midlife, there is evidence that PTSD is associated with lower volumes in the hippocampus, amygdala, anterior cingulate, corpus callosum, pons, and cerebellum, as well as total cerebral volume. There is some limited evidence that fractional anisotropy, indicating diffusivity of water in the white matter tracts, is lower in the corpus callosum in adults with PTSD compared to those without. Very few studies have investigated the association between PTSD and white matter hyperintensities, a marker of cerebral small vessel damage, and none have found a significant association between these factors unlike in many other psychiatric disorders in late life including bipolar disorder and major depression. Finally, machine learning algorithms using normative data to estimate an individual’s brain age found accelerated brain aging in PTSD compared to those without PTSD.

There is even more limited research on brain activity (using functional MRI or other such methods) in late-life PTSD. In midlife, these can be classified into three categories: (1) symptom provocation where participants are shown stimuli related to their trauma, (2) activation tasks where participants are shown negative imagery that is unrelated to their trauma, and (3) affective priming tasks where participants are shown some affective priming stimulus then shown a neutral stimulus where a decision is made (to test priming effects of negative stimuli). In symptom provocation paradigms, participants are read scripts of the traumatic events described in their own words or presented stimuli (images, words, sounds, etc.) related to it and these studies have largely identified mixed findings—amygdala, ventromedial prefrontal cortex (vmPFC), insula, and dorsal anterior cingulate cortex (dACC) hyper- and hypoactivation, but some have also identified hippocampal hypoactivation related to these paradigms. Studies using trauma-unrelated tasks, where participants are shown fearful faces or negative imagery have been more consistent in identifying that in participants with PTSD, compared to those without, there was hypoactivation of the vmPFC, rostral ACC, and dorsal ACC as well as hyperactivation of the amygdala. Using affective priming tasks, researchers have found that individuals with PTSD have hyperactive amygdala compared to trauma-exposed participants and non–trauma-exposed participants, but other regions have not been consistently identified (hyperactive insula, rostral ACC, vmPFC, and hippocampus).

There are three prevailing models related to PTSD brain activity. The functional neurocircuitry model posits that a hyperactive amygdala contributes to heightened processing of threatening stimuli and is combined with a hypoactive PFC (vmPFC, ACC, medial PFC) that contributes to persistent fear of trauma and nontrauma stimuli, which is all further combined with a hypoactive hippocampus that prevents adaptive fear learning and extinction. The triple network model posits a hypoactive executive control and default mode network and a hyperactive salience network resulting in impoverished cognitive function combined with deficits in self-referential processing (e.g., heightened rumination) and worsened processing between those networks (salience network). The dissociative subtype model posits that individuals with PTSD and chronic dissociation have hyperactivity in the midline prefrontal, dACC, and limbic regions that underlies states of detachment and depersonalization from their emotional experiences. These models have been developed on midlife PTSD and they do not take into account the age-related effect of white matter deterioration and cortical thinning. More data is needed to integrate these models as it relates to late-life PTSD.

Generalized Anxiety Disorder

In recent years, there has been an increased interest in the neural mechanisms of late-life GAD. There are now several studies that explore cortisol, inflammation, neurotrophic factors, hormones, and brain structure and function.

In late life, GAD is associated with elevated levels of cortisol when compared with non-GAD participants. In addition, elevated levels of worry are also associated with levels of cortisol regardless of the categorical diagnosis. Response to treatment with escitalopram in individuals with late-life GAD was associated with lowered levels of cortisol that correlated with lowered levels of worry. In another late-life GAD study, response to cognitive behavioral therapy in addition to escitalopram or escitalopram monotherapy alone was associated with lowered levels of cortisol.

Both animal and human studies have shown that anxiety is associated with an increase in cell-mediated immune activity and inflammation. In a large study, participants with varying anxiety disorders in late life (including GAD, panic disorder, and social phobia) had greater levels of proinflammatory cytokines compared to those without anxiety after adjusting for other factors including depression. Other studies have found that late-life GAD is associated with greater levels of CRP, TNF-α, homocysteine, fibrinogen, and IL-6). There are several behavioral factors that may explain some of these differences including diet, lifestyle, and substance use in addition to autonomic differences like heart rate, blood pressure, vasoconstriction, and platelet aggregation. These factors may also help explain differences in risk for cardiovascular and cerebrovascular diseases. Inflammation has been shown to be involved in hippocampal neurogenesis, reduced BDNF expression (which has also been shown in late-life GAD), and potentiation of both glutamate neurotoxicity and amyloid formation.

Estradiol may be another important factor involved in the increased incidence of GAD later in life. Studies have shown that during menopausal transition, hormone levels mediated increased levels of anxiety. Panic attacks worsen during menopause and, in one study, women randomized to oral estrogens showed improved depression and anxiety over 4 years.

Studies have shown that estradiol modulates fear learning and extinction, as well as the activity in both the amygdala and vmPFC. Of note, one prospective study in postmenopausal women showed no association between hormone therapy and anxiety.

Several imaging studies have identified neuroanatomical correlates of late-life GAD using MRI. These studies have largely found that individuals with GAD, compared to those without, had lower volumes in the dorsolateral prefrontal cortex (dlPFC), orbitofrontal cortex, inferior frontal gyrus, supplemental motor area, dorsal ACC, hippocampus, middle and superior temporal gyrus, rostral and pregenual ACC, amygdala, and putamen. In addition, worry severity has been independently associated with lower limbic, medial temporal cortex, and prefrontal regions of the brain, regardless of diagnosis. In another study, GAD was associated with reduced cortical thickness in the entire prefrontal cortex. In addition, one study found that greater pretreatment left dorsal ACC volumes was associated with greater improvement of worry in participants with late-life GAD treated with citalopram. There are some conflicting findings that show greater orbitofrontal cortex, ACC, amygdala, and dorsomedial PFC volumes in late-life GAD compared to older adults without GAD. More recently, one study identified that individuals with higher worry and rumination in late life had more advanced brain age, a marker of accelerated aging. The authors put forth a model where repetitive negative thoughts (worry and rumination) may be neurotoxic to the gray matter, though longitudinal studies will be needed to confirm the causal nature of this association.

Individuals with late-life GAD had lower fractional anisotropy in the uncinate fasciculus and middle cingulum. Another study found that greater worry severity was associated with lower mean diffusivity in the prefrontal cortex. Both studies indicate some deterioration of white matter in tracts salient for emotion regulation in mood and anxiety disorders. One study, however, failed to identify any differences between older adults with and without GAD using diffusion imaging.

Unlike mood disorders, studies on white matter hyperintensities have failed to identify any direct association with late-life GAD (similar to late-life PTSD). One study did find that greater activation of the limbic cortex during emotion reactivity did indirectly tie white matter hyperintensities to late-life worry; they found that greater white matter hyperintensities were significantly correlated with lower vmPFC and higher amygdala activation during emotion reactivity, which were associated with greater worry severity. Thus, white matter hyperintensities may modify neural activity, which in turn may affect late-life symptoms of anxiety.

In fMRI studies, authors have utilized both emotion reactivity and emotional Stroop tasks. During an emotional Stroop task, one study identified lower activation in the orbitofrontal cortex, dlPFC, and superior frontal gyrus in late-life GAD participants. During emotion reactivity, one study was able to show that greater worry was associated with lower activity in the medial PFC and precuneus/posterior cingulate, while greater global anxiety was associated with greater activity in the visual cortex, parahippocampus, hippocampus, precuneus, and posterior cingulate. These studies show dysregulated limbic systems and attenuated prefrontal systems that fail to regulate worry.

Using arterial spin labeling, some studies have identified alterations in cerebral blood flow using worry-specific scripts. One study found that older nonanxious participants had greater blood flow during worry induction in the insula, amygdala, and temporal– occipital cortex, while older anxious participants did not show increased blood flow in the amygdala and insula. Using a similar paradigm in another study, they additionally found that worry induction activated the visual cortex, thalamus, caudate, and medial frontal cortex, and participants with GAD had greater blood flow in the visual and parietal cortex and middle and superior frontal gyri. They also found that the thalamus increased in activation with subsequent worry induction during the paradigm. The authors identify three major hubs of worry: emotion generation (dACC, SMA, visual cortex), a subcortical process that maintains worry (amygdala, insula, hippocampus, caudate/thalamus), and regions related to emotion regulation (supplemental motor area, vmPFC, and dACC).

Another feature is altered connectivity during resting state and worry induction. One study showed hyperconnectivity in the default mode network in participants with GAD compared to those without GAD. During reappraisal of worry, nonanxious individuals showed greater mPFC to the insula and dlPFC connectivity, while those with GAD had greater paraventricular–amygdala connectivity. After treatment with citalopram, individuals with GAD had normalized connectivity patterns similar to participants without GAD. A more recent analysis also showed an inverse U-shaped association between worry and limbic–PFC connectivity, where this connectivity is maximized with moderate levels of worry.

In conclusion, these studies indicate a complex neural network supporting the GAD psychopathology, with some specific structural changes driven by older age.

Other Anxiety Disorders

Older adult female social anxiety disorder (social phobia) patients in one sample showed greater hemodynamic responses, cardiac vagal attenuation, and sympathetic vasomotor activation than those without social anxiety disorder (social phobia) in response to a public speaking challenge. Older adults with specific phobias show greater and longer-lasting cortisol response to stress than nonanxious older adults. Older adults with depression and comorbid anxiety disorders show greater memory declines over time than those with depression alone.

Genetic Factors

Anxiety disorders have been found to have a 30% to 60% heritability in twin studies. In genome-wide studies and meta-analyses, there are several studies that have focused on genetic factors associated with anxiety disorders, but there is fairly limited research on individual diagnoses. These studies have found shared genetic risk factors with neuroticism, depression, and antidepressant treatment response. BDNF receptor gene, NTRK2, has been identified as a potential genetic marker of anxiety, which has also been implicated in autism, suicide, AD, insomnia, coronary artery disease, and alcohol dependence. NTRK2 and BDNF are thought to regulate short- and long-term potentiation of brain synapses. In one meta-analysis, the transmembrane protein 106B (TMEM106B) was associated with anxiety disorders—a single-nucleotide polymorphism (SNP) within the region coding for the calmodulin-lysine N-methyltransferase (CAMKMT) was associated with anxiety, which acts in the formation of trimethyllysine in calmodulin, the gene encodes the protein methyltransferase is involved in calcium-dependent signaling.

In a case-control design, GAD was associated with the short allele of the 5-HTTLPR, which codes for the serotonin transporter and has been widely implicated in several psychiatric disorders, including depression. Comorbid depression and GAD have been associated with the minor G allele of the functional 5-hydroxytryptamine receptor 1A (5-HT1A), conferring diminished serotonergic signaling. Case-control studies have also identified genes associated with monoamine oxidase A (MAOA, involved in the deamination of neurotransmitters including dopamine, norepinephrine, and serotonin), neuropeptide S receptor (NPSR1, which is involved in arousal and has anxiolytic effects), catechol-O-methyltransferase (COMT, which degrade neurotransmitters like dopamine, norepinephrine, and serotonin), and corticotropin-releasing hormone receptor 1 (CRHR1, which binds corticotropin-releasing hormone—involved in stress). There is also some evidence that anxiety is associated with shorter telomere lengths, indicating accelerated processes of aging.

All these studies were done on midlife anxiety; thus, it is less clear how they extrapolate to anxiety in late life.

DIFFERENTIAL DIAGNOSIS

The most common differential diagnoses are medical conditions and depression. In primary care or specialty medical settings, somatic symptoms of anxiety may be difficult to distinguish from those of cardiovascular, endocrine, gastrointestinal, metabolic, neurologic, or pulmonary disorders or from the effects of medications, including antihypertensives, β-agonists, corticosteroids, digitalis, nonsteroidal anti-inflammatory drugs (NSAIDs), and thyroid medication. A workup to identify medical conditions that may be contributing to anxiety in older adults should comprise a complete medical history and physical examination, and laboratory tests, including complete blood count, thyroid function, levels of electrolytes, vitamin B_{12}, and folate. It should be noted, however, that anxiety can coexist with medical conditions and may lead to excess disability and, potentially, exacerbation of the medical condition itself. Treatment of anxiety as a co-occurring disorder may be appropriate, particularly when the medical condition is stable.

Several symptoms of major depression and GAD overlap (e.g., irritability, decreased concentration, sleep changes, and fatigue), and rates of comorbid anxiety disorders are as high as 40% to 50% in samples of depressed older adults presenting to mental health clinics. Abuse of opioids and benzodiazepines—often comorbid with anxiety disorders—is becoming increasingly more common in older adults. Cognitive disorders may also complicate diagnosis, especially in patients with dementia, where irritability, anger, and agitation are more common.

Other differentials include bipolar disorder in hypomanic phases, especially for patients with irritable mood, insomnia, difficulties concentrating, and racing thoughts. The differential is made more difficult by the relatively high rates of comorbidity (lifetime prevalence of any anxiety disorder in bipolar disorder is 43% to 45%). Among the modalities available to differentiate clinically are a past history consistent with bipolar disorder, alternations between the hypomanic clinical presentation and depression, and the distinct quality of thought process. Thus, racing thoughts in hypomania/mania follow a tangential/flight of ideas process, while racing thoughts in anxiety are usually perseverative on specific negative themes. Rarely, patients with delusional disorders may be initially misdiagnosed as GAD, given their intense preoccupations and "worries" related to selective themes.

Specific training about anxiety disorders for geriatric mental health providers can boost recognition rates, especially in depressed older adults. Table 55.3d–2 summarizes differences in the typical presentation of depression and anxiety in older adults. Because anxiety typically develops before depression and follows a chronic rather than episodic course, questions about symptoms such as worry, feeling tense or nervous, or finding it difficult to relax at times when the individual was not depressed may clarify whether a depressive episode is superimposed on a pre-existing anxiety disorder. Neurocognitive screening tools may also help to identify underlying cognitive impairment.

COURSE AND PROGNOSIS

The course of anxiety disorders in older adults is typically chronic, with frequent relapse and uncommon remission. Prognosis is often poor; recent investigations have found higher mortality risk associated with generalized anxiety and PTSD symptoms, and panic attacks, even after controlling for medical factors and depression. This is worsened as some patients are reluctant to take medications for long intervals of time, thus making relapse common. Patients may turn to alternative treatments with low success.

With respect to specific anxiety disorders, GAD may be characterized by periods of remission and relapse occurring over decades but with longer symptomatic periods in older than in younger adults. One influential view holds that late-life GAD and depression exist on a continuum, with

GAD representing the mildest form of the disorder and comorbid GAD and depression representing the most severe syndrome. Providing more support for the hypothesis that GAD and depression exist on a continuum, one epidemiologic study found that only 12% of individuals with GAD at baseline continued to have GAD without major depression at a 3-year follow-up. Almost half (48%) had remitted, and 40% went on to develop major depression, either with or without comorbid GAD. Ten-year follow-up data from the same study indicated that excess mortality was associated with depression in men but not in women; GAD, either with or without comorbid depression, was not associated with mortality. A 40-year follow-up of individuals retrospectively diagnosed with GAD found evidence that the disorder may convert to a somatoform disorder over the lifespan. There is mounting evidence that GAD may be associated with greater cognitive decline and greater conversion to AD from mild cognitive impairment, which we discuss later in this chapter.

Table 55.3d–2.
Characteristics of Depression and Anxiety in Older Adults

Typical Presentation	**Depression**	**Anxiety**
Somatic symptoms	Vegetative (e.g., poor appetite, fatigue)	Autonomic (e.g., palpitations, dizziness, gastrointestinal distress)
Activity level	Typically psychomotor retardation	Agitation or subjective feeling of restlessness
Enjoyment of life	Diminished	Not diminished, but patient may claim that they have no time to enjoy life
Sleep disturbance	Early-morning awakening	Difficulty falling or staying asleep
Thought content	Rumination about past losses or failures, suicidal ideation	Worry about current or future threats or dangers
Behavior	Decreased activity	May exhibit excessive activity (e.g., cleaning, taking on too many projects or responsibilities)
Family relationships	May be isolated or disconnected from family	May be overinvolved or intrusive with family
Social networks	Diminished	Often not diminished, may hide symptoms from network
Onset	May have early or late onset	More typically early onset
Course	Episodic	Chronic
Attribution	Patient may believe that symptoms are part of normal aging	Patient may believe that symptoms are part of their personality or a

	or a natural response to life events	natural response to life events

A 50-year follow-up study of OCD patients found remission rates of approximately 60%. By contrast, a 50-year follow-up of panic disorder patients found only 6% of patients not meeting the criteria for any psychiatric disorder at the final assessment point. In this study, the prevalence of panic disorder with and without agoraphobia began to decline after approximately age 40 years, whereas the incidence of somatoform disorders and agoraphobia without panic rose steadily from age 20 through 60 years and then leveled off. Although PTSD symptoms are most prominent immediately after the trauma, they can persist for decades, and some evidence suggests that they may emerge for the first time or increase in later life, especially in the context of cognitive decline. Late-life PTSD is associated with an increased risk of major cardiovascular events and all-cause mortality, as well as all-cause dementia, including AD, vascular dementia, dementia with Lewy bodies, and frontotemporal lobe dementia. PTSD has also been associated with an increased risk of suicide. Late-life PTSD is often underrecognized and undertreated.

Consequences

Consequences of anxiety disorders include activity limitations, higher numbers of general practitioner and home health visits, loneliness, poor life satisfaction, poor perceived health, worse cognitive function, and perceived lack of help. Older GAD patients report poorer role functioning, social functioning, and energy levels than individuals with other chronic conditions such as heart disease and diabetes. Anxiety, especially GAD and its cardinal symptom, severe worry, has been associated with worse cognitive decline compared to those without GAD or low worry. In one 3-year longitudinal study, people with Mild Cognitive Impairment (MCI) and worry had a fivefold increased risk of AD compared to those with just MCI. In general, anxiety increases the risk of developing major depressive disorder as well, including in older adults. Older adults with PTSD, even in the absence of depression, experience more days spent in bed and days with limited activity and lower levels of life satisfaction than those without PTSD after controlling for factors including age, chronic diseases, functional limitations, social network size, and cognitive functioning. They also make more visits to mental health providers and have higher rates of benzodiazepine use. The 2013 State of Health report in the United States listed anxiety as one of the four diseases with the largest number of years lived with disability.

In nursing home patients, anxiety disorders and depression are associated with lower levels of well-being, although not necessarily with higher levels of disability. Patients with depression, either alone or combined with anxiety, report lower levels of well-being than patients with pure anxiety. Individuals with anxiety and/or depression receive less assistance with activities of daily living, make more visits to medical specialists, and take more medications, especially antidepressants. The use of antidepressants is typically more common in patients with combined anxiety and depression relative to others. Patients with anxiety make more visits to medical specialists than those without anxiety or depression.

Anxiety and Dementia

While late-life depression is well-known to play a major role in incident cognitive decline, incident dementia, and conversion from MCI to AD, there is increasing interest in anxiety as a prodromal factor associated with dementia. Over the last decade, several studies have shown that both GAD and PTSD are associated with dementia. Anxiety is the most common neuropsychiatric symptom (NPS) in dementia.

In a meta-analysis, the authors reported that anxious participants showed a decline in executive function over time but not in memory when compared to those without anxiety. The authors also verified that anxiety was associated with incident cognitive impairment and dementia. PTSD has also been associated with incident dementia (including AD, vascular dementia, frontotemporal lobe dementia, and Parkinson disease). One study showed that AD and vascular dementia increased the risk for delayed emergence, re-emergence, and worsening of PTSD symptoms.

The direction of the relationship between anxiety and dementia remains a matter of debate, given the paucity of longitudinal studies using adequate anxiety measures and following up participants for a long enough time to detect subtle cognitive changes. The “anxiety as prodrome” hypothesis posits that anxiety symptoms are triggered by the awareness of worsening cognitive functioning. These tend to decrease as neurodegenerative processes advance as a consequence of more advanced cognitive dysfunction. The “anxiety as a risk factor” hypothesis posits that chronic anxiety and worry convert the immediate psychological and physiologic responses to daily stressors into prolonged physiologic activation of the regulatory systems, which produces moderate but chronic levels of activation of the cardiovascular, HPA, and immune systems. Thus, anxiety is correlated with higher levels of stress, cortisol, BDNF, proinflammatory factors, and heart rate/blood pressure. In a large study, anxiety was associated with higher levels of proinflammatory cytokines adjusting for depression, lifestyle factors, and demographic factors. Several studies have shown that inflammation is associated with worse cognitive performance and cognitive decline, including dementia. Anxiety is also correlated with higher levels of cortisol, and there is evidence that glucocorticoid overstimulation in the medial temporal lobe is associated with hippocampal atrophy and cognitive dysfunction. There are several studies that show anxiety is independently associated with lower levels of BDNF, which is associated with lower levels of synaptic plasticity important in learning, memory, and neuronal repair. Greater anxiety has also been correlated with greater deposition of amyloid and tau, an association that may be mediated by inflammation and increased cortisol. In animal models, administration of cortisol was associated with increased depositions of amyloid and tau. In a longitudinal study, participants with anxiety and significant amyloid depositions had a more rapid cognitive decline. In another study, APOE ε4 allele carriers who had significant amyloid depositions in the brain were more likely to have more severe anxiety, especially among women. Finally, anxiety is associated with greater cardiovascular disease, cerebrovascular disease, and vascular dementia—this may be driven by increased, blood pressure, vasoconstriction, and platelet aggregation commonly associated with anxiety.

Anxiety as a Risk Factor for Other Medical Conditions

Anxiety may also serve as a risk factor for common medical conditions of later life. Studies have reported an increased risk of vascular problems, respiratory disease, chronic pain, gastrointestinal conditions, and cancer associated with PTSD.

Substantial evidence links anxiety to heart disease and stroke. GAD is independently associated with a higher risk of overall mortality, stroke, cardiovascular mortality, heart failure, and coronary heart disease independent of other risk factors, including depression. Symptoms of phobic anxiety may predict the onset of ventricular arrhythmias in patients with coronary artery disease. GAD, panic attacks, and PTSD symptoms lead to an elevated risk of coronary heart disease, even in the absence of major depression. PTSD symptoms are also associated prospectively with nonfatal myocardial infarction. The reported lifetime prevalence of anxiety disorders is considerably higher than the current prevalence of coronary heart disease, suggesting that anxiety might represent a risk factor rather than a consequence of heart disease.

TREATMENT

Anxiety disorders in older adults are frequently undetected and untreated (e.g., one study found that only 36% of patients with late-life GAD received any treatment). Older adults are significantly less likely than younger adults to receive treatment for anxiety and even less likely to seek treatment. When treatment takes place, it usually involves benzodiazepines or other anxiolytic medications rather than antidepressants or psychotherapy. Treatment response rates tend to be worse in older adults compared to younger adults. Several factors may be associated with worse response (e.g., worse cognitive function), but there are also changes in pharmacokinetics (reduced glomerular filtration and hepatic metabolization), which may also increase both side effects and effectiveness. Longitudinal studies have shown recurrence and chronicity rates up to 39% to 52% after 3 to 6 years. For late-life GAD, persistent worry is a significant component of relapse. Older adults tend to prefer psychotherapeutic approaches, though meta-analyses have shown that they are less effective than pharmacotherapies.

Pharmacotherapy

Research on the treatment of anxiety in older adults is limited. Selective serotonin reuptake inhibitors (SSRIs) and serotonin–norepinephrine reuptake inhibitors (SNRIs) are the first-line treatment for late-life anxiety disorders with similar side-effect profiles in younger adults and are relatively well tolerated. Recent evidence suggests that SSRI pharmacotherapy is superior to psychotherapy as an acute treatment for geriatric anxiety. Data from controlled trials support the use of escitalopram (Lexapro), duloxetine (Cymbalta), pregabalin (Lyrica), venlafaxine XR (Effexor), citalopram (Celexa), and sertraline (Zoloft) for late-life GAD. In a randomized trial, pregabalin was associated with a 2-point greater reduction in overall anxiety, a difference statistically significant although clinically trivial. In general, there are specific potential risks that should be monitored in older adults, including the risk of falls, gastrointestinal bleed, bone loss, hyponatremia, and monitoring of blood pressure in SNRIs (especially venlafaxine). Another study also found that Buspirone was effective and well tolerated in late-life GAD. The antipsychotic Quetiapine has demonstrated efficacy and tolerability both as a monotherapy and as adjunctive therapy to SSRIs, but one study reported multiple side effects, including drowsiness, dry mouth, dizziness, headache, and nausea. Mirtazapine is a popular choice for its effects on sleep and appetite, but the evidence of its efficacy is limited and inconsistent. Tricyclic antidepressants or monoamine oxidase inhibitors (MAOI) may be efficacious; however, they should be considered only for treatment-resistant patients due to their side-effect profiles and safety concerns.

Data are even more sparse with respect to the treatment of other anxiety disorders, but evidence from controlled trials and open-label studies suggests the efficacy of sertraline, citalopram, and escitalopram, as well as of imipramine (Tofranil) and alprazolam (Xanax), for panic disorder in older adults. The SSRIs citalopram, escitalopram, and sertraline may be particularly effective treatments for anxiety in older adults based on their low propensity for drug interactions and short half-lives. Fluvoxamine (Luvox) has demonstrated efficacy in an open-label trial, including older patients with GAD, panic disorder, or OCD.

For PTSD, only one small open-label pharmacotherapy study has been conducted in adults with a mean age greater than 60 years, and studies with mean ages of 50 years or older have veteran samples comprised almost exclusively of men. Randomized, controlled trials in this age group support the use of sertraline and mirtazapine, and an open-label trial supports the efficacy of citalopram. Prazosin may improve sleep quality and PTSD-related nightmares; its impact on other PTSD symptoms is less clear.

Benzodiazepines are often the only mental health treatment received by older adults with anxiety disorders, and they are prescribed more frequently to older adults than to younger adults. Although benzodiazepines are effective in reducing anxiety and may be useful as an adjunctive, as-needed, or short-term treatment strategy, long-term daily use is associated with falls, fractures, addiction/misuse, and cognitive decline in older adults. Among benzodiazepines, lorazepam has a relatively short half-life and no active metabolites and is, therefore, preferable to alprazolam or clonazepam, which are more likely to accumulate and achieve toxic levels in older adults. Benzodiazepine taper should be very gradual in older adults, even after a short duration of use.

Overall, in anxiety disorders, the focus of research on pharmacotherapy has shifted from serotonin and norepinephrine pathways to other neurotransmitters such as glutamate, GABA, and neuropeptides. Ketamine, riluzole, oxytocin, neuropeptide Y, psilocybin, and others have been studied for GAD or PTSD. None have been tested in older populations.

Comorbid Anxiety and Depressive Disorders. Older adults with depression and comorbid anxiety disorders or significant anxiety symptoms typically demonstrate a delayed acute treatment response and increased risk of relapse of depression. Older adults with depression and concurrent anxiety symptoms require 50% more time to respond to antidepressants and are more likely to discontinue treatment. Anxiety symptoms often persist after treatment and remission of depression, increasing the risk for depressive relapse. They may also report more suicidal ideation. In one study of older primary care patients, suicidal ideation increased by 8% in the presence of anxiety disorders (compared to no anxiety disorders) and by 18% in the presence of comorbid depression and anxiety. For these reasons, it is especially important to treat older adults with depression and coexisting anxiety to complete remission. Open-label trials support the potential utility of escitalopram for comorbid depression and GAD in older adults. Both anxiety and depression are associated independently with worse cognitive function, which is associated with treatment resistance in anxiety and depression.

Augmentation and Maintenance Treatment. No randomized trials have tested augmentation strategies with anxious older adults who fail to respond to first-line treatment. A group of older adults with anxiety disorders (mostly GAD) who did not respond to an initial trial of an antidepressant or benzodiazepine responded to adjunctive risperidone (average dose 1.2 mg per day).

Although data suggest the continued efficacy of citalopram for GAD after 24 weeks of treatment, with an additional 20% of patients responding in the second 12 weeks, no research has examined maintenance therapies for late-life anxiety. Given the chronic course of anxiety disorders and their association with relapse of depression, studies on this topic should be a high priority.

Clinical Management. Because SSRIs may initially exacerbate symptoms of anxiety, older adults should be started on a low dose (Table 55.3d–3). In clinical practice, however, older adults with anxiety or depression are frequently maintained on minimum or even subtherapeutic doses, with corresponding limited benefits. If the patient is tolerating the medication, the dose should be increased at least to the minimum target dose within 4 weeks. If the patient has not made a full response (e.g., achieved at least a 50% decrease in symptoms) within another 4 to 6 weeks, the dose should be gradually increased, with continued increases up to the maximum until a response is achieved or the patient is no longer able to tolerate the medication.

Table 55.3d–3.
Usual Starting, Minimum Target, and Maximum Dose of Commonly Used Antidepressant Older Adults With Anxiety Disorders

Medication	Starting Dose (Total mg/day)[a]	Minimum Target Dose (Total mg/day)	Maximum Dose (Total mg/day)[b]
Citalopram	10	20	20[c]
Escitalopram	5	10	20
Sertraline	12.5–25	50	200
Fluvoxamine	25	100	300
Venlafaxine XR	37.5	75	225
Mirtazapine	7.5	15	45
Duloxetine	20–40	60	120
Pregabalin[d]	75–150	300	600
Buspirone[d]	15	30	60
Quetiapine[d]	25–50	50	300

[a] Older anxious patients should be started on the lowest available dose and titrate slowly (increments q5–7 days) to avoid side effects. Use of liquid form is recommended, when available, to start at very low doses.

[b] Higher doses are sometimes used on a case-by-case basis; lower doses are often used in the oldest old.

[c] The U.S. FDA revised recommendations for citalopram: maximum recommended dose is 20 mg/day for patients older than 60 years of age.

[d] Divided BID/TID.

Older adults with anxiety are frequently more resistant to the idea of pharmacotherapy and more sensitive to side effects than younger adults with anxiety or older adults with depression, sometimes attributing to the drug symptoms that preceded the initiation of a medication trial. Therefore, the specific agent may be less important than good clinical management, including information about the safety and nonaddictive potential of the medication, reassurance about the transient and noncatastrophic nature of side effects, availability of the physician at any time in the event of bothersome symptoms, and frequent visits initially (ideally weekly for the first month and biweekly thereafter until a response is achieved). Patients may need continued reassurance at every dose change and, in the event of emergent side effects, should be encouraged to stay the course for at least 1 or 2 weeks. Clinicians may need to think creatively to facilitate this process, for example, by suggesting that patients split the dose between morning and evening administration or take over-the-counter medications temporarily to counteract side effects.

Psychotherapy

Some evidence suggests that SSRI medications may be more effective than psychotherapy as an acute treatment for late-life anxiety disorders. Research, mostly with older GAD patients, provides support for the effectiveness of cognitive behavioral therapy (CBT) relative to a waiting list and active control conditions such as supportive therapy and attention placebo. One study has demonstrated the superiority of CBT to supportive psychotherapy in a well-characterized sample of older adults with anxiety disorders (52% with panic disorder), and all participants in this investigation were taking medications but had failed to respond prior to receiving psychotherapy. Response to CBT for GAD in older adults may be compromised by memory impairment, executive dysfunction, or failure to complete at-home practice. A recent study has demonstrated that the sequence of antidepressant medication augmented with CBT in older adults with GAD leads to worry reduction in the short term. This study also found that both maintenance medication and CBT could prevent relapse compared with a placebo, providing two strategies for improving outcomes in GAD in older adults.

For older individuals who refuse to take medications or for whom they are contraindicated, relaxation training may be the most effective component of a behavioral treatment for GAD or distressing anxiety symptoms that do not meet the criteria for a specific anxiety disorder. Recommending a good relaxation tape with breathing or meditation exercises, progressive muscle relaxation, or imagery may be helpful; if, after a few weeks of practice, the patient does not report sufficient gains, it may be worthwhile to reopen a discussion about medications. One study showed that for older patients, relaxation therapy might be the most efficacious aspect of CBT. While older adults can learn and effectively use reappraisal, this may be more challenging in a population with diminishing cognitive abilities. Mindfulness-based stress reduction has also been shown to reduce anxiety symptoms in older adults, even showing preliminary efficacy in individuals with cognitive dysfunction. Finally, acceptance and commitment therapy has shown improvement in worry severity in late-life GAD.

For anxiety disorders other than GAD, some evidence supports the use of CBT for panic disorder in older adults, although results from a study that included a substantial proportion of older panic disorder patients suggest the superiority of sertraline over CBT. There are no published data beyond case reports documenting the behavioral treatment of specific phobias or OCD, all of which use exposure-based protocols, and no data on the treatment of older adults with agoraphobia without panic disorder or social anxiety disorder (social phobia). Case reports of PTSD treatment in older adults have described gains associated with psychodynamic and supportive approaches, CBT, eye movement desensitization and reprocessing life review, and exposure therapy. A pilot study of prolonged exposure therapy in older veterans with PTSD showed a significant decrease in their symptoms of PTSD.

Recommendations specific to late-life anxiety include (1) ensuring that sufficient time is devoted to learning skills such as relaxation training; (2) practicing skills multiple times in a session to ensure that the rationale and procedures are well understood prior to assigning at-home practice; (3) ensuring that anxiety levels become sufficiently high during exposure exercises, which may be facilitated by, for example, having patients hyperventilate briefly before engaging in imaginal exposure; (4) not avoiding difficult or painful issues or worries if appropriate (e.g., imaginal exposure to issues surrounding an ill spouse’s death); (5) de-emphasizing more abstract cognitive restructuring techniques in favor of simpler strategies (e.g., coping statements); and (6) between-session reminder phone calls and weekly review of concepts. There is also some evidence that incorporating religious or spiritual themes may also help, especially in older Black individuals.

Mrs. H was a 69-year-old married White woman with a college degree. She reported that she was “always a worrier” but believed that she was acting in the best interests of her family by looking out for their well-being. She indicated that her worrying had caused problems within her family; for example, her daughter was annoyed with Mrs. H for sending her classified ads and articles about employment resources to encourage her to find a new job. She further stated that her concern about her children led her to telephone them daily, and she had once contacted the police when her call was not returned by the next day.

In addition to worrying about her children, she indicated that she was concerned about her husband’s health, finances, and her health, particularly her recent diagnosis of gastroesophageal reflux disease. Although she believed that her worry was normal under the circumstances and she was resistant to the idea of taking medication, she indicated that sleep disturbance, fatigue, and difficulty concentrating were making it difficult for her to accomplish daily tasks.

Mrs. H was started on 50 mg per day of sertraline but called the psychiatrist’s office after 3 days to report that she had discontinued the drug after experiencing an upset stomach and a burning sensation in her throat. She agreed with the doctor’s recommendation for her to restart the medication at a reduced dose of 25 mg per day and to take chewable calcium tablets as often as needed. He called her back 2 days later to inquire about the success of this strategy. Although Mrs. H reported that the symptoms continued, she agreed to continue the medication for another week until her next scheduled appointment. At that point, she was unwilling to consider raising the dose but agreed to stay the course for another 2 weeks, at which point she reported that the side effects were “unpleasant but manageable” and reluctantly agreed to try 50 mg per day. Over the next 3 months, her dose was gradually increased to 200 mg per day, with diminishing side effects. Mrs. H reported that although she continued to worry about her family, the thoughts were less distressing, and she was sleeping better and had more energy.

Mrs. P was an 82-year-old Black woman whose daughter brought her to the clinic out of concern that she was depressed. Although Mrs. P lived independently and had been sleeping and eating well, she had refused to leave her home unaccompanied since experiencing a noninjurious fall 3 months before the appointment. She denied feeling sad and reported that she enjoyed watching television, listening to music on the radio, and having family and friends from her church visit her in her home. She stated that although she used to take walks for exercise and do her grocery shopping at the corner market, she now experienced considerable anxiety about not having help available should she fall outside of her home. Her family was trying to decide whether she should come to live with one of her children. Because of her strong preference to continue living on her own, she agreed to come to the clinic for an evaluation.

After consulting with her physician, who reported that she did not have orthostasis or other medical conditions increasing her risk of falling, the psychiatrist decided first to have Mrs. P evaluated by a physical therapist, who tested her gait and taught her to use a cane. Although Mrs. P reported that she was using the cane to feel more secure within her home, she continued to avoid unaccompanied trips outside of her home. She was then referred to a psychologist for behavioral treatment.

Mrs. P and the psychologist collaboratively developed a hierarchy of situations involving exposure to trips outside of the home that varied in duration, distance, and availability of family members for support. Her family was enlisted to help with the treatment by initially being available by telephone while she was engaged in an exposure. As treatment progressed, she was “weaned” from making calls to family members immediately before, during, and after her walks. At this point, which took six sessions over an 8-week period, Mrs. P reported that she was able to leave her home alone, taking her cell phone but not using it routinely during every excursion. Treatment was deemed successful and terminated.

Transcranial Magnetic Stimulation

Transcranial magnetic stimulation (TMS) is a safe, effective, noninvasive, and nonconvulsive treatment alternative that has been approved for the treatment of treatment-resistant depression as well as OCD. Currently, TMS is only FDA-approved for treatment-resistant depression and OCD. For OCD, only high-frequency deep repetitive transcranial magnetic stimulation (dTMS) over the dorsomedial prefrontal and anterior cingulate cortices (dmPFC-ACC) with an H7 coil using the BrainsWay system has been approved. Several studies have shown that this approach can help reduce obsessional ideation. Research aiming to develop other targets, including in the dlPFC, supplemental motor area, and orbitofrontal cortex, is under way.

Studies have shown that in patients with comorbid depression and generalized anxiety, TMS can help improve symptoms of depression and anxiety. However, TMS is not approved for GAD. In a meta-analysis of four studies, it was found that there is consistent evidence that low-frequency stimulation over the right dlPFC reduces anxiety symptoms. There is mounting evidence for the use of TMS in PTSD. In a meta-analysis of nine studies, it was found that high-frequency TMS of the right dlPFC resulted in improvement of PTSD and, in some studies, comorbid depression. There is more limited evidence for the efficacy of TMS in panic disorder and specific phobia. In one randomized, double-blind study, low-frequency TMS over the right dlPFC in patients with panic disorder and depression was associated with a greater response to panic disorder, depression, and social adjustment compared to sham TMS. There is even more limited evidence for a specific phobia, with most studies only looking at changes in biomarkers after a single session of TMS, which is not sufficient to look at treatment efficacy.

In general, TMS is well tolerated and is a promising new approach showing efficacy for GAD, PTSD, and OCD and it may be a well-tolerated option for older anxious individuals. Currently, there are no studies regarding TMS's efficacy and tolerability in late-life anxiety.

FUTURE DIRECTIONS

Anxiety disorders are more common than usually recognized in older adults. In the mental health clinic setting, they are typically seen as a co-occurring diagnosis with depression, although this has changed as public awareness about anxiety and its treatment has grown. Coexisting anxiety and depression typically represent a more severe disorder than depression alone. Because onset is often early in life, older adults with anxiety disorders have usually experienced a chronic course. Distress and impairment associated with these disorders can be substantial. Data on biologic factors involved in late-life anxiety implicate HPA axis dysregulation and genetic, cognitive, neuroanatomical influences are under investigation. There is also some preliminary evidence regarding altered brain activity and connectivity. SSRI medications appear to represent the first-line acute treatment for geriatric anxiety. Very few maintenance studies have been performed, and more research is needed. Psychotherapy has shown some efficacy in older adults for GAD and panic disorder with promising pilot studies in PTSD and has not been adequately investigated for phobias, including agoraphobia. A focus on anxiety disorders across the life span is needed; one practical step toward this is including older participants in clinical trials within multiple settings, including community and residential facilities.

FURTHER READINGS

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.

Andreescu C, Lee S. Anxiety disorders in the elderly. *Adv Exp Med Biol*. 2020;1191:561–576.

Andreescu C, Varon D. New research on anxiety disorders in the elderly and an update on evidence-based treatments. *Curr Psychiatry Rep*. 2015;17(7):53.

Bower ES, Wetherell JL, Mon T, Lenze EJ. Treating anxiety disorders in older adults: current treatments and future directions. *Harv Rev Psychiatry*. 2015;23(5):329–342.

Cirillo P, Gold AK, Nardi AE, et al. Transcranial magnetic stimulation in anxiety and trauma-related disorders: a systematic review and meta-analysis. *Brain Behav*. 2019;9(6):e01284.

Gulpers B, Ramakers I, Hamel R, Köhler S, Voshaar RO, Verhey F. Anxiety as a predictor for cognitive decline and dementia: a systematic review and meta-analysis. *Am J Geriatr Psychiatry*. 2016;24(10):823–842.

Lenze EJ, Wetherell JL. Bringing the bedside to the bench, and then to the community: a prospectus for intervention research in late-life anxiety disorders. *Int J Geriatr Psychiatry*. 2009;24:1–14.

Le Roux H, Gatz M, Wetherell JL. Age at onset of generalized anxiety disorder in older adults. *Am J Geriatr Psychiatry*. 2005;13(1):23–30.

Ly M, Andreescu C. Advances and barriers for clinical neuroimaging in late-life mood and anxiety disorders. *Curr Psychiatry Rep*. 2018;20(1):7.

Miller K, Driscoll D, Smith LM, Ramaswamy S. The role of inflammation in late-life post-traumatic stress disorder. *Mil Med*. 2017;182(11–12):e1815–e1818.

Purves KL, Coleman JR, Meier SM, et al. A major role for common genetic variation in anxiety disorders. *Mol Psychiatry*. 2020;25(12):3292–3303.

Robb CE, De Jager CA, Ahmadi-Abhari S, et al. Associations of social isolation with anxiety and depression during the early COVID-19 pandemic: a survey of older adults in London, UK. *Front Psychiatry*. 2020;11:591120.

Wetherell JL, Petkus AJ, White KS, et al. Antidepressant medication augmented with cognitive-behavioral therapy for generalized anxiety disorder in older adults. *Am J Psychiatry*. 2013;170(7):782–789.

55.3e Geriatric Mood Disorders

TALYA SHAHAL, M.D., JENNIE DAVIDOW, M.D., JENNIFER R. GATCHEL, M.D., PH.D., AND BRENT P. FORESTER, M.D., M.SC.

Geriatric mood disorders cause suffering, increase the medical burden, worsen the outcomes of many medical conditions, and contribute to disability. Depression increases the risk of obesity, frailty, diabetes, cognitive impairment, and mortality. Investigators at the World Health Organization concluded that depression produces a greater decrement in health than angina, arthritis, asthma, and diabetes. This finding holds across countries and demographic characteristics. The complexity of geriatric mood disorders and comorbid medical illnesses requires coordination of care among health professionals and continuity of care. Moreover, the behavioral disability of persistent geriatric mood disorders often requires expensive support systems. Outpatient costs of depressed older adults were found to be about 50% higher compared to nondepressed older adults, even after adjustment for medical comorbidities.

Despite the availability of relatively efficacious treatments, geriatric depression continues to be underdiagnosed. The failure to recognize geriatric depression stems from the clinical complexity of the syndrome, symptom overlap with medical comorbidities, social bias, and barriers in the settings in which most depressed older adults are treated. Depression in older individuals does not have pathognomonic symptoms or signs. Older adults are more likely to describe somatic symptoms as their primary concern, such as those related to sleep, fatigue, decreased appetite and body weight, gastrointestinal distress, pain, hypochondriasis, and exhibit psychomotor slowing or agitation, as well as cognitive symptoms such as deficits in memory, concentration, processing speed, and executive functioning. Many of the symptoms of depression (e.g., loss of interest, lack of energy, loss of appetite, and sleep problems) can be caused by medical illnesses. Moreover, depressed patients often have several chronic medical illnesses, which makes it difficult to determine whether a symptom is caused by depression, a medical illness, or both. Depressive disorders and medical illnesses frequently co-occur, share symptoms, and often have a bidirectional risk relationship–a depressive disorder is associated with poor health, and many medical illnesses are associated with an increased risk for a depressive disorder. In addition, depressed older adults often experience anhedonia (depression without sadness) or irritability rather than a depressed mood, which can be difficult to ascertain during a brief examination.

Although the prevalence of major depressive disorder (MDD) among older adults is lower compared to younger adults, older adults have higher incidence rates of other depressive disorders (i.e., minor, subsyndromal, or subthreshold depression, dysthymia, or adjustment disorder with depressed mood). Subthreshold depression syndromes are associated with a greater odds of lifetime psychiatric disorders and of developing a MDD or an anxiety disorder in subsequent years. In addition, subthreshold depression syndromes also cause significant impairment and decreased quality of life, such as increased disability, morbidity, and mortality. This underscores the importance of detecting and aggressively treating subclinical depressive syndromes in older adults. Depression in older adults is associated with a worse trajectory compared to younger adults, likely moderated by depression severity, number of previous episodes, and medical comorbidity. Additional risk factors for a poor depression outcome are older age, severe depression at baseline, comorbid dysthymia, and younger age of onset. Older adults who experience social isolation are at greater risk for increased morbidity, diminished immune function, depression, and cognitive decline. Loneliness specifically has been linked closely to depressed mood, decreased well-being, and higher mortality.

Ageism among patients and health professionals may lead to diagnostic complacency based on the view that older individuals who lose their physical prowess and social status have valid reasons to be depressed. Two-thirds of depressed older adults are treated in busy primary care offices. Such settings are geared to acute care of medical problems creating barriers to effective diagnosis and referral to mental health treatment. In addition, many older adults are reluctant to receive mental health treatment, which may be related to high levels of perceived public stigma toward mental illness as well as negative attitudes toward mental health treatment. Furthermore, racial and ethnic minority older adults often experience a greater burden of unmet mental health needs related to patient, provider, and health care system barriers. Shortages in the geriatric and geriatric psychiatry physician workforce exacerbate access to mental health treatment.

Incorrectly attributing physical symptoms to depression may lead to premature discontinuation of standard medical assessment and failure to identify a medical cause for physical complaints. Although late-life depression is underdiagnosed, many nondepressed older patients receive antidepressants. Inappropriate use of antidepressant medication may increase the cost of care and lead to adverse side effects or drug interactions. To address these issues, the collaborative care model seeks to improve the quality of assessment and treatment of depression in primary care settings. In this model, the behavioral health team, consisting of a behavioral health support specialist and supervising psychiatrist, deploys standardized assessment and measurement-based care within a primary care setting. Nearly two decades of research have demonstrated that the implementation of a collaborative care model in primary care practices across various healthcare systems leads to substantial clinical improvement of depression and anxiety in older adults and an overall reduction in healthcare costs.

The past decade has witnessed an acceleration of the understanding of the neurobiology of late-life mood disorders. Future studies designed using the NIMH Research Domain Criteria (RDoC) guidelines will hopefully identify critical neurobiologic targets paving the way for novel pharmacotherapy, and psychotherapy treatment approaches for late-life mood disorders.

EPIDEMIOLOGY

The National Comorbidity Survey Replication discovered that the overall prevalence of major depression among persons aged 65 years or older was estimated to be 1.5% in women and 0.2% in men, with an overall prevalence of 1%, approximately one-fourth of that in younger adults. Recent data identified MDD in 1% to 5% of community-dwelling older adults and 6.5% to 9% of the patients treated within primary care clinics.

However, when combining symptoms of major or minor depression with reported treatment for depression, the cumulative depression prevalence in later life is noted to be as high as 11.19%, with similar rates for men (10.19%) and women (11.44%). Further, in this population study of late-life depression, Caucasians and Hispanics had nearly three times the prevalence of depression found in African Americans.

Between 15% and 36% of community-dwelling older adults endorse subthreshold, though clinically significant, depressive symptoms. Subthreshold depression is lower in community settings versus primary care settings and highest in long-term care settings. Approximately 8% to 10% of older adults with subthreshold depression develop major depression per year.

The prevalence of major depression has been increasing in recent birth cohorts compared to cohorts born in the early part of the 20th century. Depression appears to have an increasingly earlier onset in cohorts born in recent years, and there is a narrowing in the differential risk for depression between genders.

Depression frequently occurs in medical populations. For patients treated in primary care settings, depression was identified in 17% to 37% of patients. Approximately 30% of these patients had major depression, and the remainder had a variety of depressive syndromes that could benefit from medical attention. Major depression occurs in roughly 25% of older home care patients, 11% of medically hospitalized patients, and up to 48% of individuals in long-term care settings.

As the world population ages, prevalence rates for older adults with bipolar disorder (BD) are also rising, now estimated to account for 25% of all bipolar patients. While the estimated lifetime prevalence of BD in the U.S. population for individuals aged 60 years and older is 1% to 2%, rates in specific clinical populations vary from 3% of nursing home residents up to 7% of psychiatric outpatients, 10% of psychiatric inpatients, and 17% of individuals presenting for psychiatric treatment in ER settings. Further, approximately 70% to 95% represent cases with onset prior to age 50 years that have persisted into later life. Approximately 69% of late-life bipolar patients are women.

In summary, mood disorders occur in the older geriatric population at a rate lower than that of younger adults. However, significantly higher rates of mood disorders occur in medically ill, disabled, or institutionalized older adults. With the aging of birth cohorts who appear to be at particular risk for depression, an unprecedented number of older adults are expected to require psychiatric attention.

SUBTYPES OF LATE-LIFE DEPRESSION

Depression in Medically Ill Patients

Geriatric depression typically occurs in the context of medical disorders. Sometimes these conditions predispose or trigger depression. For example, subclinical hypothyroidism increases the risk for depression more than fourfold, while nearly a third of patients develop depression within a year after myocardial infarction. The prevalence of depression is also higher among patients with hypertension. Overall medical severity and disability appear to be important risk factors for depression. For these reasons, geriatric depression is viewed as a heterogeneous entity.

Older adults with chronic disease have a 50% greater risk for depression than older, medically healthy adults. The prevalence rate of depression among patients with chronic disease is up to 50%. Depression frequently occurs due to comorbid medical conditions such as thyroid disease, diabetes, cardiac disease, and other chronic medical conditions and can predict cardiovascular disease and all-cause mortality independently of a wide range of potential confounders. Depression increases the risk of coronary disease by 60% to 80% while also increasing the risk of cancer and a shortened survival time, regardless of cancer etiology. Individuals with malnutrition, metabolic syndrome, and posthip fracture surgery are all at elevated risk of depression. Depression not only is a contributor to hypertension but also may affect the outcomes and prognosis of hypertension and the efficacy of drugs. Data from a large health maintenance organization population suggest that the health care cost of depressed medical patients is twice that of nondepressed patients with similar levels of medical morbidity, though only a small proportion of the increased costs are spent on specialty mental health care.

Depression impairs social function more than heart and lung disease, arthritis, hypertension, and diabetes and leads to greater impairment in physical function than arthritis, hypertension, and diabetes. This effect seems to be mediated through decreasing adherence to treatment, as well as through biologic mechanisms such as changes in sympathetic–parasympathetic balance, changes in serotonin-mediated thrombogenesis, and changes in the immune system.

Depression is underdiagnosed and undertreated in primary care settings. Approximately 11% of depressed high utilizers of primary care services receive adequate antidepressant treatment, whereas 34% receive inadequate treatment and 55% receive no treatment. The pattern of inadequate treatment of depression also extends to long-term care settings, where 23% of nursing home residents diagnosed with depression receive no treatment.

Depression in Patients With Cancer

Among patients with cancer, depressive symptoms are associated with decreased quality of life, significant deterioration in physical activities, relationship difficulties, and greater pain. For older adults, the combined impact of aging and cancer can be devastating, and older adults with cancer have some of the highest rates of completed suicide. Depression among patients with cancer has a negative impact on treatment engagement and adherence. Nevertheless, older adults are far less likely to be accurately diagnosed or treated for depression compared to other age groups, resulting in decreased quality of life, poor treatment adherence, increased length of hospital stay and health service utilization, and in severe cases, suicide. Despite its importance, the area of mental health for older adults with cancer is underdeveloped. Almost half of all cancer patients who have serious depressive symptoms are not treated for these symptoms. This is partially explained by the overlap of depressive symptoms and cancer, cancer treatment-related symptoms, time and resource limitations, and inappropriate normalization of depressive symptoms among people with cancer. A recent study found that a history of depression treatment prior to a cancer diagnosis is significantly associated with the presence of elevated depressive symptoms among older adults with cancer, highlighting the importance of a detailed psychiatric history when evaluating the risk for current depressive syndromes. As there are no empirically validated guidelines that combine best practices for depression management in both oncology and geriatrics, clinicians should work closely with older patients and care partners in order to identify and treat impairing symptoms and to make referrals to psychosocial services when necessary.

Depression in Patients With Dementia

In DSM-5, cognitive disorders are described on a spectrum: mild and major neurocognitive disorders (MaNCD) and their etiologic subtypes. Despite a recent change in nomenclature, the term dementia continues to be used frequently in clinical settings and is retained here for consistency. In patients with mild neurocognitive disorder, depression prevalence is above 30%. The depression prevalence among patients with Alzheimer disease (AD) varies from 5% to 48% based on patient sampling, dementia severity, and diagnostic criteria used. Examination of the interval between the occurrence of depression and the development of dementia suggests that depression may represent an early symptom, risk factor, or prodrome of dementia. A meta-analysis concluded that a history of depression doubles the risk of developing dementia in late life. Depression occurs in approximately 25% of patients with cerebrovascular disease and in up to 50% of those with Parkinson disease. The similarity of depressive manifestations with symptoms and signs of neurocognitive disorders, for example, loss of interest, decreased energy, agitation or psychomotor retardation, poor concentration, and apathy, often poses diagnostic problems. Sad mood and psychic rather than vegetative features have been found useful in distinguishing depressed patients with dementia from patients with dementia alone.

Several studies suggest that AD pathology, amyloid beta (Aβ) or tau, may be associated with late-life depression. Among cognitively normal community-dwelling older individuals without major depression, Aβ burden (measured by positron emission tomography [PET] imaging) was associated with worsening anxious-depressive symptoms over a 1- to 5-year follow-up. Regional tau burden (measured by PET imaging) was also cross-sectionally associated with subclinical depressive symptoms independent of brain amyloid in this community-dwelling sample. Patients with amnestic mild neurocognitive disorder and a history of major depression had higher Aβ deposition, mainly in the frontal cortex, compared to patients with amnestic mild neurocognitive disorder without a history of major depression. Individuals with AD who have a history of depression had a greater amyloid plaque burden in the hippocampus than those without depression. Moreover, one study found evidence that citalopram can decrease Aβ plaques in the brains of mouse models of AD mice and stop the growth or slow the production of pre-existing plaques in AD mouse models and in humans. A recent pilot study found that amyloid status predicted poor antidepressant response among cognitively intact older adults, suggesting that alternative treatment approaches may be needed for amyloid-positive depressed older adults. However, further investigation, including prospective controlled trials, is needed to understand the relationship between antidepressant medication and AD pathology.

Finally, depression increases the risk of conversion of mild neurocognitive disorder to AD, and long-term treatment with antidepressants delays the conversion of mild neurocognitive disorder to Alzheimer dementia.

Despite these findings supporting depression as a risk factor or prodrome of dementia, there have been contrasting findings among studies. Indeed, a recent study comparing cortical amyloid deposition (measured by PET imaging) in older adults with major depression versus nondepressed cognitively normal older adults found that those with major depression had less Aβ deposition than nondepressed individuals. Aβ deposition was not associated with depression history characteristics in this sample, but it was associated with worse memory performance. These variable findings may, in part, reflect the heterogeneity of late-life depression, encompassing both early-onset MDD and late-life onset depressive symptoms that occur as part of the prodrome of dementia syndrome. Future work disentangling differential neurobiology underlying these presentations is critical to clarify prognosis and treatment and ultimately to reduce dementia risk.

Depression With Reversible Dementia

Some depressed older adults develop a dementia syndrome that improves or largely subsides after the remission of depression. This syndrome has been termed *pseudodementia*, *dementia syndrome of depression*, or *depression with reversible dementia*, and it has attracted considerable clinical interest. Abnormal executive function, processing speed, memory consolidation, and working memory may persist after the remission of mood symptoms in many patients with geriatric depression. Older adults with both depression and diabetes were found to be particularly vulnerable to cognitive deficits.

Depressed older adults who remain with some cognitive impairment even after improvement of depression usually have early-stage dementia whose cognitive manifestations are exaggerated when the depressive syndrome is superimposed. Although some disagreement exists, it appears that even those individuals with nearly complete cognitive recovery develop high rates of irreversible dementia (about 20% per year) on follow-up. Many of these patients may remain nondemented for 1 to 2 years before the development of irreversible dementia. About 40% of older adults with depression and mild cognitive impairment progressed to AD within 27 months. A presence of active depression within the last 2 years compared to those with a more remote history of depression was found to be a marker of even greater risk, possibly reflecting a more aggressive progression of underlying Alzheimer pathology. The heterogeneity of the course of depression with reversible dementia suggests that patients suffering from this syndrome can be ordered along a continuum. At the one end of the continuum lie cases in which the cognitive disturbance results predominantly from the depressive syndrome itself. On the other end are cases in which the neurocognitive impairment originates from a progressive subclinical neurodegenerative disorder with some contribution by the depressive syndrome. Therefore, the identification of cognitive dysfunction in older adults with depression constitutes an indication for a thorough diagnostic workup and frequent follow-up aimed at the identification of treatable neurocognitive disorders.

Psychotic Depression

Psychotic depression occurs in 20% to 45% of hospitalized depressed older adults and in 3.6% of depressed older adults living in the community. Patients with psychotic depression more commonly present with delusions, whereas hallucinations are less frequent. The usual themes of depressive delusions are guilt, hypochondriasis, nihilism, persecution, and sometimes jealousy. Depressive delusions can be distinguished from the delusions of patients with dementia in that the latter are less systematized and less congruent to affective disturbance. Patients with psychotic depression have more severe depressive symptomatology and a poorer prognosis compared to patients with nonpsychotic depression. In addition, symptoms of malnutrition, dehydration, and cognitive deficits are often present.

Late-Life Bipolar Disorder

Although community prevalence rates of BD are lower in older adult populations than in younger adult ones, recurrent episodes of mania and hypomania persist into later life and affect 5% to 10% of psychiatric patients. Late-onset mania often occurs in patients who have had recurrent depression episodes since early life. Late-life BD occurs in the context of cognitive impairment, high medical burden, disability, and limited psychosocial support. These factors influence the course of late-life BD. Importantly, decrements in physical health-related quality of life in this population are not an inevitable consequence of aging, and adequate attention to medical comorbidity at any age may have a tangible impact.

Cognitive impairment occurs in 40% to 60% of symptomatic bipolar older adults and may persist after improvement of mood symptoms. Deficits in attention, executive functions, and memory have been noted. Older adults with BD manic episodes may have more severe executive function deficits than older adults with unipolar depression. These impairments may reflect, in part, processes involved in the pathophysiology of mood disorders, including dysfunction of corticostriatal circuits and other limbic pathways. Cognitive impairment may be a prodrome of dementia and may also be associated with poor outcomes in late-life mania. Evidence regarding the connection between late-life BD and the progression of cognitive dysfunction has been inconsistent. While epidemiologic and cross-sectional studies have shown that BD increases the risk for dementia later in life, a recent review and meta-analysis examining longitudinal cognitive performance among patients with late-life BD did not identify any evidence of accelerated deterioration in cognitive functioning.

Older adults with BD have higher stroke risk factors than both aged controls and unipolar older adults. Late-onset mania is especially associated with vascular burden and high mortality due to cerebrovascular disease. Neuroimaging studies have revealed a high frequency of signal hyperintensities and silent strokes in older bipolar patients. Structural abnormalities are most prominent in BD with late age at the onset of the first episode. Cardiac pathology, renal impairment, and thyroid dysfunction are also common in aged bipolar patients. While evidence supporting an association between cerebrovascular disease and late-onset BD is reasonably strong, the mechanisms underlying this relationship are less clear. The presence of cardiovascular risk factors, which increase stroke risk, is also more common among patients with late-life BD. However, it remains unclear whether ischemic brain disease is a cause and/or consequence of late-life BD. Regardless, these findings highlight the importance of cardiovascular care as part of routine clinical management in older adult BD.

BD leads to behavioral disability that worsens the symptom course in BD. Disability in older adults can lead to institutionalization and influence morbidity and mortality. Geriatric bipolar patients report low perceived social support compared to same-age controls. Bipolar patients with a history of frequent or lengthy episodes of illness lack social support. The effect of social support on treatment outcomes of BD in older adults requires further investigation.

Depression and Disability

According to the World Health Organization, unipolar depression alone is responsible for more than 1 in 10 years lived with disability worldwide. The morbidity and mortality associated with major depression render it the number one cause of disability worldwide and exerts an extraordinary economic burden on society in terms of lost productivity. Depression is among the 10 diseases that increase disability-adjusted life years (DALYs) the most. DALY is a measure of years of life lost to premature death and years lived with a disability. In a 12-year follow-up study, individuals with major depression at baseline were four times more likely to report incident disability after controlling for medical illness and sociodemographic factors. Whereas depression increases the risk of disability, the opposite is also true. Increases in disability over time predict the emergence of depressive symptoms. Impairment in instrumental activities of daily living is related to depressive symptomatology and particularly to anxiety, helplessness, hopelessness, and worthlessness. Disability is associated with executive dysfunction and with late-onset depression.

Bereavement

Depressive symptoms or syndromes often follow a significant loss. Older adults may be at heightened risk for negative bereavement outcomes, including depression due to changes in social, financial, and practical support, isolation and loneliness, needing to simultaneously cope with pre-existing chronic illness, reduced mobility, and cognitive decline. The prevalence of bereavement-related psychological disorders varies, with approximately 9% to 25% of bereaved older adults developing prolonged grief disorder, 15% experiencing depressive symptoms, and 9% to 16% developing bereavement-related PTSD. Approximately 7% of bereaved older adults will develop continued distress and impairment lasting for years after the death, termed “persistent complex bereavement disorder” in DSM-5.

SUICIDE RISK

Older adults are at significant risk of death by suicide relative to younger adults. White men older than 85 years have the highest rates of suicide deaths, almost four times the rate in the general population (i.e., 47.17 per 100,000 vs. 14.21 per 100,000). Older adults have high suicide rates worldwide, with more than 7,000 yearly deaths by suicide of people aged 60 and older in the United States alone. Suicide ranks ninth among causes of death in the United States, ahead of sepsis, liver disease, and hypertension. Among older Americans who reported suicidal ideation in the past year, 12.7% reported at least one suicide attempt in the past year. Among older adults who died by suicide, lifetime suicide attempts prevalence was higher among people ages 65 to 74 years compared to ages 75 to 84 years. In this population that was studied, the prevalence of psychiatric disorders, legal and financial stressors, and relationship problems decreased with age. In contrast, physical conditions and bereavement increased with age. The severity of depression, psychotic depression, alcoholism, recent bereavement, abuse of sedatives, physical illnesses, and recent development of disability is the most important risk factors in depressed older adults. Accumulation of chronic medical conditions, social isolation, psychiatric and neurocognitive disorders (which might contribute to cognitive disinhibition and altered decision making), as well as physical and psychological pain, may also contribute to the increased risk in this population. Aging reduces suicide attempts but increases their lethality. Elder frailty might also contribute to suicide lethality. It is estimated that within the United States, there are 25 attempts for every suicide death. Among older adults, there are only four attempts for every suicide death, making nonlethal suicide attempts less common. Therefore, it is critically important to identify suicidal ideation and risk early in this age group before it escalates into an attempt. Suicide victims are demographically different from individuals who attempt suicide. About 60% of those who commit suicide are men, while 75% of those who attempt suicide are women. More than 70% of suicide victims in the United States use firearms compared to only 10% in Italy. Additional leading suicide methods in the United States include drug overdose and self-inflicted lacerations. In late life, suicide attempts are often more carefully planned and are preceded by fewer warnings of suicidal intent.

Indeed, unlike the high frequency of suicide attempts in late life, reports of suicidal ideation decrease in older adults. Despite its lower frequency, suicidal ideation is a risk factor for suicide in older adults. The severity of depression is the strongest predictor of the course of suicidal ideation. In older adults with major depression, suicidal ideation is most common in those with previous serious suicide attempts, high severity of depression, and poor social support.

Depression is the most common psychiatric diagnosis in older adult suicide victims, unlike younger adults in whom substance abuse alone or with comorbid mood disorders and psychosis are the most frequent diagnoses. Major depression is identified in 80% of suicide victims older than 74 years of age. Many older adult suicide victims are widowed, and fewer are single, separated, or divorced, compared with younger adults. Depressive symptoms, medical status, disability (pain and functional impairment), perceived health status, cognitive difficulties, social disconnectedness, and perceived burdensomeness and hopelessness, as well as access to deadly means, were identified as risk factors for suicidal ideation and death in older adults. Insomnia, one of the most common symptoms of depression at any age and particularly in late life, is an independent risk factor for suicidal ideation in adults and suicidal death in older adults. It is unclear whether the treatment of insomnia can mitigate this risk. Studies suggest that although older adults may exhibit less avoidance of death-related topics and greater overall acceptance of death, it is problematic when older adults express a desire for their life to end, even without active thoughts of ending one’s own life. They should not be viewed as normative, as even passive suicidal ideation is associated with an increased risk for suicidal behavior.

Suicide among patients with dementia is 8 times more common than in the general older adult population. Suicide occurs mostly in the early stages of dementia and is related to disinhibition, impaired problem-solving, and poor decision-making associated with cognitive decline.

Suicide is also a leading cause of death in patients with bipolar illness. By one estimate, 25% to 50% of all patients with BD will attempt suicide at least once in their life, and 8% to 19% will die of suicide. Risk factors for suicide in BD include younger age of illness onset, history of past suicide behavior, family history of suicide acts, comorbid borderline personality disorder, substance use disorders, and hopelessness.

The clinical characteristics of depressed older adult suicide victims suggest that identification of depression and appropriate care can reduce the incidence of suicide. The majority of older adult suicide victims see their physicians within a few months of their death, and over a third, within a week of their suicide. Among depressed older adults seen in primary care in the 12 months prior to a suicide attempt, less than 5% received appropriate depression treatment prior to the attempt. Therefore, a reliable assessment of suicide risk is critical since protective measures may avert suicide. There is no evidence that questions about suicide can increase the likelihood of suicidal behavior. Brief interventions such as safety planning in addition to aggressive treatment of depression may prevent suicide, as suicidal ideation is closely associated with depressive symptomatology in older adults. Worsening depression and suicidal ideation may indicate the need for psychiatric hospitalization and/or ECT.

BIOLOGIC DYSFUNCTION

The Late-Onset Hypothesis

An early hypothesis of late-life depression postulated that major depression with the onset of the first episode in late life includes a large subgroup of patients in whom neurologic brain abnormalities contribute to the development of a depressive syndrome. Clinical findings supported the “late-onset hypothesis.” Relative to older adults with early-onset major depression, individuals with late-onset major depression have a less frequent family history of mood disorders, a higher prevalence of dementing disorders, and a higher rate of dementia development on follow-up, although some disagreement exists. Most older adults with major depression and initially reversible dementia have late-onset depression. Despite the improvement of cognitive functions with the remission of a depressive syndrome, many of these patients develop irreversible dementia during long-term follow-up. The “late-onset depression” hypothesis pointed out the importance of aging and disease-related brain changes in late-life depression. Additional research has highlighted mild behavioral impairment, a constellation of affective and behavioral symptoms with onset later in life, as a prodrome for dementia syndromes.

The Depression-Executive Dysfunction Hypothesis

This hypothesis postulates that a “depression-executive dysfunction (DED)” syndrome exists with a clinical presentation resembling medial-frontal lobe syndrome. Clinical studies documented that the DED syndrome is common and characterized by psychomotor retardation, reduced interest in activities, apathy, impaired insight, suspiciousness, and severe behavioral disability, but less pronounced depressive ideation and vegetative signs.

Recent clinical studies found that DED is common, with approximately 30% of depressed older adults who have abnormal performance in tests of verbal fluency, response inhibition, novel problem solving, cognitive flexibility, working memory, and/or ideomotor planning. Patients with depression and executive dysfunction have a poor, slow, and unstable response to antidepressants as well as early relapse and recurrence. Similarly, depressed older adults who reported executive functioning complaints at baseline had a slower response to escitalopram treatment than those without executive functioning complaints.

A recent systematic review identified a triad of depressive symptoms, executive dysfunction, and impaired gait speed in up to 17% of older adults with depression. This may represent a distinct entity with shared etiology and has significant implications on fall risk, daily activities, quality of life, and overall morbidity and mortality. This triad is associated with vascular risk factors and white matter hyperintensities (WMH), suggesting that frontal-subcortical dysfunction might be the underlying cause.

The DED hypothesis assumes that executive dysfunction is the clinical expression of frontolimbic dysfunctions predisposing to late-life depression. The limbic system evaluates reward-related stimuli and organizes stress responses, and higher neocortical structures exert control over these processes. Insula metabolism may be a treatment-specific biomarker in major depression. Specifically, insula hypometabolism was associated with remission in response to cognitive-behavioral therapy and poor response to escitalopram, while insula hypermetabolism was associated with remission to escitalopram and poor response to cognitive-behavioral therapy in a mixed-age sample. Reduced bilateral activation of the dorsal-anterior cingulate and the hippocampus was demonstrated in severely depressed, nondemented older adults performing a word-activation task. Low resting functional connectivity within the cognitive control network and high-resting functional connectivity within the default mode network distinguished depressed from normal older adult subjects. Low resting functional connectivity within the cognitive control network predicted a low remission rate and persistence of depressive symptoms and signs, apathy, and dysexecutive behavior after treatment with escitalopram. In contrast, resting functional connectivity within the default mode network was correlated with pessimism but did not predict treatment response.

Patients with DED syndrome often present with apathy. Resting functional connectivity studies of apathy in late–late-life depression showed that apathy is characterized by abnormalities in the cognitive control network as well as in reward and salience networks.

Depression frequently develops in disorders of subcortical structures, including vascular dementia, Parkinson disease, Huntington disease, supranuclear palsy, basal ganglia calcification, and stroke of the caudate head. WMH are common in geriatric depression and are often located in subcortical structures and their frontal projections, compromising connections between limbic and dorsal cortical structures. WMH may interfere with limbic–dorsal balance and lead to chronic depressive syndromes. Subcortical WMHs have been associated with both executive dysfunction and nonremission of late-life depression.

Complementing these findings, diffusion tensor imaging studies documented that reduced white matter integrity in distributed cerebral networks that connect the prefrontal cortex with subcortical and posterior cortical regions (dorsal and rostral anterior cingulate cortex [ACC], dorsolateral-prefrontal cortex [DLPFC], hippocampus, posterior cingulate, insula, neostriatum, and the midbrain as well as select temporal and parietal regions) are associated with executive dysfunction and poor response of late-life major depression to a serotonin reuptake inhibitor. Similarly, lower activation in the DLPFC and other brain regions during an in-scanner performance of the Wisconsin Card Sorting Test of executive function predicted a less favorable response to cognitive behavioral therapy in depressed older adults. Depressed, older, short serotonin transporter 5-HTTLPR allele carriers had both a lower remission rate than long allele homozygotes and more microstructural white matter abnormalities in frontolimbic areas. These abnormalities may be caused by inflammatory responses, vascular disease, prominent aging-related changes, degenerative brain disease, or an accumulation of those and other factors.

The Inflammation Hypothesis

The inflammation hypothesis postulates that aging-related immune processes promote changes in the emotional and cognitive neural networks predisposing to geriatric depression and/or triggering metabolic brain changes mediating the depressive syndrome in late life. The hypothesis is supported by emerging findings suggesting that: (1) aging leads to a proinflammatory state; (2) immune responses of the CNS can influence the function of some of the emotional and cognitive networks pertinent to geriatric depression; (3) elevation of peripheral cytokines and acute phase proteins (mainly IL-6, Tumor Necrosis Factor (TNF), and C-reactive protein) is associated with depressive symptoms in older adults; (4) some antidepressants reduce the expression of several inflammation markers in the periphery; (5) blocking peripheral cytokines from disrupting the blood–brain barrier (BBB) exhibited antidepressant actions, and (6) pretreatment of cytokine plasma levels may influence the response to antidepressants. Dysregulation of both the innate and adaptive immune systems occurs in depressed patients and hinders favorable prognosis, including antidepressant responses.

Aging promotes a chronic state of neuroinflammation, and animal and human studies suggest that this proinflammatory state may compromise neural systems related to depression. In aged rodents, the cerebral cortex produces more inflammatory cytokines than the cortex of adult animals. In older animals, the inflammatory cytokine IL-6 has higher concentrations in the hippocampus, cerebral cortex, and cerebellum than in the same regions of younger animals. In normal human adults, enhanced activity of the subgenual anterior cingulate cortex and reduced functional connectivity (FC) of the subgenual anterior cingulate with the amygdala, medial-prefrontal cortex, and nucleus accumbens during emotional face processing was modulated by a peripheral inflammatory response (IL-6). Cytokines induce indoleamine 2,3-dioxygenase, an enzyme that reduces serotonin production. They also dysregulate the glutamate system, promote excitotoxicity and decrease the production of neurotrophic factors, neuroplasticity, and neurogenesis. Inflammatory responses to immune challenge influence the function of emotional networks. A single nucleotide polymorphism (SNP) encoding IL-1b has been associated with both reduced activity of the anterior cingulate and the amygdala in response to emotional probes and with poor response of major depression to an antidepressant. Patients treated with the cytokine interferon-α (IFN-α) have greater activation of the dorsal-anterior cingulate than controls. Administration of IFN-α also increased glutamate in the basal ganglia of nondepressed subjects with hepatitis C (mean age 45). Changes in glutamate concentrations were, in turn, associated with anhedonia and psychomotor slowing. Cytokines contribute to oxidative stress, which damages glial cells in the prefrontal cortex and the amygdala. Inflammation may cause resistance to glucocorticoids in immunocytes and their cellular targets.

Peripheral inflammatory markers are elevated in late-life depression, and their levels are associated with the severity of depression and with cognitive symptoms of depression. A recent meta-analysis of 82 studies showed that peripheral levels of IL-6, TNF-α, IL-10, soluble IL-2 receptor, C–C chemokine ligand 2, IL-13, IL-18, IL-12, IL-1 receptor antagonist, and soluble TNF receptor 2 were elevated and interferon-gamma levels were lower in individuals with major depression compared to controls. Several of these inflammatory markers (TNF-α, IL-6, and IL-1b) were found to correlate with the severity of depressive and cognitive symptoms. Increased inflammatory markers have been associated with atypical symptoms of depression and with suicidal behavior in MDD. In brain autopsies of suicide victims, increased cytokine concentrations have been found in the orbitofrontal cortex and the dorsolateral prefrontal cortex. Higher levels of IL-6 predicted the chronicity of depression, as well as higher severity of depression at follow-up. Elevated IL-6 levels have also been associated with increased suicide risk, with the highest levels of IL-6 correlating with the most violent suicide attempts.

Proinflammatory changes occur in various medical illnesses and health risk factors predisposing to late-life depression. Increased circulating inflammatory cytokines have been found in cardiovascular disease. High body mass index and smoking have been associated with increased inflammatory markers in major depression. Chronic stress, a precipitant of depression, increases circulating IL-1β and IL-6 as well as cognitive impairment in older adults. Childhood trauma has been shown to cause long-lasting effects on peripheral inflammation later in life and is associated with an increased risk of depression. Although peripheral cytokines do not cross the blood–brain barrier, they send signals via molecular, cellular, and neural routes, which ultimately enhance brain inflammation. Aging may exacerbate the effects of stress in the brain, leading to behavioral and cognitive changes similar to those of depressive syndromes.

A large community-based prospective cohort study spanning over 20 years monitored high-sensitivity C-reactive protein (CRP) among individuals from middle-to late-life. Those with repeated or chronic elevations in systemic inflammation had greater symptoms of depression than older adults. These associations occurred independently of medical comorbidity and did not differ based on late-life cognitive status or *APOE* ε4 genotype.

The role of inflammation in depression treatment response has been studied. Tricyclic antidepressants (TCAs) may reduce some proinflammatory cytokines, increase anti-inflammatory cytokines, and reduce polysaccharide-induced “sickness behavior” in animals. Pretreatment cytokine plasma levels may influence the response to antidepressants in humans. SSRI-resistant patients had higher TNF-α plasma levels, while euthymic patients with a history of depression had TNF-α levels similar to those of normal controls. A higher level of TNF-α was shown to predict failure of depression to respond to escitalopram. The largest study to date reported that patients with baseline low C-reactive protein (CRP <1 mg/L) levels had greater improvement in depression when treated with escitalopram than with nortriptyline. In contrast, patients with higher CRP levels had greater improvement when treated with nortriptyline than with escitalopram. In treatment-resistant major depression, baseline concentrations of high-sensitivity CRP (>5 mg/L), TNF, and TNF-soluble receptors were associated with a favorable response to the TNF antagonist infliximab.

In addition to influencing treatment response, neuroinflammation has also been speculated to underlie the relationship between depression and neurodegenerative diseases. The exaggerated release of proinflammatory cytokines may lead to neuronal damage. Such damage can interfere with anti-inflammatory and immunosuppressant regulation, leading to cognitive impairment, neurodegeneration, and decreased neurogenesis in different brain areas, including the hippocampus. Reduced hippocampal volume is one of the early brain changes associated with dementia.

The Vascular Depression Hypothesis

This hypothesis postulates that cerebrovascular disease predisposes, precipitates, or perpetuates some late-life depressive syndromes. The hypothesis was based on the high comorbidity of depression and vascular risk factors; the high incidence of depression poststroke; the high prevalence of WMHs in late-onset depression; the high frequency of cognitive impairment in depressed patients with vascular risk factors; and the similarity of cognitive abnormalities of depression to those associated with WMH. Recent studies show greater WMH severity in specific white matter fiber tracts in late-life depression, including the cingulum bundle, uncinate fasciculus, and superior longitudinal fasciculus. Fiber tract structural connectivity is positively correlated with resting state functional connectivity of connected regions.

A syndrome of “subcortical ischemic depression” has been proposed based on WMH literature. This syndrome is defined by the combination of depression and MRI evidence of subcortical ischemic changes. Age, lassitude, and a history of hypertension were positively associated with the diagnosis of subcortical ischemic vascular depression, whereas a family history of mental illness and loss of libido were negatively associated with this diagnosis.

MRI stigmata of cerebral small vessel disease (i.e., WMHs, lacunes, microbleeds, perivascular spaces, and cerebral atrophy) are associated with depression and stroke. Studies found that a history of depression increases the risk of vascular disease and stroke and may promote inflammation or epigenetic changes of genes related to vascular integrity, suggesting that depression has a bidirectional relationship with vascular disease.

Patients with vascular depression often present with psychomotor retardation, anhedonia, lack of insight into their illness and disability, and are less likely to report feelings of guilt. WMHs are associated with executive dysfunction and reduced activation of brain regions related to executive and psychomotor functions. Executive dysfunction was associated with bilateral WMH in the inferior frontal white matter, temporal-occipital periventricular white matter, and the anterior limb of the internal capsule, as well as scattered clusters in the prefrontal white matter. WMHs were correlated with impairments in goal maintenance during a continuous performance task but also with reduced activity in DLPFC and reduced connectivity of the DLPFC with task-relevant brain regions, including the middle frontal gyrus and supramarginal gyrus.

WMH and microstructural brain abnormalities predict a poor response of late-life depression to antidepressants. Increases in WMH volume over

2- and 4-year intervals are associated with nonremission or relapse. Microstructural white matter abnormalities in multiple frontolimbic brain areas, including the rostral- and dorsal-anterior cingulate, DLPFC, genu, hippocampus, posterior cingulate, left superior corona radiata, and right inferior longitudinal fasciculus are associated with nonremission.

The disconnection of structures regulating mood and cognition may be one of the mechanisms by which cerebrovascular disease promotes late-life depression. Another mechanism may be vascular dysregulation leading to reduced cerebral blood flow (CBF). Mild CBF reduction may impair cognitive and affective processes, while greater CBF reduction in the context of autoregulatory deficits may cause ischemic injury.

Vascular diseases impair vascular regulation. Vascular changes, including increased intima-media thickness, increased arterial stiffness, and endothelial dysfunction, are pronounced in late-life depression, and endothelial function may be most impaired in nonresponders to antidepressants. In late-life depression, perfusion deficits are more severe, particularly in the medial- and lateral-prefrontal cortices and subcortical and temporal structures.

Arguably, at least three interconnected processes may lead to “vascular depression,” that is, disconnection of networks responsible for the regulation of mood and cognition, reduced regional CBF, and inflammatory responses related to vascular disease. As an example, inflammation leads to vascular remodeling, accelerates vascular damage, and may contribute to perfusion deficits. In turn, regional perfusion deficits adversely affect gray matter function and contribute to the development of WMHs that disrupt regional structural and functional connectivity.

Genetic Studies

It is well established that the risk for MDD is partially mediated by genetic factors. A meta-analysis quantified the increased risk of MDD as an odds ratio (OR) of 2.84 (95% CI = 2.31–3.49) for first-degree relatives of individuals with MDD. A large twin study in a Swedish population estimated the heritability of MDD to be approximately 37%, which is consistent with estimates made in previous studies. In more recent years, technologic advances have enabled researchers to move beyond family-based heritability studies of MDD to identify disease-associated susceptibility loci using genome-wide association studies (GWAS). What has become clear from these studies is that MDD is a polygenic disorder, with multiple loci being identified, but each with a small effect. A recent genome-wide meta-analysis of MDD analyzed data from 807,553 individuals of all ages had identified 102 independent variants, 269 genes, and 15 gene-sets associated with depression. In an independent replication sample of 1,306,354 individuals, 87 of the 102 associated variants remained significant following multiple testing corrections. This meta-analysis identified significant genetic correlations among the general population for depression with neuroticism (which was bidirectional), anorexia nervosa, attention-deficit hyperactivity disorder (ADHD), schizophrenia, and BD. Interestingly, previously identified depression-associated genes were not identified in this study. These genes include those associated with the serotonergic system, such as SLC6A4, GNB3, HTR2A, and TPH2. This is unexpected, as the serotonergic system forms the basis of most antidepressant treatments but could indicate a functional separation between genetic pathways of depressive disease and pathways of antidepressant treatment.

In many cases, it is possible to identify associations between particular genes and characteristics (endophenotypes) that may predispose an individual toward the development of depression. One such endophenotype is hippocampal volume loss. In a cohort of 138 older adults aged 60 and above, genetic variants at three *AGTR1* SNPs (a gene encoding the angiotensin-II type 1 receptor of the renin-angiotensin system [RAS]) were independently associated with accelerated hippocampal volume loss in the right hemisphere over a 4-year follow-up period. In addition, the *AGTR1* risk alleles predicted worse episodic memory performance and interacted with the *APOE4* allele to accelerate right hippocampal volume loss. These initial findings suggest that targeting molecules in the RAS may have the benefit in late-life depression. Future work will involve investigating the effects of genetic variation on other relevant endophenotypes of mood disorders in older adults.

Genetic studies using SNP heritability approaches have provided evidence suggesting that some loci may contribute to both unipolar major depression and BD, particularly those involving calcium channel signaling. GWAS in BD has identified several risk-conferring variants, including markers within the *CACNA1C* gene (encoding the α subunit of the L-type voltage-dependent calcium [Ca^{2+}] channel CaV1.2) and the *ANK3* gene (encoding ankyrin G, a cytoskeletal protein involved in scaffolding in the axon initial segment and in nodes of Ranvier), which may be mechanistically linked through providing signaling integrity in GABAergic neurons. Initial findings have shown that variants in these genes may functionally impact neuronal circuitry involved in affective processing in individuals with BD carrying the risk variants. An additional study analyzing four independent genome-wide data sets along with data contributed to the Psychiatric Genomics Consortium Bipolar Group identified the *CACNA1C* gene as well as pathways involved in hormonal regulation, second messenger systems, glutamate systems, and neuronal development, as contributing to the genetic predisposition and underlying neurobiology of BD.

In addition to work studying candidate loci, recent evidence also suggests that epigenetic mechanisms, processes affecting gene expression and translation that do not involve changes in the DNA sequence, could play a role in depression neurobiology and risk. Through one model, epigenetic mechanisms could mediate the lasting increases in depression risk following exposure to an adverse life event, such as childhood trauma, which is strongly associated with depression in adulthood. An additional model suggests that underlying genetic and environmental factors, and interactions between the two, drive aberrant epigenetic mechanisms targeting stress response pathways, neuronal plasticity, and other behaviorally relevant pathways that have been implicated in MDD.

A Model of Late-Life Depression

Accumulation, and perhaps a synergy of noxious factors, is consistent with the clinical reality of geriatric populations in whom a variety of processes contribute to brain dysfunction, leading to depression and to cognitive impairment. Etiologic factors, directly or by causing brain abnormalities increasing vulnerability to depression, promote metabolic changes mediating the expression of the depressive syndrome. Etiologic factors include cerebrovascular lesions, perfusion-related abnormalities, and inflammatory and repair responses. Abnormalities in frontolimbic and frontostriatal functional architecture resulting from these processes may serve as vulnerability factors for depression. Heredity, epigenetic factors, and psychological vulnerability represent additional predisposing factors. Biologic responses to stress may promote processes serving as etiologic or vulnerability factors of late-life depression.

A working model of late-life depression postulates that depressive syndrome represents the clinical expression of dysfunction in reward, salience, and cognitive control networks. The degree of dysfunction may determine the intensity of symptoms related to mood, cognition, and/or motoric behavior and account for the heterogeneous clinical presentations of late-life depressive syndrome. Abnormalities in the frontolimbic system are involved in functional abnormalities, chronicity, and relapse of depressive symptoms. Genetic factors, aging, and disease-related processes (e.g., inflammation, vascular disease, neurodegenerative disease pathology accumulation [e.g., Aβ and tau]) may cause dysfunction in reward, salience, and cognitive control networks. Many of these factors, as well as psychological factors, start in mid-life and continue into late-life. These stressors might lead to depression directly, through inflammatory responses and structural brain changes, or indirectly through neglect of health. This model seeks to consolidate hypotheses surrounding etiologic, predisposing, and stress-related factors and mechanisms mediating the behavioral expressions of late-life depression and the course of illness. Although the hypotheses are described separately, some of their mechanisms overlap, and additional mechanisms may be at play.

Brain Structure Studies

Neuroradiologic studies suggest frontostriatal and limbic impairment in depression. WMHs prevalent in geriatric depression are mainly located in subcortical structures and the medial-orbitofrontal cortex. An increase in the volume of WMHs over a 2-year period predicted lower rates of sustained late-life MDD remission during the same period, and worsening of white matter grade in older adults over a 5-year period has been associated with cognitive decline.

Histopathologic and neuroimaging studies of individuals with mood disorders have provided evidence of microscopic and macroscopic changes in several brain regions. Low volumes of the subgenual anterior cingulate, the head of the caudate nucleus, the putamen, and the hippocampus have been noted in mixed-age depressed patients. Reduced hippocampal volume has also been found in depressed older adults. Small left hippocampal volume in geriatric depression was associated with the later development of dementia. Reduction in orbitofrontal cortex volume has been reported in depressed older adults compared to normal controls. Abnormalities in neurons of the DLPFC have been observed in unipolar and BD. Moreover, a reduction in glia of the subgenual prelimbic anterior cingulate gyrus has been reported in unipolar-depressed patients. Finally, amygdala volume findings in depressed patients have been mixed.

Brain Function Studies

Increased metabolism has been noted in limbic regions, including the amygdala, the rostral-anterior cingulate, and the posterior orbital cortex as well as anterior (right and left superior frontal gyri) and posterior (precuneus, inferior parietal lobule) cortical regions. In contrast, dorsal neocortical areas, such as the lateral and DLPFC, and the dorsal-anterior cingulate, as well as the caudate, have reduced blood flow during depression. Reduced bilateral activation of the dorsal-anterior cingulate and the hippocampus was demonstrated in severely depressed, nondemented older adults performing a word-activation task.

The limbic system evaluates reward-related stimuli and organizes stress responses. Reciprocal pathways link limbic structures with striatal and neocortical sites. It has been proposed that abnormalities in the reciprocal relationship between ventral limbic and dorsal cortical neural systems are critical physiologic disturbances in depression. White matter abnormalities compromising connections between limbic and dorsal cortical structures may interfere with limbic–dorsal balance and lead to chronic depressive syndromes.

A recent study provided evidence for insula metabolism as a treatment-specific biomarker in MDD. In this positron emission-based study, insula hypometabolism (relative to whole-brain levels) was associated with remission in response to cognitive-behavioral therapy and poor response to escitalopram, while insula hypermetabolism was associated with remission to escitalopram and poor response to cognitive-behavioral therapy in a mixed-age sample. This finding will require replication in an independent prospective sample and also in an older-aged sample but offers promise for the use of brain-based biomarkers in guiding treatment selection in MDD. Other studies found that treatment with SSRI or total sleep deprivation (TSD) resulted in decreased metabolism in anterior cortical and limbic regions and increased posterior cortical metabolism in geriatric depression. The metabolic changes after SSRI treatment were associated with improvement in depressive symptoms. In behavioral and imaging studies of the brain as people age, patterns of resilience to negative emotional experiences and reactivity to positive emotional experiences were noted. Older adults exhibit less reactivity to negative stimuli and greater memory for positive information. However, older adults who have damage to areas of the prefrontal cortex related to emotion regulation are more likely to experience negative outcomes such as depression.

COURSE AND OUTCOME

Geriatric Depression

Knowledge of risk factors and periods of high risk for particular adverse outcomes enables the clinician to use appropriate diagnostic methods and offer preventive treatments with favorable risk-benefit ratios. Despite advances in antidepressant therapies, longitudinal studies of 1 to 6 years duration suggest that 7% to 30% of geriatric patients have chronic major depression. If partially remitted subjects are considered chronic, the rate of chronicity reaches 40%. Chronicity of depression may be predicted by a history of a long current episode or long previous episodes, coexisting medical illness, high severity of depression, nonmelancholic presentation, and delusions.

Naturalistic treatment studies of geriatric populations observed a 13% to 19% rate of relapse/recurrence at 1 year, while a 15% relapse rate is observed in patients receiving controlled antidepressant treatment. This percentage is lower than the relapse/recurrence rate (34%) reported in younger adults. However, if the follow-up of older adults with depression is extended from 3 to 6 years, the recurrence rate increases to 38%. A 20-year follow-up study of mixed-age patients with depression showed that 95% of patients had a recurrence. Patients with only minor or subsyndromal depression have an increased risk (approximately fivefold) of major depression compared with unaffected individuals after 1 year. Depression outcomes are also predicted by initial medical burden, self-rated health, and social support- both objective and perceived.

Relapse/recurrence appears to have different predictors than a recovery of depression. The likelihood of relapse/recurrence may be high in patients with a history of frequent episodes, late age of illness onset, history of dysthymia, intercurrent medical illness, greater functional impairment, and higher baseline depression, and those in whom there is a perceived external locus of control.

Geriatric Bipolar Disorder

In geriatric BD, relapse and hospital readmission after manic episodes are common as episodes become more frequent, and the interval between episodes becomes shorter with age. In comparison to younger patients with BD, there is evidence to suggest that geriatric BD patients have less severe manic episodes, are more likely to relapse into depression after a manic episode, are less likely to experience psychotic symptoms, are more likely to experience rapid cycling between episodes, and are less likely to attempt suicide.

EVALUATION AND TREATMENT OF GERIATRIC DEPRESSION AND BIPOLAR DISORDER

Overview of Geriatric Depression Evaluation and Treatment

The objectives of geriatric depression treatment include:

1. Remission of depression
2. Reduction in the risk of relapse and recurrence
3. Improvement of cognitive and functional status
4. Development of skills or provision of support needed for coping with handicaps and psychosocial adversity.

Clinicians should expect an improvement in both disability and excess medical morbidity with adequate treatment.

Geriatric depression can be effectively treated with psychotherapy, pharmacotherapy, or combination treatment. Appropriate behavioral rehabilitation techniques should be combined with antidepressant treatment in order to enable older adults to regain function as depressive syndrome subsides.

Long-term outcomes of depressed older adults can be improved by

1. Identifying and treating comorbid psychiatric and medical conditions
2. Minimizing medication side effects
3. Striving for full remission of acute symptoms

4. Providing psychoeducation for patients, families, and health care professionals about the illness and the importance of adhering to treatment recommendations, both acutely and after symptom remission.

The evaluation of geriatric depression should include assessment of psychopathology, medical and neurologic status, pharmacotherapy, functional impairment, and psychosocial factors, as well as suicidal risk. For instance, neurocognitive disorders, anxiety, sleep disorders, use of drugs that may cause depression (e.g., steroids), and benzodiazepine and alcohol abuse often contribute to late-life depression and worsen its course. Therefore, attention to these problems should be part of the assessment of depressed older adults. A further complexity of assessment includes the risk of missing a bipolar depression diagnosis. As such, a diagnosis would alter clinical management and contraindicate the use of antidepressant monotherapy. Rating scales, including the Hamilton Depression Rating Scale, the Cornell Scale for Depression in Dementia, the Geriatric Depression Scale, and the Mini-Mental State Examination or Montreal Cognitive Assessment (MOCA), can be useful in assessment.

Adults with comorbid medical illnesses, including both chronic/deteriorating processes and acute/debilitating events, are preferentially affected by late-life depression. Further, a number of other medical conditions, such as viral infections, endocrinopathies (e.g., thyroid and parathyroid abnormalities), vitamin B_{12} deficiency, and malignancies (e.g., lymphoma and pancreatic cancer), are often associated with depression. Vitamin D deficiency has also been linked to late-life depression. Medical conditions likely to respond promptly to treatment (e.g., amphetamine withdrawal) should be treated first, and antidepressants should only be used if depression persists. For many other medical conditions, however, depression may not remit until an antidepressant agent is used. For example, depression in a hypothyroid older adult patient rarely responds to thyroid supplementation alone. Similarly, an antidepressant trial often is ineffective before hypothyroidism is corrected. Thyroid supplementation and an antidepressant drug, however, offer the highest probability of antidepressant response.

Some drugs can contribute to or cause late-life depression, such as steroids, antiparkinsonian drugs, benzodiazepines, and β-blockers. If the depression occurs in a patient receiving methyldopa (Aldomet), cimetidine (Tagamet), clonidine (Catapres), hydralazine, propranolol (Inderal), or reserpine, switching to another medication is recommended. If the depression occurs in a patient treated with digitalis, estrogens, tamoxifen (Soltamox), vinblastine, or vincristine, the medication should be continued, and depression should be treated with antidepressants. There is no agreement on the best strategy for depression that occurs in patients treated with benzodiazepines, corticosteroids, progesterone, propoxyphene (Darvon), or psychostimulants. As interferon-α treatment frequently is associated with depression, increased frequency of psychiatric follow-up and starting or increasing antidepressant treatment prophylactically should be considered before initiating such treatment in patients with a history of depression.

The evaluation of geriatric depression should also include assessments of the psychosocial situation and functional capacity of a patient, as functional difficulties can lead to physical limitations and social isolation, which can further worsen the course of depression. As a more extreme example, poststroke patients experience a “psychosocial storm” characterized by individual strain (e.g., difficulties with ADLs, demanding rehabilitation for the patient) as well as family/caregiver strain (e.g., alteration of usual family roles in accommodating the increased need of the poststroke family member) that increase feelings of helplessness, hopelessness, and depression in both the patient and the family/caregiver. Indeed, depressive symptoms seem to be increased in caregivers of patients with chronic illness, which is particularly pronounced in caregivers of patients with dementia. Other psychosocial factors to consider include race and ethnicity, as there is evidence for disparities in antidepressant medication prescriptions and symptom severity between different racial and ethnic groups. Further, evaluation of adverse childhood experiences (e.g., physical, sexual, and emotional abuse) should be a part of the assessment, as the negative effects of repeated adverse childhood experiences do not diminish in later life and are associated with depression in older adults.

Psychotherapy for Geriatric Depression

For the treatment of major depression in later life, current practice guidelines recommend a combination of pharmacotherapy and psychotherapy, with psychotherapy alone often sufficing as a first-line treatment for minor depression in later life. Psychotherapy may play an essential role in the care of bereaved depressed older adults or in patients who have significant life crises, lack social support, develop disabilities, or lack coping skills to deal with their life situations. Psychotherapy may be helpful in supporting older adults to cope with their symptoms and adhere to treatment before antidepressant drugs take effect, and it should be considered in patients who cannot tolerate biologic treatments. Telepsychiatry may offer an efficacious option for psychotherapy delivery for homebound older adults. One study showed that an integrated telehealth service for older adults with depression was better than treatment as usual. Telepsychiatry was widely adapted during the COVID-19 pandemic to help meet the mental health needs of the geriatric population at a time when in-person visits posed significant exposure and health risks. A study of behavioral activation delivered via telepsychiatry found that it was effective at increasing social interaction, satisfaction with social support, and reducing disability.

A number of psychotherapeutic modalities, including interpersonal therapy, cognitive-behavioral therapy, and problem-solving therapy (PST), have been found effective in depressed older adults without significant cognitive impairment or medical burden. For patients with executive dysfunction, a number of psychotherapies have been found effective, including a modification of PST where the therapist is more directive, the focus is on less complex problems to facilitate learning, and the program is more structured. PST has been shown to be more effective than supportive therapy in reducing depression and disability in older adults with depression and executive dysfunction. Additionally, problem-adaptation therapy (PATH) was developed for older individuals with major depression and intermediate cognitive impairment and disability; this modality combines PST, environmental adaptation tools, and caregiver involvement to target the behavioral limitations inherent in depression and disability.

Extending the focus on the environment, ecosystem-focused therapy (EFT) has been developed to support patients through the “psychosocial storm” of an acute medical illness where depression is often a major contributor to disability and a hindrance to recovery. EFT has been developed as a treatment for poststroke depression. A meta-analysis showed that antidepressants are only weakly efficacious in poststroke depression and the efficacy of conventional psychotherapies is questionable. The aim of EFT is to help the patient develop new perspectives and skills to confront the significant changes to their ecosystem brought on by these acute changes. A preliminary study showed that EFT is more efficacious than education in stroke and depression.

According to a meta-analysis of psychotherapy studies in a wide range of late-life depressive syndromes, psychotherapy, and pharmacotherapy have comparable efficacy. Although psychotherapy has proven to be an effective treatment in this population, it is often an underutilized resource. One idea behind the underutilization is that the complexity of available psychotherapies makes their implementation difficult in a general community setting.

To streamline both the delivery and language of psychotherapy interventions, it has been proposed that neurobiologic constructs be used to identify biologically meaningful behavioral targets that can serve as the basis for simplified and effective treatments. This view of treatment is supported by the RDoC project and has been successfully applied to the treatment of late-life depression with the development of “Engage,” a psychotherapy using simple-to-learn strategies that target accepted neurobiologic domains of late-life depression. Initial studies suggest that ENGAGE can be taught to community-based therapists, is noninferior to PST, and engages its behavioral targets.

Pharmacotherapy for Geriatric Depression

Overview. Older adults benefit from the same psychopharmacologic agents as younger patients. Benzodiazepines should generally be avoided. Residual anxiety should be treated with an increase of the dosage of the antidepressant to maximum levels rather than prescribing a benzodiazepine. Residual insomnia in depressed older adults can be treated with alternate agents such as trazodone (Desyrel). Orexin antagonists, including suvorexant and lemborexant, may also be considered for short-term treatment of insomnia in the geriatric population. Data shows suvorexant is well tolerated in the geriatric population with adverse events ratings comparable to placebo, with mild adverse effects including daytime sedation and sleepiness. Nonbenzodiazepine medications, including zolpidem (Ambien) or zaleplon (Sonata), have improved safety profiles compared to benzodiazepines. However, they also have side effects in this population, including cognitive impairment, serious injury, and fractures.

Clinical practice guidelines recommend that SSRIs and other second-generation antidepressants (e.g., serotonin-norepinephrine reuptake inhibitors [SNRIs], bupropion, mirtazapine) are the first-line treatment for late-life depression. Late-life depression responds more slowly to antidepressants than the depression of younger adults. Executive dysfunction, slow cognitive processing speed, WMH, microstructural white matter abnormalities, and small hippocampal volumes each predict a slower and less complete response to antidepressant treatment. Older adult patients who respond to antidepressant treatment usually require 8 to 12 weeks to achieve full remission. The minimal length of an antidepressant trial should be 3 to 4 weeks at an adequate dose before a switch to another antidepressant or an augmentation agent is used. Patients who show no signs of improvement by the fourth week, however, will not likely respond to further treatment with the same agent.

Selective Serotonin Reuptake Inhibitors. Studies of older adult outpatients have shown that SSRIs are equally effective as tertiary amine tricyclics in the acute treatment of late-life depression. Recent findings suggest that SSRIs may delay the onset of AD in individuals with a previous history of depression, another reason to consider their use in this population. The dosages of SSRIs should be increased gradually. The starting daily dosages are 10 to 20 mg for fluoxetine, paroxetine, and citalopram, 5 to 10 mg for escitalopram and vortioxetine, and 25 to 50 mg for sertraline. For some patients, daily dosages of 20 mg of fluoxetine, paroxetine, and citalopram, 10 mg of escitalopram, and 100 mg of sertraline are sufficient, although higher dosages may be required. The most frequent side effects of SSRIs are insomnia, akathisia, nausea, anorexia, and pseudoparkinsonism. Inappropriate secretion of antidiuretic hormone (SIADH) leading to hyponatremia is much more common in older adults than in younger adults. Further, SSRIs may lead to an increased risk of bleeding, fragility fractures, poor mobility, and stroke.

Drug interactions should be considered in older adult patients receiving SSRIs. Fluoxetine and paroxetine inhibit the cytochrome P450 2D6 liver cytochrome isoenzyme. This pathway is essential for the hydroxylation of nortriptyline and desipramine and the metabolism of antipsychotics and type 1A antiarrhythmic drugs (encainide, flecainide), β-blockers, and verapamil, whose plasma levels may be raised when fluoxetine or paroxetine is administered. The metabolite of fluoxetine, norfluoxetine, is an inhibitor of the 3A4 isoenzyme. The 3A4 isoenzyme is responsible for the metabolism of alprazolam (Xanax), carbamazepine (Tegretol), quinidine, erythromycin, and astemizole and may lead to an increase in the plasma level of these agents. Sertraline, citalopram, and escitalopram have a weak inhibitory action of the 2D6 isoenzyme and do not inhibit the 3A4. Fluvoxamine (Luvox) inhibits the P450 3A4 and the 1A2 cytochrome isoenzymes but does not significantly inhibit the 2D6 isoenzyme. Drugs metabolized by the 3A4 should be avoided in patients receiving fluvoxamine. Similarly, theophylline should be cautiously prescribed in fluvoxamine-treated patients since fluvoxamine may produce a threefold decrease in theophylline clearance by inhibiting the 1A2 isoenzyme.

Vortioxetine is a newer SSRI that is a serotonin 1A receptor agonist and a serotonin 3 receptor antagonist. It is also an antagonist of the serotonin 1D and 7 receptors and a partial agonist of the serotonin 1B receptors, although with somewhat lower affinity. Vortioxetine has been found to be effective in late-life depression. It may improve cognitive domains, including memory and executive dysfunction in the context of depression, The general starting dose is 10 mg daily, and a range of 5 to 20 mg daily is efficacious. Vortioxetine neither inhibits nor induces liver enzymes. It is metabolized by the 2D6 isoenzyme, and its dosage should be adjusted when coadministered with 2D6 inhibitors or inducers.

Vilazodone was approved by the FDA for the treatment of MDD in 2011 and is both an SRI and a partial agonist at the 5-HT1A receptor. Data on vilazodone in the geriatric population is limited, though vilazodone performed comparably to paroxetine in late-life depression in one study. Vilazodone is metabolized by CYP450 3 A4 and is highly protein bound, so it should be used with caution in combination with other drugs that are highly protein bound, such as digoxin or warfarin. The starting dose is 10 mg/day and is then titrated to a therapeutic dose of 20 to 40 mg/day. Adverse effects include diarrhea, nausea, and headaches.

Serotonin-Norepinephrine Reuptake Inhibitors. Venlafaxine XR (extended-release) leads to high remission rates (asymptomatic states) and has been found effective in hospitalized depressed patients as well as in drug-resistant depression and depressed patients with chronic pain. For this reason, venlafaxine should be considered in severe depression and in depression unresponsive to other agents. Older adult patients appear to require dosages comparable to those of younger adults. Daily dosages of 112.5 to 225.0 mg are adequate for the majority of older adults. Nausea, a rather frequent side effect of venlafaxine, can be minimized by a slow increase in the daily dosage. Blood pressure should be monitored, especially in patients receiving dosages above 225 mg daily, since venlafaxine can increase blood pressure. Venlafaxine has limited drug–drug interactions and may be selected in older adults who require treatment with drugs interacting with other antidepressants. Venlafaxine has a short elimination half-life, but the extended-release preparation can be administered in a single dose. Nonetheless, abrupt discontinuation of venlafaxine XR may cause a withdrawal syndrome.

Duloxetine, an SNRI with relatively balanced serotonergic and noradrenergic action, has demonstrated efficacy in late-life depression as well as the treatment of comorbid pain. Following a week of 30 mg/day, the dose can usually be increased to 60 mg/day after a week of treatment if tolerated, with additional time on that does not associated with an increased chance of remission. For patients who remit on a dose of 60 mg, remission will typically occur by week 12 of treatment; additional time at that dose is not associated with an increased chance of remission.

Levomilnacipran is an SNRI with higher selectivity for norepinephrine reuptake inhibition than other SNRIs and has proven effective for major depression, although studies specific to late-life depression are ongoing. There is some evidence to suggest that individuals older than 60 years of age may have a greater antidepressant response to this medication than younger individuals. There is also evidence, however, suggesting that levomilnacipran may be less well tolerated in older adults with depression. Levomilnacipran has also been shown to inhibit the beta-site amyloid precursor protein cleaving enzyme-1. An investigation is currently underway to determine whether it may prevent beta-amyloid plaque formation. Levomilnacipran is largely metabolized by cytochrome P450 3A4. The medication is intended for once-daily dosing with a starting dose of 20 mg daily. The target dose range is 40 to 120 mg daily.

Tricyclic Antidepressants. The secondary amines, nortriptyline, and desipramine are the most frequently used TCAs in geriatric depression. They have lower anticholinergic and sedative effects than the tertiary amines amitriptyline, doxepin, and imipramine. Nortriptyline appears to have a lower potential for orthostatic hypotension than other TCAs. The plasma levels of antidepressants required for the treatment of depressed older adult patients are similar to those needed by young adults (nortriptyline plasma levels of 60 to 150 ng/mL and desipramine above 115 ng/mL). Older adult patients often develop therapeutic blood levels of nortriptyline or desipramine while on low daily dosages (nortriptyline 1.0 to 1.2 mg/kg of body weight and desipramine 1.5 to 2.0 mg/kg). Older patients are more likely than younger patients to develop delirium, constipation, urinary retention, dry mouth, and orthostatic hypotension from these medications. For this reason, the dosages of antidepressants should be increased at a slower pace than in younger adults. Older adults on nortriptyline more often develop higher plasma levels of the nortriptyline metabolite 10-hydroxy-nortriptyline than do young adults, even when the plasma levels of the parent compound are similar. High 10-hydroxy-nortriptyline levels may contribute to cardiac conduction defects in older adults.

Orthostatic blood pressure, the electrocardiogram (ECG), and urination should be monitored frequently in older adults receiving nortriptyline or desipramine. Pretreatment systolic orthostatic hypotension has been found to correlate with antidepressant response to nortriptyline in some older adults. Therefore, nortriptyline, rather than desipramine, should be considered in such patients, and orthostatic blood pressure and subjective symptoms of orthostasis should be monitored carefully. TCAs have anticholinergic properties. For this reason, these drugs should be avoided in patients with prostatic hypertrophy or patients with narrow-angle glaucoma. Nortriptyline and desipramine have properties similar to those of type 1A antiarrhythmic drugs (quinidine-like drugs). When administered to patients with right or left bundle branch block, tricyclics may cause a second-degree block in approximately 10% of cases. For this reason, an ECG should always precede the use of tricyclics in older adults. A multicenter study (CAST-II) demonstrated that type 1A antiarrhythmics increase cardiac mortality in postmyocardial infarction patients. Although TCAs were not used in this study, the type 1A antiarrhythmic properties of TCAs suggest cautious use in depressed patients with ischemic heart disease.

Monoamine Oxidase (MAO) Inhibitors. The MAO inhibitors have been studied in depressed patients who are in their late 60s to early 70s.

However, limited information exists on the use of these agents in the old-old. MAO inhibitors are effective in patients with major depression as well as depressed patients with panic attacks. Low dosages of MAO inhibitors (e.g., phenelzine 30 to 45 mg daily or tranylcypromine 20 to 30 mg daily) should be used in older adults. Orthostatic hypotension is the most frequent important side effect of MAO inhibitors. This side effect is of concern in older adults because it may lead to falls and fractures, especially of the hip or the humerus. Other side effects include weight gain, insomnia, lack of energy, and daytime somnolence in phenelzine-treated patients and nervousness, insomnia, and excessive perspiration in tranylcypromine-treated patients. Peripheral neuropathy occurs in a small percentage of patients on MAO inhibitors and often responds to pyridoxine. Sympathomimetic amines, monoamine precursors, TCAs, SSRIs, venlafaxine, concomitant administration of two MAO inhibitors, and tyramine-rich food may cause a hypertensive crisis and should be avoided in patients on MAO inhibitors. The selegiline patch does not require diet restrictions at the lowest 6-mg dose and may be safer during diet violation than other MAOIs at higher doses. The advantage of transdermal administration is that intestinal monoamine oxidase is not inactivated to the same degree as oral administration, reducing the risk of a hypertensive crisis triggered by ingested monoamines.

Other Antidepressants. Bupropion, noradrenaline, and dopamine reuptake inhibitors have proven efficacy in younger adults as well as older adults with major depression based on a number of limited trials of the IR, SR, and XL formulations. Bupropion does not cause cognitive impairment or sedation, is safe in overdose, and lacks significant cardiotoxicity; it should be considered in older adult cardiac patients. Bupropion has few drug interactions, but it should not be prescribed in patients receiving MAO inhibitors. Bupropion may exacerbate pre-existing hypertension, and regular monitoring of blood pressure is required. Seizures have been reported in 0.4% of patients treated with bupropion, but the risk of seizures can be minimized by the use of sustained-release preparations, slow introduction of bupropion (100 mg daily), and restriction of the total daily dosage to 450 mg. Most older adults require a total daily dose of 300 mg.

Mirtazapine has been increasingly used in frail, depressed nursing home patients with anorexia. Mirtazapine antagonizes the α-2 presynaptic inhibitory receptor as well as the 5-HT2 and the 5-HT3 serotonin receptors. Mirtazapine is an antagonist of histaminergic receptors and has a mild antagonist effect on the α-1 adrenergic and muscarinic receptors. These effects account for the sedative, hypotensive, and anticholinergic side effects of mirtazapine. Mirtazapine is not an inhibitor of the P450 cytochrome. Side effects include somnolence, dizziness, increased appetite, weight gain, and hyponatremia. A mild increase in cholesterol and triglycerides has been observed in 15% and 6% of patients, respectively. The starting dosage of mirtazapine is 7.5 to 15 mg and should be administered at bedtime to reduce daytime sedation. Daily dosages from 30 to 45 mg are sufficient for the treatment of late-life depression.

Psychostimulants may be helpful as augmentation agents in the treatment of late-life depression. Dextroamphetamine and methylphenidate appear to improve apathy and anergy in medical patients or cognitively impaired patients. Dextroamphetamine or methylphenidate at dosages of 5 to 10 mg in the morning and at noon is sufficient. Psychostimulants have a rapid onset of action, minimal side effects, limited potential for tolerance, and low risk for addiction. In patients with Parkinson disease, pramipexole, a dopamine receptor D2/D3 agonist, has been shown to improve depressive symptoms, including in treatment-resistant depression.

Complementary and Alternative Medicines. There are a number of natural supplements with some evidence of efficacy in geriatric populations. For instance, omega-3 fatty acids, St. John’s wort, and *S*-adenosyl-*L*-methionine (SAMe) have shown some benefit for depression in a limited number of small studies in older adults. St. John’s wort must be used with caution in geriatric populations due to drug–drug interactions through the induction of cytochrome P450 3A4. In addition, there has been limited study of nonpharmacologic alternative therapies in older adults. Tai Chi, and mind-body exercise practice, have been found beneficial in augmenting escitalopram in late-life depression in a randomized control trial. Bright light treatment is another noninvasive treatment with some evidence of efficacy in geriatric depression. Evidence also suggests that exercise confers a modest benefit in late-life depression compared with usual care or nonactive control interventions. Adherence to the Mediterranean diet and tea drinking was also associated with having a beneficial effect on depressive symptoms in older adults in one study. Future study of these alternative approaches and their optimal delivery in older adults is needed.

Somatic Treatments

Electroconvulsive Therapy (ECT). ECT should be considered in both geriatric depression and mania because of its efficacy, rapid onset of action, and safety. In 2018, the FDA approved ECT for the treatment of severe major depressive episodes associated with MDD or BD in patients aged 13 years and older with treatment-resistant symptoms or who require a rapid response to treatment due to the severity of their psychiatric or medical condition. ECT may therefore be chosen in patients with severe mood syndromes unable to tolerate the long waiting imposed by the gradual introduction of antidepressants or mood stabilizers and the slow onset of drug action in older adults. In addition, ECT can be effective in treating neuropsychiatric comorbidities such as catatonia and parkinsonism, often present in depressed older adults.

ECT treatment characteristics must be chosen with an eye to maximizing treatment efficacy while minimizing cognitive side effects. An example of one such consideration is that ECT administration as often as three times weekly appears beneficial in producing a more rapid recovery than ECT given twice or once a week, yet less frequent ECT is associated with less cognitive impairment in the time period shortly after ECT. No differences in cognition have been found. However, 1 month after ECT completion between groups that had ECT thrice or twice weekly. ECT treatment variables to consider are electrode placement (right unilateral vs. bilateral), treatment frequency, stimulus intensity (energy), duration, pulse width, and pulse frequency. The effects of these choices are described elsewhere in this book. High-dose, ultra-brief, right unilateral ECT treatment has become a common choice due to reports of comparable efficacy and reduced cognitive side effects compared with bilateral ECT.

The mortality rate for all ages following a single ECT treatment is approximately 1 per 80,000 treatments, corresponding to 1 per 10,000 ECT-treated patients. This rate may be higher in older adults with severe medical conditions, but it appears to be lower than that for some medication treatments (TCAs), and there is some evidence that it may be compensated for by decreases in mortality following hospitalization (compared with depressed patients receiving other forms of treatment or no treatment at all). The majority of deaths (67%) related to ECT are due to cardiac complications and occur immediately after ECT or within a few hours of treatment. Compared to younger adults, older adults and patients with pre-existing cardiac disease are more prone to cardiovascular events occurring in the context of ECT, including ischemic syndromes, arrhythmias, decompensation of heart failure, and severe transient rise in blood pressure. With adequate medical evaluation, however, monitoring during and after ECT, and appropriate intervention, most cardiovascular events related to ECT have a benign outcome.

Geriatric patients require a longer time to recover their memory than younger adults, especially after bilateral ECT. Moreover, geriatric patients sometimes develop prolonged confusion after ECT. The increase in falls with ECT has been reported as up to 15% in older adults. Controlled studies, however, are needed in order to examine whether and to what extent ECT-related confusion is responsible for these falls. The practice of favoring ECT over antidepressants in medically ill depressed patients may have led to inflated reports of ECT-related morbidity. Anecdotal literature suggests that ECT was used uneventfully in patients over 100 years old and in older adults with aortic aneurysm, cardiac pacemaker, history of myocardial infarction or stroke, transient ischemic attacks, severe hypertension, arrhythmias, or need for anticoagulants.

Other Therapies

Vagal nerve stimulation (VNS) is an FDA-approved treatment for chronic and recurrent depression. Through an implanted generator, electrical signals are delivered to the left cervical vagus nerve, where it is postulated that stimulation of the afferent fibers leads to modulation of mood-related networks (e.g., amygdala and locus coeruleus). Although studies of efficacy in geriatric depression are lacking, VNS does not carry an increased risk of surgical complications and does not necessarily increase the risk for cardiac abnormalities in this population.

Transcranial magnetic stimulation (TMS) is also FDA-approved for the treatment of depression. In this noninvasive procedure, an electromagnetic field is generated that depolarizes neurons leading to modulation of cortical activity. This procedure is well tolerated with minimal cognitive side effects and has been shown to lead to a higher remission rate than sham TMS in studies. One study of repetitive transcranial magnetic stimulation (rTMS) in older adults with unipolar and bipolar depression found that it led to significant improvements in depression and anxiety from baseline to post-treatment in all age groups, with those aged 60 years and older showing statistically greater improvement than younger participants. Another preliminary study with TMS showed a significant reduction of depressive symptoms in poststroke depression, and rTMS has been found efficacious in vascular depression.

Treatment of Bereavement

For many bereaved older adults, symptoms of bereavement will resolve on their own over time. For this reason, prophylactic treatment is not recommended for the entire bereaved older adult population. Self-help groups, counseling, group therapy, and individual psychotherapy can be considered for individuals with an intense reaction to loss or a history of depression. A recent systematic review of treatment options for grieving older adults reported that viable treatment approaches are those incorporating behavioral activation and guidance on coping with issues related to secondary changes brought on by loss. Overall, grief interventions typically had small effects on grief-related symptoms in older adults.

Bereaved persons with depressive symptomatology that do not meet the criteria for major depression may benefit from brief focused psychotherapy. Interpersonal therapy lends itself well to this purpose since it addresses issues related to loss, depression, and change of roles.

Although antidepressant drugs can be effective in bereaved older adults with depressive symptoms, clear guidelines are not available. If major depression develops during the bereavement period, antidepressant treatment should be considered.

Complicated grief—a persistent complex bereavement disorder characterized by prolonged acute grief, feeling that life has no meaning, avoidance behaviors, and dysregulated emotions—has been identified in DSM-5 as a condition in need of further study. Complicated grief treatment (CGT) has been developed as a 16-week treatment with components of prolonged exposure, interpersonal therapy (IPT), and motivational interviewing and has been found to be more efficacious in symptom improvement than IPT in older adults. It appears that antidepressants may enhance the response to CGT but that antidepressants alone may be ineffective.

Treatment of Psychotic Depression

For the treatment of psychotic depression, the recommended treatment includes combinations of antidepressants (SSRIs or venlafaxine XR) and atypical neuroleptics (except clozapine), although there are limited clinical trials for older adults. In one multicenter 12-week double-blind, randomized controlled trial (STOP-PD) and a follow-up study (STOP-PD II), combination pharmacotherapy with olanzapine and sertraline was associated with higher remission rates during the trial period compared to olanzapine plus placebo in patients. ECT treatment is another option, especially if combination drug treatment fails or is poorly tolerated.

Treatment of Depression in Dementia

In older adults with major depression and mild-to-moderate dementia, the historical treatment of choice was a combination of an antidepressant with psychosocial intervention, although there are limited data to support the use of antidepressant medication. Indeed, in a large, well-powered study, not only were antidepressants (i.e., sertraline and mirtazapine) found ineffective in depressed demented patients but they also were associated with more adverse effects than placebo. A review article in 2017 found that evidence does not support the use of pharmacologic treatment of depression in patients with dementia. Further, a more recent meta-analysis in 2021 found that nonpharmacologic interventions were more efficacious than drug interventions in reducing depressive symptoms in older adults with depression symptoms and comorbid dementia without an MDD diagnosis. These findings suggest that psychosocial interventions such as PATH and caregiver-focused treatment may be better first-line treatment options for this population.

Treatment-Resistant Depression

Change in antidepressant strategy should take into consideration that depressed older patients take a longer time to respond to antidepressants than younger adults. If a patient has had little or no response to an adequate antidepressant trial, switching to a different antidepressant rather than using an augmenting agent is an appropriate first step. If a patient has had little or no response to an initial SSRI, the patient may be switched to an SNRI or bupropion (SR or XL). If a patient has had little or no response to venlafaxine, the patient may be switched to an SSRI. If the patient has had little or no response to a TCA or bupropion, switching to an SSRI or SNRI is appropriate. Although studies on older adults are lacking, research on younger patients suggests that switching from one SSRI to another is an effective strategy.

Augmentation therapies are recommended in depressed older adults with partial response to the initial antidepressant, although there is limited agreement on the choice of augmentation therapies. Bupropion, lithium, or nortriptyline may be added if the initial antidepressant is an SSRI. If the initial antidepressant is a TCA, lithium or an SSRI may be added. If the patient was on bupropion, an SSRI or lithium may be added. An additional augmentation strategy in older adults is the combination of an SSRI or SNRI plus mirtazapine. In a study of middle-aged adults (mean age in mid-40s), mirtazapine (30 mg) plus fluoxetine (20 mg), venlafaxine (225 mg), or bupropion (150 mg) achieved greater remission rates (46% to 58%) after 6 weeks than fluoxetine monotherapy (25%). Among those who responded, double-blind discontinuation of one agent led to relapse in 40% of the cases.

The addition of methylphenidate may be helpful to accelerate treatment response to an SSRI antidepressant and may improve mood and well-being and lead to a higher remission rate compared to either drug alone.

Treatment with an atypical neuroleptic, alone or in combination with other medications, is sometimes effective for treatment-resistant depression. The augmentation of antidepressants with aripiprazole has been found to be effective in older adults. In one large RCT, aripiprazole augmentation (dose range 2 to 15 mg daily, target dose 10 mg daily) led to a higher remission rate compared to placebo, with remission rates sustained at a 12-week follow-up. Adverse effects reported included increased dream activity, weight gain, tremor, and extrapyramidal symptoms, including akathisia and parkinsonism. Quetiapine XR has been found effective as an augmenting agent for depression in younger adults, with more limited data available for its efficacy in older adults. In one placebo-controlled RCT of quetiapine in older patients with MDD, quetiapine monotherapy was associated with higher remission rates of depression than placebo and was well tolerated, with adverse events including dizziness, headache, and somnolence.

Brexpiprazole is a novel antipsychotic that is structurally similar to aripiprazole and is a partial agonist of D2 and 5-HT1A receptors and a strong antagonist of 5HT2A and norepinephrine alpha1B and alpha 2 C receptors. Brexipiprazole is metabolized by cytochrome P450 CYP3A4 and CYP2D6, so the dose should be reduced when prescribed with inhibitors of these enzymes. A lower starting dosage is recommended in older adults (0.5 mg/day) with a maximum of 3 mg/day. While efficacious in studies on younger adults with depression, its efficacy has not been studied specifically in older adults. It should be noted that the use of antipsychotic agents (both atypical and typical) in older individuals with dementia has been associated with a greater risk of mortality and increased risk of cerebrovascular events. It is unclear if this risk extends to older adult patients with mood disorders who do not have dementia. When medication treatment fails, ECT may offer the best chance for response and remission.

Subcutaneous ketamine, an intervention for depression of emerging interest in the general adult population, was well tolerated (at doses of 0.1 to 0.5 mg/kg) and found to lead to remission of symptoms in a recent randomized control trial (n = 16) for treatment-resistant depression in older adults. Repeated treatments (up to five treatments administered at intervals of 1 week or greater apart) resulted in a higher likelihood of remission or a longer time to relapse.

In contrast, in one study, esketamine (the S-enantiomer of ketamine racemate, which has shown efficacy in younger adults with treatment-resistant depression) in combination with an antidepressant did not differ significantly from antidepressant plus placebo nasal spray when studied in older adults with treatment-resistant depression. The adverse event profile was similar to that seen in esketamine studies of younger patients. Future work is needed to evaluate rapidly-acting antidepressants in older adult populations.

Continuation and Maintenance Treatment

The intent of continuation treatment (6 months) is to prevent relapse of the acute episode, while maintenance treatment is offered after the first 6 months of remission in order to prevent a new depressive episode. Continuation and maintenance treatment should use the same dosages of antidepressants as those of acute treatment. Continuation treatment should be offered to all patients with late-life depression. Patients at high risk for recurrence should have maintenance treatment. The strongest predictor of recurrence is a history of three or more lifetime episodes of depression. For such patients, experts recommend treatment for 3 years or longer. Disagreement exists concerning the optimal maintenance treatment duration in older adults with fewer episodes. Other predictors of depression recurrence are high severity of the initial episode and persisting anxiety, and in mixed-age populations, late age of illness onset is a strong predictor of recurrence. In patients with psychotic depression who achieve remission following treatment with an antidepressant and an antipsychotic, continuation of the antipsychotic drug is recommended, typically for 6 months, but some experts recommend shorter or longer treatment. For depressed patients who failed antidepressants but responded to ECT, antidepressants (other than those to which the patient previously failed to respond) plus a mood stabilizer should be used for continuation/maintenance treatment. Continuation/maintenance ECT is another option, with the PRIDE study showing that four additional treatments after remission was beneficial in sustaining mood improvement for most patients.

Adherence. Approximately 70% of patients fail to take 25% to 50% of their medications. Older patients and their families may often have difficulty accepting the diagnosis of depression and understanding the need for long-term treatment. Further, cognitive impairment, physical disability, and financial hardship faced by many older adults may present a challenge to treatment adherence. Concurrent medical illnesses may interfere with an antidepressant response or the attainment of adequate dosages. Alcohol and other substance use disorders may undercut pharmacotherapy. Difficulties accessing health care may hinder the ability of older adults, especially functionally dependent older adults, to obtain adequate treatment.

Care Delivery

Several interventions have been developed to improve the treatment of geriatric depression in the primary care setting. The Improving Mood-Promoting Access to Collaborative Treatment (IMPACT) study focused on older adults and offered access for up to 12 months to a depression care manager who was supervised by a psychiatrist and an expert primary care physician. The depression care manager offers education, care management, and support of antidepressant management by the primary care physician or PST (brief structured psychotherapy). Depressed primary care older adult patients receiving the IMPACT intervention were more likely to receive antidepressants or psychotherapy than usual care patients. They had lower depressive symptomatology and higher overall function, response and remission rates, quality of life, and satisfaction with treatment. These benefits were sustained 12 months after the end of the intervention. A recent systematic review showed that the collaborative care model for depressive disorders among older adults is feasible and beneficial in diverse settings.

Positive outcomes have also been reported by the Prevention of Suicide in Primary Care Elderly: Collaborative Trial (PROSPECT). The lynchpin of the PROSPECT intervention was the use of care managers, who were trained social workers, nurses, and psychologists who worked alongside primary care physicians. They followed the patients and monitored their psychiatric symptoms and signs, treatment adherence, and side effects, and provided appropriately timed and targeted recommendations to physicians according to an algorithm. Patients treated in practices offering the PROSPECT intervention had a higher likelihood of receiving antidepressant treatment, a greater decline in the rate of suicidal ideation, and a more favorable course of depression over a period of 24 months. The beneficial effects of the PROSPECT intervention were most pronounced in primary care patients with major depression rather than minor depression.

The concept of an integrated care model has been extended to depressed patients with chronic medical illnesses. The personalized adherence intervention for depression and COPD (PID-C) is an integrated approach implemented by care managers to improve adherence to treatment for both COPD and depression with a focus on helping patients to overcome barriers to treatment and offering interventions that support patients in using treatment prescribed by their own physicians. PID-C reduced depressive symptoms and dyspnea-related disability more than usual care.

Clinical case management has been shown to reduce the severity of depression and improve disability in low-income, disabled older adults suffering from major depression.

Overview of Geriatric Bipolar Disorder Evaluation and Treatment

The goals of geriatric BD include symptom amelioration (i.e., resolution of mania, remission of depression, and maintenance of euthymia) as well as improvements in functional status and prevention of recurrences. The evaluation of geriatric BD should include a comprehensive assessment exploring comorbidity (psychiatric, medical, and neurologic), functional status, and psychosocial factors. Further, substance use disorders are common in geriatric BD patients and should be routinely assessed and treatment offered in the evaluation period. Medications should be evaluated as well, given that medications such as corticosteroids and dopaminergic agents such as stimulants and antiparkinsonian drugs can precipitate mania.

Older adults with BD often have limited, slow, or fragile treatment responses. Guidelines for dosage and duration of treatment with antimanic agents cannot be extrapolated directly from younger adult studies because aging, comorbid neurologic diseases, high medical burden, and poor social support may directly or indirectly, through poor adherence, interfere with the treatment response of late-life BD and increase vulnerability to side effects.

Pharmacotherapy for Geriatric Bipolar Disorder

Lithium is the best-investigated drug in geriatric mania. Data shows that in older adults with bipolar depression or treatment-resistant MDD, lithium users experienced less depressive symptomatology, lower clinical severity, and less frequent psychiatric admissions compared to those not prescribed lithium. Lithium may be less effective in mania with comorbid neurologic and medical disorders than in uncomplicated mania. Age-associated decline in renal clearance in older adults leads to high lithium plasma levels even when low dosages are used. Lithium should be introduced slowly, starting with daily dosages of 150 mg. About one-half of the dosage required for young adults is sufficient for older adults. The half-life of lithium is about 24 to 36 hours in patients who are in their 70s. Therefore, steady-state pharmacokinetics is anticipated 5 or more days after the stabilization of the daily dosage. In some older adults, lithium plasma levels of 0.3 to 0.6 mEq/L were found to be effective. Controlled studies, however, are needed to establish the efficacy of low lithium plasma levels in older adults. The onset of action of lithium is slow and may require several days or weeks.

The side effects of lithium can be pronounced in geriatric patients. Older adult patients often develop a fine tremor or myoclonus at lithium plasma levels considered nontoxic in young adults. Lithium can produce parkinsonian symptoms or worsen pre-existing Parkinson disease.

Delirium can develop at plasma levels below the therapeutic range. Neuroleptics increase the risk of delirium in older adult patients on lithium. Lithium may worsen cognitive function, especially in individuals with dementia, though some research suggests that long-term lithium treatment might have neuroprotective effects. Longer lithium exposure has been associated with greater white matter integrity in older adults with bipolar depression. Lithium inhibits glycogen synthase kinase 3, a key enzyme in the metabolism of amyloid precursor protein and in the phosphorylation of tau protein, which may reduce the risk of AD. A case-control study of older adults with BD reported that patients treated continuously with lithium had a significantly lower prevalence rate of dementia than patients who did not receive continuous lithium treatment. Sinoatrial block may be caused by lithium in predisposed patients; digitalis and β-blockers increase the risk for sinoatrial block. Salt depletion, thiazide diuretics, and nonsteroidal anti-inflammatory drugs (NSAIDs) may raise the lithium plasma levels and result in toxicity. Chronic kidney disease appears to be common in older adults treated with lithium.

The antiepileptics sodium divalproex, carbamazepine, and lamotrigine are effective for preventative treatment in BD. Anticonvulsants are effective in both lithium responders and lithium nonresponders. The GERI-BD trial demonstrated that lithium and divalproex were found to be equally well tolerated and efficacious, with lithium associated with a greater reduction in manic symptoms over 9 weeks. However, anticonvulsants may be more effective than lithium in patients with rapidly cycling mania or dysphoric mania. Although there is limited research on anticonvulsants in older adult populations with BD, anticonvulsants should be considered, especially in patients at high risk for lithium toxicity or history of poor response to lithium. Drowsiness and sedation are frequent dose-dependent side effects of valproate and carbamazepine. Many patients are habituated to sedation. Therefore, the slow introduction of valproate or carbamazepine may reduce this problem. In older adults, the principal side effects of valproate are gastrointestinal distress and ataxia. Frequent side effects of carbamazepine are ataxia and confusion. Carbamazepine can also cause hyponatremia. Approximately 0.4% of patients receiving valproate and 2.0% of patients receiving carbamazepine develop leukopenia with a white blood cell count below 4,000/mm^3. Valproate may cause thrombocytopenia in older patients much more frequently than in younger adults. Carbamazepine-treated patients have a reduction in their white blood cell count during the first 2 weeks. For this reason, periodic monitoring of the white count is indicated, especially during the early phases of treatment with anticonvulsants.

Lamotrigine is generally well tolerated and does not cause weight gain, sedation, sexual side effects, or cognitive impairment. There is some preliminary evidence to suggest that lamotrigine might be associated with improved cognitive status in adults with bipolar depression. It helps mainly for depression, which for bipolar patients is often a greater concern than mania or hypomania. A mild rash is common, and the risk of serious but not life-threatening rash is approximately 1 in 250. The risk of Stevens–Johnson syndrome is approximately 1 in 1,000, or perhaps lower, with cautious uptitration as recommended to reduce the rash risk. Lamotrigine should be stopped immediately if a rash develops. Simultaneous treatment with valproate can more than double lamotrigine's half-life, so such patients should start at no more than 25 mg every other day with dose increases no more than once every 2 weeks in the first weeks. Effective doses are usually 100 to 200 mg/day. Thus, a drawback of lamotrigine is that it can take weeks to reach a therapeutic dose.

Agitated manic older adult patients may be treated with low dosages of lorazepam. Lorazepam is favored over other benzodiazepines because it undergoes only phase II metabolism (glucuronidation), a process minimally affected by aging. Effective alternatives to lorazepam are atypical and high-potency neuroleptics in oral (e.g., risperidone 1 to 2 mg, olanzapine 5 to 10 mg daily) or injectable (e.g., haloperidol 0.5 to 5 mg daily) form. These drugs should be used early in the treatment of acute mania and should be withdrawn when the agitation is controlled, and therapeutic plasma levels of a mood stabilizer are established.

In addition to agitated mania, antipsychotic agents may also be useful in the treatment of mania as well as bipolar depression. For instance, there is evidence that atypical antipsychotics, including olanzapine, risperidone, and quetiapine, may be effective in the treatment of geriatric mania, while quetiapine and lurasidone may have some efficacy in bipolar depression. Lurasidone has been associated with sustained improvement in depression symptoms in older adults with bipolar depression, with common adverse effects including headache, nasopharyngitis, and insomnia. Adjunctive lurasidone therapy (used in combination with lithium or valproate) led to higher rates of tremor and akathisia than monotherapy and did not differ from placebo when used as adjunctive therapy in one study.

Paliperidone and cariprazine have also been found effective in the treatment of acute mania for the general adult population alone or in combination with a mood stabilizer, though data on these agents is limited in the geriatric population.

Asenapine has been studied in a small number of open-label trials with manic patients aged 60 years and older, with results showing efficacy in reducing manic and hypomanic symptoms. Side effects of asenapine observed in these trials included QTc prolongation, restlessness, gastrointestinal discomfort, and extrapyramidal symptoms. Clearance of asenapine appears to decline with age, though the safety profile does not seem to differ significantly between older and younger patients.

Aripiprazole may be efficacious in treating symptoms of mania when used in conjunction with a mood stabilizer for older adults who do not respond well to other first-line treatments, though current data specific to older adults with this medication are limited. An open-label trial of aripiprazole adjunctive to traditional mood stabilizers in 20 older adults found significant improvement in mania symptoms over 12 weeks. The increased risk of weight gain, hyperlipidemia, and diabetes associated with at least some atypical neuroleptics raises the risk of cardiovascular disease.

Antidepressant use in BD is controversial due to concern about antidepressant-induced mania or mixed states, although studies of such complications are limited in older populations.

Other Treatments for Geriatric Bipolar Disorder

ECT is efficacious in late-life bipolar depression. In fact, a recent mixed-age observational study found that higher age was associated with a better prognosis for response to ECT in patients with BD. Limited data support the efficacy of ECT in the treatment of mania in older adults. TMS has been shown to be a treatment option for geriatric bipolar depression, with one study showing that rTMS leads to greater improvement in unipolar and bipolar depression and anxiety symptom scores in patients older than 60 and appeared to be well-tolerated. Cognitive-behavioral therapy, family-focused therapy, and interpersonal social rhythm therapy have been shown to be effective as adjuncts to medications in shortening the time to remission from a depressive episode. Psychosocial interventions targeting adherence to treatment, such as customized adherence enhancement (CAE) therapy, have been found effective in geriatric BD in improving treatment adherence, reducing mood-related symptoms and signs and global psychopathology, and improving functioning.

The Helping Older People Experience Success (HOPES) program is a 2-year program for older adults with severe mental illness that aims to improve self-efficacy, increase community engagement, and decrease reliance on healthcare services by offering weekly social skills training, health care management by a nurse, and bimonthly community outings to practice skills. The HOPES intervention was shown in an RCT to improve patient outcomes at the conclusion of the 2-year program and at 3-year follow-up.

An enhanced clinical intervention (ECI) intensive, including educational modules and illness management components (such as the development of crisis plans; and psychoeducation about symptoms and side effects), was shown to help older adults with BD improve global functioning and quality of life.

A 12-week group intervention called “Medication Adherence Skills Training of Middle-Aged and Elderly Adults with Bipolar Disorder” (MAST-BD) comprised of four modules—educational, motivational, medication management, and symptom management—was developed to target medication nonadherence. This intervention was shown to improve adherence, as well as depressive symptoms and quality of life.

FUTURE DIRECTIONS

Although medical and neurologic comorbidity of geriatric mood disorders may complicate care, they have served as a window to mechanisms contributing to geriatric mood disorders. The clinical association of depressive symptoms with impairment in specific cognitive functions has been a central stimulus to research on geriatric depression. Studies of the integrity of white matter in distinct brain regions have begun to identify abnormalities leading to disconnection states contributing to late-life depression and executive dysfunction. Advances in neuroimaging and electrophysiology make possible the identification of structural, functional, and processing abnormalities predisposing depressed older adults to poor response to antidepressants or to mood stabilizers. Genetics research identified some polymorphisms that increase the risk of resistance to some antidepressants. Neuroimaging studies are underway to identify the mechanisms by which genetic factors increase resistance to antidepressants and also to investigate endophenotypes of geriatric mood disorders and biomarkers that may predict response to specific interventions.

With the recently advanced hypotheses of geriatric depression, there is the promise to guide novel treatment development. For example, studies of the “DED syndrome” demonstrated that executive dysfunction and structural and functional abnormalities of the cognitive control system are associated with poor response to antidepressant drugs. However, appropriately modified PST reduced both depression and disability more than supportive therapy in depressed, nondemented patients with major depression and executive dysfunction. PATH was developed in order to address the clinical needs of depressed older adults with cognitive impairment up to moderate dementia. A neuroplasticity-based computerized cognitive remediation-geriatric depression treatment (nCCR-GD) was developed to target executive dysfunction in late-life depression. The underlying assumption was that remediation of these deficits may improve the functional brain networks shared by executive dysfunction and depression. Two studies have now shown that this software improves both depressive symptoms and executive function in late-life major depression. Because it is standardized and self-administered, nCCR is not subject to therapists’ skill drift and may be helpful in minimizing barriers to access to care due to its low cost and administration in a patient’s home. Future work may investigate whether nCCR alone or in combination with TMS or pharmacotherapy may ameliorate specific cognitive dysfunction and improve outcomes in geriatric depression and BD.

While the inflammation hypothesis of late-life depression has generated interest in the role of anti-inflammatory agents in the treatment of late-life depression, many questions remain. The position of inflammation in the chain of biologic events leading to late-life depression needs to be clarified. Assuming that the brain’s inflammatory responses have regional specificity, it is unclear whether the localization of these responses influences the clinical profile of late-life depression and its course. In addition, the role of peripheral inflammatory indices in determining the choice of antidepressants needs to be further understood. A review of meta-analyses suggests that in the subgroup of individuals with MDD and evidence of increased inflammatory biomarkers, NSAIDs, omega-3 fatty acids, and cytokine antagonists may have antidepressant properties.

Minocycline, a microglia inhibitor, has been shown to exert an antidepressant effect in a meta-analysis of clinical trials, though it was noted that studies had a small sample size and only a small number were RCTs. A larger RCT of minocycline and celecoxib as adjunctive treatments for bipolar depression did not show an effect compared to a placebo. Etanercept, an agent that reduces TNF-α–mediated cellular response, reduced symptoms of depression in a placebo-controlled study of patients with psoriasis. The cyclooxygenase-2 inhibitor celecoxib has been found to augment the efficacy of reboxetine and fluoxetine in patients with MDD. In treatment-resistant major depression, baseline high concentrations of high-sensitivity CRP predicted a favorable response to the TNF antagonist, infliximab. Adequately powered randomized controlled trials need to clarify which anti-inflammatory agents have antidepressant action and for whom.

In addition, future work is needed to ascertain whether antihypertensive, antihypercholesterolemic, and antiplatelet agents and diabetes control can improve the outcomes of vascular depression. Longitudinal assessment of depressive symptomatology during antiplatelet secondary prevention trials can evaluate the efficacy of these agents in improving the course of vascular depression. Modulators of the RAS improve cerebral perfusion and may improve cognition. Angiotensin receptor blocker (ARB) use has been associated with improvement in depression, anxiety, and quality of life in one study. Direct studies are needed to clarify the role of ARBs in the treatment of vascular depression. Few studies suggest that calcium channel blockers may reduce depressive symptoms, and in an early study of vascular depression treated with a variety of antidepressants, the addition of nimodipine improved depressive symptoms more than the addition of an inactive comparator. Finally, anti-inflammatory agents may have a role in vascular depression, but studies are lacking.

Advances in biology and psychotherapeutic interventions, along with clinical and psychosocial studies of the comorbidity and social complexity of geriatric mood disorders, are generating knowledge on which to build comprehensive models of care. For programs such as PID-C to become part of the mainstay of clinical treatment of late-life depression, their efficacy, implementation potential, and fiscal feasibility need to be demonstrated. The IMPACT and PROSPECT interventions are concrete examples of treatment models that address the clinical complexity of geriatric depression and the barriers imposed by the primary care setting. Reimbursement constraints remain a barrier to the implementation of such programs. Studies are underway that build on the principles of these models and focus on existing services supported by the available reimbursement system. Such studies offer focused training to clinicians and interventions at the organizational level to increase functional collaborative relationships between primary care practices and already funded independent mental health clinics. Further research is needed to find efficient interventions and to weigh costs versus benefits in various older adult populations with distinct needs.

FURTHER READINGS

*Alexopoulos GS. Mechanisms and treatment of late-life depression. *Transl Psychiatry*. 2019;9(1):188.

Bartels C, Wagner M, Wolfsgruber S, Ehrenreich H, Schneider A, Alzheimer's Disease Neuroimaging Initiative. Impact of SSRI therapy on risk of conversion from mild cognitive impairment to Alzheimer's dementia in individuals with previous depression. *Am J Psychiatry*. 2018;175(3):232–241.

Beurel E, Toups M, Nemeroff CB. The bidirectional relationship of depression and inflammation: double trouble. *Neuron*. 2020;107(2):234–256.

Bridle C, Spanjers K, Patel S, Atherton NM, Lamb SE. Effect of exercise on depression severity in older people: systematic review and meta-analysis of randomised controlled trials. *Br J Psychiatry*. 2012;201(3):180–185.

Dham P, Colman S, Saperson K, et al. Collaborative care for psychiatric disorders in older adults: a systematic review. *Can J Psychiatry*. 2017;62(11):761–771.

Flint AJ, Meyers BS, Rothschild AJ. Effect of continuing olanzapine vs placebo on relapse among patients with psychotic depression in remission: the STOP-PD II randomized clinical trial. *JAMA*. 2019;322(7):622–631.

Gallagher D, Kiss A, Lanctot K, Herrmann N. Depression and risk of Alzheimer dementia: a longitudinal analysis to determine predictors of increased risk among older adults with depression. *Am J Geriatr Psychiatry*. 2018;26(8):819–827.

Haigh EAP, Bogucki OE, Sigmon ST, Blazer DG. Depression among older adults: a 20-year update on five common myths and misconceptions. *Am J Geriatr Psychiatry*. 2018;26(1):107–122.

Howard DM, Adams MJ, Clarke TK, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343–352.

Ismail Z, Elbayoumi H, Fischer CE, et al. Prevalence of depression in patients with mild cognitive impairment: a systematic review and meta-analysis. *JAMA Psychiatry*. 2017;74(1):58–67.

Kiosses DN, Ravdin LD, Gross JJ, Raue P, Kotbi N, Alexopoulos GS. Problem adaptation therapy for older adults with major depression and cognitive impairment: a randomized clinical trial. *JAMA Psychiatry*. 2015;72(1):22–30.

*Kok RM, Reynolds CF 3rd. Management of depression in older adults: a review. *JAMA*. 2017;317(20):2114–2122.

Mackin RS, Insel PS, Landau S, et al. Alzheimer's disease neuroimaging initiative and the ADNI depression project. late-life depression is associated with reduced cortical amyloid burden: findings from the Alzheimer's disease neuroimaging initiative depression project. *Biol Psychiatry*. 2021;89(8):757–765.

Masana MF, Haro JM, Mariolis A, et al. Mediterranean diet and depression among older individuals: the multinational MEDIS study. *Exp Gerontol*. 2018;110:67–72.

McIntosh AM, Sullivan PF, Lewis CM. Uncovering the genetic architecture of major depression. *Neuron*. 2019;102(1):91–103.

Mura G, Carta MG. Physical activity in depressed elderly: a systematic review. *Clin Pract Epidemiol Ment Health*. 2013;9:125–135.

Ochs-Ross R, Daly EJ, Zhang Y, et al. Efficacy and safety of esketamine nasal spray plus an oral antidepressant in elderly patients with treatment-resistant depression-TRANSFORM-3. *Am J Geriatr Psychiatry*. 2020;28(2):121–141.

Okereke OI, Singh A. The role of vitamin D in the prevention of late-life depression. *J Affect Disord*. 2016;198:1–14.

Patel K, Abdool PS, Rajji TK, Mulsant BH. Pharmacotherapy of major depression in late life: what is the role of new agents? *Expert Opin Pharmacother*. 2017;18(6):599–609.

*Patrick RE, Heintz HL, Skurla MD, et al. A comprehensive review of the evaluation, diagnosis, and treatment of older adult bipolar disorder. *Curr Treat Options Psychiatry*. 2020;7:508–530.

Penner-Goeke S, Binder EB. Epigenetics and depression. *Dialogues Clin Neurosci*. 2019;21(4):397–405.

Sachs GS, Nierenberg AA, Calabrese JR, et al. Effectiveness of adjunctive antidepressant treatment for bipolar depression. *N Engl J Med*. 2007;356(17):1711–1722.

Schut H, Stroebe MS, van den Bout J, Terheggen M. The efficacy of bereavement interventions: determining who benefits. In: Stroebe MS, Hansson RO, Stroebe W, Schut H, eds. *Handbook of Bereavement Research: Consequences, Coping, and Carei*. American Psychological Association; 2001;705–737.

Sonsin-Diaz N, Gottesman RF, Fracica E, et al. Chronic systemic inflammation is associated with symptoms of late-life depression: the ARIC study. *Am J Geriatr Psychiatry*. 2020;28(1):87–98.

Taylor WD, Aizenstein HJ, Alexopoulos GS. The vascular depression hypothesis: mechanisms linking vascular disease with depression. *Mol Psychiatry*. 2013;18(9): 963–974.

Watt JA, Goodarzi Z, Veroniki AA, et al. Comparative efficacy of interventions for reducing symptoms of depression in people with dementia: systematic review and network meta-analysis. *BMJ*. 2021;372:n532.

Wei J, Ying M, Xie L, et al. Late-life depression and cognitive function among older adults in the U.S.: the National Health and Nutrition Examination Survey, 2011–2014. *J Psychiatr Res*. 2019;111:30–35.

*Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord*. 2018;20(2):97–170.

*Young RC, Mulsant BH, Sajatovic M, et al. GERI-BD: a randomized double-blind controlled trial of lithium and divalproex in the treatment of mania in older patients with bipolar disorder. *Am J Psychiatry*. 2017;174(11):1086–1093.

Zhang Y, Chen Y, Ma L. Depression and cardiovascular disease in elderly: current understanding. *J Clin Neurosci*. 2018;47:1–5.

55.3f Alzheimer Disease and Other Neurocognitive Disorders

GARY W. SMALL, M.D., AND STEPHEN T. CHEN, M.D.

In part because of medical technologic advances, people in industrialized societies are living longer, and the number of older persons is growing rapidly. Because age is a major risk factor for developing Alzheimer disease and other neurocognitive disorders (NCDs), the number of persons suffering from these illnesses has increased along with the aging of our society in recent years. With the exception of rare genetic mutations, the cause of Alzheimer disease is not known. Scientists are actively searching for disease-modifying treatments and possible cures, and improved knowledge of potential risk and protective factors and biomarkers may contribute to preventive strategies. Along with such approaches, accurate early diagnosis, symptomatic treatments, and environmental interventions are essential to minimize the suffering of patients and family members and to optimize the quality of life for patients.

HISTORY

In 1906, the German psychiatrist and neurologist Alois Alzheimer first described a middle-aged patient who had suffered from a progressive dementia that affected language, memory, and behavior. After the patient's death at age 55 years, Alzheimer applied new staining techniques to the patient's brain tissue and demonstrated the presence of what is now termed neurofibrillary tangles and neuritic plaques in the neocortex and other brain regions. For years, Alzheimer disease was considered presenile dementia because it was not known that plaques and tangles were present in most older adults with dementia. However, in the late 1960s, when the degree of dementia was shown to correlate with the number of neuritic plaques in neocortical association areas in cases of senile dementia, late-onset Alzheimer disease was finally recognized.

Over the last several decades, the recognition that cognitive decline may be pathologic rather than normal in aging has helped to drive scientific advances but has also served to stigmatize the condition. Many older people have been reluctant to seek help for memory symptoms because of fear that they may be suffering from Alzheimer disease. Thus, symptomatic treatments are often begun after patients are already suffering from moderate degrees of dementia. However, if treatments that delay the onset of Alzheimer disease can be developed, there may be less emphasis on the treatment of the disease state and greater attention to prevention and deceleration strategies, reducing the stigma and increasing access to care.

COMPARATIVE NOSOLOGY

A definitive diagnosis of Alzheimer disease can be made only by histopathologic examination of brain tissue, generally postmortem. Therefore, most cases are diagnosed based on clinical information and, in some settings, additional diagnostic studies.

Current diagnostic criteria for dementia and Alzheimer disease are provided by two important references (Table 55.3f–1). In the fifth edition of the *Diagnostic and Statistical Manual* (DSM-5) published in 2013, the syndrome of dementia from the revised fourth edition (DSM-IV-TR) is subsumed under the new category “Major Neurocognitive Disorders.” A major NCD requires a significant decline in at least one cognitive domain and daily functioning based on subjective concern and clinical assessment. DSM-5 provides additional diagnostic criteria for specific NCD subtypes, which include vascular, frontotemporal, traumatic brain injury, and Lewy bodies.

In 2011, a work group formed by the National Institute on Aging and the Alzheimer Association (NIA-AA) published diagnostic guidelines for “probable” and “possible” Alzheimer disease. Criteria for both are similar to those for a major NCD but require cognitive impairment based on history and objective cognitive assessment and a minimum of two cognitive or behavioral domains affected. Probable Alzheimer disease further specifies that cognitive impairment is insidious in onset, progressive, and prominent in memory, language, visuospatial, or executive functioning. Possible Alzheimer disease should be considered when the course is atypical or cannot be reliably documented, and there is evidence of other etiologies, for example, cerebrovascular disease or Lewy body dementia.

Both DSM-5 and the NIA-AA group recognize milder levels of cognitive impairment or a predementia stage of Alzheimer disease. In DSM-5, “Mild Neurocognitive Disorders,” which feature “modest” cognitive decline and no functional impairment, was added as a new diagnostic category to encourage earlier diagnosis and intervention, which may more effectively modify the course of illness than that in the later stages. The NIA-AA guidelines refer to the predementia stage as “mild cognitive impairment (MCI).” Core clinical criteria for MCI include concern for cognitive changes, impairment in one or more cognitive domains, preferably measured by serial objective testing, preserved independence in daily activities, and no evidence of dementia and other etiologies of cognitive impairment. Not all patients with MCI progress to dementia, though somewhere between 6 and 25% convert to dementia each year. While either mild NCD or MCI can be diagnosed with only clinical data, the DSM-5 and the NIA-AA guidelines both refer to research biomarkers for Alzheimer disease that may aid diagnosis, particularly in the earlier stages of the disease. Such biomarkers include structural neuroimaging with magnetic resonance imaging (MRI), molecular neuroimaging with positron emission tomography (PET), cerebrospinal fluid (CSF) analysis, and genetic testing. Biomarker research and its potential impact on the nosology of Alzheimer disease are discussed below.

Table 55.3f–1.
Comparison of Diagnostic Criteria for Mild and Major Neurocognitive Disorders in Fifth Edition of the Diagnostic and Statistical Manual (DSM-5), Mild Cognitive Impairment, and National Institute on Aging-Alzheimer’s Association Criteria for Probable Alzheimer Disease Dementia

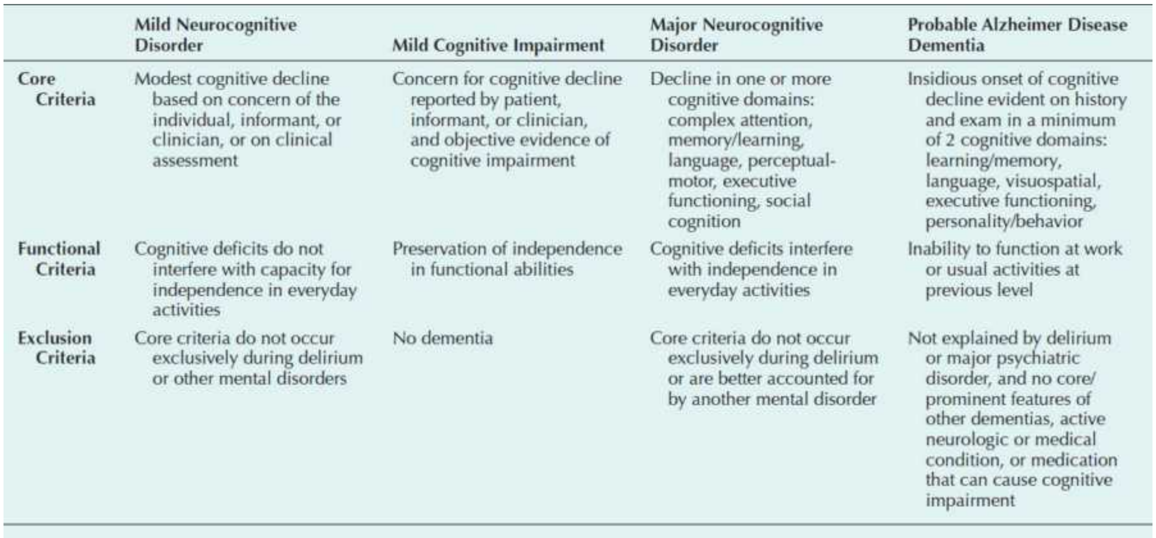

	Mild Neurocognitive Disorder	Mild Cognitive Impairment	Major Neurocognitive Disorder	Probable Alzheimer Disease Dementia
Core Criteria	Modest cognitive decline based on concern of the individual, informant, or clinician, or on clinical assessment	Concern for cognitive decline reported by patient, informant, or clinician, and objective evidence of cognitive impairment	Decline in one or more cognitive domains: complex attention, memory/learning, language, perceptual-motor, executive functioning, social cognition	Insidious onset of cognitive decline evident on history and exam in a minimum of 2 cognitive domains: learning/memory, language, visuospatial, executive functioning, personality/behavior
Functional Criteria	Cognitive deficits do not interfere with capacity for independence in everyday activities	Preservation of independence in functional abilities	Cognitive deficits interfere with independence in everyday activities	Inability to function at work or usual activities at previous level
Exclusion Criteria	Core criteria do not occur exclusively during delirium or other mental disorders	No dementia	Core criteria do not occur exclusively during delirium or are better accounted for by another mental disorder	Not explained by delirium or major psychiatric disorder, and no core/prominent features of other dementias, active neurologic or medical condition, or medication that can cause cognitive impairment

EPIDEMIOLOGY

NCDs are the most common psychiatric disorders in older persons and a growing public health problem in the United States. According to data compiled by the Alzheimer's Association in 2021, Alzheimer disease accounts for 60% to 80% of dementia cases and afflicts 6.2 million people in the United States or more than 1 in 9 people over the age of 65. As the baby boom generation (born between 1946 and 1964) ages, this number is expected to more than double by 2050 to 12.7 million (Fig. 55.3f–1). Alzheimer disease is the fifth leading cause of death as well as the leading cause of disability and poor health among older adults. Individuals with Alzheimer disease can live a relatively long time, on average 4 to 8 years after diagnosis, and require significant caregiving for much of this time. Alzheimer disease and other NCDs cost society over $355 billion in direct health care and long-term care costs and $257 billion in unpaid caregiving annually. These costs do not account for the additional health and financial burdens of dementia caregivers, who suffer more stress, depression, chronic health conditions, and loss of employment than other caregivers and adults. Direct costs are projected to exceed $1.1 trillion in 2050 if current trajectories hold. Recent research is focusing on novel treatments that effectively delay the onset of disease. Treatment that delays onset by 1 to 5 years is projected to reduce direct health costs by 14% to 39%.

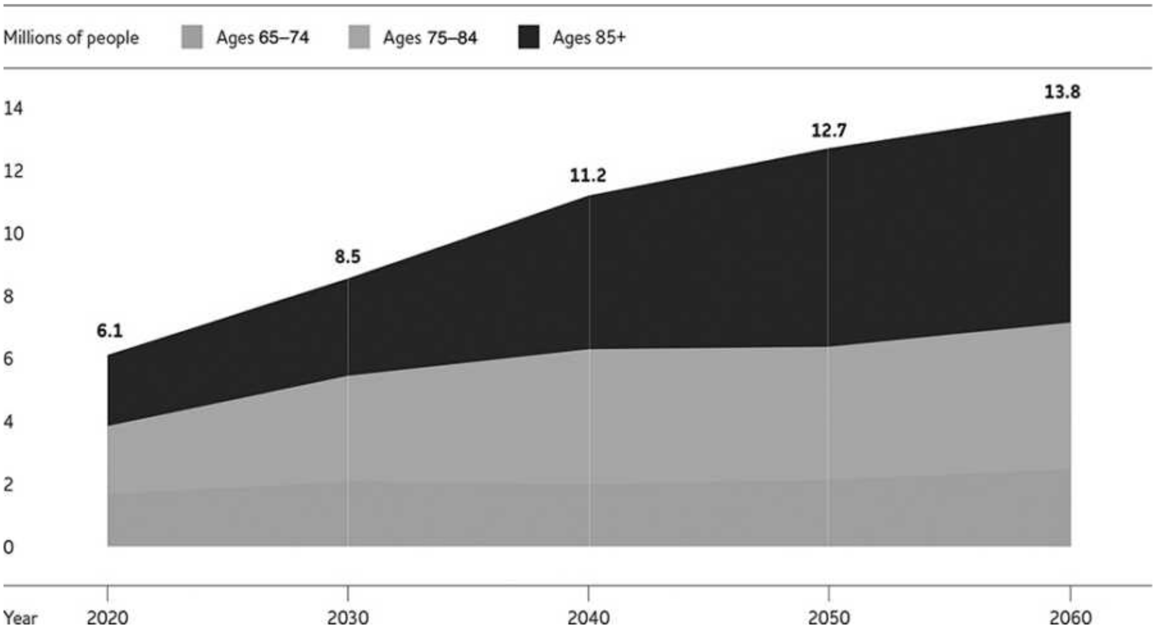

FIGURE 55.3f–1. Projected number of people aged 65 and older (Total and by Age) in the U.S. population with dementia due to Alzheimer disease, 2020 to 2060. (Data from Rajan KB, Weuve J, Barnes LL, McAninch EA, Wilson RS, Evans DA. Population estimate of people with clinical AD and mild cognitive impairment in the United States [2020–2060]. *Alzheimers Dementia*. 2021;17:1966–1975.)

ETIOLOGY

The search for possible causes for Alzheimer disease has seen significant advances but remains inconclusive. Scientists are approaching this search from various angles, which include genetics, neurochemistry, neuropathology, and neuroimaging studies. Hopefully, this search will eventually lead to treatments that may delay or even halt the disease process.

Alzheimer disease was first characterized and remains pathogenetically defined by its histopathologic features, neuritic plaques (extracellular deposits of amyloidogenic proteins), and neurofibrillary tangles (abnormal intracellular cytoskeletal filaments). Reduction in neuron number and synapses and an accumulation of plaques and tangles occur in the frontal, temporal, and parietal lobes; that finding is consistent with the involvement of association and limbic brain structures.

Extensive study of the formation of plaques and tangles and other neurobiologic changes associated with neurodegeneration have led to two prevailing hypotheses to explain the mechanisms that lead to Alzheimer disease: the cholinergic and the amyloid cascade hypotheses. The cholinergic hypothesis posits that a dysfunctional cholinergic system causes the symptoms of Alzheimer disease. Clinical evidence suggests that anticholinergic medications impair cognitive function in older adults, and biochemical and neuropathologic evidence confirm a presynaptic cholinergic deficit in brains with Alzheimer disease. The amyloid hypothesis states that Alzheimer disease is caused by the excess accumulation and deposition of β-amyloid to form neuritic plaques. Additionally, the accumulation of β-amyloid is believed to result in neuronal dystrophy and neurofibrillary tangles and induce cholinergic dysfunction, oxidative stress, inflammation, and vascular damage.

Discoveries in the area of genetics and molecular biology have been critical to our understanding of the etiology of Alzheimer disease and have shown relationships between genes and pathogenesis. The high frequency of Down syndrome (trisomy 21) patients who invariably develop Alzheimer disease neuropathology and dementia by their 40s led to finding significant linkage to chromosome 21 in some early-onset families. Excitement escalated when the gene coding for the amyloid precursor protein (APP) found in senile plaques was localized to the same chromosome 21 regions. APP mutations, however, are relatively rare. Most early-onset pedigrees not segregating APP mutations show mutations of a chromosome 14 gene (presenilin 1). Another form of early-onset Alzheimer disease is represented by families of Volga German origin, in whom chromosome 1 mutations (presenilin 2) have been identified.

Studies of the common late-onset disease (symptoms beginning after age 60 years) showed evidence for linkage or association or both for a chromosome 19 region. Apolipoprotein E (APOE) was a candidate Alzheimer disease susceptibility gene partly because APOE and CSF bound to immobilized β-amyloid peptide and antisera to APOE stain senile plaques and neurofibrillary tangles. In addition, APOE localized to the same region of chromosome 19 identified in linkage studies.

APOE has three allelic variants (2, 3, and 4). Everyone inherits one allele from each parent so that five common genotypes are possible (2/3, 3/3, 2/4, 3/4, and 4/4). The sixth genotype (2/2) is rare. In the general population, approximately 3% of persons have the 4/4 genotype, 20% have the 3/4 genotype, and most people have the 3/3 genotype. The APOE-4 allele increases risk and decreases age at dementia onset in a dose-related fashion. That is, Alzheimer disease risk is lowest for the 3/3 genotype, higher for the 3/4 genotype, and highest for the 4/4 genotype.

By contrast, the APOE-2 allele has a protective effect, so Alzheimer disease risk is lower for people with the 2/3 genotype than for those with the 3/3 genotype. Susceptibility to Alzheimer disease from APOE affects many races and has been confirmed worldwide.

More recent research has focused on inflammatory pathways and mediators that contribute to the neurodegenerative process. Pathologic β-amyloid can activate phagocytic microglia and astrocytes to secrete cytokines, chemokines, and other proinflammatory molecules, such as prostaglandins and free radicals, which can have deleterious effects on the brain. Several inflammatory processes are over-represented in brains with Alzheimer disease. Microglia are highly activated and surround β-amyloid plaques. Interleukin-1 overexpression promotes the synthesis and processing of APP and, thus, amyloid production and plaque deposition. It also induces the activity and mRNA expression of acetylcholinesterase, which reduces cholinergic activity. Chemokines and their receptors are upregulated and may contribute to plaque-associated inflammation and neurodegeneration. Complement activation fragments, reactive astrocytes, and activated microglia are all highly colocalized with plaques containing aggregated β-amyloid. Chronic inflammation is a well-known contributor to vascular disease and may be the common pathway to the prevalent coexistence of vascular and Alzheimer disease pathology. Another common link may be APOE-4, a risk factor for developing not only late-onset Alzheimer disease but also atherosclerosis. These relationships suggest that neurovascular damage is a primary occurrence and that subsequent injuries, including Aβ deposition, amplify and/or exacerbate vascular damage that then leads to neurodegeneration and, eventually, cognitive decline.

Our current database suggests that Alzheimer disease may be a heterogeneous condition and therefore has multiple causative factors. Investigations comparing familial (one or more affected first-degree relatives of a proband or large autosomal dominant kindreds) and sporadic Alzheimer disease patients, as well as patients with early-onset (60 years or younger) and late-onset dementia, have demonstrated clinical and biologic differences among them. Other studies suggest differences in disease expression between men and women, cerebral glucose metabolism, neuroendocrine function, and a possible interaction among age at onset, sex, and family history of dementia. Case-control studies have identified prior head trauma and depression as risk factors for Alzheimer disease, and higher educational achievement, larger brain size, and cardiovascular conditioning as protective factors. In some cases, the entire cause may lie in a single genetic defect, while in others, a mixture of genetic and environmental factors may play a role. The likelihood of multiple causative factors suggests that there are multiple approaches for developing effective prevention and treatment strategies.

CLINICAL EVALUATION

NCDs in late life usually begin with milder forms of cognitive decline and progress gradually. Early on, pathologic cognitive decline may be difficult to distinguish from normal aging, in which there is commonly an awareness of mild memory changes but with minimal objective cognitive deficits. However, as neurodegeneration progresses, MCI, often an intermediate state between normal aging and NCD, develops and places patients at increased risk for developing an NCD. Patients with primarily memory deficits may have the amnestic subtype of MCI, which is more likely than other MCI subtypes to progress to dementia due to Alzheimer disease.

The clinical hallmark of Alzheimer disease is memory impairment. Short-term memory is affected initially. Patients experience difficulty in learning new information, as shown by their inability to recall three objects after 5 minutes or even after just two distractions. Long-term memory may be affected, and patients may not recall important past personal information. Other cognitive areas may be impaired, including abstract thinking, judgment, personality, and language.

In a typical case of Alzheimer disease, the onset is so insidious that family members have difficulty recognizing when the impairment began. Diagnosis is often missed or delayed because of the insidious onset and gradual progression of symptoms, but also due to lack of attention to cognitive issues during medical screenings. Prevalence studies of community samples have documented high proportions of undiagnosed cases. A complete history of the patient’s cognitive and everyday functioning from someone who knows the patient well and physical, neurologic, and mental status examinations are essential in the clinical evaluation. The history should include screening questions for behavioral changes, including apathy, paranoia, agitation, insomnia, anxiety, and depression. On examination, the patient may show *apraxia* (inability to carry out motor commands, even though comprehension and motor function are intact), *agnosia* (inability to recognize objects, despite intact sensory function), or difficulties in visuospatial skills (inability to copy two-dimensional and three-dimensional figures or to assemble blocks). Brief standardized mental status tests (e.g., the Mini-Mental State Examination, Montreal Cognitive Assessment) are useful screening tools, and more extensive neuropsychologic batteries provide details on the nature and degree of cognitive deficits.

IMAGING AND LABORATORY EXAMINATION

Structural neuroimaging, preferably MRI, is appropriate to detect potentially treatable lesions that may affect cognitive functioning. MRI can also detect atrophy suggestive of a neurodegenerative process, though, in clinical practice, radiologists interpret structural scanning results based on visual readings, which may be insensitive in detecting subtle atrophy.

Clinical laboratory studies, including complete blood count, comprehensive metabolic panel, urinalysis, thyroid hormone levels, and vitamin B12 levels, should be ordered to detect potentially treatable medical conditions that may cause cognitive impairment.

Biomarkers

Recent advances in biomarker research have moved the field away from defining Alzheimer disease as a clinical syndrome toward defining it by its postmortem neuropathology or in vivo biomarkers. In 2018 the NIA-AA created a “research framework” that defines Alzheimer disease by the presence of certain imaging and CSF biomarkers based on studies that have shown that these biomarkers are valid proxies for the neuropathology of Alzheimer disease. The framework recognizes three groups of biomarkers: those that measure β-amyloid plaques, fibrillar tau, and neurodegeneration or neuronal injury. The authors’ intent in shifting from a syndromal to a biologic construct was primarily for research purposes, to “enable a more precise approach to interventional trials where specific pathways can be targeted in the disease process and in the appropriate people.”

The earliest currently used imaging biomarker is brain atrophy, which likely reflects loss and shrinkage of the neuropil. Atrophy in the entorhinal cortex and hippocampus is associated with an increased risk for progressive memory decline and the development of Alzheimer disease. MRI, in conjunction with image analysis programs, can detect relatively small volumes of brain atrophy. These analyses have shown that medial temporal or hippocampal atrophy measures can distinguish patients with a clinical diagnosis of Alzheimer disease from controls. PET scans measuring regional glucose metabolism after injection of fluorodeoxyglucose (FDG) also can detect brain atrophy and differentiate patients with Alzheimer disease from cognitively intact individuals and those with MCI and other NCDs. Patients with Alzheimer disease show hypometabolism in the posterior cingulate, parietal, temporal, and frontal regions, and these patterns can be observed in patients with MCI as well (Fig. 55.3f–2). Combining genetic risk assessment with PET is a potential strategy for the early detection of Alzheimer disease. Imaging studies using FDG-PET have shown brain metabolic deficits in nondemented middle-aged and older persons with the APOE-4 allele. In clinical practice, FDG-PET scans have been approved for reimbursement by the U.S. Centers for Medicare and Medicaid Services to assist with the differential diagnosis of Alzheimer disease and frontotemporal dementia, making this technique more accessible to patients and their physicians.

Several small-molecule probes for use with PET imaging have been found to provide in vivo measures of amyloid plaques and tau tangles. PET studies using Pittsburgh Compound-B (PIB), a derivative of thioflavin-T amyloid dye, show significantly greater cortical retention in patients with Alzheimer disease compared with controls (Fig. 55.3f–3). PET scans using FDDNP, which can measure both amyloid plaques and tau tangles, have been shown to differentiate patients with Alzheimer disease from those with MCI and cognitively intact controls and can predict future cognitive decline (see Fig. 55.3f–2). FDDNP-PET scans also can be displayed as three-dimensional cortical surface projections, which show signal patterns remarkably similar to those expected from autopsy studies demonstrating regional brain accumulation patterns of plaques and tangles (Fig. 55.3f–4). PET studies have shown that elevated binding of the tau ligand flortaucipir to medial temporal lobe structures and the neocortex is associated with clinical impairment across the normal aging to dementia spectrum and the presence of postmortem tau pathology. Newer PET ligands are being developed to, more specifically, image amyloid plaques or tau tangles.

CSF examination is another tool for the early detection of Alzheimer disease pathology. CSF levels of proteins associated with plaques and tangles, such as elevated phosphorylated tau and ratio comparisons of different β-amyloid peptides, have been shown to distinguish patients with Alzheimer disease from normal controls and other types of dementia, as well as correlate with imaging biomarkers and disease progression. While

CSF study is not used extensively in clinical settings, this technique may ultimately be more accessible and affordable than imaging. Scientists are investigating other potential fluid biomarkers that would circumvent the potential risks of lumbar puncture. A promising blood biomarker using 30 serum proteins showed 88% sensitivity and 82% specificity for Alzheimer disease.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of Alzheimer disease can be a challenge. Other neurocognitive, psychiatric, neurologic, and medical disorders can present similarly to Alzheimer disease. Except for measures of regional cerebral glucose metabolism using PET scanning, most biomarkers are still available only for research purposes, and the clinician is left with conventional methods of evaluation: the patient's history, mental status and physical examination, structural imaging, and laboratory tests. The history provides important information, and informants familiar with the patient's cognitive function should be sought for corroboration. Even with the most thorough evaluation, diagnostic errors occur. When patients are followed for an autopsy for neuropathologic confirmation, the accuracy of prior clinical diagnosis can be as low as 60%.

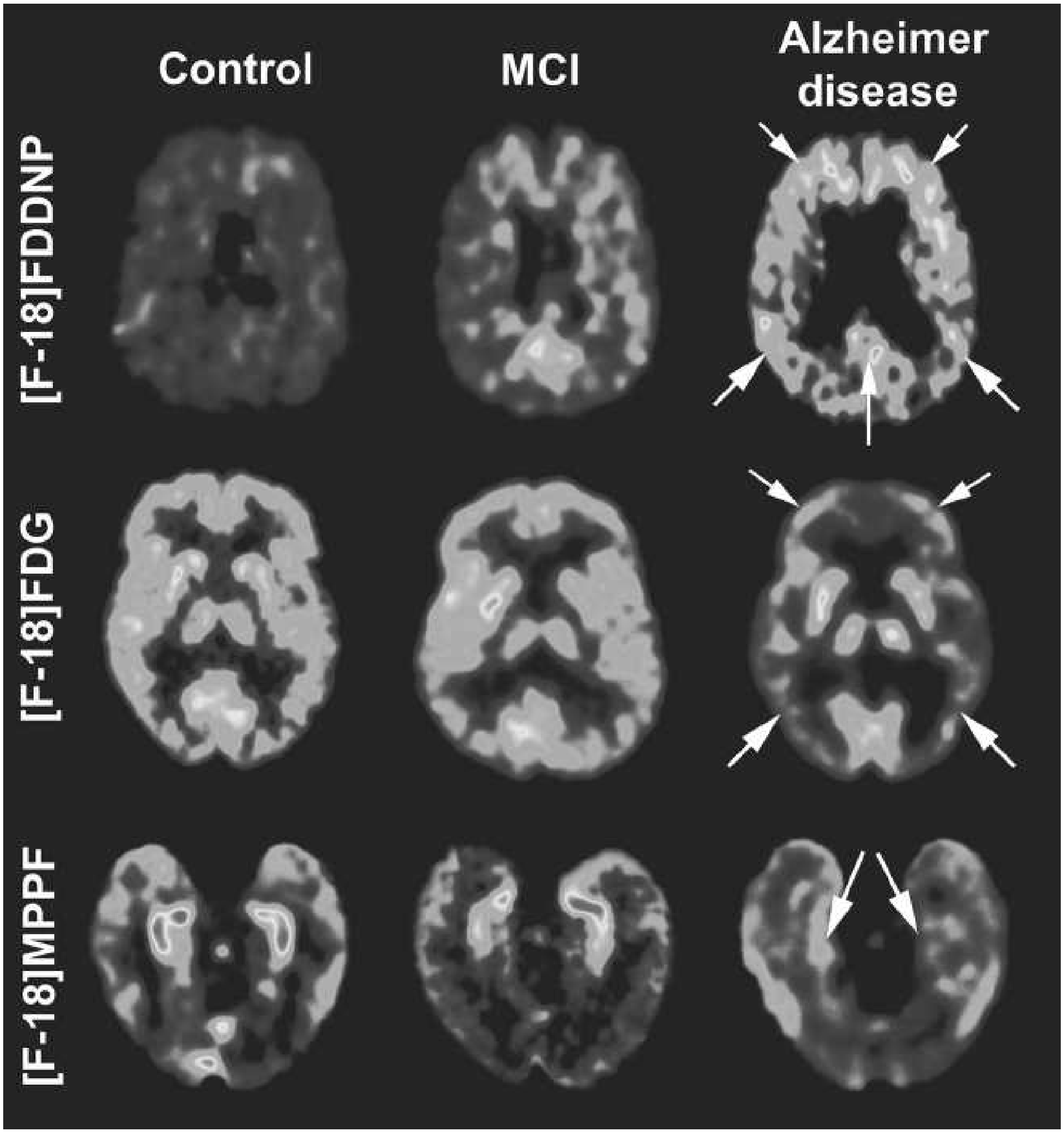

FIGURE 55.3f–2. Representative examples of brain positron emission tomographic (PET) images from a control subject **(left)**, a patient with mild cognitive impairment (MCI) **(center)**, and a patient with Alzheimer disease **(right)**, including the use of the agents FDDNP **(top)**, FDG **(middle)**, and MPPF **(bottom)**. For each imaging method, the patient with MCI shows intermediate values between those of the control and the patient with Alzheimer disease. Absolute scales are comparable within subjects with a specific probe, and lighter colors indicate higher values. In the patient with Alzheimer disease, the arrows point to frontal, parietal, and posterior cingulate **(upper)**, frontal and temporal-parietal **(middle)**, and medial temporal **(lower)** regions. (Adapted with permission from Kepe V, Barrio JR, Huang SC, et al. Serotonin 1A receptors in the living brain of Alzheimer's disease patients. *Proc Natl Acad Sci USA*. 2006;103(3):702–707.)

Major vascular NCD is estimated to account for 10% to 20% of major NCDs in late life. Clinically, the disorder is characterized by a sudden onset of illness and stepwise decline in cognitive function rather than the insidious onset and gradually progressive course of Alzheimer disease. The frequent coexistence of vascular disease and Alzheimer pathology (estimated to occur in about one-third of all major NCDs) impairs diagnostic accuracy.

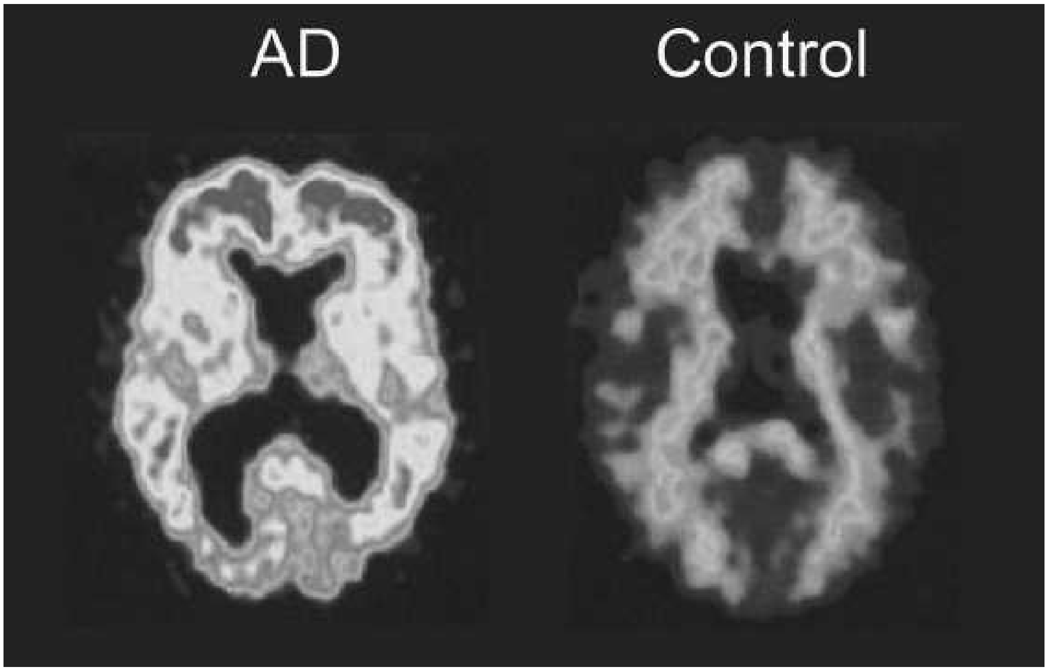

FIGURE 55.3f–3. Pittsburgh Compound-B (PIB)-standardized uptake images showing higher PIB retention in a 79-year-old patient with Alzheimer disease than in a 67-year-old healthy control. (Reproduced with permission from Klunk WE, Engler H, Nordberg A, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann Neurol*. 2004;55(3):306–319. doi: 10.1002/ana.20009.)

Even though Alzheimer disease and vascular disease account for most cases of major NCDs, other conditions must be considered and sometimes can be differentiated from Alzheimer disease by their clinical presentation. For example, frontotemporal lobar degeneration is characterized by marked personality and behavioral changes, executive dysfunction, and relative preservation of other cognitive domains. By contrast, patients with Lewy body disease may evidence visual hallucinations, delusions, fluctuating mental status, REM sleep behavior disorder, and sensitivity to antipsychotic medications.

Substances, including prescribed medications, may cause memory loss and cognitive dysfunction. For older patients, medications deserve special emphasis as a cause of cognitive impairment because of widespread polypharmacy and age-related changes in body function and pharmacokinetics. The American Geriatrics Society updates its Beers criteria for potentially inappropriate medication use in older adults every 3 years. These criteria include many medications that can cause cognitive impairment (JAGS). An older person's brain can be extremely sensitive to changes in its environment, and almost any physical illness can impair brain function and produce a dementia-like syndrome. Numerous potentially treatable physical and mental conditions and offending medications must be considered whenever the diagnosis of dementia is entertained (see Table 55.3f–1).

The clinical presentation of late-life depression may mimic that of an NCD. The earlier term "depressive pseudodementia" is no longer considered useful because the cognitive impairment associated with depression is real. The reported frequency of mistaking depressed patients for patients with an NCD ranges from 5% to 15%. Psychomotor retardation, impaired concentration, and apathy characteristic of the depressed patient may be mistaken for symptoms of an NCD, especially when these symptoms are accompanied by memory loss. In the dementia syndrome of depression, however, the complaints of memory tend to exceed the actual memory impairment observed on neuropsychologic testing. The fact that NCDs and depression often coexist further clouds the diagnostic picture. About 30% of patients with a progressive NCD also suffer from some form of depression.

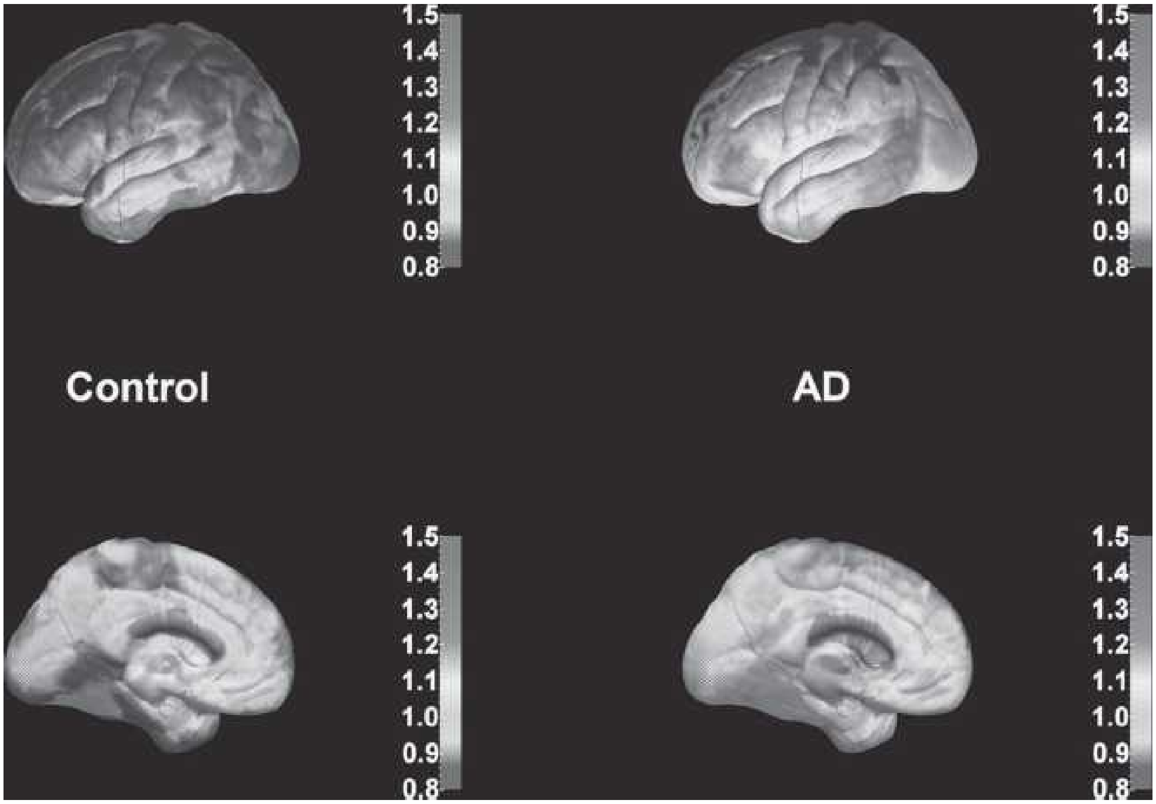

FIGURE 55.3f–4. Three-dimensional cortical surface projection images of FDDNP-positron emission tomography (PET) scans from a control subject and a patient with Alzheimer disease showing lateral **(upper)** and medial **(lower)** brain surfaces. Brighter colors indicate higher levels of plaques and tangles. (Reprinted from Small GW, Bookheimer SY, Thompson PM, et al. Current and future uses of neuroimaging for cognitively impaired patients. *Lancet Neurol*. 2008;7(2):161–172. doi: 10.1016/S1474-4422(08)70019-X.)

Depression is a risk factor and a prodrome of dementia, and shared neurobiologic mechanisms may influence both conditions. Although depression occurring in the course of dementia may reflect a psychologic reaction to cognitive losses, PET scanning of patients with late-life depression shows increased signals indicating higher levels of amyloid plaques and tau tangles in patterns that overlap with those observed in Alzheimer dementia (Fig. 55.3f–5).

Many people, as they age, complain of forgetfulness and worry about its meaning. Age-associated memory impairment, as defined by a workgroup convened by the National Institute of Mental Health (NIMH) in 1987, applies to healthy persons 50 years and older who have subjective complaints of memory loss and mild deficiencies in memory performance compared to young adults. About 40% of persons in their 50s and up to

85% of those aged 80 years and older suffer from age-associated memory impairment. Data suggest that the condition is stable over intervals of up to 4 years and thus is presumed to reflect normal aging and does not usually progress to dementia. Although it may take them longer to remember, most older adults continue to recall information they have learned well and important events from the past. Because age-associated memory impairment is so common, it is also referred to as normal aging.

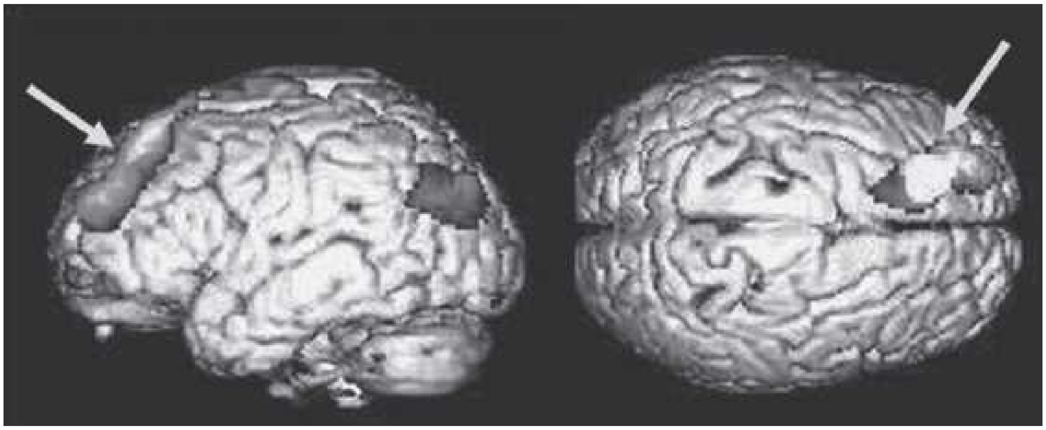

FIGURE 55.3f–5. Statistical parametric mapping of results of fluorodeoxyglucose (FDG)-PET scanning shows a 5% decline in metabolic activity in the left dorsolateral prefrontal cortex after a 2-week healthy lifestyle program, including memory training, physical conditioning, and healthy diet, whereas no such decline was observed in the control group. The darker areas highlight the location of all cortical voxels showing significantly greater decline ($P < .01$) in the intervention group compared with the control group. The left image is a left lateral viewpoint; the right image viewpoint is from the top of the brain. The arrows point to the voxels of peak significance ($Z = 3.82$, $P < .001$). (Reprinted from Small GW, Silverman DH, Siddarth P, et al. Effects of a 14-day healthy longevity lifestyle program on cognition and brain function. *Am J Geriatr Psychiatry*. 2006;14(6):538–545. doi: 10.1097/01.JGP.0000219279.72210.ca.)

COURSE AND PROGNOSIS

Alzheimer disease is a progressive disease that eventually leads to death, not directly from its impact on cognition, but usually from intercurrent illnesses. The rate of progression varies, and it is difficult to estimate how rapid a decline a patient may have. Some patients become extremely impaired within 1 year; others experience plateaus for several years. The majority experience behavioral disturbances such as agitation and psychosis at some point during the illness. Other causes of NCDs, including frontotemporal lobar degeneration, vascular disease, and Lewy body disease, are gradually progressive as well.

TREATMENT

The first step in treating NCDs is attempting to identify any underlying illnesses that may be causing cognitive impairment and other symptoms of the syndrome. Occasionally, dementia can be effectively resolved when the underlying cause (e.g., drug toxicity, hypothyroidism) is treated. If none is found or if the symptoms persist after appropriate specific treatment, then symptomatic treatments are indicated.

Psychopharmacotherapy

Several pharmacologic treatments are available for Alzheimer disease (Tables 55.3f–2 and 55.3f–3). Many of these approaches diminish symptoms and increase the quality of life, but none has proven to be a disease-modifying treatment or a treatment that has an enduring effect on the clinical progression of Alzheimer disease by interfering with the pathophysiologic processes that lead to cell death. Most pharmacologic agents now available or in development target, a specific symptom (e.g., agitation or memory loss) and are derived from the known neurobiology of the disease (e.g., a specific neurotransmitter deficit) or hypothesized antidementia approaches (e.g., antiamyloid, antitau, anti-inflammation, or antioxidation).

Table 55.3f–2.

Substances and Potentially Reversible Conditions That Might Cause Neurocognitive Impairment Resembling Alzheimer Disease

Substances
Anticholinergic agents
Antihypertensives
Antipsychotics
Corticosteroids
Digitalis
Opiates
Nonsteroidal anti-inflammatory agents
Phenytoin
Polypharmacy
Benzodiazepines
Psychiatric Disorders
Anxiety
Depression
Mania
Delusional (paranoid) disorders
Metabolic and Endocrine Disorders
Addison disease
Cushing syndrome
Hepatic failure
Hypercarbia (chronic obstructive pulmonary disease)
Hypernatremia
Hyperparathyroidism
Hypo/hyperthyroidism
Hypoglycemia
Hyponatremia
Renal failure
Volume depletion
Miscellaneous Conditions
Fecal impaction
Hospitalization
Impaired hearing or vision

One approach to improving memory function, based on the cholinergic hypothesis of Alzheimer disease, has been to enhance cholinergic activity in the brain. The U.S. Food and Drug Administration (FDA) has approved four cholinesterase inhibitors (ChEIs)—tacrine (Cognex), donepezil (Aricept), rivastigmine (Exelon), and galantamine (Razadyne)—and one *N*-methyl D-aspartate–receptor antagonist memantine (Namenda) for the treatment of Alzheimer disease. These drugs have proven benefits for memory, other aspects of cognition, behavioral disturbances, overall functioning, and even caregiver burden. Tacrine caused high rates of hepatotoxicity and was discontinued in 2013.

ChEIs were originally indicated only for mildly to moderately ill patients with Alzheimer disease, but their efficacy has since been demonstrated in severe Alzheimer disease, Lewy body disease, and dementia associated with Parkinson disease. Memantine is indicated for moderately to severely ill patients with Alzheimer disease. Data from clinical trials have demonstrated that combining a ChEI with memantine provides a beneficial response compared to either medication alone.

Table 55.3f–3.
Medications Used or Under Investigation for the Treatment of Alzheimer Disease

Cholinesterase Inhibitors
Tacrine[a]
Donepezil[a]
Rivastigmine[a]
Galantamine[a]
Sustained-release physostigmine
Eptastigmine
Cholinergic Agonists
Xanomeline
Milameline
Antioxidants
Vitamin E (α-tocopherol)
Hormones
Conjugated estrogens
***N*-Methyl D-aspartate Antagonist**
Memantine[a]
Other Potential Cognitive Enhancers
Acetyl-L-carnitine
Curcumin
Ginkgo biloba
Nonsteroidal anti-inflammatory drugs
Omega-3 fatty acids

[a] Approved by the U.S. Food and Drug Administration.

Clinical research also points to the importance of early detection and treatment with ChEIs. In randomized, placebo-controlled inpatients with Alzheimer disease, these medications result in less cognitive decline compared with a placebo after 6 months of treatment. During the following 6-month open-label treatment periods, the patients who were originally treated with a placebo were given active drug, and at 1 year, better cognitive performance was observed in patients who began drug treatment from the beginning of the trial compared with those who had been placebo-delayed for 6 months.

Studies using ChEIs in patients with MCI have been inconclusive. For example, a randomized, placebo-controlled trial in patients with MCI showed that donepezil delayed progression to Alzheimer disease compared with placebo after 1 year of treatment; however, no difference between the drug and placebo was found after 3 years of treatment.

Studies have examined a variety of other agents that may improve cognitive functioning, including estrogen, nonsteroidal anti-inflammatory agents, and antioxidant supplements. Unfortunately, any clinical benefit from these various treatments has been inconclusive. The widespread use of over-the-counter preparations should be evaluated routinely, as some do carry risks. For example, while vitamin E (α-tocopherol) at 2,000 IU daily lowered rates of functional decline in patients with moderately severe Alzheimer disease in one study, meta-analyses have found an increased risk of myocardial infarction in older patients taking daily vitamin E supplementation, particularly doses greater than 400 IU.

Research on treatments that are superior to current therapeutics (the last one approved by the FDA in 2003) has focused on disease-modifying therapies. As of 2020, nearly a hundred potentially disease-modifying treatments were under investigation, representing 80% of all agents under investigation for Alzheimer disease. Seventeen agents were in Phase 3 clinical trials. Mechanisms of action target a variety of pathways, including amyloid (monoclonal antibodies, vaccines, antiaggregation), tau (antiaggregation), inflammation (protease inhibitor, omega-3 fatty acid), metabolic (insulin sensitizer), vascular (antihypertensives, statin), synaptic plasticity (synaptic vesicle protein and glutamate modulators), and neurogenesis. Developing these treatments is quite challenging, as these studies require more participants, biomarkers, longer trials, and greater cost compared with symptomatic treatments.

Psychiatric medications may be helpful for treating behavioral disturbances associated with dementia, such as depression, agitation, psychosis, and anxiety, although the FDA has approved none for such use. (Table 55.3f–4) Patients who have dementia with concurrent depression may benefit from antidepressant medications, though data are limited. A 2018 Cochrane review of 10 randomized placebo-controlled trials of different antidepressants with a total of 1,592 participants with dementia and depression suggested questionable benefit. The authors found high-quality evidence of little or no difference in depression symptom rating scales between antidepressants and placebo. There was high-quality evidence of no effect of antidepressants on activities of daily living and moderate-quality evidence of no effect on cognition. There was moderate quality evidence of higher remission rates in the antidepressant group than in the placebo group. The data for antidepressants are even more limited for treating agitation. One randomized controlled trial of 186 participants with Alzheimer disease and agitation showed that citalopram reduced agitation significantly more than placebo. A smaller study of patients with Alzheimer disease on donepezil showed that sertraline reduced overall behavioral disturbances more than placebo.

Table 55.3f–4.

Medications Used or Under Investigation for the Treatment of Behaviors Associated With Alzheimer Disease and Other Neurocognitive Disorders

Antipsychotics
Aripiprazole
Haloperidol
Olanzapine
Quetiapine
Risperidone
Brexpiprazole
Anticonvulsants/Mood Stabilizers
Carbamazepine
Divalproex sodium
Gabapentin
Lithium
Antidepressants
Citalopram
Escitalopram
Paroxetine
Sertraline
Trazodone
Anxiolytics
Buspirone
Clonazepam
Lorazepam
Antiparkinsonian Drugs
Selegiline
α-Adrenergic Receptor Antagonists
Propranolol
Pindolol
Prazosin
Others
Dextromethorphan/Quinidine
Dronabinol
Methylphenidate
Pimavanserin

Antipsychotic drugs may be effective in treating psychotic symptoms and agitation. Multiple controlled trials and meta-analyses of controlled trials of antipsychotic treatment in dementia indicate that antipsychotics have a significantly greater effect than placebo, but the effect sizes are mostly less than 20%. As there is no evidence that one agent is superior to others, the side-effect profile is a primary consideration when choosing a specific agent.

First-generation or “conventional” antipsychotics may be generally categorized as “high-potency” or “low-potency” antipsychotics, according to how potently they occupy dopamine-2 receptors. High-potency agents such as haloperidol are more likely to cause parkinsonian symptoms, whereas such low-potency drugs as chlorpromazine cause sedation, postural hypotension, and anticholinergic effects. Less frequently, tardive dyskinesia and neuroleptic malignant syndrome may develop with either. Clozapine, while considered an “atypical” antipsychotic, can have similar side effects as low-potency antipsychotics and requires routine blood-count monitoring because of the risk of agranulocytosis. Because of the lower associated risk for neuromotor side effects, second-generation atypical antipsychotic drugs, such as quetiapine, risperidone, olanzapine, and aripiprazole, have received greater attention in recent years. However, these atypical agents are associated with other adverse effects, including metabolic syndrome, vascular events, and cardiac arrhythmias, which warrant vigilant monitoring. Overall, conventional and atypical antipsychotic medications have been associated with increased mortality rates in older adult patients with dementia, which prompted the FDA to issue black box warnings for all antipsychotics. Investigators have more recently challenged the basis for these warnings, with data showing no significant differences in mortality in this population between antipsychotics and placebo. Pimavanserin, which has FDA approval for treating psychosis in Parkinson disease without dementia, has shown potential efficacy and an acceptable tolerability profile in treating psychosis in Alzheimer disease and other dementias.

Benzodiazepines and anticonvulsants have been used and studied for the treatment of agitation. However, there is little evidence for their use in patients with dementia, and they have undesirable adverse effects, such as sedation, confusion, memory impairment, dysarthria, and ataxia. Moreover, chronic benzodiazepine use has been associated with an increased risk for Alzheimer disease.

Dementia Care Management

While disease-modifying treatments remain elusive, patients with Alzheimer disease and other progressive dementias require a holistic approach to care, given the many ways these diseases affect their lives. In 2015, the American Academy of Neurology (AAN) and the American Psychiatric Association (APA) formed a Dementia Measurement Set Update Work Group to review and update previous guidelines on dementia care. The Work Group designed a comprehensive dementia management strategy that includes measures for disclosure of dementia diagnosis, education and support of caregivers, functional status assessment, screening and management of behavioral and psychiatric symptoms, safety concern screening and follow-up, driving screening, advance care planning and palliative care counseling, pain assessment, and pharmacologic treatment. Descriptions of these interventions are in Table 55.3f–5. Some of the desired outcomes from this strategy include: increased patient and caregiver engagement in disease education, better treatment planning and advanced decision making, improved quality of care and caregiver burden, maintenance or improvement of cognitive, physical, and functional levels, decreased complications and behavioral and psychiatric symptoms, improved safety awareness and risk management, and recognition of and adherence to end of life wishes.

Attention to the Environment. Attending to patients' environments is essential to their well-being and can mitigate the need for pharmacotherapy for behavioral disturbances. Patients with cognitive losses are sensitive to their surroundings and do their best with optimal stimulation. Understimulation may cause withdrawal; overstimulation may cause confusion and agitation. Familiar and consistent surroundings and caregivers will maximize the patient's cognitive functions and minimize behavioral disturbances. If moves cannot be avoided, placing familiar objects (e.g., photographs and furniture) in the new environment helps create a more recognizable, comfortable, and homelike atmosphere. Daily routines—regularly scheduled awakening, bedtime, meals, and physical exercise—increase a patient's sense of security. Memory and orientation can be facilitated by prominent displays of clocks and calendars, daylight exposure, a night light, checklists, and diaries. Medication schedules should be simplified as much as possible. Newspapers, radio, and television, appropriate to the patient's cognitive level, can be useful in maintaining a patient's contact with and awareness of the outside world. Daily simple yet meaningful tasks that the patient can complete provide a valuable sense of usefulness and accomplishment for the patient. Specific therapies such as music and activity therapies have also been shown to improve behavior.

Family Intervention. Addressing the needs of the patient with Alzheimer disease would not be complete without addressing the needs of the patient’s care network. While the identified patient is the one suffering from the disease and in need of treatment, often the spouse or adult child initiates contact with the physician and has their own reasons for evaluation and treatment. Considerable skill and sensitivity are necessary to maintain an alliance with family members while respecting the patient’s autonomy, dignity, and privacy. Psychotherapeutic intervention with family members is a critical aspect of treatment. Caregiver and family stress can result in suboptimal care for the patient, as they themselves might experience frustration, anxiety, fatigue, anger, resentment, guilt, and even depression. Interpersonal conflicts among the family members may arise and require resolution to help the patient. Children and siblings might become anxious about their own risk of developing Alzheimer disease. Education and counseling about the nature of the patient’s illness help relatives to cope with their feelings. Relatives may need reassurance that their emotional reactions are common and that talking about them can bring relief. Many relatives also need help in grieving the loss of the person they once knew. The Alzheimer’s Association, a national organization of family members and stakeholders with local chapters throughout the United States has been at the forefront of providing educational and emotional support for family members. Sometimes, a family member’s needs may exceed this level of support and require individual attention in the form of treatment for anxiety or depression or a full physical and neuropsychologic evaluation, which may either allay anxiety about having the disease or occasionally uncover early signs of cognitive impairment, which may lead to more anxiety, but also the opportunity for early intervention.

Table 55.3f–5.
2015 Dementia Management Measurement Set Update

Measure Title	Measure Description
Disclosure of dementia diagnosis[a]	Percentage of patients with a diagnosis of a qualifying dementing disorder or disease whose diagnosis has been disclosed to them and, if available, their primary caregiver
Education and support of caregivers for patients with dementia	Percentage of patients with dementia whose caregivers were provided with education on dementia disease management and health behavior changes and were referred to additional resources for support in the last 12 months
Functional status assessment for patients with dementia	Percentage of patients with dementia for whom an assessment of functional status was performed at least once in the last 12 months
Screening and management of behavioral and psychiatric symptoms associated with dementia (BPSD)	Percentage of patients with dementia for whom there was a documented screening for behavioral and psychiatric symptoms, including depression, and for whom, if screening positive, there was also documentation of recommendations for management in the last 12 months
Safety concern screening and follow-up for patients with dementia	Percentage of patients with dementia or their caregivers for whom there was a documented safety screening in 2 domains of risk, dangerousness to self or others and environmental risks, and for whom, if screening positive, there was documentation they were provided with recommendations for their mitigation, which may include referral to other resources, in the last 12 months
Driving screening and follow-up for patients with	Percentage of patients with dementia for whom there was a documented screening

dementia	for driving risks and for whom, if screening positive, there was also documentation they were informed of alternatives to driving in the last 12 months
Advance care planning and palliative care counseling for patients with dementia	Percentage of patients with dementia who have an advance care plan or surrogate decision maker documented in the medical record or documentation in the medical record that an advance care plan was discussed but the patient did not wish or was not able to name a surrogate decision maker or provide an advance care plan and percentage of patients with dementia or their surrogate decision maker who received comprehensive counseling regarding ongoing palliation and symptom management and end-of-life decisions within 2 years of initial diagnosis or assumption of care
Pain assessment and follow-up for patients with dementia[a]	Percentage of patients with dementia who underwent documented screening for pain symptoms at every visit and if screening positive also had a documentation of a follow-up plan
Pharmacologic treatment of dementia[a]	Percentage of patients with dementia or their caregivers with whom available guideline-appropriate pharmacologic treatment options and nonpharmacological behavior and lifestyle modifications were discussed at least once in the last 12-month period

[a] New measure added in 2015.

Reprinted from Sanders AE, Nininger J, Absher J, Bennett A, Shugarman S, Roca R. Quality improvement in neurology: dementia management quality measurement set update. *Neurology*. 2017;88(20):1951–1957. doi: 10.1212/WNL.0000000000003917.

Prevention Strategies. The MacArthur Foundation's studies on aging have shown that early lifestyle choices, more than genes, determine our health and vitality as we age. Epidemiologic, survey, and animal data suggest that these lifestyle choices can specifically impact brain health. Reducing modifiable cardiovascular risk factors, depression, low education, and physical inactivity by just 10% to 25% could potentially prevent millions of cases of Alzheimer disease worldwide. Some of these same risk factors have been shown to be associated with subjective memory impairment, a possible precursor to MCI and Alzheimer disease, in cognitively intact persons of all ages. These same individuals reported better memory if they engaged in healthy lifestyle habits such as regular exercise, a nutritious diet, and not smoking. One controlled study demonstrated that participating in a healthy lifestyle program combining mental and physical exercise, stress reduction, and a healthy diet for just 2 weeks was associated with significant positive effects on cognitive performance and brain metabolism (see Fig. 55.3f–5). Subsequent primary prevention studies have similarly focused on multidomain prevention strategies, incorporating a combination of nutritional education, physical activity, and vascular care, and found mixed results, either no benefit or potential improvement or maintenance of cognition, particularly among individuals with increased risk for dementia. Results from three large multidomain dementia prevention trials suggest that targeting preventive interventions to at-risk people is an effective strategy. Modifying one's lifestyle to incorporate some of these healthy behaviors comes with minimal risk and offers other proven health benefits. Until more effective pharmacotherapies are available, prevention strategies through healthy lifestyle choices may be the safest and most beneficial way to reduce one's risk for developing Alzheimer disease and other NCDs.

SUMMARY

Alzheimer disease and other NCDs continue to challenge clinicians and researchers to find effective ways to reduce the suffering of patients afflicted with these disorders. We have learned a great deal about their neurobiology and how to diagnose these illnesses. Scientists are working rigorously to discover evidence for disease-modifying treatments. Until such a discovery is made, clinicians can still have a meaningful impact on the lives of patients and their families through a combination of treatment modalities, including the use of available symptomatic medications, preventive medical care, a healthy lifestyle strategy, education, and comprehensive dementia care.

FURTHER READINGS

Adelman RD, Tmanova LL, Delgado D, Dion S, Lachs MS. Caregiver burden: a clinical review. *JAMA*. 2014;311(10):1052–1060.

Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011:7(3):270–279.

Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol*. 2011;10:819–828.

Bjelakovic G, Nikolova D, Gluud LL, Simonetti RG, Gluud C. Antioxidant supplements for prevention of mortality in healthy participants and patients with various diseases. *Cochrane Database Syst Rev*. 2012;2012(3):CD007176.

Burns A, Iliffe S. Dementia. *Br Med J*. 2009;338:b75.

By the 2019 American Geriatrics Society Beers Criteria® Update Expert Panel. American Geriatrics Society 2019 updated AGS Beers Criteria® for potentially inappropriate medication use in older adults. *J Am Geriatr Soc*. 2019;67(4):674–694.

Chen ST, Siddarth P, Ercoli LM, Merrill DA, Torres-Gil F, Small GW. Modifiable risk factors for Alzheimer disease and subjective memory impairment across age groups. *PLoS ONE*. 2014;9(6):e98630.

Crous-Bou M, Minguillón C, Gramunt N, Molinuevo JL. Alzheimer's disease prevention: from risk factors to early intervention. *Alzheimers Res Ther*. 2017;9(1):71.

Cummings J, Ballard C, Tariot P, et al. Pimavanserin: potential treatment for dementia-related psychosis. *J Prev Alzheimers Dis*. 2018;5(4):253–258.

Cummings J, Fox N. Defining disease modifying therapy for Alzheimer's disease. *J Prev Alzheimers Dis*. 2017;4(2):109–115.

Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer's disease drug development pipeline: 2020. *Alzheimers Dement (N Y)*. 2020;6(1):e12050.

Donix M, Small GW, Bookheimer SY. Family history and APOE-4 genetic risk in Alzheimer's disease. *Neuropsychol Rev*. 2012;22(3):298–309.

Dubois B, Feldman HH, Jacova C, et al. Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol*. 2010;9:1118–1127.

Dudas R, Malouf R, McCleery J, Dening T. Antidepressants for treating depression in dementia. *Cochrane Database Syst Rev*. 2018;8(8):CD003944.

Jack CR Jr, Bennett DA, Blennow K, et al. Contributors. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14(4):535–562.

Kales HC, Kim HM, Zivin K, et al. Risk of mortality among individual antipsychotics in patients with dementia. *Am J Psychiatry*. 2012;169:71–79.

Kepe V, Bordelon Y, Boxer A, et al. PET imaging of neuropathology in tauopathies: progressive supranuclear palsy. *J Alzheim Dis*. 2013;36:145–153.

Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nat Rev Neurol*. 2018;14:653–666.

Kumar A, Kepe V, Barrio JR, et al. Protein binding in patients with late-life depression. *Arch Gen Psychiatry*. 2011;68:1143–1150.

Lee JC, Kim SJ, Hong S, Kim Y. Diagnosis of Alzheimer's disease utilizing amyloid and tau as fluid biomarkers. *Exp Mol Med*. 2019;51(5):1–10.

McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:263–269.

Merrill DA, Siddarth P, Kepe V, et al. Vascular risk and FDDNP-PET influence cognitive performance. *J Alzheimers Dis*. 2013;35:147–157.

Miller KJ, Siddarth P, Gaines JM, et al. The memory fitness program: cognitive effects of a healthy aging intervention. *Am J Geriatr Psychiatry*. 2012;20:514–523.

Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. *J Intern Med*. 2014;275:214–228.

Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and meta-analysis. *Alzheimer's Dement*. 2013;9:63–75.e2.

Raji CA, Eyre H, Wei SH, et al. Hot topics in research: preventive neuroradiology in brain aging and cognitive decline. *AJNR Am J Neuroradiol*. 2015;36:1803–1809.

Sanders AE, Nininger J, Absher J, Bennett A, Shugarman S, Roca R. Quality improvement in neurology: dementia management quality measurement set update. *Am J Psychiatry*. 2017;174(5):493–498.

Small GW. Treating dementia and agitation. *JAMA*. 2014;311:677–678.

Small G, Bullock R. Defining optimal treatment with cholinesterase inhibitors in Alzheimer's disease. *Alzheimers Dement*. 2011;7:177–184.

Small G, Vorgan G. *The Alzheimer's Prevention Program*. Workman; 2012.

Small GW, Siddarth P, Ercoli LM, Chen ST, Merrill DA, Torres-Gil F. Healthy behavior and memory self-reports in young, middle-aged, and older adults. *Int Psychogeriatr*. 2013;25:981–989.

Small GW, Siddarth P, Kepe V, et al. Prediction of cognitive decline by positron emission tomography of brain amyloid and tau. *Arch Neurol*. 2012;69:215–222.

Vandenberghe R, Adamczuk K, Dupont P, Laere KV, Chételat G. Amyloid PET in clinical practice: its place in the multidimensional space of Alzheimer's disease. *Neuroimage Clin*. 2013;2:497–511.

Yokomizo JE, Simon SS, Bottino CM. Cognitive screening for dementia in primary care: a systematic review. *Int Psychogeriatr*. 2014;26:1783–804.

55.3g Delirium

JAYSON VARUGHESE, M.D., M.S., AND SUSAN W. LEHMANN, M.D., M.S.

Delirium is an acute cognitive disorder of attention and awareness which disproportionately affects geriatric patients. Reports indicate that delirium is present in one-third of hospitalized older patients and occurs in 15% of older patients presenting to emergency departments. The incidence of delirium is high among individuals with dementia, especially Alzheimer disease. Delirium is usually an indicator of severe underlying medical illness or metabolic derangement. However, it may also occur in vulnerable individuals with relatively minor medical provocations, such as constipation or mild dehydration.

Delirium is frequently unrecognized, yet the costs of not detecting delirium are significant. Delirium is associated with poor outcomes, both in the hospital and after discharge. Patients who are delirious are more likely to have longer lengths of hospitalization. Older patients with delirium are at increased risk for nursing home placement and are at increased risk for prolonged cognitive impairment and dementia. Delirium is associated with high rates of morbidity as well as an increased risk of mortality. Indeed, it has been estimated that the U.S. healthcare costs which are attributable to delirium range from 40 to 150 billion dollars annually. Studies indicate that delirium may be preventable in up to 30% to 40% of cases, and so assessment for delirium has become a significant focus for patient safety management in acute care hospitals.

DEFINITION

Delirium is a multifactorial syndrome whose hallmark features are that of disturbance in attention, awareness, and cognition. The symptoms develop acutely, usually over hours or days, in the setting of serious medical illness or in the postoperative period. Symptoms include impaired ability to focus or sustain attention as well as disorientation to the environment. Delusions or perceptual disturbances such as hallucinations (usually visual) are common, as are incoherent and disorganized speech. Patients usually have disrupted sleep–wake cycles and may be lethargic and drowsy or agitated and hyperalert. A key feature of delirium is that these disturbances tend to fluctuate in severity during the day.

The *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition, text revision (DSM-5-TR), classifies delirium as a neurocognitive disorder. Historically, other terms used to describe delirium include “acute brain failure,” “acute confusional state,” “organic brain syndrome,” or “encephalopathy.” However, there is some consensus advocating the use of two terms: delirium and acute encephalopathy. Acute encephalopathy is not a syndrome; rather, it is defined as a rapidly developing pathologic process and could be the underlying disorder causing delirium.

HISTORY

The syndrome of delirium has been described in medical writing for over 2,000 years. It was recognized as early as the fifth-century BC, when Hippocrates recorded a detailed description of the clinical features of delirium, noted that the syndrome was linked to physical disease, and suggested that it carried a poor prognosis. In the first century AD, the Roman Celsus distinguished delirium from major mood disorders and hysteria. In the second century, Galen also noted that the brain could be disturbed *secondarily* in diseases such as pneumonia. In antiquity and in medieval times, delirium was recognized, usually in the context of febrile illness, and was called by a variety of different names.

In his classic textbook, *General Psychopathology,* first published in 1913, the German psychiatrist Karl Jaspers reported his observations of delirious patients as follows:

> Delirium is characterized by the patient turning away from external reality . . . He is dominated by an anxiety which is often extreme and a drive for aimless activity . . . When left alone the patient is always on the verge of sleep which, however, is never fully achieved.

In 1935, Harold Wolff and Desmond Curran published a study of 106 patients with delirium hospitalized in the United States and the United Kingdom, with careful cataloging of psychobiologic functions and case reports for 20 of the patients. These authors made several observations of note—for example, that patients recede into vagueness and hallucinations when external stimulation ceases and that summoning relatives to the bedside of the delirious patient does much to restore function. Most important, however, was that they established for this cohort of patients that symptomatology was independent of etiology in delirium.

In the 1940s, John Romano and George Engel performed a series of EEG studies on delirious patients and demonstrated that the degree of EEG slowing was correlated with the severity of the disturbance in consciousness. These authors hypothesized that delirium represented a state of “cerebral insufficiency.”

Zbigniew J. Lipowski wrote extensively about delirium and emphasized both the prevalence and seriousness of delirium as a public health problem. His work ushered in the contemporary era of research about delirium, and his book, *Delirium: Acute Confusional States,* published in 1990, is considered a classic in the field. Sharon Inouye and her colleagues in geriatric medicine have carried out a series of investigations regarding the prevention of delirium, which led to the development of a successful preventive multistep approach to delirium prevention and management. This work has highlighted the importance of recognizing delirium as an indicator of the quality of care.

COMPARATIVE NOSOLOGY

The *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition, text revision (DSM-5-TR) defines delirium as a neurocognitive disorder whose key features are disturbances of attention and awareness, which develop acutely and which fluctuate over time. Typically, these changes in attention and awareness ebb and flow throughout the day such that the patient may appear lucid and talkative at one point in time and then a few hours later might appear lethargic and confused. In addition, there are impairments in other aspects of cognitive functioning, such as orientation to time and place, memory, expressive and/or receptive language, and visuospatial abilities. Abnormal perceptions, especially visual and tactile hallucinations, are common. There must be evidence from the patient's history and physical examination or laboratory investigations to indicate that these disturbances are due to an underlying medical illness or due to substance intoxication or withdrawal.

The DSM-5-TR criteria include specifications about whether the delirium is acute (i.e., lasting a few hours or a few days) or if it is persistent (i.e., lasting weeks or months). In addition, the DSM-5-TR recognizes three psychomotor subtypes of delirium: hyperactive, hypoactive, and mixed levels of activity.

For delirium, due to another medical condition, the DSM-5-TR requires that there be evidence from the history, physical examination, or laboratory findings that the delirium is due to a specific underlying medical condition. In clinical settings, it is unusual for there to be one single, identifiable cause that is clearly the etiology of the delirious state, and, more commonly, delirium is attributable to multiple causes. DSM-5-TR also specifies substance intoxication delirium and substance withdrawal delirium, which should be used when the symptoms of disturbed attention, awareness, and impaired cognition are most significant from a clinical standpoint and require clinical attention.

In the *International Classification of Diseases*, 10th revision (ICD-10), delirium due to medications or substances is specified separately by the causative medication or substance.

EPIDEMIOLOGY

Incidence and prevalence rates for delirium vary widely, depending on the clinical setting and patient characteristics. Prevalence in the community is approximately 0.4% of adults and 1.1% of people aged 55 and older. Among patients over age 70 admitted to a hospital, about one-third have delirium. Of these individuals, about half are delirious at the time of admission, and the rest develop delirium during their hospitalization. The onset of the presenting symptoms of delirium is frequently a reason that a patient is brought to an emergency department for evaluation. Overall, the incidence of delirium during hospitalization ranges from 6% to 56%, with higher rates being more commonly seen among patients with dementia. Incidence rates of delirium also vary across hospital services. Incidence rates of delirium are highest in intensive care units, postoperative clinical units, and palliative care settings. Prevalence rates of delirium have been reported to range from 23% in hospitalized older adults in inpatient settings to 35% in palliative care settings and up to 70% in mechanically ventilated patients. Among residents in nursing homes who are aged 75 and older, the prevalence of delirium is as high as 60%. Among patients with terminal illnesses, up to 80% of individuals become delirious around the time of death. The wide range of prevalence rates across settings reflects the interaction between the individual's risk factors and clinical settings.

Risk factors for delirium are listed in Table 55.3g–1. These include pre-existing characteristics which predispose an individual to delirium as well as factors that precipitate an episode of delirium. Predisposing risk factors include pre-existing cognitive impairment, advanced age, frailty, previous history of delirium, severe medical illness, alcohol or sedative dependence, hearing or vision impairment, and poor nutritional status. Precipitating factors include bed rest or immobilization, medications (especially opioids, anticholinergic medications, and benzodiazepines), physical restraints, sleep deprivation, and low hematocrit. Awareness of risk factors for delirium, especially those which are modifiable, is critical for delirium prevention efforts.

Table 55.3g–1.
Risk Factors for Delirium

Pre-existing cognitive impairment (dementia, executive dysfunction)
Previous history of delirium
Multiple medical problems
Multiple medications (especially opioids, anticholinergics, and benzodiazepines)
Advanced age
Structural brain disease (e.g., stroke or traumatic brain injury)
Alcohol or sedative dependence
Chronic liver disease
Severe medical illness
Terminal illness
Chronic psychiatric illness
Undernutrition (low albumin and prealbumin)
Compromised cerebral circulation
Deafness or hearing impairment
Blindness or visual impairment
Sleep deprivation
Immobilization
Frailty
Stress, including psychosocial stress

ETIOLOGY

While delirium can be caused by a single etiologic factor, most commonly, the etiology is multifactorial. Predisposing risk factors, as described above, confer vulnerability to delirium. The development of delirium depends upon the complex interplay of predisposing and precipitating factors. Aging and neurodegenerative conditions can affect key processes, including brain connectivity, neuroinflammation, glial cell alterations, and vascular changes. These factors may underlie an increased vulnerability to delirium due to a relatively minor condition, such as a urinary tract infection. At the same time, delirium often develops as a consequence of serious medical and/or metabolic derangement and, therefore, is considered a medical emergency. Recognition of the multifactorial etiology of delirium is essential to clinical management because a multidimensional approach to both prevention and treatment is needed.

Table 55.3g–2 presents a list of etiologies for delirium frequently seen in older patient populations. Among hospitalized older adults, the most important causes are infections and medications. Infections can include urinary tract infections, pneumonia, and sepsis. Many medications have been implicated as being deliriogenic in vulnerable populations, and their use should be avoided if possible. In particular, opioids, anticholinergic medications, and benzodiazepines are frequent causes of delirium in older patients. Anticholinergic medications include atropine, diphenhydramine, benztropine, and oxybutynin (Ditropan), as well as medications with strong anticholinergic side effects such as imipramine, ranitidine (Zantac), dipyridamole (Persantine), warfarin (Coumadin), isosorbide, theophylline (Theo-Dur), nifedipine (Adalat, Procardia), and prednisolone (Prelone). Among opioid medications, meperidine appears to be most likely to induce delirium, and hydromorphone is the least likely to cause delirium. Among the benzodiazepines, the most lipophilic drugs with the longest elimination half-lives are most likely to cause delirium, including chlordiazepoxide (Librium), flurazepam (Dalmane), and diazepam (Valium).

Table 55.3g–2.
Etiologies of Delirium in Older Adult Patients

Systemic Illnesses

Infections: Pneumonia, urinary tract infection, sepsis, influenza

Metabolic derangements: Dehydration, hypoxia, hypoglycemia, hyperammonemia, uremia, hyponatremia, thiamine deficiency, hyperthyroidism

Cardiovascular conditions: Arrhythmia, congestive heart failure, myocardial infarction, severe hypertension

Medications

Anticholinergics
Benzodiazepines, other sedative hypnotics (e.g., barbiturates)
Opioids
Antiarrhythmics
Certain antibiotics (e.g., fluoroquinolones, clarithromycin)
Digoxin
Interferons

Withdrawal States

Alcohol
Benzodiazepines, other sedative hypnotics

Primary Brain Diseases

Stroke or transient ischemic attack
Trauma: Brain injury, subdural hematoma
Infection/inflammation: Abscess, meningitis, encephalitis, vasculitis
Aneurysm
Epilepsy: Nonconvulsive status epilepticus, postictal states

Surgery or Trauma

Hip fracture repair
Open heart surgery (e.g., coronary artery bypass grafting)

A 79-year-old woman with metastatic ovarian cancer was admitted to the hospital with acute onset of vomiting, inability to stand, and delirium. She had been treated with warfarin for a hypercoagulable state. On admission workup, she was found to have a cerebellar hemorrhage with evidence of increased intracranial pressure. Anticoagulation was reversed, the hemorrhaging stopped, and intracranial pressure was normalized. Digoxin was started for newly diagnosed atrial fibrillation. She returned to her baseline level of cognitive function, which was completely normal. She was transferred to a rehabilitation facility to improve her ability to ambulate, which was still affected by the cerebellar hemorrhage. Soon after her admission to the rehabilitation unit, she was reported to be disoriented. She began to accuse the staff of waking her up at 3 AM. Labs were ordered to determine the cause of her mental status change. Her digoxin level was elevated at 2.4 ng/mL (normal range 0.5 to 2.0 ng/mL). Her dose of digoxin was reduced, and her symptoms of delirium resolved over several days without other intervention.

The definitive pathophysiology and etiology of delirium remain unclear, but evidence indicates that several sets of biologic factors appear to be involved. Major mechanisms leading to delirium include systemic inflammation (involving increased inflammatory cytokines like IL-1, IL-1β, and TNF), hypoxia, impaired blood flow, impaired tissue perfusion, and impaired metabolism. Pre-existing brain pathology and ongoing inflammation can damage vascular endothelium and the blood–brain barrier (BBB). The precise causal relationship between inflammation and delirium is not clearly understood. It is thought that inflammation affects the BBB to increase the permeability of inflammatory mediators into the brain, which affects neuronal cells (e.g., microglial cells, astrocytes) that produce both neuronal cell dysfunction and death. In addition, there is in vitro evidence suggesting that these neuronal cells can be primed by neurodegeneration to show an exaggerated inflammatory response which suggests a causal link with dementia syndromes. Vascular supply of glucose and oxygen can also lead to impaired function of brain structures. It is thought that delirium represents a reversible syndrome of neuronal dysregulation caused by disruption of oxidative metabolism. Certain brain structures, such as the cerebral cortex and hippocampus, are relatively more sensitive to this disruption than other parts of the brain. At the cellular level, disruption of oxidative metabolism results in an inability to maintain ionic gradients. In addition, there is abnormal synthesis, release, and metabolism of neurotransmitters and an inability to dispose of neurotoxins.

Given the altered arousal states present in different subtypes of delirium, it is widely assumed that the reticular activating system (RAS) pathway, which regulates alertness, is affected. There are neocortical and limbic inputs to this system that control attention. Many neurotransmitters have been implicated in affecting this pathway, as well as other brain areas, that can precipitate delirium, including acetylcholine, dopamine, histamine, norepinephrine, and GABA (gamma-aminobutyric acid). Cholinergic deficiency and dopamine excess have both been implicated in delirium. There is evidence suggesting that increased anticholinergic burden increases delirium risk, which is correlated with adverse cognitive effects of anticholinergic medications. Cholinergic, dopaminergic, and serotonergic neurons are especially sensitive to oxidative stress. In hypoxia, not only are derangements of these systems found, but there is also an increase in the release of glutamate. It has been hypothesized that the presence of dopamine is required for at least some glutamate-mediated neuronal injury in hypoxia. Increased dopaminergic neurotransmission is also believed to underlie the perceptual disturbances often seen in delirium, and antidopaminergic antipsychotic drugs can effectively treat these disturbances. Neurotransmitter disturbances of histamine and norepinephrine may also be involved in delirium pathogenesis. Histamine activates neuronal connections in the prefrontal cortex, limbic system, and basal ganglia that influence arousal and wakefulness. First-generation antihistamines (H1 and H2 receptor antagonists) have sedative effects that can lead to delirium. Regarding norepinephrine, both excess norepinephrine and diminished norepinephrine activity can impair frontal cortical function. During stress, the amygdala activates locus coeruleus firing triggering high noradrenergic activity that impairs attention. During inflammatory events such as trauma and sepsis, sympathetic nervous system activation results in high noradrenergic activity, which is associated with postoperative delirium in older adult patients. Increased noradrenergic drive is also present during alcohol withdrawal, which accounts for withdrawal symptoms like high blood pressure, tremor, and agitation. Delirium tremens is associated with GABA hypoactivity and hyperactivity of glutamate and norepinephrine. GABA hyperactivity is implicated in hepatic encephalopathy. Sedative hypnotics like benzodiazepines can produce cognitive impairment by enhancing GABA tone in corticothalamic networks, which has also been implicated in delirium pathogenesis.

DIAGNOSIS AND CLINICAL FEATURES

Delirium is a clinical diagnosis based on history, physical examination, and mental status examination. However, the first step is a suspicion of the disorder based on clinical observation. Because the features of delirium are quite variable and, by definition, tend to fluctuate over time, it is often overlooked or unrecognized by clinicians. Patients with delirium may seem so drowsy that they fall asleep while eating or speaking. At other times, the patient may seem inattentive to questions, with poor eye contact, easy distractibility by irrelevant stimuli, and requiring questions to be repeated because of difficulty sustaining focus on the conversation. These fluctuations tend to occur throughout a 24-hour period so that a patient might be relatively alert and lucid with appropriate orientation to the environment for a few hours during different parts of the day, alternating with periods of confusion, incoherence, or somnolence. This waxing and waning of orientation and level of awareness might not be evident to an observer who interacts with the patient during only one of these periods during the day. In the acute hospital or nursing home setting, where clinical staff change with each shift, it is important for physicians to be alert to changing descriptions of the patient's mental state and behavior across different shifts to specifically look for this pattern of symptom fluctuation. It is not uncommon to find that hospital staff who work evening shifts and day shifts have different views of the patient's level of cognitive and behavioral disturbances.

The speech of the patient with delirium is often incoherent, rambling, irrelevant to the topic, and disorganized. Specific language impairments include dysarthria (impaired speech articulation), dysnomia (impaired naming), and dysgraphia (impaired writing ability). The patient's ability to reason abstractly, retain information and direct their thinking is impaired. As a result, in the medical setting, patients with delirium may temporarily lose the capacity for informed consent. Frequently, patients who are delirious have sleep–wake cycle reversals and may wake during the night believing it is time to get washed and dressed. Disorientation to the environment is common, with patients mistaking the hospital for home and beginning to wander out of the hospital bed looking for the kitchen or a family member.

Other features commonly associated with delirium include transient delusional thoughts and emotional disturbances such as paranoia, fear, and disruptive behaviors. Delusional thinking is usually paranoid in nature, with patients becoming fearful that routine or specialized medical care is being done to them with a nefarious purpose. Perceptual disturbances due to delirium can be among the most frightening symptoms to patients and to their families. These problems vary in severity and frequently include misinterpretations of items and people in the patient's environment (e.g., belief that the physical therapist is a family member or mistaking the bang of a closing door for a gunshot). Illusions, which are misperceptions of actual stimuli, are also common in delirium. An example of an illusion might be a delirious patient misperceiving an intravenous line to be a snake. Hallucinations are perceptions in the absence of actual stimuli and are usually visual, but they may also be tactile or auditory in nature. The hallucinations which occur in delirium range in severity from simple colored spots or geometric figures to moving objects, such as white handkerchiefs falling, to more complex phenomena, such as animals, people, monsters, panoramic views, and scenes acted out. Typically, these hallucinations are felt as threatening to patients. It is easy to see how they can engender feelings of fear and anxiety.

Disruptive behaviors can be present more or less continuously, or they can occur unpredictably. These behaviors range from disorganized, purposeless actions such as picking at bandages or repetitive vocalizations to pulling out Foley catheters or attempting to get out of bed and ambulate without assistance. Distorted perceptual experiences can lead a patient to incorrectly interpret a medical situation as dangerous, leading them to attempt to pull out intravenous lines or endotracheal tubes or to respond aggressively to a nurse trying to assist the patient. While disruptive behaviors can occur at any time of the day, they are more common at night.

When delirium is suspected, the process of making the diagnosis includes a careful history, focusing on the timeline of symptom development. A key feature that distinguishes the impairments of delirium from those of dementia is the acute development of symptoms. Delirium can present almost immediately after head trauma, or it may develop over a period of hours or days in a hospitalized patient with pneumonia or other infection. An older adult outpatient with metabolic disturbances or medication toxicity might develop symptoms more gradually over a period of days to weeks. Unlike other cases of medical diagnosis, important aspects of the patient’s history usually come from individuals other than the patient themselves. Speaking with family members of the patient or other knowledgeable informants who can provide a clear description of the patient’s baseline level of cognitive functioning is crucial in establishing that an acute change has occurred. When delirium occurs in a hospitalized patient, it is important to review nursing notes to help establish the timeline of the abrupt onset of symptoms and the course of fluctuating symptoms. Other important aspects of the history include a careful review of all medications prescribed to the patient and the timing of when new medications were begun or when dosages were changed. A complete medication history review must include over-the-counter medications as well as prescription medications. In older patients, medications are the most common cause of delirium, with sedative-hypnotic, anticholinergic, opioid analgesic, and antidepressant medications being the most potentially problematic. A thorough history must also include an assessment of alcohol use, history of pain and its treatment, and any new physical symptoms.

The physical examination of the patient should include careful evaluation for possible underlying causes, with an assessment of vital signs, including oxygen saturation, and thorough cardiac and pulmonary examination. The neurologic examination should look for evidence of focal findings. Motor system abnormalities that can be seen in delirium include tremors, myoclonus, asterixis, increased deep tendon reflexes, and increased tone. Evidence of adrenergic system overactivity can be seen in delirium due to alcohol or sedative withdrawal and also with anticholinergic toxicity states. Signs of hyperadrenergia include tachycardia, hypertension, sweating, pallor, flushing, and pupil dilation.

The mental status examination of the patient begins with an assessment of the patient’s level of awareness and arousal. It should be noted whether the patient is somnolent or lethargic, and if so, how much stimulation is required to arouse the patient. The patient’s thinking and speech should be assessed, including whether speech is coherent or rambling and whether answers to questions are logic or disorganized. A number of simple tests can be used to evaluate the patient’s level of attention. Often it is evident through interviewing the patient that they are having difficulty sustaining focus on the conversation and/or is being easily distracted by other stimuli in the room. A common test known as digit span requires patients to repeat five digits forward and three digits backward. An alternative approach would be to ask patients to recite the days of the week or months of the year backward. In addition to impairment of awareness and attention, patients with delirium will have a change in at least one other area of cognitive functioning. This might include memory, learning, or disorientation. Perceptual disturbances, such as misinterpretation, illusions, and hallucinations, should be assessed. A bedside cognitive test can be used to screen patients who are alert enough to cooperate with formal testing. Commonly used bedside cognitive tests are the Mini-Mental State Examination (MMSE), Mini-Cog, and Montreal Cognitive Assessment (MoCA).

Several psychomotor phenotypes of delirium are described in DSM-5. Patients who have the hyperactive subtype of delirium exhibit hypervigilance, restlessness, distractibility, and agitation. Individuals with hyperactive delirium are easier to identify but may be more difficult to manage because of problematic behaviors, which can include refusing to cooperate with medical care, combativeness, and wandering. Patients with hypoactive delirium present primarily with withdrawal, lethargy, sparse speech, and decreased motor activity. Patients with mixed levels of activity may appear to have a normal level of psychomotor activity even though attention and awareness are impaired. This subtype also includes individuals who have a rapidly fluctuating level of psychomotor activity between hypoactive and hyperactive states.

A number of assessment instruments have been developed to diagnose delirium and evaluate its severity. A commonly used bedside tool is the Confusion Assessment Method or CAM. The CAM identifies key features of delirium and has been validated in many studies. Another quick tool that can be used is the 4A’s Test (4AT). Unlike the CAM, the 4AT does not require special training to administer. The CAM-ICU and the ICDSC (Intensive Care Delirium Screening Checklist) have both been validated for use in the ICU setting for critically ill patients. The Delirium Rating Scale-Revised-98 (DRS-R-98) has also been validated in both general medical populations and in geriatric populations. While an assessment tool does not take the place of a mental status examination, it can be very useful as a screening measure to help detect patients who are likely to be delirious. Daily use of a standardized screening instrument, such as the CAM, can be important for prompt identification of acute changes and fluctuations in cognition and attention in hospitalized older patients who are at risk of developing delirium. This can also be used to train nurses or other caregivers to recognize delirium.

PATHOLOGY AND LABORATORY EXAMINATION

The syndrome of delirium is diagnosed on the basis of the clinical history and the examination of the patient at the bedside. The diagnosis of delirium requires a prompt and careful laboratory evaluation to determine the etiology or etiologies. Delirium frequently heralds serious medical illness and, therefore, should be responded to as a medical emergency. A standard laboratory evaluation of delirium is shown in Table 55.3g–3. As indicated in Table 55.3g–3, if the basic evaluation fails to reveal a likely cause, then other selected laboratory investigations may be initiated. Depending on the particular case, further workup may include a brain imaging study or a lumbar puncture for cerebrospinal fluid studies. For instance, a lumbar puncture should be done if there is concern about possible CNS infection, autoimmune disorder, or subarachnoid hemorrhage. A brain computed tomography (CT) or magnetic resonance imaging (MRI) should be done if there is a high possibility of stroke or head trauma causing brain hemorrhage. Arterial blood gas evaluation should be done if there is a concern for hypercarbia in a patient with chronic obstructive pulmonary disease. Detailed neuropsychologic testing is not usually obtained because the attentional impairment of delirium compromises the validity of higher-order testing.

Electroencephalography (EEG) has a characteristic pattern in delirium and typically shows generalized slow-wave activity in the delta and theta range. In hyperadrenergic states, such as alcohol or sedative withdrawal, the EEG will show increased low-voltage fast (beta) activity. However, the EEG is rarely useful in determining the underlying cause of delirium and is not routinely performed unless there is a suspicion of seizure activity or nonconvulsive status epilepticus.

Table 55.3g–3.
Laboratory Workup for Delirium in the Older Adult Patient

- Basic evaluation
  - Blood tests
    - Complete blood count (CBC) with differential and platelets
    - Chemistries:
      - Albumin
      - Blood urea nitrogen (BUN)
      - Creatinine (Cr)
      - Creatine phosphokinase (CPK)
      - Electrolytes: Sodium (Na^+), potassium (K^+), chloride (Cl^-), bicarbonate (HCO_3^-), calcium (Ca^{2+}), phosphate (phos), magnesium (Mg^{2+})
      - Glucose
      - Liver function tests
    - Erythrocyte sedimentation rate or C-reactive protein
    - Syphilis serology
    - Medication levels
    - Toxicology screen (serum and urine)
- Urinalysis (with culture and sensitivity testing, if indicated)
- Electrocardiogram
- Chest radiograph

Other laboratory studies to consider, depending on clinical indications: brain imaging (computed tomography or magnetic resonance imaging), electroencephalogram, arterial blood gas, cerebrospinal fluid studies, vitamin B_{12} level, folate level, thyroid-stimulating hormone, heavy metal screening, human immunodeficiency virus testing, and serum anticholinergic activity (available through research laboratories).

DIFFERENTIAL DIAGNOSIS

Although the symptoms of delirium can mimic a number of psychiatric conditions, including psychosis, depression, or mania, in the older adult patient, the most important differential diagnosis is between delirium and dementia. Population studies of brain donors have shown that there is an increased risk of accelerated cognitive decline in patients with both delirium and pathologic dementia burden (neurofibrillary tangles, amyloid plaques, vascular lesions, and Lewy bodies) compared to delirium alone or dementia alone. To distinguish between delirium and dementia requires obtaining a thorough history from family members and caregivers or from nursing staff who have recently been caring for the patient and who can provide a description of the patient's baseline level of function and timeline of recent acute changes. Since dementia is a recognized risk factor for delirium, the two diagnoses often coexist in the same patient. Information from family members about the patient's premorbid functioning—such as when the patient last drove a car, balanced a checkbook, or dressed independently—can help identify a more slowly developing dementia. Description of a more rapid decline in general and cognitive functioning over days to weeks and/or a rapidly fluctuating course suggests delirium evolving in a patient with pre-existing dementia. For patients hospitalized on acute medical units, it is imperative to review reports from evening and night shift nursing staff in addition to day shift staff to look for fluctuations in attention, cognitive abilities, and behavior over the course of a 24-hour period. Table 55.3g–4 lists features that can help distinguish delirium from dementia.

Patients who are delirious can appear depressed or manic and frequently are anxious and psychotic. Patients with the hypoactive subtype of delirium tend to be withdrawn, minimally reactive, with minimal speech, poor oral intake, and flat affect, and can be mistaken as being depressed. While older adult patients who are depressed frequently exhibit impairment in short-term memory and in concentration, they do not experience fluctuations in awareness and rapid fluctuations in cognitive functioning that are characteristic of delirious patients. Conversely, patients with the hyperactive subtype of delirium may be agitated, incoherent, and psychotic and may be mistaken as manic. A patient who presents with an acute change in attention and cognition with fluctuating symptoms should be evaluated for delirium, whether or not the patient has a history of depression or bipolar illness, to prevent the possibility of not detecting a serious underlying medical disorder.

Table 55.3g–4.
Differential Diagnosis: Delirium versus Dementia

	Delirium	Dementia
Onset	Acute	Often insidious
Duration	Variable; usually brief	Chronic, unless reversible
Consciousness	Impaired, fluctuating	Normal
Orientation	Impaired	Variable
Memory	Impaired	Impaired
Attention	Impaired	May be normal
Thought	Disorganized; transient delusions; content-rich	Often stereotyped; content poor
Perception	Florid hallucinations (especially visual)	No hallucinations or simple hallucinations
Insight and judgment	Poor, but may return partially during lucid intervals	Variable, depending on etiology and stage
Mood and affect	Often anxious; may appear depressed	May be depressed or anxious
Sleep	Disturbed	May be normal
Electroencephalogram	Abnormal: generalized slowing and slow posterior dominant frequency	Normal or mild slowing of posterior dominant frequency

COURSE AND PROGNOSIS

An episode of delirium can be observed to progress through prodromal, overt, and resolving phases. During the prodromal period, there are typically symptoms such as anxiety, restlessness, irritability, mild confusion, difficulty concentrating, poor sleep, nightmares, daytime drowsiness, social withdrawal, increased sensitivity to light and noise, illusions, and hypnopompic and hypnagogic hallucinations. There is growing recognition of subsyndromal delirium, in which patients present with some, but not all, of the characteristic features of delirium. The prodrome may persist as subsyndromal delirium, or it may evolve into overt delirium over a period of days or weeks. Patients with subsyndromal delirium have outcomes intermediate between those with no symptoms of delirium and those who meet the full diagnostic criteria.

Delirium that develops in the context of severe medical illness, such as sepsis, may progress to stupor, coma, and death if appropriate, and prompt treatment of the illness is not administered. With appropriate treatment of the underlying medical illness and metabolic abnormalities, delirium will generally resolve. During the resolution phase, psychotic symptoms lessen, sleep cycles normalize, and clarity of awareness returns but cognitive deficits may be more persistent. The stage of resolving delirium in older adult patients may last from days to months. Persistent delirium is not uncommon and has been seen in 20% of older adult patients exhibiting residual symptoms for up to 6 months. Persistent delirium is associated with poor functional recovery from surgery and from medical illnesses and is also associated with increased length of hospital stays and with higher medical costs. Patients who are recovering from delirium will usually require additional assistance for activities of daily living and close monitoring of their medical condition and cognitive functioning.

In the absence of neurodegenerative disease or dementia, the signs and symptoms of delirium usually resolve completely. In patients with Alzheimer disease or Parkinson disease, cognitive deficits can remain such that patients function below their previous baseline levels. Recent research has indicated that patients with Alzheimer disease may experience more rapid cognitive decline following an episode of delirium. In addition, there is also growing evidence that in vulnerable older adults, delirium can lead to permanent cognitive impairment and increases incident dementia. The pathophysiologic mechanism underlying delirium-induced dementia is not clear but may involve processes other than those associated with Alzheimer disease. A serious medical illness that leads to delirium can also be thought of as a “stress test” for the brain that can unmask underlying cognitive deficits that were not obvious previously.

Most patients who have had a serious episode of delirium retain little or no recall of events they experienced while delirious. Patients who had mild-to-moderate episodes of delirium are more likely to have memories of their experience. Some patients develop residual anxiety or even a post-traumatic stress disorder (PTSD) as a consequence of the fearful emotions and frightening perceptual disturbances they experienced when they were delirious.

The presence of delirium is associated with a significant elevation in mortality risk for older adult patients, both during hospitalization and in the 6 to 12 months after discharge from the hospital. This association remains even after controlling for the severity of underlying illness and other confounding factors. Delirium symptoms that persist at discharge from the hospital are associated with a high rate of nursing home placement. Patients who have experienced an episode of delirium and have recovered remain at increased risk for a recurrent future episode of delirium.

PREVENTION

In recent years, there has been considerable interest in efforts to prevent delirium. Nonpharmacologic delirium prevention protocols have been shown to reduce the incidence of delirium. The most widely used strategy is a multicomponent nonpharmacologic approach devised by Sharon Inouye and colleagues, known as the Hospital Elder Life Program (HELP). This multidimensional strategy has been used for hospitalized patients with diverse medical problems. The HELP interventions include cognitive reorientation, therapeutic activities and early mobilization (such as daily exercise and walking), reduced use and reduced doses of psychoactive drugs, nonpharmacologic promotion of sleep, maintenance of oral hydration and adequate nutrition, and provision of appropriate hearing and vision aids.

Delirium prevention efforts have also focused on patients undergoing surgery as well as in the ICU since delirium is a common syndrome in these settings, especially in older adults. Studies have looked at patient baseline characteristics as well as intraoperative and postoperative management and their association with delirium. Evidence indicates that the route of intraoperative anesthesia (e.g., general, spinal, or epidural) does not affect the development of delirium. However, the total dose of anesthetic drugs may have a role in causing delirium. A double-blind, randomized study found that using lighter sedation for older adult patients undergoing hip fracture repair surgery led to a 50% drop in the risk of postoperative delirium. In addition, minimizing the use of benzodiazepines and certain opioid pain medications, especially meperidine, in the postoperative period can help decrease the incidence of delirium. Studies in critical care settings have shown early physical and occupational therapy during breaks in sedation in mechanically ventilated patients led to shorter delirium duration. Similarly, earlier mobilization of surgical patients led to improved functional status and decreased the number of days spent in the ICU. These evidence-based preventative interventions have been grouped into recommended care bundles targeted to optimize patient recovery and outcomes. One such care bundle used in the ICU, called the ABCDEF bundle, has been associated with decreased risk of delirium.

ABCDEF is an acronym that stands for

- Assess
- Both spontaneous awakening and breathing trials
- Choice of sedation/analgesia
- Delirium assessment/management
- Early mobility/exercise
- Family engagement/empowerment.

Multiple studies have looked at pharmacotherapy as a means to prevent delirium in high-risk patients in ICU. A recent systematic review on antipsychotic agents found limited evidence supporting the routine use of haloperidol or second-generation antipsychotic drugs (olanzapine, risperidone) for the prevention of delirium. Similarly, there is no clear evidence for the use of cholinesterase inhibitors or melatonin in delirium prevention in non-ICU settings.

TREATMENT

The most important aspect of treatment is to identify and treat underlying medical illnesses and derangements causing delirium. In many cases, timely and appropriate medical treatment will be lifesaving to the patient. When delirium is detected in older adult patients, it should be considered a medical emergency, resulting in immediate workup to determine the underlying cause or causes and to initiate treatment. In older patients, a single cause for delirium usually cannot be identified, as delirium is more commonly multifactorial in etiology. Most of the time, an older patient with delirium will be hospitalized in order to adequately evaluate the patient and address treatable causes.

Table 55.3g–5 lists critical medical interventions which should be used in managing older adult patients with delirium. Besides diagnosing underlying medical causes, attention should be given to all possible contributors to the delirious state, especially modifiable predisposing factors. These include the elimination of all unnecessary medications, treating hypoxemia and anemia, correcting vision and hearing impairments that lead to sensory deprivations, maintaining good hydration and nutrition, preventing urinary retention and fecal impaction, and minimizing pain. Reducing the use of indwelling urinary catheters can prevent catheter-associated urinary tract infections and associated delirium. Close monitoring is necessary for the patient’s safety and to prevent aspiration, falls, or inadvertent self-harm. Physical restraints should be avoided as much as possible. The use of physical restraints has been associated with the worsening of psychotic symptoms and with medical complications such as deep vein thrombosis, pulmonary embolism, and cardiac dysrhythmias in delirious patients.

Table 55.3g–5.
Medical Interventions for Delirium in Older Adult Patients

Ensure Safety	Close Observation by a Family Member or Staff May Be Indicated
Ensure adequate hydration and nourishment	Patient cannot monitor intake or feeding; this usually means intravenous hydration is indicated Eating and drinking should be observed to ensure adequate intake and to avoid choking/aspiration
Ensure adequate oxygenation	Consider supplemental oxygen via nasal cannula to keep O_2 saturation >95%
Attention to bowel/bladder function	Watch intake/output Ensure regular bowel movements via an appropriate bowel regimen
Investigate common etiologies of delirium	Perform the following: Chest radiograph to rule out pneumonia Urinalysis to rule out urinary tract infection Physical/lab exam to rule out dehydration/electrolyte disturbance Consider medication effects (not just toxicity, but ANY medication linked temporally to delirium onset) Hold nonessential medications
Treat pain adequately	Recommended drugs include scheduled acetaminophen for mild pain and low-dose hydromorphone for severe pain Avoid meperidine
Increase activity level as appropriate and tolerated	Get the patient out of bed to the chair as soon as possible Include a written plan in the chart to maintain mobility

Table 55.3g–6 lists environmental and nonpharmacologic interventions that have been shown to be beneficial in the management of patients with

delirium. These include taking steps to help restore a normal sleep–wake cycle by minimizing light, noise, and disruptions during the night while increasing exposure to light and activity during the day. Many patients benefit from being moved closer to a window, where natural daylight and night cues can be appreciated. It is important for staff and family to be educated about appropriate ways to interact with the patient to promote cognitive functioning. Delirious patients should not be left alone for long periods. Family members and close friends of the patient should be encouraged to visit and should be allowed to stay with the patient overnight if they so desire. Hospital staff and family members should reorient the patient to time, place, and person frequently during the day. Both overstimulation and understimulation should be avoided, as either extreme can exacerbate symptoms of delirium. The patient should be given only one task to do at a time rather than multistep instructions. For patients who are anxious or fearful, frequent calm reassurance by family and staff can minimize agitation.

There are no FDA-approved medications to treat behavioral disturbances in delirium. The first-line approach to managing agitation and psychotic symptoms in delirium is to try the nonpharmacologic interventions previously described, including reorientation, support, and reassurance. Currently, there is limited evidence on the use of medications in treating or preventing delirium. However, pharmacologic treatment may be necessary for agitated delirious patients for whom verbal reassurance is inadequate and for patients who exhibit behaviors that are dangerous to themselves or others. Pharmacologic treatment may also be indicated for patients with psychotic symptoms causing significant fear and anxiety.

Table 55.3g–6.
Environmental Interventions for Prevention and Management of Delirium in Older Adult Patients

Appropriate Lighting	**Bright During Day** **Dark at Night**
Promote sleep at night	Low nighttime noise level Pagers and other electronic devices on silent mode Telephone ring turned down Television turned off Voices kept low Doors closed as patient care permits Minimize necessary awakenings
Frequent reorientation	Staff to reorient patient to time, place, and person verbally each shift Clock and calendar in room Familiar possessions in room
Education	Family to be educated about the cause, course, and prognosis of delirium Staff to be educated about behavioral dysregulation in delirium and its management
Promote sensory function	Hearing aids on and function checked Glasses on and cleaned
Comfort/reassurance	Personal effects from home at bedside Family and close friends encouraged to visit and stay with patient, if possible Frequent reassurance provided to patient

In deciding which medication to use in treating behavioral symptoms of delirium, it is important to determine whether alcohol or sedative withdrawal is a cause of delirium. For patients experiencing alcohol or sedative withdrawal delirium, benzodiazepines are the medications of choice. Lorazepam or oxazepam may be preferable to chlordiazepoxide, especially in patients with decreased liver metabolism, because they are not metabolized by the liver and have shorter half-lives. It should be noted that benzodiazepines are not without adverse effects, especially in older adult individuals. These drugs can affect respiration and hemodynamics and can cause dysrhythmias. They can increase confusion and the risk of falls and, therefore, hip fractures. Paradoxical reactions, in which increased agitation and disinhibition are seen rather than sedation, are observed with greater frequency among patients with liver disease or low serum albumin levels. In the treatment of alcohol or sedative withdrawal, a common dosing guideline for lorazepam is as follows. The severity of withdrawal is monitored using a structured instrument such as the Clinical Institute Withdrawal Assessment for Alcoholism, revised (CIWA-Ar).

Clinical Institute Withdrawal Assessment for Alcoholism, revised (CIWA-Ar):

Day 1: Lorazepam 1 to 2 mg is given orally, intravenously, or intramuscularly every 4 hours on a scheduled basis.
Day 2: This dose is decreased by 50%.
Day 3: The dose is decreased again by 50% on day 3.

Score the CIWA-Ar every 4 hours. If the score increases to greater than 8 to 10, then the frequency of lorazepam doses is increased to every 1 to 2 hours until symptoms are controlled, then the original protocol is restarted. Patients with full-blown delirium tremens will likely require higher and more frequent doses.

A 78-year-old male was admitted to the emergency department after a 5-week alcohol binge. He was known to have a long history of alcohol dependence, with multiple prior episodes of delirium tremens and probably alcohol-related dementia. On admission, clinical examination and laboratory testing revealed malnutrition, urinary tract infection, and pneumonia. The patient was started on oral lorazepam but, within 24 hours, was transferred to the intensive care unit with tachycardia and visual and auditory hallucinations. A lorazepam infusion was started, along with a beta blocker for control of heart rate. Over the next 24 hours, the patient's condition stabilized. He was switched from the lorazepam infusion to intravenous bolus lorazepam with good control of symptoms. As his medical problems were treated, his mental state improved. The beta blocker was discontinued, and the lorazepam was tapered slowly over several weeks. He was discharged off medications 25 days after admission.

For delirious patients who are not experiencing alcohol or sedative withdrawal, benzodiazepines are avoided because they may exacerbate delirium and increase the risk of falls. Antipsychotic medications are the most studied class of drugs for delirium treatment. In non-ICU patients, the efficacy of these drugs is unclear, and review studies show no effect on delirium severity, symptom resolution, or mortality. Similarly, studies of patients in ICU indicate no significant treatment effects of antipsychotic medications on the duration of delirium, hypoactive/hyperactive delirium, or the number of days alive with/without coma. Other agents such as melatonin, ramelteon (a melatonin receptor agonist), and cholinesterase inhibitors have been suggested for delirium treatment, but there is limited consistent evidence supporting their use. However, expert consensus does support a limited role for drugs in the treatment of delirium, especially in patients with severe distress

For psychotic and behavioral symptoms, haloperidol, a typical antipsychotic, has been used due to its availability (it can be dosed orally and intravenously delivered) and, when dosed appropriately, has minimal adverse effects. It does not cause significant hypotension, is not anticholinergic, does not depress respiration, and does not cause or impair glucose regulation. The risk of QTc interval prolongation is minimized by staying within dosing guidelines and adhering to recommendations regarding QTc interval checks. Before haloperidol is initiated, a baseline QTc interval is obtained. If the interval is 450 ms or greater in a male or 470 ms or greater in a female, intravenous haloperidol should not be used. Any patient whose QTc interval is prolonged beyond 25% of baseline during treatment should have haloperidol discontinued. There is evidence that intravenous haloperidol is less likely to cause extrapyramidal symptoms than oral doses of haloperidol.

A conservative but usually effective dosing schedule for an older adult delirious patient is as follows: haloperidol 0.25 mg to 0.5 mg intravenously every 8 hours on a scheduled basis, and 0.25 mg to 0.5 mg intravenously every 6 hours as needed to keep agitation and psychotic symptoms under control. If oral dosing of haloperidol is used, higher doses are given (e.g., 0.5 mg to 1 mg by mouth every 8 hours and 0.5 mg to 1 mg by mouth every 6 hours as needed). Younger patients may require and tolerate higher doses of medication. How long the acute dose of haloperidol is maintained depends on how the patient responds clinically. If the patient becomes calm and psychotic symptoms lessen within the first 12 to 24 hours (as usually happens), this dose of haloperidol is continued for 2 to 3 days and then is tapered over the next 2 to 3 days. If the patient does not become calm within this time frame, consideration should be given to a missed or inadequately treated medical condition prolonging delirium. In this case, haloperidol would likely be continued for a longer period.

Atypical antipsychotic medications have been shown to be just as effective as haloperidol in controlling symptoms of agitation and psychosis in delirious patients. These medications are generally well tolerated when used for acute episodes of delirium. One limitation is that they are given orally rather than intravenously.

The two atypical antipsychotic agents which are most commonly used for delirium in older adult patients are risperidone and quetiapine, with dosing schedules as follows:

Risperidone 0.25 to 0.5 mg twice daily on a scheduled basis and 0.25 to 0.5 mg every 6 hours on an as-needed basis.
Quetiapine 25 mg at bedtime to 25 mg twice daily on a scheduled basis and 25 mg every 6 hours on an as-needed basis.

An 84-year-old man was admitted from home with acute coronary syndrome. Medical history was significant for coronary artery disease, myocardial infarction, left carotid endarterectomy, chronic renal insufficiency, and renal cell carcinoma. He underwent three-vessel coronary artery bypass grafting. On the first postoperative day, he was noted to be irritable, combative, and confused. He told his family that he was brought to the hospital for “experimentation” and had persistent tactile hallucinations of leaves and other objects in his mouth. Nighttime restlessness progressed to nocturnal insomnia and agitation. The patient had marked fluctuations in alertness during the day, with worsening agitation in the evening hours. During his agitated episodes, he pulled out his intravenous line several times. During a period of relative lucidity, an Mini-Mental State Examination (MMSE) was performed, and he scored 10/30. A psychiatric consultant recommended 0.5 mg of haloperidol orally every 8 hours on a scheduled basis for symptomatic treatment of his hallucinations and agitation. Medical complications (which included dehydration, atrial fibrillation, hypotension, and pneumonia) were managed as they arose. Although agitation and psychotic symptoms were improved, the patient required intermittent one-to-one observation by staff. After 7 days, his mental state began to improve as his medical condition stabilized. He was discharged on postoperative day 17 with some residual memory problems and insomnia but without evidence of psychotic symptoms or agitation.

FUTURE DIRECTIONS AND COVID CONSIDERATIONS

Despite progress over the years in understanding delirium, we still need more studies to elucidate the neuropathology of delirium, which can aid in preventing and treating delirium. *Animal models* could be helpful in understanding the underlying pathophysiology of delirium. Studies have looked at the role of interleukins, neurotransmitters, hypoxia, and brain pathology on delirium—but more studies are needed. An animal model could also be helpful in testing preventive or treatment strategies. Understanding pathophysiology may aid in devising pharmacologic targets. More studies of pharmacologic agents are needed to help with delirium prevention or treatment, including psychotropics that have limited evidence but continue to be used in multiple settings. Dexmedetomidine has also been looked at in the prevention or treatment of delirium in postoperative patients and should be explored further. Different approaches to sedation for surgery also are promising areas for investigation. In particular, using light sedation, where possible, seems to be less likely to cause delirium.

COVID-19, the coronavirus disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2), has also been implicated in causing high rates of delirium (up to 79.5% in ICU settings in the general population and up to 40% in non-ICU hospitalizations in older adults). There is limited evidence to elucidate the pathophysiologic role of COVID-19 in the development of delirium, but multiple mechanisms have been implicated, including systemic inflammation and/or direct brain entry of the virus. During the COVID-19 pandemic that arose in 2020, there were fears that best practices for delirium management would be compromised due to quarantine restriction requirements for inpatient care. Particularly early in the pandemic, shortages of personal protective equipment (PPE), changes to staff hierarchy and priorities, and restricted physical bedside presence affected adherence to delirium preventative protocols. These changes may have favored increased use of sedation and prolonged use of already limited ventilator resources. Early reports of COVID-19 cohorts suggested more widespread use of benzodiazepines for sedation and that COVID-19 patients often developed a severe delirium phenotype. At the present time, little is known about the impact of possible deviations from usual delirium management protocols on the course of COVID-19 delirium in older adults.

Clearly, more research is needed to fully understand the impact of the virus on the geriatric population.

Another area requiring future research is evaluating methods of improving the quality of life following delirium, as delirium is associated with functional decline and worsening disability in older adults. A collaborative approach to delirium prevention needs to be advocated with a focus on delirium prevention/treatment protocols. Strategies like the ABCDEF bundle and HELP protocol have shown benefits in decreasing delirium, but more studies of interventions that can help prevent or treat delirium are needed. Improved approaches to the prevention and treatment of delirium are needed to limit impairments and adverse outcomes on the physical, cognitive, and psychological functioning of older adults.

FURTHER READINGS

Chazon PL, Johnston L, McAuley E, Bonner S. Histamine and delirium: current opinion. *Front Pharmacol.* 2019;10:299.

Cole MG, Ciampi A, Belzile E, Zhoung L. Persistent delirium in older hospital patients: a systematic review of frequency and prognosis. *Age Ageing*. 2009;38:19–26.

Cole MG, McCusker J, Ciampi A, Belzile E. The 6- and 12-month outcomes of older medical inpatients who recover from subsyndromal delirium. *J Am Geriatr Soc*. 2008;56:2093–2099.

Davis DHJ, Muniz-Terrera G, Keage HAD, et al; Epidemiological Clinicopathological Studies in Europe (EClipSE) Collaborative Members. Association of delirium with cognitive decline in late life: a neuropathological study of 3 population-based cohort studies. *JAMA Psychiatry*. 2017;74(3):244–251.

Davis DHJ, Muniz-Terrera G, Keage H, et al. Delirium is a strong risk factor for dementia in the oldest-old: a population-based cohort study. *Brain*. 2012;135:2809–2816.

de Rooij SE, van Munster BC, de Jonghe A. Melatonin prophylaxis in delirium: panacea or paradigm shift? *JAMA Psychiatry*. 2014;71(4):364–365.

Ely EW, Shintani A, Truman B, et al. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. *JAMA*. 2004;291:1753–1762.

Grover S, Kumar V, Chakrabarti S. Comparative efficacy study of haloperidol, olanzapine and risperidone in delirium. *J Psychosom Res*. 2011;71:277–281.

Hawkins M, Sockalingam S, Bonato S, et al. A rapid review of the pathoetiology, presentation, and management of delirium in adults with COVID-19. *J Psychosom Res.* 2021;141:110350.

*Inouye SK, Bogardus ST Jr, Charpentier PA, et al. A multicomponent intervention to prevent delirium in hospitalized older patients. *N Engl J Med*. 1999;340:669–676.

*Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med*. 1990;113:941–948.

Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. *Lancet*. 2014;383:911–922.

Laurila JV, Laakkonen ML, Tilvis RS, Pitkala KH. Prediposing and precipitating factors for delirium in a frail geriatric population. *J Psychosom Res*. 2008;65:249–254.

Lipowski ZJ. *Delirium: Acute Confusional States*. Oxford; 1990.

Marcantonio ER. Postoperative delirium: a 76-year-old woman with delirium following surgery. *JAMA*. 2012;308:73–81.

McAvay GJ, Van Ness PH, Bogardus ST Jr, et al. Older adults discharged from the hospital with delirium: 1-year outcomes. *J Am Geriatr Soc*. 2006;54:1245–1250.

Meagher DJ, Moran M, Raju B, et al. Motor symptoms in 100 patients with delirium versus control subjects: comparison of subtyping methods. *Psychosomatics*. 2008;49:300–308.

Nadelson MR, Sanders RD, Avidan MS. Perioperative cognitive trajectory in adults. *Br J Anaesth*. 2014;112:440–451.

Oh ES, Needham DM, Nikooie R, et al. Antipsychotics for preventing delirium in hospitalized adults: A Systematic Review. *Ann Intern Med*. 2019;171(7):474–484.

Saczynski JS, Kosar CM, Xu G, et al. A tale of two methods: chart and interview methods for identifying delirium. *J Am Geriatr Soc*. 2014;62:518–524.

*Schuckit MA. Recognition and management of withdrawal delirium (delirium tremens). *N Engl J Med*. 2014;371:2109–2113.

Shehabi Y, Bellomo R, Reade MC, et al. Early goal-directed sedation versus standard sedation in mechanically ventilated critically ill patients: a pilot study. *Crit Care Med*. 2013;41(8):1983–1991.

Siddiqi N, Young J, Cheater FM, Harding RA. Educating staff working in long-term care about delirium: the Trojan horse for improving quality of care? *J Psychosom Res*. 2008;65:261–266.

Sieber FE, Zakriya KJ, Gottschalk A, et al. Sedation depth during spinal anesthesia and the development of postoperative delirium in elderly patients undergoing hip fracture repair. *Mayo Clin Proc*. 2010;85(1):18–26.

Strijbos MJ, Steunenberg B, van der Mast RC, Inouye SK, Schuurmans MJ. Design and method of the hospital elder life program (HELP), a multicomponent targeted intervention to prevent delirium in hospitalized older patients: efficacy and cost effectiveness in Dutch health care. *BMC Geriatr*. 2013;13:78.

The ASAM Clinical Practice Guideline on Alcohol Withdrawal Management. *J Addict Med.* 2020;14(3S):1–72.

Trzepacz PT, Mittal D, Torres R, Kanary K, Norton J, Jimerson N. Validation of the Delirium Rating Scale-revised-98: comparison with the delirium rating scale and the cognitive test for delirium. *J Neuropsychiatry Clin Neurosci*. 2001;13:229–242.

Wilson JE, Mart MF, Cunningham C, et al. Delirium. *Nat Rev Dis Primers*. 2020;6(1):90.

Xia ZQ, Chen SQ, Yao X, Xie CB, Wen SH, Liu KX. Clinical benefits of dexmedetomidine versus propofol in adult intensive care unit patients: a meta-analysis of randomized clinical trials. *J Surg Res*. 2013;185:833–843.

Young J, Leentjens AF, George J, Olofsson B, Gustafson Y. Systematic approaches to the prevention and management of patients with delirium. *J Psychosom Res*. 2008;65:267–272.

55.3h Schizophrenia and Other Psychotic Disorders

Carl I. Cohen, M.D., and Paul D. Meesters, M.D.

Over the past century, the vast majority of schizophrenia studies have primarily involved young adult patients, and only about 1% of the literature on schizophrenia is dedicated to issues of older adults. In this chapter, we examine early-onset and late-onset schizophrenia (LOS) in older adults and how they can be distinguished from other psychotic disorders in later life.

HISTORY AND COMPARATIVE NOSOLOGY

Emil Kraepelin’s description of “dementia praecox” as a disorder of progressive decline with onset during adolescence or early adulthood greatly influenced the subsequent conceptualization of schizophrenia. Nonetheless, Kraepelin later questioned the term *dementia praecox* after observing that the onset of this disorder was not limited to younger patients, nor was the course necessarily progressive. He used the term *paraphrenia* to describe a subset of patients with an onset of delusions and hallucinations that were primarily paranoid and with a better prognosis. This group of patients had several symptoms like those of dementia praecox but with less alteration of emotion and volition. Kraepelin was said to refer to paraphrenia as the “darkest area of psychiatry.”

Approximately 20% to 25% of patients with schizophrenia have an onset of the disorder after age 40. There has been considerable vacillation in diagnostic criteria regarding the age of onset. Martin Roth and colleagues introduced the idea that LOS could be a distinct entity after applying the term “late paraphrenia” to a group of patients with an onset of the disorder after the age of 60. Years later, the term “late paraphrenia” was included in the ninth edition of the International Statistical Classification of Diseases and Related Health Problems (ICD-9). The use of the terms “paraphrenia” and “late paraphrenia” led to some confusion because Kraepelin did not consider late age of onset a feature of paraphrenia, and paraphrenia is a more inclusive term and not synonymous with schizophrenia.

The early editions of the Diagnostic and Statistical Manual of Mental Disorders did not specify the age of onset criteria for schizophrenia. DSM-I and DSM-II included the category “Involutional Psychotic Reaction” for psychotic symptoms occurring during the involutional period, and DSM-II used the terms “involutional state” or “involutional paraphrenia” to describe a delusional condition with onset during the involutional period. In contradistinction, the DSM-III restricted the diagnosis of schizophrenia to onset before age 45, whereas the DSM-III-R included a late-onset category for initial presentations after age 44. The ICD-9 included the diagnosis of “Paraphrenia” and listed “Involutional Paranoid State” and “Paraphrenia, Involutional” as approximate synonyms. However, the ICD-10, DSM-IV, DSM-IV-TR, DSM-5, and the DSM-5-TR did not include separate diagnostic categories or criteria for LOS.

The omission of a late-life category is problematic since there is evidence that late-onset disorders, especially those arising after age 60, may differ clinically from earlier-onset cases. In 2000, an international expert group published a consensus report that proposed that schizophrenia be termed “Late-Onset Schizophrenia” and “Very–Late-Onset Schizophrenia-Like Psychosis” for disorders that have an onset between ages 40 and 60 and after the age of 60, respectively.

EPIDEMIOLOGY

Epidemiologic studies from the United States and the United Kingdom place the prevalence of schizophrenia at 0.6% to 1% in persons aged 45 and over and 0.1% to 0.5% in those aged 65 and over. Based on insurance claims data of persons aged 65 years and older, the prevalence estimates were 0.55% for men and 0.84% for women, with an overall prevalence of 0.71%. In the United States, by 2025, roughly three-quarters of a million people, or approximately one-fourth to one-third of all persons with schizophrenia, will be aged 55 and over; by 2060, their proportion may reach 40%. Worldwide, the number of people aged 60 years and older with schizophrenia will expand to an estimated 10 million people by 2050.

EARLY-ONSET SCHIZOPHRENIA

Schizophrenia typically develops in the second or third decade of life. Despite lower life expectancies than their age peers, increasingly more persons with schizophrenia are surviving into later life. It is estimated that nearly two-thirds of persons with schizophrenia aged 60 and over have an early-onset disorder.

Course and Outcome of Early-Onset Schizophrenia

There are several key points about course and outcome that have emerged in recent years that are relevant to the care of older adults with schizophrenia:

First, determining the outcome of schizophrenia can be a daunting task since it is dependent on definitions of the outcome, historical context, methods of assessment, diagnostic criteria, geographical settings, sociodemographic differences, age of onset, responsiveness to medications, and length of hospitalization Thus, Hegarty and colleagues looked at longitudinal studies of outcome in schizophrenia (all age groups) across the 20th century and found a mean improvement rate of 28% during the first quarter of the century, 49% in mid-century, and 36% in the last two decades.

Second, there is an increasing consensus that outcomes in schizophrenia are multidimensional and encompass clinical and social elements. A comprehensive view of outcomes in older adults should include symptoms, functionality, a combination of symptoms and functionality (“clinical recovery”), and positive health (“successful aging”) based on physical, social, and psychological well-being. The symptom outcomes include positive and negative symptoms (sometimes combined into “clinical remission”), depression, and cognitive functioning. The social functioning outcomes include community or social integration, adaptive functioning (e.g., activities of daily living), and quality of life (QoL).

Third, there is substantial heterogeneity in the outcomes of schizophrenia in later life. Despite scarce data, some findings suggest that the diversity seen in younger individuals might be even greater in older age. Long-term studies observing the symptoms and functioning in EOS suggest that the course of schizophrenia may not be as pessimistic as previously thought, and findings tend to challenge the notion implied in the term “dementia praecox.” Ten studies of hospitalized patients who were reinterviewed after hospital discharge (mean of 22 to 37 years postdischarge) conducted 30 to 50 years ago found recovery rates ranging from 12% to 55% based on a single assessment in later life. The investigators used varying diagnostic criteria and social indicators. Several cross-sectional studies conducted during the early part of this century found clinical remission rates ranging from 29% to 49% and clinical recovery rates of 22% in older adults with schizophrenia, nearly all of whom were living in community settings. However, longitudinal studies were less optimistic, with only about a quarter of persons displaying persistent clinical remission and about one-eighth attaining persistent clinical recovery. The latter was consistent with findings in young and mixed-age samples that found a median rate of clinical recovery of 13.5%.

Fourth, outcome measures are relatively distinct from each other and may need to be viewed as separate targets for intervention. A study in New York City of community-dwelling older adults with schizophrenia aged 55 and over who had early-onset disorder found that the shared variance among clinical remission, depressive symptoms, cognitive functioning, and community integration ranged from 0% to 22%, with a median value of 5%. The relative independence among outcomes suggests a variegated pattern of outcome and, consequently, the need for more targeted, personalized care. Indeed, when the investigators subsequently looked at five outcome measures over 52 months (remission, community integration, depression, cognitive functioning, and self-health), they found various combinations of favorable outcomes, as well as differences in the number of favorable outcomes, attained: 9%, 17%, 22%, 24%, 22%, and 6% had favorable outcomes in none, one, two, three, four, or all five outcome measures, respectively.

Fifth, although schizophrenia has been thought to be largely quiescent or have a steady end-stage in later life, recent longitudinal data shows this is not true. About two-fifths of persons fluctuate between being in and out of clinical remission, and between one-fourth and one-third of persons fluctuate between varying states of clinical recovery, depression, cognitive dysfunction, and community integration.

The final point is that the course in later life can be understood within the recovery model of mental illness. As noted above, while most persons do not achieve persistent clinical recovery, many display fluctuations in their symptoms and functioning. Based on longitudinal data from the NYC community study described above, Cohen and Reinhardt devised a five-tier taxonomy of late-life course trajectories among persons living in the community. On the two extremes, 12% showed persistent recovery, and 18% were unable to ever attain symptom remission or social integration. The remainder showed various combinations of remission and social integration. Hence, most older persons are in various states of recovery (“recovering”), and targeted approaches to address their specific clinical and social needs may advance their progress along the recovery continuum. Although the extremes of this continuum attract the most attention, most older patients with schizophrenia can be identified at some point in between these extremes, moving with time in one direction or the other.

While a prediction of the long-term course of the illness in an individual cannot be accurately made, the general factors associated with favorable outcomes include female gender, later onset of symptoms, less severe negative symptoms, and better premorbid functioning. However, with age, the strength of association between these variables and outcome weakens. Consequently, predicting a probable long-term course for schizophrenia patients is difficult. Apart from the relatively small group of poor outcome patients who worsen over time and who are largely found among long-term institutionalized patients in state hospitals or nursing homes (about 15% of all older adults with schizophrenia), most older community-dwelling persons with schizophrenia have been found to have a nondeteriorating course accompanied by fluctuations over time in individual outcome dimensions.

Symptom Dimensions of Outcome. In the previous section, we discussed some of the findings of hybrid or global measures of outcome, such as “clinical remission,” which combines various positive and negative symptoms, and “clinical recovery,” which combines remission and community integration. Below, we focus on key individual measures that comprise the dimensional framework in Section III of the DSM-5-TR. The DSM-5-TR authors contend that the dimensional assessment may assist with treatment planning, prognostic decision-making, and research on pathophysiology. In time, the dimensional model is expected to gain more traction in the care and research affecting older adults with schizophrenia.

POSITIVE SYMPTOMS. Long-term studies of outcomes have demonstrated that there is an improvement in positive symptoms with age. In the NYC community sample described above, roughly three-fifths of patients living in the community demonstrated favorable outcomes for positive symptoms. On 52-month follow-up, nearly half maintained persistently low or absent positive symptoms, and about one-third fluctuated between the presence and absence of positive symptoms.

NEGATIVE SYMPTOMS. Negative symptoms are thought to predominate the clinical picture of schizophrenia in later life. However, estimating negative symptoms in older schizophrenia patients can be challenging, as the clinical picture is often confounded by depression, demoralization, cognitive decline, medication effects, and poverty. The few studies of negative symptoms in community-dwelling older adults found prevalence rates of between one-quarter to one-half of the sample. Notably, longitudinal data from the New York City sample found that slightly over two-fifths of people fluctuated between the presence and absence of negative symptoms. Thus, contrary to earlier notions of schizophrenia progressing to a “burnt-out” state in later life, negative symptoms do not typically dominate the clinical picture, appear to be no higher than in younger schizophrenia populations, and fluctuate more than positive symptoms over time.

COGNITION. It is well established that older adults with schizophrenia have more cognitive deficits than their age-matched peers, with levels of cognitive deficits being nearly two standard deviations below control participants. Rajji and coauthors found that compared to healthy controls, persons with schizophrenia, aged 19 to 79, had lower scores in overall cognition and each cognitive domain except for social cognition. However, the rate of decline across the age groups was similar to that observed in the healthy control groups. Consequently, schizophrenia is no longer viewed as a neurodegenerative disorder (“dementia praecox”) but rather as a neurodevelopmental disorder in which deficits arise around the time of onset and do not progress beyond the expected effect of aging. While the rate of cognitive decline in schizophrenia may be similar to age-matched controls, the pre-existing deficits accrued at disease onset result in an earlier crossing of the diagnostic threshold into a major neurocognitive impairment (dementia).

Within-group longitudinal data of middle age and older adults with schizophrenia in the community showed considerable heterogeneity: a majority of persons seemed to be cognitively stable; however, about one-fifth of persons showed a more rapid decline, and one-fifth showed improvement, often rising from dementia to near-normal levels. Worse cognition (global and/or domain-specific) is strongly associated with diminished functional competence and community adaptation; more modest correlations have been found with older age, female gender, lower education, living in an institution, longer duration of illness, earlier age of onset, and higher levels of positive and negative symptoms.

Several recent epidemiologic studies have found prevalence rates of dementia diagnosis in persons with schizophrenia of 2 to 20 times that of their age peers, although the actual prevalence rates ranged from 1.8% to 64%, with rates of 70% at age 80 in one study. Friedman and coinvestigators found accelerated cognitive decline experienced by chronically institutionalized persons with schizophrenia beginning at age 65, but the decline in cognition was substantially less than in Alzheimer disease (AD). Patterns of decline, neuropathology, and polygenetic findings did not resemble those of AD, even among more chronically hospitalized groups. Rajji and associates’ review of older schizophrenia patients concluded that the findings of cognitive decline at age 65 found in institutionalized persons do not seem to apply to those who have been living in the community.

The fact that the course of decline and the underlying neuropathology differs from AD and that as many as one-fifth of persons show substantial improvements in cognition over time suggests that rendering a diagnosis of a major neurocognitive disorder may be inappropriate. As in the case of depression described below, the employment of the dimensional perspective described in the DSM-5-TR may help resolve these thorny diagnostic issues.

DEPRESSION. Among the studies on depression in older adults with schizophrenia, prevalence rates of depression ranged from 11% to 75%, depending on the severity criteria that were used. At all levels of depression, the prevalence was significantly greater than their age peers in the general population, although rates were similar to those of younger persons with schizophrenia. Because there is a substantial minority of persons who do not experience depression, it may be inappropriate to consider depression to be a core feature of the disorder, although it is a common feature with prevalence rates that remain high throughout the lifespan. Longitudinal data from the New York City sample showed that persistent depression (defined as the presence of syndromal or subsyndromal depression) occurred in 44% of persons, 30% remained persistently nondepressed, and 26% fluctuated between depression and nondepression.

The potential for suicide remains high in later life. Among depressed persons aged 40 and over with schizophrenia, 23% had current suicidal ideation, and 49% had made a previous attempt. In another study that included depressed and nondepressed persons with schizophrenia aged 55 and over, 13% had current suicidal ideation, and 30% had made a suicidal attempt during their lifetime.

Social and Adaptive Functioning

COMMUNITY INTEGRATION AND SOCIAL FUNCTIONING. Social functioning is a critical outcome indicator. Social functioning refers to the ability to establish and maintain relationships with friends and family as well as to undertake work and leisure activities. Adaptive functioning refers to the ability to function independently. A review by Meesters and colleagues concluded that the majority of older schizophrenic patients are well behind their healthy age-peers in various aspects of social and adaptive functioning, such as social roles, social support, and social skills. At the same time, considerable heterogeneity among patients can be found, and cognitive functioning had a major impact on social functioning, outweighing clinical symptoms. Other factors associated with better adaptive functioning include fewer negative symptoms, better physical health, and independent living. Long-term studies indicate that social functioning often improves in older schizophrenia patients, although, as with other domains, there may be movement in favorable and unfavorable directions over time.

A more recent body of work has focused on community integration in schizophrenia. Community Integration can be conceptualized as a "normalization" of functioning in society and comprises assessments of functioning and perceived QoL within several dimensions, viz., independence, physical, psychological, and social integration. In a cross-sectional study in New York City, a community comparison group of older adults had roughly twice the level of high community integration than those with schizophrenia (41% vs. 23%). In a longitudinal study using the same sample, 34% of schizophrenia persons fluctuated between high and low community integration, suggesting that there is a movement toward improved social integration in later life but also the potential for increased social isolation and diminished functioning.

QUALITY OF LIFE. QoL indicates an individual's self-perception of their position in life concerning their goals, expectations, and worries. It is an indicator of individuals' self-appraisal of their condition. Among older individuals with schizophrenia, lower QoL is associated with depression, positive and negative symptoms, cognitive deficits, financial strain, poor social support, and impaired social skills. Lower QoL may also predict suicidality. Although QoL indices are lower in older community-dwelling schizophrenia patients than their age peers matched for race and income, the absolute difference between the groups is small, and both scored in the moderate range of QoL and substantially higher than persons with chronic pain. Like other outcome measures, there is a fluctuation between good QoL and poor QoL. The New York City group found that 28% of persons moved between favorable and less favorable QoL outcomes, 40% had persistently favorable QoL, and only 32% had persistently less favorable QoL. Over time, many older adults with schizophrenia appear to find ways to adapt to their illness and, despite ongoing symptoms, value their QoL and find "wellness within illness."

In summary, the research data on the symptom and social outcome dimensions of schizophrenia in older patients provide some support for optimism in later life. However, based on recent longitudinal studies, we have a more nuanced understanding of the dynamics of various domains. Thus, as noted above, a favorable outcome in one category does not necessarily result in a favorable outcome in another category. This should lead to a more personalized care approach to treatment that will include a combination of biologic and psychosocial methods.

Medical Comorbidity and Mortality

Life expectancy is reduced by 15 to 20 years for people with schizophrenia. The mortality rate in schizophrenia is two to three times higher than in the general population, and this mortality gap has increased in recent decades. Although persons with EOS who live into old age should be considered to be survivors, the differences in mortality rates between persons with schizophrenia and their age peers persist into later life, perhaps even into their 80s. On the other hand, a review by Copeli and Cohen found that older adults with schizophrenia consistently showed lower age-adjusted mortality rates compared to middle-aged and younger persons with schizophrenia. While the mortality rate from suicides and accidents is elevated in older individuals with schizophrenia versus their age peers, natural causes of death account for the largest part of the reduction in life expectancy.

It is unclear as to why there are differences in mortality rates between older adults with schizophrenia and their age peers. Copeli and Cohen's review found inconsistent findings regarding differences in the prevalence of serious medical disorders between them. For example, compared to their age peers, older adults with schizophrenia had increased mortality rates and greater prevalence of respiratory diseases; no differences in mortality and prevalence of diabetes; higher mortality rates but no differences in the prevalence of cardiovascular disease; and considerable variability with respect to cancer, with incidence rates being equal or perhaps lower, but mortality rates being higher. Some of the differences in mortality can be attributed to lifestyle factors (e.g., poor diets, smoking, diminished physical activity), being in lower socioeconomic classes, and the impact of antipsychotic medications, especially second-generation antipsychotics. Some of the excess morbidity and mortality may be due to high rates of inflammation and oxidative stress reported among persons with schizophrenia. Moreover, they seem to be at risk for undetected or inadequately treated medical conditions that may be secondary to individual or systemic factors. A Danish Registry study found that persons with schizophrenia aged 70 and over did not differ from controls regarding chronic medical illnesses, but they were significantly less likely to receive medication for cardiovascular diseases; hospital admissions and the number of days hospitalized were equal to controls, but with significantly fewer general medical outpatient contacts. Finally, diminished insight regarding their symptoms and illness management may contribute to inadequate healthcare.

Biologic Features

Numerous neuroimaging cross-sectional studies support structural brain alterations in schizophrenia patients and controls. These include diminished gray matter in frontal, temporal, and postcentral cortical regions and increased ventricular size, and white matter reductions in the corpus callosum, although these findings have been inconsistent. Schnack and coinvestigators compared MRI scans for healthy subjects and persons with schizophrenia (age range 16 to 67 years) and found that there was an “accelerated” pattern of brain aging in people with schizophrenia at illness onset and during the initial 5 years; after that time, the rates of decline were comparable between the two groups, although the brain gap stayed constant.

Concerning neuropathology studies, a review by Centorino and Schultz concluded that postmortem studies have been even more mixed and complex than the neuroimaging studies, and no clear neuropathologic lesions have been demonstrated in older populations. Nguyen and colleagues’ review found that a majority of studies reported abnormal biomarker levels among older individuals with schizophrenia, including senescence-related indices of inflammation, telomere length, oxidative stress, metabolic health, gene expression, synaptic function, and testosterone, suggesting that schizophrenia is associated with aging-related biochemical abnormalities. About one-fourth of these studies observed age-related changes in schizophrenia, especially in markers of receptor/synaptic function and gene expression. Schizophrenia patients with greater disease severity and longer illness duration had higher levels of inflammatory and oxidative stress biomarkers and shorter telomere lengths. These studies must be interpreted cautiously because of inconsistencies in study designs, especially in the failure to control for potentially confounding variables such as smoking, substance abuse, exercise, diet, access to health care, and use of antipsychotic agents.

Treatment

There are only a few controlled studies assessing treatment approaches in older schizophrenia patients. Optimal treatment for older patients with schizophrenia will likely involve a targeted combination of pharmacology, nonpharmacologic therapy, and community-based approaches.

Somatic Approaches. Reviews of various open studies, along with the few controlled studies in older adults with schizophrenia, found somewhat mixed results for first-generation antipsychotics but more consistently favorable results with second-generation antipsychotics. Khan and colleagues provided an update on an earlier consensus paper and recommended risperidone (2 to 3 mg/day) as the first-line treatment for late-life schizophrenia; the second-line treatments are quetiapine (100 to 200 mg/day), olanzapine (5 to 15 mg/day), and aripiprazole (10 to 15 mg/day). For those with a late-onset disorder, the starting dose for antipsychotics should be ¼ of the typical adult dose, and the maintenance dose should typically be ⅓ to ½ of the adult dose.

Several studies found that estrogen treatment in postmenopausal women with schizophrenia resulted in diminished symptom severity. One controlled study suggested that citalopram may be effective for treating middle-aged and older schizophrenia adults with subsyndromal depression.

A literature review of electroconvulsive therapy (ECT) in older persons with schizophrenia identified six prospective studies and one case series; participants were between 49 and 96 years old. The authors concluded that bilateral ECT was effective as an acute and maintenance treatment in this population that had the onset of their illness earlier in life. ECT was not effective in those with very–late-onset psychosis, possibly due to the high prevalence of comorbid brain pathology other than schizophrenia. Transcranial magnetic stimulation and deep brain stimulation have not been studied systematically in the older age group.

The long-term use of antipsychotic medications has raised numerous concerns. Because of the pharmacokinetic and pharmacodynamic changes associated with aging, older people are more prone to the adverse effects of medications, for example, cardiovascular, orthostasis, metabolic, anticholinergic, extrapyramidal, sedation, falls, and neuroleptic malignant syndrome. In a middle-aged sample, higher lifetime antipsychotic dose years were associated with poorer cognitive performance. Gray matter volume decreases on MRI in people with schizophrenia are correlated with cumulative exposure to antipsychotic treatments, while no effects were observed for the duration of illness or illness severity.

Some older adults with schizophrenia may need little or no medication, either because symptoms have abated or because medication has been ineffective. Consequently, several writers have proposed trials to gradually taper or discontinue medications. Graff-Guerrero and coauthors' study of persons aged 50 and over found antipsychotic dose reduction was feasible in patients with stable illness and was accompanied by diminished adverse effects and improved illness severity measures. This may have been due to the lower therapeutic window of D2/3R occupancy that was needed in older persons (50% to 60%) than what had been reported in younger patients (65% to 80%).

Nonpharmacologic Approaches. Because many persons with schizophrenia, especially early-onset cases, do not make full recoveries with medication alone, a variety of nonpharmacologic strategies have been devised specifically for this population. The first generation of nonpharmacologic approaches focused on delivering group-based psychosocial skills training that enhances social cognition and perception, verbal communication, social skills, and interpersonal problem-solving skills. Random control studies found significantly favorable outcomes for several programs that have targeted middle-aged and older adults. The second generation of model programs focused on illness self-management and wellness techniques. Self-management strategies train older patients to assume greater responsibility for their physical health as well as improve coping abilities to deal with psychiatric stress.

There are emerging technologies that will enable older adults with schizophrenia to improve their health, age in place, and receive age-friendly care. These technologies encompass a variety of modalities, such as humanoid care robots, smart home systems and devices, personal devices, and software that enable telehealth connectivity. In-home sensing and monitoring devices address everything from medical needs such as fall detection and medication adherence to personal or “quantified self” measurements targeted to calorie and step counting, sleep quality, and air quality.

Community-Based Approaches. In older adult patients with chronic mental illnesses, Assertive Community Treatment (ACT) teams have engaged patients in treatment more successfully and had fewer dropouts from treatment versus comparison groups. Several model programs have combined collaborative care and case management strategies to address the medical and psychiatric issues facing many patients, as well as facilitate communication between the mental health team and the primary care providers. Collaborative care integrates mental health, primary care, and pertinent community supports and has the potential to ease the accessibility issues faced by older adults and to foster cross-discipline knowledge transfer to address complex clinical situations such as severe depression or suicidality. Case manager strategies typically use persons with clinical backgrounds to provide guidance, counseling, and skills training to older clients concerning psychiatric and physical health. They can also assist in guiding more appropriate care by primary care providers and enhance interdisciplinary communication.

LATE-ONSET SCHIZOPHRENIA

Epidemiology

A review of schizophrenia in later life found that slightly less than one-quarter of patients with schizophrenia had an onset of the disorder after age 40, with 13% in the fifth decade of life, 7% in the sixth decade, and 3% in later decades. A Dutch case-register study reported a 1-year prevalence rate of 0.55% for schizophrenia in persons over age 59 years, of which 64% had EOS and 35% had LOS. These figures suggest that the loss of early-onset individuals (to death or recovery) may be partially compensated by the influx of cases with a later onset. The epidemiologic study of LOS is complicated by several factors, including the use of various definitions of “late onset,” uncertainty in the determination of the age of onset, and questions regarding a correct diagnosis.

LOS affects women between 2 and 10 times more often than men. The robustness of this finding, coupled with the higher incidence rates of EOS in men, has led to the hypothesis that estrogen-mediated dopaminergic inhibition may protect younger women from schizophrenia, whereas the relative estrogen deficiency associated with menopause may be related to the increased incidence of LOS in women. A majority of treatment interventions with estrogen in postmenopausal women with schizophrenia have shown an improvement in symptoms. However, the incidence of schizophrenia is still considerably higher in younger versus older women.

Thus, an alternative hypothesis is that older men may be protected in some way.

The precise etiology of LOS, with an age of onset of between 40 and 60, and very late-onset schizophrenia-like psychosis (VLOS), with onset over age 60, is largely unknown. The International LOS Group concluded that schizophrenia with onset in middle age is a neurodevelopmental disorder and that differences with EOS are more of degree than of kind. The group also postulated that VLOS had some features that suggested a neurodegenerative component, including more brain abnormalities and neuropsychological deficits. A potential problem with this tripartite classification system is its reliance on age rather than clinical presentation, and some persons over 60 can resemble earlier-onset cases.

Overall, symptom presentation in EOS and LOS is more similar than different. Most studies have found that patients with LOS did not differ significantly from those with EOS regarding the severity of positive symptoms, although a large study by Vahia and colleagues found that late-onset patients had less severe positive symptoms and general psychopathology. This study also found that both groups had similar negative, depressive, and deficit symptoms.

Both EOS and LOS have cognitive deficits at presentation. Most studies have reported that, like EOS, cognitive deficits in LOS generally remain stable over time. There is less consensus on the nature and severity of these cognitive deficits. Some studies have reported fewer severe cognitive deficits in LOS as compared to EOS; others have found no differences in severity. No consistent biologic markers of schizophrenia with a later onset have been identified, although one study found raised C-reactive protein concentrations in people with LOS and VLOS.

A review of neuroimaging studies that compared EOS and LOS found differences in 8 of 10 studies, but there were inconsistencies across the studies. Similarly, measures of event-related brain potentials and neurofibrillary tangle densities found no differences between EOS and LOS groups.

Limited data exist specifically examining functioning in LOS. Patients demonstrated significantly greater disability in various areas of social functioning compared to healthy controls. Although marriage rates are higher in persons with LOS compared to EOS, social isolation was greater when comparing LOS patients to the general population. While some studies found that LOS was less frequently associated with a family history of schizophrenia than EOS, a more recent study reported no appreciable differences. Familial association of schizophrenia appears to be more common in EOS and LOS than in VLOS.

While differences between early-onset schizophrenia and LOS are more subtle, VLOS is a special case. VLOS is rare, with a reported annual incidence of 27 per 100,000 men and 48 per 100,000 women. Previous findings examining age have been mixed, showing both increased and decreased VLOS incident rates with advancing age. The clinical presentation of VLOS includes many striking features. One of them is a high prevalence of partition delusions. In this type of delusion, a person is convinced that people, objects, or radiation can pass through impermeable structures (walls, windows, doors). Hallucinations often involve multiple sensory domains (e.g., auditory, visual, tactile, olfactory). Formal thought disorders and negative symptoms, core symptoms of schizophrenia at a younger age, are infrequent, which has further fueled criticism of whether VLOS should be considered as a subgroup of schizophrenia. Mood disruption often occurs in VLOS, but adequate research to confirm this is lacking. Patients tend to be agitated and combative about their paranoid psychotic experiences.

The pattern of cognitive dysfunction is a focus in the debate on whether VLOS is a primarily neurodegenerative disorder. Nonetheless, like schizophrenia starting at a younger age, cognitive dysfunction in VLOS affects mainly frontal-subcortical cognitive domains. Memory disorders relate to learning new information, but retention of information is usually unaffected. This finding stands in contrast to the retention disorders that characterize cortical dementias like AD. Limited data exist on the longitudinal course of cognitive dysfunction in VLOS, and most studies had relatively short follow-ups. An often-cited Australian study that assessed cognition over 5 years found conversion to dementia in 9 of the 19 subjects. However, these findings were criticized for possibly including patients with early-stage dementia with psychotic features. Neuroimaging studies in VLOS have been very few but did not convincingly demonstrate a neurodegenerative substrate. Genetic factors have not been implicated in the etiology of VLOS.

In clinical practice, patients for whom VLOS is considered a differential diagnosis constitute a heterogeneous population, with only a proportion of them developing dementia with time. Like LOS, in VLOS, there is a broad differential diagnosis with other disorders involving psychotic symptoms in later life. A more conclusive diagnosis of VLOS, therefore, can be made only after a sufficient follow-up period.

Table 55.3h–1 provides an outline of the features that distinguish the three categories of schizophrenia.

Course and Prognosis

The course of LOS is often chronic but may be interrupted by partial remissions and exacerbations. Most studies have reported that in LOS, as in EOS, cognitive deficits tend to remain relatively stable over time, apart from changes associated with normal aging. The overall prognosis of LOS may be better than that of EOS. Patients may be quite responsive to antipsychotics used in lower doses. However, mortality rates, especially from suicide, exceed that in the general population and are probably comparable to those in EOS.

Treatment

Concerning LOS, a Cochrane report, updated in 2013, concluded that there is a paucity of trial-based evidence upon which to base guidelines for the treatment of LOS and that there is a need for good quality-controlled clinical trials. Howard and colleagues reported that a low dose of the antipsychotic drug amisulpride (100 mg) was effective and well-tolerated as a treatment for VLOS. Many of the pharmacologic and psychosocial recommendations for EOS can be applied to LOS, although expert guidelines suggest lower doses of antipsychotic medications may be needed in the latter.

Psychosocial treatment is a vital aspect of the management of LOS, and treatments specifically designed for older adults with schizophrenia, such as cognitive behavioral therapy and social skills training, have shown promise. The ability to engage patients who often lack insight into their disorder requires an assertive and multidisciplinary approach.

DIFFERENTIAL DIAGNOSIS OF PSYCHOTIC DISORDERS IN LATER LIFE

Making a diagnosis of schizophrenia in later life can be challenging. The onset of hallucinations and delusions in older adults can arise de novo or be associated with mood disorders, sensory deficits, polypharmacy, substance abuse, medical disorders, or dementia. The lifetime risk of developing psychotic symptoms in late life has been estimated to be up to 23%, primarily due to incident dementia. The prevalence of psychotic symptoms in noncognitively impaired older people has been estimated to be between 1% and 13.4%, with a median prevalence of 3.6%. One review study from Western Europe found the prevalence of psychotic symptoms among noncognitively impaired older adults increased with advancing age, so that rates were under 2% in persons aged 65 to 74, but 4% and 7% in those aged 85 to 94 and 95 to 104, respectively. The prevalence of psychosis in late-life in clinical populations is appreciable and is observed in 5% to 15% of geriatric psychiatry inpatients and up to 62% of nursing home patients. The clinician must be able to distinguish between "primary" psychosis (not attributable to other medical disorders) and "secondary" psychosis (attributable to other medical disorders), which account for 40% and 60% of cases in older adults, respectively. It is essential to obtain an accurate history and appreciate the time course of the disorder. Table 55.3h–2 lists the principal etiologies and time trajectories of psychotic disorders in older adults.

Table 55.3h–1.
Features Differentiating the Schizophrenia Subcategories

	Early-Onset Schizophrenia (EOS)	**Late-Onset Schizophrenia (LOS)**	**Very–Late-Onset Schizophrenia-Like Psychosis (VLOS)**
Age of onset	<40	40 to 60	>60
1-year prevalence among persons age 60+ (Dutch Catchment Area Study)	0.55%	0.14%	0.05%
Gender	Male = Female	Female > Male	Female >> Male
Family history of schizophrenia spectrum disorder	Present	Present	Absent
Cognitive decline	Deficits present at illness onset with age-related decline (stable vs. peers)	Deficits present at illness onset with age-related decline (stable vs. peers)	Substantial decline in a significant minority
Positive symptoms	▶ Severity of positive symptoms at onset but some improvement in later life ▶ Variety of positive symptoms ▶ Auditory hallucinations > visual hallucinations	Current level and characteristics of positive symptoms similar to older adults with EOS	▶ Lower severity of positive symptoms vs. EOS ▶ Visual hallucinations > auditory hallucinations

			▸ Other hallucinations such as tactile and olfactory are more common ▸ More somatic delusions ▸ More partition and persecutory delusions Auditory hallucinations with accusatory running commentaries
Negative symptoms	Present	Present	Absent
Thought disorder	Present	Present	Absent
Depression	Present	Present	Less common
Mortality vs. age peers	Higher	Higher	Higher
Antipsychotic dosage	May need higher dosages	Lower doses may be effective	Very low doses may be effective
Nonpharmacologic treatment	Psychosocial skills training; self-management techniques	No controlled studies	No controlled studies

Mood Disorder

Differentiating schizophrenia from mood disorder with psychotic features and schizoaffective disorder can be difficult because mood disturbances are common during the prodromal, active, and residual phases of schizophrenia. As described earlier, prevalence rates of cooccurring clinical depression have been found in up to three-quarters of older persons with schizophrenia, although mood symptoms may not meet the criteria for major depression. Furthermore, mood-congruent or mood-incongruent psychotic features may occur in major depression or bipolar disorder, although in both disorders, psychotic features tend to be more mood-congruent, for example, grandiosity, erotomania, special powers among bipolar individuals and delusions of poverty, guilt, sin, or physical illness in depressed persons. Auditory hallucinations occur less commonly in depression and may be vague or condemnatory. Catatonic features are more commonly seen in severe depressive disorders, and a history of mood symptoms may help to discriminate between conditions based on DSM-5-TR criteria. Mood disorder with psychotic features beginning after 40 years of age is differentiated from LOS by its affective symptoms that precede psychotic symptoms.

Table 55.3h–2.
Causes of Psychotic Symptoms in Older Adults

Etiology	Primary vs. Secondary Psychosis	Prevalence Among Psychotic Disorders	Time Course
Affective disorders	Primary	38%	Weeks to months
Schizophrenia spectrum and delusional disorders	Primary	3%	Months to decades
Major neurocognitive disorders (all types)	Secondary	40%	Months to years
Medical disease	Secondary	7%	Days to months
Delirium	Secondary	7%	Hours to days
Drug toxicity (substance abuse/medications)	Secondary	4%	Hours to months

Adapted from Reinhardt M, Ghoneim D, Murante T, et al. Schizophrenia spectrum and other psychotic disorders. In: Tampi RR, Tampi DJ, Boyle LL, eds. *Psychiatric Disorders in Late in Life. A Comprehensive Review.* Springer Nature; 2018:205–216.

Schizoaffective Disorder

It can be tricky to distinguish schizophrenia with depressive symptoms from schizoaffective disorder. The key difference is that in schizoaffective disorder, a *major* mood disorder must be present. The mood symptoms must occur simultaneously with the active-phase symptoms of schizophrenia, and they must be present for a substantial proportion of the total active and residual periods of the illness. Additionally, delusions or hallucinations must be present for at least 2 weeks in the absence of prominent mood symptoms. If the mood symptoms have a duration that is brief in relation to the total duration of the disturbance or do not meet the full criteria for a major depressive disorder, manic disorder, or mixed episode, then a diagnosis of schizoaffective disorder cannot be made. In the DSM-5-TR, clinicians must determine if a major depression has been present for a “majority” of the active and residual portions of the schizophrenia illness. Thus, clinicians must decide based on the life history of the illness, whereas previously, clinicians tended to look at discrete episodes. For geriatric psychiatrists, accurately determining what proportion of a 30 or 40-year illness involved major depression or manic symptoms is nearly impossible. Also, over the course of the illness, some older persons who had major depressive symptoms may attain persistently lower levels of depression, and consequently, they may no longer be considered to have schizoaffective disorder. The DSM-5-TR acknowledges the potential transience of the diagnosis and that some individuals will convert from schizoaffective disorder to another disorder as mood symptoms become less prominent. For these reasons, the DSM-5-TR states that there is growing evidence that schizoaffective disorder is not a distinct nosologic category.

Delusional Disorder

Delusional disorders can generate diagnostic conundrums. Persecutory, somatic, or erotomanic delusions in the older adult patient are typically secondary to another neuropsychiatric disorder (e.g., schizophrenia, mood disorder, or dementia); however, occasionally, a primary delusional disorder is implicated. Delusional disorders are rare in later life, with a prevalence rate of 0.03% to 0.04% and a lifetime risk of between 0.05% to 0.46% (the DSM-5-TR estimates the lifetime prevalence at 0.2%), which is well below lifetime risks for schizophrenia and bipolar illness. Delusional disorders usually first appear in the middle to late adulthood. The overall prevalence of delusional disorders is slightly higher among women than among men, and the average age of onset is earlier for men (40 to 49 years) than for women (60 to 69 years). There are no clear neuroanatomical changes associated with delusional disorder. There is mixed evidence that hearing or visual abnormalities might play a role in the development of delusional disorders. Family, twin, and adoption studies have suggested that a close association between delusional disorder and schizophrenia is unlikely, although rates of familiar schizophrenia may be higher than in the general population.

Delusional disorder is differentiated from schizophrenia by a lack of prominent hallucinations and an absence of deterioration in areas of functioning outside the delusional scope. The delusions are necessarily nonbizarre and involve situations that may occur in real life, such as theft, suffering from a disease, or spousal infidelity. For example, delusions that are seen commonly in older adult patients include skin infestation by insects; that internal organs are cancerous; that the person is being spied on, followed, or poisoned; and that the spouse is unfaithful. One study found that nearly half of patients had persecutory delusions. Consequently, a delusional disorder may be difficult to differentiate from schizophrenia with prominent paranoid ideation, especially in instances in which disorganized speech, disorganized behavior, or flat or inappropriate affect are less prominent. Nonetheless, the presence of auditory hallucinations or bizarre delusions seen in schizophrenia should differentiate the disorder. Furthermore, poor psychosocial functioning in delusional disorder is directly related to the delusional beliefs themselves, but marital and social functioning may be compromised. While subtle cognitive deficits may be found, intellect is generally maintained. Like schizophrenia, the course of the disorder varies among those afflicted, with perhaps one-third experiencing remission and two-thirds a prolonged course. Many avoid treatment, although among those who take antipsychotic medication, improvement has been reported in more than four-fifths of patients.

Psychosis Associated With Cognitive Disorders, General Medical Conditions, or Delirium

A patient who presents with psychotic symptoms after the age of 40, particularly if the symptoms develop after the age of 60, requires a thorough evaluation for underlying medical and neurologic disorders. Dementia is a common disorder in older adult persons, and 50% to 60% with a diagnosis of dementia of the Alzheimer type may experience psychotic symptoms at some point during their dementia. Delusions in AD tend to be nonbizarre, episodic, and preceded by cognitive decline. Typically, delusions in AD involve someone stealing from them or entering their house. Visual hallucinations are more common in AD than in schizophrenia, while auditory hallucinations are less common. Other neurocognitive disorders such as Lewy body Dementia and Parkinson Disease with Dementia may have higher rates of hallucinations and delusions than AD, whereas vascular and frontotemporal dementias have lower rates of psychotic symptoms.

Psychosis due to a general medical condition is diagnosed from the history, physical examination, or laboratory tests that show the delusions or hallucinations are the direct physiologic result of a general medical condition. Metabolic disturbances and neurologic and endocrinologic disorders are the most common causes of psychotic symptoms.

The diagnosis of delirium is established through observation of an acute change in mental status. and is characterized by a fluctuating deficit in attention, with changes in cognition that are not better explained by another neurocognitive disorder. The diagnosis of delirium remains reliant on clinical acumen, despite multiple efforts to establish diagnostic biomarkers. Behavioral symptoms are very common during delirium, including hallucinations and delusions, sleep disturbances, aberrant behavior, and affective lability. Hallucinations and delusions are common in delirium, with prevalence rates ranging from 40% to 70% for hallucinations, which are more frequently visual, and from 24% to 79% for delusions, which are more commonly paranoid types.

Substance-Induced Psychotic Disorder

About half of persons aged 65 and over drink alcohol, and one-fourth of those aged 85 and over continue to drink. Illicit drug use has been estimated at 10% to 15% of persons aged 50 to 64 and 5% among those aged 65 and over. At least one prescribed medication is taken by 87% of older adult adults, and 38% take five or more medications; about two-fifths use over-the-counter medications. These older adults are at greater risk for adverse effects because of polypharmacy, changes in metabolism and neuroreceptor sensitivity, and comorbid physical and neurologic disorders. Several different substances, such as amphetamines, cocaine, and phencyclidine, can produce symptoms like those of schizophrenia, such as delusions, hallucinations, or a mixture of positive and negative symptoms. Many substances can induce psychotic symptoms during withdrawal states, that is, benzodiazepines, barbiturates, and alcohol. Ideally, substance-induced psychotic disorders would be differentiated from schizophrenia after a period of abstinence; however, other evidence, such as whether the psychotic symptoms appear to be exacerbated by the substance and decrease when it has been discontinued or whether the psychotic symptoms have been instigated and maintained by the substance use, need to be considered.

FURTHER READINGS

Bartels SJ, Dimilia PR, Leutwyler H. Model programs and interventions for older adults with schizophrenia. In: Cohen CI, Meesters PD, eds. *Schizophrenia and Psychoses in*

Later Life: New Perspectives on Treatment, Research, and Policy. Cambridge University Press; 2019.

Brink M, Green A, Bojesen AB, Lamberti JS, Conwell Y, Andersen K. Physical health, medication, and healthcare utilization among 70-year-old people with schizophrenia: a nationwide Danish register study. *Am J Geriatr Psychiatry*. 2017;25(5):500–509.

Centorino M, Schultz SK. Biological changes in older persons with schizophrenia. In: Cohen CI, Meesters PD, eds. *Schizophrenia and Psychoses in Later Life: New Perspectives on Treatment, Research, and Policy*. Cambridge University Press; 2019:55–68.

Cohen CI, Abdallah CG, Diwan S. Suicide attempts and associated factors in older adults with schizophrenia. *Schizophr Res*. 2010;119(1):253–257.

Cohen CI, Meesters PD, Zhao J. New perspectives on schizophrenia in later life: implications for treatment, policy, and research. *Lancet Psychiatry*. 2015;2(4):340–350.

Cohen CI, Vengassery A, Garcia Aracena EF. A longitudinal analysis of quality of life and associated factors in older adults with schizophrenia spectrum disorder. *Am J Geriatr Psychiatry*. 2017;25(7):755–765.

Cohen CI. Assessing outcomes in schizophrenia in later life. In: Cohen CI, Meesters PD, eds. *Schizophrenia and Psychoses in Later Life: New Perspectives on Treatment, Research, and Policy*. Cambridge University Press; 2019:95–111.

Cohen CI, Freeman K, Ghoneim D, Vengassery A, Ghezelaiagh B, Reinhardt MM. Advances in the conceptualization and study of schizophrenia in later life: 2020 Update. *Clin Geriatr Med*. 2020;36:221–236.

Cohen CI, Reinhardt MM. Recovery and recovering in older adults with schizophrenia: A 5-tier model. *Am J Geriatr Psychiatry*. 2020;28:872–875.

Copeli F, Cohen CI. Medical issues in older persons with schizophrenia. In: Cohen CI, Meesters PD, eds. *Schizophrenia and Psychoses in Later Life: New Perspectives on Treatment, Research, and Policy*. Cambridge University Press; 2019:79–94.

Friedman JI, Harvey PD, Coleman T, et al. Six-year follow-up study of cognitive and functional status across the lifespan in schizophrenia: a comparison with Alzheimer's disease and normal aging. *Am J Psychiatry*. 2001;158(9):1441–1448.

Gogos A, Sbisa Alyssa M, Sun J, Gibbons A, Udawela M, Dean B. A role for estrogen in schizophrenia: clinical and preclinical findings. *Int J Endocrinol*. 2015;2015:615356.

Graff-Guerrero A, Rajji TK, Mulsant BH, et al. Evaluation of antipsychotic dose reduction in late-life schizophrenia: a prospective dopamine D2/3 receptor occupancy study. *JAMA Psychiatry*. 2015;72(9):927–934.

Hegarty JD, Baldessarini RJ, Tohen M, Waternaux C, Oepen G. One hundred years of schizophrenia: a meta-analysis of the outcome literature. *The Am J Psychiatry*. 1994;151(10):1409–1416.

Howard R, Rabins PV, Seeman MV, Jeste DV. Late-onset schizophrenia and very-late-onset schizophrenia-like psychosis: an international consensus. The international late-onset schizophrenia group. *Am J Psychiatry*. 2000;157(2):172–178.

Jeste DV, Wolkowitz OM, Palmer BW. Divergent trajectories of physical, cognitive, and psychosocial aging in schizophrenia. *Schizophr Bull*. 2011;37(3):451–455.

Khan AY, Redden W, Ovais M, Grossberg GT. Current concepts in the diagnosis and treatment of schizophrenia in later life. *Current Geriatrics Reports*. 2015;4(4):290–300.

Liu AY, Rajji TK, Blumberger DM, Daskalakis ZJ, Mulsant BH. Brain stimulation in the treatment of late-life severe mental illness other than unipolar nonpsychotic depression. *Am J Geriatr Psychiatry*. 2014;22(3):216–240.

Madhusoodanan S, Coleman A. Treatment of schizophrenia and psychoses in older adults: psychopharmacological approaches. In: Cohen CI, Meesters PD, eds. *Schizophrenia and Psychoses in Later Life: New Perspectives on Treatment: Research, and Policy*. Cambridge University Press; 2019:146–168.

Maglione JE, Vahia IV, Jeste DV. Schizophrenia spectrum and other psychotic disorders. In: Steffens DC, Blazer DG, Thakur ME, eds. *Textbook of Geriatric Psychiatry*. 5th edition. American Psychiatric Publishing; 2015.

Meesters PD, Comijs HC, de Haan L, et al. Subjective quality of life and its determinants in a catchment area based population of elderly schizophrenia patients. *Schizophr Res*. 2013;147(2):275–280.

Meesters PD, de Haan L, Comijs HC, et al. Schizophrenia spectrum disorders in later life: prevalence and distribution of age at onset and sex in a Dutch catchment area. *Am J Geriatr Psychiatry*. 2012;20(1):18–28.

Meesters PD, Stek ML, Comijs HC, et al. Social functioning among older community-dwelling patients with schizophrenia: a review. *Am J Geriatr Psychiatry*. 2010;18(10):862–878.

Muñoz-Negro JE, Ibáñez-Casas I, de Portugal E, Lozanno-Gutierez V, Martinez-Leal R, Cervilla JA. A psychopathological comparison between delusional disorder and schizophrenia. *Can J Psychiatry*. 2018;63:12–19.

Murante T, Cohen CI. Cognitive functioning in older adults with schizophrenia. *FOCUS*. 2017;15(1):26–34.

Nguyen TT, Eyler LT, Jeste DV. Systemic biomarkers of accelerated aging in schizophrenia: a critical review and future directions. *Schizophr Bull*. 2018;44(2):398–408.

Rajji TK, Ismail Z, Mulsant BH. Age at onset and cognition in schizophrenia: metaanalysis. *Br J Psychiatry*. 2009;195(4):286–293.

Rajji TK, Voineskos AN, Butters MA, et al. Cognitive performance of individuals with schizophrenia across seven decades: a study using the MATRICS consensus cognitive battery. *Am J Geriatr Psychiatry*. 2013;21(2):108–118.

Reinhardt M, Ghoneim D, Murante T, et al. Schizophrenia spectrum and other psychotic disorders. In: Tampi RR, Tampi DJ, Boyle LL, eds. *Psychiatric Disorders in Late in Life. A Comprehensive Review.* Springer Nature; 2018:205–216.

Schnack HG, van Haren NE, Nieuwenhuis M, Hulshoff Pol HE, Cahn W, Kahn RS. Accelerated brain aging in schizophrenia: a longitudinal pattern recognition study. *Am J Psychiatry*. 2016;173(6):607–616.

Sigström R, Gustafson D. Epidemiology of psychotic disorders: methodological issues and empirical findings In: Cohen CI, Meesters PD, eds. *Schizophrenia and Psychoses in Later Life: New Perspectives on Treatment, Research, and Policy*. Cambridge University Press; 2019:1–12.

Stroup TS, Olfson M, Huang C, et al. Age-specific prevalence and incidence of dementia diagnoses among older us adults with schizophrenia. *JAMA Psychiatry*. 2021;78(6):632–641.

Suzuki T, Uchida H. Successful withdrawal from antipsychotic treatment in elderly male inpatients with schizophrenia–description of four cases and review of the literature. *Psychiatry Res*. 2014;220(1–2):152–157.

Vahia IV, Palmer BW, Depp C, et al. Is late-onset schizophrenia a subtype of schizophrenia? *Acta Psychiatr Scand*. 2010;122(5):414–426.

Zisook S, Kasckow JW, Golshan S, et al. Citalopram augmentation for subsyndromal symptoms of depression in middle-aged and older outpatients with schizophrenia and schizoaffective disorder: a randomized controlled trial. *J Clin Psychiatry*. 2009;70(4):562–571.

55.3i Personality Disorders

Marc E. Agronin, M.D.

Multiple normative and pathologic changes during the aging process affect personality, and these same forces apply to the consideration of personality disorders in late life. These disorders represent a heterogeneous array of chronic affective and behavioral disturbances that cause enormous personal and interpersonal distress and pose significant challenges to clinical management, especially in the setting of comorbid medical and psychiatric conditions. Changing and inherently limited diagnostic schemes combined with a general lack of previous study make it difficult to understand and plot the trajectory of a given personality disorder in late life. Clinical perspectives on personality stability versus change will vary depending on the methods of assessment. There is limited empirical evidence and a range of unsubstantiated opinions as to whether personality disorders persist with minimal or no change, convert into more age-appropriate forms with similar levels of pathology, or remit over time. Understanding all of these factors is critical to improving diagnosis and optimizing clinical management.

DEFINITIONS

Defining the concept of *personality* itself is key to a fuller understanding of personality disorders. According to the *Bantam Medical Dictionary*, personality is “an enduring disposition to act and feel in particular ways that differentiate one individual from another.” George Vaillant qualifies this definition, suggesting that the enduring aspects of personality, which he terms *temperament,* consist largely of various inherited elements such as personality traits and IQ. He bases this belief on longitudinal studies of core personality traits (e.g., extroversion, introversion) that have generally found stability with aging. On the flip side, Vaillant holds that elements of personality or *character* defined by adaptive styles can change significantly with aging, affected by both “environment and maturation.” Putting together both elements of stability and change, Henry Murray’s statement that “the history of the personality is the personality” is most helpful when considering aged individuals. Lacking accurate longitudinal data from an older individual’s younger days, the average clinician is forced to make decisions about personality functioning based on more recent history.

The distinction between normal and abnormal personalities is based largely on the expression of certain personality *traits* (as opposed to more acute psychopathologic symptoms) that are repeatedly associated with unhappiness, interpersonal conflict, and/or occupational dysfunction. Although everyone has a unique personality with relative strengths and weaknesses, the current diagnostic model requires the categorization of certain personalities as pathologic along the lines of various subtypes. This categorical model has been criticized for the ways in which it provides pejorative labels and obscures the richness of personality. In late life, it is particularly problematic given the large number of physical and psychosocial stresses that can distort an individual’s pattern of adaptation.

An alternative dimensional approach was proposed but not adopted in the *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition (DSM-5). Personality disorders are defined in DSM-5 as chronic psychiatric disorders characterized by enduring and pervasive patterns of inflexible and maladaptive inner experiences and behaviors. These dysfunctional features are rooted in early development, deviate from cultural expectations, and lead to significant social and occupational impairment across several spheres of function: perception and interpretation of experiences, emotional expression, interpersonal relations, and impulse control.

There are 10 subtypes of personality disorders listed in the DSM-5 classification, grouped into three broad phenomenologic clusters: cluster A, the *odd* cluster, includes paranoid, schizoid, and schizotypal subtypes; cluster B, the *dramatic* cluster, includes antisocial, borderline, histrionic, and narcissistic subtypes; and cluster C, the *anxious* cluster, includes avoidant, dependent, and obsessive-compulsive subtypes. Depressive and passive-aggressive (or negativistic) personality disorders were two subtypes that were listed in the appendix of the *Diagnostic and Statistical Manual of Mental Disorders,* fourth edition text revision or DSM-IV-TR but are no longer included in DSM-5. Each personality subtype is defined by 8 to 10 diagnostic criteria, and an individual must meet a specified number of them in any combination in order to qualify for a diagnosis. Individuals may demonstrate a single disorder or a combination of maladaptive traits from several disorders. Individuals who have symptoms characteristic of a personality disorder in general, but do not meet the criteria for any specific disorder, may be diagnosed as having an *unspecified personality disorder.*

Although there are several designated pediatric categories for personality disorders, there are no equivalent, age-adjusted geriatric subtypes, despite growing evidence that changes in both personality and personality disorders in late life can lead to enduring but *different* patterns of personality dysfunction. However, when personality change results from physical damage to the brain from substance abuse or a medical condition, the term *personality change due to another medical condition* can be used (with the medical condition specified and then coded separately), with several designated subtypes: labile, disinhibited, aggressive, apathetic, and paranoid. Common causes of personality change include traumatic head injury, stroke, endocrine dysfunction, encephalitis, and Alzheimer disease.

EPIDEMIOLOGY

Community-based epidemiologic studies have found rates of personality disorders in adults ranging from 10% to 20%, with the 2001–2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) Data for 16,884 adults ages 50 and older yielding a rate of 10.7% for at least one personality disorder. Another set of NESARC data found a rate of 14.5% of one or more personality disorders in older adults. These rates increase two- to threefold in individuals being treated in both outpatient and inpatient psychiatric settings for comorbid psychiatric diagnoses, particularly major depression. For example, a study of 546 older adult psychiatric inpatient admissions found personality disorders in 13% of all patients, which increased to 24% of those diagnosed with major depression. In all studies, the method of diagnosis has produced variations in prevalence rates. Epidemiologic surveys and more impressionistic clinical diagnoses have generally found higher rates of personality disorders and yet a more limited spectrum of subtypes compared to criterion-based structured clinical interviews.

The most common personality disorders seen across late-life studies have been paranoid, avoidant, dependent, and obsessive-compulsive subtypes. The most common subtype in the NESARC data sets was obsessive-compulsive personality disorder at 6.5% to 7.6%, followed by narcissistic personality disorder at 3.9% and paranoid personality disorder at 2.5%. In late life, nearly two-thirds of those diagnosed with a personality disorder also meet the criteria for two or more subtypes or for the unspecified category. It is important to bear in mind that the lifelong course of specific subtypes will influence prevalence. For example, a decrease in the prevalence of antisocial personality disorder in late life no doubt results from increased rates of remission, death, institutionalization, and avoidance of psychiatric settings with age. Decreasing rates of other personality disorders in late life may be due to the fact that they are difficult to differentiate from other specific psychiatric diagnoses or may change in ways that obscure their recognition.

There are limited data on the prevalence of personality disorders in long-term care settings, due to the great difficulty obtaining longitudinal data, in addition to the challenge of interviewing individuals who may be suffering from varying degrees of cognitive impairment. One survey that based diagnoses on clinical interviews found that nearly 11% of a modest sample of nursing home residents had a primary or secondary diagnosis of personality disorder. An identical diagnostic rate was found in a sample of intellectually disabled adults older than 40 years who were admitted to a long-term care institution—although, in that study, the congenital diagnosis might call into question the personality disorder diagnoses.

A review of referrals to a nursing home found that 15% were for unspecified personality problems, although it is not clear how many of these individuals had actual personality disorders. Rates for personality disorder subtypes in long-term care facilities have not been established in any study.

DIAGNOSIS

In spite of the very specific diagnostic criteria provided by DSM-5, clinicians immediately encounter a plethora of diagnostic challenges when assessing older individuals. Sometimes the older personality, viewed from an ageist stance, is erroneously believed to be more dependent, stubborn, hypochondriacal, and less adaptive in the face of stress compared to younger personalities. Such an attitude may lead clinicians to view many disordered behaviors as normal in late life, preventing accurate diagnosis. At the same time, many clinicians might not consider the possibility of a personality disorder in an older patient or might be daunted by the large number of diagnostic criteria. As a result, these diagnoses are often deferred or merely lumped under an unspecified category. Lack of knowledge either of specific criteria or how to apply them in late life has been shown to create discrepancies between largely impressionistic clinical diagnoses and those determined through more rigorous structured interviews.

The criteria themselves foster misdiagnosis since they were developed for adults and do not take into consideration age-associated changes in lifestyle and behaviors that render many of them obsolete in later life. For example, the diagnosis of antisocial personality disorder includes criteria such as “repeatedly performing acts that are grounds for arrest,” “repeated physical fights or assaults,” and “consistent irresponsibility” with work or financial obligations. For an older person who is retired from work, residing in congregate living or a long-term care facility with limited mobility, or physically frail, these criteria will simply not apply. Diagnostic criteria for borderline personality disorder, such as self-injurious behaviors and self-damaging impulsivity with respect to spending, sex, driving, or eating, are also less applicable to older adults, who in general tend to be less impulsive. As of yet, there are no geriatric criteria that can be substituted, and attempts by clinicians to make diagnoses based on fewer criteria or to “age-adjust” other criteria threaten the very validity of the diagnosis. Several age-specific tools with reasonable accuracy have been developed to address evolving diagnostic symptoms, including the Gerontological Personality Disorders Scale. Researchers have also suggested that dimensional rather than categorical models may be more useful across ages.

Not surprisingly, an analysis of data from the National Epidemiologic Survey on Alcohol and Related Conditions, which studied over 43,000 individuals, found that young and old adults endorsed nearly 30% of personality disorder criteria differently, despite having equal degrees of personality dysfunction. The end result was that older adults were more likely to have schizoid and obsessive-compulsive personality disorder diagnoses, while younger adults were more likely to have avoidant and dependent personality disorder diagnoses. The same diagnostic criteria, then, yielded different disorders based on age.

Another diagnostic dilemma is the need for accurate longitudinal descriptions of a personality on which to base the diagnosis. Unlike symptoms of most other psychiatric disorders, which are readily reported or observed, personality disorder symptoms can be more difficult to elicit through self-report given their pejorative nature, and when observed, require repeated encounters across different settings to be accurate and reliable. Instead, clinicians often have to rely on clinical records, which typically lack longitudinal data, or on existing informants who can rarely provide critical information dating back to an older patient’s young adulthood.

An important caveat for any clinician attempting to make a late-life diagnosis is that personality dysfunction does not necessarily indicate a personality disorder. Significant, repeated, and sometimes devastating losses in late life, such as the death of a spouse or of a series of close friends, a chronic debilitating illness, chronic pain, physical disability, the loss of finances or livelihood, or the overall loss of personal attributes that once defined an individual (e.g., strength, beauty, athletic, or sexual prowess), can lead to severe adjustment reactions that mimic personality disorders. Without clear longitudinal information, it is difficult to determine whether a patient with pervasive dysfunctional interpersonal relations, for example, is demonstrating a lifelong pattern or a more recent reaction to life events. Many symptoms of major depression and generalized anxiety disorder overlap with depressive, negativistic, avoidant, and depressive symptoms of several personality disorders. Similarly, symptoms of persistent bipolar mania can mimic antisocial, narcissistic, and borderline traits. For this reason, personality disorder diagnoses are best deferred during episodes of acute psychiatric syndromes. Perhaps one of the most difficult decisions to make is whether changes in demeanor brought on by neurocognitive disorders and associated neuropsychiatric syndromes such as apathy are merely expressions of the underlying disorder or represent a true personality change.

CLINICAL FEATURES AND COURSE

Since its inception in 1952, iterative versions of the DSM have provided a heterogeneous and fluid array of personality disorder subtypes up until the more enduring categorization that appeared in the third edition of the DSM (DSM-III) in 1980. Despite the relative stability in the nomenclature since then, the DSM has never been clear on the course of personality disorders into late life, only suggesting that several remain relatively unchanged while others attenuate. The possibility of a late-onset personality disorder is not addressed in the DSM, aside from the diagnosis of a personality change disorder. Unfortunately, there are limited data on most subtypes, and current thinking rests largely on case studies.

For the clinician working with older adult individuals with known or suspected personality pathology, the main question is: What do these disorders look like in late life? Are they fundamentally the same as in previous years, or does aging change them in ways that either attenuate symptoms and cause remission or lead to similar degrees of pathology but with novel, age-appropriate manifestations? Consider the following clinical vignette.

Mr. L was an 82-year-old man who suffered from brittle diabetes, morbid obesity, and chronic obstructive lung disease. After his wife died, Mr. L was unable to manage his own care needs and had to be admitted to an assisted living facility (ALF). He was oxygen dependent and used an electric scooter to get around. Mr. L quickly established a reputation as a rude and aggressive man. He would badger other residents and staff in the elevator when they did not get out of his way quickly. He would make unrealistic demands of the waiters in the dining room and insult them when he did not get his way. He accused the nursing staff of deliberately withholding treatment after one of his medications went missing from his tray. Over time, the staff grew increasingly tired of his verbally abusive and disruptive behaviors and was afraid to help him since he often threatened to sue them for shoddy care. They described him as a difficult, paranoid, and hateful person.

These behaviors built into a crisis one day when Mr. L became enraged at a deaf resident standing in front of him in the elevator who could not hear him berating her for moving out of his way. He pushed her down and ran over her foot with his scooter, causing a serious skin laceration and bruise. He completely denied his actions, however, accusing the injured woman of deliberately feigning injury after blocking his scooter and getting bumped. He screamed at all staff who tried to intervene and threatened to kill anyone who got in his way. As a result, Mr. L was hospitalized involuntarily and not allowed to return to the ALF.

Additional information on Mr. L indicated that he had been an extremely aggressive but quite successful real estate broker as a young and middle-aged man. His daughter estranged herself from him due to his abusive behaviors, and his son limited his contact with his father, whom he described as arrogant, paranoid, and verbally abusive to those who disagreed with him. Mr. L's behaviors were chronic but seemed to have improved in the several years prior to his becoming ill, perhaps because his business success enabled him to get involved in several gratifying philanthropic endeavors. However, with the onset of illness and then the death of his wife, he could not cope with being dependent on others for help. He was completely unable to trust staff or forge stable relationships in the ALF and became increasingly angry, depressed, and paranoid. Having to adjust to institutional living proved impossible, and his brittle coping style was eventually overwhelmed, resulting in a crisis characterized by aggression and paranoid accusations. Putting together known history and current observations, Mr. L's clinical team diagnosed him with a paranoid and narcissistic personality disorder.

In this clinical vignette, several themes appear that characterize the course and prognosis of many personality disorders throughout life. First, dysfunctional behaviors do not arise de novo but have a history and have pervaded the individual's life in one form or another since young adulthood. In many cases, there may be long periods of increased stability in middle age that represent symptomatic remission, typically bolstered by stable relationships, successful occupational pursuits, and intensive therapeutic assistance. Not all personality disorders, however, will benefit from these stabilizing factors—particularly those with avoidant, schizoid, and schizotypal features. As seen in the illustrative case, major late-life stresses can exacerbate existing pathologic traits or lead to the reemergence of previous traits in ways that are different in kind but equally problematic as earlier in life. Lacking a longitudinal history and based strictly on observed behaviors, many personality-disordered individuals are typically diagnosed with major depression, bipolar disorder, or even late-onset schizophrenia, given the intensity of their symptoms. Although these diagnoses may present comorbidly, they are driven in part by the underlying personality disorder.

Cluster A Disorders

In late life, the diagnostic criteria for the three odd cluster disorders coalesce around a lack of desire for and openness toward interpersonal relationships. The tendency of the paranoid personality disorder to read malevolent intent into the actions of others leads to a lifetime of angry conflicts with others, while the aloof and strange demeanor of the schizoid and schizotypal individual precludes the formation of close relationships. As a result, odd cluster personality disorders tend to enter later life without the benefit of past interpersonal experiences that can, for less dysfunctional individuals, teach social skills and provide emotional support. It is believed that the pathologic cognitive and behavioral patterns of the odd cluster disorders tend to remain the same or even worsen with aging, with some conversion to schizophrenia spectrum disorders.

Paranoid, schizoid, and schizotypal individuals face age-associated stresses with variable responses. On the one hand, the lack of close relationships means that these individuals may feel a less emotional impact from the inevitable loss of family members or longstanding social contacts. However, when medical illness and disability occur, these same individuals often have great difficulty seeking help and coping with forced dependency on others, especially when placed in long-term care institutions. The result is often the increased expression of longstanding dysfunctional behaviors and defense mechanisms, including paranoid ideation, social isolation, and odd or idiosyncratic beliefs or rituals. There may also be an increased risk of psychosis to the extent that these disorders are considered potential precursors for late-onset schizophrenia.

Cluster B Disorders

Although lumped together under the “dramatic” cluster domain, these four personality disorders are heterogeneous in both expression and course over the life cycle. A central theme, however, is that while their psychopathology results in severe and, in some cases, deadly consequences, there is potential for symptomatic quiescence, improvement, or even remission in middle age when there are significant personal or occupational achievements or the presence of consistent interpersonal relationships. In addition, individuals with more personal, social, and/or financial resources and with less psychiatric comorbidity tend to fare better. By the same token, age-associated losses in relationships, social and occupational roles, physical attributes such as strength and beauty, and financial status can reverse many of these gains and prove devastating to the personality. For those individuals who had improved over their lifetimes, the stress of major losses can lead to the reemergence of previous dysfunctional cognitive and behavioral patterns, but in ways not always recognizable by the current diagnostic nomenclature. These relapses are inevitably complicated by comorbid psychiatric disorders such as major depression, neurocognitive disorders, anxiety disorders, and substance-related and addictive disorders.

There are several factors that limit the number of older individuals with antisocial personality disorder who are seen in clinical settings. First, their history of reckless and criminal behaviors and high rates of substance abuse often results in either death or incarceration long before they reach middle age. Second, these individuals have poor insight into their problems, few supportive relationships, and a lack of motivation to get help unless there is an acute issue that prevents their ongoing activities. As a result, both young and old antisocial individuals tend to avoid psychiatric settings.

Despite this rather grim assessment, clinical data reveal, somewhat surprisingly, that up to a third of antisocial individuals demonstrate significant clinical improvement if they survive into middle age, and about 12% to 33% no longer meet diagnostic criteria after 10 to 15 years. Individuals with poor social skills and higher degrees of aggression tend to do worse. The most salient changes include a shift away from aggressiveness, promiscuity, and criminal behaviors but replaced, in part, by substance abuse, depression, and somatization. All of these changes may reflect a general trend of decreased impulsivity and aggression with aging.

In long-term care settings, antisocial behaviors may include resistance to care, surreptitious substance abuse, abuse of other residents, and disregard for institutional rules regarding personal safety, hygiene, and schedules. Based strictly on observable behaviors, these older individuals might not meet the criteria for antisocial personality disorder, even though a longstanding lack of conscience—the hallmark of the disorder—is driving more age-appropriate dysfunctional behaviors.

In a similar fashion, individuals with borderline personality disorder demonstrate the potential for significant improvement and even remission in middle age, with one longitudinal study of 100 individuals finding that 75% no longer met the criteria after 15 years. This improvement was bolstered by personal and occupational successes and associated with increased physical attractiveness, higher intelligence, and self-discipline. Poorer outcomes were associated with a history of abuse, antisocial tendencies, high levels of aggression, and substance abuse. Even with improvement, however, there remained a significant risk for depression, substance abuse, and suicide.

Clinical case studies have described borderline individuals who demonstrate reemergent symptoms later in life but with a different appearance. Two key stressors are the loss of a spouse or partner and admission to a long-term care setting. Earlier defining symptoms of self-mutilation, impulsive and risky behaviors, and poor self-identity are attenuated and replaced by mood lability and depression, nonadherence to treatment, overuse of prescription medications, and alcohol abuse. In long-term care settings, however, the erratic and rageful interpersonal style seen in younger individuals with borderline personality disorder is often still present, creating havoc in the milieu.

Over their lifetimes, individuals with a narcissistic personality disorder often leave a wake of hurt and disgruntled personal and business relationships in their path, and though aware of their own anger or estrangement from others, they rarely attribute it to their own behaviors. Their sense of entitlement and demands on others may bring success in middle age, but with time they are particularly vulnerable to age-associated losses of personal attributes, gratifying occupational and social roles, and supportive if not somewhat subservient relationships. It becomes increasingly difficult for them to replace lost or jettisoned relationships over time, and they may find themselves in great need due to illness or disability but without adequate social support. The very process of aging proves to be a significant narcissistic injury, only worsened when they have1 to rely on the help of others they deem below them. In long-term care, the narcissistic personality disorder often presents as simultaneously depressed, rageful, and imperious toward staff, leading to frequent conflicts.

Histrionic personality disorder is probably the least studied in late life and has not been reported much in epidemiologic surveys or in clinical case studies. Like narcissistic individuals, however, histrionic individuals often lack deep, supportive social relationships and are equally buffeted by age-associated losses in physical appearance that allowed them to maintain some state of psychological equilibrium in the past. It has been suggested that somatization may be one dysfunctional avenue for older histrionic individuals to obtain attention and social gratification.

Cluster C Disorders

The anxious cluster disorders are thought to involve a lifetime of chronic personality dysfunction without significant remission, perhaps explaining why they tend to be some of the most commonly seen personality disorders in late life. Comorbid major depression is common and is often the presenting diagnosis when the underlying personality disorder is not appreciated. Avoidant and dependent individuals, in particular, often enter later life without having been able to cultivate many supportive relationships. As a result, the loss of a spouse or someone who served as their main support can be devastating and precipitate severe anxiety, depression, and incapacity. Given their poor social skills and inability to activate the social resources to meet their daily needs, these individuals may end up in dire circumstances, bereft of support, and at high risk in the community of abuse and neglect. The same psychopathology may be somewhat adaptive in long-term care settings, however, where forced dependency is a welcome respite for them and alleviates the difficult responsibility of having to assume self-care.

Individuals with obsessive-compulsive personality disorder tend to be less isolated in later life than avoidant or dependent individuals and have often cultivated sufficient social support due to their rigid insistence on routines and schedules. These individuals are more self-sufficient and may do relatively well with aging as long as they are able to find activities, part-time jobs, or other tasks to keep them busy and organized. Cognitive and physical disability can threaten this stability, however, especially when obsessive-compulsive personalities are unable to control their daily life according to preset rules and routines and have to depend on less reliable caregivers or long-term care staff. The latter circumstances can lead to significant conflicts with caregivers, as well as angry behavioral outbursts, anxiety and panic attacks, and depression.

TREATMENT

Given both the chronicity and pervasiveness of personality disorders, the initial goal of treatment is not to provide a cure but to reduce the frequency and intensity of symptoms. True symptomatic improvement is challenging and takes months, if not years, to accomplish, and will always be slowed down when comorbid medical and other psychiatric disorders are present. In turn, the presence of comorbid personality pathology, particularly Cluster B traits, has been shown to render cognitive behavioral therapy (CBT) less effective for late-life anxiety disorders. Similarly, treatments for major depression have been found to take longer and to be less effective in the presence of underlying personality disorders. Thus, realistic target symptoms should be established at the outset of treatment so as not to fall into the trap of building up excessive expectations that end up disappointing the patient and their social support as well as burning out the therapist.

There are several necessary treatment approaches for any older individual with a personality disorder. First, a general survey of active medical, neurologic, and medication-related factors that may be causing acute changes in personality expression must be ascertained to ensure that adequate treatment is being provided.

Many of these factors, including cerebral dysfunction, metabolic abnormalities, pain, and unmet physical or emotional needs, may be triggering and/or fueling interpersonal conflicts or disruptive behaviors that represent the personality disorder. Normal psychotherapeutic or pharmacologic strategies will fail unless these factors are addressed. This first step will necessitate significant involvement of multidisciplinary treatment team members, other treating physicians, and caregivers in order to implement the treatment plan as well as to serve as informants on response—particularly for individuals with cognitive impairment and/or physical disabilities.

Second, comorbid psychiatric disorders such as major depression, panic disorder, and substance-use disorders, among others, must be under treatment in order to reduce the psychological stress on the personality, as well as to clarify the personality disorder diagnosis. Diagnostic false positives are more likely during episodes of other psychiatric disorders, prompting many clinicians to defer personality disorder diagnoses during this time. A particular challenge at this juncture involves dealing with personality disorders within the setting of Alzheimer disease or other types of neurocognitive disorders, which can change personality itself or distort previously dysfunctional expressions of personality in ways both positive and negative. Some individuals may be more labile and impulsive, while others may be more apathetic and less likely to act out, especially when there is frontal or temporal lobe dysfunction. Unfortunately, it is difficult to predict the course of any given individual personality or personality disorder within the setting of major neurocognitive disorders, even when the specific regions of cerebral or subcortical damage are known. Although there are a number of existing acetylcholinesterase agents and one glutamate-receptor antagonist that can treat the cognitive, behavioral, and functional symptoms of Alzheimer disease and other major neurocognitive disorders, there is no evidence that any of these agents will have a discernible effect on personality disorders.

Third, clinicians must be willing and able themselves or have available to them a multidisciplinary team that can actively engage the older personality-disordered individual over time and through a variety of treatment modalities in order to foster a stable alliance. For older borderline, histrionic, narcissistic, dependent, and obsessive-compulsive subtypes, episodic clinical visits with an unpredictable schedule and therapy with too narrow a focus on medication management are liable to impair the therapeutic relationship. It is best to have several team members working together to provide an adequate density of time and attention in order to maintain stability and manage crises. On the flip side, older schizoid, schizotypal, paranoid, and avoidant subtypes might do best with shorter and more intermittent clinical visits, with a specific focus geared toward their stated needs. In all circumstances, there are inevitable medical and social issues—and often crises—that arise and require the involvement of a clinician who has made a firm commitment to the patient. Although such an approach seems obvious, it is all too often overlooked in a time-limited and overstretched mental health system.

Fourth, older individuals with personality disorders are often seen in long-term care settings, providing a structured setting in which to engineer environmental or behavioral plans. These plans require the input and creativity of mental health, social work, and nursing staff to be developed, implemented, and revised if necessary. The following clinical vignette illustrates how these plans may work.

Mr. P was a 78-year-old man with severe spinal stenosis who suffered from chronic back pain and was confined to a wheelchair. He had previously been diagnosed with both narcissistic and obsessive-compulsive personality disorders, as well as major depression. Shortly after being admitted to a nursing home, he frequently became enraged and verbally threatening toward nurses whom he accused of not adequately maintaining the battery on his electric wheelchair. On one occasion, he struck a nurse who was trying to reconnect a lead to the battery, accusing her of failing to follow his detailed instructions. A multidisciplinary clinical team met to discuss the situation and decided to condition Mr. P's access to the electric wheelchair on his cooperative behavior. A single staff member was assigned to work with Mr. P to maintain the battery, which limited the targets of his wrath. He earned hours in the electric wheelchair for each day that he went without cursing or abusing staff. As a result, the treatment team noticed a significant reduction in his abusive behaviors.

The key to success with any plan lies in the concerted and well-coordinated efforts of staff to implement its elements, monitor for improvement, and maintain communication. Pitfalls include disputes among staff members that often reflect the personality disorder's inner dynamics, lack of consistency, and lack of patience. When plans fail, staff members may inappropriately resort to punitive actions to deal with dysfunctional behaviors or attitudes, such as withholding attention or care, neglecting legitimate needs, using chemical restraints unnecessarily, or seeking alternative placement.

Psychotherapy

In older individuals with several subtypes of personality disorders and minimal to no cognitive impairment, both individual and group psychotherapeutic modalities may be helpful. Schizoid, schizotypal, paranoid, antisocial, and avoidant individuals tend to be more reluctant to engage in insight-oriented therapy and might do best through the indirect fostering of therapeutic relationships. In long-term care settings, this can be achieved through nonpsychiatric therapies (e.g., physical or occupational therapy) or activities (small jobs, arts and crafts, and music) that provide vital involvement, socialization, and meaning. Older individuals with borderline, narcissistic, histrionic, obsessive-compulsive, and dependent personality disorders will do better with direct forms of psychotherapy, especially CBT. Dialectical behavior therapy (DBT) is a particularly useful form of CBT that was originally developed for individuals with borderline personality disorder but has been adapted for individuals with major depression and personality disorders. DBT focuses on helping individuals become mindful of both the triggers and the consequences of their behaviors and then develop strategies to change them. Studies in younger individuals with DBT have found less overall depression and anxiety, fewer suicidal and self-injurious behaviors, and a decreased number and length of psychiatric hospitalizations.

Group psychotherapy can take on many forms for older adult individuals, ranging from psychoeducational forums to single-focus groups for socialization, bereavement, or caregiving to more formal therapy groups such as are found in a structured outpatient program. Group selection for the older personality-disordered individual will depend on their tolerance for a group setting, ability to participate without either monopolizing the group or acting out, and level of motivation. Cognitively intact individuals with reasonably good social skills are the best candidates.

Pharmacotherapy

Psychotropic medications can be used across several levels in the older patient with a personality disorder: (1) crisis management; (2) treatment of comorbid psychiatric disorders, such as major depression; (3) treatment of the personality disorder itself; and (4) treatment of the most salient target symptoms associated with the personality disorder. The first two levels are covered extensively in other parts of this text and are inevitably faced with every personality disorder. Treatment of the personality disorder itself is quite difficult given the heterogeneity of presentations, and the literature has almost exclusively focused on borderline, schizotypal, and antisocial subtypes. Even in the majority of those studies, the main focus of treatment was on selected target symptoms that were most problematic, such as impulsivity, aggression, and psychosis. All three of these common symptoms in both dramatic cluster and odd cluster personality disorders have been found to be partially responsive to antipsychotic agents, antidepressants, and mood stabilizers such as divalproex (Depakote) and carbamazepine (Tegretol). Many of the early studies looked at the use of tricyclic antidepressant agents and monoamine oxidase inhibitors to treat symptoms of personality disorders. Although they demonstrated some degree of efficacy, these latter agents are less practical in older adult patients due to their anticholinergic properties and propensity to cause orthostasis. In general, there is no evidence to date that clearly supports one particular agent or medication class as more efficacious than another for any particular symptom or personality disorder. As it stands, given the lack of data for older individuals, the selection of specific psychotropic medications for both young and old individuals with personality disorders will follow the same guidelines.

In older individuals, clinicians must hearken to the familial caveat to “start low, go slow, but go” when prescribing medications. However, another equally important caveat is to know when the use of medications might actually be counterproductive. It is risky to prescribe to patients with poor adherence to medications or hazardous use of them due to cognitive impairment, denial of illness, or impulsivity. In all of those cases, more structure is needed to administer the medications, either through a trusted family member or a visiting nurse or by transferring the patient to an assisted living or long-term care facility. Patients with obsessive-compulsive and dependent traits sometimes become fixated on medication doses and dosing schedules and spend excessive time discussing this with the prescribing doctor instead of dealing with more relevant issues.

Taking medication becomes a way for them to pathologically structure their relationship with the doctor, such as through increased dependency. In all of these situations, the clinician must weigh the potential benefits of the medication against the potential harm to the patient through misuse or the resultant exacerbation of dysfunctional behaviors.

FURTHER READINGS

Abrams RC, Spielman LA, Alexopoulos GS, Klausner E. Personality disorder symptoms and functioning in elderly depressed patients. *Am J Geriatr Psychiatry*. 1998;6:24–30.

Agronin M. Personality disorders in the elderly: an overview. *J Geriatr Psychiatry*. 1994;27(2):151–191.

*Agronin ME, Maletta G. Personality disorders in late life: understanding and overcoming the gap in research. *Am J Geriatr Psychiatry*. 2000;8(1):4–18.

*Balsis S, Eaton NR, Cooper LD, Oltmanns TF. The presentation of narcissistic personality disorder in an octogenarian: converging evidence from multiple sources. *Clin Gerontol*. 2011;34(1):71–87.

*Balsis S, Segal DL, Donahue C. Revising the personality disorder diagnostic criteria for the Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (DSM-5): consider the later life context. *Am J Orthopsychiatry*. 2009;79:452–460.

Balsis S, Woods CM, Gleason ME, Oltmanns TF. Overdiagnosis and underdiagnosis of personality disorders in older adults. *Am J Geriatr Psychiatry*. 2007;15(9):742–753.

Black DW, Baumgard CH, Bell SE. A 16- to 45-year follow-up of 71 men with antisocial personality disorder. *Compr Psychiatry*. 1995;36(2):130–140.

Bohus M, Stoffers-Winterling J, Sharp C, Krause-Utz A, Schmahl C, Lieb K. Borderline personality disorder. *Lancet*. 2021;398(10310):1528–1540.

Grant BF, Hasin DS, Stinson FS, et al. Prevalence, correlates, and disability of personality disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions. *J Clin Psychiatry*. 2004;65(7):948–958.

Gulpers B, Oude Voshaar R, Kampman M, Verhey F, VAN Alphen S, Hendriks GJ. The impact of personality pathology on treatment outcome in late-life panic disorder. *J Psychiatr Pract*. 2020;26(3):164–174.

Holzer KJ, Huang J. Physical health-related quality of life among older adults with personality disorders. *Aging Ment Health*. 2019;23(8):1031–1040.

Ingenhoven T, Lafay P, Rinne T, Passchier J, Duivenvoorden H. Effectiveness of pharmacotherapy for severe personality disorders: meta-analyses of randomized controlled trials. *J Clin Psychiatry*. 2010;71(1):14–25.

Kirchner SK, Roeh A, Nolden J, Hasan A. Diagnosis and treatment of schizotypal personality disorder: evidence from a systematic review. *NPJ Schizophr*. 2018;4(1):20.

Laheij-Rooijakkers LAE, van der Heijden PT, Videler AC, Segal DL, van Alphen SPJ. Development of a tool to detect older adults with severe personality disorders for highly specialized care. *Int Psychogeriatr*. 2020;32(4):463–471.

Lynch TR. Treatment of elderly depression with personality disorder comorbidity using dialectical behavior therapy. *Cogn Behav Pract*. 2000;7:468–477.

McGlashen TH. Schizotypal personality disorder. Chestnut Lodge follow-up study: IV. Long-term follow-up perspectives. *Arch Gen Psychiatry*. 1986;43:329–334.

McGlashen TH. The Chestnut Lodge follow-up study. III. Long-term outcome of borderline personalities. *Arch Gen Psychiatry*. 1986;43:20–30.

Meuwissen-van Pol ECH, Rossi G, de Weerd-Spaetgens CMEE, van Alphen SPJ. Screening for personality disorders in geriatric medicine outpatients. *Eur Geriatr Med*. 2020;11(2):289–295.

*Oltmanns TF, Balsis S. Personality disorders in later life: questions about the measurement, course, and impact of disorders. *Annu Rev Clin Psychol*. 2011;7:321–349.

Paris J, Brown R, Nowlis D. Long-term follow-up of borderline patients in a general hospital. *Compr Psychiatry*. 1987;28:530–535.

Penders KAP, Peeters IGP, Metsemakers JFM, van Alphen SPJ. Personality Disorders in Older Adults: a Review of Epidemiology, Assessment, and Treatment. *Curr Psychiatry Rep*. 2020;22(3):14.

Reynolds K, Pietrzak RH, El-Gabalawy R, Mackenzie CS, Sareen J. Prevalence of psychiatric disorders in U.S. older adults: findings from a nationally representative survey. *World Psychiatry*. 2015;14(1):74–81.

Rosowsky E, Lodish E, Ellison JM, van Alphen SPJ. A Delphi study of late-onset personality disorders. *Int Psychogeriatr*. 2019;31(7):1007–1013.

*Schuster JP, Hoertel N, Le Strat Y, Manetti A, Limosin F. Personality disorders in older adults: findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *Am J Ger Psychiatry*. 2013;21(8):757–768.

Segal DL, Coolidge FL, Rosowsky E. *Personality Disorders and Older Adults: Diagnosis, Assessment, and Treatment*. John Wiley & Sons, Inc.; 2006.

Stevenson J, Datyner A, Boyce P, Brodaty H. The effect of age on prevalence, type and diagnosis of personality disorder in psychiatric inpatients. *Int J Geriatr Psychiatry*. 2011;26(9):981–987.

Thompson LW, Gallagher D, Czirr R. Personality disorder and outcome in the treatment of late-life depression. *J Geriatr Psychiatry*. 1988;21(2):133–153.

van Alphen SP, Engelen GJ, Kuin Y, Hoijtink HJ, Derksen JJ. A preliminary study of the diagnostic accuracy of the Gerontological Personality Disorders Scale (GPS). *Int J Geriatr Psychiatry*. 2006;21(9):862–868.

Videler AC, Hutsebaut J, Schulkens JEM, Sobczak S, van Alphen SPJ. A life span perspective on borderline personality disorder. *Curr Psychiatry Rep*. 2019;21(7):51.

Zweig RA, Agronin ME. Personality disorders. In: Agronin ME, Maletta GJ, eds. *Principles and Practice of Geriatric Psychiatry*. 2nd ed. Lippincott Williams & Wilkins;2011:523–544.

55.3j Alcohol and Substance Abuse in Older Adults

Steve Koh, M.D., M.P.H., M.B.A., Hilary Gould, Ph.D., Omar Ghosn, M.D., and Michael Hann, M.D., M.B.A., M.S., C.P.E., F.A.C.H.E., F.A.P.A.

The most recent United States and global data projects rapid population growth of older adults. An increase in longevity contributes to this growth, even though countries with higher socioeconomic status have shown a decrease in fertility rates. In 2019, approximately 16.5% of the U.S. population was 65 years or older; projections based on U.S. Census data indicate that by the year 2050, this number is expected to reach 22%. This aging trend is not unique to the United States. The United Nations Department of Economic and Social Affairs projects that those above age 65 will account for about 16% of the total global population by 2050. Even within the older adult group, the proportion of the oldest old (above age 80) is increasing. Furthermore, the global ratio of males to females is narrowing due to improvements in male life expectancy. In 2013, this ratio in those above 60 was 76:100. By 2050, this will increase to 87:100. Within the United States, there is a clustering of older adults with disability, in poverty, or socially isolated in central urban centers or in remote areas. In this context, there is a likelihood of higher rates of substance dependence. Population trends also indicate that the aging cohort of adults who are currently considered middle age will reflect the substance use and misuse trends of younger cohorts.

As the U.S. population aged, classically held views on alcohol and substance misuse have evolved in the clinical setting. Alcohol and drug misuse continue to significantly contribute to worldwide disability and negative health outcomes. Older adults who started misusing in their younger years are now facing severe consequences from medical, psychiatric, and social perspectives. Once considered an "invisible epidemic" of substance misuse in the older population, clinicians are now faced with higher rates of co-occurring disorders that are complicated by the natural aging processes and cognitive decline. Most recent data from the Substance Abuse and Mental Health Services Administration (SAMHSA) indicates overall substance use of 165.4 million Americans 12 years and older, with alcohol use accounting for 50.8%. While alcohol use continues to be the highest contributor to substance use, nicotine use continues at 21.1%, and illicit drug use is on the rise at 13%. The National Household Survey on Drug Abuse from 2016 data suggested that by 2020, 2.7 million Americans aged 65 years or older will misuse prescription drugs, and 4.4 million older adults will require some form of substance abuse treatment. Despite the rise in substance use among older adults, forms of ageism continue to be common in clinical practice.

Several common barriers impede the identification of substance misuse in older adults, some of which are outlined in Table 55.3j–1. Many stereotypes of the typical older adults with substance abuse are prevalent in clinical care. For example, an older patient needing addiction treatment is often deemed to have been a life-long heavy drinker with failing health related to alcohol use and a limited to a nonexistent social network. The relatively healthy older adult who is drinking "moderately" is not considered to be in need of help and is often not assessed fully for the potential effect of alcohol on overall health and well-being. Also, substance use is addressed in a cross-sectional manner (i.e., if the patient is using now), and past chronic use history is often ignored. As the baby boom cohort enters the geriatric population, unique challenges face providers; along with the increased size of this cohort with changes in age distribution, there will be higher rates of excessive drinking, binge drinking, and illicit drug use.

Table 55.3j–1.
Barriers to Detection of Substance Misuse in the Geriatric Population

Barriers to the Clinician	Barriers to the Older Adult
Assumptions of lower prevalence of substance misuse in the geriatric population	Poor insight into substance misuse and reluctance to voluntarily seek help
Failure to screen appropriately for substance misuse in the older population	Lack of knowledge about the mental health system and treatment options
Masking of symptoms of substance misuse by other comorbid physical and mental/cognitive ailments	Feelings of shame, denial, desire to continue using, pessimism about treatment and recovery
Discomfort with addressing substance misuse with older adults	Cognitive problems—substance-induced amnesia, underlying neurocognitive disorder
Absence of collateral information from family members and caretakers	Family members and/or caretakers may not adequately report concerns of substance misuse
Assumption that older adults may not benefit from substance misuse treatment and prevention	Belief that other medical issues may be more important to address than substance misuse

Adapted from Crome I, Dar K, Janikiewicz S, Rao T, Tarbuck A. *Our Invisible Addicts: First Report of the Older Persons' Substance Misuse Working Group of the Royal College of Psychiatrists*. College Report CR 165, 2011.

An additional potential barrier to identification is the relative rarity of severe use disorders in older adults, where the prevalence of moderate to severe alcohol use disorder (AUD) is less than 4% and even lower in women, and illicit drug use disorders are currently less than 1%. On the other hand, 10% to 20% of older adults are drinking above recommended limits and are therefore at risk of having or developing problems even if they do not meet the criteria for a use disorder.

For older adults, a lack of understanding regarding the definition of at-risk use and who may benefit from treatment interferes with assessment or treatment-seeking behavior, which may be further compounded by denial. Similar to providers, patients may perceive physical symptoms (e.g., fatigue, sleep problems, anxiety, confusion) as normal or attribute them to a comorbid condition. Furthermore, age and substance-related declines in memory may result in underreporting of past and current substance use.

Mr. Doe is a 74-year-old man who was evaluated in the Emergency Department (ED) after a fall in which the patient sustained a hip fracture. In addition to hip pain, the patient endorses some ill-described upper abdominal discomfort. The patient's blood alcohol level (BAL) was 0.04 in the ED. The patient lives alone but takes care of himself and has a small social network of a few friends. He has a family, but they do not live nearby, and he is financially independent. When asked about general health habits, he reports that he does not smoke but drinks every day at dinner and before bed. The ED physician asked the CAGE questions, which upset the patient, who responded negatively to all items. The patient is taking hydrochlorothiazide and Lisinopril for hypertension and was recently prescribed Ambien as a sleep aid and a histamine antagonist for gastritis by his primary care physician (PCP). The ED physician documented that the patient's fall was likely secondary to the use of Ambien and declined to investigate the patient's drinking habits further due to the patient's irritation with the topic.

A review of the patient's medical record indicated that the patient received a mental health evaluation 6 months earlier at the request of his family. The family was worried about him living alone and wanted him to move into an assisted living facility. He was determined to be able to perform daily activities, although he was somewhat slow, and he rarely left home. There was no sign of dementia, but there was minor impairment in reaction time and problem-solving. When asked about alcohol use, he reported that he drank one drink at dinnertime and two drinks of sherry before bed. He described this pattern of drinking since he was widowed 15 years previous.

The case study illustrates several diagnostic dilemmas faced by clinicians who treat older drinkers:

- Mr. Doe did not seem to meet any dependence criteria.

- The patient, the ED physician, and the psychiatrist failed to make the connection that alcohol misuse was probably contributing to Mr. Doe's gastritis and to his presentation in the ED.

Mr. Doe was, however, drinking well above the recommended limits and was, therefore, an at-risk drinker, even if he did not present as the typical "alcoholic."

Take-home points:

- Diagnostic criteria and symptoms for alcohol and prescription drug misuse are not as applicable for older adults and may be confounded with comorbid medical illnesses.
- Lack of recognition by health care providers is often compounded by a lack of understanding of the benefits of reduced substance use when the patient is not physiologically dependent, or there is a general belief that there are few, if any, effective, accessible treatments to target substance misuse.
- Improved dissemination of current treatment methods and their efficacy, as well as continued development of other effective treatments, is essential.

HISTORY

Substance misuse among older adults as a distinct entity was first described in the research literature in 1964. In the 1960s and 1970s, researchers believed that substance misuse was rare in older adults and thus devoted little attention to this population as compared to younger adults. Health providers assumed that individuals with life-long addictions either died early or spontaneously improved in later life, and older adults with alcohol misuse were considered to likely have had a long history of drinking. Even the DSM-III-R, published in 1987, reads: "In males, symptoms of alcohol dependence or abuse rarely occur for the first time after age 45." Thus, until the past three decades, alcohol and drug misuse in older adults was rarely discussed in substance abuse or in the gerontologic literature. As the frequency of substance problems climbs in older adults, while it remains largely underdiagnosed and undertreated, attention to this issue has been increasing over time. Since 2013, there have been over 34,000 PubMed publications on the topic of older adults and substance use, with a peak number of publications in 2019. In routine primary care practice, however, screening and preventative measures for older populations are often still missing. More recently, studies have been conducted examining alcohol and prescription misuse in older adults and investigating treatment approaches better suited to this population. Recent studies indicate that while alcohol use continues to be the most common, prescription drug misuse has been on the rise and accounts for number two. Most importantly, prescription opioids and benzodiazepines are most concerning due to polypharmacy, sensitivity, and risks of harm, overdose, and death.

DSM-5 TERMINOLOGY

The Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) criteria for substance use disorder are considered identical for younger and older adults. Whereas distinctions were made between substance abuse and dependence in previous editions of the DSM, DSM-5 combines these diagnoses into a single disorder measured on a continuum from mild to severe. The essential feature of a substance-use disorder is a cluster of cognitive, behavioral, and physiologic symptoms indicating that the individual continues using the substance despite significant substance-related problems. Each of 10 separate classes of substances is addressed as a separate use disorder (e.g., AUD, stimulant use disorder) and is assessed clinically along a severity spectrum scale.

DSM-5 defines substance use disorder as a maladaptive pattern of substance use occurring within the previous 12 months leading to clinically significant impairment or distress. Diagnostic criteria include recurrent substance use resulting in failure to fulfill major roles or obligations, recurrent substance use in physically hazardous situations, and continued substance use despite persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of the substance. Other criteria include the presence of tolerance and withdrawals, intense cravings, consumption of the substance in larger amounts or time periods than intended, persistent desire or unsuccessful attempts to cut down, and excessive time spent obtaining, using, or recovering from the substance. Finally, an individual with a substance use disorder may reduce or sacrifice important social, occupational, or recreational activities and continue to use the substance despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance. A patient's substance-use disorder is categorized as "moderate" if they meet two or three criteria and "severe" if they meet four or more criteria.

For the purposes of terminologic clarity and to reflect the transition to DSM-5 criteria, the term *abuse* will only be used generally in this chapter and will not be used to refer to the clinical entity described in previous versions of the DSM. However, *physiologic dependence*, which is defined as evidence of tolerance or withdrawal (Severity Spectrum Scale Items 4 and/or 5), continues to be clinically relevant and will be used in this chapter in accordance with the DSM-5 conceptualization of this term.

Although the diagnostic criteria for substance-use disorders are the same for younger and older adults, some of the criteria may be less relevant for the latter. For example, problems related to work or the family may not be relevant for a retired or widowed older adults. Older women tend to drink at home alone and thus would be unlikely to have legal or work-related difficulties. Similarly, an older adults with physiologic dependence on a benzodiazepine with chronic insomnia may have no other symptoms of a sedative, hypnotic, and anxiolytic use disorder and may be more prone to falls and cognitive impairment from the benzodiazepine use itself. On the other hand, health problems are often more prominent in the geriatric population and may be a motivating reason for changing alcohol and substance use.

ALCOHOL

Spectrum-of-Use Categories

As the DSM-5 criteria may be less ideally applied for older adults, alternate categories of substance use may be more appropriate, particularly for alcohol use. Individuals who drank less than one or two alcoholic drinks in the previous year are categorized as *abstainers*. Approximately 50% to 70% of older adults report abstinence, making this the most common late-life drinking pattern. The reasons for abstaining, however, may have implications regarding treatment and counseling about drinking. Although many individuals report life-long abstinence, other individuals stop drinking alcohol due to the onset or presence of an acute or chronic illness, while others abstain due to a previous history of alcohol problems. Identification of the latter group allows preventive monitoring and screening to help avoid future relapse and also alerts the clinician to increased vulnerability for mental health problems and cognitive declines. *Low-risk*, *social*, or *moderate drinkers* include individuals who drink within the recommended guidelines and are not using any contraindicated medications or having any problems related to use.

An *at-risk* or *excessive drinker* is defined as someone who does not meet AUD criteria but is drinking at a level that increases the likelihood of problems or negative consequences. Drinking above recommended limits would be included in this category (more than one drink per day and, in particular, more than three drinks in a day). *Problem drinking* is the term often used in place of abuse for older adults drinking at a level that is causing detrimental medical, psychological, or social effects.

Table 55.3j–2.
Common Signs and Symptoms of Potential Substance Misuse and Abuse in Older Adults

Anxiety
Blackouts, dizziness
Depression, mood swings
Disorientation
Family problems
Financial problems
Headaches
Incontinence
Increased tolerance to alcohol or medications
Legal difficulties
Memory loss
New difficulties in decision making
Poor hygiene
Falls, bruises, and burns
Idiopathic seizures
Sleep problems
Social isolation
Poor nutrition

Adapted from Barry KL, Blow FC, Oslin DW. Substance abuse in older adults: Review and recommendations for education and practice in medical settings. HRSA/AMERSA monograph; 2002, http://www.informaworld.com

The Spectrum-of-Use Categories and the DSM-5 Severity Spectrum Criteria are both based on knowledge from young and middle-aged adults, and the applicability and validity in older persons are questionable. Physiologic tolerance may occur in older adults without increased consumption over time due to the physiologic effects of aging. DSM-5 diagnostic criteria are also susceptible to self-report biases, including memory impairment, lack of insight, and unwillingness to acknowledge symptoms. Other signs and symptoms may alert the physician to potential alcohol and substance misuse in older adults (Table 55.3j–2). In addition, collateral information, particularly when cognitive impairment is present, from family, friends, employers, or others is an important source of information to supplement or confirm the patient's report.

Epidemiology

Several epidemiologic studies indicate that the prevalence of AUDs in the geriatric population is underestimated. This represents an unfortunate bias, as AUDs in older adults are shown to have a greater probability of responding to treatment than those developed at an earlier age. While drinking tends to decrease with age, recent studies have shown that aging cohorts' rate of drinking is rising. Furthermore, the aging baby-boom generation, with its higher rates of alcohol and substance use, is more likely than previous generations of older Americans to misuse alcohol. About 20% of adults aged 60 to 64 and around 11% over age 65 report current binge drinking (defined as five or more drinks on the same occasion in the past 30 days). It should be noted that these are definitions developed for the general population and are not those used for older adults. AUDs afflict 1% to 3% of the older adult population and significantly worsen their physical, psychiatric, and social condition. In addition, up to 30% of inpatient geriatric patients on medical floors and up to 50% of geriatric psychiatry admissions present with a concurrent AUD.

Gender-related changes in alcohol consumption are also quite relevant diagnostically in screening for alcohol misuse. Women have historically tended to curb their drinking as they age more than men; however, data analyzed from the past two decades suggest that the prevalence of drinking and binge drinking is increasing at a more rapid rate than it is for men. Older women are more susceptible than their male counterparts to the adverse effects of alcohol misuse, including medical morbidity, and are more apt to have comorbid medical conditions, depression, anxiety, and abuse of psychoactive medications. Furthermore, as men live longer, it will be important to continue the screening for alcohol misuse and consider preventative measures.

Rates of AUDs are greater in health care settings relative to community-based samples. Mental health clinics also have higher rates of AUDs compared to the community. Nevertheless, alcohol use and abuse in older people are often overlooked in the clinic. Ageism can unconsciously affect care givers, who normalize the patient’s complaints. Sleep problems can often be misperceived as always being a bad sleeper or a low appetite can be attributed to always being a picky eater. Older adults may also be wary of stigma and minimize their alcohol use, fearing being identified as “alcoholic.” On busy clinic days, symptoms that can be suggestive of an AUD, like poor energy and changes in mood, may be quickly explained and credited to a mood disorder or a medical condition. Furthermore, older women’s alcohol use can be overlooked secondary to stereotyping.

Etiology

The etiology of problem drinking in older adults has not been well studied. It is a common myth that older problem drinkers have a life-long history of alcohol-related problems. This myth is more accurate for illicit drugs and nicotine, whereas as many as one-third of older adults seeking treatment for alcohol misuse develop their problems in later life. Thus, although drinking patterns are often stable over time, increased drinking in later life is not unusual due perhaps in part to more free or unstructured time, a rise in disposable income, and losses of loved ones. Older adults also potentially face several stressful life transitions (e.g., retirement, menopause, the “empty nest” phenomenon, loss of friends), worsening medical conditions (e.g., chronic pain, multiple medical conditions leading to a reduction in mobility and decrease in functioning), and the adoption of new roles (e.g., widow[er], caretaker) as they approach an older age, thereby increasing the vulnerability to alcohol misuse. Treatment approaches for late-onset and life-long problem drinkers are currently the same; however, the distinction is important for clinicians as the former are more likely to go undiagnosed and, therefore, untreated.

Recommended Limits

Certain age-related factors, including physiologic factors that increase sensitivity to alcohol and drugs in older adults, have led the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the Center for Substance Abuse Treatment (CSAT) to set lower recommended limits for alcohol use for older adults. The maximum drinking limit for women and men over the age of 65 is no more than an average of one standard drink per day, or seven standard drinks per week which is 50% less than the recommended limit for younger men. Table 55.3j–3 provides a description of how to measure a standard drink. In addition, older adults should not consume more than three standard drinks in 1 day (i.e., binge drinking).

Aging leads to a decrease in lean body mass and total body water, an increase in blood–brain barrier permeability, and a heightened neuronal receptor sensitivity to alcohol. These physiologic changes increase the blood alcohol concentrations of older adults and lead to worse impairment compared to younger adults drinking the same amount. Another physiologic change that contributes to 20% to 50% higher alcohol blood levels in older adults is a reduction in the activity of gastric and liver aldehyde dehydrogenase. In addition, older adult individuals are subject to polypharmacy which puts pressure on the microsomal ethanol oxidizing system (MEOS) in the liver. The prescribed medications compete with alcohol to be metabolized, which ensures both alcohol and drugs accumulate in the bloodstream and lead to further toxicity.

Table 55.3j–3.
Alcohol Conversion Chart

"1 Standard Drink"

Beverage Type	Quantity
Beer	12 oz
Wine	5 oz
Fortified wine	3 oz
Hard liquor (80-proof distilled spirits)	1½ oz (i.e., "a shot")
Malt liquor	8 oz
Liqueur or aperitif	4 oz

Additional Conversions

Beverage Type/Quantity	# Standard Drinks
Beer:	
▶ 1 (6) pack of 16 oz cans/bottles	8
▶ 1 quart	3
▶ 1 liter (33.8 oz)	2.8
Wine (e.g., red, white, chianti):	
▶ 1 bottle (25 oz)	5
▶ 1 bottle (40 oz)	8
▶ Fifth	6
▶ 1 magnum	12
▶ ½ gallon	16
Fortified wine (e.g., sherry, port):	
▶ Pint	7
▶ Fifth	12
Hard liquor (e.g., rum, gin, tequila, vodka):	
▶ Fifth	18
▶ Pint	11
▶ Quart	20
▶ ½ gallon	40
▶ 1 bottle (25 oz)	17
▶ 1 bottle (40 oz)	27
Malt liquor (e.g., Old English):	
▶ 12 oz	1.5
▶ Forty (40 oz)	5

Screening

Utilizing brief, free, and validated instruments can be useful for screening for alcohol misuse to help in the prevention or early intervention of older adults at risk. If screening indicates misuse, further assessment and detail is needed to accurately diagnose a substance use disorder and develop an individualized treatment plan. Screening measures may be administered by verbal interview or by paper-and-pencil or computerized questionnaire with equal reliability and validity. Additional information may be helpful with screening, including attending to changes in behavior, mood, cognition, or appearance; unexplained symptoms; slurred speech; and lab results, including liver-function test abnormalities. If screening instruments are not sensitive to older age, it is generally recommended to reduce the threshold. Screening should also assess the quantity and frequency of alcohol use and is often more accurate when administered by an interviewer versus self-administration due to variation in individuals' definitions of a standard drink (see Table 55.3j–3 for a description of a standard drink).

Alcohol misuse may be a contributory factor to all of the geriatric common presenting problems in primary care: impaired intellectual function, immobility, instability, and incontinence. The PCP is in a prime position to implement those screening instruments to monitor the patient regularly, and even if the patient has stopped alcohol use, there may be the temptation to substitute with other substances such as prescribed medications.

In the PCP's office, a variety of presentations should raise suspicion and warrant screening for AUD:

- Worsening of an otherwise stable chronic medical condition (diabetes mellitus type II, gastritis, gout, hypercholesterolemia, hypertension, macrocytic anemia, osteoporosis, and Parkinson disease);
- A change in a stable medication regimen efficacy;
- Acute emergent medical problems like aspiration pneumonia, dehydration, fecal or urinary incontinence, gastrointestinal disorders, heart failure, hypothermia, malnutrition, orthostasis, syncope, and unexplained falls;

- Onset or worsening of psychiatric disorders anxiety, confusion, dementia, depression, insomnia, and Wernicke–Korsakoff syndrome.

The most appropriate screening assessments used in the geriatric population are the Short Michigan Alcoholism Screening Test–Geriatric version (SMAST-G) and the Alcohol Use Disorders Identification Test (AUDIT). The SMAST-G was developed specifically for screening older adults in a variety of settings, and its psychometric properties exceed those of other screens for the identification of AUDs in this population (Table 55.3j–4). The AUDIT and the abbreviated version, the AUDIT-C (Table 55.3j–5), assess the frequency of drinking and binging in the past year. The AUDIT-C score ranges from 0 to 12, with a higher score indicating an increased likelihood of alcohol use impacting health and safety. For older adults, a score of 2 or greater suggests the need for further assessment or recommendation of a brief alcohol intervention.

Table 55.3j–4.

Short Michigan Alcohol Screening Test—Geriatric Version (SMAST-G)

In the past year:	YES	NO
1. When talking with others, do you ever underestimate how much you actually drink?	(1)	(0)
2. After a few drinks, have you sometimes not eaten or been able to skip a meal because you didn’t feel hungry?	(1)	(0)
3. Does having a few drinks help decrease your shakiness or tremors?	(1)	(0)
4. Does alcohol sometimes make it hard for you to remember parts of the day or night?	(1)	(0)
In the past year:		
5. Do you usually take a drink to relax or calm your nerves?	(1)	(0)
6. Do you drink to take your mind off your problems?	(1)	(0)
7. Have you ever increased your drinking after experiencing a loss in your life?	(1)	(0)
8. Has a doctor or nurse ever said they were worried or concerned about your drinking?	(1)	(0)
9. Have you ever made rules to manage your drinking?	(1)	(0)
10. When you feel lonely, does having a drink help?	(1)	(0)
TOTAL SMAST-G SCORE (0–10)		

Two or more positive responses are indicative of an alcohol use problem.

Reprinted with permission of the University of Michigan Alcohol Research Center.
© 1991 The Regents of the University of Michigan.

Table 55.3j–5.
AUDIT-C Alcohol Screening

1. How often did you have a drink containing alcohol in the past year?

—	Never	(0 points)
—	Monthly or less	(1 point)
—	Two to four times a month	(2 points)
—	Two to three times per week	(3 points)
—	Four or more times a week	(4 points)

If you answered “never,” score questions 2 and 3 as zero.

2. How many drinks did you have on a typical day when you were drinking in the past year?

—	1 or 2	(0 points)
—	3 or 4	(1 point)
—	5 or 6	(2 points)
—	7 to 9	(3 points)
—	10 or more	(4 points)

3. How often did you have six or more drinks on one occasion in the past year?

—	Never	(0 points)
—	Less than monthly	(1 point)
—	Monthly	(2 points)
—	Weekly	(3 points)
—	Daily or almost daily	(4 points)

Possible range = 0–12. For older adults, a score of 3 or more is considered positive.

Other screening instruments that are commonly used in younger patients are not ideal for older adults for various reasons. The CAGE questionnaire is the most widely used alcohol-screening test. However, while the CAGE questionnaire has shown high specificity for detecting AUDs, there is a relatively low sensitivity for problem drinking. Poor insight into drinking is often a barrier for older adults' self-reporting of the amount of drinking, and for isolated older adults, there may not be enough social interaction to report if others are criticizing the drinking pattern. If a CAGE questionnaire is used, a modified version that only inquires about recent problems is most appropriate for older adults, and the threshold for indicating possible problems is one positive response.

The CAGE questionnaire and the SMAST-G can be used together to increase their sensitivity. Some laboratory markers of alcohol abuse can also be helpful in screening: liver enzymes like aspartate aminotransferase (AST) and γ-glutamyl transferase (GGT), mean corpuscular volume (MCV) of red blood cells, and specific alcohol biomarker carbohydrate-deficient transferrin (CDT). The sensitivity of MCV and GGT for alcohol use detection is higher in older adults compared to their younger counterparts. Laboratory values may indicate long-term abuse or recent use or misuse, including blood alcohol or blood acetate, a metabolite of alcohol. High blood levels of alcohol without signs of intoxication indicate high tolerance to alcohol.

Differential Diagnosis

Clinicians must be careful when diagnosing psychiatric conditions that occur only in the presence of intoxication or withdrawal, as both of these states may mimic psychiatric symptoms, including depressive, anxiety, and psychotic symptoms. Diagnosing based on historical self-report of symptoms leads to questionable reliability and validity, particularly with older patients and alcoholics who may have cognitive deficits as part of their illness.

Alcohol problems and depression are common comorbid conditions in late life. Depression symptoms can precede (primary or independent depression), occur simultaneously (indeterminate), or develop as a consequence (reactive or secondary depression) of significant alcohol use. Repeated use of depressants can induce mood disturbance and may affect the mood enough to meet the criteria for a major depressive episode. A diagnosis of primary depression may lead to a treatment plan targeting both disorders, whereas a secondary depression diagnosis would suggest treating the addiction and monitoring the depression. In the latter case, treatment with antidepressants is not likely to be effective, whereas termination of substance use can resolve mood disturbance. In one notable study of 42 patients who began detoxification with significant symptoms of depression, only 6% reported continued significant depressive symptoms at 4 weeks.

The high prevalence of chronic comorbid medical conditions in older adults may confuse the clinical picture and affect the ability of the physician to conclusively determine deteriorating functioning due to substance misuse. Alcohol problems contribute to cognitive deficits in later life. Wernicke–Korsakoff syndrome has a specific triad and is mostly caused by alcohol dependence. Alcohol-related dementia, however, can be difficult to delineate from Alzheimer disease due to overlapping symptoms and presentation. Dementia in older adult alcoholics is five times more prevalent than in abstinent older adults What’s more, around a quarter of patients who suffer from dementia have a comorbid AUD. In older adults, sedative intoxication can also be confused with progressive dementia. Withdrawal of the substance tends to stabilize cognitive and functional status for substance-related cognitive impairment but not for dementias.

Sleep disorders also frequently co-occur with excessive alcohol use. Alcohol use and aging are both associated with changes in sleep patterns. Alcohol has been shown to decrease sleep latency, decrease stage IV sleep, and precipitate or aggravate sleep apnea. Age-related changes can be exacerbated by alcohol use, depression, comorbid medical conditions, and medication use, including increased rapid eye movement (REM) episodes, decrease in REM length and stage III and IV sleep, and increased awakenings. On the other hand, abstinent alcoholics with poor sleep due to insomnia, frequent awakenings, and REM fragmentation demonstrate normalization of sleep periodicity and temporary suppression of REM sleep upon drinking, suggesting that alcohol may also be used to self-medicate sleep disturbances. Differential diagnosis, again, involves a historical account of the initiation of sleep problems and determining if symptoms subside with substance abstinence.

Course and Prognosis

Drinking patterns often are stable over time, but increased drinking in later life is not unusual due to factors idiosyncratic to an older population. In addition, aging patients can experience increased problems due to alcohol without increased consumption due to physiologic changes that increase sensitivity to alcohol and drugs, increased medication use, and the development of chronic health conditions.

Alcohol misuse is related to adverse effects in several physical health and psychiatric domains. Intoxication can lead to an increased risk of falls and motor vehicle accidents, especially in conjunction with medication interactions or other substance misuse. Chronic alcohol misuse is associated with multiple disease outcomes in several organ systems. More than 90% of older patients with a diagnosis of alcohol misuse have a history of depression and depressed older adults are three to four times more likely to have an AUD compared with nondepressed older adults. Like depression, late-life anxiety disorders can both be a consequence of problematic alcohol use as well as a primary psychiatric illness. Generalized anxiety disorder is the most common type of anxiety disorder in the older population, and those with the disorder are more than twice as likely to have a substance use disorder as those without an anxiety disorder.

Following depression, substance use disorders, particularly AUDs, are the second most common category of psychiatric disorders associated with completed suicide in the older population. Drinking among older adults elevates suicide risk through interactions with many factors that are more prevalent in this age group than in younger adults, factors such as depressive symptoms, medical illness, negatively perceived health status, and low socioeconomic support. Together, alcohol misuse and comorbid psychiatric illness are a potentially lethal combination. If appropriately addressed, however, positive outcomes and recovery can be expected. Studies have shown that more than 20% of treated older adult alcohol-dependent patients remain abstinent after 4 years.

Consequence of Moderate Use

Before discussing the negative prognoses for alcohol and substance abusers, it is important to acknowledge that moderate alcohol consumption may have positive effects on physical health and mental well-being. Low-risk alcohol consumption has been linked with a reduction in cardiovascular disease risk for men and women. For older adults without cardiovascular disease, moderate alcohol use has been associated with lipoprotein subclass distribution, including fewer small, low-density lipoprotein particles and increasingly large and medium-sized high-density lipoprotein particles. A cardiovascular protective effect of light to moderate alcohol consumption (1 drink/day for women or up to 2 drinks/day for men) has been reported in both the adult and older adult populations, but the findings are largely correlational. In addition, moderate alcohol use has been correlated with a lower incidence rate of type 2 diabetes mellitus in older adults and may reduce the risk of cognitive impairment and dementing illnesses.

Despite these potential health benefits, abstinence is recommended for individuals taking particular medications and those with a history of alcohol or drug abuse. In addition, although moderate drinking may protect against certain health problems, any alcohol consumption may be detrimental after a diagnosis of a chronic health condition (such as cardiovascular disease or diabetes). Furthermore, there is no evidence that initiation of drinking by nondrinkers will translate into reduced health risks, and moderate use may contribute to several negative health outcomes. Moderate alcohol consumption may reduce stroke risk due to blocked blood vessels but may increase stroke risk due to hemorrhaging or bleeding. The risk of breast cancer in women is increased by about 50% in women who drink three to nine drinks per week compared to fewer than three drinks per week. However, other sources show that even three to six alcoholic drinks per week are an independent risk factor for breast cancer in older women. In addition, there is a linear relationship between alcohol use and bone mineral density at the hip. Even limited alcohol consumption has been shown to impair one’s driving ability and may increase the chance of accidents and fatal injuries due to falls, motor vehicle accidents, and suicides. In addition, depression, memory problems, and other cognitive changes, liver disease, cardiovascular disease, and sleep problems have been associated with moderate alcohol use. Mental health providers also need to be cognizant that alcohol can negatively impact other mental health disorders. Even in the moderate use range, greater alcohol consumption adversely affects depression treatment. The converse is also true; the reduction of moderate alcohol use during depression treatment enhances treatment outcomes. Limited alcohol use may also negatively impact the course and prognosis of dementing illnesses and may precipitate or exacerbate the onset of personality changes and behavioral disturbances in patients with dementia. Depression, sleep problems, memory, and cognitive disturbances, and hepatic and cardiovascular diseases, have all been associated with moderate alcohol use.

Consequence of Severe Use

Severe AUD is associated with an increased probability of morbidity and mortality from disease-specific disorders, including acute pancreatitis, alcohol-induced cirrhosis, and alcohol-related cardiomyopathy. Older adults are also more susceptible than young persons to having worse hypertension in the setting of severe alcohol use. In addition, individuals having more than six standard drinks per day for a duration of more than 5 years are at an increased risk of heart failure secondary to nonischemic dilated heart disease and ventricular dysfunction. Alcohol also impacts the central nervous system (CNS) directly and indirectly. The prevalence of dementia is almost five times higher in severe AUD than in people with no AUD, and about 25% of the geriatric patients with dementia present with comorbid AUD. Alcohol directly modifies cell membrane fluidity, increasing the inhibitory activity of GABA receptors and inhibiting the excitatory activity of NMDA receptors. Alcohol also has indirect effects by causing thiamine, nicotinic acids, B vitamins, and folate deficiencies. Direct neuronal damage from alcohol leads to alcohol dementia, while indirect neuronal damage through thiamine deficiency causes Wernicke–Korsakoff syndrome and Marchiafava–Bignami disease.

Comorbidity of substance misuse with other medical and psychiatric conditions may impact the course, treatment, and prognosis of this condition. Substance misuse in earlier life can also impact later life, including increasing the likelihood of developing late-life mental health problems. Comorbid depression and alcohol misuse are not only common in later life but may have a reciprocal effect on each other. Depressed alcoholics are more likely than nondepressed alcoholics to have a complicated course of depression, manifested by increased suicide risk and social dysfunction. In addition, a history of alcohol abuse as a younger adult is predictive of a more severe and chronic depression course in later life. Almost 20% of people above the age of 65 who suffer from depression have a concomitant AUD. On the other hand, more than 90% of older subjects with AUDs also suffer from depression. Due to the complexity of the relationship, it is hard but important for a successful treatment plan to differentiate which disorder preceded or is driving the other. Moreover, dual diagnoses of internalizing disorders (depression and anxiety) and AUDs are the cause of more than 70% of completed suicides. They are also major contributing factors to household distress: domestic abuse and violence, separation and divorce, and socioeconomic drift. Depressed patients with AUD who successfully become abstinent use have progressively markedly improved psychiatric conditions compared with those who continue to drink. Vice versa, addressing anxiety and depression is also a critical target when treating AUDs.

Pharmacologic Treatment

Similar to the general population, AUD treatment for older adults consists of two phases, an acute detoxification strategy and a long-term prevention approach. The risk of complications, however, is higher for the older adult population. During alcohol withdrawal, older adults are at higher risk of both medical (e.g., cardiac arrhythmias and ischemias, fluctuating blood pressure, aspiration pneumonia) and neurologic complications (e.g., alcohol hallucinosis, delirium tremens, confusion, seizures, and Wernicke–Korsakoff syndrome) compared to the younger population, requiring closer medical monitoring. Wernicke–Korsakoff, characterized by peripheral neuropathy, ataxia, ocular paralysis, confusion, and confabulation, requires particular attention. Triggered by low serum levels of magnesium, Wernicke–Korsakoff is best treated with magnesium and thiamine supplementation for 5 days. Alcohol withdrawal is best managed with a benzodiazepine replacement and taper during the duration of the treatment. Scales like the Clinical Institute Withdrawal Assessment of Alcohol (CIWA) and Minnesota Detoxification Scale (MINDS) are used as a guide for health care providers to dose the benzodiazepine to avoid both excessive sedation from high doses or withdrawal complications from low doses. Short-acting benzodiazepines like lorazepam and oxazepam are recommended in older adults to avoid further risks of sedation. Nonbenzodiazepine options are less commonly used but remain useful in the management of agitation or in reducing the burden of high benzodiazepine requirements. Antipsychotics like haloperidol are often used in the case of severe agitation and hallucination. Beta-blockers such as atenolol and alpha-2 agonists like clonidine have been useful in managing agitation in the setting of tachycardia and hypertension. They, however, need to be judiciously dosed to avoid bradycardia and hypotension. Gabapentin and phenobarbital are most commonly used when a previous history of delirium tremens and intensive care unit hospitalization has been documented. Both medications enhance GABA-mediated activity, complementing the action of benzodiazepine to reduce excessive dosing in the acute phase.

After completing the detoxification phase, the challenge of remaining abstinent and preventing relapse arises. Evidence is suggestive that older adults have higher success remaining abstinent compared to their younger counterparts. More than 20% of treated alcohol-dependent patients remain abstinent 4 years later. What's more, older women seem to maintain abstinence better than men. Pharmacologic interventions are efficacious in reducing both the frequency and intensity of relapses, with modest effect sizes, however. Having a good safety profile, naltrexone is favored in relapse prevention strategies for older adults. Its antagonistic effect on the mu-opioid receptor is thought to decrease cravings and decrease the reward signals when drinking alcohol. Naltrexone's monthly injectable formulation may also come in handy to ensure compliance, especially when concerns of cognitive decline and forgetfulness are present. Acamprosate, an N-methyl-D-aspartic acid glutamate receptor inhibitor, is also a commonly used agent that has shown similar efficacy to naltrexone in preventing alcohol relapse and decreasing the severity of alcohol binges. Some studies have suggested synergistic effects and slightly improved outcomes when combining naltrexone and acamprosate. Pregabalin, topiramate, and baclofen are also commonly used for relapse prevention, albeit with less robust studies and evidence of efficacy.

TOBACCO

Epidemiology

Despite decreasing smoking trends, cigarette smoking kills more than 480,000 Americans every year and continues to be the leading cause of preventable disease and death in the United States. Approximately 70% of those deaths occur among those over the age of 60 years. Although often overlooked in older adults, smoking contributes significantly to poor health and disability in older adults. Smoking is the biggest risk factor for chronic obstructive pulmonary disease (COPD), which includes chronic bronchitis and emphysema. Smokers are 12 to 13 times more likely to die from COPD as compared to nonsmokers. Smoking contributes to frailty in older adults with chronic kidney disease and increases the risk of cardiometabolic, musculoskeletal, and cerebral complications. Among older adults, cigarette smoking also remains an important predictor of cancer mortality.

While smoking rates tend to decrease with age, one in five smokers is 50 years or older. Of community-dwelling individuals, 15.2% of adults aged 65 to 74 and 8.4% of those older than 75 are smokers. This rate is even higher in older African American men, where 22.1% of individuals 65 to 74 and 12% of 75 years old and older smoke. The importance of clinicians continuing to have a low threshold for smoking screening and working with patients toward smoking cessation does not change while taking care of older patients.

Etiology

Despite being wrong, impressions that late-life smoking cessation does not provide benefits and does not lead to overt changes in health outcomes may explain why older adults continue to smoke. In fact, it has been shown that compared with younger adults, a lower number of older adults express a desire to quit or make an attempt to quit smoking; however, they are more likely to succeed in their attempts. Adverse health events, including hospitalization or being diagnosed with a chronic illness, are associated with increased odds of smoking cessation among older adults.

Screening

Screening for tobacco should be included as part of the routine mental and physical health care for adults over the age of 60 and should be updated annually or whenever problems arise that may be related to tobacco use. There is no standard screening tool for tobacco in clinical settings, nor any adaptations for older adults. One method used for assessing the quantity and frequency of substance use (including alcohol, tobacco, and other substances) involves prospective recording and monitoring of use, retrospective accounts of daily use over a defined period of time (i.e., time-line follow-back method [TLFB]), or questions regarding average use. Although the prospective diary method is the gold standard, it is often impractical for screening assessments, as it requires more time and multiple visits. The TLFB method is the most commonly used technique in addiction studies, and 7-day TLFB assessments are highly correlated with the prospective method. Two difficulties arise with the TLFB. First, the week of measurement may not represent the typical week. Second, the TLFB method tends to underestimate use for less frequent users (compared to nonusers or frequent users). The TLFB also takes longer to administer than general questions and may vary due to variations in potency and modes of use.

Several tools currently exist that may be useful for screening tobacco use in older adults, including the Alcohol, Smoking, and Substance Involvement Screening Test (including ASSIST, ASSIST-LITE, and NIDA-Modified ASSIST) developed by the World Health Organization, the Tobacco, Alcohol, Prescription medication, and other Substance use (TAPS) Tool from NIAA, and the CAGE-AID which is adapted to include questions about other drugs besides alcohol. These measures screen for all categories of substance misuse. Compared to younger adults, older adults' reasons for quitting often include health risks to other family members, interactions with other medications, and changes in health status. Counseling on the health and longevity benefits of quitting is important to counteract commonly held misbeliefs that "the damage is done."

Course and Prognosis

Quitting smoking at a later age does not automatically cleanse the body from chronic smoke and toxin exposure and insults to the body. It can, however, reduce the risk of developing or dying from cancer or other medical conditions. Even quitting smoking at the age of 80 or older reduces the risk of developing or dying from cancer or other diseases. Among those diagnosed with cancer, smoking cessation improves treatment prognosis. On the other hand, continued smoking after a cancer diagnosis worsens the prognosis. It can lead to poorer response to treatment, increased chance of recurrence, increased chemotherapy side effects and toxicity, and increased chances of consequent primary tumors.

Impressions that late-life smoking cessation does not provide benefits and does not lead to changes in health outcomes are patently false. Even in older age, smoking cessation can decrease overall morbidity and improve quality of life. Older adults who quit smoking also experience many other health benefits, such as a lower risk of cognitive decline, a lower risk of polypharmacy, improved overall health outcomes, and improved quality of life. Despite the proven benefits, only 1 in 20 older adults attempting to quit smoking succeed. Smoking and nicotine addiction are chronic, relapsing conditions. Successful smoking cessation often happens following numerous interventions and attempts to quit. Treatment plans can be more successful when they include educating older adults about the benefits of quitting smoking. Older adults should also be encouraged to use evidence-based pharmacologic treatments, especially if they have failed to quit smoking on their own multiple times in the past.

Pharmacologic Treatment

Nicotine replacement therapy (NRT) is based on substituting the nicotine ingested by smoking with various forms of nicotine delivery methods with the goal of minimizing withdrawal symptoms when attempting to quit. NRT is currently the most studied pharmacologic treatment in older adults, and the limited available evidence suggests NRT is efficacious for smoking cessation among this population. NRT is available in different forms: patch, gum, inhaler, sublingual tablet, lozenge, and spray. All forms of NRT improve the chances of success by 50% to 70%, with no difference in effectiveness between the various forms.

The Food and Drug Administration (FDA) has also approved two non-nicotine–containing medications for smoking cessation: bupropion and varenicline. Varenicline facilitates smoking cessation through its partial agonist properties at the α4 β2 nicotinic acetylcholine receptor that both minimizes nicotine withdrawal symptoms and reduces the rewarding properties of nicotine. It is well tolerated in older adults and does not require dose adjustment based on age alone. Varenicline, however, should be used with caution in the presence of psychiatric disorders as it may worsen depression and increase the risk of suicidal thinking and behavior. Bupropion's effects on smoking cessation are thought to be independent of its antidepressant properties; its mechanism of action to decrease cravings has been attributed to the inhibition of the reuptake of dopamine in the CNS. Bupropion should be used with a reduced dose in older adults and should not be used in the case of comorbid epilepsy. Combining medications has synergistic effects and increases the chances of success but requires a cautious approach with the older adult to avoid the risks of polypharmacy.

Adverse health events, such as hospitalization or being diagnosed with a chronic illness, are associated with increased odds of smoking cessation among older adults. Among older adults attempting to quit smoking or who recently succeeded in quitting, more than a third (37%) used counseling and/or FDA-approved cessation medication to quit; 9% used at least one type of counseling from telephone quitlines, one-on-one counseling, cessation clinic, class, or support group. Very few used counseling without medication or used a specific type of counseling to have a reliable estimate. Thirty-four percent used at least one type of FDA-approved cessation medication, with the largest proportion, 29%, using NRT. Ten percent used varenicline, and too few used bupropion to be reliably estimated. Only 6% used both counseling and FDA-approved cessation medication. Cessation counseling, including individual, group, and telephone counseling, and FDA-approved cessation medications were found to be effective methods individually.

However, counseling and medications were found to be more effective when combined.

Because nicotine addiction is a chronic, relapsing condition, successful smoking cessation often requires repeated intervention and multiple attempts to quit. In particular, older adults may benefit from being informed by health care providers that they can still derive substantial health and quality-of-life benefits from quitting and that, even though they may have tried to quit unsuccessfully many times, they can still succeed in quitting, and that their chances of doing so are better if they use evidence-based tobacco cessation treatments.

TETRAHYDROCANNABINOL

Since 1970, marijuana has been classified as a Schedule I substance under the Controlled Substances Act (CSA). However, the United States is in the midst of a cultural shift regarding its perception of and the potential medicinal applications considered appropriate for marijuana. As of 2021, the medicinal use of marijuana is legal in 36 states, the District of Columbia, and four U.S. territories. Additionally, eight states, the District of Columbia, and two U.S. territories have passed legislation to regulate marijuana for nonmedicinal use. The wave of legalization is occurring in conjunction with a societal shift in the perception of marijuana. In the United States, there has been a general decline in stigma associated with marijuana use as well as a decline in the perceived risk of marijuana use, although these perceptions vary by state and region. The increasingly legal use of and easing access to marijuana, as well as a decrease in stigma associated with its use, is also affecting the substance use patterns of older adults.

Epidemiology

Marijuana is the most commonly used illicit substance among older adults, and older adults currently represent the fastest-growing demographic of marijuana users. According to a secondary analysis of data from the National Survey on Drug Use and Health, the prevalence of past-year marijuana use by older adults is estimated to have increased from 2.4% in 2015 to 4.2% in 2018. The rate of increase in marijuana use among older adults was particularly significant for women, racial/ethnic minorities, those with higher family incomes, and those with mental health problems. Multiple factors contribute to the rise in marijuana use by older adults, including a decrease in stigma and an increased interest in the possible benefits of marijuana for chronic health conditions. Additionally, the risk associated with the co-use of alcohol and marijuana in the older adult population was observed to be higher than the risk of using either independently.

Etiology

Older adults are increasingly using marijuana for medicinal purposes to treat conditions such as pain, insomnia, neuropathy, and anxiety. Although medicinal marijuana is typically well-tolerated in older adults, aging-related factors may pose specific risks. With advancing age, lean body mass and total water volume decrease relative to total fat volume, thereby decreasing total body volume and subsequently increasing the serum concentration, absorption, and distribution of alcohol and drugs in the body. Furthermore, reduction in liver mass and decreased hepatic blood flow results in slower drug metabolism, and changes in renal function can lead to the build-up of some substances. At the same time, the response to the CNS-active drugs is frequently exaggerated due to age-related decline in CNS function and increased pharmacodynamic sensitivity for some psychoactive drugs, thus amplifying the response to sedative-hypnotic and mind-altering drugs. The presence of dementia also increases the likelihood of intoxication and impaired functioning. Paradoxically, however, recent research suggests the possibility that cognitive functioning may actually improve when patients use medical marijuana due to factors such as improved sleep and pain control. However, age-related changes in drug sensitivity vary by drug, and pharmacodynamic changes associated with aging continue to be an area of investigation.

Screening

As marijuana increases in popularity, potency, and medicinal purposes, it is imperative to screen older patients for current use. When screening for substance misuse, using nonstigmatizing language and an empathic nonconfrontational style can be especially important for older adults. The White House Office of National Drug Control Policy released guidance around terminology in 2016, such as replacing labels like “addict” or “abuser” with the terminology of “substance use disorder,” “unhealthy use,” or “harmful use.” Cannabis use can be particularly challenging to obtain standardized measurements due to variations in the potency of strains, forms (e.g., flowers, oil, waxes), and modes of use (e.g., smoking, vaporizing, ingesting, topical). Similar to screening for alcohol misuse, many cannabis screeners target functioning and sequelae of use and do not take into consideration developmental changes in older adults. For example, the Marijuana Problem Scale measures negative consequences related to motivation and productivity, work and finances, physical health, and memory impairment which are easily confounded with the aging process. Most screening measures for problems related to marijuana use target adolescents and young adults. The Cannabis Use Disorder Identification Test-Revised (CUDIT-R) developed from the AUDIT is often recommended for use in older adults, although it has not been studied extensively. The CUDIT-R is a 20-item self-report measure of cannabis misuse in the past 6 months.

Risks

Older adults frequently suffer from chronic comorbid health conditions and are more likely to be taking multiple medications than the younger population. Polypharmacy and medication interactions in the context of marijuana use increase the possibility of medical complications in older adults using marijuana. Marijuana is chemically complex, but the two primary active ingredients are Tetrahydrocannabinol (THC) and Cannabidiol (CBD), both of which may alter hepatic metabolism and could alter blood levels of other medications. Older patients taking antiepileptic or blood thinners should be particularly cautious when considering marijuana use, as various studies have shown that blood levels of these medications may be altered when using marijuana. Marijuana use has also been shown to increase heart rate and blood pressure, although there is minimal evidence suggesting a link between marijuana use and an increased risk for cardiovascular events. Nonetheless, it is recommended to screen patients who use marijuana for cardiovascular disease.

Withdrawal

Although there is emerging research studying cannabis withdrawal, there is a paucity of research on cannabis withdrawal in older adults specifically.

Cannabis withdrawal syndrome (CWS), as defined in the DSM-5, occurs after the cessation of prolonged or heavy cannabis use. CWS symptoms include:

1. Irritability, anger, or aggression
2. Nervousness or anxiety
3. Sleep disturbance
4. Appetite or weight disturbance
5. Restlessness
6. Depressed mood
7. Somatic symptoms, such as headaches, sweating, nausea, vomiting, or abdominal pain

A recent meta-analysis suggests that up to 47% of patients who regularly use marijuana or demonstrate physiologic dependency experience CWS after discontinuing marijuana consumption. Cannabis withdrawal has also been associated with insomnia in one-third of patients after discontinuing cannabis. Cannabis withdrawal has been associated with lower total sleep time, worse sleep efficiency, longer sleep latency, shorter REM latency, lower slow wave sleep, and periodic limb movement. However, further research is required to study this phenomenon, more specifically in the older adult population.

Pharmacologic Treatment

Various studies have been conducted to evaluate the efficacy and safety of medical cannabinoids in a variety of conditions in adult patients, and in a substantial number of studies, older subjects were included. However, there is a general lack of evidence concerning the use of cannabinoids, specifically in older patients, resulting in a scarcity of data to guide treatment decisions. A review of studies performed that included older patients in the study population indicated no efficacy of medical marijuana on dyskinesia, breathlessness (versus placebo), chemotherapy-induced nausea and vomiting (versus prochlorperazine), or oral THC in the symptomatic treatment of behavioral problems in dementia. However, sample size and methodologic quality may have biased these conclusions. Further research on cannabinoids, specifically in the older population, is needed.

BENZODIAZEPINES

Epidemiology

Benzodiazepines are one of the most common classes of prescription medications in the United States. Benzodiazepine use increases with age and tends to be chronic in late life, which may increase the risks for fractures, subtle cognitive decline, and benzodiazepine dependence. According to the CSAT, 23% of all prescriptions written for older adults are for benzodiazepines. Despite the wide use of benzodiazepines, the American Geriatrics Society Beers Criteria recommends benzodiazepine use be limited in the older population to seizure disorders, REM sleep disorders, benzodiazepine withdrawal, alcohol withdrawal, severe generalized anxiety disorder, periprocedural anesthesia, and end-of-life care. Commonly, sleep disorder is the indication cited for benzodiazepine use in the older adult population, accounting for approximately 59% of total prescriptions; this is despite the fact that long-term benzodiazepine use does not help with sleep restoration. Sleep disorders occurring later in life are thought to be related to decreased activity of the suprachiasmatic nucleus, contributing to a disruption of circadian rhythms in older adults, including the nocturnal secretion of endogenous melatonin. However, the iatrogenic contribution of medications, including stimulants, antihypertensives, decongestants, chemotherapy, and respiratory medications, cannot be overlooked.

Etiology

While it is well established that benzodiazepines increase the risk of cognitive impairment, delirium, fall injuries, fractures, and motor vehicle accidents in older adults, there remains insufficient research on the impact of these medications on the older population. Studies on this topic are made more difficult for a variety of reasons, including a low overall 12-month prevalence of sedative, hypnotic, or anxiolytic use disorder diagnoses in older adults and less frequent referral to mental health resources than younger patients. This occurs despite the high rate of benzodiazepine prescriptions and commonly occurring clinical “red flags” of physiologic benzodiazepine dependence, including:

- Long-term use
- Rebound anxiety and insomnia on withdrawal from the drug
- Strong desire to use benzodiazepines
- Driving while under the influence of benzodiazepines
- Use of benzodiazepines despite falls
- Use of benzodiazepines in addition to other hypnotics
- Continuing use of benzodiazepines despite physicians’ recommendations to discontinue

Screening

In 2020, the U.S. Preventive Services Task Force revised previous guidelines and now recommends screening all adults for unhealthy drug use when appropriate treatment services can be offered. Benzodiazepines that were previously clinically appropriate for patients may become problematically related to changes in health status or aging. All prescribed medications should be reassessed as patients approach the age of 65, and prescription drug misuse should not be overlooked. As with all substances, attention to functioning and impairment is paramount, and problems can develop gradually and subtly. Specific to benzodiazepine use, assessing cognitive and memory-related impairments, diminished psychomotor abilities, deleterious events such as automobile accidents, falls, behavioral disinhibition, and worsening psychiatric symptoms are relevant. When problematic substance use or risk of misusing substances is suspected, it may be appropriate to provide routine toxicology screening (urine, saliva) to determine if additional assessment is necessary or to monitor progress. It is important to note, however, that urine drug screening does not detect all benzodiazepine use. Furthermore, there is consensus that routine use of laboratory tests as screening tools in the primary care setting is not recommended.

Differential Diagnosis

The DSM-5 diagnostic criteria for substance use disorders, including sedative, hypnotic, and anxiolytic use disorders, are the same for both young and older adults. However, differences in life circumstances may make the detection of a sedative use disorder more challenging for older adults. Older adults with physiologic dependence on a benzodiazepine with chronic insomnia may have no other symptoms of a sedative, hypnotic, and anxiolytic use disorder and may be more prone to falls and cognitive impairment from the benzodiazepine use itself. Age-based biases are also commonplace among physicians and caregivers and contribute significantly to the misdiagnosis of comorbid conditions and the misuse of benzodiazepines. Physicians and family members often overlook clinical signs of sleep disorders in older adults, including somnolence and fatigue, as signs of normal aging. Common sleep disorders, including sleep-disordered breathing and periodic limb movement syndromes, have been shown to benefit from nonbenzodiazepine evidence-based treatments. Overnight recording of the apnea–hypopnea index is indicated in patients with symptoms of sleep disorder breathing, including high snoring severity, unintentional napping, and/or excessive daytime sleepiness. Primary insomnia in older adults may be initially treated using numerous nonpharmaceutical tools to improve sleep hygiene, such as limiting naps, adjusting medications, avoiding stimulants like coffee, environmental improvements, examining sleep problems, and other specific treatments. Nonbenzodiazepines sedatives like zolpidem are often considered to be safe and effective in older adults, but recent research shows that they may lead to as many fractures, falls, and delirium as benzodiazepines.

Course and Prognosis

Despite clear indications that benzodiazepines are not the preferred agent for long-term treatment of anxiety, this is the second most common reason for the prescription of a benzodiazepine to an older adult, often in the absence of first-line antidepressant medications. Risk factors for anxiety disorder in older adults include female gender, low socioeconomic status, family history, external locus of control, the current level of perceived stress, poor primary support, and poor social support. While SSRI and SNRI medications are associated with falls and an increased risk of hyponatremia in older adults, their adverse effect profile is favorable in the chronic treatment of anxiety compared to benzodiazepines. Evidence for the nonpharmacologic treatment of anxiety in the older adult population is conflicting regarding cognitive behavioral therapy, physical therapy, and other behavioral approaches.

Withdrawal and Treatment

Abrupt or excessively rapid discontinuation of benzodiazepines can lead to benzodiazepine withdrawal, particularly if use has been long-term. Benzodiazepine withdrawal poses clinical risks, including rebound anxiety and seizures. Conversely, safely tapering benzodiazepines in the older adult population may have several benefits, including a decreased risk of falls. However, despite multiple guidelines and consensus statements advising against the use of benzodiazepines for older adults, prescription rates remain high. Effective clinical strategies for tapering long-term benzodiazepine use include combining clinical education and medication review as well as consultation focused on a clinical reassessment of benzodiazepine benefits and harms. Various tapering protocols have been studied, ranging from 4 to 22 weeks in duration; however, there is a general lack of research in older patients regarding the ideal rate for tapering benzodiazepines.

OPIOID ANALGESICS

Epidemiology

In the United States, the prevalence of substance use disorders in older exceeds 5.7 million people, with opioid use disorders increasing at an alarming rate. Data from multiple agencies, including the Centers for Disease Control (CDC), have documented an alarming rise in prescription opioid analgesic abuse with a concordant rise in attributed mortality. The opiate's impact on the older population is second only to alcohol, accounting for up to 22% of substance-related inpatient admissions. While a large percentage of the older population suffers from some form of chronic pain, often resulting in pain management with opioid analgesics, there is currently no evidence for the long-term effectiveness of opioid analgesics in the treatment of persistent noncancerous pain conditions in older patients. Despite growing concerns for opioid analgesic misuse, it is important to note that pain continues to be undertreated in adults.

Etiology

Declining physical health and physiologic changes associated with aging affect the metabolism of medications and, thus, the potential for medication misuse. Previous and comorbid psychiatric diagnoses, poor health status, social isolation, and female gender also increase the risk for prescription drug misuse. Many cases of psychoactive medication misuse have been shown to be unintentional on the part of the patient or physician. The drugs are usually obtained with an appropriate prescription for a specific health-related purpose, either a physical condition or emotional distress. An older patient may misunderstand the directions, their confusion may be compounded by multiple medications, or misuse may occur due to a prescription that is not appropriate for an older adult. Unintentional misuse can progress into abuse to obtain the desirable effects. Physiologic dependence on psychoactive medications without meeting additional DSM-5 substance use disorder criteria can also result from even therapeutic dosages for brief periods. Aging also involves increased sensitivity to over-the-counter and prescription medications, with women demonstrating even greater sensitivity.

Screening

Given the high utilization of medical services by older adults and because a large percentage do not self-identify as substance misusers, physicians and health care staff can be critical for identifying individuals with alcohol and drug problems. Screening can be verbal, written, computerized, and by general query or semistructured questions. Involvement of collateral or caretakers may be clinically appropriate. Assessing for dysfunction in age-appropriate domains and functioning, such as interference in social activities rather than problems with work, should be utilized. Screening measures previously mentioned, such as the ASSIST, TAPS, and CAGE-AID may provide clinical utility, although research studies in older adults are limited. Screeners specific to opioid misuse, opioid use disorder, and calculating benefit–harm ratios in older adults have not been well-studied or replicated. Some research suggests that the Opioid Risk Tool (ORT) has clinical validity in older adults and measures demographic variables and family and personal history to alert providers who may be at the most risk. If screening indicates misuse or concern for risk, further assessment is needed for diagnosis and treatment.

Withdrawal and Pharmacologic Treatment

Factors such as opioid withdrawal, opioid-induced hyperalgesia, and chronic noncancer pain make treatment of opioid use disorders challenging for older adults. Medication-assisted treatment (MAT) involves the use of U.S. FDA-approved medications in combination with counseling and behavioral therapies to treat opioid use disorders. Although MAT has been shown to be effective in the treatment of opioid use disorders for older adults, its utilization only ranges from 7.9% to 9.8%. This underutilization may stem from a failure to clinically recognize opioid use disorders in older adults, given the differences in life-factors compared to a younger population. The Canadian Coalition of Seniors' Mental Health's guidelines for the treatment of opioid use disorder in older adults recommends the use of buprenorphine as a first-line medication for both withdrawal management and maintenance therapy, with second-line agents including methadone, slow-release oral morphine, and naltrexone. Studies have examined the use of methadone in older adults for the treatment of chronic opioid misuse and physiologic dependence and found that they were no different from their younger counterparts in terms of medical and psychiatric problems or employment but did significantly better in treatment. Despite this more positive prognosis, older adults are thought to be more susceptible to the stigma of drug addiction, old age, psychotropic medication use, depression, poverty, race, and HIV status associated with methadone treatment.

The American Geriatric Society recommends that acetaminophen be used as a first-line agent in the treatment of pain in the older adult population. An appropriate second line for pain management includes SNRIs, gabapentin, and pregabalin alone or in combination with acetaminophen. NSAIDs should be used sparingly as older adults are considered more susceptible to the potential gastrointestinal adverse effects of these agents. Tricyclics have been shown to be somewhat efficacious in the treatment of pain in the general population; however, their use in the older population is not recommended, given their significant adverse effect profile.

ILLICIT DRUGS

Epidemiologic data for illicit drug use, such as amphetamines, heroin, hallucinogens, and cocaine, in older adults are limited. Data from the early 1990s in the Epidemiologic Catchment Area Study estimated lifetime prevalence rates for substance misuse to be 0.12% for older men and 0.06% for older women and a lifetime history of illicit drug use to be 2.88% and 0.66%, respectively. More recent work, however, suggests that illicit substance abuse is increasingly common in adults seeking treatment. Most recent studies indicate a gradual increase in the prevalence of illicit drugs in the older population. Studies looking at perspectives of illicit drugs show an overall decrease in negative opinions about commonly misused substances like marijuana and stimulants. It is important to note that the overall rise in prevalence in this age group is relatively unique when compared to other age groups. This trend likely reflects the aging of the baby boom cohort and is predictive of further increases in illicit substance use and abuse in older adults in the next decades.

The data on illicit stimulants (either amphetamine or cocaine use) in older adults is not very consistent. Some report an overall rise in misuse in the last decade, with prevalence rates between 2% and 3% of all substances. Data indicate higher rate increases of cocaine and methamphetamine use in underrepresented communities in recent years. What is unclear is if these higher reports indicate actual new incidences of use, results from higher screening from clinicians, or natural aging of the younger population with past, ongoing use. Additionally, it is not clear if the prescription stimulants are being cited as misuse by those surveyed. For example, over the past decade, there has been a significant increase in attention-deficit/hyperactivity disorder diagnoses in middle to older age adults. This has led to an increase in prescription stimulant use. The risk of stimulants in older adults is not trivial. Cardiovascular accidents in the form of myocardial infarctions and strokes are commonly reported by clinicians. Hallucinogen usage rate and prevalence in older adults are even more difficult to review. Either because of a decrease in access to these substances as one ages or a decrease in interest in these compounds in older age, most studies indicate less than 1% of misused substances being hallucinogens. However, there does seem to be increasing interest in treatments with ketamine and LSD usage in clinical settings with the older population. It is unclear how this may translate to misuse in the years to come.

In the aging population, while the data are largely unclear on specific illicit drug use, what is certain is that the misuse rate has been increasing. How these substances will interact and impact the overall health of older adults in the context of age-related biopsychosocial changes is of critical importance for the clinician to understand. It is unclear how stimulants and hallucinogens will impact age-associated cognitive decline or cardiovascular changes. It is unclear how different new onset of use will compare to long-term use. The important factor is that there is enough evidence to encourage clinicians to screen for illicit drug use early and continue to monitor even into older age. Illicit drug use can often be determined by a simple urine test. The majority of drugs are detectable in a urine drug screen for 4 days or greater, with some detectable after several weeks. Other physical signs of intoxication for different substances can be determined by physical examination. Intoxication with sedative, hypnotic, or anxiolytic substances or opioid medications results in decreased blood pressure, respiratory rate, pulse, and body temperature, while amphetamine or cocaine intoxication generally induces the reverse.

NONPHARMACOLOGIC THERAPIES

Older adults with substance misuse are underrepresented in research, policy initiatives, and clinical settings. Many research studies exclude patients over the age of 65, resulting in a significant gap of knowledge regarding the neurobiology, treatment response, and prognosis of addiction in later life. On the clinical side, older adults are less likely to be screened for substance use disorders, and clinicians may misattribute declines in functioning and quality of life to age or comorbid medical conditions. Compounding this, older adults are the least likely to perceive treatment needs and use treatment services. Even once identified for treatment, many face ageism when seeking services; for instance, substance use treatment centers often discriminate against older adults, not allowing them to enter into treatment programs. Only one in five treatment centers offers services specifically designed for older adults.

Despite the paucity of research and limited clinical services, evidence suggests that older adults are capable of change and benefit from treatment. This is particularly true when interventions are adapted to age-related physical, cognitive, and psychosocial needs. Studies have suggested that older adults are more likely to complete treatment, have improved outcomes relative to middle-aged adults, and can benefit from varied intervention approaches. In addition to pharmacologic therapies discussed in each subsection above, the following section includes information on brief interventions, psychosocial interventions, 12-step recovery programs, moderators and correlates of treatment response, and partial-hospitalization (PHPs) and intensive outpatient programs (IOPs).

Brief Interventions

Low-intensity, brief interventions or brief therapies are cost-effective, practical techniques that may be useful as an initial or adjunctive treatment approach with at-risk and problem users across a range of clinical settings. Brief interventions are time-limited and nonconfrontational and range from unstructured counseling and feedback to more formal therapy. As the interventions were developed using concepts and techniques from the behavioral self-control literature, one trademark of brief interventions is to encourage individuals to change their behavior through motivational interviewing. As older adults with substance misuse often experience shame about their substance problem, nonconfrontational and supportive approaches are well suited, and labels (e.g., addict, alcoholic) are best avoided.

Six elements under the acronym FRAMES have been identified by the SAMHSA Consensus Panel as critical to brief interventions:

- Feedback about personal risk
- Responsibility for change
- Advice from the clinician
- Menu of treatment options
- Empathic style
- Self-efficacy or optimistic empowerment

Brief interventions or therapies may be provided individually, with family, or in group settings.

Randomized clinical trials of brief interventions for alcohol problems among older populations have demonstrated the efficacy of the treatment, that is, reduced alcohol consumption in at-risk drinkers compared to controls, while also demonstrating that older adults can be engaged in brief intervention protocols and finding the protocols reasonable. Older patients who receive intervention in integrated care conditions have reduced drinking compared to those who are referred out. Most trials to date have been conducted in primary care settings using only one to four sessions that last from 15 to 60 minutes, including some telephone contacts and occasionally follow-up over 1 year.

Brief interventions for older adults are effective in mental health settings as well; thus, geriatric mental health care providers are encouraged to gain familiarity with brief intervention therapy both as a treatment tool and as a way to increase adherence to specialty treatment. Components of the intervention may utilize interactions outside of the clinic (e.g., mail or computer-based) which can help avoid many of the limitations of interventions targeting unhealthy drinking or substance use in clinical settings. For example, such interventions may: provide persons who do not regularly participate in the health care system access to important health information; offer the flexibility of use; allow the patient to maintain anonymity, thereby reducing potential stigma; provide a unique opportunity to individually tailor the intervention, and be far less costly than in-person interventions. Brief screenings are recommended by the U.S. Preventative Task Force as they have been shown to be practical in identifying people with unhealthy substance use and, coupled with brief interventions, can improve outcomes.

Psychosocial Interventions

Beyond screening and brief interventions, some older adults, such as those with moderate to severe use, may need more formal specialized substance use disorder treatment such as cognitive behavioral therapy, family-based therapies, and group and mutual-help support interventions. Few substance use disorder or co-occurring disorder treatments have been designed or studied specifically for older adults. Therefore less is known about best practices or recommendations. One study of older veterans found that individuals who completed a 16-week relapse-prevention group intervention, including cognitive-behavioral and self-management approaches, were more likely to be abstinent at a 6-month follow-up than noncompleters. Another study that included older adults (7% were age 60 or over) demonstrated the effectiveness of three different manual-driven individual psychological treatments (cognitive behavioral therapy, motivational enhancement therapy, and 12-step facilitation) that spanned 12 weeks. In this study, alcohol consumption was reduced 1 year post-treatment in adults with alcohol dependence.

Recommendations for treatment modification for older adults include attending to age and generational factors, psychosocial stressors, and cognitive and physical limitations. Generational factors may affect beliefs and attitudes toward substance use ranging from moralistic failure to cultural permissiveness. Developmental factors facing older adults often include the death of friends and family members, role changes in occupational and home settings, and normal age-related declines in physical and cognitive abilities. Flexible treatment modalities that offer transportation, in-home, or telehealth services are often needed for older adults who need accommodations. Similar to substance use treatments with younger adults, involving family members/caregivers, strengthening or expanding nonsubstance-using social networks, and addressing comorbid psychiatric conditions is often essential to recovery.

Twelve-Step Recovery Programs

Community-based programs for many years have centered on the traditional 12-step peer support model of recovery and rehabilitation, which includes mutual support and encouragement to help group members achieve and maintain sobriety. Studies have shown that older adults engage in and attend 12-step programs (e.g., Alcoholics Anonymous including Seniors in Sobriety, Narcotics Anonymous, Self-Management and Recovery Training [SMART], Women for Sobriety) with comparable frequency to other age groups. On the other hand, the degree of involvement in the programs has been shown to be less in older adults, and the likelihood of engagement in subsequent formal aftercare is reduced. The 12-step model has clearly been successful for many individuals; however, more treatment options are available now than there were when 12-step programs were developed. In addition, certain factors need to be considered, including perceived stigma, level of comfort disclosing personal information in a group setting, the degree to which age-appropriate issues are discussed, and transportation and health problems that may interfere with the likelihood of attendance.

Moderators and Correlates of Treatment

Certain factors may impact treatment response and adherence in older adults receiving treatment. Age-appropriate treatment is considered important in improving treatment completion and attendance rates at group meetings as compared to treatments that do not consider the unique aspects of aging. Characteristics of age-sensitive treatment include: being supportive and nonconfrontational, flexible, sensitive to gender differences, sensitive to cultural differences, holistic, and focused on coping and social skills. Older adults in a study of male veterans with alcohol problems randomly assigned after detoxification to age-specific or standard mixed-age treatment were 2.9 times more likely at 6 months and 2.1 times more likely at 1 year to report abstinence when in the age-appropriate treatment group. Programs that offer treatments specific to age (e.g., Smokefree60+, AA for the Older Alcoholic) should be considered first for older patients.

Matching the treatment to the patient also affects treatment outcome, although this has not been well evaluated in older adults. In one study of older primary care patients who were at-risk drinkers, 60.4% of patients referred to an integrated care model using a brief intervention modality attended at least one session versus 33% of patients referred to specialty care. This result occurred despite the elimination of copayments, availability of insurance claims at most sites, and appointment availability within 2 weeks of the initial assessment.

Certain patient-level characteristics may also impact treatment outcomes. At a 6-month follow-up assessment of patients treated at a private outpatient chemical dependency program, older women with severe AUD were significantly more likely to report abstinence from drugs and alcohol for the previous 30 days than men (79.3% vs. 54.0%). Among patients who were not abstinent, men reported a mean of four heavy drinking days, whereas women denied any heavy drinking. Favorable outcomes for older adults have also been associated with longer treatment retention and the absence of social network members who encourage substance use.

Intensive Outpatient Programs and Partial-Hospitalization

Each year more older adults are admitted to the hospital for substance use disorder treatment. However, treatment services outside of the hospital have failed to keep up. Few substance use treatment programs (approximately 18% of 14,000) provide services designed specifically for older adults, and most that do are hospitals or affiliated with inpatient psychiatric services. IOPs and PHPs are an important component of the continuum of care for the treatment of substance use disorders, but few offer services directed at older adults, and several have histories of excluding this age group. There are no research studies examining the utilization of a stepped approach for older adults with substance misuse.

IOP and PHP programs vary widely but generally include at least 9 hours of treatment per week and cover comprehensive assessment, medical and psychiatric treatment, individual/family/group therapies, 24-hour crisis services, and case management and discharge planning. When treating older patients, IOP and PHP programs need to be actively involved in local aging services, including residential care facilities and adult day health centers.

FUTURE DIRECTIONS

In the past 5 years, significant work has been done to address substance use, its consequences, and its long-term effects in the older population. Some of this work has been necessary as those in middle age transition to older age groups, and clinicians become more astute in identifying the impact of substance use on the overall physical and mental health of older adults. Even so, there continues to be a need for improved research to decrease the knowledge gap to understand the etiology and neurobiology of addiction for this population. Future studies, including individuals over the age of 65, are necessary to improve the understanding of treatment outcomes in older adults, including the impact of nutrition, vitamin supplementation, and comorbid medical illnesses. Further research is needed to validate DSM-5 diagnostic criteria for substance use disorders in older adults and to define the point at which patients with substance misuse should be treated in specialty addiction care as opposed to primary care or general geriatric mental health clinics. With the aging cohort and substance misuse pattern, more attention needs to be paid to prescription medication overuse and other illicit substance use and their impact on the older population. The expansion of the knowledge base on the potential risks and benefits of marijuana use and misuse in the geriatric population, given the current trend of legalization and expanding use, will be important. There may be a need for age-specific substance screening tools that can address the limitations of our current screening process in primary care settings. These tools should pay special attention to the interplay between substance use and normal and pathologic aging processes. As the population continues to age, the impact of higher incidences of diagnoses like ADHD needs to be monitored for potential misuse of prescription stimulants. Additionally, as novel treatments like ketamine and hallucinogenic agents like LSD become more common, how they may be misused by older adults must be reviewed over time.

FURTHER READINGS

Alsherbiny MA, Li CG. Medicinal cannabis-potential drug interactions. *Medicines (Basel)*. 2018;6(1):3.

Amini R, Sahli M, Ganai S. Cigarette smoking and cognitive function among older adults living in the community. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*. 2021;28(4):616–631.

Arora K, Bhagianadh D. Smoking and alcohol consumption following a new dementia diagnosis. *J Gerontol B Psychol Sci Soc Sci*. 2021;76(4):745–755.

Bahji A, Stephenson C, Tyo R, Hawken ER, Seitz DP. Prevalence of cannabis withdrawal symptoms among people with regular or dependent use of cannabinoids: a systematic review and meta-analysis. *JAMA Netw Open*. 2020;3(4):e202370.

Ballantyne JC, Sullivan MD, Koob GF. Refractory dependence on opioid analgesics. *Pain*. 2019;160(12):2655–2660.

Bares CB, Kennedy A. Alcohol use among older adults and health care utilization. *Aging Ment Health*. 2021;25(11):2109–2115.

Caputo F, Vignoli T, Leggio L, Addolorato G, Zoli G, Bernardi M. Alcohol use disorders in the elderly: a brief overview from epidemiology to treatment options.

Exp Gerontol. 2012;47(6):411–416.

Fenollal-Maldonado G, Brown D, Hoffman H, Kahlon C, Grossberg G. Alcohol use disorder in older adults. *Clin Geriatr Med*. 2022;38(1):1–22.

Gerlach LB, Wiechers IR, Maust DT. Prescription benzodiazepine use among older adults: a critical review. *Harv Rev Psychiatry*. 2018;26(5):264–273.

Gruber SA, Sagar KA, Dahlgren MK, et al. The grass might be greener: medical marijuana patients exhibit altered brain activity and improved executive function after 3 months of treatment. *Front Pharmacol*. 2018;8:983.

Henley SJ, Asman K, Momin B, et al. Smoking cessation behaviors among older U.S. adults. *Prev Med Rep*. 2019;16:100978.

Hughes A, Lipari RN, Williams MR. Marijuana use and perceived risk of harm from marijuana use varies within and across states. The CBHSQ Report. July 26, 2016. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, Rockville, MD.

Jakabek D, Yücel M, Lorenzetti V, Solowij N. An MRI study of white matter tract integrity in regular cannabis users: effects of cannabis use and age. *Psychopharmacology (Berl)*. 2016;233(19–20):3627–3637.

Joshi P, Shah NK, Kirane HD. Medication-assisted treatment for opioid use disorder in older adults: an emerging role for the geriatric psychiatrist. *Am J Geriatr Psychiatry*. 2019;27(4):455–457.

Levin KH, Copersino ML, Heishman SJ, et al. Cannabis withdrawal symptoms in non-treatment-seeking adult cannabis smokers. *Drug Alcohol Depend*. 2010;111(1–2):120–127.

Lloyd SL, Striley CW. Marijuana use among adults 50 years or older in the 21st century. *Gerontol Geriatr Med*. 2018;4:2333721418781668.

Manini AF, Yiannoulos G, Bergamaschi MM, et al. Safety and pharmacokinetics of oral cannabidiol when administered concomitantly with intravenous fentanyl in humans. *J Addict Med*. 2015;9:204–210.

Maree RD, Marcum ZA, Saghafi E, Weiner DK, Karp JF. A systematic review of opioid and benzodiazepine misuse in older adults. *Am J Geriatr Psychiatry*. 2016;24(11):949–963.

Markota M, Rummans TA, Bostwick JM, Lapid MI. Benzodiazepine use in older adults: dangers, management, and alternative therapies. *Mayo Clin Proc*. 2016;91(11):1632–1639.

Maust DT, Lin LA, Blow FC. Benzodiazepine use and misuse among adults in the United States. *Psychiatr Serv*. 2019;70(2):97–106.

Mugunthan K, McGuire T, Glasziou P. Minimal interventions to decrease long-term use of benzodiazepines in primary care: a systematic review and meta-analysis. *Br J Gen Pract*. 2011;61(590):e573–e578.

Paquin AM, Zimmerman K, Rudolph JL. Risk versus risk: a review of benzodiazepine reduction in older adults. *Expert Opin Drug Saf*. 2014;13(7):919–934.

Rieb LM, Samaan Z, Furlan AD, et al. Canadian guidelines on opioid use disorder among older adults. *Can Geriatr J*. 2020;23(1):123–134.

Schuster JP, Hoertel N, von Gunten A, Seigneurie AS, Limosin F; CSA Study group. Benzodiazepine use among older adults with schizophrenia spectrum disorder: prevalence and associated factors in a multicenter study. *Int Psychogeriatr*. 2020;32(4):441–451.

Scott EP, Brennan E, Benitez A. A systematic review of the neurocognitive effects of cannabis use in older adults. *Curr Addict Rep*. 2019;6(4):443–455.

Subbaraman MS, Kerr WC. Subgroup trends in alcohol and cannabis co-use and related harms during the rollout of recreational cannabis legalization in Washington state. *Int J Drug Policy*. 2020;75:S0955–3959(19)30181-1.

Sznitman SR, Vulfsons S, Meiri D, Weinstein G. Medical cannabis and insomnia in older adults with chronic pain: a cross-sectional study. *BMJ Support Palliat Care*. 2020;10(4):415–420.

Topiwala A, Ebmeier KP. Effects of drinking on late-life brain and cognition. *Evid Based Ment Health*. 2018;21(1):12–15.

55.4 Treatment of Psychiatric Disorders

55.4a Geriatric Psychopharmacology: General Principles

Bruce G. Pollock, M.D., Ph.D., F.R.C.P.C.

Pharmacokinetic and pharmacodynamic changes associated with aging and illness render older patients more sensitive than younger adults to the adverse effects of medications. Importantly, psychotropics are among the most common medications associated with preventable adverse drug events among the aged. Those 65 years and older, who represent 16.5% of the U.S. population, consume more than 40% of all prescription medications. Approximately one-third of people over age 60 are taking ≥5 prescription drugs. Additionally, they take on average four over-the-counter drugs or herbal preparations. All of this polypharmacy puts them at particular risk for drug–drug interactions.

Preventable drug interactions, inappropriate medications, excessive dosages, and unrecognized adverse drug reactions all contribute substantially to morbidity and mortality. Prescribing cascades occur when side effects from one medicine are treated with another medication. It is not uncommon for a prescriber not to be aware of all the medications a patient is taking. For these reasons, it has been observed that the most common treatable illness in older patients is iatrogenic and carefully stopping or reducing dosages (deprescribing) is often indicated.

Despite, or perhaps because of, their extensive medication use, the old are often excluded from clinical and regulatory trials. Without randomized, prospective drug trials, it is not truly possible for clinicians to decide whether a treatment is efficacious or even harmful. The absence of controlled studies makes it difficult to determine the optimal drug class or drug within a class for treating older adult psychiatric patients. For example, there were limited and somewhat contradictory anecdotal reports on the efficacy and safety of IV ketamine in older adults. However, 194 subjects over age 65 years were included in a double-blind, randomized trial of esketamine (Spravato), and no significant improvement was found in this age group compared to placebo, a fact that was included in the label. Similarly, when manufacturers of atypical antipsychotics were seeking an indication for psychiatric symptoms accompanying dementia, sufficient pooled placebo-controlled data were available to enable the FDA to determine that these medications increased mortality. Encouragingly, there has been continued progress in pharmacotherapy trials for late-life depression and generalized anxiety disorder as well as for the cognitive and psychiatric symptoms accompanying Alzheimer disease.

Although older adult patients frequently require lower doses of medications to mitigate side effects, age, alone, does not entirely guide appropriate dosage. Often, there is greater physiologic variation among individual older adults than there are group differences between young and old. It is important to consider individual patient characteristics, such as cardiovascular or renal function and genetic and acquired differences in body composition or drug metabolism, together with a drug’s physicochemical properties.

In addition to using rational pharmacotherapy, the physician must form an alliance with patients and their families to improve adherence to treatment. Nonadherence with therapy is a major problem for psychiatric patients, and this dilemma is exacerbated with age. The cost and complexity of medication regimens, inherent to the multiple illnesses of the aged, are further complicated by communication difficulties arising from impaired hearing, cognitive impairment, and language or cultural differences.

PHARMACOKINETIC CONSIDERATIONS

Pharmacokinetics provides a way of describing and predicting drug concentrations in plasma and various tissues over time. It has been estimated that 70% to 80% of adverse drug reactions are dose, and therefore concentration, related. Concentration relationships for adverse effects were clearly evident with tricyclic antidepressants (TCAs) and neuroleptics. Cholinesterase inhibitors have been associated with dose-related, cholinergic (nausea, vomiting, diarrhea) side effects, which can be particularly severe if doses are escalated too rapidly, or if cytochrome P450 (CYP) 3A4 metabolism is inhibited for those treated with donepezil (Aricept) or galantamine (Reminyl). For newer antidepressants and antipsychotics, explorations of drug concentration variability, let alone response relationships, were deemed unnecessary. Nonetheless, therapeutic indices for most drugs narrow with age and regulatory trials conducted in younger, healthier subjects are often conducted at or near their maximum-tolerated dosages to increase the probability of demonstrating efficacy. Failure to address this issue resulted in notable difficulties with the introduction of bupropion (Wellbutrin) and clozapine (Clozaril) (seizures at high doses), as well as fluoxetine (Prozac) and risperidone (Risperdal) (agitation, akathisia, extrapyramidal syndromes with excessive doses) in older patients. Olanzapine (Zyprexa) has shown a curvilinear dose–response relationship for treating older adult patients and quetiapine (Seroquel) is currently used over a wide dose range in older adults. In fact, more than 25% of neuropharmacologic drugs introduced in the past two decades have undergone recommended postmarketing dosage reductions for safety reasons.

The selective serotonin reuptake inhibitors (SSRIs) are first-line treatments for depression in older adults. Although they have relatively flat dose–response curves, unnecessarily high SSRI drug concentrations may increase side effects and cost. For example, SSRI plasma concentrations have been linearly associated with adverse cognitive and psychomotor effects in older subjects. Dose-dependent effects of SSRIs have also been reported for fragility fractures, falls, and decrements in bone density. Concern has also been raised about dose-dependent QT interval prolongation with citalopram (Celexa) in those older than age 60 or who are poor CYP 2C19 metabolizers.

It is important to appreciate the critical lack of pharmacokinetic information relevant to the majority of older adults encountered in clinical practice. The limited information that does exist for older subjects is largely derived from “classic” or “two-stage” pharmacokinetic modeling. Although single- and multiple-dose pharmacokinetic studies are a standard part of phase I drug development, they are almost exclusively carried out on young, physically healthy, unmedicated men. For example, the only published geropharmacokinetic data on the disposition of fluoxetine (which has a geriatric indication) is limited to a study of single doses in 11 healthy subjects (ranging in age from 65 to 77 years). Traditional pharmacokinetic methodology requires an intensive sampling regimen (i.e., a large number of precisely timed samples) usually obtained from a small number (e.g., 6 to 12) of volunteers. The pharmacokinetic parameters are determined for each individual separately and average pharmacokinetic parameters are calculated as the mean of the individually determined parameters. Single-dose studies often fail to provide accurate estimates of pharmacokinetic parameters for older adults. Multiple-dose kinetic studies are especially important to detect nonlinear phenomena wherein metabolic clearance ceases to be proportional to dose.

An alternative method is that of population pharmacokinetics using mixed-effects modeling. This statistical approach permits identification of individual and overall population pharmacokinetic parameters based on only a few samples per subject and using each data point to inform the entire analysis. Because a greater number of typical patients are evaluated, the results are much more applicable to clinical practice than those of traditional pharmacokinetics. Incorporating population pharmacokinetic analyses into clinical trials with older subjects may prove to be a significant addition to the ability to determine sources of variability in psychotropic plasma concentrations and response in older adults, particularly those with medical comorbidity.

Age-associated pharmacokinetic differences may be due to changes in absorption, distribution, metabolism, or elimination of a drug.

Absorption

With age, there is a decrease in the rate of gastric emptying and intestinal blood flow. Most studies evaluating drug absorption, however, have shown no difference in the rate or extent of absorption in geriatric versus younger patients. Nevertheless, antacids, high-fiber supplements, laxatives, and the cholesterol-binding resin cholestyramine (Questran) may significantly diminish the absorption of medications, such as TCAs and anticonvulsants. Drugs with anticholinergic properties may decrease absorption of L-dopa (Larodopa), owing to delayed gastric emptying in conjunction with the gastric metabolism of L-dopa. The potential gain in compliance achieved by administering medications at the same time should be weighed against the possibility of drug interactions in the absorption phase.

There is increasing interest in the genetic differences and drug interactions affecting functioning of the important drug efflux transporter, P-glycoprotein (MDR1), which can transport a large number of clinically used drugs out of the intestine, thereby playing a significant role in absorption. For example, variants in the ABCB1 gene, which encodes P-glycoprotein, have been found to influence the bioavailability of digoxin (Lanoxin). P-glycoprotein can be induced by carbamazepine (Tegretol), and inhibited by calcium channel blockers as well as macrolide antibiotics.

Distribution

Clinicians frequently overlook the fact that a drug’s half-life is directly proportional to its apparent volume of distribution. Adipose tissue increases from 33% of body mass in younger women to 48% in the old, and from 18% to 36% in men. Therefore for drugs that distribute in body fat (the vast majority of psychotropics), the volume of distribution will increase with age, leading to a longer half-life and drug accumulation. For water-soluble drugs, such as lithium or digoxin, volumes of distribution will be diminished in older patients, reducing the margin of safety (i.e., transient increases of these drugs' plasma concentrations will not be buffered to the same degree in older adults). Lipid-soluble benzodiazepines (e.g., diazepam [Valium] and clonazepam [Klonopin]) with long half-lives have been associated with an increased risk of falls and hip fractures in contrast to more water-soluble benzodiazepines with intermediate half-lives, such as lorazepam (Ativan).

The volume of distribution is also affected by the extent of drug bound to plasma proteins. Age-associated illness may cause reductions in serum albumin and possible increases in α-1-acid glycoprotein. The majority of psychotropics are transported in the blood extensively bound to these plasma proteins. Drug displacement interactions from plasma proteins were previously the subject of considerable concern because the protein-bound drug acts as an enormous, pharmacologically active reservoir. It is now recognized that not only is free drug pharmacologically active, but it is also more available for metabolism and tissue distribution. Therefore, an increase in the absolute concentration of free drug, caused by reduction in plasma proteins or a displacement interaction, would be buffered and equilibrium rapidly reestablished. When plasma drug monitoring is used to adjust TCA or anticonvulsant doses, it should be appreciated that the total drug concentrations (free + protein bound) are usually reported. In the face of hypoalbuminemia or a displacement interaction (e.g., phenytoin and aspirin), the total drug level may appear low and doses may, in consequence, be inappropriately increased. If the displacing drug also causes inhibition of drug metabolism (e.g., valproic acid [Depakote] and aspirin), total drug levels may appear to be acceptable, while free-drug concentrations are in fact significantly elevated. The use of unbound (free) drug levels in these situations has been found to be essential for lidocaine (Xylocaine), theophylline (Theo-Dur), valproic acid, phenytoin (Dilantin), and digitoxin (Crystodigin).

The blood–brain barrier forms an interface between the circulating blood and the brain and possesses various carrier-mediated transport systems for small molecules to support and protect central nervous system (CNS) function. For example, the blood-to-brain influx transport systems supply nutrients, such as glucose and amino acids. Age and illness, such as diabetes, may affect the integrity of the blood–brain barrier, increasing the drug exposure of the brain. P-glycoprotein, located at the luminal membrane of endothelial cells, functions as a clearance system for drugs entering the brain as well as for metabolites and neurotoxic compounds produced in the brain. Drugs recognized by these transporters are expected to show lower blood–brain barrier permeability. Examples of drugs now known to be substrates for P-glycoprotein include digoxin, erythromycin (Ery-Tab), lidocaine, morphine (Astramorph), nifedipine (Adalat), nortriptyline (Pamelor, Aventyl), propranolol (Inderal), risperidone, and verapamil (Calan). Several P-glycoprotein substrates may also inhibit its function, such as erythromycin, risperidone, and verapamil, permitting greater CNS penetration.

Metabolism

The purpose of drug metabolism is to transform lipid-soluble compounds into more water-soluble derivatives that can be more effectively eliminated and may be less active pharmacodynamically. Biotransformation of drugs is categorized as belonging to either phase I (oxidation or reduction) or phase II (conjugation). A large group of heme-containing enzymes, which vary in substrate preferences, collectively called CYP enzymes, is responsible for oxidative (phase I) drug metabolism. As more is learned about both the diversity and specificity of the CYP enzymes, a structure is emerging for categorizing potential drug–drug interactions and possible age-associated changes. Available evidence suggests that there is not a uniform age-associated decline in the quality of activity of these enzymes. Nonetheless, reductions in hepatic mass and blood flow with aging place greater emphasis on interindividual differences in drug metabolic capacity. These metabolic differences may be genetic or the result of interactions from multiple medications. Isoenzyme specificity suggests that inhibition or induction of a given CYP will affect all drugs metabolized by that specific enzyme. A drug may act as a substrate for one CYP at a time or simultaneously for more than one. A drug can act as both a substrate and a competitive inhibitor for the same isozyme, or it can act as a noncompetitive inhibitor for a different enzyme.

CYP 2D6 is the isozyme specifically responsible for hydroxylation of many antidepressants (e.g., nortriptyline, desipramine [Norpramin, Pertofrane], duloxetine [Cymbalta], paroxetine, venlafaxine [Effexor], and vortioxetine [Brintellix]) and several antipsychotics (perphenazine [Trilafon], thioridazine [Mellaril], brexpiprazole [Rexulti], iloperidone [Fanapt], and risperidone). Approximately 7% of the Caucasian population are genetically poor 2D6 metabolizers, which has been shown to impact on nortriptyline doses and adverse effects of perphenazine in older patients. Concern has also been raised about the potential cardiotoxicity of venlafaxine in poor 2D6 metabolizers or those taking concomitant diphenhydramine (a CYP 2D6 inhibitor). Studies conducted in older, unmedicated volunteers have not found an age-associated decline in CYP 2D6 activity. Nevertheless, many drugs commonly used by older patients competitively (e.g., perphenazine) or noncompetitively (e.g., paroxetine, fluoxetine, bupropion) inhibit this enzyme, leading to potentially significant interactions. The CYP 2D6 mediated metabolism of dextromethorphan and its inhibition by quinidine has led to the introduction of a combined product (Nuedexta) for the treatment of pseudobulbar affect.

CYP 2C19 is responsible for the metabolism of diazepam, mephenytoin (Mesantoin), citalopram, escitalopram (Lexapro), and omeprazole (Prilosec), as well as several barbiturates. In addition, the demethylation of tertiary TCAs is also partly dependent on 2C19. The incidence of the poor metabolizer phenotype in a young Caucasian population is approximately 3% but was 18% in a Japanese sample. There is some evidence that 2Cl9 functioning may decline with age. CYP 2C19 is responsible for metabolizing the narrow-therapeutic–index drugs phenytoin, tolbutamide (Orinase), and warfarin (Coumadin) and is inhibited by fluvoxamine (Luvox) and fluoxetine. Clopidogrel (Plavix) is an anticoagulant which requires 2C19 metabolism to be an active agent, It is mainly used to ensure patency after stent procedures. Given the high proportion of Asians in Hawaii who are poor CYP 2 C19 metabolizers, it was found that there were a disproportionate number of stent failures in Hawaii.

CYP 3A4 is the most prevalent drug-metabolizing enzyme and is responsible for metabolizing the most medications. Because it is normally a high-capacity enzyme, serious and surprising toxicity has occurred when the 3A4-mediated clearance of terfenadine (Seldane), astemizole (Hismanal), cisapride (Propulsid), midazolam (Versed), triazolam (Halcion), mibefradil (Posicor), and cerivastatin (Baycol) was diminished. Also, because hepatic metabolism of CYP 3A4 drugs is typically so rapid, many of these drugs are perfusion limited (i.e., dependent on hepatic blood flow, which may decline by 40% between ages 25 and 65 years). Therefore, it often appears that drugs specifically metabolized by CYP 3A4 decline with age—for example, alprazolam (Xanax) and triazolam clearances in older adults are only 50% to 80% of those in young adults. In particular, gender- and age-related pharmacokinetic differences have been demonstrated for several 3A4-typed drugs, such as sertraline (Zoloft), mirtazapine (Remeron), and nefazodone (Serzone). CYP 3A4 activity may be inhibited by grapefruit juice, protease inhibitors, “mycin” antibiotics, and “azole” antifungals. For example, pimavanserin (Nuplazid) concentrations are substantially increased by ketoconazole (Nizoral).

Among antidepressants, nefazodone and fluvoxamine are the most potent inhibitors, followed by fluoxetine, through its demethylated metabolite. The very long half-life of norfluoxetine may result in interactions occurring many weeks after initiation of fluoxetine. Similarly, the inhibition of the CYP 3A4-mediated clearance of “statins,” resulting in muscle weakness, may be misattributed in depressed patients. The 3A4 enzyme also may be induced by other drugs, such as carbamazepine, phenytoin, topiramate (Topamax), modafinil (Provigil), barbiturates, steroids, and St. John’s wort, increasing the risk of therapeutic failure for CYP 3A4 drugs. Many CYP 3A4 inhibitors (e.g., verapamil) and inducers (St. John’s wort) also have been found to interact at the P-glycoprotein drug transporter, amplifying their CYP 3A4-mediated effects.

CYP 1A2 metabolizes clozapine, olanzapine, fluvoxamine, asenapine (Saphris), and theophylline and contributes to the demethylation of some tertiary TCAs. This isozyme is induced by cigarette smoking and eating cruciferous vegetables and charcoaled meats, as well as by certain medications, such as omeprazole and phenobarbital (Luminal). Estrogen replacement therapy in postmenopausal women has been found to inhibit CYP 1A2 metabolism, whereas treatment with growth hormone causes induction. Inhibition and induction magnify an already marked, 60-fold variation in 1A2 activity. This degree of metabolic variation mandates that theophylline and clozapine be closely monitored.

There is more limited information about CYP 2B6, which substantially metabolizes bupropion. In vitro evidence has suggested that fluoxetine, paroxetine, and sertraline may cause CYP 2B6 inhibition. Given bupropion’s risk of concentration-related seizures and the frequency of its use with SSRIs, this may be of potential concern.

Beyond oxidative metabolism, phase II metabolism in older patients has been a neglected area for systematic study. In general, drugs such as lorazepam and oxazepam (Serax) accumulate less than more lipid-soluble drugs and undergo single-step conjugative elimination. Glucuronidation, however, may not be an immutable process, being significantly affected by malnutrition. Despite the common assertion that phase II metabolism does not change with age, significant reductions in oxazepam’s elimination have been found in those older than 80 years.

Excretion

The average glomerular filtration rate decreases by 50% between the ages of 20 and 70 years. Age alone, however, cannot be used to predict renal function, because at least one-third of older patients may maintain relatively stable renal function. Serum creatinine may not appear elevated in the face of declining renal function because there is often a decline in muscle mass, the major source of creatinine, with age and debilitation. Minimally, the use of estimates from standard nomograms or equations should be used to estimate creatinine clearance, given that it is often impractical to collect a 24-hour urine sample. Decline in renal clearance may directly affect excretion of such medications as lithium, gabapentin (Neurontin), pregabalin (Lyrica), topiramate, amantadine (Symmetrel), paliperidone (Invega), memantine (Namenda), rivastigmine (Exelon), and desvenlafaxine (Pristiq). The clearance of hydrophilic (polar), psychotropic drug metabolites may also be reduced in older patients. Paliperidone is the active 9-OH metabolite of risperidone and accumulations of active hydroxylated metabolites of antidepressants, such as nortriptyline, bupropion, and venlafaxine, have been observed in older patients with uncertain clinical consequences.

The nonsteroidal anti-inflammatory drugs (NSAIDs) and angiotensin-converting enzyme (ACE) inhibitors reduce lithium's clearance (possibly through effects on renal blood flow). Given the widespread use of these medications in older adults and their increased risk of lithium toxicity, this potential interaction deserves emphasis. Low-salt diets and the possibility that SSRIs may cause inappropriate antidiuretic hormone secretion, with concomitant hyponatremia, are additional sources of interactive complications with lithium.

PHARMACODYNAMICS

As patients age, there is a general reduction in homeostatic mechanisms (e.g., postural control, orthostatic circulatory responses, thermoregulation, visceral muscle function, laryngeal reflexes, hypoxic responses, cognitive function). This reduction may interfere with adaptation to environmental changes and may manifest as adverse drug experiences. All psychotropics may increase risk of falls and hip fractures and may increase cognitive impairment. Physiologic (postural or action) tremors may be exacerbated by numerous drugs, including β and dopamine agonists, xanthines, valproic acid (valproate), steroids, and antidepressants. Reduced β-receptor sensitivity with aging contributes to diminished baroreceptor reflex responses, whereas coronary artery disease, reductions in the heart's intrinsic sinus rate, and damage to conduction pathways predispose to arrhythmias. The fact that SSRIs may exacerbate bradycardia should also be borne in mind. In older patients chronically treated with antipsychotics, the prevalence of obesity, poor nutritional status, and increased triglycerides may contribute additional cardiovascular risk factors.

Specific receptor and neurotransmitter changes associated with senescence, such as reductions in dopamine or acetylcholine content, result in increased sensitivity to D2 antagonists and antimuscarinic agents. Side effects from conventional antipsychotics and tricyclics were common and frequently troubling in older patients. In particular, the association of these medications with significant cognitive impairment declines in activities of daily living, and significant falls (with concomitant risk of hip fracture) was of grave concern in the frail older adults. Prospective studies demonstrated that there was a very rapid and high incidence of onset of tardive dyskinesia in older patients taking conventional antipsychotics for relatively brief periods (cumulative incidence reported as 26% to 45% over 1 year). Although atypical antipsychotics are extensively used in older patients, meta-analyses of 17 controlled trials found that these medications increase mortality (4.5%) compared to placebo (2.6%) in older adults suffering from dementia within the relatively short duration of the trials conducted (typically 10 weeks). Moreover, the side effects of these medications may outweigh their potential benefits. Some of these side effects are concentration related, such as extrapyramidal symptoms with risperidone and anticholinergic symptoms with clozapine or olanzapine. Concentrations of atypical antipsychotics may be significantly affected by concomitantly administered medications that inhibit or induce CYPs IA2 (asenapine, clozapine, olanzapine), 3A4 (aripiprazole, iloperidone, quetiapine, ziprasidone), or 2D6 (aripiprazole, risperidone).

Two groups of drugs cause anticholinergic effects: those used therapeutically for their anticholinergic properties (e.g., oxybutynin [Ditropan], benztropine [Cogentin]) and those that have anticholinergic effects that are incidental to their main actions (e.g., amitriptyline [Elavil], cimetidine [Tagamet], digoxin, paroxetine). Peripheral manifestations of anticholinergic action include dry mouth, tachycardia, blurring of vision, urinary retention, and constipation. Common difficulties in older patients, such as constipation, xerostomia, glaucoma, and urinary retention, may be markedly exacerbated by even modestly anticholinergic medications. Anticholinergic effects on the CNS may range from subtle changes, such as mild confusion and impairment of recent memory, to physical decline. Patients with a dementing illness are most at risk from these central anticholinergic effects. Nonetheless, studies have found that even medications with modest anticholinergic effects are additive and affect cognitive performance in a wide cross-section of older adults. In fact, understanding of the nature of cognitive decrements associated with "normal" aging may be distorted by modest anticholinergic burden. It should also be noted that medications with prominent noradrenergic actions can also cause dry mouth, urinary retention in older men, and tachycardia, in addition to exacerbating hypertension and tremor.

Older patients have been found to require smaller doses of anticoagulants to achieve the same degree of anticoagulation as obtained in younger individuals. Although confounded by effects of other medications and hepatic dysfunction, increased sensitivity of older adults to anticoagulants is probable. Anticoagulants figure prominently in drug-related morbidity and mortality. Genetic tests may, in the near future, help to detect the estimated one-third of patients with variants that make them slow metabolizers of warfarin (CYP 2C9) or more sensitive to its effects (vitamin K epoxide reductase activity [VKORC1]). Anticoagulant–antidepressant interactions may be pharmacokinetic as well as pharmacodynamic. Fluvoxamine poses a risk for a pharmacokinetic interaction through inhibition of CYP 2C9, reducing the clearance of warfarin's active S-enantiomer. In addition, inhibition of CYPs 1A2 (fluvoxamine), 2C19 (fluvoxamine, fluoxetine), and 3A4 (fluvoxamine, fluoxetine) will cause R-warfarin to accumulate, which also contributes to reduced S-warfarin clearance. However, increased bleeding times with SSRIs alone or in combination with anticoagulants also may be possible owing to SSRIs' effects on platelet serotonin and thereby attenuating aggregation. Concurrent therapy need not be avoided, but close monitoring is advisable whenever an SSRI is combined with warfarin. Both concerns (increased risk of gastrointestinal bleeding in older adults) and potential benefits (protective effects against myocardial infarction in smokers) have been raised by epidemiologic studies of SSRIs.

In addition to problems that may be anticipated, because most of the data on a new drug's effects have been obtained in younger patients, unexpected side effects are not infrequent in older adults. In older patients, SSRIs and venlafaxine have been associated with the syndrome of inappropriate antidiuretic hormone secretion (SIADH). Although SIADH is a rare adverse effect of many psychotropic medications, the majority of reports of antidepressant-induced SIADH have involved patients older than 65 years. In an analysis of 760 reports of hyponatremia associated with SSRIs, the median time to onset of the hyponatremia was 13 days (range, 3 to 120 days). SSRIs in older patients also have been associated with Parkinsonism or worsening of motor disability in patients who have idiopathic Parkinson disease. It is possible that this is a result of serotonergic-influenced reduction in dopaminergic transmission in a population with diminished dopamine reserves. Concern has also been raised recently from large population studies. One found chronic use of SSRIs associated with a twofold increased risk of fragility fractures and lower bone mineral density. Another associated SSRI therapy with a nearly fivefold increased risk of completed suicide in older patients during the first month of treatment compared with other antidepressants. This finding is somewhat paradoxical since the combined results of regulatory studies with SSRIs suggested an attenuation of suicidal ideation in older subjects early in treatment. Although the absolute risk of suicide may be low, there is a clear need for close monitoring of older patients, especially males with comorbid medical illness.

In addition to the CYP isozymes discussed previously which may impact on pharmacogenetics, There are genetic polymorphisms that may impact response at drug targets. The best studied is SLC64A which incodes for the serotonin transporter promoter which in turn determines the availability of serotonin transporters which are the sites of action of SSRIs. Those possessing the long form (L/L) of SLC64A may have a more rapid and robust response to an SSRI as compared to those with the S/S genotype.

Finally, a genetic polymorphism in the human leucocyte HLA antigen B*1502 has been linked to severe dermatologic reactions such as Stevens–Johnsons syndrome in those prescribed carbamazepine, oxcarbazepine (Trileptal), lamotrigine (Lamictal), and phenytoin.

Among the greatest threats to successful pharmacotherapy in older patients are patient nonadherence with medication regimens and drug–drug interactions. High-risk medications, including anticoagulants, antihypertensives, and antiarrhythmics, have been consistently emphasized in long-term, comprehensive surveys of adverse drug reactions in older patients. Physicians need to be particularly vigilant and knowledgeable when psychotropics are coprescribed with these high-risk medications. As patients age, pharmacodynamic changes and concomitant illness will significantly alter their response to medication. Psychiatrists treating older patients should use a set of routine procedures as follows: (1) identify patient risk factors, such as medical status, use of high-risk medications, and alcohol use; (2) take a thorough drug history involving caregivers and pharmacists, being sure to include over-the-counter medications and update regularly in a prominent place in the patient's record; (3) educate patients and caregivers about their medications and dosing regimens and monitor adherence; (4) be knowledgeable about drug actions, interactions, and clearance; (5) use the lowest effective doses and carefully graduated increases, and avoid unnecessary polypharmacy, but do not undertreat; (6) frequently reassess psychomotor function, including mental status and balance, as well as orthostatic blood pressures and specific laboratory tests when appropriate; and (7) if there is a deterioration in patients' physical or cognitive states, always suspect an adverse drug reaction.

FUTURE DIRECTIONS

Developments in clinical trial design and the completion of large effectiveness trials serve further to heighten concerns about the efficacy and side effects of currently available pharmacologic therapy for the symptomatic relief of psychiatric symptoms in older adults.

Controlling for heterogeneous drug exposure in conjunction with improved understanding of the genomic basis of cognitive, mood, and psychotic dysfunction and its complex interactions with age-related ischemic and other environmental CNS changes, may lead to more targeted molecular therapeutics. Regulatory policy and public health initiatives must be amplified to increase the representation of typical older subjects in clinical trials and improve the relevance of the therapeutic evidence base.

Recently, there have been a number of significant rigorous trials conducted by pharmaceutical companies for disease-modifying agents for Alzheimer disease. Recently, there has been promise shown for the first time in delaying the progression of Alzheimer's by three anti-amyloid drugs, aducanumab (Aduhelm), lecanemab (Leqembi), and donanemab. Federally funded projects focused on treating the agitation accompanying dementia are proceeding. Also, small pharmaceutical companies have taken an interest in this area and there are number of phase 3 trials which are ongoing at the time of writing. It is to be hoped that we will soon have evidence-based alternatives to the extensive use of lethal antipsychotics for this tragic consequence of dementia.

For medications approved for release, it is to be hoped that postmarketing reporting and the cross-linking of electronic medical and pharmacy records will continue leading to improved prescribing for older psychiatric patients.

ACKNOWLEDGMENT

The Peter & Shelagh Godsoe Endowed Chair in Late-Life Mental Health.

FURTHER READINGS

Aftab A, Lam JA, Liu F, Ghosh A, Sajatovic M. Recent developments in geriatric psychopharmacology. *Expert Rev Clin Pharmacol*. 2021;14(3):341–355.

American Geriatrics Society 2019 Beers Criteria Update Expert Panel. American Geriatrics Society updated Beers criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc*. 2019;67:674–694.

Andrade C, Sandarsh S, Chethan KB, Nagesh KS. Serotonin reuptake inhibitor antidepressants and abnormal bleeding: a review for clinicians and a reconsideration of mechanisms. *J Clin Psychiatry*. 2010;71:1565–1575.

Cascorbi I. Drug interactions-principles, examples and clinical consequences. *Dtsch Arztebl Int*. 2012;109:546–555.

Chew ML, Mulsant BH, Pollock BG, et al. Anticholinergic activity of 107 medications commonly used by older adults. *J Am Geriatr Soc*. 2008;56:1333–1341.

Coupland CAC, Hill T, Dening T, Morriss R, Moore M, Hippisley-Cox J. Anticholinergic drug exposure and the risk of dementia: a nested case-control study. *JAMA Intern Med*. 2019;179:1084–1093.

Danielsson B, Collin J, Bergman GJ, Borg N, Salmi P, Fastbom J. Antidepressants and antipsychotics classified with torsades de pointes arrhythmia risk and mortality in older adults—a Swedish nationwide study. *Br J Clin Pharmacol*. 2016;81:773–783.

Flockhart DA. Cytochrome P450 Drug Interaction Table. Indiana University School of Medicine, Division of Clinical Pharmacology. Accessed March 7, 2021. https://medicine.iu.edu/internal-medicine/specialties/clinical-pharmacology/drug-interaction-flockhart-table

Fox C, Smith T, Maidment I, et al. Effect of medications with anticholinergic properties on cognitive function, delirium, physical function and mortality: a systematic review. *Age Ageing*. 2014;43:604–615.

Fralick M, Bartsch E, Ritchie CS, Sacks CA. Estimating the use of potentially inappropriate medications among older adults in the United States. *J Am Geriatr Soc*. 2020;68:2927–2930.

Hefner G, Wolff J, Hahn M, et al. Prevalence and sort of pharmacokinetic drug-drug interactions in hospitalized psychiatric patients. *J Neural Transm (Vienna)*. 2020;127:1185–1198.

Ho T, Pollock BG, Mulsant BH, et al. R- and S-citalopram concentrations have differential effects on neuropsychiatric scores in elders with dementia and agitation. *Br J Clin Pharmacol*. 2016;82:784–792.

Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. *Lancet*. 2014;383:911–922.

Islam F, Maciukiewicz M, Freeman N, et al. Contributions of cholinergic receptor muscarinic 1 and CYP1A2 gene variants on the effects of plasma ratio of clozapine/N-desmethylclozapine on working memory in schizophrenia. *J Psychopharmacol*. 2021;35:31–39.

Jin Y, Pollock BG, Frank E, et al. Effect of age, weight and CYP2C19 genotype on escitalopram exposure. *J Clin Pharmacol*. 2010;50:62–72.

Lau SWJ, Schlender JF, Slattum PW, Heald DL, O'Connor-Semmes R. Geriatrics 2030: developing drugs to care for older persons-a neglected and growing population. *Clin Pharmacol Ther*. 2020;107:53–56.

Meyers BS, Jeste DV. Geriatric psychopharmacology: evolution of a discipline. *J Clin Psychiatry*. 2010;71:1416–1424.

Nebes RD, Pollock BG, Perera S, Halligan EM, Saxton JA. The greater sensitivity of elderly APOE ε4 carriers to anticholinergic medications is independent of cerebrovascular disease risk. *Am J Geriatr Pharmacother*. 2012;10:185–192.

Pollock B, Forsyth C, Bies R. The critical role of clinical pharmacology in geriatric psychopharmacology. *Clin Pharmacol Ther*. 2009;85:89–93.

Pollock BG, Ferrell RE, Mulsant BH, et al. Allelic variation in the serotonin transporter promoter affects onset of paroxetine treatment response in late-life depression. *Neuropsychopharmacology*. 2000;23:587–590.

Pollock BG, Laghrissi-Thode F, Wagner WR. Evaluation of platelet activation in depressed patients with ischemic heart disease after paroxetine or nortriptyline treatment. *J Clin Psychopharmacol*. 2000;20:137–140.

Pollock BG, Mulsant BH. Between Scylla and Charybdis: antipsychotic and other psychotropic medications in older nursing home residents. *CMAJ*. 2011;183:778–779.

Porsteinsson AP, Drye LT, Pollock BG, et al. Effect of citalopram on agitation in Alzheimer disease: the CitAD randomized controlled trial. *JAMA*. 2014;311:682–691.

Qato DM, Alexander GC, Conti RM, Johnson M, Schumm P, Lindau ST. Use of prescription and over-the-counter medications and dietary supplements among older adults in the United States. *JAMA*. 2008;300:2867–2878.

Reuben DB, Herr KA, Pacala JT, et al. *Geriatrics at Your Fingertips*. 22nd ed. The American Geriatrics Society;2020.

Richardson K, Loke YK, Fox C, et al. Adverse effects of Z-drugs for sleep disturbance in people living with dementia: a population-based cohort study. *BMC Med*. 2020;18:351.

Romskaug R, Skovlund E, Straand J, et al. Effect of clinical geriatric assessments and collaborative medication reviews by geriatrician and family physician for improving health-related quality of life in home-dwelling older patients receiving polypharmacy: a cluster randomized clinical trial. *JAMA Intern Med*. 2020;180:181–189.

Sandson NB, Marcucci C, Bourke DL, Smith-Lamacchia R. An interaction between aspirin and valproate: the relevance of plasma protein displacement drug-drug interactions. *Am J Psychiatry*. 2006;163:1891–1896.

Smeets CHW, Smalbrugge M, Koopmans RTCM, et al. Can the proper intervention reduce psychotropic drug prescription in nursing home residents with dementia? Results of a cluster-randomized controlled trial. *Int Psychogeriatr*. 2021;33(6):577–586.

Steinmetz KL, Coley KC, Pollock BG. Assessment of geriatric information on the drug label for commonly prescribed drugs in older people. *J Am Geriatr Soc*. 2005;53:891–894.

55.4b Antidepressants and Mood Stabilizers

CARL SALZMAN, M.D.

GOALS AND RATIONALE FOR THE PHARMACOLOGIC TREATMENT OF DEPRESSION IN LATER LIFE

Clinically significant depression in late life is a disorder that encompasses a spectrum of severity ranging from debilitating major depression that has serious consequences for the life of the patient to minor depression that can impair quality of life. Most depressed older individuals have had prior depression, sometimes many, and sometimes as part of a bipolar spectrum. As in younger people, the course of depression in older adults is characterized by exacerbations and remissions as well as chronicity. Multiple pathways contribute to late-life depression, including neurobiologic and psychological factors, physical illnesses and their treatments, cognitive and sensory decline, as well as common social and economic problems. Depression is common in medical diseases and may be linked to inflammatory cytokines. First-onset depression in an older adult is usually associated with the increasing presence of cortical white matter hyperintensities and cognitive impairment. Depression also commonly accompanies late-onset dementias. Clinicians should also be alert to other medications and alcohol causing or exacerbating depressed mood. Among community-dwelling residents over the age of 65, approximately 13% to 27% may suffer from subsyndromal depression, 15% exhibit depressive symptoms but do not meet diagnostic criteria for any disorder, 1% to 2% suffer from major depression, and 2% are living with dysthymia. The prevalence of depression in all forms is much higher in nursing homes and in hospitals.

There appears to be considerable heterogeneity in the onset and course of late-life depression. Diagnostic patterns are also variable among older individuals. True clinical depression should be distinguished from unhappiness and grief that can be part of normal aging. When symptoms are serious enough to impair normal age-appropriate mood and behavior and persist for more than 2 weeks, the diagnosis of a treatable late-life depression should be considered. Clues to depression include frequent medical visits with persistent complaints of pain, headache, fatigue, and insomnia; and high utilization of services. Additional clues to depression in older adult hospitalized patients include delayed recovery, treatment refusal, and problems with discharge. In the nursing home patient, depression should be suspected in the presence of apathy, failure to thrive, agitation, and delayed rehabilitation. Older individuals who are depressed may also have structural and functional changes in the brain, including ventriculomegaly, white matter hyperintensities, cortical thinning, and vascular changes. Apathy, as part of late-life depression, is particularly resistant to antidepressant treatment and may be a result of changes in cortical thickness and communication between elements of the salience network.

Depression in later life should be treated because it usually persists in the absence of treatment, causing distress to patients' caregivers, as well as leading to excess healthcare use and increased substance abuse. Mood disorders also play a greater role in the suicides of older adult patients than in those of younger patients; the group at the highest risk for suicide is white males aged 80 to 84 years. A majority of suicide victims visit a primary care physician in the last month of life.

Aging is associated with many important changes in life, such as the loss of loved ones, retirement, and changes in living arrangements that may contribute to depression. For some depressed older individuals, psychosocial interventions and various forms of psychotherapy are available for treatments, although they have not been studied as intensively as pharmacotherapy. Music and pets are often helpful.

Pharmacologic treatment of depression in older adults may improve day-to-day functioning and self-care and can improve physical health and cognitive functioning. Although drug treatment benefits usually outweigh the risks, late-life depression is often untreated with medication, especially in frail older adults, either because the depression is not recognized or from fear that treatment will further damage an older individual. Poor medication adherence, concomitant substance use, and comorbid neurologic illnesses decrease response to medication for late-life bipolar disorder. The risks of no treatment or inadequate treatment, however, include chronic depression, increased cognitive impairment, worsening of physical illness, poor compliance with medical treatment, social impairment, reduced quality of life (satisfaction, ability to cope), and suicide. Anxiety is a common component of late-life depression, and high levels of anxiety may delay the response to antidepressant treatment. Although the research evidence for nonpharmacologic treatments is scarce, the few available studies, as well as clinical experience, suggest that for those older adults who cannot take medications or who prefer not to take a drug, nonpharmacologic treatments and interventions should be implemented.

The goals of pharmacologic treatment of both major and minor depressive disorders in later life include remission, if not substantial reduction, of depressive symptoms, as well as the prevention of relapse and recurrence. Hence, long-term use of antidepressants is probably the rule, not the exception. Treatment goals also include restoring psychosocial functioning, relieving excess disability, and improving patient compliance with other medical therapies that might be refused because of hopelessness or pessimism caused by depression. Antidepressants may also ease the adaptation to irreversible losses. Recent meta-analyses and extensive clinical experience suggest the greater efficacy of antidepressants compared with placebo in older adults. Expert consensus guidelines now recommend that antidepressant treatment should be combined with psychotherapy when possible. Since older individuals with mood disorders are often taking multiple medications, it is also necessary for the prescribing clinician to remain alert for possible adverse drug interactions. These will be discussed in each of the following sections.

PHARMACOLOGIC TREATMENT OF DEPRESSION

Antidepressants used to treat older depressed adults include the same drugs that are given to younger adults. The overall efficacy of antidepressants for the treatment of late-life depression is well established, although many, if not most older adult patients do not experience full remission from their depressive symptoms and may be more sensitive to side effects and drug interactions. In general, aging results in altered metabolism and excretion of some antidepressants, prolonging their effect. For example, the clearance of some antidepressants, such as citalopram, is delayed, causing drug accumulation and high blood levels. The sensitivity of the central nervous system to antidepressant drugs is also increased by age: most (but not all) older individuals are more sensitive to the therapeutic effects of antidepressants as well as their side effects. Since older individuals are often taking multiple medications, there is an increased risk of adverse drug interactions with antidepressants.

Accumulating data suggest that older adult patients respond as well to antidepressant treatments as younger adults. Data also suggest that the time to respond does not differ from younger adults, but clinicians often notice that it may require longer time periods to achieve remission or full therapeutic response. Response to treatment of late-life depression is associated with severity, current episode severity, and executive dysfunction. In frail older adults, treatment response to antidepressants may not differ from placebo response, suggesting that the older adult patient’s physical status may also predict response or nonresponse. The presence of anxiety is associated with increased severity of depression but does not affect response to antidepressant treatment. Antidepressants are sometimes given to older individuals with Alzheimer disease (AD), but the data for their usefulness is negative, and there may be anticholinergic side effects that might worsen dementia.

SELECTIVE SEROTONIN REUPTAKE INHIBITORS

Selective serotonin reuptake inhibitor (SSRI) antidepressants are now considered the first choice for the treatment of older adult depression because of their efficacy and favorable side-effect profile. Studies of SSRIs do not demonstrate the superior antidepressant efficacy of any one member of this class over another. Late-life depression associated with dementia, as well as with a medical illness, responds to SSRIs. Most SSRIs can be taken once daily, which increases compliance, especially for older patients who are taking multiple medications. Although SSRIs do not cause orthostatic hypotension, they have been implicated in an increased prevalence of falls in older adult nursing home residents. At therapeutic doses, they are less likely to cause cognitive impairment or anticholinergic and cardiovascular adverse effects. The limitations of these agents include nausea, headache, restlessness, sexual dysfunction, decreased heart rate, occasional insomnia, and occasionally inappropriate antidiuretic hormone secretion. Some SSRIs, especially fluoxetine, fluvoxamine, and paroxetine (and sertraline, to a lesser extent), interfere with cytochrome-metabolizing enzymes. Since hepatic metabolism is the primary route for drug detoxification and elimination, inhibition of these enzymes (especially CYP450 2D6 and 3A4) may raise the blood levels of other medications, causing unwanted side effects and toxicity. Medications whose metabolism may be inhibited include cardiac antiarrhythmic drugs, antidepressants and neuroleptics, some benzodiazepine anxiolytics, steroids, some antihistamines and antibiotics, and antifungal agents. Fluoxetine and paroxetine, in particular, are strong inhibitors of the metabolizing hepatic enzyme CYP450 2D6. This enzyme is necessary to convert tamoxifen to its active metabolite; inhibition by either of these SSRI antidepressants will prevent this conversion and render the tamoxifen useless. Citalopram and escitalopram do not affect hepatic metabolizing enzymes and may be less likely to cause drug interactions. For some clinicians, this lack of enzyme interaction suggests either of these two drugs is a first-choice SSRI antidepressant for older adults who may be taking other medications.

Examination of the role of polymorphisms of the serotonin transporter protein gene indicates that in older individuals (as in young adults), the “s” variant is associated with a decreased reduction in serotonin reuptake and a less robust response to SSRI antidepressants.

Starting doses of SSRIs should be low (paroxetine 10 mg, sertraline 25 mg, citalopram 10 mg, escitalopram, or fluoxetine 10 mg) followed by gradual dosage increases, particularly in those above age 75. Therapeutic dosages will vary, ranging from 20 to 60 mg a day of paroxetine, 50 to 200 mg a day of sertraline, and 10 to 60 mg a day of fluoxetine, citalopram, or escitalopram. Therapeutic blood concentrations of SSRIs are not available for routine clinical use. Studies of augmentation of an SSRI antidepressant with other medications such as methylphenidate, bupropion, thyroid, or estrogen have shown equivocal efficacy in older adults. However, recent data have demonstrated effective therapeutic augmentation of SSRIs with aripiprazole. Escitalopram has also been combined with the glutamatergic drug memantine (usually used for the treatment of AD) for augmentation of its therapeutic effect.

Newer antidepressants with selective serotonin reuptake inhibiting properties are now available. All are effective antidepressants; there are no data to suggest that any are superior to existing SSRIs. These drugs include vilazodone (given in the usual daily dose of 40 mg taken with food), which combines serotonin reuptake blockade with direct effects on synaptic serotonin receptors.

DUAL-ACTING ANTIDEPRESSANTS

There are 10 currently available antidepressants that enhance neuronal transmission in central nervous system mood-regulatory circuits by inhibiting the presynaptic reuptake of both norepinephrine and serotonin: venlafaxine, desvenlafaxine, duloxetine, milnacipran, levomilnacipran, mirtazapine, trazodone, nefazodone, vilazodone, and vortioxetine.

Venlafaxine, an inhibitor of noradrenaline and serotonin reuptake, is an effective and safe antidepressant for older adults. Compared with other antidepressants, it may induce slightly higher rates of remission of depressive symptoms in nonolder adults; data are scant for older adults. Patients respond to a relatively wide range of daily doses, and no dosage adjustment is necessary based on age. The use of venlafaxine entails several precautions. Headache and nausea may be minimized by using lower dosages of 25 mg daily and using the extended-release formulation, gradually titrating up to a therapeutic dosage. Venlafaxine may cause modest increases in diastolic blood pressure, which should be monitored. Venlafaxine use is not associated with drug interactions or enzyme inhibition. Desvenlafaxine is also available for the treatment of depression. This metabolite of venlafaxine is an effective antidepressant at 50 mg/day. Duloxetine is an effective antidepressant at doses similar to those used for younger adults (up to 60 mg/day). Its side-effect profile is mild, and there have been clinical reports that this medication (like other dual-acting antidepressants) may be helpful in controlling pain from arthritis or rheumatoid disorders. Milnacipran is an effective antidepressant for older adults at doses of 75 to 100 mg/day. At these doses, it does not adversely affect the cognitive function or psychomotor performance. Levomilnacipran is now available and is effective in geriatric patients at doses of 20 to 40 mg/day. This antidepressant is primarily cleared through the kidneys, so older patients with renal impairment probably should not be treated with this particular antidepressant. Mirtazapine, through its unique receptor activity, is also considered a dual-acting drug. Mirtazapine is an effective antidepressant in older adults at doses of 15 to 30 mg/day. However, it is quite sedating, so starting doses should be low; mirtazapine may be helpful as a sleep-inducing medication even in the absence of depression. A rapidly dissolving tablet is very useful for nursing home patients who have difficulty swallowing a pill.

Although not officially classified as dual-acting drugs, trazodone and nefazodone also affect both norepinephrine and serotonin neurotransmission. These drugs are weak serotonin reuptake inhibitors and are less commonly used to treat depression. Trazodone, however, is an effective drug for inducing sleep onset. Nefazodone strongly inhibits the CYP450 3A4 metabolizing enzyme;

it should not be prescribed together with benzodiazepine anxiolytics, steroids, antibiotics, antifungal medications, and certain antibiotics, and has been decreasingly used and available for the treatment of depression in all ages. When prescribed mood-stabilizing medications (antiseizure drugs such as carbamazepine, lamotrigine, and divalproex sodium), blood levels of the mood stabilizer may rise, producing unwanted side effects.

Vilazodone and vortioxetine are newer antidepressants that are available for clinical use. Research has been scant on the clinical efficacy of these two drugs in the depressed geriatric population. They appear to be equally effective with each other in older adults and as therapeutically effective as sertraline.

SSRIs and dual-acting antidepressants, such as venlafaxine but not tricyclic antidepressants, are associated with an increased rate of bone loss in older women, especially in the hip. This bone loss is associated with a two-fold increase in clinical fragility fracture since there is an increased risk of falls in nursing home residents who are taking SSRI (or tricyclic) antidepressants.

In general, these dual-acting antidepressants are not likely to produce serious drug interactions with other medications. It is good practice, however, to minimize the total number of medications that are coprescribed with these antidepressants and avoid using more than one antidepressant at the same time. Drugs that increase serotonin function may interact with these dual-acting drugs to produce a serotonin syndrome and should be avoided.

BUPROPION

Bupropion is an effective and safe antidepressant that is often prescribed to older patients because of its activating properties and relatively benign side-effect profile (no weight gain or significant anticholinergic properties). Advantages of bupropion include efficacy, especially in psychomotorically retarded older patients, and no adverse effects on cognition. Bupropion does not cause sedation and may even be stimulating for some older adult patients, which may interfere with sleep. Bupropion may be helpful in depressed older patients with Parkinson disease. However, because of its activating properties, bupropion may increase anxiety in some older individuals, especially those who are already anxious. Bupropion is a strong inhibitor of the metabolizing hepatic enzyme CYP450 2D6. This enzyme is necessary to convert tamoxifen to its active metabolite; inhibition by bupropion will prevent this conversion and render the tamoxifen useless. As with other antidepressants, starting doses should be lower than those recommended for younger adults, and dosage increments should be gradual.

TRICYCLIC ANTIDEPRESSANTS

Tricyclic antidepressants, nortriptyline, and desipramine, in particular, are effective medications for major depressive disorder in older adults. Blood levels of these two drugs necessary to achieve a therapeutic effect in older adults are the same as in younger adults, thus providing the clinician with a laboratory value to guide dosing. Hydroxy metabolites of the tricyclics are potentially cardiotoxic in older individuals and may be present at higher plasma levels than in younger adults because their renal excretion is decreased by advancing age. An electrocardiogram (EKG) can be used to monitor patients in order to avoid approaching cardiotoxicity, which is signaled by a widening of the QTc interval.

Annoying peripheral anticholinergic side effects of tricyclics, particularly constipation and dry mouth, can become severe problems in older adults leading to bowel obstruction and poor dental hygiene. Central nervous system anticholinergic side effects, which usually consist of decreased memory and attention, can progress to a delirious state with severe confusion, disorientation, and paranoid ideation. Orthostatic hypotension, another common adverse effect of tricyclic treatment of late-life depression, may increase the risk of falls; oversedation may interfere with participation in daytime activities.

To prevent these common tricyclic antidepressant side effects, relatively low starting doses with small dosage increments are usually employed. The prescribing maxim “start low (initial dose) and go slow (subsequent dosage increases)" applies to older individuals as a class of patients, although not necessarily to any individual patient.

Plasma levels of imipramine, desipramine, and nortriptyline can be routinely obtained. These levels correlate with therapeutic effects as well as with side effects. Prescribing clinicians who wish to use tricyclic should follow plasma levels of the tricyclic (preferably the secondary amines, desipramine, or nortriptyline) as doses are adjusted to the therapeutic range. Using these levels, along with an EKG to monitor the appearance of a potentially dangerous cardiac arrhythmia, provides the prescribing clinician with two sources of data to guide the use of these medications.

MONOAMINE OXIDASE INHIBITORS

Monoamine oxidase (MAO) inhibitors may be effective in depressed older patients. Their advantages include few anticholinergic effects and no effects on cardiac conduction. MAO inhibitors are potentially dangerous for older individuals and should not be used except by prescribing clinicians who have experience with these medications. The potential for toxic interactions between these drugs and other medications (e.g., serotonin-enhancing drugs; meperidine or other opiates, tricyclic antidepressants) or certain commonly eaten foods (e.g., aged cheese, pickled meats, and fish) is high in older adult population. Consequently, if these medications are to be used, they must be prescribed only to those older patients who are able to follow the dietary and drug interaction limitations.

For appropriate and compliant older patients, problem interactions with food and drugs can be minimized through patient education and collaboration among treating physicians. Limitations in their clinical utility include both hypotension and hypertension, insomnia, and somnolence. The starting dosages of tranylcypromine and phenelzine are 10 mg a day and 15 mg a day in two doses, respectively. In general, maximum dosages do not typically exceed 40 to 50 mg a day for tranylcypromine and 75 to 90 mg a day for phenelzine.

Transdermal selegiline (Emsam) is now available as a skin patch, although there is scant literature regarding its use in older adults. This

MAO inhibitor decreases the problems of drug and food interactions that may occur when high doses of the patch or the interacting medication or food are ingested.

STIMULANTS

Stimulants, such as methylphenidate (Ritalin), are not classified as antidepressants and should not be used as a solo medication to treat an older depressed individual. However, low doses of methylphenidate (2.5 to 5.0 mg/day in divided doses) may be helpful in activating a withdrawn, apathetic older patient, as well as increasing motivation and social participation. Methylphenidate may also enhance the therapeutic effect of antidepressants which have only produced a partial response. Low doses are not associated with tachycardia, hypertension, or cardiac toxicity; doses administered after 3 or 4 PM may interfere with sleep. Other activating drugs (modafinil, armodafinil) may also be used for mild activating effects. Stimulants and other activating drugs should not be given after 2 to 3 PM in order to avoid interfering with sleep onset at bedtime.

TREATMENT OF PSYCHOTIC DEPRESSION IN LATER LIFE

Psychotic depression is a severe illness with profound depressive symptomatology accompanied by delusions, often of somatic dysfunction, hopelessness, or guilt. Psychotic depression poses a health risk from progressive weight loss and exhaustion as well as from suicide, usually by violent means. For this reason, recognition and treatment of psychotic depression can be lifesaving. Both young and older patients with psychotic depression respond poorly to short-term treatment with antidepressants given alone. Combinations of antidepressants and antipsychotic medications are effective in psychotic depression in both controlled studies of younger patients and naturalistic treatment studies of older patients. Although tricyclics were the most frequently prescribed antidepressants in conjunction with an antipsychotic drug, clinicians are now using a combination of

SSRIs or dual-acting antidepressants together with atypical neuroleptics to reduce the side effects of older medications.

Electroconvulsive therapy (ECT), rapidly effective and safe, remains the treatment of choice for severe geriatric psychotic and treatment-resistant severe depressions. Concern over potential cognitive impairment from ECT, although often transient, has limited its usefulness. Recent data on a study of ultrabrief pulse width and right unilateral electrode placement demonstrated fewer neurocognitive adverse effects. When combined with venlafaxine in this older adult sample, cognition remained relatively modest or stable.

TREATMENT OF DEPRESSION ASSOCIATED WITH MEDICAL ILLNESS AND COGNITIVE IMPAIRMENT

The National Institutes of Health (NIH) Consensus Development Conference on the Diagnosis and Treatment of Depression in Late Life stated that “the hallmark of depression in older people is its comorbidity with medical illness.” Compelling evidence from placebo-controlled clinical trials indicates that antidepressant medications are efficacious in patients with common comorbid medical disorders as diverse as stroke, Parkinson disease, AD, cancer, ischemic heart disease, chronic obstructive pulmonary disease, and arthritis. It appears that all antidepressants are therapeutic, at least to some degree, even in rehabilitation units where treatment of depressed patients can facilitate recovery from acute illness or injury and in a residential care setting, where patients are characterized by extreme old age and the interacting effects of many of the common disabling medical disorders of later life. Until recently**,** tricyclic antidepressants were the most widely studied drugs for the treatment of depression associated with another illness. They have demonstrated safety and effectiveness for poststroke patients, as well as for older depressed patients with Parkinsos disease, cardiovascular illness, and cancer. However, interactions between tricyclic antidepressants and other nonpsychiatric medications used to treat medical or neurologic illnesses can make older patients highly vulnerable to the development of side effects. Thus, careful monitoring of both potential adverse effects and plasma concentrations is necessary to optimize the safety of these agents in frail older patients. SSRIs and other nontricyclics are now more widely used in order to avoid this potential for adverse drug events; SSRIs have been shown effective for depressed patients with cardiovascular illness, stroke, Parkinson disease, and cancer.

Depression and cognition intersect in two ways: older individuals with failing cognition or who are suffering from dementia commonly are depressed; depressed older individuals frequently complain of impaired memory and concentration. This interaction between depression and diminished cognitive function has been well studied. Regardless of whether the primary illness is dementia or depression, depression should be treated following the basic principles that apply to older patients. Medication use should be supervised, and caregivers should assess the benefits and monitor for adverse effects of treatment. Patients with dementia of Alzheimer type are highly sensitive to increased cognitive impairment from anticholinergic medications. Therefore, antidepressants with the lowest anticholinergic effects are preferred, which include SSRIs, low doses of bupropion, venlafaxine, mirtazapine, and vilazodone. Low doses of nortriptyline are sometimes helpful as well. Memory and concentration often improve as depression responds to medication; in mild to moderate degrees of dementia, successful antidepressant treatment can restore or even transiently improve memory and concentration. However, as dementia worsens, antidepressants may cause increased agitation and must be discontinued. The study by Lisanby et al. discussed above suggests that new ECT administration techniques combined with a dual-acting antidepressant may offer clinical improvement for a patient with late-life depression without further increasing cognition.

ESSENTIALS OF PATIENT MANAGEMENT

Several principles of clinical practice should guide the clinician who wishes to treat a depressed older individual. Careful psychiatric and medical diagnoses go hand in hand; older depressed individuals must be medically evaluated, especially for thyroid dysfunction and levels of B_{12} and folate. For some older people, this evaluation should include a neurologic examination to rule out Parkinson disease or dementia. It is well established that inflammation and inflammatory disorders are commonly associated with mood disturbance, especially depression. Chronic illnesses such as diabetes are also commonly associated with depression. Poststroke and postmyocardial infarction depressions are similarly well known, and failure to treat the depression may increase morbidity or subsequent mortality. A comprehensive list of all medications (prescribed or self-purchased), including herbs, holistic remedies, and homeopathic preparations, must be obtained since medication-induced (or worsened) depression is common in older adults. The clinician must also determine whether or not alcohol (or other drugs) is being used excessively or inappropriately before embarking on a treatment program. It may be necessary for the clinician to obtain this information from a spouse, child, or other sources, as the depressed older person may not offer an accurate history.

Treatment-Resistant Depression

Recurrent depressive illness often becomes progressively resistant to treatment as a person ages and has more recurrences. Consequently, the usual medication recommendations may be insufficient to achieve therapeutic response or remission. The use of ECT has already been discussed, and the use of ketamine in older adults is currently under investigation. Augmentation of antidepressant therapeutic effects has become more common among geriatric patients, although the risk of drug interactions and potential drug toxicity must be considered. Among substances that have been found useful in younger treatment-resistant patients may be useful in older adults if attention is paid to the need for age-adjusted dosing, the potential for polypharmacy, and behavioral and metabolic drug interactions. More drugs are necessarily better for older depressed patients, but judicious augmenting agents may be helpful. Lithium has recently been shown to have therapeutic augmenting effects, as has thyroid, low-dose stimulants, and atypical antipsychotic agents.

Bipolar Depression

As individuals with bipolar illness age, the pattern of their mood dysregulation commonly shifts from a euphoria/dysphoria oscillation to a mixed picture of both moods and, finally, in old age, primarily a dysphoric mania or treatment-resistant depression. Older patients with a history of bipolar episodes in younger life must be very carefully treated with an antidepressant, usually in conjunction with a mood-stabilizing medication such as lithium or, preferably, an anticonvulsant such as lamotrigine. Monotherapy with an antidepressant frequently causes a switch in mood from depression to a hypomanic mixed, or dysphoric hypomania that may be difficult to treat and medically toxic to the older individual. Therefore, unless there a contravening reasons, any antidepressant given to an older bipolar depressed patient should also receive an adjunctive mood stabilizing medication or an antipsychotic drug. No specific antidepressant is indicated for the treatment of bipolar depression, but lamotrigine is a frequently used mood stabilizer for protection against a mood switch, and there are reports of its therapeutic effects for depression. Lurasidone, asenapine, and aripiprazole have also been reported to be effective in late-life bipolar depression.

Managing the depressed geriatric patient also includes educating the patient and family about the nature of depression as a true illness, emphasizing that it is treatable. Patients and their families need to understand that response to medication may take longer than for younger adults, that treatment adherence is essential, and that adverse effects can be expected but are amenable to supportive countermeasures. Weekly contact during the early phases of short-term treatment is recommended for individualizing treatment, providing encouragement and support, reviewing questions and concerns, and monitoring safety (including suicidal potential). Common errors in practice include underdosing, discontinuing treatment too soon, failure to consider possible drug interactions, and polypharmacy. If a patient has not begun to respond by 6 to 8 weeks, compliance should be re-evaluated, dosage or plasma concentration should be optimized, and referral to a geriatric psychiatrist experienced in treatment resistance should be considered.

Late-life depression is often a chronic illness and tends to recur, especially among those whose first depression develops after the age of 60. One must bring a longitudinal model of depression to treatment. Short-term treatment is intended to induce remission, but incomplete remissions are not unusual and commonly lead to relapse. Recurrences after a longer period of recovery are also common. Studies continue to demonstrate the therapeutic benefits of long-term treatment of depression, and it is recommended, therefore, that treatment continue at least 6 months beyond recovery and for longer periods after a second or third episode. Many older patients have chronic depression that requires indefinite maintenance therapy. Long-term prevention of new episodes is best achieved by maintaining patients on the same dosage of antidepressant medication that was used successfully to treat the acute episode.

Treatment of Bipolar Disorders in Later Life. Bipolar disorders constitute 5% to 10% of the diagnoses of older psychiatric inpatients. The diagnosis of a manic episode is usually not difficult in older individuals, although advanced age is associated with greater degrees of irritability and less euphoria. Most manic older adult patients have a history of prior manic episodes; the onset of a first manic episode in late life is unusual and often is associated with another illness, such as thyroid dysfunction or alcohol or substance abuse. Older individuals who have suffered repeated episodes of mania and depression over the course of their lives are likely to be quite dysfunctional when manic, with rapid shifts of mood. Episodes of euphoria decrease with advanced age and are replaced by “dysphoric” or “mixed” mood states characterized by anger, irritability, paranoid ideation, and depression**.** Diagnostic principles follow the suggestions for evaluation of an older depressed patient, therefore, including a careful and comprehensive medical evaluation, laboratory screening, and history of past or current drug and alcohol use. Older patients with bipolar disorder have a higher body mass index (BMI) and a greater burden of endocrine, metabolic, and respiratory disease.

Despite growing clinical experience in treating older individuals with bipolar disorder, the research and clinical trials literature is less extensive than the geriatric antidepressant literature. In principle, therefore, the management of mania in older patients is similar to that in younger patients, although there may be a relatively poor response to standard pharmacologic treatment in patients with long-standing bipolar illness. Treatment of the acute manic episode is focused on two separate categories of dysfunction: (1) management of manic behavior and impaired sleep, and (2) regulation of excessive or inappropriate mood states. Agitated, hyperactive older manic patients or those who are psychotic, as well as patients with impaired sleep, may be helped by treatment with antipsychotic drugs. All antipsychotics are effective for these symptoms, although current practice favors the use of second-generation drugs in order to avoid excessive extrapyramidal symptoms caused by conventional neuroleptics. Risperidone, olanzapine, and quetiapine are commonly used for this purpose; clozapine is also therapeutic, but its use in older patients is associated with numerous potentially serious side effects; other medications (e.g., ziprasidone, aripiprazole, lurasidone, olanzapine asenapine) may also be helpful, but research studies in older adults of any of these medications are scant and do not provide adequate prescribing information. In general, therefore, starting doses of an atypical antipsychotic drug should be low, with gradual dose increases; clinicians should resist the pressure to rapidly increase doses to control behavioral disruption. When an atypical neuroleptic is given at bedtime, sleep is often promoted, which, in turn, may help diminish the manic symptoms.

Lithium

Regulation of mood is accomplished by the use of mood-stabilizing medications; principles of their use follow guidelines for younger adults, although older individuals are significantly more sensitive to side effects. Lithium, for example, commonly produces side effects at doses and blood levels that are not toxic in younger people. For example, delirium can develop at lithium concentrations at or even below the young adult therapeutic lithium plasma concentration range. Delirium and cerebellar dysfunction may last for weeks after discontinuation of lithium use. Lithium-induced tremor is more frequent in older patients than in younger patients. Lithium can worsen Parkinson disease or produce parkinsonism in neurologically intact older patients and can also worsen cognitive function, especially in demented patients. Sinoatrial block can be caused by lithium, a risk that is increased by the concurrent use of digitalis and β-adrenergic receptor antagonists (β-blockers). Older patients often receive drugs that increase lithium plasma concentrations, including thiazide diuretics, nonsteroidal anti-inflammatory drugs, or angiotensin-converting enzyme inhibitors. Age-induced decrements in glomerular filtration rates and renal blood flow reduce lithium excretion, and a decline in glomerular filtration rate may accompany long-term lithium use. Because of the potential for toxicity and impairment of renal function, lithium use should be restricted to selected older patients whose physical health and ability to comply with prescribing guidelines and laboratory studies are evident. Starting dosages of 75 to 150 mg a day may reduce adverse effects and improve compliance. These very low doses may successfully treat the disturbed mood with lower, nontoxic blood levels than in younger adults (e.g., 0.2 to 0.4 mM/L). Chronic lithium treatment may also adversely affect thyroid function leading to a hypothyroid state that may resemble depression or aggravate pre-existing depression. It is essential, therefore, that lithium use in older adults be accompanied by regular checks of renal and thyroid function.

Anticonvulsant Mood Stabilizers

It is now well established that some anticonvulsant medications also are excellent mood stabilizers and may be effective and safe for older individuals. Like lithium, however, anticonvulsants are used primarily to prevent the recurrence of dysregulated mood states. For example, the anticonvulsants carbamazepine (Tegretol) and divalproex sodium

(Depakote) have demonstrated prophylactic effectiveness in older individuals. However, these two anticonvulsants affect cytochrome P450 metabolizing enzymes and may affect blood levels of other medications given for psychiatric or medical conditions. Carbamazepine induces enzyme metabolism, while Depakote inhibits metabolism. Current prescribing practice favors the use of lamotrigine (Lamictal) as a stabilizing anticonvulsant for older adults. Although a few individuals may develop a skin rash with this drug (less likely in older adults), the side-effect profile is benign, and with low therapeutic doses (12.5 to 100.0 mg/day), there may be few or no side effects. Oxcarbazepine (Trileptal) is also sometimes used to stabilize mood in younger adults, but there are no studies of its effectiveness in older adults.

Most anticonvulsants are metabolized by hepatic metabolizing enzymes, primarily CYP 3A4. Hematologic and liver function tests are needed prior to treatment with anticonvulsants to ensure their safe use. Carbamazepine induces its own metabolism, so blood levels may need to be adjusted during treatment. This drug can also cause leukopenia (white blood cell counts below 4,000 /mm) in approximately 2% of cases; it should not be used for routine geriatric patient treatment. Valproate inhibits the CYP450 3A4 metabolizing enzyme and can cause elevations of other medications that use this metabolic pathway (e.g., lamotrigine, some benzodiazepines, and antidepressants). Valproate can also cause increased ammonia levels and toxic neurologic states, although it is not clear that older individuals are more susceptible to this side effect. Lamotrigine, currently the preferred mood-stabilizing anticonvulsant for older adults, is metabolized by CYP450 3A4, and its blood levels can be affected by other medications (e.g., valproate, fluoxetine) that inhibit this enzyme. Oxcarbazepine can cause a sudden decrease in serum sodium, followed by a change in mental status that may be mistaken for a worsening psychiatric condition. Levetiracetam has been used as a mood stabilizer in a small number of older patients. Other new anticonvulsants have not been tried on older adults and cannot be recommended.

Prescribing guidelines for lithium and anticonvulsant treatment of late-life mood dysregulation follow the general geriatric prescribing principle: “start low and go slow.”

Continuation and maintenance treatment is necessary for older bipolar patients who achieve remission of a manic or depressive episode. Although studies of geriatric populations are lacking, studies of younger patients suggest that lithium, carbamazepine, and valproate are effective in preventing both depressive and manic episodes. For this reason, continuation and maintenance of treatment with these drugs should be considered for older patients. Older patients, in particular, should be carefully and frequently monitored for alterations in medical conditions, especially hepatic function, as well as cognitive function. As the older bipolar patient continues to age, dosage modifications are likely to be necessary.

FURTHER READINGS

Behlke LM, Lenze EJ, Carney RM. The cardiovascular effects of newer antidepressants in older adults and those with or at high risk for cardiovascular diseases. *CNS Drugs*. 2020;34:1133–1147.

Beyer JL, Johnson KG. Advances in pharmacotherapy of late-life depression. *Curr Psychiatry Rep*. 2018;20:34.

Borhannejad F, Shariati B, Naderi S, et al. Comparison of vortioxetine and sertraline for treatment of major depressive disorder in elderly patients: a double-blind randomized trial. *J Clin Pharm Ther*. 2020:45:804–811.

Cassano T, Calcagnhi S, Carbone A, et al. Pharmacological treatment of depression in Alzheimer’s disease: a challenging task. *Front Pharmacol*. 2019:10:1067.

Dols A, Kupka RW, van Lammeren A, Beekman AT, Sajatovic M, Stek ML. The prevalence of late-life mania: a review. *Bipolar Disord*. 2014;16:113–118.

Ellison JM, Sivrioglu EY, Salzman C. Pharmacotherapy of late-life depression: evidence-based recommendations. In: Ellison JM, Kyomen HH, Verma S, eds. *Mood Disorders in Later Life*. 2nd ed. Informa Healthcare USA, Inc.; 2009:239–290.

Forester BP, Jordan B. Bipolar disorder in later life. In: Ellison JM, Kyomen HH, Verma S, eds. *Mood Disorders in Later Life*. 2nd ed. Informa Healthcare USA, Inc.; 2009:65–90.

Garnock-Jones KP. Vortioxetine: a review of its use in major depressive disorder. *CNS Drugs*. 2014;28:855–874.

Gedulig ET, Kellner CH. Electroconvulsive therapy in the elderly: new findings in geriatric depression. *Curr Psychiatry Rep*. 2016;18:40.

Jacobson SA. *Clinical Manual of Geriatric Psychopharmacology*. 2nd ed. American Psychiatric Publishing, Inc.; 2014:536.

Kok RM, Reynolds CF 3rd. Management of depression in older adults. A review. *JAMA*. 2017;317:2114–2122.

Krause M, Gutsmiedl B, Bighelli I, Schneider-Thoma J, Chaimani A, Leucht S. Efficacy and tolerability of pharmacological and non-pharmacological interventions in older patients with major depressive disorder: a systematic review, pairwise and network meta-analysis. *Eur Neuropsychopharmacol*. 2019;29:1003–1022.

Krause-Sorio B, Siddarth P, Michaela M, et al. Regional white matter integrity predicts treatment response to escitalopram and memantine in geriatric depression: a pilot study. *Front Psychiatry*. 2020;11:548904.

Lisanby SH, McClintock SM, Alexopoulos G, et al. Neurocognitive effects of combined electroconvulsive therapy (ECT) and venlafaxine in geriatric depression: phased 1 of the PRIDE study. *Am J Geriatr Psychiatry*. 2020;28:304–316.

Mallery L, MacLeod T, Allen M, et al. Systematic review and meta-analysis of second-generation antidepressants for the treatment of older adult with depression: questionable benefit and considerations for frailty. *BMC Geriatr*. 2019;19:306.

Masse-Sibille C, Djamila B, Julie G, Emmanuel H, Pierre V, Gilles C. Predictors of response and remission to antidepressants in geriatric depression: a systematic review. *J Geriatr Psychiatry Neurol*. 2018;31:283–302.

Maust DT, Oslin DW, Thase ME. Going beyond antidepressant monotherapy for incomplete response in nonpsychotic late-life depression: a critical review. *Am J Geriatr Psychiatry*. 2013;21:973–986.

Mellen EJ, Harper DG, Ravichandran C, Jensen E, Silveri M, Forester BP. Lamotrigine therapy and biomarkers of cerebral energy metabolism in older age bipolar depression. *Am J Geriatr Psychiatry*. 2019;27:783–793.

Patel K, Petal SA, Tarek KR, Mulsant BH. Pharmacotherapy of major depression in late life: what is the role of new agents? *Expert Opin Pharmacother*. 2017;18:599–609.

Pimontel MA, Solomonov N, Oberlin L, et al. Cortical thickness of the salience network and change in apathy following antidepressant treatment for late-life depression. *Am J Geriatr Psychiatry*. 2021;29(3):241–248.

Rawson KS, Dixon D, Civitelli R, et al. Bone turnover with venlafaxine treatment in older depressed adults. *J Am Geriatr Soc*. 2017:65:2057–2063.

Robinson DS, Kajdasz DK, Gallipoli S, Whalen H, Wamil A, Reed CR. A 1-year, open-label study assessing the safety and tolerability of vilazodone in patients with major depressive disorder. *J Clin Psychopharmacol*. 2011;31:643–646.

Saade YM, Nicol G, Lenze EJ, et al. Comorbid anxiety in late-life depression: relationship with remission and suicidal ideation on venlafaxine treatment. *Depress Anxiety*. 2019;36:1125–1134.

Sajatovic M, Dines P, Fuentes-Casiano E, et al. Asenapine in the treatment of older adult with bipolar disorder. *Int J Geriatr Psychiatry*. 2015;30:710–719.

Sajatovic M, Gildengers A, Al Jurdi RK, et al. Multisite, open-label, prospective trial of lamotrigine for geriatric bipolar depression: a preliminary report. *Bipolar Disord*. 2011;13:294–302.

Salzman C. Treatment of depression. In: Salzman C, ed. *Clinical Geriatric Psychiatry.* 4th ed. Williams and Wilkins; 2005.

Sheline YI, Disabato BM, Hranilovich J, et al. Treatment course with antidepressant therapy in late-life depression. *Am J Psychiatry*. 2012;169:1185–1193.

Shulman KI, Almeida OP, Herrmann N, et al. Delphi survey of maintenance lithium treatment in older adults with bipolar disorder: an ISBD task force report. *Bipolar Disord*. 2019:21:117–123.

Siddarth P, Funes CM, Laird KT, Ercoli L, Lavretsky H. Predictors of cognitive improvement following treatment for late-life depression. *J Geriatr Psychiatry Neurol*. 2021;34(2):162–168.

Tedeschini E, Levkovitz Y, Iovieno N, Ameral VE, Nelson JC, Papakostas GI. Efficacy of antidepressants for late-life depression: a meta-analysis and meta-regression of placebo-controlled randomized trials. *J Clin Psychiatry*. 2011;72:1660–1668.

55.4c Antianxiety Drugs

ALASTAIR J. FLINT, M.B., AND KATHLEEN S. BINGHAM, M.D., PH.D.

Anxiety disorders are common in late life, with 5% to 10% of community-dwelling older adults meeting the criteria for a current anxiety disorder. Despite this prevalence, there has been relatively little research on the pharmacologic treatment of anxiety in older adults, and most treatment decisions in this group need to rely on research conducted in younger adults. Extrapolating findings from younger to older adults is, however, potentially problematic, given the pharmacokinetic and pharmacodynamic changes that occur with aging. In particular, the efficacy, tolerability, and safety of drugs in older adults can be affected by age-related physiologic changes (such as the increased volume of distribution of lipophilic drugs, decreased function of the liver and kidneys, and changes in organ receptor sensitivity), the increased frequency of comorbid medical and neurologic conditions, and the fact that older adults take more medications than their younger counterparts, thereby increasing the risk of drug–drug and drug–body interactions. This chapter will highlight issues that are of particular relevance to older adults in the pharmacologic treatment of anxiety.

Double-blind, randomized controlled trials (RCTs) are the gold standard for evaluating treatment efficacy. Nonblind ("open-label") and uncontrolled studies are of limited use in guiding treatment because of the significant biases and confounds that can affect their findings. There have been 11 double-blind RCTs of the pharmacologic treatment of anxiety disorders in persons aged 60 years or older (Table 55.4c–1), but only nine studies pertain to drugs approved for sale in North America. Nine of the 11 RCTs focused on generalized anxiety disorder (GAD), while two studies included participants with nonspecific anxiety.

Given that the diagnostic category of anxiety disorders has traditionally included phobic disorders, panic disorder, GAD, post-traumatic stress disorder (PTSD), and obsessive-compulsive disorder (OCD), why have treatment studies in the elderly focused almost exclusively on GAD? Knowledge of the epidemiology of anxiety disorders in later life helps answer this question. First, phobias and GAD are the most prevalent anxiety disorders in older adults. However, the most common presentations of phobia in later life, that is, agoraphobia without panic and specific phobia, tend not to come to psychiatric attention. In addition, the treatment of choice for these two disorders is behavioral therapy, not pharmacotherapy. Therefore, GAD is the most frequently occurring late-life anxiety disorder in psychiatric practice that is medication-responsive. Second, panic disorder, OCD, and social anxiety disorder (social phobia) have a low incidence and prevalence in later life, and therefore clinicians and researchers may view them as less compelling to study compared with disorders with a higher incidence and prevalence. In contrast, approximately half of the cases of syndromal generalized anxiety in older adults are late onset, often in association with major depression. Finally, syndromal and subsyndromal generalized anxiety is the most common manifestation of anxiety associated with medical illness and cognitive impairment in older persons, again frequently in association with depression.

The 11 RCTs in Table 55.4c–1 have evaluated the efficacy and tolerability of antidepressants, benzodiazepines, nonbenzodiazepine anxiolytics, an anticonvulsant, and an atypical antipsychotic in the treatment of late-life GAD. The remainder of this chapter will discuss these classes of drugs, as well as comment on miscellaneous drugs that have been examined in younger persons with anxiety.

Table 55.4c–1.
All Published Double-Blind Randomized Controlled Trials of the Pharmacologic Treatment of Anxiety Disorders in Persons Aged 60 Years or Older (Circa 2022)

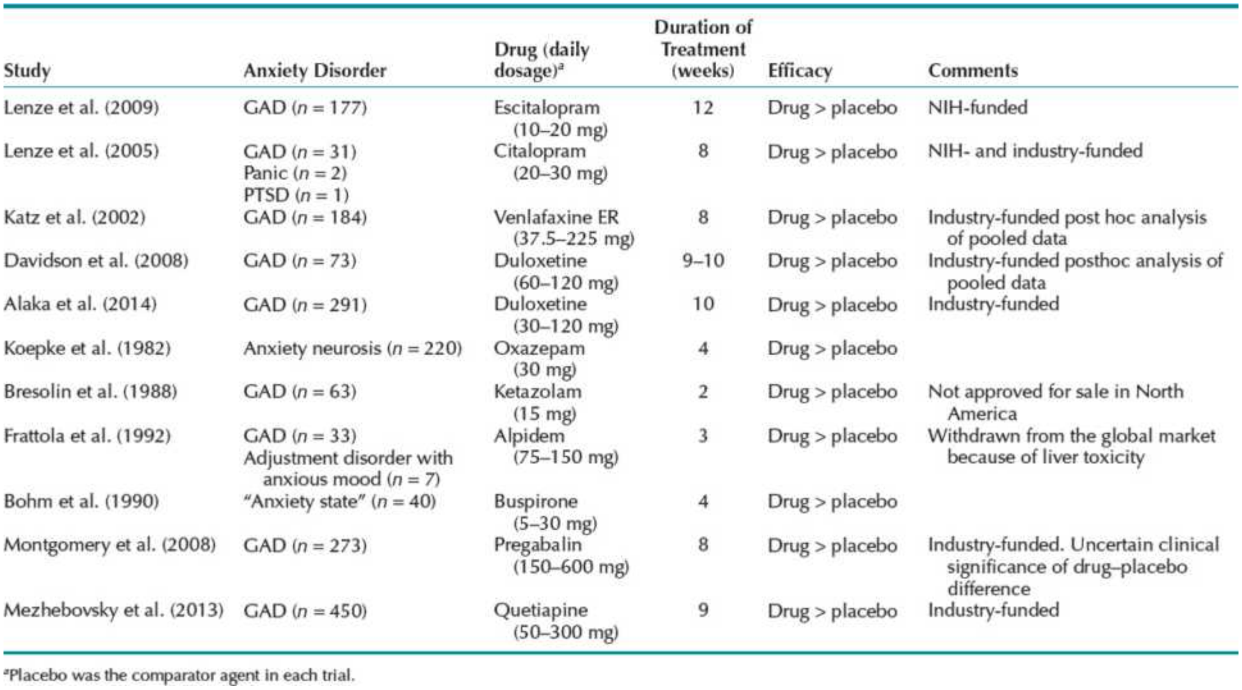

Study	Anxiety Disorder	Drug (daily dosage)[a]	Duration of Treatment (weeks)	Efficacy	Comments
Lenze et al. (2009)	GAD (*n* = 177)	Escitalopram (10–20 mg)	12	Drug > placebo	NIH-funded
Lenze et al. (2005)	GAD (*n* = 31) Panic (*n* = 2) PTSD (*n* = 1)	Citalopram (20–30 mg)	8	Drug > placebo	NIH- and industry-funded
Katz et al. (2002)	GAD (*n* = 184)	Venlafaxine ER (37.5–225 mg)	8	Drug > placebo	Industry-funded post hoc analysis of pooled data
Davidson et al. (2008)	GAD (*n* = 73)	Duloxetine (60–120 mg)	9–10	Drug > placebo	Industry-funded posthoc analysis of pooled data
Alaka et al. (2014)	GAD (*n* = 291)	Duloxetine (30–120 mg)	10	Drug > placebo	Industry-funded
Koepke et al. (1982)	Anxiety neurosis (*n* = 220)	Oxazepam (30 mg)	4	Drug > placebo	
Bresolin et al. (1988)	GAD (*n* = 63)	Ketazolam (15 mg)	2	Drug > placebo	Not approved for sale in North America
Frattola et al. (1992)	GAD (*n* = 33) Adjustment disorder with anxious mood (*n* = 7)	Alpidem (75–150 mg)	3	Drug > placebo	Withdrawn from the global market because of liver toxicity
Bohm et al. (1990)	"Anxiety state" (*n* = 40)	Buspirone (5–30 mg)	4	Drug > placebo	
Montgomery et al. (2008)	GAD (*n* = 273)	Pregabalin (150–600 mg)	8	Drug > placebo	Industry-funded. Uncertain clinical significance of drug–placebo difference
Mezhebovsky et al. (2013)	GAD (*n* = 450)	Quetiapine (50–300 mg)	9	Drug > placebo	Industry-funded

[a] Placebo was the comparator agent in each trial.

ANTIDEPRESSANTS

In recent years, there has been growing awareness that antidepressant medications have a broad spectrum of efficacy in the treatment of anxiety disorders in younger adults. RCTs in mixed-age patients have found that antidepressants of various classes are beneficial in the treatment of GAD, panic disorder, social anxiety disorder, PTSD, and OCD (Table 55.4c–2).

RCTs of antidepressant pharmacotherapy in older people with an anxiety disorder found that escitalopram (Lexapro), citalopram (Celexa), and Duloxetine (Cymbalta) are more efficacious than placebo in treating GAD (Table 55.4c–1). In addition, a post hoc analysis of data pooled from RCTs found that venlafaxine (Effexor) had greater efficacy than placebo in older persons with GAD. In the escitalopram, citalopram, and venlafaxine studies, approximately two-thirds of participants treated with the drug were classified as responders after 8 to 12 weeks of treatment, compared with placebo response rates of 24% for citalopram, 41% for venlafaxine, and 51% for escitalopram. In the duloxetine clinical trial, the 10-week rate of response was 71.3% for duloxetine and 45.5% for placebo. Thus, there is a high rate of placebo response in clinical trials of later-life GAD, highlighting the importance of nonpharmacologic factors contributing to symptomatic improvement in these studies. The venlafaxine analysis also compared response rates between younger and older participants and found almost identical rates of response to drug and placebo between the two age groups. Compared to placebo, statistically significant adverse events were: fatigue/somnolence, sleep disturbance, and urinary symptoms with escitalopram; gastrointestinal symptoms with venlafaxine; and dry mouth, nausea, and weight loss with duloxetine. There was no significant difference in adverse effects between citalopram and placebo. However, the analysis was limited by the small number of participants in that study.

Table 55.4c–2.
Drugs Found Efficacious in Randomized Controlled Trials Treating Anxiety Disorders in Mixed-Age Samples

Disorder	Drug
Generalized anxiety disorder (GAD)	SSRI: escitalopram[a], fluoxetine, paroxetine[a] SNRI: duloxetine[a], venlafaxine[a] TCA: imipramine Other Antidepressants: trazodone Benzodiazepine: various Azapirone: buspirone[a] Other: hydroxyzine, pregabalin, quetiapine
Panic disorder with or without agoraphobia	SSRI: citalopram, fluoxetine[a], fluvoxamine, paroxetine[a], sertraline[a] SNRI: venlafaxine[a] TCA: clomipramine, desipramine, imipramine MAOI: brofaromine, phenelzine Benzodiazepine: alprazolam[a], clonazepam[a], diazepam, lorazepam
Social anxiety disorder	SSRI: fluvoxamine, paroxetine[a], sertraline[a] SNRI: venlafaxine[a] MAOI: brofaromine, moclobemide, phenelzine Benzodiazepine: alprazolam, clonazepam
Post-traumatic stress disorder (PTSD)	SSRI: fluoxetine, paroxetine[a], sertraline[a] SNRI: venlafaxine TCA: amitriptyline, imipramine MAOI: phenelzine Other: prazosin (nightmares)
Obsessive-compulsive disorder	SSRI: fluoxetine[a], fluvoxamine[a], paroxetine[a], sertraline[a] TCA: clomipramine[a]

[a] FDA-approved.

SSRI, selective serotonin reuptake inhibitor; SNRI, serotonin and norepinephrine reuptake inhibitor; TCA, tricyclic antidepressant; MAOI, monoamine oxidase inhibitor.

Selective Serotonin Reuptake Inhibitors

Of the selective serotonin reuptake inhibitors (SSRIs), escitalopram and citalopram are well suited to older persons. They have linear pharmacokinetics across the dose range (that is, an increase in dose results in a predictable increase in plasma concentration), no clinically active metabolites, elimination half-lives of 33 hours or less, and little propensity for drug–drug interactions. In addition, sertraline (Zoloft) is also suitable in older persons because of its favorable pharmacokinetic profile. Although data from RCTs in mixed-age patients have demonstrated the efficacy of sertraline in several anxiety disorders (Table 55.4c–2), there are no controlled data pertaining to sertraline in late-life anxiety disorders, although it has been found to be efficacious, safe, and well-tolerated in late-life depression. On the other hand, paroxetine and fluoxetine have less favorable pharmacokinetic profiles in older adults, and paroxetine has the potential to cause anticholinergic adverse effects such as worsening cognitive function, blurred vision, dry mouth, constipation, and urinary hesitancy; therefore, these drugs are not recommended as first-line SSRIs in older adults.

If started at too high a dose, SSRIs may initially exacerbate symptoms of anxiety and, therefore, a lower starting dose (for example, 2.5 mg/day of escitalopram, 5 mg/day of citalopram, or 25 to 50 mg of sertraline) is often indicated in older adult anxious individuals. Once it has been established that the patient is tolerating the starting dose, the dose should then be gradually increased to the usual therapeutic range (Table 55.4c–3). Both citalopram and escitalopram have been associated with dose-dependent prolongation of the QTc interval on EKG and, as a result, it is recommended not to exceed 20 mg/day of citalopram in persons aged 60 years or older and 10 mg/day of escitalopram in persons 65 years or older. Anxious patients frequently misattribute somatic symptoms of anxiety to the adverse effects of medication. Thus, in addition to starting with a low dose of antidepressant, adherence to medication can be enhanced by discussing potential side effects in advance, frequent monitoring of the patient during the first few weeks of treatment, and availability of the clinician by telephone between appointments to address any questions or concerns that the patient may have about treatment.

Based on the above considerations, escitalopram is a reasonable first-line choice for the treatment of GAD in later life, given that it has FDA approval for this condition and there are efficacy and tolerability data specific to older adults. In the absence of pharmacotherapy data pertaining to anxiety disorders other than GAD in older adults, sertraline would be a reasonable first-line SSRI for panic, social anxiety, PTSD, and OCD, given that it has FDA approval for the treatment of these disorders based on studies of mixed-age samples, it has a favorable pharmacokinetic and tolerability profile in older patients, and there is considerable flexibility in dosing without the concern of QTc prolongation at higher doses.

Table 55.4c–3.
Usual Starting Dose, Initial Target Dose, and Recommended Maximum Dose of Selected SSRIs and SNRIs in Older Adults with Anxiety Disorders

Drug	Usual Starting Dose (mg/day)	Initial Target Dose (mg/day)	Recommended Maximum Dose (mg/day)
Escitalopram	2.5	5	10
Citalopram	5	10	20
Sertraline	25	50–100	200
Venlafaxine ER	37.5	150	300
Duloxetine	30	60	120

SSRI, selective serotonin inhibitor; SNRI, serotonin and norepinephrine reuptake inhibitor.

Serotonin and Norepinephrine Reuptake Inhibitors

Venlafaxine is a serotonin and norepinephrine reuptake inhibitor (SNRI), although it has been postulated that its noradrenergic effect occurs only at higher doses (possibly 150 mg/day or higher). The doses used to treat GAD in older persons in RCTs were relatively low (75 to 150 mg/day in the majority of older subjects), raising the possibility that the serotonergic action of the drug may be sufficient to treat GAD in many older patients. Like escitalopram, citalopram, and sertraline, venlafaxine has linear pharmacokinetics and lacks a propensity for significant drug–drug interactions. The usual starting dose of venlafaxine ER in older persons is 37.5 mg/day, with dose increments of 37.5 mg every 4 to 5 days to reach a minimum target dose of 150 mg/day. The dose range of venlafaxine ER is wide (75 to 375 mg/day), and the usual therapeutic dose for GAD has not been established in older patients. However, if there has not been a satisfactory improvement of GAD after 4 weeks of treatment at a dose of 150 mg/day, the dose should be gradually increased, guided by response and tolerability (Table 55.4c–3). Venlafaxine is a suitable antidepressant for older persons with GAD, and on the basis of RCTs conducted in mixed-age samples, it can also be considered as a first-line option for panic, social anxiety, and PTSD. Given that Venlafaxine and SSRIs are different drug classes, switching from one class to the other (i.e., from an SSRI to venlafaxine or vice versa) is a reasonable approach in the case of inadequate response to a therapeutic trial of the first drug.

Duloxetine has a more circumscribed role in the treatment of anxiety since it is indicated only in the treatment of GAD (Table 55.4c–2). Duloxetine has linear pharmacokinetics across a dose range of 30 to 120 mg/day. In older adults, a starting dose of 30 mg/day is recommended, increasing to an initial target dose of 60 mg/day if tolerated. If there has been insufficient improvement after 4 weeks, the dose can be gradually increased by 30 mg increments to a maximum of 120 mg/day, guided by response and tolerability. Duloxetine is both an inhibitor and substrate of cytochrome P450 2D6 (CYP2D6) and thus has the potential for interaction with other drugs metabolized by this enzyme. Given that CYP2D6 catalyzes the metabolism of a large number of drugs, this unwanted property of duloxetine should be taken into account when deciding on the choice of an SNRI to treat late-life GAD.

Other Antidepressants

In addition to SSRIs and SNRIs, selected tricyclic antidepressants (TCAs) and monoamine oxidase inhibitors (MAOIs) have been found to be efficacious in a range of anxiety disorders in younger adults (Table 55.4c–2), although there have been no comparable studies in older adults. With the exception of desipramine (Norpramin), the TCAs that have been examined in younger persons have been tertiary amine compounds, which are potentially problematic in older adults. In particular, older adults may be vulnerable to the anticholinergic, sedative, hypotensive, and adverse cardiac effects of tertiary amine TCAs, especially if they have concomitant physical or cognitive problems that could be aggravated by these effects. A post hoc analysis of data pooled from three RCTs of the treatment of poststroke depression found that nortriptyline (Aventyl) (a secondary amine TCA that is preferred over tertiary amine TCAs in older adults) was more effective than placebo in treating generalized anxiety associated with the depression. Thus, if a TCA is considered for an older anxious person, the recommendations would be desipramine in the case of panic disorder and either nortriptyline or desipramine (which is a metabolite of imipramine [Tofranil]) in the case of GAD. Although OCD rarely starts as a primary disorder in later life, older persons with chronic OCD may present for treatment. Clomipramine (Anafranil), a serotonergic tertiary amine TCA, can be a useful treatment for OCD, especially in cases that have failed to adequately improve with SSRIs. While the use of clomipramine has potential limitations in older adults because of its side-effect profile, it is worth considering, in combination with cognitive behavior therapy, in selected older persons who have OCD and that has failed to respond to other medications. Similarly, first-generation MAOIs such as tranylcypromine (Parnate) or phenelzine (Nardil) should be kept in mind as a potential treatment for refractory panic disorder, social phobia, or PTSD, or for refractory anxious depression. However, the need to avoid tyramine-containing foods and certain medications to prevent the development of a hypertensive crisis, and the potential for these drugs to cause significant supine and orthostatic hypotension (as well as weight gain associated with phenelzine), will limit the use of MAOIs to carefully selected individuals.

Vortioxetine (Brintellix) is a serotonergic antidepressant that differs from SSRIs in neurochemical activity in that it modulates a variety of serotonin receptors in addition to inhibiting the reuptake of serotonin. A recent meta-analysis of four RCTs found no difference between vortioxetine and placebo in rates of response and remission of GAD in mixed-age adults. To our knowledge, vortioxetine has not been studied in RCTs as a treatment for other anxiety disorders.

BENZODIAZEPINES

Despite their widespread use, benzodiazepines have a limited role in the treatment of late-life anxiety. First, they are not a useful treatment for two of the three most common anxiety disorders in later life, that is, specific phobias and agoraphobia without panic. Second, the frequent comorbidity of GAD with depression in later life suggests the need for medication with both antidepressant and anxiolytic effects. Benzodiazepines are not an effective treatment for the core symptoms of depression. On the other hand, antidepressant medication is an effective treatment for GAD even when depression is absent. Third, there are a number of risks associated with benzodiazepines in older adults, including cognitive impairment, psychomotor impairment, excessive daytime sedation, instability of gait, falls, and hip fractures. Finally, anxiety disorders tend to be chronic conditions, with most patients requiring long-term treatment. Long-term use of benzodiazepines can be associated with psychological and physical dependence and tolerance, so it can be difficult to discontinue the drug if its benefits diminish and/or adverse effects develop. Persons with a history of alcoholism or drug abuse may be at particular risk of becoming dependent on a benzodiazepine.

In situations where benzodiazepines are indicated (for example, acute situational anxiety or as a short-term adjunct to an antidepressant until the antidepressant becomes effective), lorazepam (Ativan) or oxazepam (Serax) are preferred in older adults because they have relatively short elimination half-lives (10 to 15 hours), no active metabolites, and they do not undergo oxidative metabolism in the liver (thus, their clearance is not affected by age). As a result, these drugs do not have the potential for cumulative toxicity in older adults, which is a risk associated with longer-acting compounds such as diazepam or clonazepam. Oxazepam is slowly absorbed, so it is not very effective as a hypnotic unless it is taken at least an hour before bedtime. The usual dosage range in older adults is lorazepam 0.5 to 1.5 mg/day and oxazepam 10 to 30 mg/day. Although alprazolam (Xanax) is FDA-approved for the treatment of GAD and panic disorder, there are no data pertaining to its risk-benefit profile in older persons with these conditions. Alprazolam has a high potential for dependence and abuse. Its relatively short elimination half-life can result in rebound anxiety, which can lead to patients taking more frequent doses and higher daily doses than prescribed. For these reasons, alprazolam is not recommended for older adults.

As noted, most older adults who are prescribed a benzodiazepine do not require the drug on an ongoing basis. However, there are occasions when long-term benzodiazepine treatment may be necessary. In particular, some patients who have been maintained on a benzodiazepine for years experience a significant worsening of anxiety when an attempt is made to withdraw the medication. In such cases, if the benzodiazepine is not causing significant side effects, it is usually the least disruptive to leave the individual on the medication and monitor them on a regular basis. If side effects do become an issue as the person grows older, dose reduction rather than complete discontinuation of the drug may be the best compromise. In addition, some patients do not experience adequate relief of anxiety symptoms with a nonbenzodiazepine medication alone, and in these cases, longer-term treatment with an adjunctive benzodiazepine, given at the lowest possible dose, may be necessary.

BUSPIRONE

Buspirone (Buspar), an azapirone derivative and a serotonin-1A receptor partial agonist, is a nonbenzodiazepine anxiolytic that is more efficacious than a placebo in the treatment of GAD. It is not indicated in the treatment of other anxiety disorders. Although buspirone and benzodiazepines have comparable efficacy in patients with GAD who persist with treatment, several studies have found that patients are more likely to discontinue buspirone than a benzodiazepine for reasons other than side effects. It has been suggested that the higher rate of discontinuation of buspirone compared with a benzodiazepine may be due to buspirone’s delayed onset of action.

Although buspirone is an efficacious treatment for GAD, it has a limited role in the treatment of generalized anxiety in late life. First, it is not a substitute for antidepressant medication in the treatment of generalized anxiety that is comorbid with, or secondary to, major depression. Second, because buspirone has a delayed onset of action, it is not suitable as an acute adjunctive treatment for generalized anxiety symptoms associated with major depression. Third, there is evidence that buspirone may have diminished efficacy in persons who have previously been treated with a benzodiazepine. Therefore, its usefulness in the management of older adult patients with chronic GAD, many of whom have been treated with benzodiazepines, is unclear. Finally, the twice-a-day to three-times-a-day dosing schedule of buspirone increases the risk of poor adherence to this medication.

Unlike benzodiazepines, buspirone does not cause sedation, psychomotor impairment, cognitive impairment, or depressed respiratory drive. This side-effect profile suggests that it would be preferable to a benzodiazepine in the ongoing management of generalized anxiety in nondepressed patients who have chronic obstructive airway disease, sleep apnea, or a neurologic disorder. However, select antidepressants may also be effective for anxiety and well tolerated in these patients.

The usual starting dose of buspirone in older patients is 5 mg twice a day, followed by dose increments of 5 mg every 4 to 5 days to reach a target dose of 20 to 30 mg/day, given in divided doses. If a person fails to show significant improvement after 4 to 6 weeks at this target dose, the dose can be further gradually increased to a maximum of 45 mg/day, as tolerated. The main side effects of buspirone are dizziness, headache, and nausea. Gradual dose titration is essential in order to minimize the risk of these effects.

Sometimes a physician elects to switch a patient from long-term benzodiazepine therapy to buspirone. Because buspirone has a delayed onset of action and does not suppress benzodiazepine withdrawal symptoms, a patient should not be abruptly switched from a benzodiazepine to buspirone. Instead, the patient should take a therapeutic dose of buspirone for 2 to 4 weeks before the benzodiazepine is gradually tapered. Despite this approach, some patients are less satisfied with the anxiolytic effect of buspirone compared with that of a benzodiazepine.

OTHER DRUGS

Pregabalin

Pregabalin (Lyrica) is an anticonvulsant drug that binds to the alpha-2-delta subunit of the voltage-dependent calcium channel in the central nervous system and decreases the release of neurotransmitters, including glutamate, norepinephrine, and substance P. Unlike benzodiazepines, pregabalin does not bind to gamma-aminobutyric acid (GABA) receptors and does not alter GABA metabolism. Several industry-funded RCTs in mixed-age patients have found that pregabalin is more efficacious than a placebo in the treatment of GAD, although pregabalin does not have FDA approval for this indication. Although a statistically significant difference between 150 to 600 mg/day of pregabalin and placebo was found in patients aged 65 years or older with GAD, the clinical significance of the difference is uncertain: there was only a 2-point greater reduction in HAM-A scores with drug versus placebo (12.9 vs.

10.7). Moreover, pregabalin and placebo groups did not differ in response or remission rates at the end of the 8-week study. Pregabalin was well tolerated in this group of older patients.

Quetiapine XR

First-generation antipsychotic medications have long been used in the acute treatment of agitation, but they are not recommended as a treatment for primary anxiety disorders because of the risk of tardive dyskinesia associated with their longer-term use. Quetiapine XR (Seroquel XR) is a second-generation atypical antipsychotic with a low risk of neurologic side effects. An industry-funded study found that 50 to 300 mg/day of quetiapine XR was more efficacious than placebo in treating GAD in older adults, with significant differences in response and remission rates by the end of 8 weeks of treatment. While quetiapine has a low risk of parkinsonism and tardive dyskinesia, and it can be associated with weight gain and metabolic dysfunction and, as a result, cannot be recommended as a routine first-line treatment of GAD in older adults.

Beta Blockers and Antihistamines

Beta blockers (e.g., 10 to 20 mg of propranolol [Inderal]) can be effective for performance anxiety when given about one hour prior to the feared performance. Therefore, they can be useful in patients with specific types of social anxiety disorder, such as public speaking or eating in public. However, beta-blockers should be avoided in anxious patients with sinus bradycardia, hypotension, or asthma.

RCTs conducted in Europe have found that the antihistamine drug hydroxyzine (Atarax) is more efficacious than a placebo and comparable in efficacy to a benzodiazepine and buspirone in the treatment of GAD in mixed-age adults. Hydroxyzine has not been studied as a treatment of GAD in late life, but its sedative and anticholinergic effects (increasing the risk of daytime somnolence, cognitive impairment, and falls and fractures) limit its usefulness in older patients.

FURTHER READINGS

Alaka KJ, Noble W, Montejo A, et al. Efficacy and safety of duloxetine in the treatment of older adult patients with generalized anxiety disorder: a randomized, double-blind, placebo-controlled trial. *Int J Geriatr Psychiatry*. 2014;29:978–986.

Bandelow B, Reitt M, Röver C, Michaelis S, Görlich Y, Wedekind D. Efficacy of treatments for anxiety disorders: a meta-analysis. *Int Clin Psychopharmacol*. 2015;30:183–192.

Flint AJ. Generalized anxiety disorder in elderly patients: epidemiology, diagnosis, and treatment options. *Drugs Aging*. 2005;22:101–114.

Gerlach LB, Wiechers IR, Maust DT. Prescription benzodiazepine use among older adults: a critical review. *Harv Rev Psychiatry*. 2018;26:264–273.

Gosmann NP, Costa MA, Jaeger MB, et al. Selective serotonin reuptake inhibitors, and serotonin and norepinephrine reuptake inhibitors for anxiety, obsessive-compulsive, and stress disorders: a 3-level network meta-analysis. *PLoS Med*. 2021;18:e1003664.

Hovenkamp-Hermelink JHM, Jeronimus BF, Myroniuk S, Schoevers RA. Predictors of persistence of anxiety disorders across the lifespan: a systematic review. *Lancet Psychiatry*. 2021;8:428–443.

Iaboni A, Bronskill SE, Reynolds KB, et al. Changing pattern of sedative use in older adults: a population-based cohort study. *Drugs Aging*. 2016;33:523–533.

Katz IR, Reynolds CF, Alexopoulos GS, Hackett D. Venlaxafine ER as a treatment for generalized anxiety disorder in older adults: pooled analysis of five randomized placebo-controlled clinical trials. *J Am Geriatr Soc*. 2002;50:18–25.

Kimura M, Tateno A, Robinson RG. Treatment of poststroke generalized anxiety disorder comorbid with poststroke depression: merged analysis of nortriptyline trials. *Am J Geriatr Psychiatry*. 2003;11:320–327.

Lenze EJ, Rollman BL, Shear MK, et al. Escitalopram for older adults with generalized anxiety disorder: a randomized controlled trial. *JAMA*. 2009;301:295–303.

Maust DT, Chen SH, Benson A, et al. Older adults recently started on psychotropic medication: where are the symptoms? *Int J Geriatr Psychiatry*. 2015;30:580–586.

Mezhebovsky I, Mägi K, She F, Datto C, Eriksson H. Double-blind, randomized study of extended release quetiapine fumarate (quetiapine XR) monotherapy in older patients with generalized anxiety disorder. *Int J Geriatr Psychiatry*. 2013;28:615–625.

Paquin AM, Zimmerman K, Rudolph JL. Risk versus risk: a review of benzodiazepine reduction in older adults. *Expert Opin Drug Saf*. 2014;13:919–934.

Pinquart M, Duberstein PR. Treatment of anxiety disorders in older adults: a meta-analytic comparison of behavioral and pharmacological interventions. *Am J Geriatr Psychiatry*. 2007;15:639–651.

Qin B, Huang G, Yang Q, et al. Vortioxetine treatment for generalised anxiety disorder: a meta-analysis of anxiety, quality of life and safety outcomes. *BMJ Open*. 2019;9:e033161.

Ramos K, Stanley MA. Anxiety disorders in late life. *Clin Geriatr Med*. 2020;36:237–246.

Tannenbaum C, Martin P, Tamblyn R, Benedetti A, Ahmed S. Reduction of inappropriate benzodiazepine prescriptions among older adults through direct patient education: the EMPOWER cluster randomized trial. *JAMA Intern Med*. 2014;174:890–898.

Wetherell JL, Petkus AJ, White KS, et al. Antidepressant medication augmented with cognitive-behavioral therapy for generalized anxiety disorder in older adults. *Am J Psychiatry*. 2013;170:782–789.

Witlox M, Garnefski N, Kraaij V, et al. Prevalence of anxiety disorders and subthreshold anxiety throughout later life: systematic review and meta-analysis. *Psychol Aging*. 2021;36:268–287.

55.4d Antipsychotic Drugs

KELLY C. LEE, PHARM.D., M.A.S., B.C.P.P., AND VIVIAN HOOK, PH.D.

Antipsychotic medications constitute the treatment of choice for psychoses associated with schizophrenia or schizoaffective disorder. Antipsychotics are frequently prescribed for older adults in the treatment of psychosis and severe behavioral disturbances such as agitation and aggression. Although nonpharmacologic therapies may be helpful in milder states of behavioral disturbances associated with cognitive disorders such as dementia, antipsychotics are often necessary to treat severe symptoms. Age-related physiologic changes in older adults can increase sensitivity to adverse events from these medications. Treatment is also often complicated by the frequent presence of comorbid medical conditions and medication interactions in older adults. Despite the complexities, antipsychotic medications are commonly prescribed to older patients in inpatient, outpatient, residential, and nursing home settings.

Antipsychotics are broadly classified into two groups. The older agents (i.e., introduced in the United States before 1990) are referred to as first generation and primarily block dopamine receptors (primarily D2) to exert antipsychotic effects. First generation antipsychotics (FGAs), while effective for positive symptoms of schizophrenia, are associated with undesirable extrapyramidal symptoms and endocrine changes.

Second generation antipsychotics (SGA) have an absence or decreased incidence of extrapyramidal effects and tardive dyskinesia, as well as an expanded spectrum of activity against negative symptoms associated with schizophrenia. Because of their improved neurologic profile, SGAs have generally replaced FGAs. Unfortunately, due to studies that ultimately led to black box warning labeling for antipsychotics when used in older adult patients with dementia, the benefits of using these agents must be carefully weighed against the potential risks.

PHARMACOLOGIC ACTIONS

Pharmacokinetics and Pharmacodynamics

Pharmacokinetics can generally be described as the body's effect on the absorption, distribution, metabolism, and elimination of drugs. Numerous age-related changes in physiology may affect pharmacokinetics. Changes in pharmacodynamics (the drug's effect on the body) are also likely to occur; however, these mechanisms are not as well understood relative to the knowledge of age-related pharmacokinetic changes. Neurotransmitter and receptor function changes in the central nervous system may increase sensitivity to drugs, and toxicity may occur at lower plasma levels. In general, these pharmacokinetic and pharmacodynamic changes may lead to higher blood levels of medications, prolonged pharmacologic effects, and a greater risk for side effects in older adults.

Receptor Activity

The specific mechanism of action of antipsychotics has not been elucidated, but most of the research attention has focused on the receptor blockade of dopamine and serotonin. While the FGAs have been shown to work through the blockade of the central dopamine receptors (specifically D2), SGAs have been referred to as serotonin–

dopamine antagonists as they block both 5-HT$_{2A}$ and D2 receptors. Antipsychotic drugs also affect other receptors, usually resulting in adverse effects rather than an enhanced therapeutic response. For example, histamine-1 (H$_1$) blockade is responsible for sedation and possibly weight gain. Blockade of the 5-HT$_{2C}$ receptor may also play a role in weight gain associated with some SGAs. Blockade of the α$_1$ receptors causes orthostatic hypotension, whereas α$_2$ blockade plays a role in sexual dysfunction. Muscarinic receptor blockade results in classical anticholinergic effects such as dry mouth, blurred vision, constipation, and urinary retention. An understanding of the receptor-binding properties of the FGAs and SGAs can predict their adverse effect profiles (Table 55.4d–1).

Table 55.4d–1.
Relative Adverse Effects for Selected Antipsychotics

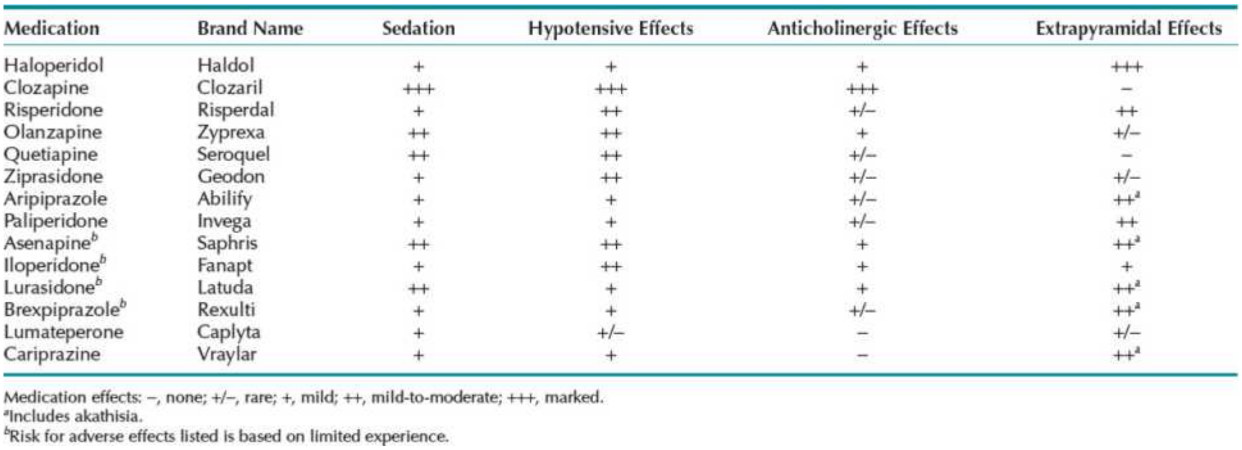

Medication	Brand Name	Sedation	Hypotensive Effects	Anticholinergic Effects	Extrapyramidal Effects
Haloperidol	Haldol	+	+	+	+++
Clozapine	Clozaril	+++	+++	+++	–
Risperidone	Risperdal	+	++	+/–	++
Olanzapine	Zyprexa	++	++	+	+/–
Quetiapine	Seroquel	++	++	+/–	–
Ziprasidone	Geodon	+	++	+/–	+/–
Aripiprazole	Abilify	+	+	+/–	++[a]
Paliperidone	Invega	+	+	+/–	++
Asenapine[b]	Saphris	++	++	+	++[a]
Iloperidone[b]	Fanapt	+	++	+	+
Lurasidone[b]	Latuda	++	+	+	++[a]
Brexpiprazole[b]	Rexulti	+	+	+/–	++[a]
Lumateperone	Caplyta	+	+/–	–	+/–
Cariprazine	Vraylar	+	+	–	++[a]

Medication effects: –, none; +/–, rare; +, mild; ++, mild-to-moderate; +++, marked.

[a] Includes akathisia.

[b] Risk for adverse effects listed is based on limited experience.

THERAPEUTIC INDICATIONS FOR ANTIPSYCHOTICS

Early-Onset Schizophrenia

Historically, it has been thought that all patients with schizophrenia uniformly follow a poor prognostic course with no return to a premorbid level of functioning. Fortunately, long-term studies in chronic schizophrenia have been more optimistic. Approximately one-third of patients either undergo remission or have only mild symptoms over the long term, while 15% to 20% of patients get worse. The remaining patients appear to have a stable course. Patients who have benefited from maintenance antipsychotic treatment with prolonged absence of psychotic symptoms may warrant a trial of medication. Older patients typically have fewer and less severe positive symptoms than younger patients, while negative symptoms tend to persist into late life. Relapse and exacerbation may occur in older patients with chronic schizophrenia just as they do in younger patients, and investigators have struggled to identify factors that may predict an outcome. Hence, if a drug withdrawal is attempted, careful monitoring is required.

Late-Onset Schizophrenia

Approximately 20% to 30% of patients have onset of symptoms after 40 years of age and are considered to have late-onset schizophrenia (LOS). In general, regardless of the age of onset of schizophrenia (LOS vs. early-onset schizophrenia [EOS]), patients with schizophrenia are similar in terms of positive symptoms, such as delusions and hallucinations, chronicity of the course of illness, and family history. However, LOS seems to be more common in women, and those with LOS have better premorbid functioning with successful occupations and marital histories, fewer negative symptoms, and less severe neurocognitive impairment. Additionally, patients with LOS typically require lower antipsychotic doses and have fewer positive symptoms compared with EOS patients. Onset after age 65 usually signifies very LOS-like psychosis, which is typically secondary to general medical conditions such as dementia or neurodegenerative disorder. Additional studies are needed to identify genetic variants, biomarkers, and changes in neuroimaging studies that may differentiate LOS from EOS.

Delusional Disorder

Delusional disorder refers to the presence of one or more persistent, nonbizarre delusions and may be misdiagnosed as schizophrenia. Delusions commonly involve situations that may occur in real life, such as being poisoned, having a disease, or experiencing spousal infidelity. Delusional disorder generally presents in middle or late life and can be differentiated from schizophrenia by the absence of prominent auditory or visual hallucinations and by the lack of functional deterioration outside the area of delusions. Psychiatric comorbidities, especially depressive disorders, are common among those with delusional disorders. Mild symptoms may be treated without medications, but antipsychotics may be helpful for the agitated patient whose delusions have a negative impact on social relationships or functional behaviors. Antipsychotic nonadherence is a common problem in those with delusional disorders, although some suggest that women may be more adherent to pharmacologic treatments than men.

Dementia With Psychosis or Severe Behavioral Disturbances

The most common types of dementia are Alzheimer and vascular types. Patients who develop dementias have cognitive impairment resulting in memory and executive functioning decline but also noncognitive dysfunction that can be more distressful. Many older adult patients with dementia develop behavioral disturbances such as aggression, agitation, delusions, and other neuropsychiatric symptoms during their course of illness. These changes in mood, affect, and behavior can lead to early institutionalization, increased caregiver burden, and increased healthcare costs. When behavioral problems arise, early recognition and prompt treatment can improve the quality of life of patients, their families, and caregivers and reduce healthcare costs. In addition to medications used for cognitive dysfunction, nonpharmacologic interventions and antipsychotics have been widely used for treating behavioral disturbances associated with dementia. Treatment guidelines have generally recommended careful assessment, diagnosis, and nonpharmacologic treatment approaches, such as psychosocial and environmental interventions, prior to pharmacologic approaches. When nonpharmacologic approaches fail, the provider, together with the patient and their family, should examine pharmacologic options carefully when severe agitation and psychosis symptoms are dangerous and/or cause significant distress to the patient. The family should be informed about the relative risks and benefits of antipsychotics and other treatments, as well as the risks of not using medication, so that they can participate effectively in clinical decision-making. Documentation of informed consent should be obtained.

The use of antipsychotics in those with dementia is controversial due to the FDA black box warning regarding the increased incidence of cerebrovascular events and death associated with antipsychotic use. While the effectiveness of antipsychotics for noncognitive symptoms in dementia may be marginal, a short-term trial is generally warranted to control symptoms of mania, psychosis, hallucinations, and aggressive behaviors that could endanger others. While there is no significant difference in effectiveness between FGAs and SGAs, SGAs are generally preferred due to better side effect profiles, particularly in terms of extrapyramidal side effects and tardive dyskinesia. SGAs with evidence for managing behavioral symptoms of dementia include aripiprazole [Abilify], olanzapine [Zyprexa], quetiapine [Seroquel], and risperidone [Risperdal]. Antipsychotics should be used in the lowest effective doses and carefully monitored for adverse events.

The decision to discontinue antipsychotics in patients who have responded to the drug must be carefully weighed against the possibility of worsening symptoms. If the patient experiences significant adverse effects or there is no significant response after 4 weeks of treatment, consideration should be made to taper or withdraw the medication. A Cochrane Database review concluded that chronic antipsychotics may be withdrawn from patients without risking worsening of behavioral symptoms, especially in those with mild neuropsychiatric symptoms. Due to the low number and quality of studies, it is unknown whether the discontinuation of antipsychotics affects cognitive function, quality of life, or mortality rates.

Parkinson Disease With Psychosis

Psychotic symptoms can occur in up to 40% of patients with Parkinson disease. The symptoms may be present continuously or intermittently over the course of a month and predominantly consist of hallucinations. Most hallucinations are visual, and of these, half of the patients also have auditory hallucinations. Hallucinations also tend to be different from schizophrenia and generally have less “emotional content.” Delusions can occur in 5% to 10% of patients treated with antiparkinsonian drugs, and patients are typically paranoid in nature, such as in spousal infidelity. Psychosis is stressful for patients and can potentially cause more distress to caregivers. Psychosis is a poor prognostic indicator for institutionalization, dementia, and mortality. Managing psychosis is one of the most difficult challenges in the care of patients with Parkinson disease. Dopaminergic drugs that can improve the motor disability caused by Parkinson disease are associated with the emergence of psychotic symptoms and appear to be dose-related. Conversely, the use of antipsychotic drugs to reduce psychosis can exacerbate motor symptoms of parkinsonism. Additionally, anticholinergic medications can cause delirium with agitated behavior or psychosis. Management of psychosis in Parkinson disease begins with withdrawal or lowering of the dose of the dopaminergic agent without disrupting motor signs. Due to the complexity of balancing the dopaminergic requirement to control both symptoms, psychosis presents a serious threat to the patient’s ability to maintain independence. If psychosis persists, antipsychotic medications with the lowest liability for exacerbating parkinsonism may be helpful (clozapine [Clozaril] and quetiapine). Pimavanserin [Nuplazid] is an antipsychotic with inverse agonist and antagonist activity at 5-HT2A and 5-HT2C receptors. It is approved for use in the treatment of hallucinations and delusions associated with Parkinson disease psychosis. The most common adverse reactions include gastrointestinal disorders, peripheral edema, a confusional state, and hallucination.

Delirium

Delirium is a cognitive disorder that is characterized by fluctuating levels of consciousness, inattention, memory impairment, and perceptual disturbances. Delirium may be accompanied by illusions, visual hallucinations, and delusions. Delirium can arise from the acute onset of a new medical illness, an exacerbation of a pre-existing chronic illness, a drug reaction or interaction, fever, or trauma. A careful medical, neurologic, and laboratory evaluation to identify the conditions that are contributing to the mental status change is required. Drugs with anticholinergic properties are common causes of delirium in older adults. Recognition and treatment of delirium are important to reduce the risks of morbidity and mortality in these patients. Nonpharmacologic treatment, such as environmental changes, behavioral support, and redirection, are key strategies for treating delirium. Antipsychotics should be prescribed only when necessary to maintain safety and basic care. The lowest effective dosage should be used for a minimum duration for high-risk symptoms that pose safety concerns to the patient or staff. Antipsychotics should also be discontinued or have a plan for tapering once delirium symptoms have resolved.

Depression

Patients with delirium and dementia are at high risk of depression, and depression may be a predictor of cognitive decline. Suicide risk assessment is also important since suicidal ideation can be seen in up to 30% of older adults with major depression. Psychotic symptoms may also occur in patients suffering from depression. In contrast to nonpsychotic depression, older adult patients with psychotic depression are at increased risk for relapse, suicide, and hospitalization. Major depressive disorder with psychotic features responds better to treatment with a combination of an antipsychotic and an antidepressant than to treatment with either agent alone. Electroconvulsive therapy is highly effective in major depressive disorder with psychotic features and, at times, may be considered a first-line treatment, even in older adults. Aripiprazole, brexpiprazole (Rexulti), quetiapine, and olanzapine in combination with fluoxetine (Prozac, Symbyax for combination product), are also indicated for adjunctive treatment for nonpsychotic depression; however, the benefit of using multiple medications that may present adherence challenges and drug–drug interactions in older adult patients must be clearly warranted.

Bipolar Disorder

Bipolar disorder is not uncommon in older adults and often leads to functional impairment and substantial utilization of health care resources. While conventional mood stabilizers such as lithium and valproate are still widely used for bipolar disorder, SGAs have also emerged as first-line options for the control of manic symptoms. Currently, aripiprazole, asenapine (Saphris), cariprazine (Vraylar), olanzapine, quetiapine, risperidone, and ziprasidone (Geodon) are approved for mania. Aripiprazole, quetiapine, and ziprasidone are also indicated for bipolar maintenance treatment, while quetiapine, olanzapine/fluoxetine, and lurasidone (Latuda) are also indicated for bipolar depression. SGAs may offer advantages over drugs like lithium due to their wider therapeutic range and fewer drug–drug interactions. However, metabolic and cardiovascular considerations should still be considered in older adult patients. Unfortunately, comparative effectiveness or safety data for antipsychotics in older adults with bipolar disorder are lacking.

TREATMENT PRINCIPLES

Pretreatment Evaluation

Older adult persons who present with new-onset psychosis, an exacerbation of psychosis, or severe behavioral disturbance should have a thorough medical examination (including neurologic assessment) prior to initiating a new or modifying a pre-existing antipsychotic regimen. At a minimum, the evaluation should include a comprehensive history, physical examination, and laboratory screening, including a toxicology screen for substances of abuse. This may identify medical or substance-induced problems that may be contributing to the behavioral disturbances (e.g., infection or pain) or that may represent a special concern during treatment (e.g., recent myocardial infarction).

Antipsychotic Selection

The choice of an antipsychotic is usually based on a number of factors, and selection should be highly individualized. Factors include specific symptoms, experience with previous treatment (efficacy and side effects), simplicity of regimen (available dosage forms, avoidance of drug interactions, ease of obtaining a therapeutic dose), and cost considerations. With few exceptions, there is limited evidence for the superior efficacy of one antipsychotic over another, and clinicians should base their decision on the above patient-specific factors.

Dosage Forms

Although all antipsychotics are available in oral formulations, only a few are available as injectables (e.g., olanzapine, aripiprazole, ziprasidone, risperidone, and paliperidone [Invega]). Short-acting intramuscular preparations of an antipsychotic may be preferred over oral medications if a patient is agitated and not likely to be cooperative. Intramuscular preparations bypass the gastrointestinal tract and first-pass metabolism and have a faster onset of action. For example, most short-acting intramuscular antipsychotics reach a maximum plasma level within 30 to 60 minutes, and clinical effects are seen within 15 minutes. While short-acting intramuscular preparations offer administration advantages, care should be taken to avoid repeating doses without first assessing for a response or adverse effects. Alternatively, rapidly dissolving tablet formulations (e.g., clozapine, risperidone, olanzapine, aripiprazole) and sublingual tablets (e.g., asenapine) offer the advantage of improved adherence in some patients who have difficulty with swallowing, resistant to taking medications, or prone to behaviors such as “cheeking” medications. Except for people with schizophrenia or schizoaffective disorder with a history of poor medication-taking behaviors, long-acting depot medications are not usually prescribed for older adults with psychosis.

Table 55.4d–2.
Recommended Dosing Guidelines for Antipsychotics

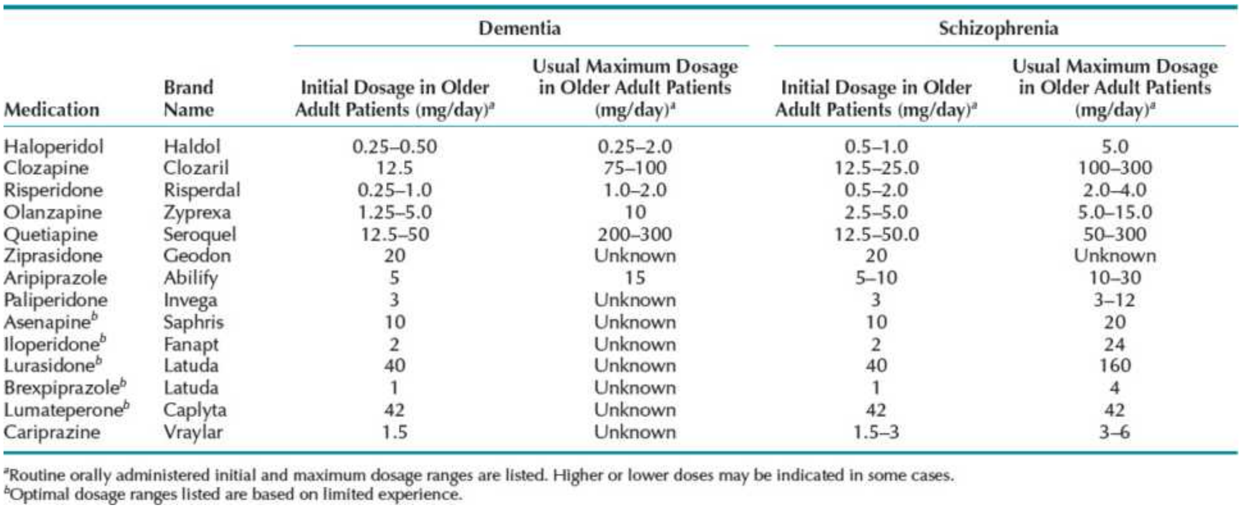

		Dementia		Schizophrenia	
Medication	Brand Name	Initial Dosage in Older Adult Patients (mg/day)[a]	Usual Maximum Dosage in Older Adult Patients (mg/day)[a]	Initial Dosage in Older Adult Patients (mg/day)[a]	Usual Maximum Dosage in Older Adult Patients (mg/day)[a]
Haloperidol	Haldol	0.25–0.50	0.25–2.0	0.5–1.0	5.0
Clozapine	Clozaril	12.5	75–100	12.5–25.0	100–300
Risperidone	Risperdal	0.25–1.0	1.0–2.0	0.5–2.0	2.0–4.0
Olanzapine	Zyprexa	1.25–5.0	10	2.5–5.0	5.0–15.0
Quetiapine	Seroquel	12.5–50	200–300	12.5–50.0	50–300
Ziprasidone	Geodon	20	Unknown	20	Unknown
Aripiprazole	Abilify	5	15	5–10	10–30
Paliperidone	Invega	3	Unknown	3	3–12
Asenapine[b]	Saphris	10	Unknown	10	20
Iloperidone[b]	Fanapt	2	Unknown	2	24
Lurasidone[b]	Latuda	40	Unknown	40	160
Brexpiprazole[b]	Latuda	1	Unknown	1	4
Lumateperone[b]	Caplyta	42	Unknown	42	42
Cariprazine	Vraylar	1.5	Unknown	1.5–3	3–6

[a] Routine orally administered initial and maximum dosage ranges are listed. Higher or lower doses may be indicated in some cases.
[b] Optimal dosage ranges listed are based on limited experience.

General Dosing Recommendations

There is considerable heterogeneity in age-related pharmacokinetic and pharmacodynamic alterations. Thus, the efficacy and adverse effects of a particular antipsychotic dosage may vary considerably. In addition, the time required for antipsychotic response is typically longer than in younger patients. The initial dosage should be relatively low, and dose increases should be titrated cautiously but up to an effective level. Antipsychotic medication dosage in an older patient with a primary psychotic disorder (e.g., schizophrenia) is usually one-half the usual dosage in young patients, although some older patients will respond best to the higher dosages typically prescribed for younger patients. Optimal medication dosage for older patients with a primarily cognitive or neurologic disorder (e.g., dementia) and psychosis is usually 25% to 50% of doses typically prescribed to younger patients (Table 55.4d–2).

Adherence Issues

Older patients commonly take several medications simultaneously and therefore have an increased risk of adverse drug reactions. A complete drug history should be obtained, and concomitant medications simplified, if possible, prior to implementing antipsychotic treatment. Medication adherence is a problem in all age groups, but older adults may be especially susceptible to adherence challenges. Physical problems, such as impaired vision, may increase the likelihood that instructions will be misread or that one medicine will be mistaken for another. Patients with cognitive impairment may forget to take their medication, and drug regimens with complex administration schedules should be avoided. Drug administration issues also present challenges for older adult patients, such as special handling requirements for asenapine (handle with dry hands, no food or liquid for 10 minutes after dose) and dietary requirements for ziprasidone and lurasidone (350 to 500 kcal required with dose). Psychological factors, such as fear of side effects or becoming dependent on medications, may also affect adherence.

Table 55.4d–3.
Dosage Formulations and Drug Interaction Potential for Selected Antipsychotics

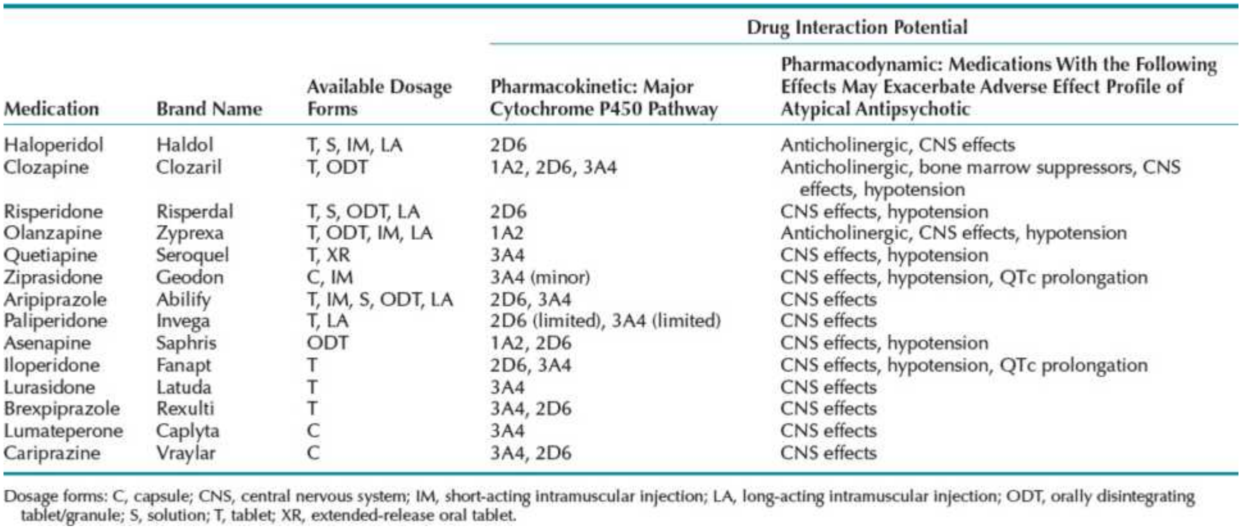

			Drug Interaction Potential	
Medication	Brand Name	Available Dosage Forms	Pharmacokinetic: Major Cytochrome P450 Pathway	Pharmacodynamic: Medications With the Following Effects May Exacerbate Adverse Effect Profile of Atypical Antipsychotic
Haloperidol	Haldol	T, S, IM, LA	2D6	Anticholinergic, CNS effects
Clozapine	Clozaril	T, ODT	1A2, 2D6, 3A4	Anticholinergic, bone marrow suppressors, CNS effects, hypotension
Risperidone	Risperdal	T, S, ODT, LA	2D6	CNS effects, hypotension
Olanzapine	Zyprexa	T, ODT, IM, LA	1A2	Anticholinergic, CNS effects, hypotension
Quetiapine	Seroquel	T, XR	3A4	CNS effects, hypotension
Ziprasidone	Geodon	C, IM	3A4 (minor)	CNS effects, hypotension, QTc prolongation
Aripiprazole	Abilify	T, IM, S, ODT, LA	2D6, 3A4	CNS effects
Paliperidone	Invega	T, LA	2D6 (limited), 3A4 (limited)	CNS effects
Asenapine	Saphris	ODT	1A2, 2D6	CNS effects, hypotension
Iloperidone	Fanapt	T	2D6, 3A4	CNS effects, hypotension, QTc prolongation
Lurasidone	Latuda	T	3A4	CNS effects
Brexpiprazole	Rexulti	T	3A4, 2D6	CNS effects
Lumateperone	Caplyta	C	3A4	CNS effects
Cariprazine	Vraylar	C	3A4, 2D6	CNS effects

Dosage forms: C, capsule; CNS, central nervous system; IM, short-acting intramuscular injection; LA, long-acting intramuscular injection; ODT, orally disintegrating tablet/granule; S, solution; T, tablet; XR, extended-release oral tablet.

Drug Interactions

Both pharmacokinetic and pharmacodynamic drug interactions can occur with antipsychotic agents. Cytochrome P450 (CYP) enzymes, especially 1A2, 2D6, and 3A4, are responsible for the metabolism of many antipsychotics. In general, antipsychotics are primarily metabolized hepatically by CYP enzymes but are not major inducers or inhibitors of CYP-mediated metabolism. Thus, antipsychotics do not commonly affect the pharmacokinetics of other medications but are themselves affected by other medications. Induction or inhibition of these enzymes by other drugs may occasionally provide clinically important drug interactions. Drugs that have the potential to have serious interactions with antipsychotics include fluvoxamine (1A2 inhibitor); erythromycin, ketoconazole, and ritonavir (3A4 inhibitors); quinidine, risperidone, fluoxetine, and paroxetine (2D6 inhibitors); and cimetidine (multiple systems). Conversely, enzyme inducers such as carbamazepine (3A4) or cigarette smoking (1A2) can reduce plasma concentrations of antipsychotic drugs by increasing the metabolism of the antipsychotic agent and lead to less-than-effective plasma concentration. For example, carbamazepine has been shown to reduce plasma concentrations of haloperidol by 50%. Cigarette smoking (1A2) has been shown to increase the clearance rate of clozapine and olanzapine. A pharmacodynamic interaction occurs when medications with similar unwanted properties are concurrently prescribed, such as low-potency FGAs or clozapine, with other drugs possessing strong anticholinergic effects (e.g., diphenhydramine or hydroxyzine). Pharmacodynamic drug interactions of greatest concern involve additive orthostasis, anticholinergic effects, QTc prolongation, and sedation (Table 55.4d–3).

PRECAUTIONS AND ADVERSE REACTIONS

Mortality

The black box labeling of antipsychotics by the FDA was prompted by 15 placebo-controlled trials that showed a 1.6- to 1.7-fold increased risk of death among those with dementia taking atypical antipsychotics. Specific causes of these deaths were attributed to heart-related events (e.g., heart failure, sudden death) or infections (mostly pneumonia). In a study of older Veterans, the risk of 10-year all-cause mortality was highest among those with dementia and prescribed antipsychotics (78%) compared to those without dementia or exposed to antipsychotics (42%). In a meta-analysis of 17 studies, there was no difference among antipsychotics for the primary safety outcome of death. In another systematic review and meta-analysis of 352 placebo-controlled studies in schizophrenia, there was no difference between antipsychotic drugs and placebo in mortality by any cause.

Cerebrovascular Events

Cerebrovascular adverse events, such as strokes, transient ischemic attacks, and fatalities, have been observed in older adult patients with dementia-related psychosis prescribed with aripiprazole, olanzapine, quetiapine, or risperidone. In 17 placebo-controlled trials, a higher number of cerebrovascular adverse events were observed in patients treated with SGAs compared to patients treated with a placebo. Data analyses have suggested that increased age (older than 80 years), concurrent use of benzodiazepines, sedation, pulmonary infections (pneumonia), and medical risks (prior strokes, hypertension, coronary artery disease, and hyperlipidemia) may increase the risk of experiencing a cerebrovascular adverse event.

Whether the mortality and cerebrovascular adverse event risks have a similar underlying mechanism is yet unclear. In a study of over 30,000 Medicare and Patient Assistance Program enrollees, FGAs were associated with higher risks of stroke and death compared with SGAs over 180 days. Interestingly, stroke did not explain the difference in mortality risk between the two drug groups. Orthostatic hypotension seems a likely mediating variable in some patients, as well as falls and oversedation.

Weight Gain

Weight gain has been associated with FGAs and SGAs in young and older adults. Weight gain is most prominent within the first 4 to 12 weeks of treatment with antipsychotics, especially SGAs. The cause of antipsychotic-induced weight gain is multifactorial and may include a change in food preferences, increased food or fluid intake, carbohydrate craving, or a sedentary lifestyle. Blockade of the histamine and 5-HT_{2C} receptors are also contributory. Histamine H_1

blockade causes sedation that may produce inactivity. H_1 blockade may also increase weight by interfering with normal satiety signals from the gut, resulting in overeating. In the CATIE-AD study, older adult patients who were treated with olanzapine, quetiapine, or risperidone had an average weight gain of 1.0, 0.7, and 0.4 lb, respectively, compared with a monthly weight loss of 0.9 lb for those who received placebo. Weight gain was more prominent in those with a BMI of less than 25 at baseline, and this is consistently seen in other age groups. Clozapine has also been associated with significant weight gain. Weight gain associated with antipsychotic treatment should be taken seriously because it may contribute to or exacerbate medical conditions such as diabetes, coronary artery disease, and hypertension. If significant weight gain is observed, a switch to another antipsychotic with lower weight gain liability may be considered as long as treatment effectiveness is not compromised.

Metabolic Effects

Diabetes mellitus and other metabolic disorders have been shown to be more common in adult and older adult patients with schizophrenia than in the general population. FGAs and SGAs that are used to treat schizophrenia can also worsen glucose intolerance, diabetes, and hypertriglyceridemia in those with or without a history of diabetes mellitus. In rare incidences, these events may lead to diabetic ketoacidosis and death. These complications may or may not be associated with weight gain. Clozapine and olanzapine have been associated with a 1.3- to 1.4-fold higher prevalence of diabetes, followed by quetiapine and risperidone. Potential mechanisms for glucose dysfunction include H_1-receptor antagonism, 5-HT_{2C}-receptor antagonism, reduced functioning of pancreatic cells or pancreatic cell damage, and altered regulation of insulin resistance. In older patients with various psychiatric disorders (e.g., schizophrenia, mood disorders, PTSD, or dementia), treatment with antipsychotics (aripiprazole, olanzapine, quetiapine, and risperidone) for at least 6 months was associated with a high incidence of metabolic syndrome. Body mass index, blood pressure, fasting blood glucose, LDL and

HDL cholesterol, and triglycerides were the primary metabolic markers studied. Patients also experienced serious and nonserious adverse events at the 2-year follow-up (23.7% and 50.8%, respectively). In addition, there was no significant change in psychopathology associated with any of the SGAs in the study. Hypertriglyceridemia seems to be due to increased adiposity resulting in higher free fatty acid concentration, insulin resistance, and increased VLDL production. Clozapine and olanzapine, again, have been associated with higher risks of hypertriglyceridemia compared to other antipsychotics in non–older adult patients. In the CATIE-AD trial and other systematic reviews, the association of weight gain and BMI was demonstrated with SGAs, but there was no association with glucose, total cholesterol, and triglyceride levels. The impact of metabolic syndrome and other adverse events on the overall health of older adult patients must be considered in the decision to use these drugs, and careful and proactive monitoring of weight, glucose levels, and lipid panel is warranted.

Tardive Dyskinesia

Tardive dyskinesia is a syndrome characterized by abnormal involuntary movements, typically involving the orofacial region, limbs, and trunk. Essential features of tardive dyskinesia include involuntary movements of the tongue, face, and jaw. Most tardive dyskinesia cases are relatively mild. In some older patients, a form of tardive dyskinesia may develop that is severe enough to impair functioning. Severe oral dyskinesia may result in dental and denture problems that can progress to ulceration and infection of the mouth as well as muffled or unintelligible speech. Severe orofacial tardive dyskinesia can impair eating and swallowing, which in turn can produce significant health problems. Gait disturbances due to limb dyskinesia may leave patients vulnerable to falls and injuries. Increasing age is the most consistently observed risk factor for the development of tardive dyskinesia. The annual incidence rates of developing tardive dyskinesia are three- to fivefold greater in older patients compared to younger adults. In addition, older adult patients seem to have earlier tardive dyskinesia symptoms and have more persistent symptoms than younger patients. Other risk factors include higher mean daily and cumulative antipsychotic doses and the presence of extrapyramidal signs early in treatment. The risk of antipsychotic-induced tardive dyskinesia is suggested to be lower with SGAs compared to FGAs. This lower risk was confirmed in a 2-year study of risperidone and olanzapine in antipsychotic-naïve older adult patients. There is some indication that the risk of tardive dyskinesia may be higher in women, non-Whites, patients with affective psychiatric disorders, and patients taking concomitant anticholinergic agents; however, the heterogeneity of available studies makes it challenging to determine this association conclusively.

Anticholinergic Side Effects

Older adult patients are much more sensitive to the anticholinergic side effects of medications than younger adults. Anticholinergic medications may aggravate existing comorbidities or concomitant medications with anticholinergic properties in older adult patients. Peripheral anticholinergic symptoms may include dry mouth, constipation, urinary retention, tachycardia, blurred vision, and aggravation of narrow-angle glaucoma. Dry mouth may cause problems with dentures; constipation may lead to ileus and perforation; urinary retention may be further aggravated by prostatic hypertrophy, and blurred vision may further aggravate pre-existing visual problems.

The central muscarinic blockade in older adults also increases their risk for delirium and impaired cognition, attention, memory, and self-care capacity. The higher risk of falls and hip fractures associated with antipsychotics in older adults is due in part to the anticholinergic, sedative, and orthostatic properties of these medications. In some cases, anticholinergic toxicity may be mistaken for an exacerbation of psychosis, and the antipsychotic dose may be increased inappropriately. Severe anticholinergic symptoms due to high doses of antipsychotics may manifest as increased irritability, aggressiveness, and visual hallucinations.

Sedation

Sedation is a common side effect of antipsychotics that can increase the risk of confusion and falls in older adults. Antipsychotics with sedative properties are often prescribed to calm agitated patients or to treat sleep disturbances. In a comparative effectiveness review, sedation was more common in older adult patients taking olanzapine or quetiapine than risperidone, but this was not statistically significant. Sedation is likely to occur in older adults at much lower doses than those prescribed to younger patients, and older adult patients are at risk for developing daytime sedation from the residual effects of nighttime dosing of the antipsychotic. Older adult patients may appear confused or have other cognitive impairments due to oversedation and should be evaluated appropriately.

Hypotension

Because of decreased adrenergic tone and cardiac output, older adult patients may be more vulnerable to orthostatic effects and may have severe orthostasis at relatively low doses. Orthostatic episodes are most likely to appear early in treatment, and the patient should be monitored closely during this period. Patients (or their caregivers) should be instructed to rise slowly after being seated or supine and take extreme care if they awaken during the night. The lowest doses of antipsychotics should be used, and patients should be adequately hydrated throughout the day.

Cardiovascular Effects

In addition to orthostatic hypotension, antipsychotics may also cause tachycardia and electrocardiogram (ECG) changes. Tachycardia may occur as a result of the anticholinergic effects of antipsychotic medications on vagal inhibition or secondary to orthostatic hypotension. ECG changes such as prolongation of the QT and PR intervals, ST-segment depression, and T-wave flattening have been observed with antipsychotics. The most clinically important ECG change is a prolongation of the QTc interval (which is the QT interval adjusted for heart rate). QTc prolongation may lead to the development of ventricular tachyarrhythmias such as torsades de pointes and ventricular fibrillation, that may result in syncope, cardiac arrest, or sudden cardiac death. All antipsychotics have the potential to prolong the QTc interval but to varying degrees. Thioridazine, pimozide, IV haloperidol, ziprasidone, and iloperidone (Fanapt) appear to have the largest effect on QTc interval. Effects are usually dose-related, although that may not be the case for ziprasidone or iloperidone. Other risk factors for QTc prolongation and torsades de pointes include advanced age, female gender, and electrolyte imbalance (hypokalemia, hypomagnesemia). Patients with a history of syncope, cardiac disease, electrolyte imbalance, or a family history of early sudden cardiac death should have an ECG before antipsychotic treatment. While thioridazine and pimozide are no longer clinically used, IV haloperidol, ziprasidone, and iloperidone are still therapeutic options for patients with schizophrenia. These drugs should be avoided in patients with a known history of QTc prolongation, recent acute myocardial infarction, or uncompensated heart failure. Clinical and ECG monitoring of cardiac function is recommended during antipsychotic therapy, particularly in patients treated with antipsychotics with known higher risks of QTc prolongation or those who have symptoms that could indicate torsades de pointes (e.g., dizziness, palpitations, syncope). These agents should also not be prescribed with other drugs that prolong the QT interval.

Other Adverse Effects

Clozapine use is associated with a dose-related risk of seizures and is the only antipsychotic agent associated with the rare development of myocarditis. In rare cases, orthostatic hypotension can be accompanied by cardiac collapse and/or cardiac arrest. Due to clozapine's significant risk of agranulocytosis, mandatory monitoring of white blood cell count and absolute neutrophil count is required. Age and certain ethnicities appear to be risk factors for the development of agranulocytosis. Clozapine, thus, is used specifically in patients for whom multiple therapies are inadequate or not tolerated or who have severe tardive dyskinesia.

FURTHER READINGS

Atti AR, Ferrari Gozzi B, Zuliani G, et al. A systematic review of metabolic side effects related to the use of antipsychotic drugs in dementia. *Int Psychogeriatr*. 2014;26(1):19–37.

Ballard C, Orrell M, YongZhong S, et al. Impact of antipsychotic review and nonpharmacological intervention on antipsychotic use, neuropsychiatric symptoms, and mortality in people with dementia living in nursing homes: a factorial cluster-randomized controlled trial by the Well-Being and Health for People With Dementia (WHELD) program. *Am J Psychiatry*. 2016;173(3):252–262.

Bingham KS, Whyte EM, Meyers BS, et al; STOP-PD Study Group. Relationship between cerebrovascular risk, cognition, and treatment outcome in late-life psychotic depression. *Am J Geriatr Psychiatry*. 2015;23(12):1270–1275.

Declercq T, Petrovic M, Azermai M, et al. Withdrawal versus continuation of chronic antipsychotic drugs for behavioural and psychological symptoms in older people with dementia. *Cochrane Database Syst Rev*. 2013;(3):CD007726.

Downing LJ, Caprio TV, Lyness JM. Geriatric psychiatry review: differential diagnosis and treatment of the 3 D's—delirium, dementia, and depression. *Curr Psychiatry Rep*. 2013;15:365.

Dubovsky SL, Frobose C, Phiri P, de Greef R, Panagides J. Short-term safety and pharmacokinetic profile of asenapine in older patients with psychosis. *Int J Geriatr Psychiatry*. 2012;27:472–482.

Friedman JH. Atypical antipsychotic drugs in the treatment of Parkinson's disease. *J Pharm Pract*. 2011;24(6):534–540.

Gareri P, De Fazio P, Manfredi VG, De Sarro G. Use and safety of antipsychotics in behavioral disorders in elderly people with dementia. *J Clin Psychopharmacol*. 2014;34:109–123.

Graff-Guerrero A, Rajji TK, Mulsant BH, et al. Evaluation of antipsychotic dose reduction in late-life schizophrenia: a prospective dopamine D2/3 receptor occupancy study. *JAMA Psychiatry*. 2015;72(9):927–934.

Jackson JW, Schneeweiss S, VanderWeele TJ, Blacker D. Quantifying the role of adverse events in the mortality difference between first and second-generation antipsychotics in older adults: systematic review and meta-synthesis. *PLoS One*. 2014;9(8):e105376.

Jackson JW, VanderWeele TJ, Viswanathan A, Blacker D, Schneeweiss S. The explanatory role of stroke as a mediator of the mortality risk difference between older adults who initiate first-versus second-generation antipsychotic drugs. *Am J Epidemiol*. 2014; 180(8):847–852.

Jalbert JJ, Daiello LA, Eaton CB, Miller SC, Lapane KL. Antipsychotic use and the risk of diabetes in nursing home residents with dementia. *Am J Geriatr Pharmacother*. 2011;9:153–163.

*Jin H, Shih PA, Golshan S, et al. Comparison of longer-term safety and effectiveness of 4 atypical antipsychotics in patients over age 40: a trial using equipoise-stratified randomization. *J Clin Psychiatry*. 2013;74(1):10–18.

Kheirbek RE, Fokar A, Little JT, et al. Association between antipsychotics and all-cause mortality among community-dwelling older adults. *J Gerontol A Biol Sci Med Sci*. 2019;74(12):1916–1921.

Lenze EJ, Mulsant BH, Blumberger DM, et al. Efficacy, safety, and tolerability of augmentation pharmacotherapy with aripiprazole for treatment-resistant depression in late life: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2015;386(10011):2404–2412.

*Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet*. 2013;382:951–962.

Lin S-T, Chen C-C, Tsang H-Y, et al. Association between antipsychotic use and risk of acute myocardial infarction: a nationwide case-crossover study. *Circulation*. 2014;130(3):235–243.

Maglione JE, Thomas SE, Jeste DV. Late-onset schizophrenia: do recent studies support categorizing LOS as a subtype of schizophrenia? *Curr Opin Psychiatry*. 2014;27:173–178.

Mehta S, Chen H, Johnson M, Aparasu RR. Risk of serious cardiac events in older adults using antipsychotic agents. *Am J Geriatr Pharmacother*. 2011;9:120–132.

Rado J, Janicak PG. Pharmacological and clinical profile of recently approved second-generation antipsychotics: implications for treatment of schizophrenia in older patients. *Drugs Aging*. 2012;29:783–791.

Rosenheck RA, Leslie DL, Sindelar JL, et al; Clinical Antipsychotic Trial of Intervention Effectiveness-Alzheimer’s Disease (CATIE-AD) investigators. Cost-benefit analysis of second-generation antipsychotics and placebo in a randomized trial of the treatment of psychosis and aggression in Alzheimer disease. *Arch Gen Psychiatry*. 2007;64:1259–1268.

Sajatovic M, Forester BP, Gildengers A, Mulsant BH. Aging changes and medical complexity in late-life bipolar disorder: emerging research findings that may help advance care. *Neuropsychiatry (London)*. 2013;3(6):621–633.

Schneider-Thoma J, Efthimiou O, Huhn M, et al. Second- generation antipsychotic drugs and short-term mortality: a systematic review and meta-analysis of placebo-controlled randomized controlled trials. *Lancet Psychiatry*. 2018;5(8):653–663.

Seitz DP, Gill SS, Herrmann N, et al. Pharmacological treatments for neuropsychiatric symptoms of dementia in long-term care: a systematic review. *Int Psychogeriatr*. 2013;25(2):185–203.

Van Leeuwen E, Petrovic M, van Driel ML, et al. Stopping or continuing long-term antipsychotic drug use for behavioural and psychological symptoms in older people with dementia. *Cochrane Database Syst Rev*. 2018;3(3):CD007726.

Vigen CL, Mack WJ, Keefe RS, et al. Cognitive effects of atypical antipsychotic medications in patients with Alzheimer's disease: outcomes from CATIE-AD. *Am J Psychiatry*. 2011;168(8):831–839.

*Woerner MG, Correll CU, Alvir JM, Greenwald B, Delman H, Kane JM. Incidence of tardive dyskinesia with risperidone or olanzapine in the elderly: results from a 2-year, prospective study in antipsychotic-naïve patients. *Neuropsychopharmacology*. 2011;36:1738–1746.

Wustmann T, Pillmann F, Friedemann J, Piro J, Schmeil A, Marneros A. The clinical and sociodemographic profile of persistent delusional disorder. *Psychopathology*. 2012;45:200–202.

Yunusa I, Alsumali A, Garba AE, Regestein QR, Eguale T. Assessment of reported comparative effectiveness and safety of atypical antipsychotics in the treatment of behavioral and psychological symptoms of dementia: a network meta-analysis. *JAMA Netw Open*. 2019;2(3):e190828.

55.4e Antidementia Drugs

IQBAL AHMED, M.D., F.R.C.PSYCH.

The medications available for prescription in the United States for treating Alzheimer disease include cholinesterase inhibitors, donepezil rivastigmine, galantamine, and the noncompetitive *N*-methyl-D-aspartate (NMDA) antagonist memantine. The cholinesterase inhibitors are specifically indicated in their U.S. Food and Drug Administration (FDA)–approved labeling for "the treatment of mild to moderate dementia of the Alzheimer type." Donepezil and rivastigmine transdermal are further indicated for severe Alzheimer disease as well. Rivastigmine also has approved labeling for "the treatment of mild to moderate dementia of Parkinson's disease." Memantine is indicated for the treatment of moderate to severe Alzheimer disease; the FDA specifically did not approve an indication for its use in mild Alzheimer disease. A capsule containing both donepezil and memantine is also marketed for moderate to severe Alzheimer disease. These drugs, thus, are considered "antidementia" drugs.

Antidementia drugs are, by definition, medications intended to improve cognitive function in patients with dementia. The term has regulatory significance with the FDA because it differentiates these medications from “cognitive enhancers,” medications that might be expected to improve cognition without reference to a particular illness, and from medications that might improve such behavioral syndromes as depression, aggression, or psychosis in people with dementia.

Worldwide, a range of putative cognitive enhancing agents is offered with and without prescription, including piracetam derivatives, cholinesterase inhibitors, an NMDA antagonist, memantine, selegiline, antioxidants, vitamins, food supplements, so-called nutraceuticals, and various plant derivatives, including *Ginkgo biloba* standardized extract. In the United States, food supplements and nutraceuticals are regulated by the FDA only with respect to apparent safety and inappropriate advertising claims. No specific health claim or disease-treatment claim should be made by the marketers of nutraceuticals and food supplements, and efficacy evidence for the conditions being treated does not have to be addressed. Rather, marketers of nutraceuticals and food supplements may make a “structure/function claim,” characterizing the supplement as intended to affect normal structure or function or the mechanism by which it acts to maintain such structure or function (21 CFR 101.93).

Although it has been suggested that other medications available in the United States might have efficacy in patients with Alzheimer disease, clinical trial results indicate that selegiline, vitamin E, *Ginkgo biloba* extract, conjugated estrogens, and anti-inflammatories do not have salutary effects—in some cases, they have cognitively impairing effects—and should not be used to treat dementia.

Alzheimer disease is the most common dementia associated with aging and hence is the most common target for antidementia therapeutics. Vascular dementia, dementia with Lewy bodies, Parkinson disease dementia, frontotemporal dementia, and dementias associated with hippocampal sclerosis are less common, although there is a large degree of overlap of the pathology of Alzheimer disease with these other pathologies.

Although the causes of Alzheimer disease are unknown, except for rare early-onset familial and genetic cases, a substantial part of the pathophysiology is sufficiently understood to allow for the rational development of symptomatic or potentially disease-modifying treatments.

The framework for clinical trials in Alzheimer disease developed slowly during the last two decades after the advent and general acceptance of research-based diagnostic criteria for the dementia of Alzheimer disease (i.e., the essentially identical National Institute of Neurological Communicative Disorder [NINCDS] Alzheimer's Disease and Related Disorders Association [ADRDA] criteria for probable Alzheimer disease, and the *Diagnostic and Statistical Manual of Mental Disorders,* fourth edition text revision [DSM-IV-TR] criteria for dementia of the Alzheimer type), a better understanding of its underlying pathology, and the development of mechanism-based pharmacologic therapeutics. For example, the pathophysiology of Alzheimer disease is characterized, in part, by early loss of synapses, death of neurons, the deposition of insoluble β-amyloid peptide fragments, atypical inflammatory responses, and the excess production of toxic, oxidizing free radicals, thus providing targets for drug development.

Until the 1980s, medications were proposed for dementia treatment based on clinical experience or prevailing theories of dementia and aging. Examples of such earlier claimed treatments (and their rationale) include nonspecific psychostimulants (to increase arousal), vasodilators (to increase blood flow), and hyperbaric oxygen (to enhance metabolism). One medication, ergoloid mesylates, was approved in the United States for the ill-defined condition "senile mental decline," a condition based on vague diagnostic constructs, such as "cerebral insufficiency" or "cerebral deterioration." The mechanism of action of ergoloid mesylates is unknown but may involve effects on complex neurotransmitter and vascular effects. However, it is no longer available or marketed in the United States.

REGULATORY AND METHODOLOGIC CONSIDERATIONS

FDA guidelines for establishing that a drug has antidementia efficacy include requirements that patients meet accepted diagnostic criteria for a primary well-accepted dementia, such as Alzheimer disease; that trials use double-blind, placebo-controlled parallel-group designs; that at least one trial is of 6 months duration; and that improvement is obtained on an accepted cognitive instrument and a clinician’s global assessment, to ensure a clinically meaningful effect as compared to a potentially measurable but trivial cognitive effect. Improvements in activities of daily living also may be considered as an index of clinical meaning in lieu of clinical global assessments.

From a regulatory perspective, behavioral changes as a result of treatment are of secondary importance despite the clinical importance of such improvement. Specifically, improvement in behavior or functional activities in the absence of improved cognition is not a sufficient outcome for allowing a drug to be labeled “antidementia.” A concern is that antidepressants or antipsychotics might be misleadingly advertised as being effective for dementia when their intended use is to treat depressive or psychotic symptoms or aggression.

For clinical trials in Alzheimer disease, the Alzheimer’s Disease Assessment Scale—cognitive subscale (ADAS-cog) has become the standard cognitive change measure, and a clinician’s interview-based impression of change instrument or activities of daily living inventory has become the measure of “clinical meaningfulness.” The ADAS-cog provides a summary score reflecting measures of word recall, naming, following commands, constructional and ideational praxis, orientation, word recognition, spoken language, comprehension of spoken language, word finding, and recall of test instructions. Although it has proven itself sufficiently sensitive to change with cholinesterase inhibitor treatments in mild to moderate Alzheimer disease, it is not as sensitive to change in mild dementia and mild cognitive impairment (MCI) due to Alzheimer disease (also known as prodromal Alzheimer disease), and people with severe dementia have difficulty performing the test. More recently, another cognitive battery, the Severe Impairment Battery, has been used as an index of cognitive change in Alzheimer disease trials of patients who are moderate to more severely cognitively impaired. It is a group of brief tests assessing social interaction, memory, orientation, language, attention, praxis, and construction.

For mild Alzheimer disease and MCI due to Alzheimer disease, a Neuropsychological Test Battery (an NTB) has been offered that—in its most common form—consists of three tests of immediate and delayed memory, verbal and category fluency, and digit span; and the Repeatable Battery For The Assessment of Neuropsychological Status (RBANS) which similarly assesses memory, language, attention, and visuospatial/constructional ability. Regulatory guidelines for marketing approval do not exist for the treatment of other dementias (e.g., vascular dementia, dementia with Lewy bodies, frontotemporal dementia), MCI (other than MCI due to Alzheimer disease), other cognitive disorders, or age-related cognitive decline or for cognitively unimpaired persons, mainly because of a lack of general expert consensus on the validity of the proposed diagnostic criteria. One exception is Parkinson disease associated mild to moderate dementia for which rivastigmine is approved by the FDA for this indication.

In 2018, FDA proposed draft guidelines for the clinical development of drugs for early-stage Alzheimer disease, including the entities MCI due to Alzheimer disease and preclinical Alzheimer disease. The guidelines allow for only a composite outcome, including assessments of cognition and function combined in order to approve a drug for MCI and a single cognitive outcome for provisional approval for a drug for preclinical Alzheimer disease. No drugs have been approved under these guidelines.

Although there are efforts to gain recognition of functionally significant but mild cognitive loss, such as age-associated memory impairment, in which cognition is lower than a young adult comparison, or age-related cognitive decline, in which memory is impaired compared to an age-appropriate comparison, the vagueness of the definitions and underlying pathology limits their general acceptance, and hence the acceptance of “treatments” for such unrecognized conditions. People with age-associated memory impairment generally do not worsen cognitively at a rate greater than the population, whereas those with MCI have an increased rate of worsening and developing dementia. The results of several clinical trials for improving cognition in MCI using the cholinesterase inhibitors donepezil, galantamine, and rivastigmine and vitamin E, anti-inflammatories, and piracetam (Nootropil) have not shown significant cognitive effects for the drugs.

With respect to unimpaired older persons, nearly all of whom can claim that memory, learning, and other aspects of cognition are not as good as they used to be, there are societal questions about the appropriateness of treatment with cognitive enhancers even if they are demonstrated to be effective. Similarly, generally accepted guidelines do not exist for the treatment of cognitive deficits associated with schizophrenia, attention-deficit disorder, or depression, despite general recognition that these cognitive deficits are important areas of clinical attention. There have been attempts to provide clinical criteria for “cognitive impairment associated with schizophrenia” and cognitive outcomes criteria for clinical trials.

PHARMACOLOGIC TREATMENT OF SYMPTOMATIC BEHAVIORS OR BEHAVIORAL AND PSYCHOLOGICAL SYMPTOMS OF DEMENTIA

The majority of patients with Alzheimer disease experience such behavioral symptoms as depression, anxiety, psychosis, agitation, aggression, apathy, withdrawal, and vegetative disturbances (e.g., sleep or appetite change). These signs and symptoms frequently lead to the use of psychotropic medication treatment (both appropriate and inappropriate), physical restraint, and institutionalization. Behavioral symptoms burden caregivers and are considered among the most stressful aspects of caregiving. A variety of psychotropic medications, however, may be effective for pharmacologic intervention and improve many of these symptoms. Evidence suggests that these symptoms occurring in the context of Alzheimer disease can represent manifestations of two syndromes: psychosis of Alzheimer disease and depression of Alzheimer disease.

The treatment of symptomatic behaviors begins with careful characterization of the target symptoms. Such symptoms as agitation, aggression, or delusions can result from chronic or recurrent psychiatric disorders, represent the first manifestation of a new psychiatric disorder, or indicate a change in medical status. (The term *agitation* often is used to describe a range of nonspecific behaviors in older adults beyond excess psychomotor activity, including wandering, pacing, uncooperativeness, and verbal outbursts. Similarly, the commonly used phrase *behavioral and psychological disorders of dementia* is vague and unspecific and probably should not be used.) A comprehensive medical evaluation might uncover potentially treatable underlying physical conditions that can exacerbate behaviors. Moreover, patients with dementia are at increased risk for developing superimposed delirium and consequent disruptive behaviors.

Patients with dementia are sensitive to a variety of factors that may influence behavior, including sensory impairment, environmental disruption, and altered daily routines. Consequently, environmental manipulations may be helpful. Many dementia patients experience anger, sadness, and denial, which may improve with psychosocial and psychotherapeutic interventions. Various behavior management techniques, including cognitive-behavioral management, may be helpful for more moderately to severely impaired individuals. Supportive psychotherapy and counseling may be helpful for patients with more MCIs. Caregivers often need greater awareness that their own behaviors influence the patient. Counseling about the nature of the patient’s illness not only helps relatives cope better but also is associated with better outcomes for patients, such as delayed institutionalization.

Although the nonpharmacologic and pharmacologic treatment of symptomatic behaviors in dementia patients is beyond the scope of this section, the suggested interventions are necessary preparatory approaches to the pharmacologic treatment of cognitive impairment as well.

DRUGS TO TREAT DEMENTIA

From a conceptual consideration, antidementia drugs can be grouped into two broad categories: those intended to improve cognitive function symptomatically and those intended to modify the course of illness over the longer term, as manifested by the potential attenuation of cognitive and functional decline. In practice, these two categories overlap. Medication development for dementia typically focuses on one or the other, depending on the mechanism of action of the new drug. The only drugs currently approved, however, are for symptomatic treatment, and their efficacy and FDA marketing approval were based on 6-month-long, placebo-controlled clinical trials. Table 55.4e–1 lists available drugs to treat the cognitive symptoms of dementia.

Cholinergic Hypothesis

The cholinergic hypothesis of dementia is based on observations of a massive decline of the cholinergic projections from the basal forebrain to the cortex; reductions in the activity of choline acetyltransferase, the key enzyme in acetylcholine synthesis; a marked loss of cholinergic cell bodies in the nucleus basalis in Alzheimer disease; correlations between choline acetyltransferase reduction or nucleus basalis cell reduction and neuritic, β-amyloid–containing plaque densities in cortical areas; and correlation of the cholinergic deficits with the decline in cognitive test performance.

Table 55.4e–1. U.S. Food and Drug Administration–Approved and Marketed Medications for Alzheimer Disease

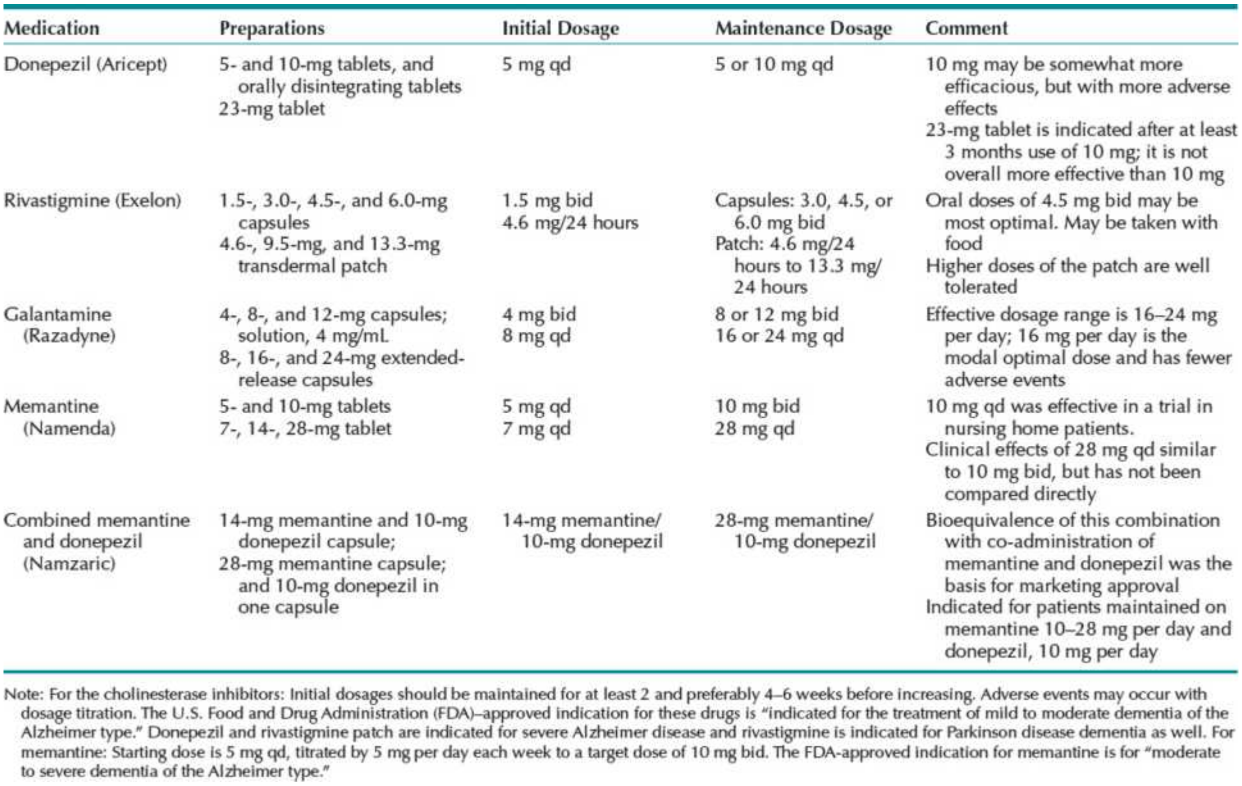

Medication	Preparations	Initial Dosage	Maintenance Dosage	Comment
Donepezil (Aricept)	5- and 10-mg tablets, and orally disintegrating tablets 23-mg tablet	5 mg qd	5 or 10 mg qd	10 mg may be somewhat more efficacious, but with more adverse effects 23-mg tablet is indicated after at least 3 months use of 10 mg; it is not overall more effective than 10 mg
Rivastigmine (Exelon)	1.5-, 3.0-, 4.5-, and 6.0-mg capsules 4.6-, 9.5-mg, and 13.3-mg transdermal patch	1.5 mg bid 4.6 mg/24 hours	Capsules: 3.0, 4.5, or 6.0 mg bid Patch: 4.6 mg/24 hours to 13.3 mg/24 hours	Oral doses of 4.5 mg bid may be most optimal. May be taken with food Higher doses of the patch are well tolerated
Galantamine (Razadyne)	4-, 8-, and 12-mg capsules; solution, 4 mg/mL 8-, 16-, and 24-mg extended-release capsules	4 mg bid 8 mg qd	8 or 12 mg bid 16 or 24 mg qd	Effective dosage range is 16–24 mg per day; 16 mg per day is the modal optimal dose and has fewer adverse events
Memantine (Namenda)	5- and 10-mg tablets 7-, 14-, 28-mg tablet	5 mg qd 7 mg qd	10 mg bid 28 mg qd	10 mg qd was effective in a trial in nursing home patients. Clinical effects of 28 mg qd similar to 10 mg bid, but has not been compared directly
Combined memantine and donepezil (Namzaric)	14-mg memantine and 10-mg donepezil capsule; 28-mg memantine capsule; and 10-mg donepezil in one capsule	14-mg memantine/ 10-mg donepezil	28-mg memantine/ 10-mg donepezil	Bioequivalence of this combination with co-administration of memantine and donepezil was the basis for marketing approval Indicated for patients maintained on memantine 10–28 mg per day and donepezil, 10 mg per day

Note: For the cholinesterase inhibitors: Initial dosages should be maintained for at least 2 and preferably 4–6 weeks before increasing. Adverse events may occur with dosage titration. The U.S. Food and Drug Administration (FDA)–approved indication for these drugs is "indicated for the treatment of mild to moderate dementia of the Alzheimer type." Donepezil and rivastigmine patch are indicated for severe Alzheimer disease and rivastigmine is indicated for Parkinson disease dementia as well. For memantine: Starting dose is 5 mg qd, titrated by 5 mg per day each week to a target dose of 10 mg bid. The FDA-approved indication for memantine is for "moderate to severe dementia of the Alzheimer type."

Many other neuronal systems are affected besides the cholinergic system, including pronounced declines in the noradrenergic projections from locus ceruleus to cortex and general changes in glutamate systems. Notably, there is greater research emphasis on more fundamental pathogenetic processes affecting neurodegeneration, including β-amyloid peptide production and deposition, tau-protein phosphorylation resulting in neurofibrillary tangles, inflammatory responses, and the use of medications that might be neuroprotective.

Nevertheless, the observation that cholinergic deficits represent the most pronounced transmitter deficit throughout the dementia process has led to a considerable focus on cholinergic restitutive strategies. The cholinergic hypothesis of dementia treatment implies that if the cholinergic deficit is responsible for cognitive or behavioral changes, then potentiation of central cholinergic function should improve cognitive and behavioral impairment associated with Alzheimer disease.

Thus far, there have been three main cholinergic approaches to Alzheimer disease treatment: use of precursor loading, use of direct cholinergic agonists, and cholinesterase inhibition.

Precursor Loading and Direct Cholinergic Stimulation

Early investigations focused on the use of acetylcholine precursors, including choline and phosphatidyl choline (lecithin), analogous to the use of dopamine precursors in Parkinson disease. Numerous trials with precursors, however, have not shown significant results, failing to show a reliable improvement in cognitive function. Thus, even though these products are sold in health food stores, their use cannot be recommended.

Direct cholinergic agonists enhance responsiveness to released acetylcholine and minimize the presynaptic inhibition of further release by intrasynaptic acetylcholine concentrations. Thus far, however, direct muscarinic or nicotinic receptor agonists have not proven useful in enhancing cognition in Alzheimer disease. The older and newer, relatively more muscarinic receptor subtype-specific investigational medications have not proven useful either. Marketed cholinergic agonists, such as arecoline and bethanechol (brand name, Urecholine), cannot be recommended because of their inadequate central nervous system (CNS) penetration, high frequency of cholinergic side effects, and minimal, if any, efficacy.

Although nicotinic agents currently under development, including transdermal nicotine, nicotinic α_7 receptor modulators, and newer nicotinic α_4, and β_2 receptor modulators, may improve attention, arousal, and concentration in adults, their usefulness in such cognitive impairment syndromes as Alzheimer disease has not been adequately investigated. They also may have a role in attention-deficit disorder and cognitive impairment associated with schizophrenia. So far, none of the two α_7 nicotinic receptor modulators which have been studied have had positive effects on cognition or function.

Cholinesterase Inhibitors

The most successful approach to the treatment of Alzheimer disease thus far has been cholinesterase inhibition. In early studies, physostigmine demonstrated modest cognitive improvements but required individualized and very frequent dosing because of its very short duration of action (1 to 2 hours), thus limiting any potential clinical usefulness.

Table 55.4e–2.
Pharmacodynamics and Pharmacokinetics of Marketed Cholinesterase Inhibitors

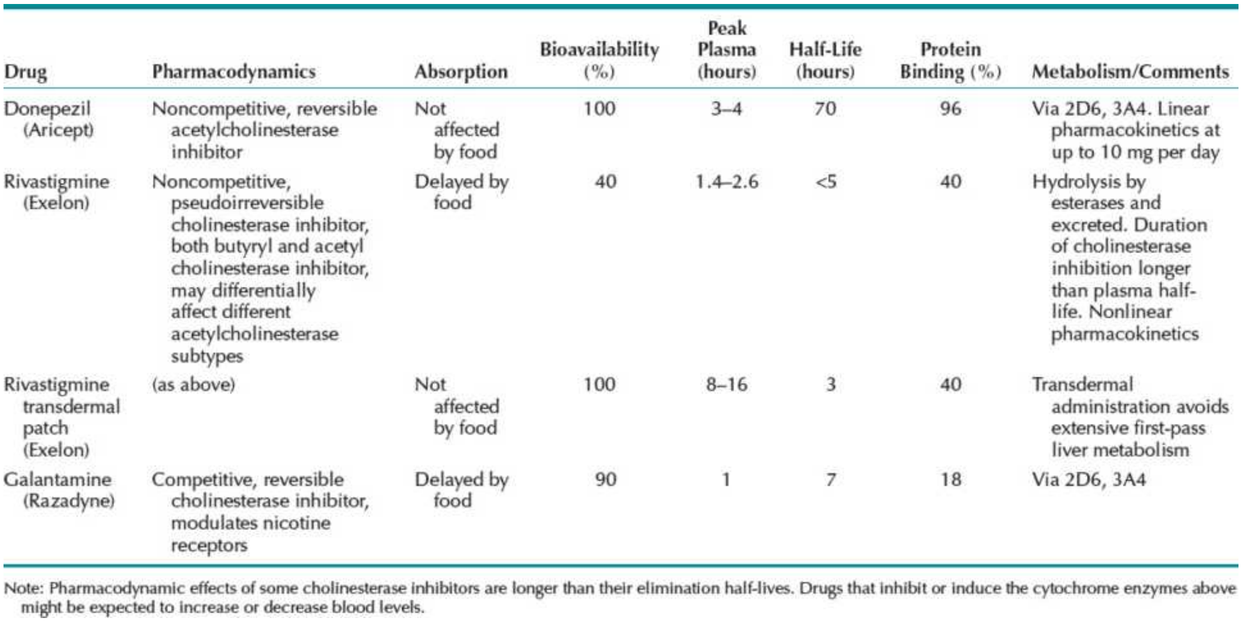

Drug	Pharmacodynamics	Absorption	Bioavailability (%)	Peak Plasma (hours)	Half-Life (hours)	Protein Binding (%)	Metabolism/Comments
Donepezil (Aricept)	Noncompetitive, reversible acetylcholinesterase inhibitor	Not affected by food	100	3–4	70	96	Via 2D6, 3A4. Linear pharmacokinetics at up to 10 mg per day
Rivastigmine (Exelon)	Noncompetitive, pseudoirreversible cholinesterase inhibitor, both butyryl and acetyl cholinesterase inhibitor, may differentially affect different acetylcholinesterase subtypes	Delayed by food	40	1.4–2.6	<5	40	Hydrolysis by esterases and excreted. Duration of cholinesterase inhibition longer than plasma half-life. Nonlinear pharmacokinetics
Rivastigmine transdermal patch (Exelon)	(as above)	Not affected by food	100	8–16	3	40	Transdermal administration avoids extensive first-pass liver metabolism
Galantamine (Razadyne)	Competitive, reversible cholinesterase inhibitor, modulates nicotine receptors	Delayed by food	90	1	7	18	Via 2D6, 3A4

Note: Pharmacodynamic effects of some cholinesterase inhibitors are longer than their elimination half-lives. Drugs that inhibit or induce the cytochrome enzymes above might be expected to increase or decrease blood levels.

Cholinesterase inhibitors inhibit CNS acetylcholinesterase in the synaptic cleft, thus preventing the degradation of endogenously released acetylcholine (Table 55.4e–2). However, progressive presynaptic cholinergic neuron loss continues, so there is a progressively diminishing production of acetylcholine over time.

The three cholinesterase inhibitors that are currently marketed for the treatment of the cognitive symptoms of Alzheimer disease, and tacrine which is no longer marketed, can be grouped into two classes according to their structure and mode of inhibition: (1) reversible inhibitor tertiary and quaternary amines (i.e., tacrine, donepezil, or galantamine) and (2) pseudoirreversible inhibitor carbamates (i.e., rivastigmine). The irreversible organophosphate cholinesterase inhibitor metrifonate was clearly efficacious at higher doses but was not approved for marketing because of myasthenia and respiratory depression, possibly owing to its effects at the myoneural junction in muscles of respiration.

INDIVIDUAL CHOLINESTERASE INHIBITORS: CLINICAL STUDIES, DOSING, AND ADVERSE EFFECTS

For the most part, the several published multicenter trials of cholinesterase inhibitors were designed with the objective of obtaining marketing approval from the FDA. The protocols were fairly similar to each other, generally selecting outpatients with mild to moderate Alzheimer disease, usually with Mini-Mental State Examination (MMSE) scores between 10 and 26, inclusively. Patients in these trials were generally physically healthy, usually treated for 6 months or less, and more likely to be female, and they had a mean age of approximately 72 to 74 years, a decade lower than the median age of Alzheimer disease patients in the United States.

Tacrine

Tacrine is a noncompetitive reversible inhibitor of cholinesterase. It binds near the catalytically active site of the acetylcholinesterase molecule to inhibit enzyme activity. It has other actions as well, including blocking sodium and potassium channels and direct activity at muscarinic receptors. Statistically significant treatment effects for the 30 mg qid and 40 mg qid dosage groups were found on the ADAS-cog, a clinician’s interview-based impression of change rating, and activities of daily living in 3- and 6-month clinical trials. Tacrine will not be discussed further because it is no longer marketed. Additional information can be found in the previous editions of this book.

Donepezil

Donepezil is a long-acting, piperidine-based, highly selective, and reversible acetylcholinesterase inhibitor. Two phase III clinical trials provided pivotal evidence for the drug's FDA approval in late 1996. Additional randomized clinical trials include a trial of 6 months duration, a Scandinavian study of 12 months, a study of nursing home patients, a 6-month trial in which a large proportion of patients were more severely impaired than in previous trials, and an assessment of activities of daily living.

One non–industry-sponsored, randomized, placebo-controlled trial followed patients over several years, reporting very modest cognitive effects over 2 years but no significant effects on loss of function, nursing home placement, or health economic measures.

Pharmacokinetics. Oral bioavailability approaches 100%, with peak concentrations in 4 hours and linear pharmacokinetics. Donepezil is extensively bound to plasma proteins, including α_1-acid glycoproteins. The drug accumulates at a constant rate, reaching a steady state in approximately 2 weeks. It has a relatively slow clearance, with a long elimination half-life of 70 hours. It is both excreted unchanged in the urine and extensively metabolized by cytochrome P450 (CYP) 2D6 and 3A4 hepatic enzymes to active and inactive metabolites.

Drug Interactions. Inhibitors of CYP 3A4 and CYP 2D6 (such as ketoconazole [Nizoral] and quinidine [Quinaglute]) and inducers of CYP 3A4 and CYP 2D6 (such as phenytoin, carbamazepine, phenobarbital, and dexamethasone) could inhibit donepezil metabolism or increase its rate of elimination but may not be clinically significant in view of its high protein binding and very low therapeutic concentrations of 2 to 12 ng/mL.

Dosing. Donepezil is initiated at 5 mg per day and then increased to 10 mg per day after 4 to 6 weeks. Raising the dose earlier increases the risk for cholinergic adverse events. Five or 10 mg per day is an effective dose; 10 mg tends to be somewhat more effective than 5 mg when the various trials as a group are evaluated. The efficacy or appropriateness of higher doses has not been established.

A 23-mg formulation was marketed in 2010 intended for patients who are being maintained on 10 mg and for whom a higher dose is thought desirable. In a randomized, controlled trial, efficacy compared to patients who continued on 10 mg per day was not demonstrated.

Rivastigmine

Rivastigmine is a pseudoirreversible, selective acetylcholinesterase inhibitor and butyrylcholinesterase inhibitor. Four phase III randomized, placebo-controlled, 26-week-long clinical trials were completed of similar design but differing mainly in dosing methods. Two have been published, and the results of the other two have been included in secondary reports.

In the two published trials, doses were titrated weekly during the first 7 weeks to one of two dosage ranges, 1 to 4 mg per day or 6 to 12 mg per day, and dose decreases were not permitted, possibly contributing to lesser tolerability and seemingly more side effects during these stages of treatment.

A transdermal patch formulation was developed based on a placebo-controlled trial in mild to moderate Alzheimer disease comparing a 17.4-mg patch, 9.5-mg patch, and 6 mg of orally administered rivastigmine twice per day, and another trial using a 13.3-mg patch in severe Alzheimer disease.

Pharmacokinetics. Rivastigmine is well absorbed; very little binds to protein. Although the elimination half-life is less than 2 hours, enzyme inhibition lasts approximately 9 hours. Rivastigmine is not metabolized by the hepatic microsomal system. Rather, after binding to acetylcholinesterase, the carbamate portion of rivastigmine is slowly hydrolyzed (hence the “pseudoirreversibility”), conjugated to a sulfate, and excreted renally. Thus, its metabolism is essentially extrahepatic and unlikely to have significant pharmacokinetic interactions.

The pharmacokinetics of the transdermal patch shows a time to a maximum concentration of 8 to 16 hours and a 3-hour elimination half-life after the patch is removed.

Drug Interactions. Rivastigmine's metabolism does not depend on liver CYP enzymes; therefore, no drug interactions related to hepatic cytochromes have been observed.

Dosing. The starting dose of rivastigmine is 1.5 mg twice a day, taken with meals. If this dose is well tolerated after a minimum of 2 weeks of treatment, it may be increased to 3 mg twice a day. Subsequent increases to 4.5 mg and then 6 mg twice a day should be based on good tolerability of the current dose and may be considered after a minimum of 2 weeks of treatment. Higher daily doses are associated with better efficacy than lower doses. The initial transdermal patch is applied at 4.6 mg (5 cm^2) per day and then increased to one 9.5-mg (10-cm^2) patch per day after 4 weeks. The dosage level may be increased to a 13.3 (15-cm^2)-mg patch per day after a subsequent 4 weeks, and the usual maintenance doses may be 9.5-mg or 13.3-mg patch per day.

Galantamine

Galantamine, an alkaloid originally extracted from *Amaryllidaceae* (*Galanthus woronowii,* the Caucasian snowdrop) but now synthesized, is a reversible, competitive inhibitor of acetylcholinesterase with relatively less butyrylcholinesterase activity.

Because competitive inhibitors compete with acetylcholine at acetylcholinesterase-binding sites, their inhibition is, theoretically, dependent on intrasynaptic acetylcholine concentration in that they will be less likely to bind to sites in brain areas that have high acetylcholine levels. Another characteristic is allosteric modulation of nicotinic receptor sites, possibly enhancing cholinergic transmission by presynaptic nicotinic stimulation.

At least four large clinical trials involving more than 2,400 subjects have been published, generally showing efficacious doses between 8 mg twice a day and 16 mg twice a day, with few adverse effects at the low dose. Four milligrams twice a day was not an effective dose.

Pharmacokinetics. Galantamine is well absorbed and has approximately 90% bioavailability, little protein binding, and linear pharmacokinetics, with peak concentrations in approximately 1 hour and an elimination half-life of approximately 7 hours. It is metabolized by CYP 2D6 and 3A4, glucuronidated, and excreted in the urine. There is an approximately 25% decrease in clearance of galantamine in CYP 2D6 poor metabolizers.

Drug Interactions. Drugs that are potent inhibitors for CYP 2D6 or 3A4 may increase the availability and plasma levels of galantamine. Coadministration of paroxetine and ketoconazole increases the bioavailability of galantamine by 40% and 30%, respectively. Galantamine clearance is decreased by approximately 25% to 33% by concurrent administration of CYP 2D6 inhibitors, such as amitriptyline, fluoxetine, and quinidine; cimetidine increased bioavailability by approximately 16%. Galantamine itself does not inhibit CYP 3A4 or 2D6 and has little effect on the metabolism of other drugs.

Dosing. Initial dosing is 4 mg twice a day and should be raised to 8 mg twice a day after 2 to 4 weeks. For patients who are tolerating medication but not responding, the dose can be raised to 12 mg twice a day after another 4 weeks. The initial dose of the extended-release formulation is 8 mg per day and raised to 16 mg per day after 2 to 4 weeks, with the option to raise it to 24 mg per day.

Adverse Effects of Cholinesterase Inhibitors

The most common adverse events with cholinesterase inhibitors are cholinergically mediated and involve those related to cholinergic excess, including nausea, vomiting, diarrhea, dyspepsia, anorexia, fatigue, and weight loss. In addition, myalgia or muscle cramps occur in approximately 7% to 9% of patients, indicating direct myoneural effects (but do not seem to be associated with rivastigmine), which specifically inhibits only certain CNS acetylcholinesterase subtypes. The frequency of the cholinergic side effects generally is less than 15% and often no greater than placebo in clinical trials. In general, they tend to be mild and short lived, lasting less than 2 or 3 days, and occur around the time that dosage is increased. Medication dosage can usually be maintained or temporarily lowered and then raised. Fatigue also tends to occur, and occasionally there is a decrease in heart rate of 4 to 7 beats per minute, possibly due to enhanced vagal effects.

Direct reversible hepatotoxicity was associated only with tacrine, more commonly in women than men, and is the most common abnormal finding with this drug which is no longer marketed.

Contraindications to Cholinesterase Inhibitors

There are few absolute contraindications to cholinesterase inhibitors, and they relate to the unique chemical properties of each drug. Because of the actions of cholinesterase inhibitors, these drugs require caution when used in patients with significant asthma, significant chronic obstructive pulmonary disease, gastrointestinal ulcer disease, cardiac conduction defects, or clinically significant bradycardia. Appropriate considerations are involved in general anesthesia as well because they may prolong the effects of succinylcholine-type drugs.

Donepezil. The most common gastrointestinal adverse effects of donepezil include nausea, vomiting, diarrhea, and anorexia. In addition, some patients develop muscle cramps, headache, dizziness, syncope, flushing, insomnia, weakness, drowsiness, fatigue, agitation, and sleeplessness. Weight loss occurred at twice the rate of placebo in the nursing home study but not in outpatient trials. Adverse effects occurred at a higher rate when the titration from 5 to 10 mg was made in 1 week compared to 6 weeks.

Rivastigmine. Adverse effects with orally administered rivastigmine were primarily gastrointestinal and occurred in the high-dose (6 to 12 mg per day) group. They occurred mainly during dose escalation and led to withdrawal in one study in 23% of the high-dose group, 7% of the low-dose group, and 7% of the placebo group. Adverse effects that occurred in the higher-dose group were sweating, fatigue, asthenia, weight loss, malaise, dizziness, somnolence, nausea, vomiting, anorexia, and flatulence. In the maintenance phase, dizziness, nausea, vomiting, dyspepsia, and sinusitis occurred more in the 6 to 12 mg per day group than in the placebo group.

Adverse effects with the transdermal patch at the 17.4-mg dose were similar to 6 mg twice per day of oral rivastigmine, with the 13.3-mg dose somewhat less, and with the 9.5-mg dose were not significantly greater than placebo.

Galantamine. Galantamine's principal adverse effects were nausea, vomiting, diarrhea, anorexia, weight loss, abdominal pain, dizziness, and tremor. They were more frequent earlier in the course of treatment and during the dosage titration from 16 mg a day to 24 mg a day and higher. In one trial, the effective dosage of 16 mg per day was not associated with overall greater adverse events than experienced by the placebo-treated group.

Clinical Application and Treatment Implementation for Cholinesterase Inhibitors

The differential pharmacology of the cholinesterase inhibitors requires treatment considerations, such as once-per-day versus twice-per-day dosing, oral versus transdermal administration, and acetylcholinesterase inhibition versus butyrylcholinesterase inhibition. Answers are not known.

Donepezil and rivastigmine transdermal patches have been demonstrated to be effective in more severe patients as well as in outpatients with mild to moderate Alzheimer disease (MMSE scores between 10 and 26). Some experts believe that cognitive efficacy also is apparent in the more mildly and more severely cognitively impaired, but very few studies have been done on mild Alzheimer disease or MCI and did not show advantageous effects.

Duration of Treatment and Discontinuation

Placebo-controlled clinical trials with cholinesterase inhibitors have lasted only 6 months, with only a few exceptions. In some trials, when medication is discontinued, patients return within 3 to 6 weeks to the cognitive level of the placebo patients, indicating an overall symptomatic effect and that treatment efficacy is longer than 6 months. Three placebo-controlled trials suggest that, in some cases, efficacy could be longer than 1 year on cognitive and functional measures. A few analyses of databases or open-label extensions of clinical trials suggest that patients who tolerate and continue to take cholinesterase inhibitors over at least 1 year might experience a delay in nursing home placement compared to those who cannot tolerate or do not take them. However, these analyses are biased and probably subject to the “healthy patient effect,” whereby patients who are destined to have a milder course of illness or who have active caregivers are the ones who *can* take these medications, so that the apparent long-term effect compared to historical controls is illusory. These observational studies have not been supported by a recent, long-term, placebo-controlled trial.

Discontinuation of cholinesterase inhibitors occasionally has been associated with worsening cognition and occasional episodes of confusion in some patients. It is generally recommended that medication be tapered before discontinuing. In one trial in which donepezil was tapered and discontinued or continued, in moderate to severe outpatients who had been treated for at least 3 months, continuing donepezil was associated with slightly better cognitive and functional scores over the course of a year.

MEMANTINE: CLINICAL STUDIES, DOSING, AND ADVERSE EFFECTS

Overstimulation of NMDA receptors has been implicated as a causative mechanism in Alzheimer disease dementia. In preclinical studies, memantine, a moderate-affinity, uncompetitive NMDA receptor antagonist, protects against the effects of glutamate and calcium-mediated neurotoxicity. It has been available in Germany for the treatment of spasticity, Parkinson disease, and other neurologic disorders since 1982 and was approved by the European Union in

2002 and the United States in 2003 for treating moderately severe to severe Alzheimer disease.

Mechanism of Action

Memantine's action is hypothesized as being an open channel blocker without appreciable activity until increased glutamate levels trigger the receptor and cause the ion channel to open. It then enters the channel, occupying a site within it, preventing calcium influx and the depolarization and hyperactivation of the neuron. Because it is cleared from its site of action quickly, it might provide partial blocking of the channel, depending on intrasynaptic levels of glutamate. As glutamate increases, causing channels to open, memantine may block more, and as glutamate decreases, it may block less, allowing greater neuronal activity. In some respects, memantine can be viewed as a modulator of glutamatergic activity.

Clinical Studies in Moderate to Severe Alzheimer Disease

The effect of memantine on cognition generally has been mixed in clinical trials in moderate to severe Alzheimer disease (i.e., MMSE scores <15). Memantine (10 mg bid) was investigated for the treatment of moderate to severe Alzheimer disease in a 28-week, placebo-controlled trial of 252 patients. The decline of cognitive function was lessened in memantine patients compared to the deterioration observed in placebo.

Another 24-week, randomized, placebo-controlled trial of memantine in moderate to severe Alzheimer disease showed greater improvement in cognition, global assessment, and function compared to placebo. In this trial, memantine or placebo was administered to patients who had been maintained on stable doses of donepezil for at least 6 months. In fact, the average time on donepezil was approximately 2.5 years, with nearly 90% on it for more than 1 year before being randomized. Notably, the placebo-treated patients (i.e., patients who continued on donepezil throughout the 6 months of the trial) continued to decline at a rather rapid rate, suggesting that memantine substantially stabilized symptoms over this time, whereas any effect of the cholinesterase inhibitor may have been lost earlier over the disease course. Based on these two trials, memantine received marketing approval in the United States.

A third moderate to severe Alzheimer disease trial, however, which was similar in design to the other 6-month trials, failed to show significant effects for memantine compared to placebo.

A fourth moderate to severe Alzheimer disease trial failed to show that memantine was more effective than placebo over the course of a year when given to patients being maintained on donepezil. A fifth clinical trial, also in moderate to severe Alzheimer disease, did not find an overall clinical effect for memantine compared to placebo over the longer course of 2 years on average, again in patients maintained on cholinesterase inhibitors, while by contrast, it showed positive effects on function for vitamin E.

In a 12-week, randomized, placebo-controlled trial of memantine 10 mg per day in nursing home patients with Alzheimer disease or vascular dementia, memantine showed significant improvement over placebo in global assessment and activities of daily living.

Using the higher dose formulation, memantine, 28 mg, and intended for once-per-day administration showed efficacy over 24 weeks. The 28-mg-per-day dose has not been directly compared to the 10 mg bid dose in a single trial. The placebo-controlled trial of 28 mg per day did not show greater effects than the previous 10 mg bid placebo-controlled trials.

Clinical Studies in Mild to Moderate Alzheimer Disease

Two of three mild to moderate Alzheimer disease trials (i.e., MMSE scores from 10 to 23), one with patients taking cholinesterase inhibitors, did not show significant results, and a third showed significance only with one analysis. Based on these results, the FDA did not approve memantine for mild Alzheimer disease.

Pharmacokinetics

Memantine absorption is not affected by food intake; bioavailability is 100%. Plasma protein binding is 45%, and peak plasma concentrations occur 3 to 7 hours after oral intake. It is widely distributed and rapidly diffuses across the blood–brain barrier. The majority of the parent compound is excreted unchanged in the urine, in part by tubular secretion. There is minimum hepatic metabolism; the elimination half-life is 60 to 80 hours.

The 28-mg formulation, an extended-release capsule, shows a C_{max} and 24-hour Area under the Curve (AUC) that is 48% and 33% higher than the original memantine formulation, with an elimination half-life from 60 to 80 hours. The extended-release formulation is also available in combination with donepezil.

Drug Interactions

Interactions with memantine are few. There are potential interactions with hydrochlorothiazide, resulting in a lower plasma level of the diuretic or a potentially reduced effect. There are no known interactions with cholinesterase inhibitors. Drugs that alkalinize urine would reduce the renal clearance of memantine.

Dosing

Dosing is fairly straightforward, with titration to 10 mg bid over the course of 3 weeks, starting at 5 mg qd and increasing the daily dose by 5 mg each week; or with 7-mg capsules qd, and increasing by 7 mg qd each week to a dose of 29 mg qd.

Adverse Events

Interestingly, agitation occurred less often in memantine treatment groups than in placebo groups. In the memantine add-on trial, confusion, generally rated as mild, occurred four times more frequently in the memantine group, and headache twice as frequently. Of note, gastrointestinal symptoms were less frequent, with diarrhea or fecal incontinence occurring half as frequently in the memantine group, suggesting that the actions of memantine may mitigate some of the cholinergic effects of donepezil.

Clinical Application and Treatment

Memantine is indicated for moderate to severe dementia of the Alzheimer disease type, which generally corresponds to patients with MMSE scores of 14 or less. It generally can be considered as initial treatment for this group if patients have not been treated previously. If they have, then memantine can be added to a cholinesterase inhibitor as well. Because the clinical trials that showed efficacy were no longer than 6 months, treatment duration is unclear, and issues are similar to those of the cholinesterase inhibitors. Whether discontinuation is associated with worsening symptoms is not known.

Although memantine is commonly used with or without a cholinesterase inhibitor in mild Alzheimer disease, clinical trials assessing efficacy in mild to moderate Alzheimer disease have not demonstrated cognitive efficacy for memantine. This is due to a lack of efficacy observed in the mild subgroup of patients in these trials, and memantine is not FDA approved for mild Alzheimer disease. Nevertheless, it is commonly used in mild Alzheimer disease, with nearly half of mild Alzheimer disease patients in observational research studies receiving it.

OTHER MEDICATIONS AND APPROACHES TO ALZHEIMER DISEASE AND COGNITIVE ENHANCEMENT

Although cholinesterase inhibitors have been the most studied class of medications for Alzheimer disease, a broad range of other drugs and supplements are available for prescribing in some countries or are undergoing clinical trials. Some of these may be approved for marketing in the United States, but for other indications. These approaches might improve symptoms or slow neurodegenerative processes, or both.

Atypical inflammatory responses, the production of oxidizing free radicals, and the production from amyloid precursor protein of insoluble and possibly toxic β-amyloid peptide fragments are part of the neurodegenerative process. Therefore, anti-inflammatories and antioxidants might have a role in the treatment of progressive dementias. The neurotrophic and cholinergic properties of estrogens were hypothesized to be useful in women.

Huperzine A

Huperzine A is a plant extract that has selective acetylcholinesterase inhibition. It is sold as a food supplement and is a traditional Chinese treatment. A phase II trial failed to demonstrate efficacy at a lower dose of 200 mcg bid, but with a suggestion of a cognitive effect at 400 mcg bid.

Monoamine Oxidase Type B Inhibitors

Some positive effects have been reported with selegiline, a monoamine oxidase (MAO) inhibitor that relatively selectively inhibits MAO type B at a 5- to 10-mg daily dose. At this dose, its overall effect may be to increase CNS levels of dopamine and trace neurotransmitters, such as phenylethylamine, without affecting norepinephrine levels. Selegiline is marketed for the treatment of Parkinson disease, for which it has demonstrated effects in maintaining motor function and, in some Eastern European countries, as a cognitive enhancer.

From a different perspective, selegiline and other MAO type B inhibitors may have neuroprotective potential. Elevated levels of MAO type B in the brains of Alzheimer disease patients might lead to elevated levels of neurotoxins, such as 6-hydroxydopamine, quinones, and free radicals. It has also been postulated that exposure to substances in the environment could lead to neurotoxic products (via oxidative biotransformation). In either case, inhibition of MAO on a chronic basis could theoretically retard the progression of the illness based on antitoxic mechanisms. Also, selegiline may have neuronal “rescue” effects based on neurotrophic-like properties not involving MAO inhibition. It is notable that rodents and dogs treated with selegiline have increased survival. Selegiline is also marketed in the United States as an FDA-approved veterinary medicine for “canine cognitive dysfunction.”

A 12-month placebo-controlled trial of transdermal selegiline in patients with milder Alzheimer disease revealed no significant effects overall. Oral selegiline, 5 mg twice a day, along with and separately from vitamin E, 1,000 IU twice a day, has been associated with prolonged maintenance of activities of daily living and avoidance of nursing home placement in moderately to severely impaired Alzheimer disease patients, but with no improvement in cognitive function. An adequate efficacy trial has yet to be completed in Alzheimer disease.

Another recent, 12-month, phase II, placebo-controlled trial of the MAO-I, RO4602522 (EVT 302), did not show efficacy in 495 patients with moderate Alzheimer disease.

Ladostigil is a derivative of rasagiline and rivastigmine. It acts as a cholinesterase inhibitor and MAO-I; it has antioxidant properties and modulates the processing of amyloid precursor protein. It was assessed in a mild to moderate Alzheimer disease trial and in a phase II trial in MCI. In both trials, ladostigil missed its primary endpoints. In sum, there is little evidence for the efficacy of MAO type B inhibitors.

Ergoloid Mesylates

Ergoloid mesylate was a commonly used putative cognitive-enhancing drug until the advent of the cholinesterase inhibitors and is no longer marketed in the United States. It was used primarily for patients with dementia, age-associated cognitive symptoms, or memory complaints. The agent has mixed and complex actions on central α-adrenergic, dopaminergic, and serotonergic receptors.

Nootropics

The term *nootropic* is derived from the Greek *noos* (mind) and *troepin* (forward) and refers to a drug class with diverse actions structurally related to piracetam. They are pyrrolidine derivatives of γ-aminobutyric acid (GABA), although they do not appear to have GABA-like effects and are postulated to have a neuroprotective effect on the CNS against hypoxia, electroconvulsive treatment, and drug intoxication. They also may enhance the CNS microcirculation by reducing platelet activity and adherence of red blood cells to vessel walls and may stimulate central cholinergic activity. They include piracetam, oxiracetam, pramiracetam, and aniracetam. They may improve memory and learning, but an antidementia mechanism has not been established.

Clinical trials of nootropics in Alzheimer disease have shown mixed results with piracetam in the treatment of dementia in older adults. Pramiracetam and oxiracetam have been evaluated in large-scale multicenter studies with no sufficiently significant clinical effects in dementia or in MCI. Nootropics, however, are marketed for the treatment of dementia and cognitive impairment in countries other than the United States. There is some evidence for efficacy in attention-deficit disorder.

Ginkgo Biloba

The ability of an extract from the leaves of *Ginkgo biloba* to improve cognitive function in Alzheimer disease patients is the subject of ongoing work. Preclinical evidence suggests that certain flavonoid components in *Ginkgo* might have neuroprotective effects, including scavenging of oxygen free radicals; protection of membranes against lipid peroxidation; inhibition of β-amyloid formation, aggregation, and toxicity; and antiapoptotic effects, and may increase hippocampal muscarinic receptors and enhance choline uptake.

A preparation (EGb 761) is approved in Germany for "disturbances of cerebral function related to dementia syndromes," supported by various clinical studies. However, a clear role for *Gingko,* as a symptomatic agent or in the prevention of Alzheimer disease has not been determined. One 52-week placebo-controlled, double-blind trial of 309 patients with Alzheimer disease or vascular dementia demonstrated that 120 mg per day of a standardized *Ginkgo* extract (EGb 761) conferred a statistically significant improvement on the ADAS-cog scores compared to placebo. A 6-month placebo-controlled trial of 120 and 240 mg per day of EGb 761 for mild to moderate Alzheimer disease, however, did not show efficacy. A Cochrane meta-analysis did not support the advantageous effects of

Ginkgo biloba extract for Alzheimer disease or as a cognitive enhancer. A randomized, placebo-controlled trial (the Gingko Evaluation of Memory Study [GEMS]) funded by the National Center for Complementary and Alternative Medication evaluated *Ginkgo biloba* EGb 761 extract, 240 mg per day, in delaying the onset of Alzheimer disease or MCI in 3,072 subjects aged 75 years and older who were followed over 5 years and found no effect. Another prevention study in 200 subjects 85 years or older, also sponsored by the National Center for Complementary and Alternative Medication, did not find an effect for 240 mg per day of EGb 761 in delaying progression to MCI. Finally, a multicenter prevention trial in France involving 2,854 patients followed for 5 years did not show a positive effect for EGb 761 for delaying dementia or improving cognition.

Based on these studies, there is little or no evidence to recommend *Ginkgo biloba* for the treatment of Alzheimer disease or vascular dementia. Studies in normal older adult subjects have been inconsistent, presenting a confusing picture for its use, perhaps partly owing to variations in dose, duration, and methodologic issues such as lack of blinding.

Vitamin E

The antioxidant effects of vitamin E (DL-α-tocopherol) might have an impact on reducing clinical progression. In the prototypical antioxidant trial, vitamin E (1,000 IU twice a day) and selegiline (5 mg twice a day), each given alone to moderately impaired patients with Alzheimer disease, delayed the time until patients required nursing home placement died or lost activities of daily living. Overall, the delayed time to these end points was approximately 7 months compared to the placebo. There were no effects on cognition in this group of patients, who were more severely impaired than in other studies. Adverse effects included falls and syncope in selegiline patients (9% and 10%, respectively), α-tocopherol patients (14% and 7%, respectively), and the combination (22% and 16%, respectively).

A multicenter trial in amnestic MCI patients reported no significant delay in conversion to Alzheimer disease or in cognitive effects with α-tocopherol over 3 years of treatment. The outcomes of the previous moderate to severe Alzheimer disease trial were supported by a trial comparing vitamin E with memantine and placebo over 2 years on average and showing improved function with vitamin E and not with memantine.

Taken together, considering the absence of cognitive efficacy and the potential risks for treatment, vitamin E cannot be recommended as a treatment.

B Vitamins and Homocysteine Reduction

Elevated levels of homocysteine are a risk factor for cardiovascular disease, vascular dementia, and Alzheimer disease. Plasma concentrations rise when homocysteine's metabolism to methionine or cysteine is impaired, which may occur under a number of circumstances, including natural aging. Increased blood levels of homocysteine are associated with damage to vascular endothelial cells through lipid peroxidation and release of endothelium-derived relaxing factor and may increase the risk of vascular thrombosis. More relevant to Alzheimer disease, homocysteine potentiates the neurotoxic effects of β-amyloid peptides in vitro. Homocysteine plasma levels can be reduced by as much as 30% using regimens of B vitamins (e.g., B_{12} 1 mg per day, B_6 25 mg per day, and folic acid 5 mg per day). A placebo-controlled clinical trial of these vitamins over 18 months in patients with mild to moderate Alzheimer disease confirmed that a substantial reduction in homocysteine could be achieved but did not show cognitive or functional effects compared to placebo.

HMG-CoA Reductase Inhibitors

Epidemiologic evidence suggests that cholesterol lowering and the use of 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors ("statins") may impact the development or progression of Alzheimer disease, possibly because cholesterol levels modulate the enzymatic processing of amyloid precursor protein and consequently the production of β-amyloid, a main component of neuritic plaques in the Alzheimer disease brain. Furthermore, cholesterol-fed rabbits demonstrate increased β-amyloid in the brain, and hypercholesterolemia induced by a high-cholesterol diet results in increased levels of β-amyloid in the CNS. Statin administration in vitro reduces intra- and extracellular β-amyloid levels in neuronal cultures.

Observational studies of patient records demonstrated a 60% to 73% lower prevalence of Alzheimer disease in lovastatin (Mevacor)- or pravastatin (Pravachol)-treated patients and a markedly reduced risk for Alzheimer disease as assessed in an analysis of a clinical database.

Clinical trials, however, generally have not supported the use of statins for Alzheimer disease. A randomized, placebo-controlled trial of pravastatin (the PROSPER study) in 5,804 subjects aged 70 to 82 years with risk factors for vascular disease reported favorable cardiovascular outcomes but no significant effect on cognitive function or disability during an approximately 3-year follow-up period. In a randomized, placebo-controlled trial of the effect of simvastatin (Zocor) on cardiovascular outcomes in a subgroup of 5,806 patients aged 70 or more years, there was no significant difference in the proportion of patients with cognitive impairment or in the incidence of dementia over the average 5-year follow-up period.

A small-scale trial of atorvastatin (Lipitor) indicated some positive cognitive effects. Two larger placebo-controlled trials, one with simvastatin and another with atorvastatin, including 400 and 641 patients, respectively, treated over 18 months, did not show positive effects for the drugs in symptoms or in slowing the clinical decline of Alzheimer disease patients.

More recently, case reports have brought attention to patient reports of the possible adverse effect of statin use on cognition. However, a recent systematic review and meta-analysis of observational studies appeared to confirm the absence of a neurocognitive risk associated with statin treatment and suggested a potential favorable role of statins in reducing the risk of dementia. Randomized clinical trials with an ad hoc design are needed to explore this potential neuroprotective effect. However, at this time, these drugs should not be considered useful for preventing or treating Alzheimer disease.

Docosahexaenoic Acid

A reduced risk of Alzheimer disease associated with fish consumption in epidemiologic studies is also associated with omega-3 fatty acids. Docosahexaenoic acid (DHA) is an omega-3 fatty acid found in fish oil, a component of neural membrane phospholipids. Several animal studies show DHA to have antiamyloid, antioxidant, and neuroprotective mechanisms. A carefully conducted multicenter, randomized, placebo-controlled trial of DHA in patients with mild to moderate Alzheimer disease did not show efficacy for DHA.

Souvenaid

Souvenaid (Nutricia N.V.) is marketed outside the United States as a medical food. It contains uridine, choline, phospholipids, eicosapentaenoic acid, DHA, selenium, and vitamins E, C, B_{12}, and B_6; folic acid. The rationale is that these are precursors and cofactors for the development of neuronal membranes and are intended to support the formation of synapses. This represents both a structure/function food supplement claim and a medical food claim. Clinical trials in Alzheimer disease, however, have not been supportive of a clinical or functional effect.

Axona

A particular liquid formulation of medium-chain fatty triglycerides composed of glycerin and caprylic acid (AC-1204 or Axona) is intended to create chronic ketosis in an effort to enhance mitochondrial metabolism. A rationale is that a reduction in cerebral glucose uptake and utilization is observed in Alzheimer disease. Therefore, the AC-1204 formulation is intended to improve cellular metabolism by providing an alternative energy substrate to glucose. Caprylic acid is metabolized into the ketone bodies acetoacetic acid and β-hydroxybutyric acid, which in turn are converted to acetyl-CoA and energy by the tricarboxylic acid (TCA or Krebs) cycle.

The formulation was initially marketed as a medical food for Alzheimer disease and then withdrawn by the FDA because of absent clinical trials efficacy evidence and claims. A subsequent phase III trial to assess potential efficacy for its use as a drug rather than as a medical food also failed.

UNPROVEN AND UNSAFE MEDICATIONS TO SLOW CLINICAL PROGRESSION OR DELAY ONSET OF ALZHEIMER DISEASE

Anti-inflammatories

Markers of inflammation in the brains of Alzheimer disease patients suggest a potential therapeutic benefit of anti-inflammatories. Epidemiologic studies with nonsteroidal anti-inflammatory drugs (NSAIDs) indicate that these agents may be potentially effective for reducing the incidence or delaying the onset of Alzheimer disease.

Nonsteroidal Anti-Inflammatory Drugs and Alzheimer Disease. For example, (1) patients with rheumatoid arthritis who receive chronic anti-inflammatory treatment have a reduced risk for Alzheimer disease compared to controls without rheumatoid arthritis, (2) twins who used anti-inflammatory medications had a lesser risk for Alzheimer disease than did their co-twins, and (3) the risk of Alzheimer disease was reduced in cohorts of people using anti-inflammatories in middle age years before the typical onset of Alzheimer disease.

A placebo-controlled trial of indomethacin (Indocin) initially supported a potential role for at least NSAIDs. Subsequent randomized, placebo-controlled trials of NSAIDs in Alzheimer disease patients have failed to show a clinical benefit. Moreover, 1-year-long, placebo-controlled trials of low-dose prednisone (10 mg) or NSAIDs by the National Institute of Aging Alzheimer's Disease Cooperative Study and long-term trials of celecoxib (Celebrex) and rofecoxib (Vioxx) did not show advantageous effects.

To investigate the effect of NSAIDs in delaying the onset of Alzheimer disease, an ongoing multisite trial (the Alzheimer's Disease Anti-Inflammatory Prevention Trial) randomized approximately 2,528 patients aged 70 years or older with a family history of dementia to naproxen (Aleve), celecoxib, or placebo, and followed them for 5 to 7 years before the trial was ended early due to cardiovascular toxicity concerns. Results, however, indicated nontherapeutic and cognitively toxic effects with increased risk for Alzheimer disease with both treatments.

In view of the negative clinical trials with NSAIDs and considering their frequent adverse effects, including gastrointestinal bleeding, especially with long-term use, NSAIDs cannot be recommended for the prophylaxis or treatment of Alzheimer disease.

Estrogens

Preclinical evidence amply demonstrates that estrogens improve cognitive function through cholinergic neuroprotective and neurotrophic effects. For example, estradiol (Estrace) replacement increases choline acetyltransferase and acetylcholine after ovariectomy and enhances learning in ovariectomized rats. Clinical evidence of a possible effect of estrogen replacement therapy (ERT) has included observations of an inverse relationship between ERT and death certificate diagnoses of dementia, several case-control and cohort studies in which 40% to 50% reductions in the risk for developing Alzheimer disease were reported, and several smaller clinical reports suggesting that estradiol, estrone (Estrol), or conjugated equine estrogens (CEE, Premarin) enhance cognitive function in Alzheimer disease.

Acute adverse events with ERT are frequent and generally dose dependent and include endometrial hyperplasia, uterine bleeding (when taken with a progestin), breast tenderness, nausea, and venous thrombosis, among others. Because of the enhanced risk for endometrial cancer when estrogens are given to women with intact uteri, administration of progestins with estrogens has been recommended for women with intact uteri.

During the last few years, a number of clinical trials have been carried out, and except for one very small trial using transdermal estradiol, results with CEE (without a progestin) did not substantiate their efficacy. A 20-week-long and a 1-year-long trial of CEE in hysterectomized women were negative. Indeed, there was a trend in women receiving hormone replacement to worsen cognitively and—although a relatively small sample size—have greater adverse events, including a few instances of deep vein thrombosis.

Without evidence for efficacy, then safety is of great concern. The Women’s Health Initiative trial followed a large number of women randomly assigned to hormone replacement therapy (CEE with or without the progestin medroxyprogesterone [Provera or PremPro when combined with CEE]) or placebo to assess potential cardioprotective effects of the hormone replacement. In 2002, the arm of the trial that included combination treatment was halted because of a substantial increase in cardiovascular adverse effects. The results of the cognitive outcomes in 4,532 women in the subgroup older than 65 years were reported in 2003, finding that the CEE/medroxyprogesterone-treated women had an increased risk for developing dementia. In 2004, the arm of the trial that included CEE treatment alone also was halted for lack of effect and increased strokes with estrogens. CEE alone was associated with an increased incidence of dementia and worse cognition over the approximately 5-year follow-up period.

It is possible but not at all proven that estrogen treatment in midlife around menopause could be neurotrophic and preventative of the onset of Alzheimer disease, but beyond epidemiologic studies, evidence is lacking. The lack of any salutary cognitive effects, combined with the substantial cardiovascular, stroke, and breast cancer risk over the long term, strongly suggests that there is no role for the use of combined estrogen/progestin or CEE-alone intervention in women over 65 years of age.

ADJUNCTIVE AND COMBINATION THERAPY

It is not known whether combining medications such as cholinesterase inhibitors with other drugs or vitamins confers added efficacy. There have only been pilot studies comparing combinations of cholinesterase inhibitors with such drugs as selegiline, suggesting only their feasibility and not efficacy. The combination of selegiline and vitamin E did not confer any added benefits over either alone in one study and was associated with greater adverse effects.

The combination of a cholinesterase inhibitor with memantine is another example of the potential use of combined therapy. Although such issues as what drug to use first or whether to start both simultaneously have not been worked out, nor is there evidence that each drug separately has an effect that the other does not, it appears that the medications can be used together safely.

A recent network meta-analysis of memantine plus donepezil showed superior outcomes for cognition, global assessment, daily activities, and neuropsychiatric symptoms, but lower acceptability than monotherapy and placebo.

TREATMENT OF MILD COGNITIVE IMPAIRMENT

Operational criteria for MCI are generally taken to include a subjective cognitive (memory) complaint; MMSE within normal limits; neuropsychological tests for memory greater than 1.5 standard deviations below age-appropriate norms; activities of daily living within normal limits; and Clinical Dementia Rating Scale score of 0.5, indicating no dementia. Criteria for MCI due to Alzheimer disease and used for clinical trials, however, are slightly different, requiring as well progressive worsening and a delayed memory deficit on the Logical Memory II delayed recall below 9 in most cases.

In research clinics, a diagnosis of “amnestic mild cognitive impairment” or “mild cognitive impairment of the amnestic type,” marked by predominant memory impairment, implies a conversion rate to Alzheimer disease of about 12% to 15% per year. However, MCI is not necessarily a predementia stage of Alzheimer disease because many people who fulfill the criteria do not progress to dementia, and some improve to not being cognitively impaired. The meaning of MCI is further confounded as it is also viewed as a transitional state on the road to the development of Alzheimer disease, as an “at-risk” state with an increased likelihood for dementia over a specified time, and as the very early diagnosis of Alzheimer disease before the full dementia syndrome is recognized.

Although criteria for MCI can be defined by consensus, the diagnosis as variously used suffers from vagueness in conceptualization, with the condition overlapping with Alzheimer disease. The application of MCI criteria seems to vary, however, and the diagnosis does not necessarily predict the onset of Alzheimer disease. For example, although all MCI trials nominally used the Mayo Clinic criteria for MCI, when the criteria from three clinical trials were applied to the same clinic cohort, differing prevalences were found from 17% to 51%. For clinical diagnostic purposes, MCI has broadly been defined as mild neurocognitive impairment in DSM-5.

MCI has been a therapeutic target for several clinical trials of cholinesterase inhibitor medications marketed for Alzheimer disease. These trials have ranged from 2 to 4 years in duration. None showed positive effects for the cholinesterase inhibitor in delaying time to dementia onset or improving cognitive scores beyond 6 to 12 months. Given the vagaries of the definition of MCI, that many people with MCI will not progress, the adverse effects of the medications, and the lack of efficacy in the trials, the use of cholinesterase inhibitors for people with MCI cannot be recommended.

A development for clinical research and trials is to define MCI “due to Alzheimer’s disease” or the equivalent “prodromal Alzheimer’s disease” by the above clinical and cognitive criteria but to also require a positive amyloid-β PET scan or low amyloid-β CSF levels to enhance the likelihood that the research participant will progress to Alzheimer disease.

TREATMENT OF VASCULAR DEMENTIA

Although a range of treatments has been proposed for the cognitive deficits of ischemic vascular dementia, more recently conducted trials using particular research criteria for vascular dementia have yielded the best efficacy evidence. Six-month trials of the cholinesterase inhibitors donepezil and galantamine have generally demonstrated very modest cognitive efficacy similar to that shown in Alzheimer disease but often without evidence of global improvement or improvement in activities of daily living. Similarly, two 6-month multicenter trials of memantine also tended to show improvements in cognition. These trials relied on a cognitive battery designed for patients with Alzheimer disease.

Overall there is insufficient evidence for the efficacy of cholinesterase inhibitors or memantine in vascular dementia. One limitation is that vascular dementia is a heterogeneous condition that is difficult to define clinically and often overlaps with Alzheimer disease. These drugs are not approved for use in vascular dementia. The results of the trials demonstrate that cognitively enhancing medications might be useful in other conditions besides Alzheimer disease.

TREATMENT OF PARKINSON DISEASE DEMENTIA

Patients with Parkinson disease have an increased risk for dementia, and about 25% may develop dementia after the onset of the illness. The dementia, occurring after the onset of parkinsonian motor symptoms, can be characterized by a particular pattern of impairments in executive function, memory, and attention that is different from other dementias such as Alzheimer disease.

From an FDA perspective, however, a diagnosis of Parkinson disease dementia is made in patients in whom “a progressive dementia syndrome occurs at least 2 years after a diagnosis of Parkinson’s disease” in the absence of other causes of the dementia. This definition came about as a result of the FDA approving the cholinesterase inhibitor rivastigmine for “mild to moderate dementia associated with Parkinson’s disease.” Thus, rivastigmine used in the same dosing range as with Alzheimer disease is the only drug indicated for treating the dementia.

β-Amyloid–Based Interventions

Aducanumab is an amyloid-β–directed antibody, which was approved by the FDA in 2021 for the treatment of Alzheimer disease through their Accelerated Approval pathway. This pathway allows for the approval of treatments for serious or life-threatening illnesses where there is the expectation of clinical benefit despite some uncertainty and when the therapy may provide an advantage over existing treatments. This approval was based on a surrogate endpoint of the reduction of β-amyloid plaque, as there was conflicting data in two phase III clinical trials surrounding the reduction of clinical decline. The treatment will now need a postapproval trial demonstrating clinical benefit. After an initial titration, aducanumab is administered as an intravenous infusion every 4 weeks at a maintenance dose of 10 mg/kg. This is the first disease-modifying therapy approved by the FDA for Alzheimer disease and is the first medication approved by the FDA for Alzheimer disease since 2003. Intravenous lecanemab (Leqembi) was also granted accelerated approval for the treatment of Alzheimer disease in January 2023, after a phase 3 trial demonstrated a reduction in brain amyloid levels and less cognitive decline. It was granted a traditional approval after a confirmatory trial verified clinical benefit. It is recommended that it should be used in patients with mild cognitive impairment or mild dementia, the population in which treatment was studied in clinical trials. Due to concerns about developing amyloid-related imaging abnormalities or ARIA, a type of brain swelling or bleeding brain, imaging monitoring is recommended for both these agents.

Other drugs with a broad range of mechanisms are in development. They include β- and γ-secretase inhibitors that prevent the cleavage of amyloid precursor protein into insoluble β-amyloid; vaccines such as synthetic β-amyloid in order to generate β-amyloid antibodies; the use of other manufactured antibodies against β-amyloid blocking the receptor for advanced glycation endproducts (RAGE) from binding amyloid; preventing amyloid fibril formation; and a range of other “antiamyloid” approaches.

Moreover, there is a new focus on the development of neuroprotective medications or drugs that might cause neurogenesis or enhance plasticity; and drugs and antibodies directed against tau-protein targets. Research and development efforts, of course, continue.

FUTURE DIRECTIONS

Current pharmacologic treatment options for improving cognition in Alzheimer disease include cholinesterase inhibitors and memantine. Other available drugs, food supplements, and vitamins, including anti-inflammatories, available estrogens, B-complex vitamins, vitamin E, medium-chain fatty acids, and *Ginkgo biloba* extract, have not been proven efficacious. Some show considerable toxicity in clinical trials and cannot be recommended for use.

Ideally, it would be best to treat Alzheimer disease early before the onset of dementia; however, there are no current drugs for this, and cholinesterase inhibitors have been demonstrated to be ineffective in this regard.

Although no currently available antidementia treatment has shown dramatic effects in improving the cognitive impairments of Alzheimer disease, the improvements nonetheless may be clinically meaningful for some people. Therapeutic results are usually modest, affecting a minority of patients. Cholinesterase inhibitors are the best proven efficacious symptomatic treatments for Alzheimer disease. Patients assessed and treated in clinical trials, however, were usually outpatients with mild to moderately severe dementia and few concomitant medical illnesses. Duration of effect beyond 6 months to 1 year (2 years in one trial) and long-term safety is not known.

The efficacy of these drugs for other cognitive disorders remains to be established, although the FDA has approved rivastigmine for Parkinson disease dementia. Disease-Modifying Therapy and Prevention trials are ongoing. Future treatments with greater efficacy in Alzheimer disease and for cognitive enhancement in older adult patients will be dependent on a greater understanding of possible neurobiologic underpinnings of cognition, brain imaging, and the various neurocognitive disorders. The challenge will be to develop treatments that have a reasonable benefit-to-risk ratio.

ACKNOWLEDGMENTS

This chapter represents an update and revision of previous chapters in this series, most recently by Lon S. Schneider, M.D., from the last edition. The author is indebted to Dr. Schneider for his work and retains much of his prose in this current edition.

FURTHER READINGS

ADAPT Research Group; Lyketsos CG, Breitner JCS, Green RC, et al. Naproxen and celecoxib do not prevent AD in early results from a randomized controlled trial. *Neurology*. 2007;68:1800–1808.

Aisen PS, Cummings J, Schneider LS. Symptomatic and nonamyloid/tau based pharmacologic treatment for Alzheimer disease. *Cold Spring Harb Perspect Med*. 2012;2:a006395.

*Birks J. Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Database Syst Rev*. 2006;1:CD005593.

*Birks J, Grimley Evans J, Iakovidou V, Tsolaki M. Rivastigmine for Alzheimer's disease. *Cochrane Database Syst Rev*. 2000;(4):CD001191.

Birks J, Grimley Evans J. Ginkgo biloba for cognitive impairment and dementia. *Cochrane Database Syst Rev*. 2007;(2):CD003120.

*Birks JS, Harvey RJ. Donepezil for dementia due to Alzheimer's disease. *Cochrane Database Syst Rev*. 2006;(1):CD001190.

*Courtney C, Farrell D, Gray A, et al; AD2000 Collaborative Group. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): randomised double-blind trial. *Lancet*. 2004;363(9427):2105–2115.

Cummings J, Aisen P, Apostolova LG, Atri A, Salloway S, Weiner M. Aducanumab: appropriate use recommendations. *J Prev Alzheimers Dis*. 2021;8(4):398–410.

DeKosky ST, Williamson JD, Fitzpatrick AL, et al; Ginkgo Evaluation of Memory (GEM) Study Investigators. Ginkgo biloba for prevention of dementia: a randomized controlled trial. *JAMA*. 2008;300:2253–2262.

Dysken MW, Sano M, Asthana S, et al. Effect of vitamin E and memantine on functional decline in Alzheimer disease: the Team-AD VA cooperative randomized trial. *JAMA*. 2014;311:33–44.

Emre M, Aarsland D, Albanese A, et al. Rivastigmine for dementia associated with Parkinson's disease. *N Engl J Med*. 2004;351(24):2509–2518.

Emre M, Aarsland D, Brown R, et al. Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov Disord*. 2007;22(12):1689–1707; quiz 1837.

Farlow MR, Salloway S, Tariot PN, et al. Effectiveness and tolerability of high-dose (23 mg/d) versus standard-dose (10 mg/d) donepezil in moderate to severe Alzheimer's disease: a 24-week, randomized, double-blind study. *Clin Ther*. 2010;32:1234–1251.

Feldman HH, Ferris S, Winblad B, et al. Effect of rivastigmine on delay to diagnosis of Alzheimer's disease from mild cognitive impairment: the InDDEx study. *Lancet Neurol*. 2007;6(6):501–512.

Gilman S, Koller M, Black RS, et al; AN1792(QS-21)-201 Study Team. Clinical effects of Abeta immunization (AN1792) in patients with AD in an interrupted trial. *Neurology*. 2005;64(9):1553–1562.

Grossberg G, Manes F, Allegri RF, et al. The safety, tolerability, and efficacy of once-daily memantine (28 mg): a multinational, randomized, double-blind, placebo-controlled trial in patients with moderate-to-severe Alzheimer's disease taking cholinesterase inhibitors. *CNS Drugs*. 2013;27:469–478.

Guo J, Wang Z, Liu R, Huang Y, Zhang N, Zhang R. Memantine, donepezil, or combination therapy-what is the best therapy for Alzheimer's disease? A network meta-analysis. *Brain Behav*. 2020;10(11):e01831.

Henderson ST, Vogel JL, Barr LJ, Garvin F, Jones JJ, Costantini LC. Study of the ketogenic agent AC-1202 in mild to moderate Alzheimer's disease: a randomized, double-blind, placebo-controlled, multicenter trial. *Nutr Metab (Lond)*. 2009;6:31.

Howard R, McShane R, Lindesay J, et al. Donepezil and memantine for moderate-to-severe Alzheimer's disease. *New Engl J Med*. 2012;366:893–903.

*Kavirajan H, Schneider LS. The efficacy and adverse effects of cholinesterase inhibitors and memantine in vascular dementia: a systematic review and meta-analysis of controlled trials. *Lancet Neurol*. 2007;6:782–792.

Knight R, Khondoker M, Magill N, Stewart R, Landau S. A systematic review and meta-analysis of the effectiveness of acetylcholinesterase inhibitors and memantine in treating the cognitive symptoms of dementia. *Dement Geriatr Cogn Disord*. 2018;45(3–4):131–151.

Kozauer N, Katz R. Regulatory innovation and drug development for early-stage Alzheimer's disease. *New Engl J Med*. 2013;368:1169–1171.

*Leber P. Criteria used by drug regulatory authorities. In: Qizilbash N, Schneider L, Chui H, eds. *Evidence-Based Dementia Practice*. Blackwell Science; 2002.

Loy C, Schneider L. Galantamine for Alzheimer's disease and mild cognitive impairment. *Cochrane Database Syst Rev*. 2006;1:CD001747.

McKeith I, Del Ser T, Spano P, et al. Efficacy of rivastigmine in dementia with Lewy bodies: a randomised, double-blind, placebo-controlled international study. *Lancet*. 2000;356:2031–2036.

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease. Report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology*. 1984;34:939–944.

*McShane R, Areosa Sastre A, Minakaran N. Memantine for dementia. *Cochrane Database Syst Rev*. 2006;(2):CD003154.

Olin J, Schneider L, Novit A, Luczak S. Hydergine for dementia. *Cochrane Database Syst Rev*. 2001;(2):CD000359.

Olmastroni E, Molari G, De Beni N, et al. Statin use and risk of dementia or Alzheimer’s disease: a systematic review and meta-analysis of observational studies. *Eur J Prev Cardiol*. 2022;29(5):804–814.

Petersen RC, Thomas RG, Grundman M, et al; Alzheimer’s Disease Cooperative Study Group. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med*. 2005;352(23):2379–2388.

*Qizilbash N, Schneider LS, Chui H, Tariot P, Brodaty H. *Evidence-Based Dementia Practice*. Blackwell Science; 2002.

Quinn JF, Raman R, Thomas RG, et al. Docosahexaenoic acid supplementation and cognitive decline in Alzheimer disease: a randomized trial. *JAMA*. 2010;304:1903–1911.

*Rabins PV, Blacker D, Rovner BW, Rummans T, Schneider LS. American Psychiatric Association Practice Guidelines Work Group on Alzheimer’s disease and other dementias. Practice guidelines for the treatment of patients with Alzheimer’s disease and other dementias, second edition. *Am J Psychiatry*. 2007;164(12):3.

Rafii MS, Walsh S, Little JT, et al; Alzheimer’s Disease Cooperative Study. A phase II trial of huperzine A in mild to moderate Alzheimer disease. *Neurology*. 2011;76:1389–1394.

Reisberg B, Doody R, Stöffler A, Schmitt F, Ferris S, Möbius HJ; Memantine Study Group. Memantine in moderate-to-severe Alzheimer’s disease. *N Engl J Med*. 2003;348(14):1333–1341.

Sano M, Ernesto C, Klauber MR, Schafer K, Woodbury P. A two-year, double-blind randomized multicenter trial of selegiline and α-tocopherol in the treatment of Alzheimer’s disease. *N Engl J Med*. 1997;336:1216.

*Schneider LS. Issues in design and conductance of clinical trials for cognitive-enhancing drugs. In: McArthur RA, Borsini F, eds. *Animal and Translational Models of Behavioural Disorders*. Vol. 2. Neurological Disorders. Elsevier; 2008.

Schneider LS, DeKosky ST, Farlow MR, Tariot PN, Hoerr R, Kieser M. A randomized, double-blind, placebo-controlled trial of two doses of Ginkgo biloba extract in dementia of the Alzheimer’s type. *Curr Alzheimer Res*. 2005;2:541–551.

Schneider LS, Insel PS, Weiner MW; Alzheimer’s Disease Neuroimaging Initiative. Treatment with cholinesterase inhibitors and memantine of patients in the Alzheimer’s Disease Neuroimaging Initiative. *Arch Neurol*. 2011;68:58–66.

Schneider LS, Mangialasche F, Andreasen N, et al. Clinical trials and late-stage drug development for Alzheimer’s disease: an appraisal from 1984 to 2014. *J Intern Med*. 2014;275:251–283.

Schneider LS. Treatment of Alzheimer’s disease with cholinesterase inhibitors. In: Galasko D, ed. *Clinics in Geriatric Medicine: Alzheimer’s Disease and Dementia*. WB Saunders; 2001:337.

Shah RC, Kamphuis PJ, Leurgans S, et al. The S-Connect study: results from a randomized, controlled trial of Souvenaid in mild-to-moderate Alzheimer’s disease. *Alzheimers Res Ther*. 2013;5:59.

Shumaker SA, Legault C, Kuller L, et al; Women's Health Initiative Memory Study. Conjugated equine estrogens and incidence for probable dementia and mild cognitive impairment in postmenopausal women: Women's Health Initiative Memory Study. *JAMA*. 2004;291:2947–2958.

Tariot PN, Farlow MR, Grossberg GT, Graham SM, McDonald S, Gergel I; Memantine Study Group. Memantine treatment in patients with moderate to severe Alzheimer's disease already receiving donepezil: a randomized controlled trial. *JAMA*. 2004;291:317–324.

Vellas B, Coley N, Ousset PJ, et al; GuidAge Study Group. Long-term use of standardised Ginkgo biloba extract for the prevention of Alzheimer's disease (GuidAge): a randomised placebo-controlled trial. *Lancet Neurol*. 2012;11:851–859.

55.4f Electroconvulsive Therapy and Neurostimulation Treatments

Mustafa M. Husain, M.D., Tarek K. Rajji, M.D., F.R.C.P.C., and Robert Garrett, M.D.

Chapters 34.20a and 34.20b comprehensively detail the application of electroconvulsive therapy (ECT) and other neurostimulation therapies, respectively. This particular section applies to the utility of ECT and other neurostimulation therapies in a geriatric population.

BRIEF HISTORY AND BACKGROUND

Convulsive therapy was introduced in the early part of the 20th century by Ladislas von Meduna as a treatment for severe catatonia in dementia praecox by use of camphor oil injections. In the late 1930s, Ugo Cerletti and Luigi Bini refined convulsive therapy by using electrical stimulation to induce therapeutic convulsions, and this formed the precursor of modern-day ECT. Due to its broad applications and, at times, misapplications to a variety of psychiatric diseases in the 1940s, as well as its associated cognitive adverse effects, ECT received substantial criticism from both inside and outside the psychiatric community. These critiques of ECT resulted in its limited use, but they also procured its scientific and systematic development as a clinically beneficial and safe treatment modality.

A report of older adult patients who received Medicare between the years of 1987 and 1992 found that the rate of ECT use increased each successive year. For example, from 1987 to 1992, the number of patients treated with ECT increased from 12,000 to 15,560. In the state of Texas, annual reports of ECT use showed that in 2007 a total of 1,772 patients received ECT, and of those, 22% were over the age of 65, with the oldest patient treated at age 94.

Unfortunately, the use of ECT has been declining in the United States. A study of general hospital stays for ECT for the period from 1993 to 2009 showed that the average number of stays declined from 12.6 to 7.2 per 100,000 adult U.S. residents. The percentage of general hospitals conducting ECT also decreased—from 15.8% to 10.6%. The decline for geriatric patients (over 65 years old) was significant. The number of inpatient stays for this population declined from 41.2 per 100,000 U.S. residents in 1995 to 12.3 per 100,000 in 2009. While this data may not reflect all ECT conducted in the United States, the trend is concerning, especially in light of the proven efficacy of ECT. This occurs in the backdrop of a rapidly growing U.S. population of adults 65 years old and older. In 2020 there were approximately 56.1 million in this age group, but this is expected to grow to 80.8 million in 2040.

In a survey of 306 geriatric psychiatrists in the United Kingdom, the majority felt that ECT is a well-tolerated and beneficial treatment for geriatric patients with severe psychiatric disorders, even those patients who are 80 years of age and older. On the other hand, a 2004 survey of practitioners in the Netherlands suggested that there is hesitation to utilize ECT in older adult patients as a first, second, or even as third option. While 84% of the 152 psychiatrists noted that the use of ECT should be broadened in older adult patients, approximately 54% agreed that it should be a first- or second-line treatment option. Equally important, only 15.1% and 13.9% indicated ECT as a first-line treatment for major depression with psychotic features or severe suicidal risk, respectively. Collectively, these surveys suggest that the practice of ECT is influenced by practitioner opinion and experience and that usage estimates vary per country.

PRECLINICAL STUDIES

Preclinical studies have shown that with ECT stimulation, there is increased hippocampal expression in the dentate gyrus of neurotropic and transcription factors, including the neurotropic factors: brain-derived neurotrophic factor (BDNF), vascular endothelial growth factor (VEGF), and fibroblast growth factor (FGF) and transcription factors, including c-Fos (which plays an important role in signal transduction, cell proliferation, and differentiation) and the arachidonic acid pathway (COX 2). There are cell proliferation and neurotropic effects of ECT also in the prefrontal cortex, amygdala, and hypothalamus. After several stimulations, there is a release of even more neuropeptide factors, including neuropeptide Y, thyrotropin-releasing hormone, and an increase in nerve growth factors. Chronic ECT works by changing gene expression. Thus, relapse after ECT is thought to be due to disrupted realteration of genes. Chronic ECT also helps strengthen second messenger systems. With the emergence of treatment-resistant depression, defined as depression that does not improve after adequate and substantial antidepressant treatment, ECT has found a new role in antidepressant algorithms.

CLINICAL INDICATIONS

ECT is one of the most effective somatic therapy for neuropsychiatric disorders. In the older adult, it is often the treatment of choice for severe depression. Depressed geriatric patients are especially at high risk for sequela of depression, including psychotic symptoms, malnutrition, weight loss, and inactivity. As a result, ECT is commonly employed in geriatric depression with the added benefit that it is relatively safe and well-tolerated compared to medications. ECT often becomes the first-line treatment recommendation for patients who are acutely suicidal, psychotic, or present with severe melancholia. Recent research studies suggest that among patients receiving ECT, older age is associated with rapid response and remission. The PRIDE 1 (Prolonging Remission in Depressed Elderly) study of ECT for geriatric depression included 240 patients in the acute phase. In this study, the response rate was 70%, the remission rate was 62%, and the mean time to remission was 2.5 weeks.

Relapse after the index course of ECT is of major clinical significance; it is thought to be due to disrupted realteration of genes. With the emergence of treatment-resistant depression, defined as depression that does not improve after adequate and substantial antidepressant treatment, ECT has found a new role in antidepressant algorithms.

Mrs. C is a 70-year-old Caucasian woman with a 2-year history of severe recurrent major depression and a three-and-a-half-year history of Parkinson disease, initially treated with Sinemet. She was referred by her outpatient psychiatrist for an ECT consultation. She was taking Remeron 45 mg nightly and Wellbutrin XL 150 mg daily. She had been switched from Sinemet to Ropinirole, which due to lack of effect, was discontinued in November 2014.

Prominent symptoms included: decreased self-care, decreased appetite with 20-pound weight loss within a year, decreased sleep, marked social withdrawal, spending 22 hours a day in bed, and decreased mobility. She was, per family, “alive, but not living,” and “her smile was not there anymore.” She stopped watching TV and preferred keeping the shades down. Mrs. C’s Parkinson disease was apparently masking her comorbid depression. She was initiated with ECT treatments. She received 11 acute course treatments over 6 weeks and her MADRS improved from 47 to 9 with only mild memory dysfunction, namely with concentration and brief postictal confusion. After the 11th treatment, she began maintenance ECT. She reported feeling less tension and sadness. Her family reported increased interest, initiating outdoor activities with her husband, and overall a brighter affect. She also had improvement with Parkinson disease symptoms after the fifth treatment—both her tremors and handwriting improved. Her outpatient psychiatrist decided once she was in the maintenance phase to switch her from Remeron and Wellbutrin XL to Brintellix, which was titrated from 5 to 10 mg. At the last ECT treatment, the husband reported continued improvement, especially in motivation.

Severe depression is a common psychiatric illness in geriatric patients, with estimates in the range of 2 to 5 million patients with severe and subsyndromal depression, respectively. Depression in all forms late in life results in significant mental and physical complications, impaired neurocognitive function, decreased activities and instrumental activities of daily living, and increased morbidity. The World Health Organization reported that among all medical and psychiatric diseases, severe depression was the second leading cause of worldwide disability. Of significant health interest, patients with depression later in life have an increased risk of developing psychosis, suicidal ideation, suicidal intent, and suicidal completion. Thus, it is critical to provide a therapy that has a rapid onset of action in order to abate acute crises.

Older adult patients present with multiple psychiatric and medical illnesses. The frequency of reoccurrences and severity of depressive symptoms often increases with age. Severe depression in the older adult confers high levels of functional impairment and morbidity. Older adults are vulnerable to dehydration, weight loss, immobility, and complications from an array of comorbid medical conditions. ECT is indicated for older patients with severe psychiatric illnesses, including affective and psychotic disorders, catatonia of any etiopathology, and, in certain cases, neurologic disorders (Table 55.4f–1). Within the affective disorders spectrum, ECT can be used to treat both unipolar and bipolar depression, with and without psychotic features. For unipolar depression, ECT may be especially beneficial for older adult patients with psychotic or melancholic symptoms.

Geriatric patients with schizophrenia spectrum illnesses may also benefit from ECT, particularly if the psychotic disorder is new onset, acutely exacerbated, or treatment refractory. Schizophrenia affects approximately 1.1% of the U.S. population, and most geriatric patients with schizophrenia tend to manifest severe positive and negative symptoms that complicate their clinical prognosis and limit their adaptive functioning. Originally introduced for the treatment of schizophrenia, ECT is helpful in ameliorating both the positive and negative symptoms in patients who are nonresponsive to typical or atypical neuroleptic treatments.

Table 55.4f–1.
Indications for Electroconvulsive Therapy (ECT)

Major depressive disorder with melancholic features
Major depressive disorder with acute suicidality
Major depressive disorder with psychotic features
Treatment-resistant depression (TRD)
Bipolar disorder with acute mania
Bipolar disorder with depression
Bipolar disorder resistant to or intolerant of pharmacotherapy
Catatonia (of any etiology)
Schizophrenia resistant to or intolerant of pharmacotherapy
Parkinson disease resistant to or intolerant of pharmacotherapy
Major Neurocognitive Disorder due to Lewy Body disease resistant to or intolerant of pharmacotherapy
Major Neurocognitive Disorder due to Alzheimer disease resistant to or intolerant of pharmacotherapy
Epilepsy resistant to or intolerant of pharmacotherapy

Older adult patients often present with affective or psychotic disorders, which are secondary to a medical diagnosis. A thorough examination of the etiology of the psychiatric pathology should be undertaken to determine the association. In those instances where the underlying cause is related to the primary medical illness, medical care should be implemented. If no improvement is seen in the psychiatric illness, treatment should be initiated, beginning with less invasive options (i.e., pharmacotherapy, psychotherapy, social interventions) and incrementally building up to the administration of ECT. However, in cases where there is increased lethality (i.e., catatonia), ECT may be considered a first-line treatment. Delirium requires an exhaustive evaluation to identify and address underlying etiologies. However, ECT may improve the neuropsychiatric symptoms of this serious syndrome.

Neurologic disorders such as Parkinson disease and refractory epilepsy, which afflict older adult patients and result in severe affective or psychotic disorders, are also indications for employing ECT. Older adult patients with Parkinson disease may be at increased risk for severe depression, which may improve with ECT. These patients also often have side effects from drug–drug interactions; hence, ECT may be a favorable alternative to pharmacotherapy. Caution needs to be taken, though, when administering anesthetic agents during the ECT treatment course. Of note, some of the antiparkinsonian medications can result in depressive and psychotic symptoms. A collection of case studies reporting on a total of 22 patients with Parkinson disease found ECT to be beneficial for depressive and parkinsonian motoric symptoms, though considerable relapse in both motoric and affective symptoms occurred. Neuroleptic malignant syndrome, a life-threatening condition related to the use of neuroleptic medications, historically, in some cases, has been treated with ECT. However, ECT is not recommended as a first-line treatment, and prudent use requires careful monitoring of cardiac adverse effects due to pre-existing unstable blood pressure and muscular rigidity, which could be exacerbated during ECT sessions. Because ECT has anticonvulsant properties, as seen with subsequent increases in the seizure threshold after each treatment, ECT can be used to treat older adult patients with treatment-resistant epilepsy or those patients with status epilepticus. The risks and benefits must be weighed when applying ECT as other treatments may be more optimal such as vagus nerve stimulation (VNS), which is approved by the U.S. Food and Drug Administration (FDA) for treatment-resistant epilepsy.

Although not routinely performed, ECT can be administered to older adult patients with dementia. Patients with dementia, such as Lewy body or Alzheimer dementia, frequently develop severe affective symptoms, which can be treated with ECT if other therapeutic options fail. Several published case reports found ECT to be safe and tolerable, and efficacious, with a 40% response rate in older adult patients with dementia. Nonetheless, in patients who are diagnosed with dementia, the utmost attention must be paid to the progression of dementia and other neurocognitive effects in order to minimize cognitive impairments. As such, it is prudent to have a neuropsychological evaluation before commencing with ECT. While ECT is rarely utilized to treat combative or agitated behavior associated with dementia, even when other treatments are ineffective, there is limited research that has documented efficacy for those behavioral symptoms. The use of ECT as a therapeutic modality with patients who are diagnosed with dementia may be legally limited in certain states, and the treatment team would benefit from legal consultation regarding consenting procedures and the inclusion of the family in acute and longitudinal care plans.

Table 55.4f–2.
Medical Evaluation for Older Adult Patients Undergoing Electroconvulsive Therapy

Physical examination
Neurologic evaluation
Complete blood count
Electrolytes
Thyroid function
Electrocardiogram
Anesthesia workup
Oral/dental examination
Brain scan (such as MRI or CT)
Chest x-ray
Spinal series (if clinically indicated)

PRE-ECT EVALUATION

A comprehensive evaluation for older adult patients undergoing ECT should include a complete psychiatric history, neurocognitive evaluation, and medical/surgical history (Table 55.4f–2). The evaluation should examine potential proscriptive factors that could limit the effectiveness of ECT in geriatric patients.

A thorough and comprehensive psychiatric interview should be conducted that ascertains historical and current psychiatric diagnostic and treatment information. Given that older adult patients may have an extensive psychiatric history, it will be advantageous to obtain prior psychiatric records and to include a family member in the interview process to provide corroborating information.

Documentation of the psychiatric treatment history is of vital importance as it will have implications for the ECT treatment course. The historical use of ECT in patients has been found to be a predictor of future safety and efficacy. For instance, patients who benefited from ECT are more likely to benefit from future use of ECT, and those patients who showed no benefit may not benefit from ECT. Nonetheless, the prior ECT procedures

(BL vs. RUL, Stimulus dosing) need to be examined to ensure that ECT was properly conducted and followed American Psychiatric Association–recommended guidelines.

Psychotropic medications that were and were not beneficial need to be recorded in order to determine the level of treatment resistance and to plan the prospective acute, continuation, and maintenance treatment courses. Those psychotropic medications that were found to be beneficial may possibly be added to the acute ECT course to enhance the speed of response and therapeutic advantages and can be used in the continuation and maintenance phases to prolong remission. Collection of the antidepressant treatment history can be facilitated by standardized forms such as the Antidepressant Treatment History Form (ATHF). For example, the ATHF allows for the quantification of antidepressant medications in order to determine the adequacy of the treatment and the level of therapeutic resistance. Collaboration between the ECT practitioner and the geriatric psychiatrist at the initial psychiatric interview can lay the foundation for successful psychiatric treatment and provide an accurate understanding of the older adult patient's psychiatric needs.

A clinician privileged to administer ECT and an anesthesiologist should collect comprehensive medical and surgical histories. Ideally, this evaluation will occur as close as possible to the initial ECT session. Prior surgeries, ECT treatments, and exposure to anesthesia should be ascertained. Specifically, personal or family histories of malignant hyperthermia or other difficulties with anesthesia must be queried. Older adult patients may have multiple gastrointestinal, dental, cardiac, respiratory, musculoskeletal, or neurologic disorders that could complicate anesthesia or ECT procedures. Cardiovascular diseases are a major area of concern for increased morbidity and mortality in the treatment of older adult patients. These include recent myocardial infarctions (generally occurring within 6 months), congestive heart failure, coronary artery disease, aneurysms, and valvular disease. Brain tumors or conditions that increase intracranial pressure also present unique safety concerns. These conditions require careful clinical judgment and, in some cases (i.e., large, rapidly growing, or metastatic tumors), will present a clear contraindication to ECT. Recent hemorrhagic and ischemic strokes or conditions that predispose patients to cerebral vascular bleeding also pose significant safety risks. Poor dentition, loose teeth, and dentures should be evaluated as special precautions (i.e., customized bite block) may be required during the ECT treatment.

Older adult patients also require a careful physical examination with emphasis on the neurologic examination (including fundoscopic examinations to assess for papilledema) and dental, cardiac, respiratory, gastrointestinal, and musculoskeletal systems. Medical or specialty consultation for older adult patients can assist the ECT practitioner with the optimization or understanding of an illness rather than providing only “clearance” for ECT. The laboratory evaluation should be individualized and congruent with clinical indications. However, ECT centers should mandate minimum standards for this patient population, which should include a complete blood count (CBC), electrolytes, chest radiographs, and an electrocardiogram. Spinal radiographs should be considered in geriatric patients with a history of back or neck pain. Brain computed tomography (CT), magnetic resonance imaging (MRI), and electroencephalogram (EEG) are not routine or mandated but may be more beneficial and prudent in this population as compared to young adults.

The older adult patient’s current and prior pharmacotherapy should be reviewed due to interactions between different medications and ECT. Mood stabilizers, such as lithium, taken during the ECT treatment can increase the incidence of delirium after the ECT treatment and can impact seizure expression. The unique classes of antidepressant medications differentially impact ECT, with certain classes enhancing and others interfering with ECT. Monoamine oxidase inhibitors (MAOIs) can cause drug–drug interactions with anesthetic agents and should be discontinued before beginning ECT treatment. Tricyclic antidepressant medications (e.g., nortriptyline) are usually safe to use with ECT and may increase the therapeutic benefits. The use of selective serotonin reuptake inhibitors (SSRIs) in combination with ECT has received limited attention, but the use of selective serotonin-norepinephrine reuptake inhibitors (SNRIs), such as venlafaxine, has been shown to be beneficial when combined during the acute and maintenance ECT course.

Most antideprassants and antipsychotic medications have a synergistic effect with ECT; that is, the combined use of psychotropic medications and ECT can magnify the therapeutic effects. However, augmenting ECT with reserpine should be avoided as it could increase acute hypertension and the incidence of mortality. For patients with comorbid anxiety who use benzodiazepines, usage of those medications can reduce the seizure duration and should be tapered or discontinued before treatment initiation. Also, anticonvulsant medications should be tapered or discontinued 24 hours before each ECT treatment as they inhibit seizure induction and may result in increased stimulus dosage relative to the seizure threshold in order to elicit beneficial seizure activity. More research is warranted to decipher pharmacotherapy and ECT augmentation strategies, particularly with newer antidepressant and antipsychotic medications.

NEUROCOGNITIVE ASSESSMENT

A major concern regarding ECT by patients and their families is the perception that ECT causes cognitive impairment. Neurocognitive assessment of an older adult patient who receives ECT should be conducted before, during, and after the treatment course by a qualified clinician. The assessment before and after the treatment course should comprehensively assess neurocognitive functions, including premorbid intellectual ability, processing speed, attention, verbal and visuospatial memory, working memory, and executive function. Such a comprehensive evaluation allows for the establishment of a baseline to quantify the neurocognitive impact after the treatment course. Also, the baseline testing will aid in ruling out a diagnosis of primary dementia or delirium. Measuring neurocognitive functioning during the treatment course provides treatment-specific information to gauge the effects of ECT.

An exhaustive neurocognitive examination of patients considered for ECT may not be necessary unless there are indications of severe neurocognitive impairment. For example, research has found that certain cognitive domains, such as executive function, attention, and processing speed, are minimally impacted by ECT. The cognitive domain that has received the majority of attention is memory. An acute course of ECT has an immediate impact on anterograde learning and memory, but these effects tend to be transient and resolve within a week of the end of the acute phase. Thus, the practitioner designing and administering the neurocognitive battery needs to balance the information to be acquired with a patient burden. Especially for older adult patients, neurocognitive assessments can be taxing and counterproductive as the information can be compromised due to fatigue and decreased effort.

At a minimum, before and after the ECT course, the neurocognitive domains to be assessed include global cognitive functioning and retrograde memory (Table 55.4f–3).

In most clinical practices, global cognitive functioning is typically assessed using a standardized measure such as the Mini-Mental State Examination (MMSE) or the Montreal Cognitive Assessment (MoCA). A short measure such as the MMSE or MoCA gathers information to determine whether further tests need to be administered to clarify if any other neurocognitive deficits are present. Succinctly measuring global cognitive function taps into many cognitive domains in a short period of time (e.g., within 5 to 10 minutes) and is usually well tolerated by most older adult patients. Unfortunately, the MMSE, MoCA, and other common measures of global cognitive function are insensitive to the effects of ECT. Consequently, the baseline information will provide limited data to guide clinical treatment. In certain cases, it would be beneficial to use age and education-adjusted normative data, such as Mayo’s Older Americans Normative Studies, to account for the age-specific effects to fully appreciate the ECT-associated effects.

Table 55.4f–3.
Neurocognitive Assessment Before, During, and After Electroconvulsive Therapy (ECT)

Before and After ECT	**Global Cognitive Function**
	Attention
	Executive Function
	Learning and memory
	Confrontational naming
	Visuospatial construction
	Retrograde Memory
	Personal information
	Impersonal information
	Executive Function
	Cognitive flexibility
	Social Interaction
During ECT	**Time to Orientation**
	Eyes open
	Name
	Age
	Day of week
	Place
	Date of birth

The data generated from global cognitive function measures can be strengthened by also assessing retrograde memory. To accomplish this, standard scales can be employed that measure autobiographical information for personal and impersonal details. Personal information regards unique information, such as a prior birthday or employment history, whereas impersonal information regards information more of a “pop-culture” nature, such as a famous person, a famous television program, or a well-known event. If a standard scale is not used, the patient can be quizzed using information gathered from the clinical interview and corroborating information from a significant other. An investigation of 50 patients who were randomized to receive either right unilateral or bitemporal ECT found that patients showed a propensity to recall more personal information as opposed to impersonal information, regardless of the ECT stimulus dosage or efficacy outcome. However, relative to patients randomized to receive right unilateral ECT, those who received bitemporal ECT showed greater retrograde amnesia for personal and impersonal information. An important aspect of this study was the measurement and quantification of retrograde amnesia with a novel scale, which had the advantage of being directly translated into clinical practice. To increase the validity of the retrograde memory assessment, corroborating information should be gathered from a significant other.

During the treatment course, an important domain of cognitive function to assess is orientation, especially the time required to achieve full reorientation after each ECT session. Research has consistently shown that ECT has an impact on orientation and causes a period of disorientation immediately after each ECT session. A historical study compared bitemporal and right unilateral ECT and found that bitemporal ECT resulted in lengthier times to recover full orientation compared to right unilateral ECT. Although the study involved a small number of patients, it generated invaluable information and set a precedent for quantifying disorientation with an easy-to-use structured method. This method involves a clinician administering an orientation questionnaire. With a stop-watch or other method to measure time, timing starts as soon as the patient is able to breathe without assistance and ends when the patient accomplishes five out of six orientation criteria. These criteria include opening eyes and correct verbalization of name, place, day of the week, age, and date of birth. In one study, time to orientation was associated with the magnitude of retrograde amnesia (i.e., the lengthier time required to achieve orientation results in greater retrograde amnesia). Factors such as type of lead placement, stimulus dosing, and pulse waveform have been found to affect time to orientation, with bitemporal placement, higher dosages relative to seizure threshold, and sine wave resulting in longer times to achieve orientation. Ultimately, neurocognitive functioning is important to assess in order to quantify the adverse effects of ECT and to assist in clinical decision-making.

CONSENT

Informed consent is not a one-time event; rather, it is a continual process. As such, the initial consenting meeting is a vital component of the ECT treatment as it sets the stage for the consent process. Consenting to procedures before ECT is initiated is important, particularly for geriatric patients, given that there may be neurocognitive compromises related to the underlying illness. Many patients with psychiatric illnesses may lack the complete ability to provide informed consent, particularly those with a psychotic disorder, and this can conflict with the provision of treatment. Table 55.4f–4 provides suggested recommendations when obtaining informed consent.

Table 55.4f–4.
Suggested Informed Consent Procedures for Geriatric Patients Receiving Electroconvulsive Therapy (ECT)

Follow all local, state, and federal guidelines for the proper administration of ECT.
Qualify the patient’s capacity to provide informed consent.
Engage in an open conversation with the patient and a family member (if available) regarding the ECT treatment and the rationale for administration.
Utilize multimedia aids such as videos documenting an ECT treatment. Elicit feedback from the patient and answer any questions that arise.
The consent form must describe all aspects of the ECT course, including indications and contraindications in individual patients for ECT, ECT treatment procedures, anesthesia, adverse effects and respective benefits, alternatives to treatment, and rights to decline ECT treatment.
Use a standardized questionnaire to assess if the patient understands the consent form.
Have a witness cosign the informed consent form.
Document the consent process in the medical record.
Provide the patient with a copy of the signed consent form and place a copy in the chart.
Informed consent is a process, not a one-time activity, and must be monitored throughout the ECT course.

Prior to signing the consent form: Different states may have different requirements for informed consent. Prospective patients need a complete and thorough explanation of what ECT consists of, the procedures that will take place, the associated effects (both positive and negative), the length of the treatment, alternatives to treatment, and the continued care after the acute ECT course is completed. All of this should be written in the consent form that the patient will sign. Having an open forum of questions and answers makes this process informative and respects the patient's autonomy. This initial meeting can be further enhanced by watching a video of the ECT treatment with the patient and clarifying any related questions. The inclusion of the video ensures a comprehensive explanation of the ECT treatment and utilizes multimodal means of learning to help ensure that the patient is fully informed. Further, the initial consenting session can facilitate the establishment of trust between the patient and the doctor, clarify misconceptions, and help demystify ECT, as many older adult patients may view it as a negative treatment.

Building a supportive network for the patient is advantageous during and after the ECT treatment, and the consenting session is an ideal setting to lay that framework. The physician obtaining consent should invite a family member or caregiver to attend the informed consent meeting. This will allow for a thorough discussion of the ECT treatment process, allow the patient to feel that they are not alone, and provide the patient with someone to talk to after the discussion. Although it is not required, having a family member cosign the consent form provides further assurances that consenting procedures are being followed. For example, a retrospective chart review of 62 geriatric patients who received ECT in the New York state area found that 75% of those charts contained a family member's cosignatory, even in cases where the patient was documented to have the capacity to provide informed and voluntary consent. For older adult patients who are cared for at home, the family member is a vital component of the consent process and the overall treatment, as they will be primarily responsible for ensuring that the patient receives the appropriate aftercare. Also, the family member may have unanswered questions and negative stigmas associated with ECT that can be clarified in the consenting session. An additional step is to obtain the patient's consent for the clinician to contact a family member or the primary psychiatrist to obtain additional information on the patient over the course of ECT.

A common procedure to perform before treating a patient is to assess the patient's capacity to provide informed consent, especially for those patients who may have neurocognitive impairments. To provide informed consent, a patient must have the ability to provide voluntarily and fully informed consent. To assess the capacity to consent, the clinician (i.e., nurse, physician) should review the medical chart, acquire the patient's educational history and information, and administer a global cognitive examination such as the MMSE. These procedures will help guide the clinician in determining the level of capacity by which prospective plans can be arranged. For instance, should the patient have a limited understanding of their condition and treatment options, additional education regarding ECT, including a family member, and having the patient answer a questionnaire may help to increase the decision-making ability. An empirical study conducted at the Mayo Clinic comparing geriatric (mean age: 75.6, standard deviation [SD] = 5.87) and nongeriatric patients (mean age: 41.9, SD = 8.01) randomized to an additional half-hour session of education regarding ECT provided in a one-on-one format with a psychiatrist found that the provision of extra education increased the patient's knowledge base to inform their treatment decision. An elaborate battery of neuropsychological tests can be administered to comprehensively assess for any neurocognitive impairment; however, those procedures may not be necessary and could overburden the patient. Nonetheless, they are available to the physician and could be a source of additional information.

Some patients may be unable to voluntarily consent. This could be due to myriad factors, including psychiatric illness (i.e., catatonia), limited ability to make decisions, or blatant refusal. For those patients unable to voluntarily provide consent (different states may have different rules), the consenting physician should ask for an ECT consultation. The outside ECT consultation, usually performed by the chief of ECT services for the hospital, provides a second opinion of the patient's decisional ability and the immediacy for treatment. Particularly for those patients who refuse ECT, the secondary opinion provides further justification for the dire need for treatment based on the safety of the patient and others and can be used in legal proceedings to attain court-ordered treatment. As in voluntary consent, it is beneficial to include a family member in this process. The inclusion of the family member can help to mitigate any coercion or perception of coercion of the patient.

The consent process should be completely documented in the patient's medical record, including the number of meetings, who was in attendance, educational aids used, neuropsychological tests administered (if any), and the signed consent form. A copy of the signed consent form should be given to the patient. Since the consenting requirements and regulations can vary by state, it is critical to understand the respective laws for the consenting of the patient to receive ECT. For example, in the state of Texas, the patient is reconsented before each ECT treatment is provided. The reconsenting of the patient at each treatment further substantiates that consent is not a one-time decision; rather, it is a continual process, and the patient or family member could have new questions arise through each phase of the treatment.

ECT PROCEDURES

Anesthesia

All ECT treatments are provided while the patient is under general anesthesia, and for older adult patients, an anesthesiology consultation is of significant importance in order to determine the optimal anesthetic doses and minimize side effects. The anesthesia consultation should occur before the first treatment. Anesthetic doses should be re-evaluated before each treatment, with subsequent dosage adjustments based on the outcome of the prior treatments. Different classes of anesthetic medications are delivered, including neuromuscular blocking agents and hypnotic agents. The interactions between anesthesia and ECT can affect clinical outcomes, making it essential to have input and collaboration between the anesthesiologist and the ECT practitioner. For example, certain hypnotic agents can raise the seizure threshold, thereby inhibiting the production of seizure activity and resulting in a postponement of clinical efficacy. Hence, the provision of ECT in older adult patients is dependent upon multidisciplinary teamwork.

Age-related physiologic changes must be taken into account when providing anesthesia during ECT. Many older adult patients aged 65 years and older will have slow conduction in peripheral nerves, increased systolic blood pressure, reduced cardiac function, decreased hepatic metabolism, and decreased circulation time. These physiologic changes can result in potential delayed onset of anesthetic action as well as prolonged elimination. This can affect the ECT treatment in various ways. For example, as the seizure threshold is already high in older adult patients, proper dosage and timing of administering hypnotic agents must take into account the eliciting of a therapeutic seizure; otherwise, the ECT treatment may be ineffectual. Also, the administration of succinylcholine, a muscle relaxant, must be timed accordingly so that the body is relaxed upon ECT initiation in order to minimize muscle contraction and allow the patient to regain muscle control in parallel with regaining consciousness.

Different hypnotic agents have been used during ECT treatments, each with their respective physiologic properties that impact the ECT expression. A small study that compared the effects of different doses of methohexital, propofol, and etomidate in 10 patients with an average age of 54 years (range: 27 to 81 years) who received bilateral ECT found that propofol and methohexital at high doses significantly suppressed EEG seizure activity. Also, the use of etomidate, although it had minimal effects on seizure expression, resulted in the prolonged recovery of cognitive function. With the different anesthetic agents available and various administration techniques, expertise is paramount in synergistically pairing ECT and anesthesia.

With the use of anesthetic agents, there is the potential risk of increased confusion or disorientation after the ECT treatment. This side effect becomes critical when providing ECT to older adult patients as they are at an increased risk of ECT-associated adverse cognitive side effects in addition to normal age-related cognitive decline. One small study compared the cognitive adverse effects of anesthesia with etomidate or ketamine in 10 patients and found that short-term memory, as assessed by asking the patients to recall the names of four objects, was less affected by ketamine. The significance of this study is that it showed that different types of anesthetic medications could differentially affect cognitive function after the ECT treatment; however, the conclusions were limited by the small sample size, under dosage of the ECT treatment (i.e., provided right unilateral ECT at 200% above seizure threshold), and the fact that short-term memory was only measured with one nonstandardized cognitive test. On the contrary, in small documented cases, ketamine has been found to result in visual hallucinations and thus is rarely used in current clinical practices. A recent study of 79 patients in Britain found that low-dose ketamine with ECT did not provide any benefit in efficacy or verbal learning skills but increased the risk of adverse side effects. This study was powered only to detect large benefits.

Treatment Parameters

ECT treatment parameters can significantly affect its efficacy and tolerability. The three parameters that interact are pulse waveform, electrode placement, and stimulus dosage. Regarding pulse waveform, the first waveform introduced and has the greatest neurocognitive adverse effects is the sine waveform. Although this waveform is rarely used due to technologic enhancements in ECT devices and newer pulse waveforms (e.g., ultra-brief), a large-scale research study found that it is employed in some psychiatric practices and in some underdeveloped countries. However, the majority of academic and public institutions that provide ECT do so with brief or ultra-brief pulse waveforms. These two waveforms result in equivalent efficacy of sine wave but produce minimal neurocognitive side effects. Ultra-brief is the latest waveform developed and has been associated with faster rates of time to orientation after treatment, less neurocognitive adverse impact, and equivalent efficacy.

There has been, and there still remains, a debate as to what is the appropriate site placement for the ECT electrodes. There have been various electrode placement sites studied over the years of ECT research, and three sites have received considerable attention, those being bitemporal, right unilateral, and bifrontal (Fig. 55.4f–1). For all electrode placements, the patient's scalp will need to be prepared by applying a topical prep solution to clean the scalp, which will reduce impedance and allow maximum passage of the electrical stimulus.

For bitemporal electrode placement, each electrode is placed on each temple 1 inch above the midpoint of the line extending from the tragus of the ear to the external canthus of the eye. For right unilateral, one electrode is placed on the right temple 1 inch above the midpoint of a line extending from the tragus of the ear to the external canthus of the eye, and the other electrode is placed 1 inch to the right of the vertex. For bifrontal, each electrode is placed 2.5 inches above the outer external canthus of the eye. Bitemporal electrode placement was the first electrode placement used and has been shown to be significantly advantageous relative to the other two placement sites in terms of speed of response and remission. However, right unilateral placement, six times above the seizure threshold, has been shown to produce an equivalent response and remission rates despite requiring more time while also having minimal adverse neurocognitive effects. Bifrontal placement has received nominal attention but has shown high efficacy rates coupled with low adverse neurocognitive effects (Fig. 55.4f–2).

Stimulus dosing can be performed with one of two methods, either the age-based method or the empirical titration procedure. Both procedures have been found to produce successful inductions of therapeutic seizures.

In general, with the age-based method, stimulus dosage is adjusted by age, with younger age groups treated at low doses and older age groups treated with higher doses. Some clinicians prefer age-based titration methods because they may limit the number of subconvulsive stimuli. Regarding this method, either full-age or half-age dosing formulas can be applied. The full-age formula (dose = patient age multiplied by five) is more appropriate to use with right unilateral ECT as it typically equates to a dose close to six times the seizure threshold. On the other hand, the half-age formula (dose = patient age multiplied by 2.5) is more fitting with bilateral placement and results in doses approximately 2.5 times the seizure threshold. The age-based methods, though, do not account for individual variability, which could equate to the provision of markedly elevated stimulation with significant cognitive adverse effects.

The empirical titration procedure has been shown to sustain efficacy while limiting adverse effects through the containment of unnecessary energy. This procedure involves the delivery of an initial stimulus dose, which is nontherapeutic, followed by restimulation with increased intensity until a 20- to a 25-second interval of motoric or EEG manifestation of a seizure is observed. As found in clinical and research practice, the seizure threshold increases with age; thus, older adult patients require higher doses to achieve grand mal seizures. Even titration procedures are altered for older adult patients. Whereas with younger patients, those below the age of 50, the first stimulus may be administered at 5%; for the older adult, the starting stimulus is administered at 10%. Stimulus dosing is also influenced by choice of electrode placement. For subsequent treatments after the ECT titration session, bilateral and bifrontal electrode placement dosing is administered at moderately suprathreshold levels, defined as 1.5 times the seizure threshold. On the other hand, right unilateral ECT placement dosing is administered at moderately to markedly suprathreshold stimulation, which is 2.5 to 6 times the seizure threshold. These dosing strategies have been extensively researched and tested in order to ascertain their therapeutic effects.

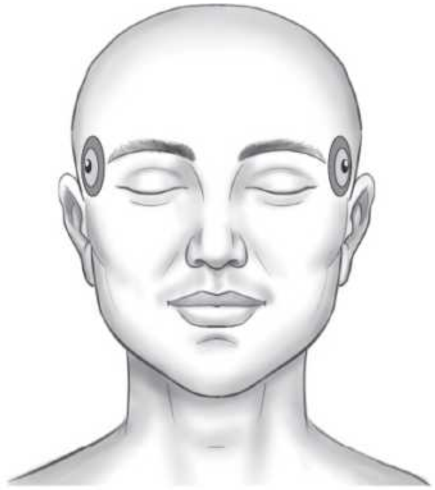

A Bilateral

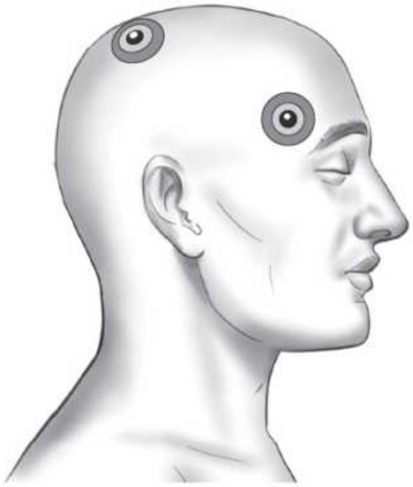

B Right unilateral

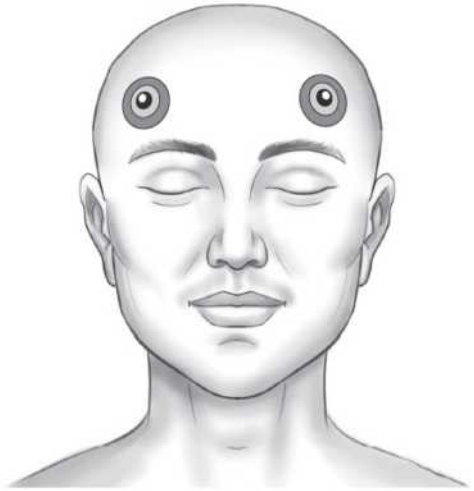

C Bifrontal

FIGURE 55.4f–1. Electrode placement sites. (Adapted from Lisanby SH. Electroconvulsive therapy for depression. *N Engl J Med.* 2007;357(19): 1939–1945.)

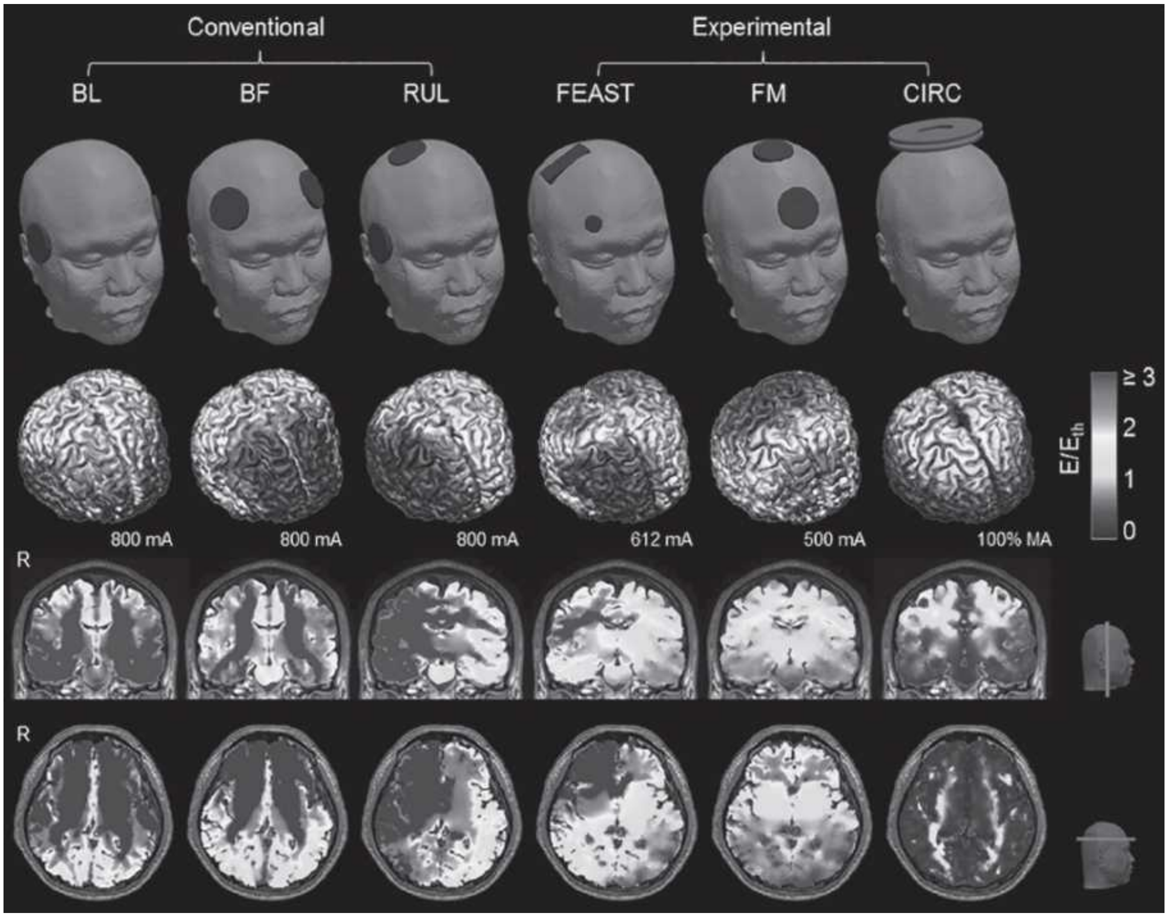

FIGURE 55.4f–2. Electrical field distribution ECT versus MST. (See eBook for color images.) (Reprinted with permission from Lee WH, Lisanby SH, Laine AF, Peterchev AV. Comparison of electric field strength and spatial distribution of electroconvulsive therapy and magnetic seizure therapy in a realistic human head model. *Eur Psychiatry*. 2016;36:55–64. doi:10.1016/j.eurpsy.2016.03.003.)

ECT Monitoring

During the administration of ECT, the induced seizure should be monitored via EEG and motoric response recordings. The EEG monitoring shows characteristic patterns of the grand mal seizure as well as the postictal suppression (Fig. 55.4f–3). The sites for recording the EEG are the frontal–mastoid and frontal–frontal montage, which will maximize the EEG seizure expression. During the tonic phase, hypersynchronous polyspike complexes are followed by polyspike and slow-wave complexes, and during the clonus phase, they are followed by sudden or gradual suppression. Most ECT-induced seizures last less than 90 seconds and consequently are deemed inadequate if they do not show bilateral generalization on EEG manifestations. For monitoring motoric expressions, the cuff technique can be used. This procedure involves applying a blood pressure cuff to a limb (i.e., arm or leg) before the administration of a muscle relaxant. Commonly, the cuff is applied just above the ankle of the left leg in order to observe ECT motor manifestations.

Continuous monitoring of hemodynamic functioning is maintained throughout the ECT treatment session. This is essential for anesthesia and the rapid shifts in parasympathetic and sympathetic outflow that occur after the electrical stimulus. As the patient receives the electrical stimulus, increased parasympathetic tone can produce asystole or significant bradycardia. These hemodynamic effects can be aggravated by subconvulsive stimuli or medications. The ECT-associated bradycardia can be mitigated by pretreatment with anticholinergics such as atropine or glycopyrrolate. During the seizure, heart rate and blood pressure generally elevate and may follow by a period of bradycardia. As a result, geriatric patients with any cardiac condition often require additional monitoring, management, and medications to dampen these effects.

Patients With Comorbid Medical Diagnoses

There are no absolute contraindications to ECT except cochlear implants. It is safe and effective in patients who have comorbid medical disorders with appropriate consultation of relevant medical specialists. Many older adult patients have pre-existing cardiovascular diseases such as hypertension, angina, arrhythmias, aneurysms, conduction abnormalities, and previous myocardial infractions. Consultation with a cardiologist prior to ECT for optimum management is warranted in these situations. Possible concerns involve the increased vagal (parasympathetic) tone during and shortly after each electrical stimulus, which can evoke brief bradycardia or asystole. The onset of the therapeutic seizure leads to elevations in heart rate and blood pressure. Postictal, there is a second period of heightened vagal tone, evidenced by bradycardia and ectopic beats in some patients. With careful management, cardiovascular risk can safely be mitigated in these situations. Anticholinergics such as glycopyrrolate and atropine given prior to ECT can subvert cardiac arrhythmias. Short-acting β-blockers such as esmolol and labetalol are often helpful in controlling hypertension and tachycardia during treatments. Other cardiac agents such as nifedipine, hydralazine, trimethaphan, and nitroglycerin can be used if indicated.

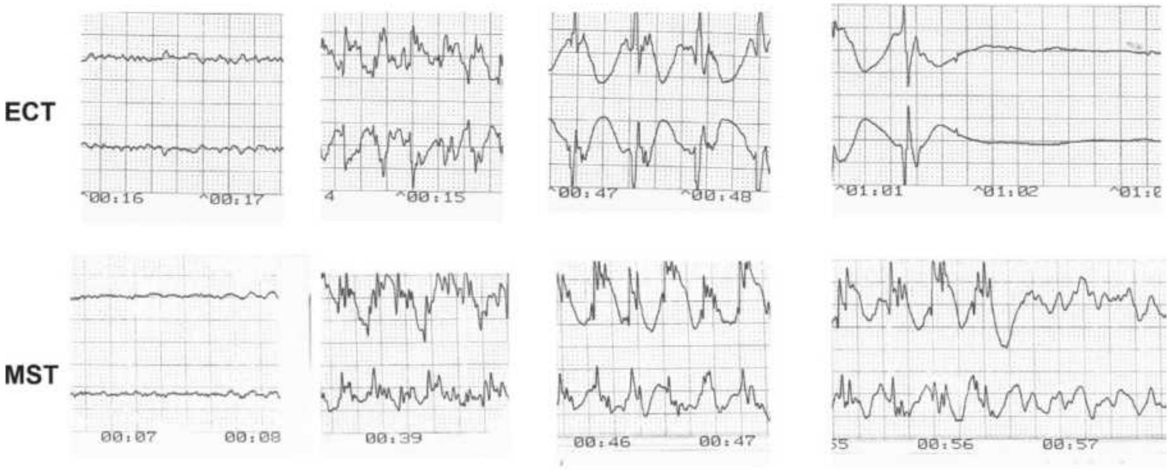

(1) Postinduction baseline

(2) Early seizure activity

(3) Grand mal seizure

(4) End seizure and postictal suppression

FIGURE 55.4f–3. EEG profile ECT versus MST. (Reprinted with permission from White PF, Amos Q, Zhang Y, et al. Anesthetic considerations for magnetic seizure therapy: a novel therapy for severe depression. *Anesth Analg*. 2006;103(1):76–80. doi:10.1213/01.ane.0000221182.71648.a3.)

Intracranial mass lesions or increased intracranial pressure presents a serious risk factor for ECT. Optimally, subdural hematomas should be removed prior to treatment. Older adult patients with depression and who present with previous cerebral infarctions can benefit from ECT. However, the amount of time that should pass after a stroke to maximize safety is unknown. Otherwise, the incidence of cerebrovascular complications is minimal.

Osteoporosis, spinal disc disease, and spondylolysis may necessitate a higher level of muscle relaxation during ECT. Neurosurgical consultation for ideal management prior to treatment is advisable in these situations. Pulmonary status must also be maximized prior to treatment. Unique ventilation strategies may be employed for patients with lung disease.

ECT leads to transient increases in intraocular pressures. Patients with glaucoma should receive eye drops prior to treatment. Collaboration with ophthalmologists for patients with retinal detachments and acute close-angle glaucoma is essential. Patients with loose, uneven, or decayed teeth are at risk for dental injury with ECT. Dental consultation and unique block strategies may need to be employed in this setting.

ACUTE CLINICAL SAFETY AND EFFICACY

ECT is regarded as one of the safest and most effective somatic treatments for severe major neuropsychiatric disorders, particularly those for which alternate treatments have been ineffective. Its safety compares favorably to other procedures that require general anesthesia. The American Psychiatric Association Task Force on ECT, the National Network of Depression Centers Task Force on ECT, and the National Institute for Health and Clinical Excellence in the United Kingdom suggest that ECT is an efficacious acute treatment. In many systematic controlled trials, ECT has consistently been found to be superior to sham ECT (i.e., patients received only anesthesia) and pharmacotherapy.

The remission rates following an acute course of ECT have been found to vary between 20% and 80%, with an effect size as large as 0.9. This large range is dependent on an interaction of both ECT treatment parameters and patient clinical factors. For instance, in order for ECT to be effective, it must be appropriately applied with the respective therapeutic dosage, pulse waveform, and electrode placement. A controlled trial that compared fixed high-dose right unilateral and bitemporal ECT in older adult patients with major depression showed both electrode placements to be similar in terms of clinical outcome with comparable remission rates. However, bitemporal ECT was associated with higher cognitive deficits. In the Consortium for Research in ECT (CORE) study, which investigated the safety and efficacy of continuation ECT, approximately 80% of the total sample achieved remission during the acute phase with bitemporal, dose-titrated, brief pulse waveform ECT. Patients were treated three times a week, and standardized depressive rating inventories were collected within 24 to 48 hours after each treatment to systematically monitor mood change. Thus, a major conclusion of this study was that ECT could result in substantial acute efficacy when treatment is judiciously applied following a standardized therapeutic protocol with continual monitoring of patient symptom severity.

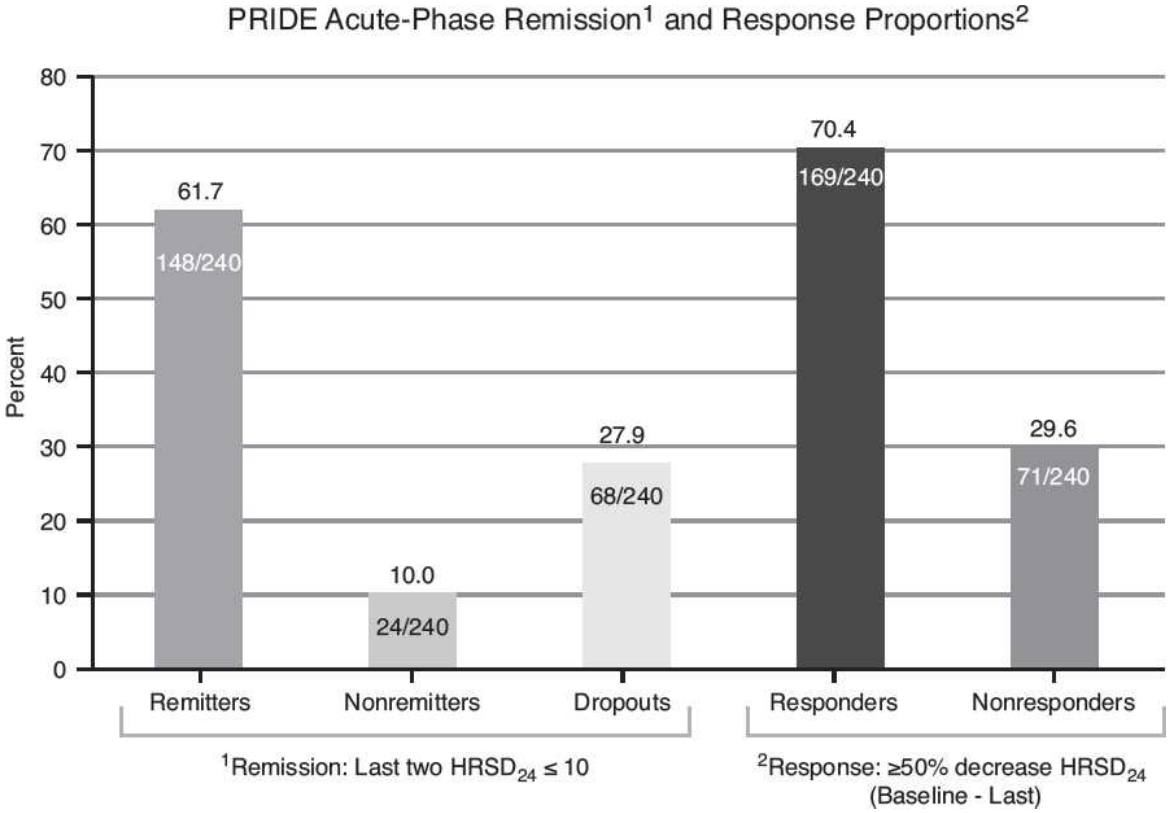

FIGURE 55.4f–4. PRIDE Acute-Phase Remission and Response Rates. (Adapted and reprinted with permission from Kellner CH, Husain MM, Knapp RG, et al. Right unilateral ultrabrief pulse ECT in geriatric depression: phase 1 of the PRIDE study. *Am J Psychiatry*. 2016;173:1101–1109.)

The PRIDE Acute Phase ECT for geriatric depression included 240 patients in the acute phase. In this study, the response rate was 70%, the remission rate was 62%, and the mean time to remission was 2.5 weeks (Fig. 55.4f–4). The PRIDE Maintenance Phase included 122 remitters from PRIDE Acute Phase and showed a decreased relapse rate in patients maintained on ECT + medication (13.1%) versus those maintained on medications alone (20.3%).

The significant benefit of ECT rests not only with its therapeutic benefits but also with the speed at which response and remission are achieved. Analyses from the CORE trial of 253 patients found that approximately 54% and 83% of the sample achieved a response (decrease in depression severity as measured by a standardized rating scale by a minimum of 50%) by the third and sixth ECT treatments, respectively. In terms of remission (level of depression severity found in the normal population), 65% of the sample met the criteria for remission by the tenth ECT treatment. Elucidating the speed with which ECT works, between the fourth and ninth ECT treatment, remission rates increased from 10% to 60%. As noted above, the rapidity of clinical benefit in ECT was supported in the PRIDE 1 trial of 240 patients, where the mean time to remission was approximately 8 to 9 ECT treatments (3 weeks). Given that pharmacotherapy may take many weeks to achieve a steady state and produce results, and psychotherapy likewise may take time for results to show, ECT is the most important therapeutic choice for those patients who are in critical condition due to acute and severe psychiatric illness and in need for quick response.

Early retrospective studies that utilized chart reviews found conflicting results, with some indicating ECT to be ineffective for older adult populations and others finding it to be both safe and efficacious. With technologic improvements and advances in ECT practice, recent controlled clinical trials have found ECT to be highly effective in older adult patients. Comparisons with tricyclic antidepressants, MAOIs, and even SSRIs found ECT to be superior. In fact, age has been found to be an influencing factor for the therapeutic effects of ECT. A report from the CORE trial showed that age was positively associated with treatment response and subsequent remission. In that report, 90% of the older adult cohort, relative to 64% of the younger cohort, responded with ECT. Further examination of the data discovered that the greater response was only valid for those patients with psychotic features and suggested that the presence of psychosis and older age interacted and resulted in higher response rates. That is, older adult patients without psychotic features and younger patients showed comparable response rates, but older adult patients with psychotic features showed higher rates of response relative to younger patients. The PRIDE 1 study also came to a similar conclusion with data that patients 70 and older had an odds ratio of remitting 1.89 greater than those aged 60 to 69.

ECT is a highly effective treatment for the majority of older adult patients, even those who are 75 years and older. A prospective study of 268 patients examined three categories of older adult patients: those younger than 60 years of age, those between 60 and 74 years, and those aged 75 or older. Although the younger cohort had relatively lower seizure thresholds, no differences were found between groups concerning the number of treatments or electrode placement. The middle cohort was found to have higher response rates compared to the younger cohort, but no efficacy differences were found between the middle and older groups.

This study found that ECT is beneficial for the oldest old and confirmed earlier clinical reports that older age is positively associated with higher seizure threshold levels.

The effectiveness of ECT is paralleled by its timeliness in the onset of response, which may result in shorter and less costly hospital stays. A small controlled trial of 13 patients over the age of 65 found a 21% reduction in depression severity (as rated on a standardized scale) after the first ECT treatment and a 36% reduction after the second. The depressive symptom impacted most after the first treatment, with a decrease of 36%, was “reported sadness.” On average, four treatments resulted in a decrease in depression severity by 50%, though the investigators cautioned that response varied per individual case and that further research was required to assess the time to respond and remission in older adult patients. To a degree, these data are synonymous with the findings of the CORE and PRIDE trials, which found ECT to have a quick onset of action.

ECT can be judiciously applied to a spectrum of geriatric patients with comorbid illnesses, including cerebrovascular disease or dementia; however, there is a dearth of information from systematic randomized controlled trials. Retrospectively, chart reviews and clinical reports support ECT to be efficacious in the treatment of severe mood disorders in older adult patients with concomitant dementia, neurocognitive impairments, or stroke. The presence of cerebrovascular abnormalities did not minimize the therapeutic response; nonetheless, when applying ECT to geriatric patients in the presence of concomitant cerebral pathologies and dementias, prudent observation and monitoring of symptomatology and adverse neurocognitive effects become paramount.

The optimal benefit from ECT in older adult patients is dependent upon eliciting therapeutic seizures. Both clinical and research practice has shown that older adult patients have higher seizure thresholds and thus require higher doses of electrical stimulation. One controlled trial reported that in 20 patients, 2 required stimulus doses that exceeded the levels available on commercial ECT devices in the United States. The FDA has limited the ECT devices to 576 millicoulombs; as such, a small percentage of older adult patients with remarkably high seizure thresholds may not benefit due to the inability to elicit seizure activity.

CONTINUATION AND MAINTENANCE ECT SAFETY AND EFFICACY

A cogent constraint of ECT is the high relapse rate after completion of the acute treatment course, with proneness for relapse being greater during the first week and the month after remission. Consequently, though older adult patients may show better response and remission with ECT relative to younger patients, they may also have higher rates of relapse. One study that followed 51 older adult patients for 1 year during and after treatment with either ECT or pharmacotherapy treatment for major depression found that 41% of the total sample was rehospitalized at least once within the first year, and 46% (6 of 13) of the ECT group was rehospitalized during the first month after acute treatment.

Many factors may contribute to the high relapse found following a successful acute course of ECT, including residual depressive symptoms, concomitant physical illness, and psychosocial factors. Residual depressive symptoms are those symptoms that did not abate with antidepressant treatment that could substantially increase the risk of reverting back to baseline status or could be a precursor to the development of a new depressive episode. Many older adult patients who suffer from comorbid physical illness may be burdened with additional stressors that complicate continued recovery from the acute phase. Poor physical health, specifically cardiovascular disease and diabetes, has been linked to both affective and anxiety disorders. Thus, older adult patients with comorbid physical medical illnesses may be more at risk for re-experiencing a return of psychiatric pathology after a period of remission. Finally, older adult patients with increased psychosocial stressors such as lack of autonomy, living alone, or decreased activities and instrumental activities of daily living may be more susceptible to relapse occurrence. Nonetheless, the goal of continuation and maintenance therapy is to sustain the period of remission while simultaneously limiting the associated adverse effects.

The CORE trial was designed to determine whether continuation pharmacotherapy (lithium plus nortriptyline) or ECT (bilateral electrode placement 150% above the seizure threshold) resulted in a better outcome, and the results indicated that both were equivalent in prolonging remission. Of the patients randomized to the continuation ECT condition, 37.1% relapsed, and 46.1% retained remission. For those in the continuation pharmacotherapy group, 31.6% relapsed, and 46.3% retained remission. The time to relapse was similar in both groups, with patients in the ECT arm relapsing after an average of 9 weeks and those in the pharmacotherapy arm after an average of 6 weeks. Although both of these therapeutic modalities were deemed better than no treatment after attaining remission in the acute ECT phase, they were moderately effective in retaining the antidepressant benefits produced during the acute ECT phase.

A probable benefit of maintenance therapy with ECT is the lack of drug–drug interaction. Older adult patients usually receive multiple medications for the management of physical health-related issues. The resulting polypharmacy from mixing a multitude of medications, such as antidepressants, mood stabilizers, analgesics, and cardiac medications, could result in iatrogenic effects. Moreover, older adult patients may be sensitive to the pharmacologic properties of medications, and proper administration must be monitored to ensure the attainment of the full therapeutic effects while limiting harmful reactions. In such cases where older adult patients are extremely sensitive to the effects of medications or they have a coexisting medical condition that limits the use of psychiatric medications, ECT can be a viable alternative. However, given the cardiovascular adverse effects of ECT in older adult patients, especially those with pre-existing hypertension or unstable cardiac activity, continued ECT use needs to be combined with careful and comprehensive monitoring of cardiovascular functioning. For instance, one study of patients over the age of 75 found the development of transient hypertension and sinus tachycardia during and after the administration of ECT. Although the transient phenomena were considered “normal,” they still merit monitoring and evaluation.

ECT ADVERSE EFFECTS

ECT is one of the safest and most well-tolerated somatic treatments, with a documented mortality rate of 0.002%. The majority of side effects are limited to the time during and immediately after an ECT treatment, with the majority of side effects persisting for a short period of time after a complete acute course and others lasting for a longer period of time (i.e., 6 months and longer) (Table 55.4f–5). Three broad side-effect categories for ECT are systemic, neurophysiologic, and neurocognitive.

Systemic Effects

Mortality associated with ECT is most commonly attributed to cardiovascular or respiratory complications. Death can result from myocardial infractions or ruptured arrhythmias, ruptured aneurysms, and brain herniation. Pulmonary aspiration is rare but particularly important to prevent in patients with chronic lung disease. Short-lived increases in blood pressure, pulse, cerebral blood flow, and intracranial pressure present risks for older adult patients but, in the majority of cases, are well tolerated. Ideally, the modern practice of using muscle relaxants prevents fractures that often occurred in the early days of ECT practice. Patients can still suffer from dental complications (i.e., fractured teeth, temporomandibular joint pain) or complain of myalgia, but these risks are extremely rare with the advent of anesthesia and placement of bite blocks. Muscle pain can be related to the use of depolarizing agents such as succinylcholine, and headaches can be successfully treated with over-the-counter analgesics and often improve with time. Pretreatment for musculoskeletal pain can also be beneficial in certain cases. Some older adult patients complain of nausea and vomiting following ECT sessions. Antiemetics such as ondansetron, metoclopramide, and prochlorperazine can be administered to treat these symptoms or prophylactically prior to ECT.

Table 55.4f–5.
Side Effects Related to Electroconvulsive Therapy (ECT)

Headache
Hypertension
Cardiac arrhythmias (bradycardia, tachycardia, asystole)
Nausea
Tongue bite
Temporomandibular joint pain
Disorientation
Delirium
Anterograde amnesia
Retrograde amnesia
Prolonged seizure
Precipitation of mania
Drug–drug interactions
Anesthetic complications
Aspiration
Breathing complications
Myalgia
Dental complications
Postseizure myocardial infarction
Postseizure stroke
Tardive seizure
Scalp burn (very rare)
Bone fractures (very rare)
Spinal disc prolapse (very rare)
Compression fracture (very rare)

Neurophysiologic Effects

Research involving humans and animals indicates that neither ECT nor brief seizures impart structural brain damage. This is further substantiated by carefully controlled imaging studies. However, seizures lasting more than 180 seconds should be aborted with an anticonvulsant anesthetic agent or a benzodiazepine. Lengthy seizures and status epilepticus could theoretically lead to neuronal injury if the patient becomes hypoxic. This is unlikely, though, as ECT patients are well oxygenated, and most prolonged seizures are easily suppressed. Spontaneous seizures after treatment (tardive seizure) are rare but, in their event, require careful evaluation and management. Postictal delirium can involve disorientation, agitation, and confusion and is best managed with supportive interventions, minimization of environmental stimuli, and judicious use of medications such as benzodiazepines.

Neurocognitive Effects

Concerns about the neurocognitive effects of ECT are likely the most limiting factor for the provision of this therapeutically beneficial treatment. Numerous studies have systematically examined the neurocognitive sequelae associated with ECT, taking into account various ECT parameters and neurocognitive domains, as well as the effects of the underlying psychiatric disease. Although research and clinical findings have varied per patient sample and method of ECT administration, the overarching conclusion is that ECT does result in transient negative cognitive effects, particularly in the memory domain.

The ECT parameters studied have included sites of electrode placement, stimulus intensity, and pulse waveform. There are three prominent electrode placement sites used clinically: bitemporal, right unilateral, and bifrontal. Anecdotally, bitemporal is thought to produce the fastest speed of response but also carries the burden of having the most neurocognitive adverse impact. Right unilateral, on the other hand, is thought to have similar properties in terms of efficacy yet a more benign neurocognitive profile. Bifrontal, too has been suggested to have high efficacy with benign neurocognitive effects. Numerous clinical trials studying the impact of bitemporal and right unilateral ECT have consistently found right unilateral to be more cognitively sparing relative to bitemporal, notably with regard to memory functions.

Examinations of bifrontal ECT have been limited due to small sample sizes and inconsistent procedures in terms of dosing strategies. However, the CORE group, in its large-scale, controlled trial of three electrode placements, compared the cognitive effects of bitemporal, right unilateral, and bifrontal placements administered with brief pulse waveform during an acute ECT course. The study found that all three electrode placements were equivalent both in efficacy and neurocognitive adverse effects. However, that study was limited by missing data and patients who exited the study early. A study conducted in Belgium compared bifrontal and right unilateral electrode placements administered with ultra-brief pulse width during the acute course. That study found that both bifrontal and right unilateral were equivalent in both efficacy and neurocognitive outcome.

Studying electrode placement must take into account the interactions of stimulus dosing and pulse waveform. Research has found a consistent direct correlation between stimulus dosage and neurocognitive impairment. With the introduction of stimulus dose titration in the 1980s, the dosage of ECT is limited to the amount needed to provide a therapeutic seizure. This has helped to minimize the negative neurocognitive side effects that were found in earlier use of ECT when higher doses were applied. An empirical example of the effects of stimulus dosage was found in a controlled, double-masked clinical trial that compared bitemporal and right unilateral ECT. The treatment conditions were bitemporal ECT at 150% above the seizure threshold and right unilateral ECT at 50%, 150%, or 500% above the seizure threshold. Efficacy and neurocognitive assessments were performed before the treatment, after the sixth and last treatments, and 1 week after the last treatment. The investigators found that the neurocognitive effects of right unilateral ECT were dependent on the stimulus dosage utilized to provide treatment. When administered at higher doses, statistically and clinically significant differences were found between the right unilateral groups, with the higher dose (i.e., 500% above seizure threshold) resulting in greater time to reorientation and more anterograde memory impairment (i.e., memory for new information). However, even at the highest right unilateral dose, bitemporal electrode placement resulted in more cognitive impairment in the domain of retrograde amnesia.

Another controlled investigation examined the acute neurocognitive effects of only right unilateral ECT delivered at either a fixed high dose (403 millicoulombs) or a moderate titrated dose (2.25 times the initial seizure threshold). The findings mimicked those of other studies and showed that right unilateral ECT at a fixed high dose produced greater neurocognitive effects than the moderate titrated dose in the domains of time to orientation, global cognitive function, and retrograde memory. Thus, when deciphering the neurocognitive effects, electrode placement and stimulus dose must both be examined in order to sufficiently understand the interaction. With the recent advent in studying the interaction effects of pulse waveform, there has been inconsistent information. For example, a randomized, controlled clinical trial found that ultra-brief, relative to brief pulse, was associated with fewer neurocognitive effects, regardless of electrode placement. However, in another randomized controlled clinical trial that compared brief and ultra-brief pulse waveforms in patients treated only with right-unilateral electrode placement, empirical dose-titrated ECT found that both pulse waveforms produced similar clinical and neurocognitive outcomes. That is, the study suggested that ultra-brief pulse waveform has no neurocognitive advantages relative to brief pulse waveform. Further research is warranted to clarify both the clinical and neurocognitive effects of the ultra-brief pulse waveform.

Time to reorientation (i.e., the time needed after the end of each ECT session to regain orientation) is greatly impacted immediately after each treatment, with resolution following in time after treatment. Like other cognitive domains, orientation time is affected by electrode placement, stimulus dosage, and pulse waveform with bitemporal electrode placement, higher stimulus dosages, and sine waveform resulting in lengthier times to reorientation. In a study that compared bitemporal and right unilateral ECT, both electrode placements showed relatively similar patterns of orientation recovery over the course of treatment, indicating that subsequent ECT treatments do not incrementally prolong orientation recovery time. For instance, in that study, the first and last two treatments were compared, and no within-group differences were found regarding reorientation time, even though between-group differences were observed. Of interest, with right unilateral ECT, age was positively related to recovery time, with older patients regaining orientation at a faster rate than younger patients. The positive association between age and reorientation time was mitigated by the lengthy reorientation time in the bilateral group. Research of the pulse waveforms has consistently found that sine waveform results in the longest time to regain orientation, followed by brief pulse, and finally, ultra-brief pulse.

When assessing neurocognitive effects, acute and long-term effects must be examined. Acutely, ECT results in modest impairment of many domains, including visuospatial and verbal learning and memory and global cognitive function. However, those effects typically resolve quickly within 1 week to 1 month after the end of the acute ECT course. A study of 45 depressed patients who received ECT (administered with right unilateral or bitemporal electrode placement) compared those above and below the age of 60 on a battery of neurocognitive tests at set time points. The study found that immediately after and at long-term follow-up, patients showed improvement in neurocognitive measures of attention, verbal fluency, and memory for visual shapes. The investigators concluded that the improvement in neurocognitive function may have been related to decreased depression severity or the ECT-associated activation of dopaminergic and serotonergic neurotransmitters.

Consequently, in terms of short- and long-term neurocognitive effects, the domain of neurocognitive function is greatly impacted is retrograde memory. Initially, retrograde memory was thought to resolve over time after the end of the acute phase, with a substantial recovery of personal memories. Earlier research found that retrograde amnesia could be predicted by the global level of cognitive functioning before the initiation of ECT.

A community study of ECT practices in the New York area showed evidence that retrograde amnesia is a persistent effect of ECT and that some autobiographical information may be lost. Seven New York hospitals (three community hospitals, two private hospitals, and two university hospitals) were included in the study, and a total of 347 patients who were longitudinally followed for 6 months provided outcome data. Focusing on the neurocognitive outcome, the seven treatment facilities differed due to ECT practice differences at the respective sites. For instance, one center administered ECT with a sine waveform that was associated with more effects on psychomotor speed and memory for autobiographical information. The significance of the site differences indicated that there is no one uniform practice of ECT administration and that outdated methods of ECT treatments that result in greater neurocognitive effects are still used in community practice.

More importantly, the study found that retrograde memory deficits persisted for at least 6 months and were related to ECT technique, female gender, lower premorbid intelligence, and advancing age. Of equal importance, the number of ECT treatments with bitemporal electrode placement was associated with the amount of retrograde amnesia, even at the 6-month time point. However, that study was critiqued as it used an autobiographical memory interview that has not been shown to have validated psychometric properties. Further research is needed with psychometrically sound measures of autobiographical information to confirm or refute those findings. Nonetheless, older adult patients may be more prone to ECT-associated adverse cognitive effects, and treatment with right unilateral electrode placement, empirical dose titration, and ultra-brief pulse waveform may help to mitigate the cognitive effects.

Autobiographical memory, the area mainly affected by ECT, can be categorized into two types: One that focuses on personal memories and one that focuses on impersonal memories (i.e., public events). A clinical trial of patients treated with either right unilateral or bitemporal ECT, who were compared to a healthy control group, found that memory for impersonal events was more affected than memory for self, and greater deficits were found in those patients treated with bitemporal ECT. The pattern of memory loss showed that more recent events, those closest in time to the delivery of ECT, were the most vulnerable to the ECT effects. Also, those events that were patient determined to have low salience were most vulnerable to the effects of ECT.

The most reliable data on the neurocognitive effects of acute ECT come from the PRIDE Phase 1 study. In this study, cognition was assessed at baseline and within 72 hours after the last ECT session. Some domains showed no significant change from baseline. These include verbal learning, delayed free recall of words, and initiation/perseveration. Other domains showed statistically significant declines, which were assessed to NOT be clinically meaningful. The assessment of modest impact was based on small effect sizes and the cognitive domains remaining intact/average to mildly impaired. The affected domains include autobiographical memory consistency, short-term free recall and recognition of learned words, phonemic fluency, processing speed, visual scanning, and cognitive flexibility. The assessment of autobiographical memory consistency should be considered in light of research showing that this domain typically declines over a period of time equivalent to an index ECT course, meaning that the decrease observed in this study does not necessarily imply neurocognitive impairment.

A neurocognitive domain that is often overlooked in both clinical and research arenas is the patient's view of their cognitive experience, which is commonly referred to as subjective cognitive functioning. Many older adult patients with affective disorders will complain of memory difficulties, short attention spans, and slow psychomotor speed. Although the neurocognitive functions may be intact based on formal neurocognitive testing, the patient's subjective experience needs to be considered, especially with regard to ECT. Due to the negative stigma associated with ECT, which some older adult patients refer to as the “shock,” their subjective viewpoint is a factor in the efficacy of the treatment. Prior research found that patient self-rating of neurocognitive function was related to a depression severity self-rating scale, with higher scores on the depression scale correlating to greater subjective cognitive impairment scores. During and following an acute ECT course, patients showed significant improvement in subjective cognitive function, indicating that their memory had improved regardless of electrode placement type (bitemporal or right unilateral). The change in self-evaluation of neurocognitive functions was correlated with improvement in mood and was also related to the degree of retrograde amnesia.

Finally, it is important to note that ECT produces marked and significant improvement in health-related quality of life measures, with the largest gains in the mental health component of this measure.

In summary, the PRIDE Studies were an important addition to the clinical literature on ECT. Acute Phase focused on the efficacy and tolerability of Right Unilateral Brief Pulse ECT (RUL-BP). Maintenance Phase focused on neurocognitive side effects and how to prolong remission.

Pride Acute Phase was a multisite randomized trial of 240 adults who were 60 years old or older and in a depressive episode of major depressive disorder. Patients were given an acute course of Right Unilateral Ultra-Brief Pulse ECT combined with flexible-dose venlafaxine targeted to 225 mg qd. ECT was given 3 times per week, and the average number of sessions to produce remission was 7.3 ECT treatments. Results of Acute Phase, 62% of patients met remission criteria, 10% did not remit, and 28% dropped out. The vast majority of nonremitters met response criteria with a total response rate to ECT of 70%. Remission was quite rapid, with a mean time to remission of approximately 2.5 weeks. The largest contributor to remission was ECT, as the cohorts with the highest and lowest venlafaxine dose had no differences in efficacy outcomes. PRIDE Acute Phase 1 showed that ECT is an effective treatment for moderately severe and severe depression in geriatric patients. Insomnia was also studied in PRIDE Maintenance Phase with the finding that ECT + venlafaxine had no modest benefit for insomnia which was linked to the general improvement in depression symptoms. A comprehensive neurocognitive battery was given 72 hours after the last

ECT to assess for neurocognitive side effects. One hundred and seventy-two patients completed the study.

The PRIDE Phase 2 trial was designed to determine if pharmacotherapy alone was equivalent to pharmacotherapy + ECT. In Maintenance Phase, remitters (n = 120) were randomized to either pharmacotherapy (venlafaxine + lithium) or pharmacotherapy plus RUL-UBP (Ultra brief pulse) ECT. ECT was administered in a symptom-titrated algorithm-based longitudinal ECT protocol referred to as STABLE. STABLE mandated 4 additional ECTs in the first month after completion, as well as symptom-titrated ECT as needed. Results of the study indicated that pharmacotherapy + ECT was more effective than pharmacotherapy alone.

OTHER NEUROSTIMULATION THERAPIES

Although ECT is a highly effective treatment for major affective disorders and other severe treatment-resistant psychiatric conditions, its use is limited due to associated adverse cognitive effects and high relapse rates after the acute phase. In order to remedy the negative effects of ECT and to increase the delivery of efficacious treatments, other neurostimulation modalities are being developed as antidepressant therapies as well as therapies for other mental health and brain conditions. Many of these neurostimulation modalities have mechanisms of action that differ from ECT and, thus, could provide an understanding of the underlying pathophysiology of neuropsychiatric disorders. These modalities include transcranial direct current stimulation (tDCS), transcranial alternating current stimulation (tACS), Vagus nerve stimulation (VNS), rapid transcranial magnetic stimulation (rTMS), magnetic seizure therapy (MST), and deep brain stimulation (DBS). Most of these newer neurostimulation therapies are in an experimental phase of development, especially in geriatric patients.

Transcranial Direct Current Stimulation and Transcranial Alternating Current Stimulation

Transcranial direct current stimulation (tDCS) applies a weak direct current (1 to 2 mA) with scalp electrodes. Depending on how it is delivered, anodal stimulation increases cortical excitability, while cathodal stimulation increases cortical inhibition. tDCS has variable effects on long-term potentiation (LTP) or long-term depression (LTD). As a result, this tool has been applied to a wide array of studies investigating neuroplasticity and cognitive neuroscience. Generally, tDCS is well tolerated and has the potential to alter cortical activity beyond the stimulation session. Therapeutic research has focused on depression, stroke rehabilitation, cognitive remediation, and, more recently, studies of dementia. Contemporary work suggests that clinical outcomes with tDCS may be enhanced by coupling it with cognitive remediation, antidepressants, or other concurrent treatments. In late life, most tDCS studies have focused on enhancing cognition among patients with dementia or at risk for dementia, such as those with mild cognitive impairment. In a large ongoing study (Prevention of Alzheimer dementia with Cognitive remediation plus tDCS in Mild cognitive impairment and Depression (PACt-MD), older patients with remitted depression, mild cognitive impairment, or both are randomized to receive tDCS to the frontal cortex combined with cognitive training versus sham tDCS plus sham cognitive training. The main objective is to assess whether enhancing prefrontal cognitive function with tDCS and cognitive training could prevent cognitive decline in these high-risk populations.

Transcranial alternating current stimulation (tACS) is another form of transcranial electrical stimulation that delivers an alternating instead of a direct current to targeted regions. Also focusing on cognition in older adults, tACS delivered to the frontal and temporal regions of the brain seem to be a promising strategy to enhance cognition, especially when the frequency of the alternating current is matched to naturally occurring brain oscillations, for example, theta oscillations.

Rapid Transcranial Magnetic Stimulation

Rapid (or repetitive) rTMS is a noninvasive neurostimulation technique that applies alternating magnetic currents, which then produce an electrical current in the underlying cortical areas. rTMS has been under significant investigation as a treatment for major affective disorders and other neuropsychiatric illnesses.

The use of rTMS is attractive to many clinicians as it has been shown to be safe and well-tolerated. In 2008, the FDA-approved rTMS for treating adult patients with unipolar major depressive disorder for whom one antidepressant treatment has failed to work. The main side effects reported include temporary headaches and scalp pain at the site of stimulation. As rTMS has been investigated for more than 20 years, safety guidelines have been generated by expert scientists to guide clinical decision-making regarding rTMS treatment parameters. A limitation of using rTMS in a geriatric sample is the presence of cortical atrophy in the prefrontal cortex (the usual site of stimulation). Cortical atrophy is inversely related to treatment outcome with rTMS and has been found to be a predictor of nonresponse. Deep TMS (dTMS) claims to have deeper cortical penetration and may address this issue. However, to date, there are no neuroimaging corelated studies in this regard.

While fewer and smaller studies have been conducted to evaluate rTMS in older adults with depression, the current literature suggests that it is as effective in this population as in mid-life and that it is well tolerated.

To date, two studies examined the effects of rTMS in older adults with depression in randomized controlled trials.

The first study examined the results from two previous clinical trials by selecting older adults. It showed that only bilateral high-frequency rTMS, and not unilateral or sham rTMS, was effective in achieving remission in older adults with depression.

The second study examined the efficacy, tolerability, and cognitive effects of high-dose deep rTMS in older adults with depression. In this study, older adults with depression were randomized to sham or active deep rTMS delivered over the dorsolateral and ventrolateral prefrontal cortex 5 days per week over 4 weeks. The study found that the remission rate was higher with active than sham rTMS (40.0% vs. 14.8%). However, there was no change in cognition and no serious adverse events.

Recently a novel form of rTMS called theta-burst stimulation (iTBS) that clusters TMS pulses into high-frequency bursts of TMS pulses at 50-Hz repeated every 200 milliseconds for 2 seconds on and then 8 seconds off for a total of 600 pulses has been shown to be as effective as rTMS for adults with depression. The advantage of iTBS over traditional rTMS is that each treatment is about 3 minutes instead of 30 minutes, and multiple-session-per-day protocols are now being investigated to accelerate treatment. However, to date, no studies on geriatric depression are available.

Magnetic Seizure Therapy

Magnetic seizure therapy (MST) is a neurostimulation therapy currently being studied that utilizes rTMS at high frequencies (i.e., 50 to 100 Hz) to induce therapeutic seizures. The goal of MST is to focally induce therapeutic seizures, containing the cortical spread of the seizure, in order to maximize efficacy while limiting adverse effects. Initial research in primate and human models found MST to be safe and well tolerated, with evidence of limited negative effects. Several pilot follow-up studies in humans further suggested that MST is effective for the treatment of acute depression and comparable to ECT but with fewer cognitive side effects. For example, in one trial, 20 patients with unipolar or bipolar major depressive disorder were randomized to receive focal MST, with the magnetic stimulating coil placed over the frontal cortex, or nonfocal MST, with the coil placed over the vertex. The results found the treatment to be modestly efficacious, with response and remission rates of 50% and 15%, respectively. Of benefit, MST showed limited cognitive effects in the domain of retrograde amnesia for autobiographical memories. Also, there were no adverse effects in the domains of global cognitive functioning, verbal and visual anterograde learning and memory, or time to orientation. Patients also reported subjective improvement in cognitive functioning.

In another study, 37 patients with depression completed at least nine ECT or MST (100-Hz) treatments in a randomized, double-blind protocol. While response rates were low for both arms, MST did not differ from ECT in effectiveness. However, MST was associated with improvements in psychomotor speed, verbal memory, and cognitive inhibition and with no cognitive decline.

A large study is currently underway that includes patients over 65 to compare MST to ECT to confirm whether they are equivalent and to compare their cognitive effects: the Confirmatory Efficacy and Safety Trial of Magnetic Seizure Therapy for Depression (CREST-MST) study. This study is a randomized, double-blind, noninferiority clinical trial comparing MST for treatment-resistant depression to RUL-UB-ECT. Treatment is administered 2 to 3 days per week until reaching remission or up to 21 treatment sessions. Remission from suicidal ideation is also being measured. MST also holds promise to have antisuicidal effects in treatment-resistant depression. Given the higher risk of suicide among older adults suffering from depression, its use in this population may also be further justified by this specific effect.

Vagus Nerve Stimulation

Vagus nerve stimulation (VNS) was approved by the FDA in 2005 as an adjunctive antidepressant treatment for patients with treatment-resistant depression. VNS involves implanting a device similar to a pacemaker that produces brief, intermittent electrical stimulation to the vagus nerve. This neurostimulation modality was investigated in two large-scale, randomized, controlled clinical trials. The first, an open-label trial, found VNS to be safe and tolerable, producing limited side effects grouped into two categories, surgery related and stimulation related. These side effects were transient and resolved over time; even those related to stimulation lessened over time, and many patients reported substantial adaptation to side effects such as increased cough and hoarseness during active stimulation.

In terms of efficacy, the second trial, double-masked and sham-controlled, found no difference between active and sham VNS during the acute phase. However, during the long-term open-label phase in which all patients received active treatment, depression severity decreased as measured by both clinician and patient reports, and modest response (29.8%) and remission (17.1%) rates were found. Although VNS has shown some benefits, it remains unclear how to adequately provide dosing. In the initial study, the majority of the patients received high-dose VNS, whereas, in the second study, VNS doses were lower.

Of significance, there were no negative cognitive side effects found, and conversely, patients showed minimal improvement in the cognitive domains of psychomotor speed and executive functions. However, this improvement was supposed to be related to the improvement in depression as opposed to directly related to VNS. Nonetheless, VNS could be an option for geriatric patients with treatment-resistant depression who have not benefited from other available antidepressant therapies and who, for any medical reasons, are unable to undergo ECT.

VNS has also been studied in pilot investigations for cognitive enhancement in Alzheimer dementia, and a large study is ongoing to evaluate it as a potential treatment for MCI using a noninvasive transcutaneous VNS (tVNS). tVNS stimulates the afferent auricular branch of the vagus nerve, located medial of the tragus at the entry of the acoustic meatus. Given that it does not require surgery, it is likely to be safer for older adults.

Deep Brain Stimulation

Deep brain stimulation (DBS) involves the surgical implantation of a pacemaker-like device that sends continuous electrical stimulation to a specifically targeted region in the brain. DBS has been safely used with beneficial results in patients with severe movement disorders such as Parkinson disease. The procedure is considered invasive but safe, as the surgery takes place while the patient is awake, and the electrode leads are placed directly in cortical tissue.

There has been recent interest in using DBS to treat severe affective disorders based on initial anecdotal observations that patients with movement disorders appeared brighter during the course of treatment. DBS has been granted FDA approval and Humanitarian Device Exemption for the treatment of obsessive-compulsive disorder (OCD). DBS is currently being studied for a variety of other psychiatric and neurologic conditions, including addiction, anorexia nervosa, depression, and schizophrenia. While these studies are not specific to geriatric populations, one geriatric condition for which DBS is actively being investigated with some promising early results is Alzheimer dementia. Lazanos reported a case with improved cognition with DBS that has prompted interest in studying DBS for dementia.

Preliminary studies in Alzheimer dementia suggest that DBS of the fornix increases glucose metabolism in the temporal and parietal lobes and that these changes tend to be associated with clinical benefits among patients older than 65. Larger and confirmatory studies are underway. Other preliminary studies investigated DBS of the Nucleus Basalis of Meynert, with findings suggesting a cognitive and metabolism-stabilizing effect. Similar preliminary evidence of a clinical stabilizing effect is suggested by DBS to the ventral capsule/ventral striatum.

FUTURE DIRECTIONS

Algorithmic treatment development is a crucial future step to optimize the application of ECT and neuromodulatory interventions in the older adult. This work will aid clinicians in sequential treatments or complex regimens with multimodal approaches. Further work to identify risk factors for neurocognitive side effects and minimize this burden for patients will improve outcomes. Another important area is optimal continuation and maintenance treatment which is still a major problem and poorly understood.

With the advent of newer somatic noninvasive therapies such as tDCS, rTMS, and MST, and surgically invasive therapies such as VNS and DBS, additional research is warranted to understand their effects in geriatric populations. Research thus far has found these neurostimulation therapies to be cognitively benign and possibly cognitively enhancing with promising clinical efficacy in relatively younger cohorts, and this will have to be tested in the geriatric population and can have significant therapeutic implications when treating older adult patients, especially those with pre-existing cognitive deficits. MST, in particular, though presently in developmental stages, could be a new frontier in convulsive therapies through its proposed mechanism of action that facilitates therapeutic seizure induction coupled with little neurocognitive effects.

With newer neurostimulation and convulsive therapies added to the psychiatric therapeutic armamentarium, future directions will need to begin to address the issue of developing high-caliber treatment algorithms tailored to the individual geriatric patient. This could involve utilizing different neurostimulation modalities in different treatment courses. For example, patients who benefited from acute ECT may receive continuation TMS to prolong the period of remission. Also, the combination of neurostimulation modalities with pharmacotherapy requires an understanding of how to refine that administration in order to enhance clinical efficacy, particularly in older adult patients who are prescribed multiple medications. The goal of these treatment algorithms should focus on defining the optimum delivery of somatic therapies that maximize therapeutic benefits and minimize unnecessary medications and adverse reactions. In addition, these evidence-based algorithms for geriatric patients should underscore the importance of neurocognitive assessments and effects when determining clinical management with somatic and multimodal therapies. These scientific advances in geriatric psychiatry will contribute to the continued evolution of novel and advanced somatic therapies for older adult patients.

FURTHER READINGS

Anderson IM, Blamire A, Branton T, et al. Ketamine augmentation of electroconvulsive therapy to improve neuropsychological and clinical outcomes in depression (Ketamine-ECT): a multicentre, double-blind, randomized, parallel-group superiority trial. *Lancet Psychiatry*. 2017;4(5):365–377.

Bjolseth TM, Engedal K, Benth JS, et al. Clinical efficacy of formula-based bifrontal versus right unilateral electroconvulsive therapy (ECT) in the treatment of major depression among elderly patients: a pragmatic, randomized, assessor-blinded, controlled trial. *J Affect Disord*. 2015;175:8–17.

Blumberger DM, Hsu JH, Daskalakis ZJ. A review of brain stimulation treatments for late-life depression. *Curr Treat Options Psychiatry*. 2015;2:413–421.

Bosboom PR, Deijen JB. Age-related cognitive effects of ECT and ECT-induced mood improvement in depressive patients. *Depress Anxiety*. 2006;23:93–101.

Brunoni AR, Valiengo L, Baccaro A, et al. The sertraline vs. electrical current therapy for treating depression clinical study: results from a factorial, randomized, controlled trial. *JAMA Psychiatry*. 2013;70(4):383–391.

Chopra A, Abulseoud OA, Sampson S, et al. Mood stability in Parkinson disease following deep brain stimulation: a 6-month prospective follow-up study. *Psychosomatics*. 2014;55(5):478–484.

Feliu M, Edwards CL, Sudhakar S, et al. Neuropsychological effects and attitudes in patients following electroconvulsive therapy. *Neuropsychiatric Dis Treat*. 2008;4:613–617.

Gardner BK, O'Connor DW. A review of the cognitive effects of electroconvulsive therapy in older adults. *J ECT*. 2008;24:68–80.

Husain MM, Rush AJ, Fink M, et al. Speed of response and remission in major depressive disorder with acute electroconvulsive therapy (ECT): a Consortium for Research in ECT (CORE) report. *J Clin Psychiatry*. 2004;65:485–491.

Kellner CH, Husain MM, Knapp RG, et al. Right unilateral ultrabrief pulse ECT in geriatric depression: phase 1 of the PRIDE study. *Am J Psychiatry*. 2016;173:1101–1109.

Kellner CH, Husain MM, Knapp RG, et al. A novel strategy for continuation ECT in geriatric depression: phase 2 of the PRIDE study. *Am J Psychiatry*. 2016;173:1110–1118.

Kellner CH, Knapp R, Husain MM, et al. Bifrontal, bitemporal and right unilateral electrode placement in ECT: randomized trial. *Br J Psychiatry*. 2010;196(3):226–234.

Kimball JN, Rosenquist PB, Dunn A, McCall V. Prediction of antidepressant response in both 2.25 x threshold RUL and fixed high dose RUL ECT. *J Affect Disord.* 2009;112:85–91.

Lee WH, Lisanby SH, Laine AF, Peterchev AV. Minimum electric field exposure for seizure induction with electroconvulsive therapy and magnetic seizure therapy [published online ahead of print December 9, 2016]. *Neuropsychopharmacology.* 2017;42:1192–1200.

*Lisanby SH. Electroconvulsive therapy for depression. *N Engl J Med*. 2007;357:1939–1945.

Lisanby SH, Husain MM, Rosenquist PB, Maxiner D, Gutierrez R. Daily left prefrontal repetitive transcranial magnetic stimulation in the acute treatment of major depression: Clinical predictors of outcome in a multisite, randomized controlled clinical trial. *Neuropsychopharmacology.* 2009;34:522–534.

Lisanby, SH, McClintock SM, Alexopoulos MD, et al. Neurocognitive effects of combined electroconvulsive therapy (ECT) and venlafaxine in geriatric depression: Phase 1 of the Pride study. *Am J Geriatr Psychiatry*. 2020;28:3.

Lozano AM, Mayberg HS, Giacobbe P, et al. Subcallosal cingulate gyrus deep brain stimulation for treatment-resistant depression. *Biol Psychiatry*. 2008;64:461–467.

Mantovani A, Lisanby SH. Transcranial magnetic stimulation in major depression. In: George MS, Belmaker RH, eds. *Transcranial Magnetic Stimulation in Clinical Psychiatry*. American Psychiatric Publishing, Inc.; 2007.

McCall WV, Lisanby SH, Rosenquist PB, et al. Effects of a course of right unilateral ultrabrief pulse electroconvulsive therapy (ECT) combined with venlafaxine on insomnia symptoms in elderly depressed patients. *J Clin Psychiatry*. 2018;79(2):16m11089.

McCall WV, Lisanby SH, Rosenquist PB, et al. Effects of continuation electroconvulsive therapy on quality of life in elderly depressed patients: a randomized clinical trial. *J Psychiatr Res*. 2018;97:65–69.

McClintock SM, Choi J, Deng ZD, et al. Multifactorial determinants of the neurocognitive effects of electroconvulsive therapy. *J ECT*. 2014;30(2):165–176.

McClintock SM, Trevino K, Husain MM. Vagus nerve stimulation: Indications, efficacy and methods. In: Swartz C, ed. *Electroconvulsive and Neuromodulation Therapies*. Cambridge University Press; 2009.

Navarro V, Gasto C, Torres X, et al. Continuation/maintenance treatment with nortriptyline versus combined nortriptyline and ECT in late-life psychotic depression: a two-year randomized study. *Am J Geriatr Psychiatry*. 2008;16:498–505.

Nordenskjold A, von Knorring L, Engstrom I. Predictors of the short-term responder rate of electroconvulsive therapy in depressive disorders—a population based study. *BMC Psychiatry*. 2012;12:115.

Ostergaard SD, Speed, MS, Kellner, CH, et al. Electroconvulsive therapy (ECT) for moderate-severity major depression among the elderly: data from the Pride study. *J Aff Disord*. 2020;274:1134–1141.

Oudega ML, van Exel E, Stek ML, et al. The structure of geriatric depressed brain and response to electroconvulsive therapy. *Psychiatry Res*. 2014;222:1–9.

Rhebergen D, Huisman A, Bouckaert F, et al. Older age is associated with rapid remission of depression after electroconvulsive therapy: a latent class growth analysis. *Am J Geriatr Psychiatry*. 2015;23:274–282.

Sackheim HA, Haskett RF, Mulsant MD et al. Continuation pharmacotherapy in the prevention of relapse following electroconvulsive therapy: a randomized controlled trial. *JAMA* 2001;285(10):1299–1307.

Segrave RA, Arnold S, Hoy K, Fitzgerald PB. Concurrent cognitive control training augments the antidepressant efficacy of tDCS: a pilot study. *Brain Stimul*. 2014;7:325–331.

Semkovska M, McLoughlin DM. Objective cognitive performance associated with electroconvulsive therapy for depression: a systematic review and meta-analysis. *Biol Psychiatry*. 2010;68:568–577.

Sienaert P, Vansteelandt K, Demyttenaere K, Peuskens J. Randomized comparison of ultra-brief bifrontal and unilateral electroconvulsive therapy for major depression: cognitive side-effects. *J Affect Disord*. 2010;122:60–67.

Spaans HP, Sienaert P, Bouckaert F, et al. Speed of remission in elderly patients with depression: electroconvulsive therapy v. medication. *Br J Psychiatry*. 2015;206:67–71.

Spaans HP, Verwijk E, Comijs HC, et al. Efficacy and cognitive side effects after brief pulse and ultrabrief pulse right unilateral electroconvulsive therapy for major depression: a randomized, double-blind, controlled study. *J Clin Psychiatry*. 2013;74:1029–1036.

Sutton TA, Sohrabi HR, Rainey-Simth SR, et al. The role of APOE-ε4 and beta amyloid in the differential rate of recovery from ECT: a review. *Transl Psychiatry*. 2015;5:e539.

Tielkes CE, Comijs HC, Verwijk E, Stek ML. The effects of ECT on cognitive functioning in the elderly: a review. *Int J Geriatr Psychiatry*. 2008;23:789–795.

Vaughn McCall W, Lisanby SH, Rosenquist PB, et al; CORE/PRIDE Work Group. Effects of a right unilateral ultrabrief pulse electroconvulsive therapy course on health related quality of life in elderly depressed patients. *J Affect Disord*. 2017;209:39–45.

Vespa J, Medina L, Armstrong D. Demographic turning points for the United States: population projections for 2020–2060. US Census Bureau, P25-1144 Issued March, 2018. Revised February, 2020.

Wilkins KM, Ostroff R, Tampi R. Efficacy of electroconvulsive therapy in the treatment of nondepressed psychiatric illness in elderly patients: A review of the literature. *J Geriatr Psychiatry Neurol.* 2008;21:3–11.

Yew JD, Mulsant BH, Haskett, RF, et al. Relapse during continuation pharmacotherapy after acute response to ECT: a comparison of usual care versus protocolized treatment. *Ann Clin Psychiatry.* 2007;19(1):1–4.

55.4g Psychosocial Factors in Psychotherapy of the Older Adults

JOEL I. SADAVOY, M.D., F.R.C.P.(C)

The term *psychosocial,* as used here, refers to social and environmental elements in an individual’s life that interact with and affect the individual’s mental health and intrapsychic processes. Such elements may be conceptualized at two levels: an intermediate level of factors arising within the immediate environs of the patient (e.g., friendship or family relations) which may change and evolve over time and a macro level of factors associated with more stable societal, ethnic, and cultural norms. A clinician’s knowledge of history can be an important element in establishing a therapeutic alliance with a patient in psychotherapy. The impact of historical events remains a vibrant but unseen influence on the internal world of the older adult patient who often feels that this aspect of their life experience and identity is not valued or understood. While history is not modifiable it is an important aspect of a psychotherapist’s tool kit. It informs therapy by providing context for therapeutic understanding and decision making. For example, for the oldest old, key earlier life decisions such as whether to marry or have children may have been heavily influenced by the experience of the trauma of victimization or combat during wars, local conflicts, terrorism, worldwide economic depression, or political oppression. In turn, such factors can strongly affect the degree to which an individual will age successfully, since marital status and family support are strong predictors of health and well-being in late life. The aging demographic continues to evolve as older adults live longer. The number of centenarians in the United States grew from 53,000 in 2010 to over 90,000 in 2020, with similar increases in numbers of individuals living to extreme old age in other industrialized countries. We are still in the early stages of understanding the psychosocial implications of living to extreme old age. Perhaps of special relevance in this group are the psychological impacts of isolation and loneliness in the context of recurrent loss as individuals outlive most of those around them who once provided love, identity, and companionship.

RELATIONSHIPS

The passage of time leads to major changes in a patient's social network. In youth, time is often viewed as an ally, an opportunity for growth, and a time when new interpersonal roles and novel social partners are preferred. In old age, time is no longer an ally, eroding many life achievements and increasing the preference for emotionally close social relationships. Perhaps most profound are the changes associated with the loss of relationships through death or other factors.

While old age does not inevitably lead to isolation and loneliness, these are important targets of psychotherapeutic and social intervention in late life. Social disconnection is associated with suicide ideation, attempts, and deaths in later life and has been identified as a modifiable intervention target to reduce suicide risk. Loneliness is a related, but slightly different concept which may be defined as a state of emotional distress from lacking desired interpersonal relationships. In numerous studies loneliness and isolation have been correlated with depression, dementia, substance and chronic physical health conditions common in older adults such as hypertension, heart disease, diabetes, and conditions sensitive to chronic inflammatory stress response. Understanding loneliness and isolation as key determinants of physical illness, and psychopathology in older adults can guide psychotherapy. Targeted Interventions include behavioral activation which has been shown to increase social interactions and decrease loneliness among homebound older adults and engage psychotherapy for depression in later life.

There is a convincing body of research in support of the concept that better social relations is related to better health and longer life in older adults. Loneliness and the absence of better social relations are risk factors for psychiatric disorders, particularly depression, and stress-related physical illnesses and disability. A recent U.S. study of over 3,000 adults found that social disconnectedness predicted higher levels of perceived isolation, which in turn predicted higher depression and anxiety symptoms. Conversely, higher levels of depression and anxiety symptoms predicted perceived isolation and social disconnectedness.

What is less clear in the literature on social relations is which elements are most critical, for example, social support, social network, social integration, or social participation. The majority of data support the concept that older adults who are part of what have been termed more socially endowed, diverse, or resourceful networks have better mental health. Such networks are characterized by a greater number and variety of social ties, with whom the individual is intensely involved. It is also unclear whether all relationships confer health benefits. It may be that social relationships with friends and confidantes (so-called discretionary or nonkin relationships) have a greater effect on mental health than relationships with kin, particularly children. There remains some debate as to whether it is the nature of the network itself or whether it is the individual's perception of and satisfaction with support that is the most important protective factor. Both seem to be important factors and should be taken into account when evaluating individual patients.

It seems fairly clear that beyond the nature of the network, the presence of negative interactions is a source of stress while perceived positive relations buffer stress. Hence, it is perhaps not surprising that strong social support helps to buffer depression, and married adults are at a lower risk of depression than unmarried or widowed adults. There is an ongoing debate over the elimination of the bereavement exclusion criterion from a diagnosis of depression in the Diagnostic and Statistical Manual 5. A recent study, for example, has found a high rate of significant depressive symptoms in bereaved older adults who lost their spouse (although major depressive disorder was not associated with bereavement). It is estimated that one in six men and almost half of women over the age of 65 are widowed, and of these about a third will experience complicated bereavement, increasing their risk of depression and associated morbidity and mortality. Moreover, even uncomplicated bereavement can be associated with depressive symptoms which, while they may not reach threshold criteria for a diagnosis of major depression, nevertheless also may impair function and increases their risk of medical comorbidity and mortality. However, psychosocial resources including religious and spiritual belief, and practice, social and family support and Erikson’s concept of vital involvement expressed in commitment to life tasks are moderating or protective factors which might be modifiable and enhanced by therapy.

Bereavement, while almost always difficult, may also be a release from an oppressive relationship, liberating the remaining spouse to resume old pursuits or take on new roles and responsibilities. For example, one spouse spoke eagerly about her ability to return to the academic research and teaching she had relinquished due to the demands of providing care to her ailing spouse.

During care of older adults, as with younger adult patients, it is important to make no assumptions about the significance of loss but rather to explore the specific meaning of loss to each patient. It is important, too, to be aware that the bereaved spouse may need time and energy to develop new roles and relationships and they may struggle with how to identify themselves beyond being a spouse and partner. Other social changes may emerge that are supportive following bereavements. For example, widowed older women receive more help from friends than married women while widowed older men are the most likely to receive support adjusting to practical and emotional changes following the loss of their spouse.

Beyond bereavement, marital relationships in old age also are changed by physical or cognitive impairment, and shifting patterns of dominance and responsibility. For example, the basic nurturing role previously carried out by a loving wife may now fall into the unaccustomed hands of her husband in the face of her decline with Alzheimer disease. He experiences not only loss, but also the need to offer comfort and care, behaviors and feelings with which he has always been awkward. Guilt, resentment, and some anger sometimes emerge as a consequence. Conversely, the physical or mental decline of one spouse may elicit kindness and nurturing late in life that was absent in earlier years. For example, a haughty, narcissistic wife became “chastened” by her awareness of her own decline. While she agonized over her own losses, she also realized that the husband she previously demeaned and devalued was now her greatest ally. He welcomed the new role and the respect he began to feel from her, revealing a nurturing, loving side she had never seen before.

Sociocultural trends in relationships will impact the aging process. For example, in North America, there has been a trend, especially among well-educated, high socioeconomic groups, to delay child-bearing in favor of focusing on career and financial advancement, only to find that fertility has been significantly reduced. This, in turn, often leads to inadvertent and unwanted childlessness and deep late-life regrets. In addition, greater participation of women in the workforce may preclude them from taking on the more traditional role of caregiver and further contribute to the relative lack of children able to support aging parents. Cultural patterns of marriage have also changed. Japan’s almost universal marriage tradition stood in contrast to patterns in Europe, where marrying at a later age or not marrying has been much more common. Interestingly, however, Japanese women are increasingly delaying marriage and childbirth, with a high percentage of the unmarried expressing the intention to never marry. This inevitably leads to a dramatically lower birthrate. The result will be a growing cohort of childless older adults in many societies. Childlessness would seem to disadvantage individuals as they age, but research suggests that childless individuals compensate for not having the support of children by strengthening relationships with friends and collateral kin such as siblings, nieces, and nephews. Nevertheless, in times of increased need for support in old age, childless older adults are left without adequate levels of care. Women are more likely to be childless than men, and childless older adults of both sexes are much more likely to live alone or in institutions and have fewer face-to-face contacts than their peers who have children. However, interestingly there appears to be little difference in psychological well-being between childless seniors and those who are parents.

A social safety net can mitigate social pressures associated with old age. For example, in the United States social insurance provides the financial resources to allow many single older adults to live alone. Similarly, housing alternatives for seniors including congregate living, supportive housing, and community-based social welfare services reduce isolation and dependency on others. However, it is not only practical support that is crucial to the mental well-being of older adults. Rather, for many, it is the presence of true confidants that is key to warding off demoralization and depression. In this regard, informal care networks may be equally or more relevant.

Research on the impact of interventions to enhance social inclusion and reduce isolation is of limited quality. Still the evidence that exists suggests that interventions such as music and singing, intergenerational programs, art and culture, and multiactivity interventions may be associated with positive health outcomes. Such data may support psychotherapeutic interventions that address depressive withdrawal, age-related demoralization, and poor social engagement.

Despite enhanced social support, older adults, especially women, without supportive families often face an emotionally impoverished old age. This is particularly true if governments erroneously assume that families will care for older adults. Childless older adults or those emotionally or geographically distant from their children will obviously be disproportionately disadvantaged by such policies and more likely to be alone. Clearly, these issues have a central effect on the focus of psychotherapy. Problem-solving or interpersonal psychotherapy approaches for isolated older adults will construct and address a different array of problems and issues than for older adults who have strong support structures, such as the management of issues of abandonment or deficits in intimacy.

The nature of intimacy similarly shifts, often in the direction of distance and isolation. The equal partnership between married spouses may shift to a caregiver—care recipient relationship with consequent loss of mutuality and intimacy. Sexuality is a particular victim of the switch from spouse to caregiver. Friendships, easily maintained during younger, more active years, often become much more difficult to sustain, especially in the face of physical impairments. Driving is more of a challenge, evening and night travel more fearsome, and gathering places less welcoming and accommodating. Healthy and active older adults are more able to maintain equitable friendships where both friends can contribute equally. More frail older adults receive more than they give, becoming so-called “overbenefited” in their friendships. Equitable friendships or relationships in which older adults (women) give more than they receive lead to higher morale than in overbenefited women. Interestingly, older men seem to establish more equitable relationships than women.

Affiliation and intimacy seem to be more important to women than to men, although such data should be approached with caution as there have been few studies on the needs of men for intimacy in old age. Some work has shown that, when compared to married men and married and widowed women, widowed men less often have relatives living close by and are less likely to be members of community organizations but may be no more likely to describe themselves as lonely. The data on the importance of intimacy has not been well verified for all cultural groups. For example, there's some evidence that Black women place more importance on functional aid (as opposed to intimate affiliation) than White women. Studies of Japanese older adults have demonstrated an association between social support and self-rated health and mortality, results that were also noted in studies in Sweden and Finland.

Friendships are more than just emotionally supportive. Data suggest that the existence of a close friend and to a lesser extent kinship contacts increase the likelihood of maintaining functional independence.

The pattern and meaning of change in friendship networks in old age are different for men and women. For example, married men may derive much of their self-esteem from their career and family and value friendships significantly less than married women. In old age, men may establish increased casual friendships but there is little indication of increased close relationships and overall data suggest that men (in contrast to women) decline in many measures of friendship. After retirement, men may establish casual friendships but turn to their spouses for personal meaning and self-worth previously derived from their work. With the death of a spouse, married men may be particularly vulnerable to isolation because of their relative lack of friendships and other community connections developed earlier in life. In addition, their spouse may have been their primary or only confidante and may also have been their link to the community. Married women may be relatively protected in bereavement because they are more likely to have other confidantes, both relatives and friends, and belong to community organizations. As a result, they have the opportunity to find other meaningful roles and relationships that may mitigate the loss of their spouse. However, for both men and women, social support from family or friends can only go so far. It is more difficult to lessen the internal loneliness that stems from the loss of the person who is often the spouse’s most significant attachment figure and companion. Indeed, other relationships may not be able to alleviate that loneliness and the loss of the life partner must be mourned and therapy may provide an important venue for working through. While loneliness is a natural human experience it may not be easy to reliably measure using standard instruments. Why is this? While lack of social contact and support are key antecedents that can lead to loneliness, it is not equivalent to social isolation. Social isolation is an objective state and more easily identified and quantified while loneliness is an internal dysphoric and unpleasant experience which is related to the experience of deficiencies in social relationships; for example the feeling, right or wrong, that while there may be many people around to talk to, nobody truly understands. This is a common expression of loneliness among older adults such as those who are caring for a spouse with dementia. There are a number of loneliness measures both direct (self-report of loneliness) and indirect (measuring factors associated with loneliness without using the word itself to avoid the problem of stigma of self-identification) which yield different results. For example, a direct question revealed greater prevalence of loneliness among older women while an indirect questionnaire showed greater incidence among older adult men.

Illness and other age-related events like bereavement take their toll on friendship networks. These networks become depleted although the number of contacts may be less important than their quality. The maintenance of fewer emotionally meaningful relationships may provide sufficient social connectedness and protection against loneliness. Nonetheless, even for naturally outgoing personalities, the avenues for establishing and maintaining new friendships are often blocked. One outcome is the erosion of relationships in which the old person’s history, accomplishments, desires, and ambitions are known by their friends. Increasingly, aging fragments and isolates. Data on adaptation in the oldest old are not consistent. The oldest old, in contrast to the young-old, often become overburdened by illness and frailty, and even previously adaptive personality types with strong positive self-assessments often cannot successfully adjust to their disabilities. Some data indicate that subjective well-being can be sustained through positive self-referential attitudes. Results of these changes include physical distance from longstanding friendships and often, although not invariably, reduced frequency of contacts, which increase feelings of aloneness and create reliance on more casual opportunistic relationships such as with neighbors. While certainly not universal, this process of change contributes to creating a special context imposed by old age on the therapeutic process. Old patients, often through no fault of their own and despite their efforts, may present with a burden of aloneness. This is especially true for those who are unmarried or bereaved. This can create in therapists the anxious sense that they must be more than a treater, responding to an unspoken request for friendship and becoming for the patient the only place and person to offer a safe, intimate emotional haven. The therapist may find these challenges to be somewhat daunting as they may stir up the therapist's own fears about illness, incapacity, and death. The therapist may feel overwhelmed and worry about being engulfed by the patient's helplessness and dependency. In response, the therapist may distance themself from or reject the patient, labeling them as chronically ill, beyond help, and incapable of change. The reverse reaction may also occur in the therapist who wishes to halt or turn back the aging process. In this situation, the therapist may see therapy as life sustaining or even lifesaving. As a result, they may avoid termination rather than face the inevitability of the patient's and, perhaps, their own death.

Blanket rules about friendship and its impact in old age are inadvisable since the nature of friendships differs for each person depending on their psychological makeup. People differ in their need for, as well as their capacity to have, friends; their ability to maintain them; the nature and intensity of their friendships; their ability to do without them; and their standards for with whom they allow intimate confiding relationships. For example, Donald Winnicott suggests that those who never fully invested in friendships earlier in life may be better able to "make do" in old age with more superficial acquaintances without experiencing aloneness. Such individuals may be more invested in activities and personal expressions of creativity that are not dependent on contact with others (i.e., a well-developed capacity for self-soothing and self-sustenance). The implication for psychotherapy is that therapists are most effective when they explore in depth with patients the nature of their friendship networks, their satisfaction with the level of intimacy and support, and strategies either for mourning lost relationships that cannot be replaced or for ways of enhancing their friendship networks to fulfill their needs. Many patients do not volunteer such information in therapy or may not be aware of their needs; hence, it must be elicited and explored actively. No specific psychotherapeutic techniques have been developed for dealing with these issues, but by extrapolation from other work (e.g., on depression), interpersonal psychotherapy, cognitive-behavioral approaches, problem-solving therapy, or perhaps reminiscence therapy may be effective tools.

Assessing social isolation includes evaluating ease of relating to others; feeling of isolation from others; extent of sharing feelings with a significant other; connecting with others when needed; feelings of engagement with others; and feelings of aloneness and friendlessness.

Beyond friendship, sexual and other physical intimacy may diminish or disappear from the life of the old person—no one to touch or be touched by. The literature suggests that sexuality among older adults can remain vibrant and pleasurable although the determinants of sexual activity are varied and may change over time. For example, in one large study positive determinants of sexual activity included a positive attitude toward sexuality, sexual debut before age 20, having a very happy relationship and a physically and mentally healthy partner, good global health and mental health, being married/cohabiting, and drinking alcohol more than three times a week. Sexual frequency and interest is also enhanced when there is a change of partner in old age. Sexuality in older men is less researched than women and often focused on the impact of erectile dysfunction and there is limited cross-cultural, gender identity, and sexual preferences research. In the psychotherapeutic context, sexuality is often rich therapeutic territory but discussion of sex in old age may challenge the openness of the patient and the skill and beliefs of the therapist. Data continues to support a conclusion that seniors are embarrassed or otherwise avoidant of talking about sexual problems with health care professionals who may be viewed as uncomfortable, not knowledgeable, or disinterested.

Positive attitudes to sexuality appear to be maintained in old age. While true as an isolated piece of data, findings also must be seen in the full psychosocial context, in which older adults over a certain age are often widowed and without partners and with limited opportunity to form new close bonds. The lack of a suitable partner is the key predictor of sexual activity among women. One study found that 82% of residents of subsidized independent-living facilities had not had intimate sexual contact in the past year, although many had at least one experience of hugging or other touching. Hence, for example, intimate touching may become the domain of the medical encounter during physical examination or the nonsexual hugging of children, grandchildren, or friends. Physical isolation becomes especially evident in retirement or nursing homes or other institutions where predominantly old women live in an unusually gender-homogenous environment. Whatever sexual or intimate interest remains is put in cold storage, often forever. It is also true that sexual activity and pleasure vary dramatically, often directly associated with the availability of sexual opportunities in relationships. Psychotherapists and other caregivers need to be sharply aware of the patient's needs and reactions under these circumstances, elicit the impact on the patient, and incorporate the realities into the treatment, either working toward change or acceptance. The therapist must also be aware of countertransference reactions in this area. Unresolved conflicts over parental sexuality may lead the therapist to avoid any discussion of sexuality or to regard the older client as asexual. As in all psychotherapies, the therapist must monitor their reactions to the patient through active self-examination. By bringing these reactions into awareness, the therapist can minimize their potential to have detrimental effects on the therapy and, instead, use them to better understand the patient.

The psychosocial phenomenon of loss of intimate relationships in old age may have special impact on the form of the psychotherapeutic relationship. Therapists are often and easily experienced by the isolated patient as a wished-for substitute for the lost relationship. Therapists may feel alone with the patient in their depleted world as well as uncomfortable pressures to adopt the roles or behaviors of friend, confidant, protector, advisor, or advocate. The therapist who identifies too closely with the patient's problems may be immobilized by pity or sadness that impairs their ability to empathize with or promote change in the patient. On the other hand, the therapist may shrink away from the patient's sadness, thereby increasing rather than easing the client's isolation and grief.

While old age is a time of substantial change and requires great adaptability, satisfactions may still abound. However, they tend to be encountered within a narrower frame, focused more on family relationships and very close friendships. For many, the loss or unavailability of friendships/relationships promotes a kind of centripetal emotional force that draws some older adult persons into their families and somewhat away from other relationships particularly when faced with the challenges of failing health. The loss of confiding relationships may be especially clinically relevant, since some research suggests that these relationships are protective against depressive disorder. Effective psychotherapy and environmental interventions that support the patient's need for new relationships can counter some of the effects of loss of confidants. While attention to relationships during therapy is important regardless of the patient's age, it is especially necessary with older adults. Assessment and intervention must elicit not only the presence of instrumentally supportive relationships, but also which of them are sustaining of self-esteem and maintenance of identity.

The increasing importance of family relationships with advancing age adds an element to the psychotherapy of older adults not found as prominently among younger adult patients. Therapists often must take into account age-induced dependency issues in families and the conflicts these issues engender, such as forced intimacy and caregiving situations with adult children. For example, conflicts arise over the different perceptions of need that the older adult parent has. Children may be overly anxious and protective, aging parents inappropriately resistive of needed help or exaggerating of their helplessness. While family assistance and caring are crucial supports for the aging person, the therapist must be wary of assuming that family can substitute for friends. Indeed, friendship interaction is positively correlated with self-esteem, while kinship interaction is not. Whatever the pattern, the psychotherapist of older adults must take the family system into account, sometimes involving the other family members directly in the therapy or helping the older adult person develop appropriate interpersonal patterns of interaction. Flexibility of psychotherapeutic technique and interventions is essential. The usual adult–therapist stance broadens when treating older adults as psychosocial issues associated with frailty, dependency, and relational shifts impinge, often rendering individual assessment and focal intervention inadequate. Impelled by these realities, the therapist often must combine overt roles of advocate, intermediary, and educator with the more traditional psychotherapeutic roles and interventions. Sometimes, this means reaching out actively to families, social agencies, or medical caregivers, roles that are often unaccustomed for psychotherapists. While therapists need to be flexibly supportive, they also need to examine their motivations when intervening on their patients' behalf. Transference- and countertransference-induced forces can lead therapists to inappropriately cross therapeutic boundaries, unwittingly adopting the idealized roles of adult–child or spouse seemingly in response to real psychosocial needs, but actually impelled by dimly understood or unconscious dynamic forces in the therapeutic relationship. Therapists dealing with relationship changes in aged patients also must be prepared to deal with their own feelings of loss, grief, fear, and loneliness and wishes to idealize, dominate, or avoid painful psychological issues that may be generated during treatment of the psychosocial stresses of old age.

ROLE CHANGES

Occupational role changes impose relatively unique challenges on older adults. This group is almost universally forced to give up important sources of self-esteem, creative activity, social networks, and the organizing effects of job expectations. The importance of work is exemplified by the finding that nonemployed older women have lower self-esteem. Reactions to retirement differ. Retirement may be eagerly anticipated and handled well, especially among those who used to be called blue-collar workers and those with the resources, resilience, and creativity to use newfound leisure time in a satisfying way. Others, such as some professionals, often avoid retirement, and may find the loss of their occupational roles difficult. Either way, the psychotherapist must be keenly aware of the centrality that work-related roles play in structuring life and the changes to this structure imposed by aging. These factors are especially relevant in the so-called young-old, 65 to 75 years, who are more likely to have freshly encountered the issues of occupational role change. For example, an emotionally distant marital relationship may have been quite stable and acceptable when work substituted for unavailable marital closeness. After retirement the emptiness of this form of relationship becomes much more evident and can lead to depressive reevaluation of life and choices. While difficult, this is also a time of therapeutic opportunity. Since retreat into work-related defenses is no longer possible, patients are forced to confront themselves and reexamine their long-held beliefs about others, especially their spouses.

Here, examining with patients the sources of their feelings both longstanding, deriving from early, perhaps unconscious developmental factors, and more immediate, for example, associated with distorted modes of perception and thoughts about a partner who seems to want to care but is constantly rebuffed, is important. In the more advanced years, these issues generally recede, either adapted to and worked through or overwhelmed by other more pressing issues. However, when fresh, the psychosocial pressures of occupational role change, especially when combined with other events, such as spousal bereavement, often emerge as states of vulnerability, fearfulness, and impaired capacity to seek new, alternative sources of gratification.

Beyond its function as a source of personal gratification and self-esteem, the workplace serves an important social function. For example, one of the important considerations in retiring for women is the finding that they tend to give up work-related relationships after retirement. Working may have disrupted other familial and social roles and may not have allowed time for employed women to nurture or even maintain other relationships. In retirement, women may feel disconnected and that they have no role or purpose in life, particularly as other relationships may no longer be active. Hence, retirement creates a more impoverished social life for women, increasing their vulnerability to emotional reactions that accompany lost relationships.

As implied already, parenting roles may change dramatically in old age. For example, well-established patterns of parental dominance and reliance by children on a strong authoritative parent often suddenly reverse as parents become more frail and dependent. Conflict, guilt, or fear may invade a previously stable system. Particularly vulnerable are family systems whose members adapted to earlier conflicts by establishing distance, for example, children moving away from the pressures created by demanding, unreasonable, critical parents. In such circumstances, the increasing needs of aging parents may impose pressures on children to care for parents for whom they have had longstanding conflicted feelings. The psychotherapist must be careful to examine and understand the parent–child system and beware of taking stereotyped approaches to aging parents since the patterns of parental role changes are quite varied. For example, frailty and fear of being alone and incapacitated often lead older adult patients to complain about the inattentiveness of adult children. Taken at face value, the psychosocial needs of the patient often are indisputable. However, the pattern of the family members' reactions to the patients' needs is rooted in the long history of their relationships with each other.

One daughter found herself trapped providing care to her aging parents and subjected to a pattern of mistreatment well established earlier in her life. She found visits to her parent's home very difficult, but if she didn't go, her mother would repeatedly attempt to contact her and berate her when a visit eventually occurred. The daughter chose to go regularly to forestall intrusions and having to cope with her mother's disapproval and agitation. Her sense of obligation towards her parents kept her involved in a situation where she felt that there was little or nothing she could do to bring about positive change for herself or her parents. Another child spoke about her sadness at what she saw as the "changing of the guard," meaning she was now caring for the people who previously cared for her. To be effective, the therapist must be sensitive to the emotional forces that are preventing the family from adapting to the new, age-induced needs of the parent.

Many older adults, mostly women, become caregivers to a spouse who develops dementia. They are subject to an array of stressors that lead to what has been termed caregiver burden. Burden is associated with disproportionate rates of both psychiatric and physical morbidity. Psychosocial factors as well as personal characteristics impact the resilience of caregivers and their capacity to manage this new and generally unwelcome new role. Dementia caregiving is particularly arduous. It differs from other forms of caregiving in that caregivers of persons with dementia have more unmet needs and strain than nondementia caregivers with higher levels of physical, psychological, social, and economic consequences. They spend significantly more hours per week providing care, report greater impact on employment, greater strain, mental and physical health problems, less time for leisure and other family members, and family conflict. Behavioral and psychological symptoms of dementia (BPSD) are well-recognized sources of stress and burden on family and other informal caregivers of persons with dementia. The emotional effects of caregiving include facing helplessness in the face of this disease while concurrently dealing with chronic grief and mourning. This form of chronic grief is often unrecognized or diminished even by well-meaning others and hence has been called disenfranchised grief. Such grief can become "complicated" and the longer it goes on the more likely it is to induce depression.

Family caregivers concurrently balancing employment with caregiving roles for older adults are a vulnerable subset of caregivers. In the United States, as of 2013, about 60% of family caregivers of people with dementia were employed. A recent survey examined 1,839 Canadian employees who were also caregivers. It revealed that those who were dealing with high or medium levels of caregiving demand for family members with dementia had significantly greater role demands and challenges on measures of perceived stress, depressed mood, burden, self-efficacy, mastery, control, absenteeism, job–caregiving conflict, and role overload than nondementia or low dementia demand caregivers. These results suggest that clinicians dealing with patients and their families will enhance clinical care by identifying these dementia caregiver employees and intervening with psychosocial interventions known to be effective in sustaining and supporting them.

As with other stressors of old age, social support and engagement are strong modifiers of burden as are the availability of instrumental support such as homecare, respite care, and easy access to professional support. These factors are often modifiable elements in the treatment of caregiver burden when incorporated into the management of caregivers. However, beyond these instrumental psychosocial factors, management of caregivers must be integrated with therapeutic interventions to strengthen adaptive, coping skills, and self-efficacy together with enhancing skills for the management of dementia behaviors. It has been shown that fully integrated comprehensive approaches to caregivers which incorporate skills training, problem-solving training, instrumental psychosocial support, and psychotherapeutic principles can be effective in enhancing key caregiver capacities. Recent reviews of the effectiveness of psychosocial and behavioral interventions for informal dementia caregivers revealed pooled effect sizes in favor of the intervention groups on burden, depression, stress, and sense of competence.

SOCIETAL ATTITUDES AND RESPONSES

Biases about aging and the aged abound, and the older individual is confronted with them. Some of the North American perspectives on aging are quite negative—for example, older adults are perceived as no longer productive and therefore marginal to mainstream life or, because of frailty, imposing undue economic burdens on overstretched pension and public support systems. As a result, there may be a tendency to dismiss the problems of older adults or a reluctance to provide help. Perhaps not surprisingly, negative attitudes to aging, both societal and internal self-perception, are risk factors for poor health and well-being.

The effective psychotherapist of older adults incorporates these societally engendered issues into the therapy, recognizing with the patient the nature and impact of the messages and working on appropriate responses. In this role, the therapist can help to buffer the sometimes harsh messages the old person hears daily about those in their age group. A particular issue is a difference between the patient's self-perception—more youthful, involved, and competent—and the perception of younger adults who can only see and react to the apparently universal aging appearance—lost skin turgor, decreased muscle mass, gray hair, and postural change. In other words, the older person's external and age-related life issues obscure the underlying depth of feeling and identity and can delay a more complete and deeper understanding by the young of the truer self-perceptions held by the older individual. Older patients in psychotherapy often feel both misunderstood by society at large and quietly ashamed of themselves for their diminished capability and resultant helplessness, loss of purpose or role in life, or the burden they may feel they impose on others. In the presence of shame, individuals may fear new social situations and avoid them. The therapist will be most effective if these issues are articulated and explored with aging patients, helping them counter the societally induced loss of self-esteem and feelings of alienation by confronting distortions and developing techniques to defend against inappropriate stereotyping by those around them.

The COVID-19 pandemic of 2020 led to abrupt societal changes and imposed life transitions on everyone. Older adults have been particularly exposed to the implication of social distancing measures which have disrupted social networks and provoked isolation and dislocation. These changes have been widely recognized as detrimental to the mental health of older adults. Interestingly, however, studies of the impact of COVID-19 on older adults in several western countries early on in the pandemic suggested that this group may be psychologically more resilient than younger groups despite their vulnerability to more severe complications, higher mortality, disruptions to daily routines, and access to care. Similar to findings of studies in Spain, Netherlands, and Canada, the Centers for Disease Control and Prevention (CDC) survey of 5,412 community-dwelling adults across the United States in June 2020 found that mental health resilience was higher among the older adult group. In this study, participants aged 65 years or older reported significantly lower percentages of anxiety disorder (6.2%), depressive disorder (5.8%), or trauma- or stress-related disorder (TSRD) (9.2%) than young participants. It should be emphasized however, that these data do not take into account those from underrepresented minorities, lower socioeconomic groups, informal caregivers, or those with pre-existing mental health problems like depression or dementia. However, these findings are broadly consistent with studies that have shown that older adults are as resilient and perhaps more so as younger age groups.

Psychotherapy that addresses resilience can be helpful in uncovering sources of personal strength in the background and experience of the older adult patient and map them onto the current challenges. In isolation-inducing COVID-19, for example, exploring with the patient their natural kinship and friendship networks and helping them determine strategies and routines for reaching out safely and appropriately can be useful.

It is unwise to interpret old age as a risk factor in itself. For example, in this disguised but real situation, an 85-year-old man was locked down in his home because of COVID-19. He had pre-existing mobility problems and had personal support worker (PSW) home care. His wife had died 2 years previous, and his vibrant social life had come to a halt when COVID-19 shut down all social options. His therapist was especially worried about his ability to cope as this shutdown was a reprise of his prolonged quarantine with his family in the 1940s polio epidemic, crippling his father, killing his mother, and disabling him for a time. Remarkably, the prior experience, while devastating at the time, did not overwhelm or weaken him. Rather it promoted in him a sense that he had already gone through the worst and could face this new challenge.

Even the massive trauma of the Holocaust strengthened the coping and resolve of another woman, now in her 90s, who had been a hidden child when her parents were murdered by the Nazis. In her mind, the adversity of being locked down and isolated to her room in a supportive housing unit while evoking early memories of her war experiences was “nothing” compared to what her parents had gone through during transport and in the concentration camp almost 80 years before. The re-evoked memories of the Holocaust, while not easy for her to bear, were already a familiar part of her emotional landscape, and she showed remarkable resilience in the face of the new COVID-19–induced social disruption and isolation.

ECONOMIC AND ENVIRONMENTAL FACTORS

For many, old age is a time of economic and environmental change. Retirement and the many difficulties sometimes imposed by living on fixed incomes, social security, or supplements provided by children can induce real or fantasized anxiety and preoccupation about financial security. These factors can influence psychotherapy in a number of ways, for example, inducing envious transferential feelings in patients or guilty countertransference in therapists.

Financial restrictions, often real but sometimes imagined, lead to restricted social activity, feelings of isolation, humiliation, and shame, as well as uncertainty and fear of the future.

Like advancing frailty or bereavement, economic difficulties must be actively dealt with in psychotherapy. This involves determining with patients their actual financial situation, separating distortions from reality, uncovering latent meanings about financial security, and then introducing practical approaches to coping and interpretive understanding of the underlying meaning of the situation to patients.

The living accommodation is another important psychosocial variable that often changes in old age. The key element that induces psychological stress associated with moving to a new setting in old age is that it is generally a forced move, precipitated by physical or cognitive decline, economic factors, or isolation. Even more than death, the specter of abandonment to an impersonal chronic care facility is most feared by older adults. The prospect of such a move provokes various forms of abandonment anxiety, including the knowledge that this probably will be their last abode before death. This emotional reaction becomes more acute for married couples when the move necessitates the separation of the partners and placement in different settings appropriate to each one's care needs.

Moving to a long-term care facility may also generate internal conflicts associated with forced interaction with other older adult residents. Emotional responses may occur for a variety of reasons. For example, aged patients routinely express their anxiety at being confronted with the pain and disability of their age peers. In them, they see mirrored their own future and become afraid. These feelings can be powerful motivators, preventing patients from accepting realistic moves to settings where they can receive needed care. Pressing a reluctant, frail patient to accept needed help in an institution can pose ethical dilemmas for the therapist. Not the least of these is the reality that many or most of these institutions are barely adequate to the task of satisfying the needs of their residents. Concerns about the capacity of nursing homes to care for and protect vulnerable older adults were brought into stark and tragic relief by the COVID-19 pandemic. By May to June 2020, prior to vaccinations which began to protect this most vulnerable sector of the population, seniors in nursing homes accounted for a disproportionate number of deaths in North America and Europe. Canada had the bleak distinction of leading the way, with almost 80% of COVID-19–related deaths occurring in nursing homes, while in the United States, the figure was 40%. These statistics inevitably added to the struggle of seniors and their families as they weighed the risks and benefits of moving to a long-term care facility, including whether they could trust the system to watch over their safety and welfare.

Inevitably, no matter what the setting, the new resident entering long-term care will encounter severe psychosocial stressors, including some degree of loss of individual identity, regimentation to the institutional routines, and loss of privacy. Confronting these realities can be difficult not only for the patient but also for the sensitive therapist who is only too aware of the limitations of the resources upon which they must rely.

The psychotherapist of an aged patient facing environmental dislocation will be most effective if they can uncover the patient's fears and anxieties —abandonment and loneliness, shame, infirmity, pain, or death—and work toward both realistic appraisals of the future and development of skills to combat the specific feared outcome of the environmental transition.

ETHNOCULTURAL ISSUES

Culture, ethnicity, and race have a profound impact on aging, which must be understood and dealt with by psychotherapists, especially when practicing in geographical areas made up of diverse communities. For example, older adults among racial/ethnic minorities are less socially engaged, with reduced access to social groups and nonchurch-related activities. This appears to be associated with greater vulnerability to a less robust recovery from illness and mortality. Data suggest that racial/ethnic minority residents of long-term care facilities are less socially engaged than white residents, although nursing home residents, in general, have low levels of social engagement. The staff of nursing home facilities are from diverse cultures, and there may be a mismatch of cultural knowledge, values, and norms between staff and residents, which adds to isolation. In the United States, language ability may be an important factor in the mental health of seniors. For example, low English language proficiency has been associated with increased levels of depression among Mexican American older adults.

Depending on the specific community, the presence of mental illness may carry with it a profound stigma that stops both patients and their families from seeking care. These traditional cultural values are likely to be more evident in less acculturated members of ethnocultural communities. The therapist must be careful to evaluate the level of acculturation of a given patient and avoid stereotypical presumptions about cultural values, for example, assumptions about family support and filial piety.

When patients from minority ethnocultural communities do seek care, they often come with socioculturally sanctioned feelings of shame at their inability to cope with their emotional problems. Compounding the more universal problematic aspects of aging, aging members of ethnocultural communities also must deal with barriers to mental health care imposed by language- and immigration-related features, such as the feelings associated with leaving their homeland, unanticipated family conflicts, loss of status, or inaccessible spiritual support. Often, newly arrived or even well-established but poorly integrated, aged immigrants are carrying a heavy burden of isolation, for example, unfamiliar with how to travel beyond their neighborhood, unduly reliant on often overburdened family members, feeling like aliens when they may have expected to feel like respected older adults, encountering racist attitudes, or struggling with inadequate finances. While perhaps self-evident, it is crucial to the establishment of an effective therapeutic alliance for psychotherapists to convey to the patient that they are sensitive to these issues by enquiring about them in an empathic, caring, and specific manner. Therapists must be aware of the influences of their own culture on their choices, assumptions, and prejudices, guard against cultural superiority, and be willing to adopt an attitude of cultural humility and flexibility in adjusting their practice to accommodate cultural concerns. Most important is the overt demonstration to patients that the therapist recognizes them as individuals who will not be stereotyped by virtue of their culture, ethnicity, or race.

Psychotherapy and alliance building with the patient is aided when the therapist is sensitive to culturally determined reactions to and perceptions of illness and its causes, such as beliefs about traditional treatments or approaches to those who are ill, acceptance and trust of mainstream services, culturally embedded terms and expressions about illness, and values about openness versus shame with health care professionals.

Therapeutic approaches and techniques are influenced by cultural factors. Traditional cultural perspectives on aging and mental health forge the individual’s view of the determinants of mental health and, in turn, the acceptability of and trust in therapeutic interventions. To be effective, the therapist suspends preformed assumptions about the older adult's background and culture and explores in an open, inquiring manner how cultural values influence explanations for symptoms and illnesses, diagnosis, and treatment.

Therapists evaluate the degree to which culturally determined communication patterns affect therapy and adapt accordingly. These factors include four forms of communication patterns:

1. Proxemics, the study of the perception and use of personal space (e.g., Anglos are more distant, while Latinos, Black Americans, and the French want much closer stances)
2. Kinesics, the study of bodily movements—for example, facial expression or eye contact (e.g., among some Asians smiling is often an expression of embarrassment rather than mirth, and smiling may be seen by them as a sign of weakness rather than friendliness)
3. Paralanguage, which refers to vocal cues such as loudness of voice or inflections, and is easily misinterpreted (e.g., loud, rambunctious behavior is common in some groups such as Americans but is easily misunderstood as aggressive in low-volume cultures such as Japanese)
4. High–low context communication, which refers to overt, intellectualized communication (low context) versus subtle or hidden communication rooted more in situations or internalized in the person. High-context communication is exemplified by how a mildly spoken "yes" can actually mean "no" when spoken by a Filipina challenged by her superior. In this context, her "yes" is the expression of someone who is afraid to contradict her superior.

Once the patient's cultural perspectives are understood, the therapist's task becomes one of synthesizing this understanding and helping the patient integrate their traditional views with Western medical concepts (e.g., integrating the patient's view that they is suffering because of failure to honor ancestors with Western concepts of guilt). The successful, culturally competent psychotherapist will be able to translate the patient's metaphors and belief systems, understanding both their concrete and symbolic meaning. For example, understanding that because physical symptoms are acceptable while mental symptoms are not, the patient may only feel able to convey distress in physical terms such as, "I have no energy," "My heart is uncomfortable or heavy," and so on.

Decision making in many cultures is group (not individually) based, as is more common in Western culture. It is necessary for therapeutic success that the therapist recognizes the importance of family hierarchy in decision making in some cultures and maintain an authentic stance of nonjudgmental collaboration with both patient and family.

An example of cultural knowledge that can be essential to the psychotherapy of a traditional, older adult Native American, for example, is the centrality to the Native American of sacred tribal land. Tribal land usually refers to that which is occupied by an individual’s tribe prior to their resettlement. Being torn away from traditional lands was a universal experience for Native Americans, disrupting many aspects of basic culturally based practices. For example, most religious ceremonies, cultures, and customs revolve around tribal lands. Clearly, this form of dislocation, even though it occurred long ago, has had a profound effect on the personal identity of individuals, their sense of belonging in the world, their feeling of trust in those who are in power, and their sense of belonging within the wider society.

Of course, no one can hope to know all about every culture, but one basic ingredient is recognition that the older adult patient struggles with many culturally determined issues and conflicts that they may feel they have no right to speak about openly. For example, among groups who experience racism, older adults may feel uncertain about how safe it is to reveal their anger about white dominance and may view the treating doctor as a symbol of the very oppression and loss that they suffered in the past. Early in the therapy, therefore, it is useful to clearly indicate acceptance of the patient’s uncertainty and, where appropriate, actively open such issues for discussion. This can be reinforced in the first assessment when the therapist can ask about the patient’s feelings about traditional beliefs, experiences with the mainstream, including racism, the difference between their private thoughts and their public beliefs, and their sense of belonging in society and in therapy.

Psychiatrists and other therapists will be most effective if they are able to take a truly inquiring attitude toward the patient and be open to the patient’s explicit attempts to educate them. In so doing, the therapist must also take care that the patient is not taking refuge in cultural differences to explain away all emotional reactions. It is important to carefully distinguish between patients’ reluctance to reflect on and talk about themselves and true difficulties with expression. Having said this, it is also true that patients who fluently speak a language do not always use that language effectively for emotional expression. Patients may not be able to fully express their emotional state in English if it is their second language and can only convey the true meaning when they revert to their language of origin. This sometimes is relevant, especially in the early stages of dementia if language skills begin to decline slightly.

Another important cultural competence of the therapist is the ability to accurately gauge the degree of closeness or distance required by the patient. The willingness or ability of individual patients to delve into important areas such as sexuality or interpersonal conflict with family members is often colored by feelings of shame and humiliation and prohibition against discussing such matters beyond the family, some of which are culturally determined. It may be helpful to preface inquiry into these areas by first acknowledging with patients their potential sensitivity and asking how they feel about exploring such issues. In other words, it is important to address the underlying culturally determined resistance that the patient may bring to the therapy before attempting to get to the substance itself.

The effective therapist treating patients from another culture will have the capacity to match the nature of the therapy to the level of belief and sophistication about Western methods that the patient holds. For example, the idea that there are social and interpersonal origins for feeling is not universally held around the world. Moreover, the implication that early life experience affects later life development may pose culturally determined problems in the therapy.

Factors associated with immigration may have special relevance to psychotherapy. For example, often hidden but of great psychological importance, especially in refugee groups, are histories of severe deprivation and trauma associated with war, political oppression, or famine. Because psychological trauma is a function of life experience, the prevalence of psychological trauma in older adult will vary based on the life experiences of particularly aged cohorts, for example, those who lived in times of political oppression, war, or deprivation versus those who lived in relative peace and tranquility. One recent Canadian study, for example, found that 11.1% of the older adult patients presented with posttraumatic stress symptoms, but less than 2% met DSM IV criteria for full posttraumatic stress disorder. The data are clear that late-life effects of earlier adult trauma endure, but their presence must be actively sought out by the therapist, as patients frequently keep silent about their experiences and their persisting effects.

SPIRITUAL ISSUES

Frequently omitted from psychotherapy inquiry, spiritual issues may play a profound role in the lives of many older adult patients. Religious or philosophical beliefs often are a source of great comfort in enabling appropriate acceptance of change, facing aloneness, or confronting death. Aging patients may return to religion in later years, both as a source of comfort and as a social resource. Some suggest that isolated older adults may benefit from becoming more involved in religious activity. While attending to spirituality and religious needs may be important, it is often not addressed in psychotherapy, and the use of pastoral counseling is limited. Psychotherapy will be more acceptable to the patient and often more effective when it is carried out with a carefully elicited understanding of the patient's beliefs and attempts to use the strengths inherent in the patient's faith as an ally in treatment. However, the therapist must take care to deal with religion from the patient's perspective, not their own.

FUTURE DIRECTIONS

Regardless of the type of psychotherapy employed, the therapist of the aged patient must understand, acknowledge, and incorporate into the treatment the array of psychosocial factors that form the context of the patient's problems. The complex set of factors taken together is the matrix of the patient's life within which they will be supported or failed. In psychodynamics terms, this matrix may be viewed as the holding environment, the adult version of that which is provided by an infant's parents. As with children, when the holding environment fails in its supportive and enhancing functions, the individual comes under psychological stress. It is the person in their environment that is reflected in the environment within the person. Psychotherapists will be most helpful and effective when they work simultaneously with the external and internal worlds of the older adult patient.

FURTHER READINGS

Beckman N, Waern M, Östling S, Sundh V, Skoog I. Determinants of sexual activity in four birth cohorts of Swedish 70-year-olds examined 1971–2001. *J Sex Med*. 2014;11:401–410.

Bruce LD, Wu JS, Stuart S, Russell DW, Nemecek DA. Loneliness in the United States: a 2018 national panel survey of demographic, structural, cognitive, and behavioral characteristics. *Am J Health Promot*. 2019;33:1123–133.

Erzen E, Çikrikci Ö. The effect of loneliness on depression: a meta-analysis. *Int J Soc Psychiatry*. 2018;64(5):427–435.

Ginsberg TB, Pomerantz SC, Kramer-Feeley V. Sexuality in older adults: behaviours and preferences. *Age Ageing.* 2005;34(5):475–480.

Gitlin LN, Marx K, Stanley IH, Hodgson N. Translating evidence-based dementia caregiving interventions into practice: state-of-the-science and next steps. *Gerontologist*. 2015;55(2):210–226.

Hawthorne G. Measuring social isolation in older adults: development an initial validation of the friendship scale. *Soc Indic Res*. 2006;77:521–548.

Inzitari M, Risco E, Cesari M, et al. Nursing homes and long term care after COVID-19: a new ERA?. *J Nutr Health Aging*. 2020;24:1042–1046.

Kim G, Kim M, Park S, Jimenez DE, Chiriboga D. Limited English proficiency and trajectories of depressive symptoms among Mexican American older adults. *Gerontologist*. 2019;59:856–864.

Koenig HG. The role of religion/spirituality in the mental health of older adults. In: Sadavoy J, Jarvik L, Grossberg G, Meyers B, eds. *Comprehensive Textbook of Geriatric Psychiatry*. W.W. Norton;2004.

Kreager P. Where are the children? In: Kreager P, Schrüder-Butterfill E, eds. *Ageing Without Children: European and Asian Perspectives*. Berghahn Books; 2004:1.

Lamont RA, Nelis SM, Quinn C, Clare L. Social support and attitudes to aging in later life. *Int J Aging Hum Dev*. 2017;84:109–125.

Leigh-Hunt N, Bagguley D, Bash K, et al. An overview of systematic reviews on the public health consequences of social isolation and loneliness. *Public Health*. 2017;152:157–171.

Leong F, Lee S-H. A cultural accommodation model for cross-cultural psychotherapy: illustrated with the case of Asian Americans. *Psychotherapy (Chic)*. 2006;43(4):410–423.

Li Y, Cai X. Racial and ethnic disparities in social engagement among US nursing home residents. *Med Care*. 2014;52:314–321.

Lonnèe-Hoffmann RA, Dennerstein L, Lehert P, Szoeke C. Sexual function in the late postmenopause: a decade of follow-up in a population-based cohort of Australian women. *J Sex Med*. 2014;11:2029–2038.

Lutz J, Van Orden KA, Bruce ML, Conwell Y; Members of the NIMH Workshop on Social Disconnection in Late Life Suicide Social Disconnection in Late Life Suicide. Social Disconnection in late life suicide: an NIMH Workshop on State of the Research in identifying mechanisms, treatment targets, and interventions. *Am J Geriatr Psychiatry*. 2021;29(8):731–744.

Mendes de Leon CF, Gold DT, Glass TA, Kaplan L, George LK. Disability as a function of social networks and support in elderly African Americans and Whites: the Duke EPESE 1986–1992. *J Gerontol B Psychol Sci Soc Sci*. 2001;56(3):S179–S190.

Nicolaisen M, Thorsen K. Who are lonely? Loneliness in different age groups (18-81 years old), using two measures of loneliness. *Int J Aging Hum Dev*. 2014;78(3):229–257.

Pin S, Guilley E, Spini D, d'Epinay CL. The impact of social relationships on the maintenance of independence in advanced old age: findings of a Swiss longitudinal study. Z *Gerontol Geriatr*. 2005;38(3):203–209.

Polenick CA, DePasquale N, Eggebeen DJ, Zarit SH, Fingerman KL. Relationship quality between older fathers and middle-aged children: associations with both parties' subjective well-being. *J Gerontol B Psychol Sci Soc Sci*. 2018;73:1203–1213.

Préville M, Lamoureux-Lamarche C, Vasiliadis H-M, et al. The 6-month prevalence of posttraumatic stress syndrome (PTSS) among older adults: validity and reliability of the PTSS scale. *Can J Psychatr*. 2014;59:548–555.

Ronzi S, Orton L, Pope D, Valtorta NK, Bruce NG. What is the impact on health and well-being of interventions that foster respect and social inclusion in community-residing older adults? A systematic review of quantitative and qualitative studies. *Syst Rev*. 2018;7(1):26.

Sadavoy J. Survivors. A review of the late life effects of prior psychological trauma. *Am J Ger Psychiatry*. 1997;5(4):287–301.

Sadavoy J, Sajedinejad S, Chiu M. A quasi-experimental study of the effectiveness of the Reitman Centre CARERS group intervention on family caregivers of persons with dementia. *Int J Geriatr Psychiatry*. 2021;36(6):811–821.

Sadavoy J, Sajedinejad S, Duxbury L, Chiu M. The impact on employees of providing informal caregiving for someone with dementia. *Aging Ment Health*. 2022;26(5):1035–1043.

Santini ZI, Jose PE, Cornwell EY, et al. Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older Americans (NSHAP): a longitudinal mediation analysis. *Lancet Public Health*. 2020;5(1):e62–e70.

Sikorski C, Luppa M, Heser K, et al; AgeCoDe Study Group. The role of spousal loss in the development of depressive symptoms in the elderly—Implications for diagnostic systems. *J Affect Discord*. 2014;161:97–103.

Sinković M, Towler L. Sexual aging: a systematic review of qualitative research on the sexuality and sexual health of older adults. *Qual Health Res*. 2019;29(9):1239–1254.

Tang F, Heo JG, Weissman M. Racial differences in social engagement and health status among older women. *Soc Work Public Health*. 2011;26:110–122.

Walker RB, Luszcz MA. The health and relationship dynamics of late-life couples: a systematic review of the literature. *Ageing Soc*. 2009;29:455–480.

55.4h Individual Psychotherapy

JOEL I. SADAVOY, M.D., F.R.C.P.(C), AND LAWRENCE W. LAZARUS, M.D.

Psychotherapy for the older adult has evolved in step with the evolution of general adult psychotherapy. Any debate over whether psychotherapy is as effective for the older adult as it is for younger patients is now largely resolved. The data are clear from an array of outcome studies and meta-analyses that most forms of psychotherapy are demonstrably effective in older adults treated for depression and anxiety. However, as with psychotherapy in general adult populations, there are few evidence-based guidelines that help the clinician determine which form of psychotherapy will work for a given patient. Some suggest that older adults prefer shorter-term therapies, although clinical experience indicates that there remains wide variation in what older adults want and expect from therapy. Hence, therapists will be most effective if they have an array of techniques to choose from and determine on a case-by-case basis the specific form of therapy to be used.

Overall meta-analytic reviews of randomized controlled trials (RCTs) of various psychotherapies for older adults have confirmed that the psychotherapies are generally effective. Effect sizes (ES) derived from meta-analytic studies of psychotherapy in older adults range from 0.63 to 0.78, with higher values for specific interventions, such as 0.85 for 7 studies of cognitive therapy and 1.05 for 8 studies of reminiscence therapy compared to no intervention or placebo. Life review (LR) is a more systematic technique of reminiscence (R), and data indicate it is more effective. In LR therapy for depression in older persons, the ES is more than twice that of reminiscence (LR-ES 1.28 vs. R-ES 0.52) and four times greater when well-being is measured (LR-ES 1.02 vs. R-ES 0.24).

Because of the complexity of the challenges of late life that impact the adaptation of older adults and their vulnerability to emotional disturbance and disorders, individual psychotherapy often needs to integrate other interventions into the therapeutic process. In addition to manualized models of psychotherapy in older adults, models that incorporate combinations of therapy may be necessary to address the complex factors affecting older adults' adaptation. For example, social support, sometimes mediated through the use of simple technology like the telephone and psychoeducational models, has shown positive synergistic effects with psychotherapy on depression outcomes, and systematic telephone-based psychotherapy has been effective in reducing distress in aging dementia caregivers (CGs). More recently, video conferencing for psychotherapy has become more common. However, the use of this technology has introduced complicated elements for treating older adult individuals. Data suggests that the older adult in the United States often experience internet-based interventions as unacceptable or inaccessible because of complicated technology that they have difficulty learning, lack of access to the internet, or cost considerations. Dealing with intense, emotionally charged therapeutic issues online challenges both patient and therapist to maintain a therapeutic connection without the human proximity of in-person contact. These challenges have been brought into sharper focus by the COVID-19 epidemic of 2020–2021. Because of public health restrictions on travel and face-to-face interactions necessitated by COVID-19, practitioners were forced to turn to technology to deliver psychotherapeutic services. Funding limitations, which have been a barrier to the provision of online services, were significantly eased during the COVID-19 epidemic by Medicare and Medicaid in the United States and some other jurisdictions. For example, the Canadian Federal Government provided funding to support online access to health care, and in Ontario, the publicly funded Ontario Health Insurance Plan (OHIP) has funded all live technology-enabled services during the COVID-19 epidemic. While broadened funding regulations have been introduced as temporary measures, the wide acceptance and apparent effectiveness of technology-enabled clinical service for the older adult, including psychotherapy, suggests that some measures may be adopted long term. One problem is that evidence for the effectiveness of these interventions has not caught up with the widespread use, although initial studies are promising. For example, a recent systematic review has found that online cognitive-behavioral therapy (CBT), problem-solving therapy (PST), and behavioral activation are both acceptable and effective for the older adult.

Psychotherapies are effective treatments for depression in late life and may sometimes be considered a first-line intervention. Studies comparing the efficacy of various individual therapeutic approaches for depression in geriatric outpatients (e.g., cognitive-behavioral, interpersonal, problem-solving, reminiscence and LR, and psychodynamic psychotherapy) have not yet been able to demonstrate the superior efficacy of one approach over another. At the Palo-Alto V.A., for example, a comparison of cognitive, behavioral, and brief psychodynamic therapy for depressed older adult outpatients found comparable remission ratios (53% had remitted after about 18 sessions). At 2-year follow-up, 77% of those whose depression had initially remitted were still depression free. The majority of patients who did not respond to the initial treatment remained depressed at the 2-year follow-up despite continued treatment. An important variable limiting reliable generalizability of the effectiveness in depression of the most well-researched psychotherapies in the older adult, CBT and PST, is the fact that these studies have generally been conducted on young-old cohorts which are well white and well educated.

Having said that, data suggest that older adults prefer psychotherapy to medication for depression treatment. Depression is often comorbid with other medical comorbid conditions, and psychotherapy has shown effectiveness in depression with various comorbidities, including chronic obstructive pulmonary disease, heart failure, Parkinson disease, and stroke, among others. Similarly, evidence suggests that psychotherapy can be effectively delivered in a variety of health care and social settings. While choosing specific forms of psychotherapy for individual patients of any age remains a refined and somewhat unreliable clinical skill, choosing an intervention for the sometimes complex presentations of the older adult requires even more care since it must take into account psychiatric, medical, functional, and cognitive status, together with the settings in which depression is likely to present, for example, long term care, hospital wards, and community health care centers.

For more severely depressed outpatients, a combination of psychotherapy with an antidepressant may be more effective than either alone. Interpersonal psychotherapy (IPT), for example, has shown effectiveness in enhancing social adjustment when used in combination with an antidepressant. Even here, however, initially, positive research outcomes have been called into question by more recent data. One problem in interpreting the data is that the variables of psychotherapy, such as frequency of intervention, are not always standardized from one study to another, limiting the generalizability of specific studies.

Creative, neurobiologically informed behavioral psychotherapeutic interventions that require little training and are readily delivered in community settings show promise. An example is Engage, an evidence-based, manualized behavioral intervention that promotes engagement in previously meaningful, rewarding activities that may have been impeded by depression through simple interventions. Trained Engage therapists use “reward exposure” to promote the activation of positive valence systems, dysregulation of which may promote depression through increased negativity bias, apathy, and emotional dysregulation. In each session, similar to PST methodology, patients prioritize two to three activities that are meaningful to them, brainstorm ideas on how to pursue them, and make a concrete implementation plan. Data suggest that this approach is effective, readily learned, and adopted by clinicians of various types in community settings.

Meta-analysis of evidence-based psychosocial therapies for anxiety has revealed support for CBT, cognitive therapy, relaxation training, and supportive therapy both for self-reported as well as diagnosed anxiety in older adults at post-treatment. One comprehensive analysis reported an overall mean ES for treatment versus no treatment or placebo control of 0.55, just above the median (0.5), providing some, although not robust, support for psychological treatment of late-life anxiety.

The strongest support is for CBT for generalized anxiety disorder (GAD) but is weaker for subjective anxiety, where relaxation training has demonstrated more efficacy especially when practiced long term. Meta-analytic study shows support for the effectiveness of CBT for GAD but CBT does not clearly emerge as a better treatment when compared to active controls.

Age-related brain changes probably affect the response of older adults to psychotherapy although empirical data are modest. For example, reduced bilateral cortical thickness may be a marker of individuals at higher risk for poor response to psychotherapy.

In this section, the authors review the history of, and theoretical basis for, psychotherapy with the older adult, give brief descriptions of various therapies with illustrative clinical vignettes and suggestions for future research.

HISTORY

Modern psychotherapy began with Sigmund Freud’s attempt to construct a universal model of understanding human behavior. In the prebiologic era of psychiatric treatment, psychoanalytic theory formed the basis for most psychotherapeutic interventions. Unfortunately, Freud held a particularly gloomy but highly influential view of the potential for the psychoanalytic treatment of patients over the age of 50. He believed that there was too much material to be dealt with and that the older adult had inelastic mental processes and hence were not educable. These ideas produced inordinate pessimism about the potential for psychological change in older adults. As the psychotherapies matured and began to be applied to a broader range of patients, therapists came to recognize that age per se does not preclude therapeutic change. Indeed, in the 1950s and 1960s, several analysts wrote enthusiastically about their positive experiences in treating older adult patients. In retrospect, this enthusiasm resulted from the treatment of generally vigorous older adult patients who were good candidates for analytically oriented psychotherapy.

THEORETICAL ISSUES

Effective psychotherapy, at a minimum, appears to require a strong, early, working alliance regardless of which specific psychotherapy technique is used. The working alliance is commonly understood to be agreement on the goals and tasks of therapy in the context of a positive affective bond between patient and therapist. Older adults, not unlike younger patients, vary in their expectations and understanding of psychotherapy. They often come to psychotherapy with general wishes to feel better but uncertain what to expect. As with younger patients, the alliance in psychotherapy is enhanced when: the therapist steps back with the patient to inquire about the nature and form of the problems; helps the patient define agreed-upon goals for therapy, and articulates them clearly such as reduction of anxiety symptoms or improved relationship with a significant other; there is a personal bond between therapist and patient, in which the patient views their therapist as caring, understanding, and knowledgeable; both patient and therapist agree on the means to achieve goals.

Alliance is further strengthened when the therapist implicitly or explicitly conveys to the patient their understanding and knowledge of the age-related tasks and challenges facing the patient. There are many models for framing this understanding, depending on the type of therapy being used, but at the root, each model makes use of common themes now understood to underpin the adaptational challenges of those negotiating the geriatric phase of life.

A major developmental task of aging is finding restitution for the inevitable biopsychosocial losses associated with this stage of the life cycle. Therefore, loss is one of the most prominent issues discussed by older adult patients. For many, it is the cumulative effect of multiple repeated losses, before sufficient time has passed for mourning and resolution, that is so devastating. The older person struggles to maintain ego integrity and a sense of purpose in life at a phase of the life cycle often characterized by diminishing coping resources, increasing self-doubt, and the onslaught of assaults and challenges to self-esteem and a positive self-image. The magnitude of an older adult person's reaction to a loss depends, to some degree, on the amount of pride and emotional investment in the lost function. Beyond lowered self-esteem, older adults experience life changes that can damage their sense of security and safety. For some, the world becomes uninviting and dangerous, particularly if they have encountered real events such as falls, personal attack, or inaccessible help at a time of need.

Perhaps counterintuitively, fear of death is not as prominent in the older adult as in younger adults and hence is not a major theme in geriatric psychotherapy. Clinical experience suggests that other fears especially focused on health concerns, incapacitation, memory loss, pain, abandonment, and helplessness are more evident. For example, in one psychotherapy study cohort of seniors over age 60, reasons for seeking treatment were depression related to interpersonal relationships, health conditions, grief/loss, finances, housing, and challenges due to executive dysfunction.

Self-worth, security, and independence depend in part on the individual's ability to maintain physical capacities. Beyond obvious challenges posed by physical illness or decline, impairment of important psychological functions becomes pronounced when frailty or illness forces restricted function and imposes a new, different, and often unacceptable self-image and personal identity.

Psychological adaptiveness in old age varies widely. Vulnerability to states of depleted self-esteem, insecurity, fear, or anxiety is heightened when an individual cannot adapt flexibly and is too wedded to maintaining or retrieving that which must be relinquished. Problems of psychological adaptation and diminished self-esteem occur in late life because of such factors as overdependence on social or work roles, overemphasis on the maintenance of control and mastery in all situations, and unresolved problems with self-esteem arising from earlier developmental stages. Difficulties with self-esteem may be rooted in maladaptive, unmodified beliefs; unconscious issues such as dependency conflicts; or interpersonally generated psychological reactions like unresolved grief.

The majority of older adult persons adapt quite well to the challenges imposed by negative societal stereotypes about aging and the losses noted above that threaten to erode self-esteem. Several psychological self-protective mechanisms help to maintain a positive self-image, such as focusing on past successes, discounting messages that do not fit with lifelong self-concepts, refusing to apply general myths and misconceptions about aging to the self, and choosing to interact with those who provide pleasurable self-esteem–enhancing experiences. Some selectively screen out unwanted perceptions such as those that challenge impaired cognitive abilities. While often adaptive and self-protective, such selectivity may also distort reality and promote maladaptive denial.

GENERAL TECHNIQUES

In the initial interview, the psychiatrist explores the major reason the patient has come for the interview and is active and empathic in exploring patient and family resistance to therapy. The patient and family are asked to bring with them all currently used prescription and over-the-counter medications. Psychotherapy with the older adult can only be successful if it is integrated with effective management of relevant somatic factors affecting adaptation and emotional responses. Hence, careful attention is paid to previously overlooked medical problems and medication adverse effects that may contribute to current psychiatric signs and symptoms.

The older adult patient is usually interviewed individually before important information and concerns are obtained from the family with the consent of competent patients. Older adult patients respond positively to the therapist's listening, empathic, and encouraging attitude. Some patients like to sit close to the therapist and may welcome such expressions of spontaneous warmth as a natural touch on the patient's arm. However, the therapist evaluates each patient's sensitivity to intimacy and does not assume that familiarity is welcome just because the patient is old. Inquiring about how the patient has successfully coped with adversity in the past helps to recall effective coping strategies and offers hope that current stresses and losses can be similarly overcome.

Psychologically less sophisticated patients, who are accustomed to talking about their physical complaints when visiting a physician, initially may feel more at ease if the therapist begins with a medically oriented focus. The therapist may then shift the focus to psychological and social problems.

In general, the therapist of the older adult patient is more active, engaging, supportive, and flexible than with younger patients. The therapist is knowledgeable about the normative, age-related changes that occur in the psychological, cognitive-intellectual, and social life of older people; geriatric medicine; psychopharmacotherapy; family therapy; and community resources. In both office and hospital settings, it is especially useful to work with a mental health colleague, such as a social worker, who is knowledgeable about family therapy and community resources. Patients must feel that they have obtained something from the initial interview (e.g., empathic understanding; hopefulness) to increase the likelihood that they will return for subsequent sessions. Telephone reminders to selected patients or family members in advance of the next appointment impart the therapist’s interest and improve patient compliance. Older adult patients, particularly the frail and dependent, may hesitate to express any negative reactions or doubts about therapy. The therapist might wonder out loud whether the patient may find it difficult, for a variety of reasons, to return for a second visit. Such exploration increases the likelihood that therapy will continue.

Realistic, explicit treatment goals are negotiated with the patient and, as appropriate, family. If treatment goals are overly ambitious, both patient and therapist may experience frustration and disappointment. Many older adult patients, whether consciously or unconsciously, have a specific treatment goal in mind, which once accomplished, leads them to terminate treatment.

SUPPORTIVE PSYCHOTHERAPY

More frail patients, such as those who are cognitively impaired or chronically ill, may require such psychotherapeutic techniques as support of healthier defense mechanisms, ventilation, advice, help with acceptance of diminished capacities, and increased dependency needs. Rather than trying to achieve intrapsychic change, the therapist strives to improve and refine the environment to the particular needs of the patient. For example, the therapist may help the family to arrange a more caring, nurturing environment and assist them to become better CGs.

The patient’s life history and psychodynamics (obtained from the patient or family members) enable the therapist to institute specific supportive therapeutic strategies.

An older adult woman with progressive cognitive impairment and lifelong personality traits of dependency and passivity is anxious and depressed about entering a nursing home. She may adapt well to the home if assigned to a roommate who is nurturing and protective.

Alvin Goldfarb has described a brief, supportive therapy technique for institutionalized, cognitively impaired patients. The therapist improves patients' foundering self-esteem, sense of control, and safety by permitting them to develop an apparent special relationship with the therapist, who is perceived as benevolent and powerful. The patients believe they have some control over the benevolent physician. This is accomplished in small, subtle ways. For example, the physician may elicit the patient's preferences for the frequency of sessions, daily timetables, or diet. Then in a realistic or symbolic way, the therapist acquiesces to the patient's wishes while maintaining a quiet caution about being perceived as being unduly manipulative.

INTERPERSONAL PSYCHOTHERAPY

There are not a lot of data supporting the efficacy of IPT in older adults. However, what there is, suggests it is useful as either a primary intervention for depression and possibly as an adjunct to antidepressants although this latter data are inconsistent.

Nevertheless, older adults often have to cope with loss of friends and family, to adapt existing social skills to changing life circumstances, and to resolve conflict with significant others, especially adult children. Hence, a therapy that helps the older patient cope with and resolve interpersonal problems would seem well suited to older adult patients. Recent cross-sectional study data in older adults confirmed that loneliness is a significant risk factor for depression in both men and women although especially so for women. Interestingly and perhaps counterintuitively, social network size was not associated with the risk of depression, although this is consistent with other findings, that is, it is the level of perceived intimacy and social support that is most critical rather than the overall size of the social network.

IPT is a time-limited (usually 16 weeks), manualized psychotherapy that focuses on one or two interpersonal problems, such as the death of a family member. It has been adapted and studied in the acute, continuation, and maintenance treatment of late-life depression.

In a series of controlled studies at the University of Pittsburgh, 79% of older adult outpatients with major depression who were treated in the acute and continuation phases with a combination of IPT and nortriptyline achieved full remission. Over the subsequent 3 years, 80% of the patients with recurrent major depression who had remitted and who continued on combined treatment, compared with 10% who received placebo, were still in remission.

LIFE REVIEW OR REMINISCENCE THERAPY

There is considerable empirical support from many RCTs of the efficacy of individual LR and reminiscence therapy for the treatment of depression, compared to no treatment control groups although support for group intervention is not strong. Along similar lines, there is some evidence that autobiographical memory practice can reduce symptoms of depression when compared with “usual care.”

Reminiscence is characterized by the progressive return of memories of past experiences, especially those that were meaningful and conflictual. It is theorized that reviewing the past helps the patient find greater meaning in the present. To varying degrees, older adult patients in therapy reminisce about the past, search for meaning in their life, and strive for the resolution of past interpersonal and intrapsychic conflicts. LR therapy is the therapeutic promotion of reminiscence by systematically enhancing this reminiscing process making it more conscious and deliberate. The therapist may guide the process by encouraging the patient to write or tape an autobiography. Techniques include reunions with family and friends and looking through memorabilia such as scrapbooks or picture albums. This technique has been reported to resolve old problems, increase tolerance of conflict, relieve guilt and fears, and enhance self-esteem, creativity, generosity, and acceptance of the present. Perhaps obvious is that the effectiveness of reminiscence interventions requires sensitivity to the individual and sociocultural history of older adults, which may be obscured for younger therapists by the passage of time or unfamiliarity of therapists with the backgrounds of immigrant older adults. Culturally appropriate reminiscence techniques have shown efficacy for symptoms of depression, anxiety, and low self-esteem among Taiwanese older adults.

Some authors have cautioned about possible negative consequences of encouraging LR, particularly for a depressed patient with obsessive tendencies or life histories filled with multiple misdeeds not amenable to a positive review. Evoking memories in patients who have suffered massive trauma or whose lives have been seriously affected by deprivation, overwhelming abandonments or humiliation will be emotionally challenging for patients, and therapeutic technique is directed to providing an environment of safety and trust so that the evoked emotion can be contained, modulated, and perhaps reframed to enhance self-acceptance.

COGNITIVE-BEHAVIORAL THERAPY

Older adults are a heterogeneous population. There is no consensus on how much adaptation is necessary for the CBT technique for older adults. However, it is probable that when frailty, sensory impairment, or cognitive decline is present, therapy will be facilitated if some adaptations are employed. For example, with some older adults, the therapist is often more active with a focus on here-and-now issues. The pace of therapy is sometimes slower than with younger adults, and when appropriate, the therapeutic alliance may be enhanced by permitting the patient to adopt a senior teaching role with the younger therapist. To compensate for specific deficits such as hearing loss or cognitive slowing, information may be presented in different modalities, such as written material or taped sessions for the patient to take home. Homework has been found to be feasible with many older adults and helpful to the process, including the use of formal thought records. With older patients, material may have to be reviewed more often, and termination is often managed gradually rather than abruptly.

There is some evidence to suggest that CBT is most effective for older outpatients with mild–moderate depression who are relatively healthy with high motivation for treatment and absence of a personality disorder. These characteristics are not unlike those of candidates for psychodynamic therapy and IPT.

PROBLEM-SOLVING THERAPY

Four RCTs investigated the effectiveness of PST for older adults. Three found that PST was better than other therapies and better than no treatment although not all trials are positive. There is some indication that PST is especially useful for depression associated with executive dysfunction where problem solving and decision making may be particularly impaired. A large comparative trial of PST and supportive therapy found that the more structured and focused intervention of PST versus a less structured supportive method was more effective for treating late-life depression associated with executive dysfunction.

PST has been adopted for use in a variety of disorders and older adult populations with demonstrated effectiveness. It has been shown to be effective in reducing symptoms of poststroke apathy compared to placebo, late-life major depression compared to supportive therapy and reminiscence therapy, minor depression and dysthymia compared with paroxetine and placebo, and vision-related quality of life in patients with macular degeneration.

PST is a brief solution-focused intervention that draws on methods commonly employed in other modes of structured therapy such as CBT or IPT, that is, direction and coaching, hierarchical lists created, homework assigned, relating problems encountered to evoked emotions, situational analysis of antecedents and precipitants for the problem, and patient-generated solutions guided by the therapist are core to the method. Problems can range from major adverse life events, such as bereavement or illness, to everyday coping such as transportation, banking, or shopping. Once a problem is identified and its determinants are specified, a brief list of the pros and cons of potential but realistic solutions is generated by the patient and therapist taking into account the focal nature of the therapy and the time constraints of the brief format. Patients with mild cognitive impairment can utilize PST if they retain sufficient cognitive capacity to absorb the method and retain key elements from one session to the next. They may require additional supportive strategies, such as written notes, memory devices, environmental adaptations, and CG involvement.

PST may be most effective when delivered face to face either individually or in group formats but other acceptable and feasible methods of PST delivery have been shown to be effective for harder-to-reach patients, such as Internet-based therapy for depressed home-care older adults. Improvement in quality of life and a reduction in anxiety symptoms have been demonstrated through the use of videophone in hospice CGs.

PST has been widely adopted for use with informal CGs of those with dementia improving coping skills, self-efficacy, and depression. Frontline workers dealing with patients with dementia and their families in the community are often overwhelmed by the complexity of problems they must address during relatively short interventions. When professional workers in the community are trained in PST, it has been shown that, compared to a control group of CGs in standard care, PST training helps them organize their approach to complex problems. This, in turn, appears to enhance their ability to help CGs, thereby significantly improving the capacity of CGs on key measures of function and emotion, such as coping, mastery, and competence, while significantly reducing emotion-oriented coping, burden, and stress.

BRIEF (TIME-LIMITED) PSYCHODYNAMIC PSYCHOTHERAPY

Brief or time-limited psychodynamic therapy can be considered for older adult patients with clearly defined, age-related problems that can be expected to resolve within a brief period of psychotherapy, such as an adjustment disorder, unresolved grief reaction, or new-onset anxiety disorder. Setting a time limit at the beginning reinforces the patients' confidence in their ability to resolve problems within a circumscribed period of time, focuses and accelerates the therapeutic process, diminishes the fear of dependency, and accommodates limited financial resources. Brief therapy uses the therapist's psychodynamic understanding of the patient and the patient's transference to clarify and help resolve the patient's emotional reactions to current life stress.

In a pilot study of the process and outcome of brief psychodynamic psychotherapy of patients with an adjustment disorder with depressed mood, the women showed greater improvement and were more active and assertive than the men, who tended to become more dependent on the therapist. The patients used the therapist to validate competency and normalcy, which helped to support and restore feelings of mastery and self-esteem. Symptomatic improvement often exceeded the achievement of insight or self-understanding.

INSIGHT-ORIENTED PSYCHOTHERAPY

While studies of dynamic psychodynamic therapies are relatively sparse in number, several RCTs of psychodynamic therapy (including brief dynamic therapy) for older adults have shown that it is an effective treatment for depression comparable to other forms of therapy, such as CBT in one head-to-head trial. Overall, data suggest that about 70% of people who had received dynamic psychotherapies were in remission at a 2-year follow-up, broadly consistent with meta-analytic data on the overall efficacy of psychotherapy in older adults. A rich array of both earlier and recent literature describing successful clinical cases of psychotherapy with the older adult convincingly supports empirical data and the position that, for appropriately selected cases, age per se does not interfere with the effectiveness of psychodynamic therapy, including psychoanalysis.

The selection of patients and the basic approach to insight-oriented therapy for older patients who are physically and cognitively intact is largely the same as that for younger patients. Appropriate patients have the ability to form a therapeutic alliance, the capacity for introspection and to mourn losses, and the motivation for change. Older patients may be particularly suited for this therapy given their appreciation of the finitude of life, increased tendency toward introspection, and a more philosophical view of themselves and life.

During longer-term psychotherapy, the condition and circumstances of the older patient evolve and change. The development of physical illness may require a medical referral, collaboration with family members, or advocacy on behalf of the patient for community and other support services. Such active intrusions into the patient’s real-life problems are not commonly dealt with by psychodynamic therapists, but when treating older adult patients, techniques often need to be adapted to their changing needs. Therapists are most effective when they learn to integrate different therapeutic modalities while maintaining a psychodynamic focus on the therapeutic relationship, the transference, and the goal of inner conflict resolution. Indeed, the necessity to effectively integrate different modalities of treatment is one of the hallmarks of the therapist’s flexibility necessary to treat and manage the evolving therapeutic issues of older adult patients.

MULTICOMPONENT PSYCHOTHERAPEUTIC INTERVENTIONS FOR FAMILY CGs

Older adults are commonly confronted with undertaking caregiving roles with spouses who develop dementia. It has been estimated that 75% of community-based caregiving is provided by informal family CGs. In the United States, most of these family CGs are older female spouses who are disproportionately vulnerable to the well-established negative effects, both physical and psychological, of dementia caregiving leading this group to be so-called hidden or invisible second patients. These dementia CGs must cope with the emotional and practical problems posed by a progressive, untreatable disorder that gradually erodes the fabric of the relationship with their care recipient (CR), causing attendant emotional loss and grief. Over the years, the CG must learn to navigate a new, unfamiliar, and unwelcome lifestyle, increasingly trapped in a role that forces them to cope with an illness that induces behavioral and psychological distress as a disturbance in the CR, alters their behavior and personality, and impairs independent functioning, capacity for communication and decision making. Attention to the psychological morbidity of CGs is clearly important, both to prevent care breakdown and institutionalization of the CR and to sustain the CG.

Intervention methods for supporting the health and well-being of dementia CGs have included education and psychoeducation, peer-support groups, CBT, and family counseling. A recent review concluded that there is support for the clinical and cost-effectiveness of multicomponent programs. CBT and multicomponent intervention outcomes for CGs with depressive disorders were positive, with a large ES.

Group and family-focused psychotherapy have been shown to be effective for dementia CGs incorporating psychoeducation, cognitive reframing, communication skills training, and professionally led group interventions. Examples of evidence-based, successful, scalable multicomponent programs are the Canadian program called CARERS, an acronym for **C**oaching, **A**dvocacy, **R**espite, **E**ducation, **R**elationship, and **S**imulation, and the New York University Caregiver Intervention (NYUCI).

CARERS is a manualized, integrated multicomponent group psychotherapeutic, skills-training intervention demonstrated to be an effective intervention for improving coping, mastery, depression, and burden in family CGs. The psychotherapeutic group intervention is designed to systematically address three key areas of CG need: problem-solving, communication, and behavioral management and the emotional impact of dementia caregiving. Following an in-depth intake assessment, each CG participates in eight weekly 2-hour closed group sessions of four to six CGs led by a trained mental health clinician. A predetermined agenda structures the outline of each group while the content of each group is determined by the CGs, based on an initial go-round check-in to elicit current caregiving problems and challenges and emotional reactions and foster group cohesion. In sessions 1 through 4, CGs are taught a form of problem-solving therapy adapted for dementia CGs; communication and behavioral management are addressed in sessions 5 to 8. Specially trained standardized patients participate with each CG in tightly controlled, live therapeutic simulation interventions during which the group therapists coach CGs and lead group interaction focused on the observed emotional consequences of communication breakdown and isolation. Group psychotherapy principles inform all aspects of CARERS to foster the presumed curative factors of group psychotherapy, including ensuring safe emotional expression, instillation of hope, universality and reduced isolation, imparting information, altruism, interpersonal learning, and group cohesiveness. Since its inception in 2009, more than 300 therapists have completed a systematic, manualized course of face-to-face and e-learning training. All aspects of group intervention and training have been adapted to be delivered both face-to-face and live online. CARERS is demonstrably scalable and has maintained effectiveness and fidelity when disseminated to multiple sites in Canada and the United States. It is the core intervention program for the Ontario Ministry of Health-funded program called Enhancing Care for Ontario Care Partners.

The NYUCI for CGs is a well-researched, evidence-based, demonstrably effective program comprised of a detailed intake assessment followed by two individual and four family counseling sessions, support group participation, and telephone counseling for both the primary CG and family members. Postintervention phone or email support is provided indefinitely to address evolving changes as dementia progresses. Several studies have shown its effectiveness in reducing CG depression and delaying nursing home admission.

OTHER THERAPIES

Mindfulness-based cognitive therapy (MBCT) and mindfulness-based stress reduction (MBSR) have been increasingly used interventions for the management of anxiety, depression, and stress-related symptoms. Recent data suggest that, when compared to waitlist controls, MBSR may be effective in older adults, although the quality of studies is often low. In a province-wide intervention in Ontario, Canada, for family CGs of persons with dementia living at home, a manualized telemedicine version of MBCT has been successfully modified for delivery online and is currently in the early stages of service rollout. Participants, many of whom are older women, value the accessibility of the online telemedicine format and readily adopt the intervention. A small preliminary feasibility study has shown that the intervention is acceptable and effective in improving self-compassion and coping among CGs when compared to usual care controls.

CLINICAL ISSUES

Indications for Individual Psychotherapy

Older adults are a heterogeneous population. The continuum ranges from the normative or healthy older patient whose overall functioning is retained to the frail older adult patient. Five categories are normative older adults; normative older adults in crises; independent frail older adults; dependent, physically frail older adults; and dependent, cognitively impaired, frail older adults. Indications for various modalities of individual therapy are strongly influenced by the position of the patient on the normative–frail continuum. Moreover, because capacities change over time, therapeutic approaches for a given patient may also need to change.

With advancing age, a crisis is frequently the stimulus for seeking out psychiatric care, often initiated by a family member. A small qualitative study found that common reasons for older adults seeking psychotherapy included depression related to interpersonal relationships, health conditions, grief/loss, finances, housing, and challenges due to executive dysfunction, that is, mild cognitive dysfunction.

Individual therapy during crisis focuses on mourning for lost important relationships or personal capacities and working through conflicts over separation, redirecting energy to more realistic goals, redefining relationships, especially with regard to accepting appropriate dependency and reestablishing or developing appropriate strategies to enhance mastery, sense of control, and self-esteem. Family education about ways of helping the impaired older adult is of critical importance. Psychotherapy during a crisis is often integrated with medication and environmental therapies. Problem-solving techniques, brief IPT, and CBT are useful approaches during crises, although longer-term supportive or psychodynamically oriented therapy may evolve.

Behavioral techniques using operant behavioral conditioning have been applied to a wide range of target symptoms in chronically ill older adult patients, including incontinence, pain, insomnia, wandering, socially inappropriate behaviors, reduced level of activity and participation, toileting, bathing, and feeding. Environmental cues are manipulated to increase, decrease, or shape behaviors. These techniques are useful for institutionalized and frail older adult patients to reduce the excess disability that often accompanies dementia. An ABC approach to behavioral management of disturbed behavior may be useful. This technique includes accurate neuropsychological evaluation; defining specific measurable and observable behavior for modification; determining the antecedents (A) to the onset of the disturbed behavior and systematically recording them; recording the type, frequency, and duration of the baseline (B) conditions; tailoring a treatment program (A and B) to alter the behavior, which includes CG counseling, the introduction of reinforcing, and extinguishing techniques, that is, consequences (C); instituting environmental changes and psychotherapy to assist CG; and program evaluation. A major deficiency of behavioral interventions is their labor intensiveness and the need to maintain the program. Once the interventions are withdrawn, behaviors tend to return to pretreatment levels in these institutionalized patients. Cognitive-behavioral interventions in long-term care settings, while often effective when introduced by a consultant, generally fail if the consultant leaves. Frontline staff in these settings usually lack the training to sustain sophisticated cognitive-behavioral programs on their own.

Various forms of individual psychotherapy have shown therapeutic value for states of bereavement. Bereavement can lead to chronic states, sometimes called complicated or prolonged grief, which is characterized by the persistence of acute grief symptoms complicated by cognitive, emotional, and behavioral symptoms. Complicated grief is common, found in 9% of bereaved older women. Psychotherapeutic interventions are particularly effective for bereavement complications such as chronicity, depression, or anxiety disorders. Most interventions include some education, support, attention to the relationship with the deceased before death and as it has changed since, and supportive assurances of “normality” for bereaved individuals who need to be reminded that their feelings and experiences are not deviant or abnormal. Complicated grief therapy (CGT) has structured this process and includes grief monitoring, imaginal and situational revisiting (promoting memory and using aids such as pictures or situations), imaginal conversations, building personal emotional and instrumental supports, defining and planning personal aspirations and opportunities for self-care, and promoting positive emotions. RCT data suggest that, when compared to IPT, CGT may be highly effective for the older adult.

Psychotherapy of various types has been recommended for the treatment of anxiety disorders. Preliminary data suggest that CBT can be quite effective in older patients with panic disorder and may thus be considered as an alternative to pharmacologic treatment for selected patients or as an adjunct to pharmacologic treatments when patients have behavioral symptoms that can be targeted.

Psychotherapy has shown some value in the management of older adult patients struggling with alcohol abuse. The inclusion of cognitive-behavioral training that addresses themes of self-advocacy, self-esteem, and relapse prevention may aid recovery.

Individual psychotherapy is often helpful in the treatment of sexual dysfunction. For example, premature ejaculation usually responds readily to sex education, reassurance, and sex therapy techniques practiced at home alone or with a cooperative partner.

Limitations

Older adults are underrepresented in psychiatric outpatient clinics (2% to 4% of patients seen) and even more so in private psychiatric practice. Older patients are often not referred for therapy, possibly because primary care physicians are not fully aware of its effectiveness in the age group. Overall a minority (25% in one study) of depressed older adults receive psychotherapy, but even of these, only one-third got consistent, ongoing therapy. Young–old, more highly educated older adults are more likely to receive psychotherapy.

The application of psychotherapy, therefore, appears to be limited by the willingness of practitioners, both psychiatric and primary care, to use it for older adult patients. Completion of a series of regular therapy sessions, whether short or long-term, is often limited by practical considerations such as difficulties with insurance and personal finances, transportation, or weather.

Raised in an era when shame was associated with psychiatric problems and treatment, some older adult persons feel embarrassed and indignant when their physician or family suggests consultation with a psychiatrist. Many older adult persons, as well as health professionals, believe that depression and anxiety are expected concomitants of aging or routinely attribute abnormal psychological states to physical illness. The symptoms themselves, such as depression, apathy, and hopelessness, prevent older persons from seeking treatment. Patients whose ethnoracial backgrounds differ markedly from that of their therapist may have quite different notions and expectations about a visit to a psychiatrist. Many cultures view mental disorders as shameful. Individuals may fear damage to their families' reputations. For example, older adult patients from some ethnic backgrounds have expressed fear that their grandchildren may be less marriageable if there is a mental illness in the family. Beyond cultural barriers, other access barriers limit care for these groups, including language incompatibilities, lack of knowledge of where to obtain psychiatric services, economic issues, a tendency to seek care only after a crisis, and fragmented, difficult-to-access mental health care services.

Adult children of an aging parent may minimize the parent's psychiatric symptoms, be ambivalent about assuming a parenting role, and may share their parent's negative attitudes toward psychiatry. All these barriers result in impaired access to care, poor compliance with medication and psychotherapy, and a high dropout rate.

Strategies to overcome these formidable barriers in the very first session include active exploration of patient and family resistances to treatment, discussion of how ethnic differences between therapist and patient can interfere with treatment, ambivalent and negative attitudes about psychotropic medication, and encouraging the patient to share reactions to the first session. If the therapist intuitively suspects the patient is resistant to treatment, active inquiry about problems that may interfere with subsequent sessions, such as shame or practical issues like transportation, are explored. It is important that the patient feels that something has been achieved in the very first session. In this way, the seeds of the therapeutic alliance are planted with both the patient and family in the initial visits. Even for the very impaired, depressed patient bereft with hopelessness, conveying hope and a willingness to help can facilitate the beginning of a therapeutic relationship.

Complications

Whatever the form of psychotherapy, the course may be complicated by issues of transference and countertransference. The patient's feelings about therapy will be influenced by education, socioeconomic status, culture, and ethnicity, which can restrict or enhance the impact of treatment.

Transference is age-independent (of both patient and therapist), the determinants stemming from the timeless unconscious. However, developmental theory expands the traditional perspectives by demonstrating that transference derives not only from significant infantile experience but also from internalized objects encountered in adult life, for example, a spouse or a child.

While there are no age-specific transferences, the psychodynamic impact of life experiences during old age may increase the emergence of certain forms. For example, frailty leads to dependency conflicts and fears of abandonment, emotions that can easily evoke parental transference reactions toward the health professional derived from the reactivation of the patient's need for a protector (parental figure). This form of transference emerges readily in dependent, institutionalized, or otherwise isolated patients, even when psychotherapeutic contact is relatively brief and intermittent. Bereavement, especially of a spouse, or other loss (e.g., sexual capacity) may lead to an eroticized spousal transference to the therapist in which the patient unconsciously relates to the therapist as an erotic love object who becomes a fantasy substitute for the lost object.

Reverse transference describes a phenomenon of geriatric psychotherapy in which the patient experiences and relates to the therapist as a child and takes a parental role toward the therapist in the treatment dyad. Therapists may experience this transference interaction as demeaning or patronizing and feel uncomfortable with the seeming reversal of dominance. Young therapists may find this especially challenging. It is generally a transient phase of therapy that defensively protects the patient from deeper levels of transference-based affect associated with conflicts over dependency.

Somewhat different is filial transference, in which patients deal with the therapist as one of their own children, enacting such unconscious conflicts as rage for not fulfilling wishes for caring or guilt derived from earlier failures in the parental role.

Countertransference responses of therapists are often intense in the treatment of older adult patients, sometimes blocking the therapist's willingness to engage the patient in individual psychotherapy. The complexity or chronicity of problems of late life may lead therapists to be unduly concerned that older patients will become dependent on them and engender some therapists' reluctance to embark on treatment for fear of the patient's dependency. Similarly, it is not uncommon for younger therapists to avoid treating older adults because of their anxiety over having to deal with frailty or decline in older patients. Uncertainty about their ability to help the older person may induce feelings of guilt or anxiety in the therapist and lead to withdrawal from or avoidance of the provision of psychotherapy.

Engaging at an emotionally deep level with older patients becomes even more challenging for therapists who are struggling with problems that echo those of their patients. This dynamic has been brought into stark focus during the recent COVID-19 pandemic as patients and therapists alike have been forced to confront the emotional, psychological, and practical challenges of isolation, social distancing, shutdowns, restrictions, and anxieties. For example, a therapist whose practice is focused on working with spouses struggling with the emotional impact of their partner's dementia went through a period when they were unable to take on patients that reminded them of their mother, who was devastated by the evolving dementia of the therapist's father. Supportive supervision helped the therapist separate their own feelings from their patients' and gradually enabled them to reengage effectively.

Less experienced therapists may initially doubt the applicability of their skill to older adult patients, thus reinforcing preexisting uncertainties about the mastery of their craft. The patient's frailty sometimes induces especially strong withdrawal from using psychotherapy because therapists are unwilling to encounter and confront their own hopelessness and fears of abandonment or mortality. Some therapists withdraw because of fear of being engulfed by the patient's demands and dependency. These factors may be compounded by the added therapeutic burdens of working at a slower pace and dealing with sensory deficits and cognitive decline.

Countertransference parallels transference reactions in that therapists may develop strong urges to rescue the patient and do too much, reactions that may derive from unresolved guilt toward their own parents.

Therapists must deal within themselves with the repetitive need for therapy to focus on mourning and acceptance of loss. This is in contrast to the more hopeful perspective of ego development and mastery, which is common in psychotherapy of younger adults and easier for therapists to deal with. Treating older adult patients holds the danger that helpless anxiety or despair may emerge in therapists who have not successfully dealt with their own issues of loss and/or advancing age or frailty.

Therapists of older adult patients are often surprised by the emergence of erotic countertransference during more intensive therapy and may recoil from or deny its presence or feel uncomfortable about aspects of the patient's sexual feelings, thereby depriving the therapy of crucial areas of work.

The sometimes-overwhelming effects of cumulative life trauma are painful for therapists, who may retreat into mechanistic or formulaic approaches, inhibiting their own empathetic relatedness.

A young therapist was surprised when a difficult older adult woman complained angrily about feeling manipulated and exposed when sent home to write her life story, a technique the therapist had read was suitable for the older adult. In a discussion of the case with his supervisor, it became evident that he felt overwhelmed by the patient's painful life and uncertain of what would be helpful. Rather than engaging her to understand her feelings, he prematurely employed a cognitive intervention that made her feel infantilized and rejected.

The process of termination may mobilize the very conflicts and issues that led to therapy, such as the loss of a loved one. Nevertheless, the termination process affords another opportunity for the patient to work through conflicts about the finitude of life and relationships, feelings about abandonment and loneliness, and dependency.

In approaching termination, older adult patients need to be assured of the therapist's continued future availability should the stresses associated with advancing age precipitate regression, conflict, and decompensation. While some older adult patients terminate successfully, for many, therapy is an intermittent process to be resumed as needed. The therapist can ease the patient's return to therapy by anticipating, with the patient, reasons that may inhibit reinstating therapy, such as fear of disappointing the therapist and feeling ashamed because symptoms have recurred. Termination is sometimes imposed by unanticipated events, such as illness. Both patient and therapist must deal with the emotional impact. The focus of the relationship shifts from coping with life to dealing with illness and, in the extreme, death.

Older therapists may find that patients who terminated years before successfully return to therapy as they become older. Age-related emotional disruption may emerge superimposed on deeper psychodynamically and developmentally determined vulnerabilities which had been dealt with in younger years. Examples of age-related themes that have precipitated a return to therapy have included forced retirement and loss of purpose and status, fears of age-related physical illness, loss of a spouse through dementia or death, and anxiety or depression often associated with subtle early cognitive decline. Psychotherapists must be technically agile in adapting interventions to the new aging context and capacities of such returning patients, often layering cognitive, behavioral, problem-solving, and interpersonal foci onto their prior understanding of the patient's lifelong internal emotional conflicts.

Contraindications

There are few contraindications to psychotherapy, although the form of treatment is carefully tailored to the needs of the patient. Psychoanalytic psychotherapy is approached with caution. The therapist first ensures that the patient retains sufficient ego capacity to deal with regression during therapy, work through adaptive mourning, and has the capacity to replace lost sources of self-esteem.

As in younger populations, CBT is indicated for the treatment of acute or chronic depressive disorders, concurrent with antidepressant medications when indicated, and for symptoms associated with a personality disorder. Mild cognitive impairment secondary to dementia is not a contraindication to CBT, although suicidal depression or severe cognitive decline requires alternative therapies.

Goals of Treatment

Psychotherapy is most effective when it is used for specific indications with defined goals. Goals are modified to fit the function, cognition, and psychological capacity of each patient. Generally, the greater the frailty, the more modest the goals are. Goals for patients able to undertake dynamic psychotherapy include psychological mastery over the past and its unresolved grief; resolution of adult and childhood sources of shame and guilt; resolution of current interpersonal conflicts, especially with spouse or adult children; working through conflicts associated with loss of self-esteem and establishing substitute sources of gratification for those that had to be relinquished; coming to terms with failure to achieve ideal aspirations; and, at the most advances level, processing and integrating unconscious conflicts experienced in the therapeutic transference relationship. Pollock has termed the creativity and energy that may be released by the successful resolution of losses *mourning-liberation.*

In crisis therapy, the goal is to reestablish the patient's best premorbid level of function. Examples include acceptance of appropriate levels of dependence or controlling excess, psychologically induced morbidity.

For the more debilitated, frail, older adult patients, psychotherapy goals are often symptom-focused behavioral changes. These include focal interventions, for example, dealing with dementia-related behaviors such as aggression, excess withdrawal, agitation, or delusions. With these patients, therapy goals include strengthening the roles of family CGs and fostering a more appropriate, supportive environment for the patient.

ETHICAL ISSUES

Frail, institutionalized older adult patients have limited capacity to use traditional psychotherapies. They are easily viewed by therapists as beyond help. In allocating scarce resources, therapists struggle with whether to expend effort on this population or reserve it for their more vigorous older adult patients. This is a difficult ethical dilemma since frail older adult patients are the most dependent and helplessly reliant on CGs, who in turn are burdened with the role of deciding who shall or shall not receive certain forms of care.

While most forms of psychotherapy require the active participation of patients and, therefore, their inherent consent, behavioral interventions may be applied to the patient externally. In using behavioral techniques with frail older adult patients, therapists are advised to exercise caution in distinguishing between externally imposed goals of therapy, that is, a focal symptom that is troublesome to staff, and the overall welfare of an often helpless patient whose ability to consent to, and participate in, setting treatment goals is limited.

Clinicians are often confronted with ethical concerns over confidentiality and the patient's decision-making capacity. Both family members and therapists may inadvertently infantilize more frail patients and override their right to participate fully in the therapeutic decision-making process. The psychotherapist must carefully evaluate a patient's ability to make decisions. Similarly, the family's right to know about the patient and participate in therapy is carefully balanced by the preeminent right of the competent older adult patient to confidentiality in the therapeutic relationship. In practice, this is often a difficult issue for therapists, especially when families are intrusive, demanding, or litigious.

FUTURE DIRECTIONS

Conclusions about the efficacy of psychotherapy have remained restricted to older adult patients motivated to participate in research who are typically healthy, community-residing Caucasians in their 60s and 70s. While the evidence for the effectiveness of interventions continues to grow and support psychotherapy in various conditions, it is somewhat narrowly focused on CBT, PST, and reminiscence therapies. Given the current understanding that motivation and compliance with psychotherapy are major factors governing success, it would be helpful to understand older adult persons' resistance to therapy and strategies to prevent premature termination.

Further Readings

Alexopoulos GS, O'Neil R, Banerjee S, et al. "Engage" therapy: prediction of change of late-life major depression. *J Affect Disord*. 2017;221:192–197.

Areán PA, Raue P, Mackin RS, Kanellopoulos D, McCulloch C, Alexopoulos GS. Problem-solving therapy and supportive therapy in older adults with major depression and executive dysfunction. *Am J Psychiatry*. 2010;167(11):1391–1398.

Ayers CR, Sorrell JT, Thorp SR, Wetherell JL. Evidence-based psychological treatments for late-life anxiety. *Psychol Aging*. 2007;22(1):8–17.

Brenes GA, Danhauer SC, Lyles MF, Anderson A, Miller ME. Long-term effects of telephone-delivered psychotherapy for late-life GAD. *Am J Geriatr Psychiatry*. 2017;25:1249–1257.

Choi NG, Marti CN, Conwell Y. Effect of problem-solving therapy on depressed low-income homebound older adults' death/suicidal ideation and hopelessness. *Suicide Life Threat Behav*. 2016;46:323–336.

Choi NG, Marti CN, Bruce ML, Hegel MT, Wilson NL, Kunik ME. Six-month postintervention depression and disability outcomes of in-home telehealth problem-solving therapy for depressed, low-income homebound older adults. *Depress Anxiety*. 2014;31(8):653–661.

Choudhury TK, John KC, Garrett RK, Stagner BH. Considering psychodynamic therapy for older adults. *Psychodyn Psychiatry*. 2020;48(2):152–162.

Fauth EB, Jackson MA, Walberg DK, et al. External validity of the New York University Caregiver Intervention: key caregiver outcomes across multiple demonstration projects. *J Appl Gerontol*. 2019;38:1253–1281.

Francis JL, Kumar A. Psychological treatment of late-life depression. *Psychiatr Clin North Am*. 2013;36(4):561–575.

Gentry MT, Lapid MI, Rummans TA. Geriatric telepsychiatry: systematic review and policy considerations. *Am J Geriatr Psychiatry*. 2019;27:109–127.

Gonçalves DC, Byrne GJ. Interventions for generalized anxiety disorder in older adults: systematic review and meta-analysis. *J Anxiety Disord*. 2012;26:1–11.

Gould RL, Coulson MC, Howard RJ. Cognitive behavioral therapy for depression in older people: a meta-analysis and meta-regression of randomized controlled trials. *J Am Geriatr Soc*. 2012;60(10):1817–1830.

Gustavson KA, Alexopoulos GS, Niu GC, McCulloch C, Meade T, Areán P. Problem-solving therapy reduces suicidal ideation in depressed older adults with executive dysfunction. *Amer J Geriatr Psychiatry*. 2016;24(1):11–17.

Hall J, Kellett S, Berrios R, Bains MK, Scott S. Efficacy of cognitive behavioral therapy for generalized anxiety disorder in older adults: systematic review, meta-analysis, and meta-regression. *Am J Geriatr Psychiatry*. 2016;24:1063–1073.

Hazlett-Stevens H, Singer J, Chong A. Mindfulness-based stress reduction and mindfulness-based cognitive therapy with older adults: a qualitative review of randomized controlled outcome research. *Clin Gerontol*. 2019;42(4):347–358.

Hsieh HF, Wang JJ. Effect of reminiscence therapy on depression in older adults: a systematic review. *Int J Nurs Stud*. 2003;40:335–345.

Jonsson U, Bertilsson G, Allard P, et al. Psychological treatment of depression in people aged 65 years and over: a systematic review of efficacy, safety, and cost-effectiveness. *PLoS ONE*. 2016;11(8):e0160859.

Karlsson B, Johnell K, Sigstrom R, Sjoberg L, Fratiglioni L. Depression and depression treatment in a population-based study of individuals over 60 years old without dementia. *Am J Geriatr Psychiatry*. 2016;24(8):615–623.

Kiosses DN, Alexopoulos GS. Problem-solving therapy in the elderly. *Curr Treat Options Psychiatry*. 2014;1(1):15–26.

Knavelsrud C, Bottche M, Pietrzak RH, Freyberger HJ, Renneberg B, Kuwert P. Integrative testimonial therapy: an Internet-based, therapist-assisted therapy for German elderly survivors of the World War II with posttraumatic stress symptoms. *J Nerv Men Dis*. 2014;202(9):651–658.

Laidlaw K. A deficit in psychotherapeutic care for older people with depression and anxiety. *Gerontology*. 2013;59:549–556.

Li SYH, Bressington D. The effects of mindfulness-based stress reduction on depression, anxiety, and stress in older adults: a systematic review and meta-analysis. *Int J Ment Health Nurs*. 2019;28(3):635–656.

Linde K, Sigterman K, Kriston L, et al. Effectiveness of psychological treatments for depressive disorders in primary care: systematic review and meta-analysis. *Ann Fam Med*. 2015;13:56–68.

Newton NA, Jacobowitz J. Transferential and countertransferential processes in therapy with older adults. In: Duffy M, ed. *Handbook of Counseling and Psychotherapy with Older Adults*. John Wiley and Sons; 1999.

Pinquart M, Forstmeier S. Effects of reminiscence interventions on psychosocial outcomes: a meta-analysis. *Aging Ment Health*. 2011;16:541–558.

Plotkin DA. Older adults and psychoanalytic treatment: it’s about time. *Psychodyn Psychiatry*. 2014;42(1):23–50.

Raue PJ, McGovern AR, Kiosses DN, Sirey JA. Advances in psychotherapy for depressed older adults. *Curr Psychiatry Rep*. 2017;19:57.

Sabir M, Henderson CR, Kang S, Pillemer K. Attachment-focused integrative reminiscence with older African Americans: a randomized controlled intervention study. *Aging Ment Health*. 2016;2(5):517–528.

Sadavoy J. Personality disorders in late life. In: Laidlaw K, Pachana N, eds. *Oxford Handbook of Clinical Geropsychology. Great Clarendon*: Oxford University Press; 2014.

Sadavoy J, Sajedinejad S, Chiu M. A quasi-experimental study of the effectiveness of the Reitman Centre CARERS group intervention on family caregivers of persons with dementia. *Int J Geriatr Psychiatry*. 2121;36(6):811–821.

Sadavoy J, Sajedinejad S, Chiu M. Evaluation of the Reitman Centre CARERS Program for Supporting Dementia Family Caregivers: a pre-post Intervention study. *Int Psychogeriatr*. 2022;34(9):827–838.

Schachter J, Kächele H, Schachter JS. Psychotherapeutic/psychoanalytic treatment of the elderly. *Psychodyn Psychiatry*. 2014;42(1):51–63.

Shear MK, Wang Y, Skritskaya N, Duan N, Mauro C, Ghesquiere A. Treatment of complicated grief in elderly persons: a randomized clinical trial. *JAMA Psychiatry*. 2014;71(11):1287–1295.

Westerhof GJ, Bohlmeijer ET. Celebrating fifty years of research and applications in reminiscence and life review: state of the art and new directions. *J Aging Stud*. 2014;29:107–114.

Yen HY, Lin LJ. A systematic review of reminiscence therapy for older adults in Taiwan. *J Nurs Res*. 2018;26(2):138–150.

55.4i Cognitive-Behavioral Therapy

JOHN R. MCQUAID, PH.D., ERIC GRANHOLM, PH.D., AND PATRICIA A. AREÁN, PH.D.

DEFINITION

Considerable progress has been made in developing cognitive-behavioral therapy (CBT) interventions for use in older people with anxiety, depression, and psychotic disorders.

CBT is a form of psychotherapy that attempts to change the interpretations of experiences that determine feelings and behaviors with the goal of reducing symptoms and improving functioning. The premise of CBT is that cognition (the process of acquiring knowledge and forming beliefs) can influence feelings and behaviors. CBT interventions help people to identify and modify thoughts and misinterpretations of experiences that are at the root of problematic behavior. Behavioral interventions (increasing rewarding activities, practicing new skills) both improve mood and provide opportunities to test the accuracy and benefit of beliefs that may contribute to psychological symptoms and impaired functioning.

ADAPTING CBT FOR OLDER CLIENTS

Cognitive therapy is often described as a manualized and structured form of psychotherapy. It is true that there is a strong emphasis on using standardized and validated interventions along with an ongoing assessment of change in targeted symptoms and behaviors. However, when provided optimally, CBT is individualized for the specific client, regardless of age or aging-related issues. Problems and goals are identified in a collaborative manner, activities and skills are designed for the specific needs of the client, and the therapist and client continually review progress to determine whether treatment is providing the desired benefit. Clinicians should keep in mind that treatment modifications should not be made solely on the basis of a client’s age but should be made when issues such as cognitive impairment, disability, and illness are present that can influence the effectiveness of treatment. These issues can occur at any age but are particularly prevalent in older adults and especially common among those who are 75 years old or older. Therefore, before making any major accommodations, a good clinician assesses the need to make them first rather than solely making assumptions about the client’s age.

Ageist Beliefs and Age-Related Challenges

One of the age-specific modifications clinicians may need to consider to do with cohort beliefs about seeking help and ageism. Older adults are less likely to seek help for mental health problems than younger adults and have been noted to hold erroneous beliefs about aging. Cohort-related beliefs can devalue therapy (e.g., “Don’t air your dirty laundry”) and ageist beliefs held by both therapists and clients (e.g., “I’m too old to learn or change”) can interfere with participation in therapy. Just as they can be applied to beliefs that increase psychiatric symptoms, thought-challenging interventions can also be used to modify ageist beliefs that interfere with treatment participation. Similarly, using CBT techniques to address interpersonal and behavioral challenges related to aging is also necessary, such as role-playing age-relevant interpersonal situations (e.g., talking to a doctor about health concerns or eyeglasses) and practicing solving age-specific concerns (e.g., finding transportation, coping with hearing or vision problems).

Transportation Barriers

Transportation and ambulation problems, which are more common among older clients, can further exacerbate problems with isolation and withdrawal from society. It may be necessary to provide transportation to therapy sessions or conduct therapy at convenient locations in the community. Providing treatment in the community can also reduce the stigma associated with going to a mental health clinic, which can especially deter older clients.

Cognitive Impairment

Addressing neurocognitive impairment is another important factor to consider when designing CBT interventions for older adults. Due to normal aging and lifetime comorbid factors that can impact cognition in older clients, especially those with severe mental illnesses (e.g., poor education, lack of work, physical illness, poor nutrition, poverty), cognitive functioning may be impaired to a greater extent in older clients. Simplifying and slowing down discussions, repeated practice of skills, and neurocognitive compensatory aids (e.g., reminder notes, treatment workbooks, writing down information in session, and acronyms to help with skill recall) can be used to help alleviate cognitive impairments. Given fine motor and vision difficulties that are also common in older clients, cognitive compensatory aids and educational materials should use large fonts and writing spaces. Clinicians who have worked with younger populations may need to modify their expectations about the speed at which change may take place, without assuming that older clients cannot make change.

Medical Problems

Many medical impairments increase in frequency as people age. With longer life spans, there is increased exposure to a variety of medical ailments, including cancer, heart disease, musculoskeletal disorders, and other chronic health problems. These changes in health have a number of direct and indirect effects on psychiatric symptoms and treatment. Physical suffering can directly influence mood, anxiety, and psychotic symptoms. In addition, reductions in physical health can interfere with engaging in rewarding activities, demands of health care can reduce financial resources, and sensory impairments secondary to health problems can interfere with both daily activities and interpersonal interactions. All of these problems can also interfere with participation in CBT, which emphasizes the active practice of new skills. Therefore, behavioral interventions need to focus on increasing client activities within the constraints of their medical condition. In addition, cognitive interventions can be beneficial if they help clients to shift perspective from the negative aspects of illness and to focus on thinking that can improve their experience in the moment, even in the context of medical problems.

Role Changes

Aging is associated with transitions through different roles across stages of life. Some of these changes can be disruptive for people, particularly if they are unexpected or involve loss of important sources of reinforcement. For example, retirement may be a positive experience for many, but for some, it can be very disconcerting, particularly individuals who defined their worth primarily through their productivity. Other role transitions involve profound loss, such as changing from a spouse to a widow(er) or divorcee, or from a homeowner to a member of an assisted-living community. In addition, for some individuals role changes include unexpected demands of caregiving at an older age, either of a spouse, child, or grandchild, or alternatively losing independence and needing the aid of a caregiver. These role shifts can involve both changes in self-image as well as physical and emotional demands. These changes can contribute to depression by either reducing access to rewarding activities (e.g., work) or by activating beliefs the client has that can be depressogenic (e.g., “I’m unproductive and useless”). They can exacerbate anxiety symptoms as individuals face uncertainty and increased perceived risk (e.g., “my wife handled the finances, how can I manage without her”?).

When clients are facing such role changes, CBT can be particularly beneficial. On the pragmatic side, behavioral exercises such as scheduling, information gathering, and contacting support resources can help patients more effectively navigate the role change. Cognitive interventions can address depressogenic thoughts, challenge beliefs that are inconsistent with the new role, and generate thoughts that are likely to improve functioning. These efforts can work in service of modifying schema that are less beneficial in the new role (e.g., seeing one’s value as connected to productivity; feeling dependent on others) to a schema that is more functional given the realities of the situation (e.g., seeing one’s self as innately valuable; remembering that new skills can be learned).

Interpersonal Events

Older adults are vulnerable to changes in their relationships that can have important implications for their mood and functioning. Deaths of spouses, other family members, friends, and acquaintances can both provoke grief and represent loss of important social supports. Interpersonal relationships can end through other means as well, including adult children moving away from home or the patient leaving the work environment. In addition, changes in relationships associated with aging, such as having a child become a caregiver of the older person or needing to become a caregiver of an older person, can strain those relationships.

Issues of loss, isolation, and improving social support, leisure activities, and interpersonal communication skills, therefore, are even more important for older clients. Again, the combination of behavioral activities, such as communicating with supports or developing new activities that can increase positive contacts, as well as thought challenging to reduce cognitions that exacerbate the pain associated with interpersonal events can be important themes in CBT for older adults.

Culture and Age Cohorts

Broadly defined, culture provides the context that guides the development of beliefs and expectations. That context changes over time. Among geriatric populations, different age groups have different expectations. The oldest members of society, having lived through the Great Depression and World War II, grew up with different cultural expectations than their children, who were born during and after World War II and came of age during the 1950s and 1960s. It is important for therapists to assess each person’s developmental history with a focus on understanding the beliefs and values that arose from that history. This assessment can then help in developing interventions that address unhelpful thoughts and behaviors while supporting the core values of the client. In addition, the assessment is critical to prevent clinicians from relying on general assumptions about a particular age cohort, but rather to have an understanding of how the experiences shaped the person with whom they are working.

In reviewing the domains above, two major themes arise: *Loss* (of health, role, loved ones, finances) and *danger* of further losses in these domains. Loss tends to be associated with risk for depression and perceived danger with risk for anxiety. Treatment for older individuals in particular needs to address coping with such losses in a constructive manner, helping clients differentiate between likely and unlikely threats, and identifying coping strategies for upcoming risks.

Traditional CBT emphasizes change orientation, attempting to change how one acts and thinks to improve mood. However, emphasizing change alone can be limited in effectiveness and potentially invalidating the client, if the therapist does not also acknowledge the unchangeable nature of some challenges in the older person's life. Some CBT experts refer to this as "the healthy management of reality." From this perspective, the role of CBT is to help clients accurately understand the challenges they face, and then when appropriate modify their actions and thoughts so they can live a meaningful and rewarding life within the context of those challenges. Recent models that have derived from CBT and incorporate acceptance and mindfulness approaches (including dialectical behavior therapy, acceptance and commitment therapy, and mindfulness-based cognitive therapy) all emphasize both skills in change as well as acceptance of unchangeable realities. These treatments tend to incorporate skills for distress tolerance and mindful observation to aid in managing unmodifiable aspects of life experience.

Implications of the COVID-19 Pandemic

Most of these factors have been exacerbated by the COVID-19 pandemic. Older adults have been most vulnerable to this disease, suffering the highest mortality rates. They have experienced disproportionate levels of isolation and loss and increasing psychiatric symptoms. However, the pandemic has also created pressure for adoption of novel approaches to communication and care delivery, particularly the use of telemental health/video conferencing. While this has been broadly adopted and there is research demonstrating CBT is effective via telemental health, some have raised concern over whether older clients can and will use technology such as video conferencing to conduct psychotherapy. While there is limited research in this domain, there is growing evidence that older clients are both willing and able to adopt this technology an dissatisfied with geropsychiatry services in general delivered by telemental health. The authors have had experience of successfully implementing CBT via video in the context of the COVID-19 pandemic with older clients and generally found broad acceptance and appreciation of the benefits of reduced risk and increased access to care. This highlights the need to avoid allowing biases regarding older adults to guide decision making and to engage in collaborative coordination with clients.

TECHNIQUES

Stages of Cognitive-Behavioral Therapy

In general, CBT can be framed in three stages. The first stage is assessment and initial treatment and includes definition of clients' current problems and goals, education about the model of treatment, establishment of a collaborative relationship, and agreement on a treatment plan. In addition, the therapist begins to develop a case formulation. The formulation is a model that attempts to explain the current difficulties in terms of a schema (an organizing set of deeply held beliefs that define the client's perception of self and determine the information that the client attends to, remembers, and recalls, and how that information is interpreted). One goal of this phase is to instill hope that improvement can occur, if possible through the client experiencing some symptom relief secondary to an initial exercise or homework assignment. The middle stage of therapy focuses on skills acquisition, with the primary goal of the client learning to implement CBT interventions independently. The final stage, relapse prevention and termination, emphasizes consolidation of skills, problem-solving obstacles to skill use, identifying potential risks for symptom exacerbation, and developing a plan to apply learned skills in the case of symptom recurrence.

Assessment and Initial Treatment

CBT emphasizes objective measurement of constructs whenever possible, and this begins in the assessment phase. When working with older clients, this will include assessing not only symptom severity, but also the resources available and the challenges faced, and then trying to define all these factors in objective terms. Number of supports, income relative to expenses, health limitations, and caregiver responsibilities are all-important aspects to assess. In addition, standardized measurement of symptoms provides a baseline to assess progress in treatment. In particular, the use of measures with normative data for older populations (such as the Geriatric Depression Scale) is important because several symptoms of depression and anxiety (e.g., sleep changes, reduced energy, memory, and concentration difficulties) can occur in older populations independent of depressive symptoms, and therefore measures need to account for these potential confounds.

As noted above, a key goal of the initial assessment is the development of a case formulation. A case formulation provides a framework for choosing interventions that match the specific needs of the client. Case formulation is particularly important when dealing with special populations such as older adults because it guides both the assessment and the treatment. A well-defined case formulation increases the likelihood that areas such as those reviewed above are addressed and incorporated into treatment.

Juanita was a 69-year-old divorced, unemployed, heterosexual, cisgender Mexican American woman who supported herself on her Social Security benefits. She lived in a one-bedroom apartment. she was serving as guardian of her 5-year-old grandson following his mother’s incarceration. She presented with the following concerns:

1. Depressed mood: Juanita reported daily crying, anhedonia, increased appetite and weight gain, feeling tired “all the time,” feelings of guilt and shame, and thoughts of death.
2. Anxiety symptoms: Juanita met criteria for generalized anxiety disorder. She reported constant excessive daily worry about both major stressors (her daughter’s legal troubles; financial stress) as well as other factors (her children, her friends, safety).
3. Daughter’s arrest and custody of grandson: Her daughter had been arrested approximately 1 year prior to treatment, leading to Juanita receiving custody of her grandson. Her grandson frequently cried at night for his mother, and Juanita felt overwhelmed by trying to provide for his care. Her daughter was incarcerated several hours from where Juanita lived, limiting visitation.
4. Health problems: Juanita had type II diabetes. She had difficulty managing her symptoms in part due to anxiety about checking her glucose levels. Juanita met criteria for obesity and worried that her diabetes would progress and lead to amputation of her foot.
5. Social isolation: Although Juanita had family and friends in the area, she rarely contacted them or answered her phone.

In the intake interview, Juanita stated, “I tried to be a good mother, but I failed. And now I’m going to fail my grandson.” When discussing goals, Juanita stated that her greatest concern was her daughter’s and grandson’s futures. In discussing this goal in terms of what Juanita could influence, she agreed that she would like to provide a safe and loving environment for her grandson and to be a support for her daughter. These goals required her to take several actions around caring for herself, including:

1. Reconnect with family and friends to increase pleasant activities and get help with caring for her grandson
2. Improve her diabetes care and reduce weight so that she could be physically better able to care for her grandson
3. Generate additional income by re-establishing her home-based business sewing traditional dance dresses for a local dance studio
4. Make arrangements to visit her daughter in prison with her grandson on a regular schedule

In this case, several of the described factors that are associated with depression and anxiety in older people were apparent. Juanita had clear medical problems that limited her activities, she had limited resources, and she was isolated from support (albeit by her own actions). She also experienced a clear loss (her daughter being incarcerated) and a demanding role change (becoming caregiver to her grandson). In the context of the meaning she ascribed to these events ("I failed"), it is not surprising for her to be experiencing depressive symptoms. Of particular note, her anxiety and depressive symptoms were exacerbating both her health problems (interfering with diabetes management) and interpersonal functioning (avoiding supports).

In reviewing the goals, the initial theme expressed by Juanita was a desire to care for others. However, the therapist was concerned that without taking care of Juanita's own issues, she would be unable to achieve her goals. Therefore, the therapist helped focus on self-care in terms of necessary preparation for her goal of caring for others. In this case, the proposed formulation identified a potential core belief held by Juanita of "I am undeserving; I need to care for others to get love." Rather than directly challenging this initially, the therapist first worked with strategies to achieve the goals Juanita had identified. Over the course of therapy, an additional goal of schema change (moving from "I'm undeserving and need to care for others to get love" to "I am a good and deserving person who loves to care for others") was instituted, but initially therapy focused on engaging Juanita in treatment so that she could gain some symptom improvement before taking on the more difficult tasks of acceptance and schema change.

In working with the therapist, these general goals of providing for her grandson were translated into more specific actions that could meet her goals and be beneficial to her in terms of improving her mood and functioning.

Middle Stage Interventions

Although the interventions in this stage of therapy are generally divided into behavioral and cognitive interventions, such interventions frequently have both cognitive and behavioral aspects. Behavioral interventions can lead to changes in cognition (e.g., doing an activity and finding out it is more enjoyable than predicted), while cognitive interventions can be useful in helping patients to do behavioral interventions (e.g., challenging the belief that one has to "feel" like doing something before doing it). Research has shown that conducting either cognitive or behavioral exercises leads to change in both cognition and behavior. In general, interventions focus initially on assessing baseline thoughts and activities and their relationship to mood and other outcomes. Clients then learn change strategies and coping skills in this stage of therapy and assess whether these skills are effective in improving outcomes.

Psychoeducation. This initial assessment of thoughts and activities and their relationship to mood and other outcomes provides an opportunity to introduce the cognitive model and the idea that thoughts influence feelings and behaviors. The groundwork for future cognitive work is laid by using the client's own experience to illustrate how thoughts relate to moods and behaviors. With all clients, but particularly with older clients with cognitive impairments, it is helpful to draw on a pad or marker board or use a handout to illustrate associations between thoughts, feelings, and behaviors, as shown in Figure 55.4i–1.

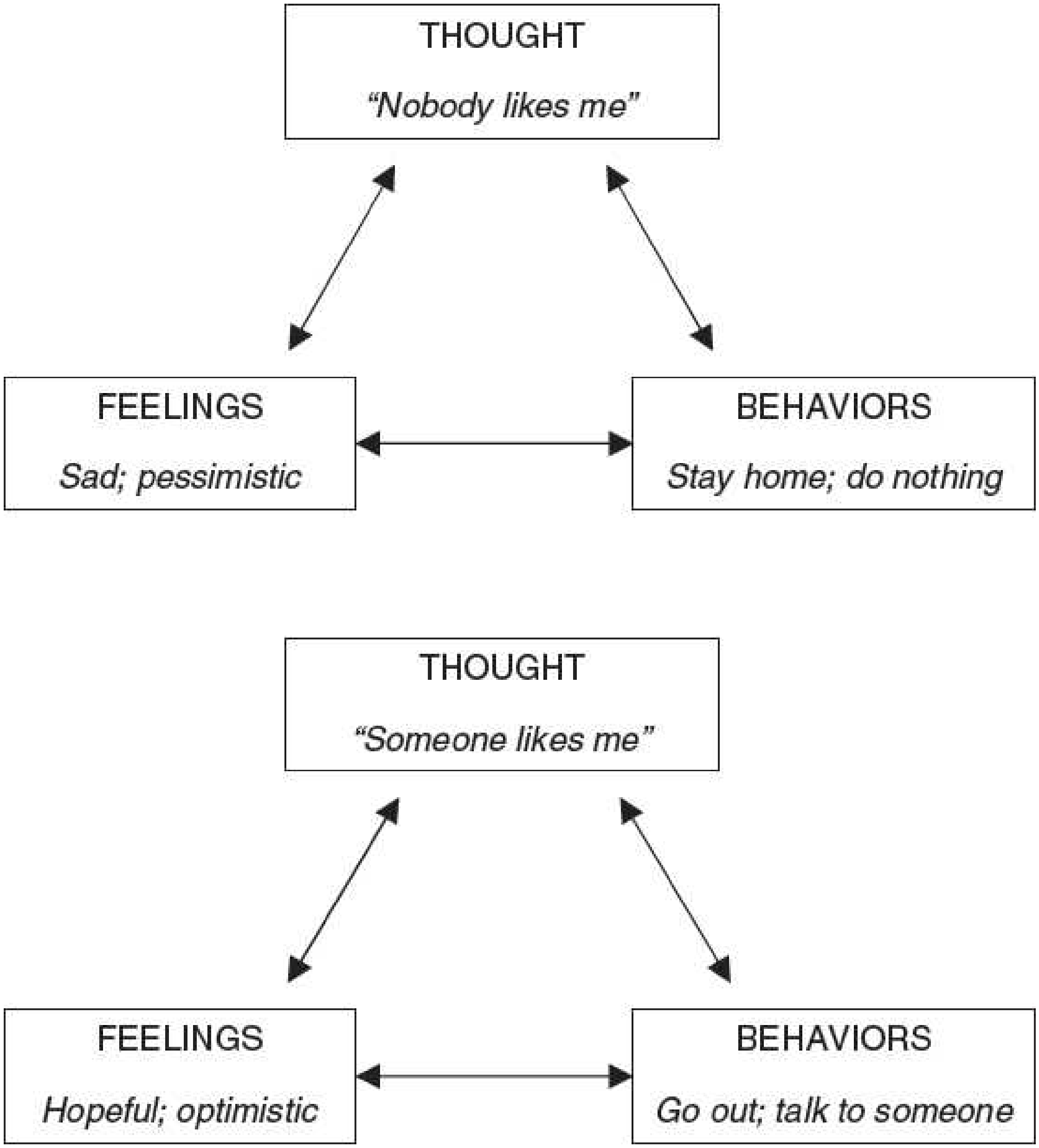
THOUGHT
"Nobody likes me"
FEELINGS
Sad; pessimistic
BEHAVIORS
Stay home; do nothing
THOUGHT
"Someone likes me"
FEELINGS
Hopeful; optimistic
BEHAVIORS
Go out; talk to someone

FIGURE 55.4i–1. Generic cognitive model illustrating relationships between unhelpful (**top**) and helpful (**bottom**) thoughts, feelings, and behaviors.

Depending on individual client needs, it may also be important to provide education about symptoms associated with specific diagnoses, the action of medications, medication adherence strategies (especially in the context of memory impairments), and complementary interactions between medication and psychotherapy treatments. It is often useful in this context to provide education about the vulnerability-stress-coping model of mental illness, focusing on the interaction of biologic vulnerabilities, stress triggers, and coping strengths in symptom exacerbation or relapse. Finally, it is generally best practice in all mental health interventions to provide education about triggers and signs of crisis, assess risk for suicide, and develop an individualized safety plan for prevention of suicide or relapse/hospitalization as indicated.

Behavioral Interventions. Behavioral interventions in CBT for mood disorders tend to emphasize activities that increase positive affect, gather evidence to challenge unhelpful beliefs or both. Initial interventions can solely consist of *activity tracking* (e.g., recording a baseline of all activities that a person is doing throughout the day). Activity tracking helps clients recognize their current activity level, potentially challenge unhelpful beliefs like "I never do anything" and identify opportunities for further intervention. *Activity scheduling* is simply identifying one or more activities that clients may find rewarding and planning them into their day. Keys to successful activity scheduling are generating multiple options, working with the clients to identify choices they believe they can complete, and problem-solving potential obstacles to the activity. In addition, tracking relevant emotions (e.g., depression, anxiety, happiness) before and after the activity can help clients notice whether there is any benefit. In older adults, medical problems can significantly limit mobility and activity, and some beliefs (e.g., "But, I've been doing it this way for years") can make changing routines particularly challenging. However, activities do not have to involve strenuous physical activity to enhance mood or alter cognitions, and most clients can manage some change in routine if care is taken to address interfering cognitions and problem-solve around limitations.

A variant of activity scheduling is *pleasure predicting*. This is a useful technique when clients doubt that they will enjoy an activity. In addition to rating mood before and after the exercise, clients are asked before they do the activity to predict how much they will enjoy it and then rate the actual enjoyment once completed. Comparing the predicted outcome with the actual outcome helps challenge the belief “I won’t enjoy it” and helps to highlight the tendency to underestimate the benefit of pleasant events.

For anxiety disorders, behavioral interventions include graduated exposure, which involves engaging in feared activities in an organized fashion. Using principles of systematic desensitization, clients break down a feared activity (e.g., shopping in a store) into specific steps that are arranged from least anxiety-provoking to most (e.g., shopping in a large grocery store early when no one else is there versus shopping in a small crowded shop at the busiest time of the day). Clients practice doing less anxiety-provoking tasks first and then move on to the next task. Clients may become less physiologically reactive to the feared situation, as well as develop more helpful cognitions (e.g., “I can handle this.”) although research indicates that the violation of the expectation of a negative outcome is key to overall improvement, even if anxiety in the specific exposure does not decrease.

Clients in anxiety treatment can also be taught a variety of relaxation techniques, such as progressive muscle relaxation or breathing retraining, to facilitate management of their anxiety symptoms. The limited research available suggests that relaxation alone can be as effective as CBT for reducing anxiety symptoms in older clients. Recommendations specific to late-life anxiety include (1) ensuring that sufficient time is devoted to learning relaxation training skills, (2) emphasizing in vivo rather than imaginal exposure exercises, and (3) de-emphasizing more abstract cognitive restructuring techniques (e.g., directly disputing dysfunctional thoughts).

A related exercise that has been incorporated into both anxiety and depression treatment, and to some extent acceptance interventions for psychosis, is mindfulness meditation. Mindfulness is defined as paying attention in a purposeful way, in the moment, nonjudgmentally. Clients are taught to allocate their attention to a particular stimulus (e.g., their breath), and to just notice this. As distractions arise (e.g., internal thoughts, external sounds), clients learn to just notice these, as well as their own reaction to them. They then return attention to the target stimulus. Mindfulness is a particularly useful skill in depression and anxiety for helping clients learn to observe their thoughts. When people notice their thoughts, they can then question the accuracy of them and do not need to act on those thoughts.

Juanita had stopped communicating with other family members, even though this had been a rewarding activity in the past. Therefore, therapy focused on this as an opportunity for behavioral activation.

Therapist: One thing you told me is that you used to enjoy time with your son and his family. Why don't we work on arranging some time with them?
J: Okay.
T: Okay, so what day and time do you think makes sense?
J: Well, my son works, so it would be best on the weekend, maybe in the afternoon.
T: How do you usually arrange to get together?
J: I just call him up and tell him I'm coming over.
T: Is there anything stopping you now?
J: I just don't want to be around people. I just feel like I'll bring everyone down. That would just make things worse.
T: Remember, part of the goal is to test beliefs like that. Would you be willing to go visit and then write down your evidence as to whether you brought everyone down and things got worse?
J: Okay.

This example highlights several important skills: Activity planning (spending time with the family), identifying obstacles (thoughts about her response), and challenging thoughts (gathering evidence from the experience).

Cognitive Interventions. One of the key skills that clients learn is thought challenging. The key steps to thought challenging are:

1. Noticing clients’ thoughts and identifying their effect on mood and behavior
2. Learning to test the helpfulness and/or accuracy of those thoughts
3. Generating a new, more accurate, or useful thought
4. Examining the effect of the new thought on mood and behavior to determine if it is an improvement

For some clients, the initial task of noticing thoughts is challenging. For clients who have difficulty even noticing their thoughts, a form such as the automatic thoughts checklist can be helpful. This is a list of thoughts associated with depression. Clients read through the list and check those that they experience. Once clients can recognize their thoughts, there are a variety of forms for teaching clients to examine and change them. With older clients, a straightforward model, such as the “3C’s” form, can be useful because it is a simple model (Table 55.4i–1). Clients are initially taught to write down thoughts in the “catch it” column and the relationship to their mood in the “check it” column, as well as any errors or problems with the thought. Once clients can recognize their cognitions, they learn strategies to question and challenge thoughts that are unhelpful and write down more helpful thoughts in the “change it” column. Strategies include identifying common unhelpful thoughts (e.g., all-or-nothing thinking, predicting the future) and examining evidence both in support of the unhelpful thought and in contradiction to it. Throughout these exercises, clients are taught to determine the effectiveness of these efforts by rating changes in their mood in response to the new thought (Table 55.4i–2).

Table 55.4i–1.
3C's Homework Form

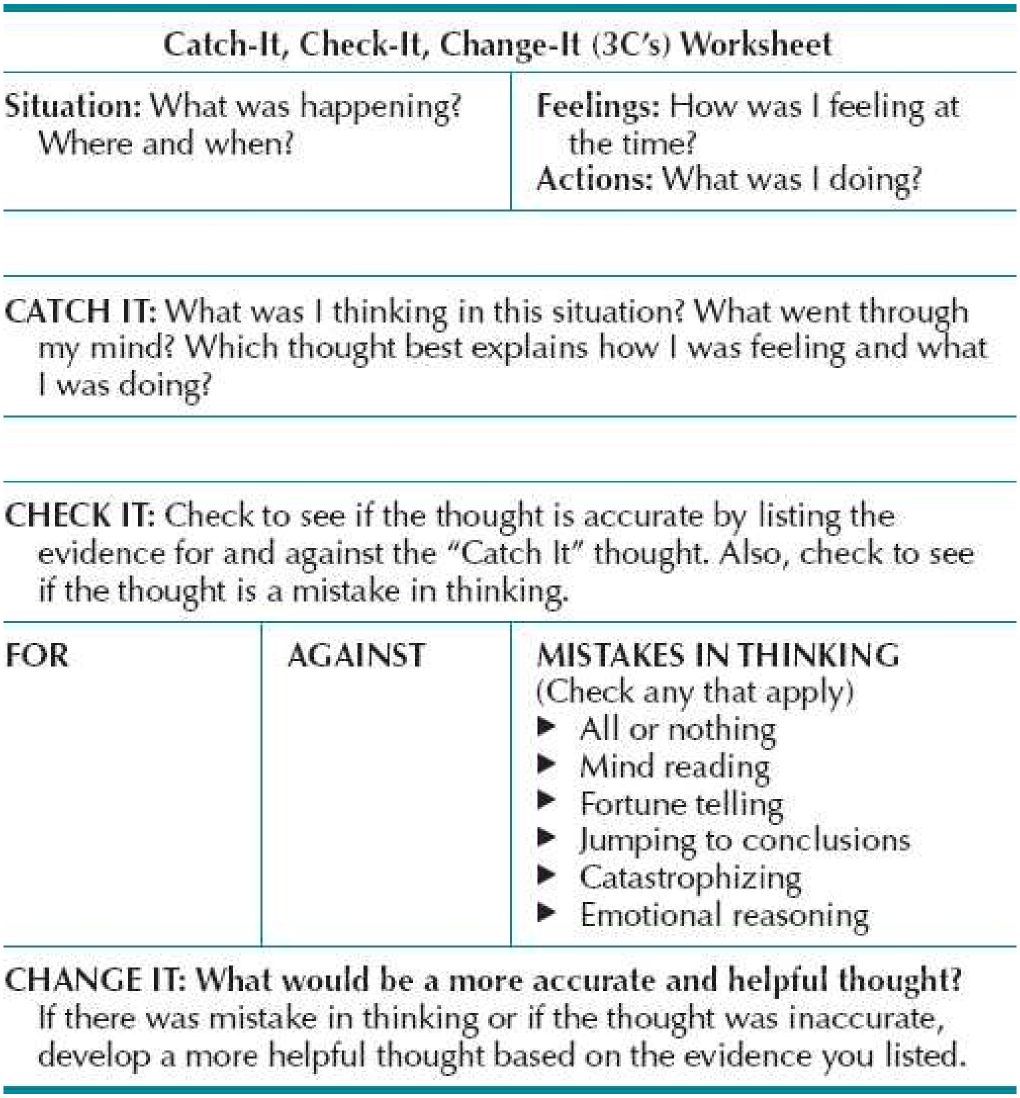

<table>
<tr><th colspan="3">Catch-It, Check-It, Change-It (3C's) Worksheet</th></tr>
<tr><td colspan="2">Situation: What was happening? Where and when?</td><td>Feelings: How was I feeling at the time?
Actions: What was I doing?</td></tr>
<tr><td colspan="2"></td><td></td></tr>
<tr><td colspan="3">CATCH IT: What was I thinking in this situation? What went through my mind? Which thought best explains how I was feeling and what I was doing?</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">CHECK IT: Check to see if the thought is accurate by listing the evidence for and against the "Catch It" thought. Also, check to see if the thought is a mistake in thinking.</td></tr>
<tr><td>FOR</td><td>AGAINST</td><td>MISTAKES IN THINKING
(Check any that apply)
▶ All or nothing
▶ Mind reading
▶ Fortune telling
▶ Jumping to conclusions
▶ Catastrophizing
▶ Emotional reasoning</td></tr>
<tr><td colspan="3">CHANGE IT: What would be a more accurate and helpful thought? If there was mistake in thinking or if the thought was inaccurate, develop a more helpful thought based on the evidence you listed.</td></tr>
</table>

Table 55.4i–2.
Common Types of Unhelpful Thinking

All or nothing	Viewing situations in terms of absolutes: “I am a complete failure”
Selective attention	Attending to negative information and/or ignoring positive information (e.g., noticing health problems and ignoring supportive family members)
Shoulds	Holding rigid rules and expectations; refusing to accept realities outside one’s control: “I should be able to do all the things I used to do.”
Emotional reasoning	Judging a situation based on one’s emotional response, rather than objective evidence: “I was feeling anxious; therefore, I must have looked foolish.”
Mind reading	Making assumptions about others’ perceptions: “They think I’m too old.”
Predicting the future/probability estimation errors	Overestimating the likelihood of a negative event: “I know that I will fall and break my hip if I go to the store.”
Catastrophizing	Underestimating one’s ability to cope with a negative event: “I can’t stand it if I get sick.”
Over-responsibility	Feeling guilt for situations outside of one’s control: “It is my fault my grandchildren are unhappy.”

Many people, including therapists and older people themselves, hold unhelpful beliefs about older adults. These can include thoughts such as, “Older people can’t learn/change,” “Others do not want to be around older people,” “Older people are a burden,” or “Older people don’t have much to give.” Beliefs about managing suffering (“Don’t air your dirty laundry”) or about therapy (“Therapy is for crazy people”) can also interfere with the acceptability of treatment. Therapists need to attend to both of these sets of beliefs, both in how therapy is initially presented and in the course of the actual treatment.

Mike was a 75-year-old retired, gay-identified cisgender White male whose long-time partner and husband had died approximately 2 years before initiating treatment. He was referred for care by a close friend who was concerned that he was still grieving the death of his husband and had become isolated.

Mike had grown up in a traditional Christian family on a farm in a small Midwest town, the middle of three siblings (an older brother and a younger sister). He reports having a loving mother and distant father, but that he disconnected from his parents and siblings when he started to identify as gay because he anticipated they would reject him based on their religious beliefs. He joined the Navy in part to "escape" small town life. Mike had served as a Navy corpsman including deployment on an aircraft carrier during the Vietnam War. He had no direct combat exposure but was involved in caring for wounded evacuated to the ship.

Mike found his role rewarding and served in the military for approximately 15 years. However, he ended up leaving the military prior to retirement, for which he would have been eligible after 20 years, because rumors of his sexual orientation were beginning to spread among his unit. Upon leaving the military he attended a state university in a large city and obtained his bachelor's degree and teaching credential. He then pursued a successful career as a middle school teacher.

Mike had very limited dating or sexual experience prior to or during his military service, fearing consequences for his career. However, while in school he was active in the gay community of his city, dating frequently. He also reports drinking heavily at this time. He met his partner, Bill, and reports that within a year they had developed a committed relationship. Bill was approximately 10 years older than Mike, and Mike describes him as a partner as well as a mentor helping him navigate the "gay scene." Over the next 10 years with Bill's support, he was able to disclose his sexual orientation to his parents and siblings, and his brother and sister in particular were supportive and developed a close bond.

His parents were initially “disappointed” but eventually “came around.”

Mike reported that his relationship with Bill was marred by several periods of heavy drinking during which Mike had anonymous sexual encounters with other men. When Bill found out about these incidents, they had a series of arguments and confrontations that led to Mike attending Alcoholics Anonymous (AA). They maintained their relationship and married when gay marriage was legalized, and Mike remained sober for the following 30 years. However, Bill would frequently remind Mike of his infidelities and criticize him during times of conflict.

Approximately 10 years prior to Mike initiating CBT, Bill had developed a neurodegenerative disease that quickly limited his physical functioning, and left him bedridden for the last 4 years of his life. Mike served as his primary caretaker throughout his decline, retiring from his job as a teacher to take care of Bill. During this time, as Bill developed symptoms of dementia, he would continue to recall Mike’s earlier infidelities and remind him of how he had hurt Bill in the past.

When initially presenting for treatment, Mike shared that since Bill had died he continued to feel both extreme sadness at the loss and guilt about his previous behavior. However, he also reported anger at Bill’s inability to forgive him, which in turn led to additional guilt. He also reported having “no purpose” now that Bill had passed. Although he and Bill had a large network of friends, he was not interacting with them. While he was still attending AA, he no longer took leadership roles in the organization, sponsored other members, or participated in social events. He only spoke to his brother and sister when they called and had started screening their calls to avoid them.

At the beginning of therapy, Mike was meeting criteria for major depression, including daily sad mood, anhedonia, decreased energy, pervasive guilt, poor appetite, insomnia, and thoughts of death (he denied suicidal ideation, plan, or intent, but thought that the world would be better off without him). He stated he was irritable at AA meetings and took benign comments as criticisms directed at him. He also reported he had not cleared out any of Bill's belongings, and that he criticized himself for this as well.

Therapy focused on three areas: Acknowledging the loss of Bill, developing goals for Mike that were consistent with his values, and increasing his engagement with people who cared about him. Based on his history, the therapist developed a case formulation that focused on the core belief "I'm defective." This manifest in secondary assumptions of "I deserve to be punished" and "I can never be forgiven for the harm I've done." This in turn led him both to be very self-critical and to interpret those around him as being critical as well.

In the course of therapy, Mike learned to examine his thoughts and the relationship between thoughts, actions, and feelings. As an example, he stated that one of his goals was to start cleaning up his home, including donating most of Bill's clothes to charity. However, in the session, Mike reported that he had been unable to donate the clothes and got upset when he thought about it. The therapist helped Mike to examine his thoughts using the 3C's form: In reviewing the thoughts, Mike was guided in identifying thinking errors, in this case, "should" thoughts. He was then guided through questioning the assumptions underlying the thoughts to generate alternative, more helpful thoughts:

Therapist: What do you tell yourself when you looked at Bill's clothes?

Mike: If I get rid of them, I'm being unfaithful. I'm betraying him.

T: When you say that, what emotions come up?

M: Shame, guilt . . . and some anger

T: I can see where the shame and guilt come from; how does that thought lead to anger?

M: I get angry at myself for even wanting to give away his clothes. I shouldn't want to.

T: I see. You're telling yourself there's something wrong with you if you want to give his clothes away; "I shouldn't want to do this." So what would be a more helpful thing to say to yourself?

M: I don't know.

T: Well, what is the reason you want to give away the clothes? Is it because you didn’t love Bill?

M: No! It’s just that it is so hard to see those reminders. I think it would help me move forward.

T: Do you think that Bill would not want you to move on?

M: No, he told me that he wanted me to be happy after he was gone, or at least he did when he wasn’t upset with me.

T: So it sounds like the “should” thought is not true; there are good reasons for wanting to give the clothes away, and it makes sense to want to do so. What would be a more helpful thing to say?

M: Maybe that it’s ok to give the clothes away; it is doing something for others, and Bill would like that.

T: And what’s your evidence that this is true?

M: Well, he told me to move on when he was gone, and he did like donating stuff to charity.

T: Anything else? If the roles were reversed, what would you want Bill to do, if you were gone and your stuff caused him pain?

M: Oh, I’d want him to donate it. I wouldn’t want him to hold onto it.

T: So what might be a more useful thing to tell yourself?

M: He’d be ok with me giving his stuff away because it would be helping others

T: On a scale of 0 to 100, how much do you believe that’s true?

M: I would say about an 80; we had talked about this.

T: When you say that to yourself, what happens to your emotions?

Mike: They get better.

In this example, the therapist asked questions to guide Mike in questioning his assumptions, and to notice evidence that alternative thoughts may be true. The therapist then helped Mike formulate a more helpful thought and tested its usefulness by asking how it affected his feelings, using the 3C’s form.

Situation	Catch it	Check it	Change it
Deciding on what to do with Bill's things	"If I get rid of his clothes, I'm betraying him." "I shouldn't want to give his things away."	Emotions: Shame (90) Guilt (90) Anger (60) Thinking error: "Should" thought	"Giving away Bill's clothes would be helping others, and he'd like that." Emotions: Shame (20), Guilt (20), Anger (30)

Socratic Questioning and Techniques for Getting at Thoughts. Also called "guided discovery," the overall purpose of Socratic questioning is to challenge the accuracy and completeness of thinking. These are open-ended questions phrased, so people rarely respond with a "yes" or "no." It is important that the tone of the discussion remains collaborative and the therapist does not betray their judgment of the client's interpretations. Socratic questioning may be used at almost any stage of therapy and may be particularly useful when resistance is detected. There are several different categories of Socratic questioning, defined below with examples:

CONCEPTUAL CLARIFICATION QUESTIONS. These are basic "tell me more" questions that encourage deeper thinking:

- *What exactly does this mean?*
- *How does this relate to what we have been talking about?*
- *What is the nature of . . . ?*
- *What do we already know about this?*
- *Can you give me an example?*
- *Are you saying . . . or . . . ?*

PROBING ASSUMPTIONS. These are directed at uncovering assumptions underlying beliefs. This can be particularly helpful for identifying underlying core schema.

- *What would it mean about you if . . . ?*
- *You seem to be assuming . . . ?*
- *How did you choose those assumptions?*
- *Please explain why/how . . . ?*
- *How can you verify or disprove that assumption?*
- *What would happen if . . . ?*
- *Do you agree or disagree with . . . ?*

PROBING RATIONALE, REASONS, AND EVIDENCE. These uncover reasoning in arguments.

- *Why is that happening?*
- *How do you know this?*
- *Why do you think X causes Y . . . ?*
- *Are these reasons good enough?*
- *Would it stand up in court?*
- *How might it be refuted?*
- *What evidence is there to support what you are saying (or an alternative)?*
- *When did you first decide this was happening?*

QUESTIONING VIEWPOINTS AND PERSPECTIVES. Most arguments are given from a particular position. Show that there are other, equally valid viewpoints.

- *Another way of looking at this is . . . ; does this seem reasonable?*
- *What is an alternative way of looking at this?*
- *What is the difference between . . . and . . . ?*
- *Why is it better than . . . ?*
- *What are the strengths and weaknesses of . . . ?*
- *How are . . . and . . . similar/different?*
- *What if you compared . . . and . . . ?*
- *How could you look at this another way?*

PROBING IMPLICATIONS AND CONSEQUENCES. The argument given may have logical implications that can be forecasted. Do these make sense? Are they desirable?

- *Is . . . a helpful thought? Is . . . helping you reach your goal?*
- *How does this thought affect how you feel or what you do?*
- *Then what would happen?*
- *What are the consequences of that assumption?*
- *What are the implications of . . . ?*
- *How does . . . affect . . . ?*
- *Why is . . . important?*

Mindfulness and Acceptance-Based Interventions. These interventions can be particularly helpful in helping clients to observe their thoughts and emotions, and to also tolerate distressing situations. Key exercises include body scans (where the client learns to notice each part of the body in order, in a nonjudgmental way) and breath-focused meditation (paying attention to breathing without trying to change it). Clients then learn to use the skills of nonjudgmental observation practiced in these exercises in attending to experiences in their everyday lives.

Schema Change. CBT also addresses schema, defined as core beliefs and associated cognitive biases that are held by people. Common core beliefs in mood disorders focus on achievement ("I am worthless unless I am productive") or affiliation ("I am unlovable"), while in anxiety disorders they focus on vulnerability ("I can't cope; I am insufficient"). Due to the stigma and multiple failure experiences associated with severe mental illness, people with schizophrenia may have both types of core beliefs. In paranoia, core beliefs may also focus on persecution as an explanation for failure ("poor me") or as a deserved punishment ("bad me"). Schema change involves the other treatment techniques previously defined. In addition, the hypothesized schema is generally reviewed with the client, although how this is done and timed is in itself guided by the formulation. For example, someone who has a core belief of "I am incompetent" and a secondary assumption of “I can’t let others see who I really am” may find discussion of a core belief very disrupting early in therapy, whereas someone with the same core belief but a secondary assumption of “others need to help me because I can’t help myself” may find an early discussion positive and supportive. The costs and benefits of the schema are discussed in the context of the current problems and goals, and an alternative schema is derived that is more functional for clients while still being consistent with their values.

Mike frequently expressed the thought “I’m a screw-up; I’m not good enough.” His therapist noted that Mike seemed to believe that there was something inherently flawed or defective about himself and he deserved to be punished. Mike agreed with this. He noted that intellectually, he knew he was trying to be a good person, but emotionally he always felt there was something innately “broken” about him. Through the use of guided discovery, Mike tied this sense of himself as flawed to struggling with his sexual orientation as a young man and the teachings of the church in which he was raised, as well as his infidelities with Bill and Bill’s lasting anger about them. The therapist suggested to Mike that he consider another perspective; what would the people in his life say they valued most about him—Would they focus on his flaws or his gifts? Mike was able to share that his family and friends had all said how much they respected the love and dedication with which he cared for Bill. He also still received cards from former students sharing how important he had been to them as a teacher. In addition, although Bill frequently criticized him, Bill also was very loving and appreciative. In fact, most of the time, they were very loving toward each other, and Mike realized he tended to focus on the negative times. Finally, Mike and his therapist reflected on the strength and commitment it took to have stayed sober for 30 years, and the support he had provided to others in AA, particularly younger gay men struggling with their own sexual orientation.

Mike was willing to consider a more helpful view of himself, and came up with an alternative self-statement; “I am a caring guy.” He developed some new secondary assumptions; “I’ve done things that I’m ashamed of, but I have worked hard to acknowledge my mistakes and make amends” and “I’ve been a good influence in a lot of lives.”

While Mike agreed that these statements were true, he stated “they don’t feel true.” This is a common experience for clients when they generate more accurate thoughts that are in contradiction to long-held schema. Therefore, assignments were designed where Mike would act as if the new formulation was true, and see how it worked for him. This included increasing his contacts with family members and friends, and behaving like he deserved to be happy (e.g., doing enjoyable things). He found contacting people hard initially, but he was greeted with enthusiasm. As he started joining friends for activities, his mood started to improve and he found it easier to act on tasks around the house. He still missed Bill, but now had more support to cope and friends who expressed understanding about his loss.

Termination and Relapse Prevention. As clients achieve symptom improvement and functional gains, the emphasis of CBT shifts to preparing for termination. Treatment emphasizes clients developing their own interventions and experiencing mastery of skills. In addition, the therapist helps the client identify both persistent problems that need to be managed as well as potential future problems, and then identify the skills that can be used to address those problems. Particularly for older clients, termination can be stressful if they have lost other relationships recently. Therapists need to emphasize the goal of the client developing skills, and the importance of independent practice to adequately learn the skills. However, it can also be helpful to schedule follow-up “booster” sessions. This both provides the client with some sense of support and allows a chance to problem solve difficulties in implementation of relapse prevention plans.

Modifications for Psychosis

Psychosis is the most expensive mental disorder to treat, with hospitalization costs accounting for the majority of expenditures. The cost of care for psychosis patients over 65 years old is even higher than for patients aged 30 to 64. Best practice guidelines and nationalized healthcare systems recommend or mandate that CBT be delivered to consumers with schizophrenia as part of a comprehensive treatment program that includes pharmacotherapy and support resources. For example, CBT is among the recommendations made by the American Psychiatric Association, Department of Veterans Affairs, and the Schizophrenia Patient Outcomes Research Team (PORT), a project sponsored by the National Institute of Mental Health to develop evidence-based treatment guidelines for schizophrenia in the United States. PORT client survey findings, however, indicate poor conformance with this recommendation, especially for older people with psychotic disorders. Providers with expertise in delivering these interventions for older people with psychosis are very much needed.

In principle, all of the stages and techniques of CBT discussed above can be applied to the treatment of older people with psychosis, with some modifications but not fundamental differences. All of the techniques discussed above are appropriate for the treatment of the multiple problems faced by consumers with psychotic disorders. In this section, a brief discussion of the vulnerability-stress model and the general cognitive model of schizophrenia is provided, followed by descriptions of specific CBT interventions applied to psychosis, negative symptoms, and poor functioning in older people with schizophrenia.

Cognitive Model of Schizophrenia. In the context of the vulnerability-stress-coping model of schizophrenia, personal enduring vulnerability traits, like cognitive and neurobiologic dysfunction, interact with environmental stressors and coping responses to contribute to symptom formation. Neurocognitive deficits exacerbated by stress hyper-reactivity can lead to decompensation and the emergence of delusional beliefs and hallucinations, as well as diminished capacity to accurately evaluate beliefs. People with psychosis, like healthy individuals, have great difficulty distancing themselves from their highly charged beliefs but also have the added problem of insufficient cognitive resources available to evaluate and correct them. Rigid thinking, perseveration, cognitive biases, impaired abstract reasoning and memory, and cognitive distortions can play an important role in the development of delusional beliefs by leading to exaggerated appraisals of negative life events that become highly charged delusional belief systems.

In addition, multiple failure experiences due to cognitive deficits, odd behavior, and social withdrawal can lead to dysfunctional performance beliefs and social aversion attitudes (e.g., “I am inferior,” “Why bother; I will fail,” “If you can’t do something well, why do it at all”? and “People are against me”), which can exacerbate maladaptive behaviors, such as social withdrawal and suspicious avoidance. In the context of generic cognitive model applied to schizophrenia, dysfunctional attitudes like these may contribute to negative symptoms and poor functioning. Negative appraisals about one’s self and one’s ability to perform goal-directed tasks contribute to passivity and avoidance of productive activities, as a defense against anticipated failure and negative evaluations by others. This avoidance and lack of effort manifest as diminished motivation (avolition-apathy), loss of pleasure (anhedonia-asociality), impoverished speech (alogia), and blunted affect. CBT can be used to address and challenge beliefs about psychotic symptoms, social aversion attitudes, and defeatist performance beliefs, including expectancies (“It won’t be fun”), self-efficacy beliefs (“I always fail”), and anomalous beliefs (“Spirits will harm me”) that interfere with functioning. In addition, by including behavioral interventions, such as social skills training, skills deficits that interfere with social functioning can be addressed. By challenging self-defeating beliefs that interfere with skill performance while addressing skill capacity, people are more likely to engage in functional behaviors and use the skills they have.

Psychoeducation for Psychosis. Education is a key component in helping clients with psychotic symptoms make sense of their experiences. Psychoeducation is best accomplished as a two-way process, whereby the clients themselves come up with psychological explanations of their symptoms using guided discovery with inclusion of information from the therapist only where necessary. It is typically helpful to review the history and symptoms at the onset of psychotic illness in the context of the initial stressors that may have triggered it, using reflections like, “So, with all these stressors going on, including your father’s death, you suddenly began hearing this voice, and tried to make sense of it." Further, "Given that it seemed that nothing else could explain where this voice was coming from, I can see why you decided it was your father talking." Comments like this emphasize both that data were gathered that led to a conclusion about the source of the voice (an empirical approach) and that there were life events that contributed to the experience. This can set the stage for gathering evidence that might dispute these long-held conclusions, as well as identifying potential relapse triggers for relapse prevention plans. It can be challenging to review the first episode in older clients when the first episode was often decades prior. Collateral informants and record reviews can be helpful here, but the most important information is the evidence and context recalled by the client. In older clients, reviewing the impact of recent losses and bereavement (and sometimes anticipated losses or illness of loved ones) on symptom exacerbation can be useful to illustrate the impact of stress on symptoms. In the context of the vulnerability-stress model, it is often useful to focus on "normalization," which involves presenting the idea that any individual is vulnerable to experiencing psychotic symptoms if exposed to certain circumstances and sufficient stresses (e.g., sleep deprivation, unusual severe stressors, or substance abuse). This "normalization" discussion can serve to reduce stigma associated with severe mental illness. The focus on normalization and open questions and reflections directed at an honest understanding of how the client arrived at delusional beliefs often facilitates rapport and builds a trusting relationship. Time and patience are needed in this process, as trust can be slow to develop with some clients (e.g., with paranoia).

Clients are introduced to the general concepts of CBT, including the relationship between thoughts, behaviors, and feelings (cognitive model). Multimedia presentation of information can help compensate for neurocognitive impairment associated with both psychotic illness and aging. For example, drawing figures, like the one below, on a pad (or handout) for clients to take home can help compensate for cognitive impairment (Fig. 55.4i–2).

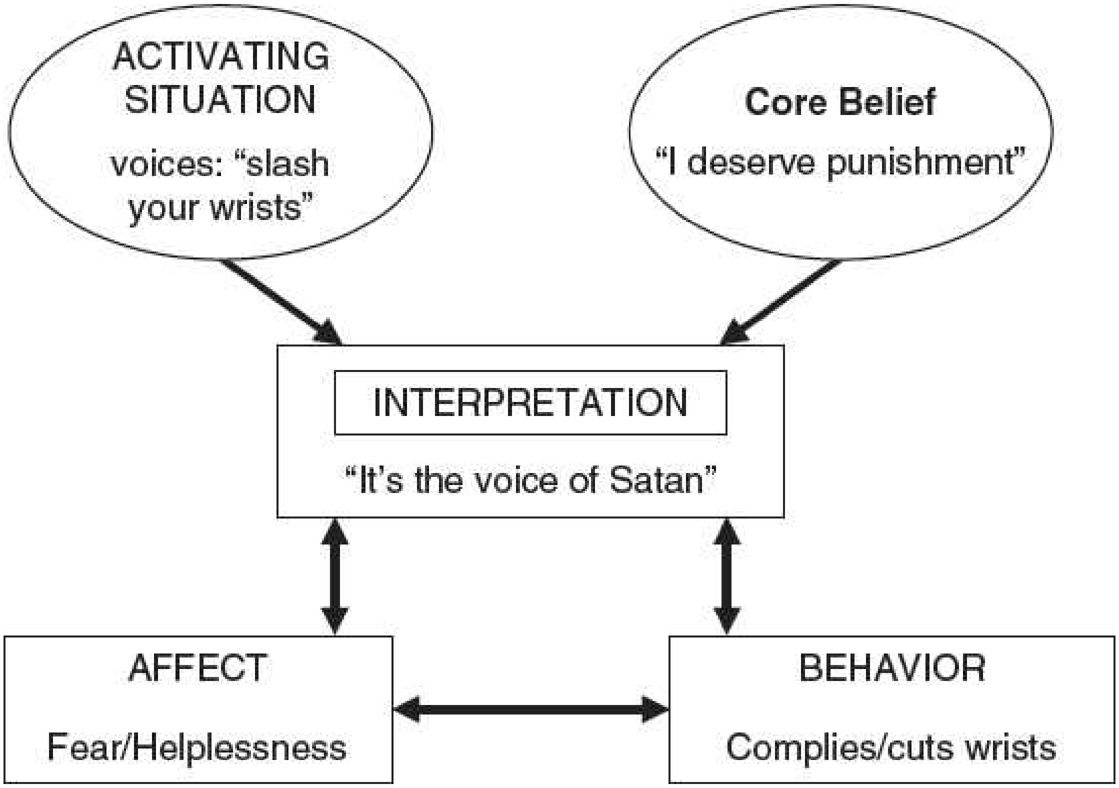
ACTIVATING SITUATION
voices: "slash your wrists"
Core Belief
"I deserve punishment"
INTERPRETATION
"It's the voice of Satan"
AFFECT
Fear/Helplessness
BEHAVIOR
Complies/cuts wrists

FIGURE 55.4i–2. Generic cognitive model applied to psychosis.

This example helps to illustrate the importance of interpretations or thoughts about experiences, such as auditory hallucinations. It is important to stress that it is not the voice itself that leads to unwanted feelings or behaviors, but the thought about the voice. It is more distressing to think Satan is making threatening demands (Satan is a powerful, controlling, external agent) than to think voices are due to brain changes caused by illness. If the interpretation is that voices are due to illness (no power to harm, controllable, internal origin), then fear, helplessness, and risk of self-harm can be reduced. The core belief ("I deserve punishment") is provided in Figure 55.4i–2 to illustrate the importance of developing a formulation about highly charged beliefs that drive interpretations. It is typically not useful to complicate the model by including this aspect in the psychoeducation process early in therapy, although it may be added later.

Behavioral Interventions. Behavioral activation, pleasure predicting, and relaxation training techniques described above are all relevant to the older client with psychosis, and these techniques are used in essentially the same way. Particularly important treatment targets for older people with schizophrenia or other psychotic disorders are communication skills and psychosocial interactions (e.g., asking someone to share a leisure activity, making positive requests for support from friends, family, or professionals, expressing anger and assertiveness). The predominant therapeutic technique used to improve communication skills is behavioral role-play. Role plays are also useful for generating thoughts that can interfere with adaptive social interaction (e.g., "He thinks I'm dumb"; "He won't want to talk to me"; "He wants something from me"). These thoughts can then be targeted for change using thought-challenging interventions (e.g., the 3C's). Low self-efficacy and defeatist performance beliefs ("I will forget what to say"; "I will mess it up") can also be identified by using scaling techniques, where the therapist asks clients to rate their expectations for success in the role play on a scale of 1 to 10 before practicing and again rate performance after practice. The therapist can then compare these ratings to point out how people perform better than they think they will (pre–post ratings almost always increase), so clients can raise expectations for success to increase the likelihood that they will use the skills in the real world. In this way, the role-play itself can serve as a behavioral experiment to test out defeatist beliefs. By teaching new social skills while changing the defeatist thoughts that interfere with skill performance, negative symptoms and functioning can improve.

Role plays can focus on a number of topics. Clients are more motivated to learn skills when role-play topics are linked to the client's personal goals (e.g., role-playing how to ask someone to coffee if the client's goal is to make a friend). Expressing positive and negative feelings and making a positive request can be emphasized to improve the assertive, nonaggressive, clear, and comfortable sharing of feelings in social interactions. Improving everyday leisure activities and psychosocial functioning are common role-play topics for these skills (e.g., asking a roommate to change their behavior; or asking someone to go to the movies). Another common focus of role plays is on identifying warning signs of psychotic symptom exacerbation and practicing reporting these to the doctor. It is sometimes useful to make a video recording of role plays and play back the recording in session. With the help of the therapist, clients critique the strengths and weaknesses of their own performance according to a list of key skills displayed on a handout or poster board (e.g., eye contact, speech volume). Clients sometimes have little insight into the impact of their negative symptoms (e.g., low voice volume, poor eye contact, blunted affect) on their social skills until they watch themselves on video. The group therapy setting is ideal for role plays. Members with better skills model for others, and this leads both to a sense of mastery for the skilled client and an increased perception that the skill is learnable for the unskilled client. Clients can sometimes accept feedback better from peers than from therapists, and situations played between group members are more similar to in vivo experiences than role plays between therapist and patient.

Given the loss and isolation experienced by older clients, one important focus is to help clients recognize and identify support persons in their lives. Older clients with psychosis often have a difficult time recognizing that some people in their lives truly are support persons, such as health care providers, board-and-care workers, and peers. Often, clients believe that only family or relatives can act as support persons. Because many older clients with schizophrenia no longer have connections with their family or relatives, it becomes increasingly important to identify support persons and help clients to recognize that many others in their life outside of family can act as supports.

Cognitive Interventions. Thought-challenging interventions (e.g., using the 3C’s) for clients with psychosis are delivered in essentially the same way as for clients with depression or anxiety. Clients are introduced to the general concepts of CBT, including the relationship between thoughts, actions, and emotion (generic cognitive model); identifying automatic thoughts; and challenging unhelpful thoughts by examining evidence for beliefs and mistakes in thinking (e.g., jumping to conclusions, mind reading, all-or-none thinking). Through in-session discussions and homework assignments, clients are taught to identify thoughts; identify relationships between thoughts, feelings, and behaviors; and identify mistakes in thinking. Thought records are used to elucidate the relationship between situational triggers (e.g., people speaking outside the client’s room) and thoughts (e.g., “They’ve come to hurt me”) and subsequent changes in feelings (e.g., fear) or behaviors (e.g., isolation, aggression) in response to thoughts. Thoughts that lead to distressing feelings or maladaptive behaviors are labeled as “unhelpful,” and are examined for accuracy. Socratic questioning is used to help clients examine the logic of beliefs and generate more adaptive alternatives to mistakes in thinking or thoughts without sufficient evidence. The primary thoughts targeted are typically beliefs about voices (e.g., “My voices are from an external being,” “My voices have power to harm me,” “My voices cannot be controlled”), beliefs about daily events that clients use as evidence to confirm delusions (“That person on the bus was part of the conspiracy”), and dysfunctional performance beliefs (e.g., “Why bother, I will fail,” “I am too sick”) that interfere with adaptive functioning.

Using the 3C’s technique for thought challenging is particularly helpful to compensate for neurocognitive impairment that is common in older clients with severe mental illness. The 3C’s technique simplifies learning and serves as a mnemonic device to help clients remember and use cognitive techniques in everyday life. Table 55.4i–3 shows examples of a 3C’s thought record for beliefs about psychosis. Thought records like this are completed both in session and as skills practice assignments between sessions. Situational antecedents are typically first linked to the feelings and behaviors that arise in them. This is done to emphasize that clients need to use these unwanted feelings and behaviors as an alert or “red flag” reminder to examine their thinking (“Catch it”). Once the thought linked to the feeling or behavior is identified and written down, the impact of the thought in terms of emotions is rated, possible mistakes in thinking are identified, and the evidence for and against the thought is reviewed (“Check it”). If the thought does not “check out” (e.g., there is evidence the thought is unhelpful), it is replaced in favor of a more balanced, more helpful thought (“Change it”). The group setting can facilitate thought challenging because other members are often better able to notice delusional thinking and articulate logical flaws, contrary evidence, and alternative thoughts than the client, who has an emotional conviction attached to the inaccurate thought.

Table 55.4i–3.
Example of 3C's Thought Record for Beliefs About Psychosis

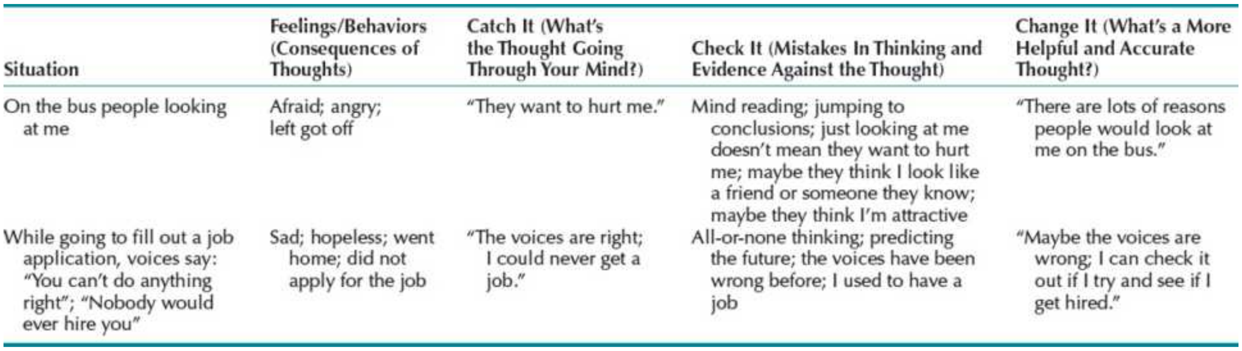

Situation	Feelings/Behaviors (Consequences of Thoughts)	Catch It (What's the Thought Going Through Your Mind?)	Check It (Mistakes In Thinking and Evidence Against the Thought)	Change It (What's a More Helpful and Accurate Thought?)
On the bus people looking at me	Afraid; angry; left got off	"They want to hurt me."	Mind reading; jumping to conclusions; just looking at me doesn't mean they want to hurt me; maybe they think I look like a friend or someone they know; maybe they think I'm attractive	"There are lots of reasons people would look at me on the bus."
While going to fill out a job application, voices say: "You can't do anything right"; "Nobody would ever hire you"	Sad; hopeless; went home; did not apply for the job	"The voices are right; I could never get a job."	All-or-none thinking; predicting the future; the voices have been wrong before; I used to have a job	"Maybe the voices are wrong; I can check it out if I try and see if I get hired."

To gather evidence to evaluate thoughts, behavioral experiments are conducted inside and outside of sessions (homework). For example, one common belief is that a voice is from an external agent that is omnipotent (e.g., the voice of God or the Devil). A behavioral experiment to test this belief might involve training clients in distraction techniques to stop or quiet hallucinations, such as humming, listening to music, shadowing what the voices say, watching TV, or counting. If clients can control the voices to any extent using these techniques (in session if hallucinations are triggered or at home for homework), this is discussed as evidence against the omnipotence of the voices. How can simply humming or turning on the TV control God or Satan? The following therapist–client interaction illustrates a behavioral experiment.

Pat had been experiencing some paranoia that was causing her to isolate herself at home.

P: I'm really confused because I had so many coincidences happen that I was writing them down. Then my friends will say, "Are you taking your medication"? when I tell them about what's been happening. But then they can't deny that there are so many of them and it is really weird. And then it becomes so weird that mathematically it would be like getting hit by lightning two or three times. It's just that the odds are such long shots that all this would be happening. Then I start thinking that I must be under some kind of mind control and it's preplanned because the odds of this happening couldn't possibly be this strong. So this has been difficult. This is confusing.

T: What kinds of things have been coincidences in the past few weeks?

P: Just like a matter of 2 days. What I'm thinking about or what I'm doing, the next thing I'll see it in print, or it will be said on the TV. Whatever I'm thinking about, one thing or another after another. Just like, I was thinking about hermaphrodites and people who are male and female and I open my book and it's right there. Talking about androgynous people, male and female people as I was just thinking it, and then several conversations like that came up. What I was talking about or what I was thinking happened right then and there, at that moment, and I keep trying to ignore it 'cause people have said something about free associating or something like that, too much. It just got to be too much of it happening. Either I'm at some really strong path with God or I'm under some kind of mind control. I don't know which.

T: How does that impact your life? How does it get in the way of you obtaining your goals?

P: When I feel my mind controlled, for me, I end up deciding not to move at all. Like the only way is to contain it. Like it's a disease, infection. I just feel diseased, like I have an infection. Like nobody should be near me.

T: Because we might catch it?
P: Yeah, or something like that.
T: So we could catch the coincidence disease?
P: Yeah.
T: What would happen to us?
P: You'd be in the same situation.
T: Which is?
P: Being controlled.
T: So we would feel the same way you're feeling. Well, are you feeling that way right now?
P: Yeah, pretty much.
T: What if I came and sat next to you? Can we try this out and see if I catch what you've got? How about it?
P: I don't think you would know right away.
T: How long would it take me to know?
P: Just until you start noticing those types of coincidences all happening to you.
T: How long would it take me to know?
P: I don't really know.
T: Approximately?
P: I don't know. I'm just guessing how all this works.
T: Well, would you mind? I mean, because you're going to be in therapy for several months. So do you think it could happen during that time?
P: Yeah.

The therapist sat next to Pat for the next 2 months and revisited the idea weekly that the "coincidence disease" was contagious. Over time, Pat continued to notice coincidences, but the inferred controlling nature of these occurrences decreased over time, in part, because her conviction in the belief that it was a disease diminished. This is an example of thought chaining—if others don't catch it, maybe it is not a disease, and if I don't have a disease, then the cause of coincidences is not a controlling disease. As conviction in this delusional belief diminished, Pat was willing to try more activities outside the home, which provided additional opportunities for behavioral experiments to test out possible causes of coincidences, with a focus on generating alternative hypotheses.

The following case example illustrates how cognitive-behavioral interventions can be used to address command hallucinations, paranoia, and medication adherence.

Mark is a 61-year-old unmarried Caucasian male who is unemployed and taking an art class at a local community college. He served in the U.S. Army for 6 years, from ages 18 to 24 but had not worked since his discharge from active duty. He was not in combat and denied trauma (he was trained in a logistics/support role). He was medically discharged from the military for Psychosis, Not Otherwise Specified (NOS) and received benefits income from the Department of Veterans Affairs (VA). He received care from the local VA facility, including antipsychotic medication, and was diagnosed with schizophrenia, paranoid type.

He lived in a group home in a financially depressed area of his city. He had several psychiatric hospitalizations, often stemming from physical altercations, including assaulting an individual he thought was going to hurt the pastor of his church. He was arrested but not charged with this assault. He reported frequently hearing voices, including command hallucinations, especially when he forgot to take his medications, and ascribed these assaults to the voices (“The voices tell me to stop bad people from hurting other people”). He also experienced difficulty with paranoid ideation, reporting he often thought people were watching and talking about him on public transportation and at his group home.

Mark had a strong identity as an artist, and pursued art consistently for many years since leaving the military. He sculpted using clay in his bedroom and showed his therapist pictures of his sculptures. However, he reported that the limited space in his room prevented him from working on his art consistently. He also recognized that when his symptoms increased, it interfered with his working on his art. Therefore, the consumer and therapist developed goals to increase his opportunity to practice art and better manage his hallucinations and paranoia by learning cognitive-behavioral strategies and improving his medication adherence.

Treatment involved using the 3 C's and behavioral experiments to examine thoughts about the omniscience of his command hallucinations (e.g., "They know who will hurt my friends") and medications (e.g., "It's no big deal if I forget my meds"). For example, he would *C*atch thoughts about the omniscience of his voices ("They can tell who has evil in them"), *C*heck the evidence against the thought ("They were wrong about the guy they said would hurt my pastor") and the consequences ("I get arrested when I listen to them"), and *C*hange the thought to be more realistic and helpful ("Maybe the voices are wrong"). After his conviction about the omniscience of the voices was reduced, he was more willing to ignore them and not comply with their commands to protect others in order to test whether their threats would come true. By not complying, he was conducting a behavioral experiment that produced further evidence that he did not have to do what they said to protect himself or others. He also reported that his voices were from "good spirits; maybe angels" who would warn him about "people with evil in them." The fact that he thought the voices were benevolent and might protect him from harm contributed to poor medication adherence because he did not want to take medications that quieted or eliminated the voices that he perceived as helpful. The therapist used Socratic questioning to challenge these beliefs about the voices as benevolent and omniscient external agents:

T: You mentioned that you have noticed that the medications can make the voices go away, right?

M: Yes, I have noticed that.

T: That makes me a little confused, though. How can something you put in your body—medications that change your brain chemistry—make angels or something outside of your body stop talking?

M: Well, I'm not sure but it does.

T: But does it make sense that a chemical inside you could change an angel or some kind of powerful supernatural spirit? How does that work?

M: I don’t know but I get nervous when it’s been a while since the voice has given me any warnings or told me who to watch out for.

T: I can see how you might want to stop taking the medication for a couple of days when you get nervous about that, so the voices might come back, but does it make sense that medications would make angels or spirits come and go? What if the voice isn’t from an angel or spirit? What if it is from inside you and not an outside spirit? That would help explain why changing your brain chemistry can stop it. What do you think about that possibility?

M: Well, that part adds up, I guess, but I’m just not sure.

T: And we already discovered that the voices are not always correct about who is dangerous. That also makes me think the voices are from you rather than an angel because people are more likely to be wrong about things than angels. We all make mistakes, but angels probably don’t.

CLINICAL ISSUES

Cognitive-Behavioral Therapy and Medications

The use of CBT forms one part of a package of care that often includes the use of medications, where relevant and appropriate, as well as other services (e.g., case management, vocational rehabilitation, couples/family therapy). This is particularly true for psychosis. For depression, meta-analyses of CBT compared to antidepressants indicate that effects are similar (d = 0.35), and that response to CBT among older adults is similar to effects seen in younger adults. A meta-analysis including head-to-head comparisons of antidepressants and CBT in older adults found no differences between treatments, but the authors concluded the number and quality of studies were too low to make definitive conclusions. Although older adults report a preference for talk therapies over medications, the two are often prescribed together. This generally means that cognitive therapy is used alongside medication in a synergistic way, such that CBT improves medication adherence and medication controls acute symptoms that might interfere with CBT. It is recommended that, if the CBT therapist is not the prescriber, a two-way release of information be obtained from each client in order to facilitate continuity of care and communication.

Individual versus Group Format

There are several advantages to using these interventions in a group format. As noted above, group CBT is one of the first-line recommendations for the treatment of depression in older adults. Groups are widely used in mental health clinics (increasing the likelihood that these interventions will be acceptable to such sites); treatment can be provided for a greater number of clients in groups than could be accomplished in individual sessions; groups can impact social support systems and allow the practice of communication skills; groups provide a shared identity with others (i.e., healing power of connections with others, “me too” and “not alone” influences); confronting or observing others in groups may provide better insight into one’s own symptoms and skills deficits, and groups can provide social reinforcement for change. One important difference between group and individual CBT formats is that groups tend to emphasize skills training over detailed case formulation. There is less time in groups to fully explore the unique content and history of each client’s presenting concerns in developing a formulation. There is also greater flexibility to tailor therapy to the individual needs of the client in individual therapy.

Goal-Setting Challenges

Setting goals for therapy with older clients who have been ill for decades can present unique challenges encountered less frequently with more recent onset clients. Particularly for older clients who have had a chronic or recurrent course of a disorder, there may be low expectations for treatment success. Information-processing biases also often lead to a focus on failures and ignoring past achievements. By acknowledging a goal, older clients must accept the risk of yet another failure. This risk assessment is often exaggerated into risk of symptom exacerbation and rehospitalization. For patients with psychosis in particular, by this stage of illness, older people have also been informed that they have a severe, chronic brain disease with little hope of recovery. Older clients have had much less exposure to modern recovery and psychosocial rehabilitation models that have only recently begun to transform health care systems to focus more on strengths and hope for functional achievement. Hopelessness in providers can be passed on to older chronic clients. While not as potentially profound as psychosis, other depressive and anxiety disorders that have followed a chronic course can also facilitate hopelessness for change.

These factors can create challenges for eliciting a goal for therapy. Setting a treatment goal may require patience across several sessions. Cognitive therapy tools can be used to address dysfunctional performance beliefs, negative expectancies, and ignoring the positive and other attributional styles that can lead to hopelessness and reluctance to set a meaningful goal. Older clients may also initially offer a goal that they have learned providers want to hear (e.g., "I want to take my medications more regularly"), rather than a more personally meaningful goal. Goals may need to be revised as more meaningful goals emerge. It may be useful to ask clients to recall a time when they were functioning better in the past or imagine a time years in the future when they are much better and ask about daily behaviors. Ask, "What would you be doing if you weren't feeling this way"? This can stimulate a helpful discussion about desired, meaningful functional behaviors (relationships, volunteer work, leisure activities). The key is to get a personally meaningful goal.

We have found it most helpful to focus on specific, functional behavioral goals (increasing socialization, relationships, leisure activities, living skills to promote independent housing, school, work/volunteering) chosen by clients. Emotional or symptom goals (e.g., "I want to stop being sad" or "I don't want to hear voices") are less helpful unless framed in terms of their impact on functional behaviors. For example, one can ask, "How would you know you weren’t sad anymore? What would you be doing”? The response is often something like, “I would be getting up and doing more things.” The therapist can then ask what things, and frame the functional behavior as the goal (e.g., doing things with friends, taking a class). Goals should also be specific and stated as attainable behaviors. For example, the therapist can ask, “How will you know the goal has been achieved? What will you be doing differently each day if you achieve the goal”? The therapist’s goal is to help older clients set meaningful, realistic, attainable, specific goals stated in terms of functional behaviors.

Importance of Homework

One goal of CBT is to provide skills that generalize and are used by the client outside therapy in everyday life. To that end, treatment includes both skill practice in sessions as well as homework assignments. In-session practice is a crucial means by which clients understand the sometimes difficult concepts in this intervention. Homework assignments are crucial for generalizing these skills to the “real world,” but homework adherence can sometimes be challenging. The term “at-home practice” is often used, rather than “homework,” to emphasize the importance of practice outside of sessions and to reduce the negative connotation associated with “homework” assignments back in school days. Adherence to at-home assignments can vary considerably between clients. Older clients in particular may have expectations that the active treatment factor in therapy is disclosure rather than practice, and this can interfere with willingness to complete outside assignments. To facilitate homework completion, the therapist guides, instructs, and explains the rationale for completing home activities, and activities are practiced in session. The therapist goes over several examples so that clients can more easily do the remainder of the assignment at home. An analogy with learning to play a musical instrument can be used to stress the importance of practicing skills (e.g., “Just like learning to play the guitar: if you only practice when you are at your music lesson, you won’t learn songs as fast as if you practice in between lessons”). Principles of operant conditioning are also used to facilitate homework completion (e.g., positive attention and praise to those who complete homework). The therapist makes clear to clients that homework completion is expected and failure to complete assignments is the topic of thorough problem solving to help the client do a better job next time. The authors have had success in obtaining 80% homework completion in 80% of older clients with schizophrenia in their CBT for psychosis groups. Previous research with other populations has shown that CBT incorporating homework assignments helps clients improve about 60% more than those in treatment without homework.

Indications

CBT treatment targets may change as people age. There is evidence that CBT is efficacious for a range of disorders in older adults, including depressive disorders, anxiety disorders such as panic disorder and social phobia, and psychosis. In addition, while there are a number of pharmacologic treatment options available, clients often desire psychotherapy options. Of note, although positive symptoms tend to be reduced in older clients and may often be fairly well controlled through pharmacologic treatment, these clients commonly still experience negative symptoms and severe psychosocial functioning deficits, which are important treatment targets of CBT in older people with psychosis.

Contraindications

A number of factors may interfere with psychotherapy overall. For all older adult clients, it is important to assess the possible causes of many psychological symptoms, such as fatigue, appetite, and sleep disturbance, because medical comorbidity and normal age-related changes may contribute. Comorbidities, such as substance abuse and medical illness, may interfere with attendance in CBT. This barrier can be overcome by conducting therapy in the client’s home or senior center in the neighborhood, in the presence of a caregiver, by finding a resource for transportation, or by providing treatment via videoconferencing. In addition, CBT interventions can include comorbidities as treatment targets (e.g., reducing substance use through harm reduction techniques, improving health status through addressing obstacles to treatment adherence). Poor insight (awareness of having a mental illness that requires treatment) is prevalent in some psychotic disorders and predicts poor course of illness and functioning. There is also some evidence that insight moderates treatment outcome in CBT for psychosis. Severe disorganization in psychotic clients presents a number of challenges to the logical, hypothesis-testing process of cognitive therapy, but some of these clients succeed despite thought disorder, and there is no clear data on the impact of this potential barrier. The authors have found that disorganization frequently fluctuates, as well, so a more behavioral intervention focus can be used until clients are more able to communicate more linearly. Neurocognitive impairment is often assumed to interfere with CBT, but this assumption may not always be accurate. Progressive dementia will likely interfere with CBT and neurocognitive impairment does predict poorer outcome in serious mental illness, in general, as well as treatment outcome in psychosocial interventions. However, there is no evidence that neurocognitive impairment reduces the benefit of CBT in serious mental illness relative to standard pharmacologic care alone. That is, even participants with moderately severe neurocognitive impairment show added benefit from receiving CBT plus pharmacotherapy, relative to receiving only pharmacotherapy. With little clear evidence about which clients will and won’t benefit from CBT, the authors recommend that willing clients be given a chance to succeed. The biggest barrier to success may be provider hopelessness; clients who are never referred for CBT because they are thought to be too old or too impaired in some way can never succeed.

Cognitive Therapy versus Cognitive Remediation

Two broad categories of cognitive interventions have been developed: Cognitive remediation, which focuses on rehabilitation of specific cognitive functions (e.g., memory, attention), and CBT, which trains participants to use metacognitive strategies to identify and challenge unhelpful or inaccurate thoughts. These areas of intervention have several factors in common and may lie on a continuum from more “molecular” to more “molar” cognitive rehabilitation approaches, but they are often incorrectly equated. At one end, cognitive remediation targets specific problems of neurocognitive processing, and at the other end, CBT targets more global cognitive constructs, such as thinking and self-concept. Clients with severe mental illness typically have neurocognitive impairment and may benefit from cognitive remediation interventions and/or medications that attempt to improve neurocognition. Combining these interventions may enhance outcome in CBT, but cognitive remediation is not cognitive therapy.

RESEARCH AND EVALUATION

Depression

Recent studies indicate that CBT is as effective as pharmacotherapy for treatment of major depression. A meta-analysis of treatments for late-life depressive disorders found that pharmacotherapy and psychotherapy (most often, CBT) were similarly effective in terms of symptom reduction. The American Psychological Association (APA) depression treatment guidelines for depression listed group CBT as a recommended treatment for older patients with depression, and individual CBT, either alone or in combination with pharmacotherapy, as a conditional recommendation if a first-line recommendation is not available or acceptable. There is currently little evidence of differential effectiveness of pharmacotherapy versus psychotherapy in the treatment of major depression and related mood disorders, and therefore, recommendations for treatment should be based on factors such as patient preference and possible medical implications. In particular, older clients may have more difficulty with side effects of medications or have concerns about interactions with other medications, making CBT a particularly appropriate option. Finally, treatment preference studies overwhelming suggest that older adults prefer talk therapies over medications, so these treatments should be accessible in settings that serve older adults.

Anxiety

There are fewer well-designed studies available on CBT for late-life anxiety disorders, especially trials comparing CBT with pharmacotherapy for late-life anxiety disorders, and the combination of medication and psychotherapy has not yet been formally tested in an adequate clinical trial. Most randomized, controlled trials have compared CBT to other psychotherapies or control conditions, and most have examined generalized anxiety disorder or enrolled clients with a range of anxiety diagnoses without differentiating between them. A recent meta-analysis of 14 randomized controlled trials (RCTs) for generalized anxiety disorder (GAD) found that CBT was superior to no treatment or treatment as usual, but there was insufficient evidence to draw conclusions in comparison to other active treatments. An earlier review suggested that medications, particularly selective serotonin reuptake inhibitors, appear to be more effective than psychotherapy as an acute treatment for late-life anxiety disorders. In general, results of psychotherapy trials suggest that CBT has some benefit, although clients do not on average reach a symptom-free status. Cognitive-behavioral interventions that include relaxation training appear to be more effective than those without. For older individuals who refuse to take medications or for whom they are contraindicated, relaxation training may be the most effective component of a behavioral treatment, but direct comparisons have not been conducted. At this point, there is need for better-controlled trials of CBT for specific anxiety disorders, and in particular, trials that compare CBT with pharmacotherapy as well as trials examining combined pharmacotherapy and psychotherapy, especially in older clients. Nonetheless, the current literature indicates that CBT is a potentially useful approach for older clients with anxiety disorders.

Schizophrenia

Only a handful of studies have examined any form of psychotherapy interventions for older people with schizophrenia. Over 50 clinical trials of CBT for psychosis have been conducted with middle-aged and younger people with schizophrenia. These clinical trials found significant improvements in hallucinations and delusions (typically the primary treatment target), community functioning, negative symptoms, social anxiety, and mood. Available clinical trials, however, have concentrated on the adult population up to 65, so their generalizability to older groups needs to be considered cautiously. The authors developed, manualized, and conducted clinical trials of a novel psychosocial rehabilitation intervention called cognitive-behavioral social skills training (CBSST) in older people with schizophrenia (mean age 58). CBSST combines techniques from both CBT and social skills training to target the multidimensional deficits that contribute to disability in aging people with schizophrenia. In four separate clinical trials, the authors have found significant improvements in social skills and community functioning in CBSST relative to standard care and relative to a goal-focused supportive contact control group, and these gains in functioning were maintained at 1-year post-treatment follow-up. A recent review of psychosocial interventions for improving functioning in middle and older age clients with severe mental illness concluded that the strongest evidence was for skills training interventions that incorporate either cognitive or behavioral techniques and use a group format. However, there is clearly a need for expanded research, particularly in the older population.

FURTHER READINGS

*American Psychological Association. *Clinical Practice Guidelines for the Treatment of Depression Across Three Age Cohorts*. American Psychological Association; 2019.

Ayers CR, Sorrell JT, Thorp SR, Wetherell JL. Evidence-based psychological treatments for late-life anxiety. *Psychol Aging*. 2007;22:8–17.

Beck AT. *Cognitive Therapy and the Emotional Disorders*. International Universities Press; 1976.

Burns DD. *The Feeling Good Handbook: Using the New Mood Therapy in Everyday Life*. Morrow; 1989.

Craske MG, Treanor M, Conway CC, Zbozinek T, Vervliet B. Maximizing exposure therapy: an inhibitory learning approach. *Behav Res Ther*. 2014;58:10–23.

*Cuijpers P, Karyotaki E, Pot A, Park M, Reynolds C 3rd. Managing depression in older age: psychological interventions. *Maturitas*. 2014;79:160–169.

Cuijpers P, van Straten A, Smit F, Andersson G. Is psychotherapy for depression equally effective in younger and older adults? A meta-regression analysis. *Int Psychogeriatr*. 2009;21(1):16–24.

Cuijpers P, van Straten A, Warmerdam L, Andersson G. Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. *Depress Anxiety*. 2009;26(3):279–288.

Dwight Johnson M, Apesoa-Varano C, Hay J, Unutzer J, Hinton L. Depression treatment preferences of older white and Mexican origin men. *Gen Hosp Psychiatry*. 2013;35:59–65.

Gatz M, Fiske A, Fox LS, Kaskie B, Kasl-Godley JE. Empirically validated psychological treatments for older adults. *J Ment Health Ageing*. 1998;4:9.

Gould RL, Coulson MC, Howard RJ. Cognitive behavioral therapy for depression in older people: a meta-analysis and meta-regression of randomized controlled trials. *J Am Geriatr Soc*. 2012;60:1817–1830.

Gould RL, Coulson MC, Howard RJ. Efficacy of cognitive behavioral therapy for anxiety disorders in older people: a meta-analysis and meta-regression of randomized controlled trials. *J Am Geriatr Soc*. 2012;60:218–229.

Granholm E, Holden J, Link PC, McQuaid JR, Jeste DV. Randomized controlled trial of cognitive behavioral social skills training for older consumers with schizophrenia: defeatist performance attitudes and functional outcome. *Am J Geriatr Psychiatry*. 2013;2:251–262.

*Granholm E, Holden J, McQuaid JR. *Cognitive Behavioral Social Skills Training (CBSST) for Schizophrenia: A Practical Guide and Consumer Workbook*. The Guilford Press; 2015.

Gum AM, Arean PA, Hunkeler E, et al. Depression treatment preferences in older primary care patients. *Gerontologist*. 2006;46(1):14–22.

*Hall J, Kellett S, Berrios R, Kaur Bains M, Scott S. Efficacy of cognitive behavioral therapy for generalized anxiety disorder in older adults: systematic review, meta-analysis, and meta-regression. *Am J Geriatr Psychiatry*. 2016;24(11):1063–1073.

Hantke N, Lajoy M, Gould CE, et al. Patient satisfaction with geriatric psychiatry services via video teleconference. *Am J Geriatr Psychiatry*. 2020;28(4):491–494.

Jayasekara R, Procter N, Harrison J, et al. Cognitive behavioural therapy for older adults with depression: a review. *J Ment Health*. 2015;24(3):168–171.

Kingdon DG, Turkington D. *Cognitive Therapy of Schizophrenia*. The Guilford Press; 2005.

*Laidlaw K, Thompson LW, Dick-Siskin L, Gallagher-Thompson D. *Cognitive Behaviour Therapy with Older People*. Wiley and Sons; 2003.

Secker DL, Kazantis N, Pachana N. Cognitive behavioral therapy for older adults: practical guidelines for adapting therapy structure. *J Rat Emot Cogn Behav Ther*. 2004;22:93–109.

Smart EL, Brown L, Palmier-Claus J, Raphael J, Berry K. A systematic review of the effects of psychosocial interventions on social functioning for middle-aged and older-aged adults with severe mental illness. *Int J Geriatr Psychiatry*. 2020;35:449–462.

Stanley MA, Bush AL, Camp ME, et al. Older adults' preferences for religion/spirituality in treatment for anxiety and depression. *Aging Ment Health*. 2011;15:334–343.

Thase ME, Kingdon D, Turkington D. The promise of cognitive behavior therapy for treatment of severe mental disorders: a review of recent developments. *World Psychiatry*. 2014;13:244–250.

*Thompson LW, ed. *Handbook of Behavioral and Cognitive Therapies with Older Adults*. The Guilford Press; 2008.

Thorp SR, Ayers CR, Nuevo R, Stoddard JA, Sorrell JT, Wetherell JL. Meta-analysis comparing different behavioral treatments for late-life anxiety. *Am J Geriatr Psychiatry*. 2009;17:105–115.

Wetherell JL, Gatz M, Craske MG. Treatment of generalized anxiety disorder in older adults. *J Consult Clin Psychol*. 2003;71:31–40.

Wolitzky-Taylor KB, Castriotta N, Lenze EJ, Stanley MA, Craske MG. Anxiety disorders in older adults: a comprehensive review. *Depress Anxiety*. 2010;27:190–211.

55.4j Family Intervention and Therapy with Older Adults

LAUREN N. DECAPORALE-RYAN, PH.D., ANN ELIZABETH CORNELL, PSY.D., C.E.A.P., JESSICA GOODMAN, PH.D., AND CAROL ANN PODGORSKI, PH.D., M.P.H., M.S.

Over the past decade, demographic trends have altered the composition of the older adult population in substantive ways. The older adult profile began to change in 2011, when the Baby Boom generation, currently one-quarter of the U.S. population, began to turn 65. With an increase of nearly 35% over the past decade, nearly 17% of the population is currently over age 65, with racial and ethnic minorities constituting 23%. Older adults are now comprised of two generational cohorts—the "traditionalists" born before 1946 and the "Baby Boomers" born between 1946 and 1964. These groups differ in terms of experiences, values, attitudes, and preferences. The Baby Boom generation is likely to have families that are more geographically dispersed, experience more divorces, have fewer biologic children, amassed more blended families, and are increasingly living in multigenerational households. Baby Boomers are more likely to be healthier, but with the increased longevity of “the traditionalists,” 70% of those over 65 have functional limitations. Similar to previous generations, their needs are met predominantly by families and friends.

The infiltration of Baby Boomers into the over-65 population has also triggered shifts in health care preferences and in relationships with medical professionals. This information-seeking, authority-challenging generation was instrumental in forming and shaping the movement toward consumer-directed health care. In addition, the newest older adults are more accepting of the notion of “mental health” and less resistant to seeking mental health care. According to the Substance Abuse and Mental Health Services Administration (SAMHSA), 14% of those over 65 had any mental illness in 2019, and 47% of them received mental health services. A growing body of evidence indicates that older adults generally accept and are satisfied with telegeropsychiatry services. Thus, the demand for mental health services among older adults is increasing, and the nature of the older adult patient–mental health clinician relationship will continue to evolve to reflect multiple generational and sociocultural influences.

Other changes in health care delivery approaches have increased the need for clinicians to hone their skills for working with families. The Patient Protection and Affordable Care Act, passed in 2010, placed a greater emphasis on integrated care through establishment of federally qualified patient-centered medical homes. According to the Agency for Healthcare Research and Quality, “integrated behavioral health care” is the care a patient experiences as a result of a team of primary care and behavioral health clinicians, working together with patients and families, using a systematic and cost-effective approach to provide patient-centered care for a defined population. Another widely adopted health care practice affecting health care delivery is the emergence of patient- and family-centered care. In 2012, influential organizations, including the Institute of Medicine and The Joint Commission, adopted patient-centered care as a fundamental aim toward enhancing quality of care. An outgrowth of patient-centered care has been the emergence of family-centered care. The primary goal of patient- and family-centered care is to promote patients and families as active participants in care and decision making by treating them with respect and dignity; sharing complete and unbiased information in ways that are useful and affirming; and encouraging their participation and collaboration. A current nationwide initiative developed to improve care for older adults by focusing, in part, on what matters most to them is the Age-Friendly Health System, a program launched by the John A. Hartford Foundation and the Institute for Healthcare Improvement. Inherent in all of these efforts is the need for clinicians to be skilled in building and navigating relationships within families, as well as among families and health and social service providers. The adoption of integrated health care and patient- and family-centered care have created a shift in health care delivery from *serving* patients to *partnering* with patients and families.

Older adults today often have modified extended families as byproducts of single-parenthood, multiple marriages, and blended families. While families may span great geographical distances, they nevertheless are often able to mobilize support between and among older adults, their adult children, and other family members as needed. Those at greatest risk for isolation include Hispanics and men, who more often have fewer connections and report that they do not receive the social or emotional support that they need. When older adults do not have relatives nearby, they often form intimate social bonds with friends, neighbors, or others who are not related by blood or legal ties. Thus, an appropriate definition of the aging “family” must be flexibly inclusive.

Family intervention and family therapy are not one and the same. Family intervention is the process of bringing together two or more family members to address problems causing distress or impairing function. In late-life families, family intervention involves educating about the unique medical and psychosocial challenges of aging and their effects on the older adult and family. Family intervention also may involve facilitating communication among family members regarding age-related concerns, but it does not involve intentionally changing the way a family functions. In contrast, family therapy is an intervention that brings family members together to alter the family structure or functioning, thus facilitating adaptation to the challenges of aging. Family therapy is considered when assessment reveals problems in the family's pattern of organizing itself or interacting in response to age-related concerns and when family members explicitly agree to engage in a deliberate process of change. The decision to engage in family therapy typically reflects an acknowledgment that the family's natural adaptive capabilities are overwhelmed by the nature or intensity of challenges or perhaps that the family does not have the knowledge or skills necessary to deal with the situation at hand.

HISTORY

Family therapy with older adults has its roots in two distinct but overlapping theoretical approaches to the understanding of human behavior. First, family systems theory maintains that the family functions as a network of interconnected elements so that change in any one individual's status or behavior will affect that of others in the family. Accordingly, there is a natural tendency for the family system to attempt to maintain balance or homeostasis. Second, family life cycle theory views the family system as moving as a unit through time, encountering both predictable, normative transitions and unique, family-specific challenges. Life events such as a change in health or functional status, loss of a loved one, or remarriage of a parent challenge the family's natural homeostatic mechanisms and require a dynamic process of adaptation in order to reestablish equilibrium. The clinician working with aging families interprets the behavioral or emotional problems of an older adult as indications that the family system is struggling with transitions or crises associated with the life cycle stage. The therapist's task is to help family members organize or reorganize their resources and family roles in a manner consistent with the life cycle challenge they are confronting.

Although much of family developmental and systems theories have focused on the first half of the family life cycle, more recent depictions consider the structural and cultural heterogeneity of aging families. Life cycle theorists now consider late-life families as being comprised of blended families, families with single parents or grandparents raising children, and families of gay, lesbian, and transgendered individuals. Moreover, life cycle theories have started to include descriptions and differentiation of developmental tasks during the second half of life, including adjustment to retirement, attainment of grandparenthood, coping with loss and illness, and passing on family legacy. Recent theoretical accounts also address stages of family reorganization when its older adults decline due to progressive illnesses. Whether voluntary or involuntary, shifts in power within or between generations often create instability in families which often culminate in reorganization of roles and responsibilities. Such new models help therapists frame interventions that are appropriate to the tasks associated with a family's needs for adaptation.

THEORETICAL ISSUES

A variety of specific approaches to family intervention and therapy with older adults have their foundations in systemic and life cycle themes. The contextual approach emphasizes family ethics and seeks to help families balance the intergenerational "ledger" so that patterns of caregiving are congruent with the family's history and resources. The developmental-interactional model categorizes late-life family problems as situational, transitional, or chronic, integrating life cycle and structural-strategic approaches to help the family frame solutions to the problems in a manner consistent with their unique history and style. Recent approaches to grief therapy combine elements of family systemic, symbolic interactional, and narrative approaches to facilitate family storytelling, or "meaning making," thus strengthening family ties among the living as well as helping families redefine their relationships with those who have died. Cognitive behavioral therapy approaches, such as problem-solving therapy, are increasingly being used with older adult couples, caregivers, and families alike.

Strength-Vulnerability Model of Health and Illness

The strength-vulnerability model postulates that the mental health of individual family members is determined over time by the interaction of individual and family life cycle processes. Although not empirically validated, this clinical approach builds on systemic and life cycle theories, as well as research linking health outcomes to a variety of personality and psychosocial factors. Specifically, the model focuses on strengths and vulnerabilities at the level of the individual older adult (e.g., biologic health, personal themes, and individual life history) as these interact with health-enhancing and risk factors at the larger levels of family and other social systems (e.g., family events, themes, relational processes).

According to this theory, attachment and caregiving form the bond of affection between family members that must be present as a foundation for the development of other interpersonal processes in the family. Communication is the development of shared meaning between family members. This process includes sharing important themes and content, as well as constructive communication of affect. Shared problem solving is the ability to develop solutions to problems and carry them out cooperatively. Implicitly or explicitly, this process involves the delineation of family roles as they pertain to decision making and authority. Mutuality involves a long-term commitment to the relationship and the dynamic ability to reinvent the relationship over time in the face of life cycle transitions. These processes are thought to develop sequentially in families so that some element of attachment and caregiving is necessary for effective communication. Likewise, some effective communication is needed to successfully and collaboratively solve problems. Finally, effective problem solving is a prerequisite to the ability to reorganize and reinvent family relationships with mutual understanding through periods of crisis and life change. The epigenetic pattern is actually circular and dynamic, involving periodic reworking of various relational abilities rather than being a linear progression through a series of discrete stages. Thus, if late-life transitions and crises are met with successful mutuality, attachment and caregiving will take a new and strengthened form. With an unsatisfactory resolution of change, the bedrock of attachment and caregiving will be undermined.

TECHNIQUES

Family therapists are tasked with developing interventions that are congruent with the family's stages of relational and life cycle development. Thus, a detailed family assessment is the foundation upon which family interventions are developed.

Assessment

During the initial sessions, the therapist evaluates the family's unique strengths and vulnerabilities, including, at a minimum, the following key points, which can be easily remembered using the acronym LIFE: *Life cycle transitions* faced by the family and by individual family members—for example, raising children, job demands, financial stressors, retirement, grandparenthood, loss of a spouse, or onset of chronic or degenerative illness; *Individual roles* assumed by family members—for example, caregiver, breadwinner, matriarch or patriarch, and scapegoat; *Family relational processes*—that is, the quality of attachment or caregiving bonds, communication, and problem solving; and *Ethnic or cultural factors*, including spiritual beliefs and practices impacting the family's functioning.

Levels of Intervention

The family assessment determines whether the intervention will most appropriately be psychoeducational, with a focus on providing information pertinent to a problem or illness, or therapeutic, with a goal of changing the family's structure and quality of relational functioning. Several levels of therapy are considered, each reflecting work at successively deeper strata of family relational functioning. These levels serve as general guidelines for considering the type and intensity of treatment appropriate for a particular family.

Level One: Psychoeducational Intervention and Community Referral. Families who exhibit solid attachment bonds, open communication, and good problem-solving skills do not require ongoing therapy but, rather, benefit from brief education regarding a specific late-life issue. For example, when older adults develop new medical disorders, such as Alzheimer disease, families benefit from information specific to the disorder and its management as well as referral to community agencies and support groups. Families can usually organize themselves for effective problem solving regarding the care and support of the older adult. Despite their potential benefit, only a small percentage of families avail themselves of these resources. In the case of chronic, deteriorating illnesses, however, the family may benefit from brief, intermittent contact with the therapist at those transition points when the disease worsens or they acknowledge readiness to face the next challenge. Such families typically do not require regular, ongoing professional support.

Level Two: Restoring Effective Problem Solving. Families often have solid, positive bonds of attachment with good communication skills and histories of effective problem solving, yet they temporarily are overwhelmed by the intensity and unfamiliarity of a late-life problem, such as severe illness, functional decline, or the approaching death of a previously capable family member. In these cases, effective problem solving may be temporarily stymied because of paralyzing grief reactions or sudden disequilibrium in the previously stable family structure. The family may need to reorganize lines of authority and decision making to bring them into accordance with the demands of a later-life transition. For example, families typically go through a natural transition during which the younger generation gradually assumes more of a role in an older adult’s affairs. This reciprocal process requires not only commitment and involvement on the part of the younger family members, but also willingness on the part of the older adults to allow this transition to occur without becoming overly demanding, resentful, or needy. Therapists recognize that the history of parental authority and caregiving continues to influence intergenerational dynamics well beyond the actual parenting years.

One common example of family difficulty at this level is when the matriarch or patriarch experiences a decline in health or cognitive capacity, throwing the surviving family members into a state of disequilibrium. The spouse and adult children in this situation are unaccustomed to assuming decision-making authority and typically need assistance in reorganizing family roles. After receiving clear and direct information about the status of the older family member, reflecting on the meaning that they as well as the older adult assign to this situation, and having an opportunity to grieve these changes, adaptive adult children at this level of relational development can often mobilize their resources to solve family problems.

Level Three: Improving Patterns of Communication. Some families who care deeply about each other and have solid bonds of attachment communicate little about serious problems. Some may have a history of arguments and resentment stemming from past crises. Some family rules prohibit the exchange of certain emotions or information. Others have been able to function with little overt communication in earlier stages of the life cycle, because they were not confronted with major loss or illness. When faced with challenges of aging, they have difficulty developing a common understanding of the problem and thus cannot work collaboratively to find and carry out solutions. Families at this level benefit from active modeling of open, positive communication. Basic communication techniques, such as active or reflective listening, can greatly benefit these families, as can clear and consistent modeling of how to talk openly but respectfully about the challenges before them.

QUESTIONS FOR THE PATIENT

Communication and Problem Solving. How is information about important events shared in your family? Who typically participates in making decisions? Is one person responsible for making a final decision? Is anyone excluded from important family decisions? How do family members resolve conflict when it occurs?

Family Roles and Expectations. What role does the older adult play in the family? Who are the individuals responsible for caring for the older adult? Who, if anyone, is primarily responsible for providing care? Has the older adult appointed a health care proxy? Does anyone live with an older adult? How much time in person, on the phone, or otherwise is each person able to contribute to the older adult's care? Who, if anyone, is responsible for helping older adults with their activities of daily living (ADLs)/instrumental activities of daily living (IADLs)? Are you aware of local resources that may be able to help the older adult? What have your experiences with the medical system been like in the past? What are your expectations for the older adult's care? What do you know about their end-of-life wishes if the older adult is unable to express them?

Level Four: Improving Bonds of Attachment and Caregiving. Families with negative, weak, or ambivalent bonds of attachment typically have a history of conflict, marginal functioning, or insufficient nurturance between family members. In some cases, adult children who were abused or neglected as children struggle to take care of the one they perceive as a former abuser. Such families are at risk for elder abuse, and some may require referral to professional caregivers or long-term care facilities. In many instances, however, techniques, including family life review or contextual family therapy interventions, can be used to induce empathy for the older adult's position, thereby decreasing negativity, increasing feelings of safety, and making it possible to forge more positive bonds of affection between the generations.

Family life review is the process of bringing multiple generations together to weave a coherent intergenerational family story. Typically, this process is facilitated by constructing a family genogram, which is a graphic, multigenerational picture (family tree) that encapsulates family history, structure, and relationships over time. The therapist encourages the family to begin with the family of origin stories of those in the oldest generation, gradually working down through the themes and experiences of subsequent generations. Learning about their upbringing, cultural context, and early family circumstances often gives younger family members a different, more empathic perspective. Some children or grandchildren then can establish more mature "adult-to-adult" relationships with them, rather than continuing to relate solely from the vantage point of children whose needs for nurturance and care were unmet.

CLINICAL ISSUES

Clinical Applications Across Health Care Settings

The strength-vulnerability model of health and illness is an appropriate framework to use for conceptualizing and delivering patient- and family-centered care to families with older adults in multiple therapeutic contexts. Applications of this model across health care settings in which mental health clinicians practice are described here.

Outpatient Psychiatry

The majority of older adults with behavioral health symptoms receive mental health services through some type of outpatient setting. Outpatient services are provided in a variety of settings, including private practices, community mental health centers, hospital-based clinics, or programs embedded within nonprofit organizations. Geriatric specialty clinics offer a range of services, including psychotherapy, social work, and case management, which can help to support the work of mental health clinicians. Many older adults gain access to psychiatry through consultation models that exist in some integrated care settings. In this model, primary care physicians and psychiatrists are colocated, and physicians are able to consult with a psychiatrist onsite, or they can refer patients within the practice to see the psychiatrist for an evaluation and treatment recommendations. Occasionally, independent living communities may help organize or provide access to psychiatric services for residents. In any type of outpatient setting, psychiatry can play an important role in assessing and identifying family, social, contextual, environmental, and physiologic factors that can support or exacerbate the patient’s mental health issues.

In any outpatient setting, clinicians should focus on: conducting a comprehensive evaluation of mood, cognition, and functional capacity; paying particular attention to issues that may impair the patient’s ability to perform ADLs or IADLs; gathering information about the patient’s social support network including family, friends, and community resources; provide psychoeducation on the identified mental health issue and how such issues may impact the older adult’s ability to care for themselves; identify potential barriers to treatment including financial resources, transportation, housing, and access to health care; and gather information about the older adult’s physical health and collaborate with the patient’s physician or other members of the health care team. As with any population, there are challenges involved when treating older adults in an outpatient setting. The most salient issue is the constraint of time. Completing a comprehensive evaluation that includes an assessment of family history and function within a 30-minute intake poses a formidable challenge. The medical comorbidities with which many older adults present often add to the complexity of the assessment and treatment planning processes.

Mrs. D was an 89-year-old, Caucasian, widowed female of Italian descent, referred to an older adult outpatient specialty clinic for symptoms of depression and anxiety. She reported a long history of depression and anxiety and estimated that her current symptoms began six months ago when her family decided she could no longer live alone after she gave herself food poisoning. This incident made the family realize how difficult it was for the patient to complete her ADLs due to arthritic pain, decreased mobility and stamina, and an unstable gait.

The patient's family consisted of her four daughters, three of whom lived locally. Each of the daughters contacted the psychiatrist separately to provide information about the patient's mental and physical health. During the intake, which two of the daughters attended, the patient was irritable and expressed her concerns about her eldest daughter's ability to care for her properly. The daughters argued openly about the patient's needs until one daughter left. As the psychiatrist gathered more information it was clear that this family had deep emotional wounds that negatively impacted all of their relationships. The daughters felt "burnt out" and stated that they cared for their mother "out of obligation, not love." Two of the daughters reported having been physically and emotionally abused by the patient.

Based on the strength-vulnerability model, this family is best described as having negative, weak, and ambivalent bonds of attachment (Level 4). Communication and, therefore, problem solving were ineffective. They were unable to make decisions as a group and each wanted to be in a position of authority. Working on the hypothesis that the family needed to work on a family life review to encourage more positive family communication, the psychiatrist referred the patient and her family for family therapy. Meanwhile, the psychiatrist took active steps to create opportunities for positive communication and problem solving. The patient and her daughters were asked to create a list of activities with which the patient would require assistance, and the degree to which she would need assistance. The psychiatrist provided guidance regarding the ADLs and IADLs to consider. Family members were encouraged to review the list and to decide which type of help each was able to provide. An information system including the patient's needs, medical status, and psychological status was devised and shared with all family members on a biweekly basis. The combination of these interventions helped the family learn how to communicate and problem solve more effectively while working with a counselor to improve the overall quality of the attachment bonds.

Inpatient Medicine Service

Over 40% of older patients seen in the emergency department are admitted to the hospital. These patients not only face physiologic problems but also suffer psychosocially, which is even more likely if diagnosed with chronic medical conditions. Estimates suggest that nearly 75% of older adults who enter the hospital functionally independent will leave requiring support with ADLs and many of them will return to the emergency department for repeat admissions. The need for adequate discharge planning and improved long-term disposition is dependent upon assessment of their holistic needs, including functional status, cognitive capacity, mood, and social support. Though hospitalists are often relied upon to conduct such assessments, their time is limited and many have not been trained extensively in the unique needs of geriatric patients or in psychiatric interview techniques. Psychiatry can play a supportive role in acute medical settings by offering consultation to the medical team and to patients and their families, thereby ensuring a safe transition from hospital to home or an alternate placement. In addition to this consultative role, psychiatrists offer valuable help to patients coping with biopsychosocial changes and transitions.

To do this work most effectively, it is important to utilize a patient- and family-centered approach that recognizes the older adult as an activated patient who wishes to participate in the development of treatment and discharge plans. Psychiatry supports the patient and family by recognizing the patient's preferences, helping them to make informed decisions, listening to their concerns and questions, providing them with education about the mind–body connection in the context of acute medical illness, and informing them of available resources. In an acute medical setting, psychiatrists can anticipate conducting a comprehensive assessment of mood, cognition, and functional capacity, paying attention to the illness story and changes that have occurred as a result of hospitalization; evaluating the patient’s social supports and availability of caregiving; providing consultation on medication management of symptoms, including labile mood, poor appetite, and insomnia and identifying resources to address other patient and family needs.

Conducting a comprehensive assessment with an older adult in the medical setting is similar to the health assessments psychiatrists conduct with adult patients, but requires additional detailed attention to the functional capacity of the patient as outlined in Recommended Questions for the Patient.

Recommended Questions for the Patient

PERSONAL/SOCIAL FUNCTIONING. Where do you currently live? Where will you go after discharge? Where do you hope to go? Is there anything you (patient/family) worry about regarding discharge? Is there anything about the physical layout of your home that concerns you? Prior to your hospitalization, did you require help with ADLs/IADLs (specify each)? How does your illness limit your ability to engage in routine activities? Who has helped you in the past when you’ve become ill? Who will help when you go home? Are you the caregiver to a loved one (and, if yes, who will help you care for them now)? Have you experienced any other transitions or changes recently? Do you have transportation needs? Do you have access to and/or are you able to use a computer or your phone to connect with telehealth services?

EXPERIENCE OF PAST AND PRESENT ILLNESS. Have you recently been hospitalized? What caused you to become sick enough to return to the hospital? Why were you readmitted? What could have been done differently before you left the hospital last time? What has gotten in the way of you taking care of yourself? What needs to happen while you are here/before you are discharged so that you remain healthy? How well do you follow medical recommendations? What gets in the way of taking care of yourself?

Psychiatry is also often tasked with providing psychoeducation to the treatment team, patient, and family. It is well established that a patient's physical functioning, number of comorbid medical conditions, and history of depressive episodes can result in comorbid mood and anxiety disorders in late life. When patients and families receive adequate information it serves to validate the experienced change in mood, normalizes the experience of lability, acknowledges the mind–body connection, explains the anticipated recovery from changes in personality and mood, and provides patients and families with hope for improved quality of life following illness.

Clinicians can also provide supportive listening, paying attention to the patient and family experience of illness, the subsequent loss of function and independence, and fears of death. Through dialogue the consultant can identify whether the development of improved coping skills and implementation of relaxation techniques at the bedside would be beneficial. The family draws support from acknowledgment that they are also impacted by the hospitalization. Caregivers are described as "invisible patients," as their needs are often not assessed or addressed in health care. They report that hospitalization causes disruptions to their everyday routines, distress, and difficulty meeting their own personal needs. Caregivers are often older themselves and require assistance with their own care in order to best serve the one needing care. Consideration for caregiver burden, burnout, risk of elder abuse, and determination of the family's capacity to provide care are essential components of a discharge plan.

Mr. B was a 76-year-old man of Jamaican descent. He was hospitalized following a stroke, suffering from dysarthric speech and changes in mobility. His wife of 47 years reported that their youngest daughter, who lived with them, had been admitted to the same hospital that week for management of congestive heart failure. Their six other children expressed concern about their mother's capacity to care for both her husband and daughter postdischarge.

The patient was very tearful throughout his hospitalization. Mrs. B was increasingly distressed by the patient's continued dysarthric speech, which was exacerbated by his tearfulness. She wanted to know what was "wrong" with him, articulating that he had never suffered from such a labile mood and expressing uncertainty about how best to support him. Though the Medicine service recognized the patient's symptoms as common following a stroke, they did not have adequate time to address Mr. or Mrs. B's distress or understand it fully. The Medicine service requested a consult from Psychiatry to better meet the patient's and family's needs.

A psychiatry resident met Mr. B, asking questions about his medical experience. Mr. B expressed regret that he had not been a more spiritual man, noting that his survival would not have been possible without religious intervention. His wife frequently nodded her head, saying little except to occasionally repeat what Mr. B had said when concerned the resident had not understood him. Amidst the couple's sobbing, the resident asked, "How much do you understand about why you feel this way?" Again, Mr. B noted religious reasoning. The resident was able to validate the family's tie to spirituality while simultaneously providing psychoeducation about the mind–body connection and the impact of stroke on mood.

Mr. B heard this information and asked if he would always be a burden on his wife in the aftermath of this acute event. By then one of their sons had arrived and inquired about what types of care his father would likely require. The resident facilitated a conversation focused on their collective observations of limitations, eventually engaging them in a process of problem solving, and exploration of their hopes for his future recovery. At the close of the visit, the resident ensured that Mr. and Mrs. B were given resources regarding therapeutic services, including support groups and respite services for Mrs. B. The referring provider was given recommendations for medication management of symptoms, which had already been shared with the family.

In this case, Psychiatry identified what the family understood about the patient's health and mood and provided education on both topics. The family learned new coping strategies and received referrals to appropriate services both in the hospital and out. Through these interventions, Psychiatry supported Medicine and completed tasks that would have otherwise been the responsibility of the hospitalist and may not have been addressed as thoroughly. Psychiatry's efforts enabled the patient to become active in his own recovery and his family to become participants in discharge planning. Psychiatry helped Medicine transition the patient safely home, reduced the likelihood of readmission, and increased patient–family satisfaction.

Long-Term Care

While the number and variety of housing and residential care options for older adults have evolved to include retirement communities and assisted living facilities, this proliferation has not obviated the need for traditional long-term care facilities. Among the 1.4 million long-term care residents in the United States, 80% have one or more psychiatric diagnoses. Since the 1990s, nursing homes have absorbed increasing numbers of older adults with severe mental illness and serious behavioral problems and, hence, the prevalence of mental illnesses in long-term care populations has increased. Despite the high prevalence of mental disorders in this population, less than one-fifth of nursing home residents with psychiatric diagnoses receive mental health treatment. The recent proliferation of telemedicine and telementoring has increased access to psychiatric and other mental health services for this population. In facilities that do provide mental health services, psychiatrists typically work in one of two models—as consultants or as members of multidisciplinary teams. At present, the consultant model is used more widely, and psychiatrists are most often consulted to address behavioral manifestations associated with adjustment issues, disease and infirmity, and dementia. A survey of geriatric psychiatrists in the United States found that the most common treatment recommendations in long-term care included psychiatric medications followed by changes in the general medical regimen, staff support interventions, medical diagnostic testing, behavioral interventions, staff training, individual or group psychotherapy, and family psychotherapy. These data suggest that geriatric psychiatrists tend not to rely solely on pharmacotherapy, but instead recommend a more diverse range of interventions in accordance with evidence-based practices. In some facilities psychiatry is the discipline responsible for intervening with family issues, especially conflicts between staff and families; in others, disciplines such as social workers, marriage and family therapists, and mental health counselors are available to serve in this role. Since 40% of long-term care residents generally reside with family prior to placement, family involvement after admission is commonplace. Additionally, long-term care placement is a significant life cycle event for older adult families that often triggers the manifestation of a spectrum of family dynamics.

Maria was a 51-year-old Latinx woman of Mexican-American descent and daughter of her 83-year-old father and nursing home resident. Her father, Mr. F., lived in the facility for 6 months following a discharge from the hospital after a car accident that left him with limited function on his left side. Prior to the accident, Mr. F. had lived in the in-law apartment attached to the home where Maria lives with her husband and son and was driving independently within a 5-mile radius of the home. Mr. F. moved into the apartment reluctantly 2 years ago following his wife's death. He did not want to leave his home, his church, or the diner in his neighborhood where he met "the guys" for coffee a couple of times a week. Mr. F. had been well known in the community "in the day" because he owned and operated a popular barbershop for many years. Maria ultimately persuaded him to move so that she could help him more and he would be near family. He continued to grieve the loss of his wife and spent most of his time watching baseball games and old movies and helping his daughter with her garden. His health was stable as Maria supported his medication management of hypertension and high cholesterol, but Maria began to become concerned when he started to lose his appetite, appeared noticeably thinner, and started to become withdrawn. Shortly thereafter Mr. F. was involved in a car accident, was hospitalized and then went to a rehab facility for 2 months. He recovered some function while in rehab but not enough to be able to return home safely.

Mr. F. did not resist moving into a nursing home as he believed he would never be "home" again anyway. After a month or so of adjusting to his recent physical limitations, he began to eat dinner in the dining room. Being an "haz lo que tengas que hacer" ("do what you gotta do") kind of man, the nursing staff and therapists from rehab liked him. He was quiet but friendly and very engaged in therapy. He met other residents on the unit, began to watch baseball with others, and occasionally played cards or checkers when invited. The nursing home staff was pleased with his adjustment to the facility.

Maria struggled with the adjustment. She didn’t understand why he was not getting stronger. She repeatedly advocated for rehab services for her father at least 4 days per week. The social worker overheard Maria commenting to another visitor that “they were giving up on him.” The social worker did her best to validate Maria’s feelings while explaining that her father was receiving as much rehab as he could tolerate. Maria felt better after speaking with the social worker and continued to advocate for more rehab now that “the social worker said that she agreed with her.”

Maria was also displeased with the food her father was served. She was concerned about his cholesterol and displeased that they were not serving him traditional meals like she would have prepared for him. She questioned his dietician about the menu choices he recommended. She advocated for fresh fruits and vegetables and talked about how important gardening had always been to him. She questioned the physician about alternative medicine and use of herbal supplements to reduce his risk of stroke. She decided to back down on that fight and took matters into her own hands by supplementing his diet with lots of green, leafy vegetables from her garden.

At the time of the 6-month family meeting the nurse manager opened by reporting that Mr. F. was doing really well, that he had developed a few friendships on the unit, that his mood had improved, and then declared, “We are all fond of him and I think he is happy here.” Maria became agitated and started to voice other complaints. She asked if his bed could be moved because the sun glare obstructed his view of the TV. The nurse responded politely that she had discussed this with Mr. F. and he said that he preferred it the way it is because he did not wish to give up his view of the outdoors. This confused Maria because she felt that her father no longer needed her to advocate for him.

OPPORTUNITIES FOR INTERVENTION WITH MARIA

Provide Education

- To better understand the physical therapy treatment plan

- To assure her that the meals meet nutritional standards (provide her a sample meal with nutritional breakdown)
- To address her questions about herbal supplements

Improve Communication

- Develop and deliver “team” messages uniformly as a way to reinforce that all team members understand and can explain the plan of care
- Respect the delicate balance between allowing Maria to be “the expert” about her father and assuring her that the staff have relationships with him and are getting to know him also

Use “teach back” or “ask-tell–ask” around topics like his nutrition and therapy

Create Collaborative Team Meeting Formats

- Avoid “Team shares information–Family shares concerns” hierarchy
- Allow Maria an opportunity to add items to the agenda
- Come prepared with answers to questions Maria typically asks

Provide Maria With Opportunities to Be a Partner In Care

- Assure Maria that she maintains a special role in her father’s life
- Encourage her to provide care in alternate ways (e.g., bring in meals her father used to enjoy, buy toiletries, pick out his clothes, decorate the room for holidays)
- Show Maria how her contributions have shaped her father’s plan of care or the way in which care is delivered

Palliative Care Service

Current research demonstrates that patients with advanced, life-threatening illness benefit from care that addresses their psychosocial needs, including loss, grief, quality of life, and suicidal ideation. Though rates of depression are as high as 58% in this population, psychiatric needs often go unrecognized or untreated. Lack of treatment is partially the result of limited availability of psychiatry consultation and clinicians’ perceptions of their insufficient competence to meet the needs of this population.

Patients and families facing terminal illness are often seeking improved quality of life and maximum comfort. Psychiatry can support a patient–family-centered experience that enhances shared decision making about end-of-life decisions, honors patient preferences regarding place of care, and recognizes caregiver fatigue. Families require continued education not only about the disease process, but also about palliative care treatment options. Though palliative care teams are equipped to have these conversations, the dialogue can sometimes come as a surprise to families, and clinicians in psychiatry and medicine can continuously assess the family's readiness for palliative interventions. Additionally, psychiatry is uniquely positioned to help address family conflicts that may arise with respect to different perceptions around end-of-life options and to help families transition from their in-hospital experience to postloss bereavement experience and provide grief counseling.

Ms. P was a 71-year-old Asian American when she presented to the hospital. Widowed for over a decade and suffering from amotivation and anhedonia, she had stopped attending to her own health. She had once enjoyed preparing fresh meals, but she had discontinued this, finding it difficult to cook for herself. She eventually moved in with her daughter and grandson. Her daughter worked two jobs and was a single parent; she found cooking to be taxing and barely found the time to buy groceries. Often, her daughter would return home with fast food or takeout. Ms. P began eating increasingly large amounts of fried foods, sweet snacks, and carbohydrates despite warnings that they were bad for her health. Ms. P suffered from uncontrolled diabetes for several years, episodes of gastroparesis, and chronic kidney disease. Upon hospital admission, she was diagnosed with end-stage kidney disease, and there was a discussion of starting dialysis; alternatively, they noted that without dialysis, Ms. P would likely die. In the interim, palliative care was monitoring the patient's medication management of pain and had started a patient controlled analgesia (PCA). Ms. P was quiet whenever members of the medical team entered her room, as was her daughter; however, her daughter shared her preferences quite openly outside of the hospital room, adamant that her mother should receive dialysis.

Psychiatry was consulted to enhance communication between mother and daughter and to help the team better assess Ms. P's interest in dialysis. Knowing that Ms. P and her daughter remained silent in front of the team, the psychiatrist recommended that a family meeting take place to include psychiatry, medicine, and palliative care. In doing so, the team was able to model (and subsequently encourage) effective communication. The psychiatrist demonstrated the use of open-ended questions, facilitating a conversation that prompted participation from all team members, Ms. P, and her daughter. Ms. P eventually articulated that she was not interested in living a life that required intensive medical care and wanted to be with her husband. Her daughter was tearful but able to verbally accept her mother's choice. Palliative care was then able to provide insight into expectations of the disease process. With this information, the psychiatrist then invited Ms. P to consider other personal preferences and wishes. The psychiatrist supported further dialogue between mother and daughter. At the conclusion of the conversation, Ms. P expressed gratitude that she might die with dignity and was happy to know that she could live her remaining days at home with supportive services and her daughter and grandson nearby.

When Ms. P died a few weeks later, her daughter mailed a letter of thanks to the palliative care team and psychiatrist. She explained the value she perceived in preparing her for the future, avoiding treatment her mother viewed as futile, and improving their communication. She lamented the fact that a family meeting had not occurred sooner in her mother's care, wondering if the last decade of their relationship might have been different. She shared that she previously found her mother's quiet, stoic nature to be one of avoidance rather than resilience and that the family meeting helped her view her mother more positively. She concluded that in giving her mother dignity at the end of her life, the psychiatrist had reduced the family's experience of guilt and grief.

Indications

Family therapy is most commonly used in older adults when the presenting problem is one in which the patient's distress or capacity for problem solving is rooted in a relational context. Family therapy is also indicated when the clinical situation at hand could likely be resolved with brief, solution-focused interventions that are designed to address specific therapeutic goals. Most serious mental health problems of older adults are more successfully addressed when family members are involved in the treatment. Although caution and clinical judgment must be exercised as to the type and timing of family interventions, substantial evidence indicates that psychiatric problems worsen or resolve less completely when conflictual or chaotic family environments are not addressed in the treatment. A major developmental task that families face throughout the life course is finding a balance between supporting individual members to be competent and independent while at the same time remaining connected to and part of the family as a whole. Families are often confronted with this challenge as they encounter life cycle transitions, such as the birth of a child, marriage, retirement, an aging parent, or the loss of a loved one.

Limitations

Older adults are most likely to share psychiatric complaints with their primary care physician, and even then, their report of symptoms is often in the form of somatic concerns. It is, therefore, quite difficult to reach aging families and implement therapeutic interventions without flexibility, integration, and collaboration. Though family therapy is most commonly provided by marriage and family therapists, social workers, and clinical psychologists with specialty training in family psychology, making referrals to family therapists is not always straightforward as requirements for licensure differ by state, and the availability of clinicians is often limited. Reimbursement for family therapy services varies by payer, type of mental health provider, setting, and state. Psychiatrists face similar barriers in the implementation of this work. In addition to complexities regarding access, a major limitation is a pervasive attitude among clinicians that the potential for change in late-life families is negligible.

Contraindications

Family therapy with older adult clients is contraindicated if a patient’s insight or capacity to change is impeded by cognitive impairment, paranoia, or active substance abuse. Conjoint family therapy is typically contraindicated in situations that involve extreme anger on the part of any involved family member or situations in which an assessment has indicated that therapy may increase the likelihood of abuse or other harm. Additionally, it is important to be mindful of the setting in which therapeutic interventions are being applied. For example, in an inpatient medical setting where an episode of care is brief, longstanding transgressions and disagreements are unlikely to be resolved. It is contraindicated to initiate exploration of such work without a plan for follow-up after a hospital discharge or if a referral for outpatient care is not yet established.

Goals of Treatment

Family therapy interventions are used to reduce clinical symptoms that impact the individual and the family system. As noted, such interventions can be utilized to improve function by effectively changing the family's structure, patterns of interaction, problem-solving abilities, capacity for empathy and understanding, communication skills, and sometimes the meaning that a family ascribes to a certain event or situation. In addition, addressing a family's response to medical illness, death and dying, and the aging process is critical. The cases presented in this chapter provide examples of the multiple layers of family dynamics that accompany the transitions older adult families often confront.

ETHICAL ISSUES

Providing treatment to older adult families at critical times of transition is embedded within layers of complexities. The settings in which the care is provided often introduce additional ethical considerations that may interfere with the provision of optimal patient- and family-centered care. Some ethical issues that clinicians are likely to encounter are discussed here.

End-of-Life Care

Patients, families, and care teams encounter difficult life-and-death decisions, which encompass medical decision making, comfort management, and emotional and spiritual well-being. Though formal documents and legal statutes in the United States emphasize the importance of individual autonomy and independent decision making, research suggests that older adults want family members involved in end-of-life decisions. Those suffering from a chronic and terminal illness may experience cognitive changes that impede their ability to make decisions independently and benefit from familial support. Despite this fact, the research indicates that older adults and their children often disagree on key factors about the end of life. Clinicians can support patients and families by facilitating open dialogue about end-of-life preferences, especially before patients lose the capacity to share their wishes.

Clinicians further assist families in making decisions collectively by addressing the family's relational dynamics and life cycle considerations. Such conversations may occur rapidly in acute medical situations that require immediate decision making or more slowly at the outset of a chronic illness. In some instances, family members do not support their loved one's preferences regarding the use of life-sustaining measures. Such scenarios may require the development or enhancement of families' natural problem-solving abilities, modeling of adaptive communication, or grief work. In such situations, clinicians are challenged to balance the patient's needs while acknowledging the larger family experience and preferences.

Capacity and Decision Making

Decisional capacity refers to a person's ability to make their own decisions in a variety of contexts, including those related to health care. The issue of consent or refusal of treatment regularly arises when working with older adults, particularly in the inpatient setting. Some have argued that the medical conditions older adults face can contribute to feelings of helplessness, thus rendering them more open to the undue influence of others. Though clinical methods are available to assist with assessing capacity, this determination is made amidst multiple contextual factors. Psychiatrists are thoroughly trained to assess capacity, but it is important to not only determine the patient's ability to understand and reason but also to ascertain a sense of the patient's values prior to the onset of illness. Such evaluations and subsequent information sharing with the patient attempt to honor patient autonomy and appropriate family inclusion. It is important to understand shared versus conflicting values held by the patient and family and the risk of conflict and potential coercion when values and wishes do not align. Further, if a patient is deemed "incompetent," clinicians also consider whether the surrogate decision maker is acting with the patient's values and best interests in mind.

Cultural Considerations

Each generation of older adults gives rise to a unique set of values, beliefs, and family dynamics. These factors impact how older adults and families interact with the healthcare system and can play an important role in treatment decisions and outcomes. As increasing numbers of Baby Boomers reach the age of 65, it is important for clinicians to recognize the different social and demographic characteristics that define this unique cohort. Individuals in this group tend to be more educated than their predecessors, have more access to health care information, hold higher and more specific expectations of the health care system, and have a tendency to challenge authority. Baby Boomers are more open to receiving mental health services than previous generations. Despite these overarching themes, Baby Boomers comprise a heterogeneous group. Thus careful evaluation and treatment planning are fundamental to providing care that honors and respects the cultural and generational considerations of patients and their families.

In addition to generational considerations, U.S. Census data from 2019 indicate that nearly one-quarter of the over-65 population is comprised of members of racial or ethnic minority populations. There is a need for practice guidelines addressing the feasibility and acceptability of family therapy approaches across these populations.

Elder Abuse, Neglect, and Exploitation

Elder mistreatment, including abuse and neglect, is a public health problem of significant concern. Estimates indicate that one in 10 older adults in the United States alone are victims of abuse or neglect. Community-based studies indicate that one in six individuals over the age of 65 experienced abuse in the prior year, and over half lived with the perpetrator of abuse. Older adults who are cognitively or functionally impaired are at higher risk for abuse, and some who experience severe disability are unable to report maltreatment. Elder abuse is one of the strongest predictors of long-term care placement. Advanced age, female gender, being in a spousal relationship, African-American race, living with others, physical and economic dependency on caregivers, isolation, intergenerational conflict, substance misuse, and mental illness have all been associated with increased risk of abuse. Additionally, published reports suggest that family members who perceive a high level of caregiver burden, have a history of their own mental health needs, or have a history of substance misuse, may be more likely to engage in abusive behaviors, particularly when seeking compliance from the one needing care.

Clinicians must be alert to the needs of patients and their caregivers. It is helpful to look for signs of abuse or neglect, including repeated, unexplained bruises or injuries, malnutrition, fearfulness around particular family members, or, in the case of those who are being financially exploited, a discrepancy between their ability to pay for personal commodities and their reported level of income. Clinicians can reduce the risk of abuse by conducting thorough assessments of caregiver burnout and by attending to the caregivers' experience of an aging relative. Family therapy is recommended to mediate conflicts that might give rise to mistreatment, but more research is needed to understand the treatment techniques that are most effective.

The majority of states now have mandatory reporting laws for elder abuse, as well as state or county-funded protective service programs for older adults who have been abused or are at risk. Practitioners are responsible for knowing state reporting laws. Although estimates indicate that one-third of actual cases of child abuse are reported, some studies estimate that only 1 in 24 cases of elder abuse is reported. Older adults are often reluctant to disclose the problem for fear of retaliation from an abusive family member. Further limitations to the frequency of reporting are thought to be related to the professional’s fear of offending the patient or family or placing the patient at greater risk for harm.

RESEARCH AND EVALUATION

Despite the growing knowledge base on family systems and life cycle approaches to the treatment of older adults, evidence of the utility of these methods remains sparse. A recent examination of therapeutic outcomes across the lifespan found that while older adults differed from younger adults on key therapeutic issues, age was not significantly related to the amount of change observed over time across all individual and relational measures. In contrast, there are numerous studies on the effectiveness of psychoeducational interventions and cognitive-behavioral psychotherapies that are focused on family caregivers of older adults. Many of the most promising advances in the field are rooted in ecosystemic interventions that are delivered outside of the traditional healthcare practitioner’s office. These interventions have been demonstrated to have positive effects on such outcomes as caregiver burden, depression, subjective well-being, knowledge and skills, and physical health, as well as on the level of symptoms expressed by the older adult receiving care. At present, research on caregiving, social isolation, elder abuse, and end-of-life care are the most fertile areas for investigations of family therapy approaches.

FURTHER READINGS

Achterberg W, Pot AM, Kerkstra K, et al. Prevalence of psychiatric disorders among older adults in long-term care homes: a systematic review. *Int J Geriatr Psychiatry.* 2006;21:1156–1162.

Binstock RH, George LK. *Handbook of Aging and the Social Sciences*. Academic Press; 2011.

*Capezuti EA, Malone ML, Khan A, Bauman SL. Assessing needs of family caregivers. In: Capezuti EA, Malone ML, Gardner, et al., eds. *The Encyclopedia of Elder Care: The Comprehensive Resource on Geriatric Health and Social Care*. 4th ed. Springer Publishing; 2017:70–73.

Carr D, Utz RL. Families in later life: a decade in review. *J Marriage Fam*. 2020;82:346–363.

Clissett P, Porock D, Harwood RH, Gladman JR. Experiences of family carers of older people with mental health problems in the acute general hospital: a qualitative study. *J Adv Nurs*. 2013;69:2707–2716.

Deardorff WJ, Grossberg GT. Psychiatric consultations in long term care: an evidence-based practical guide. *Curr Psychiatr*. 2016;15(11):39–48.

Devine M, DeCaporale-Ryan L, Lim M, Berenyi J. Psychological issues in medically underserved patients. *Prim Care*. 2017;44(1):99–112.

Fairman N, Irwin SA. Palliative care psychiatry: update on emerging dimension of psychiatric practice. *Curr Psychiatry Rep.* 2013;15(7):374.

Fisher KA, Seow H, Brazil K, et al. Prevalence and risk factors of depressive symptoms in a Canadian palliative home care population: a cross-sectional study. *BMC Palliat Care*. 2014;13(1):10.

Fulmer T, Mate KS, Berman A. The age-friendly health system imperative. *J Am Geriatr Soc*. 2018;66(1):22–24.

Gilligan M, Stocker CM, Jewsbury Conger K. Sibling relationships in adulthood: research findings and new frontiers. *J Fam Theory Rev.* 2020;12(3):305–320.

Gossner J, Anihia H, McPhee DP. Contextual therapy for couples who are primary caregivers for a parent(s) with Alzheimer's Disease. *Contemp Fam Ther.* 2021;43:248–258.

Hantke N, Lajoy M, Gould CE, et al. Patient satisfaction with geriatric psychiatry services via video teleconference. *Am J Geriatr Psychiatry*. 2020;28(4):491–494.

King DA, Quill T. Working with families in palliative care: one size does not fit all. *J Palliat Med*. 2006;9:704–715.

*Kissane DW, Bloch S. *Family Focused Grief Therapy: A Model of Family-Centered Care during Palliative Care and Bereavement*. Open University Press; 2002.

Lachs MS, Pillemer K. Elder abuse. *Lancet*. 2004;364:1263–1272.

*Lin M, Giles H. The dark side of family communication: a communication model of elder abuse and neglect. *Int Psychogeriatr*. 2013;25:1275–1290.

McMahan RD, Barnes DE, Ritchie CS, et al. Anxious, depressed, and planning for the future: advance care planning in diverse older adults. *J Am Geriatr Soc*. 2020;68(11):2638–2642.

*Podgorski C. Neurocognitive disorders: systemic functionality and interconnectedness. In: Russo J, Coker JK, King JH, eds. *DSM-5 and Family Systems: An Applied Approach*. Springer Publishing; 2017:443–472.

Pruchno R. Not your mother's old age: baby boomers at age 65. *Gerontologist*. 2012;52(2):149–152.

*Qualls SH, Williams AA. *Caregiver Family Therapy: Empowering Families to Meet the Challenges of Aging*. American Psychological Association; 2013.

Rolland JS. Helping families with anticipatory loss. In: Walsh F, McGoldrick M, eds. *Living Beyond Loss: Death in the Family.* 2nd ed. WW Norton; 2004.

Sabey AK, Jensen J, Major S, Zinbarg R, Pinsof W. Are older adults unique? Examining presenting issues and changes in therapy across the life span. *J Appl Gerontol.* 2020;39(3):250–258.

Shibusawa T, Jones T. Aging adults needing care: interventions to respond to family needs and stress. In: Wampler KS, Rastogi M, Singh R, eds. *The Handbook of Systemic Family Therapy*. John Wiley & Sons Ltd; 2020.

*Shields CG, King DA, Wynne LC. Interventions with later-life families. In: Mikesell RH, Lusterman DD, McDaniel SH, eds. *Integrating Family Therapy: Handbook of Family Psychology and Systems Theory*. American Psychological Association; 1995.

Sorensen S, King DA, Pinquart M. Care of the caregiver: individual and family interventions. In: Loboprabhu SM, Molinari VA, Lomax JW, eds. *Supporting the Caregiver in Dementia: A Guide for Health Care Professionals*. Johns Hopkins University Press; 2006:168.

The Joint Commission. *Advancing Effective Communication, Cultural Competence, and Patient- and Family-Centered Care. A Roadmap for Hospitals*. The Joint Commission; 2010.

Van Orden KA, Bower E, Lutz J, et al. Strategies to promote social connections among older adults during ‘social distancing’ restrictions. *Am J Geriatr Psychiatry*. 2020;10:16.

*Walsh F. Families in later life: challenges and opportunities. In: Carter B, McGoldrick M, eds. *The Expanded Family Life Cycle: Individual, Family, and Social Perspectives*. 4th ed. Allyn and Bacon; 2010.

55.4k Group Therapy

Kenneth M. Schwartz, M.D., F.R.C.P.C., Ingrid Söchting, Ph.D., R.Psych., C.G.P., Shayne Vitemb, M.A., L.M.F.T., C.G.P., and Alvin Keng, M.D., F.R.C.P.C.

Relevant models of new applications of group therapy are emerging steadily to meet the increasing needs of older adults with complex medical and mental health issues, although evaluation and research have yet to keep pace with these clinical developments. Foci of intervention include the aging-related losses of interpersonal, emotional, social, physical, and cognitive capacities. The group therapies provide demonstrably effective treatment achieved through improving social support and engagement while enhancing individual patient and family member coping capacities and skills. Group therapy is provided in ambulatory, hospital, and institutional settings and ideally should also be available in community and primary care settings. Online delivery of group therapy also deserves special mention, with unique considerations around access, privacy, and equivalence to in-person delivery. In all instances, the group leader's belief in and commitment to the continued potential for growth and change in later life is of overarching importance and is at the heart of clinical effectiveness in groups for older adults.

HISTORY

The roots of group therapy can be traced back to the work of social workers in a variety of settings for seniors, including homes for the aged, community centers, and mental health institutions in the earlier part of the 20th century. Group interventions combined activity and psychosocial treatments, emphasizing didactic instruction, inspiration, and social support, often referred to as a therapeutic milieu. In the following decades, interest in psychodynamic group approaches with older adults gained prominence as Maurice Linden, an American psychiatrist and pioneering group therapist, wrote a series of transformative articles on the topic. His careful observations led to the conviction that group therapy provided institutionalized patients with psychotic and various degrees of cognitive decline opportunities for resocialization, happiness, and some measure of self-sufficiency. Linden's perceptions have stood the test of time. The group approach was well tolerated by patients, satisfying to staff, and economical. Group therapy has since broadened its reach with the provision of treatment in acute care settings, day hospitals, ambulatory, and community settings. Accompanying this has been a shift in the emphasis from the powerful and caring leader to interpersonal models emphasizing the importance of relationships between group members in healing and change. Also, not only are psychotherapeutic interventions routinely integrated with pharmacotherapy, but the recognition of the limits of singular models has encouraged integration within group psychotherapy itself, synthesizing the benefits of cognitive-behavioral, dynamic, and developmental models. As is common in the field of group therapy, practitioners continue to respond to emerging clinical needs with diverse populations with a renewed emphasis on creating evidence-based models of intervention that are applied and then evaluated.

Current models include (1) verbal groups for the cognitively intact, which include groups for burdened caregivers; (2) verbal groups for the cognitively impaired, including cognitive stimulation, resocialization, remoralization, and reality orientation groups; (3) verbal virtual groups; (4) activity- and creativity-centered groups, including dance movement, drama, music, art, and poetry groups; (5) a variety of time-limited homogeneous groups addressing issues such as retirement, depression, anxiety, bereavement, loneliness, and medical illness (i.e., HIV/AIDS, chronic disease, pain); and (6) self-help and advocacy groups.

THEORETICAL ISSUES

Older adult individuals cover a large range, from baby boomers to centenarians, with an increasingly diverse range of diagnoses, abilities, and racial, gender, cultural, and religious backgrounds. The need for more effective psychological treatments will become more, not less important, as the increasingly aging population becomes more psychologically sophisticated, having utilized psychotherapy at earlier stages of life.

The themes encountered in older adult groups include reactions to loss, both developmental and traumatic, and to decline, with the resultant discontinuity with a prior, more valued sense of self. Consequently, themes revolve around issues of maintaining dignity, increased need for help and dependency, social isolation and loneliness, changing roles and/or conflictual relationships with adult children, mourning of losses of spouse and/or health, and existential struggles with retaining a sense of meaning and purpose. Certain ageist and stereotypical attitudes of society and some group leaders regarding, for example, late-life sexuality and intimacy contribute to the neglect of these subjects being rarely discussed in older adult groups.

Depression among older adults is a major public health concern with prevalence estimates for severe depression up to 9% and for milder up to 37%. Depression is often comorbid with other mental or medical issues, chronic, and with high health care costs for both the individual and society. Loneliness and the lack of social support are profound risk factors for depression. Group therapy provides opportunities to grieve losses and to re-engage with interests and relationships. It provides real relatedness alongside the acquisition of skills needed to generate continuing social integration beyond treatment. Because the experience of relatedness generated in the group is the result of the members' own actions, feelings of dependence may be reduced and feelings of personal competence may be increased. Enhanced social support produces a primary effect of improved well-being and a secondary, stress-buffering effect, which is particularly important for older adults. The importance of these supports and networks was highlighted during the COVID-19 pandemic. Higher levels of social support are associated with improved treatment outcomes for depression.

Often, group members experience shame at having lost their prior sense of self, viewing the current "incompetent" mental and/or physical replacement to which they must adjust as "bad" or "defective." In groups, they can be assisted to learn that their presumed defect is merely human and that other members also struggle with similar thoughts and feelings. The sharing of painful and deeply personal concerns can reduce the shame and stigmatization that isolate the individual. Through the therapeutic mechanism of altruism, members help one another, simultaneously raising self-esteem for both the receiver and provider of support. Realistic hope can similarly be nurtured countering feelings of demoralization and despair. Hope is malleable and can evolve with support to focus on new wishes and aspirations despite the undeniable decline, loss, and limited time. The vignette of Mr. G illustrates a group process leading to a more hopeful and renewed sense of self.

Mr. G, an 80-year-old man voiced displeasure at others in wheelchairs even though he himself needed one. He spoke with disdain about initially needing a cane, then a walker, and finally a wheelchair “because I do not want others to feel sorry or look down on me.” A second group member agreed while adding that she had only started using a wheelchair recently “against my wishes.” A third member, also in a wheelchair asked “Is it possible to retain one’s independence and pride while in a wheelchair living in a nursing home”? She then answered her own question stating “I do whatever I can for myself and I have ways to help others.” Other members in the group, both with and without wheelchairs, shared their own struggles in both seeking and accepting help, with other health issues as well, such as the loss of hearing or vision. Most acknowledged they were better at giving than receiving because they were raised that way. After more discussion, all the members, including Mr. G., agreed that accepting help when needed was a sign of strength and something that they wished at this later stage of life to learn to accept without shame or feeling like a burden to others.

Combined pharmacotherapy and psychotherapy are progressively employed in the acute, maintenance, and continuation phases of treatment. Combining drug and psychosocial therapies demands that attention be paid to the meaning each treatment has to the individual while being cognizant of the risks of devaluing one modality over the other. For example, a medication group where participants are encouraged to discuss their shame and fears associated with taking antidepressant medications can lead to better medication compliance and thus improvement in symptoms of depression. Group members realize through their interactions that despite negative feelings associated with taking medication, they are better able to accept the use of medication.

A metatheoretical synthesis of dynamic and developmental approaches along with cognitive-behavioral approaches into a seamless whole emerges from the recognition that effective psychotherapy is facilitated by understanding of the individual's private, subjective experience residing beneath more manifest interpersonal behavior, bridging emotional insight of oneself and others with skill acquisition. In all instances, the overarching goal is to increase each group member's sense of self-efficacy and dignity. Restoring a sense of personal effectiveness and the recognition of cause and effect in one's life is essential. This integrative model emphasizes the individual's sense of agency, as opposed to viewing oneself as a victim, of their own relational world. It combines interpersonal feedback within the here-and-now of the group and the processing of extragroup events and assigned homework, adding intensity and immediacy that enlivens cognitive and behavioral approaches. Behavioral activation and skills training build atop this. Allied with these more traditional and recent approaches are mindfulness-based cognitive therapy (MBCT) groups and dialectical behavior therapy (DBT) aimed at increasing reflective capacity, reducing reactivity, and enhancing skills in regulating negative emotions. Adapted for older adults, these approaches also focus on teaching skills to enhance openness to new experiences and in creating meaning over the course of life.

TECHNICAL INTERVENTIONS AND GROUP APPROACHES

Technical expertise in the application of group therapies is surpassed in importance only by the therapist's posture of respect, hopefulness, genuine warmth, and empathy. Study after study has reiterated the findings that effective group therapy is predicated upon a strong sense of group cohesion and the leader's capacity for both receptive (understanding) and expressive (communicative) empathy. Beyond the relationship between member and leader, successful treatment hinges on the establishment of group cohesion. Modifications of therapy approaches are sometimes necessary for group work with older adults in order to achieve and maintain cohesion. These include flexible use of protocols, the adaptation of cognitive restructuring techniques, additional supportive techniques, and practical arrangements for problems in hearing, vision, mobility, and other disabilities. Nonetheless, cohesion may be difficult to achieve with demoralized, withdrawn, and self- or other-devaluing patients. The group therapist, or therapists in groups with cofacilitation, need to be mindful that cohesion emerges both from the emotional bond shared by group members and the group's effectiveness with its designated tasks. Therapist interventions follow from the understanding of the tasks and goals of the therapy. The rationale for the group and the therapist's intervention should be transparent and clear. Cotherapy may be useful in providing an appropriate level of focus and activation, continuity during therapist absences, and the added opportunity to address countertransferential reactions through cotherapy discussion and the coconstructed space for thinking. Within institutional settings, cotherapy also facilitates the transmission of information between the group and other institutional care providers to ensure maximum treatment integration and reduce institutional resistance to group interventions.

Older adult groups focus on helping members cope with medical, psychological, mental health, and societal challenges associated with aging. Emotion-based coping in verbal groups helps reduce or manage the intensity of the emotional distress through the sharing of thoughts and feelings. Too often these feelings end up being expressed through somatization or the use of drugs, alcohol, or other less healthy ways. Problem-based coping is emphasized in situations in which something constructive needs to be done. Based on problem-solving therapy, a derivative of cognitive-behavioral principles, it is a more practical approach to help address problems by systematically examining and finding external solutions to problems without depth exploration. Existential-based coping is an opportunity to make use of the older adult's awareness of the limits imposed by aging and decline but also of the revitalization that can occur by focusing less on regrets over things undone or unsaid and more on living in the present before time runs out. The overarching goal of the therapist is to increase each group member's sense of self-efficacy. Restoring a sense of personal effectiveness and the recognition of cause and effect in an older adult's life is essential. The integrative model emphasizes the individual’s sense of agency, as opposed to viewing oneself as a victim, of their own relational world. It combines interpersonal feedback within the here-and-now of the group and the processing of extragroup events.

Attention must be paid to the maintenance of group norms and group integrity, which may be challenged by patient or institutional resistance. Patient resistance may be difficult to discern and can be interwoven with the severity of illness and comorbid illness, as well as on practical issues such as reliance on others for transportation to and from the group. In institutional settings, a positive alliance between the group therapist and administration is essential. Group therapy inevitably reflects the culture and norms of the larger milieu in which it is embedded. The principle of isomorphy notes that important concerns, values, and fears are transmitted throughout the institutional system and gain expression at every level, from patient to leader to group to ward to institution. Hence, an opportunity for dialogue and interchange between all the stakeholders in the treatment environment is invaluable while striving to ensure appropriate privacy and confidentiality within the circle of care.

A key aspect of this is to ensure that group sessions in institutional settings are scheduled in an accessible location without conflict with other scheduled activities. Active outreach including the arrangement of portering may be required to promote regular attendance. Groups for the older adults may often be held at midday in ambulatory settings and may be contiguous with lunch, to encourage opportunities for social interaction around mealtime. In institutional settings and day hospital programs, the treatment frame is also affected by the fact that group members will often have involvement with one another outside of the group. Opportunities for engagement and interaction outside of the group can be employed as part of the prescription of behavioral interventions and to facilitate homework. In all instances, a group norm should be established that protects the confidentiality of the group members and encompasses extragroup contact within the treatment frame.

Therapy groups generally consist of 6 to 10 patients, but activity groups or groups for depressed or frail populations may be larger and benefit from the increased energy in the room. Duration of sessions is shaped by members' capacities and may vary from 30 minutes daily for cognitively impaired patients to 90 minutes weekly for more cognitively intact patients. Many group formats are ongoing, but the majority of cognitive-behavioral groups and groups for burdened caregivers are closed and time-limited, lasting from 6 to 20 sessions. The limits of briefer treatments are being recognized, and treatments are often extended to include follow-up sessions that meet at a reduced frequency, but over a longer period of time, to extend treatment into maintenance and continuation phases. It is not uncommon in some settings to see members attend a monthly maintenance phase meeting for many years or for group participants to continue to meet informally after the group sessions end.

Group models are also developing outside of traditional venues. The COVID-19 pandemic has also impacted group therapists who have had to move groups from a circle of chairs in offices to online virtual platforms where members view frames of each other, from the circle to the square! Learning new technology and attending to new privacy concerns has been especially challenging for older therapists. A similar challenge exists for older adults to not withdraw into a "gero-pause" which is an avoidance of pursuing new interests or skills because it is too hard or complicated.

The ideal treatment setting will offer an array of group therapy options, phasing and tailoring interventions to patient requirements with graded levels of intensity and depth. The therapy should be experienced as a self-esteem–enhancing and sustaining experience and a nonfailure experience for each member. Therapist activity must focus on promoting group interaction in ways that maintains the dignity of participants. The therapist should achieve a balance between support and expression and may need to intervene with active prompting, reinforcement, reframing, and modulation of emotional intensity, emphasizing the invitation for what is desired rather than the rebuke for what is not. An important therapeutic task is the empathic confrontation of projective defenses and resistance while at the same time showing empathy for a group member's wish to protect themselves during an increased time of stress. That may result in disengagement that isolates individuals and undermines group cohesion. The therapist may need to be quite active in reinforcing progroup or procohesion exchanges, bolstering patient feelings of worth and belonging. It is essential not to permit depressive affect to overwhelm the group as depressed older adults are particularly prone to the effect of demoralization.

Below we review various group approaches and techniques including cognitive-behavioral therapy (CBT), existential, life review, activity, caregivers, and cognitive impairment. While there are other approaches, these are the major ones for older adult groups.

Cognitive-Behavioral

CBT emphasizes conscious cognition and new learning. The CBT model lends itself easily to a group setting for older adults. Cognitive techniques can help patients realize long-standing patterns of unhelpful (usually overly self-critical) thinking patterns that reinforce inaccurate and self-sabotaging beliefs about the self, others, and one's environment. In the above case of Mr. G, he erroneously believed that "the world" pitied him because of his confinement to a chair. Behavioral approaches focus on remediating the patient's reduction of environmental or social reinforcement often a result of their withdrawal, diminished interpersonal interactions, and thus "rustier" skills, but also an increase in legitimate adverse life events such as losing people to death and losing one's independence and abilities.

The cognitive behavioral approach emphasizes the teaching of skills with a consistent focus on patients assuming responsibility for practicing new behaviors including new thoughts, crediting themselves for improvements, and stimulating a sense of self-efficacy and mastery. Larger problems are deconstructed into smaller, more manageable components, as realistic goals are set, also referred to as the "chunking" approach. Behavioral techniques include goal-setting, but also other interventions such as modeling, role-play, psychoeducation, behavioral exposure to anxiety stimuli, and the review of homework with the provision of feedback about each group member's efforts in using prescribed skills to re-engage with pleasurable experiences or approaching previously feared situations.

Therapeutic strategies also include education about common "thinking traps" people of all ages are prone to. These include *catastrophizing* (e.g., "It will be terrible if I end up in a care home"), *mind-reading* (e.g., "They clearly feel pity for me"), and *black-and-white* thinking (e.g., "If I can't sing the way I used to why bother at all."). Patients are helped to examine evidence for and against their beliefs and to find alternative explanations that are more fair, accurate, and for sure helpful. The example of Mr. G illustrates a cognitive shift to a belief of still being able to contribute and enjoy mental stimulation despite physical mobility limitations. It is noteworthy that this shift was brought about by help from fellow group members who presented evidence countering Mr. G's previously held, more rigid beliefs. It is imperative that skilled group facilitators activate the entire "group brain" during the cognitive restructuring process and avoid relying on themselves to persuade a patient to "think positively."

In another example of Mrs. A from a CBT group, we see how the group can gently challenge a member to take a step out of their comfort zone.

In a day hospital for depression, Mrs. A, a 78-year-old widow, commented upon entry to the group that she was depressed and that “the group will not help me.” Other members of the group suggested a thought-challenging record could improve Mrs. A’s mood. Group members then assisted Mrs. A in finding evidence that did not support her belief of “being beyond help.” In a later session focused on creating personal action plans, Mrs. A observed she had withdrawn from her friends, was not returning phone calls, and was declining offers to socialize. Encouraged by the other group members, she decided on the action plan of returning a phone call to a friend. In the next session, she reported that she was able to complete the action plan of phoning the friend. Her friend was happy to hear from her and Mrs. A noticed her mood improved as a result. Having achieved this success and feeling the improved sense of well-being from having had a meaningful connection with a friend, Mrs. A decided that she would call a friend several times a week. Most importantly, she decided to stay with the group which contrary to her initial belief now seemed helpful to her.

CBT groups are quite structured, usually following a protocol with preset session content, check-in, and check-out, homework review, new learning, assignment of new homework, and discussion of anticipated obstructions to completion of the homework. It does require a baseline of cognitive capacity but can easily accommodate normal cognitive decline and early stages of dementia. Group therapists must respect the impact of slower cognitive processing and ensure adequate pacing, appropriate repetition, and reinforcement of new learning. The group setting is ideal for this model, linking together mutual support, positive modeling between group members, and the collaborative challenging of common cognitive and behavioral contributions to distress and depression. Group therapy outcome research indicates that the group format is just as effective as individual CBT, if not superior in some cases.

Although CBT groups are effective across the lifespan, an important drawback is the lack of active use of the group process. Finding ways to augment CBT by capitalizing on the valuable group process presents a challenge for most group CBT therapists. Promising efforts to integrate both a structured and a process approach have been made. A program at an outpatient clinic at a hospital in Western Canada developed an intensive, time-limited group therapy program for patients aged 65 and above. A CBT and an interpersonal therapy, IPT, group are each held on two consecutive days (IPT on Monday and CBT on Tuesday) over a 16-week period. Each group lasts for 90 minutes. Patients thus receive 3 hours of weekly group therapy for 16 weeks. The groups are closed, and size ranges between 7 and 10 members. Standard CBT interventions as described above are integrated with an IPT protocol adapted for a group format.

IPT is derived from attachment theory and recognizes that interpersonal function is a critical component of psychological adjustment and well-being. Group IPT therapists use stage-appropriate interventions and lead the group through the initial engagement stage, to differentiation, to work, and to termination. Similar to individual IPT, group IPT helps patients gain a sense of mastery and restore a sense of social belonging and meaning in life, with the emphasis primarily on current interpersonal relationships. Focusing on the elaboration of affect and self-disclosure is useful in gaining an understanding of the individual’s internal world, and the process of reminiscing (life review) may further flesh this out. The facilitators do not preplan thematic discussions or set agendas, but they help the group organize emerging discussions by identifying themes and may also encourage the group to carry over or revisit a particular topic. For example, goals set in the CBT group may be addressed in the IPT group from perspective of lacking motivation, and missing one’s spouse to do things with.

Existential

Contrary to the common reputation of Existentialism as a gloomy way of thinking, obsessed with death and dread, existential thinking can be seen as a dynamic, invigorating approach that encompasses a range of states of being in the world—from alienation and isolation to connectedness and transcendence. These aspects of existential thought can deepen the work of group therapists working with older adults and can be integrated into or inspire a variety of types of treatment in group therapy.

Existential psychology as well as existential philosophy can be said to share their early roots in a kind of response to Descartes' famous statement, "I think, therefore, I am." This statement, revolutionary in its time, helped a chaotic Europe emerge from a world of violence into a realm of rationality and civilization. But over time, the full context of Descartes' statement was lost, placing rationality in center stage, while leaving the "I am" part of this statement neglected. Existentialists placed the emphasis on the "I am" part, insisting on examining existence or being, not just thinking, as an opening into a discussion of human existence.

Group therapy for older adults has the unique capability of providing its members with an opportunity to examine what it means "to be," as well as what it's like to "be in the world." Group members can explore such questions as, what do I make of the world I am now living in, and what does it mean to "be" in this world? Is there a place for me? Do I have a contribution to make? Can there still be passion in life, and are new friendships possible? What about the past?

Some of the more common understandings of existentialism focus on death awareness and the finiteness of life as a path to the appreciation of life in the present. With older adults, there is less of a need—perhaps no need–to heighten one's awareness of death—they know it all too well. Group members are likely to have already experienced profound loss and, due to their age, know that they have more years behind than ahead of them. Many have been treated for serious illnesses or have loved ones who suffer from very specific life-threatening problems such as cancer or heart disease. For them, awareness of life's finiteness may be a source of very real and present pain. With this in mind, processing around issues of death need not be emphasized, nor should they be feared or avoided.

The focus of existential group therapy with older adults can focus on such existential concepts as isolation, everyday life, existential anxiety, and life's meaning as identified by psychological thinkers such as Irv Yalom and Rollo May. Making good use of the awareness of the finiteness of time in the group, such as the 90-minute session as a metaphor for life with its beginnings and endings is of great value. Members may seek support in repairing damaged relationships with adult children, or reframing what is meaningful to them, as well as forming new connections that bridge isolation in the here and now of the group therapy experience. Ideally, the group can be a haven for an "I-Thou" experience as inspired by Martin Buber, in which one is connected to another being in a very genuine way, with less pressure from external constraints including social status and pressure to be "nice." When bringing existential influences into one's group work, it is important that the concepts not be rigidly or dogmatically imposed, since its essence is an honoring of the authentic experience of individuals and the freedom and responsibility that entails.

Life-Review

Life review and reminiscing techniques employ the naturally occurring process of patients' recollections and reviews of their lives. It can serve as a vehicle for making meaning of one's life, viewing it as valuable and of consequence. Furthermore, the life review approach can facilitate the resolution of old conflicts and acceptance of losses and the unchangeable facts of one's life. When used in a group format, it also creates a sense of social identification among group members. Reminiscing can be used as the main treatment focus or to stimulate cohesion in beginning groups as an introductory exercise. Reminiscing can be spontaneous or prescribed in a series of sessions, following a chronologic life course sequence. Either mile stoning, which emphasizes only positive events, can be utilized, or the review process can be comprehensive and autobiographical, providing an opportunity for mourning, conflict resolution, and acceptance. Encouraging a focus on the past is not without risk. Facilitating the group members' return to dialogue within the here-and-now of the group, eliciting feedback and interaction, ensures that morbid self-absorption, preoccupation with painful memories or regrets, and guilt do not exacerbate the individual’s withdrawal from the group or adherence to the past as a way of avoiding engagement with the present.

Activity and Creativity

Activity- and creativity-centered groups (e.g., dance, movement, pottery, music, or drama) encourage social engagement through shared participation in a pleasurable activity that may rekindle a connection with the previous sense of self through the use of the creative process. Group activities elicit feelings of pleasure and mastery, challenge self-denigrating and constrictive distortions that individuals may have about themselves, and counter the trend toward homogenization. Although the focus in the group is on physical activity, integrating a psychodynamic appreciation of the meaning of the activity groups amplifies the use of a physical activity to promote social interchange. This has been demonstrated to assist in role transition, acceptance of loss in body function, and integration of new goals, interests, and meaning in life.

Caregivers

Groups for caregivers are used widely in response to the increasing number of frail or cognitively impaired individuals being cared for in the community by their families. Increasing longevity is a challenge for both society and family members, generally unprepared with little knowledge and few skills to deal with their loved one’s physical and especially cognitive decline associated with dementia. These caregivers are often highly stressed, isolated, overwhelmed, and have an increased risk of psychological, physical, and financial problems, yet they remain the key resource to the implementation of any interventions for the identified patient. Groups may provide an opportunity for working through feelings of loss and abandonment, as well as anger at the burdens imposed on the caregiver. Attention to the process of the giving and receiving of mutual care within the actual groups provides a direct opportunity to address self-management and self-care issues. Education, social support, and skills training may all be integrated into these groups. Using a simulation of emotionally challenging caregiver scenarios is also being applied to good effect in facilitating the direct application of specific and relevant coping skills. Group therapy interventions for caregivers work best when they integrate a variety of techniques that are structured and intense; require the active involvement of caregivers, include education, support, and respite; promote knowledge transfers, skill building, and competency; and are combined with individual support.

Cognitive Impairment

Growing attention to individuals with mild cognitive impairment is evident in groups that provide psychoeducation, support, and coping strategies for couples with the objective of reducing anxiety, aligning expectations, and jointly coping while reducing marital stress. Groups for cognitively impaired patients focus on patient engagement, with the structure of the group activity employing multiple modes of cognitive stimulation. Interpersonal and group processes are generally not addressed. These groups employ a broad range of sensory and cognitive stimuli to activate and engage patients, with consistent reinforcement and validation provided by the leader(s). The therapist should be guided by the principle of conferring meaning to every patient's communication rather than dismissing these expressions as irrelevant. Cognitive exercises, activities, word games, life reviews, discussions, participation in arts and crafts, pleasant aromas, music; baking; guided imagery; relaxation; and physical exercise are components of these groups. What links all such groups is the moment-to-moment focus on esteem-enhancing patient engagement with the world around them. Groups of 6 to 10 participants work well. The group should be small enough that each member actively participates. Future growth in these approaches is likely, reflecting the growing role and positive effects of nonpharmacologic interventions for dementia recommended in contemporary practice guidelines.

CLINICAL ISSUES

Countertransference Challenges

The depth of the despair associated with some older adults' struggles with their living situations in the community or long-term care settings can be occasionally difficult to endure for the therapist and lead to withdrawal from working with these populations. Patients and therapists must understand that demoralization is a state—it is a specific emotional reaction and not an obligatory trait of aging. While recognizing age-related limitations of time and energy, the belief that an older adult's life will be one of unhappiness and despair can be challenged, often through the countervailing experience in the here and now of the group, and successful aging and improved mood can be resumed. Identification with a member's position of hopelessness loses the therapeutic perspective. While instilling or reinforcing hope is essential, therapists must be careful in a wish to avoid inflicting pain on members or themselves by offering unrealistic hope, which may indicate discomfort in hearing about the struggles of certain members. When this occurs, members will experience their emotional pain as being dismissed and not being shared in the group. Realistic expectations help therapists guard against such negative countertransference associated with profound difficulties with hopelessness, self-criticism, and beliefs about one's competence. Fortunately, there are enough rewards, including older members offering an explicit appreciation of the therapist's understanding and acceptance of who they were and still are, to make work with this increasing and still underserviced population so worthwhile.

Another critical issue involves age discrepancy, as the therapists are almost always younger than the group members. Working with older adults can have an impact on therapists who are younger or perceived as younger (e.g., "you could be my grandchild") than their patients. This consideration is an important aspect of the provision of group therapy to older adults. The countertransference experienced by younger therapists warrants full and fair exploration in supervision. Prior to commencing group therapy with older adults, younger therapists should be prompted and encouraged to explore their motivations, apprehensions, and preconceived notions of the upcoming work. During the course of therapy, strong reactions may develop among younger therapists. For example, patients may be reminded of parental or authority figures from their past or present. Conversely, patients may also be experienced as benign, frail “grannies” who need to be “taken care of” based on the therapists’ experience and memories with their own grandparents. Age in countertransference is also fluid, and such roles are reversed. Supervision around these topics, therefore, should not be presumptuous and should be provided on an ongoing basis. Younger therapists are also more likely to engage with this modality for the first time as a trainee. Regular supervision allows for addressing their unique educational and professional growth needs and will ensure future generations can continue this work with aptitude and confidence free of therapy-interfering projections.

Indications and Contraindications for Group Participation

The choice of a group modality for older adults may take into consideration both practical issues, such as the availability of various group types, along with affordable and convenient transportation, and patient issues, such as the type and extent of psychiatric illness and/or cognitive impairment.

Group therapy may be used as an adjunct to pharmacotherapy, or individual treatment, or as an independent treatment. With respect to selection criteria, structured groups, especially those that are more homogeneous in terms of membership, ego functioning, and cognitive capacities and not age alone, are more effective than unstructured groups. There is a potential for poorer treatment outcomes in groups that lack either focus or appreciation for the potential to amplify depression by intensifying affects related to illness and death beyond some members’ capacity. As with all treatments, therapeutic success is enhanced by the linking of patient goals and abilities to the tasks and techniques of the treatment. Mixing cognitively impaired and intact patients together is likely to jeopardize cohesion and group effectiveness by generating a wish among members to disidentify and withdraw from the more impaired participants.

Although pregroup preparation has not been extensively researched with older adults, the evidence for its usefulness in other patient populations is quite substantial. Pretherapy preparation outlining the group treatment's goals and methods demystifies an otherwise anxiety-provoking situation, normalizes and destigmatizes treatment, helps to establish a therapeutic alliance, and sets group norms regarding regular attendance, confidentiality, and extragroup socialization. This process may increase group therapy tenure and task adherence and foster patient hopefulness. Role-play or simulations may be utilized to induct the patient into the therapy role.

A written summary of the group therapy logistics, norms, tasks, and expectations may be useful to supplement the preparation interview. Pregroup preparation should also explore treatment resistance. Identification of ego-syntonic character difficulties, in particular narcissistic pathology or interpersonally alienating behaviors may be useful in containing behaviors that can result in early peer rejection from the group.

Dropout rates in group therapy, in general, range between 10% and 50% in younger adults and are likely the same with older adults. The likelihood of premature termination from treatment increases with increasing severity of depression, comorbid physical illness, and comorbid characterologic difficulties of devaluation and externalization.

Burdened caregivers are a notably vulnerable group. Limits in health care resources coupled with an aging population have resulted in the common situation of caregivers routinely providing at home the kind of care that in a previous generation was provided by professionals. Caregiver stress emerges from strain around safety, managing patient problematic behavior, self-care, social support, and personal emotional well-being. Caring for the cognitively impaired individual generates greater psychological strain and burden than caring for the medically ill. Family caregivers are often disproportionately vulnerable to both psychological and medical illness.

Contraindications

In general, individuals who require acute and active management because they are in crisis, suicidal, actively abusing substances, or acutely psychotic, do not do well in general groups. Severe cognitive or sensory impairment or medical illness that obstructs the individual's engagement with the group enterprise may also require exclusion from group therapy. Some contraindications are relative and can be remedied by employing groups that are homogeneous for that condition. A program that has very few dropouts or therapy failures likely has set the bar for entry too high and may be missing opportunities to treat some patients who could benefit.

Goals

Notwithstanding the range of group formats described, consensus exists regarding realistic goals of treatment. Although obviously shaped by particular patient needs, in general, the goals of treatment include reduction of symptomatic disturbance; reduction of isolation; restoration of interpersonal skills and engagement within a social network; correction of psychological distortions and cognitive beliefs; restoration of self-esteem; improved cognitive function; acquisition of problem-solving and coping skills; grieving and adaptation to loss; and realistic acceptance of one's needs and limitations with effective use of available resources. In institutional settings, group therapy may provide a forum to deal with the imposed inhibition of personal autonomy associated with routine, regulation, and restricted opportunity, thereby enhancing the quality of life by addressing the dissonance between the individual's wishes and the imposed environment. At other times, addressing interpersonal tensions between residents may be the focus of treatment. A men's group may have particular value in institutional settings in which men are in the minority and feel isolated and particularly distant from former pursuits and interests.

Objectives of groups for burdened caregivers include education about the illness process and specific ways of providing care with the acquisition of strategies for dealing with disruptive behavior and problem-solving and stress-reduction skills. Working through loss and grief and facilitating appropriate disengagement that may lead to more realistic decisions about placement are also important objectives. Additional areas of focus include legitimizing the needs of the caregiver and promoting self-care, working through anger and guilt, and addressing relationships with healthcare professionals. These objectives may be integrated or addressed in a series of modules emphasizing important aspects sequentially.

Groups for the cognitively impaired aim to reduce excess disability and foster more appropriate social behavior and social functioning, thereby optimizing the quality of life. Groups that emphasize creative expression aim to restore a sense of competence and mastery through the activity be it art, drama, sculpting, or dance. Wellness and exercise groups aim to maintain patient well-being through physical activity and stress reduction provided in the context of a social milieu.

ETHICAL ISSUES

Proper patient selection and preparation and competent therapist functioning contribute to an environment that is generally well tolerated and experienced by participants as respectful and helpful. There is growing awareness of the differential impact group therapists can achieve, even when applying the same model of intervention. The interplay of therapist attributes and models of care may result in wide differences in impact.

Ethical care demands the therapist commit to ongoing training, evaluation of one’s clinical impact, and supervision or consultation. The ending of a group may be resisted because of the important social support and self-restorative experience offered by the group and because it provides a sense of stability and consistency in the face of the participant’s many life changes. This necessitates an active focus on the transfer of skills gained in treatment into the external world, recognizing that treatment is a bridge and not a destination. Therapeutic groups must be well integrated collaboratively into the overall milieu in which the individual resides.

Alterations in the frame of group therapy required within institutional settings, along with other factors that contribute to a higher likelihood of extragroup contact both occurring and even being endorsed, warrant special attention to issues of confidentiality. Extragroup contact and the risk of subsequent subgrouping may be managed by diminishing secretiveness around the extragroup relationships and actively bringing them to the attention of the group

Societal stereotypes of aging often lead to barriers hindering older adults from being referred and appropriately treated. For too many, the idea of aging stirs up feelings of loss, deterioration, and chronic medical illness, and challenges the needs of a culture obsessed with youth and youthful physical beauty as opposed to the beauty that reflects self-acceptance and the wisdom from a life well lived. Ageist beliefs among group members themselves can interfere. It is common to hear myths such as “depression or dementia is an inevitable part of aging” and “I am too old to change.” The misperceptions must be identified and worked through in the group. As a result, attention to therapist countertransference and reactions to the clinical work is particularly important in the practice of group psychotherapy. Realistic expectations guard against such negative countertransference as denial, hopelessness, and therapeutic despair. A reflective capacity for self-examination is particularly important in working with older adults because of potential issues of identification with patients, filial countertransference, and avoidance of the issue of sexuality.

An example of reflective capacity on the part of the group leader is offered in the Mr. K vignette. The group leader permits addressing the topic of romantic and sexual desires. The ethical issue involves balancing the needs of individuals with dementia with the needs and concerns of those around them, including their adult children.

Mr. K, a 75-year-old widower with minimal cognitive and physical impairment, became attracted to Mrs. B, a 73-year-old widow with moderate dementia of Alzheimer type. They began spending much time together and even expressed a mutual wish to get engaged. Mr. K in the group shared that “We want to be always together. We kiss now, maybe one day we’ll do more, but I wouldn’t want to touch her if she wouldn’t want me to. I would respect her feelings.” The couple went on to discuss how Mrs. B’s family wanted to break up the relationship as they believed Mr. K would exploit her and that if she was more cognitively intact, she would remain loyal to her deceased husband. Group discussion focused on how to honor both the obligation to promote autonomy and dignity and to offer protection when needed and whether it is possible to balance both ethical principles. Several ethical issues are raised: Is it appropriate to apply the usual consent to medical treatment to consent to sex? Should cognitive capacity be the sole and/or main determinant of whether or not a sexual relationship should be interfered with by adult children or nursing home staff or management? Under which circumstances, if any, should the family be notified; and should a person’s prior self take absolute precedence over the desires of the person now with some degree of dementia who still retains the ability to consent?

Group therapy is compatible with clinically indicated pharmacotherapy or individual therapy. Its concurrent use is complementary and not antagonistic. However, a collaboration between treatment providers is the critical, at times neglected component of effective conjoint therapy.

An additional important ethical concern relates to the removal of barriers to access to care that are rooted in ethnic and cultural patient factors. The concept of cultural competence is gaining prominence in the field of mental health as an important variable influencing treatment effectiveness. Cultural competence relates to the therapist’s appreciation for and understanding of the patient’s specific cultural context regarding illness, aging, mental health, family relationships, and, of course, treatment. The ideal group therapy will be provided in the primary language of the participants and in accessible settings within the specific communities. Recognition of crosscultural factors may require attention to the culturally resonant translation of therapy models.

RESEARCH AND EVALUATION

Group therapy in older adults is effective in treating a variety of disorders. There is a substantial evidence base behind modalities such as cognitive-behavioral (CBT) and mindfulness-based CBT treatments. There is also emerging literature on other group-based modalities, such as reminiscence therapy (RT), dialectical-behavioral therapy, schema-focused therapy, facilitated support groups, and care management for patients and their caregivers. Although the majority of this literature describes the delivery of these modalities in person, some literature describes the use of telephone and online platforms.

In a meta-analysis comparing the effects of psychotherapies for depression between different age groups, there was no meaningful difference between the effect size of psychotherapy between middle-aged adults (24 to 55 years), older adults (55 to 75 years old), and older adults (75 years+). In a systematic review ascertaining the effects of group therapy on geriatric depression, group therapy led to an improvement in geriatric depression, as well as other psychological variables, in a variety of settings, protocols, and participant characteristics. In the treatment of depression, therefore, older age alone does not prohibit from benefitting from psychotherapy. Furthermore, there may be common factors across modalities of group therapy that result in the improvement of older participants.

Group CBT is perhaps the best-studied modality among older adults and across a variety of psychiatric conditions. In a randomized control trial of older adults, group CBT demonstrated faster and sustained improvements in depression and anxiety in comparison to a control discussion group up to 6 months post-treatment. In another study, the use of group CBT, combined with duloxetine, showed a significant treatment response for generalized anxiety disorder in comparison to duloxetine alone. Telephone-based CBT also has demonstrated superiority in comparison to supportive therapy, treating anxiety symptoms among older adults up to 1-year post-treatment. Furthermore, culturally adapted models of CBT have shown promise in alleviating worries and anxiety over dementia among older Chinese adults when compared to treatment-as-usual. Group-based CBT also has evidence in treating post-traumatic stress disorder (PTSD) among older veterans and substance-use disorders. Lastly, an adapted version of CBT has been used for patients with mild cognitive impairment and their caregivers and has been found to help patients accept their illness, with female patients showing a decrease in helpless cognitions. Overall, group CBT appears to be helpful for older adults.

Group mindfulness-based CBT treatments, often referred to as “third-wave” therapies, have also been shown to be effective in treating older adults. One meta-analysis of mindfulness-based therapies for older adults showed a moderate effect size in treating depressive symptoms. Another meta-analysis of mindfulness-based stress reduction (MBSR) has shown effectiveness in reducing depression in older adults but concluded there was no clear evidence about the results extending to anxiety and stress. Another version of mindfulness-based CBT, MBCT, has been tested in randomized control trials and shown to be effective for anxiety symptoms in older adults in residential care. Overall, various group mindfulness-based treatments are effective in treating depression and possibly anxiety in older adults. Limited evidence exists for their effect on other psychiatric disorders, but it may be promising work in the future.

RT was introduced in dementia care in the 1970s and involved the use of discussion of past activities, events, and experiences, with the aid of tangible prompts. Reminiscence may be carried out in individual or group formats. While individual RT has been shown to have a small positive impact on cognition and quality of life, the studies of group RT have not shown this, suggesting that the mode of delivery is important in RT. However, group RT may have a positive effect on the communication of the patient with dementia, though the effect is small. Similarly, there is limited research to support group RT on depression, and anxiety in older adults in long-term care, as found in a systematic review. Group RT may have an adjunctive role in dementia care and other disorders but may not be sufficient as a stand-alone therapy.

A review of two randomized clinical trials using DBT skills training groups as an augmentation strategy of antidepressant medication to treat chronic depression and personality disorders in older adults demonstrates significant advantages for decreasing remission rates for medication and DBT compared to medication alone in the treatment of chronic depression. As well, DBT plus medication demonstrates superiority compared to medication alone in decreasing interpersonal sensitivity and interpersonal aggression at post-treatment and 6-month follow-up in a sample of older adults with comorbid depression and personality disorder. Research is underway to explore the effectiveness of the DBT model that has been modified to better meet the needs of the older adults.

A multicenter RCT of group-schema–focused therapy enriched with psychomotor therapy for older adults with cluster B and C personality disorders is currently underway and offers to enrich our understanding of treating personality disorders in later life.

A systematic review of group and individual cognitive intervention programs for individuals with amnestic mild cognitive impairment demonstrates comparable efficacy of both training formats. Statistically significant improvement after either treatment is obtained in objective memory and on subjective measures of memory, quality of life, or mood. It is suggested for further research to combine group and individual approaches as both approaches have unique benefits.

Participatory group-based case management seemed to decrease loneliness among those living alone in a randomized control trial of older people living alone. Furthermore, groups for burdened caregivers generally demonstrate that participants are very satisfied with the group experience. Complexities in the measurement of subjective and objective burden, coupled with the likely progressive deterioration of the care recipient, may confound other outcome measures. The acquisition of knowledge is easy to measure and demonstrate. Reductions in burden are harder to demonstrate. Both professional- and peer-led groups show significant effects, with the former tending to emphasize the provision of information and problem solving and the latter emphasizing peer support and networking. The best results emerge from integrative interventions that employ a broad focus, recognizing that there are multiple domains of concern for caregivers. This is consistent with a summary of the reviews of the caregiver intervention literature that demonstrates that more recent intervention studies are more effective when they employ combined interventions that target both the caregiver and the patient.

A pilot study of evidence-based interventions for family caregivers of individuals with behavioral and psychological symptoms of dementia demonstrates improvements in caregivers' self-perception of competence, significant reductions in emotion-focused coping, enhanced mastery, and reduced burden. Caregivers with more compromised baselines benefited most from the intervention comprised of three integrated components: group psychotherapy, problem-solving techniques, and skill acquisition for specific current challenging interactions in caregiving. The systematic use of standardized patients to role-play the spouse or parent with dementia is employed. Limitations of the study include the lack of a control group of carers and no outcome measures for care recipients.

Online delivery is an emerging modality for leading group psychotherapy but is scarcely researched. The COVID-19 (coronavirus) pandemic resulted in the rapid application of online groups for treating psychiatric conditions. Online groups provide several benefits, including improved access to care and cost-effectiveness, though they also pose several challenges, including confidentiality, access to technology, and managing core therapeutic tenets (i.e., managing the therapeutic frame, adapting to a lack of physical presence, and body language). There may be additional challenges for older adults as well who may not be familiar with online technology.

FUTURE DIRECTIONS

We can reliably anticipate an ever-increasing need for effective and meaningful group therapy interventions for the older adults and their families. Challenges and opportunities that lie ahead include creating models of intervention that are culturally accessible to a heterogeneous population of seniors. The field will also need to develop and test interventions for mood disorders that address common comorbid conditions, including anxiety, substance abuse, personality disorders, and, of course, cognitive impairment. In an allied fashion, we will need to develop integrative group interventions for the treatment and management of patients with dementia and their families, both together and separately. Groups for caregivers, both family and professional, of older adults with late-life psychiatric and medical illness and behavioral and psychological symptoms of dementia also need to be developed. In all domains, greater efforts will need to be made in the training of group therapists as we link theory to practice to empirical evaluation of the therapy.

FURTHER READINGS

Burlingame GM, Strauss B, Joyce AS. Change mechanisms and effectiveness of small group treatments. In: Lambert MJ, Bergin SE, Garfield SL, eds. *Bergin and Garfield’s Handbook of Psychotherapy and Behavior Change*. 6th ed. Wiley-Blackwell; 2013:640–689.

Chiu M, Wesson V, Sadavoy J. Improving caregiving competence, stress coping, and mental well-being in informal dementia carers. *World J Psychiatry*. 2013;3:65–73.

Cuijpers P, Karyotaki E, Eckshtain D, et al. Psychotherapy for depression across different age groups: a systematic review and meta-analysis. *JAMA Psychiatry*. 2020;77:694–702.

Fain DS, Sharon A, Moscovici L, Schreiber S. Psychotropic medication from an object relations theory perspective: an analysis of vignettes from group psychotherapy. *Int J Group Psychother*. 2008;58:303–326.

Heidegger M. *Being and Time*. Translated by Joan Stambaugh. State University of New York Press; 2010.

Kishita N, Takei Y, Stewart I. A meta-analysis of third wave mindfulness-based cognitive behavioral therapies for older people. *Int J Geriatr Psychiatry*. 2017;32:1352–1361.

Leszcz M. Group therapy. In: Sadavoy S, Lazarus LW, Jarvik LF, Grosberg GT, eds. *Comprehensive Review of Geriatrics Psychiatry II*. 3rd ed. American Psychiatric Press Inc.; 2004.

Li SYH, Bressington D. The effects of mindfulness-based stress reduction on depression, anxiety, and stress in older adults: a systematic review and meta-analysis. *Int J Ment Health Nurs*. 2019;28:635.

Lo Coco G, Melchiori F, Oieni V, et al. Group treatment for substance use disorder in adults: a systematic review and meta-analysis of randomized-controlled trials. *J Subst Abuse Treat*. 2018:99:104–116.

Lynch TR, Cheavens JS, Cukrowicz KC, Thorp SR, Bonner L, Beyer J. Treatment of older adults with comorbid personality disorder and depression: a dialectical behavior therapy approach. *Int J Geriatric Psychiatry*. 2007;22:131–143.

Pierce BS, Perrin PB, Tyler CM, McKee GB, Watson JD. The COVID-19 telepsychology revolution: a national study of pandemic-based changes in U.S. Mental Health Care Delivery. *Am Psychol*. 2020;76:14–25.

Qiuling A, Wang K, Sun F, Zhang A. The effectiveness of modified, group-based CBT for dementia worry among Chinese elders. *J Affect Disord*. 2020;274:76–84.

Ristolainen H, Kannasoja S, Tiilikainen E, Hakala M, Närhi K, Rissanen S. Effects of ‘participatory group-based care management’ on well-being of older people living alone: a randomized controlled trial. *Arch Gerontol Geriatr*. 2020;89:104095.

Schwartz K. Concurrent group and individual psychotherapy in a psychiatric day hospital for depressed elderly. *Int J Group Psychother*. 2004:54:177–201.

Schwartz K. Remembering the forgotten: psychotherapy groups for the nursing home resident. *Int J Group Psychother*. 2007;57:497–514.

Schwartz K. Sexuality, intimacy and group psychotherapy with older adults. *Int J Group Psychother.* 2019;69(1):126–144.

Schwartz K, Schwartzberg SL. Psychodynamically informed groups for elders: a comparison of verbal and activity groups. *Group*. 2011;35:17.

Seritan AL, McClud, NK, Hinton L. Geriatric depression: review for primary care. *Curr Psychiatry Rev*. 2009;5:37–142.

Sloan DM, Unger W, Lee DJ, Beck JG. A randomized controlled trial of group cognitive behavioral treatment for veterans diagnosed with chronic posttraumatic stress disorder. *J Trauma Stress*. 2018;31(6):886–898.

Söchting I, O’Neal E, Third B, Rogers J, Ogrodniczuk JS. An integrative group therapy model for depression and anxiety in later life. *Int J Group Psychother*. 2013;63:502–523.

Syed Elias SM, Neville C, Scott T. The effectiveness of group reminiscence therapy for loneliness, anxiety and depression in older adults in long-term care: a systematic review. *Geriatr Nur*. 2015;36(5):372–380.

Tavares LR, Barbosa MR. Efficacy of group psychotherapy for geriatric depression: a systematic review. *Arch Gerontol Geriatr*. 2018;78:71–80.

Vitemb S. Talkin’ bout my generation: existentialism, aging, and newly emerging issues in group therapy. *Int J Group Psychother.* 2018;68(3):337–351.

Vitemb SA, Vitemb SA. Existentialist thinking and its influence on group therapy. *GROUP*. 2012;36(2):107–119.

Weinberg H. Online group psychotherapy: challenges and possibilities during COVID-19-A practice review. *Group Dyn Theory ResPract*. 2020;24:201–211.

Wilfley DE, MacKenzie KR, Welch RR, Ayres VE, Weissman MM. *Interpersonal Psychotherapy for Group*. Basic Books; 2012.

Woods B, O'Philbin L, Farrell EM, Spector AE, Orrell M. Reminiscence therapy for dementia. *Cochrane Database Syst Rev*. 2018;3:CD001120.

Wuthrich VM, Rapee RM, Kangas M, Perini S. Randomized controlled trial of group cognitive behavioral therapy compared to a discussion group for comorbid anxiety and depression in older adults. *Psychol Med*. 2016;46:785–795.

Xie ZJ, Han N, Law S, et al. The efficacy of group cognitive-behavioural therapy plus duloxetine for generalised anxiety disorder versus duloxetine alone. *Acta Neuropsychiatr*. 2019;31(6):316–324.

55.4I Counseling and Support Needs of Dementia Caregivers

Liat Ayalon, Ph.D., and Patricia A. Areán, Ph.D.

INTRODUCTION AND DEFINITION OF CAREGIVING

What does caregiving mean? Webster's online edition defines a caregiver as "a person who provides direct care (as for children, older adults, or the chronically ill." A caregiver may be either a family member or a friend who provides informal care or a formal caregiver who receives money for their services. While anyone may assume the function of a caregiver, a primary caregiver is a family member or a friend who provides care to a person who can no longer take care of themselves. Although the reasons for assuming these responsibilities vary between the two types of caregivers, the burden of caring for a person with dementia (PWD) is high for both types of caregivers. Currently, informal caregivers are the majority of the caregiving population, and in most cases, care is provided either informally or through a collaboration of formal and informal carers. According to the Alzheimer's Association, in 2020, more than 16 million informal caregivers provided care to individuals with Alzheimer disease (AD) and other dementias. This has a tremendous economic value amounting to $305 billion and 18.6 billion hours of care. Caregivers of PWD provide substantially more care compared with caregivers of individuals with cognitive impairment or caregivers of people who are cognitively intact. Given the increase in life span, the decrease in childbirth, the entrance of women into the workforce, the high divorce rate, and the transition into nuclear family living, it is expected that formal long-term care alternatives will become more popular in future years.

Who is most likely to be a caregiver? Based on the Alzheimer’s Association, one in three caregivers of PWD is 65 and older, and almost 66% are women. About two-thirds of the caregivers live with the PWD. About one-quarter care for a child under the age of 18 in addition to caring for a PWD, thus representing the “sandwich generation.” According to the National Alliance for Caregiving (NAC/AARP), 30% of all caregivers also cared for children younger than 18 and, thus, are torn between caring for an older relative and younger children.

Researchers found no demographic differences between caregivers of individuals with dementia and those with milder cognitive impairments. Most caregivers of dementia patients care for people with AD, although a fair number also care for people with Lewy body disease, frontotemporal dementia, and vascular dementia. Each of these illnesses poses a unique set of problems for caregivers. AD is typically characterized initially by memory impairments, but as the illness progresses, it can result in personality and behavioral changes that may be upsetting to caregivers. Lewy body disease shares this same course but is coupled with physical disability related to tremors and unusual hallucinations that many caregivers find hard to understand and manage. Frontotemporal dementia is particularly vexing, as patients who develop this illness tend to be younger and physically healthy (men in their 50s), and while their memory remains intact for a long time in the course of the illness, behaviorally these patients become impulsive, inappropriate, and highly agitated. These disorders are not mutually exclusive. A person with AD may also have vascular disease; a person with Lewy body disease may also develop frontotemporal dementia. The more complicated the dementing illness in the patient, the more caregivers have to contend with. Because AD is the most common dementia afflicting older people, caregivers of persons with AD have received the most attention, with the majority of research focused on informal caregiving.

The Mental Health Consequences of Caregiving

Caregivers are generally at greater risk for mental health and health problems than noncaregivers. A recent systematic review has shown that women, married caregivers, and those providing intensive care are most likely to be negatively affected. Consistently, those caring for older adults with dementia tend to present with more negative emotional consequences compared with those who care for older adults who do not suffer from dementia. The most common mental health consequences of caring for a PWD are depression, anxiety, and a condition called *burnout.* These three conditions are often lumped together under the term *caregiver distress.* Studies find mixed results as to which condition is most common. It is largely believed that depression is the most common mental health consequence of caregiving. The Aging, Demographics, and Memory Study (ADAMS) found that as many as 44% of the caregivers of PWD were depressed, but only 27% of the caregivers of older adults with cognitive impairment but no dementia indicated depressive symptoms. Others have shown that as many as 40% of caregivers of older adults with AD are depressed, but only 5% of noncaregivers report depressive symptoms and meet the clinically significant cutoff of depression.

It is important to note that burnout is not necessarily the same condition as depression or anxiety, but these conditions do co-occur. According to Schulz, caregivers of PWD are particularly susceptible to burnout, a psychological and physical state of exhaustion resulting from excessive and prolonged stress. Caregivers who are family members are particularly susceptible to burnout, and many speculate that this emotional consequence of caregiving explains why 90% of elder abuse is committed by family members. Stress and burnout are not one and the same. People under stress can generally feel hopeful about the future, and when no longer in the stressful situation, they begin to feel like their normal selves. Burnout is far more severe and is characterized by frustration; feeling powerless and hopeless; feeling drained of emotional energy; becoming detached from others, which results in withdrawal and isolation; and feeling trapped, irritable, and sad.

Caregiver distress, be it depression, anxiety, or burnout, does not necessarily remit after caregiving responsibilities end. According to Carol S. Aneshensel and colleagues, caregivers who are less distressed before the death of the care recipient tend to have better mental health outcomes, whereas caregivers with significant emotional distress while caregiving tend to become even more depressed after the loss of the care recipient. Caregiver burden before death tends to increase the odds of becoming depressed; self-esteem and emotional support tend to have protective roles. A recent systematic review has highlighted the need to acknowledge that there also are positive aspects associated with caring for a PWD, which should be acknowledged.

The Effects of Caregiving on Physical Health

In addition to mental health consequences, there is a notable increase in poor health outcomes in caregivers. Compared with caregivers to older adults without dementia, caregivers to older adults with dementia are more likely to say that caregiving worsens their health. Some researchers speculate that poor health outcomes are the result of poor healthcare behaviors; caregivers generally do not have sufficient time to attend to their own medical appointments, and therefore potentially preventable health problems arise, such as hypertension and heart disease. Recent research also suggests that the stress related to caregiving may also influence health outcomes. Brent T. Mausbach and colleagues found that an exaggerated plasma norepinephrine response in caregivers is linked to an increased allostatic load related to cardiovascular disease. These researchers have also associated caregiver stress with reduced immune cell ss2-receptor sensitivity and increased D-dimer marker of hypercoagulability. Roland von Känel and colleagues have also found that caregiver distress is associated with increased frailty, affecting interleukin-6, C-reactive protein, and D-dimer. Janice Kiecolt-Glaser and colleagues found that caregiver distress has a direct effect on immune function through increases in the proinflammatory cytokine interleukin-6. Taken together, these studies suggest that morbidity and mortality related to caregiving have a physical basis and that beyond mental health consequences, caregiving distress has significant and direct impacts on health.

HISTORY AND THEORETICAL ISSUES

Attention to the needs of informal caregivers has been increasing over the past three decades. This is primarily due to the changing demographics and the realization that caregiving to older adults with dementia carries substantial economic, emotional, physical, and relational tolls. Several theoretical models have been offered to account for the negative effects associated with caregiving. Most theoretical models explain caregivers' burden, depression, and ill health by proposing a relationship between primary stressors, such as the older adult's functional or cognitive impairment, and the caregiver's outcomes. This relationship is thought to be mediated through the amount of informal support provided by the caregiver and the use of formal support.

TECHNIQUES FOR PREVENTING AND MANAGING CAREGIVER DISTRESS

Intervention models are often based on the above-detailed conceptualization of caregiver distress, and many are multicomponent models that address several risk factors, including education about the dementing illness, emotional support, information about resources, and counseling in the form of skills training, anger management, and mood management. The most widely studied interventions are skills-training classes, educational interventions, and support groups. Interventions that are administered over a long period of time and target the entire family are considered particularly effective.

Skills-Training Classes

There is considerable evidence in support of the efficacy of structured classes aimed at educating caregivers in stress management, anger management, and self-care. These interventions typically are multicomponent in that they teach caregivers about the causes of dementia, explain the reasons for behaviors they will see in the care recipient, and provide them with strategies for managing their anger, mood, and burden. Most of these interventions are behaviorally focused; they teach caregivers how to refocus their understanding of the care recipient’s behavior, make use of stress-reduction strategies like meditation and relaxation training, increase pleasant activities, and find opportunities to take a break from their caregiving responsibilities. The success of these interventions is very well documented. According to Mary Mittelman and colleagues, interventions aimed at mediating the factors related to caregiver distress (coping, appraisals, support) have significant effects on distress even 4 years after treatment. Interventions aimed at anger management and burnout are also effective. According to David Coon and colleagues, anger management and skills-building interventions have positive impacts on caregiver appraisal and self-efficacy.

Skills-training classes appear to be effective across ethnic groups. According to one study, a multicomponent intervention that targets stress management and self-care strategies was better than education and support alone in Caucasian, Hispanic, and African-American caregivers. Caregivers with low-pretreatment self-efficacy seemed to respond better to these interventions than those with high self-efficacy, although, across groups, everyone experienced benefit from these skills-building groups. One of the largest studies of caregiver treatment is the REACH trial, a study of 167 caregiver–recipient dyads, which found that coping and stress management classes were far more effective in managing caregiver depression and burden than interventions focused solely on how to manage behavior related to dementia. These findings were robust for 2 years after caregivers participated in the coping skills interventions. The REACH trial has also been shown to improve self-reported health status.

Education and Support Groups

Research shows that even minimal education classes that focus on caregiver-coping strategies and information about dementia have benefits for caregivers in mitigating distress. In one study, a minimal information control condition was compared to a skills-building program for African-American and Caucasian caregivers. The Caucasian caregivers benefited from the information control condition better than the skills-building condition, indicating that even the provision of information about dementia and caregiving can have substantial benefits to caregivers. Robert Woods and colleagues compared a mental health nursing program to support caregivers to community educational programs and found that community educational programs were as effective in reducing caregiver distress as the more intensive nursing program. A recent review has suggested that psychotherapeutic techniques in conjunction with educational interventions are most effective.

CLINICAL ISSUES

Indications

Not everyone who cares for a PWD feels the experience is a negative one. Some people, while experiencing some stress, are able to cope well with the role of caregiver. Spouse caregivers are more likely to report depressive symptoms, greater financial and physical burden, and lower psychological well-being. This is only partially accounted for by the fact that spouse caregivers also provide more hours of care. Compared with men, women caregivers are more likely to report depressive symptoms. Those caring for older adults with frontotemporal dementia are more likely to report distress and strain compared with caregivers of older adults with AD. Other studies have shown that certain ethnic groups cope better with caregiving than others; for instance, Hispanic and Black caregivers are more likely to report high levels of burden compared with White and Asian caregivers. A number of caregiver characteristics seem to be related to whether or not the caregiver will be distressed by their responsibilities. The most notable and well-studied characteristics related to the development or prevention of caregiver distress are the caregiver's culture, coping style, social support, and formal support; the relationship between the caregiver and care recipient; and, finally, characteristics of the care recipient. While each of these factors contributes to distress, most researchers agree that these factors interact with one another, and no one factor alone contributes to caregiver distress.

Limitations

One of the most complicated aspects of treatment for distress is access to these effective interventions. Several studies show that caregivers rarely feel they have time to devote to their own well-being. A number of studies have been done to see if telemedicine-type support could benefit caregivers. The studies have largely been mixed, indicating that simple telephonic support does not result in as good outcomes as in-person group support. However, a study of 299 caregivers receiving a multimedia caregiver intervention over the Internet found that Internet interventions of this nature provided a low-cost, convenient, and effective means of overcoming caregiver distress, as well as being flexible enough to tailor the intervention to the needs of the caregivers. Further, web-based support combined with in-person caregiving groups appeared to have a substantial impact on depression in 225 caregivers. Finally, Sara Anne Smith and Paul Bell studied a CD-ROM plus printed material intervention for rural caregivers and found this intervention to result in greater use of available support services, use of behavioral interventions to manage care-recipient aberrant behavior, and reductions in distress. Research of this nature is beginning to show that access to effective psychosocial interventions is possible for caregivers. However, a recent systematic review found only low levels of support for the efficacy of Internet-based supportive interventions.

Mrs. Y was a 56-year-old Caucasian woman whose husband of 30 years was diagnosed with AD 3 years ago. Over the years, Mr. Y had become increasingly forgetful and impulsive, often leaving food on the stove to burn, walking to the corner store and becoming lost, and buying expensive items that the family did not need. In the last year, Mr. Y bought two new cars and three high-definition televisions, despite the fact that he already owned a 2-year-old car in good condition and the Y did not watch television. Since Mr. Y's dementia became more apparent, he had to stop working, and Mrs. Y, as a consequence, also had to cut back on her working hours to accommodate Mr. Y's needs. In addition to Mr. Y's forgetfulness and impulsivity, his sleep patterns also became disturbed, and he often woke at night and wandered around the house, which disturbed Mrs. Y's sleep.

Mr. and Mrs. Y had two children, both of whom lived in different states; thus, Mrs. Y had to cope with caregiving demands on her own. As a result of a part-time job and an increasing worry that leaving Mr. Y at home alone would result in either a fire in the home or more unnecessary excursions to car dealers and electronics stores, Mrs. Y became quite distressed and experienced several colds and cases of flu that she was not able to care for. One such illness resulted in a case of pneumonia, and it was at this point in time that Mrs. Y came to the attention of the clinical team.

Although Mr. Y had ongoing neurology appointments in a prestigious university hospital, it was clear that Mrs. Y had not made use of the resources available to her through this service. Her reason for not attending the support groups, working with the social worker to link her to respite care, and making use of community resources was her lack of time between taking care of Mr. Y and her work to help support them financially. While in the hospital, the team social worker visited Mrs. Y to talk to her about the importance of making use of available services and the need to care for herself first so that she could then better care for Mr. Y. From this discussion, it was clear that Mrs. Y had become depressed and burnt out. Not only was she suffering physically from her caregiving responsibilities, but she was also feeling less interested in usual activities, had lost her appetite, had begun feeling hopeless and helpless, had trouble concentrating at work, and wished she or her husband would die. This last symptom left her feeling incredibly guilty; she loved her husband, and although forgetful and impulsive, he remained a loving companion to her. She was also clearly resentful toward her children; when she entered the hospital, her daughter had reluctantly flown in from her home to care for Mr. Y and had made it clear she would only stay until Mrs. Y returned from the hospital.

A clinical team began to work on reducing Mrs. Y’s caregiver distress in the following way. First, the social worker began to make arrangements for Mr. Y to attend a day treatment program that would provide services within the Y’s financial means. Mr. Y’s schedule would be mapped to Mrs. Y’s work schedule so that she would not have to worry about Mr. Y being home alone. Temporary evening respite care was also arranged while Mrs. Y recovered from her pneumonia. This respite care was eventually reduced to 1 day a week so that Mrs. Y could have time for herself and attend a skills-building class for dementia caregivers. Mrs. Y was reluctant to join the class—in fact, she only attended intermittently; however, she did make use of the neurology department’s online caregiver support program in conjunction with the occasional classes and reported that she found the website helpful.

Six months after her hospitalization for pneumonia, Mrs. Y was feeling both physically and emotionally better. She still struggled with her husband’s increasing dementia but had become more active, re-engaged with friends, and was more productive at work. One year after her hospitalization, Mr. Y was moved to an assisted-living facility because of his rapidly deteriorating condition. Mrs. Y then joined a caregiver support group, as she missed her husband and was dealing with guilt related to assisted-living facility placement. This group was particularly useful to Mrs. Y in helping her see this placement as a necessary step in her and Mr. Y's health, gave her strategies for making her visits to Mr. Y pleasant, and reinforced her need to remain connected to her friends.

Goals of Interventions

Given the diverse nature of the various caregiving interventions, potential goals for treatment are also diverse. Goals may include increasing knowledge about dementia, improving the caregiver's coping skills with problem behaviors, reducing caregiver burden, depression or anxiety, increasing well-being, reducing the care recipient's problem behaviors, or increasing social support. The exact intervention goals should determine the intervention offered.

ETHICAL ISSUES

Caregivers of older adults with dementia face multiple ethical challenges. As the disease progresses, dementia caregivers have to make decisions about various treatment options, end-of-life care, and participation in drug trials. Some of these issues are likely to surface, during treatment, as a source of great distress and agony to some caregivers. A different ethical issue that should be addressed by researchers is the amount of burden put upon caregivers of PWD, who often accompany their loved ones during drug trials and, as a result, become secondary participants in dementia research.

RESEARCH AND EVALUATION

Successful interventions exist, with the majority of caregiving interventions including both a supportive and an educational component. Longer interventions tend to be more effective than shorter ones. More attention should be placed on clinically significant interventions rather than merely statistically significant ones. There is also increasing interest in establishing consensus on outcome measures for caregivers' interventions. In the past, access to interventions was complicated by the fact that caregivers, who rarely feel they have time to attend classes, were not likely to take the time to attend these sessions. Recent research suggests that interventions can be successfully delivered over the telephone and the Internet. However, these interventions may not meet the need for social support so often present among caregivers of older adults with dementia, and their efficacy should be further examined.

Further Readings

Alzheimer's Association. Alzheimer's Disease Facts And Figures. 2020. https://www.alz.org/media/Documents/alzheimers-facts-and-figures.pdf

Aneshensel CS, Botticello AL, Yamamoto-Mitani N. When caregiving ends: the course of depressive symptoms after bereavement. *J Health Soc Behav*. 2004;45(4):422–440.

Ayalon L, Green O. Live-in versus live-out home care in Israel: satisfaction with services and caregivers' outcomes. *Gerontologist*. 2015;55:628–642.

Balducci C, Mnich E, McKee KJ, et al. Negative impact and positive value in caregiving: validation of the COPE index in a six-country sample of carers. *Gerontologist*. 2008;48:276–286.

Bom J, Bakx P, Schut F, van Doorslaer E. The impact of informal caregiving for older adults on the health of various types of caregivers: a systematic review. *Gerontologist*. 2019;59(5):e629–e642.

*Boots LM, de Vugt ME, van Knippenberg RJ, Kempen GI, Verhey FR. A systematic review of Internet-based supportive interventions for caregivers of patients with dementia. *Int J Geriatr Psychiatry*. 2014;29(4):331–344.

Bremer P, Challis D, Hallberg IR, et al; RightTimePlaceCare Consortium. Informal and formal care: substitutes or complements in care for people with dementia? Empirical evidence for 8 European countries. *Health Policy*. 2017;121(6):613–622.

Cheng ST, Au A, Losada A, Thompson LW, Gallagher-Thompson D. Psychological interventions for dementia caregivers: what we have achieved, what we have learned. *Curr Psychiatry Rep*. 2019;21(7):59.

Egan KJ, Pinto-Bruno ÁC, Bighelli I, et al. Online training and support programs designed to improve mental health and reduce burden among caregivers of people with dementia: a systematic review. *J Am Med Dir Assoc*. 2018;19(3):200–206.

*Elliott AF, Burgio LD, Decoster J. Enhancing caregiver health: findings from the resources for enhancing Alzheimer's caregiver health II intervention. *J Am Geriatr Soc*. 2010;58(1):30–37.

Friedman EM, Shih RA, Langa KM, Hurd MD. US prevalence and predictors of informal caregiving for dementia. *Health Aff (Millwood)*. 2015;34(10):1637–1641.

Haley WE, Bergman EJ, Roth DL, McVie T, Gaugler JE, Mittelman MS. Long-term effects of bereavement and caregiver intervention on dementia caregiver depressive symptoms. *Gerontologist*. 2008;48:732–740.

Haley WE, Gitlin LN, Wisniewski SR, et al. Well-being, appraisal, and coping in African-American and Caucasian dementia caregivers: findings from the REACH study. *Aging Ment Health*. 2004;8(4):316–329.

Kiecolt-Glaser JK, Dura JR, Speicher CE, Trask OJ, Glaser R. Spousal caregivers of dementia victims: longitudinal changes in immunity and health. *Psychosom Med*. 1991;53:345–362.

Mahoney R, Regan C, Katona C, Livingston G. Anxiety and depression in family caregivers of people with Alzheimer disease: the LASER-AD study. *Am J Geriatr Psychiatry*. 2005;13(9):795–801.

Mausbach BT, Ancoli-Israel S, von Känel R, et al. Sleep disturbance, norepinephrine, and D-dimer are all related in elderly caregivers of people with Alzheimer disease. *Sleep*. 2006;29(10):1347–1352.

*Mausbach BT, Chattillion EA, Roepke SK, Patterson TL, Grant I. A comparison of psychosocial outcomes in elderly Alzheimer caregivers and noncaregivers. *Am J Geriatr Psychiatry*. 2013;21(1):5–13.

Mausbach BT, Dimsdale JE, Ziegler MG, et al. Depressive symptoms predict norepinephrine response to a psychological stressor task in Alzheimer’s caregivers. *Psychosom Med*. 2005;67(4):638–642.

Mausbach BT, Mills PJ, Patterson TL, et al. Stress-related reduction in personal mastery is associated with reduced immune cell beta2-adrenergic receptor sensitivity. *Int Psychogeriatr*. 2007;19(5):935–946.

Mittelman MS, Roth DL, Haley WE, Zarit SH. Effects of a caregiver intervention on negative caregiver appraisals of behavior problems in patients with Alzheimer’s disease: results of a randomized trial. *J Gerontol B Psychol Sci Soc Sci*. 2004;59:P27–P34.

Overton E, Appelbaum PS, Fisher SR, Dohan D, Roberts LW, Dunn LB. Alternative decision-makers’ perspectives on assent and dissent for dementia research. *Am J Geriatr Psychiatry*. 2013;21(4):346–354.

*Pinquart M, Sörensen S. Spouses, adult children, and children-in-law as caregivers of older adults: a meta-analytic comparison. *Psychol Aging*. 2011;26(1):1–14.

Riffin C, Van Ness PH, Wolff JL, Fried T. Family and other unpaid caregivers and older adults with and without dementia and disability. *J Am Geriatr Soc*. 2017;65(8):1821–1828.

Schulz R, Newsom J, Mittelmark M, Burton L, Hirsch C, Jackson S. Health effects of caregiving: the caregiver health effects study: an ancillary study of the Cardiovascular Health Study. *Ann Behav Med*. 1997;19:110–116.

Smith SA, Bell PA. Examining the effectiveness of the Savvy Caregiver Program among rural Colorado residents. *Rural Remote Health*. 2005;5:466.

*Sörensen S, Conwell Y. Issues in dementia caregiving: effects on mental and physical health, intervention strategies, and research needs. *Am J Geriatr Psychiatry*. 2011;19(6):491–496.

von Känel R, Dimsdale JE, Mills PJ, et al. Effect of Alzheimer caregiving stress and age on frailty markers interleukin-6, C-reactive protein, and D-dimer. *J Gerontol A Biol Sci Med Sci*. 2006;61(9):963–969.

Wong C, Merrilees J, Ketelle R, Barton C, Wallhagen M, Miller B. The experience of caregiving: differences between behavioral variant of frontotemporal dementia and Alzheimer disease. *Am J Geriatr Psychiatry*. 2012;20(8):724–728.

Woods RT, Wills W, Higginson IJ, Hobbins J, Whitby M. Support in the community for people with dementia and their carers: a comparative outcome study of specialist mental health service interventions. *Int J Geriatr Psychiatry*. 2003;18:298–307.

55.5 Financial Issues in the Delivery of Geriatric Psychiatric Care

Steve Koh, M.D., M.P.H., M.B.A., Helen H. Kyomen, M.D., M.S., Michael Hann, M.D., M.B.A., M.S., C.P.E., F.A.C.H.E., F.A.P.A., and Hilary Gould, Ph.D.

Mental health financing for the aging population is complex and compounded by multiple economic factors. While Medicare and Medicaid are cited as examples of a single-payer system in the United States, the costs of an aging population are increasing, and mental health care remains underfunded. This is further complicated by growing financial concerns for individuals who are facing not only rising health care costs but also higher living expenses during a time of decreased income and diminished savings. Despite the passage and implementation of legislation that ensured mental health parity—the Patient Protection and Affordable Care Act (PPACA), health care costs continue to be a significant percentage of total personal and governmental expenditures. One reason for this is that monitoring and enforcement of parity are inadequate and often left up to each state. Consequently, mental health care continues to be carved out from general health care coverage, and older patients may find themselves picking between certain medical treatments and daily living expenses. Additional financial burdens on households may emerge, including becoming caregivers to loved ones with chronic illnesses. Economic factors such as payment mechanisms and funding sources are key determinants of access to geriatric psychiatric care and health care delivery services and the actions of providers.

Since the 1990s, government, and industry have sharpened their focus on health care expenditures, and it has become clear that cost-effective health care cannot ignore comorbid mental illness. Now widely recognized as a cost multiplier when present as a comorbid condition, the need for improved mental health treatment to achieve better outcomes has been reported by multiple sources. The mental health treatments requiring different financial support can include those related to physician services, levels of care, collaborative care, substance use treatments, memory care centers, pharmacy costs, and preventative care. Ongoing attention to the Triple Aim of health care continues to drive innovation: improving the health of a population, increasing quality outcomes of individuals, and lowering the cost of care delivery. While still far from an ideal solution, the rise of accountable care organizations (ACO), financial models pertaining to value-based outcomes, and cost-effective treatment models serve to give hope for the future. These developments highlight the relationship between patients, providers, and larger systems crucial in providing optimal care for older adults.

DEMOGRAPHIC TRENDS AND THE COST OF HEALTH CARE FOR OLDER ADULTS

Rational planning for mental health care services for the future requires an understanding of the magnitude and diversity of the projected growth patterns in the older adult population in the United States. Furthermore, the differences between sexes and ethnic groups are important to note as there continue to be significant financial resource differences between them. In the mid-1950s, only 1 of every 11 Americans was older than the age of 65. In each of the last several decades, the proportion of older adults has grown substantially. By 2000, there were more than 35 million people aged 65 and older, constituting approximately 12% of the American population. Since then, the growth of the baby boomers has contributed to more than 52 million adults over the age of 65 in 2018. By 2034, for the first time in the history of this country, older adults are expected to outnumber children, accounting for approximately 20% of the total population. By 2050, it is projected that there will be 88.5 million individuals over the age of 65. This is not unique to the United States. The United Nations estimates that by 2050, 21% of the total global population will be above the age of 60. Additionally, the number of oldest-old adults (i.e., 85 years of age and above) continues to increase. It is expected that by 2040, 14.4 million Americans will be among the oldest.

The differences in the number of older men versus women will have important economic consequences. The economic effects of lower educational levels, traditional work roles, and income discrimination experienced by women in young and middle age will be exaggerated as retired women depend on pensions and social security derived from preretirement income. Currently, older women outnumber men by 29.1 million to 23.3 million. While in 2013, the ratio of older males to females was 76:100, by 2050, it is projected to increase to 87:100. This is due to continued increases in life expectancy for both sexes and improved health care. With increased longevity often comes greater health care needs. Many older adults need home care at some point, and it is estimated that even after liquidating all assets, only 40% of older adults with disabilities can afford just two years of home care. As the cost of care increases, the ability to keep older adults in their homes will diminish, and the expense will shift to governmental agencies. Additionally, elements that make up social determinants of health cannot be ignored. Research indicates that stable housing and home ownership have a direct correlation to self-rated health. Poor self-rated health ratings were directly related to higher health care costs and poor outcomes.

Similarly, ethnic minorities make up a growing proportion of the older-than-65 population. In 1990, approximately 11% of people older than 65 in the United States were of non-White ethnicity (non-European descent). In 2018, 23% of older Americans were members of a racial or ethnic minority, which is approximately 12.3 million individuals. This is expected to increase to 27.7 million by 2040. Currently, older African Americans make up 9% of the population, while Hispanics represent 8%. It is expected that older Hispanics will become the largest ethnic minority group by 2030. This change has been due to decreases in the rates of human immunodeficiency virus, heart disease (in women), homicide, and unintentional injuries. Still, many older African American individuals have considerably higher rates of poverty, illness, and violent deaths than Caucasian individuals in the same age group. The economic and social discrimination of the chronic underprivileged associated with American minority status is exaggerated by the socioeconomic consequences of older age. In this population, accrued social and financial resources are limited; access to preventive and acute health care is poor; and the likelihood of exposure to tobacco, alcohol, other toxins, and violence is higher. In addition, barriers to government services such as housing, transportation, meals, and income maintenance are more prevalent. All of this results in a cycle of medical, economic, and social poverty that affects the quality of life and has a substantial impact on the provision and financing of geriatric psychiatric care for older adult minorities. Further compounding this situation is the fact that there are significant state-to-state discrepancies in ethnic minority life expectancies. Currently, there are 3,428 medically underserved areas/populations in the United States. For mental health services, about 37% of the U.S. population is living in areas with mental health professional shortages. In these areas, the utilization of emergency and a higher level of care is high and adds to the financial cost of care.

Retirement from employment has important effects on economic status, access to health care, and insurance. Retirement reduces personal income and increases reliance on government assistance and retirement plans. Reduction in income after retirement threatens the stability of social class for older people and increases the likelihood that some people will experience economic poverty for the first time in their later life. These poverty prevalence rates are noticeably higher for women and minorities. In 2013, over 4.2 million older adults were below the poverty level based on income alone. If regional variants in cost of living, noncash benefits received, and nondiscretionary expenses are considered (the so-called Supplemental Poverty Measure from the U.S. Census Bureau), the poverty level of older persons reaches 14.6%. In 2019, these numbers increased to over 4.9 million people aged 65 and older living in poverty. The percentage of those living in poverty increased with age. For example, 11.1% of those above the age of 80 were in poverty in comparison to 8.4% of those aged 65 to 69. Additionally, the poverty rate was 18.0% among those identifying as Black or African American compared with 17.1% for those identifying as Hispanics, 9.3% for the Asian population, and 6.8% for the non-Hispanic White population. A major contributor to this increase is high medical out-of-pocket expenses. This cost is magnified in geriatric mental health care and in long-term care, where out-of-pocket copayments, deductibles, and direct purchase of services account for a substantial component of total costs.

As long-term financial stability is questioned, many individuals are electing to delay retirement or continue to work part-time. Since the Great Recession (spanning from 2007 to 2009), the average retirement age has increased. Twenty percent of Americans continue to work past 65 years of age, compared with 10% in 1985. Even though median income typically peaks between the ages of 45 and 54, in more recent years, those above age 65 saw percentile increases in their median income. This indicates that more Americans are working into their 60s, particularly those with higher education and higher earnings. Yet, many indicate that there is added pressure (related to changing work environment and ageism) to retire despite their worry about long-term security. The single most important factor for concern about retirement is health care cost. Retirement directly affects health insurance premiums, the availability of certain insurance plans, and access to some health care delivery systems.

Overall fiscal impact on health care by the older population is undeniable. While their earnings are decreasing, the health care needs of the growing older population continue to rise. The cost of health care in the United States is far greater than in other industrialized nations. Older adults make up a significant proportion of this high cost. Almost half of all health care spending is consumed by just 5% of the population. In fact, older adults account for more than 36% of total health care costs. Those between 65 and 70 years of age represent 29%, and those above the age of 80 account for 14% of the top 5% of spenders. In general, medical spending doubles between ages 70 and 90. Among those with Medicare–Medicaid coverage, the older adults and disabled account for more than 70% of Medicaid spending despite representing just 25% of the Medicaid population. The top five most expensive health conditions to treat are heart disease, cancer, trauma, mental disorders, and pulmonary conditions. Twenty-five percent of the U.S. population reportedly has one of the top five chronic conditions (i.e., mood disorder, diabetes, heart disease, asthma, and hypertension). The rise in Medicare and Medicaid costs can be explained partially by the aging population and mental illness.

HISTORICAL BACKGROUND

Historically, government policy has been important in shaping the status of psychiatric care for older people. During the 19th century, financial responsibility for the care of the chronically mentally ill and people with major neurocognitive disorders was transferred from local to state authorities, resulting in the replacement of asylums and almshouses by the massive state hospital systems of the first part of the 20th century. Ultimately, deinstitutionalization in areas of limited community resources resulted in the transfer of a substantial number of older people with severe and persistent mental illnesses from state hospitals to community nursing homes with sparse psychiatric resources. Beginning in approximately 1965, reimbursement for nursing home care under Medicaid effectively shifted some of the financial responsibility for this population from states alone to the states and federal government.

The introduction of Medicare and Medicaid in the mid-1960s created a private market for older adult health care. Private practitioners and institutions were reimbursed fee-for-service at market prices by these insurance programs. Adequate financial returns for hospital-based services, reinforced by the growth of commercial indemnity insurance, drove the development of for-profit hospital care, particularly in freestanding psychiatry. Simultaneously, severe limitations in reimbursement for ambulatory psychiatric services under Medicare impeded the growth of private practice interest in geriatric psychiatric care while creating imbalanced incentives for the use of hospital resources. Federal-level sensitivity to these new expenditures and social concerns about inappropriate care led to the nursing home reform provisions of the latter part of the 1980s and to the formulation and implementation of a set of guidelines for the use of psychiatric interventions in that environment. The health care cost reform of the last generation has similarly reshaped geriatric psychiatric service delivery. In the recent past, Medicare was more attractive than some commercial payers to hospitals and individual providers. Nonpsychiatrist mental health specialists, newly eligible for participation in Medicare and other payment programs, created the availability of a greater diversity of providers, with psychotherapeutic services much more broadly available to older adults.

The passage of PPACA has led to high hopes for better health care coverage and cost savings. For older adults, it is not clear how remarkable financial improvements have been. In part, this is because the application of PPACA has been inconsistent across the states, and it did not significantly overhaul Medicare services for older adults. Access to mental health care seems to have improved with more funding for Federally Qualified Health Centers and federally funded programs like the Program of All-Inclusive Care for the Elderly (PACE), as well as the increased emphasis placed on mental health care services. The Medicaid expansion to cover more Americans at or below the poverty line has increased preventative care that has improved the lives of those entering older age. However, inequality in mental health care coverage continues due to geriatric provider workforce shortages, continued high out-of-pocket costs of mental health care, and poor access to geriatric psychiatric care. It is hoped that the ongoing rise in health care costs and the significant impact of comorbid mental illness will lead to more attention to innovative care models that will emphasize outcomes and overall value to the older population’s health.

MEDICARE SYSTEM

The Medicare program was established in July of 1965 to provide medical care benefits for older people participating in the Social Security program. Since then, a number of modifications have been made to the Medicare program, in both the Hospital Insurance program, Part A, and the supplemental Medical Insurance program, Part B, and through the development of Medicare Part C and Part D. The Medicare Prescription Drug, Improvement, and Modernization Act of 2003 provided the most extensive changes to Medicare since its initiation. In addition, the 2010 PPACA has contributed to fundamental changes in overall health care delivery, including Medicare.

In 1972, Medicare was expanded to cover younger people with disabilities and older adults outside of social security who were willing to pay a premium. In 1973, the program started to provide medical coverage for people with end-stage renal disease. In 1997, the Balanced Budget Amendment of 1997 (BBA 1997) established Medicare Part C, then known as Medicare+Choice. As part of Medicare Part C, the BBA authorized the Centers for Medicare and Medicaid Services (CMS, the replacement agency for the Health Care Finance Administration) to offer various health care options. Medicare Part C includes Medicare managed care plans (such as health maintenance organizations [HMOs], provider-sponsored organizations [PSOs], and preferred provider organizations [PPOs]); Medicare private, unrestricted fee-for-service plans, Special Needs Plans (SNP), Religious Fraternal Benefit (RFB) plans, and medical savings account plans.

In 2003, Medicare was extensively updated to include:

- Expansion and renaming of the Medicare+Choice program as Medicare Advantage
- Prescription drug programs and benefits, including a new program called “Medicare Part D”
- Medication therapy management programs for those with multiple chronic conditions, multiple prescriptions, and likely high drug

expenses
- Coverage of certain screening and disease management programs such as initial physical examinations and cardiovascular disease and diabetes screening
- Changes in reimbursements to hospitals and physicians, including ongoing refinement of the prospective payment system (PPS) for inpatient psychiatric care
- Improved rural health care provisions
- A demonstration project for recovery audits to identify underpayments or overpayments and collect overpayments, and
- Changes in the acquisition systems for durable medical equipment and other products to provide for a competitive bidding process

Medicare was designed to provide coverage for the treatment of acute conditions, and this has been the primary orientation of the program over the years. Medicare has not covered most preventive and pharmaceutical services since the program's inception. The Medicare Prescription Drug, Improvement, and Modernization Act of 2003 provided for some change in this focus with coverage of some preventive services and the addition of pharmaceutical benefits. Long-term care and dental services continue to be excluded from traditional indemnity Medicare. Other services, including mental health care, had been subject to limitations in payment and substantial copayments. The Mental Health Parity Act of 1996 (MHPA) ensured that large-group health plans could not set annual or lifetime dollar limits on mental health benefits, which are less favorable than any such limits for medical or surgical benefits. The Wellstone-Domenici Mental Health Parity and Addiction Equity Act (MHPAEA) of 2008, enacted into law on October 3, 2008, and effective for most health insurance plans on January 1, 2010, added additional protections. This law required that group health plans (with over 50 employees that offer both medical/surgical and mental health/substance use benefits) provide coverage for treatments of mental health/substance use disorders that is no more restrictive than the coverage for treatments of medical/surgical conditions.

Equity coverage is applied to financial requirements as well as treatment limitations. However, the MHPAEA does not require large group health plans or health insurance issuers to offer mental health or substance use disorder benefits.

In 2020, Medicare provided insurance coverage for 34 million older adults (97% of older adults in America). Growth in the older adult population, advances in medical technology, more extensive use of pharmaceutical treatments, and other increases in the cost of health care have contributed to the extraordinary growth in health care expenditures, both in the Medicare program and otherwise, that has occurred since the mid-1990s. Although health care costs have continued to grow, especially among older adults, various efforts to contain Medicare costs have been successful in slowing health care inflation rates. Each of these changes has profoundly affected access to appropriate psychiatric care for older people, and each has impacted the scope and configuration of geriatric mental health specialty care. Aggregate Medicare cost reductions have resulted from decreases in reimbursement levels of hospital and physician providers, increased requirements for out-of-pocket payments by Medicare beneficiaries, and substantial shifts from indemnity insurance to managed care. In addition, as part of the Deficit Reduction Act of 2005, CMS has been allowed since 2008 to begin adjusting payments for “never events.” Never events are “serious and costly errors in the provision of health care services that should never happen” and are characterized by “unambiguous,” “usually preventable,” and “serious” events that are “adverse” and/or indicate “a problem in a health care facility’s safety systems” and/or are “important for public credibility or public accountability.” Examples of never events include hospital-acquired severe decubiti, death, or serious disability related to a medication error, and death associated with falling while being cared for in a health care facility. Never events lead to increased health care costs due to the need to treat the consequences of the error. Additional Medicare cost reduction has been attributed to the PPACA of 2010 and sequestration, which started in 2013. Still, Medicare spending represented 20% of national health spending in 2013 and grew 3.4% to about $586 billion, which was a slowdown from a growth of 4.0% in 2012. This slowdown was thought to be due largely to slower enrollment growth and the impacts of the PPACA and sequestration. In 2020, Medicare spending grew 3.5% to $829.5 billion, about 20% of the total National Health Expenditure. From 2019 to 2028, Medicare is expected to have a spending growth of 7.6% per year due to projected high enrollment.

The Medicare program has several components: Hospital Insurance, or Part A, covers inpatient hospital care, skilled nursing facility care, limited home care, and hospice services. Part B, the supplemental Medical Insurance program, covers preventative care, services provided by physicians and other professional providers, durable medical equipment, ambulatory medical laboratory tests, outpatient rehabilitation care, and group therapy services in psychiatric hospitals. Part C includes Medicare managed care plans (such as HMOs, PSOs, PPOs, Medicare private, unrestricted fee-for-service plans, SNP, RFB plans, and medical savings accounts). Under Medicare Part D, Medicare beneficiaries have been eligible to obtain prescription drug coverage by enrolling in a private prescription drug plan or in an HMO or other plan that accepts Medicare capitation payment and offers a drug benefit.

Hospital Insurance (Part A)

Americans older than age 65 who participate in the Social Security program are entitled to Part A Medicare coverage. The Part A Hospital Insurance Trust Fund is supported by social security payroll taxes paid by employers, employees, and the self-employed. In 2020, over 39 million people had Medicare coverage, and at the end of 2020, the Hospital Insurance Trust Fund had assets of $134 billion.

Medicare coverage for inpatient care is limited to relatively brief inpatient stays, presumably for the stabilization of acute symptoms. Part A provides coverage for episodes, which are time-limited “spells” of illness. A spell is an episode of inpatient care that starts with hospital admission and ends after the first period of 60 consecutive days outside the hospital after hospital discharge. If a person is discharged and readmitted numerous times during a given episode of illness and has not been out of hospital for 60 or more consecutive days, they are considered to have remained within the same spell. For treatment of any condition in a general hospital, there is no limit on the number of spells or days covered within an individual's lifetime. However, no more than 150 inpatient days are covered during each episode.

The first 60 days of each spell are fully paid after a deductible equal to the national average cost of 1 day of hospitalization has been met. The next 30 inpatient days are paid at the same rate, less a 25% copayment. The last 60 days of covered hospitalization require a 50% copayment. Days 91 to 150 are identified as "lifetime reserve days." Medicare pays for those days during any single spell of illness but only once in a beneficiary's lifetime. For example, if a patient requires five hospitalizations totaling 110 days without a period of discharge of at least 60 days, they would have 40 days of remaining coverage in their lifetime for future episodes that last more than 90 days. Individuals may save these days for future hospitalizations and use other resources to pay for any part of the costs of days 91 to 150.

At the time that Medicare was first signed into law, there were large populations of older adult patients in state psychiatric hospitals. Federal legislators feared at that time that the enactment of Medicare would shift the burden of care for that population from state-supported resources to acute care facilities reimbursed with federal monies. Consequently, Medicare coverage for inpatient treatment in freestanding psychiatric hospitals was limited to a total of 190 days over the course of a person's lifetime. The law provides that if an individual becomes eligible for Medicare during a psychiatric hospitalization, the Medicare intermediary may deem that less than the full 150 days allocated for that spell of illness will be covered. Given improvements in treatment and increased life expectancy, this limitation can easily create undue hardship and compromise continuity of care. For example, a person with a recurrent major affective disorder who experiences several depressive episodes requiring inpatient care after age 65 could easily exhaust the lifetime freestanding psychiatric hospital benefit.

Part A provides coverage for nonacute services, including skilled nursing facility care, if these services are used briefly and for “remediable” conditions. For a beneficiary to obtain benefits for skilled nursing facility care, they must have been hospitalized for at least 3 consecutive days for an acute condition, and skilled nursing facility admission must follow within 30 days of hospital discharge. Medicare may fully cover up to 20 days and partially cover days 21 to 100 of skilled nursing facility care for each episode of illness. Similarly, coverage for home health services is limited to care for acute and remediable conditions.

From 1966 until 1983, Medicare Part A paid for hospital care using a retrospective cost-based reimbursement system. This system compensated providers on a fee-for-service basis for all those costs deemed allowable under Medicare without an incentive for utilization management. In 1982, after years of rapid and sustained growth in Medicare expenditures for inpatient care, the Reagan administration signed into law the Tax Equity and Fiscal Responsibility Act (TEFRA). TEFRA created incentives to reduce hospital payments by setting targets for operating costs, and it provided supplemental payments for compliance with specified objectives. Reimbursement was transitioned from a retrospective cost-based approach to a payment system designed to reflect patient mix, geographic location, and training costs. This TEFRA-mandated system is known as the PPS.

The PPS established by TEFRA uses diagnosis-related groups (DRGs) to classify patients who require similar care into major diagnostic categories and assigns each Medicare patient discharge to a diagnostic group. These groups are derived from principal and secondary diagnoses, procedures rendered, and, to a lesser degree, age, sex, complications, and discharge status. The DRG system was revolutionary in creating a “case-based” approach to reimbursement. Under the PPS, hospitals are paid a fixed amount for each case according to the corresponding DRG, regardless of the resources that are consumed during the hospitalization. If patients require extraordinary resources or extremely prolonged hospital stays, Medicare classifies the patient as an “outlier” and provides an additional payment to the hospital, although at a per diem rate considerably below actual costs. This set of conditions creates an incentive for hospitals to provide efficient care and create an infrastructure of ambulatory services that can shorten or substitute for hospitalization.

Several studies have indicated that, although DRGs are modestly effective in predicting use for medical and surgical diagnoses, their ability to predict use for psychiatric diagnoses is poor. Accordingly, the Health Care Finance Administration, the federal agency responsible at that time for Medicare/Medicaid fiscal management, exempted freestanding psychiatric hospitals and distinct psychiatric units in general hospitals from the DRG system until a fairer system could be ascertained. In 2004, a newly designed Medicare PPS for inpatient psychiatric services with updated psychiatric DRGs was established by the CMS. Under this new system, a federal per diem rate was established, providing for adjustments for changes in the hospital market basket index, age, comorbidities, and psychiatric diagnosis. In this system, 15 psychiatry-related DRGs were established, with weights ranging from 22% above average for DRG 876 (“Operating Room Procedure with Principal Diagnosis of Mental Illness”) to 12% below average for DRG 897 (“Alcohol/Drug Abuse or Dependence without Rehabilitation Therapy and without Complications and Comorbidities”). Currently, there are 12 psychiatry-related DRGs (i.e., DRG Codes 876–887; 896, 897).

Until the new PPS for inpatient psychiatric services was initiated, freestanding psychiatric hospitals and psychiatric units in general hospitals were reimbursed retrospectively by Medicare. Ever since TEFRA was implemented in 1983, payment for treatment in psychiatric units had been capped at a rate established for each facility based on use during its first fiscal year of operation after October 1, 1983. Until 1998, if the actual cost of care per case was greater than the capped rate, the hospital absorbed the loss. However, if the cost was less than the capped rate, a small incentive payment was retained by the hospital. The Balanced Budget Act

(BBA) of 1997 modified and limited these reimbursements by establishing differential caps, with or without wage adjustment. It also provided less generous bonuses and relief payments for PPS-exempt facilities. The Balanced Budget Refinement Act (BBRA) of 1999 addressed some of the flawed policies and excessive payment reductions resulting from BBA 1997. The BBRA of 1999 allowed for wage adjustments at the 75th percentile limit established by the BBA of 1997 for PPS-exempt facilities that received their first Medicare payment before October 1, 1997, and increased bonus payments for psychiatric hospitals. The BBRA of 1999 also stipulated that a new Medicare PPS for inpatient psychiatric services was to be initiated by April 1, 2004. The final rule establishing the PPS for inpatient psychiatric facilities was published on November 15, 2004, in the Federal Register. Since then, CMS has made periodic updates to the inpatient psychiatric facilities PPS. Under the Code of Federal Regulations 42 CFR 412.428, annual updates include revisions to the federal per diem base rate, hospital wage index, International Classification of Diseases (ICD) coding and DRG classifications, electroconvulsive therapy (ECT) payment per treatment, fixed dollar loss threshold amount and national cost-to-change medians and ceilings in urban and rural areas.

Supplemental Medical Insurance (Part B)

Medicare Part B is a medical insurance program that covers preventative care, services provided by physicians and other professional providers, durable medical equipment, ambulatory medical laboratory tests, outpatient rehabilitation care, and group therapy services in psychiatric hospitals. Medicare Part B is a voluntary program and requires a monthly premium payment for participation. Approximately 93% of Medicare Part A beneficiaries also participate in the Part B program. Part B is supported by the Supplementary Medical Insurance Trust Fund, which includes general revenues from the Treasury, as well as premiums. In 2022, the standard monthly premium for Medicare Part B is $170.10, and the annual deductible is $233. The Medicare Part B premium, deductible, and coinsurance rates are determined each year according to the Social Security Act.

Since the enactment of the MHPAEA, group health plans and health insurance issuers are required to ensure that financial requirements (such as copays and deductibles) and treatment limitations (such as visit limits) applicable to mental health or substance use disorder benefits are no more restrictive than the predominant requirements or limitations applied to substantially all medical/surgical benefits. The MHPAEA supplements prior provisions under the MHPA, which required parity with respect to annual and aggregate lifetime dollar limits for mental health benefits.

From 1984, after TEFRA and the PPS were initiated, until 1992, the rate of growth of Medicare physician costs was more than double the rate of growth for inpatient hospital costs. Policy analysts attributed this growth to uncontrolled price escalation and substantial increases in the number and intensity of services consumed. In 1989, inflation in physician-related costs and mounting evidence that physicians’ behaviors in treating complex medical illnesses were influenced by the high prices of the procedures they performed led to a sweeping reform of Medicare Part B through the Omnibus Budget Reconciliation Act of 1989 (OBRA-89).

OBRA-89 created standards for expenditures and procedure volumes to bring the total cost of physicians’ services under control. The law required the secretary of the United States Department of Health and Human Services (HHS) to identify specific volume performance standard growth rates for surgical, primary care, and nonsurgical nonprimary care services at the beginning of each fiscal year. These standards must be related to medical inflation and changes in the size of the Medicare population, as well as to changes in service volume and intensity. At the end of each year, the secretary must compare actual growth and expenditures with this volume performance standard and determine any changes in fee schedules for the following year. OBRA-89 also created a national uniform Medicare fee schedule to replace local methods that were traditionally used by intermediaries to set fees for each procedure and for each provider. In 1992, a Resource Based Relative Value

Scale (RBRVS) was developed to determine the relative values of approximately 7,000 physician procedures and services. In this scale, the relative value of each of these services is derived from three components: work, practice, and malpractice expenses, and the opportunity costs of postgraduate specialty training required. The proportion of each component in determining the relative value is based on a weighted average of specialty-specific practice expense and malpractice data. To change provider behavior and reduce incentives to perform highly reimbursed technical procedures, the work component is derived from estimates of the time intensity and stress associated with the delivery of the procedure. Approximately 60% of the value of each fee is adjusted for local costs. The product of each relative value and a monetary conversion factor determine the fee for each service.

OBRA-89 was a landmark for behavioral health in its provision for direct reimbursement of services provided by psychologists and social workers. It also sets criteria for nonphysician providers regarding consultation with a physician. Nonphysician providers must document that a patient has been informed that consultation with the patient's primary care physician (PCP) is desirable for the PCP to evaluate medical conditions that could contribute to a patient's well-being. The provider must also document any written or verbal communications with the PCP about treatment unless the patient specifically prohibits such contact. The law does not require consultation with a psychiatrist. The law stipulates that reimbursement of services provided by psychiatric nurses and other therapists requires direct physician supervision.

In general, direct patient contact is required for any reimbursable service. Consultation or other services rendered by telephone and renewals of prescriptions that do not include direct evaluation of a clinical condition have not been reimbursable. Medicare Part B allows billing for services required to obtain clinical information from relatives or close associates of unreliable or uncommunicative patients. Family counseling services are reimbursable only when their purpose is to facilitate patient care. With the direct in-person care limitations imposed by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and SARS-CoV-2 variant outbreaks starting in 2019, there has been more support for the reimbursement of telehealth services, especially on a state-by-state basis.

There are two ways in which providers may participate in Part B of the Medicare program:

1. A provider may accept the assignment and agree to accept the applicable Medicare fee as payment in full. Medicare's fiscal intermediary pays the provider the approved amount less copayment and deductible directly. The provider is legally required to bill the patient directly for the copayment and the deductible.
2. A provider may choose not to accept an assignment. Providers who do not accept assignments may charge patients up to 115% of the approved Medicare fee. Medicare reimbursement for providers who do not participate in Medicare Part B is limited to 95% of the approved fee, with fewer deductibles and copayments. The provider may then bill the patient directly for the balance of charges up to the 115% limit.

All providers must accept assignments from Medicare beneficiaries who also participate in the Medicaid program. The assignment must be accepted for all Medicare recipients who are at or below the federal poverty level, regardless of whether they have Medicaid as a secondary payer. In an increasing number of states, medical licensure authorities and insurance commissions require that all physicians who participate in the Medicare program accept assignments for patients seen within that state.

Medicare Risk in Managed Care

Since the early 1970s, Medicare has encouraged older Americans to enroll in capitated health plans. Since that time, HMOs have been able to contract directly with Medicare. Between 1985 and 1997, Medicare payments to risk-contracting HMOs for aged and disabled beneficiaries were based on adjusted actuarial estimates of the per capita cost of such services incurred in a beneficiary's county of residence, using adjusted average per capita cost (AAPCC) methods. The AAPCC had been based on the individual's county of residence, age, and sex and whether a patient is covered by Medicaid, lives at home or in an institution, and works or is retired or disabled. The county per capita payment rates are adjusted for the average beneficiary and have been published each year as the County Rate Book.

HMOs, although allowed to earn "normal profit margins," must use savings above a predetermined rate to provide extra services for older adult members. For many years, to penetrate a local market, HMOs frequently waived premiums and offered to include usually uncovered services in the benefits package to induce enrollment. Unlike traditional Medicare, HMOs allow access to primary care and other preventive services in the absence of a primary diagnosis. In addition, managed care plans usually provide coverage for some portion of outpatient pharmaceutical costs.

The nature of HMO risk-contracting with Medicare provides an incentive for HMOs to market to healthy older adult populations who are likely to consume services valued at considerably less than the 95% of AAPCC criterion used to set capitation rates. Indeed, healthy populations, even if older than 65 years, yield substantially greater profit margins for HMOs, leaving sicker, more disabled people to be cared for via direct payments from CMS in the traditional Medicare program. Several studies have indicated that capitated plans generally enroll healthier and less costly beneficiary populations than those who remain in fee-for-service Medicare. In one study, in the 6 months preceding enrollment, new members of capitated plans had health care costs that were 35% lower than the average costs of patients in traditional fee-for-service Medicare. Some plans profit through marketing to "younger older adults" (people aged 65 to 70) in settings in which there is a greater likelihood of health and independence. Studies indicate that Medicare has overpaid some capitated plans by 5% to 20% because selection factors resulted in enrollees with better health, on average.

To reduce financial incentives rewarding the enrollment of healthy persons and to pay more fairly for the care of frailer populations, the

BBA 1997 required Medicare to link payments to HMOs to the health status of their beneficiaries by the beginning of the year 2000. In this system, frail enrollees are eligible for higher payments than healthier cohorts are. This encouraged plans to discontinue avoiding the enrollment of more chronically ill populations and to bring a broader base of the general older adult population into managed care. In enacting Medicare Part C in 1997 to create Medicare+Choice, a new methodology was introduced to pay organizations that enter Medicare Part C contracts. The equivalent of the County Rate Book is based on the greatest of three amounts: (1) a new blended payment rate based on a calculation method combining local and national rates in setting county rates, (2) a new minimum specified rate amount, or (3) an amount that is based on a 2% increase over the prior year’s rates, with the 1997 County Rate Book serving as the baseline.

The percentage of beneficiaries enrolled in Medicare managed care increased steadily into the 1990s. Whereas only 3.6% of older Americans were enrolled in HMOs in 1990, this proportion increased to 10% in 1996 and to more than 15% in 1998. Since 2001, enrollment has been declining due to the departure of many health insurance programs from the Medicare managed care market due to reimbursement rates that are increasingly inadequate to cover medical utilization costs. The Medicare+Choice program, a major component of the BBA 1997, renamed in 2004 to Medicare Advantage, is the Medicare managed care component. It is designed to make it easier for various kinds of organizations, including PSOs and PPOs, to become participants in Medicare managed care. Medicare Advantage has been organized to encourage greater competition among health plans enrolling Medicare beneficiaries.

Managed care plans caring for Medicare recipients have applied the same paradigms to behavioral health care that they have used for their younger commercial populations. Many of these plans, perhaps a majority, “carve out” the behavioral health benefit and capitate the risk for these services to managed care organization (MCO) vendors. Most of the MCOs are large for-profit entities that function as psychiatric and substance use insurers in numerous markets regionally and nationally. MCOs generally establish local networks of ambulatory and inpatient providers; establish fee-for-service pricing with those providers; provide medical management, care management, utilization review, and credentialing functions; and pay claims. These MCO systems are generally very effective in reducing inpatient use. This effectiveness allows them to negotiate significantly discounted prices for services and to limit the size of available networks, thereby reducing the management effort required to control service use.

A review of the credentialing and procedure manuals for a group of the largest MCOs nationally found virtually no mention of requirements for specialty training or expertise in the care of older adults or for special consideration of medical comorbidity or frailty. There were no requirements for services to be sited at locations in which older adults reside or where access to care may be facilitated for frail or disabled people. This review suggests that the current model for delivery of managed behavioral health services to older adult Medicare beneficiaries is unlikely to reflect recent advances in behavioral neuroscience. Similarly, many carve-out programs exclude or limit comprehensive assessment, neuropsychological testing, and other routine components of major neurocognitive disorder evaluation.

Managed care continues to be important in Medicare policy. However, reductions in payments to managed care entities by Medicare coupled with complex paperwork requirements have prompted many managed care plans to leave the Medicare program. The 2010 Affordable Care Act provides for health care reform with additional opportunities for geriatric mental health care through ACO, patient-centered medical homes, and other structural changes in health care delivery.

Insurance Coverage Supplementing Medicare

Around 90% of older Medicare beneficiaries have supplemental Medicare coverage for service costs not covered by Medicare directly. These supplemental coverages include employer or union-sponsored retiree health plans, Medigap policies, Medicare+Choice/Medicare Advantage, and dual Medicaid–Medicare coverage.

In July 1991, Congress enacted mandatory federal Medigap standardization requirements as part of the OBRA of 1990 (OBRA-90). It provided requirements for 10 basic policies. All these policies must include payment of the Medicare Part A hospital and Part B provider coinsurances. In addition, all policies must pay any applicable 50% copayment for outpatient psychiatric services. More comprehensive policies may elect to cover deductibles, skilled nursing facility coinsurance, outpatient prescription pharmaceuticals, and preventive in-home services. Medigap regulations provide incentives for the use of PPOs by Medicare recipients. This allows insurers to offer lower premiums to members who agree to use a PPO for all their care. An increasing number of states have authorized insurers to offer this option.

The Medicare Prescription Drug, Improvement, and Modernization Act of 2003 impacts Medigap policies that have a drug benefit. People currently enrolled in a Medigap plan without a drug benefit are not affected. However, those who have a Medigap policy with drug benefits have two choices: (1) They may keep their current Medigap plan with drug benefits. However, if they later decide to enroll in a Medicare Part D plan, a late enrollment penalty may be charged if the Medigap policy is not equivalent to Medicare Part D. (2) They may enroll in a Medicare Part D plan. Those who choose to do this can either enroll in another Medigap plan without drug coverage to cover Medicare's cost-sharing for other services or keep their current Medigap plan but drop its drug benefits and pay a lower premium. People who reach the age of 65 after January 1, 2006, can enroll in Medigap policies without drug benefits. After January 1, 2006, people were unable to purchase a supplemental policy to help offset their deductibles and copayment for drug policies. Regarding employer-sponsored retiree plans, the government pays 28% of the costs that companies incur for each retiree to encourage companies to maintain their drug coverage for retirees; this amounted to between $480 and $9,850 per retiree in 2021.

On April 16, 2015, the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) was authorized. Among various actions, this replaced the Sustainable Growth Rate formula (never actually put into action due to multiple short-term Congressional “patches” since 2002). Starting in 2020, it affected insurance coverage supplementing Medicare by decreasing funding to supplemental Medicare plans, reducing “first dollar” coverage of supplemental Medigap plans so that individuals new to Medicare could not purchase Medigap plans that pay for the Part B deductible.

MEDICAID SYSTEM

Enacted in 1965 as a social insurance program to pay for the medical care of Americans with limited financial means, Medicaid is funded through a combination of federal and state matching funds. To receive these matching funds, states must provide a basic benefits package, including hospital, physician, and nursing facility services, as well as a wide range of supportive services. The program allows states to impose restrictions on the services that they fund and on the level of reimbursement. Medicaid is a means-tested and needs-based program; thus, an individual must meet certain criteria to be eligible. The eligibility varies widely from state to state, but the most important criteria are poverty and financial resources. Medicaid was expanded upon PPACA being signed into law to include all adults under the age of 65 with income up to 133% of the poverty line. Although initially a legislative requirement, mandatory Medicaid expansion was blocked by the Supreme Court of the United States (SCOTUS) in its June 2012 *National Federation of Independent Business (NFIB) v. Sebelius* decision, effectively allowing states to voluntarily participate in or decline Medicaid expansion. As of August 2021, 38 states and the District of Columbia have elected to participate.

Medicaid expansion contributed to increases in insurance coverage overall, as well as enrollment in the Medicaid program, regardless of a state’s participation status in Medicaid expansion. In states that elected to expand Medicaid, enrollment increased by 13 million beneficiaries by 2020, approximately 81.2% of whom were newly eligible individuals and entitled to the newly eligible enhanced federal matching rate. The increase in coverage observed in states that declined Medicaid expansion is likely due to a “welcome-mat” effect, in which individuals who were previously eligible but not enrolled may have subsequently elected to enroll due to the increased outreach and marketing associated with Medicaid expansion. Insurance coverage for individuals with serious psychological distress (SPD) or probable depression (PD) in Medicaid expansion states increased at an even greater rate than the general population. Medicaid expansion has been associated with increased access to health services, increased quality of medical care delivered, and reduced mortality. Initial data from 2020 to 2021 indicate that Medicaid enrollment has also increased during the COVID-19 global pandemic, during which states have been required to provide continuous coverage as a condition of increased federal funding provided under the declared public health emergency. In addition to increasing health care coverage for millions of Americans, Medicaid expansion has also reduced out-of-pocket spending on mental health services for low-income adults, particularly for adults with psychological distress (SPD). However, despite a reduction in out-of-pocket costs, Medicaid expansion does not appear to have increased the use of mental health services.

For older Americans with mental illness, Medicaid is important in two ways. First, it is the most important payer for long-term care services. The nursing home is an important site of care for older adults with severe and persistent mental illnesses, including major neurocognitive disorder, schizophrenia, and depression associated with medical illness and disability. Deinstitutionalization from state facilities shifted the financial burden of care for many people with these diagnoses from the states (in state hospitals) to cost sharing with the federal government via its partial support of Medicaid. The nursing home reform provisions of the OBRA of 1987 (OBRA-87) require preadmission screening for nursing home applicants to determine if an active psychiatric illness is present. If active treatment for an identified psychiatric disorder cannot be provided appropriately in the nursing home, alternative referral for this care is mandated. Nursing homes must demonstrate efforts to provide appropriate care for behaviorally disturbed people, and appropriate training in the management of difficult behaviors is required of all staff members of long-term care facilities participating in the Medicaid program.

Currently, Medicaid continues to be the major coverage option for long-term services and supports (LTSS). Medicaid finances approximately half of all long-term care in the United States, and nearly two-thirds of all Medicaid spending focuses on older adults and adults with disability. Medicaid's scope is broader than Medicare and encompasses services delivered in hospitals, institutions, homes, and communities. As Americans continue to age, it is estimated that over 70% of those above the age of 65 will need long-term services, and those above the age of 85 are four times more likely to need long-term services. In 2013, the most recent year with comprehensive data for both programs, 51% of the estimated cost of $310 billion for long-term services was covered by Medicaid (19% was out-of-pocket). To best manage these health system costs, state Medicaid agencies are increasingly turning to managed care to provide LTSS, resulting in more than a doubling of spending on managed long-term services and supports (MLTSS) from 2012 to 2015. As of 2017, 41 different MLTSS programs operate in 24 states, with a total enrollment of 1.8 million beneficiaries.

The second area of importance is the dual Medicare–Medicaid-eligible patients. These are individuals who not only meet the eligibility criteria for Medicare but also Medicaid (often due to financial hardship or disability). In 2020, 12.3 million Americans met this dual eligibility, and these patients account for a disproportionate amount of health care spending. The *Dual Eligible Standards* are updated annually by CMS and outline standards for dual eligibility. Even with Medicaid expansion, dual-eligible individuals do not constitute most of the Medicaid population. However, they account for over 70% of Medicaid spending. Because of this, multiple states are developing Medicare Advantage Dual-Eligible Special Needs Plans (D-SNPs), which have been shown to potentially promote more efficient health care utilization. D-SNPs are a form of Medicare Advantage plan that was authorized under the Medicare Prescription Drug, Improvement, and Modernization Act of 2003 and are specially designed to meet the needs of dual eligible beneficiaries. Approximately 3 million dual-eligible beneficiaries in 43 states and the District of Columbia were enrolled in D-SNPs as of 2021.

OUT-OF-POCKET COSTS

As discussed above, maintaining one's health is the single most expensive cost during old age. Despite governmental support in the form of Medicare and Medicaid, many older Americans bear out-of-pocket costs in health care. Part of the cost is due to the simple fact that increases in the federal budget and growth in Medicare spending have prompted political and policy interest in raising the cost-sharing obligations of older adults. Out-of-pocket costs can include the following: monthly premiums for physician services under Medicare, prescription drug coverage under Medicare, custodial long-term care services and supports, routine dental care (including dentures), routine vision care or eyeglasses, and hearing exams and aids. To cover these expenses, many opt to get supplemental insurance to cover the cost not covered or only partially covered by Medicare (e.g., Medigap, employer or union-sponsored retiree coverage, or employer-sponsored insurance for those who are still working). Those who cannot obtain supplemental insurance due to the high cost and are below the federal poverty level can receive Medicaid supplements. However, even with Medicaid expansion under PPACA, some older adults with low income do not qualify for Medicaid due to their saved assets.

Although there have been measurable reductions in recent years in per capita out-of-pocket Medicare beneficiary spending due to the growth of managed care and cost containment measures, the effects of out-of-pocket payments on older adults, particularly those with limited resources, continue to be problematic. According to the 2018 Medicare Current Beneficiary Survey, beneficiaries with traditional Medicare spent an average of $6,168 on insurance premiums and medical services, with the top quarter paying an average of $14,123. Premiums and supplemental insurance account for about 47%, and of the remaining, long-term facility costs are the highest, accounting for about 18%. This is followed by medical providers/supplies (14%), prescription drugs (11%), and dental care (6%). These out-of-pocket expenses represent a potentially significant financial burden for many households, with half of traditional Medicare beneficiaries reporting at least 16% of income being spent on out-of-pocket health care costs. The cost rises with age, and there are notable differences between men and women as they approach and pass the age of 85. For women, who pay more out-of-pocket expenses than men, the difference can be attributed to rising long-term care costs. Chronic illnesses, such as Parkinson disease, that are associated with aging have profound impacts on rising health care costs. Additionally, out-of-pocket costs can increase for those with mental illness requiring more Medicare-covered services. As discussed further in the Mental Health Parity section, coverage of mental health treatment is not equitable; hence older adults with chronic mental health issues requiring frequent or long psychiatric hospitalizations can face catastrophic costs when obtaining adequate treatment.

Since Medicare and Medicaid do not cover different levels of long-term care, privately paid long-term insurance plans are available and add to additional personal costs. Prescription drug costs continue to be the fastest-growing segment of health care costs for older Americans. To mitigate these costs, PPACA closed Medicare Part D’s “donut hole” in 2020, capping the beneficiary’s responsibility for pharmaceutical payments at 25% once they enter the coverage gap. Furthermore, PPACA restructured Medicare Part B and D premiums to reduce costs for those in low-income brackets and increase contributions from the wealthy. These measures, as well as coverage of preventative care, may work to decrease overall out-of-pocket costs.

PATIENT PROTECTION AND AFFORDABLE CARE ACT AND MENTAL HEALTH PARITY

Historically, a significant financial barrier to mental health care has been the inequality of coverage of mental treatment by insurers. While President Kennedy's administration first required the Federal Employees Health Benefits Program (FEHBP) to cover psychiatric illness at the same level as general medical care, by 1975, even FEHBP scaled back on mental health coverage. Efforts in the 1980s to improve private insurance benefits were largely driven by individual states, and no consistent parity-level care or financial coverage existed. The MHPA originated during President Clinton's administration in the context of his ultimately failed health care reform. By the time it was passed in 1996 as an amendment to the Veterans Affairs, Housing and Urban Development legislation, it just provided mental health parity for annual and lifetime dollar limits for group plans with 50 or more employees. It did not apply to annual hospitalization day or visit limits nor alcohol/substance use treatments. Most health plans found loopholes to circumvent the law, and restrictions were placed on psychiatric hospitalization days and outpatient visits. They also put a larger emphasis on cost sharing. Despite its limitations, MHPA helped push the agenda toward mental health parity. After significant work, in 2008, MHPAEA was passed to address the loopholes present in the MHPA. It underwent the implementation stage and became effective on April 5, 2010. It became applicable beginning or after July 1, 2010. MHPAEA requires that if an employer offers behavioral health coverage, it must be equal to those for general medical/surgical benefits. It continues to prohibit the use of annual and lifetime dollar limits for mental health benefits and now includes addiction benefits.

Since the passage of Medicare and Medicaid in 1965, the most significant health care reform legislation in America is the PPACA. Its overarching goals are to increase the quality and affordability of health insurance, lower uninsured rates, and reduce overall costs of health care. After its passage, it successfully underwent challenges in court on the individual mandate and federal subsidies mechanisms, most recently in 2021 when the Supreme Court upheld it for the third time. By 2016, the Department of Health and Human Services estimated that 20 million fewer Americans were without insurance, and polls indicated a decrease in the percentage of uninsured by nearly half. Most of these newly insured adults were between the ages of 18 to 64, as a high proportion of older adults were already enrolled in Medicare and, therefore, not eligible for coverage expansion. However, with changes in political support and rescinding of the federal tax penalty, the number of enrollees had gone down by 2017. Currently, 38 states and the District of Columbia have expanded coverage with Oklahoma and Missouri beginning in 2021 after voters approved ballot measures. The continued changes to PPACA continue to plague what is already a complicated healthcare landscape. With the Tax Cuts and Jobs Act of 2017, the individual mandate was repealed, and it is estimated that 13 million fewer people will have health insurance in 2027. While it is outside the scope of this chapter to give an overview of all provisions, several, as it pertains to care for older Americans, are worth noting.

For older Americans with mental illness, keeping out-of-pocket health care costs low, ensuring mental health parity, helping low-income individuals, and ensuring adequate coverage of mental health treatments are important. PPACA works on several levels to help achieve these goals. First, the law prohibits insurers from denying coverage due to pre-existing conditions and sets minimum standards for health insurance policies. As most states have set certain mental health coverage as a minimum standard at the parity level, PPACA works to ensure that mental illness treatment is adequately covered. The law also makes significant changes to the health insurance market by creating exchanges that allow individuals and employers to compare policies and prices of insurance. This exchange allows for a more affordable and robust insurance market, as do sliding scale mechanisms for low-income individuals and families (between 100% and 400% of the federal poverty level). Additionally, Medicaid eligibility has been dramatically increased. Now individuals and families with incomes up to 133% of the federal poverty level qualify. This expansion of Medicaid is arguably the single most important aspect of the law, which has worked to decrease the number of uninsured in America. Furthermore, PPACA made several important reforms to Medicare. The reimbursement model has been restructured from fee-for-service to bundled payments. This encouraged cost savings at the point of care. It also created an Independent Payment Advisory Board to help reduce costs and created several measures to monitor for and control Medicare fraud and abuse. PPACA improves preventative care coverage, which will be important to improve the lives of older adults by potentially screening for and possibly preventing chronic diseases earlier. It closed out the “donut hole” in Medicare Part D and reduced prescription costs. Mechanisms of PPACA can help serve older Americans with mental illness by improving outreach and education, improving care transitions, creating Medicare–Medicaid Coordination Office, and encouraging innovation.

The law helps to offset the cost of health care in several ways. It reduced the number of bankruptcies due to medical care costs and, in some cases, allowed individuals to maintain employment without worrying about health coverage. Innovations like medical homes, ACO, expansion of telehealth, and integrated/collaborative care models have been implemented, and their outcomes are being reviewed. Furthermore, it expanded community health centers, invested in public health programs, and promoted caregiver support. Some of these changes have been embraced by the public at large, and despite ongoing changes to the law, it is likely that significant portions of the law will continue. However, some of the existing evaluation studies do not show robust gains in every area. For example, there is evidence to suggest that while PPACA improved health care coverage overall, the incremental change to the underserved population has only been moderate. The gap between the haves and the have-nots continues to be large. The significant rise in health care coverage has increased the demand for mental health care services, yet capacity in the number of clinics, providers, and psychiatric beds has been inadequate. In some circumstances, this has led to a dramatic increase in mental health emergency room visits that can be indicative of insufficient community resources. The rapidly expanded health care coverage has also resulted in some unintended consequences. For example, to meet the demand, many states have started to embrace lower-cost mental health providers such as psychiatric mental health nurse practitioners and physician assistants.

There has been a long history of benefit inequality between mental health and general health, including higher mental health insurance premiums, limitations on mental health care coverage, and carving out mental health care from general health insurance. Importantly for older adults, the lack of mental health care parity in Medicare meant that outpatient mental health treatment services required up to a 50% copayment while general care required just 20%. This posed significant barriers for older Americans to access mental health care. With the passage of PPACA, much of the mandated parity-level coverage for mental health has been reinforced. Importantly, it also mandates that health plans that provide out-of-network coverage for medical/surgical benefits also provide equal mental health benefits. After the implementation of MHPAEA, Medicare payment liability limitations were gradually reduced, and by 2014, a 100% limitation was in effect. Thus, just like other medical treatments, mental health treatment had a cost-sharing of 20%. Unfortunately, as of this writing, the implementation of mental health parity continues to be played out on the ground level with inconsistent application across the states. Even with the passage of PPACA, there still is no specific funding or mechanism to monitor, enforce, or implement parity at the local level. This forces many psychiatrists and patients to continue advocating and working toward mental health parity. The ongoing issues related to mental health carve-out cannot be overstated. Even with Medicare and Medicaid, older adults must navigate a complex care system and often find themselves unable to get treatment at the same place as their primary care. This poses challenges to implementing care coordination, collaborative care models, and streamlined health data access. These ongoing issues prove to be a gap in care from which older adults often fall into and receive disjointed and inconsistent mental health care.

ACCOUNTABLE CARE ORGANIZATIONS AND VALUE-BASED OUTCOMES

A critical challenge in American health care can be summed up by attempts being made to achieve the Triple Aim: improving patient experience of care, improving the health of populations, and reducing the cost of health care being delivered. Recently, there have even been calls to add 4th and 5th dimensions to these goals, namely: improvements in medical health providers' well-being and implementation of health care equity into clinical care. Arguably, these are responses to the problem of rising health care costs that have not translated to positive outcomes in terms of patient experience, overall health, provider job satisfaction, and equitable care to all strata of society. When it comes to mental health care, the potential impacts of achieving these goals cannot be overstated. Overall, mental health care continues to suffer from marginalization and isolation from the rest of medicine. It is notable then that focus on mental health care can contribute significantly to a health care system to achieve these goals.

Some emerging models that will impact geriatric psychiatry in the future are ACO, integrated care, patient-centered medical homes, and population health. All these models attempt to provide cost-effective, highest-quality care–not only to the individual but to large populations. Integrated care is one response to the often disjointed and uncoordinated health care being delivered. The World Health Organization defines it as a concept that brings together inputs, delivery, management, and organization of services related to diagnosis, treatment, care, rehabilitation, and health promotion. The integration model is then designed to improve the overall access, quality, and efficiency of care delivery. One of its outcome goals is to achieve improved patient satisfaction as well. The patient-centered medical home is a model developed by PCPs where a health care provider leads a team centered on comprehensive and continuous care of an individual with the goal of maximizing health outcomes. It embraces the whole person, care coordination by a personal physician, and emphasizes quality and safety. Finally, population health focus is an approach that tries to improve large patient populations by reducing health disparities, influencing social determinants of health, and developing strategies to have specialized but hard-to-find providers focused on certain diagnoses or patient subgroups. Population health management endeavors to focus interventions on specific patients and to avoid worse outcomes in the future. By focusing specialists, like geriatric psychiatrists, on complicated patients as identified within a patient registry and/or within a large organization, it seeks to maximize the impact of the specialist. The overall result would be the identification and prevention of potentially high-cost patients.

One illustrative model for the future is mental health care in ACO and valued-based outcome models. As defined by the Centers for Medicare and Medicaid Services (CMS), an ACO is “an organization of health care practitioners that agrees to be accountable for the quality, cost, and overall care of Medicare beneficiaries who are enrolled in the traditional fee-for-service program who are assigned to it.” ACOs take on large patient populations and demonstrate lower costs to CMS compared to projected costs from usual models taking care of a similar population. The cost savings then is shared between CMS and the ACO. Conversely, if ACO shows a higher cost, it would owe the difference. Additionally, the ACOs must show performance improvements on 33 quality measures across five domains: patient/caregiver experience, care coordination, patient safety, preventative health, and at-risk population/frail older adult health. The initial demonstration programs showed $183 million in savings and improvement in clinical quality scores from 70.8% to 84%. There also has been research showing overall positive experiences from patients and providers in ACO models. After the relative success of the pioneer ACOs, PPACA established Medicare Shared Savings Program (MSSP), which is a mechanism for providers and organizations to establish ACOs, as well as different model tracks for ACOs to interact with federal agencies. As of 2015, 404 MSSP ACOs have been approved, and in the 2014 reporting period, there has been a combined $338 million in savings. Much of these savings come from a decrease in hospitalizations.

The potential benefit to ACOs in paying attention to mental health care is clear. It is now well-established that co-occurring mental illness leads to a higher degree of morbidity. Better screening and treatment of depression have been cited as critical elements in an overall improvement in health outcomes for a population. The importance of this is evidenced by the fact that depression screening and treatment are part of ACOs value outcome measures. For older populations, ACOs should be embraced. However, during its early stages of adoption, the data suggest that many psychiatric providers did not take part. Even in ACOs that had a greater focus on mental health care, the outcomes showed minimal improvements, with many citing access and provider availability as reasons for less-than-desired outcomes. Another challenging aspect of incorporating mental health care into ACOs is related to the ongoing issues of carved-out insurance, the disparity in care, and difficulties in measuring mental health outcomes. However, these are not insurmountable challenges. As part of Medicare, integrated, parity-level mental health coverage should be the standard, and advances in information systems and data-driven care can help drive quantitative outcome measures in psychiatry. Notably, because of the large scope of the problem posed by mental illnesses in older populations, it is likely that ACOs will start focusing on preventative mental health care and develop means to show improvements in provider well-being.

CURRENT AND FUTURE DIRECTIONS

Mental health care has been and will continue to be intimately related to the financial aspects of health care delivery models. As of this writing, the financial implications and impact on geriatric psychiatric care are actively evolving. PPACA and changes to traditional CMS care models have made significant changes to the insurance market (private and public), and their impact will be felt for years to come. The growth of the older adult population will continue to be a major challenge faced by health care policymakers and providers. Without continued innovation and change, older adults will continue to consume health care services at an unsustainable rate. The recent emphasis on cost reduction in service delivery, while technology advances and the size of the frail older population grows at an unprecedented rate, has created conflict, threatening stability and coherence in humanistic, scientific, and economic spheres. Advances in behavioral neuroscience and, in particular, geriatric psychiatry can improve quality of life while extending lifespans and improving the economic productivity of the fastest-growing segment of the population. Despite some concerns, there are ample opportunities for geriatric psychiatrists to get involved and take on leadership roles in ACOs, value-based care models, integrated health care systems, and data-driven care.

One significant issue that has not been adequately addressed by the ongoing legislative push to create high-quality mental health care is that of the workforce. There simply are not enough geriatric psychiatrists to take on the increases in the volume of work, and without the means to meet the demand, many insurers are turning to cheaper and more abundant provider groups like advanced practice providers. The importance of quality for our patients needs to be addressed while also dealing with the realities of our limited workforce. Additionally, the growing emphasis on quality outcome measures begs the question of how to adequately represent quality geriatric psychiatric care when chronic mental illness follows a different time course and expected outcomes. The push to integrate mental health with general health also challenges the traditional geriatric psychiatry practice model. Geriatric psychiatrists must be more transparent and collaborative with their primary care colleagues and other mental health providers. From leadership roles in ACOs to taking on the challenge of creating new models of geriatric psychiatric care, it is important that psychiatrists recognize the unique opportunities being posed in the current drive for health care parity and improved outcomes. By leveraging technologic and medical advances, geriatric mental health care is well suited to achieve these goals.

FURTHER READINGS

Altman D, Frist WH. Medicare and medicaid at 50 years: perspectives of beneficiaries, health care professionals and institutions, and policy makers. *JAMA*. 2015;314(4):384–395.

Barath D, Amaize A, Chen J. Accountable Care Organizations and preventable hospitalizations among patients with depression. *Am J Prev Med*. 2020;59(1):e1–e10.

Bartels SJ, Gill L, Naslund JA. The Affordable Care Act, accountable care organizations, and mental health care for older adults: implications and opportunities. *Harv Rev Psychiatry*. 2015;23(5):304–319.

Breslau J, Han B, Lai J, Yu H. Impact of the affordable care act medicaid expansion on utilization of mental health care. *Med Care*. 2020;58(9):757–762.

Busch SH, Tomaino M, Newton H, Meara E. Access to mental health support services in Accountable Care Organizations: a national survey. *Healthc (Amst)*. 2022;10(1):100613.

Chang CH, Mainor A, Colla C, Bynum J. Utilization by long-term nursing home residents under Accountable Care Organizations. *J Am Med Dir Assoc*. 2021;22(2):406–412.

Chen C, Ackerly D. Beyond ACOs and bundled payments: medicare's shift toward accountability in fee-for-service. *JAMA*. 2014;311(7):673–674.

Cook BL, Flores M, Zuvekas SH, et al. The impact of medicare's mental health cost-sharing parity on use of Mental Health Care Services. *Health Aff (Millwood)*. 2020;39(5):819–827.

Davis K, Stremikis K, Squires D, Schoen C. *Mirror, Mirror on the Wall. How the Performance of the US Health Care System Compares Internationally*. CommonWealth Fund; 2014.

Drake C, Busch SH, Golberstein E. The effects of federal parity on Mental Health Services use and spending: evidence from the medical expenditure panel survey. *Psychiatr Serv*. 2019;70(4):287–293.

Figueroa JF, Lam MB, Phelan J, Orav EJ, Jha AK. Accountable Care Organizations are associated with savings among Medicare beneficiaries with frailty. *J Gen Intern Med*. 2021;36(12):3891–3893.

Figueroa JF, Phelan J, Orav EJ, Patel V, Jha AK. Association of mental health disorders with health care spending in the Medicare population. *JAMA Netw Open*. 2020;3(3):e201210.

Frenier C, Nikpay SS, Golberstein E. COVID-19 has increased Medicaid enrollment, but short-term enrollment changes are unrelated to job losses. *Health Aff (Millwood)*. 2020;39(10):1822–1831.

Fung V, Price M, Nierenberg AA, Hsu J, Newhouse JP, Cook BL. Assessment of Behavioral Health Services use among low-income Medicare beneficiaries after reductions in coinsurance fees. *JAMA Netw Open*. 2020;3(10):e2019854.

Gertner AK, Rotter J, Cruden G. Effects of the mental health parity and addiction equity act on specialty outpatient behavioral health spending and utilization. *J Ment Health Policy Econ*. 2018;21(3):91–103.

Gunasinghe C, Gazard B, Aschan L, MacCrimmon S, Hotopf M, Hatch SL. Debt, common mental disorders and mental health service use. *J Ment Health*. 2018;27(6):520–528.

Keohane LM, Zhou Z, Stevenson DG. Aligning Medicaid and medicare advantage managed care plans for dual-eligible beneficiaries. *Med Care Res Rev*. 2022;79(2):207–217.

Koh S, Kyomen HH. Financial issues in the delivery of geriatric psychiatric care. In: Sadock BJ, Sadock VA, Ruiz P, eds. *Comprehensive Textbook of Psychiatry/X*. Lippincott Williams and Wilkins; 2017:4212–4221.

Mathis J. Housing is mental health care: a call for Medicaid demonstration waivers covering housing. *Psychiatr Serv*. 2021;72(8):880–884.

McDowell MJ, Busch AB, Sen AP, et al. Participation in Accountable Care Organizations among hospitals offering substance use disorder and Mental Health Services. *Psychiatr Serv*. 2018;69(11):1131–1134.

Miller S, Wherry L. Four years later: insurance coverage and access to care continue to diverge between ACA Medicaid expansion and non-expansion states. *AEA Papers and Proceedings*. 2019;109:327–333.

Ng-Mak D, Rajagopalan K. Examining quality of care for individuals treated for mental health using the HEDIS mental health quality measures. *Curr Med Res Opin*. 2019;35(1):87–95.

Nundy S, Cooper LA, Mate KS. The quintuple aim for health care improvement: a new imperative to advance health equity. *JAMA*. 2022;327(6):521–522.

Riley C, Roy B, Herrin J, et al. Association of the overall well-being of a population with health care spending for people 65 years of age or older. *JAMA Netw Open*. 2018;1(5):e182136.

Seo V, Baggett TP, Thorndike AN, et al. Access to care among Medicaid and uninsured patients in community health centers after the Affordable Care Act. *BMC Health Serv Res*. 2019;19(1):291.

Sommers BD, Maylone B, Blendon RJ, Orav EJ, Epstein AM. Three-year impacts of the Affordable Care Act: improved medical care and health among low-income adults. *Health Aff (Millwood)*. 2017;36(6):1119–1128.

Theriault KM, Rosenheck RA, Rhee TG. Increasing emergency department visits for mental health conditions in the United States. *J Clin Psychiatry*. 2020;81(5):20m13241.

Wetzler S, Schwartz B, Wetzler S, Patel U, Counts N. Treatment of serious mental illness in medical and mental health settings. *Psychiatr Serv*. 2020;71(8):789–795.

Zuvekas SH, McClellan CB, Ali MM, Mutter R. Medicaid expansion and health insurance coverage and treatment utilization among individuals with a mental health condition. *J Ment Health Policy Econ*. 2020;23(3):151–182.

55.6 Special Areas of Interest

55.6a Psychiatric Aspects of Long-Term Care

ABHILASH K. DESAI, M.D., AND GEORGE T. GROSSBERG, M.D.

The prevention and treatment of psychiatric disorders is a critical component of comprehensive, state-of-the-science care for all individuals who live in long-term care (LTC) homes. More than 90% of the LTC population, at some point, will experience clinically significant psychiatric symptoms. Table 55.6a–1 shows the prevalence of common psychiatric disorders in the LTC population. Thus, psychiatric care is a core component of the day-to-day care of most residents in long-term care facilities (LTCFs).

Table 55.6a–1.

Prevalence of Common Psychiatric Disorders in Long-Term Care

Psychiatric Syndrome (Point Prevalence)	Key Example(s)	Comments
Major neurocognitive disorders (MNCD) (60%–90%)	Alzheimer disease (AD) Vascular cognitive impairment (VCI) Parkinson disease dementia Dementia with Lewy bodies Frontotemporal dementia Traumatic brain injury and chronic traumatic encephalopathy Hypoxic/Anoxic encephalopathy Alcohol-related dementia Long COVID-19	Many residents have both AD and VCI. Most residents are in advanced stages of major neurocognitive disorder. Behavioral and psychological symptoms of major neurocognitive disorder are the most common reason for psychiatric consultation. Many residents, especially in subacute care, have mild neurocognitive disorder and, when there is superimposed depression or delirium, may be misdiagnosed as having major neurocognitive disorder.
Depressive disorders (20%–40%)	Depression due to MNCD, stroke,	Adequate control of pain is needed to

	cerebrovascular disease, or another general medical condition (most common) Major depression (single episode and recurrent) Persistent depressive disorder Adjustment disorder with depressed mood or with mixed depression and anxiety Depression due to medications	improve depression, and optimal treatment of depression is needed to improve pain for residents who have pain and depression. Drug–drug interactions between antidepressants (e.g., SSRIs) and drugs used to treat medical comorbidity (e.g., tramadol, diuretics, warfarin) are common and need to be monitored. A schedule of pleasant events and physical exercise have been found to be effective in the treatment of depression in residents with MNCD.
Delirium (5%–15%)	Hypoactive delirium (most common) Hyperactive delirium Mixed delirium	Delirium is especially prevalent among residents newly transferred from an acute care hospital (subacute settings

		in skilled nursing facilities). Subsyndromal delirium is even more common and commonly manifests with new onset of agitation and/or new onset of falls. Delirium is often the only initial manifestation of COVID-19 in the LTC population.
Psychotic disorders (5%–10%)	Psychotic symptoms due to MNCD (most common) Psychotic symptoms due to Parkinson disease and dopaminergic agents used to treat Parkinson disease Schizophrenia and schizoaffective disorder	Psychotic symptoms are also common in the context of delirium. Psychotic symptoms are a common cause of persistent physical aggression. As adults with schizophrenia live longer, their prevalence in LTC is expected to rise further.
Hypomania and mania (2%–5%)	Bipolar disorder (types I and II) (most common)	As adults who have bipolar disorder live longer, their prevalence in LTC is

	Medication-induced hypomania or mania Hypomania or mania due to stroke or another general medical condition	expected to rise further. Corticosteroids are the most common cause of medication-induced hypomania or mania.
Anxiety disorders (lack of rigorous data on prevalence; our clinical experience suggests high prevalence)	Anxiety symptoms and bathing phobia in the context of major neurocognitive disorder (most common) Generalized anxiety disorder Panic disorder Fear of falling	Anxiety symptoms are often comorbid with depressive symptoms. Premorbid subsyndromal generalized anxiety disorder is often exacerbated by the development of an MNCD.
Trauma-related disorders (lack of rigorous data on prevalence; our clinical experience suggests growing prevalence)	Posttraumatic stress disorder (PTSD) Nightmare disorder Subsyndromal PTSD	War veterans, victims of abuse (physical, sexual, emotional), and victims of a motor vehicle accident should be identified at the time of admission to an LTCF so that investigation for the presence of PTSD can be done soon after admission.
Impulse-control disorders (5%–	Sexually inappropriate	Dopamine agonists (pramipexole,

15%)	behaviors in the context of MNCD (especially VCI)	ropinirole, rotigotine) may cause impulse-control problems, including sexually inappropriate behaviors and excessive gambling or spending sprees.
Substance-use disorders (lack of rigorous data on prevalence; clinical experience indicates growing and significant prevalence)	Alcohol (most common) Tobacco Prescription opioids Benzodiazepines Cannabis Caffeine	Hazardous use of alcohol, tobacco, prescription opioids, and benzodiazepines that do not meet the criteria for a substance use disorder is prevalent in long-term care and should be treated vigorously. As more baby boomers enter LTC, the prevalence of substance-use disorders (especially alcohol and marijuana) in the LTC population is expected to grow.
Other psychiatric disorders with a growing prevalence	Disorders of the sleep–wake cycle Eating disorders Intellectual disability	These disorders are often comorbid with other psychiatric disorders. They need to be identified

	Autism spectrum disorder Personality disorder	and the treatment plan modified accordingly.

SSRIs, selective serotonin reuptake inhibitors.

APPROACH TO EVALUATING AND MANAGING PSYCHIATRIC SYMPTOMS AND DISORDERS

A systematic approach to evaluating and managing psychiatric disorders using evidence-based practices and interventions has the potential to reduce suffering dramatically and prevent unnecessary visits to the emergency department and hospitalization. Table 55.6a–2 lists the role of psychiatric consultant (psychiatrist, psychiatric nurse practitioner, psychiatric physician assistant) in LTC. The authors recommend four steps to evaluate and manage psychiatric disorders in the LTC population.

Assessment of Safety

The first step is to assess safety. The psychiatric consultant may need to address behaviors that pose an imminent danger to the resident or others (behavioral emergencies), such as violent or suicidal behaviors, before attempting to assess underlying psychiatric problems. Suicidal behaviors can be direct (e.g., suicidal ideas, such as voicing thoughts to cut oneself with a knife or scissors, or suicide attempts, such as attempting to strangle oneself with a cord) or indirect (e.g., refusing dialysis, refusing food, or life-preserving medications such as warfarin). Completed suicides, although rare in LTC, may be more common than in the community. Psychopharmacological medications (PPMs) may need to be used judiciously to address behavioral emergencies. Table 55.6a–3 lists PPMs that are commonly used to manage behavioral emergencies. If a behavioral emergency cannot be resolved, the resident may need to go to the nearest emergency department for urgent evaluation and stabilization.

Team-Based Evaluation to Identify Causes of Psychiatric Symptoms and Disorders

The second step is for a multidisciplinary team (resident, family members, primary care provider, psychiatric consultant, nurse, nursing assistant, social worker, and activity therapist) to identify all factors that may contribute to psychiatric symptoms. Psychiatric symptoms are typically due to several specific reversible or modifiable causes. Table 55.6a–4 lists common reversible causes of psychiatric symptoms in the LTC population. Obtaining the resident's perspective is essential, even if the resident has significant cognitive impairment. Diagnoses listed in the LTC resident record need to be validated as they are often incorrect. In complicated cases (e.g., presence of chronic mental illness, refractory and severe behavioral and psychological symptoms of neurocognitive disorder, concerns of abuse and neglect, complex polypharmacy), the psychiatric consultant may need to consult a professional who has geriatric expertise (e.g., geriatric psychiatrist). Telepsychiatry may improve access to geriatric psychiatrists who may be able to be present via audio/video to assess the resident along with the primary care provider.

Action Plan

In the third step, the team collaboratively devises and institutes an action plan (treatment plan) that is individualized and directed at the specific target psychiatric symptoms, diagnoses, and differential diagnoses identified in the second step. An action plan generally consists of a multimodal approach combining a variety of psychosocial environmental interventions targeted at caregivers and residents to address psychiatric symptoms that are most distressing or problematic. An action plan should address all the potentially modifiable causes of psychiatric symptoms identified during the evaluation. Professionals from allied fields (e.g., recreational therapist, dietician, psychologist, pharmacist, occupational therapist, physical therapist, speech therapist, music therapist, art therapist, board-certified behavior analyst) and professionals who have knowledge of complementary and alternative medicine (e.g., aromatherapy, massage therapy, horticultural therapy) can play valuable roles, especially in complicated and or treatment-resistant cases. PPMs may be the first line in behavioral emergencies related to violent behaviors, relapse of pre-existing psychiatric disorder (e.g., bipolar disorder), new onset of major depression, persistent and distressing psychotic symptoms, and persistent aggressive behavior. Table 55.6a–5 lists key principles of psychopharmacology with the LTC population.

Response to Treatment

The fourth step involves appropriately monitoring the adverse effects of interventions as well as response to treatment so that the psychiatric consultant can evaluate whether to modify or continue the interventions. If the problem is not improving, the psychiatric consultant should discontinue the intervention. If the intervention included an antipsychotic for the management of severe persistent aggression in residents with advanced dementia, the psychiatric consultant should make systematic attempts to taper the dosage and discontinue the medication after 4 to 7 months.

DELIRIUM AND PSYCHIATRIC ASPECTS OF COMORBID MEDICAL CONDITIONS

Age above 85, presence of a neurocognitive disorder, and polypharmacy are the three most important risk factors for delirium in the LTC population. Residents who have advanced stages of Major Neurocognitive Disorder (MNCD) have low cognitive reserve and thus are prone to develop delirium even with mild conditions such as urinary tract infection or a new medication that has anticholinergic activity. The psychiatric consultant should consider an acute change in mental status the sixth vital sign (pain being the fifth vital sign and heart rate, respiratory rate, temperature, and blood pressure being the first four vital signs). New onset of falls and agitation are the two most common early signs of developing delirium in LTC populations.

New onset of agitation is one of the most common reasons for psychiatric consultation, and the psychiatric consultant may be the first health care professional to diagnose delirium.

There are three types of delirium: hypoactive (most common), hyperactive, and mixed. Psychotic symptoms are common during delirium and may cause the resident to become severely agitated or physically aggressive, further complicating management. The Confusion Assessment Method (CAM) is a useful tool for improving the accurate diagnosis of delirium in the LTC population. Table 55.6a–6 lists common causes of delirium in the LTC population. Mania can also mimic delirium. Management of delirium involves ensuring safety and identifying and treating its various causes. Judicious short-term use of low doses of atypical antipsychotics or haloperidol may be appropriate and necessary for the management of severe physical aggression and psychotic symptoms in the context of delirium.

Table 55.6a–2.
Role of the Psychiatric Consultant in Long-term Care Homes

Areas of Concern	**Role of the Psychiatric Consultant**
Medicalization of human suffering (e.g., antidepressant given for adjustment disorder with depression)	Normalize distress related to losses and adjusting to living in an institution; deprescribe psychiatric medications given for adjustment problems
Inappropriate treatment of behavioral emergencies (e.g., suicidal ideas, aggression)	Guide PCP and team in prompt assessment, understanding the behavioral emergency and appropriate treatment
Inaccurate psychiatric diagnoses (e.g., dementia-related psychoses diagnosed as schizophrenia)	Conduct a comprehensive psychiatric evaluation and provide accurate diagnoses
Irrational psychopharmacologic medication use (e.g., quetiapine used for insomnia)	Rational psychopharmacologic medication use
Medication-induced psychiatric problems (e.g., anticholinergic medications causing cognitive impairment)	Identify medications that could be contributing to or causing psychiatric problems and, in collaboration with PCP and consultant pharmacist, institute appropriate interventions (e.g., rational deprescribing of anticholinergics)
Under-recognition of medical conditions causing agitation in persons with MNCD	Help PCP identify potential medical conditions (e.g., infection, delirium, pain) that may be causing or worsening agitation in persons with MNCD
Inadequate optimization of	Work with family and staff to

psychosocial environmental approaches to manage agitation in persons with MCND	problem solve, provide education, and optimize personalized strengths-based psychosocial environmental approaches to effectively reduce distress in persons with MNCD
Low utilization of individual and group counseling for psychiatric disorders (e.g., depression, chronic pain, chronic insomnia, chronic anxiety)	Help the team identify residents who could benefit from individual and group counseling (e.g., cognitive-behavioral therapy)
Delayed recognition of the need for hospice and palliative care in residents with advanced dementia	Help the team recognize the need for hospice and palliative care in residents with advanced dementia
Lack of knowledge and skills to manage “difficult” residents (e.g., residents with trauma, personality disorders, and addictions)	Staff and family education, guidance, and support in appropriate management of trauma, personality disorders, and addictions
PCP needs help assessing the resident’s capacity to make medical decisions	Psychiatrists can assess resident’s decision-making capacity and guide the PCP and team

Table 55.6a–3.
Guide to Key Psychopharmacological Options for Management of Acute Agitation in LTC[a]

Psychopharmacological Options (Including Dosages per 24 Hours, Routes of Administration)*b*	**Common Conditions**	**Clinical Pearls**
Haloperidol (0.5 mg–2 mg; tablet, liquid, intramuscular injection)	Delirium Acute exacerbation of chronic mental illness (e.g., Schizophrenia, Bipolar Mania)	Drug of choice in delirium (while causes are being investigated) Avoid in residents with pre-existing parkinsonism or a history of extrapyramidal syndrome (EPS) due to antipsychotics
Risperidone (0.5–2 mg; tablets, liquid, orally disintegrating tablet [ODT])	Delirium Acute exacerbation of chronic mental illness (e.g., Schizophrenia, Bipolar Mania)	Avoid in residents with pre-existing parkinsonism or a history of EPS due to antipsychotics
Olanzapine (5–10 mg; tablet, ODT) (2.5–5 mg intramuscular injection)	Delirium Acute exacerbation of chronic mental illness (e.g., Schizophrenia, Bipolar Mania)	Avoid in residents at high risk of anticholinergic toxicity (e.g., residents with high pre-existing anticholinergic burden, residents with urinary retention)
Ziprasidone (10–20 mg;	Delirium	Avoid in residents

intramuscular injection)	Acute exacerbation of chronic mental illness (e.g., Schizophrenia, Bipolar Mania)	with prolonged QTc interval (especially if QTC is longer than 500 msec)
Aripiprazole (5–10 mg; ODT) (9.75 mg intramuscular)	Delirium Acute exacerbation of chronic mental illness (e.g., Schizophrenia, Bipolar Mania)	Monitor for akathisia
Asenapine (5–10 mg; sublingual)	Delirium Acute exacerbation of chronic mental illness (e.g., Schizophrenia, Bipolar Mania)	Cost may be a barrier to its use
Chlorpromazine (25–50 mg; tablets) (10–25 mg intramuscular injection)	Delirium Acute exacerbation of chronic mental illness (e.g., Schizophrenia, Bipolar Mania)	Used if other options are not feasible or not effective Monitor for hypotension
Lorazepam (0.5–5 mg; tablet, liquid, intramuscular injection)	Delirium related to alcohol withdrawal or benzodiazepine withdrawal Panic attack Acute exacerbation of chronic mental illness (e.g., Schizophrenia, Bipolar Mania)	Avoid in residents with advanced lung disorders In some situations (e.g., acute exacerbation of schizophrenia), a combination of lorazepam and haloperidol or lorazepam and

		risperidone may be considered
Midazolam (10–20 mg; liquid; higher doses as subcutaneous infusion may be needed in some situations)	Palliative sedation for refractory agitation during end-of-life care	Training in the use of buccal midazolam is needed. Buccal midazolam is generally used for the management of seizures as an alternative to rectal diazepam

---

[a] Identification and treatment of underlying causes and nonpharmacologic interventions (e.g., de-escalation strategies) are always first-line interventions. No medication has been approved by the FDA for the management of delirium. Intramuscular olanzapine, ziprasidone, and aripiprazole have been approved by the FDA for the management of acute agitation in patients with schizophrenia, but they have not been studied in LTC populations. Start with lower dosages and may repeat the dose in 30–45 minutes if the initial dose is not effective.

[b] In frail elderly, dosages may need to be lower, and for some elderly (e.g., young-old [age 65–75 years] who are not frail), dosages may need to be higher than suggested here. Oral routes of administration are preferred over parenteral; ODT, liquid, and sublingual are preferred over tablets.

Table 55.6a–4.

Common Reversible Causes of Psychiatric Symptoms in Long-Term Care

Cause	Key Example(s)	Potential Psychiatric Symptoms
Medications	Benzodiazepines Anticholinergics Corticosteroids Dopamine agonists Opioids	Cognitive impairment Agitation Hypomania, mania Psychotic symptoms Impulse control disorders
Physical health conditions	Adverse drug events, pain, neuropathy, constipation, urinary tract infection, dehydration, delirium, anemia, hypoglycemia, vitamin B_{12} deficiency, vitamin D deficiency, low blood pressure due to overtreatment of hypertension, allergies, dry skin	Agitation Depression Cognitive impairment
Sensory deficits	Hearing deficit Vision deficit	Depression Agitation Paranoia Visual hallucinations Cognitive impairment
Disorders of the sleep–wake cycle	Insomnia Obstructive sleep apnea	Increased or decreased sleep Night-time agitation Cognitive impairment

	Excessive daytime sleepiness REM sleep behavior disorder Restless leg syndrome	
Psychiatric conditions	Major neurocognitive disorders Major depression (single episode or recurrent) Generalized anxiety disorder Post-traumatic stress disorder (PTSD) Bipolar disorder Schizophrenia	Agitation/aggression Social withdrawal Insomnia Cognitive impairment Nightmares Poor food and fluid intake Delusions Hallucinations Hypomania/mania
Long-term care staff approach	Impatience Lack of empathy Over expectation	Agitation/aggression Depression Anxiety
Long-term care staff's understanding of psychiatric symptoms/disorders	Staff's lack of understanding of aging-related disability ("he can stay in bed all day, as he is 90 years old and is tired")	Agitation Depression Insomnia
Long-term care staff distress	Caregiver anxiety and depression Caregiver burnout Excessive caregiver burden (having to	Agitation Depression

	care for too many residents)	
Negative environmental factors	Nighttime noise Low lighting Boredom Overstimulation Improper or inadequate incontinence care routines Therapeutic diet Lack of places to wander safely Physical restraint	Agitation Depression Anxiety Poor food and fluid intake Insomnia
Unmet needs of residents	COVID-19 and pandemic-related restrictions Toileting, hunger Physical activity Spend time with nature To be productive, to be needed Spiritual needs To express and vent Freedom from restraint	Loneliness Agitation Depression Anxiety

Case example 1: Ms. A is a 65-year-old woman who recently transferred from a local hospital to a skilled nursing facility. She was referred for psychiatric evaluation because she was not eating, appeared depressed, and was not sleeping well. The nursing staff suspected depression. During the assessment, the psychiatric consultant discovered that Ms. A had undergone a colonic resection for ulcerative colitis and had excessive liquid drainage from an ileostomy. During the examination, Ms. A was inattentive and somewhat drowsy, and had psychomotor retardation. The psychiatric consultant diagnosed hypoactive delirium, possibly due to disorders of fluids and electrolytes. A subsequent basic metabolic panel showed hyponatremia, elevated blood urea nitrogen, and serum creatinine (higher than baseline), confirming the diagnosis. Treatment with intravenous fluids at the facility quickly brought resolution of the presenting symptoms.

Clinical pearl: Hypoactive delirium often presents as depression.

Table 55.6a–5.
Key Principles of Psychopharmacology in Long-term Care

1. Prescription of even a single psychopharmacologic agent is a complex decision. It should be undertaken judiciously and, in most situations, after psychosocial environmental interventions have been tried and failed.
2. Start low, go slow, and keep going. Initial doses of psychotropic drugs need to be low or, for frail residents, very low. The dose increase needs to be slower than for younger adults. Final therapeutic doses for some residents may need to be as high as those for younger adults.
3. The benefit–risk ratio for all pharmacologic agents (including psychopharmacologic drugs) is lower than that for younger adults, necessitating close monitoring for adverse effects and prompt discontinuation of the drug if adverse effects develop or if there is no therapeutic response.
4. The psychiatric consultant should evaluate for periodic dose reduction and discontinuation of psychotropic drugs, and documentation should include the justification for continued use of psychopharmacologic agent(s).
5. The psychiatric consultant should promptly reassess the benefit-risk ratio of all psychotropic drugs whenever there is a significant decline in the resident’s overall health and functioning or a decline in renal and/or hepatic function.
6. Psychopharmacologic therapies should comply with regulatory guidelines for the use of medications in nursing homes.
7. The psychiatric consultant should provide residents and their surrogate decision makers optimal information regarding the risks and benefits of psychopharmacologic therapies and alternatives before initiating therapy and periodically thereafter.

Malnutrition

Malnutrition is prevalent in the LTC population. It commonly manifests as unintentional weight loss or, less commonly, as excessive weight gain and is a common reason for psychiatric consultation. Weight loss is usually accompanied by loss of appetite (including general loss of appetite as well as loss of appetite for food offered at the facility) or refusal of food. Depression is one of the most common causes of a general loss of appetite. Medications such as cholinesterase inhibitors also can cause anorexia and weight loss. Not having someone available to assist with meals, especially in residents with MNCD, is also a common reason for weight loss in the LTC population. Liberalizing diets, increasing assistance during feeding, optimizing dental health, cutting food into smaller and manageable sizes (finger foods), giving smaller portions at each meal, and increasing the availability of favored snacks throughout the day can help address the problem of food refusal. The psychiatric consultant should evaluate all residents who refuse food for depression as well as for psychotic symptoms (especially delusions of food being poisoned). For residents who have depression accompanied by loss of appetite and weight loss, the psychiatric consultant may consider prescribing mirtazapine as the first choice antidepressant. Nutritional supplements should be given between meals rather than with meals.

The prevalence of obesity and excessive weight gain is increasing in the LTC population. This is typically accompanied by excessive consumption of calorie-dense food. Along with a sedentary lifestyle, PPMs (e.g., valproate, olanzapine, quetiapine) and other pharmacologic medications (e.g., gabapentin) are common causes of excessive weight gain. It may be necessary to discontinue PPMs that are thought to cause or contribute to weight gain and, as necessary, replace them with alternative PPMs that have a lower propensity for weight gain.

Table 55.6a–6.
Common Causes of Delirium in Long-Term Care

Cause	Key Example(s)	Comments
D: Drugs	Anticholinergic drugs Benzodiazepines Opioids Steroids	Drugs listed in the Beers list of drugs that are inappropriate for older adults often cause cognitive impairment and delirium. Many drugs have low anticholinergic activity and, if used with other drugs that have anticholinergic activity, may increase anticholinergic load sufficient to cause delirium. Benzodiazepine withdrawal can cause delirium.
E: Embolic—thrombotic	Pulmonary emboli Deep vein thrombosis (DVT) Cerebrovascular accident (CVA) COVID-19	Residents who spend a large amount of time in bed (e.g., those who have depression or those who are in pain) are at high risk of DVT and its complications. Neuroimaging may be necessary to diagnose CVA accurately.
L: Last phase of life	Residents in an end-of-life state	Delirium is common in the terminal stages of major neurocognitive disorder.
I: Infections	Urinary tract infection	Asymptomatic bacteriuria and asymptomatic

	Bronchitis Pneumonia Cellulitis COVID-19	pyuria are prevalent, especially among women, and often misdiagnosed as urinary tract infections. Delirium may be a more common manifestation of pneumonia than fever and cough.
R: Respiratory conditions R: Renal conditions	Acute exacerbation of chronic obstructive pulmonary disease (COPD) Acute renal failure (ARF)	Acute bronchitis is one of the common causes of exacerbation of COPD. Dehydration is a common cause of ARF.
I: Intestinal conditions I: Injury	Acute intestinal obstruction, including fecal impaction Hip fracture Head injury	Inadequate treatment of constipation often leads to intestinal obstruction.
U: Urologic	Urinary retention	Drugs that have anticholinergic activity are the most common cause of acute urinary retention.
M: Myocardial dysfunction M: Metabolic	Acute myocardial infarction (AMI) Dehydration Hypothyroidism or hyperthyroidism Hyponatremia	Delirium may be a more common manifestation of AMI than chest discomfort. Laboratory tests may be the only way to

	diagnose dehydration accurately.

Macro- and micronutrient deficiencies can often lead to depression, insomnia, and neurocognitive disorders. Many commonly used medications can cause micronutrient deficiencies (e.g., metformin and proton pump inhibitors are associated with vitamin B12 and magnesium deficiency). Checking vitamin levels (e.g., B12, D, B1), magnesium levels, and tests for celiac disease may be necessary in some cases. Vitamin supplementation without checking for blood levels may improve mental health outcomes.

Dieticians may be consulted to address nutritional deficiencies but also for education related to dietary strategies for boosting brain function (e.g., Mediterranean diet, MIND diet) and for drug–food interactions (e.g., dietary needs of individuals on lithium).

Many psychiatric medications need to be taken with food for adequate absorption (e.g., ziprasidone needs to be taken on 500 calories meal, lurasidone needs to be taken on 350 calories meal), and some psychiatric medications should not be taken with meals (e.g., quetiapine extended release).

Chronic Pain

Psychiatric consultation can provide valuable help in improving outcomes and the quality of life of residents who are experiencing chronic pain. Depression is common, underrecognized, and undertreated in residents who have chronic pain. Pain is one of the most common causes of agitation in residents who have MNCD, especially verbal agitation (e.g., yelling, screaming, verbally abusive behavior) and resisting personal care. The psychiatric consultant should consider pain the fifth vital sign. We recommend the *ATMAN* approach for the management of chronic noncancer pain without opioids and tramadol. Refer to Table 55.6a–7 for the description of the *ATMAN* approach. This involves the use of topical analgesics and nondrug interventions as first line, nonopioid oral analgesics, and interventional pain management as second line, and only in intractable cases, the use of tramadol or other opioids. Antidepressants, especially serotonin-norepinephrine reuptake inhibitors (SNRIs, such as duloxetine), can help reduce the perception of pain and improve function as well as the quality of life, even in the absence of significant depression, for residents who have chronic pain. Serious adverse drug–drug interactions between psychotropic drugs and analgesics (e.g., selective serotonin reuptake inhibitors [SSRIs] and tramadol [increased risk of seizure and serotonin syndrome], benzodiazepines and opioids [increased risk of delirium, falls, and injury]) can be prevalent and underrecognized in the LTC population. The psychiatric consultant should consider recommending a trial of scheduled acetaminophen to treat agitation (especially verbal agitation) in individuals with advanced MNCD before prescribing PPM. Psychiatric care should be a key component of an integrated approach to chronic pain management.

Table 55.6a–7.
ATMAN Approach to Managing Chronic Noncancer Pain Without Opioids and Tramadol

A:	Acetaminophen, Antidepressants (especially selective serotonin-norepinephrine reuptake inhibitors)
T:	Topical agents (e.g., diclofenac, lidocaine, capsaicin)
M:	Muscle relaxants (e.g., baclofen, baclofen pump)
A:	Anti-inflammatory agents (e.g., NSAIDs, intra-articular glucocorticoids), anticonvulsants (e.g., gabapentin, pregabalin, carbamazepine)
N:	Nondrug interventions (e.g., cognitive-behavioral therapy for pain, exercise, multimodal rehabilitation, physical therapy, self-management, heat, and cold remedies)

SPECIFIC DISORDERS AND LONG-TERM CARE

Major Neurocognitive Disorders (Dementias)

Evaluation and treatment of MNCDs are key aspects of psychiatric care for LTC residents. Residents who have MNCD and have not had a thorough evaluation to clarify its cause(s) may need workup to rule out potentially treatable or reversible factors of cognitive decline, including thyroid dysfunction, vitamin deficiency (e.g., B12 deficiency), subdural hematoma, normal pressure hydrocephalus, central nervous system infection, and brain tumor. The authors recommend using the Montreal Cognitive Assessment (MoCA) or Saint Louis University Mental Status (SLUMS) to assess residents' cognitive function. The Brief Mental Status Interview is not as thorough as the previous two scales, although it is part of the Minimum Data Set used to assess all residents in nursing homes in the United States.

Behavioral and psychological symptoms of MNCD (also called behavioral and psychological symptoms of dementia [BPSD]) are one of the most common reasons for psychiatric consultation. They are a heterogeneous group of symptoms that tend to cluster into five distinct groups—depression, apathy, psychotic symptoms (delusions and hallucinations), sleep disturbance, and agitation—and are distinct from the cognitive deficits that are the hallmarks of all MNCD. Agitation, defined as "inappropriate verbal, vocal, or motor activity that is not explained by needs or confusion per se," can be characterized as either aggressive or nonaggressive. Aggressive behaviors can occur with or without agitation and include fighting, throwing, grabbing, destroying furniture, cursing, and screaming, whereas nonaggressive agitated behaviors include restlessness, pacing, wandering, hoarding, repetitive questioning, inappropriate disrobing, and verbal outbursts. Agitation is best understood as an expression of a resident's efforts to communicate unmet needs. The majority of individuals who have MNCD (including those who have advanced MNCD) can speak for themselves to a much higher degree than has been recognized to date. The psychiatric consultant should make all efforts to understand the resident's perspective and have the treatment plan guided by that perspective. The psychiatric consultant may use the Agitation Subscale of the Neuropsychiatric Inventory nursing home version, Cohen Mansfield Agitation Inventory, or Pittsburgh Agitation Scale to evaluate the severity of agitation and to document response to treatment. The impact of the COVID-19 pandemic is disproportionately high on individuals with advanced dementia. Tables 55.6a–8 and 55.6a–9 describe the impact and potential psychosocial interventions to mitigate stress and address complications.

Management of MNCD should begin with a discussion with the resident and/or their family regarding the potential benefits of using cholinesterase inhibitors and memantine. Some more-capable residents (e.g., residents who have mild neurocognitive disorder due to probable Alzheimer type and highly motivated residents who have mild MNCD) may benefit from and often prefer interventions to slow cognitive decline. For these residents, creating and implementing an individualized brain fitness program that includes regular exercise, a MIND diet, and cognitive and social strategies in various combinations is appropriate and recommended.

Individualized, strengths-based psychosocial environmental interventions (also called nonpharmacologic interventions) are recommended as the first line of treatment for all behavioral and psychological symptoms related to MNCD because they are effective (at least modestly) and the risks are minor. Table 55.6a–10 lists psychosocial environmental interventions for neurocognitive disorder, and Table 55.6a–11 lists key staff training programs.

Table 55.6a–8.
Impact of COVID on Residents With Dementia

Stressor	Potential Biologic–Psychosocial Mechanisms	Potential Negative Consequences
Physical distancing, masks, lockdown, and other infection-prevention interventions	Loneliness Difficult understanding spoken and nonverbal facial communication Trauma	Depression Apathy Agitation Anxiety
Acute COVID infection	Elevated cytokines Hypoxia Ischemia and blood clots Adverse effects of medications used to manage COVID (e.g., dexamethasone)	Delirium Fatigue Psychosis (rare) Agitation Increased mortality
Post-COVID sequelae (Long COVID)	Persistent inflammation in cerebral blood vessels Loss of taste	Persistent neurocognitive impairment (“Brain Fog”) Reduced food intake and weight loss

The psychiatric consultant should institute PPMs to treat severe and persistent agitation related to MNCD only after thoroughly evaluating and treating all potentially treatable medical conditions and other modifiable factors, and psychosocial environmental interventions have not been adequate in addressing them.

Discussions should make it clear that any psychopharmacologic intervention for agitation in the context of neurocognitive disorder is off-label (i.e., is not approved by the U.S. Food and Drug Administration for its treatment) with the exception of suvorexant, that is approved by the U.S. FDA for insomnia due to Alzheimer's disease dementia and brexpiprazole, that is approved by the U.S. FDA for the treatment of agitation due to Alzheimer's disease dementia.

Antipsychotics are one of the most commonly prescribed PPMs for the treatment of agitation and aggression among residents who have MNCD, and the prevalence of use in the LTC population is high (20% to 25%). There is a strong initiative from the Center for Medicare and Medicaid Services (CMS) in the United States to reduce the use of antipsychotics in the LTC population. In addition, the United States Congress has passed legislation mandating the reduction of antipsychotic use in nursing homes. This is in response to strong evidence that antipsychotics are often used inappropriately in the LTC population (the majority of its use is by residents who have MNCD), and the use of antipsychotics by individuals who have MNCD carries high risks of morbidity (e.g., cerebrovascular event, pneumonia, deep vein thrombosis, hip/femur fracture) and increased mortality. Table 55.6a–12 shows a stepped-care approach to the appropriate use of antipsychotic medications for MNCD. The psychiatric consultant may consider prescribing short-term use of low-dose brexpiprazole, risperidone, olanzapine, quetiapine, and aripiprazole to treat severe agitation, aggression, and psychosis with dementia in certain clinical situations (e.g., clear risk of harm to the resident or others and psychosocial environmental interventions have not been effective). Risperidone is approved for short-term use (up to 6 weeks) to manage severe aggression in individuals with MNCD in Europe. Atypical antipsychotics are preferred over conventional antipsychotics (e.g., haloperidol), which are associated with higher risks. High-potency antipsychotics (e.g., risperidone, haloperidol) should be avoided in individuals with parkinsonism (e.g., Parkinson disease (PD), Lewy Body Dementia). It is important that the psychiatric consultant explain to the surrogate decision maker(s) that severe psychotic symptoms and/or persistent aggression are associated with considerable emotional distress, increased mortality, falls, and hospitalization; hence, in certain situations, the use of antipsychotics may be justified. To minimize risks, the authors recommend prescribing the lowest possible dose of antipsychotics for the shortest possible time and regular monitoring of metabolic parameters when appropriate and for extrapyramidal symptoms.

Table 55.6a–9.
Examples of Some Innovative Psychosocial Interventions to Address Loneliness and Trauma

1. Small-group activities (including opportunities for exercise)
2. Distanced-communal dining
3. Compassionate care visits
4. Window visits with family and friends
5. Video visits with family and friends (Facetime, Skype, Zoom, and other platforms)
6. Reading books, poems, scriptures
7. Audiobooks
8. Music-based interventions (including personalized music on an MP3 player or iPod)
9. Encouraging residents to apply lavender (or another soothing scent) lotions to hands, feet, and neck
10. Simulated presence therapy: playing select audio–video recordings of family and friends
11. One-to-one interaction (includes supervised visits outdoors to look at squirrels and birds).

Among mood stabilizers, there is some support for the use of carbamazepine to treat physical aggression in MNCD, but its use carries considerable risks, including falls, drug–drug interaction, and delirium. Valproate has been the most studied of all anticonvulsants for the treatment of behavioral and psychological symptoms of MNCD, but the results to date have been inconsistent or negative, and its use carries considerable risks (especially falls and sedation).

Hence, the authors do not recommend the use of valproate to treat behavioral and psychological symptoms of MNCD.

SSRIs (especially citalopram and escitalopram) and low-dose trazodone are options for the treatment of persistent anxiety, irritability, agitation, and aggression not responding to psychosocial environmental interventions. The psychiatric consultant may also consider prescribing antidepressants to treat severe emotional lability. Antidepressants, although safer than antipsychotics, carry considerable risks, including the risk of increased agitation, nausea, insomnia, hyponatremia, bleeding (especially if used concomitantly with anticoagulants), and falls. Some antidepressants (e.g., duloxetine) also pose a risk of orthostatic hypotension. The psychiatric consultant may consider prescribing a trial of a combination of dextromethorphan and quinidine to treat pseudobulbar affect. The psychiatric consultant may consider a judicious trial of methylphenidate for residents who have MNCD or traumatic brain injury and are experiencing severe apathy when psychosocial environmental interventions have not been effective, and there are no cardiac contraindications. The use of benzodiazepines carries a high risk of serious adverse effects (especially delirium, falls, and injury) and thus should be minimized. Short-term use of low-dose short-acting benzodiazepines such as (lorazepam or oxazepam) is an option for the treatment of prominent anxiety or infrequent severe agitation. Hydroxyzine carries a significant anticholinergic effect, and hence its use to treat anxiety in individuals with MNCD should be avoided.

Table 55.6a–10. Psychosocial/Environmental Interventions for Neurocognitive Disorders

Psychosocial Interventions Directed Toward Caregivers

Type of Psychosocial Environmental Intervention	Examples	Practical Suggestions
Staff education	Education about dementia Education about Parkinson disease Education about common triggers for agitation Education about adverse psychiatric effects of commonly prescribed medications	Case-based education is preferred. Recording the education and making it available on a website for caregivers to review on their own time. Providing work credit for caregivers who complete the education.
Staff training	Training in approach to the management of behavioral and psychological problems in long-term care residents Training in effective communication and appropriate documentation	Hands-on training, especially for new staff and staff with minimal experience. Video recordings of training for staff to review on their own time.

Staff support and incentives	Informal support to staff members Formal support groups for staff members Opportunity to take on a leadership role in psychiatric aspects of long-term care	During visits, a psychiatric consultant can provide direct and situation-specific verbal support to caregivers. A break room for staff members where they can “chill” (massage chairs, relaxing music, yoga videos, and mats).

Psychosocial Environmental Interventions Directed Toward Residents

Type of Psychosocial Environmental Intervention	Examples	Comments/Resources
Environmental	Design of physical environment that accommodates cognitive impairment and promotes emotional well-being Robust activity programming Structured routines Therapeutic gardens	www.healthdesign.org Consultation, staff education, and periodic supervision of the programs by recreational therapist Activity programs run by activity therapist Cognitive stimulation therapy (www.cstdementia.com) www.timeslips.org www.thegreenhouseproject.org
Music-based	Individualized	Consultation, staff education,

strategies	music Group music	and periodic supervision of the programs by music therapists Availability of piano for the residents to play, music CDs, iPod, mp3 players, radio
Physical activity and exercise	Walking programs Chair yoga Tai Chi Dance exercise programs	Guidance, staff education, and periodic supervision of the programs by a physical therapist Staff encouraged to make physical activity a normal part of the resident’s routine
Aromatherapy combined with massage therapy	Hand or foot massage with a fragrant lotion	Consultation, staff education, and periodic supervision of aromatherapy programs by certified aromatherapist Availability of common therapeutic lotions (e.g., lavender and citrus lotions) for staff to use
Animal-assisted therapy	Dogs and cats visiting	Eden Alternative programs (www.edenalt.org)
Emotional support	Support groups	Stroke support group Early-stage dementia support group Chronic pain support groups Parkinson disease support groups Veterans support groups
Counseling	Problem-solving therapy	Facility administration encouraging licensed

	Reminiscence therapy Interpersonal therapy Cognitive-behavioral therapy	psychotherapists/counselors to provide psychological therapies to cognitively capable residents Internet-based self-psychotherapy for cognitively capable residents
Spiritual	Maintaining religious rituals	“Postcards to God” program

Table 55.6a–11.
Eight Key Training Programs for Long-Term Care Staff

Training Program	Resource
Person-centered language training	https://practicetransformation.umn.edu/clinical-tools/person-centered-language/
Bathing without a Battle training	https://bathingwithoutabattle.unc.edu
Mouth care without a Battle training	https://www.mouthcarewithoutabattle.org
Namaste care training	https://namastecare.com/namaste-care-individual-program/
The DICE approach—systematic approaches to address behavioral expressions (aka behavioral and psychological symptoms of dementia [BPSD]) Online training:	https://diceapproach.com
TimeSlips Creative Engagement Training	https://timeslips.org
Cognitive stimulation therapy	https://www.ucl.ac.uk/international-cognitive-stimulation-therapy/international-cognitive-stimulation-therapy-cst-centre

Staff Mindfulness and Resilience Training (SMART sessions)	A psychiatric consultant can provide such training or guide the social worker in providing this training

Table 55.6a–12.

Approach to Appropriate Use of Antipsychotic Medications to Manage Severe Behavioral and Psychological Symptoms of Major Neurocognitive Disorder

Step 1: Presence of at least one of the following:

- Persistent severe agitation and or physically aggressive behavior

or

- Psychotic symptom causing severe emotional distress

Step 2: Assess safety

- The resident poses an imminent danger to themself and/or others. (Address safety first with the following options individually or in combination: one-to-one staff monitoring; use of oral or intramuscular antipsychotic medication; transfer to Emergency Department. When safety is ensured, proceed to Step 3.)
- The resident does not pose an imminent danger to themself and/or others. (Proceed to Step 3.)

Step 3: The treatment team completes a thorough evaluation to look for treatable or reversible causes of the target problems

- Primary care clinician or psychiatric consultant completes history and examination (including pertinent neurologic and detailed mental status examination)
- Complete team discussion to identify treatable or reversible causes
- Order laboratory and other tests as necessary, and results reviewed or testing not required

Step 4: Treat agitation and/or emotional distress with an individualized treatment plan and without psychopharmacologic intervention

- Correct identified causes of target problems
- Team discussion identifies and institutes individualized psychosocial environmental interventions

Step 5: Assess response to treatment

- Adequate improvement and/or the resolution of target problems
- Lack of adequate response to treatment (Proceed to step 6.)

Step 6: Attempt a trial of pharmacologic interventions other than antipsychotics

- Attempt a trial of acetaminophen if there are no contraindications

- Attempt a trial of citalopram (up to 20 mg/day) or escitalopram if there are no contraindications
- Attempt a trial of low-dose trazodone if there are no contraindications
- Attempt trial of a combination of the above medications if there are no contraindications

Step 7: Assess response to treatment

- Adequate improvement and/or resolution of target problem(s)
- Lack of adequate response to treatment (Proceed to step 8.)

Step 8: Make a decision to prescribe antipsychotic medications

- Obtain consent from surrogate decision maker for the use of antipsychotics after adequately explaining the risks and potential benefits of their use

Step 9: Attempt trial of antipsychotic medications

- Attempt trial with one of three atypical antipsychotics: brexpiprazole, risperidone, olanzapine, quetiapine, aripiprazole[a]

Step 10: Assess response to treatment

- Adequate improvement and/or resolution of target problems without significant adverse effects
- Adequate improvement and/or resolution of target problems but associated with significant adverse effects (Proceed to step 11.)
- Lack of adequate response to treatment (Proceed to step 11.)

Step 11: Attempt trial of another of the atypical antipsychotic medications mentioned above

- Adequate improvement and/or resolution of target problems (Proceed to step 13.)
- Lack of adequate response to treatment (Proceed to step 12.)

Step 12: Team discussion (including with surrogate decision maker) regarding further care

- Decide to transfer the resident to a local Emergency Department for acute inpatient psychiatric treatment or to transfer the resident directly to an acute inpatient psychiatric unit in a hospital (preferably a geropsychiatric unit)

- If in keeping with the residents' values and wishes, decide to switch to palliative care (with more liberal use of antipsychotics and, if necessary, short-acting benzodiazepines such as lorazepam to relieve the burden of intractable suffering, even at the cost of excessive sedation) as the resident is in severe to terminal stages of dementia and burdens of transfer to Emergency Department and hospitalization are high and potential benefits low

Step 13: Attempt to reduce and discontinue antipsychotics after 3–6 months of stabilization

- Attempt to reduce and discontinue antipsychotic medication

Step 14: Assess response to dose reduction and discontinuation

- Lack of relapse of target symptoms
- Relapse of target symptoms (Proceed to step 15.)

Step 15: Reinstitute antipsychotics and attempt Step 13 after 3–6 months of stabilization

---

Note: The psychiatric consultant may need to skip steps 5 and 6 if the resident who has a major neurocognitive disorder has behavioral and psychological symptoms that are posing an immediate danger to self and or others. In behavioral emergencies (e.g., violent behavior), immediate use of one or two doses of antipsychotic medication may be necessary to stabilize the emergency situation before stepped-care approach is initiated. If, during step 3, the team concludes that the resident qualifies for end-of-life care, proceed to palliative care directly. Step 12 can be instituted at any point if there is a radical change in the resident's clinical status. Collaboration with a pharmacist is encouraged at every step and in creating a facility-wide initiative to reduce inappropriate use of antipsychotics.

[a] In residents with Parkinson disease dementia or Lewy Body Dementia, Pimavanserin (first line) should be considered for the management of severe psychotic symptoms with quetiapine (second line) as a second-line option. Clozapine may also be considered for treatment-refractory Parkinson disease psychoses.

Case example 2: Mr. M is a 76-year-old veteran living in a rural nursing home in Idaho. His wife is very devoted to him and his primary support. Mr. M has advanced dementia, probable Alzheimer type. Due to the COVID-19 pandemic, the nursing home had to go on lockdown for several weeks. Mr. M became increasingly sad, irritable, and withdrawn. He developed a reduced appetite and lost 15 pounds over 5 weeks. He stopped coming out of his room. Staff instituted a comprehensive individualized pleasant activity schedule that included the following: more frequent phone calls and daily video chats with his wife (this helped his wife cope with stress also), scheduled time to watch his favorite movies on TV (especially Westerns), his favorite food and drinks every day, window visits with his wife three times a week, daily one on one visits with activity therapist (with physical distancing—staying 6 feet away) for 10 minutes where they would engage in a variety of activities depending on his preference (e.g., coloring, listening to music, watching videos of his pet dog that his wife had provided to the nursing home), and bright light therapy 20 minutes daily in the morning. Mr. M has significant receptive and expressive aphasia, due to which cognitive strategies (e.g., reminiscence) could not be done. Once the lockdown status was lifted, his wife was allowed to visit him in person as part of compassionate care visits. Gradually, Mr. M's mood improved, his appetite and weight improved, and he started coming out of his room and smiling more. Initiating antidepressant mirtazapine was considered, but the team, in collaboration with his wife, decided to try nonpharmacologic interventions first.

Clinical pearl: Individualized pleasant activity schedule is an effective intervention to treat depression in individuals with dementia.

Case example 3: Mr. C is an 80-year-old male, married for 55 years, a retired schoolteacher, and father of three children. He has had MNCD, probable Alzheimer type, for the last 8 years. Mr. C had been experiencing severe agitation and aggressive behaviors over the last several months. His wife reported that Mr. C would be “Dr. Jekyll and Mr. Hyde,” at times pleasant and nice but then suddenly would “have a look of hatred” toward her and accuse her of infidelity. Mr. C had been physically aggressive toward his wife, which prompted admission to an assisted living (AL) home. At the AL, Mr. C attempted to choke a female resident, twisted the arm of a female staff member, and made several threats that he would “destroy” the place. Mr. C had also voiced paranoid thoughts that “they” (meaning the staff and residents) were trying to steal his money and belongings. Mr. C’s PCP had done a thorough workup and found no reversible cause, so he had prescribed escitalopram (10 mg daily), donepezil, and as-needed alprazolam. The staff had attempted various individualized psychosocial environmental interventions (e.g., taking him for walks, playing his favorite music in his room), but the aggressive behaviors persisted. At this point, psychiatric consultation was requested to facilitate transfer to a local acute inpatient geropsychiatric unit. After completing a thorough assessment and review of records, the psychiatric consultant encouraged the facility to give a trial of antipsychotics before transferring Mr. C to an acute inpatient psychiatric unit. The facility agreed. The psychiatrist recommended initiation of olanzapine 5 mg at bedtime, discontinued alprazolam, started scheduled acetaminophen as there were some reports from the staff that Mr. C occasionally complained of lower back pain, and added short-term (for 1 week) p.r.n. orally dissolvable 5 mg olanzapine up to twice a day. Over the next 2 weeks, Mr. C’s aggressive episodes improved modestly, and the staff was able to engage him in more individualized psychosocial environmental interventions. After 4 weeks, his wife and the staff felt that Mr. C was much better, although he still had episodes of anger.

Clinical pearl: Judicious use of atypical antipsychotics may help reduce aggressive behaviors by residents who have MNCD and prevent a transfer to the ED.

Depressive Disorders

The prevalence of depressive disorders is high in LTC residents. Hence, the authors recommend assessing all residents for depression within 1 month of admission to an LTC facility and every 6 months thereafter. Residents who have a history of depression, psychiatric hospitalization, and/or antidepressant treatment have shown the strongest association with new onset of depression after admission to an LTCF and thus would benefit from closer monitoring. The period of adjustment to living in LTC can range from a few weeks to 6 months as residents cope with multiple losses, including independence, home, privacy, possessions, and control over their personal schedule. Additional challenges to living in LTC include interaction with peers who have severe MNCD. High staff turnover and a lack of robust activity programming may also contribute to depression.

Standardized scales recommended for assessing depression in LTC residents include the Geriatric Depression Scale, used for residents who have mild cognitive impairment or mild dementia, and the Cornell Scale for Depression in Dementia, used for residents who have moderate to severe dementia. The Patient Health Questionnaire—9 or the Saint Louis University “AM SAD” (appetite, mood, sleep, activity, thoughts of death) may be used in place of the Geriatric Depression Scale.

Workup should include identifying medications that may cause depression (e.g., interferon, propranolol), medical conditions (e.g., pain, hypothyroidism, hyperthyroidism), and neurologic conditions (e.g., stroke) that may cause or contribute to depression. Androgen deficiency syndrome (testosterone deficiency) may cause depression in older men and should be considered in the differential diagnosis. Apathy and hypoactive delirium are often misdiagnosed as depression.

Treatment of depression begins with psychoeducation, rational deprescribing, and optimizing the treatment of medical and neurologic conditions.

Psychosocial and environmental strategies are appropriate and necessary and may be sufficient for the treatment of mild to moderate depressive disorder. Such interventions include psychotherapy (counseling), physical exercise, behavioral activation therapy with an individualized pleasant activity schedule, music therapy, bright light therapy, increasing social support, and complementary and alternative treatments such as yoga, Tai Chi, meditation, massage therapy, journaling, gratitude-based interventions, and aromatherapy. Psychotherapy with the LTC population primarily involves supportive psychotherapy (including reminiscence therapy and life review) with problem-solving strategies (especially for residents who have executive dysfunction) and/or interpersonal psychotherapy components (especially for residents coping with grief or family conflict). With more-capable residents, cognitive behavior therapy may also be appropriate. Meaning-centered psychotherapy is recommended for LTC residents experiencing demoralization. Malnutrition, dehydration, constipation, and pain are common comorbid conditions among residents who have depression and need to be optimally treated for successful outcomes.

Although rigorous studies have not consistently found antidepressants to be effective for the treatment of depressive disorders among LTC residents, most experts and expert consensus guidelines support the use of antidepressants to treat moderate to severe depressive symptoms. The starting doses of antidepressants need to be low, but final effective doses may need to be comparable to those used by younger adults. One of the SSRIs, SNRIs, or mirtazapine is the preferred first-line treatment for depression if the psychiatric consultant opts for pharmacotherapy. Bupropion and nortriptyline are usually the second line of treatment. SSRIs are generally used as first line and SNRIs as second line for the treatment of depression in individuals with PD. Nortriptyline may be preferred over other antidepressants to treat depression for some residents who have PD. The psychiatric consultant should also consider prescribing pramipexole to treat depression for residents who have PD; it may even be preferred over nortriptyline and other antidepressants, as it also improves motor symptoms of PD. The psychiatric consultant may consider prescribing short-term use of low-dose methylphenidate to treat depression with severe apathy to speed up the response to an antidepressant, especially in the context of rehabilitation if the resident is not participating in physical therapy. The psychiatric consultant should avoid prescribing monoamine oxidase inhibitors as well as antidepressants with high anticholinergic activity, such as amitriptyline, imipramine, and doxepin, because of the high risk–benefit ratio. For residents who have an inadequate response to an antidepressant, the psychiatric consultant should consider adding transcranial magnetic stimulation for milder symptoms and aripiprazole as an adjuvant to treat moderate to severe major depressive disorder. Quetiapine and brexpiprazole may also be considered for treatment-resistant depression, although aripiprazole has been studied more rigorously in older adults with major depression compared to quetiapine and brexpiprazole. Electroconvulsive therapy (ECT) is the treatment of choice for severe major depression, especially in medically complex patients, older adults, and for treatment-resistant symptoms. ECT is faster acting than any other intervention for severe depression and can be lifesaving in individuals with severe malnutrition due to severe depression and individuals at high risk of suicide. Intranasal esketamine spray and intravenous ketamine may be considered in select individuals (e.g., individuals with severe depression receiving end-of-life care). Lithium augmentation is effective for treatment-resistant major depression, but its use is limited in the LTC population because of high medical comorbidity (e.g., significant kidney disease and cardiac arrhythmias). T3 thyroid supplementation is also an option for the treatment of major depression not responding adequately to antidepressant therapy. Optimizing hypothyroidism treatment (e.g., getting to a TSH level between 1 and 2) may also improve depression outcomes. A combination of antidepressants and antipsychotics (e.g., sertraline plus olanzapine) may be considered for psychotic depression. Psychotic symptoms in the context of major depression are more common in older adults compared to younger adults. All residents with major depression need to be closely screened for psychotic symptoms as morbidity and mortality are high, and a lack of response to antidepressants is very likely. ECT is the most effective treatment for psychotic depression in older adults.

Case example 4: Mr. L is an 88-year-old widowed male, a retired construction worker living in a nursing home. He has a long history of alcohol addiction and alcoholic liver disease. Psychiatric consultation was made, as the staff felt Mr. L was depressed. The staff reported that for the last several weeks, Mr. L had been refusing to get out of bed, was not eating well, lost 15 pounds, did not enjoy visits from his grandchildren, and had voiced multiple somatic complaints, including dizziness. During the evaluation, the psychiatric consultant found out that Mr. L had had multiple recent losses: two of his children had died of heart disease in the last year, and his pet dog had died 6 months ago (just before he moved into the nursing home), and 3 months ago his house was sold to pay for the nursing home. Mr. L became tearful during the interview. Workup for reversible conditions detected mild vitamin B12 deficiency and mild dehydration. There was no evidence of MNCD, and Mr. L was open to recommendations from the psychiatric consultant, including antidepressants. The psychiatric consultant recommended vitamin B12 supplementation, improved intake of daily fluids, and weekly individual psychotherapy with a psychologist who visits the facility; the psychiatrist did not recommend antidepressants. Over the next 6 weeks, Mr. L showed significant improvement. He started to come out of his room, participated in activities, improved his food intake, gained 2 pounds, and would look forward to visits from his grandchildren.

Clinical pearl: Psychotherapy should be routinely offered to treat depression for cognitively capable residents, and LTC facilities should arrange with local psychotherapists to visit the facility and provide psychotherapeutic services.

Psychotic Disorders

Psychotic symptoms in the context of MNCD are the most common psychotic disorders in the LTC population, followed by psychotic symptoms due to medication and due to general medical and neurologic conditions (e.g., PD, stroke, traumatic brain injury, COVID-19). Dopaminergic agents used to treat PD, anticholinergics, and corticosteroids are the three most common categories of medication that cause and or worsen psychotic symptoms. Psychotic symptoms in the context of affective disorder (major depression as well as bipolar disorders), schizophrenia, schizoaffective disorder, and delusional disorder are becoming increasingly prevalent due to the aging of adults who have a chronic mental illness. Misdiagnosis of schizophrenia, bipolar disorder, and schizoaffective disorder is prevalent in LTC populations. Psychotic symptoms may also be the initial or most disturbing manifestation of delirium. Psychotic symptoms are among the most serious complications of MNCD and are associated with a high degree of emotional distress and accelerated cognitive decline. The psychiatric consultant should consider the new onset of psychotic symptoms among LTC residents as equivalent to delirium, with a thorough workup, including, as appropriate, neuroimaging.

Psychotic symptoms associated with MNCD are typically mild to moderate in severity and are best managed with psychosocial/environmental interventions. Antipsychotics have not been shown to reduce psychotic symptoms related to neurocognitive disorder, and the authors do not recommend their use unless psychotic symptoms are accompanied by severe aggression and or agitation. Management of psychotic symptoms in the context of PD is challenging, as the majority of residents who have PD have PD-related MNCD. Furthermore, only pimavanserin and clozapine have been found to be effective for the treatment of psychotic symptoms in the context of PD. Pimavanserin is approved by the U.S. FDA and is generally preferred over clozapine, although the cost is a significant concern with the use of pimavanserin. The use of clozapine carries even higher morbidity and mortality risk for LTC residents who have PD than for community-dwelling individuals who have PD due to frailty and the presence of significant cognitive impairment. The first strategy is to reduce the dosage of antiparkinsonian agents (such as anticholinergics, amantadine, catechol-O-methyl transferase (COMT) inhibitors, and dopamine agonists with Levodopa reduction as a last resort) and use psychosocial environmental interventions to manage psychotic symptoms. If psychotic symptoms are causing severe distress and agitation, the psychiatric consultant may consider prescribing pimavanserin. For residents who have neurocognitive disorder due to Lewy body disease and are experiencing psychotic symptoms, a trial of rivastigmine or donepezil may help both cognitive and psychotic symptoms (especially visual hallucinations) to a modest extent.

Case example 5: Ms. T is a 71-year-old divorced female, a homemaker, and mother of one daughter and one son, who has a long history of schizophrenia. She was admitted to a nursing home after a stroke. Psychiatric consultation was requested to adjust Ms. T's antipsychotic medication, as she was refusing to participate in physical therapy and making paranoid statements. After completing a thorough evaluation, including a review of records from previous psychiatric hospitalization, the psychiatric consultant confirmed the diagnosis of schizophrenia. Ms. T also had mild neurocognitive disorder due to cerebrovascular disease. The psychiatrist noticed that many staff members felt that Ms. T should be transferred to another facility as they were not equipped to manage chronic mental illness. Ms. T's primary care clinician was also not comfortable managing chronic mental illness by adjusting antipsychotics. Ms. T was taking quetiapine 200 mg daily at bedtime and risperidone 0.5 mg once daily. The psychiatrist increased risperidone to 0.5 mg twice daily for the first week and then 1 mg twice daily and decreased and discontinued quetiapine over 2 weeks. Ms. T showed significant improvement in psychotic symptoms over the next 3 weeks and began to participate in physical therapy.

Clinical pearl: The majority of LTCFs are not prepared to meet the mental health needs of residents with severe mental illness. Psychiatric consultants can play a key role in improving their acceptance by facility staff and help improve the staff's understanding of these illnesses and knowledge about evidence-based strategies to treat them.

Disorders of the Sleep–Wake Cycle

Insomnia, excessive daytime sleepiness, and reversal of the night–day cycle are common among LTC residents. Medications are the most common cause of disorders of the sleep–wake cycle (e.g., SSRI worsening restless leg syndrome [RLS]). RLS and REM sleep behavior disorder (RBD) are prevalent in the LTC population, are often underdiagnosed, and are amenable to treatment. Treatment for insomnia involves discontinuing medications causing or contributing to insomnia (when feasible), optimizing treatment of physical health conditions (e.g., hypothyroidism), and psychosocial environmental interventions (e.g., reducing daytime napping, increasing physical activity, and exposure to sunlight or bright light, following a regular sleep time, soothing bedtime rituals, reducing exposure to noise and light at night). Cognitive-behavioral therapy for insomnia (CBT-Insomnia) and Brief Behavior Therapy for Insomnia (BBTI) are evidence-based interventions for chronic insomnia and are recommended as first-line treatment. Suvorexant is approved by the FDA for the treatment of insomnia related to Alzheimer dementia. Ramelteon and suvorexant may also be considered for the treatment of chronic insomnia. They are safer than other sedative-hypnotics, but the cost is a barrier to their use. The psychiatric consultant may consider prescribing low-dose trazodone (25 to 50 mg) to treat persistent insomnia not responding to psychosocial environmental interventions. The use of hypnotics such as zolpidem, zaleplon, and eszopiclone should be restricted to low doses for short periods of time (days to 2 weeks) (e.g., for a new resident adjusting to living in an LTCF and experiencing significant insomnia) due to a high risk–benefit ratio and regulatory concerns. The psychiatric consultant may consider prescribing low-dose clonazepam and melatonin to treat RBD and may consider prescribing dopamine agonists and gabapentin enacarbil to treat RLS. Dopamine agonists are often used to treat RLS but may cause psychotic symptoms or agitation. Switching to gabapentin, correcting an iron deficiency, leg massages, and evening walking programs may be safer alternatives to dopamine agonists for RLS. Free high-quality apps such as CBT-i Coach from the Veterans Administration are recommended to educate and train the individual in relaxation exercises

Posttraumatic Stress Disorder

Posttraumatic stress disorder (PTSD), nightmare disorder, and subsyndromal PTSD are increasingly seen in the LTC population, not only among veterans but also among civilians who have a past history of abuse or motor vehicular accident, or recent stay in an intensive care unit. LTC staff need education to understand psychological and behavioral manifestations of past trauma and training in care practices to accommodate trauma-related challenges residents typically face. For example, male staff members may not be involved in caring for a female resident who has a history of sexual or physical abuse by a male. Physical restraints are not recommended in general but may have additional negative psychological effects on residents who have a history of abuse. The treatment of PTSD typically involves trauma-centered psychotherapy and, if necessary, antidepressants (especially SSRIs) for anxiety and depression and trazodone for insomnia. Free high-quality apps such as PTSD-coach from the Veterans Administration are recommended to educate and train individuals with PTSD in relaxation exercises. Short-term or as-needed use of short-acting benzodiazepines may be necessary for residents who have PTSD and are experiencing acute anxiety attacks. Low-dose prazosin may be appropriate for the treatment of nightmares in the context of PTSD, although LTC residents are at higher risk than the general population of adverse effects of prazosin, such as hypotensive effects, orthostasis, and falls.

Anxiety Disorders

Generalized anxiety disorder (GAD), panic disorder, and phobias, especially bathing phobia and fear of falling, are prevalent in the LTC population. A bathing phobia is especially common among residents who have MNCD and is best managed by staff training in creative and alternative ways to maintain hygiene using strategies such as towel bathing instead of using a shower. The authors recommend the DVD “Bathing without a Battle,” created by the University of North Carolina, for training all staff members in person-centered strategies to bathe residents who have advanced MNCD. Treatment of GAD and panic disorder typically involves cognitive-behavioral therapy for anxiety (CBT-anxiety), relaxation exercises, and, if necessary, antidepressants (especially SSRIs). Free high-quality apps such as Cleveland Clinic Wellness Mindful Moments and UCLA Mindful are recommended to educate and train the individual in relaxation exercises.

PSYCHIATRIC ASPECTS OF PALLIATIVE CARE

More than 25% of older adults currently die in LTCFs, and the number is expected to rise to more than 40% by 2030. The majority of residents have at least one life-threatening illness, and at least 50% of these have a life expectancy of 2 years or less. Thus, it is imperative that all these residents admitted for long-term stay be offered palliative care as soon after admission as possible. Some residents admitted for short-term stay for rehabilitation may also be appropriate for education about palliative care. Psychiatric consultants can help the staff appreciate that most residents and their family members find the goal of increasing comfort much more satisfying and reasonable than aggressive life-prolonging care. The psychiatric consultant can play a key role in clarifying a resident’s advance care directives, especially in the context of advanced MNCD, where it may be difficult to predict life expectancy, but the burdens of traditional medical care may still be high.

Delirium and depression are two psychiatric syndromes residents frequently experience at the end of life. Psychiatric consultants may be the first health care professionals to identify the resident as in the process of dying and help initiate palliative and hospice care. During end-of-life care, it is appropriate and often necessary to prescribe benzodiazepines to manage anxiety and insomnia, morphine to manage pain and dyspnea, and atypical antipsychotics such as risperidone and olanzapine to manage psychotic symptoms associated with agitation or aggressive behavior. The use of methylphenidate to treat depression during end-of-life care has been recommended for its rapid effect, but among LTC residents, who often have advanced MNCD, methylphenidate may cause delirium or agitation, and thus the psychiatric consultant should closely monitor its use. Palliative sedation may be necessary to relieve the intractable suffering some residents experience during the end stages of life.

Another common reason for referral to a psychiatrist is to evaluate decision-making capacity. Some common clinical situations for such a referral are a resident who has end-stage renal disease and wants to discontinue dialysis; a resident who has depression and is requesting a Do Not Resuscitate status.

FURTHER READINGS

American Geriatrics Society 2023 Beers Criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc*. 2023;71:2052–2081.

American Psychiatric Association. *American Psychiatric Association Practice Guideline for the use of Antipsychotics to Treat Agitation or Psychosis in Patients with Dementia*. The American Psychiatric Association Publishing; 2016.

Bethell J, Aelick K, Banineau J, et al. Social connection in long-term care homes: A scoping review of published research on the mental health impacts and potential strategies during COVID-19. *J Am Med Dir Assoc*. 2021;22:228–237.

Brodaty H, Aerts L, Harrison F, et al. Antipsychotic deprescribing for older adults in long-term care: The HALT study. *J Am Med Dir Assoc*. 2018;19:592–600.

Desai AK, Grossberg GT. Geriatric psychiatry. In: Sinclair AJ, Morley JE, Vellas B, Cesari M, Munshi M, eds. *Pathy's Principles and Practice of Geriatric Medicine.* 6th ed. John Wiley & Sons Publishing 2022;904–918.

Desai AK, Lo D, Grossberg GT. Depression in older adults receiving hospice care. In: *Late-life Mood Disorders*. In: Lavretsky H, Sajatovic M, Reynolds III CF, eds. Oxford University Press; 2013.

Desai AK, Grossberg GT, McFadden S. Understanding experiences and perspectives of persons with dementia. *Dementia Care: An Evidence-Based Approach*. In: Boltz M, Galvin JE, eds. Springer Publishing; 2016.

Desai AK, Grossberg GT. *Psychiatric Consultation in Long-Term Care: A Guide for Health Care Professionals*. 2nd ed. Cambridge University Press; 2017.

Dunn LB, Cassidy-Eagle EL. *Practical Strategies in Geriatric Mental Health. Cases and Approaches*. American Psychiatric Association Publishing; 2020.

Han EJ, Park M, Park S, Giap TTT, Han D. Randomized controlled trial of the caregiver orientation for mobilizing personal assets and strengths for self-care (COMPASS) for caregiving journey: A national family caregiver support program in a long-term care insurance system. *J Am Med Dir Assoc*. 2020;21:1906–1913.

Hui C, Sulzer D. Depression in long-term care. In: *Late-life Mood Disorders*. Lavretsky H, Sajatovic M, Reynolds III CF, eds. Oxford University Press; 2013.

Kotlyar M, Gray SL, Maher Jr RL, Hanlon JT. Psychiatric manifestations of medications in the elderly. In: Agronin ME, Maletta GJ, eds. *Principles and Practice of Geriatric Psychiatry*. 2nd ed. Lippincott Williams & Wilkins; 2011.

McFadden S. *Dementia Friendly Communities. Why We Need Them and How We Create Them*. Jessica Kingsley Publications; 2021.

Taylor WD, Schulltz SK, Panaite V, Steffens DC. Perspectives on the management of vascular depression. *Am J Psychiatry*. 2018;175(12):1169–1175.

The Gerontological Society of American. Dementia-Related Psychosis: Gaps and Opportunities for Improving care. Report. August 2019.

Thorn RP, Levy-Carrick NC, Bui M, Silbersweig D. Delirium. *Am J Psychiatry*. 2019;176(10):785–793.

Vahia IP, Jeste DV, Reynolds III CP. Older adults and the mental health effects of COVID-19. *JAMA*. 2020;324(22):2253–2254.

Veldwijk-Rouwenhorst AE, Smalbrugge M, Zuidema SU, Hanssen SAJ, Koopmans RTCM, Gerritsen DL. Continuous palliative sedation in nursing home residents with dementia and refractory neuropsychiatric symptoms. *J Am Med Dir Assoc*. 2021;22:305–311.

Weintraub D, Mamikonyan E. Impulse control disorders in Parkinson’s disease. *Am J Psychiatry*. 2019;176(1):5–11.

55.6b Forensic Aspects

David Naimark, M.D., Keith A. Davidson, J.D., Ariana Nesbit, M.D., M.B.E, and Ansar M. Haroun, M.D. (Retired)

The American Board of Psychiatry and Neurology has recognized both geriatric psychiatry and forensic psychiatry as subspecialties, each meriting its own fellowship training program and subspecialty board examination. The field of “forensic geropsychiatry” is where these two subspecialties overlap, and this subspecialty is growing in its knowledge base and literature. Unlike clinical geropsychiatry, which evolves slowly based on the natural progression of knowledge in the basic psychiatric sciences, forensic geropsychiatry may grow rapidly with abrupt changes based on new laws. Whereas geriatric psychiatry is generally defined by a demographic characteristic (that of age), forensic psychiatry is defined by its ethical duty, to tell the truth, regardless of the effect the psychiatrist’s truth-telling may have on the individual they’re evaluating. This contrasts with the duty of clinical psychiatry, where the clinician’s primary ethical obligation is to the patient’s welfare.

CRIMINAL ASPECTS OF FORENSIC GEROPSYCHIATRY

Epidemiologically, there is a relationship between criminal behavior and age, with an overrepresentation of youthful offenders and an underrepresentation of older adult offenders. Although criminal behavior remains a primary problem for the young, criminality is increasing more rapidly in older adults. This trend may partially be accounted for by the population increase in the total number of older adults; however, there may be an independent increase in criminal behaviors in older adults for other reasons. Possible explanations include the improved efficiency of law enforcement to detect crimes or new patterns of crimes (such as Internet-based crimes) that may be easier for physically disabled elders to commit.

It can be useful to look at the relationship of age to subtypes of criminal behavior. One typology of criminal behavior contrasts crimes that are affective (based on emotion, impulsive) with those that are predatory (coldly calculated, premeditated). While youthful offenders are often hotheaded and at high risk for the former, the older adults are more mature, are less hotheaded, and may be at greater risk for the latter.

The expertise of a geriatric psychiatrist may be needed at several stages of a criminal proceeding. For example, a forensic geropsychiatrist may be asked to opine as to whether the defendant understood their legal rights when they were arrested. Officers are required to inform arrestees of their right to avoid self-incrimination (i.e., to remain silent) as well as their right to counsel. Older individuals or any other persons who may be confused or are cognitively impaired are at risk of not understanding these rights and their associated warnings. If the arrestee does not understand these rights, any statements made in response to police questioning may not be admissible in a court proceeding. In other words, the defendant’s understanding (or lack of understanding) can change the outcome of a legal case by allowing (or disallowing) the statements or confessions. It may be difficult for a psychiatrist to determine whether the defendant understood their rights because many elders are in a gray zone of partial understanding. As a result, what may be relevant was whether the elder spoke to the police voluntarily or through coercion. Although it may be rare for the police to coerce using physical pain, officers may be tempted to coerce using psychological pressures to which the older adults may be especially susceptible (see the following discussion of undue influence). A practical problem that may arise from such possible coercion is that the elder may “confess.” If the confession is voluntary, it is, of course, admissible, but if coerced, it may be challenged. Analysis of the voluntariness of such a confession requires knowledge not only of the behavior of those who elicited the confession (usually the police) but also of the unique psychological and physical strengths and frailties of the confessor.

Competency to Stand Trial

When a criminal trial starts, the court must be satisfied that the defendant is not only physically present but also mentally present and, therefore, able to participate in the trial process (i.e., mentally competent to stand trial). The constitutional competence to stand trial standard requires the defendant to be able to understand the criminal proceedings and to assist their attorney in their own defense in a rational manner. Based on these requirements, a psychiatrist evaluating a defendant’s competence to stand trial will give an opinion on the cognitive status of the elder (whether there is any deficit in *understanding* the nature of the criminal proceedings), the social capacities of the elder (ability to *assist* an attorney), and the rationality of the elder (able to understand and assist in a *rational* manner). Furthermore, in most states, if the psychiatrist detects a deficit in one of these abilities, the evaluator must opine whether the deficit is attributable to a mental disease—thereby making the defendant incompetent to stand trial—or from some other cause. For example, an elder who cannot assist their attorney because of a hearing loss is “audiologically” incompetent but not mentally incompetent. Similarly, psychiatrists must carefully consider the etiology of an older adult defendant’s forgetfulness. If the defendant’s memory deficits are attributable to a mental disorder such as a neurocognitive disorder, then the defendant may postpone or even avoid trial. In contrast, if the forgetfulness is attributable to expected age-related cognitive decline, then it will not qualify for a finding of mental incompetence.

JO, a 72-year-old man with dementia, had a superimposed psychotic disorder at the time of the trial. He had numerous delusions and hallucinations. He seemed to have a basic understanding of the trial process, with the exception of two unusual beliefs: While he could tell the evaluator that the judge oversees the trial and tells the jury the law and the jury is the body that decides, after hearing the evidence, whether he is guilty or innocent, he also had the belief that the judge was an “interloper” and the jury members were “four-footed animals.”

Is this defendant incompetent to stand trial? JO’s competency to stand trial is likely to be controversial. Is his understanding of the function of judge and jury adequate? Does it matter if his conception of the judge and jury is irrational if he knows their function? On the other hand, perhaps he thinks their status as interlopers and animals will make them perform their jobs in a severely substandard way—a way that will deprive him of due process.

Competence to Stand Trial Restoration

If the court finds that a defendant is incompetent to stand trial, the defendant is typically committed to an inpatient treatment setting for competency restoration. Competency restoration generally involves not only the treatment of the individual’s mental health disorder but also education about the legal system and the defendant’s case.

Prior to this treatment beginning, the forensic psychiatrist may be asked to opine as to whether the court may or may not order involuntary medications in order to restore the defendant’s competence to stand trial. In order to order involuntary medications for this purpose, the following criteria must be met:

1. An important government interest is at stake.
2. The medication significantly furthers the state’s interest (e.g., the medication should be likely to restore the defendant’s competence to stand trial and not have severe side effects that would interfere with the trial competency).
3. The medication must be the most appropriate and least intrusive method of restoring the defendant’s competency.
4. The medication must be medically appropriate.

In geriatric cases, the forensic evaluator must be especially careful to balance the risks of medication side effects against the likelihood that the medication will restore the defendant’s competence to stand trial. Geriatric patients may have different responses to medications, which must be separately analyzed in each case. In the case of the elder who has many medical conditions complicating the side-effect profile of the intended psychotropic medication, a different conclusion may be reached than in the case of a nonelder, making it more likely that the court may reject the application for involuntary medication. Some individuals may be restored to competence without involuntary medications. However, if pharmacotherapy is required to restore competence to stand trial, but the court rejects the application allowing involuntary medication, the defendant may be determined to be “permanently incompetent.” Similarly, if the defendant’s incompetence is attributable to a mental health disorder not easily treatable with psychotropic medications (e.g., major neurocognitive disorder due to Alzheimer disease), they may also be deemed permanently incompetent.

Currently, such cases of permanent incompetence are rare because most cases of trial incompetence arise from psychotic disorders, which usually respond to antipsychotic medications.

However, as more older adults enter the criminal justice system, and as more older adults are found to be incompetent to stand trial, it is likely that there will be more cases of trial incompetence from cognitive rather than psychotic disorders. Because some cognitive disorders are reversible and some are not, geriatric psychiatrists will increasingly be asked to separate these two subpopulations. Ultimately, there will be a large group of older adult persons who are permanently incompetent to stand trial. The Supreme Court has ruled that the state cannot hold a person incompetent to stand trial for more than the reasonable period of time necessary to determine whether there is a substantial probability that they will attain that capacity in the foreseeable future. In other words, once a defendant is determined to be permanently incompetent, they can no longer be held under the guise of competence restoration. That being said, permanently incompetent defendants may be able to be civilly committed, which would allow many to remain hospitalized based on their mental disorder, their dangerousness, and, in many states, their grave disability.

Mental Health Defenses

Most criminal defendants are immediately or eventually found competent to stand trial and are, therefore, able to work with their attorney to put forth a defense. During the criminal proceedings, competent defendants may assert any one of several mental health defenses. Mental health defenses vary across jurisdictions, with not guilty by reason of insanity (NGRI) being the most common.

An NGRI defense requires a confession of behavioral guilt ("I did it") but an absence of moral guilt based on a cognitive—and in some jurisdictions, volitional—mental deficit. The most common test of insanity in the United States is known as the M'Naghten standard. The M'Naghten standard states that to establish insanity, the defendant must prove that, at the time of the offense,

> the party accused was laboring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or if he did know it, that he did not know he was doing what was wrong.

Because older adult defendants are at a greater risk for cognitive deficits, the geriatric psychiatrist should clarify the legal standard in the jurisdiction of practice, especially noting what types of cognitive deficits represent a condition qualifying for an NGRI defense. Although amnesia for the alleged offense is common in older adults, such memory deficits will not qualify for an NGRI defense. In contrast, cognitive deficits in the areas of understanding wrongfulness are more likely to be excused. For example, if a defendant with a neurocognitive disorder fails to recognize his wife and believes that she is an intruder trying to kill him, and if he injures her in self-defense, he may be found NGRI because he did not know that what he was doing was wrong.

Restoration of Sanity

If an elder is indeed found to be NGRI, the state is legally required to make efforts to restore sanity. Generally, a defendant is considered to be restored to sanity when they will no longer pose a danger to the safety of others due to their mental health disorder. In the case of youthful offenders, most of the NGRI acquittals are based on treatable psychotic disorders. But in the case of the elder, the acquittal may be based not on a treatable psychotic disorder but rather on a neurocognitive type, which may or may not respond to treatment. If the defendant's cognitive decline is medically reversible, the court should be informed so that they can order the appropriate treatment. However, if the defendant has an irreversible neurocognitive disorder, this will create a problem for the court: Although the defendant may be confined for treatment, they may not be confined in the absence of treatment because no punishment may be imposed on a defendant who is "not guilty." That being said, in many court cases, the right to treatment is extremely minimal; for example, enough to help keep the defendant out of restraints. In other words, the fact that certain people are effectively untreatable does not mean that they cannot be confined if they are dangerous or gravely disabled. Notwithstanding this, a skillful defense attorney may argue that the elder must be treated or released, thus creating a dilemma for the court.

Sentencing

The forensic geropsychiatrist should be aware of the five possible purposes and philosophies of sentencing: retribution, incapacitation, deterrence, rehabilitation, and restoration. If the purpose of the sentence is punitive (to punish), it is uncommon for a court to seek a psychiatric opinion, although a skillful defense attorney may argue that the very punishment that may be appropriate for a youthful offender may be cruel or unusual for an elder, who may have a reduced capacity to tolerate discomfort. If the court instead focuses on restitution, incapacitation, or rehabilitation, the geropsychiatrist can offer an opinion as to the merits of any of these.

More specifically, a forensic geropsychiatrist may be asked to write a presentencing report to help the court assess for the presence of psychiatric factors that would suggest that the defendant is at a higher or lower risk for reoffending. During sentencing, judges may consider age-related factors like the epidemiologic risk of future harm (lower in older adults), the degree of culpability based on gullibility (older adults may be more gullible and hence less blameworthy), the likelihood of physical violence (lower in older adults), and the likelihood of good compliance with medication to achieve remission of mental illness (which may be lower in medically impaired subjects, whose medical status may make tolerating psychotropic medications more difficult).

Judges may also consider “STAR” factors related to the defendant’s decision making. These objective factors should be noted in a presentencing report for a judge. Subjective factors, such as whether the elder did or did not inspire pity in the examining doctor, should not be given weight. If the doctor is aware of positive or negative feelings toward the defendant that may introduce a bias, the doctor should consult with a colleague and either avoid giving an opinion or give the opinion but clarify why these feelings may have interfered with the psychiatrist's ability to provide an objective opinion.

STAR Factors:

- S: Speed of decision making, with the older adults tending to be more reflective and less impulsive.
- T: Time preference, with mature seniors tending to be more future-oriented than more present-oriented and immature youth.
- A: Elders who have practiced autonomy for most of their lives may be less rebellious toward authority and less determined to demonstrate their independence/autonomy than youthful offenders.
- R: Risk estimation is a function of intelligence and may be poor in both immature youth as well as in cognitively impaired elders. That being said, risk-seeking behaviors tend to decrease with age.

CIVIL ASPECTS OF FORENSIC GEROPSYCHIATRY

In contrast to criminal cases, which involve an action that is considered harmful to society as a whole, civil matters involve a private dispute between persons or organizations. Some of the most common civil matters that a forensic geropsychiatrist may be asked to consult on include testamentary capacity (i.e., the capacity to make a will) and financial and contractual capacity. Other types of medicolegal cases that a geropsychiatrist might encounter include those involving advance directives and medical decision making. Although beyond the scope of this chapter, forensic neuropsychiatrists may also be consulted on cases involving elder abuse or hospital or nursing home malpractice.

GENERAL CONCEPTS IN CIVIL FORENSIC GEROPSYCHIATRY

Decision-Making Capacity and Informed Consent

Civil forensic cases generally require the assessment of decision-making capacity and/or undue influence. Decision-making capacity usually involves the ability to understand a body of knowledge and to rationally apply that knowledge to a decision-making process. In contrast, competence is the quality or condition of being legally qualified to perform an act and/or make decisions. Geropsychiatrists assess an individual's capacity as it relates to their competence. In other words, while the doctor analyzes the decisional capacity and forms a penultimate medical opinion, the trier of fact forms a legal judgment about the ultimate issue: competence.

The principle of respect for autonomy directs psychiatrists to respect patients' choices when they are intentional and made with understanding and without controlling influences (i.e., autonomous). However, in patients with mental health disorders, choices may not always be autonomous due to the fact that they are influenced by the internally controlling influence of a mental health disorder. In these cases, the psychiatrist should recognize that the choice is not truly autonomous and therefore override their decision based on the principles of beneficence and nonmaleficence.

Capacity. In clinical and research settings, a psychiatrist may be asked to opine on an individual's capacity to provide informed consent to medical treatment, to be a research subject, and to die. Capacity evaluations are generally done by analyzing four factors:

1. The patient's preference and the stability of that preference over time
2. Whether the patient has a factual understanding of their medical condition and proposed interventions
3. Whether the patient appreciates the implications of their decision
4. The patient's ability to rationally manipulate the information

Decision-making capacity is required for a patient to be able to provide informed consent. Informed consent is a cornerstone of ethical medical practice, and its goal is to protect autonomous choice. The informed consent process can provide a meaningful chance for a physician and elder to participate in a personalized discussion of risks, benefits, and alternatives associated with a decision in question. In this process, the patient becomes a partner with the physician in the management of the situation, thereby engaging in reasoned decision making. Unfortunately, the informed consent discussion may morph into a procedure done to the patient or subject (i.e., “consenting” the individual). This implies that the elder is a passive recipient of the information who then acquiesces to the decision.

In life outside of medicine, issues of decision making present themselves constantly. Unlike in the medical world, however, the rules around the need for the transactor to obtain informed consent from the transactee become markedly murkier (e.g., does a department store employee need to obtain informed consent before he allows an elder to purchase socks?). Although an individual’s sock-buying capacity is not typically evaluated, geroforensic psychiatrists are often asked to opine on an elder’s capacity to make financial decisions and enter into contracts, make a will (i.e., testamentary capacity), and drive.

Over the last half-century, multiple state and federal court decisions have greatly influenced the analysis of the concept of informed consent. What the doctor must disclose is measured by two standards. The older rule, and one contained in some statutes, is that the doctor must disclose what a reasonable doctor would disclose. The newer rule is the “reasonable patient” standard: The doctor must disclose all material information that they could reasonably expect a patient would want to consider in determining whether to undergo the procedure. Informed consent requires not only that the patient is informed but also that they have the capacity to decide and that their decision should be voluntary. The latter was called into question in one landmark case when an involuntarily committed forensic patient was deemed unable to consent to experimental psychosurgery because the inherently coercive institutional environment prevents voluntary consent.

The first element of informed consent refers to "informed" and asks whether the information from the database (regarding the specific decision in question) was transmitted to the decision maker. It does not assess the state of mind of the transactee and, as such, is the responsibility of the transactor. With inadequate or absent information, the consent would be uninformed, not informed.

The second element of informed consent addresses the "sent" (from the same root as *science*, as in knowledge) and requires the elder to have the mental capacity to understand and appreciate the information provided. The best information in the universe is useless to the elder whose cognitive plumbing is clogged to the degree that it simply can't pass through the pipes of mental functioning. It is here that the forensic evaluator will perform the analysis of decision-making capacity through a mental functioning examination. Mental status deficits, in and of themselves, do not lead to mental incapacity to make a decision. It is only when the observed defect prevents the information from penetrating into the awareness of the subject that the defect becomes relevant. Put another way, mental state defects are a "threshold test": some decisions are complicated and require a maximally functional mental state, whereas others are simple and probably require no more than a marginally intact mental state.

The third and perhaps most misunderstood element of informed consent concerns the "con" part of consent. This distinguishes consent from assent and requires the assessment of whether the elder was able to give a voluntary agreement to the transaction. Lack of voluntariness can be as simple as the overt threatening of an elder unless they agree to a decision or as subtle as the insinuation into a susceptible elder's life to the degree that the elder becomes unduly influenced and thereby unable to give voluntary consent. It is important to understand that "influence" is generally construed to be acceptable, as all human beings influence each other repeatedly every day. It is only when the influence progresses to the point that the will of the elder has been substituted by another's will (generally for the purpose of extracting money or other property) that the court may declare that the influence is undue.

Undue Influence. Although there may be considerable disagreement about when undue influence may exist, there are some circumstances where undue influence likely cannot occur. These circumstances include whether the elder (1) knows their own mind; (2) can distinguish between their interests and the interests of another; and (3) can distinguish a neutral, disinterested assertion from an active, persuasive one.

These conditions emphasize cognitive constructs (i.e., knowing and distinguishing) but leave out issues of emotional knowing. If there is no emotional vulnerability leading to emotional dependence, undue influence is unlikely to occur.

The three elements of informed consent clearly overlap (i.e., mental capacity is related to the ability to give voluntary consent), and the distinction is, therefore, somewhat artificial. It is helpful for clarity, though, for the forensic evaluator to separate the elements as much as possible.

Within the informed consent process, much interest is now focused on the process of informing rather than on the content of information. "Nudging" occurs when information is presented in a different manner in order to improve an individual's decision-making process. Previously, it was assumed that competent patients were rational and, therefore, blind to how information was framed. It is now recognized that most individuals are not blind to framing effects but rather influenced by them. For example, when a doctor informs a patient of the risks of medical treatment, the doctor has the choice of framing the information in positive terms ("75% probability of success") or in the alternative ("25% probability of failure"). The two pieces of information are mathematically equivalent, and a rational agent should be indifferent to how the information is presented. However, most patients process these two pieces of information differently and are likely to reach different conclusions depending on the presenting frame. Whether "nudges" are ethical remains controversial and depends on whether the specific nudge supports or undermines the individual's dignity, welfare, and autonomy.

GENERAL PROCEDURE TO ASSESS UNDUE INFLUENCE. The question of undue influence frequently arises in civil geriatric forensic cases. Undue influence can be defined as the pressure applied unfairly by an individual attempting to manipulate or deceive with the intent of causing that person to make a certain decision. Although every jurisdiction has different elements for undue influence, undue influence generally requires an element of coercion that overcomes the free will of the individual being influenced, thereby robbing the individual's autonomy.

EVALUATING FOR UNDUE INFLUENCE. A forensic geropsychiatrist should consider four elements when assessing the presence of undue influence.

1. Vulnerability of the victim
2. The influencer's apparent authority
3. The actions or tactics used by the influencer
4. The equity of the resulting decision or action

Undue influence may impact an elder's ability to make autonomous decisions related to finances, contracting, medical, and will or trust creation.

Vulnerability of the Victim. This element is the most relevant to geropsychiatric involvement because it requires a finding of at least one of the following conditions: incapacity, illness, disability, impaired cognitive function, emotional distress, isolation, or dependency. Typically, it is the judge or jury who will decide whether this element has been met in a legal trial; however, courts rely heavily on medical expert opinions. These opinions are required in most cases for two primary reasons: (1) hearsay; and (2) interpretation of medical information.

First, nearly every diagnostic conclusion or opinion contained in a written medical record is considered hearsay or an out-of-court statement. Therefore, medical records cannot be used as evidence at the time of trial. Instead, a medical expert must review the records and then give an opinion as to the victims' vulnerability or susceptibility to undue influence. Second, oftentimes medical records have legal terms and jargon that are not easily understood by judges or juries. A medical expert can help interpret medical records to enhance (or perhaps decode) their understanding at trial.

It is interesting to note that judges and juries are allowed to accept either party's expert opinion or reject them both. Under the law, the trier of fact (judge or jury) has the final say as to whether a victim was, in fact, vulnerable to undue influence. The law notwithstanding, medical experts provide a valuable service to help the trier of fact reach their final decision.

Influencer's Apparent Authority. An influencer's apparent authority focuses primarily on the relationship between the influencer and the victim. Relationships that create an increased risk of undue influence include fiduciaries, family members, care providers, health care professionals, legal professionals, spiritual advisers, financial advisers, and experts. Unlike the vulnerability element discussed above, this element focuses on the wrongdoer's position of power over the victim.

Actions or Tactics Used by the Influencer. Actions and tactics focus on the bad acts undertaken by the influencer to suppress the free will of the victim. Actions and tactics include controlling necessities of life, medicine, and access to family and friends, using affection, intimidation, or coercion, initiating changes in personal or property rights, using secrecy in making financial changes, or making changes at inappropriate times and places.

Equity of the Decision or Action. Equity is a legal concept that refers to the fairness of the result. If a person has a long history of benefitting family members under their will and then suddenly changes that plan to benefit a caretaker, this could be viewed as an inequitable result under the circumstances. While equity, by itself, is not enough to establish undue influence, an apparent inequitable result can be persuasively used to help support the other elements of undue influence.

PRESUMPTION OF UNDUE INFLUENCE. Most jurisdictions have a general presumption that undue influence exists where: (1) a confidential relationship exists between the influencer and the victim (much like the apparent authority element discussed above); (2) the influencer actively participated in the procurement of the decision or action; and, (3) the influencer unduly benefitted from the decision or action taken. Where a presumption exists, it is then the burden of the influencer to prove the absence of undue influence, which can be difficult to accomplish in most cases.

Competency to Make Medical Decisions or Consent to Be a Research Subject

While the "meat" of the medical/research capacity evaluation is the forensic interview, there are certain standardized assessment instruments that may be useful. More specifically, the MacArthur Competence Assessment Tools have been shown to have good validity and reliability in multiple populations. Included are semistructured interviews for evaluating the four aspects of decision-making capacity, as described in the previous section. Recently, a new instrument for assessing decisional capacity for clinical research was validated: The University of California, San Diego, Brief Assessment of Capacity to Consent (UBACC) is a 10-item scale focusing on understanding and appreciation of information contained in a research protocol. This Assessment was found to have good reliability and validity while taking less than 5 minutes to administer.

Recent research suggests that certain procedures for "enhancing" informed consent for research and treatment may be helpful. Effective interventions include corrected feedback, multiple learning trials, and well-organized simplified consent forms. Even in older adult individuals with psychotic disorders, enhanced consent procedures involving a computerized slideshow have recently been shown to improve understanding of research consent. When older adult persons are incapable of deciding whether to participate in research, the question arises whether their family—or some other proxy—may consent on their behalf. The federal regulations allow a legally authorized representative (LAR) to do so while leaving the question of who may serve as a LAR to the states. A study of the state regulations shows variability in the statutes, with most states not weighing in on whether families may consent. The important questions that remain are whether proxies are/should be permitted to consent to research on behalf of an incompetent subject; if so, who may/should serve as available proxies and in what order; and whether there are/should be limits on when proxies may decide in the research context. Further study of these issues is warranted, as is gathering data on what is actually being done and how well it is working.

Competency to Make Decisions About Death and Dying

Over the last two generations, there has been a change in societal attitudes to death and dying, and this is reflected in the way in which geriatric patients may view this topic. Throughout history, life has been mostly valued and cherished, while death has been feared. The respect given to the healing professions stemmed partly from the ability of the healer to prolong life and postpone death. But recently, scholars from both medicine and medical ethics have looked more critically at what should be valued in life, its quantity or quality. There is now a school of thought that argues that it is the quality of life that is more valuable and that if a good enough quality of life is not possible, death should not be feared and postponed at all costs but rather considered and possibly welcomed.

Older adult patients may raise questions about treatment refusal, suicide, physician-assisted suicide, and euthanasia. If an older adult patient indicates that they want to refuse treatment in an attempt to promote death, the psychiatrist should determine whether the individual has decision-making capacity. If the decision to seek death is causally related to a mental disorder, then the ethical task is simple: treat the mental disorder, and reassess. On the other hand, if the elder's decision to seek death is rational and not related to a mental disorder, the patient may have the capacity to refuse treatment, even if that refusal may hasten death.

In such cases where there is a clash (real or apparent) between law and ethics, the clinical geropsychiatrist should seek consultation from a forensic colleague. The forensic specialist may then advise the clinician that the principle of respect for autonomy requires that the health care provider respect patients' choices when they are intentional and made with understanding, without controlling influences. The Doctrine of Informed Consent also balances the competing demands from law, ethics, and psychiatry. The clinical psychiatrist should ensure that all three elements of informed consent are met. Because such cases have the potential for careful scrutiny by legal agencies, ethics committees, and the media, the psychiatrist should document how the elements of informed consent were met in detail.

Physician Aid in Dying. Suicide has historically been condemned on moral and religious grounds. With the emergence of psychiatry in the mid-19th century, suicide started being conceptualized under a medical framework, and it was thought of not as a choice by an autonomous agent but rather as a symptom of a mental health disorder. However, over the past several decades, the medical model has been challenged, and some argue that suicide can be rational in some cases. This movement is reflected by the fact that, since 1997, 10 states have adopted statutes that allow individuals to request lethal prescriptions that they can choose to ingest if two physicians agree that the patient has decision-making capacity and an irreversible, terminal illness likely to lead to death within 6 months.

In states where physician aid in dying (PAD) is legal, geropsychiatrists may be asked to evaluate a terminal patient's capacity to consent to PAD. One of the key questions that the geropsychiatrist will attempt to answer is whether the patient's request for PAD is driven by a diagnosis of depression or whether it reflects a rational and autonomous choice free from the controlling influence of a mental health disorder. This can be a complicated determination, given that suicidal ideation is itself a symptom of depression. These evaluations are also complicated because no guidelines for the assessment of capacity to consent to PAD have been broadly agreed upon. Forensic psychiatrists should be aware that, without set standards for assessing such decision-making capacity, they may use themselves as the “reasonable person” by which to judge the patient’s request for PAD. In other words, their own personal beliefs about PAD may influence their capacity evaluation.

Competency to Make a Will

A forensic geropsychiatrist is most often called upon to evaluate an individual’s capacity to make a will after the elder is deceased and the will is being contested. However, some lawyers arrange psychiatric evaluations of their clients at the time of the will creation in order to avoid later disputes about testamentary capacity.

The level of capacity to create a will, referred to as “testamentary capacity,” is measured using two different standards. First, there is a three-part test in most states that determines whether the person creating the will understands (1) what they are doing in making a will; (2) the nature and extent of their property (i.e., a general recitation of what a person owns, not specific details); and, (3) the “natural objects of their bounty” (i.e., their relation to living descendants, spouse, and parents, and others whose interests are affected by the will). Most jurisdictions require evidence of a mental defect that has a direct correlation to the individual’s ability to understand one of these three elements. A series of eccentric or odd behaviors will not, in and of itself, establish a lack of capacity under this standard. However, a mental defect, such as impaired cognitive functioning that directly affects at least one of the three parts listed above (e.g., an individual’s ability to understand the natural objects of his bounty), would qualify as a lack of capacity under the three-part test. A person must understand all three parts of the capacity standard to have testamentary capacity.

Second, the person must not suffer from a mental disorder that involves delusions or hallucinations that have an effect on the particular bequest. Under this standard, a delusion or hallucination alone is enough to find a lack of testamentary capacity, but only if the delusion or hallucination has a direct correlation to the dispositions under the will. The “insane delusion” standard is hard to meet: The delusion must be extreme, and the individual must cling to the delusion in the face of uncontroverted contrary facts. For example, a belief that one’s family no longer loves one would not be an insane delusion even if it is untrue and not well grounded in evidence: It *could* be true. By contrast, a belief that one’s family is trying to kill one with ray guns would be an insane delusion. Similarly, a belief that one’s child is the product of adultery and not really one’s child will typically not be considered an insane delusion. However, a belief that the child has been taken over by aliens, and is not really one’s child, would. In addition, the delusion must affect the bequest. If the individual has another independent (sane) reason for the bequest, then the delusion does not vitiate their testamentary capacity.

Mr. Jones was an unmarried 75-year-old male with three children. One year prior to his death, he began exhibiting strange behavior. He would repeatedly visit his neighbors and offer to buy all of the furniture in their homes even though the furniture was never for sale. Mr. Jones would also make crank calls to local pizza parlors. Over the course of 12 months, he refused to let any family members or friends into his home and would send gifts through the mail that typically consisted of a whole fish or raw meat. Mr. Jones was never diagnosed with any form of dementia or mental defect causing cognitive impairment. Upon his death, Mr. Jones' will was discovered. The will was created 6 months before Mr. Jones' death, and he left his entire estate to his church. Did Mr. Jones have the legal capacity to create his will?

Mr. Jones likely had the requisite legal capacity to make his will. The law in this area presumes that decedents have capacity until proven otherwise. Even though Mr. Jones exhibited eccentric and strange behavior prior to his death, there is no evidence that he had a mental defect that directly affected his ability to understand the objects of his bounty, the fact that he was creating a will, and the general nature of his property.

Undue influence is commonly used in attempts to overturn a decedent's will. Undue influence requires (1) vulnerability of the victim, (2) apparent authority of the wrongdoer, (3) actions and tactics of the wrongdoer, and (4) equity of the result. Lessened capacity generally makes a subject more susceptible to undue influence. Although the lessened capacity may not be enough to prove a lack of capacity, it could provide support for a successful undue influence claim, provided the other elements of undue influence are present. In the example above, there are no facts to support the other elements of undue influence, but if a wrongdoer was present and exerting pressure on Mr. Jones (see box above) to leave all his assets to him, then there could be a claim for undue influence.

Roger is an 83-year-old wealthy business executive. Roger is married to 81-year-old Jenny, and they have two daughters, Wendy and Sally. For the past 20 years, Sally has worked in Roger's business, where she oversees all of his finances. Roger was advised to start gifting some of his assets to his children and grandchildren so he would not be subject to estate tax on his death. He instructed Sally to prepare checks of $10,000 each for Wendy, Wendy's two children, Sally, Sally's one child, and Sally's husband. Wendy became enraged when she learned that Sally's husband received a $10,000 check, but her husband did not. A few weeks after the checks were written, Roger died. Wendy then went to Jenny and told her that Sally was stealing money, could not be trusted with the funds, and that Sally was trying to take Jenny's assets. Jenny becomes alarmed and immediately fires Sally from the business. She cut off all communications with Sally. Wendy then moved Jenny into her home to care for her. Wendy was given a power of attorney to manage the finances, and Wendy began overseeing the family business. Jenny then created a new will, leaving all assets to Wendy, and disinheriting Sally. To note, all previous wills gave everything equally to the two daughters. Sally attempted to talk to her mother, but Wendy refused to let them speak and refused Sally entry to her home to see Jenny. During this time, Wendy frequently told Jenny that Sally was trying to be nice to her so that she could steal more money. Jenny died, and Wendy believed she was entitled to everything.

Was Jenny unduly influenced to create her will leaving everything to Wendy? The first question is whether Jenny was susceptible to undue influence. There are signs that Jenny may indeed be susceptible to undue influence due to the death of her husband, her age, and the fact that she so readily believed Wendy's wild claims. A forensic geropsychiatrist evaluating this claim would investigate Jenny's medical records to determine her level of mental functioning, remembering that Jenny need only have a weakened mental state, not a complete lack of capacity because the remaining elements of undue influence focus on the actions of Wendy. Wendy had apparent authority over Jenny both as a trusted daughter and as an agent under Jenny's durable power of attorney. Wendy used suspicious tactics in keeping Jenny secluded and refusing to allow Sally to see Jenny. The resulting will was unlike any past will, where the estate was divided equally.

This example also highlights the "poisoning of the mind" that occurs when an elder is secluded and manipulated by a single child. In this example, undue influence was present, and the will could therefore be set aside.

Competency to Make Financial Decisions or Contract

In contrast to testamentary capacity, which is often questioned retrospectively when the elder is deceased, the question regarding whether an individual has the competency to make a financial decision usually occurs when the patient is alive. For example, a forensic geropsychiatrist may be asked to provide an opinion on an elder's capacity to make financial decisions in the context of a dispute involving an elder's right to give money to a person of their choosing, particularly when there is a concern that the elder is being financially exploited.

Generally, in order to be found legally competent to make financial decisions or contracts, the subject must be able to understand and appreciate, to the extent applicable:

1. The rights, duties, and responsibilities created by or affected by the decision
2. The probable consequences of the decision
3. The significant risks, benefits, and reasonable alternatives involved in the decision

A financial decision or contract can also be overturned based on the legal concept of undue influence. In this context, undue influence focuses more on the actions of the wrongdoer and looks to mental functioning only to determine if the subject is vulnerable or susceptible to undue influence. In other words, a person can be susceptible to undue influence but still have the requisite legal mental capacity to make a financial decision or contract. Courts tend to be wary of overturning an elder’s capacity to make decisions regarding whether to enter into contracts or make financial decisions, whereas they are more liberal with protecting an elder from possible undue influence perpetrated by another.

Diane is an 85-year-old female who has always been independent and refuses to talk about her finances with her only daughter, Tina Smith. Tina lives in San Francisco, while Diane lives in Los Angeles. Diane owns the home she lives in as well as three other homes on the same block, which she rents out to various tenants. One of Diane's neighbors, Ned Riley, is a handyman who has helped her repair her properties and various other odd jobs throughout the past 10 years. Diane's mental functioning has slowly declined, and she has started having trouble walking. Ned and his wife, Brenda, started assisting Diane with grocery shopping, taking her to medical appointments, and obtaining her prescription medicine. After several months, Brenda takes over Diane's checkbook and starts paying her bills. Brenda writes all the checks by hand and then has Diane sign the checks. After 10 months of assisting Diane, Ned and Brenda create a corporation and enter into a contract between Diane and the new corporation to "buy" Diane's three rental properties by paying a $1,000 down payment and agreeing to make payments of $1,000 per month. They set the total purchase price for all three properties at $300,000. Each rental property is worth over $250,000, making this sale arrangement a below-market bargain for Ned and Brenda. After the initial down payment, no other payments are made, and all rental income goes directly to Ned and Brenda. After a year of this arrangement, Diane runs low on funds and cannot pay for her prescriptions, which is the first time Diane mentions her finances to her daughter. Tina hires a lawyer and discovers the "sale" arrangement as well as the neighbors' actions with Diane's checkbook.

A forensic geropsychiatrist evaluating Diane may be asked to consider both her contractual capacity as well as whether she was unduly influenced to make the financial decisions described above. Her contractual capacity would be evaluated using the elements pertaining to her capacity to make a decision. Although the above facts do not provide enough details to fully evaluate capacity, the mere fact that the properties were sold at a substantial discount suggests that Diane may have had difficulty understanding the consequences of that decision.

In addition, the facts of this case are typical of the pattern of behavior found in many undue influence claims: the neighbors started by being helpful and slowly took over more and more jobs for Diane. Such actions build trust and goodwill with Diane. Over time, the neighbors used that trusting relationship to convince Diane to enter into a bargain sale of the properties. The first step in evaluating an undue influence claim is to determine Diane's susceptibility to undue influence. The forensic geropsychiatrist would then consider the actions and tactics of the neighbors and their position of trust with Diane, and the inequity of the result. Capacity and undue influence are separate claims that can be evaluated at the same time; one claim can succeed where the other fails, or they both can be established where the facts are sufficient.

Competency to Drive

Loss of driving privileges is, after nursing home placement, the most feared event for an older adult person. Age alone does not cause impairment behind the wheel; however, impairments associated with neurocognitive disorders can. Geriatric psychiatry training rarely provides specific instructions on how to complete a driving assessment. Therefore, many states require that a physician make a referral to the Department of Motor Vehicles (DMV) if a major neurocognitive disorder is diagnosed but do not require the physician to state whether the patient can or cannot drive.

After a case is referred to the DMV, screening test administrators evaluate whether the elder has the four mental requirements to drive:

1. Proper integration of sensory input
2. Focus of attention
3. Proper association of thought

4. Appropriate judgment

Advanced Directives

Advance directives are instructions regarding the patient’s preferences for medical care if they become incapacitated that are designed to promote an individual’s autonomous decision making. One type of advance directive is a living will, which addresses a patient’s wishes for life-sustaining procedures and end-of-life treatment. The second type of advance directive is a durable power of attorney for health care, where a patient designates another person the authority to make decisions for themselves in the event of incapacity. Although hospitals are required not only to ask patients whether they have medical advance directives at the time of admission but also to provide their patients with information about advance directives, most patients nevertheless still do not have such documents.

More recently, psychiatric advance directives have been promoted as a means by which patients with mental health disorders can describe their preferences for psychiatric treatment in the event that they become incapacitated. Psychiatric advance directives can include information such as medication preferences, helpful de-escalation, coping techniques, and a list of individuals that the patient would like involved in their care. Currently, only 25 states have specific psychiatric advance directive statutes. However, for patients in states without specific psychiatric advance directive laws, general medical advance directives can be used to direct some types of psychiatric treatment.

ADMINISTRATIVE ASPECTS OF FORENSIC GEROPSYCHIATRY

The Older Adult Impaired Physician

Life expectancy is rising in most Western societies, and upper-income, highly educated persons like physicians are likely to live and work even longer than the national average. Impediments to working in old age may include physical infirmities (e.g., a Parkinsonian tremor may cut a surgical career short) or mental infirmities (e.g., mild cognitive impairment). If the affected individual is a physician, this is most likely to come to the attention of a medical board through a complaint made by an unhappy patient, usually after a bad clinical outcome.

The geriatric psychiatrist may be asked to evaluate the older adult physician in order to help the medical board reach a conclusion regarding the physician's license. The evaluating geropsychiatrist should avoid giving a conclusion (i.e., "the doctor should keep or not keep their medical license") because this ultimate opinion is the decision of the medical board. Instead, the forensic geropsychiatrist should answer only the specific question(s) asked. For example, the forensic geropsychiatrist may be asked to diagnose the physician in question. They may also be asked to evaluate the various strengths and weaknesses in the elder's capacity to plan, organize, sequence and abstract (i.e., the executive functions), and identify any related impairments.

Older Adult Applicants for Disability Benefits

Public or private provisions protect working individuals against losing all income in the case of a physical or mental disability. In the public sector, most citizens are covered by the Social Security Administration's (SSA's) policies. In the private sector, there are many insurance companies that offer disability benefits that will provide the individual with greater financial benefits than what would otherwise be offered through the SSA. Unlike in the private sector, SSA provisions are not designed to generate a profit but instead are built to ensure that no disabled person is unable to provide for themselves if they lack the capacity to work. Forensic geropsychiatrists may be asked to evaluate individuals applying for disability in order to determine whether the applicant meets the SSA's or the private insurance company's definition of a disability.

If consulted on an SSA disability case, the geriatric psychiatrist should resist the temptation to give an opinion about the ultimate issue (i.e., the claimant can or cannot work), which is an administrative instead of medical judgment based on social security law. Rather, the doctor should give an opinion regarding the penultimate medical issues. These evaluations can be complicated because there is a normal decline in performance with age in concentration, persistence, and pace. The SSA will then ask the psychiatrist to rate any impairments in severity.

Evaluating for Disability. The geriatric psychiatrist should determine whether the applicant meets the criteria for a qualifying mental disorder, as well as the quality and quantity of impairments in four areas:

1. Activities of daily living
2. Social functioning
3. Concentration, persistence, and pace
4. Decompensations

The SSA guidelines are sensitive to the demands of age. In other words, the same level of impairment may result in a failed claim if the applicant is young but succeeds if the applicant is an older adult. If an impairment is present, the doctor should relate it to a causal factor, which could be a mental disorder (e.g., major neurocognitive disorder) or another complicating factor (e.g., lack of English language skills). Based on the evaluating psychiatrist’s opinion, the SSA will then determine whether the applicant does or does not qualify for disability benefits.

CORRECTIONAL ASPECTS OF FORENSIC GEROPSYCHIATRY

Opportunities for the geriatric psychiatrist to provide clinical services within the correctional system abound and are likely to increase exponentially over the ensuing years as the population ages. It is important to recognize that the correctional psychiatrist *treats* the incarcerated elder and hence plays a different role than the forensic psychiatrist, who *evaluates* the incarcerated elder to truthfully answer a medicolegal question. Over the past 10 years, awareness of the impending problem of the rising number of older adult prisoners has increased. Virtually every major newspaper in the United States has run a story on how the “graying” of the incarcerated population is likely to present a serious ethical or moral challenge as well as an economic crisis for the penal system. There are currently more than 125,000 prisoners over the age of 50 and 35,000 prisoners over the age of 65 in U.S. correctional facilities. Due to sentencing laws and demographic shifts in the population, the U.S. Census Bureau estimates that in 2030, one-third of the jail population will be over 55 years of age.

The geriatric correctional psychiatrist may consider whether an aging prisoner should be released. There are a number of considerations: 97% of older adult inmates are nonviolent offenders, and less than 1% of those on probation or parole who are returned to jail or prison are older than 60. Furthermore, the poor health conditions of prisoners combined with harsh prison life make the prison environment particularly conducive to fostering disease. There is, however, no legal statute that is likely to provide elders with any specific rights beyond the general right to medical care and the right to be free from cruel and unusual punishment. A related question is whether age should be a basis for execution exemption. There are presently over 100 people aged 60 or over on death row. The Supreme Court has ruled that it is unconstitutional to execute certain special populations (e.g., juveniles and those with intellectual disabilities), given that executing these vulnerable populations would violate the Eighth Amendment’s prohibition against cruel and unusual punishment. However, there is no specific prohibition against executing older adults. Therefore, unless the older adult defendant does not understand why they are being executed (i.e., unless they are found incompetent to be executed), they are unlikely to be exempt from the ultimate punishment.

NOSOLOGIC/THEORETICAL ISSUES

In routine clinical practice, most geropsychiatrists use the Diagnostic and Statistical Manual (DSM) or equivalent International Statistical Classification of Diseases and Related Problems (ICD) and follow the “spirit” of the text.

Testifying in Court

In court testimony, the doctor swears to tell the whole truth, and if the diagnoses do not follow the letter of the text, the doctor must truthfully acknowledge that. But the truth may be difficult to determine because forensic geropsychiatrists are often asked to clarify boundaries that may be fuzzy, including the boundaries between the following:

1. Cognitive disorder/age-related cognitive decline
2. Antisocial personality disorder/adult antisocial behavior
3. Depressive disorder/bereavement

In the accompanying box, the first example is a mental disorder, and the second is not. The boundary between the two is often vague and ill-defined, and there may be temptations to place the boundary in a manner that promotes values that promote the well-being of the evaluee; for example, for policy-based (access to care) rather than scientific (nosology) reasons. Although some may consider it admirable for the clinical geropsychiatrist to be compassionate or to advocate for greater access to health care for untreated patients, the forensic geropsychiatrist subscribes to a code of ethics that emphasizes truth-telling rather than the traditional ethical values of promoting health.

Future research in forensic geriatric psychiatry may therefore be of two types: quantitative (i.e., data-based) or qualitative (i.e., conceptual). Quantitative research may focus on the sciences pertaining to nosology, such as clarifying the proper boundaries of the various mental disorders that may be a basis for forensic opinions. Future quantitative research may also help to better delineate the pathways—cognitive, social, moral, logical—of good and bad judgment that often form the basis of findings of mental incompetence in older adults. Future qualitative research may help integrate findings from the biomedical aspects of geropsychiatry with the implications of such findings. For example, whether or not a neurocognitive disorder is curable may alter the legal protections for the rights of the older adult mentally ill and may raise ethical concerns regarding the allocation of limited health care resources to a population of older adults that may have unlimited growth.

FURTHER READINGS

Anfang SA, Appelbaum PS. Forensic evaluation of the older patient. In: Agronin ME, Maletta GJ, eds. *Principles and Practice of Geriatric Psychiatry*. Lippincott Williams & Wilkins; 2006.

Anfang S, Bonnie R, Brendel R, et al. APA Resource Document on Physician Assisted Death. 2017. Accessed on March 14, 2022. www.psychiatry.org/psychiatrists/search-directories-databases/library-and-archive/resource-documents

Appel JM. Locating and identifying third-party decision makers. *J Am Acad Psychiatry Law*. 2022;50(1):84–96.

Brendel RW, Schouten R, Nesbit A, Levenson JL. Legal and ethical issues. In: Levenson JL, ed. *The American Psychiatric Publishing Textbook of Psychosomatic Medicine: Psychiatric Care of the Medically Ill*. 3rd ed. American Psychiatric Publishing; 2019.

Burkle CM, Mueller PS, Swetz KM, Hook CC, Keegan MT, et al. Physician perspectives and compliance with patient advance directives: the role external factors play on physician decision making. *BMC Med Ethics*. 2012;13:31.

California Probate Code sections 811 (evidence of mental defects required to prove a lack of capacity) (1996); 812 (capacity to make decisions) (1996); and 813 (capacity to give informed consent to medical treatment) 1995.

California Welf. & Inst. Code section 15610.70 (effective January 1, 2014).

Clark RV. *28 Cal.4th*. 2002;89:96.

Cohen AB, Wright MS, Cooney L. Guardianship and end-of-life decision making. *JAMA Intern Med*. 2015;175(10):1687–1691.

Darby RR, Dickerson BC. Dementia, decision-making, and capacity. *Harv Rev Psychiatry*. 2017;25(6):270–278.

DeMartino ES, Dudzinski DM, Doyl CK, et al. Who decides when a patient can’t? Statutes on alternate decision makers. *N Engl J Med*. 2017;376(15):1478–1482.

Drogin EY, Barrett CL. Evaluation for guardianship. In: Roesch R, Zapf PA, eds. *Forensic Assessments in Criminal and Civil Law: A Handbook for Lawyers*. 135–

147. Oxford University Press.

Edersheim J, Murray ED, Padmanabhan JL, Price BH. Protecting the health and finances of the elderly with early cognitive impairment. *J Am Acad Psychiatry Law*. 2017;45(1):81–91.

Garner BA, ed. Blacks Law Dictionary. 11th ed. West Publishing Company; 2019.

Gold LH, Anfang SA, Drukteinis AM, et al. AAPL practice guideline for the forensic evaluation of psychiatric disability. *J Am Acad Psychiatry Law*. 2008;36(4):s3–s50.

Grisso T. *Evaluating Competencies: Forensic Assessments and Instruments*. 2nd ed. Plenum Publishing; 2003.

Horwitz HL. California's new statutory definition of undue influence—modernization or game changer? *California Trusts Estates Quarterly*. 2013;19(3):1–8.

*Jeste DV, Palmer BW, Appelbaum PS, Golshan S, Glorioso D. A new brief instrument for assessing decisional capacity for clinical research. *Arch Gen Psychiatry*. 2007;64:966–974.

Mart EG. Testamentary capacity and undue influence: Assessments and problematic techniques. *Psychological Injury and Law*. 2020;13:1–10.

McCarthy KE, Rose C. State initiatives to address aging prisoners. U.S. Department of Justice. Accessed March 14, 2022. www.ojp.gov/ncjrs/virtual-library/abstracts/state-initiatives-address-aging-prisoners

Miller MK, Bornstein BH. Elders on death row. *APA Online Monitor Psychol*. 2007;38(4):60.

*Naimark D. Financial exploitation of the elderly: the evaluation of mental capacity and undue influence. *Am J Forens Psychiatry*. 2001;22(3):5–19.

Naimark D, Dunn L, Haroun A, Morris G. Informed consent and competency-legal and ethical issues. In: Jeste DV, Friedman JH, eds. *Current Clinical Neurology: Psychiatry for Neurologists*. Human Press Inc.; 2006.

National Resource Center on Psychiatric Advance Directives (2018). Accessed April 24, 2021. https://www.nrc-pad.org/getting-started/

Nishimura A, Carey J, Erwin PJ, Tilburt JC, Murad MH, McCormick JB. Improving understanding in the research informed consent process: a systematic review of 54 interventions tested in randomized controlled trials. *BMC Med Ethics*. 2013;14:28.

Plotkin DA, Spar JE, Horowitz HL. Assessing undue influence. *J Am Acad Psychiatry Law*. 2016;44(3):344.

Rinkert J, Naimark D. The elderly driver. *Psychiatr Times*. 2005;12:7.

Saks ER, Dunn LB, Wimer J, Gonzales M, Kim S. Proxy consent to research: the legal landscape. *Yale J Health Policy Law Ethics*. 2008;8:37–92.

Sunstein CR. The ethics of nudging. *Yale J Regul*. 2015;32:413–450.

Superior Court of California. County of San Francisco, Undue Influence: Definitions And Applications, A Project Supported by the Borchard Foundation Center on Law and Aging, Final Report (March 2010). Accessed March 14, 2022. http://www.courts.ca.gov/documents/UndueInfluence.pdf

The Trusts & Estates Section of the State Bar of California. Preserving Senior Independence, An Outline for Speaker Presentations (June 6, 2013). p. 2.

Thoreen VL, David KG. What every trusts and estates practitioner needs to know about elder financial abuse. California Trusts Estates Quarterly. 2014;20:1.U.S. Gov't Accountability Office, Elder Justice. Stronger Federal Leadership Could Enhance National Response to Elder Abuse, GAO-11-208 (March 2011) p. 48 (Citing MetLife Mature Market Institute et al., Broken Trust: Elders, Family, and Finances: A Study of Elder Financial Abuse Prevention (2009) (notes 1 & 4).

U.S. Gov't Accountability Office, Elder Justice. National Strategy Needed to Effectively Combat Elder Financial Exploitation, GAO-13–110 (November 2012).

55.6c Ethical Issues

Nathan Fairman, M.D., M.P.H.

Elsewhere in this volume, one can find both a broad overview of psychiatric bioethics and a more particular treatment of key ethcial issues at the intersection of palliative care and mental health. Given the excellent coverage of these and other aspects of psychiatric bioethics in this textbook, it's worth asking: *why include a separate chapter on ethical issues in geriatric mental health?* After all, the ethical challenges in old-age psychiatry are, at their core, no different from the ethical challenges one can encounter in the care of nonolder adult mental health patients or, for that matter, in the care of patients without mental illness. However, the myriad changes that accompany aging—developmental, physical, social, cognitive, and others—intersect with mental illness in ways that repeatedly bring ethical concerns into the foreground of patient care. As a result, taking care of older adults with mental illness involves navigating a set of common ethical issues which, while similar in kind to those seen in other populations, differ in their complexity and consequences.

The chapter begins with a reflection on the Principlism paradigm and the way in which it focuses our attention and organizes how we've come to understand ethical obligations in geriatric mental health. Next, the chapter considers two related ethical constructs that are particularly important in geriatric mental health: informed consent and decision-making capacity. The rationale for this focus is that many of the common ethical challenges in geriatric mental health involve questions about respect for autonomy and personhood, particularly in cases marked by cognitive deficits or affective states that can influence decision-making. Careful attention to decisional capacity and informed consent can help ensure ethically responsible care in a population that is sometimes vulnerable to impaired decision-making. A further rationale for this area of focus is that, in many cases, older adults with chronic, life-limiting illnesses eventually face decisions that can influence the timing of death—a circumstance that may seem to "raise the stakes" surrounding questions of informed consent and decisional capacity. Finally, the chapter closes with a patient case that illustrates some of these challenges and presents an ethically responsible approach to caring for a depressed older adult with serious illness and a desire to die.

THE FOCUSING EFFECT OF A FRAMEWORK

Principlism has, since its emergence in the late 1970s, been the predominant ethical paradigm in western medicine. In their *Principles of Biomedical Ethics,* Tom Beauchamp and James Childress identified four core principles of bioethics: respect for autonomy, nonmaleficence, beneficence, and justice. Each principle gives rise to particular duties or responsibilities that health care providers must meet when they care for patients. A true ethical dilemma occurs when the duties that arise from two or more principles are in direct conflict with one another.

Principlism has been criticized from several directions. In regard to geriatric psychiatric ethics, one particular criticism warrants mention: Principlism assigns a high level of moral value to personhood and individualism, and as a result, the framework has a difficult time accommodating the values of nonwestern, communitarian cultures and worldviews. Hence, some of the circumstances that create an apparent ethical dilemma from the view of Principlism may become less salient ethically when viewed through the prism of a different ethical framework. Our social embeddedness as human beings and the very particular, idiosyncratic details of any individual circumstance aren’t given much real estate in the Principlist landscape of ethical value. Nonetheless, Principlism continues to be the predominant bioethical framework in contemporary western medicine, and it does provide a useful way to conceptualize common tensions that occur in geriatric mental health care. The important thing to notice is that the Principlism framework itself focuses our attention, repeatedly, on particular circumstances of geriatric mental health that often produce tension between the obligations of two core principles: respect for autonomy and beneficence.

This brings us to two important reasons that ethical challenges (according to Principlism) frequently occur in the setting of geriatric mental health: first, conditions that are common in old age and in mental illness often undermine an individual’s ability to exercise autonomy. In an ethical sense, such individuals can become vulnerable because various diseases (mental, physical, etc.) erode their capacity to make autonomous, informed choices. Second, there is a unique way in which cognitive decline and mental illness can destabilize the notion of “personhood,” which is central to Principlism. Whose autonomy is to be respected, for example, when severe Major Depression or an advanced Neurocognitive Disorder can so thoroughly and fundamentally alter the present “personhood” of a previously healthy individual? Other medical conditions do not have the power to confuse us about the target of our ethical obligations in the same way that mental illness and severe cognitive changes can.

GENERAL APPLICATIONS OF THE CORE PRINCIPLES TO GERIATRIC MENTAL HEALTH CARE

Several general applications of the principles of bioethics to geriatric mental health care warrant mention. For example, dignity is often a paramount consideration in geriatric mental health. As a result of the stigma associated with neuropsychiatric disorders, as well as the loss of function and status that can be associated with neurodegenerative conditions, patients and loved ones often view health-related decisions through the prism of concerns about dignity. A clinician's efforts to preserve her patient's dignity reflect an ethical commitment to respect her patient's autonomy and personhood. Similarly, the right to confidentiality can also be viewed as an extension of the core principle of respect for autonomy. In the context of work with decisional-impaired adults, confidentiality issues can be particularly complex, as there is often a need to work closely with family members and caregivers, sometimes requiring the disclosure of information at a level that would be uncommon (and perhaps inappropriate) with a patient whose decisional capacity was intact.

Justice considerations, for example, can be relevant to the development and testing of psychotropic medications in older adult populations. Historically, older age patients have been underrepresented in clinical trials of antidepressants and other psychotropic medications. Given potential age-related changes in pharmacokinetics and pharmacodynamics, as well as the increased likelihood of drug interactions from polypharmacy, the need for clinicians to extrapolate from data collected from samples of younger patients raises ethical concerns in terms of justice (in addition to nonmaleficence). The example also illustrates that ethical dilemmas are not always best conceptualized or addressed at the level of individual clinician and patient interactions but may also occur and require change at the level of public policy—in this case, with respect to setting national (or international) research requirements and priorities.

OPERATIONALIZING RESPECT FOR AUTONOMY: DECISION-MAKING CAPACITY AND THE DOCTRINE OF INFORMED CONSENT

As noted above, ethical dilemmas in geriatric mental health often arise from apparent opposition between two core principles: respect for autonomy and beneficence (and its corollary, nonmaleficence). In other words, the conditions of mental illness and advancing age frequently combine to produce circumstances that make it seem impossible to know how to meet those ethical obligations or impossible to satisfy both of them at once. What's more, such conflicts can be particularly difficult to resolve when the weighing of considerations from competing principles is highly subjective, that is, in those instances wherein reasonable people might disagree and come to divergent conclusions regarding the most appropriate course of action. Consider the controversies around the "right to die" near the end of life, often exemplified in discussions about withdrawing or withholding treatment, palliative sedation, or medical aid-in-dying. One can readily identify tensions among the core bioethical principles in such issues; wherein reasonable people might disagree about how to best respect autonomy, balance obligations to beneficence and nonmaleficence, and even address concerns about justice. Some of these tensions are brought to light in the following case:

Sylvia is a 68-year-old woman with advanced chronic obstructive pulmonary disease, a history of breast cancer, and major depression. Though she lives alone, over the past year, she has found it increasingly difficult to maintain her independence: she's been in and out of the hospital several times, and after the most recent hospitalization, she was discharged to a nursing facility for rehabilitation. She found the brief stay in the nursing facility to be demoralizing—not so much because of the quality of care she received, but rather because the idea of being dependent on others for her most basic needs ran contrary to her own sense of self-worth. She counts her ex-husband, her daughter, and a former colleague among her closest supports, and she has vowed to them that she would "rather die" than experience "the indignity of wasting away" in a nursing home.

After another hospitalization—the result of yet another bout of pneumonia—Sylvia appears more depressed than usual, and there are some new concerns about delirium. She makes an oblique reference to her psychiatrist, suggesting that she wishes her life were over: "I'm tired of going through this; when will it end? There must be something you can do to help me. . ."

From the standpoint of bioethics, how should Sylvia's doctor respond? What course of action will strike the right balance among the different—and seemingly opposed—obligations? For instance, how should her doctor negotiate the two components of respect for persons (honoring Sylvia's right to autonomous decisions while protecting her in the event that her capacity for authentic autonomous decision-making is diminished due to what may be depression or cognitive impairment)? Further, apart from concerns about autonomy, how should Sylvia's doctor understand "the good" that should be promoted or "the harm" that should be avoided (in order to respect the dictates of beneficence/nonmaleficence) in a case where a suffering patient, near the end of life, desires that her death come more quickly?

Not surprisingly, in medical ward settings, psychiatrists are frequently asked to provide consultation in cases like Sylvia's—to evaluate the degree to which a patient's refusal of potentially life-extending treatment (or, more generally, their desire for hastened death) may reflect a compromised capacity for autonomous decision making due to depression, cognitive impairment, or even psychosis. From the perspective of the geriatric mental health clinician, the question of decisional capacity—a component of informed consent relating fundamentally to the principle of respect for persons—is likely to be instrumental in helping to sort out some of these issues.

Three Components of Informed Consent

The doctrine of informed consent operationalizes the duty to respect autonomy: it ensures that the patient is capable of coupling the physician's expert knowledge with their own personal values and preferences—to arrive at an informed medical decision about their own body. In most formulations, informed consent includes three interrelated components:

1. Voluntariness—Valid informed consent requires that the individual's decision is made in the absence of coercion (i.e., the decision must be truly voluntary).
2. Disclosure—Informed consent also requires disclosure of all the information relevant to the decision at hand, such as the purpose and nature of the proposed treatment or procedure, viable alternatives, and the foreseeable risks and benefits, from each possible course of action (including doing nothing). Such disclosure is generally documented with a printed consent form that the patient or his or her surrogate signs; however, as cogently noted in a report from the Institute of Medicine:

> Conceptually, informed consent should be construed as an evolving decision-making process, rather than as simple "permission." This model of "ongoing conversation and decision making" allows for reinforcement of previously disclosed information, the introduction of new information, and respect for autonomous participant decision making. The act of obtaining *written* consent is tangential to the *process* of informed consent and merely provides a mechanism to document and record that ethically relevant communication has taken place. (Institute of Medicine. Responsible Research: A Systems Approach to Protecting Research Participants. National Academy of Sciences. 2003:108–135. doi: 10.17226/1050.)

3. Decisional Capacity—The individual must have an adequate level of cognitive and reasoning faculty to make an authentic, informed decision.

Four Components of Decision-Making Capacity

Decision-making capacity derives from a set of mental abilities, which together enable a patient to engage in the process of informed consent. Having decisional capacity is a necessary—but not sufficient—element of informed consent. In most formulations, decision-making capacity comprises four conditions:

1. Understanding—the ability to comprehend information relevant to the decision at hand;
2. Appreciation—the ability to apply that information to one’s own situation and condition;
3. Reasoning—the ability to manipulate and consider pertinent information so as to rationally assess the different options available; and
4. Communication—the ability to express a stable choice over time.

Many experts draw a distinction between “capacity” (a determination made by a physician) and “competency” (a determination made by a judge), but the two terms are often, confusingly, used synonymously. In general, questions of capacity should relate to specific interventions at a specific point in time (e.g., a trial of dialysis, or surgical repair of an aneurysm, etc.), rather than to a more generalized ability since the capacity required to engage in a simple, straightforward decision may be substantially different from the capacity required to engage in a highly complex or risky decision (or the capacity to decline a highly beneficial, low-risk intervention).

In most states, decisional capacity can be determined by any licensed physician and sometimes also by a small set of nonphysician mental health specialists, such as clinical psychologists. Hence, most consultation requests for “capacity evaluation” can be framed as providing a second opinion when the primary clinician’s assessment is indeterminate or uncertain. In practice, psychiatrists are frequently consulted when serious affective conditions, problems with cognition, or deficits in reality testing are felt to interfere with decision-making capacity.

Capacity determinations are ultimately based on clinical assessment, which is often complemented by an objective measurement of several domains of cognition. Unfortunately, empirical studies suggest poor reliability, even among experienced clinicians, in terms of categorical determinations of an individual patient’s capacity to consent to treatment based on standard clinical interviews. Interrater reliability may improve when the clinicians providing such ratings are given explicit operational guidelines. Several tools have been developed to assist with the assessment of decisional capacity, such as the MacArthur Competence Assessment Tool for Treatment (MacCAT-T), the Capacity to Consent to Treatment Instrument (CCTI), and the Aid to Capacity Evaluation (ACE). The advantages and disadvantages of any instrument should be considered in reference to the specific purpose of the capacity assessment for an individual patient or research subject. Finally, while published reviews and educational resources often leave the impression that capacity determinations are merely a matter of applying certain criteria in the context of a clinical examination, such assessments in complicated cases are rarely straightforward, and they often involve an uncomfortable degree of subjectivity and judgment.

Respect for Autonomy in Patients With Impaired Decision Making

Where does this leave the geriatric mental health specialist when faced with a person like Sylvia, who, in the advanced stage of a progressive and terminal illness, experiencing some form of depression as well as possible cognitive deficits, begins to give voice to a desire to die? If Sylvia has decisional capacity, she might, for example, choose to decline antibiotics for the pneumonia and be discharged home with hospice care, with an expectation that she would soon die. If she lacks capacity, however, should antibiotics be given without her consent? Should she be “forced,” against her objections, to return to a nursing facility? If she is depressed, should her tiredness with living trigger the standard mental health safety protocols: an evaluation for involuntary detainment, for example? From the standpoint of bioethics, uncertainty about these questions reflects tensions arising from the dual obligation to respect Sylvia’s autonomy while optimizing beneficence (and minimizing nonmaleficence). Hence, in very practical ways, what is done for (and to) Sylvia will hinge on the assessment of decisional capacity.

Often, a careful and systematic process can help to achieve an acceptable solution, even in some of the most complicated cases. A “process-oriented” solution to this ethically challenging situation might involve thoughtful clarification of Sylvia’s decision-making capacity, as well as a comprehensive assessment of other sources of suffering, close communication and collaboration with family, realistic assessment of prognosis and the effectiveness of treatment, and therapeutic interventions provided in time-limited trials with clear outcomes. In Sylvia’s case, such a process might look like this:

1. Empathic clarification of her request: what, precisely, is she referring to when she says, “I’m tired of going through this…”? What does she have in mind when she says, “there must be something you can do to help me…”? As Susan Block and Andrew Billings have observed: “Most patients, in saying that they want to die, are asking for assistance in living—for help in dealing with depression, anxiety about the future, grief, lack of control, dependence, physical suffering, and spiritual despair.”
2. Attention to potential sources of coercion (e.g., from financial pressures or burdens of caregiving).

3. Careful assessment of decisional capacity. When severe depression influences capacity, it often does so through distorted judgments, leading patients to undervalue the likelihood of a meaningful future, underestimate the possibility that depression could be relieved, or even be unaware of the presence of depression itself.
4. A thorough exploration of Sylvia’s long-held values; how she has made decisions in the past; confirmation from a surrogate that her current choices are “authentic” in the sense that they are in alignment with her lifelong preferences and values.
5. Identification and treatment (if consistent with the goals of care) of any potentially reversible causes of delirium.
6. Thorough, expert characterization and appropriate treatment of “depression.” Is this another major depressive episode? Grief? Demoralization? An adjustment disorder?
7. A comprehensive review of other forms of distress: physical, emotional, interpersonal, spiritual, practical, etc. Reliance on an interdisciplinary team to respond to the patient’s suffering and deliver the plan of care.
8. Interventions that are structured in time-limited trials with clear outcomes.
9. Continuous, frank communication with the patient, her surrogates, and other relevant stakeholders.

Although treatment “against one’s will” is sometimes supported by the obligation to protect vulnerable individuals with impaired decisional capacity, clinicians should also be open to the possibility that, in many cases, one can simultaneously respect the person’s autonomy and dignity while also protecting him or her as a person vulnerable to impaired decisional capacity. In particular, efforts to explain and obtain the person’s assent, wherever possible, will foster dignity and respect, even when consent must be obtained from the patient’s surrogate.

Ethical challenges are commonly seen in geriatric psychiatry because advancing age and mental illness can combine in unique ways that undermine autonomy and threaten personhood. When such challenges arise in older adult mental health care, they are often the result of tension between two core ethical principles: respect for autonomy and beneficence. The doctrine of informed consent and the concept of decision-making capacity are two constructs that operationalize the duty to respect autonomy. Some of the most difficult ethical challenges in geriatric mental health can be addressed by adopting a process that optimizes and preserves a patient’s ability to express their autonomy while engaging an interdisciplinary team to assess and address the multiple dimensions of distress that can occur in this population.

FURTHER READINGS

American Psychiatric Association. The principles of medical ethics with annotations especially applicable to psychiatry. 2009. https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/Ethics/principles-medical-ethics.pdf

*Beauchamp T, Childress J. *Principles of Biomedical Ethics*. Oxford University Press; 1979.

Bloch S, Green SA. *Psychiatric Ethics*. 4th ed. Oxford University Press; 2009.

Block SD. Psychological issues in end-of-life care. *J Palliat Med*. 2006;9(3):751–772.

Block SD, Billings JA. Patient requests for euthanasia and assisted suicide in terminal illness. The role of the psychiatrist. *Psychosomatics*. 1995;36:445–457.

Community tools: Aid to Capacity Evaluation (ACE). University of Toronto Joint Centre for Bioethics. Available at https://jcb.utoronto.ca/wp-content/uploads/2021/03/ace.pdf

*DuBois JM. *Ethics in Mental Health Research: Principles, Guidance, and Cases*. Oxford University Press; 2008.

Giron MST, Fastbom J, Winblad B. Clinical trials of potential antidepressants: to what extent are the elderly represented: a review. *Int J Geriatr Psychiatry*. 2005;20:201–217.

*Grisso T, Appelbaum PS. *Assessing Competence to Consent to Treatment: A Guide for Physicians and Other Health Professionals*. Oxford University Press; 1998.

Grisso T, Appelbaum PS. *MacArthur Competence Assessment Tool for Treatment (MacCAT-T)*. Professional Resource Press; 1998.

Federman D, Hannam K, Lyman Rodriguez L. *Responsible Research: A Systems Approach to Protecting Research Participants*. Institute of Medicine, National Academies Press; 2003.

Kim SYH, Karlawish JH, Kim HM, Wall IF, Bozoki AC, Appelbaum PS. Preservation of the capacity to appoint a proxy decision maker: implications for dementia research. *Arch Gen Psychiatry*. 2011;68(2):214–220.

Marson DC, Earnst KS, Jamil F, Bartolucci A, Harrell LE. Consistency of physicians' legal standard and personal judgments of competency in patients with Alzheimer's disease. *J Am Geriatr Soc*. 2000;48:911–918.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. 1979. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html

Oppenheimer C. "Ethics in old age psychiatry." *Chapter 21 in: Bloch S, Green SA. Psychiatric Ethics.* 4th ed. Oxford University Press; 2009.

Rabins PV, Black BS. Ethical issues in geriatric psychiatry. *Int Rev Psychiatry*. 2010;22(3):267–273.

Sessums LL, Zembrzuska H, Jackson JL. Does this patient have medical decision-making capacity? *JAMA*. 2011;306(4):420–427.

Sullivan MD, Youngner SJ. Depression, competence, and the right to refuse lifesaving medical treatment. *Am J Psychiatry*. 1994;151:971–978.

Walaszek A. Clinical ethics issues in geriatric psychiatry. *Psychiatr Clin North Am*. 2009;32(2):343–359.

55.6d Gender Issues

HELEN H. KYOMEN, M.D., M.S.

Gender and age are important factors in the daily practice and science of medicine and are key components in establishing diagnoses, evaluating risks, and determining treatment, including the safe and effective dosages of medications. Due to differences in life expectancy, there are more older women than men. In 2018, there were approximately 125 older women for every 100 older men in the United States; in those 85 years and older, there were about 181 women for every 100 men. This has significant clinical and scientific consequences. For example, in current psychiatric practice, psychopharmacologic interventions are widely used. In general, prescription drugs are used more frequently by older adult women than men. According to global postmarketing drug surveillance data, overall, women report more adverse drug reactions than men. Yet women, specifically those of older age, were underrepresented in medical investigations. Studies designed to establish the safety, tolerability, and efficacy of medications were historically biased for incorporating younger men as study subjects. As a result, established therapeutic guidelines may have been inadequate or inappropriate for older women.

Great efforts have been made toward filling the knowledge gap that resulted from the historical underrepresentation of women in research studies. In 1985, the U.S. Public Health Service Task Force issued a report on women’s health issues that documented that women were disadvantaged in health care and research. Based on this report, the National Institutes of Health (NIH) released a policy encouraging the inclusion of women in clinical research. In 1990, following an investigative study by the Government Accounting Office (GAO) and urging from the Congressional Caucus on Women’s Issues, the NIH established the Office of Research on Women’s Health to take responsibility for establishing policy and promoting research on women’s health. In March 1993, the U.S. Food and Drug Administration (FDA) stated that it would “end its ban on women’s participation in most drug safety tests and require companies to carry out analysis by gender in virtually all applications for new drugs.” This declaration followed the GAO October 1992 report “FDA Needs to Ensure More Study of Gender Differences in Prescription Drug Testing.” Further guidance was issued in 1993, followed by regulations in 1998 and 2000. Federally funded clinical intervention studies would be required to include women, with data analysis by gender. These legal pronouncements, with continued oversight, help to ensure that women are represented in health care research, which informs current evidence-based medical practice and patient-centered care.

Clearly, patients are better served when gender and age are considered. This is especially important for older women as after menopause, morbidity and mortality from conditions such as cardiovascular disease, major neurocognitive disorder, or depression reach or exceed rates seen in men. These differences will be highlighted and illustrated in the sections that follow.

IMPACT OF GENDER ON PSYCHIATRIC CONDITIONS

Major Neurocognitive Disorders

Major neurocognitive disorders are important health concerns in aging populations. Major neurocognitive disorders due to Alzheimer disease and major vascular neurocognitive disorder together account for over 60% to 80% of all major neurocognitive disorders. Of those diagnosed with major neurocognitive disorder due to Alzheimer disease, about two-thirds are women. Being a woman has been associated with an increased risk of developing major neurocognitive disorder due to Alzheimer disease, although the evidence is not as clear for major vascular neurocognitive disorder. Research using autopsy data suggests that over half of patients with major neurocognitive disorder due to Alzheimer disease have vascular lesions.

There is a growing body of scientific literature on how cardiovascular risk factors may play a role in the development of major neurocognitive disorder due to Alzheimer disease. Obesity in midlife may influence, to a greater degree in women than in men, the development of major neurocognitive disorder. Hypertension and hypercholesterolemia in both women and men predict a higher risk of developing major neurocognitive disorder due to Alzheimer disease with aging. In women older than 75 years of age, rates of hypertension and hyperlipidemia are higher than in men of similar age. Thus, with older age, hypertension and hyperlipidemia may play a more prominent role in the development of major neurocognitive disorder due to Alzheimer disease in women. The apolipoprotein Eε4 (ApoEε4) genotype seems to be a risk factor for major neurocognitive disorder due to Alzheimer disease primarily in women and seems to have a more deleterious effect on hippocampal gross pathology and memory performance in women than in men. Diabetes, which increases in frequency with age in women to a greater extent than in men, is associated with an increased risk for cognitive impairment and may influence this trend as well. Whether prevention or reversal of these associated factors can cause a decreased risk of major neurocognitive disorder due to Alzheimer disease in women has yet to be clearly established.

Women with major neurocognitive disorder due to Alzheimer disease are at greater risk of having comorbid conditions undetected and inadequately treated, potentially affecting mortality risk. In a cross-sectional study of 340 community-dwelling older adults with major neurocognitive disorder due to Alzheimer disease, women were at greater risk of having a malnourished state that was undetected by clinicians, despite having a low body mass index (BMI) (less than 21). In a retrospective cohort study of 5,831 men and 17,918 women nursing home residents with major neurocognitive disorder due to Alzheimer disease, women were underweight, less likely to be hospitalized, and received fewer medications for each disease compared to men. The women in this cohort were also older and more severely cognitively impaired than the men. In an observational study of 28,367 nursing home residents with major neurocognitive disorder due to Alzheimer disease, it was found that although psychotic and depressive symptoms were equally prevalent in both women and men, women were more likely to receive antidepressants and men were more likely to receive antipsychotics. This may be due, in part, to the fact that men were more likely than women to have behavior problems such as wandering, abusiveness, and socially inappropriate behaviors in this cohort. These findings suggest that greater attention by gender to the nutritional, health, and psychiatric status of older adult patients with major neurocognitive disorder due to Alzheimer disease may be needed so that appropriate, timely interventions may be proactively implemented.

As major neurocognitive disorder due to Alzheimer disease progresses, a decline in cognitive, functional, and overall health status occurs, with the gradual loss of the ability to live independently. Women and men with major neurocognitive disorder due to Alzheimer disease, if admitted to a nursing home, seem to have different reasons for nursing home admission. A 2-year, four-wave follow-up study with 573 patients with major neurocognitive disorder due to Alzheimer disease and their family caregivers showed that men entered an institution primarily for medical and caregiving needs, but women, who showed less physical and cognitive impairment than men in nursing home admission, entered due to family caregiving patterns. This suggests that providing additional education and support to affect family caregiving patterns might delay nursing home placement for women with major neurocognitive disorder due to Alzheimer disease.

Mortality rates in older adults with major neurocognitive disorder due to Alzheimer disease are higher in women than in men. In a 15-year prospective epidemiologic study of 1,670 community-dwelling older adult individuals, it was found that having a major neurocognitive disorder due to Alzheimer disease increased mortality risk overall in women but not in men.

Cardiovascular risk factors, the lack of detection and adequate treatment of comorbid conditions, and patterns of family care provided to older adult women with major neurocognitive disorder due to Alzheimer disease may possibly influence mortality. If so, interventions to influence these factors might reduce mortality rates in older adult women with major neurocognitive disorder due to Alzheimer disease.

Depression

Depression in older adults is prevalent and a significant public health problem. It is two to three times more common in older adult women than in older adult men overall. The comorbidity associated with depression in older adult women is significant. Depressive symptoms statistically significantly increased the risk for major neurocognitive disorder in women 85 years and older. The prevalence of generalized anxiety disorder is higher in older adults with depression. In studies with fairly large cohorts, the association of depression with decreased health status, disability, and decreased quality of life in older women is prominent. In a prospective study of 4,657 older adult women, depression, along with underweight status, obesity, and smoking, was associated with the development of frailty. A greater prevalence of depression was found to contribute substantially to greater disability and decreased quality of life in older adult women compared to men in a study of 1,348 older adult patients. In addition, depression, and frailty in older adults, especially in older women, have been associated with an increased risk of death.

In population-based cohort studies, depression was found to be associated with increased rates of hip bone loss and lower bone mineral density (BMD) in older women compared to younger women. Higher rates of hip fracture in older adult women have been attributed to declining bone strength with age; depression may exacerbate this trend by potentially influencing rates of bone loss and BMD. Of note, some antidepressants and antipsychotics that may be used to treat depression may exacerbate bone loss by causing hyperprolactinemia and changes in bone metabolism, leading to bone demineralization.

It has yet to be clearly determined whether preventing or reducing depression might improve the health status, disability, comorbid conditions, and quality of life in older women, or vice versa. In addition, there may be complex, fluctuating, interactive relationships among these factors, so affecting one may or may not clearly lead to an association with decreased depression. In such cases, multipronged interventions that influence all implicated factors together may be most effective clinically.

Depression has been associated with psychosocial and lifestyle features. In a cross-sectional study of 580 geriatric hospital patients, depression was associated with fewer social contacts and support. This association persisted even after adjusting for physical and cognitive impairment and was stronger for women than for men. In a study of 860 community-dwelling older adults from 60 to 96 years of age, decreased depression was associated with both light and strenuous exercise status in both women and men, with light exercise having a somewhat stronger effect on depression in women than in men. These findings, albeit associations without clearly established causality, raise questions as to whether programs that promote social support and light exercise might affect depression even more in women than in men.

Although depression and some associated conditions are more prevalent in older adult women than in older adult men, there are significant adverse comorbid conditions associated with older adult men. Suicidality is associated with depression, and because depression is more prevalent in older adult women than in older adult men, it may be expected that older adult depressed women have higher suicide rates. However, the highest rates of successful suicide in the United States had occurred in older adult white men, possibly due to men using more lethal means in attempting suicide. In a community-based sample of 4,051 older adult individuals, major depression with psychotic features was found to increase mortality risk in both women and men, but milder “neurotic depression” was found to increase mortality risk in men only. Determining why women seem to be more “protected” with respect to these factors might help lead to more effective interventions for both men and women who are older adults and depressed.

Psychosis

In women, the onset of schizophrenia peaks at about age 25 to 30 years, with a second, smaller peak after age 45. In men, the age of onset is earlier, at about 20 to 25 years of age. Psychosis is often considered to be uncommon in older adults without major neurocognitive disorder. However, a 3-year longitudinal prospective cohort study of 347 older adult individuals living in community or institutional settings found that the prevalence of any psychotic symptom was 10.1%, higher than previously reported. Moreover, hallucinations, delusions, and paranoia were each related to an increased incidence of major neurocognitive disorder in older adults aged 85 to 88 years. Hallucinations and paranoia were associated with increased 3-year mortality in women but not in men.

A descriptive, cross-sectional study of 90 outpatients with early-onset schizophrenia and 34 with late-onset schizophrenia found that a greater proportion of women had late-onset schizophrenia compared to men (41% vs. 20%). Women overall had more severe positive psychotic symptoms than men, and women with late-onset schizophrenia seemed to have less severe negative psychotic symptoms than men or women with earlier-onset schizophrenia.

In general, these data suggest that women may be more prone to developing psychosis later in life compared to men and that positive psychotic symptoms are more severe and associated with greater mortality in women than in men. However, antipsychotics are frequently used to treat psychotic signs and symptoms, and the use of these medications may affect mortality rates. Both typical and atypical antipsychotics are associated with a greater risk of death, so cautious use of these medications becomes very important.

Other Psychiatric Conditions

Generalized anxiety disorder is much more prevalent among the older adult population than previously believed. In general, anxiety has been reported to be more frequent in women than in men. Among the older adult population, this trend seems to continue. In one study of 103 older adult individuals with generalized anxiety disorder, women comprised 54% (25/46) of patients with generalized anxiety disorder and 70% (40/57) of patients with both generalized anxiety disorder and major depression.

Regarding bipolar disorder, one epidemiologic study found that older adult patients with bipolar disorder reported having higher lifetime and 12-month rates of alcohol use, dysthymia, generalized anxiety, and panic disorder than those without bipolar disorder. Older adult women with bipolar disorder reported more panic disorder, whereas older men with bipolar disorder reported more alcohol dependence. Much more needs to be understood about anxiety and mania in the older adult population, especially with respect to gender issues, before specific recommendations for treatment can be made confidently. However, if the trend toward more anxiety disorders in older adult women persists, it suggests that the judicious use of antianxiety agents might be especially pertinent for women compared to men.

Deterioration of the quality of sleep occurs in both women and men with aging. However, when women enter menopause, they are at higher risk for even greater disruptions in sleep patterns. Gender-appropriate treatment may include time-limited hormone replacement therapy used in the early stages of menopause.

TREATMENT ISSUES

The management of psychiatric conditions includes psychosocial, behavioral, milieu, and other psychotherapeutic interventions, as well as biologically based treatments such as medication or brain stimulation therapies like electroconvulsive therapy and transcranial magnetic stimulation.

Psychotherapeutic Interventions

When treating geriatric patients, there are unique aspects of transference and countertransference that must be addressed. Older patients may view younger therapists as being like their children and less experienced or knowledgeable about life issues. The older adult may feel more stigmatized at having to seek psychiatric help. Although the therapist's understanding of the older patient's relationships with both parents continues to be important, it may be even more essential to determine what the patient's relationships with children and other mental health professionals have been to effectively treat the older patient. Therapists must pay special attention to their attitudes and biases regarding aging, sexuality, disabilities, losses, and death so that these vital concerns in aging are not avoided, and optimal adaptations can be facilitated with the older adult patient. Ageist and stereotyped views of older adults as "shipwrecks at the end of their voyage—hard of hearing, with low vision, physically disabled, asexual, unattractive, slow, unproductive, and close to death"—deny many older adults the professional psychiatric help they need in a respectful, gender-sensitive, reality-oriented, compassionate, and knowledgeable manner that promotes ongoing growth, development, adaptation, friendship, and enjoyment of life.

Gender-specific developmental theory for later life has not yet been established. However, clinically, eclectic psychotherapeutic techniques incorporating aspects of psychodynamic, cognitive-behavioral, interpersonal, and problem-solving therapy have been found to be useful for older women and men. Individual, couples, family, and group therapies may all be helpful in appropriate situations, and peer support and education are vital.

A psychotherapeutic issue of focus in older adult patients includes the need to adapt to being more dependent than in their younger adult years.

Older women and men who, during their younger adult lives, prided themselves on being self-sufficient and able to nurture others may find dependency on others especially difficult. When assistance with personal care is necessary, older patients may have more difficulty accepting direct care from caregivers of the opposite gender, especially if the caregiver is a son caring for a mother or a daughter caring for a father. Excessive resistance to dependency can result in self-injurious or self-neglectful behaviors. Not wanting to be a burden to others (especially one's children) is a common concern and can often be successfully addressed in psychotherapy by pointing out the vital contributions that the older patient has made to the family over a lifetime. Unrealistic and excessive feelings of entitlement and dependency on others beyond actual need and capacity, often with an excessive focus on losses and somatic concerns, may also require therapeutic attention. The individual etiologies and meaning of such behaviors need to be explored and clarified to patients in a gender-sensitive manner.

Family participation is a vital component of psychotherapy with dependent older adult persons. More women in their advanced years live alone, have more chronic illnesses, and may need assistance from adult children. The readiness for emotional, physical, and, at times, financial commitment to parental caretaking can vary widely among adult children in a family. The skilled and sensitive psychotherapist can assess and effectively distribute tasks and can help families function as a quasi-multidisciplinary team while facilitating access to appropriate community resources. By not assuming that the women in the family will take on all housekeeping and caretaking responsibilities, gender barriers to participating in caregiving can, at times, be overcome. Family involvement in medication management and treatment adherence, as well as many other activities of daily living for the older patient, remains a professional standard in the therapeutic interventions for dependent, frail older adults. Even if professional caregiving services are paid for by insurance, family involvement often is essential for the appropriate access and deployment of such services.

Another psychotherapeutic issue of focus for older patients relates to the developmental phase of becoming a grandparent. This subject needs attention in theory, research, teaching, clinical practice, and policymaking. The stereotyped image of the grandmother as an unemployed or retired babysitter, meal preparer, cookie baker, and ever-available individual has become less and less of a reality. Often, no such expectations are made of the grandfather, whose interest and assistance are often appreciated at whatever level he chooses. Grandfathers, especially when retired, may enjoy the new roles of helper and caregiver in which they may not have participated earlier in their lives. A therapist's role in a grandparent's life can include the resolution of intergenerational conflicts and facilitation of the development of this new and challenging identity. Grandmothers who live alone and at a distance are less likely to have the financial resources to visit other family members than grandfathers or grandparents who live together. Clinical observations have revealed that depriving older women of the opportunity of regular visits with adult children and grandchildren can contribute to depression and considerable distress. Other clinical observations reveal that older parents who survive alcohol use disorders from their younger years tend to get depressed when rejected by adult children who cannot forgive their abusive behaviors from earlier years. This rejection inadvertently leads to a deprivation of the grandparent role.

A third psychotherapeutic issue of focus involves addressing cultural, societal, and economic predicaments that the older adult patient may be facing. The older population includes persons born around the turn of the previous century, as well as those born after World War II. Many members of the aging population came to the United States during various immigration waves. Thus, cultural heterogeneity is vast in both older adult women and older adult men. Intergenerational and gender-based variations in cultural practice and language may lead to conflict, calling for cross-cultural psychotherapeutic intervention.

Older adults have experienced many societal changes in the United States. Gender roles for women have changed from being primarily defined by men in power and authority to being less stereotyped and more self-determined. Yet, it is common knowledge that women who succeed in this male-dominated world, be it in political, corporate, academic, or other venues, often remain in the minority. Most of these women learn the "male rules of the game," and many modify their lifestyles by remaining unattached, delaying childbearing, enduring the trials and tribulations of fertility specialization technologies, adopting children, or having no children. Communication styles, behaviors, and responsibilities within and between genders, families, and generations have changed markedly. Older adult patients may benefit from psychotherapy in dealing with the causes and consequences of these tremendous fluctuations in societal roles and circumstances.

Coping with financial issues, including payment for health care and extended care facility placement, is often a focus of psychotherapeutic intervention for older adults. Money management is becoming more complex. Greater numbers of women rise to this task earlier in life, and more women without experience in financial matters must rise to the task in later life. The economic effects of lower educational levels, traditional work roles, and income discrimination experienced by women in young and middle age become exaggerated as retired women depend on pensions and Social Security derived from preretirement income. Women are about one-half as likely as men to have pensions, and the amount of their pension is one-half as much as that of men. When men die, they are three times as likely to leave an impoverished spouse. Single, widowed, and divorced women often must purchase the in-home care that married men were provided through their spouses. Women live longer, have more disabilities, and live alone more often than men. Thus, they often require the services of nursing homes more than men do. Seventy-five percent of nursing home residents are women aged 65 years or older. Women are four times more likely than men to use Medicaid for nursing home care because of their relative indigent status. Due to relatively low Medicaid reimbursement to nursing homes, these women have little choice in the selection of a nursing home. They must share a room, their beds can only be held for a limited period in case of hospitalization, and medications are often generic. Older men with spouses to look after them at home use more acute care hospital Medicare dollars than women. Medicare does not pay for skilled nursing home care for more than 100 days (with days 21 to 100 requiring a coinsurance payment), only covers rehabilitative nursing home care after a 3-day hospitalization, and provides limited coverage for prescriptions. In the past, Medicare provided only 50% coverage for outpatient psychotherapy visits; with the movement for parity in mental health, Medicare covers 80%. Many managed care companies with Medicare contracts have discontinued care provision for the older adult population, leaving many without continuation of much-needed medical care from their primary care physicians. Education and problem-solving regarding these issues, and psychotherapeutic support that promotes the patient’s psychological strengths and resources, may empower patients to cope with these difficult situations with a greater sense of hope and mastery.

Over the last few decades, online (Internet) computer use has grown in the population in general and may be a source of psychoeducational and peer support for older adults, especially for those who are more isolated. One study of 174 subjects found no significant main effects for gender and online communication and relationship preferences for friends, family, coworkers, or unknown individuals. However, frequent Internet users had higher scores on online communication and relationship building. Another study found that although access to computers was equal for both women and men, in the year 2000, women were 0.7366 times, and in the year 2002, they were 0.5675 times less likely than men to use computers to obtain web-based health information. Yet another study found that in 2012, women used the Internet with greater social motives to search for health-related information, while men were more open to establishing a virtual relationship with a general practitioner.

Psychotherapeutic treatment for family caregivers of frail or disabled older adults is important. Family caregivers often continue to work full-time and may have children for whom they are caring. These caregivers are at times referred to as “the sandwich generation,” caught between the tasks of caring for children and older adults. Many sites of employment continue to be unfriendly or economically unable to provide their employees with assistance with family caregiving. Extended family members, including grandmothers, who traditionally might have been able to provide more care, might be employed and less available to help in child and dependent elder care or need care themselves. Reconstituted families in their second and third marriages are spread thin between children of spouses. The availability and quality of community resources for caregiving have remained inconsistent because of inadequate funding and staffing. The consequences of many of these changes are manifested in the crisis of informal caregiving for older adults who are challenged by mental illness, physical conditions, cognitive disorders, and/or functional deficits.

Pharmacologic Interventions

The complexities of incorporating gender issues are often not addressed in the scientific literature specifically, especially among older adults. However, there has been some research evaluating gender differences in the pharmacokinetics and pharmacodynamics of drugs in animals and younger humans.

Pharmacokinetics. Pharmacokinetic investigations in animals and younger patients have revealed gender differences in various factors that affect drug absorption and bioavailability, drug distribution, metabolism, and elimination. Gender differences have been documented for gastrointestinal motility and metabolizing enzyme effects, but studies have not consistently shown a difference in drug bioavailability between women and men. There are few studies examining differences in bioavailability, and these are generally confounded by variability in other pharmacokinetic factors, such as distribution, metabolism, and excretion. The available evidence suggests that gender-based differences in bioavailability are not clinically significant.

Gender-related differences in drug pharmacokinetics may be due primarily to differences in drug distribution, given that many of the factors influencing the volume of distribution differ by gender. In general, women, compared to men, have a higher percentage of body fat, lower mean body weight, decreased plasma volume, and less organ blood flow, leading to differences in the rate and extent of drug distribution. In women, compared to men, due to differences in body fat, lipophilic drugs may have greater volumes of distribution, and water-soluble compounds may have less. For example, a larger volume of distribution for diazepam has been observed in women compared to men, thought to be due to differences in proportions of body fat and gender-related protein-binding changes.

Gender-related differences in plasma protein binding may lead to pharmacokinetic differences for certain agents. Although albumin levels usually are not affected by gender, endogenous and exogenous estrogens decrease levels of α-1 acid glycoprotein (AAG) in plasma and induce hepatic AAG-glycosylation. Exogenous estrogens also increase serum-binding globulins, such as sex-hormone–binding globulins, corticosteroid-binding globulins, and thyroxine-binding globulins. Variations in levels of such plasma-binding proteins may change the amount of active agent that is available. This may be important to consider when interpreting serum drug levels.

Pharmacokinetic variability is influenced greatly by gender-related differences in drug metabolism. Hepatic drug metabolism is influenced especially by hepatic enzyme activity, as well as by liver blood flow, which is less in women than in men. The cytochrome P450 3A (CYP 3A) subfamily, especially CYP 3A4 in humans, is active in the initial hepatic metabolism of many medications. Other prominent cytochrome P450 enzymes include CYP 1A2, CYP 2B6, CYP 2C9, CYP 2C19, CYP 2D6, and CYP 2E1. Drugs also can be metabolized through intestinal CYP enzymes. The frequency of hepatic metabolizing enzyme variant alleles has been reported to differ among ethnic groups and by gender in younger humans and animals. Gender-based differences in drug levels have been attributed to gender differences in hepatic enzyme activity, which also are influenced by endogenous sex hormone production and hormonal changes associated with exogenous sex hormone use, pregnancy, and menopause. Enzyme CYP 2D6 has been found to be important in metabolizing antipsychotic agents. It has been reported that tardive dyskinesia develops more frequently in Chinese women with schizophrenia than in men due to the increased frequency of defective CYP 2D6 alleles in Chinese women. The activity of the enzyme CYP 1A2 has been found to be prominent in the metabolism of antipsychotics and is higher in females than in males.

Drug excretion usually is through the kidney or the liver. Renal clearance of drugs depends on the glomerular filtration rate, which is directly proportional to weight and thus higher in men than women in general. Gender-related differences in drug renal excretion rates are most likely due to weight differences.

Pharmacodynamics. Pharmacodynamic changes in drug response based on gender have not been as well studied as pharmacokinetic differences. Gender-related pharmacodynamic changes are differentiated from pharmacokinetic effects by showing that a given plasma level of active drug in both women and men does not result in the same pharmacologic outcome.

Many psychotropic medications have been reported to have gender-related modifications in pharmacodynamics. Women have greater impairment of psychomotor abilities from diazepam compared to men. This is thought to be due to heightened diazepam pharmacodynamic sensitivity in women. Compared to men, women show more improvement in psychotic signs and symptoms and greater severity of adverse side effects from typical antipsychotic agents. Further studies need to be done to determine whether these differences are caused primarily by pharmacodynamic or pharmacokinetic processes. Some but not all research suggests that women and men appear to show different effects of various antidepressants, although more work is needed to clarify these findings.

FUTURE DIRECTIONS

The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) includes sections on Gender-Related Diagnostic Issues for many diagnoses. Yet, it could greatly benefit from more evidence-based gender- and age-specific data, especially regarding neurocognitive disorders. Incorporating gender into the evaluation and management of psychiatric conditions in an evidence-based fashion is important but extremely complex, given that the differences or similarities in responses and outcomes that are observed between older adult women and men may be due to multiple factors. These factors may be affected by contextual, sociocultural roles and attitudes, genetically based, affected by prenatal hormone organizational effects, related to natural or drug-induced hormone fluctuating activational effects, associated with variable circadian/sleep patterns, or influenced by pharmacokinetic/pharmacodynamic characteristics. In older adults, the sum of these, as well as the effects of aging, may play a role. To evaluate these factors carefully, studies ideally would include standardized psychiatric and functional rating scales that are gender-sensitive and incorporate sociocultural contextual measures (such as those evaluating immigration status, ethnocultural and religious values, family and caregiving roles, employment status, and socioeconomic level), assess genetic factors, have greater numbers of observations to account for potential behavioral fluctuations that may be related to changes in hormone levels and circadian/sleep patterns of behavior, evaluate pharmacokinetic and pharmacodynamic factors using gender-appropriate methods, be prospective and longitudinal in nature, and enroll both women and men in adequate numbers to be able to carry out meaningful comparative analyses by gender. Routinely incorporating functional neuroimaging investigations into descriptive and psychotherapeutic/drug interventional studies also may enhance knowledge regarding gender-based differences and similarities in brain function. Data on outcomes with absolute and proportional indices, sociocultural contextual measures, genetics, pharmacokinetics, pharmacodynamics, adverse effects, dropouts, onset and rates of psychotherapeutic or drug intervention effects, and placebo response should be reported by gender. Such all-inclusive studies are challenging to design, carry out, and complete. Due to constraints in resources, time, subject participation, and other practical considerations that can affect scientific rigor, most studies are planned using partial models, focusing thoroughly on a few aspects of any broadly comprehensive study. Approaches such as study replication/semireplication with the convergence of outcomes and statistical techniques using meta-analysis or physiology-based whole-body population models for assessing the influence of individual variability on the pharmacokinetics of drugs are used to help shed light on complex issues. Still, information garnered from the appropriate analysis of primary data of well-designed, highly powered studies would be most convincing.

The Society for Women's Health Research (https://swhr.org/) highlights many specific alarming gender gaps in knowledge that hinder effective prevention, treatment, and cure for conditions that are prevalent in women. Research, clinical, and educational attention and resources need to be expanded to include the older population in ways that are equitable to their numbers, especially for older women.

Gender-based life-cycle studies, of which Erik Erikson, George Vaillant, and others have been the pioneers, should be carried out simultaneously for women and men, rather than sequentially for each gender, to provide the necessary information regarding which lifestyles and conditions promote or hinder healthy aging for women and men. Appropriate and useful engagement of aging men and women regarding the benefits of rapidly changing information technologies needs ongoing attention and development. As psychosocial processes affect illness outcomes, so does illness affect psychosocial processes. Thus, affective, cognitive, and functional status require routine assessment in primary care, and all medical records from primary care physicians, including gynecologists, need to be available to the psychiatrist before the patient's first visit. The managed care interfaces with necessary state-of-the-art medical and mental health care need serious attention and overhaul. Interdisciplinary work, rather than mental health carve-outs and fragmentation of medical, and mental health care, has been shown to benefit patient care and to be cost-effective.

The caregiving crisis needs major policy attention. Education for all primary care and specialties in age- and gender-sensitive diagnosis and interventions needs to start early in medical school and residencies.

Much progress has been made in promoting gender-sensitive research and the prevention, evaluation, and treatment of illnesses. However, tremendous amounts of work still need to be done to promote a greater understanding of the similarities and differences between genders in older adult patients relative to the description, etiology, treatment, and intervention of psychiatric illnesses and the enhancement of healthy behavior.

ACKNOWLEDGMENTS

The author expresses thanks to Dr. Marion Zucker Goldstein (1932–2019), who graciously invited me to coauthor the original version of this chapter. I would also like to recognize the contributions of Harshita Pinnamaneni, a Tufts University School of Medicine Visiting Student Program fourth-year medical student from Wayne State University School of Medicine in Detroit, MI, who participated in researching topics for this current chapter.

FURTHER READINGS

Abramson BL, Srivaratharajah K, Davis LL, et al. Women and Hypertension: Beyond the 2017 Guideline for Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults. Accessed September 27, 2021. https://www.acc.org/latest-in-cardiology/articles/2018/07/27/09/02/women-and-hypertension

Administration on Aging. 2019 Profile of Older Americans. 2020. Accessed September 27, 2021. https://acl.gov/sites/default/files/Aging%20and%20Disability%20in%20America/2019ProfileOlderAmericans508.pdf.

Alzheimer' Association. 2020 Alzheimer's disease facts and figures. *Alzheimer's Dement*. 2020;16:391–460.

Baker FC, Lampio L, Saaresranta T, Polo-Kantola P. Sleep and sleep disorders in the menopausal transition. *Sleep Med Clin*. 2018;13:443–456.

Bidmon S, Terlutter R. Gender differences in searching for health information on the Internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. *J Med Internet Res*. 2015;17:e156.

Buckley RF, Mormino EC, Amariglio RE, et al. Sex, amyloid, and APOEe4 and risk of cognitive decline in preclinical Alzheimer's disease: findings from three well-characterized cohorts. *Alzheimer's Dement*. 2018;14:1193–1203.

Byers AL, Lui L, Vittinghoff E, et al. Burden of depressive symptoms over 2 decades and risk of nursing home placement in older women. *J Am Geriatr Soc*. 2018;66:1895–1901.

Federal Interagency Forum on Aging-Related Statistics. 2020 Older Americans. Key indicators of well-being. Accessed September 27, 2021. https://agingstats.gov/docs/LatestReport/OA20_508_10142020.pdf

Girgus JS, Yang K, Ferri CV. The gender difference in depression: are elderly women at greater risk for depression than elderly men? *Geriatrics (Basel)*. 2017;2:35.

Haring B, Leng X, Robinson J, et al. Cardiovascular disease and cognitive decline in postmenopausal women: results from the Women's Health Initiative Memory Study. *J Am Heart Assoc*. 2013;2(6):e000369.

Johnson RW. What is the lifetime risk of needing and receiving long-term services and supports? Published April 3, 2019. Accessed September 27, 2021. https://aspe.hhs.gov/reports/what-lifetime-risk-needing-receiving-long-term-services-supports-0

Kerwin DR, Gaussoin SA, Chlebowski RT, et al. Interaction between body mass index and central adiposity and risk of incident cognitive impairment and dementia: results from the Women's Health Initiative Memory Study. *J Am Geriatr Soc*. 2011;59:107.

Kyomen HH, Lieberman ME. Gender issues. In: Sadock BJ, Sadock VA, Ruiz P, eds. *Comprehensive Textbook of Psychiatry*. 10th ed. Lippincott Williams & Wilkins; 2017.

Li R, Ma X, Wang G, Yang J, Wang C. Why sex differences in schizophrenia? *J Transl Neurosci (Beijing)*. 2016;1:37–42.

Love S, Miners JS. Cerebrovascular disease in ageing and Alzheimer's disease. *Acta Neuropathol*. 2016;131:645–658.

Mirza SS, Wolters FJ, Swanson SA, et al. 10-year trajectories of depressive symptoms and risk of dementia: a population-based study. *Lancet Psychiatry*. 2016;3:628–635.

National Institute of Mental Health. Suicide. Accessed September 27, 2021. https://www.nimh.nih.gov/health/statistics/suicide

Nebel RA, Aggarwal NT, Barnes LL, et al. Understanding the impact of sex and gender in Alzheimer's disease: a call to action. *Alzheimers Dement*. 2018;14:1171–1183.

Neu SC, Pa J, Kukull W, et al. Apolipoprotein E genotype and sex risk factors for Alzheimer disease: a meta-analysis. *JAMA Neurol*. 2017;74:1178–1189.

Richard W. Johnson Urban Institute. What is the lifetime risk of needing and receiving long-term services and supports? 2019.

Scommegna P. Older women live longer, but with more disability and financial challenges than men. 2019. Accessed September 27, 2021. https://www.prb.org/resources/women-live-longer-than-men/

Sundermann EE, Katz MJ, Lipton RB. Sex differences in the relationship between depressive symptoms and risk of amnestic mild cognitive impairment. *Am J Geriatr*

Psychiatry. 2017;25:13–22.

Trenaman SC, Rideout M, Andrew MK. Sex and gender differences in polypharmacy in persons with dementia: a scoping review. *SAGE Open Med*. 2019;7:2050312119845715.

United States General Accounting Office. Women's health. Women sufficiently represented in new drug testing, but FDA oversight needs improvement. 2001. Accessed September 27, 2021. https://www.gao.gov/assets/gao-01-754.pdf

Watson S, Caster O, Rochon PA, den Ruijter H. Reported adverse drug reactions in women and men: aggregated evidence from globally collected individual case reports during half a century. *EClinicalMedicine*. 2019;17:100188.

Williams LJ, Pasco JA, Jackson H, et al. Depression as a risk factor for fracture in women: a 10 year longitudinal study. *J Affect Disord*. 2016;192:34–40.

Woods NF, LaCroix AZ, Gray SL, Aragaki A, Cochrane BB. Women's Health Initiative. Frailty: emergence and consequences in women aged 65 and older in the Women's Health Initiative Observational Study. *J Am Geriatr Soc*. 2005;53:1321–1330.

Zucker I, Prendergast BJ. Sex differences in pharmacokinetics predict adverse drug reactions in women. *Biol Sex Differ*. 2020;11:32.

55.6e The Legal and Ethical Analysis of Older Adult Abuse

PAUL RICHARD GREENWOOD, L.L.B. (HONS.), ANSAR M. HAROUN, M.D. (RETIRED), DAVID NAIMARK, M.D., AND ARIANA NESBIT, M.D., M.B.E.

Americans who are 65 years or older are the fastest-growing population in the United States and in much of the developing world. For example, according to Governor Newsom's Master Plan for Aging, California's over-60 population is projected to diversify and grow faster than any other age group. More specifically, by 2030, 10.8 million Californians will be older adults, making up one-quarter of the state's population. The mounting statistics suggest that we are on the verge of essentially becoming a "graying world." Unfortunately, as the population ages, older adult abuse cases have become one of the most common crimes.

In the past, law enforcement has typically minimized the importance of cases involving older adult victims: police officers have tended to stereotype older victims, prosecutors have been reluctant to allow seniors to testify, and judges have underestimated the impact of abuse upon the victim. Within the medical profession, there has been a reluctance to accept that an obligation is owed not only to the older adult patient but also to the community to report suspected incidences of older adult abuse.

Most states have specific statutes in their criminal codes addressing "older adult abuse" crimes. However, few prosecutorial agencies take advantage of such laws, in part because of outdated perceptions about seniors. Some of the classic stereotyping of older citizens portrays them as being senile, fragile, long-winded, grumpy, hard of hearing, and unable to see well. Due to these stereotypes, police officers may write incomplete crime reports when the alleged victim is older adult because of their unfounded beliefs that the victim is incapable of providing sufficient information that could lead to a successful prosecution, that he may not have a reliable memory, that his senses are not adequate, or that he may not be strong enough physically to pursue prosecution. A number of communities have recognized both the failings of such misconceptions as well as the fact that no one discipline will be able to combat escalating older adult abuse alone. In response, many communities have developed multidisciplinary teams (MDTs) to review cases of abuse and consider different perspectives and methods of addressing the issue. Psychiatrists are an essential component of MDTs because they can provide much-needed information on cognitive issues for law enforcement, adult protective workers, and prosecutors.

DEFINITIONS

For the sake of clarity, a few definitions are proposed. There are three general categories of older adult abuse: physical, financial, and emotional/psychological. Within these broad types, there are several other subcategories.

Physical older adult abuse includes assault and battery, neglect, and sexual assault.

Assault and battery are primarily inflicted by either

1. Opportunists who target the older adult in the streets with muggings and robberies
2. Family members (typically children and grandchildren) who wield aggression against the older adult relative
3. Care providers who, out of a sense of frustration, lash out in anger against the victim

A care provider neglects an older adult when he or she willfully or deliberately fails to provide basic services of hygiene, medical support, or nutrition to the victim. The term “willful” encompasses circumstances where a caregiver intentionally “turns a blind eye” to an obvious condition that requires positive action in order to alleviate a situation that is endangering the health of the older adult patient. Examples of neglect include failing to turn a bedridden older adult, causing Stage 4 pressure sores, and deliberately withholding crucial medications resulting in a deterioration in the older adult's health.

Sexual assault typically occurs through force, threats, or deception and culminates in unwanted sexual activity. Such conduct can include rape, nonconsensual kissing or fondling, and the taking of sexually explicit photos or videos of the older adult victim.

Financial older adult abuse manifests in several forms:

- *Theft of property:* Most commonly occurs when a care provider steals personal items such as jewelry and antiques that inevitably wind up in the local pawn shop.
- *Theft of individual checks:* Also most commonly committed by a trusted care provider who rips checks from the back of the victim's checkbook and, when cashed, are typically out of sequence.
- *Misuse of an ATM card:* Most likely to be perpetrated by somebody who has either gained the trust of the victim or has threatened and intimidated the older adult into allowing the perpetrator access to the ATM card.
- *Credit card fraud:* Going beyond just the misuse of personal credit, the perpetrator successfully forges the victim's details into a new credit application that has been obtained without the victim's consent or knowledge.
- *Theft of assets:* Involves large-scale pilfering of assets such as savings accounts, stocks, or real property. Trusted individuals may employ a power of attorney (POA) or quitclaim deed, either of which will be obtained through stealth or intimidation. A POA is a legal document that permits another person to make legal decisions on behalf of the older adult. Depending on how the document is drafted, it may permit the other party to sign checks, open new accounts, or transfer assets belonging to the older adult. A quitclaim deed is a legal document that transfers an interest in real property to another person.
- *Scams:* Today's uncertain financial climate, greater accessibility to victims through the internet, and isolation brought about by the Coronavirus pandemic have created fertile ground for crooks and conmen to target the older adult with a wide variety of schemes. These schemes may include
    - Telemarketing fraud, such as bogus charity schemes, illegal sweepstakes, and false investments.
    - Door-to-door salespeople who convince the older adult victim to pay up-front for labor services such as a new roof, driveway, or other home improvements. These con artists may make off with the up-front money or stay around to bilk the senior out of his or her savings.

  - Thieves pretending to be representatives of utilities or roofing companies or some other trustworthy agency, thereby gaining access to a senior's home with the intention of either stealing what they can or casing the residence for a return visit and a more comprehensive theft opportunity.
  - Con artists exchanging a senior's assets in return for a false promise of "lifetime care." This scam takes advantage of the fact that seniors often fearfully contemplate the future when they may be less self-reliant.
  - Telephone scams involving a caller pretending to be a relative, such as a grandchild, who is in some kind of trouble and needs money.
  - Romance scams where the perpetrator meets a lonely, vulnerable senior through a dating website and then proceeds to persuade the victim to send money.
  - Threatening telephone calls demanding immediate payment with a threat that noncompliance will result in the senior citizen's arrest or a cutting off of the older adult's utilities.
  - Scammers posing as IRS agents who are investigating delinquent tax payments.
- *Emotional/Psychological older adult abuse:* Normally presents as verbal threats by a family member. Psychological abuse may include threats, intimidation, humiliation, or other such degradation. Threats may encompass hints of personal violence against the victim, a loved one, or a pet if the victim fails to deliver on financial "commitments."

NOSOLOGY

DSM-IV-TR contained a section on "Problems Related to Abuse or Neglect," which included a subsection on "Abuse of Adult." This subsection included both physical and sexual but not psychological abuse. The more recent DSM-V renames the section "Adult Maltreatment and Neglect Problems."

Like the DSM-IV-TR, the DSM-V includes Adult Physical Abuse and Adult Sexual Abuse; however, it also adds Adult Psychological Abuse. This change reflects the growing realization that maltreatment may take many forms, including:

- *Physical and sexual abuse:* recognized in both DSM-IV-TR and DSM-V;
- *Psychological abuse:* now recognized in DSM-V; and,
- *Neglect:* which has a DSM-V code for both child neglect and for spouse or partner neglect, but not for adult neglect (where neglect of older adults may be coded).

There remains no DSM-V code for:

- *Self-neglect:* where there is no perpetrator, but the older adult fails to provide for self in a proper manner, thus threatening personal health or safety;
- *Abandonment:* when a caretaker or legal guardian leaves the older adult to fend for himself or herself; or,
- *Exploitation:* where a vulnerable older adult is the victim of some scam, usually material or financial.

FACTORS ASSOCIATED WITH OLDER ADULT VULNERABILITY

Seniors often have psychological and financial characteristics that enable exploitation.

The older adult are likely to have:

- Desirable assets (i.e., money in various forms)
- Legal rights to control those assets through their decision making
- Some impairment in decisional capacities
- Psychological desire to exercise autonomy

Because older adults often recognize their limitations in their cognitive functioning and in their decision-making capacity, they may rely on others to help them make financial decisions.

Older adults often select their advisors based on:

- *Trust:* for example, family members who are trustworthy but not necessarily wise
- *Expertise:* for example, lawyers and financial planners who may be clever but not trustworthy
- *Legitimacy:* for example, individuals blessed by some authority, like a legal guardian or a priest
- *Knowledgeable:* for example, their treating doctor
- *Control:* for example, their caregiver or whoever gives them spending money
- *Charisma:* for example, whoever is able to charm the older adult

The characteristics the older adult choose for an advisor might have no relationship to the moral character of the advisor, who may be determined to promote the interests of the older adult or motivated only by self-interest.

Exploitation is likely to occur when a vulnerable older adult receives financial advice from an individual motivated only by self-interest. Such exploitation may violate the law, ethics, or both. Illegal exploitation is usually easily recognizable because the law is written in language that is precise, with a crisp boundary between what is legal and illegal. It may be more difficult to recognize ethically questionable but legal behaviors as exploitative.

MULTIDISCIPLINARY TEAMS

In recent years, in response to escalating incidents of reported older adult abuse, many communities have developed MDTs. MDTs are comprised of individuals from different disciplines who meet on a regular basis to review cases of abuse and neglect, provide resources and different perspectives on the case, and address systemic problems contributing to the case. Key members of MDTs are representatives from Adult Protective Services (APS), law enforcement, prosecutors, financial institutions, medical examiners, hospital personnel, mental health case workers, the Public Guardian, psychiatrists, and organizations that provide services to older adult community members (e.g., Meals on Wheels). Some states have created MDT teams in order to address specific types of abuse; for example, a Financial Abuse Specialist Team (FAST) or an Older adult Death Review Team (EDRT).

APS is the agency in the majority of states that provides social services to older adults and adults with disabilities. However, unlike their counterpart, Child Protective Services, APS relies heavily on the cooperation of the "client." If the victim refuses services offered by APS, then typically, the investigator has no option other than to walk away and close out the case.

There is no uniform definition of an older adult who falls under the jurisdiction of APS. Some states use age as definitive. For example, California defines an older adult as anyone 65 years of age or older, while Illinois regards an older adult as anyone 60 years or older. In further contrast, Colorado defines an older adult as anyone 70 years of age or older. Other states instead use a variety of definitions, such as "endangered adult" or "vulnerable adult."

Psychiatrists should become familiar with their local APS agency and with their state's definition of "older adult." A useful starting point is to contact the National Adult Protective Services Association (NAPSA). Additionally, an inquiry should be made of the local APS as to whether an older adult abuse prevention MDT exists. Psychiatrists can provide valuable insight at such meetings, particularly when there are discussions about undue influence and consent. Most MDTs have a Memorandum of Understanding (MOU) that clarifies issues of confidentiality and potential conflicts of interest.

An important role of the MDT is to provide ongoing education to professionals about current older adult abuse issues. Special attention should be given to improving interviewing techniques, the importance of photographs and videotaped victim interviews, and spotting red flags and indicators of abuse.

Mandated reporters need to understand their duty to report suspected older adult abuse. There should be ongoing discussions about enhancing or establishing a public awareness campaign that informs the community of the correct channels for reporting suspected cases of abuse. MDT members should orchestrate community efforts to create billboards, posters, and public service announcements on local television and radio. Slogans such as “abuse is getting old” have been found to be an effective means of getting the message out to the public.

The MDT sessions should use every opportunity to educate personnel who deal with older adults in need, such as APS caseworkers, emergency room nurses and doctors, paramedics, and fire officers, in awareness, recognition, and response. Specific educational initiatives may be as follows:

- *APS workers:* After an initial referral is made via a “1-800” telephone number, a caseworker is assigned to pay an unannounced call to the senior. These workers need to be tuned into looking for possible signs of abuse in order that they will know when to bring in law enforcement.
- *Emergency room and other medical personnel:* Sometimes, an older adult patient will appear at a hospital suffering from injuries sustained from an assault or neglect. Medical practitioners need to be able to differentiate between accidental and intentionally inflicted injuries. The victim may be unwilling to explain the cause of such injuries truthfully, and it will be up to the triage nurse to obtain as many details as possible regarding the patient’s living conditions and the events leading up to the patient’s arrival at the hospital.
- *Paramedics and Fire Officers:* These professionals may be the first responders to a domestic call for assistance when an older adult person is in need. If the paramedic is not trained to notice certain indicators of abuse or neglect, police may not be called to the scene.
- *Financial workers:* Financial older adult abuse often occurs at a bank, credit union, or stockbroker firm. The perpetrator will often take the older adult victim into a financial institution in order to prepare a POA or arrange for a transfer of money. The staff at these institutions should be trained to spot various techniques used to exploit seniors. MDT members

dealing with personnel from financial institutions can help alert them to red flags and suggest possible responses. For example, financial staff should make it a policy to discuss any alterations in financial arrangements or substantial withdrawals with older adult customers in private, away from the older adult’s companion. Some states—including Massachusetts and Oregon—have already established such protocols. These protocols provide excellent resources for other states. Several states have also enacted laws that provide immunity to banks when reporting a suspected case of financial older adult abuse to APS or to law enforcement. And some states, such as California, have decreed that any employee of a financial institution is a mandated reporter of suspected older adult financial exploitation. In addition, in recent years, the Security Exchange Commission has approved rules for placing temporary holds on the disbursement of funds or securities where the investment broker believes that the client may be the victim of older adult financial exploitation.

- *Attorneys who provide legal assistance to older adult clients:* Attorneys need to be educated about the signs of dementia and the impact that cognitive impairment can have on the older adult client’s decision-making. Psychiatrists can give valuable insight to these professionals about the need for sensitivity in gauging an older adult’s capacity to make sound legal decisions.

THE RED FLAGS AND INDICATORS OF OLDER ADULT ABUSE

Signs of older adult abuse may be mistaken for the initial symptoms of dementia or merely the result of increasing frailty. For example, caregivers may offer logical-sounding explanations for everything from unusual bruising to abnormally frequent check writing. Although many caregivers are warm, caring, and trustworthy people, professionals who interact with older adults need to examine situations with a keener eye than the general population.

A victim of older adult abuse may become increasingly withdrawn from the community and dependent upon the perpetrator. If an unannounced visit by a caseworker from APS is made, the perpetrator may insist that “this is not a good time.” Excuses might include that the victim is sleeping or is unwell. Even if access is given, the caregiver may try to prevent a private conversation between the victim and the social worker by leaving the door open or by hovering over the victim. Similarly, family and friends who attempt to make telephone calls to or visit the older adult are often prevented from speaking directly to the victim. Alternatively, the perpetrator may monitor the visit or phone call.

Other notable signs of abuse include:

1. Unexplained signs of injury such as bruises, welts, or scars. It is especially concerning if they appear to be symmetrical or shaped like a specific object.
2. Broken bones, sprains, or dislocations.
3. Overdose on prescription drugs or an indication that a medication is not being taken regularly (e.g., a prescription has more pills remaining than it should or remains unfilled).
4. Broken eyeglasses or frames.
5. Signs of physical restraints (e.g., marks around the wrists or ankles).
6. Unaccounted weight loss, signs of malnutrition, and dehydration.
7. Untreated physical problems (e.g., bed sores, urinary tract infections, etc.).
8. Unsanitary living conditions (e.g., dirt, bugs, soiled bedding, and clothes).
9. An unbathed and unkempt older adult.
10. Unsuitable clothes for the weather.
11. Living without essentials (e.g., no heat or running water, faulty electrical wiring, and other fire hazards).
12. Older adult left alone in a public place.

During financial transactions at a bank, the suspect caregiver will often attempt to control the conversation with the bank official and will endeavor to speak on behalf of the victim. The bank clerk may be told that the older adult customer is sick or hard of hearing. The suspect will justify the need for a POA or will insist that any withdrawals are made in cash.

Other financial red flags include larger-than-usual withdrawals, sudden changes in the older adult’s financial condition, items or cash reported missing from the senior’s home, and suspicious changes in wills, POA, titles, and policies. In addition, sometimes names are added to the senior’s signature card, and bills are uncharacteristically left unpaid. There may be evidence of financial activity that the senior could not have initiated because of physical impairments. There may also be contracts for unnecessary services, goods, subscriptions, or evidence of over- or undermedication.

SENIORS AND THE COURT SYSTEM

For many seniors, their own abuse will be their first exposure to the criminal justice system. Because older adults may have fears, misconceptions, embarrassment, and feelings of reluctance about pursuing their case, a treating psychiatrist should discuss such feelings with their patient prior to a hearing where the older adult will be required to testify. With the older adult's permission, the psychiatrist may consider reaching out to the prosecutor or victim advocate prior to the hearing to discuss any major concerns that the older adult may have.

If a question regarding the older adult's competency to testify arises, the psychiatrist should know that the threshold to demonstrate the witness' capacity is low. More specifically, the witness typically only needs to show that he or she understands the significance of taking the oath and is able to appreciate the difference between telling the truth and a lie.

ESTABLISHING A "THEFT" IN OLDER ADULT FINANCIAL ABUSE

It is sometimes difficult to ascertain whether sufficient evidence exists to prove that theft has occurred. In other words, it can be challenging to demonstrate that a suspect has taken property from the victim without consent and with the intent to permanently retain the said property.

Three different scenarios often occur:

Scenario One: The victim is able to testify competently that the defendant took money or property without consent. The victim may be asked to look at a check and confirm that the signature is a forgery or that a series of ATM transactions were not authorized.

Scenario Two: The victim is incompetent to testify because of a severe major neurocognitive disorder. The transaction took place at a time when the victim was already diagnosed with that advanced condition. A treating physician for the victim will need to lay the appropriate medical testimony foundation to establish that the victim would have been unable to provide the necessary consent at the time of the transaction.

Scenario Three: The victim is marginally competent to testify, although there may be some signs of short-term memory loss. The victim appears to have voluntarily transferred assets to the suspect and appears to understand the nature and extent of the transaction. The suspect may describe the transfer as either a gift or a loan. Law enforcement may have refused to investigate such a scenario on the basis that no apparent crime has been committed. The scenario may be further

> complicated by the fact that the victim, prior to the transaction, was taken to meet with an attorney (unsurprisingly, chosen by the suspect) to discuss the impending transfer. That attorney will likely testify that the client understood the significance of the transaction and wished to proceed. This last scenario is the most difficult one to prove as theft, even if the suspect has clearly manipulated the older adult victim to bring about the intended result. Lawyers refer to this conduct as "undue influence."

Historically, the concept of undue influence has been used widely in civil actions such as will contests or contract disputes. However, a growing number of states recognize undue influence as a criminal concept. As a result, creativity is required in order to introduce evidence of manipulation that will demonstrate that the transaction was completed without the victim's consent. Often a psychiatrist can provide valuable testimony either as an expert witness to explain undue influence or as a percipient witness to discuss the victim's susceptibility to manipulation.

Consent is the cornerstone of many financial older adult abuse cases. In California, the trial judge in a criminal financial older adult abuse case can instruct the jury that, in order to consent to a transaction, a person must:

- Act freely and voluntarily and not under the influence of threats, force, or duress;
- Have knowledge of the true nature of the act or transaction involved; and,
- Possess the mental capacity to make an intelligent choice regarding whether or not to do something proposed by another person.

The judge will then conclude, "Consent requires the free will and positive cooperation in act or attitude."

In grappling with the question of whether an older adult was susceptible to undue influence; certain factors should be weighed:

1. The length of the relationship between the victim and the suspect. The longer the friendship, the less likely it is that fraud can be established.
2. The nature of the relationship between the victim and the suspect. If the suspect performed certain services for the victim without any other evident remuneration, an assessment must be made to determine whether the amount of the transfer was a reasonable exchange for services rendered.
3. The spending habits of the victim prior to the transfer. If it can be established that the victim had a history of frugal spending, it might be

possible to show that this particular transaction was completely out of character.
4. The amount of the transfer compared to the residual left for the victim. Sometimes the victim is left with insufficient funds to provide for unexpected emergencies after the transaction has been completed.
5. The documentation that accompanied the transaction. If the suspect maintains that the monies were given in the form of a loan, then it might be expected that supporting paperwork evidencing a loan was drafted at the same time.
6. The question of whether the lawyer who prepared the documentation was introduced by the suspect to the victim.
7. The issue of whether the victim demonstrated any changes in behavior prior to the transaction that might suggest short-term memory loss, confusion, or a decreased ability to rationally weigh the risks and benefits of different decisions.
8. The methods used by the suspect to develop a friendship with the victim. Did the suspect create a false persona to be more interesting to the older adult? Did the suspect gradually isolate the victim from other friendships and family, causing the victim to become increasingly dependent upon the suspect? Did the suspect ply the victim with alcohol to wear down the victim’s defenses?

EXCUSES AND JUSTIFICATIONS IN OLDER ADULT ABUSE

The psychiatrist should be familiar with common reactions to accusations of older adult abuse. The alleged perpetrator may first deny that there was abuse. Because older adult abuse requires some combination of actual harm, intended harm, or bad motive, the alleged perpetrator may argue that these factors are absent. But if confronted with evidence of proven abuse, the perpetrator may employ some excuse or justification.

In an excuse, the perpetrator admits that abuse occurred but should not have. However, the perpetrator explains their conduct with a statement such as the following:

- “It was an accident” (i.e., with no intent to perform the action);
- “It was a mistake” (i.e., the action was indeed intended but executed in some faulty manner);
- “The victim provoked me”;
- “I was impaired” (e.g., tired, intoxicated, overwhelmed); or,

- “Most nursing homes treat their older adults just as I did, so I was following the norm.”

In contrast, if a perpetrator invokes a justification, they allege that, although the abuse was regrettable; given the circumstances, it was the right thing to do:

- “I did provide substandard care (i.e., abuse), but if I had demanded more resources for better care, I would have been fired. If that were to happen, then the nursing home would not have any doctor, so the residents would be without any medical coverage.”
- “Because some substandard medical coverage is better than no medical coverage, given the circumstances, I was justified in doing what I did.”

ETHICAL ANALYSIS OF SUSPECTED OLDER ADULT ABUSE

The Difference between Threats and Offers

Although most people would agree that threats are ethically problematic while offers are ethically acceptable, the distinction may be difficult to see. For example, a child says to his parents, “either leave all of your assets to me in the will, or I will put you in the cheapest nursing home because you cannot take care of yourself.” This is clearly a threat, the child is exploiting his parents’ lack of capacity for self-care, and this may qualify as older adult psychological abuse. In contrast, a child may say to his parents, “if you leave all your assets to me in the will, I will not put you in a nursing home. Instead, I will buy a larger home, and you can live with me in your old age.” Because this is an offer and thereby not exploitative, it likely does not qualify as older adult psychological abuse. Although in both cases, the parents are invited to consider substantively similar deals, in one case, the proposition is manipulative, while in the other, it is not.

The Difference between Shoves and Nudges

An advisor or caregiver should presumably respect the autonomy of the older adult; violating such autonomy is unethical. “Shoving” refers to forcing a decision onto an unwilling older adult and is clearly unethical. However, the ethics of “nudging”—where information is presented in a different manner in order to improve an individual’s decision-making process—remains controversial and depends on whether the nudge applied supports or undermines an individual’s welfare, dignity, and autonomy. Competent older adults should be masters of their own agency, but such agency can be changed, altered, or influenced by reframing information or overcoming queasiness or procrastination.

An example of a nudge is how information may be reframed if an older adult's medical interest would be promoted by taking a first-generation antipsychotic, but the older adult refuses such treatment. If the patient will only consent to medications that help him "relax," the doctor may mention that a specific antipsychotic medication will cause him to feel "mellow." If the patient then consents to take this medication, the patient maintains autonomy. In this scenario, the doctor nudged the patient into taking a beneficial medication by reframing truthful but influential information.

Similarly, if an adult child knows that his parents want to change their will so that he will become the sole beneficiary, but the parents keep postponing the will update due to procrastination, the child may encourage them to do the paperwork by reminding them that it will otherwise be a significant hassle for him to be forced to get his inheritance through probate court. This would be an example of a nudge that promotes the autonomy of the older adult parents by increasing the likelihood that they overcome their procrastination in order to make a decision that is in line with their autonomous wishes and values.

CORONAVIRUS 19 (COVID-19) AND OLDER ADULT ABUSE

Psychiatrists should also be aware that the global Coronavirus Disease 2019 (COVID-19) pandemic has made it significantly easier for perpetrators to commit older adult abuse.

COVID-19 and older adult abuse:

- Older adults have been forced to stay at home, isolated from other family members. The telephone and computer have become their lifelines, and scams have become more common.
- Scams have emerged involving so-called miracle cures for the virus or using the vaccine rollout as an excuse to obtain personal information of the older adult that is then used to commit identity theft.
- APS caseworkers have been unable to make personal visits to new clients where there are reports of alleged abuse. Instead, the initial

assessment is done over the telephone, preventing the social worker from assessing living conditions and the suspected abuser's behavior.

- Some residents in long-term care facilities have been exposed to COVID-19 and suffered an untimely demise because of a failure by the facility to reduce the risk of transmission of the disease.
- Suspects who have been arrested in "nonviolent" financial older adult abuse cases have typically been given a notice to appear in court at a much later date, often months in the future. By the time the defendant faces either a preliminary hearing or trial, the older adult victim may already have died or may be experiencing serious memory problems.
- Defendants who have been convicted of "nonviolent" financial older adult abuse are often granted noncustodial sentences so as to avoid overcrowding local jails.
- Some older American Asians have been subjected to unprovoked physical assaults in the street; such crimes are being described as hate crimes stemming from misguided beliefs about the origins of the pandemic.

CONCLUSION

Over the last few years, older adult abuse has become one of the most common crimes in the United States. The responsibility to watch out for the senior population is a collective one, and psychiatrists will play an increasingly important role in the fight against older adult abuse as the population ages. It is important for psychiatrists to learn how to identify risk factors for and signs of abuse so that older adults will be protected and the abusers held accountable. Psychiatrists should also consider joining local MDTs that develop protocols to combat older adult abuse.

CASES

Profile of a Physical Abuser of the Older Adult

Very often, the abuser of an older adult is the son of a widowed woman. He may be either divorced and has come back to live at home with his mother because of financial pressures borne by the divorce, or he is single and has never left home. He tends to be in his 40s or 50s, while his mother is usually in her mid-to-late 70s. In the majority of cases, the son is lazy and unemployed. He will have used a variety of excuses as to why he is unable to work, and his mother has accepted such excuses. He will be addicted to alcohol, narcotics, or gambling. In order to feed his habit, the son extracts money from his mother. There may come a time when the mother refuses to provide any more money, and this tends to be the flash point for violence. Violence may begin with him breaking objects in the home that he knows his mother values. As the violence escalates, he may grab his mother's arms, shake her, push her into furniture, or throw an object at her. Characteristically, the mother will not immediately call the police, perhaps out of a sense of failure or shame. Often, neighbors hear the altercations between the son and his mother, but they may also be reluctant to involve the police.

Profile of a Financial Abuser of the Older Adult

The profile of the financial abuser is more difficult to describe. The exploiter can be a family member who has been distant for many years and, after the death of the older adult's spouse, surfaces to befriend the widow or widower in the hope of acquiring some or all of the anticipated inheritance. Alternatively, the abuser can be a trusted individual such as the victim's attorney, realtor, accountant, pastor, or physician. Ultimately, many financial abusers will portray the transfer of assets as either a "gift" or a "loan." For that reason, many alleged thefts from older adults are mistakenly judged to be "civil" rather than criminal violations.

Guardianship abuse has gained national attention over the past 3 years after a high-profile case out of Nevada involving defendant April Parks. Parks was a professional guardian appointed by the probate court in Las Vegas to manage and control the finances and lives of multiple older adult clients. She used her position to drain many of her victims' bank accounts and isolate them from their families. In January 2019, Parks was sentenced to 16 to 40 years in state prison. As a result of this case and others like it, many states have begun a process of examining lesser restrictive alternatives to guardianship called supported decision-making. Psychiatrists need to be involved in decisions regarding whether an individual will benefit more from guardianship or supported decision-making.

FURTHER READINGS

Breaking Down Barriers: A Multidisciplinary Team Approach to Strengthening California's Elder Justice System. Accessed March 14, 2022. https://ovc.ojp.gov/funding/awards/15povc-21-gg-01077-nonf

California Health & Human Services Agency Master Plan for Aging 2020. California Office of the Attorney General. Bureau of Medi-Cal Fraud and *Elder* Abuse. Accessed March 14, 2022. https://oag.ca.gov/dmfea

Dubble C. A policy perspective on elder justice through APS and law enforcement collaboration. *J Gerontol Soc Work*. 2006;46:35–55.

Dyer CB, Pavlik VN, Murphy KP, Hyman DJ. The high prevalence of depression and dementia in elder abuse or neglect. *J Am Geriatr Soc*. 2000;48(2):205–208.

Financial Exploitation of Specified Adults-2165. Accessed March 14, 2022. www.finra.org/rules-guidance/rulebooks/finra-rules/2165

Han SD, Mosqueda L. Elder Abuse in the COVID-19 Era. *J Am Geriatr Soc*. 2020;68(7):1386–1387.

HelpGuide.Org, Older adult Abuse. Types, signs, symptoms, risk factors, and prevention. Accessed March 14, 2022. http://www.helpguide.org/mentall eldr_abuse-JhysicaLemotionaLsexuaLneglect.htm

Kahan FS, Paris BEC. Why elder abuse continues to elude the health care system. *Mt Sinai J Med.* 2003;70(1):62–68.

Koenig RJ 3rd, DeGuerre CR. The legal and governmental response to domestic elder abuse. *Clin Geriatr Med*. 2005;21(2):383–398.

Lachs MS, Pillemer KA. Elder Abuse. *N Engl J Med*. 2015;373:1947–1956.

Makaroun LK, Bachrach RL, Rosland A. Elder Abuse in the Time of COVID-19—Increased Risks for Older Adults and Their Caregivers. *Am J Geriatr Psychiatry*. 2020;28(8):876–880.

Mistreatment of Asian American and Pacific Islander (AAPI) Elders. https://ncea.acl.gov/NCEA/media/Publication-4.0/NCEA_RB_AAPI_2021.pdf.

National Committee for the Prevention of Elder Abuse: Best practice models. Accessed March 14, 2022. http://www.preventelderabuse.org/communities/best.html

National Research Council. *Elder Mistreatment: Abuse, Neglect, and Exploitation in an Aging America*. National Academies Press; 2003.

Rowan JM, Anetzberger GJ, et al. *Elder* abuse multidisciplinary teams. In: Factora RM, eds. *Aging and Money*. Springer, Cham; 2021.

Senior Safe Act Fact Sheet. Accessed March 14, 2022. www.investor.gov/senior-safe-act-fact-sheet#_ftnref1

State of Nevada vs. April Parks 2019.

Sunstein CR. The ethics of nudging. *Yale J Regul*. 2015;32:413–450.

55.6f Geriatric Psychiatry: Sexuality and Aging

ABESH KUMAR BHATTACHARJEE, M.D., PH.D., NIRAJ K. ASTHANA, M.D., AND DANIEL D. SEWELL, M.D.

The proportion of older adults in the United States is rapidly increasing. Data from the Pew Research Organization and the U.S. Center for Disease Control and Prevention (CDC) reveal that an estimated 10,000 Americans turn 65 years old every day and are joining the pool of over 54 million individuals who are presently over 65. Currently, the percentage of

Americans who are 65 years old or older is 15%. By 2030, when all surviving members of the Baby Boomers cohort will have reached or exceeded the age of 65, this percentage will increase to 21% and by 2060 to 25%. Because of these unprecedented demographic shifts, issues impacting the health and well-being of older adults, including sexual health, are becoming increasingly more important.

Traditionally, aging has been viewed as a period of progressive decline in physical, psychosocial, and cognitive domains, which consequently increase the health care burden and decrease quality of life. This negative view of old age contrasts starkly with the recently ascendant concept of healthy aging or successful aging, which is supported by existing research on older adults who continue to function well and be active, productive, and engaged in life. Self-ratings reveal that successful aging is a subjective perception of an individual's quality of life, which comprises physical, social, and spiritual dimensions. Sexual activity and enjoyment are considered components of these quality of life dimensions, particularly the physical and psychosocial dimensions.

Regardless of how long a person lives, experiencing human closeness and intimacy remains important. Individuals and societies react in a variety of ways to sexual intimacy among older people. Despite the rapid growth of the older population and the aggressive marketing of products aimed at sexual health in later life, sexuality in older adults has received relatively limited research attention. A review of cross-cultural data from 106 traditional societies determined that sexual behavior in older men and women was very common though not invariably observed. For example, 84% of the studies, which included information about sexual behavior in older females, documented continued sexual activity and interest among older females. The authors of this review concluded that the continued sexual activity among older members of many traditional societies and the cessation of sexual behavior among older members of other traditional societies suggested that cultural as well as biologic factors might be important determinants of sexual expression in later life in these societies.

Cultural and biologic factors also appear to be important determinants of sexual behavior in the industrialized world, where many still believe that older people stop having sex or believe that it is dangerous for older people to be sexually intimate. For some, certain cultural taboos from the Victorian era may even today inhibit sexual behavior and provoke both intrapsychic and interpersonal conflict. Older people may internalize the myth of sex in older age being "wrong" or "inappropriate," and some may choose to repress their sexual urges rather than become objects of derision or disapproval (e.g., being considered “a dirty old man” or in the case of women, “a cougar.”)

Fortunately, over the past 60 years, attitudes about later life sexuality have been evolving. Contrary to persistent stereotypes, recent studies have indicated that many older men and women remain sexually active, experience sexual desire, and regard sexual activity as an important aspect of their lives. Moreover, multiple aspects of sexual health in older adults are positively associated with self-rated health and indicators of physical function. The Women’s Health Initiative Study found that activity and satisfaction with sexuality were associated with higher self-perceived successful aging and quality of life. The acceptance of hormone replacement therapy for menopausal and postmenopausal women, the increasing popularity of oral erectogenic agents, the incorporation of “ED” into the common vernacular, and the increasing frequency of examples of later-life sexual intimacy in the media all indicate changing attitudes about sexuality in older adults.

It is possible to live a happy and fulfilling life without sexual experiences. A clinician who is knowledgeable about sexuality in later life and willing to discuss sexuality with older patients has the challenge of creating an open, safe environment for discussion without implying that any problem necessarily exists. Ironically, a therapist who addresses these issues must be aware that the introduction of the topic may be perceived as very helpful or may unintentionally trigger feelings of inadequacy or incompleteness in patients for whom sex never was or is no longer a central part of life. Sexual disinterest is a matter of concern only if a person finds it troubling or if it causes problems relating to others. All things considered, it is better to ask with sensitivity about sexual behavior, listen carefully to determine if sexual problems exist, and remember that whatever reaction ensues will serve as an opportunity for a deeper understanding to develop between the patient and the clinician.

In order to help older individuals who are experiencing a discrepancy between their desire for sexual intimacy and their actual experience of it, knowledge about the following may be useful: (1) the neurobiology of sexual function; (2) the effects of aging on sexual function; (3) the impact of gender and/or sexual minority group membership on later-life sexual health; (4) the recommended guidelines for how to take a sexual history; (5) the types of sexual dysfunction, their causes, including dementia, and their recommended treatments; and (6) the impact of institutional living on sexual expression.

NEUROBIOLOGY OF SEXUAL FUNCTION

Sexuality and the Sexual Response Cycle

The neuroendocrine perspective of sexual behavior can be summarized in a five-component model focused on neurologic and hormonal mechanisms that lead to behavioral and motivational changes during human development. These components of sexuality are sensorimotor (prenatal and neonatal spinal reflexes, such as penile or clitoral erections), valence or salience (sensory stimuli developed in the late juvenile period), attractiveness (a wide range of behaviors intended to increase sexual attractiveness), receptivity (the willingness to engage in sexual activity), and solicitation or proceptivity (the urge to seek out partners). With reproductive maturity, all become tightly coordinated until later in life, when these components may become uncoupled again because of different rates of change associated with aging.

The neuroendocrine model acquires a new level of nuance and complexity when considered through the lens of sociologic and psychological conceptualization of sexual motivation. Historically, the scientific literature tends to consider sexual interest and motivation in terms of desire and arousal and places sexual desire in a nested framework of contexts, from dyadic relationships to communities, to cultures. This perspective on sexual motivation harmonizes well with the Interactive Biophysical Model of Health, which posits that health is a function of biophysical and psychocognitive dynamics between individuals over time.

Various models of sexual arousal and response have been proposed over the past century. Masters and Johnson described four sequential phases to the “sexual response cycle”: excitement, plateau, orgasm, and resolution. Kaplan proposed a triphasic model of sexual response, which highlighted the interdependence of the three components (desire, excitement, and orgasm) and proved to be clinically helpful because the phases correlate well with various common sexual problems. For example, low libido or hypersexuality are problems with Phase I or the desire phase, erectile dysfunction (ED) or dyspareunia are problems with Phase II or the arousal phase, and anorgasmia or premature ejaculation (PE) are problems with Phase III or the orgasm phase. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM 5) uses these phases as a basis for classifying sexual dysfunctions.

Critics of more traditional, linear models for understanding the human sexual response have identified certain limitations. The first of these critiques argue that traditional models rely on the assumption that sexual desire is fundamentally spontaneous and automatic, which may not accurately describe the sexual experience for many women and may unnecessarily pathologize normal female sexual behavior. For example, it is not clear how the use of this model may have shaped the current view that hypoactive sexual desire disorder is significant among women. One of the more recent refinements of the sexual response cycle has been the addition of the concepts of “spontaneous desire,” which is believed to be more typical of men, and “stimulus-driven desire,” which is believed to be more significant for women and refers to desire aroused by stimuli such as the behavior or appearance of the partner. To build upon this, scholars have also argued that traditional models may not fully capture the role cultural, emotional, and psychological factors and relationship dynamics may play in the human sexual response.

Another concern with the older models is they fail to account for the possibility that normal sexual responses may not always progress in a linear fashion. For instance, in some individuals, arousal or excitement may occur before the development of desire. To address these potential limitations, Rosemary Basson described a nonlinear model, focusing on the female sexual response and capturing the complex biopsychosocial interactions involved. This nonlinear model describes desire as something that is not necessarily spontaneous but, instead, can be generated by both sexual and nonsexual stimuli, such as emotional intimacy, love, and other relational factors. Additionally, while desire can precede arousal, this linear progression is not always required; instead, arousal can precede desire, or sexual stimuli can increase both desire and arousal simultaneously. The nonlinear model also notes that both physical and psychological satisfaction drive emotional intimacy, which in turn may increase an individual’s likelihood of both seeking out sexual stimuli as well as being receptive to them. Since the development of Basson’s nonlinear model, researchers have also noted that, although the nonlinear model was originally developed to address shortcomings associated with describing the female sexual experience, its principles can likely be applied to both sexes. Ultimately, the nonlinear model provides a framework that includes individuals that may not have spontaneous sexual desire and provide a means for understanding additional ways that sexual desire can be developed and interact with other components of the human sexual response, like arousal and stimuli.

Recently, there has been an increasing amount of research using neuroimaging to better understand the neurobiology of sexual function. Functional neuroimaging research efforts looking at the different phases of the sexual response cycle have found characteristic patterns of brain activation during different phases. Researchers have identified that in the desire phase of the sexual response cycle, they see coactivation that is described as a “sexual wanting pattern” and involves activation of the occipitotemporal cortex, superior parietal lobule, ventral striatum (VS), amygdala, hippocampus, orbitofrontal cortex (OFC), anterior cingulate cortex (ACC), and anterior insula. Meanwhile, in the arousal state, researchers have observed what has been termed a “sexual liking pattern” involving activation of the hypothalamus, anterior and posterior insula, ventral premotor cortex, middle cingulate cortex, and inferior parietal lobule. Notably, there are very similar patterns of brain activation across gender groups and sexual orientations, as long as preferred sexual stimuli are used. Additionally, patterns of brain activation observed in these various phases were also found to be stable over time, and even when subjects were scanned a year or more apart, the brain response remained similar. Additional information about the neurobiology of the sexual response cycle can be found elsewhere in this text.

AGE-RELATED CHANGES IN SEXUAL FUNCTION AND BEHAVIOR

Aging may have both positive and negative impacts on sexual expression. Normal age-associated changes in sexual function are summarized in Table 55.6f–1. For women during and after menopause, lowered estrogen levels cause atrophy of the vaginal epithelium, labia majora, and Bartholin glands (the source of fluid that lubricates the vagina), as well as shortening and loss of elasticity of the vaginal canal. Reductions in vaginal blood flow, genital engorgement, and vaginal lubrication lead to a delay in achieving sexual arousal. Loss of lubrication may also contribute to dyspareunia. Notably, over half of postmenopausal women experience a genitourinary syndrome of menopause, which has been found to be one of the most important factors responsible for increasing the odds of female sexual dysfunction.

While orgasmic capacity is retained with age, decreased muscle tone may decrease sensitivity, orgasm may require a longer period of direct stimulation, there may be reductions in the number and intensity of vaginal contractions, and some women may experience painful uterine contractions during orgasm. Moreover, postmenopausal hormonal changes also increase the risk of pelvic organ prolapse, which has been associated with both physiologically and psychologically negative impacts that impair sexual function. In particular, chronic pelvic pain, dyspareunia, reduced self-image, and increased partner avoidance have been associated with pelvic organ prolapse.

Testosterone in the central nervous system influences female sexual behavior. Women experience declines in androgen levels with age, especially between the third and fifth decades of life. To date, scientific support for a direct relationship between reduced testosterone levels and change in sexual functioning in women is inconsistent, but, in general, lowered levels are thought to contribute to decreases in libido, arousal, responsiveness, genital sensation, and orgasm. Some recent research suggests that both estrogens and androgens may work together to enhance sexual arousal and response in women. For example, the administration of estrogen plus small amounts of testosterone to menopausal women may provide greater improvement of psychological symptoms such as a sense of well-being, depression, and fatigue, as well as sexual problems related to the effects of estrogen deprivation on the genitalia, decreased libido, and anorgasmia. Further strengthening the understanding of this relationship, recent systematic reviews of randomized controlled trials have found that, across several studies, transdermal testosterone therapy is associated with increased sexual activity and improved symptoms of sexual dysfunction in postmenopausal women. The authors do note that it is still unclear whether there is a direct relationship between circulating testosterone levels and symptoms of sexual dysfunction and whether the effect of androgen therapy on sexual function may be due to increased aromatization of estrogens.

Table 55.6f–1.
Normal Age-Associated Changes in Sexual Function

Function	Male Changes	Female Changes
Sex hormones	Decline in testosterone levels	Cessation of estrogen production Decline in testosterone levels
Genitals	Fewer and less functional sperm Decreased penile blood flow Decreased penile smooth muscle relaxation	Cessation of ovulation Decreased pelvic blood supply Shortening and narrowing of vagina Atrophy of vaginal mucosa, labia majora, Bartholin glands Decreased vaginal elasticity Decreased muscle tone
Fertility	Reduced rates of conception	Ceases after menopause
Libido	For some, may decline due to lower testosterone	For some, may decline due to lower testosterone
Arousal	Requires more tactile stimulation Erections take longer to achieve Erections are less firm Erections more difficult to sustain	Requires more tactile stimulation Decreased vaginal swelling/lubrication Decreased clitoral sensitivity
Orgasm	Takes longer to achieve Ejaculation is less forceful Decreased ejaculate volume	Takes longer to achieve Fewer, less forceful vaginal contractions
Refractory period	Lengthened by hours to days	Lengthened

Perhaps even more than in men, however, variability of sexual response in women is likely to be contextual, partner-based, and relatively independent of biologic control systems. In general, for women, the response to sexual stimuli is largely influenced by the quality of emotional intimacy with the partner. Frequently, emotional well-being and the health of the relationship contribute more to sexual enjoyment than the physiologic response. Interestingly, recent research assessing the psychobiologic factors involved in sexual function in aging women found that, although aging and menopause were associated with decreased sexual function, the physiologic and endocrine correlates of these processes, including estrogen and progesterone levels, were not predictive of overall sexual functioning. Meanwhile, psychosocial factors, including self-esteem, optimism, life satisfaction, emotional support, and relationship satisfaction, did significantly predict specific important aspects of sexual function, including pain, arousal, contentment, and orgasm.

In men, testosterone is necessary for normal libido and optimal erectile function, and, as in women, testosterone levels also progressively decline with age, particularly beginning around the fifth decade of life, while levels of sex hormone-binding globulin increase. These changes result in more bound testosterone and less free or physiologically available testosterone and are believed to be responsible for the gradual decline in sexual responsiveness in men as they age. Specifically, psychic and tactile stimuli are less effective, and full penile erections occur more slowly. Maintenance of erections also requires more direct genital stimulation. The force and volume of ejaculation may diminish, and ejaculation may not occur with every sexual encounter, but difficulties with rapid ejaculation tend to lessen with increasing age. Other age-related changes include possibly less intense orgasms, retrograde ejaculation, more rapid penile detumescence, and longer refractory periods.

Despite the clear role that testosterone plays in male sexual behavior in later life, the relationship between them remains unclear. For example, the observation that men with a deficiency of testosterone may be capable of an erectile response demonstrates some independence of sexual arousal from testosterone levels and underscores that a lack of sexual interest in men does not result only from a lack of testosterone. As with women, many psychological and relationship factors may contribute to libido in men, such as the perceived level of attractiveness of the partner, various feelings toward the partner such as resentment or hostility, issues related to maintaining control or feeling controlled, and difficulties dealing with sexual dysfunction in himself or his partner.

Research exploring the influence of race, ethnicity, and membership in a sexual and/or gender minority on older adult sexual behavior is emerging but not yet abundant. As a result, available findings should be viewed cautiously. Although supportive scientific information in not yet available, it seems very likely that the psychological factors that contribute to libido in women and men also contribute to libido in those who identify as nonbinary (NB) and who may experience a gender identity that is neither exclusively woman or man or is in between or beyond both genders. It is important to note that some NB individuals may identify as genderfluid, agender (without gender), genderqueer, or something else entirely. Karraker et al. found that in adults aged 57 to 72 years old, non-White and/or Hispanic ethnic status was positively associated with increased sexual frequency in men and decreased sexual frequency in women. Research has also indicated that, in comparison to White individuals, Black older adults may have a higher odds ratio of being in nonmonogamous relationships compared to White and Hispanic groups.

Importantly, across the spectrum of gender and sexual identities, there are a number of chronic medical illnesses that are associated with sexual dysfunction, including Parkinson disease, diabetes, chronic kidney disease, cancer, and cardiovascular disease. Moreover, sexuality in adults suffering from cognitive impairment or dementia is a complex issue that raises important ethical considerations regarding consent and capacity. This is further complicated by the fact that dementia-associated sexual disinhibition can be difficult to differentiate clinically from a cognitively impaired patient's desire to fulfill their sexual needs. Further discussion of this issue can be found in the "Sexual Behavior in Long-term Care" section of this chapter.

Frequency of Sexual Expressions in Older Adults

Until relatively recently, surveys of the frequency of sexual function in later life have, in general, been limited by methodologic short comings such as relatively small, nonrandom subject samples of convenience and questions which narrowly define sex as coitus. In addition, the studies of the effects of aging mainly involve white, middle-class, well-educated, predominantly heterosexual populations, so there is currently much less information about the role that race, culture, sexual orientation, gender identity, education, and socioeconomic conditions play in determining sexuality.

According to several major surveys, 50% to 90% of men and women over 60 continue to be *sexually active*, usually defined as having sexual intercourse at least once in the past year. In surveys of older adults who meet this definition of being sexually active, estimates of their *sexual frequency* have been obtained using multiple choice questions with options such as: once a month, two to three times a month, and once a week or more. Recent explorations of this issue have come from the National Social Life, Health, and Aging Project (NSHAP), which, in Wave 1 of this study, provided researchers with data from a nationally representative probability sample of 3,005 community-dwelling older U.S. adults (1,550 women and

1,455 men) who were 57 to 85 years old, defined sex or sexual activity as "any mutually voluntary activity with another person that involves sexual contact, whether or not intercourse or orgasm occurs," and included in-home interviews conducted in English and Spanish by professional interviewers. Analysis of this data revealed that sexual activity over the past 12 months declined with age. Overall, men were more likely than women to be sexually active, report a good quality of life, and be interested in sex. These gender differences increased with age and were greater among the 75- to 85-year-old group. For men, 84.4% for those 57 to 64, 67.4% for those 65 to 74, and 38.9% for those who were 75 to 85 years old were sexually active, whereas, for women, 61.9% for those 57 to 64, 39.8% for those 65 to 74 and 16.8% for those who were 75 to 85 years old were sexually active.

Wave 2 of this study focused on measuring sexual interest, behavior, sexual experience, frequency, and appeal of physical contact in 3,377 community dwellers who were 62 to 90 years old, and of whom 73.7% (*n* = 2,487) had a spouse or cohabiting partner. It is important to note that 54.8% (*n* = 1,365) engaged in touching the genital organs of their partners. Fifty-nine individuals who were 80 to 90 years old had sexual intercourse in the last 12 months, compared with 383 individuals who were 62 to 69 years old. In general, the measures for the salience/valence and receptivity (terms defined in an earlier paragraph under Neurobiology of Sexual Function: Neuroendocrine Model) components decrease with age. Other noteworthy findings from this study included: the inclination to make one's self sexually attractive to others is stronger in women, two-fifths of the men and women reported dissatisfaction with their partner's frequency of caring behavior that makes later sexual behavior pleasurable, a fifth of women and quarter of men had vaginal sex in the past year, and physical contact is frequent among those living with partners, whereas it is rare among those living alone.

Wave 3 of the NSHAP, completed between 2015 and 2016, provided additional insights into trends in sexual function and behavior among older adults. For instance, analysis of Wave 3 data highlighted the value of affectionate touch among older adults and noted that even nonsexual physical touch was associated with increases in relationship satisfaction, life satisfaction, and mental health for both men and women. Newer studies assessing this data have also looked at sexual well-being in older adults who participate in age-heterogeneous relationships, an area with historically limited literature. Findings from this research revealed that on most measures of relationship well-being and life satisfaction, older adults in age-heterogeneous relationships did not significantly differ from those in age-homogeneous relationships. One important finding, however, was that older women partnered with younger men did have higher odds of reporting an unhappy relationship compared to those in age-homogeneous relationships. The authors theorized that this finding might be due to issues with communication and the increased stigma older women face compared to older men when engaging in age-heterogeneous relationships. Particularly enlightening analysis from Wave 3 of the NHSAP has been that the relationship between quality of sexual life and mental health may be even stronger than previously thought. An analysis looking at data from all 3 waves of NHSAP found that being sexually active was positively related to levels of happiness and general mental health, in addition to being negatively related to levels of depression. Furthermore, higher sexual frequency and better sexual quality were also associated with better mental health outcomes.

The English Longitudinal Study of Ageing (ELSA) provided data on sexual activity and enjoyment of life in a sample of 3,045 men and 3,834 women. Analysis of this data found that for both men and women, sexual activity in the past year significantly predicted a higher mean rating of life enjoyment. Interestingly, this analysis also revealed that forms of romantic or sexual contact other than intercourse, including frequent kissing, petting, or fondling, were significantly associated with higher mean enjoyment of life for both men and women. As the openness and acceptance of the role of sexuality in late life have expanded, the proportion of older adults that remain sexually active may be increasing.

A National Poll On Healthy Aging survey conducted by the University of Michigan/AARP was administered online in October 2017 to a randomly selected, stratified group of older adults aged 65 to 80 (n = 1,002) and provided data on the frequency of sexual behaviors in older adults which was similar to the data from the NHSAP. This study found that 65% of people in this age group described themselves as still being interested in sex, and more than half, 54%, said sex is important to their quality of life. Overall, 40% of 65- to 80-year-old adults are sexually active. The specific percentages by age range were 65 to 70 years (46%), 71 to 75 (39%), and 76 to 80 (25%). Men were more likely to be sexually active than women: 51% versus 31%.

A review undertaken by Beckman et al. (2014) found that the important factors associated with being sexually active at age 70 are: Positive attitude toward sexuality; being married or cohabitating; having a happy relationship; having a partner who was physically and mentally healthy; good physical and mental health; getting good sleep, and drinking alcohol more than three times per week.

Perhaps as the result of biologic factors, social conditioning, or some combination of both, some older people were not especially interested in sex when they were younger, and, even with the increased acceptance of later life sexuality, growing older has not increased their interest. For other older individuals, sexual intimacy has become associated with painful emotional conflict resulting from or causing difficulties with their partners. One member of a couple may have a significantly different level of interest or functional ability than the other. For these individuals, the stereotype of “sexless old age” may provide a comfortable shelter from what is experienced as vexing and/or insurmountable obstacles to the achievement of sexual satisfaction. Still, other older individuals may have simply grown tired of sex and have replaced sexual intimacy with other satisfying nonsexual activities. Some older individuals become celibate as the result of a deliberate decision to share sex only with a single life-time partner who may have become too ill or disabled to have sex or who may have died. The sexual behavior of some older people may be influenced by a religious belief that sex is solely for procreation and, therefore, ends with the completion of menopause. Although it is important to be aware of older people’s sexuality, clinicians must be careful neither to oversexualize the aging process nor to overmedicalize declining sexual function and interest.

Sexually active older adults need and benefit from sex education. Table 55.6f–2 includes strategies for enhancing sexual function in later life. Older adults also need education about HIV and other sexually transmitted diseases (STDs), as well as safe sex practices, including condom use. HIV infection and acquired immunodeficiency syndrome (AIDS) are no longer problems associated only with those who are members of high-risk groups such as gay men, sex workers, and intravenous drug users. Older individuals have also become part of the epidemic, although, fortunately, in the United States, the estimated annual numbers and rates of HIV/AIDS diagnoses in the 65-year-old and older-aged groups have been decreasing. Specifically, according to the CDC, from 2008 to 2018, the estimated rate of new HIV infections among persons 65 years old and older in the United States decreased from 2.4 per 100,000 persons to 1.7. In addition, the number of adults 65 years old and older living with HIV infection who had ever been classified as stage 3 (AIDS) decreased from 748 (2 per 100,000 persons) in 2008 to 673 (1.8 per 100,000) in 2012. In terms of racial and ethnic minority differences, individuals who self-identify as Black/African American or White continue to rank highest among racial and ethnic minorities in terms of the numbers diagnosed with HIV infection and with AIDS.

Table 55.6f–2.
Strategies for Enhancing Sexual Function in Later Life

Optimize Your Physical Health

Maintain a heart-healthy diet

Get sufficient rest

Have your hearing evaluated and corrected if necessary

Learn and practice good dental hygiene, which includes flossing and seeing a dentist on a regular basis

Take good care of your skin

Avoid tobacco products

Minimize alcohol intake

Engage in regular physical exercise

Learn and perform specific exercises that can improve your appearance and function such as crunches and knee raises for a protruding abdomen and Kegel exercises for weakened pelvic muscles

Obtain optimal treatment of any medical illnesses

Inform the clinicians who prescribe medications for you that you wish to avoid medications that are likely to cause problems with sexual function

Pretreat symptoms of pain or dyspnea that interfere with sex

Use lubrication to counteract vaginal dryness and sensitivity

Optimize Your Psychological Health

Foster a positive attitude toward sexuality in later life

Put aside unrealistic expectations of sexual performance

Optimize the Health of Your Relationship with Your Sexual Partner

Maintain open, honest communication

Identify and discuss changes in sexual function

Discuss ways to adapt sexual practices to these changes

Focus extra time and energy on sensual foreplay

Select sexual practices and positions that account for declines in strength and endurance

Recently, there has also been increased research interest in the effects of ageism on sexual function in older adults. Ageism, or age discrimination, has been found across numerous studies to have negative impacts on older adults' self-image, level of emotional distress, and overall psychological well-being. Ageism may negatively impact sexual function and behavior by promoting the internalization of negative stereotypes about older adults.

Mainstream culture can depict older adults as asexual, less sexually desirable, or as impotent. Older men may face negative stereotypes about being “dirty” if they retain an interest in sexual activity as they age or no longer being masculine enough to be considered sexually desirable. Older women may also face stereotypes about sexual activity being “inappropriate” after a certain age, as well as discriminatory attitudes about no longer being sexually attractive enough. Interestingly, a recent study in 2020 assessing public views among younger to middle-aged adults about late-life sexuality found that roughly half of participants (48.9%) held overall negative views of late-life sexuality, confirming that there still exists significant ageism in how sexuality in older adults is viewed. Clinicians also appear to be at risk of ageism biases. Multiple studies have shown that health care professionals are generally less likely to discuss sexual health issues with older adults, which in turn can decrease the chance of older patients seeking help for these issues. Additionally, it has been found that this decreased engagement with older patients about sexuality may be driven by stereotyped views of aging adults, like the belief that sex was less relevant or less important in the lives of older adults. Research findings indicate that older adults who have not had sex in the past 2 years, or were not sure about their hopes or plans for sex in the future, were more likely to have experienced episodes of ageism.

Importantly, as internet and social media technologies have continued to evolve over time, the use of these technologies among older adults in relation to sexual function and behavior has continued to develop. Overall, scientific literature in the area of internet use among older adults is currently limited but growing gradually. Recent research has found that older adults are using the internet for a variety of functions related to sex and relationships. In a recent study of European adults ages 60 to 75 years old, researchers found that 36% of men and 15% of women reported using the internet for sex and love purposes, including watching pornography, finding committed partners, and seeking information or advice. Beyond this, research has also found that internet use can increase the perception of social support in older adults by making it easier to stay in touch with friends and family.

GENDER AND SEXUAL MINORITIES IN LATER LIFE

Terminology and Demographics of the Older LGBTQ+ Population

In an ongoing effort to include all gender and sexual minorities under one umbrella due to their shared formative experiences of discrimination and victimization, the term LGBT continues to expand and evolve. A more recent version of this term is LGBTQ+, where the “Q” stands for questioning or queer, and “+” represents all those who do not identify with any of the groups explicitly included in this initialism (e.g., intersex, asexual, ally, pansexual, agender, gender queer, bigender, gender variant, and pangender). When the “Q” is used to denote questioning, it refers to individuals who remain uncertain of their sexual orientation or gender. In some instances, the “Q” is used as a short-hand for “queer,” which historically has been a pejorative term for gay but which is now reclaimed by some gay men, lesbians, bisexuals, and transgendered persons as a self-affirming umbrella term. In this manuscript, the term LGBTQ+ will be used unless the referenced information relates specifically to a subset of the groups represented under the LGBTQ+ umbrella. Keeping the geriatric theme of this chapter in mind, the authors recommend caution when using the term “queer” when caring for older adults because this term may still be experienced as offensive when used as an epithet, especially among older members of a gender or sexual minority group.

For a variety of reasons, obtaining an accurate count of the number and proportion of individuals who belong to one or more of the gender and sexual minorities has been challenging. The three main sources of information about the size of the LGBTQ community are the U.S. Census, research efforts by individual researchers and academic institutions, most significantly The Williams Institute at the UCLA College of Law, and surveys conducted by the Gallup organization.

Every 10 years since 1790, the U.S. Census Bureau (USCB) has counted everyone who lives in the country, as required by the U.S. Constitution. The USCB collects basic information, including age, sex, and race, for every person. The U.S. Census Survey currently contains questions about a person’s sex, race, ethnicity, and age, but not about their gender identity or sexual orientation. Some helpful information about the LGBTQ+ community became available after The American Community Survey (ACS), a U.S. Census project, began counting same-sex households in the 2005 survey. In 2016, four federal agencies (the Department of Justice, the Department of Housing and Urban Development [HUD], the Department of Health and Human Services, and the Environmental Protection Agency) requested that the USCB include questions about gender identity and sexual orientation in the upcoming 2020 Census Survey to support the efforts by these agencies to prevent or prosecute hate crimes and reduce discrimination in housing, employment, and lending. Although initially approved by the Obama Administration in 2017, the Trump Administration ordered the Justice Department to reverse this decision. The 12-question 2020 U.S. Census Survey asked respondents to identify themselves as either “male” or “female,” which left no accurate option for individuals who are NB and do not identify as either. In June 2021, The LGBTQ Data Inclusion Act (H.R. 4176) was introduced into the House of Representatives. This act would require Federal surveys, including the U.S. Census, to include voluntary, confidential self-report questions on gender identity and sexual orientation. Although support for the LGBTQ Inclusion Act has been increasing, whether it will be enacted by Congress is uncertain.

Although the U.S. Census does not yet ask about gender identity (or sexual orientation), in 2010, a measure of married versus unmarried households was added. For the 2019 ACS and the 2020 Census, the USCB revised the household relationship question. The relationship question in these surveys asked couples living together to define their relationship with their partners in a new way—“same-sex” or “opposite-sex.” The new response categories are allowing the USCB to better understand the characteristics of same-sex households, including their number and geographic distribution, which will help produce the most comprehensive and accurate national data yet on same-sex couples, which can then be used to better inform public policy affecting LGBTQ+ people. Unfortunately, this improved methodology will not provide the number of LGBTQ+ individuals who are living alone or who are not in a relationship, nor will it provide the number of individuals in each of the groups signified by the initialism.

Data collected by the USCB has already begun to create a better understanding of the lives of same-sex couples. Contrary to the long-held societal belief that older members of the LGBTQ+ community are destined to be single, lonely, and unhappy, many have long-term relationships, are emotionally stable, and may be just as likely (or as unlikely) to be successful and happy in their later years as heterosexual individuals. There were almost a million same-sex couple households in the United States in 2017, and just over half (51.7%) of them were female-coupled households. Same-sex households make up 0.8% of all U.S. households, but some cities, such as San Francisco, have a higher percentage of same-sex households—3.0%—than the national rate, according to the ACS. The geographic distributions of same-sex households vary geographically. Interestingly, while there are slightly more female same-sex couple households in the United States, male same-sex couple households are more common in cities with the most same-sex households. For example, about 77.4% of Washington, D.C.’s same-sex households are male, compared to only 56.4% of those in Phoenix. Compared with opposite-sex couples, same-sex couples are more likely to have higher incomes, have both people employed, and be more educated.

For the first time in 2019, the Current Population Survey's Annual Social and Economic Supplement from the USCB's America's Families and Living Arrangements table included estimates of same-sex couples. At that time, there were 543,000 same-sex married households and 469,000 same-sex unmarried partner or cohabiting households. That's a small number compared to the 61.4 million opposite-sex married and eight million opposite-sex unmarried partner households. In addition, there were 191,000 children living with two same-sex parents.

As has been the case for the USCB surveys, until very recently, most scientific studies did not include NB gender or sexual orientation as demographic variables. The consistent inclusion of NB gender and sexual orientation options in future large-scale research efforts will contribute to a more detailed and accurate understanding of the LGBTQ+ population, including those who are older. In addition, there is growing awareness that the LGBTQ+ community is not monolithic, and, as previously noted, separate information about each gender and sexual minority group is needed. For example, to date, no *nationwide* survey has attempted to enumerate the older transgender community.

Another challenge has been determining how best to ask about gender and sexual and minority status. Some experts have suggested that asking about behavior (e.g., With whom have you had sexual intimacy?) rather than identity (e.g., Are you gay?) may provide more accurate information. By far, however, the most common methodology has been self-identification which depends on the perception that honestly answering is safe and will have no negative consequences. Another methodologic concern is the impact of age on a respondent's candor. Older members of the LGBTQ+ community lived through a more intensely homophobic era, may have more internalized homophobia, and may be significantly less likely to feel safe answering questions about their gender and/or sexual identities. For example, 6.5% of the individuals aged 65+ who responded to a 2012 Gallup poll either refused to answer the question about sexual orientation or stated that they did not know. It seems likely that the response rates from studies that use self-identification may underestimate the number of older LGBTQ+ individuals. As more people believe that it is safe to answer questions about sexual orientation and gender identity on large surveys, the percentages will likely increase. Fredriksen-Goldsen and Kim (2014) published a recent study that points to an increasing willingness among older adults to self-identify as LGBT on surveys.

Other than the U.S. Census, a 25-year review of the literature undertaken by Fredriksen-Goldsen and Muraco (2010) revealed that prior to 2000, information about LGBTQ+ adults came from small research studies, all of which employed self-selected samples. The first large-scale research study to include a question about sexual orientation was the inaugural California Health Interview Survey (CHIS) of 2001, and, for the very first time, the 2015 CHIS survey added a question on gender identity. In 2005, scholars at The Williams Institute at the UCLA College of Law began publishing analyses of the data from the USCB and the ACS. Subsequently, The Williams Institute has become the epicenter of LGBTQ+ demographic research.

In 2012, when the nationwide Gallup Poll added a question about sexual orientation (“Do you personally identify as lesbian, gay, bisexual, or transgender?”), 3.4% of the over 120,000 individuals responded that they were LGBT. The Gallup organization has continued to study the LGBT population. In 2017, Gallup estimated that 4.5% of U.S. adults identified as LGBT. In 2020, when the Gallup organization undertook its most recent effort to quantify the size of the LGBT population, the question used was expanded from two options (straight or LGBT) to five options: straight or heterosexual; lesbian; gay; bisexual, or transgender. This survey found that 5.6% of U.S. adults identify as LGBT, with 3.1% identifying as bisexual, 1.4% as gay, 0.7% as lesbian, and 0.6% as transgender.

Research that extends beyond just the quantification of the size of the LGBTQ+ community and includes information about variables such as their sexual behaviors, health, education, occupations, and families has the potential to amplify our understanding of the lives of older LGBT individuals. Beginning in 2005 with The National Social Life, Health and Aging Project (NSHAP), some reasonably sound scientific data about the lives of older LGBT individuals have been published, including information from the National Health Interview Survey (NHIS), the National Survey of Sexual Health and Behavior (NSSHB), and the Aging and Health Report (also known as The Caring and Aging Project) which was launched in 2010 and has yielded the largest sample to date of older LGBT adults, (n = 2,560, aged 50 to 95). The 2013 NHIS, which included 34,557 adults age 18+, was the first large health survey that asked a question about sexual orientation. While the number of individuals in the LGB group was too small (n = 795, 2.3%) to obtain statistically significant results, the LGB group members did show greater rates of cigarette smoking and alcohol consumption and higher levels of psychological distress than their straight peers.

Generational Differences and a Brief Review of the LGBT Equality Movement

People are shaped, in part, by the defining cultural mores and social climate of their formative years. The current population of older LGBTQ+ adults is composed of three generational groups, and each of these groups experienced distinctly different cultural conditions. Those born between 1910 and 1925 (currently 95 years old and older), the “Greatest Generation,” grew up in a world where public or private same-sex behaviors could result in being arrested and sent to prison and hiding one’s sexual orientation was considered self-protective. Those born between 1926 and 1945 (currently age 76 to 95 years old), the “Silent Generation,” came of age in the same repressive environment but witnessed and participated in the social upheavals of the 1960s and 1970s. In 1948, Alfred Kinsey and his associates published *Sexual Behavior in the Human Male*, which was widely discussed by the public, and purported that 10% of adult males in the U.S. were exclusively homosexual. Those born between 1946 and 1964 (currently 71 to 57 years old), the “Baby Boom” generation, lived through their teenage and young adult years at the same time as the “Gay Rights Movement” was born. The Aging and Health Report provided examples of how age/generation impacts the experience of older LGBT adults. For example, data from this study indicated that, among older LGBT adults, the rate of victimization increases with increasing age, and the rate of internalized stigma for those 80 and older is higher than for those 50 to 64 and 65 to 79.

The “Sexual Revolution” of the mid-1960s to 1970s affected both heterosexuals and those who were not heterosexual. The Gay Rights Movement began in the early 1970s and grew into a powerful force pushing for civil protections. One result of this push was the decision in 1973 by the American Psychiatric Association (APA) to remove “Homosexuality” from its list of mental disorders. The evolution of social attitudes toward LGBTQ+ people born during the “Sexual Revolution” began to evolve with unprecedented speed in the early 2000s. In 2003 the U.S. Supreme Court struck down the remaining laws against sodomy in 14 states (Lawrence v. Texas), decriminalizing male same-sex behaviors. Subsequently, one-by-one, states began to legalize same-sex marriage or civil unions. The rapid shift in attitudes towards same-sex marriage resulted in landmark decisions by the U.S. Supreme Court in 2013 (United States v. Windsor) and 2015 (Obergfell v. Hodges), which made same-sex marriage legal throughout the United States.

The current older generations of the LGBTQ+ community grew up during a time of pervasive homophobia, heterosexism, and associated discrimination and hate crimes. Before the turn of the twentieth century, few LGBTQ+ individuals had any sense of connection to others like themselves or acceptance by society at large. Most ended up marrying opposite-sex partners and either lived double lives or suppressed all homosexual sexual inclinations. Not until the Stonewall riots in 1969 gave birth to the modern gay liberation movement did the LGBTQ+ community begin to develop a sense of solidarity and systematically begin to pursue greater social acceptance. As a result, many older LGBTQ+ individuals have endured major traumas personally, politically, and socially. Many older LGBTQ+ couples are still living completely or partially in the closet. The complicated challenges of the AIDS crisis have, in some ways, both helped and hindered efforts to increase acceptance and have diverted both attention and resources away from other important issues which impact the LGBTQ+ community, such as cancer research.

The majority of potential challenges faced by older LGBTQ+ couples are identical to those faced by older heterosexual couples, but a few of the challenges may be different. The key unique challenges for older LGBTQ+ individuals are:

1. A lifetime of discrimination, especially in housing and employment, and a long-term lack of legal and social recognition combine to create deep economic insecurity for LGBTQ+ older adults
2. A reliance on chosen family, due to family rejection and legalized discrimination, creates social isolation and vulnerability, and
3. Long-term discrimination, combined with a lack of competent, inclusive health care, leads to specific mental and physical health disparities.

An older member of the LGBTQ+ community may experience a lack of support from friends, relatives, neighbors, and coworkers, as well as various obvious and subtle forms of discrimination from hospitals, various levels of government, and other institutions such as insurance companies, home health care agencies, hospice organizations, proprietary residential care (PRC) providers, and faith-based organizations. This isolation and unequal treatment can become particularly difficult when a partner is ill or dies. Some of the legal hurdles were mitigated in 2015 by the U.S. Supreme Court, which legalized same-sex marriage in all fifty states, and required states to honor out-of-state same-sex marriage licenses. This decision has helped eliminate forms of discrimination and unequal treatment in a variety of situations, such as hospital visitation, income tax benefits, insurance, retirement benefits, and inheritance. Though the legal rights of LGBTQ couples in same-sex marriage are set clear, the consistent enforcement of these rights is uneven. In addition, other civil rights of LGBTQ+ individuals remain unclear and/or unprotected in some states.

A number of organizations now exist which may provide support and guidance to older members of the LGBTQ community. The existence of these organizations may be especially important for anyone who begins the process of *coming out* as an older adult. *Coming out of the closet*, or simply *coming out*, is a figure of speech for self-disclosure of belonging to sexual orientation or gender identity minority group. Framed and debated as a privacy issue, coming out of the closet is a multifaceted experience in which each person may have both shared and unique experiences. Established in 1978, Senior Action in the Gay Environment (SAGE) is the country's largest and oldest organization dedicated to improving the lives of LGBTQ+ older adults. SAGE offers a variety of social services to older members of the LGBTQ+ community and has independent affiliates in various cities in the United States and Canada. The Lesbian and Gay Aging Issues Network (LGAIN), which is sponsored by the American Society on Aging and the National Association of Lesbian and Gay Gerontologists, has created a resource guide for older LGBTQ+ people. In addition, LGBTQ+ community centers around the country often sponsor or support social organizations for older members of the LGBTQ+ community. For example, in San Diego, California, the LGBTQ+ community center supports a social organization for older members of the LGBTQ+ community called Fellowship of Older Gays (FOG).

HOW TO OBTAIN A SEXUAL HISTORY

Knowing how to take a sexual history is the first step in helping an older person with their sexual quality of life. Many patients think that all clinicians, including psychiatrists, are knowledgeable, skillful, and interested in addressing sexual concerns. Unfortunately, education in human sexuality and on how to take a sexual history is not always emphasized in the curriculums of institutions that train health care professionals, including medical schools and residency training programs.

Though the ability to take a sexual history is not especially difficult to master, doing so may provoke certain fears such as:

1. the clinician or patient experiencing arousal while talking about sex

2. not knowing what questions to ask or how to help with a patient's sexual problems
3. unexpectedly becoming aware of one's own sexual concerns
4. having the patient see one's moral repugnance or personal distaste about certain sexual practices
5. fearing that questions about an individual's sexual health may be experienced as intrusive, inappropriate, or disrespectful, especially when caring for an older adult

   These apprehensions will be reduced by:
    a. identifying them in yourself
    b. thinking about them rationally
    c. learning about the broad range of sexual expressions
    d. understanding professional boundaries
    e. acquiring more knowledge and experience regarding the diagnosis and treatment of sexual dysfunctions; and
    f. increasing one's experience with collecting the sexual histories of older adults (i.e., practicing).

Lindau and colleagues, in 2007, reported that the prevalence of discussing sex with a physician since age 50 for men and women was 38% and 22%, respectively. Other recent research has found that psychiatrists are less likely to obtain a sexual history from older patients presenting with depressive symptoms compared with younger patients, and they are also less likely to refer to appropriate services if sexual dysfunction is identified. On the other hand, many older people are reluctant to seek help for sexual problems even if they have a significant effect on their quality of life.

The evaluation of the sexual functioning of an older individual is like the evaluation of a younger adult, although for the younger clinician, it may be more difficult due to ageist beliefs or fears of offending. General considerations include:

1. establishing rapport which may require that this topic not be explored until at least the second or third contact
2. meeting in a private and comfortable environment
3. introducing the topic by putting it in context (e.g., sexual behavior is both an indicator of health and a contributor to health)
4. intentionally asking patients for permission to talk about their sexual health
5. accepting that the first few times you introduce the topic, you may feel uncomfortable and may even blush but not let this stop you

6. being aware of your own potential prejudices and subconscious feelings
7. regardless of the age of the patient, remember that you are a trained medical expert and may have helpful knowledge to share
8. not waiting for the patient or patient’s partner to introduce the topic because they may assume that you are only interested in “medical problems” or may view sexual problems as trifling (individuals caring for a partner with dementia may be especially prone to make this assumption because of fears that asking about sex may portray them as selfish or as inadequate caregivers with misplaced priorities)
9. beginning with a neutral, nonjudgmental statement like “Have you experienced any changes in your sexual life?” or, when the older individual is being assessed for depression, “sometimes when people feel very depressed, they lose all interest in sex. Do you think this is an issue for you?”
10. using open-ended questions and active-listening techniques
11. demonstrating sensitivity to the patient’s values and beliefs
12. allowing sufficient time for discussion, which means bringing up the issue no later than the midpoint of the time period set aside for the meeting
13. not making assumptions like the person is heterosexual (perhaps as few as 10% of older persons are questioned about their sexual orientation) or that the person has a sexual problem
14. being clear that sexual activity includes both coital and noncoital behaviors
15. providing information about HIV and other STD), including the fact that condom use not only prevents conception but also prevents transmission of STDs
16. educating patients regarding age-related changes that may affect sexuality, like “often women around the time of the menopause can suffer not only with the hot flashes you have described but also with sexual problems such as vaginal dryness. Is that something you have experienced?” or “people at your age may experience ED. Is that something you have experienced?” and, if clinically indicated, discussing noncoital methods of achieving sexual intimacy
17. possibly using humor. (Note: A phenomenologic study with experienced nurses in long-term care found that the use of humor was an effective method of addressing the sexual needs of older adults).

Disclosing sexual history may be particularly sensitive for transgender people because of prior experiences of discrimination in health care. Physicians can ask patients for their gender pronouns simply by inquiring, “What name and pronouns would you like us to use?” or “I would like to be respectful—how would you like to be addressed?” As with all patients, consider assuring transgender people that the questions being asked are important for their medical care and well-being. It is important to avoid assumptions about patients’ sexual orientation or the gender or gender identity of their partners. Transgender people may be of any sexual orientation, and their partners may be cisgender men or women or other transgender people. In addition, as with all patients, it is crucial to avoid assumptions regarding the types of sexual activity (e.g., oral, anal, vaginal, or no sexual activity) in which transgender people engage. Although HIV and other STDs disproportionately affect transgender women compared to other populations, some transgender women may not be sexually active or may be in long-term, mutually monogamous relationships and thus have a low risk of HIV and other STDs. Health care providers should not assume that a patient is at high risk based on gender identity; rather, a risk assessment should be based on each individual’s sexual history.

The CDC has provided guidance regarding helpful questions for discussing sexual orientations and activities, like:

1. What types of sex do you have?”
2. “Have you had any anal intercourse?” If yes, “Was it receptive? Insertive? Was a condom always used?”
3. “Can I ask you a few questions about your sexual partner(s)?”
4. Instead of “Do you use condoms?” ask, “What do you think about using condoms?”; and
5. “Some of my patients find it difficult to talk to a potential sexual partner about HIV. What do you find to be the best way to disclose your HIV status?”

Clinicians taking comprehensive sexual histories serve to enhance the quality of life among older adults by identifying sexual concerns. Sexual dysfunction is common among older adults—approximately 25% for men and 47% for women, especially those with chronic medical conditions. Furthermore, patients prompted by their health care provider are more likely to report issues with sexual functioning, which makes communication with health care providers that much more important.

HOW TO ASSESS SEXUAL DYSFUNCTIONS IN OLDER PATIENTS

Because sexual dysfunction in later life can result from psychological, psychiatric, medical, social, or some combination of these causes, an optimal evaluation should cover each of these areas. For example, psychological factors that impact sexual function in later life include decreased self-esteem or a decreased sense of sexual attractiveness due to physical changes, which may be natural changes associated with aging or changes brought about by illness or the treatment of illness, such as a mastectomy for breast cancer. Sexual dysfunction may also develop when a member of a couple experiences an illness, such as depression, dementia, or stroke, which precludes emotional intimacy. Social issues may also trigger sexual dysfunction. The lifespan of men in the United States is shorter than that of women (75.1 versus 77.8 years), which often leaves older women without sexual partners.

Research has consistently shown a decline in sexual activity with age for both men and women, with the reasons differing by gender. For men, the decline is primarily due to age, poor health, and medications, all of which have been related to ED. For women, health is less of a factor; having a functioning partner is a more important variable. Menopause has some negative impact on women's sexual interests and desires, but psychosocial factors appear to be more important.

In a large recent survey of community-dwelling older adults who were sexually active, about half of both men and women reported at least one troubling sexual problem. The most common forms of sexual dysfunction in older women include female sexual interest/arousal disorder, female orgasmic disorder, and genitopelvic pain/penetration disorder. A 1994 survey, the Massachusetts Male Aging Study (MMAS), revealed that male sexual dysfunction, presenting as ED, diminished libido, or abnormal ejaculation, first emerges as a common problem for men in their early 40s and increases with advancing age. At age 40 years, as an example, 40% acknowledged some level of impaired sexual function, and another 10% recognized a waning sexual prowess or interest with each succeeding decade. The reality is that ED is a natural part of aging, and the prevalence increases with age. In the MMAS, the percentage of participants with ED at the ages of 50, 60, and 70 years were 50%, 60%, and 70%, respectively. Thus, it seems reasonable to conclude that nearly all men who live long enough will likely develop ED. A 9-year longitudinal follow-up study of this same cohort confirmed the age-associated declines in most domains of sexual function: sexual intercourse, erection frequency, sexual desire, satisfaction with sex, and orgasm. Rating one's health as poor was associated with a decreased likelihood of being sexually active and, among those sexually active, being more likely to report sexual problems.

The following are some guidelines on how to optimally assess sexual dysfunction in an older patient. When a sexual problem involves both the person with the dysfunction and their partner, the inclusion of both individuals in the assessment is essential. Often each person will have a somewhat different version of the problem. To initiate the interview, the clinician should ask the identified patient to describe the problem and then ask the partner, "What do you think about this report?" Such open-ended questions can be followed up by more specific questions designed to pinpoint which phases of the sexual response cycle (desire, arousal, or orgasm) are involved for the patient and the patient's sexual partner. Illnesses that cause pain, malaise, or actual physiologic or anatomical limitations may impact any of the phases of the sexual response cycle.

After the joint interview, each person should be interviewed individually. For widowed patients who are in second relationships, it may be difficult to acknowledge in front of the new partner that guilt regarding the deceased partner is contributing to the sexual problem. In fact, the loss of a partner may be the biggest barrier to sexuality for older individuals, especially for older women. Certain societal taboos, though waning, may still prevent a surviving partner from seeking a new mate. Even when a new mate is found, an older person may experience sexual difficulties as a result of guilt arising from a perception of being unfaithful to a former partner and anxiety about overcoming old patterns of physical intimacy and establishing new ones. Masters and Johnson called this "The Widower's Syndrome."

It is essential to learn about the role of sexuality in each person's life and about past sexual experiences. Potentially useful questions include: What would be an optimal frequency for you? Have there been any problems before? What has changed in your relationship? What is the meaning or value of intimacy? Some individuals may view emotional intimacy as an essential precursor to quality physical or sexual intimacy, while for others, the primary value of sex is to assuage libido.

Elucidating the time course of the development of the sexual problem may help uncover underlying etiologic factors. For example, if the problem's appearance coincided with the onset of an illness of the initiation of medication, especially if abrupt, then a causal connection should be strongly considered. Asking a question like "sometimes a person taking these medications notices sexual problems; is this something that has affected you at all?" may be helpful. Many medications, including medications with psychiatric indications like selective-serotonin reuptake inhibitors (SSRIs), have the potential to impact one or more aspects of sexual function.

In general, a gradual onset suggests an etiology related to chronic medical illness. For example, does osteoarthritic pain limit movement or physical position? Does cardiovascular disease reduce physical stamina? Are physical changes associated with aging or illness, such as a colostomy or stress-associated urinary incontinence, provoking embarrassment in the person or impacting sexual interest in the partner? Is a previously unrecognized or unsuccessfully treated psychiatric or substance use disorder present? What is the patient's current sexual ability? It may be helpful to elicit a detailed history of the patient's most recent previous sexual experience. The ability to achieve orgasm during masturbation suggests that the sexual problem may have a psychological origin.

THE TREATMENT OF SEXUAL DYSFUNCTION IN OLDER ADULTS

Just as in a younger adult, sexual dysfunction in an older patient may cause significant distress. The clinician must integrate all of the pertinent information into a compassionate and reasonable treatment plan. If there is a contributing psychological problem, individual and/or couples psychotherapy should be provided. If the sexual problem has a relevant physical or biologic component, then education (e.g., using lubricant to address pain associated with vaginal dryness), medication dose reduction or substitution (e.g., SSRIs), or surgery (e.g., the chronic phase of Peyronie disease) may be helpful.

Desire Phase Problems

The treatment of sexual dysfunctions may be viewed through the lens of the sexual response cycle. The first phase of the sexual response cycle is "desire." Desire may be influenced by psychological, psychiatric, or medical problems. For example, illnesses whose optimal treatment requires surgery may result in physical disfigurement (e.g., mastectomy), which, in turn, triggers low self-esteem and feelings of being no longer sexually attractive or desirable. Psychiatric illnesses which may decrease desire include major depression and alcoholism. Mania is often associated with hypersexuality, and the various forms of dementia may lead to increases or decreases in libido.

A variety of medical illnesses have the potential to reduce libido. One common pathway for certain medical illnesses to trigger decreased libido is a reduction in androgen levels. Illnesses associated with lowered androgen levels include chronic renal failure, chronic pulmonary disease, hepatic cirrhosis, hyperthyroidism, and HIV. Medications such as anticonvulsants and tamoxifen may also lead to lowered testosterone levels. Several medical specialty professional organizations, including the U.S. Endocrine Society, the American College of Physicians, and the American Urological Association (AUA), periodically publish practice guidelines regarding the evaluation and management of male testosterone deficiency. The AUA guidelines emphasize that up to a third of men who are treated with testosterone therapy do not meet the criteria for a testosterone deficiency diagnosis, while a large percentage of men who need testosterone therapy fail to receive it due to clinical concerns about the development of prostate cancer and cardiovascular events, even though currently available evidence does not clearly support these concerns. All of the guidelines emphasize the importance of making a diagnosis of androgen deficiency according to accepted criteria and initiating testosterone treatment only in men with consistent symptoms and signs and two early morning unequivocally low total serum testosterone baseline levels obtained on separate occasions. In some patients, measurement of free or bioavailable testosterone levels is necessary. The initiation of testosterone therapy is recommended for men with symptomatic androgen deficiency to induce and maintain secondary sex characteristics and to improve their sexual function, sense of well-being, muscle mass and strength, and bone mineral density. Testosterone therapy should not be initiated in patients with breast or prostate cancer, a palpable prostate nodule or induration, or prostate-specific antigen serum level greater than 4 ng/mL or greater than 3 ng/mL in men at high risk for prostate cancer, such as African-American men or men with first-degree relatives with a history of prostate cancer without further urologic evaluation, a hematocrit greater than 50%, untreated severe obstructive sleep apnea, severe lower urinary tract symptoms with International Prostate Symptom Score above 19, or uncontrolled or poorly controlled heart failure. Men receiving testosterone therapy should be monitored regularly and require total serum testosterone levels every 6 to 12 months. The panel members who created this guideline favored a cutoff for men of levels less than 300 ng/mL. Finally, the method by which testosterone levels should be measured is an unresolved controversy. The gold standard is equilibrium dialysis, but this test is expensive, labor intensive, and not widely available. Most laboratories measure free testosterone by the analog method, which often yields less reliable results.

The Endocrine Society's guideline for treatment with testosterone is detailed elsewhere in the text (Table 55.6f–3).

Based on currently available data, sex steroids influence female sexual function, but the measurement of androgen levels in women presenting with low sexual function is not helpful and should not be used for the purpose of diagnosing androgen insufficiency in women. Although a recent report suggests that many women following oophorectomy have no decline in sexual function, a subset of women with lowered androgen may experience lowered libido, and there is some evidence that testosterone therapy in this setting may be helpful. A recent review and meta-analysis of testosterone treatment in women provided robust support for a trial of testosterone in postmenopausal women with concerns regarding sexual function. In postmenopausal women, testosterone supplementation improved libido and several other domains of sexual response. For premenopausal women with low libido, the currently available evidence does not support measurement of testosterone serum levels or treatment with testosterone. Currently, there are no approved medications for female hyposexual desire disorder. The Endocrine Society Clinical Practice Guideline reappraised androgen therapy in women after the Task Force established by the Endocrine Society commissioned two systematic reviews of published data and reviewed several other existing meta-analyses.

The Endocrine Society Task Force recommended against:

1. making a diagnosis of androgen deficiency syndrome in healthy women because there is a lack of a well-defined syndrome and data-correlating androgen levels with specific signs or symptoms were unavailable
2. the general use of testosterone for the following indications: infertility; sexual dysfunction other than hypoactive sexual desire disorder; cognitive, cardiovascular, metabolic, or bone health; or general well-being
3. the routine use of dehydroepiandrosterone due to limited data concerning its effectiveness and safety in normal women or those with adrenal insufficiency
4. the routine prescription of testosterone or dehydroepiandrosterone for the treatment of women with low androgen levels due to hypopituitarism, adrenal insufficiency, surgical menopause, pharmacologic glucocorticoid administration, or other conditions associated with low androgen levels because there are limited data supporting improvement in signs and symptoms with therapy and no long-term studies of risk

Table 55.6f–3.
Signs and Symptoms of Low Testosterone

Men
- Erectile dysfunction
- Infertility
- Reduced beard growth

Women
- Impaired arousal
- Decreased genital sensitivity

Both
- Breast discomfort
- Chronic fatigue/lethargy
- Decreased intellectual function, including memory and concentration
- Decreased libido
- Decreased muscle strength
- Decreased physical endurance
- Depressed mood
- Hot flushes, sweats
- Increased fat mass, especially centrally
- Irritability
- Lethargy
- Loss of body hair
- Mild anemia
- Sarcopenia
- Sleep discomfort

Evidence supports the short-term efficacy and safety of high physiologic doses of testosterone for postmenopausal women with sexual dysfunction due to hypoactive sexual desire disorder. Importantly, endogenous testosterone levels did not predict response to therapy. At present, physiologic testosterone preparations for use in women are not available in many countries, including the United States, long-term safety data are lacking, and evidence suggests that transdermal forms of testosterone may be preferred. Any woman receiving testosterone therapy should be monitored for signs and symptoms of androgen excess.

Arousal Phase Problems

Sexual problems in the arousal phase of the sexual response cycle include ED, vaginismus, and dyspareunia. ED is defined as the inability to achieve and/or maintain an erection satisfactory for sexual function. Studies have found that the prevalence of ED is significantly higher among older men as compared to those in younger age groups. The MMAS found a tripling of the prevalence of complete ED among men aged 70 years as compared to those aged 40 years. Compared with men in their 20s, those in the oldest age group (70 to 75 years) had a 14-fold higher relative risk for ED. In this study, age was found to be the strongest predictor of ED. Other age-adjusted risk factors significantly associated with ED included hypertension, diabetes, heart disease, obesity, smoking, hypercholesterolemia, lack of physical exercise, poor health, depressed mood, and prostate disease.

There are now a handful of medications that may be helpful to men with sexual dysfunctions involving the arousal or orgasm phases of the sexual response cycle. Phosphodiesterase V inhibitors act peripherally on the penile vascular system to improve tumescence. These medications increase cyclic guanosine monophosphate (cGMP) in the corpus cavernosum, which relaxes the smooth muscle in the penile vasculature and fosters blood flow. Three medications in this category are now available: sildenafil, vardenafil, and tadalafil, which essentially only differ in their half-lives. Other agents for ED include transurethral and injectable prostaglandin E1, which are less convenient to administer but may appeal to some men who value the immediate and predictable response and who show poor response or intolerance to Phosphodiesterase V inhibitors. Alprostadil, a synthetic form of prostaglandin E1, is currently used, and it has a unique role in men with ED secondary to diabetes and secondary to radical pelvic surgery (e.g., radical prostatectomy).

Orgasm Phase Problems

PE and anorgasmia are sexual dysfunctions of the orgasm phase of the sexual response cycle. Increasing age often diminishes, or sometimes even resolves, problems with PE. The treatment of PE mainly includes drug therapy, which many consider first-line, and psychological and behavioral therapies. SSRIs are effective for the treatment of PE, including fluoxetine, paroxetine, sertraline, and dapoxetine. Dapoxetine hydrochloride is a short-acting SSRI and is the first and only medication approved for *on-demand* treatment of PE in European countries but not in the United States. As treatment with fewer side effects and lower cost, behavioral therapy aims to improve self-confidence and relieve anxiety and depression by continuously training men to master certain sexual skills to delay ejaculation time. In a prospective randomized trial, the combination of dapoxetine with behavioral therapy was more effective in treating lifelong PE patients than the use of dapoxetine alone. Psychotherapy uses counseling to detect and correct any interpersonal problems that can lead to PE. Physical techniques include “stop-start,” “squeeze,” “sensate focus,” and pelvic floor muscle rehabilitation. Studies have shown that the short-term effect of behavioral therapy can reach 45% to 65%, but the long-term effect is still unclear. Finally, there are data that bupropion may have some value for improving sexual function in men. In one study, both normal subjects and patients who took bupropion reported improvements in desire, arousal, and orgasm, and these effects appeared to be independent of antidepressant efficacy.

The definitions of female sexual dysfunction have recently been revised in the DSM 5, with the merging of hypoactive sexual desire disorder and female sexual arousal disorder into female sexual interest/arousal disorder. There are no FDA-approved medications for female arousal disorders or orgasmic disorders. The value of phosphodiesterase inhibitors for women with sexual dysfunction remains unclear. An open-label, nonrandomized study found that sildenafil 50 mg was well tolerated in postmenopausal women with sexual dysfunction and was associated with statistically nonsignificant improvement in vaginal lubrication, clitoral sensitivity, and orgasm frequency, and overall sexual desire or satisfaction with sexual intercourse did not improve. A controlled study of sildenafil 25 or 50 mg as needed found that both doses resulted in greater arousal, more frequent orgasms, and greater sexual satisfaction compared with placebo. Like for men, bupropion may be useful for sexual dysfunction in women. One small study ($n = 20$) found that women had significant improvement relative to placebo in all aspects of sexual function, including intensity of orgasm on either bupropion-SR 150 mg or 300 mg daily. It is important to note that several small-scale studies have assessed the effects of the above-mentioned drugs. The results are conflicting, and the drugs are not devoid of adverse effects. Currently, ongoing Phase I and II clinical trials for drugs to treat female sexual dysfunction include topical alprostadil, bremelanotide (BMT), intranasal testosterone (TBS-2), intravaginal dehydroepiandrosterone (DHEA), sublingual testosterone with sildenafil, apomorphine, bupropion, and trazodone.

Illnesses Associated with Sexual Dysfunction

The treatment of sexual dysfunction can also be viewed through the lens of specific illnesses. Almost any illness may cause sexual dysfunction. Some of these illnesses may be gender specific, for example, “senile” vaginitis and prostate cancer, and others may impact both genders. Those most often associated with sexual dysfunction are illnesses that negatively impact physical stamina or cause pain, those that require medications that adversely impact sexual function, or those that impact the endocrine system, the reproductive system, or the nervous system. Table 55.6f–4 lists some of the illnesses which may cause sexual dysfunction in later life. Given that the prevalence of serious medical illness increases with age, it is not surprising that sexual dysfunctions are more common in older patients.

Table 55.6f–4.

Some of the Medical and Psychiatric Conditions Associated With Sexual Dysfunction in Later Life

Diseases That Negatively Impact Physical Stamina or Cause Pain

- AIDS
- Chronic fatigue
- Chronic obstructive pulmonary disease
- Congestive heart failure
- Coronary artery disease
- Degenerative joint disease
- Fibromyalgia
- Irritable Bowel syndrome
- Peripheral vascular disease
- Renal failure

Diseases That Require Medications That Adversely Impact Sexual Function

- Cancer
- Hypertension

Disease of the Endocrine System

- Diabetes mellitus
- Hypercortisolism (Cushing syndrome)
- Hyperpituitarism (Acromegaly)
- Hypopituitarism
- Hyperthyroidism
- Hypothyroidism

Reproductive System Diseases

- Breast cancer
- Cervical cancer
- Ovarian cancer
- Prostate cancer
- Testicular cancer
- Uterine cancer

Diseases of the Nervous System

- Alzheimer dementia
- Epilepsy
- Generalized anxiety disorder
- Major Depressive disorder
- Multiple sclerosis
- Obsessive-compulsive disorder
- Panic disorder
- Parkinson disease
- Schizophrenia

In a patient with chronic renal disease, sexual function may be diminished due to fatigue and malaise. Chronic renal disease may also lead to sexual dysfunction due to associated vascular disease, neuropathies, and endocrine abnormalities. For example, both free and total testosterone levels are lowered in chronic renal disease but may return to normal following transplantation. In patients with chronic renal disease, sexual dysfunction may also be related to abnormalities between the pituitary and the Leydig axis. Both vitamin E and erythropoietin have been reported to improve sexual function in men with renal failure.

Studies have found that cardiovascular disease, including myocardial infarction (MI), may physically and psychologically interfere with sexual functioning and negatively impact all three phases of the sexual response cycle in older women as well as older men. Following an MI, many patients have concerns about intercourse that include anxiety about the inability to perform or fear that one may die during coitus. Angina may contribute to fears of having an MI during sex and reduce libido. In addition, the anxiety and depressive disorders that commonly occur after an MI may themselves cause sexual dysfunction.

Sexual intercourse places roughly the same demands on the heart as a vigorous walk around a city block. There is great intersubject variability in blood pressure and heart rate increases during sexual intercourse. Generally, an individual who can walk vigorously or pedal 600 kpm per minute on a stationary bicycle without symptoms, increased blood pressure, or electrocardiogram changes can safely engage in sexual intimacy, including coitus. Sexual abstinence is recommended for 8 to 12 weeks after MI, although there is minimal evidence to justify this in uncomplicated patients. After this period of abstinence, gradual resumption of sexual activity is recommended, often beginning with sexual experiences other than sexual intercourse. Because superior positions may heighten cardiac demand, side-by-side or patient-inferior positions may be recommended when coitus is initially resumed. In selected patients, the use of beta blockers and sublingual nitrates prior to intercourse may reduce angina. Patients who benefit from the use of nitrates must be instructed not to use phosphodiesterase inhibitors.

Chronic pulmonary disease impacts sexual function in at least two ways: chronic arterial hypoxia is believed to cause testosterone reductions and associated diminished desire in both men and women, and shortness of breath may limit physical ability to engage in sexual activity.

Both insulin-dependent (Type I) and noninsulin-dependent (Type II) diabetes mellitus (DM) may contribute to sexual dysfunction in men and in women. ED has been reported in up to 50% of men with Type I DM and in up to 33% of men with Type II DM. As men living with DM age, the rate of ED may approach 70%, and comorbid smoking and alcohol abuse predict earlier onset of ED. Men living with DM are believed to develop ED secondary to neurologic or microvascular lesions, and, as a result, the presence of any of the “opathies” will be associated with ED. Sexual dysfunction in women with DM is often overlooked but may occur at a rate as high as 50%. Dyspareunia secondary to decreased vaginal lubrication and difficulty achieving orgasm are among the problems reported by women living with DM.

A patient living with cancer may experience sexual dysfunction due to psychological factors, physical problems directly related to cancer, including pain, and side effects of the treatments. Women diagnosed with breast cancer may experience grieving and fears of death and disfigurement that decrease libido even before any treatment has been provided. Women whose treatment includes mastectomy and who have a history of marital difficulties, and who do not discuss with their spouse the possibility of mastectomy are at increased risk for sexual problems. A recent study compared women who had a lumpectomy to women who had a mastectomy and found that having a lumpectomy was associated with a greater sense of sexual desirability but no difference in the frequency of sexual experiences or orgasms. The treatment of sexual dysfunction following mastectomy includes minimizing medical problems that are inhibiting sexuality. Medical issues that contribute to sexual dysfunction in women undergoing treatment for breast cancer include dyspareunia and other forms of pain, fatigue, ovarian dysfunction from chemotherapeutic agents such as tamoxifen, and nausea and vomiting associated with chemotherapy. If possible, open communication between the patient and her sexual partner is helpful during all stages of treatment. It is helpful for the partner to observe the mastectomy wound soon after surgery. Rapid resumption of sexual activities is also believed to be helpful, and alternative sexual positions and vaginal lubrication may be helpful. Lastly, breast reconstruction may help improve self-image and restore self-esteem, both of which may contribute to improving sexual functioning.

Although testicular cancer is relatively uncommon in older men (incidence is highest at about 35 years of age, then declines to nearly zero by the age of 80), staging surgery may result in ejaculatory dysfunction and ED. Even in the absence of these problems, surgical removal of one or both testicles may cause problems with self-esteem. However, the increasing use of testicular prostheses has greatly reduced this concern for most men. Prostate cancer increases in frequency with age. The relative success of nonsurgical interventions and the introduction of surgical techniques, which are nerve-sparing, have helped to decrease ED and other sexual problems in men being treated for prostate cancer. Patients may need to wait 2 to 3 years following surgery for maximum sexual function to return. One study reported that radiation therapy caused ED in over 50% of the men who were studied.

As is true for male genital malignancies and for breast cancer in women, female genitourinary cancers are often associated with serious emotional problems which impact sexual function. Sexual difficulties with cervical carcinoma vary with the treatment approach. In general, surgery has fewer side effects than radiation therapy. Radical hysterectomy compromises sexual function because of vaginal shortening, but this may improve when there is ongoing sexual activity. The most important variable predicting sexual dysfunction following hysterectomy is whether the ovaries are removed. Radiation therapy promotes vaginal atrophy and stenosis due to impaired ovarian function. The extent of surgical treatment for vulvar carcinoma often determines sexual side effects. From a medical perspective, local excision often does not interfere with sexuality. More extensive surgeries may cause numbness.

Sexual problems in patients diagnosed with colon cancer are usually related to chemotherapy-associated side effects or to surgeries resulting in ostomies. Fear of fecal spillage and odor may inhibit sexual enjoyment. Women with ostomies can develop dyspareunia due to fistula formation. Men who require posterior abdominal perineal resection may become impotent due to the ablation of autonomic nerves to the penis.

The desire or need to express one’s sexuality does not expire with age, and for many older people, including those living in an aged care facility, sexuality continues to be important. A variety of outcomes have been observed when a spouse or significant other has dementia. Research indicates that 73% to 80% of the time, the presence of dementia has a negative impact on a couple’s sex life. Conversely, about one-quarter of the couples who have been studied continue a warm and satisfying sexual relationship. A significant number of sexually inactive caregiving partners (39% in one study) are dissatisfied with the inactivity. Clearly, there is no one right answer for everyone. Ballard and Poer have written that in the absence of a cure or treatment, the most significant thing that can be done for a patient with dementia is to adapt our approaches with each change in symptoms to make life as meaningful and satisfying as possible. Though written with the patient in mind, this quote and underlying sentiment seem equally relevant to caregiving spouses or partners. For patients with dementia and/or their caregiving partners, the opportunity to discuss their sexual concerns with a health care professional can be tremendously helpful. According to research, however, health care professionals only rarely inquire about the impact that dementia is having on a couple's sex life.

Among the factors that may influence whether a couple continues to have satisfying sexual experiences are:

1. the type of dementia
2. the stage of the dementia
3. the couple's previously established patterns of sexual intimacy
4. the impact of dementia on the patient's ability to maintain two-sided intimacy
5. the impact of dementia on the patient's knowledge of how to engage in and follow through with various sexual activities and on the sexual activities may be preferred or even demanded
6. the influence of dementia on the patient's personality, libido, grooming, and personal hygiene
7. the impact of dementia on the caregiving partner's view of the patient
8. the effect of caregiving on the spouse or sexual partner of the patient
9. whether the patient with dementia is able to continue living at home or requires institutionalization

Regarding the partner with dementia, possible changes include remaining interested but developing problems with follow through or with apraxia for one or more sexual acts, becoming hypersexual, or making new or unusual requests regarding sexual activities. At the time of diagnosis, ED may be present in around half of the men living with dementia. At some point, the erosion in intellectual function may make it difficult or impossible to accurately assess the sexual health of the relationship and to provide emotional intimacy.

For a variety of reasons, a caregiving spouse or partner may lose interest in sex. This loss of interest may be the result of overwhelming grief or physical exhaustion from the demands of caregiving. Sometimes caregiving partners deny their own needs, including their sexual needs to assuage survivor guilt or as a way of showing their respect for and loyalty to the person with dementia. It is also possible that dementia may cause such drastic changes in the patient's values and personality that the caregiving partner sees the patient as a child or in other ways so different that sexual intimacy begins to feel wrong or uncomfortable. Caregiving partners may sometimes need to be reminded that they do not have to participate in an activity that is frightening, upsetting, or no longer satisfying. A caregiving partner has the right not to be abused. Due to the level of the patient's cognitive impairment, the caregiving spouse may wonder if having sex with the partner is a form of rape. Ballard and Poer suggest that if the dementia patient initiates the experience and if the experience is mutually satisfying to both, then it should not be viewed as rape. It is quite possible that a patient might be incapable of giving informed consent for medication or research because of the complexity and novelty of the decision but still might be capable of giving informed consent to participate in a sexual relationship, a decision that is based more on old learning.

Patients with Alzheimer dementia will usually experience dramatic changes in short-term memory relatively early in the course of the illness. But personality and social graces do not erode until relatively late in the illness. The preservation of personality and social skills may allow a couple living with Alzheimer dementia to maintain satisfying sexual experiences through most of the illness as long as ED is not problematic. One of the earliest studies of sexual function in patients with Alzheimer dementia found that loss of erection was reported in 53% of male patients around the time of their initial diagnosis. In addition to ED, another potential situation that may negatively impact the quality of sexual life of a couple living with dementia is when short-term memory deficits trigger delusions of infidelity because the patient with dementia has trouble remembering where the caregiving spouse has gone or how long the partner has been absent and then begins to suspect that the partner is having an affair.

Patients with frontotemporal dementia (FTD) very often experience personality changes or profound language problems early in the course of their dementia, and either of these can negatively impact the sexual quality of life. For example, personality changes may cause the patient with dementia to behave so differently that their partner begins to feel as if they are with a stranger. Sometimes the expressive or receptive aphasia seen in patients with FTD is so severe that it is no longer possible to maintain two-sided intimacy that, for some patients and their partners, is an essential ingredient for safe and satisfying sexual experiences.

The later the stage of the dementia illness, the more difficult it becomes to maintain sexual intimacy. One study of 20 Caucasian male Alzheimer patients and their nondemented caregiving wives found that the patients with dementia reported no reduction in quality of sexual function regardless of the stage of illness. The wives of these patients, however, reported that the more advanced the stage of dementia, the worse their quality of sexual life became. The life expectancy of patients with many forms of dementia may exceed 12 years or more. For some healthy, active spouses, this may become an unacceptably long time to live without satisfying emotional and sexual experiences. Certainly, giving and receiving affection, and physical intimacy, are needs that do not die with the overwhelming responsibilities of becoming a caregiver. As a result, some caregiving spouses may begin a new intimate relationship while continuing to provide for the needs of the ill partner. When this occurs, the adult relatives of the patient with dementia may have a variety of reactions. Some family members may be very sympathetic. Others may fail to understand the healthy partner's need for an intimate relationship. Other family members may feel the need to protect the ill partner's interests or may have fears that the well partner's new alliance will threaten the financial assets of the ill partner or of their own anticipated inheritance. Sometimes the most difficult aspect of the well partner's decision to begin a new intimate relationship is that this decision makes it impossible for other family members to deny the reality of the tragic illness that has befallen the patient with dementia.

Dementia may also include the development of inappropriate sexual behaviors and the development of behaviors that appear sexual but are actually not. Caregivers of patients with dementia, whether a partner, some other family member, or a paid professional, may experience distress, discomfort, and uncertainty about how to manage inappropriate sexual behaviors. Professionals should assess these behaviors with the assumption that there is a valid need to be met, keeping in mind that not all behaviors that appear sexual may be sexual. For example, touching the genital area repeatedly may not represent masturbatory efforts but may be a signal that help with toileting is needed. Climbing in bed with the patient in the room next door may not be an act of sexual aggression but the expression of a desire for comfort, companionship, or protection or simply the expression of a lifelong habit of having a bed partner. The long-term care section of this chapter contains further discussion of this topic.

Medications Associated With Sexual Dysfunction

A variety of classes of medications and specific individual medications may cause sexual dysfunction in adults of all ages, including older adults. Table 55.6f–5 provides a list of some of the medications which have the potential to cause sexual side effects. Several classes of psychotropic medications can negatively impact sexual function. Men may experience ED from SSRIs. Both men and women may experience decreased sensitivity of the genitalia, delayed orgasm, or anorgasmia from SSRIs. Any of the psychotropic medications with significant dopamine receptor blocking ability may also cause a variety of sexual problems, which are believed to be mediated, at least in part, through medication-associated hyperprolactinemia and associated decreased androgen levels. Drugs, which increase gamma amino butyric acid (GABA) activity, such as benzodiazepine, have been reported to cause less satisfying orgasms for women.

Table 55.6f–5.

Some of the Medications Associated With Sexual Dysfunction in Later Life

Analgesics

- Methadone
- Morphine sulfate

Anti-Infective Medications

- Ketoconazole

Antineoplastic Medications

- Tamoxifen

Cardiovascular Medications

- Amiodarone (Cordarone)
- Angiotensin-converting enzyme inhibitors
- Calcium channel blockers
- Clonidine (Catapres)
- Digoxin (Lanoxin)
- Disopyramide (Norpace)
- Guanethidine (Ismelin)
- Prazosin (Minipress)
- Reserpine
- Spironolactone
- Thiazide diuretics

Endocrine Medications

- Adrenal corticosteroids
- Leuprolide (Eligard, Lupron)

Gastrointestinal Medications

- Cimetidine (Tagamet)

Neurologic and Psychiatric Medications

- Antipsychotic medications
- Benzodiazepines
- Carbamazepine (Carbatrol, Equetro, Tegretol)
- Carbidopa/levodopa (Sinemet)
- Divalproex sodium (Depakote)
- Lithium carbonate (Eskalith, Lithobid)
- Monoamine oxidase inhibitors
- Phenytoin (Dilantin)
- Primidone (Mysoline)
- Selective serotonin reuptake inhibitors
- Trazodone
- Tricyclic antidepressants
- Venlafaxine (Effexor)

Other

Diphenhydramine (Benadryl)
Cocaine
Hydroxyzine (Atarax, Vistaril)
Nicotine
Alcohol

Antihypertensive medications may also impair sexual function, although recent research suggests that beta-adrenergic blockers may not be as problematic as originally believed. Digoxin may decrease sexual desire. Agents which block the histamine receptor subtype 2 (H2 blockers) frequently interfere with penile erections, although not all medications in the class may do this equivalently. For example, cimetidine is more likely than ranitidine to cause ED. Lastly, alcohol, though not a medication, may seriously impair sexual function.

SEXUAL BEHAVIOR IN LONG-TERM CARE

In the United States, 4.2% of individuals 65 years old and older live in nursing homes (NH). This estimate does not include older adults living in other residential care settings. Over the past several decades, a spectrum of terms have been coined to describe the various levels of care provided by, and the configurations of, these residential care facilities for older adults, including independent living, assisted living, subacute care hospitals, rehabilitative facilities, and retirement communities. In addition to these terms, other terms describing housing options for older adults include senior community, 55 and up community, senior retirement community, senior living community, multilevel senior community, and continuing care retirement community (CCRC). In this manuscript, the abbreviation PRC (defined earlier in this manuscript) has been used when referring to the above spectrum of housing options for older adults. If something is unique to a specific housing option, this has been highlighted. As the older adult population increases, our understanding of each PRC setting's unique cultural environment will be critical to implementing successful transitions and promoting the health and well-being of older adult residents.

The exact numbers of older adults living in each type of senior housing across the United States are not available due to the heterogeneity of labels and defining features, as well as the lack of central data collection. In 2017, the not-for-profit senior living organization National Senior Campuses reported over 19,000 senior living units across the United States, up from the

15,500 estimates provided in 1998 by The American Senior Housing Association (ASHA), currently estimates that one million Americans live in some type of senior living community and that number is expected to double by the year 2030.

Most older adults, even those residing in long-term care environments, maintain some level of sexual interest, albeit at decreased levels. Future cohorts of older adults are expected to have even greater sustained interest in sex, as well as a higher frequency of participation in sexual acts, as generational acceptance and the perception that sexuality is normal for older individuals increases. Evidence suggests, however, that sexual desire is mediated not only by age but also by psychosocial factors, including partner availability, frequency of contact, and generational perceptions of the appropriateness of sexual activity in later life, as well as biologic factors, including the presence of chronic diseases and the medications used to treat them. For residents living in long-term care settings such as assisted living, one or more of these factors are likely to be highly relevant. A survey of 250 NH residents in Texas determined that only 8% of them had been sexually active (masturbation or intercourse) during the previous month, but 17% of the inactive residents said they would have liked to be active but lacked the opportunity. Preferred forms of sexual expression vary widely. Sexuality and intimacy are manifested in various ways, including intercourse, but findings overwhelmingly reveal that intimate touch, hand holding, and other less physically intense expressions are common and valued, just as significantly as younger individuals might enjoy more vigorous physical expressions of sexuality.

Although addressing the continuing sexual needs of older residents of PRC settings can be a difficult task to accomplish, these should be addressed with the same priority as nutrition, hydration, personal hygiene, and other well-accepted needs. Older members of gender and sexual minorities are especially at risk for challenges and difficulties with comfortably meeting their sexual needs in a residential care environment. Sexual behavior in PRC facilities is no longer a private matter and is often open to public scrutiny. In one way or another, sexual behavior in PRC facilities has an impact on staff members, other residents, and the families of residents and is, therefore, much more often inhibited by other residents, staff, and family.

The biggest potential barriers to achieving sexual satisfaction for residents are:

1. the culture of the residential care community, including the discomfort nurses, physicians, and other health care providers may have in discussing sexuality with older adults
2. the physical design of the PRC environment
3. existing government policies and regulations
4. internalized ageism, heterosexism, homophobia, and transphobia among the residents
5. attitudes and preferences of family members

Historically, the culture of most residential care environments for older adults has not supported the sexual needs of their residents. Although there is an increasing emphasis on the application of person-centered approaches to care delivery, many providers of residential care services for seniors still neglect to recognize and address sexuality as a component of well-being. The cultural beliefs and attitudes of personal and professional caregivers are major contributors to the culture of the PRC community and may become another common obstacle to sexual expression. Many administrators and caregivers are unaware of the importance of facets of the care environment that impact the ability of residents to express their sexuality and are unaware of what sort of support may be required. Studies have shown that NH staff members' attitudes about later-life sexuality are primarily negative. Such attitudes often lead caregivers and members of the management team to view any attempt at sexual expression as inappropriate, even by residents without cognitive impairment.

The policies of PRC facilities regarding sexuality are sometimes informal and vague. For example, some assisted living directors instruct their staff to leave quietly when they observe sexual activity in residents' apartments. The responses of PRC community administrators and staff members to sex and intimacy among residents are context related. Directors and staff members rely on the wishes and directives of the family members and powers of attorney. Directors can be permissive if the family is supportive of the sexual relationship but can reinforce limit setting if the family is not accepting the behaviors. If an assisted living setting has a low census of patients, policies may become more flexible: Problematic sexual behavior may be discouraged, but impulsive patients will not be banned. Advance directives are generally used to give guidance concerning medical care, not sexual relationships. As directives become more widely used, however, they may also be used to specify other aspects of care the patient wishes to receive when incompetent, including sexual behavior. In fact, some individuals may feel so strongly about loyalty to a spouse or religious belief that they may include directives concerning future sexual relationships. Cultural training around the healthy expression of sexual behaviors in older adults may be needed to ensure that older adults who want to remain sexually active receive proper acceptance, education, understanding, and support, even after moving to a residential care community. Another barrier to satisfying sexual experiences for long-term care residents is the lack of knowledge and understanding among the residents themselves about later-life sexuality, and so residents may also benefit from education about later life sexuality. Physical aspects of life in a residential care environment are another common and significant barrier to sexual intimacy, such as living in a shared room with only a curtain between the beds for privacy.

Government policies and regulations also have a large impact on the culture of PRC communities. Eighteen states in the United States address the issue of married couples sharing rooms or having conjugal visits unless there is any medical contradiction (Federal regulation 483.10). Only four states allow room for sexual relationships between persons of the same gender, as well as persons of the opposite gender who are not married. Such recognition seems to be important to assure that NH residents have the same rights as other citizens in this regard.

Despite the difficulties associated with methods and measurement, evidence to date suggests that quality of life in residential care environments for older adults is largely determined by the presence of mental health problems and one's subjective sense of well-being. Research also suggests that both the individual's cultural identity, the sensitivity of the treating team towards this identity, and the culture of the residential care environment can significantly impact the mental health and perceived quality of life of older adults living in a residential care home. As a result, mental health care providers play an integral part in the care of older adults residing in NHS and other PRC environments.

LGBTQ+ Individuals and Long-Term Care

According to the National Gay and Lesbian Task Force, there are currently between 1.4 and 3.8 million lesbian, gay, bisexual, and transgender (LGBT) Americans over the age of 65. They remain particularly at risk of both homophobia, transphobia, and heterosexism, especially those who are residents of PRC facilities. Homophobia is defined as the unreasonable hatred, prejudice, and fear of LGB people. Transphobia is the term to describe hatred, prejudice, and fear of transgender individuals. Heterosexism and cisgender status function as default assumptions; it is a bias where the heterosexuality and cisgender of most people are presumed. Institutional heterosexism relies on the assumption of heterosexuality as the norm and results in the health needs of LGBT people being neglected. Since the U.S. Supreme Court legalized same-sex marriage in all fifty states, federal recognition of same-sex unions has allowed them to receive the financial and other benefits that come with opposite-sex marriages. As LGBT people age and become more dependent on federal benefits, such as Social Security, Medicare, and Medicaid (none of which were previously extended to same-sex partners), and become more enmeshed in health care systems that do not grant their partners the same rights as legal spouses, they experience the effects of a lifetime of inequality. These effects can follow a lesbian or gay person into the NH or other PRC setting. In fact, “invisible” is a recurring description of older LGBTQ+ individuals. Although “passing” is used as a survival tactic to protect against discrimination, the trauma for those who have lived their lives openly and honestly but find themselves going back into hiding when “they become ill, vulnerable and dependent on others” should not be minimized. Transgender persons are at particular risk of avoiding institutions and situations where they would be forced to rely on assistance from insensitive and transphobic providers, for example, in vulnerable situations such as bathing.

Many LGBT older adults are concerned about entering NHs due to fears that they will be forced into the closet to avoid being shunned by other NH residents for being openly lesbian, gay, bisexual, or transgender. Others fear mistreatment by administrative staff, care staff, and peers. Furthermore, for a long time, most retirement communities in the United States were faith-based; this meant they could discriminate without bounds because of their exemption from most nondiscrimination statutes. As an alternative to faith-based homes, several retirement communities catering to the LGBT community have now opened around the country. The most well-known of these is called RainbowVision, the first U.S. retirement village designed specifically for gay men and lesbians in Santa Fe, New Mexico. Similar communities now exist in several major American cities.

Scientific evidence regarding the clinical preparedness for dementia care providers to care optimally for members of the LGBTQ+ community remains scarce. A nationally representative mail-in survey of NH social service directors, however, revealed that LGBT cultural training is still not the norm. More than 75% of responders reported not receiving any training during the previous 5 years. Recent work by Nowaskie and Sewell (2020) found that dementia care providers endorse only moderate clinical preparedness and knowledge about LGBTQ+ health care, especially with transgender-specific care. However, they do endorse strong positive attitudes about the LGBTQ+

population, which may facilitate the recognition of the need for improved education and training. The current state of dementia care providers' LGBTQ+ cultural humility has, nonetheless, significant yet modifiable gaps. For those engaged as dementia caregivers, better education and more LGBTQ+ patient exposure (e.g., through patient panels and cultural humility training) are necessary in order to improve the care being provided to members of the LGBT community impacted by dementia illness. The LGBT Cultural Competency Project is an example of a collaborative initiative between multiple partnerships with a primary goal of providing education and awareness of the unique needs and concerns of LGBT older adults who are aging into long-term care communities. Some NHs, including those that do not specifically cater to the LGBT community, have endorsed the HEALE curriculum. The HEALE curriculum sets a standard for best practices in nursing management and for LGBT cultural competency in geriatric education. The nurses' HEALE training has significantly increased knowledge and, as a result, nurses' confidence in providing culturally sensitive care to older LGBT individuals. In 2015, OutCare was established as the nation's first comprehensive resource for LGBTQ+ health care. The support provided by OutCare includes provider and health care resource directories, including an online, nationwide directory of health care providers who identify as culturally competent in the care of the LGBTQ+ community, mentorship, and cultural competency training.

Lastly, family members may also be a barrier to sexual satisfaction for older adults living in a long-term care environment. Family members may have difficulty understanding and accepting that their older relative is having a sexual relationship with someone new. Sometimes their concerns may be legitimate (e.g., worry about STDs), and sometimes their concerns may have the appearance of being self-serving (e.g., concerns about the relationship's potential impact on inheritance).

Cognitive Capacity for Decisions Regarding Sexual Behaviors

Determining whether two patients, both of whom have dementia, should be allowed to share sexual intimacy may provoke feelings of anxiety or discomfort, but this has and will continue to be required of health care professionals. There are no universally accepted criteria or processes for determining the capacity to consent to sexual experiences. The assessment process should be initiated when there is evidence that the presence of a mental health or neurocognitive condition is impairing decision-making abilities, undertaken according to legal standards, and with the burden of proof on those alleging incapacity. In addition, upholding the basic human right of autonomy requires that cognitive capacity be viewed as specific to a situation or decision. Individuals who have legally established guardians or another form of proxy decision maker may still retain the capacity to consent to engage in intimate relationships.

Lichtenberg and Strzepek have published the only formal capacity assessment tool in this area, a semistructured interview tool based on the principles of informed consent. The tool requires that older adults have a basic level of cognitive functioning, and they chose a score of 14 or greater on the Mini-Mental State Examination (MMSE) as confirmation. Subsequently, an interviewer of the same sex as the older adults must assess each patient according to three criteria each patient's awareness of the relationship (e.g., Is the patient aware of who is initiating the sexual contact or is the person being misidentified as someone else such as a deceased spouse?); the potential risks (Can the patient describe how they will react if the relationship ends?); and the ability to avoid exploitation (Does the patient have the ability to state what level of sexual intimacy is desired and to say no to any uninvited or unwanted sexual act?). The results of both interviews are then reviewed by the members of the interprofessional team, and members vote to determine if the criteria have been met.

The Vancouver Coastal Health Authority (VCHA) recently created guidelines to assess intimate relationship decision making when one or both older adults may lack cognitive capacity. These guidelines are based on the premise that PRCs have an ethical and legal obligation to recognize, respect, and support the sexual lives of their residents and were created as a guide for PRCs to determine how best to achieve this obligation while also supporting staff and care providers. If an older adult is found to lack capacity, then a team that includes a family member should assemble to decide whether the intimate relationship will be permitted.

The three key recommendations from the VCHA guidelines are:

1. If a person has been appointed by the court or the client to make decisions specifically about sexual activity, then that person is the resident's decision maker regarding intimacy;
2. If no one with legal authority has been appointed by the court or the client to make decisions specifically about sexual experiences, then a group including a family member, assessing clinician, and PRC community representative should convene to determine whether sexual activity is in the best interest of the resident; and
3. The resident should be included in the decision-making process to the fullest extent possible.

Additional helpful guidance is available from The Hebrew Home at Riverdale in Riverdale, New York, which has posted online a copyright-protected document titled: “Policies and Procedures Concerning Sexual Expression at the Hebrew Home At Riverdale.” The intent of this document is to create a culture that recognizes and respects the importance of emotional intimacy in the lives of older adults. Such close human interactions are viewed as a normal and natural aspect of daily life.

The Hebrew Home at Riverdale policies specify that residents have the right to sexual activity, providing in each instance that the activity does not involve:

1. nonconsensual acts;
2. acts with minors;
3. acts between persons if there is a possibility of transmission of an STD; and
4. acts that impact negatively on the resident community as a whole through public display.

The guidelines for assessing consent to sexual activity contained in this document are available at the Hebrew Home’s website: www.hebrewhome.org.

Recommendations for PRC Communities

A federal appeals court recently upheld a finding by the CMS Appeals Board that NHs have an obligation to intervene when residents with dementia or Alzheimer disease engage in sexual activity and that the level J violation and the immediate jeopardy findings against the NH were correct.

The appeals court ruled that the NH had failed to properly intervene when three residents living with dementia engaged in sexual intimacy because the facility failed to;

1. talk to the residents about their feelings regarding these relationships,
2. document the residents’ capacity for consent, and
3. obtain medical assessments of how the residents’ cognitive deficits affected their capacity to consent.

In order to provide the highest quality experience possible for residents while also protecting the PRC community from allegations of medicolegal or ethical misconduct, it is important for every PRC to create written and readily accessible policies to ensure that a resident’s right to safe and healthy sexual expression is recognized, respected, supported and maintained while ensuring that the ability of residents living with cognitive problems has been properly confirmed and documented. Optimally created policies should include the requirement for ongoing and recurring staff training and case discussions. In addition, sexuality, including gender, preferred pronouns, sexual orientation, relationship status, and sexual needs, desires, and concerns, should be routinely assessed during the resident's initial visit with the community's nurse or physician and no less than annually thereafter. The community should create and maintain an atmosphere where the privacy of residents is respected. Staff should be required to document any relationship or intimacy concerns, including any proposed corrective actions.

The topic of sexual behavior in older adults often reveals ageistic beliefs and may sometimes illicit polarized viewpoints among patients, their family members and friends, and those who care for them. It is recommended that family members be informed of the community's policies regarding intimate relationships and sexual expressions. In the absence of capacity concerns, the resident's permission should be obtained for one or more family members to be included in any discussions surrounding resident intimacy issues. Like the need for staff training, the family members and friends of older adults may need education about the fact that, in spite of frailty or dementia, many older adults continue to have sexual needs that are unique and real.

Inappropriate sexual behavior among long-term care residents includes using sexually suggestive language, exposing oneself, masturbating, or openly participating in sexual activity with another resident. Whether in the long-term care environment or at home, the approach to sexually inappropriate behaviors is the same. If the problem sexual behaviors are new, a careful evaluation should take place. Underlying medical (e.g., bilateral medial temporal lobe injury), psychiatric (e.g., hypomania or mania), psychosocial (e.g., loneliness), and environmental (e.g., use of confusing terms of endearment by staff members such as "honey or "sweetheart") need to be investigated and addressed while concurrently implementing behavioral interventions. In order to avoid precipitating a "catastrophic reaction," it is important not to scold, argue or shame the person. Praising and encouraging acceptable behaviors and ignoring inappropriate behaviors may be helpful. Remaining calm helps to avoid rewarding unwanted behaviors. Sometimes inappropriate sexual behaviors may be triggered by underlying feelings of loneliness, isolation, or anxiety. Sometimes unmet sexual needs may trigger anxiety, frustration, or restlessness. Either way, providing appropriate forms of attention and physical affection proactively may prevent, reduce or eliminate problematic sexual behaviors.

Depending on the specific behavior, different solutions should be explored. For example, for the patient who repeatedly masturbates in public places, providing privacy as well as other things like a vibrator, personal lubricant, or condom may be helpful. If a patient disrobes in public, dressing the patient in clothing that makes disrobing difficult may help. For patients who expose themselves unnecessarily during personal care or become sexually aroused, it is important to set limits verbally, involve more than one caregiver, and, depending on the patient's sexual orientation, use caregivers whose gender is not the gender that stimulates the patient, and make the interactions as clinical as possible by wearing medical attire like a white coat and gloves, and performing these activities in an examination room rather than in the resident's room.

If treatment of all potential causes of inappropriate sexual behaviors coupled with behavioral interventions is unsuccessful, a medication trial may be needed. Table 55.6f–6 lists medications that may be helpful, including various forms of estrogen and SSRIs. Mood stabilizers may be useful even if hypersexuality is occurring in the context of a manic-like syndrome that does not fulfill all of the DSM 5 diagnostic criteria for mania. Antipsychotic medication may be helpful if the sexual behavior is driven by a delusion of mistaken identity in which the patient is attracted to a person misidentified as the patient's spouse or former sexual partner.

CASE EXAMPLE

The following case highlights a number of key issues associated with sexually inappropriate behaviors in patients living with dementia:

1. Although these behaviors can arise in patients living with a variety of types of dementia, they are most often seen in patients whose dementia has preferentially injured the frontal lobes resulting in problems with judgment and impulse modulation.
2. These behaviors are more likely to occur in later stages of dementia, resulting in sexually inappropriate behaviors occurring in the midst of other types of behavioral challenges.
3. Wide variation in both sex hormone levels and their impact on physical characteristics and behavior often limits their value in establishing the etiology of sexually inappropriate behaviors (this patient's measured serum testosterone level was actually low).
4. Medroxyprogesterone acetate (MPA) may help reduce but may not eliminate sexually inappropriate behaviors completely.

5. The use of MPA requires careful monitoring due to potential side effects that limit its usefulness and safety.

Table 55.6f–6

Medications for Hypersexuality in Patients With Dementia

Medication	Preparation	Suggested Dosage	Potential Side Effects
Hormonal Agents			
Antiandrogens			
Medroxyprogesterone acetate (Depo-Provera)	IM injection 150 mg/mL	150–300 mg once a week	Sedation, weight gain, fatigue, hot and cold flashes, depression
Estrogens			
(Premarin)	0.3-, 0.45-, 0.625-, 0.9-, 1.25-mg tablet	0.625–2.5 mg/day	Increased risk of cardiac death in men with prostate cancer, fluid retention, nausea, vomiting, gynecomastia
(Enjuvia)	0.3-, 0.45-, 0.625-, 1.25-mg tablets		
(Climera)	0.025-, 0.0375-, 0.05-, 0.06-, 0.075-, 0.10-mg/day transdermal patch	0.05–0.1 mg/day	
Gonadotrophin-releasing hormone analogs			
Leuprolide acetate (Eligard)	Injectable suspensions: 7.5-, 22.5-, 30-, 45-mg/month	7.5 mg every 4 weeks or every 8 weeks	Hot flashes, erectile dysfunction, irritation at injection site
(Lupron Depot)	3.75-, 7.5-mg/month		

	11.25 mg every 3 months		
Serotonergic Agents			
Serotonin reuptake inhibitors			
Citalopram	10 mg/5 mL oral solution, 10-, 20-, 40-mg tablets	20 mg/day	Headache, nausea, diarrhea, loss of appetite, insomnia
Paroxetine	10 mg/5 mL oral suspension 10-, 20-, 30-, 40-mg tablets 12.5-, 25-, 37.5-mg controlled-release tablets	20 mg daily	
Tricyclic antidepressants			
Clomipramine	25-, 50-, 75-mg tablets	150–200 mg/day	Orthostatic hypotension
Trazodone hydrochloride	50-, 100-, 150-, 300-mg tablets	100–500 mg/day	Sedation, hypotension, priapism
Antipsychotics			
Atypical antipsychotics			
Quetiapine	25-, 100-, 200-, 300-mg tablets	25 mg/day	Orthostasis, sedation, increased death due to cardiovascular events
Mood Stabilizers			
Carbamazepine	100-mg chewable	50–300 mg/day	Ataxia, dizziness, rash, sedation,

	tablet 200-mg tablet 100-, 200-, 400-mg extended-release tablets 100-, 200-, 300-mg capsule 100 mg/5 mL and 200 mg/10 mL suspensions		nausea, vomiting, hyponatremia, agranulocytosis, aplastic anemia, falls
Valproate	250-mg capsules 125-mg sprinkle capsule 250-, 500-mg extended-release tablets 250 mg/5 mL syrup 100 mg/mL injection	500–1,500 mg/day	Nausea, sedation, falls, tremor
Other Agents			
Pindolol	5-, 10-mg tablets	40 mg/day	Fatigue, hypotension
Cimetidine	200-, 300-, 400-, 800-mg tablets 300-mg/L oral solution	600–1,600 mg/day	Headaches, nausea, arthralgias, delirium

The primary care physician (PCP) of a 91-year-old widowed male, a retired engineer with vascular dementia residing in a dementia-specific residential unit, requested a geriatric psychiatry consultation to evaluate depression and inappropriate sexual behaviors. Because of the severity of the patient's cognitive deficits, most of the history was obtained from collateral sources.

Prior to the emergence of the patient's dementia symptoms, the patient had no psychiatric problems or psychiatric treatment of any kind. His cognitive deficits were first recognized several years prior to the time of the consultation request following a stroke which left the patient with residual ataxia and pseudobulbar affect. About 5 months after his stroke, behavioral symptoms, including wandering and elopement attempts, required the patient to move from an independent living apartment to a dementia-specific assisted living unit. Several months later, a geriatric psychiatry consultation was obtained after the patient slapped the hand of a male peer because this peer had touched the patient's food. Following this incident, the patient told a nurse, "if it happens again, I'll break his hand off." The consulting geriatric psychiatrist (CGP) found no undiagnosed medical factors that might have caused the problem behaviors and observed that the patient was struggling with frustration related to both a loss of control and problems with short-term memory. The CGP recommended discontinuation of quetiapine, given its questionable efficacy and unfavorable risk-benefit profile in patients living with dementia, and initiation of treatment with dextromethorphan/quinidine to target the patient's emotional lability and impulsive behaviors. In addition, the CGP suggested that, if needed, a trial of divalproex might also be reasonable.

Several months later, when sexually inappropriate behaviors emerged, the patient's PCP independently changed the patient's psychiatric medications to escitalopram 10 mg PO daily and divalproex delayed release 250 mg PO twice daily (which yielded a trough serum level of 30 mcg/dL) and encouraged the staff to use a variety of behavior-modifying techniques including avoidance of the use of ambiguous terms of endearment such as "honey" or "sweetie" and responding to inappropriate sexual behaviors with simple, firm verbal limit setting in order to avoid unintentionally reinforcing these behaviors with excessive positive (laughter) or negative (anger) affect. The patient, however, continued to have frequent and disruptive inappropriate sexual behaviors involving staff members, including verbalizing requests for sex, kissing and groping staff members, and masturbating in public.

The patient had no known past psychiatric history of hypomanic or manic episodes. The patient's medical problems included congestive heart failure, diabetes mellitus, type 2, hypothyroidism, benign prostatic hypertrophy, a history of sinus surgery, and a stroke 22 months previously. His medications were: acetaminophen 500 mg PO TID, bumetanide 0.5 mg PO BID, clopidogrel 75 mg PO daily, divalproex 250 mg PO BID, doxazosin mesylate 8 mg PO Q PM, escitalopram 10 mg PO QD, lansoprazole 30 mg PO QD, levothyroxine 100 mcg PO QD, metoprolol 12.5 mg PO BID, probiotic containing saccharomyces boulardii lyo 250 mg PO BID, sitagliptin 100 mg PO QD, and vitamin B12 1,000 mcg PO QD.

On examination, the patient's vital signs were within normal limits. The CGP examined the patient in the common area and observed that he was a casually dressed and groomed Caucasian man who was seated in a wheelchair and had no observed abnormal movements. He was distant, detached, and disinterested. He reported feeling "pretty normal," denied any current symptoms of depression or anxiety but was unable to describe his current medications other than saying that he was taking "a bunch" and that he was not aware of any side effects. His affect revealed frustration and irritability, his speech was normal, his process of thought was linear, and the content of his speech was noteworthy for concreteness, sarcastic responses to most questions, and the absence of clear answers to questions about suicidal ideation, homicidal ideation, delusions, or hallucinations. He was oriented to self and attentive but refused to cooperate with a standardized bedside cognitive screening test.

None of the patient's medications were felt likely to be triggering the patient's sexually inappropriate behaviors, and the results of a wide range of laboratory tests obtained by his PCP 1 week prior did not reveal a medical problem that would explain the patient's inappropriate sexual behaviors. These lab results included a random glucose of 123 mg/dL (high), creatinine 1.64 mg/dL (high), albumin 2.9 g/dL (low), total protein 5.3 g/dL (low), hemoglobin 8.6 g/dL (low), hematocrit 25.7% (low), TSH 3.09 µIU/dL, prolactin 14.3 µg/L, FSH 15.5 IU/L (high), LH 7.5 IU/L, progesterone 0.5 ng/mL, testosterone 187 ng/dL (low). After obtaining informed consent from the patient's family, a trial of medroxyprogesterone acetate (MPA) 150 mg IM Q monthly was initiated. No other changes in his medications were made in order to clearly assess the impact of the addition of MPA.

One month later, the CGP learned from staff members that the patient was no longer propositioning or attempting to kiss or grope members of the staff, although he was masturbating in his room occasionally. When interviewed, the patient reported feeling “in and out” and being “pushed around” by “people like you [the psychiatrist].” The CGP concluded that the patient was demonstrating a favorable response to the treatment with MPA and recommended continuation of MPA along with ongoing use of nonpharmacologic measures. Two months after initiation of treatment with MPA, the patient had not had any problematic sexual behaviors but had been spending most of the day sleeping, had become apathetic about meals, would eat if spoon-fed, and had lost 17 pounds over the preceding three months. When interviewed, the patient was oppositional and antagonistic, denied depression and anxiety but reported feeling angry due to “people like you.” When the CGP asked the patient to elaborate, he replied, “You had better get your ass out of here and never come back.”

Due to concerns about potential divalproex-associated nausea and weight loss, as well as the absence of any clear benefit, divalproex was discontinued, and a trial of mirtazapine 7.5 mg PO QHS was initiated to help with appetite stimulation. Three weeks after this visit, the patient was transferred from the dementia unit to a skilled nursing unit due to poor safety awareness associated with a number of falls. During his first week in the new location, he had two episodes of inappropriate sexual behavior: he grabbed a nursing assistant’s buttocks and asked her if she wanted to see his “banana,” and he was observed patting his groin in a hallway. The CGP examined the patient in a public area of the skilled nursing unit and observed that he was hostile, intermittently scratched himself, and patted his exposed abdomen. When asked about the recently concluded music-based activity, the patient replied, “I was there because they dragged me in there,” and he reported feelings of depression because “I have to put up with people like you.” No changes in the patient’s treatment were recommended.

When the patient was assessed approximately 4 months after the initiation of treatment with MPA, the CGP learned that no episodes of sexually appropriate behavior had occurred during the past month, but several episodes of verbal and physically aggressive behaviors had been observed, including smearing feces, resisting attempts to change his soiled clothing and attempting to spit and urinate on staff members. The patient informed the CGP that he saw a “spirit.” His blood glucose at the time of this interaction was determined to be 587 mg/dL. After consultation with the patient’s PCP, a decision was made to discontinue MPA due to concerns that the patient had developed delirium secondary to MPA-associated metabolic/endocrinologic dysregulation.

FURTHER READINGS

Ballard EL, Poer CM. *Sexuality and the Alzheimer’s Patient*. The Duke Family Support Program; 1993.

Basson R, Schultz WW. Sexual sequelae of general medical disorders. *Lancet*. 2007;369:409–424.

Beckman N, Waern M, Östling S, Sundh V, Skoog I. Determinants of sexual activity in four birth cohorts of Swedish 70-year-olds examined 1971–2001. *J Sex Med*. 2014;11(2):401–410.

Bitzer J, Platano G, Tschudin S, Alder J. Sexual counseling in elderly couples. *J Sex Med*. 2008;5(9):2027–2043.

Breen S, Carlson M, Clements G, Everett B, Young J. *Supporting Sexual Health and Intimacy in Care Group Homes in British Columbia*, Vancouver Coastal Health Authority: 2009.

Fredriksen-Goldsen KI, Kim HJ. Count me in: response to sexual orientation measures among older adults. *Res Aging*. 2015;37(5):464–480.

Fredricksen-Goldsen KI, Kim HJ, Emlet CA, et al. *The Aging and Health Report: Disparities and Resilience Among Lesbian, Gay, Bisexual, and Transgender Older Adults*. Institute for Multigenerational Health; 2011.

Fredriksen-Goldsen KI, Muraco A. Aging and sexual orientation; a 25-year review of the literature. *Res Aging*. 2010;32(3):372–413.

Galinsky AM, McClintock MK, Waite LJ. Sexuality and physical contact in National Social Life, Health, and Aging Project Wave 2. *J Gerontol B Psychol Sci Soc Sci*. 2014;69:S83.

Gates GJ. *Same-Sex and Different-Sex Couples in the American Community Survey: 2005–2011*. The Williams Institute at UCLA College of Law: 2012.

Gewirtz-Meydan A, Ayalon L. Shades of grey: exploring public opinion about later-life sexuality. *Can J Aging*. 2020;39(4):647–661. https://www.muse.jhu.edu/article/774970

Harawa NT, Leng M, Kim J, Cunningham WE. Racial/ethnic and gender differences among older adults in nonmonogamous partnerships, time spent single, and human immunodeficiency virus testing. *Sex Transm Dis*. 2011;38(12):1110–1117.

Jeste DV, Depp CA, Vahia IV. Successful cognitive and emotional aging. *World Psychiatry*. 2010;9(2):78–84.

Kaplan HS. *Disorders of Sexual Desire and Other New Concepts and Techniques in Sex Therapy*. Brunner/Hazel Publications; 1979.

Karraker A, DeLamater J, Schwartz CR. Sexual frequency decline from midlife to later life. *The J Gerontol*. 2011;66B(4):502–512. doi.org/10.1093/geronb/gbr058

Kinsey AC, Pomeroy WB, Martin CE. *Sexual Behavior in the Human Male*. W.B. Saunders Company; 1948.

Kinsey AC, Pomeroy WB, Martin CE, Gebhard PM. *Sexual Behavior in the Human Female*. W.B. Saunders Company; 1953.

Leavitt CE, Leonhardt ND, Busby DM. Different ways to get there: evidence of a variable female sexual response cycle. *J Sex Res*. 2019;56(7):899–912.

Lichtenberg PA, Strzepek DM. Assessments of institutionalized dementia patient's competencies to participate in intimate relationships. *Gerontologist*. 1990;30:117–120.

Lichtenberg PA. Sexuality and physical intimacy in long-term care. *Occup Ther Health Care*. 2014;28(1):42–50.

Lindau ST, Gavrilova N. Sex, health, and years of sexually active life gained due to good health: evidence from two US population based cross sectional surveys of ageing. *BMJ*. 2010;340:c810c810.

Mace NL, Rabins PV. *The 36-hour Day* 6th ed. Johns Hopkins Press; 2017.

Masters WH, Johnson VE. *Human Sexual Response*. Little, Brown and Company; 1966.

Mulhall JP, Trost LW, Brannigan RE, et al. Evaluation and management of testosterone deficiency: AUA guideline. *J Urol*. 2018;200:423–432.

Ruesink GB, Georgiadis JR. Brain imaging of human sexual response: recent developments and future directions. *Curr Sex Health Rep*. 2017;9(4):183–191.

The GeniUSS Group. *Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Responders in Population-Based Surveys*. The Williams Institute at UCLA College of Law: 2014.

Waite L, Cagney K, Dale W, et al. National Social Life, Health and Aging Project (NSHAP): Round 3, and COVID-19 Study [United States], 2015-2016, 2020-2021 (ICPSR 36873) v.7.

Ward BW, Dahlhamer JM, Galinsky AM, Joestl SS. Sexual orientation and health among U.S. adults: National Health Interview Survey, 2013. *Natl Health Stat Report*. 2014;77:1–10.

Wierman ME, Arlt W, Basson R, et al. Androgen therapy in women: a reappraisal: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2014;10:3489–3510.

Winn RI, Newton N. Sexuality in aging: a study of 106 cultures. *Arch Sex Behavior*. 1982;11:283–298.

55.6g HIV and Aging

Laura M. Campbell, M.S., María Marquine, Ph.D., and Raeanne C. Moore, Ph.D.

Once considered a death sentence, HIV has been transformed into a chronic disease as a result of highly active antiretroviral therapy (ART). The life expectancy of adults living with HIV that are receiving ART is now approaching that of the general population, and today more than half of persons with HIV in the United States are over age 50, with aging trends projected to continue. This remarkable success has brought with it a new set of challenges. As people with HIV are becoming older, HIV has become a chronic disease with multiple interacting causes of morbidity and mortality, generating significant public health interest in understanding the multiple factors that can impact aging with HIV.

Recognizing the issues and needs of the increasing older population of persons with HIV, a White House Conference on HIV and Aging was held in 2010, and the issue of HIV and Aging was again spotlighted at the 2015 White House Conference on Aging. In addition, the Office of AIDS Research of the National Institutes of Health (NIH) commissioned a working group to develop an outline of the current state of knowledge and areas of critical need for research in HIV and Aging. The findings and recommendations of the working group are summarized in a published report. While much research is needed to better understand health outcomes in older persons with HIV, the American Academy of HIV Medicine, in collaboration with the AIDS Community Research Initiative of America, developed a series of guidelines for clinicians managing older patients with HIV.

EPIDEMIOLOGY

In 2018, one in six (17%) of new HIV diagnoses in the United States were among older adults, and this rate has remained stable in recent years. Similar to younger age groups, members of racial/ethnic minority groups are at increased risk of HIV, with rates of infection in 2018 being 39.2 for blacks/African Americans, 16.4 for Hispanics/Latinos, 13.3 for persons of multiple races, 11.3 for Native Hawaiians/Pacific Islanders, 7.7 for American Indians/Alaska Natives, 4.8 for whites, and 4.7 for Asians. Older males are also at increased risk as compared to older females, with men making up the majority of new HIV diagnoses among people over the age of 50. As shown in Figure 55.6g–1, the most common forms of transmission vary by sex.

Older persons in the United States are more likely to be diagnosed with HIV later in the course of the disease compared to younger persons, which can result in worse disease burden and prognoses. For example, in 2015, over 50% of those diagnosed with HIV aged 55 and over had HIV for over 4.5 years prior to diagnosis, which is the longest diagnosis delay for any age group. However, in persons over age 55 with HIV, 71% received some HIV care, 57% were retained in care, and 64% were virally suppressed in 2018, which is better than the overall estimates for persons with HIV (65%, 50%, and 56% respectively).

Deaths of persons with diagnosed HIV may or may not be related to HIV. Of the estimated 15,484 deaths among people living with HIV in the United States in 2018, 11,425 (72%) were among people aged 50 and older. From 2014 to 2018, the death rate of persons with HIV aged 50 to 54 (rate = 11.9 in 2018) and 55 to 59 (2018 rate = 12.7) decreased over time, and the death rate of those aged 60 to 64 remained stable (2018 rate = 11.9), whereas the rate of death for persons aged 65 (2018 rate = 6.7) and older increased.

PREVENTION

Persons over the age of 50 appear to have several of the same risk factors for HIV as younger persons (e.g., sexual intercourse without protection, sharing drug equipment), which can be exacerbated by issues unique to older adults. Many older adults are sexually active, yet they may not consider themselves at high risk for HIV, may have less knowledge about HIV, and ultimately may be less likely to engage in HIV prevention. For example, older persons who are no longer concerned about pregnancy may be less likely to use condoms and/or other HIV prevention tools (e.g., pre-exposure prophylaxis [PrEP]). Additionally, age-related thinning and dryness of the vaginal tissue may place older women at an increased risk of acquiring HIV.

While there is much left to understand regarding the reasons driving later HIV diagnoses among older persons, some of the key reasons may be that older persons are less likely than younger persons to discuss their sexual habits or drug use with their doctors and that they might mistake HIV symptoms for those of normal aging. In turn, health care providers may also be less likely to ask about these issues and to test older persons for HIV. It also appears that HIV stigma might be of particular concern for older persons with HIV and may prevent them from seeking care and disclosing their HIV status.

Given these prevention challenges, the Center for Disease Control and its partners are implementing a High Impact prevention approach which includes new funding through the *Ending the HIV Epidemic* initiative, with the goal of reducing the number of new infections by 90% by 2030. This program includes a number of initiatives aimed at providing support and technical assistance to health departments and community-based organizations for implementing prevention programs, raising awareness about HIV, supporting HIV testing and optimizing linkage to care, increasing the percentage of those with viral suppression, and increasing PrEP prescriptions for people who may benefit from PrEP.

NEUROCOGNITIVE OUTCOMES OF HIV AND AGING

HIV and aging are independently associated with central nervous system (CNS) dysfunction. The co-occurrence of HIV and older age appears to have additive or synergistic negative effects on the CNS, with several studies observing elevated rates of neurocognitive impairment (NCI) and/or HIV-Associated Neurocognitive Disorders (HAND; see next paragraph) in older cohorts of persons with HIV. For example, in a meta-analysis from 2021 of 21 cross-sectional studies primarily from high-income countries, persons with HIV aged 50 and over were more likely to be cognitively impaired than HIV-negative controls (OR = 2.6). This meta-analysis found that older persons with HIV had lower scores in the domains of executive functioning, processing speed, verbal fluency, memory, and motor as compared to older HIV-negative adults. Several studies have also examined the issue of whether the combined effects of HIV and aging on NCI are evidence of premature (i.e., worse cognitive performance compared to HIV-negative peers), accentuated (i.e., higher rates of NCI compared to HIV-negative peers), or accelerated (i.e., a steeper decline in cognition compared to HIV-negative peers) aging. While the topic does remain controversial, a systematic review from 2020 of 37 studies found evidence for premature and accelerated but not accentuated aging.

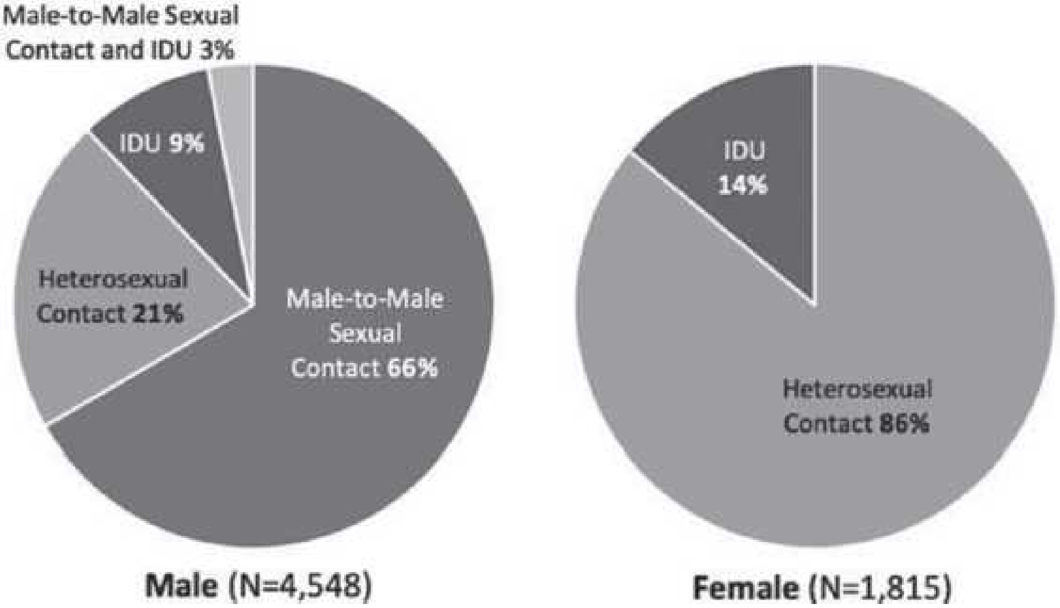

FIGURE 55.6g–1. New HIV diagnoses among adults aged 50 and older in the United States by sex* and transmission category in 2018. *Based on sex at birth and includes transgender people. IDU, injection drug use. *Due to rounding, percentages may not add up to 100%.* (From Center for Disease Control and Prevention: Diagnoses of HIV Infection in the United States and Dependent Areas, 2018 [Updated], 2020;31.)

Another factor to bear in mind for future studies in the area is that definitions of “neurocognitive impairment” can differ from study to study and often differ from the clinical definition of a “cognitive disorder.” For example, the Diagnostic and Statistical Manual of Mental Disorders-5th Edition (DSM-5) uses different criteria for defining a neurocognitive disorder than the current research criteria for HAND (i.e., Frascati criteria). For example, only problems on one neurocognitive domain are required by DSM-5 rather than two (per the Frascati criteria) to define impairment in “global” neurocognitive function implied by the term “cognitive disorder” (whether major or minor). Moreover, the exclusion criteria required for a diagnosis of neurocognitive disorder are limited to other psychiatric disorders per the DSM-5 but explicitly include CNS illnesses specific to the setting of immunodeficiency, as noted by both the original American Academy of Neurology (AAN) criteria and the current Frascati criteria. Moreover, different research criteria for cognitive impairment use different cut-points to define “impairment.” For example, Frascati criteria define impairment as performance at least 1 standard deviation below a demographically adjusted normative score in at least 2 domains, whereas some studies may employ stricter cut-offs. These differences in the research make it difficult to compare across studies and obscure the findings of future research.

While much is yet to be known regarding key predictors of NCI among older persons with HIV, a recent study investigated predictors of longitudinal neurocognitive decline in 436 adults with HIV (Age in years: Mean = 43.9, Standard Deviation = 8.4; Mean follow-up = 35 months). Results from multivariable models showed that key predictors of an earlier time to neurocognitive decline included Hispanic ethnicity, notable non-HIV comorbidity, being off ART, having low albumin and low hematocrit, having a lifetime methamphetamine use diagnosis, and increased depressive symptoms. These, along with other findings, suggest that a number of different factors (HIV and non-HIV specific) are probably important predictors of NCI in older persons with HIV.

As the population of persons with HIV continues to age, there is an increased need to understand the intersection of HIV and age-related neurodegenerative diseases such as Alzheimer’s disease. It is important to be able to differentiate between HAND and Alzheimer’s disease given the different prognoses: Alzheimer’s disease is associated with progressive cognitive and functional decline, whereas HAND is usually nonprogressive. Given that memory dysfunction is common in both HAND and Alzheimer’s disease, early signs of Alzheimer’s disease in persons with HIV may be missed or be attributed to HAND. This research is still underway, but it is an important topic of research that is necessary to be able to meet the health care needs of older persons with HIV.

MENTAL HEALTH OUTCOMES OF HIV AND AGING

Older persons living with HIV face unique emotional issues, including stigma, marginalization, and discrimination. In addition to emotional factors, these adults also often face unique social factors (e.g., lack of appropriate treatment facilities, poor access to health care, financial problems, disability) and unique family factors (e.g., HIV in a spouse, poor family support). These issues may make persons with HIV more vulnerable to psychiatric illness, decreased well-being, and poorer health-related quality of life. Depression is the most common psychiatric diagnosis among persons with HIV, with lifetime major depression disorder estimated between 25% and 45% among persons with HIV. However, despite depression being common in older persons with HIV, prevalence rates are not well studied, and depression is thought to be significantly underdiagnosed and undertreated in older persons with HIV. Additionally, the prevalence of other psychiatric conditions such as anxiety is thought to be increased in older persons with HIV as compared to HIV-negative peers, but again this is not well studied among older persons with HIV. Nevertheless, the cooccurrence of mental illness and HIV is associated with poorer health outcomes, such as increased rates of comorbid conditions and reductions in ART adherence. Efforts to integrate primary health care at Infectious Disease or General Medicine clinics with mental health treatment and to combat the societal stigma that is associated with mental illness and HIV are likely to improve the health and well-being of persons living and aging with this disease.

The role of mental health professionals in treating older persons living with HIV is threefold: (1) identify behavioral factors most saliently impacting the course of HIV and well-being outcomes to determine how and when to intervene; (2) work with the patient’s geriatrician and other primary care team members to determine the most appropriate course of treatment; and (3) implement treatment plan, closely monitor progress, and modify as needed.

Emotional Functioning

Depression management is a priority to achieve optimal health outcomes among those aging with HIV. Older persons with HIV do not show the same age-related declines in depression that are observed among older adults without HIV, making depression all the more relevant in this population. Additionally, research has indicated moderately high levels of overall distress among older persons with HIV, distress that is compounded by the addition of other chronic health problems. Moreover, depression in persons with HIV is associated with worse physical and mental health-related quality of life. For example, one recent study found that the correlation between depression and physical quality of life is the strongest in older persons with HIV with depression as compared to younger persons with HIV. These findings further emphasize how important it is to detect and manage depression in older persons with HIV in order to promote both mental and physical well-being.

Both antidepressant medications and psychotherapy have shown effectiveness in treating depression among those living and aging with HIV. Effectively treating depressive symptoms can lead to better everyday outcomes, including better medication adherence, sustained virologic suppression, reduced likelihood of HIV transmission, improved health-related quality of life, and longer survival. However, despite the existence of effective treatments for depression, underdiagnosed and untreated depression is highly prevalent. This is likely due to numerous factors complicating the clinical picture of aging with HIV, including individual patient characteristics, somatic symptom overlap between depression and chronic medical illnesses, and competing demands of primary care visits. Physiologic changes associated with medication metabolism that may cause symptoms that look like depression also complicate the clinical picture and care planning. Studies have found patient race/ethnicity to be a contributor to depression treatment, with depressed African American patients with HIV less likely to receive treatment than their non-African American counterparts. Additionally, investigators have found that patients with HIV diagnosed as severely depressed in a primary care setting are more likely to receive mental health treatment than patients meeting clinical criteria for mild or moderate depression. These investigators also found that if depression was not treated on the day of a patient’s primary care assessment, it was unlikely to be treated within the following 90-day period, regardless of depression severity.

It is also important to consider cohort effects when diagnosing and treating older adults living with HIV. The mental health of older adults who have been newly diagnosed with HIV will look different than the mental health of adults who have successfully survived into older age with HIV. In addition to the duration of HIV disease, individual patients may be experiencing mental health problems for the first time in their lives, or they may have a long-term history of mental health symptoms. These cohort effects impact how older adults perceive mental health problems and treatment and self-identify, experience, and describe their emotions and distress. For example, while not recognized by the DSM-5 at this time as a clinical syndrome, many older persons with HIV who are long-term survivors have what some of these patients refer to as AIDS Survival Syndrome. *AIDS Survival Syndrome* has been defined as a latent trauma response that emerges among those who survived the AIDS epidemic. These individuals represent a unique group of older persons with HIV who lived through a period of history in which they suffered the loss of many of their friends, partners, and community and were faced with the real threat of death to their own personhood. *AIDS Survival Syndrome* is thought to manifest as a cluster of symptoms including one or more of the following: depression, personality changes, flashes of anger, survivor's guilt, anxiety, emotional numbness, insomnia, social withdrawal and isolation, hopelessness, substance abuse, sexual risk-taking, and a lack of future orientation. Symptoms can range from mild to severe. While empirical evidence is currently lacking on AIDS Survival Syndrome, the authors believe this nomenclature can help clinicians understand the lived experience of this particular cohort of older persons with HIV and provide education and develop appropriate treatment plans for these individuals. Moreover, for those with symptoms of AIDS Survivor Syndrome, the term may normalize their feelings while providing a way to connect with others with similar lived experiences.

Substance Use

Prevalence rates of substance use are higher in older persons with HIV compared to their HIV-negative peers and range from approximately 20% to 50%. Substance use is a known primary risk factor for HIV medication nonadherence, particularly in those using hard drugs (i.e., cocaine or amphetamines) or people who inject drugs, or those with multiple-substance use such as combined cocaine and methamphetamine or combined cocaine and heroin. Excessive alcohol use, with and without other substance use, is also a predictor of poor adherence. A large study of substance use patterns among persons with HIV aged 50 and older found the poorest adherence and greatest self-perceived impairment from substance use among those using alcohol and cocaine/crack or those using multiple substances. Another large study examining excessive alcohol use among older persons with HIV with poor antiretroviral adherence found the following clinical characteristics to be related to the most severe unhealthy drinking patterns: African American race, less than full-time employment, alcohol treatment within the past year, and a positive depression screen. The pronounced role of substance use among older persons with HIV, together with the known negative impact of substance use on medication adherence and other functional outcomes, underscores the need for ongoing substance use screening and treatment in addition to increased prevention efforts on the part of both primary care clinicians and mental health professionals.

HIV and Severe Mental Illness

A meta-analysis published in 2016 found the prevalence of HIV among persons with severe mental illness (SMI) in the United States was 6.0% which is markedly higher than the general population (0.6%). Additionally, in African countries, the prevalence of HIV among persons with SMI was 19%. Despite these high rates, the literature on the co-occurrence of SMI and HIV is limited, particularly in older persons with HIV, but limited research does suggest that older persons with HIV and SMI, particularly psychosis, are at greater risk for mortality than with either illness alone. The association between SMI and HIV differs at an individual level and may be due to their simple coexistence or SMI preceding the acquisition of HIV. Alternatively, among those with psychosis, HIV may have a direct impact on the development of psychosis due to the effects of the virus or antiretroviral medications on the CNS.

Patients with schizophrenia, in particular, may have difficulty describing their symptoms and may be more susceptible to HIV-related distress. Clinical evaluation of older persons with HIV with schizophrenia requires a comprehensive assessment of the onset and course of symptoms. Additionally, caution should be taken when prescribing antipsychotic medications and take into account medications' side effects and interactions with antiretroviral medications.

Some evidence exists for the beneficial effects of antidepressant medication on improving antiretroviral adherence among patients with co-occurring HIV, bipolar disorder, and current mental health symptoms. For example, research examining antiretroviral adherence among patients with HIV and bipolar disorder found antiretroviral adherence was significantly better than adherence to psychiatric medications. These investigators also found a correlational relationship between antiretroviral adherence and psychiatric medication adherence, indicating increasing adherence to one medication may have an impact on adherence to the other. In other research, a disconnect between memory performance on standard neuropsychological tests, subjective memory complaints, and depressive symptoms in this population has been observed, such that memory complaints were not associated with memory performance but were associated with depressive symptoms. In turn, memory complaints were predictive of problems with everyday functioning (e.g., unemployment, dependence on activities of daily living [ADL]). Compared to persons with HIV without bipolar disorder, those with cooccurring bipolar disorder had less awareness of their actual memory abilities. Investigators concluded that both subjective memory complaints and objective memory ability should be considered when treatment planning in this subset of patients with HIV, as both appear to impact everyday functioning behaviors.

Successful Aging With HIV

To date, most studies have focused on factors associated with aging with HIV that are predictive of poor functioning, including physical illness burden, immune dysfunction, cognitive impairment, and psychiatric problems. Nonetheless, there is a growing body of literature that suggests the possibility of successful aging in individuals with HIV, and understanding HIV in this light may help identify individual factors related to a better quality of life and well-being. Studies have shown that there is a potential for successful cognitive and emotional aging among older adults with HIV, despite significant medical comorbidities and disease burden. Additionally, emerging research has shown that successful aging among persons with HIV appears to be buffered by protective psychological factors such as resilience and optimism.

There is also a rapidly growing body of healthy aging literature demonstrating the benefits of physical activity on physical and mental health and well-being outcomes. In older persons with HIV, studies have also demonstrated a positive impact of physical activity on improved physical function and lower rates of cognitive impairment. While physical activity is one of the most well-studied modifiable lifestyle factors, other modifiable lifestyle factors such as a cognitively active lifestyle, good nutrition, and sleep have also garnered attention as possible ways to improve or sustain physical and mental health among older persons with HIV.

COMORBIDITY AND MULTIMORBIDITY

As persons with HIV are living longer, they are increasingly experiencing fewer AIDS-defining conditions and a wider range of non-AIDS diseases. Some of the more common comorbidities in older persons with HIV include hypertension, diabetes mellitus, insulin resistance, cardiovascular disease, pulmonary hypertension, liver disease, kidney failure, non-AIDS cancers, osteoporosis, peripheral neuropathy, and “frailty,” which is characterized by weight loss, weakness, and increased risk of disability and death.

Although comorbidities in persons with HIV on ART are often defined as “non-AIDS,” they are associated with HIV disease (HIV-associated non-AIDS or HANA), HIV treatment, and/or behaviors, conditions, and demographics more common among those with HIV. Thus, while they are also common in HIV-negative older persons, these conditions may act differently among older persons with HIV. Several mechanisms have been suggested, including microbial translocation, chronic inflammation, oxidative stress, and immune senescence. Yet, much work remains to be done to understand the mechanisms underlying comorbidity in older persons with HIV.

Guidelines are available for the management of common comorbid conditions in older persons with HIV. It is important to bear this in mind. However, that multimorbidity is common among older persons with HIV and is more than simple comorbidity. Multimorbidity is conceptualized as the presence of several health problems that have the propensity to synergize to worsen a person’s functional capacity and increase mortality more than might otherwise seem apparent. When considering diagnosis and treatment in multimorbidity, the sum is greater than the parts. This multimorbidity can result in overlapping injury to multiple organ systems transforming HIV into a complex chronic disease. This has prompted the need for better methods to track the impact of comorbidities and multiple organ system dysfunction on the health of people with HIV.

The VACS Index was developed in this context as a tool to monitor the impact of dysfunction involving multiple organs commonly affected in HIV, using objective measures that are collected routinely in HIV clinics. It was initially published in 2013 and then was recently updated in 2019 (i.e., the VACS Index 2.0) to include additional clinical markers that improved discrimination of mortality. The VACS Index 1.0 includes age, traditional HIV biomarkers (i.e., HIV-1 plasma RNA and current CD4 count), and non-HIV biomarkers (i.e., indicators of renal and liver function, anemia, and Hepatitis C coinfection), and the VACS 2.0 added in albumin, white blood cell count, and body mass index (BMI). The VACS Index 1.0 and 2.0 have been validated for estimating the risk for mortality among people with HIV, including in diverse cohorts. The VACS Index 1.0 has also been associated with other outcomes, such as NCI, but the strength of this association is weak among Hispanics/Latinos, and more research is still needed to further examine associations with the newer VACS Index 2.0.

IMPACT ON EVERYDAY FUNCTIONING

Real-world functioning impairment is a known sequela of HIV disease, and functional disability expenditures have made HIV among the most expensive chronic illnesses. As such, the NIH Office of AIDS Research and the U.S. Senate Special Committee on Aging has emphasized the significant public health priority for research improving functioning in those aging with HIV. Real-world functioning in aging and HIV is complex and multidetermined. Age and HIV are associated with increased risk for NCI, which, along with numerous risk factors (i.e., certain demographics, depression, anxiety, memory complaints, social isolation), result in older persons with HIV having a disproportionate disruption in real-world daily functioning. *Real-world functioning* is defined as what a person actually does in their daily lives. In particular, decreases in job-related activities, self-reported problems with basic (BADLs) and instrumental activities of daily living (IADLs), worse clinician-reported functional outcomes, ART nonadherence, and driving accidents have been found in older persons with HIV relative to their younger counterparts. These age-related impairments have been found irrespective of neurocognitive status. In terms of *functional capacity*, defined as a person’s abilities that are essential to function independently in the community (i.e., what they are capable of doing, not what they actually do), a synergistic effect of HIV and neurocognitive status has been observed, such that older, cognitively impaired adults with HIV perform worse on tasks of medication management and finance management than younger and unimpaired groups.

While older age has been found to exacerbate HIV-associated disability in real-world functioning, not all older persons with HIV evidence functional impairment, which raises important questions about individual clinical characteristics impacting poor functional outcomes. The most well-studied predictor of disruptions in real-world functioning is NCI. The association between neurocognitive and real-world functional impairment is highly complex, and even subtle neurocognitive deficits, such as shallow encoding and forgetting, have been found to be associated with IADL dependence in this population. Factors believed to influence this relationship include the match between the individual's profile of neurocognitive strengths and weaknesses and the specific demands of the real-world activity (e.g., its complexity and cognitive and functional requirements). Motivation, awareness of cognitive and functional deficits, cognitive reserve (i.e., effects of enrichment of cognitive resources through educational and vocational experiences together with a person's innate intellectual ability), and the use of compensatory strategies also play a role in determining the nature of this relationship. There is also a multitude of additional static and dynamic (i.e., changes in one area may impact other areas of functioning) biologic, psychological, and social factors influencing the impact of neurocognitive ability on real-world daily functioning. For example, having sufficient home and community resources, along with appropriately adjusting daily routines to account for physical and cognitive symptoms, have been associated with successful real-world daily functioning among adults with HIV. To begin to examine these complexities, some have started to utilize technology such as ecologic momentary cognitive testing in order to examine cognition and everyday functioning in real-time in participants' natural environments. For example, one study found that doing a cognitively stimulating activity (e.g., reading a book or working) is associated with better performance on a smartphone-based executive functioning task in older persons with HIV. Moreover, this study demonstrates the utility of technology in helping to understand the complexities of everyday functioning.

Cognitive reserve may be especially important to consider within the context of aging with HIV. Research has found lower cognitive reserve is associated with risk for HIV-associated NCI and greater declines in IADLs. Conversely, higher cognitive reserve has been found to be the sole significant clinical co-factor of successful cognitive aging among older persons with HIV, whereas, among younger persons with HIV, several co-factors affecting successful cognitive aging have emerged (i.e., higher cognitive reserve, Caucasian race/ethnicity, shorter duration of HIV, and lower prevalence of lifetime alcohol dependence). Thus, the evidence suggests that among older persons with HIV, lower cognitive reserve may interfere with the adaptive ability to initiate the use of compensatory strategies when engaging in behaviors in their daily lives. Apathy and depression are also two known reliable predictors of adverse functional outcomes among older persons with HIV. A link between more severe depressive symptoms and overreporting neurocognitive and functional impairments has been identified. Furthermore, a link has also been found between less severe depressive symptoms and underreporting impairments. Associations between higher self-reported apathy with greater severity of IADL declines and more numerous cognitive complaints have been observed, suggesting that apathy may make a unique contribution to important everyday functioning outcomes. Fortunately, these risk factors can easily be screened for and are amenable to change. One rigorously designed study found no relationship between antidepressant use and better daily functioning among older persons with HIV with a current or lifetime major depressive disorder diagnosis. While, to the authors' knowledge, the relationship between antidepressant use and daily functioning among older persons with HIV has only been explored in one study, it may be important for clinicians to consider treatment incorporating a combination of pharmacotherapy and psychotherapy.

When assessing and treating functional impairments in this population, the patient's medical history, medical comorbidity burden, HIV disease burden, culture, and cohort effects should also be carefully considered. In the current era of effective ART, older adults recently diagnosed with HIV may not have the same downstream effect of HIV disease-specific biomarkers on real-world daily functioning that can be observed in older survivors; thus, disability risks may vary based on the duration of the disease. In terms of the cultural impact of age and HIV on real-world functioning, assessing the sociocultural context in which HIV occurs is important to determine factors impacting functional outcomes. These sociocultural factors may include (but are not limited to) access to health care and antiretroviral treatment, the value of employment versus disability, HIV-associated stigma, and behavioral expressions of distress.

THE SPECIAL CASE OF THE VETERAN POPULATION

The Veterans Health Administration (VHA) is the largest single provider of HIV care in the United States, with 1 out of every 250 Veterans in care at VHA living with HIV. The VHA faces challenges in meeting the needs of the growing population of older Veterans with HIV. Up to 63% of veterans with HIV have at least one mental health diagnosis, and many Veterans in need of mental health treatment are not receiving treatment. Primary barriers to treatment include identifying targets for intervention, scheduling and availability of a mental health provider, time and travel to and from a VA facility, and privacy concerns.

Older Veterans with HIV are at increased risk for depression as well as loneliness and social isolation. Depression affects over half of Veterans with HIV, making it the most common comorbidity in this population. The Veterans Aging Cohort Study found social isolation is associated with a greater risk of hospitalization and death in older Veterans with and without HIV and reported that this effect may be greater in Veterans with HIV. In a retrospective cohort study using data from the national VHA HIV Clinical Case Registry, researchers examined the impact of mental illness among Veterans with HIV on progression to AIDS-defining illness and death and found substance use to be predictive of AIDS-defining illness, whereas schizophrenia, bipolar disorder, and substance use were predictive of all-cause mortality. In summary, data from these studies indicate the need for increased knowledge of the mental health needs of older Veterans with HIV and increased access to care, specifically care targeting the aforementioned risk factors for poor outcomes.

TREATMENT AND MANAGEMENT

With longer life expectancy and improved health outcomes, it is no longer assumed that adults aging with HIV will need to collect disability payments as income. But instead, many may seek to continue working until retirement age. Given the public health burden of functional impairments associated with HIV, it is imperative to understand the risk and protective factors impacting real-world daily functioning. Additionally, it is equally important to understand these risk and protective factors in the context of sociodemographic characteristics, including sex, age, cohort effect, racial/ethnic minority status, and socioeconomic status. A targeted understanding of the most potent factors impacting real-world daily functioning at an individual level will aid clinicians in determining how and when to intervene in order to improve functioning and health-related quality of life, as well as identify functional rehabilitation interventions that can be incorporated into clinical care.

Despite the greater availability of knowledge regarding the neurocognitive consequences of HIV disease compared to the psychiatric effects, there remains a paucity of literature on treating HAND and improving overall health outcomes. Evidence has consistently shown that patients with histories of significant immunocompromise (e.g., low nadir CD4) have an increased risk of poor neurocognitive outcomes. Therefore, current practice standards recommend prescribing ART at the time of HIV diagnosis as a prevention strategy against the neuroinflammatory cascades associated with poor immune functioning. While it is accepted that people with HAND should be on ART, it is also known that certain antiretrovirals are particularly neurotoxic and, if possible, it may be best to opt for a less neurotoxic antiretroviral if it is thought this may be contributing to HAND. Unfortunately, clinical trials using neuroenhancing drugs (e.g., psychostimulants, memantine) for the treatment of HAND have been largely unsuccessful. However, treatment of other comorbid conditions (e.g., HCV infection, depression) may be beneficial. Overall, there is currently no first-line clinical pharmacotherapy for HAND among adults aging with HIV. While cognitive and behavioral rehabilitation strategies are increasingly being explored, published studies on cognitive neurorehabilitation approaches for HAND are scarce. Solely using pharmacotherapy for HAND treatment is contraindicated, as there is a strong correlation between neurocognitive deficits and medication nonadherence among persons living with HIV. Psychotherapeutic interventions incorporating cognitive and behavioral components with adherence (i.e., cognitive behavioral therapy for adherence and depression [CBT-AD]) are potentially efficacious supplements or alternatives to pharmacotherapy. Additionally, there is some promising early evidence of the potential efficacy of restorative cognitive rehabilitation approaches.

Lifestyle modification is another potential functional rehabilitation treatment approach. Evidence-based behavioral treatments exist for a vast array of lifestyle factors, including physical activity, stress reduction, mental activity, social engagement, sleep hygiene, and substance use reduction. For example, although the research is somewhat mixed, a number of studies have found that mindfulness-based interventions are efficacious in reducing depression and anxiety symptoms and improving quality of life. While most studies have not examined older persons with HIV specifically, a meta-analysis found that older age actually conferred greater benefit from mindfulness-based interventions and thus may be particularly beneficial for older persons with HIV. Although many of the aforementioned lifestyle factors have yet to be tested (via controlled trials), specifically among older adults with HIV, many of these factors protect and promote cognitive reserve, increase brain-derived neurotrophic growth factors, and reduce neuroinflammation and oxidative stress, which could have a direct benefit on cognitive and real-world everyday functioning.

FUTURE DIRECTIONS

As long-term survival among persons living with HIV continues to improve, the need for identifying personalized targets for improving outcomes in this population is clear. Comorbid health problems are the norm among persons living with HIV who are 50 years of age and older. Continued efforts to integrate primary health care, geriatricians, HIV providers, and mental health treatment are needed to address the needs of this population. Although recommended, many primary health care providers and HIV providers are unlikely to assess for HAND or mood symptoms, which have a direct negative impact on treatment. Comprehensive assessment and integrated care planning in this population, regardless of age, is necessary. It is increasingly important for the patients, their care providers, and the health care system to expand the focus of clinical research and practice beyond the alleviation of symptoms to consider issues of real-world functioning and health-related quality of life.

FURTHER READINGS

American Academy of HIV Medicine: Updated—The HIV and Aging Consensus Project: Recommended Treatment Strategies for Clinicians Managing Older Patient with HIV. 2014

American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders. 5th Ed. American Psychiatric Publishing; 2013

Antinori A, Arendt G, Becker JT, et al. Updated research nosology for HIV-associated neurocognitive disorders. *Neurology*. 2007;69:1789–1799.

Aung HL, Aghvinian M, Gouse H, et al. Is there any evidence of premature, accentuated and accelerated aging effects on neurocognition in people living with HIV? A Systematic Review. *AIDS Behav*. 2021;25(3):917–960.

Balderson BH, Grothaus L, Harrison RG, McCoy K, Mahoney C, Catz S. Chronic illness burden and quality of life in an aging HIV population. *AIDS Care*. 2013;25:451–458.

Blackstone K, Tobin A, Posada C, et al. HIV-infected persons with bipolar disorder are less aware of memory deficits than HIV-infected persons without bipolar disorder. *J Clin Exp Neuropsychol*. 2012;34:773–781.

Campbell LM, Paolillo EW, Heaton A, et al. Daily activities related to mobile cognitive performance in middle-aged and older adults: an ecological momentary cognitive assessment study. *JMIR Mhealth Uhealth*. 2020;8:e19579.

Center for Disease Control and Prevention: Diagnoses of HIV Infection in the United States and Dependent Areas, 2018 (Updated), 2020.

Deng L, Zhang X, Gao Y, et al. Association of HIV Infection and Cognitive Impairment in Older Adults: A Meta-analysis. *Ageing Res Rev*. 2021;68:101310.

Eggers C, Arendt G, Hahn K, et al. HIV-1-associated neurocognitive disorder: Epidemiology, pathogenesis, diagnosis, and treatment. *J Neurol*. 2017;264:1715–1727.

Fazeli PL, Doyle KL, Scott JC, et al. Shallow encoding and forgetting are associated with dependence in instrumental activities of daily living among older adults living with HIV infection. *Arch Clin Neuropsychol*. 2014;29:278–288.

Greyson R, Horwitz LI, Covinsky KE, Gordon K, Ohl ME, Justice AC. Does social isolation predict hospitalization and mortality among HIV+ and uninfected older veterans? *J Am Geriatr Soc*. 2013;61:1456–1463.

Heaton RK, Franklin DR Jr., Deutsch R, et al. Neurocognitive Change in the Era of HIV Combination Antiretroviral Therapy: The Longitudinal CHARTER Study. *Clin Infect Dis*. 2015;60(3):473–480.

High KP, Brennan-Ing M, Clifford DB, et al. HIV and aging: state of knowledge and areas of critical need for research. A report to the NIH Office of AIDS Research by the HIV and Aging Working Group. *J Acquir Immune Defic Syndr*. 2012;60(Suppl 1):S1–18.

Hooshyar D, Goulet J, Chwastiak L, et al. Time to depression treatment in primary care among HIV-infected and uninfected veterans. *J Gen Intern Med*. 2010;25:656–662.

Hughes E, Bassi S, Gilbody S, Bland M, Martin F. Prevalence of HIV, hepatitis B, and hepatitis C in people with severe mental illness: a systematic review and meta-analysis. *Lancet Psychiatry*. 2016;3:40–48.

Marquine MJ, Sakamoto M, Dufour C, et al. The impact of ethnicity/race on the association between the Veterans Aging Cohort Study (VACS) Index and neurocognitive function among HIV-infected persons. *J Neurovirol*. 2016;22:442–454.

Marquine MJ, Montoya JL, Umlauf A, et al. The Veterans Aging Cohort Study (VACS) index and neurocognitive change: a longitudinal study. *Clin Infect Dis*. 2016;63:694–

702.

Montoya JL, Jankowski CM, O’Brien KK, et al. Evidence-informed practical recommendations for increasing physical activity among persons living with HIV. *AIDS*. 2019;33:931–939.

Moore DJ, Fazeli PL, Moore RC, et al. Positive psychological factors are linked to successful cognitive aging among older persons living with HIV/AIDS. *AIDS Behav*. 2018;22:1551–1561.

Moore DJ, Posada C, Parikh M, et al. HIV-infected individuals with co-occurring bipolar disorder evidence poor antiretroviral and psychiatric medication adherence. *AIDS Behav*. 2012;16:2257–2266.

Moore RC, Fazeli PL, Jeste DV, et al. Successful Cognitive Aging and Health-Related Quality of Life in Younger and Older Adults Infected with HIV. *AIDS Behav*. 2014;18:1186–1197.

Morgan EE, Iudicello JE, Weber E, et al. Synergistic effects of HIV infection and older age on daily functioning. *J Acquir Immune Defic Syndr*. 2012;61:341–348.

Nurutdinova D, Chrusciel T, Zeringue A, et al. Mental health disorders and the risk of AIDS-defining illness and death in HIV-infected veterans. *AIDS*. 2012;26:229–234.

Parsons JT, Starks TJ, Millar BM, Boonrai K, Marcotte D. Patterns of substance use among HIV-positive adults over 50: Implications for treatment and medication adherence. *Drug Alcohol Depend*. 2014;139:33–40.

Rooney AS, Moore RC, Paolillo EW, et al. Depression and aging with HIV: associations with health-related quality of life and positive psychological factors. *J Affect Disord*. 2019;251:1–7.

Rubin LH, Sundermann EE, Moore DJ. The current understanding of overlap between characteristics of HIV-associated neurocognitive disorders and Alzheimer’s disease. *J Neurovirol*. 2019;25(5):661–672.

Scott-Sheldon LAJ, Balletto BL, Donahue ML, et al. Mindfulness-based interventions for adults living with HIV/AIDS: a systematic review and meta-analysis. *AIDS Behav*. 2019;23:60–75.

Sueoka K, Goulet JL, Fiellin DA, et al. Depression symptoms and treatment among HIV infected and uninfected veterans. *AIDS Behav*. 2010;14(2):272–279.

Tate JP, Sterne JAC, Justice AC, Veterans Aging Cohort Study (VACS) and the Antiretroviral Therapy Cohort Collaboration (ART-CC). Albumin, white blood cell count, and body mass index improve discrimination of mortality in HIV-positive individuals. *AIDS*. 2019;33:903–912.

Thames AD, Becker BW, Marcotte TD, et al. Depression, cognition, and self-appraisal of functional abilities in HIV: An examination of subjective appraisal versus objective performance. *Clin Neuropsychol*. 2011;25:224–243.

Thames AD, Kim MS, Becker BW, et al. Medication and finance management among HIV-infected adults: the impact of age and cognition. *J Clin Exp Neuropsychol*. 2011;33:200–209.

United States Senate Special Committee on Aging: Older Americans: The Changing face of HIV/AIDS in America, in One Hundred Thirteenth Congress, first session, September 18, 2013. Washington, D.C, 2013.

Watkins CC, Treisman GJ. Neuropsychiatric complications of aging with HIV. *J Neurovirol*. 2012;18:277–290.

Williams EC, Bradley KA, Balderson BH, et al. Alcohol and associated characteristics among older persons living with human immunodeficiency virus on antiretroviral therapy. *Subst Abus*. 2014;35:245–253.

55.6h Technology for Seniors

AVERY QUYNH, B.S., VANESSA SCOTT, B.S., COLIN A. DEPP, PH.D., AND IPSIT V. VAHIA, M.D.

With the population of older adults expected to reach two billion worldwide by 2050, there will be a serious shortage of skilled care providers to manage the needs of this population. To cope with the increasing demands for health and social care, there is a great need to reorganize care processes and leverage different technologic solutions. Simultaneously, with the rapid maturing of the "aging in place" movement, there is an unprecedented incentive to optimize an array of technologies to help adults maintain independence in their own homes and avoid or forestall the transition to institutionalization. Over the past decade, there has been a revolution in how technology has both evolved and impacted the global population at large. Canadian social scientists Wellman and Quan-Haase used the term "hyperconnectivity" to describe multimodal human-to-human and human-to-machine interaction using a variety of platforms, including face-to-face, email, mobile phone technology, and the Internet. The implication of this term is that most persons will communicate via a multitude of channels facilitated by technology and that human interaction with multiple forms of technology will grow substantially.

In response to this societal shift, there has been an explosion in technologies over the past decade, facilitated not just by seamless communication but also by advances in computing and sophisticated sensors that can increasingly provide continuous objective data on a number of parameters. Although older adults have traditionally been considered late adopters of emerging technology, there is a growing consensus that this demographic group, with its needs for increased care and monitoring, especially among the oldest old with multiple medical needs and reduction in independent functioning, represents a significant opportunity for technology-based innovation.

COMPARATIVE NOSOLOGY

Defining Technology in Aging

The term *technology* is derived from the Greek term "*tékhne-*," which means art or skill. It is defined by the Collins English Dictionary as the "total knowledge and skills available to any human society for art, skills, etc." This is a very broad term, subject to multiple interpretations, and inclusive of technology in all its forms and considerations, from the use of flint to make fire to the latest developments in robotics. More recently, within the context of aging, researchers have favored the use of the term "gerontechnology," which has been described as "an interdisciplinary field of scientific application of technologic applications that are directed toward the aspiration and opportunities of the older person." These are broad terms, and as may be expected, they cover a very broad range of technologies. Table 55.6h–1 presents an overview and broad classification of types of technology based on application. This chapter will focus primarily on Information and Communication Technology and Assistive Technology. These terms encompass a range of approaches, such as virtual reality (VR), robotics, and tools that incorporate sophisticated data analytics methods such as machine learning. Other technologies, including Smart Home Technology, are also discussed.

Table 55.6h–1.
Classification of Technology for Aging

Type	Example	Target Population	Role
Medical technology	Pacemakers, implanted medication infusion systems, vagal nerve stimulators	Focused: Older adults with specific medical issues	Targeted medical role
Assistive Technology	Telemedicine systems, sensors (e.g., accelerometers), monitors (e.g., glucometers)	Narrow: Older adults with medical conditions that require monitoring	Maintenance of physical health and well-being
Information and Communication Technology	Wi-Fi, smartphones, tablet devices (e.g., iPads)	Broad: Older adults with need and ability to connect to the internet or use cellular technology for communication	Staying in touch with friends or family, more integrated social functioning, access to information
Smart Home technology	Personal emergency response systems (PERS), environmental sensors	Broad: Older adults with cognitive or functional impairments, invested in aging in place	Facilitating remote monitoring of older adults living independently, minimizing risk of in-home adverse events

There has been a growing cross-disciplinary effort to develop and apply information and communication technologies (ICTs) for acquiring, processing, presenting, and managing health-related data in older adults. The goal of these technologies is to improve health, quality of life, and social connectivity for less impaired older adults and assist in clinical care and monitoring of more impaired older adults. Although older adults do not use ICTs as commonly as younger individuals, they do use devices such as smartphones in comparable ways, that is, to communicate, access information (health-related and otherwise), entertainment, and financial management. Emerging data suggest that certain factors, such as the perceived advantage of technology and design issues, contribute to the lower rates of technology use in older adults. These issues may serve to guide the modification and development of ICTs tailored to engage and benefit older adults.

Assistive technologies have a more narrow scope and focused application since they are usually developed to target specific medical conditions or functional deficits. It is important to note that a specific technology may be classified in more than one category. For example, a smartphone is usually considered a platform for an ICT. However, it may be used as a form of assistive technology, if a specific app has a targeted role in monitoring. In addition, there is a new generation of cellphone accessories that aim to leverage the computing capabilities of smartphones to transform them for use as medical devices. For example, a 2016 FDA-approved device, the Steth IO, is a digital stethoscope linked to an iPhone application that can measure heart rates and listen to lung sounds. Portable and more widely accessible health devices like the AliveCore portable EKG scanner are also technologic advances in mobile medicine. Such devices have the potential to facilitate remote consultations and rapid diagnoses.

Technology and Aging in Place

An important goal of technology for older adults is to compensate for age-associated losses in physical and cognitive functioning. There is growing evidence to support the use of assistive technologies as a way to allow older adults to “age in place.” Aging in place is defined by the Centers for Disease Control and Prevention (CDC) as “the ability to live in one’s own home and community safely, independently, and comfortably, regardless of age, income, or ability level.” Social networking, remote monitoring, and assistive devices have shown promise in enhancing opportunities for aging in place. Since a significant proportion of increased health care costs associated with aging is related to the need for intensive monitoring and institutionalized care, promoting aging in place is an important approach to preserving independence and reducing costs. Both ICTs and assistive technologies have a key role in promoting aging in place. The American Association for Retired Persons (AARP), which has taken on a prominent role in advocacy for aging in place, describes case scenarios on its website of how technology, including pillbox sensors, refrigerator sensors, and personal emergency response systems (PERS) can help promote aging in place. Applications complement such technologies by offering services like organizing transportation, tracking medication, and meal assistance for older adults,

SPECIFIC APPLICATIONS OF TECHNOLOGY

Telehealth

Medical providers are increasingly using telecommunication devices in the diagnosis and care of patients in remote locations. The COVID-19 pandemic has been primarily responsible for the scaling up of telehealth to the extent that it is now a default modality for providing care in many settings, particularly in higher-income countries. These technologic modalities include video, voice, robotic, and remote-access technology to facilitate interactions between providers and patients. Telehealth approaches to observe patients and provide intervention services have direct relevance to psychiatry and geropsychiatry. The use of telemedicine demonstrates its ability to be cost-effective and improve the ease with which patients may be observed and assessed.

A large and rapidly growing body of literature supports the reliability and feasibility of using videoconferencing as an assessment modality for patients with psychiatric illnesses. Furthermore, this approach is a useful method of delivering interventions in this population and is associated with good functional and psychiatric outcomes. Studies conducted since the COVID-19 pandemic also validate the significant pragmatic benefits reported by recipients, including quicker access to providers and the convenience of not needing transportation to get care.

Building on this explosion of telecare, academic institutions, government agencies, and several privately operating companies are developing concepts, systems, and devices in a geropsychiatry context. Telehealth approaches mitigate the unique challenges faced by this population. For example, distance from providers may restrict regular health care visits for older patients with limited mobility. Services such as video consultation and remote monitoring devices facilitate the transfer of information between the patient and doctor and monitor the appropriateness of clinical interventions, thus improving the quality and access to health care. A majority of interventional telemedical studies in the geriatric population have focused on monitoring older adults. This includes routine personal monitoring accompanied by supportive contact between the health care provider and the patient and/or measurements of key biomarkers using technologic modalities. Overall, such interventions appear to have a positive impact on important health outcomes such as adherence and behavioral change.

Older patients tend to have physical and cognitive limitations and thus have distinctive needs that must be addressed by telehealth protocols. It is particularly promising that telehealth approaches appear to be feasible and useful in a wide variety of geriatric subgroups. For example, Loh and colleagues developed and examined a videoconferencing protocol for diagnosing Alzheimer disease and found that there was good diagnostic agreement between the telemedical modality and the face-to-face assessment. Telepsychiatry has also been investigated in nursing home populations. Not only do telepsychiatry visits have a positive impact on the quality of patient care, but they also appear to be associated with significant financial savings (approximately $30,000 saved over the course of 278 telepsychiatry visits). During the COVID-19 pandemic, it became increasingly difficult for older adults to receive care due to physical and social distancing restrictions. However, the rise of telemedicine allowed patients to get the care they needed without having to leave their homes. Studies performed using telemedicine with adults with dementia showed the feasibility and overall acceptance of these methods of health care. Taken together, these data support the potential benefit of telemedical applications in aging cohorts and highlight the need for greater empirical attention to the development and use of these novel approaches.

The application of telehealth concepts to geropsychiatry requires attention to the clinical benefits of these approaches, as well as their feasibility, acceptability, and cost–benefit ratio. As these systems spread in use, it is worthwhile to consider whether demographic factors (e.g., age, gender, race, level of education, and employment status) impact the ease with which older adults use these telehealth modalities. Furthermore, the influence of factors such as dementia, psychiatric illness, and caregiver availability also needs to be better characterized. These variables can guide the efforts to maximize the usability and acceptability of telehealth interventions in older adults and, potentially, subsequent public policy.

Telemedicine-based approaches also continue to find application in providing clinical services in underserved areas, and one of the major consequences of the COVID-19 pandemic has been the explosion of ways in which telemedicine-based mental health services are accessible. Beyond video-based interactions between providers, services such as guided therapy apps, mindfulness programs, and chatbots have all found niches that expand the offerings and availability of mental health care, even in remote areas. A challenge for the future will be to ensure that there is equity in how telecare is provided and ensure that older adults without access to technology or a lack of proficiency in its use are not left behind.

Mobile Health

The term *mobile health* (or mHealth) loosely refers to the use of mobile phone technology for health care. In the past 10 years, there has been an explosion of interest in the uses of mobile devices in collecting health-related data, delivering health care, and facilitating interaction among patients, clinicians, and family stakeholders in the service of improving health. There are several reasons why mobile phones are specifically attractive as a potential platform for health care delivery. For one, recent estimates indicate that the penetration of mobile phones is approximately 90% in the United States. This penetration has increased at a rapid pace over the past 10 years and is nearing complete saturation. Older adults are, as with other technologies, less likely to own a mobile device than their younger counterparts, but as of 2014, 74% owned a mobile device. A review of 28 studies by Wilson and colleagues found that while older adults experienced some barriers to smartphone use (e.g., being a “digital native,” sensory or motor problems with interaction), smartphones and tablets were central to the lives of older adults sampled. The use and ownership of cell phones cut across other sociodemographic strata, with poor and underserved subpopulations often more likely to rely on mobile devices. This ubiquity of mobile devices provides a remarkable new conduit for innovation in health care delivery, and both academic and consumer developers have responded with the creation of a wide variety of mHealth applications.

The “value proposition” of mHealth is that it could enable delivering care outside of the typical clinic setting, either by augmenting interventions by providing additional interaction to support everyday health-sustaining behaviors (e.g., medication adherence) or by extending them beyond the time frame typically associated with the intervention. Remote mobile technologies can automate tasks that are typically performed by clinicians, such as appointment reminders, symptom and risk assessments, and scheduling, which could increase access to health care by reducing the need for face-to-face interaction. Additional benefits of mHealth are the ability to help compensate for prospective memory problems and facilitate the self-management of chronic illnesses. Mobile phones can provide timely reminders for beneficial health behaviors and deliver “just-in-time” interventions that could diminish the time gap between symptom emergence and intervention. Given that large volumes of data are available both through patient reports and sensors linked with the devices, it may also be possible to identify actionable patterns of symptoms over time in chronic recurrent illnesses and then intervene based on accumulating knowledge of individual patients so as to personalize care to individuals. Finally, provider-facing applications may assist in managing patient needs and diagnoses, reducing the need for brick-and-mortar clinic settings, which may be particularly attractive in the developing world. To temper enthusiasm somewhat, while collectively, these advantages offer great potential, the mHealth field is nascent, and so few mHealth interventions are yet embedded in routine care.

mHealth interventions typically involve a number of elements, which include the following: (1) a patient-facing platform. This is an interface that provides interaction, and typical platforms include the feature phone that is not connected to the Internet yet capable of voice or SMS interaction, a smartphone, or a dedicated device that is developed for the mHealth intervention, (2) one or more active communication channels to the patient, which may include SMS, live or asynchronous voice, email or in-app interaction. Alternatively, communication can involve passive sensing, or the collection of data without the need for patient response. Sensors are highly varied but could include activity measured by GPS antenna or accelerometer embedded in the device or ancillary sensors that couple with the device, such as blood pressure or glucose, (3) a set of algorithms that govern the interaction between data captured on the device and a clinical response. These may be automated and do not involve direct clinical oversight or might flag or prompt interaction from clinicians tasked with monitoring the incoming data, (4) a server that stores and allocates interactive elements to individual users in the intervention network, and, finally (5) a clinician facing dashboard or utility that enables a supervising clinician the capacity to manage, monitor, and interact with patients in the mHealth intervention. Many iterations of the above architecture are possible, with some apps that are completely consumer-facing without clinician oversight and others offering complex layered interaction with a variety of clinicians and other technologic assets such as the electronic medical record.

A number of reviews have addressed the question of whether mHealth interventions are effective. To date, the evidence is mixed. While recent systematic reviews have indicated that mHealth interventions seem effective for certain commonly researched areas, in particular smoking cessation and HIV medication adherence, other areas were limited in terms of evidence. A key limitation of the literature is the lack of high-quality randomized clinical trials. Very few of the applications that have actually been subjected to rigorous randomized trials are actually available to consumers. As such, the mHealth field is somewhat bifurcated into unproven apps directly available to consumers and academic products that are not translated to products available to consumers.

mHealth solutions for mental health and aging are far fewer in number than are ones for chronic medical illness. Nonetheless, there are some examples of apps designed for chronic mental illness, including schizophrenia, bipolar disorder, and substance abuse. Review studies have indicated that mHealth applications could provide important resources to support access, initiation, engagement, and sustainment of psychiatric care. A study published by Depp and colleagues indicates that even among older adults with schizophrenia, the rate of past exposure to mobile phones is high (78% had ever used one), and the great majority felt comfortable with the use of mobile devices. Despite high rates of prior use, few patients currently own mobile devices, which creates a barrier to future implementation without subsidization of mobile devices. Moreover, data gathered by mobile devices on symptoms appear to be valid—one study indicated that daily mood ratings obtained in a sample of patients with bipolar disorder appear to be more consistent with clinician ratings than were identical mood ratings recorded on paper-and-pencil forms. As such, evidence, albeit limited, suggests that patients with chronic and severe psychiatric impairments are capable of using mobile devices and can report accurately on symptoms that would be the likely focus of intervention.

There have been few randomized trials of mHealth interventions in psychiatric illnesses. Depp and colleagues conducted a 24-week randomized controlled 81-patient trial of a four-session psychoeducational intervention for bipolar disorder augmented by automated personalized reminders contrasted to a four-session intervention without mobile device augmentation. Effect sizes for this trial indicated that the mobile augmented condition was associated with a greater reduction in depression by a medium effect size; the magnitude of this effect is notable given the challenges met by many interventions in impacting bipolar depression. However, these positive effects seemed to dissipate at follow-up, suggesting that gains made by mobile augmentation were not sustained once mobile intervention ceased. Separate studies by Granholm and colleagues, as well as Ben Zeev and colleagues, have each reported positive changes in medication adherence, psychiatric symptoms, and socialization in open trials of SMS and smartphone application interventions in schizophrenia. In addition, Depp found in an adjacent study that a similar population will self-initiate mobile intervention even when not prompted to engage. Gustafson and colleagues evaluated a smartphone application designed to extend the impact of rehabilitation for alcohol use disorder and found significant reductions in alcohol use in a randomized trial of more than 300 patients. This was later supported in a follow-up study led by Mckay assessing a phone call versus application intervention, in which both were effective but not any more powerful when combined. As such, these emerging clinical trials offer an encouraging indication of the feasibility, acceptability, and effectiveness of mobile interventions that augment psychiatric care. Given the large number of ongoing clinical trials of mHealth interventions in mental health conditions (see ClinicalTrials.gov), it is likely that data on the effectiveness of mental health targeting mobile applications will proliferate in the near future.

There are a variety of exciting and emerging directions for mHealth mental health applications. For one, there is increasing evidence that passive sensors (described in detail in the next section) can identify patterns consistent with a change in symptoms. For example, vocal prosody appears to covary with mood symptoms in bipolar disorder. If passive sensors can supplant patient reports, it may be quite possible to efficiently monitor and allocate preventative care to patients without the need for patient-initiated interaction. Second, and relatedly, there is increasing interest in the use of predictive analytics in leveraging the intensive longitudinal data generated by mHealth applications to develop dynamic, personalized risk profiles for serious adverse events like suicide. Thompson and colleagues reported preliminary data on a statistical technique called functional data analysis that is optimized for the noisy repeated data that are generated in mHealth and found that a simple algorithm modeling negative affect highly accurately predicted subsequent increases in suicidal ideation in a small sample of bipolar patients. Other approaches have used natural language processing to identify patterns predictive of suicide. Other frontiers include the use of more sophisticated algorithms to intervene with patients depending upon social contexts or the enlistment of social networks and peers in managing mental illnesses. As with mHealth in general, while exciting, each of these novel approaches would need to be adequately evaluated in well-designed clinical trials to move beyond proof-of-concept stages to translate to impact on psychiatric care.

In summary, the field of mHealth is a rapidly evolving one that, while yet to fully deliver on its promise, holds great potential for impacting the landscape of mental health care in older adults with psychiatric illnesses. Globally, older adults are increasingly adopting mobile technologies, and the development and tailoring of mHealth interventions to meet the needs of older adults will be a subject of research for years to come.

Sensors/Self-Monitoring Devices

With accelerometers available readily, monitoring of movement, behavior, balance, as well as physiologic variables such as heart rate and oxygen saturation is becoming simpler. Researchers are developing sensors that can perform blood tests by connecting through ports to the computing chips in cell phones. Qualcomm—a San Diego-based communications company—had initiated a competition encouraging developers to produce a “tricorder,” a noninvasive handheld diagnostic device that can monitor multiple physiologic functions and perform diagnostics without the availability of sophisticated labs. In 2017 the winning Final Frontier team created DxtER, an AI engine used to diagnose medical conditions using noninvasive sensors that collect vital signs and other biometrics.

Table 55.6h–2 outlines various types of sensors and the types of roles they can play in promoting aging care.

The use of sensor technology in health care especially for older adults, is a rapidly evolving field. Wrist-worn and body-worn accelerometers have found multiple uses in health care to infer behavior from physical activity. Most of these sensors rely on accelerometry to detect motion. However, the new generation of sensors can utilize more sophisticated approaches, such as computer vision or radio sensing, to extract markers of behavior from large quantities of data.

Chemical sensors and body-worn sensors are noninvasive alternatives that have shown promise in vitro testing. These sensors can be easily produced on a mass scale using 3D printer technology and can be mounted on a variety of platforms such as mouthguards for measuring electrolytes and metabolites in saliva), textiles (for measuring sweat or urine), bandaids (for measuring chemicals in interstitial fluids). These sensors can also be integrated with microneedles that are minimally invasive and painless. Currently, sensors in development can measure a wide range of electrolytes and metabolites, including glucose, lactate, sodium, potassium, pH, and urea.

Wearable sensors can already measure vital signs, and commercially available devices such as the recently launched Apple Watch can measure and transmit heart rate in real time. These sensors have led to the creation of novel “personal health data.” Philips Respironics developed the Actiwatch Spectrum, an actigraphy device that measures daily activity as well as sleeping patterns. Several health care systems are working on developing infrastructure that can incorporate such data with electronic health records to have a more complete set of data on patient health and well-being.

Wireless electroencephalography (EEG) devices are also gaining popularity. These devices consist of a semicircular band with electrodes on either end, which can be placed bitemporally to collect bitemporal EEG readings. These devices are also being utilized to gather EEG data to predict emotional states, pain, and anxiety.

As the use of wearable devices and sensors grows, there is a need to develop algorithms that can generate relevant clinical alerts based on changes in physiologic and environmental data. Further, medical providers, caregivers, and the wearers themselves will need tools to visualize these data and use that information to predict or respond to abnormal events.

Table 55.6h–2.
Classification of Sensor Technology

Type of Sensor	Examples	Role	Stage of Development (at time of writing)
Implantables	Nanosensors	Monitoring of parameters in very constrained environments, e.g., small blood vessels, GI	Conceptual development and prototyping. Clinical validity not yet established
Neurosensors	Bitemporal EEG monitors	Portable noninvasive monitoring of neurologic (e.g., seizures), psychiatric (e.g., anxiety), and physiologic (e.g., pain) parameters	Commercially available. Clinical validity under assessment
Wearable physiologic monitoring devices	Microneedle-based sensors, tattoo sensors	Noninvasive monitoring of electrolyte and metabolite levels	Prototypes developed. In-vitro device validity and reliability under assessment
Wearable external monitors	Accelerometers, gyroscopes	Monitoring of psychical activity patterns and sleep	Commercially available and clinically validated

Environmental sensors	Motion sensors, gas sensors	External continuous monitoring of activity and functioning	Motion sensors are currently commercially available. Gas sensors are in prototype stage

Smart Homes and Senior Design

Advances in ICT have led to a growing interest in the development of “smart homes,” that is, living environments equipped with multiple sensors that can monitor and detect a person’s activity. The goal of these technologies is to improve the quality of life and promote the physical independence of older adults. Within this context, multiple devices and technologies may be embedded within the infrastructure of the residence and may thus function without the need to train residents on their use. For example, a heat sensor mounted on a stove can detect if the stove has accidentally been left on for a prolonged period. Motion sensors can keep track of a person’s mobility in the house and may provide information on clinical issues such as triggers of insomnia and markers of agitation such as pacing. Au Yeung and colleagues utilized passive environmental sensing techniques to monitor the behaviors of patients with late-stage dementia. They used a combination of motion sensors, actigraphy devices, and bed pressure mats to monitor the behaviors of older patients with dementia in assisted living facilities and found that passive environmental sensing techniques can help monitor patients without obstructing their daily activities and help providers adjust the patients’ medications based on their activity. Existing prototype smart homes (e.g., the Awarehome at the Georgia Institute of Technology) serve as research environments that guide the development of technologies that may aid the development of senior-friendly living. Smart homes may be beneficial for individuals living alone who are unable to seek help in emergencies (e.g., falls, strokes, unconsciousness), older adults with cognitive (e.g., memory loss, visuospatial deficit) and/or physical (e.g., hearing, visual, or mobility) impairments, and individuals needing assistance in performing basic and instrumental activities of daily life. Additionally, smart homes may provide integrated telehealth tools that can connect people living in rural or underserved areas with medical providers. With regard to the use of smart homes, important questions such as health outcomes, clinical algorithms for predicting potential health problems, user perception and acceptance, and privacy issues are yet to be fully addressed. Ideally, a smart space, depicted in Figure 55.6h–1 for health monitoring, should be able to integrate implanted, body-worn, and environmental sensors to measure changes in the individual's body, their behavior, and changes in the environment. Such a space would also incorporate mHealth approaches (described above) and machine learning and big data analytics (described below) to generate relevant health information with the capability to be predictive of health events. Smart home technologies are also increasingly being deployed in settings such as Assisted Living facilities and dementia care facilities to facilitate a higher level of monitoring without increasing the need for human resources.

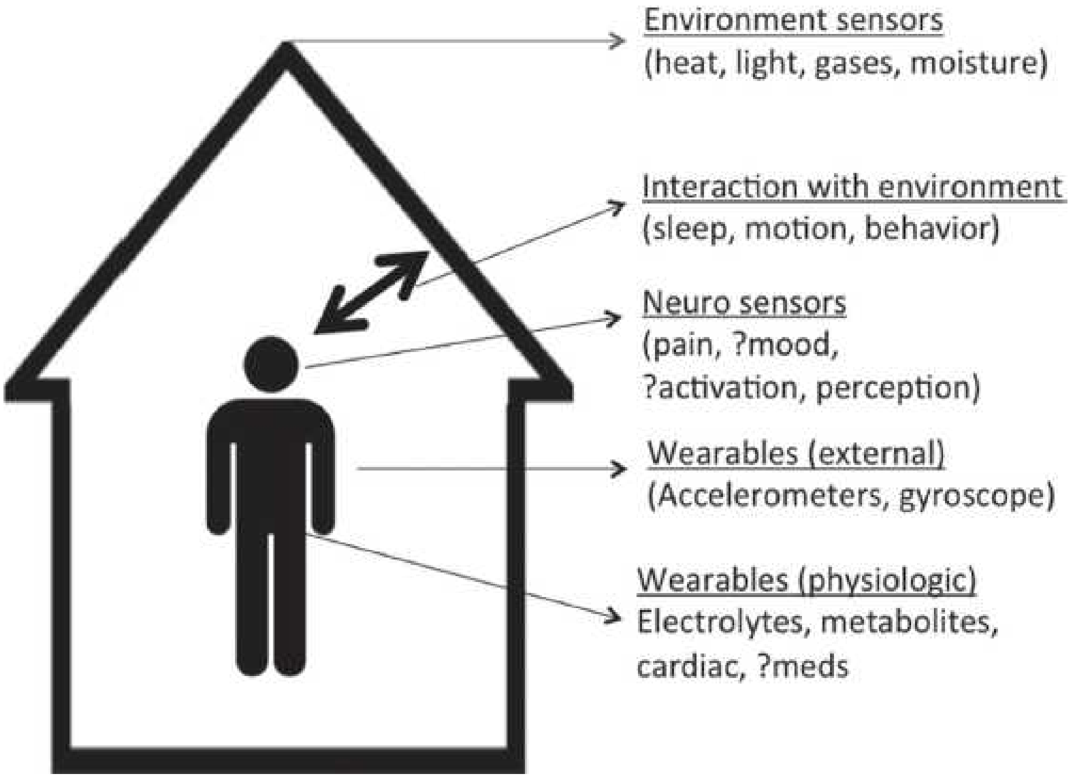
Environment sensors
(heat, light, gases, moisture)
Interaction with environment
(sleep, motion, behavior)
Neuro sensors
(pain, ?mood,
?activation, perception)
Wearables (external)
(Accelerometers, gyroscope)
Wearables (physiologic)
Electrolytes, metabolites,
cardiac, ?meds

FIGURE 55.6h–1. Smart room. This figure indicates the multiple parameters of interaction between persons and their environment that can be quantified by sensors. This figure includes both existing and in-development sensor technology.

PERS are another set of important applications of ICTs. These are relatively inexpensive devices designed to facilitate emergency responses for older adults. They may reduce mortality rates and hospital utilization and may be especially useful in rapid detection and early response to falls, and the use of this technology is a subject of great interest and study.

Overall, it remains to be seen how sophisticated home-based solutions and the growing number of advanced ICT applications are tolerated and utilized by older adults in general and specific groups such as those with disabilities or dementia.

Games

Interest in clinical and cognitive applications of video games and game-based interventions continues to grow.

Broadly, computerized games can be classified into one of four types:

- Games developed for or played on commercially developed gaming systems such as the Sony Playstation, Microsoft Xbox, or Nintendo Switch
- Commercially available cognitive games such as Lumosity or Posit Science modules
- VR games
- Games developed for research purposes

In one study conducted by Rosenberg and colleagues in 2010, researchers utilized the Nintendo Wii system, which relies on the persons using the system to perform physical tasks, and administered an intervention of physical exercise conducted indoors to assess improvements in subsyndromal depression in older adults. In a small sample of 19 patients, they noted that exercise using this game, performed for 35 minutes three times a week, resulted in statistically and clinically significant improvements in depressive symptoms.

Several cognitive games have been marketed in the past 10 years, aimed at improving cognitive function in older adults who have experienced cognitive decline and also at younger adults who are cognitively intact but invested in maintaining a high level of cognitive function. However, evidence of the efficacy of such games remains mixed. Despite early promise, large meta-analyses of data from studies of such games indicate that, at best, they may result in some domain-specific improvement in the cognitive abilities specifically targeted by each game. However, there is no evidence so far to suggest that such gains may be more generalizable to other domains.

There is also existing evidence of the efficacy of a game that was custom-developed to require multitasking. This game—a 3D racing game called Neuro Racer—was administered to older adults, and the authors noted that by performing the multitasking required by the game, older adults demonstrated benefits in multiple cognitive domains, including those such as sustained attention and working memory, which were not targeted by the game. Moreover, the resultant cognitive enhancement persisted for up to 6 months. This study provides a counterpoint to previous studies on cognitive games and suggests that rather than targeting specific domains, games that tap multiple skills may be efficacious in improving cognitive outcomes. More recent studies are examining how gamification-based therapeutic approaches can increase adherence and stimulate brain changes that can lead to improvement in depression.

In a 2019 systematic review, Skurla and colleagues highlight the saliency of using VR in aging adults experiencing disorders or impairments as an alternative approach to traditional cognitive measures, interventions, and training. The culmination of existing VR literature from 2008 to 2020 suggests a need for creating standardized VR simulations to use in future research. The feasibility of using VR to measure cognitive impairments across publications yields promising results and has value for utilizing more life-like ecologic assessments within laboratory settings. For example, a grocery store simulation task can evaluate things like working memory, spatial cognition, or daily living activities within the safety of a laboratory setting. Generally, this review emphasizes that studies found significant differences in VR cognitive assessment tasks between healthy and impaired samples. Additionally, VR studies could distinguish different cognitive domains, suggesting the possibility of future VR diagnostic elements. Given the increased availability of this technology and its shift from computer-mouse simulation to headset and navigation technology, there is a lot of room for innovation in using this as a research methodology in the future.

Overall, the use of a computer gaming interface has been found to be an acceptable and enjoyable method for a wide range of tasks and tests for older adults. Given the ease with which games can now be accessed on ICTs, this application is likely to grow rapidly over the coming years.

Tablets

Given the larger screen and the consequent versatility of the touchscreen, tablets are increasingly being used for a variety of applications in older adults.

Several existing cognitive screening tools and other performance-based standardized tests are being adapted to iPad versions, including the UCSD Performance Skills Assessment. Clinical applications, especially for comprehensive neuropsychological testing using tablets as the platform for administration, are beginning to gain acceptance. A study by Mielke and colleagues has demonstrated that the use of iPad tablets may be especially acceptable to older adults and that testing scores are comparable to those from regular pen-and-paper testing. With the proliferation of telemedicine, device-based health care, including testing, is gaining vogue for older adults.

In addition to applications developed for delivery using tablets, there is evidence that simply learning to use a tablet device may be beneficial. Older adults are believed to be more receptive to learning these devices over computers because of larger screens, intuitive interfaces, and touch screens. Learning to use an iPad may result in improvement in cognitive domains, including episodic memory and processing speed, though there is no evidence it can improve visuospatial function. Programs created for teaching older adults to use technology, including tablets, may also promote well-being by enhancing intergenerational activity and new learning. In one study, Kamat and colleagues demonstrated that iPads may be effectively used to control behavioral symptoms in older adults with dementia. In this study, the authors noted that because the iPad is a versatile device that can be used as a platform for a broad range of tasks, it may be an especially effective mechanism for administering nonpharmacologic interventions. Its primary advantage would be the ability to administer multiple interventions (e.g., reminiscence through pictures, music therapy, alternative engagement through videos, and cognitive stimulation) using a single device.

Social Media

Older adults are the fastest-growing segment of users of social media such as Facebook; this population is increasingly leveraging this technology to remain connected with family and friends. A 2021 survey by the Pew Research Center found that 50% of adults over 65 use Facebook. Additionally, 45% of this population also uses social media in general, such as Instagram, TikTok, Snapchat, and Facebook. There is growing interest in creating Facebook interfaces that are simplified to facilitate use by older adults. Improving social connectedness among older adults can have a broad range of implications for enhancing well-being. The behavioral and psychological correlates of social media use in older adults warrant close empirical attention.

Caregiver Applications

ICT applications have the potential to be useful not only to older adults but also to their caregivers. These tools include peer networking, professional support, and resource identification. One innovative intervention for caregivers utilized a computer-mediated, automated, interactive system to reduce caregiver stress and aid with managing disruptive behaviors exhibited by patients with Alzheimer's disease. While these applications are promising, given the rising number of people serving in caregiver roles, the outcomes associated with these technologies are yet to be fully characterized.

FUTURE DIRECTIONS

Robotics and Nanotechnology

With the rapid growth of technology, there has been a greater emphasis on developing robotic devices that can take over many physically demanding tasks of daily life. Promising ICT applications such as robotic-assistive devices for activities of daily living and other voice-activated technologies have been receiving greater attention for their potential use to reduce the physical demands of daily activities such as cooking, but further research is required to examine how they are suited for the needs of older adults. Robotic devices may even be implantable or flexible and wearable and may play an assistive role in limb or organ function. Similarly, nanotechnology, which is growing rapidly, is considered a modality for intervention at the cellular level for aging-related disorders, especially those related to telomere dysfunction, reversal of oxidative cellular damage, and fabrication of stem cells and stem-cell–based therapies.

In addition to approaching robotics as a means of physical health, research surrounding social robots is exploring this technology and its utility in social interaction and mental health intervention. A 2021 review explored the potential for social robots in geriatric mental health as social companions, means of social interaction research, possible intervention mediums, and more widescale daily living utility. By investigating how humans interact with robots' researchers may identify how to optimize these relations for mental health interventions. The capacity for programming robots to identify emotions, collect speech patterns, and provide a social companion is an intriguing solution to things like increased independence in older adults and complements existing mental health support.

Big Data

Big data is a broad term used to loosely denote large volumes of data, whether collected in a structured or unstructured manner, that can be mined and analyzed for information. The rise of big data has paralleled the development of sensor technology and rapid real-time analytics. As sensors allow one to collect extremely large volumes of behavioral and other data in real time, big data analytics will be essential to allow for rapid, effective analysis of this data to generate information and decision-making. Big data analytics are also likely to play a key role in developing applications aimed at the prediction of health events and preventative medicine.

BARRIERS AND LIMITATIONS

Although there have been significant technologic advances, older adults continue to experience barriers to using technology effectively. These include inadequate training to use the devices, lack of knowledge about the technologies available, inadequate design and equipment suitability, and cost/payment issues. Some of these are reviewed below.

Integration With Health Care System

Despite rapid growth and interest in technology and aging, significant issues persist. For technologies that involve gathering personal health data, there remains incomplete integration with electronic health records. As a result, physicians often view such data (e.g., accelerometer data, sleep data) as burdensome as it involves the additional step of accessing, processing, and interpreting data. However, several companies that develop electronic medical records (e.g., Epic) are in the process of developing health apps that integrate with patients' electronic medical records. Manufacturers of commercial health care monitoring devices are also in the process of establishing ways for smoother integration of device data health with records.

Ultimately, it is likely that personal health data will be automatically integrated with electronic medical records as a routine procedure, allowing physicians to review both sets of data simultaneously.

Billing

The ICT applications for older adults hold great promise for the improvement in their quality of life and physical wellness. However, these technologies are often prohibitively expensive. In the United States, a large number of assistive technologies are not reimbursed by public or private insurance systems. The prospect of restricted payment hinders the development and widespread use of technology by those who can benefit from it the most. Congress and the Federal Government may promote the use of health technologies by incentivizing their use by providers and changing the policies regarding reimbursement through the Medicare program.

Literacy and Usability

Although ICTs have considerable potential for benefit in aging contexts, literacy, and physical limitations may pose a barrier to their utility in this population. Unsurprisingly, older adults' perceptions regarding the usability of ICTs appear to be related to their acceptance of technology. Factors such as perceived usefulness, perceived ease of use, and subjective norm have a strong impact on older adults' acceptance and use of ICTs.

Privacy and Ethics

Privacy concerns regarding the data generated by wearable and environmental monitoring systems have a significant bearing on the motivation of older adults and their caregivers to adopt these technologies. The security of these data is often trusted by a third party, and this may be associated with real or perceived disadvantages, which at present are not well characterized.

Information Overload

Although the large volume of data generated from wearable sensors and smart home devices is central to the appeal of these technologies, there are important issues that warrant consideration. The devices available tap different modalities (e.g., gait, skin conductance, diet, sleep) and provide heterogeneous longitudinal data. There is no consensus regarding the best methods to appropriately aggregate, analyze, and interpret the data collected. Episode Discovery, an unsupervised algorithm, has shown early promise in monitoring the health of older adults and identifying unusual or dangerous changes in behavior. Algorithms such as this are yet to be fully characterized with clinical populations.

Age-Related Impairments

Declines in working memory, attention, and processing speed commonly accompany normal aging. These age-related declines in cognition have important implications for technology access and use in older adults. For example, training older adults to use new platforms or technologic skills may be impacted by declines in working memory and attention, while their ability to discriminate and select targets from complex displays may be impacted by declines in information processing skills. These are important aspects that warrant close consideration during the development of technology for older adults.

Lack of Regulation and Evidence

Sensors and other health care devices remain both understudied and unregulated. Although the FDA is stepping up its effort at monitoring health sensor technology, significant work will be required to establish the safety, validity, and reliability of these devices before they are widely deployed. A further issue is that despite a recent wave of technology targeting older adults, the majority of these technologies are developed by technology firms and marketed without sufficient evidence to establish the safety, efficacy, reliability, and validity of these devices to perform the tasks they are designed for. There is a great need for data on how these technologies are used and how effective they may be. Geriatric psychiatrists have a great opportunity to be leaders in the area of establishing parameters for technology use and development in medicine.

FURTHER READINGS

AARP. Technology, tablets, social media and more - AARP [Internet]. 2015. http://www.aarp.org/home-family/personal-technology/

Armontrout JA, Torous J, Cohen M, McNiel DE, Binder R. Current regulation of mobile mental health applications. *J Am Acad Psychiatry Law*. 2018;46(2):204–211.

Au-Yeung WM, Miller L, Beattie Z, et al. Monitoring behaviors of patients with late-stage dementia using passive environmental sensing approaches: a case series. *Am J Geriatr Psychiatry*. 2022;30(1):1–11.

Beishon LC, Elliott E, Hietamies TM, et al. Diagnostic test accuracy of remote, multidomain cognitive assessment (telephone and video call) for dementia. *Cochrane Database Syst Rev*. 2022;4(4):CD013724.

Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mHealth research: lessons from 3 mobile intervention studies. *Adm Policy Ment Health*. 2015;42(2):157–167.

Carotenuto A, Traini E, Fasanaro AM, Battineni G, Amenta F. Tele-neuropsychological assessment of Alzheimer's disease. *J Pers Med*. 2021;11(8):688.

Chaudhuri S, Kneale L, Le T, et al. Older adults' perceptions of fall detection devices. *J Appl Gerontol*. 2017;36(8):915–930.

Chaudhuri S, Oudejans D, Thompson HJ, Demiris G. Real-world accuracy and use of a wearable fall detection device by older adults. *J Am Geriatr Soc*. 2015;63(11):2415–2416.

Choi YK, Lazar A, Demiris G, Thompson HJ. Emerging smart home technologies to facilitate engaging with aging. *J Gerontol Nurs*. 2019;45(12):41–48.

Choi YK, Thompson HJ, Demiris G. Internet-of-things smart home technology to support aging-in-place: older adults' perceptions and attitudes. *J Gerontol Nurs*. 2021;47(4):15–21.

Depp CA, Harmell AL, Vahia IV, Mausbach BT. Neurocognitive and functional correlates of mobile phone use in middle-aged and older patients with schizophrenia. *Aging Ment Health*. 2016;20:29–35.

Depp CA, Moore RC, Perivoliotis D, Granholm E. Technology to assess and support self-management in serious mental illness. *Dialogues Clin Neurosci*. 2016;18(2):171–183.

Elbaz S, Cinalioglu K, Sekhon K, et al. A systematic review of telemedicine for older adults with dementia during COVID-19: an alternative to in-person health services?

Front Neurol. 2021;12:761965.

Gustafson DH, McTavish FM, Chih M-Y, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. *JAMA Psychiatry*. 2014;71(5):566–572.

Kasckow J, Felmet K, Appelt C, Thompson R, Rotondi A, Haas G. Telepsychiatry in the assessment and treatment of schizophrenia. *Clin Schizophr Relat Psychoses*. 2014;8(1):21A–27A.

Kulpa E, Rahman AT, Vahia IV. Approaches to assessing the impact of robotics in geriatric mental health care: a scoping review. *Int Rev Psychiatry*. 2021;33(4):424–434.

McKay JR, Gustafson DH, Ivey M, et al. Efficacy and comparative effectiveness of telephone and smartphone remote continuing care interventions for alcohol use disorder: a randomized controlled trial. *Addiction*. 2022;117(5):1326–1337.

Min J, Sempionatto JR, Teymourian H, Wang J, Gao W. Wearable electrochemical biosensors in North America. *Biosens Bioelectron*. 2021;172:112750.

Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. *Gen Hosp Psychiatry*. 2013;35(4):332–338.

Poulin C, Shiner B, Thompson P, et al. Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS ONE*. 2014;9(1):e85733.

Rabinowitz T, Murphy KM, Amour JL, Ricci MA, Caputo MP, Newhouse PA. Benefits of a telepsychiatry consultation service for rural nursing home residents. *Telemed JE Health*. 2010;16(1):34–40.

Redmond SJ, Lovell NH, Yang GZ, et al. What does big data mean for wearable sensor systems? Contribution of the IMIA wearable sensors in healthcare WG. *Yearb Med Inform*. 2014;9(1):135.

Rosenberg D, Depp CA, Vahia IV, et al. Exergames for subsyndromal depression in older adults: a pilot study of a novel intervention. *Am J Geriatr Psychiatry*. 2010;18(3):221–226.

Satariano WA, Scharlach AE, Lindeman D. Aging, place, and technology toward improving access and wellness in older populations. *J Aging Health*. 2014;26(8):1373–1389.

Skurla MD, Rahman AT, Salcone S, et al. Virtual reality and mental health in older adults: a systematic review. *Int Psychogeriatr*. 2022;34(2):143–155.

Soulas J, Lenca P, Thépaut A, eds. Monitoring the habits of elderly people through data mining from home automation devices data. In: *Progress in Artificial Intelligence. Proceedings of the 16th Portuguese Conference on Artificial Intelligence, EPIA 2013, 9–12 September, 2013*. Springer; 2013;16:343–354.

Taha J, Czaja SJ. Aging, cognition, and technology systems. In: Patel V, Arocha J, Ancker J, eds. *Cognitive Informatics in Health and Biomedicine. Health Informatics*. Springer; 2017.

Thompson WK, Gershon A, O'Hara R, Bernert RA, Depp CA. The prediction of study-emergent suicidal ideation in bipolar disorder: a pilot study using ecological momentary assessment data. *Bipolar Disord*. 2014;16(7):669–677.

Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? *PLoS Medicine*. 2013;10(2):e1001382.

Topo P. Technology studies to meet the needs of people with dementia and their caregivers a literature review. *J Appl Gerontol*. 2009;28(1):5–37.

Torous J, Summergrad P, Nassir Ghaemi S. Bipolar disorder in the digital age: new tools for the same illness. *Int J Bipolar Disord*. 2016;4(1):25.

Vahia IV, Kamat R, Vang C, et al. Use of tablet devices in the management of agitation among inpatients with dementia: an open-label study [published online ahead of print August 4, 2016]. *Am J Geriatr Psychiatry*. 2017;25(8):860–864.

Vahia IV, Sewell DD. Late-life depression: a role for accelerometer technology in diagnosis and management. *Am J Psychiatry*. 2016;173(8):763–768.

Van den Berg N, Schumann M, Kraft K, Hoffmann W. Telemedicine and telecare for older patients—a systematic review. *Maturitas*. 2012;73(2):94–114.

Wilson SA, Byrne P, Rodgers SE, Maden M. A systematic review of smartphone and tablet use by older adults with and without cognitive impairment. *Innovation in Aging*. 2022;6(2):igac002.

55.6i Positive Psychiatry of Aging

Barton W. Palmer, Ph.D., and Dilip V. Jeste, M.D.

There is a growing recognition that medicine should encompass not only the treatment of symptoms but also the promotion of well-being. For geriatric psychiatry, this means greater attention to the promotion of “successful aging” as well as to other aspects of well-being among older adults. The notion that mental health is more than the absence of symptoms or disorders is not new. In 1946 the World Health Organization (WHO) published a definition of *health,* which was not merely the absence of disease or disability, but rather “a state of complete physical, mental, and social well-being.” While seemingly self-obvious, this definition is potentially revolutionary in its implications for the practice of geriatric as well as general psychiatry. Yet, approximately 75 years later, the implications of that revolution have still not become fully weaved into mainstream psychiatry. Contemporary geriatric and general clinical psychiatric practice and training remain primarily restricted to the diagnosis and treatment of mental illnesses, while psychiatric research mainly focuses on elucidating the underlying pathology and ameliorative interventions to treat and prevent relapses of mental illnesses. Within the field of Positive Psychology, founded in the late 1990s by Martin Seligman and colleagues, a growing body of research strongly suggests that positive psychosocial factors (PPSFs) such as resilience, optimism, and social engagement are associated with higher levels of subjective well-being as well as lower morbidity and greater longevity. With the founding of positive psychiatry approximately 10 years ago, subsequent psychiatric research has shown the association of PPSFs with better well-being, lower morbidity, and longevity also generalize to those living with serious mental illnesses. Although many key questions still need to be addressed through empirical research, the last decade has finally brought together the empirical foundation for positive psychiatry as a potentially key aspect of 21st-century psychiatry.

DEFINING POSITIVE PSYCHIATRY

Positive psychiatry is the science and practice of psychiatry that seeks to understand and promote well-being through assessment and interventions aimed at enhancing behavioral and mental wellness. As a part of psychiatry, itself rooted in the field of medicine, positive psychiatry is grounded in biology and ultimately seeks to determine the neurobiologic underpinnings of the PPSFs associated with positive outcomes and to use that information to develop more effective biologic and psychosocial preventive, therapeutic, and rehabilitative interventions. As a branch of medicine, an important facet of positive psychiatry that at least partially distinguishes it from the complementary field of positive psychiatry is that it is primarily focused on the promotion of PPSFs and positive outcomes among those with or at high risk for psychiatric or physical illnesses.

HISTORY

In 1906 William James, a physician and psychologist by training, spoke of the need for a new approach to psychology that would study and apply principles underlying the success of the “mind cure.” The concept of *mind cure* referred to the purported healing powers of positive emotions and beliefs. Although in retrospect, the mind-cure movement itself appears an anachronism of questionable validity, James’s emphasis on positive emotions and beliefs illustrates an early effort to include PPSFs during the emergence of psychology as an independent field. Yet these inclinations were largely neglected until the mid-20th century, which brought the Humanistic Psychology movement of Abraham Maslow and colleagues. Their focus was on the study and promotion of health and creativity among individuals. But that movement also failed to fully shape psychiatry, as well as mainstream clinical psychology. Instead, positive psychiatry traces its most immediate lineage and debt to the Positive Psychology movement initiated by Martin Seligman and colleagues in the mid to late 1990s. In his 1998 Presidential Address, Seligman spoke of the need for “a reoriented science that emphasizes the understanding and building of the most positive qualities of an individual: optimism, courage, work ethic, future-mindedness, interpersonal skills, capacity for pleasure and insight, and social responsibility.” In the ensuing decades, Positive Psychology has grown and flourished into an international movement, but much of the research has focused on promoting well-being among the psychiatrically and physically healthy, and it has had less impact on the day-to-day practice of clinical psychiatry. This situation began to change over the past decade. In his May 2013 address as President of the American Psychiatric Association, Dilip Jeste commented “I expect that the future role of psychiatry will be much broader than treating psychiatric symptoms. It will seek to enhance the well-being of people with mental or physical illnesses. That is positive psychiatry.” positive psychiatry is allied with and partially overlaps with Positive Psychology in concepts and goals, but also complements in that psychiatrists are first trained as physicians, an applied biologically rooted profession focused on treating diseases and disabilities.

POSITIVE MENTAL HEALTH OUTCOMES

There has been a growing recognition that medicine should encompass not only the treatment of symptoms but also the promotion of well-being. Similarly, the goal of psychiatric interventions should not be just a reduction in psychopathology but also the promotion of positive mental states. Some of the better-known outcomes are summarized below.

Successful Psychosocial Aging

There is no single consensus definition of successful aging. A 2006 review of the empirical literature revealed 29 definitions in 28 published empirical reports. But there were some common themes. Traditional notions of successful aging have emphasized how much one has not aged, that is, the relative absence of physical, cognitive, and social disability. However, more recent studies focused on the perspectives of older adults themselves showing that self-perceived successful aging is not dependent on physical health. Large proportions of older adults with chronic physical illnesses, who would not meet objective disability-based criteria of successful aging, perceive themselves as aging successfully. Qualitative studies indicate that older adults consider the ability to adapt to circumstances and a positive attitude toward the future as more important than an absence of physical disease and disability. This distinction between physical disability and self-perceived successful aging may help explain a now well-documented paradox of aging. Specifically, while physical health and functioning, as well as some aspects of cognitive performance, tend to decline with increasing age, subjective well-being begins a steady improvement from youth that continues into old age.

Other Positive Outcomes

Well-Being. Echoing the WHO definition of health, positive psychiatry construes *well-being* as representing the presence of pleasant, positive experiences, and states such as life satisfaction and happiness (hedonic well-being), as well as eudemonic well-being, including a sense of purpose or meaning in life and self-acceptance. The concept of well-being has risen to prominence in recent years in economics, with self-reported well-being considered a measure of national development alongside traditional constructs such as the gross domestic product. Well-being is associated with better health outcomes and with reduced mortality/greater longevity.

Low Perceived Stress. The concept of stress has a complex history; the term may be used to refer to objective pressure or tension or the body's objective response to demands or change. In the context of positive psychiatry, however, the emphasis is on the individual's perception of stress, which appears to be a better predictor of deleterious changes in some biomarkers, such as telomere length, than objective stress. Perceived stress specifically represents the degree to which individuals perceive demands, problems, or challenges as exceeding their ability to cope with them. Recurrent unhealthy stress responses lead to cumulative wear and tear.

Post-Traumatic Growth. The link between severe stress and post-traumatic stress disorder (PTSD) is self-evident but not inevitable. Some people who experience significant traumatic events experience a form of personality growth that is referred to as *Post-Traumatic Growth*. It is characterized by an increased appreciation of life, altered priorities, closer interpersonal relationships, a greater sense of personal strength, and/or recognition of new possibilities for one’s life. Post-traumatic growth is fostered by the individual’s resilience in the face of severe stress–it is not the opposite of PTSD, as they may be copresent within the same individual in response to the same stress or trauma.

Recovery from Serious Mental Illnesses. Contemporary empirical data paint a brighter picture of the course of psychopathology and disability among older people with serious mental illnesses than traditional conceptualizations. For example, relative to their younger counterparts, older adults with schizophrenia tend to have better psychosocial functioning, including better adherence to medications and self-rated mental health, and a lower prevalence of substance abuse and psychotic relapse. There is no consensus definition of recovery (e.g., there are differences of opinion regarding the need for full symptom remission), but there is a general consensus that recovery involves decreased functional dependence and reduction of the impact of the illness on one’s life. Elyn Saks has described this capacity as finding “wellness within illness.”

Prevention. Prevention is one of the key positive outcomes of positive psychiatry. The focus on prevention has long roots in community psychiatry, where there have been efforts directed at primary, secondary, and tertiary prevention. In recent decades, medicine and psychiatry have experienced a paradigm shift from an almost exclusive focus on treating pathology to promoting health, as well as from caring for individual patients to enhancing the health of communities through preventive measures. Prevention is no less meaningful in the context of geriatric psychiatry. A wealth of empirical data indicate that prevention in later life can be practical and cost-effective.

Prevention can occur at multiple levels.

a. Primordial: Strategies aimed at altering partially malleable risk factors such as social determinants of mental health like social isolation and discrimination can prevent or at least delay the occurrence of depression.
b. Primary: Treatment or prevention of depression following stroke, myocardial infarction, and/or bypass surgery results in significant reductions in morbidity and mortality.
c. Secondary: Examples include prevention of suicides following successful treatment of suicidal behavior (suicide attempts); prevention of PTSD with counseling after personal or community disasters; and prevention of relapse in schizophrenia, bipolar disorder, or substance abuse with very early detection of prodromal symptoms and signs.
d. Tertiary: Reversing dementia due to vitamin B deficiency with appropriate vitamin supplementation illustrates tertiary prevention.
e. Quaternary: This is exemplified by the prevention of iatrogenic conditions such as antipsychotic-induced tardive dyskinesia, anticholinergic confusion, or delirium.

POSITIVE PSYCHOSOCIAL FACTORS

A key tenet of positive psychiatry is that positive outcomes such as those described above are at least partially mediated by and can be fostered by the enhancement of intrapersonal positive psychosocial traits as well as positive environmental/social factors, together subsumed under the rubric of PPSFs. Most of these factors are valued highly by society, yet until recently, they have been generally neglected in the biomedical sciences as imprecise, amorphous concepts that lack relevance to traditional medicine or neuroscience research. There have been numerous studies in recent years suggesting that the PPSFs are tightly linked to human biology and are associated with positive health outcomes.

POSITIVE PSYCHOSOCIAL CHARACTERISTICS

PPSFs may have as strong(if not stronger) an impact on patient well-being as common health risk factors such as obesity, hypertension, sedentary lifestyle, smoking, and even objective overall health status. The provision of effective interventions to strengthen PPSFs in older psychiatric patients could substantially increase the prevalence of recovery in seriously mentally ill older adults. Some of the key intrapersonal constructs (positive psychosocial traits) and positive environmental factors are considered below.

Positive Psychosocial Traits

Resilience. *Resilience* refers to ready recovery from or positive adaptation to illness, depression, or other adversities. Resilience is associated with greater well-being and reduced mortality risk among older adults. People in their 90s who endorse higher levels of resilience have a 43% higher likelihood of living up to 100 years compared to their peers with lower resilience. Among people with medical illnesses, resilience is also associated with better treatment adherence, better self-care behaviors, and improved quality of life. Among physically ill patients, resilience is associated with improved physical health, health-related quality of life, self-care, and treatment and exercise adherence. Resilience is also associated with better emotional health and well-being.

Optimism. The term *optimism* refers to a characteristic of people who tend to expect good outcomes. Optimism is associated with a lower risk of depression and higher overall subjective well-being, lower mortality, as well as less pain, better cardiovascular, oncologic, pregnancy, and other physical health outcomes, and better immune function. The precise mechanisms through which optimism has these effects are not fully clear, but it may partially operate through its effects on health behaviors. Older adults with higher optimism are more likely than less optimistic age-comparable peers to consistently engage in healthy behaviors.

Personal Mastery, Coping Self-Efficacy, and Meaning in Life. *Personal mastery* refers to an individual's expectations of her/his ability to achieve self-desired outcomes. A closely related concept is that of *coping self-efficacy*, that is, the individual's self-perceived ability to perform the requisite coping behaviors needed to overcome life challenges. Individuals high in personal mastery or coping self-efficacy have a belief in control over their life circumstances and future—that is, they have a general sense of control over things that happen to them and have greater confidence they can solve problems in their lives. In contrast, *optimism* more broadly denotes a tendency to expect positive outcomes regardless of whether or not one is the agent of those outcomes. In the context of chronic life stress, personal mastery and coping self-efficacy may promote specific, adaptive, goal-oriented behaviors, as well as maintenance of those behaviors, despite ongoing stresses. High levels of personal mastery tend to protect people from the potentially negative effects of stress on general health.

There is also growing evidence showing that *meaning or purpose in life* is associated with better mental and physical health. Developing a purpose leads to greater well-being and productivity.

Wisdom. *Wisdom* is a construct with heuristic value in that it appears across most cultures and contrasts with factual knowledge or even intelligence. Although traditionally, this construct was thought to be the province solely of philosophy and religion, recent successful efforts to develop standardized quantitative measures have also advanced the scientific study of wisdom. Wisdom may be defined as a complex trait with several domains, including prosocial attitudes and behaviors (compassion, empathy, and altruism), social decision making or problem solving, capacity for self-reflection and understanding/insight, decisiveness, acknowledgment and acceptance of uncertainty, emotional regulation, tolerance of divergent value systems, openness to new experience, spirituality, and a sense of humor. What unites these components is their utility for the self (e.g., greater well-being) and for others (better relationship with others and serving the common good). The cumulative experience of life with age may promote some of the dimensions associated with wisdom. Some, although not all, studies suggest an increase in several domains of wisdom with aging that may be important for an older adult’s ability to survive and even thrive in spite of worsening physical health. The Grandmother Hypothesis suggests that wisdom transmitted to younger generations may provide species fitness advantage and help neutralize the evolutionary pressure associated with loss of fertility in old age.

Recent research has also shown that wisdom has a strong inverse relationship with loneliness and may thereby protect against the deleterious effects of loneliness on medical morbidity and mortality. Although sometimes wisdom is contrasted with knowledge and “book learning,” there is emerging evidence that key components of wisdom, such as empathy, compassion, and self-compassion, may be enhanced with compassion training, mindfulness, and other contemplative interventions.

Spirituality and Religiosity. Religious and spiritual practices are reportedly associated with greater well-being as well as better health outcomes throughout the lifespan. It may be argued that *religiosity* is better conceptualized under the umbrella of positive environmental factors (see next section) as the term *religiosity* is usually reserved in reference to involvement in organized religion and thus has an inherent social connotation, whereas spirituality can be personalized. However, religiosity ultimately rests on the beliefs and behaviors/choices of an individual. The term *spirituality* refers to an internal experience, that is, the degree to which one's personal beliefs, thoughts, and behaviors focus on transcendent topics such as the meaning of life, belief in a higher being, power or level of consciousness, awareness, or existence. Religious and spiritual practices are associated with positive outcomes such as greater well-being as well as better health outcomes. These positive effects of religious practices may partially rest on the emotional support derived from the religious community. Religious and spiritual beliefs also tend to confer a sense of optimism. Spirituality has been reported to increase with age.

Positive Environmental Factors

In addition to intrapersonal positive psychological traits, the interaction of the person with the environment and the nature of that environment also impact the likelihood of positive outcomes. These environmental factors are described below.

Social Engagement and Support. *Social Engagement* refers to integration into a social network or social participation—that is, how well-integrated one is into a social network. Within this construct are the number and quality of close relationships, the frequency of socialization, and the degree to which one finds meaning or pleasure from social integration/interactions. A related construct of *social support* refers to the degree to which other people are available for emotional and physical support. Empirical research has shown that the quantity and quality of one's social relationships are associated with better mental and physical health outcomes, including lower mortality, as well as more specific health outcomes such as lower prevalence of hypertension, cardiovascular disease, dementia, depression, anxiety, and substance use disorders in young and old adults.

Family Dynamics. A particularly influential form of social engagement and support is the family. There is a large amount of empirical literature showing that family dynamics and the degree of a supportive family environment affect mental and physical health, including symptom severity, treatment response, and recovery.

Other Environmental Factors. Access to quality physical and mental health care, as well as environmental factors such as availability of safe parks and walking paths, community centers, and access to grocery stores for healthy eating alternatives, can promote or discourage physical activity and exercise, nutrition, and stress reduction, especially in older adults.

BIOLOGY OF POSITIVE PSYCHIATRY

As noted above, psychiatry, as a branch of medicine, is rooted in biology, and so is positive psychiatry. Biology of relevant constructs includes neuroscience as well as genetics and epigenetics involving interactions among genes, environment, and structure and function of the brain. This type of research has implications for developing biologic interventions to enhance positive traits such as wisdom, resilience, and optimism, and thereby benefit mental and physical health. Work on biologic aspects of PFSFs has accelerated during the past decade, but there remains an obvious need for additional investigations in this arena.

Potential Biomarkers of Positive Mental Health

Allostatic load and telomere length are among putative biomarkers of stress and aging but can also be conceptualized as potential biomarkers of well-being in reverse.

Allostatic Load. Allostatic load reflects the body’s cumulative physiologic response to frequent and sustained exposure to environmental stressors; it represents dysregulation of one or more biologic systems, multiple pathways to disease, and the cumulative cost of antecedent challenges which together can serve as an indicator of morbidity, as well as affecting cognitive functioning and psychological factors. There is no single measure of allostatic load, but a commonly used composite includes systolic and diastolic blood pressure, body mass index, serum cholesterol (with an emphasis on HDL), and plasma norepinephrine and epinephrine.

Telomere Length. Telomere length is an indicator of biologic age and may represent a cumulative index of the number of cell divisions and a cumulative record of exposures to genotoxic and cytotoxic processes such as oxidation. Telomere length is maintained by the enzyme telomerase, and there is some preliminary evidence that healthy lifestyle changes, as well as psychosocial stress reduction interventions, can increase telomerase activity.

Oxidative Stress. Oxidative stress, caused by an over-production of free radicals or underproduction of anti-oxidant molecules, damages DNA, protein, lipids, and other macromolecules in many tissues, with the brain being particularly sensitive. Oxidative damage to cellular components by free radicals may be an important component of the aging process. Thus, prolonged or repeated exposure to oxidative stress may accelerate aspects of biologic aging, leading to greater comorbidity and mortality.

Inflammation and Immune Function. Neuroinflammation and immune function also relate to several PPSFs. For instance, a low level of optimism has been reportedly associated with increased inflammation markers. Reduced personal mastery resulting from chronic stress may provoke excessive activation of the autonomic nervous system, which may, in turn, affect immune functioning. Several studies report direct associations of engagement in pleasant events with better health, including reduced plasma concentrations of inflammatory biomarkers and lower catecholamine response to stress.

Genetics and Epigenetics. Genes clearly play a role in determining the lifespan and health status in older age. Many PPSFs are also influenced by genetic factors. For example, resilience and optimism are both partially heritable. Beginning in development, an individual's genes and their interactions with environmental factors may shape the neural circuitry and neurochemical function expressed in PPSFs. Various genetic polymorphisms affect a person's limbic reactivity and prefrontal-limbic connectivity. Adaptive functioning of fear, reward, emotional regulation, or social-behavior circuits underlies a resilient person's capacity to face stress. Stochastic epigenetic changes occurring during brain development that drive resilience and hardiness may affect response to periods of high stress.

About 20% to 30% of the variability in human lifespan appears to be attributable primarily to genetic factors. Overall, longevity seems to be promoted by an absence of genes that predispose people to diseases rather than by the presence of "life-extending" genes. The risk for cognitive impairment in older age is clearly raised by the presence of the ApoE4 allele, but this may be attenuated by higher educational attainment or greater "cognitive reserve." In short, genetic vulnerabilities may be amenable to modification through intentional or incidental environmental changes.

Neuroplasticity Across the Lifespan

Throughout most of the history of neurology, the prevailing assumption was to see childhood as the period of brain development and later life as a period of neurodegeneration. But the past two decades of neuroscience research have shown reality to be more complex as both processes, development, and degeneration, occur throughout the lifespan. For example, adolescence is a period of programmed pruning of neural synapses, yet new synapses and even neurons can form in certain subcortical regions in older adults who stay active physically, cognitively, and psychosocially.

There are now convincing empirical demonstrations that neuroplasticity, with associated brain growth and development, continues into old age. Mechanisms underlying neuroplasticity include compensation for aging-associated degenerative changes (e.g., involvement of larger numbers of neuronal networks in the performance of activity) as well as synaptogenesis, increase in the number of dendritic spines, and even neurogenesis in selected areas of the brain such as the hippocampal dentate gyrus and periventricular region. Such neuroplasticity is related less to the genes inherited than to the behavior, attitude, and environment that affect the expression of those genes. It is this neuroplasticity that may enable PPSFs to affect an individual's well-being throughout the lifespan, although the brain is admittedly less plastic in later life than it is in childhood and early adulthood.

INTERVENTIONS

Research on interventions aimed at promoting PPSFs is in the early stages but seems highly promising. Treatment trials have shown beneficial effects of behavioral strategies buttressing resilience, changing attitudes toward aging, or incorporating meditation on outcome measures such as well-being, as well as alcohol consumption, biologic response to stress, and mortality rates in older adults**.** There is considerable diversity in the types of psychosocial or behavioral interventions reported, which include writing gratitude letters, practicing optimistic thinking, acts of kindness, replaying positive experiences, socializing, life review, mindfulness, cognitive behavior therapy, life coaching, goal training, and forgiveness therapy. Cognitive behavior therapy has been shown to be effective in enhancing optimism and facilitating the reappraisal of traumatic events in a more positive light. Increasing optimism has been found to increase adherence to pharmacotherapy in formerly pessimistic individuals. Interventions aiming to improve social engagement, while not specifically targeting depression, appear to have antidepressant effects in older people. Studies of the longer-term effects of meditation have reported reduced mortality rates in older adults at risk for negative cardiovascular outcomes. A number of randomized controlled trials have shown the efficacy of psychosocial interventions to enhance components of wisdom such as empathy and compassion, emotional regulation, and spirituality.

Interventions to Enhance Healthy Aging

A key component of positive psychiatry in the context of working with older adults is the concept of successful or healthy aging. The quest for "antiaging" therapies and interventions has gone on for thousands of years, with as yet little in the way of a "cure" for aging being found. However, there are interventions that can delay the onset of disability and disease and prolong "healthspan." In contrast to earlier assumptions, many of these interventions, such as physical exercise, appear to produce benefits even among older adults.

Lifestyle Interventions. Lifestyle behaviors are a logical place to start when thinking about enhancing positive outcomes in older people. Physical exercise, proper nutrition, activity participation, and regular health screenings and vaccinations are among the most powerful and modifiable predictors of the health span. In addition to the positive effect of healthy behaviors on the cardiovascular system, recent research has provided compelling data that improvements in brain structure and functioning can be attained from improving lifestyle habits.

PHYSICAL ACTIVITY. The axiom “what’s good for the heart is good for the brain” has gained research support in recent years. There is strong evidence from randomized controlled clinical trials conducted among sedentary older people that increasing aerobic physical exercise can enhance performance on cognitive tests. Together, the empirical literature suggests that initiating exercise participation in later life may have significant effects that extend beyond physical outcomes and include cognitive and emotional improvements.

Using animal models and examining the biologic effects of exercise, it is likely that physical activity has multiple pathways leading to improvement in cognitive and emotional health. By reducing hypertension, the risk for cerebrovascular diseases decreases. By reducing obesity, stress on the skeletal system is diminished. Physical activity may also have the secondary benefit of increasing social engagement. Fall risk may be decreased by increasing muscle mass. Research from animal models suggests that at the cellular level, exercise may provide a mild stressor that may act as a kind of “inoculation” against other kinds of stress and may promote neuronal growth.

The Centers for Disease Control and Prevention suggests that adults engage in 3 to 5 days per week of moderate-intensity exercise (equivalent to brisk walking) for 30 minutes or more, stretching every day, and strength-building activities 2 to 3 days per week. Only about one-third of Americans meet the minimum exercise recommendations, and the prevalence of meeting these criteria decreases with age. At particular risk for not engaging in recommended levels of physical exercise are older people from ethnic minorities, who have lower educational attainment, or who have sub- or supernormal body mass indices.

A central challenge to increase exercise participation is identifying the barriers to exercise in routine life. The “walkability” of the built environment is an important consideration for enhancing engagement in physical activity, especially for older adults because they appear to prefer walking over other exercises.

DIET AND NUTRITION. Diet and nutrition are clearly related to health and other positive outcomes, and no other strategy for healthy aging has received as much research support as caloric restriction. Caloric restriction has been examined in worms (Caenorhabditis elegans), rodents, primates, and humans. Reducing caloric intake has been associated with increasing the lifespan of certain strains of mice and primates. In addition to extending the lifespan, deceleration in other aging phenotypes accompanies caloric restriction. The exact mechanism of caloric restriction is unclear, but it may involve reducing the rate of metabolism, perhaps limiting cellular wear and tear. Whether and how calorically restricted diets can be safe and palatable in humans remains to be seen; however, there are ongoing studies examining the impact of caloric restriction in older, overweight, or obese adults.

The content of dietary intake is also of interest to promote health, and in an epidemiologic study, a “low-risk” diet has been recommended one that includes a higher ratio of polyunsaturated to saturated fats, high in cereal fiber and marine fatty acids, low in glycemic load, minimal trans fats, and multivitamins. Additionally, moderate alcohol drinking (1 to 2 drinks per day), especially red wine, appeared to be better in the Nurses’ Health Study. As with exercise participation, only a minority of older adults conform to such a diet, and a growing proportion of people, including older people, are obese. Indeed, the rising incidence of obesity in the United States may significantly slow the rate at which life expectancy increases over the next 30 years–that is, obesity may decelerate the dramatic gains that have been made in life expectancy. Practitioners working with older adults need to provide dietary education, explaining the risks associated with unhealthy diets and encouraging healthier alternatives.

Biologic Treatments. The potential for biologic treatments in positive psychiatry is largely awaiting further research, but there does seem to be potential for advances in this arena. Recent research suggests that it might be possible to use neurobiologic approaches to enhance resilience in the context of anxiety and mood disorders. With an increased understanding of the neural circuitry underlying resilience, new interventions may also be developed using findings indicating that individuals can be trained to modulate their own brain activity with the help of real-time functional magnetic resonance imaging (fMRI)-based neurofeedback. Biologic interventions may be developed on the basis of an improved understanding of neuroplasticity as well as gene variations related to positive traits. In addition, findings from animal and human studies might eventually yield new pharmacologic agents that can maximize the adaptive function of the HPA axis as well as monoamine, neuropeptide, and other neurochemical stress response systems.

FURTHER READINGS

Aftab A, Lee EE, Klaus F, et al. Meaning in life and its relationship with physical, mental, and cognitive functioning: a study of 1,042 community-dwelling adults across the lifespan. *J Clin Psychiatry*. 2019;81(1):19m13064.

Bejerholm U, Roe D. Personal recovery within positive psychiatry. *Nord J Psychiatry*. 2018;72(6):420–430.

Brown L, Huffman JC, Bryant C. Self-compassionate aging: a systematic review. *Gerontologist*. 2019;59(4):e311–e324.

Chakhssi F, Kraiss JT, Sommers-Spijkerman M, Bohlmeijer ET. The effect of positive psychology interventions on well-being and distress in clinical samples with psychiatric or somatic disorders: a systematic review and meta-analysis. *BMC Psychiatry*. 2018;18(1):211.

Eglit GML, Palmer BW, Jeste DV. Overview of measurement-based positive psychiatry. *Nord J Psychiatry*. 2018;72(6):396–403.

Gage FH. Adult neurogenesis in mammals. *Science*. 2019;364(6443):827–828.

Heckenberg RA, Eddy P, Kent S, Wright BJ. Do workplace-based mindfulness meditation programs improve physiological indices of stress? A systematic review and meta-analysis. *J Psychosom Res*. 2018;114:62–71.

Jebb AT, Morrison M, Tay L, Diener E. Subjective well-being around the world: trends and predictors across the life span. *Psychol Sci*. 2020;31(3):293–305.

Jeste DV. A fulfilling year of APA Presidency: from DSM-5 to positive psychiatry. *Am J Psychiatry*. 2013;170:1102–1105.

Jeste DV, Koh S, Pender VB. Perspective: social determinants of mental health for the new decade of healthy aging. *Am J Geriatr Psychiatry*. 2022:30(6):733–736. S1064-7481(22)00010-0.

Jeste DV, Lee EE, Palmer BW, Treichler EBH. Moving from humanities to sciences: a new model of wisdom fortified by sciences of neurobiology, medicine, and evolution. *Psychol Inq*. 2020;31:(2):134–143.

Kim SH, Park S. A meta-analysis of the correlates of successful aging in older adults. *Res Aging*. 2017;39(5):657–677.

Lam J, Aftab A, Lee E, Jeste D. Positive psychiatry interventions in geriatric mental health. *Curr Treat Options Psychiatry*. 2020;7(4):471–488.

Lee EE, Bangen KJ, Avanzino JA, et al. Outcomes of randomized clinical trials to enhance social, emotional, and spiritual components of wisdom: a systematic review and meta-analysis. *JAMA Psychiatry*. 2020;77(9):925–935.

Lohr JB, Chang H, Sexton M, Palmer BW. Allostatic load and the cannabinoid system: implications for the treatment of physiological abnormalities in posttraumatic stress disorder (PTSD). *CNS Spectrums*. 2020;25(6):743–749.

Nguyen TT, Jeste DV. Ageism: the brain strikes back. Cerebrum: emerging ideas in brain science. *Summer*. 2021:15–21.

Nguyen TT, Zhang X, Wu TC, et al. Association of loneliness and wisdom with gut microbial diversity and composition: an exploratory study. *Front Psychiatry*. 2021;12:648475.

Scheier MF, Carver CS. Dispositional optimism and physical health: a long look back, a quick look forward. *Am Psychol*. 2018;73(9):1082–1094.

Treichler EBH, Glorioso D, Lee EE, et al. A pragmatic trial of a group intervention in senior housing communities to increase resilience. *Int Psychogeriatr*. 2020;32(2):173–182.

Verdolini N, Vieta E. Resilience, prevention and positive psychiatry. *Acta Psychiatr Scand*. 2021;143(4):281–283.

56

Public Psychiatry

56.1 Public and Community Psychiatry

Nidal Moukaddam, M.D., Ph.D., and Asim A. Shah, M.D.

"I am proposing a new approach to mental illness and to mental retardation. This approach is designed, in large measure, to use Federal resources to stimulate state, local, and private action. When carried out, reliance on the cold mercy of custodial isolation will be supplanted by the open warmth of community concern and capability. Emphasis on prevention, treatment, and rehabilitation will be substituted for desultory interest in confining patients in an institution to wither away."—John F. Kennedy, Message from the President of the United States Relative to Mental Illness and Mental Retardation, February 5, 1963.

WHAT IS COMMUNITY PSYCHIATRY?

Community psychiatry, historically, blossomed and evolved in parallel with the push for deinstitutionalization and closure of large psychiatric facilities in the 60s and 70s. The overarching vision included geographically centered coordinated care for the most vulnerable and severely ill members of society. The target population typically suffers from triple serious illness: psychiatric, medical, and addictive, and often has a history of traumatic life experiences, compounding the treatment complexity. Based in public sectors, community psychiatry has evolved to encompass a diffuse network of treatment locations and modalities that diverge from its original character. Since the passage of the 2010 Patient Protection and Affordable Care Act (ACA), the experience of public, social, and community psychiatry has been transforming, with multiple changes in behavioral health care delivery and conceptualization. The term public can refer to psychiatric programs, treatment, or institutions paid for by public funds or as objects of public policy, whether paid for or not. Traditionally, community, or public psychiatry contrasted with private psychiatry, the latter being the entrepreneurial practice of delivering services to individuals having some means of purchasing them. As discussed below, expansion in inclusion criteria of recipients of community psychiatry has led to changes in its theoretical and practical scopes.

Funding for public and community psychiatric services tends to be provided by federally legislated appropriations that are passed through state, county, and municipal government agencies (such as departments of mental health; substance use treatment services; children, youth, and family services; and public health, social services, education, adult corrections, and juvenile justice agencies). Ultimately, most public psychiatry and community services are provided by not-for-profit community mental health, substance use treatment, child guidance, social services, or health care organizations that either are funded by or subcontracted to government agencies. Thus, the very existence of public and community psychiatric services and the policies and resources determining how they are delivered are extremely dependent on legislative mandates and fiscal appropriations from all levels of public government.

Psychiatry practitioners and administrators working in the public sector must be knowledgeable and savvy concerning the laws, regulations, and funding mechanisms relevant to their services. There is recognition of the prevalence of substance use disorders, both as primary disorders (often in combination with mental disorders, that is, “co-occurring” substance use and psychiatric disorders) and as secondary disorders (when mental disorder[s] are the condition of the first onset of primary concern, i.e., “comorbid” psychiatric and substance use disorders), and community psychiatry clinical workflows must reflect the various layers needed for effective treatment.

THE ROOTS OF COMMUNITY PSYCHIATRY

The precursor for community psychiatry, described below, was termed *boarding out*. As early as 1864, Sir Arthur Mitchell, in his work titled “The Insane in Private Dwellings,” discussed how mental health care did not have to be limited to asylums and how this expansion did not necessarily have to relate to limited asylum bed capacity, but rather a true ability to provide proper community-based care. The term community, as historically applied to psychiatry, connotes the mental health status of a specific, geographically defined population in a specific geographical catchment area. It is synonymous in all respects with “not the state hospital” and is contrasted to institutional care, especially in the move to community mental health in the second half of the 20th century.

The concept of “boarding out” and living with family despite chronic mental illness was tested in Scotland in the late 1800s as well. A major survey of individuals with mental illness in 1862, based on 7,000 home visits, revealed that the number of people living in asylums was about equal to the ones outside asylums and that homelessness was a major issue, even at the time. The other trend that Mitchell described, even in 1862, was that individuals with mental illness suffered often precipitous socioeconomic decline, sometimes affecting their entire families. These discoveries were in parallel with epidemiologic mortality surveys of individuals with psychiatric illness in the community and with mortality surveys. They are also echoed in modern literature that gave rise to community psychiatry: one example is the thought that individuals with chronic psychotic disorders such as schizophrenia become less employable and end up in lower socioeconomic classes than peers of comparable age. The social drift hypothesis was re-explored using population, twin, and molecular genetic data, with the conclusion that the downward sociofunctional drift was largely due to genetic, not environmental influences, though the role of "social defeat" and trigger events preceding first psychotic episodes, has long garnered interest as a factor in desensitization of the mesolimbic dopamine system. Regardless of the reasons for this drift, the community psychiatrist often encounters complex patients with limited employment histories and limited social functioning, reflecting the impairments of negative and cognitive domains of schizophrenia.

The concept of chronic mental illness being compatible with a noninstitutional life rhythm was espoused by writers such as Susser, who outlined modern applications of the concepts in the 1960s and the experiment of community psychiatry in Salford, Britain. The goal was, and still is, to treat mental illness on an outpatient basis and, by providing psychosocial support as well as psychiatric care, to aim for full integration of individuals with severe mental illness (SMI) in society.

The term psychiatry originally described the medical specialty focused on the care of persons with mental illness in the state hospital setting. Its domain was strengthened by the dominant role of the medical superintendents of the state hospitals (who would eventually be called "psychiatrists") and their medical colleagues in the latter half of the 19th century and the first half of the 20th century. However, in the era of community mental health in the latter half of the 20th century, the term exploded to cover a host of disciplines, and new personnel was recruited to staff the community-based supportive treatment systems that were supposed to take the place of the institutions. Nursing, social work, clinical psychology, various therapeutic specialties, and non-degreed direct care providers all became labeled as part of psychiatry. One of the most dramatic changes emerging since the late 1990s is the role of consumers of psychiatric services as providers in their own right through networks of peer-support services and advocacy initiatives. Family members of people with these disorders are also increasingly taking a more assertive role in treatment and support.

Different phases are noted in how mental illness was viewed in the United States: Colonial times and the early years of the Republic were the era of the poorhouse, during which the great majority of persons with mental illness found refuge or were confined for their own and others' protection as part of local community management. Care and custody were the operative functions, and treatment was nonexistent. In the 18th century, the ideals of the Enlightenment formed the basis for a paradigm shift toward fostering a better life for persons with mental illness rather than condemning and isolating them, ushering in an era of “moral treatment” (aimed at restoring the capacity for ethical conduct) through a return to the pastoral rural environment of the preindustrial age (“the Retreat”). In the mid-19th century, the protection and moral rehabilitation of the mentally ill was scaled up and institutionalized through the pioneering work of Dorothea Dix, resulting in an age of the asylum, during which public inpatient facilities, generically called state hospitals in the United States, were constructed in every state. A century later, advances in medicine, psychodynamic psychiatry, biologic psychiatry, and cognitive-behavioral psychology, along with a remarkable policy consensus, moved federal legislation to shift from institutionally based care to an era of community mental health. Finally, as health care costs began to soar at the end of the 20th century, and as employer purchasers and government moved to create market-driven health care, mental health policy moved conceptually to tightly organized systems of care that focused on prevention, early intervention, and intensive community care for children with severe emotional disturbances and their families and community support and recovery for adults with serious, acute, and persistent mental illnesses. The policy and concept, although ideal, remain partially operational with the smooth provision of care constantly jostled by bureaucratic, regulatory, and economic potholes.

In the American colonies and in the early days of the Republic, the care of the mentally ill was modeled on the practices of 18th-century England, where the Enclosure Laws disrupted the rural social order and set thousands on the road to seek shelter and employment in the cities. People with mental disabilities were at a particular disadvantage. They were saved from vagrancy by confinement in poorhouses, prisons, and jails, along with criminals, debtors, and other people with disabilities. In the American colonies, most towns would establish a poorhouse, and persons with mental illness, especially those who were found to be disruptive, would be committed by the actions of local government to extended stays in them. It is not that there were no hospitals. The construction of a public hospital was first proposed in Virginia in 1766. The money was appropriated in 1770, and the Public Hospital at Williamsburg opened in 1773, which was the first public mental hospital in the colonies. In Massachusetts, McLean Asylum opened in Charlestown in 1818. The first large state hospital opened in Worcester, Massachusetts, in 1833.

The physician superintendents of these early hospitals were the reformers of the era, following the ideas of moral management put forth by the French reformer Philippe Pinel. Pinel believed that if people were removed from their urban environments, relocated to an asylum in a beautiful pastoral setting, engaged in agriculture and the trades, lived in close proximity with staff, and were treated with kindness, then soon they would fully recover from the behavioral changes that they manifested as a result of having lived in a stressful, demanding environment. Benjamin Rush, a signer of the Declaration of Independence and considered the founder of American psychiatry, applied many of these theories and methods of his own in the treatment of the insane in the years after 1783 in the Philadelphia Hospital.

The great majority of those who experienced mental illness, however, were kept in poorhouses or jails under deplorable conditions and became a cause for reform. The Unitarian reformer, Dorothea Dix, had begun a 2-year survey of the conditions in which these people lived. In 1843, she compiled her findings in a Memorial to the Legislature of Massachusetts: “I tell what I have seen, to present the state of the insane persons confined within the Commonwealth in cages, closets, cellars, stalls, pens! Chained, naked, beaten with rods, and lashed into obedience!”

After the Massachusetts Legislature appropriated funds for a substantial expansion of the Worcester Hospital between 1844 and 1854, Dix traveled more than 30,000 miles, visiting hundreds of prisons, almshouses, jails, and hospitals, collecting data, and arousing public opinion to her cause. In 1848, with the example of federal support given to education, she sought federal support for national hospitals. President Franklin Pierce, who stated, “I cannot find any authority in the Constitution that makes the Federal Government the great almoner of public charity throughout the United States,” and vetoed the bill for federal support. Public mental health was, for the next century, exclusively in the hands of the states and the state hospitals that continued to be built only to be quickly filled and become hopelessly overcrowded.

Early American Psychiatry and Ties to the Community

It has been said that U.S. medicine reached a critical juncture toward the beginning of the 20th century. Science and medicine had established a strong alliance in the major university medical schools, and the profession had established powerful public credibility as breakthroughs transformed the practice of medicine. This has started the era of evidence-based practice. Psychiatrists were not at the forefront of this transformation, but they were profoundly affected by it. One of the most prominent psychiatrists of the period, was Adolf Meyer, and his life and career may be thought of as representative of what was happening to psychiatry. Meyer was born in Switzerland in 1866 and was trained as a neurologist. He came to the United States in 1892 and worked for 8 years in state hospitals in Illinois and Massachusetts. In this sense, he came from the ranks of the asylum psychiatrists who were the bulwark of the state hospital system that had, by that point, spread throughout the country.

The first public psychiatrists were state hospital superintendents, responsible for all institutional clinical and administrative functions. Within a few short years, the populations of state hospitals grew so that their treatment function was eclipsed by care and custody concerns. The superintendent became more of a manager and administrator than a treatment provider. The ability to organize and to bring order to the institution became a necessary skill.

The New York State Hospital system had established, in 1895, the Pathological Institute, which had two mandates: the study of mental diseases from the standpoint of cellular biology and the offer of instruction in brain pathology to state hospital physicians. The Institute sought to become the cutting edge of reform—not without opposition from the asylum psychiatrists, who were more oriented toward managerial and administrative concerns and clear demarcation of the diseased state of institutional inmates from the normality of functioning people in the community. Meyer became its director in 1901. When Henry Phipps endowed a psychiatric clinic at Johns Hopkins, Meyer was chosen as its first head. From Johns Hopkins, he played a leading role in American psychiatry for more than three decades.

Twentieth Century U.S.

At the beginning of the 20th century, the concept of dynamic psychiatry opened the possibility of psychiatry practiced in the community. Dynamic psychiatry blurred the distinction between health and disease and saw mental illness on a continuum. This allowed the center of attention for psychiatric practice to move beyond the institutions and to include troubled individuals in the early stages of illness. This approach elevated the significance of the life history and prior experience of each person. It was a short leap of faith to suggest that early community treatment might prevent the onset of SMIs that, up to then, had required institutionalization. The once strong bond between the asylum and psychiatry was unraveling. Dynamic psychiatry expanded the jurisdictional boundaries to include psychologically troubled individuals, as well as allegedly dysfunctional social structures and relationships. From 1880 to 1940, psychiatrists identified new careers outside of institutions. Psychiatrists articulated novel and sometimes bizarre theories and therapies; expanded jurisdictional boundaries to include not only mental disorders but also the problems of everyday life; proposed solutions for society’s ills; and defined a preventive role. In retrospect, it was an over-reach that subsequently led to questions about the credibility of the field and its foundations. What had been hypotheses initially were interpreted by some as factual and often damaging (e.g., the so-called “schizophrenogenic mother,” asthma as “a cry for help,” etc.). It was also during this “pre-chlorpromazine era” that psychiatry alienated itself from the rest of medicine. Despite its good intentions, it was, in retrospect, gross overstatement based more on conjecture and opinion than a scientific basis.

Meyer was a leading figure in advocating “a fresh and courageous pragmatism” and held “a biological conception” of human beings: “We must…accept the statement that all mental activity must have its physiological side and its anatomical substratum.” He developed a genetic-dynamic or developmental approach to mental illness, which he named psychobiology. He stressed the interaction of the organism with its environment and defined mental disorders in behavioral terms, tracing their origins to defective habits. He recognized the importance of consumer advocacy and supported Clifford Beers, a former mental patient, who wrote a celebrated work, A Mind That Found Itself, and who founded the Mental Hygiene movement.

Meyer was an empiricist, urging his colleagues to assemble and catalog facts. He was not a theoretician or a systematizer. Perhaps he realized that biology did not yet have the scientific tools to construct adequate theories about human mental processes. In this regard, he did not favor Sigmund Freud and what he viewed as the tendencies to dogmatic pronouncement. The eminent social historian of American mental health policy, Gerald Grob, observed that "psychoanalytic concepts…had far greater influence on culture and thought than on medicine in general or psychiatry in particular." This seems to overstate the case, given the great enthusiasm for psychoanalysis that came after World War II and that dominated popular perceptions of psychiatry.

On the eve of World War II, as Meyer retired, public psychiatry had advanced on three fronts. First, research institutes and psychopathic hospitals were developed alongside the state asylums in which research and training would benefit the care of persons with mental illness. These hospitals and institutes had limited success in that their research efforts to establish the efficacy of psychodynamic treatment progressed slowly, and they suffered a degree of mission drift away from the seriously and persistently mentally ill in the asylums and toward less severe mental health problems in the community. Second, an active consumer advocacy movement began in the founding of the National Committee for Mental Hygiene in 1909, which developed into a national movement focusing first on the plight of individuals in the asylums and then on mental health issues in the community. Finally, the ideas of dynamic psychiatry made it possible to think of the various stages of mental illness and the possibility of the prevention, amelioration, and potentially even cure of mental disease.

The 1950s were a time of great therapeutic optimism. Biologic psychiatrists pointed to the efficacy of drug therapy; their psychodynamic colleagues could point to psychological, psychosocial, and environmental therapies that would reach the most severely disabled institutionalized patients. Milieu therapy (the therapeutic community) arose in this period and had much in common with moral therapy. It appeared that medications were going to allow patients to engage in the therapeutic community, and then other innovations, such as unlocked wards and partial or day hospitals, were going to help some chronic patients transfer to the community.

As all of this momentum for change was building, a remarkable partnership for political action was developing among Robert Felix, the head of the newly formed National Institute for Mental Health, Senator Lister Hill, and Representative J. Percy Priest. Hearings were held aimed at focusing public attention on mental health problems and thus strengthening the case for an enhanced federal role. The Mental Health Study Act was passed in 1955, creating the Joint Commission on Mental Illness and Mental Health (JCMIMH). That Commission's final report, Action for Mental Health, did not make a clear-cut distinction between individuals with serious and persistent mental illness and psychologically troubled individuals. Although the JCMIMH report was a compromise among the 26 participating constituency organizations, nonetheless, there was strong support for fundamental change. Faith in the effectiveness of treatment in community programs was based on belief in the efficacy of early intervention, psychodynamic theories, epidemiologic efforts to elucidate the prevalence of mental illness in the community, and the putative identification of environmental variables.

At the end of the decade, with a new administration with an aggressive domestic agenda, the National Institute of Mental Health (NIMH) convened the Interagency Task Force on Mental Health, which developed the Blueprint for Action and, more importantly, activated the political coalition that made President Kennedy's initiative possible. The final policy sidestepped the provisions recommended by the JCMIMH for the refurbishment of the state hospitals and the provision of acute institutional care for the most severely disabled and devoted its attention to community mental health clinics, which it recommended to be enhanced to comprehensive community mental health treatment centers—the community mental health centers (CMHCs).

The national program for comprehensive CMHCs was in place. For the next 15 years, grants flowed to the localities and then through the states to the localities to the new centers, and states proceeded to build two community systems while beginning the process of phasing out the old state hospitals. The immediate discovery was that the CMHCs were able to expand mental health services for the communities but were not at all prepared to care for persons with serious and persistent mental illness who had spent most of their lives in state institutions. The discharge of 90% of the state hospital population (the deinstitutionalization process) without adequate community resources available was chaotic and often precipitous. It led to a class of homeless and generally untreated or transinstitutionalized individuals (put into nursing homes, prisons, squalid rooming houses, and hotels, etc.) who deteriorated outside the state hospitals. There also arose a peripatetic subpopulation with serious and persistent mental illness who often had comorbid substance abuse and generally avoided any beneficial involvement in treatment for a significant portion of the early phase of their illness. However, the states, to varying degrees, were set on a course of dismantling state hospitals that were becoming increasingly expensive. As federal staffing grants to the CMHCs were phased out and state contract dollars replaced them, the states gradually narrowed the focus of the community programs to its priority clients—the former residents of the state hospitals. Special programming models were developed to provide supportive programming for them that focused on community support and rehabilitative approaches rather than clinical ones.

As community mental health programs developed, it became clear that psychiatrists were, for the most part, playing limited roles: reimbursement could not support them full time, and they gravitated to private practice models that were appropriate for individuals with less severe mental health issues. The psychiatrists who were involved in the early CMHCs were quickly outnumbered by practitioners from the other disciplines and often sidelined themselves in private practice models in which they could treat educated and functional people with relatively mild disorders. One tongue-in-cheek publication during this era described the “YAVIS” individual as the “ideal” psychotherapy candidate…Young, Attractive, Verbal, Insightful, Successful. Many psychiatrists and others in behavioral health gravitated toward this profile for their ideal caseloads. These practice models limited their ability to treat persons with severe mental health disabilities without appropriate outreach and supportive care.

ORGANIZED SYSTEMS OF CARE

The psychiatric hospital is a special target of health care corporations. Its profit margins are large, investment costs are low, inpatient psychiatric care is now widely insured, and the complexity (and perhaps ambiguity) of psychiatric diagnosis and treatment makes cost control efforts by insurers and the government difficult. In 1968, for-profit hospital chains were nonexistent in psychiatric care. In 1980, investor chains owned 25% of private psychiatric hospitals, and, just 2 years later, these corporations owned a full 43% market share. Between 1982 and 1983, the number of investor-owned psychiatric hospitals rose 42.5%, from 106 to 151. The increase was partly attributable to investor groups’ acquisition of existing facilities, but the new construction of psychiatric hospitals, which grew 53.3%, accounted for most of the growth. “Ad-Psych” (Adolescent Psychiatry) units are so lucrative that corporations have closed units providing other types of services in order to establish the more profitable adolescent units. Admission data reflect the current domination of the private sector in the provision of adolescent inpatient psychiatric services. Lois A. Weithorn, in the Stanford Law Review, noted that in 1980, for example, private facilities admitted 61% of all adolescents who underwent psychiatric hospitalization; since that time, juvenile admissions to these facilities have increased substantially.

U.S. health care costs began to increase dramatically in the 1970s. By the end of the 1980s, they were spiraling. The article just cited by Lois Weithorn in the Stanford Law Review addressed the possible violation of children’s rights posed by their involuntary hospitalization in commercially operated psychiatric hospitals. It documented a trend of which U.S. employers—and the insurance companies that they hired to administer their health benefits—were all too aware. While employee health care costs overall were climbing, behavioral health expenditures were increasing by 50% each year. Venture capital had discovered a chink in the armor of health care finance. Even though there were lifetime limits for mental health benefits, a substantial amount of money lay in those coffers. Psychiatric hospitals were relatively cheap to build and could be designed to be converted to older adult housing facilities once the gap was closed. The criteria for mental illness were sufficiently permeable, and standards of care were sufficiently undefined that decisions for hospitalizations were made on behalf of many patients when more considered judgments would have found other alternatives. In fact, many hospitals launched deliberate marketing campaigns to fill their beds, and many practitioners participated. The trend was an abuse. Its most cynical aspect was that benefit dollars that could have been devoted to the care of persons with SMI were quickly extracted—a $50,000 lifetime limit could be spent in 60 days of private hospital care—so that sick patients were discharged to the care of public sector agencies.

Although several examples of health maintenance organizations (HMOs) had demonstrated their effectiveness in improving the health status of enrolled populations for several decades, U.S. employers were reluctant to require their employees to sign up for a plan instead of the traditional indemnity (or fee-for-service) health insurance. However, the growth of behavioral health expenditures was so dramatic that they overcame their reluctance for this sector. Within a few years, an industry of managed behavioral health organizations was created by the contracting activities of human resources executives of the major U.S. corporations. The new companies were staffed by a corps of seasoned mental health professionals, most of whom had been prepared in the ranks of the community mental health movement. They introduced, in a business-like fashion, utilization review techniques reinforced by elaborate information and communication technology. They selected panels of providers who could be trusted to care for enrolled members without inappropriate or extended use of hospital or other expensive care. Their impact was dramatic: Within a few years, behavioral health expenditures stabilized, at least temporarily.

The success of the carve-out of behavioral health care inspired employers to apply their contracting techniques to the rest of health care. Even though in 1987, only 27% were in managed care, by 1995, 73% of insured employees were enrolled in some form of managed care. Simultaneously, in the early 1990s, President Bill Clinton undertook a health care reform that attempted to provide universal health care as a highly regulated public entitlement. The Clinton health plan was labeled “managed competition” as a nod to the rapid move to competitive health care markets. However, the originators of that term—an assembly of business and health care experts known as the Jackson Hole Group—were much clearer about an approach in which government and government regulation would play a minimal role.

Nonetheless, a health care marketplace was created in the course of the failed Clinton health care reform. A wide variety of health care entrepreneurs would organize providers to bring their products to the markets, and consumers were encouraged to make choices that would ensure the value of the care that they would receive. The employers, through their contracting practices, sought to apply the principles of industrial organization to health care delivery to advance the ideal of organized systems of health care that were accountable and attractive to value-driven consumers. Although certainly, some excellent health care delivery systems have developed, most have fallen far short of these ideals in a full array of organizational structures. All of these structures and models fell under the term “managed care.”

In the early 1990s, managed care's utilization review techniques and benefits had restrained providers; HMOs and other plan types restructured health care delivery; the insurance industry transformed itself, acquiring HMOs and creating a new array of "products" for employers to purchase; risk was cascaded to hospitals, in many cases making them insurers and encouraging them to consolidate their positions in physician hospital organizations (PHOs); and physicians reorganized their practices into independent practice associations (IPAs) and multi- and single-specialty groups. By the mid-1990s, the rise of health care costs leveled off (temporarily), but there was widespread dissatisfaction among consumers and the rest of the system. By 2000, many of these measures were relaxed. Instead of benefits that restrained professionals, copays for consumers restrained their utilization. Employers encouraged broader networks, preferred provider organizations, and eliminated gatekeeper access to specialists. The insurance industry consolidated itself through cross-market mergers and large national contracts.

Quality and accountability began to dominate the health care market by the end of the 1990s, and consumers were encouraged to look to quality ratings by major accreditation organizations in making their choices. Managed care has had an indelible impact on the U.S. health care system, introducing issues of quality and accountability, reinforced by the Institute of Medicine's report Crossing the Quality Chasm.

How Community Psychiatry Became Its Current Self

As the process of deinstitutionalization proceeded and patients were moved from state hospitals to care in community programs, the financing also moved from institutional budgets to the categorical budgets of the state mental health agencies to fund the community programs. As employers moved dramatically to managed care in the 1990s, however, state Medicaid agencies followed suit and contracted with the Behavioral Health Managed Care Organizations (BHMCOs) to manage their mental health benefit. States quickly realized that by moving mental health categorical dollars into their Medicaid program, they could leverage Federal Financial Participation (FFP) into their state's general fund. Almost all of the states shifted large portions of their mental health budgets for direct services into the Medicaid program—hence the substantial increases in mental health funding. However, these services ended up in the stewardship of Managed Behavioral Health Organizations (MBHOs), whose products tended to be modeled on private-sector mental health care for general populations. Currently, patients are diffuse in communities, with the fragmentation of U.S. social welfare programs that is reflected in disparate payment systems.

Three contemporary trends have influenced the conceptualization of public and community psychiatry in the daunting climate facing U.S. health care. The first is the realization that the effect and pervasiveness of psychiatric disorders impose a global burden on society. The second trend has been the shift in funding for public psychiatric services from state expenditures, tied to the historic funding for state hospitals, to Medicaid, a health insurance model driven by the “medical necessity” for intervention. Finally, the power of reform efforts in general medical care and psychiatry have led to increasing demand for the use of evidence-based interventions defining quality, safety standards, and efficiency in the delivery of health care.

The first trend—the recognition of the burden of these disorders—can be most directly linked to a landmark study called the Global Burden of Disease. This study, funded by the World Health Organization (WHO), the World Bank, and several foundations, considered the leading causes of disability and premature death worldwide as a result of various health and situational problems. Of the ten leading causes of disability worldwide, five are neuropsychiatric, with the leading cause of disability being depression. When considering the leading causes of healthy years of life lost using the concept of disability-adjusted life years (DALYs), which reflects premature death and disability, depression was found to be in the top 3 in the world and in the United States. Mental health disorders were the fourth most costly health problem in the United States in 2000, 2004, and have only gotten worse in recent years. Schizophrenia, though not a highly prevalent disease, afflicts a substantial burden. When the cost to society of trauma-related disorders (which include post-traumatic stress disorder [PTSD]) are added in—and exposure to traumatic stressors has been found to play a role in both causing and moreover severely worsening most psychiatric disorders—the cost to the United States is ~50% more than the annual cost of heart disease and almost double the annual cost of cancer. New WHO data indicate that rates of various mental illnesses are increased to 22.1% in areas with war conflict (depression, anxiety, PTSD, bipolar disorder, and schizophrenia).

Second, after a period of critical evaluation of U.S. medical practice, the IOM, in its 2001 report entitled Crossing the Quality Chasm: A New Health System for the 21st Century, brought much clarity. This is a call to American medicine to reform inadequate service delivery systems and to apply effective practices systematically. It documented the need for a fundamental overhaul of the entire U.S. health care system, addresses the problems caused by insufficient quality in the health care system, and offers a road map for creating a new health system. A key principle in enhancing the quality of health care services is that treatments and practices must have a clinical and scientific evidence base in order to justify their utilization and costs. The Quality Chasm report was a primary influence among a series of reports from the IOM and other authoritative agencies in the two decades that laid the foundations for the health reform that has been codified in the ACA (2010).

Third, ACA-related changes to financing brought about an avalanche of changes in community mental health provision. The role of Medicaid in shaping public services for people with mental illnesses is difficult to overstate. For state governments, the appeal was almost immediate. When the states assumed the onus of care for people with mental illnesses, largely through the mechanism of state institutions, 100% of those expenditures came from state budgets—competing with all of the other priorities facing state legislatures. Medicaid, which is a state-federal partnership, allows states to shift between 50% to 80% of their cost for covered services to the federal budget. More recent federal laws and regulations mandating parity in eligibility and reimbursement for mental health (compared to medical/physical health) have attempted to direct more services to more individuals with mental illnesses. The Triple Aim: Better Care, Better Health, Lower Costs, has been the goal all along. Better Care refers to evidence-based psychiatry, integrated with primary care, in the context of coherent health care delivery systems.

The developments leading up to the ACA, as well as the statute itself, have laid the groundwork for the agenda of improving health care for public and community psychiatry. This is to be accomplished by CMHCs applying for status as Community Health Centers, or Federally Qualifies Health Centers (FQHCs), as well as strengthening the behavioral health services in the existing FQHCs. Fidelity to research-validated practices in implementing treatment and replication as the therapeutic model is aimed for, with some funding funneled through the Patient-Centered Outcomes Research Institute (PCORI). The ACA included the concept of a “Primary Care Residency” program to be based in community-based practices as Teaching Health Centers, including the FQHCs and CMHCs. Psychiatry is included as a specialty in those programs.

THE RECIPIENTS OF COMMUNITY PSYCHIATRIC CARE

Federally Qualified Health Centers (FQHCs) deliver care to the underserved populations in the United States. According to the Health Resources and Services Administration (HRSA), the health center program has 255,000 full-time staff and serves nearly 29 million patients. The number of patients served has tripled since 2000. More than half of FQHC patients belong to racial and ethnic minority groups, and almost 90% live in poverty, according to HRSA data.

At its inception, community psychiatry aimed to help communities as a whole, but over the years, much operationalization of what qualifies for treatment in community settings has taken place. Initial selection criteria reflected the population that was admitted to institutions and seeking care. Subsequent cost-cutting measures and refinement of workflows and protocols led to many restrictions but also much heterogeneity. The basic definition of serious mental illness aims to include the most vulnerable populations. In the Unied States, the term ***serious mental illness*** (SMI) originated from the National Plan for the Chronically Mentally Ill, a product of the President’s Commission on Mental Health (1978). SMI typically involves symptom severity and functional disability. Attempts to refine the concept of SMI resulted in redefinition by Substance Abuse and Mental Health Services Administration, published in the ADAMHA Reorganization

Act, Public Law 102–321, 1992, but states are free to use modified definitions, and more than 20 of them have adopted their own definitions. SMI definitions vary in Europe as well. Those definitions shape the funding received by organizations delivering community psychiatry in various states. The only common element of all SMI definitions is a psychiatric diagnosis, but the nature of the psychiatric disorder, disability, and duration vary by location. Psychotic disorders appear inclusionary in reviewed criteria, with mood disorders (major depressive and bipolar disorders) being the second most common category included in SMI classifications. Personality disorders are deemed by many to meet disability and duration criteria for SMI but are not accepted as entry diagnoses in some localities. This heterogeneity could have significant impacts on the purity and consistency of psychiatric nosology used in communities. This may lead to erroneous or exaggerated diagnoses of psychotic disorders, mood disorders or omission of substance-induced disorders, neurocognitive disorders, and organic causes of mental illness.

Financial Thresholds

Financial threshold to define recipients of community psychiatry services have also been used. Over the past two decades, entry into community programs has been linked to dual criteria incorporating both diagnosis and financial status. Recipients of community services often do not meet criteria for Medicaid, but after the ACA-related Medicaid expansion, some were able to get coverage; thus, coverage for community psychiatry now includes a spectrum of funding from uninsured, 100% to 138% of poverty/ eligible for subsidized private coverage, and Medicaid-eligible. Medicaid-only enrollees can be categorized as foster care young adults (ages up through 21), very–low-income adult parents, low-income peripartum women, or Supplemental Security Income (SSI) recipients. Income below 138% of the federal poverty level, meeting SMI criteria, and uninsured status are often needed to benefit from indigent resources. *Thus, community psychiatry now serves a larger swath of the population than it did at its inception in the 1960s*. It now serves individuals with SMI, but also individuals one could think of as the “working poor.” This is the group of workers who purchase health insurance because they do not receive health benefits from their employer, earn too much to qualify for Medicaid, and do not qualify for Medicare. With this expansion, the kinds of services needed and the spectrum of issues encountered become more variable, not limited to SMI, and therefore need a careful reassessment of the face and identity of community psychiatry.

THE ROLE OF THE PSYCHIATRIST IN COMMUNITY PSYCHIATRY

The psychiatrist, in their role as the medical specialist, trained in the discipline of medical sciences, had a very narrow focus, exercising medical and legal authority to take defined actions on behalf of individual patients. In the community psychiatry paradigm, the public psychiatrist practices psychiatry in publicly funded services and carries out programs that are mandated by public policy, interfaces with other team members, and is the team leader in the care of the patient with mental illness. As a public and community psychiatrist, they are engaged with a specific geographically defined population using community programs for the care of individuals with mental illness in order to improve the health of the community as broadly defined.

Psychiatrists in the public sector should work within a team that includes professionals from several disciplines (e.g., psychology, social work, nursing, occupational therapy, rehabilitation or addiction counseling, social services, housing, and employment specialists) and non-degreed direct care workers (e.g., bachelor's degree–level counselors or case managers, high school graduate, indigenous outreach workers, peer-support specialists, and family advocates), each of whom brings unique skills and experience to address the varied needs of people with severe and persistent mental illnesses. The team, rather than any single individual, assumes responsibility for the ongoing care of each patient across the many levels of services and often for many years. Its success is based on effective communication and coordination. Communication and coordination should be explicitly focused on the individual's (i.e., patient's and/or family's) stated goals, as well as on the community resources available and the team's technical and logistical capacities. Maximizing the individual's motivation to participate actively and productively in a comprehensive treatment plan that is patient-centered and collaborative is essential.

Psychiatrists can play four primary roles in multidisciplinary psychiatric rehabilitation teams: (1) conducting psychiatric diagnostic and treatment evaluations with significant involvement from members of the team; (2) providing thoughtful pharmacotherapy; (3) coordinating care with physical medicine in order to promote both physical and mental health; and (4) serving as the team's medical director, mentor, sometimes clinical supervisor, peer, and colleague.

The psychiatrist is being called upon to play a special role as the project or program medical director—with the attendant responsibilities of monitoring the medical safety and well-being of all patients and establishing or supporting management and clinical procedures that support quality of care and a cohesive treatment team. As the team's clinical Mentor, or simply as a team member, the psychiatrist serves as a role model for compassionate and professional behavior in relation not only to patients but also to all other team members. Formal or informal leadership is especially important when psychiatry residents and medical students work on a team as a training experience. The team psychiatrist serves not only as a mentor and a role model for the core aspects of psychiatric practice but also should demonstrate the values and skills necessary to integrate psychiatry within the framework of psychosocial interventions that are multimodal, community-based, and longitudinal.

The connotation associated with public and community psychiatry has been synonymous with care for the poor and underprivileged, care for special populations normally not attended to by the private sector, overwhelmed systems, low reimbursement and inadequate resources, inferior and unresponsive care, and a final common pathway for individuals no longer able to afford private care. This is not for want of their providers' dedication: it is the paucity of the physicians' numbers and the general public neglect of this sector. The patients of public and community psychiatry as a group lack the most basic resources to sustain themselves (food, clothing, housing, financial support, social support, etc.), and they are arguably the group least likely to be able to advocate for themselves and to have the capacity to recover from this default position. The needs of the patients, the inadequacies of the delivery systems, the lack of an adequate workforce, and the lack of coverage make the sector of public and community psychiatry a poster child for the attention of the ACA reforms. It also makes the practice of community psychiatry both noble and challenging. The shift from a narrow focus on the treatment of a solo therapist to a broader psychiatric rehabilitation model, and now to models that stress the recovery of the individual, requires solid training (and continuing education) and hard work, as well as a combination of personal qualities, competencies, and expertise that have been summarized as "persistence, realistic optimism, a collaborative style, teaching ability, and networking and system coordination skills."

A major challenge to psychiatrists serving adults who have the dual burden of mental illness and socioeconomic disadvantage is the identification and enhancement of the psychological and sociocultural resources that can promote illness management and reintegration in a support system. The goal, of course, is recovery with the help of fully operational programs of assisting adults with severe and persistent mental illness to make the best possible adjustment in the least restrictive public and community settings.

THE INPATIENT TO OUTPATIENT AND COMMUNITY PSYCHIATRY CONTINUUM

Community psychiatry best exists in a system of care that attends to the full needs of persons with psychiatric illness. This need exists given the complexity and high frequency of co-occurring disorders in the patients typically served by community psychiatry teams. This would include inpatient care, emergency psychiatric services, substance use treatment modalities, social services, and functional rehabilitation support.

The ideal setup for treatment is that majority of treatment gets conducted on an outpatient basis, with flare-ups and crises being handled in transitional intensive programs and partial day hospitals, and for severe symptoms, in inpatient psychiatric units for short stays. Partial hospitalizations were found to be equivalent to inpatient care in meta-analysis. The advantage of preponderant outpatient treatment is both cost-driven but also mindful of the need to keep patients in their natural environment. An essential component for the success of community psychiatry is a linkage between various service components, so patients do not end up needing higher levels of care inappropriately. When patients are released, the cohesiveness of the conditions of release and careful considerations for re-entry in the community are essential to avoid re-hospitalization or treatment nonadherence. This area has had repeated formulas tried for the last two centuries, dating back to the state of Massachusetts in 1897. Economic drivers of decreased length of acute hospitalization overlay other factors that had driven community and social psychiatry to the forefront of psychiatry, namely optimistic beliefs that improved treatment options would lead to less need for acute care and that options for other conditions, such as addictive disorders, are available. Trends for treatment, however, show low usage of anticravings drugs and pharmacologic options for addictive disorders and poor treatment adherence to antipsychotics following hospitalizations.

The reality, however, is that there is a shortage of outpatient psychiatrists, a shortage of hospital beds, and many patients, up to 20%, awaiting hospitalization end up enduring emergency center boarding. Boarding affects patients with mental illness disproportionately compared to those with medical illnesses. Boarding is noted to be longer for patients with co-occurring illnesses, intellectual disabilities, aggression, and lack of insurance funding, all characteristics that tend to be more prevalent in community psychiatry samples. Once hospitalized, these patients have longer lengths of stays, and this is predicted by severity of illness, psychosis or bipolar disorder, polypharmacy, and need for emergency restraints/interventions.

The majority of caregivers for individuals with mental illness are family members, placing strain on communities. Care for a significant number of these individuals is poorly coordinated, and there is a recurring whorl of decompensation, emergency room visits, hospitalizations, various outpatient and sometimes residential programs, variable compliance and funding struggles with treatment, diagnostic confusion (e.g., schizophrenia vs. bipolar disorder, borderline personality vs. PTSD vs. depression, etc.), polypharmacy with lack of attention to guidelines or to metabolic consequences leading to the iatrogenic production of physical health problems for these individuals who then enter the cycle multiple times. Involvement with the judicial system is often part of the course. Thus, more recent focused attempts in community psychiatry have focused on jail diversion programs and better interface with the legal system. Additionally, within the legal system, there is a push to maintain treatment continuity and ensuring inmates receive psychiatric care consistently.

Case management remains a cornerstone of community psychiatry. The roles of the case manager include establishing rapport, coordinating with families, and connecting the patient to community resources. Some definitions of case management work are more holistic and focus on comprehensive care, with some models requiring case managers to work on social skills training and psychoeducation, for instance. The role of the case manager will also depend on the treatment model, with more intensive involvement, for instance, in ACT teams: in a comparison on ACT versus the standard care model, the intensity and frequency of visits varied, with 13 visits on average for the ACT team in the first 3 months, but 5 in the standard care model, and this difference reflects in the panel that a case manager can carry.

STIGMA AND DISCRIMINATION–AN OPPORTUNITY FOR COMMUNITY PSYCHIATRY

The Relationship Between Community Psychiatry and Health Maintenance

Community psychiatry has been evolving and adapting as psychiatry, as a discipline itself evolves. The worst-case scenario is the evolution of community and social psychiatry aspects to mere delivery of psychiatric services, with accompanying reductionism in diagnosis, treatment options, and views of mental illness. But this evolution also carries the potential to truly help our most vulnerable populations. The establishment of community psychiatry infrastructure and network can help further other needs of the population, including medical, substance use, and social. In that scope, community psychiatry, a branch of social psychiatry, can also serve as a springboard for wider wellness and theoretical, structural changes in health care by linking the social, psychological, and biologic spheres. In fact, with changes in managed care reimbursement and restrictions imposed by newer forms of health care, it is suggested that private psychiatry is starting to veer toward the models initially pioneered by community psychiatry. The historical separation of psychiatry from the rest of medicine, partly self-imposed and partly extruded by mainstream medicine, will hopefully be corrected as delivery systems change and as greater knowledge comes to influence more focused and effective interdisciplinary efforts.

The elements that make community psychiatry structures (in its more ideal form) a desirable model are: community psychiatry often involves teamwork with physicians, mid-level providers, nursing and case management, social work staff. These teams often interact with the legal system and families and (again, in ideal circumstances) with the patients’ medical teams. Additionally, community psychiatry often interacts with and serves racial and ethnic minorities. This offers the multidisciplinary interactions our patients need. It is intuitive to match wellness initiatives, educational initiatives, and medical monitoring and social services measures to existing multidisciplinary infrastructures. HRSA aims for coordinated, comprehensive, and patient-centered care, integrating medical, dental, behavioral health, substance use disorder, and patient services. Individuals with SMI (schizophrenia, bipolar disorder, major depressive disorder) are at increased risk of cardiovascular disease-related death versus controls, as evidenced by a meta-analysis of 92 studies totaling more than 1 million individuals. Risk factors for mortality include antipsychotic use, elevated body mass index. Additionally, other types of diseases such as infectious, endocrine/nutritional/ metabolic, circulatory, respiratory were elevated in all SMI categories (including borderline personality disorder). Excess mortality from cancer is present in SMI as well. It is estimated that individuals with SMI die 8 to 10 years earlier than age-matched cohorts, often from natural causes. Therefore, it is essential that care systems combine as many interventions and preventive services as possible in order to truly help patients who often lack self-efficacy to mitigate their own problems. Integrative care, combining medical, social, and psychiatric, and improving treatment adherence to antidepressant and antipsychotic regimens may have some protective effect on premature mortality.

An example of the interface of medical and psychiatric management in the goal of comprehensive care for patients is the Delivery System Reform Incentive Payment (DSRIP).

DSRIP represents system transformation aiming to improve clinical management across settings, aiming for larger-scale population health improvement and, ultimately, reduction in avoidable hospital use. In Texas, DSRIP was designed as part of Medicaid waiver, §1115 of the Social Security Act. Nested in the 2011 Texas Healthcare Transformation Quality Improvement Program, DSRIP measures were spread and planned through regional health care partnerships (RHPs). It targets Medicaid enrollees and low-income uninsured individuals and includes regional assessments, goals/rationale for projects/expected results/annual milestones, and metrics. DSRIP aims to transform care delivery by integrating systems of care, delivering high-quality care that meets guidelines, and ensuring patients get prevention and primary care services. An obstacle to this, of course, is access, so one of the aims is to increase patient access, provide patients with a positive health care experience, and offer timely, proactive, coordinated medical home care.

DSRIP applies to mental health, which has a high chronic disease burden and extensive need for care navigation upon hospital discharge. This leads to unmet mental and behavioral health needs, which often get compounded by health care provider shortage and access to care. Some of the DSRIP mandates include follow-Up after Hospitalization for Mental Illness (FUH) within 7, then 30 days, and initiation of Alcohol and Other Drug Dependence Treatment on a continuum with hospitalization. Health systems have a choice in what measure to implement for their populations as well within the DSRIP context. The example of the Harris Health System in Harris County, Texas (the third largest county in the United States) was an alignment of elements needed for physical care and to meet guidelines for antipsychotic prescribing. The measures chosen were HbA1C and physical exams. Periodic reminders were built in the electronic health records. A second-level intervention was to add lipid profiles and a patient health questionnaire (PHQ-9) added.

Other opportunities for care alignment include using home visits by case management to screen for unfavorable social determinants of health, domestic and intimate partner violence, and other issues. A poorly explored area in community psychiatry and its nexus with medical care is the provision of end-of-life care. SMI patients have limited engagement in end-of-life planning and endure much suffering, often with concomitant physical and mental clinical decompensation. A potential role of community mental health and accompanying social services would be to promote advance directives and prepare patients for potential medical issues which they are not receptive to when they are ill.

From a broader perspective, psychiatry embedding in community centers can also be useful for prevention. Prevention is a natural impulse for any public health model. A disease or debilitating condition that can be avoided through early intervention and timely treatment can save a neighborhood or community and secure the overall health of the population. Health plans and community care systems should adopt carefully designed benefit packages for their enrollees, and that include preventive services and early interventions. The major public payers, Medicare and Medicaid, have incentive-based funding mechanisms encouraging some accountability for the most medically vulnerable patients. Smart public and community psychiatry will position itself in accountable clinical entities that are actively engaged in early intervention and intensive, individualized care management.

Application of the traditional public health concepts of primary, secondary, and tertiary prevention has been lacking in psychiatry but is highly possible. Primary prevention involves addressing the root causes of illness with healthy individuals, with the goal of preventing illness before it occurs. Secondary prevention involves the early identification and early treatment of individuals with acute or subclinical disorders or high-risk persons to reduce morbidity. Tertiary prevention attempts to reduce the effects of a disorder on an individual through rehabilitation and chronic illness care management.

Universal interventions are those intended for the general public, such as immunizations or media campaigns providing information about illnesses, early warning signs, and resources for health promotion and timely treatment. Selective interventions focus on individuals at higher-than-average risk (e.g., persons with prodromal symptoms or a family history of psychiatric disorders) to reduce morbidity by enhancing resilience and preventing the onset of illness. Indicated interventions target individuals who are experiencing impairment as a result of illness as early as possible in the course of the illness to reduce the burden of the illness on the individual, family, community, and treatment system.

Psychiatric disorders most often follow a course over time that begins with an often lengthy period in which prodromal or subthreshold symptoms or functional problems precede the full onset of an impairing disorder. Psychiatry integrated into primary care is in a position to identify these early symptoms and intervene effectively. In primary care, surveillance for very early universal intervention is possible, and selective interventions can be easily made as symptoms are recognized. Intensive psychiatric services can be organized in the form of indicated interventions when disabling symptoms appear to establish a path to recovery. A good example is early detection for individuals at high risk for psychosis and providing early interventions to prevent conversion to first break psychosis.

A public health-oriented strategy for disease and disability prevention will involve early identification and intervention in cooperation with consumers, families, health care professionals, and educators. Psychiatric professionals can play a key role by bringing their expertise to the development of appropriate screening, assessment, education, and treatment strategies that, if implemented on a wide scale, may reduce the burden of mental health disability and facilitate recovery. Psychiatrists also can become involved in the social marketing of prevention services and programs to add expertise and credibility to support legislation, regulations, and commercial investment that will provide the economic base necessary for broad-based prevention programs.

Prevention interventions have been found to be effective with adults with a variety of risk factors or preclinical problems. For example, women who have been raped are less likely to develop PTSD if they receive a five-session cognitive-behavioral treatment than if their recovery is left to chance. Men and women identified with subthreshold symptoms of depression in primary care settings are more likely to remain free from the full syndrome of depression or, if they do become clinically depressed, they will be able to recover rapidly with treatment if their standard medical treatment is enhanced by education about depression and learning skills for coping actively with depressive symptoms or stressors. This includes women who are at risk for postpartum depression, who have been shown to benefit from in-home cognitive behavior therapy in the first year after giving birth. Children who show early signs of severe behavioral problems (e.g., aggression, impulsivity, stealing) have been found to be less likely to become severely and chronically behaviorally impaired and antisocial when they and their caregiver(s) participate in therapeutic programs designed to enable the child and caregiver to deploy executive functions (i.e., anticipating consequences, inhibiting reactive behaviors, problem-solving) while simultaneously strengthening their attachment bond with each other.

The Medicaid program includes a mandated program for Early Periodic Screening Diagnosis and Treatment (EPSDT) that broadly covers very early intervention and extends services through the 21st year. States have been slow to meaningfully implement the periodic screenings and full use of the program because of the potential financial liability of the program to the states. Consent decrees have, in some cases, been used to expand these services for children with severe emotional disturbances. EPSDT could provide a complete framework for the early identification of young children with psychiatric risk factors and would support the effective targeting of comprehensive mental health and parenting services in the primary care setting of the family's "medical home."

COMMUNITY PSYCHIATRY AND THE PUBLIC SECTOR

Psychiatry's relationship to public sector agencies has, with some specific exceptions, been one of detachment because of the dominance of the private practice models before the advent of managed care and also largely because of the nature and structure of U.S. social welfare programs. Many psychiatry training programs lack meaningful rotations in public sector settings. In contrast to most industrialized nations in which comprehensive social welfare reforms were integrated, U.S. social welfare programs grew incrementally and categorically, that is, one category of service at a time. Large-scale initiatives, such as President Johnson's War on Poverty, have been implemented in a piecemeal fashion by fragmented federal, state, and local government bureaucracies. Various social service agencies have been created through a process and set of alliances that has been called the iron triangle. Advocates form an organization to champion a particular cause, such as blindness, developmental disabilities, primary care, or mental health, among others. They find key legislative sponsors who advance the cause through legislation and appropriation. This creates a bureaucracy, and then bureaucrats join the alliance. Through successive legislative sessions, the alliance of advocates, lawmakers, and bureaucrats builds a stronger categorical agency; for example, adults and families with dependent children or people with biologically based, SMI. From the 1930s through the end of the 20th century, U.S. social welfare agencies were created on this pattern: The system has resulted in generous funding and hopeless fragmentation of effort at local service delivery levels in which each agency depends on separate funding silos that also deliver conflicting rules and regulations.

It has been said that "mental health is not a place." Psychiatrists have a role at the receiving end of every categorical silo because the clients of each categorical agency experience mental illness or various addictions. Services for children with severe emotional, mental, or behavioral disturbances are a dramatic case of the importance of psychiatrists taking a leadership role in promoting comprehensive rather than fragmented services. Five categorical agencies—child welfare, education, primary health, substance abuse, and juvenile justice—all have a responsibility to care for these children. Especially for the children with the most severe disabilities, protective services, special education, the juvenile probation officer, and the child's pediatrician all need the consultation and support of a skilled psychiatrist. The situation is similar for adults with SMI, for whom case managers, vocational rehabilitation counselors, basic needs benefits specialists, social workers, psychotherapists, substance abuse counselors, parole or probation officers, legal conservators, visiting nurses, peer specialists, and physicians all may be mandated to deliver services. The need for psychiatric leadership is equally important for those who care for persons with mental retardation or developmental disabilities, for the older adult, for persons in vocational rehabilitation programs, and for persons in correctional programs. Virtually all public resources come through this complex of categorical agencies.

Health policy is formulated between government and employer purchasers and expressed unevenly through public health functions in communities, consumers, and advocates. Academic psychiatry bears responsibility for research, training, education, and workforce development for a behavioral health system that is only loosely integrated into the health care system. It will be impossible to achieve the ideals of the public health perspective on public and community psychiatry unless ways are deliberately devised to engage committed and proficient psychiatrists in meaningful roles in the operations of these agencies. Again, the factors at play in health care reform provide an opportunity to integrate the efforts of the complex and fragmented array of health and social service agencies by an emphasis on the primacy of the clinical process (maintained within the PCMH), the revision of funding mechanisms to encourage clinical sponsorship by accountable clinical entities, and emphasis upon the implementation of evidence-based practice standards.

BARRIERS TO ACCESS

The National Comorbidity Survey replication reports that less than 40% of those with severe psychiatric disorders have had an encounter with a psychiatrist in the last year; that one in six (15%) has received minimally adequate care; that one in twenty persons with substance use disorders received care. Even when mental illness is identified, people with socioeconomic adversities often do not, or cannot, get adequate mental health services in their communities. For example, although it is estimated that more than a quarter of a million incarcerated adults in the United States have psychiatric disorders, few were detected or received treatment until they reached jail or prison. Federal and state correctional systems have instituted mental health screening and treatment programs to address psychiatric disorders as a health problem for incarcerated adults and to manage the problematic behavior that can occur in controlled settings as a result of mental illness. On returning to the community, the vast majority of prisoners with psychiatric disorders cease to receive more than minimal mental health services: A recent study found that fewer than one in six (16%—strikingly similar to the National Comorbidity Study finding) received steady mental health services, and only one in twenty with addictions received steady substance abuse recovery services. Thus, access to mental health and addiction treatment services is far better in prison than in the community. This obviously causes great concern because of the possibility that correctional facilities may be a de facto system of care for low-income people (often of minority backgrounds) with serious mental illness.

Evidence of serious and pervasive barriers to accessing mental health care can be found in several kinds of crisis situations. Children and adults in poverty increasingly are deferring physical and mental health care until illnesses become chronic and severe and then find it difficult to gain entry into any setting for services except public hospital emergency departments. Children with severe psychiatric or behavioral disorders are often held in emergency department facilities for days and even weeks because the staff cannot locate any treatment facilities with an appropriate level of care that have an opening or are willing to take the child as a patient (or to accept the child or family's insurance coverage or lack thereof). People who must rely on the public sector services thus face daunting obstacles when seeking appropriate mental health care because of the serious underfunding of practitioners and programs. The economic forces and public policy dilemmas that drive the ever-increasing cost of health care bear directly on public sector psychiatry, as well as on the lives of tens or hundreds of thousands of people who do not receive adequate care.

COMMUNITY PSYCHIATRY AND RECOVERY

It is important to survey where and how Public and Community Psychiatry will participate in the changes that are afoot for American health care and how it will pioneer and support recovery for vulnerable patients. The state hospitals were the first publicly funded health care, but they were separated in a strange way from the rest of health care as American medicine found its scientific legs in the reforms sparked by the Flexner Report in 1910. The strides that were made by medicine and science in the early 20th century were not possible for psychiatry because the tools were not yet available for the science of the brain. Psychiatry remained in the state hospitals with patients supported with public funds, and the separation between psychiatry and the rest of medicine widened. By the end of the century, this had changed dramatically, and psychiatry was building on a flourishing base of science. The change was advanced in the middle of the century when biologic psychiatry and the new medications began to be used and understood. But the hospitals were still operating, and the strange separation of psychiatry from the rest of medicine persisted. The former Commissioner of Mental Health in Massachusetts and Director of the NIMH's Joint Commission on Mental Health and Illness (1961), Jack Ewalt, MD, who had played a key role in the development of community mental health, wrote a small book that methodically detailed the operational decline of the state hospitals at the same time that the Hill-Burton legislation was funding sparkling new general hospitals in communities throughout the country.

The term Public and Community Psychiatry may change just as dramatically in the next decade as it serves more people and has access to newer medications but also the advantage of technology, wearables, and personalized medicine. The junction we are currently experiencing is that of aiming for remission or recovery.

ATTITUDES TOWARD COMMUNITY PSYCHIATRY

The premise of community psychiatry is to move care from institutions to the community. Thus, a direct correlate is to place clinics and other community resources in neighborhoods. An ideal scenario is full community acceptance and support of its most vulnerable individuals, but in fact, community psychiatry could face opposition from some neighborhoods; in this case, the opposition is linked to negative attitudes such as stigma toward mental illness, fear that community clinics would bring down the value of the neighborhood or bring about crime or violence, etc. Questionnaires such as the Community Attitude toward Mentally Ill questionnaire (CAMI) are designed to measure these attitudes and help guide public policy. Attitudes toward mental illness and community psychiatry can be viewed as individual versus societal. Different influences impact these two currents. Societal stigma links to systemic barriers, which are meant to be fought by community psychiatry establishments and systems. Individual stigma and discrimination are but one level of systemic stigma.

Overall, most people have neutral to positive attitudes to the idea of community psychiatry when it does not affect their lives unless they believe mental illness is associated with crime and violence. Increasing acceptability of community psychiatry would benefit from focused messaging and from starting mental services in a stepwise fashion. An important factor in increasing acceptability is also to increase individuals' direct contact with those who suffer from mental illnesses, as this step, increasing familiarity, has been found to help the acceptability of services.

As with other new concepts, the personality traits of recipients shape their responses on an individual level (for instance, using a big five-factor model, openness and agreeableness are associated with more acceptability and positive attitudes toward community psychiatry, whereas neuroticism is associated with less). Negative attitudes toward mental illness itself, such as believing that an individual with mental illness should be avoided and that marrying or partnering with someone with a mental illness is a bad decision (CAMI items), correlate with negative attitudes toward psychiatry, help-seeking, and community psychiatry perceptions. These opinions may be more negative in rural areas and in less industrialized countries.

Negative attitudes toward community psychiatry denote a poor understanding of the role and scope of the discipline. The advantage of having community services embedded in neighborhoods is to have accessible services bypassing access barriers and avoiding disconnecting the patient from their sources of support, typically families. Espousing negative attitudes to help-seeking and to community services easily leads to delay in care, presenting with more severe symptoms, and therefore reinforcing the view that mental illness is incompatible with normal everyday life. Of note, attitudinal barriers affect referral in mild to moderate cases, whereas structural barriers seem to be much more influential in severe cases, according to a WHO survey of 63,678 individuals with mental illness. As individuals with severe illness are the main recipients of community psychiatry services, it is essential that systemic and structural barriers be addressed, including societal ones.

A meta-analysis of studies on attitudes and perceptions toward mental illness shows an increasing trend toward better mental health literacy, moving toward a more biologic model of mental illness and perhaps greater acceptance of biologic, psychotropic treatments in middle- and high-income countries, but no significant change in stigma or attitudes toward mentally ill individuals.

Interventions to reduce stigma range from mental health education or provision of information, direct contact (shown to disconfirm stereotypes related to mental illness), indirect contact (increase familiarity), entertainment, and art (to replace stereotypes with healthier, more balanced views and more objective information), psychoeducation and even psychotherapy. In a review of more than 8,000 studies, Thornicroft et al. found that contact (direct) produces short- to medium-term knowledge improvement, but attitudinal improvement or does not occur. The impact of these attitudinal, stigma, and discrimination-based views can be far-reaching and affect funding and expansion of community resources, not just utilization. This area is not well researched and could benefit from manualization and increased fidelity in intervention delivery. Additionally, different campaigns could be needed for different types of mental illness and for low-, middle-, and high-income countries.

Attitudes of police officers are integral to community psychiatry response. For example, in an examination of how different agencies respond to borderline personality disorder patients, reviewers found that different agencies will assign different diagnoses to the same patient and hence deal with them differently, assigning them to different workflows. This illustrates the tremendous need to coordinate care frequently and across disciplines and institutions for successful community psychiatry outcomes.

Stigma and discrimination in community psychiatry also show in how police officers deal with mental illness, reflective societal prevailing attitudes. A common example is the use of restraints (handcuffs) when transporting patients from community settings to the emergency center or the hospital, even when the patient is voluntary.

COMMUNITY PSYCHIATRY AND PSYCHIATRIC TRAINING

According to the National Survey of Psychiatric Practice, more early and mid-career psychiatrists were working in the public sector/community psychiatry than in private settings, an increasing trend since the 1990s, with 2002 being the first year for more psychiatrists being in the former. In 1957, only 1,700 of the 10,000 American Psychiatric Association members worked in state hospitals. Within less than 20 years, the overwhelming majority of psychiatrists in the United States went from working in state hospitals to working in other settings that had little or no interface with the populations having an SMI or serious and persistent mental illness or with children with serious emotional disorders.

Training experiences in community psychiatry are required by the Accreditation Council for Graduate Medical Education. The swath of psychiatric care covered by community psychiatry has grown, and more psychiatrists practice in community-funded settings than in private practice or academic and research settings. Community psychiatry training tracks remain few and far between, with only four tracks being active in the United States in 2014. Creating alternatives to community psychiatry exposure in residency has relied on experiences with ACT teams or integrated psychiatry, though these two settings do not offer the same experience but rather some similarities. Community psychiatry fellowships focus on integrating care with shelters and primary care physicians to cover the spectrum of care the SMI population requires; the experience also focuses on building a human appreciation of the community psychiatry patient, an often traumatized individual dealt heavy blows by life. Despite all these limitations, there are views in favor of incorporating more community psychiatry experiences in psychiatry rotations, given the expansion of the field and the remarkable diversity of skills and collaboration opportunities it teaches.

What Skills Ought to Be Taught in Community Psychiatry Experiences?

The first is to improve knowledge of what to expect in the community psychiatry population, which has a particular profile of illness, as described in the SMI section. Patients in community clinics may often have complex trauma and overlapping symptoms, for which careful history taking is required. Familiarity with disease comorbidities is essential as most community psychiatry patients will suffer from co-occurring disorders. A common finding is that patients with psychotic symptoms of any nature will be labeled with a diagnosis of a primary psychotic disorder, such as schizophrenia or schizoaffective disorder, even when a proper diagnosis is a stressor or trauma-related disorder, or a substance-induced disorder. This label has implications for treatment recommendations and the adverse effect burden one may get from prolonged exposure to antipsychotics. Additionally, the stigma of having a psychotic disorder may cause detrimental long-term effects for self-image, hopes of recovery, and employment-seeking. More importantly, a mistaken diagnosis can lead to missed treatment opportunities, for example, focusing on pharmacologic measures and not referring to psychotherapy for trauma or giving antipsychotic medication but not an antidepressant.

The second skill essential to success in community psychiatry practice is efficient time management. As time is limited in most encounters, it is easy to skip elements of the interview that can be very revealing. Many learners are tempted to skip a formal mental status exam, recall testing, or to ask about first-rank psychotic symptoms, preferring to ask about hallucinations only. While time constraints are very real, the adequacy of history-taking remains the cornerstone of diagnostic delivery, and learning this skill is essential.

The adequacy of cognitive testing and of carefully surveying a prior history of violence and traumatic brain injuries has to also be emphasized. Individuals with SMI are at a higher likelihood to be victims of domestic and sexual violence, and individuals who have a family representative as a payee for their benefits (e.g., social security disability) are also noted to have a higher risk of family violence and victimization. Individuals who are homeless have a higher risk of traumatic brain injury compared to the general population. When rushing through a complex plan of care, often including off-label medication usage, the combination of head injuries, interpersonal violence, and untreated neuropsychiatric sequelae of these issues can largely contribute to poor understanding of instructions and poor follow-through. The same pattern can be found in patients with a history of heavy alcohol use. Thus, it is really important for learners to know how to complete thorough data gathering and select the types of cognitive tests, albeit brief, they will need to construct a proper formulation of the patient’s level of impairment.

Third, the essence of community psychiatry is an ability to absorb complex situations and histories and to work through them in teams. Learners need to practice interprofessional data gathering, sharing, and decision-making. The physician’s role in the community psychiatry team is pivotal but may be somewhat underemphasized. Learners will often find other team members thankful for a collaboration and for a nuances biopsychosocial formulation with proper use of terminology to accompany what they have already observed in their clinical interactions.

Fourth, the cultural aspects of community psychiatry, including engagement, understanding of the patients’ background, and respect of cultural beliefs, have to be taught. There may be, in some subpopulations, a high preponderance of issues such as war violence, a refugee experience, or human trafficking that learners may not be familiar with.

Last, the structure of busy clinics in community settings, away from academic support and supervision, takes adjustment and modification of workflow. It is not studied whether the rates of burnout differ for community versus noncommunity psychiatrists, but the volume and severity of psychopathology seen in community psychiatry necessitate healthy coping mechanisms and often involves introspective reflection on severe illness and social justice.

TELEPSYCHIATRY IN THE COMMUNITY

The advent of telepsychiatry has the potential to improve delivery to community psychiatry patients but also highlights disparities in internet access and the ability to benefit from virtual psychiatry experience. In community psychiatry, financial restrictions have made encounters with psychiatrists shorter than they had traditionally been. There has been a shift toward more biologic aspects of psychiatric treatments and a dissociation of therapy and medication management sessions. So the baseline, typical visit in a community psychiatry setting has been described as brief and focused. These changes have been consolidated since the managed care movement in the 1980s.

Adept community psychiatrists maximize the use of the limited time they spend with patients and maximize engagement in those encounters. Using virtual encounters, however, has been challenging for the patient population typically served in community psychiatry settings. To achieve a proper result in a physician–patient encounter, aspects of rapport-building, diagnosis, and treatment decisions must be maintained; and to do so, the level of immersion in the visit, the level of engagement/emotional distance, and privacy have to be managed. The physical elements of the encounter, referred to by the psychoanalyst and artist Milner as the “frame,” shift from in-person to virtual, playing the responsibility of privacy and adequate space to conduct the visit, and placing the onus of working out technical requirements and difficulty on the patient. Patients who do not have an adequate space to conduct the visit because of homelessness and lack of privacy in crowded living situations that many patients served in community psychiatry endure. This obstacle can translate into discomfort about discussing issues pertaining to the visit and potentially, because of increased sensory input, distractibility during the visit.

By and large, telepsychiatry is highly accepted, with some studies indicating that show rates in clinical settings improve with the use of virtual visits, as transportation barriers are bypassed and it is overall easier for patients. But it is not the case for individuals with SMI such as schizophrenia. Levels of engagement in therapy services and continuation of therapy use virtually may suffer as well in the SMI patient population. This fits with already known data from the Results From the National Comorbidity Survey Replication showing that by and large, individuals with significant mental illness remain under or untreated, especially in traditionally underserved groups, including older adult persons, racial–ethnic minorities, those with low incomes, those without insurance, and residents of rural areas, and therefore the patient populations that is the main target of community psychiatry. This is also the population hypothesized to not be able to fully utilize virtual visits. To engage in telepsychiatry, one has to have access to the platform needed and learn how to use the technology involved. The SMI population often has trouble with both, though once SMI patients are converted to telepsychiatry, they often remain and show equivalent levels of satisfaction to non-SMI patients. Thus, the difficulty may be to establish proper structural elements to help this transition. Of note, even though internet-enabled smartphone ownership is almost ubiquitous, patients with SMI in general, and psychosis in particular, hover around 78% in use, with even less owning a smartphone currently, and these parameters correlate with negative symptoms and cognitive functioning.

The Study to Promote Innovation in Rural Integrated Telepsychiatry (SPIRIT) study, currently underway in 12 FQHC in three states, aims to examine how to optimize engagement in care in patients with bipolar disorder or PTSD in a sequential, multiple assignments, randomized Trial design, and which of these patients would benefit from Telepsychiatry Collaborative Care. Preliminary data show that engagement in psychiatric care improved outcomes, but that success hinged on the presence of case management and psychology services in conjunction with telepsychiatry services; further, the telepsychiatry aspect is supplemented by an in-person case manager to facilitate adherence and implementation of treatment recommendations virtually given.

Virtual psychiatry applications for SMI can also clash with the nature of the disease targeted during the visit. The presence of delusions, such as mind-reading or persecutory delusions, belief of having an implanted chip or tracking device, could impair engagement with the psychiatrist on screen and potentially worsen the pre-existing delusion. Hallucinations would make engagement difficult as well and may be harder to redirect on screen compared to in-person interactions. Poorly controlled mood or personality symptoms, with acting out on screen (self-harm, using substances) is harder to mitigate virtually than in person as well. Potential negative outcomes of poor engagement could show as nonadherence with a treatment plan, dropping out of treatment altogether, and having to receive emergency or inpatient services because of clinical decompensation.

The addition of cultural and language barriers may further impair the successful use of telepsychiatry in non-English speaking patients. The Cultural and Linguistic Appropriate Services, via the U.S. Health and Human Services Department, outlines 15 standards of culturally competent care that have to be intermeshed with telepsychiatry structures and equipment and the ACGME framework. There are gaps in knowledge relating to applications of telepsychiatry to cultural aspects of community psychiatry. There are also severe gaps in determining the ideal clinical workflows to reliably help transitional the patients in community psychiatry clinics to virtual platforms successfully. It is likely that structural/organizational efforts have to be set in place for this purpose, including staff help to train patients on technology use, connect them prior to physician visits, and carefully review discharge instructions. These measures would help avoid further decreasing physician time and would solidify understanding of treatment recommendations. Special considerations should be paid to patients with substance use disorders needing regular drug testing, patients who receive long-acting injectable medications, and patients on medication regimens needing regular lab work monitoring.

ASSETS AND ADVANTAGES OF ONGOING PROGRAMS

The Harris County Experience

Houston, Texas, is the fourth largest city in the United States. Harris County, the third-largest county in the United States, has two networks of community psychiatry that work in a complementary fashion. One is a state-funded mental health system targeting the three main SMI categories (psychotic disorders, major depressive disorders, and bipolar disorders) and intellectual disabilities and strictly adhering to financial restrictions, thus serving the most vulnerable population in the county. Funded by the legislature, it follows all the above rules and provides an invaluable safety net to the community but providing outpatient services, a comprehensive psychiatric emergency program, and case management, and a crisis outreach service.

The second service, The Harris County Community Behavioral Health Program (CBHP), funded by the county itself, has had to grow and fill the gaps that state legislation does not cover. In doing so, an alternative model of providing community psychiatry services has evolved. In 2004, CBHP placed behavioral health staff, therapists and psychiatrists, into 21 community health centers in order to better meet the behavioral health and primary care needs of the uninsured county population, winning the 2007 American Psychiatric Association Achievement Award in the category of community-based programs. The program proved to be cost-effective, highly rated by specialists and primary care alike, and well-liked by patients who appreciated direct access to psychiatric services. As CBHP has continued to grow, options for specialized programs have emerged. Lessons from CBHP include the following:

- Direct access to psychiatrists works well for most patients, avoiding care delays and allowing quick treatment initiation and improvement.
- The absence of restrictions on diagnosis does not place a heavy burden on volume as initially feared. Past the first stabilization period, numbers have been manageable. Lack of diagnostic restrictions means that psychiatrists can do a proper diagnostic job and not feel compelled to add a diagnosis so the patient can qualify for services.
- Colocation with primary care and specialty medical services works well for patients who often have comorbid disorders. As discussed above, patients with SMI tend to be medically complex. Careful care coordination is often required for medication regimens, as antipsychotics while bringing benefits, may also worsen chronic medical issues.
- Substance use treatments are a must given the large percentage of patients who have SUD.
- Psychotherapy, both individual and group, are necessary for successful outcomes in community psychiatry given the high prevalence of traumatic experiences in this population.
- Culturally competent psychiatry is essential given the diverse makeup of our population.
- Colocation with other services offers opportunities to educate and provide informal support to colleagues, fostering collaborations.
- Multidisciplinary efforts allow better ability to screen for domestic violence, intimate partner violence, and human trafficking.

- Community psychiatry is very local in flavor. Every clinic has a particular population, certain diagnoses that predominate, or an ethnic or racial group. The microcosm of a clinic is often difficult to describe, but psychiatrists embedded in each one usually have a feel for the needs and characteristics of their specific subpopulations.

The positive growth of CBHP has paved the way for more specialty programs that have emerged to fill the needs of the community, based on careful gap analysis and a desire to truly and decisively improve the lives of our patients. Specialty programs include the follow.

Multi-Visit Patient Program (MVP)

As discussed above, the goal of a successful community psychiatry experience is the smooth transition of patients from inpatient acute hospitalizations when needed to outpatient care. This lofty goal is not always possible, and the literature is rife with descriptions of patients who often present to emergency services. These patients are often substance users but sometimes have psychosis, mood, or personality disorders. Derogatory terms such as frequent flyers were much used in the 1980s and were gradually replaced by more neutral terms such as MVP. The MVP program provides linkage of patients who have more than six emergency center visits per year to rehabilitation resources, housing resources, and other community resources as indicated. The main drivers of utilization seem to be divided into behavioral, social, and medical. Solving these issues is often not within the hands of the medical system or the psychiatrists, testifying to the interdependence of agencies on each other in the community psychiatry structure. The MVP program has been successful in mitigating recidivism for multiple individuals, much to the relief of staff and patients alike.

Maternal Perinatal Addiction Treatment Clinic

Based on the Maternal Opioid Misuse model (MOM) funded by Medicaid, the MPAT clinic aims to serve the needs of pregnant women with histories of drug use. Women who use drugs have a much higher chance of unanticipated pregnancy, shorter durations between pregnancies than recommended by the WHO for optimal health, and frequent negative outcomes in relation to domestic violence and intimate partner violence. Pregnant women will often try to stop drug use when they discover they are with a child, but often fail: women who are pregnant tend to use less compared to those who are not, use tends to decrease over the trimesters, but often quickly rebounds postpartum. They report feeling judged and embarrassed, not just because of substance use but also because of associated behaviors such as prostitution. The percentage of pregnant women who use illicit substances (drugs, alcohol, tobacco) is 20.8%, according to National Survey on Drug Use and Health, from SAMHSA. Beyond just numbers, substance use disorders have been gaining press as state after state (UT, CO) highlighted concerns about the significant contribution of SUD to maternal mortality. The Maternal Mortality and Morbidity Task Force and Department of State Health Services, in its 2018 Biennial report, recommends the following:

- ensure that women with substance use disorders and mental health conditions have insurance coverage for necessary care services;
- support models of maternal health care that integrate behavioral health services; and
- develop strategies to ensure that behavioral health and community services to support long-term disease management are readily available, including expanding the number of behavioral health care providers and services across the state.

Thus, there was clearly a need for a specific intervention. The group of women targeted for this intervention are frequently victims of the intersectionality of trauma, poor social determinants of health, risky practices, like substance use, and high-risk medical pregnancies, in addition to potential other medical conditions. The MPAT clinic provides multidisciplinary support from neonatology, anesthesia, addiction, and general psychiatry, psychology, social work, maternal–fetal medicine, and from social and community works in inpatient treatment centers. The program has linkage to substance treatment centers, the jail system, the child and family protective systems and is involved in outreach efforts. Since its inception, the program has seen a tremendous increase in the number of referrals, and has been able to implement Universal Substance Use Screening in all county clinics, as screening is found to be superior to urine drug testing, is not confounded by bias and false positives, and gives the opportunity to intervene during the screening. Substance use is likely underrecognized without screening, and the MPAT program is an example of an evidence-based, multidisciplinary intervention that is unique to community psychiatry.

THE LARGER PICTURE

Community psychiatry has the opportunity to shape perceptions of mental illness and fight stigma as its reach increases. By interacting with multidisciplinary teams and working on social determinants of health, it is possible to generate a positive, evidence-based picture of appropriate, successful treatments. To achieve success, community psychiatry has to tackle societal issues affecting its patients, including gun violence, homelessness, the role of the legal system, and the limitations of psychotropic medications. As discussed above, the groups of patients served by community psychiatrists no longer just represent the sickest individuals with SMI. Therefore, community psychiatry must become flexible and integrate treatment for co-occurring disorders and personality pathology. It also has to include health promotion and accommodate insurance system payments.

A little-discussed aspect of community psychiatry is the fact that it is the de facto safety net for chronic childhood disorders and intellectual disabilities but also deals with emergency center recidivism, inpatient hospitalization recidivism, etc. The next decade will see specific subspecialties burgeoning within community psychiatry. The examples of programs illustrated above give an idea of the growing scope of community psychiatry and the potential it has. Community psychiatry is not the same it was 10 years ago. Patients who are employed, but want ease of access and someone with the ability to handle difficult cases, may be the next community psychiatry beneficiary, and services will have to be flexible to accommodate those needs.

FURTHER READINGS

Baxter AJ, Harris MG, Khatib Y, Brugha TS, Bien H, Bhui K. Reducing excess mortality due to chronic disease in people with severe mental illness: Meta-review of health interventions. *Br J Psychiatry*. 2016;208(4):322–329.

Begley CE, Hickey JS, Ostermeyer B, et al. Best practices: Integrating behavioral health and primary care: The Harris County Community Behavioral Health Program. *Psychiatr Serv*. 2008;59(4):356–358.

Charlson FJ, Ferrari AJ, Santomauro DF, et al. Global epidemiology and burden of schizophrenia: Findings from the global burden of disease study 2016. *Schizophr Bull*. 2018;44(6):1195–1203.

Charlson F, van Ommeren M, Flaxman A, Cornett J, Whiteford H, Saxena S. New WHO prevalence estimates of mental disorders in conflict settings: A systematic review and meta-analysis. *Lancet (London, England)*. 2019;394(10194):240–248.

Correll CU, Solmi M, Veronese N, et al. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: A large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. *World Psychiatry*. 2017;16(2):163–180.

Crossley N, Sweeney B. Patient and service-level factors affecting length of inpatient stay in an acute mental health service: A retrospective case cohort study. *BMC Psychiatry*. 2020;20(1):438.

Fortney JC, Bauer AM, Cerimele JM, et al. Comparison of teleintegrated care and telereferral care for treating complex psychiatric disorders in primary care: A pragmatic randomized comparative effectiveness trial. *JAMA Psychiatry*. 2021;78(11):1189–1199.

Fortney JC, Heagerty PJ, Bauer AM, et al. Study to promote innovation in rural integrated telepsychiatry (SPIRIT): Rationale and design of a randomized comparative effectiveness trial of managing complex psychiatric disorders in rural primary care clinics. *Contemp Clin Trials*. 2020;90:105873.

Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the global burden of disease study 2017. *Lancet (London, England)*. 2018;392(10159):1789–1858.

Goddu PA, O’Conor KJ, Lanzkron S, et al. Do words matter? stigmatizing language and the transmission of bias in the medical record. *J Gen Intern Med*. 2018;33:685–691.

Grassi L, Riba M. Cancer and severe mental illness: Bi-directional problems and potential solutions. *Psychooncology*. 2020;29(10):1445–1451.

Gratzer D, Torous J, Lam RW, et al. Our digital moment: Innovations and opportunities in digital mental health care. *Can J Psychiatry*. 2021;66(1):5–8.

Hilty DM, Gentry MT, McKean AJ, Cowan KE, Lim RF, Lu FG. Telehealth for rural diverse populations: Telebehavioral and cultural competencies, clinical outcomes and administrative approaches. *Mhealth*. 2020;6:20.

Joseph R, Kester R, O'Brien C, Huang H. The evolving practice of psychiatry in the era of integrated care. *Psychosomatics*. 2017;58(5):466–473.

Kozloff N, Mulsant BH, Stergiopoulos V, Voineskos AN. The COVID-19 global pandemic: Implications for people with schizophrenia and related disorders. *Schizophr Bull*. 2020;46(4):752–757.

Little J, Trauer T, Rouhan J, Haines M. Borderline personality disorder and interagency response. *Australas Psychiatry*. 2010;18(5):441–444.

Martínez-Martínez C, Richart-Martínez M, Ramos-Pichardo JD. Operational definition of serious mental illness: Heterogeneity in a review of the research on quality-of-life interventions. *J Am Psychiatr Nurses Assoc*. 2020;26(3):229–244.

Mckinney JR, Russell M, Avellaneda-Ojeda A, et al. A comprehensive care approach for pregnant persons with substance use disorders. *Int J Mental Health Addiction*. 2022:1–12.

Mendiola CK, Galetto G, Fingerhood M. An exploration of emergency physicians' attitudes toward patients with substance use disorder. *J Addict Med*. 2018;12(2):132–135.

Mishkind MC, Shore JH, Bishop K, et al. Rapid conversion to telemental health services in response to COVID-19: Experiences of two outpatient mental health clinics. *Telemed J E Health*. 2021;27(7):778–784.

Miu AS, Vo HT, Palka JM, Glowacki CR, Robinson RJ. Teletherapy with serious mental illness populations during COVID-19: Telehealth conversion and engagement. *Couns Psychol Q*. 2020:1–18.

Neilson S, Chittle A, Coleman T, Kurdyak P, Zaheer J. Policies and procedures for patient transfers from community clinics to emergency departments under the mental health act: Review and policy scan. *Int J Law Psychiatry*. 2020;71:101576.

Reardon CL, Buhr KA, Factor RM, Walaszek A, Zeidler Schreiter E. Integrated care: Should it count as community psychiatry training for psychiatry residents? *Community Ment Health J*. 2019;55(8):1275–1278.

Sariaslan A, Fazel S, D'Onofrio BM, et al. Schizophrenia and subsequent neighborhood deprivation: Revisiting the social drift hypothesis using population, twin and molecular genetic data. *Transl Psychiatry*. 2016;6(5):e796.

Schneider F, Erhart M, Hewer W, Loeffler LA, Jacobi F. Mortality and medical comorbidity in the severely mentally ill. *Dtsch Arztebl Int*. 2019;116(23-24):405–411.

Selten J-P, Cantor-Graae E. Hypothesis: Social defeat is a risk factor for schizophrenia? *Br J Psychiatry*. 2018;191(S51):s9–s12.

Talley RM, Brunette MF, Adler DA, et al. Telehealth and the community SMI population: Reflections on the disrupter experience of COVID-19. *J Nerv Ment Dis*. 2021;209(1):49–53.

Tesfaye Y, Agenagnew L, Terefe Tucho G, et al. Attitude and help-seeking behavior of the community towards mental health problems. *PLoS One*. 2020;15(11):e0242160.

Topolovec-Vranic J, Ennis N, Colantonio A, et al. Traumatic brain injury among people who are homeless: A systematic review. *BMC Public Health*. 2012;12(1):1059.

Wolska A, Malina A. Personality and attitudes towards people with mental disorders: Preliminary studies results. *Int J Soc Psychiatry*. 2020;66(3):270–278.

56.2 Reforming Health Care

NIDAL MOUKADDAM, M.D., PH.D., AND ASIM A. SHAH, M.D.

Reform is the process by which one achieves health and optimal utilization of health resources. Health care is provided in public versus private sectors, with the balance differencing by country and sociopolitical agenda of governing political parties. Yet the realities of care provisions are also heavily impacted by a country's financial resources and organizational capacities and broader geopolitical forces such as conflicts/wars and natural disasters. Thus, health care delivery, and therefore health reform, has to include a somewhat stable/static, long-term planning element and more dynamic adaptive capacity reflecting budgetary and sociopolitical realities.

Health care delivery occurs in acute, primary, secondary, tertiary, and community-based services. When thinking of reform, different strategies and goals apply to each setting, with the overarching goal of maintaining an efficient, nimble, and cost-effective health care continuum.

In the United States, the goals of reforming health care are

- improving the experience of care
- improving the health of populations
- reducing per capita costs of health care.

Another way to consider reform is to think of it as

- insurance market reform
- delivery reform
- payment reform.

One would also be remiss not to mention that the pandemic will have long-lasting impacts on reimbursement and medical models. A central theme of reform, to be conducted successfully, is to maintain the balance between innovation and collaboration and stability. Reform is inextricably linked to politics.

The COVID-19 pandemic has shown the perils of systems where health care coverage is not easily accessible and the financial needs of keeping a population healthy. This chapter will discuss how reform can gain from discrete, well-defined measures and how those measures interface with politics and the law. The interface of medical reform with legal issues can impact the environment in which physicians practice, via tort reform and change caps on economic and noneconomic damages, with improving availability of medical review panels.

Puzzling aspects of health care reform center around the simple fact that health care recipients are *individuals.* In contrast, the stakeholders range from patients, researchers, policymakers, and the general public at the receiving end of reform, even when in a healthy state. Health care delivery has economic, financial, logistical, and political dimensions and a public health and social scope. Health care delivery is typically a significant expense for a country.

The Commonwealth Fund report "Mirror Mirror: reflecting poorly" (https://doi.org/10.26099/01DV-H208) compares the health care systems of 11 high-income countries by analyzing 71 performance measures across five domains—access to care, care process, administrative efficiency, equity, and health care outcomes. These are drawn from surveys and administrative data from the Organisation for Economic Co-operation and Development and the World Health Organization. Last published in 2021, the report still points to the United States as ranking last on access to care, administrative efficiency, equity, and health care outcomes, but second on care process measures. It is estimated that 16.8% of U.S. GDP is spent on health care.

In the United States, an individual who cannot afford private insurance, for example, from a job loss, applies for Medicaid if they meet socioeconomic criteria, apply for Medicare if they received social security income for at least 2 years, or buy insurance through the Affordable Care Act. The transition from personal experience to structural health coverage expansions is where the sociopolitical realities interfere. For the example listed in this paragraph, the options experienced by the individual rely on states expanding care coverage. This expansion can be done in multiple ways that reflect a country's, state, or community structure and political leanings. In the United States, previously suggested ways to ensure expanding coverage include expanding Medicare coverage to individuals starting the age 60 instead of 65, Medicare for all or other inclusive plans, or expanding Medicaid. The tenet of Medicare for all (or any name that one chooses to select for a reform campaign) would be controlling the prices paid to physicians, hospitals, and drug companies. This would theoretically be achieved by reducing administrative costs, including insurance overhead and profits. Thus, Medicare for All or equivalent plans could be the equivalent of quasi-universal coverage.

To put things in perspective, let us look at the dimensions from the Commonwealth Fund report, then compare them to the triple aim espoused by the ACA:

- **Access to care** includes measures of whether an individual can afford proper health care and if they can get this health care in a timely fashion. Affordability of health care is linked to timeliness because seeking emergency services will usually be linked to worse care outcomes and higher expenses. Countries that do well in that respect usually provide some sort of universal health care formula. While never without criticism, those universal access programs avoid desperate emergency service use as a last resort. This measure includes access to after-hours and weekend care and insurance denying payments of a claim or paying less than expected, therefore leading to financial issues for patients.
- **Care process** includes measures of preventive care, safe care, coordinated care, engagement, and patient preferences. This aspect has improved tremendously in the United States since the ACA and Patient-Centered Outcomes Research Institute (PCORI) initiatives were implemented. U.S.-specific challenges are cost-effectively maintaining care processes while minimizing care disparities. Additionally, the U.S. ranks lowest in terms of

patients having continuity with the same doctor, which is a concerning measure for individuals, as continuity is associated with better long-term patient outcomes.

- **Administrative efficiency** refers to the bureaucratic processes by which health care provision is managed. Can a physician get their patient’s medications or referrals efficiently or promptly? Can they implement required guidelines quickly when needed? In this measure, the U.S. ranks poorly compared to other wealthy countries, and most physicians rank administrative efficiency as suboptimal. Two aspects of administrative efficiency are subject to reform potentially. One is the administrative burden aspect and how much physicians have to do to schedule and implement treatment. The other is the administrative burden becoming a deterrent for patients to seek care. This links to patients getting more emergency care and further bloating and excess expenditures in the system.
- **Equity and Income-Related Disparities** is a well-known, well-advertised issue in the United States that remains at the forefront of the administrative and political agenda. The ACA implementation has helped bridge the gap, but as of the 2021 data, minorities in the United States still have subpar access to care and poorer health outcomes than their nonminority counterparts and those in other countries. One notable poor statistic in the United States is that maternal mortality is 17.4 deaths per 100,000 live births, whereas the next high-income country (France) has a rate of 7.6 per 100,000 live births.

Table 56.2–1.
Five Periods of American Legislative Activity for National Health Insurance

Dates: Proposed Reform	Benefit	Administration/Financing	Model/Outcome
1912–1919 American Association for Labor Legislation (AALL): Opposed by American Medical Association (AMA), American Federation of Labor, businesses	Medical care, sick pay, funeral expenses for lower-paid workers (<$1,200/year) and dependents	Administered by states; financed by payroll tax collected in regional funds	Model: Social insurance Failed to pass
1946–1949 Wagner–Murray–Dingle Bill: Supported by President Truman; opposed by AMA	Retirement, disability, unemployment, and health insurance: Physician and hospital care	Expansion of Social Security; employer/employee contributions to “social insurance trust fund”	Model: Social insurance Failed to pass
1963–1965 Medicare and Medicaid	Medicare: Hospital care (Part A) physician care (Part B); medications (Part D) Medicaid: Comprehensive health care, long-term care	Medicare administered federally, Medicaid by the states; Medicare financed through Social Security, federal income taxes, and individual premiums; Medicaid financed by general state and federal tax revenues	Medicare: Social insurance model Medicaid: Public assistance (welfare) mode Both programs were enacted
1970–1974 Kennedy–Griffiths Health Security Act:	National health insurance for the entire population	Administered federally, eliminating all public and private insurance; Nixon plan featured employer	Model: Combined social insurance and

President Nixon countered with an employer mandate plan; Physicians for National Health (PHN) proposed “single payer”		mandate, privately administered, plus Medicaid and Medicare; PNH plan financed by federal tax with single fund in each state	public assistance; failed to pass Nixon plan: Maintained public assistance; added social insurance; failed to pass PNH plan: Merged social insurance and welfare; failed to pass
1991–1994 Clinton Health Security Act: Opposed by insurance and pharmaceutical industries and small employers, single-payer advocates; proposed voluntary approaches	Comprehensive health care	Administered federally; financed through governmental funds, employer mandate, and individual mandate (with subsidies); voluntary: Tax incentives for employers; medical savings accounts (MSAs) for individuals	Model: Integrated social insurance and public assistance Failed to pass

From a mental health perspective, suicide rates in the United States have continued to rise for the past 20 years; and while U.S. patients are more likely to accept mental health care, they have less access to mental health professionals and social services. Almost a quarter of individuals in the U.S. report having been told they have anxiety, depression, or another mental issue, but the abundance in diagnosis is not accompanied by effective treatment or mitigation of adverse outcomes such as suicide (https://www.commonwealthfund.org/publications/issue-briefs/2020/may/mental-health-conditions-substance-use-comparing-us-other-countries).

HEALTH CARE REFORM IN THE 20TH CENTURY: A HISTORICAL OVERVIEW

The reform of the U.S. health care system has been periodically and repeatedly proposed since and before the social reforms of President Franklin Roosevelt. Presidents Harry Truman, Jimmy Carter, Gerald Ford, George H. W. Bush, and Bill Clinton all considered and proposed reform plans. Earlier proposals for national health insurance set the framework for these proposals. Table 56.2–1 summarizes the significant periods of legislative activity since the early 20th century and the complex elements involved in the proposals. Earlier proposals coupled health with other social welfare programs; later proposals narrowed in on health care as health care costs increased dramatically in the later part of the 20th century. Proposals were designed on either the model of social insurance or that of public assistance or welfare. As the importance of universal coverage became clear, these models merged. The locus of administration or control of the proposed health insurance programs—by employers, the federal government, or the states—became important. In particular, the most critical issue was the financing of the system: through government funds, the requirement that employers pay for their employees ("employer mandate"), or the requirement that individuals buy their own insurance ("individual mandate").

In the last decade of the 20th century, as health care costs continued to spiral upward, the proposed Clinton Health Security Act generated a tremendous amount of energy but eventually failed to be enacted. Large employers changed their methods of purchasing care, hiring companies to introduce a variety of cost control techniques, generally described as "managed care." Meanwhile, "incremental" measures were taken in the 1996 Health Insurance Portability and Accountability Act, the 1997 Balanced Budget Act, which limited payments to hospitals, and the 2003 Medicare Prescription Drug Improvement and Modernization Act. In addition to providing partial coverage to Medicare enrollees for medications, the latter act provided the Medicare program with the flexibility to introduce more aggressive health care purchasing techniques to control cost. These measures also supported individual incentives to control cost through new insurance models such as medical savings accounts, which combine tax-exempt personal savings for health care with catastrophic insurance coverage. The movement toward "managed care" purchasing did slow the rising costs for a few years, but its requirements were relaxed due to a strong reaction against control by both patients and providers.

Managed Care: A Phenomenon on the Path to Health Care Reform

The freewheeling spirit resulting from having generous public and private financial support for health care from roughly 1950 to 1980 led to a heady time for U.S. medicine. Society supported the clinical advances built on research that promised breakthroughs just over the horizon. Practitioners were, by association, part of this new era, and, as authorities in their professions, their determination of what patients needed was funded almost without question by public and private sources. As resources became limited owing to many factors, such as inflationary pressures and legislation, payers began to question what it was that health care was delivering for the dollar; that is, the issue of value was raised for the first time in the health economic equation by those paying for health care. A corollary of the issue of value was the question of the accountability of practitioners. Subsequent landmark studies demonstrated geographically differing practice patterns and consequently different costs but similar outcomes for the same disorder or disease. This raised the general question of why this should be the case? Other fundamental issues that have emerged include the following: How can access to care be broadened? How can clinical outcomes as a result of health care be improved? What about the uninsured? How can costs be controlled?

Utilization Review

Employers and purchasers embraced the concepts of utilization review and utilization management as strategies to try to ensure that the most appropriate and most cost-effective clinical decisions with the best outcomes were being made on behalf of any given patient. For this discussion, utilization management is considered the global concept that subsumes various aspects of the review process. These days, virtually all care, except for the increasingly rare full fee for service, is under some form of review and management process, whether public or commercial payer.

Some practitioners consider the review process an intrusion into the clinician–patient relationship. The process can be cumbersome, and disagreement about the decisions reached during a review often contributes to resentment. With some justification, practitioners are irked because third parties with no clinical relationship to the patient have the authority to influence the process of care, challenging the practitioners' clinical autonomy. At one point, the clinical criteria used in the review process were considered "proprietary" and not available to clinicians, so they might see how their care process was being reviewed. At the same time, if practices were reasonably standardized and if practitioners had training (or retraining) that was uniform with a focus on (1) diagnostic accuracy, (2) evidence-based, cost-effective practice, and (3) attention to clinical outcomes, perhaps a great deal of the variance seen in practice might diminish, thereby reducing the need for third-party reviews.

Selective Closed Panels and Networks

Another strategy to improve care while reducing costs was to create selective or preferred provider networks in which clinicians applied to become members of the clinical panel of a managed care organization. By applying to be a network provider, clinicians agreed to practice within the guidelines established by the managed care organization. In theory, the network had some degree of compliance and fidelity to the standards of practice of that particular managed care company. However, this was more of an idealized system of care for several reasons. In many parts of the country, the clinical marketplace is highly competitive because of a glut of practitioners vying for limited clinical dollars. Functionally, this means that many individual practitioners may apply to a particular managed care panel because of simple economics (i.e., they need the income) rather than an interest or willingness to practice within a managed care framework. So, there may be little selectivity about a particular panel of providers because the process is more open-ended until the managed care organization has determined there are sufficient providers for a given geographical area. Whether justified or not, some managed care organizations have reputations for being difficult to work with, reimbursing at what is viewed as “low rates.” As a result, such organizations may have difficulty attracting sufficient providers and may be less selective than is preferable. In more affluent areas, some practitioners, especially well-established ones, do not accept insurance of any kind, so utilization review is not part of the overview of their patients’ clinical care.

Care Management and Intensive Case Management

Much psychiatric care is done with persons with chronic mental health conditions. Driven mainly by the demographic realities of an aging population, health reform considerations focus on chronic illness care and the concept of care management and intensive case management, virtually identical concepts. By analyzing the process of care for various disease entities, care management attempts to introduce discrete improvement steps to improve the efficiency and coordination of care, enhance patient compliance, achieve better outcomes, and reduce costs. Care management has been used with chronic illnesses, such as diabetes or hypertension. In psychiatry, care management has been successfully applied to mood and anxiety disorders. However, the high degree of practice variability not only by the provider but also by a geographic section of the country reflects the historic inability and failure of the health care system to incorporate readily and to standardize quality improvement and evidence-based ideas into the system in which clinician, patient, and health system engage.

Managed Behavioral Health Care

In the late 20th century, costs for psychiatric and substance-abuse services rose considerably faster than the costs for physical health care. The effort to instill accountability among practitioners, expect value for payment, improve access, provide cost-effective services, and use to the extent possible evidence-based approaches to treatment to improve outcomes created a backlash by mental health practitioners against managed care. Scandals in the 1980s and 1990s of for-profit managed behavioral health care companies engaging in unethical practices overshadowed the equally scandalous accounts of practitioner excesses of use and inappropriate, often misguided, treatment. Managed behavioral health care became synonymous with denial of care and, in for-profit systems, the image of fat-cat corporate leaders with large salaries and generous stock options derived from the misfortune of patients and beleaguered practitioners, the latter mired in an endlessly confusing system of external review, was readily accepted by many as the only reality. What was missing in the tumult was the perspective that managed behavioral health care was a way station on the path to behavioral health care reform.

Public Sector Managed Behavioral Health Care

Governmental agencies purchase 60% of all behavioral health care at federal, state, and local levels. As resources followed patients out of the state hospitals, much of the state expenditures remained controlled by agencies responsible for specific populations such as mental health, criminal justice, child welfare, and social services. Medicaid, the entitlement program for those unable to afford health care, began in 1965 and is the primary funding source for local governments to pay for public behavioral health care services. Medicaid's overall costs rose dramatically by 2000, accounting for more than 15% of all U.S. health care costs, with behavioral health care accounting for about 10% of Medicaid spending.

Medicaid's behavioral health costs became a severe liability at the state or county government level because of yearly cost overruns. As a result, government agencies turned to managed care contracting of Medicaid services to shift risk to nonprofit behavioral health care provider organizations, thus limiting the state's liability.

State governments also soon learned that they could redirect the bulk of their behavioral health expenditures for severely disabled persons in the categorical agencies into the Medicaid program, thus increasing their match base for the Medicaid Federal Financial Participation (FFP) and leveraging federal revenues into their state general fund. Medicaid expenditures continued to increase, and this was a major contributing factor to the 100% increase in public mental health dollars from 1956 to 2003. The downside from the point of view of the seriously and persistently mentally ill was that the control of these services went from the public agencies responsible for behavioral health care to commercial managed care companies whose networks and capacities were more in tune with the needs of employed populations.

Safety-net hospitals, community mental health centers, and academic medical centers were under extreme financial distress and endured substantial service limitations. Thus, while public sector experiments with large-scale managed behavioral health care produced some successes, costs, and quality of care continued to be problematic.

Private Sector Managed Behavioral Health Care

Government deregulation of the hospital industry enabled the astronomical growth of for-profit mental health and substance abuse facilities, many organized as large hospital chains. As a result, the use of mental health and substance abuse treatment increased dramatically, exemplified by the fourfold increase in adolescent inpatient admissions between 1982 and 1987. The proliferation, financial success, and perceived abuses of proprietary psychiatric hospitals caused great consternation among employers, insurers, legislators, and the public. Many hospitals, including presumably less exploitative, not-for-profit hospitals, seemed to provide treatment based on reimbursement potential rather than actual medical necessity. One of the physician pioneers of the managed behavioral health care industry commented that it was a myth that substance abuse treatment required a fixed length (usually 28 days) of inpatient treatment. Rarely has such a fixed false belief had so little factual support or received such fervent adherence. As early as 1983, a report from the Office of Technology Assessment that the literature failed to demonstrate a general superiority of inpatient over alternative treatment settings, as measured by outcome. Subsequent studies have confirmed that hospital or residential treatment is neither necessary nor sufficient for most people with substance use disorders.

Nevertheless, long inpatient stays for adolescents with behavioral or substance abuse problems became common. Inappropriate hospitalizations for excessive periods became exposed as proprietary systems flourished. Health care payers (employers, insurance companies, Medicare, and Medicaid) became quite rightly concerned by the rising costs and the suspect practices. They questioned the actual medical necessity of costly services such as extended psychiatric inpatient stays. The psychiatric profession also successfully argued that inpatient stays for psychiatric patients could not be actuarially determined with any precision, thus making federal Medicare mental health inpatient criteria exempt from federally determined diagnostic-related groups (DRGs) for reimbursement purposes. In this climate, it became more and more challenging to maintain adequate oversight or regulation of mental health and substance abuse treatments. Private employers became the first to respond proactively to these concerns by contracting with the pioneer entrepreneurs of the first generation of managed behavioral health organizations (MBHOs) led by mental health professionals who had the credibility and specialized knowledge and skill to provide oversight of mental health and substance abuse practitioners.

This initial phase of managed behavioral health care (approximately 1983 to 1987) began with traditional indemnity insurers changing benefit designs in the hope of discouraging excess use. Tactics included limiting mental health coverage for specific disorders or treatments and instituting deductibles, copayments, and benefit maximums. These efforts were ultimately all unsuccessful. Entrepreneurial companies, led by psychiatrists and psychologists, began their entry into the health care marketplace through contracts with HMOs struggling to administer behavioral health care benefits. These initial efforts' dramatic success was driven by the limitation of access to mental health and substance abuse services. Access was limited through tight medical management (e.g., instituting strict medical necessity criteria for level-of-care determinations, limiting authorizations for inpatient admissions and continuation of stay, or the proliferation of denials of payment for care), as well as by restricting the network of caregivers and institutions credentialed to provide care for the population covered.

The first phase of limitation of access, aimed at restricting access for patients, was successful but led to dramatic customer complaints. As large employers (not HMOs) became the target for MBHO marketing, patient choice was increased by expanding provider networks, but overall costs were now saved through a dramatic reduction in rates paid to providers. Typical discounts were 20% to 25% of professional fees and 25% to 35% of hospital charges. These discounts seemed Draconian at the time and were widely seen as threats to behavioral health care providers' solvency, but more significant fee reductions were to come. This network expansion and reduction of provider rates characterized the second phase of the managed behavioral health care revolution (approximately 1993 to 1998).

The third phase (1998 to the present) has been termed the phase of industry consolidation. The proliferation and financial success of MBHOs in the 1980s and early 1990s led to the development of administrative models, information system infrastructure, extensive provider network management techniques, and sophisticated medical management strategies that resulted in significant economies of scale. Nevertheless, provider rates continued to fall, along with profit margins. Utilization patterns also changed dramatically. In tightly managed markets (such as Minneapolis and Los Angeles), the average commercial inpatient days per 1,000 members declined: from approximately 150 in 1983, 75 in 1990, and 10 to 15 in 2000. The cost to employers of behavioral health care under traditional indemnity plans was approximately $9 per member per month. By 2000, these costs fell dramatically to less than $3 per member per month. To sustain these reductions, the MBHO industry experienced dramatic consolidation. Within a few years, through mergers, acquisitions, and market growth, the industry became concentrated into a small number of very powerful behavioral health care corporations.

HEALTH CARE REFORM IN THE 21ST CENTURY

Whether admired or not, many of managed care's innovations are irreversible. Managed care's financial success was dramatic (i.e., decreased behavioral health care costs). Utilization management by managed care companies led to significant decreases in the average length of inpatient stays for psychiatric patients and fewer total inpatient days for populations of patients as a whole. Despite widespread complaints of diminished physician, hospital, and patient autonomy, constraints of decreased physician and hospital reimbursement, and anecdotal horror stories of sacrificed patient health care quality, there is no systematic way of convincing lawmakers that managed care destroyed the quality of the U.S. health care system and no easy way to reverse the changes that have happened to medical practice in general and psychiatric practice in particular. One could reasonably argue that the health care system was already imploding, which was the impetus for yet another effort at reforming the system. The National Committee on Quality Assurance (NCQA) points to significant improvements in health care quality resulting from efforts at self-improvement by the managed care industry when these are conducted properly. Managed care contributed to health care quality by emphasizing population-, evidence-, and guideline-based health care, chronic illness care management, and information technology to evaluate and change the process and outcomes of care.

Table 56.2–2.
Comparison of the U.S. Health Care System to Other Industrialized Countries

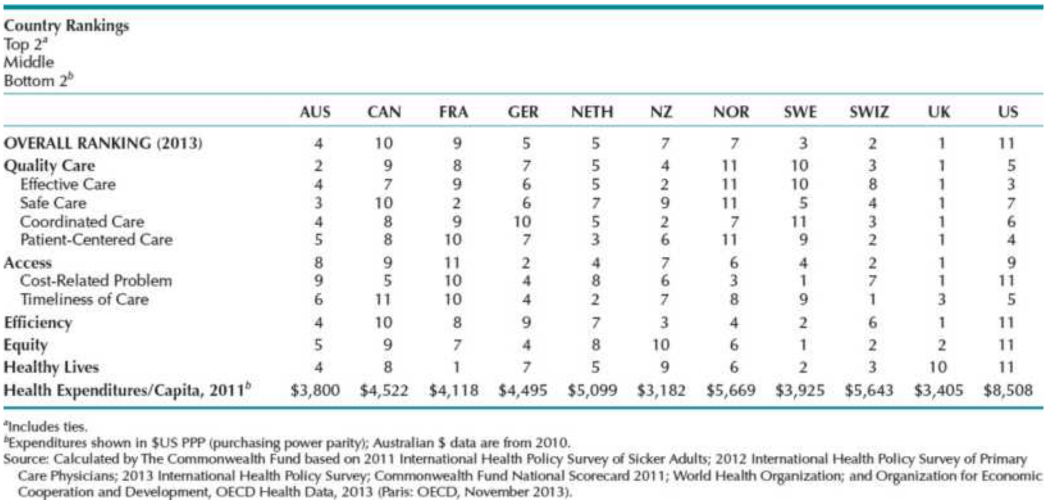

Country Rankings
Top 2[a]
Middle
Bottom 2[b]

	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US
OVERALL RANKING (2013)	4	10	9	5	5	7	7	3	2	1	11
Quality Care	2	9	8	7	5	4	11	10	3	1	5
Effective Care	4	7	9	6	5	2	11	10	8	1	3
Safe Care	3	10	2	6	7	9	11	5	4	1	7
Coordinated Care	4	8	9	10	5	2	7	11	3	1	6
Patient-Centered Care	5	8	10	7	3	6	11	9	2	1	4
Access	8	9	11	2	4	7	6	4	2	1	9
Cost-Related Problem	9	5	10	4	8	6	3	1	7	1	11
Timeliness of Care	6	11	10	4	2	7	8	9	1	3	5
Efficiency	4	10	8	9	7	3	4	2	6	1	11
Equity	5	9	7	4	8	10	6	1	2	2	11
Healthy Lives	4	8	1	7	5	9	6	2	3	10	11
Health Expenditures/Capita, 2011[b]	$3,800	$4,522	$4,118	$4,495	$5,099	$3,182	$5,669	$3,925	$5,643	$3,405	$8,508

[a] Includes ties.
[b] Expenditures shown in $US PPP (purchasing power parity); Australian $ data are from 2010.
Source: Calculated by The Commonwealth Fund based on 2011 International Health Policy Survey of Sicker Adults; 2012 International Health Policy Survey of Primary Care Physicians; 2013 International Health Policy Survey; Commonwealth Fund National Scorecard 2011; World Health Organization; and Organization for Economic Cooperation and Development, OECD Health Data, 2013 (Paris: OECD, November 2013).

However, by the beginning of the 21st century, health care costs continued to spiral upward in the United States, and health care quality remained substantially below standards set in other comparable affluent societies (see Table 56.2–2). With overall health care costs rising at more than twice the inflation rate, employer groups could not sustain the resultant increase in premiums.

Waste in Health Care

In recent years, considerable attention has been paid to eliminating waste in U.S. health care, not just redesigning a dysfunctional health care system with improved access, better quality, improved efficiency, expanded coverage to enhance equity, and lowered costs. It has been estimated that a third of the annual health care expenditure ($2.8 trillion in 2014) could be eliminated without any adverse health consequences to individual patients. Unnecessary administrative expenses, care that provides little or no value, costs associated with medical errors, and fraud are examples of waste. There have been six areas identified that are responsible for the waste:

1. poor care delivery
2. poor care coordination
3. excessive care
4. excessive administration
5. pricing failures
6. fraud and abuse

These six areas were projected to push U.S. health care expenditures to 20% of the gross domestic product by 2020. Not surprisingly, therefore, in the first two decades of the 21st century, the urgency of health care reform has intensified as the growth of U.S. health care spending is projected to double in 8-year intervals representing almost 18% of the gross domestic product in 2011. The competitive position of the United States in the global economy is threatened, and many argue that the economy cannot be fixed until health care is fixed. Influential reports show that the United States spends considerably more on health care than any other country. This growth in spending consistently outstripped collective income. Yet the United States ranks poorly among industrialized countries in standard measures such as life expectancy and infant mortality. Moreover, it cannot provide health insurance to at least 15% of its population. Therefore, health care reform will remain at the forefront of the domestic agenda as long as the path of sustainability has not been achieved.

THE IMPACT OF THE AFFORDABLE CARE ACT (ACA) AND THE TRIPLE AIM DEFINITION—APPLICATIONS TO MENTAL HEALTH

“Between the health care we have and the care we could have lies not just a gap, but a chasm.” IOM, 2001

The key provisions of the ACA are designed to:

- Provide coverage for more people
- Provide more benefits, better protection, and transparency
- Lower costs for consumers and government
- Provide quality and efficiency
- Provide a stronger workforce and health care infrastructure.
- Provide a greater focus on public health, prevention, and chronic illness care
- Support innovative techniques and technologies

Meeting the three inter-related concepts (1) improving the experience of care, (2) improving the health of populations, and (3) reducing per capita costs of health care is required to achieve high-value care as reform shifts from a *volume-based* to a *value-based* system. The passage of the ACA as law in 2010 laid the foundation for health care transformation. Key policy reports in the first decade of the 21st century sparked the need for the ACA, including the influential report of the Institute of Medicine (IOM), in 2001: “Crossing the Quality Chasm: A New Health System for the 21st Century.” The report highlighted the disparity between best practice and current practices in the United States and the effect on widespread variability and escalating health care costs. It established Six Aims for the 21st-century health care system, espoused by the ACA.

The larger health care systems in which psychiatrists and other behavioral health specialists work profoundly impact the delivery and outcomes of psychiatric and physical health care services. Internationally, and over the last two decades in the United States, the dominant theme in health care system reform shifted from an emphasis on managed care to a tripartite focus on patient-centered care (as a complement to provider-driven disease management), improvement of the health of entire populations with interventions tested in comparative effectiveness research (CER) (as opposed to individual-level observational or clinical trial research), and the development of incentives to maximize health care cost-effectiveness (as opposed to containing costs by restricting access to or reimbursement for health care). The Triple Aim has served as a foundation for health care reform, and we are now in a place to examine some of its effects a decade later (see Fig. 56.2–1).

The guiding principle of the Triple Aim is that health care quality should be defined based upon the vantage of the recipient, *the patient*, and not externally by technical or clinical considerations. Beginning with a patient-centered (as opposed to disease-based) perspective, health care must be teamed up with public health (i.e., prevention and health promotion) and social services to fully meet the needs of individuals and families and create a healthy overall population. This biopsychosocial approach integrates otherwise fragmented or missing services and makes primary (rather than specialty) health care the driver for delivering services—containing and reducing costs and promoting health. The implementation of this system has been arduous and carried its own challenges. The concepts of patient-centeredness and equity are not new or radical but have been recurrent themes in health care reform efforts in the United States for more than a century. They have been the bedrock and the foundation in the more successful health care systems in other affluent and industrialized countries internationally.

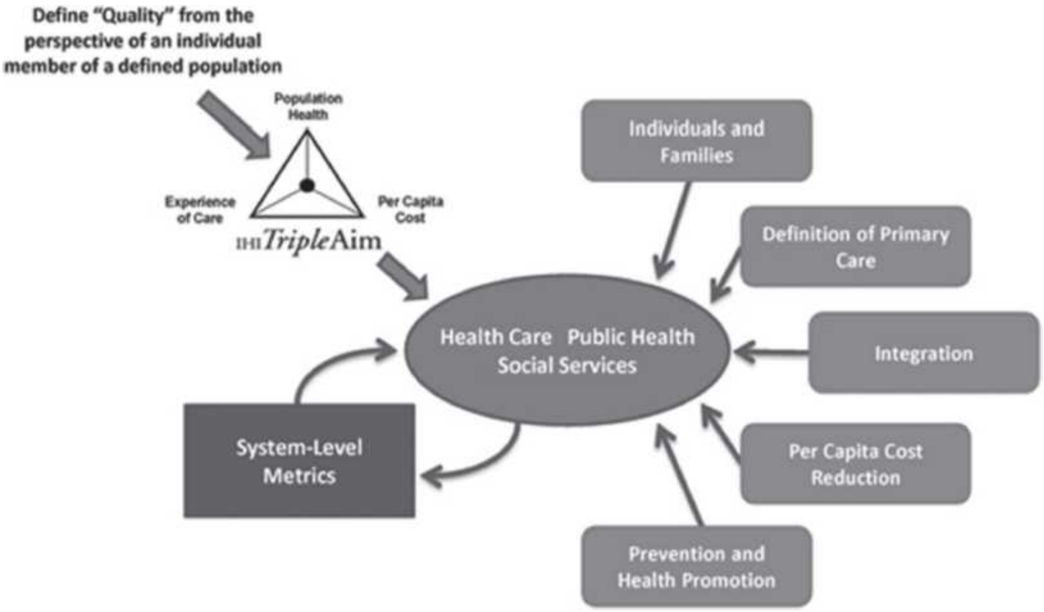
Define "Quality" from the perspective of an individual member of a defined population
Population Health
Experience of Care
Per Capita Cost
IHI TripleAim
Individuals and Families
Definition of Primary Care
Integration
Per Capita Cost Reduction
Prevention and Health Promotion
Health Care Public Health Social Services
System-Level Metrics

FIGURE 56.2–1. Design of a triple aim enterprise. (Institute for Healthcare Improvement. Design of a Triple Aim Enterprise. 2012. https://www.ihi.org/Engage/Initiatives/TripleAim/pages/default.aspx.)

For psychiatry, specifically, the impact of patient-centered care is far-reaching given the fact that individuals with serious mental illness (SMI) represent an undue large portion of underinsured or uninsured patients and have multiple co-occurring nonpsychiatric disorders, e.g., medical, addiction disorders, in addiction to negative social determinants of health.

Expanding the concept of best practice to include public health and prevention efforts and coupling this with chronic illness care management are among the key points of the IOM monograph and are aligned with ACA triple aim, as are the pivotal roles of preprofessional training the retraining of the existing workforce. With finite resources for health care amid competing priorities, the reformed health care system and, specifically, the reformed behavioral health care system must find ways to integrate effectively multiple factors, such as (1) practitioner, system, and patient accountability; (2) economic support for public health efforts, such as prevention and early identification; (3) the linkage of acute care with chronic illness care/collaborative care/coordinated care management; (4) improved access; (5) effective integration with primary and specialty care; (6) development of models of the most clinically effective and cost-effective care (e.g., integrated with public and physical health, interdisciplinary in practice, with continuity of care across venues); and (7) meaningful involvement of patients and, where feasible, their families, in the process of care. Managed care can be viewed as a sign and a symptom within the broader construct of health care reform. The creation of medical homes has been the *first step* toward this integration.

Numerous studies have shown the high prevalence of behavioral health conditions in physical health. Patients with comorbid illnesses (mental, substance use, and physical health disorders) represent the most costly group in health systems. Many of these individuals are seen in public sector systems, often incarcerated or homeless with inadequate physical and behavioral health care.

A dramatic example of the need to find ways to integrate primary care and behavioral health care in coherent delivery systems is the “dual eligibles,” that is, the Medicare and Medicaid eligible population. This group is among the most costly of any population of patients, not only because a large portion of the poor and very older adult patients are in long-term care (nursing homes and hospice), but also because those in the community find it difficult to negotiate consistent access to health care due to their disabilities. Typically, a given state might find that 40% of their dual-eligible population have serious and persistent mental illnesses.

Sadly, real obstacles remain in reforming health care related to mental health outcomes. The National Committee for Quality Assurance (NCQA) reported that despite essential strides made in the general medical and surgery sectors, there continued to be disturbing exceptions to overall health care quality improvement in mental health, that the situation remained poor and concerning with individuals afflicted with SMI dying 7 to 10 years earlier than their non-SMI counterparts. The IOM’s recommendations centered first on linking mental health and substance use treatment more closely with general medical care; second, the details of the original Quality Chasm report, its aims, rules, and redesign approaches, should be operationalized in the treatment of mental health and substance use disorders. Inadequate leadership and resources have been cited along with these areas having fewer objective, standardized diagnostic methods compared to physical health disorders; weakness of the evidence base for supporting quality measures; lack of coherent strategies for adopting quality measures that get back to the problems in leadership in assuming this responsibility to move the field along. *Fewer than 5% of the National Quality Forum’s list of 650 quality indicators relate to the care for individuals having mental health and substance use problems*.

The 2001 IOM Quality Chasm Report articulated a strategy for health care improvement in a separate report for mental and substance use disorders because it has numerous distinctive characteristics. These distinctive characteristics include the greater stigma associated with mental and substance use conditions, more frequent coercion of these patients into treatment, a less developed system for measuring quality and improvement, a more educationally heterogeneous workforce, the fragmentation of the behavioral health care system with inadequate communication and coordination of care among its components, and a perverse funding structure.

In its report “Improving the Quality of Health Care for Mental and Substance Use Conditions” (2006), the 10 items listed as guidelines/priorities for mental health and substance use conditions are as follows:

1. Care is based on continuous healing relationships.
2. Care is customized to patient needs and values.

3. The patient is the source of control.
4. Knowledge is shared, and information flows freely.
5. Decision-making is evidence-based.
6. Safety is a system responsibility.
7. Transparency is necessary.
8. Needs are anticipated.
9. Waste is continuously decreased.
10. Cooperation among clinicians is essential

The report also noted the considerable societal and workplace burden that mental and substance use conditions bring with substantial employer costs from absenteeism, “presenteeism” (i.e., being at work but essentially not functioning fully), sub-par educational achievement by children having such conditions leading to a variety of emotional and behavioral problems. Children with poor educational achievement enter a cycle of emotional and behavioral problems leading to risk for delinquent and antisocial behaviors, dropping out of school, eventually affecting the child welfare and juvenile justice systems. This, in turn, becomes a feeder phenomenon into the adult justice system that has seen an increasing number of inmates in prisons and jails with mental and substance use conditions. The U.S. Bureau of Justice Statistics estimated that 16% of all persons in jails and prisons have a mental or substance use disorder. Systematic assessment studies document prevalence estimates that are twice as high, with as many as one in three incarcerated adults and youths having a behavioral health disorder.

An overriding theme throughout the report was that treatment can be effective and that recovery is possible. The report noted that quality problems hinder effective treatment and recovery, noting that numerous studies have shown the discrepancy between what is considered effective care and what is provided, echoing the Quality Chasm report. Low adherence rates to clinical practice guidelines and evidence-based practice standards were noted across major diagnostic categories. For example, adequate adherence rates to established practice guidelines occurred in only 27% of studies covering alcohol withdrawal, bipolar disorder, depression, panic disorder, psychosis, schizophrenia, and substance use.

Parallel programs targeting the same priorities for integrative mental health care in the United States are discussed below.

The Chronic Care Initiatives in Health Maintenance Organizations

These aim to change physician and patient behaviors in several ways, including

- Moving from a focus on a cure to symptom relief
- Expanding the boundaries of care among providers and between traditional medical and social services

- Making the patient a critical part of the care process through shared decision-making:
    - Creating active roles for the patient’s significant others
    - Tying patient satisfaction to clinical outcomes

Building Health Systems for People With Chronic Illnesses

Sponsored by the Robert Wood Johnson Foundation, this program attempts to better coordinate medical and supportive services for the chronically ill and, in so doing, challenges the fact that most insurance systems tend to fund acute care rather than supportive care for individuals with chronic conditions. The program attempts to do the following:

- Integrate services that care is provided across diverse settings; systems of care link clinical and nonclinical support services.
- Reallocate resources by redirecting existing and acute care—focused delivery and financing to serve the chronically ill better.
- Promote early medical and nonmedical interventions.
- Help individuals maintain independence as long as possible in their homes and communities.
- Give the chronically ill and their caregivers a role in designing and improving their health care and support services.
- Support the development of models of care that people with different chronic health conditions can use.

The demonstration program in the accompanying text box set the stage for two operational models for fundamental system changes to improve quality and outcomes. Many proposed and expected system reforms surround efforts to implement the chronic illness care model and the collaborative care model. The chronic illness care model was first elaborated and tested for chronic general medical illnesses such as diabetes, congestive heart failure, and asthma. Recent efforts have attempted to apply the principles of chronic illness care to depression in primary care settings. On the other hand, the collaborative care model was first described and tested for depression in primary care and has been broadened for other psychiatric illnesses that can also be treated in general medical settings with care managers and psychiatrists’ support, such as panic posttraumatic stress disorder.

The 2008 Mental Health Parity and Addiction Equity Act

The Mental Health Parity and Addiction Equity Act enacted in 2008—the Wellstone-Domenici Parity Act—tightened the requirements for parity laid out in the Mental Health Parity Act of 1996. Parity was designed to bring individuals with behavioral health disorders into the mainstream U.S. health care system to correct the historical inattention and inequality of care in this population and reduce the stigmatization such individuals experienced compared to people having physical health problems alone. The intent was to have behavioral health care benefits equivalent to medical and surgical care benefits. In addition, about two-thirds of primary care physicians had reported difficulty in accessing outpatient behavioral health services for their patients, a rate twice as high as that for other services. Lack of coverage or inadequate coverage, health plan barriers, and a shortage of behavioral health providers were cited by primary care physicians as problems contributing to accessing good behavioral health care. The Mental Health Parity and Addiction and Addiction Equity Act required group health plans with behavioral health benefits to provide the same level of coverage as physical health plan coverage. Access to behavioral health services has been particularly problematic for low-income uninsured and Medicaid patients. Ironically, primary care physicians practicing in areas with many psychiatrists had a greater probability of not securing behavioral health services for their patients because of various plan barriers and the inadequacy of coverage these patients had. Recent studies show that psychiatrists, compared to all other areas of medicine, have the lowest acceptance rate of Medicare (i.e., less than 50% of psychiatrists accept Medicare, with figures lower still for Medicaid). The speculation about such an observation its effect on how other specialties, professionals, trainees, policymakers, and the general population, view adherents to such a practice is important to consider. Some have advocated tying partial or total medical school debt-forgiveness, or even state medical licensure, to having a reasonable percentage of one's practice tied to practitioners seeing individuals covered by Medicare and Medicaid.

The ACT's effect was the practical disappearance of quantitative treatment limits (QTLs), which insurance companies frequently used to limit such treatments and improve coverages for substance use and mental health conditions. More help-seeking in adults suffering from mental health disorders, substance abuse, and eating disorders were noted after parity laws were mandated but with the main caveat that out-of-pocket expenditures also increased. From a limited perspective, the parity acts of 1996 and 2008 were successful. However, from a more global view, they still did not meet the needs of reform in terms of care access, administrative burden, and, more importantly, equity needs (Table 56.2–3).

IMPROVING MODELS IN HEALTH CARE: THE ACA A DECADE LATER

Patient-Centered Medical Homes and Medical Neighborhoods

In 2007, the American Academy of Family Physicians, the American Academy of Pediatrics, the American College of Physicians, and the American Osteopathic Association issued joint principles for the Patient-Centered Medical Home (PCMH), a model to provide comprehensive primary care to children and adults, where appropriate care is structured, delivered, and coordinated around the specific needs of each patient. The American Academy of Pediatrics originally developed the medical home model in 1967 for children with special health needs, where it was essential to have a centralized location for a child's medical records. The joint principles of PCMHs (personal relationship with licensed providers, team approach, comprehensive/whole-person approach, coordination and integration of care, quality and safety, expanded access to care, added value recognized, are all hallmarks) cover many of the characteristics of Accountable Care Organizations (ACOs) where PCMHs are seen as the building blocks for ACOs.

Table 56.2–3.

Summary of the Institute of Medicine's Recommendations of Care for Individuals With Mental and Substance Use Condition

1. Emphasize patient-centered care and the importance of supporting patient decision making and preferences	The current health systems place a number of barriers to patient-centered care. Insurance discrimination is manifested by less benefit coverage for mental and substance use (M/SU) conditions. There are higher copays for these services. ▸ Clinicians, organizations, accrediting bodies, health plans, and payers need to support a patient-centered care (PCC) approach that respects decision-making capacity of patients. ▸ Stereotypes of patients being unable to arrive at decisions have resulted in coercion of care and discrimination historically.
2. Improve the infrastructure for quality improvement and evidence-based care	Significant problems with the current infrastructure for quality improvement. Assessment and treatment practices not well codified and captured in administrative data sets. Outcome measurements not widely applied; performance measurement for M/SU health care receives insufficient attention in the private sector; public sector efforts have not yet achieved consensus. Quality improvement methods are not permeating day-to-day service. This leads to systems of care that are not based on the principles of quality measurement, and they fail to demonstrate continuous improvement.
3. Coordinate care across systems of care	Poor linkages across systems of care; there is general fragmentation between M/SU and general health care and between mental and substance use care for both children and adults. Unclear accountability for coordination of care among providers.
4. Ensure that the National Health Information Infrastructure Initiative (NHIII) benefits people with mental and substance use conditions	M/SU conditions are not well incorporated into the strategic design of the NHIII. The proposed solutions are as follows: ▸ Charge the Office of the National Coordinator of Health Information Technology and the Substance Abuse and Mental Health Services Administration (SAMHSA) to jointly develop a plan that addresses M/SU health care.

	▸ Leaders in M/SU health care should become proactive in the national process to set health care data and information technology standards. ▸ Public and private purchasers of health care should decrease the burden of variable reporting and billing requirements. Public and private purchasers of health care should encourage widespread adoption of electronic health records, computer-based clinical decision support systems, computerized order entry, and other forms of useful information technology.
5. Increase the workforce capacity	A crisis in the M/SU workforce exists. There is greater variation in the quality of the M/SU workforce than in that for general health care. There are across-the-board deficiencies in education, and there is no “core knowledge” base across disciplines. The resulting variation in licensure, credentialing, and continuing education across disciplines does not assure agreed-on core competencies or adherence to the same standards of practice. The remedies include the following: ▸ Creation of an ongoing, federally funded public–private Council on the Mental and Substance Use Health Care Workforce. ▸ Collaboration of the Council with institutions of higher education, licensing and accrediting bodies, purchasers, and other private sector initiatives developed by the Association for Medical Education and Research in Substance Abuse (AMERSA) and the Annapolis Coalition (http://www.annapoliscoalition.org).
6. Restructure the marketplace to promote incentives for quality and improvement	The Institute of Medicine (IOM) report finds problems in the way in which M/SU care is purchased. Government (state and local) purchasers dominate, and frequently insurance is purchased for M/SU health care separately from general health care. This results in “carve-out” arrangements. Private insurance tends to avoid coverage or offers more-limited coverage to

individuals with M/SU illnesses. Government purchasers often use direct purchasing by contract rather than insurance arrangements.

- Government should pass laws designed to enact coverage parity by purchasers so that individuals with M/SU conditions are not discriminated against when enrolling in health plans.
- Examples of tools that could be used to bring about parity are risk adjustment, payer carve-outs, risk-sharing or mixed-payment contracts, and benefit standardization across health plans.
- State-level procurement for M/SU care should give the greatest weight to quality and use M/SU health care quality measures in procurement and accountability.
- State and local governments should reduce emphasis on grant-based systems of financing and increase the use of funding mechanisms that link some funds to measures of quality.

Data from the Institute of Medicine. *Crossing the Quality Chasm: Adaptation to Mental Health and Addictive Disorders*. National Academies Press; 2006.

Some of the more recent literature suggested that early PCMH interventions led to only modest improvement in quality and patient experience, with little evidence of effects on costs and no reductions in utilization of hospital, emergency, or ambulatory care services over 3 years. At the same time, these studies suggested that refinement of the outcomes associated with PCMHs was needed and that the development of the NCQA standards for medical home recognition had not been fully developed applied in the early PCMH development. Nonetheless, the studies concluded that further PCMH development and function refinement was necessary.

As the conceptualization of the organization of health care delivery systems developed, it was logical that the PCMH model would expand to include a constellation of entities and activities responsible for the care of populations of patients. Thus, the concept of the Medical Neighborhood was spawned encompassing PCMHs with comprehensive primary, secondary, tertiary services that are community-based, responsive to the needs of individual patients at different times, a truly integrated system of care with good coordination, effective communication, attention to patient education, and cost-effective. Such a concept forms the basis for accountable care.

Accountable Care Organizations

With the advent of the ACA, the Department of Health and Human Services (DHHS) began developing new approaches in conjunction with the private sector to improve care and to meet the Triple Aim for Medicare and Medicaid recipients. ACOs are voluntary groups of physicians, hospitals, and other health care providers that assume responsibility for a defined population of patients based on the population's use of primary care services. The ACO is responsible for the entire continuum of care that the defined population of patients might need, including access to specialists, social services, public health, and assessment of other socioeconomic factors such as housing, transportation, education, poverty, and nutrition. Providing preventive care for all patients and case management for individuals with chronic diseases is a component of an ACO. If the ACO succeeds in delivering high-quality, improved care at a lower cost, there is an incentive built into the model whereby the ACO can share in the savings achieved through meeting the Triple Aim. Under the ACA, an ACO agrees to manage the health care needs of a minimum of 5,000 Medicare beneficiaries for at least 3 years. Under an ACO, providers are held accountable for the health of their patients and have strong incentives to cooperate and create savings by avoiding unnecessary care (see below). Functionally, ACOs need to be fully integrated through an electronic medical record to share needed information among providers. Different ACO models exist to give hospitals and clinicians choices in working with one another. This process, initiated in 2010, is still ongoing with many hurdles.

The Teaching Health Center

There has been interest in providing post-graduate medical education (post-GME) in the primary care environment of the Federally Qualified Health Centers for some years. These programs have demonstrated that physicians trained in these settings are three times more likely to work in Federally Qualified Health Centers (FQHCs) and twice as likely to serve underserved areas of the country. The ACA provided funds to establish Teaching Health Centers training physicians in nine primary care–related disciplines (including Psychiatry). In the same vein, the IOM issued a report in 2014, "Graduate Medical Education That Meets the Nation's Health Needs," calling for the reform of financing, governance, and transparency of GME. The Report called for a 10-year transition period in which the current funding for more traditional residency programs will be modified. As a result, transformative and innovative residency programs will be created and supported in primary care settings and interdisciplinary organizations providing intensive care programs for persons with severe disabilities and multiple chronic illness conditions.

On the surface, the idea of involving learners in the new model was attractive, but the funding was not consistent, and funding instability limited the impact of this program. It remains an attractive target for further reform. The Teaching Health Center can also meet the need for medical innovation programs in rural areas, improving processes on multiple levels and using alternative funding.

Comparative Effectiveness Research and Patient-Centered Outcomes Research

CER and patient-centered outcomes research (PCOR) are essential to fulfilling each of the three goals of the Triple Aim and correspondingly have been allocated substantial funding support by the ACA through the Patient-Centered Outcomes Research Trust Fund (PCOTF). PCORTF received funds from the general fund of the Treasury and a small fee assessed on Medicare, private health insurance, and self-insured plan revenues. The PCORI received approximately $3.5 billion from the PCORTF over ten years ending September 30, 2019 and used those funds to build the capacity of the health care system to provide and evaluate patient-centered and outcome-driven services. The remaining 80% of PCORTF funds were transferred to the Agency for Healthcare Research and Quality (AHRQ) to support and evaluate PCOR in U.S. health care organizations.

PCOR is intended to provide methods and data repositories that can inform health care services by monitoring key health outcomes. Critical research questions for this agenda include how patients, families, and clinicians deal with the complex and often confusing choices involved in achieving and sustaining health, preventing and treating diseases and other health problems, and accessing and navigating health care systems. Traditional medical research has produced vital advances in understanding disease conditions and their treatment. However, it has not been intended or designed to enhance the current understanding of how patients subjectively experience, participate in and are affected by health care and interactions with health care providers. Patient-centered medical research does exist, but its results typically are not systematically disseminated to either clinicians or patients and families. Research linking the patient's involvement in health care and health outcomes also exists. However, its findings have not been extended beyond demonstrating that a collaborative provider–patient relationship plays a crucial role in patient motivation, adherence, and clinical and functional outcomes. PCOR takes on these critical questions that patients and their clinicians face daily in health care encounters and their personal lives. PCOR is a systematic approach to patient-centered CER, in which patient and clinician preferences, choices, actions, and interactions are studied to determine tools, strategies, resources, and courses of action associated with cost-effective health care outcomes. For patients, PCOR involves testing and then providing guidance about which approaches to care work best, given their particular health conditions, personal and familial concerns, socioeconomic circumstances, and sociocultural preferences. PCOR is designed to provide evidence-based guidance addressing the questions they face daily in practice for clinicians. For insurers, PCOR produces evidence to help them make the best decisions on improving health outcomes for their members while also containing costs, especially with high-burden, high-impact disease conditions.

For psychiatry and the broader mental health field, PCOR provides opportunities and an impetus to conduct real-world clinical research on topics such as these recently funded PCORI projects:

- Integrative Medicine Group Visits: A Patient-Centered Approach to Reducing Chronic Pain and Depression in a Disparate Urban Population
- A Community-Based Executive Function Intervention for Low-Income Children with ADHD and ASD
- Improving Transitional Care Experience for Individuals with SMI
- A Helping Hand to Activate Patient-Centered Depression Care among Low-Income Patients (AHH)
- Integrated Care and Patient Navigators for Latinos with SMI
- Patient-Centered Trauma Treatment for PTSD and Substance Abuse: Is It an Effective Treatment Option?

Pay for Performance, Public Reporting of Performance, and Meaningful Use

The effort to improve quality and, therefore, outcomes has taken many guises, from revamping training to produce clinicians and systems that practice in a quality paradigm to imploring clinicians and systems to do the right thing. At its most basic level, change can also be brought about by behavioral shaping and the use of incentives and disincentives to achieve the desired result—in this case, quality improvement and patient-focused, high-quality care. Two such initiatives, Pay for Performance (P4P) and Public Reporting of Performance (PR), are designed to achieve these results. P4P and PR have been common in physical health care for several years. CMS has adopted P4P and has launched initiatives to demonstrate different ways to implement it. More than half of commercial health maintenance organizations use P4P, and the acceptance of performance measurement among physicians and organized medicine has been acknowledged if not fully realized. P4P and PR, in essence, link physician reimbursement and nonmonetary rewards to improve the quality of care. In addition, physicians recognized for providing high-quality care and more efficient care deliver it at a lower cost than physicians who do not participate.

P4P and PR can reward health plans, provider groups, and individual clinicians in varying ways. Payment is made based on meeting prespecified goals. P4P has had a limited presence in behavioral health, but efforts to expand it into common disorders such as depression and substance use disorders have received considerable attention. For example, a P4P paradigm in behavioral health might involve the percentage of patients diagnosed with a new episode of depression who were treated with an antidepressant at an adequate dosage, who were given the medication for a full 12 weeks; or patients who are given at baseline and subsequently, the Patient Health Questionnaire (PHQ), where the scores are used, along with routine clinical assessment, to track response to treatment over time. If P4P and PR achieve similar penetration in behavioral health care as in physical health care, then the modeling for this approach ideally would begin in training programs.

In mental health, P4P is risky as it may enhance specific outcomes but undermine the essence of psychiatric treatment and the comprehensive totalistic humane view of the patient. Additionally, it has been linked in nonmedical organizations to increasing employee stress and anxiety, so the effect on physician burnout remains unknown but could be harmful if P4P becomes widely implemented or the de facto practice model. A large meta-analysis of P4P in 12 countries between 2010 and 2018 shows that the effect of P4P is significant if both cost and quality are included but often has unintended consequences and smaller effect sizes than intended.

Behavioral Health Care Services for Children With Serious Emotional Disorders

Children and adolescents with serious emotional disorders (SEDs), such as those with suicidal depression or early-onset psychotic disorders, represent another specialized population. On an annual basis, approximately 2% to 3% of children 3 to 5 years of age and 6% to 9% of children 6 to 17 years of age use mental health services. Of these, only a small percentage is defined as having SED, approximately 0.5%. These children, particularly adolescents, often require complex case management and intensive (e.g., residential or inpatient) services that are costly to provide and pay for.

The most problematic aspect of caring for children and adolescents with serious emotional disturbances is that five categorical agencies are responsible for providing it. Child Welfare has the largest group of these children and the largest budget. Special Education obligations in local school districts are second in expenditures. Mental Health agencies have a clear mandate but less funding despite usually caring for those in greatest need. Juvenile Justice programs serving youth who come before the courts and General Health and Primary Care systems often are the first to identify children with SEDs. However, these tend to operate separately from mental health specialty services and rely on psychotropic medications even though psychosocial treatments are considered the first-line intervention for most pediatric and adolescent psychiatric disorders.

This division of responsibility for child and adolescent services makes it very difficult to organize a coherent delivery system, and in most jurisdictions, confusion and fragmentation are the rule. In the 1990s, following a decade of National Institute of Mental Health (NIMH) planning grants, the Robert Wood Johnson Foundation developed the Mental Health Services Program for Youth (MHSPY) to come up with coherent systems of care for children with SEDs and their families, integrating the efforts of the responsible agencies at the community level. This was followed by a large grant program by the Substance Abuse and Mental Health Services Administration (SAMHSA). A number of the more than 80 communities that received these grants have developed advanced care management program models providing individualized, intensive care services to children (and their families) who might otherwise have been confined in institutional settings. Services under these models are often funded by risk-adjusted case rates derived from the blending of funds from different agencies' budgets and often administered by the Medicaid program. A compelling high-cost, high-need group that can potentially benefit from the integrated care provided under health homes is children with SEDs were identified in a 2014 policy brief by the Center for Health Care Strategies as:

> The Medicaid health home state plan option is designed to provide a comprehensive system of care coordination across medical, behavioral, and long term services and supports for Medicaid beneficiaries with chronic physical or serious mental health conditions . . . states can tailor behavioral health-focused health homes to meet the unique needs of both children and adults. (Center for Health Care Strategies 2014 policy brief. https://www.chcs.org/media/Developing_Health_Homes_for_SED_02_27_14.pdf.)

The health home is envisioned as a comprehensive set of services that wrap around the patient and address the full range of complex health problems. This could include pediatric, psychological, educational, behavior modification, case management, social work, family therapy, child welfare, legal/juvenile justice services, and psychiatry for children with SEDs. When working within the health home rubric with children, psychiatrists play a vital leadership role in collaboratively informing and coordinating with multiple behavioral health and social/educational services providers as well as with the pediatric primary care provider.

HEALTH INFORMATION TECHNOLOGY (HIT) AND ELECTRONIC HEALTH RECORDS (EHR)

> "By computerizing health records, we can avoid dangerous medical mistakes, reduce costs, and improve care." *President George W. Bush, State of the Union Address, January 20, 2004.*

Two successive administrations, Bush and Obama, have taken on the task of improving health records and communications. The 2009 American Recovery and Reinvestment Act (ARRA) and the related Health Information Technology for Economic and Clinical Health (HITECH) Act were designed to improve the widespread implementation of electronic health records. These acts represent wide-ranging public investments to help expand EHR utilization and expand HIT use to big data analysis. By modernizing the health care field, the hope is to improve population health, decrease medical errors, streamline financial reimbursement, and streamline billing activities. Another advantage of HIT adoption is the creation of more jobs in that health care technology sector, thereby simplifying and vitalizing a new economic sector that, until ~2008, was fairly limited. Expansion of that sector benefits HIT vendors, the data analytics sector, academic medical centers, professional bodies, and universities.

On the other hand, HIT could see major expansions with the advent of machine learning and the possibility of data mining through EHRs to find population health trends and improve them.

The most basic use of HIT is to transition from paper charting to electronic charting. The transition was spurred by the belief that EHRs would help decrease medical errors and enhance care collaborations. To achieve high use, governmental push included redefining “meaningful use,” then financial incentives and later penalties for nonuse.

The concept of meaningful use is embedded within the ACA. Meaningful Use entails financial incentives for providers (professionals and hospitals) in their “meaningful use” of certified *electronic health records* (EHR) to improve patient care. Examples of Eligible Professional Core Objectives are: use of computerized physician order entry (CPOE) for medication, laboratory, and radiology; generating and transmitting permissible prescriptions electronically (eRx); recording specific demographics (preferred language, sex, race, ethnicity, date of birth); recording smoking status for patients 13 years old and older; use of clinical decision support to improve performance on high-priority health conditions; a provider who receives a patient from another setting of care or provider of care should perform medication reconciliation; a provider who transitions their patient to another setting of care or provider of care or refers their patient to another provider of care provides a summary of care record for each transition of care or referral.

In addition, to receive an incentive payment, providers are required to submit Clinical Quality Measure (CQM) data through the EHR that measure and track the quality of health care services. These include health outcomes, clinical processes, patient safety, efficient use of health care resources, care coordination, patient engagements, population and public health, and adherence to clinical guidelines.

The effect of EHR implementation on medical practice has been profound. The first impact has been a change in how physician–patient encounters are conducted. Over the past two decades, the length of the average ambulatory experience has decreased to less than 13 minutes. In parallel, rates of burnout in physicians have grown as EHR use has become commonplace. An analysis of time spent with patients versus completing EHR/bureaucratic tasks indicates that up to 50% of encounter time is used to complete those tasks, and most physicians chart 1 to 2 hours after the workday is over. No EHR seems to have significantly better satisfaction or less burnout than others. The trend of EHR adoption also coincided with an increased emphasis on patient satisfaction. However, patient satisfaction depends mainly on a higher percentage of gaze level from the physician. Some literature from 2000 to 2015 encourages physicians to multitask better to meet the need for patient eye contact and keep the human touch in the clinical encounter. However, there is a growing awareness of the transformation of the clinical encounter in the past few years and that burnout was reaching nationwide severe proportions. Physicians who have to continue/finish their charting at home have increased odds of burnout and HIT-related stress. Patient satisfaction tends to drop when there is a switch from paper charting to HER, potentially reflective of the change in human interaction level. Therefore, as a significant reform move, HIT has transformed medicine as we know it, with some positive and some very significant negative, unintended consequences. Some of these consequences can be mitigated with training to enhance documentation quality, readability, and clinical accuracy. Training can lead to fewer medical errors and increased efficiency in chart review and data retrieval; however effective trainings are few and far between, and on a national level, physicians still rate EHR as a major reason for burnout. This is especially true in primary care specialties, and job satisfaction is related to HER satisfaction in some cases. The impact of trainings and other HER improvement strategies on burnout range from about 5%. In psychiatry, specifically, HER has transformed and quasi-annihilated the therapy experience, though, for fairness, it is necessary to point out that the demise of psychotherapy administered by psychiatrists started with managed care, another wave of reform that preceded the HIT/HER implementation. Most depression treatment has turned to psychotropic prescribing in the past two decades with a concomitant decrease in psychotherapy. While one cannot pinpoint a single factor as responsible for decreasing psychotherapy rates, a plethora of factors, including the change in physical settings needed to implement HER, may have contributed, as well as workflow disruptions. Interestingly, a measure of dissatisfaction with HER is that patients consistently overestimate the amount spent by physicians on HER.

The experiment with HER is significant as HIT can further modify the medical experience, HER modifications focusing on higher-level cognitive tasks and improved team communication may help, and improved patient portal usage to help with health maintenance.

The next stages for HIT expansion are big data analysis and health outcomes optimization via data mining and patient-reported outcomes (PROs). From a systems perspective, the quasi-universal use of HER and the level of acceptability in patients makes it possible to use it for (1) PROs so that clinicians do not have to ask questions in session as much, (2) Diagnostic validation, and (3) Interface with applications and wearables outside of session. All these goals are still years away but are abundantly discussed as the next frontiers in HIT and the improvement of health care. The PCORI is an excellent resource to guide this field development effort. The Users’ Guide to Integrating Patient-Reported Outcomes in Electronic Health Records aims to outline a map of PRO-EHR integration across the planning, selection, and engagement areas.

In order to properly use PROs, careful selection of what outcomes to use has to occur first. The outcomes must be clinically meaningful, have some expected impact over the disease course or illness treatment decisions, or patient self-efficacy, and present a reasonable burden for patients regarding time and accessibility. For psychiatric illnesses, substance use disorders, mood symptoms, and medication adherence are obvious choices. Regardless of challenges, the potential for large-scale data collection and standardization is obvious with HER integration. This innovation will pave the way for effective population surveillance.

Second, patient engagement is crucial and may highlight socioeconomic and educational status disparities. While smartphone ownership is quasi-ubiquitous, technical literacy is still subject to socioeconomic and age disparities, with younger, wealthier individuals being more technology-savvy. Individuals with psychotic disorders and severe depression, catatonia, or mania may not fully engage in PRO collection or use and may need separate workflows; these challenges are specific to psychiatry and specifically to those working with populations with severe mental illness. Specifically, smartphone ownership and app usage are shown to be much less in individuals with psychotic disorders than those without, and this could be related to the nature of their illness, with paranoia and delusions centered around the information being collected, the effect of magnetic waves from the phone or wearable, and the way information will be used.

Third, the field currently lacks guidelines to support PRO meaningful interpretation. For example, patients logging their mood or filling out a mood questionnaire cannot get a personalized, treatment-driven interpretation of their symptoms. The information must be available to their clinician to be usable, which requires interoperability with EHR. Last, the liability of clinicians and health care systems of PRO-generated info is not tested yet. One can think of scenarios when PRO include mood deterioration or suicidal ideations preceding an adverse outcome such as a completed suicide. This area is unexplored yet, and there are no guidance measures to lead psychiatrists to use apps, wearables, or PRO portals from their institutions for patient care.

Accessing patients to their PRO and interpretation of information provided is a field of its own, currently undergoing extensive growth. The federal laws protecting patient access to their medical record should cover that, but again, this is an expanding area with no formal guidance. In order for the next HIT revolution and reform to occur, specifically in the context of having patients fill in PRO, several steps have to be formalized:

- The health system culture has to acknowledge the importance and usefulness of PRO and support the use of PROs.
- Health systems have to invest in PRO collection as a necessary means to help patients and clinicians improve population health one individual at a time.
- Training and efforts to promote this culture change, including cost and time requirements, must be provided. On a more psychiatry-specific and disease-specific level, a framework to build notification procedures and modalities and identify the appropriate recipients to receive notifications for PRO data will be needed (who will be on call for late evening patient digital alerts?). In certain situations, alert fatigue may become a theoretical possibility, as with personality disorders patients.

A Further EHR-HIT reform element is the allocation of wearables to monitor health outcomes when patients are in their natural environment. This is far from reality in psychiatry but is, like PROs above, being discussed and planned as the next step in patient-centered population improvement relying on real-time data. Areas of interest identified in order for this transition to be successful include “device type, communication interface, data transfer protocol, smartphone and/or PC integration, direct integration to open health platform, 3rd platform integration with open health platform, support for health care system/middleware connection, recorded health data types, integrated sensors, medical device certification, whether or not the user can access collected data, device developer access, and device availability on the market.” At this point, those measures are best suited for physical illnesses with well-defined physiologic markers that can be used to build machine learning algorithms, for example, heart rate, blood pressure, or blood sugar.

In conclusion, the issues involved in reforming health care are complex and transformational. They have the potential to change the way medical encounters are conducted.

Effective reform requires the expertise of a broad range of disciplines, from economics to ethics to basic public philosophy, and effects take decades to show. Therefore, psychiatrists must remain an integral part of the conversation as users of the system and as physician leaders and be ready with a clear vision and meaningful participation in the national health care dialogue.

FURTHER READINGS

Ahlgrén-Rimpiläinen AJ, Arffman M, Suvisaari J, et al. Excess mortality from breast cancer in female breast cancer patients with severe mental illness. *Psychiatry Res*. 2020;286:112801.

Alejandria M. *The Impact of the Mental Health Parity and Addiction Equity Act (MHPAEA) on Behavioral Health Care Utilization and Spending*. California State University; 2021.

Bauer M, Glenn T, Geddes J, et al. Smartphones in mental health: a critical review of background issues, current status and future concerns. *Int J Bipolar Disord*. 2020;8(1):1–19.

Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Aff*. 2008;27(3):759–769.

Correll CU, Solmi M, Veronese N, et al. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. *World Psychiatry*. 2017;16(2):163–180.

Couper MP, Gremel G, Axinn W, Guyer H, Wagner J, West BT. New options for national population surveys: The implications of internet and smartphone coverage. *Soc Sci Res*. 2018;73:221–235.

Dahl MS, Pierce L. Pay-for-performance and employee mental health: Large sample evidence using employee prescription drug usage. *Acad Manage Discov*. 2020;6(1):12–38.

Das-Munshi J, Ashworth M, Gaughran F, et al. Ethnicity and cardiovascular health inequalities in people with severe mental illnesses: protocol for the E-CHASM study. *Soc Psychiatry Psychiatr Epidemiol*. 2016;51:627–638.

Domaney NM, Torous J, Greenberg WE. Exploring the association between electronic health record use and burnout among psychiatry residents and faculty: a pilot survey study. *Acad Psychiatry*. 2018;42(5):648–652.

Gardner RL, Cooper E, Haskell J, et al. Physician stress and burnout: the impact of health information technology. *J Am Med Inform Assoc*. 2019;26(2):106–114.

Gensheimer SG, Wu AW, Snyder CF. Oh, the places we’ll go: patient-reported outcomes and electronic health records. *Patient*. 2018;11(6):591–598.

Gottschalk A, Flocke SA. Time spent in face-to-face patient care and work outside the examination room. *Ann Fam Med*. 2005;3(6):488–493.

Gribben V, Bogetz A, Bachrach L, Blankenburg R. The golden ticket Project for peer recognition. *Clin Teach*. 2019;16:203–208.

Hodgkin D, Horgan CM, Stewart MT, et al. Federal parity and access to behavioral health care in private health plans. *Psychiatr Serv*. 2018;69(4):396–402.

Horn ME, Reinke EK, Mather RC, O'Donnell JD, George SZ. Electronic health record-integrated approach for collection of patient-reported outcome measures: a retrospective evaluation. *BMC Health Serv Res*. 2021;21(1):626.

Kennedy-Hendricks A, Epstein AJ, Stuart EA, et al. Federal parity and spending for mental illness. *Pediatrics*. 2018;142(2):e20172618.

Kurz T, Liaw W, Wingrove P, Petterson S, Bazemore A. Funding instability reduces the impact of the federal teaching health center graduate medical education program. *J Am Board Fam Med*. 2017;30(3):279–280.

Luther L, Buck BE, Fischer MA, Johnson-Kwochka AV, Coffin G, Salyers MP. Examining potential barriers to mHealth implementation and engagement in schizophrenia: Phone ownership and symptom severity. *J Technol Behav Sci*. 2020;7:13–22.

Mojtabai R, Olfson M. National trends in psychotherapy by office-based psychiatrists. *Arch Gen Psychiatry*. 2008;65(8):962–970.

Muzny M, Henriksen A, Giordanengo A, et al. Dataset of wearable sensors with possibilities for data exchange. *Data Brief*. 2020;28:104978.

North F, Pecina JL, Tulledge-Scheitel SM, Chaudhry R, Matulis JC, Ebbert JO. Is a switch to a different electronic health record associated with a change in patient satisfaction? *J Am Med Inform Assoc*. 2020;27(6):867–876.

Papakostas GI, Ionescu DF. Updates and trends in the treatment of major depressive disorder. *J Clin Psychiatry*. 2014;75(12):1419–1421.

Rieselbach RE, Epperly T, Nycz G, Rockey P. Teaching health centers can meet objectives for state Medicaid innovation. *J Grad Med Educ*. 2018;10(3):362–366.

Robinson KE, Kersey JA. Novel electronic health record (EHR) education intervention in large healthcare organization improves quality, efficiency, time, and impact on burnout. *Medicine*. 2018;97(38):e12319.

Rotenstein LS, Torre M, Ramos MA, et al. Prevalence of burnout among physicians: A systematic review. *JAMA*. 2018;320(11):1131–1150.

Ryu B, Kim N, Heo E, et al. Impact of an electronic health record-integrated personal health record on patient participation in health care: development and randomized controlled trial of MyHealthKeeper. *J Med Internet Res*. 2017;19(12):e401.

Sieja A, Markley K, Pell J, et al. Optimization sprints: improving clinician satisfaction and teamwork by rapidly reducing electronic health record burden. *Mayo Clin Proc*. 2019;94(5):793–802.

Stewart RE, Lareef I, Hadley TR, Mandell DS. Can we pay for performance in behavioral health care? *Psychiatr Serv*. 2017;68(2):109–111.

Thalmayer AG, Friedman SA, Azocar F, Harwood JM, Ettner SL. The mental health parity and addiction equity act (MHPAEA) evaluation study: Impact on quantitative treatment limits. *Psychiatr Serv*. 2017;68(5):435–442.

Zaresani A, Scott A. Is the evidence on the effectiveness of pay for performance schemes in healthcare changing? Evidence from a meta-regression analysis. *BMC Health Serv Res*. 2021;21(1):1–10.

56.3 The Role of the Hospital in the Care of Persons with Mental Illness

ERICA TAYLOR, M.D., M.P.H.

Throughout the 19th century and for the first half of the 20th century, public psychiatric hospitals (and some private ones as well) provided all the roles necessary to take care of psychiatric patients. The state hospital was the source for inpatient, outpatient, case management and outreach, and emergency services. If someone needed admission, a family member could simply transport him to the front door. With the shift from inpatient to community-based services that began in earnest in the mid-20th century, admission to, discharge from, and services within the psychiatric hospital became both more complex and more convoluted. Layer upon layer of agencies, services, and regulations facilitated and impeded the psychiatric hospital in its mission (Fig. 56.3–1).

The reasons for the evolution of this relationship between the psychiatric hospital and its prospective and actual patients are multifactorial. As Figure 56.3–1 illustrates, there are entities with overlapping but competing agendas at every turn. Fundamentally, there are two predominant factors that account for the contemporary, tortuous, labyrinthine interface between the psychiatric hospital and the community. First, the psychiatric hospital, in a manner quite distinct from the general hospital, plays two roles: clinical services and public safety, thus, the extensive interface between the psychiatric hospital and the criminal justice system (see Fig. 56.3–1). And second, psychiatry, more than any other field of medicine, has been inundated by gyrations of conflicting actions to decrease the cost of care and treatment, thus the barriers and alternatives to psychiatric hospital admission (see Fig. 56.3–1).

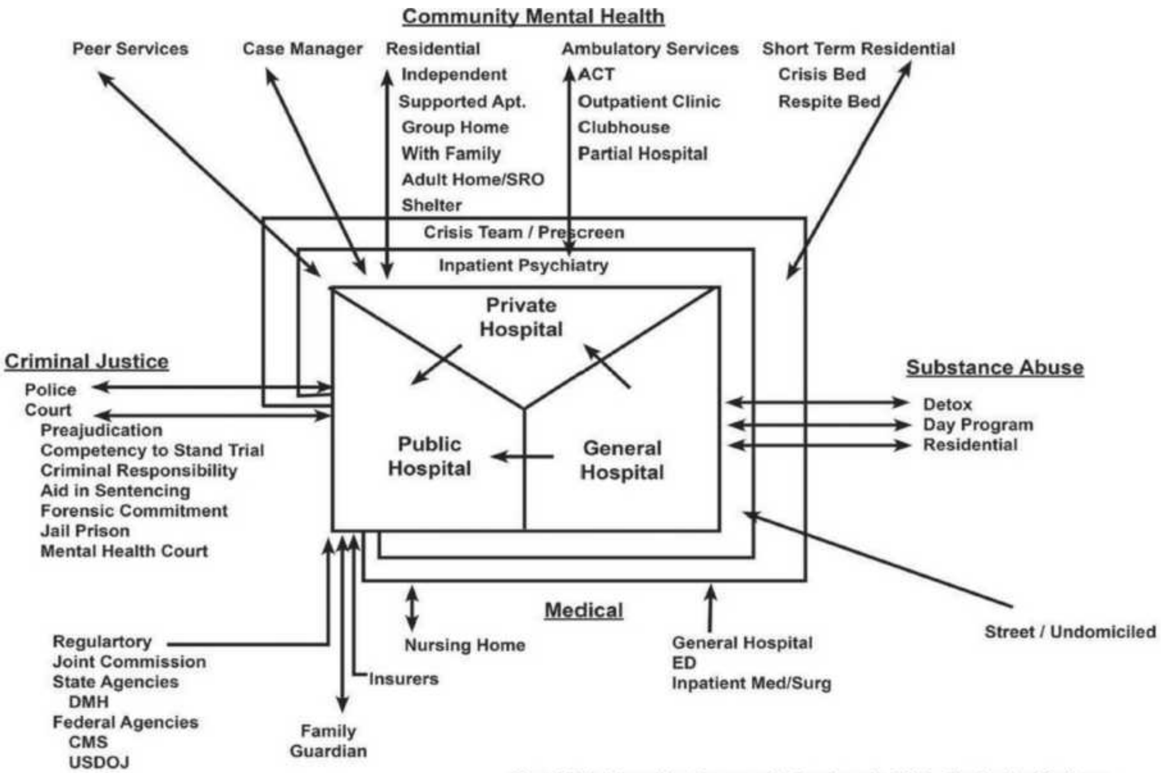
Community Mental Health
Peer Services
Case Manager
Residential
Independent
Supported Apt.
Group Home
With Family
Adult Home/SRO
Shelter
Ambulatory Services
ACT
Outpatient Clinic
Clubhouse
Partial Hospital
Short Term Residential
Crisis Bed
Respite Bed
Crisis Team / Prescreen
Inpatient Psychiatry
Private Hospital
Public Hospital
General Hospital
Criminal Justice
Police
Court
Preajudication
Competency to Stand Trial
Criminal Responsibility
Aid in Sentencing
Forensic Commitment
Jail Prison
Mental Health Court
Substance Abuse
Detox
Day Program
Residential
Medical
Nursing Home
General Hospital
ED
Inpatient Med/Surg
Street / Undomiciled
Regulartory
Joint Commission
State Agencies
DMH
Federal Agencies
CMS
USDOJ
Insurers
Family
Guardian

Key: ACT - Assertive Community Treatment; CMS - Center for Medicare and Medicaid Services; DMH - Department of Mental Health; USDOJ - United States Department of Justice

FIGURE 56.3–1. The psychiatric hospital in the context of the community.

The first one-and-one-half decades of the 21st century have not been kind to psychiatric hospitals. Virtually every state in the United States is experiencing a shortage of psychiatric beds. These shortages include acute care beds, intermediate and long-term care beds, and forensic beds. The effects of these bed shortages are increased waiting lists for state hospital beds, increased waiting lists for other psychiatric beds like general hospitals and Veterans Administration (VA) psychiatric beds, overcrowding in state hospitals, backups, and boarding in general hospital emergency departments (EDs), and premature discharges from acute care settings. These shortages have also meant a shift of persons with serious mental illness (SMI) behind locked doors from public hospitals to jails and prisons. These psychiatric bed closures come from all types of hospitals. Sixty-four percent of states have experienced declines in the number of general hospital specialty unit psychiatric beds, and 19 states have experienced a decline in the number of private psychiatric hospital beds. Nationwide, the number of state and county hospital beds declined by 86% (369,999 to 52,612) between 1969 and 2002; the number of VA medical center (VAMC) psychiatric beds declined by 84% (51,696 to

8,386) within the same time period. The inaccessibility of psychiatric beds is a regular feature in news stories and editorials, often with a link between that and a local or national tragedy.

Lengths of stay have plummeted. Between 1990 and 2010, the length of stay of hospitalizations for SMI in acute care settings decreased by 50% (from 14 days to 7 days). Private insurers represent a shrinkage percentage of payments for hospital services, especially for substance abuse services. About 50% of all payments for psychiatric inpatient services were from government programs; state governments' funding increased by 65% just between 2000 and 2001.

Hospital-based outpatient services grew by 78% (242,301 episodes to 1,104,300) between 1969 and 2002. In 2002, however, hospital-based outpatient services represented only 31% of all outpatient episodes of care for psychiatric disorders.

The shift of resources from hospital to nonhospital services is reflected by the changing distribution of expenditures by state mental health authorities (SMHAs). Inpatient services received 63% of the SMHA budget in 1981, 50% in 1993, 28% in 2004, and 23% in 2012. Although the overall inpatient percentage decreased, the proportion going to forensic patients increased to 33% of all state hospital expenditures.

While there has been a shift in public expenditures from inpatient to outpatient services, there has been an overall decrease in public expenditures for psychiatric treatment. In fiscal years 2010 and 2011, the biggest cuts were directed at state hospitals with 20 and 15 states, respectively, cutting state hospital funding. Less funding meant fewer individuals could be serviced in state hospitals, with 31 states decreasing the persons they admitted comparing 2007 to 2009.

At the same time, the demand for state hospital utilization increased in 18% of the states, in part because states were also making cuts in all manner of community-based services. Between fiscal years 2009 and 2011, 17 states had cuts in their overall state mental health expenditures by greater than 1%, with 5 of those states exceeding 20% (1 state approached 50%).

Marshaling scarce resources has led to synergies among various sectors within a state. In just more than 75% of the states, community-based programs (often at general hospitals) perform a gate-keeping function to state hospital entry. Some states (38%) require general hospital admission prior to a transfer to a state hospital. And in all but five states, there are assisted outpatient commitment (AOT) (involuntary outpatient commitment) laws on the books with an expectation these can be used to decrease both inpatient and overall mental health services costs.

The picture does not look optimistic, at least for the immediate future. About half of the states are reorganizing their state hospitals, including downsizing, reconfiguring, closing, and/or consolidating. Despite this, 11 states still have 6 or more state hospitals each. And a surprising number of states are building new state hospital buildings, but these generally replace one or more existing state hospitals and usually do not reflect an increase in beds (see Table 56.3–1). As more patients are admitted under forensic commitments, lengths of stay increase, and so does the unavailability of beds to others. In 2006, about 40% of patients found not guilty by reason of insanity had lengths of stay of 1 to 5 years, and about 38% had lengths of stay of greater than 5 years.

While the place of psychiatric hospitals as a tool in the resources box of treatment for mental illness is shrinking, the need is growing. On any given day in the fiscal year 2010, the number of noninstitutionalized individuals with an episode of serious psychological distress in the past was 3% of the population. In 2012, 4.1% of the adult U.S. population had an episode of SMI in the past year. The 12-month prevalence for schizophrenia in adults was 1.15%, while it was 3.5% for post-traumatic stress disorder (PTSD) (one-third of these classified as severe), 9.1% for any personality disorder, and 18.1% for anxiety disorder (23% of these classified as severe). The mismatch between need and services is no better represented than by the fact that in 2008 over 40% of adults in the United States with an SMI did not receive any treatment for a mental health problem. Of those that did, 7.5% had an inpatient, and 40.5% had an outpatient episode of care.

The state of hospital psychiatry raises questions about how it came to play the roles it does. What were the historical influences that brought the psychiatric hospital to where it is, and what keeps it there? How does the psychiatric hospital (this term will include general hospital psychiatry units) fulfill its current role? And what should be the roles of psychiatric hospitals in the future?

THE HISTORICAL EVOLUTION OF THE ROLE OF THE HOSPITAL

Prior to the 19th century, the care of insane persons was custodial in nature, took place largely in almshouses, and was predominantly a local endeavor. State care, in the beginning, meant an institution established, built, and managed by the state but whose operation was not funded by the state. Payers were, in descending order of priority, the patient, the family, and the county. The earliest public facilities—in Williamsburg, Virginia, and Columbia, South Carolina—functioned in this manner. The state taking responsibility for all insane persons and performing its responsibility through state-managed and funded hospitals began with Worcester State Hospital in 1833.

Before there were any more than a few public hospitals, there was already a well-established group of private psychiatric hospitals. Three of these began as parts of general medical hospitals: The Institute of the Pennsylvania Hospital, Bloomingdale Hospital (New York), and McLean Hospital (Massachusetts). Three more were freestanding institutions: Butler Hospital (Rhode Island); Brattleboro Retreat, which began as the Vermont Asylum for the Insane; and the Hartford Retreat, which subsequently was renamed the Institute of Living (Connecticut). Two of the hospitals began under the aegis of religious organizations: The Friends Asylum (Pennsylvania) was initiated by a Quarterly Meeting of the Philadelphia Society of Friends, and the Mount Hope Institution (Maryland) was a Catholic hospital that began as the Mount Saint Vincent Hospital.

Of these larger private facilities, all but one preceded the state's first public hospital, sometimes by more than 50 years. The private hospitals helped the state to meet its obligations to the insane poor by admitting "public" patients. Payments were made by the state to the private facility, and some treatment was given without charge. Thus, the states' purchase of beds from the private sector was not a novel concept of the late 20th century but is actually how U.S. hospital psychiatry began.

Many public and some private hospitals throughout the country were built under a design put forth by Thomas Kirkbride, a founding member of the American Psychiatric Association (APA). In his book, *On the Construction, Organization, and General Arrangements of Hospitals for the Insane*, Kirkbride prescribed every aspect of hospital construction and operation. His prototype hospital, variations of which can still be seen across the United States (but are fast being torn down), is shown in Figures 56.3–2 and 56.3–3. Noteworthy among his recommendations is that a 250-bed hospital would have 16 wards, with about 15 patients per ward.

Table 56.3–1.

A Sample of Public Psychiatric Hospitals Built Since January 2000

State	Hospital	Opening Date for Patients	No. of Hospital Beds
AK	Alaska Psychiatric Institute	July 2005	80
AR	Arkansas State Hospital	July 2008	210
AZ	Arizona State Hospital–Adolescents	July 2002	16
AZ	Arizona State Hospital–Adults	February 2003	200
CA	California Health Care Facility–Stockton	July 2013	1,772
CA	Coalinga State Hospital (Forensic)	September 2005	1,500
CO	Robert L. Hawkins High Security Forensic	June 2009	200
DC	St. Elizabeth’s Hospital	April 2010	292
GA	Central State Hospital—Payton B. Cook Forensic	January 2009	184
IN	Evansville State Hospital	September 2003	168
IN	Logansport—Isaac Ray Treatment Center	August 2005	88
KY	Eastern State Hospital	September 2013	239
MA	Worcester Recovery Center and Hospital	October 2012	290
ME	Riverview Psychiatric Center	2004	92
MI	Center for Forensic Services	2005	210
MO	Center for Behavioral Medicine	2004	50
MO	Fulton State Mental Hospital	2018	300
MS	South Mississippi State	2000	50

	Hospital		
MT	Montana State Hospital	2000	174
NC	Broughton	2016	382
NC	Central Regional Hospital	July 2008	398
NC	Cherry Hospital	2015	314
NJ	Greystone Park Psychiatric Hospital	July 2008	450
NV	Rawson–Neal Psychiatric Hospital	August 2006	190
OH	Northcoast Behavioral Healthcare Hospital	2016	258
OH	Summit Behavioral Healthcare Hospital	2006	291
OR	Oregon State Hospital—Salem	January 2011	620
OR	Oregon State Hospital—Junction City	March 2015	174
SC	William S. Hall Psychiatric Institute	2015	515
TN	Memphis Mental Health Institute	October 2007	75
TN	Western Mental Health Institute	March 2010	162
TX	Rusk State Hospital: Maximum Security Unit	June 2012	60
VA	Eastern State Hospital Geriatric	April 2008	150
VA	Eastern State Hospital Adult	September 2008	150
VA	Western State Hospital	2013	250

Additional Facilities Opened Since 2005.

California: Coalinga State Hospital—opened Sep 2005, beds 1,260.

Connecticut: Silver Hill Hospital, Chronic Pain and Recovery Center Program—opened 2012 (no bed count).

Silver Hill Hospital, Eating Disorder Program for Adults—opened 2015 (no bed count).

Indiana: NeuroDiagnostic Institute—opened March 2019, beds 159.

Texas: University Behavior Health Denton—opened 2005, bed 104.

By the 1880s, most psychiatrists agreed that “insane patients can best be managed away from their own homes.” Parsons, an alienist of the era, opined, “well-equipped and well-managed asylums for the insane are among the most important of the outgrowths of modern civilization and benevolence.” By the end of the 19th century, it was clear that private facilities were for the wealthier, more genteel class of citizens, who were able to pay fees and were provided remarkably comfortable accommodations. Public hospitals were entities of the state, county, or municipal governments that served not only the poor but also the middle class, given that the latter could not afford private hospitals.

A major difference between the public and private facilities is that public institutions sought to lower their censuses as state hospital populations grew and patients had progressively longer lengths of stay, whereas private facilities sought patients to maximize their occupancy rates. A fundamental criticism of private facilities was pointed out by Parsons: “It is not for the interest of the proprietor of a private asylum to cure his patients, and hence they may not make an honest effort to promote their cure.” Much of what many consider to be modern developments in hospital psychiatry were well under way before the end of the 19th century: Oversight of private hospitals by state boards, advertising by private hospitals, involuntary admissions to private hospitals, more perks for higher fees, and the “dumping” of patients from the private to the public sector.

Work and recreation were both seen as important to the mental health of patients in the 19th century. Generally, the public hospitals did better with the former, and the private ones with the latter. Recovery, employment, and wellness are not novel ideas of contemporary hospital psychiatry. Each was endorsed more than 150 years ago in both public and private hospitals.

Nor is evidence-based psychiatry a new thought. Andrews articulated in 1887 that “the unparalleled progress in neurology, cerebral anatomy, physiology, pathology, and localization of function has enlarged the horizon of our knowledge of disease and of the action of causes.” This progress, he said, “furnished a scientific and positive basis for treatment in many cases of insanity which before was unattainable.” Inpatient treatments near the end of the 19th century included such modalities as electricity, massage, oophorectomy, enforced rest, hydrotherapy, and physical training.

Changes in hospital psychiatry over the first half of the 20th century were dramatic. Psychoanalysis was introduced to the inpatient setting in the 1920s and somatic treatments such as insulin coma, Metrazol shock, and electroconvulsive therapy in the 1930s. In the 1940s, attention was focused on developing milieu therapy, whereas in the 1950s, the introduction of the first antipsychotic medication, chlorpromazine, fundamentally altered inpatient psychiatry.

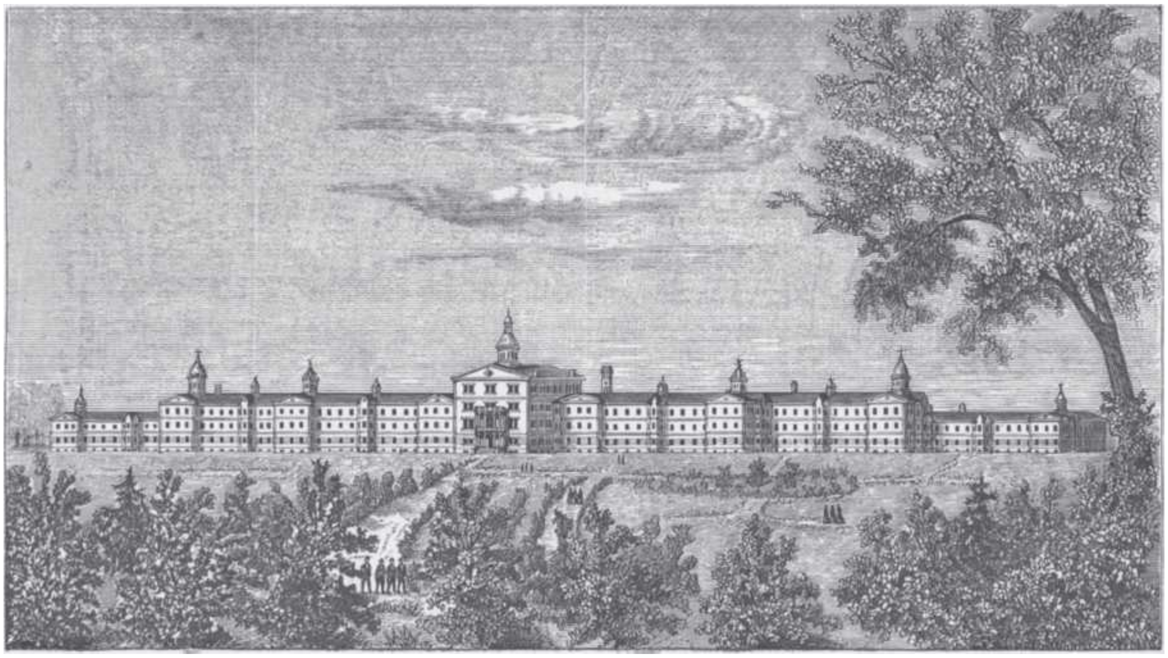

FIGURE 56.3–2. Exterior of hospital prototype designed by Thomas Kirkbride. (From Kirkbride TS. *On the Construction, Organization, and General Arrangements of Hospital for the Insane*. J.B. Lippincott; 1880.)

The mid-1950s began the shift from hospital-based to nonhospital-based care and treatment of persons with mental illness, a movement retrospectively labeled “deinstitutionalization” but perhaps better termed “dehospitalization.” General hospital psychiatry and day hospitals became major loci of care and treatment, and the role of the psychopathic hospital expanded. Units in general hospitals were not always welcomed by hospital medical staff, who were concerned that psychiatric patients, whom the physician thought would overwhelmingly be noisy, destructive, violent, or suicidal, would disrupt the general hospital.

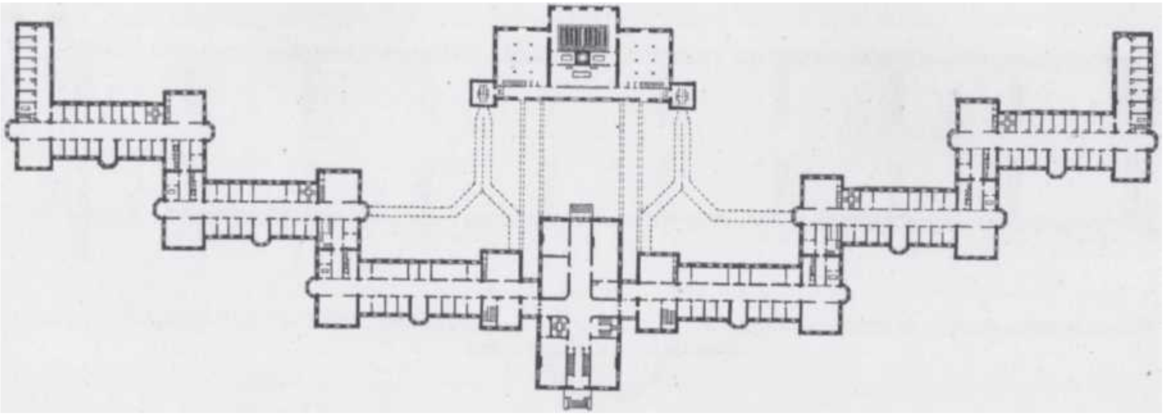

FIGURE 56.3–3. Interior of hospital prototype designed by Thomas Kirkbride. (Kirkbride TS. *On the Construction, Organization, and General Arrangements of Hospital for the Insane*. J.B. Lippincott; 1880.)

The darkest era in U.S. hospital psychiatry was in the 1940s, a time when the care of persons with mental illness was described as a national disgrace. Attempts were made to improve inpatient treatment through the promulgation of principles and standards, such as those put forward by the Division of Mental Hygiene of the U.S. Public Health Service, the National Committee for Mental Hygiene, and the APA's Committee on Psychiatric Standards and Policies. These probably had less effect than exposés such as Mary Jane Ward's *The Snake Pit*, a fictionalized first-person account of life in a state hospital.

Whereas the public sector was overburdened and overrun, by 1950, the private sector was also facing financial difficulties. The Medical Director of McLean Hospital stated at the 1950 APA Institute, "The time when you could run a private hospital as a hospital and make money has gone. Operating expenses have so increased that it is almost impossible to collect the cost from our patients."

In the 1960s, as health insurance for hospitalization expanded to include psychiatric treatment, psychiatric hospital treatment outside of state hospitals became more available. Misuse of insurance for hospitalization for mental illness was quickly recognized. To remedy the manipulation of hospital benefits, interventions were introduced: Peer review, reimbursement for partial hospitalization, and reimbursement for other alternatives to inpatient treatment. “Dumping” and “cost-shifting” emerged as part of the everyday life of inpatient psychiatry.

In the 1970s, states were contemplating phasing out state hospitals. General hospital psychiatric units were becoming more of a presence. The expanding role of general hospital psychiatry drew caution from some who warned workers in general hospitals to be watchful of efforts to convert general hospital psychiatric units into mini-state hospitals. Declining numbers of individuals willing or able to pay for private hospital treatment and more available third-party coverage led private hospitals to admit patients of different socioeconomic classes and ethnic backgrounds and to keep all patients for shorter lengths of stay.

In the 1980s, there was a remarkable growth in private sector hospital beds. Private corporations were rushing to build psychiatric hospitals as less regulation, lower overhead, and greater latitude in insurance reimbursement meant there were huge profits to be made in the “psychiatric hospital industry.” For-profit psychiatry grew, but with it came a new set of problems. For-profit hospitals were the least likely to treat the most functionally disabled; would treat few if any patients who could not pay for treatment; usually had poorer staff-to-patient ratios than nonprofit hospitals; failed to provide education and training; failed to contribute adequately to research, and overall provided a lower quality of care. These hospitals tended to be totally disengaged from the communities in which they were located; they paid higher salaries and thus hired away graduating residents from academic and public sector positions.

The heyday of for-profit hospitals (most private psychiatric hospitals) did not last long. Throughout the 1990s, the total and the average number of beds per private hospital decreased, as did the average length of stay. Between 1992 and 1998, private psychiatric hospital beds declined by 23%. For-profit companies had built “chains” of hospitals so rapidly that there quickly came to be too many private psychiatric beds, and hospital chains began experiencing major financial losses. Lawsuits against the chains for unscrupulous practices such as recruitment of patients and manipulative billing practices were settled in the hundreds of millions of dollars.

The 1990s also witnessed the privatization of public mental health services. Medicaid, Medicare, and state funds were directed to the private sector to greater degrees during the 1980s and 1990s. States assisted freestanding private hospitals in obtaining waivers of the Institution for

Mental Disorders (IMD) rule (see later discussion). A greater reliance on the private system for those formerly treated in the public system created many unintended consequences that varied state by state, including inpatient treatment further from one's community, less continuity of inpatient treatment, new patterns of recidivism, and certain patients being blacklisted at some hospitals.

In 2001, 37% of general hospitals had psychiatric inpatient care. This percentage is higher than those with a psychiatric emergency, consultation-liaison (C-L), or even outpatient services. Hospital psychiatry in the United States continued to be a *business*. However, in the business of psychiatry, nonprofit psychiatric care providers compared favorably with for-profit providers.

It is interesting to note how the practice of hospital psychiatry appears to have shifted in response to shorter lengths of stay. A study at McLean Hospital in 2002 that compared inpatient antipsychotic medication use in 1989, 1993, and 1998 showed the chlorpromazine-equivalent final total dose of all antipsychotic medications used at discharge in 1998 was 29.3% greater than in 1993 and 46.1% greater than in 1989. The average time from admission to first antipsychotic dose decreased dramatically from 6.8 days in 1989 to 1.5 days in 1993 and .9 days in 1998. Two or more antipsychotic medications were used during 20% of inpatient days in 1998, compared to 5.7% days in 1993 and 1.7% in 1989. Pressures to discharge patients reduced not only the length of treatment but also the content of treatment.

In the first decade of the 2000s, the prospects for the private psychiatric hospital and for general hospital psychiatric units did not look good. Although overall health care spending increased by 15.7% between 1992 and 1999, mental health and substance abuse spending decreased by 17.4%. Behavioral health care spending went from 7.2% of total private health insurance spending to 5.1% over this same time period. Reimbursement rates fell such that they often did not even meet the total costs of care and treatment. While reimbursement rates fell, costs to deliver treatment increased.

Private hospitals were offering fewer services. Those hospitals offering partial programs decreased from 82.5% to 66.7%, and those offering outpatient services dropped from 64.2 to 49.2% between 2000 and 2001. At the same time, oversight increased. The Office of the Inspector General (OIG) of the Department of Health and Human Services, the Joint

Commission on Accreditation of Healthcare Organizations, and the Center for Medicare and Medicaid Services all stepped up investigations of hospitals. Medicaid's progressive decrease in disproportionate share payments (DSHs) to public mental hospitals is taxing some states' ability to operate their public system at its pre-2000 scale.

The landscape of U.S. inpatient psychiatry became so rough that well-established private hospitals closed or moved. Those in the former category include Taylor Manor and Chestnut Lodge in Maryland. In the latter category, the best known is the Menninger Clinic, formerly of Topeka, Kansas, and now in Houston, Texas. In the public sector, state hospitals, well established since the 19th century, began reinventing themselves, usually by the consolidation of multiple public hospitals and by new construction such as Worcester State Hospital (Massachusetts), Oregon State Hospital, and Western State Hospital (Virginia). Furthermore, the need to serve and contain persons with mental illness and histories of sexually problematic behaviors is diverting resources from the treatment of psychiatric disorders. California built a 1,500-bed facility for forensically committed individuals—mostly sexually violent predators—representing the first new California facility in more than 50 years. Even older adult individuals with such histories can often not be discharged from the public psychiatric hospital as federal law prohibits nursing homes from admitting persons who are classified as the highest level of sex offenders.

As components of psychiatric services in the second decade of the 21st century, psychiatry hospitals are shrinking, growing, being torn down, and being newly built all at the same time. As hospitals are unable to meet the needs of the population requiring inpatient level of care, local police have become the triage agents for persons with serious mental illness in the streets; jails and prisons are unofficial replacement facilities for state hospitals, and mental health courts provide more case management than anything else. One potential positive note is the growing number of new public psychiatric hospitals built with the hope that the building's design will facilitate patients' recovery (see Table 56.3–1). It is well worth remembering, however, the observation made in the wake of newly constructed community mental health centers (CMHCs) with inpatient units in the 1970s: the latest in "psychoarchitecture" and design elements do not result in behavioral remediation since clinical improvement does not result simply from changing bricks, mortar, and furniture. That being said, newly built hospitals, most of which have had significant problems with translating programmatic and life safety design concepts into actual hospital structures, have brought more pleasant surroundings, in an overall safer environment, with more available space for in- and outdoor programming (see Figs. 56.3–4 and 56.3–5). These figures are presented as an example and not as a model.

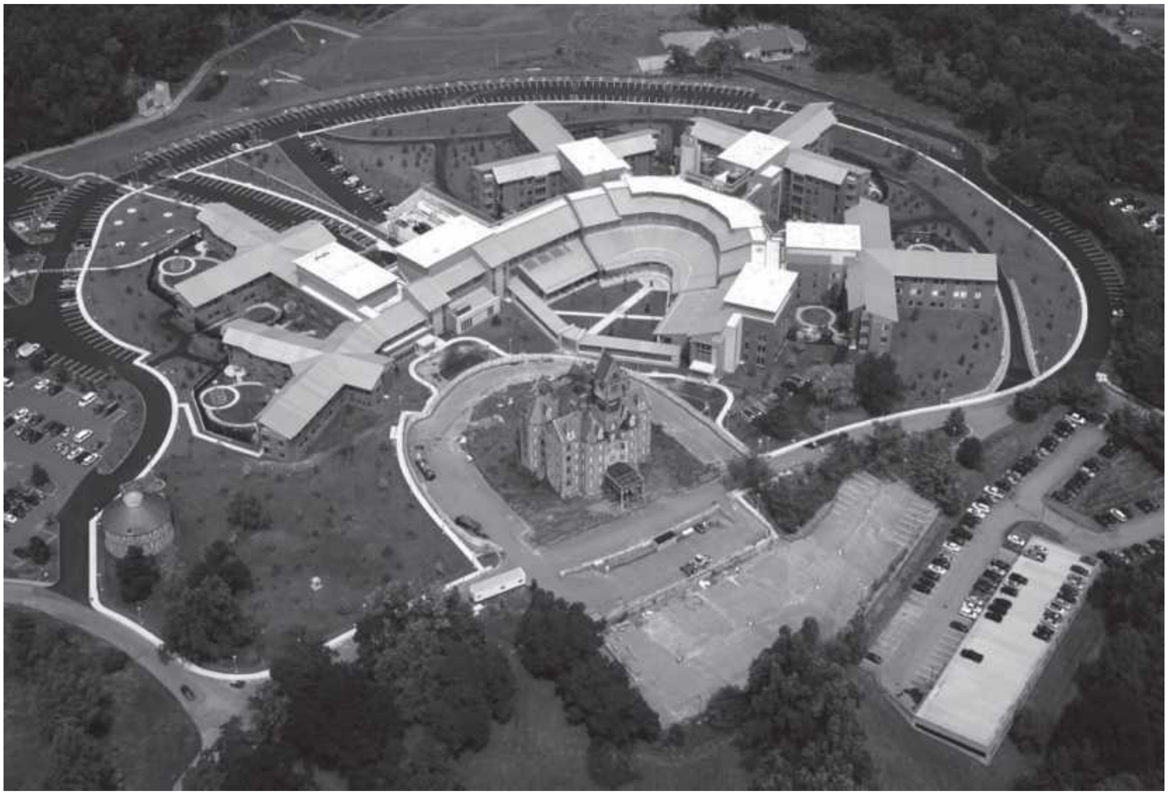

FIGURE 56.3–4. Worcester Recovery Center and Hospital (WRCH). Exterior view.

Available psychiatric beds in general and private hospitals appear to be made inaccessible to patients by insurance providers because limiting admission to hospitals with which they have negotiated rates is one means to control cost. As reimbursement rates drop below the actual expenditures of the general hospital, these hospitals limit the third-party payers with which they will contract or further reduce or eliminate psychiatric beds, exacerbating the bed shortage.

Public psychiatric beds are highly disproportionately occupied by forensically committed patients. This makes beds for civil admissions less available, and patients are stuck in general or private psychiatric hospital beds, which in turn delays the movement of psychiatric patients needing admission out of the ED. Waiting time in an ED can be up to 1 month for a psychiatric bed, or the patient is prematurely discharged from the ED, sometimes with lethal consequences.

Despite ascriptions of causality for the current state of hospital psychiatry to malevolence, malfeasance, and malversation, it would seem more accurate to see the role of the psychiatric hospital in the second decade of the 21st century as still the result of a basic tension between clinical benefit, social control functions, and parsing out of costs. Each governmental level and each government agency wants another one to fund the treatment of mental illnesses.

Despite all of these uncertainties, there is no indication that psychiatric hospitals will cease to play some role in the future. The remainder of this section focuses first on defining the role of the hospital and then on implementing that role.

DEFINING THE ROLE OF THE PSYCHIATRIC HOSPITAL

Clinical Rationale

The clinical rationale for the role of the psychiatric hospital was well established by the mid-18th century. Edward Jarvis, a nonpsychiatrist physician, well enmeshed in the psychiatric care of his era, stated it as well as anyone. First, Jarvis discussed his thoughts on “diseases of the mind”:

> Maladies of the mind, like those of the body, require that the causes that produced them shall cease to act upon them. Unfortunately the diseases of the mind frequently find their origin in the circumstances and associations of home, in the cares and anxieties of business, in the relations of neighborhood, in the affairs of the town, in the movements of religious, political or other associations, in the habits of indulgences which are practiced. Or when the diseases are once established, they may be kept up by some of these. Moreover, the natural perceptions and moral sensibilities being disordered, the scenes, circumstances, and the persons, which are usually agreeable and favorable, in health, become objects of aversion and sources of irritation and keep the disorder alive. (Jarvis E. *Address Delivered at the Laying of the Corner Stone of the Insane Hospital, at Northampton, Massachusetts.* J&L Metcalf; 1856.)

Then Jarvis provided his thinking on why hospitals were needed to treat persons with these diseases:

> As the lunatic cannot be separated from these, while in his own house, he must go from them and generally among strangers. It is needful that the mentally disordered not only go away from home, but be placed where all the circumstances and associations are different from those to which he has been recently accustomed, and where all the influences that may reach him, can be controlled and modified, so that none but such as are favorable may affect him. They will be found in the hospital [where] the officers and the attendants are selected on account of their peculiar fitness for their work. They have the high intelligence, the scientific knowledge, and the moral endowments of gentleness, and firmness, and sympathy that are required, and they know how to occupy the diseased mind and lead it from its morbid fancies, and fix it upon cheerful and satisfactory subjects. (Jarvis E. *Address Delivered at the Laying of the Corner Stone of the Insane Hospital, at Northampton, Massachusetts.* J&L Metcalf; 1856.)

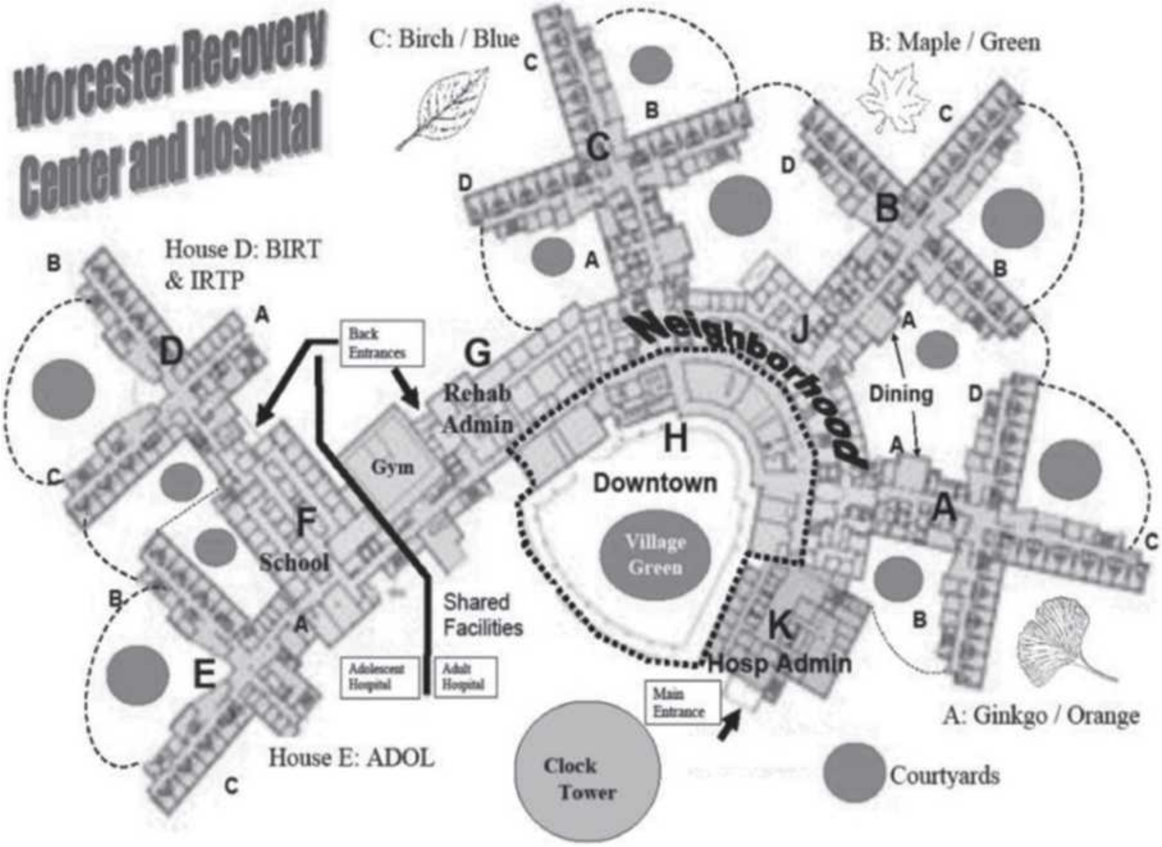
Worcester Recovery Center and Hospital
C: Birch / Blue
B: Maple / Green
House D: BIRT & IRTP
Back Entrances
G
Rehab Admin
Gym
F
School
E
House E: ADOL
Shared Facilities
Adolescent Hospital
Adult Hospital
Neighborhood
J
H
Downtown
Village Green
K
Hosp Admin
Main Entrance
Clock Tower
Dining
A
A: Ginkgo / Orange
Courtyards
A
B
C
D

FIGURE 56.3–5. Worcester Recovery Center and Hospital (WRCH). Interior view.

Jarvis provided a quite optimistic prognosis if his recommendations were followed, but one quite consistent with those of his professional colleagues: “If insane persons are allowed to enjoy the means of healing in the early stages of their disorder, within one year after it appears, about

75 to 90 percent, can be restored to health.” Furthermore, he argued for prompt hospitalization with the onset of symptoms:

> But in this class of maladies, time makes rapid havoc and diminishes the chances of cure. If the means of healing are not tried within the first year, the ratio of cures is reduced perhaps a half in the second and still more in the third year, and when five years of neglect shall have elapsed, the hope of restoration is reduced to a mere accident, which no human skill can promise to accomplish. (Jarvis E. *Address Delivered at the Laying of the Corner Stone of the Insane Hospital, at Northampton, Massachusetts.* J&L Metcalf; 1856.)

By the mid-19th century, it was also clear to those who treated persons with insanity that the hospital had much to offer those in need of acute treatment due to advances in the field. Samuel Woodward, the first president of the APA, noted the following in 1846:

> The abandonment of depletion, external irritants, drastic purges, and starvation, and the substitution of baths, narcotics, tonics, and generous diet, is not less to be appreciated in the improved condition of the insane, than the change from manacles, chains, by-locks and confining chairs, to the present system of kindness, confidence, social intercourse, labor, religious teaching, and freedom from restraint. In this age of improvement, no class of mankind have felt its influence more favorably than the insane. (Woodward SB. Observations on the medical treatment of insanity. *Am J Psychiatry.* 1850;7(1):1–34. doi: 10.1176/ajp.7.1.1.)

Hospital treatment continued to improve in the mid-19th century. Charles Nichols, superintendent of the Government Hospital for the Insane (St. Elizabeth’s), commented as follows in the 1870s: “Much less mechanical restraint, and much less seclusion are now used in our institutions for the insane than were formerly resorted to in them. . . . I trust that we shall be able to still further reduce the proportion of cases in which restraint will be deemed advisable.”

The role the physician could play in the hospital was also clear by the mid-19th century. Luther Bell, superintendent of the McLean Asylum, indicated in 1849:

> Domiciliated with our patients, ready at all hours, by day or night, to watch the effect of our prescriptions, sure of their being fairly and fully carried into effect, without hazard of the interference of timid friends or unfaithful nurses; obliged to make no disguise, no prescription for form or appearance sake, it is in our power to follow out the best convictions of our judgment, uninfluenced by aught extraneous to the object desired. (Bell LV. On a Form of Disease resembling some advanced stages of mania and fever, but so contradistinguished from any ordinarily observed or described combination of symptoms, as to render it probable that it may be an overlooked and hitherto unrecorded malady. *Am J psychiatry.* 1849;6(2):97–127. doi: 10.1176/ajp.6.2.97.)

That the psychiatric hospital played a far greater role than just treating acutely ill insane persons and restoring them to health was quite clear by that time. William Awl, superintendent of the Ohio Lunatic Asylum in Columbus, wrote in 1849:

> More effort and skill is often demanded, to better the condition and make some improvement in the shattered intellects of those who are deemed hopeless, than to effect recoveries. . . We must never permit ourselves to lose sight of humanity, but always remember the importance of the institution to all classes, and the advantages which may be secured to those whom we are obliged to regard as incurable.

Thomas Kirkbride expanded on this note in 1864, indicating the importance of the safety function played by the hospital:

> Owing to various causes, but especially to a neglect of treatment in the early stages of the disease, there are a large number of the insane, in every community, whose chances of restoration are slight, but whose comfort and happiness are promoted, and their protection from ill treatment secured by a residence in these institutions, where they are under circumstances which give them the best chance for permanent improvement. To provide for this numerous class is another important object, attained by the establishment of such hospitals. Next to these come the protection of the community and the relief of private families.

These roles quickly became compromised by the services being overwhelmed by the number of patients needing to be served. By 1890, W. W. Godding, the APA's seventeenth president, was observing:

> The overcrowding goes on, cots are brought out at night and laid down on the corridor floors, at first one or two in nooks and alcoves that seem quite designed for this sort of thing, but the business grows, the special adaptation of recesses and alcoves becomes less apparent as the line of beds side by side stretches in lengthening vista down the hall. And still the floors fill up until one, two, three hundred are thus nightly not accommodated but provided for. From ground floor to attic your hospital, spread with its white beds seems a very bivouac of lunatics . . . Day by day, year after year, I have seen the individualized treatment of special cases swamped by the rising tide of indiscriminate lunacy pouring through the wards, filling every crevice, rising higher and higher until gradually most distinctions and landmarks have been blotted out. (Godding WW. Aspects and outlook of insanity in America. *Am J Psychiatry.* 1890;47(1):1–16.)

These trends persisted, such that by 1954 Kenneth Appel, the eighty-second president of the APA, echoed Godding:

> There is overcrowding 20–30% in 70% of…hospitals, and 50% overcrowding in 8% of the hospitals. Hospitals are 40% understaffed in physicians; 15,000 more psychiatrists are needed; 37,000 patients in 29 hospitals have no psychiatrists, and 2,000 patients are in 5 hospitals which do not employ a full-time physician. (Appel KE. President's page. *Am J Psychiatry.* 1954;110(11):859–864. https://doi.org/10.1176/ajp.110.11.859)

Four years later, Harry Solomon, director of the Boston Psychopathic Hospital, proclaimed:

> After 114 years of effort, in this year 1958, rarely has a state hospital an adequate staff as measured against the minimum standards set by our Association . . . I believe therefore that our large mental hospitals should be liquidated as rapidly as can be done in an orderly and progressive fashion.

Clinical concerns fueled the changing role of the hospital and the changing role of various classes of psychiatric hospitals (state, county, private, general hospital), but clinical concerns were never separate from fiscal concerns. It has never been clear which has been the predominant concern, but there is ample evidence accumulated over the last 150 years that economic considerations have been of no less importance than clinical considerations in defining the role of the hospital.

Social Control Rationale

Every state has statutes that govern the admission of persons with mental illness, over their objection, to hospitals within their state; the involuntary continued hospitalization of these individuals should they want to leave; the commitment of persons who have been charged with a crime and there is a question of their competency to stand trial and/or their criminal responsibility; and the transfer of inmates from jail and prisons to hospitals. Every state sets forth criteria under which an individual can be held on emergency detention and then civilly committed, and these criteria are uniformly danger to self, danger to others, and some variant of gravely disabled. The statutes for civil commitment are generally uniformly applied across hospital types, while forensic commitments are to government-run hospitals or in a few states to nongovernment hospitals with contracts with the state to provide such services.

In no other field of medicine could a competent person be admitted to a hospital over his objection. This occurs only in psychiatry and is the component of psychiatry that is social control. This social control function is the most fractious aspect of hospital psychiatry. It is this aspect of psychiatric hospitals that has been the brunt of antipsychiatry movements, whether they originate with psychiatrists themselves, for example, Thomas Szasz, R.D. Laing; religious groups; or individuals who label themselves “consumer survivors.”

So engulfing had the social control function of the psychiatric hospital become that until the early 1980s, simply being an involuntary patient at a psychiatric hospital meant one could be involuntarily medicated. The argument was a simple one: what was the purpose of holding a person in a hospital against his will if you could not treat him? This changed when, through litigation, the point was made that commitment is based on dangerousness, while the right to refuse treatment is based on competency (see, e.g., *Rogers v. Commissioner of Mental Health*, 390 Mass.489, 458 N.E.2d 308 [1982]).

The objection to the social control function of the hospital is hardly new. One-hundred and seventy-five years ago, Elizabeth Stone protested her admission to the McLean Asylum; 150 years ago, Elizabeth Packard made a career out of her protestations; and 70 years ago, actress Frances Farmer tried to rally the nation to support her self-proclaimed unjust involuntary admissions and treatment.

This week, in every state in the United States, there are individuals in various stages of civil and criminal commitment. There are also individuals being medicated over their objections. And what some believe is a glorious evolutionary step forward in mental health services and others think is a malevolent regression, a progressively larger percentage of involuntary treatment is taking place outside of hospitals, through AOT, mental health courts, conditional release, representative payeeships, leveraged housing, and probation.

So entrenched have these competing interests of social control versus autonomy become that each has its own set of slogans. Those in favor of treatment over objection in a defined set of circumstances talk about “rotting with your rights on.” Those in favor of free choice, no matter what the circumstances, chant: “Nothing about us without us.”

There is one area where both sides together object to the use of the psychiatric hospital as a social control agent—for individuals who do not meet the criteria for having a mental illness, or whose danger is not on the basis of a mental illness, or who are stuck in a hospital because they have no place to go—yet are committed to the psychiatric hospital. All states have their population of individuals who are there because their psychopathic behaviors land them in hospitals; individuals whose high-profile crimes make them “political prisoners,” and individuals whose dangerous sexual behaviors means they must be locked up somewhere. The latter situation is the basis for the Supreme Court decision in Kansas v. Hendricks (see Table 56.3–2).

It is perhaps the irony of the social control function of psychiatric hospitals that law enforcement cannot get individuals into hospitals who need to be there because district attorneys, forensic review boards, families, and victims fight to keep individuals in psychiatric hospitals who do not need to be there. The social control function is highly susceptible to stigma and prejudice.

Fiscal Rationale

In 1912, there was a debate in Congress on a government health insurance program. President Franklin Roosevelt almost established a federal health plan as part of the 1935 Social Security Act. Presidents Harry Truman and John Kennedy had health care proposals before Medicare and Medicaid passed during President Lyndon Johnson's term in office. However, before all that, President Franklin Pierce played a major role in the early eschewal of federal responsibility for the health care of persons with mental illness.

The individual most responsible for bringing the issue to President Pierce's desk was Dorothea Dix, a well-known reformer. On June 23, 1848, Dix delivered a Memorial Lecture to the U.S. Congress. Dix's subjects, as stated in her address, were as follows: "Allowing at the present time 22,000,000 inhabitants in the United States and supposing only one in every thousand to be insane or idiotic, we have then 22,000 to take charge of." She described the state they were in:

> a majority of whom are in needy or necessitous circumstances . . . destitute of appropriate care and protection . . . bound with galling chains, bowed beneath fetters and heavy iron balls, attached to drag chains, lacerated with ropes, scourged with rods, and terrified beneath storms of profane execrations and cruel blows . . . abandoned to the most loathsome necessities, or subject to the vilest and most outrageous violations.

Dix's remedy was hospitals: "No disease is more manageable or more easily cured than insanity; but to this end special appliances are necessary, which cannot be had in private families, nor in every town and city; hence the necessity for hospitals." Dix appealed to the legislators' sense of doing what is right: "Humanity requires that every insane person should receive the care appropriate to his condition." She articulated the social control function of the hospital by spelling out the civil and social obligation to consult and secure the public welfare: "First, in affording protection against the frequently manifested dangerous propensities of the insane." Furthermore, she made a powerful economic argument, indicating the insane and the community at large would benefit by "assuring seasonable and skillful remedial care, procuring their restoration to usefulness as citizens of the republic, and as members of communities."

Edward Jarvis best articulated the 19th-century position on the cost of care versus the cost of treatment that was fundamental to Dix's campaign:

> A man of twenty years of age, if sane, has an average life of 39.48 years, while if insane he has but an average life of 21.31 years if not restored to health. The average time for restoring to health the insane who apply for treatment upon the early symptoms of disease is twenty-six weeks. At $4 per week, which was the average cost in the three State Lunatic Asylums in Massachusetts for the past year, this amounts to $104, to which is added $30 for each patient, for the cost of rent or interest on the value of the hospital, etc., for six months, making an average cost of $134 for restoration to health. If not restored to health, the family or State must be at an expense of $156 a year for 21.31 years, and must also lose the patient's earnings for the 39.48 years which he would have made if well. The cost of the patient's support is estimated at $2,121, while the loss of his future labor, if he becomes insane at twenty years of age, is estimated at $2,665.37, making a total loss of $4,786.37 if not cured; while, if cured in the average time of twenty-six weeks at a cost of $134, there will be a gain to the family or to the State of $4,652. (Jarvis E. Treatment of insanity. In: Mann EC, ed. *A Manual of Psychological Medicine and Allied Nervous Diseases.* P. Blakiston, Son & Co; 1883.)

Dix lobbied for her land grants bill through four presidential administrations. The bill was passed by both houses of Congress during Franklin Pierce's term. However, Pierce vetoed the bill. In his Veto Message on May 3, 1854, first, he laid out the issues as he saw them:

> It cannot be questioned that if Congress has power to make provision for the indigent insane without the limits of the District it has the same power to provide for the indigent who are not insane, and thus to transfer to the Federal Government the charge of all the poor in all the States. It has the same power to provide hospitals and other local establishments for the care and cure of every species of human infirmity . . . The whole field of public beneficence is thrown open to the care and culture of the Federal Government.

Later in his Message, he explained the basis for his veto:

> If the several States, many of which have already laid the foundation of munificent establishments of local beneficence, and nearly all of which are proceeding to establish them, shall be led to suppose, as, should this bill become a law, they will be, that Congress is to make provisions for such objects, the fountains of charity will be dried up at home, and the several States, instead of bestowing their own means on the social wants of their own people, may themselves, through the strong temptation which appeals to states as to individuals, become humble supplicants for the bounty of the Federal Government.

From the veto by President Pierce into the early decades of the 20th century, the federal government was basically disinterested in the psychiatric and substance abuse populations. It eventually turned to the issues around the treatment of this population due to concerns about drug addiction. In 1914, the Harrison Narcotics Act allocated funds to study drug use. Subsequently, the Narcotics Division within the U.S. Public Health Service was created. It was renamed 2 years later the Division of Mental Health. In 1996, the federal government focused on mental illness with the passage of the National Mental Health Act, which established the National Institute of Mental Health (NIMH).

In 1989, the Omnibus Budget Reconciliation Act (OBRA) required the secretary of the Department of Health and Human Services to conduct a study of the IMD exclusion. The IMD exclusion dated from the 1950 Social Security Act Amendments, in which payments on behalf of those in IMDs were excluded from old-age assistance. This exclusion continued unchanged until the enactment of Medicaid under the Social Security Act Amendments of 1965. For the first time, states could use Federal Financial Participation (FFP) for a population in IMDs—in this case, those 65 years of age or older. States could bill the federal government to help pay for the treatment in psychiatric hospitals of patients who were older than the age of 65 years. Other populations were not included, but Congress did expand FFP for services to persons with mental illness in general hospitals.

The 1965 changes to the IMD exclusion began to allow states to shift the cost of mental illness expenditures to the federal government. In a state hospital, the state bears 100% of the cost of care and treatment. The state cannot bill Medicaid. In any facility that is eligible for Medicaid reimbursement, each state pays considerably less than 100%, according to the FFP formula: Nonfederal percentage = 45% × (per capita state income) ÷ (per capita national income). Historically the range of federal participation has been between 50% and 83%—thus the fiscal incentive to move patients out of state hospitals.

An IMD is a facility with more than 16 beds. In addition, any one of the following defines an IMD: (1) the facility is licensed as a psychiatric facility, (2) the facility is accredited as a psychiatric facility, (3) the facility is under the jurisdiction of the SMHA, (4) the facility specializes in providing psychiatric/psychological care and treatment, or (5) the current need for institutionalization for more than 50% of all of the patients/residents in the facility results from mental diseases.

In 1971, Congress expanded Medicaid to include coverage of services rendered in intermediate care facilities (ICFs). In 1972, Congress expanded FFP for inpatient psychiatric treatment in IMDs to individuals younger than 21 years of age so that now those younger than 21 and older than 65 years of age are eligible.

The IMD exclusion sets up a tension between the states and the federal government, each wanting to lower its share of the costs for psychiatric hospital treatment. Between 1972 and 1988, the federal government was busy with evaluating the proper application of the IMD exclusion by the states, disallowing certain states' practices, and defending itself in court. This culminated in the 1985 U.S. Supreme Court decision in *State of Connecticut Department of Income Maintenance v. Heckler,* in which the Court unanimously found for the federal government, indicating that a hospital, a skilled nursing facility (SNF), or an ICF could be an IMD.

Thus, the IMD exclusion profoundly affects the role of the psychiatric hospital and the distribution of patients across types of hospitals. States would rather have persons with mental illness in community residential settings than in state hospitals and in general hospitals rather than state hospitals. In both situations, the former means Medicaid payments; the latter means no Medicaid reimbursement.

No matter what the sector or who the payer, however, in the last quarter of the 20th century, inpatient psychiatric treatment was getting more and more costly—hence the rise of managed care. Starting in the 1970s and continuing through the early 1980s, the principal methods for cost containment were actuarial manipulations designed to manage benefits. There was little regard for the effect these changes in benefits had on the indirect costs borne by beneficiaries, employers, and/or the community. Changes were made without consideration of their effects on the displacement of care onto the general medical sector. Modifications in benefit plan design included the imposition of restrictions, exclusions, limitations, and penalties; provision of incentives; requirements for prior authorization; and retrospective denial of claims. This methodology is managing benefits.

Managing benefits changed into cost control through management of care, a methodology claimed to be more patient-focused. By limiting authorization of expenditures to only those services deemed medically necessary and appropriate, delivered in the least restrictive and least intrusive treatment settings, and only by previously designated qualified practitioners, advocates of managed mental health care believed that proper care could be delivered to patients at the appropriate times, in designated settings, by the right type of provider, in skillfully titrated amounts, and at considerable cost savings. Techniques used to achieve this end included management through contracting, profiling, credentialing, and educating hospitals; organizing hospitals into a specialized, accessible, and interactive network of behavioral health care providers; using mandated standardized pretreatment assessment and treatment planning methods; and requiring the use of standardized forms, restricted formularies, and outcomes management.

Though not yet achieved, the more progressive goals of managed behavioral health care are disease management, demand reduction, health promotion, and wellness. Health maintenance organizations (HMOs) and other capitation-based arrangements have fiscal incentives to use these strategies, but both efforts and results have yet to be really impressive. Nonetheless, the techniques used mean that managed care has and continues fundamentally to affect the role of the psychiatric hospital.

The federal government sought further cost containment. The Tax Equity and Fiscal Responsibility Act of 1982 (TEFRA) led to new methods to control Medicare costs, the most significant being diagnostic-related groups (DRGs) for reimbursement. With the end of cost-plus reimbursement and its replacement with reimbursement based on diagnosis, hospital profits for medical and surgical treatment began to decline. The behavioral health care industry was initially unaffected by DRGs, largely because psychiatric diagnoses were thought to lack sufficient power to predict the course of treatment. Thus, there appeared to be tremendous financial opportunities available in the relatively unregulated behavioral health care market. These opportunities were short-lived; the rise and fall of for-profit psychiatric hospitals were equally steep, as described earlier. With the application of DRGs to psychiatry, this fiscal advantage also dissipated. Thus, the federal government, through Medicare and Medicaid, fundamentally affects the role of the psychiatric hospital.

Although the modern era of managed behavioral health care began in the private sector, these principles and practices spread to the public sector. Medicaid Managed Care was initiated due to the spiraling costs of the Medicaid program. Between 1985 and 1993, Medicaid costs tripled, and the number of beneficiaries increased by more than 50%.

Other Governmental Considerations

Although state governments most directly affect the roles of psychiatric hospitals—through state laws, regulations, licensure, inspections, and investigations—the federal government has profoundly changed the roles of psychiatric hospitals through federal legislation and court decisions. Table 56.3–2 provides a summary of these influences over the last six decades.

The federal government's influence can be examined through the example of the right to treatment. Much of the right-to-treatment litigation has focused on establishing standards of care and concomitant staffing patterns. In the 1940s, the APA maintained there should be no less than 1 psychiatrist for every 150 hospitalized patients, 1 graduate nurse for every 40 patients, and 1 attendant for every 8 patients. No state hospital of the day met all of the APA's standards.

The right to treatment was initially articulated by Morton Birnbaum, a lawyer, and physician in 1960. There followed a series of lower-court cases: *Lake v. Cameron* (1966), *Rouse v. Cameron* (1966), *Nason v. Superintendent of Bridgewater State Hospital* (1968), *Wyatt v. Stickney* (1971), and *New York State Association for Retarded Children v. Rockefeller* (1973), the so-called Willowbrook case. Each of these is briefly explained in Table 56.3–2.

The U.S. Supreme Court decided three cases between 1975 and 1982 that might have established a right to treatment, meaning states that would be obligated to treat committed patients. In *O'Connor v. Donaldson* (1975), the Court found "A State cannot constitutionally confine, without more, a nondangerous individual who is capable of surviving safely in freedom by himself or with the help of willing and responsible family members or friends." In *Halderman v. Pennhurst State School & Hospital* (1977), the Court again did not reach the issue of a constitutional basis for a right to treatment or habilitation. In *Youngberg v. Romeo* (1982), the Court went the furthest it has to date when it held that persons with mental retardation who were involuntarily committed to institutions have "constitutionally protected liberty interests under the Due Process Clause of the Fourteenth Amendment to reasonably safe conditions of confinement, freedom from unreasonable bodily restraints, and such minimally adequate training as reasonably may be required by their interests.” Had the Supreme Court established a right to treatment, the roles of psychiatric hospitals might have been considerably expanded.

The major advances in the right to treatment in the 1980s occurred pursuant to the Civil Rights of Institutionalized Persons Act (CRIPA) of 1980. This act authorized the U.S. government to institute a civil action against any state whose officials, employees, or those acting on their behalf are “subjecting persons residing in or confined to an institution to egregious or flagrant conditions which deprive such persons of any rights, privileges, or immunities secured or protected by the Constitution or laws of the United States causing such persons to suffer grievous harm.” Under this authority, the Civil Rights Division of the U.S. Justice Department has been able to substantially bring the right to treatment closer to a reality for thousands of persons in a state, county, or municipal psychiatric hospitals by applying the standards of *Youngberg* to negotiate consent decrees. A consent decree is a document filed with the court in which the two parties sign a binding agreement in lieu of litigation.

Table 56.3–2.

Federal Government and Court Actions Defining the Role of Hospital Treatment Over the Last Six Decades

1955	The Mental Health Study Act of 1955 (Public Law 182) calls for a nationwide analysis and re-evaluation of the human and economic problems of mental illness. It leads to the formation of the Joint Commission on Mental Illness and Health.
1963	President John Kennedy submits to Congress a “Special Message on Mental Illness and Mental Retardation” that calls for a “wholly new national approach” to those two health problems.
1960	In *Dusky v. United States,* U.S. Supreme Court affirmed a criminal defendant’s constitutional right to have a competency evaluation before proceeding to trial and setting the standard for determination of such competence. This has subsequently meant many individuals are sent to state hospitals every year for competency evaluations.
1964	*Dixon v. Weinberger*, filed in U.S. District Court of the District of Columbia, claims that patients at St. Elizabeth’s Hospital have a statutory right to treatment and that those involuntarily committed under the 1964 Hospitalization of the Mentally Ill Act of the District of Columbia must be placed in the least restrictive setting consistent with suitable treatment.
1965	Medicare legislation is passed. It includes limited coverage for patients receiving active treatment in state hospitals in addition to those in general hospitals.
1966	The Social Security Amendments of 1965 (Public Law 89–97) adds Title XIX, Medicaid, to the Social Security Act. Medicaid funds psychiatric treatment in general hospitals for those under categorical assistance programs and who qualify as medical indigents, and it improves coverage for psychiatric illness under Medicare.
	In *Lake v. Cameron*, a U.S. court of appeals rules that an individual could not be committed to the hospital until

	hospital officials determined that there was no less-restrictive facility available to care for her.
	In *Rouse v. Cameron*, a U.S. court of appeals finds that a criminal defendant who is acquitted by reason of insanity and involuntarily committed to a psychiatric hospital has a right, legally enforceable through habeas corpus, to adequate and suitable treatment.
1970	*Wyatt v. Stickney* is filed in federal district court in Alabama, addressing the issue of detention without treatment of involuntarily civilly committed persons. The court subsequently finds that three fundamental conditions are necessary for adequate and effective treatment in public psychiatric hospitals: A humane psychological and physical environment, enough qualified staff to administer adequate treatment, and individualized treatment plans.
1972	In *Lessard v. Schmidt*, a federal district court decides that under the due process provisions of the Constitution, persons facing involuntary civil commitment are guaranteed the full array of procedural safeguards formerly guaranteed only to individuals charged with a crime.
1973	In *Souder v. Brennan*, a federal district court finds that patient–workers are covered by the Federal Fair Labor Standards Act, entitling them to minimum wages and overtime compensation.
1974	In *Donaldson v. O’Connor*, the U.S. Fifth Circuit Court of Appeals rules that a person who is involuntarily civilly committed to a psychiatric hospital has a constitutional right “to such treatment as will help him be cured or to improve his mental condition.” The U.S. Supreme Court issues its decision in this case on June 26, 1975, finding that “a State cannot constitutionally confine, without more, a nondangerous individual who is capable of surviving safely in freedom by himself or with the help of willing and responsible family members or friends.”

1975	The first class-action suit on the right to refuse treatment is filed for Boston State Hospital patients in a case originally known as *Rogers v. Okin* (right-to-refuse-treatment cases were basically decided on state-by-state basis to the end of the century, with very different rulings among the states).
1975	In *Drope v. Missouri*, the U.S. Supreme Court rules when deciding whether to evaluate a criminal defendant's competency, the court must consider any evidence suggestive of mental illness, even one factor alone in some circumstances. Therefore, the threshold for obtaining a competency evaluation is low. When the issue is raised, the motion should be granted. This means even more individuals sent to the state hospital for evaluation.
1976	In *Tarasoff v. Regents of the University of California*, the California Supreme Court rules that "when a therapist determines, or pursuant to the standards of his profession should determine, that his patient presents a serious danger of violence to another, he incurs an obligation to use reasonable care to protect the intended victim against such danger."
1978	In *Rennie v. Klein*, the U.S. Supreme Court rules that an involuntarily committed, legally competent patient who refuses medication has a right to professional medical review of the treating psychiatrist's decision. The Court left the decision-making process to medical professionals. This decision changed the long-standing practice whereby a committed patient could be treated over his object simply because he had been committed.
1979	In *Addington v. Texas,* the U.S. Supreme Court holds that the standard of proof in civil commitment hearings is "clear and convincing," a lower standard than "beyond a reasonable doubt."
1980	The Civil Rights of Institutionalized Persons Act allows the federal government to initiate actions against states whose public institutions, such as hospitals, prisons,

	nursing homes, and jails, deny residents their constitutional rights.
	In *Suzuki v. Yuen,* the U.S. Court of Appeals for the Ninth Circuit rules that involuntary civil commitment solely on the grounds of danger to property is unconstitutional.
1981	In a 5-to-4 decision, the U.S. Supreme Court rules that inpatients of public psychiatric institutions are not eligible for Supplemental Security Income (SSI) payments granted to mentally ill patients in other settings.
1982	In *Youngberg v. Romeo*, the U.S. Supreme Court decides that a person in an institution has a constitutionally guaranteed “right to personal security,” “a right to freedom from bodily restraint,” and the right to receive “such training as an appropriate professional would consider reasonable to ensure his safety and to facilitate his ability to function free from bodily restraints.”
	Congress enacts the Tax Equity and Fiscal Responsibility Act (Public Law 97–248), which sets the stage for a prospective payment system for Medicare.
1985	In *Ake v. Oklahoma,* the U.S. Supreme Court rules indigent criminal defendants have a right to a competency evaluation. The potential pool of persons who can be sent to state hospitals for evaluation expands again.
1990	The Americans with Disabilities Act (ADA) is enacted to eliminate discrimination against disabled persons. Title II says, “No qualified individual with a disability, shall, by reason of such disability, be excluded from participation in or be denied the benefits of the services, programs, or activities of a public entity, or be subjected to discrimination by any such entity.”
1991	The Patient Self-Determination Act, a part of the Omnibus Budget and Reconciliation Act of 1990, requires health care facilities that receive Medicare and Medicaid

	funding to provide information to adult patients about their right to make their own health care decisions, including the right to accept or refuse treatment and to execute advance directives about medical care.
1997	In *Kansas v. Hendricks*, the U.S. Supreme Court rules that sex offenders can be civilly committed because they “suffer from a volitional impairment rendering them dangerous beyond their control” and because “far from any premature objection, the confinement’s duration is instead linked to the stated purpose of the commitment, namely to hold the person until his mental abnormality no longer causes him to be a threat to others.”
1999	In *Olmstead v. L.C. and E.W.*, the U.S. Supreme Court rules that the ADA requires states to provide community placement for persons with mental disabilities if the state’s treatment professionals have determined it is appropriate, if it is not opposed by the individuals affected, and if it can be reasonably be provided considering state resources and the needs of other disabled persons.
	Publication of *Mental Health: A Report of the Surgeon General.*
2000	The Ticket to Work and Work Incentives Improvement Act permits states to increase the level of income disabled persons can earn before they become ineligible for Medicaid and extends the period that disabled persons can be employed and the income they can earn before losing Medicare coverage.
	Wyatt v. Sawyer (formerly *Wyatt v. Stickney*) settled.
	The Drug Addiction Treatment Act allows office-based heroin treatment by permitting qualified physicians to prescribe take-home doses of buprenorphine.
	Federal legislation authorizes $10 million over 4 years to establish up to 100 mental health courts.
	Federal rules are issued establishing a national uniform standard for protecting the privacy of medical records.

2001	The Surgeon General promulgates a national strategy for suicide prevention: *National Strategy for Suicide Prevention: Goals and Objectives.*
	The Health Care Financing Administration is restructured and renamed the Center for Medicare and Medicaid Services (CMS).
2002	The Department of Health and Human Services releases *Report to Congress on the Prevention and Treatment of Co-occurring Substance Abuse and Mental Disorders*.
2003	The President's New Freedom Commission on Mental Health puts out its report, *Achieving the Promise: Transforming Mental Health Care in America*.
2004	A federal district court in Oregon requires the state to eliminate a $250 copayment for hospital visits for the standard population of the Oregon Health Plan.
2005	The federal government, jointly through the Department of Health and Human Services (DHHS), the Department of Housing and Urban Development, the Department of Justice, the Labor Department, the Veterans Administration, and the Social Security Administration, put out *Transforming Mental Health Care in America. The Federal Action Agenda: First Steps*. This starts a multiyear venture to alter the forms and functions of the mental health system in the United States.
2006	The 2006 Budget Reconciliation Act alters Medicaid rules, such that states can impose premiums and higher cost sharing on beneficiaries, as well as allows states to provide scaled-back Medicaid benefit packages. One consequence is transferring some Medicaid beneficiaries to the rolls of the uninsured, with hospitalization going from Medicaid reimbursement to free care.
	The Substance Abuse and Mental Health Services Administration (SAMHSA) provides a training curriculum to reduce (or eliminate) the use of seclusion and

restraint: *Roadmap to Seclusion and Restraint Free Mental Health Services*.

The CMS publishes final rules on the use of seclusion and restraint in the *Federal Register*. The new limits (previously there were none) are: Adults, 4 hours; children and adolescents age 9–17 years, 2 hours; and children younger than age 9 years, 1 hour.

Implementation of Medicare Part D.

2007 The Agency for Healthcare Research and Quality (AHRQ) of the DHHS reports that there is no evidence base for the off-label use of atypical antipsychotics.

SAMHSA, in *An Action Plan for Behavioral Workforce Development*, reports "an anemic pipeline of new recruits to meet the complex behavioral health needs of the growing and increasingly diverse population in this country."

The AHRQ, in *Care of Adults with Mental Health and Substance Abuse*, reports that 6% of all community hospital stays were primarily for psychiatric or substance use disorders and nearly 25% included such a disorder.

Federal judges in Wisconsin question diagnoses invented to civilly commit sex offenders after expiration of their criminal sentence. In *Brown v. Watters* the diagnosis was "paraphilia not otherwise specified-nonconsent"; in *McGee v. Bartow* the diagnosis was "personality disorder not otherwise specified with antisocial feature." Neither is a *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV) diagnosis.

2008 The Paul Wellstone and Pete Domenici Mental Health Parity and Addiction Equity Act of 2008 (MHPAEA) amended the Public Health Service Act, the Employee Retirement Income Security Act (ERISA) and the Internal Revenue Code, to require group health plans and health insurance issuers to ensure that financial requirements (such as copays, deductibles) and

	treatment limitations (such as visit limits) applicable to mental health or substance use disorder benefits are no more restrictive than the predominant requirements or limitations applied to substantially all medical and surgical benefits.
2010	The Affordable Care Act was passed by Congress and then signed into law by President Obama on March 23, 2010. Titles I. Quality, Affordable Health Care for All Americans and II. The Role of Public Programs and V. Health Care Workforce have direct impact on psychiatric hospitals.
2012	U.S. Supreme Court rules in *National Federation of Independent Business v. Sebelius*on June 28, 2012, in a 5–4 ruling, that the Affordable Care Act is constitutional. The court defines the individual mandate as a tax and not a mandate and provides a choice for States to Opt-Out of Medicaid Expansion.
2013	A final regulation implementing MHPAEA is published in the Federal Register on November 13, 2013.
2013	Representative Tim Murphy (R-PA) introduces, on December 12, 2013, H.R.3717—Helping Families in Mental Health Crisis Act of 2013 which is the first comprehensive federal legislation addressing the needs of those with serious mental illness since President Kennedy’s Community Mental Health Act in 1963.
2014	Representative Ron Barber (D-AZ) introduces, on May 2, 2014, H.R. 4574—The Strengthening Mental Health in Our Communities Act, the House Democrats’ alternative to the bill filed by representative Murphy.
2014	U.S. Supreme Court Justice Ruth Bader Ginsburg files a dissenting opinion in *Burwell v. Hobby Lobby Stores, Inc.*, a decision that exempts employers from covering birth control (as required by the Affordable Care Act) if the employer objects on religious grounds. One of Ginsburg’s queries was, could Scientologist employers not cover antidepressants on religious grounds?

2014	On April 1, 2014, President Obama signed into law the Protecting Access to Medicare Act (H.R. 4302), which includes a demonstration program based on the Excellence in Mental Health Act. This act (1) creates criteria for "Certified Community Behavioral Health Clinics," (2) provides $25 million that will be available to states as planning grants to develop applications to participate in the 2-year pilot, (3) stipulates that 8 states will be selected to participate in the 2-year pilot program, and (4) requires participating states to develop a Prospective Payment System for reimbursing Certified Behavioral Health Clinics for required services they provide.

On July 26, 1990, the Americans with Disabilities Act (ADA) passed and established a "clear and comprehensive prohibition of discrimination on the basis of disability." Title II is of greatest interest because this title covers state and local government services, and so includes public hospitals. Title II indicates "no qualified individual with a disability shall, by reason of such disability, be excluded from participation in or be denied the benefits of the services, programs, or activities of a public entity, or be subject to discrimination by any such entity." One regulation promulgated to operationalize this Act states, "A public entity shall administer services, programs, and activities in the most integrated setting appropriate to the needs of qualified individuals with disabilities." This marks a shift from the long-standing standard, "least restrictive alternative," to the new standard, "most integrated setting." The ADA was seen not so much as a way to improve services within hospitals as a way to force states to move patients out of these institutions.

The hallmark Supreme Court decision to date applying the ADA to persons in hospitals is *Olmstead v. L.C.* (1999). The case involved two women who were in a Georgia state hospital despite the fact that treatment professionals had concluded each could be appropriately treated in community-based programs. The Court found that the "unjustified isolation . . . [of persons with mental disabilities] is properly regarded as discrimination based on disability." The basis for this finding was that "Institutional placement of persons who can handle and benefit from community settings perpetuates unwarranted assumptions that persons so isolated are incapable or unworthy of participating in community life.” The justices then made a leap to the conclusion that institutional confinement “severely diminishes the everyday life activities of individuals, including family relations, social contacts, work options, economic independence, educational advancement, and cultural enrichment.” Thus, began the most recent wave of translocating persons from a public psychiatric hospital to alternative settings. Apparently, general hospitals and private hospitals were deemed “in the community.” So too, based on the most recent data, are jails and prisons.

FULFILLING THE ROLE

The hospital fulfills its role in the care and treatment of psychiatric patients in a prescribed way. Efficacious treatment derives from effectively performing each step.

Assessment

Assessments are done by individual practitioners of different disciplines. An admission assessment is conducted by a physician, preferably a psychiatrist. This assessment, done at the patient’s presentation to the hospital, documents the chief complaint; the basis for admission and history of the present illness; relevant past psychiatric history, family history, and social history; a complete mental status examination; a substance abuse screen; and a risk assessment (suicide, homicide, fire setting, others). This assessment should lead either to admission or to nonadmission with an outpatient plan.

A nursing assessment is completed as soon after the patient’s arrival on the ward as possible, but no later than 8 hours. The nursing assessment overlaps in part with the physician admission assessment but covers some areas in less detail and others in more detail. The nursing assessment includes a pain assessment, a falls risk assessment, skin care concerns, and special nursing care needs.

The third assessment that needs to be completed proximate with admission is the medical history and physical examination. This is done either by the admitting psychiatrist or by an internist, depending on the staffing of the hospital. Besides attending to general medical history, particular attention needs to be paid to how medical disorders and physical findings might be affecting the psychiatric presentation.

All three of these assessments need to cover the areas of alcohol and substance use/abuse. Positive screens lead to an alcohol and drug abuse assessment, preferably completed by staff with expertise in this area.

Other disciplines need to complete their assessments prior to the interdisciplinary team (IDT) meeting at which the Treatment and Recovery Plan (TRP) is developed. These assessments include those by social work—the psychosocial assessment—and by rehabilitation services (often inappropriately referred to as "activities therapy"). The psychosocial assessment needs to provide information from collateral sources when such information is necessary to complete the database, there are collateral sources to contact, and the patient has provided permission (release of information) to contact specific individuals. Some hospitals include a psychology assessment on all admission; others use psychologists who function as consultants, and psychological assessments are conducted only on referral. Dietician assessments are often done routinely; medical specialty assessments are done on referral.

Formulation

Traditionally, the weakest part of the process is from assessment to intervention, and one is often left out entirely. The formulation needs to be seen as a fundamental part of the process. The formulation pulls the data together into a coherent picture of the individual. It is not a brief restatement of the facts of the case. The formulation takes the symptoms, which are the basis for an understanding of the *patient*, and adds significant historical information to provide an understanding of the *person*.

Leaving out or improperly constructing the formulation leads to the loss of important information that does not directly translate into diagnoses but is crucial for effective treatment planning. Formulations may be recorded as components of each assessment—for example, psychiatric formulation, nursing formulation—and then brought together at the IDT meeting, or they may be written at the IDT meeting by the psychiatrist drawing from all of the assessments.

Treatment Planning

Because treatment must commence at the onset of hospitalization, the Initial Treatment and Recovery Plan (ITRP) is, of necessity, a multidisciplinary process. The ITRP is initiated by the admitting psychiatrist and nurse based on their assessments. Other disciplines add to the ITP as they conclude their assessments.

Whereas the ITRP provides the basis for directing treatment during the first few days of hospitalization, the rationales for specific interventions must be included in either the assessments or in the progress notes. For example, each medication ordered by the admitting physician must have a documented rationale. It is not sufficient simply to state that medications taken prior to admission are being continued because this does not provide any basis for the medications. Stating that aripiprazole and clozapine, for instance, are being prescribed fails to justify polypharmacy without providing a basis beyond recording that these were the medications prescribed by the CMHC psychiatrist. Furthermore, any as-needed (PRN) medication ordered at admission needs to have a rationale, one that is specific to the patient. Ordering lorazepam 2 mg orally for agitation and haloperidol 5 mg orally for severe agitation is not sufficient. Such orders are neither specific to the patient nor do they inform the nurse what the prescriber means by “agitation” and what he means by “severe agitation.”

The TRP expands upon the ITRP and derives from an *interdisciplinary* process. The IDT is composed of professional and paraprofessional staff, as well as the patient. The patient, assisted by staff, is free to invite any other individuals to this meeting, for example, family members, friends, or advocates. The purpose of the IDT meeting is to *write* the TRP. It is not a time to sanction a TRP written before the meeting by one member of the IDT. The meeting uses multidisciplinary assessments as the basis to create an interdisciplinary treatment plan.

The creation of the TRP follows an orderly process. The IDT establishes, with the patient as the lead, the problems, which in the inpatient setting are predominantly barriers to discharge. Other problems may be addressed, depending on the expected length of stay of the patient. For each problem, there is a long-term goal (LTG) and a set of short-term goals (STGs) or objectives. The goals are the patient’s goals, not the staff’s goals for the patient. Hence, problems and goals belong to the patient. For each STG, there is a series of interventions. Interventions are what the staff does, not what the patient does. Each intervention under an STG must specifically assist the patient to move toward the achievement of that STG.

STGs need to be specific, measurable, attainable, realistic, and timely (often referred to by the acronym SMART). STGs are intermediate steps toward the accomplishment of the LTG. Small steps that the patient reaches at intervals reinforce the patient and the staff.

Interventions need to specify the frequency of the intervention, the duration of the intervention, and the name of the intervener. For example, “Nurse Nightingale provides Medication Education Group, Tuesday and Thursday, 10:00–10:45 a.m.”

Common errors in interventions include the following:

1. Writing job descriptions: “Nurse Nightingale will administer medication and monitor for side effects.”
2. Failing to document intervener and/or duration: “Anger Management Group Wednesdays.”
3. Including interventions unrelated to the STG, such as “Dr. Woodward will meet with patient Mondays, 1:00–1:20 p.m. to review concerns about weight gain from clozapine,” when the STG is “Patient will have one five-minute telephone conversation per day with mother for three consecutive days without telling her mother to drop dead or piss off.”
4. Using terms that communicate no specific intervention (and can often be dangerous to staff if done wrong) such as “encourage patient to . . . ” or “assist patient to . . . ” If such interventions are used, the intervention must be explicit about how the “encourage” is to be done.

The ideal sets of interventions are clear, concise, and explicit. If the entire IDT were out sick one day, covering staff should be able to treat the patient by performing the interventions.

The TRP is the road map for the patient’s treatment. Active treatment—a concept or phrase used by many certifying entities—is defined as that treatment found as an intervention on the TRP that moves the patient toward the achievement of an STG that moves the patient toward the achievement of an LTG that addresses one of the patient’s problems. A simple rule when one is calculating active treatment hours for a patient is, “if it’s not on the treatment plan, it’s not active treatment.”

Progress is recorded in progress notes. These are written pointedly addressing STGs. Each note needs to be labeled with the discipline of the intervener and the STG being addressed. Thus, there would be no nursing staff note reading “Patient ate well and slept well” unless eating and sleeping were problems on the TRP with specific LTGs and STGs.

The initial TRP is completed within a specified time frame, varying between 1 and 10 days, and specified by the facility's policies and procedures. The TRP is reviewed through IDT meetings at which treatment plan reviews (TRPRs) are conducted. The TRPR builds on the TRP, uses progress or lack thereof noted in the progress notes, and modifies the TRP as needed. TRPRs also occur on a schedule. In units with short lengths of stay, the TRPR occurs from daily to every 3 days. In intermediate length of stay units, TRPRs occur weekly for 60 days, then every 2 weeks for 60 days, then monthly. In longer stay units, TRPRs may occur monthly after the TRP done at 10 days.

Besides occurring on a scheduled basis, TRPRs need to occur in response to a patient's clinical need. Many hospitals have thresholds that, when met, require a TRPR to be held within one business day. Such thresholds included (1) any episode of seclusion or restraints; (2) a specific number of episodes of seclusion or restraint or a specific number of hours of seclusion or restraint within a designated time frame; (3) the administration of any psychiatric medication PRN or a given number of PRNs within a specified time frame; (4) the initiation of 1:1 observation; and/or (5) any serious incident, for example, patient-on-patient assault, patient-on-staff assault, or suicide attempt.

The TRP needs to be comprehensive or account for why it is not (either in the TRP or in the hospital's policies and procedures) in all but acute care settings. Thus, the TRP would be expected to address all of the patient's psychiatric and medical problems on longer-term units. A TRP could provide interventions for those problems that are impediments to discharge and address other problems as resources allow; this is often the case on intermediate-length-of-stay units. A CTP could address only those problems that are impediments to discharge with the expectation other problems will be addressed in treatment after discharge; such an approach is often employed on short-term units. This does not mean that maintenance treatments will not be provided for acute or chronic problems that are not the focus of the hospitalization, such as antihypertensive medications, treatments for gastroesophageal reflux disease or constipation, and hypnotic medications.

There is a new emphasis on tying the hospital TRP to a community TRP for persons with SMI in systems of care (traditionally CMHCs, but now also health homes and other integrated care settings). In the person-centered approach, the community TRP starts with the individual's "life goals.” Community clinicians, support and outreach staff (peer and nonpeer workers), and naturalistic resources are channeled toward assisting the individual to move ever closer to his life goals. Hospitalization is usually a detour on the journey toward one’s life goal, so focusing on problems that are impediments to discharge is in the service of returning the person to the community to take up the pursuit of his individualized life goals. To do so effectively, the hospital TRP should record the life goals as anchor points.

Treatment

Treatment planning is necessary but not sufficient for effective treatment to ensue. Members of the IDT plus other staff need to provide integrated treatment that follows the plan.

Psychopharmacologic treatment is a mainstay of contemporary inpatient treatment but must be prescribed (physician), dispensed (pharmacist), and administered (nurse) with awareness of contemporaneous interventions. Either initiating psychotropic medication or adjusting medication regimens is considered the sine qua non of inpatient psychiatry. Of course, not all admissions require medication, and distinguishing these patients from all others is an important function of the assessments and observations that are unique to the inpatient setting.

In prescribing medication, the attending psychiatrist must communicate the reasons for the medication and its side effects not only to the patient (obtaining informed consent before ordering the medication) but also to all members of the IDT. Hence, the importance of monitoring progress and informative progress notes accompanying each order.

PRN medications warrant particular attention. PRN medication transfers de facto prescription authority from physician to nurse because the nurse decides when and how much medication the patient will receive (within the parameters set by the physician). Unless the physician checks daily on the total dose of each PRN medication given, it is not possible for the physician to relate the patient’s progress to the medication administered. Hospitals have taken various approaches to PRN psychotropic medication, including eliminating them entirely, limiting the orders to short-term orders, for example, 3 or 5 days, or allowing PRNs only on admission units and not on other units within the hospital.

Well-written progress notes by the psychiatrist assist the nurse in writing useful nursing progress notes. Writing “patient remained calm and cooperative throughout shift” may convey no useful information if, for example, the patient is depressed with a recent suicide attempt and is externally calm and cooperates because it takes less energy and concentration to do so than not to do so. A physician progress note detailing the specific target signs and symptoms for which antidepressant X has been prescribed provides the foundation for the nurse to write progress notes that focus on those signs and symptoms. If the physician simply writes antidepressant X for “dysphoria” or “depression,” progress notes by other staff members are likely to be equally uninformative.

Another function shared predominantly by the psychiatrist, and the nursing staff (Registered Nurse [RN], Licensed Practical Nurse [LPN], Mental Health Technicians [MHTs]) is patient safety. Maintaining a safe milieu and maximizing the safety of each patient and staff depends on (1) active, individualized treatment, (2) levels of observation and privileges geared to the patient’s level of risk, (3) prompt and skillful response to patient aggression, (4) modification of therapeutic interventions as the nature and/or level of risk changes, and (5) continuing staff education and training.

Level of observation and privileges serve the function of patient safety and should not be used as rewards for compliance with process variables like attendance at groups. When the level of observation is for the risk of harm to self, the most noteworthy harm being suicidality, the hospital needs a process by which increasing the level of observation, that is, more staff oversight can easily be instituted by a physician or RN, but decreasing the level of risk can only occur on a written physician’s order after a face-to-face evaluation. In some hospitals, this can only be done by the attending psychiatrist. Telephone orders and verbal orders should never be accepted for lowering the observation level.

In addition to medication, inpatient treatment involves individual sessions and group interventions. Individual psychotherapy is the exception these days, although this modality continues to be used in some public and some private hospitals. There are a few hospitals in which intensive psychotherapy is the mainstay of inpatient treatment. The programming of individual sessions with the social worker to focus on discharge planning is the rule. Individual sessions with nursing staff to practice skills taught in groups should occur at least through the day and evening shifts (and the night shift if appropriate to the patient’s needs).

Group interventions fall under the heading of psychosocial rehabilitation. In general hospitals, these groups occur on the unit. In state hospitals and in some private hospitals, these interventions are delivered off-ward in areas of the hospital, often referred to as the “treatment mall.” There are several advantages to setting up groups or classes in a centralized location, including the following:

1. Patients follow a variant of the “work order day”; patients, like staff, “commute” to where they spend their daytime hours; in the case of patients, they live in one place (the unit) and get their treatment in another location (the mall); lunch is often provided at the mall.
2. Pooling patients and consequently pooling staff means that a large number of different groups can be provided per hour, rather than having the same group occur simultaneously in multiple locations.
3. The issue of patients remaining in bed, watching TV, or participating in other diversions is handled by institutional policy and practice rather than day by day, case by case.
4. The safety of staff and patients is increased because there is more staff available for an emergency.
5. Professional and paraprofessional staff spend more time with patients because patients and staff are at the treatment mall during mall hours.

Newly constructed hospitals (see Table 56.3–1) are designed with accessible treatment mall locations. The use of perimeter security with significant freedom of travel for patients within the building eases access to mall-based treatment and maximizes patients’ autonomy (see Fig. 56.3–5). One controversy surrounding the use of treatment malls at hospitals is whether or not the hospital can require all patients to attend the mall. Does the right to refuse treatment also mean the right to not go where treatment is provided? In one model, patients have a right to refuse treatment, but patients do not have a right to determine where they are at the hospital at any given time. If this were a right, patients could refuse to remain behind any locked door. Patients must go to the treatment mall, but the hospital must provide a place for them on the mall if they refuse to attend the groups they are scheduled to attend. A corollary controversy in this model relates to the issue of whether staff can lay hands on a patient if the patient refuses to go to the mall. Staff can lay hands on a patient if there is an emergency, and leaving a patient on a ward with no staff because of the patient’s refusal to leave the ward would constitute an emergency. In another model, where patients cannot be compelled to leave the unit for groups, hospitals have created units with and without an off-unit requirement. There are benefits and drawbacks of each type of unit and absent safety concerns; a patient makes his choice based on his valuation of the different benefits of each type of unit.

Although there are no controlled studies of the operation and effectiveness of treatment malls, data collected at individual hospitals seem to support (1) rapid adjustment of patients to a change to the treatment mall approach; (2) very low use of hands-on procedures to move patients; (3) decreased rates of seclusion and of restraint without accompanying increases in the total daily dose of antipsychotic and/or antimanic medications or of benzodiazepines; and (4) improvement in patients' sleep with fewer doses of hypnotics. Most importantly, patients do seem to learn skills through the use of mall-based groups.

Clinicians doing hospital-based treatment, just like their colleagues doing outpatient treatment, need to be mindful of three contemporary concepts: recovery, evidence-based treatment, and trauma-informed care. Practitioners of all clinical disciplines have struggled to achieve consensus definitions of these percepts. SAMHSA's definitions are as follows:

Recovery: A process of change through which individuals improve their health and wellness, live a self-directed life, and strive to reach their full potential.

Evidence-based: Demonstrating effectiveness in empirical research that meets a standard of scientific rigor as embodied in minimum review requirements, which include the stipulation that an intervention must have demonstrated one or more positive behavioral outcomes ($P \leq .05$) in substance abuse and/or mental health in at least one study using an experimental or quasi-experimental design.

Trauma-informed care: It is a strengths-based service delivery approach that is grounded in an understanding of and responsiveness to the impact of trauma, that emphasizes physical, psychological, and emotional safety for both providers and survivors, and that creates opportunities for survivors to rebuild a sense of control and empowerment.

A well-functioning psychiatric hospital would incorporate these principles into its therapeutic offerings by having expertise in and providing, among other offerings: cognitive behavior therapy for psychosis (CBTp) cognitive enhancement and restructuring therapy (CERT); cognitive restructuring for PTSD; cognitive therapy for psychosis (CT-R); dual recovery (mental illness and substance abuse); dialectical behavior therapy; family support and treatment; fire-setting treatment; grief therapy; illness management and recovery (IMR); individualized exercise plan (IExP); leisure skills development; medical care and treatment; mental illness and problematic sexual behavior treatment (MIPSB treatment); motivational interviewing and motivational enhanced therapy; person-centered approach; pharmacotherapy; positive behavioral support plan; readiness for change interventions (individual and group); sensory modulation; vocational rehabilitation; and wellness management recovery action plan (WRAP). Behavioral interventions, generally under the direction of a psychologist with particular expertise in this area, need to be available at all hospitals. This is of paramount importance in hospitals that have units with intermediate- or long-stay patients. The behavioral plan is based on a functional assessment of the target behavior to be modified or eliminated. The plan provides a series of responses aimed at reinforcing some behaviors and extinguishing others. All direct care staff must be trained in the specific interventions to be used for a specific patient. Preferably, the hospital has a way to document such training and to ensure that any staff assigned to work with a patient who has a behavior plan is trained in that plan.

During hospital treatment, a major controversy has been patients' "right to smoke." It is unclear how this is a right. If it were, patients would also have a "right to drink." No one argues the latter, so why the former? Forbidding smoking on the unit has generally occurred at all hospitals. Moving that restriction to a designated area within the hospital was the next step. Patients can often be seen at the perimeters or in the courtyards of the hospital smoking. Many hospitals have progressed to prohibiting smoking anywhere in the building, and some even prohibit smoking on the hospital grounds. Hospitals should neither be in the business of selling cigarettes nor use staff to purchase cigarettes for patients—a practice referred to as "making a cigarette run." Given that there is now a range of aids to stopping smoking—gum, patches, pills, support groups—these should be made available to patients and to staff. Nicotine replacements should be used as a transition to becoming nicotine-free; too often, they become permanent fixtures on the patient's pharmacologic plan. It was not that long ago that state hospitals received free tobacco from cigarette companies, purchased cigarette papers, and paid staff, and/or had patient groups whose focus was to roll cigarettes. The rolling machines had been purchased with state funds. Progress in smoking cessation needs to continue, particularly given the increased morbidity and mortality of persons with chronic mental illness. The role of the hospital should be to teach patients and support their efforts to maximize a healthy lifestyle. Wellness programs are being conducted at many hospitals and are well worth supporting.

Discharge Planning

The aphorism “Discharge planning begins at the time of admission” is, of necessity, particularly true today, given all of the pressures surrounding the length of stay. Although the discharge process is primarily the responsibility of the IDT’s social workers, it is a mistake to assume it is their task alone. Discharge planning is an interdisciplinary process.

Discharge planning is processed by the social worker interacting with the patient and IDT members until the TRP. At this point, the discharge plan becomes an integral part of the TRP. Whether discharge-related issues are identified as one of the problems on the TRP depends on the patient’s specific circumstances. Discharge might be a problem if the patient is homeless, cannot return to the residence from which they were admitted, or is an illegal immigrant with no eligibility for entitlements. An often-missed problem related to effective discharge is recidivism; that is, the patient has multiple admissions, generally with short community tenure between them. Recidivism is often too quickly ascribed to noncompliance with medication without the IDT doing a thorough investigation of the etiology of repeated admissions. Noncompliance may well be a factor in any short time period between admissions, but this may also be accounted for by a number of other factors. Often ignored is the contribution to stabilization made by the structured environment of the hospital. This is often the primary factor in recommending discharge to step down over “home” to allow for transition with continued structure. Often commented on but inadequately addressed is the use of alcohol, marijuana, other street drugs, and other prescribed medication even in a medication-compliant patient. There is also the contributing factor of secondary gain—the hospital may be the locus of the best quality of life the patient has experienced for years or even decades.

Discharge planning involves the establishment not only of sites of service and appointment times but also of efficacious linkage strategies. The steps that need to be taken to maximize the likelihood that the patient will actually make it to their first appointment with the outpatient psychiatrist are of paramount importance. Linkage also involves providing to the outpatient psychiatrist, outpatient therapist, and other appropriate personnel pertinent information in writing. A copy of the instructions given to the patient should reach the outpatient sites *before* the patient arrives for their first appointment; the discharge summary needs to follow shortly thereafter. So too does the TRP, so that the community can incorporate the work the hospital has done into the community TRP, just as the hospital incorporated the work the community had done into the hospital TRP. Phone contact from inpatient psychiatrist to outpatient psychiatry is advisable for any case seen as needing more than the usual written communications.

SPECIAL POPULATIONS

This section has focused on the role of the hospital in serving the general adult psychiatric population. The hospital serves many other populations, defined either by their demographics or by specific service needs.

Children and Adolescents

Twenty-one percent of U.S. children aged 9 to 17 years have a diagnosable mental or addictive disorder causing at least minimal impairment, and 5% to 9% of children this age have a serious emotional disturbance that causes extreme functional impairment. Significant numbers of children and adolescents in need of treatment do not receive behavioral health services. When children's mental health services are unavailable, unaffordable, or inappropriate, many young people end up in the child protection or juvenile justice systems.

Children and adolescents represent approximately 7% of the total mental health inpatient population. However, children's services receive a disproportionally low fraction of what adults receive for inpatient services, despite the fact, that on average, children's inpatient treatment is about 1.5 times as expensive as adult inpatient treatment.

Although the number of available inpatient psychiatric beds decreased, the admission rates of children and adolescents to psychiatric hospitals increased over the last few decades. Between 1986 and 1997, the number of youth admitted to inpatient services increased by 143%, and the rate per 100,000 youths increased by 119%. As the number of inpatient beds and specialized staff to care for children and adolescents with serious emotional disturbance has decreased, the number of families relying on hospital EDs for psychiatric services has increased. ED physicians report seeing a large number of children whose parents have simply run out of options for care.

Inadequate hospital-based and other services have created cohorts of labeled children. “Boarder kids” are children stuck for days, weeks, or even months in pediatric wards because there are no beds in a psychiatric hospital or ward. “Wait-list kids” wait months for case management or outpatient services. “Stuck kids” are in psychiatric hospitals but cannot leave because there is no place for them to go.

There are many factors that complicate the role of the hospital in the care and treatment of children and adolescents. Alcohol and illicit drugs are major problems. Current alcohol use is 36.2% for 17-year olds, 19.9% for 15-year olds, and 6.5% for 13-year olds. Among youth ages 12 to 17 years, about 12% are current illicit drug users. Annual suicide attempts in 15- to 24-year olds have surpassed 1 million in the United States, and suicide is the third-leading cause of death for this age group. Age-specific mortality rates from suicide are 1.6 per 100,000 for 10- to 14-year olds, 9.5 per 100,000 for 15- to 19-year olds, and 13.6 per 100,000 for 20- to 24-year olds. Almost half of the students with mental disorders drop out of grades 9 through 12, and of those dropouts, 73% are arrested within 5 years of leaving school. Children and adolescents with serious emotional disturbance often live in families with multiple risk factors and vulnerabilities, including single-parent households, poverty, and family history of substance abuse, violence, mental disorders, and/or felony convictions.

Although the challenges for families of children and adolescents with emotional and behavioral problems are great, best practices emphasize collaboration with families when coordinating care for the child. Inpatient facilities have placed an increasing emphasis on incorporating the family and other relevant support systems into a child’s treatment plan.

Today, children and adolescents are likely to receive mental health services in what is referred to as “systems of care.” Systems of care provide a spectrum of mental health and other necessary services organized into a coordinated network from community-based and least restrictive treatment settings and case management to intensive inpatient psychiatric care. Systems of care rely on “wraparound services.” The values underlying systems of care include a focus on the individual child and family, the provision of services within the most normalized environment that meets the youth’s need, the creation of partnerships with families, the use of a strength-based orientation, the creation of an environment of cultural competence, the provision of unconditional care, and the use of 24-hour institutional care at various points in time. Systems of care do not eliminate the need for inpatient level of care but often shorten lengths of stay.

As the wraparound model appears to have evolved into a best practice in the care of the most difficult children and adolescents, it is important to note that inpatient hospitalization treatment is not a failure of the wraparound model but one of its service components. Thus, the hospital plays a key role in wraparound services.

A challenge that most states face is providing services to children with emotional and behavioral disturbances when they age out of the child system. The eligibility requirements for services in a state’s adult system are many times narrower than the eligibility in the child system, so some transition-age youth who have been receiving services for years are screened out, leaving them without care and treatment. A vulnerable time for a young adult is exacerbated by the states’ need to restrict its eligibility pool for publicly funded mental health services.

Veterans

The veterans’ hospital system was established after World War I when veterans were entitled to health care benefits and compensation. States were unable to handle the needs of this population, given that there were hundreds of thousands of U.S. soldiers “wounded not mortally” and uncounted numbers who suffered from neuropsychiatric conditions.

On his last day in office in 1921, President Woodrow Wilson signed into law a bill supporting $18.6 million for the establishment of hospitals for veterans. The years 1921 to 1923 saw the provision of more than 6,000 beds in new veteran’s hospitals across the United States. From 1923 to 1945, there was a considerable expansion of hospital beds. Following World War II, the VA was entirely reorganized. In the decades following World War II, VA hospitals accounted for only a small percentage of the hospitalized population. Trends in bed use shifted, however, such that between 1955 and 1975, the number of VA inpatient episodes increased by 143%, whereas state and county mental hospitals decreased by 37%.

In the mid-1990s, the VA shifted its traditional focus on inpatient care and dramatically expanded its outpatient services. Between 1995 and 2001, the number of occupied annual inpatient psychiatry beds fell by 58% in VA hospitals, the length of stay decreased by 52%, and the number of veterans treated in the VA outpatient system increased by 27%. This shift, and the reorganization of the Veterans Health Administration (VHA) into 22 smaller, semiautonomous networks, may be contributing to the VA's apparent inability to meet the needs of veterans from combat in Afghanistan and Iraq. And that need is substantial. A nationally representative sample drawn from the National Health and Resilience in Veterans Study showed that the current and lifetime prevalence of PTSD in veterans was 4.8% and 8%, respectively. PTSD came from combat and noncombat situations and was associated with elevated risks of severe psychiatric disorders and suicide.

Efforts were initiated in 2006 to substantially improve the capacity of VA hospitals to meet the increasing psychiatric need of veterans. The VHA struggled to provide the necessary service. A 2010 report estimated that 18 veterans a day die by suicide and 950 suicide attempts occur each month among veterans receiving services through the VAH. Between 2010 and 1014, VA hospitals were the subject of scandals and problems, with systemic problems including turning away veterans who needed services and who then suicide, maintain secret waiting lists to create the appearance that delay in service was much shorter than it was, putting out false performance reports, covering up suicide attempts and running facilities in which illicit drug dealing was routine practice. As one commentator said, “the VAH is a bright, shining example of misguided priorities and terrible mismanagement.”

Veterans seek and/or receive treatment out of VHA settings for a number of reasons: The veteran is not service-connected for a psychiatric disability, the veteran prefers treatment out of the VHA, the veteran is a committed pursuant to a forensic status, or in some jurisdictions, the veteran refuses voluntary admission and requires civil commitment. Thus, continuity of care can be particularly problematic as the veteran moves in and out of the VA, public, and private systems of care.

The challenges facing the VAH are significant. In 2014, there were 24 million veterans in the United States, about one-third of whom were enrolled in VA care. There are 21,000 mental health professionals to meet this need, but only 3,000 are psychiatrists. The expectation is that the Veterans Access, Choice, and Accountability Act of 2014 will improve the ratio of VA psychiatrists, non-VA psychiatrists, and other mental health clinicians available to treat veterans with psychiatric disorders. The VA MISSION Act authorizes or expands several programs with the intention of recruiting and retaining health care providers in the Veteran's Health Administration. Among other things, the Act increases the maximum amount of student loan debt that may be reduced under VA's Education Debt Reduction Program (EDRP); authorizes designated scholarships for physicians and dentists under the VA Health Professional Scholarship Program (HPSP); establishes the VA specialty education loan repayment program to incentivize VHA employees to pursue education and training in medical specialties for which VA determines there is a shortage; establishes a pilot Veterans Healing Veterans Medical Access and Scholarship Program, and extends eligibility for VA's EDRP to clinical staff working at Vet Centers. The Veterans Access, Choice, and Accountability Act of 2014 (VACAA) required an increase in the number of GME positions by up to 1,500 over a 10-year period, beginning July 1, 2015, through 2024, with an emphasis on primary care, mental health, and other specialties the VA Secretary deems appropriate. The VACAA also required that the expansion focus on creating new programs at VA facilities where there had not previously been GME training and that the new positions be awarded for training in rural or otherwise underserved areas. As of September 2017, the VA had allocated about one-third of these new positions; it has not completed the expansion. Instead, the VA's objective is to complete the expansion by 2025.

Individuals active in the military receive their psychiatric inpatient treatment in Department of Defense hospitals. Of major concern for this population is suicide. The U.S. Army suicide rate exceeds the civilian suicide rate with substantially elevated suicide rates following psychiatric hospitalization—an eightfold increase for 3 months following a psychiatric hospital discharge and a fivefold increase for 12 months after psychiatric hospitalization. The Patient-Centered Community Care (PCCC) program contracts were awarded in September 2013 to provide inpatient and outpatient specialty care and mental health care for eligible veterans when the local VAMC cannot readily provide the services, such as when there is a lack of available specialists, there are long wait times, or there is an extraordinary distance from a veteran’s home.

Substance Abuse Comorbidities

Persons whose alcohol and/or substance use require hospital-based treatment fall into two categories: Those with so-called “straight substance abuse,” meaning that there is no other mental illness that is the focus of treatment, and persons who are dually diagnosed with a mental illness and a substance use disorder both of which require active treatment.

Treatment for substance use in either population requires detoxification when such is necessary and then treatment initiating prevention of relapse and recovery. Relapse and recovery are lifelong enterprises that may or may not involve inpatient treatment. With the exception of detoxification of persons at high medical risk, all the services for persons with substance abuse or individuals dually diagnosed can be found at a hospital and nonhospital loci of care.

Perhaps the major shift in the treatment of persons with dual disorders that occurred in the latter part of the 20th century was the movement from consecutive treatment to concurrent treatment. For decades, there was the belief that one could not address other psychiatric disorders until one had first addressed alcohol and/or substance abuse or dependence. Currently, efficacious treatment is considered to be treatment that addresses both at the same time.

Because the rate of alcohol and substance abuse is quite high in persons with other *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5) disorders, hospitals need to have expertise in substance abuse assessments and treatment at all levels of service. Many a diagnosis will be missed if the attending psychiatrist or the C-L psychiatrist is not thinking about the potential roles that alcohol and illicit drugs may be playing in the patient’s presentation. The legalization of marijuana or medical marijuana in some states is presenting new dilemmas for psychiatric hospitals and C-L psychiatrists. It is too soon to tell how this will be dealt with in different states, and within one state, in psychiatric inpatients settings in different types of hospitals.

Geriatrics

Patients older than the age of 65 years and those deemed “functionally geriatric” are treated in units with mixed-aged populations in some hospitals and on psychogeriatric units in others. The involuntary admission of older adult individuals varies greatly from state to state, in that some state statutes prohibit the involuntary admission of persons with any form of “organic brain syndrome” as the basis for the commitment, whereas others do not.

Older adult patients have special needs that the hospital must attend to (1) greater comorbidity with medical disorders; (2) increased likelihood of delirium; (3) possibility of their having another disorder that is confused with dementia, especially major depressive disorder; (4) greater need that all other medications they are taking be considered when the assessment is done and when psychopharmacologic recommendations are made; (5) higher suspicion for impaired capacity to consent to treatment; and (6) more likely to need additional social work care and services to assist for example in Adult Protective Services (APS) reports, identifying and contacting alternative decision markers, addressing end of life care concerns.

Of particular concern is the use of antipsychotic medication in older adult patients. In 2005, the FDA warned that second-generation antipsychotic medication was associated with an increased risk of death in older patients with Alzheimer’s disease and other age-related dementias. In 2008, the warning was expanded to include first-generation antipsychotics. Subsequently, a Harvard Medical School study of older adults, new users of antipsychotic medication showed that haloperidol had twice the mortality rate of the atypical antipsychotics, and among the atypicals, quetiapine had the lowest mortality rate. Another class of medications of concern is benzodiazepines, ubiquitous in their use in hospital settings to quiet agitated patients, including older adults. A case-controlled study using Quebec, Canada health insurance data showed chronic use of benzodiazepines by older adult patients was associated with about a 50% increase in Alzheimer disease 5 to 10 years after a period of 90 or more days of use, with benzodiazepines with half-lives of greater than 20 hours carrying the greatest risk. Thus, caution should be used in the inpatient management of older adult patients by using benzodiazepines for longer than brief—up to 90 days—treatment.

Staff in any of the hospital services should consider the need for, and implement if necessary, a complete organic medical workup, formal cognitive assessments, and referral for follow-up services that are age-sensitive.

Developmental Disabilities Comorbidities

Most states have deinstitutionalized persons with developmental disabilities/delays at an even greater pace than their population with chronic mental illnesses. That means that former residents of developmental disability facilities who would not have been seen for hospital treatment now are.

Persons who present for services at hospitals with anywhere from mild to profound developmental disabilities may be coming from their home where they live with parents, from community residential programs where they live with staff, or, in the case of mild mental retardation, from an independent living arrangement. In each instance, the person needs individualized services that are informed by their capacities. Especially important is the use of concepts and language the individual can comprehend.

Serving persons with developmental disabilities in the hospital has evolved in many different ways. In most hospitals, these individuals receive services in the same areas of the hospital as all others. There are certain disadvantages to this approach in that the essential skills to meet the needs of the population may be lacking. There are a few general hospitals that have specialty units to serve persons with developmental disabilities and even more state hospitals that do.

When the person with developmental disabilities presents for services at the hospital, it is most often due to unmanageable behavior. That is not to say that individuals with developmental disabilities do not become psychotic, depressed, or anxious; they surely do, and these conditions need psychiatric attention. However, a majority of persons dually diagnosed with a mental illness and developmental delays present because they can no longer be dealt with in the environment in which they live. Commonly the individual receives such diagnoses as impulse control disorder not otherwise specified, intermittent explosive disorder, or, if adolescents or younger children, conduct disorder or oppositional defiant disorder.

Persons with developmental disabilities should leave the inpatient setting with a functional analysis of the problematic behavior that permits the outpatient setting to follow through with treatment that will facilitate habilitation. Often this entails the use of behavioral interventions. All hospital treatment of persons with developmental disabilities entails close collaboration with the team of persons providing all outpatient and residential services. Collaborative treatment is essential in the hospital's achieving successful outcomes with this population.

Medical Comorbidities

Psychiatric disorders complicated by or complicating medical disorders are a common occurrence. Psychiatric services are often required when there is a diagnostic question related to mentation or mood; when a patient is too "out of control" or their behavior is so "inappropriate" that management of the patient is not going well on the medical or surgical services; when the medical or surgical practitioner cannot communicate with the patient or cannot obtain informed consent; or when there is an ethical question about the course of treatment.

There is perhaps no better example of the psychiatric hospital's role in treating persons with medical illness than HIV infection. There is a bidirectional relationship between severe mental illness and HIV infection; persons with psychiatric disorders are at higher risk of acquiring HIV, and persons who are HIV positive have increased rates of psychiatric illness. The interactions between disorders and treatments are myriad. Psychiatric disorders may impair judgment and heighten the risk of contracting HIV (e.g., mania in bipolar disorder) or lead more directly to increased risk (e.g., intravenous drug use). HIV can induce, both psychologically and by a direct effect on the brain, anxiety, depression, psychosis, and dementia. Antiretroviral medications can induce psychosis. Drug–drug interactions—drugs prescribed for HIV status, psychiatric status, and medical disorders due to HIV status—make treating the comorbid psychiatric disorders quite complex. Being HIV positive per se, the symptomatology of HIV, the psychiatric disorders that predated HIV-positive status, and the psychiatric disorders resultant from being HIV positive may all diminish compliance with both retrovirals and psychotropics.

The hospital plays a unique role in meeting the clinical needs of persons who are HIV positive. Although the general hospital is the mainstay of treatment for this population, the state hospital is not infrequently called on to provide care and treatment to symptomatic and nonsymptomatic HIV-positive individuals.

Although HIV infection has been used as an example, a comparable presentation could be made for scores of medical disorders. Some, such as systemic lupus erythematosus, temporal lobe epilepsy, and space-occupying brain lesions, are dramatic examples. On a more frequent basis, hospital-based psychiatric and medical services draw together to treat such conditions as cancer, renal failure, and trauma.

Psychiatry has known for over 150 years that individuals with SMI die younger than persons without these disorders. And the reason for this is still unknown. It is known that medical comorbidities can be significantly exacerbated by psychiatric medications. This is no better illustrated than metabolic syndrome in persons taking some of the second-generation antipsychotic medications. It behooves the inpatient psychiatrist to work closely with his medical colleagues in the hospital and other staff like dieticians and occupational therapists to create and monitor wellness programs for his patients.

A service that has become standard in many state hospitals is hospice care. These services are provided to the state hospital patient who has spent some time in the hospital and without the patient needing to be transferred to another facility or even to change units within the hospital. The advantages to the patient are many, but the fundamental advantage is the patient's ability to remain in a familiar environment with staff he is comfortable with. It is important to have the patient's informed consent or the guardian's if a guardian of the person is in place, for all services that will not be provided and those that will. Hospice care in the hospital works closely with families and significant others throughout the patient's stay to process bereavement and loss.

Deaf Individuals

In treatment as orally based as is psychiatric treatment, the clinical approach to persons who are deaf presents a unique set of challenges. Although in many states, deaf individuals who need psychiatric treatment secure that treatment within the hospital alongside all others with the addition of translation services, some states have a statewide deaf unit at one of its hospitals. Here, persons who are deaf can interact with staff without the imposition of a translator and can be in an environment that is sensitive to the deaf culture. There are very few private hospitals that offer this service.

The range of the hospital's roles in treating deaf persons is no different than for hearing persons. To fulfill that role, hospital staff needs to be aware of and sensitive to the deaf culture. For this population, the state hospital-based deaf unit may be the least restrictive, most integrated setting when inpatient treatment is required. In states with deaf units that also require civil admissions only come from other hospitals, this restriction is waived for the deaf population who can be direct admits from multiple loci of care, including general hospital EDs.

Forensic Patients

Persons admitted to the hospital pursuant to the commitment of a crime are the states' responsibility and one that the state generally fulfills directly at state hospitals. There are a few exceptions when the state contracts out for these services.

Forensic patients are in the hospital basically for one of three reasons: (1) pretrial, for a determination of competency to stand trial and/or of criminal responsibility for the crime; (2) after a finding of being incompetent to stand trial, guilty but mentally ill, or not guilty by reason of insanity; or (3) jail or prison could not manage the inmate due to his behaviors as a result of psychopathology. Restoration of competency to stand trial is a unique forensic task, and the hospital needs to have specialty functions to accomplish this. Much of the rest of the clinical treatment of the forensic population is the same as that of the nonforensic population. Security on forensic units is generally much tighter than in other places at the hospital.

Often the role that the hospital is required to fulfill for many forensic patients is one that it does not play very often for other patients—housing forensic patients for years to decades beyond the time the patient needs hospital-level care. The degree to which this occurs is heavily dependent on the state commitment statutes and procedural handling of its forensic patients.

Although there is a clear dichotomy between forensic and nonforensic patients based on the state statute under which the patient was admitted, the clinical variation and treatment need to show no such distinction. Whether a patient is civil or forensic, especially for minor crimes, has more to do with the circumstances of the patient's initial encounter with the police than it does with their clinical presentation. The boundary is also blurred these days because of states admitting more persons at the end of a criminal sentence, especially those convicted of sexual offenses.

The hospital has multiple roles with the forensic patient: (1) to meet the person's needs as a psychiatric patient, (2) to meet the state's needs to proceed with due process in trials, and (3) to maintain the public's safety in the face of those who break or are charged with breaking state or federal law. Hospitals are considerably more comfortable with the first role than with the other two.

With forensic patients, the judicial or quasi-judicial systems control the admittance and discharge of patients. In many states, this contributes to both overcrowding at hospitals and to an ever-increasing percentage of forensic patients in the hospital's census. It is an irony of current times that jails and prisons feel overwhelmed by inmates sheriffs believe should be psychiatric inpatients, and psychiatric hospitals feel overwhelmed by inpatients psychiatrists assess to need jails and prisons.

OTHER HOSPITAL ROLES

This section has focused on the inpatient functions of the hospital, which comprise the service that is unique to it. The hospital fills a number of other roles, many of which are also performed in some fashion by entities that are not hospitals.

Emergency Services

Psychiatric treatment is provided on an emergency basis in hospital EDs (see Fig. 56.3–1). The arrangements under which this is done vary greatly. In community hospitals, the person presenting with a psychiatric problem is generally treated in the same fashion as all individuals—the same waiting area, the same examining rooms, sometimes the same physicians. In larger hospitals, all of these functions are often segregated into a designated area of the ED where patients are seen by nonphysician evaluators who consult with psychiatrists who may or may not see the patient. Like many other hospital functions, crisis evaluation services have moved out of the hospital and to freestanding crisis services in some jurisdictions. Not uncommon is a hybrid model in which the hospital ED plays host to an outside crisis service. Here, the hospital performs medical screening evaluations.

The ED functions may include psychiatric holding beds, that is, the ability to keep a person in the ED for up to 23 hours, thus avoiding an admission. The ED may include crisis beds, a service allowing a short number of days that again avoid admission to the psychiatric unit. Crisis beds are another service the hospital shares with nonhospital settings such as CMHCs, which offer freestanding crisis beds.

Psychiatric patients “stuck” in ED’s awaiting transfer to an inpatient bed have become a nexus for consternation over inadequate inpatient beds. These patients can wait hours, days, weeks, or in some cases over a month, during which time little is done for them other than room and board in the ED. In August 2014, the Supreme Court of Washington State found (*IN RE the Detention of DW et al.*) it unlawful, under Washington’s involuntary treatment act, to board patients in EDs just because the appropriate evaluation and treatment facilities are overcrowded.

Only about half the states have processes in place to systematically identify open beds for psychiatric patients who need admission. Even those states that can effectively search for bed availability are stuck with the inadequate number of psychiatric beds nationwide. A 2012 study by the Treatment Advocacy Center, “No Room at the Inn,” demonstrated that not one state had sufficient numbers of public hospital psychiatric beds.

Step Down/Admission Diversions Services

Hospitals offer a range of services that function between inpatient and outpatient levels of care. Such services include day hospital, partial hospital, and night hospital. In treatment arrangements that are during the day, the individual attends a full day of active psychiatric treatment but lives outside of the hospital, for example, in their home, with family, or in a community residential program. In the night hospital program, the person leaves the hospital for school or work during the day but spends overnights at the hospital. All of these programs are also used as attempts at admission diversion, that is, to avoid inpatient hospitalization for this episode of illness. Step-down programs and admission diversion programs are also offered by nonhospital entities, with the latter also offering respite beds. These have a longer length of stay than crisis beds and are not generally provided by hospitals.

Partial hospital programs, once enthusiastically endorsed by insurers as less expensive alternatives to inpatient treatment for their use as both alternatives to admission and step-downs to decrease the length of stay, have lost favor. There is a dwindling supply of such programs. Just between 2000 and 2001, the number of facility members of the National Association of Psychiatric Health Systems (NAPHS) offering partial programs decreased from 83% to 67%. Between 2003 and 2013, the number of hospitals offering partial programs decreased by 20% (from 508 to 410). There is expected to be further loss of partial programs if CMS decreases their reimbursement rate, as it is proposing to do.

Outpatient

Outpatient services are offered in more settings than any other service. Besides the hospital, outpatient services are found in CMHCs; group settings, either integrated with primary medical services or at their own sites; private practices; family services agencies; and hosted in locations such as shelters, specialty clinics, mental health residential programs, and mobile units. Payment for these services is most often by third-party insurers—private or government—with a copay by the patient. In larger metropolitan areas particularly, some outpatient providers have established practices that take no insurance.

The major challenge facing outpatient psychiatry at hospital locations is the high cost of delivering the treatment within hospital settings and the poor reimbursement rates. It has reached the point where expanding business in the outpatient department (OPD) is the last objective of the hospital because each new patient adds to the deficit.

On a more practical note, hospitals are seen as unfriendly places to the individual seeking outpatient psychiatric services. Parking is often difficult, far from the OPD, and expensive. Negotiating the corridors of the hospital is a challenge that many would prefer to avoid. In addition, the patient may have to go through a registration process that simply adds another layer of annoyance.

The prognosis for hospital-based outpatient services is not good. Often the victim of cherry-picking by others (the most difficult patients are directed to the hospital OPD), the hospital psychiatric OPD is on the hit list of many hospital chief executive officers. One remedy may be hospital-based integrated clinics where outpatient psychiatry and medicine are collocated. Such attempts have been around for 40 years, but the momentum for such integration is accelerating.

Consultation-Liaison

Although many hospital executives would gladly replace the psychiatric inpatient service with more profitable medical-surgical beds and make much better use of the space devoted to the psychiatry OPD (which is often banished from the main hospital buildings anyway), the C-L function is usually highly valued. In part, that is because the C-L service directly affects the hospitals' medical and surgical services by evaluating patient's capacity to consent, assessing cases of delirium and psychosis, and assisting with difficult dispositions. The C-L service can be invaluable in making differential diagnoses that will fundamentally alter treatment and disposition. Some have argued that it is because a hospital could not effectively function without C-L psychiatry that much of the rest of hospital-based psychiatry is tolerated. This may be a somewhat cynical overstatement, but it certainly has some elements of truth. A well-functioning C-L service is a fountain of goodwill for the psychiatry department.

Community Education and Patient/Family Support

One role of the hospital that is largely one of goodwill is the hospital's role as host to a remarkable array of self-help and support groups. Those that fall under the penumbra of mental health and substance abuse activities are frequent in hospitals; such groups are Alcoholics Anonymous, Narcotics Anonymous, and the National Alliance on Mental Illness. Less directly related to mental illness but using principles of support are groups centered on all manner of chronic conditions and issues, such as breast cancer, colostomy, and scleroderma. There are also groups for individuals in specific life circumstances, such as parents without partners and parents who lost a child to cancer.

Hospitals also play a community education role. Members of the psychiatry staff may provide substantive lectures geared to the lay public. These are different yet complement the hospital's psychiatry grand rounds, which themselves are usually open to the public. Another form of education is informing the public about the services the hospital has to offer.

FUTURE DIRECTIONS

The challenges that face hospitals in the second decade of the 21st century are not new. Prior to 1950, virtually every problem and many of the solutions that define contemporary hospital psychiatry were noted in the professional literature. One need only read such sources as the *American Journal of Insanity*, *Alienist and Neurologist*, or state hospital annual reports to find them all.

Homelessness, the overpopulation of jails and prisons with persons with mental illness, the eschewal of persons with mental illness from any locale because that jurisdiction does not want financial responsibility for the person, an inadequate focus on treatment and overreliance on management, and a propensity to not spend tax money on persons with mental illness are all conditions Dorothea Dix found in the 1840s when she began her lifelong work of reform. Samuel Woodward, APA's first president and the first superintendent of what is now the Worcester Recovery Center and Hospital (see Figs. 56.3–4 and 56.3–5), would have recognized them all.

The loss of the person in meeting the acute needs of the patient, the reliance on modern technologies for diagnoses that have diminished one-to-one time with patients, and the general sense of the dehumanization of psychiatry and medicine, in general, were noted by William Allison White in 1925. White remarked in his APA presidential address that "the introduction of these highly technical methods has been accomplished by a corresponding lack in development, at least, if not actual loss, of any adequate dealing with the patient as an individual or as a human personality."

Even hospitals bending to the dictates of payers to minimize costs is not new. More than 125 years ago, Kirkbride lamented that "it will be a sad day for these institutions, and still sadder for the patients in them, when the rivalry of hospitals and their offices shall be, rather to discover for how little inmates can be kept, than to secure what is best, and most thoroughly promotes the great objects for which they were established."

State hospitals are disappearing by design. Whether that is the result of thoughtful or ill-conceived planning is a matter of debate. General hospital psychiatry programs, however, are disappearing through a thoughtless process of survival of the fittest.

Reimbursement and operating margins for psychiatric services lag behind those of other hospital-based medical services. The diminishing financial returns from psychiatry have led hospital management consultants to recommend converting psychiatric beds to higher-margin medical and surgical beds. Even to remain fiscally viable now, psychiatric hospital inpatient revenues need to cross-subsidize emergency, consultation, and outpatient services.

Today, psychiatrists generally fail to acknowledge those times when the psychiatric hospital is a forerunner or role model for the rest of psychiatry, if not the rest of medicine. One of the foci of contemporary medicine is the integration of medicine and psychiatry or medical and mental health services. As practitioners in psychiatry and in medicine focus on methodologies to integrate care and treatment, they need look no further than the public psychiatric hospital, which has been an integrated care organization for over 100 years and still is a fine example of integrated care and treatment. Outpatient integration, it seems, is mimicking business as usual at the state hospital but without its walls.

Attributes of a state hospital that contribute to its high degree of integration include:

- Psychiatric and medical services collocated
- An integrated psychiatric and medical treatment plan
- A single shared medical record, now an electronic medical record in many institutions
- Ease of communication among a patient’s physicians and between those physicians and other disciplines
- Case or care management integrated into treatment
- Professional and paraprofessional staff assisting patients to follow through with psychiatric and medical recommendations in an integrated housing and care model
- On-site wellness programs, for example, dietary, exercise, organized leisure
- Other health promotion activities, for example, weight loss, smoking cessation, treatment of substance abuse
- Transportation to any off-site medical appointments
- Workforce development, for example, student placements, pre- and postdoctoral internships, residency rotations
- Paid time off for staff education
- On-site dental, podiatry, optometry, radiology, phlebotomy, and neuropsychiatry services
- Peer supports
- Referrals to community and social supports
- Services for all independent of the ability to pay

- Funding supports not only direct service but time for collaboration
- Unitary billing system
- Transition services for relocation provided and funded
- Interfaces with vocational services
- Provision of in vivo training, for example, bus riding, banking, using the library
- On-site hospice care with direct care staff the patient is already familiar with.

The state psychiatric hospital, as an integrated care system, should have outcome measures that reflect these attributes. The current, common outcome measures fail to take this into account and currently include the number of assaults patient on patient and patient on staff; the number of serious assaults; the number of episodes and time in any form of seclusion or restraint (manual, mechanical, chemical); two or more antipsychotic medications at discharge; five or more psychiatric medications prescribed; time out of the hospital on unsupervised passes; and timeliness of all manner of documentation. Many of these measures are open to interpretation, especially what a facility chooses to label as an episode of restraint. What about improvements in health and wellness as outcome measures?

Virtually no knowledgeable commentator claims that the system of care and treatment for mental illnesses can be sustained without the roles played by psychiatric hospitals. At the same time, too few informed people are working on methods to interdigitate the hospital's roles into the fabric of an integrated system of care, one that looks less like Figure 56.3–1 and more like one that achieves effective outcomes, through a person-centered approach, with fiscal management that does not compromise the very care and treatment psychiatrists are attempting to deliver.

FURTHER READINGS

Andrews JB. The distribution and care of the insane in the United States. *Am J Insanity*. 1887;44:192–211.

Butler JS. The individualized treatment of the insane. *Alienist Neurol*. 1886;7:435.

Center for Substance Abuse Treatment. *Screening Assessment, and Treatment Planning for Persons With Co-Occurring Disorders. Overview Paper 2 (DHHS Publication No. (SMA) 07–4164)*. Substance Abuse and Mental Health Services Administration, and Center for Mental Health Services; 2007.

Committee on Crossing the Quality Chasm. *Adaptation to Mental Health and Addictive Disorders: Improving the Quality of Health Care for Mental and Substance Use Conditions*. National Academies Press; 2006.

Deutsch A. *The Mentally Ill in America*. Columbia University Press; 1949.

Dorwart RA, Schlesinger M. Privatization of psychiatric services. *Am J Psychiatry*. 1988;145:543–553.

Fisher WH, Geller JL, Pandiani JA. The changing role of the state psychiatric hospital. *Health Aff (Millwood)*. 2009;28(3):676–684.

Geller JL. Issues in American psychiatry reflected in remarks of APA presidents, 1844–1994. *Hosp Commun Psychiatry*. 1994;45:993–1004.

Geller JL. Excluding institutions for mental diseases from federal reimbursement for services: strategy or tragedy. *Psychiatr Serv*. 2000;51:1397–1403.

Geller JL. The last half-century of psychiatric services as reflected in *Psychiatric Services*. *Psychiatr Serv*. 2000;51:41–67.

Geller JL. A history of private psychiatric hospitals in the USA: from start to almost finished. *Psychiatr Q*. 2006;77:1–41.

Geller JL, Biebel K. The premature demise of public child and adolescent inpatient psychiatric beds. Part I: overview and current conditions. *Psychiatr Q*. 2006;77:251–271.

Geller JL, Harris M. *Women of the Asylum*. Doubleday; 1994.

Greenblatt M, York RH, Brown EL. *From Custodial to Therapeutic Care in Mental Hospitals*. Russell Sage Foundation; 1955.

Grob GN. *Mental Illness and American Society, 1875–1990*. University Press; 1983.

Haskins BG. Servicing deaf adult psychiatric inpatients. *Psychiatr Serv*. 2004;55:439–441.

Heisler, EJ, Panangala, SV. The Veterans Health Administration and Medical Education: In *Brief*. Congressional Research Service Report; 2018.

Honberg R, Diehl S, Kimball A, Gruttadaro D, Fitzpatrick M. *State Mental Health Cuts: A National Crisis*. National Alliance on Mental Illness; 2011.

Hurd HM, ed. *The Institutional Care of the Insane in the United States and Canada*. Johns Hopkins University Press; 1916.

Liptzin B, Gottlieb GL, Summergrad P. The future of psychiatric services in general hospitals. *Am J Psychiatry*. 2007;164:1468–1472.

McLoughlin KA, Geller JL. Interdisciplinary treatment planning in inpatient settings: from myth to model. *Psychiatr Q*. 2010;81:263–277.

Menninger RC, Nemiah JC, eds. *American Psychiatry After World War II, 1944–1994*. American Psychiatric Press; 2000.

National Institute of Mental Health. Transforming the understanding and treatment of mental illnesses. http://www.nimh.nih.gov/health/statistics/prevalence/use-of-mental-health-services-and-treatment-among-adults.shtml. Accessed December 21, 2014.

Parks J, Radke AQ, eds. *The Vital Role of State Psychiatric Hospitals*. National Association of State Mental Health Program Directors; 2014.

Report of the Surgeon General's Conference on Children's Mental Health: A National Action Agenda. Department of Health and Human Services; 2000.

Sharfstein SS, Dickerson FB. Hospital psychiatry for the twenty-first century. *Health Aff (Millwood)*. 2009;28:685–688.

Sledge WH, Tebes J, Wolff N, Helminiak TW. Day hospital/crisis respite care versus inpatient care. Part II: service utilization and costs. *Am J Psychiatry*. 1996;153:1074–1083.

Stevens R. Can the government govern? Lessons from the formation of the veteran's administration. *J Health Politics Policy Law*. 1991;16:281–305.

Torrey EF, Fuller DA, Geller J, Jacobs C, Ragosta K. *No Room at the Inn: Trends and Consequences of Closing Public Psychiatric Hospitals*. Treatment Advocacy Center; 2012.

Yazgan IC, Greenwald BS, Kremen NJ, Strach J, Kramer-Ginsberg E. Geriatric psychiatry versus general psychiatry inpatient treatment of the elderly. *Am J Psychiatry*. 2004;161:352–355.

56.4 Mental Health Services Research

Lisa B. Dixon, M.D., M.P.H., and Milton Wainberg, M.D.

The imperative for mental health services research arises from some basic and alarming facts. Over half of the persons with mental disorders receive no mental health treatment. For those who do, the vast majority—approximately 75%—do not receive treatment that can be judged of good quality by evidence-based standards. Those outside of the health services research field, such as clinicians and other stakeholders, may not see how mental health services research can improve the substandard quality of mental health care. In fact, they may not even understand what health services research is. The Association for Health Services Research broadly defines it as:

> Implementation Science is the multidisciplinary field of scientific investigation that studies how social factors, financing systems, organizational structures and processes, health technologies, and personal behaviors affect access to health care, the quality and cost of health care, and ultimately our health and well-being. Its research domains are individuals, families, organizations, institutions, communities, and populations.

This chapter reviews the specific scope of mental health services research and illustrates through several case studies how such research addresses critical questions that arise in the routine provision of services, laying the groundwork for improving the quality of mental health care.

Notably, implementation science or the study of strategies and approaches to promote the use of evidence-based practices has become a key element of mental health service research.

THE SCOPE OF MENTAL HEALTH SERVICES RESEARCH

Clinicians seek to address identified clinical problems, disorders, and diseases. Clinicians turn to biomedical research to answer two basic questions: (1) What is known about the nature of this problem? (2) What can be done to remedy it or alleviate the suffering associated with it? Research answers to the first question derive from *basic sciences research*, whether biologic (e.g., What neural pathways are affected in bipolar disorder?), behavioral (e.g., How do organisms respond to different types of reinforcement schedules?), psychological (e.g., How do negative automatic thoughts affect mood?), or sociologic (e.g., What are the pathways to homelessness?). Research answers to the second basic question arise from *clinical research* involving well-controlled treatment trials designed to determine efficacy and safety, as well as from the more or less systematic trial and error of clinical experience. Broadly, the basic and clinical sciences deal with the knowledge needed prior to the initiation of routine care. In other words: What is the health problem, and what treatments might relieve this problem? *Health services research* questions, however, arise once clinical services get underway. Clinicians have extensive exposure to the basic and clinical sciences during their training but may have little or no exposure to, and hence little understanding of or appreciation for, the value of health services research and implementation science.

What kinds of questions do mental health services researchers ask? These questions can be divided into two general categories—*clinical services research* and *service systems research*. Like clinical research, *clinical services research* focuses on “front-line” clinical issues and uses individual service recipients (patients, families) as the participants and units of study. Examples of questions asked by *mental health clinical services research* include the following: How effective are antidepressant agents in producing illness remission among typical patients with major depression seen in routine office practices? What considerations influence individuals to seek treatment and to agree to complete a course of prescribed treatment? What range of treatments for persons with schizophrenia are “evidence-based?” Do persons with substance abuse and mental illness have better outcomes when they receive care from integrated addiction mental health teams or programs with parallel services? What are the effects of a routine depression screening measure in primary care on detecting and treating major depression in such settings?

Mental health services system research examines questions at the service system level and typically uses aggregated service components (providers, treatment agencies, state mental health systems, etc.) as the units of study. Examples of questions asked by service systems researchers include the following: How many persons with severe mental illness who are being treated in a public mental health system have access to routine primary care? Does this differ between two states with different Medicaid policies? What are the various sites in a city where homeless persons with severe mental illnesses receive services and are these coordinated? What are the effects of a decision by Medicaid to cap the total number of prescribed medications that recipients can receive? Do restricted formularies for antipsychotic medications reduce overall costs? Do they affect quality? Do local mental health authorities improve coordination, continuity, and quality of care for high-need patients?

Implementation science research typically integrates multiple levels to ask how to best implement evidence-based practice.

Given this broad scope of mental health services research, it is also helpful to note what it does not include (Fig. 56.4–1). The distinction has already been mentioned between basic and clinical research on one hand and mental health services research on the other. The interface between clinical research and clinical services research begins at the transition between treatment efficacy and treatment effectiveness studies. *Efficacy studies* examine the outcomes of treatments within narrowly defined samples of patients who are carefully selected to control for variables that might interfere with determining a treatment effect (e.g., the exclusion of patients with comorbid substance abuse or medical disorders). Efficacy study designs attempt to maximize internal validity. *Effectiveness studies* examine treatment outcomes among more inclusive, diverse, and “typical” clinical populations. External validity assumes a more important role. Generally, effectiveness studies only occur after treatment efficacy has been established. Mental health services research, while somewhat distanced from the efficacy studies with which many clinicians are familiar, is still *research*, using standardized scientific procedures to answer specific research questions to generate new knowledge that is generalizable beyond the system being evaluated. It should be distinguished methodologically from routine activities undertaken by service systems and service agencies to monitor policy implementation, program or system administration, or quality assurance.

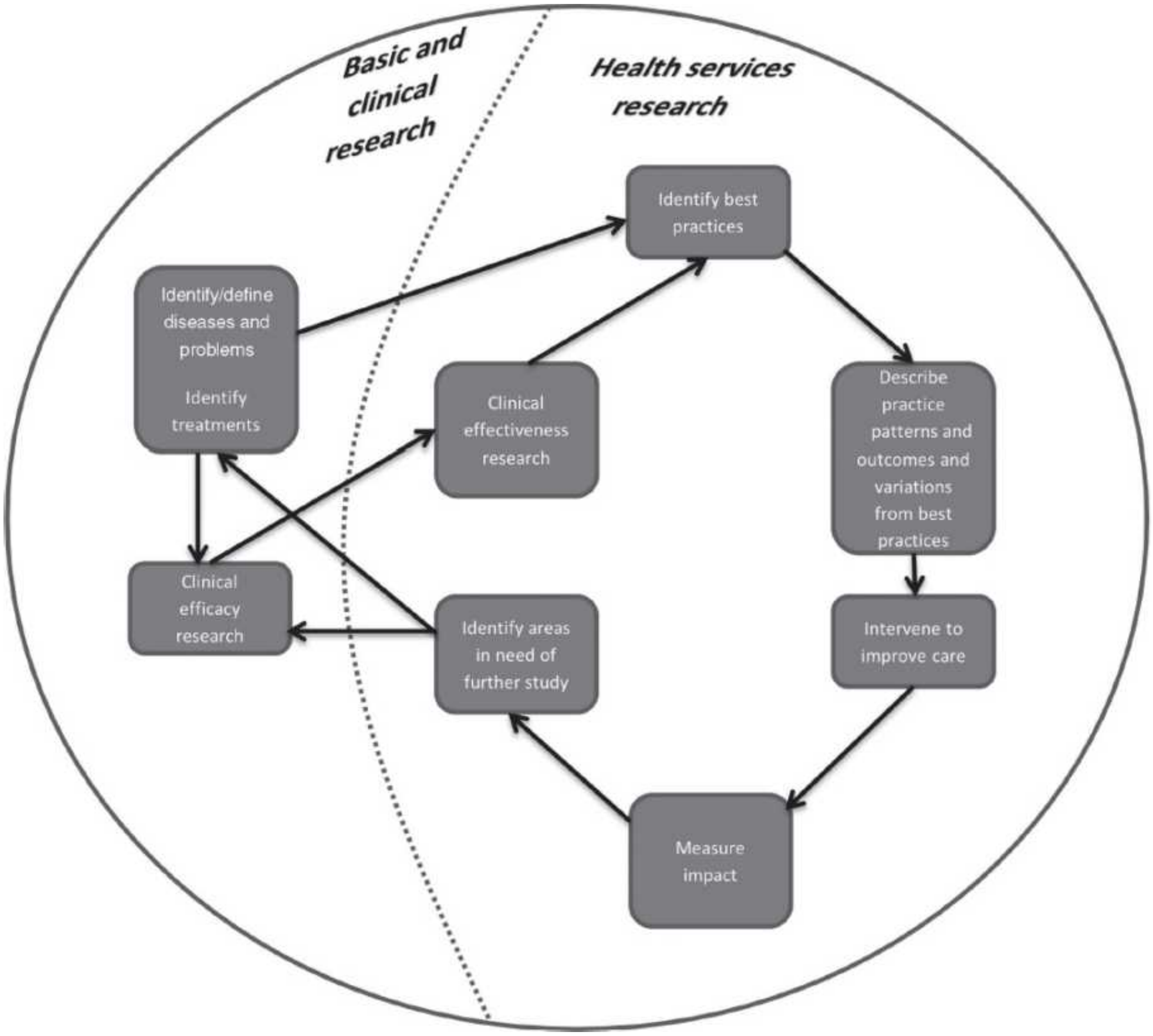
Basic and clinical research
Health services research
Identify best practices
Identify/define diseases and problems
Identify treatments
Clinical effectiveness research
Describe practice patterns and outcomes and variations from best practices
Clinical efficacy research
Identify areas in need of further study
Intervene to improve care
Measure impact

FIGURE 56.4–1. The relationship between basic and clinical research and health services research.

Figure 56.4–1 illustrates the distinction between basic and clinical research and health services research, the flow between these larger research domains, and depicts various activities conducted within health services research. For example, one way in which clinical practice is affected by services research occurs as the results of clinical efficacy trials and effectiveness trials are applied to routine practice—a form of "translational research." Given that a primary goal of basic and clinical research is to inform clinical practice when a body of scientific knowledge has accumulated, it should be adopted into clinical practice. However, this translation does not happen as it should; the gaps between existing scientific knowledge about treatment efficacy and effectiveness on one hand and routine clinical practice on the other have been referred to collectively as the "quality chasm" in medical care encompassing both access and quality.

A major question for mental health services research, then, is how to promote the adoption of scientifically validated treatments (evidence-based practices) and, by doing so, bridge the "quality chasm." One response has been the development of evidence-based practice guidelines, variously referred to as evidence-based practices, evidence-based treatment recommendations, or "best practices." The identification of evidence-based best practices is critical to examining current practice patterns, evaluating their quality, identifying areas in need of improvement, and ultimately identifying areas in which current treatments fall short. The methodology for developing such evidence-based practice guidelines emphasizes the results of efficacy and effectiveness research, augmented by expert opinion when research evidence is underdeveloped in essential areas.

Once a set of research-based best practices has been identified, health services researchers can ask a variety of questions about how current practices compare to best practices, including variations in practices across treatment settings, practitioners, and patient subtypes, as well as the relationship to such factors as insurance coverage, health policies and regulations, and cultural diversity. Such research can also examine patterns in outcomes of care across different kinds of services and relate these to patterns in the implementation of evidence-based practices. Such studies are essentially descriptive, but they allow comparisons of actual practices to standards of care.

Identifying gaps between actual and best practices can lead to developing health services interventions to improve care. Such interventions fall into one of two general categories. *Supply-side* interventions focus on individual providers or service programs. These interventions seek to change the care offered in some way. Some examples of supply-side interventions include provider education and training programs, systematic feedback to providers regarding their current practices, consultations to agencies about the implementation of best practices, financial incentives to alter practices, and so on. *Demand-side* interventions focus on the demand for services and are directed at changing patient behavior. Examples of such interventions include educating patients about best practices, training patients in how to advocate for themselves in health settings, and providing service vouchers to patients to enable them to select services that they identify as most helpful.

Both supply- and demand-side interventions can occur at distinct levels involving individual providers and patients, local service agencies, local or regional service systems, and local, regional, or national policies. *Shared decision-making* is a strategy that involves both supply-side and demand-side approaches at the level of the individual patient–provider interaction. *Policy-level interventions* occur at the level of a system of care with the intention of changing patterns of service delivery—for example, changing the services covered by insurance, creating local mental health authorities to improve the coordination and outcomes of treatment in a given locale, mandating the integration of mental health and substance abuse services, or offering Medicaid care through a coordinated and managed “health home.”

Such interventions, however, are not in themselves health services research. Rather, health services research evaluates the effects of such interventions on a range of outcomes, including patient outcomes, changes in patterns of care, and costs. In health services research, the interventions must be clearly defined, the fidelity of their implementation closely monitored, and the outcomes measured in a standardized form. Research designs to minimize threats to internal and external validity—for example, randomized clinical trials—are critical, as in clinical research. The research questions, although based on specific interventions, focus on knowledge that can be generalized to other settings. As in other areas of research, findings from any phase of health services research optimally feedback to raise new research questions.

Recent work has focused on methodologies for implementation research, a burgeoning domain of health services research that attempts to understand how effective interventions can be implemented. Studies of implementation problems in carrying out interventions at the level of systems or individual clinical care depend upon various conceptual frameworks but generally acknowledge the need to study factors including (1) the structural, political, and cultural context as well as internal organization of the settings in which the intervention will take place, (2) characteristics of the individuals involved, (3) the process of implementation of the intervention, and (4) the nature of the intervention itself. Such research takes a pragmatic approach, seeking to understand how real-world conditions affect the implementation of the evidence-based practices for which clinical and other research has established evidence of efficacy and effectiveness.

The following “case reports” concretely illustrate the application of mental health services research to critical questions that arise in the routine provision of services.

Case 1: Treatment of Depression in Primary Care Settings

Depression is one of the most common mental disorders and accounts for a staggering amount of disability worldwide. Over the last several years, there have been substantial improvements in methods for diagnosing depression, and a broad range of efficacious treatments for depression exist, including antidepressant medications and evidence-based forms of psychotherapy. A number of evidence-based practice guidelines have been developed, so the opportunities to effectively diagnose and treat depression are substantial.

Given the state of diagnostic and treatment methods for depression, how successful is the United States in treating depression? This health services research question raises some basic epidemiologic queries. How many people suffer from depression? Do they seek treatment, and where? What treatments do they receive? The most comprehensive estimates of the prevalence of mental disorders derive from the Epidemiologic Catchment Area (ECA) Study and the National Comorbidity Survey (NCS), which estimated the 1-year prevalence of major depressive disorders at 6% to 7%. The vast majority of depressed individuals (about 80%) contact a health care provider over a 1-year period, but only about 25% receive any appropriate treatment for depression. Most patients (about 80%) who visit a health care provider for depression see a primary care provider; only about 20% of patients with depression visiting a health care provider see a mental health specialist. Among such patients with a treatment contact for depression, 80% of those treated in primary care received no appropriate treatment for depression (pharmacotherapy or counseling). The opposite prevailed in mental health treatment settings, where most patients (nearly 90%) received an appropriate form of depression treatment. Such data illustrate the challenges in providing quality care for individuals with depression. Most persons receive no depression treatment (despite some contact with a health care provider). Most of those who do receive some treatment receive inadequate care. The predominant opportunity for the treatment of depression is in primary care. How can clinicians and administrators address these problems in access to and quality of depression care? How can mental health services research assist in these efforts?

Health services researchers have examined interventions targeting providers (the supply-side of care), individuals affected by depression (the demand-side), and policies. By far, the majority of health services research in this area has focused on the supply-side through attempts to change provider management of depression, especially primary care providers, who see most of the patients with depression. Such research has approached the under-treatment of depression in primary care with different interventions of interest. For example, a multicomponent health care system–level intervention called “collaborative care” structures depression treatment in primary care settings by using case managers who educate and follow patients, track outcomes, and adjust treatments for patients who are not responding, consultation and decision support for primary care providers by mental health specialists (psychiatrists or psychologists), and the use of technology-based resources to support the system. Numerous randomized controlled trials on collaborative care for depression have demonstrated meaningful effects on depression symptom improvement, adherence, response to treatment, remission, and recovery from depression, as well as satisfaction with care. However, only a small effect has been established on patient quality of life, including functional status.

A second health services research approach has studied the implementation of quality improvement (QI) programs for managed primary care settings with the goal of increasing the percentage of clinically depressed patients receiving appropriate depression care by national guidelines. The QI intervention consists of four components: institutional commitment of in-kind resources, training of local clinical leaders to implement interventions (cognitive behavior therapy, pharmacotherapy), training of local treatment staff, and patient identification. In one 12-month study, depressed patients in QI intervention clinics were more likely to receive counseling or medications, visit a mental health specialist, have a remission from the depressive episode, and retain employment than depressed patients in “control” clinics. Researchers have evaluated another approach in “behavioral health labs” that improves screening and detection of depression in primary care settings. All three of the above depression treatment models demonstrate the potential for improving depression care in primary care settings through “supply-side” interventions. Major challenges remain, however, in promoting the widespread adoption of these approaches.

Far less research on depression care has been conducted on demand-side interventions with patients, although one can certainly argue that even the predominantly supply-side interventions just reviewed increase demand by educating patients about their symptoms of depression and about the value of treatment. Negative attitudes about depression and its treatment among the general population can deter persons from seeking help. Such negative attitudes arise from personal health beliefs, social norms, and stigma, as well as from prior negative experiences with help-seeking. True demand-side interventions are aimed at directly increasing the likelihood that individuals with depression will seek and participate in effective treatment. Such demand-side interventions focus on awareness of the signs and symptoms of depression, self-assessment for these signs and symptoms, the effectiveness of treatment, and skills in seeking help. Some common examples of such efforts include National Depression Screening Day campaigns, free depression screenings at employee health fairs, free pamphlets in physician offices and clinics, antistigma campaigns through the media, and direct-marketing media campaigns on depression, such as those currently put forward by pharmaceutical companies. Health services research that examines the effects of such programs remains sparse, and little is known about the effects of these types of demand-side interventions.

A third domain of intervention to improve detection and treatment for depression in primary care is that of policy, system organization, and financing approaches. A prime example of this type of intervention is the current effort to promote health insurance parity for the treatment of mental disorders. Inadequate insurance coverage for the treatment of a mental disorder poses major barriers to treatment. The federal Affordable

Care Act (ACA) enacted in 2010 expands coverage of mental health and substance use disorders by including benefits for such disorders in the Essential Health Benefits required of all new small group and individual market insurance plans and by applying federal parity protections more broadly, requiring insurance plans to cover treatment for mental health and substance use disorders at parity with medical and surgical benefits; a projected 62.5 million individuals will benefit by access to these federal parity protections. In addition, the ACA provides states a new option for organizing care in their Medicaid programs: “health home” services for Medicaid beneficiaries with severe chronic conditions who can benefit from enhanced coordination and management of physical and behavioral health care as well as long-term community-based services and supports. The “health home” model aims to improve the quality of health care and the patient experience while providing more cost-effective care. Health services research on service organization and financing seeks to answer critical questions about the effects of policy interventions such as parity laws and “health homes” on issues including the quality and experience of behavioral health care—such as depression care—delivered through reorganized systems.

Case 2: Services for Families of Persons with Schizophrenia

During the second half of the 20th century, clinical research on families in the treatment of individuals with schizophrenia dispelled early assertions that disturbed family relationships play a major role in the etiology of the disorder. Instead, research led to the recognition that families provide major support networks for most persons with this disorder and that they need education and support to enable them to engage as partners in care. The potential advantages of engaging and educating families optimally were suggested by early studies linking high levels of “expressed emotion” among families to a greater likelihood of patient relapse after hospital discharge. Research spanning from the 1970s to the 1990s clearly established family psychoeducation (FPE) as an evidence-based practice in the care of individuals with schizophrenia. The Schizophrenia Patient Outcomes Research Team included FPE as a treatment recommendation, stipulating that patients who have ongoing contact with their families should be offered a family psychosocial intervention that lasts at least 9 months and includes education, support, and some training in problem-solving skills. Other statements of best practices also include a recommendation for FPE. FPE was found to reduce relapse rates, improve medication adherence, reduce symptoms, and improve functional outcomes for individuals with schizophrenia. Specifically, more than 30 randomized clinical trials showed that relapse rates for individuals receiving FPE were 25% to 50% lower than those observed for individuals receiving standard individual services.

In line with the health services research framework (see Fig. 56.4–1), investigators have examined the prevailing patterns of practices as they relate to families with regard to both access to any services and, more specifically, access to evidence-based practices consistent with the FPE model. A variety of studies conducted in community settings and in Veterans Administration facilities has shown that not only is FPE rarely offered, but in fact, a far lower standard of family contact with clinicians occurs in the majority of cases. Chart reviews, clinician interviews, and consumer interviews all have produced the same conclusions. Thus, major problems in the access to and quality of family-oriented services exist.

Following the health services quality-of-care framework, researchers have undertaken efforts to improve practices by disseminating FPE. Supply-side approaches have been the most common. Numerous initiatives to disseminate FPE have been attempted throughout the country. The New York State Office of Mental Health (OMH) mounted an impressive effort to implement programs at 35 different sites throughout the state and provided extensive training. Although numerous family groups were started and the potential benefit was large for those families and patients who participated, OMH decided not to continue the project because only a small number of families and patients throughout the state actually received the program. The Substance Abuse and Mental Health Services Administration included FPE as part of an effort to disseminate evidence-based practices by creating toolkits that include informational resources for all stakeholders, a workbook for practitioners, and a fidelity scale. The implementation phase of this project resulted in a limited selection of FPE as a practice to disseminate and limited penetration of FPE in those sites that selected FPE.

The marginal success of these dissemination initiatives for this evidence-based practice demands a greater understanding of how to influence changes in practice. Research must strive to understand and overcome observed roadblocks to the dissemination of family services, whether these roadblocks exist on the supply-side among providers and service systems or on the demand-side among patients and their families. Such roadblocks can take many forms. In identifying such barriers, health services researchers seek critical points for intervention to improve dissemination. For example, practitioners may be unaware of the value of the intervention, or they may lack the skills to deliver it. Families may be suspicious of such services as a result of unpleasant experiences based on older family therapy models. Payment systems may not provide support for the programs. However, even the FPE demonstration projects that had the capacity to overcome most or all of these barriers failed substantially to affect dissemination. In such a case, health services researchers have at least two other strategies at their disposal. The first is to develop more inventive dissemination strategies. The second is to decide that a modified or new evidence-based practice is needed—one that is more acceptable to providers, patients, and service systems.

This first task of developing more inventive dissemination strategies requires reevaluating the barriers to dissemination. Such barriers exist at the disparate levels of patient and family, providers, administrators, and payers. Supply-side strategies, such as providing higher rates of reimbursement for providing FPE, as well as more intensive and targeted training for providers, could be successful. Health services researchers investigate these strategies by using qualitative methods to interview providers and payers where FPE initiatives have already been started to determine whether they think that increased rates of payment or training would help. Strategies under study also target barriers to FPE at the consumer and family levels, thus focusing on developing the demand for FPE. A new approach called Recovery-Oriented Decisions for Relatives' Support (REORDER) uses a shared decision-making process to structure patients' exploration of whether they want to include families in their mental health care. REORDER is personalized by attending to patient preferences for family involvement, focusing on the mental health recovery goals of the patient, and exploring how family members could join with the patient to meet those goals. The REORDER clinician begins by working with the consumer to identify the pros and cons of family involvement and then uses education, support, and decision-making exercises to assist and empower patients to meet their goals. A randomized trial showed that this approach markedly increased rates of family contact with clinicians while it also increased a measure of recovery among participants assigned to the REORDER condition.

The second strategy to address difficulties in implementing FPE is to develop new evidence-based practices or to test revised models of FPE that attempt to reduce barriers to implementation. For example, a meeting of family researchers and practitioners identified the need to determine the critical ingredients of FPE so that a simpler and briefer model could be tested. The National Alliance on Mental Illness (NAMI) Family to Family Education Program may provide a second approach. This program is not a clinical service but is peer-driven. Trained family-member volunteers teach this manualized 12-week program in weekly 2- to 3-hour sessions. The program provides information about mental illnesses, treatments, medication, and rehabilitation. Participants also learn self-care, mutual assistance and communication skills, problem-solving strategies, and advocacy and develop emotional insight into their responses to mental illness. Although the program does not aim to reduce patient relapse, it may provide some benefits to families. Notably, the program had been widely disseminated in spite of the dearth of systematic research demonstrating its effectiveness. After its dissemination, a partnership between health services researchers and NAMI conducted three sequential studies, including a randomized trial have provided evidence that the program is helpful to families—in improving family functioning, coping, empowerment, and knowledge. This demonstrates that in the pursuit of dissemination of evidence-based practices, health services researchers may tackle the challenge of conducting research on existing practices that are in wide use but as yet relatively unstudied.

Case 3: Addressing the Burden of Mental Illness and Mental Health Disparities Using Technology and Implementation Science

Burden and Treatment Gap. Since the 1990s, we have been aware that mental and substance use disorders (henceforth, mental health and/or disorders) are the largest contributor to the global burden of disease everywhere. This problem will persist as long as a minority of individuals with mental and substance use disorders never receive treatment. This treatment gap is vast and global. In high-income countries, 35% to 50% of people in need of mental health care do not receive care. Despite being home to more than 80% of the world’s population, low- and middle-income countries have less than 20% of the world’s mental health resources. On average, only 0.5% of national health care budgets in low- and middle-income countries (vs. 5.1% in high-income countries) are devoted to mental health care, and 75% to 85% of those with mental disorders in low- and middle-income countries have no access to care. Mental health disparities exist everywhere, and the majority of those in need of mental health care worldwide lack access to high-quality mental health services. Stigma, structural discrimination, human resource shortages, fragmented service delivery models, and lack of research capacity for implementation and policy change contribute to the current mental health treatment gap.

Shortage of Humans Resources/Task shifting. One of the most significant challenges in decreasing the mental health treatment gap is the significant scarcity of trained mental health workers in health care systems worldwide. The shortage of certified mental health specialists is vast and global, especially those who are skilled in using evidence-based practices. "Task-sharing or task-shifting"—the process of delegation whereby clinical tasks are moved from a highly specialized workforce to less specialized health workers, including trained lay health workers—is an effective strategy to deal with the mental health personnel shortage and is a great option to enhance efficiency on delivery of health care services. Multiple studies throughout the world have proven that lay health workers can effectively provide evidence-based treatment for common mental disorders. Task sharing requires a significant, ongoing role for psychiatrists and psychologists to properly train, continually supervise, ensure quality assurance, and support for the nonspecialist health workers.

COVID. Globally, the already limited access to mental health services has worsened as health care systems prioritize COVID-19 care. People with mental disorders are highly vulnerable to Coronavirus 2 (SARS-CoV-2) infection and vice versa. History of psychiatric illness is a risk factor for COVID-19 infection, hospitalization, and death. COVID-19 survivors have elevated rates of depression, anxiety, psychosis, PTSD, mania, and sleep disorders. It has exacerbated the structural, social, economic, health, and mental health vulnerabilities of disadvantaged populations both in the United States and globally (e.g., racial/ethnic minorities, migrants, and displaced people, low socioeconomic background), placing them at elevated risk for COVID-19 and its consequences, including increased mental health service needs.

Telehealth became necessary to protect providers and patients against SARS-CoV-2 infection and maintain ongoing care for patients already receiving care or new patients. However, health care systems struggle to make these transitions in providing care. Developing remote-based methods to triage and provide mental health services became crucial. Importantly, wherever this became available via phone or internet, patients previously in care with access to resources and digital infrastructure were more easily adherent to care, as some of the limitations to access to care could be addressed (e.g., travel, childcare, work, school). Unfortunately, the mental health services access for new patients or those without resources was negatively affected. The COVID-19 pandemic calls for a new model of mental health service delivery in a system that was already incompetent to provide mental health care.

Comorbidity among psychiatric diagnoses is the rule rather than the exception. It produces greater functional impairment and worsens social determinants. Both psychopharmacology and/or evidence-based psychotherapy trials have shown efficacy and effectiveness in high-, middle-, and low-income countries, suggesting that both modalities have promise in expanding treatment for mental disorders where services are scarce. Yet, to date, studies integrating mental health in primary care (e.g., Collaborative care) address one or two disorders through individual-level interventions and show low enrollment of racial/ethnic minorities and underserved populations. Decreasing mental health disparities necessitates sustainably implementing comprehensive mental health services for multiple disorders through community-health systems-policy makers partnerships, which in turn can improve functioning and social determinants and inform new policies.

Technology. Leveraging novel technology to support the implementation and supervision of evidence-based treatments can help meet mental health needs. Digital technology in the mental health sector has proliferated globally. Digital patient and provider-facing solutions can give unprecedented access to high-quality mental health care and can effect scalable, cost-saving, reliable, and sustainable, comprehensive mental health services delivery, especially in low-resource settings, yet they are seldom used with disadvantaged communities and public health systems. Digital technology can improve overall cost-effectiveness by strengthening the mental health capacity of existing nonspecialists and supporting task-shifted delivery of high-fidelity evidence-based treatments.

Closing the Gap—Implementation Science. The gap between mental health evidence-based treatment research and practice in real-world communities and health systems is vast; less than half of clinical innovations make it into routine practice. Reasons include the fact that researchers tend to disseminate their work through traditional academic conferences and publications. Further, the research settings where researchers develop and test evidence-based treatments do not reflect the complex, less-resourced real-world settings in which they ultimately must be delivered to have a population impact. Dissemination and implementation science address this research-to-practice gap aiming to accelerate the speed with which translation and real-world uptake and impact of testing, evidence-based treatments occur and improve public health. Meeting the population's mental health needs requires employing implementation science methods to integrate evidence-based treatments into routine practice. Importantly, solutions implemented in low- and middle-income countries can offer innovative options that might be applied to improve mental health services across diverse settings worldwide, including high-income countries.

Two ongoing NIMH-funded studies, one in a low-income country, Mozambique, and another in New York City, address these challenges and needs using a stepped-care approach to improve clinical outcomes—symptoms, functioning, quality of life, among patients with any mental disorder guided by the Proctor's implementation framework. The framework asserts that implementation outcomes serve as necessary preconditions for attaining subsequent desired changes in service or clinical outcomes and that clearly distinguishing implementation outcomes from treatment effectiveness is critical for transporting interventions from research settings to practice. The two studies examine three distinct yet interrelated outcomes—implementation, service, and patient outcomes—to determine the overall system-level impact of implementing stepped-care comprehensive mental health services. Certain implementation factors, such as acceptability, appropriateness, and fidelity to the approach, are more salient early in the implementation process and may be leading indicators of reach, retention, sustainability, and cost, the most important implementation outcomes. The extent to which services are safe, effective, patient-centered, timely, efficient, and equitable as they are being implemented ensures the quality of care that fosters better clinical care and outcomes and patients' satisfaction. Ultimately, sustainability is the goal.

Both studies evaluate implementation, service, and patient outcomes using a digital platform, the Mental Wellness Everywhere Digital Training and Services Platform (mWell Everywhere), to create digital patient records, screen for all mental disorders and suicide risk using a validated ultra-brief screening (3 items) and triage (10 items) Mental Wellness Tool (mwTool), and guide task-shifted evidence-based treatment delivery with fidelity while monitoring clinical, services and implementation outcomes, and tracking individual and community-level social determinants. The instrument developed in Mozambique combines valid measures and has been tested elsewhere to comprehensively assess all psychiatric disorders and the severity of the illness to guide triage and treatment. The 3-item screener with proxy respondents detects any mental disorder among household relatives, thus allowing for further community reach. This permits efficient and convenient community assessment—an informant can identify with three questions the need to assess a family member living in the same household. Me-Well meets COVID-19 safety requirements (e.g., telehealth); and provides psychosocial support, self-help, and guidance to communities and providers. Using Me-Well, patients with common mental disorders (depression, anxiety, PTSD) receive Interpersonal Counseling, four structured sessions designed for use in nonspecialist settings, and focusing on the relationship between interpersonal problems (e.g., disputes, life changes, grief, loneliness) and the onset of psychiatric symptoms. For patients with severe mental disorders, providers use a tablet-based Medication Diagnostic Algorithm to deliver psychopharmacologic treatment according to specific diagnoses. For treatment of patients with substance use disorders, providers use Screening, Brief Intervention (Motivational Interviewing), and Referral to Treatment (SBIRT), a brief, one to four–session therapeutic approach that employs empathic listening to explore patients' goals, resolve ambivalence, and elicit motivation to change. Patients with moderate to high risk of suicide work with the provider to develop a suicide Safety Plan Intervention (SPI) that includes identifying warning signs, coping strategies, and sources of support to alleviate a suicidal crisis when hospitalization is not clinically indicated.

All of these evidence-based treatments have mobile applications that guide facilitators in standardized treatment steps and allow remote monitoring of patient outcomes and treatment fidelity. The Me-Well Everywhere platform guides and supports providers using interactive methods, knowledge tabs with tips and sample scripts, and automated processes to enable both supervisors and providers to recognize clinical “red flags” or fidelity issues. Meta-data from the platform tracks treatment fidelity and individual provider platform uptake implementation outcomes, for example, number of patients seen, the degree to which key activities are completed, and the average number of sessions per patient. Any user data that appears to deviate from the standard sequence of activities is automatically flagged for the supervisor to intervene and provide corrective feedback to the provider. The platform generates a detailed clinical record after each encounter and can communicate with any Electronic Medical Record platform to enable access to patient characteristics and medical needs (e.g., pregnancy, multimorbidities, current medications).

Implementing Me-Well Everywhere in Mozambique will expand the current mental health system that currently only provides access to care to the 30% of the population living in urban areas. In New York City, like in many high-income countries, shortages in mental health resources and how services are rendered create long waiting periods with the ultimate expectation of ongoing rather than evidence-based, time-limited treatment, requiring expansion of human resources to increase access. In reality, more than half of patients who are able to access mental health services drop out of care after 3 to 6 sessions, unrelated to symptomatic improvement. Thus, can we meet people where they are and provide indicated and tested treatment within that brief time frame by using short-term, evidence-based treatments, and reassessing the need for more care, and offer sequential treatments accordingly. Evaluating how to provide services with a more efficient approach using sequentially targeted and tested short-term treatments with proper follow-up assessment to determine the need for continued services may inform solutions to increasing access to those in need waiting for care.

Mental health disparities are associated with social determinants and substantial individual, familial, and community-level suffering. The public health system is yet to address the population's mental health and social determinant needs, which in turn worsen mental health and functional outcomes at the individual and community level. Systems of care must tackle all mental health disorders, especially when comorbidities are the norm. However, the scarce siloed mental health resources are insufficient to address these mental health disparities. Health care systems require integrating comprehensive mental health services, implementing brief, valid community, and individual screening approaches, and evidence-based treatments with fidelity. Principles of implementation science provide a road map for how to do so sustainably and inform needed policies to address the vast mental health treatment gap globally. Implementation of evidence-based treatments demands expanding the mental health workforce using task-shifting by rigorously training and supervising human resources with diverse training backgrounds. A multilevel transformative integration of evidence-based, comprehensive mental health services stepped-care leveraging rigorous technologies into health systems in partnership with the affected communities and policymakers can meet these challenges, ensure sustainable, comprehensive mental health services with associated policies to decrease mental health disparities, and improve communities' functioning and social determinants.

OVERVIEW

Pressures are mounting to ensure that the fruits of biomedical research translate into treatments and services. Mental health services research has substantial power to both inform and advance this process. In the end, all mental health services research can be summarized in three questions. Do all persons with mental health needs have access to treatment? Is the treatment they receive of adequate quality, incorporating the latest in scientific knowledge? Does the treatment offer good value—that is, good outcomes for the resources expended? These questions are posed by the field in such a manner as to maximize generalizability and to identify subgroup variations and needs. These questions must also be asked within systems concerned with health care equity and social policy to inform clinical care and maximize the benefit to all individuals—the ultimate goal.

Mental health services research, necessarily interdisciplinary, must consider how social factors, structural barriers, financing systems, organizational structures and processes, health technologies, and personal behaviors affect access to health care, the quality and cost of health care, and ultimately health and well-being. Health services researchers work closely not only with biomedical colleagues but also with investigators from other scientific disciplines (sociology, anthropology, psychology, economics, education, and political science) and with service system and policy leaders to implement and evaluate changes.

Much of biomedical science tends to be occupied with internal validity—the strength of the claim that the experimental manipulation causes the outcome. This leads to the desire to control all potentially confounding variables, but it reduces external validity. External validity, or generalizability, lies at the center of the translational process and health services research. At the same time, generalizability encompasses sensitivity to variations among subgroups of patients, taking into account demographic and cultural differences in disease, health care beliefs, and behaviors and access, as well as how health systems address barriers and leverage facilitators of services implementation.

Moving forward, mental health services research faces several challenges. First, while the advance of “evidence-based practices” embodies the principle of translating scientific knowledge into practice, more attention must be given to the complementary principle of “practice-based evidence.” In essence, how can one inform the directions of science through evidence that emerges from clinical practice and experience? Second, challenges for treatment systems and for researchers seeking to identify and promote “best practices” arise from the growing diversity of U.S. culture. Third, the field needs a better understanding of how to implement system changes and disseminate best practices. For this, the field must look beyond the traditional medical and social sciences to such disciplines as marketing science. Fourth, research designs that treat service providers as research subjects require careful consideration to ensure the protection of human subjects that have not been traditionally addressed by institutional review boards.

Finally, major challenges exist in the juxtaposition of the comparatively slow progress of science with the needs of practitioners, planners, administrators, and policymakers, who must respond quickly to the needs of patients and systems of care. The Recovery After an Initial

Schizophrenia Episode (RAISE) project of the NIMH provides a current example of a multidimensional partnership serving to hasten the translation of scientific findings to practice. The NIMH funded two RAISE projects to study the effectiveness and implementation of evidence-based, coordinated specialty care programs for young adults experiencing the onset of nonaffective psychosis. Attention to policy and practice while these research studies were being conducted led to the relatively rapid expansion of resources for evidence-based programs as part of the mental health block grant in 2014. For health services research to remain relevant, new models of partnership will continue to be needed—models that provide timely information in an iterative and collaborative manner among all stakeholders.

FURTHER READINGS

Alegria M, NeMoyer A, Falgas Bague I, Wang Y, Alvarez K. Social determinants of mental health: Where we are and where we need to go. *Curr Psychiatry Rep*. 2018;20(11):95.

Boldrini M, Canoll PD, Klein RS. How COVID-19 affects the Brain. *JAMA Psychiatry*. 2021;78(6):682–683.

Croghan TW, Schoenbaum M, Sherbourne CD, Koegel P. A framework to improve the quality of treatment for depression in primary care. *Psychiatr Serv*. 2006;57:623–630.

*Demyttenaere K, Bruffaerts R, Posada-Villa J, et al. The WHO World Mental Health Survey Consortium. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA*. 2004;291:2581–2590.

Dixon LB, Goldman HH, Srihari VH, Kane JM. Transforming the treatment of schizophrenia in the United States: The RAISE initiative. *Annu Rev Clin Psychol*. 2018;14:237–258.

Dixon LB, Glynn SM, Cohen AN, et al. Outcomes of a brief program, REORDER, to promote consumer recovery and family involvement in care. *Psychiatr Serv*. 2014;65(1):116–120.

Dixon LB, Lucksted A, Medoff DR, et al. Outcomes of a randomized study of a Peer-Taught Family To Family Education Program for mental illness. *Psychiatr Serv*. 2011;62(6):591–597.

Dixon L, Lyles A, Scott J, Lehman A, Postrado L. Services to families of adults with schizophrenia: from treatment recommendations to dissemination. *Psychiatr Serv*. 1999;50:233–238.

Essock SM, Goldman HH, Von Tosh L, et al. Evidence-based practices: setting the context and responding to concerns. *Psychiatr Clin North Am*. 2003;26:919–938, ix.

*Goldman HH, Frank RG, Burnam AM, et al. Behavioral health insurance parity for federal employees. *N Engl J Med*. 2006;354:1378–1386.

*Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. National Academies Press; 2001.

Katon W, Robinson P, Von Korff M, et al. A multifaceted intervention to improve treatment of depression in primary care. *Arch Gen Psychiatry*. 1996;53:924–932.

Kessler RC, Berglund P, Demler O, et al.; National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. 2003;289:3095–3105.

Mental Health Atlas 2017. World Health Organization; 2018.

Organization WH. *mhGAP: Mental Health Gap Action Programme: scaling up care for mental, neurological and substance use disorders*. World Health Organization.

Pitschel-Walz G, Leucht S, Bäuml J, Kissling W, Engel RR. The effect of family interventions on relapse and rehospitalization in schizophrenia—a meta-analysis. *Schizophren Bull*. 2001;27:73–92.

President's New Freedom Commission on Mental Health. Achieving the promise: Transforming mental health care in America. *Final Report (DHHS Pub. No. SMA-03-3832, Vol. 2004)*. Substance Abuse and Mental Health Services Administration; 2003.

Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health*. 2011;38(2):65–76.

Shelton RC, Lee M, Brotzman LE, Wolfenden L, Nathan N, Wainberg ML. What is dissemination and implementation science?: An introduction and opportunities to advance behavioral medicine and public health globally. *Int J Behav Med*. 2020;27(1):3–20.

Thota AB, Sipe TA, Byard GJ, et al. Collaborative care to improve the management of depressive disorders: a community guide systematic review and meta-analysis. *Am J Prev Med*. 2012;42(5):525–538.

Wainberg ML, Gouveia ML, Stockton MA, et al. Technology and implementation science to forge the future of evidence-based psychotherapies: the PRIDE scale-up study. *Evid Based Ment Health*. 2021;24(1):19–24.

Wainberg ML, Lovero KL, Duarte CS, et al. Partnerships in research to implement and disseminate sustainable and scalable evidence-based practices (PRIDE) in Mozambique. *Psychiatr Serv*. 2021;72(7):802–811.

Wainberg ML, Scorza P, Shultz JM, et al. Challenges and opportunities in global mental health: a research-to-practice perspective. *Curr Psychiatry Rep*. 2017;19(5):28.

Wells KB, Sherbourne C, Schoenbaum M, et al. Impact of disseminating quality improvement programs for depression in managed primary care. *JAMA*. 2000;283:212–220.

Young AS, Klap R, Sherbourne CD, Wells KB. The quality of care for depressive and anxiety disorders in the United States. *Arch Gen Psychiatry*. 2001;58:55–61.

56.5 The Psychiatric Hospitalist

ANGEL CIENFUEGOS, M.D.

The “psychiatric hospitalist” model has been adopted for over a quarter of a century by the American healthcare delivery system and other systems around the world. As opposed to the nonhospitalist traditional models, these are medical providers who are part of inpatient programs in medicine, pediatrics, surgery, and now mental health in which the medical care is provided. The hospitalist is a specialist in inpatient treatment based in an inpatient facility or hospital. This type of practice specializes in acute and serious illness rather than chronic ambulatory medicine and health maintenance. They are knowledgeable about comorbidities in acute and critical clinical presentations. The number of hospitalists has grown consistently, and it has been suggested that the generalist who cares for patients across the full continuum of care may be a vanishing breed.

Traditionally, patients had been cared for in the hospital by their prior primary care physicians, in most cases family practitioners or general internists in a “Continuity of Care” model. Over the past several years, hospitalized patients’ care is now more likely to be provided by a hospitalist. In especially complex cases that require multiple consultations with subspecialists, it is the hospitalist who coordinates care. When the patient is discharged, they return to the care of the primary care physician and routine follow-up by other medical specialties. This change parallels the change of practice patterns in psychiatry, in which, in many instances, a separate psychiatrist may care for a patient while admitted for inpatient treatment. Historically, these individuals in psychiatry have been called unit directors or ward chiefs of large state mental institutions. However, in their current role, they are the psychiatric parallel of the hospitalist found in medicine and pediatrics. In both cases, outpatient physicians relinquish the care of hospitalized patients to inpatient specialists.

The development of the hospitalist model has occurred similarly for medicine and psychiatry in the context of the changing role of acute care hospitals. As both medicine and psychiatry become more ambulatory in orientation, inpatient institutions have moved away from their central role in health care. They have become specialized institutions for acutely ill patients. Generally, the level of severity of illness in the inpatient setting has continued to increase as well as the number of comorbidities. Although the hospitalist model was widely accepted previously in Canada and the United Kingdom, the first modern hospitalist program was established in the United States in 1994, overcoming the opposition of many primary care physicians who objected to the model because of the disruption in continuity of care. In any case, it was a fact that inpatient care had become an economically inefficient use of their time.

The hospitalist has emerged in response to the increasing complexity of care associated with hospitalized patients. These individuals, with time and training, develop significant experience with acutely ill patients. They develop an understanding of continuity and coordination of care since part of their responsibility is to work on the "hand-off"—the patient's transition from outpatient to inpatient and, after treatment, back again to the prior outpatient physician.

The initial growth in the number of hospitalists occurred because leaders at health maintenance organizations (HMOs), some hospitals, and medical groups believed that they would lead to a reduction in inpatient costs without a reduction in patient satisfaction or quality of care. Reduced residency work hours contributed to the growth of hospitalists as hospitals' need for more on-site physicians grew. Recent estimates indicate that the current number of hospitalists in the United States is nearly 30,000. In 2010, 57% of U.S. hospitals had hospitalists on staff, including 87% of hospitals with a number of beds equal to or above 200. Around 80% of VA hospitals currently have hospitalists. Membership of the professional organization for hospitalists, the Society of Hospital Medicine, is about 6,000.

Because hospitalists have become common at the major teaching hospitals in the United States, they have also become a major provider of instruction to medical students and residents in internal medicine and pediatrics. These two groups now receive much of their inpatient education from hospitalists working in their role as attending physicians.

THE DEFINITION OF HOSPITALIST PHYSICIANS

Hospitalists are defined as physicians who have general inpatient care as the primary focus of their professional activities, which include clinical care, teaching, research, or administration. Hospitalists typically spend more than 25% of their time working in hospital settings. Commonly they are internists taking the place of primary care physicians in providing inpatient care. In addition to patient care responsibilities, many hospitalists become involved in clinical-administrative tasks. Commonly these include quality improvement efforts and peer review. Academic hospitalists may explore issues such as outcomes research that explores the cost and quality of inpatient care. Many lead the development of new inpatient educational programs for house staff and medical students. Most hospitalists work in community hospitals; a smaller number work in teaching hospitals. In those academic settings, their responsibilities expand to include teaching medical students and residents.

In general medicine, most hospitalists have been trained as internists. There is no separate training program for hospitalists, although some residency tracks and 1-year fellowships have begun to emerge. Such residency tracks typically focus on training in inpatient medicine in addition to hospital-oriented quality improvement, patient safety, and end-of-life care. Fellowships usually focus on providing future academic hospitalists with the research skills they will need in areas related to inpatient care.

Hospitalists are usually salaried physicians with defined work hours. They care for patients with complex medical illnesses rather than the chronicity found in outpatient practice. Many of them have interests in health systems and want a chance to improve the quality of services and control costs associated with inpatient practice. The increasing complexity of medical care demands greater specialization. In many settings, it is increasingly regarded as unrealistic to expect physicians to have expertise in both inpatient and outpatient medicine. Supporters of the hospitalist movement emphasize the benefits of a specialized knowledge base. Hospitalists specialize in inpatient care, allowing primary care physicians to focus on ambulatory care. A consequence of the increase in the number of hospitalists has been a proportionate increase in the number of “office-only” physicians as well.

Hospitalists in academic settings are on-site experienced physicians. Their presence helps hospital wards to run more smoothly and, by doing so, increases patient and nursing staff satisfaction. Such individuals are usually familiar with the ward dynamics. They also usually have a good relationship with the ancillary staff, appreciate their contributions, and have long-term commitments to the hospital. Such involvement facilitates the management of clinical situations to prevent them from evolving into significant problems. Such individuals also usually have a breadth of clinical knowledge and experience from working in hospitals and are skillful in the expeditious treatment of the seriously ill, comfortable in managing multiple minor inpatient problems, and familiar with consulting various medical and surgical specialists. These attending physicians are also best positioned to provide the supervision that interns and residents need when managing acutely ill inpatients. Due to their inexperience and rapid turnover, interns and residents benefit from the support that a responsible physician provides when making major medical decisions.

It has been suggested that the minimum amount of inpatient work expected of a hospitalist is 25% of their practice. This minimal amount allows the hospitalist to become experienced in the management of common inpatient conditions. Relationships with care providers in affiliated disciplines can also develop more easily with this degree of availability. Such individuals are also more likely to become committed to improving the hospital by participating in quality improvement efforts and helping to control costs.

There are two common models of hospitalist care. The most common consists of full-time inpatient physicians who practice as a group. In the second model, a group of physicians rotate responsibility for inpatient care. The rotational periods vary, depending on the practice of the group. This method allows a group to maintain skills in both inpatient and outpatient practice. It can, however, produce frequent periods of discontinuity in outpatient care.

An additional aspect of the hospitalist model is whether it is voluntary or involuntary. In some managed care organizations, the outpatient physician is required to transfer care to a hospitalist when the patient is admitted for inpatient treatment. Commonly under this model, the outpatient physician is not permitted to write orders and is relegated to visiting the patient in the hospital. In contrast, many physicians voluntarily transfer responsibility for the care of their patients to hospitalists, cardiologists, or specialists in intensive care medicine. For example, it is common for primary care physicians to transfer care of their patients to intensivists (physicians specializing in intensive care) when their patient is admitted to the intensive care unit (ICU). In a similar manner, geriatricians working in nursing homes commonly transfer care to internists when their patients are admitted to the hospital. Although not all hospitalists may have the same degree of interest or ability in studying patient outcomes or developing practice guidelines, most of them participate in these creative activities.

ROLES OF THE HOSPITALIST

The predominant role of the hospitalist is to take care of hospitalized patients. Over the years, inpatient care has become more complex and specialized, and while the number of hospitalized patients has decreased, the acuteness of their conditions has risen. Simultaneously, the scrutiny of the length of stay (LOS) and associated hospital costs has increased. It is common to see hospitalists involved in projects to reduce the LOS and cost of admission. Generally, reports show that inpatient care of medical patients by hospitalists reduces LOS and hospital costs. Hospitalists are able to respond rapidly to changes in a patient’s condition because they tend to be readily available and are able to expedite the timing of treatments and consultations.

The hospitalist is also the receptor for inpatient admissions. Increasingly, outpatient physicians prefer to transfer hospital care to a specialist. Common reasons for this preference include the convenience of uninterrupted outpatient work, the convenience of not making rounds in the hospital, and a lack of expertise or comfort with the acuteness of illness and clinical skills associated with inpatient work. There is also some desire to avoid the intensive scrutiny associated with providing care to hospitalized patients. Similarly, hospitalists provide pre- and postoperative medical care to surgical patients. Their availability allows the surgeon to be in the operating room all day and ensures that the patient receives appropriate medical care. Many health care systems are having hospitalists become the physician of record for nonmedical inpatients with medical comorbidities.

An analogous situation has long existed with ICUs. Most ICUs are closed, meaning that they use dedicated critical care physicians who accept the patient from the generalist on the ward and transfer care back to them when the patient leaves the unit. These physicians have the skills, abilities, and availability to care for critically ill patients, and because of these skills, they have improved outcomes and efficiencies.

In many specialties, including critical care, oncology, bone marrow transplantation, cardiology, infectious diseases, and psychiatry, the hospitalist model is common. The attending physician is dedicated to the inpatient service. In academic centers, this may be for multiple rotations throughout the year or full-time.

An additional role for hospitalists is providing physician participation in hospital management. Commonly this occurs by their joining in quality management operations, the development of hospital policies and procedures, management of the formulary, participation in credentialing activities, and utilization management.

When considering the role of hospitalists, it is useful to realize that they also care for outpatients, often in a consultative role. Some hospitalists participate in follow-up clinics in which patients are seen once or twice following discharge. Such outpatient activities, however, represent only a small fraction of their time.

Hospitalists also play a role in communication. Most systems expect the hospitalist or their designee to contact the primary care physician within 24 hours of admission and to ensure that either a summary letter or the discharge summary is sent to the outpatient clinician at discharge.

Another key role for hospitalists in communication is in the transmission of clinical information from the outpatient care physician to the inpatient provider on hospitalization and the subsequent transfer of information back to the outpatient care physician by the hospitalist on discharge. This “hand-off” is symbolic of the transfer of clinical responsibility that occurs during hospitalization. The hospital-based physician assumes responsibility for diagnosis and management. For this reason, most hospitalists believe that there is an obligation to validate the diagnosis. This is, in essence, the provision of a second opinion about the diagnosis and subsequent management. Much has been written about the hand-off. It is one of the two controversies of the model (the other controversy being the obligatory use of hospitalists). The disadvantage of the hand-off is obvious: important information can be lost and result in discontinuity. There may also be patient dissatisfaction because of the assignment of a new physician.

The hand-off can also be viewed as producing one of the strengths of the hospitalist model. It results in the patient being under the care of a physician in the hospital throughout the day. This individual reacts to clinical data in real time. They are the coordinators of inpatient care. However, some surveys show that patients are at least as satisfied with the care from hospitalists as they are with care from their primary care physician. The data suggest that patients may be willing to trade familiarity for availability. It is also noteworthy that many hospitalized patients do not have primary care physicians, especially among the uninsured. In some systems, when hospitalists are introduced, there is an increase in patient satisfaction with their outpatient physician, suggesting that the improved ability to focus on their outpatients is a benefit of this model for the primary care physician.

Hospitalist systems that focus on inpatient–outpatient communication have fewer difficulties. Commonly, programs fax the discharge summary to the outpatient clinician on the day of discharge. Some programs fax daily progress notes as well. Increasingly, programs have a continuity clinic in which patients may be seen once or twice soon after discharge to ensure stability before the patient reappears at the primary care physician’s office.

THE PSYCHIATRIC HOSPITALIST

It could be argued that psychiatry was the first of the medical disciplines to adopt the hospitalist model on a large scale. Of interest, however, full-time inpatient psychiatrists have never called themselves hospitalists. Psychiatric hospitals and asylums date back to the early Middle Ages. Until the rise of psychoanalysis, the most widely available career opportunity available to doctors treating the mentally ill was working in a hospital. The choice of career was to work in a publicly supported state mental hospital or a private psychiatric hospital. The physicians then resembled what we would consider to be hospitalists.

In the twentieth century, the advent of antipsychotic drugs led to the trend of deinstitutionalization. Psychiatry, together with numerous other disciplines, adopted the model of continuity of care, in which physicians follow their ambulatory patients in outpatient settings and while hospitalized in the event of a needed inpatient stay.

As the field has expanded and the effect of third-party payers has restructured the profession, the proportion of hospitalists has declined while the absolute number has increased. One of the earliest surveys of career areas for psychiatrists dates to 1953; it noted the role of psychiatrists working full-time in hospitals. Later surveys noted that the percentage of psychiatrists working in mental hospitals was decreasing. In 1965, for example, about 30% of psychiatrists worked in hospitals, but by 1979 this had dropped to less than 14%.

One of the problems that nonhospitalist-based psychiatrists encountered was that admitting and caring for hospitalized patients was very disruptive to their careers as outpatient clinicians, academicians, teachers, and researchers. The role of unit director or ward director has been a traditional role in psychiatry. At one time, these individuals were relatively junior psychiatrists who, in academic settings, attempted to succeed in the traditional “triple threat” role. That is, they pursued success in research, patient care, and teaching. In any setting, academic or not, these individuals are the epitome of the psychiatric hospitalist. Their job functions are highly congruent with the descriptions of hospitalists noted earlier.

In 2010, a hospital part of the Veterans Affairs Healthcare System transitioned its mental health inpatient units from nonhospitalist or continuity of care (nonhospitalist based) to a hospitalist-based model of care. This change resulted in an increase in the total number of discharges and a decrease in the number of readmissions within 30 days. Patients discharged by hospitalists were more likely to follow up with after-care appointments.

From a financial standpoint, psychiatric healthcare delivered by hospitalists decreases the LOS and the cost. It also facilitates the care of patients with medical comorbidities since medical consults are readily available, and hospitalists can easily coordinate care with hospitals’ medical and surgical disciplines. In turn, psychiatric hospitalists provide psychiatric consultation to medical and surgical units, reducing those specialties’ cost of care.

Psychiatric hospitalists are familiar with the assessment and treatment of the most severely mentally ill patients, including patients with acute psychotic agitation and patients who attempt to harm themselves or others. They manage patients whose psychiatric illnesses put them in physical danger, such as patients who are suicidal or engage in self-injurious behaviors. They are comfortable with protocols to manage delirium and to assess and treat withdrawal from alcohol and other substances.

Psychiatric hospitalists also have extensive training in the legal requirements for involuntary detention and involuntary medication administration, and they frequently testify in judicial evidentiary hearings. Civil commitment statutes vary across states, and psychiatric hospitalists are familiar with the statutes for their state, including the requirement for civil commitment hearings within a specified time and the rights of the detained person. For example, some states give detained patients the right to refuse treatment, while others do not. Psychiatric emergency hold laws vary in length from 23 hours to 10 days across the United States, and psychiatric hospitalists know the criteria that must be met for the initial emergency hold and for psychiatric holds to be extended. Psychiatric hospitalists spend most of their time in inpatient units. They carry a significant share of resident and medical student training. Trainees working with hospitalists are frequently more satisfied with the amount of time and dedication they get from their attendings, which normally are part of the clinical faculty of the associated academic institution. These faculty members are expected to be educators in addition to clinicians and are not subjected to the same publication and funding expectations of research faculty. The inpatient units are the center for education and training during the first years of residency and offer the appropriate diversity of clinical presentations providing students with the appropriate amount of learning material.

Psychiatric inpatient units are environments where multidisciplinary approaches involving psychiatrists, nurses, psychologists, social workers, caseworkers, substance use counselors, and, as needed, medical consultants from other medical disciplines. Due to the multidisciplinary nature of inpatient hospital work, psychiatric hospitalists are able to treat medical comorbidities more effectively and are well-qualified to assess appropriateness of transfers from medical to psychiatric units.

HISTORY OF THE MODERN HOSPITALIST MOVEMENT

Several factors have led to this practice pattern. For outpatient clinicians, the arrangement frees them from the disruptive burden of seeing a small number of inpatients. It allows them to focus on their area of expertise: outpatient practice. It avoids the disruptions in the outpatient schedule that can occur when inpatient rounds are a daily responsibility. For the hospital and third-party payers, the model reduces inpatient expenditures with similar clinical outcomes. This model does not adversely affect patient satisfaction with care or trainees' satisfaction with their educational experience. Although this practice pattern may benefit hospitals or group practices, its professional fees are not usually sufficient to make the program self-supporting. To remain viable over the long term, such programs require support from the programs such as hospitals or group practices that receive benefit from the efficiencies that these programs produce.

The traditional model for inpatient care was outpatient physicians who made hospital rounds early in the morning and again at the end of the day. There could be calls from the emergency room or the hospital. The calls would, at times, interrupt outpatient practice and interfere with the efficient treatment of outpatients. Many clinicians noted a decline in office productivity when they were obliged to make hospital rounds for a small number of inpatients.

A significant factor that led to the development of hospitalists was the desire to focus efficiently on outpatient treatment and allow inpatient treatment to be provided by those who would give it full-time attention. This allows outpatient physicians to avoid the frequent interruptions, paperwork, and preauthorizations associated with inpatient care.

The acuteness and diagnostic complexity of inpatients have been increasing for the last several decades. At the same time, new and more complex treatment modalities have emerged in inpatient care. Those who only infrequently care for inpatients have found it increasingly difficult to remain current in these treatment methods. Part of the appeal of the hospitalist movement has been that it frees outpatient physicians from the responsibility to remain current with treatments they would only rarely use.

In England and Canada, the early recognition of the benefits of focusing on either ambulatory or inpatient care led very early in this process to the concept of distinct outpatient and inpatient physicians. The term "hospitalist" was first used in the United States by Wachter and Goldman to describe the concept.

In academic medical centers, hospitalists have become a third career pathway in addition to the prior pathways of clinical educator and clinical researcher. The development of the clinical educator had been due to the need to care for outpatients and to teach students and residents at ambulatory care sites. How an individual physician spends their time is highly variable and tends to depend on the needs of the institution and the skills and experience of the individual physician. Time is typically divided among providing direct care to outpatients, supervising students and residents in providing ambulatory care, and various scholarly activities. Many also spend several months each year attending an inpatient unit, during which time outpatient responsibilities are reduced. Educational activities may include responsibility for specific courses in the medical school curriculum or the teaching of specific knowledge or skills, for example, the patient examination. Most academic medical centers expect their clinician teachers to do scholarly work to meet standards for promotion. These activities commonly include developing new educational programs, writing review articles or case reports, conducting research, and presenting original work at scientific meetings.

In Europe and Canada, there have long been specialists in inpatient general medicine who have been active in urban hospitals. A clear demarcation between inpatient and outpatient practice is new in the United States. In the United Kingdom, the consultant physician would take over care from the general practitioner when the patient was admitted to the hospital.

The increasing use of such hospitalists was initially spurred on by the emphasis on efficiency that came with the rise of managed care. The managed care environment stimulated the development of the hospitalist model in the United States. The emphasis on efficiency and cost containment took precedence over continuity of care.

Up to the 1970s, most primary care was provided by a general practice physician who was not board-certified in any particular specialty. Most medical care at that time was provided in a hospital setting. In the inpatient setting, the general practitioner was often assisted by the internist, who was often more experienced in inpatient medicine. By the 1980s, family practice had emerged as a specialty and came to define the general practitioner. The role of the internist in inpatient practice was diluted by the rise of medical specialists who could assist the general practitioner with inpatient care. Until the mid-1990s, hospitalists were uncommon in general medicine in the United States. Rare examples of hospitalists in the United States before the 1990s were called house doctors, hospital rounders, or inpatient managers. By the 1990s, however, general internists and family practitioners had begun to focus on outpatient care and had a progressively smaller role in inpatient care.

Internists began to manage inpatients, and intensivists emerged to manage ICUs. Emergency departments began to be staffed by specialists in emergency medicine.

Although multiple forces were responsible for this change, two major factors were predominant—cost pressures on hospitals and physicians and the increasing acuteness of inpatients. The shift of most medical care to the ambulatory setting resulted in most physicians having little ongoing experience with inpatient care. Practice effects meant that those who did large volumes of inpatient care had superior outcomes.

Many felt that the hospitalist model conflicted with the traditional notion of a generalist who would treat patients in both inpatient and outpatient settings and, in so doing, provide treatment that was less fragmented and more holistic.

THE STAGES OF HOSPITAL CARE

It has been suggested that hospital care in the United States has gone through four stages. In the first stage, all inpatients were cared for by their primary care physicians. This worked best when the acuteness of illness was low, most practitioners had large numbers of inpatients, the cost was not as important as it is today, and hospitals were adjacent to practitioners' offices.

The second stage occurred when these conditions no longer existed. Specifically, hospitals are no longer nearby, the acuteness of inpatients is significant, costs are managed, and inpatient care has a significantly higher degree of complexity. Furthermore, outpatient offices are busy due to the shift of care from inpatient to ambulatory. At this point, physicians may form a group to rotate responsibilities for inpatient care. This system introduces an inpatient–outpatient discontinuity. The compensation for this is the enhanced availability of the inpatient physician.

A third stage is the development of voluntary hospitalists who are dedicated inpatient physicians paid by the hospital, medical group, or managed care company. Under this model, not all inpatient care is provided by hospitalists. Primary care physicians retain the opportunity to manage their patients when they are admitted to the hospital.

A fourth stage is the mandatory hospitalist stage. At this point, all inpatients are cared for by hospitalists. This model is most common when there is strong institutional control of physicians. When there is a loose affiliation between the hospital and primary care providers, the relationship is more likely to be at stage three if outside physicians wish to retain control of their patients.

BENEFITS FROM THE USE OF HOSPITALISTS

Many studies of hospitalists have demonstrated a reduction in the cost of care without a change in the quality of care. Hospitalists are more efficient at hospital resource utilization than physicians who spend the majority of their time practicing in ambulatory settings. Many factors are believed to contribute to the improved efficiency of hospitalists. These include the practice effect of being familiar with common inpatient conditions and being more familiar with the hospital system. They are also more available to acutely ill patients and may respond more promptly to changes in their clinical state or changes in laboratory data. Through a rotating on-call system, a hospitalist is on-site at the hospital at all times.

Ambulatory care physicians experience a significant benefit from the presence of hospitalists. Hospitalists allow outpatient-based clinicians to attend to their patients without the disruption in their schedules that can result from the unpredictable demands associated with caring for acutely ill patients in the hospital. Furthermore, many ambulatory care-based clinicians may lack the familiarity with the inpatient management techniques that they had when they were trainees.

For the typical internist, the inpatient census is small. However, this small number of patients typically requires more effort than in the past because the patients tend to be more acutely ill and their care more complex. The pace of hospital work has also quickened, and physicians are under pressure to minimize the LOS. Adding to these challenges is the fact that many conditions that were once an indication for inpatient admission are now commonly treated in the outpatient setting, whether that be the office, the home, or the nursing home. The result is that outpatient practice is also more acute and busier, making the disruptions of providing care for some inpatients also more challenging.

Because most primary care physicians spend decreasing amounts of time caring for inpatients, many feel they have lost some of their skills in inpatient care. This reduction has occurred both because fewer patients are admitted and because the LOS is shorter than in past decades. With fewer inpatients, there are fewer interruptions in a tightly scheduled day of outpatients. What emerges is an increasing split between inpatient and outpatient practice.

Hospitalists reduce inpatient resource utilization, resulting in a shorter LOS and a lower cost of inpatient care without a reduction in quality of care or patient satisfaction. The reduction in resource utilization probably is a practice effect combined with the ready availability of hospitalists for diagnostic and management decisions.

The rapid increase in the number of hospitalists in the United States suggests that there are significant benefits to this position. In terms of patient care, their presence in the hospital and ability to be available on short notice allows patients to be managed rapidly. Their availability permits them to assess patients frequently and to promptly follow up on test results. This is thought to be one of the significant factors in how hospitalist care produces a reduction in LOS. Hospitalists may also shorten a patient's stay in the emergency department by being available throughout the day to admit a patient. An experienced hospitalist usually becomes comfortable caring for acutely ill patients with a variety of medical conditions. In some cases, this may result in a decreased use of laboratory studies and consultative services. This may also contribute to a reduction in the cost of inpatient care.

There are a large number of uninsured patients in the United States. When they present in the emergency department seeking admission, there usually is no primary care physician with whom they have a prior relationship. Therefore there is no outpatient physician available to manage their inpatient care. An increasing role for hospitalists is to provide care for these patients.

An additional benefit of the hospitalist approach is the ability of outpatient physicians to dedicate themselves to outpatient care without the distractions and interruptions of acute inpatients. Many office-based physicians feel that providing inpatient care is not cost-effective. The poor reimbursement for hospital visits is a common reason that many no longer provide hospital care. Outpatient physicians are also increasingly seeing sicker patients. Their volume of inpatient work has also increased. Due to these factors, many physicians feel that it is increasingly important to focus on improving their availability and skills in ambulatory medicine. These goals are facilitated by focusing exclusively on outpatient medicine. Inpatient medical care is increasingly complex. Due to this increasing complexity, inpatient care requires a higher degree of specialization than in the past. For this reason, it is increasingly difficult for physicians with limited inpatient practices to maintain their expertise in inpatient care. Good communication between inpatient and outpatient physicians also has enhanced the acceptance and perception of the benefits of this model. A key aspect of communication is the prompt delivery of a discharge summary to the outpatient physician.

In all areas of medicine, there has been a shift away from relatively healthy individuals with acute illness being admitted to the hospital toward individuals with increasing levels of medical complexity. The hospitalist is increasingly seen as the specialist focused on the medically complex inpatient. Due to their training and experience, these inpatient physicians become proficient in the efficient management of acute illness. Commonly patients are hospitalized because they are profoundly ill or suffer from a rare or atypical illness. They benefit from their hospitalist's inpatient skills.

Hospitalists also work with surgeons to provide primary inpatient care to surgical patients in the pre- and postoperative periods. These combined hospitalist–surgeon teams have been associated with reductions in postoperative complication rates with no significant difference in cost or LOS.

Hospitalists improve the efficiency of inpatient care while maintaining or enhancing the quality of care. Their expertise comes from familiarity with the management of common inpatient conditions. Efficiency consists of reductions in resource use, usually measured as the average LOS and hospital cost. This occurs because hospitalists are familiar with clinical pathways, are experienced in the planning of postdischarge care, and are available throughout the day to respond to changes in diagnostic data or clinical conditions.

Hospitalists relieve office-based physicians of the unpredictable demands of hospital practice. Patients are no longer kept waiting, and physicians no longer have the inconvenience of recurrent trips to the hospital. Outpatient physician productivity increases following the removal of inpatient responsibilities. Outpatient physicians are more reliably available. They are not interrupted by hospital calls or the need to abruptly leave to see acute inpatients. They are able to expand their outpatient hours because they no longer make inpatient rounds. Their workday is more predictable.

Hospitalists standardize the quality of inpatient care. In many academic medical centers, there has been a concern about the wide variation in inpatient abilities found in community-based clinicians. Many programs have worried about granting these physicians admitting privileges or opportunities to participate in the teaching service. When a hospitalist system is used, it allows both the community-based physicians and the outpatient academic faculty to enjoy the same quality of inpatient care that is provided by hospitalist faculty.

Hospitalists produce clinical science with an emphasis on outcomes research, quality assessment, and quality improvement; clinical trials to improve treatment or diagnosis; and the development of clinical guides and pathways.

There is a positive correlation between experience and quality of care in a number of areas of medicine and surgery. When implemented, the hospitalist model has resulted in fewer but more appropriate specialty consultations, lengths of stay, and hospital costs. From familiarity, hospitalists develop good working relationships with ancillary hospital staff, including discharge planners, nurses, consulting specialists, and outpatient physicians. Retention and recruitment improve in the hospitalist system.

HOSPITALISTS AS CLINICIAN-EDUCATORS

Large academic institutions are the leaders in hospitalist-based care. In these institutions, one of the hospitalist functions is the training of residents and medical students. Some are considering or implementing hospitalist fellowships. This means house staff receive training from those with expertise in hospital care. There is an expert readily available throughout the day to confer with about patient management. There is improved teaching and expert guidance. Hospitalists provide role modeling to trainees for careers in hospital work.

The effect of hospitalists on academic medical centers may further the separation between researchers and clinicians. The researchers will further focus on research and obtaining grants. The clinician-educators will concentrate on clinical work and teaching. With the group of clinician-educators, there will be a further subdivision, with some focusing on outpatients and others on inpatients.

Hospitalists also improve the educational experience for medical students and house staff. Because the faculty supervisors are skilled and experienced in modern inpatient care, they provide more effective teaching and oversight. Generally, when academic programs have evaluated the teaching abilities of hospitalists, there has been enhanced satisfaction by students and residents when compared to prior models. Some trainees have also noted a reduction in workload associated with the hospitalist model. House staff report that hospitalists are more effective at teaching than general psychiatrists.

Many academic medical centers have hired hospitalists as clinician-educators. This need has arisen because primary care physicians are able to spend very little time in the hospital. This results in less time with trainees for educational activities. The expectation is that hospitalists will devote more time to education, performance feedback, and supervision.

Resident evaluations of teaching by hospitalists are higher than those of traditional attendings who briefly rotate onto the teaching service.

Clinical research done by hospitalists has traditionally consisted of studying the hospitalist movement. This included studies of the efficiency, cost, and quality of hospitalist-run services, including inpatient–outpatient communication. Research has begun to expand to include prevention of nosocomial infections and hospital quality assessment, including medical error and pain management, health services and outcomes research, and ethics. Newer areas also include clinical trials that focus on conditions common to the hospital and genetic epidemiology.

DISADVANTAGES AND OBJECTIONS TO THE HOSPITALIST MODEL

There are two potential models for the hospitalist system: Voluntary and closed. In a closed system, only hospitalists have admitting privileges, and ambulatory care physicians are excluded. An open system allows both groups to admit their patients. Concerns have been raised that a closed system may engender bad feelings among community physicians.

A potential disadvantage of the hospitalist model is that outpatient physicians will lose their skills in treating inpatients (although some have questioned how serious a disadvantage this may be if they are no longer providing inpatient care). It is clear, however, that by focusing on outpatient care, they can be expected to improve their skills in that area.

Subspecialist consultants may be apprehensive that hospitalists, due to their familiarity with common inpatient conditions, may produce fewer consultations. A reduction in consultations would be associated with a lower cost of care. Because of the severity of illness in hospitalized inpatients, consultations will continue to occur.

During the transition from outpatient to inpatient and again with the transition back to outpatient status, there is a risk of poor communication between the treating clinician and the receiving clinician. With appropriate phone calls and copies of discharge summaries and letters, it is possible to minimize this problem and avoid a significant loss of medical information. An aspect of this coordination of information flow is the risk of fragmentation of care and the loss of a care coordinator who has comprehensive knowledge about the patient’s health. In addition, the discontinuity of care and associated lapses in information flow may be associated with a higher risk of medical error.

An additional issue is that the number of acute care hospitals is diminishing, and with them, so is the relative percentage of patients admitted to the hospital.

Implementation of the hospitalist model has an effect on the roles of other physicians operating in the hospitalist system. Hospital-based specialists typically serve as consultants to admitting physicians. Because hospitalists, due to their increased experience, are more familiar with common inpatient conditions, their need for specialty consultations is less than that of nonhospitalist admitting physicians. This has an effect on the volume of consults and the income of consultants.

Generally, residents have been happy with the quality of teaching provided by hospitalists, ranking them higher than traditional attendings. However, because of their greater expertise in patient work, it can also be expected that hospitalists will threaten the autonomy of residents. They tend to have an increased physical and intellectual presence on the wards.

Discontinuity of care is a consequence of changing physicians when moving from the outpatient to the inpatient setting. The associated concern is a gap in information about the patient’s wishes and clinical state during the transition. The outpatient clinician loses care of the inpatient, and the hospitalist loses follow-up care when the patient is discharged. This discontinuity is often addressed by hospitals by developing programs to enhance communication. These efforts commonly include the prompt calling of outpatient clinicians on admission and discharge and faxing discharge summaries.

Evidence suggests that patients accept hospitalists without significant difficulty. Generally, they are satisfied with the quality of care and do not perceive significant differences between hospitalist care and that provided by their outpatient clinician or traditional ward attendings. It has been suggested that patients may be willing to trade the familiarity of the outpatient clinician for the enhanced availability of their hospitalist. It has also been suggested that patients may no longer be as emotionally bound to their physicians. Many patients have no regular outpatient physician. Furthermore, in the current insurance environment, patients may switch health plans and providers with frequency.

Increasingly, training programs in psychiatry, medicine, and pediatrics emphasize ambulatory care. This raises questions about the adequacy of preparation of recent graduates for inpatient attending. Training programs may need to be revised to provide appropriate preparation for the hospitalist role.

Physicians who wish to provide both inpatient and outpatient care may be unhappy with the hospitalist model. For many ambulatory physicians rotating responsibility allows physicians to maintain inpatient care skills. Some ambulatory physicians are uncomfortable with the potential loss of these skills.

PROFESSIONAL ORGANIZATIONS

The National Association of Inpatient Physicians is a nonprofit organization founded in 1997. It is dedicated to advancing the practice of inpatient medicine and promoting excellence and efficiency in clinical care. It is affiliated with the American College of Physicians. There is no analogous professional organization for psychiatrists.

SPECIALTY TRAINING AS A HOSPITALIST

Increasingly, training programs are geared toward providing only enough inpatient training so that graduates are capable of caring for acutely ill outpatients. Special tracks may be developed to provide adequately trained hospitalists. These programs would focus on developing clinician-teachers with the skills and knowledge necessary for inpatient practice. Sufficient outpatient experience would be provided so the hospitalist would comprehend what level of severity can be managed in the outpatient clinic. In medicine, tracks exist at both the resident and fellow levels.

As the training of hospitalists becomes common, one emerging possibility is the development of two types of residencies. One would emphasize ambulatory care, and the other would focus on hospital-based care. This would evolve through the use of different tracks for residents. In psychiatry, this separation would be expected to occur after the first two years, which usually emphasizes inpatient and consultative work.

FINANCIAL, POLITICAL, AND ORGANIZATIONAL ISSUES

In many hospitals, the reimbursement for inpatient care is such that revenues from hospitalist services are insufficient to provide for the costs of physicians to staff the hospitalist program. In many cases, the benefits to the hospital or, in the academic medical center, to the training program are sufficient to justify additional program support. In academic programs, the training and service obligations are often thought to justify the additional cost.

The willingness of ambulatory care physicians to forgo inpatient care has been heavily influenced by the relatively low compensation. Many have realized that they can increase their clinical income by substituting relatively higher-paying outpatients for lower-paying inpatients and avoiding the unreimbursed commutes to the hospital.

However, it is the low rate of reimbursement for inpatient care that complicates the implementation of the hospitalist model. Low reimbursement along with a payer mix that often includes uninsured patients means that few hospitalist programs can support themselves solely on professional fees. Further complicating the picture is the fact that clinicians are reimbursed a single professional fee each day despite commonly interacting with the patient several times a day and also working with allied professional staff and families. The net result is, commonly, a significant shortfall in the revenue needed to pay hospitalist physicians at a rate commensurate with community norms for other physicians. Generally, hospitals are able to generate about 60% to 80% of their salaries by fee-for-service billings and/or capitation dollars. Because of the benefits they provide to hospitals, practice groups, and other sponsoring agencies, these groups commonly supplement the shortfall in professional fee income.

The hospitalist model will have an effect on a hospital's income. Cases that are paid on a per diem basis may produce less income because of the shorter LOS that occurs with hospitalist attendings. In a capitated case rate in which the hospital is paid according to diagnosis-related groups or when a medical group is paid on a contract rate, there are clear benefits from the reduced lengths of stay and reduced use of ancillary tests associated with hospitalist attendings. In such situations, the savings that result from the use of hospitalists may exceed the salary support provided by the institution. This significant savings is the rationale for supporting the salaries of hospitalists.

To avoid the perception of a conflict of interest, most institutions avoid paying hospitalists to reduce the utilization of services. However, many programs have created incentives related to outcomes measures and patient satisfaction.

HOSPITAL MEDICINE AS A NEW SPECIALTY

Fellowships are currently very limited. Typically they focus on training clinical educators and provide a curriculum on teaching theory, curriculum development, and evaluation methods and opportunities to strengthen teaching skills, along with opportunities to receive feedback about teaching activities. Some also provide instruction on research methods and encourage clinical research activities.

In many institutions, hospitalist positions are staffed by recent graduates and are intended to be transient. In these settings, the opportunities for long-term career development are unclear. In contrast, academic settings that intend to develop a stable cadre of hospitalists provide protected time for promoting scholarly activities and working on improvements in the hospital system.

Traditionally academic medical centers have been challenged in attempting to recognize the contributions of clinician-educators. It is these accomplishments that are the basis of being rewarded with promotion through the professorial ranks. Customary measures of accomplishment include publications and grants, along with other objective evidence of regional, national, and international recognition. The challenge has been to apply these assessment standards to the provision of clinical care or teaching ability. Traditionally, teaching capabilities have been assessed by peer and student evaluations. More challenging is the assessment of clinical excellence.

Hospitalists, similar to emergency medicine physicians and critical care specialists, are site-defined specialists. That is, their area of expertise is not defined by an organ or procedure but by where they deliver care. As such, they care for patients with a variety of illnesses and ages within that location.

In medicine, there are other attributes associated with a subspecialty. There is a large group that identifies themselves as hospitalists. There is a professional society with regular meetings. There is a clinical textbook. Clinicians agree that there is a core skill set and body of knowledge.

FUTURE DIRECTIONS

The hospitalist movement is rapidly growing in psychiatry, medicine, and pediatrics. In the latter two areas, where the numbers are more clearly known, the number of hospitalists is larger than those in many board-certified subspecialties. The hospitalist movement, similar to prior trends in emergency medicine and critical care medicine, has a defining environment and a correspondingly separate body of knowledge and skill set.

Hospitalists can be expected to have an increasing role in improving care for hospitalized patients. Due to their extensive familiarity with the workings of the hospital, they are able to identify system problems and help both hospital workers and staff in dealing with stressful situations, including catastrophes, pandemics, and others. They are able to work with administrators and nurses to ensure that necessary equipment is available and proper protocols are in place to facilitate quality care.

Hospitalists have developed related research areas. Typically these are outcomes-oriented studies that focus on hospital care. Examples include quality of care, cost containment, resource utilization, and LOS. Such studies are typically conducted by academic hospitalists at university medical centers.

The digital health movement is changing healthcare across all disciplines, and hospitalists are likely to contribute to the enhancement of electronic health records (EHRs), particularly to ensure their reliability and portability to outpatient providers. While documentation in EHRs can be time-consuming, ultimately, EHRs improve hospital care by providing clear and reliable documentation, in contrast to the hand-written notes of the pre-digital era. Since these EHRs are often accessible to outpatient clinicians, EHRs also facilitate continuity of care.

Currently, the term hospitalist does not refer to a subspecialty. In medicine, there is a growing effort to make the hospitalist a subspecialty with residency requirements and board certification. The future may see similar trends in psychiatry.

Unresolved questions about the future of the hospitalist movement remain. For example, should hospitalists be exclusively inpatient providers, or should there be some component of outpatient or consultative work? Should all inpatient care be delivered by hospitalists, or is there also a role for outpatient clinicians in the inpatient setting? How will hospitalists be compensated, and who will employ them? What is the optimal training to be a hospitalist?

As the acuteness of inpatients increases, the demands of the job increasingly resemble emergency medicine department attendings or intensive care specialists and are less like outpatient primary care clinicians. Whereas current compensation for inpatient and outpatient physicians is similar, it can be expected that compensation for hospitalists will also increasingly resemble that of intensivist or emergency medicine physicians.

Unique to psychiatry, there is a certain lack of recognition of the hospitalist physician as a differentiated area of practice. In other areas of medicine, the hospitalist role is widely recognized. Despite a long history of the hospitalist role in psychiatry, a role that dates back to the origins of the profession, few psychiatrists who work predominantly in a hospital setting would consider themselves to be hospitalists. They are more likely to call themselves unit directors, ward directors, or attending physicians. As the recognition of psychiatric hospitalists increases, a further appreciation of the similarities and differences between psychiatric hospitalists and hospitalists in other areas of medicine can be expected. One can expect, for example, that the activities of colleagues in other areas of medicine will provide guidance for improving the career path for hospitalists, starting with fellowships and progressing through various levels of seniority following training.

Under-recognized and poorly characterized, psychiatric hospitalists can be expected to be at the forefront of interactions with third-party payers that seek to alter the nature of psychiatric hospitalization. Improved recognition of their role, clarity of their training pathway, and identification of the stages of professional career development will promote this career pathway to trainees and ensure that there are well-trained individuals providing expert inpatient care.

FURTHER READINGS

Auerbach AD, Nelson EA, Lindenauer PK, Pantilat SZ, Katz PP. Physician attitudes toward and prevalence of the hospitalist model of care: Results of a national survey. *Am J Med*. 2000;109:648.

Bellet PS, Whitaker RC. Evaluation of a pediatric hospitalist service: impact on length of stay and hospital charges. *Pediatrics*. 2000;105:478.

Bess J, Casher M. The academic psychiatric hospitalist: a model for treatment and training in the academic inpatient setting. *J Psychiatr Pract*. 2011;17(2):145–151.

Casher MI, Bess JD. *Manual of Inpatient Psychiatry.* 2nd ed. Cambridge University Press; 2020.

*Craig DE, Hartka L, Likosky WH, Caplan WM, Litsky P. Implementation of a hospitalist system in a large health maintenance organization: the Kaiser Permanente experience. *Ann Intern Med*. 1999;130:355.

Davis KM, Koch KE, Harvey JK, Wilson R, Englert J. Effects of hospitalists on cost, outcomes, and patient satisfaction in a rural health system. *Am J Med*. 2000;108(8):621.

Diamond HS, Goldberg E, Janosky JE. The effect of full-time faculty hospitalists on the efficiency of care at a community teaching hospital. *N Engl J Med*. 1998;129(3):197.

Egan JM, Webber MGT, King MRD, Boyd M, Preez-Wilkinson G, Brock D. The hospitalist: a third alternative. *Med J Aust*. 2000;172:335.

*Everett GD, Anton MP, Jackson BK, Swigert C, Uddin N. Comparison of hospital costs and length of stay associated with general internists and hospitalist physicians at a community hospital. *Am J Manag Care*. 2004;10:626.

Freese RB. The Park Nicollet experience in establishing a hospitalist system. *Ann Intern Med*. 1999;130:350.

Goldman L. The impact of hospitalists on medical education and the academic health system. *Ann Intern Med*. 1999;130:364.

Haftel HM, Bozynski ME. Changing teaching for changing times: the effect of a hospitalist program on the education of students. *Acad Med*. 2000;75:524.

Hedman LC, Petrila J, Fisher WH, Swanson JW, Dingman DA, Burris S. State laws on emergency holds for mental health stabilization. *Psychiatr Serv*. 2016;67(5):529–535.

Hoffman A, Hatefi AAB, Wachter CR. Hospitalists, value, and the future. *Future Hosp J*. 2016;3(1):62–64.

Kelley MA. The hospitalist: a new medical specialty? *Ann Intern Med*. 1999;130:373.

Kisuule F, Howell EE. Hospitalists and their impact on quality, patient safety, and satisfaction. *Obstet Gynecol Clin North Am*. 2015;42(3):433–446.

*Levinson W. What is an academic general internist? *JAMA*. 2002;288:2045.

Lindenauer PK, Pantilat SZ, Katz PP, Wachter RM. Hospitalists and the practice of inpatient medicine: results of a survey of the National Association of Inpatient Physicians. *Ann Intern Med*. 1999;130:343.

Lipsitt DR. Psychiatry and the general hospital in an age of uncertainty. *World Psychiatry*. 2003;2(2):87–92.

*Lurie JD, Miller DP, Lindenauer PK, Wachter RM, Sox HC. The potential size of the hospitalist workforce in the United States. *Am J Med*. 1999;106(4):441.

*Lye PS, Rauch DA, Ottolini MC, Landrigan CP, Chiang VW. Pediatric hospitalists: report of a leadership conference. *Pediatrics*. 2005;117:1122.

Melzer SM, Molteni RA, Marcuse EK, Rivara FP. Characteristics and financial performance of a pediatric faculty inpatient attending service: a resource-based relative value scale analysis. *Pediatrics*. 2001;108:79.

Pantilat SZ, Alpers A, Wachter RM. A new doctor in the house ethical issues in hospitalist systems. *JAMA*. 1999;282:171.

Paris J. *Prescriptions for the Mind: A Critical View of Contemporary Psychiatry*. Oxford University Press; 2009.

Saint S, Fowler KE, Krein SL, et al. An academic hospitalist model to improve healthcare worker communication and learner education: results from a quasi-experimental study at a Veterans Affairs medical center. *J Hosp Med*. 2013;8(12):702–710.

Schroeder SA, Schapiro R. The hospitalist: new boon for internal medicine or retreat from primary care? *Ann Intern Med*. 1999;130:382.

Sloan PA, Asghar-Ali A, Teague A, Body E, Kunik ME. Psychiatric hospitalists and continuity of care: a comparison of two models. *J Psychiatr Pract*. 2010;16(3):170–176.

*Srivastava R, Stone BL, Murphy NA. Hospitalist care of the medically complex child. *Pediatr Clin North Am*. 2005;52(4):1165.

Van Walraven C, Seth R, Austin PC, Laupacis A. Effect of discharge summary availability during post-discharge visits on hospital readmission. *J Gen Intern Med*. 2002;17:186.

Wachter RM, Goldman L. The hospitalist movement 5 years later. *JAMA*. 2002;287:487.

*Wachter RM, Katz P, Showstack J, Bindman AB, Goldman L. Reorganizing an academic medical service: impact on cost, quality, patient satisfaction, and education. *JAMA*. 1998;279:1560.

White HL, Glazier RH. Do hospitalist physicians improve the quality of inpatient care delivery? A systematic review of process, efficiency and outcome measures. *BMC Med*. 2011;9:58.

Wood R, Wand APF. The effectiveness of consultation-liaison psychiatry in the general hospital setting: a systematic review. *J Psychosom Rese*. 2014;76:175–192.

56.6 Psychiatric Rehabilitation

LUJAIN ALHAJJI, M.D., VANESSA PADILLA, M.D., AND RADU V. SAVEANU, M.D.

Mental illness represents one of the leading causes of disability, with depression being a major cause of disability worldwide. Serious mental illness (SMI) results in severe significant impairment in functioning. Psychiatric rehabilitation can have a dramatic impact on improving functioning, independence, recovery, and integration in society. The field of psychiatric rehabilitation continues to grow and evolve. While initially thought of as a series of psychosocial interventions that were aimed at augmenting psychopharmacologic approaches, it has now become collaborative, integrative, multi- and interdisciplinary. Multidisciplinary care ensures that patients' biologic, psychological, and social factors are recognized and interlinked. Patient-centered approaches are imperative in reaching personal recovery outcomes.

For a successful outcome of psychiatric rehabilitation, a patient-centered, holistic, and strength-based approach needs to be implemented. A comprehensive understanding of the patient's psychiatric conditions, treatment regimen, social support, contextual cues in the environment, and individual strengths and limitations needs to be incorporated.

Psychiatric rehabilitation is intended for people living with severe mental illnesses (SMIs) such as schizophrenia, bipolar disorder, or schizoaffective disorder, as well as for those with treatment-resistant major depression, severe anxiety disorders, and substance abuse (alone or as comorbidity). Desired long-term outcomes should involve increased autonomy, subjective sense of well-being, financial independence, and improved quality of life in addition to focusing on remission and symptom control. Many persons with schizophrenia may achieve control over positive symptoms, such as hallucinations and delusions, but still have to deal with negative symptoms that impact communication, initiative, and interpersonal interactions. Neurocognitive symptoms of schizophrenia have been implicated in reduced ability to function in the home, workplace, and community. However, these cognitive effects extend well beyond issues with memory, processing speed, and executive dysfunction. Indeed, some of the greatest difficulties with community functioning are related to deficits in social cognition, including perception and processing of affect, knowledge of social rules and conventions, inferring another person's intentions, as well as attributing causality for positive or negative outcomes. For example, people living with schizophrenia may struggle with inferences or attributional biases. Lack of insight into one's own processes and misinterpreting the emotions and intentions of others can lead to isolation, limited meaningful interpersonal contact, difficulties dealing with others in the community, and problems with interacting with others in the workplace.

Psychiatric rehabilitation strategies are increasingly holistic, integrative, respect the autonomy of the individual in decision-making, and specify incremental, graduated goal attainment with the goal of optimal function in the community. Functional outcomes and improved quality of life remain important aspects of psychiatric rehabilitation. Providing integrative and early psychiatric rehabilitation services that incorporate mental health, substance abuse, and medical health care service delivery can ameliorate the burden of disease, reduce symptomatology, increase employment and productivity, and improve the overall sense of well-being.

PSYCHIATRIC REHABILITATION IN THE MEDICAL SETTING

People living with SMI have an overall poorer health condition than the general population, with a higher prevalence of common conditions including cardiovascular disease, metabolic syndrome, and HIV. They are also at a higher risk of comorbid substance use disorders (SUD), most commonly nicotine use, which may further precipitate, worsen, or exacerbate medical issues. Persons living with schizophrenia are more likely to suffer from obesity, and treatment with antipsychotics further increases this risk. In addition, lifestyle and social factors such as poor access to healthcare services, lack of education or access to nutritious food, financial constraints, and lack of exercise can all contribute to poorer overall health. Inflammation may also play a role in elevating risk factors for chronic medical conditions. Moreover, by facing stigma in their personal lives, persons with SMI may be dismissed by healthcare providers who attribute their somatic complaints as psychiatric in nature. Persons with psychiatric disabilities can face barriers in advocating for themselves and communicating effectively with their health providers. General poor health among individuals with psychiatric disabilities negatively impacts their ability to pursue education, gain employment, become financially independent, and be productive members of their communities.

Psychiatric rehabilitation was previously considered a psychosocial treatment program that focuses on aspects such as social skills training (SST), vocational remediation, and building environmental supports. It is now recognized that many psychiatric disorders need to be conceptualized as chronic conditions with relapses and remissions; hence, psychiatric rehabilitation has also evolved to include psychosocial as well as medical approaches. Persons with SMI may struggle with treatment adherence which can lead to exacerbations of underlying psychiatric, substance use, and medical disorders. As a result, a major focus of many current psychiatric rehabilitation approaches is to facilitate adherence to medical treatments and medication regimens as a primary focus of psychiatric rehabilitation. This is typically achieved by using advanced therapeutic and behavioral techniques. Motivational interviewing is also often used to elicit a patient’s ambivalence relating to taking medications as recommended. With growing technologic innovations, the integration of electronic reminders and innovative smartphone solutions such as mobile apps have also expanded. An emphasis on cognitive and behavioral training with regard to prospective memory, or “remembering to remember,” an intended action is emphasized to promote medication adherence. Further, it is well accepted that medication compliance is likely going to be higher with concomitant efforts to build a supportive environment by preparing the individual to deal with environmental stressors and promoting a physically and emotionally healthy lifestyle. For example, there is a bidirectional relationship between diabetes and depression, with depressed patients having increased challenges to comply with self-management of their diabetes (glucose monitoring, taking medications, following strict dietary recommendations, and exercising, among others). Collaborative care models shared medical appointment models, and cognitive-behavioral therapy (CBT) for adherence and depression (CBT-AD) are available strategies to assist patients with diabetes to improve their medical care management. Integrating approaches to care requires the coordination of several health care disciplines.

UNIFYING CONCEPTS OF PSYCHIATRIC REHABILITATION AND TREATMENT

Successful approaches to psychiatric rehabilitation involve the integration of medical, psychological, and psychosocial approaches. As described below, psychiatric rehabilitation is more person-centered and needs to be tailored to individual needs and goals. Collaboration between rehabilitation team members and patients is vital as this enhances therapeutic alliance, partnership, autonomy, and shared decision-making on behalf of providers, patients, and their families. This partnership in shared decision-making takes into consideration the patients' and their families' specific needs, as well as their cultural beliefs and values. This individualized approach to treatment also emphasizes targeted goals that are unique for an individual patient with the appropriate combination of therapies that are specific to the person's needs.

It is of essential importance for both the patient and the treatment team to (a) identify targeted short-term and long-term goals that are operational and obtainable; (b) plan a sequence of interventions and support systems that are appropriate to an individual's stage of recovery; (c) build in social and environmental supports, stress management, coping strategies, and other continuing therapies that will reduce the chances of relapse; and to (d) continue a collaborative treatment milieu between the patient, family, and care providers that provide the greatest opportunity for mutual respect and decision making. Collaborating with the patient as an active participant in their rehabilitation journey means that treatment becomes educational as well as treatment-focused.

PSYCHIATRIC REHABILITATION AND INCORPORATION OF PERSON-CENTERED RECOVERY MODELS

In the United States, the New Freedom Commission on Mental Health (2003) recognized the voices of the recovery movement and mandated that recovery be consolidated into all facets of mental health policy. The person-centered approach was cited as a vital aspect of recovery, as defined by the U.S. Substance Abuse and Mental Health Services Administration (Department of Health and Human Services, 2004). Collaborative, person-centered approaches to care planning are increasingly recognized as instrumental in accelerating personalized recovery goals. Person-centered care core principles include the development of respectful relationships between providers and people experiencing SMI; the integration of the person in identifying individualized goals; the significance of involving personal, family, and community-based resources to reach those goals; a recognition of responsible risk-taking as a necessary part of growth and recovery; and a focus on acknowledging the strengths and capacities of the person as an essential step to recovery.

This integrative method requires a multidisciplinary or interdisciplinary team approach. The person-centered approach builds upon this by actively engaging the patient and their family in shared decision-making to cocreate an action plan toward recovery such as medication use and incorporates their personally relevant goals and unique psychosocial circumstances into the formulation of treatment plans. Family involvement may especially be important for individuals from diverse cultural backgrounds who value family input. Family-focused recovery may play a key role for some people living with mental illness, as their individual and family needs may be better understood by the family as a group and incorporated into the recovery plan. Encouraging the individual to actively engage in recovery planning can help them develop the confidence and skills needed to regain control over their lives.

The person-centered approach is in line with a more optimistic vision for the long-term recovery of the seriously mentally ill as it offers opportunities to generate higher levels of motivation in patients. These progressive changes in ideology have mostly been championed through the ongoing leadership of federal, state, and local agencies. Many training programs are more cognizant about teaching emerging healthcare practitioners about the person-centered, recovery-oriented mode of service delivery, which is expected to continue to intensify and develop in the coming years given that this approach is expected to strengthen the therapeutic alliance, promote greater motivation and collaboration, and achieve better outcomes. It is vital to follow the same person-centered approaches with all patients, with a focus on valued outcomes, while also tracking and measuring outcomes of success of services provided over time.

DEFINING RECOVERY

Outcomes of psychiatric rehabilitation can vary given the heterogeneity of SMI and the wide variety of services available. Reaching recovery goals can hence vary based on the patient’s own targets, as well as personal attributes, social environment, continuity and quality of treatment, and subjective experience.

Sustained recovery is a long-term process and refers to long-term abstinence or remission of symptoms and return to normal psychosocial functioning. Recovery allows persons living with SMI to improve their overall health and wellness, lead self-fulfilled lives, and empower themselves to reach their full potential as contributing members of their community. It can be measured by a set of processes, stages, and outcomes that are based on mediating and moderating variables that facilitate its attainment and maintenance. That is, recovery is an outcome of the process of recovering.

Hope, destigmatization, empowerment, self-acceptance, insight, awareness, collaboration with professionals, sense of autonomy and self-control, as well as participation in self-help and patient-run programs, are all attributes that may help individuals progress toward recovery. Subjective experiences and attributes mediate the process leading to recovery and also sustain recovery. For example, a treatment or self-help program may engender a feeling of empowerment, which can be instrumental in motivating a person to sustain treatment and rehabilitation until recovery can be achieved. Once recovery has been achieved, empowerment may be even more firmly experienced through continuously improved functioning and cognitive, social, and emotional stability.

Many factors can influence recovery, including overcoming and managing one’s symptoms, choosing healthy outlets and appropriate treatment to support one’s general well-being, having access to healthcare and stable housing, obtaining purpose and meaning from one’s life, including having enjoyable daily activities, financial independence, and contribution to one’s community, and having supportive relationships. Rehabilitation services must remain flexible in the services provided to individuals and continue to advocate for their patients, reduce health disparities, improve access to care, and ensure that care delivery supports their patient’s sustained recovery.

Similarly, having hope for future improvements in quality of life may be reinforced by the enthusiasm and collaborative alliance embedded in a therapeutic relationship. Many patients who have written first-person accounts of their recovery have pointed to the importance of a long-term relationship with a practitioner who refused to give up hope for the patient’s ultimate improvement and recovery. Hope can serve to fuel motivation for change and for active participation in clinical services or self-help, which are stepping stones toward recovery. Subjective experiences of empowerment and hope continue to grow after a state of recovery is attained, contributing to further strengthening and protecting individuals from stress-induced relapse. Encouragement and hope also enhance efficacy expectations that empower the individual to assume greater control and responsibility for their lives.

EMPIRICALLY BASED MODELS OF PSYCHIATRIC REHABILITATION

Psychiatric rehabilitation has evolved from being considered an add-on treatment if conventional measures with psychotropic medications or psychotherapy fail, to a way to “fill the day” with activities for persons with SMI, with descriptions of the token economy of other reinforcement programs. Psychiatric rehabilitation has moved to become an evidence-based, highly specialized specialty within psychiatry, offering structured interventions, assessment tools, and defined target goals in order to reach desired individualized outcomes in persons with SMI. Evidence-based practice recommendations are based on well-designed research studies that show effective interventions in improving illness course and outcome.

Over 50 years of intervention research has led to the expectation that even patients with the most severe mental disabilities can improve their overall daily functioning when provided with empirically validated, evidence-based behavioral interventions. Some examples of evidence-based practices (EBPs) for the recovery of SMI include assertive community treatment (ACT), cognitive-behavioral therapy for psychosis (CBTp), cognitive rehabilitation, family psychoeducation, illness self-management training, SST, and supported employment, among others. Some other forms of interventions are gaining more evidence, though not yet evidence-based, such as expressive and art therapy in patients with schizophrenia.

In 1991, the Americans with Disabilities Act was signed into law, requiring equal employment opportunities and reasonable employer accommodations for workers with mental illness. Moreover, fulfillment of interventions for therapeutic improvement has been facilitated by political advocacy, destigmatization of mental illness, and self-help activities of consumer organizations such as the National Alliance for the Mentally Ill, the Schizophrenia Fellowship, and networks of mental health clients.

Recovery is now more than ever a realistic goal for many persons with mental disabilities with enduring biologic vulnerabilities such as schizophrenia; however, it is necessary that the full range of evidence-based biopsychosocial services be available in a continuous, accessible, comprehensive, coordinated, person-centered, culturally appropriate, and competently delivered system of care. Despite the fact that the field of psychiatry produces a wide array of evidence-based treatments and services, public and private funding has eroded for organizing and delivering these services with high fidelity. With that in mind, psychiatric rehabilitation is confronted with the major challenge of fiscal limitations, inadequate resources, and poor coordination. Two core principles of psychosocial rehabilitation include: (i) all resources allocated to the recovery process should be coordinated and integrated with the goal of optimizing treatment, and (ii) a consistent therapeutic team should provide continuity of care, taking into account the person's life context.

ACCESSIBILITY TO EVIDENCE-BASED TREATMENT AND SERVICES

Despite the growing evidence available on psychiatric rehabilitation, there is a wide gap between the scientific literature and the design and practice of many psychiatric rehabilitation programs. This means that evidence-based psychiatry rehabilitation interventions are not widely available, nor easily accessible, in day-to-day clinical practice. Funding opportunities may be limited and can hinder the development of these programs. In addition, there are many differing opinions among psychiatry professionals in relation to views on recovery and treatment goals, as well as patient and family expectations, which can further prevent the restructuring of existing programs and the new development of quality-based programs.

One of the biggest hurdles in implementing evidence-based rehabilitation programs and improving access is organizational level challenges, competition in allocating funds to resources, a lack of administrative and financial integration of state and local community mental health and substance abuse services, and, at times, a limited understanding of mental health needs beyond symptom control. While the

United States revolutionized the concept of evidence-based rehabilitation services, it appears that the existing services fall behind other high-income countries. The consequences of those system-wide problems have meant that too few individuals with mental disabilities are receiving state-of-the-art services and too many thousands end up living on the city streets or in jails and prisons.

Implementing recovery-oriented and person-centered programs can take a long time, requiring the commitment of providers to not only practice but also train and educate others entering the field. Conducting continuous measures and tracking outcomes is vital for quality measures. Engaging learning environments, such as training events and user-friendly practical guidelines that support clinical decision-making, can be helpful in training future rehabilitation providers. An example of this includes the development of a free evidence-based resource center from the Substance Abuse and Mental Health Services Administration (SAMHSA), which provides toolkits available on their website: https://www.samhsa.gov/resource-search/ebp. These efforts, such as the project implementing EBPs, have mostly been sponsored by federal, state, and foundation agencies. Standardized guidelines for the application of the evidence-based treatments are provided in workbooks that are included in the toolkits focused on improving the functioning, quality of life, and recovery of patients, as are training materials and procedures to provide means for assuring high-quality treatments.

The American Psychiatric Association (APA) has also provided free-access training materials for psychiatrists on basic understanding of recovery from SMI and SUDs and recovery-oriented care. These training videos can be accessed at https://www.psychiatry.org/psychiatrists/practice/professional-interests/recovery-oriented-care.

The ultimate purpose of any available EBP toolkit—“to promote the implementation of evidence-based psychosocial treatments in routine mental health settings’’—is a goal that has great potential to improve the lives of people diagnosed with serious mental health disorders. Access to EBP Toolkit helps clinicians, agencies, and organizations to be educated in facilitating best practices in the training, implementation, and monitoring of evidence-based treatments for mental health. However, implementing and directly monitoring fidelity to the model can be a costly and time-consuming process. A balance needs to be achieved between monitoring the accurate implementation of each EBP while promoting its dissemination in a way that allows institutions to tailor the innovation to fit the specific function, mission, patient population, local resources, and other contingencies they face. Continuous quality improvement has been used effectively as an approach to implement EBPs in community-based settings.

At the level of the community agency implementing EBPs, specific expectations for clinician performance and ongoing program evaluation, as well as a culture change, are required to implement and sustain the provision of EBPs. Thus, evaluating the cost-effectiveness of the EBP toolkits using a longitudinal time frame and a methodology that captures the organizational characteristics and the total expenditures to the systems of care in using the EBPs for the patients within their catchment areas is essential to the long-term viability of these programs.

With improved and more reliable diagnostic methods, as well as symptom and functional assessment and an empirically validated definition of recovery, the population and needs of patients requiring psychiatric rehabilitation are becoming better defined. Methods for determining readiness for rehabilitation also have been validated, enabling administrators and providers to be more efficient in allocating rehabilitation programs and resources. Because of its growing technology and its conceptual and empirical base, psychiatric rehabilitation has attracted an interdisciplinary group of practitioners and investigators whose work has contributed to its evolving scope and stature.

The challenges to community-based care for persons with serious and chronic mental illness have been met partially by new initiatives at the clinical and policy-making levels. New interventions have been designed and validated, such as psychoeducational and behavioral family therapies, SST, CBT for medication-resistant psychotic symptoms, intensive case management models for assuring continuing care, supported housing, and employment services. At the level of psychiatric politics, consumers and families have become empowered by the advent of peer-directed self-help organizations; the use of social media for the education of the public about mental health disorders and the possibilities for recovery, including noted political, sports, and entertainment figures going public with their experience of living with mental health disorders and successful treatment; legislative lobbying, assertive advocacy, peer support, and family involvement in treatment; efforts to gain parity in funding for mental health services; and the Americans with Disabilities Act regulations.

The COVID-19 pandemic has brought on many challenges in mental health care delivery. Provision of rehabilitative care for persons living with SMI is vital to monitor COVID-19–related recovery outcomes, including the accessibility of services, healthcare disparities, rate of relapses and symptoms exacerbations, rates of telehealth encounters, patient and family satisfaction with the quality of care, functional, social and vocational outcomes.

REHABILITATION IN THE PRODROMAL, ACUTE, AND STABLE PHASES OF SEVERE MENTAL ILLNESS

Effective treatment and rehabilitation of persons with SMI require clinical services that are linked to the phase of their disorder. For most stress-related, recurring, and disabling mental disorders, the following phases are relevant for practitioners' initiating and/or adjusting their treatment:

- Motivation and engagement in treatment
- Prodromal
- Acute
- Stabilizing
- Stable
- Recovery
- Refractory

There is broad-based variability in the manner in which individuals experience a particular phase in that no two individuals experience any phase in the same way. Individual differences influence each phase in terms of severity, duration, degree of disability, resolution, course, and boundaries. The following cases depict two persons in the stabilizing phase. This serves to exemplify the need to adapt treatments to individual diversity within a particular phase of illness.

Mr. B, a 19-year-old male, was admitted for involuntary psychiatric hospitalization after he locked himself in his room for three days and refused to eat because he was afraid someone was poisoning him. His father would hear him talking and yelling to himself incoherently. When his father was able to enter his room, he found the room to be in disarray with the windows covered with newspapers, as Mr. B thought he was being watched. After six days of stabilization, he was released from the hospital and returned home to his father but was still very anxious about being alone, had difficulties with hearing voices telling him he was “cursed,” and was extremely concerned that his movements were being monitored. Because he was resistant to oral medications (he thought they might be meddled with and might poison him), he was prescribed a long-acting injectable antipsychotic, which helped to stabilize his symptoms. Upon returning to school, he completed a semester taking a reduced course load and living at home. However, during the next semester, he insisted on taking a full course load and started to smoke cannabis and other drugs to help manage the resultant distress. He began to feel on edge and felt that others in his class were judging him critically and talking about him. His grades started to deteriorate, and Mr. B developed another acute psychotic episode. He was readmitted to the psychiatric hospital with extreme paranoia and had to again drop out of school. This time, treatment became integrative, and, in addition to adjusting his psychiatric medications, he received supportive psychotherapy and was engaged in several cognitive-behavioral and skills-building training that provided him with ongoing support and structure. He learned to be more assertive and comfortable in his interactions through assertiveness, social skills training, and relaxation/stress management techniques, which greatly helped his coping skills. Mr. B eventually returned to school, earned his degree, and is currently in a supportive work program.

Ms. A, age 24 years, had persistent depressive symptoms that increasingly became worse, particularly when she graduated from medical school and accepted a role as an internal medicine resident in a busy county hospital. She became increasingly convinced that she was an unqualified physician despite graduating with honors and felt hopeless and worthless. She consistently felt that she was subpar compared to her peers, which further perpetuated her sadness, hopelessness, and her sense of failure. As a physician intern, she struggled to complete necessary clinical duties due to fatigue and impaired attention and concentration. On top of existing insomnia with difficulty falling and staying asleep, transitioning to night shifts further impaired her sleep. Her husband noticed that she was becoming more isolative, tearful, and was not interested in joining her residency group activities or completing house chores. Her family started to become increasingly concerned about her lack of communication and began to worry about her engaging in self-harm behaviors. Ms. A. started to feel that she was a burden to others and began to think about death. On an evening shift, she encountered a verbally hostile patient who demanded to transfer his care to a different provider. She then became acutely suicidal and was subsequently admitted to an inpatient psychiatry unit for treatment and stabilization of her severe depression. Following discharge from the hospital four days later, she was placed on a combination of antidepressant and antipsychotic medications and immediately referred to a psychotherapist who focused on CBT techniques. It took several weeks before she was adequately stabilized to engage in interventions for major depression and suicide prevention. In addition, Ms. A and her husband started couples' therapy with another therapist that helped the husband better understand her illness, supported her behavioral homework, and helped the couple to communicate better. Ms. A, with the help of her supportive residency program director, is now back to work, with an adjusted on-call schedule, and is involved in maintenance therapy where she sees her psychotherapist weekly and continues to adhere to all her mental health treatment.

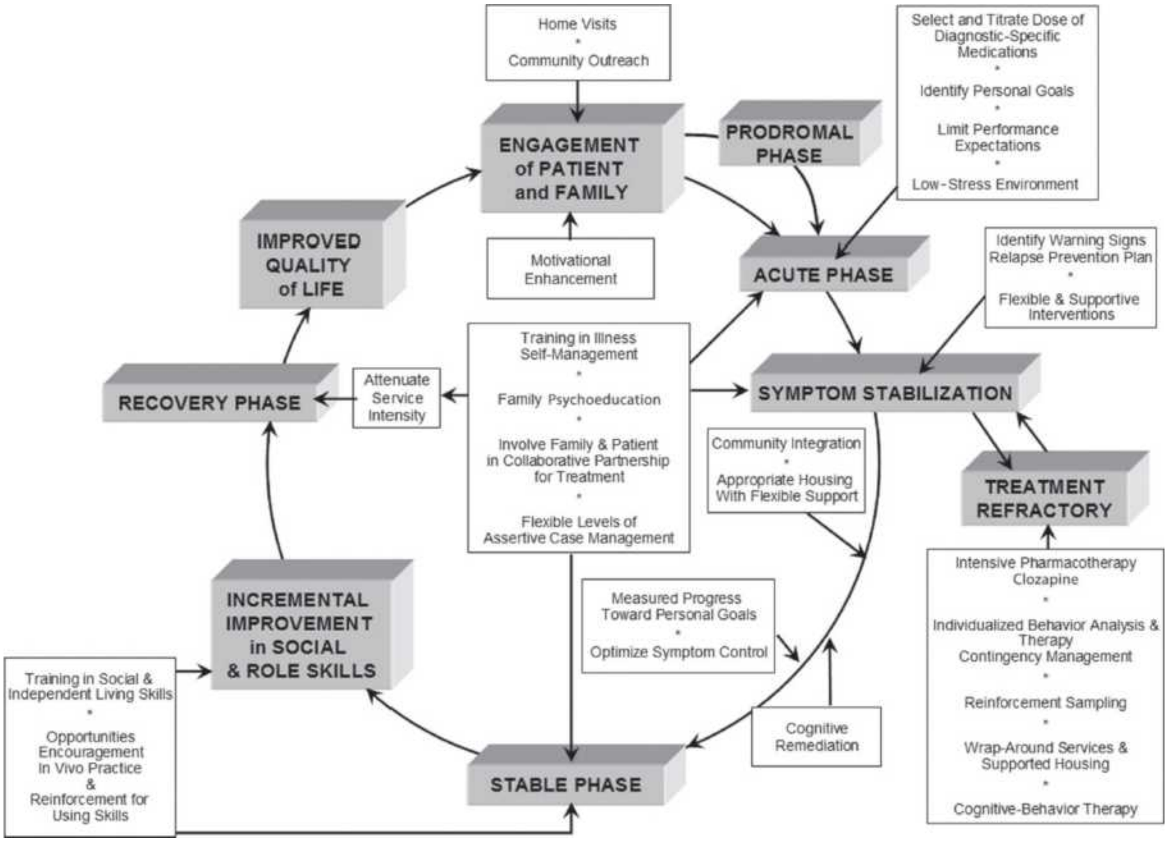
Home Visits
•
Community Outreach
ENGAGEMENT of PATIENT and FAMILY
PRODROMAL PHASE
Select and Titrate Dose of Diagnostic-Specific Medications
•
Identify Personal Goals
•
Limit Performance Expectations
•
Low-Stress Environment
IMPROVED QUALITY of LIFE
Motivational Enhancement
ACUTE PHASE
Identify Warning Signs Relapse Prevention Plan
•
Flexible & Supportive Interventions
Training in Illness Self-Management
•
Family Psychoeducation
•
Involve Family & Patient in Collaborative Partnership for Treatment
•
Flexible Levels of Assertive Case Management
RECOVERY PHASE
Attenuate Service Intensity
SYMPTOM STABILIZATION
Community Integration
•
Appropriate Housing With Flexible Support
TREATMENT REFRACTORY
Intensive Pharmacotherapy Clozapine
•
Individualized Behavior Analysis & Therapy Contingency Management
•
Reinforcement Sampling
•
Wrap-Around Services & Supported Housing
•
Cognitive-Behavior Therapy
INCREMENTAL IMPROVEMENT in SOCIAL & ROLE SKILLS
Measured Progress Toward Personal Goals
•
Optimize Symptom Control
Training in Social & Independent Living Skills
•
Opportunities Encouragement In Vivo Practice & Reinforcement for Using Skills
Cognitive Remediation
STABLE PHASE

FIGURE 56.6–1. Multidimensional framework for phase-linked treatment and rehabilitation.

These two examples illustrate the need to conceptualize the treatment of these two individuals into several components (a) stabilization; (b) dealing with ongoing acutely psychotic, mood, or anxiety symptoms; (c) using the therapeutic alliance to provide the maximum motivation for treatment; (d) using integrative therapeutic support to augment and strengthen the possibility of positive treatment results; (e) ongoing assessment and strategies aimed toward relapse prevention; and (f) reinforcement and maintenance of gains.

The framework for phase-linked treatment and rehabilitation is depicted in Figure 56.6–1. It should be noted that EBPs are the backbone of services generally found to be useful for patients as they traverse the various phases of their disorder. There is considerable variability in how fast each individual moves from one phase to another, and progression may not be linear in fact. Some may even skip a phase, becoming stable within one week of an acute relapse, without the gradual stabilization of symptoms and functioning that is more frequently the case. People also move back and forth between phases as stressors impinge on their vulnerability, and the individual has not compiled sufficient protective factors.

Prodromal Phase of SMI

Early identification, engagement, and treatment are the mainstays of this phase, which may be associated with individuals who are at risk of their first episode of SMI or are experiencing the warning signs of a recurrence of their disorder. Preventive interventions for young people at high risk of developing a psychotic illness remain in the early stages of development and need further investigation. Psychosocial and pharmacologic interventions have not been empirically documented as fully effective in averting schizophrenia or other psychotic disorders when provided to young people experiencing prodromal signs and symptoms. The sensitivity and specificity of identifying the early signs of a developing psychosis have not been fully validated; however, recent review studies suggest that “basic symptoms” (defined as ‘subtle, subclinical disturbances in one’s experience of oneself and the world that prospectively predict first episodes of psychosis’) can improve the prediction of psychotic relapse. Further prospective studies are still needed to compare basic symptoms and early signs as predictors of psychotic relapse. Moreover, in the digital era, some studies have focused on social media activity (primarily via patterns of linguistic and behavioral changes) and the use of machine-learning models to identify and predict psychotic relapse in young individuals with recent-onset psychosis and imminent risk of hospitalization.

Youth who are possibly moving toward their first psychotic episode may experience sleep problems, strange, unusual, unwanted, and intrusive thoughts, depression, irritability, hypomania, or anxiety. Other prodromal signs and symptoms include ideas of reference, paranoid worries and fears, significant social and emotional withdrawal, suspiciousness, apathy and loss of drive, disturbances of attention and concentration, functional deterioration, odd behavior, perceptual abnormalities, and derealization. These young people almost always refuse to accept professional intervention, preferring to think that their condition is transient and will dissipate without psychiatric treatment. Moreover, the stigma of mental illness is a major deterrent to obtaining treatment.

For these individuals and their families, a patient-centered, nonconfrontational stance is desirable. Because it is not even certain that they are moving to an acute psychotic episode, it is necessary to monitor the presenting symptoms and how they might be interfering with the individual's current life. Then, depending on the specific problems that are interfering with the individual's reaching their goals, the psychiatrist points out that treatment may very well help to overcome the obstacles to progress toward these goals. This approach is known as motivational interviewing and represents a strategy for engaging patients in early treatment. It may be necessary for the practitioner to ally with the family so that the latter can join in the motivational enterprise. The treatment approaches to this early developmental phase of schizophrenia are delineated in Table 56.6–1.

Symptoms Prodromal to a Relapse. Because each psychotic episode places the individual at greater risk of prolonged psychosis, disability, and refractoriness to treatment, prevention of relapse has been assigned a high priority. It is often the case that the individual with recurrent illness does not recognize or acknowledge their symptoms; in this case, it is best to simply work on building a relationship with the patient that will eventually pay dividends for active treatment once trust has been achieved or restored.

Table 56.6–1.
Interventions for the Prodromal Phase of Schizophrenia in Young Persons Not Yet Diagnosed with the Disorder

1. A psychoeducational approach is directed to the patient, family, other caregivers, primary care physicians, school counselors, and staff at college counseling centers. The psychoeducation must be keyed to the particular needs, readiness, and role of the individuals who are in a supportive relationship with the individual who is experiencing the prodromata to a serious mental illness. Outreach and liaison with “gatekeepers” to mental health services—primary care physicians, clergy, schoolteachers, and school counselors—is important for mental health facilities and solo practitioners to mount a successful effort in identifying young individuals at high risk of a serious mental disorder. There should be ongoing liaison, education, and consultation with the family and referring professionals and agencies.
2. Nonstigmatizing, age-appropriate, culturally congruent services should be offered in a youth-friendly, nonclinical environment.
3. Specific treatment for symptoms and behavioral problems that are interfering with the young person’s goals and functioning should be adopted:
    a. Motivational interviewing.
    b. Stress management.
    c. Cognitive-behavioral therapy or interpersonal therapy for anxiety and depression.
    d. Cognitive-behavioral therapy for prepsychotic symptoms.
    e. Supportive relationship that permits open discussion of the young person’s concerns, worries, symptoms, and progress toward goals.
    f. Antipsychotic medication at low dose for symptoms that do not respond to psychosocial interventions and continue to impede the young person’s functioning.
4. There should be frequent re-evaluation of the patient’s clinical status to determine whether a schizophrenia prodrome is actually operating.

If possible, prior to a recurrence of psychosis, the practitioner should provide illness management services that include training in recognizing and monitoring early warning signs of relapse together with developing an emergency plan for preventing full-blown relapse. This educational effort is important because recurrent episodes of schizophrenia, major depression, bipolar disorder, and other SMIs carry a poorer prognosis and a higher risk of:

- More frequent acute illness episodes in the future
- Higher likelihood of nonadherence to treatment
- Greater difficulty treating each recurrent episode
- More serious disruption of social, occupational, educational, and family roles
- Greater likelihood of stigma and family isolation
- More frequent involvement of police, involuntary admission to hospital, homelessness, and legal problems or incarceration.
- Increased financial and emotional burden for the family.

Early Intervention and Relapse Prevention. One component of illness management is educating patients and their families to recognize the prodromal signs and symptoms of relapse, monitor them, bring them to the attention of the mental health provider, and follow through with their emergency plan to avert relapse. Early intervention during the initial onset of a psychiatric disorder should be implemented to minimize the risk of developing chronic disability.

An acute episode of mental illness, including relapse for persons who have had previous episodes, frequently follows prodromal signs and symptoms for patients with mood and psychotic disorders. Regardless of where a patient resides during an acute episode of illness—crisis home, family home, inpatient unit, or partial hospitalization—the major task is to mitigate overstimulation and provide a safe, secure, supervised, and predictable environment. This is accomplished by offering clear expectations for daily activities, a posted schedule of daily events, discharge criteria, standards for each privilege level, and a structured routine and treatment environment. Staff experienced in working with acutely ill patients offer individualized, clear expectations for behavior and low task demands that are not unnecessarily burdensome but are flexible in terms of how much social interaction is prescribed. In addition, the staff should offer low-key planned and scheduled activities, orientation, and guidance on the rules of the facility by posting visual aids on the bulletin board and whiteboards. One staff member should be designated to help each patient conduct problem-solving and offer supportive therapy at all hours.

In this intervention phase, pharmacologic and behavioral treatments of acutely ill patients are intended to stabilize or eliminate symptoms and improve cognitive status so that the patient can move toward focusing on the development of skills and utilizing the social and professional supports that will help the individual move into the next phase of treatment. Close and prompt postdischarge monitoring is vital. Discharge planning should include not only the assessment of psychiatric symptoms but also consist of access to safe housing, outpatient medical and mental health care, employment, community resources, and social support.

Most patients in the acute phase can begin learning illness management skills, including understanding their symptoms, describing their mental illness and treatment plan, and creating strategies for minimizing its impact on their daily functioning and reducing rehospitalizations. During the acute treatment phase, it is essential to customize skills training and support to symptomatic and cognitively impaired individuals; thus, in educating, it is very important to use strategies such as motivational enhancement, repetition, responsive reading, modeling, role-playing, and problem-solving. It is also necessary to be flexible in using relevant and culturally appropriate communication in educating patients about the value of psychotropic medication use.

Continuity of Care From Hospital to Community. Continuity of care with mental health and other medical providers is essential for improving provider–patient relationships and reducing fragmentation of care. It is associated with greater patient satisfaction, higher adherence to treatment, and reduced hospital service utilization. Due to hospital stays becoming shorter than in the past, patients may be discharged with ongoing active symptoms and may not be sufficiently optimized. Despite efforts by hospital providers to ensure continuity of care and create a postdischarge plan, numerous studies have confirmed the frequent lack of follow-up when transitioning from inpatient to outpatient care.

Patients struggling with persistent acute, distressing symptoms postdischarge from inpatient care may face obstacles when attempting to engage in outpatient services. Cognitive deficits are common in SMI and interfere with the patient taking the initiative or planning to set up a follow-up appointment. Psychoeducation with frank and honest discussions should be conducted with the family involved, as at times, the patient’s support system may believe that definitive treatment has been completed in the hospital, fail to encourage follow-up services, or even suggest discontinuing medications due to lack of psychoeducation or understanding of recovery, as well as the presence of stigma toward mental health care services.

Case managers can help facilitate this transition by coordinating a community mental health liaison to visit the patient while hospitalized to introduce themselves, discuss services, and answer any questions. An outpatient case manager or representative can also accompany the patient to visit their community mental health center for follow-up. This familiarization can reduce anxiety, as the patient will no longer need to repeat their histories which may often be related to traumatic experiences during past acute episodes, and increase the likelihood of follow-up. Providing transportation from the patient's home directly to the facility for their appointments increases attendance and participation.

Training programs, such as the Community Re-Entry Module, empower patients transitioning out of the inpatient psychiatry setting to learn skills and attitudes that support continuity of care. Some of the skills taught include: connecting with the community, planning a daily schedule, making and keeping appointments, and managing medications. These skills allow the patient to take responsibility for making the connection with outpatient services with the purpose of enhancing their quality of life by functioning more autonomously.

In this continuity of care program, family members are involved in the training as much as their schedules permit. A randomized, controlled study comparing the continuity in reaching outpatient services for inpatients who participated in the Community Re-Entry Module versus those who received standard discharge planning by a social worker showed that more than 80% of the individuals participating in the module succeeded in getting to an outpatient facility after leaving the hospital, compared to only 37% of those receiving a standard discharge. This study confirms the capabilities of patients with schizophrenia to take responsibility for maintaining their treatment.

During the acute phase of treatment, some common patient responsibilities include such functional and symptomatic outcomes as engaging in independent self-care and personal hygiene without prompting, socializing with peers without hostility or social withdrawal, voluntarily accepting medication, and learning the reasons why medications are being prescribed—both for dampening the current symptoms that led to the hospitalization and for prophylactic purposes to reduce risk of relapse. The patient and physician review the daily behavioral progress, which is the cornerstone of motivation, hope, and shared decision-making. Most inpatient units have “level systems” that are used to provide incentives to patients to take more responsibility for their self-care, reliability, participation, and behavior. Level systems can be customized to meet the individual needs of specific patients by using tokens, credits, or points superimposed on the level system. Another way to motivate patients to reach their goals and discharge readiness during the acute phase of illness is to use therapeutic contracts. Therapeutic contracts invite collaboration in that both parties describe their goals and responsibilities for the inpatient period of treatment.

The psychiatrist and other staff must ensure that the patient is familiar with and has a plan for reaching follow-up care as an outpatient. As discharge approaches, the responsible psychiatrist should reach out to the family and invite them to attend a psychoeducation meeting with the psychiatrist present to answer questions and give feedback on how their relative is progressing. Having the family present with the psychiatrist and treatment team during a psychoeducational meeting in the acute phase of illness provides opportunities to acquire another source of information about the patient’s illness and disability. Family members may be more reliable assessors, and many of them keep written records or diaries of the course of illness, medications, and responsiveness and side effects of the illness. Home visits seldom occur in the current era; however, this can provide considerable information about the quality of life of the patient in their current environment prior to hospitalization, thereby enabling more accuracy and clinical acumen to be brought to bear on the diagnosis, treatment experiences, and social functioning of the patient.

Stabilization Phase

Stabilization of symptoms and monitoring of cognitive deficits almost always continue after acute discharge into the stabilization phase. Most of the considerations noted previously in the acute phase are continued in the outpatient setting; hence, the treating psychiatrist has the responsibility for ensuring that continuity of care is updated.

Once the patient is discharged from the inpatient setting, it is likely that medications are to be continued. Monitoring symptoms using quantitative scales such as the Brief Psychiatric Rating Scale (BPRS), the Positive and Negative Syndrome Scale (PANSS), or the Target Symptom Scales during regular appointments can help the prescribing practitioner make informed decisions regarding dosing or type of medication changes. Responsiveness to medication is also affected by the emotional climate in the family, which needs to be considered by the prescribing psychiatrist. The use of a quantitative rating scale for the informed use of medications is illustrated by the stabilization of the patient in the following case, a young woman with schizophrenia.

Ms. C, a 30-year-old woman, had three previous hospitalizations with relapse of acute psychosis, exhibiting paranoid delusions and auditory hallucinations. These symptoms were stabilized with antipsychotic medication. The positive symptoms of schizophrenia were much improved after her last hospitalization; however, she reported having “no interest or motivation” and no desire to interact with others. During the first month of outpatient visits, her psychiatrist gradually titrated up her medication, with evidence of stability in that her delusions and hallucinations remitted. However, serial monitoring with the PANSS indicated worsening in her negative symptoms of schizophrenia. A family meeting revealed that the family did not comprehend the effect that negative symptoms could have on Ms. C’s daily functioning despite the lack of active psychosis. As a result, they criticized Ms. C for not returning to work, leading to frequent heated arguments. The psychiatrist scheduled a series of weekly and then biweekly educational sessions for Ms. C and her parents to explore the family structure, expectations of treatment and provide psychoeducation about Ms. C’s illness and treatment. This intervention gradually lowered the “emotional temperature” of the family household, resulting in Ms. C receiving the assistance that she needed in order to engage in vocational rehabilitation.

During the stabilization phase, learning illness management skills, including gaining a more solid understanding of the role of medication, can be very useful. This includes skills such as how to communicate assertively with the psychiatrist about side effects or lack of symptom reduction, as well as how to communicate with friends and family about their illness. Supportive treatment sessions are provided on a weekly or biweekly basis with extended invitations to the family or other caregivers to participate in the training sessions for illness management focusing on proper medication use and relapse prevention.

Illness self-management education involves teaching individuals with SMI about the nature and course of their medical and psychiatric conditions, engaging them in motivational interviewing to explore implementing lifestyle choices that can promote better illness management, instructing them how to monitor their symptoms, how to self-administer medications, and preparing them on signs and symptoms that warrant contacting their treatment providers or seek emergency treatment. Behavioral skills training is useful to address nonadherence to help incorporate taking treatment as prescribed into the individual's day-to-day routine. Developing a relapse prevention plan can be useful to provide a physical, concrete reminder for the patient to monitor their symptoms. Helping the patients develop coping strategies for persistent symptoms can reduce emotional distress.

Collaborative Relationship With Patient. The development of a collaborative and therapeutic relationship is essential if the patient is to remain in long-term treatment and make progress toward their personal goals. Personal, relevant goals are motivating for the patient's participation in the phases of stabilization, stable, and recovery. Depending on the needs and desires of the patient, these goals can be connected to areas of functioning such as self-management, friendship, and intimacy, work or school, family relations, housing, recreation, financial, and spiritual needs. The psychiatrist and other mental health professionals working with the patient must help the patient set incremental, stepwise goals that will bring the patient closer to their overall personal goals.

As symptoms and cognitive impairments stabilize, the patient may benefit from weekly SST that gives the patient opportunities to learn from social modeling, role-play practice, and homework assignments to facilitate achieving short-term goals that are on the pathway to overall goals. For example, a man in his 30s who has reached the recovery phase of schizophrenia may have a personally relevant goal to find an apartment and live independently. With the guidance of the skills trainer, he may proceed through a series of steps, including the following:

- Learning and demonstrating the ability to engage in medication self-management and other Independent Living Scales (ILS) required for living in an unsupervised residence.
- Through problem-solving training, identifying various ways to locate apartments that are for rent, and generating options.
- Evaluating the pros and cons of each identified apartment—for example, costs, location, and availability of public transportation.
- Practicing in the group how to ask questions assertively about the apartment with the property manager or owner, then applying the skills practiced to real-life situations.
- Reviewing the advantages, disadvantages, and feasibility of each apartment and making a provisional choice with second and third preferences.
- Role-playing in the group session on how to request a loan from parents to help subsidize the rent, then carrying out the request with parents.
- Role-playing in the group session finalizing and signing a lease agreement with the property manager, then doing it in real life.

Once he settles into an apartment, there are many more related goals, including furnishing the apartment, meeting neighbors, and identifying and engaging in social events available in the community. Once the individual is successfully established in the apartment, he may again set new personal goals such as making friends and finding a part-time job.

Increasing motivation through repeated successful reinforcing experiences can encourage the patient to establish broader functional goals and reinforce therapeutic alliance with the rehabilitation team. As therapeutic relationships strengthen, so does the willingness of patients to confide in their treatment team.

Involving Families in Stabilization. Families play a major role in facilitating or impeding improvement in patients with SMI—including schizophrenia, mood disorders, disabling anxiety, or somatoform disorders. In many instances, family members become caregivers for persons living with psychiatric disabilities. Involving family members can help prevent relapse, encourage treatment adherence, facilitate a supportive home environment, reduce distress, and promote a fitting level of patient independence. Through rehabilitation efforts, families can be included in planning, monitoring, and promoting stabilization. In fact, mobilizing families as partners in the treatment enterprise is relevant for all phases, including stable and recovery periods.

Relatives and other caregivers can be involved in treatment sessions that are organized either for individual families or multiple family groups, with goals limited to psychoeducation about the nature, course, and treatment of the psychiatric condition or, more extensively, teaching skills in communication and problem-solving, coping with persistent symptoms, managing crises, and implementing relapse prevention plans.

In the stabilization phase, patients should be involved in family sessions as their symptoms and cognitive impairments permit. It is essential that family services be fully integrated with other treatment modalities being offered to the patient, such as social and vocational rehabilitation and medication management. Multidisciplinary integration is accomplished most effectively through regular consultations and interactions among all those responsible for the patient's services. Communication among mental health providers can be initiated and led by the psychiatrist through phone calls, case conferences, team meetings, and virtual platforms. Families can also benefit from participation in educational programs offered by their local National Alliance on Mental Illness (NAMI) affiliates that include Family-to-Family seminars and Homefront meetings for families of veterans, as well as family support groups, educational lectures, and annual statewide and national conferences. For the specific individuals living with SMI, NAMI also offers a free Peer-to-Peer educational program to better understand their illness and their recovery process.

Contact with families should continue indefinitely because they often serve as the most important continuous thread of support that weaves together the various phases of illness. It is now well established that patients living with relatives who learn coping skills from behavioral family management have two to three times fewer relapses than patients who are treated with medication and case management alone.

Stable Phase

Patients are considered stable when their symptoms no longer interfere with their functioning in everyday activities. This does not mean that their symptoms are in remission, although they are usually rated as "moderate," "mild," or "very mild" on the BPRS. Some patients living with schizophrenia, for example, are able to participate in work, school, social, and recreational activities in normative settings even when delusions and hallucinations are present on a daily basis. Psychological interventions can displace symptoms, such as when patients are actively engaged in conversations or tasks that shift their attentional capacity from isolated self-preoccupation. Cognitive functioning, such as attention and planning, can also be improved by cognitive remediation (CR) in this phase of treatment.

Incremental steps toward achieving long-term, person-centered goals continue in the stable phase. Progress toward goals can be promoted by participation in training programs that use behavioral learning principles to assure generalization of skills learned in skills-training sessions that take place in clinical settings. The key to generalization is for clinicians, family members, or peers to promote opportunities and encourage, as well as reinforce the use of skills in the natural environment.

Vocational Rehabilitation. Vocational rehabilitation programs exist to bridge the divide between persons with SMI who are interested in working to gain employment. During the stable and recovery phases, patients may be able to begin job-seeking or return to school. Although supported employment (or supported education) is the most popular evidence-based service for vocational rehabilitation, it is not the only occupational goal that is clinically relevant. A wide variety of vocational and educational activities can be compatible with the stable and recovery periods of functioning. Before some patients are ready for work in competitive jobs, they inevitably need to transition through work-related activities short of regular jobs in the community. The Clubhouse Model of Psychosocial Rehabilitation is an example of a model in place for persons living with severe mental illness who engage in a nonclinical, integrated therapeutic working community to achieve recovery by strengthening work-mediated relationships. Studies have found evidence of the efficacy of the Clubhouse Model in promoting employment, reducing hospitalization, and improving quality of life for those with SMI.

At the same time, it is important to recognize the value of supported employment because patients with mental illness who gain jobs in the competitive sector can work half-time or more in competitive lines of work and can gain satisfaction with their renewed or achieved status as workers. Based on direct comparison analysis and network meta-analysis, supported employment and augmented supported employment have been found to be the most effective interventions for obtaining and maintaining employment for those living with SMI. Supported employment provides persons with SMI an opportunity to obtain competitive employment through a rapid job search and retention, whereas traditional vocational rehabilitation programs provide a “training-place” prevocational training and experience in noncompetitive, structured staff-supported settings with other SMI individuals prior to work placement. Individual placement and support model of supported employment (SE-IPS) also includes collaboration with psychiatric services, counseling on benefits, and counseling on job challenges such as disclosure of psychiatric illness, transportation challenges, and meeting job expectations.

However, not all individuals with mental disabilities wish to participate in supported employment. Many persons living with SMI show ambivalence toward obtaining competitive jobs. Approximately half of those with disabling forms of schizophrenia, mood disorders, and personality disorders do not engage in supported employment. Moreover, supported employment is not always associated with increased self-esteem, quality of life, global functioning, or reduced psychiatric symptoms. There are alternative ways to lead a working life, including an array of opportunities that can be viewed as prevocational in nature, such as volunteer opportunities and on-the-job training.

Various work circumstances can lay the groundwork for recovery. A 2018 meta-analysis reviewed 21 SE-IPS programs randomized controlled trials around the world and found that participants in SE-IPS programs were more likely to obtain a competitive job compared with alternative vocational rehabilitation services (mean risk ratio = 2.31 [95% CI 1.99–2.69]). Competitive employment is associated with improvements in quality of life and with increased self-esteem, reduced stress, increased socialization, and reduced stigma. When seeing themselves as self-directed agents contributing purposefully to their own patient-centered work goals, patients can make progress toward independence and self-sufficiency in a variety of ways as they are able to see themselves working toward recovery.

Recovery Phase

Recovery does not mean “cure”; rather, patients in this phase are empowered to live meaningful and satisfying lives, having reached many of the overall goals that bring them into normal community settings and interactions. In this phase, individuals are more autonomous and self-directed in their choice of person-centered goals and taking steps and garnering resources to continue achieving their goals. They are equal partners with their professional providers of services, making shared decisions and serving increasingly as their own case managers. They often assume paying jobs as peer counselors and peer advocates, in addition to joining with mental health professionals as active proponents of public education and political action for decreasing stigma and improving funding for mental health services in their states and communities.

Stigma toward mental health may lead to social withdrawal and the perception of being devalued or discriminated against, which in turn can affect the self-esteem of patients and become a barrier to recovery. Individuals can also develop a sense of embarrassment and shame. Many social factors may mediate the presence of stigmatization toward patients in recovery, including but not limited to: negative beliefs of family members or support systems toward mental illness itself, social media influence and stereotypes related to mental health, negative attitudes of society in general, attitudes of a healthcare professional during crisis or emergencies, as well as structural discrimination experienced by the individual. There is limited evidence about the effectiveness of specific interventions to reduce stigma related to mental health. However, psychoeducation, contact with the mentally ill, stigma self-management, protest, and social activism, as well as legislative reform and advocacy, have been mentioned as potential interventions. Reducing stigma is crucial to decrease barriers during the recovery phase.

Recovered patients are capable of managing routines, work, school, independent living, and illness self-management without direct supervision by relatives or mental health professionals. That being said, it is strongly recommended that recovered persons sustain their connection to mental health services, even if this consists of only occasional “check-ins” with their treatment providers, given the increased vulnerability to relapse and stress sensitivity often experienced by individuals who have experienced serious mental health disorders. Refresher courses for patients’ emergency plans to implement relapse prevention are one mechanism to extend recovery for longer periods of time.

ASSESSMENT IN PSYCHIATRIC REHABILITATION

The assessment of persons with SMI involves a thorough and holistic assessment of a number of important domains with specific well-validated tools. The goal of any good assessment is to render an accurate diagnosis and then to suggest areas that can be targeted with specific interventions. The key to a comprehensive assessment of SMI is a thorough clinical assessment of the presence of symptoms, including the nature, frequency, intensity, and duration of particular symptoms. It should be noted that the management of certain symptom clusters (e.g., reduction of hallucinations and/or delusions) but not others (e.g., apathy, social cognition, and interpersonal skills) may not be specifically addressed by specific pharmacologic and behaviorally-based treatments. In addition, reduction of symptoms is not analogous to a “cure” for most individuals since neurocognitive and behavioral deficits in conditions such as schizophrenia may persist after symptomatic relief of positive symptoms. A number of symptom rating scales include the PANSS, the BPRS, and the Scale for the Assessment of Negative Symptoms (SANS). These instruments can provide useful and important information on symptom dimensions and provide a common basis by which to come to inform clinical and research diagnosis as well as to assess outcomes. This appraisal is vital as an initial step in creating a collaborative recovery plan.

As with many assessment instruments, interpretation of these measures depends on the operationalization and validity of specific symptom constructs as well as their ability to designate which cluster of symptoms meets the criteria for a particular psychiatric condition. It is important to take into account the fact that many psychiatric syndromes have overlapping features and differences in symptom presentation can be a function of the stage of illness, time of onset of a psychiatric episode, and the context and environment in which the patient is assessed. Many persons with SMI have comorbid psychiatric and medical disorders such as co-occurrence of SUD. Further specific endocrine, metabolic, and neurologic disorders may produce symptoms that mimic psychiatric symptomatology. In addition, an assessment of coping skills, daily living skills, housing arrangements, level of education and employment, family involvement, and social, sexual, and spiritual needs should be integrated into the holistic assessment of SMI.

One of the big challenges in the field is that controlled clinical trials often use relatively stringent inclusion and exclusion criteria that do not account for all of the comorbid psychiatric conditions, medical conditions, polypharmacy, and important mediating factors such as adherence, supported employment services (SES), environmental supports, and cross-cultural issues. As a result, it is increasingly important to ensure that findings obtained from placebo-controlled, randomized clinical trials generalize to individuals in different medical, hospital, and community mental health settings. While controlled randomized clinical trials may possess internal validity, the true test of an intervention is its effectiveness in real-world community settings.

Neurocognitive Symptoms

Cognitive assessment is a crucial component of the assessment of SMI, particularly in schizophrenia, given the number of impairments associated with the disorder and its challenging effects on activities of daily living and functioning. Neurocognitive deficits are thought to be a core feature of schizophrenia and often include emotional processing, perceptual disturbances, attention, sensorimotor integration, executive function, context processing, social perception, processing speed, and working memory. Up to 40% of persons with bipolar disorder may have completely normal cognitive profiles. In contrast, up to 75% of persons with schizophrenia perform below general population standards on neurocognitive tasks. Early identification of neurocognitive deficits in individuals living with schizophrenia is important as these deficits strongly correlate with chronic disability.

A tremendously positive development in schizophrenia research has been the development of the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) and Consensus Cognitive Battery (MCCB) for the assessment of schizophrenia. As a gold-standard assessment battery, the MCCB tests for ten cognitive domains take approximately 60 to 70 minutes to administer and allow researchers to standardize trials. The MCCB has not only been successful in research on schizophrenia but has also been able to differentiate Bipolar I and Bipolar II patients from controls. It has been recommended that because of the more subtle nature of cognitive deficits in schizophrenia, additional tests that should be considered (besides the MCCB) include the California Verbal Learning Test-II (CVLT-II). In addition, the Wisconsin Card-Sorting Task, Trails B, and the Stroop Color-Word task are all measures that should be considered. Moreover, a shorter (10-minutes) easily-administered battery, the Brief Cognitive Assessment Tool for Schizophrenia (B-CATS), has been developed to assist clinicians in the cognitive assessment of individuals living with schizophrenia, with proven good test-retest reliability and correlating 0.76 ($P < .01$) with the MATRICS battery.

An interview-based assessment of cognition specifically for schizophrenia is the Schizophrenia Cognition Rating Scale (SCoRS). The SCoRS assesses cognitive functioning based on interviews with both patients and their caregivers. The assessor then incorporates both sources of information to determine a final rating. Performance on the SCoRS has been found to be associated with both actual performance on cognitive tests and measures of real-world functioning. One study indicated that patients with schizophrenia who perform “normal” on cognitive testing do not indicate an association between cognitive performance and social or vocational functioning.

The challenge with any neuropsychological assessment is that there are confounding factors in SMI other than underlying putative brain damage that may contribute to poor performance. For example, poor motivation, sedation and other cognitive effects of medications, social anxiety, paranoia or hypervigilance, and interactions with the test examiner may have significant effects on test outcomes. In some settings, attention and motivation to perform may be so low that the neuropsychological battery with the Tests of Malingering and Memory (TOMM) or Green’s Word Memory Test (WMT) is augmented. It is important that any neuropsychological assessment that is performed is not in the acute stages of SMI. It is much more helpful to conduct neurocognitive evaluation after ample stabilization has occurred. There is a tendency to think that neuropsychological testing taps orthogonal, independent cognitive domains but deficits in memory and measures of verbal fluency may all be related to generalized cognitive slowing, attentional deficits, and reduced speed of processing. While there have been attempts such as the National Institute of Mental Health (NIMH) initiative (Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia [CNTRICS]) aimed to develop measures derived from cognitive neuroscience, with explicit links to known neural systems and psychopharmacology, the field is still hampered by a locationist notion of the brain where a function is thought to occur in specific brain regions (e.g., amygdala) rather than distributed networks of neuronal activity that involve multiple brain regions. This highlights the importance of relating neurocognitive function to functional measures such as fMRI and PET in order to identify working neural networks in the brain.

Social Cognition

Social cognition involves a variety of socially relevant domains, including emotion perception, social problem-solving, understanding social cues, and interpretation of others' attitudes and intentions. There is increasing evidence that negative symptoms of schizophrenia have an impact on both social competence and social outcomes but do not exert an influence on social cognition. However, social competence and social cognition have significant effects on real-world social outcomes. The result of the Expert Panel and Rand Survey identified four major domains of social cognition that include (a) emotional processing, (b) attributional style/bias, (c) social perception, and (d) theory of mind (ToM)/mental state attribution. Measures that the committee agreed upon for final inclusion in the battery to assess the aforementioned domains included the abbreviated version of the Ambiguous Intentions and Hostility Questionnaire (AIHQ), Bell Lysaker Emotion Recognition Task (BLERT), Penn Emotion Recognition Test (ER-40), Hinting Test, abbreviated version of Relationship Across Domains (RADs), Reading the Mind in the Eyes Test, The Awareness of Social Inferences Test (TASIT), and the abbreviated version of the Trustworthiness Test.

As part of the Social Cognition Psychometric Evaluation (SCOPE) study, the results of the final validation study showed that the BLERT, a test of affect recognition, the ER-40, and the Hinting Task (ToM) had the highest psychometric properties among a number of social cognition tasks and are the ones recommended for use in clinical trials. Meanwhile, the TASIT, Intentionality Bias Task (IBT), and Eyes Test showed weaker psychometric properties and remain in need of further study. On the other hand, the Mini Profile of Nonverbal Sensitivity (MiniPONS) and the Social Attribution Task-Multiple Choice (SAT-MC) showed poorer psychometric properties, and caution should be exerted when considered for clinical trials.

Nowadays, pharmacologic interventions may be used in conjunction with social cognitive training to facilitate the improvement of neurocognitive symptoms. Aerobic exercise has been found to have positive social cognitive effects and may be useful when introduced in conjunction with social cognitive interventions.

Communication and Social Skills

Deficits in metacognition, the ability to construct complex ideas about oneself and others, and the ability to respond to social and psychological challenges underlie many of the communication and social skills deficits found in schizophrenia. This mismatch between perceptions of oneself and the world and communication with others in the environment can cause considerable difficulties for an individual with schizophrenia. Impaired communication and language skills have been identified as common in schizophrenia, and it has been shown that individuals with schizophrenia have greater social skills deficits than those diagnosed with major depression or bipolar disorder. Those with schizophrenia appear to have social skills deficits that are intermediate.

SST is based on behavioral principles of modeling; shaping, rehearsal, and positive reinforcement are all important parts of successful training paradigms. SST aims to develop or improve social interactions, social performance, or interpersonal skills. It should be noted that SST has advocated for helping an individual interact with his/her world. Both social cognition and negative symptoms in schizophrenia carry the most variance in explaining real-world social outcomes. SST has a positive impact on improving the negative symptoms of schizophrenia. This suggests that targeting both social cognition and negative symptoms may be as important as teaching social skills to individuals. The problem with both social skills and expressive and receptive communication deficits in schizophrenia is that there is a complex interaction between social perceptions, integration of thinking about self and others, and behavioral outputs that can all serve to produce negative outcomes affecting social, vocational, and community life. It is likely that the most effective approaches are going to have to focus on the complex interplay between individual perceptual and social cognition deficits, how these interact with more complex metacognitive processes and help modify behavioral repertoires that will foster optimal real-world functioning.

Assessing Functional Outcomes

Broadly speaking, the goal of any psychiatric rehabilitation or remediation effort is to enhance and support optimal functioning in the environment.

Reaching milestones such as stable romantic relationships or full-time employment are not sensitive measures of functioning. Despite the considerable overlap, there are varying types of functions, including interpersonal communication, the ability to manage one's household affairs, managing one's finances, community affairs, and working in a vocational setting. For example, managing one's household affairs may involve arranging for transportation to the bank or supermarket and understanding bus, train, or subway routes. There are important decisions to be made on how to prioritize income, making sure that income is used for food, housing, and medication before it is allotted for other purposes. Persons must interact and effectively communicate with others and meet daily responsibilities in the community and the workplace. Understanding the individual's psychiatric condition, treatment regimen, adherence to treatment, natural social supports, contextual cues in the environment, and an understanding of a patient's strengths and limitations is critical to helping the individual achieve optimal functioning within the community. This requires assessment techniques that allow for appropriate targeted interventions and remediation strategies. Proper assessment allows for optimal treatment and psychiatric rehabilitation. These are not two distinct entities but rather a continuous effort to promote wellness and maintain optimal function in many contexts.

A critical distinction between functional capacity and functional performance is crucial. Functional capacity refers to the ability and the underlying nonsocial skills required to successfully perform activities of daily living in the community. Functional capacity can encompass a wide range of activities such as (a) the ability to count money or balance a checkbook, (b) having sufficient social cognition to understand implicit rules regarding social interactions or behaviors, (c) adequate social skills to successfully navigate a work or community environment, (d) requisite cognitive abilities such as intellect, memory, reasoning, language abilities to engage in meaningful home, community or occupational activities. Examples of functional capacity include balancing a checkbook, preparing a meal, understanding, completing work tasks adequately (i.e., taking medications as prescribed), and using the public transit system. If a person does not possess adequate functional capacity (requisite skills required for independent or even assisted living), then the focus of psychiatric rehabilitation has to be on attempting to build and foster those skills that will result in the greatest independence and quality of life. On the other hand, functional performance, while often dependent on underlying functional capacity, refers to the actual ability to consistently perform a variety of skills in a way that maximizes autonomy, quality of life, and the ability to cope with real-world demands. The functional performance then represents the actual performance of the individual in the interpersonal, social, financial community, and vocational domains.

Individuals with a psychiatric condition such as bipolar disorder may have an excellent underlying financial capacity, intellectual skills, social cognition, and awareness and be able to live independently and be gainfully employed. However, because of the episodic recurrence of a wide variety of mood symptoms, there may be certain periods where they may fail to adequately manage their manic, depressive, or mixed episodes and may mismanage finances, have difficulties meeting the demands of home, work, or family and be incapacitated despite their high premorbid levels of functioning. Individuals living with schizophrenia may have intact financial and adequate cognitive skills to manage to work and live independently in some settings. However, persistent difficulties with negative symptoms and social cognition can lead to problems that may negatively affect their ability to support themselves and their families. Primary and comorbid substance abuse can radically cause a patient to make poor choices and to allocate financial resources to their addiction or dependency, thus threatening their living situation. Psychiatrists are confronted with the unfortunate challenge that many unsheltered, socially impoverished, and disadvantaged individuals have underlying psychiatric conditions and lack proper support systems that could substantially help to improve their well-being and quality of life.

The distinction between functional capacity and general real-world functional performance is critical in that these constructs dictate the approaches that one will take toward assessment and subsequent management. Skill-building and other approaches are necessary to help increase functional capacity (i.e., via individual or small groups SST). However, the effectiveness of these approaches can only be seen when they form the building blocks to enhance functional performance (i.e., generalizing SST to real-world community and/or occupational settings). Thus, assessment of functional capacity and real-world performance are critical elements in psychiatric rehabilitation. Functional capacity and performance can further be subcategorized with regard to particular domains such as cognitive, financial skills and judgment, social cognition, social skills, and judgment that affect individual autonomy and quality of life. Further, specific types of functional skills required for everyday living may pertain to specific domains such as caring for one's self in the home, successfully interacting with the community, or learning and maintaining job-specific skills. This suggests that functional assessment strategies need to be tailor-made for each individual and display reliability (consistency of measurement) and validity (assessing the construct that needs to be measured).

Types of Functional Assessment. Functional assessment can be employed in a number of ways. The most direct method of assessing functional capacity is direct observation of the patient performing behaviors in the clinical setting. One widely used "paper and pencil" measure, such as the Performance-Based Skills Assessment (UPSA) developed by investigators at the University of California, directly assesses the individual's functional capacity with regard to finances, communication skills, household chores, transportation, and planning recreational and other activities. The advantage of such a measure is that it is easy to administer, has demonstrated reliability and validity, and can provide direct and quantifiable information on functional capacity.

Knowledge of deficits in functional capacity can be critical for initial assessment as well as response to treatment. However, a performance-based assessment may not always map onto the individual's performance in the community. In this case, a qualified informant such as a friend, family member, or healthcare provider can provide essential information based on repeated observations of the individual in the context of different environmental supports and contextual cues. Specifically, a reliable collateral informant can provide essential information that may simply not be available by performance-based functional assessment. However, the quality of the informant matters and an assessment of the frequency and nature of the relationship may be beneficial. If a collateral informant has little opportunity to observe the patient, has a cognitive, psychiatric disability themselves, or self-serving biases, their report may be of little value in providing useful functional information. While it is always important to obtain and respect the individual's evaluation of their real-world functional performance, it is important to note that in schizophrenia and other neuropsychiatric conditions such as Alzheimer disease (AD) and other neurodegenerative disorders, reports are often positively biased (i.e., the patient often reports a much higher level of functional performance than may actually be the case). Anosognosia, or the "inability to appreciate what one does not know," is a neuropsychological condition that is seen in traumatic brain injury (particularly within the right hemisphere), AD, and psychotic disorders such as schizophrenia. A 2020 study showed that overconfidence in a social cognitive emotion recognition task in participants with schizophrenia was a predictor of poorer performance on every performance-based task, including other social cognitive tests and measures of neurocognition.

As a result, it is often very useful for the clinician to compare the report of the patient and reliable informant and to also use performance-based functional assessment tools and rating scales. In addition, computer-based performance assessments have emerged, such as the Virtual Reality Functional Capacity Assessment Tool (VRFCAT). VRFCAT involves assessing function in a simulated virtual world, including tasks such as: preparing a shopping list, riding a bus, shopping in a store, and managing money. One study has shown that it is feasible to use technology-based functional capacity assessment protocols with older adults and patients with schizophrenia by conducting the assessment on a touch screen computer, with a battery of tests including ATM banking/financial management, prescriptions refill via telephone/voice menu system, and forms completion (simulating a clinic and patient history form). However, it is important to note that an individual's ability to complete these skills in a simulated and controlled environment differs from completing these skills in day-to-day life. An individual with prominent negative symptoms of schizophrenia may struggle with avolition and anhedonia and thus would have a lack of motivation for completing such tasks in the real world. Defeatist attitudes in individuals with schizophrenia have been shown to affect variance in negative symptoms, as well as neurocognition and functional outcomes. It is important to assess the individual's motivation in engaging in these rehabilitative assessments and interventions.

Functional Assessment Scales. In addition to the UPSA, the performance-based measure of functional capacity, a study also notes important scales for functional capacity such as the ILS and Test of Adaptive Behaviors (TABS). Everyday functional rating scales selected by the RAND expert panel include the Independent Living Skills-Survey

(ILSS), Specific Levels of Functioning Scale (SLOF), Life Skills Profile (LSP), Birchwood Social Functioning Scale (SFS), Social Behavior Schedule (SBS), and Heinrichs-Carpenter Quality of Life Scale (QLS).

The most widely employed measure of functional capacity, the UPSA, is a particularly compelling scale of functional capacity that has shown correlations of 0.6 to 0.7 with neurocognition in a number of studies. The USPA requires specialized personnel for administration. It includes simulations of everyday activities of daily living such as communication (i.e., dialing the telephone to schedule a physician appointment, obtaining and dialing a number from a telephone operator); comprehension/planning (reading and planning a recreational outing such as to the beach on a hot sunny day); finance (counting currency, making change, and paying a bill); household management skills (e.g., shopping for grocery items), and transportation (using a bus schedule in planning a trip and understanding charges and obtaining route information). Some versions of the UPSA include a medication management subtest and are referred to as the UPSA-II. USPA-B is a shorter version of this instrument that focuses on communication skills and finances.

The SLOF relies on self or informant reports but is especially useful when employed by a high-contact clinician. The scale taps the following domains: Participation in Community and Household Activities, Interpersonal Relationships, Work Skills, Self-Care, Physical Functioning, and Socially Acceptable Behavior subscales.

The UPSA has the advantage of being an objective performance-based functional measure and is preferable to self-reports of patients who often overestimate their abilities or questionable informants. Critics of the UPSA maintain that the UPSA measures tap functional behaviors but that they are more psychometrically similar to cognitive measures. There are contextual cues in the real-world environment that dictate real-world functional performance apart from demonstrating capacity in contrived clinical and laboratory settings. Measures such as the SLOF are most effective when used with informants but may in some cases be susceptible to reporter biases. The SLOF and other functional rating scales are typically most effective in the hands of a high-contact clinician who knows the patient well and can weigh informant and patient reports against his/her own clinical judgment. While the SLOF has been extensively validated with schizophrenia patients, studies have shown that 64% of patients with bipolar disorder showed evidence of deficits in one or more SLOF domains. The UPSA-1, UPSA-II, or UPSA B has been employed with a large number of patient groups, including those with major depression, bipolar disorder, and schizophrenia. Importantly, the instrument has also been employed in diverse cultural and language groups. A systematic review found that SLOF had the strongest observed correlation with the UPSA.

Other important scales in psychiatric rehabilitation include the Maryland Assessment of Social Competence (MASC), a structured role-play designed to assess social skills. Other scales use time-sampling techniques consisting of numerous observations of patient social behaviors, such as the Time Sample Behavioral Checklist (TSBC), Clinical Frequencies Recording System (CFRS), or Staff-Resident Interaction Chronograph (SRIC) but are usually only appropriate for institutional and residential treatment settings since these involve extensive ratings by staff members.

SPECIFIC APPROACHES

Cognitive Remediation in Community and Occupational Function

Neurocognitive deficits are core features of schizophrenia, and moderate to severe cognitive impairments are present in the vast majority of cases. CR may lead to small to moderate improvements in negative symptoms and social functioning, in addition to cognition. CR can be especially beneficial for persons experiencing subjective cognitive symptoms with schizophrenia and for individuals experiencing vocational challenges. This approach generally starts with simple information processing tasks, and then progresses toward more complex tasks such as targeting working memory and planning.

Cognitive remediation therapy (CRT) is an intervention that enhances cognitive performance in individuals with schizophrenia. However, other psychosocial interventions may be needed to improve functional outcomes. In the last two decades, studies have focused on identifying psychological and nonpsychological approaches to enhance social cognitive interventions and improve functioning. Even early in the course of psychotic disorders, social cognitive interventions can be helpful, and evidence shows that social cognitive deficits are present during the prodromal phase of individuals at high risk of developing psychotic disorders. As not all individuals benefit from social cognitive training, combining this approach with psychopharmacology may be promising.

The most commonly used measures are memory scales such as the Rey Auditory Verbal Learning Test (RAVLT) or CVLT-II. Executive function is often measured by the Wisconsin Card Sorting Test (WCST) or Trails B (although this measure taps many aspects of neurocognitive function) or measures of the Delis-Kaplan Executive Function System **(**DKEFS). Digit symbol of the WAIS-IV is a good measure of processing speed, as is Trails A. Examples of widely used neurocognitive batteries include the MATRICS MCCB, the Brief Assessment of Cognition in Schizophrenia (BACS), and Repeatable Battery for the Neuropsychological Assessment of Schizophrenia (RBANS).

When combining CR with functional skills training, better results in community living skills occurred than with functional training alone. An important review indicated that poor neurocognitive function had significant effects on vocational and occupational status within the community setting. Subsequent investigations revealed that CR plus SES enhanced positive employment outcomes, whereas other studies have not shown such effects. The notion is that training component cognitive abilities in the hopes that they will generalize to specific real-world behaviors may be less important than directly training target behaviors of interest. As an example, consider an adult with schizoaffective disorder who is working in a supported occupational setting. One approach might be to train specific cognitive component abilities in the hopes that they will generalize to the work setting. An alternative approach would be training those skills required for job-related performance. For example, CR strategies would specifically target actual job-related activities such as organization, strengthening procedural memory for certain role activities, and employing strategies to monitor performance. This targeted functional training would involve the cognitive aspects of the skills related to employment but directly integrate this training in an integrated manner and in a manner that would more likely generalize to the employment setting. Some studies have indicated that CR can be associated with neuroadaptation, though it is unclear if there are other factors influencing these results and if results persist with time.

Rehabilitation of Social Cognition

There has been an increased interest in moving beyond merely framing core features of schizophrenia in terms of positive and negative symptoms and neurocognitive impairment. There has been emerging evidence that deficits in social cognition can also be considered core and stable features of schizophrenia. Social cognition pertains to the perceptions, beliefs, and interpretation of other people's behaviors. Poor social cognition can significantly impair a person with SMI to appropriately respond to others in their environment. Green and others described five major aspects of social cognition; (1) ToM generally refers to an individual's ability to infer the beliefs and perceptions of others; (2) Social knowledge consists of the understanding of rules and conventions that underlie social behavior and guide social interactions; (3) Social perception is the ability to recognize general rules concerning roles and societal rules; (4) Emotional processing involves the experiencing and use of emotions and has been linked with emotional intelligence; and (5) Attributional bias involves attributing causality to both positive and negative outcomes. Each of these components is critically important in understanding how an individual relates to others in the community and their vocational performance. It should be noted that deficits in social cognition are not only observed in those with schizophrenia but also in individuals with bipolar disorder and, to a lesser extent, those with major depressive disorder.

There is increasing evidence that social cognition accounts for far more variability in community function than neurocognitive abilities. Moreover, a meta-analysis of studies on social cognition revealed that social cognitive training procedures have relatively large effects on emotion recognition, small to medium effects on measures tapping ToM effect, and no effects on attributional bias or social perception. One of the most important findings in this study is that these kinds of techniques have a moderate to large effect on community functioning. However, there may be less of an effect on higher-order social cognitive functions, which may improve more by using virtual reality such as the RC2S (CR of social cognition), which involves navigating a computerized character through situations that require emotional processing and learning contextual cues.

A 2018 meta-analysis of neuroimaging studies examining brain activation during facial emotion recognition and ToM tasks in patients with schizophrenia found neural-related circuit disruptions (i.e., reduced recruitment of the right insula, anterior cingulate and medial prefrontal cortex, and increased activation in the bilateral parietal cortex) leading to complex impairments in social cognition. Later on, a 2020 meta-analysis further described neural correlates known to play a role in social cognition of those with schizophrenia (i.e., bilateral anterior insula, right temporoparietal junction, left amygdala, and dorsolateral prefrontal cortex), suggesting that these brain areas should be studied as a potential target for therapeutic interventions in this population. Other small studies examining effects of social cognition intervention found neuroplasticity changes, including a reduction in grey matter loss, increase in regional brain activity, and decrease in alpha-range EEG activity, though these studies require to be replicated.

Cognitive-Behavioral Therapy

CBT is a type of psychotherapy based on principles of conditioning, learning, and cognition. As applied to schizophrenia, CBT engages patients collaboratively, challenging their interpretations of events or experiences. In schizophrenia, CBT aims to reduce the intensity of delusions, hallucinations, and related anxiety, to develop more adaptive behavior, and to encourage engagement of the person in relapse prevention. In many cases, the reduction in anxiety that accompanies modified attributions regarding voices leads to a reduced frequency and intensity of hallucinations. Individuals who are able to reframe or find alternative explanations for their positive symptoms can experience relief from these symptoms. If an individual resists challenging these symptoms, the CBT provider can encourage the individual to identify coping mechanisms to reduce distress related to these symptoms. In the case of negative symptoms of schizophrenia, CBT is often used to challenge anhedonic beliefs, schedule activities, and administer self-rewards for increasingly engaging in pleasurable activities. CBT can be provided in individual or group settings and works best with patients who are willing to actively participate in treatment, often as the result of distress from symptoms and/or at least minimal insight into the abnormal nature of their experiences.

A 2014 meta-analysis of 14 studies including 1,484 patients found that compared to other adjunctive measures, CBT showed a significant reduction in positive symptoms. The specific symptoms alleviated with adjunctive CBT include delusions, distress, or delusions related to auditory hallucinations, paranoia, and negative symptoms. Improvements in distress related to auditory hallucinations and delusions were maintained at follow-up. A more recent meta-analysis of CBT for schizophrenia conducted in 2018, including 33 studies with a total of 1142 participants with schizophrenia and positive symptoms, measured response rates to CBT by the degree of improvement. In this meta-analysis, 44.5% of patients exhibited a 20% (minimally improved), and 13.2% of patients exhibited a 50% (much improved) reduction of overall symptoms. However, the degrees of improvement reported on positive symptoms showed 52.9% (minimally improved) and 24.8% (much improved) in patients.

It seems that CBT may help patients presenting with psychosis. However, the effects may be small to moderate. Further research is needed to assess the efficacy of CBT for schizophrenia on social functioning, quality of life, relapse prevention of psychotic episodes, insight, and other outcomes. Meanwhile, it is worth noting that in addition to problems uniquely associated with psychotic disorders, various forms of CBT are effective for addressing generalized anxiety, panic, social anxiety, depression, obsessive-compulsive symptoms, and SUD. These problems may co-occur with chronic psychotic disorders, and there is no reason to believe that CBT interventions are any less effective in treating them in people with schizophrenia than in the general population.

Most patients with schizophrenia are able to engage in CBT despite exhibiting acute psychosis and cognitive deficits, as long as they are somewhat motivated to engage in CBT. Motivated patients typically demonstrate good treatment responses, whereas unmotivated patients often do not respond or drop out of treatment. A major challenge is to find ways to engage patients exhibiting first-episode psychosis in CBT. Often, more supportive efforts are needed to engage first-episode patients in the treatment system prior to initiating CBT.

Social Learning and Social Skills Training

The social learning program (SLP) is a comprehensive, integrated set of learning-based techniques (typically including a token economy) and skills-training activities. In most cases, an SLP is delivered within the context of a residential facility, usually for patients with severe and chronic behavioral disabilities. A key feature of the SLP is that all staff members on all shifts are trained to structure their interactions with patients in specific ways during all waking hours. Research indicates that SLP is the most effective form of inpatient treatment for people with schizophrenia, including the so-called “treatment-refractory” patients. The SLP has demonstrated effectiveness with a wide range of people with disabilities in a variety of settings (e.g., state and private hospitals and maximum security forensic hospitals). The components of the SLP form an extremely positive teaching environment that assists clients in developing adaptive behaviors and skills relevant to successful return to the community. Contemporary versions of this approach incorporate skills-training technologies, psychiatric rehabilitation techniques for engaging clients, CBT, vocational rehabilitation, cognitive rehabilitation, and substance abuse treatment.

A guiding assumption of the SLP is that all persons with severe mental disorders, no matter how severe the impairments, can learn new behaviors and skills and, therefore, progress to less structured environments. To promote behavior change and skill acquisition, the SLP incorporates a wide variety of techniques based on fundamental learning principles. Differential reinforcement forms the basis of many of the procedures of the SLP. Staff at all levels and disciplines are highly trained and carefully supervised in the timely and consistent delivery of verbal praise, positive social attention, and material reinforcement for adaptive behaviors and the application of extinction or response-cost techniques for bizarre, aggressive, or rule-violating behaviors. Effective SLPs deliver much more positive reinforcement than extinction or response costs. The result is a highly reinforcing, positive, and supportive environment.

Although most comprehensive SLPs include some sort of token economy or point system to deliver feedback and behavior-related consequences, some use individual behavioral contracts, contingency management plans, or social reinforcement alone. However, in all well-designed SLPs, specific objectives and goals of clients are individualized, as are the specific interventions for assisting them with acquiring the skills necessary to accomplish these goals/objectives. Another key feature of all effective SLPs is that, as patients begin to demonstrate improvements in behaviors, there is a planned and systematic withdrawal of tokens or points and an emphasis on generalization training so that clients will be able to apply acquired skills and knowledge successfully in “real-life” community settings under more natural, primarily social reinforcement conditions.

An effective SLP is based on a wealth of data that is generated during all waking hours by all staff, using a number of observational rating procedures (described in the section on assessment). The availability of data on all critical behaviors (e.g., grooming, attendance at groups and activities, social contacts, and aggressive behavior) allows for sophisticated data-driven treatment planning. Moreover, the ability to demonstrate specifically to patients why their privileges or other consequences of their behavior are occurring and what needs to change, and how much to receive different consequences is a powerful tool. In effect, this type of data-driven system serves to instill responsibility for behavior change in the patient, instills a sense of agency and mastery regarding behavior change, and reduces the institutionalized and dependent attitude that staff arbitrarily determine privileges, which is so often found in state hospital settings. Moreover, as demonstrated in a recent study, patient satisfaction with treatment is often high in SLPs.

Given the complexity, specificity, and range of procedures involved in the SLP, staff training is of paramount importance. All staff members are involved in extensive initial and ongoing training, using both didactic instruction and in vivo practice. Competency-based supervision is provided through regular performance feedback, using observational data on staff-patient interactions and structured supervisory observations.

Another important feature of SLPs is the focus on the generalization and maintenance of gains after program discharge. As patients acquire skills and develop goals for community living, they are provided with opportunities to apply these skills and pursue their goals in community settings. After discharge, declining contact aftercare is provided. Some SLPs arrange to train staff at community residential facilities to maximize the maintenance of gains after discharge.

SLPs have traditionally been provided in long-term treatment units. Recently, however, modifications of the full program have been tried in other settings, such as acute inpatient units and community-based residential settings for outpatients with SMI, with demonstrated success.

Illness/Wellness Management Skill Training

Gaining the ability to manage one’s psychiatric illness is central to the rehabilitation and recovery perspective. In the rehabilitation and recovery literature, skill training in illness/wellness management has gradually been differentiated from related social, and living skills approaches, reflecting a growing recognition that specialized skills are needed to self-manage psychiatric disorders, comparable to skills needed to self-manage physical conditions such as diabetes. When providing persons with SMI the knowledge and skills, they are empowered to manage their own symptoms, make informed decisions about their treatment plans, and identify their recovery plan. People learn about the episodic and persistent symptoms of their illness, the relationship between symptoms and functional impairments, pharmacologic options and other behavioral techniques (e.g., relaxation and stress management) for controlling symptoms, medications side effects, identification of “warning signs” of an impending relapse, creating a relapse prevention plan, skills to collaboratively communicate with the rehabilitation team, mental health providers and family who can support their recovery journey, and various other aspects of their disorder and its management. Behavioral skills indirectly relevant to disorder management include, for example, the assertive skills necessary for dealing with the doctor and the doctor’s receptionist in getting an appointment for a medication visit. Integrated Illness Management and Recovery (I-IMR) programs target patients who have psychiatric disorders and comorbid medical conditions. Wellness self-management includes directing individuals with issues relating to lifestyle choices, nutrition, physical activities, relationships, spiritual and pleasurable activities.

Several developer groups have packaged illness/wellness management skill training materials for testing and dissemination. A number of original studies and reviews confirm the effectiveness of skill training focused on illness/wellness management for improving adherence to treatment in people who live with SMI. The SAMHSA has produced packages for public dissemination that fall under this category, including What Individuals in Recovery Need to Know About Wellness, Illness Management, and Recovery, and Medication Management Approaches in Psychiatry. Both packages include materials for all participants in the illness/wellness management process, including the person in recovery, physician, other service providers, family, and friends. The Agency of Healthcare Resource and Quality (AHRQ) created a Self-Management Support Resource Library with training material for providers as well as patient material on self-management. These resources can be accessed on AHRQ’s website: https://www.ahrq.gov/ncepcr/tools/self-mgmt/home.html

Illness/wellness management training for people with schizophrenia benefits from relapse prevention and related techniques. Relapse prevention programs work with individuals to identify environmental stressors that can trigger warning signs of relapse, develop steps that can be taken to prevent relapse, and teach individuals stress management techniques. Well known for its application in substance abuse, many of the techniques of relapse prevention are well suited to the episodic nature of the lived experience of people living with schizophrenia, the role of stress in precipitating psychotic symptoms, and the important role of the recovering person in managing those episodes. The original application of relapse prevention is also of interest and should be further studied, given that people with schizophrenia often have substance abuse problems as well. Peer involvement is also especially relevant to illness/wellness management.

Supported Education

Supported education prepares people with psychiatric disabilities to achieve educational goals in an academic setting in a person-centered approach, including starting college or continuing interrupted college education due to mental health issues. A retrospective study examining medical records of 114 patients with the first episode of psychosis (FEP) found that 82% of individuals enrolled in college experienced interruptions in their education following FEP. Supported education can help forge a path for individuals with SMI to develop future careers or engage in meaningful work. A randomized controlled trial in 2017 found that a comprehensive, team-based treatment program that includes supported employment and education for individuals following FEP was associated with greater improvement in work or school participation.

Supported education addresses ways to overcome obstacles related to achieving educational success, such as improving academic skills and problem-solving, bolstering self-confidence, and engaging in career development. Every college has its own policies regarding disability support services, with accommodations varying depending on the individual’s needs. Registering with the college’s disability resource office, providing documentation of the diagnosis, prior and requested academic accommodations, and collaborating with the college disability resource center are some of the steps needed to be taken by individuals with SMI to obtain supported education.

Many models of supported education exist, including contained classrooms, onsite support, and mobile support. In the mobile support approach, a counselor meets one-on-one with a student to tailor an individualized plan to maximize academic success. This may include assistance in accessing services already in place in the educational setting, help with locating and using school social and recreational services, preparatory coursework, case management support, ongoing support and check-in meetings at the local student center, study support, visits to the campus library, linkage support to other campus services, and individual counseling (didactic, supportive, and/or exploratory). Other models of supported education include didactic group meetings in a classroom setting focusing on study skills, supportive group meetings, and the provision of standard study guides. Evidence from these programs suggests that distress levels are higher in people with schizophrenia than in other students. This should lead to further efforts to introduce stress management techniques into supported education activities for patients with schizophrenia. In addition, some rehabilitation programs offer wellness and resilience training, as well as academic coaching for individuals with SMI who are on college leave.

Sex Education

Many individuals with SMI aspire to and achieve meaningful education, work, and enjoy mutually supportive family and intimate relationships. The topic of intimacy and sexuality has been given the least attention in this population by mental health professionals. Surveys have discovered that 60% to 70% of adults with SMI report being sexually active in the last year, with approximately 30% reporting two or more partners. Persons with SMI are more likely to be involved in risky sexual behaviors, including inconsistent condom use, multiple sexual partners, and transactional sexual acts. In addition, some psychotropic medications are associated with sexual side effects. Persons with chronic mental illness are also more likely to experience intimate partner violence as well as sexual exploitation.

Sexuality and sexual expression should be recognized in all individuals living with chronic mental illness. The World Professional Association for Transgender Health (WPATH) has created a standard of care guideline for health professionals caring for transgender and gender diverse, including guidelines on psychological well-being, social support, and individual fulfillment. Sexual health education needs to be tailored for lesbian, gay, bisexual, transgender, queer, intersex (LGBTQI) individuals. Peer-to-peer support can be beneficial in sexual education.

Clearly, sex education should be an integral part of a comprehensive rehabilitation program. The current impetus for increasing patients' sexual knowledge derives in part from the alarmingly high rates of HIV infection and AIDS found among individuals with SMI. Not surprisingly, the majority of sex education programs have been designed to increase patients' awareness of the risk of infection, dispel misinformation and stereotypes, and provide the knowledge and skills needed to decrease their chances of becoming infected. A meta-analysis examining prevalence data of blood-borne infections in people with SMI found that people with SMI are at risk of blood-borne viral infections, but due to methodologic limits, these prevalence rates may be overestimated. Other contributing factors include low socioeconomic status, substance use, and sex. Emphasizing teaching techniques that use procedural learning and memory can mitigate any cognitive disabilities that may impact learning. In psychiatric rehabilitation, implicit or procedural learning is enhanced by video, and live demonstrations termed social modeling. Other techniques that build on procedural learning and memory are role-playing, interacting with a computer as in computer games, using pictures and attractive graphics to amplify written information, and engaging in repeated practice.

These educational methods can also be used for sex education programs that have broader goals than simply teaching persons with serious and persistent mental illness how to reduce their chances of becoming infected by HIV. Some of the educational objectives that comprise more ambitious and enlightened efforts in this realm include the following:

- Increasing patients' knowledge and comfort about sexuality
- Encouraging patients to identify and clarify their values and attitudes about sexuality
- Informing patients about alternative methods of birth control
- Teaching patients decision-making skills regarding sexual activity
- Overcoming medication-related sexual dysfunction and preventing deterioration or cessation of sexual activity
- Improving intimacy skills
- Imparting basic education on sexually transmitted diseases
- Instructing patients in proper condom use

COMORBID SUBSTANCE USE DISORDERS

The field of psychiatric rehabilitation is confronted with the need to identify and treat SUD as epidemiologic data show high prevalence rates among individuals with concurrent SMI. SUD have been associated with medication nonadherence, underachievement, decreased work productivity, poor health, neuropsychological impairment, social dysfunction, and violence, among other negative biopsychosocial effects. Prospective studies have shown that treatment outcomes, hospitalization rates, housing stability, and functional status are worse among patients with dual disorders. Studies have also shown that problems related to substance use tend to persist over the long term among patients with SMI. Lifetime treatment or help-seeking behavior is uncommon, particularly among those with psychiatric comorbidity. Thus clinicians are encouraged to be proactive with assessing for SUD to mitigate the negative impact on successful recovery and community adjustment.

Historically, individuals with SMI such as schizophrenia and co-occurring SUD received treatment for one disorder but were excluded from the other. Eventually, distinct and parallel treatment programs emerged; however, treatment outcomes were poor using this model. As a result, integrated treatment models combine substance abuse and mental health interventions in one clinical program. This model provides comprehensive outpatient dual disorders interventions as well as integrated residential treatment programs. In integrated treatments for patients with dual disorders, mental health treatments, and substance abuse treatments are brought together by the same clinician, or team of providers, in the same program to ensure that the patient receives a consistent, specific, and holistic explanation of illness and a coherent prescription for treatment rather than a contradictory set of messages from different providers. Integrated treatment aims to improve collaboration between treatment providers, eliminate the patient’s burden of attending two separate programs, and remove financial and other barriers to access and retention. Comprehensive programs have frequently included not only standard mental health interventions, such as medication monitoring and support services, but also assertive outreach to engage patients in treatment; intensive case management; trauma interventions; housing supports; and individual, group, and family substance abuse counseling.

Motivational interviewing, as well as CBT, are often utilized in patients with comorbid substance use and chronic mental illness. Harm reduction such as needle exchange services and pharmacologic therapy such as opioid replacement therapy can be incorporated as well on an individualized basis. Relapse prevention, which is often included in these approaches, focuses on preventing exacerbations of both mental illness as well as substance use relapse. Literature in this area shows that long-term treatment is associated with better overall outcomes. Research efforts are still being made to test and refine individual components and combinations of integrated dual-diagnosis treatments. As with all treatment, effective interventions targeting substance use should be individualized to address personal factors and the stage of motivation.

TOWARD MORE INTEGRATIVE MODELS OF CARE

Current psychiatric rehabilitation extends far beyond previous psychosocial models that were added to medication management and other somatic therapies. Psychiatric rehabilitation today is an active, collaborative, and person-centered endeavor, which necessitates a strong therapeutic relationship, trust, and collaboration between the patient, family, and multidisciplinary treatment team. Excellent treatment is predicated by using state-of-the-art and empirically validated assessments. While neurocognitive deficits may be pervasive in many types of schizophrenia, persons with bipolar or unipolar affective disorder may have less but still significant impairment in this area. In some cases, comorbid substance abuse disorders can further impair already pre-existing cognitive deficits. Equally important is the accurate assessment of functional capacity (based on performance-based tests) and, most importantly, real-world functional performance. Functional performance cannot be limited to merely higher- and lower-order activities of daily living but must also encompass social and vocational functions. There is also increasing evidence that deficits in social cognition and meta-cognition create considerable stress for individuals with schizophrenia beyond typical memory impairments, slow cognitive processing, and other neurocognitive deficits. As previously discussed, there is ample literature to suggest that improving communication skills, enhancing social skills, and behavioral goal attainment are all effective means to help the individual cope with stressors that might increase the likelihood of relapse.

This chapter has focused on empirically oriented approaches that can help in these efforts. As previously mentioned, it is of essential importance for both the patient and the treatment team to (a) identify targeted short-term and long-term goals that are both operational and obtainable, (b) plan interventions and build supports that are appropriate to a person's stage of recovery, (c) build coping strategies and stress management techniques that will reduce the chances of relapse, and (d) foster a collaborative treatment milieu involving the patient, family, and care providers.

Every person with SMI is a unique individual. While evidence-based approaches can provide guidance in the selection of the best instruments and interventional tools, this can never be done in isolation and without knowledge of the person's strengths, weaknesses, resources, support, and unique personality and environment. The best chances for optimal recovery and stabilization are through a multimodal approach that encourages adherence to medication and targeted treatment goals, stress management, and coping techniques, as well as developing optimal environmental supports.

SMI conditions are like many other chronic medical conditions that need constant attention and management. However, unlike many physical ailments, persons with SMI still have to manage additional stigma and lack of understanding afforded those with chronic medical and neurologic conditions. The most effective interventions will be those with the most empirical support, targeting the many different aspects of deficits in SMI (including but not limited to social isolation, deficits in social cognition, problematic meta-cognition, functional and vocational deficits, as well as problems with communication and social skills). Focusing solely on these targets without recognizing the importance of contextual cues and social supports risks fragmenting care.

Psychiatric rehabilitation is based on providing hope, self-empowerment, resilience, and recovery outlook and works to provide a patient-centered approach for individuals with psychiatric disabilities to accomplish their desired goals. There needs to be continued emphasis on more accessible, inclusive, culturally sensitive, integrated, and comprehensive models of care. Patients deserve no less.

FURTHER READINGS

Beck AT, Himelstein R, Bredemeier K, Silverstein SM, Grant P. What accounts for poor functioning in people with schizophrenia: a re-evaluation of the contributions of neurocognitive v. attitudinal and motivational factors. *Psychol Med*. 2018;48(16):2776–2785.

Bighelli I, Huhn M, Schneider-Thoma J, et al. Response rates in patients with schizophrenia and positive symptoms receiving cognitive behavioural therapy: a systematic review and single-group meta-analysis. *BMC Psychiatry*. 2018;18(1):380.

Birnbaum ML, Ernala SK, Rizvi AF, et al. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. *NPJ Schizophr*. 2019;5:17.

Campos C, Santos S, Gagen E, e al. Neuroplastic changes following social cognition training in schizophrenia: a systematic review. *Neuropsychol Rev*. 2016;26(3):310–328.

Cella M, Preti A, Edwards C, Dow T, Wykes T. Cognitive remediation for negative symptoms of schizophrenia: a network meta-analysis. *Clin Psychol Rev*. 2017;52:43–51.

Czaja SJ, Loewenstein DA, Lee CC, Fu SH, Harvey PD. Assessing functional performance using computer-based simulations of everyday activities. *Schizophr Res*. 2017;183:130–136.

da Silva AG, Baldaçara L, Cavalcante DA, Fasanella NA, Palha AP. The impact of mental illness stigma on psychiatric emergencies. *Front Psychiatry*. 2020;11:573.

Deckler E, Hodgins GE, Pinkham AE, Penn DL, Harvey PD. Social cognition and neurocognition in schizophrenia and healthy controls: intercorrelations of performance and effects of manipulations aimed at increasing task difficulty. *Front Psychiatry*. 2018;9:356.

Eisner E, Drake R, Lobban F, Bucci S, Emsley R, Barrowclough C. Comparing early signs and basic symptoms as methods for predicting psychotic relapse in clinical practice. *Schizophr Res*. 2018;192:124–130.

Firth J, Stubbs B, Rosenbaum S, et al. aerobic exercise improves cognitive functioning in people with schizophrenia: a systematic review and meta-analysis. *Schizophr Bull*. 2017;43(3):546–556.

Frost BG, Tirupati S, Johnston S, et al. An Integrated Recovery-oriented Model (IRM) for mental health services: evolution and challenges. *BMC Psychiatry*. 2017;17:22.

Granholm E, Harvey PD, Social skills training for negative symptoms of schizophrenia. *Schizophr Bull*. 2018;44(3):472–474.

Harvey PD, Strassnig MT, Silberstein J. Prediction of disability in schizophrenia: symptoms, cognition, and self-assessment. *CNS Spectr*. 2019;10(3).

Horan WP, Green MF. Treatment of social cognition in schizophrenia: current status and future directions. *Schizophr Res*. 2019;203:3–11.

Hurford IM, Ventura J, Marder SR, Reise SP, Bilder RM. A 10-minute measure of global cognition: validation of the Brief Cognitive Assessment Tool for Schizophrenia (B-CATS). *Schizophr Res*. 2018;195:327–333.

Jáni M, Kašpárek T. Emotion recognition and theory of mind in schizophrenia: a meta-analysis of neuroimaging studies. *World J Biol Psychiatry*. 2018;19(sup3):S86–S96.

Matsuda Y, Makinodan M, Morimoto T, Kishimoto T. Neural changes following cognitive remediation therapy for schizophrenia. *Psychiatry Clin Neurosci*. 2019;73(11):676–684.

Metcalfe JD, Drake RE, Bond GR. Economic, labor, and regulatory moderators of the effect of individual placement and support among people with severe mental illness: a systematic review and meta-analysis. *Schizophr Bull*. 2018;44(1):22–31.

Miller E, Stanhope V, Restrepo-Toro M, Tondora J. Person-centered planning in mental health: a transatlantic collaboration to tackle implementation barriers. *Am J Psychiatr Rehabil*. 2017;20(3):251–267.

Moreno C, Galderisi S, Nordento M, et al. How mental health care should change as a consequence of the COVID-19 pandemic. *Lancet Psychiatry*. 2020;7(9):813–824.

Padilla VL, Rustad JK, Harvey PD, Wasserman B, Ziemer D, Musselman DL. Neurocognitive impairment in patients with comorbid diabetes mellitus and depression. *Pers Med Psychiatry*. 2017;1–2:2–10.

Perez MM, Tercero BA, Penn DL, Pinkham AE, Harvey PD. Overconfidence in social cognitive decision making: correlations with social cognitive and neurocognitive performance in participants with schizophrenia and healthy individuals. *Schizophr Res*. 2020;224:51–57.

Pinkham AE, Harvey PD, Penn DL. Social cognition psychometric evaluation: results of the final validation study. *Schizophr Bull*. 2018;44(4):737–748.

Rosenheck R, Mueser KT, Sint K, et al. Supported employment and education in comprehensive, integrated care for first episode psychosis: effects on work, school, and disability income. *Schizophr Res*. 2017;182:120–128.

Shinn AK, Cawkwell PB, Bolton K, et al. Return to college after a first episode of psychosis. *Schizophr Bull Open*. 2020;1(1):sgaa041.

Stevens M, Ratheesh A, Watson A, Filia K, Donoghue BO, Cotton SM. Rates, types and associations of sexual risk behaviours and sexually transmitted infections in those with severe mental illness: a scoping review. *Psychiatry Res*. 2020;290:112946.

Suijkerbuijk YB, Schaafsma FG, van Mechelen JC, Ojajärvi A, Corbière M, Anema JR. Interventions for obtaining and maintaining employment in adults with severe mental illness, a network meta-analysis. *Cochrane Database Syst Rev*. 2017;9(9):CD011867.

Szabo S, Merikle E, Lozano-Ortega G, Powell L, Macek T, Cline S. Assessing the relationship between performance on the University of California Performance Skills Assessment (UPSA) and outcomes in schizophrenia: a systematic review and evidence synthesis. *Schizophr Res Treatment*. 2018;2018:9075174.

Vita A, Barlati S. The implementation of evidence-based psychiatric rehabilitation: challenges and opportunities for mental health services. *Front Psychiatry*. 2019;10:147.

Vucurovic K, Caillies S, Kaladjian A. Neural correlates of theory of mind and empathy in schizophrenia: an activation likelihood estimation meta-analysis. *J Psychiatr Res*. 2020;120:163–174.

56.7 A Sociocultural Framework for Mental Health and Substance Abuse Service Inequities

Margarita Alegría, Ph.D., Bernice A. Pescosolido, Ph.D., Kiara Alvarez, Ph.D., and Glorisa Canino, Ph.D.

The Institute of Medicine defined disparities in health care as "racial or ethnic differences in the quality of healthcare that are not due to access-related factors or clinical needs, preferences, and appropriateness of interventions." Responding to an important gap in the definition, the National Institute on Minority Health and Health Disparities expanded the definition in 2019 to include "a pattern of poorer health outcomes, indicated by the overall rate of disease incidence, prevalence, morbidity, mortality, or survival in the population as compared with the general population" and defined "racial/ethnic minorities, socioeconomically disadvantaged populations, underserved rural populations, and sexual and gender minorities" as those more likely to experience the disparities. Although reports since 2000 have strengthened public concern about glaring discrepancies in patterns of health service use and outcomes for racial and ethnic minorities in the United States, there has been limited progress in disparities reduction from 2000 to 2018.

Most entrenched disparities continue to be associated with the patient's race, ethnicity, and income profile. Ignoring these nonmedical factors represents a lost opportunity to eradicate the augmented age-adjusted mortality and increased burden of disease for minority populations. Despite the attention focused on health services disparities, particularly in view of demographic projections suggesting that ethnic/racial minorities will become the majority of the U.S. population by 2045 and by 2020 for children, we lack strategic approaches that match the complexity of the problem. Understanding the policy, social, and cultural factors that account for and reduce these disparities provides guidance for crafting effective individual interventions and systemic policies. Without frameworks and models, implementing the science-to-services translation called for by the National Institutes of Health is challenging.

To date, numerous conceptual frameworks have attempted to chart the potential causes of mental health service disparities. Most are based on the understanding that disparities in mental health and mental health care have no single underlying cause and that mechanisms likely vary across racial/ethnic groups and geographic contexts. Many scholars argue that the *fundamental causes* of health service disparities emerge from macrosocial factors such as historical and contemporary structural racism. Structural racism refers to “the totality of ways in which societies foster racial discrimination through mutually reinforcing inequitable systems (e.g., in housing, education, employment, earnings, benefits, credit, media, health care, and criminal justice, etc.) that in turn reinforce discriminatory beliefs, values, and distribution of resources.” Given the entrenchment of structural inequities in resources, policy, and practice, many scholars argue that a targeted focus on structural racism offers a direct and effective approach to finding a solution to health service disparities.

Additionally, some frameworks detail a small set of factors related to a single, particularly important, intermediary mechanism (e.g., residential segregation or employment) of health service disparities. Others seek to provide an overview of a broad range of possible mechanisms. A basic critique of service disparities frameworks is that too many have centered closely on individual-level mechanisms (i.e., supportive social relationships, family networks) and individuals’ responses to those circumstances (i.e., optimism, personal agency), which impact psychosocial reserve capacity and cognitive-emotional processes. These models tend to be reductionistic, failing to incorporate dynamic, system-level factors or consider reciprocal effects of different interventions, the context of dynamic reciprocal relations, or issues such as the availability of resources or services in the area. These system forces are vital to grasp the essential issues and to design system interventions. However, targeted factors should be those that matter substantially, are modifiable, and advance the science and public policy practice. Building from past theoretical and empirical work, this chapter updates a culturally informed, complex, population health systems framework for mental health and substance abuse service disparities which integrates a new definition of health care inequities.

In our effort to integrate sociocultural elements into a health service inequities frame, our definition of health care inequities or disparities refers to *racial and ethnic differences in access, health care quality, or health care outcomes that are not due to clinical needs or the appropriateness of treatment* and that *are a result of systemic, avoidable, and unjust social and economic policies and practices that create barriers to equity in mental health and substance abuse treatment services and outcomes.* Disparities occur across many dimensions, including race/ethnicity, socioeconomic status, age, location, gender, disability status, and sexual orientation. Patient preferences are also included as potential sources of disparities since they do not stem from "undue differential treatment on the basis of race and ethnicity." Past experiences and availability of services patterns might influence racial and ethnic minorities' preferences in treatment. Thus, preferences cannot be disentangled from people's available options in care or from health literacy factors and can contribute to inequities.

This approach to understanding disparities in service delivery emphasizes that sociocultural processes refer not only to race/ethnicity as a particular set of values, norms, and beliefs but also to the historical and concurrent circumstances and experiences leading to a group's social, economic, and political status. In addition, culture creates repertoires or "cultural toolboxes" that provide accepted "go-to" ways to behave, including how to respond to health problems. To understand the relationship between culture and health, a simplistic view of culture or cultural differences cannot work. Cultural processes are fundamental threads connecting communities, health systems, providers, and patients. They "veil" the entire complex system that deals with population health. Nowhere in the Sociocultural Framework (SCF) is the nature and operation of "culture" partitioned in a separate box or, even more importantly, proxied by sociodemographic categories. Racial, ethnic, and national categories are just proxies for markers of culture, pre- and postimmigration histories, and social and economic stratification, and they should be acknowledged as such. Since each individual has a profile of statuses, not just one, that could be subject to discrimination (e.g., member of a minority group, identifies as LGBTQ+, and is an immigrant), understanding and addressing their joint effects on mental health, substance use, treatment, and outcomes is complicated. When researchers are trying to reflect intersectionality (i.e., the convergence of multiple devalued statuses) rather than including an ever-expanding set of interacting statuses, Lamont and colleagues suggest focusing on cultural capital and cultural processes that sustain and reproduce societal and health inequities.

THE SOCIOCULTURAL FRAMEWORK: A COMPLEX, POPULATION HEALTH SYSTEMS APPROACH

Many of the recent agenda-setting reports in medicine see many contexts implicated in the cause, treatment, and outcomes of health and disease for individuals, providers, and societies. In many ways, seeing the relevance of everything from genetic risk to national health care policies is not new. From Bronfenbrenner's early ecologic systems theory to the current popularity of socioecologic models, the idea that different contexts matter for human growth, development, and well-being has become the "metaphorical blueprint" for much of basic and translational science. There is a general acceptance that the "environment" or the "environmental context" in its many guises "gets under the skin" and "inside the mind." What is novel here, and often missing in these standard models, is the recognition that much of what happens across these contexts results from the active forces that operate at their borders. Context here is composed of two types: across-level contexts (e.g., place or geography, health care organization, and individuals) and across-system contexts (the community and the health care system). Complex systems theory offers the view that these across-level and across-system interactions tend to persist over time, adapt to changing circumstances, and, as a result, are difficult to alter by addressing only one level or one system. Thinking through how major contexts meet, clash, cooperate or otherwise interact across levels and systems provides novel insights for critical, potential levers of change. Even more importantly, without establishing what happens at the touchpoint of major levels and systems, unintended consequences are likely.

As depicted in Figure 56.7–1, the Socio-Cultural Framework (SCF) for Mental Health and Substance Abuse Service Inequities has three essential elements—systems, levels, and interacting mechanisms. Two key systems in the complex population health system that can produce disparities or buffer inequities are the health care system and the community system. Each operates at three levels—the societal level (larger policy or environmental contexts), the organizational level (formal organizations or lay sectors), and the individual level (provider or patient). At each level, these contexts interact or interface to produce the fundamental mechanisms of inequality and bias that cumulate to inequities in health, health care, and quality of life.

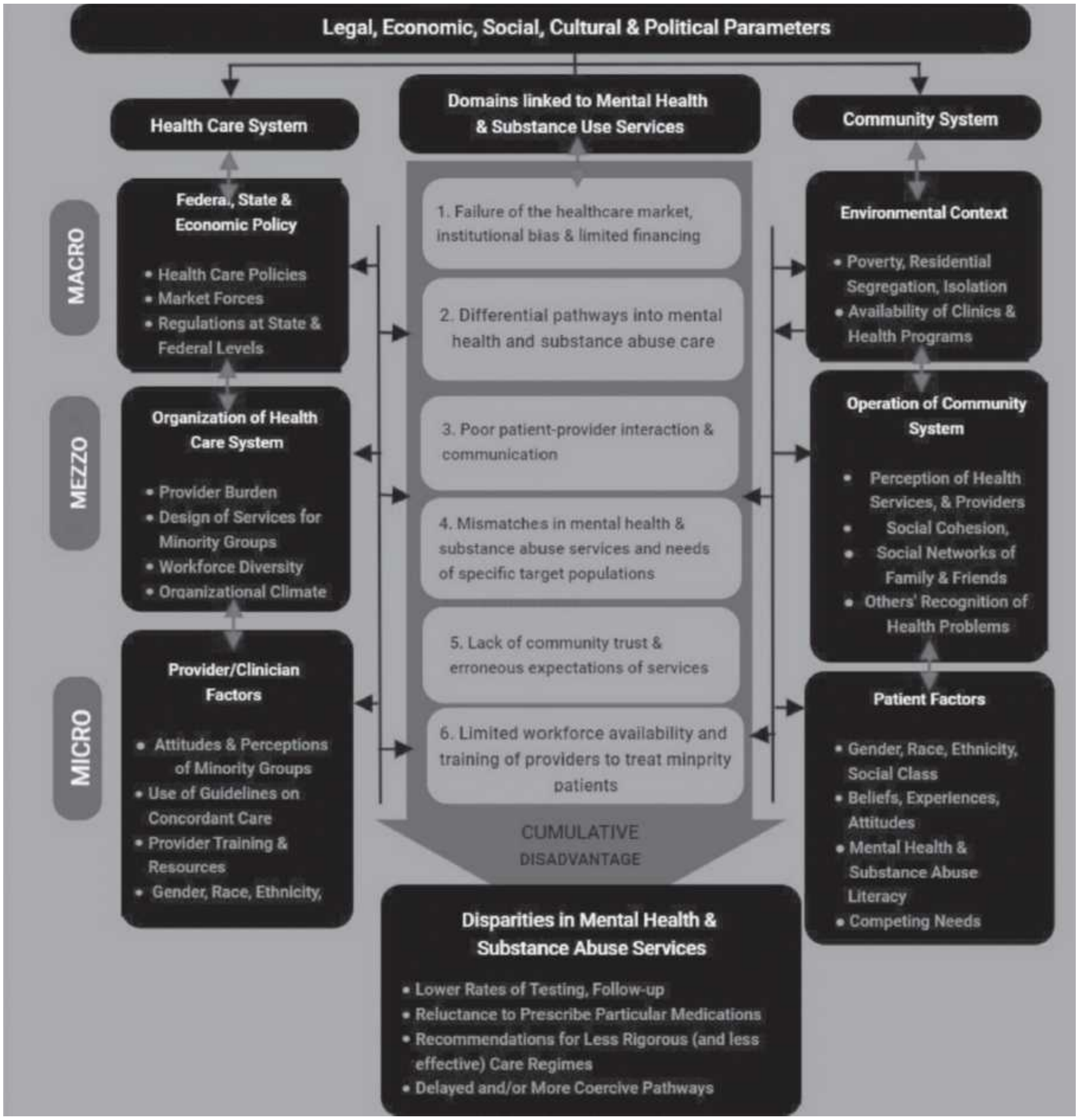
Legal, Economic, Social, Cultural & Political Parameters
Health Care System
Domains linked to Mental Health & Substance Use Services
Community System
MACRO
MEZZO
MICRO
Federa., State & Economic Policy
• Health Care Policies
• Market Forces
• Regulations at State & Federal Levels
Organization of Health Care System
• Provider Burden
• Design of Services for Minority Groups
• Workforce Diversity
• Organizational Climate
Provider/Clinician Factors
• Attitudes & Perceptions of Minority Groups
• Use of Guidelines on Concordant Care
• Provider Training & Resources
• Gender, Race, Ethnicity,
1. Failure of the healthcare market, institutional bias & limited financing
2. Differential pathways into mental health and substance abuse care
3. Poor patient-provider interaction & communication
4. Mismatches in mental health & substance abuse services and needs of specific target populations
5. Lack of community trust & erroneous expectations of services
6. Limited workforce availability and training of providers to treat minprity patients
CUMULATIVE
DISADVANTAGE
Environmental Context
• Poverty, Residential Segregation, Isolation
• Availability of Clinics & Health Programs
Operation of Community System
• Perception of Health Services, & Providers
• Social Cohesion,
• Social Networks of Family & Friends
• Others' Recognition of Health Problems
Patient Factors
• Gender, Race, Ethnicity, Social Class
• Beliefs, Experiences, Attitudes
• Mental Health & Substance Abuse Literacy
• Competing Needs
Disparities in Mental Health & Substance Abuse Services
• Lower Rates of Testing, Follow-up
• Reluctance to Prescribe Particular Medications
• Recommendations for Less Rigorous (and less effective) Care Regimes
• Delayed and/or More Coercive Pathways

FIGURE 56.7–1. A sociocultural framework for mental health and substance abuse service inequities. (Adapted from Alegria M, Pescosolido BA, Williams S, Canino G. Culture, race/ethnicity and disparities: Fleshing out the socio-cultural framework for health services disparities. In: Pescosolido B, Martin J, McLeod J, Rogers A, eds. *Handbook of the Sociology of Health, Illness, and Healing*. Handbooks of Sociology and Social Research. Springer, New York, NY.)

We follow the Institute of Medicine's (IOM) *Unequal Treatment* Report in focusing on three levels within the health care system:

1. Federal and economic health care policies and regulations at the state and federal level
2. The operation of the health care system and provider organization
3. Provider/clinician-level factors.

These domains are depicted on the left-hand side of Figure 56.7–1. The SCF extends the IOM framework by adding the parallel focus to the other key system—the community system. These additional levels are drawn from utilization, “take-up” or help-seeking models and include:

1. The environmental level, specifically social and economic forces affecting minority populations
2. The operation of the community networks, including family, friends, and the lay sector
3. Patient-level factors.

These domains are depicted on the right-hand side of Figure 56.7–1.

The touchpoint of the two systems at each contextual level produces six main mechanisms listed in the center portion of Figure 56.7–1. These include:

1. Failure of health care markets, institutional bias, limited financing, and lack of ancillary services leading to decreased access and quality of mental health care and substance abuse treatment services
2. Differential pathways into mental and substance abuse health care leading to differential experiences of treatment, no treatment, or inadequate treatment
3. Poor patient–provider interaction leading to miscommunication, clinical uncertainty, low therapeutic alliance, bias, stereotyping, and clinical errors; and high dropout rates
4. Mismatches in mental health and substance abuse service offerings and minorities’ needs leading to resistance or lack of compliance in health care and under-treatment
5. Lack of community trust and erroneous expectations of services, leading to disengagement and poor collaboration in treatment
6. Limited workforce availability and training of mental health and substance abuse service providers to treat ethnic and racial minorities, leading to reliance on informal or poorly trained providers.

In sum, at each level, the health care system and community system interact at diverse levels to produce mechanisms that lead to differential health care outcomes in mental health and substance use care, including greater illness severity and diminished quality of life. The levels and systems are linked by and through these mechanisms in a complex population health system. All mechanisms and contexts are set within a shaded region that represents culture. Below, we discuss how the interaction between systems works at each level and operates to impact mental health and substance abuse service inequities. We provide concrete examples of how the six main mechanisms may produce inequities. Because service inequities may also result from interactions between levels, some overlap is inevitable in our descriptions. Where appropriate, we bring in evidence-based concepts and findings not specific to mental health and substance abuse care to identify potential leverage points to reduce inequities that lead to these negative mental health and substance use outcomes or to create testable models tailored for diverse racial and ethnic groups. Finally, we acknowledge that the mechanisms linked to access may differ from those leading to poor quality of care or differential treatment outcomes.

HEALTH CARE SYSTEM DOMAINS

Federal, State, and Economic Policy

Federal, state, and economic policies impacting health and mental health care have the potential to create, exacerbate, or reduce health inequities. Due to their over-representation in government-sponsored programs like Medicaid, populations may face additional challenges as a result of state and federal regulations aimed toward cost control. For example, state differences in implementation of the Patient Protection and Affordable Care Act (ACA), federal financing of Medicaid payments, and the Children's Health Insurance Program (CHIP), all have a disproportionate effect on minority populations. We discuss each, in turn, below.

Affordable Care Act and Medicaid Expansion. A key provision of the ACA is the expansion of Medicaid to include adults with incomes up to 138% of the federal poverty level (FPL, $16,104 for an individual or $27,310 for a family of three in 2014), in order to increase the number of Americans who are insured. A review of 65 studies from 2014 to 2020 found that states which expanded Medicaid significantly narrowed racial and ethnic disparities in coverage among Latinx and Black individuals compared to White individuals. Few studies examined effects among other racial and ethnic groups, which is an important gap in the research. Medicaid expansion improved disparities in some indicators of access to care, quality of care, and health outcomes (e.g., infant and maternal health). Substance abuse treatment increased, with 36% more people receiving treatment after 4 years of the expansions in expansion states, compared to nonexpansion states. Despite ACA provisions, individual states are allowed to decline the Medicaid expansion, and there are still those who do not qualify for Medicaid in their state and those who earn just enough that they are ineligible for the insurance tax credits available. Since minorities tend to be overrepresented among the poor, the flexibility of states to set their own eligibility criteria has implications for minorities' access to and quality of care. An estimated 4 in 10 uninsured African Americans fall into the coverage gap. Latinxs, who have the highest percentage of uninsured when compared to other racial/ethnic groups, are also affected.

Immigrants of all backgrounds face particular challenges under current health insurance laws. Those who are lawfully present are subject to a 5-year waiting period for Medicaid and CHIP, as well as exclusions from these programs for some classes of immigration status. States may waive the waiting period for pregnant women and children, but not all states choose to do so. Undocumented immigrants are further disadvantaged, as they are not eligible for Medicaid at all, are not eligible for tax credits used to make private health insurance more affordable, and are not eligible to purchase health insurance via the federal health insurance exchange. Lack of access due to not having insurance serves as a potential mechanism promoting service inequities. Not surprisingly, it has been argued that expanding insurance coverage for minorities would partly reduce inequities.

Low Reimbursement Levels and Provider Shortages. A major challenge for Medicaid beneficiaries is having a sufficient number of providers that accept Medicaid and CHIP payments. Alexander and Schnell find that increasing payments for new patient office visits reduces reports of providers turning away beneficiaries. According to their results, closing the gap in payments between Medicaid and private insurers would reduce more than two-thirds of disparities in access among adults. The implementation of the Mental Health Parity and Addiction Equity Act (MHPAEA) was expected to reduce the large unmet need for services. According to statistics by the Medicaid and CHIP Payment and Access Commission (MACPAC), only 35.7% of psychiatrists accepted Medicaid reimbursement rates as of 2019. Low participation of physicians is particularly concerning given the expansion of Medicaid to 85.3 million people in poverty.

Low reimbursement rates, payment delays, and billing requirements have been noted as a primary reason for not taking Medicaid patients in physician surveys. Systemically, providers within organizations serving large numbers of Medicaid patients may experience high expectations for productivity in order to make services financially feasible, which can increase provider burnout. There may also be a need for providers to spend additional time coordinating care for more vulnerable patients. The implementation of Accountable Care Organizations (ACOs), which bring together groups of providers and provide more flexible payment structures to incentivize positive outcomes, may be beneficial in reducing health care inequities. Expansion of safety-net providers such as Community Health Centers (CHCs), another provision of the ACA, is another avenue by which health inequities may be reduced due to their ability to promote screening, meet cultural sensitivity requirements, provide transportation assistance, and treat all patients regardless of immigration status or ability to pay. A review of mostly quasi-experimental studies from 2000 to 2017 found that greater funding for CHCs improved access to care (including mental health care), especially among low-income and racial/ethnic minority individuals. An additional consideration is that the long-term sustainability of CHCs is partially dependent on the revenue that they can continue to generate from patients, which is more challenging in states with high proportions of uninsured patients.

Children’s Health Insurance Programs. CHIP is designed to make free or low-cost health insurance available to children without insurance but who are ineligible for Medicaid. As such, it plays a crucial role in reducing access barriers that create health service inequities among minority youth. In the first 10 years after CHIP was implemented, rates of uninsured children dropped from 30% of Latino children to 20%, 19% of Black children to 12%, and 17% of Asian, Native Hawaiian, and Pacific

Islander children to 12%. A study of CHIP implementation in New York found that preexisting racial/ethnic inequities in access, unmet needs, and usual sources of care between white, Latino, and African American youth were virtually eliminated during the first year of enrollment. However, racial and ethnic minority children continue to be less likely to be enrolled in health insurance when compared to white children, even when eligible for CHIP, and disparities are substantial among Latinx children with immigrant parents. The Current Population Survey indicates that in 2018 the uninsured rate among Latinx children with only foreign-born parents (13.1%) was over triple the rate for non-Hispanic white children (3.8%) and 2.5 times higher than the rate among Hispanic children with at least one U.S.-born parent (5.2%). It was 4.6% among Black children.

In 2009, the Children's Health Insurance Authorization Act (CHIPRA) established provisions aimed at reducing health inequities, including providing outreach grants to increase enrollment, increasing federal matching rates for interpreters and translators, providing federal funding for medical care for immigrant children in states that waive the 5-year insurance waiting period, establishing quality of care measures to track health outcomes for children, and including specific provisions addressing health concerns that disproportionately affect low-income children and children of color. Across states that enacted the additional CHIPRA coverage for immigrant children, health insurance rates for this population increased by 14.5%, and unmet health needs decreased significantly.

Mental Health and Substance Abuse Parity. Laws aimed at achieving parity between insurance benefits for general health conditions and insurance benefits for mental health conditions have been enacted in various ways since 1996. In 2008, MHPAEA was passed, and the ACA subsequently expanded the reach of the act to more types of insurance plans. The MHPAEA does not require plans to offer specific mental health benefits but rather to ensure that the benefits offered are equal to those offered for other health conditions. Implementation of the law has been a complex process, resulting in final regulations being issued in 2013. While parity provisions are a promising step forward in access to care, it remains to be seen how the enactment of current regulations may or may not impact health service inequities. Poor behavioral health outcomes persist among people from minoritized, marginalized, and oppressed populations;

thus, it is time for the model of care and advocacy in the behavioral health field to shift its priorities from parity to equity.

National Standards for Cultural and Linguistically Appropriate Services. Policies can directly address issues relevant to minorities. The National Standards for Cultural and Linguistically Appropriate Services (CLAS) were published in 2000 by the Department of Health and Human Services (DHHS) and revised in 2013. The CLAS standards provide guidelines for health care organizations addressing three areas: governance, leadership, and workforce; communication and language assistance; and engagement, continuous improvement, and accountability. These standards have been used to shape service delivery systems, as well as by some states to propose or pass legislation. Organizations are urged to provide trained interpreters, as well as linguistically and culturally adapted medical forms for non-English speakers. Among public mental health agencies in 43 states and Washington, DC surveyed in 2016, 10 had no policies corresponding to the 3 CLAS national standards, and 8 had policies in either 1 or 2 of the domains. Twenty-five agencies had policies in all 3 domains; only 8 of these had adopted all 15 CLAS standards. Improving communication within health care organizations and consistency of accountability measures may contribute to more widespread implementation. While the CLAS standards are not a federal law, failure to meet language access recommendations may be a violation of Title VI of the Civil Rights Act of 1964, which prohibits discrimination in organizations receiving federal funding. DHHS-funded organizations must provide meaningful access to services to non-English-speakers, including trained interpreters.

Policy Supporting Innovation. Policy changes can also facilitate the use of innovative strategies to address health inequities. In 2013, the Centers for Medicare and Medicaid Services adopted a regulation change allowing states to choose to reimburse for physical and mental health services provided by unlicensed providers, as long as a physician or other licensed practitioner has recommended the service. This opens a possibility for states to fill in gaps in service availability by engaging providers such as community health workers (CHWs) and other lay providers (e.g., addiction recovery coaches or peer support specialists). The use of lay providers to deliver mental health services has been successful in several major studies in underresourced communities within the United States. Further research is needed to examine the extent to which evidence-based interventions delivered by lay providers can increase equity in access *and quality* of behavioral health care.

Technology used to track the delivery of patient care and implement quality measures is another promising avenue via which to reduce health inequities. Advances in health information technology, including electronic medical records, have been a critical component of national health policies. The ACA calls for improved reporting requirements, while CHIPRA has promoted the development of standard quality measures as a way of tracking and addressing health care needs. The IOM noted in a 2009 report that more detailed information on racial, ethnic, and linguistic background collected in public health data systems would help target efforts to eliminate health service inequities. Little had changed by 2020 when many states' incomplete reporting of racial and ethnic data related to COVID-19 cases contributed to the overall delayed and inadequate response in addressing racial disparities. Combining detailed data collection with quality measures allows to identify gaps in care for specific groups and to respond with specific programmatic improvements.

In sum, the larger parameters of society shape the health care system. At the federal and state level, specific policies on health care, as well as economic policies and regulations, create a set of market forces. In turn, the nature of these health care system policies interacts with the contours of the community system to impact the experiences of ethnic and racial minorities. We describe both the intended and unintended consequences of these policies on health care organizations next.

Operation of the Health Care System and Provider Organization

As a result of the system-level factors described above, health care systems face differential costs and opportunities associated with treating ethnic and racial minority populations. As noted above, reimbursement by the state or the federal government holds a central role in shaping provider burden, service design, workforce diversity, and the organizational climate of health care organizations.

Cultural and Linguistic Competency in the Operation of the Health Care System. Health care policies designed to encourage cultural and linguistic competence, such as the CLAS guidelines, are expressly designed to counter discrimination. Yet, the ways in which a policy is implemented by organizations and providers impact client care. Language barriers for non-English speaking immigrant groups create major challenges in access to care. In one study alone, language barriers between provider and patient accounted for a $38 increase in charges for testing and increased length of stay in the emergency department. Patients who speak a different language from their providers report worse outcomes and higher dropout rates and may experience misdiagnosis or poor treatment based on misunderstandings in assessment. The use of interpreter services by competent providers has been shown in some instances to result in lower-quality care when the interpreter is not knowledgeable about mental health conditions or tries to make their own interpretations. However, a study of 452 Spanish-speaking patients seen for behavioral health visits in primary care found that interpreter use did not impact therapeutic alliance when compared to working with bilingual providers. Qualitative data from a subset of patients indicated a preference for working with bilingual providers when available, but found patients were satisfied with behavioral health care services delivered via interpreters and saw them as a way to increase access to care.

At the same time, improving linguistic competence within a system results in additional costs and time investment. Health systems may be reluctant or simply unable to cover these resources. Furthermore, even if willing, individuals may not be readily available or trained to fulfill the required services. Medical interpreters may not receive specific training in mental health care interpretation. Policy efforts such as state legislation based on the CLAS standards may be beneficial in reducing health inequities, but resources and training must be available to health care systems to finance and implement organizational change. A review of health care organizations' implementation of the CLAS standards suggested that barriers to CLAS implementation include a lack of awareness of the benefits of CLAS, perceived complexity of implementation, and organizational resistance to change. Recommendations identified across studies included providing sufficient language access services (i.e., professional interpretation and translation); ensuring data collection on race, ethnicity, and language as related to patient services; and obtaining buy-in from leaders and administrators. Community engagement, workforce development and training, and strategic implementation across all organizational levels were also highlighted.

Integrated Health Care. The integration of behavioral health services into the primary care setting has been advanced as a strategy to improve behavioral health outcomes for minority populations, given studies that indicate that members of racial and ethnic minority groups are more likely to seek mental health care via their primary care providers (PCPs). Yet, only limited data are currently available on how to best implement high-quality behavioral health services for minority populations in primary care. Studies have found that PCPs are less likely to detect mental health concerns among racial and ethnic minority patients compared to white patients. The pressure to treat behavioral health challenges in the time-limited setting of primary care, in addition to constraints in specialty care options available for PCP referral, may result in situations in which clients are offered lower-quality services or do not follow up on treatment. One study found that Latino primary care patients diagnosed with depression were more likely to follow up on and complete treatment via medication alone, despite the majority of the patients expressing a preference for psychotherapy over medication. These patients were able to access medication directly via their PCP but may have experienced challenges in accessing therapy services.

Research on models of integrated and collaborative care, as well as their impact on specific racial and ethnic minority populations, is needed in order to explore how this approach may or may not reduce health service inequities. As one of the most studied models of integrated care, collaborative care shows promise as a strategy to reduce inequities. A review of studies of collaborative care for depression among limited English proficiency patients (LEP) concluded that collaborative care delivered by bilingual providers was more effective than usual care. Weaknesses in the evidence base included a low number of studies (particularly for languages other than Spanish), lack of information about whether cultural adaptations were made in addition to language adaptations, and lack of research examining why LEP patients are less likely than English-proficient patients to be enrolled in collaborative care programs. A review of collaborative care for racial/ethnic minority patients similarly found it to be a promising approach to improve depression outcomes but found that studies encountered challenges in recruitment of racial/ethnic minority patients.

In other ways, the design of health care systems and services contributes to inequities. When health care organizations do not offer effective treatment for behavioral health problems, issues such as alcoholism, substance dependency, or posttraumatic stress disorder affect health status directly and compound other health problems indirectly. Further, offering a service-level spectrum, whereby good quality health services are offered for services that attract the healthy, but poor-quality services are offered for the sick or those with chronic conditions (referred to as “adverse selection”), tends to discourage use and hinders good outcomes.

Organizational Climate and Institutional Racism. Outcomes research has demonstrated that the climate of attitudes and beliefs and culture (both normative expectations and actual behaviors) in the provider organization toward minoritized populations affects the receipt of quality care. For example, epidemiologic evidence indicates that racial and ethnic minority patients have lower rates of mental health disorders overall but higher levels of severe illness and a lower likelihood of receiving mental health care, reflecting structural inequity. As a result, racial and ethnic minority patients may come to mental health and substance abuse care at higher levels of illness severity and terminate treatment early. This may reinforce a cycle of bias and discrimination, in which providers view patients as requiring more resources and intensive treatment or are seen as typically less adherent. In addition, these patients may also require referrals and networking with other social and service agencies—work for which the health care provider is not compensated. Rather than recognizing structural racism as a root cause of these inequities, providers may maintain negative attitudes towards patients or ascribe cultural explanations to structural conditions. Structural competency has been advanced as a paradigm shift in medical education and practice, away from a focus on stigma and cultural competency towards a focus on recognizing how structures, including social and economic forces, shape clinical interactions and how these conditions in turn call for structural interventions. At the organizational level, the use of evidence-based strategies for organizational change, such as the ARC (Availability, Responsiveness, Continuity) intervention model, may be an avenue by which to change organizational attitudes. In this model, “change agents” are trained to educate providers, advocate for systems-level change, provide implementation data and feedback, and facilitate communication within the organization and with the broader community.

Design of Services for Minority Groups. While the evidence is inconclusive on whether the lack of targeted services for diverse patient populations plays a central role in service inequities, a meta-analysis showed that culturally appropriate interventions were four times more effective than those not targeted to specific cultural groups. Further, interventions conducted in the client's native language were twice as effective as those conducted in English. Yet, there is a serious shortage of culturally appropriate behavioral health services for ethnic minority youths. In a review of psychosocial interventions, Pina and colleagues found that there were no well-established interventions effective for treating trauma or stress reactions, self-injurious behavior and suicidality, or comorbid disorders or multidisorders among ethnic minority youth. There is also a need to address the disproportionately high rates of suicidal behavior and homelessness among BIPOC youth who are LGBTQ. Policies cocreated in partnership with these youth are likely to be more effective in meeting their needs.

Structuring for racial and ethnic concordance or matching between patient and provider is one area of heated debate. Race concordance for African Americans has been associated with higher patient participation in decision-making and greater satisfaction with care. Patient–provider matching also tends to improve retention of clients and intervention outcomes and may increase entry and adherence to treatment. If this is, indeed, the case, professional services in culture-specific clinics might be a tool for providing better services to ethnic minorities. However, there is limited evidence on whether or not they reduce service inequities. While ethnic similarity may facilitate retention in therapy, it may have only a minor role in therapy outcomes. Similarly, physician's race concordance with patients does not appear to be associated with either the use or quality of care, only with preventive care.

Clients are likely to respond to the therapy in ways that are consistent with their cultural socialization regarding care. Therefore, culture most likely defines the options perceived to be available, useful, and/or desirable behaviors of what is culturally expected of the therapist and client roles. Patient preference for one theoretical orientation over another might characterize members of cultural groups. Researchers hypothesize that therapists and clients who are ethnically similar may be more analogous in terms of problem perception and attitudes about coping and treatment goals, and some studies have found that patients delay treatment based on the perception that an ethnically dissimilar clinician will not understand their needs. As a result, these coinciding factors may improve the effectiveness of therapist/client interactions.

This is not to say that all racially/ethnically concordant patients and providers will share the same attitudes and beliefs. Research has demonstrated that *cognitive* match between patient and provider is most important, with evidence of this significance persistent even after controlling for ethnic match and the client's language preference. These findings have implications for practice/training and for the need to assess discrepant orientation and treatment ideas. Across studies, reasons for patients not initiating treatment included interpreting that the mental health system could not help them, the belief that somatic symptoms were not related to mental health conditions, and stigma (though interpretations of stigma varied). Reasons for not continuing treatment included feeling discomfort discussing emotions with strangers, preferring a particular clinician communication style, and having concerns about clinician power and authority.

Provider-Level Factors

Use of Guideline Concordant or Evidence-Based Treatment. Providers' reluctance to adopt evidence-based treatments (EBT) when serving minorities has been well-documented, leading to service inequities. The question remains as to whether the deviation from EBT is greater among providers who serve primarily ethnic and racial minorities or whether providers are less likely to provide EBT to their minority clients. Evidence has been found for both; clients who are poor or African American and Latinxs have been found to be less likely to be offered the more rigorous or intensive care options provided to those who were more advantaged, but also facilities serving mostly ethnic and racial minority patients were less likely to offer evidence-based care. For example, a study of services received by low-income adult patients with anxiety disorders in a community mental health center found that the majority of patients received pharmacotherapy, and that of the minority receiving psychosocial treatment, they were receiving supportive therapy rather than the established evidence-based practice of cognitive-behavioral therapy (CBT). Only 3.8% of patients with an anxiety disorder received the most effective treatment, exposure-based CBT. However, many health care institutions do not provide sufficient support for providers to adopt quality behavioral health care within the pressures for productivity.

Research examining the use of medical procedures across chronic illnesses, such as depression, has consistently found a lack of referral of minorities for EBTs when compared with non-Latino whites, even when minorities meet clinical appropriateness criteria for EBTs and when age, gender, and insurance status are controlled. Although longitudinal studies examining changes in trends for inequities in EBT among Medicare-managed care patients indicate that racial inequities in EBT have declined for some illnesses, for mental health care, the gap in EBT for minority patients remains.

Provider's Attitudes and Perceptions: Uncertainty, Stereotyping, and Bias. Providers seldom receive training issues surrounding stigma, either in the community system or in the health care system. As a result, there is reason to assume that they carry the prejudices of the communities they live in and the health care systems in which they work. Stigma attributes causes of behavior (genetics, socialization, character), assign characteristics to individuals (e.g., violent, unintelligent), and ascribes beliefs about health behaviors (noncompliant) that are embedded in individuals, organizations, and places. Stigma interferes with how the problems of ethnic and minority group members are conceptualized, how treatments are planned, and what expectations are held for recovery. From the start, providers do not have the cultural tools to motivate minority patients to voice their symptoms, treatment concerns, or preferences. Providers display fewer patient-centered behaviors, are less receptive to question asking, and tend to demonstrate greater verbal dominance with minorities than with white patients. This can lead to miscommunication and misunderstanding between the patient and provider, resulting from cultural or language differences. In turn, this contributes to uncertainty in the clinical encounter, which is well documented to increase provider hesitancy in treatment which may manifest in disparities in clinical decision-making, treatment planning, care provided, and outcomes. When cultural differences have diagnostic implications for clinical symptoms or for signs or tests due to atypical symptom presentation, providers have incomplete information about the efficacy of a therapeutic intervention. This is not surprising because research on these deep cultural differences is not supported or regarded as highly as standard treatment-based research, and, even in the latter, minorities are under-represented in clinical trials.

Stereotyping of minoritized patients by health care professionals, documented in a number of recent studies, is another potential source of service disparities. Even if stereotyping stems from unconscious bias, reflecting limitations in cognitive mechanisms, the effects on inequities are real, nonetheless. Categorizations such as race, age, sex, and socioeconomic status are overgeneralizations and produce statistical discrimination, which goes directly against the more tailored precision approaches. For example, providers tend to see Black patients as less educated than non-Latino white patients. Providers' implicit bias has consistently been linked to worse perceptions of interactions with health care providers among African American patients. Finally, provider-based biases have also been posited as negatively influencing clinical decision-making.

Both over and underdiagnosis could result from these stereotypes and negative biases. Overdiagnosis occurs when appropriate behaviors in one culture are interpreted as pathologic by a professional from another culture. Incorrectly diagnosing a psychotic disorder where none exists could result in inappropriate hospitalization or in the use of potentially dangerous antipsychotic medication. For example, in more extreme cases, certain cultural or religious beliefs and practices common among Latinxs and African Americans accept, and sometimes encourage, visual and hearing perceptions of supernatural experiences, which are interpreted by professionals in the dominant culture as pathologic hallucinations. In an epidemiologic study in Puerto Rico, 13% of the population reported some kind of hallucination. Yet, a qualitative analysis of these symptoms by trained local psychiatrists revealed that they occurred mostly in the context of religious or spiritualist experiences and were not pathologic. Not surprisingly, recent epidemiologic and clinical research has raised questions about the relationship between the endorsement of psychotic symptoms and the diagnosis of psychotic disorders among culturally diverse groups.

Underdiagnosis occurs when clinicians ignore genuine manifestations of mental illness, misguided by a misinterpretation of cultural sensitivity. The absence of cultural and contextual information contributes to an elevated probability of misdiagnoses and inappropriate treatment formulation. A growing body of research has documented that minority patients were more likely to report that their provider did not discuss end-of-life care and were less likely to receive recommendations for how to quit smoking or engage in diet and exercise. Providers may unintentionally or intentionally dismiss minority patients’ concerns or may not communicate about health promotion, psychoeducation, or early screening for mental disorders. Only one-quarter of Latinxs surveyed in the National Latino and Asian American Study reported ever being asked about alcohol or drugs. Fewer than 20% of Latino patients reported ever being asked about emotions, nerves, or mental health problems. This limited inquiry may lead to missed opportunities for early identification of mental health or substance use disorders and a lack of referral to specialty services.

Clinician Training in Cultural Responsiveness. The concern with provider-based bias does not translate into the notion that there are no beliefs in ethnic and racial communities that interfere with treatment. In a recent study on COVID-19 vaccination intentions, the fact that individuals from diverse racial and ethnic groups worked in a health care setting did not offset disparities. Asian, Black, and Latinx individuals, as well as those who identified as multiracial, expressed a lower likelihood of vaccine uptake. This provides a clear caution about assuming the health stance of individuals based on relevant sociodemographic categories. While cultural responsiveness training has been offered as a solution to the problems described above, it is not yet an integral part of the education of many health professionals. Health care personnel typically have limited training in the cross-cultural formulation and cultural humility, leading to a lack of concordance in defining the problem and treatment goals. There is also evidence that providers signal their expectations about patients’ outcomes and whether they believe there will be improvements, which may be related to actual patient health outcomes. For example, health care professionals may communicate lower expectations of treatment outcomes for socially and economically disadvantaged patients, which could discourage minority patients and lead to poor outcomes and low treatment retention. Furthermore, clinicians and researchers have called for a move beyond cultural competency to structural competency, calling for increased focus in medical education and health care practice on understanding how social, economic, and political forces, including structural racism, produce downstream health care conditions.

THE COMMUNITY SYSTEM

Community systems differ dramatically in a series of characteristics that both can facilitate the use of services and offer protection against health problems. In the SCF, the community is defined as “a group of people with a similar cultural origin and diverse characteristics that are linked by social ties, share common perspectives, and engage in joint action in geographical locations or settings.”

Environmental Level

Neighborhood context influences both individual health and health behaviors. Residing in socioeconomically disadvantaged neighborhoods has been linked to poor health outcomes due to built-in environment factors (e.g., increased exposure to pollutants and decreased access to green spaces); limited access to health care facilities and health care providers; poor quality of available health care; stress related to poverty and neighborhood violence; limited opportunity for economic mobility; disproportionate contact with the criminal legal system; and decreased social connectedness. These circumstances are not accidental. Political policies, such as redlining predominantly Black neighborhoods, decreased residents’ access to secure mortgages and opened up their neighborhoods to natural risks (e.g., flooding).

The geographical area in which minority populations live helps to create a social context which, in itself, fosters a lack of connection with treatment systems and the dominant culture. Geographical segregation continues, perhaps even at greater levels than would be expected for minority populations. In a nationwide study of data from over 25 years, Chetty et al. found the black–white gap in income mobility persists even among boys who grew up in the same neighborhood. Fewer than 5% of Black children grow up in low-poverty neighborhoods, which also have low levels of racial bias among whites and high rates of father presence. Members of immigrant groups have a variety of experiences with resettlement depending on the time period in which they arrived in the United States, the circumstances of entry, and the political and social climate towards the particular group at the time. Thus, the social context and the geographical context are often intertwined and overlapping. This impacts health on multiple levels, from the prevalence of particular conditions to access to and use of mental health and substance abuse services.

Level of Social Cohesion

Community social cohesion may play a role in health inequities by influencing factors that affect service outcomes. For example, there is evidence that organized and cohesive communities are able to ensure that state budget cuts do not affect their community and this, in turn, allows them to maintain increased access to funding for health and education. Neighborhood social cohesion has also been associated with greater use of preventive health care services. Beyond service use, social cohesion may impact health outcomes more directly as well—it has been shown to moderate the impact of neighborhood built environment (e.g., structural disadvantage, urban green space) on mental health.

Social Capital and Social Networks

While it is difficult to draw a clear line between the “community” as “place” and the “community” as “people,” the difference is important. While a place can shape the personal connections that individuals have, social scientists also focus on the community as individuals’ “small worlds”—that is, the actual nature of social ties and social interactions that individuals have in a geographic area. Family, friends, and even casual contacts can recognize a problem, may facilitate or inhibit help-seeking, and may influence adherence in future medical contacts, filling prescriptions, and adhering to lifestyle advice. Community ties shape how people recognize, interpret, and guide an individual’s reaction to the onset of mental health and substance abuse problems. They also continue to exert influence throughout the entire illness career. That is, not only do social networks constitute individuals’ “community of meaning,” but they are also their “community of influence.” Research has found social networks to play a key role in whether individuals deal with mental health and substance use problems or dismiss them, whether they rely on the family alone or seek professional help, and whether they are welcomed back into the community or rejected. All considered, social networks are critical in the course and outcome, especially in recovery from substance use disorders. For individuals from cultures that are wary of the formal treatment system and who are immersed in a strong family system, this can result in delays in care and/or the burnout of their social networks. However, when the health care system has a known concern for and tradition of understanding the population groups under their purview, social networks can facilitate care, communication, and recovery.

Service systems working with minoritized families need to consider the important role of extended family and “fictive kin,” that is, individuals who are not related by marriage, birth, or adoption but who are also possible sources of support or burden in treatment. For example, in strong and tightly meshed family networks, such as those of most Latinxs, African Americans, and Asians, individuals may initially be discouraged from seeking professional health services, feeling that the family should handle the situation. Only when family support systems are lacking, or become ineffectual in handling a crisis, do some culturally diverse families turn to professional help, even then activating social networks contacts to find trusted sources of care. Yet family support and supportive social networks have also been commonly shown to *facilitate* service use among African American adolescents, particularly when caregivers have had positive experiences with mental health services and/or refer their children.

Access to mental health services is a *process* that is relational and socially embedded. One way that nonprofessional support can lead to formal support is through group social identification. Self-identifying as a person with a mental illness or addiction and identifying with other people who have the same condition can promote professional help-seeking, in part by reducing stigma. Innovative strategies that leverage the influence of community networks can facilitate health and health care. Certainly, the role of churches and community centers in African American neighborhoods has been leveraged as a way to reduce vaccine hesitancy during the COVID-19 pandemic, as was the case for HIV education and testing. Trusted community organizations offer a way to embed health care services in the settings in which individuals and families naturally seek help.

Individual Factors Other than Need or Clinical Appropriateness

General Education. The number of years of schooling is directly linked to the development of effective capabilities that influence subsequent life conditions that have an effect on health, including economic welfare, healthy lifestyles, and full-time employment. Higher rates of education have been shown to positively impact mental health by increasing one’s resources to cope with adversity or increasing a sense of control over one's destiny. Low educational attainment is not only a strong predictor of poor mental health and substance abuse outcomes for children but remains significant for mid-age and older cohorts.

Health Literacy and Numeracy. Health literacy has two components—knowledge of health issues and knowledge of what to do if a health problem arises. The first is a skill involving both oral and written communication that may or may not be related to educational attainment. It is a twofold concept that includes the individuals' basic literacy skills and the application of those skills within the health care system. A related concept, health numeracy, involves not only understanding mathematics but also quantitative concepts such as "reduce/increase." The IOM has highlighted that health literacy is not solely an individual skill; rather, it results from the interaction between individual capacity and the demands of the health care system. It can also result from what "others" around individuals know. Inequality in educational opportunities and resources also contributes to why public health researchers conceptualize health literacy as a "midstream" determinant of health.

Health literacy may indirectly create barriers to health care use by inhibiting access to care due to a patient's inability to register for health insurance; difficulty in interpreting coverage benefits and rules; avoidance or delays in accessing health services due to the inability to adequately fill out forms in physician offices or difficulty following directions to service facilities. Once individuals with poor health literacy enter care, they face inequities in health outcomes associated with literacy limitations, with people who read at lower levels being 1.3 to 3 times more likely to experience adverse health outcomes than individuals with high literacy levels. They are also more likely to report poor health, have an increased risk of hospital admission, and have incomplete understanding of their health problems and treatment.

On the second dimension of health literacy, awareness of relevant resources, inequities in health care among some racial and ethnic minority children appear to be linked to parental knowledge of the disorders and available treatments. Zuckerman and colleagues have found that compared with non-Latino white parents, parents who were Latino or other ethnicities had less familiarity or personal experience with developmental disabilities, like Attention Deficit and Hyperactivity Disorder (ADHD) and Autism Spectrum disorder (ASD). One study suggests that parents' level of English proficiency and parents' knowledge about ASD are related to disparities in Latinx children's ASD diagnosis and treatment in the United States. There is also evidence that different ethnic groups in the United States have varied perceptions of what is considered a mental health problem in their offspring. In one study, Latino and African American parents rated their adolescents' mental health and life satisfaction as better than their non-Latino white counterparts and overall perceived fewer emotional or behavioral problems in their children. These differences existed even after adjusting for the clinical needs of the child, the child's level of impairment, and the sociodemographic characteristics of the parents. Thus, while a lack of knowledge about a problem may lead to health inequities, it is also possible that different perspectives on a situation may lead to different patterns of seeking care.

Negative Provider Stereotypes. Patient views of providers may prove to be influential factors in causing inequities for racial and ethnic minority populations. For example, research revealed that individuals holding strong negative stereotypes of physicians are more likely to delay care or only seek treatment once symptoms become severe or impede daily functioning. Therefore, if delays in treatment are due in part to negative stereotypes, service inequities in illness prevention or early intervention may emerge. Negative stereotypes of physicians may result from patients' own history of negative treatment in care or the distrust many minorities feel towards the health care system. However, recent population-level research in the United States has documented very high endorsement of general and specialty providers for unlabeled situations that meet DSM criteria for psychiatric diagnoses of schizophrenia or depression. While there were few, if any, differences between majority and minority group populations, the response to individuals with alcohol dependence revealed continued and even increasing stigma.

Patient's Competing Needs. Some ethnic and racial minority patients may come from low-income backgrounds with complex social needs. Lack of access to paid sick leave, inflexible work scheduling, irregular hours, lack of transportation or unreliable transportation, and difficulty obtaining childcare all pose additional challenges to utilizing health care services. Studies have found that these barriers account for both missed appointments and lack of follow-through on specialty referrals.

WHEN HEALTH CARE SYSTEMS AND COMMUNITY SYSTEMS MEET: MECHANISMS PRODUCING INEQUITIES

Taken together, inequities arise when disadvantages in the health care system interact with those in the community. At each level, the SCF suggests that six fundamental mechanisms produce inequities for minoritized populations. These are represented as the center stream in Figure 56.7–1. While they link with causes of inequities, they also suggest points of intervention. They represent potential leverage points for systems, providers, and patients to change the health care system, existing policies, clinical training, and community programs. Below, each interface between community and treatment systems is explored.

1. Failure of health care markets, institutional bias, and limited financing leading to decreased access to and utilization of mental health and substance use services. State and federal regulations aimed toward cost control also contributed to the increase in the number of uninsured or underinsured. For example, individuals are typically excluded from Medicaid as states respond to increased costs by enacting more stringent eligibility criteria. Yet, the expansion of public insurance programs and subsidized private plans to low- and middle-income individuals offered under the ACA can increase access to care. Nonetheless, some minority groups may confront other hurdles to getting insurance, such as lack of citizenship or documentation status, or may still have difficulty paying for subsidized coverage, leading to a lack of insurance or underinsurance. While health insurance coverage facilitates access, it is not a sufficient condition for ensuring behavioral health care since lack of available appointments is a serious barrier to mental health and substance use treatment. The low number of mental health and substance use providers accepting Medicare and Medicaid reduces access to quality care for low-income and vulnerable populations. However, other factors such as increased HMO penetration, augmented provider supply, and the presence of a community mental health center in neighborhoods of color have been shown to reduce service inequities.
2. Differential pathways into services leading to differential experiences of service, delayed utilization, no service, or inadequate service. Language barriers can lead patients to drop out of care, opt for

alternative services, or delay care. Inaccuracies in assessment can also result in mistreatment or misdiagnosis of minority patients. There is mounting evidence that minority patients receive lower quality treatment and receive treatment for a shorter period of time. For example, white youth are more likely than minority patients to receive adequate care following a major depressive episode, while minority youth are less likely to be prescribed psychotropic medications, even after adjusting for need factors. Equally importantly, the larger cultural context determines how similar social networks can facilitate or delay treatment. Competing demands on time from work and childcare for minority parents may constrain minority families from engaging in treatment, given more urgent matters.

Thus, differential pathways into services may represent a key mechanism leading to service inequities among racial/ethnic minorities, with lack of patient problem recognition, differential referral and treatment by providers, and barriers to care serving as potential precipitating factors to inequities. Noting the insufficient attention to dynamics of pathways to treatment, Alang and McAlpine found that individuals who were “ordered” into care (i.e., followed a coercive or heavily recommended pathway) either for a mental illness or substance use disorder were less likely to rate treatment as effective compared to those who made an independent decision. Not surprisingly, they also found that individuals who are ordered into care are more likely to be African American (86% more likely) or Latinx (twice as likely), with the former significantly less likely to report effectiveness. They conclude that pathways to care reflect differential positions in the social hierarchy.

Other pathways result in inequities that may have both short- and long-term effects. For example, Latino and African American children are overrepresented in the juvenile social service sectors. They receive fewer therapeutic services compared to clinically comparable white children and remain longer in services. As a result, data analytics are being used to detect patients at risk of receiving lower-quality care based on demographic and illness-related factors, and payment systems used to incentivize providers for providing high-quality care to these patients. Nonetheless, there is considerable debate over the utility of pay-for-performance measures in achieving specific goals at the level of hospitals and the health system as a whole. ACOs may help eliminate inequities by rewarding health care systems for efficient use of resources and positive outcomes and penalizing them for inefficiency and poor results.

3. Poor patient–provider interaction leading to miscommunication, clinical uncertainty, low therapeutic alliance, stereotyping and clinical errors. Patient–provider match and expectations for a visit are important in determining treatment outcomes. However, attempts to address research documenting problems arising from poor patient–provider interaction have inadvertently produced additional problems. Ethnocentric models of health service delivery may overpathologize minority behaviors that differ from those of the majority culture. Racial and ethnic minority patients report being treated with disrespect or being looked down upon more often than non-Latino white patients. The critical link to inequities lies not only with what happens in the clinical encounter but with the resulting effects on the patient's utilization of health care and on the advice they offer to others in their lay network. For example, patients who reported being treated with disrespect were less likely to go for a routine physical exam within the last year, to have received optimal care for a chronic condition, and to have followed their doctor's advice.

   Patients may also have different expectations of treatment and provider behavior that impede effective communication. A number of studies have reported that differences in race and ethnicity between patients and their providers can present barriers to effective communication and partnerships for care. Inequities resulting from barriers to effective communication between racially/ethnically mismatched patients and providers include: discordance in the understanding that patients and providers have regarding disease causation and effectiveness of treatments, significantly lower reports of physician participatory decision-making styles by minority patients as compared to nonminority patients, and lower rates of receiving preventative and necessary medical care among Black patients with non-Black physicians compared to Black patients with Black physicians. These examples demonstrate that patient–provider match and expectations for a visit are important in determining treatment outcomes.

4. Mismatches in mental health and substance use service offerings and minorities' services needs and living circumstances leading to dropout and low or no service use. Services inequities have been linked to higher disability and greater disease burden in minority populations.

One of the mechanisms by which service inequities take place is through the mismatch of services to the patient’s needs. As we described earlier, there is substantial evidence of the potential for clinicians to engage in “statistical discrimination”—that is, to incorrectly assign a classification or social categorization to a patient that aligns with negative stereotypes, particularly when the patient’s request or symptoms are ambiguous. Under such conditions, clinicians may inadvertently incorporate attitudes and beliefs from prior clinical encounters or common stereotypes. This may lead clinicians to disregard individual data and consequently mismatch services to needs. This mismatch of services to needs can result in decreased utilization of and access to health services and health resources as well as differential outcomes in health and greater disease burden. In the absence of cultural and contextual information that supplants stereotypes, there is an elevated probability of misdiagnoses and inappropriate treatment formulation. Minoritized populations are more likely to have nonconcordant psychiatric diagnoses in psychiatric emergency services and to be undetected for mental health disorders in primary care. Research reveals that symptom presentation varies across racial and ethnic groups and can differ from what most clinicians are trained to expect. Not surprisingly, racial and ethnic minority patients are less likely to be offered “talk therapies” such as counseling and tend to be prescribed higher drug dosages than their white counterparts.

Dissimilarities in the typical conditions in which many minority populations live as contrasted to their providers (i.e., including poverty, lack of economic opportunity, food insufficiency, inadequate housing, social exclusion, low access to educational opportunities, and unsafe neighborhoods) can lead to social inequality that needs to be tackled by the health care system. As such, there has been a call for health equity research that prioritizes multilevel interventions, addressing structural and contextual factors at the community, organizational, and individual levels. Multilevel interventions to amplify social connection in minority neighborhoods could increase behaviors that have a positive or protective effect on health, such as exercise, expansion of social networks, and better problem-solving.

5. Lack of community trust and erroneous expectations of services leading to disengagement and poor collaboration in health care service systems. Local perceptions of health services affect utilization and

depend on prior experience with the provision of services and the efforts that the local service system has made to understand and adapt to the cultural and need demands of its clientele. Studies, including a qualitative analysis of an Afro-Caribbean community, revealed that their experience of mistreatment and social exclusion by health professionals reverberated on future utilization and on community sentiments toward the mental health system. In addition, African American families report fearing that their children would be institutionalized and express this fear as an important reason for not taking their children to treatment. Repeated negative contacts with the health system or with a lack of trust in key gateway providers such as teachers result in community-held beliefs that decrease help-seeking and produce health service inequities for minoritized populations.

6. Limited workforce availability and training of service providers to treat ethnic and racial minority youth leading to reliance on informal or poorly trained providers. The disproportionately low geographic supply of providers, particularly multilingual service providers, in communities with ethnic and racial minorities as compared to nonminority communities translates into market failure to provide more and better-quality care to minorities. For example, communities of color have been found to be four times more likely to have a shortage of providers, regardless of community income. Low reimbursement rates from public and private payers can lead to a shortage of services and provider burnout. Health care organization arrangements for distributing providers to minority Medicaid beneficiaries have also been suggested to be influential in-service inequities. Only 7.8% of psychologists, 12.9% of social workers, and 21.4% of psychiatrists are from racial/ethnic minority groups. Medicaid beneficiaries, particularly those with limited English language skills, may not be able to access physicians with a high caseload of privately insured patients.

Communities with high proportions of black and Latino residents were four times as likely as communities with non-Latino whites to have a shortage of physicians regardless of community income. In parallel, health care services available through the Indian Health Services vary widely across tribes. The Indian Health Services partially offsets the lack of insurance for some uninsured American Indian and Alaska Natives (AIANs), but important needs are left unmet. This is problematic since AIANs were more than two times as likely to be uninsured than non-Latino whites. These findings have led to the conclusion that the elimination of racial/ethnic health inequities is an issue of distributive justice which entails redistributing valuable material resources and opportunities to those in the lower echelons of society.

In sum, the interaction of the health care system and the community system at different levels produces a mismatch in cultures—not only the culture of medicine and the culture of everyday life but in the mismatch of dominant cultural ideas and those that may be unique to minoritized populations. Oftentimes, the beliefs of both patients and providers are based on generalized cultural beliefs and do not cut through those prejudices to get to individualized care. Equally important, both actual experiences and a family's, provider's, or community's experiences can produce a view of treatment that limits the receipt of effective care.

FUTURE DIRECTIONS: RESEARCH AND INTERVENTION

The SCF is intended to help conceptualize and carry out research directed at eliminating ethnic and racial inequities, to assist in the training of health and social agency personnel, and to supply service planners and policymakers working hypotheses, initially, and a science-base, eventually, on how to intervene and eliminate service inequities. As evidenced from our review of extant theory and research, traditional service models should be restructured in order to reduce service inequities for racial and ethnic minority populations.

However, it is important to understand the role of a framework and the difference between a framework and a model. As Nobel Laureate Elinor Ostrom pointed out, theory provides an understanding of a complex whole. A framework brings together extant findings from different branches of science to provide a picture of how the entire system operates and what are the key subsystems and factors at work in each subsystem. Doing so is complicated because different disciplines have their own very different frameworks, theories, and models that they employ to understand their parts of the complex system. But without an organizing whole, findings are not likely to cumulate, discrepant findings are not explicable, and unintended consequences in translation to policy and programs are more likely.

The framework, as an organizing picture, must be translated into a testable model for the particular problem at hand and the research question to be answered. The framework offers an overview of potentially relevant variables to be considered, orients researchers to issues in research design, data collection instruments and approaches, fieldwork protocol, and even potential analytic strategies. For example, a complex population systems framework aligns with the current discussion of the utility of multilevel modeling. This is a relatively new tool that now allows the consideration of different contexts in the same analyses. This is important because when there are nested levels (e.g., samples of individuals who live in different cities), these specialized analytic approaches, for example, correct for the violation of independence of cases assumed in regression analyses. Without these corrections, the determination of statistical significance is off. However, the data requirements to be able to analyze the effects of different contexts are very high, requiring a large number of contexts (e.g., neighborhoods, counties) at the higher level. It would be impossible, or nearly so, to develop a feasible research study that could incorporate everything in a framework.

That said, from the SCF, we posit six mechanisms that form a foundation for rethinking intervention strategies: (1) increased client/patient activation/empowerment; (2) information exchange between client and provider; between systems and communities addressing social networks rather than individuals alone; (3) new and/or different policies to increase access to services and quality of services; (4) better treatment matching; (5) restructuring of service systems to reduce reliance and action based on stereotypes, prejudice, and clinical uncertainty; and (6) quality collection of data on racial and ethnic disparities to improve system accountability.

1. **Patient activation.** Patient activation, defined as learning the knowledge, skills, attitudes, and self-awareness necessary to have greater decision-making and efficacy in health care behavior, has been associated with improved health behaviors such as exercise, disease self-management behaviors such as adherence to drug regimens, and an array of health outcomes including improved mental health. There is evidence that patient activation enables patients to take more charge of their health behavior, elicit more information from their provider, achieve greater satisfaction in the clinical encounter and better outcomes and increase engagement and retention in care.

2. **Information exchange.** Information exchange is based on a dialogue between patient and provider or between community and health service system so that the provider learns about the social context (i.e., where patients live, what they do, what matters to them, their social position, and their environment and cultural perspective of the patient). There is evidence that information exchange promotes a more egalitarian view of the information needed for a patient-centered encounter and to structure treatment recommendations or proposed behavioral changes acceptable in the patient's daily lives. If service use and provider action depend on a social influence process in the community as well as individual predispositions as has been documented since the 1970s, then understanding and using effective diffusion methods are likely to be more effective and efficient.

   However, information has a larger system component as well. Further, a recent *JAMA* editorial laid out the role of technology, especially the digital divide, in the differential rate of vaccination between minority and majority populations and among age groups. They noted that while there have been important recommendations offered to promote equity in the distribution of COVID-19 vaccines, they have missed the information component, which requires not only information and skills to sign up and do so quickly, but a recognition of the gaps in home-based broadband (40% among older people) and the greater likelihood of dead zones in minority communities. Yet, national guidelines and expert recommendations have neither considered nor proposed solutions to technology-related equity challenges.
3. **Policies.** Policies include decisions by organizations such as routinely providing assistance to patients in communicating with their providers or regulations by governments or public authorities such as deciding eligibility criteria for Medicaid services. Recent research has demonstrated the larger health effects of policy changes that are critical for disadvantaged populations. When the U.S. Department of Agriculture mandated that the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) move from a paper voucher to an electronic benefits transfer (EBT) debit card, participation in this program increased substantially (7/8%). This was, even more, the case in states that heretofore had the lowest WIC participation rates. This has direct implications for health. Research has clearly documented that WIC improves health outcomes for pregnant women,

infants, and young children. An example of a policy change could be a new requirement that interpreter services be readily available in the clinical encounter if English is not the primary language of the patient. Another policy to increase equity could be mapping areas where few mental health providers accept public insurance programs to provide additional incentives (i.e., higher reimbursement rates) for providers to offer care in those areas.

4. **Better treatment matching.** Better treatment matching includes the provision of linguistically appropriate services as well as treatment and outreach options that are socioculturally compatible with the circumstances of the target population. There are three basic criteria on which to judge whether there is treatment matching of services: the treatment/services are based on the cultural values of the groups of interest; the strategies involved in the treatment/services are congruent with the culture and context of the particular ethnic/racial group, and the components of the service/treatment strategy are based on the expectations and behavioral preferences of the ethnic/racial groups. It is argued that the treatment matching of the services/treatments with the patient/community is necessary for quality improvements to be institutionalized and sustained by the service organization and acceptable to the community.
5. **Restructuring of services.** Restructuring of services involves an organizational change to better accommodate the needs of clients. Settings that are accessible to communities and which they trust (i.e., community-based organizations, churches) might be more acceptable as settings for health or mental health service delivery or outreach. Changing stereotypes, prejudices, and uncertainty in provider–patient relations includes creating interventions to change preconceived ideas of either the patient or the provider that may adversely affect the communication or interaction in the clinical encounter.
6. Requiring public and private health systems to more thoroughly collect and report patient outcomes by race, ethnicity, and primary language. Requiring public and private health systems to more thoroughly collect and report patient outcomes by race, ethnicity, and primary language would improve transparency and accountability around health service inequities and the success of related equity-focused interventions. As research progresses in addressing the complexity of the levels and domains that produce health and health care inequities, such data can help equity interventions become more tailored and effective.

These six areas for intervention are just some of many potential ways to increase equity in mental health and substance use disorder service outcomes within the lens of the SCF. The overall goal of the SCF is to help stimulate the research and policy thinking that can foster change for health care systems and communities.

ACKNOWLEDGMENTS

Isabel O'Malley assisted with the research and revision of this chapter for the current edition. This paper revision was supported by NIH Research Grants #1R01MH117247, funded by the National Institute of Mental Health, Grant #1R01AG046149, funded by the National Institute of Aging, and Grant #1R01DA044526, funded by the National Institute of Drug Abuse.

FURTHER READINGS

Agency for Healthcare Research and Quality. *2018 National Healthcare Quality and Disparities Report*. 2019. https://www.ahrq.gov/research/findings/nhqrdr/nhqdr18/index.html

Aggarwal NK, Cedeno K, John D, Lewis-Fernandez R. Adoption of the national CLAS standards by state mental health agencies: a nationwide policy analysis. *Psychiatr Serv*. 2017;68(8):856–858.

Aggarwal NK, Pieh MC, Dixon L, Guarnaccia P, Alegria M, Lewis-Fernandez R. Clinician descriptions of communication strategies to improve treatment engagement by racial/ethnic minorities in mental health services: a systematic review. *Patient Educ Couns*. 2016;99(2):198–209.

Alang SM, McAlpine DD. Pathways to mental health services and perceptions about the effectiveness of treatment. *Soc Ment Health*. 2019;9(3):388–407.

Alexander D, Schnell M. The impacts of physician payments on patient access, use, and health. *National Bureau of Economic Research.* 2019 (0898-2937).

Alvidrez J, Castille D, Laude-Sharp M, Rosario A, Tabor D. The national institute on minority health and health disparities research framework. *Am J Public Health*. 2019;109(S1):S16–S20.

Barksdale CL, Rodick WH 3rd, Hopson R, Kenyon J, Green K, Jacobs CG. Literature review of the national CLAS standards: policy and practical implications in reducing health disparities. *J Racial Ethn Health Disparities*. 2017;4(4):632–647.

Barnett ML, Gonzalez A, Miranda J, Chavira DA, Lau AS. Mobilizing community health workers to address mental health disparities for underserved populations: a systematic review. *Adm Policy Ment Health*. 2018;45(2):195–211.

Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med*. 2011;155(2):97–107.

Chetty R, Hendren N, Jones MR, Porter SR. Race and economic opportunity in the United States: An intergenerational perspective. *Q J Econ*. 2020;135(2):711–783.

Garcia ME, Ochoa-Frongia L, Moise N, Aguilera A, Fernandez A. Collaborative care for depression among patients with limited English proficiency: a systematic review. *J Gen Intern Med*. 2018;33(3):347–357.

Guth M, Artiga S, Pham O. Effects of the ACA Medicaid Expansion on Racial Disparities in Health and Health Care. 2020. https://www.kff.org/report-section/effects-of-the-aca-medicaid-expansion-on-racial-disparities-in-health-and-health-care-issue-brief/

Guzman L, Chen Y, Thomson D. The Rate of Children Without Health Insurance Is Rising, Particularly among Latino Children of Immigrant Parents and White Children. 2020. https://www.hispanicresearchcenter.org/wp-content/uploads/2020/09/children-without-health-insurance-sep-2020.pdf

Holgash K, Heberlein M. Physician acceptance of new Medicaid patients: what matters and what doesn't. *Health Aff Blog*. 2019;2019:10.

Hu J, Wu T, Damodaran S, Tabb KM, Bauer A, Huang H. The effectiveness of collaborative care on depression outcomes for racial/ethnic minority populations in primary care: a systematic review. *Psychosomatics*. 2020;61(6):632–644.

Kim ES, Kawachi I. Perceived neighborhood social cohesion and preventive healthcare use. *Am J Prev Med*. 2017;53(2):e35–e40.

Krieger N. Discrimination and health inequities. *Int J Health Serv*. 2014;44(4):643–710.

Lamont M, Beljean S, Clair M. What is missing? Cultural processes and causal pathways to inequality. *Socioecon Rev*. 2014;12(3):573–608.

McGregor B, Li C, Baltrus P, et al. Racial and ethnic disparities in treatment and treatment type for depression in a national sample of Medicaid recipients. *Psychiatr Serv*. 2020;71(7):663–669.

Metzl JM, Hansen H. Structural competency: theorizing a new medical engagement with stigma and inequality. *Soc Sci Med*. 2014;103:126–133.

Nutbeam D, Lloyd JE. Understanding and responding to health literacy as a social determinant of health. *Annu Rev Public Health*. 2021;42:159–173.

Pescosolido BA. Of pride and prejudice: the role of sociology and social networks in integrating the health sciences. *J Health Soc Behav*. 2006;47(3):189–208.

Planey AM, Smith SM, Moore S, Walker TD. Barriers and facilitators to mental health help-seeking among African American youth and their families: a systematic review study. *Child Youth Serv Rev*. 2019;101:190–200.

Saloner B, Maclean JC. Specialty substance use disorder treatment admissions steadily increased in the four years after medicaid expansion. *Health Aff (Millwood)*. 2020;39(3):453–461.

Saloner B, Wilk AS, Levin J. Community health centers and access to care among underserved populations: a synthesis review. *Med Care Res Rev*. 2020;77(1):3–18.

Villalobos BT, Bridges AJ, Anastasia EA, Ojeda CA, Hernandez Rodriguez J, Gomez D. Effects of language concordance and interpreter use on therapeutic alliance in Spanish-speaking integrated behavioral health care patients. *Psychol Serv*. 2016;13(1):49–59.

Wolitzky-Taylor K, Zimmermann M, Arch JJ, De Guzman E, Lagomasino I. Has evidence-based psychosocial treatment for anxiety disorders permeated usual care in community mental health settings? *Behav Res Ther*. 2015;72:9–17.

Yearby R. Structural racism and health disparities: Reconfiguring the social determinants of health framework to include the root cause. *J Law Med Ethics*. 2020;48(3):518–526.
Zestcott CA, Blair IV, Stone J. Examining the presence, consequences, and reduction of implicit bias in health care: a narrative review. *Group Process Intergroup Relat*. 2016;19(4):528–542.
Zuckerman KE, Lindly OJ, Reyes NM, et al. Disparities in diagnosis and treatment of autism in Latino and non-Latino white families. *Pediatrics*. 2017;139(5):e20163010.

56.8 Criminalization of Persons With Serious Mental Illness

LINDA E. WEINBERGER, PH.D., AND H. RICHARD LAMB, M.D.

The enormously increased presence of persons with serious mental illness in the criminal justice system is one of the great problems of the day. As a result, mental health professionals and society have become much more concerned about the number of persons with mental illness in jails and prisons, as well as the treatment provided to these persons both in such facilities and after release. These issues are relatively recent. Reports of large numbers of persons with mental illness in U.S. jails and prisons began appearing in the 1970s, a phenomenon that had not been reported since the 19th century.

For the purposes of this section, the focus will be on persons with serious mental disorders, that is, schizophrenia, schizoaffective disorder, bipolar disorder, and major depressive disorder. This section discusses the definition of criminalization of such persons and how criminalization came about, the extent of criminalization, the role of the police with regard to persons with mental illness, and the characteristics that are often found in persons with mental illness who have been arrested. The difficulties in treating this population both before incarceration and after release are addressed, including the relationship between serious mental illness and violence and the problems presented by persons who are not able to recognize that they have a major mental illness. The section presents a summation of ways proposed to reduce criminalization. These ways include the following: increased community mental health services that are effective and appropriate to treat persons who are or are at risk of becoming criminalized; better coordination among the police, courts, and mental health professionals; mobile crisis teams; increased mental health training for police officers; and diversion programs for offenders with serious mental disorders to prevent them from entering the criminal justice system, as well as transferring those who are already arrested into the mental health system. The necessary components of treatment for such persons will be discussed. These include having clear goals of treatment and practice, establishing a close liaison between treatment staff and the justice system; understanding the need for structure; having a focus on managing violence; and appreciating the crucial role of case management, appropriate living arrangements, and involving family members. Finally, the section ends with a brief discussion of recently enacted criminal justice efforts to decrease the total number of persons incarcerated and recommends possible strategies to reduce the criminalization of persons with serious mental illness.

MAGNITUDE OF THE PROBLEM

The U.S. prison population for both federal and state prisons was 1,430,800 inmates as of December 31, 2019. In addition, at midyear 2019, there were 734,500 inmates held in county and city jails. Although these numbers reflect a small but continuing decline in the rate of incarcerations, the percentage of jail and prison inmates assumed to be seriously mentally ill (schizophrenia, schizoaffective disorder, major depressive disorder, and bipolar disorder) has been estimated at about 16%. Using these numbers yields an estimate of 346,448 incarcerated persons with serious mental illness in jails, and state and federal prisons. Although the actual number may be somewhat higher or lower, depending on the accuracy of the percentage, these figures represent a substantial number.

Moreover, research has shown that approximately one-fourth of persons with serious mental illness in the community have a history of at least one arrest. The association between serious mental illness and arrest is a widespread phenomenon that has affected both the mental health and criminal justice systems.

Ms. G was a 28-year-old single woman who was diagnosed with schizophrenia at age 22. Prior to her diagnosis, she had a chaotic childhood and adolescence. Her mother abused alcohol and often left her alone in their apartment for days at a time. Ms. G was placed in multiple foster home placements where she frequently eloped, looking for her mother. Attempts at reunification were always unsuccessful due to the mother's continuing neglect and Ms. G's increasing disruptive behavior. At age 18, Ms. G was no longer a ward of the court, and she began living with her 23-year-old boyfriend. She tried to work at fast-food restaurants but found the demands of the job increasingly stressful. She also started to experience paranoid ideas about her coworkers with whom she often fought and heard voices telling her that they wanted to sexually molest her. On one occasion, when she was 20, she fought with her manager and threw his glasses so hard against the wall that they broke. The police were called, and Ms. G was arrested.

During her incarceration, the jail personnel did not recognize that she had a mental illness. She was given probation with no requirement that she obtain psychiatric treatment. She continued to live with her boyfriend until he said she could no longer live with him because of her constant hallucinations and delusions about him. She was now homeless and without money. One day, she walked into a grocery store and stole a loaf of bread and a package of sliced turkey. She was apprehended by a security guard, and the police arrested her. During this incarceration, she was identified as having a mental illness and was treated with antipsychotic medication in jail. Ms. G received probation, but this time was to attend psychiatric outpatient treatment. She went to the community mental health clinic once but did not return because she did not believe she had a mental illness and refused to take any medications.

Shortly thereafter, Ms. G got into a fight with another homeless person and stole his blanket in front of several bystanders who called the police. This led to another arrest where she spent a few months in jail and was again given probation with a condition of outpatient psychiatric treatment. Following her release, she attended outpatient treatment at a mental health clinic for several weeks. Initially, she took her medications which resulted in a dramatic improvement of her symptoms. However, several weeks later, Ms. G stopped taking her medications, wherein her symptoms returned, and she became very threatening to the staff. The court was informed of this, and she was remanded to jail. Prior to her release, Ms. G was encouraged to return to her previous outpatient clinic for treatment. Sometime later, Ms. G appeared at the clinic, but the intake team concluded that her history of nonadherence to psychiatric medications along with her past violent behavior could make her a problem in management beyond the capabilities of the clinic.

A few months later, Ms. G was arrested outside a department store for having stolen a sweater and pushing a customer out of her way. This time, she was charged with petty theft with priors and assault, which was a felony. The judge believed he had no option but to send Ms. G to state prison. Criminalization was now confirmed.

Most outpatient mental health treatment facilities are not structured to successfully treat persons like Ms. G, and, as described in this section, community mental health staff is often reluctant to try.

DEFINING CRIMINALIZATION

Many factors come into play in determining why a person with mental illness is arrested rather than taken to a psychiatric hospital.

Generally, persons who are thought to have committed a felony are arrested and brought to jail regardless of their mental condition. The criminal justice system, charged by society with the responsibility for removing from the community persons accused of committing serious crimes, sees no alternative but to first place the person in custody in a secure setting and then arrange for psychiatric treatment, if necessary. If the person is thought to have committed a serious crime, the police and the criminal justice system generally do not want to place this person in a psychiatric hospital because such persons may well pose a number of problematic issues that cannot be treated and managed adequately by the mental health system and thus place society at risk.

Examples of this may be a person with schizophrenia who was found by the police in possession of a loaded gun; he was brandishing the weapon and threatening others because he believed they were poisoning him. Another person with untreated bipolar disorder was arrested after he drove his car erratically on the street and onto a sidewalk full of pedestrians because he wanted to be recognized for his race-car-driving skills.

For persons charged with misdemeanors, the situation is more complex. In 1972, the term “criminalization of the mentally ill” was first used. It was observed that persons with mental disorders who engaged in minor crimes were increasingly subject to arrest and prosecution. Later, many authors applied the concept of criminalization to persons with mental disorders who were arrested for serious crimes. The distinction between arrest and incarceration of persons with mental illness who have committed minor offenses and those who have committed serious offenses is an important one.

This section takes the position that the term criminalization should be used primarily in connection to persons with serious mental illness who are arrested and incarcerated for minor offenses instead of being placed in the mental health system. However, it should be acknowledged that many persons with serious mental illness who commit serious crimes and enter the criminal justice system might not have engaged in such behavior if they had been receiving adequate and appropriate mental health treatment at that time.

Concern about the inappropriate placement of persons with serious mental illness in the criminal justice system is not new. As far back as 1939, researchers in Europe advanced the thesis that a relatively stable number of persons are confined in any industrial society. Using prison and mental hospital census data from 18 European countries, they found an inverse relationship between prison and mental hospital populations. They theorized that if one of these forms of confinement is reduced, the other will increase. According to this theory, when prison populations are extensive, mental hospital populations will be small, and vice versa. Thus, if there is room in prisons and a shortage of hospital beds, many persons with mental illness who come to the attention of law enforcement might well be directed to the criminal justice system. Recent research in South America is compatible with this hypothesis. As will be seen in this section, this has happened in the United States and other countries; the census of state hospitals has dropped precipitously while the number of persons incarcerated in the criminal justice system has risen to record proportions.

One of the major concerns in present-day psychiatry is that placement in the criminal justice system poses a number of important problems and obstacles for the treatment and rehabilitation of persons with serious mental illness. Even when quality psychiatric care is provided in jails and prisons, the inmate/patient still has been doubly stigmatized as both a person with mental illness and a criminal. Furthermore, jails and prisons have been established to mete out punishment and to protect society; their primary mission and goals are not to provide treatment. The correctional facility's overriding need to maintain order and security, as well as its mandate to implement society's priorities of punishment and social control, greatly restrict the facility's ability to establish a therapeutic milieu and provide all the necessary interventions to treat mental illness successfully. Indeed, such a setting can exacerbate symptoms of mental illness as well as behavior that is noncompliant with the correctional facility's rules; thus, possibly resulting in longer periods of incarceration.

CAUSES OF CRIMINALIZATION

Many factors have been cited as causes for persons with mental illness being placed in the criminal justice system. The most common ones are deinstitutionalization and the unavailability of intermediate and long-term hospitalization in state hospitals for persons with serious mental illness; more formal and rigid criteria for civil commitment; the lack of adequate support systems for persons with serious mental illness in the community, as well as the difficulty such persons coming from the criminal justice system, have in gaining access to community mental health treatment; persons with co-occurring disorders of a serious mental illness and a substance abuse problem being caught up in the war on drugs (being arrested and incarcerated on the basis of the use, sale, and possession of drugs); and a belief by many law enforcement personnel that deviant behavior can be dealt with more quickly and efficiently within the criminal justice system than in the mental health system.

Deinstitutionalization

Beginning in the late 1950s, the inclination to provide hospital beds declined precipitously. This was because of the introduction of psychoactive medications; the development of more efficacious community treatment interventions; the creation of federal programs (e.g., Supplemental Security Income, Social Security Disability Insurance, Medicaid, and Medicare), which would fund community treatment and housing for persons with mental illness; the influence of the civil rights movement; and the high cost of institutionalizing persons with mental illness. For example, in 1955, when the number of patients in state hospitals in the United States reached its highest point, 559,000 persons were institutionalized in state mental hospitals out of a total national population of 165 million (339 beds per 100,000 population). By 2016 (as a result of hospital closures and bed eliminations), the number of persons in state mental hospitals dropped to 37,679 for a population of approximately 324,000,000, or 11.7 beds per 100,000 population. This rate is similar to that found in 1850, when persons with serious mental illness received little care and concern. Their plight led to the beginning of the movement, spearheaded by Dorothea Dix and others, to provide more humane care by treating persons with serious mental illness in hospitals. With respect to freestanding and county facilities that provide acute psychiatric inpatient treatment, the number of available beds has also been dwindling. For example, in California, there were approximately 17.16 beds per 100,000 residents in 2017, a loss of more than 42% per capita since 1995.

Deinstitutionalization has thus been viewed as setting the stage for increasing numbers of persons with mental illness entering the criminal justice system. It must be remembered that the community mental health system was developed in the 1960s and 1970s for persons with mental illness who had more moderate needs. Although deinstitutionalization held the promise of persons with serious mental illness being able to live successfully in the community, that hope was not achieved for a sizeable number of individuals. Part of its failure was attributed to a lack of planning before or during deinstitutionalization, as well as a lack of adequate funding. As a result, communities were not able to provide a sufficient community care system (i.e., housing, medical and psychiatric care, social services, and social and vocational rehabilitation) for the formerly hospitalized patients. Despite this, the majority of deinstitutionalized patients were able to adapt successfully in the community; however, this was not the case for a considerable minority who were arrested and placed in jails and prisons or who became homeless (between one-fourth and one-third of homeless persons have a serious mental illness). Some of these individuals presented challenges in treatment—such as not seeing themselves as mentally ill, not taking their medications, abusing substances, and in many cases, becoming violent when stressed. Many of these persons needed highly structured care to replace that which had been provided to them, albeit imperfectly, in psychiatric hospitals. The flawed implementation of deinstitutionalization would thus appear to be a significant factor accounting for the migration to jails and prisons of many persons with mental illness who otherwise could have been treated with more adequate services in the community or in a hospital. The criminal justice system had become, by default, a safety net for the mental health system.

Restrictive Civil Commitment Criteria

Many believe that more formal and rigid civil commitment criteria have contributed not only to deinstitutionalization but also to an increased number of persons with mental illness in jails and prisons. In 1969, California’s then-novel civil commitment law, the Lanterman–Petris–Short Act, went into effect. Within a decade, every state made similar modifications in their commitment codes. Such a rapid and complete consensus among legislatures is virtually unprecedented. More importantly, it reflected a nearly universal view that past inattention to the rights of persons with mental illness needed to be corrected.

In effect, the new civil commitment laws resulted in three major changes. First, the laws modified the substantive criteria for commitment from more general criteria that simply embodied concepts of mental illness and need for treatment to more specific criteria that embodied either dangerousness to self or others resulting from mental illness or the incapacity to care for oneself. Second, the laws changed the duration of commitment from indefinite and extensive periods to specific and brief periods. Third, the new laws explicitly provided that persons civilly committed have rapid access to the courts, to attorneys, and, in some cases, to jury trials; this access ensured the kinds of due-process guarantees to civilly committed persons that criminal defendants had obtained over the previous decade.

These procedural safeguards and clear commitment standards resulted in fewer, as well as shorter, commitments. Consequently, many individuals with mental illness who otherwise would have been civilly committed by family or others were now left to reside in the community where they might not have been able to care for themselves and control their impulses. Thus, some of these individuals might well have committed criminal acts and entered the criminal justice system.

Lack of Adequate Community Support Systems

Another factor that both leads to and perpetuates the criminalization of persons with serious mental illness is the lack of adequate support systems in the community. An essential part of support systems for these persons is the availability of community treatment resources. Moreover, the type of community mental health resources available may not be suitable for all of the population to be served.

The availability or lack of availability of treatment resources in the community has three important aspects. First, in most jurisdictions, mental health treatment, housing, and rehabilitation resources are insufficient to serve the very large numbers of persons with serious mental illness in the community. For instance, case management has come to be viewed as one of the essential components of an adequate mental health program. However, the mental health system is ill-prepared to provide quality case management services to all persons with serious mental illness who require it, including those leaving jails and prisons.

Second, community mental health resources may be inappropriate for some of the population to be served. For instance, persons with mental illness may be expected to come to outpatient clinics when the real need for a large proportion of this population is outreach services, in which mental health professionals go to see patients where they live. In addition, some service providers may lack the ability to provide the degree of structure required by many persons with serious mental illness. The issue of structure will be discussed later.

Third, persons with mental illness who have been in jail may not be able to gain access to community treatment even when it is available. Many of these persons have been described as resistant to treatment, dangerous, and seriously substance-abusing, characteristics generally not considered desirable by most community mental health agencies and staff. Furthermore, because many of these agencies may not have the capability to provide the needed structure, limit setting, and security and safety for staff necessary to treat these persons successfully, their reluctance to treat them may be appropriate.

A large proportion of persons with mental illness who commit criminal offenses are found to be highly resistant to psychiatric treatment. They may refuse referral, may not keep appointments, may not be adherent with psychiatric medications, may not abstain from substance abuse, and may refuse appropriate housing placements. There is evidence that many of these persons suffer from a disorder called anosognosia (a biologically based inability to recognize that one has a mental illness, and thus a biologically based lack of insight). This is discussed in greater detail later.

Thus, the mental health system finds offenders with mental illness extremely difficult to treat and may be hesitant to serve them. This reluctance may extend to virtually all areas of community-based care, including psychiatric treatment, therapeutic housing, social and vocational rehabilitation, and general social services. Moreover, many offenders with mental illness can be intimidating because of previous violent and fear-inspiring behavior. Treating this group is very different from helping passive, formerly institutionalized patients to adapt quietly to life in the community. Community mental health professionals are not only reluctant but may also be afraid to treat them, especially when measures are not adopted to ensure staff safety. Consequently, these persons are left for the criminal justice system to manage. However, there are outpatient facilities in which structure is provided, security personnel is present, and thus staff is protected, and mental health and criminal justice staff collaborate closely; under such circumstances, many of these persons are successfully treated.

The Role of the Police

The police play an important role in the criminalization of persons with serious mental illness, and this will be discussed in the context of the more in-depth exposition of the police and mental health that follows.

Police and Mental Health. Since the advent of deinstitutionalization and the exodus of persons with mental illness into the community, law enforcement agencies have had an increasingly important function in the management of persons who are experiencing psychiatric crises. The rationale for the police to intervene in the lives of persons with mental illness derives from two common-law principles: the power and the authority of the police to protect the safety and welfare of the community, and the state’s paternalistic or *parens patriae* authority, which dictates protection for citizens with disabilities who cannot care for themselves, such as those who are acutely mentally ill.

The police are typically the first responders in the community to urgent situations involving persons with mental illness. They provide services 24 hours a day, 7 days a week. Therefore, they play a major role as a mental health resource. They are responsible for either recognizing the need for the treatment of an individual with mental illness and connecting the person with the proper treatment resources or making the determination that the individual has committed a type of illegal act for which the person should be arrested. This responsibility thrusts them into the position of primary gatekeepers who determine whether the individual will enter the mental health or the criminal justice system.

With respect to persons with mental illness, police in all states have the power to transport persons for psychiatric evaluation and treatment when there is probable cause to think that they are a danger to themselves or to others because of their mental condition. As a result, law enforcement officers have assumed the role of “street-corner psychiatrist” by default. Many officers have grown accustomed to this role and consider it one of their duties; however, other officers do so reluctantly, and some do it with resentment. A major problem with having to fulfill this role is that the police have little training in performing this kind of triage. This lack of training is one of the factors that has played an important part in the criminalization of persons with mental illness. The training of law enforcement officers is discussed later in the subsection entitled *Diverting Persons with Mental Illness from the Criminal Justice System.*

The police have a great deal of discretion in the exercise of their duties, including determining what to do when they encounter a person with acute mental illness in the community. In most cases, the police use informal tactics, such as trying to “calm” the person or taking the person home. This practice has been referred to as “psychiatric first aid.” In situations that cannot be handled informally, the police may have to take persons with mental illness to hospitals or jails.

In some cases, however, public policy limits the police officer’s discretionary power. For instance, if the person with mental illness is alleged to have committed a major crime, the disposition is clear—that person is taken to jail because of the seriousness of the offense.

Often, the interaction between the police and the person with mental illness is initiated by citizens. In such cases, the citizens’ demands also may come into play and limit the discretion of the police. For example, many retail stores have a policy that anyone caught shoplifting should go to jail, and store managers are instructed to make a citizen’s arrest and call the police without exception. In another kind of situation, people who have just been assaulted by a person with psychosis frequently are not inclined to be sympathetic to their assailant, even when a mental disturbance is evident. The result may be an angry citizen insisting that the person be arrested and taken to jail.

Police and Criminalization. As discussed previously, the police play an important role in the criminalization of persons with serious mental illness. Even when the police consider the behavior of a person who has come to their attention to be the result of a mental illness, they may encounter a number of problems and irritants when trying to place the person in an inpatient psychiatric facility. For example, there may be long waiting periods in psychiatric emergency rooms during which police officers cannot attend to other duties. Mental health professionals may question the judgment of the police and refuse admission, or they may admit for only a brief hospital stay a person who just a short time before constituted a clear menace to the community.

However, the police are well aware that if they arrest a person with mental illness, that person will be dealt with in a more systematic and predictable way under the criminal justice system. The individual will be taken into custody; it can be hoped that the person will be seen by a mental health professional attached to the court or in jail and will probably receive psychiatric evaluation and treatment. Thus, an arrest is a response with which police are familiar, one over which they have more control and one that they believe will lead to an appropriate disposition.

Although a person with mental illness suspected of committing a serious crime will most certainly be arrested and transported to jail, a number of factors in addition to those just mentioned have been proposed to explain why a person with serious mental illness suspected of committing a minor offense is arrested rather than taken to a hospital. A person who appears mentally ill to a mental health professional may not appear so to police officers, who, despite their practical experience, have not had sufficient training in dealing with this population and are still laypersons. In addition, mental illness may appear to the police as simply alcohol or drug intoxication, especially if the person with mental illness has been using drugs or alcohol at the time of the interaction with the police. Moreover, in the heat and confusion of an encounter with the police and other citizens, which may include forcibly subduing the person with mental illness, signs of mental illness may go unnoticed. Furthermore, studies have shown that the occurrence of violence at the time of a police encounter increases the chances that the person with mental illness will be taken to jail.

Another instance is when law enforcement officers may charge persons with mental illness with a misdemeanor and take them to jail if they think that no other appropriate alternatives are available, a practice that has been referred to as “mercy booking.” Thus, the scarcity of psychiatric inpatient beds and/or limited community mental health services may be a factor contributing to the criminalization of persons with serious mental illness.

MENTAL ILLNESS AND VIOLENCE

In countries that have experienced mass deinstitutionalization since the late 1960s, studies have shown that persons with serious mental illness were more likely to commit criminal offenses and even more likely to commit violent offenses than the general population. This growing body of evidence demonstrates a relationship between mental illness and violence, primarily among persons who do not take their medications and are psychotic. Substance abuse has also been shown to increase the risk of violent behavior, particularly in combination with serious mental illness. Thus, nonadherence to medication, acute psychosis, and substance abuse account for most of the increased violence among persons with serious mental illness. For instance, one study examined persons with serious mental illness in the United States who were arrested over a 10-year period and found that more than one-third were arrested for violent crimes (assault and murder), and more than one-fifth were arrested for drug offenses.

The relationship between serious mental illness and violence is most easily seen in relatively nonviolent societies, such as in Scandinavia. The Scandinavian countries maintain centralized case registries that record comprehensive data on everyone born in the country or who immigrates there. Included in this wealth of data are psychiatric registers and national police registers. In Denmark, the psychiatric registry includes data such as admissions to psychiatric wards, dates of admission and discharge, and discharge diagnoses. There is also a national police register that records all arrests and convictions, as well as information on all criminal proceedings. These registers allow for the computerization and analysis of the data.

For instance, it was found that males in Denmark with a serious mental disorder who were admitted to a psychiatric hospital by age 44 years represented only 5% of the total population of males but were responsible for about 30% of all the violent offenses committed by males. Likewise, female mental patients in Denmark constituted about 5% of the female population but were responsible for 50% of all the violent offenses committed by females.

In another Danish study, the risk of violent offending was found to be 4.6 times higher for male subjects with schizophrenia and two times higher for those with affective psychoses than for male subjects who were never admitted to a psychiatric ward. For female subjects with schizophrenia, the risk was 23.2 times higher, and for those with affective psychoses, the risk was 3.9 times higher than for female subjects who were never admitted to a psychiatric ward.

Similar findings were noted in Sweden in a study that examined all 15,117 individuals who were born in Stockholm in the same given year. They were evaluated through case registries 30 years later to determine how many were diagnosed with a serious mental illness and were convicted of a violent crime. It was found that 15% of the men and 6% of the women who had a major mental disorder were convicted of a violent offense. A recent study of persons with mental illness in Sweden who were prescribed antipsychotics or mood stabilizers found that violent crime dropped by 45% for those receiving antipsychotics and 24% for those receiving mood stabilizers.

It has been asserted that violent behavior by a subgroup of persons with serious mental illness increases community stigmatization and rejection of persons with mental illness generally. These violent episodes, often highly publicized, may result in the association of mental illness with violence in the public’s mind. Such stigmatization can result in the public’s reluctance to support funding for adequate mental health services for persons with mental illness. Moreover, there may be less tolerance for integrating persons with mental illness into community housing and finding them appropriate employment. It will be difficult to reduce the stigma of mental illness and its consequences until the violence of this subgroup is addressed. Furthermore, to safeguard the welfare of persons with serious mental illness, as well as the safety of the public, mental health professionals, and society must not ignore the evidence that mental illness is associated with some increased risk of violence.

Although there is a subgroup of persons with serious mental illness who are violent, it is extremely important to take into consideration data that demonstrate that an even larger group of persons with serious mental illness are the victims of violence. For example, in a review of studies of violent victimization, 35% of persons with serious mental illness were victims of violence within the last year. For persons with serious mental illness, their victimization by violence tends to receive less public attention than the violence that is perpetrated by them.

Anosognosia

Anosognosia is a biologically based inability to recognize that one has a mental illness and thus is a biologically based lack of insight. Researchers have identified various parts of the brain (e.g., ventricles, frontal and temporal lobe subregions, medial temporal and inferior parietal regions, and subcortical structures) where damage is likely to result in anosognosia. Impaired insight may be mediated by deficiencies in conceptual organization and the lack of flexibility in abstract thinking. Generally, lack of insight that one has a serious mental illness is one of the risk factors for violence in persons with schizophrenia, thus demonstrating a relationship between poor insight and violent behavior.

The importance of anosognosia is demonstrated by a study investigating 907 adults who met the criteria for psychotic and affective disorders and received public mental health services. This study found that 26% of these persons with serious mental illness reported being arrested or violent in the last year. These individuals were significantly more likely to deny needing psychiatric treatment.

Overall, these findings suggest that many of the people who are most in need of psychiatric treatment are the least likely to believe that they need it. Treatment nonadherence is a key barrier to implementing evidence-based interventions. Several studies confirm that treatment involvement is related to a lowered risk of arrest and violence in persons with serious mental illness by reducing their psychiatric symptoms and substance abuse. Interventions to improve insight and treatment adherence in this population are warranted and needed.

SOME CHARACTERISTICS OF PERSONS WITH SERIOUS MENTAL ILLNESS WHO ARE

INCARCERATED

It is known that there are very large numbers of persons with serious mental illness in jails and prisons. What is known about them in terms of their criminal histories and legal status, the psychiatric services they use while incarcerated, and the challenges they might present in psychiatric treatment after release?

Two studies attempted to answer these questions. The first was a retrospective study of inmates with serious mental illness who were arrested and placed in a large, urban county jail. It should be noted at the outset that these findings do not necessarily represent what one would find in a similar study conducted in state and federal prisons or even in other jails. This study revealed that 76% of these inmates required and received psychiatric inpatient care or its equivalent for part of their time in jail during the current offense. Clearly, a large number of people with serious mental illness are receiving their acute psychiatric inpatient treatment in the criminal justice system rather than the mental health system.

With respect to these inmates' history before the current arrest, at least 92% were known to be nonadherent to psychiatric medications, 94% had prior arrests, 72% had prior arrests for violent crimes, and 76% were known to have a history of substance abuse. Given these data, in addition to the fact that three-fourths of these inmates required inpatient psychiatric care in the jail, it would appear that the jail had acquired the responsibility of handling many persons with serious mental illness who were among the most difficult and expensive to treat.

Another study conducted on individuals who had a serious mental illness and were transferred from jail to a psychiatric inpatient unit found that 75% had a history of physical assault in the community, 72% engaged in substance abuse, and 78% were generally nonadherent to medications. Findings revealed that poor insight into illness and nonadherence to medications were associated independently with violence toward others.

That such large percentages of persons in both studies had a history of being nonadherent to psychiatric medications suggests that treatment on re-entry into the community should include such measures as evaluation and supervision by mental health courts, assertive community treatment (ACT), and assisted outpatient treatment (AOT) (see later discussion).

Moreover, short-term, intermediate-term, or long-term hospitalization may be needed for many.

It was observed in these studies, as it frequently is in jails generally, that a number of persons with serious mental illness are arrested when it appears that their offending conduct was due primarily to their illness. The nature and circumstances of the offense suggest that these persons should have been psychiatrically hospitalized instead of being taken to jail. The following two cases serve to illustrate this point.

A person with psychotic symptoms and a previous diagnosis of Schizophrenia was walking in circles, talking to himself, and then urinating outside a grocery store; he was arrested and charged with indecent exposure. Another person with manic and paranoid symptoms was panhandling at a gas station and yelling that a customer was stealing his money, whereupon he pushed this customer and others. The attendant at the gas station called the police, who arrested him for assault. These individuals would probably have been better served if they were taken to a psychiatric hospital and placed under the jurisdiction of the mental health system rather than being arrested and placed in jail. Incarcerating such persons in settings where it is difficult to provide good treatment, such as jail, does not appear to meet the needs of society by protecting society from dangerous people. Moreover, efforts to reduce the possibility of recidivism by these persons can generally be accomplished by treatment and management within the mental health system.

Using Cognitive Behavioral Therapy to Address Antisocial Characteristics

Some professionals have attributed criminal acts by many persons with serious mental illness primarily to the individual's criminal characteristics. These professionals believe that although there are incarcerated individuals who may have a serious mental illness, many of them are in correctional facilities largely because of their antisocial attitudes and a lack of problem-solving and self-control skills that contribute to their criminal behavior. Consequently, a primary intervention suggested to reduce recidivism is to focus on the antisocial cognitions of these persons.

Indeed, a number of persons with serious mental illness in the criminal justice system may have antisocial tendencies, often to the extent of justifying a diagnosis of antisocial personality disorder. While cognitive-behavioral therapy (CBT) may be a necessary component of the treatment regimen for many such offenders in reducing their recidivism, if such persons' thinking and judgment are impaired by their untreated psychotic and manic symptoms, they may not be able to profit from CBT interventions. Therefore, if CBT is to be effective, it should be combined with psychiatric treatment.

Diverting Persons with Mental Illness from the Criminal Justice System

With the very large and increasing number of persons with serious mental illness found in jails and prisons, there have been large-scale efforts to divert them from the criminal justice system to the mental health system. Various stakeholders are involved or affected by diversion: criminal justice personnel, mental health service providers, victims, the arrested individuals' families, and the community-at-large.

Diversion, before the person is actually booked into jail or prebooking diversion, is exemplified by large-scale efforts to create community mobile crisis teams of police officers and/or mental health professionals. In addition, coordination between police and mental health professionals, as well as mental health training for law enforcement officers, may help in reducing arrests

Diversion after booking takes a variety of forms. These include specialized mental health courts, also known as problem-solving courts, that are therapeutically oriented to handle issues related exclusively to offenders with mental illness. Mental health consultation with other courts may also assist the judge by offering recommendations for treatment in lieu of incarceration.

Mobile Crisis Teams. Psychiatric emergencies are managed effectively in communities in which close formal liaisons between law enforcement and mental health agencies are established. These arrangements help to facilitate the resolution of crises of persons with mental illness in the field without the need to resort to hospitalization or incarceration. When resolution in the field is not possible, the existence of these liaisons increases the number of persons with mental illness who are referred to the mental health system rather than jailed, thus reducing criminalization. For those who need psychiatric hospitalization, these relationships tend to increase the rate of acceptance by hospitals.

Different strategies have been developed to provide a mobile team of police, mental health professionals, or both to respond in the community to persons with mental illness who are in crisis. The following are two of these strategies that are used most commonly.

An increasing number of jurisdictions use sworn police officers who have special and extensive mental health training to provide crisis intervention services as part of crisis intervention teams (CIT programs) and to act as liaisons to the mental health system. This approach is often referred to as the “Memphis Model” because it was developed in Memphis, Tennessee. These specially trained officers may deal with mental health emergency situations on-site or act as consultants to the officers at the scene. This model places a heavy reliance on psychiatric emergency services that have agreed to a no-refusal policy for persons brought to them by the police. Although this strategy has a close liaison with mental health, it does not require the actual participation of mental health professionals in the field.

Another model that some jurisdictions use deploys teams composed of both specially trained sworn police officers and mental health professionals employed by the local community mental health department. These teams have been shown to be effective in resolving emergency situations in the community involving persons with mental illness and in diverting them to the mental health system rather than to the criminal justice system. An advantage of this model is that the mental health professionals on the team may have access to the person’s documented mental health history. Having such data available is an important factor that can help the team better understand the person’s mental condition and make more informed decisions.

The major goals of both of these specialized mobile crisis teams are to resolve the crisis and reduce criminalization. Studies evaluating such teams found arrest rates ranging from 2% to 13% (with an average of less than 7%), in contrast to an arrest rate of 21% for contacts between nonspecialized police officers and persons who were apparently mentally ill. This finding lends credence to the idea that a specialized response lowers the incidence of inappropriate arrests.

Regardless of whether police officers act as part of a team with members of the mental health system in dealing on-site with emergency situations, or police act on their own in dealing with these situations, it is imperative that they obtain all available information about past police contacts with the individual when mental illness is thought to be present. Such information may include aspects of the person’s actions during previous police contact, such as suicide attempts and violence, and how the previous contact was resolved.

A crucial part of developing and maintaining any specialized mobile response team is that a rigorous and ongoing evaluation of the program must be conducted. For instance, response times—the time from the call to the arrival of the team on the scene—should be measured on a routine basis. If response times are too long, police officers in the field tend not to call on the mobile teams. The teams’ arrest rates, and the mental health alternatives they choose, need to be evaluated regularly to be sure that appropriate dispositions are being made and that mental health facilities are cooperating with these efforts.

Increased Coordination and Need for Training. From the standpoint of the police, it is clear that officers need and want rapid on-site assistance from mental health professionals when they are called on to deal with difficult or complex situations involving persons with mental illness who are acutely psychotic, behaving bizarrely, or exhibiting violent behavior, or for persons who have attempted suicide or made a suicide gesture. Similarly, mental health professionals who are working as members of psychiatric emergency teams without police support may feel ill-equipped to handle such individuals in the field. Thus, it has become increasingly apparent in recent years that when persons with mental illness in the community are in crisis, neither the police nor the emergency mental health system alone can serve them effectively and that it is essential for the two systems to work closely together in their planning, implementation, and continuous problem-solving.

Training of Police. Most police departments provide some type of officer training in mental illnesses; however, training in how officers should respond to persons with mental illness who are in crisis is less common. It has been found that when police respond to crisis situations, the risk of injury or death for individuals with serious mental illness is 16 times greater than for those with no serious mental illness. Consequently, mental health training is needed for all police officers, not just for those who are part of the specialized mobile crisis teams. Benefits of training include assisting the police in their interactions with persons who have serious mental illness as well as the potential to reduce injuries.

Training for police officers about mental health issues and individuals with mental illness should begin in police academies as a regular part of the curriculum. This training should focus on enhancing the understanding of mental illness by law enforcement personnel. Furthermore, it should aim at increasing their knowledge about such mental health issues as how to recognize mental illness, how to handle violence or potential violence among persons with mental illness, what to do when a person is threatening suicide, and when to call the specialized mobile crisis team. In addition, information should be provided as to what mental health resources are available, both inpatient and outpatient, and when a person should be arrested and incarcerated as opposed to being taken to a mental health facility.

For the training to be successful and relevant, trainers should come from both the mental health and criminal justice systems. Trainers from the mental health system should be persons who have experience with crisis intervention in the field and/or in crisis centers, as well as knowledge and field experience relating to the criminal justice system. These mental health professionals must understand and appreciate the challenges and risks that law enforcement officers experience every day in the field. Whereas trainers from the criminal justice system already have credibility with their trainees and knowledge about such issues as officer safety and enforcement protocols that noncriminal justice system persons would not have.

Another important issue to consider when law enforcement officers encounter persons with mental illness is how the individual perceives the officers. Persons with mental illness may have negative attitudes toward the police; they may be fearful of police, believing that the officers do not listen to them, and may use force against them. Such perceptions can lead to dangerous escalating interactions. Consequently, the attitudes and behaviors of both the officer and person with mental illness can influence the outcome of their contact. Understanding how persons with mental illness may perceive law enforcement officers, particularly if they had a previous negative experience with police, is essential.

It needs to be emphasized that mental health training should be an ongoing process, not only for police officers in special mental health crisis teams but also for all police officers.

MENTAL HEALTH SERVICES AFTER ARREST

To reverse criminalization, it is extremely important to identify persons with mental illness as quickly as possible after they have been arrested and placed in jail. A facility may routinely screen all incoming detainees for serious mental disorders. Initial screenings may be conducted by custody personnel who notice the inmate engaging in odd or bizarre behavior. Medical/psychiatric questionnaires are also used in many correctional institutions as screening devices. Such questionnaires typically consist of less than 20 questions aimed at identifying inmates who may need psychiatric care. Those individuals who endorse items on the questionnaire or who appear to have a mental illness may then be seen by a clinician for a more extensive evaluation to determine what type of housing and/or mental health treatment they should receive while incarcerated. Even more important, if the person is one who should have been treated in the mental health system in the first place, discharge planning should begin as soon as possible after an arrest.

Given the large numbers of persons with serious mental illness who are arrested, jail administrators should be encouraged to negotiate programmatic relationships with mental health agencies to provide multidisciplinary teams of clinicians to work in the jail setting. These teams should include psychiatrists so that psychoactive medications can be prescribed. The teams should be established inside jails to provide short-term crisis intervention, treatment, referral to a psychiatric hospital if necessary, and discharge planning. Such information can be useful in helping the judicial system make decisions concerning the person’s legal disposition.

Mental Health Courts

Post-booking diversion strategies are being used increasingly by special courts called mental health courts. The first mental health court that received national attention was established in Broward County, Florida, in 1997, and since then, the number of mental health courts in the United States has increased greatly. Initially, these courts were set up to hear cases of persons with mental illness who were typically charged with misdemeanors. In recent years, these courts have increased their purview to serve persons with mental illness charged with felonies. In mental health courts, all the courtroom personnel (i.e., judge, prosecutor, defense counsel, and other relevant professionals) have experience and training in mental health issues and available community resources. These mental health courts have a particular set of characteristics: they hear specialized cases involving defendants with mental illness; they use a nonadversarial team of professionals (e.g., judge, attorneys, and mental health clinician); they are linked to the mental health system that will provide treatment, and they use some form of adherence monitoring that may involve sanctions by the court.

Underlying the concept of mental health courts is the principle of therapeutic jurisprudence, which emphasizes that the law should be used, whenever possible, to promote the mental and physical well-being of the people it affects. The concept of therapeutic jurisprudence operates on the belief that the application of the law can have therapeutic consequences. It should be emphasized that therapeutic jurisprudence does not diminish the importance of public safety, which is fully considered by the court.

In a system characterized by therapeutic jurisprudence, people with serious mental illnesses charged with crimes may be diverted into programs designed to address their treatment and service needs rather than simply being incarcerated with their treatment needs being neglected. Even individuals with serious mental illnesses convicted of serious crimes can be provided with humane and appropriate treatment while incarcerated. Generally, mental health courts facilitate linking offenders with serious mental illness to appropriate needed services and supports on discharge from jail in order to enable them to successfully re-enter their communities.

Mental health courts were developed as a strategy to divert persons with mental illness from the criminal justice system into the mental health system. When offenders with mental illness are booked into jail, their case is usually handled by mental health courts in one of two ways. In the first situation, the prosecution of their cases may be postponed by the judge so that they may be involved in a specified treatment program run by the mental health system.

If they successfully complete the program, their criminal charges are dismissed (or, in some cases expunged), and the individuals are no longer under the jurisdiction of the criminal justice system. If they fail to complete the treatment program successfully, the charges are reinstated, and the criminal proceedings resume. In the second case, the offender with mental illness offers a plea of guilty or *nolo contendre,* which is accepted by the court. The person is sentenced (e.g., sentenced to 180 days in jail, which may be held in abeyance) and given probation, which includes treatment by the mental health system as a condition of probation. If the person is nonadherent with treatment or other probationary conditions, probation may be withdrawn, and the sentence imposed. There is evidence that these types of jail diversion may reduce recidivism for persons with serious mental illness.

The mental health court system works in a collaborative effort to devise and implement a treatment plan that includes not only medications and therapy but also housing, social and vocational rehabilitation, income assistance, and family support if possible. These interventions are to address the individual's mental illness and reduce their risk of recidivism. Research has suggested that a factor contributing to reducing recidivism and improving clinical outcomes is when the court and mental health personnel work as a "team" to coordinate and provide services. Clearly, the quality of services plays an influential role in the success of mental health courts. However, there may be unforeseen limitations of these programs with respect to available resources. That is, there are often insufficient community psychiatric treatment, rehabilitation, and housing capabilities in the existing mental health system to accommodate those persons with mental illness who are diverted from the criminal justice system. Another issue that has already been discussed is that many community resources are not appropriate, adequate, or willing to serve a large proportion of those who are diverted from the criminal justice system. Not surprisingly, there is research evidence that this lack of appropriate resources to which diverted persons can be referred results in recidivism.

Another problem resulting from the services recommended by these courts for offenders with mental illness is the possibility that they might be provided at the expense of programs for persons with mental illness who have not committed crimes. When this happens, an undesirable situation may be set up whereby committing a crime offers the patient with mental illness the best chance of getting psychiatric treatment.

These specialized mental health courts have been given high priority because they appear to be part of the answer to the problem of criminalization. However, for these diversionary programs to be truly effective and not overwhelm the already existing limited budgets of the departments of mental health, the treatment to which the courts have referred the offender with mental illness should be funded by additional monies. These monies could come, at least in part, from savings in the criminal justice system, which would no longer have to incarcerate and treat these persons.

The Importance of Mental Health Consultation to the Court

Consultative arrangements with mental health professionals may help to correct a problem that involves the computerized criminal histories, or "rap sheets," available to the court of persons with serious mental illness. Criminal justice personnel, such as prosecutors, probation officers, and parole officers, access the person's criminal history. It is important to consider what effect this may have on the court's decision-making regarding the disposition of the individual when there is no mental health input. These computerized criminal histories seldom include a narrative as to what really happened during the previous arrests. Even when a person with serious mental illness commits nonserious offenses, a series of them can result in a lengthy computerized criminal history for the individual. Thus, when such individuals subsequently appear in court, they may look like habitual criminals when their criminal histories are accessed.

By providing mental health evaluation and consultation, it may become clearer in these instances that these are persons with serious mental illness who were not receiving adequate treatment and who repeatedly acted in an inappropriate and/or aggressive manner when stressed, which caused them to come to the attention of law enforcement officers and be arrested. An example would be a woman with a history of bipolar I disorder who has a history of arrests for crimes including disturbing the peace (playing loud music throughout the early morning hours), destroying public property (defacing doors in a county building), resisting arrest for disorderly conduct (screaming at her neighbors in the middle of the night), and hitting a man on the bus whom she falsely believed was making sexual overtures toward her. Her most recent arrest was for throwing rocks at her neighbor because she mistakenly believed he was spying on her while she was taking a shower. Without the input of a mental health professional to evaluate this individual, the judge might believe that she should be incarcerated instead of referred to a hospital or other psychiatric treatment. Thus, this person with serious mental illness would continue to be criminalized.

OUTPATIENT TREATMENT TO REVERSE OR PREVENT CRIMINALIZATION

To decriminalize persons with serious mental illness, it is necessary to find ways to help them become stabilized outside of jails and prisons and, to the extent possible, not enter the criminal justice system at all. Thus, the community treatment of persons with serious mental illness who are or may become offenders has developed into an increasingly important and urgent issue. There are greatly increased numbers of persons with serious mental illness who find their way into the criminal justice system, are released into the community, and unfortunately, revolve back into the criminal justice system.

A study examined all persons (*N* = 79,211) who were incarcerated in one state's prison system during a 1-year period. Ten percent were diagnosed as having major depressive disorder, bipolar disorders, schizophrenia, or nonschizophrenic psychotic disorders. Twenty-one percent of these inmates with serious mental illness were currently incarcerated for a violent offense. These inmates also had a statistically higher percentage of assault, robbery, and property offenses than the inmates who had no psychiatric disorder. In addition, inmates with serious mental illness were far more likely to have a history of incarceration than those without a serious mental illness. As a result of this study's findings, the authors acknowledged that there was a need to not only develop programs to treat persons with serious mental illness in correctional facilities but also to find effective alternatives to incarceration in the community, including pre- and postbooking diversion programs, increasing discharge planning or continuity of care programs, drug and alcohol screening and treatment, and access to psychiatric inpatient treatment.

The purpose of this section is to review community treatment of persons with serious mental illness who, without treatment, have been or might become criminalized and to discuss the principles of treating this population as viewed generally by clinicians who conduct such treatment. This includes those who specialize in forensic mental health care. In addition, because such large numbers of persons with serious mental illness have become criminalized and then released into the community, many mental health professionals who previously may not have been involved in treating offenders with mental illness are now finding themselves with treatment responsibilities for this population.

Probably, many criminalized persons with serious mental illness can be treated at mainstream mental health clinics on their release from jails and prisons. This is exemplified by nondangerous persons with psychosis who are arrested for minor crimes, such as disturbing the peace or not paying for a meal in a restaurant, and who are taken to jail rather than the more appropriate disposition of taking them to a hospital. Another example would be individuals with serious mental illness who committed felonies and were sent to prison but are nondangerous, such as a person with bipolar disorder who was sentenced to prison for forging checks.

However, it must be recognized that there are many who are discharged from correctional institutions who have multiple problems that cannot be adequately treated in traditional community-based facilities. An example of this would be persons with serious mental illness who have a history of dangerousness; placing such persons in a traditional community clinic often results in exacerbations of their illness and acts of violence that lead to rehospitalization or reincarceration. For these individuals, it is important to have special, highly structured clinics staffed by professionals who understand dangerous offenders with mental illness and their treatment. Moreover, adequate security in such clinics is essential; this should include metal detectors, the presence at all times of trained security personnel, and emergency alarm buttons within easy reach of the mental health staff. These clinics may be in the mental health system, with the patients' treatment monitored by the criminal justice system (e.g., court and/or department of probation), or the clinics may be an actual part of the criminal justice system (e.g., run by parole departments).

Treatment of Co-Occurring Disorders

It is estimated by professionals and other personnel in the criminal justice system who are knowledgeable about incarcerated persons with serious mental illness that at least three-fourths of them also meet DSM-5 criteria for Substance Use Disorders. Clearly, if treatment after release is to be successful, both the mental illness and the substance abuse must be addressed. These services should be integrated into the community, and it is generally held that the same clinical personnel should provide and coordinate both mental health and substance abuse treatments. Treatment of co-occurring disorders very frequently needs to be a long-term process.

Assisted Outpatient Treatment

Another important treatment modality that is available in almost all states is AOT, often called outpatient commitment. AOT is a court-ordered civil commitment in which the patient is not under the jurisdiction of the criminal justice system. Rather, AOT is initiated by the mental health system to ensure that persons with mental illness and a history of hospitalizations or violence participate in services in the community that are appropriate to their needs. AOT is for persons with mental illness who can live in the community with the help of family, friends, and mental health professionals but are resistant to psychiatric treatment, including medications. Without such treatment, they may relapse and become violent and/or dangerous to themselves and require hospitalization. Because of these characteristics, this population is also prone to be arrested, incarcerated, and criminalized. To prevent decompensation, which may result in hospitalization or criminalization, and to help these persons stabilize in the community, such at-risk individuals are ordered to participate in outpatient psychiatric treatment, with their progress closely monitored by the court.

When AOT is successful, it is because there is usually a requirement that it be accompanied by an intensive and proven form of community treatment, such as ACT or forensic assertive community treatment (FACT). ACT is a community-based program with mobile mental health treatment teams that provide an array of treatment, rehabilitation, and housing services that are available 24 hours a day. FACT is similar to ACT but also includes legal leverage via collaboration with the criminal justice agencies in an effort to reduce recidivism. The goal of ACT and FACT is to help persons with serious mental illness stay out of the hospital and develop skills for living in the community. It should be noted that AOT, ACT, and FACT are evidence-based practices; that is, they have been shown to be effective, using scientifically valid studies.

There are many who believe that the widespread use of AOT would reduce the criminalization of persons with serious mental illness because it mandates treatment for individuals who otherwise are likely to be arrested and enter the criminal justice system.

Challenges of Mandatory Outpatient Treatment

Individuals in AOT with serious mental illness, as well as those who have committed criminal offenses, represent a challenge to outpatient clinicians. The problem lies not only in ensuring safety of the community but also in working with individuals who may be resistant to treatment.

A large proportion of persons with serious mental illness who commit criminal offenses have a history of being highly resistant to psychiatric treatment before their involvement in the criminal justice system; this is also true for persons who are considered to need AOT. They may have refused referral, may not have kept appointments, may have been nonadherent with psychoactive medications, and may have refused appropriate housing placements. Many have anosognosia and are not able to recognize that they have a mental illness.

For many offenders with mental illness, problems of resistance may continue after release from incarceration, which can place them at risk to the community. This is a fact that the treating professional must always keep in mind.

The mental health system finds many resistant persons with mental illness extremely difficult to treat and is often reluctant or unable to serve them. The reluctance becomes even greater after these persons have committed offenses, become involved in the criminal justice system, and are referred to community agencies. The disinclination to serve these persons extends to virtually all areas of community-based care, including therapeutic housing, social and vocational rehabilitation, and general social services. Many offenders with serious mental illness are intimidating because of previous violent, fear-inspiring behavior. Treating them is very different from helping submissive, nonthreatening, formerly institutionalized patients to adapt quietly to life in the community. Thus, community mental health professionals not only are reluctant but also may be afraid to treat offenders with mental disorders. In addition, professionals may work in treatment facilities that do not adequately provide for staff safety, do not possess the authority and leverage of the criminal justice system, and do not provide treatment interventions with an adequate structure for this population. The reluctance to treat potentially dangerous persons in such treatment facilities may be warranted. Another issue related to mental health professionals' reluctance to treat such individuals is their belief that they do not have adequate training and experience in risk assessment to ensure public safety. As discussed later, outpatient programs need to increase their ability to treat persons who do not see themselves as having a mental illness and/or have a potential for dangerous behavior.

Another important obstacle to receiving outpatient treatment for persons with serious mental illness who are in AOT or are offenders is that community mental health resources may be inappropriate. For instance, these persons may be expected to come to outpatient clinics when the real need for many in this population is for outreach services where professionals come to them.

As discussed later, a crucial part of the treatment of these persons with serious mental illness is adequately and appropriately structured and supportive therapeutic housing arrangements.

When the Criminal Justice System Is Involved

When the criminal justice system is involved in the outpatient treatment of offenders with serious mental illness, a need exists for a clear understanding of the criminal justice system's perspectives and goals. These perspectives and goals include an emphasis on concerns about public safety, control of violence, extensive use of authority, and close cooperation between the mental health and criminal justice systems.

There are a variety of circumstances under which offenders with mental illness in the community remain under the jurisdiction of the criminal justice system. One such group consists of individuals given probation by the court that includes a condition of outpatient treatment, with the possibility of being sent to jail or prison if they are nonadherent to treatment. Another consists of individuals released from prison who are referred for treatment by their parole officer with the understanding that failure to comply may result in a revocation of parole and return to custody.

Treatment Under the Jurisdiction of the Criminal Justice System

For those offenders who are mandated by the court or an agent of the criminal justice system (probation officer, parole officer) to attend outpatient treatment, the criminal justice system retains jurisdiction over the individual. In such situations, both the offender with mental illness and the therapist must satisfy legal requirements, such as regular attendance and periodic progress reports. The patient must comply with legal restrictions, such as abstinence from drugs and alcohol. Therapists who treat such individuals are often asked to provide reports to the court, probation officer, or parole officer on the patient's treatment, including issues such as attendance, diagnosis, progress, and prognosis.

Both the criminal justice system and forensic clinicians generally expect that offenders with mental illness given treatment will gain some understanding of the role of their psychiatric disorder in the past and potential future dangerous behavior and that they will avoid behavior or situations that might increase the risk of criminal activity or deterioration in their clinical condition. Moreover, both society and the criminal justice system expect that treatment will be conducted under conditions that can, to the greatest extent possible, ensure safety of the community.

Thus, a balance must exist between individual rights and public safety. However, it has been argued that courts place a greater emphasis on the potential dangerousness of the offender with mental illness than on the individual's rights. In so doing, the courts place the burden on the offender to demonstrate that they no longer pose a danger to the community. Consequently, the person's risk of violence, to a large extent, dictates whether such an individual will be placed into and remain in the community.

Outpatient treatment for offenders with serious mental illness is not designed to make presently dangerous individuals nondangerous. Rather, the criminal justice system presumes that offenders with mental illness placed in outpatient treatment will not be dangerous to others while under supervision and treatment in the community. A primary concern in the outpatient treatment of these individuals is to assess any changes in the mental condition that may indicate violence and to reduce the potential threat of harm. Therefore, features that contribute to a patient's risk of harm are addressed first in treatment. The psychiatrist must assess dangerousness and incorporate this assessment into an intervention strategy. Thus, outpatient clinicians must have as great possible an understanding of each patient's potential for violence as a function of their history and psychiatric condition. Moreover, the clinicians should feel comfortable and qualified to engage in such assessments.

In contrast, in nonforensic psychiatric treatment of nonviolent persons with serious mental illness, the primary focus is usually on alleviation of symptoms. Clinicians who treat offenders with mental illness need to recognize that they have assumed responsibility not only to the patient but also to society in ensuring the patient's safety in the community. This responsibility is not as central when clinicians are treating nonoffenders.

Identifying a Treatment Philosophy

To work effectively with this extremely difficult group of psychiatric patients, both those in AOT and those who are offenders, the necessity of identifying and articulating a treatment philosophy of both theory and practice should be emphasized. This philosophy, as already mentioned, should strike a balance between individual rights and public safety and use treatment services that take both into account. A reality-based treatment philosophy is needed, one that includes clear treatment goals, with attention paid to goals expressed by the patient; a close liaison with the court or other criminal justice agency monitoring the patient, including access to each patient's database from the criminal justice and mental health systems; and an emphasis on structure and supervision.

The importance of psychotropic medications for this population with a serious mental disorder, most of whom have problems centering on control, cannot be overemphasized. In many cases, if adherence is to be ensured, the patient must live in a community facility in which each dose is dispensed by staff.

A clear conception of the clinical uses and therapeutic value of authority appears to be a cornerstone of successful community treatment for persons with serious mental illness in AOT or for those who are offenders. When treatment is effective, the staff is not ambivalent about the use of authority. They are comfortable about insisting consistently and reasonably that the imposed conditions be followed, monitoring patients' adherence to prescribed psychiatric medications, and monitoring patients to detect the use of alcohol or illegal drugs. They have no problems with insisting that patients live in appropriately structured and supportive residential settings, to the extent needed by each individual, as a condition for remaining in the community. They are willing to rehospitalize patients in community facilities promptly at times of crisis and are willing to recommend revocation of patients' community status and return them to the court with the opinion that community treatment is no longer appropriate.

It is also important to emphasize the legal and ethical aspects of treating persons under the jurisdiction of the court. Before persons in AOT and offenders with mental illness are asked to consent to outpatient treatment, they should be apprised of all the conditions and limitations that will be imposed on them, why they will be imposed, and what will happen if they do not comply. Areas to be addressed include limits to confidentiality, with respect to both past and present treatment and criminal history, and conditions under which such information must be shared with criminal justice and mental health system personnel; supervision and monitoring by various authority figures, such as probation or parole officers, judges, therapists, and case managers; mandatory adherence with treatment and other imposed conditions; and residence in an appropriate living situation. The patient must understand that nonadherence to the terms and conditions may result in revocation of outpatient status. It is also imperative that the treatment staff understand the patient's legal status and conditions for community placement fully and agree to monitor and uphold them. Staff members must accept their role as agents of social control.

Treatment Goals

Generally, mental health professionals believe that community treatment of persons with serious mental illness who are in AOT and/or have been offenders should focus on stabilization of the illness, enhancement of independent functioning, and maintenance of internal and external controls that prevent patients from acting violently and committing offenses. It is hoped that patients will share these goals. Patients should at least have the goal of avoiding involvement with the criminal justice system. Important points for discussions between the clinician and the patient are the patient's understanding of which behaviors and symptoms are of concern, why they are of concern, what is expected of the patient both by the clinician and by the supervising court or criminal justice agency, and how the treatment can help the patient to meet these expectations.

Even patients with severe psychiatric symptoms may be able to understand these elements of treatment and adopt appropriate attitudes toward treatment. Thus, a full resolution of symptoms is not always needed for treatment to be viable in a community setting. For some persons with long-term, serious mental illness, total elimination of symptoms is not a realistic goal. Rather, the primary prerequisite for safe and effective community treatment may be patients' ability to understand and accept their clinical needs and the system's legal requirements and to demonstrate adherence.

Liaison Between Treatment Staff and the Court

Intense supervision and monitored treatment increase successful outcomes. Therefore, an essential aspect of the treatment of offenders with mental illness, especially those under the jurisdiction of the criminal justice system and of persons in AOT, is a close liaison among treatment staff and the court, the district attorney's office, the departments of probation and parole, and the patient's counsel. At the core of the liaison is a complete and relevant database, which is fundamental in understanding the extent of the patient's problems, determining whether outpatient treatment is appropriate for the patient, and developing a treatment plan. The database should include arrest reports, computerized criminal histories ("rap sheets"), hospital records, evaluations by court-appointed psychiatrists or psychologists, results of psychological testing, probation reports, and records from previous hospitalizations, outpatient treatment, and incarcerations.

A successful liaison requires open, frequent, and continuing contact between the treatment staff and the various aforementioned personnel in the legal system. In addition, all must respect each other's perspectives and accept that they are all working toward the same goals. Together, the mental health and legal systems lend their expertise to developing and modifying the most effective community treatment program for patients. Both must accept that such a program cannot be effective for certain individuals who need hospitalization or are too dangerous and thus are not appropriate for outpatient treatment.

The Need for Structure

Usually, patients referred for mandatory outpatient treatment lack internal controls; they need external controls and structure to organize them to cope with life's demands. For instance, mental health professionals who treat this population generally believe that staff should insist that patients' days be structured through meaningful, therapeutic activities such as work, day treatment, and various forms of social therapy. Another basic element of structure for this population is that treatment is mandatory and under the jurisdiction of the court.

What effect does involuntary treatment have on persons with serious mental illness? Adherence has been shown to increase when persons with serious mental illness are required to undergo involuntary treatment. Adherence is important because it has been demonstrated that persons with serious mental illness who do not comply with treatment present an increased risk to themselves and to the community. Thus, treatment nonadherence may in and of itself result in rehospitalization or incarceration.

For offenders with mental illness, the criminal justice system tends to view outpatient treatment or conditional release programs as provisional; such programs are not regarded as the right of the offender with mental illness. Thus, the courts are not obliged to grant this form of treatment; if granted, the courts allow the state to quickly and easily revoke it. The possibility of revocation of outpatient status gives the treatment staff powerful leverage to ensure adherence to the treatment conditions.

Finally, it is essential that staff understand and accept the therapeutic value of structure and their use of authority if community treatment of offenders with serious mental illness is to be successful.

Management of Violence

A major focus of therapy should be on high-priority issues, such as the need for the patient to control impulses and inappropriate expressions of anger. Persons whose violence is rooted in a major mental illness often experience their violence as a frightening loss of control. Because their violent acts often occur in the context of psychosis, they tend to perceive a clinician who is not aware of their destructive potential as unable to protect them. They tend to establish that knowledge by testing the clinician for limits. The clinician must be continuously alert and firm in order not to risk being perceived as uncaring and unable to protect the patient from their destructiveness. Treatment contracts made by the psychiatrist and the patient that deal with expectations and consequences openly and from the beginning provide a helpful structure for working with these potentially violent patients.

Persons with serious mental illness, especially those with histories of violent behavior, generally need continuous rather than episodic care. Thus, regular monitoring is needed, especially when symptoms are absent or at a low ebb, in order to deal with individual and situational factors that may arise and result in violence.

Persons with serious mental illness who have been hospitalized or incarcerated frequently enter outpatient treatment while they are still taking medication and may, for the first time in years, find themselves symptom-free, in control of their violent impulses, and able to function in the community. After a while, they may believe that they can succeed without medication; they discontinue its use, decompensate, and perhaps engage in violent behavior. Thus, for this population of persons with serious mental illness, many of whom have problems involving control and violence, the importance of continued adherence to antipsychotic medications and medications for bipolar and depressive disorders cannot be overemphasized.

In addition, behavioral, cognitive, and psychoeducational techniques emphasizing anger management have been widely used and are successful in the treatment and management of violence.

Integrating Treatment and Case Management

The purpose of case management is to help patients obtain the services they need in a timely manner and for as long as they need them. Almost all of these patients need the basic elements of case management, which starts with the premise that each patient has a designated professional with overall responsibility for their care.

The case manager formulates an individualized treatment and rehabilitation plan with the patient’s participation. As care progresses, the case manager monitors the patient to determine whether they are receiving treatment, have an appropriate living situation, have adequate funds, and have access to vocational rehabilitation. The treatment plan emphasizes helping the patient to deal with practical problems of daily living. In addition, the case manager provides outreach services to the patient wherever they are living, whether alone, with family, in a board-and-care home, or in another residential setting.

Outreach services may take the form of assertive case management. An assertive case management program deals with patients on a frequent and long-term basis. It takes a hands-on approach that may necessitate meeting with patients on “their own turf” or even seeing a patient daily. This form of contact and familiarity with the patient helps the case manager anticipate and prevent significant decompensation. Low caseloads for case managers of potentially violent persons with mental illness—probably not more than ten cases—have been recommended, as have 24-hour-a-day, 7-day-a-week availability. Many violent acts and arrests occur during evenings, nights, and weekends, when traditional treatment programs are closed.

Patients who perceive the case manager as mainly an agent of the court and whose primary intention is to make them “toe the line” will be less likely to form a positive relationship with a case manager. They may be guarded and defensive in their interactions. They may also feel a need to portray themselves in a good light, even though they have problems for which they need help. Therefore, case managers, in addition to enforcing the terms and conditions of the court and the treatment, need to be persons who will go beyond this role to secure and provide services and support. It is crucial that patients see the case manager as their advocate to further their treatment and rehabilitation, particularly when agencies in the criminal justice system are dealing simultaneously with them in more coercive or authoritarian ways.

Case managers may experience considerable stress in working with patients in mandatory outpatient treatment. They may feel, with some justification, that they are in danger if they work with persons with serious mental illness who were violent in the past or if their work requires visits to high-crime areas, where many persons with serious mental illness live. Additional problems may include unusual working hours, which may disrupt their family and social relationships, and lack of upward career mobility.

Therapeutic Living Arrangements

Survival in the community for most persons with serious mental illness appears to depend on an appropriately supportive and structured living arrangement. Such an arrangement can often be provided by family members. In many cases, however, the kind and degree of structure the patient needs can be found only in a living arrangement outside of the family home with a high staff–patient ratio, dispensing of medication by staff, enforcement of curfews, and therapeutic activities that structure most of the patient’s day.

Some patients need a great deal of structure and supervision in their housing situation. Others need only a minimal amount, and most fall somewhere in between. The treatment staff member assigned to the patient or the patient’s case manager must decide whether a particular living arrangement has the appropriate amount of structure to meet the patient’s needs. However, it is important to first discuss the suggested living arrangement with the court or the responsible agent in the criminal justice system and obtain their approval.

A consideration for all patients who move from a closed or locked setting such as a hospital, jail, or prison is the patient’s ability to handle the transition. It is generally not advisable for persons who have been hospitalized or incarcerated for a long time to then be placed in the community in a living situation with little or no structure. Such individuals are frequently unable to cope with the immediate stress and demands of these arrangements, and they either decompensate or commit subsequent offenses.

For many persons with serious mental illness, it is helpful to release them to graduated lower levels of structure—for example, from a hospital or prison to a locked community facility to a halfway house and, when they are ready, to family or independent living. In the graduated release process, close attention must be paid to patients' coping skills and the need for monitoring.

Working With the Family

The role of family members or significant others can be critical in the treatment of offenders with mental illness or for those in AOT. The treatment team should determine whether these individuals were the victims of the patient's aggression, whether they have maintained contact with the patient, and whether they are interested in continuing such contact. The team should also learn whether other social support systems are available while the patient resides in the community. Social support can be an extremely important part of community treatment for persons with mental illness. Assessing problems that may develop between the patient and family members or significant others is essential if contact between them is anticipated.

Another important consideration is the family members' need for guidance and support, especially when they have been victimized by the patient. Clinicians should help them understand the patient's mental condition, teach them to recognize symptoms of decompensation, demonstrate methods for self-protection, and explain the patient's current legal situation.

Moreover, family members should be involved in support groups to help them during crises. In self-help programs, they can benefit from the experience of other families in similar situations.

Diversity, Awareness, and Sensitivity

It is essential that mental health professionals understand how the roles of ethnicity, race, culture, gender, religion, sexual orientation, and age can affect one's mental health. Equally important is an understanding of the kinds of treatment and how treatment is presented that will be acceptable to these groups. Only then can the mental health system provide to racial, cultural, religious, and ethnic minorities as well as persons of all ages, sexual orientations, and genders services that are appropriate for their needs and ones to which they will be responsive.

Pitfalls

Mental health professionals must understand their potential liability for the actions of their patients on community status who are prone to violence. They need to be aware that in some states, courts have imposed on therapists a duty of care that extends to the foreseeability of harm. Moreover, the litigious nature of society adds further pressures, stress, and risks. Treating clinicians may be held accountable for patients' suicides, assaults, homicides, and other crimes, even when treatment was mandated, sound, and met the professional community's standards of care. In both public agencies and private practice, malpractice insurance coverage must be adequate so that clinicians feel comfortable, personally and professionally, in undertaking the treatment of persons with serious mental illness who need mandatory outpatient treatment.

Another important concern of those clinicians involved in the community treatment of offenders with serious mental illness, especially persons who have committed high-profile crimes, is the possibility of notoriety and unfavorable publicity. Clinicians may fear that if patients commit further sensational crimes, their professional reputation will suffer, and their facility may receive negative media coverage.

Twenty-Four–Hour Structured Inpatient Care

Not all persons with mental illness can be treated effectively in the community. There is a substantial minority of persons with serious mental illness who need the structure and support of acute, intermediate, or long-term care in a hospital setting or a highly structured, locked 24-hour care community facility. It is important to identify high-risk patients so that scarce community treatment resources are used for those most likely to succeed and for those who are least likely to incur the high costs of lengthy hospitalization or rearrest for serious offenses.

It is essential to facilitate access to hospital care for patients who need it. In the early years of deinstitutionalization, many believed that if the counter-therapeutic practices that were exposed in some state psychiatric hospitals could be eliminated, this would simultaneously eliminate the need for hospitals altogether. However, experience has shown that some persons with serious mental illness require hospitalization, and some may need it for an extended period. This may be true even in communities that offer a complete array of excellent and integrated community-based services.

It is important to acknowledge that the community is not necessarily the most benign treatment site for all people with mental illness at all times. Access to hospital care for those who need it for as long as they need it is absolutely essential to the success of deinstitutionalization and the reduction in criminalization for this population.

For those persons with serious mental illness who do not require a setting as highly structured as that of a locked psychiatric hospital but who are not ready for an open community setting, intermediate care facilities have proven extremely valuable. An example is the Institutes for Mental Disease (IMDs) in California. IMDs are locked, private, community intermediate care facilities with a moderate but lesser degree of structure than that found in long-term psychiatric hospitals, such as state hospitals. The facility provides very close medication supervision, with many patients receiving their medication in either liquid or crushed form. The staff–patient ratio is high, with at least 27 hours of activities a week for each patient. These activities include group therapies, therapeutic, social, and recreational activities, and instruction in activities of daily living.

Intermediate care facilities have been used widely for both civil and criminally committed patients. These intermediate care facilities have helped to stabilize persons who might well have been criminalized; for those persons coming from jails and prisons, they might well have prevented recidivism.

Decarceration of Jails and Prisons

Imprisonment rates in the United States increased 500% over the past 40 years, with a peak occurring between 2006 and 2008. Currently, the United States has the highest rate of incarceration in the world. This resulted in prison overcrowding and great financial costs. Consequently, many efforts are underway currently to decrease the number of persons imprisoned. Some of these are reductions in state and federal sentences for less serious crimes, including drug offenses, the use of alternatives to prison by means of diversion programs, offering early release to prisoners who participate in rehabilitation programs, and legal mandates to reduce prison overcrowding.

Beginning in 2009, the number of prisoners in U.S. prisons and jails slightly dropped each year. The greatest reduction in prisoners was in California, which exemplifies many of the factors that led to decarceration. In the case of *Brown v. Plata*, the U.S. Supreme Court in 2011 found that overcrowding in California's prisons resulted in a violation of prisoners' constitutional rights and upheld an order to significantly reduce the number of state prisoners. Legislation known as "Public Safety Realignment" was enacted that year which used two primary methods to reduce overcrowding, costs, and recidivism. The first method focused on reducing admissions to prison by having many nonserious, nonsex, and nonviolent felony offenders serve sentences, generally of 3 years or less, in county jail instead of prison. The second method focused on having fewer prisoners return to prison; that is, inmates who previously were released on state parole were now released under county probation to be supervised and provided services by county agencies. If they violated their conditions of probation (but without committing a new crime), they could not be returned to prison but could be placed in jail.

"Realignment" became effective on October 1, 2011. At that time there were 143,710 state prisoners. By May 26, 2021, the number was 92,503, a reduction of 51,207 state prisoners. Persons with serious mental illness may account for a sizeable number of those released. As mentioned previously, recent studies estimate that approximately 16% of persons in prisons have a serious mental illness. If this percentage holds for the 51,207 reduction in state prisoners, then approximately 8,193 (51,207 × 16%) fewer persons with a serious mental illness are incarcerated in California prisons.

In 2018, California added a new statutory scheme (Sec. 1001.35 to 1001.36 Penal Code) to continue reducing the number of persons with mental illness in the criminal justice system by granting pre-trial diversion for individuals with mental disorders. It is not limited to specialty courts, can apply to individuals arrested for misdemeanors and felonies, offers a variety of care settings, and would result in the expungement of the arrest if successfully completed within the 2-year period of diversion. This form of diversion is for defendants whose mental condition played a significant role in the charged offense, who would respond to mental health treatment, and who would not pose a significant risk of danger to the public if treated in the community. There are codified exclusionary criteria, including certain serious charged crimes (e.g., murder, rape) and diagnoses (antisocial personality disorder, borderline personality disorder, and pedophilia). States across the nation are addressing the financial costs and effectiveness of incarceration for those currently in prison. Relying on local jails instead of prisons to incarcerate offenders is also likely to result in high costs and overcrowding conditions in the jails and lead to releasing the least dangerous jail inmates for more serious offenders. As a result, arresting persons with serious mental illness who commit or are alleged to commit nonserious, nonsex, and nonviolent crimes and placing them in jails may become less common than in the past.

Consequently, the era of the criminalization of large numbers of persons with serious mental illness may be ebbing in both state prisons and jails. The criminal justice system will not be able to absorb the great numbers that it did formerly. As a result, during decarceration, the mental health system may be expected to assume the care and treatment for far more persons with serious mental illness. Some of these individuals may be among the most difficult and expensive to treat, as well as at great risk of decompensation and/or reoffense. If, however, society gives a high priority to successfully treating and reducing the hundreds of thousands of persons with serious mental illness who are currently in jails and prisons, or who are being released, or who are at risk of being arrested, some similar problems encountered during deinstitutionalization may be avoided.

Some Possible Strategies

The list of problems is formidable, but they can be resolved. There would need to be a very large increase in the number of community mental health services, including ACT and FACT teams and intensive case management; co-occurring disorder programs; administration of antipsychotic medications and mood stabilizers (especially important in reducing arrests for those who have problems involving control and violence); crisis response teams; appropriately structured and supportive housing; vocational rehabilitation; social rehabilitation; availability of acute care in the community; access to acute, intermediate, and long-term inpatient care when needed; more flexible civil commitment laws; and

AOT when indicated. Many of these interventions are evidence-based practices. With a sufficient amount of all these services being made available, the number of persons with serious mental illness being criminalized should drop significantly.

With the large numbers of persons with serious mental illness who first come to the attention of law enforcement officers, the opportunity to enter the mental health system should be given to those who are appropriate for diversion. Toward this end, there needs to be many more mobile crisis teams, such as CITs of specially trained police officers, who can find alternatives other than arrest and incarceration for such persons in crisis. There should also be increased education and training on mental illness and available mental health resources for law enforcement officers in general. Diversion after arrest may be accomplished in mental health courts (or even general courts), where persons with mental illness are sent to the mental health system to obtain treatment in lieu of incarceration. It needs to be emphasized that pre- and postbooking diversion can be effective only if there are adequate and appropriate mental health resources to which persons with serious mental illness can be diverted.

For those persons with serious mental illness who enter and remain under the jurisdiction of the criminal justice system, creating a successful treatment program for them is a difficult task that demands the input and cooperation of correctional and mental health professionals who are knowledgeable and accepting of the tenets of both the criminal justice system and mental health treatment. A close liaison among such varied professionals helps to maximize a successful outcome for offenders with mental illness. These persons should be provided with appropriate treatment while they are incarcerated and after they have been released into the community. Moreover, all such programs should be evaluated independently to determine which are effective and which are not to ensure the best possible use of scarce resources.

Finally, a significant increase in mental health services, ranging from outpatient treatment and case management to highly structured 24-hour care, would no doubt result in far fewer persons with serious mental illness committing criminal offenses. There is an enormous stigma attached to people who have been in jails and prisons or in a forensic hospital. Such persons have been categorized as both mentally ill and as offenders, which makes it extremely difficult to place them in community treatment and housing.

In this time of decarceration, psychiatrists are poised to provide persons with serious mental illness the mental health treatment and supportive social services that were inadequately provided previously for so many and thus led to their decompensation and criminal behavior. If the goals of reducing the criminalization of persons with serious mental illness are to be accomplished, the mental health and criminal justice systems must not only work collaboratively to problem solve but also be provided with all the necessary resources and funding to identify and treat these individuals in the most appropriate setting. It cannot be emphasized enough that the criminal justice system should not be used as a substitute for the mental health system.

FURTHER READINGS

Baillargeon J, Binswanger IA, Penn JV, Williams BA, Murray OJ. Psychiatric disorders and repeat incarcerations: the revolving prison door. *Am J Psychiatry*. 2009;166:103–109.

Beach C, Dykema L, Appelbaum PS, et al. Forensic and nonforensic clients in assertive community treatment: a longitudinal study. *Psychiatr Serv*. 2013;64:437–444.

*Brown V. Plata 131 S. Ct. 1910. 2011.

Carson EA. Prisoners in 2019. *Bureau of Justice Statistics*. 2020;NCJ-255115:1–38.

Fazel S, Zetterqvist J, Larsson H, Långström N, Lichtenstein P. Antipsychotics, mood stabilisers, and risk of violent crime. *Lancet*. 2014;384:1206–1214.

Harcourt BE. Mass incarceration: causes, consequences, and exit strategies: reducing mass incarceration: lessons from the deinstitutionalization of mental hospitals in the 1960s. *Ohio St J Crim L*. 2011;9:53–88.

Kubiak S, Comartin EB, Ray B, Tillander E. The effect of systems collaboration on the individual outcomes of mental health court participants: a multi-site study. *Int J Law Psychiatr*. 2018;60:64–72.

Lamb HR, Weinberger LE. Decarceration of our jails and prisons: where will persons with serious mental illness go? *J Am Acad Psychiatry Law*. 2014;42:489–494.

Lamb HR, Weinberger LE. Some perspectives on criminalization. *J Am Acad Psychiatry Law*. 2013;41:287–293.

*Lamb HR, Weinberger LE. Rediscovering the concept of asylum for persons with serious mental illness. *J Am Acad Psychiatry Law*. 2016;44:106–110.

Lamb HR, Weinberger LE, Marsh JS, Gross BH. Treatment prospects for persons with severe mental illness in an urban county jail. *Psychiatr Serv*. 2007;58:782–786.

Lamberti JS. Preventing criminal recidivism through mental health and criminal justice collaboration. *Psychiatr Serv*. 2016;67:1206–1212.

Landess J, Holoyda B. Mental health courts and forensic assertive community treatment teams as correctional diversion programs. *Behav Sci Law*. 2017;35:501–511.

Lehrer DS, Lorenz J. Anosognosia in schizophrenia: hidden in plain sight. *Innov Clin Neurosci*. 2014;11:10–17.

Lurigio AJ. Co-occurring disorders: mental health and drug misuse. In: Leukefeld C, Gullotta TP, Gregrich J, eds. *Handbook of Evidence-Based Substance Abuse Treatment in Criminal Justice Settings*. US Springer Science + Business Media; 2011:279–292.

Markowitz FE. Mental illness, crime, and violence: risk, context, and social control. *Aggress Violent Behav*. 2011;16:36–44.

Mundt AP, Winnie SC, Arduino M, et al. Psychiatric hospital beds and prison populations in South America since 1990: does the Penrose hypothesis apply? *JAMA Psychiatry*. 2015;72(2):112–118.

Munetz MR, Griffin PA, Kemp K. Jail diversion: using the sequential intercept model. In: Yeager KR, Cutler DL, Svendsen D, Sills GM, eds. *Modern Community Mental Health: An Interdisciplinary Approach*. Oxford University Press; 2013:453–472.

Pettinati HM, O'Brien CP, Dundon WD. Current status of co-occurring mood and substance use disorders: a new therapeutic target. *Am J Psychiatry*. 2013;170:23–30.

Pinals DA, Fuller DA. The vital role of a full continuum of psychiatric care beyond beds. *Psych Serv*. 2020;71:713–721.

*Prins SJ. Criminogenic or criminalized? Testing an assumption for expanding criminogenic risk assessment. *Law Hum Behav*. 2019;43:477–490.

*Redlich AD. The past, present, and future of mental health courts. In: Wiener RL, Brank EM, eds. *Problem Solving Courts: Social Science and Legal Perspectives*. Springer; 2013.

Silver E, Felson RB, Vaneseltine M. The relationship between mental health problems and violence among criminal offenders. *Crim Justice Behav*. 2008;35:405–426.

Sisti DA, Segal AG, Emanuel EJ. Improving long-term psychiatric care: bring back the asylums. *JAMA*. 2015;313:243–244.

*Slate RN. Deinstitutionalization, criminalization of mental illness, and the principle of therapeutic jurisprudence. *S Cal Interdis L J*. 2017;26:341–355.

Steadman HJ, Redlich A, Callahan L, Robbins PC, Vesselinov R. Effect of mental health courts on arrests and jail days. *Arch Gen Psychiatry*. 2011;68:167–172.

*Swartz MS, Wilder CM, Swanson JW, et al. Assessing outcomes for consumers in New York's assisted outpatient treatment program. *Psychiatr Serv*. 2010;61:976–981.

Torrey EF, Fuller DA, Geller J, Jacobs C, Ragosta K. No room at the Inn: Trends and consequences of closing public psychiatric hospitals. *Treatment Advocacy Center*. 2012:1–28.

Torrey EF, Zdanowicz M, Kennard A, et al. *The Treatment of Persons With Mental Illness in Prisons and Jails: A State Survey*. Treatment Advocacy Center; 2014. Accessed July 1, 2021. http://tacreports.org/storage/documents/treatment-behind-bars/treatment-behind-bars.pdf

Van Dorn R, Volavka J, Johnson N. Mental disorder and violence: is there a relationship beyond substance use? *Soc Psychiatry Psychiatr Epidemiol*. 2012;47:487–503.

57

Psychiatric Education

57.1 Graduate Psychiatric Education

George Alan Keepers, M.D.

Over the last 100 years, the education of physicians in the specialties of medicine has evolved from informal, unsystematized training of variable length and content into the highly regulated residency training programs that constitute the normal educational path for medical school students throughout the United States today. At times gradual, the development of these programs was changed dramatically by the influence of the Flexner report, which resulted in the standardization of both undergraduate and graduate medical education and the adoption of a residency training model begun at Johns Hopkins University. Today, graduate medical education (GME) is undergoing another disruptive episode of evolution championed by the ACGME, the organization responsible for the accreditation of all residency training programs in the United States. In the past, training programs were standardized around the processes and contents of their clinical and didactic curriculum. The Accreditation Council on Graduate Medical Education's (ACGME) competency movement and its more advanced elaboration, the Milestones Project, seek to use measures of the intermediate and final outcomes of residency training to judge the adequacy of training and of training programs.

This textbook has always covered the subject of psychiatric training, but each author has developed the topic in a somewhat different fashion. In earlier editions, the role of the American Board of Psychiatry and Neurology (ABPN) in influencing training programs was emphasized by Lester Rudy and subsequently by Steven Scheiber. James Eaton, in contrast, emphasized the historical development of the field and the influence of the National Institute of Mental Health (NIMH) in providing funding and direction for the development of high-quality training programs. Most recently, Larry Faulkner reviewed the major organizations that influence psychiatric training programs, provided an overview of the organizational structures that regulate and determine the processes and content of training programs, and addressed issues in psychiatric education that were then relevant.

As Faulkner reviewed in detail, psychiatric educators faced a number of controversial issues in the past. Some of these issues have been resolved, while others continue to be very actively discussed. The interference of service requirements in residency with residents' educational activities has been effectively eliminated by ACGME regulations which limit work hours, require protected time for education and incorporate residents' judgments regarding this issue into the ACGME Psychiatry Residency Review Committee's (PRRC) evaluation of training programs. However, certain new problems have arisen as a result of these regulations. Work compression, which is the pressure to accomplish all patient care activities within a limited time, has created new stressors for residents. Some fear that these pressures result in the loss of adequate training experiences since residents can no longer thoroughly evaluate patients or observe the full course of their illnesses.

Conflict in the training curriculum between biologic psychiatry and psychotherapy education has been reduced by the field's consensus regarding the importance of both and by the specification in the PRRC's Program Requirements (PRs) and residency training milestones (discussed below) of specific training requirements in both areas. The conflict between resident support and resident evaluation has also been softened by recent developments. Academic medical centers are increasingly developing resident support systems that exist outside of departmental structures. The movement toward competency-based education embodied in the milestones supports a developmental rather than summative or judgmental approach to residency training.

Nonetheless, balancing the provision of specific feedback based on high expectations while maintaining a positive, nurturing environment for residents remains a challenge.

While ABPN certification following residency remains nominally voluntary, two developments in the field have pressured residents and programs toward 100% participation in certification. First, many, if not most, health systems are now requiring American Board of Medical Specialties (ABMS) specialty certification as a condition of credentialing. Second, the PRRC's PRs, which took effect in July of 2019, require that programs maintain an aggregate 3-year pass rate above the bottom fifth percentile of programs in that specialty.

One issue raised in the previous edition remains a significant area of controversy. The limitation of duty hours and reduction in the clinical responsibilities assigned to residents has had the unintended consequence of reducing residents' continuous contact with patients and delaying their maturation as independent practitioners. It is even uncertain as to whether the reductions in work hours have led to the intended improvements in patient safety, perhaps because of the multiplication in transitions of care required in the new paradigm. It is clear that extended shifts of greater than 16 hours have a negative effect on attentional processes and reaction time. But, thus far, studies have not demonstrated increased patient mortality or morbidity from extended resident duty hours. The ACGME and others continue to study this issue.

This chapter adopts a historical approach to the topic of Graduate Psychiatric Education (GPE). The history of psychiatric training will be described, the development of the organizations responsible for regulating it will be reviewed, and the content and structure of residency training as it now exists will be described. Several recent changes in the conduct of GPE will be discussed, and the above and other controversial topics reviewed.

HISTORY OF PSYCHIATRIC EDUCATION

Early History

Prior to Abraham Flexner's 1910 report for the Carnegie Foundation, medical education in the United States was in an abysmal state. There were no established standards for undergraduate medical education, no requirements for GME, and little to no regulation of the practice of medicine. Different medical traditions, including allopathy, osteopathy, and naturopathy, provided highly varied training experiences. Physicians might then go directly into practice or seek employment, sometimes under the guidance of a more senior physician who had a particular interest. It was not unusual for physicians to practice in fields for which they had no particular training and for their interests and practice to shift during the course of their careers. Those engaged in what is now called psychiatric practice in the United States fell into two camps—on the one hand, the alienists, and on the other, the physicians who conducted ambulatory practices in nervous conditions. These divisions were not original to the United States but were rooted in developments in the field in Europe. The "spa physicians" made great efforts to assert that their clients and practices were different from the mentally ill. The "alienists" explicitly dealt with severely ill patients and were involved, often as superintendents, with newly developed institutions, frequently state-sponsored, which specialized in the care of severely mentally ill individuals. Originally innovative and providing access to moral treatment, these institutions became so overcrowded that they were at best able to provide custodial care.

Many physicians of varying interests practiced and trained in these settings. For example, Dr. James Hawthorne, who established such an institution in Portland, Oregon, was typical of the physicians of this era. A native of Pennsylvania, he began his study of medicine under the tutelage of a local physician Dr. Bascom, later obtaining a medical degree from the Medical University in Louisville, Kentucky. He opened a general practice in Louisville, which became very successful, and repeated this success in California, starting in 1850. He moved to Portland in 1857, opened a general practice, and subsequently established the Oregon Hospital for the Insane. The Hawthorne hospital eventually had 450 psychiatric beds and was supported on the basis of a contract with the state legislature. It was regarded as one of the finest such institutions in the United States, and Dr. Hawthorne, who had no formal training in psychiatry, was known as an expert in the illnesses of the mind. He was a member of the Association of Medical Superintendents of American Institutions for the Insane (organized in 1844), which became the American Psychiatric Association in 1921. Many physicians practiced at the Hawthorne hospital. Most notably, Simeon Josephi came to the hospital and worked under Dr. Hawthorne's supervision. Dr. Josephi left his employment at the hospital to found the Oregon Medical School in 1863. This type of fluid arrangement for training and practice was typical of the times.

Influence of the Flexner Report

Unfortunately, but not surprisingly, this unregulated system led to the public being subjected to poorly trained physicians, naturopaths, and snake oil salesman, all claiming to be legitimate medical practitioners. The American Medical Association (AMA) was organized in 1847 originally to deal with the lack of regulations and standards in medical education and medical practice. Despite the establishment of residency training programs at Johns Hopkins University in 1889, progress was quite limited until the Carnegie Foundation employed Abraham Flexner to study the condition of medical education in the early 1900s. The current system of undergraduate and graduate medical training was founded on Flexner's 1910 report.

Governmental and professional regulatory mechanisms soon followed and have proliferated into a complex group of organizations affecting GPE today. This resulted in the closure of inadequate schools and established allopathy and osteopathy as the only legitimate basis for medical training. Flexner strongly favored the residency program established by Johns Hopkins, and this became and has remained the framework for all residency training in the United States (Table 57.1–1).

Table 57.1–1.
Major Organizations Affecting Graduate Medical Education

American Association of Directors of Psychiatric Residency Training Program	AADPRT	Promotion of excellence in psychiatric training; resources for program directors; provides input to ACGME and ABPN from the field
Association of American Medical Colleges	AAMC	Association medical schools, teaching hospitals, and academic medical centers promoting the improvement of health through enhancing the effectiveness of academic medicine
American Board of Medical Specialties	ABMS	Improvement of health care through setting professional standards for lifelong certification in partnership with the specialty boards
American Board of Psychiatry and Neurology	ABPN	Certification, lifelong learning, and maintenance of certification of psychiatrists and neurologists
American College of Psychiatrists	ACP	Honorary society; develops Psychiatry Resident-in-Training Examination (PRITE) and the Psychiatrist-in-Practice Examination (PIPE) and grants awards to outstanding residents, training programs, and scholars
Accreditation Council of	ACCME	Approval of continuing

Continuing Medical Education		medical education (CME) programs
Accreditation Council for Graduate Medical Education	ACGME	Development of residency training Program Requirements (PRs), Milestones, accreditation of institutions and residency training programs
American Medical Association	AMA	Largest professional medical association. Advocacy for patients and the profession
American Osteopathic Association	AOA	Accreditation of osteopathic schools
American Psychiatric Association (and subsidiary state associations)	APA	Professional association for psychiatrists. Provides multiple education efforts for students, residents, and practitioners. Advocacy for patients and the profession
Educational Commission for Foreign Medical Graduates	ECFMG	Credentialing of international medical graduates for training in the United States
Federation of State Medical Boards	FSMB	Coordinates activities of state licensing boards
Liaison Committee on Medical Education	LCME	Accreditation of Medical Schools
National Board of Medical Examiners	NBME	U.S. National Medical Licensure Exam (USMLE) for medical students and residents
National Residency	NRMP	Regulates recruitment

Matching Program	activities. Matches residents and programs

Following the Flexner report, several specialties (Ophthalmology, 1916; Otolaryngology, 1924; Obstetrics and Gynecology, 1930; Dermatology and Syphilology, 1932) established boards for the purpose of describing training in the specialty and examining candidates for certification. The ABPN was established in 1934 as a collaborative effort of the APA, the American Neurological Association (ANA), and the AMA. Originally, like other Boards, the ABPN both accredited training programs and certified practitioners.

Establishment of the Accreditation Council on Graduate Medical Education

The burden imposed on the Boards by the accreditation of training programs was substantial, and in 1940 the AMA Council on Medical Education, the American Board of Internal Medicine, and the American College of Physicians established the Conference Committee on Graduate Training in Internal Medicine, a forerunner of the ACGME Residency Review Committees (RRCs). The American College of Surgeons and the American Board of Surgery established their own accreditation bodies in 1949.

Academic educators soon realized that, with the establishment of Medicare and federal funding for residency training, the several accrediting organizations then in existence with widely varying standards for approval of training programs were not tenable and required reform. The case for uniformity of standards with a national scope was supported by the AMA, other medical groups, and the government. In 1972, the AMA sponsored the ABMS, the AHA, the AAMC, and the Council of Medical Specialty Societies (CMSS) to form a Liaison Committee for Graduate Medical Education (LCGME) to coordinate the activities of the independent RRCs then in existence. The failure of this effort indicated that an independent, self-sustaining organization was required to accomplish these tasks, and the ACGME replaced the LCGME in 1981. The ACGME was charged with the responsibility for the accreditation and monitoring of residency programs throughout the United States. In the subsequent years, the current system of RRCs, PRs, and periodic accreditation site visits was developed.

The evolution of the ACGME PRs in psychiatry has been a collaborative effort between the ACGME, the ACGME PRRC, psychiatric professional organizations, and the ABPN. This has resulted in more uniform, higher-quality training throughout the United States and an alignment between completion of an accredited residency and eligibility to take the ABPN certification examination. Beginning in 2010, the ACGME undertook a major revision of the methods used to evaluate residents and training programs. Entitled the Next Accreditation System (NAS), this new system exchanges yearly reviews of data from all programs with the selection of outliers for review for the old system of in-depth reviews of all programs every 5 years following site visits. The details of this system are discussed below.

The American Board of Psychiatry and Neurology

The ABPN retained the task of certifying graduates of accredited training programs in psychiatry and neurology. The ABPN, like the other 24 specialty boards of the ABMS, is a nongovernmental, independent, nonprofit 501(c) 3 organization governed by a board of directors (Table 57.1–2).

The board is comprised of eight psychiatry directors and eight neurology directors, each of whom serves two 4-year terms (Table 57.1–3). Currently, the Board includes nine neurology directors due to a unique circumstance. Candidates for selection as psychiatry directors are solicited from the APA, the American College of Psychiatry (ACP), and the AMA. The ANA and the American Academy of Neurology (AAN) nominate candidates in neurology. Current directors of the board may also nominate candidates.

Table 57.1–2.
The 24 Boards of the American Board of Medical Specialties. (Year approved in parentheses)

Allergy and Immunology (1971)[a]
Anesthesiology (1941)
Colon and Rectal Surgery (1949)
Dermatology (1932)
Emergency Medicine (1979)
Family Medicine (1969)
Internal Medicine (1936)
Medical Genetics (1991)
Neurologic Surgery (1940)
Nuclear Medicine (1971)
Obstetrics and Gynecology (incorporated 1930)
Ophthalmology (incorporated 1916)
Orthopedic Surgery (1935)
Otolaryngology (incorporated 1924)
Pathology (1936)
Pediatrics (1935)
Physical Medicine and Rehabilitation (1947)
Plastic Surgery (1941)
Preventive Medicine (1949)
Psychiatry and Neurology (1934)
Radiology (1935)
Surgery (1937)
Thoracic Surgery (1971)
Urology (1935)

[a] Conjoint board sponsored by the American Board of Internal Medicine and the American Board of Pediatrics.

Source: American Board of Medical Specialties (ABMS). Data abstracted from the online ABMS Guide to the Medical Specialties, 2020.

Certification methods have changed through the years, and these changes are detailed elsewhere in this text. Two most recent developments are of great importance to psychiatric educators. Ending a 75-year tradition, the last ABPN oral examination in general psychiatry was given in Portland, Oregon, at Oregon Health & Science University in 2015. The oral examination in child psychiatry was discontinued not long afterward, with the last examination conducted at Northwestern in Chicago. The new certification examination in General Psychiatry has proven to have very high reliability, with passage rates that are similar to those of the old Part I examination (Table 57.1–4).

Table 57.1–3.
American Board of Psychiatry and Neurology Directors. Larry Faulkner, M.D., President and CEO

2023 Psychiatry Directors

Joan M. Anzia, M.D., Chicago, IL, Chair
Sheldon Benjamin, M.D., Worcester, MA
Robert J. Boland, M.D., Houston, TX
Josepha A. Cheong, M.D., Gainesville, FL
Sandra DeJong, M.D., Cambridge, MA
David C. Henderson, M.D., Boston, MA
Andrea G. Stolar, M.D., Houston, TX
Christopher R. Thomas, M.D., Galveston, TX

2023 Neurology Directors

Allison Brashear, M.D., M.B.A., Buffalo, NY, Vice Chair
Imran I. Ali, M.D., F.A.A.N., Toledo, OH
Amy Brooks-Kayal, M.D., Sacramento, CA
Terrence L. Cascino, M.D., F.A.A.N., Rochester, MN
Erika P. Greene, M.D., Houston, TX
Jaffar Khan, M.D., Atlanta, GA
Steven L. Lewis, M.D., Allentown, PA
Nina F. Schor, M.D., Ph.D., Bethesda, MD
Benn E. Smith, M.D., Scottsdale, AZ

American Board of Psychiatry and Neurology (ABPN). Available at: https://abpn.org/about/board-of-directors-and-test-committees/.

Table 57.1–4.
Pass Rates on the New ABPN Exam

Year	Candidates	Passed (%)
2016	1,496	1,353 (90)
2017	1,501	1,339 (89)
2018	1,556	1,354 (87)
2019	1,619	1,422 (88)
2020	1,537	1,369 (89)
2021	1,821	1,628 (91)
2022	1,966	1,752 (89)

Data from American Board of Psychiatry and Neurology (ABPN). Initial Certification Statistics report. Available at: https://www.abpn.org.

Going forward, all of the ABPN's examinations in the core specialties of psychiatry, child psychiatry, and neurology, as well as the subspecialties, will be single certification examinations conducted at testing centers throughout the United States. Psychiatric residency training programs must certify the credentials of their graduates through the computerized ABPN PreCert system as a condition of eligibility to take the examination. Data from this system make it feasible to accurately ascertain the quality of residency training programs as measured by certification examination passage rates for individual programs.

Also of interest is the ABPN's return to a role in the accreditation of residency training programs. The ABPN collaborates with other ABMS boards to develop a number of combined educational programs that are not ACGME accredited. The core programs contributing to the training are ACGME accredited programs, but the training track for residents in the two programs is not. Residents complete training in these programs in less time than it would take to complete the specialty training programs separately. These programs include training that leads to certification in psychiatry and neurology, psychiatry and internal medicine, and psychiatry and family medicine. The Triple Board Program was established as a collaboration between the ABPN and the American Board of Pediatrics (ABP) in 1984. Graduates of these programs may take the certification examinations in psychiatry, child and adolescent psychiatry, and pediatrics.

A small but significant number of institutions offer these combined training programs. These programs cannot be accredited by the ACGME, although the specialty programs which participate in them are reviewed and accredited by their respective RRCs. The ABPN has agreed to provide regular review and approval of the combined training programs in order to assure quality training for the residents in these programs and to assure that graduates will be eligible to take the certification examinations in each specialty (Table 57.1–5).

Table 57.1–5.
Combined Training Programs

Name	Number of Programs
Psychiatry/Internal Medicine	16
Psychiatry/Family Medicine	6
Psychiatry/Neurology	4
Pediatrics/Psychiatry/Child and Adolescent Psychiatry	11
Pediatrics Portal	54

Data from American Board of Psychiatry and Neurology (ABPN). Combined Programs and Alternative Pathways. Available at: https://abpn.org/access-residency-info/combined-programs-and-alternative-pathways/.

Table 57.1–6.
American Board of Psychiatry and Neurology (ABPN) Subspecialties and First Year of Examination

Subspecialty	Year of First Examination
Child and Adolescent Psychiatry	1959
Geriatric Psychiatry	1991
Clinical Neurophysiology[a]	1992
Addiction Psychiatry	1993
Forensic Psychiatry	1994
Pain Medicine[a]	2000
Consultation Liaison Psychiatry	2005
Sleep Medicine[a]	2007
Hospice and Palliative Medicine[a]	2008
Brain Injury Medicine[a]	2014
Addiction Medicine[a]	2017

[a] Multispecialty examination.

Modified from American Board of Psychiatry and Neurology (ABPN). ABPN website history. Available at: https://abpn.org/about/mission-and-history/.

The Subspecialties of Psychiatry. Child and Adolescent Psychiatry was the first subspecialty of psychiatry for which the ABPN developed a credentialing process and certification examination. Subsequently, nine other subspecialties were identified, and the ABPN now certifies subspecialists in these areas (Table 57.1–6).

The ABMS has specific criteria for the establishment of new subspecialties, and approval for their inclusion must be obtained from that organization. All of the subspecialties in psychiatry require completion of an ACGME accredited residency, except Brain Injury Medicine, which is still in the grandfathering period extended to 2022, and Addiction Medicine, with a grandfathering period until 2025. The ACGME PRRC accredits Child and Adolescent, Geriatric, Forensic, Consultation-Liaison (changed from Psychosomatics in 2017), and Addiction residencies. The other residencies are administered by other boards. All of the subspecialties are 1-year fellowships, except for Child and Adolescent Psychiatry, which is a 2-year program. Residents may enter Child and Adolescent Psychiatry during any part of a General Psychiatry program. One year of training may be counted for both programs, enabling the completion of both programs in 5 years rather than 6. A previous discussion in the field was whether to allow completion or partial completion of the other subspecialty fellowships in the fourth year of the general residency. Although the discussion continues, none of the steps that would be required for such a change have been initiated. All of the subspecialty certifications require taking a single computerized certification examination except for Child and Adolescent Psychiatry. That specialty had traditionally had a Part II oral examination, which was eliminated in 2016 after the completion of the last oral examination at Northwestern in Chicago.

Many subspecialties are developed collaboratively between several ABMS boards. Sleep medicine is an example of such a subspecialty. One of the boards will take the lead in administering the subspecialty and developing the certification examination. Each member board is responsible for credentialing candidates from their own specialty. As new subspecialties are recognized by the ABPN and ABMS, a grandfathering period, usually 5 years, is allowed during which candidates who meet established criteria for experience in the subspecialty area may take the certification examination without completing a residency. After this period expires, candidates must complete appropriate residency training programs to be eligible to take the exam. Of particular recent interest is the development of the multidisciplinary subspecialty of Addiction Medicine, which is administered, oddly, by the American Board of Preventative Medicine. There are now more fellowship programs in Addiction Medicine than in Addiction Psychiatry, and many previously uncertified addiction specialists have taken advantage of the grandfathering period to become certified. The Practice Pathway for such individuals will remain open until 2025.

Table 57.1–7.
Psychiatry and Subspecialty Training Programs Numbers and Resident Census 2015–2020

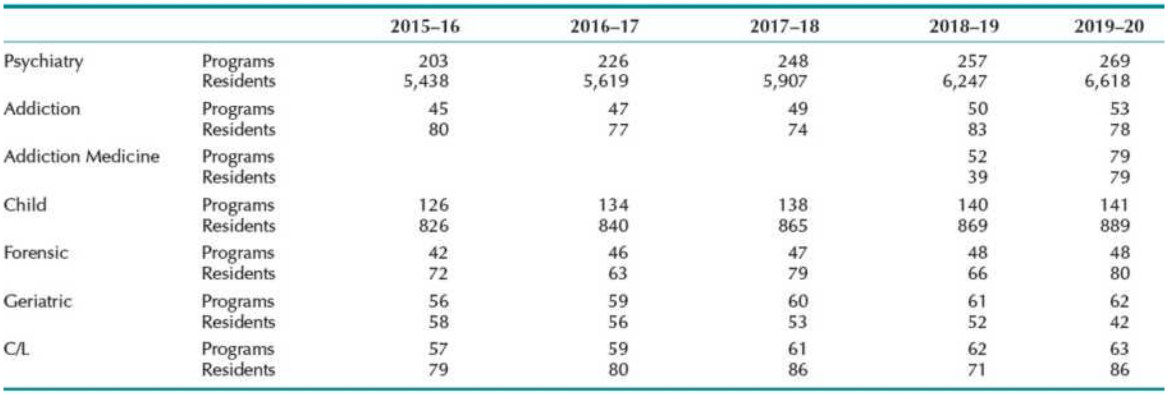

		2015–16	2016–17	2017–18	2018–19	2019–20
Psychiatry	Programs	203	226	248	257	269
	Residents	5,438	5,619	5,907	6,247	6,618
Addiction	Programs	45	47	49	50	53
	Residents	80	77	74	83	78
Addiction Medicine	Programs				52	79
	Residents				39	79
Child	Programs	126	134	138	140	141
	Residents	826	840	865	869	889
Forensic	Programs	42	46	47	48	48
	Residents	72	63	79	66	80
Geriatric	Programs	56	59	60	61	62
	Residents	58	56	53	52	42
C/L	Programs	57	59	61	62	63
	Residents	79	80	86	71	86

Modified with permission from Accreditation Council for Graduate Medical Education (ACGME). Data Resource Book Academic Year 2019–2020. Available at: https://www.acgme.org.

The number of core psychiatry subspecialty programs and subspecialty fellows has not kept pace with the increase in the numbers of general programs and residents. Except for Child and Adolescent Psychiatry, these programs have fewer than 90 fellows in training each year. This is problematic from the standpoint of providing an adequate workforce, particularly in geriatric and addiction psychiatry. The failure to interest sufficient numbers of residents in subspecialty fellowship training is the subject of considerable discussion and distress for the profession (Table 57.1–7).

The American Board of Medical Specialties

In 1933, the American Hospital Association (AHA), the Association of American Medical Colleges (AAMC), the Federation of State Medical Boards (FSMB), the AMA, the National Board of Medical Examiners (NBME), and representatives of the four specialty boards in existence at that time (Ophthalmic Examinations, Otolaryngology, Obstetrics and Gynecology, and Dermatology and Syphilology), established the Advisory Board for Medical Specialties. Originally the ABMS was to serve as an advisor to the Boards, a forum for discussion of common problems, and a coordinator of board work when feasible. When expressly delegated by the Boards, the ABMS was to assume control over common policies and procedures. The organization was restructured in 1970 into the ABMS. The AHA, AAMC, FSMB, AMA, and NBME, along with the CMSS, the ACGME, and the Educational Commission for Foreign Medical Graduates

(ECFMG), continue to serve as associate members of the ABMS. The ABMS derives its funding through levies on the Specialty Boards and, in recent years, has assumed a more directive role with its 24 member Boards. In particular, the ABMS has established standards that the boards are required to follow, specifying what steps diplomats of the boards must take in order to maintain their certification. The ABMS has proposed a revision to these standards that emphasize quality improvement in health care and professionalism. The proposed ABMS requirement that the Boards assess professionalism in their diplomats is controversial for practical and legal reasons.

These Maintenance of Certification (MOC) or Life Long Learning programs currently consist of four elements: professional standing (medical licensure), continuing medical education (CME), self-assessment (SA), and performance in practice (PIP), which includes patient and peer feedback and systematic evaluation of the diplomat's clinical care. Additionally, all diplomats of the boards must document the completion of a patient safety course.

In conjunction with the ACGME, the ABMS and some member boards have begun to offer U.S. training standards for residencies and practitioner certification abroad. Singapore was the first country to request this program and now has a robust program of residency training and certification for several specialties that is similar to the U.S. equivalents. ACGME international programs are now training residents in China, Guatemala, Haiti, Kenya, Lebanon, Oman, Pakistan, Qatar, Saudi Arabia, and the United Arab Emirates, in addition to Singapore. ABMS International has helped to develop certification in examinations in 12 specialties in Singapore, including a certification examination in Psychiatry.

The Liaison Committee on Medical Education

Interestingly, accreditation processes for undergraduate medical education were unified later than the initial wave of certification for specialists. The Liaison Committee on Medical Education was not founded until 1942. Prior to its founding, the AMA's Council on Medical Education and Hospitals inspected and approved medical schools and approved hospitals for internships. The AAMC approved medical schools for membership in the AAMC. Leaders of the AAMC and the AMA, in an effort to avoid duplication of effort, assure the quality of training, protect medical students from the wartime draft, and counteract pressures to shorten medical training, established the LCME at a meeting in Chicago. The LCME established and has maintained the standard that made psychiatric education and rotations one of the required experiences for all medical students. This standard has been critical to maintaining an appropriate place for psychiatric education in the medical school curriculum when it has been challenged on intellectual and, more recently, economic grounds.

Table 57.1–8.
PGY1 Match Rates

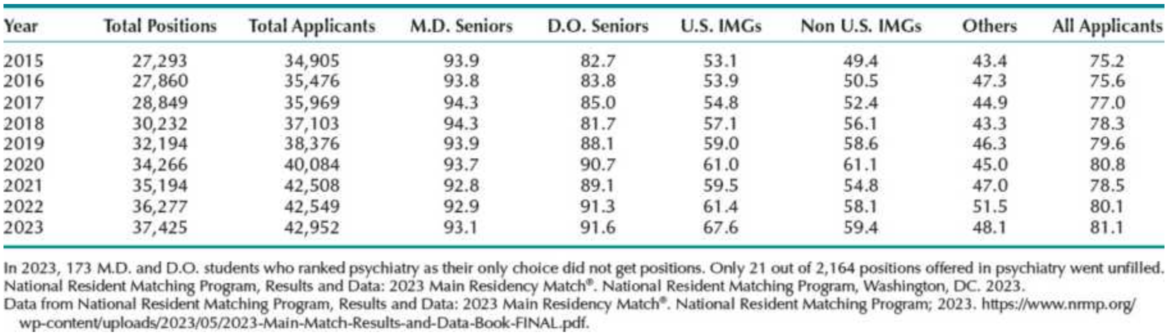

Year	Total Positions	Total Applicants	M.D. Seniors	D.O. Seniors	U.S. IMGs	Non U.S. IMGs	Others	All Applicants
2015	27,293	34,905	93.9	82.7	53.1	49.4	43.4	75.2
2016	27,860	35,476	93.8	83.8	53.9	50.5	47.3	75.6
2017	28,849	35,969	94.3	85.0	54.8	52.4	44.9	77.0
2018	30,232	37,103	94.3	81.7	57.1	56.1	43.3	78.3
2019	32,194	38,376	93.9	88.1	59.0	58.6	46.3	79.6
2020	34,266	40,084	93.7	90.7	61.0	61.1	45.0	80.8
2021	35,194	42,508	92.8	89.1	59.5	54.8	47.0	78.5
2022	36,277	42,549	92.9	91.3	61.4	58.1	51.5	80.1
2023	37,425	42,952	93.1	91.6	67.6	59.4	48.1	81.1

In 2023, 173 M.D. and D.O. students who ranked psychiatry as their only choice did not get positions. Only 21 out of 2,164 positions offered in psychiatry went unfilled. National Resident Matching Program, Results and Data: 2023 Main Residency Match®. National Resident Matching Program, Washington, DC. 2023.

Data from National Resident Matching Program, Results and Data: 2023 Main Residency Match®. National Resident Matching Program; 2023. https://www.nrmp.org/wp-content/uploads/2023/05/2023-Main-Match-Results-and-Data-Book-FINAL.pdf.

PSYCHIATRIC EDUCATION TODAY

Resident Recruitment and the NRMP Match

Postgraduate psychiatric education today begins with what everyone calls “the Match.” Prior to the establishment of The National Resident Matching Program (NRMP), enrollment in a residency training program depended on individual negotiations between graduating medical students and the training program’s director. Students were often pressured to accept a position in a program immediately, frequently before they had had an opportunity to interview for and consider other programs. The NRMP was established in 1952 in an effort to prevent these abuses and provide a level playing field for all the programs and applicants. The NRMP is governed by a board of directors, with nominees from the ABMS, the AMA, the AAMC, the AHA, the CMSS, the AAMC Organization of Student Representatives, the American Medical Student Association, the AMA Medical Student Section, and the Consortium of Medical Student Organizations.

Each year, the recruitment process, in anticipation of the NRMP match day in March, preoccupies most GPE program directors (PDs), senior medical students, and many current residents who are very interested in the younger colleagues who might join them in their residency program. Selecting and recruiting strong candidates for residency training is a critical task for departments. During the recruitment season in 2020 and 2021, all recruitment interviews were done through virtual means due to the COVID pandemic. It remains to be seen whether this methodology will become a permanent feature in future recruitment years. Successful recruitment benefits the department’s training program, its future faculty, and the communities which depend on the department for the services of its graduates. Ideal candidates, like those described by Johann Christian Reil in 1803, are the subject of intense competition between training programs.

> Perspicacity, a talent for observation, intelligence, good will, persistence, patience, experience, an imposing physique, and a countenance that commands respect. All those qualities required for the cure of the insane are so rare as to make the staffing of institutions difficult.

It is not much of an exaggeration to say that good residents make good teachers and good training programs. Bright, enthusiastic, conscientious residents are rewarding to teach and contribute to a stimulating scholarly environment. Unfortunately, residents who display the opposite characteristics can cast a pall over the entire training program.

The NRMP has specific guidelines that govern the interview and recruitment process for residency applicants, and all GPE programs and medical students must agree to comply with them in order to participate in the NRMP match. The NRMP has established a uniform date of appointment for GME positions and a system for matching medical students and residency training programs that depends on lists that rank program preference (by the applicants) and applicant preference (by the programs). These lists are computer-matched to produce the best possible outcome for both parties. In 2021, almost 27,000 U.S. medical students (M.D. and D.O.) participated in the NRMP match, and almost 22,000 “independent” applicants applied for approximately 38,000 positions in over 4,000 GME programs across the United States. “Independent” applicants include former graduates of U.S. medical schools, U.S. osteopathic students, Canadian students, and graduates of international medical schools.

Since 1995, the number of applicants for these positions has substantially exceeded the number of positions. This trend has accelerated over the last 10 years with the inclusion of many more U.S. IMGs and Osteopathic medical students in the Match. This has produced substantial anxiety on the part of today's applicants regarding their likelihood of obtaining a residency position. In fact, the probability of a graduate of a U.S. institution succeeding in the Match has changed little over the last 25 years. About 90% of these graduates will match with a program they have ranked. Other candidates have cause for considerable anxiety. They will succeed in obtaining a position less than 50% of the time. In the last several years, the number of psychiatry training programs and applicants has grown dramatically. Almost all residency training programs in General Psychiatry fill their positions in the match. Recently, many graduates of U.S. medical schools have not been able to obtain a residency position in psychiatry. The proportion of residents from U.S. medical schools is now higher than what is found in Internal Medicine and Family Medicine residency training programs (Tables 57.1–8 and 57.1–9).

Table 57.1–9.
Psychiatry Residency Training Positions, Number of Positions Filled, and Number of Positions Filled by U.S. Graduates 2017–2021

Year	Positions	Positions Filled	Filled by U.S. Graduates
2017	1,495	1,491 (99.7)	923 (61.7)
2018	1,556	1,540 (99.0)	982 (63.1)
2019	1,740	1,720 (98.9)	1,054 (60.6)
2020	1,858	1,838 (98.9)	1,138 (61.2)
2021	1,907	1,904 (99.8)	1,205 (63.2)

Data from National Resident Matching Program, Results and Data: 2023 Main Residency Match®. National Resident Matching Program; 2023, Tables 7A and 7B. https://www.nrmp.org/wp-content/uploads/2023/05/2023-Main-Match-Results-and-Data-Book-FINAL.pdf.

Funding of Residency Positions

The residency positions for which these candidates are competing are funded by a variety of means. The federal government is by far the biggest sponsor of training stipends through two means. First, teaching hospitals receive direct and indirect graduate medical education funding (DME and IME) directly from the Centers for Medicare and Medicaid Services (CMS). These payments are based on a complicated formula that takes into account the percentage of Medicare patients treated at the facility, the historical number of positions when the program was started, and regional cost variations. Many hospitals and universities have many more residents than are approved and funded under the “cap” allowed by this formula, and there are dramatic differences in reimbursement by region. The second mechanism is funds provided by the Veteran’s Administration for residency positions, which in many programs provide a substantial number of stipends. State mental health divisions, community mental health centers, private hospitals, and other organizations also fund some residency positions. All of these organizations expect clinical services in return for the funds they provide. It is often very challenging for a PD to construct a program that satisfies the clinical service demands of the funders and provides the educational program required by the PRs of the ACGME PRRC.

Program Requirements for Psychiatric Training

The ACGME’s process for determining the requirements that programs must meet involves obtaining extensive input from the field, revisions by the RRC followed by more feedback from the field, and subsequent revisions. The ACGME Committee on Requirements reviews the PRs and has the authority to withhold final approval. Increasingly the ACGME has sought to standardize the language and content of the PRs across specialties, so that much of what is found in the psychiatric and psychiatric subspecialty PRs is the same as in other specialties. The PPRs cover the issues of institutional sponsorship and resources, qualifications of the PD and faculty, creation of an academic environment, necessary clinical experiences, mandatory educational curriculum, work conditions, evaluation of residents, evaluation of the program and faculty, and specific outcomes in the language of the ACGME competencies. Recently, the process of evaluating residents has become much more systematic with the development of the milestones (discussed later in this chapter).

Institutional Requirements. Institutions must make a substantial investment in their residency training programs in order to become ACGME-accredited. Institutional accreditation is required before any residency training program can be approved. The institution must appoint a Designated Institutional Official (DIO) who has authority and responsibility for the oversight and administration of the institution's ACGME accredited programs and is ultimately responsible for compliance with the ACGME Institutional, Common, and Specialty PRs. The institution must also establish a Graduate Medical Education Committee (GMEC), which conducts annual program reviews on all residency training programs and, in general, is responsible for the maintenance of a safe and effective learning environment. Many other aspects of institutional investment are mandated, including sections on resident contracts, work environment, learning environment, and responsiveness to ACGME oversight. The ACGME has recently started a new program of institutional review entitled Clinical Learning Environment Review (CLER). The purpose of this review is to assess the GME learning environment of each sponsoring institution and its participating sites. A site visit is conducted during which the site visitors assess the institution through direct observation and interviews of residents and other personnel. This program duplicates much of the work of the Joint Commission on Hospital Accreditation (JCAHO).

The PPRs further elaborate on what is required to conduct a psychiatric residency training program. The subspecialty PPRs are similar and will not be separately covered. The requirements include special provisions addressing the physical facilities which are necessary for psychiatric training. Adequate office space for interviewing patients and equipment for recording sessions with patients are examples of these types of requirements. The types, length, and supervision of clinical experience specific to psychiatry are detailed in the PPRs, as are certain aspects of the ACGME competencies.

The Program Director. The PPRs specify that the residency training PD must be an ABPN-certified psychiatrist with sufficient administrative and educational experience to take on the challenging task of leading the program. The PD is expected to have full authority over the program and to have sufficient tenure in the position to be effective. The core duties of the PD are mandated in the PRs. These include the selection of training sites and the development of program letters of agreement (PLAs) between external sites and the sponsoring institution. The PD must appoint a site coordinator and faculty for each rotation. The PD is responsible for conducting regular evaluations of the rotations and of the faculty and for removing faculty from their teaching roles if necessary. The PD must select residents for the program, evaluate their performance, monitor their stress, and provide twice-yearly formal evaluations. Compliance with ACGME and institutional policies and procedures, particularly those concerning limitations of duty hours and grievance procedures, is required. A summative evaluation at the end of residency training is required, and the resident's rotations and educational experiences must be conveyed to the ABPN at the end of residency to assure eligibility to take the certification examination. The PRRC has, not surprisingly, required that the institution fund and the PD devote at least one-half of their efforts to the training program. Success in this highly technical and demanding position requires at least this much effort.

The Faculty. Psychiatric faculty in an RTP must also be certified by the ABPN, have a strong interest in education, and devote sufficient time to fulfill their supervisory and educational duties. Other teaching physicians should be certified by their specialty board. Non-physician faculty must have completed training, certification, and licensure appropriate to their profession. The maintenance of an academic, scholarly environment is critical to the success of the educational program. All of the faculty should regularly participate in clinical teaching rounds, conferences, grand rounds, journal clubs, and other activities common to the academic setting. Some of the faculty must be engaged in research activities. Demonstration of the faculty's scholarship may be through acquisition of peer-reviewed funding, publication of peer-reviewed articles, other publications, presentations at regional and national meetings, participation in quality assurance activities, and participation and leadership in regional and national professional organizations.

Clinical Experiences. The clinical experiences that psychiatric residents must have during residency are listed in Table 57.1–10. Note that in the first year of residency, the internship must include 4 months of experience in a primary care setting. The purpose and format of the primary care experience are subjects of considerable debate in the field at this time. Currently, most residents complete these experiences in inpatient Internal Medicine wards caring for very ill patients for a brief period of time. These clinical situations are unlikely to be encountered regularly in practice by most psychiatrists. They do, however, have a formative effect on the resident's identity as a physician. Experience in ambulatory or integrated care settings may prove to be more valuable to the psychiatrists of the future. Currently, the PPRs do not require training in the collaborative care environments that many psychiatrists and the APA consider essential to the future delivery of psychiatric care.

Table 57.1–10.
Required Psychiatric Graduate Education Clinical Experiences

Experience	Months Required
Primary Care	4
Neurology	2
Inpatient Psychiatry	6–16
Outpatient Psychiatry	12
Child and Adolescent Psychiatry	2
Geriatric Psychiatry	1
Addiction Psychiatry	1
Consultation Liaison Psychiatry	2
Forensic Psychiatry	_[a]
Community Psychiatry	_[a]
Emergency Psychiatry	_[a]

[a] Required experiences without specific time requirements.

Data from Accreditation Council for Graduate Medical Education (ACGME) Program Requirements for Graduate Education in Psychiatry. Available at: https://www.acgme.org.

Most programs include the required 2 months of experience in neurology in the first year, leaving 6 months which is usually comprised of required psychiatric rotations. Inpatient experiences must include at least 6 months with direct responsibility for the care of acutely ill general psychiatric patients. A wide variety of settings are acceptable for this experience, including traditional inpatient wards, research wards, and specialty care units. The outpatient requirement of a year must include longitudinal treatment of patients with multiple modalities. Some of the resident's patients must be seen weekly under supervision for psychotherapeutic treatment. Forensic, emergency, and community psychiatry experiences are not required to have a minimum duration but must include certain key experiences. For example, in their emergency psychiatry experience, residents must evaluate, manage, and triage acutely ill psychiatric patients. Elective opportunities that enhance and specialize residents' training experiences must be available. These are usually provided in the more senior years of the residency (Table 57.1–10).

A unique aspect of the PPRs is the ability to double-count certain experiences for the completion of both the general and child and adolescent residencies. This enables residents to complete training in general and child psychiatry in only 5 years. One month of child neurology, pediatric consultation, addiction psychiatry, and 20% of the resident's ambulatory psychiatry experience may be double-credited, and the experiences in community and forensic psychiatry can count for both programs. This feature, recognized by the ABPN as a legitimate path to eligibility for board certification, has been important to the successful recruitment of residents into child and adolescent training programs.

Didactic Curriculum. A formal didactic curriculum must be provided for the residents during training. The PD is expected to establish goals and objectives for the curricular components that are couched in terms of the ACGME competencies. These have now been considerably elaborated and are covered in the section below entitled Psychiatric Milestones. The curriculum must be designed in a progressive manner that enables resident progression in their knowledge of the field. It is not acceptable to provide the same seminar material to all years of residents simultaneously. The time for the didactics must be protected from clinical duties so that resident attendance is not compromised. The program must document that residents attend at least 70% of these didactic sessions.

Evaluation of Residents. Due to recent radical changes in the process of program evaluation discussed below, the process of resident evaluation in training programs has drastically changed. The ACGME now requires programs to report on each resident's progress twice yearly. A requirement for a Core Competencies Committee (CCC) has been established. The PD is expected to appoint the CCC, which must consist of at least three faculty members. Residents may not be members of the CCC. The committee must utilize evaluations provided by clinical supervisors and other personnel on rotations, test scores from the PRITE examination, Clinical Skills Examinations (CSEs), and other evaluative methods. The CCC must rate each resident on the milestones semiannually, and these ratings must be reported to the ACGME. These evaluations are developmental in approach and indicate areas of strength and weakness. The resident, PD, and faculty are expected to work collaboratively to improve resident performance in areas of relative weakness.

Summative evaluations at the end of each residency year and at the end of residency must be recorded so that there is a record of the resident's completion of all required rotations and attainment of sufficient competency on the ACGME competencies to enable independent practice and ethical and professional behavior. Uniquely in psychiatric (and neurologic) training, successful completion of clinical examinations, in which the resident interviews a patient while attaining and maintaining rapport, conducts a mental status examination, and presents the data collected in an organized fashion, is required by the ABPN for eligibility to take the board examination. Psychiatry residents must successfully complete at least three such evaluations during residency. This in-residency oral examination was established by the ABPN as a requirement for eligibility to take the certification examination and is in the PPRs. It replaces Part II of the old board examination. The residency program must provide all of this data to the ABPN through the PreCert system to ensure the eligibility of the resident to take the board examination.

The Psychiatry Milestones

In 1999, the ACGME and ABMS, realizing that medical practice required more than medical knowledge and skill in patient care and procedures, established additional competencies which physicians are expected to attain during the course of their training. These competencies should be regarded as desired outcomes that medical training produces. The six Core Competencies are Patient Care, Medical Knowledge, Practice-Based Learning and Improvement, Interpersonal and Communication Skills, Systems-Based Practice, and Professionalism. Requirements for inclusion of training in these areas are included in all of the ACGME PRs and describe behaviors, attributes, or outcomes that must be demonstrated by residents as they advance in training. For example, in the competency of Professionalism: “Residents must demonstrate a commitment to carrying out professional responsibilities and an adherence to ethical principles. Residents are expected to demonstrate: compassion, integrity, and respect for others; responsiveness to patient needs that supersedes self-interest; respect for patient privacy and autonomy; accountability to patients, society and the profession; sensitivity and responsiveness to a diverse patient population, including but not limited to diversity in gender, age, culture, race, religion, disabilities, and sexual orientation; and (specialty-specific outcomes).”

Table 57.1–11.
Authors of the Psychiatry Milestones 1.0

Working Group	Advisory Group
Chris Thomas, M.D., Chair	Chris Thomas, M.D., Chair
Sheldon Benjamin, M.D.	Timothy Brigham M.Div., Ph.D.
Adrienne Bentman, M.D.	Carol Bernstein, M.D.
Robert Boland, M.D.	Beth Ann Brooks, M.D.
Deborah Cowley, M.D.	Larry Faulkner, M.D.
Jeffrey Hunt, M.D.	Deborah Hales, M.D.
George Keepers, M.D.	Victor Reus, M.D.
Louise King, M.S.	Richard Summers, M.D.
Gale Manos, M.D.	
Donald Rosen, M.D.	
Kathy Sanders, M.D.	
Mark Servis, M.D.	
Kallie Shaw, M.D.	
Susan Swing, Ph.D.	
Alik Widge, M.D., Ph.D.	

Source: Accreditation Council for Graduate Medical Education. Data abstracted from The Psychiatry Milestones.

The milestones for residency training represent the next iteration in specifying the outcomes rather than the processes of residency training. Detailed accounts of the development of the ACGME milestones project have been published in the *Journal of Graduate Medical Education* and the *New England Journal of Medicine*. In brief, public and political concern about the outcomes of medical training resulted in a call for reform from the Institute of Medicine (2000), Council on Graduate Medical Education (2007), Josiah Macy Jr. Foundation (2011), and Medicare Payment Advisory Commission (MedPAC) reports (2011). The ACGME response to this crisis in 2008 was to require the creation of milestones describing resident progress in the Core Competencies for all specialties. This project was conceived as a method for evaluation of individual resident progress that would be uniformly applied across the United States and as a method to evaluate the efficacy of residency programs in producing the desired outcomes.

Internal Medicine, Pediatrics, and Surgery were the first specialties to begin milestone development in 2009. In 2011, psychiatry and the other remaining specialties began development with the goal of completing all specialty milestones by 2014. Representatives from the ABPN, APA, American Association of Directors of Psychiatric Residency Training (AADPRT), and ACGME selected individuals to participate in milestone development for psychiatry, and the Working Group and the Advisory Group were formed (Table 57.1–11).

The Working Group was charged with milestone development and began meeting in December of 2011. The Advisory Group reviewed this work and suggested potential improvements. The development of milestones required five meetings and numerous conference calls over the course of the next 2 years. Extensive feedback from the field was sought as part of this process and was incorporated into the final documents. Details of the Psychiatry Milestone development process have been published in Academic Psychiatry. In 2016, the ACGME began the process of revising the milestones due to dissatisfaction with the complexity and inconsistencies in Milestones 1.0. The rationale and processes involved were described in the Journal of Graduate Medical Education. A Milestones 2.0 Group was selected for the revision of the Psychiatry Milestones and developed a set of milestones in which the subcompetencies in System-Based Practice, Problem-Based Learning and Improvement, Professionalism, and Interpersonal Communication were more closely aligned with other specialties. Additionally, the Workgroup developed a Supplemental Guide to assist faculty in assessing resident progress (Table 57.1–12).

Table 57.1–12.
Authors of the Psychiatry Milestones 2.0

Elie Aoun, M.D.
Jeff Aronowitz, M.D.
Adrienne Bentman, M.D.
Lynneice Bowen, M.D.
Brittany Campbell
Deborah Cowley, M.D.
Sandra DeJong, M.D., M.Sc.
Laura Edgar, Ed.D., C.A.E.
Lillian Joy Houston, M.D.
Furhut Janssen, D.O.
George Keepers, M.D.
Mark Kinzie, M.D.
Matthew Macaluso, D.O.
Patcho Santiago, M.D.

Source: Accreditation Council for Graduate Medical Education. Data abstracted from The Psychiatry Milestones 2.0.

The milestones are intended to allow faculty to rate the development of a resident’s skills, ability, knowledge, attitudes, and behavior in achieving the ACGME competencies during training. They are developmental in nature. Basic skills in an area are expected at the beginning of training. Residents are expected to enhance these skills and, by the end of the residency, attain a level of accomplishment that will allow independent practice. The milestones are organized for the purpose of rating resident accomplishments into five levels. Level 1 corresponds to what is expected of a beginning resident. As residents complete their rotations and educational experiences, they are expected to attain Levels 2 and 3 as intermediate stages of accomplishment. It is important to note that the levels in the milestones do not correspond to PGY levels. Residents will progress at different rates within a program, and programs differ in what experiences they offer residents in each year of residency training. By the completion of residency, residents will, on average, have progressed to Level 4 on most milestones. This accomplishment, however, is not sufficient for graduation. The milestones do not measure every quality that is important, and the PD must make an independent judgment on the readiness of the graduating resident for independent practice (Table 57.1–13).

Table 57.1–13.
Description of the ACGME Milestones Levels

Has not achieved Level 1. The resident does not demonstrate the milestones expected of an incoming resident

Level 1. The resident demonstrate milestones expected of an incoming resident

Level 2. The resident is advancing and demonstrates additional milestones but is not yet performing at a mid-residency level

Level 3. The resident continues to advance and demonstrate additional milestones; the resident demonstrates the majority of milestones targeted for residency in this subcompetency.

Level 4. The resident has advanced so that they now substantially demonstrate the milestones targeted for residency. This level is designed as the graduation *target* and does not represent a graduation *requirement.* Making decisions about readiness for graduation is the purview of the residency program director.

Level 5. The resident has advanced beyond performance targets set for residency and is demonstrating “aspirational” goals, which might describe the performance of someone who has been in practice for several years. It is expected that only a few exceptional residents will reach this level.

Accreditation Council for Graduate Medical Education. Data abstracted from discussions in the Psychiatry Milestones Committees.

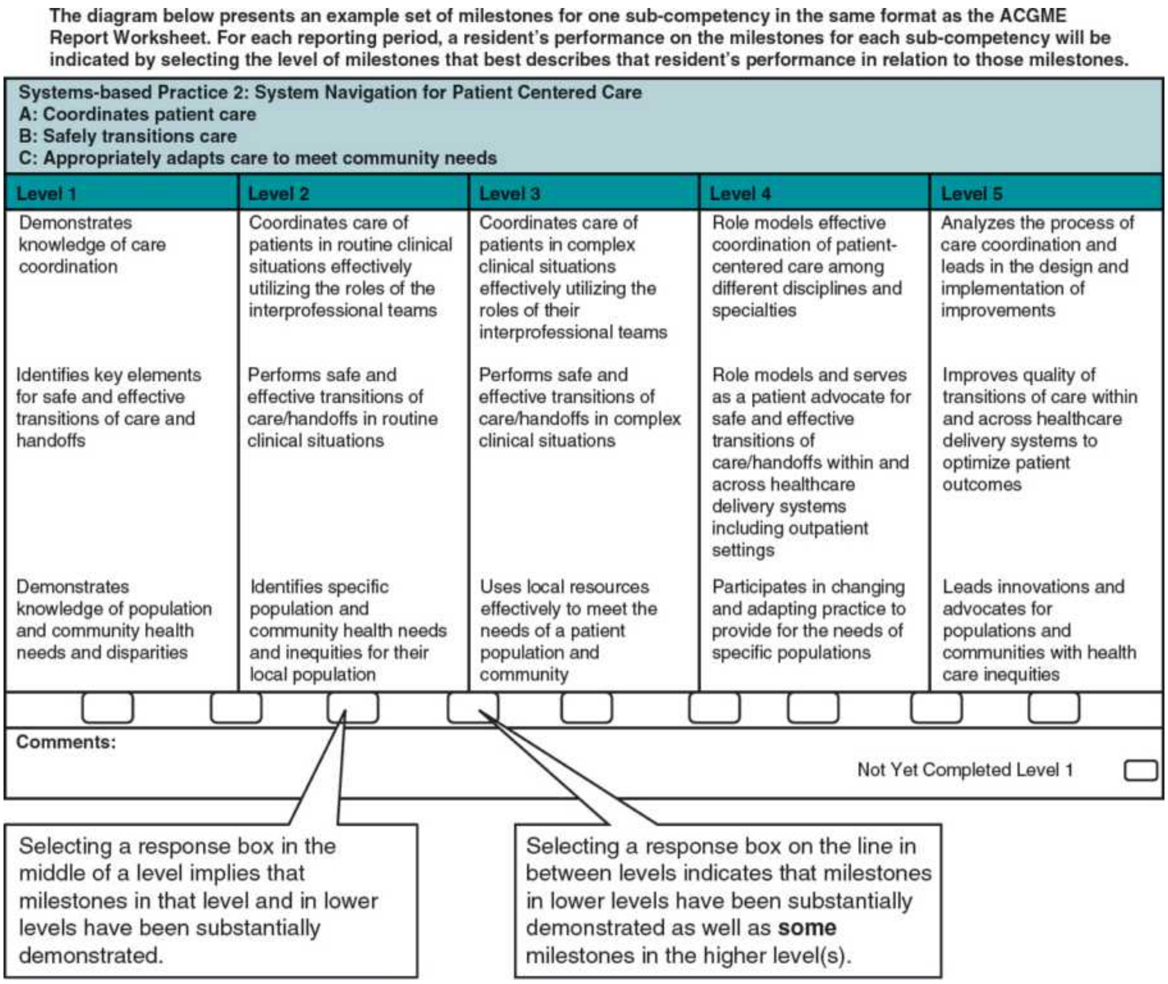

The diagram below presents an example set of milestones for one sub-competency in the same format as the ACGME Report Worksheet. For each reporting period, a resident's performance on the milestones for each sub-competency will be indicated by selecting the level of milestones that best describes that resident's performance in relation to those milestones.

Systems-based Practice 2: System Navigation for Patient Centered Care
A: Coordinates patient care
B: Safely transitions care
C: Appropriately adapts care to meet community needs

Level 1	Level 2	Level 3	Level 4	Level 5
Demonstrates knowledge of care coordination	Coordinates care of patients in routine clinical situations effectively utilizing the roles of the interprofessional teams	Coordinates care of patients in complex clinical situations effectively utilizing the roles of their interprofessional teams	Role models effective coordination of patient-centered care among different disciplines and specialties	Analyzes the process of care coordination and leads in the design and implementation of improvements
Identifies key elements for safe and effective transitions of care and handoffs	Performs safe and effective transitions of care/handoffs in routine clinical situations	Performs safe and effective transitions of care/handoffs in complex clinical situations	Role models and serves as a patient advocate for safe and effective transitions of care/handoffs within and across healthcare delivery systems including outpatient settings	Improves quality of transitions of care within and across healthcare delivery systems to optimize patient outcomes
Demonstrates knowledge of population and community health needs and disparities	Identifies specific population and community health needs and inequities for their local population	Uses local resources effectively to meet the needs of a patient population and community	Participates in changing and adapting practice to provide for the needs of specific populations	Leads innovations and advocates for populations and communities with health care inequities

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Comments:

Not Yet Completed Level 1 ☐

Selecting a response box in the middle of a level implies that milestones in that level and in lower levels have been substantially demonstrated.

Selecting a response box on the line in between levels indicates that milestones in lower levels have been substantially demonstrated as well as **some** milestones in the higher level(s).

FIGURE 57.1–1. An example of milestones subcompetency performance level reporting. (Reprinted with permission from Psychiatry Milestones, 2nd Revision. © 2020 Accreditation Council for Graduate Medical Education (ACGME). Available at: https://www.acgme.org.)

In the milestones, each of the six ACGME competencies contains subcompetencies that describe specialty-specific elements of the general Core Competency. In the Psychiatry Milestones 2.0, there are 20 subcompetencies. There are an additional 32 subheadings in the milestones, which are called “threads.” The Milestones 2.0 revision reduced the number of areas requiring evaluation by 50% (Fig. 57.1–1).

Subcompetencies describe the progression of specific skills, knowledge, or attributes in a single area, such as psychiatric interviewing. The underlying milestones are organized into threads that describe the component skills of the subcompetency and the progression of these skills across the levels (Fig. 57.1–2).

The psychiatry milestones, subcompetencies, and threads are presented in Table 57.1–14.

Milestones for each of the subspecialties of psychiatry were completed after work on the general psychiatry milestones was finished. This work utilized the experience of the original milestone committee members who participated and chaired the subspecialty development committees to rapidly complete this task. A schematic outline of the subspecialty milestones was created to assist the committees in their task. Committee members met for a single face-to-face working session at the ACGME headquarters in Chicago. The Advisory Committee reviewed the work accomplished at the meeting and provided suggestions. The Working Group subsequently accomplished the revisions through conference calls and online participation. Subspecialty milestones are currently being revised (Table 57.1–15).

The psychiatry milestones represent a major paradigm shift in the overall approach that programs take to resident training and assessment because of their focus on outcomes and their specificity. Because of the process of development, they represent a true consensus of psychiatric educators regarding what residents need to accomplish and what programs need to teach. This shift in emphasis is not, however, without its critics. They have pointed out correctly that there is no evidence that utilization of the milestones will actually produce better psychiatrists. Detractors advocate resident assessment methods that are based on actual patient outcomes. The critics would prefer that there be no change in educational methods until prospective studies establish that there will be a positive outcome. Acceding to such objections would essentially guarantee that postgraduate education remains frozen in its current state. The ACGME has instead adopted a program based on modern quality improvement methods, essentially a PDSA cycle.

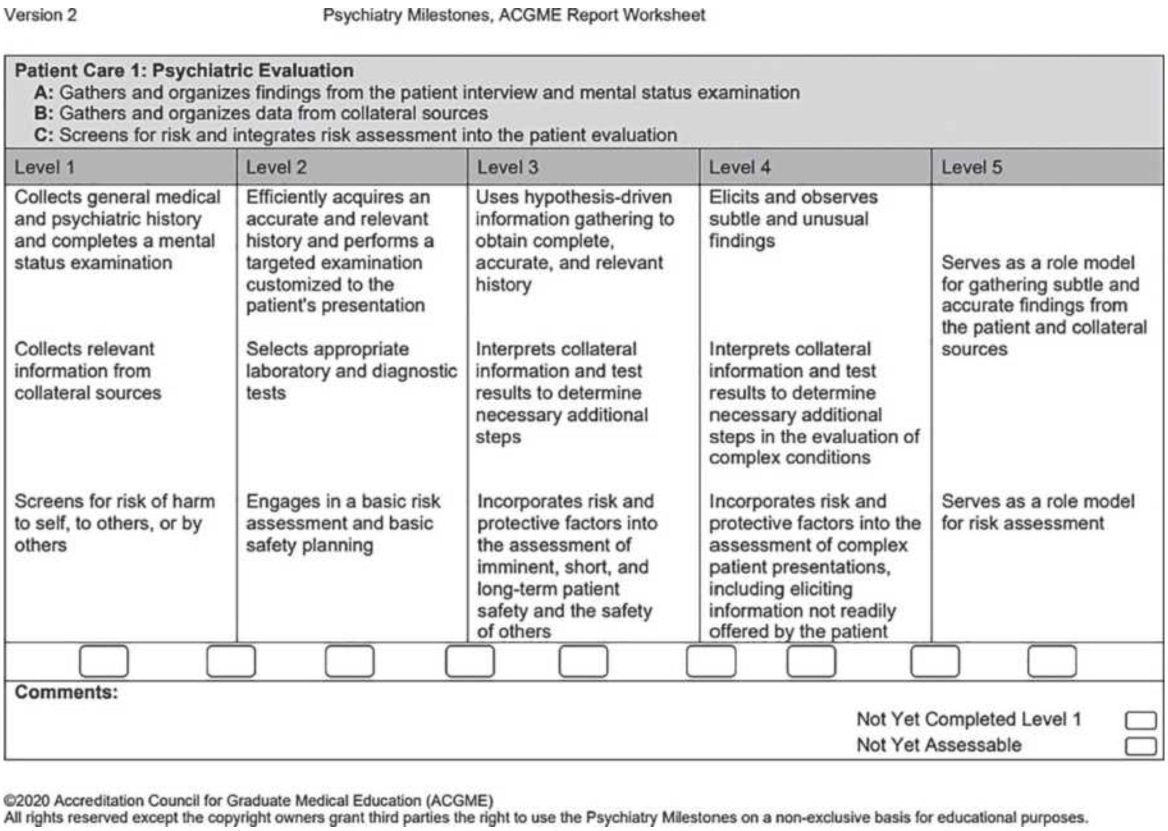

Version 2 Psychiatry Milestones, ACGME Report Worksheet

Patient Care 1: Psychiatric Evaluation
A: Gathers and organizes findings from the patient interview and mental status examination
B: Gathers and organizes data from collateral sources
C: Screens for risk and integrates risk assessment into the patient evaluation

Level 1	Level 2	Level 3	Level 4	Level 5
Collects general medical and psychiatric history and completes a mental status examination	Efficiently acquires an accurate and relevant history and performs a targeted examination customized to the patient's presentation	Uses hypothesis-driven information gathering to obtain complete, accurate, and relevant history	Elicits and observes subtle and unusual findings	Serves as a role model for gathering subtle and accurate findings from the patient and collateral sources
Collects relevant information from collateral sources	Selects appropriate laboratory and diagnostic tests	Interprets collateral information and test results to determine necessary additional steps	Interprets collateral information and test results to determine necessary additional steps in the evaluation of complex conditions	
Screens for risk of harm to self, to others, or by others	Engages in a basic risk assessment and basic safety planning	Incorporates risk and protective factors into the assessment of imminent, short, and long-term patient safety and the safety of others	Incorporates risk and protective factors into the assessment of complex patient presentations, including eliciting information not readily offered by the patient	Serves as a role model for risk assessment
☐ ☐	☐ ☐	☐ ☐	☐ ☐	☐

Comments:

Not Yet Completed Level 1 ☐
Not Yet Assessable ☐

©2020 Accreditation Council for Graduate Medical Education (ACGME)
All rights reserved except the copyright owners grant third parties the right to use the Psychiatry Milestones on a non-exclusive basis for educational purposes.

FIGURE 57.1–2. Example of an actual subcompetency and its threads from the ACGME psychiatry milestones 1.0. Note how the threads describe progression of skills across the levels. (Reprinted with permission from Psychiatry Milestones, 2nd Revision. © 2020 Accreditation Council for Graduate Medical Education (ACGME). Available at: https://www.acgme.org.)

Assessment of Residency Training Programs

The milestones were not only conceptualized as a method for evaluation of resident progress but are also intended to be used eventually as a method of programs evaluation, and there have been many other changes to the accreditation process.

Incorporation of Osteopathic Training Programs. The ACGME is now responsible for the accreditation of all residency training programs in the United States. In 2014, the American Osteopathic Association (AOA) entered into an agreement with the ACGME to consolidate the accreditation of osteopathic residency training programs under the ACGME. AOA accredited programs could elect to apply for ACGME accreditation until 2020. AOA specialty board certification for faculty in these programs was acceptable to the ACGME during the transition period. The programs transitioned to the ACGME may have a co-director certified by AOA specialty boards as long as the other co-director is ABMS-certified. The AOA and AACOM were integrated into the governance structure of the ACGME and are a nominating organization for candidates for membership on the RRCs. All of the RRCs that review AOA programs for accreditation have osteopathic members. By all accounts, this transition has been very successful. AOA programs that applied for ACGME accreditation achieved that status in 98.6% of cases. Graduating DOs success in the NRMP match has increased dramatically, with 99.29% placed either in the Match or through the NRMP's Supplemental Offer and Acceptance Program (SOAP). DOs now serve on the ACGME Board and on 20 of its RRCs (Table 57.1–16).

The Next Accreditation System (NAS). Previously, the ACGME fulfilled its accreditation responsibilities through a system of regular review of programs by the RRC. Data were collected every 5 years for programs in good standing. The PD wrote a Program Information File (PIF), which covered every aspect of program operation in relation to the ACGME PRs. A site visitor would travel to the program and confirm the PIF with interviews with the PD, residents, and faculty. The RRC would then review the data at its meetings, issue an accreditation decision, and issue citations for areas judged out of compliance with the PRs.

The accreditation process has been drastically altered under the NAS. Implemented for seven pilot specialties (emergency medicine, internal medicine, neurologic surgery, pediatrics, diagnostic radiology, and urology) in July of 2013, this system is now being used for all 26 ACGME core specialties. The NAS changes what has been described as a "biopsy" model into a program of annual data collection and review for every training program in the United States. The data sources for this review include an annual resident survey, faculty survey, and data entered by the program into the Accreditation Data Systems (ADS). For many specialties, case logs and board certification pass rates are reviewed annually. It is anticipated that aggregate data from the milestones will be incorporated into the accreditation system after enough experience has been gained.

Table 57.1–14. Psychiatry Milestones

Milestone Subcompetencies	**Threads**
Patient Care 1: Psychiatric Evaluation	A: Gathers and organizes findings from the patient interview and mental status examination B: Gathers and organizes data from collateral sources C: Screens for risk and integrates risk assessment into the patient evaluation
Patient Care 2: Psychiatric Formulation and Differential Diagnosis	A: Organizes and summarizes findings and generates differential diagnosis B: Identifies contributing factors and contextual features and creates a formulation C: Uses the emotional responses of clinician and patient as diagnostic information
Patient Care 3: Treatment Planning and Management	A: Creates treatment plan B: Monitors and revises treatment when indicated C: Incorporates the use of community resources
Patient Care 4: Psychotherapy	A: Establishes therapeutic alliance and manages boundaries B: Selects, sets goals, and provides psychotherapies including supportive, psychodynamic, and cognitive-behavioral C: Manages therapeutic process
Patient Care 5: Somatic Therapies (including Psychopharmacology and Neurostimulation Therapies)	A: Understands the mechanisms of action, indications, and evidence base for somatic therapies and appropriately applies them to patient care

	B: Educates patients about somatic therapies including access to accurate psychoeducational resources C: Appropriately monitors patient’s response to treatment
Patient Care 6: Clinical Consultation	
Medical Knowledge 1: Development through the Life Cycle (including the Impact of Psychopathology on the Trajectory of Development and Development on the Expression of Psychopathology)	A: Knowledge of human development B: Knowledge of pathologic and environmental influences on development
Medical Knowledge 2: Psychopathology (includes Knowledge of Diagnostic Criteria, Epidemiology, Pathophysiology, Course of Illness, Comorbidities, and Differential Diagnosis of Psychiatric Disorders, including Substance Use Disorders and Presentation of Psychiatric Disorders across the Life Cycle and in Diverse Patient Populations)	A: Knowledge to identify and treat psychiatric conditions B: Knowledge at the interface of psychiatry and the rest of medicine
Medical Knowledge 3: Clinical Neuroscience (includes Knowledge of Neurology, Neuropsychiatry, Neurodiagnostic Testing, and Relevant Neuroscience and their Application in Clinical Settings)	A: Neurodiagnostic and neuropsychological testing B: Neuropsychiatric comorbidity C: Application of neuroscientific findings in psychiatry
Medical Knowledge 4: Psychotherapy	A: Fundamentals B: Practice and indications C: Evidence base
Systems-Based Practice 1: Patient	A: Analyzes patient safety events

Safety and Quality Improvement	B: Appropriately discloses patient safety events C: Participates in quality improvement
Systems-Based Practice 2: System Navigation for Patient-Centered Care	A: Coordinates patient care B: Safely transitions care C: Population and community health needs
Systems-Based Practice 3: Physician Role in Health Care Systems	A: Understanding and working within the health care system B: Health care financing and advocacy C: Transition to practice
Practice-Based Learning and Improvement 1: Evidence-Based and Informed Practice	
Practice-Based Learning and Improvement 2: Reflective Practice and Commitment to Personal Growth	
Professionalism 1: Professional Behavior and Ethical Principles	
Professionalism 2: Accountability/Conscientiousness	
Professionalism 3: Well-Being	
Interpersonal and Communication Skills 1: Patient- and Family-Centered Communication	
Interpersonal and Communication Skills 2: Interprofessional and Team Communication	
Interpersonal and Communication Skills 3: Communication within Health Care Systems	

Modified from Accreditation Council for Graduate Medical Education. The Psychiatry Milestones 2.0. Available at: https://www.acgme.org.

Table 57.1–15. Authors of the Subspecialty Milestones

Subspecialty Working Groups (Chris Thomas, Chair)	
Child and Adolescent	Jeffrey Hunt, Chair, Sandra Dejong, Laura Edgar, Howard Liu, Cynthia Santos
Geriatrics	Iqbal Ahmed, Chair, Josepha Cheong, Laura Edgar, Maria, LLorente, Susan Maxner, Sandra Swantek
Forensics	Andrea Stolar, Chair, Laura Edgar, Richard Frierson, Stephen Noffsinger, Charles Scott, Howard Zonana
Consultation-Liaison	Robert Boland, Chair, Madeleine Becker, Catherin Crone, Laura Edgar, James Levenson, Mark Servis
Addiction	Marian Fireman, Chair, Laura Edgar, Kyle Kampnam, Robert Ronis, Andrew Saxon, Jeffrey Wilson
Subspecialty Advisory Group (George Keepers, Chair)	Larry Faulkner, Frances Levin, Christopher Varley

Data from Accreditation Council for Graduate Medical Education. The Psychiatry Milestones.

The data from these sources are statistically analyzed by the ACGME, and outliers are identified for review on the basis of decision rules which are formulated by the RRCs. The data, program history, and previous ACGME actions are then reviewed by ACGME staff for presentation to the RRC. An RRC member conducts a review of the program and presents recommendations at the RRC meeting. The RRC may elect to continue the program's accreditation, continue accreditation with a warning, or ask for a site visit. Programs may be placed on probation following a site visit if problems with the program a very severe. The usual outcome of these reviews is to communicate to the program and its sponsoring institution that there are Areas for Improvement (AFIs) or to cite programs for noncompliance with the PPRs. AFIs are issues that may be dealt with by the program and its associated GMEC. Citations require, in addition, communication with the ACGME concerning the program's plan to remedy the problem. Programs with citations are automatically subject to annual review by the RRC.

Table 57.1–16.
Members of the Psychiatry RC

Suzanne Sampang, M.D., Chair	Adrienne Bentman, M.D., Vice Chair
Lilanthi Balasuriya, M.D., Resident Member	Deborah Simpson, Ph.D., Public Member
Joan Anzia, M.D.	Brigette Bailey, M.D.
Catherine Crone, M.D.	Peter Daniolos, M.D.
Richard Frierson, M.D.	Erick Hung, M.D.
Jeffrey Hunt, M.D.	Furhut Janssen, D.O.
Mark Kinzie, M.D.	Maria Lapid, M.D.
Robert McCarron, D.O.	Art Walaszek, M.D.
John Young, M.E.	
Larry Faulkner, M.D., Ex-Officio	Tristan Gorindo, M.D., Ex-Officio
Tanya Lopez, M.S., M.P.P., Ex-Officio	

Data from Accreditation Council for Graduate Medical Education. The Psychiatry Review Committee.

This system has the potential to identify and correct problems in training programs much earlier than the previous accreditation system. Additionally, its developers believe that it will encourage a more rapid pace of program improvement. Criticism of the new system has focused on the increased workload and expense required to comply with the new requirements and on the validity of the data, particularly the resident survey. PDs have expressed concern that residents are often confused by the questions on the resident survey and may answer the questions differently when properly oriented. Some studies do indicate that significant differences exist between the results from internal resident surveys and the ACGME resident survey. In 2019, the ACGME undertook the process of revising the resident and faculty surveys in response to these criticisms. A multidisciplinary task force and a commercial firm specializing in survey design completed this task. The revised surveys have been administered, and the results of the revision are being evaluated.

RELATION TO CHANGES IN MEDICAL EDUCATION

These changes in residency training are not occurring in isolation. The traditional system of medical education has been resistant to the incorporation of scientific developments in educational theory and performance improvement that have long been implemented in other educational and corporate settings. Future physicians can expect to be educated quite differently. Medical schools are incorporating the concepts of competency-based education and selecting students who will adapt better to the modern practices of team-based medicine. The requirement that students demonstrate competence at each stage of education is becoming universal. This will require a more flexible approach to the medical school curriculum than currently exists. For example, students matriculate to medical school at varying degrees of readiness. Flexible preparatory curricula are needed to meet varying student needs for additional education prior to their first day of medical school. Ideally, the goal is for every student and every resident to be uniformly excellent—to deserve an A grade. The purpose of competency-based undergraduate and graduate medical education is to raise the level of performance and reduce the variability in educational outcomes.

Following residency education and board certification, it is becoming mandatory that physicians maintain and improve their competence, incorporate new knowledge, and develop new skills throughout their careers. The ABMS and specialty board MOC programs are oriented to this goal. These developments in undergraduate, graduate, and postgraduate medical education are a veritable revolution in the way doctors will be educated. Medical education’s past has been one of discrete episodes of learning with little assurance that competence has been attained or will be maintained. Its future is as a lifelong learning process with feedback mechanisms ensuring that physicians practice at a uniformly excellent level.

FURTHER READINGS

Adamowski S. The ACGME and ABMS initiatives: toward the development of core competencies. In: Scheiber S, Kramer T, Adamowski S, eds. *Core Competencies for Psychiatric Practice: What Clinicians Need to Know*. American Psychiatric Publishers; 2003:43–54.

Adams MA, Willett LL, Wahi-Gurunaj W. Usefulness of the ACGME resident survey: a view from internal medicine program directors. *Am J Med*. 2014;127(4):351–355.

Arrighi JA, Keepers GA. Great expectations: Should we anticipate that changes in work hour standards will impact health care outcomes? *J Grad Med Educ*. 2019;11(2):156–158.

Balon R. The unspoken tyranny of regulatory agencies: a commentary on the ACGME Resident Survey. *Acad Psychiatry*. 2012;36:351–352.

Benjamin S, Widge A, Shaw K. Neuropsychiatry and neuroscience milestones for general psychiatry trainees. *Acad Psychiatry*. 2014;38:275–282.

Burchiel KJ, Zetterman RK, Ludmerer KM, et al. The 2017 ACGME common work hour standards: promoting physician learning and professional development in a safe, humane environment. *J Grad Med Educ*. 2017;9(6):692–696.

Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296:1094–1102.

Edgar L, Roberts S, Holmboe E. Milestones 2.0: A step forward. *J Grad Med Educ*. 2018;10(3):367–369.

Eid SM, Ponor L, Reed DA, Beydoun MA, Beydoun HA, Wright S. Associations between in-hospital mortality, health care utilization, and inpatient costs with the 2011 resident duty hour revision. *J Grad Med Educ*. 2019;11(2):146–155.

Ey S, Moffit M, Kinzie JM, Brunett PH. Feasibility of a comprehensive wellness and suicide prevention program: A decade of caring for physicians in training and practice. *J Grad Med Educ*. 2016;8(5):747–753.

Fahy BN, Todd SR, Paukert JL, Johnson ML, Bass BL. How accurate is the accreditation council for graduate medical education (ACGME) resident survey? Comparison between ACGME and in-house GME survey. *J Surg Educ*. 2010;67:387–392.

Feldman AM, Ludmerer KM. Solving one of graduate medical education's problems: A loss of thoroughness. *J Grad Med Educ*. 2017;9(3):287–289.

Goitein L, Ludmerer KM. Resident workload—let's treat the disease, not just the symptom: comment on "Effect of the 2011 vs 2003 duty hour regulation–compliant models on sleep duration, trainee education, and continuity of patient care among internal medicine house staff." *JAMA Intern Med*. 2013;173(8):655–656.

Hunt JR, Thomas CR. ACGME milestone development in general psychiatry: patient care and medical knowledge. *Acad Psychiatry*. 2014;38:261–267.

Kassebaum DG. Origin of the LCME, the AAMC-AMA partnership for accreditation. *Acad Med*. 1992;67:85–87.

Kinzie JM, DeJong SM, Edgar L, et al. Psychiatry milestones 2.0: Using the supplemental guide to create a shared model of the development of professional identity and expertise. *Acad Psychiatry*. 2021;45(4):500–505.

Macaluso M, Kinzie M, Cowley D, et al. The psychiatry milestones 2.0: How did we get from 1.0 to 2.0 and what can users expect? *Acad Psychiatry*. 2020;44(6):779–784.

Mantosh D, Manring J, Satish U. The new milestones: do we need to take a step back to go a mile forward? *Acad Psychiatry*. 2015;39(2):147–150.

Nasca TJ, Philbert I, Brigham T, Flynn TC. The next GME accreditation system-rationale and benefits. *N Engl J Med*. 2012;366(11):1051–1056.

National Resident Matching Program: Match results by state, specialty and applicant type. Accessed June 2021. https://www.nrmp.org/wp-content/uploads/2023/05/2023-Main-Match-Results-and-Data-Book-FINAL.pdf.

Norman G, Norcini J, Bordage G. Competency-based education: milestones or millstones? *J Grad Med Ed*. 2014;6(1):1–6.

Rajaram R, Saadat L, Chung J, et al. Impact of the 2011 ACGME resident duty hour reform on hospital patient experience and processes-of-care. *BMJ Quality & Safety*. 2016;25:962–970.

Reil JC. *Rhapsodieen über die anwendung der psychischen curmethode auf geisteszerrüttungen*. Curtschen Buchhandlung; 1803:19.

Rosen DE. The development of outpatient milestones. *Acad Psychiatry*. 2014;38:283–287.

Sanders K, Servis M, Boland R. The four general competencies. *Acad Psychiatry*. 2014;38:268–274.

Shorter E. *A History of Psychiatry: From the Era of the Asylum to the Age of Prozac*. John Wiley and Sons; 1997:17.

Swing SR, Beeson MS, Carraccio C, Coburn M, Iobst W, Selden NR, et al. Educational milestone development in the first 7 specialties to enter the next accreditation system. *J Grad Med Educ*. 2013;5:98–106.

Swing SR, Cowley DS, Bentman A. Assessing resident performance on the psychiatry milestones. *Acad Psychiatry*. 2014;38:294–302.

Thomas CR, Keepers GA. The milestones for general psychiatry residency training. *Acad Psychiatry*. 2014;38:255–260.

57.2 Examining Professional Patients

M. SAGMAN KAYATEKIN, M.D., GEORGE BOMBEL, PH.D., AND KIM-LAN TRAN CZELUSTA, M.D.

In this chapter, we will talk about the development of a professional during his adulthood. More specifically, this is an attempt to think about human development, with a focus on professional development and how to think about the pauses and regressions in this ever-evolving process—if/when it may go off the tracks during later phases in life due to phase-specific challenges or personal problems like mental illness, substance abuse.

To illustrate the salient points in the professional maturation and growth chart, we will take a psychiatry resident during his residency training and then shift to a professional in their mature adult years, facing various roadblocks. The first section is specific to psychiatric residents. However, much of the discussion would generalize to other professions. Then, in the second section, we broaden our archetypical patient to include other professions.

Despite unique professional development trajectories, we believe that all professionals share expected psychological growth and cognitive development aspects. One crucial common denominator is that professional growth and development occurs in an interpersonal setting. This setting includes not only didactic learning but some form of apprenticeship. This approach is the solid foundation of any profession where there is a need for a supervised training period to develop competent professionals. This is more recognizable for those trained in psychiatry than in sister clinical fields such as medicine, psychology, social work, and nursing. However, it applies to many other fields of human education and professional skills; postgraduate studies in various areas, progression in military ranks or any organization, arts, becoming an artisan, craftsman, or a farmer.

In sum, humans learn and internalize their professional identities as they work, while being embedded in a matrix of a maturationally necessary asymmetric relation with others where information is transmitted through explicit (didactic education, bedside teaching) and implicit ways (identification with teachers as they are performing the tasks they are teaching the student).

PSYCHIATRIC RESIDENCY AS A MODEL OF PROFESSIONAL LEARNING, GROWTH, AND MATURATION

Starting Residency

Residency training is often the first employee situation for residents since many go directly from high school to college and medical school without a regular and sustained job. Unlike the occasional summer jobs, they may have held, working as a doctor, albeit in training, creates a highly complex nexus of relations, set of rules, regulations, and expectations that the young resident must learn, adapt to, and internalize. Thus, from the end of social role expectancies, these dynamics will reveal many capabilities and fragilities that were relatively dormant due to the person's straightforward, significantly simple interpersonal, and educational life during the "student" years.

It is an interesting dilemma that residents are professionally hybrid persons; although they are employees, they are also learners, which creates a complex web of supervision, milestones, and didactic education. Ultimately, the residency director is responsible for ensuring each resident will competently practice independently and with professionalism. Management of resident difficulties is one of the most time-consuming yet essential responsibilities for a residency director. In addition, the residency director balances complex and often conflicting issues related to fairness to the resident, the program, and the public. As the administrative requirements for residency training continue to increase, having a transparent process in place is essential to support both the residency director and the resident and promote the best outcome possible for all involved.

It is not uncommon for many trainees to have difficulty during these formative years of their professional development during residency. Some difficulties are straightforward and quickly addressed, while others can be extremely challenging, time-consuming, and stressful to all involved. The resident's status as both an employee and a trainee simultaneously guides and complicates addressing problems.

The residency program is poised to make training accommodations to support struggling trainees toward completing residency training. Common modifications include changing rotations or call schedules, increasing supervision, decreasing workload, or repeating a clinical experience. In addition, when approaching a resident's difficulty, the residency director is continually balancing three essential tenets: (1) fairness to the resident, (2) safety to the public, and (3) adherence to program policies to help each resident complete residency training.

Early Detection of Resident Difficulties

Data sources about resident performance include attending supervisors and chief residents, site-based directors, chiefs of service, on-call faculty, resident peers, administrative staff, patients, nurses, and other hospital staff. Documentation of the concern and subsequent intervention is essential as the concern may be the beginning of a series of later difficulties. In some cases, presenting concerns to the resident and supporting their efforts to reflect on their role in a conflict is a sufficient intervention. This will depend on the residents' openness and responsiveness to feedback.

A resident concern can have multiple contributing factors, and the residency director may not be able to address all etiologies at play. For instance, if the resident concern is a result of family or personal stressors (divorce, sick family member, family conflict), the residency director can offer emotional support in addition to referring to the institution's Psychiatric Counseling Service or Employee Assistance Program. The residency director could also consider temporarily reducing workload or work hours to allow the resident more time to address their situation. However, determining how long such a modified, reduced schedule is reasonable while ensuring the resident meets standard residency training and graduation expectations is not always easy to decipher.

If the resident's difficulty results from problems within the clinical unit or team members, the residency director can help elucidate and address the external factors resulting in residency difficulty. For instance, if the problem stems from a specific staff member of the clinical team with known concerns, then the residency director can provide feedback to the staff member's supervisor, who can then help intervene with the problematic staff member. Interpersonal difficulties with team members sometimes result from the confluence of different interpersonal styles that collide. In such cases, enlisting the help of the responsible attending or chief of the service may be warranted.

Most residents have a predictable developmental course during residency training, but some struggle more than others during the transition from medical student to intern. The increased patient care responsibilities combined with administrative tasks, new electronic medical records systems, teaching expectations, and more advanced specialty-specific clinical skills can be overwhelming and result in difficulties, including time management issues. Moving between large urban environments and small towns during medical school to residency training can also be jarring for some trainees. Lastly, leaving a familiar environment with established sources of support to move to a new surrounding with limited early social supports can be isolating.

Fit with the specific residency program may be a factor contributing to a difficult adjustment for residents. Fit encompasses many components and can include geographic location and program size, number of clinical sites and faculty, program culture, patient volume, typical work hours, and call type and frequency. Fit with specialty may be another factor contributing to resident difficulties as it is not uncommon for residents to change specialties in the middle of or after completing training in another specialty.

Finally, a resident’s psychopathology or medical illness can play a significant role in resident difficulties. Referral to the institution’s Psychiatric Counseling Service, Employee Assistance Program, or community provider offers treatment for the resident and provides information to assist the residency director in determining how to best support the resident. A medical leave of absence may be necessary to allow residents time to address their own health needs. The resident may not successfully meet the demands of residency training and successfully meet graduation requirements depending on the severity of the resident’s condition. Residents who have limited insight into their condition or are noncompliant with treatment recommendations can be especially vulnerable to difficulties.

When problems were noted, Dr. A was a PGY2 psychiatry resident who had excelled academically in her premed studies. She graduated Phi Beta Kappa from college, attended a competitive medical school where she was elected to the Alpha Omega Alpha Honor Medical Society, and then in the first few months, continued to perform at a clinical level that was higher than expected for residents at her level of training. However, her feedback became somewhat uneven in the Clinical Competency Committee meetings over the first 2 years of her residency training. The feedback ranged from "quick thinker, organized, focused, responsible, extreme fund of knowledge" to "challenging, immature, abrasive, poisonous attitude, disrespectful to team. doesn't see how her actions/words affect others."

This inconsistent feedback was difficult to convey, and much of it was informal, being relayed verbally to the program director. However, the resident's written evaluations were average or above average. When her residency director finally provided this feedback, Dr. An attributed the negative comments to mostly jealous and competitive peers and pointed out supervisors' above-average evaluations. During outpatient rotations of PGY3, Dr. A's work performance deteriorated further; she was frequently late to clinics and routinely late to close out electronic medical records. Since Dr. A worked at various outpatient clinic settings, the pattern of tardiness was not captured until the end of the PGY3 year.

Despite these various obstacles, the PGY4 year, which mainly was research electives, went relatively smoothly. Dr. A managed to graduate. She was not the stellar student her residency had hoped for but a young clinician with strengths and yet a fragility to develop split relationships. These relationships ranged from positive to negative, depending on the interpersonal context and task; she fared better in environments with a clear frame and comparatively limited close patient or staff contact. Patients' capacity to activate regressive individual tendencies and evoke split group dynamics with teams was overwhelming for Dr. A. After a process of beneficial psychotherapy and numerous consultations with her director and peers, following graduation, she chose to work in pharmacology research in a major corporation. It was a much less complicated interpersonal context where she could adapt with her introverted sides, and she fared well in this subbranch of her field.

PROFESSIONAL CRISES IN LATER YEARS

Clinical Phenomenology of the Loss of Sense of Self

We will not talk about specific pathologies and their treatment in this chapter—these are described in other chapters. Instead, we will focus on aspects of the clinical presentation and the treatment of "Professionals" that differentiate them from other social groups.

A standard, almost universal theme in the patients we see is the "loss of a sense of self." A lack of sense of self commonly accompanies a major depression or other major Axis I pathology. Nevertheless, some of our professional patients come without any severe pathology but just with a deep, painful sense of loss of self-definition, mainly in the context of a developmental-stage–related challenge.

Almost all of these patients we see have identified very deeply with their professional roles; their professional identity is "the" defining, core aspect of their self-images. Furthermore, this psychically vital element collapses due to exhaustion, common developmental challenges such as promotion or retirement, or major psychopathology.

The roots of this central and psychologically vital professional identification are somewhat complicated. Nevertheless, we describe some of the most common ones in the associated case descriptions.

Identification With Predecessors

A patient was in his mid-50s and was the heir apparent to a very successful business developed by her father and uncle. She was groomed very early on for this prestigious role, and the patient prepared herself well by attending Ivy League schools, doing a graduate degree in the field, and spending considerable time working on different parts of the organization during any off times from school. However, once she was actually in the leadership role, she had difficulties filling her predecessors' shoes and was constantly being critical of herself and being criticized by the family, hoping she would continue the family heritage. This ongoing discrepancy between the expectation and inability to fulfill it was one central root of her deterioration and hospitalization.

Disidentification

A 40s entrepreneur was from a humble background. His family had lower-class origins where the parents were rather lackadaisical about work, were into using substances, and did not care too much about raising family standards. The patient strongly disidentified with these aspects of his parents and developed a multimillion-dollar company over the years with hard work and luck. However, his only brother had identified with the parents, dropped out of college, and lived a marginal, parasitic life, lost in the world of drugs and multiple unsuccessful short-lived marriages. He committed suicide in his early 40s and following this unexpected horrific loss, around mid-career, the patient crumbled and became a person with no motivation for work, very similar to his brother, before he came to treatment.

Traumatic Abandonment

A successful doctor in his mid-40 s was hospitalized because of increasing exhaustion over excessive work. He described a sudden paternal loss at the age of early teens/puberty, following which his mother developed severe depression and the patient had to “grow” on his own and very fast; he worked two jobs to make ends meet, took care of his mother and continued his education. It was of note that he could remember little of his relationship with his father, and there was a strong undercurrent of chronic depression/grief that was helped enormously by repressing the painful grief defensively/adaptively in a hypomanic/obsessive way at work. He had accumulated enough wealth to retire by his mid-40s but then crumbled into a depression with a deep sense of “who am I.” He had lost a central, defining element of self that led to a 6-month hospitalization and then ongoing psychoanalytic treatment posthospitalization.

Central Defensive/Adaptive Styles of the Archetypal Professional Patient: the Hypomanic-Obsessive Style

A patient had a background of personal and transgenerational trauma due to immigration from a war zone, parental suicide, and neglect. He described an important adaptive aspect of excessive work quite poignantly; “I wake up with a deep sense of sadness and grief. Then in a matter of an hour, after taking a shower and having my breakfast, I immerse myself in my work, and all these painful feelings disappear. I work excessively and methodically till I exhaust myself and usually others around me. I can maintain my pace on 4 to 5 hours of sleep every day. This cycle repeats itself every morning, and when I can’t work, it becomes intolerably painful.”

Besides the centrality of defining oneself as a professional, we believe in a universal characterologic defensive/adaptive aspect of successful professional patients seen in specialty mental health programs geared toward professionals. This pattern includes the following typical constellation: prominent premorbid hypomanic character traits usually accompanied by obsessive-compulsive personality organizational aspects and a pattern of counter-dependence. For the sake of clinical shorthand, we name it as hypomanic-obsessive character organization. The collapse of this structure with defensive and adaptive elements leads to regression and the need for treatment.

The abundance of hypomanic-obsessive presentations in the patient population of our particular Professionals Program provides insights into this relatively neglected style of personality organization. It is as if these hypomanic traits, which allow a broad range of affect to be displayed in a diffusive way, are contained/constrained by obsessive substructures that are the other prominent aspect of personality organization. These patients can work endless days and hours—it is not unusual that a patient takes off only a few days over many years. As one of our patients suggested, "There is something very attractive about working and being successful at it. It is as if work has a way of expanding by itself, intruding into the person's life once the fulfillments are there". Moreover, these two elements of their personality structure fit well with their extreme independent/counter-dependent style of relating to others and problem-solving. One patient described this independent cognitive style as having "an internal debate" in decision-making instead of consulting with others.

The literature is abundant for mood disorders, yet hypomanic/hyperthymic presentations are somewhat in the margins of this literature. In addition, there is a debate on whether it is a bipolar spectrum disorder or a temperament. There is evidence that it has genetic connections to creativity. We do know that any illness can be used for defensive and adaptive purposes. Beyond its impact on the human body or psyche, in a person struggling with a chronic or repetitive illness, the person in consideration's body and mind adapt to this infringement and uses it in an adaptive/defensive manner. Here though, we see some temperamental/learned traits used for defense and adaptation and becoming the central defining element of the patient's personality.

In common parlance, this is what is coined rather loosely as the concept of "workaholism." What the typical jargon, of course, misses is the fact that this is not an aspect of the character structure that can be extirpated for treatment, but it is an essential capability of the human mind that can be used more adaptively if it can be brought under the proverbial "domain of the ego." In an interpersonal level, when operating properly, hypomanic/obsessive style provides the person with extremes of independence, a capacity to make significant decisions on their own without the need for group-think that is more common in successful professionals that lead teams of various sizes—from a clinical team to a large organization.

In their detailed life histories, we notice that there are some standard developmental patterns in these patients. Bypassing multiple adolescent developmental tasks is a common element. There is a distorted, stunted, bypassed adolescent growth; a typical pattern is a very early love relationship; they become devoted lovers/friends and join their lives with the person they met in high school or early college years. Thus, they lack a core group of friends; their spouse usually is their “best friend.”

Family relationships, time with friends, and other “fun” activities are sorely missing for these patients. Everything they do is tied to an ultimate aim of developing a perfect result—as a professional or parent. They are pervasively result-focused, and, in their worldview, “fun” activities with no purpose are an enigma for them. Paradoxically, they display social anxieties and almost a tendency to schizoid isolation when there is no central work-related task. These core mild schizoid-social phobic aspects are somewhat dormant side by side with impressive capacities to work and be very social within workgroups and lead them to success.

They usually treat their nonwork-related aspects of life with the same principles they successfully utilize in work. Thus, having a final product of a perfect marriage and shining kids becomes a powerful motive. These interpersonal patterns are “counter dependent” in that the patient relates to persons as if they were objects. This approach is very similar to their successful cognitive work modus operandi—they treat job-related issues in meticulous detail with an impressive capacity to make decisions almost singlehandedly—in a fiercely independent intellectual functioning. Similarly, they treat relationships in a nonpersonal distant way and try to fix, find algorithmic solutions, and may even be extremely helpful with their close relationships’ decision-making.

Finally, in the background developmental history, quite regularly, there is a history of attachment trauma, parental loss, the transmission of trauma across generations, and unresolved grief, which usually begins around the preadolescent teens.

Clinical Phenomenology of the Decompensation of the Hypomanic-Obsessive Defensive Style

Once regression sets in, three aspects of the presentation stand out: The mood becomes somber, depressed with an agonizing absence of “who they are." The obsessive sides shift to a deep incapacitating ambivalence, ambitendence, and ruminations, losing all its impressive analyzing/problem-solving qualities. Finally, the patient develops the presentation of deep sadness, apathy, preoccupation with guilt-ridden ruminations, and extremes of dependence which is the polar opposite of their usual functioning self.

The hypomanic/obsessive style deteriorates to an obsessive/ruminative style and may even lead to precatatonic, and in rare instances, catatonic states.

One of our patients, a 40-year-old businessman, described the progression into catatonia wonderfully. The first sign was a sense of increased anxiety—contextual, concerning conflict within a love relationship, or work-related conflict. Then he recognized a slowing and perseveration of his thought processes and speed. This slowing led to difficulty in building sentences and a general, painful slowing of his psychomotor functioning. Moreover, the crux of the crisis was when he became entirely immobile with thoughts, themes circling in his mind, unable to decide to move around or stay still. He could stay like that for a day. He would be exhausted from keeping a rigid posture and not eating or drinking. As a result, he was hospitalized a few times. Occasionally he would be able to break this agonizing immobile state and pull himself out of the freezing. This syndrome was clearly articulated by earlier clinicians like Kahlbaum and resembled an advanced form of obsessive indecisiveness, ambivalence, and ambitendence.

An example of a milder case is a 70-year-old patient who had fully retired after a successful career. He was observed "stuck" in the doorway to his room on our professional treatment unit. He attempted to go in, took a few steps toward the bedroom, and then walked out of the bedroom a few steps toward the hallway. When we inquired about this, he said he was torn between wanting to go toward and away from his room. He thus repeatedly walked in and out endlessly.

On a more profound level, the following case exemplifies the tormenting state of ambivalence. The patient, a gentleman in his 60s with a very typical hypomanic/obsessive personality style, was tremendously successful in his line of work. Following retirement, he became depressed. This depression was characterized by the loss of self-definition around his identity as a professional. He decided to come to the hospital after some years of toying with the idea, and it took him about another year-plus to make up his mind after making a final decision to come to the campus. However, it was almost impossible for him to initiate the admissions process once in the building. He could not get in through the doorway of the admissions building. He was stuck. He was begging to be admitted, and a second later, he was telling us that he could not be admitted because he was too sick and needed to be admitted to a locked facility. He then begged that we send him to such a unit as he began lying down and writhing, rolling around on the floor in agonizing anxiety. Finally, after the patient spent a few days in involuntary treatment (that we had initiated), he returned and managed to bring himself to the unit despite a paralyzing sense of indecisiveness and ambivalence.

Once in the unit, he displayed similar symptoms on more subdued and subtler intrapsychic and interpersonal levels. He managed to keep himself in the program but wore down the staff’s patience by constantly asking whether he would be transferred again to a locked unit. His indecisiveness was at a level where he would seek guidance on ordinary daily chores, such as on which medications to choose or which foods to eat. Once he got the recommendation on the medication or the food, he would immediately start thinking about the ill effects of the pill or the food he had chosen, become regretful, and try to change his mind again. No ordinary logic counterargument was of any help, and he was aware of this, and this self-awareness made him more ashamed of his behavior. With intense treatment, he slowly cleared from this depressive/obsessive/indecisive pit and maintained a much higher level of functioning in the unit.

For many patients, the painful ruminations flooding their minds when regressing were rooted in beneficial aspects of their cognitive organization when focused and controlled. Premorbid, these patients usually do not have an obsessive-compulsive disorder but rather an obsessive-compulsive personality structure. Their meticulousness with details and ability to weigh options were such that they were admired for their analytical skills and their precise decision-making capabilities. However, this solution-oriented cognitive functioning began displaying a “broken record” quality when decompensated, and there was no processing but futile repetition of similar themes. So, the crumbling of the hypomanic-obsessive defensive/adaptive structure and the accompanying loss of sense of a central definition of self would add to the patient’s problems and worsen the patient’s depression.

TREATMENT: UNFOLDING INTO THE SYSTEM, REPEATING THE STORIES OF THEIR LIVES

This chapter will focus on some unique aspects of the treatment specific to this clinical population. However, standard treatments, such as psychopharmacologic agents, should be used when appropriate. Also, sleep problems are pervasive and can be treated with cognitive-behavioral therapy for insomnia (CBTi) or appropriate medications. For older professionals, cognitive evaluations are essential. In our experience, it is rare not to treat patients with some standard psychiatric treatment.

Our repeated experience is that psychotherapeutic approaches and biologic treatments work on similar neural substrates and, if appropriately combined, tend to potentiate one another. In the last few decades, the worldwide increase in moving away from monolithic and exclusionary/antagonistic “biologic” versus “psychological” theories toward more integrative theories of mind and mental illness is a witness to this fundamental, paradigmatic shift.

Some Central Therapeutic Elements of This Unique Setting

Repetition is one primary key—one way to evaluate these patients’ intrapsychic/interpersonal scripts is how they unfold into professionals’ programs and repeat their central issues: conflicts, deficits, interpersonal modii of operandi. A central dynamic is the externalization of internal representational scripts in the form of; seeing the self in staff and peers, seeing the object in staff and peers, group dynamics that unfold with peers, and transferences to hospital. Enactments and clearing from enactments are central to treatment.

Due to the therapeutically regressive elements (limited time, relative confinement to a physical space with other peers, intense group dynamics, regressive pulls of therapies), there is rapid individual/group regression, and the patient presents intrapsychic and interpersonal functioning with relative transparency in the milieu.

The psychotherapeutic hospital is an environment where the patient constantly tests functional beliefs, distortions, and expectations. As a result, various transference, countertransference, and enactment scenarios develop and are lived through and interpreted.

Patients report that it is much easier to open up and talk about their experiences honestly in the hospital setting because they feel a sense of belonging when they hear the stories of fellow patients—the pathologically grandiose fantasy that they are alone, the only one with this kind of embarrassing mental health problem dissipates very rapidly. It becomes a more ordinary and shared experience.

As well-established research demonstrates, psychotherapeutic hospitals, which offer more extended immersion into various psychotherapy and biologic therapies, are conduits of impressive and promising change. However, in essence, we probably still do not have a refined, satisfactory theory of how these unique human ecologies work.

As a demonstration of an approach to treat professional patients, the remainder of this chapter will describe a representative case in detail.

Treatment Example: Marvin

Marvin was a divorced, 60-year-old and CFO of a technology company with a vague history of anxiety, dysphoria, perfectionism, and aloofness. From his perspective, he was primarily content and healthy aside from intermittent “blue spells” over the years. These spells were quickly ameliorated with medication changes; however, 6 months before the current series of inpatient psychiatric treatments, he decided to “preretire” to escape a painful work situation and spend more time with his wife, two adult children, and several grandchildren. This retirement involved stepping down from CFO to a part-time consultant to the company’s board, a move that immediately left him feeling empty, directionless, worthless, and psychologically collapsed. In turn, he became overwhelmed with regret and severe rumination about the decision to preretire. Dread, anxiety, and dysphoria overtook him, and he became suicidal—eventually culminating in suicide attempts. After the most recent (benzodiazepine overdose), he was admitted to a crisis-stabilization unit in his city, then transferred to a partial hospital program.

PSYCHOLOGICAL ASSESSMENT. As a part of a thorough evaluation, Marvin’s treatment team requested a psychological assessment to clarify stressors, defenses, interpersonal relations, thought processes, coping capacities, and diagnosis, and to receive treatment recommendations. Marvin participated in a clinical interview, then completed the Millon Clinical Multiaxial Inventory—4th Edition (MCMI-IV), the Thematic Apperception Test (TAT), and the Rorschach Inkblot Test, using the Rorschach Performance Assessment System (R-PAS) administration and scoring guidelines. During the assessment sessions, Marvin presented with a fragile but determined grin that was insufficient to mask his underlying dysphoria. His thought process was ruminative, focusing on the loss of his CFO role, and recent marital conflicts. Nevertheless, he remained tense, polite, cooperative, and engaged throughout the process, though mildly condescending at times and doubtful that the assessment would provide any new insights about how his mind works.

The assessment data suggested that, behind his defensive functioning, Marvin felt fundamentally inadequate, broadly ineffective, emotionally weak, and flawed in significant ways. For example, despite his successes, he described feeling socially inept and awkward in most life spheres (with friends, family, colleagues). Concomitantly, he tended to suspect that others are ever-poised to hurt, criticize, reject, shame, humiliate, or deceive him—so he was pretty suspicious, though he did not appear to reach the threshold for clinical paranoia. These perceptions of self and others left him shy, apprehensive, and racked with self-doubt. Marvin's primary strategy for managing this pain, uncertainty, and fear was an avoidance pervasive and robust enough to resemble schizoid withdrawal. Thus, on the surface, he appeared distant, aloof, emotionally detached, or even indifferent to others—and although willing to engage in some obligatory family and business engagements, he appeared to prefer solitary activities. However, this presentation was more reflective of assiduous efforts to deny longings for closeness and approval to neutralize the internal conflict, for example, if he had no desire, he did not need attachments and therefore could not truly be hurt. Broadly, in concert with natural human longings for social contact and acceptance, Marvin's insecurities oriented him toward others (e.g., for soothing, comforting, approval, mirroring), while the suspiciousness motivated him to avoid or withdraw, behaviorally and emotionally. Therefore, he was highly conflicted about closeness, or even about establishing superficial relationships, unless approval and deep appreciation (and safety) could somehow be guaranteed.

Upon this defensive stratum rested another that addressed Marvin's painful self-concept directly, fueled in part by internalized family (of origin) expectations that he succeeds tremendously, attends the best schools and receive academic honors, become a CEO, and be the family's source of pride and healing; that is, he felt compelled to succeed to save the family from its shame or humiliation based on their immigration status (originally from Eastern Europe)—or else he risked drowning in his shame and humiliation (and theirs), as a disappointment and a failure. This painful internal terrain prompted him to adopt an idealized self-image as an elite super-achiever: An enterprising, enlightened, and inspired businessman who worked tirelessly, motivated colleagues and mentees, and accomplished what others could not. Other facets of this self-image were more compulsive; for instance, he worked to see himself—and be seen by others—as a serious-minded, disciplined, efficient, and industrious individual who was reliable and committed to work, productivity, and the bottom line. Accordingly, he was loath to appear frivolous or rash, terrified of making mistakes, or perceived as exercising poor judgment. Self-control was paramount; he presented as “buttoned up,” self-sufficient, and entirely in control of his needs, impulses, emotions, and behavior. Any acknowledgment or display of emotional pain or vulnerability would be too exposing, dangerous, shameful, and unbearable since, for him, it would be tantamount to broadcasting the inadequacy, weakness, and defectiveness he secretly felt but worked so intensely to escape or deny.

Given the stakes—instability of self-esteem/worth and the potential for excruciating exposure and shame—Marvin was compelled to extend his domain of control to include others. As suggested, this included managing his public and even private personas so that others only perceived what he wanted them to perceive (a superior, industrious, controlled super-achiever). In this way, he could have his greatness reflected, an unconscious psychological maneuver that effectively dispatched his most profound painful self-concept. Ideally, essential others (e.g., family members, mentees) would shine as well, as good reflections of him, but they were devalued or pushed away if they disappointed him or began to approach or eclipse his status. Marvin generally managed others by setting high expectations and being pushy and demanding. However, when his status or authority was challenged, control over others was somehow lost, or he felt disrespected, he was capable of direct or indirect retaliatory verbal aggression. These outbursts served to discharge his rage but also to push others back into line.

However, these measures intermittently failed during the “blue spells” or when stressors severely taxed Marvin’s limited coping resources. For example, when he had few mentees at work and felt less idealized. Alternatively, when a new junior manager publicly questioned his decisions, which challenged his sense of superiority and made him feel criticized, humiliated, and quietly furious. Similarly, when his adult children complained that he showed little interest in spending time with his grandchildren, he was left feeling criticized, ashamed, and resentful. These factors combined to motivate the “preretirement” to escape the painful work environment and have more opportunities to demonstrate that he was the “super grandpa” associated with his persona of exceptionality. However, he was unable to predict how the part-time consultant position would ultimately fail to “fill [his] sails,” provide worth and direction, neutralize his emptiness, and buttress his fragile sense of self (as inadequate, weak, defective, socially inept). He subsequently collapsed and was overrun by waves of regret and uncontrollable rumination about having made “the worst decision of my life!” Unfortunately, his defense of aloofness, superiority, self-sufficiency, and invulnerability made him a closed system, that is, he desperately longed for soothing, comfort, and relief from his pain within loving relationships wherein he is known but was unable to drop his armor. Ultimately, this has left him feeling dreadfully alone, misunderstood, resentful, and disillusioned with life, leading to his suicidal ideation and attempts.

Marvin’s assessment data and presenting symptomatology suggested the following diagnoses for consideration (from Diagnostic and Statistical Manual of Mental Disorders—5th Edition [DSM-5]): Major depressive disorder, recurrent; and other specified personality disorder with narcissistic and obsessive-compulsive personality traits.

Based on the assessment data, Marvin received treatment recommendations for short and long-term psychotherapy work. The team suggested that Marvin would benefit from time-limited cognitive-behavioral approaches (CBT) in the short term. These approaches could focus on modifying his painful intermediate and core beliefs (e.g., I am inadequate, weak, and defective; letting others see my vulnerabilities would be disastrous; anything less than being a super-achiever is an utter failure; I know best, so people should do things my way). These approaches could also help him retrain his attentional processes to register more accurate data that would allow him to develop reality-based, realistic, compassionate concepts of self and others.

In the longer-term psychotherapy work to initiate structural change (2 to 3× week psychodynamic psychotherapy), it was suggested that Marvin address his relationship templates and attachment style more immediately within the psychotherapy relationship itself—since they would inevitably surface there. After a basic alliance was established, the therapy focus could become more expressive, during which the therapist could explore Marvin’s feelings and relationship patterns more deeply. Given the nature of Marvin’s developmental arrest, it would be necessary for the therapist to communicate empathy about Marvin’s need for mirroring and to help him understand the links between (a) early deprivations, (b) his current hunger and emotional pain, and (c) the efforts to elicit signals from others that he is great, superior, admirable—including how he utilizes control to achieve this objective. It would also be imperative for the therapist to register and acknowledge their intermittent failures at providing this empathy. This cyclical therapy process—empathizing/mirroring, then recognizing and processing Marvin's painful feelings (e.g., grief, shame, humiliation, hopelessness, rage) when mirroring failures occur—would allow, over time, his current insecure sense of self to gain traction and resume its developmental trajectory.

PSYCHOTHERAPY. Marvin's individual psychotherapy consisted of a series of 15 sessions with a psychodynamic therapist, occurring three times per week throughout treatment. His original goals were to learn techniques to "manage the anxiety and depression and get back to have a contented life with my family"—a typical set of goals for the large subset of professional patients who present with a blend of compulsivity or perfectionism and narcissism. These individuals tend to fantasize that their struggles are mechanical problems that can be easily loaded into an algorithm that provides a quick solution. Not surprisingly, therefore, Marvin grumbled during the developmental history because he could not understand how his childhood experiences—including early, crucial attachment relationships—had any bearing on the current struggles. Similarly, he minimized the importance of addressing his recent suicidal ideation and suicide attempts because he no longer felt suicidal and was confident he would control any ideation and impulses that might occur again in the future. Consequently, he declined offers for a Collaborative Assessment and Management of Suicidality (CAMS) session.

He begrudgingly tolerated these deeper inquiries by the therapist but occasionally protested the lack of focus on immediate symptoms. A few times, he went so far as to suggest that the therapist was either incompetent or withholding effective (quick) treatments that were otherwise available. Here, Marvin responded moderately well to efforts by the therapist to highlight and challenge (a) his underlying fantasy that his mind is merely a contraption to be hacked and solved, and (b) his distorted perception that the therapist would be interested or motivated to withhold better treatments than were being provided. Marvin preferred to repeatedly describe his immediate psychiatric symptoms, such as anergia, anhedonia, low motivation, intermittent insomnia due to rumination, hopelessness, self-criticism, guilt, and regret, difficulty concentrating, and anxiety and tension secondary to catastrophic thinking. However, the most painful symptoms were the most difficult for him to describe and label. This process became a focus during the initial phase of psychotherapy. For example, he started making conceptual and emotional links between (a) the subjective emptiness, purposelessness, and worthlessness he felt overwhelmed by (and could now label) and (b) the decision to give up the CFO position that provided such buttressing and direction, but also opportunities to receive precious admiration and mirroring.

These insights paved the way for Marvin to begin noticing how robust was his use of control—to restrict his emotions, to ensure that others only ever perceived him as an astute leader, responsive mentor, and as an individual who was admirable, impressive, serious, disciplined, industrious, and entirely in control of himself. Marvin and the therapist explored how this defensive strategy manifested itself with family members, coworkers, in the milieu with fellow patients, and in the therapy relationship itself. For example, the therapist repeatedly pointed out when Marvin tried to impress her with tales about business adventures that indicated how accomplished, productive, and efficient he was; or when he tried to present himself as unusually informed about the epidemiology of specific mental disorders. Here, she made sure to interpret Marvin’s tendency to present his super-persona as a desperate effort to flee direct contact with his deeper self-image as inadequate, weak, defective, and shameful. Around this time in the individual work, he received push-back from fellow patients about his aloofness, invulnerability, and efforts to impress. In concert, these external relational forces moved Marvin to begin softening his compulsive-narcissistic armor, which then paved the way for him to become a little more vulnerable, and therefore more willing to authentically express despair, regret, and sadness about having decided to preretire. Notably, only then did he become capable of making emotional links between the painful self-concept (defective, weak), his subjective experience of emptiness, and the pressure he felt to save the family (of origin) from their shame and humiliation about their immigration status by becoming a super-achiever.

These insights altered the trajectory of Marvin’s healing journey. He eventually replaced his original psychotherapy goals (“manage anxiety and depression”) with a set of more nuanced and sophisticated goals for the psychotherapy work while still on the professionals unit and going forward. These new goals focused on addressing his underlying sense of inadequacy, correcting his distorted perceptions of others, and reducing his reliance on compulsive-narcissistic armor. He also started to come to terms with the fact that he lived most of his life feeling utterly responsible for saving the family system from shame and humiliation—and in the process, disowned or discarded parts of himself.

PSYCHOTHERAPY PROCESS GROUP. Throughout his treatment, Marvin attended a twice-weekly psychotherapy process group (80 minutes each session) based on a group-as-a-while interpretation model based broadly on the work of Wilfred Bion. In this adapted model, the group explores interpersonal communication, group dynamics, and group-as-a-whole processes, with particular attention to how the group manages anxiety about connecting in the here-and-now. Reliably, when group members feel safest, the group is most cohesive, so unconscious worries about the group's disintegration are in their lowest ebb—a condition that inclines group members to explore how they are relating to each other during the session. This level of connection requires the most significant vulnerability and is relatively rare because it involves exploring perceptions of each other and associated emotions in real time. This can include acknowledging and expressing caring and warmth, regret, sadness, hurt and anger, or envy. In this "safe" mode, group members are more inclined to share a great deal of detail about their internal struggles and the suffering that brought them to residential treatment.

In contrast, when group members feel less safe, the group is less cohesive, and unconscious worries about the disintegration of the group are ascendant—a mode that leaves members feeling insecure, too vulnerable to share details about their suffering, and more inclined to discuss superficial topics (e.g., the weather). Alternatively, they might unconsciously decide to complete the task of connecting in the here-and-now by forcing themselves into faux cohesion by finding a common enemy to rail against, for example, a political figure, a scapegoated staff member, the mental health system. Invariably, individual group members end up occupying roles (such as peacemaker, leader, outsider, helpless one) they adopted in their families of origin to reduce anxiety by nudging the process group toward cohesion or faux cohesion. In this model, the group "therapist" is not a facilitator, per se, but a consultant to working on the task. Consultations or interpretations are primarily directed toward the group-as-a-whole, for example, pointing out how the members have silently colluded to enlist one particular member to take up all the space with a monologue about his business history so that the group can avoid exploring and processing sadness they feel about the impending discharge of another beloved member. The ultimate goal of this approach is to help group members create the conditions for authentic connection and healing through emotional closeness and vulnerability. This goal is more likely to occur as group members learn more about how groups, in general, manage anxieties about working on a task and as each group member learns more about how they participate in these group-as-a-whole processes.

Given Marvin's defenses of aloofness, superiority, self-sufficiency, and invulnerability, he was slow to warm up to the psychotherapy process group experience. He declined initial offers by group members to share his reasons for coming to treatment. As they shared, he mainly remained quiet and disengaged, frequently looking out the large windows, lost in thought. After a few sessions, he began joining the group in some lighter conversations about experiences at work before treatment, though he limited his participation to noting some of his business accomplishments and describing how he has needed to "put some of my disrespectful underlings in their place." Members did not respond with open admiration, which frustrated him. Perhaps in response to this deprivation, he appeared to try moving into the role of wise elder by offering platitudes and problem-solving when members discussed their struggles. Some members were receptive and appreciative, while others saw through his defensive maneuvering and resented his superior stance, especially given that he had thus far avoided sharing anything deeply personal or painful. However, this resentment remained underground despite a few consultations about how the group seemed terrified of conflict and determined to ignore or deny palpable emotions in the room. Eventually, the dynamic shifted when the group added a new member (Libby) who was not particularly conflict-avoidant and was willing to articulate her frustration with Marvin about his persona of wisdom and invulnerability. In one group, she exclaimed: "You're like Gandalf or something, above it all…but I don't buy it. I think you're actually lonely. You'd have to be!" Her tone was both critical and caring. Marvin responded by referring to some of her recent dramatic behavior on the unit as "earth-mother histrionics." Other group members, emboldened by Libby's bravery, stepped into voice some of their pent-up frustrations about Marvin's superiority and inaccessibility, but also their desire to know him better, which he seemed to register. Notably, the discussion continued after the group ended and the consultant left the room. Three days later, much of the group tension was gone in the following group session, and Marvin began talking for the first time about the waves of regret and loss following his decision to preretire. His rumination and reassurance-seeking began to emerge, and members provided understanding and verbal support, which enabled him to become tearful during a later group session. In response to interventions in the group and milieu, he was becoming less of a closed system, for example, less aloof, more vulnerable, more able to allow himself to receive from others.

TERMINATION. After an 8-week stay, Marvin was asymptomatic. He left the hospital with an alternative plan of developing a part-time individual consulting practice and a self-help group therapy for higher managers going through burnout—it was an interesting outcome, a way of sharing his experience in the hospital with the less informed and perhaps less fortunate. On a personal level, Marvin continued his treatment with intensive psychodynamic work, group therapy, and medication management. Two years later, we had a call from him and learned that he was doing quite well.

FURTHER READINGS

Akiskal HS, Akiskal KK. In search of Aristotle: temperament, human nature, melancholia, creativity and eminence. *J Affect Disord*. 2007;100:1–6.

Allen JG, Fowler JC, Madan A, Ellis TE, Oldham JM, Frueh CF. Discovering the impact of psychotherapeutic hospital treatment for adults with serious mental illness. *Bull Menninger Clin*. 2017;81(1):1–38.

Blatt SJ, Ford RQ, Berman W, Cook B, Meyer, R. The assessment of change during the intensive treatment of borderline and schizophrenic young adults. *Psychoanalytic Psychol*. 1988;5(2):127–158.

Carroll BT. Kahlbaum’s catatonia revisited. *Psychiatry Clin Neurosci*. 2001;55(5):431–436.

Cooper AM. The representational world and affect. *Psychoanal Inq*. 2005;25(2):196–206.

Erikson EH. The Problem of Ego Identity. *J Amer Psychoanal Assn*. 1956;4:56–121.

Fico G, Caivano V, Zinno F, et al. Affective temperaments and clinical course of bipolar disorder: an exploratory study of differences among patients with and without a history of violent suicide attempts. *Medicina (Kaunas)*. 2019;55(7):390.

Gabbard GO. The treatment of the ‘Special’ patient in a psychoanalytic hospital. *Int Rev Psycho-Anal*. 1986;13:333–347.

Gabbard GO, Roberts LW, Crisp-Han H, Ball V, Hobday G, Rachal F. *Professionalism in Psychiatry*. American Psychiatric Association Publishing; 2012.

Hershberg SG. Interfaces among neurobiology, cognitive science, and psychoanalysis: implicit and explicit processes in therapeutic change. commentary on papers by Allan N. Schore, Wilma Bucci, and James L. *Fosshage. Psychoanal Dial*. 2011;21(1):101–109.

Jobes, DA. *Managing suicidal risk: A collaborative approach*. Guilford Press; 2016.

Kayatekin MS. *Psychoanalytic Conversations. From the Psychotherapeutic Hospital to the Couch*. Rowman & Littlefield; 2016.

Morrison J, Muñoz RA *Examining Psychiatrists and Other Mental Health Professionals in Kaplan and Sadock's Comprehensive Textbook of Psychiatry 59.2*. In: Sadock BJ, Ruiz P, Sadock VA, eds. Wolters Kluwer Health; 2017.

Nersessian E, Solms M. Editors' introduction. *Neuro-Psychoanalysis*. 1999;1(1):3–4.

Parish M. From couch to culture through the therapeutic community. *Psychoanal Psychol*. 2012;29(3):330–345.

Rihmer Z, Akiskal KK, Rihmer A, Akiskal HS. Current research on affective temperaments. *Curr Opin Psychiatry*. 2010;23:12–18.

Rosen D, Stukenberg KW, Saeks S. The group-as-a-whole object relations model of group psychotherapy. *Bull Menninger Clin*. 2001;65(4):471–488.

Silverman DK. Extenders or modifiers: a discussion of Weiss and Sampson's control-mastery theory. *Psychoanal Psychol*. 1990;7(1):125–136.

Strong CM, Nowakowska C, Santosa CM, Wang PW, Kraemer HC, Ketter TA. Temperament-creativity relationships in mood disorder patients, healthy controls and highly creative individuals. *J Affect Disord*. 2007;100:41–48.

Sutherland JD. Bion revisited: Group dynamics and group psychotherapy. In M. Pine, ed. *Bion and Group Psychotherapy*. Jessica Kingsley Publishers; 1985;47–86.

Thys E, Sabbe B, De Hert M. Creativity and psychiatric illness: The search for a missing link—an historical context for current research. *Psychopathology*. 2013;46:136–144.

58

Ethics and Forensic Psychiatry

58.1 Clinical-Legal Issues in Psychiatry

CLARENCE WATSON, J.D., M.D., SPENCER ETH, M.D., AND GREGORY B. LEONG, M.D.

Among the medical specialties, psychiatry plays a particularly distinct role within the province of the law, given its potential application across a wide range of legal issues. Contained at the core of most legal principles is the general presumption that adults of sound mind possess the capacity to understand, appreciate, reason, and choose throughout a boundless range of life activities. When those mental capacities are called into question at the behest of a legal quandary, psychiatry may be summoned to assist with reaching an equitable resolution. The presence or absence of a mental health or cognitive disorder may significantly impact the manner that a legal issue, whether civil or criminal, is perceived and ultimately managed. For instance, the presence of major mood or psychotic symptoms may impact an individual's legal right to refuse a recommended surgical procedure. Similarly, the presence of significant psychiatric symptoms at the time of a criminal act may negate an individual's culpability for that offense. The enduring relationship between the law and psychiatry has evolved over the centuries as a necessary element of humankind's aspiration toward a just resolution of legal disputes. In no other place is this more evident than with the historical development of the legal concept of insanity in criminal law. Once reserved only for those individuals whose minds were equated with that of a wild beast, the legal concept of insanity has progressed to varied jurisdictional approaches in determining the degree of mental impairment necessary to absolve an individual of criminal responsibility. Equally as significant is the historical development of the doctor–patient relationship. Once considered a lopsided paternalistic arrangement, in which patients were expected to passively accept the medical care directed by their physicians, the doctor–patient relationship has matured into a dynamic interaction that centers on patients' autonomy, including a patient's right to refuse treatment.

From an eagle's eye view, the interplay between psychiatry and the law can be broadly divided into situations involving clinical treatment interactions and situations involving specific legal matters that require the input of psychiatrists to settle a contested legal issue. The complexities within this broad division cannot be overstated, and accordingly, the subspecialty of forensic psychiatry was, in essence, produced from necessity and the ever-developing legal landscape. Clinical-legal issues such as patient confidentiality, informed consent, the right to refuse treatment, and medical standards of care are ubiquitous in the psychiatrist–patient therapeutic relationship. Potential involvement by psychiatrists as experts in contested legal matters, such as criminal cases, personal injury lawsuits, employment-related claims, testamentary disputes, or child custody cases, requires familiarity with legal principles and jurisdictional laws and the ability to apply psychiatric knowledge to the relevant legal inquiry at hand.

Over the course of centuries, the unique pirouette shared between psychiatry and the law has at times lacked grace, but nonetheless, has steadily moved toward a more elegant cooperative existence. Often marked by tensions inherent in speaking different native tongues of the respective realms of medicine and law, the cooperative existence has benefited from both medical and legal literature aimed at bridging the gap. This chapter is meant to contribute to those literary efforts and offers common ground for the understanding of clinical-legal issues involving psychiatry.

THE DOCTOR–PATIENT RELATIONSHIP

The concept of the doctor–patient relationship is naturally bound to the ancient roots of medicine itself. Evolving from a period when physicians were not distinguished from priests, magicians, or other ancient healers, the doctor–patient relationship initially resembled an archaic priest–supplicant relationship that hinged on a parental figure's power to manipulate nature on behalf of his ward. During the 4th century BC, the Hippocratic Oath set forth a code of ethics outlining principles that remain essential features of modern-day medical ethics, including patient confidentiality, abstaining from sexual contact with patients, and preventing harm to patients. The familiar dictum, *Primum non nocere* or first, do no harm, further distilled the fundamental ethical principle at the core of the doctor–patient relationship.

Clearly, at the heart of these traditional ethical doctrines, physicians are charged with acting only in the service of patients' medical benefit. The ethical principle of beneficence, where the ultimate medical benefit is paramount, was traditionally used to justify a paternalistic approach to medical care that was primarily physician-driven. However, during the late 19th and early 20th centuries, there was a growing sentiment that patients had the right to actively influence the course of their medical care. The concept of patient autonomy gradually gained ground and ultimately displaced the long-accepted institution of medical paternalism.

Patient autonomy was clearly endorsed in the 1914 case, Schloendorff v. Society of New York Hospital, which served as the figurative "canary in the coalmine" for medical paternalism in the dawn of an evolving informed consent doctrine. In that case, Justice Benjamin Cardozo ruled, "Every human being of adult years and sound mind has a right to determine what shall be done with his own body . . ." While not formally established in Schloendorff, the informed consent doctrine would become a more clearly defined legal requirement four decades later in the 1957 Salgo v. Leland Stanford Jr. University Board of Trustees case.

Certainly, at the inception of any doctor–patient relationship, there attaches a body of legal and ethical obligations owed by the physician to his patient. The shift away from medical paternalism toward the incorporation of patient autonomy permitted a more appropriate balance within the therapeutic relationship. Such a balance between a physician and his patient naturally facilitates a treatment partnership; however, there are clinical situations that call for prioritizing medical benefit over patient autonomy. Justice Cardozo's words in Schloendorff indicated his awareness of two important caveats regarding the “right to determine what shall be done” with one’s own body: adulthood and soundness of mind. It follows, of course, that the impact of existing psychiatric symptoms must be considered in the full clinical context when issues related to patient autonomy arise. Medical decision-making capacity, involuntary psychiatric treatment considerations, and patient confidentiality matters may all be influenced by the presence of significant psychiatric symptoms that temporarily deemphasize autonomy concerns.

MEDICAL DECISION-MAKING AND COMPETENCE

As initially framed in Salgo v. Leland Stanford Jr. University Board of Trustees, the law requires that physicians provide facts to their patients regarding treatment options that are “. . . necessary to form the basis of an intelligent consent.” Through this case law, the informed consent doctrine was established and formed a basis for physician liability when patient harm occurs in the setting of inadequate communication of the risks associated with treatment. In order to sufficiently obtain a patient’s informed consent, three elements must be present: the physician’s disclosure, the patient’s voluntariness, and the patient’s mental competence. Exceptions to the informed consent requirement include medical emergency, patient’s waiver of the physician’s disclosure requirement, and invocation of therapeutic privilege in situations where the disclosed information would be detrimental to the patient.

Generally, the informed consent disclosure requirement dictates that patients receive information regarding their diagnosis, the proposed treatment with the associated risks and benefits, treatment alternatives, and likely consequences of declining treatment. The 1972 case, Canterbury v. Spence, established that physician liability may be found if there is a causal connection between a physician’s failure to disclose and a patient’s injury, where the disclosure of the significant risks of treatment would have resulted in the patient’s refusal of that treatment. According to the court in Canterbury, that causal connection is based on “…what a prudent person in the patient’s position would have decided if suitably informed of all perils bearing significance.”

Depending on the state jurisdiction, the legal standard for an adequate disclosure may vary between a physician-based standard and a patient-based standard. In states that utilize a physician-based approach, doctors are required to provide information to patients that a reasonable physician in his specialty would disclose. States that utilize a patient-based approach may require an objective standard that considers the information that a reasonable patient similarly situated would want to know to reach a decision. Some states may require a subjective analysis of the patient-based standard that considers the information that the specific patient would want to know given their circumstances.

The voluntariness requirement directs that a patient be permitted to weigh the disclosed information and reach a decision unencumbered by coercion. Accordingly, threats, manipulation, or intentionally misleading information during a patient's decision-making efforts would nullify informed consent. A patient must be given the opportunity to freely consider the physician's disclosure and freely reach a medical decision.

The competency element requires that a patient possesses the mental capacity to understand, appreciate, and utilize disclosed information to reach a rational decision. Legally, all adults enjoy a presumption of competence in medical decision-making and other routine life activities until adjudicated otherwise. Traditionally, the term "competence" is a judicial determination regarding decision-making capability. Conversely, the term "capacity" refers to a clinical determination regarding a patient's decision-making abilities. Accordingly, the psychiatrist's role in competency evaluations is to determine an individual's mental capacity, which may later be formally adjudicated in a legal competency proceeding. In contested matters, courts have upheld the rights of a competent adult to refuse treatment. Courts may, however, adjudicate a person as incompetent if psychiatric symptoms or cognitive deficits impair an individual's mental capacity to communicate a choice, to demonstrate factual understanding and appreciation of his situation, or to rationally manipulate factual information related to his situation. In cases of adjudicated incompetence, treatment decisions generally shift from the patient into the hands of a substitute decision-maker or designated guardian.

Significant symptoms of mental illness or cognitive disorders are the typical culprits that undermine competence. However, the presence of those symptoms alone is not sufficient to establish a lack of mental capacity or to prove incompetence. In the 1973 case, In re Maida Yetter, a patient who was committed to a state psychiatric hospital refused treatment for possible breast cancer. Although delusional, the patient offered a rationale that justified her treatment refusal and allowed the court to determine that she was competent to make that decision independently and did not require a substitute decision-maker.

A controversial area related to medical decision-making capacity involves questions surrounding a terminally ill patient's right to die and physician-assisted death. Currently, ten states (Oregon, Washington, Vermont, California, Colorado, Hawaii, New Jersey, Maine, Montana, and New Mexico) and the District of Columbia legally permit physician-assisted death. In jurisdictions allowing terminally ill patients to choose physician-assisted life-ending measures, psychiatrists may be called upon to perform mental capacity assessments relating to those decisions.

INVOLUNTARY HOSPITALIZATION AND TREATMENT

Two traditional legal justifications, *parens patriae* and police power, provide state jurisdictions with the authority to involuntarily hospitalize or civilly commit mentally ill people for the purpose of treatment. *Parens patriae* refers to governmental power to care for individuals who are unable to care for themselves. The initial application of the parens patriae principle for the care of the mentally ill in the United States is found in the 1845 case, In re Oakes, which established the basis for courts to involuntarily hospitalize mentally ill persons for treatment and to prevent harm. The police power principle refers to the societal interest of permitting governmental authority to protect the general public from potential harm.

Jurisdictional statutes regarding involuntary psychiatric hospitalization vary from state to state but generally require the presence of a mental illness and a degree of risk of harm to the patient or to others due to the patient's symptoms of mental illness. Typically, the clinical determination for the necessity of civil commitment weighs a patient's current symptoms, relevant historical information, and collateral information sources (e.g., family members, friends, etc.) in the setting of a patient's inability or unwillingness to cooperate with the examination. During these assessments, the critical issue that must be addressed is whether patients' current psychiatric symptoms place them at risk of harming themselves or others. Additionally, some jurisdictions specifically require that involuntary treatment be implemented via the least restrictive alternatives available. Accordingly, clinical consideration must be given to alternatives to involuntary inpatient hospitalization, such as day treatment programs, group home placement, or other outpatient settings, depending on the circumstances.

Usually, these civil commitment assessments involve emergency situations in which family members, ambulance personnel, or police bring the patient to the emergency department because of concerns that the patient will imminently engage in suicidal, self-injurious, or violent behaviors. In addition, a voluntary patient might demand an immediate hospital discharge against medical advice and compel the psychiatrist to consider whether a civil commitment is necessary in lieu of discharge. Due to the urgency of these clinical situations, civil commitment laws permit clinicians to initiate involuntary emergency hospitalization without an initial court determination. This allows for immediate and necessary treatment interventions without the delay that awaiting a court hearing would cause. Following a statutorily determined period, a judicial review hearing is typically required to weigh the patient's need for continued involuntary hospitalization versus a patient's right to be discharged from inpatient treatment.

The legal standard of proof in civil commitment proceedings is "clear and convincing," which is less than the "beyond a reasonable doubt" standard in criminal trials. It is important to note that involuntary hospitalization does not per se permit medical treatment over patients' objections. Some jurisdictions require additional legal proceedings to permit the involuntary administration of medication over patients' objections. Following adequate treatment targeting the symptoms that necessitated the civil commitment, the patient may be discharged to the community to continue outpatient treatment. In some states, patients with extensive histories of treatment nonadherence and frequent hospitalizations may be evaluated prior to discharge for the appropriateness of outpatient civil commitment in order to improve community functioning.

In addition to standard civil commitment laws for psychiatric treatment, some states have enacted specialized civil commitment laws for the involuntary treatment of violent sex offenders who continue to pose a risk of danger following incarceration. In 1997, the U.S. Supreme Court upheld the constitutionality of sex offender civil commitment laws. Under these specialized provisions, these convicted and incarcerated sex offenders are evaluated prior to their prison release to determine if commitment criteria are present, including the existence of a congenital or acquired mental abnormality, the number and severity of previous sexual offenses, and dangerousness.

Psychiatrists cannot lose sight of the inherently intrusive nature of involuntary treatment. Accordingly, clinicians are ethically obligated to exercise due diligence in obtaining all relevant clinical data prior to determining the need for civil commitment. The psychiatrist continues to bear the ethical obligation of avoiding unnecessary breaches of patient confidentiality. However, testimony and communications to mental health courts required by law for the purpose of civil commitment are permitted without patient consent. Nonetheless, clinical efforts to preserve patient dignity and foster a therapeutic alliance with the patient remain paramount.

CONFIDENTIALITY AND PRIVILEGE

Confidentiality and privilege are related concepts that center on privacy rights belonging to patients regarding information shared with their physicians during the course of therapeutic relationships. Confidentiality represents physicians' obligation to their patients to not disclose information to third parties without a patient's consent. Privilege also deals with the protection of patient information; however, it refers to the more specific protection against physicians being compelled to provide protected patient information during court proceedings. Privilege is a right that specifically belongs to the patient, in that a physician cannot claim the privilege and refuse to disclose patient information once a patient has waived the privilege.

The doctrine of patient confidentiality is an ethical and legal duty that has been long recognized throughout the tradition of medicine, as evidenced by the Hippocratic Oath's charge to preserve patient privacy. Certainly, it is not difficult to understand the basis for this duty. The very nature of the doctor–patient relationship prompts patients to trust physicians with intimate details of their lives, relationships, and bodies for the purpose of medical treatment. Patients' candid communication of these details facilitates effective medical treatment and would likely be discouraged if physicians were free to reveal those details publicly.

Physicians may be held liable for breaches of patient confidentiality through legal theories of medical malpractice, invasion of privacy, and breach of contract for implied warranty of confidentiality. State legislatures have also enacted statutes that prohibit the unauthorized release of patient information. Accordingly, physicians may also be held liable for the breach of statutory duty of confidentiality. In addition, the U.S. Congress enacted the Health Insurance Portability and Accountability Act of 1996 (HIPAA), which provides a set of privacy regulations that govern the electronic transmission of health information for insurance eligibility and billing purposes. Psychiatrists should note that the practice of telepsychiatry also falls under HIPAA regulations. Accordingly, psychiatrists must ensure that their videoconferencing systems are compliant with HIPAA and state privacy requirements. Violations of HIPAA privacy regulations call for the imposition of fines. However, these regulations do not provide compensation to patients impacted by confidentiality breaches.

There are several exceptions to the confidentiality duty owed to patients by their psychiatrists. First, disclosures by psychiatrists without patient consent may be permitted in emergency situations in order to prevent imminent harm. Second, psychiatrists may breach confidentiality when acting to protect third parties from a potentially dangerous patient. In 1976, the California Supreme Court decided Tarasoff v. Regents of the University of California, which set into motion a nationwide trend establishing the duty of mental health professionals to protect or warn their patients' potential victims or to notify police of a risk of danger posed by a patient. Laws regarding the duty to warn or protect third parties vary by state jurisdiction and are defined by state statutory and case law. These laws provide an avenue for clinician liability if harm befalls a third party as a result of a mental health professional's failure to disclose the danger posed by their patient. Generally, disclosures made in good faith for the purpose of protecting the third party will not result in liability for breach of confidentiality. Third, psychiatrists may disclose patient information when adhering to state reporting requirements involving concerns such as child or elder abuse, unsafe drivers, or impaired health care professionals.

The physician–patient and psychotherapist–patient privilege doctrines evolved from the historical legal prohibition of compelled testimony from attorneys against their clients. The basis of that legal tradition was to protect the sanctity of the attorney–client relationship, which would be effectively eviscerated if lawyers could be forced to testify against the very clients that they represented. Much like the justification for requiring patient confidentiality, the physician–patient, and psychotherapist–patient privilege facilitate a more open and effective therapeutic alliance. In all state jurisdictions, patients possess the legal privilege to prevent their psychiatrists and psychotherapists from testifying about them. Federal courts also recognized patients' privilege following the U.S. Supreme Court's 1996 ruling in Jaffee v. Redmond. As the privilege belongs solely to the patient, psychiatrists have no basis for refusing to testify if their patients choose to waive that privilege.

However, the privilege is not absolute and may be pierced under a number of circumstances, thereby paving the road to compelled testimony by treating psychiatrists. Patients may lose the privilege in cases where they raise their mental status as an issue in a civil or criminal matter. Medical malpractice claims initiated by patients against their psychiatrists or therapists would also remove the privilege. The privilege does not apply in civil commitment proceedings. In child custody cases, the privilege may be set aside if the best interests of the child outweigh the negative impact of the psychiatrist's testimony. Finally, the privilege could be potentially nullified if the patient shared the information in question with their physician in the presence of a third party who is not involved in the patient's care.

PSYCHIATRIC MALPRACTICE

Psychiatrists tend to dread malpractice lawsuits in excess of the statistical likelihood of being sued and having a judgment entered against them. Across all medical specialties, there has been a major reduction in the rate of paid malpractice liability claims during the last decade, which, unfortunately, has not been accompanied by commensurate decreases in the cost of malpractice insurance policies. There are approximately ten paid claims for every 1,000 American allopathic physicians each year. By contrast, however, the Institute of Medicine estimates that more than 100,000 Americans die annually from medical mistakes, and about 6% of patients are injured by medical mismanagement. Recent data reveal that among an extremely large sample of physicians, malpractice actions cluster in a relatively small percentage: 89% of physicians had no malpractice claims, about 9% had only one claim, while the remaining 2% or so had two or more claims that amounted to nearly 40% of the total claims. Psychiatrists experienced about 4% of all claims, commensurate with their representation of 3.6% of the total sample of physicians.

About half of malpractice claims filed are litigated. Of these claims, courts dismiss more than half, and another 40% are resolved before a trial verdict. The remaining small minority of cases become jury trials, of which 80% result in a defense verdict favoring the physician. Nationally, the mean time for resolution of a litigated malpractice claim is 2 years, which has grown considerably because of the backlog of cases associated with COVID restrictions affecting many aspects of the legal process, particularly jury trials. In one large state, about 2% of all malpractice settlement reports were for psychiatrists, compared with 9% for internists and 10% for general surgeons. The majority of malpractice actions filed against psychiatrists derive from accusations of negligent treatment, adverse effects of prescribed medications, and suicidal and assaultive behavior of patients. The remainder mostly concerns alleged mistakes in diagnosis, breaches of confidentiality, and boundary violations. A psychiatrist’s participation in a malpractice settlement exceeding a threshold amount will result in a mandatory report to the federal National Practitioner Data Bank (NPDB), an entity that was first authorized in 1986 as a repository for information such as adverse licensure and hospital actions, and health care providers’ malpractice payments. NPDB entries are routinely accessed by hospitals and licensing boards and may trigger inquiries that cause discomfort for physicians who may have felt unfairly coerced to agree to a settlement.

The theory of medical malpractice refers to professional negligence that falls within the civil law domain, as defined by state statutes and case law. For example, in Florida, patients or their representatives may file lawsuits for personal injury or wrongful death against health care providers. In those malpractice actions, plaintiffs must prove by a preponderance of the evidence that the actions of the provider represented a breach of the prevailing professional standard of care. The standard of care in this context is a legal term that refers to skillful treatment, which in light of all relevant circumstances, is recognized as acceptable and appropriate by reasonably prudent providers in the same specialty, such as practicing psychiatrists.

The law recognizes that the mere existence of an injury does not automatically imply that provider negligence occurred without proof that the provider did, in fact, breach the standard of care. A bad outcome does not per se equate to bad care. An exception to this rule is the legal concept, *res ipso locquitor* or the thing speaks for itself, where the commission of an egregious error, such as failing to remove an abdominal clamp during surgery, shifts the burden of proof from the plaintiff to the defendant physician to show that negligence did not occur. For psychiatrists, such an egregious error would certainly include sexual misconduct during a therapy session under the guise of treatment.

The necessary elements for patients to successfully establish malpractice claims have been neatly summarized as encompassing four D’s: duty, dereliction, direct causation, and damages. The duty element requires that a doctor–patient relationship had been established, thereby creating a duty of care. Although seemingly straightforward for most treatment situations, there are some instances in which the nature of the professional relationship can be vague. Does a duty begin with the first contact between a potential patient and psychiatrist setting an appointment over the phone or only after the initial in-person or telehealth virtual platform meeting? When does the duty terminate? Does a covering physician assume responsibility for a patient if the patient calls for a medication refill or is seen briefly on rounds over the weekend?

The element of dereliction refers to a breach of the duty to provide reasonable and appropriate care. Psychiatrists are not legally expected to offer perfect, state of the art, or even excellent diagnostic and treatment services. All that is required is for the psychiatrist to perform as would other psychiatrists facing the same patient. Being board certified in psychiatry is neither necessary nor sufficient to establish that no deviation from the standard of care has occurred. It may be useful to demonstrate compliance with authoritative treatises, such as the Comprehensive Textbook of Psychiatry, recent peer-reviewed publications, or professional association treatment guidelines. However, it is an accepted principle that every person is unique and, consequently, patient-centered care may call for modifications in the usual therapeutic approach. Prescription medications are dispensed with package inserts containing information drug companies have submitted to the Food and Drug Administration (FDA). However, physicians are under no legal requirement to adhere to what is in the package insert and have substantial deference in prescribing “off label” as called in their clinical practices. Indeed, perhaps half of patients prescribed the common psychotropic medications are receiving them for off-label indications. Accordingly, a clinical judgment that adequately accounts for a patient's full clinical context, despite an eventual bad outcome, would likely assist in buffering against a claim of dereliction. When proposing an unusual treatment plan that may be later questioned, the psychiatrist would ordinarily be expected to obtain informed consent from the patient. The details of the consent discussion should be carefully inscribed in the medical record, and it may be helpful for patients to sign a note documenting their understanding and agreement.

The direct causation element requires a direct causal connection between a breach of the standard of care and a subsequent patient injury. The term "causation" is one that may have conceptually different meanings in medicine and law. For scientific-minded physicians, a cause–effect relationship identifies that a particular prior event is the physical cause if it results in the later effect. The law takes a very different approach. The salient issue is the proximate cause, by which the focus of the legal inquiry is whether the act in question was necessary to produce the injury. A common test for direct causation is whether the harm would not have occurred "but for" the prior action.

Legal commentators have argued that direct causation is that necessary element of a set of conditions that are jointly sufficient for the result. For example, a psychiatrist's older adult patient endorsed symptoms of depression, weakness, postural instability, and osteoporosis. The patient fell to the floor while rising from the office couch at the end of a session and immediately complained of severe pain. The patient was taken to a local hospital emergency room and diagnosed with a hip fracture. Surgery was performed, but the patient developed a postoperative infection that ultimately proved fatal. This tragic chain of events undoubtedly had multiple underlying medical causes. However, from a legal perspective, the psychiatrist could be found to be negligent and the proximate cause for all of the patient's subsequent injuries and death. Depending on the facts, the crux of the matter may be that, but for the mishap while rising from the psychiatrist's couch, the patient would not have had a broken hip, surgery, an infection, and death. In this regard, the patient was surely an "eggshell" whose fall was the "straw that broke the camel's back." As lawyers are fond of saying, one takes plaintiffs just as you find them.

The damages element refers to the plaintiff's alleged injuries that resulted from the physician's dereliction of duty. Civil law recognizes several categories of damages that may accrue to victims of malpractice, all of which are exclusively financial. Monetary awards to successful plaintiffs are intended to compensate the plaintiff for the harm suffered. Compensatory damages address what the patient has suffered as a direct consequence of the substandard care and include present and future medical costs, lost income, physical and mental impairment, pain and suffering, and decreased life expectancy. Punitive damages, also known as exemplary damages, may be awarded in addition to compensatory damages when a defendant's behavior was found to be particularly egregious or malicious. Punitive damages are not meant to compensate a plaintiff's loss; instead, they are meant to punish the defendant and to serve as a deterrent to the defendant and others regarding similarly wanton behavior. Punitive damages, which are not covered by insurance, rarely arise in malpractice actions against physicians.

The specter of serial medical malpractice claims, analogous to mass tort litigation, became a grim reality in the aftermath of the conviction of a Michigan State University (MSU) sports physician for sexually abusing multiple young women under the guise of treatment. MSU agreed to pay $500 million to 332 alleged victims, thereby settling their lawsuits without a trial. The University of Southern California (USC) was faced with at least 1,700 claims of sexual misconduct during gynecologic examinations by a physician practicing for many years in the campus student health clinic. The two negotiated and court-approved settlements for the USC student plaintiffs totaled $1.1 billion. These universities had employed the physicians and were thus ultimately responsible for the damages inflicted under the legal theory of *respondeat superior* that holds an employer liable for employees' wrongful or negligent acts. The "deep pocket" universities were the target of civil legal action, while the physicians were pursued separately and aggressively by the criminal justice system.

Certain aspects of psychiatric practice involve a greater risk of liability. Electroconvulsive therapy (ECT) is a procedure that usually calls for additional training and special precautions and may be associated with a greater likelihood of adverse events than conducting a psychotherapy session. Consequently, psychiatrists who wish to perform ECT may need to purchase insurance policies that specifically cover this procedure. Prescribing certain psychiatric medications also warrants a cautious approach. Conventional antipsychotic medications place patients at risk of developing tardive dyskinesia, which, when severe, can be disfiguring and disabling. Successful malpractice cases have been filed when psychiatrists failed to alert patients of this side effect risk and neglected to perform and document routine examinations for its detection. Similarly, the use of atypical antipsychotic medications may call for routine assessments for weight gain and abnormalities of blood lipid and glucose metabolism in addition to patient discussions regarding these associated treatment risks. Increasingly, psychiatrists use formal education materials to supplement informed consent discussions when prescribing common psychotropic medications, such as benzodiazepines, stimulants, and lithium.

The specter of patient suicide affects all practicing psychiatrists; indeed, it is a truism in the field that there are two kinds of psychiatrists—those who have already had a patient commit suicide and those who will. Rather than encouraging fatalism, the ever-present danger of suicide for certain groups of patients should motivate psychiatrists to remain cognizant of the basic suicide-reducing strategies. Knowledge of the evolving epidemiology of suicidal behavior is fundamental in assessing static and dynamic risk factors for each patient. There is no question that asking patients about the presence and severity of depressive symptoms and suicidal thoughts is essential for proper care. Such direct inquires do not suggest to or predispose a patient to become actively suicidal. The Joint Commission has suggested the implementation of standardized suicide risk assessment instruments, such as the Columbia-Suicide Severity Rating Scale (C-SSRS), that can be embedded as mandatory fields in electronic note templates so that clinicians must explicitly address the issue of suicidal ideation in settings such as the emergency room. Psychiatrists must respond to the escalating risk of suicidality with very careful monitoring and appropriate modifications in the treatment plan.

No discussion of psychiatric malpractice is complete without consideration of boundaries. Much has been written about the conscious and unconscious forces that are in play in both patient and psychiatrist during a session, especially in the context of psychotherapy. Maintaining the integrity of the therapeutic frame, however challenging, is the linchpin of treatment. Of particular concern are boundary violations that constitute frank lapses in care, such as psychiatrists becoming romantically involved with a current or former patient. This boundary violation is usually characterized by a progression from a seemingly innocent conversation, to more frequent social contacts, to physical affection, to sexual activity. In these cases, psychiatrists can exhibit remarkable denial regarding the serious consequences of their own behavior. Not infrequently, a transgressing psychiatrist, such as one who may be lonely, depressed, and substance-abusing, begins a destructive relationship with a vulnerable patient, whose judgment may be impaired by an idealizing transference and self-defeating tendencies. Such unethical behavior may result in malpractice and licensing board actions that jeopardize a professional career.

THE PSYCHIATRIST AS AN EXPERT

Expert Witness Testimony

Psychiatrists may be asked to participate in civil and criminal court matters by providing professional opinions based on their psychiatric knowledge and experience as applied to a particular legal question. Typically, the opinions are provided to a requesting party, generally the court or an attorney, either as a written or oral report. In some instances, the opining psychiatrist is required to testify in court, with a jury or judge considering such testimony along with other evidence in order to reach a verdict. Psychiatric expert testimony takes place in the public purview and is often viewed by the public as a “battle of the experts” or, more pejoratively, battle of “hired guns.”

In some cases, expert witness testimony is legally required, such as in cases involving allegations of medical malpractice. For example, in order to support a medical malpractice claim that a defendant psychiatrist’s treatment led to a patient’s suicide, the plaintiff must present expert testimony to show what a psychiatrist of ordinary skill, care, and diligence would have done differently under similar circumstances. Of course, should this malpractice case go to trial, there may also be psychiatric expert witnesses testifying on behalf of the defendant psychiatrist. Two court cases have governed the admissibility of scientific expert testimony in the United States, Frye v. United States and Daubert v. Merrell Dow Pharmaceuticals, Inc. As both cases indicate, it is ultimately the judge’s decision as to the admissibility of the expert witness testimony.

Role Conflict as a Psychiatric Expert Witness

The fundamental professional role of psychiatrists involves a clinical or therapeutic perspective. In the clinical or therapeutic setting, the treating psychiatrist deals with the patient's narrative truth and makes no further attempts at confirming a patient's statements. In contrast, when a psychiatrist works outside of the health care setting in the legal arena, the clinical or therapeutic perspective loses its primacy, and an individual's self-report should be viewed with objectivity and hence skepticism. Assuming both the therapeutic and expert witness roles for the same individual invite erosion of the psychiatrist–patient relationship, whether it involves completing a disability evaluation in which the treating psychiatrist does not find support for the requested disability, or especially in the case of testifying in court as a psychiatric expert witness.

Separating the therapeutic and forensic roles is the best solution, though, at times, the dual role cannot be avoided, as in the case of involuntary civil commitment or involuntary medication hearings in which the treating psychiatrist is usually the only professional to testify. There are times in which a treating psychiatrist could be called to testify as a fact witness regarding a patient, but this would entail testifying about symptoms, diagnosis, prognosis, and similar clinical matters versus providing the ultimate psychiatric-legal opinion such as those involving capacity or causation.

Testamentary Capacity

With the growth of the older adult population, psychiatrists are being confronted with an increasing incidence of its emblematic mental condition, dementia, and the related clinical-legal issues of capacity and competence. Although adults are assumed to have capacity for the various activities of daily living and life tasks, that presumption is rebuttable, especially in the presence of neurocognitive impairment. Each specific deliberative legal act, such as testifying in court as a witness, executing a financial transaction, and even getting married, has its own criteria for mental capacity that must be satisfied in order for that action to be valid. Among the older adults, the process of writing a last will and testament is a common and highly valued event that has important implications that will arise after the death of the testator.

Legislatures and courts recognize that making a will is frequently in the province of persons whose mental acuity is significantly less than young individuals engaged in other tasks. Consequently, the cognitive standard for testamentary capacity is generally set particularly low. In most jurisdictions, the necessary elements are that persons understand: (1) the fact that they are actually making a will, (2) the nature and extent of their property, (3) the disposition of that property, and (4) who are their natural heirs. From a mental status perspective, this domain of knowledge does not require that the person making the will is fully oriented and, indeed, having the diagnosis of Alzheimer disease per se would not disqualify a person from having sufficient cognitive ability to make or change a will if the criteria for testamentary capacity could be established.

Although simple to state, the prerequisite criteria for testamentary capacity may be, in practice, quite difficult to ascertain after a person's death when the will is finally read. For instance, a person could have been suffering from severe dementia at the time of death. However, the document was actually prepared years earlier, and it is the state of mind at the time when the will was written that is dispositive if the document is later contested. A retrospective assessment conducted after death is challenging when there is inadequate evidence of the individual's prior knowledge and abilities. Medical records, deposition testimony from relatives and acquaintances, and information about activities of daily living can contribute to a determination of capacity.

Since wills may be the product of debilitated older adult persons who depend on the physical and medical care of others, there is another requirement that must be satisfied before the terms of the will are honored. Testamentary capacity demands that the testator not be subject to the undue influence of another person. For example, the philanthropist Brooke Astor died at the age of 105 in 2007. Her older adult son was convicted in criminal court of preying upon his mother's diminished mental state by defrauding her in life and coercing her to change her will in his favor rather than directing her assets to her favorite art institutions. An adult child who spends considerable time with a vulnerable parent may be expressing love and devotion. Alternatively, that situation may be exploited to benefit the child at the parent's expense.

Psychiatrists may be asked to opine whether a litigant heir exerted control over the person making a will. Certainly, whether there existed a confidential relationship and whether the will was rewritten to inappropriately benefit that person are critical factors. In this setting, a "confidential relationship" between two individuals occurs when one of the parties has an obligation to act in good faith for the benefit of the other party. The psychiatrist's expertise will be most relevant to the question of the extent to which the testator was vulnerable to being influenced due to medical or psychiatric illness and the power of the asymmetric relationship.

Criminal Competencies

The presence of competent decision-making is necessary for a criminal defendant navigating the criminal justice process, starting from the initial arrest to the implementation of the death penalty. While competence is presumed for adults, the presence of a mental disorder may negate competence. The criminal competencies include: competency to waive Miranda rights, competency to confess, competency to stand trial, and competency to be executed.

The U.S. Constitution and its Amendments, along with subsequent federal and state case law and statutes, delineate the legal safeguards for individuals accused of committing crimes. At the entry point into the criminal justice system, law enforcement officials must inform a criminal suspect of their Miranda rights, such as the right to remain silent and the right to counsel. The presence of an active mental disorder or intellectual disability can negate an individual's capacity to voluntarily, knowingly, and intelligently waive their Miranda rights. Though very few adults lack the capacity to waive their Miranda rights, the majority of incompetence to waive Miranda rights involves juveniles where factors such as immaturity and emotional vulnerability are more likely to adversely impact voluntariness, knowledge, and/or intelligence. The competency to confess might be considered under the broader competence to waive Miranda rights. In 1986, the U.S. Supreme Court in Connelly v. Colorado ruled that absent physical coercion by police, an actively psychotic individual, who approached a police officer to confess about committing a homicide as a result of command auditory hallucinations to confess, had provided a valid confession.

The most common forensic psychiatric evaluation in the criminal justice system is that of competence to stand trial. The criteria for competence to stand trial is guided by several U.S. Supreme Court decisions, with two having the most significance, Dusky v. U.S. (1960) and Godinez v. Moran (1993). The Dusky case established that competency to stand trial is composed of two prongs: (1) understanding of the criminal proceedings; and (2) reasonable ability to consult with defense counsel. In the Godinez case, the court ruled that if criminal defendants are competent to stand trial, they are also competent for other activities in criminal proceedings, such as the competence to waive counsel, competence to enter a plea (waive trial), and competence to be sentenced.

For jurisdictions that permit capital punishment, the U.S. Supreme Court has mandated certain exclusions to the imposition of the death penalty, including mental retardation (now referred to as Intellectual Disability in the DSM-5-TR) in Atkins v. Virginia in 2002 and for those under age 18 at the time of committing the capital offense in Roper v. Simmons in 2005. The U.S. Supreme Court in Ford v. Wainwright in 1986 and Panetti v. Quarterman in 2007 prohibited the execution of condemned inmates who are incompetent to be executed. Incompetence to be executed involves a condemned inmates' inability due to a mental disorder to understand the death penalty and the reason it is being imposed on them. Although cases of competence to be executed are quite rare, it has raised several ethical issues for society and the medical profession.

Criminal Responsibility and Psychiatric Illness

The determination of culpability and subsequent societal sanctions for those committing a criminal act finds its origins at least as far back as the dawn of recorded human history, along with considerations given to those with abnormal mental conditions. The evolution of English law in the area of criminal responsibility of the mentally disordered over the past millennium has exerted great influence on how criminal responsibility has been determined in the United States. Although the number of psychiatric evaluations for defendants claiming that a mental condition negated their criminal responsibility for an offense accounts for a small proportion of evaluations performed for the criminal justice system, the issue of criminal responsibility of those with mental disorders is arguably the most visible to the general public due to extensive media coverage, both fictional and nonfictional.

To be proven guilty of committing a crime, the prosecution must prove beyond a reasonable doubt that the accused person not only committed a prohibited act (*actus reus*) but also had the requisite mental state or "guilty mind" (*mens rea*) for the forbidden act.

The insanity defense is used to prove that an individual did not possess the necessary mental state for criminal responsibility due to mental illness at the time of the act. From the preceding millennium, the concepts of the “wild beast test” and the ability to distinguish between good and evil (right and wrong) represented society’s efforts to legally define “insanity.” These efforts evolved and culminated in the formulation of the M’Naghten insanity standard in 1843 by the English House of Lords following an insanity acquittal in the English trial of Daniel M’Naghten. The M’Naghten insanity test required the defendant to prove that they were “laboring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or, if he did know it, that he did not know he was doing what was wrong.” The M’Naghten standard was quickly adopted for use by the U.S. judicial system. The federal government, the District of Columbia, and 46 states permit some form of the insanity defense. Among the 46 states that permit the insanity defense, approximately half use a M’Naghten type of insanity test. Four states (Montana, Idaho, Utah, and Kansas) disallow the insanity defense. In 2020, the U.S. Supreme Court affirmed that states are not required to provide an insanity defense (Kahler v. Kansas).

While the majority of state jurisdictions use the M’Naghten insanity standard, some states use other insanity tests. Currently, the most common other type of insanity test is that of the American Law Institute (ALI). To qualify for insanity under the ALI insanity test, a criminal defendant must demonstrate that as a result of a mental disorder, they lack “substantial capacity to either appreciate the criminality of his conduct or to conform his conduct to the requirements of the law” (Model Penal Code §4.01, 1962). When comparing the two current major insanity tests used in the United States, the M’Naghten test has two “cognitive” prongs (knowledge of the “nature and quality” and “wrongfulness”) versus ALI’s one “cognitive” prong (“appreciate the criminality”) and one “volitional” prong (“conform conduct”). Interestingly, some states that use an ALI-type of insanity test do not use the “volitional prong,” thereby functionally making the insanity standard more like the M’Naghten test. Wording differences between the M’Naghten and ALI insanity tests would suggest that an insanity defense would be more likely successful in jurisdictions using the ALI insanity test.

In actuality, the overall use of the insanity defense is infrequent. The defense is raised in approximately 1% of felony cases with an estimated 25% success rate. The vast majority of insanity acquittals involve prosecutorial concurrence prior to trial. It is estimated that juries reject the insanity defense in 93% of cases going to trial. It has been argued that the legal system's requirement to not inform juries of the potential legal consequence of an insanity acquittal, that is, that a defendant acquitted by insanity would not be automatically released into the community, but would be subject to commitment in a state mental health facility, adversely impacts jurors' willingness to accept insanity defenses.

CHILDREN AND THE LAW

Children and adolescents can become enmeshed in the U.S. legal system in a multitude of ways, some of which are comparable to adults, and some are fundamentally different. In fact, the number of youth involved in court actions has increased dramatically, driven by criminal activity in gangs and on their own, by the escalating rates of foster care, divorce, and custody disputes, and by the proliferation of child abuse neglect, and victimization. Further, there continues a veritable parade of personal injury lawsuits filed on behalf of child victims that alleged harm from sexual misconduct by generations of clergy, teachers, and others. Psychiatrists would benefit from a familiarity of the basic concepts of children and the law, if for no other reason than to better understand their adult patients whose early histories may be replete with such encounters and whose younger family members may become entangled in some aspect of the law.

Children, Juvenile Justice, and Violence

Juveniles can be both perpetrators and victims of crimes. The indisputable developmental differences between prepubertal children and adolescents, and especially adults, have legal implications insofar as cognitive phase-specific limitations of youth accused of criminal behavior would seem to necessitate special measures if not an entirely different method of adjudication. Indeed, states created parallel juvenile justice systems for youthful offenders that, in many ways, were responsive to the developmental immaturity that characterizes children and younger adolescents. In 1899, Illinois established the first juvenile court separating the adjudication and confinement of youth as delinquents from the conviction and imprisonment of adults as criminals. The scope of juvenile courts expanded to encompass both criminal acts and behaviors that would not be considered illegal for adults, such as curfew violations, drinking alcohol, and truancy, which have been called status (as a minor) offenses. The rationale for juvenile courts' broad jurisdiction is the promotion of social warfare by prioritizing education and rehabilitation rather than punishment.

Every state proceeded to adopt a system of juvenile courts, whose procedures were then significantly modified by the Supreme Court's landmark cases under Chief Justice Earl Warren. In 1967, the U.S. Supreme Court held that, although it was permissible to treat juveniles and adults differently in court, youthful offenders facing possible incarceration are entitled to the same constitutional rights of notice of charges, counsel, protection from self-incrimination, and the opportunity to confront and cross-examine witnesses, as afforded to adults. Three years later, the Supreme Court also ruled that the "beyond a reasonable doubt" standard of proof also applied in juvenile court. Extending constitutionally based procedural rights to juveniles did not imply that juveniles could escape conviction and punishment, especially if they committed serious crimes. Prosecutors have the discretion to file for a judicial waiver that moves the youth from juvenile court to adult court in order to deny the protections that juvenile jurisdictions provide, including the preclusion of long prison sentences.

Most states exclude first-degree murder charges from the juvenile system, which for many years effectively allowed a convicted juvenile murderer to be sentenced to execution. The eligibility of a person who committed their crime as an adolescent for capital punishment proved to be controversial even among death penalty advocates. The Supreme Court reviewed the question of juvenile capital punishment and decided in Roper v. Simmons (2005) that defendants were not eligible for execution if their crimes were committed under the age of 18 years. The court's opinion in this decision drew heavily on scientific data concerning neuropsychological development that contrasted adolescents with adults, which were presented in amicus briefs submitted by psychiatric and psychological professional organizations. Five years later, the Court, in Graham v. Florida (2010), held that the U.S. Constitution prohibits sentences of life without parole for juvenile offenders who did not commit homicide. In those cases, a juvenile offender is not guaranteed eventual release, but if a life sentence is imposed, the juvenile must be afforded a realistic opportunity to obtain release before the end of that sentence. In Miller v. Alabama (2012), the U.S. Supreme Court extended its reasoning to preclude as unconstitutional a mandatory sentence of life without the possibility of parole for juveniles convicted of homicide offenses.

The Columbine High School tragedy in 1999 engendered an awareness of a looming threat of mass killings occurring in a school, possibly perpetrated by an adolescent or young adult. Despite heightened vigilance of the danger of a lethal attack on campus, a mentally ill college student at Virginia Tech University in 2007 murdered 33 persons, and in 2012 a mentally ill adolescent opened fire on and mortally wounded 20 children at the Sandy Hook Elementary School in Connecticut. More recently, in February 2018, a 19-year-old former student expelled from Marjory Stoneman Douglas High School in Parkland, Florida, killed 17 people in that school. There have been strident demands to develop strategies to profile potentially deadly students who may be targeting schools for deadly violence. Youths with the risk factors of a history of psychiatric illness, aggression, suicidal behavior, and/or family strife and who have access to weapons may indeed be at a higher likelihood for violence. However, less than 5% of crimes overall can be attributed to mental illness. Identifying those individuals at risk with any degree of certainty must overcome the daunting challenge of the high false-positive rate associated with predicting a rare event.

Most experts are skeptical of the wisdom of deploying armed security in elementary and high school buildings. Further, implementing zero-tolerance policies for aggressive behavior may disproportionally impact minority, disadvantaged, and special needs students who generally benefit the least from suspensions or expulsions. Alternatively, school-based interventions designed to decrease bullying and stigma, provide enhanced access to students for counseling and mental health care, expand communication among students, faculty, and parents, and eliminate or at least safeguard guns in the home all hold the promise of more effectively addressing the threat of interpersonal and self-inflicted violence.

Child Custody

Each year in the United States, there are about half the number of divorces as there are marriages. Divorce proceedings sever the marriage contract between spouses, and although children may be a focus of the case, they are not considered formal parties to the legal action. The court has the responsibility to protect the well-being of dependent children affected by divorce, which may be straightforward when the parents are in general agreement or hotly contentious when they are in battle. Child custody subsumes both legal custody, referring to the responsibility for making major decisions about health and education, and physical custody, referring to which parent is spending time with the child. Many jurisdictions favor awarding joint legal custody to both parents unless there is evidence that this would be frankly detrimental to the child's best interests. Under those circumstances, the court may grant sole legal custody to one parent, with or without time-sharing with the other parent. A determination of the best interests of the child is made by evaluating all of the factors affecting the welfare and interests of the minor child and the circumstances of that family, including, but not limited to: the demonstrated capacity of each parent to facilitate and encourage a close and continuing parent–child relationship and to determine, consider, and act upon the needs of the child as opposed to the needs or desires of the parent; the moral fitness and mental and physical health of the parents; evidence of domestic violence, sexual violence, child abuse, abandonment, or neglect; and the developmental stages and needs of the child and the demonstrated capacity of each parent to meet the child's developmental needs.

Treating psychiatrists of a parent or child in custody or visitation litigation may have important information that could influence the court's ruling. The scope and limits of confidentiality in such matters may become a difficult and contentious question. Further, the availability of the psychiatrist to write reports, testify, or advocate in these adversarial legal proceedings is fraught with conflicting ethical, legal, and therapeutic duties. Many treating psychiatrists have sought to avoid becoming entrapped in these dilemmas by limiting their documentation and recollection, though it may not be possible to predict which patient will be the source of a subpoena. Consultation with an attorney and referral to a forensic psychiatrist may be advisable for the treating psychiatrist confronted with such circumstances.

Child Abuse

Reports of suspected maltreatment, which are mandated by law for health care and school personnel, will trigger investigations by the police or a child protective agency. If probable cause for maltreatment is found, such as bruises or malnutrition, the child can be removed from the home, and the agency might petition the court for alternative placement in foster care. Allegations of sexual abuse usually result in criminal charges against a parent, family member, or nonfamilial perpetrator. As the victim of interpersonal violence, children and adolescents suffer enduring psychological injuries, which may also lead to civil court actions seeking to recover compensation for the injured child. Younger children are more vulnerable since, as a rule, they are less able to protect themselves from being exploited by adult predators.

Society and the medical profession have traveled far in the last 100 years from a time when physical child abuse was unrecognized to the current era of far greater sensitivity to multiple forms of maltreatment, including neglect, emotional and physical abuse, and sexual molestation. However, significant difficulties arise in the evaluation of allegations of sexual abuse, especially in young children. The psychiatrist examining a child about whom there is an accusation of sexual abuse must face the challenge of infusing scientific objectivity into the practical task of providing judgments regarding the validity of symptoms, their putative etiology, the harms suffered by the child, the prognosis, and what present and future treatment interventions are indicated.

FURTHER READINGS

Amols L, Eth S. Childhood trauma, PTSD, and the cautious forensic expert. *J Am Acad Psychiatry Law*. 2017;45:70–74.

Appelbaum PS, Gutheil TG. Confidentiality and privilege. In: Appelbaum PS, Gutheil TG, eds. *Clinical Handbook of Psychiatry & the Law*. 4th ed. Lippincott Williams & Wilkins; 2007.

Atkins v. Virginia, 536 US 304 (2002).

Canterbury v. Spence, 464 F.2d 772 (DC Cir. 1972).

Colorado v. Connelly, 479 US 157 (1986).

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 US 579 (1993).

Death with Dignity: in your state. Available at: www.deathwithdignity.org. Accessed July 29, 2021.

Dusky v. United States, 362 US 402 (1960).

Ford v. Wainwright, 477 US 399 (1986).

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

Godinez v. Moran, 509 U.S. 389 (1993).

Graham v. Florida, 560 U.S. 48, 82 (2010).

Greenberg SA, Shuman DW. Irreconcilable conflict between therapeutic and forensic roles. *Prof Psychol Res Pract*. 1997;28:50.

Harvard Law Review Association. Developments in the law: civil commitment of the mentally ill. *Harv Law Rev*. 1974;87:1190–1406.

In re Maida Yetter, 62 Pa. D. & C. 2d 619 (1973).

Jaffee v. Redmond, 518 US 1 (1996).

Janofsky J, Hanson A, Candilis P, et al. American Academy of Psychiatry and the Law. Practice guideline: forensic psychiatric evaluation of defendants raising the insanity defense. *J Am Acad Psychiatry Law*. 2014;42(4 Suppl):S3–S76.
Kahler v. Kansas, 140 S.Ct. 1021, 589 US ___, (2020).
Meyer DJ. Psychiatric malpractice and administrative inquiries of alleged physician misconduct. *Psychiatr Clin North Am*. 2006;29:615–628.
Miller v. Alabama, 567 U.S.460 (2012).
M’Naghten’s Case, 8 Eng. Rep. 718 (HL 1843).
Model Penal Code §4.01 (1962).
Mrad DF, Watson C. Civil commitment. In: Drogin EY, Dattilio FM, Sadoff RL, Gutheil TG, eds. *Handbook of Forensic Assessment: Psychological and Psychiatric Perspectives*. Wiley; 2011.
Panetti v. Quarterman, 551 U.S. 930 (2007).
Roper v. Simmons, 543 US 551 (2005).
Salgo v. Leland Stanford Jr. Univ. Bd. Trustees, 317 P.2d 170 (Cal. 1957).
Schloendorff v. Society of New York Hospital, 105 NE 92 (N.Y. 1914).
Shariful A, Dixson BA, Constantino E, Regan J. The law and practice of off-label prescribing and physician promotion. *J Am Acad Psychiatry Law*. 2021;49:53–59.
Strasburger LH, Gutheil TG, Brodsky A. On wearing two hats: role conflicts in serving as both psychotherapist and expert witness. *Am J Psychiatry*. 1997;154:448–456.
Studdert DM, Spittal MJ, Zhang Y, Wilkinson DS, Singh H, Mello MM. Changes in practice among physicians with malpractice claims. *N Engl J Med*. 2019;380:1247–1255.
Tarasoff v. Regents of the University of California, 551 P.2d 334 (Cal. 1976).
Voskou P, Douzenis A, Economou A, Papageorgiou SG. Testamentary capacity assessment: legal, medical, and neuropsychological issues. *J Geriatr Psychiatry Neurol*. 2018;31:3–12.
Weinstock R, Leong GB, Silva JA. Competence to be executed: an ethical analysis post Panetti. *Behav Sci Law*. 2010;28:690–706.
Williams SC, Schmaltz SP, Castro GM, Baker DW. Incidence and method of suicide in hospitals in the United States. *Jt Comm J Qual Patient Saf*. 2018;44:643-650.

58.2 Ethics in Psychiatry

LAURA WEISS ROBERTS, M.D., M.A.

Ethics is a scholarly discipline that brings a greater understanding of moral aspects of human experience and action. Psychiatry, more than any other field of medicine, concerns itself with studying and strengthening the capacities that make humans unique and that allow people to attach, learn, love, reason, feel, live with one another, and make choices. Separately and together, the fields of ethics and psychiatry examine human meaning, intention, and consequences, and both are concerned with the expanse and limits of the self in the world. Thus, by their natures, ethics and psychiatry are overlapping fields of inquiry.

With so much in common, it is perhaps surprising that the field of psychiatry enjoys more than its share of ethical controversy. Psychiatry is viewed with skepticism by society at large, and the practice of psychiatry presents a number of concerns that can be seen as ethically problematic. Even psychiatrists themselves disagree on matters of ethical importance within the profession. Long-standing ethical issues related to power and intimacy in psychodynamic–psychotherapy relationships are now joined by intriguing new ethical questions about privacy and boundaries in treatment in the modern era of telehealth and smartphone deliverables. Similarly, older dilemmas regarding the potential exploitation of vulnerable individuals in psychiatric research have now been joined by new debates regarding patient autonomy, unconscious bias, and equity in healthcare. Old controversies regarding patient confidentiality in publishing patient case reports are now joined by new uncertainties regarding privacy in technologies that use machine learning or artificial intelligence, for example, to identify individuals who are at risk of suicide based on social media activity. Novel issues pertaining to topics as widely varying as psychiatric genetics, interpersonal violence, sexuality, and global mental health disparities arise each day among the global context of a changing climate, the aftermath of a novel pandemic, and increasing demands for racial justice, highlighting the complex and evolving nature of advancing mental health care.

Learning about ethics is helpful preparation for all physicians, including psychiatrists and psychiatrists-in-training who encounter complex dilemmas in their everyday work. Physicians might begin by familiarizing themselves with relevant ethical guidelines, including those of the American Medical Association (AMA), the American Psychiatric Association (APA), and the American Academy of Child and Adolescent Psychiatry. Historical ethical codes, such as the Hippocratic Oath of the 4th-century BC and Percival’s *Medical Ethics* from the late 1800s, shed light on how standards of conduct for physicians have evolved. Psychiatrists must have a sound working knowledge of ethics resources from related areas, as well, such as professionalism, legal concerns, gender-related issues, and cultural competencies.

Are general codes of ethics for physicians sufficient for an understanding of the ethical requirements of psychiatrists? Should psychiatrists be held to a higher or different standard of ethics than other physicians? Many psychiatrists argue that a more complete knowledge of ethics is particularly important for mental health practitioners because of the nature of mental disorders and the impact of stigma. Others argue that greater mastery of ethics is essential because of the influence that psychiatrists may have on their patients' lives as a result of transference, the set of emotions and thoughts brought to the therapeutic relationship from the patient's prior life experience. Yet others argue that true ethics expertise, robust decision-making skills, and disciplined ethics practices are of special importance for psychiatrists because of their professional power to initiate involuntary treatment in certain situations. For all of these reasons, psychiatrists and psychiatrists-in-training will ideally possess a thorough understanding of the field of ethics, including its underpinnings and its applications.

Toward this aim, this chapter introduces the foundations of present-day ethics in psychiatry. Outlined are key bioethics principles and illustrations of approaches to identifying and resolving psychiatric ethics dilemmas.

ETHICAL SCHOOLS OF THOUGHT

Several ethical schools of thought have influenced how physicians understand ethical aspects of medical practice. Examples include the deontologic school, utilitarianism, justice-based ethics, virtue-based ethics, and care ethics.

Table 58.2–1.
Examples of Ethical Models of Thought

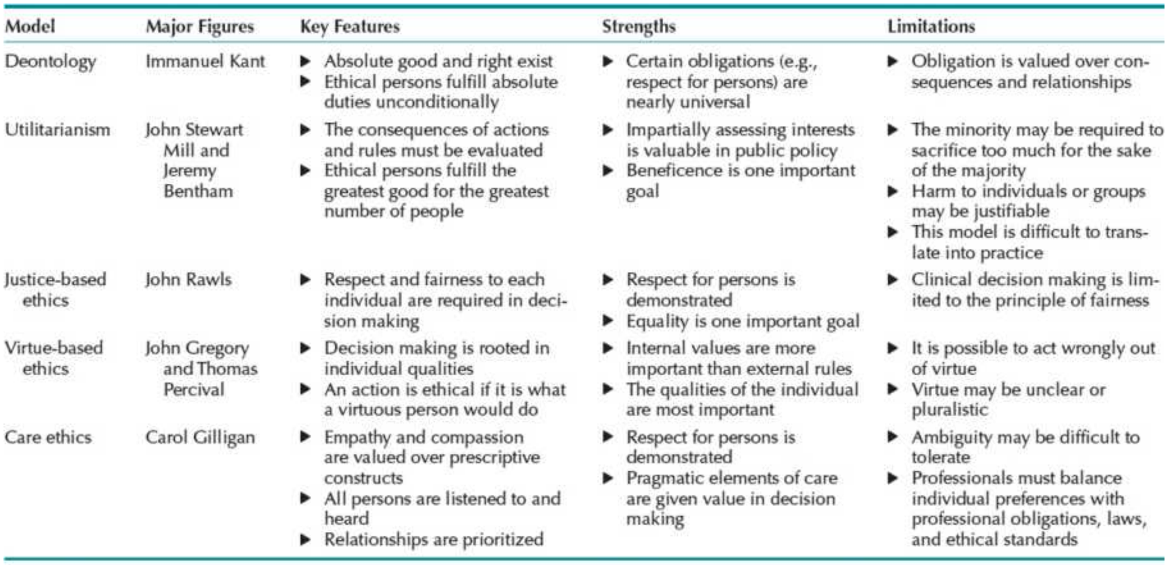

Model	Major Figures	Key Features	Strengths	Limitations
Deontology	Immanuel Kant	▶ Absolute good and right exist ▶ Ethical persons fulfill absolute duties unconditionally	▶ Certain obligations (e.g., respect for persons) are nearly universal	▶ Obligation is valued over consequences and relationships
Utilitarianism	John Stewart Mill and Jeremy Bentham	▶ The consequences of actions and rules must be evaluated ▶ Ethical persons fulfill the greatest good for the greatest number of people	▶ Impartially assessing interests is valuable in public policy ▶ Beneficence is one important goal	▶ The minority may be required to sacrifice too much for the sake of the majority ▶ Harm to individuals or groups may be justifiable ▶ This model is difficult to translate into practice
Justice-based ethics	John Rawls	▶ Respect and fairness to each individual are required in decision making	▶ Respect for persons is demonstrated ▶ Equality is one important goal	▶ Clinical decision making is limited to the principle of fairness
Virtue-based ethics	John Gregory and Thomas Percival	▶ Decision making is rooted in individual qualities ▶ An action is ethical if it is what a virtuous person would do	▶ Internal values are more important than external rules ▶ The qualities of the individual are most important	▶ It is possible to act wrongly out of virtue ▶ Virtue may be unclear or pluralistic
Care ethics	Carol Gilligan	▶ Empathy and compassion are valued over prescriptive constructs ▶ All persons are listened to and heard ▶ Relationships are prioritized	▶ Respect for persons is demonstrated ▶ Pragmatic elements of care are given value in decision making	▶ Ambiguity may be difficult to tolerate ▶ Professionals must balance individual preferences with professional obligations, laws, and ethical standards

Reprinted with permission from Laura Weiss Roberts, Dyer AR. *Concise Guide to Ethics in Mental Health Care.* American Psychiatric Publishing; 2004.

These perspectives differ greatly on what constitutes the beliefs and actions of a good person. The philosophers who developed these paradigms sought to explain as much as possible with a few simple rules. Such simplicity is elegant academically, though harder to apply in real life, as Table 58.2–1 notes. Each school of thought, taken alone or taken to an extreme, may lead to conclusions that most people in a pluralistic society would find unacceptable. Nevertheless, these models provide a framework for discerning and evaluating moral issues encountered in everyday experience.

The deontologic school centered on the work of Immanuel Kant (1724 to 1804), who articulated the inherent sacredness of the individual person and the primary intention of a human being as a moral agent. He believed that certain ways of behaving are absolute from an ethical perspective, irrespective of consequences. Kant articulated a rule called the “categorical imperative” guiding all moral behavior. His first formulation of the categorical imperative was that a person should “act only on that maxim through which you can at the same time will that it should become a universal law.” For example, in deciding whether to lie or steal or kill, individuals should think about whether they want others to engage universally in these behaviors. Kant’s second formulation of the categorical imperative was to “act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.” In other words, it is unethical to use another person solely as an object to obtain a goal without concern for the effect of one’s actions on the person. Moral value, in Kant’s system, relates to respect, rationality, and duty in relation to the moral agent.

The utilitarian school of thought originated with Jeremy Bentham (1748 to 1832) and John Stuart Mill (1806 to 1873) in 19th-century England. They argued that the value of an act should be judged solely on its effect or consequences. Acts are ethically right, in this view, if their effect is to produce the greatest good for the greatest number of people. Acts are not right or wrong in and of themselves, according to act-utilitarians. Lying, stealing, and killing are not inherently wrong, for example. In this school, the ends justify the means. Rule-utilitarianism moves beyond act-utilitarianism to suggest that good individuals will undertake decisions guided by rules that—if followed by all people—would bring about the greatest good for the greatest number of individuals. In this formulation, there is more “forgiveness” for decisions with unpredictably negative outcomes. Rule-utilitarianism does not address the role of the intentions of individual moral agents, nor does it answer the question of what is “good.”

More recently, in contrast with utilitarianism, John Rawls (1921 to 2002) developed a school of thought drawing from his concept of justice. It followed upon strengths of Kant’s deontologic approach in that it highly valued the individual as a moral agent. Rawls supported the rights of the individual as a matter of fairness. He gave primacy to the intrinsic worth of each person rather than to a metric that assessed the greatest utility for the group as a whole. He stated, “Each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override.” Rawls asserted that in deciding what to do and what rules and systems to establish, people should place themselves behind a “veil of ignorance” and assume that they do not know who they are in the system. For example, in deciding what welfare benefits a society should offer and the amount of taxes to impose on the rich, people should assume that they do not know whether they will be rich or poor. In deciding what authority doctors have over patients, people should assume that they do not know whether they will be in the role of the doctor or in the role of the patient. Rawls’ approach, as with that of Kant, does not rely on the consequences of actions but, rather, their primary intent.

Virtue-based theory has contributed greatly to the ethical framework of medicine and the key concepts of professionalism. John Gregory (1724 to 1773) was a Scottish physician–ethicist, and Thomas Percival (1740 to 1804) was an English physician–ethicist. Their work, taken together, created the foundation for the modern ethical concept of medicine as a profession. Concerned about exploitative, self-interested, and entrepreneurial practices of physicians at the time, they articulated the virtues of scientific and clinical competence, dedication to the interests of the patient, and service to the public through ensuring the skills of peers and students or apprentices. Gregory described the moral qualities of an astute and compassionate physician as follows:

> The chief of these is humanity; that sensibility of heart which makes us feel for the distresses of our fellow-creatures, and which, of consequence, incites us in the most powerful manner to relieve them. Sympathy produces an anxious attention to a thousand little circumstances that may tend to relieve the patient; an attention which money can never purchase. . . If the physician possesses gentleness of manners, and a compassionate heart, and what Shakespeare so emphatically calls 'the milk of human kindness,' the patient feels his approach like that of a guardian angel ministering to his relief: while every visit of a physician who is unfeeling, and rough in his manners, makes his heart sink within him. (Gregory, John. *Lectures on the Duties and Qualifications of a Physician*. Carey & Sons; 1817:22,23.)

Gregory and Percival individually emphasized the virtues of the honorable physician as demonstrating self-sacrifice and self-effacement and as seeking at all times to advance the field of medicine, not for financial gain but the benefit of others in society. The modern notion of integrity derives from this line of thought, and the work of Gregory and Percival continues to inform the discussion of conflicts of interest and how they may be avoided or resolved.

Carol Gilligan (born 1936) developed care ethics as a skeptical response to the history of ethical thought, in which concepts of "moral development" elevated predominately by empowered white men were naturalized while historically marginalized voices were excluded, amounting to arbitrary moral indifference. While deontologic, utilitarian, and justice- and virtue-based ethics primarily employ prescriptive ethical constructs, care ethics emphasizes modes of cognition, such as empathy and compassion, and pragmatic elements of care, such as attentiveness, responsiveness, and interdependence, that require epistemic openness, humility, and tolerance of ambiguity. Care ethics affirms a decision-making process in which all individuals are valued according to these ideals.

These schools of thought have been highly influential in present-day principles of ethics that the AMA and APA advance. Other schools of thought exist in the United States and throughout the world, with rich and long-standing traditions and different implications for current practice. From legal and ethical points of view, physicians engaging in professional activities in the United States are held to the AMA and APA codes of ethics and standards.

BASIC ETHICAL PRINCIPLES

The modern field of biomedical ethics takes a primarily principle-based approach. In their influential text *Principles of Biomedical Ethics,* Beauchamp and Childress defined the most critical principles to be autonomy, beneficence, nonmaleficence, and justice. The Belmont Report on human research ethics elevated the concept of respect for persons, and ethics scholarship has emphasized the principle of respect for the law. Other principles, such as compassion, confidentiality, fidelity, integrity, and veracity, are also central to the practice of psychiatry. See Table 58.2–2. Ethical analyses examine the interplay of these principles and seek to identify and resolve when ethical principles are in conflict.

APPROACHING ETHICAL DILEMMAS

Ethical dilemmas do not typically emerge in simple situations of doing good versus doing bad. In such situations, there is no dilemma. Indeed, ethical dilemmas occur when two seemingly good things are at stake. Consider the example of a psychiatrist who receives a request from a worried and loving mother to provide information about issues arising in her son’s therapy sessions. The son has a good relationship with his mother but has been bullied at school, feels embarrassed, and does not want his parents to know what is bothering him. For the psychiatrist, the good of responding empathically and with accurate information to the distressed mother competes with the good of protecting the fragile patient alliance with the son. Dilemmas often come when two or more of the basic bioethics principles are in conflict or cannot be completely fulfilled given the circumstances of a particular situation.

Table 58.2–2.
Examples of Biomedical Ethics Principles

Autonomy	The notion of self-rule; the capacity to make authentic decisions related to one's body and mind.
Beneficence	Doing good; the commitment to seek to bring about benefit.
Compassion	Deep regard for the experiences and suffering of others.
Confidentiality	A legal privilege associated with the right of privacy; the obligation not to disclose information obtained from a patient or observed or gathered in caring for a patient.
Fidelity	"Faithfulness" or loyalty to ethical ideals.
Integrity	The notion of being whole or complete; the capacity to adhere wholly to the principles of the profession.
Distributive justice	Equitable distribution of benefits and burdens in society.
Nonmaleficence	Avoiding harm or injury to others.
Respect for the law	The obligation to adhere to the law.
Respect for persons	Fundamental regard for the dignity, sacredness, and value of the individual.
Veracity	"Honesty," involving the positive duty to tell the truth and the negative duty to avoid deception.

Recognizing Ethical Issues and Tensions

The ability to identify ethical issues and tensions present in the patient care situation is thus the first and most basic of the ethics skills important to the sound clinical practice of psychiatry. Indeed, it is increasingly appreciated that every aspect of clinical care can have ethical significance. The judgment of a clinician that a patient's characteristics, symptoms, or challenges warrant a psychiatric diagnosis has ethical meaning. Moreover, the observation that individuals who belong to underrepresented minority groups are more likely to receive more serious and more stigmatized clinical diagnoses, such as schizophrenia, than individuals who belong to majority groups– despite having the same symptom profiles and clinical histories– reveals the role of bias and potential discrimination in assessment and diagnostic procedures.

Recognizing the ethical features of a patient’s case requires sensitivity, and this capacity is enhanced by knowledge of ethical principles and how they may come into conflict.

Seeing how ethical principles may compete or collide is informative and can help practitioners to discern a path forward. For instance, consider the case of a newly diagnosed cancer patient who, overwhelmed and fearful, attempts suicide by overdose. She requests aid in dying from her oncologist, who also feels overwhelmed and fearful because of the patient’s reaction, and refers the patient for a psychiatric evaluation. The psychiatrist learns that the patient’s mother died from cancer and that she was her mother’s only caregiver over the long and difficult course of her illness. The psychiatrist finds that the patient likely had a pre-existing depression before learning of her cancer diagnosis. The patient is decisionally capable, despite her distress, and the psychiatrist feels great empathy for the patient. In this example, many principles are in play, and several are in tension: most clearly, honoring the patient’s wish to die (autonomy) conflicts with the physician’s responsibility to avoid harm (nonmaleficence). A superficial analysis would suggest that the patient’s right, especially in an autonomy-oriented society, would trump all other considerations.

Sound standards for clinical and ethical care are those that are informed by the full set of ethical principles that govern medicine. By recognizing the two principles that are in opposition—in this case, autonomy and nonmaleficence—and considering the other principles at play in the situation, the clinician begins to see additional options for intervention that align with standards of medicine. In this case, identifying ways of reducing the patient’s physical and psychological suffering, treating her pre-existing condition, providing accurate information about the nature of her condition and potential treatments, and adhering to state legislation regarding assisted suicide all may contribute to an improved physical and mental health outcome for this patient. Ideally, clinical care and ethical reflection work together, as illustrated in this case.

The disciplined exercise of identifying ethically salient concerns in a clinical situation and framing and exploring the relevant bioethics principles can help open up the thinking of the clinician. Reflecting on related areas that influence ethical facets of the case, such as considerations regarding the law, the patient’s background or identity, and the patient’s culture is also valuable. These efforts, taken together, represent an important ethical skill and the initial step in clarifying ethically sound strategies for optimal patient care.

Gathering Additional Information and Expertise

Ethical questions sometimes arise in a clinical situation when there is simply too little information or when clinicians are working at the boundary of their experience or scope of practice. In such situations, additional information or broadened and deepened expertise may lead to a ready resolution of the problems at hand. Data or expert guidance could come from the patient’s prior medical records or clinical practice and ethics guidelines. In addition, speaking with an experienced supervisor, seeking advice from a colleague with specialized expertise, requesting ethics or legal consultation, and/or reviewing supplemental clinical data from outside sources may help.

Consultation-liaison psychiatrists see the value of additional data-gathering and consultation as resolving apparent ethical issues frequently in their work. Psychiatric consultation services have long fulfilled an important ethical function in most hospitals. For example, if a patient is declining potentially life-saving care for what appears to be an irrational reason and is demanding to return home against medical advice, a psychiatric evaluation may reveal an underlying condition, such as an addiction, delirium, early dementia, or poor pain control, interfering with the patient’s decision making. Once this underlying condition is treated and the symptom burden is lessened, the patient’s decisional capacities may improve, leading to what appears to be a more rational decision to the caregiving team. If the patient is not experiencing a threat to their decisional capacity, it is important to explore the patient’s values, background, and reasons for making a decision that contrasts with the medical recommendation. Cultural and religious preferences, fears based on past experiences, implicit bias, or issues related to family stresses may be influencing the decision. Seeking additional background information to evaluate these factors may be very helpful in developing health-preserving strategies for this patient, such as consulting a religious leader, inviting emotional support from a family member or counselor, or identifying appropriate social service resources. In all of this, the caregiving team is wise also to obtain consultation from the legal service of the hospital. Information on the role of state or federal law in the case will help the physicians to determine whether temporary or long-term guardianship may be appropriate to pursue.

Thus, in approaching an ethically complex clinical situation, recognizing the ethical dimensions of a patient's care and then seeking additional information or expertise are the two initial steps to take in resolving dilemmas and developing ethically astute treatment plans.

Identifying Risky Situations and Arranging for Appropriate Safeguards

Many situations in patient care present predictable ethical challenges, and with experience, astute clinicians begin to anticipate when these ethically risky situations are likely to occur. A clear illustration is when a patient is impulsive, and the caregiving team must ensure the safety of the patient and others, for example, family members or the clinic staff. The clinician may be tempted to make choices that differ from usual clinical practices because of the situational threat. When the patient is a celebrity or an influential person, other ethical risks can enter into the clinical picture. Staff members may be tempted to review the patient's electronic medical record, and caregiving physicians may begin to change their usual clinical practice, consciously or not, in order to please or garner favor with the special patient. A very different example is when there is a need to report child abuse or neglect. In this case, the psychiatrist may feel emotionally conflicted, wanting to abide by the law but feeling worried about the impact of the intervention on the therapeutic alliance with the family or about the impact of the investigation on the child. When entrusted with the care of vulnerable individuals, such as trauma survivors, children, or extremely ill patients, psychiatrists will recognize numerous ethical risks that may arise.

Ethical dilemmas signal the need for additional safeguards in order to ensure that an appropriate standard of care is upheld. Additional safeguards may include incorporating a broader health care team, introducing alternative decision-makers for vulnerable patients who lack decision-making capacity or obtaining permission from the patient to engage the family more fully. When a psychiatrist routinely works with high-risk patients, the clinical setting itself must have more safeguards, such as physical safety training for clinical staff. A geriatric psychiatrist who works with many individuals with advanced dementia may need to have a close working relationship with other clinicians, legal advocates, and the court in order to facilitate optimal patient care. In sum, heightened safeguards help address the ethical risk that is associated with psychiatric practice in all settings. When such risk is predictable because of the context or the patient population, ethically astute psychiatrists will build safeguards into their routine practices for the good of the patients and the caregiving team.

Working Within One’s Scope of Practice and Engaging in Continuous Learning

Working within one’s scope of clinical competence is considered ethically important because of the risk of doing harm when practicing with insufficient expertise to care for a patient properly. If physicians are confronted with patient care situations that are outside of their scope of competence, then referral or, at the very least, consultation and supervision are appropriate. Sometimes physicians practice in settings where referral is not feasible. In rural settings, for example, telehealth services may bring additional expertise to the care of the patient. In training settings, where early-career physicians may lack expertise, close supervision helps to ensure an appropriate standard of care. In emergency situations, clinicians are expected to do their best and not to abandon the patient in extreme need.

Related to the ethical obligation to work within one’s areas of competence is the responsibility to continue to learn. This imperative relates to the fundamental concepts of serving the best interest of the patient and avoiding harm in clinical care. Remaining aware of advances in one’s field of practice and domains of competence is critical to optimal patient care and ensures that one’s practice remains within community standards.

Reflecting on One’s Values and Perspectives

Psychiatrists are taught to pay close attention to their responses to a patient, including how they respond to a patient’s appearance, emotions, and personal story. Implicit bias greatly shapes interpersonal interactions in everyday life and in clinical care. Awareness of potential sources of bias and of countertransference (the set of emotions and thoughts that the psychiatrist brings to the therapeutic relationship from prior experiences) is valuable in clinical care and provides insights into the ethical dimensions of a patient’s care. Indeed, to conduct ethically sound patient care, psychiatrists must be able to see how their experiences, attitudes, and feelings influence their perceptions of the ethical aspects of a clinical situation. Awareness of one’s reactions may help prevent ethical blind spots and even breaches of ethical conduct. When a psychiatrist finds a patient very attractive or when a patient makes the psychiatrist very uncomfortable or angry, these emotional cues should be seen as clinical and ethical signs that require acknowledgment, clinical interpretation, and a therapeutic response.

Medical professionals are just as susceptible to biased thinking as the general population. Implicit and explicit biases held by health professionals perpetuate health disparities and nonequitable systems of care. A commitment to addressing discrimination and identify and eliminate structural bias through clinician training and other methods is increasingly viewed as a professional value and necessary to uphold ethical principles.

Ideally, team-based care and teaching settings will create opportunities to consider how different team members or supervisors, and trainees view the ethical issues present in a patient care situation. Astute team leaders will facilitate explicit discussion of these issues and guide their colleagues in defining the conflicting ethical principles. Clarifying differences in the points of view of team members may prevent later misunderstandings. Team leaders can demonstrate the importance of speaking with the patient, gathering additional data, working within one’s scope, introducing specialized expertise, and building in appropriate safeguards.

DECISION-MAKING STRATEGIES

Jonsen, Siegler, and Winslade created a four-part model for clinical decision making that many health care settings have adopted. The four elements to consider when making a decision include (1) clinical indications, (2) patient preferences, (3) quality of life, and (4) socioeconomic or external factors. In this model, clinical indications have the greatest weight on decision making. Clinical indications refer to the physician’s obligation to diagnose a patient’s condition carefully, to inform and educate the patient about the illness process, to consider the available treatment options in light of appropriate standards of care, to make therapeutic recommendations that adhere to these standards, and to ensure that treatments are provided competently and fulfilled properly in partnership with the patient.

The second element of this decision-making model, patient preferences, is grounded in the principles of respect for persons and autonomy. Once the physician presents the patient with recommendations based on what is clinically indicated, the physician elicits the preferences of the patient. In the practice of psychiatry, the nature of mental disorders may interfere with patients’ decisional capacity, voluntarism, or ability to identify or express their preferences. The ethically astute psychiatrist will thoroughly evaluate decisional capacity, keeping in mind that this capacity may change with time and under certain circumstances. When a patient is unable to provide informed consent, appropriate decision-making safeguards, such as an alternative decision-maker, may be put in place in accordance with state law.

Quality of life is the third element in the Jonsen, Siegler, and Winslade model, referring to the sense of personal satisfaction individuals express and/or experience with regard to their physical, mental, and social situations. In assessing quality of life, physicians must directly engage with patients regarding their values and points of view. The physician must not make assumptions about what represents a good life for the patient. Clinical indications and patient preferences are weighted more heavily than quality of life in the decision-making model because, when overweighted or incorrectly assessed, quality of life could be used to justify decisions that do not meet current ethical standards in society.

The fourth element in this clinical decision-making model entails socioeconomic or external factors. These factors are diverse and include the interests of society, the role of family members, cultural factors, and costs and limitations in access to care. This element is last primarily because of the obligations of beneficence. The United States defines denying a patient life-saving treatment in an emergency on the basis of ability to pay as unethical and illegal, for example. In less acute situations, however, a discussion with patients about the costs of care and how these costs may affect their choices of treatment is important and ethically acceptable.

EXAMPLES OF ETHICAL PRACTICE STANDARDS IN PSYCHIATRY

Several texts and resource documents provide comprehensive overviews of ethical practice standards in psychiatric practice. Key concepts related to several ethical practice standards for psychiatry are briefly outlined below and draw upon the schools of thought, biomedical ethics principles, and the four-part ethical decision-making model described in this chapter.

The Therapeutic Relationship and Professional Boundaries

The therapeutic relationship presupposes that the physician is worthy of the patient's trust. The physician's expertise and the physician's intention to do good, avoid harm, and place the patient's best interest above all other interests are the basis of that trust. Professional boundaries are behaviors that align with these fundamental commitments to serve the wellbeing of the patient.

Interactions with patients fulfill or gratify the interests of the physician rather than the interests of the patient can fracture or destabilize the foundation of trust. One example is the ethical problem of intimate relationships arising between physicians and patients. The APA is explicit in its direction that psychiatrists must not engage in sexual relations with patients, including former patients. Such intimacies are inherently exploitative; they interfere with therapy, risk harm to the patient, and undermine confidence in the profession, thereby damaging the ability of others to benefit from psychiatric care. The profession of psychiatry sees sexual contact during therapy or ending therapy in order to start a relationship as indefensible.

Some have argued that the power differential between a former therapist and patient is the primary reason to forbid relationships between them. Many relationships, however, have power differentials because of the different earning capabilities of the partners, different levels of educational achievement, or a society that gives more power to one gender, one class, or one ethnicity over another. The reason to bar sexual relationships between psychiatrists and present and former patients is more complex than that. The prohibition protects patients from being lured into a relationship that the patient would not otherwise desire because of a genuine need for treatment. The prohibition also helps to ensure that the psychiatrist stays focused on the patient's wellbeing. Self-effacement is an important virtue that asks physicians to put away their own needs or wishes for the sake of the patient’s health and wellbeing. The modest and self-effacing psychiatrist is willing to forgo relationships with patients outside of standard treatment. The Exploitation Index developed by Epstein helps therapists to identify potentially exploitative practices in the care of their patients.

Therapeutic boundary violations are behaviors that subordinate the interests of the patient to the interests of the physician and are not compatible with the ethical practice of psychiatry. These violations are, by definition, exploitative, although they need not be sexual or romantic. Boundary violations include seeing patients outside of normal office hours or in nonclinical locations, engaging in social or business relationships with patients, accepting gifts from patients, and making nonsexual physical contact with patients. “Friending” or following a patient on social media or completing a Google search for a patient’s name are generally considered boundary violations because these actions are likely to serve the interests of the therapist rather than the best interests of the patient. These types of violations carry the potential to take advantage of the patient or to harm the treatment relationship and should therefore be avoided.

Therapeutic boundary crossings are behaviors that are intended to serve the interests of the patient but differ from usual ethical practices. A psychiatrist disclosing personal information to build rapport with a distant and troubled patient or accepting a small handmade gift at the holidays to demonstrate respect for a patient’s culture are two examples of boundary crossings.

Traditional views of the boundaries in the therapeutic relationship derive from reflection on long-term, dynamically oriented psychotherapy relationships, typically occurring in urban contexts. Novel questions arise when considering other roles, interventions, and settings. For example, is a child psychiatrist caring for impoverished and traumatized youth in a culturally unique community held to the same expectations as a therapist caring for an older, successful businessperson in a major metropolitan location? Are psychiatrists who see patients once in an emergency department held to the same standard of personal–professional role separation as their ambulatory clinic counterparts? Does a psychiatrist who is consulted about the suitability of a patient with Parkinson disease for neurosurgical intervention establish a therapeutic relationship with the patient in the same manner as with a patient needing treatment for depression? Intuitively, these situations require more nuanced responses than strict rules or inflexible codes of ethics. For this reason, medical ethics is understood to evolve by its very nature. This fact does not permit psychiatrists to deviate from expected standards but, rather, permits psychiatrists to apply standards (and, more importantly, their underlying intent) to novel situations.

Confidentiality

The importance of confidentiality in tending to illness dates back to Hippocrates, who said, “What I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account one must spread abroad, I will keep to myself, holding such things.” Confidentiality refers to the physician’s duty to not release information learned in the course of treatment. Salient information may relate to disclosures by the patient, observations of the physician, or laboratory findings, or notes from prior treatment in the medical record. Protection of patient confidentiality fosters the therapeutic alliance and is a prerequisite for patients to be willing to speak freely and openly with their doctors. Table 58.2–3 notes illustrations of optimal practices for protecting confidentiality in psychiatric practice.

Table 58.2–3.

Examples of Optimal Approaches to Confidentiality in Clinical Practice

- Provide accurate information to patients about the realities of confidentiality in the clinical setting.
- When appropriate, remind patients that the purpose of the medical record is to be shared as a way of providing optimal care by a team of health professionals.
- Openly discuss with patients the legal limitations of confidentiality and avoid assuring patients that their disclosures will be “completely” confidential.
- Do not speak to others or, if working in a team-based care setting, avoid talking with others beyond the health care team, about patients’ care without explicit, written permission to do so.
- Actively work within organizations and professional societies to develop state-of-the-art confidentiality safeguards as science, clinical care, health policy, the law, and health care systems evolve.
- Seek consultation and expert guidance when confidentiality safeguard questions arise.
- Model best practices in confidentiality to colleagues, team members, and trainees.

Psychiatric records, including even the identification of a person as a patient, must be protected with particular diligence. A psychiatrist may release confidential information only with the authorization of the patient or when compelled by law. Even when patients have given permission for their personal information to be used, safeguards are still needed. It is important to disclose only that information that is truly necessary for the situation at hand. In teaching and research, one should not unnecessarily release patients’ names or identifiable information. Presenters of patient material in ward rounds and case conferences should remind attendees not to repeat what they hear. Confidentiality and the ethical obligation to withhold information endure even after a patient’s death.

In an era of electronic health records, digitized medical information, and a push toward national health identifiers, whether, to what degree, and how special privacy protections for mental health treatment will continue to be guarded remains unclear. As patients are increasingly able to access their entire medical record, new issues come to light about the extent to which mental health records may receive additional or special safeguards. Although “break the glass” protocols for accessing mental health records are now common, such protocols still allow many different providers access. Mental health providers must remain aware of who has access to the documentation in a patient’s electronic medical record. Transparency of health records has been adopted in many countries. Psychiatrists thus have a responsibility to consider the mental health impact that patients may experience if they access their health records. Psychiatrists may elect to review health record documentation with their patients.

New ethical questions arise every day regarding confidentiality in novel forms of treatment, whether medical or psychosocial. For instance, what are appropriate privacy protections for psychotherapy occurring by telephone, video, Internet, email, or texting? Should health data collected via wellness applications and passive symptom tracking be awarded the same protections and safeguards as data collected in a hospital or clinic setting? How will innovators in clinical research and genetics protect the personal information of specific individuals whose biomarkers, taken together, represent a unique biologic signature?

Informed Consent

Informed consent is an ethical and legal standard that exemplifies the principle of respect for persons and seeks to bring the principle of autonomy to life in clinical medicine. Informed consent is the idea that capable and knowledgeable individuals may freely choose a course of action pertaining to their health. Informed consent can involve accepting a recommended treatment or choosing an alternative, including no treatment (i.e., informed refusal).

Informed consent involves three components: information-sharing, decisional capacity, and voluntarism. With respect to information-sharing, physicians should provide patients accurate and balanced information regarding (1) the nature of their illness or condition; (2) the need for intervention; (3) the anticipated risks or benefits and their relative likelihood, including rare but very serious risks; (4) appropriate alternatives and their attendant risks and benefits; and (5) future consequences of the different choices, to the extent possible. Physicians should offer this information at the level and in a language that the patient understands and, to the extent possible, in an unpressured environment.

Physicians should build opportunities to clarify points of information and to correct misimpressions into the conversation. Individuals who cannot provide consent must receive information at the level at which they can understand it, and physicians should seek the individuals' acceptance when possible. See Table 58.2–4.

Grisso and Appelbaum characterized decisional capacity as having four subcapacities. The first is the physical and mental ability to communicate. This subcapacity does not refer to the ability to speak a particular language but, rather, the ability to express—including through eye blinks or hand signals—a choice. The second subcapacity is the ability to understand relevant factual information. The third subcapacity is the ability to reason, that is, to weigh the different factors and risks and to think through the choice at hand in an objective and flexible manner. The fourth subcapacity is the ability to appreciate the meaning of the decision and to evaluate the choices in the context of one's life. This capacity is close to the concept of insight, which usually refers to patients' ability to understand the nature and impact of their illness. The capacities should be viewed as a continuum, not simply as present or absent.

Table 58.2–4.
Examples of Optimal Approaches to Informed Consent in Clinical Practice

---

- Provide accurate information regarding the nature of the illness and the need for clinical intervention; anticipated risks and benefits of the proposed intervention, including rare but serious risks; alternatives to the proposed treatment and their likely outcomes, including no intervention.
- Avoid using overly complicated language and overwhelming patients with too much technical information.
- Be attentive to patients' comfort, communication style, and interpersonal cues.
- Invite patients to ask questions and double-check their understanding of what was explained.
- Provide additional information if requested by patients.
- Use translators when appropriate and, if desired by patients, include family members in the information-sharing session(s).
- Supplement the information-sharing dialogue with written materials as appropriate.
- Evaluate each of the four components of decisional capacity: capacity to communicate, capacity to understand, capacity to reason, and capacity to appreciate.
- If symptoms or stressors interfere with optimal decision making, seek to ameliorate patients' discomfort to help support their ability to approach the decision process.
- Invite patients to express their preferences.
- Explore the reasons and hopes behind these preferences.
- Provide emotional reassurance and eliminate, to the extent possible, coercive pressures that patients may be feeling.
- Clarify patients' values and patterns of past personal decisions.
- Create a context of nonjudgmental support and open and ongoing communication with patients.

Many psychiatric conditions erode the cognitive strengths that underpin decisional capacities. Patients with severe and persistent mental illness may suffer episodic, fluctuating, or progressive impairments of decisional capacity. Ideally, the informed consent process consists of repeated opportunities over time to gather relevant clinical information and to discuss and clarify patients' values, preferences, informational needs, decisional abilities, and decision-making processes. Engaging with consent in a manner that is ongoing and continuous enhances patients' abilities to express greater autonomy, voluntarism, and informed decision making, and more deeply and honestly fulfills the ethical obligations of the physician.

Voluntarism is the capacity that relates to one's ability to act freely, in the absence of coercion, in expressing one's authentic wishes. “Deliberateness, purposefulness of intent, clarity, genuineness, and coherence with prior life decisions” are key elements of the capacity for voluntarism.

Roberts has conceptualized voluntarism as having four cardinal dimensions: (1) developmental factors, (2) illness-related factors, (3) psychological issues and cultural and religious values, and (4) external features and pressures. These four dimensions help to bring additional specificity and nuance to the discussion of autonomy. For example, the voluntariness of an adolescent is understood in clinical practice and in law as less than that of a legal adult except in certain circumstances. Some health conditions render affected individuals as less able to identify (e.g., in relation to alexithymia), to voice (e.g., in relation to mutism), or to enact (e.g., in relation to avolition, amotivation) their personal preferences or decisions. In some cultures, individuals are not encouraged to hold their own distinct and personal choices but instead to subordinate any individual needs to a larger group. And, finally, in some settings, such as institutional, clinical, or legal contexts, individuals may have diminished freedoms or subjectively experience themselves as having diminished freedoms, which affects individual voluntarism.

Informed consent standards are more rigorous for high-risk decisions than for other decisions. This concept is referred to as the “sliding scale” for informed consent. When patients make a choice that may endanger them (e.g., choosing one form of chemotherapy over another), the standard for decisional capacity is higher than when they make a choice that poses minimal or virtually no risk (e.g., having their blood pressure taken). Because declining recommended treatment usually entails heightened risk, the standard for informed refusal also is higher. Physicians must give extra care in providing information and counseling patients regarding alternatives and in verifying decisional and voluntarism capacity if a decision under consideration, whether accepting or refusing an intervention, is high risk.

When an individual is not capable of informed consent or refusal, others who are legally authorized may make decisions for the patient. In some cases, the patient may designate the alternative decision-maker. For young children, a parent or guardian may make the decision. At times, a formal legal process must identify a suitable decision-maker. State laws differ in how to identify decision-makers and what direction decision-makers receive, such as whether to make the decision that appears to be in the best interests of the patient or the decision that most closely resembles what the patient would choose if it were possible (i.e., “best interests” vs. “substituted judgment” approaches).

Overlapping Roles and Conflicts of Interest

Differing and overlapping roles are common for professionals who serve society as clinicians and also as teachers, consultants, administrators, and public policy experts. Having more than one role generates ethical questions and often ethical tensions because the duties of each role, by definition, do not align perfectly. (Having more than one role creates dual role, or dual agency, problems.) When roles receive compensation from different sources, issues also arise in relation to competing interests and divided loyalties. Many controversies related to conflicts of interest have become a focus of governmental policy and media reports. Ethically, conflicts of interest matter greatly in that patient needs may be subordinated to the interests of physicians who are motivated by greed rather than serving patient wellbeing. Conflicts of interest occurring at multiple levels—among individual physicians, institutions, and systems of care and research—have eroded public trust in the medical profession.

To illustrate, dual agency conflicts may arise for psychiatrists who work in the military, occupational health programs, or correctional settings. For example, psychiatrists may be called on to use their expertise to make a recommendation on a patient’s fitness for duty, competency to stand trial, or disability. As a result of the evaluation, the patient ultimately may be returned to front-line military or police work or to prison or may receive (or not) financial compensation. The ethical question of what decision serves the best interests of the patient becomes moot in these situations in which the other duties of the psychiatrist, such as protecting the public, become predominant. Particularly vexing dilemmas arise in the correctional setting, where psychiatrists may be called on to treat patients against their will when treatment would improve their disorder at the expense of their interests. Psychiatrists may not ethically force treatment on patients to make them competent to be executed. The issue is less clear concerning making patients better able to stand trial when doing so might lead them to spend more time in jail. Honesty, transparency, objectivity, and rigor—always important—are especially salient in the context of dual agency. The key issue to ethical behavior in such situations is to provide accurate information about the nature of the evaluation process, including information about the physician's responsibilities beyond those to the patient and about how usual expectations regarding patient confidentiality are not in force. Before the examination begins, physicians must notify the individuals explicitly about how they will disclose or report the information and conclusions from the evaluation.

Physicians can manage many conflicting roles and conflicts of interest in professional roles through the conscientious use of safeguards such as role separation, recusal, financial limits, responsibility limits, and additional supervision or oversight. Because public trust is at stake in issues of conflict of interest, broad dialogue and scrupulous attention to emerging guidelines and principles are very important for all fields of medicine, including psychiatry. Integrity—namely, intending good, being transparent, adhering to professional standards, and simply behaving in a manner that is worthy of trust—is key. Psychiatrists, moreover, should play an active role in educating the public on the value of professionals participating in broad activities in society, such as research, education, administration, and policymaking, in order to bring the greatest positive value of medical expertise to broader society.

Psychiatric Opinions

In 1964, Barry Goldwater was the Republican candidate for president of the United States. *Fact* magazine sent a questionnaire to U.S. psychiatrists asking whether Goldwater was psychologically fit to serve as president. Despite any direct knowledge of Goldwater derived from an in-person interview, approximately half of the psychiatrists who responded to the survey said he was not fit to serve in the role. The magazine then published an article proclaiming, "1189 Psychiatrists Say Goldwater is Psychologically Unfit to be President." In response, the APA asserted that it is unethical to render psychiatric judgments about a person in the media.

It is clearly unethical for psychiatrists to render a definitive psychiatric opinion about someone for whom they have only very limited information or only a biased source of information, such as the media. The practice appears nevertheless to be quite common. Invoking one's expertise as a psychiatrist without full knowledge of a patient is misleading and inappropriate. It is, however, ethical to render an opinion about someone one has not evaluated if one has an adequate source of unbiased information and is prudent in stating the limits of one's knowledge. In fact, psychiatrists commonly render opinions about people they have never met. When senior psychiatrists supervise residents or other mental health practitioners, they render opinions and make suggestions on treatment, although they may not have met the person in question. Psychiatrists are sometimes asked to render an opinion on testamentary capacity—whether someone who is now deceased had the capacity to make an informed decision and whether the person was subject to undue influence. Moreover, esteemed psychiatrists have studied political leaders and historical figures and thereby advanced the understanding of history and human behavior. Similarly, it is ethical for psychiatrists to reanalyze the data of a neutral forensic evaluator and make conclusions about parties they have not interviewed if there is adequate information in the neutral forensic evaluator's report to do so.

Thus, the ethics of rendering an opinion about a person whom one has not examined depends on several factors. Key issues include the purpose of rendering the opinion, the quality of information, and the reason for not interviewing the person.

Responsibilities to Third Parties

Several court rulings have stated that psychiatrists have a responsibility to third parties. These concerns are more properly forensic in nature but are relevant to good ethical practices because of the ethical notion of respect for the law. Court rulings create case law, not ethics. Nevertheless, case law leads to new expectations and indicates society's view of professional responsibilities. Legal rulings also have value because they point ethicists' attention to issues that need deeper consideration. In the original and subsequent *Tarasoff* decisions, for example, the courts upheld the position that psychiatrists must act overtly to protect the community. Courts have held psychiatrists liable when their patients, or former patients, have injured or killed people with guns or by reckless driving (e.g., 1983 *Jablonski v. US;* 1988 *Naidu v. Laird;* 1980 *Lipari v. Sears, Roebuck and Co.;* 1983 *Petersen v. Washington*). Subsequent legislative rulings have limited psychiatrists' liability in such matters. Psychiatrists who provide incorrect information place themselves at risk if such action harms the interests of a third party (e.g., *Molien v. Kaiser Foundation Hospitals* and *Ramona v. Isabella*). When clinicians give direct suggestions that affect third parties (such as the advice to get a divorce) rather than helping patients evaluate their choices in the context of their lives, clinicians leave themselves liable for the consequences. The liability is increased if they become directly involved by participating in a confrontation with the third party.

Self-Care and Impairment

Stress, distress, and illness can be occupational hazards of the work of psychiatrists and other physicians. Beyond these concerns, general mental health and wellbeing concerns are increasingly prevalent among medical students, residents, and practicing clinicians, as is the phenomenon of professional "burnout." Burnout has been defined as a syndrome tied to chronic workplace stress involving emotional exhaustion, depersonalization, and lack of personal accomplishment. Extreme stress can cause significant mental health concerns and challenge personal relationships. Health care professionals are also at risk for making errors in patient care when under extreme stress. For these reasons, the imperative to engage in self-care has arisen as a key ethical commitment of health professionals. Individuals who identify as nonmajority, who have greater financial debt, or who are experiencing health problems may also experience greater stress in their professional work. In all of these situations, there can be a "double bind" in which the clinician is responsible for "self-care," and yet the factors that lead to burnout are outside of the control of the individual. For these reasons, self-care alone is not the answer for eliminating stress, distress, and related illness conditions. Greater equity, flexibility, and support in the workplace and in the broader culture of medicine have been identified as constructive approaches that can help diminish the emergence and perpetuation of burnout among health professionals.

As noted by the AMA's *Principles of Medical Ethics*, professionals have a duty to report colleague misconduct and impairment. Misconduct is defined as professional acts that violate ethical standards, while impairment is the inability to practice medicine due to physical or mental illness or deterioration or the excessive use of addictive substances, such as alcohol. Both may result in harm toward patients, colleagues, or others and should be guarded against, in keeping with the ethical principle of nonmaleficence.

A perceived lapse in professionalism may or may not constitute misconduct, depending on whether or not an ethical standard is violated. Professionalism is a fluid concept informed by cultural and societal mores. Recently, physicians and trainees have drawn attention to the ways in which professionalism is weaponized against individuals who are perceived as being nonmajority or different. For example, women trainees and faculty perceive that more "necessary behavior" is required for them to gain recognition in the workplace. Trainees and physicians of color experience real harm in the form of microaggressions or disciplinary actions when, for example, their dress or appearance is seen as nonconforming to white norms. Professionalism requires rigorous efforts to earn the trust of all stakeholders, including colleagues and patients.

ETHICS AS FUNDAMENTAL TO MEDICAL PROFESSIONALISM

Service to society is the *raison d'être* of the professions. In medicine, physicians provide expertise to advance human health and to lessen the burden of suffering associated with illness. Medical professionals, including those in the field of psychiatry, have the opportunity to develop distinct expertise to help others. These professionals have the responsibility for creating and applying new knowledge, training others, and participating in leadership throughout the profession and broader society. These professionals also have the duty to ensure that colleagues uphold the clinical and ethical standards of their disciplines. Conscientious self-care is increasingly recognized as an ethical responsibility for health professionals, as is vigilance to avoid emotional and physical exhaustion and to address emerging health issues. The role of injustice and inequity in society in health disparities is a central concern of the field of psychiatry, especially in light of the potential vulnerability of people living with mental illness.

FURTHER READINGS

American Medical Association. *Code of Medical Ethics Chicago*. American Medical Association; 2021. Accessed April 22, 2021. https://www.ama-assn.org/topics/ama-code-medical-ethics

American Psychiatric Association. *The Principles of Medical Ethics with Annotations Especially Applicable to Psychiatry 2013 Edition*. American Psychiatric Association; 2013. Accessed April 22, 2021. https://www.psychiatry.org/psychiatrists/practice/ethics

Appelbaum PS, Grisso T. Assessing patients' capacities to consent to treatment. *N Engl J Med*. 1988;319:1635–1638.

Baker RB, McCullough LB. *The Cambridge World History of Medical Ethics*. Cambridge University Press; 2008.

Beauchamp TL, Childress JF. *Principles of Biomedical Ethics*. 8th ed. Oxford University Press; 2019.

Bloch S, Green SA*. Psychiatric Ethics*. 5th ed. Oxford University Press; 2021.

Char DS, Shah NH, Magnus D. Implementing machine learning in health care: addressing ethical challenges. *N Engl J Med*. 2018;378:981–983.

Epstein RS, Simon RI. The Exploitation Index: an early warning indicator of boundary violations in psychotherapy. *Bull Menninger Clin*. 1990;54(4):450–465.

FitzGerald C, Hurst S. Implicit bias in healthcare professionals: a systematic review. *BMC Medical Ethics*. 2017;18:19.

Gelso CJ, Hayes JA. *Countertransference and the Therapist's Inner Experience: Perils and Possibilities*. Lawrence Erlbaum Associates, Inc.; 2007.

Gengoux G, Zack SE, Derenne JL, Robinson A, Dunn LB, Roberts LW. *Professional Well-Being*. American Psychiatric Association Publishing; 2020.

Gilligan C. *In A Different Voice: Psychological Theory and Women's Development*. Harvard University Press; 1982.

Gregory J. *Lectures on the Duties and Qualifications of a Physician*. W. Strahan and T. Cadell; 1772. Reprinted in McCullough LB: John Gregory's Writings on Medical Ethics and Philosophy of Medicine. Dordrecht: Kluwer; 1998.

Jonsen AR, Siegler M, Winslade WJ. *Clinical Ethics: A Practical Approach to Ethical Decisions in Clinical Medicine*. 8th ed. The McGraw-Hill Companies, Inc.; 2015.

Kant I. *Fundamental Principles of the Metaphysics of Morals [1785]*. Prometheus Books; 1990.

Lo B. *Resolving Ethical Dilemmas: A Guide for Clinicians*, 6th ed. Lippincott Williams & Wilkins; 2019.

Mill JS. *Utilitarianism [1861]*. Everyman's Library; 1996.

National Institutes of Health. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*; 1979. Accessed April 22, 2021. http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

Percival T. Medical Ethics: Or a Code of Institutes and Precepts, Adapted to the Professional Conduct of Physicians and Surgeons. J. Johnson & R. Bickerstaff; 1803.

Rawls J. *Theory of Justice*. Harvard University Press; 1971/1999.
Roberts LW. Informed Consent and the Capacity for Voluntarism. *Amer J Psychiatry*. 2002;159(5):705–712.
Roberts LW. *A Clinical Guide to Psychiatric Ethics*. American Psychiatric Association Publishing; 2016.
Roberts LW, Termuehlen G. *Professionalism and Ethics: Q&A Self-Study Guide for Mental Health Professionals*. 2nd ed. American Psychiatric Publishing, Inc.; 2022.
Roberts LW, Reicherter D. *Professionalism and Ethics in Medicine: A Study Guide for Physicians and Physicians-in-Training*. Springer Science+Business Media, LLC; 2015.
Sabin JE, Harland JC. Professional ethics for digital age psychiatry: boundaries, privacy, and communication. *Curr Psychiatry Rep*. 2017:19:55.
Shenoy A, Appel JM. Safeguarding confidentiality in electronic health records. *Camb Q Healthc Ethics*. 2017:26:337–341.
Shim R, Vinson S. *Social (In)Justice and Mental Health*. American Psychiatric Association Publishing; 2020.
Sperry L. *Dictionary of Ethical and Legal Terms and Issues: The Essential Guide for Mental Health Professionals*. Routledge; 2007.
Torous J, Roberts LW. The ethical use of mobile health technology in clinical psychiatry. *J Nerv Ment Dis*. 2017;205:4–8.
US Department of Health and Human Services. *Code of Federal Regulations, Title 45, Public Welfare, Part 46, Protection of Human Subjects*. *Revised*, 2018. Accessed April 22, 2021. http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html

58.3 Correctional Psychiatry

ARYA SHAH, M.D., AND JHILAM BISWAS, M.D.

Correctional psychiatry functions at the boundaries of law and medicine and requires an understanding of ethics and the myriad ways that law impacts the practice of psychiatry. As a subset of forensic psychiatry, correctional work requires being aware of issues of ethics and morality and can be fraught with potential bias and conflicts of interest. Given a large portion of the psychiatric population interacts with correctional settings, psychiatrists benefit greatly from learning about how corrections and law enforcement engage with inmates.

Today, certain county jails in large cities are considered the largest psychiatric hospitals in the United States. Multiple factors, including legal, social, fiscal/health care reimbursement, and political, over the past 70 years have led to the current substantial presence of psychiatric and substance use disorders seen within American jails and prisons.

Some examples of these factors include reducing beds at state inpatient facilities, “zero tolerance” arrest, sentencing policies as part of the “war on drugs,” and the deinstitutionalization of many individuals with chronic and persistent mental illness without adequate community resources. The unintended consequence of these national policies is that it has criminalized mental illness.

Mental and substance-related disorders, as well as chronic medical conditions, are prevalent public health problems affecting individuals in U.S. correctional facilities, and these settings have become the default mental health system for high-risk populations. The sheer number of inmates with severe psychiatric disorders in the U.S. prison system and the “revolving door phenomena” in which many mentally ill people move continuously between homelessness and the criminal justice system represents a national public health crisis. This risk of reincarceration also exists among prisoners with co-occurring severe mental illness and substance use disorders.

In the postpandemic era, equitable health care among races is a key area of correctional psychiatry and must be addressed in the treatment, research, and advocacy of this patient population. Incarceration disproportionately affects Black Americans, and the data shows that living in a predominantly brown and black community heightens the risk a person will interact with law enforcement and corrections. Providing high-quality mental health care in corrections may be one way to address the inequities faced by disenfranchised minorities and a core area of criminal justice reform.

Over the course of the last decade, the imprisoned population has slowly come down from its height in 2010, but data shows the United States continues to hold a quarter of the world’s incarcerated population. The COVID-19 pandemic did cause a drop in incarceration rates from 2.1 to 1.8 million. While this has been a positive development, many argue the high reincarceration rate will again increase numbers if policies stay the same. Among the incarcerated population, around 20%, or an estimated 356,000, have a severe and persistent mental illness, that is, carry a diagnosis of a serious mental disorder such as schizophrenia, bipolar disorder, or severe depression.

Notably, not all correctional settings are the same, and there are many types. These include lockups, jails, state and federal prisons, U.S. Immigration and Customs Enforcement (ICE) detention centers, Bureau of Indian Affairs detention centers, U.S. military prisons, jails, and detention centers. There have been recent concerns regarding access to care and other medicolegal issues for persons with severe mental illness in the U.S. ICE system, which is discussed elsewhere in the literature.

The most common correctional settings for adults are jails and prisons. Jails are city- or county-operated correctional facilities that confine individuals involved in the criminal justice system who are awaiting trial or serving short sentences for misdemeanors. There are approximately 3,500 jails in the United States. Detention refers to the period following arrest in which an individual is held in secure custody prior to or after court proceedings. The amount of time a person is in jail varies (e.g., hours, days, weeks, typically 1 year or less). Some individuals may remain in jail for years for more serious charges while awaiting legal disposition of their case, or they may serve their sentence in a jail setting. There is significant variability in size, with some small rural jails holding 10 to 50 detainees and some major metropolitan jails holding 7,500 to 10,000 detainees.

A prison is a facility operated under the auspices of a state or federal government for the confinement of adults convicted of a felony whose sentences generally exceed 2 years. Individuals in prisons may be serving one or more life sentences, and some states have the death penalty. There are approximately 1,300 state and federal prisons in the United States.

Similar to adults, youth in juvenile justice settings face significant mental and co-occurring substance-related disorders, which represent a serious public health concern, and the need for services in the juvenile justice system is high. While involved with the juvenile justice system and family courts, youth may be housed within juvenile detention centers, reception or diagnostic centers, community-based or other residential treatment programs, or juvenile correctional facilities. Some juveniles may be held in county jails, and some youth may be transferred/sentenced to adult prison settings. In many states, certain felony offenses committed by adolescents are moved to adult criminal court. This practice has led to an increase in the detention and incarceration of younger offenders in adult jails and prisons. Data reflect rates of SMI among incarcerated adolescents that are substantially above those found in the community.

This chapter will present an overview of American correctional systems and provide background on the demographic and mental health landscape of the correctional population. The chapter will identify ongoing practical, legal, and ethical challenges that a psychiatrist may face while providing care in correctional settings. Finally, the chapter will also aim to make recommendations for psychiatric assessment and treatment in the correctional setting, discuss the roles of mental health clinicians in corrections, and identify phenomena that might influence the future of correctional psychiatric practice in the United States.

PSYCHIATRIC PATIENTS IN CORRECTIONAL FACILITIES

Mental Illness in the Correctional System

The importance of demographic studies of psychiatric disorders in correctional facilities cannot be overemphasized. For advocacy purposes and for public health considerations, the results of this research dramatically portray the critical need for psychiatric services in jails and prisons.

However, the true incidence and prevalence of psychiatric disorders in correctional facilities are difficult to obtain. Well-designed, carefully done epidemiologic studies are costly and complex to carry out, and because of the high inmate turnover rates, doing a point prevalence study of mental illness in jails is difficult. The best way to detect the incidence and prevalence of psychiatric problems in jails is to find cases on admission. Prison inmates stay for longer terms, and conducting a point-prevalence epidemiologic study of mental illness in this setting is more feasible. Despite these difficulties and limitations, the epidemiology of mental illness in correctional facilities is gradually becoming known. In a 2017 special report on mental health problems in prisoners and jail inmates, the Bureau of Justice Statistics estimated that 37% of prisoners and 44% of inmates had been diagnosed with a mental illness by a mental health professional. In general, most authorities now estimate that about 20% to 40% of the correctional population has serious and persistent mental illnesses. In 2010, the Treatment Advocacy center issued a report identifying that jails and prisons housed more than three times as many individuals with serious mental illness than did hospitals. Thus, in 2012, more than 350,000 seriously mentally ill offenders were incarcerated in the nation's jails and prisons. (This is contrasted to an estimated 35,000 patients in state psychiatric hospitals.)

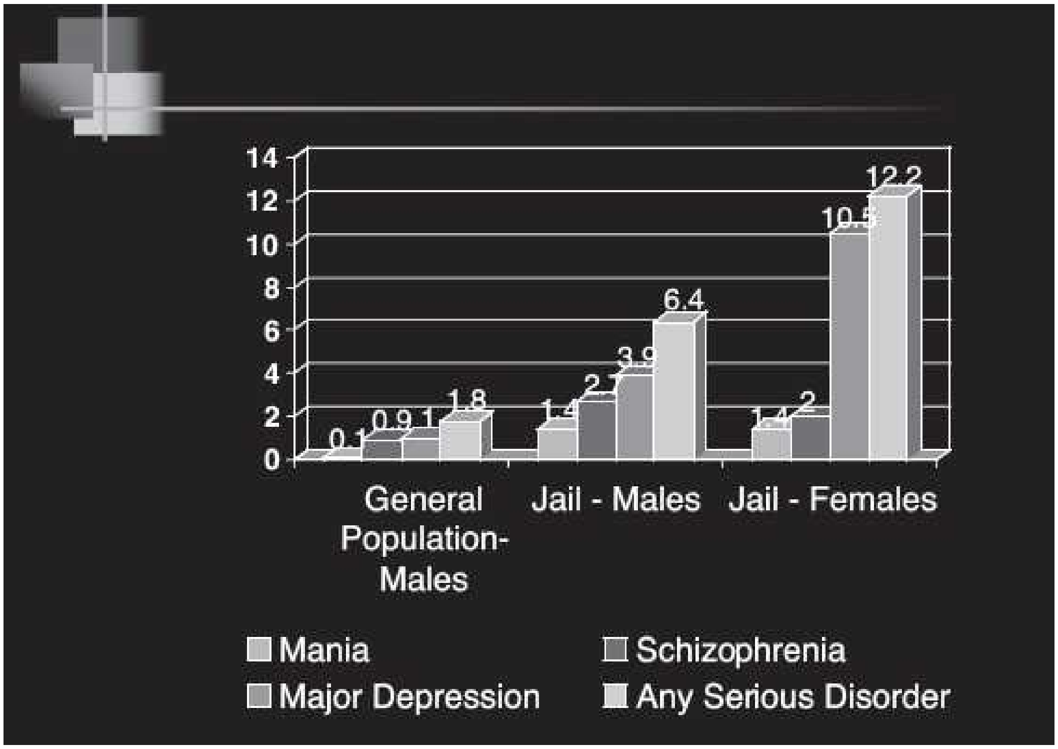

FIGURE 58.3–1. Serious mental disorders among the general population and among jail admissions.

Jails. With regard to jails, the few available epidemiologic jail studies highlight that after controlling for demographic differences between jail samples and the epidemiologic catchment area (ECA) samples, jail prevalence rates of mental illness are two to three times higher than those in the general population (Fig. 58.3–1).

Prisons. Because prison inmates stay for longer terms, there are more prison population-based epidemiologic studies of mental illness than there are in jails. The ECA study found that 6.7% of inmates in prisons had experienced schizophrenia at some point, compared to 1.4% of the U.S. household population. A number of studies reveal that, at any given time, a little more than 2% of the inmate prison population is in inpatient or residential psychiatric care.

More than 30% of male inmates and 78% of female inmates with mental illness reported prior physical or sexual abuse, and inmates with mental illness in correctional facilities are more likely than other offenders to have a higher prevalence of homelessness, unemployment, alcohol and drug abuse, and physical and sexual abuse before their current incarceration.

HISTORY OF CORRECTIONAL PSYCHIATRY IN THE UNITED STATES

Until recently, the judicial system maintained a “hands-off” policy with regard to conditions in jails and prisons, a doctrine that regarded the inner workings of correctional facilities as beyond the responsibility and authority of the courts and best left to correctional professionals.

It was during the 1970s, during the tidal wave of the civil rights movement, that various studies and court cases exposed the abominable conditions of health and mental health care in the jails and prisons of the United States. These findings led to significant reform of the correctional system and of correctional psychiatry and health care. The confluence of a number of interrelated forces spearheaded this reform, including professional organizational advocacy by the American Medical Association (AMA) and the American Psychiatric Association (APA), “conditions of confinement” lawsuits, which were now entertained by the courts, and the development of standards and guidelines for health and mental health care in jails and prisons.

Table 58.3–1.
American Psychiatric Association Position Statement on Psychiatric Services in Adult Correctional Institutions (2021)

The American Psychiatric Association Position on mental health services for individuals in adult incarcerated settings should minimally include:

1. Timely and effective access to mental health services screening, evaluation, and treatment, including an appropriate range of therapeutic interventions.
2. Necessary staffing levels to ensure access to care and to provide appropriate services.
3. Quality assurance and improvement reviews to identify and resolve inadequacies of care.
4. As therapeutic an environment as possible, optimized by support, guidance, and advocacy from mental health leadership and collaboration with correctional leadership. Mental health administration should be incorporated into the overall management of the facility.
5. Proper procedures, based on appropriate legal guidance, and staffing to safely administer involuntary medications when clinically indicated and when reasonable attempts at voluntary engagement in medication adherence have not been successful. This should not preclude psychiatric hospitalization if available and clinically indicated.
6. A balance and an effective partnership between security and treatment needs to ensure the effective delivery of mental health services.

Reprinted with permission from the Position Statement on Psychiatric Services in Adult Correctional Facilities (Approved by the Board of Trustees, December 2021). American Psychiatric Association. All Rights Reserved.

In 1974, the AMA carried out a survey of sheriffs administering local jails, which revealed that, in addition to frequently lacking mental health screening and testing resources, fewer than 20% of jails had facilities for caring for individuals with alcohol use disorder, only 10% had any facilities for inmates with drug addiction, and only 14% had facilities for inmates with mental illness. Studies of prison correctional health care revealed that prison correctional health care was no better than that in jails. The APA subsequently published a Position Statement on providing psychiatric care in correctional settings, which was prepared by the APA Committee on Psychiatry and the Law, with the most recent version published in 2018 (Table 58.3–1). In 1975, the Federal Law Enforcement Assistance Administration (LEAA) provided a grant to the AMA to upgrade correctional health care. The purpose of the grant was to develop a model correctional health care system and to distribute information on correctional health care issues. Another important goal was the development of a national accreditation program with correctional health care standards. In 1977, the AMA published its first jail correctional health standards; 16 jails were accredited for compliance that year. In 1979, the prison and juvenile correctional health care standards were published.

At the same time, the judicial system was establishing the unique legal context of psychiatric services in jails and prisons—that such services are constitutionally guaranteed. In 1976, the U.S. Supreme Court, in the case of *Estelle v. Gamble*, held it unconstitutional for prison officials to be deliberately indifferent to the serious medical needs of those in their custody, and in 1977, a federal court, in the case of *Bowring v. Godwin*, extended this requirement to psychiatric care.

In 1983, as an outgrowth of the LEAA-funded AMA program, the National Commission on Correctional Health Care (NCCHC) was established. The NCCHC consists of 37 professional organizations, including the AMA and the APA. It assists correctional institutions in upgrading their correctional health care services and, most importantly, has established correctional health care standards and correctional health care accreditation processes. These standards, which are revised regularly, represent the state of the art for correctional health care, and with their recently revised 2018 publication of the NCCHC's Standards for Mental Health Services in Correctional Facilities, the state of the art, as well as evidence-based practice, for mental health services in jails and prisons.

In 1989, the APA published Psychiatric Services in Jails and Prisons. This document, referred to as the APA Guidelines, was developed by correctional psychiatrists for correctional psychiatrists and described in detail the principles governing the delivery of psychiatric services in jails and prisons; the specific requirements for an adequate process of screening, evaluation, and treatment; the specific roles of the correctional psychiatrists; and the application of the principles and guidelines to specific populations. The second edition of the APA Guidelines was published in 2000; the third edition, now titled *Psychiatric Services in Correctional Facilities*, was published in 2016.

PERSISTING CHALLENGES IN CORRECTIONAL PSYCHIATRY

Despite the significant burden of psychiatric needs in correctional settings, there exist significant gaps and shortcomings in care within jails and prisons.

Shortage of Correctional Psychiatrists

The National Health Service Core (NHSC) maintains a database of correctional institutions that are designated Health Professional Shortage Area (HPSA) facilities that have a minimum of 250 inmates and a ratio of psychiatrists to inmates of at least 1:2,000. More than 450 correctional facilities nationally have been found to meet the criteria for Health Professional shortages in the database, with such figures likely representing an underestimate of the true count of correctional facilities with a shortage in psychiatric providers. Consequences of such gaps in care have been called to public attention; in the more recent landmark case of *Brown v Plata,* the Supreme Court recognized that as many as 72% of suicides in correctional settings likely related to inadequacies in mental health care, with wait time for receiving care as long as 12 months (Supreme Court Ruling 536, 2011).

The lack of adequate numbers of clinicians in correctional settings may be related to multiple factors. Currently, forensic exposure is sporadic in residency training. This lack of exposure to psychiatric practice in corrections can result in graduates feeling ill-equipped to practice in forensic settings. Furthermore, misconceptions related to justice-involved individuals receiving treatment, apprehension regarding provider safety, and belief that clinical care may be less fulfilling in correctional settings may all contribute to the lack of psychiatric clinicians choosing to build careers in the field. Exposure to forensic psychiatry in training has been proposed as a solution for countering ambivalence and increasing provider comfort with providing care in these settings.

Overcrowding

In 2006, more than 7.2 million people were on probation, in jail or prison, or on parole. At year-end 2006, this represented 3.2% of all U.S. adult residents or one in every 31 adults. State and federal prison authorities had jurisdiction over 1,570,861 inmates at year-end 2006: 1,377,815 in State jurisdiction and 193,046 in Federal jurisdiction. Local jails held 766,010 persons awaiting trial or serving a sentence at year-end 2006. An additional 60,222 persons under jail supervision were serving their sentence in the community.

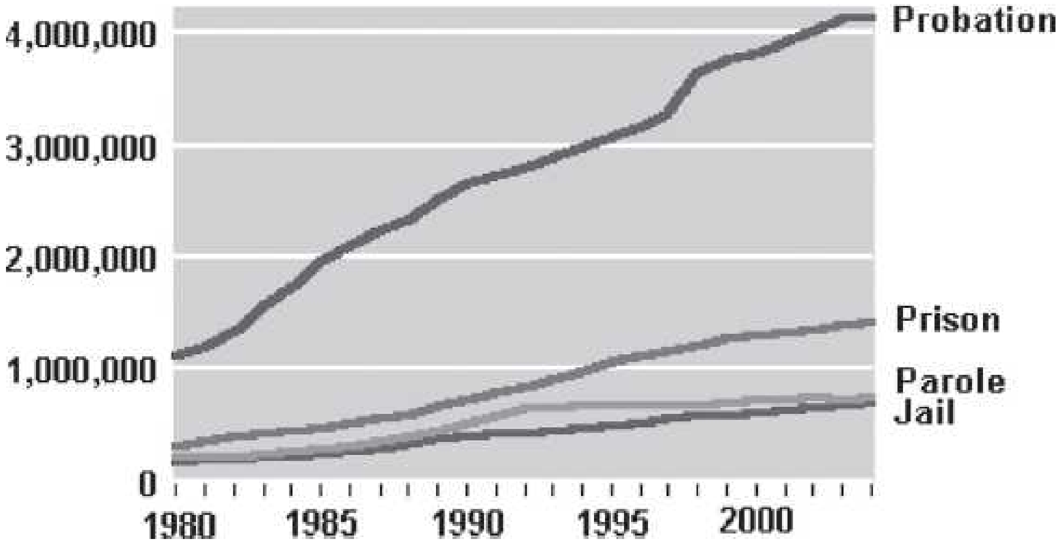

FIGURE 58.3–2. Adult correctional populations from 1980 to 2004.

The overcrowding that has resulted from the increase in the correctional population over the last three decades has made many large jails, such as the Cook County Jail in Chicago, the Los Angeles County Jail, and Harris County Jail in Houston, Texas, the country's largest mental institutions (Fig. 58.3–2).

Multiple historical factors have contributed to the current problem of overcrowding and over-representation of mental illness in our county's jails and prisons. With the discovery of chlorpromazine and its antipsychotic properties, there was growing optimism about the possibility of community-based care for individuals with serious mental illness who had previously been institutionalized. The phenomenon of deinstitutionalization in the 1960s involved the closure of state hospitals in hopes of transitioning psychiatric care into community-based mental health treatment programs. However, a variety of social, financial, and legislative factors ultimately contributed to inadequate funding of community mental health treatment centers, resulting in a system that was ill-equipped to manage the mental health needs of individuals requiring access to care. The combination of untreated mental illness and lack of access to care in the community have since contributed to the described “revolving door” of incarceration faced by many individuals with mental illness, contributing to the phenomena of “transinstitutionalization” and the over-representation of mental illness within our country’s justice system (Fig. 58.3–3).

The Criminalization of the Mentally Ill

Some of the factors contributing to the high rates of people with serious mental illness in the criminal justice system include, as reasons for the increased entry, the process of deinstitutionalization and its accompanying phenomenon of “transinstitutionalization,” the inadequate access to the seriously and persistently mentally ill to community treatment, the fact that these patients are arrested at disproportionately higher rates and often have co-occurring substance-related disorders, the prejudice that the mentally ill are more violent than other persons, and the use of jails and prisons as housing of last resort.

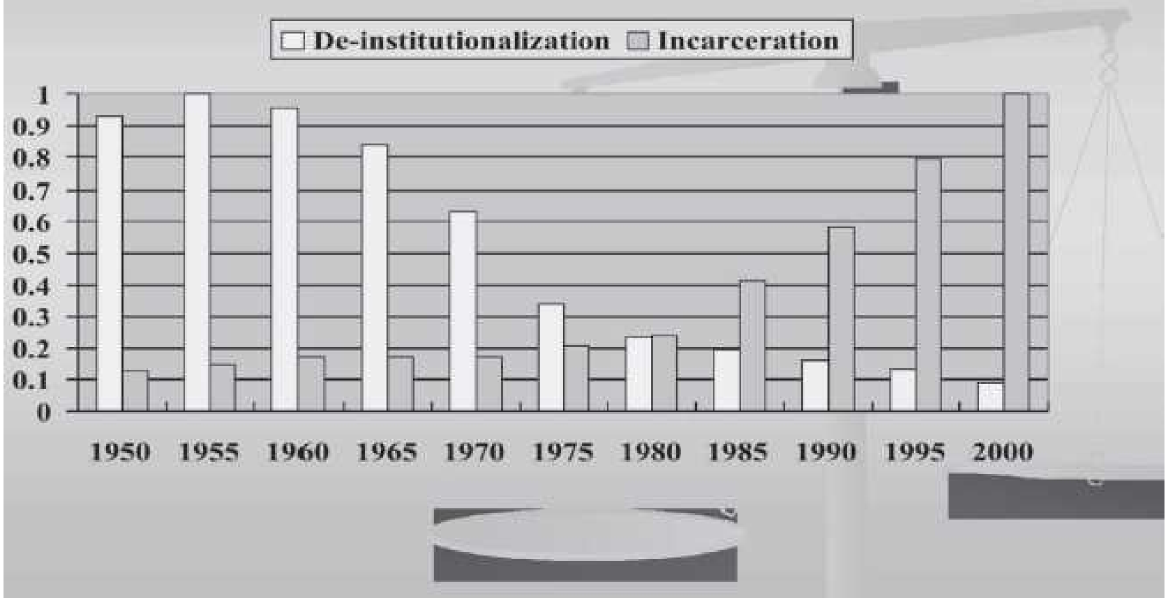

FIGURE 58.3–3. Transinstitutionalization from 1950 to 2000.

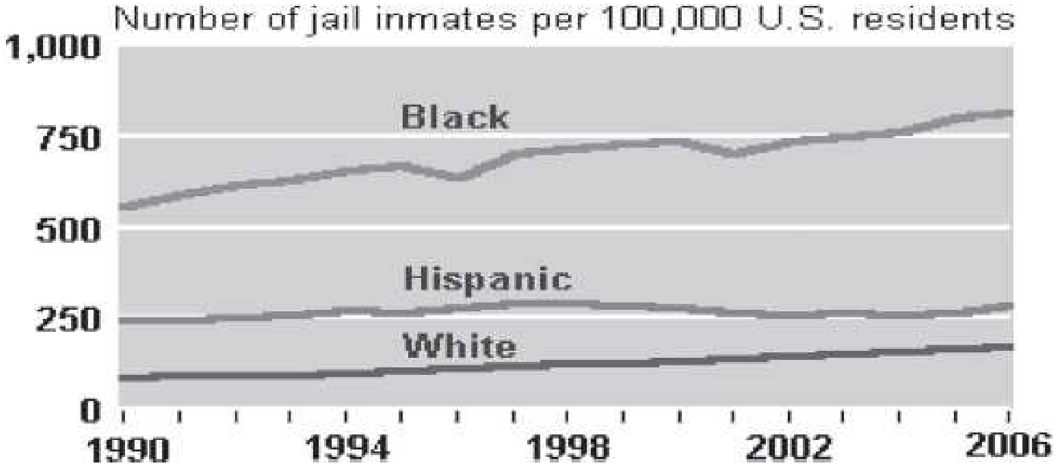

FIGURE 58.3–4. Jail incarceration rates by race and ethnicity, 1990 to 2006.

There are additional problems with discharge planning and reentry into the community, including the fact that these patients spend long periods of time incarcerated, the pathogenic nature of incarcerated environments, the lower rates of parole, and the higher rates of recidivism due to inadequate community mental health services.

The United States has the largest percentage of its population incarcerated than any other country in the world. Since 1989, the National Drug Control Strategy, which called for a mandatory minimum sentence for drug crimes, caused the jail and prison populations to grow at a rapid rate. In 1985, the percentage of sentenced drug offenders was approximately 13%. Currently, it is more than 30%. The importance of the relationship between the increase in the number of inmates incarcerated for drug offenses and the disproportionate increases in the numbers of female, nonwhite, and foreign-born inmates cannot be overstated.

Racial Inequity

In addition to the well-recognized phenomenon of mental illness being over-represented in the correctional system, racial disparities among individuals housed in jails and prisons have been viewed as evidence of significant care inequity (Fig. 58.3–4).

While the presence of mental illness alone is a risk factor for justice involvement and incarceration, disparities in access to community treatment may contribute to additional racial inequities. A report from the Office of the Surgeon General identified that racial and ethnic minorities not only have less access to community mental health treatment than do white individuals but that even when received, the care is of poor quality. It is possible that such disparities in the community influence the disproportionate number of people of color in the correctional system. Despite making up 13% of the general population, African Americans make up 40% of individuals who are currently incarcerated. Furthermore, there is evidence that individuals of color may be less likely to receive mental health treatment within jails and prisons, with mental health symptoms more likely to be viewed as criminal and to be met with punitive action. Such phenomena highlight the importance of the continued study of structural barriers to equitable access to treatment and enhanced recognition and treatment of mental illness in populations who have historically had inequitable access to care.

THE LEGAL AND ETHICAL PRINCIPLES UNDERLYING THE PRACTICE OF CORRECTIONAL PSYCHIATRY

Landmark Cases in Correctional Psychiatry

In the 1970s, the crucial years of the reform of the correctional system, a series of court decisions established the principle that psychiatric services in the jails and prisons of the United States are constitutionally guaranteed. This legal doctrine holds that under the Eighth Amendment of the U.S. Constitution, which prohibits “cruel and unusual punishment,” whenever a state or the federal governmental entity takes custody of a person, it must provide for the necessities of life that the person otherwise is unable to obtain. These include food, clothing, shelter, and medical care, which include psychiatric care.

In 1976, the Supreme Court, in the case of *Estelle v. Gamble*, held it unconstitutional for prison officials to be deliberately indifferent to the serious medical needs of those in their custody, and in 1977, the federal court case of *Bowring v. Godwin* extended this requirement to psychiatric care. In 1980, another federal court case, *Ruiz v. Estelle*, listed a number of specific components of mental health care components that are required in correctional facilities, including but not limited to the involvement of trained mental health professionals, utilization of systematic screening and evaluation, adequate monitoring of methods and amounts of medications administered, maintenance of a confidential medical record, and implementation of a suicide prevention program.

In 1989, the case of *Langley v. Coughlin* listed a number of other specific criteria that, if not met, could provide a basis for successful inmate legal claims. Many of these criteria explicitly required providing for so-called secondary or supportive rights, such as taking psychiatric histories and maintaining medical records, in recognition of the fact that such rights are required to facilitate the primary right to diagnosis and actual care. *Langley v. Coughlin* also explicitly requires that care not be limited to psychotropic medication.

Madrid v. Gomez, in 1995, a State of California Department of Corrections legal case, added several factors that may determine the constitutionality of a correctional mental health system. These included that inmates should be able to communicate needs to medical staff, have access to individualized treatment, and reserve the right to a hearing and procedural protections prior to involuntary treatment; furthermore, institutions were required to implement symptoms of quality assurance and hire staff that are competent and trained in the needs of those with mental illness within the correctional system.

In addition to these constitutional mandates, jail and prison mental health services may be required to meet other legal requirements, such as those associated with federal laws (such as the Americans with Disabilities Act), state laws, and other regulations related to liability for medical malpractice.

Mental health in correctional facilities is considered a constitutional guarantee. A lack of it is considered a violation of the Eighth Amendment of cruel and unusual punishment as determined by *Estelle v. Gamble*. However, substance use is variably included as part of mental health and often depends on local courts and governments. The position of the courts has been that the provision of care and treatment in cases of substance use disorder (e.g., for drug abuse) in jails and prisons is not considered an aspect of psychiatric care and, therefore, is not considered constitutionally required. However, there is such an obvious need for the treatment of substance use disorders in jails and prisons that it is usually included in correctional programs.

Ethical Considerations of Correctional Psychiatry

Correctional mental health providers should follow the standards of professional ethics as set out in the APA's Annotations Especially Applicable to Psychiatry to the AMA's Principles of Medical Ethics.

Although the Preamble to the AMA's Principles states, "A physician must recognize responsibility to patients first and foremost," it then adds the words "as well as to society." Correctional psychiatric practice presents unique challenges related to how the psychiatrist's duties to the patient may clash with their duties to the department of correction, which can be regarded as an aspect of the correctional psychiatrist's duties to "society."

Ethical conflicts may manifest in various features of mental health care in correctional settings. In the correctional setting, there is a conflict between the two distinct roles of the correctional psychiatrist: As the inmate's treating physician and as an employee and agent of a local, state, or private correctional agency. This problem is known as that of "dual loyalty" (also called the "the double-agent problem"). Therefore, while the concepts of autonomy and nonmaleficence are central tenants to the practice of medicine, both of these values may be at odds with correctional perspectives and policies, and the correctional environment might threaten to coerce professional judgment and actions into meeting security needs rather than the mental health needs of individual inmates. Such conflicts can affect the doctor–patient relationship, and the unique power dynamics existing within the correctional setting necessitate that mental health providers consistently monitor unintended identification with correctional authorities and countertransference with inmates. The APA has set forth in its guidelines that the inherently coercive settings of jails and prisons might create additional difficulties for the correctional psychiatrist in regard to obtaining informed consent. It is interesting to note that one legal case stated that it is impossible to obtain true informed consent in this setting.

Confidentiality also presents unique considerations in the correctional setting. Although the usual principles of psychiatric confidentiality apply to the delivery of psychiatric services in jails and prisons, there are special needs presented by these settings, especially in relation to security and safety. The APA Guidelines state, "The rights of the patient to . . . confidentiality must be weighed against the needs of other inmates as well as the institutional needs of safety and security." Furthermore, the limitations of privacy in correctional settings have been acknowledged, given the unique organizational structure within the facility. While the NCCHC Standard on "Privacy of Care" in corrections requires that "mental health services are conducted in private and carried out in a manner designed to encourage the patient's subsequent use of services . . . in private, without being observed or overheard by staff or inmates," the standard goes on to acknowledge the limits of privacy with imminent risk is of concern, as well as acknowledging that "in corrections mental health triage commonly occurs at cell-side. As the triage discussion becomes more involved than a simple triage, the clinician should be aware of the need to maintain privacy and move the patient to a clinical setting."

It is critical for psychiatrists and other mental health staff working in correctional settings to define and maintain their role as a clinician, working in the best interests of the patient, as opposed to a forensic evaluator, or as an agent of the court, law enforcement, or of the state or local jurisdiction. Without a release of information, treating psychiatrists should avoid verbal or written communication with attorneys and other court personnel, and they should avoid inappropriate communication with the media. It is also recommended that correctional psychiatrists specify clearly any limits on the usual precepts of confidentiality before initiating treatment, except in emergencies. However, this notice may be presented to the inmate in general terms, such as applying to situations in which the patient or the institution may be at risk.

On the contrary, the correctional psychiatrist must be able to share relevant confidential information with facility administrators (on a "need-to-know" basis) when it relates to the medical management of a patient, security issues, or a patient's ability to participate in programs. Similarly, it is appropriate that there be communication with nonclinical staff regarding an inmate's significant mental health needs that must be considered in classification decisions to preserve the health and safety of that inmate, other inmates, or staff.

The correctional psychiatrist should be sensitive to the fact that confidentiality is especially important to those for whom the lack of confidentiality might have serious legal and other consequences, such as those detainees with pending legal charges. Some practical suggestions for psychiatrists and other mental health staff might include the avoidance of exploration into the details or circumstances of the alleged criminal act(s) and the individual's state of mind, criminal intent, mitigating factors, or other defense strategies. Similarly, at risk are those who have experienced sexual violence, sex offenders, or persons with HIV/acquired immunodeficiency syndrome (AIDS).

Collecting Forensic Psychiatry Information in a Jail or Prison

The collection and use of forensic information by the correctional psychiatrist present a unique clinical and ethical challenge for mental health clinicians in this setting. The APA Guidelines set out the "distinction [that] must be drawn between information obtained by a correctional psychiatrist in the course of treatment, and information obtained from the inmate in the course of an evaluation for non-treatment purposes."

When correctional authorities ask correctional psychiatrists to collect information for forensic purposes, the NCCHC standards support the refusal of the correctional psychiatrist to do evaluations of inmates for use in adversarial proceedings or to order blood tests to determine drug or alcohol use (unless it is for clinical reasons). These matters are of such concern to the NCCHC that it has promulgated an explicit standard prohibiting health care workers from collecting "forensic information." The standard on Forensic Information in the Standards for Mental Health Services in Correctional Facilities states, "Mental health staff are prohibited from participating in the collection of forensic information," which is defined as "physical or psychological data collected from an inmate that may be used against him or her in disciplinary or legal proceedings." The intent of the standard is to ensure the preservation of the treatment relationship between the provider and the patient.

The NCCHC standard goes on to state that "where state laws and regulations require that such acts be performed by health/mental health staff, the services of outside providers or someone on the institution's staff who is not in a therapeutic relationship with the inmate is obtained." Relevant also are the Ethics Guidelines for the Practice of Forensic Psychiatry of the American Academy of Psychiatry and the Law (AAPL), which deal with this issue explicitly, noting, "When a patient is involved in parole, probation, conditional release, or in other custodial or mandatory settings, psychiatrists should be clear about limitations on confidentiality in the treatment relationship and ensure that these limitations are communicated to the patient."

PSYCHIATRIC TREATMENT IN JAILS AND PRISONS

The Policy Goal of Psychiatric Treatment in Jails and Prisons

According to the APA Psychiatric Services in Correctional Facilities (Third Edition):

The fundamental policy goal for correctional mental health care is to provide the same quality of mental health services to each patient in the criminal justice system that should be available in the community. This policy goal is deliberately higher than the "community standard" that is called for in various legal contexts, because resource restrictions on community care can sometimes limit the adequacy of available services. Correctional systems provide additional challenges not present in the community, given that incarcerated persons have no options for care other than what is provided in the correctional facility.

The Purposes of Psychiatric Treatment in Jails and Prisons

Psychiatric disorders and psychiatric comorbidities are overrepresented in correctional populations compared with the general population. Correctional facilities and systems are, therefore, legally obligated to address the medical and mental health care needs of individuals in their placement. The purposes of psychiatric treatment in jails and prisons as set out by the APA Guidelines are listed in Table 58.3–2. These principles are echoed in the NCCHC Standards, which add that the goal is the "best level of functioning" and "In the correctional setting, as in most other environments, the immediate objective of mental health treatment is to alleviate symptoms of serious mental disorders and prevent relapses to sustain patients' ability to function safely in their environment."

Models of Health Care in Jails and Prisons

Different correctional settings have different organizational models for health care delivery, and it is important that correctional psychiatrists become familiar with the model that their institutions follow. In many facilities, medical services and mental health services may be under the same or under separate authorities with different administrators and lines of reporting and supervision. Variations include situations in which the correctional psychiatrist is a full-time or part-time state, federal, or county employee, employed by a university or private health care system, or is a locums tenens, part-time or temporary individual contractor providing mental health services to any of the above or alternatively an employee or individual contractor for a national for-profit organization that is responsible for medical and/or mental health care services.

Table 58.3–2.
The American Psychiatric Association Guidelines on Purposes of Psychiatric Treatment in Jails and Prisons

To relieve the unnecessary extremes of human suffering
To make the correctional facility safer for everyone who lives, works, or visits there
To enable patients to avail themselves of the rights of due process
To permit inmates the opportunity to make use of programs offered
Correctional settings is public psychiatry as jails are the new psychiatric hospitals

Jails and prisons differ in the scope of mental health services that are provided. Jails process most detainees who have a short length of stay, and, accordingly, jails are primarily concerned with the stabilization and management of acute conditions such as acute psychosis, substance intoxication, withdrawal, and suicide prevention. While a person is in jail, the purpose of mental health treatment is to prioritize the detainee's health and safety, but, when appropriate, this might also include the goal and outcome of restored competency to be arraigned or restored competency to stand trial so that their pending legal proceedings may resume. Because this is a forensic and nonclinical role, typically, jail psychiatry staff are not involved in the assessment of whether an individual is competent or is not competent to stand trial. Similarly, they are typically not involved in the treatment of restoration of competency to stand trial, but this may vary by location (and whether the jail has a competency restoration program on-site or if this is a separate on or off-site program). Of course, once an individual returns to jail, ensuring continuity of care and the psychotropic medications (typically initiated in the state hospital or competency restoration program) and other treatments is essential and can be very challenging.

In a prison setting, the outcome might be to return the inmate to the general population with the ability to participate in prison programs, including clinically indicated mental health programs, psychotropic medication administration, monitoring, further psychiatric evaluation and treatment, and preparation for release. Prisons should be considered communities in which individuals spend many years of their lives. As such, prisons need a greater range of services that emphasize long-term support for particularly vulnerable individuals, including sheltered and/or residential units (for individuals with stable conditions who cannot be placed into the general population), case management, individual and/or group therapies, and other in-cell programming materials (self-study materials and the like). Programs should be available to prison inmates who repeatedly self-injure, mutilate, those with the need for frequent/repeated psychiatric hospitalizations who might require step-down programming, followed by sheltered or residential unit placement with other similar offender patients, and more intensive treatment to ensure medication compliance, monitoring of their activities of daily living (ADLs), and medication and illness teaching and education, than that which is typically available within general population settings. Sheltered housing and tailored assessment and treatment programs for individuals with developmental disabilities are also necessary.

Issues in Administration and Management

WRITTEN POLICIES AND PROCEDURES. Mental health administrative personnel in a jail or prison should establish written policies on such critical issues as (1) staffing patterns, (2) admission, (3) referral, (4) discharge criteria, (5) health care information management, and (6) interagency and intra-agency communications, especially with regard to confidential medical and mental health information.

THE FACILITY INFORMATION TECHNOLOGY SYSTEM: NEEDS ASSESSMENT AND ALLOCATION OF RESOURCES. With regard to a correctional facility's information technology system, the APA Guidelines point to a capacity for "needs assessment" and state: "It is critical that the administration of the correctional facility have the capacity to gather and assess information relevant to the prevalence of mental illness in the facility because this data forms the basis of the needs assessment that assures an adequate level of care."

THE FACILITY QUALITY ASSURANCE PLAN. One important aspect of the reform of the correctional mental health system calls for each jail or prison to have a quality assurance plan that systematically reviews the quality of mental health services and the efficiency and effectiveness of the utilization of staff and resources in the delivery of services.

ACCREDITATION AND STANDARDS. The NCCHC has become the major organization establishing correctional health care standards and providing correctional health care accreditation. In 2007, the NCCHC published separate standards for the accrediting of mental health programs in jails and prisons, Standards for Mental Health Services in Correctional Facilities. The most current NCCHC standards include the Jail (2018), Prison (2018), Mental Health (2018), and the Juvenile Standards (2018). The NCCHC also has developed standards for opioid treatment programs within correctional settings, which are linked to specific federal regulations.

Jails, prisons, juvenile detention, and other correctional facilities may be accredited by the NCCHC (an offshoot of the AMA as described previously), the ACA, The Joint Commission, or a combination of the above. Although national accreditation is typically voluntary, it is often a contractual requirement for universities, other health care systems, and private vendors who provide health care services to correctional systems. In addition, when facilities undergo investigation or litigation or are placed in receivership or federal oversight, they are often mandated to establish and maintain national accreditations.

Access to Mental Health Care and Treatment in Correctional Settings

Constitutional provision and statutory law require that an inmate in a correctional setting have access to mental health treatment, and this principle is emphasized in the APA Guidelines on Psychiatric Services in Jails and Prisons, which state, “Timely and effective access to mental health treatment is the hallmark of adequate mental health care.” In addition, it is striking to note that the very first standard of the NCCHC’s Standards for Mental Health Services in Jails and Prison is on access and requires that “inmates have access to care to meet their serious mental health needs,” defining access to care as meaning that “in a timely manner, a patient can be seen by a clinician, receive a professional clinical judgment, and receive care that is ordered” and adds that this principle is “the foundation upon which National Commission on Correctional Health Care standards are based.”

Furthermore, this section notes that access to care procedures should be flexible enough to accommodate inmates with mental illness and functional deficits and that such deficits can manifest as the following:

- Inability to recognize emerging health or mental health need
- Inability to follow usual procedures in requesting health or mental health care
- Severe anxiety or suspicion preventing follow-up care
- Behavior that makes care especially challenging

Barriers to Inmates Receiving Services. To determine whether there is adequate and appropriate access to psychiatric treatment, it is critical that the treatment delivery system be analyzed to ensure that there are no unreasonable barriers to patients' receiving services. Some examples of unreasonable barriers listed in the APA Guidelines are listed in Table 58.3–3.

Table 58.3–3.
American Psychiatric Association Guidelines on Examples of Unreasonable Barriers to Patients' Receiving Psychiatric Services

Instituting or allowing disincentives that would deter a patient from seeking care for their legitimate mental health needs
Interfering with the prompt transmittal of a patient's oral or written request for care
Allowing unreasonable delays before patients are seen by mental health staff or outside consultants
Charging fees or costs that prevent or deter patients from seeking care
Punishing an inmate for seeking or refusing care

Staffing. The effectiveness of any correctional psychiatric system is largely dependent on staffing considerations. Staffing must be sufficient to ensure that every inmate with serious mental illness or in psychiatric crisis has timely access to evaluation by a competent mental health professional. More important than the number of staff is access to adequate care. Accordingly, none of the national correctional health care standards mandates staffing patterns other than that there be an adequate number and type to meet the needs of the population.

Screening. In jails and prisons, screening is critical, and, not surprisingly, both the APA Guidelines and the NCCHC Standards emphasize the importance of appropriate screening. The APA Guidelines describe in detail the sequence of screenings and assessments that lead up to treatment (Table 58.3–4).

For example, the APA Guidelines describe the progression in regard to an individual with a substance abuse disorder who might present while receiving mental health screening with any of a variety of medical, neurologic, or psychiatric problems. The receiving mental health screening instrument should include questions relating to the individual's use of alcohol and other drugs, the patterns of use, and observations related to whether the offender is under the influence of alcohol or drugs or experiencing withdrawal symptoms. The Guidelines note that the correctional psychiatrist should have an important role in the development of the screening instruments, in the training of staff in its use, and in educating correctional security, medical, and mental health staff about the signs and symptoms of intoxication and withdrawal.

A positive mental health screening for substance intoxication should trigger an immediate mental health screen for the presence of depressed mood and/or suicidal ideas. Similarly, the intake mental health screening should include questions on an inmate's history of alcohol and other drug use, as well as observations for the presence of signs of intoxication or withdrawal symptoms. The brief mental health assessment and the comprehensive mental health evaluation are particularly important in ensuring a proper diagnosis as the basis for the appropriate clinical management of an inmate with a substance use problem.

Table 58.3–4.
American Psychiatric Association Guidelines on Sequence of Screening and Assessments Leading up to Treatment

- Screening
  - Receiving mental health screening
  - Intake mental health screening
- Evaluation
  - Brief mental health assessment
  - Comprehensive mental health evaluation
- Intervention

The NCCHC notes that "receiving screening" is the means by which "newly admitted inmates' emergent or urgent mental health needs are identified and met through an immediate referral to mental health staff."

Both the APA and the NCCHC require a subsequent “intake” screening that must take place within 14 days. The step-by-step sequence ensures that an inmate who is identified through reception screening, intake screening, or by various processes at any time thereafter (including self-referral) as possibly in need of mental health treatment is provided with a suitable mental health evaluation.

Sick Call and Nonemergency Mental Health Care Requests and Services. Although the correctional psychiatrist works with inmates by appointment in an office situation if possible, in most correctional facilities, access to general health care is by means of “sick call.” The NCCHC Standards are quite specific and require that “all inmates have the opportunity daily to request mental health care” and that “verbal or written requests for mental health care are received daily by mental health staff and triaged within 24 hours.” There is a separate NCCHC standard for emergency services, and emergency mental health care is defined as “care for an acute mental illness that cannot be deferred until the next scheduled mental health clinic or routine appointment,” which must be available 24 hours a day, 7 days a week.

Referral. Emphasizing that within a correctional facility, the referral process is critical, and noting that this can be simple or complex, depending on the facility, the urgency of the situation, and the mental health coverage provided, the APA Guidelines state that the referral process, as set out in the facility’s operating standards, should be time constrained and specifically define the roles of the various participants, that the maximum time for response to each referral should be stated, and that timely referral and response should be designated as indicators of quality care.

Consultation. An important role of the correctional psychiatrist is to take the initiative in providing consultation to staff with regard to certain situations. Some of these are listed in the NCCHC Standards, such as consulting with regard to inmates who have a chronic illness (i.e., asthma, heart disease, diabetes, and hypertension) or communicable disease (e.g., HIV, hepatitis, syphilis, gonorrhea, or chlamydia); counseling patients coping with the psychological effects of their disease and consulting with regard to inmates who are physically or developmentally disabled and either assisting staff in managing such inmates or providing behavioral consultation so that the inmate can deal with the emotional and attitudinal consequences of their disability during incarceration. Consultations might also be called for with regard to inmates who are chemically dependent; pregnant or have recently given birth, for example, to minimize the postpartum phase and separation anxiety from the newborn; and of course, older adults or terminally ill inmates.

The Therapeutic Milieu for Mental Health Treatment. Table 58.3–5 lists the APA Guidelines conditions that would contribute to a therapeutic milieu for treatment.

Evidence-Based Treatment and Prescribing

Correctional psychiatrists must strive to provide evidence-based treatment as the standard of care in jails and prisons. Evidence-based treatments can be defined as treatments that, in appropriate studies, (1) have been rigorously tested, (2) have yielded consistent, replicable results, and (3) have proven safe, beneficial, and effective for most people diagnosed with the particular mental illness. While some researchers might argue over the standards by which an intervention or treatment approach is judged to be evidence-based, there is general agreement that the term and designation imply that a given practice has withstood rigorous scientific examination

Table 58.3–5.
American Psychiatric Association Guidelines on Conditions That Contribute to a Therapeutic Milieu for Treatment

A therapeutic milieu implies the following conditions:

- A sanitary and humane environment
- Written procedures and adequate staffing to permit adequate observation
- Adequate allocation of resources for the prevention of suicide and assault
- Medical and mental health staff available to provide adequate treatment and supervision
- Social interactions that foster recovery
- The ability to transfer an inmate to an appropriate mental health facility when the foregoing conditions cannot be met

The public mental health system must take steps to ensure that practice keeps pace with research. Because correctional psychiatry is a critical part of public psychiatry, it is appropriate to quote from the Report of the Criminal Justice/Mental Health Consensus Project, which describes the development of the concept as it promotes the utilization of evidence-based practices in public psychiatry:

> *In recent years, enormous advances have been made in treatments available for persons with mental illness. New medications have emerged; new services, supports, and interventions have proven effective. Researchers have conducted studies and collected data—they have developed an "evidence-base"—which demonstrate the effectiveness and applicability of some of these treatments and approaches.*

Of note, The APA publishes Practice Guidelines that provide evidence-based recommendations for the assessment and treatment of a wide range of psychiatric disorders of considerable importance to correctional psychiatrists, including acute stress disorder and post-traumatic stress disorder (PTSD), bipolar disorder, borderline personality disorder, psychiatric disorders related to HIV/AIDS, major depressive disorder, obsessive-compulsive disorder (OCD), panic disorder, schizophrenia, substance use disorders, suicidal behaviors, and others.

It is generally recommended that psychiatrists should use psychotropic medications in a safe and clinically appropriate manner and after a review of the potential risks, benefits, side effects, and alternatives. As with any mental disorder, barring maintenance treatment, it is unwarranted to prescribe psychotropic medications in the absence of distinct target symptoms or when placement and mental health follow-up services are unclear. To avoid risk management problems, it is generally recommended that medication be used according to U.S. Food and Drug Administration (FDA)—approved indications and doses, and if this is not preferred, then it is suggested that the correctional psychiatrist obtain and document a second opinion.

Precautions against abuse are necessary, and so, for example, it is recommended that directly observed therapy (DOT), crush and float, or other strategies be implemented for medications with particular overdose or bartering risks. The risk of medication diversion is an additional consideration, including diversion of medications with CNS effects that are maximized by nasal insufflation, injection, and other means of administration.

Involuntary Treatment in Correctional Psychiatry. Because involuntary treatment is complex and changes depending on the jurisdiction, the APA Guidelines say only "the actual institutional policies and the decisions of the correctional psychiatrist in regard to the right to refuse treatment should conform with the rules and procedures of the jurisdiction in which the facility is located."

On the contrary, the importance of this subject is highlighted by the fact that there are two separate NCCHC standards that apply. The first, Informed Consent and Refusal of Mental Health Care, states, "All examinations, treatments, and procedures are governed by informed consent practices applicable in the jurisdiction" and notes that "the intent of this standard is to support the ethical principle that patients have the right to accept or refuse mental health interventions."

The second NCCHC standard, Emergency Psychotropic Medication, states, "Mental health staff follow policies developed for the emergency use of forced psychotropic medications as governed by the laws applicable in the jurisdiction" (the definition of "forced medications" is "those given without the patient's consent").

It should be noted that the terminology used in the NCCHC Standards is dramatically different from that generally used in psychiatry. As a rule, psychiatric facilities use terms such as "treatment over objection" or "refusal of medication." The Standards focus on the word "forced" and its implications. Thus, the discussion of the latter standard states:

> *This standard's intent is to have a clinically approved protocol for emergency situations when an inmate is dangerous to self or others due to a medical or mental illness and when forced psychotropic medication may be used to prevent harm, based on a prescriber's order. The standard supports the principle that psychotropic medication may not be used simply to control behavior or as a disciplinary measure.*

Treatment Planning. The correctional psychiatrist has been trained as early as medical school to understand the importance of a detailed, in-depth, professional treatment plan. In any context or location, this requirement is a crucial reflection of the necessary standard of medical and psychiatric care. This is reflected in the APA Guidelines, which, with regard to discharge planning, notes that "it is recommended that discharge planning be a part of the initial treatment plan." Similarly, the NCCHC Standards require that "mental health services are provided according to individual treatment plans." The standards examine this subject in considerable detail, requiring that "patients receive care that is tailored to their individual needs . . . [including] consideration of cultural and language differences."

Peer-to-Peer Programs. Peer-to-peer programs have proven their worth in prison ministries, chaplain volunteers, and other spiritual ministries. Other peer activities may include volunteer work, providing care and training to service dogs for Veterans, and other individuals with disabilities, for example, the Patriot Paws programs. There are other self-help peer-to-peer organizations such as Alcoholics Anonymous (AA) and Narcotics Anonymous (NA). This concept is being extended into areas of reentry and continuing postincarceration support.

The APA Guidelines state, "As safety and security allow, self-help and peer support programs or activities contribute to the overall goals of the psychiatric services and where and when available should be promoted and encouraged."

However, within the correctional community, this is not an uncontroversial area. Concerns about the power of inmate workers were cited earlier. There are valid security concerns about prior inmates working within a correctional facility.

Suicide Prevention. The risk of suicide is dramatically higher among incarcerated jail and prison inmates than among the population at large. Suicide rates for men in jails are approximately 9 to 15 times higher than for the general population, and prison suicide rates are approximately 1.5 times higher than in the general population. An increased risk of suicide associated with mental illness also has been found. Researchers examined psychiatric disorders and completed suicide in the nation's largest state prison system, and their findings highlight the importance of maintaining suicide prevention programs in correctional settings with particular emphasis on screening and monitoring of patients with severe psychiatric disorders. It is further a legal requirement that every jail and prison must have a suicide prevention program that identifies and responds to each suicidal inmate and takes an active approach to their management.

An essential standard from the Jail, Prisons, Juvenile and Mental Health Standards of the National Commission outlines that an "adequate" suicide prevention program in a particular jail or prison should include the following components:

- Training for all staff who interact directly with inmates to recognize warning signs and intervene appropriately with individuals at risk for suicide.
- A formal and detailed suicide risk assessment process.
- Identification of inmates at increased risk of suicide through screening at or near the time of admission to the facility and through referral at any time during an individual's incarceration.
- An effective and well-understood referral system that allows staff and inmates to bring a suicidal inmate to the prompt attention of a mental health clinician.
- Timely evaluation by a mental health clinician to determine the level of risk posed by a referred inmate.
- Timely implementation of monitoring interventions such as close observation (at least every 15 minutes), continuous monitoring, alternative housing, or referral to a higher level of care (e.g., infirmary or hospital). When indicated, a psychiatrist either on call or in person assists in the decision about interventions. Any mental health clinician may order an increase in monitoring level, but a decrease in level may be ordered only by a psychiatrist or doctoral-level mental health clinician after an in-person evaluation.
- Housing options that allow for adequate monitoring of suicidal inmates by staff. Suicidal inmates should not be placed in isolation settings without continuous monitoring. Continuous monitoring for active suicidality should occur regardless of housing location. Supervision aids such as video monitoring may be used as a supplement to, but not as a replacement for, active staff monitoring.
- Communication among mental health, correctional, medical, and other staff of the specific needs and risks presented by a suicidal inmate.
- Timely provision of mental health services, including medication, individual and group therapies, and crisis intervention, for chronically or acutely suicidal inmates.
- Accurate, behaviorally specific, and highlighted documentation in the medical record of behaviors or statements that indicate suicide risk.
- Quality improvement reviews with psychiatric staff participation in suicide attempts and completed suicides to help prevent future occurrences.
- Critical incident support, opportunities for peer-to-peer discussion, and availability of Employee Assistance Program support for completed

suicides to assist staff and inmates in dealing with feelings of guilt, fear, grief, and anger.

HIGH-RISK PERIODS. It is obvious that an inmate might become suicidal at any point during incarceration. The high-risk periods that both the APA Guidelines and the NCCHC Standards emphasize are listed in Table 58.3–6.

Table 58.3–6.
Periods of High Risk for Suicide in a Jail or Prison

At the time of admission to a lockup, jail, or prison
After new legal problems (e.g., new charges, additional sentences, institutional proceedings, denial of parole)
After receiving bad news regarding one's self or one's family (e.g., a serious illness in the family or the loss of a loved one)
After suffering humiliation or rejection (e.g., sexual assault)
On being placed in administrative segregation or other specialized single-cell housing assignment
In the early stages of recovery from severe depression

Monitoring for and Preventing Self-Harm. Although there are many more suicide attempts and incidents of nonlethal self-harm than completed suicides, any threats of self-harm or self-injurious behavior must be taken seriously. Self-harm behaviors, regardless of motivation, can result in significant morbidity and mortality. For this reason, even behaviors that do not lead to injury should be taken seriously and not dismissed as merely "suicidal gestures."

Because self-harm is hard to control, it is a serious challenge for correctional officials and a complex clinical situation for the correctional psychiatrist, who must determine whether the self-harm is the result of a mental disorder. This requires a variety of means to assess the problem, including risk assessment methods and protocols and might require temporary housing in a mental health unit.

When self-mutilation is determined not to be the product of a mental disorder, the inmate may be denied admission to a mental health unit. This may confuse family members, attorneys, courts, patient advocates, and correctional authorities, who view such behavior as a sign of mental instability. It is, therefore, essential that the correctional psychiatrist follows clear protocols for self-harm to determine whether it is a psychiatric symptom, a manipulative gesture to escape a dangerous situation, or an effort to manipulate the system for personal gain. Careful and thorough documentation and communication with custody personnel are the key factors. Treatment, however, must be individually tailored and calls for a team approach.

Seclusion and Restrain in Correctional Psychiatry. Both the APA and the National Commission have addressed the complex issue of seclusion and restraint in correctional facilities. In 2006, the APA published the Resource Document on the Use of Restraint and Seclusion in Correctional Mental Health Care, which provides a detailed and nuanced discussion of this complex topic. It is strongly recommended that every practicing correctional psychiatrist carefully study and review this document. It "discusses the use of seclusion or restraint for purposes of mental health intervention in jails and prisons, in contrast to its use for correctional purposes (i.e., specifically custody reasons)."

With regard to the correctional setting specifically, this APA Resource Document details such matters as the location where the restraints or seclusion are being employed, the time frames that are required, and the devices that are permissible under various circumstances. The same principles are contained in the NCCHC Standard entitled Use of Restraint and/or Seclusion for Mental Health Reasons, which states, "Clinically ordered restraint or clinically ordered seclusion are available for patients exhibiting behavior dangerous to self or others as a result of mental illness." The standard further distinguishes between "custody" restraint or punitive segregation initiated for security concerns and "clinically ordered" restraint or seclusion initiated by medical or mental health staff as a therapeutic intervention when there are clinical concerns requiring intervention.

There remains a great amount of controversy regarding the empirically validated psychological effects of segregation and what effects, specifically, this has on individuals with or without serious mental illness. In sum, it appears that prolonged segregation exposes individuals to potential psychological, physiologic, and medical risks. Those with SMI have a special vulnerability to the adverse effects of social isolation. Psychiatrists should assist correctional systems with the development of evidence-based behavioral interventions that avoid the potential harm of social isolation.

The NCCHC Standard broadly defines segregated inmates as “those isolated from the general population and who receive services and activities apart from other inmates.” These housing units can be labeled as “administrative segregation, protective custody, disciplinary segregation, or a supermax tier.” It notes that “for the purposes of this standard, the living and confinement conditions define the segregated status, not the reason an inmate was placed in segregation.” The NCCHC Standard has an additional category—“extreme isolation”—which “refers to situations in which inmates are seen by staff or other inmates fewer than three times a day.”

The APA Guidelines state that when an inmate is segregated for any reason from the general population, the responsibility to address serious mental health needs remains in effect (Table 58.3–7). Indeed, because of the stressful nature of segregation housing, facilities should make special efforts to assess and address mental health treatment needs in these settings. Although correctional psychiatrists should not be involved in the decision to place an inmate in disciplinary segregation, once an inmate is there, the psychiatric staff should monitor their mental health status because an inmate who is placed in and is unable to cope with administrative segregation or other specialized single-cell housing assignment might also be at increased risk of suicide.

Table 58.3–7.
The American Psychiatric Association Guidelines on Principles for Mental Health Services in Segregation Housing

The APA Guidelines recommend that in providing essential mental health service in segregation housing, the following principles should be observed:

- Inmates should be placed in segregation housing solely for correctional purposes and never because they exhibit the symptoms of mental illness, unless there is an immediate and serious danger for which there is no other reasonable alternative. (This principle does not refer to medical or psychiatric seclusion, which should follow state mental health law and professional practice.)
- When an inmate is placed in segregated housing for appropriate correctional reasons, the facility remains responsible for meeting all of the serious medical and psychiatric needs of that inmate. Thus, such inmates must receive any mental health services that are deemed essential, and require closer monitoring of psychiatric decompensation than in general population.
- Inmates who are in current, severe psychiatric crisis, including but not limited to acute psychosis and suicidal depression, should be removed from segregation until such time as they are psychologically able to tolerate that setting.
- Inmates who are known to have serious mental health needs or a history of one, when housed in segregation, must be assessed on a regular basis by qualified mental health practitioners to identify and respond to emerging crises at the earliest possible moment.
- Institutions should provide for regular rounds by a qualified mental health clinician in all segregation housing areas, and each inmate should be visited briefly to assess any emerging problem. The clinician should communicate with segregation security staff in order to identify any inmate who appears to be showing signs of mental deterioration or psychological problems.

Similarly, the NCCHC recommends, “inmates who are seriously mentally ill should not be confined under conditions of extreme isolation.” The Commission adds, “policies should also address how vulnerable populations like an adolescent, older adults, pregnant or mentally ill inmates are managed in segregation. Since segregation can exacerbate pre-existing conditions, staff should be extra cautious when placing vulnerable inmates in segregation." Similar position statements have been published by the APA (2017), and the SCP (2013).

Institutional Transfers. Classification committees continuously juggle custody and medical classifications of inmates with their workforce requirements and available space. Most institutions provide a list of the next day's transferees to the correctional psychiatric staff so that they can assemble the inmate's mental health records. Correctional psychiatric staff should ensure that the receiving unit has the requisite psychiatric resources to continue to meet the patient's needs. Each facility should have a policy that allows psychiatric staff to put a psychiatric hold on any interunit transfer of an inmate. Decisions regarding admission to and discharge from inpatient facilities are the sole provinces of clinicians.

Disciplinary Hearings. As a rule, correctional psychiatric staff should not participate in disciplinary hearings and should never be part of the punishment decisions. However, there are occasions when information from psychiatric staff might be helpful in protecting the inmate from inappropriate discipline. For example, a schizophrenic patient might refuse to comply with a correctional officer's directions because of negativism rather than defiance.

Discharge Planning. Studies of recidivism among persons in the justice system with mental illness indicate that effective discharge planning is essential to the continuity of care and a critical part of adequate psychiatric treatment in jails and prisons.

Despite court interventions developed for detainees with mental illness, there is still a lack of coordination and continuity of care for such persons on release. After incarceration, former inmates with mental illness have residential, financial, health, and mental health needs. Face-to-face linkage with community services has the greatest likelihood of success. Because most jails cannot control the court's release decisions, they usually do not know when an inmate is to be released, and so most jails do not provide case management or discharge planning services. To provide such services, discharge planning must start at admission. Providing uninterrupted medications after release is an important challenge. Without psychiatric medications, recidivism is almost inevitable.

As the APA Guidelines state, “Timely and effective discharge planning are essential to the continuity of care and an integral part of adequate mental health treatment.”

This is true whether the patient is being released into the community or transferred to another correctional facility. Because discharges (e.g., from jails) or transfers (e.g., between prisons) can occur on short notice, it is recommended that discharge planning be a part of the initial treatment plan. Essential services that should be included in adequate discharge planning are psychiatric assessments, appointments to be arranged, notifications, staff responsibilities, linkages to community-based services, help with the obtaining of necessary financial benefits (e.g., Medicaid) and housing (due to the high incidence of homelessness in the jail and prison populations), placements, and appropriate linkages.

The NCCHC Standards suggest a number of components for a discharge-planning program:

- Formal linkages between the facility and community-based mental health organizations
- A list of community-based mental health providers that is used for referral
- An inmate education session that emphasizes the importance of follow-up aftercare
- Arrangement for follow-up mental health services before an inmate’s release
- Preparation of mental health summaries
- Assistance to inmates with their application to entitlement programs
- If appropriate, a supply of medications provided at the time of the inmate’s release

Reentry Programs. Whereas discharge planning is an aspect of services in jails and prisons, the last few decades have seen the proliferation of reentry programs, which can provide discharge planning and transition services to people with mental illnesses being released from city jails or state or federal prisons. These programs (usually government-funded—generally state or county) evaluate, select, and then actively link (and follow-up) the inmate with mental illness reentering the community to a wide variety of services, including intensive case management, mobile services, outpatient treatment, psychiatric rehabilitation, housing, vocational and educational services, access to entitlements, family supports, and substance abuse treatment.

This important area of reform of the correctional system gained added impetus from the Second Chance Act of 2007, federal legislation designed to ensure the safe and successful return of prisoners to the community by providing the following: Grants to states and local governments to promote the safe and successful reintegration into the community of individuals who have been incarcerated, including employment services, substance abuse treatment, housing, family programming, mentoring, victims services, and methods to improve release and revocation decisions using risk-assessment tools; grants to nonprofit organizations that can be used for mentoring adult offenders or offering transitional services for reintegration into the community; grants to improve the availability of drug treatment to offenders in prisons, jails, and juvenile facilities; and grants to states, local governments, and Indian tribes to develop and implement family-based treatment programs for incarcerated parents who have minor children. It also authorizes the U.S. Justice Department's National Institute of Justice and the Bureau of Justice Statistics to conduct reentry-related research and establishes a national resource center to collect and disseminate best practices and to provide training on and support for reentry efforts.

In a 2007 Task Force report by the American Psychiatric Association's Task Force on Outpatient Forensic Services, several barriers to successful re-entry following incarceration are identified; these included significant social isolation and stigma related to both mental illness and justice involvement; additionally, loss of social benefits such as SSI/SSDI, difficulties seeking housing, and the severe shortage of outpatient psychiatric services. Such gaps in access further contribute to the risk of recidivism and reincarceration among those released to the community without adequate supports.

Several types of programing have been proposed to aid in successful re-entry and reduce the risk of repeat justice involvement. Two examples include Forensic Assertive Community Treatment (FACT) and Critical Time Intervention (CTI). FACT aims to reduce rates of recidivism by engaging high-risk individuals with histories of justice involvement, increasing engagement in care through proactive outreach and wrap-around services. In contrast, CTI represents a 9-month, staged intervention that promotes individuals being linked to community services and support agencies. It has been shown to be effective at facilitating treatment engagement at the time of release from psychiatric hospitals and shelters, and the intervention has also been described as a promising model for facilitating re-entry from prisons for individuals who have a mental illness.

Each year correctional facilities release approximately 12 million inmates. Because society inevitably shoulders the cost of their treatment, the period of incarceration is an opportunity to create an infrastructure to identify and treat these patients. Effective correctional psychiatric systems will have a positive influence on the public's mental health. So, in recent years, it has become increasingly evident that correctional psychiatry must be considered an integral part of community psychiatry and that correctional health must be conceptualized as an important aspect of public health.

Additional Treatment Considerations and Special Populations

Women in Correctional Facilities. Although the number of women in jails and prisons is relatively small compared to the number of men, their percentage of the total correctional population is rising. Thus, at year-end 2006, women made up 7.2% of the population under state or federal jurisdiction, up from 6.7% in 2000. Until recently, most psychiatric treatment programs for women have been based on treatment models designed for men. It is now recognized that female inmates require very different screening, evaluation, and treatment, and with the reform of the correctional system, treatment programming provided to women in jails and prisons is changing dramatically.

As noted in the APA Psychiatric Services in Correctional Facilities, 3rd Edition: Female inmates frequently encounter different stressors before, during, and after incarceration than their male counterparts. More than 75% of women entering correctional settings report a history of emotional, physical, or sexual abuse as a child or adult. They have higher rates of domestic violence victimization than men. One aspect of this change is the recognition that female inmates have had high rates of victimization (sexual and physical), causing PTSD. The trauma of being raped before 16 years of age was the most frequently reported antecedent of PTSD among female prisoners. Such prevalence of trauma requires that a trauma-informed lens be applied to the provision of care in correctional settings, as well as to the utilization of potentially traumatizing interventions such as seclusion and restraint.

Demographic studies have shown that compared to men, a greater percentage of women entering a jail or prison have a serious and persistent mental illness. Although the rates of bipolar disorder and schizophrenia are similar for incarcerated men and women, a much larger percentage of women are diagnosed with major depression and severe anxiety disorders, especially PTSD. In addition, a higher percentage of women than men in jails and prisons enter these settings with a history of substance abuse.

Programming for women in correctional facilities must also consider that most women in correctional settings have dependent children and at least one child for whom they have been the primary caregiver. Their concerns about their children require special counseling on parenting and perhaps child custody. It is additionally important to attend to the health and mental health care of women who are pregnant or who have given birth recently. The correctional psychiatrist should be alert for and should minimize the risks associated with peripartum depression and psychosis, other adjustment disorders, and distress, and ensure that the facility establishes a policy that a comprehensive mental health evaluation be conducted or has any mental health problems during the pregnancy, delivery, or postpartum period.

Table 58.3–8.
American Psychiatric Association Position Statement on Treatment of Substance Use Disorders in the Criminal Justice System (2016)

1. Consider drug court and similar alternatives whenever they may be viable options to divert individuals from the criminal justice system.
2. All inmates should be screened by trained professionals and assessed for the presence of substance use disorders, withdrawal symptoms, and comorbid psychiatric illness, in addition to Hepatitic C, tuberculosis, and HIV.
3. Treatment for the above should be available within correctional facilities, or alternatively the ability to arrange timely transfer to an appropriate medical facility should be available. Health professionals should be able to screen for withdrawal.
4. Pharmacologic therapies, relapse prevention strategies, and psychotherapy for substance use disorders should be available for arrestees and inmates with substance use disorders.
5. Probation or parole should be conditional on receiving treatment for the substance use disorder whenever appropriate and clinically recommended. Furthermore, community treatment resources and access to care should be ensured upon release.

Data from American Psychiatric Association. *Position Statement on Treatment of Substance Use Disorders in the Criminal Justice System*. American Psychiatric Association; 2016.

Inmates With Substance Use Disorders in Correctional Facilities. The correctional psychiatrist must have knowledge and familiarity in the assessment and treatment of incarcerated individuals with substance use disorders because it is estimated that up to 70% to 90% of offenders have a substance use disorder at the time of their entry into the criminal justice system, and criminal behavior often occurs when an individual is under the influence of alcohol or other drugs. Furthermore, notably, some serious suicide attempts and completed suicides in jails may occur within the first 24 to 48 hours of admission and are often carried out by inmates who are either intoxicated or experiencing withdrawal symptoms. In 2016, the APA Assembly and the APA Board of Trustees approved a revised position statement prepared by the APA Council on Addiction Psychiatry on the treatment of substance use disorders in the criminal justice system, as shown in Table 58.3–8.

INMATES WITH CO-OCCURRING DISORDERS IN CORRECTIONAL FACILITIES. In 2006, the Bureau of Justice Statistics stated that about 74% of inmates in state prisons and 76% of jail inmates who had a mental health problem also met the criteria for substance use disorder. Nearly 63% of state prisoners who had a mental health problem had used drugs in the month before their arrest, compared to 49% of those without a mental health problem. For the correctional psychiatrist, the nondetection of one or more co-occurring disorders in jails and prisons risks the exacerbation of any behavioral problems and elevates the risks of suicide and poor outcomes in treatment.

For these reasons, the APA Guidelines recommend that the detection of one type of disorder (i.e., substance use or mental illness) immediately triggers an investigation into the other type of disorder. The importance of a coordinated system of screening, referral, assessment, and diagnosis for co-occurring mental health and substance use disorders is that one type of disorder can mimic the other, interfere with detection of the other, or overshadow the other type, and one type might affect the presentation of the other type of disorder. Table 58.3–9 lists the recommended APA Guidelines strategies for inmates with co-occurring disorders.

Table 58.3–9.

American Psychiatric Association Guidelines on Recommended Strategies for Inmates with Co-Occurring Disorders

Services must integrate treatment that addresses the person’s mental health and substance use disorders simultaneously.
Each type of disorder must be treated as a primary disorder, with a focus on understanding how they interact with each other.
Medication must be prescribed with caution. In individuals who are not on psychotropic medications, they should be fully allowed to detoxify from substances before starting medications unless psychotic or suicidal symptoms are present.
Psychosocial problems and skill deficiencies should be addressed with individualized programming.
Treatment should be integrated with self-help groups and support networks, which may be invaluable tools in assisting treatment participants in maintaining their commitment to daily alcohol and drug abstinence.
Interventions must be designed for the particular setting, each setting requiring differing intensity, length, and types of services.
The presence of co-occurring disorders requires special attention to discharge planning, which must address housing and job needs, family reconnection, and continued treatment.

The APA Guidelines emphasize that substance abuse services should be integrated with mental health services and that psychosocial problems and skill deficiencies should be addressed with individualized programming created through comprehensive assessment and consultation with the inmate’s clinician, family (or significant others), and correctional medical and mental health clinicians.

In contrast to the APA position recommending the integration of services, the NCCHC in its standard on Substance Abuse Services relegates the roles of the medical and mental health staff to referral:

> *Increasingly, in larger correctional systems, inmates with [substance use] . . . disorders are receiving treatment for these problems from specialized treatment providers. Substance abuse counselors who treat [substance use] . . . problems should be properly qualified to do so by a recognized accrediting body. Medical and mental health care*

providers can facilitate the treatment process by referring inmates for treatment, providing counseling to motivate inmates to receive treatment, legitimizing self-help groups, tailoring controlled substances prescribing practices, and otherwise being a partner in the process of treatment.

There is an agreement, however, that inmates with co-occurring mental illness must not be barred from participation in substance abuse treatment programs and that self-help, 12-step, and peer support groups and networks play an important role in the rehabilitation of abusers of alcohol and other drugs in the correctional system.

DETOXIFICATION. When psychotropic medications are required during detoxification to manage medical or prevent impairment and sequelae, agitation and symptoms, or agitation, a physician must be responsible for managing the detoxification process.

Prescription of benzodiazepines might be useful to prevent or attenuate withdrawal symptoms and detoxify the inmate. However, their use should be short-term and on a tapering dose schedule to ensure that the prescribed medication does not serve as a substitute for the substance of abuse.

When circumstances warrant, an inmate should be detoxified or free from substance use for a reasonable period of time before psychotropic medications are prescribed unless psychotic or suicidal symptoms are present. The opiate- or other substance-dependent pregnant female presents special challenges to correctional systems, and a review of NCCHC standards and other accepted practice guidelines is recommended.

Inmates with Intellectual and Developmental Disabilities in Correctional Facilities. Offenders with co-occurring mental illness and intellectual or developmental disability encounter unique difficulties in the criminal justice system and have special treatment needs. From 4% to 10% of inmates have an intellectual developmental disorder (formerly referred to as mental retardation or borderline intellectual functioning), that is, significantly subaverage intellectual functioning and impairments in adaptive functioning, and “it is estimated that 40% to 70% of individuals with mental retardation in jails and prisons have diagnosable psychiatric disorders.”

This is a substantial number of inmates in the correctional population who have intellectual disabilities that present special problems for the correctional psychiatrist. According to the APA Psychiatric Services in Correctional Facilities, inmates with an intellectual or developmental disability and co-occurring mental disorders are unfortunately the most likely to be victimized by other inmates. Furthermore, custody and treatment staff need additional training and education on intellectual and developmental disabilities to lessen the likelihood that they misperceive behaviors as intentional rule infractions or attribute them solely to intellectual disability while an SMI goes untreated.

Geriatric Inmates in Correctional Facilities. The Federal Bureau of Justice Statistics notes that the U.S. prison population is aging and that "expressed in terms of percent change, the oldest age group, inmates age 55 or older, recorded the largest change—an increase of 85 percent since 1995." Effective access to mental health services for older inmates requires recognition of the special challenges facing inmates and mental health professionals, especially as the number of offenders older than 65 years in state and federal prisons increased at 94 times the rate of the overall sentenced population between 2007 and 2010. The aging of the correctional population will be a major problem in the 21st century.

As noted by the APA Guidelines, for the correctional psychiatrist, older inmates have medical needs that sometimes might present with psychiatric symptoms and might complicate psychiatric intervention in terms of medication, relevance of counseling, structure of time (programming) of the inmate, and housing. These problems are more likely to be chronic, permanent, and progressive, might be related to the possibility of dying in custody, and might be expensive (Table 58.3–10).

For the older inmate, the isolation can be expressed in a variety of ways and, when extreme, can exacerbate or create mental illness or psychiatric crisis. Special issues are related to (1) acceptance after release from incarceration, (2) provision of food, shelter, and clothing after release, (3) assault and potential greater sequelae of injury, and (4) isolation from older, "free" relatives and friends who might be unable to travel to visit.

Table 58.3–10.
The APA Guidelines on Psychosocial Concerns of Older Inmates

Lack of connection to or interaction with other inmates in the general population, given the relatively small (though increasing) percentage of older inmates
Physical vulnerability to more serious consequences of assault
More difficulty and greater length of time to adjust to a new environment
Higher rate of completed suicide
Greater possibility of dying during incarceration
Higher incidence of loss of external supports, for example, spouse, parents, friends, and other external supports

Some of these issues might benefit from mental health intervention, including group or peer counseling. Psychiatric and clinical staff need training in the special needs and care required by older inmates, and they can help educate custody staff to recognize and deal with cognitively impaired inmates who might not respond appropriately to a direct order, mobility issues, and those who might demonstrate early confusion and cognitive impairments and associated sight and hearing impairments.

The Dying Inmate. Facilities should have written policies and training to address the in-custody dying inmate–patient. When feasible, the care of an inmate with terminal illnesses should involve the use of a hospice approach—a facility that emphasizes providing support and humane care to a dying patient rather than trying to cure the underlying disease. This might be inside the prison (less likely to be needed in jail) or a contract facility outside of the prison. Another consideration is the use of compassionate release (also called “medical parole”), which is available in a majority of states. Compassionate release can have a number of advantages, including cost savings for the prison health care system, which can be substantial.

The use of inmate volunteers (peer trainers) who are oriented on basic health issues, universal precautions, and mobility management can reduce the dying inmate–patient’s (or their family’s) perception that “the prison isn’t doing enough.” Group therapy, realistic planning, and suicide prevention are specific interventions that the psychiatrist should either provide directly or ensure provision by other mental health professionals.

The policy of “compassionate release” refers to the ability of institutions to grant early release or parole to individuals suffering from severe illness while incarcerated; these policies are based on the recognition that in cases of severe illness, appropriate treatment is difficult, if not impossible, to achieve in correctional settings. As the older adult population within the correctional system has continued to grow in recent years, the policy of compassionate release has become increasingly relevant. However, very few individuals are successfully granted release under this policy. Ensuring evidence-based palliative care and compassionate preservation of dignity have been cited as important considerations among medical professionals advocating for the value of compassionate release programming.

Rape and Victims of Sexual Assault. A full description of the psychiatric assessment and prevention of sexual abuse and exploitation within the correctional setting is beyond the scope of this chapter. The following provides a summary overview.

In 2003, the Prison Rape Elimination Act (PREA) became federal law; however, this phenomenon continues to be a challenge to manage with regard to prevention and aftercare. The Bureau of Justice Statistics report for the year 2007 indicated that an estimated 60,500 inmates (or 4.5% of all state and federal inmates) experienced one or more incidents of sexual victimization involving other inmates or staff. Additionally, among the 146 prison facilities in the 2007 National Inmate Survey (NIS), which was conducted in 146 state and federal prisons between April and August 2007 with a sample of 23,398 inmates, six had no reports of sexual victimization from the sampled inmates, and 10 had an overall victimization rate of at least 9.3%.

The national PREA standards were released in 2012. There is now a need for compliance with this federal mandate for facilities in the fields of adult, juvenile, and community corrections. Both correctional staff and health care staff might need education, training, and psychological support in managing either victims or perpetrators of sexual assault. Both the NCCHC and the American Correctional Association (ACA) play a vital role in supporting correctional institutions in the implementation of and compliance with the health care aspects of the PREA with education, curriculum developments, and technical assistance. Specialty training on complicated PREA issues is available for the psychiatrist and medical and mental health care practitioners. The curriculum includes information on detecting and assessing signs of sexual abuse and harassment, medical forensic examination and forensic evidence preservation, and effective and professional responses and reporting. In August 2013, the U.S. Department of Justice began auditing facilities to ensure compliance with the Act.

Inmates Serving Life Sentences. Psychiatrists may encounter inmates with one or more life sentences, life without parole, referred to as “lifers.” These individuals pose a unique challenge to the correctional psychiatrist. Because these individuals will not be released from prison, many usual incentives such as parole, token incentives, and the like have less effect.

The APA Psychiatric Services in Correctional Facilities notes that “inmates draw on many resources to assist their adaptation to incarceration, such as family contact, peer support, work, education, and spiritual pursuits, including the practice of faith-based traditions.” Some individuals may lack these psychosocial resources such as no regular family contact, no visitors, no one on their visitation lists, no one to help fund their commissary purchases, family estrangement, and the like.

Inmates on Death Row. The care of “Death row” individuals, inmates housed on death row, who are awaiting exhaustion of legal appeals, or whose legal appeals are exhausted, poses a myriad of complicated clinical, ethical, and legal questions. Some examples of issues include the death row psychiatric assessment, considering “justifiable” depression, suicide risk management, individuals on death row who require psychiatric hospitalization, psychiatric treatment thereby restoring one’s competency to be executed, and how to balance ethical/role boundaries between clinical and forensic work. Because of the lengthy appeals process and high stakes involved, mental illness and intellectual disabilities are commonly raised in postconviction proceedings, typically on Eighth Amendment claims (arguing that the execution of the mentally ill and/or mentally retarded constitutes cruel and unusual punishment).

Inmates with mental illness on death row present unique and complex ethical issues. Because the participation of a correctional psychiatrist in an execution would violate the principle of *do no harm*, Annotation 4 of

Section 1 of the APA's Annotations states that "a psychiatrist should not participate in a legally authorized execution." Much has been written about the interpretation of the word "participate," and it is clear that the medical community is unanimous that any action that would further the implementation of an execution is prohibited. Furthermore, there is no doubt that this applies to evaluations of competency to be executed. However, what about the treatment of an incompetent inmate to achieve competency? This is a controversial area, and whereas in general, the correctional psychiatrist is obliged to provide treatment to patients in need to relieve suffering regardless of the subsequent nonclinical consequences, it has been suggested that physicians who are opposed to the death penalty not practice on death row. Similarly, it was held in the case of *State v. Perry* that the patient retains the right to refuse treatment and the state cannot treat over objection so as to impose the death penalty.

The NCCHC Standard "Inmate Executions" states that "Mental health staff does not participate in inmate executions" and sets out the following compliance indicators:

> *An inmate's treating clinician does not provide a competency evaluation for execution.*
>
> *Mental health services staff members do not assist, supervise, or contribute to the ability of another individual to directly cause the death of an inmate.*
>
> *Mental health staff members provide ongoing mental health services to inmates in death row as clinically indicated according to individual treatment plans.*

It is noted that the intention of this standard is that mental health staff not be involved in any action that would directly cause the death of an inmate because such involvement in inmate executions poses an ethical dilemma: "In order to be effective mental health care providers, correctional mental health care staff must maintain an appropriate professional relationship with their patients. This relationship is lost when mental health care staff assumes a role in inmate executions." The NCCHC Position Statement entitled Competency for Execution recommends that "the determination of whether an inmate is 'competent for execution' should be made by an independent expert" and not by correctional psychiatric staff.

Lesbian, Gay, Bisexual, and Transgender Persons and Gender Dysphoria. According to the APA Psychiatric Services in Correctional

Facilities, trauma and stigma experienced by lesbian, gay, bisexual, or transgender (LGBT), individuals can increase their psychological distress and decrease their willingness to seek mental health services. When incarcerated, they risk additional discrimination and abuse. Compared with heterosexual inmates, studies have reported that they have ten times the rate of sexual victimization by other inmates and up to three times the rate of sexual victimization by staff. This risk needs consideration on arrival at any correctional facility. LGBT individuals differ in degree of comfort with their sexual orientation or gender identity and whether they are “out” or “closeted.” Their choice about disclosure can change when moving from the community to jail or prison and between institutions. A respectful and nonjudgmental approach by the correctional psychiatrist can help LGBT inmates explore their choices and cope with their concerns. Transgender inmates, for example, may feel distress when housed in settings that do not allow them to express their gender identity or when they elect to stay closeted because of safety concerns. A therapeutic relationship in which they feel known by someone safe can improve emotional well-being.

The evaluation and treatment of gender dysphoria, formerly known as gender identity disorder, is a particularly challenging area. The evaluation, treatment, and decision-making processes of whether to continue, discontinue, or initiate hormonal treatment for offender patients are complicated, and even more so due to PREA audits and enforcement in corrections. Custody leadership makes decisions regarding classification, housing, transportation issues, and whether they allow access to gender-preferred clothing, toiletries, and housing assignments. Beyond hormonal treatment, many correctional systems do not have uniform policies about permissible treatment for transgender inmates. However, more recently, the courts have been ruling in favor of cases by inmates calling for sex reassignment surgery in the correctional setting, with some courts agreeing that denying such a surgery is a violation of the Eighth Amendment. Consistent with the prevailing position of professional medical organizations, including the APA, courts, and society in general, are increasingly recognizing gender dysphoria as a medical condition appropriate for treatment. The California Department of Corrections and Rehabilitation is credited as the first state system in the country to have implemented policies granting these surgeries.

Hunger Strikes in Jails and Prisons. Hunger strikes (the refusal of nutrition as a means of passive resistance) are a daunting problem in jails and prisons. Although the history of hunger strikes goes back at least to the Roman Empire and in the Western world, there is generally a right to go on a hunger strike that is protected as a matter of “privacy” or a “liberty interest” or can be regarded as an expression of free speech, clearly, in a correctional facility other factors come into play. Correctional authorities are charged with the “care and custody” of the inmates. Therefore, the following principles should apply:

- The correctional facility must provide for the evaluation of the situation for treatment if necessary and perhaps for legally allowable intervention.
- The medical services must provide for the evaluation of the medical stability of the inmate.
- The correctional psychiatrist must be available to evaluate such an inmate’s psychiatric condition and whether there is a need for treatment.
- The correctional psychiatrist must also evaluate the motivation for the hunger strike and the inmate’s capacity to understand the consequences of their actions and the medical decisions being made.

The majority of U.S. courts that have considered the question hold that an order to force-feed does not violate a hunger-striking prisoner’s constitutional rights because the institution in which the inmate is housed is responsible for his care while incarcerated, and there are other compelling governmental interests, such as the preservation of life, prevention of suicide, and enforcement of prison security, order, and discipline, that outweigh any constitutional rights asserted. It is interesting to note, however, that international authorities argue to the contrary. International courts have tended to allow hunger strikes to proceed even to death or disability. Thus, forced feeding of prisoners is prohibited by the Tokyo Declaration of 1975 by the World Medical Association and affirmed in 1992 by the Declaration of Malta.

In 1994, the ACHSA adopted a Position Statement that allows forced feeding when the legal procedure for medication over objection has been followed:

The management of a hunger-striking prisoner presents an ethical conflict between the obligation to preserve life and relieve suffering on the one hand and the obligation to respect patient autonomy on the other. Correctional health care providers should make every reasonable effort to honor patient self-determination. However, the deprivation of freedom that incarceration imposes will affect decision-making of the inmate patient. Therefore, special efforts in patient education and counseling are indicated.

Since forced feeding is a form of involuntary treatment, the responsible medical authority must seek the consent of the court or a legally appointed guardian, following the local requirements for involuntary treatment. The physician should make it clear to the court and to all others concerned; however, that ethical medical conduct in this setting demands medical intervention to preserve life from actions that are self-destructive.

ACHSA members should give careful consideration to the development of policies and procedures that provide a responsible and compassionate approach to the hunger-striking prisoner and which make it clear that efforts will be made to intervene with forced feeding in order to save a life.

The APA Psychiatric Services in Correctional Facilities recommends the following:

> *Hunger strikes are considered a medical emergency, and psychiatrists are often asked to evaluate inmates to determine if they have the mental capacity to make an informed decision regarding continuation of the hunger strike. Psychiatrists can perform evaluations to rule out underlying mental illness as a cause of the hunger strike (e.g., suicidal depression or delusions regarding prison food).*

There is currently no consensus to guide psychiatrists in the United States on the ethical standards of force-feeding to end hunger strikes in correctional facilities. Although the World Medical Association and the AMA, with limited exceptions, have taken stands against the practice, the Supreme Court of Connecticut, in *Commissioner of Correction v. Coleman* (303 Conn. 800 SC 18721 [2012]), found that the state's interest in the inmate's health and legitimate penologic interests in the safety of the facility outweighed the inmate's right to autonomy. Furthermore, the court ruled that "force-feeding did not violate constitutional rights to free speech and privacy."

THE FUTURE OF CORRECTIONAL PSYCHIATRY

In light of the significant numbers of individuals in our nation's jails and prisons, and due to the significant burden of mental illness and psychiatric needs in this vulnerable population, future directions for correctional psychiatry involve developing policies and protocols to ensure equitable treatment and access to care.

Diversion and Community-Based Interventions

There have been increasing efforts to pursue alternatives to incarceration of persons with severe mental illness and substance abuse disorders. One of the most promising alternatives is the diversion model, designed to prevent or minimize detention of the nonviolent mentally ill and to divert persons with severe mental illness to appropriate community-based mental health services in lieu of incarceration.

Interventions may occur at various points of interaction between individuals with mental illness and the justice system (Fig. 58.3–5). Pre- and postbooking diversion programs may hold promise for addressing overcrowding in jails and prisons and ensuring that individuals with mental illness are directed toward treatment.

Prebooking programs involve the utilization of first-response programming to respond to crises in the community. Examples might include specialized police response teams, mental health professionals hired by police departments, and local mental health crisis teams who often have relationships with local police departments. Other prebooking interventions might include collaboration between the above entities as well.

Postbooking interventions can be jail-based, involving recognition and assessment of detainees with mental health needs so that they can be transitioned to a community-based treatment program. In contrast, court-based diversion might involve assessments performed by courthouse clinicians who perform assessments and engage in discussion with the prosecution, defense, and a judge to advocate for release dependent on adherence to a community treatment protocol. Some individuals may be seen in specialized mental health courts, where involved individuals have specialized training in mental illness and will connect the individual to treatment, with follow-up to ensure adherence to the mandated treatment program.

Assisted outpatient therapy (AOT), also known as “outpatient commitment,” is also being studied and implemented to address the burden of serious mental illness in the community and in institutions. To date, 47 states have approved AOT programming (Connecticut, Maryland,

and Massachusetts are the three states that do not currently have AOT). AOT represents a means through which court-ordered treatment can be coordinated for patients with severe mental illness who have experienced recurrent hospitalizations, poor adherence to recommended treatment, and/or significant adverse consequences (such as violence or justice involvement) because of their mental illness. While principles of patient autonomy and beneficence undeniably necessitate continued efforts to engage patients in voluntary treatment, there are individuals who, due to severe mental illness, are unlikely to voluntarily engage in treatment for whom court-mandated treatment programs might be effective in reducing risk for adverse outcomes. In a revised position statement by the APA regarding these programs, it was recognized that involuntary outpatient commitment and AOT could be effective at promoting recovery when implemented systematically and for a sustained period of time, involving linkage to medication management and psychosocial services with proactive treatment models similar to those of assertive community treatment or intensive case management programs. The APA additionally recommends clear and consistent planning regarding consequences of treatment nonadherence in addition to ensuring that separate review and approval are obtained prior to consideration of involuntary medication administration in the community.

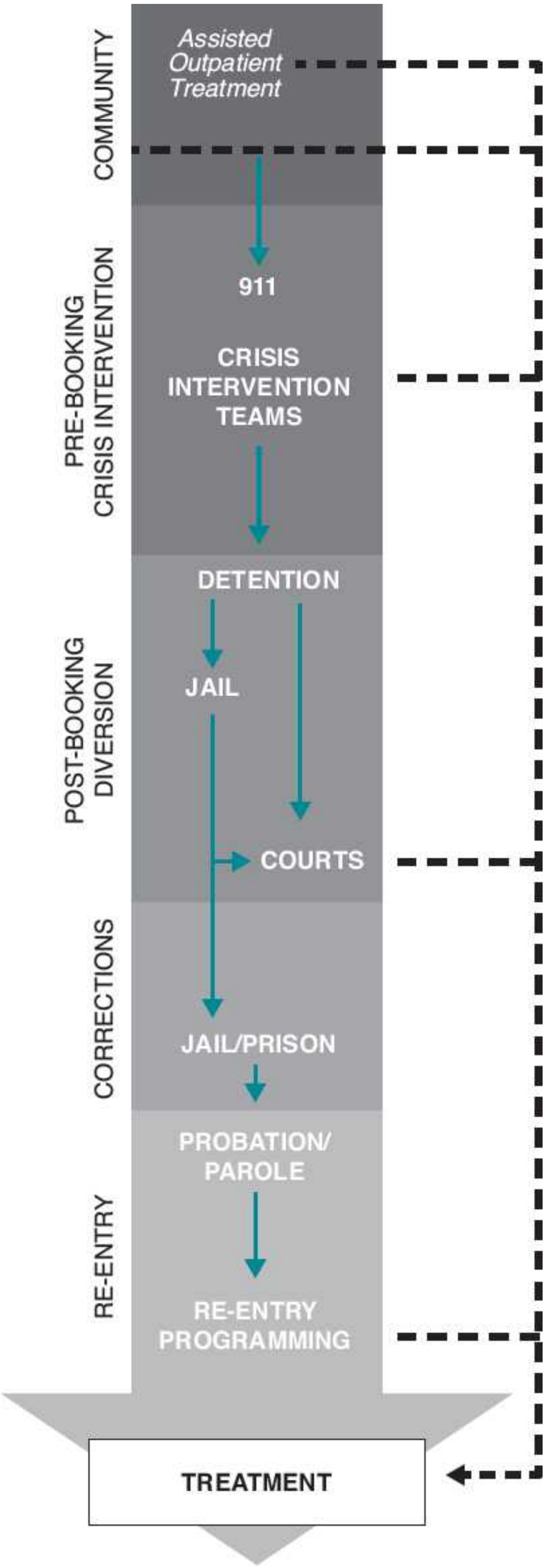
COMMUNITY
Assisted Outpatient Treatment
PRE-BOOKING CRISIS INTERVENTION
911
CRISIS INTERVENTION TEAMS
POST-BOOKING DIVERSION
DETENTION
JAIL
COURTS
CORRECTIONS
JAIL/PRISON
RE-ENTRY
PROBATION/ PAROLE
RE-ENTRY PROGRAMMING
TREATMENT

FIGURE 58.3–5. Mental Illness and the Justice System: Points of intersection.

Education

In 1996, the Accreditation Council on Graduate Medical Education (ACGME) recognized the subspecialty of forensic psychiatry, which includes substantial correctional components. The didactic curriculum for the forensic fellowship training requires that the “psychiatric curriculum includes the . . . evaluation and treatment of incarcerated individuals,” and the “clinical experience must include experiences in the . . . treatment of persons involved in the criminal justice system.” In addition, “Clinical placement must provide [forensic fellows] . . . with experience in the evaluation and management of acutely and chronically ill patients in correctional systems such as prisons, jails, community programs, and secure forensic facilities.”

However, though most psychiatrists do not choose to dedicate their careers to forensic psychiatry, the prevalence of justice involvement among individuals with serious mental illness necessitates all providers to feel comfortable managing the unique needs of this population. Despite the number of individuals receiving mental health care in correctional settings and despite the fact that many of these individuals will eventually transition their care back to community providers, only one-third of psychiatry residency training programs included a mandatory correctional rotation for all trainees. Furthermore, even in programs offering elective rotations in correctional psychiatry, very few residents elected to participate. As the field of psychiatry works toward meeting the immense mental health needs of individuals in jails and prisons, increasing trainee exposure to and awareness of correctional psychiatry will be crucial in developing solutions to this public health issue.

Using Technology in Corrections: The Electronic Medical Record and Telepsychiatry

Many county, state, and federal correctional systems have moved toward the implementation of correctional EMR systems. This is particularly helpful in multisite systems or high-volume intake facilities with rapid turnover, or alternatively in large systems where there is common movement, transfers, and turnover of inmates both intra- and intersystem. Thorough medical documentation is useful protection against litigation to help clarify an individual’s assessment and treatment plan and to ensure continuity of care.

Telepsychiatry is another tool with significant promise in relation to meeting the need for mental health care in correctional settings, with the COVID-19 pandemic highlighting the potential benefits of telepsychiatry. Greater comfort by correctional psychiatrists and other mental health professionals with this technology increases the accessibility of expeditious services for patients, and due to less need for off-site transport for psychiatric evaluation, this technology reduces the cost and burden on staff to travel with an offender. More recent evidence has demonstrated the feasibility and efficacy of telepsychiatry in correctional settings. Furthermore, recent legislation in some states requires payment parity for virtual psychiatric encounters, further increasing the sustainability of virtual care in correctional settings. Despite its promise, challenges associated with the use of telehealth technology may include difficulties with rapport building, lack of trust between patient and provider, decreased efficiency of communication, and lack of provider awareness of facility events contributing to the individual's presentation and treatment.

Advocacy

A confluence of forces is accelerating the reform of the correctional system to the benefit of correctional psychiatrists and their patients. Intense advocacy efforts, especially by family members of persons with serious and persistent mental illness in the criminal justice system, through patient advocacy groups such as the National Alliance on Mental Illness and Mental Health America, are becoming increasingly effective in lobbying legislatures to increase mental health funding, staffing, and psychiatric services in jails and prisons.

Advocating around improving mental health laws in each state is an important way to reduce the incarceration of the mentally ill. Mentally ill individuals engage in violence in rare instances. However, when they do, it is frequently in the setting of acute untreated mental illness and substance use. Because the treatment does not always follow the patient outside of hospital walls, building treatment programs monitored and supported by courts in lieu of jail time can be an important way of protecting the autonomy and freedom of mentally ill individuals and reducing the time they are in a locked setting. Better laws that support mentally ill individuals receiving proper treatment also destigmatizes both psychiatric illness and its treatment. The Treatment Advocacy Center is promoting research in this area to help the public better understand mental illness, the pathways to incarceration for the mentally ill, and solutions, through treatment, to reduce the institutionalization of this vulnerable population.

Last, the APA is joining with other professional groups to develop programs and priorities for legislative action, including such groups as the Campaign for Mental Health Reform and the Mental Health Liaison Group. This has, in turn, drawn the attention of county and state governments not only in terms of the requirements of the humane aspects of care and treatment and concerns about litigation but also the financial and societal benefits of preventing mental illness-associated recidivism through jail and prison mental health programming, diversion programs, drug courts, and mental health courts, jail-based and other competency restoration programs, and well-developed reentry programs.

ACKNOWLEDGMENT

This chapter represents an update and revision of previous chapters in this series, most recently by Henry C. Weinstein, MD and Joseph V. Penn, MD from the last edition. We are indebted to Drs. Weinstein and Penn for their work and retain much of their prose in this current edition.

FURTHER READINGS

American Psychiatric Association (APA) Official Actions. Position statement on psychiatric services in adult correctional facilities. Approved by Board of Trustees, July 2018. Approved by the Assembly, 2018.

Angell B, Matthews E, Barrenger S, Watson AC, Draine J. Engagement processes in model program for community reentry from prison for people with serious mental illness. *Int J Law Psychiatry*. 2014;37(5):490–500.

Baillargeon J, Hoge SK, Penn JV. Addressing the challenge of community reentry among released inmates with serious mental illness. *Am J Community Psychol*. 2010;46(3–4):361–375.

Batastini AB, King CM, Morgan RD, McDaniel B. Telepsychological services with criminal justice and substance abuse clients: a systematic review and meta-analysis. *Psychol Serv*. 2016;13(1):20–30.

Biswas J. When jails become hospitals: the fallout from a fragmented mental health system. *Crim Just*. 2017;32:4.

Booth BD, Mikhail E, Curry S, Fedoroff JP. Shaping attitudes of psychiatry residents toward forensic patients. *J Am Acad Psychiatry Law*. 2016;44(4):415–421.

Bureau of Justice Statistics. Special Report: Indicators of Mental Health Problems Reported by Prisoners and Jail Inmates, 2011–12, 2017. https://bjs.ojp.gov/content/pub/pdf/imhprpji1112.pdf

Carson EA, Sabol WJ. *Aging of the State Prison Population, 1993–2013*. U.S. Department of Justice Office of Justice Programs Bureau of Justice Statistics; 2016. https://www.bjs.gov/content/pub/pdf/aspp9313.pdf

Fellner J. *Old Behind Bars: The Aging Prison Population in the United States*. Human Rights Watch; 2012. https://www.hrw.org/sites/default/files/reports/usprisons0112webwcover_0_0.pdf. Accessed July 11, 2016.

Fovet T, Geoffroy PA, Vaiva G, Adins C, Thomas P, Amad A. Individuals with bipolar disorder and their relationship with the criminal justice system: a critical review. *Psychiatr Serv*. 2015;66(4):348–353.

Hayes LM. *National Study of Jail Suicide: 20 Years Later*. U.S. Department of Justice, National Institute of Corrections; 2010.

Hedden BJ, Comartin E, Hambrick N, Kubiak S. Racial disparities in access to and utilization of jail-and community-based mental health treatment in 8 US midwestern jails in 2017. *Am J Public Health*. 2021;111(2):277–285.

Kaba F, Solimo A, Graves J, et al. Disparities in mental health referral and diagnosis in the New York City jail mental health service. *Am J Public Health*. 2015;105(9):1911–1916.

Kaftarian E. Lessons learned in prison and jail-based telepsychiatry. *Curr Psychiatry Rep*. 2019;21(3):15.

Kang-Brown J, Montagnet C, Heiss J. *People in Jail and Prison in 2020*. Vera Institute of Justice; 2021.

Keram EA. Hunger strikes. In: Trestman RL, Appelbaum KL, Metzner JL, eds. *Oxford Textbook of Correctional Psychiatry*. Oxford University Press; 2015.

McGlasson T, Champion MK, Penn JV. Geriatric offenders: evaluation and treatment within correctional settings. In: Holzer J, Kohn R, Recupero P, Ellison J, eds. *Oxford Textbook of Geriatric Forensic Psychiatry*. Oxford University Press; 2017.

McKee J, Penn JV, Koranek A. Psychoactive medication misadventuring in correctional health care. *J Correct Health Care*. 2014;20(3):249–260.

Metzner JL, Tardiff K, Lion J, et al. Resource document on the use of restraint and seclusion in correctional mental health care. *J Am Acad Psychiatry Law*. 2007;35(4):417–425.

Mitchell A, Williams B. Compassionate release policy reform: physicians as advocates for human dignity. *AMA J Ethics*. 2017;19(9):854–861.

Moloney KP, van den Bergh BJ, Moller LF. Women in prison: the central issues of gender characteristics and trauma history. *Public Health*. 2009;123(6):426–430.

National Commission on Correctional Health Care. *Standards for Mental Health Services in Correctional Facilities*. National Commission on Correctional Health Care; 2015.

Penn JV, Thomas C. Practice parameter for the assessment and treatment of youth in juvenile detention and correctional facilities. *J Am Acad Child Adolesc Psychiatry*. 2005;44(10):1085–1098.

Pinals DA. Crime, violence, and behavioral health: collaborative community strategies for risk mitigation. *CNS Spectr*. 2015;20(3):241–249.

Sales CP, McSweeney L, Saleem Y, Khalifa N. The use of telepsychiatry within forensic practice: a literature review on the use of videolink—a ten-year follow-up. *J Forensic Psychiatry Psychol*. 2018;29(3):387–402.

Steiker CS. Mass Incarceration: Causes, Consequences, and Exit Strategies. 2011;1–6.

Thompson M. Gender, race and mental illness in the criminal justice system. *Corrections and Mental Health: An Update of the National Institute of Corrections*. https://www.cdcr.ca.gov/ccjbh/wp-content/uploads/sites/172/2020/06/Gender-Race-Mental-Illness_2_3.pdf

Torrey EF, Kennard AD, Eslinger D, Lamb R, Pavle J. *More Mentally Ill Persons Are in Jails and Prisons Than Hospitals: A Survey of the States*. Treatment Advocacy Center; 2010;1–18.

Trestman RL, Appelbaum KL, Metzner JL, eds. *Textbook of Correctional Psychiatry*. Oxford University Press; 2015.

U.S. Department of Justice. Prison rape elimination act final rule (28 CFR part 115). *Fed Regist*. 2012;77(119):37105–37232. http://www.prearesourcecenter.org/sites/default/files/library/2012-12427.pdf. Accessed July 11, 2016.

Wasser T, Sun A, Chandra S, Michaelsen K. The benefits of required forensic clinical experiences in residency. *Academic Psychiatry*. 2019;43(1):76–81.

Williams BA, Stern MF, Mellow J, Safer M, Greifinger RB. Aging in correctional custody: setting a policy agenda for older prisoner health care. *Am J Public Health*. 2012;102(8):1475–1481.

Yampolskaya S, Winston N. Hospice care in prison: general principles and outcomes. *Am J Hosp Palliat Care*. 2003;20(4):290–296.

59

History of Psychiatry

EDWARD SHORTER, PH.D., F.R.S.C.

Motto: “In the natural sciences old views and interpretations of observations are overtaken by new discoveries and retain historical interest only; in the humanities this is never quite the case. Psychiatry has one foot in each of these disciplines and our heritage is not something we can simply leave to the historians.” Julius Hoenig (1981). (Hoenig J. Psychiatric nosology. *Can J Psychiatry.* 1981;26(2):85.)

One of the earliest accounts of mental illness comes from the myths of the Greek gods. Hercules, under a delusive disorder, slays his children, believing them the offspring of another god.

One lad stretched forth his hands, crying, “Oh, dearest father, do not slay me; I am thy son.” But Hercules, “with savage rolling eyes,” as the chronicler tells us, “brought his club down upon the golden-haired head of the boy and broke his skull. The other children and the wife are sacrificed by the same murderous fury.”

The modern reader recognizes the diagnosis of delirious mania, which, like most other psychiatric ailments, is as old as the hills and built into the pathology of the human condition.

Across the millennia, there have been many other accounts of deluded and melancholy patients. Psychiatric illness has always been recognized as medical in nature. Yet psychiatry as a discipline arises only late in the 18th century. It was under the influence of the enlightenment that the prospect of relieving mental illness through institutional care was first broached. There had always been confinement, in the context of the noisome "hospices" that lodged indiscriminately the insane, the poor, the older adults, the medically ill, and the criminal alike. But the concept of dedicated mental hospitals stemmed from the late 18th century, surfacing first in Italy, then in France. In this manner, psychiatry became a specialty, not because its practitioners had special diagnostic or therapeutic skills but because they had the administrative abilities required to run these large institutions. For decades, the emphasis in international surveys of psychiatric facilities would be on architecture and layout rather than on diagnosis and treatment.

Before the late 19th century, there was no such thing as private practice psychiatry. Psychiatry as a discipline was based in mental hospitals, or "asylums"—a term first used in 1766 instead of "lunatic hospital"—and psychiatrists were called "alienists" on the grounds that the distracted mind had become alien. Ambulatory patients were seen in private practice by neurologists or general practitioners.

EARLY BEGINNINGS

The main storyline is the transition from traditional hospices to mental hospitals dedicated to treatment and recovery. Joseph Daquin, a psychiatrist in the Savoyard city of Chambéry, extolled in a work on The Philosophy of Madness (*La Philosophie de la Folie*) in 1792, the bringing of patients "from their fetid holes … to the freedom of a great but enclosed yard. Trained staff will accompany them everywhere and keep them in view; in the yard, warders will let them do whatever they like, simply ensuring that they do not escape, and neither endanger their own lives nor those of others. We will simply let them wander about at will for two hours, or in the interval between one meal and the next." This is not the first description of the therapeutic asylum but certainly the first that extolls the therapeutic advantages of freedom and exercise.

Italy

Most accounts of the history of psychiatry do not begin with Italy, a land that is largely silent on the historical radar. Yet, it was in Florence in the 1780s that modern psychiatry actually began. The key concept is "therapeutic institution," admitting patients in the hope of relieving their distress rather than protecting society from them. The first therapeutic asylum was launched in 1788 when Vincenzo Chiarugi ("KEY-ah-ROO-jee") opened the doors of the old Bonifazio Hospital, previously a typical hospice, exclusively to psychiatric patients. Here, he organized gardens for the patients, provided music, gave a structure to their day, and ensured proper meals and humane treatment. From 1793 to 1794, Chiarugi wrote a three-volume psychiatry textbook based on these experiences, *On Insanity (Della Pazzia)*, specifying that for psychological tranquilization, the patients should be kept in calm, quiet environments: "The sweet songs of gentle, moving music in many cases will represent the first-line treatment." As for stimulating treatments, the patients should undertake moderate exercise in the asylum and at home. In treating melancholy, Chiarugi deemed it "absolutely necessary that … [the physician] finds his way into the patient's heart, gaining his confidence and trust." Chiarugi called this "the psychological treatment" (la cura morale).

Italian psychiatry received a reformist bump-up in 1821 when Antonio Galloni became head of the venerable San Lazzaro Asylum in Reggio Emilia in the north of Italy. Founded in the 12th century as a leprosarium, in the 18th century Reggio Emilia went over to receiving psychiatric patients, and Galloni implemented many of the French reforms (see below), given that Italian psychiatry in these years stood much under the influence of France (later, of Germany). Reggio Emilia's later superintendents included Carlo Livi, who in 1871, at the beginning of his tenure, introduced Italy's first farm colony, and Augusto Tamburini (from 1877, after Livi's premature death), who began outpatient care with local families. Livi and Tamburini were among the leading figures in 19th-century Italian psychiatry, and Reggio Emilia stood as an international model for the villa system—rather than housing inpatients in giant barracks—and for the use of laboratories in patient care. Tamburini's watchwords were “*scienza e amore—studio e benevolenza*” (science and love, research and kindliness).

France

Modern psychiatry was anchored in France in two of the great Parisian hospices that Louis XIV founded in the 17th century: Bicêtre for men and the Salpêtrière for women. The seminal name in French psychiatry is Philippe Pinel, who in 1793 became chief physician of the Bicêtre hospice. Pinel is often seen (in ignorance of the contributions of Chiarugi) as the very founder of psychiatry itself. Yet the contribution of Pinel is difficult to overstate because his Medical-Philosophical Treatise on Mental Illness in 1801 (*Traité Médico-Philosophique sur l’Aliénation Mentale*) became the first international textbook and initiated the centuries-long task (not completed today) of perfecting the nosology of the field. In 1795, Pinel became the chief physician of the Salpêtrière, where he remained until his death in 1826. Pinel is also remembered as an early, influential advocate of humane care, as opposed to the chains-and-beatings approach that earlier characterized asylum life.

Pinel’s student Étienne Esquirol (“ES-kir-ole”), who in 1825, became chief physician of the French national asylum in the Paris suburb of Charenton, is seen as the virtual founder of social psychiatry, assessing social patterns in illness, and further refining Pinel’s nosology (see below). Esquirol inspired the Law of 1838 that obliged each district (department) in France to have a modern asylum; the law became worldwide a model of modern psychiatric care.

The United Kingdom

England. The first therapeutic asylum in England was St. Luke’s Hospital in London, founded by William Battie in 1750. Battie conceived admission to the hospital itself as having a therapeutic character, and in his *Treatise on Madness* (1758), he wrote,

> The regimen [in madness] is perhaps of more importance than in any distemper. It was the saying of a very eminent practitioner in such cases that management did much more than medicine; and repeated experience has convinced me that confinement alone is oftentimes sufficient, but always so necessary, that without it every method hitherto devised for the cure of madness would be ineffectual.

Yet, until the opening of the Maudsley Hospital in 1923, the anchor of English psychiatry was Bethlem Hospital ("Bedlam"), founded in 1247 as a base for the Crusaders, then gradually given over to the care of the insane. (In 1930, the hospital removed from the district of Southwark to suburban Kent, at which time the old building—one of several over the centuries—became the Imperial War Museum.) The hospital is indissolubly linked with the Monro dynasty of psychiatrists, of whom Edward Thomas Monro was in office from 1816 to 1855. Among the hospital's medical officers, or "apothecaries," was John Haslam, author of Observations on Madness and Melancholy (1809), the detailed cases of which are among the earliest patient reports in psychiatry.

In the first half of the 19th century, England led in the international movement to asylum reform, emphasizing "moral treatment" (psychotherapy), which began late in the 18th century at the Quaker Retreat in York, and the "non-restraint" system of patient care (abolishing chains and the like) that John Conolly initiated after he became superintendent of the Middlesex Asylum at Hanwell (a London suburb) in 1839. "The old system," said Conolly in 1856, in his Treatment of the Insane without Mechanical Restraints, "placed all violent or troublesome patients in the position of dangerous animals. The new system regards them as afflicted persons, whose brain and nerves are diseased, and who are to be restored to health, and comfort, and reason."

Scotland. Psychiatry in Scotland was dominated by the Edinburgh Royal Asylum (today the Royal Edinburgh Hospital), founded in 1813 at the Morningside estate, out of dismay at how badly Scottish poet Robert Fergusson was said to have been treated at the London asylum of "Bedlam." (He died there at 24.) The Royal Edinburgh was led over the years by a series of major figures in the history of British psychiatry, such as Thomas Clouston, *whose Clinical Lectures on Mental Diseases* (1883) remained for years the standard textbook.

Ireland. How direly Ireland required modern psychiatry emerges from testimony before the House of Commons in 1817:

> There is nothing so shocking as madness in the cabin of the Irish peasant, where the man is out laboring in the fields for his bread, and the care of the woman of the house is scarcely sufficient for attendance on the children. When a strong man or woman gets the complaints [severe psychiatric illness], the only way they have to manage is by making a hole in the floor of the cabin, not high enough for the person to stand up, with a crib over it to prevent his getting up. The hole is about five feet deep, and they give this wretched being his food there, and there he generally dies.

Until the foundation of the Irish Free State in 1922, Ireland was a colony of Great Britain, and Irish psychiatry rose simultaneously with English psychiatry. The Richmond Lunatic Asylum in Dublin (now St. Brendan's Hospital), the flagship of Irish psychiatry, was founded in 1816, and in 1857, Joseph Lalor became its first resident medical director. Given the absence of psychiatry in the medical curriculum until 1875, the only means of becoming a psychiatrist in Ireland, and in much of Scotland and England, was to serve as an assistant physician in an asylum. Irish clinicians of the day, such as Conolly Norman, a later superintendent of Richmond, favored the strong German emphasis on clinical psychiatry and found unhelpful the anatomical speculations of the era about causation (did the ovaries cause mental disease or not?).

German-Speaking Europe

"If French is the language of diplomacy, then German may be said to be the language of psychiatry," as Edward William Anderson, professor of psychiatry in Manchester, observed in 1962 after German had become a dead language internationally. In the years between 1850 and 1933, German was indeed the language of psychiatry, as it was the language of culture and scientific life in an enormous swath between Budapest (Hungarian psychiatrists all published in German) and Strasbourg, or between Hamburg and Vienna. Unlike England and France, where a single great city dominated the psychiatric life of the country, in German-speaking Europe, there was a profusion of small states and over 30 university psychiatric departments or psychiatric hospitals existed, most of which had scientific aspirations; by 1900, Heidelberg, Berlin, Munich, and Vienna had become world centers of psychiatric learning.

Austria-Hungary. Psychiatry in Vienna was symbolized by the famous Tower of Fools (Narrenturm), a roundish structure with 139 individual cells (that still exists today), built by Emperor Joseph II in 1784 as part of the Vienna General Hospital. Although it was soon overtaken by advances in care, it remained an emblem of the distinctiveness of Viennese psychiatry, which later gave birth to such figures as Sigmund Freud (psychoanalysis) and Manfred Sakel (insulin coma therapy—see below for both).

Germany (After 1871, United as the German Empire). In the early years, German psychiatry tended to be overshadowed by the French. Yet, there were spots of light. An important early figure was Wilhelm Griesinger, who, as an assistant physician to Ernst Albert Zeller in the Winnenthal asylum in the early 1840s, began to propagate the view that "mental disease is brain disease," a main tenet of later German biologic psychiatry. In 1845, Griesinger expressed his views in a textbook, *Mental Pathology and Therapeutics* (*Die Pathologie und Therapie der psychischen Erkrankungen*), a second edition of which in 1861, became the leading international textbook of biologic approaches to the field. In 1865, Griesinger moved from Zurich to Berlin as a professor of psychiatry at the Charité Hospital; here, he founded in 1867 the *Archives of Psychiatry and Nervous Diseases (Archiv für Psychiatrie und Nervenkrankheiten)* that was the equivalent of *the Archives of General Psychiatry* (renamed *JAMA Psychiatry* in 2013) today.

Several reformed psychiatric hospitals flourished in Germany early in the century, including Siegburg north of Bonn, founded in 1825 by

Maximilian Jacobi, who introduced to Germany the “no restraint” system—not shackling patients—and forms of psychotherapy. These reformed asylums also constituted centers of research as well as humane care, and in the days before the shift of scholarship to university psychiatric hospitals, an asylum such as Illenau in Achern (Baden), which opened its doors in 1842, under its inaugural director Christian FW Roller, was probably the intellectual center of German psychiatry; Roller tutored many clinicians who later became leaders of the field, including Richard von Krafft-Ebing (see below). In 1854, Jacques-Joseph Moreau (“de Tours”), after a visit to a number of German asylums, noted, “the great scientific movement” in German psychiatry. This was the beginning of a vast scientific take-off.

Scandinavia

In the history of psychiatry, the contributions of the Scandinavian lands have been largely neglected, partly for linguistic reasons, partly for cultural blinkeredness.

Denmark. New-style, reformed asylums arose early in the Nordic lands. Denmark had receptacles for the insane dating back centuries. The city of Copenhagen had traditionally confined the “lunatics” at a so-called plague house. Yet following the British bombardment in 1807, living conditions for these unfortunates became intolerable, and the city decided to place patients at a castle at Roskilde, not far distant from the town. The asylum, which began accepting patients in 1816, became known as St. Hans Hospital. But there were other Danish centers too. In the course of the century, neurosyphilis became a common reason for asylum admission. In 1857, Friedrich Esmarch and Peter Willers Jessen, at an asylum in the province of Schleswig (then belonging to Denmark), proposed in the *General Journal of Psychiatry* (*Allgemeine Zeitschrift für Psychiatrie*) that the mysterious fatal disease then called “progressive paralysis” seemed to be a late consequence of primary syphilis. A landmark in Danish psychiatry was the work of Copenhagen’s Carl Georg Lange, professor of pathology at the

University and considered Denmark's "first neurologist," on "periodic depressive conditions," considered by Lange to be a separate illness for which he recommended lithium therapy.

Norway. As elsewhere in Scandinavia, psychiatry in Norway rested heavily on the German tradition of biologic and genetic investigation. The first state psychiatric hospital in Norway, Gaustad Mental Hospital, in a town of that name near Christiana, was founded by Herman Major, the "first psychiatrist of Norway," in 1855. Important research came out of Gaustad, including superintendent Hans Evensen's landmark studies of schizophrenia (showing different outcomes). In 1934, Ivor Asbjörn Fölling, a biochemist at the Rikshospitalet in Oslo, the Norwegian national research hospital, discovered the metabolic cause of a disease called phenylketonuria (PKU), one of the so-called "inborn errors of metabolism" and a cause of mental retardation: It was a genetic inability to metabolize the amino acid L-phenylalanine, causing toxic amounts of phenylalanine to accumulate in the brain.

Sweden. Georg Engström founded the first reformed asylum at Vadstena in 1829; thus, as scholar Lennart Wetterberg put it, "drawing a line between barbarism and humanity in Swedish psychiatry." National mental health legislation in 1858 stipulated that an asylum must be led by a physician. The teaching of psychiatry in Swedish universities, among the earliest such teaching anywhere, began in Uppsala in 1741, with Carl von Linné, the great classifier, lecturing on medicine, including psychiatry. At the University of Stockholm, called the Karolinska Institute, Eric Godelius gave the inaugural lectures on medicine, which embraced psychiatry as well, beginning in 1811. At Karolinska, later a world-famous research institute as well as a university, Wilhelm Öhrström, the first professor of psychiatry and simultaneously head of the newly built Konradsberg psychiatric hospital, began lecturing in 1861.

Spain

Although an asylum was established at Valencia in 1409, the great clinical school of Spanish psychiatry that flourished in the Middle Ages had little continuity into the modern era. In the 19th century, Spanish psychiatry remained much under the influence of the French, and it was Madrid's Pedro Mata y Fontanet who established forensic psychiatry in Spain as the inaugural holder of the chair from 1843 when he helped introduce Esquirol's ideas. The overwhelming figure in modern Spanish neuroscience is Santiago Ramón y Cajal ("Ca-HAL"), who taught anatomy and pathology at the universities of Barcelona and Madrid. From 1888, Ramón established the neuron theory of brain histology and is generally acknowledged as the father of modern neuroscience (he received a Nobel Prize in 1906). His advice for a young investigator (1897, in Spanish) remains compelling reading for young scientists even today. Although he did not practice clinical psychiatry, many later prominent psychiatrists trained in the "Cajalian" phase of Spanish medicine (but note that he is generally referred to as Ramón, not Cajal, which was his mother's name). The main chairs of psychiatry were founded in the last years of the Second Republic (1931–1936), and the most important 20th-century clinical figure in Spanish psychiatry, Juan J. López Ibor, who originated the concept of "vital anxiety" (*timopatia ansiosa*), held the chair of legal medicine, toxicology, and psychiatry in Madrid from 1932.

United States

In 1751, the Society of Friends (the Quakers), assisted by Benjamin Franklin, founded in Philadelphia the Pennsylvania Hospital; it was the first institution in the United States to have a major psychiatric division, for "persons distempered in mind," opened by Benjamin Rush in 1793. Rush was one of the signers in 1776 of the Declaration of Independence and practiced as much general internal medicine as he did psychiatry in his years of service from 1783 to 1813. In 1812, Rush wrote the first American psychiatry textbook, *Medical Inquiries and Observations*, *upon Diseases of the Mind*—

and thus earned the sobriquet “father of psychological medicine in America.”

The first state hospital in the United States opened its doors in 1773 in Williamsburg, Virginia. Called “The Public Hospital for Persons of Insane and Disordered Minds,” it later became the Eastern State Hospital. Meanwhile, a series of private and charitable psychiatric hospitals sprang up on the model of the Pennsylvania Hospital. In 1817, the Quakers, alarmed at overcrowding in the Pennsylvania Hospital, founded a Retreat in nearby Frankford, Pennsylvania. In 1818, the psychiatric division of the Massachusetts General Hospital (incorporated in 1811) was founded at Charlestown in Boston and named (after a private donor) the McLean Hospital. The Retreat in Hartford, Connecticut, opened in 1824 and still exists today under the name Institute for Living (see below).

In 1771, a charter was granted to The Society of the New York Hospital, and in 1791, New York Hospital began accepting patients, including psychiatric; in 1808, it constructed a separate “lunatic asylum” for this purpose. At this point, the reforms of Pinel and Tuke had begun to arouse interest in the New World, and shortly proposals for “moral treatment” were circulating in New York as well. The trustees of New York Hospital, therefore, purchased a lot fronting the Bloomingdale Road in Harlem, and in 1821 the new structure named the Bloomingdale Asylum was opened, receiving private patients as well as those subsidized by funds from the Society of the New York Hospital. Years later, the Society decided that country air might be more salubrious, and in 1894 transferred the Bloomingdale patients to a new site in White Plains, Westchester, where it became the Westchester Division of The New York Hospital-Cornell Medical Center (see below for further developments).

The first large state hospital run on enlightened principles in the United States opened at Worcester, Massachusetts, in 1833, under the leadership of Samuel B. Woodward, an enlightened and forward-looking psychiatrist and close friend of Eli Todd at the Hartford Retreat. Following a lobbying campaign by Dorothea Dix, the hospital was enlarged in 1835, and on several subsequent occasions, this victory marked the beginning of Dix's career as an advocate on behalf of poor people with mental illness. Among the most influential of the early American state hospitals was the Utica Hospital, opened in 1843, part of the vast New York State system of hospitals, which served as a seedbed for successful hospital administrators in many other institutions.

The shift in U.S. psychiatry from large buildings with long corridors to small villas clustered in village-like colonies began in 1894 with the foundation of the Colony for Epileptics in Sonyea, New York ("Craig Colony" after Oscar Craig, president of the New York State Board of Charities); the first patient was admitted in 1896. With colony care, these mental hospitals adopted a more human face.

THE FIRST BIOLOGIC PSYCHIATRY

The intellectual content of this budding young discipline was overwhelmingly biologic. At the beginning, psychiatry had a profound physical orientation. The brain was seen as the seat of diseases of the mind, and treatments addressed the brain and the body as a whole. In 1672, Oxford physician Thomas Willis (who described the "circle of Willis") said of the brain's role in his famous *Two Discourses Concerning the Soul of Brutes*

> When it happens that these corporeal functions are defective or hindered, forthwith the eye of the intellect, as if covered with a veil, is wont to be very much dulled or wholly darkened . . . Because [of] their powers being hindered … the imbecility and dullness of the mind, the slowness of the ingenuity, stupidity and madness at length do often arise. The fault is either in the brain itself, or the animal spirits, or both together.

For many decades, investigations of the brain to ascertain the correctness of Willis's views were hindered by the lack of suitable stains for nervous tissue. Studies of inheritance became the main technique of biologic psychiatry. But in the 1870s, a new school of neuropathology, seconded with the aid of the microscope, began to arise, and the findings of such investigators as Theodor Meynert in Vienna and Franz Nissl in Heidelberg (see below) gave new impetus to biologic thinking.

Nerves

Intrinsic to biologic views was the assumption that disorders were caused by the nerves. Beneath the level of frank “insanity” or “madness,” depression, anxiety, obsessive-compulsive disorder, and other nonpsychotic complaints were thought nervous in nature. The view was that the nervous system was somehow physically affected and that the source of mental and behavioral symptoms lay in these ailing nerves. In 1780, Samuel Tissot of Geneva, whose practice included many patients with psychiatric disorders, said, “Among the most appropriate remedies for diminishing the irritation of the nerves, and to destroy numerous causes of their disorders, one must place the baths, cold or warm.”

By the mid-19th century, the concept of “nerves” had seized psychiatry in the way that “psychoneurosis” and “mood disorders” would later rule. English law distinguished clearly between insanity and nerves. As the asylum inspectors reported in 1844, “Homes which are licensed for the reception of Insane Persons ought to be kept exclusively for that purpose and the reception of nervous, imbecile, and dejected [depressed] persons, amongst those who are insane and often dangerous, is for obvious reasons open to serious objection.”

Psychiatrists were once called “nerve doctors,” a term that persisted in German-speaking Europe until relatively recently (Nervenarzt); a “nervous breakdown” crossed the threshold into psychosis, rendering patients incapacitated in a way that is associated with melancholic illness, acute psychosis, and delirious mania. In 1851, Henry Monro, an attending at St. Luke’s Hospital in London, said in his Remarks on Insanity, “The symptoms of insanity are simply the consequence of loss of nervous power.” He called insanity “simply a disease of deficient nervous vigour.” These words hallmarked a century of biologic psychiatry.

Predisposition, Degeneration

Psychiatrists have always believed that many mental illnesses can be passed on from generation to generation, that they are “inheritable,” or the result of “predisposition.” The psychoanalysts did not believe this and attributed reproducibility within families to “nurture” rather than “nature.” Yet, in the great sweep of things, psychiatry as a whole has believed in nature. As London’s John Haslam said in 1817, “The minute investigation of many hundred cases has fully satisfied my own mind that the offspring of a person who has labored under insanity, is peculiarly liable to become similarly affected.” By the late 19th century, it had become an accepted fact that the main cause of psychiatric disease was “heredity.” As Paris psychiatrist Ernest Billod toured Italian asylums in the early 1880s, he was struck at the mantra-like repetition of the phrase, “The predominant cause of mental illness is heredity.”

Early Pharmacotherapy

In psychiatry’s early period of biologic thinking, before the rise of psychotherapy, treatments of mental disease borrowed heavily from the “humoral” remedies employed elsewhere in medicine, mainly the therapies of “depletion”: Correcting humoral imbalances in the body by venesection (on which Philadelphia’s William Rush was very keen), leeching, and the administration of purgatives and laxatives. At the same time that depletion began to totter in the rest of medicine—toward the mid-19th century—it came under attack in psychiatry as well and was replaced by the whole world of physical and dietetic therapies. All such treatments rely mainly upon the placebo effect, but at least they are not, unlike depletion, actively harmful to the patient. As Woodward wrote in 1850,

> The abandonment of depletion, external irritants [mustard plasters], drastic purges and starvation, and the substitution of baths, narcotics, tonics and generous diet, is not less to be appreciated in the improved condition of the insane, than the change from manacles, chains, by-locks and confining chairs, to the present system of kindness, confidence, social intercourse, labor, religious teaching, and freedom from restraint. In this age of improvement, no class of mankind have felt its influence more favorably than the insane. (Woodward SB. Observations on the medical treatment of insanity. *Am J Psychiatry*. 1850;7(1):1–34.)

There was in the mid-19th century a brief but lucid anticipation of the era of psychopharmacology that would follow a hundred years later: In 1845, Paris psychiatrist Jacques-Joseph Moreau (nicknamed “Moreau de Tours,” after the medical school he attended) published the results of his therapeutic experiments with hashish. The essence of psychopharmacology is establishing differential effectiveness, and in this foundational work, Moreau established that hashish was ineffective in the treatment of dementia and stupor (*stupidité*), but that it did seem to have antimelancholic effects, lifting the mood and abolishing psychotic ideation in two melancholic patients.

As for subsequent pharmacotherapies, the opinion is widely held that psychiatry possessed no pharmacologic treatments before the psychopharmacologic revolution of the 1950s (see below). This is false. Psychiatry has always disposed of a palette of agents, admittedly of varying effectiveness (yet that is true today, as well). Until the Second World War, the Germans led the pack. In the area of medicinal innovation, in the years 1800 to 1930, Germany had 44% of the world’s total innovations, followed by France at 16% and the United States at 14. The German chemical industry began its take-off around 1860, dyes in the forefront, and from the dyes emerged the first generation of psychopharmaceuticals. The very first treatments, however, were as old as the hills.

Ethanol. People are accustomed to thinking of beverage alcohol as a source of substance abuse rather than a useful therapeutic agent. Yet before the Second World War, although the dangers of alcoholism were readily acknowledged, alcohol constituted an important treatment. In 1923, Alexander Westphal, director of the university psychiatric hospital in Bonn, said, “In milder cases of melancholia, alcohol as a sedative and hypnotic can yield good service.” The opinion was typical of the psychiatry of the day and not necessarily incorrect.

Opium. The milky juice of the unripe poppy plant *Papaver somniferum* in the treatment of melancholia goes back to the

Ancients and was revived in the 17th century by London physician Thomas Sydenham, who standardized a combination of opium and alcohol (sherry) in “laudanum,” or opium tincture. Practitioners of mental medicine began prescribing opium in the 18th century, and in 1788, William Falconer was found, in his discourse upon “the influence of the passions,” waxing enthusiastic: “Opium too, which has been introduced into practice of late years in the nervous fever, and frequently administered, and in pretty large doses, is well known to exert great cordial effects on the mind and spirits.” Tincture of opium does not seem to have been highly addictive because so much of the alkaloid is inactivated by stomach acids. But with the invention of the syringe in the 1850s, it became possible to inject opium, and the resulting epidemic of addiction caused many psychiatrists to back away from the poppy, which seems actually to be quite effective as an antimelancholic. Yet as late as the 1960s, Hoffmann-La Roche Laboratories marketed injectable opium (Pantopon) for indications including psychiatric.

Cannabis and Hashish. The derivatives of *Cannabis sativa* L have a long history of use—today quite forgotten—in psychiatric therapeutics. After Moreau’s work in 1845, hashish became widely prescribed. Auguste Voisin, the professor of psychiatry at the Salpêtrière in Paris, described hashish in 1883 as part of the panoply of agents, including potassium bromide and hydrotherapy which he deployed in the treatment of psychotic depression. Cannabis has a similar history of age-old use and was marketed as a drug in England in the 1930s under the brand name Indonad (in Germany as Indonal-Bürger). The cannabinoids became definitively stigmatized in the “war on drugs” of the 1970s and disappeared from psychological medicine.

Lithium. Lithium had drifted on and off the therapeutic palette for decades, if not centuries. Certain mineral springs in Texas containing lithium, for example, had a historic reputation as “crazy water.” In 1886, Copenhagen psychiatrist Carl Georg Lange proposed lithium in the treatment of melancholic depression. But the remedy did not catch on, as the concept behind it, a uric acid diathesis of mental illness, was soon abandoned. (On the later history of lithium, see below.)

Sedatives and Hypnotics

Medicine's encounter with sedatives goes back to 1857 when the salts of bromine ("bromides") were introduced. Among chemically synthesized sedatives, chloral hydrate followed in 1869 (and is still in use today). Paraldehyde was introduced in 1882, and a number of other synthetic sedatives were launched—all now long forgotten. The barbiturates, however, were one of the major drug classes of the 20th century. A mixture of uric acid and malonic acid, the first barbiturate, barbital sodium (Veronal), was marketed in 1903, phenobarbital (Luminal) in 1911, and then dozens of shorter half-life preparations, including amobarbital (Sodium Amytal) in 1924 and butabarbital (Butisol) patented in 1932. These agents became household names and were widely prescribed.

Despite this proliferation of agents for mild depression and anxiety, psychiatry lacked treatments for such major illnesses as schizophrenia, melancholic depression (with the exception of opium), and catatonia. In 1892, Daniel Cathell, a Baltimore physician whose practice was heavy with nervous patients, advised his colleagues,

> Should you ever have recourse to remedies intended to act chiefly through the mind, look your patient earnestly in the face … they will do double good.…You will not only find that almost anything will relieve some of these cases, but will be further surprised to learn that it has evoked great mental and spiritual enthusiasm, and that they are chanting its praise and vowing that they were cured of one or another awful thing by it.…Some, indeed, who seem to be magically benefited by doses of—nothing—will actually credit them with saving their lives. (Cathell DW. *Book on the Physician Himself: And the Things That Concern His Reputation and Success*, The Twentieth Century Edition. F.A. Davis; 1903:190.)

The jury is still out on whether this happens today. (This thread is continued below under "The Psychopharmacologic Revolution.")

PSYCHOTHERAPY, SOCIAL PSYCHIATRY

Early Psychotherapy

Although physicians have always practiced an informal kind of psychotherapy in the doctor–patient relationship itself, formal psychotherapy begins with Hippolyte Bernheim, professor of medicine at the University of Nancy in eastern France, in 1884, when Bernheim, in his book *On Suggestive Therapy under Hypnosis or in the Waking State* (*De la Suggestion dans l'État Hypnotique et dans l'État de Veille*) distinguished between hypnotic suggestion (which had been around for a century) and nonhypnotic suggestion, or a good talking-to. In 1904, Paul Dubois, professor of neuropathology in Berne, Switzerland, proposed "rational psychotherapy," a forerunner of today's cognitive-behavioral therapy (CBT), in his book The Psychoneuroses and Their Psychological Treatment (*Les Psychonévroses et leur Traitement Moral*). In 1903, Pierre Janet, professor of experimental psychology at the Sorbonne, proposed "psychasthenia" as the main mental disorder and advocated a kind of cathartic therapy for its treatment. These early systems were then overwhelmed by psychoanalysis, which by the 1920s had become the chief psychotherapy of psychiatry (see below), and were forgotten.

Psychoanalysis

Previous histories of psychiatry customarily gave great space to the history of psychoanalysis, or Freudian "depth psychiatry." Yet today, psychoanalysis has virtually disappeared from psychiatry, and such extended coverage seems pointless. For detailed accounts, the reader is referred to the author's own *History of Psychiatry* (1997) or to one of the frankly adulatory chronicles of Freud's life such as that by Peter Gay. In retrospect, psychoanalysis may be seen as a kind of hiatus in an ongoing current of biologic thinking in psychiatry that began with the old asylum physicians around the middle of the 19th century and revived with enthusiasm about drug treatments in the 1950s, resulting in the substantial displacement of psychoanalysis by psychopharmacology, neuroimaging, molecular genetics, and other forms of neuroscientific investigation in the 1970s. The following discussion is appropriately brief.

Psychoanalysis, a system of thought that attributes the origin of behavioral symptoms to conflicts within the unconscious mind, was originated by Vienna's Sigmund Freud in the late 19th century. Although *Studies in Hysteria*, the volume that Freud coauthored with Vienna family physician Josef Breuer in 1895, is commonly seen as the founding document of psychoanalysis, it merely presents psychiatry's most famous patient, "Anna O" (Bertha Pappenheim), in the context of abreactive therapies. Freud's seminal doctrines themselves come to blossom only in 1900 in The Interpretation of Dreams (*Die Traumdeutung*). In successive years, Freud slowly adumbrated the essentials of psychoanalytic treatment, explaining the role of the interpretation of dreams and of free association. Then in the early 1920s, Freud proceeded to trace the architecture of the psyche in the layers that he hypothesized: ego (conscious mind), superego (the source of morality, supposedly derived from oedipal fears toward the father), and the id (unconscious mind, harboring the sex drive and the death instinct).

Psychoanalysis as a movement began in 1902 when Freud founded a discussion group called the Psychological Wednesday Society, a group of curious physicians—many of them spa physicians and physiatrists in search of some kind of psychic bond to their patients. In 1908, this discussion group became the Vienna Psychoanalytic Society. In the same year, the first international congress of psychoanalysis took place in Salzburg, followed by a second congress in 1910 in Nuremberg, and a third in 1911, in Weimar. Meanwhile, the movement was plagued by schisms as early advocates such as Alfred Adler broke away, unable to accept Freud's doctrine of infant sexuality fully.

After the First World War, psychoanalysis became a major international movement. The dominant form of medical psychotherapy shifted from the former systems of Dubois and Janet to psychoanalysis. Psychoanalytic societies sprang up in the major western cities, and only with the Nazi seizure of power in 1933 (and Anschluss with Austria in 1938) was this expansion brutally braked.

As the movement leached out of Europe, the United States, especially New York, Washington DC, and Los Angeles, became the world epicenter of psychoanalysis.

In the New World, under the influence of such émigé analysts as Hungarian-born Sandor Rado in New York, depressive neurosis began to loom in psychoanalysis. Freud himself had little interest in depression because the patients wanted mainly to be left alone and not tell their dreams or free associate. But in New York, “depressive neurosis,” or “neurotic depression,” started to become the meat-and-drink of the Park Avenue analysts, as opposed to “hysteria,” “phobic neurosis,” and the other diagnoses that Freud himself had placed front and center. In 1931, at the request of Abraham Brill, Rado helped organize training at the newly founded New York Psychoanalytic Institute. After a disagreement with other members, in 1945, Rado headed the newly opened Columbia University Psychoanalytic and Psychosomatic Clinic for Training and Research, the first training institute within a university.

Once psychoanalysis had drifted into exile, the movement began to break into rival camps, such as ego psychology (the classic drive-defense tradition as continued in Anna Freud’s 1936, *The Ego and the Mechanisms of Defense*); Donald Winnicott and Melanie Klein’s object relations theory; and self-psychology (associated in particular with Heinz Kohut’s book *The Analysis of Self* [1971]).

In object relations, Melanie Klein’s *The Psycho-Analysis of Children* (1932) highlights the importance of introjection and projection as mechanisms. As Klein’s biographer John Arnold Lindon explained, “Projection alters the infant’s impression of his environment, and by introjection this changed picture of his environment influences what goes on in his mind. Thus an inner world is built up that is partly a reflection of the external one.”

After the 1970s, psychoanalysis became less quarrelsome and drifted even further from Freud’s original emphasis on infantile sexuality. Psychodynamic psychiatry (which is not the same thing as psychodynamic psychotherapy) agrees on the importance of unconscious conflict, psychic determinism, the role of development, transference, counter-transference, and resistance. The field is no longer opposed to medication (analysts once scorned “pills” as a hindrance to getting at the “root causes” of the complaint). Glen Gabbard at Baylor College of Medicine in Houston writes in 2009,

> A patient with a bipolar disorder who is denying that he has an illness and refusing to take lithium may ultimately have better compliance with the medication if the clinician explores the meaning of his denial and his reluctance to consider himself as someone requiring treatment.

Yet, despite the efforts of psychoanalysts to integrate findings from psychopharmacology and the neurosciences, the wind has slowly gone out of the field. There had been a time when the chair of virtually every major department of psychiatry in the United States had been a psychoanalyst. But by 1990, more than 100 of 163 residency programs for psychiatry had abandoned instruction in intensive psychotherapy. The percent of psychiatrists who were also psychoanalysts declined from around 10% in the 1950s to 2.7% by 1988. What once was the very intellectual core of the discipline of psychiatry has become a marginalized group of enthusiasts. As Arthur Yuwiler, a senior biochemist in the Department of Psychiatry at the University of California, Los Angeles (UCLA), reflected somewhat ruefully in 2004, on the exchange of the psychoanalytic paradigm for the biologic one, “I seemed to spend the first half of my scientific life arguing with psychiatrists that there was a brain as well as a mind and the last half arguing there is a mind as well as a brain.”

Nonpsychoanalytic Psychotherapies

Formal systems of psychotherapy began with Hippolyte Bernheim, an internist in the French provincial medical school at Nancy, in 1884, with his distinction between hypnotic and nonhypnotic “suggestive” therapies: the latter meant, essentially a good chat. Bernheim believed hypnosis to be an effective treatment. Jean-Martin Charcot, in Paris, however, thought hypnotizability to be merely a diagnostic marker of “hysteria.”

Group psychotherapy began with Scottish psychiatrist William Alexander Francis Browne at the Crichton Royal Hospital in

Dumfries in 1842 when he encouraged patients to stage amateur theatricals that amounted to a form of “mental therapeutics.” This strand was continued in 1922 in Vienna with Jakob Moreno-Lewy’s impromptu theater that was later tagged “psychodrama.” After Moreno immigrated to the United States in 1925, he introduced in 1931 group psychotherapy. Independent of Moreno, in these years, Louis Wender at Hillside Hospital and Paul Schilder at Bellevue Hospital, both institutions in New York, also began psychotherapy in groups. Closely related to group psychotherapy was the “therapeutic community,” which was birthed in England during the Second World War at the hands of Joshua Bierer and Thomas Main. (See section on social psychiatry.)

In the United States, the transition of psychotherapy from physicians to psychologists and psychiatric social workers began with psychologist Carl Rogers’ “client-centered therapy” after 1946. His book of that title appeared in 1951.

Family therapy is generally considered to have begun with Columbia University’s Nathan Ackerman at the founding in 1956 of the “family mental health clinic,” part of New York’s Jewish Family Service.

Social and Community Psychiatry

The main narrative arc in psychiatry is basically a first (19th-century) biologic psychiatry, then a psychoanalytic hiatus in the first half of the 20th century, followed by a second biologic psychiatry (late 20th century). Yet in the years 1940 to circa 1980, another current arose in psychiatry flowing separately from this main narrative stream: it was social and community psychiatry.

After the Second World War, the field became increasingly preoccupied with the social causation of illness and with group therapies for its treatment. A movement arose to treat patients in the context of the community rather than the asylum, and in 1946, D. Ewen Cameron, head of psychiatry at McGill University, created the first day hospital in the world at the Allan Memorial Institute in Montreal. In 1948, Joshua Bierer established the first in England, a "social club" at the Social Psychiatry Centre in London that he had founded two years previously. The movement incorporated, in 1959, Bierer's view that "Treatment must include the whole social environment of the patient and all his social relationships. He must be treated not only as a person but as part of a community." So influential did social psychiatry become in the United States that it influenced the removal of the National Institute of Mental Health (NIMH) in the years 1967 to 1992 from the National Institutes of Health, to protect, according to Joel Paris, head of psychiatry at McGill and author of *Fall of an Icon* (2005), "the social mission of the Institute from mainstream medicine." Many social psychiatrists also saw themselves as social reformers, on the grounds, as Paris continued, that "poverty and social disruption were the real causes of psychological symptoms." Therefore, "society, not the individual, should be the 'patient.'"

These ideas accorded well with the social activism of the mid-20th century. But as early as 1956, Ralph Gerard, a neuroscience leader at the University of Michigan, said presciently,

> We have all wished mental disease to be determined by sociological causes, because this situation seemed more hopeful from the therapeutic standpoint. The individual broken by the group could be healed by techniques of interpersonal interaction; but if he had bad genes, effective treatment was far in the future.

And indeed, in the United States, social psychiatry was rapidly eclipsed by mushrooming biologic perspectives.

Today, CBT is probably the most widespread form of medical psychotherapy and is associated with Philadelphia psychiatrist Aaron Beck from 1963. Popular as well is Interpersonal Psychotherapy, originated in 1967 by Yale psychiatrist Gerald Klerman and coworkers, especially for the treatment of depression. Also of interest is "dialectical behavior therapy" (DBT), a form of CBT emphasizing "mindful awareness," proposed by Marsha M. Linehan, a psychologist at the University of Washington, in 1999.

Yet while the number of psychotherapies proliferated, the percentage of psychiatric consultations in the United States that involved psychotherapy declined from 44.4% in 1996 to 1997 to

28.9% in 2004 to 2005. Clearly, the single-office psychiatrist with a primarily psychotherapeutic practice, once the very image of psychiatry with its emphasis on “the couch” and “root causes,” was giving way to the health care team with the psychiatrist’s primary purpose being diagnosis and the provision of medication.

FROM ASYLUM TO UNIVERSITY

In the last quarter of the 19th century, many national medical bodies began demanding that psychiatry be included in the medical curriculum. This created a demand for professors of psychiatry and for academic centers in which to train them. As a result of this curricular development, research began shifting from asylums to university psychiatric hospitals and academic departments of psychiatry. Investigators in many of these centers came to dominate the field, and the asylums became research backwaters, increasingly given over to warehousing the ever-growing numbers of patients rather than treating them.

German-Speaking Europe

Said British psychiatrist Malcolm Lader in 1990, “We always used to say if you go back far enough and look at German literature at the turn of the century, you’re sure to find that someone described panic disorder and everything else.” It was the closeness of the link between the academy and the psychiatric hospital that granted this extraordinary German celerity in psychopathology; much of this architecture of the delineation of syndromes still remains. This leap from asylum to the academy in German-speaking Europe sped forward in 1874 when for the first time, university professors began taking part in the Association of Southwest German Alienists, hitherto a group of influential asylum psychiatrists.

Vienna. The capital of the Austro-Hungarian Empire, Vienna was a psychiatric super-power, the global influence of which was far disproportionate to the medium size of the city. The train of distinguished psychiatrists began in the late 18th century with such figures as Franz Joseph Gall (who first speculated about cerebral localization) and continued into the 19th with Ernst von Feuchtersleben, whose 1845 Textbook of Medical Psychiatry (*Lehrbuch der ärztlichen Seelenkunde*) popularized the term "psychopathy," in the sense of a disordered personality. (Feuchtersleben also distinguished between "neurosis" and "psychosis," giving them exactly the opposite meaning they were later to have.) But the modern tradition of Vienna as a bastion of biologic psychiatry (psychoanalysis came later) began with psychiatry professor Theodor Meynert ("MY-nert") with his great 1869 treatise on neuroanatomy, *The Architecture of the Cerebral Cortex (Der Bau der Gross-Hirnrinde)*, and continued with *Illnesses Involving the Forebrain in 1884 (Erkrankungen des Vorderhirns)*, which was completely forgotten midst later shrieks of jubilation about the importance of the frontal and prefrontal lobes in schizophrenia. His successor Richard von Krafft-Ebing is remembered for his textbook of sexology Psychopathia Sexualis (1886); his successor, in turn, Julius Wagner von Jauregg ("YOW-reg"), introduced in 1917, the malarial fever cure, an epochal accomplishment for which he received the Nobel Prize in 1927. Aside from the professors, Josef Berze, head of the Vienna city asylum "Am Steinhof," described "primary insufficiency of psychic activity" in schizophrenia in 1914, and Constantin von Economo, who led a brain research institute, reported the new disease of "encephalitis lethargica" in 1917. It was in the years following 1930 that Manfred Sakel proposed insulin coma therapy (while still in Berlin; only in 1933 did Sakel return to Vienna to conduct clinical trials—see below). Finally, pediatrician Hans Asperger described "autistic psychopathy" in 1944, among the earliest descriptions of autism (which was first used as a term in pediatric psychiatry by German and Russian neurologists and psychiatrists early in the 1920s). Vienna's contribution to psychoanalysis would begin in 1900, with the publication of Sigmund Freud's *Interpretation of Dreams* (see above), and continue until the expulsion or murder of the Jewish psychoanalysts in 1938 after Anschluss with Germany. After the Second World War, Vienna never again regained this historic preeminence in psychiatry.

Heidelberg. The flagship of academic psychiatry in the German Empire berthed in Heidelberg. The “Psychiatric-Neurological Clinic” of the University (“Klinik” in German means university hospital) was founded in 1879 by Otto Fürstner; it was the first university psychiatric department with beds of its own so that patients didn’t have to be brought from a nearby asylum for demonstration in clinical lectures. Its most distinguished head was Emil Kraepelin, who had the professorship from 1891 to 1903, during which time the crucial editions of his textbook appeared (see below). The Klinik was also a provincial asylum with a catchment basin of its own. When Kraepelin left for Munich, the professorship was assumed by neuropathologist Franz Nissl, known for his demonstration in cells of the “Nissl substance.” At various points, Kraepelin’s Klinik served Alois Alzheimer (who followed Kraepelin to Munich) and Karl Jaspers. In the 1920s, such important figures as Willi Mayer-Gross (who after fleeing the Nazis to England coauthored the standard English psychiatry textbook, *Clinical Psychiatry*, in 1954); Carl Wilmanns (pioneering studies of the homeless); Hans Gruhle (premier schizophrenia researcher of his day); and August Homburger (who founded in Germany modern pediatric psychiatry). This wall of scholarship crumbled in 1933, and the names of the successors under the Nazis are either forgotten or live on in infamy. Heinz Lehmann later said in retrospect, “German medicine was an impressive thing and there is no doubt about it. When we all went away, I don’t think this tradition was transferred to America but it was abolished in Germany.”

Berlin. In 1795, a psychiatry department was founded in the Charité, the Prussian state’s great military hospital, and in 1806 Ernst Horn was called to become associate director of the hospital, with simultaneous responsibility for psychiatry. Under Horn, psychiatry at the Charité conformed to the concepts of the day of a proper reformed asylum, though Horn was much assailed for some of his practices and left under a cloud in 1818. Ten years later, he was succeeded by Carl W. Ideler, one of the so-called “Romantic” psychiatrists who saw illness as evidence of sin. This unhappy period came to an end with Ideler's death in 1860 and Wilhelm Griesinger's arrival in 1865 to convert the old psychiatric ward into a modern neuropsychiatric clinic (Klinik für Geistes- und Nervenkrankheiten). In 1867, Griesinger founded the lead German journal for biologic psychiatry, the *Archive for Psychiatric and Neurological Diseases* (*Archiv für Psychiatrie und Nervenkrankheiten*, of which the current *Archives of General Psychiatry*—now called *JAMA Psychiatry* is a distant echo). Griesinger died prematurely in 1868 of acute appendicitis. Following Griesinger came a distinguished series of biologically oriented psychiatrists, including Carl-Friedrich Westphal in 1869 (who in 1872 described agoraphobia and in 1877 obsessive-compulsive disorder). His tenure ended with his death in 1890 of what seems to have been neurosyphilis—of which so many middle-class men perished in those days (infected as medical students or law students with primary syphilis from prostitutes). Karl Bonhoeffer had the Berlin chair from 1912 to his emeriting in 1937, after which another grim period began in the history of German psychiatry. In 1909, Bonhoeffer conceived the notion of "exogenous psychosis," arising from stress or toxins from outside rather than being endogenous and inborn.

In a separate development, in 1902, Oskar Vogt founded a neurobiologic institute at the University, which in 1915 was joined with the Kaiser-Wilhelm-Institute for Brain Research. (His wife, the French-born Cécile Vogt, became an important neuroscientist, later settling in England. Their daughter Marthe discovered in 1965 that noradrenalin was present in the brain.) In 1931, the institute was transferred to the Berlin-Buch mental hospital and came in 1937, under the direction of Hugo Spatz, who was a member of the Nazi party but who goes down in history as co-discoverer of the Hallervorden–Spatz disease of neurodegeneration. Despite its vicarious history, this institute became one of the world's leading neuroscience centers.

Munich. When in 1903, Kraepelin went from Berlin to Munich, it was as the inaugural director of a newly built University Psychiatric Hospital in the Nussbaumstrasse. He proceeded to make it into a world-class institution, bringing, as stated, Alois Alzheimer from Heidelberg to head the pathology laboratory, to which young investigators streamed from all over the world.

In 1917, Kraepelin founded a freestanding psychiatric research institute, the Kaiser-Wilhelm-Institut für Psychiatrie. The psychiatric department of the adjacent Schwabing City-Hospital, headed after 1931 by Kurt Schneider, became its clinical arm. Slow to get off the ground, the Institute’s take-off began only in the mid-1920s, thanks to a group of foreign donors led by James Loeb (after whom a street in front of the present institute is now named). In 1917, Ernst Rüdin became head of the “genealogic-demographic” (genetics) section (see below). The scientific promise of the Institute was nipped in the bud by the Nazi takeover, and only after the war, when in 1954 it became the Max Planck Institut für Psychiatrie, did its international reputation blossom.

Zurich. In Zurich, a city psychiatric hospital was founded in 1860, and when Wilhelm Griesinger began lecturing there in 1863, it became a “Klinik.” Two years later, in 1865, Griesinger decamped for Berlin. In 1869, a grand new Klinik was planned at a spot called the “Burghölzli”; the “Bli,” as it was known, was helmed over the years by some important figures, including Auguste Forel (1898–1927), a co-originator of the “neuron theory” of cerebral communication; Eugen Bleuler (1898–1927), who coined the term “schizophrenia”; and his son Manfred Bleuler (1942–1969), who in 1954, determined that the prognosis of schizophrenia was often benign.

France

Paris is the great focus of French psychiatry, and in Paris, each of the main general hospitals had academic psychiatry departments chaired by professors. Yet in addition, Paris had one special institution not found elsewhere in France (or in other lands, for that matter): The special psychiatric reception service of the Prefecture of Police (la Infirmérie Spéciale des Aliénés de la Préfecture de Police).

In 1800, Paris was given a special prefect for police, and with the general mental patients law of 1838, this prefect acquired responsibility for mental patients who came to public attention. Seven years later, in 1845, this special reception service became headed by a psychiatrist. From this service flowed a good deal of the research of modern French psychiatry. The first head of note was Ernest-Charles Lasègue in 1850, who coined the diagnostic term “delusions of persecution” in 1852, and “hysterical anorexia” (anorexia nervosa) in 1873. (Lasègue also had a series of academic appointments in the teaching hospitals.) In 1872, the Infirmérie Spéciale became independent of the police headquarters and turned into an important center for research. Following his death in 1883, Lasègue was succeeded by his assistant Henri Legrand du Saulle, who in 1875 offered an early description of obsessive-compulsive disorder (*la folie du doute avec délire du toucher*).

At the Salpêtrière hospice in 1882, Jean-Martin Charcot became the inaugural holder of a chair in nervous diseases, created especially for him by the president of the Republic. Charcot’s contributions lay mainly in neurology, but his name is indissolubly connected with a certain kind of “hysteria,” involving supposed chronic stigmata and elaborate acute pantomimes that in retrospect were phenomena of suggestion and that disappeared after his death in 1893. Among the students of Charcot, best known is Joseph Babinski, who in 1896 described the plantar extension reflex that is named after him. Charcot’s successors Fulgence Raymond and Jules-Joseph Dejerine were better known as neurologists than psychiatrists.

The “clinical asylum” Sainte-Anne was opened in 1867 as the asylum of the Seine department, with Valentin Magnan, the notable advocate of “degeneration” theories, as director of its admissions department. In 1877, Benjamin Ball became the inaugural holder of the Chair of Mental and Brain Disease. Among his successors were Henri Claude in 1922, an early biologic psychiatrist, and Jean Delay in 1946, who introduced the psychopharmacological era in France. The facility became a “psychiatric hospital” in 1937, a “psychiatric center” in 1944, and a “hospital center” (shedding the unfashionable adjective “psychiatric” entirely) in 1974.

United Kingdom

London. In 1908, the asylums committee of the London County Council accepted a gift from psychiatrist Henry Maudsley to found a research hospital based on the model of the German university psychiatric clinic. The war interfered with the realization of these plans, but in 1923, the Maudsley Hospital opened its doors for psychiatry with Edward Mapother, its inaugural superintendent. The Maudsley was not to be an asylum. Mapother made clear, as historian Edgar Jones relates (in a 2003 article in Medical History), “that no patient was to be admitted under section, nor would they be certified once in the hospital.” Mapother had an academic appointment at King’s College Hospital and wanted to make the Maudsley the psychiatric arm of the college’s teaching. (Mapother received the title “professor” in 1936.) Aubrey Lewis, who joined the staff in 1928, becoming a medical officer in 1936, recalled, “The place was small, full of ambition and energy, but everybody was overworked.” In 1946, Lewis was appointed professor of psychiatry, and two years later, in 1948, the teaching unit at the Maudsley was renamed Institute of Psychiatry (IOP). In that year as well the Maudsley merged with the Bethlem Royal Hospital (the former “Bedlam”). Lewis began adding Medical Research Council (MRC) units; the first of these, the MRC Social Psychiatry Unit under Lewis himself opening in 1948. These helped make the Maudsley the premier institution in the United Kingdom for research in psychiatry. In 1997, the IOP split from the Maudsley to become a school of King’s College London. The Maudsley was a major font of social psychiatry in the United Kingdom. Psychiatrist David Healy recalled that Aubrey Lewis had stated, at a meeting in 1958 (in Healy’s words), “‘If we had a choice between the new drugs and the social treatments, such as industrial rehabilitation units, we would pick social treatments.’ And it’s interesting that in the UK under Lewis’ influence, the Maudsley remained very aloof from the new drugs.”

Accordingly, the first center of research for psychopharmacology in the United Kingdom—the first anywhere—was the Department of Experimental Psychiatry, founded by Joel Elkes at Birmingham in 1951. It was here that some of the earliest trials of the antipsychotic chlorpromazine were conducted. “I felt like a naturalist advancing into a strange continent,” said Elkes later. The notion that drugs, as opposed to psychotherapy, could prove curative in psychiatry was then very new.

On Scotland and Ireland (See in Early Beginnings Section)

Canada—McGill University

Canada contained one of the most internationally important departments of psychiatry at McGill University in Montreal. The Department of Psychiatry of McGill was the Allan Memorial Institute (named after a donor), founded in 1940 at the instigation of neurosurgeon Wilder Penfield. Chaired by D. Ewen Cameron from 1943 to 1964, it was under Cameron that the “Allan” became one of the most distinguished psychiatry departments in North America. Its clinicians and scientists included psychoanalyst Miguel Prados, who arrived from Spain in 1944; the neuropathologist Karl Stern who in 1944, established a gerontologic unit, the first in the world; the neuropathologist Vojtech Adalbert Kral, a Holocaust refugee who arrived at the Allan in 1953 and pioneered research in dementia; Clifford Scott, who became an international leader of psychoanalysis, arrived from the Maudsley in 1954. (One wag said, “Kral felt if you couldn’t see it, it didn’t exist and Clifford Scott … felt if you could say it, it wasn’t important.”)

Included as well were: Neurophysiologist Robert Cleghorn in 1946, who pioneered endocrinologic psychiatry; Charles Shagass in 1950, who conceived the notion of the “sedation threshold”; Eric Wittkower, who in 1955 set up the section on transcultural psychiatry as a joint venture with the department of anthropology; Wittkower recruited Henry BM Murphy, author of the standard textbook of transcultural psychiatry (Comparative Psychiatry, 1982), who became head of the section in 1965 after Wittkower retired; and Heinz Lehmann, who came to Montreal in 1947 as director of the Verdun Protestant Hospital and is considered one of the founders of psychopharmacology. (The World Health Organization's first training center for psychopharmacology was at the Douglas.) Psychiatrist George Simpson, a first-year resident at McGill in 1956, later said in an interview, "I would argue that the Allan was the most stimulating place in the English-speaking world at that moment, even though the rest of the city spoke French."

United States

After the Second World War, the United States came to dominate world psychiatry, partly as a function of the torrent of research funds from NIMH, partly as a result of the Diagnostic and Statistical Manual of the American Psychiatric Association becoming after 1980, the world bible of psychiatry. The Holocaust and the devastation of the Second World War also had a great deal to do with the new American primacy. Rather than flocking to the Maudsley or Vienna for training, as budding American psychiatrists did before the war, it was young psychiatrists from all over the globe who sought further training in the United States. This new American hegemony accounts for the length of the following section.

New York. New York held the greatest concentration of psychiatric firepower in the United States owing to a series of public and private teaching hospitals.

HILLSIDE HOSPITAL. Hillside Hospital, initially in Hastings-on-Hudson, was founded by Israel Strauss in 1927 as a charitable Jewish hospital operated in a nonsectarian manner. In 1941, the hospital removed to Glen Oaks, Queens. It was, as Strauss put it, "the first institution of its kind for Jews in this country." Hillside specialized in "borderline" cases, meaning the kind of psychoneurosis that usually would not receive in-patient treatment. In 1930, Louis Wender, chief physician, began one of the first programs of group psychotherapy in the country. In the 1940s, Hillside was among the earliest U.S.

institutions to offer the various shock therapies. Although mainly psychoanalytic in inspiration, in 1950, Hillside resolved upon a “biochemical” research program, and in 1952 hired Max Fink to establish the Department of Experimental Psychiatry, which was joined by Donald Klein in 1957. Fink and colleagues set up an electroencephalogram (EEG) laboratory in 1953, and with Klein, conducted pioneering trials in psychopharmacology. By 1999, following several mergers, what had become Zucker Hillside Hospital was merged into the North Shore Long Island Jewish Health System, which encompassed seventeen hospital systems.

NEW YORK/PRESBYTERIAN/WESTCHESTER DIVISION. The foundation and growth of the New York Hospital, the second oldest freestanding psychiatric hospital in the United States, is discussed above. In 1898, the hospital became associated with Weill Cornell Medical College as that institution was founded, and a 1927 endowment lent to the psychiatric division the title Payne Whitney Psychiatric Clinic. New York Hospital merged with Presbyterian Hospital in 1996, and the Payne-Whitney Clinic was renamed the New York/Presbyterian Westchester Division.

NEW YORK STATE PSYCHIATRIC INSTITUTE. In 1896, the New York State Pathological Institute was established in New York City under Ira Van Gieson. It was the first effort in the state to study the anatomical basis of insanity, receiving specimens from the various state hospitals. In 1904, the Swiss-born Adolf Meyer, previously a pathologist at several state hospitals and a professor at Clark University, succeeded him. Meyer was already a significant scientific figure and would go on in 1910 to become a professor of psychiatry at Johns Hopkins University and inaugural head of the Henry Phipps Psychiatric Clinic (see below). In 1902, the Pathological Institute moved to Ward’s Island next to the Manhattan State Hospital. Meyer changed the name to Psychiatric Institute in 1907. Following a fire at Manhattan State, planning began to convert the institute to a state mental health research hospital and to situate it at Washington Heights, where Columbia was building a new medical school, and

Presbyterian Hospital was to move. The new building of the now New York State Psychiatric Institute was dedicated in December 1929. Known as “PI,” it became the psychiatry department of Columbia University as well offering from 1931 a 10-week course for psychiatrists in the New York State Mental Hospital system. In 1936, Nolan DC Lewis, professor of neurology in the Neurological Institute, became the first academic to lead PI. He, in turn, was followed by Lawrence C. Kolb in 1954. The Psychiatric Institute innovated at all points on the treatment spectrum: In 1942, a Department of Research Psychiatry was created, with Eugene Barrera and Bernard Pacella, its main members. In 1945, a Psychoanalytic Center for Training and Research was founded, the first official psychiatric training institute to be established in a medical school. Despite the massive interest at PI in psychoanalysis, early in the 1950s, a small number of investigators led by Philip Polatin became involved in psychopharmaceutical trials; by 1956, the Institute had a department of experimental psychiatry, now led by Sidney Malitz, located at the Vanderbilt Clinic.

PI benefited greatly from the tumult of the century, giving refuge to many distinguished scientists and clinicians fleeing the Holocaust. Among the members of the informal “European Club” were Lothar Kalinowsky (from Germany and Italy), the padrone of electroconvulsive therapy (ECT) in the United States; neurochemist Heinrich B. Waelsch (Prague); Paul Hoch (Hungary and Germany), who headed experimental psychiatry and coined the diagnosis “pseudoneurotic schizophrenia”; Sandor Rado (Hungary) who became a leader of psychoanalysis in the New World; Armando Ferraro (Italy), who headed neuropathology at PI. Eric Kandel, who established the Center for Neurobiology and Behavior in 1974, was born in Vienna and later (2000) won a Nobel Prize.

The East Coast

WASHINGTON DC: THE NATIONAL INSTITUTE OF MENTAL HEALTH. The NIMH, funded by the federal government as one of the National Institutes of Health, is not a degree-granting institution, but it is very academic, encouraging original research both in-house, in the "intramural program," and by funding "extramural" psychiatric research across North America. The involvement of the federal government with psychiatric research began in 1929 with the founding of the Narcotics Division of the Public Health Service, renamed Division of Mental Hygiene in 1930. Lawrence Kolb became the chief PHS officer in charge of the division of mental hygiene, and on his retirement in 1944, he had wished to set up a national neuropsychiatric institute (NPI) within the National Institute of Health (NIH). This became possible in 1946 when the National Mental Health Act called for the creation of a National Institute of Mental Health with an in-house research program. NIMH opened its doors in 1949, and Robert Felix, who had succeeded Kolb as director of mental hygiene, became the institute's inaugural director in 1951, with physiologist Seymour Kety in charge of research.

In 1953, the first clinical NIMH ward opened at the Clinical Center for the NIH; there followed in 1955 the Laboratory of Clinical Science and in 1956 the Psychopharmacology Service Center, which in 1960 organized the Early Clinical Drug Evaluation Program, Jonathan Cole, its director. The program became a great engine of progress, conducting investigational trials in the 35 Early Clinical Drug Evaluation Units (ECDEUs) that came to be established by 1969. By 1982, among the 100 investigators most frequently cited in scientific research worldwide, 9 were in the NIMH intramural program.

JOHNS HOPKINS UNIVERSITY: THE PHIPPS PSYCHIATRIC CLINIC. Psychiatry at Johns Hopkins began in 1910 when Adolf Meyer was hired from the Pathological (later Psychiatric) Institute of the New York State Hospitals to become the inaugural professor of psychiatry and director of the new university psychiatric hospital, the Phipps Clinic, funded by Henry Phipps, which opened its doors in 1913. On his retirement in 1941, Meyer was succeeded by Nebraskan John Whitehorn in office until 1960. In these years, Rockefeller Foundation grants brought many young European psychiatrists to the Phipps to study, and it was through them that Meyer's international influence became considerable.

MCLEAN HOSPITAL (CONTINUED FROM ABOVE). In 1888, profiting from the scientific boom then sweeping U.S. psychiatry, a pathology department was organized at the McLean asylum. In 1895, the institution removed from Charlestown to Waverley (Belmont), a suburb of Boston, and was named McLean Hospital. Swiss-born August Hoch, who would shortly study with Kraepelin and Alzheimer, became leader of a chemical laboratory in 1893; Hoch became an early advocate of the new neurohistopathology; a psychology laboratory followed in 1904. After 1955, in keeping with the changing times, the hospital pioneered in the U.S. concepts of the therapeutic community under Alfred H. Stanton.

But biology lay waiting. In 1972, the “transformation of McLean,” as Alan Schatzberg later put it, began when Shervert Frazier, who had also qualified as an internist, came up from Columbia University to become a director. He recruited psychotherapist John Gunderson in 1973, who became the “borderline” specialist; Jonathan Cole and Schatzberg came in 1974 to work in depression; there arrived as well from Massachusetts General Hospital (the parent organization), Joseph Lipinsky in 1974, and Seymour Kety and Ross Baldessarini in 1977. Schatzberg said, “He recruited outstanding people and built the Mailman Research Center, an unbelievably vibrant place.”

HARVARD UNIVERSITY. The Department of Psychiatry at Harvard was founded in 1920 when Adolf Meyer’s student, the Scottish-born Charles Macfie Campbell, left the Phipps Clinic in Baltimore to become director of the Boston Psychopathic Hospital and the inaugural professor of psychiatry at Harvard. Like his mentor, Campbell leaned toward “psychobiology” and psychoanalysis. The influential post at Harvard, however, is chief of psychiatry at the Massachusetts General Hospital, one of Harvard’s teaching hospitals (of which McLean was administratively a part). This post was initially held in 1934 by Stanley Cobb, professor of neuropathology and previously director of the Harvard Neurological Unit at the Boston City Hospital. Some consider Cobb the founder of biologic approaches to psychiatry in the United States. Among

Cobb's successors were Erich Lindemann in 1955, John Nemiah in 1965, and Leon Eisenberg in 1967.

THE INSTITUTE OF LIVING (CONTINUED FROM ABOVE). In 1931, when C. Charles Burlingame was appointed superintendent, what was then the Neuropsychiatric Institute of the Hartford Retreat was in a downhill slide. Burlingame effected a revival, establishing a school of nursing, bringing in recognized figures on staff (to regain the lost accreditation), and, later in the 1930s, implementing the new "shock" therapies. In 1943, the name was changed to the Institute of Living. In 1946, Francis J. Braceland, who had studied at the Zurich University Psychiatric Hospital and the National Hospital for Nervous Diseases in London, became chief psychiatrist. (Braceland is also remembered for being a special witness at the Nuremberg Trials in 1946, whose testimony saved Rudolph Hess from execution.) Said a New York Times obituary in 1985, Braceland is "credited with changing its reputation [Institute of Living] from that of a plush sanitarium for the rich and famous to that of a highly respected psychiatric hospital for the seriously ill." In 1994, the Institute of Living became part of Hartford Hospital.

YALE UNIVERSITY. Yale had the third university department of psychiatry in the United States—after Johns Hopkins and Harvard. The central figure here was Dean Milton C. Winternitz, a neuropathologist who conceived the Institute for Human Relations, which opened in 1931. (The department of psychiatry was situated in the Institute.) Winternitz said in 1925 of the proposed Institute, "… All of these disciplines primarily interested in the problems of human behavior from the standpoint of investigation, graduate teaching, and application, i.e., therapy, [should] be brought together geographically in a single building designed to allow the greatest possible latitude for … cooperation." An embryo department of psychiatry had come to life in the 1920s. Yet only in 1930 was it led by a significant figure, when Eugen Kahn, a Kraepelin student, and collaborator with Ernst Rüdin in schizophrenia genetics, became chair and the first Stirling

Professor in psychiatry. He was followed in 1950 by another eminent Central European, Vienna's Friedrich ("Fritz") Redlich.

In 1911, Arnold Gesell created the Child Study Center at Yale, later associated with several departments, including psychiatry.

Elsewhere

CHICAGO. Chicago had a long tradition of contributions to psychiatry. In 1925, physiologist Nathaniel Kleitman organized a sleep laboratory at the University of Chicago, and from this lab emerged in 1953 a classic paper in Science co-written by Kleitman's student Eugene Aserinsky on rapid eye movement (REM) sleep. Neurologist and psychiatrist Percival Bailey had cofounded in 1946 the Society of Biological Psychiatry. And Laszlo von Meduna, the originator of convulsive therapy, arrived in Chicago from Budapest in 1939 and ultimately found a berth at the NPI of the University of Illinois.

Yet the founding figure in psychiatry is Roy Richard Grinker, who had a background both in psychoanalysis (worked with Freud in Vienna in the early 1930s) and in neurology. Grinker's career at the University of Chicago began in 1933, with his appointment as associate professor of neurology; he became associate professor of psychiatry in 1936, which established psychiatry within the department of medicine (Grinker's appointment at Michael Reese Hospital in 1936 is recalled as "taking over the department of neurology"). In 1955, psychiatry changed to a department in its own right, its first chair C Knight Aldrich. Although psychodynamics was then the watchword in Chicago, the University was not hesitant to appoint in 1966 the elfin Daniel X. Freedman from Yale as the next chair. Freedman, an enthusiastic promoter of biologic psychiatry, became chief editor of the AMA's Archives of General Psychiatry in 1970. He was succeeded as chair by Stuart C. Yudofsky in 1988 and Elliot S. Gershon in 1998. Steven Paul, who later led neuroscience research at Eli Lilly, recalled in a 2011 interview his years in the mid-1970s as a resident in psychiatry at the University of Chicago, "It was a very exciting department in those days. Herb Meltzer [biologic psychiatrist] was there and Heinz Kohut, the analyst, Bob Schuster [Charles R. Schuster, who founded the University of Chicago's Drug Abuse Research Center], and many others. A tremendous department Danny had pulled together, a small but extraordinarily fine department." Meanwhile, in 1974 Philip Holzman, professor of psychology at the University of Chicago, reported a major biologic abnormality in schizophrenia, saccadic eye-tracking movements (smooth-pursuit in tracking is normal).

UNIVERSITY OF NORTH CAROLINA. The Department of Psychiatry at the University of North Carolina (UNC) at Chapel Hill was founded by Chicago psychoanalyst George Ham in 1951. It started to become a biologic piston with the arrival in 1954 of Scotsman John Ewing as clinical instructor, then chair in 1962. Psychiatric endocrinologist Morris Lipton from the University of Chicago joined the department in 1959 and proceeded to steer it toward the endocrinologic hormones that science was just discovering. (Lipton became known as having "opened up the peptides.") Arthur Prange, who had trained at UNC, joined the staff in 1960 and proceeded to solidify UNC's growing association with psychiatric endocrinology. William McKinney (later at the University of Wisconsin) trained at UNC in the mid-1960s, and in his class were Frederick Goodwin (later an NIMH star), the Englishman Peter Whybrow (later chair of psychiatry at the UCLA NPI), and Joseph Mendels (who ended up leading biologic psychiatry at the University of Pennsylvania).

DUKE UNIVERSITY. Across the way from UNC lay Duke University in Durham. In 1953, Ewald Busse, psychoanalyst and geriatrician, became the founding chair of psychiatry at Duke. His successors H. Keith Brodie (1974 to 1982) and Bernard Carroll (1983 to 1990), ended this Freudian spin (in 1968, Australian-born Carroll introduced the dexamethasone suppression test [DST] into psychiatry for the diagnosis of melancholic depression). Duke and UNC together constituted a powerful core of psychiatric research for the southeast.

UNIVERSITY OF PENNSYLVANIA. Psychiatry at the University of Pennsylvania had begun in 1793, when Benjamin Rush, a professor of chemistry, opened a ward for insane patients at Pennsylvania Hospital. But academic psychiatry at Penn really began only in 1874, with the establishment of the Hospital of the University of Pennsylvania. A department of mental diseases received its inaugural head in 1883, and in 1901, Charles Burr was appointed Penn's first chair of mental diseases.

Karl Rickels founded the Clinical Psychopharmacology Unit in 1959. And the Center for Mental Health Policy and Services Research (CMHPSR) was established in 1986.

THE "PSYCHOPATHIC HOSPITALS". Emblematic of the new association between asylum and academy were the so-called "psychopathic hospitals," which began life in the 1880s as reception wards for newly admitted, or "acute," patients. American psychiatrists visiting Europe had learned much about humane care and the usefulness of setting these acutely ill patients aside from the chronic patients working in the farm colonies or vegetating in the back wards. (Indeed, the spotlight began to turn toward creating psychiatric wards in general hospitals, such as "Pavilion F," founded in 1902, of Albany Hospital.)

In 1901, the Michigan legislature appropriated a sum for the establishment of a "psychopathic hospital" on the grounds of the State University of Michigan, as an additional ward of the University Hospital. It was not part of the state asylum system and was entirely under the control of the university. The psychopathic hospital did not, in fact, open until 1906, under the direction of neuropathologist Albert M. Barrett.

In Massachusetts in 1912, a psychopathic hospital opened on Fenwood Road in Boston as a department of the Boston State Hospital in Dorchester; it was under Elmer E. Southard, a professor of neuropathology at Harvard. The plan was, as Henry Hurd in his great 1916 work *The Institutional Care of the Insane* put it, to establish "a hospital for the first care and observation of mental patients and the treatment of acute and curable mental diseases and for an out-patient department … for scientific research as to the nature, causes and results of insanity.” The patients were given laboratory tests, stabilized with hydrotherapy, and then either discharged as recovered or forwarded to the main wards of the state hospital. In 1920, the institution was renamed Boston Psychopathic Hospital under C. Macfie Campbell. In 1956, the name was changed again to Massachusetts Mental Health Center. Several prominent figures in U.S. psychiatry served there, including Abraham Myerson.

Fourteen further psychopathic hospitals were founded by the First World War, although their missions ranged from the conduct of research, as at the University of Michigan and Harvard’s Boston Psychopathic, to the mere reception and detention of acute patients. (On the Iowa Psychopathic Hospital, opened in 1921, see below.)

Mississippi Valley Psychiatry. A string of important psychiatry departments ran down the Mississippi valley and played a crucial role in bringing the new biologic psychiatry to the United States.

UNIVERSITY OF IOWA. Academic psychiatry began in Iowa City with the opening of the Iowa State Psychopathic Hospital in 1921 under the state board of education. The director was simultaneously chair of psychiatry at the university, and the Iowa Psychopathic is considered the first university-affiliated psychiatric department west of the Mississippi. Under the founding director, Samuel T. Orton, in 1919, a neuropathologist who previously had been clinical director of the Worcester State Hospital in Massachusetts, the department rose to become one of the powerhouses of U.S. psychiatry. Jacques Gottlieb, whose background was in neuroendocrinology and who established the effectiveness of the amphetamines in the treatment of depression, taught at the Psychopathic Hospital from 1937 to 1940, then returned to Iowa as professor of psychiatry from 1947 to 1953 before moving on. Erich Lindemann came from Heidelberg to train at Iowa in 1927 and was on staff until 1935, before accepting a post at Harvard. While at Iowa, Lindemann was among the first to investigate the effects of psychopharmaceuticals. (Lindemann is also remembered for helping pioneer the diagnosis of traumatic stress with his 1944 study of “acute grief” in the survivors of the Cocoanut Grove Fire in 1942 in Boston.) Establishing a strong bridge between Iowa and Washington University in St. Louis, George Winokur came from Wash U in 1971 to chair psychiatry at Iowa. William Coryell followed in 1977. Ming Tsuang trained in psychiatry at Iowa then served as professor from 1975 to 1982. In 1996, Winokur and Tsuang (who by now had become the Stanley Cobb Professor of Psychiatry at Harvard) published *The Natural History of Mania, Depression, and Schizophrenia*, based on the research panel “the Iowa 500.” In 1971, Nancy Andreasen completed her training at Iowa, then, while on staff, went on to pioneer the neuroimaging of schizophrenia.

WASHINGTON UNIVERSITY IN ST. LOUIS. Psychiatry at “Wash U” began in 1918, under Sidney Schwab, who headed the division of neurology and psychiatry, a “subdepartment” of the Department of Medicine. In 1938, the Department of Neuropsychiatry became an independent unit under David M. Rioch, who passed the torch to Edwin Gildea in 1942. In 1948, Harvard-trained Eli Robins became a fellow at Wash U, joining the department in 1951. Meanwhile, in 1950 Samuel Guze joined the department, and for the next 20 years, the names Robins and Guze would be closely associated with the growth of psychiatry at Wash U as it became the premier department of biologic psychiatry in the United States. In 1970, they suggested the “medical model” as an alternative to the “biopsychosocial model,” as a motor for delineating disease.

Wash U was also one of the two seedbeds (Columbia the other) of the American Psychiatric Association’s *Diagnostic and Statistical Manual of Mental Disorders* (DSM). Recalled Paula Clayton in an interview, who at the time was a resident, “We all read Kraepelin. He was our bible.” Under Eli Robins (chair 1963 to 1975) and Samuel Guze (“Goo-zay,” chair 1975 to 1989), psychiatry at Wash U opened up a series of links between basic biology and clinical psychiatry. “[It] was nearer to being a mini-NIMH,” said geneticist Irving Gottesman later, “where under one roof you could have MDs and PhDs of all sorts working together on projects.” Graduates of Wash U assumed important leadership roles throughout U.S. psychiatry. It was the consummate insiders' group and displaced the previous insiders' group, the psychoanalytically oriented Group for the Advancement of Psychiatry (GAP).

Tennessee. Frank H. Luton became the first psychiatrist at Vanderbilt in Nashville, Tennessee, when he returned in 1931, after training at Johns Hopkins under Adolf Meyer. The School of Medicine created a department of psychiatry in 1947 with the appointment of William F. Orr, Jr, who in turn was succeeded by Marc Hollender in 1969, and Michael H. Ebert in 1983. Interest at Vanderbilt passed from dynamic psychiatry to psychopharmacology in the 1970s, with the appointment of such investigators as Fridolin Sulser, Thomas Ban, and Herbert Meltzer.

Meanwhile, in Memphis, Hagop Akiskal, a graduate in 1969 of the medical school of the American University of Beirut, had organized in 1973 an affective disease clinic at the University of Tennessee, one of the first in the United States. From 1987, Akiskal propagated the idea of a "soft bipolar spectrum," heavily dependent on temperament and managed with mood stabilizers rather than conventional "antidepressants."

Tulane. Psychiatry at Tulane in New Orleans was founded by Robert G. Heath after he arrived from Columbia in 1949. In New York, Heath had trained in both neurology and psychoanalysis and established a psychoanalysis program at Tulane. But his real interest lay in the possible metabolic origins of schizophrenia. The charismatic Heath, whom Time magazine called "the Gregory Peck of psychiatry," set out to investigate intracerebral events by persuading doubtful colleagues of the wisdom of inserting indwelling electrodes. This controversial program came to nothing, but Health had several important students. In 1960, one of them, Donald M. Gallant, won for the department of psychiatry a federal grant from the ECDEU of the National Institute of Mental Health, a program that aided the introduction of psychopharmacology into American psychiatry (see above). Gallant became chair in 1969 following

Heath and expanded the entire range of research in psychopharmacology so that it became the flagship program of the Deep South in this area.

The West Coast. Several departments of international distinction on the West Coast were founded—much later than those in other parts of the country. Yet, they are part of the story of the transition from asylum to university psychiatric hospital.

STANFORD UNIVERSITY, PALO ALTO. Psychiatry at Stanford was founded in 1933 when George S. Johnson came from his post at the University of Colorado to become chair of neuropsychiatry. In 1954, neurology split from psychiatry, and Johnson continued as chair of psychiatry. Child and adolescent psychiatry, which was established in 1939, merged with the department in 1957. In 1960, Stanford University School of Medicine moved from San Francisco to Palo Alto, where it quickly became a research-oriented medical school. David Hamburg came from the NIMH in 1961 to chair the department, and he recruited a number of future leaders, such as William Dement, who founded sleep medicine as a clinical discipline as well as a research program; Jack Barchas, remembered for early work on the enkephalins and endorphins; and Helena Kraemer, who developed statistical research methods.

The department went on to become one of the main forces in U.S. psychiatry. Alan Schatzberg, a leader in psychopharmacology and coauthor with Charles Nemeroff of Textbook of Psychopharmacology (1998), came from McLean Hospital to serve as chair from 1991 to 2010. His own work centered on the adverse effects of hypercortisolemia and the development of glucocorticoid antagonists as treatments for psychosis. Laura Roberts, a leader in academic psychiatry and bioethics, was recruited from the Medical College of Wisconsin to serve as chair beginning in 2010.

LANGLEY PORTER CLINIC SAN FRANCISCO. In San Francisco, the Langley Porter Clinic, named after Dean Robert Langley Porter, who conceived it in the late 1930s, opened its doors in 1942 as a 100-bed psychiatric hospital. It later changed its name to the “Langley Porter Neuropsychiatric Institute” and then again to “Langley Porter Psychiatric Institute.” Writes historian Jurgen Ruesch in 1978, “It was based upon the idea that prevention, teaching, and research … would be enhanced if the wealth of clinical experience gained from the approximately 29,000 patients then available in the nine California state hospitals could be combined with the expertise of the staff of a medical school.” The inaugural director was Carl M. Bowman, who simultaneously became the first chair of the new division of psychiatry in the School of Medicine. Bowman was among the charter members in 1946, of the Society of Biological Psychiatry. In 1973, the hospital was transferred from the jurisdiction of the state to the University of California (as was the NPI in Los Angeles [see below]).

NEUROPSYCHIATRIC INSTITUTE LOS ANGELES. The UCLA was opened only in 1947, and Stafford Warren, the inaugural Dean, was said to have “dreamed of creating a ‘Langley Porter of the South.’” Norman Brill was brought from Georgetown in 1953 to become the founding chair of the Psychiatry Department, and in 1961, the first patients were admitted to the new NPI funded by the state government of California. A nearby Brain Research Institute also opened in 1961. Louis J. (“Jolly”) West came from the University of Oklahoma to chair the young psychiatry department, and West built staff from all around the world. In 1959, John French became the founding director of the Brain Research Institute. In 1973, the California State Department of Mental Hygiene transferred the NPI to the University of California, where it became part of the UCLA Health Sciences Center.

DIAGNOSIS

The two great arms of clinical psychiatry are diagnosis and treatment. The following sections are devoted to the histories of both.

Modern psychiatric diagnosis really begins with Philippe Pinel in Paris (see above), who in the second edition of his textbook in 1809 enumerated four psychiatric diseases: (1) “mania, or general psychosis [délire]”; (2) “melancholia, or a specific delusion”; (3) “dementia, or abolition of thought”; (4) “idiotism, or the obliteration of the intellectual and affective faculties.” What he meant by these would parse into a number of distinct modern illnesses, but the point is that he did recognize several specific forms of previously undifferentiated “madness.”

After Pinel and Esquirol (see above), the most important figure in the history of nosology is Emil Kraepelin (“Krep-uh-leen”), professor of psychiatry in Heidelberg from 1891, in Munich from 1903, and author in 1883 of the influential psychiatric textbook, Psychiatry: a Textbook for Students and Physicians (*Psychiatrie: ein Lehrbuch für Studierende und Aerzte*), which reached international influence with its fourth edition in 1893 and became worldwide the dominant voice in psychiatry by the eighth edition, completed in 1915. Kraepelin had little patience with psychiatric theory or with pathologic anatomy and placed the diagnostic interest in psychiatry squarely on the clinical side, meaning at the bedside, especially on the course and outcome of the patient’s illness.

Kraepelin was the clinician’s clinician. On the basis of clinical course, Kraepelin said in 1899 that the two specific diseases in psychiatry were “manic-depressive insanity,” a mood disorder, and “dementia praecox,” or the dementia of adolescence, a thought disorder later called “schizophrenia.” He erected a firewall between them, one that still exists today, as in the DSM system, the cardinal disorders are “major depression,” as opposed to “schizophrenia.” No other historical figure in psychiatry, including Freud, has had the influence of Kraepelin upon contemporary psychiatry, although under the influence of psychoanalysis, U.S. interest in diagnosis did wane for several decades. Henri Ellenberger, a Swiss psychiatrist transplanted to Montreal, recalled in 1955, “I remember the consternation of a German psychiatrist hearing that certain American colleagues consider manic-depressive psychosis as a form of schizophrenia; for him, this was about as fantastic as the assumption that a camel is a subspecies of the elephant.”

A page was turned with the advent of a distinctively American system of classification in 1945, the *Nomenclature of Psychiatric Disorders and Reactions*, which appeared as *War Department Technical Bulletin, Medical 203.* Spearheaded by psychoanalyst William Menninger, the head of consultation psychiatry in the U.S. Army during the war, "Medical 203" made wide use of psychoanalytic insights and underpinned the first effort of the American Psychiatric Association to classify illness, the *Diagnostic and Statistical Manual,* or "DSM-I" (the so-called subsequently) in 1952. (The APA itself was founded in 1844 as The Association of Medical Superintendents of American Institutions for the Insane, becoming the American Medico-Psychological Association in 1894 and the American Psychiatric Association in 1921.)

An even more psychoanalytically oriented volume, DSM-II, emerged in 1968.

In the 1970s, the APA drew up a new task force to make what were intended to be minor changes in DSM to bring it into accordance with the *International Classification of Diseases* (ICD) of the World Health Organization, appointing Robert Spitzer at the New York State Psychiatric Institute to lead it. Spitzer, however, disliked psychoanalysis and wanted a complete revision of the nosology to remove Freudian influences. In 1980 the revolutionary DSM-III emerged, which created such diagnoses as "major depression" and "attention deficit disorder" and anchored "bipolar disorder" firmly in the American classification. The subsequent editions of DSM—DSM-III-R in 1987, DSM-IV in 1994, and DSM-5 in 2013—changed little of this architecture, though in launching "bipolar two disorder" in 1994, DSM-IV did touch off an epidemic of "bipolar" in the pediatric population.

A rival system was the ICD of the World Health Organization. In 1938, an "international revision" conference of the League of Nations decided to add morbidity to its long-standing classification of mortality, and in 1948 the newly founded World Health Organization issued the "sixth revision" of the international lists (ICD-6). It contained such diagnoses as "delusional" insanity (other major psychiatric illnesses were cataloged under “dementia”). A distinct section on “mental disorders” was added in the ninth edition in 1975.

Over the years, the number of diagnoses in psychiatry has skyrocketed from four in Pinel to some three hundred in DSM-5. Some of these diagnoses may not stand the test of time. Yet, on a historical basis, several diagnoses do stand out as enduring and doubtless represent phenotypes of genuine underlying psychiatric diseases.

Specific Diagnoses

Melancholia. Melancholia, as distinct from generic “depression,” is one of the oldest diagnoses in medicine and goes back to the Ancients. Before the 19th century, melancholia tended to be an umbrella term for several varieties of psychosis and mood disorders that today would be recognized as distinct. Yet, at its core, melancholia has always stood for severe depressive mood disorder involving deep sadness, pervasive anhedonia, and psychomotor change, usually in the direction of slowing, but there is an anxious form of melancholia involving psychomotor acceleration as well. Timothie Bright, a Cambridge physician, in his 1586 book *A Treatise of Melancholie* described,

> the sorrowful humor which breeds in them … Of pace, they are for the most part slow … both by their members not so nimble for motion, and the mind occupied with cogitation stays [slows] the pace … Moreover, they are given to weeping sometimes … sighing, sobbing, lamentation, countenance demisse [downcast] and lowering.” Bright left a poignant image of weeping in melancholia, “It blears the eyes and makes the cheeks red; it causes the head to ache, the nose to run, and the mouth to flauer [labor], the lips to tremble; it interrupts the speech and shakes the whole chest with sighs and sobs.

In the 19th century, melancholia became differentiated from other kinds of depressive illness. Among the first to make this distinction was Richard von Krafft-Ebing in 1867, at the time a staff psychiatrist at the Illenau asylum (later, professor of psychiatry in Vienna). Krafft-Ebing distinguished between “simple depression of mood” (*die einfache Gemüthsdepression*) and severe depressions with anxiety and psychosis, including melancholia. The definitive distinction between “melancholia” and “depression” was ventured by Copenhagen psychiatrist Carl Georg Lange in his book on “periodic depressions” in 1886 (translated into German in 1892), which described these milder cases as distinct from (psychotic) melancholia.

The term “melancholia” went out of style around 1900 as Emil Kraepelin substituted “depression” for it, and Sigmund Freud decided that the depressive disorders did not qualify as the main psychoneuroses that psychoanalysis should be addressing. (But see above on a United States interested in “depressive neurosis.”) In recent decades, there has been a revival of interest in melancholia, as the field has begun to differentiate the heterogeneous basin of “major depression” on the basis of treatment response.

Psychotic Depression. In classical psychiatry, much melancholia was psychotic, and little was made of the distinction between psychotic and nonpsychotic. The difference later seemed unimportant because both groups responded equally well to ECT, introduced in 1938. Yet with the growing stigmatization of ECT after the 1960s, depression became treatable primarily with drugs, and in 1975, Alexander (Sandy) Glassman at the New York State Psychiatric Institute discovered that psychotic depression responded poorly to the standard tricyclic antidepressants (TCAs). Four years later, David Avery and Aldo Lubrano at the Palo Alto Veterans Hospital reported that psychotically depressed patients responded better to ECT than did nonpsychotically depressed. Under the leadership of Max Fink at the Stony Brook Campus of the State University of New York, ECT became defined as the most effective treatment of psychotic depression. Despite this evidence of differential responsiveness, psychotic depression continued to be downgraded as a mere “specifier” in the DSM.

Mania. Mania has undergone several transformations. For Pinel and the generations before him, it meant agitated dementia (“raving madness”). With François Fodéré in Strasbourg (1817), unbridled unreasoning outbursts became “furious mania.” In 1858, London psychiatrist Daniel Hack Tuke steered mania in the direction of an affective disorder rather than a disorder of intellect. “Although mania, in many instances, is prolonged anger, it may, likewise, be altogether pleasurable in its manifestations, presenting a condition of exhilaration and uncontrollable excitement, which is rather an excess of joy than of anger.” Then in 1881, Emanuel Ernst Mendel, senior lecturer in Berlin, anchored the term “hypomania” in the nosology, meaning elated mood, pressured speech, and increased motor activity. Mania entered DSM intact in 1952, and hypomania became official in 1994 in DSM-IV as “Bipolar II Disorder.”

Catatonia. Although individual symptoms of catatonia, such as waxy flexibility, had been described for centuries, catatonia as a clinical syndrome was recognized by Karl Kahlbaum only in 1874. Its independent existence was short, as in 1899, in the sixth edition of his Textbook, Kraepelin made it a “subtype” of schizophrenia, and for the next century, catatonia languished as an independent entity because, often when clinicians encountered it, they realized they were not dealing with schizophrenia and called it something else, or ignored it. In 1930, William Bleckwenn at the University of Wisconsin reported that catatonia responded to barbiturates, though this discovery was not widely appreciated. Karl Heinz Stauder, at the Munich University Psychiatric Hospital, described in 1934 a fatal version of catatonia, associated with hyperpyrexia.

The return of catatonia as an independent diagnosis separate from schizophrenia began in the 1970s. In 1976, Richard Abrams and Michael Alan Taylor at New York Medical College identified 20% of a sample of patients admitted to the psychiatric unit of a New York City hospital as having one more catatonia signs. Only 4 of the 55 satisfied the criteria for schizophrenia, while more than two-thirds had affective disorders, usually mania. In the context of growing interest in catatonia as independent of schizophrenia, in 1983, Gregory Fricchione at Stony Brook recommended benzodiazepines, especially lorazepam, in the treatment of catatonia instead of barbiturates. Various syndromes previously seen as separate began to be recognized as catatonic in nature: In 1996, it was suggested that toxic serotonin syndrome (TSS) might be a form of catatonia. A symposium at the American Psychiatric Association meeting in 2000 concluded that neuroleptic malignant syndrome (NMS) and malignant catatonia were the same. In 2011, Dirk Dhossche at the University of Mississippi and colleagues raised the possibility that anti-NMDA receptor encephalitis was a form of catatonia. The American Psychiatric Association responded to this series of new catatonia findings by making catatonia a virtually independent disease entity in the DSM-5 published in 2013.

Delusional Disorders

Delusional disorders, meaning fixed, systematic false beliefs, were recognized in 1816 with Etienne Esquirol's "monomania." He said that the psychosis (*délire*) is fixed on one object alone. As early as 1865, then definitively in 1874, Ludwig DC Snell, director of the asylum in Hildesheim, Germany, distinguished between primary and secondary delusions (*Wahn*), primary meaning fixed, false ideas not caused by another illness and not deteriorating; vague secondary delusions could lead to "the collapse of the entire psychic life," something Snell observed in the kind of illness later called "schizophrenia." This was the beginning of separating delusional disorders from chronic deteriorating psychosis as the former were false beliefs that were fixed, systematic, and did not affect other areas of intellectual functioning.

Anxiety. The term "anxiety" has long been used in psychiatry. In 1769, a young medical graduate of Leiden University ventured a dissertation on "Anxietate," and the term was clearly current among his professors. Yet, the German clinical tradition of the 19th century frowned on anxiety as an independent diagnosis. In Emil Kraepelin's big textbook, there is no freestanding "anxiety" disorder. But it was not just Kraepelinian obdurance. The "anxiety" entry in Daniel Hack Tuke's *Dictionary of Psychological Medicine* (1892) is only two sentences long. What placed anxiety in the center of psychiatry's radar was Freud's belief that anxiety was the basic motor of symptom formation: anxiety arose as a result of intrapsychic conflict over various drives. The entries for "anxiety" (*Angst*) in the general index of Freud's writings extended for eleven pages ("genitals," by contrast, only for four). Accordingly, anxiety has always assumed a disproportionate role in psychoanalytically oriented American psychiatry (mixed anxiety-depression has tended to loom more prominently for the British). DSM-I in 1952 prefaced its discussion of "psychoneurotic disorders" by claiming that, "The chief characteristic of these disorders is 'anxiety'…." and in the discussion that followed "anxiety" figured importantly in each, such as "dissociative reaction," "obsessive-compulsive reaction," and so forth. DSM-III in 1980, although antipsychoanalytic in inspiration, continued the centrality of anxiety while beginning its fragmentation into specific anxiety disorders, of which there were ten. Subsequent editions of DSM have continued this tradition.

Chronic Primary Psychosis (Schizophrenia). Psychiatry has struggled over the years with chronic deteriorating psychotic illness, leading either to some kind of deficit or to the dissolution of the personality (once called "dementia"). Developments in psychiatry's two major cradles are chronicled separately.

FRANCE. In 1852, Bénédict-Augustin Morel, then at the Maréville asylum in eastern France (but shortly to head the St-Yon asylum outside of Rouen) called such psychoses of adolescent-onset *démence précoce*, or premature dementia (see above). In 1857, he applied the notion of "degeneration," or worsening across the generations, to psychiatric illness (*Traité des Dégéréréscences*) and 1860, in one of the most famous declarations in the history of psychiatry, in his *Treatise on Mental Illness* (*Traité des Maladies Mentales*) he avowed premature dementia in the young to be a consequence of degeneration.

In 1883, Valentin Magnan, a staff psychiatrist at the Sainte-Anne Clinical Asylum in Paris, described "chronic psychosis"—he later added the qualifier "involving systematic evolution"—an often-deteriorating form of psychosis in patients who were not hereditarily predisposed. This remained quite influential in France, though elsewhere, it was eclipsed by “schizophrenia” (see below).

French experimentation with chronic psychosis continued, even as Magnan’s degeneration-based doctrines were discredited, and in 1910, Maurice Dide, at a psychiatric hospital in Toulouse, proposed “chronic hallucinatory psychosis,” which, though severe, did not deteriorate. To this day, the concept has remained influential in France without reaching a larger international audience.

GERMANY. Meanwhile, in Germany, a different approach to chronic psychosis was being elaborated: deterioration to “dementia,” or at least to a severe deficit. In an 1863 book on the classification of psychiatric illness, Karl Kahlbaum, a staff psychiatrist at the East Prussian provincial asylum of Allenberg, differentiated a “deteriorating psychosis,” which he called *Vesania typica*, which ran its course in four stages, the last of which was “dementia.” Vesania was a customary term at the time for chronic psychosis. In 1871, Kahlbaum and his student Ewald Hecker differentiated further a version of this Vesania beginning in adolescence that they called *hebephrenia*. Hebephrenia, like Vesania typica, usually ended in dementia. (Although Hecker’s name appeared alone on the paper, it was understood that the concept was really Kahlbaum’s.)

At this point, a considerable amount of European scholarship, dating from Morel in 1852, had accumulated that linked chronic deteriorating psychosis to onset in adolescence. In 1893, Kraepelin synthesized this literature in a concept that was to sit rock-solid in psychiatry for more than a century: *dementia praecox*. Over the next decade, he identified three subforms of it: hebephrenic, catatonic, and paranoid. They had in common a begin usually in adolescence or young adulthood, and a downhill course to “dementia.” So magisterial was the Kraepelinian synthesis that all previous writing on chronic psychosis became virtually forgotten. In 1908, Eugen Bleuler, professor of psychiatry in Zurich, baptized Kraepelin’s dementia praecox as “schizophrenia,” seeing it as a much milder disease and not linked to adolescence. Bleuler’s schizophrenia remains little changed today, a single deteriorating psychotic disorder that may onset at any point in life and that may well end in “defect” (or social remission), even if “dementia” as the end stage of the illness has stopped being spoken of. In 2013, DSM-5 dispensed with the famous Kraepelinian “subtypes.”

Many “schizophrenia” patients may spend long periods without being actively psychotic; the primary symptoms seem to be emotional blunting with avolition and a formal thought disorder.

Brief Reactive Psychosis or Brief Psychotic Disorder

Chronic psychotic illness had long been seen as something indwelling and constitutional. The alternative notion that psychosis could arise from stress was codified as *la bouffée délirante* by two students of Valentin Magnan at Sainte Anne in Paris in 1886, then by Magnan himself in 1891—who saw it as arising mainly in the “degenerate.” August Wimmer at St. Hans Hospital in Denmark christened it “psychogenic psychosis” in 1916, with an emphasis on reaction rather than degeneration. Then in 1968, Eric Strömgren in Aarhus, Denmark, revived Wimmer’s concept in an influential (in Scandinavia) paper on “reactive psychoses,” a term that he changed in 1974 to “psychogenic psychoses.” In 1980, DSM-III recognized “brief reactive psychosis,” with no reference to any of these other international terms. (These developments are charted in Shorter, *Historical Dictionary*.) The French translation in 1989 of DSM-III-R makes no reference to bouffée délirante and mentions only *psychose reactionnelle brève*, an interesting example of the collapse of these once distinctive national nosologic traditions.

Bipolar Disorder

The “bipolar disorder” of DSM is not Kraepelin’s “manic-depressive insanity,” which Kraepelin insisted included all depressions of whatever polarity, mania, hypomania, and the attendant anxiety. Rather, in his 1948 textbook Karl Leonhard, then in Frankfurt am Main and a student of Karl Kleist (who introduced the concept of *phasic psychoses*, or illnesses that alternate from one state to another), proposed the term “bipolarity” for the alternation of depression and mania as an illness separate from unipolar depression. In 1957, *in The Classification of the Endogenous Psychoses* (*Die Einteilung der endogenen Psychosen*), Leonhard asserted that unipolar depression had a different psychopathology from the bipolar variety. Filled with neologisms and lacking in quantitation, the book had a limited international impact. Yet a small cadre of investigators, such as Frank Fish in England, Helmut Beckmann in Germany, and George Winokur in the United States, became enthusiasts of the Leonhardian notion of bipolar disorder and managed to insert it in the literature, so that by 1980 when DSM-III was published, bipolar disorder was self-understood as a different disease from unipolar depression (in DSM-III called major depression).

Trauma and Its Sequelae

Ever since Napoleon’s disastrous retreat from Moscow in 1812, psychiatry has recognized that symptoms may arise from great stress. (In Napoleon’s case, of an army of 400,000 that set out, 7,000 returned. Along the way, much “nostalgia” was witnessed—an early label for fatal catatonia.) In 1866, London surgeon John Eric Erichsen originated the diagnosis of “railway spine,” vague and nonspecific symptoms supposedly attributable to the “spine” in minor railway mishaps. Carol Moeli, a Berlin psychiatrist, elaborated in 1881 upon the specific psychiatric consequences of such events, penning the first description of what became known as post-traumatic stress disorder (PTSD). In 1881 as well, Berlin neurologist Hermann Oppenheim coined the phrase “traumatic neurosis,” psychiatric and neurologic symptoms arising from shock.

The First World War spawned a whole raft of descriptions of psychological sequelae from combat, of which terms the most lasting has been “shell shock”—from Charles S. Myers in 1915. The Second World War gave birth to “combat fatigue” as a synonym for war neurosis. Meanwhile, a tragic fire in 1942 at the Cocoanut Grove nightclub in Boston gave rise in 1944 to “acute grief” as yet another term for acute stress reaction. Most recently, following a sustained political campaign by the veterans of the Vietnam War—and spearheaded by Yale psychiatrist Rober Lifton—DSM-III in 1980 coined the term “posttraumatic stress disorder.” This diagnosis has been among the most rapidly proliferating in U.S. psychiatry.

Dementia

The dementias have always been a problem for psychiatry, as the field has struggled to differentiate loss of intellectual function from personality deterioration. In 1814, Etienne Esquirol took up the idea of his teacher Pinel of dementia as a distinct disease entity, separate from melancholy, mania, and imbecility. “There exists a very distinctive kind of insanity in which the disorder of ideas, sentiments, and determinations is characterized by the weakness and the more or less pronounced abolition of all the sensory, intellectual and moral faculties.” In particular, Esquirol differentiated senile dementia from other conditions of “weakmindedness.” (Shorter, *Historical Dictionary of Psychiatry*, chronicles developments in dementia.) There followed a long series of differentiations of the organic dementias: Sergei Korsakoff distinguished ethanol-related amnesia (Korsakoff psychosis) from dementia in 1887; Alois Alzheimer distinguished “presenile” dementia from the senile variety in 1906; in 1926, Hugo Spatz, head of neuropathology at the Munich University Psychiatric Clinic, suggested that a distinctive frontal- and temporal-lobe wasting should be called “Pick’s Disease” after Arnold Pick, professor of psychiatry at the German university in Prague, who first reported it in 1892; in 1912, Friedrich Heinrich Lewy described distinctive intracytoplasmic inclusions, later called “Lewy bodies,” which he detected in the brain stem in Parkinsonism; and in 1962, neuropathologist John S. Woodard at the Camarillo State Hospital in California, proposed Lewy-body dementia as a separate form of dementia; in 1948, London psychiatrist Robert Denis Newton codified long-established German thinking that presenile Alzheimer’s and the dementia of senescence represented a single pathologic process, and suggested calling it “Alzheimer’s dementia”; Raymond

Adams at Harvard University and Maurice Victor at Case Western University, emphasized in 1977, that Alzheimer disease was not just an accelerated extension of old age but a pathologic eruption into it.

Ever since New Jersey pathologist Harrison Stanford Martland's paper "Punch Drunk" in 1928, repeated cranial trauma has been considered a distinct cause of dementia. Martland said, "Fighters in whom the early symptoms are well recognized are said by the fans to be 'cuckoo,' 'goofy,' 'cutting paper dolls,' or 'slug nutty.'" Since then, physicians have been alerted to the neuropsychiatric consequences of repeated head trauma. In 1937, in the U.S. Naval Medical Bulletin, naval surgeon J.A. Millspaugh called it *Dementia Pugilistica*. In a paper in the British Medical Journal in 1989, on "Boxing and the brain," John Arthur Nicholas Corsellis pulled this literature together in a red warning light about repeated head trauma. By the new millennium, these concerns about early dementia had been expanded to all sports involving repeated head trauma.

Neurosis

The classic distinction between psychosis and neurosis was abolished in DSM-III in 1980. Neurosis, as a term, is among the oldest in psychiatry; it was introduced by Edinburgh physician and nosologist William Cullen in his *First Lines in the Practice of Physic* in 1777 to mean diseases of the nervous system without an obvious physical lesion. In the next half-century, it came to be synonymous with brain disorder. It was Richard von Krafft-Ebing, in his *Textbook of Psychiatry* (*Lehrbuch der Psychiatrie*) in 1872, who coined the term "psychoneurosis," meaning mental illnesses that affect normal brains (in contrast to brains affected with "degeneration"). Sigmund Freud used the term psychoneurosis for the first time in his writing in 1898. But neurosis had by that time long entered medicine, and Freud proceeded to employ it as a synonym for "disorder." After Freud, neurosis remained largely in the area of psychoanalytic discourse, which was the dominant discourse in psychiatry until DSM-III in 1980.

NEW TECHNIQUES

Psychiatry is not among the most technologic of medical specialties. Yet changes in technique have expanded the range of its knowledge and enhanced the diagnostic and therapeutic abilities of its practitioners.

Epidemiology

The study of epidemiology had an early bump-up with the observation of Thomas Clouston in 1873, in the *Journal of Mental Science* that,

> The wide series of diseases which are at present called insanity prevail more in some places than in others … When one comes to look carefully into the reports of lunatic asylums in different parts of the country it is found that there are forms of brain disease (or varieties of insanity as they are called) present in abundance in one place which have almost no existence in another … This investigation must be very largely conducted on the numerical method. (Clouston TS. The local distribution of insanity and its varieties in England and Wales. *J Ment Sci.* 1873;19(85):1–19, reproduced with permission.)

Yet, it was the Germans who initially picked up this torch. At the German Psychiatric Research Institute (*Deutsche Forschungsanstalt*) in Munich, founded in 1917, by Emil Kraepelin, Swiss psychiatrist and geneticist Ernst Rüdin, inaugural director of the eugenics research institute of the DFA, led studies on psychiatric epidemiology (cross-sectional) and genetics (longitudinal). Unfortunately, in the Third Reich, Rüdin had a particularly unsavory involvement with public policy, arguing for the Nazi racial and sterilization laws, and after 1945 was briefly interned in the United States. This National Socialist involvement of the DFA and other German groups lent to psychiatric epidemiology and genetics a bad odor that retarded the development of the field for decades.

It was in Denmark that Eric Strömgren, the most important psychiatric epidemiologist of the pre-1970 period, flourished. As a student, he visited the Rüdin group in 1935; he undertook in 1938 what became an epidemiologic classic of the small island of Bornholm. He became a professor of psychiatry in 1945 and was instrumental in organizing the Danish Psychiatric Central Registry at Aarhus, which in 1969 became an electronic register of everyone in Denmark admitted to a psychiatric service. By 2011, this famous register contained over 700,000 individuals and more than 3 million treatment episodes. One offshoot of this was the Danish-American cohort study in 1968, led by David Rosenthal, Paul Wender, and Seymour Kety at NIMH. It merged two registers: that of children who had been adopted away and that of admissions to psychiatric hospitals. This remarkable database permitted the NIMH collaborators, together with Fini Schulsinger in Copenhagen and other Danish colleagues, to determine that children adopted from schizophrenia families but raised in normal homes often became schizophrenic. It was a major blow to the "nurture" school. At a conference in Dorado, Puerto Rico, where these results were first presented, Rosenthal said, "The case for heredity has held up convincingly. At the moment, I can see no grounds for withholding affirmation any longer."

In the United States, the National Institute of Mental Health sponsored, in the 1980s, the 5-site Epidemiological Catchment Area (ECA) study, the first to inquire about diseases rather than symptoms, the results of which were reported in 1991. Myrna Weisman at Columbia later said in an interview the big finding of the ECA was that "These disorders begin in the young … often in adolescence, and certainly, by young adulthood … So, the biggest contribution of the ECA was turning the focus on young people and their high rates of psychiatric disorders."

In the early 1990s, NIMH sponsored the National Comorbidity Survey (NCS) to assess the prevalence of DSM-III-R disorders in a nationally representative sample. Early in the new century, the NCS was redone in the National Comorbidity Survey Replication (NCS-R), using the international WHO Composite International Diagnostic Interview to collect DSM-IV diagnoses from a household survey and to make international comparisons possible.

Statistics

Paris psychiatrist Jean-Pierre Falret's critique of the counting method in 1864, in the preface of his On Mental Illness (*Des maladies mentales*), has not lost its relevance today: "Statistics as a procedure are in fact all the more dangerous because they offer all the appearance of truth." Falret scorned the statistics of the day as failing to include homogeneous clinical populations and using abstractions overly divorced from reality. Modern statistical theory, as Bernard Carroll traces events, originated in the last quarter of the 19th century and the first quarter of the 20th century. Regression to the mean was described in 1877; regression and correlation analysis in 1888 by Francis Galton; the Chi-squared test in 1900 was originated by Karl Pearson; the student's t-test in 1908 by W.S. Gossett; the analysis of variance (ANOVA) comes from RA Fisher in 1919; power analysis, confidence intervals, Type I and Type II error originated in the 1930s from Jerzy Neyman and Egon Pearson. But these advances were not incorporated as research tools into the clinical science of psychiatry until well beyond the 1930s.

The early use of statistics in psychiatry consisted essentially in tabular bookkeeping—simple descriptive statistics such as census details, cohort descriptions by age, age of onset, sex, or outcome. Use of statistics to test hypotheses in psychiatry came much later. Emil Kraepelin maintained detailed counting cards (Zählkarten) of cases he observed, leading up to his landmark distinction between dementia praecox and manic depressive insanity (see above). Assen Jablensky's 1993 reanalysis of those cases with modern statistical methods essentially confirmed the existence of two separate disorders that Kraepelin had recognized by informal Gestalt inferential methods. The same cannot be said of Aubrey Lewis, whose 1934 discussion of depressive illness confidently announced a single-dimensional model without the benefit of any statistical analyses. That position became dogma within Maudsley Hospital circles and led to conflict with the Newcastle school under Martin Roth. Eventually, Lewis's primary data were subjected in 1977 to modern statistical analysis by Lesley Kiloh and Roger Garside (two of the original Newcastle group), and they were found actually to support the Newcastle binary model. These two examples illustrate the absence of reliable means to validate the proposals of clinical experts before the entry of statistics into psychiatric research.

Aside from elementary analytic procedures such as group comparisons (in categorical and dimensional versions), significance testing, and correlations (all in parametric and nonparametric versions), the major statistical themes that moved psychiatric research were numerical taxonomy and psychometrics. Derived from advances in basic statistics, these approaches appeared first within academic psychology for the study of personality traits and dimensions of psychopathology. For the most part, they were slow to be taken up within the discipline of psychiatry and, when they were, the history of their adoption is inseparable from the history of clinical assessment instruments that relied on the new statistical approaches. In parallel, innovations in basic statistical theory and technique also were adopted within psychiatry for clinical trials, as well as for epidemiologic and genetic studies.

Factor Analysis and Personality Inventories

Factor analysis originated with Charles Spearman and Raymond Cattell. The psychological sciences adopted the technique in the 1930s. In parallel, the Minnesota Multiphasic Personality Inventory (MMPI) was under development by Starke R. Hathaway and J. Charnley McKinley at the University of Minnesota from the late 1930s. Local use apparently began in 1939. The original published version appeared in *JAMA* in 1943: The MMPI has seen many later revisions. A perceived advantage was casting it as a descriptive and atheoretical approach to personality and psychopathology, not derived from psychoanalytic theory. Over the years, cultural and ethnic biases in the MMPI have been alleged. Nevertheless, it attained widespread use in education, clinical psychology, and employment screening. A scoring innovation for the MMPI was the devising of T-scores—a standardized metric where the mean is set at 50 and the SD at 10. Other scales adopted the same scoring approach, which had the benefit of standardizing responses and allowing cross-scale comparisons. In parallel, Raymond B. Cattell devised in 1946 a personality inventory for which factor analysis was an essential component. His psychometric instrument is famous as the Cattell 16PF personality factors inventory.

Controlled Trials

The first controlled trial in psychiatry was reported in 1922 in the *American Journal of Psychiatry* by Clarence O. Cheney, assistant director of the New York State Psychiatric Institute, then at Ward's Island, and bacteriologist Nicholas Kopeloff. They wanted to establish whether the organ removals (colectomies, teeth, etc.) then in vogue for the treatment of mental illness were, in fact, effective, and to do so, they studied 60 patients with "focal infections" among the general admissions to the Manhattan State Hospital on Ward's Island, dividing them into two groups, the "operated group" receiving operative treatment for their " 'foci' of infection" versus those not so treated (the "control group"). They found, "On the whole, the operated group appears to have received no more benefit than the control group."

Following the tuberculosis studies of Bradford Hill after the Second World War, randomizing the patients to control or treatment groups became increasingly common. The first randomized controlled trial in psychiatry with "blinded" investigators was conducted in 1955 by David Lewis Davies and Michael Shepherd at the Maudsley Hospital, comparing reserpine to placebo in the treatment of "anxious and depressed patients."

The use of probabilistic statistics in psychiatry has a much more recent origin. Four statistics are used to measure the reduction of risk in treatment (two of them compare vs. lack of treatment). (The following discussion cribs from Andreas Laupacis and collaborators, 1988.)

- **The relative risk reduction (i.e., how much risk reduction in percent).** This is the classic measure.
- **The "odds ratio."** A probability (P) of 0.25 (one chance in four) represents odds of 1:3 or 1/3. This is insensitive to a difference in

risk in the absence of therapy. It was introduced by N. Mantel and W. Haenszel in 1959.

- **Absolute risk reduction.** The differences in event rates between the control and treatment groups. A genuine measure of absolute risk reduction but expressed as a decimal, “This measure may be difficult for [physicians] to remember and incorporate into clinical practice,” say Laupacis and colleagues.
- **Number needed to be treated (NTT).** The number of patients who must be treated in order to prevent one adverse event. Laupacis and colleagues proposed this in 1988. “It tells clinicians and patients in more concrete terms how much effort they must expend to prevent one event, thus allowing comparisons …” Although many clinicians will be put off by such exercises, these statistics mark the definitive passage from psychoanalysis with its evidentiary base of anecdotes to modern clinical psychiatry that employs the same evidentiary base as the rest of medicine.

Tests in Psychiatry

Despite protestations that psychiatry “has no diagnostic tests,” there are, in fact, quite a few. In 1906, August Wassermann and colleagues at the University of Berlin described a diagnostic test for “progressive paralysis” (what neurosyphilis was then called) based on the patient’s blood or cerebrospinal fluid. In 1913, Hideo Noguchi, a microbiologist at the Rockefeller Institute for Medical Research, succeeded in obtaining a culture of *Treponema pallidum* in a sample of the brains of neurosyphilitic patients who carried the diagnosis of “general paralysis of the insane,” the English equivalent of “progressive paralysis.” Therewith, neurosyphilis became identified as an infectious disease and passed from the purview of psychiatry, where it had lodged for the previous century. In 1931, Hans Berger in Jena used the EEG, which he had just invented, to confirm a specific pattern of brain waves in a patient with epilepsy. These were early diagnostic tests in psychiatry.

More were to come. A group of researchers led by Charles Bradley at the Emma Pendleton Bradley Home in East Providence, Rhode Island, established in 1938 that a certain subset of children who would later be called “hyperactive” (with the diagnosis of attention-deficit/hyperactivity disorder, ADHD) demonstrated a specific electroencephalographic pattern. The test remains controversial, but it has not been disallowed. In 1940, Mandel Cohen and Stanley Cobb, psychiatrists at Harvard, reported in The Journal of Clinical Investigation that carbon dioxide could precipitate what were later called “panic attacks” in patients prone to anxiety. Abnormalities in serum lactate thus became a biologic marker for panic, and there followed a considerable history of using “panicogens” to study panic.

As early as 1958, “pharmaco-EEG” was used to “classify psychopharmacologic drugs and predict their clinical applications,” as Max Fink, a pioneer of the technique, later put it. Then in 1962, in a study published in the American Journal of Psychiatry, Fink and Donald Klein, part of the Hillside Hospital Experimental Team, established in a proper randomized controlled trial of chlorpromazine and imipramine that depressed patients responded well to imipramine, schizophrenics did not. The imipramine challenge was the first biologic test for serious depression.

Bernard Carroll at the Royal Melbourne Hospital applied the DST in 1968 to confirm the presence of melancholic depression. The DST, based on a dexamethasone challenge to the hypothalamic–pituitary–adrenal (HPA) axis, establishes HPA dysregulation in patients with severe mood disorders. In 1983, Gregory Fricchione and colleagues at Harvard reported that catatonia is responsive to the benzodiazepine lorazepam. In their 2003 book on catatonia, Max Fink and Michael Alan Taylor proposed a one- or two-dose benzodiazepine challenge as a convincing diagnostic test for catatonia. Biologic psychiatry proceeds from the premise that behavioral disorders are symptoms of underlying cerebral dysfunction. It is therefore unsurprising that biochemical markers have long existed in psychiatry.

Convulsive and Coma Therapy

Convulsive therapy, probably the most powerful psychiatric treatment on offer, had its origins in the 1930s, together with the other so-called “physical therapies” that addressed the brain itself.

First of the new physical treatments was insulin coma therapy, introduced in 1930 with the discovery by Manfred Sakel, then an assistant physician in a private nervous clinic in Berlin, that the withdrawal symptoms from morphine addiction could be relieved with insulin. It soon became apparent that the curative agent was an insulin-induced coma, and when, after the Nazi seizure of power in 1933, Sakel returned to his native Vienna, he persuaded the professor of psychiatry, Otto Pötzl, to let him experiment with insulin coma in the treatment of schizophrenia. From this work issued in 1934, a series of articles in a Viennese medical weekly that turned into a book in 1935, *A New Method of Treating Schizophrenia* (*Eine neue Behandlungsmethode der Schizophrenie*). A series of insulin comas does seem to have had a certain effectiveness in relieving the symptoms of chronic psychosis, and insulin remains today an underexplored agent in psychiatry. Some of Sakel’s patients also experienced seizures, and these as well may have been a therapeutic mechanism.

It was Budapest psychiatrist Ladislaus von Meduna who, in 1935, began the systematic exploration of chemically induced cerebral convulsions as a treatment. Meduna initially used camphor to initiate brain seizures, then the heart drug pentylenetetrazol. Meduna experienced extraordinary successes, especially in patients with the symptoms of catatonia, and his 1935 report in the Journal of Combined Psychiatry and Neurology (*Zeitschrift für die gesamte Neurologie und Psychiatrie*) is of historic importance in introducing the concept of therapeutic seizures to psychiatry. Yet many patients resisted chemically induced seizures because of the anxiety they experienced in waiting for the seizure to begin.

In 1938, Rome psychiatry professor Ugo Cerletti hit upon the idea of inducing the seizures with electricity so that the patients were instantly unconscious and underwent no anxious latency period. He quickly discovered that patients with melancholic depression responded readily to ECT, and the technique spread rapidly throughout psychiatry; it was at the time relatively uncontroversial. Cerletti had demonstrated not that convulsive therapy was effective (Meduna's doing) but, as modern ECT-investigator Richard Abrams puts it in his 1988 manual of convulsive therapy, "that such convulsions could be induced safely, reliably, and inexpensively by electrical means." Cerletti and the other early electrotherapists did not restrain their patients, and as a result, the "unmodified" seizures resulted in numerous fractures. To reduce such side effects, curare was first employed as a means of blocking the nerve–muscle junction, then, in 1952, succinylcholine, together with a barbiturate anesthetic. Yet even this modified ECT became in the 1960s an object of opprobrium in the counter-culture and then in the eyes of the pharmaceutical industry, which insisted that psychopharmacology could replace convulsive therapy. But it turned out that ECT offered distinct advantages in the treatment of melancholic depression, mania, and catatonia, and in 1978 the American Psychiatry Association offered an endorsement of it to counter the ballooning stigma caused in particular by the 1975 Hollywood movie, *One Flew Over the Cuckoo's Nest*. Max Fink, at the Stony Brook School of Medicine, led a campaign to rescue ECT from threatened oblivion. In 1990, the APA issued a much more enthusiastic endorsement of ECT, and it would be fair to say (as David Healy and I do in our 2007 history of shock therapy) that ECT has now resumed its prominent place in the panoply of psychiatric treatments.

Neuroimaging

Neuroimaging dates back to Harvey Cushing's 1896 radiograph of the cervical spine of a patient who had just been shot in the neck; the history of neuroimaging includes air encephalography, introduced in 1918 and used in Germany in 1927, to identify enlargements of the ventricular system in patients with schizophrenia. Gerd Huber's classic 1957 monograph *Pneumoencephalography and Psychopathology in the Endogenous Psychoses* (*Pneumooencephalographische und psychopathologische Bilder bei* *endogenen Psychosen*) was probably the high water mark of pneumoencephalography in psychiatry.

Carotid angiography, introduced in 1927, has never had significant psychiatric uses. But radionuclide brain scanning commenced in 1948, showed great promise once radioactive ligands were introduced in the 1970s in the form of single-photon emission tomography (SPECT) and positron emission tomography (PET).

Tom Bolwig, professor of psychiatry at the University of Copenhagen, outlines the following story: Among the earliest studies, in 1974 David H. Ingvar and Göran Franzen at Lund University reported in the Acta Psychiatrica Scandinavica decreased cerebral blood flow in chronic schizophrenia. SPECT and PET were employed in the 1980s to measure regional cerebral blood flow and metabolism and, later, neurotransmitter-binding potential. SPECT and PET are used as well to quantify the regional density of receptors of a given type and receptor-binding displacement. But these methods expose patients to radioligands and demand cyclotron capacity, so they cannot be applied outside of specialized centers. In 1988, Daniel Weinberger and associates at the NIMH employed PET to help develop the "hypofrontality" hypothesis in schizophrenia.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) entered medicine in the late 1970s. In psychiatry, the first controlled study using MRI was led by Nancy Andreasen, finding in 1986 in the Archives of General Psychiatry that a population of schizophrenic males had "markedly smaller frontal lobes." Functional MRI (fMRI) examines changes in blood oxygen level, a proxy measure of neural activity. The Yvette Sheline group at Washington University in St. Louis used fMRI in 2002 to demonstrate that long-standing depression produced a decrease in 5-HT(2A) receptors in the hippocampus, a functional finding. Structural MRI measures brain and regional gray and white matter volumes and was used in 2003 by Sheline and colleagues to measure the loss of hippocampal volume in severe untreated depression (the longer untreated, the greater the loss of volume). Thus, in Sheline's terms (in the 2002 article), "Functional imaging extends the sensitivity and specificity of structural imaging."

Brain Connectivity

Brain connectivity refers to a pattern of anatomical links ("anatomical connectivity"), statistical dependencies ("functional connectivity"), or causal interactions ("effective connectivity") between distinct units within the nervous system. Using a variety of MRI and fMRI methodologies, information about the functional and structural integrity of neural circuitry—"neural profiles"—can be provided in healthy and psychiatrically unwell individuals. In 2008, M.L. Phillips and associates identified abnormalities of connectivity in the neural circuitry supporting emotion processing and regulation in bipolar disorder. In 2010, in Biologic Psychiatry, Amelia Versace and colleagues attempted to differentiate unipolar versus bipolar depression on the basis of the functional white matter integrity of the neural circuitry regulating emotion.

Quantitative EEG (QEEG) Brain Mapping

Introduced in the 1980s, the QEEG brain map produces a three-dimensional reconstruction of the brain's activity. Using advanced algorithms, it is possible to identify brain-wave patterns from deep brain structures. The method is noninvasive but has a space resolution weaker than that of the above-mentioned methods. QEEG has been used by ER John and associates to identify activity clusters in a variety of different psychotic states. One reason for the lesser popularity of QEEG compared to SPECT, PET, and MRI may be, as Tom Bolwig observes, "that understanding the elements of electrophysiology seems a bit harder for many physicians than accepting colored spots on an image of the brain." Yet QEEG represents pure neuronal transport and can record immediately and for very long periods of time, which are essential qualities for monitoring chains of events.

THE BEGINNING OF THE SECOND BIOLOGIC PSYCHIATRY

Psychiatry in Europe, with its heavy emphasis on heredity, had always been biologically oriented. But in the 1930s, psychoanalysis in the United States wiped the biologic slate clean. Then in 1946, at a meeting at the Fairmont Hotel in San Francisco, the biologic thread began to re-emerge. Two California physicians—Johannes M. Nielsen, a neurologist at the University of Southern California, and his student, George N. Thompson, chief psychiatrist at the Los Angeles General Hospital, convoked those interested in “the biological basis of behavior.” The meeting was attended by the heavyweights in U.S. neurology and psychiatry: Percival Bailey, a Chicago neurologist; Stanley Cobb, professor of neuropathology at Harvard; and Karl M. Bowman, at the Langley Porter clinic in San Francisco. From this meeting emerged the Society of Biological Psychiatry. In 1947, Nielsen and Thompson published the first textbook of biologic psychiatry in the United States, *The Engrammes of Psychiatry*.

The Revolution in Psychopharmacology

Although the term “psychopharmacology” seems to have been coined by David Macht, a pharmacologist at Johns Hopkins University, in 1920, to describe “the effect of drugs on psychological functions,” it was not until the mid-1950s, with the explosion of interest in mechanisms prompted by the success of the antipsychotics, that the term started to come into use. When Paris psychiatrist Jean Thuillier visited Joel Elkes’s newly founded department of experimental psychiatry in Birmingham in the mid-1950s, he was introduced by Willy Meyer-Gross. (Meyer-Gross, a refugee from Hitlerism who previously had been a mainstay at Heidelberg, was finishing his career in Birmingham as an associate of Elkes.) Meyer-Gross said in the introduction, “From everything that will be said to you this evening, bear in mind this new word: psychopharmacology. We already practise it here without realizing it, and everything we do tomorrow and in the future, everything we will want to do, will be gathered into this new discipline."

There has been much discussion pro and con about psychiatry's turn toward psychopharmacology. For all current concerns about polypharmacy and "cosmetic psychopharmacology," one does well to bear in mind the recollections of Robert Felix, inaugural director of NIMH in 1949, of his days as a resident in a mental hospital in 1931,

> He explained that continuous flow baths were generally used for disturbed patients. "If the patient was too disturbed to lie in the tub, either a hot or a cold pack was used," which, while not specifically identified as such, acted as an effective physical restraint. For agitated patients, whether excited or depressed, "three chemical substances were used: paraldehyde, phenobarbital, or bromide." Felix noted that their bible was chapter XI of Bleuler's Textbook of Psychiatry. 'The Treatment of Mental Diseases in General'

In the years after the Second World War, psychiatry swung on a massive hinge. "There can be no twisted thought without a twisted molecule," wrote Ralph Gerard in 1955 in the American Journal of Psychiatry, keynoting the advent of a new era in psychiatry, the easing away from psychoanalysis, and the turn to pharmacotherapeutics as the main treatment modality. Biology marked, said Gerard 2 years later in Science, "the emphatic re-entry of the brain into the mental arena." In these years, the field evolved from being the lowest prescribing of the medical specialties to the highest. (In 2005, 90.1% of all psychiatric visits ended with a prescription; the next highest was cardiovascular diseases with 85.4%.)

The revolution may have begun with the synthesis in 1938 of lysergic acid diethylamide (LSD) by Sandoz chemist Albert Hofmann. It was in 1943 that Hofmann discovered it was psychoactive; LSD went on to have a brief career in psychotherapy and the treatment of alcoholism, but it generated great scientific curiosity: Was its mechanism identical to that of schizophrenia?

In 1952, the Rhône-Poulenc Laboratories in Paris introduced the first antipsychotic (neuroleptic), chlorpromazine, of the phenothiazine class of three-ringed structures, which was brand-named Largactil in Europe, Thorazine in the United States. The staff of Paris psychiatry professor Jean Delay conducted the initial clinical trials. A number of other phenothiazine antipsychotics followed.

These two compounds—LSD and chlorpromazine—initiated the era of psychopharmacology. As Theodore Sourkes at McGill University, “a biochemist in psychiatry” as he called himself, put it later,

> The neuroleptic action of chlorpromazine and the psychodysleptic action of LSD … held out the promise, on the one hand, of a chemical treatment of mental disorders and, on the other, of exploring the possibility that schizophrenia was caused by a toxic substance. As a consequence, the fifties and sixties of the last century saw the introduction of chemical theories into psychiatry. (Sourkes TL. A biochemist in psychiatry: dropping in on the psychopharmacology era. In: Ban TA, Healy D, Shorter E, eds. *Reflections on Twentieth-Century Psychopharmacology.* Animula Publishing House; 2004:290–298.)

The proliferation of phenothiazines gave rise to the classification of “sedative” versus “incisive” neuroleptics, the former presumably for agitated patients, the latter for the withdrawn and retarded. Yet an analysis by Arthur Platz and coworkers at the Perry Point Veterans Administration Hospital in 1967 demonstrated that there was no difference in therapeutic profiles.

A more important distinction was effectiveness in patients with “positive symptoms” versus those with the apathetic withdrawal of “negative symptoms.” This was realized early. Fritz Freyhan, a German émigré considered one of the founders of psychopharmacology, observed in 1957, at the Second International Congress of Psychiatry in Zurich (the first was in Paris in 1950), “Two schizophrenic patients … may differ fundamentally, one being apathetic, the other excited and disturbed. Only the latter is a candidate for neuroleptic treatment.”

The new drugs marked the beginning of the end of the dominance of psychoanalysis. Heinz Lehmann, who conducted some of the first trials of chlorpromazine, said in a 1992 interview in the Bulletin of the Canadian Psychiatric Association,

> “The demonstration [was important] that psychotropic drugs, such as chlorpromazine … are actually effective, so effective that there remains no doubt even for the sturdiest psychoanalyst. Eventually they had to admit there is a

> physical substrate to schizophrenia and bipolar disorder, because certain physical substances, drugs, definitely are helpful in treating these disorders. (Lehmann H. Psychopharmacotherapy. In: Healy D, ed. *The Psychopharmacologists.* CRC Press; 2001:182.)

It was thus that psychiatry, after so many years out in the cold with psychoanalysis, was able to rejoin the rest of medicine. As Theodore Rothman, at the University of Southern California and Los Angeles Veterans Administration Hospital, said in 1960, at an organizational meeting that would lead to the founding of the American College of Neuropsychopharmacology (ACNP) the following year, "I can find no better keynote thought than that psychopharmacology has given us a common front with the basic medical sciences. Psychopharmacology is compelling us to re-integrate psychiatry into medicine."

Back into medicine! But how did the new drugs work? The 1960s and 1970s were captivated by one-neurotransmitter-one-disease hypotheses.

Norepinephrine was first off the blocks. In 1965, in the American Journal of Psychiatry, Joseph Schildkraut of Harvard University proposed a "catecholamine hypothesis of affective disorders" (the neurotransmitter norepinephrine has a portion with a catechol structure). "Those drugs which cause depletion and inactivation of norepinephrine centrally produce sedation or depression, while drugs which increase or potentiate brain norepinephrine … generally exert an antidepressant effect in man."

In 1968, Swedish pharmacologist Arvid Carlsson and coworkers at Gothenburg University noted in the Journal of Pharmacy and Pharmacology that the TCA imipramine blocked the reuptake of serotonin. "This action may be of importance for its antidepressive properties," the authors said, initiating the serotonin hypothesis of depression.

It remained for dopamine, the third major biogenic amine neurotransmitter, to find its distinctive disease, and this occurred in 1976, as it was reported simultaneously by psychiatrist Solomon Snyder at Johns Hopkins University (in Science) and pharmacologist Philip Seeman at the University of Toronto (in Nature) that the potency of antipsychotic drugs in schizophrenia was a function of their ability to block the dopamine receptor.

The original scientific findings underlying these reports were never overturned, yet neurotransmission is a complex phenomenon, and the field quickly lost confidence in the specificity hypothesis—although, for decades, it was marketed relentlessly by the pharmaceutical industry. In 1978, J.L. Costa and coinvestigators at NIMH said of agents affecting serotonin metabolism, "… It certainly may be that a serotonin deficit state is not a prerequisite for therapeutic efficacy with any of these drugs." It is a shame that these hypotheses, so attractive for their theoretical elegance and simplicity, have now been largely discredited because psychopharmacology has never again been able to develop a unitary theory of drug action. At this writing, the brain remains a black box.

The Golden 50s

The year 1955 saw the launch of the first blockbuster drug in psychiatry, Wallace Laboratories' meprobamate, trade-named Miltown (and Equanil by Wyeth Laboratories, which licensed it from Wallace). Launched as a "tranquilizer," Miltown became almost a household word, opening up that large market for "nerves" between the antipsychotics and aspirin (see below). It was displaced by the benzodiazepines in the 1960s.

The monoamine oxidase inhibitors (MAOIs) were probably the first effective drug class, aside from opium, for serious depression. This awareness took a while dawning. The first MAOI, iproniazid (Hoffmann-La Roche's Marsilid), reached the market in 1951 for tuberculosis, but only in 1957 did the company begin to market it for depression. A series of other MAOIs followed, the most effective of which was probably phenelzine (Warner-Chilcott's Nardil), introduced in 1959 and still in use.

"Tranquilizers"

It was London physician Edward Sutleffe, who had a substantial practice of patients with "nervous irritation" among other diagnoses, who in his *Medical and Surgical Cases* in 1824, introduced into medicine the term "tranquilizer" in connection with a "tranquilizing" drug, "the juice of ground ivy," that he was promoting. The word then slumbered in forgetfulness until 1953, when Ciba billed its new antihypertensive Serpasil (reserpine) as a "safe tranquilizer-antihypertensive." Its calming effects were soon noted in psychological medicine, and the concept of "tranquilization" became the common denominator of a range of drugs from the antipsychotic chlorpromazine, introduced clinically in the United States in 1954, to the "antiphobic" agent benactyzine (Suavitil) that Merck brought out in 1957. The tranquilizers were thought an improvement upon the barbiturates because they had the same relaxing and composing effects without the daytime sedation. Yet soon, the concept of tranquilizer was ruptured by the introduction of the benzodiazepines after 1960 (see below). These "antianxiety" agents made the concept of tranquilization, which had lumped together antipsychotics and antinervous agents, seem outmoded.

Antipsychotics

The antipsychotic era began with the introduction of chlorpromazine in clinical trials in France in 1952 by the Rhône-Poulenc Laboratories. Called Largactil in Europe and Thorazine in the United States, chlorpromazine is thought to have initiated the era of psychopharmacology by prompting research into mechanisms. A slew of similar phenothiazine antipsychotics, also called neuroleptics, followed, including Smith Kline's prochlorperazine (Compazine) in 1956 and thioridazine (Sandoz's Mellaril) in 1958. Haloperidol (Janssen's Haldol, a member of the butyrophenone class) entered practice in the United States in 1966.

Tricyclic Antidepressants

Owing to the close clinical observation of Swiss psychiatrist Roland Kuhn, Geigy Laboratories in Basel realized in 1957 that their new compound imipramine, a member of the tricyclic class, was active against "vital depression," rather than being an antischizophrenic, as they had hoped (the antipsychotic market was just taking off). Marketed in the United States in 1959, imipramine (brand-named Tofranil) initiated a half-century of “antidepressant” treatment. News of the TCAs traveled quickly. At a conference at McGill University in 1959, Hans Hoff, who had become the professor of psychiatry in Vienna after the Second World War, admitted that he and psychoanalyst colleague Paul Schilder “had once tried to destroy the Kraepelinian nomenclature because of the feeling that psychodynamics and personal reactions were more important. Now, however, since a treatment had become available, diagnosis is essential.”

“Mood Stabilizers”

The mood-stabilizer era began with lithium in 1949. The term apparently surfaced for the first time in connection with a drug called aminophenylpyridone (Maltbie Laboratories’ Dornwal) in 1960. The phrase “mood stabilizer” did not, however, come into general use until Robert Post and colleagues at NIMH floated the “kindling hypothesis” in 1986, in the British Journal of Psychiatry: recurrent depressions could tip the brain into autonomous cycles of mood disorder. Other writers then proposed that various anticonvulsants could break into the kindling process and stabilize the hyperirritable brain. Aside from lithium, considered mood stabilizers were: derivatives of valproic acid called valproate and valpromide that medicinal chemist George Carraz at Berthier Laboratories in Grenoble synthesized in 1963; carbamazepine, patented in 1960 by Geigy, was introduced as a psychoactive compound in Japan, then discovered in 1980, by NIMH investigators to have effects in manic-depression. Lamotrigine was patented by the Wellcome Foundation in 1981 and inaugurated in 2003 in the United States by GlaxoSmithKline as Lamictal (“a stabilizing force”). The judgment of David Healy and colleagues in 2003, in Bipolar Disorders, was that “These agents are now thought of as being a homogeneous group [but] they may, in fact, be quite diverse agents all of which have a certain utility when used judiciously in manic depressive states.”

After the 1950s, Relatively Few New Drug Classes

Benzodiazepines. The benzodiazepine story began with the marketing of Hoffmann-La Roche's Librium (chlordiazepoxide) in 1960. The numerous subsequent Roche benzos were also synthesized by medicinal chemist Leo Sternbach. So popular did the benzodiazepines become that entire "eras" in the 1960s and 1970s were named after such members of this class as Valium (diazepam, launched in 1963). With the exception of the SSRIs (selective serotonin reuptake inhibitors—see below), the benzodiazepines are probably the best-selling psychoactive drug class in history and moreover possessed genuine efficacy that later agents do not seem to have matched. Other benzos followed: clonazepam (Klonopin) reached the American market in 1975, and alprazolam (Upjohn's Xanax) was launched in 1981. A short-acting benzo, Upjohn's triazolam (Halcion), came out in the United States in 1982 and caused much controversy about discontinuation. There was a long period of fearfulness about addiction, although in 1983, in the Journal of the American Medical Association, Karl Rickels and associates at the University of Pennsylvania found the risks of withdrawal reactions in diazepam minimal for up to 8 months of use. In the new millennium, the benzos began a comeback. As Peter Roy-Byrne at the University of Washington School of Medicine put it in an editorial in 2014 in the American Journal of Psychiatry, "There is now general appreciation that, with careful clinical evaluation to rule out pre-existing addiction issues … benzodiazepines are safe and highly effective."

Selective Serotonin Reuptake Inhibitors (SSRIs). Marketed as agents that specifically inhibited the neuronal reuptake of serotonin, this drug class became the most successful in the history of psychiatry. Prozac (fluoxetine), which became a household word, was marketed by the Eli Lilly Company in Belgium in 1986, in the United States in 1988. Among the main SSRIs, there followed Pfizer's Zoloft (sertraline), marketed in 1992; paroxetine (Novartis's Paxil), introduced in the United States in 1993; and citalopram

(Lundbeck’s Cipramil), brought out in the United States in 1998 by Forest Laboratories (Celexa). The SSRIs were initially indicated for “depression,” then for “anxiety,” and by the time they all finally went off patent, countless were the indications for which this versatile drug class served.

Second-Generation Antipsychotics. The so-called “second-generation antipsychotics” (SGAs) have turned out not that different from the first generation and seem to be more a marketing concept. Yet, it might be noted that clozapine (Sandoz’s Clozaril), deemed the most effective of them, is not exactly a new drug, having been patented in 1963. Leo Hollister’s judgment in 1987 of the antipsychotics is of interest (Hollister led the Veterans Administration hospital in Palo Alto, California and is generally acknowledged as the dean of American psychopharmacology): “It is most discouraging that more effective pharmacotherapy for schizophrenia has not been developed in the more than three decades since the introduction of the first effective drugs.” The current drugs, he said,

> are not curative; their ameliorative effects are often limited; many patients remain totally unresponsive; they are unpleasant to take, so that many patients are less than fully compliant; and they produce major side effects.

PEDIATRIC PSYCHIATRY

Psychiatrists have always been familiar with mental diseases in children. In 1809, Haslam reported cases of “insane children,” such as a 7-year-old boy who was slow to develop and was placed on “the female side” of the hospital. Early on, psychotic illness in children was noted, sometimes as a distinctive form of what would later be called childhood schizophrenia, sometimes not. The “hebephrenia” that Ewald Hecker described in 1871 was commonly recognized in children. As Auguste Voisin, a staff psychiatrist at the Salpêtrière in Paris, commented in his 1883 textbook of “insanity in the young”: “In partial insanity, the child may be manic [extatique], subject to ideas of homicide, arson, or theft, have deluded and demonic ideas … It is in the context of this partial insanity that we find the singular cases of childhood suicide, sometimes preceded by notes from which extreme exaltation shows through; thus a child of thirteen who hanged himself left behind a note saying, ‘I bequeath my soul to Rousseau, my body to the earth.’”

Textbooks of child psychiatry were not slow in coming. In 1887, Hermann Emminghaus, professor of psychiatry in Freiburg, produced what was probably the first, *The Psychic Disorders of Childhood* (*Die psychischen Störungen des Kindesalters*). The foundational modern text of psychological approaches to disorders of childhood, *Lectures on the Psychopathology of Childhood* (*Vorlesungen über Psychopathologie des Kindesalters),* was written by August Homburger at the Heidelberg University Psychiatric Hospital in 1926. Among the first dedicated hospital divisions for children with mental illnesses was that at Bicêtre Hospice in Paris, which in 1838, received responsibility for the care of “children with mental retardation and epilepsy.” In 1842, Édouard Séguin, who had developed a system for the care of children with retardation, took over the service, and after an interregnum, in 1851, Louis Delasiauve succeeded him. Désiré-Magloire Bourneville took it over in 1879, and in the course of the next decades, made Bicêtre into one of the most progressive children’s psychiatric units in the world. (In 1892, he initiated a gymnastic competition, which became the unacknowledged forerunner of the Special Olympics.) In 1850, Séguin emigrated to the United States and initiated here the enlightened treatment of children who have what are now called intellectual disabilities.

The epidemic of encephalitis lethargica in the years 1917 to 1927 represented a strong inflection in the growing awareness of distinctive psychiatric illness in children, primarily because, unlike adults, the disease produced striking psychiatric changes in many children, transforming them from normal happy creatures into the “apache type” with severe behavioral problems. Said Leslie Colin Cook, who was a staffer at the West Park Mental Hospital in Epsom that had a center for such patients, in the Lancet in 1928, “Children of irreproachable family, hitherto of exemplary behavior and character, may in a very short period become guilty of every delinquency from mere naughtiness to cruelty and murder, and from mild bad habits to gross indecencies." This had happened, he said, in only one adult patient with the disease. The "EL" epidemic thus strengthened the interest of the field in children as a special kind of patient.

In the middle third of the 20th century, psychoanalysis became the doctrinal mainstay of child psychiatry. Freud himself did not see children and had relatively little experience with them. Psychoanalysis was first applied to the child, and adolescent population by Viennese educator and psychoanalyst August Aichhorn in his 1925 work Wayward Youth (Verwahrloste Jugend), about the ego structure of juvenile delinquency. Sigmund Freud's daughter Anna published her first book in 1927, An Introduction to the Technique of Child Analysis, and in 1936, The Ego and the Mechanisms of Defense. The Freud family fled to England after the Nazi Anschluss with Austria in 1938, and in 1941, Anna Freud helped found the Hampstead Nursery in London, where she treated children for several years. This gave way to the Hampstead Child Therapy Course and Clinic, which Anna directed until her death in 1982 (on her rivalry with Melanie Klein see above).

After the Second World War, the baton in child psychiatry passed to the United States. When the National Institute of Mental Health opened in 1949, the "quality of mothering" was put on the table—especially so by psychoanalysts such as Frieda Fromm-Reichmann at Chestnut Lodge in Maryland, who, in Psychiatry in 1948, had originated such concepts as the "schizophrenogenic mother." Among the leading child psychoanalysts in New York was Hungarian-born Margaret Mahler, who in 1950 cofounded the Masters Children's Center in Manhattan, where she developed the tripartite treatment model involving the mother in the child's treatment. Her most important work was *The Psychological Birth of the Human Infant* (1975).

Prompted by an interest in the ideas of Harvard psychologist Burrhus Frederick Skinner, in the late 1940s and 1950s, behaviorism started to become à la page. Psychoanalysis and behaviorism dominated pediatric psychiatry in the United States after the Second World War.

The child guidance clinics reached back to the First World War, and in 1948 they created an organization of clinics, the American Association of Psychiatric Clinics for Children, to develop standards of care and training. In 1953, the American Academy of Child Psychiatry was founded; in 1959, the American Board of Psychiatry and Neurology approved a 2-year child psychiatry residency to be grafted onto the regular training. The field received its own voice, The Journal of the American Academy of Child Psychiatry, in 1962, and a quarter-century later, in 1986, the Academy changed its name to the American Academy of Child and Adolescent Psychiatry.

In the heyday of psychoanalysis, “pills” were frowned on and great confidence placed in psychotherapy. Yet, the revolution in adult psychopharmacology in the 1950s had echoes in the pediatric world. “Psychopharmacology has revolutionized child psychiatry,” said Jovan Simeon, professor of psychiatry and child psychopharmacologist at the University of Ottawa, in 2004. Children had, of course, always been treated with such psychoactive agents as opium, mixed together with alcohol in laudanum (see above), which parents once routinely administered as a hypnotic. Treating children with psychopharmaceuticals entered the academic forum in 1937, when Charles Bradley, director of the Emma Pendleton Bradley Home in East Providence, Rhode Island, in the American Journal of Psychiatry reported the benefit of Benzedrine in “behavior problem” children, later defined as “hyperactive.” Loretta Bender, at Bellevue Hospital in New York, used chlorpromazine and amphetamine extensively in treatment. Systematic research in pediatric psychopharmacology began in 1958 when Leon Eisenberg at Johns Hopkins University received the first NIMH grant to study drugs in children, conducting the first controlled trial in pediatric psychopharmacology. This torch was then notably carried forward by Bender’s student Barbara Fish, first at Bellevue, then after 1972, at UCLA.

Such progress in pediatric psychopharmacology probably counts as an American “first.” There were a number of other American firsts, such as recognition by Lee Wachtel at the Kennedy Krieger Institute in Baltimore and Dirk Dhossche, in a 2010 article in Medical Hypotheses, of stereotypical movements in autism and intellectual disability as a sign of catatonia—and treatable with benzodiazepines and convulsive therapy.

As for diagnosis, disease classification in pediatric psychiatry has been in even greater turmoil than in general psychiatry, as psychoanalysis, behaviorism, and biologic approaches have dueled. Driven by parental concerns, clinicians are often pressed to make diagnoses such as ADHD. The fourth edition of DSM (DSM-IV) included the diagnosis "bipolar II disorder," characterized by depression plus "at least one hypomanic episode." Yet the field readily accepted outbursts of anger as a substitute for hypomania, touching off a virtual epidemic of "bipolar II" in the pediatric population, which the embarrassed drafters of DSM-5 were obliged to dial back in 2013, with the newly contrived "depressive" diagnosis "disruptive mood dysregulation disorder," said to be "included for children up to age 18 years who exhibit persistent irritability and frequent episodes of extreme behavioral dyscontrol." The classification of illness in children has thus still not stabilized.

NEW FORMS OF CARE

The asylum was a forbidding institution. Yet many patients outside its walls also required care, and over the years, these walls were slowly broken down.

Boarding patients out to local families achieved first this separation of care from the four walls of the institution. The earliest instance of this was seen at Gheel in the Brabant region of Belgium, where patient out-boarding went back to the 13th century. Yet the local farmers often had little tolerance for the antics of the patients, of whom they might board up to five at a time, and when Pliny Earle, of the Friends' Asylum in Frankford, visited there in the late 1830s (describing his travels in a book in 1841, he was shocked to find that the patients, "if mischievous or furious, are chained at the wrists and ankles. In every house, there is a ring beside the chimney, or attached to the beds, to which they may be chained." Boarding out was never widespread in the United States, possibly because the farms were too dispersed (unlike the compact Belgian villages). But in Scotland, around 1900, 23% of the asylum patients were boarded out.

There were other ways of breaking down the walls. The first psychiatric ward in a general hospital in the United States was established in 1902 at Albany Hospital. In 1885, Pennsylvania Hospital founded the first outpatient clinic, or very close to the first, in the United States, "so as to provide a dispensary service for the gratuitous treatment of persons suffering with incipient mental disease." (On these developments, see Henry Hurd's massive four-volume work in 1916 on the Institutional Care of the Insane.)

One of the unhappiest stories in U.S. psychiatry was that of "deinstitutionalization," under the influence of the antipsychiatric ideologies of the 1960s, discharging in-patients to a kind of community care that often turned out to be the care of the streets. This created an existential dilemma for psychiatry: what to do with the patients who were seriously ill? As Jonathan Cole said in an interview in 1996,

> We've got few places where we can serve the kind of patient who takes a long time to get better and where real rehabilitation is done. We tend to have more people who are home and crazy than we should have and nobody's going to pay for their treatment.

Beginning in the 1960s, this vast outpouring resulted in the closing of many of the classic asylums. In 1964, there were in the United States 255 public psychiatry beds per 100,000 population; by 2014, this figure had dropped to 11 per 100,000. The patients were supposed to be maintained in community settings on antipsychotics; in reality, they fed the vast ranks of the "homeless" and the imprisoned. In 2002, 64.2% of all inmates of local jails in the United States had a "mental health problem" (75.4% of the females). The figures were comparable for state and federal prisons. Donald Klein of Columbia University said later in an interview,

> "It wasn't deinstitutionalization, it was trans-institutionalization. The patients, but not the money, were dumped on the cities. There was a sudden rise in homelessness. This led to federally supported 'community care,' staffed by therapy-minded psychologists. They had a psychiatrist who did little but sign bundles of prescriptions.

By one estimate, "There are 10 times as many mentally ill persons in prisons than in state hospital beds." David Rubinow at the UNC, who penned these lines in the American Journal of Psychiatry in 2014, concluded,

> It would appear that we somehow have achieved the complete reversal of policies initiated over 170 years ago to ensure the humane treatment of those with mental illness.

Yet history has good news as well. On the threshold of the new millennium, the role of psychiatry itself was slowly changing. Once seen as a band-aid for the "walking wounded," psychiatry was becoming the main resource for the seriously ill. In the United States from 2001 to 2002, 40% of patients identified in a national household survey with a serious mental problem would be seen by a psychiatrist. As for the rest of patients who formerly would have been "in therapy," psychiatry had ceased to be the main port of call: only 21% of those with moderate and 9% of those with mild disorders saw psychiatrists; they distributed themselves instead across the entire gamut of mental and medical health services, from "general medical" to "internet" (the main resource for 30% of those with "moderate" mental problems). Alarmingly, of those seeing a psychiatrist, 29% were said to have "no disorder." Readers will give a weary sigh. Yes, it's true: some things never change.

FURTHER READINGS

Arns M, Conners CK, Kraemer HC. A decade of EEG theta/beta ratio research in ADHD: a meta-analysis. *J Atten Disord*. 2013;17(5):374–383.

Billod E. *Les Aliénés en Italie*. Masson; 1884:72.

Cathell DW. *Book on the Physician Himself*. 10th ed. Davis; 1900:150–151.

Chiarugi V. Della Pazzia in Genere e in Specie, 3 vol. 1793–94; reprint Firenze: Carlieri; 1991: I, 211–212; II, 67, 75.

Cohen RA. Interview. In: Ban TA, ed. *An Oral History of Neuropsychopharmacology*. I, 194. ACNP; 2011.

Ellenberger H. A comparison of European and American psychiatry. *Bull Menninger Clin*. 1955;19(2):43–52, 49.

Engel JW. Early psychiatry at Yale: Milton C Winternitz and the founding of the department of psychiatry and mental hygiene. *Yale J Biol Med*. 1994;67(1–2):33–47, 39.

Felix RH. Foreword: the life and times of psychiatric research. In: Earl U, et al., eds. *Biochemistry of Mental Disorders*. Dekker; 1978, vii-xvii, viii-ix.

Freyhan F. Discussion. In: Kline NS, ed. *Psychopharmacology Frontiers: Proc. of the Psychopharmacology Symp., Second Int. Congress of Psychiatry*. Little Brown; 1959:423.

Gerard RW. Orientation: analysis of program. In: Cole JO, Gerard RW, eds. *Psychopharmacology: Problems in Evaluation*. National Academy of Sciences; 1959, 9–19, xvii, 662 pp.

Haslam J. *Considerations on the Moral Management of Insane Persons*. Chapman and Hall; 1996.

Healy D. *The Psychopharmacologists: Interviews*, 3 vol. Altman; 1996–2000: I, 178.

Hoenig J. Editorial: Psychiatric nosology. *Can J Psychiatry*. 1981;26(2):85.

Hoff P. Discussion in Anon. [Proceedings of] McGill University Conference on Depression and Allied States, Montreal, March 19–21. *Can J Psychiatric*. 1959;4(spec supplement):S67–S68.

Hollister LE. Strategies for research in clinical psychopharmacology. In: Herbert YM, ed. *Psychopharmacology: The Third Generation of Progress*. Raven Press; 1987, 31–38, 31.

Lader M. Interview. In: Healy D, ed. *The Psychopharmacologists*. Chapman & Hall; 1996: I:463–481, 470. Hereafter cited as “Healy, Psychopharmacologists.”

Lehmann H. Interview. In: Healy D, ed. *Psychopharmacologists*, I, 182.

Letchworth WF. *The Insane in Foreign Countries*. Putnam’s; 1889:172.

Lindon JA. Melanie Klein. In: Alexander F, et al., eds. *Psychoanalytic Pioneers*. Basic Books; 1966:360–372, 367.

Litchfield HR. Aminophenylpyridone: a new mood-stabilizing drug. *Arch Pediatr*. 1960;77:133–137.

Moreau J-J. Notes sur les établissements d’aliénés de Siegburg, Halle, Dresde, Prague, Berlin et Vienne. *Annales Médico-psychologiques*. 2nd ser, 6. 1854;428–439, 430.

Murphy DL, Campbell I, Costa JL. Current status of the indoleamine hypothesis of the affective disorders. In: Lipton MA, et al., eds. *Psychopharmacology: A Generation of Progress*. Raven Press; 1978, 1235–1247, 1235.

Pandy K. Die Irrenfürsorge in Europa, *German trans from Hungarian*. Reimer; 1908:482.

Report of the Metropolitan Commissioners in Lunacy to the Lord Chancellor. Bradbury; 1844:37.

Rosenthal D. The heredity-environment issue in schizophrenia: summary of the conference and present status of our knowledge. In: David R, Seymour SK, eds. *The Transmission of Schizophrenia*. Pergamon; 1968:413–415, 415.

Shepherd M. Interview with Sir Aubrey Lewis (Part I). *Psychiatric Bull*. 1993;17:738–747, 743.

Shorter E. *A History of Psychiatry: From the Era of the Asylum to the Age of Prozac*. Wiley; 1997.

Shorter S. *A Historical Dictionary of Psychiatry*. Oxford University Press; 2005.

Shorter S, Fink M. *Endocrine Psychiatry: Solving the Riddle of Melancholia*. Oxford University Press; 2010.

Simpson G. Interview. In: Thomas Ban, ed. *An Oral History of Neuropsychopharmacology*. The First Fifty Years: Peer Interviews (ACNP; 2011), IV:373–384, 373; hereafter cited as “Ban, Oral History Neuropsychopharmacology.”

Snell LDC. Ueber die verschiedenen Formen des Wahnsinns. *Allgemeine Zeitschrift für Psychiatrie*. 1874;30:319–326, 320–321.

Sourkes TL. A biochemist in psychiatry: dropping in on the psychopharmacology era. In: Thomas AB et al., eds. *Reflections on Twentieth-Century Psychopharmacology*. Animula; 2004:290–298, 291–292.

Substance Abuse and Mental Health Services Administration. Mental Health: United States, 2008. *HHS Pub. No. (SMA) 10–4590*. Center for Mental Health Services, SAMHSA; 2010:123.

Thuillier J. Ten Years that Changed the Face of Mental Illness (1981), *Eng Trans*. Dunitz; 1999:139.

Tissot SA. *Traité des Nerfs et de Leurs Maladies*. Didot; 1780:II(2):358.

Voyer I. Anxietate. Mostert: Leiden med diss; 1769.

Westphal A. Manie, Melancholie, manisch-depressives Irresein. In: Otto B, et al., eds. *Lehrbuch der Psychiatrie*. Fischer; 1923:95–141, 119.

Woodward SB. Observations on the medical treatment of insanity. 1850. *Am J Psychiatry*. 1994;151(6 Suppl):220–230.

60

World Aspects of Psychiatry

SAM GNANAPRAGASAM, B.SC., M.B.B.S., M.R.C.PSYCH., MAX PEMBERTON, M.B.B.S., B.SC., M.R.C.PSYCH., PEDRO RUIZ, M.D., ANTONIO VENTRIGLIO, M.D., PH.D., AND DINESH BHUGRA, C.B.E., M.A., M.SC., M.B.B.S., F.R.C.P., F.R.C.P.E., F.R.C.PSYCH., F.F.P.H., M.PHIL., PH.D., F.R.C.PSYCH. (HON), F.H.K.C.PSYCH. (HON), F.A.C.P. (HON), F.A.M.S. (SINGAPORE), F.I.M.S.A. (HON)

Mental health disorders are highly prevalent across all regions of the world. These disorders represent a major source of disability and social burden. Even though treatments for these disorders are known and have been found to be efficacious in both high-, middle-, and low-income countries, this has not always translated into wider access, availability, and affordability of these interventions. Mental disorders remain undertreated worldwide, even when accounting for variation across societies and cultures. A treatment gap of more than 90% has been documented in some low-income countries, but even in high-income countries, high levels of gap remain. To complicate matters further, many countries do not have any mental health policies, and there are many others that have outdated policies originating in the 19th century. Consequently, both funding for treatment and research into mental health falls well below what is required or what is offered for physical disorders such as cancer and heart disease.

In recent years, there has been a focus on mental health provision around the world. This has led to the emergence of the global mental movement. Partly this is a response to rapid globalization, which has meant that people, resources, and ideas are moving much more rapidly than at any other time in the history of humanity. This has led to a clear acknowledgment that a number of factors at social and cultural levels are at play in influencing how distress is experienced and expressed. The thinking behind this movement is outlined in seminal literature such as the 2007 and 2011 Lancet Series on Global Mental Health and later revised 2018 Lancet Commission on Global Mental Health and Sustainable Development.

There remains a danger in that global mental health may follow the route of Anglo-American influences to be seen “dictating” to the low- and middle-income countries. Reassuringly, the latest commission report has reframed mental health within the Sustainable Development Goals framework and sought to establish mental health as a pillar of universal coverage. It has also sought to advocate for the use of public policies to protect mental health, including through investment and research innovation, as well as engagement with people with lived experience and strengthening of monitoring and accountability.

The international progress of collaboration has shifted considerably over the past two centuries. Initially, the rulers saw the vanquished as exotic natives who either did not have the mental capabilities to develop mental illness or, later, saw them as victims of biologic, social, and geopolitical factors. This gave way to international collaborative epidemiologic studies, which have contributed to the current understanding of both similarities and differences in etiology and outcomes of psychiatric disorders.

Although there have been many proclamations internationally recently about the importance of mental health, resources to deal with mental disorders and investing in public mental health have not followed the rhetoric. The contribution of funding to mental health care remains scarce and unequally distributed.

The authors prefer and choose the term “world psychiatry” in this chapter, as it reflects the real state of psychiatry at the present time. This chapter provides an overview of world psychiatry but focuses on epidemiology, the burden resulting from mental illness, the relationship between various illnesses, stigma, and the ethics of mental health care. There remain several serious questions in relation to the differences in rates and responses across different nations of the world. Some of these are discussed later in this chapter.

PREVALENCE AND BURDEN OF MENTAL DISORDERS WORLDWIDE

World Health Organization (WHO) estimates that more than one in four individuals worldwide develop one or more mental disorders during their lifetime. The Global Burden of Disease (GBD) project and studies that were undertaken provide specific data on multiple diseases from around the world. Herein Table 60–1, we report mental health prevalence findings from the 2017 study.

These findings in Table 60–1 are supported by a series of face-to-face community surveys conducted in 17 countries in Africa, Asia, the Americas, Europe, and the Middle East. These World Mental Health Surveys have included a total of 85,052 respondents. The lifetime prevalence of any mental disorder was found to be higher than 25% in 11 of 17 countries and higher than 15% in all countries except China and Nigeria, where prevalence rates were likely to be downwardly biased due to respondent reluctance to admit mental health problems. Whether these rates are genuinely low or whether these reflect reporting bias needs further exploration.

Not surprisingly, anxiety disorders were the most prevalent in 10 out of 17 countries, and mood disorders in 6 (Table 60–2). Anxiety disorders include agoraphobia, adult separation anxiety disorder, generalized anxiety disorder, panic disorder, posttraumatic stress disorder, social phobia, and specific phobia. Mood disorders include bipolar disorders, dysthymia, and major depressive disorder. Impulse-control disorders include intermittent explosive disorder and reported persistence in the past 12 months of symptoms of three child–adolescent disorders (attention-deficit/hyperactivity disorder, conduct disorder, and oppositional-defiant disorder). Substance disorders include alcohol or drug abuse with or without dependence. Following the *Diagnostic and Statistical Manual of Mental Health Disorders*, Fourth Edition (DSM-IV) norms, organic exclusions were excluded.

Lifetime co-occurrence of two or more mental disorders was quite common, so the sum of the prevalence rates across the various disorder types was between 30% and 50% higher than the proportion of people having at least one mental disorder. Survival analysis showed that the odds ratios for anxiety, mood, and substance use disorders were higher in recent compared to older cohorts. It is worth noting that even within the same geographical region, there appear to be major variations in the prevalence of DSM-IV/CIDI disorders.

Table 60–1.

Mental Health Disorder Prevalence from Global Burden of Disease 2017 Study

Disorder	Share of Global Population with Disorder (2017) [difference across countries]	Number of People with the Disorder (2017)	Share of Males: Females with Disorder (2017)
Any mental health disorder	10.7%	792 million	9.3% males 11.9% females
Depression	3.4% [2–6%]	264 million	2.7% males 4.1% females
Anxiety disorders	3.8% [2.5–7%]	284 million	2.8% males 4.7% females
Bipolar disorder	0.6% [0.3–1.2%]	46 million	0.55% males 0.65% females
Eating disorders(clinical anorexia & bulimia)	0.2% [0.1–1%]	16 million	0.13% males 0.29% females
Schizophrenia	0.3% [0.2–0.4%]	20 million	0.26% males 0.25% females
Any mental or substance use disorder	13% [11–18%]	970 million	12.6% males 13.3% females
Alcohol use disorder	1.4% [0.5–5%]	107 million	2% males 0.8% females

Drug use disorder (excluding alcohol)	0.9% [0.4–3.5%]	71 million	1.3% males 0.6% females

Reprinted from Skallevold HE, Rokaya N, Wongsirichat N, Rokaya D. Importance of oral health in mental health disorders: an updated review. *J Oral Biol Craniofac Res*. 2023;13(5):544–552. doi: 10.1016/j.jobcr.2023.06.003. Adapted from Dattani S, Rodés-Guirao L, Ritchie H, Roser M. Mental Health. 2021. https://ourworldindata.org/mental-health.

Table 60–2. Lifetime Prevalence Estimates of DSM-IV/CIDI Disorders in the WMH Surveys

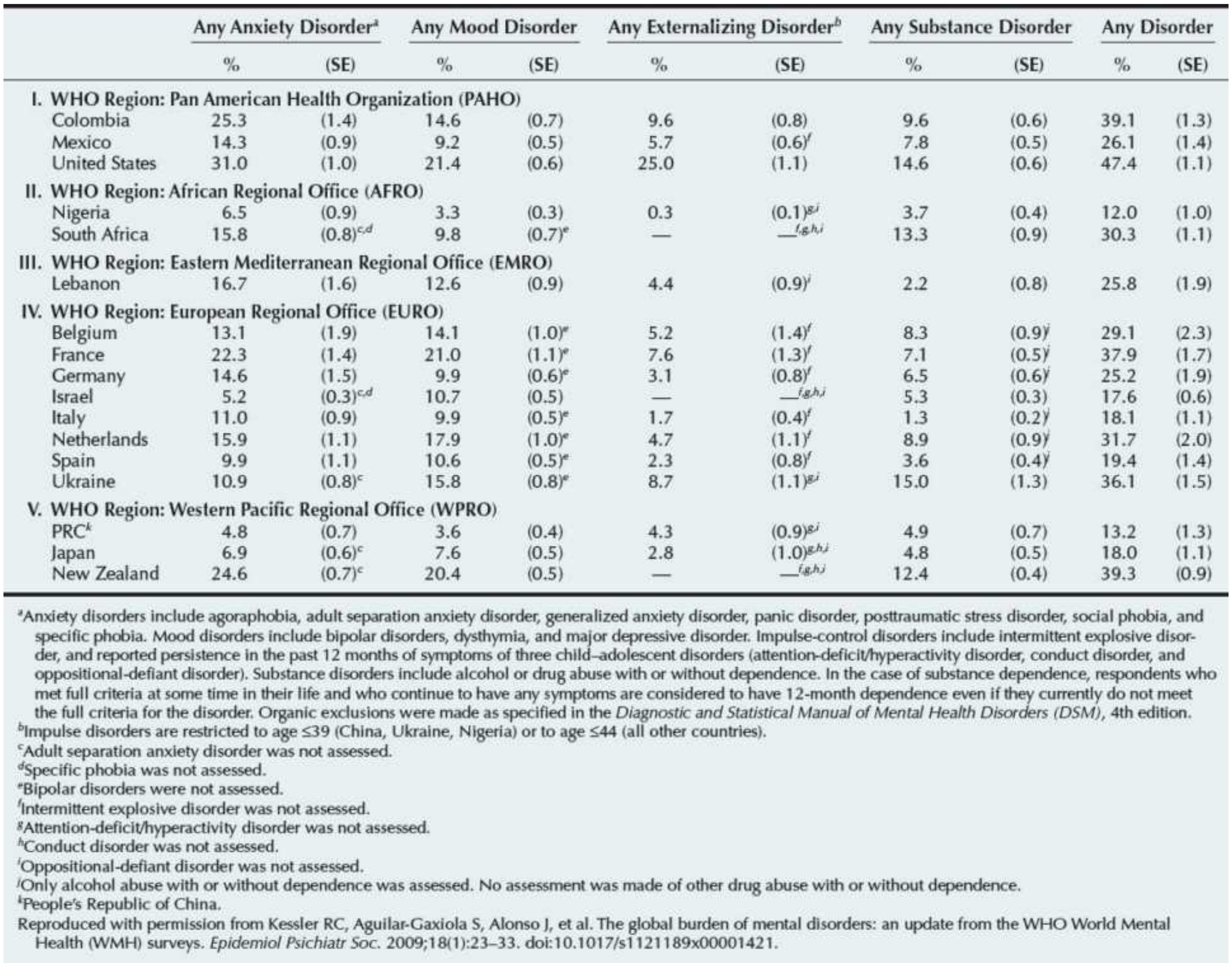

	Any Anxiety Disorder[a]		Any Mood Disorder		Any Externalizing Disorder[b]		Any Substance Disorder		Any Disorder	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)
I. WHO Region: Pan American Health Organization (PAHO)										
Colombia	25.3	(1.4)	14.6	(0.7)	9.6	(0.8)	9.6	(0.6)	39.1	(1.3)
Mexico	14.3	(0.9)	9.2	(0.5)	5.7	(0.6)[f]	7.8	(0.5)	26.1	(1.4)
United States	31.0	(1.0)	21.4	(0.6)	25.0	(1.1)	14.6	(0.6)	47.4	(1.1)
II. WHO Region: African Regional Office (AFRO)										
Nigeria	6.5	(0.9)	3.3	(0.3)	0.3	(0.1)[g,i]	3.7	(0.4)	12.0	(1.0)
South Africa	15.8	(0.8)[c,d]	9.8	(0.7)[e]	—	—[f,g,h,i]	13.3	(0.9)	30.3	(1.1)
III. WHO Region: Eastern Mediterranean Regional Office (EMRO)										
Lebanon	16.7	(1.6)	12.6	(0.9)	4.4	(0.9)[i]	2.2	(0.8)	25.8	(1.9)
IV. WHO Region: European Regional Office (EURO)										
Belgium	13.1	(1.9)	14.1	(1.0)[e]	5.2	(1.4)[f]	8.3	(0.9)[j]	29.1	(2.3)
France	22.3	(1.4)	21.0	(1.1)[e]	7.6	(1.3)[f]	7.1	(0.5)[j]	37.9	(1.7)
Germany	14.6	(1.5)	9.9	(0.6)[e]	3.1	(0.8)[f]	6.5	(0.6)[j]	25.2	(1.9)
Israel	5.2	(0.3)[c,d]	10.7	(0.5)	—	—[f,g,h,i]	5.3	(0.3)	17.6	(0.6)
Italy	11.0	(0.9)	9.9	(0.5)[e]	1.7	(0.4)[f]	1.3	(0.2)[j]	18.1	(1.1)
Netherlands	15.9	(1.1)	17.9	(1.0)[e]	4.7	(1.1)[f]	8.9	(0.9)[j]	31.7	(2.0)
Spain	9.9	(1.1)	10.6	(0.5)[e]	2.3	(0.8)[f]	3.6	(0.4)[j]	19.4	(1.4)
Ukraine	10.9	(0.8)[c]	15.8	(0.8)[e]	8.7	(1.1)[g,i]	15.0	(1.3)	36.1	(1.5)
V. WHO Region: Western Pacific Regional Office (WPRO)										
PRC[k]	4.8	(0.7)	3.6	(0.4)	4.3	(0.9)[g,i]	4.9	(0.7)	13.2	(1.3)
Japan	6.9	(0.6)[c]	7.6	(0.5)	2.8	(1.0)[g,h,i]	4.8	(0.5)	18.0	(1.1)
New Zealand	24.6	(0.7)[c]	20.4	(0.5)	—	—[f,g,h,i]	12.4	(0.4)	39.3	(0.9)

[a] Anxiety disorders include agoraphobia, adult separation anxiety disorder, generalized anxiety disorder, panic disorder, posttraumatic stress disorder, social phobia, and specific phobia. Mood disorders include bipolar disorders, dysthymia, and major depressive disorder. Impulse-control disorders include intermittent explosive disorder, and reported persistence in the past 12 months of symptoms of three child–adolescent disorders (attention-deficit/hyperactivity disorder, conduct disorder, and oppositional-defiant disorder). Substance disorders include alcohol or drug abuse with or without dependence. In the case of substance dependence, respondents who met full criteria at some time in their life and who continue to have any symptoms are considered to have 12-month dependence even if they currently do not meet the full criteria for the disorder. Organic exclusions were made as specified in the *Diagnostic and Statistical Manual of Mental Health Disorders (DSM)*, 4th edition.

[b] Impulse disorders are restricted to age ≤39 (China, Ukraine, Nigeria) or to age ≤44 (all other countries).

[c] Adult separation anxiety disorder was not assessed.

[d] Specific phobia was not assessed.

[e] Bipolar disorders were not assessed.

[f] Intermittent explosive disorder was not assessed.

[g] Attention-deficit/hyperactivity disorder was not assessed.

[h] Conduct disorder was not assessed.

[i] Oppositional-defiant disorder was not assessed.

[j] Only alcohol abuse with or without dependence was assessed. No assessment was made of other drug abuse with or without dependence.

[k] People's Republic of China.

Reproduced with permission from Kessler RC, Aguilar-Gaxiola S, Alonso J, et al. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiol Psichiatr Soc*. 2009;18(1):23–33. doi:10.1017/s1121189x00001421.

Table 60–3.
Twelve-Month Prevalence Estimates of DSM-IV/CIDI Disorders in the WMH Surveys[a,b]

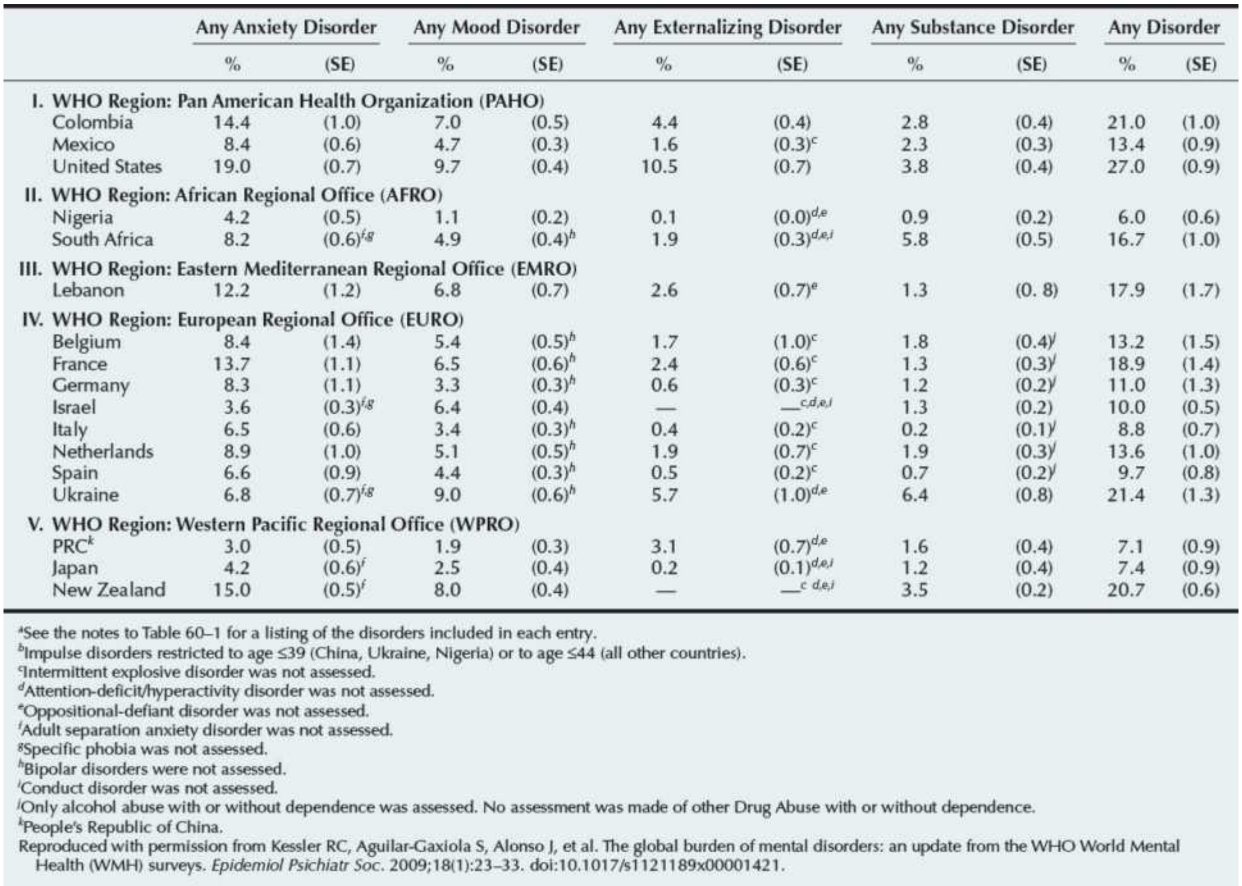

	Any Anxiety Disorder		Any Mood Disorder		Any Externalizing Disorder		Any Substance Disorder		Any Disorder	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)
I. WHO Region: Pan American Health Organization (PAHO)										
Colombia	14.4	(1.0)	7.0	(0.5)	4.4	(0.4)	2.8	(0.4)	21.0	(1.0)
Mexico	8.4	(0.6)	4.7	(0.3)	1.6	(0.3)[c]	2.3	(0.3)	13.4	(0.9)
United States	19.0	(0.7)	9.7	(0.4)	10.5	(0.7)	3.8	(0.4)	27.0	(0.9)
II. WHO Region: African Regional Office (AFRO)										
Nigeria	4.2	(0.5)	1.1	(0.2)	0.1	(0.0)[d,e]	0.9	(0.2)	6.0	(0.6)
South Africa	8.2	(0.6)[f,g]	4.9	(0.4)[h]	1.9	(0.3)[d,e,i]	5.8	(0.5)	16.7	(1.0)
III. WHO Region: Eastern Mediterranean Regional Office (EMRO)										
Lebanon	12.2	(1.2)	6.8	(0.7)	2.6	(0.7)[e]	1.3	(0. 8)	17.9	(1.7)
IV. WHO Region: European Regional Office (EURO)										
Belgium	8.4	(1.4)	5.4	(0.5)[h]	1.7	(1.0)[c]	1.8	(0.4)[j]	13.2	(1.5)
France	13.7	(1.1)	6.5	(0.6)[h]	2.4	(0.6)[c]	1.3	(0.3)[j]	18.9	(1.4)
Germany	8.3	(1.1)	3.3	(0.3)[h]	0.6	(0.3)[c]	1.2	(0.2)[j]	11.0	(1.3)
Israel	3.6	(0.3)[f,g]	6.4	(0.4)	—	—[c,d,e,i]	1.3	(0.2)	10.0	(0.5)
Italy	6.5	(0.6)	3.4	(0.3)[h]	0.4	(0.2)[c]	0.2	(0.1)[j]	8.8	(0.7)
Netherlands	8.9	(1.0)	5.1	(0.5)[h]	1.9	(0.7)[c]	1.9	(0.3)[j]	13.6	(1.0)
Spain	6.6	(0.9)	4.4	(0.3)[h]	0.5	(0.2)[c]	0.7	(0.2)[j]	9.7	(0.8)
Ukraine	6.8	(0.7)[f,g]	9.0	(0.6)[h]	5.7	(1.0)[d,e]	6.4	(0.8)	21.4	(1.3)
V. WHO Region: Western Pacific Regional Office (WPRO)										
PRC[k]	3.0	(0.5)	1.9	(0.3)	3.1	(0.7)[d,e]	1.6	(0.4)	7.1	(0.9)
Japan	4.2	(0.6)[f]	2.5	(0.4)	0.2	(0.1)[d,e,i]	1.2	(0.4)	7.4	(0.9)
New Zealand	15.0	(0.5)[f]	8.0	(0.4)	—	—[c d,e,i]	3.5	(0.2)	20.7	(0.6)

[a] See the notes to Table 60–1 for a listing of the disorders included in each entry.
[b] Impulse disorders restricted to age ≤39 (China, Ukraine, Nigeria) or to age ≤44 (all other countries).
[c] Intermittent explosive disorder was not assessed.
[d] Attention-deficit/hyperactivity disorder was not assessed.
[e] Oppositional-defiant disorder was not assessed.
[f] Adult separation anxiety disorder was not assessed.
[g] Specific phobia was not assessed.
[h] Bipolar disorders were not assessed.
[i] Conduct disorder was not assessed.
[j] Only alcohol abuse with or without dependence was assessed. No assessment was made of other Drug Abuse with or without dependence.
[k] People's Republic of China.

Reproduced with permission from Kessler RC, Aguilar-Gaxiola S, Alonso J, et al. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiol Psichiatr Soc*. 2009;18(1):23–33. doi:10.1017/s1121189x00001421.

In the same series of studies, the 12-month prevalence showed similar variations. These findings illustrated regional and national variations. Once again, Nigerian respondents reported only 6% prevalence, with similar figures from Japan and China, but the rates were 4.5 times higher in the United States (see Table 60–3).

As shown in Table 60–4, when the association between 12-month severity and days out of role were compared, again, variations within regions and across nations were substantial. Once again, Nigeria and China had similar figures, whereas the United States showed the highest levels.

When the severity of these disorders was compared across nations and regions, similar variations emerged. China, Japan, Nigeria, and this time Italy showed lower levels of serious severity and higher levels of mild severity (see Table 60–5).

Using the disability scale for common chronic physical and mental health disorders, interesting differences emerged between developed and developing countries (see Table 60–6). Cancer, asthma, hypertension, and diabetes showed higher proportions of severe disability in developing countries, whereas neck and backaches, chronic pain, and headaches were more seriously disabling in developed countries. Bipolar disorders, depression, generalized anxiety disorders, panic disorders, and posttraumatic disorder–related disability was elevated in developed countries. This variation raises interesting questions about whether these are genuinely higher levels or are rather more commonly diagnosed.

Primary Care

Among people seen by primary health care professionals, more than 20% have one or more current mental disorders. In a study carried out by the WHO at 14 sites in Africa, Asia, the Americas, and Europe, the average current prevalence of any mental disorder was 24%, without consistent differences between low- and high-income countries. Not surprisingly, the most common diagnoses were those of depression (average 10.4%) and generalized anxiety disorder (average 7.9%). Female rates were 1.89 times higher than male rates for depression, but male rates were higher for alcohol-related disorders, so there was no sex difference in the proportion of people having at least one mental disorder. Physical ill-health and educational disadvantage were both significantly associated with a diagnosis of mental disorder.

Disability-Adjusted Life Years (DALYs)

In addition to measures of prevalence, understanding the impact of mental health globally must also take into account the morbidity associated with disabling mental health conditions. To quantify the burden of various diseases and injuries, the WHO, in collaboration with the Harvard School of Public Health and the World Bank, introduced in 1993 a new measure, the disability-adjusted life year (DALY). DALYs for a given disease or injury are the sum of the years of life lost due to premature mortality plus the years lost due to disability for incident cases of that disease or injury in the general population.

Table 60–4.
Association between Severity of 12-Month DSM-IV/CIDI Disorders and Days Out of Role in the WMH Surveys

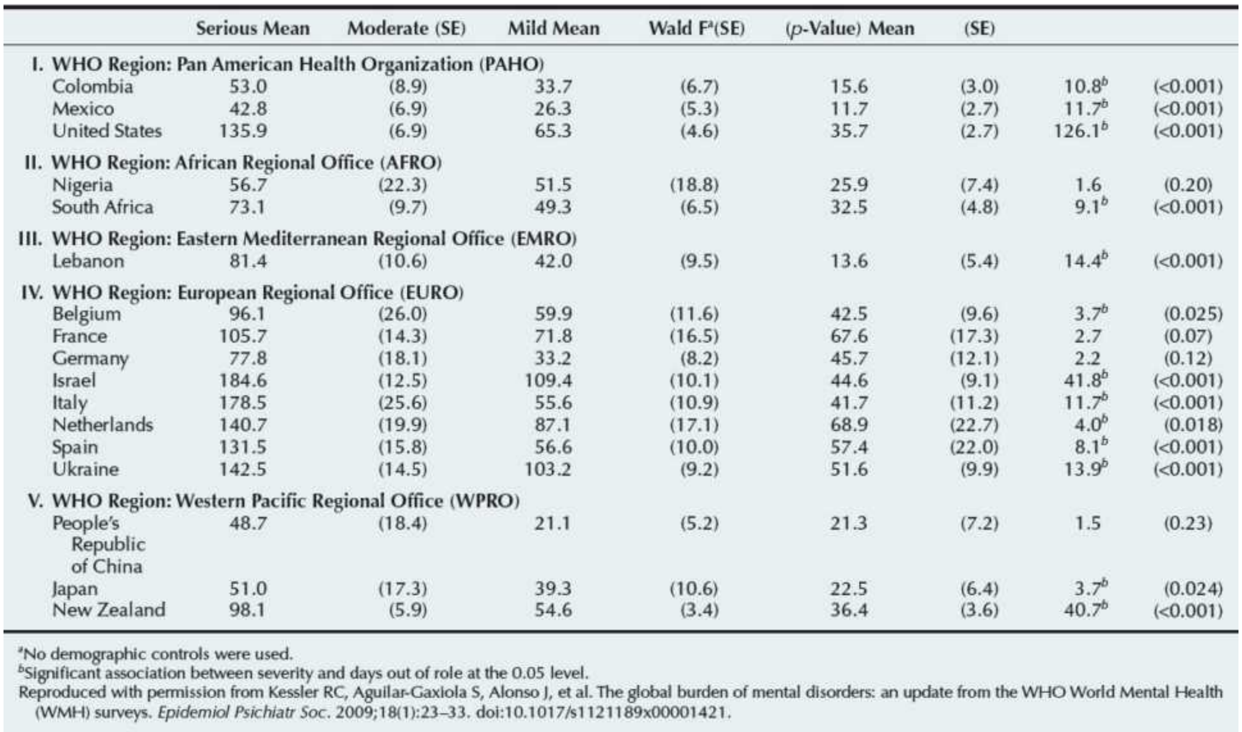

	Serious Mean	Moderate (SE)	Mild Mean	Wald F^a(SE)	(*p*-Value) Mean	(SE)		
I. WHO Region: Pan American Health Organization (PAHO)								
Colombia	53.0	(8.9)	33.7	(6.7)	15.6	(3.0)	10.8^b	(<0.001)
Mexico	42.8	(6.9)	26.3	(5.3)	11.7	(2.7)	11.7^b	(<0.001)
United States	135.9	(6.9)	65.3	(4.6)	35.7	(2.7)	126.1^b	(<0.001)
II. WHO Region: African Regional Office (AFRO)								
Nigeria	56.7	(22.3)	51.5	(18.8)	25.9	(7.4)	1.6	(0.20)
South Africa	73.1	(9.7)	49.3	(6.5)	32.5	(4.8)	9.1^b	(<0.001)
III. WHO Region: Eastern Mediterranean Regional Office (EMRO)								
Lebanon	81.4	(10.6)	42.0	(9.5)	13.6	(5.4)	14.4^b	(<0.001)
IV. WHO Region: European Regional Office (EURO)								
Belgium	96.1	(26.0)	59.9	(11.6)	42.5	(9.6)	3.7^b	(0.025)
France	105.7	(14.3)	71.8	(16.5)	67.6	(17.3)	2.7	(0.07)
Germany	77.8	(18.1)	33.2	(8.2)	45.7	(12.1)	2.2	(0.12)
Israel	184.6	(12.5)	109.4	(10.1)	44.6	(9.1)	41.8^b	(<0.001)
Italy	178.5	(25.6)	55.6	(10.9)	41.7	(11.2)	11.7^b	(<0.001)
Netherlands	140.7	(19.9)	87.1	(17.1)	68.9	(22.7)	4.0^b	(0.018)
Spain	131.5	(15.8)	56.6	(10.0)	57.4	(22.0)	8.1^b	(<0.001)
Ukraine	142.5	(14.5)	103.2	(9.2)	51.6	(9.9)	13.9^b	(<0.001)
V. WHO Region: Western Pacific Regional Office (WPRO)								
People's Republic of China	48.7	(18.4)	21.1	(5.2)	21.3	(7.2)	1.5	(0.23)
Japan	51.0	(17.3)	39.3	(10.6)	22.5	(6.4)	3.7^b	(0.024)
New Zealand	98.1	(5.9)	54.6	(3.4)	36.4	(3.6)	40.7^b	(<0.001)

[a] No demographic controls were used.

[b] Significant association between severity and days out of role at the 0.05 level.

Reproduced with permission from Kessler RC, Aguilar-Gaxiola S, Alonso J, et al. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiol Psichiatr Soc*. 2009;18(1):23–33. doi:10.1017/s1121189x00001421.

In the original estimates for the year 1990, mental and neurologic disorders accounted for 10.5% of the total DALYs lost due to all diseases and injuries. The burden of disease related to mental health disorders and related temporal change from 1990 to 2017 is shown by way of years lived with a disability (YLDs) in Table 60–7. When a comparison is made between years lived with disability due to depression between 2007 and 2017, it appears that the years have increased by 12.6% due to major depressive disorders and dysthymia by over 20%.

The WHO estimates that one in four families worldwide has at least one member suffering from a mental disorder. The objective and subjective burden related to caring for people with severe mental disorders (in terms of disruption of family relationships; constraints in social, leisure, and work activities; financial difficulties; negative effect on physical health; feelings of loss, depression, and embarrassment in social situations; and the stress of coping with disturbing behaviors) have been reported to be substantial and significantly higher than that related to caring for people with long-term physical diseases such as diabetes, heart, kidney, or lung diseases. Cross-cultural differences have been reported in some dimensions of family burden: for example, relatives in the Mediterranean area compared with those in northern European countries were found to be more resigned and to use spiritual help more frequently as a coping strategy. They reported a greater level of reduction in social interests and lower levels of support from their social network.

Children and Adolescents

The mental health and well-being of young people are important to consider, given they represent a substantial proportion of the world's residents, with one in six people aged between 10 and 19. More than half of all mental health conditions are thought to start by the age of 14. However, most are not diagnosed or treated. The consequences of not treating mental health conditions during this period have profound ramifications as the impact extends into adulthood and limits functional abilities.

The prevalence of mental disorders among children and adolescents is difficult to evaluate because it is problematic to draw a reliable boundary between psychopathological phenomena and some manifestations that are part of normal development. Population surveys may help, but access to schools and clinics is extremely variable. Hence the accuracy of figures can be challenged. Many studies have used checklists compiled by parents and teachers, which are useful in detecting cases needing special attention but may not represent an appropriate basis for prevalence estimates.

Ten percent of all children and adolescents are thought to experience a mental disorder. In adolescents (aged 10 to 19), mental health conditions are thought to account for 16% of the GBD in that group. A GBD study among children between the ages of 5 and 14 found that mental disorders were among the leading causes of both DALYs in Europe, North America, and South America. They were in the top five causes in South East Asia, the Eastern Mediterranean, and the Western Pacific. Infectious diseases held a greater burden in Africa, and as such, mental health disorders were not among the top five causes of DALYs. It is likely that as low- and middle-income countries undergo epidemiologic transitions away from communicable disease, the share of disease burden related to mental health disorders will increase among young people.

Table 60–5.
Prevalence of 12-Month DSM-IV/CIDI Disorders by Severity in the WMH Surveys[a]

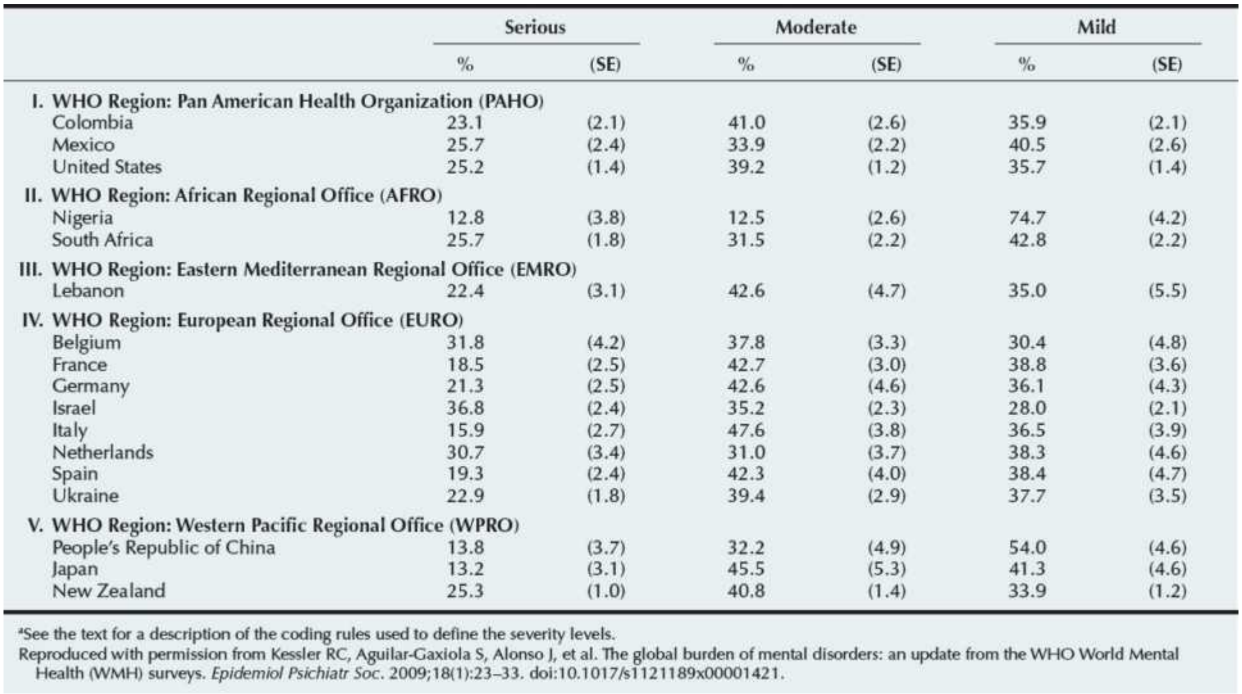

	Serious		Moderate		Mild	
	%	(SE)	%	(SE)	%	(SE)
I. WHO Region: Pan American Health Organization (PAHO)						
Colombia	23.1	(2.1)	41.0	(2.6)	35.9	(2.1)
Mexico	25.7	(2.4)	33.9	(2.2)	40.5	(2.6)
United States	25.2	(1.4)	39.2	(1.2)	35.7	(1.4)
II. WHO Region: African Regional Office (AFRO)						
Nigeria	12.8	(3.8)	12.5	(2.6)	74.7	(4.2)
South Africa	25.7	(1.8)	31.5	(2.2)	42.8	(2.2)
III. WHO Region: Eastern Mediterranean Regional Office (EMRO)						
Lebanon	22.4	(3.1)	42.6	(4.7)	35.0	(5.5)
IV. WHO Region: European Regional Office (EURO)						
Belgium	31.8	(4.2)	37.8	(3.3)	30.4	(4.8)
France	18.5	(2.5)	42.7	(3.0)	38.8	(3.6)
Germany	21.3	(2.5)	42.6	(4.6)	36.1	(4.3)
Israel	36.8	(2.4)	35.2	(2.3)	28.0	(2.1)
Italy	15.9	(2.7)	47.6	(3.8)	36.5	(3.9)
Netherlands	30.7	(3.4)	31.0	(3.7)	38.3	(4.6)
Spain	19.3	(2.4)	42.3	(4.0)	38.4	(4.7)
Ukraine	22.9	(1.8)	39.4	(2.9)	37.7	(3.5)
V. WHO Region: Western Pacific Regional Office (WPRO)						
People's Republic of China	13.8	(3.7)	32.2	(4.9)	54.0	(4.6)
Japan	13.2	(3.1)	45.5	(5.3)	41.3	(4.6)
New Zealand	25.3	(1.0)	40.8	(1.4)	33.9	(1.2)

[a]See the text for a description of the coding rules used to define the severity levels.

Reproduced with permission from Kessler RC, Aguilar-Gaxiola S, Alonso J, et al. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiol Psichiatr Soc*. 2009;18(1):23–33. doi:10.1017/s1121189x00001421.

Table 60–6.
Disorder-Specific Global Sheehan Disability Scale Ratings for Commonly Occurring Mental and Chronic Physical Disorders in Developed and Developing WMH Countries

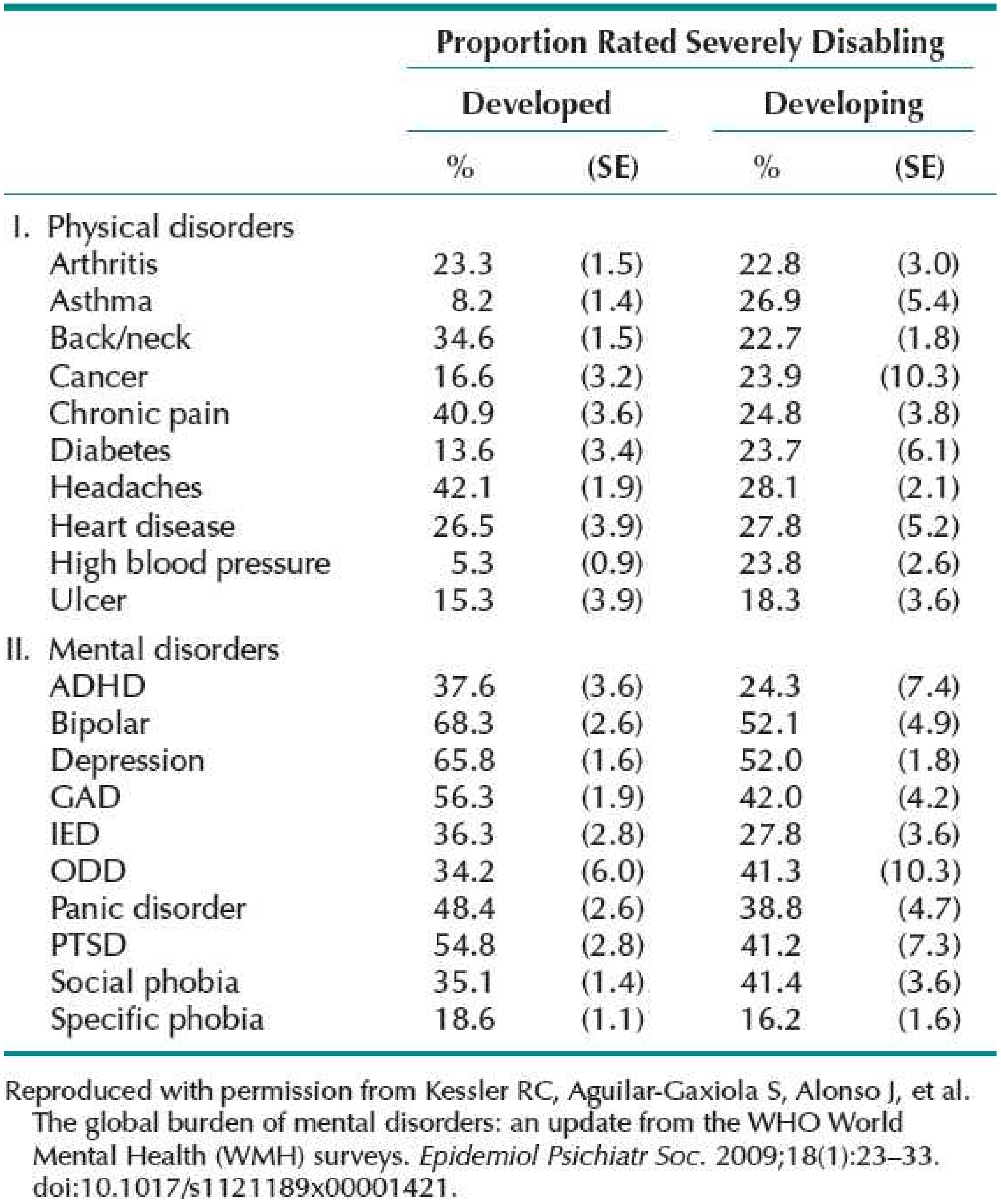

	Proportion Rated Severely Disabling			
	Developed		Developing	
	%	(SE)	%	(SE)
I. Physical disorders				
Arthritis	23.3	(1.5)	22.8	(3.0)
Asthma	8.2	(1.4)	26.9	(5.4)
Back/neck	34.6	(1.5)	22.7	(1.8)
Cancer	16.6	(3.2)	23.9	(10.3)
Chronic pain	40.9	(3.6)	24.8	(3.8)
Diabetes	13.6	(3.4)	23.7	(6.1)
Headaches	42.1	(1.9)	28.1	(2.1)
Heart disease	26.5	(3.9)	27.8	(5.2)
High blood pressure	5.3	(0.9)	23.8	(2.6)
Ulcer	15.3	(3.9)	18.3	(3.6)
II. Mental disorders				
ADHD	37.6	(3.6)	24.3	(7.4)
Bipolar	68.3	(2.6)	52.1	(4.9)
Depression	65.8	(1.6)	52.0	(1.8)
GAD	56.3	(1.9)	42.0	(4.2)
IED	36.3	(2.8)	27.8	(3.6)
ODD	34.2	(6.0)	41.3	(10.3)
Panic disorder	48.4	(2.6)	38.8	(4.7)
PTSD	54.8	(2.8)	41.2	(7.3)
Social phobia	35.1	(1.4)	41.4	(3.6)
Specific phobia	18.6	(1.1)	16.2	(1.6)

Reproduced with permission from Kessler RC, Aguilar-Gaxiola S, Alonso J, et al. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiol Psichiatr Soc*. 2009;18(1):23–33. doi:10.1017/s1121189x00001421.

Suicide

According to WHO estimates, around 703,000 people died from suicide worldwide in the year 2019. This accounted for 1.3% of total deaths that year. The global age-standardized suicide rate was 9 per 100,000. Notably, there is a wide range between countries—less than 2 to over 80.0 deaths per 100,000. Globally, the age-standardized mortality rate was higher among men (12.6 per 100,000) compared to women (5.4 per 100,000). The suicide rate, therefore, was 2.3 times in men compared to women—was notably over 3 in high-income countries, 2.8 in upper-middle–income countries, 1.8 in upper-middle–income countries, and 2.9 in low-income countries. By and large, younger people and older adults had higher rates than middle-aged adults.

Around 77% of the suicides in 2019 occurred in low-and-middle-income countries. While this was the case and related to large overall populations, high-income countries had a higher age-standardized rate (10.9 per 100,000) compared to upper-middle–income countries (7.3 per 100,000), lower-middle–income countries (10.1 per 100,000) and low-income countries (9.9 per 100,000). Suicide rates varied by region (global average of 9.0 per 100,000): Europe, Africa, and South East Asia had higher rates at 10.5, 11.2, and 10.2 per 100,000, respectively. The Eastern Mediterranean region had the lowest rate at 6.4 per 100,000. In adolescents, the number who died by suicide in those countries is greater at 88%. Among young people aged between 15 and 29, suicide was the fourth leading cause of death.

Table 60–7.
YLDs for 2017, from the Global Burden of Disease Study 2017

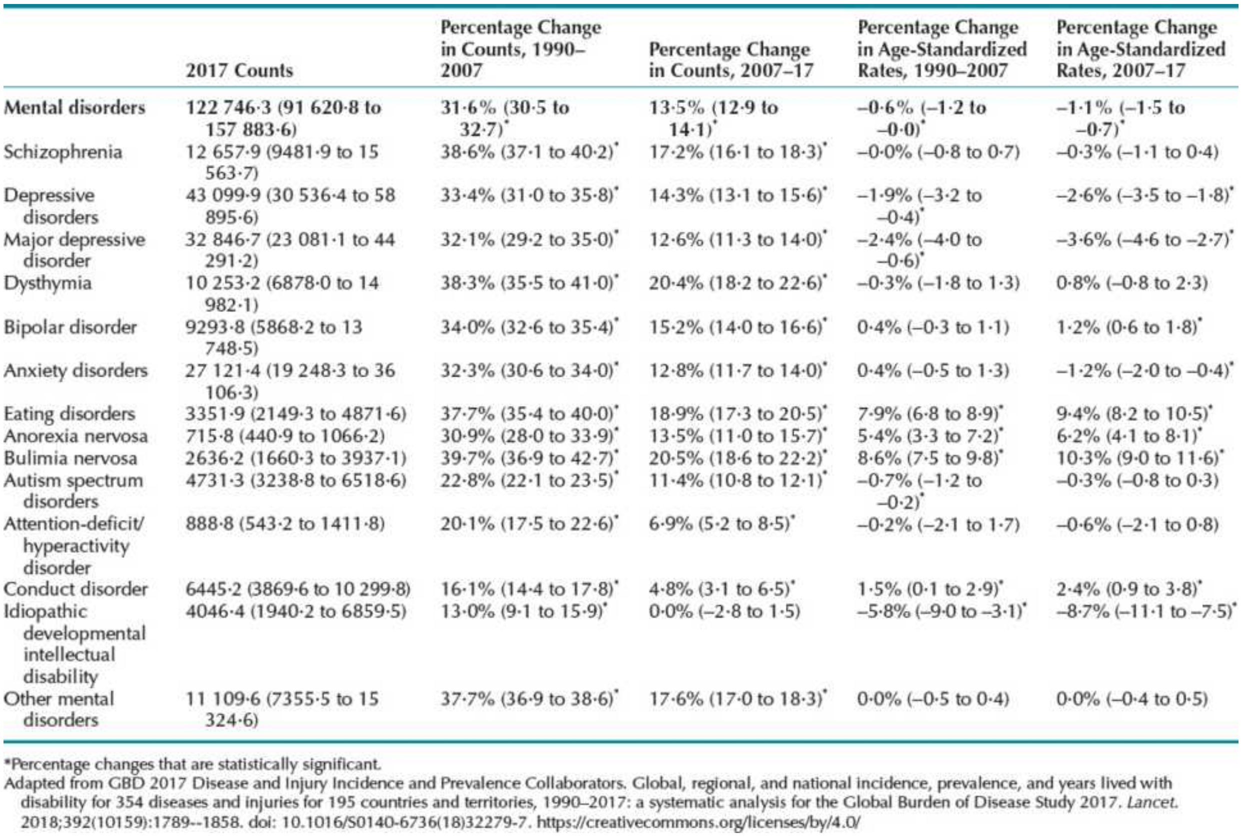

	2017 Counts	Percentage Change in Counts, 1990–2007	Percentage Change in Counts, 2007–17	Percentage Change in Age-Standardized Rates, 1990–2007	Percentage Change in Age-Standardized Rates, 2007–17
Mental disorders	**122 746·3 (91 620·8 to 157 883·6)**	**31·6% (30·5 to 32·7)***	**13·5% (12·9 to 14·1)***	**–0·6% (–1·2 to –0·0)***	**–1·1% (–1·5 to –0·7)***
Schizophrenia	12 657·9 (9481·9 to 15 563·7)	38·6% (37·1 to 40·2)*	17·2% (16·1 to 18·3)*	–0·0% (–0·8 to 0·7)	–0·3% (–1·1 to 0·4)
Depressive disorders	43 099·9 (30 536·4 to 58 895·6)	33·4% (31·0 to 35·8)*	14·3% (13·1 to 15·6)*	–1·9% (–3·2 to –0·4)*	–2·6% (–3·5 to –1·8)*
Major depressive disorder	32 846·7 (23 081·1 to 44 291·2)	32·1% (29·2 to 35·0)*	12·6% (11·3 to 14·0)*	–2·4% (–4·0 to –0·6)*	–3·6% (–4·6 to –2·7)*
Dysthymia	10 253·2 (6878·0 to 14 982·1)	38·3% (35·5 to 41·0)*	20·4% (18·2 to 22·6)*	–0·3% (–1·8 to 1·3)	0·8% (–0·8 to 2·3)
Bipolar disorder	9293·8 (5868·2 to 13 748·5)	34·0% (32·6 to 35·4)*	15·2% (14·0 to 16·6)*	0·4% (–0·3 to 1·1)	1·2% (0·6 to 1·8)*
Anxiety disorders	27 121·4 (19 248·3 to 36 106·3)	32·3% (30·6 to 34·0)*	12·8% (11·7 to 14·0)*	0·4% (–0·5 to 1·3)	–1·2% (–2·0 to –0·4)*
Eating disorders	3351·9 (2149·3 to 4871·6)	37·7% (35·4 to 40·0)*	18·9% (17·3 to 20·5)*	7·9% (6·8 to 8·9)*	9·4% (8·2 to 10·5)*
Anorexia nervosa	715·8 (440·9 to 1066·2)	30·9% (28·0 to 33·9)*	13·5% (11·0 to 15·7)*	5·4% (3·3 to 7·2)*	6·2% (4·1 to 8·1)*
Bulimia nervosa	2636·2 (1660·3 to 3937·1)	39·7% (36·9 to 42·7)*	20·5% (18·6 to 22·2)*	8·6% (7·5 to 9·8)*	10·3% (9·0 to 11·6)*
Autism spectrum disorders	4731·3 (3238·8 to 6518·6)	22·8% (22·1 to 23·5)*	11·4% (10·8 to 12·1)*	–0·7% (–1·2 to –0·2)*	–0·3% (–0·8 to 0·3)
Attention-deficit/ hyperactivity disorder	888·8 (543·2 to 1411·8)	20·1% (17·5 to 22·6)*	6·9% (5·2 to 8·5)*	–0·2% (–2·1 to 1·7)	–0·6% (–2·1 to 0·8)
Conduct disorder	6445·2 (3869·6 to 10 299·8)	16·1% (14·4 to 17·8)*	4·8% (3·1 to 6·5)*	1·5% (0·1 to 2·9)*	2·4% (0·9 to 3·8)*
Idiopathic developmental intellectual disability	4046·4 (1940·2 to 6859·5)	13·0% (9·1 to 15·9)*	0·0% (–2·8 to 1·5)	–5·8% (–9·0 to –3·1)*	–8·7% (–11·1 to –7·5)*
Other mental disorders	11 109·6 (7355·5 to 15 324·6)	37·7% (36·9 to 38·6)*	17·6% (17·0 to 18·3)*	0·0% (–0·5 to 0·4)	0·0% (–0·4 to 0·5)

*Percentage changes that are statistically significant.

Adapted from GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* 2018;392(10159):1789--1858. doi: 10.1016/S0140-6736(18)32279-7. https://creativecommons.org/licenses/by/4.0/

The age-standardized mortality rate decreased by 36% between 2000 and 2019. The rate decreased most, by 49%, in the Western Pacific Region. Although the overall decrease between 1990 and 2019 is promising and substantial, if the current trajectory continues, most countries will not meet the Sustainable Development Goals target of reducing suicide-related mortality by a third from 2015 to 2030.

Globally, the trends and varied patterns of suicidality reflect an interplay of different factors. These include socio-cultural, religious, and sociodemographic inter- and intra-country differences. In addition, economic circumstances, unemployment, exposure to violence, use of alcohol and drugs, patterns of mental disorder, and access to means of suicide (e.g., access to firearms) were all found to play a key role.

TREATMENT GAP AND PROJECTED POPULATION-LEVEL TREATMENT EFFECTIVENESS WORLDWIDE

Lifetime contact for treatment delay in various psychiatric conditions is shown in Table 60–8. Interestingly, common mental disorders showed longer delays. This is further confirmed in the 12-month service use by sectors (see Table 60–9).

The efficacy of pharmacologic and psychosocial treatments for mood, anxiety, psychotic, and substance-related disorders has been convincingly proved by clinical trials carried out in low- and middle-income as well as in high-income countries. However, the treatment gap is substantial for all mental disorders worldwide, in particular, in low-income countries.

In the World Mental Health Surveys, carried out in representative community samples in 17 countries in Africa, the Americas, Asia, Europe, and the Middle East, the proportion of lifetime cases of anxiety disorders making treatment contact in the year of disorder onset ranged from 0.8% in Nigeria to 36.4% in Israel. The proportion of cases making treatment contact by 50 years ranged from 15.2% in Nigeria to 95.0% in Germany. The median delay to treatment among cases eventually making contact ranged from 3 years in Israel to 30 years in Mexico (Table 60–10).

For Schizophrenia, a systematic review by Jääskeläinen and colleagues that was based upon 50 studies looking at outcomes found that only 13.5% of people with Schizophrenia (median) met clinical and social recovery criteria.

The proportion of lifetime cases of mood disorders making treatment contact in the year of disorder onset ranged from 6% in Nigeria and China to 52.1% in the Netherlands. The proportion of cases making treatment contact by 50 years ranged from 7.9% in China to 98.6% in France. The median delay to treatment among cases eventually making contact ranged from 1 year in Belgium, The Netherlands, Spain, and Japan to 14 years in Mexico (Table 60–11).

The proportion of lifetime cases of substance use disorders making treatment contact in the year of disorder onset ranged from 0.9% in Mexico to 18.6% in Spain. The proportion of cases making treatment contact by 50 years ranged from 19.8% in Nigeria to 86.1% in Germany. The median delay to treatment among cases eventually making contact ranged from 6 years in Spain to 18 years in Belgium (Table 60–12).

Table 60–8.

Lifetime Treatment Contacts for Mental Disorders among Survey Participants, by Contacts before or after Disorder Onset[a]

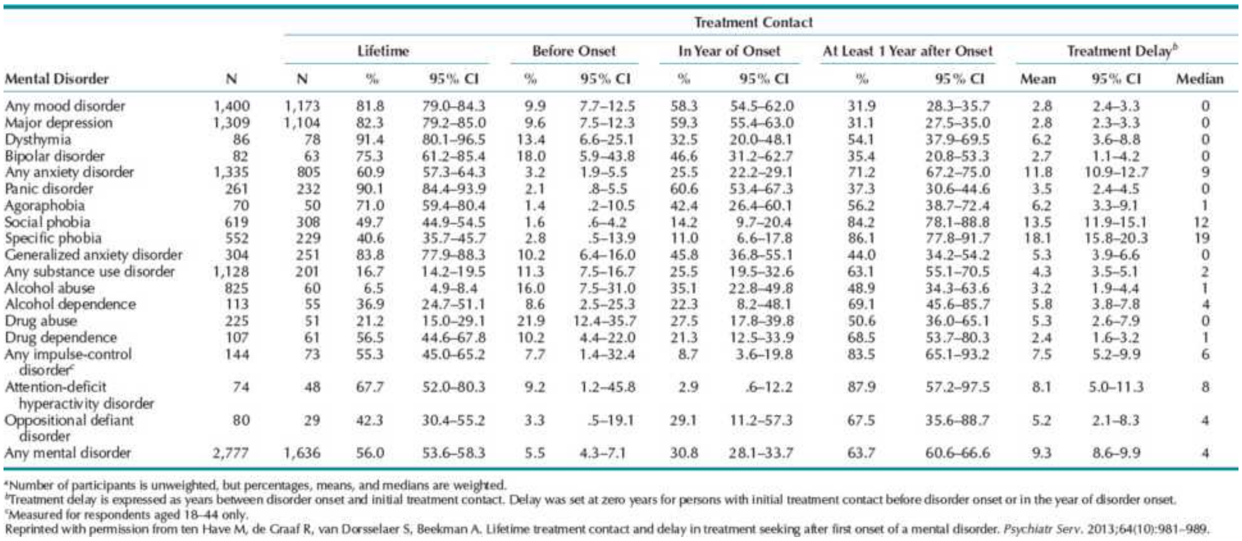

		Treatment Contact												
		Lifetime			Before Onset		In Year of Onset		At Least 1 Year after Onset		Treatment Delay[b]		
Mental Disorder	N	N	%	95% CI	%	95% CI	%	95% CI	%	95% CI	Mean	95% CI	Median
Any mood disorder	1,400	1,173	81.8	79.0–84.3	9.9	7.7–12.5	58.3	54.5–62.0	31.9	28.3–35.7	2.8	2.4–3.3	0
Major depression	1,309	1,104	82.3	79.2–85.0	9.6	7.5–12.3	59.3	55.4–63.0	31.1	27.5–35.0	2.8	2.3–3.3	0
Dysthymia	86	78	91.4	80.1–96.5	13.4	6.6–25.1	32.5	20.0–48.1	54.1	37.9–69.5	6.2	3.6–8.8	0
Bipolar disorder	82	63	75.3	61.2–85.4	18.0	5.9–43.8	46.6	31.2–62.7	35.4	20.8–53.3	2.7	1.1–4.2	0
Any anxiety disorder	1,335	805	60.9	57.3–64.3	3.2	1.9–5.5	25.5	22.2–29.1	71.2	67.2–75.0	11.8	10.9–12.7	9
Panic disorder	261	232	90.1	84.4–93.9	2.1	.8–5.5	60.6	53.4–67.3	37.3	30.6–44.6	3.5	2.4–4.5	0
Agoraphobia	70	50	71.0	59.4–80.4	1.4	.2–10.5	42.4	26.4–60.1	56.2	38.7–72.4	6.2	3.3–9.1	1
Social phobia	619	308	49.7	44.9–54.5	1.6	.6–4.2	14.2	9.7–20.4	84.2	78.1–88.8	13.5	11.9–15.1	12
Specific phobia	552	229	40.6	35.7–45.7	2.8	.5–13.9	11.0	6.6–17.8	86.1	77.8–91.7	18.1	15.8–20.3	19
Generalized anxiety disorder	304	251	83.8	77.9–88.3	10.2	6.4–16.0	45.8	36.8–55.1	44.0	34.2–54.2	5.3	3.9–6.6	0
Any substance use disorder	1,128	201	16.7	14.2–19.5	11.3	7.5–16.7	25.5	19.5–32.6	63.1	55.1–70.5	4.3	3.5–5.1	2
Alcohol abuse	825	60	6.5	4.9–8.4	16.0	7.5–31.0	35.1	22.8–49.8	48.9	34.3–63.6	3.2	1.9–4.4	1
Alcohol dependence	113	55	36.9	24.7–51.1	8.6	2.5–25.3	22.3	8.2–48.1	69.1	45.6–85.7	5.8	3.8–7.8	4
Drug abuse	225	51	21.2	15.0–29.1	21.9	12.4–35.7	27.5	17.8–39.8	50.6	36.0–65.1	5.3	2.6–7.9	0
Drug dependence	107	61	56.5	44.6–67.8	10.2	4.4–22.0	21.3	12.5–33.9	68.5	53.7–80.3	2.4	1.6–3.2	1
Any impulse-control disorder[c]	144	73	55.3	45.0–65.2	7.7	1.4–32.4	8.7	3.6–19.8	83.5	65.1–93.2	7.5	5.2–9.9	6
Attention-deficit hyperactivity disorder	74	48	67.7	52.0–80.3	9.2	1.2–45.8	2.9	.6–12.2	87.9	57.2–97.5	8.1	5.0–11.3	8
Oppositional defiant disorder	80	29	42.3	30.4–55.2	3.3	.5–19.1	29.1	11.2–57.3	67.5	35.6–88.7	5.2	2.1–8.3	4
Any mental disorder	2,777	1,636	56.0	53.6–58.3	5.5	4.3–7.1	30.8	28.1–33.7	63.7	60.6–66.6	9.3	8.6–9.9	4

[a] Number of participants is unweighted, but percentages, means, and medians are weighted.

[b] Treatment delay is expressed as years between disorder onset and initial treatment contact. Delay was set at zero years for persons with initial treatment contact before disorder onset or in the year of disorder onset.

[c] Measured for respondents aged 18–44 only.

Reprinted with permission from ten Have M, de Graaf R, van Dorsselaer S, Beekman A. Lifetime treatment contact and delay in treatment seeking after first onset of a mental disorder. *Psychiatr Serv*. 2013;64(10):981–989.

Table 60–9.

Twelve-Month Service Use by Sectors in the WMH Surveys

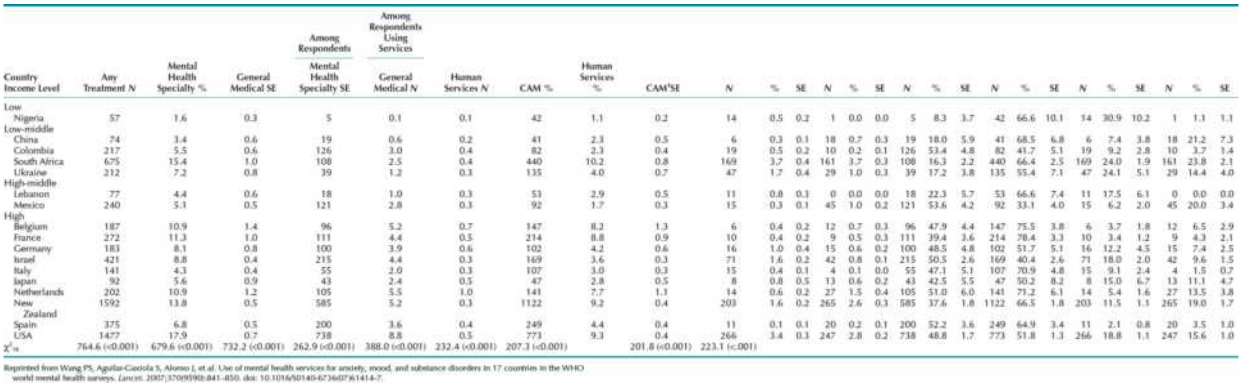

				Among Respondents	Among Respondents Using Services																						
Country Income Level	Any Treatment *N*	Mental Health Specialty %	General Medical SE	Mental Health Specialty SE	General Medical *N*	Human Services *N*	CAM %	Human Services %	CAM'SE	*N*	%	SE	*N*	%	SE	*N*	%	SE	*N*	%	SE	*N*	%	SE	*N*	%	SE
Low																											
Nigeria	57	1.6	0.3	5	0.1	0.1	42	1.1	0.2	14	0.5	0.2	1	0.0	0.0	5	8.3	3.7	42	66.6	10.1	14	30.9	10.2	1	1.1	1.1
Low-middle																											
China	74	3.4	0.6	19	0.6	0.2	41	2.3	0.5	6	0.3	0.1	18	0.7	0.3	19	18.0	5.9	41	68.5	6.8	6	7.4	3.8	18	21.2	7.3
Colombia	217	5.5	0.6	126	3.0	0.4	82	2.3	0.4	19	0.5	0.2	10	0.2	0.1	126	53.4	4.8	82	41.7	5.1	19	9.2	2.8	10	3.7	1.4
South Africa	675	15.4	1.0	108	2.5	0.4	440	10.2	0.8	169	3.7	0.4	161	3.7	0.3	108	16.3	2.2	440	66.4	2.5	169	24.0	1.9	161	23.8	2.1
Ukraine	212	7.2	0.8	39	1.2	0.3	135	4.0	0.7	47	1.7	0.4	29	1.0	0.3	39	17.2	3.8	135	55.4	7.1	47	24.1	5.1	29	14.4	4.0
High-middle																											
Lebanon	77	4.4	0.6	18	1.0	0.3	53	2.9	0.5	11	0.8	0.3	0	0.0	0.0	18	22.3	5.7	53	66.6	7.4	11	17.5	6.1	0	0.0	0.0
Mexico	240	5.1	0.5	121	2.8	0.3	92	1.7	0.3	15	0.3	0.1	45	1.0	0.2	121	53.6	4.2	92	33.1	4.0	15	6.2	2.0	45	20.0	3.4
High																											
Belgium	187	10.9	1.4	96	5.2	0.7	147	8.2	1.3	6	0.4	0.2	12	0.7	0.3	96	47.9	4.4	147	75.5	3.8	6	3.7	1.8	12	6.5	2.9
France	272	11.3	1.0	111	4.4	0.5	214	8.8	0.9	10	0.4	0.2	9	0.5	0.3	111	39.4	3.6	214	78.4	3.3	10	3.4	1.2	9	4.3	2.1
Germany	183	8.1	0.8	100	3.9	0.6	102	4.2	0.6	16	1.0	0.4	15	0.6	0.2	100	48.5	4.8	102	51.7	5.1	16	12.2	4.5	15	7.4	2.5
Israel	421	8.8	0.4	215	4.4	0.3	169	3.6	0.3	71	1.6	0.2	42	0.8	0.1	215	50.5	2.6	169	40.4	2.6	71	18.0	2.0	42	9.6	1.5
Italy	141	4.3	0.4	55	2.0	0.3	107	3.0	0.3	15	0.4	0.1	4	0.1	0.0	55	47.1	5.1	107	70.9	4.8	15	9.1	2.4	4	1.5	0.7
Japan	92	5.6	0.9	43	2.4	0.5	47	2.8	0.5	8	0.8	0.5	13	0.6	0.2	43	42.5	5.5	47	50.2	8.2	8	15.0	6.7	13	11.1	4.7
Netherlands	202	10.9	1.2	105	5.5	1.0	141	7.7	1.1	14	0.6	0.2	27	1.5	0.4	105	51.0	6.0	141	71.2	6.1	14	5.4	1.6	27	13.5	3.8
New Zealand	1592	13.8	0.5	585	5.2	0.3	1122	9.2	0.4	203	1.6	0.2	265	2.6	0.3	585	37.6	1.8	1122	66.5	1.8	203	11.5	1.1	265	19.0	1.7
Spain	375	6.8	0.5	200	3.6	0.4	249	4.4	0.4	11	0.1	0.1	20	0.2	0.1	200	52.2	3.6	249	64.9	3.4	11	2.1	0.8	20	3.5	1.0
USA	1477	17.9	0.7	738	8.8	0.5	773	9.3	0.4	266	3.4	0.3	247	2.8	0.2	738	48.8	1.7	773	51.8	1.3	266	18.8	1.1	247	15.6	1.0
χ^2_{16}	764.6 (<0.001)	679.6 (<0.001)	732.2 (<0.001)	262.9 (<0.001)	388.0 (<0.001)	232.4 (<0.001)	207.3 (<0.001)		201.8 (<0.001)	223.1 (<.001)																	

Reprinted from Wang PS, Aguilar-Gaxiola S, Alonso J, et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. *Lancet*. 2007;370(9590):841–850. doi: 10.1016/S0140-6736(07)61414-7.

Table 60–10.
Treatment Delay for Anxiety Disorders in the World Mental Health Surveys

	Making Treatment Contact in Year of Onset, % (SE)	Making Treatment Contact by Age 50 Years, % (SE)	Median Duration of Delay in Years (SE)
The Americas			
Colombia	2.9 (0.6)	41.6 (3.9)	26.0 (1.5)
Mexico	3.6 (1.1)	53.2 (18.2)	30.0 (5.1)
United States	11.3 (0.7)	87.0 (2.4)	23.0 (0.6)
Europe			
Belgium	19.8 (2.8)	84.5 (4.9)	16.0 (3.5)
France	16.1 (1.8)	93.3 (1.9)	18.0 (1.8)
Germany	13.7 (1.8)	95.0 (2.3)	23.0 (2.3)
Italy	17.1 (2.1)	87.3 (8.5)	28.0 (2.2)
Netherlands	28.0 (3.7)	91.1 (2.8)	10.0 (1.6)
Spain	23.2 (2.0)	86.6 (5.2)	17.0 (3.2)
Africa and Middle East			
Israel	36.4 (0.9)	90.7 (1.3)	3.0 (0.1)
Lebanon	3.2 (1.1)	37.3 (11.5)	28.0 (3.9)
Nigeria	0.8 (0.5)	15.2 (2.6)	16.0 (4.2)
Asia and the Pacific			
Japan	11.2 (2.4)	63.1 (6.2)	20.0 (2.4)
China	4.2 (2.0)	44.7 (7.2)	21.0 (3.1)
Oceania			
New Zealand	12.5 (.8)	84.2 (2.5)	21.0 (0.8)

SE, standard error.

Reprinted with permission from Wang PS, Angermeyer M, Borges G, et al. Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*. 2007;6(3):177–185.

Table 60–11.
Treatment Delay for Mood Disorders in the World Mental Health Surveys

	Making Treatment Contact in Year of Onset, % (SE)	Making Treatment Contact by Age 50 Years, % (SE)	Median Duration of Delay in Years (SE)
The Americas			
Colombia	18.7 (2.7)	66.6 (3.7)	9.0 (1.6)
Mexico	16.0 (2.2)	69.9 (8.5)	14.0 (3.1)
United States	35.4 (1.2)	94.8 (2.5)	4.0 (0.2)
Europe			
Belgium	47.8 (2.7)	93.7 (2.5)	1.0 (0.3)
France	42.7 (2.1)	98.6 (1.4)	3.0 (0.3)
Germany	40.4 (3.8)	89.1 (5.0)	2.0 (0.4)
Italy	28.8 (3.0)	63.5 (5.9)	2.0 (0.5)
Netherlands	52.1 (2.9)	96.9 (1.7)	1.0 (0.3)
Spain	48.5 (2.3)	96.4 (3.1)	1.0 (0.3)
Africa and Middle East			
Israel	31.9 (0.8)	92.7 (0.5)	6.0 (0.3)
Lebanon	12.3 (2.0)	49.2 (5.2)	6.0 (2.1)
Nigeria	6.0 (1.7)	33.3 (7.2)	6.0 (3.3)
Asia and the Pacific			
Japan	29.6 (4.0)	56.8 (7.3)	1.0 (0.7)
China	6.0 (2.2)	7.9 (2.6)	1.0 (2.0)
Oceania			
New Zealand	41.4 (1.3)	97.5 (1.0)	3.0 (0.2)

SE, standard error.

Reprinted with permission from Wang PS, Angermeyer M, Borges G, et al. Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization’s World Mental Health Survey Initiative. *World Psychiatry*. 2007;6(3):177–185.

Table 60–12.
Treatment Delay for Substance Use Disorders in the World Mental Health Surveys

	Making Treatment Contact in Year of Onset, % (SE)	Making Treatment Contact by Age 50 Years, % (SE)	Median Duration of Delay in Years (SE)
The Americas			
Colombia	3.6 (0.8)	23.1 (7.1)	11.0 (5.0)
Mexico	0.9 (0.5)	22.1 (4.8)	10.0 (3.3)
United States	10.0 (0.8)	75.5 (3.8)	13.0 (1.2)
Europe			
Belgium	12.8 (4.8)	61.2 (17.7)	18.0 (5.8)
France	15.7 (5.4)	66.5 (14.1)	13.0 (3.7)
Germany	13.2 (5.7)	86.1 (8.6)	9.0 (3.9)
Netherlands	15.5 (5.4)	66.6 (7.9)	9.0 (3.1)
Spain	18.6 (7.6)	40.1 (14.1)	6.0 (4.9)
Africa and Middle East			
Israel	2.0 (0.5)	48.0 (2.4)	12.0 (0.5)
Nigeria	2.8 (1.7)	19.8 (7.2)	8.0 (1.8)
Asia and the Pacific			
Japan	9.2 (5.1)	31.0 (7.8)	8.0 (4.6)
China	2.8 (1.8)	25.7 (9.0)	17.0 (3.7)
Oceania			
New Zealand	6.3 (0.8)	84.8 (15.4)	17.0 (1.3)

SE, standard error.

Reprinted with permission from Wang PS, Angermeyer M, Borges G, et al. Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*. 2007;6(3):177–185.

In the World Mental Health Surveys, failure and delays in treatment-seeking were generally greater in low-income countries, older cohorts, men, and cases with earlier ages of onset. The earlier treatment contact of people with mood disorders might be partly due to the fact that these disorders have been targeted by educational campaigns and primary care quality improvement programs in several countries.

The Disease Control Priorities Project Report estimated the cost of a mental health care package consisting of outpatient treatment of schizophrenia and bipolar disorder with first-generation antipsychotics or mood stabilizers and adjuvant psychosocial interventions and treatment of depression and panic disorder in primary care with generic selective serotonin reuptake inhibitors. The cost per capita of the package was projected to be US$ 3 to 4 in sub-Saharan Africa and South Asia and US$ 7 to 9 in Latin America and the Caribbean. The addition of brief interventions by primary care physicians for high-risk alcohol users was projected to cost a further US$ 0.04 per capita in sub-Saharan Africa and South Asia and up to US$ 0.36 per capita in Central Asia and Latin America. Treatment of common mental disorders was reported to be in the same range of cost-effectiveness as antiretroviral treatment for HIV/AIDS, secondary prevention of hypertension, or glycemic control for diabetes.

RESOURCES FOR MENTAL HEALTH CARE WORLDWIDE

According to the Mental Health Atlas 2017, 79% of the 139 countries that responded had a standalone mental health policy or plan (i.e., a document of the government or ministry of health specifying the goals for improving the mental health situation of the country, the priorities among those goals, and the main directions for attaining them). This ranged from 72% in the African region to 90% in the South East Asia region (see Table 60–13). This is a notable improvement from findings in the Mental Health Atlas 2005, whereby Africa and Southeast Asia regions had 50% and 54.5%, respectively, of countries in their regions with mental health policy. Despite improvement, there remains a difference in the existence of policy and plans, and this reflects a number of factors. This could be attributed to the low importance of mental health in many countries and settings. It also has been explained on the basis of colonial rule, where the laws have not changed even after independence. 22 of the 36 countries that stated that they did not have standalone policies or plans for mental health had others have integrated policies for mental health as part of general health or disability policies. With regard to children and adolescent mental health, of the 78 countries that responded, only 46% had a plan or strategy.

Table 60–13.
Presence of a Standalone Mental Health Policy or Plan in the Countries of Each Region of the World Health Organization (WHO)

WHO Region	Percent of Countries
Global	79
Africa	72
Americas	82
Eastern Mediterranean	78
Europe	81
Southeast Asia	90
Western Pacific	83

Modified from World Health Organization. *Mental Health Atlas 2017*. World Health Organization; 2018.

Even where mental health plans are present, they have often not been supported by sufficient human and financial resources. Less than half of the countries surveyed had policies and plans that aligned with human rights conventions related to the participation of service users in decisions concerning them and stressed the importance of the transition from inpatient to community-based care. Specifically related to suicide, only a few countries had prevention strategies—approximately a third of high-income and upper-middle–income countries and 10% of low- and middle-income countries.

The Mental Health Atlas also considered the number of countries with mental health legislation and found that only 63% of countries (111 of 175 surveyed) had one. This varied from 44% in the African region to 77% in the European region (Table 60–14). Further, in some regions such as Africa and South East Asia, most countries did not have a dedicated authority or independent body to ensure compliance of mental health legislation with international human rights principles (Table 60–15).

Table 60–14.
Presence of Mental Health Legislation in the Countries of Each Region of the World Health Organization (WHO)

WHO Region	Percent of Countries
Global	63
Africa	44
Americas	61
Eastern Mediterranean	61
Europe	77
Southeast Asia	50
Western Pacific	83

Modified from World Health Organization. *Mental Health Atlas 2017*. World Health Organization; 2018.

Table 60–15.
Existence of a Dedicated Authority or Independent body to Assess Compliance of Mental Health Legislation with International Human Rights in the Countries of Each Region of the World Health Organization (WHO)

WHO Region	Percent of Countries
Global	60
Africa	45
Americas	63
Eastern Mediterranean	53
Europe	77
Southeast Asia	33
Western Pacific	67

Data from World Health Organization. *Mental Health Atlas 2017*. World Health Organization; 2018.

The median number of adult psychiatric beds per 100,000 population ranges from 2.5 in the African region to 59.7 in the European region (see Table 60–16). When considering availability related to the income status of countries, high-income countries had 52.6 beds per 100,000 compared to 1.9 beds per 100,000 in low-income countries. This significant disparity is ongoing despite the transition toward community-based residential care in high-income settings. Community care provision continues to lag in some regions, with a median rate of 437 visits per 100,000 in the South East Asia region compared to 8,073 in the European region (see Table 60–17) significantly. When considering rates related to income status of countries, high-income countries have 36 times the rate of adult outpatient visits compared to low-income countries (7,966 vs. 000220 per 100,000).

The global median mental health expenditure per capita in 2017 was 2.5 USD. Of the overall governmental health expenditure pie, mental health was less than 2% of the global median. There was noted to be significant variation between regions, with the European region spending 20 times that of the South East Asian and African regions. Notably, most of the mental health budget (approximately 80%) in low- and middle-income countries is spent in inpatient mental health hospitals. On the other hand, high-income countries spent approximately 43% on inpatient mental health hospitals.

Out-of-pocket payment is the most important method of financing mental health care in 43% of countries in Africa and 40% of countries in Southeast Asia, whereas it is not the primary method of financing mental health care in any European country (Table 60–18). Further, national health insurance and other reimbursement schemes did not cover care and treatment of severe mental disorders in two-thirds of countries globally.

Table 60–16.
Adult Inpatient Indicators (median rate per 100,000 population) in Each Region of the World Health Organization (WHO)

WHO Region	Facilities	Beds	Admissions
Global	0.22	16.4	99.1
Africa	0.1	2.5	20.2
Americas	0.5	20.8	83.9
Eastern Mediterranean	0.1	5.1	43.3
Europe	0.7	59.7	454.4
Southeast Asia	0.1	3.2	35.7
Western Pacific	0.7	18.4	114.3

Modified from World Health Organization. *Mental Health Atlas 2017*. World Health Organization; 2018.

Table 60–17.
Adult Outpatient Care Indicators (median rate per 100,000 population) in Each Region of the World Health Organization (WHO)

WHO Region	Facilities	Visits
Global	0.90	1,601
Africa	0.07	508
Americas	1.38	3,071
Eastern Mediterranean	0.45	632
Europe	1.63	8,073
Southeast Asia	1.21	437
Western Pacific	2.08	693

Modified from World Health Organization. *Mental Health Atlas 2017*. World Health Organization; 2018.

The global median of health workers is 9 per 100,000. The median number of psychiatrists per 100,000 population ranges from 0.4 in South East Asia to 9.9 in Europe (Table 60–19). The median number of psychiatric nurses per 100,000 population ranges from 0.5 in Africa to 23.2 in Europe. The median number of psychologists working in mental health care per 100,000 population ranges from 0.1 in Southeast Asia and Africa to 5.4 in the Americas.

Nurses represent the single largest category of mental health workers, with between 30% and 50% of the overall workforce in the different regions. Notably, in the African and American regions, the number of psychiatrists and psychologists is greater than the number of nurses (see Table 60–19). Large-scale migration of psychiatrists and other mental health professionals from low- and middle-income to high-income countries, as part of the larger picture of migration of health professionals in general, has been consistently documented. India and some sub-Saharan African countries are the most important contributors to the mental health workforce in the United Kingdom, although the United Kingdom has higher rates of psychiatrists per million population compared to India and sub-Saharan African countries.

Even worse is the situation concerning child and adolescent mental health care. According to the Mental Health Atlas, only 46% of countries worldwide have a specific child and adolescent mental health policy. With regard to inpatient bed availability for this population, high-income countries have eight times the number of beds as low-income countries. In terms of outpatient care, there is a similarly striking discrepancy. The median in child and adolescent outpatient visits was 1,609 per 100,000 in high-income countries and 11 per 100,000 in low-income countries. With regard to workforce availability, there are less than 0.1 psychiatrists per 100,000 in all income groups except for high-income countries (1.19 per 100,000).

Table 60–18.
Countries in Which Out-of-Pocket Payment Is the Most Common Method of Financing Mental Health Care in Each Region of the World Health Organization (WHO)

WHO Region	Percent of Countries
Global	17
Africa	43
Americas	6
Eastern Mediterranean	17
Europe	0
Southeast Asia	40
Western Pacific	11

Data from World Health Organization. *Mental Health Atlas 2017*. World Health Organization; 2018.

PRINCIPLES FOR MENTAL HEALTH PROGRAM DEVELOPMENT WORLDWIDE

According to the WHO, the development of mental health programs worldwide should be guided by the following principles: (1) providing treatment in primary care, (2) making psychotropic medications available, (3) giving care in the community, (4) educating the public, (5) involving communities, families, and consumers, (6) establishing national policies and legislations, (7) developing human resources, (8) linking with other relevant sectors, (9) monitoring community mental health, and (10) supporting more research.

Providing treatment in primary care involves the inclusion of recognition and treatment of common mental disorders in training curricula of all health personnel, the provision of refresher training to primary care physicians, and the improvement in referral patterns in primary health care. Making psychotropic medications available includes ensuring the availability of essential drugs in all health care settings and, in high-income countries, providing easier access to newer psychotropic drugs under public or private treatment plans. Giving care in the community implies moving people with mental disorders out of prisons, downsizing mental hospitals and improving care within them, developing general hospital psychiatric units, and providing community care facilities.

Educating the public includes promoting public campaigns against stigma and discrimination; using the mass media to promote mental health, foster positive attitudes, and help to prevent mental disorders; and, in high-income countries, launching public campaigns for the recognition and treatment of common mental disorders. Involving communities, families, and consumers includes supporting the formation of self-help groups, ensuring the representation of communities, families, and consumers in services and policy making, and fostering advocacy initiatives.

Table 60–19.
Median Number of Mental Health Professionals Per 100,000 Population in Each Region of the World Health Organization (WHO)

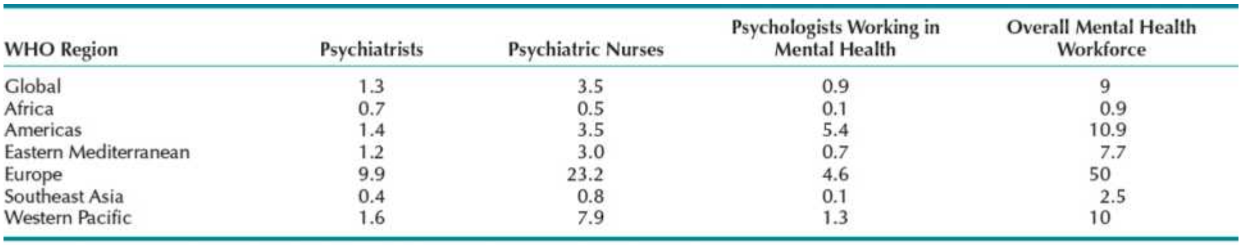

WHO Region	Psychiatrists	Psychiatric Nurses	Psychologists Working in Mental Health	Overall Mental Health Workforce
Global	1.3	3.5	0.9	9
Africa	0.7	0.5	0.1	0.9
Americas	1.4	3.5	5.4	10.9
Eastern Mediterranean	1.2	3.0	0.7	7.7
Europe	9.9	23.2	4.6	50
Southeast Asia	0.4	0.8	0.1	2.5
Western Pacific	1.6	7.9	1.3	10

Data from World Health Organization. *Mental Health Atlas 2017*. World Health Organization; 2018.

Establishing national policies and legislations includes revising legislations on the basis of current knowledge and human rights considerations and increasing the budget for mental health care. Developing human resources involves training psychiatrists and psychiatric nurses and, in middle- and high-income countries, creating national training centers for psychiatrists, psychiatric nurses, psychologists, and psychiatric social workers. Linking with other relevant sectors involves initiating school and workplace mental health programs, encouraging the activities of nongovernmental organizations in those areas, and, in high-income countries, initiating evidence-based mental health promotion programs in collaboration with other appropriate sectors.

Monitoring community mental health involves including mental disorders in basic health information systems, surveying high-risk population groups, instituting surveillance for specific disorders in the community, and, in high-income countries, monitoring the effectiveness of preventive programs. Supporting more research includes conducting studies in primary health care settings on the prevalence, course, outcome, and effect of mental disorders in the community; instituting effectiveness and cost-effectiveness studies for management of common mental disorders in primary health care; and, in high-income countries, carrying out research on the risk factors for mental disorders and on service delivery and investigating evidence on the prevention of mental disorders.

The guiding principles proposed for the prevention of suicide worldwide include (1) reducing access to means of suicide (e.g., pesticides and firearms), (2) treating people with mental disorders, (3) improving media portrayal of suicide, (4) training primary health care personnel, (5) implementing school-based programs, and (6) developing hot lines and crisis centers.

BARRIERS TO CHANGE

The most significant barriers to the implementation of the foregoing principles worldwide, according to the WHO, include the following: (1) some stakeholders may be resistant to the changes, (2) health authorities may not believe in the effectiveness of mental health interventions, (3) there may be no consensus among the country's stakeholders about how to formulate or implement the new policy, (4) financial and human resources may be scarce, (5) other basic health priorities may compete with mental health care for funding, (6) primary care teams may feel overburdened by their workload and refuse to accept the introduction of the new policy, and (7) many mental health specialists may not want to work in community facilities or with primary care teams, preferring to remain in hospitals. The suggested solutions include (1) adopting an "all-winners approach" that ensures that the needs of all stakeholders are taken into account, (2) developing pilot projects and evaluating their effect on health and consumer satisfaction, (3) asking for technical reports from international experts, (4) focusing the implementation of the mental health policy on a demonstration area and performing cost-effectiveness studies, (5) linking mental health programs to other health priorities, and (6) showing primary care practitioners that people with mental disorders are already a hidden part of their burden and that the burden will decrease if these disorders are identified and treated.

STIGMATIZING ATTITUDES TOWARD PEOPLE WITH MENTAL DISORDERS

Stigmatizing attitudes toward people with mental disorders are widespread in the general public and even among mental health professionals. Although it has been suggested that stigma may be less severe in Asian and African countries, a study carried out in India within the Stigma Programme of the World Psychiatric Association (WPA), in which 463 persons with schizophrenia and 651 family members were interviewed in four cities, reported that two-thirds of the respondents had suffered discrimination. Women and people living in urban areas were more stigmatized. Men experienced greater discrimination in the job area, whereas women experienced more problems in the family and social areas. In the family area, subtle discrimination in the form of decreased love, avoidance, rejection, distance, and excessive caution was frequently reported. In Ethiopia, among relatives of people with a diagnosis of schizophrenia or mood disorders, 75% reported that they had experienced stigma due to the presence of mental illness in the family, and 37% wanted to conceal the fact that a relative was ill. In a large survey carried out in China among people with schizophrenia and their family members, one-half of the respondents reported they had felt significantly stigmatized, and levels of stigma were found to be higher in urban areas. Stigmatization of people with mental disorders has been reported to be rare in Islamic countries.

Unlike people with physical disabilities, those with mental disorders are often perceived by the public to be in control of their disabilities and responsible for causing them. The view that “weakness,” “laziness,” or “lack of willpower” contribute to the development of mental disorders has been reported in several countries, including Turkey, Mongolia, and South Africa. The stigmatization of people with mental disorders may result in public avoidance, systematic discrimination, and reduced help-seeking behavior. In a survey carried out in 1996, in a probability sample of 1,444 adults in the United States, more than one-half of the respondents reported being unwilling to spend an evening socializing with, work next to, or have a family member marry a person with mental illness. Although most countries have some provision for disability benefits, people with mental illness are often specifically excluded from such entitlements.

Moreover, mental disorders are frequently not considered in social and private insurance schemes for health care. Shame is reported to be one of the main barriers to seeking help for mental disorders in both developed and developing countries.

A review by Bhugra and colleagues of mental health policies from Commonwealth countries found that less than half of the countries had mental health policies. They also found that there was no equity between funding for mental and physical health disorders. In further analysis global analysis, the research group found that only 11% of countries surveyed had no restriction on voting for persons with mental disorders, with 36% barring voting altogether. Related to the right to marry, 30% of countries surveyed had no restriction, and 37% barred those with mental disorders from getting married. Further, in 12% of countries, mental disorders were grounds for divorce.

Strategies for addressing the stigmatization of people with mental disorders have been subdivided into three groups: protest, education, and contact. There is some evidence that protest campaigns may be effective in reducing stigmatizing behaviors against people with mental disorders. This activist approach may be more workable in some countries rather than in others. Education may promote a better understanding of mental illness, and educated people may be less likely to endorse stigma and discrimination. Even in rural settings in low- and middle-income countries, strategies can be put into place successfully. There is no doubt that certain environments produce higher levels of stigma. These authors reported that internalized stigma and social relationships provide a risk factor for social isolation, especially among patients with schizophrenia. Ungar and colleagues observed that various strategies are needed to deliver antistigma programs, and health care systems themselves need to be included in such efforts. There is no doubt that stigma, whether it is internal or external, will influence help-seeking. Corrigan and colleagues provide a reminder that **s**tigma remains a complex construct. The people who are affected by it and also demonstrate stigma include public, self, and structural components as mentioned above in health care structures. There is no doubt that it directly affects people with mental illness, as well as their support system, provider network, and community resources. These authors also remind readers that cultures influence stigma too. It is likely that cultures see illnesses and idioms of distress in different ways, and stigma may be generated and seen in different ways. Internalized stigma, too, may be affected by cultures, although Pérez-Garín and colleagues point out that well-being can play a major role in mediating stigma. The definitions and understanding of well-being itself can be influenced by cultures. Stigma by professionals themselves deserves further exploration. Poreddi and colleagues report that stigma in medical students can be reduced after they have been exposed to psychiatric teaching and training. It has been suggested that public policy needs to change in order to reduce stigma.

An inverse relationship between having contact with a person with mental illness and endorsing stigmatizing behaviors has been documented. Several recent programs, such as Time to Change in the United Kingdom, have demonstrated that talking with patients with psychiatric illness does lead to a positive change in people's attitudes and changes attitudes in a more positive direction.

RELATIONSHIPS BETWEEN MENTAL AND PHYSICAL DISEASES

Mortality due to physical illness is significantly increased in people with severe mental disorders compared with the general population. In a follow-up study carried out in the United Kingdom, the standardized mortality ratio (SMR) for natural causes in people with schizophrenia was 2.32 (i.e., mortality was more than twice higher than in the general population). The SMR for causes "avoidable by appropriate treatment" was 4.68. The highest SMRs were those for endocrine, nervous, respiratory, circulatory, and gastrointestinal diseases. Increased nonsuicide all-cause mortality has also been reported for bipolar disorder (SMR 1.9 for males and 2.1 for females) and dementia (relative risk [RR] 2.63; 95% confidence interval [CI] 2.17 to 3.21). A meta-analysis of 15 population-based studies of the effect of a diagnosis of depression on subsequent all-cause mortality yielded a pooled odds ratio of 1.7 (CI 1.5 to 2.0). Evidence from low-income countries is limited, but a large population study conducted in Ethiopia found high mortality rates for major depression (SMR 3.55, 95% CI 1.97 to 6.39) and schizophrenia (almost 5% per year).

People with mental illness seem not to have benefited from the recent favorable trends concerning mortality due to some physical diseases. In a longitudinal study conducted in Australia, a steady decline in the mortality rate due to ischemic heart disease was observed in the general population during the period from 1980 to 1998, dropping from 209 to 143 per 100,000 person-years in men and from 139 to 117 per 100,000 person-years in women. In contrast, the mortality rate remained approximately constant in male psychiatric patients (on average 280 per 100,000 person-years) and increased from 153 to 234 per 100,000 person-years in female psychiatric patients.

The prevalence of several physical diseases is increased in people with mental disorders compared with the general population. In a study carried out in the United States, people with psychotic disorders were found to be more likely than other people to develop diabetes, hypertension, heart disease, asthma, gastrointestinal disorders, skin infections, malignant neoplasms, and acute respiratory disorders. The rate was increased even when only patients without a concomitant substance use disorder were considered. In a study conducted in Nigeria, 55.2% of persons with schizophrenia-spectrum disorders referred for the first time to a psychiatric clinic had at least one physical disease, whereas, in persons with neurotic disorders, the rate was 11.8%. A strong prospective association has been documented between depression and coronary heart disease outcomes, including fatal myocardial infarction; on the other hand, the incidence of depression is increased after myocardial infarction, especially in the first month after the event. Depression also increases the risk of type II diabetes. In South Asia, an association between maternal perinatal depression and infant malnutrition and stunting at 6 months has been repeatedly reported.

People with severe mental disorders are at increased risk of contracting HIV infection, although prevalence rates vary substantially worldwide. Conversely, a large multicenter study conducted in sub-

Saharan Africa, Asia, Latin America, Europe, and the United States reported a higher prevalence of depressive disorders among symptomatic HIV-seropositive people than in asymptomatic HIV-seropositive cases and seronegative controls. Evidence from both developed and developing countries shows that adherence to highly active antiretroviral therapy (HAART) is negatively affected by depression, cognitive impairment, and substance abuse.

The increased morbidity and mortality due to physical illness in people with severe mental disorders are first of all related to the increased prevalence in these people of several risk factors for cardiovascular, metabolic, and respiratory diseases, the most important of which are obesity, smoking, and alcohol and drug abuse.

In a study carried out in the United Kingdom, people with severe mental illness were significantly more likely to be obese (body mass index higher than 30) and morbidly obese (body mass index higher than 40) than the general population: The respective figures were 35.0 versus 19.4% and 3.7 versus 1.3%. When these figures were broken down by age and sex, 28.7% of men with severe mental illness aged between 18 and 44 years were obese, compared with 13.6% in the general population, and 3.7% versus 0.4% were morbidly obese. Even more striking were the figures concerning women of the same age: 50.6% versus 16.6% and 7.4% versus 2.0%, respectively.

In a meta-analysis of worldwide studies, a highly significant association between schizophrenia and current smoking was confirmed: The weighted average odds ratio was 5.9; it was 7.2 in males and 3.3 in females. The association remained significant when controls with severe mental illness were used (odds ratio 1.9). Heavy smoking and high nicotine dependence were also more frequent in people with schizophrenia than in the general population.

Patients with severe mental illness have also been reported to make poorer dietary choices and to be less physically active than the general population, but the relevant research evidence is not strong.

The increased risk of HIV infection in people with severe mental disorders may be explained by the high rates of sexual contact with multiple partners, the low levels of condom use, and the high prevalence of injected drug use and sexual contact with partners using injected drugs.

The access of people with mental illness to physical health care is reduced with respect to the general population. In the aforementioned study conducted in Australia, the SMR for ischemic heart disease was found to be 1.78 in men with schizophrenia. However, hospital admission for that disease was remarkably less common in both men and women with schizophrenia compared with the general population. Moreover, both men and women with schizophrenia were more than threefold less likely to receive revascularization procedures than the general population.

The quality of physical health care received by patients with severe mental illness is often worse than the general population. A study conducted in the United States found that adverse events during medical and surgical hospitalizations were significantly more frequent in patients with schizophrenia than in the other people, including infections due to medical care, postoperative respiratory failure, postoperative deep venous thrombosis or pulmonary embolism, and postoperative sepsis. All of these adverse events were associated with a significantly increased odds of admission to an intensive care unit and of death.

The decreased access of people with mental disorders to medical services has been related to several factors concerning the health care system. The effect of lack of insurance and the cost of care is well documented. In a study carried out in the United States, people with mental disorders were twice as likely as those without mental disorders to have been denied insurance because of a pre-existing condition (odds ratio 2.18). Having a mental disorder conferred a greater risk of having delayed seeking care because of cost (odds ratio 1.76) and of having been unable to obtain needed medical care (odds ratio 2.30).

Even when people with mental illness are seen by a doctor, their physical diseases often remain undiagnosed. Primary care providers may misperceive the medical complaints of people with mental disorders as “psychosomatic” or be unskilled or feel uncomfortable in dealing with this population. Underlying stigmatization may be involved. Moreover, during hospitalizations in medical and surgical wards, health care professionals may not be experienced in dealing with the special needs of patients with schizophrenia, may minimize or misinterpret their somatic symptoms, and may make inappropriate use of restraints or sedative drugs or fail to consider possible interactions of psychotropic drugs with other medications. On the other hand, many psychiatrists are unable or unwilling to perform physical and even neurologic examinations or are not up to date on the management of even common physical diseases.

To address this situation, the first step is raising awareness of the problem among mental health care professionals, primary care providers, and patients with schizophrenia and their families. The research information about the increased morbidity and mortality due to physical illness in people with schizophrenia should be appropriately disseminated.

Education and training of mental health professionals and primary care providers is a further essential step. Mental health professionals should be trained to perform at least basic medical tasks. They should be educated about the importance of recognizing physical illness in people with severe mental disorders and encouraged to become familiar with the most common reasons for underdiagnosis or misdiagnosis of physical illness in these patients. On the other hand, primary care providers should overcome their reluctance to treat people with severe mental illness and learn effective ways to interact and communicate with them; it is not only an issue of knowledge and skills but, most of all, one of attitudes.

Another essential step is the development of an appropriate integration between mental health and physical health care. A well-identified professional should be responsible for physical health care for each person with a severe mental disorder. Mental health services should be able to provide at least a standard routine assessment of their patients to identify or at least suspect the presence of physical health problems. Guidelines about the management of patients receiving antipsychotic drugs should be known and applied by all mental health services. Patients should be involved as much as possible; for instance, mental health professionals should encourage patients to monitor and chart their weight. Dietary and exercise programs should be routinely provided by mental health services. Flexible smoking cessation programs, which have shown some degree of success, could be considered in some settings.

ETHICAL ISSUES IN MENTAL HEALTH CARE

The protection and promotion of the human rights of people with mental disorders are emerging as a priority worldwide. In 1991, the United Nations issued Resolution 46/119 for the Protection of Persons with Mental Illness and the Improvement of Mental Health Care. In that resolution, the human rights of people with mental disorders and their right to treatment were codified for the first time in a UN document. The 25 principles covered the following areas: Definition of mental illness; protection of confidentiality; standards of care and treatment, including involuntary admission and consent to treatment; rights of persons with mental disorders in mental health facilities; protection of minors; provision of resources for mental health facilities; the role of community and culture; review mechanisms providing for the protection of the rights of offenders with mental disorders; and procedural safeguards protecting rights of persons with mental disorders. National governments were invited to promote the principles of the resolution by appropriate legislative, juridical, administrative, educational, and other provisions. However, violations of the human rights of people with mental disorders are still reported in many countries, and in a substantial number of low- and middle-income countries, patients in mental hospitals are physically restrained or secluded for long periods.

In Latin America, an influential document has been the Declaration of Caracas, adopted in 1990 by the Regional Conference on Restructuring Psychiatric Care in Latin America, which states that resources, care, and treatment for people with mental disorders should safeguard their dignity and human and civil rights and strive to maintain them people in their communities. The Declaration also states that mental health legislation should safeguard the human rights of people with mental disorders and that services should be organized so that these rights can be enforced.

In Africa, the Banjul Charter on Human and People's Rights, a legally binding document supervised by the African Commission on Human and People's Rights, in Article 5, addresses the right to respect for the dignity inherent in human beings and the prohibition of all forms of degradation, including cruel, inhuman, or degrading treatment.

According to the WHO, mental health legislation should cover the following issues: access to basic mental health care, least restrictive care, informed consent to treatment, voluntary and involuntary admission to treatment, competence issues, periodical review mechanism, confidentiality, rehabilitation, accreditation of professionals and facilities, and rights of families and caregivers.

In 1996, the WPA released the Madrid Declaration, which contains the ethical principles by which all national psychiatric societies are expected to abide. The Declaration includes seven general guidelines focusing on the aims of psychiatry: (1) psychiatrists must serve patients by providing the best therapy available consistent with accepted scientific knowledge and ethical principles and should devise therapeutic interventions that are the least restrictive to the freedom of the patient; (2) it is the duty of psychiatrists to keep abreast of scientific developments of their specialty and to convey updated knowledge to others; (3) the patient should be accepted as a partner by right in the therapeutic process, and the therapist–patient relationship must be based on mutual trust and respect, to allow the patient to make free and informed decisions; (4) treatment must always be in the best interest of the patient, and no treatment should be provided against the patient's will, unless withholding treatment would endanger life of the patient and/or those who surround them; (5) when psychiatrists are requested to assess a person, it is their duty to inform the person being assessed about the purpose of the intervention; (6) information contained in the therapeutic relationship should be kept in confidence and used exclusively for the purpose of improving the mental health of the patient; and (7) because psychiatric patients are particularly vulnerable research subjects, extra caution should be taken to safeguard their autonomy, as well as their mental and physical integrity.

INTERNATIONAL ORGANIZATIONS ACTIVE IN THE MENTAL HEALTH FIELD

There are many international organizations active in the mental health field. They include the WHO, which is the world's leading public health agency; some professional associations, among which the largest is the WPA, representing the psychiatric profession worldwide; and several organizations with a membership of users/families (e.g., the World Fellowship for Schizophrenia and the Global Alliance of Mental Illness Advocacy Networks [GAMIAN]) or of both mental health professionals and users/families (e.g., the World Federation for Mental Health [WFMH]).

The WHO is a UN agency with 193 member states, grouped into six regions (Africa, the Americas, European Mediterranean, Europe, Southeast Asia, and Western Pacific). It has a Department of Mental Health and Substance Abuse at its headquarters in Geneva and advisors for mental health in each of its regional offices. Its main functions are direction and coordination of international health work and technical cooperation with countries. Among the numerous recent WHO activities in the mental health field of special interest are the release of the *mhGAP Intervention Guide (mhGAP-IG)* for use in non-specialist health settings and the development of the report *Mental Health: New Understanding, New Hope*, which was completely devoted to mental health. It provided a summary of the current and projected impact of mental disorders and of the principles of mental health policy and service provision, as well as a set of recommendations for future action that can be adapted to the needs and resources of the various countries. Project Atlas aims to collect information on mental health resources across the world.

The WPA is an association of national psychiatric societies aimed at increasing the knowledge and skills necessary for work in the field of mental health and the care for the mentally ill. Its member societies number 145, spanning 121 countries and representing more than

250,000 psychiatrists. The WPA organizes the World Congress of Psychiatry every 3 years. It also organizes international and regional congresses and meetings, and thematic conferences. It has over 70 scientific sections, which are aimed at disseminating information and promoting collaborative work in specific domains of psychiatry. It has produced several educational programs and a series of books and consensus statements (including the Declaration of Madrid on ethical principles for psychiatric practice). It has an official journal, *World Psychiatry*, which is produced in English, Spanish, and Chinese and is indexed in PubMed and ranked number 1 in Social Science Citation index. Subsequently, several books and information sheets have been published (a full list is available on the WPA website).

The WFMH is a multidisciplinary advocacy and education organization aimed at promoting the advancement of mental health awareness, prevention, advocacy, and best-practice recovery-focused interventions worldwide. Among its activities is the organization of the World Mental Health Day, observed every year on October 10, each time with a different theme.

FUTURE PERSPECTIVES

Several statements have been made by various groups and organizations about priorities for future action in the mental health field at the international level. Of special interest is the aforementioned 2018 Lancet Commission on Global Mental Health and Sustainable Development. In this document, four main foundational pillars are identified: (1) Mental Health is a global public good. In this statement, there is recognition that “all countries can be thought of as developing countries in the context of mental health”, (2) Health problems exist on a continuum, (3) Health is an individual and unique product of social and environmental influences, and (4) Mental health is a fundamental human right that requires a rights-based approach. It is a good step given the focus on wider systems and sustainable development. Indeed, it is explicit in stating that “Although ‘no health without mental health’ is an important aspiration, the era of ‘no sustainable development without mental health’ has begun.”

There are key recommendations made. These include:

1. Reframe mental health within the Sustainable Development Goal Framework
2. Establish mental health care as a pillar of Universal Health Coverage
3. Use public policies to protect mental health
4. Listen to and engage people with lived experience
5. Invest more in mental health
6. Use research to guide innovation and implementation
7. Strengthen monitoring and accountability

It is imperative that local explanatory models and idioms of distress are taken into account so that culturally appropriate, culturally sensitive, and culturally accessible models of service care are developed and delivered. In virtually every country around the world, there are insufficient human and material resources dedicated to the mentally ill and mental illness. In many countries, useful models of engaging the local community exist, and it will be helpful to learn how these work. Furthermore, models using e-mental health and telemental health should be employed, particularly in countries with large geographical areas.

Among the strategies proposed to place mental health on the public health priority agenda are the development and use of uniform and clearly understandable messages for mental health advocacy and the education of decision-makers within governments and donor agencies on the evidence concerning the public health significance of mental disorders and the cost-effectiveness of mental health care. Strategies suggested to improve the organization of mental health services include, among others, the provision of incentive arrangements to overcome vested interests blocking change and the organization of international technical support to learn from countries that have experienced successful mental health reform.

To integrate the availability of mental health in general health care, it is proposed that mental health professionals be appointed and trained specifically for supporting and supervising primary health care staff. To promote the development of human resources for mental health, it is suggested that the professional and specialist workforce be increased and diversified and that the quality of mental health training be improved to ensure that it is practical and also occurs in community or primary care settings. Governments are asked to identify and scale up a priority package of service interventions and to increase mental health budget allocations, with the aim of attaining a minimum level of investment within 10 years of US$ 2 per capita in low-income countries and US$ 3 to 4 per capita in lower-middle–income countries.

International donors and governments of high-income countries, in particular those that are beneficiaries of the brain drain, are asked to place mental health on their priority agenda for health assistance to low- and middle-income countries. Finally, several indicators are proposed to monitor progress worldwide. From 2014 to 2017, WPA has launched an ambitious plan for the social justice for the mentally ill, focusing on domestic gender-based interpersonal violence; child physical and emotional abuse; prisoner mental health care; underserved populations such as asylum seekers, migrants, and refugees; those with intellectual disabilities, lesbian, gay, bisexual, and transgender individuals and older adult individuals and finally developing mental health promotion. Interestingly, World Medical Association is also focusing on child abuse and gender-based violence, and the two organizations are planning to work closely together. WPA is also working with other organizations to develop materials on peri-mental health as well as setting up the WPA-Lancet commission on psychiatry to explore the roles and responsibilities of psychiatrists in the first quarter of the 21st century.

FUTURE DIRECTIONS

This chapter has provided an overview of the burden of mental disorders worldwide, the treatment gap and the scarcity of resources for mental health care, especially in low- and middle-income countries, the WHO guidelines for the development of mental health programs, and the obstacles to their implementation, and several other issues of interest for world psychiatry. The discussion has been based as much as possible on data, although part of the data on world mental health is projections that may under- or overestimate the magnitude of the problems. Moreover, the threshold for the diagnosis of common mental disorders (e.g., depression and anxiety disorders) remains controversial, and although it can be argued that "subthreshold" conditions (not fulfilling current diagnostic criteria) might be a further neglected source of burden and disability, it can also be pointed out that some conditions fulfilling current diagnostic criteria might represent homeostatic reactions to adverse life events that are transient and do not require treatment. The threshold for the diagnosis of common mental disorders may be different in the various geographic and cultural contexts, making projections and international comparisons even more problematic. On the other hand, although several pharmacologic and psychotherapeutic interventions have been proved to be efficacious in various regions of the world, their feasibility and acceptability may not be the same worldwide, and current treatment guidelines may not be relevant to people living in several low- and middle-income countries. The same may apply to rehabilitation programs and models of community mental health care. The fact is that, although around 85% of the world's population lives in low- and middle-income countries, only a negligible proportion of research in the mental health field is conducted in those countries. Research-generated information is crucial to evaluate local needs, develop cost-effectiveness interventions, and monitor their implementation. Several initiatives are now addressing this research gap, and it can be hoped that a substantial revision of this chapter will be needed in the near future.

FURTHER READINGS

Baranne ML, Falissard B. Global burden of mental disorders among children aged 5–14 years. *Child Adolesc Psychiatry Ment Health*. 2018;12(1):1–9.

Bhugra D, Pathare S, Gosavi C, Ventriglio A, Torales J, Castaldelli-Maia J, Tolentino Jr EJL, Ng R. Mental illness and the right to vote: a review of legislation across the world. *Int Rev Psychiatry*. 2016;28(4):395–399.

Bhugra D, Pathare S, Joshi R, Kalra G, Torales J, Ventriglio A. A review of mental health policies from Commonwealth countries. *Int J Soc Psychiatry*. 2018;64(1):3–8.

Bhugra D, Pathare S, Nardodkar R, et al. Legislative provisions related to marriage and divorce of persons with mental health problems: A global review. *Int Rev Psychiatry*. 2016;28(4):386–392.

Carter R, Shimkets RP, Bornemann TH. Creating and changing public policy to reduce the stigma of mental illness. *Psychol Sci Public Interest*. 2014;15(2):35–36.

Charlson FJ, Ferrari AJ, Santomauro DF, et al. Global epidemiology and burden of schizophrenia: findings from the global burden of disease study 2016. *Schizophr Bull*. 2018;44(6):1195–1203.

Corrigan PW, Druss BG, Perlick DA. The impact of mental illness stigma on seeking and participating in mental health care. *Psychol Sci Public Interest*. 2014;15(2):37–70.

Daumit GL, Pronovost PJ, Anthony CB, Guallar E, Steinwachs DM, Ford DE. Adverse events during medical and surgical hospitalizations for persons with schizophrenia. *Arch Gen Psychiatry*. 2006;63:267–272.

Dharitri R, Rao SN, Kalyanasundaram S. Stigma of mental illness: an interventional study to reduce its impact in the community. *Indian J Psychiatry*. 2015;57(2):165–173.

Filik R, Sipos A, Kehoe PG, et al. The cardiovascular and respiratory health of people with schizophrenia. *Acta Psychiatr Scand*. 2006;113:298–305.

GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1789–1858.

Kessler RC, Aguilar-Gaxiola S, Alonso J, et al. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiol Psichiatr Soc*. 2009;18(1):23–33.

Kessler RC, Angermeyer M, Anthony JC, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization’s World Mental Health Surveys. *World Psychiatry*. 2007;6(3):168–176.

Liu Q, He H, Yang J, Feng X, Zhao F, Lyu J. Changes in the global burden of depression from 1990 to 2017: Findings from the Global Burden of Disease study. *J Psychiatr Res*. 2020;126:134–140.

Naghavi M. Global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the Global Burden of Disease Study 2016. *BMJ*. 2019;364:l94.

Oliveira SE, Esteves F, Carvalho H. Clinical profiles of stigma experiences, self-esteem and social relationships among people with schizophrenia, depressive, and bipolar disorders. *Psychiatry Res*. 2015;229:167–173.

Patel V, Boardman J, Prince M, Bhugra D. Returning the debt: how rich countries can invest in mental health capacity in developing countries. *World Psychiatry*. 2006;5:67–70.

Patel V, Saxena S, Lund C, et al. The Lancet Commission on global mental health and sustainable development. *Lancet*. 2018;392(10157):1553–1598.

Pérez-Garín D, Molero F, Bos AE. Internalized mental illness stigma and subjective well-being: the mediating role of psychological well-being. *Psychiatry Res*. 2015;228:325–331.

Pinfold V, Toulmin H, Thornicroft G, Huxley P, Farmer P, Graham T. Reducing psychiatric stigma and discrimination: evaluation of educational interventions in UK secondary schools. *Br J Psychiatry*. 2003;182(4):342–346.

Poreddi V, Thimmaiah R, Math SB. Attitudes toward people with mental illness among medical students. *J Neurosci Rural Pract*. 2015;6(3):349–354.

Prince M, Patel V, Saxena S, et al. No health without mental health—a slogan with substance. *Lancet*. 2007;370:859–877.

Stuart H, Arboleda-Flórez J, Sartorius N, et al. Special issue: stigma and mental disorders: international perspectives. *World Psychiatry*. 2005;4(Suppl 1).

ten Have M, de Graaf R, van Dorsselaer S, Beekman A. Lifetime treatment contact and delay in treatment seeking after first onset of a mental disorder. *Psychiatr Serv*. 2013;64(10):981–989.

Ungar T, Knaak S, Szeto AC. Theoretical and practical considerations for combating mental illness stigma in health care. *Community Ment Health J*. 2016;52(3):262–271.

Wang PS, Angermeyer M, Borges G, et al. Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*. 2007;6(3):177–185.

World Health Organization. *Mental Health Atlas 2017*. World Health Organization; 2018.

World Health Organization. *Mental Health Policy, Plans and Programmes*. World Health Organization; 2005.

World Health Organization. Suicide in the world: global health estimates (No. WHO/MSD/MER/19.3). World Health Organization; 2019.

World Psychiatric Association. Madrid Declaration on Ethical Standards for Psychiatric Practice. 1996. http://www.worldpsychiatricassociation.org/content/madrid-ethic-engish.shtml

Appendix A

History of Psychiatry Table

Persons and Events in Psychiatry

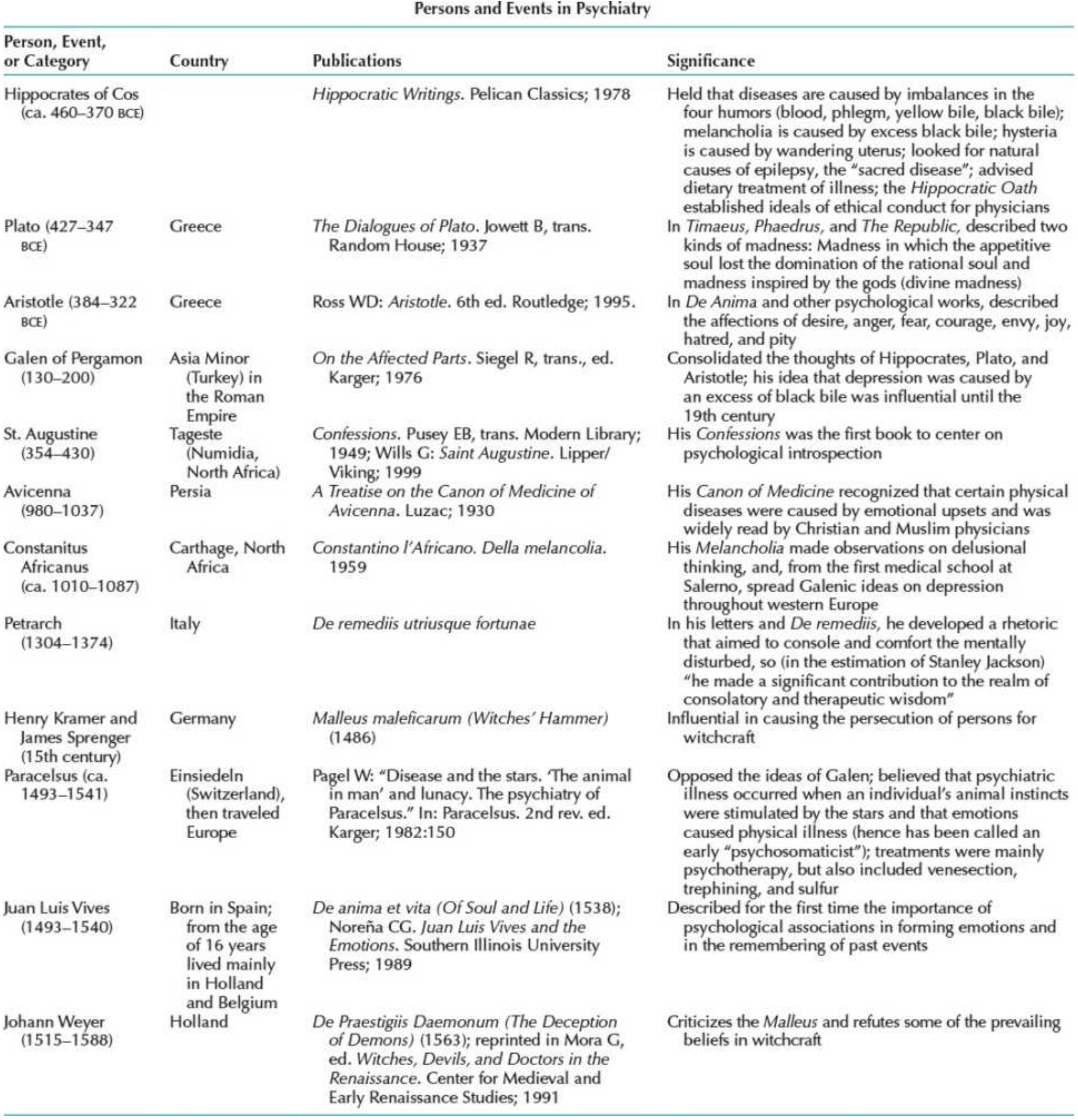

Persons and Events in Psychiatry

Person, Event, or Category	Country	Publications	Significance
Hippocrates of Cos (ca. 460–370 BCE)		*Hippocratic Writings*. Pelican Classics; 1978	Held that diseases are caused by imbalances in the four humors (blood, phlegm, yellow bile, black bile); melancholia is caused by excess black bile; hysteria is caused by wandering uterus; looked for natural causes of epilepsy, the "sacred disease"; advised dietary treatment of illness; the *Hippocratic Oath* established ideals of ethical conduct for physicians
Plato (427–347 BCE)	Greece	*The Dialogues of Plato*. Jowett B, trans. Random House; 1937	In *Timaeus, Phaedrus,* and *The Republic,* described two kinds of madness: Madness in which the appetitive soul lost the domination of the rational soul and madness inspired by the gods (divine madness)
Aristotle (384–322 BCE)	Greece	Ross WD: *Aristotle*. 6th ed. Routledge; 1995.	In *De Anima* and other psychological works, described the affections of desire, anger, fear, courage, envy, joy, hatred, and pity
Galen of Pergamon (130–200)	Asia Minor (Turkey) in the Roman Empire	*On the Affected Parts*. Siegel R, trans., ed. Karger; 1976	Consolidated the thoughts of Hippocrates, Plato, and Aristotle; his idea that depression was caused by an excess of black bile was influential until the 19th century
St. Augustine (354–430)	Tageste (Numidia, North Africa)	*Confessions*. Pusey EB, trans. Modern Library; 1949; Wills G: *Saint Augustine*. Lipper/ Viking; 1999	His *Confessions* was the first book to center on psychological introspection
Avicenna (980–1037)	Persia	*A Treatise on the Canon of Medicine of Avicenna*. Luzac; 1930	His *Canon of Medicine* recognized that certain physical diseases were caused by emotional upsets and was widely read by Christian and Muslim physicians
Constanitus Africanus (ca. 1010–1087)	Carthage, North Africa	*Constantino l'Africano. Della melancolia.* 1959	His *Melancholia* made observations on delusional thinking, and, from the first medical school at Salerno, spread Galenic ideas on depression throughout western Europe
Petrarch (1304–1374)	Italy	*De remediis utriusque fortunae*	In his letters and *De remediis,* he developed a rhetoric that aimed to console and comfort the mentally disturbed, so (in the estimation of Stanley Jackson) "he made a significant contribution to the realm of consolatory and therapeutic wisdom"
Henry Kramer and James Sprenger (15th century)	Germany	*Malleus maleficarum (Witches' Hammer)* (1486)	Influential in causing the persecution of persons for witchcraft
Paracelsus (ca. 1493–1541)	Einsiedeln (Switzerland), then traveled Europe	Pagel W: "Disease and the stars. 'The animal in man' and lunacy. The psychiatry of Paracelsus." In: Paracelsus. 2nd rev. ed. Karger; 1982:150	Opposed the ideas of Galen; believed that psychiatric illness occurred when an individual's animal instincts were stimulated by the stars and that emotions caused physical illness (hence has been called an early "psychosomaticist"); treatments were mainly psychotherapy, but also included venesection, trephining, and sulfur
Juan Luis Vives (1493–1540)	Born in Spain; from the age of 16 years lived mainly in Holland and Belgium	*De anima et vita (Of Soul and Life)* (1538); Noreña CG. *Juan Luis Vives and the Emotions*. Southern Illinois University Press; 1989	Described for the first time the importance of psychological associations in forming emotions and in the remembering of past events
Johann Weyer (1515–1588)	Holland	*De Praestigiis Daemonum (The Deception of Demons)* (1563); reprinted in Mora G, ed. *Witches, Devils, and Doctors in the Renaissance*. Center for Medieval and Early Renaissance Studies; 1991	Criticizes the *Malleus* and refutes some of the prevailing beliefs in witchcraft

Juan Huarte de San Juan (ca. 1530–1592)	Spain	*The Examination of Men's Wits* (1575)	An early account of differences in temperaments and dispositions
Giambattista Porta (1535–1615)	Italy	*De humana physiognomia* (1586)	Gave to physiognomy the air of at least a pseudoscience
Felix Plater (1536–1614)	Switzerland	*Practice of Medicine* (1602); *Observations of Diseases Injurious to Body and Mind* (1614)	New classification of diseases based on symptoms, causes, and treatments; careful description of all known psychiatric and organic diseases; the first physician to separate medicine from philosophy and make it a branch of natural science; a pioneer in studying the behavior and neuropathology of the persons with mental handicaps
Robert Burton (1577–1640)	England	*The Anatomy of Melancholy* (1621)	The most famous book on psychiatry in the 17th century; a comprehensive presentation of all previous medical-psychological thought on melancholy, also drawing on the nonmedical literature of Western civilization
Paolo Zacchia (1584–1659)	Italy	*Questiones medico-legales* (1621–1635)	Held that a physician, rather than a priest or lawyer, should evaluate a patient's responsibility for disturbed behavior; the beginning of forensic psychiatry
Thomas Sydenham (1624–1689)	England	Dissertatio epistolaris … [1682]. In: The Entire Works of Dr. Thomas Sydenham Newly Made English (1742)	Gave a comprehensive picture of the many symptoms of hysteria, believing that in the form of hypochondriacal complaints, it could exist in men and that it was caused by disturbed animal spirits; his approach to classifying the causes of symptoms resembled the modern psychiatric endeavor of constructing atheoretical classifications of psychiatric illnesses
Thomas Willis (1621–1675)	England	*Two Discourses Concerning the Soul of Brutes* (1683)	Summarized what was known about major psychiatric illnesses; recognized differences in illnesses when there was gross brain disease and when the brain seemed normal, attributing the latter to disturbed animal spirits; attributed hysteria to disturbed animal spirits acting on the brain, not to a wandering uterus
Georg Ernst Stahl (1660–1734)	Germany	*Theoria medica vera* (1707), *De animi morbis* (1708)	Theory of animism; the soul, *anima,* maintains functions of body in health and disease; psychiatric illness is caused either by inhibitions of anima or diseases of the body
William Battie (1703–1776)	England	*A Treatise of Madness* (1758)	First physician who made insanity his full work, raised "the mad business" to a respected specialty, and first used "madness" in the title of his book
Boissier de Sauvages (1706–1767)	France	*Nosologia methodica* (1765)	Nosology divided diseases into classes based on symptoms; although speculative and artificial, it stimulated a rethinking of concepts of disease
John Aiken (1747–1822)	England	*Thoughts on Hospitals* (1771)	First book on hospitals in which "lunatic hospitals" were discussed
Franz Anton Mesmer (1734–1815)	Austria, France	*Memoire sur la découverte du magnetism animal* (1779)	Showed that when a mental therapist used the so-called animal magnetism, it could cure cases of psychiatric illness; led to the discovery of hypnosis
Vincenzo Chiarugi (1759–1820)	Italy	*Regulations of the Hospitals of Santa Maria Nuova and of Bonifazio* (1789)	One of the first attempts to treat the insane inmates of asylums humanely and without restraints
William Cullen (1710–1790)	Scotland	*Nosology, or a Systematic Arrangement of Diseases* (1800)	A great 18th century nosologist who first used the terms *neurosis* and *neurotic* to describe mental diseases
Philippe Pinel (1745–1826)	France	*A Treatise on Insanity in Which Are Contained the Principles of a New and More Practical Nosology of Mental Disorders* (1801)	Classified mental illness into four main forms and established a new, humane treatment for inmates of insane asylums, which he called "the moral treatment of insanity"
Johann Reil (1759–1813)	Germany	*Rhapsodies about the Application of Psychotherapy to Mental Disturbances* (1803)	The founder of rational psychotherapy, recognizing the therapeutic value of institutional surroundings, music, psychodrama, and occupational therapy; first used the word *psychiatry,* and founded the first psychiatric journal
Benjamin Rush (1745–1813)	United States	*Medical Inquiries and Observations upon the Diseases of the Mind* (1812); Farr C. Benjamin Rush and American psychiatry. *Am J Psychiatry*. 1994;151(Suppl):65	The most famous American physician of his time and the only physician to sign the Declaration of Independence; the first American author of a book on general psychiatry; and the father of American psychiatry. A vocal abolitionist, his legacy is complicated by the fact he owned a slave, William Grubber. Similarly, although he argued for the equality of African Americans, he also wrote that Black people's skin was a type of leprosy.
Thomas Sutton (1767–1835)	England	*Tracts on Delirium Tremens* (1813)	First description of alcoholic delirium tremens

William Tuke (1732–1822); Samuel Tuke (1784–1857), grandson of William Tuke; Daniel Hack Tuke (1827–1895), youngest son of Samuel Tuke; John Charles Bucknill (1817–1895)	England	Tuke S. Description of the Retreat, an Institution near York, for Insane Persons of the Society of Friends. York; 1813; Tuke DH, Bucknill JC. *A Manual of Psychological Medicine Containing the History, Nosology, Description, Statistics, Diagnosis, Pathology, and Treatment of Insanity*. 1858	W. Tuke founded the York Retreat for the moral treatment of mentally ill Quakers in 1796; S. Tuke's *Description of the Retreat* influenced asylum treatment in England, Europe, and the United States; the *Manual* of D. Tuke and Bucknill was the first comprehensive textbook of psychiatry
Joseph Adams (1756–1818)	England	*A Treatise on the Supposed Hereditary Properties of Diseases … Particularly in Madness and Scrofula* (1814)	First book on the hereditary properties of diseases; argued that it was not a disease that was inherited but a susceptibility to diseases; therefore, prevention and cure were possible
Franz Joseph Gall (1758–1828), Johann Gaspar Spurzheim (1776–1832)	Austria, Germany	*The Physiognomical System of Drs. Gall and Spurzheim; founded on an anatomical and physiological examination of the nervous system in general and of the brain in particular* (1815)	The beginning of phrenology: Mapping parts of the brain, defining their psychological functions, and making psychological and psychotherapeutic predictions
Johann Christian Heinroth (1773–1843)	Germany	*Disturbances of the Mind* (1818); Hansen LA. Metaphors of mind and society: The origins of German psychiatry in the revolutionary era. *Isis*. 1998;89:387; Cauwenbergh LS. *J Chr A Heinroth* (1773–1843) a psychiatrist of the German Romantic era. *Hist Psychiatry*. 1991;2:365	First systematic textbook of psychiatry that attempted to formulate an actual clinical system of psychotherapy; the first to use the word *psychosomatic* and the first to hold a chair in psychological medicine at the University of Leipzig
Robert Gooch (1784–1830)	England	*An Account of … Diseases Peculiar to Women* (1829)	First account of postpartum psychosis
James Cowles Pritchard (1786–1848)	England	*A Treatise on Insanity and Other Disorders Affecting the Mind* (1835)	Standard textbook of psychiatry covering all of the literature on all known diseases; described moral insanity, later called *psychopathic personality*
Jean Etienne Dominique Esquirol (1782–1840)	France	*Mental Maladies: A Treatise on Insanity* (1838)	Coined the term *hallucination,* described "idiocy," classified insanities into monomania (partial insanity) and general delirium, and recognized both emotional and organic causes of illness; one of the authors of a comprehensive law providing an asylum for every department of France
Isaac Ray (1807–1881)	United States	*Treatise on Medical Jurisprudence of Insanity* (1838)	Founded American forensic psychiatry
James Braid (1795–1860)	England	*Neurypnology; or, the Rationale of Nervous Sleep* (1843)	Entirely separated hypnotism from animal magnetism and began the study of hypnotic phenomena
Daniel M'Naghton (1813–1865)	England	West DJ, Walk A, eds. *Daniel M'Naghton: His Trial and the Aftermath*. London Gaskell Books; 1977	Suffering from persecutory delusions, M'Naghton murdered the secretary of a government official in 1843; at his murder trial, he was ruled not guilty by reason of insanity and was committed to an insane asylum for the rest of his life; the "M'Naghton rule" has influenced English and American legal criteria for acquittal on the basis of insanity
Wilhelm Griesinger (1817–1868)	Germany	*Mental Pathology and Therapeutics* (1845)	Proclaimed that psychiatric diseases are brain diseases and that psychiatry had become a medical specialty
Jacques-Joseph Moreau de Tours (1804–1884)	France	*Du haschich et de l'aliénation mentale* (1845)	Described the effects of his taking hashish, and became the first psychiatrist to experience a drug-induced psychosis
Diversity, Equity, Inclusion, and Justice	United States	Samuel Cartwright presents a report to the Medical Association of Louisiana in 1851	Dr. Cartwright proposed the mental illness of "drapetomania," or the dissatisfaction of slaves with their condition and their resulting tendency to run away. He also coined the term dysaesthesia aethiopica, a supposed disorder that explains the lack of work ethic among slaves.
Walter Cooper Dendy (1794–1871)	England	Psychotherapeia, or the remedial influence of mind. *J Psychol Med Ment Pathol*. 1853;6:268	First introduced the term *psychotherapeia* (psychotherapy), which he defined as prevention and remedy of disease by psychical influence and which he predicted would become valuable in psychiatry

Jean-Pierre Falret (1794–1870), Jules Baillarger (1809–1890)	France	Baillarger J. La folie à double forme. *Bull Acad Imp Med.* 1853–1854;19:340; Falret JP. La folie circulaire. *Bull Acad Imp Med.* 1853–1854;19:382; Pichot P. The birth of the bipolar disorder. *Eur Psychiatry*. 1995;10:1	Baillarger and Falret called bipolar disorder "double insanity" and "circular insanity"; in 1899, Kraepelin named this illness *manic-depressive psychosis*
Thomas Kirkbride (1809–1883)	United States	*On the Construction, Organization, and General Arrangements of Hospitals for the Insane* (1854)	Set a standard for mid-19th century care of the chronically insane that commands respect today
John Conolly (1794–1866)	England	*The Treatment of the Insane without Mechanical Restraints* (1856)	Conolly's work for the nonrestraint system marked the success of a movement that began with Pinel and created a new approach to treating insanity throughout the civilized world
George Robinson (1821–1875)	England	*On the Prevention and Treatment of Mental Disorders* (1859)	Introduced the idea of looking beyond the precincts of the asylum to prevent mental illness
Gustav Theodor Fechner (1801–1887)	Germany	*Elements of Psychophysics* (1860)	Established the relationship between the intensity of stimuli and sensory reactions; called the founder of experimental psychology
Benedict-Augustin Morel (1809–1873)	France	*Traité des maladies mentales* (1860)	Presented cases of insanity and other mental illnesses caused by inherited "mental degeneration" that became worse from one generation to the next
Thomas Laycock (1812–1876)	England	*Mind and Brain* (1860)	Mentioned unconscious functional activity of the brain but did not further develop the idea
Forbes B. Winslow (1810–1874)	England	*On Obscure Diseases of the Brain and Disorders of the Mind* (1860)	First to mention psychical diagnostic tests (psychological tests) and the psychiatric interview
Ewald Hecker (1843–1909)	Germany	Die Hebephrenie. *Arch Pathol Anat Physiol.* 1871:52	First description of "hebephrenia" (later a subgroup of dementia praecox)
Karl Kahlbaum (1828–1899)	Germany	*Die Katatonie oder das Spannungsirresein* (Berlin, 1874)	First description of catatonia (later a subgroup of dementia praecox)
George Miller Beard (1840–1883)	United States	*A Practical Treatise on Nervous Exhaustion (Neurasthenia)* (1880)	Neurasthenia—a disease of mental and physical exhaustion—replaced the diagnosis of hypochondriasis; it became prevalent in the American and European upper classes from approximately 1870 to 1920
Richard von Krafft-Ebing (1840–1902)	Germany	*Psychopathia sexualis* (1886); Oosterhuis H. *Stepchildren of Nature: Kraft-Ebing, Psychiatry, and the Making of Sexual Identity*. C University of Chicago Press; 2000	Described cases of homosexuality and sexual perversions, claiming that some were caused by degeneration, and coined the terms *sadism, masochism,* and *sexual bondage; Psychopathia* was respected by lawyers and scientists, and it founded the science of sexology
Pierre Janet (1859–1947)	France	*L'automatisme psychologique: Essai de psychologie experimentale sur les formes inferieures de l'active* (Paris, 1889); van der Kolk BA, van der Hart O. Pierre Janet and the breakdown of adaptation in psychological trauma. *Am J Psychiatry*. 1989;146:1530	Showed that severe traumatic experiences caused two types of dissociative symptoms—*hysteria,* with dissociations of feelings or memories related to the experiences, and *psychasthenia,* with obsessions, phobias, and anxieties; the first systematic attempt to show how the mind can dissociate in the face of severe trauma
Herman Emminghaus (1845–1904)	Germany	*Psychic Disturbances of Childhood* (1887)	First textbook of child psychiatry
Jean-Martin Charcot (1825–1893)	France	Owen ARG. *Hysteria, Hypnosis, and Healing: The Work of JM Charcot.* Garrett Publications; 1971; by Goetz C, Bonduelle M, Gelfand T. *Charcot: Constructing Neurology*. Oxford University Press; 1995	In 1882, became the first professor of diseases of the nervous system at the University of Paris, helping to institutionalize neurology as a medical specialty and making Paris an international center for neurologic studies; although he erroneously believed that only hysterics could be hypnotized, he fully delineated and classified the protean manifestations of hysteria, and he was the first to give a psychological account of hysteria and show that some of its symptoms are the result of ideas that dominated the mind of the hysteric
Georges Gilles de la Tourette (1857–1904)	France	Etude sur une affection nerveuse caractérisée par l'incoordination motrice accompagnée d'echolalie et coprolalie. *Arch Neurol* (Paris). 1885;ix:19, 158; Yorston G, Hindley N, trans. and intro: Study of a nervous disorder characterized by motor incoordination with echolalia and coprolalia. *Hist Psychiatry.* 1998;9:97; Kushner HI. *A Cursing Brain? The Histories of Tourette Syndrome*. Harvard University Press; 1999	Described a syndrome of involuntary movements and echolalia and coprolalia in nine patients, seven of whom were at the Paris Hôpital de la Salpêtrière, where he was a resident neurologist in training; Jean-Martin Charcot, director of the Salpêtrière and France's leading neurologist, then sanctioned the use of de la Tourette's name for the syndrome

Cesare Lombroso (1835–1909)	Italy	*Genius and Insanity* (1864), *The Delinquent Man* (1876), *The Female Offender* (1893)	Emphasized the study of the individual personality, which gave rise to pathographies—psychopathologically oriented biographies; his books on delinquent men and women stimulated a school of criminal anthropology that engendered the science of criminology
Emil Kraepelin (1856–1926)	Germany	Kraepelin E. *Psychiatrie: Ein Lehrbuch für Studerende und Aerzte*. 6th ed. (1899); *Memoirs*. Springer-Verlag; 1987	Divided the major psychoses into two groups: Dementia praecox, which deteriorated to dementia, and manic-depressive psychosis, which did not deteriorate
John Hughlings Jackson (1834–1911)	England	*Selected Writings of John Hughlings Jackson* (London, 1931–1932)	From 1870 to 1900, developed the thesis that psychiatric symptoms are a regression from higher functions that are the products of evolution; the symptoms resulted from activating more primitive functions
Sigmund Freud (1856–1939)	Austria	*The Interpretation of Dreams* (1900), *Three Essays on the Theory of Sexuality* (1905), *Introductory Lectures on Psychoanalysis* (1915–1917)	Discovered the manifestations of the unconscious and how to use them in treating psychiatric patients; described infantile sexuality and how it accounted for adult sexual dysfunctions; founded psychoanalysis
Morton Prince (1854–1929)	United States	*The Dissociation of a Personality* (1905)	Early account of a multiple personality; emphasized techniques of hypnosis and manifestations of the unconscious
Clifford Beers (1876–1943)	United States	*A Mind That Found Itself* (1908)	Account of experiences in psychiatric hospitals that stimulated the mental hygiene movement in the United States
Eugen Bleuler (1857–1939)	Switzerland	*Dementia Praecox or the Group of Schizophrenias* (1911)	Coined the term *schizophrenia* and described its symptoms
Alfred Binet (1857–1911)	France	*A Method of Measuring the Development of the Intelligence of Young Children*, with T. Simon (1911); English translation, 1913	Developed the first practicable intelligence test for children; it has become the prototype for subsequent tests of intellectual ability in children and adults
Hideyo Noguchi (1876–1928)	United States	Noguchi H, Moore JW. A demonstration of the *Treponema pallidum* in the brain in cases of general paralysis. *J Exp Med*. 1913:17	The definitive demonstration, after a century of controversy, that the syphilitic organism causes general paresis; the first time that the cause of a major psychosis became known
Alfred Adler (1870–1937)	Austria	*Study of Organ Inferiority and Its Psychical Compensations* (1917); Stepansky PE. *In Freud's Shadow: Adler in Context*. Analytic Press; 1983	First psychoanalytic defector from Freud; founded the school of individual psychology and coined the terms *life style* and *inferiority complex*
Hermann Rorschach (1884–1922)	Switzerland	*Psychodiagnostik* (1921)	Inkblot test revealed unconscious motivations and ego defenses against them and was used for psychiatric diagnosis; it stimulated the development of other projective diagnostic tests
Julius von Wagner-Jauregg (1857–1940)	Austria	de Rudoli G. *Therapeutic Malaria* (1927); Wagner-Jauregg J. The history of the malaria treatment of general paresis. *Am J Psychiatry*. 1994;151(Suppl):231; Brown EM. Why Wagner-Jauregg won the Nobel Prize for discovering malaria therapy for general paresis of the insane. *Hist Psychiatry*. 2000;11:371	During the decade 1917–1927, showed that general paresis patients underwent remissions when malaria was induced; for the time, it was the most successful organic treatment of a psychosis; for that, he became, in 1927, the first psychiatrist to receive the Nobel Prize; in the 1940s, penicillin became the treatment of choice for paresis
Otto Rank (1884–1939)	Austria, France, United States	*The Myth of the Birth of the Hero* (1909), *The Trauma of Birth* (1924), *Will Therapy* (1936); Lieberman EJ. *Acts of Will: The Life and Work of Otto Rank*. Free Press, Macmillan; 1985	Made important contributions to Freudian psychoanalysis by showing how a person's wishes and fantasies influenced the formation of human myths; breaking with Freud, Rank formed his theory of how the experience of birth influenced a person's mental development and then his technique of will therapy
American Board of Psychiatry and Neurology (ABPN)	United States	Freeman W, Ebaugh F, Boyd D Jr. The founding of the American Board of Psychiatry and Neurology, Inc. *Am J Psychiatry*. 1959;115:769	In 1934, a group of prominent American psychiatrists and neurologists founded the ABPN, comprising equal numbers of psychiatrists and neurologists, with separate qualifications and examinations for psychiatrists and neurologists; the qualifications for psychiatrists included 3 years of study of psychiatry after an internship, and 2 years of hospital practice in psychiatry; the original purpose of the ABPN was, in the words of Walter Freeman, "to create standards of ethical practice and professional competence based on adequate fundamental education in the field"; since its founding, it has maintained these standards while keeping abreast of advances in information and technology

Henry A. Murray (1893–1988)	United States	*Thematic Apperception Test*. Harvard University Press; 1943	The Thematic Apperception Test, originated in 1935, is a method used for uncovering conscious and unconscious dominant themes in a person's mental life
Organizational Psychiatry/ Diversity, Equity, Inclusion, and Justice	United Kingdom	The Indian division of the Royal Medico-Psychological Association is formed in 1935.	
Charles Bradley (1902–1979)	United States	Bradley C. The behavior of children receiving Benzedrine. *Am J Psychiatry*. 1937:577; Gross MD. Origin of stimulant use for treatment of attention deficit disorder. *Am J Psychiatry*. 1995;152:298	The first physician to record the successful treatment of what came to be called attention-deficit disorder in children with stimulant medications
Cure of pellagra	United States	Roe D. *A Plague of Corn: The Social History of Pellagra*. Cornell University Press; 1973	Pellagra is a niacin-deficiency disease manifested by diarrhea, dermatitis, and dementia; in the South, it was caused by removal of niacin in the milling of corn and was a major cause of dementia; in 1912, Joseph Goldberger (1874–1929), of the U.S. Public Health Service, proved that pellagra was caused by absence of a pellagra-preventive factor in the diet; in 1937, that factor was identified as niacin; since then, with niacin-enriched diets, pellagra has largely vanished; pellagra has also occurred in other countries whose staple grain is maize (American corn)
Genetics of schizophrenia	Germany, United States	Rüdin E. *Studien über Verebung und Entstehung geistiger Störungen. I. Zur Vererbung und Neuetstehung der Dementia Praecox. [Studies on the Heredity and Genesis of Mental Diseases. I. On the Heredity and New Formation of Dementia Praecox.]* Berlin: Verlag von Julius Springer; 1916; Kallman FJ. *The Genetics of Schizophrenia*. JJ Augustin; 1938; Rainer JD. The contributions of Franz J. Kallmann to the genetics of schizophrenia. In: Cancro R, Dean SR, eds. *Research in the Schizophrenic Disorders: The Stanley R. Dean Award Lectures*. Vol. 2. Spectrum; 1985	Rüdin (1874–1952) and Kallman (1897–1965) showed that schizophrenia was transmitted by recessive genes (although they differed over how these genes exerted their effects); Kallman first worked under Rüdin in the Max Planck Institute of Psychiatry in Munich, then studied the incidence of schizophrenia among the siblings and descendants of 1,087 hospitalized schizophrenic patients in two Berlin hospitals; when he emigrated to the United States in 1936, he brought the German manuscript of his book on these patients; it was translated into English in 1938 and became a pioneering work; Kallman spent the rest of his life working at the New York Psychiatric Institute, where he established the first full-time medical genetics department in a U.S. psychiatric institution
ECT	Italy	Abrams R. The treatment that will not die. *Psychiatr Clin North Am*. 1994;17:525; Berrios G. The scientific origins of electroconvulsive therapy: A conceptual history. *Hist Psychiatry*. 1997;8:105	Two Italian psychiatrists, Ugo Cerletti (1897–1963) and Lucio Bini (1908–1964), first used ECT in Rome in 1938 to produce convulsions that alleviated symptoms in a schizophrenic patient; it was then found to be more effective in mood disorders than in schizophrenia, and although its use has declined because of psychotropic drugs, modified forms of ECT are used today
Karen Horney (1885–1952)	Germany until 1932, then United States	*The Neurotic Personality of Our Time* (1937), *New Ways in Psychoanalysis* (1939), *Neurosis and Human Growth: The Struggle Toward Self-Realization* (1950); Paris B. Karen Horney: *A Psychoanalyst's Search for Self-Understanding*. Yale University Press; 1994	Opposed Freud's theory of the castration complex in women and his emphasis on the oedipal complex and sexuality as influencing neurosis; argued that neurosis was influenced by the culture of the society in which one lived
Adolf Meyer (1866–1950)	Switzerland until 1893, then United States	Meyer A. A short sketch of the problems of psychiatry. *Am J Insanity*. 1896–1897;53: 538; Lidz T. Adolf Meyer and the development of American psychiatry. *Am J Psychiatry*. 1996;123:320; Adolf Meyer. In: Havens L, ed. *Approaches to the Mind*. Little, Brown.	From 1912 to 1940, the dominant figure in American psychiatry; his concept of psychobiology viewed the patient as a biologic and psychological unity who became ill because of internal pathology and maladaptations to the environment; he treated patients with medical and nonmedical therapies in community clinics; his work has become part of present-day American psychiatry
Constitution, personality, and mental illness	Germany, United States	Kretschmer E. Body Build and Character (1922); Sheldon WH. Varieties of Human Physique (1940)	Kretschmer (1888–1964) and Sheldon (1899–1977) postulated that thin, asthenic individuals were prone to introversion, schizoid tendencies, and schizophrenia, whereas individuals who were short and round were predisposed to be extroverts and had cyclothymic personalities and manic depression; there are many exceptions to these postulates

Führer decree (1939)	Germany	Lifton RJ. The Nazi doctors. In: "Euthanasia": *Direct Medical Killing*. Basic Books; 1986: Chapter 2; Gallagher HG. *By Trust Betrayed: Patients, Physicians, and the License to Kill in the Third Reich*. Henry Holt; 1990; Annas GJ, Grodin MA, eds. *The Nazi Doctors and the Nuremburg Code. Human Rights in Human Experimentation*. Oxford University Press; 1992	In October 1939, soon after the outbreak of World War II, a Führer decree ordered doctors to kill patients who had incurable medical illnesses (this grew out of the Nazi doctrine of preserving racial purity by eliminating those who were "biologically unfit"); during the course of the war, approximately 270,000 mental patients were killed by physicians and medical personnel; Henry Friedlander postulated in *The Origins of Nazi Genocide: From Euthanasia to the Final Solution* (1995) that these medical killings, which began in 1940, served as a stimulus and practical model for the Nazi genocide of Jews and Gypsies that began in 1941
MMPI	United States	Buchanan R. The development of the Minnesota Multiphasic Personality Inventory. *J Hist Behav Sci*. 1994;29:148	The MMPI, developed by S. R. Hathaway (1903–1984) and J. C. McKinley (1891–1950) at the University of Minnesota during the early 1940s, became the most widely used and researched psychological personality measurement device; a revised version of the MMPI, MMPI-2, was published in 1989; in 1999, a third edition of MMPI-2 was published, containing the results of new research studies along with a new chapter on forensic applications of MMPI-2
Leo Kanner (1894–1981)	Austria, United States	Child Psychiatry. Springfield, IL: CC Thomas; 1935; Autistic disturbances of affective contact. *Nerv Child*. 1943;2:217	The first English-language textbook of child psychiatry and a pioneering description of what Kanner later called "early infantile autism"
Melanie Klein (1882–1960), Anna Freud (1895–1982)	Central Europe, then England	Klein M. *The Psychoanalysis of Children* (1932); Kernberg O. Melanie Klein. In: Kaplan HI, Freedman AM, Sadock BJ, eds. *Comprehensive Textbook of Psychiatry*. 3rd ed. Williams & Wilkins; 1980; Grosskurth P. *Melanie Klein: Her World and Her Work*. Knopf; 1986; Freud A. *The Ego and the Mechanisms of Defence* (1936), *The Psychoanalytic Treatment of Children* (1946), A short history of child analysis. *Psychoanal Study Child*. 1966;21:7; Young-Bruehl E. *Anna Freud: A Biography*. Summit Books; 1988	Klein and A. Freud were pioneers in originating different concepts of child psychoanalysis; Klein's concepts of the "paranoid-schizoid and depressive positions" of very young children has influenced child psychiatry and the treatment of adult borderline and psychotic patients; A. Freud formulated a group of basic concepts of child psychoanalysis and extended the ego psychology work of her father, Sigmund Freud, by delineating the ego defenses, including "identification with the aggressor"
Helene Deutsch (1884–1982)	Europe until 1935, then United States	*The Psychology of Women*, 2 vols. (1945); some forms of emotional disturbance and their relationship to schizophrenia ("As If"). In: Deutsch H, ed. *Neuroses and Character Types*. International Universities Press; 1965: Chapter 20; Roazen P. *Helene Deutsch: A Psychoanalyst's Life*. Anchor Press/Doubleday; 1985	For several decades the most comprehensive Freudian view of the life cycle of women; her most famous clinical concept was the "as if" personality: An apparently normal person who gives a good semblance of adaptation to reality, yet is actually devoid of genuine emotion
Ivan Petrovich Pavlov (1849–1936), Andrew Salter (1914–1996)	Russia, United States	Pavlov IP. *Lectures on Conditioned Reflexes*. 2 vols. International Publishers; 1941; Babkin, BP. *Pavlov: A Biography*. University of Chicago Press; 1949; *Salter A. Conditioned Reflex Therapy*. Farrar, Straus; 1949	From 1903 to 1936, Pavlov showed how conditioned reflexes influence normal and pathologic thought; Salter then applied conditioned reflex therapy to psychiatric patients by postulating that neuroses result from an excess of cortical inhibition and by emphasizing assertion training; his therapy became one trend in the development of behavioral therapy
NMHA (1946), NIMH (1949)	United States	Grob GN. *From Asylum to Community: Mental Health Policy in Modern America*. Princeton University Press; 1991: Chapters 2–4. Romano J. Reminiscences: 1938 and since. *Am J Psychiatry*. 1990;147:785	Congressional passage in 1946 of the NMHA provided American psychiatry for the first time with generous funds for psychiatric education and research, and led to the creation of the NIMH in 1949, with goals of research in schizophrenia and training of mental health personnel
Organizational Psychiatry	India	The Indian Psychiatric Society is Established in 1947	
Popular Media	United Statues	The Snake Pit, by Mary Jane Ward is published in 1946. A movie adaptation is released in 1948, sparking a call for reforms to the U.S. state psychiatric hospital system.	

Erik H. Erikson (1902–1994)	Europe; United States since 1930s	*Childhood and Society* (1950), *Identity and the Life Cycle* (1959); Friedman LJ. *Identity's Architect: A Biography of Erik H. Erikson.* Scribner; 1999; Wallerstein R, Goldberger L, eds. *Ideas and Identities: The Life and Work of Erik Erikson.* International Universities Press; 1999	Restated Freud's concepts of infantile sexuality, developed concepts of adult identity and identity crisis, and applied psychoanalytic concepts to U.S. cultural life and political history; his *Young Man Luther: A Study in Psychoanalysis and History* (1958) influenced the writing of psychohistory
Carl Rogers (1902–1987)	United States	*Client-Centered Theory* (1951)	Showed that in many patients there existed constructive mental forces that had been previously unrecognized and unemployed and that a therapist, through the use of insight and catharsis, could aid a patient to use these forces to make beneficial changes in his or her life; the approach marked the beginning of a shift of psychotherapy from physicians and psychoanalysts to clinical psychologists and psychiatric social workers
Donald Winnicott (1896–1971)	England	Winnicott D. Transitional objects and transitional phenomena [1951]. In: *Collected Papers: Through Paediatrics to Psychoanalysis*. Tavistock, 1958; Kohon H. *The British School of Psychoanalysis: The Independent Tradition*. Yale University Press; 1986; Rodman FR. *Winnicott: Life and Work*. Perseus; 2003	Observed the transitional objects to which a young child becomes attached, which provide a bridge between the child's inner and outer worlds and then influence the development of play, creativity, and cultural life in general; a member of the "independent tradition" in British psychoanalysis (i.e., independent from the views of Klein and Anna Freud), which, instead of emphasizing instincts, stressed object relations
Jean Delay (1907–1987), Pierre Deniker (1917–1999)	France	Delay J, Deniker P. Le traitement des psychoses par une méthode neurolytique dérivée de l'hibernothérapie. *C R Congrés Med Alién Neurol France*. 1952;50:497; Thuillier J. *Ten Years That Changed the Face of Mental Illness*. Martin Dunitz; 1999; Healy D. *The Creation of Psychopharmacology*. Harvard University Press; 2002	In their first reports on chlorpromazine to French psychiatrists, Delay and Deniker emphasized that patients were quieted like animals in hibernation and called the drug "hibernotherapie"; chlorpromazine and related compounds then came into wide general use for treating patients with psychosis and helped to reduce the number of asylum patients
Harry Stack Sullivan (1892–1949)	United States	Sullivan HS. *The Interpersonal Theory of Psychiatry*. Perry HS, Gawel ML, eds. WW Norton; 1953; Perry HS. *Psychiatrist of America: The Life of Harry Stack Sullivan*: The Belknap Press of Harvard University Press; 1982	The interpersonal theory holds that a person's impulses and strivings cannot be studied in and for themselves, but only as they appear in an interpersonal situation; Sullivan also coined the terms *participant observer* (therapists must be aware of the overt and covert behavior of the patient and their own reactions); *consensual validation* (the awareness of patient and therapist of the terminology they use); and *parataxic distortion* (the patient's distortion of the real person of the therapist necessitated by the patient's personality structure)
Egas Moniz (1874–1955), Walter Freeman (1895–1972)	United States, Europe	Valenstein ES. *Great and Desperate Cures: The Rise and Decline of Psychosurgery and Other Radical* Treatments for Mental Illness. Basic Books; 1986; Pressman JD. *Last Resort: Psychosurgery and the Limits of Medicine*. Cambridge University Press; 1998	During the late 1940s and early 1950s, under the influence of Moniz and Freeman, many patients with severe psychoses and OCD were treated by lobotomy—destruction of the white matter of the frontal lobes of the brain; no adequately controlled studies were ever performed; some patients improved, but others experienced irreversible personality deterioration; with the advent of the psychotropic drugs, the use of lobotomies markedly declined
Advent of the double-blind test	United States	Gold H, Kwit NT, Otto H. The xanthines (theobromine and aminophylline) in the treatment of cardiac pain. *JAMA*. 1937;108:2173; Hampson JL, Rosenthal D, Frank J. A comparative study of the effect of mephenesin and placebo on the symptomatology of a mixed group of psychiatric outpatients. *Bull Johns Hopkins Hosp*. 1954;95:170	In the double-blind test, a drug or a procedure and a placebo are compared in such a way that neither the patient nor the therapists involved in treatment know which preparation is being administered; this type of test was first used in medicine in 1937 to evaluate the efficacy of the xanthines in relieving cardiac pain; it was used in psychiatry in 1954 to evaluate the efficacy of mephenesin, and then lithium, in relieving different psychiatric illnesses; since 1954, the main psychiatric use of double-blind tests has been to evaluate the efficacy of psychotropic drugs and psychiatric procedures

Carl Gustav Jung (1875–1961)	Switzerland	*Psychological Types* (1921), *Two Essays on Analytical Psychology* (1912–1928), *The Structure and Dynamics of the Psyche* (1916–1952), *The Archetypes and the Collective Unconscious* (1934–1955); Kirsch T. *The Jungians: A Comparative and Historical Perspective*. Routledge; 2000	After separating from Freud, Jung founded the school of analytical psychology, developing new psychotherapeutic approaches and concepts of the unconscious (especially the collective unconscious) and new personality types, such as introvert and extrovert
Antidepressant drugs	Switzerland, United States	Kuhn R. The treatment of depressive states with an iminodibenzyl derivative (G 22355). *Swiss Med J*. 1957;87:1135; Loomer HP, Saunders JC, Kline NS. A clinical and pharmacodynamic evaluation of iproniazid as a psychic energizer. *Psychiatr Res Rep*. 1957;8:129; Healy D. *The Antidepressant Era*. Harvard University Press; 1997; Kuhn R. *Hist Psychiatry*. 2007;17:253	The first reports of the effectiveness of a tricyclic drug, imipramine (G 22355), and an MAOI, iproniazid (Marsilid), in the treatment of depression
B. F. Skinner (1904–1990), Joseph Wolpe (1915–1997)	United States, South Africa	Skinner BF. *Science and Human Behavior.* Free Press; 1953; Wolpe J. *Psychotherapy by Reciprocal Inhibition*. Stanford University Press; 1958	Skinner's operant conditioning and Wolpe's reciprocal inhibition were important trends in the development of behavioral therapy
Development of child psychiatry	United States	Lewis M, ed. *Child and Adolescent Psychiatry: A Comprehensive Textbook.* 2nd ed. Williams & Wilkins; 1996: Chapters 120 (Rafferty F. Effects of health delivery systems on child and adolescent mental health care); 123 (Schowalter J. Recruitment, training, and certification in child and adolescent psychiatry in the United States); 127 (Bernstein D. The discovery of the child: A historical perspective on child and adolescent psychiatry)	After 1945, child psychiatry developed as a psychiatric specialty that separated from adult psychiatry with the formation of the American Academy of Child Psychiatry (1953) and the American Board of Child Psychiatry (1959)
Karl Jaspers (1883–1969)	Germany	*General Psychopathology* (1st ed., 1913; 7th ed., 1959) (English trans., 1963)	Delineated different mental states to show the meaningful connections between different thoughts (without looking for an underlying cause), and to have empathy for both the felt mental state of the patient and the observing mental state of the physician; this view influenced psychotherapy in general and became part of the tenets of existential psychotherapy
Kurt Schneider (1887–1967)	Germany	Schneider K. *Psychopathic Personalities.* Cassell; 1958; Schneider K. *Clinical Psychopathology*. Grune Stratton; 1959	In clinical psychiatry, Schneider carried forward—with modifications—the descriptive and phenomenologic approaches of Kraepelin and Jaspers; his work included clinical and phenomenologic descriptions of the experience of delusions and hallucinations and a classification and definition of the primary and secondary symptoms of schizophrenia; *Psychopathic Personalities* described a number of abnormal personality types that he encountered in practice and differentiated from psychotic states.
Michel Foucault (1926–1984)	France	*Histoire de la folie* (1961); English trans., *Madness and Civilization: A History of Insanity in the Age of Reason*. Pantheon; 1965	Gave an unconventional history of attitudes toward insanity in Europe from 1500 to 1800; believing that sanity and insanity are largely conditioned by history and the model on which a given society operated, he postulated that creation of the asylum and moral therapy at the end of the 18th century was not progress but an attempt to force insane people to conform to existing social values, and that the authority of the psychiatrist who headed the hospital did not derive from his knowledge of the science of human behavior, but from his association with the power of society; although important historians showed that many of his facts and interpretations contained serious errors, his ideas continue to be debated; *Histoire de la folie* is, in the words of Porter, "easily…the single most influential text of psychiatric historiography in the second half of the century"

Benzodiazepine drugs	United States	Cohen IM. The benzodiazepines. In: Ayd FJ Jr, Blackwell B, eds. *Discoveries in Biological Psychiatry*. Ayd Medical Communications; 1984:130	The benzodiazepines were developed by the pharmaceutical company of Hoffmann-La Roche; chlordiazepoxide (Librium) was first marketed in 1960 and diazepam (Valium) in 1963; the drugs became widely prescribed as anxiolytic drugs in cases of nonpsychotic anxiety
Haloperidol in the treatment of Gilles de la Tourette syndrome	France	Seignot JN. Un cas de maladie des tics de Gilles de la Tourette guéri par le R-1625. Ann Med Psychol. 1961; cxix: 578; The case of Gilles de la Tourette syndrome: A condition of nervous tics cured by R. 1625. *Hist Psychiatry*. 1997;8:434	Reported the cure of one case of Tourette syndrome with haloperidol in 1961, when the drug was an experimental compound (R. 1625); However, the patient treated had previously undergone a lobectomy, which easily could have affected his response; since then, haloperidol has become probably the most commonly prescribed drug for Tourette syndrome
Joint Commission on Mental Illness and Mental Health (1955–1961)	United States	*Action for Mental Health* (1961); Community Mental Health Centers Act of 1963	*Action for Mental Health* recommended psychiatric deinstitutionalization of the care of persons with mental illness by shifting their care from large mental hospitals into community mental health clinics; deinstitutionalization became a reality with the passage of the Community Mental Health Centers Act of 1963
Popular Media	United States	The Myth of Mental Illness, by Thomas Szasz is published.	The book was a powerful critique of contemporary psychiatric practice and is considered a classic.
Rene Spitz (1887–1974)	Europe, then United States	*The First Year of Life: A Psychoanalytic Study of Normal and Deviant Development of Object Relations*. International Universities Press; 1965	Showed that, when separated from their mothers during the first year of life, some infants exhibited overt depression, characterized by *marasmus* (a pining away of a physically healthy child), which could result in death; Spitz also observed that indiscriminate smiling at moving faces and masks began between the ages of 2 and 6 months, and that smiling at familiar faces occurred only after age 6 months
Classification of mental disorders	United States	DSM-I prepared by the Committee on Nomenclature and Statistics of the APA, G.N. Raines, chair (Mental Hospital Service, Washington, DC, 1952); DSM-II prepared by the Committee on Nomenclature and Statistics of the APA, E. Gruenberg, chair (APA, Washington, DC, 1968); Grob G. Origins of DSM-I: A study in appearance and reality. *Am J Psychiatry*. 1991;148:421	DSM-I replaced several outdated classifications of mental disorders and for the first time provided a glossary of definitions of psychiatric conditions; it was not universally accepted in the United States and most other countries; although DSM-II diagnosed new disease entities, including disorders of childhood and adolescence, its diagnostic methods and criteria were criticized by several leading psychiatrists
Epidemiologic evidence for the genetic transmission of schizophrenia	United States, Denmark	Rosenthal D, Kety SS, eds. *The Transmission of Schizophrenia*. Pergamon Press; 1968:235 (Wender PH, Rosenthal D, Kety SS. A psychiatric assessment of the adoptive parents of schizophrenics), 345 (Kety SS, Rosenthal D, Wender PH, Schulsinger F. The types and prevalence of mental illness in the biological and adoptive families of adopted schizophrenics), and 377 (Rosenthal D, Wender PH, Kety SS, Schulsinger F, Welner J. Schizophrenics' offspring reared in adoptive homes); Kety SS, Wender PH, Jacobsen B, Ingraham LJ, Jansson L. Mental illness in the biological and adoptive relatives of schizophrenic adoptees: Replication of the Copenhagen study in the rest of Denmark. *Arch Gen Psychiatry*. 1994;51:442	A study of 5,483 Danish adoptees found that 507 had been admitted to a psychiatric hospital, and 33 were classified as having schizophrenic disorders; an equal number of matched adoptees who had no history of psychiatric disorders were identified for comparison; of the 150 biologic relatives of the schizophrenic adoptees, 8.7% had a diagnosis of schizophrenia, compared with 1.9% of the 156 biologic relatives of the nonschizophrenic adoptees; that significant difference indicated that genetic factors were important in the transmission of schizophrenia; these statistics were then replicated by statistics from the rest of Denmark; this method of study was later used to show that other psychiatric disorders (bipolar disorder, suicide, psychopathy, felonious behavior, alcoholism, and hyperactivity in children) were caused by genetic factors; it stimulated molecular geneticists to search for the molecular genetic basis of those disorders
Lithium therapy for psychotic excitement	Australia, Denmark, United States	Cade J. Lithium salts in the treatment of psychotic excitement. *Med J Aust*. 1949;36:349; Schou M, Juel-Nielsen N, Stromgren E, Voldby H. The treatment of manic psychoses by the administration of lithium salts. *J Neurol Neurosurg Psychiatry*. 1954;17:250; Schou M. Forty years of lithium treatment. Arch Gen Psychiatry. 1997;54:9; J Schioldann: Mogens Abelin Schou (1918–2005)—Half a century with Lithium. *Hist Psychiatry*. 2007;17:247	After the Australian psychiatrist Cade reported that lithium quieted manic patients, the Danish psychiatrist Schou confirmed that finding in a double-blind study; Schou then developed lithium therapy for manic-depressive illness; because of lithium's toxicity and because many psychiatrists questioned its therapeutic effectiveness, it was not approved until 1970 by the U.S. Food and Drug Administration

William H. Masters (1915–2001), Virginia E. Johnson (1925–)	United States	*Human Sexual Response* (1966), *Human Sexual Dysfunction* (1970)	Both books revolutionized knowledge and attitudes about sex; *Response* showed that the female capacity for orgasm greatly exceeds that of most men, that Freud's theories of clitoral and vaginal orgasm needed revision, and that the capacity for meaningful sexual relationships continues into old age; *Dysfunction* stated that most sexual dysfunctions are amenable to some form of counseling; although that statement needed further verification, it gave hope to patients and sex counselors, and it stimulated the discipline of sex therapy
Organizational Psychiatry/ Diversity, Equity, Inclusion, and Justice	United States	Kenneth Clark becomes first African-American president of the APA in 1966.	
Margaret Mahler (1897–1985)	Central Europe until 1938, then United States	Mahler M, Pine F, Bergman A. *The Psychological Birth of the Human Infant: Symbiosis and Individuation*. Basic Books; 1975; *The Memoirs of Margaret S. Mahler*. Stepansky PE, ed. Free Press; 1988	Delineated the series of stages marking the infant's gradual intrapsychic separation from the mother and its understanding of itself as a distinct individual along with other equally distinct individuals
John Bowlby (1907–1990)	England	*Attachment* (1969), *Separation: Anxiety and Anger* (1973), *Loss: Sadness and Depression* (1980), *A Secure Base: Clinical Applications of Attachment Theory* (1988); *Holmes J. John Bowlby and Attachment Theory*. Routledge; 1993	Bowlby's attachment theory is largely based on the belief that a child's social and emotional development depends on the formation, maintenance, and renewal of attachment and affection as bonds between a child and parent (caregiver); deviant emotional and personality development is attributed to the rejection or disruption of ambivalent attachment bonds; loss, or threat of loss, of the attachment figure is viewed as a principal pathogenic agent; Bowlby furthermore believed that when the attachment behaviors of the child are nurtured by the caregiver, the child can develop a sense of security that then allows safe exploration of the world (environment), a positive sense of self, and good relations with others
Jacques Lacan (1901–1981), "the French Freud"	France	*The Language of the Self: The Function of Language in Psychoanalysis* (1968); *Turkle S. Psychoanalytic Politics: Freud's French Revolution*. Basic Books; 1978; Roudinesco E. Jacques Lacan & Co.: *A History of Psychoanalysis in France, 1925–1985*. University of Chicago Press; 1990	Emphasized language and the need to make contact with the prelanguage period in the unconscious and rejected the standard 50-minute analysis for sessions that were sometimes 10, 5, or even 3 minutes; founded a school of psychoanalysis, described as a return to Freud; at the time of his death, there were reportedly 5,000 Lacanian analysts in France; the most influential figure in French psychiatry; his ideas influenced literature, language, linguistics, economics, and mathematics; he was an integral part of the political left
Heinz Kohut (1913–1981)	Austria until 1940, then United States	*The Analysis of the Self* (1971); *The Restoration of the Self* (1977); The two analyses of Mr. Z. *Int J Psychoanal*. 1979;60:3; Baker H, Baker M. Heinz Kohut's self psychology: An overview. *Am J Psychiatry*. 1987;144:1; Strozier CS. Heinz Kohut: *The Making of a Psychoanalyst*. Farrar, Straus and Giroux; 2001	Originated the psychoanalytic school of self-psychology, which delineated a new group of developmental needs and three new views of transferences: Mirroring, idealizing, and alter ego
Jean Piaget (1896–1980)	Switzerland	Greenspan, Curry J. Piaget's approach to intellectual functioning. In: Kaplan HI, Sadock BJ, eds. *Comprehensive Textbook of Psychiatry*. 6th ed. Williams & Wilkins; 1995	Influenced by his observations of his three children, Piaget found four major stages in the formation of thought from infancy through adolescence, with each a necessary prerequisite for the one that follows: (1) *sensorimotor* (up to 2 years), child learns about space and permanent objects; (2) *preoperational* (2–7 years), verbal symbols are learned so words come to represent objects; (3) *concrete operations* (7–11 years), objects become classified according to similarities; (4) *formal operations* (11 years through end of adolescence), adult logical thinking is used more and more; although there have been criticisms and revisions of these four stages, they will (in the words of Greenspan and Curry) "serve as a basis for understanding social and emotional development and provide a framework for clinical and educational intervention"

Organizational Psychiatry/ Diversity, Equity, Inclusion, and Justice	United States	The Caucus of Gay, Lesbian, and Bisexual Members of the APA is founded in 1975.	The group quickly became an important advocacy group for LGBT mental health issues. In 1985 they changed their name to the Association of Gay and Lesbian Psychiatrists.
NAMI	United States	NAMI: *We Are Family*. Arlington, VA: NAMI; 1979; a 10-page brochure describing the work of NAMI; since then NAMI has published regular annual reports on its work	NAMI, founded in 1979 as an organization of patients with schizophrenia and their family members, has developed new approaches to working with families—counseling, individual family therapy, multifamily therapy groups, and psychoeducational programs—and is probably the most vigorous citizens' group in America that advocates to legislators and the public for the problems of persons with mental illness
Classification of mental disorders / Diversity, Equity, Inclusion, and Justice	United States	Transgender people are classified by the APA as having Gender Identity Disorder	
Sociobiologic psychiatry	United States, Holland	Van Praag H. Sociobiological psychiatry. *Compr Psychiatry*. 1981;22:441	An approach to the treatment of mental illnesses developed by the Dutch psychiatrist Herman van Praag; it analyzes the possibilities and limitations of pharmacotherapy and psychotherapy and their compatibility when used together in the treatment of psychiatric patients in various diagnostic categories
Roger Sperry (1913–1994)	United States	Some effects of disconnecting the cerebral hemispheres. *Science*. 1982;217:1223; Sperry RW, Levi-Montalcini R. In: Finger S, ed. *Minds Behind the Brain: A History of the Pioneers and Their Discoveries*. Oxford University Press; 2000	In 1981, Sperry shared the Nobel Prize in Physiology or Medicine for his work demonstrating that the left brain hemisphere contains the primary speech capacity, that the right hemisphere is involved with short-term memory, and that the two function independently when the connection between them is cut; this work (funded by grants from the NIMH) suggested that such states as autism and delusions sometimes may be caused by disturbances in the connections between the hemispheres
Brief psychotherapy	United States, England	Malan D. *A Study of Brief Psychotherapy*. Plenum Press; 1976; Mann J, Goldman R. *A Casebook in Time-Limited Psychotherapy*. McGraw-Hill; 1982; Davanloo H. *Short-Term Dynamic Psychotherapy*. Jason Aronson; 1980; Sifneos P. *Short-Term Dynamic Psychotherapy Evaluation and Technique*. Plenum Press; 1979; Crits-Cristoph P. The efficacy of brief dynamic psychotherapy: A meta-analysis. *Am J Psychiatry*. 1992; 149:151	Aims at producing insight and personality changes in a patient in a short, cost-limited time frame; it includes limiting time devoted to therapy, focusing on a particular problem, actively involving the therapist, engaging in anxiety-provoking confrontations, past–present link interpretations, and problem solving, recapitulating patient's resistances, emphasizing change and progress, and terminating early
Psychology of women	United States	Miller JB. *Towards a New Psychology of Women*. Beacon Press; 1976; Gilligan C. *In a Different Voice: Psychological Theory and Women's Development*. Harvard University Press; 1982; Levinson D, in collaboration with Levinson J. *The Seasons of a Woman's Life*. Knopf; 1996	Jean Baker Miller, Carol Gilligan, and Daniel Levinson offer different views on women's psychology and the female life cycle
Mental Health Policy		President Bill Clinton signs the Mental Health Parity Act	This act is later amended by President George W Bush in 2008 to close loopholes.
NARSAD	United States	NARSAD activities are reported in *Research* brochures and *Newsletters*	NARSAD was organized in 1986; in 1987, a funding program with 10 grants of $25,000 each was instituted; in 2003, NARSAD reported that in 15 years, it had awarded $144.4 million to fund 2,153 grants to 1,711 scientists at 312 universities and medical research institutions for research in schizophrenia and depression; in 2007 NARSAD announced that it had awarded 23 Distinguished Investigator grants and 222 Young Investigator grants; it has become the largest nongovernment, publicly funded organization that distributes funds for brain disorder research

Changes in the diagnosis and classification of homosexuality	United States	Bayer R. *Homosexuality and American Psychiatry: The Politics of Diagnosis*. Basic Books; 1981; DSM-III (1980); DSM-III-R (1987)	In 1973, in the light of new clinical information and under political pressure from the National Gay Task Force, the APA changed its diagnosis of homosexuality from a disease to a condition that can be considered a disease only if it is subjectively disturbing to the person; although some APA members protested the decision, it was upheld in a 1974 APA referendum; in the DSM-III, homosexuality was diagnosed as *ego-dystonic homosexuality;* in DSM-III-R, the reference to homosexuality was eliminated; some of the factors that influence a person to have a homosexual orientation are known (biology, family cultural environment, intrapsychic psychodynamics), but others remain unknown
Classification of mental disorders	United States	DSM-III (1980); DSM-III-R (1987); Wilson M. DSM-III and the transformation of American psychiatry: A history. *Am J Psychiatry*. 1993;150:399	The DSM-III-R reorganized the entire system of classifying diseases, leading to a more detailed, precise, and clear delineation of symptoms and a more medical and less psychoanalytic view of the symptoms; it was widely read and accepted in the United States and became the common language used by workers in psychiatry; the DSM-III-R was published because data from new studies were inconsistent with some of the previous diagnostic criteria and because of the need to review the criteria (along with systematic descriptions of diseases) for clarity and conceptual accuracy
Organizational Psychiatry	Europe	European Psychiatric Association founded in 1983	
Organizational Psychiatry/ Diversity, Equity, Inclusion, and Justice	United States	Carol Nadelson becomes first woman president of the APA in 1984.	
James Masterson (1926–), Otto Kernberg (1928–)	United States	Masterson J. *Psychotherapy of the Borderline Adult: A Developmental Approach*. Brunner/Mazel; 1976; *Narcissistic and Borderline Disorders: An Integrated Developmental*. Brunner/Mazel; 1981; Masterson J, Klein R, eds. *Psychotherapy of the Disorders of the Self: The Masterson Approach*. Brunner/Mazel; 1988; Kernberg O. *Borderline Conditions and Pathological Narcissism*. Jason Aronson; 1975; Kernberg O. *Severe Personality Disorders: Psychotherapeutic Strategies*. Yale University Press; 1984	Masterson and Kernberg developed two different psychoanalytic concepts (both emphasizing object relations) for understanding, diagnosing, and treating narcissistic and borderline personality disorders
Psychology and neurophysiology of dreams	United States	Hobson JA, McCarley RW. The brain as a dream state generator: An activation-synthesis hypothesis of one dream process. *Am J Psychiatry*. 1977;134:1335; McCarley RW. Dreams: Disguise of forbidden wishes or transparent reflections of a distinct brain state? In: Bilder RM, Lefever FF, eds. *Neuroscience of the Mind: On the Centennial of Freud's Project for a Scientific Psychology*. New York Academy of Sciences; 1998:116; Reiser MF. The dream in contemporary psychiatry. *Am J* Psychiatry. 2001;158:351	The Hobson–McCarley activation–synthesis hypothesis of dream formation, based on neurophysiologic research, postulates that dreams are caused by two reciprocally interacting groups of brainstem neurons; although this hypothesis at first tended to view dreams as products of brain physiology that lacked intrinsic psychological importance (and thus refuted Freud's psychoanalytic theory of dreams), this view has lately changed; Reiser suggested that new research has shown that the psychology of dreams is more complex than has been previously thought and postulated "a preliminary psychobiological concept of the dream process" in which "the mind exploits (for memory and cognitive functions) the special physiologic state of the activated dreaming brain"; Reiser emphasized that this concept is not a "final or complete understanding of the dream process," but that it may serve as a stimulus "for further research to both psychoanalysis and cognitive neuroscience"

Theories on the psychology of love	United States	Bergmann MS. *The Anatomy of Loving*. Columbia University Press; 1987; Sternberg RJ. *The Triangle of Love: Intimacy, Passion, Commitment*. Basic Books; 1987; Sternberg RJ, Barnes ML, eds. *The Psychology of Love*. Yale University Press; 1988; Gaylin W, Person E, eds. Passionate *Attachments: Thinking about Love*. Free Press; 1988; Person E. *Dreams of Love and Fateful Encounters: The Power of Romantic Passion*. WW Norton; 1988	These books, which appeared within 1 year of one another, differed in their accounts of the nature of love; no one dominant theory emerged, yet each book offered new information and ideas on the phenomenology of love
Evaluations of psychotherapy	United States	Bruce R, Staples FR, Cristol AH, Yorkston NJ, Whipple KS. *Psychotherapy versus Behavior Therapy*. Harvard University Press; 1975; Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM. National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Arch Gen Psychiatry*. 1989;48:971	Two methodologically sophisticated studies that concluded that groups of neurotic patients in psychotherapy showed a significant reduction in their symptoms when compared with control groups
Eric Kandel (1929–)	Austria, United States after 1939	Kandel E. Psychotherapy and the single synapse: The impact of psychiatric neurobiological research. *N Engl J Med*. 1979;301:1028; From metapsychology to molecular biology: Explorations into the nature of anxiety. *Am J Psychiatry*. 1983;140:1277; Genes, nerve cells, and the remembrance of things past. J Neuropsychiatry. 1989;1:103; Biology and the future of psychoanalysis: A new intellectual framework for psychiatry revisited. *Am J Psychiatry*. 1999;156:505; *In Search of Memory: The Emergence of a New Science of Mind*. WW Norton; 2006	Showed connections between psychiatry and neurobiology and research on the CNS of the snail *Aplysia;* developed an experimental system and set of conceptual approaches for studying the biologic basis of simple forms of learning and memory and demonstrated that "learning produces changes in neuronal architecture, changes that result from learned alterations in gene expression"; suggested that normal learning, learning of neurotic behavioral patterns, and unlearning of such detrimental behaviors through psychotherapeutic intervention might involve long-term functional and structural changes in the brain that result from altered gene expression; in his 1999 article, Kandel suggested ways of creating a discipline of neurobiology, cognitive psychology, and psychoanalysis that "would forge a new and deeper understanding of the mind" In 2000, the Nobel Prize for Physiology or Medicine was awarded to three men: Kandel for work in *Aplysia* (the second psychiatrist to receive the prize, after Wagner von Jauregg); Arvid Carlsson, a neuropsychopharmacologist at the University of Gothenburg in Sweden, known for his development of the dopamine hypothesis; and Paul Greengard, a neuroscientist at Rockefeller University in New York City, who studied how such neurotransmitters as dopamine communicate their signals at the cellular level. (Andreasen NC. Three living heroes of contemporary psychiatry. *Am J Psychiatry*. 2000;157:1911)
Aaron Beck (1921–)	United States	Beck A. *Cognitive Therapy and the Emotional Disorders*. International Universities Press; 1976; Beck A, Rush AJ, Shaw BF, Emery G. *Cognitive Therapy of Depression*. Guilford Press; 1979; Beck A, Emery G, Greenberg R. *Anxiety Disorders and Phobias: A Cognitive Perspective*. Basic Books; 1985; Beck A, Freeman A. *Cognitive Therapy of Personality Disorders*. Guilford Press; 1990; Weishaar M. Aaron T. Beck. *Key Figures in Counselling and Psychotherapy*. Sage; 1993; Barsky AJ, Ahern DK. Cognitive behavioral therapy for hypochondriasis. A randomized controlled trial. *JAMA*. 2004;291:1464	As formulated by Beck, cognitive therapy is a system of psychotherapy based on a theory of psychopathology and a set of therapeutic principles and techniques that emphasize the reorganization of a person's maladaptive processes of perceiving, thinking, and forming opinions; the study of Barsky and Ahern showed that cognitive-behavioral therapy in six 90-minute intervals "appears to have significant beneficial long-term effects on the symptoms of hypochondriasis." On November 16, 2007 Beck was one of five scientists chosen to receive the Lasker Award for his technique of cognitive therapy; the Award is considered the nation's most prestigious medical awards and carries a prize of $100,000. (*The New York Times*, November 17, 2007, Section N, p. 24)

Karl Menninger (1893–1990), William Menninger (1900–1966)	United States	Faulkner H, Pruitt V, eds. *The Selected Correspondence of Karl A. Menninger*. Yale University Press; 1988; Friedman LJ. *Menninger: The Family and the Clinic*. Knopf; 1990	The Menninger Clinic (founded in 1920 in Topeka, Kansas) became recognized for its innovative approaches to mental health, scientific excellence, and emphasis on public service; William Menninger introduced new, effective treatments for American psychiatric casualties in World War II; Karl Menninger applied psychoanalytic concepts to American psychiatry and wrote three popular and influential psychiatric-psychoanalytic books: *The Human Mind* (1930), *Man Against Himself* (1938), and *Love Against Hate* (1942); in June 2003, the Clinic moved from Topeka to Houston, Texas, where it became partners with Baylor College of Medicine and the Methodist Hospital
Frederick K. Goodwin (1936–), Kay Redfield Jamison (1946–)	United States	Goodwin FK, Jamison KR. *Manic-Depressive Illness*. Oxford University Press; 1990; 2nd ed., 2007	Offered a new diagnosis and classification of manic-depressive illness as a cluster of related illnesses, including bipolar disease, cyclothymia, and recurrent depressive disease; also offered the most comprehensive review of all information on the disease and all aspects of treatment, including medical (acute and prophylactic) and the role of psychotherapy
A memoir of depression	United States	Styron W. *Darkness Visible: A Memoir of Madness*. Random House; 1990	A national bestseller, described by the *Washington Post* as "a striking addition to the notable personal accounts of mental illness"; has been used by therapists as "an invaluable document" to help patients understand the symptoms of their depression. (*The New York Times,* November 12, 2006, Section 4, p. 11)
SSRIs	United States	"The Story of Prozac," Pharmaceutical Manufacturers Association; 1993; Grebb J. Serotonin-specific inhibitors. In: Kaplan HI, Sadock BJ, eds. *Comprehensive Textbook of Psychiatry*. 6th ed. Williams & Wilkins; 1995	SSRIs are antidepressants, generally as effective as previous antidepressants but with fewer adverse effects; the first SSRI, fluoxetine (Prozac), developed by Lilly in the 1970s, has become the most widely prescribed antidepressant in the United States; it was followed by sertraline (Zoloft), developed by Roerig in 1992, and paroxetine (Paxil), developed by SmithKline Beecham in 1993
Jerome Frank (1910–2005), E. Fuller Torrey (1937–)	United States	Frank J. *Persuasion and Healing: A Comparative Study of Psychotherapy*. Johns Hopkins University Press; 1961; 2nd rev. ed., 1973; Frank J, Frank JB. *Persuasion and Healing: A Comparative Study of Psychotherapy*. 3rd rev. ed. Johns Hopkins University Press; 1991; Torrey EF. *The Mind Game: Witchdoctors and Psychiatrists*. Emerson Hall; 1972; Torrey EF: *Witchdoctors and Psychiatrists: The Common Roots of Psychotherapy and Its Future*. Harper & Row; 1986 (a revised edition of *The Mind Game*)	These books offered perspectives on the practices of contemporary American psychotherapy and psychotherapists by delineating how psychic changes are effected in a wide range of therapist–patient interactions, including miracle cures, faith healing, Communist thought control, the so-called placebo effect in medicine, special types of individual and group psychotherapies, and the work of traditional healers in third-world countries and among Mexican and Puerto Rican Americans and Navaho
Neuroimaging	United States	Kenneth Kwong publishes his pioneering work in functional brain imaging in 1992	
Psychiatric epidemiology	United States	Robins LN, Regier DA, eds. *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*. Free Press; 1991	This NIMH Epidemiologic Catchment Area study, the largest and most sophisticated study of psychiatric epidemiology ever undertaken in the United States, showed that more than 30% of individuals have at least one psychiatric disorder (and many have more than one); that alcoholism and phobic disorder are the most common illnesses, followed by depression; that the uneducated, unmarried, and divorced are more vulnerable; that, more important than race or economic status as a predictor of most illness, the young are more vulnerable than their elders; the study permitted follow-up and evaluation of changes (both overall and in specific areas) to show how particular services are used or not used; it thus continues to provide data for an ongoing discussion of mental health needs and future psychiatric research projects

Forensic psychiatry	United States	Bursztajn H, Scherr A, Brodsky A. The rebirth of forensic psychiatry in light of recent historical trends in criminal responsibility. *Psychiatr Clin North Am*. 1994;1(3):611; Prentice S. A history of subspecialization in forensic psychiatry. *Bull Am Acad Psychiatry Law*. 1995;23(2):195	Changing concepts of criminal responsibility in mid-20th century America, and interactions between the American Academy of Forensic Sciences and the American Academy of Psychiatry and the Law, resulted in forensic psychiatry's becoming officially recognized as a subspecialty by the American Board of Medical Specialties on September 17, 1992, under the designation of "Added Qualification in Forensic Psychiatry"
Epidemic of AIDS and of persons infected with HIV (1981–)	Worldwide	Grmek M. *History of AIDS: Emergence and Origin of a Modern Pandemic*. Princeton University Press; 1990; Berridge V, Strong P, eds. *AIDS and Contemporary History*. Cambridge University Press; 1993; Feldman D, Miller J, eds. *The AIDS Crisis: A Documentary History*. Greenwood Press; 1998; Bayer R, Oppenheimer G, eds. *AIDS Doctors: Voices from the Epidemic*. Oxford University Press; 2000; Treisman G, Angelino F. *The Psychiatry of AIDS: A Guide to Diagnosis and Treatment*. Johns Hopkins University Press; 2004; Mayer K, Pizer H, eds. *The AIDS Pandemic: Impact on Science and Society*. Elsevier Academic Press; 2005	Psychiatrists are called on to manage individuals who are at high risk for HIV infection or who are HIV infected and to treat the psychiatric disorders that result from or are comorbid with HIV infection; these disorders include dementia, delirium, anxiety, and disorders of cognition, adjustment, mood, and sleep; because AIDS is largely caused by sexual contact and IV drug use, psychiatrists and mental health professionals need to teach the techniques of safe sex and safe drug use and work with addicts to withdraw from drug; the global incidence of AIDS is steadily increasing
Classification of mental disorders	United States	*American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Text rev. American Psychiatric Association; 2000 (DSM-IV-TR)	The DSM-IV, probably the most ambitious undertaking in the history of American psychiatric nosography, was based on 6 years of collecting and analyzing relevant information and making field tests of proposed changes in diagnosis; it continued the atheoretical approach to causes, made fewer diagnostic changes than the DSM-III, systematically described each disorder in terms of its associated features, and replaced the DSM-III-R as the diagnostic text used by workers in psychiatry; the DSM-IV-TR contained updated information about the associated features, prevalence, course, and familial pattern of some of the mental disorders that had been described in the DSM-IV; it also gave more comprehensive diagnostic criteria for personality change due to a general medical condition, paraphilias, and tic disorders than those given in the DSM-IV
Clozapine, risperidone, and olanzapine in the treatment of schizophrenia	United States, Canada, Europe, England, South America, South Africa	Kane J, Honigfield G, Singer J, Meltzer H, and the Clozaril Collaborative Study Group: Clozapine for the treatment-resistant schizophrenic. A double-blind comparison with chlorpromazine. *Arch Gen Psychiatry*. 1988;45:789; Kane J, Marder SR. Clozapine benefits and risks. *Schizophren Bull*. 1994;20:23; Ayd FJ Jr. Risperidone (Risperdal): A unique antipsychotic. *Int Drug Ther Newslett*. 1994;29:5; Marder SR, Meibach RC. Risperidone in the treatment of schizophrenia. *Am J Psychiatry*. 1994;151:825; Kane J. Risperidone. *Am J Psychiatry*. 1994;151:802; Beasley C, Tollefson G, Tran P, Satterlee W, Sanger T. Olanzapine versus placebo and haloperidol acute phase results of the North American double-blind olanzapine trial. *Neuropsychopharmacology*. 1996;14:111; Tollefson G, Kuntz A. Review of recent clinical studies with olanzapine. *Br J Psychiatry*. 1999;37:30; Volavka J, Czobor P, Sheitman B, Lindenmayer JP, Citrome L. Clozapine, olanzapine, risperidone, and haloperidol in the treatment of patients with chronic schizophrenia and schizoaffective disorder. *Am J Psychiatry*. 2002;159:255; Grilly J. The history of clozapine and its emergence in the U.S. market: A review and analysis. *Hist Psychiatry*. 2007;18:39	Clozapine and risperidone have been used successfully since 1989 and 1994 to treat both positive and negative schizophrenic symptoms that have been resistant to other treatments; these drugs each offer the possibility for real and continued improvement of individuals who are persistently psychotic; olanzapine has been in use since September 1996 and is effective in treating positive and negative schizophrenic symptoms; the 2002 study by Volavka et al. on 157 hospital in-patients who had a diagnosis of chronic schizophrenia or schizoaffective disorder showed that clozapine, risperidone, and olanzapine (but not haloperidol) resulted in statistically significant improvements in the positive and negative symptoms of hospital patients: "The overall pattern of results suggests that clozapine and olanzapine have similar general antipsychotic efficacy and that risperidone may be somewhat less effective. Clozapine was the most effective in treatment for negative symptoms. However, the differences were small."

Publications in psychohistory and psychobiography	United States, England	Lawton H. *The Psychohistorian's Handbook*. Psychohistory Press; 1988; Elovitz P, ed. Clio's Psyche. Understanding the "Why" of Culture, Current Events, History, and Society (1994) (a quarterly periodical published by the Psychohistory Forum, 627 Dakota Trail, Franklin Lakes, NJ 07417); Sabbadini A, ed. Psychoanalysis and History. Vol. 1, No. 1. Artesian Books; 1998–1999 (since its first issue, this *History* has appeared at regular intervals of twice a year; since 2005 edited by John Forrester, and since 2007 published by Edinburgh University Press); Szaluta J. *Psychohistory: Theory and Practice*. Peter Lang; 1999; Elovitz P. Psychoanalytic scholarship on American Presidents. *Annu Psychoanal.* 2003;31:135; Schultz WT, ed. *Handbook of Psychohistory*. Oxford University Press; 2005	Lawton defined *psychohistory* "as the interdisciplinary study of why man has acted as he has in history, prominently utilizing psychoanalytic principles" (no uniformly accepted definition exists); examined different psychohistorical writings, concepts, and methods of doing interdisciplinary research (most of which date from the appearance of Erikson's *Young Man Luther* in 1958); and called attention to the two best-established journals in the field, started in the 1970s: the *Psychohistory Review* (stopped publication in 1999) and the *Journal of Psychohistory; Clio's Psyche* has short, concise studies by various authors on greatly varied subjects (including the personalities of famous individuals, presidential character, nationalism, ethnic conflicts and the Holocaust, AIDS, movies, psychological issues, psychoeconomics, and the environment); many studies aim to combine primary research with psychological insights; notable interviews with distinguished psychohistorians (in his interview with *Clio,* psychohistorian Peter Loewenberg commented, "I'd like to get away from the idea of applying psychoanalysis to history because I think the integration of psychodynamic perceptions with historical conceptualization should take place at the moment that the historian contacts the data or the archives. Both history and psychoanalysis are fundamentally historical enterprises—they're models of explanation.") *Psychoanalysis and History* has two stated aims: (1) "historiographic studies of the psychoanalytic movement itself, as well as biographical research on aspects of the lives and works of prominent psychoanalysts. Only one century old, psychoanalysis has an already rich and articulated history, having undergone complex external and internal vicissitudes and having developed its identity in most diverse ways in different geographical regions. We intend … to present as yet unpublished documents … which will throw new light on a number of important aspects of our discipline" and (2) "We shall also attempt to apply psychoanalytic ideas—for instance the concept of unconscious processes, of developmental stages, or of relevant defence mechanisms such as repression, splitting, projective identification and 'deferred action' (Nachtraglichkeit)—to the interpretation of historical events" Szaluta's book was the first text to present both the theory and practice of psychohistory; the contributions of Freud and post-Freudians were delineated, the pros and cons of psychohistory examined, and major works of psychohistory discussed; Szaluta emphasized how psychohistory differs from traditional history and how it can be insightful in the study of the individual, the family, and the group Elovitz discussed how, in almost three decades of research on presidential psychobiography, he was influenced by psychoanalysis to develop a method of "disciplined subjectivity" in which he applied the "nonjudgmental manner of the clinician … to the politician"; he studied his subject's childhood, personality, family background, life crises, defense mechanisms, and patterns of success and failure; throughout the entire process, he carefully monitored countertransference feelings; as a result of following this method, he claimed to understand both the pressures of office and the personalities of presidents and to have insights into "understanding the contradictory desires, the ambivalence, of both the leader and the led" A *Handbook of Psychohistory* offers articles on psychohistory and psychobiography that show a variety of different psychological approaches

Paul Wender (1934–)	United States	*Attention-Deficit Hyperactivity Disorder in Adults*. Oxford University Press; 1998	As a result of his 25 years of studying ADHD in adults, Wender gave new evidence suggesting that ADHD is genetically transmitted and that approximately one-third of children with ADHD reach adulthood with persisting symptoms of the disorder; he also reviewed what was known about ADHD's symptoms, its life course, and the value of treatments with various drugs
Computers and the Internet	United States	The APA has published two books on the use of computers in psychiatry: *Lieff J. Computer Applications in Psychiatry*. American Psychiatric Publishing; 1987; and Chan CH, Luo JS, Kennedy RS. *Concise Guide to Computers in Clinical Psychiatry*. American Psychiatric Publishing; 2002; Miller MJ. Computer-based testing of the psychiatric patient. In: Sadock BJ, Sadock VA. *Comprehensive Textbook of Psychiatry*. 7th ed. Lippincott Williams & Wilkins; 2000	Computers have been applied to practically every area of psychiatry, including recording patient symptoms, facilitating the transfer of records, tracking patients in the community, and establishing databases; they have been used by insurance companies and managed care companies for accounting and monitoring patient care, and, since the 1970s, they have been used with increasing frequency in giving patients psychological tests
		Two articles about the Internet are Alessi NE. The Internet and the future of psychiatry. *Am J Psychiatry*. 1996;153:861; and Rajendran PR. The Internet: Ushering in a new era of medicine. *JAMA*. 2001;285:804	The Internet is a computer communication network that transmits information quickly and easily on a global scale, facilitating contacts between different psychiatrists, and between psychiatrists and patients, and enabling patients to obtain rapid information on the professional qualifications of the psychiatrist who is treating them and the psychiatric medications the psychiatrist has prescribed for them Although the Internet has given new information to persons who previously had not been able to obtain this information, it is information that sometimes needs to be used with caution because it can be the result of covert advertising by individuals or companies and has not been scientifically evaluated
Discoveries of genetic variations as causes of mental illnesses	United States, Canada, England, Europe, Australia	Oberlé I, Rousseau F, Heitz D, Kretz G, Davys D. Instability of a 550–base pair DNA segment in fragile X syndrome. *Science*. 1991;252:1097; Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J. Fragile X genotype characterized by an unstable region of DNA. *Science*. 1991;252:1179; Ying-Hui Fu, Kuhl DPA, Pizzuti A, Pieretti M, Sutcliffe JS. Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. *Cell*. 1991;67:1047; Schellenberg GD, Bird TD, Wijsman EM, Orr T, Anderson L. Genetic linkage of evidence for a familial Alzheimer's disease locus on chromosome 14. *Science*. 1992;258:668; Mullen M, Houlden H, Windelspecht M, Fidani L, Lombardi C. A locus for familial early-onset of Alzheimer's disease on the long arm of chromosome 14, proximal to the α-1-antichymotrypsin gene. *Nat Genet*. 1992;2:340; The Huntington's Disease Collaborative Research Group: A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosome. *Cell*. 1993;72:971; Amir RE, Van den Veyver IB, Wan M, Tran CQ, Franke U. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet*. 1999;23:185	During 1991–1993, genetic variations were identified as causes for the fragile X syndrome, a common cause of inherited mental retardation; familial early onset of Alzheimer disease; and Huntington disease; in 1999, it was discovered that the MECP2 gene causes Rett syndrome, a pervasive developmental disorder included in the DSM-IV classification

		Hyman S, Moldin S. Genetic science and depression: Implications for research and treatment. In: Weissman MM, ed. *Treatment of Depression: Bridging the 21st Century*. American Psychiatric Association; 2001:98; Andreasen NC. Schizophrenia. Chapter 8. In: *Brave New Brain: Conquering Mental Illness in the Era of the Genome*. Oxford University Press; 2001:199	The 2001 article by Hyman and Moldin stated that, although previous studies had shown that heredity plays a role in the transmission of depression, "a single major gene is an unlikely contributor to etiology; rather, multiple genes of small effect contribute with nongenetic factors in producing vulnerability to the illness"; in the same year, Andreasen observed, "Most experts now think that schizophrenia is clearly multifactorial, involving multiple genes, and possibly even different genes in different individuals, as well as many nongenetic or environmental influences. The fact that multiple genes are probably involved is the main reason why the various reports that 'the schizophrenia gene has been found on chromosome 5' (or 11 or 22 or elsewhere) have not been consistently repeated"
Three books on psychotherapy	United States	Jackson SW. *Care of the Psyche: A History of Psychological Healing*. Yale University Press; 1999; Sabo AN, Havens L, eds. *The Real World Guide to Psychotherapy Practice*. Harvard University Press; 2000; Hersen M, Sledge WH, eds. *Encyclopedia of Psychotherapy*. 2 Vols. Academic Press; 2002	Jackson's *History* is notable for its delineation of the roles of healer and sufferer throughout history; the essays by Sabo, Havens, and others suggest how, through forming relationships with patients, it is possible to do some effective psychotherapy and to counter the constraints of managed care; the *Encyclopedia* gives concise articles, written by experts, on the major psychotherapies that are now in use, along with studies of the efficacy of each of these therapies
Education of psychiatrists	United States	Luhrmann TM. *Of Two Minds: The Growing Disorder in American Psychiatry*. Knopf; 2000	Depicts contemporary American psychiatry as divided between two views of mental illness: A biomedical model, consisting of symptoms treated by psychopharmacology, and a psychodynamic model, consisting of emotional conflicts, treated by psychotherapy; although Luhrmann states that the use of both models together is necessary in treating patients, she observes that "a combination of socio-economic forces [managed care] and ideology is driving psychotherapy out of psychiatry"; she also observes that this would be bad for psychiatrists and patients and encourages "a way of thinking about mental illness that can strip humanity from its sufferers"
Theory and treatment of depression	United States, England, Switzerland	In a March 1999 New York City meeting of the American Psychopathological Association, leading clinical scientists gave papers on the topic Treatment of Depression, all of which were published as in Weissman MM, ed. *Treatment of Depression: Bridging the 21st Century*. American Psychiatric Association; 2001	The papers ranged widely in discussing different aspects of depression; Charles Nemeroff and Michael Owens, in "Contribution of modern neuroscience to developing new treatments for psychiatric disorders," reviewed the evidence that CRF is hypersecreted in depressed patients, that "untoward early life experience is associated with a persistent CRF gene expression and CRF neuronal activity, which may underlie the vulnerability of individuals exposed to child abuse or neglect to develop depression in adulthood," and that the recent use of a CRF antagonist for the treatment of major depression has been successful; Steven Hyman and Steven Moldin, in "Genetic science and depression: Implications for research and treatment," delineated the growth in knowledge from the Human Genome Project, the interactions of different genomes with their environments and protein products, *proteomics,* and then suggested that "the greatest benefit from genetic research is likely to come in the development of platforms for drug discovery and for understanding the genetic contributions to drug action"

			Other papers gave overviews of various treatments; in the United States, there are 20 FDA-approved antidepressants, which in general are "moderately effective," although none of them cures depression, and "a significant number of people either do not respond or respond only partially"; there are two new physical therapies—transcranial magnetic stimulation and vagus nerve stimulation—that remain to be evaluated; there is evidence that cognitive or behavioral therapy each is effective in the treatment of clinical depression; in an epilogue, Weissman observed, "treatment for depression is more targeted, faster, and with fewer side effects and is aimed at the underlying pathophysiology of depression"
Neuroimaging	United States	Andreasen NC. Mapping the mind. In: *Brave New Brain: Conquering Mental Illness in the Era of the Genome*. Oxford University Press; 2001:130; Morihisa JM, ed. Advances in Brain Imaging. American Psychiatric Association, 2001	Beginning in the 1970s and 1980s, techniques of CT and MRI showed how the brain was physically diminished in the dementias and other organic illnesses and how, in some cases of schizophrenia, the ventricles were enlarged; after this discovery, what were known as "functional techniques of neuroimaging"—including functional MRI, single photon emission computed tomography, and PET—were applied to studying changes in brain activity (blood flow) in mental illness; although Andreasen stated that most of these techniques are still research tools, she also observed that in some illnesses "PET permits us to study the relationship between drug dose and clinical response based on a measurement of the levels and activity in the brain"
The effects of traumatic events on individuals	United States, England, France, Germany, and Italy	Micale MS, Lerner P, eds. *Traumatic Pasts, Psychiatry, and Trauma in the Modern Age, 1870–1930*. Cambridge University Press; 2001	A collection of essays by scholars from different disciplines, describing how individuals in several countries suffered medical and psychological disturbances from the effects of railroad accidents, accidents at work, remembering traumatic events in their lives, and the aftermath of shell shock in World War I; in their essay "Trauma, psychiatry, and history," Lerner and Micale commented that there is no "single, uniform, transhistorically valid concept of psychological trauma," and that although it is "impossible to write a single, unilinear history of trauma," it is possible to write case histories of how individuals experienced trauma and to collate these experiences with the ideas about the nature of trauma that were held by scientists and social groups and that were dominant in the cultures in particular moments in the past
Diagnosis and treatment of borderline personality disorder	United States	Gunderson JG. *Borderline Personality Disorder: A Clinical Guide*. American Psychiatric Association; 2001; American Psychiatric Association. *Practice Guideline for the Treatment of Patients with Borderline Personality Disorder*. American Psychiatric Association; 2001; Oldham JM. A 44-year-old woman with borderline personality disorder. *JAMA*. 2002;287:1031	Gunderson' *Guide* and the APA's *Practice Guideline* recommend combinations of long-term psychotherapy (often consisting of cognitive therapy or psychodynamic psychotherapy) and medications that are targeted for specific symptoms, and include low-dose antipsychotics and SSRIs; Oldham observed that, after this treatment, "over time, most patients with borderline personality disorder show substantial improvement"
September 11, 2001, terrorist attacks on the World Trade Center and the Pentagon	United States	Kleinfield NR, Connelly M: 9/11 still strains New York psyche. Poll finds widespread unease and lingering fear of terror. *The New York Times*, September 8, 2003, Section A, pp. A1–A2; Sheehy G: Middletown America. *One Town's Passage from Trauma to Hope*. Random House; 2003; Cancelmo JA, Tylim I, Hoffenberg J, eds. *Terrorism and the Psychoanalytic Space: International Perspectives from Ground Zero*. Pace University Press; 2003	The 9/11 terrorist attacks were notable for their amalgam of violence, loss, and death and for the great variability of the mental and physical reactions that followed from these attacks; whereas some individuals have recovered and formed what Gail Sheehy describes as a "new life," others have suffered from persisting episodes of PTSD, anxiety, depression, physical symptoms, and fears that there will be new terror attacks; 5 years after the attack two-thirds of New Yorkers said they were still "very concerned" about another attack on their city (*The New York Times*, September 7, 2006)

Overviews of psychoanalysis	United States	Erwin E, ed. *The Freud Encyclopedia: Theory, Therapy, and Culture*. Routledge; 2002; Zaretsky E. *Secrets of the Soul: A Social and Cultural History of Psychoanalysis*. Knopf; 2004	The *Encyclopedia* consists of more than 200 entries by Freud scholars, which include biographical sketches of Freud (with an entry on his family) and of individuals who contributed to the development of psychoanalysis; entries on each of Freud's famous case histories; evaluations of Freud's theories of anxiety and defense, drives, the unconscious and repression, neuroses, and narcissism; the influence of psychoanalysis on such humanistic disciplines as anthropology and cultural studies, history and biography, literary criticism, film studies, and philosophy, as well as on biology and neurobiology; and the history and current state of psychoanalysis in various countries in Europe, the Americas, and Africa, as well as in Great Britain, Australia, Japan, Korea, China, India, and the Philippines *Secrets of the Soul* gives kaleidoscopic overviews of how Freud's ideas were influenced by the social, political, and cultural climates in which these ideas originated; Zaretsky argues that "the idea of a *personal* unconscious was new" and "gave expression to possibilities of individuality, authenticity, and freedom that had only recently emerged" and that then played an important role in "the modernism of the 1920s, the English and American welfare states of the 1940s and 1950s, the radical upheavals of the 1960s, and the feminist and gay-liberation movements of the 1970s"; at the same time, in ways that were complex and paradoxical, psychoanalysis became "a fount of antipolitical, antifeminist, and homophobic prejudice" and led to the creation of other psychotherapies and movements that in many cases replaced it; Zaretsky's book is a pioneering attempt to delineate the historical impact of psychoanalysis
Psychosomatic medicine	United States	Blumenfield M, Strain J, eds. *Psychosomatic Medicine*. Lippincott Williams & Wilkins; 2006	In 2003, the American Board of Psychiatry and Neurology granted a psychiatric subspecialty status to psychosomatic medicine, and in 2005, the first board examination in psychosomatic medicine was held; the book edited by Blumenfield and Strain consists of a collection of articles giving comprehensive accounts of psychosomatic symptoms in medical, psychiatric, and neurologic illnesses and how these symptoms can be treated
Confinement of the mentally ill in asylums	South Africa, Switzerland, France, Canada, Australia, Germany, United States, Japan, Argentina, Mexico, India, Nigeria, Ireland, and England	Porter R, Wright D, eds. *The Confinement of the Insane. International Perspectives, 1800–1965*. Cambridge University Press; 2003	This volume brings together 14 original research papers by different scholars, each of whom examines the social forces that influenced the confinement of the insane in asylums in different countries during different periods of time in the 19th and 20th centuries; summing up the contents of the papers, Porter concluded that the history of the growth of asylums "proves more complex than it might seem," revealing that "the asylum was neither just a site for care and cure, nor just a convenient place for locking up inconvenient people ('custodialism'). It was many things all at once. And far from being a weapon securely under the control of the [psychiatric] profession, or the state, it was a contested site, subject to continual negotiations among different parties, including families and the patients themselves"

Treatment of schizophrenia	United States	American Psychiatric Association: *Practice Guideline for the Treatment of Patients with Schizophrenia*. 2nd ed. American Psychiatric Association; 2004	Recommendations that are intended to help psychiatrists in assessing and treating adult patients with schizophrenia in their acute phase, treatment stabilization phase, and stable phase; each recommendation receives one of three categories of endorsement: Recommended with substantial confidence, recommended with moderate confidence, and may be recommended on the basis of individual circumstances; although the *Practice Guideline* is based on synthesis of "current scientific knowledge and rational clinical practice," it is emphasized that it is only a guideline, that it should not be construed to serve as a standard of medical care, and that "the ultimate judgment regarding a particular clinical procedure or treatment plan must be made by the psychiatrist in light of the clinical data presented by the patient and the diagnosis and treatment options available"
College Mental Health	United States	Kadison R, Geronimo I. *College of the Overwhelmed: The Campus Mental Health Crisis and What to Do about It*. Jossey-Bass; 2004	Delineates the mental pressures faced by college students as they confront academic requirements, explore sexuality, and form new interpersonal relationships; as a result of these pressures, "since 1988, the likelihood of a college student suffering depression has doubled, suicidal ideation has tripled and sexual assaults have quadrupled"; on the second half of the book, in a chapter entitled "For students only," Kadison and Geronimo offer proactive ways for students to maintain mental balance and how to seek help if things go wrong
A suppressed psychiatric scandal	United States	Scull A. *Madhouse: A Tragic Tale of Megalomania, and Modern Medicine*. Yale University Press; 2005	The madhouse Scull writes about was the Trenton State Hospital in New Jersey from 1907 to 1930, when Henry Cotton was its Superintendent; believing that infections of various body parts were a cause of psychiatric illness, Cotton treated psychiatric patients by surgical extirpation of these parts, which were mainly teeth and tonsils, but which also included stomachs, gallbladders, colons, testicles, and ovaries; when Adolf Meyer had the young psychiatrist Dr. Phyllis Greenacre investigate Cotton's treatments, Greenacre reported that these treatments resulted in deaths, disabilities, and questionable improvement of psychiatric symptoms; however, Meyer suppressed her report; Cotton was demoted from Superintendent to Director of Research; a public scandal was thus averted, and knowledge of Cotton's operations remained unknown for many years until Scull, after doing research based on documents and the recollections of several individuals, uncovered what had been unknown and published his book
Use of amphetamines as antidepressant drugs	United States	Rasmussen N. Making the first antidepressant: Amphetamine in American medicine, 1929–1950. *J Hist Med Allied Sci*. 2006;61:288	Shows how during the 1940s and 1950s amphetamine was a very commonly prescribed remedy for depression and related neurotic conditions, used by both general practitioners and psychiatrists; in the 1960s, although amphetamine was used in a variety of conditions, it largely stopped being used in most psychiatric cases, being replaced by the minor tranquilizers and the MAOI and tricyclic antidepressants
Recollections of treating schizophrenia	United Sates	Seeman M. Forty-five years of schizophrenia: Personal reflections. *Hist Psychiatry*. 2006;17:363	Seeman recollected that, in the 45 years since she has been treating schizophrenia, "although there have been many advances over this period, understanding of schizophrenia continues to be elusive, and treatments remain imperfect. Therefore, as perspectives shift and the ground moves beneath us, the psychiatric profession need to anchor itself firmly in the stories our [schizophrenic] patients tell us." Whether these stories "be of symptoms, of loss or of survival."

Mental health policy	United States	Frank R, Glide S. *Better But Not Well: Mental Health Policy in the United States Since 1950*. Johns Hopkins University Press; 2006	The mentally ill have improved due to the development of better-tolerated medications (selective serotonin reuptake inhibitors compared with tricyclic antidepressants) and the development of policies such as Medicare, Medicaid, Social Security Disability Insurance, and the expansion of private health insurance; however, the majority of the mentally ill are impoverished, and there has been no increase in their rates of employment; in addition, during the last 50 years their rates of homelessness have more than doubled, and incarcerations have gone up fivefold
George Mora (1923–2006), and his work in the history of psychiatry	Italy and after 1953 the United States	Works on the history of Italian psychiatry, a wide range of medical and psychiatric topics, a translation of Johann Weyer's *De praestigiis daemonum* (1991), and with Jeanne Brand, *Psychiatry and Its History: Methodological Problems in Research* (1970)	For a generation Mora was the foremost American authority on the history of psychiatry, as well as the author of the sections of the history of psychiatry for the first four editions of the *Comprehensive Textbook of Psychiatry*
Mood disorders	United States	In 2006–2007 two textbooks on mood disorders were published: Stein D, Kupfer D, Schatzberg A, eds. *The American Psychiatric Publishing Textbook of Mood Disorders*. American Psychiatric Publishing; 2006; Goodwin FK, Jamison KR. *Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression*. 2nd ed. Oxford University Press; 2007 (1st ed., 1990)	Each of these books gave detailed information on the latest knowledge of the causes, symptoms, and treatments of mood disorders
Albert Ellis (1913–2007)	United States	Ellis published 75 books, including *Overcoming Procrastination, How to Live With a Neurotic, Sex Without Guilt,* and *A Guide to Rational Living*	By late 1953, Ellis had begun developing Rational Emotive Behavior Therapy (REBT), a mode of therapy that aimed at helping individuals make psychological changes by teaching them to challenge their self-defeating thoughts; in 1960, he founded the Albert Ellis Institute to promulgate his work; he continued the shift of psychotherapy from psychoanalysts and from physicians to clinical psychologists and psychiatric social workers, which had been begun by Carl Rogers; in 2003, the American Psychological Association named him the second most influential psychologist of the 20th century, second only to Rogers
Diversity, Equity, Inclusion, and Justice	United States	American Psychiatric Association issues their official position supporting the care and rights of transgender and gender nonconforming individuals.	
Mental Health Policy	United States	Mental Health Parity and Addiction Equity Act of 2008. "Fact Sheet": The Mental Health Parity and Addiction Equity Act of 2008 (MHPAEA). Accessed November 19, 2015. Available at http://www.dol.gov/ebsa/newsroom/fsmhpaea.htm	Enacted in 2008 as part of the Affordable Care Act to address loopholes created by the initial Mental Health Parity Act of 1996. The act (1) includes mental health and substance use disorder benefits in the *Essential Health Benefits;* (2) applies federal parity protections to mental health and substance use disorder benefits in the individual and small group markets; and (3) provides more Americans with access to quality health care that includes coverage for mental health and substance use disorder services.
Organizational Psychiatry/ Diversity, Equity, Inclusion, and Justice	United States	Saul Levin named CEO of the APA in 2013.	This makes Dr. Levin the first openly gay person to lead a national medical specialty society in the United States.

Classification of Mental Disorders	United States	American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* 5th ed. American Psychiatric Association; 2013	In 2013, this revision to the DSM-IV was published, with several important changes: Aspergers and Autism disorders were subsumed under the term *autism spectrum disorders;* criteria were added to ADHD to facilitate application across the lifespan; symptoms must now be present before age 12 (vs. age 7 in DSM-IV); subtype specifiers of schizophrenia were eliminated due to their limited diagnostic stability, low reliability, and poor validity. In bipolar disorder, a "with mixed features" specifier has been added; Addition of disruptive mood dysregulation disorder and premenstrual dysphoric disorder; bereavement exclusion now omitted from criterion of MDD; OCD and related disorders are a new diagnostic category and include hoarding disorder excoriation (skin-picking) disorder; adjustments made to criterion for trauma and stress-related disorders; creation of gender dysphoria as a new diagnostic class. Substance disorders expanded to include gambling disorder.
Mentally ill and gun violence	United States	Metzl, Johnathan M., and Kenneth T. Machleish. "Mental Illness, Mass Shootings, and the Politics of American Firearms." *Am J Public Health.* 2015: 240–249	There is increasing focus on mental illness and gun violence in the wake of several mass shootings, which reflects decades-long history of more general debates in psychiatry about guns, violence, and mental competence. Social and political rhetoric following these violent events strengthen implications that mental illness caused these events or earlier detection of these problems might have prevented them. Many states reactively enacted legislation requiring mental health clinicians to report "dangerous patients" to local officials as a way to prevent these people from obtaining firearms. Many argue that the focus should be on prevention efforts that will have a boarder impact on communities such as reexamination of gun laws and improving mental health resources.
Changes in marijuana legislation	United States (23 states)	Wilkinson, Sauel T, Stephanie Yarnell, Rajiv Radhakrishnan, Samuel A. Ball, and Deepak Cyril D'souza. "Marijuana Legalization: Impact on Pyschicians and Public Health." *Ann Rev Med.* 2014	Between 1996 and 2015, 23 states passed laws legalizing marijuana use for medical purposes after numerous studies revealed its therapeutic potential for multiple medical conditions including epilepsy, cachexia, and chronic pain. Several states have legalized small amount of marijuana for recreational use reflecting a growing shift in public opinion regarding the safety or marijuana. Marijuana use and sale remains illegal under federal law and it is still listed as schedule I substance. Many states are calling for federal decriminalization of marijuana but the debate rages on given remaining concerns about legalization and public health effects. Despite efforts to reschedule marijuana, it continues to be classified as a schedule I substance.
United States Drug Enforcement Agency	United States	Wing N. (2015, March 18). DEA Approves Study of Psychedelic Drug MDMA In Treatment of Seriously Ill Patients. Retrieved November 17, 2015, from http://www.huffingtonpost.com/2015/03/18/deadmda-study_n_6888972.htm	MDMA was classified as a schedule I substance in 1985. In the years since, recent studies have shown that the drug can be effective for mental disorders. DEA approved study of MDMA in the treatment of anxiety in people with life-threatening illness. This comes after studies showing that MDMA can be helpful for disorders like PTSD in a supervised treatment setting.
Physician-assisted suicide	United States	Dyer, Owen, Caroline White, and Aser Garcia Rada. "Assisted Dying: Law and Practice around the World." Bmj (2015): Payne, Kate. "Consider Assisted Suicide." Tennessee Nurse 78.3 (2015): n. pag. Web 17 November 2015 Kopelman, Loretta M. Physician- assisted Suicide: What Are the Issues? Kluwer Academic Publishers, 2001	Medical professional codes have long prohibited physicians involvement in assisting a patient's suicide. However, despite ethical and legal prohibitions, calls for the liberalization of this ban have grown in recent years. Oregon's law permitting physician-assisted dying passed in 1997 is the oldest in the world. Mentally competent and physically able adults who have 6 months or less to live can get prescription for sedative medications they can take to end their lives. As of 2015, several states have enacted similar legislation, but it still remains illegal in the majority of the country.

The Human Connectome Project	United States	Hagmann P, Kurant M, Gigandet X, Thiran P, Wedeen VJ, Meuli R, Thiran, JP. Mapping human whole-brain structural networks with diffusion MRI. *PLoS ONE*. 2;e597:2007. Sporns O, Tononi G, Kotter R. The human connectome: A structural description of the human brain. *PLoS Comput Biol*. 1;e42:2005.	The Human Connectome Project aims to map all of the connections in the human brain in an effort to understand how cognitive processes emerge from their morphologic substrates. The project was launched in 2009 as part of broader efforts to increase neuroscience research. The human brain is highly complex and researchers face numerous challenges in data collection, mapping neuronal pathways, and correlating functional connectivity with anatomic connectivity. Data generated from the project could offer significant contributions to research in complex neurologic and psychological disorders.
Diversity, Equity, Inclusion, and Justice	United States	The APA issues its "Apology to Black, Indigenous and People of Color for Its Support of Structural Racism in Psychiatry."	
Classification of Mental Disorders	United States	The American Psychiatric Association publishes a text revision of DSM-5 (DSM-5-TR) in 2022	

ADHD, attention-deficit/hyperactivity disorder; AIDS, acquired immune deficiency syndrome; APA, American Psychiatric Association; CNS, central nervous system; CRF, corticotropin-releasing factor; CT, computed tomography; DSM, *Diagnostic and Statistical Manual of Mental Disorders;* ECT, electroconvulsive therapy; FDA, U.S. Food and Drug Administration; HIV, human immunodeficiency virus; MAOI, monoamine oxidase inhibitor; MMPI, Minnesota Multiphasic Personality Inventory; MRI, magnetic resonance imaging; NAMI, National Alliance for the Mentally Ill; NARSAD, National Alliance for Research on Schizophrenia and Depression; NIMH, National Institute for Mental Health; NMHA, National Mental Health Association; OCD, obsessive-compulsive disorder; PET, positron emission tomography; PTSD, posttraumatic stress disorder; SSRI, selective serotonin reuptake inhibitor; TR, text revision.

This table was conceived and developed by Ralph Colp Jr, M.D., who contributed to many editions of this book and who died in 2008. It was updated by several of the editors B. Sadock with the assistance of Kristel Carrington, and then by Robert Boland.

Appendix B

Glossary of Psychiatry and Psychology Terms

A

AA Alcoholics Anonymous.

AACAP American Academy of Child and Adolescent Psychiatry.

AACDP American Association of Chairmen of Departments of Psychiatry.

AADPRT American Association of Directors of Psychiatric Residency Training, Inc.

AAMC Association of American Medical Colleges.

AAP Association of Academic Psychiatry.

ability Capacity or competence to perform an act without previous training. *Compare* aptitude.

ablation Excision of part of the body, as in lobotomy or topectomy.

ablutomania Excessive interest in bathing and cleaning oneself; commonly observed in obsessive-compulsive disorder.

abnormal Term subject to considerable variability in definition and usage, roughly the equivalent of "psychopathologic," and defined according to several criteria: **1.** *Normative* or *statistical approach,* considers as abnormal any mental or behavioral activities that deviate from culturally determined averages or norms. **2.** *Subjective approach,* regards as abnormal or psychopathologic those phenomena that cause pain or suffering or induce a person to seek help. **3.** *Theoretical approach,* considers as abnormal any deviations from a theoretical concept of normal functioning that are indicative of psychiatric illness.

abortion Spontaneous or induced expulsion of an embryo or fetus before it has reached the stage of viability, that is, before about 20 weeks of gestation.

ABPN American Board of Psychiatry and Neurology.

abreaction A process by which repressed material, particularly a painful experience or a conflict, is brought back to consciousness; in this process, the person not only recalls but relives the repressed material, which is accompanied by the appropriate affective response. *See also* catharsis.

absence seizure PETIT MAL EPILEPSY.

absolute threshold Minimum amount of stimulus needed to produce a response.

abstinence The act of refraining voluntarily from some activity, or from the use of certain substances, such as food or drugs; in psychoanalysis, refers to refraining from sexual intercourse.

abstinence syndrome WITHDRAWAL SYNDROME; DISCONTINUATION SYNDROME.

abstract attitude ABSTRACT THINKING.

abstraction The process whereby thoughts or ideas are generalized and dissociated from particular concrete instances or material objects; concreteness in proverb interpretation suggests an impairment of abstraction, as in schizophrenia.

abstract thinking Thinking characterized by the ability to assume a mental set voluntarily, to shift voluntarily from one aspect of a situation to another, to keep in mind simultaneously various aspects of a situation, to grasp the essentials of a whole and to break a whole into its parts, to discern common properties, to plan, to assume make-believe attitudes, and to think or act symbolically. The capacity for abstract thinking is frequently impaired in patients with organic mental disorders and schizophrenia. Also called abstract attitude; categorical attitude. *Compare* concrete thinking.

abulia Inability to make decisions; lack of will.

abuse **1.** To attack or injure physically or psychologically, most commonly children, the older adults, and marital partners. *See also* battered child syndrome. **2.** To misuse substances such as drugs, illegal or prescribed, or alcohol so as to impair functioning.

acalculia Loss of ability to do calculations; not caused by anxiety or impairment in concentration. Occurs with neurologic deficit and learning disorder.

acarophobia Abnormal fear of minute animate (insects, worms) or inanimate (pins, needles) objects.

acatalepsia Severe mental deficiency characterized by the inability to reason or comprehend.

acataphasia Disordered speech in which statements are incorrectly formulated. Patients may express themselves with words that sound like the ones intended but that are not appropriate to the thoughts, or they may use totally inappropriate expressions.

acathexis Lack of feeling associated with an ordinarily emotion-charged subject; in psychoanalysis, denotes the patient's detaching or transferring of emotion from thoughts and ideas. Occurs in anxiety, dissociative, schizophrenic, and bipolar disorders. Also called decathexis. *Compare* cathexis.

accelerated interaction A term used in group therapy to refer to an extended group therapy session that may last for hours, sometimes a day. Alternate term for marathon group session.

accident proneness Tendency to have accidents as a result of psychological causes.

accreditation A voluntary process by which standards for hospitals and health facilities are approved by city or federal government agencies; may also refer to being credentialed by the American Board of Psychiatry and Neurology (ABPN). *See also* board-certified psychiatrist.

acenesthesia Loss of sensation of physical existence.

acetylcholine A neurotransmitter that causes cardiac inhibition, vasodilation, gastrointestinal peristalsis, and other parasympathetic effects; plays a critical role in memory, sleep, and the regulation of mood.

achluophobia Abnormal fear of the dark.

acid Slang for lysergic acid diethylamide (LSD).

ACLP Academy of Consultation-Liaison Psychiatry

ACP American College of Physicians.

acquired immunodeficiency syndrome (AIDS) According to the U.S. Centers for Disease Control, "a disease, at least moderately predictive of a defect in cell-mediated immunity, occurring in a person with no known cause for diminished resistance to that disease. Such diseases include Kaposi sarcoma (KS), *Pneumocystis carinii* pneumonia (PCP), and other serious opportunistic infections." The disease is caused by the human immunodeficiency virus type 1 (HIV-1). *See also* AIDS dementia complex.

acromegaly Disorder marked by hypersecretion of growth hormone by the anterior pituitary resulting in characteristic enlargement of certain body structures, such as the head, hands, feet, lips, and jaw; often associated with somnolence, moodiness, and decreased libido.

acrophobia Abnormal fear of high places.

acting out Behavioral response to an unconscious drive or impulse that brings about temporary partial relief of inner tension; relief is attained by reacting to a present situation as if it were the situation that originally gave rise to the drive or impulse. Common in borderline states.

activity group psychotherapy Type of group therapy, introduced and developed by S. R. Slavson, designed for children and young adolescents, emphasizing emotional and active interaction in a permissive, nonthreatening atmosphere; the therapist interprets behavior and stresses reality testing.

actualization

Realization of one's full potential. *See also* individuation.

aculalia Nonsense speech associated with marked impairment of comprehension. Occurs in mania, schizophrenia, and neurologic deficit.

acute stress disorder Severe emotional reaction resulting from extreme environmental stress such as the death of a loved one, a disaster, or a catastrophic life situation. It usually remits within hours or days; however, if it persists for over a month it may evolved into a condition called posttraumatic stress disorder (PTSD).

ADAMHA Alcohol, Drug Abuse, and Mental Health Administration.

adaptation Process of fitting or conforming to the environment by behavioral or psychic changes that promote optimal functioning. *See also* adjustment.

adaptational approach In analytic therapy, consonant with Sandor Rado's formulations on adaptational psychodynamics, therapy focusing on the maladaptive patterns used by patients in the treatment sessions, on how those patterns developed, and on what the patients must do to overcome them and stabilize their functioning at self-reliant, adult levels. *See also* social adaptation.

adaptive behavior Any behavior that increases an organism's ability to adjust to a specific environment or situation.

ADD Attention-deficit disorder, usually referred to as Attention Deficit/Hyperactivity Disorder (ADHD).

addiction Repeated and increased use of a drug or chemical substance, the deprivation of which gives rise to symptoms of distress and an irresistible urge to use the agent again, leads to deterioration of physical and mental health. *See also* drug dependence.

ADHD Attention-deficit hyperactivity disorder.

adiadochokinesis Inability to perform rapid alternating movements. Occurs with neurologic deficit and cerebellar lesions.

adjustment Maladaptive reaction to identifiable circumstances or life events that is expected to remit when the stress ceases. In DSM-5, divided into the following subtypes based on symptoms: with anxiety, with depressed mood, with mixed anxiety and depressed mood, with disturbance of conduct, with mixed disturbance of emotions and conduct, and unspecified.

adjustment disorder Maladaptive reaction to identifiable circumstances or life events that is expected to remit when the stress ceases.

Adler, Alfred (1870–1937) Viennese psychiatrist and one of Freud's original followers who broke with him and introduced and developed the concepts of individual psychology, inferiority complex, and overcompensation. *See also* masculine protest.

adlerian Attributed to or described by Alfred Adler.

administrative psychiatry Psychiatry concerned with public affairs, policy-making, executive decisions, and the duties of persons in executive positions.

adolescence Period of growth from puberty to maturity. Its beginning is marked by the appearance of secondary sexual characteristics, usually at about age 12; termination is marked by the achievement of sexual maturity at about age 20. *See also* psychosexual development.

adolescent mania Signs of mania occurring in an adolescent as the result of substance abuse (including alcohol) or of antisocial behavior but not as the result of a true bipolar disorder.

adrenaline EPINEPHRINE.

adrenergic 1. Relating to nerve fibers, tracts, or pathways in which synaptic transmission is mediated by norepinephrine (e.g., the sympathetic division of the autonomic nervous system), more generally, to processes activated by epinephrine or norepinephrine. **2.** Relating to drugs (catecholamines) that mimic the actions of the sympathetic nervous system.

adrenergic system Nerve cells or fibers of the autonomic nervous system that employ norepinephrine as their neurotransmitter.

Advanced sleep phase circadian rhythm disorder A circadian rhythm sleep–wake disorder characterized by sleep–wake times that are earlier than conventional times; individuals with this disorder are referred to as “early birds” or “larks.”

adventitious reinforcement Responses that are reinforced accidentally by the coincidental pairing of response and reinforcement; such events may have clinical implications in the development of phobias or other neurotic behavior.

adverse effects Secondary, unfavorable reactions or results; usually referring to side effects of drugs.

adynamia Weakness and fatigability, a characteristic of neurasthenia and depression.

adynamic ileus PARALYTIC ILEUS.

aerophagia Excessive swallowing of air. Seen in anxiety disorder.

affect The subjective and immediate experience of emotion attached to ideas or mental representations of objects. Affect has outward manifestations that may be classified as restricted, blunted, flattened, broad, labile, appropriate, or inappropriate. *See also* mood.

affect-fantasy Jung's term for an emotional-laden fantasy.

affectionate transference *See* transference.

affective disorder *See* MOOD DISORDER.

affective hallucination Hallucination in which the content is either grandiose or self-deprecatory or involves other features common to a depressive syndrome such as guilt, disease, or poverty.

affective interaction Interpersonal experiences and exchanges that are emotionally charged.

aftercare After hospitalization, a continuing program of rehabilitation designed to reinforce the effects of therapy and to help patients adjust to their environment.

ageism Prejudice or discrimination against a particular age group, especially the older adults.

ageusia Lack or impairment of the sense of taste; may be seen in depression and neurologic deficit.

aggression Forceful behavior, verbal or physical, usually directed against another person.

aggressive drive One of the two primal drives or instincts (the other is the sexual drive) in Freud's dual-instinct psychoanalytic theory. It is conceived of as an unconscious destructive drive or an impulse directed at oneself or another that aims toward dissolution and death and operates on the repetition–compulsion principle, in contrast to the sexual drive, which follows the pleasure–pain principle. Also called death instinct, Thanatos.

aging The process of growing old; in human physiology, aging is said to occur when there is a failure to replace cells in sufficient numbers to maintain full functional capacity.

agitated depression Major depressive disorder with psychomotor agitation as a primary sign and symptom, characterized by restlessness, increased motor activity, and racing thoughts.

agitation State of anxiety associated with severe motor restlessness.

agnosia Inability to understand the import or significance of sensory stimuli; cannot be explained by a defect in sensory pathways or sensorium. In strict usage, diagnosis of agnosia implies an organic cerebral lesion; however, the term has also been used to refer to the selective loss or disuse of knowledge of specific objects due to emotional circumstances, as seen in certain schizophrenic, hysteric, and depressed patients. For types of agnosia, see the specific term.

agonist In pharmacology, a substance that excites the receptor sites of the cell and promotes or enhances a biologic response. *Compare* antagonist.

agoraphobia Morbid fear of open places or of leaving the familiar setting of the home; may be present with or without panic attacks.

agraphia Loss or impairment of a previously possessed ability to write.

aichmophobia Abnormal fear of pointed objects.

AIDS Acquired immunodeficiency syndrome.

AIDS dementia complex Neurologic complication of AIDS characterized by confusion, impaired judgment, impulsive behavior, and psychosis. Also called AIDS encephalopathy.

ailurophobia Abnormal fear of cats.

akathisia State of motor restlessness manifested by a compelling need to be in constant movement; may be seen as an extrapyramidal side effect of phenothiazine or other antipsychotic medication.

akinesia Lack of physical movement, as in the extreme immobility of catatonic schizophrenia; may also occur as an extrapyramidal effect of antipsychotic medication.

akinetic mutism Absence of voluntary motor movement or speech in a patient who is apparently alert (as evidenced by following eye movements). Seen in psychotic depression and catatonic states.

Al-Anon Organization of relatives of alcoholics, patterned after Alcoholics Anonymous, which promotes the discussion and resolution of common problems.

Alateen Organization directed to children of alcoholics, through group support, so that they may better understand their parents' alcoholism. *See also* Alcoholics Anonymous.

alcohol abuse and dependence Called alcohol use disorder in DSM-5. *See also* alcoholism.

alcohol amnestic disorder Alternative term for Korsakoff psychosis.

alcohol dementia WERNICKE SYNDROME.

alcoholic blackout Amnesia experienced by alcoholics concerning behavior during drinking bouts; usually indicates that reversible brain damage has occurred.

alcoholic deterioration Dementia and mental deterioration associated with chronic excessive alcohol use. Called dementia associated with alcoholism in DSM.

alcoholic hallucinosis Occurrence of hallucinations with a clear sensorium in a person with a history of heavy drinking and alcohol dependence; usually follows a prolonged drinking bout. *See also* Delirium tremens.

alcoholic psychosis Mental disorder that results from alcoholism and that involves organic brain damage, as in delirium tremens and Korsakoff psychosis.

Alcoholics Anonymous (AA) An organization of alcoholics formed in 1935 that uses certain group methods, such as inspirational-supportive techniques, to help rehabilitate chronic alcoholics.

alcohol idiosyncratic intoxication Syndrome of marked alcohol intoxication with subsequent amnesia for the period of intoxication produced by the ingestion of quantities of alcohol that would be insufficient to induce intoxication in most people. Also called pathologic intoxication.

alcohol intoxication Constellation of specific neurologic, psychological, and behavioral effects produced by the recent ingestion of alcohol; characteristically, the effects include slurred speech, motor ataxia, disinhibition of sexual or aggressive impulses, lability of mood, impairment of attention or memory, and impairment of judgment. *See also* intoxication.

alcoholism Excessive dependence or addiction to alcohol, usually to the point that the person's physical and mental health are threatened or harmed; may be episodic or continuous. Called alcohol use disorder in DSM-5.

alcohol withdrawal syndrome Symptom complex resulting from the cessation or reduction in alcohol intake in a person who has been drinking heavily; includes tremulousness, seizures (rum fits), hallucinations, and autonomic hyperactivity (delirium tremens). *See also* withdrawal syndrome.

Alexander, Franz (1891–1964) Hungarian psychoanalyst who founded the Chicago Institute for Psychoanalysis, developed the concept of corrective emotional experience and specificity hypothesis in psychosomatic disorders, and, in collaboration with Thomas French in 1946, identified most of the basic characteristics of brief dynamic psychotherapy.

alexia Loss of the power to understand the meaning of written or printed words and sentences. *Compare* dyslexia.

alexithymia Inability or difficulty in describing or being aware of one's emotions or moods; associated with depression, substance abuse, and posttraumatic stress disorder.

algophobia Abnormal fear of pain.

algorithm Graphic representation used as a set of instructions for solving diagnostic or clinical problems. Also called decision tree.

alienation In psychiatry, used to describe a person's feelings of detachment from self or society, to denote one's avoidance of emotional experiences, or to describe efforts to estrange oneself from one's own feelings. *Compare* relatedness. *See also* isolation.

alienist Obsolete term for a psychiatrist who offers expert opinion about a person's mental health or sanity.

allergic jaundice Relatively rare, unpredictable, non–dose-dependent hypersensitivity reaction after treatment with some phenothiazines; generally appears during the second to fourth weeks of therapy, resulting in hyperbilirubinemia and deposition of bile pigment in the skin and mucous membranes which causes a yellow appearance of the patient; an obstructive (cholestatic) type of jaundice manifested by predominantly conjugated (direct-reacting) hyperbilirubinemia, the presence of bilirubin in the urine, and an abnormally high plasma alkaline phosphatase.

alliance *See* therapeutic alliance.

allied health professional *See* paraprofessional.

alloplasty Adaptation to stress by attempting to change the environment. *Compare* autoplasty.

Allport's group relations theory Theory developed by Gordon W. Allport that a person's behavior is influenced by their personality and the need to conform to social forces; illustrates the interrelationship between group therapy and social psychology.

alogia Inability to speak because of mental deficiency or an episode of dementia.

alopecia Hair loss. *See* trichotillomania.

alpha (α) error TYPE 1 ERROR.

alpha waves As measured by an electroencephalogram (EEG), brain waves that are 8 to 13 Hz, indicating a relaxed, waking state.

ALS Amyotrophic lateral sclerosis.

alternating role Pattern characterized by periodic switching from one type of behavior to another.

altruism Regard for and dedication to the welfare of others; a term originated by the French philosopher Auguste Comte (1798–1857). In psychiatry, the term is closely linked with ethics and morals; Freud recognized altruism as the only basis for the development of community interest; Bleuler equated it with morality.

Alzheimer disease DEMENTIA OF THE ALZHEIMER TYPE.

AMA American Medical Association.

ambivalence Presence of strong and often overwhelming simultaneous contrasting attitudes, ideas, feelings, and drives toward an object, person, or goal; term coined by Bleuler, who differentiated three types: affective ambivalence, intellectual ambivalence, and ambivalence of the will. Commonly seen in schizophrenia and borderline personality disorder.

ambulatory schizophrenia A descriptive term refers to schizophrenic mental illness that is sufficiently well compensated so as not to require continuous hospitalization.

amentia Lack of intellectual development as a result of inadequate brain tissue. *Compare* dementia.

American Academy of Child and Adolescent Psychiatry (AACAP) Organization of child and adolescent psychiatrists, founded in 1953, whose objectives are to study and advance medical contributions to the knowledge and treatment of psychiatric illnesses of children and adolescents.

American Association of Chairmen of Departments of Psychiatry (AACDP) Organization of chairpersons of departments of psychiatry, founded in 1967, that concentrates on promoting medical education, research, and patient care, especially as they relate to psychiatry; prompting the growth and development of psychiatry; and providing a forum for discussion among chairpersons.

American Association of Directors of Psychiatric Residency Training, Inc. (AADPRT) Organization of training directors of U.S. psychiatric residency programs, founded in 1973, whose mandate is to oversee standards of approved residency criteria and to further psychiatric education.

American Board of Psychiatry and Neurology (ABPN) Group, founded in 1934 that sets the standards for teaching and training in psychiatry and neurology in American hospitals and training centers. Its members are from the American Medical Association, American Psychiatric Association, and American Neurological Association and is a member of the American Board of Medical Specialties. It conducts periodic examinations of those who have satisfactorily completed their training; upon passing the examination, candidates are certified to practice the specialty of psychiatry and/or neurology and/or child psychiatry. It also oversees continued certification for psychiatrists and neurologists.

American College of Physicians (ACP) Honorary organization of internists and related medical specialists, founded in 1915, that sponsors postgraduate courses for physicians as well as teaching and research scholarships.

American College of Psychiatrists (ACP) Honorary organization of senior psychiatrists, founded in 1963, established to honor persons who have made a significant contribution to the study and development of psychiatry in the United States.

American Law Institute formulation of insanity Legal definition of insanity whereby a person is absolved of criminal responsibility if, as a result of a mental disorder, he “lacks substantial capacity either to appreciate the wrongfulness of his conduct or to conform his conduct to the requirements of law.” *See also* Durham rule; M’Naghten rule.

American Medical Association (AMA) National organization of medical doctors of all specialties, founded in 1847, whose mandate is to monitor the educational, training, and research concerns of the medical profession; plays an advisory role in national legislation related to health care and disseminates medical information.

American Neurological Association (ANA) National organization of neurologists, founded in 1875 that conducts research and educational programs; also elects members to the American Board of Psychiatry.

American Psychiatric Association (APA) Most important and largest professional organization of American physicians who specialize in the practice of psychiatry. The organization was founded in 1844 by 13 superintendents (“original thirteen”) as the Association of Medical Superintendents of American Institutions for the Insane; in 1891, the name was changed to the American Medico-Psychological Association and in 1921 to its present name. Its purpose is to advance the development of psychiatric education, research, and practice in the United States.

American Psychoanalytic Association (APA) National organization of graduates of various psychoanalytic training programs, founded in 1911, mandated to establish and maintain standards for the field of psychoanalysis and the training of psychoanalysts and to encourage research; historically represents orthodox schools of classical psychoanalysis derived from Freudian theory.

American Psychological Association Founded in 1892, largest and most important organization of American psychologists who specialize in experimental, clinical, research, and educational psychology, and many other subspecialties in psychology; sometimes known as the “big APA” to distinguish it from the smaller American Psychiatric Association.

amimia Lack of ability to make gestures or to comprehend those made by others.

amines Organic compounds derived from ammonia by replacing hydrogen atoms with organic groups, such as the catecholamines.

amino acid Organic acid in which one of the CH hydrogen atoms has been replaced by NH_2; structured building blocks of proteins that function as neurotransmitters similar to the biogenic amines, probably in 60% to 70% of the synapses in the brain.

amnesia Disturbance in memory manifested by partial or total inability to recall past experiences. For types of amnesia, see the specific term.

amnestic aphasia, amnesic aphasia Disturbed capacity to name objects, even though they are known to the patient. Also called anomic aphasia.

amnestic disorder Mental syndrome in which impairment of memory is the predominant cognitive defect; in DSM-5, called major or minor neurocognitive disorder due to another medical condition.

amok Condition, usually associated with Malayan men, characterized by a sudden, unprovoked outburst of wild rage usually culminating in homicide.

amphetamine Central nervous system stimulant with a chemical structure and action closely related to the sympathomimetic amines; used to treat depression, attention-deficit hyperactivity disorder, hypersomnia, and narcolepsy; may be abused to prolong wakefulness and attentiveness.

amygdala Part of the basal ganglia complex in the dorsomedial portion of the temporal lobe: it and the hippocampus are major components of the limbic system and have been implicated in the production of memory, emotions, and violent behavior.

amyotrophic lateral sclerosis (ALS) Degenerative neuromuscular disease, caused by a slow virus, with an insidious onset beginning between the ages of 50 and 70; wasting begins in upper limbs, eventually spreading to the tongue, mouth, and palate. Degeneration of both pyramidal tracts above the medulla produces pseudobulbar palsy, which, when severe, results in impaired voluntary control of emotional reactions, which consequently, may be exaggerated, explosive, and quite inappropriate. Also called Lou Gehrig disease.

Amytal interview Psychiatric interview conducted with the patient having received an intravenous infusion of amobarbital (Amytal), which produces an altered state of consciousness during which repressed memories and feelings may be recalled or reexperienced; used in the treatment of dissociative states, psychogenic amnesia, and conversion disorder.

ANA American Neurological Association.

anaclitic Depending on others, especially as the infant on the mother; anaclitic depression in children results from an absence of mothering.

anaclitic therapy Psychotherapy characterized by allowing the patient to regress; used mainly in the treatment of psychosomatic disorders.

anal character Freud's term to describe persons who are perfectionistic, parsimonious, punctual, and stubborn.

analeptic PSYCHOMOTOR STIMULANT.

anal erotism Pleasurable experience felt by a child in the anal area during the anal phase of psychosexual development.

analgesia State in which one feels little or no pain.

anal phase Second stage in psychosexual development, occurring when a child is between the ages of 1 and 3 years. During that period, activities, interests, and concerns are centered on the anal zone, and the pleasurable experience felt in that area is called anal erotism. *See also* infantile sexuality; psychosexual development.

analysand In psychiatry, the one who is being analyzed.

analysis in depth PSYCHOANALYSIS.

analysis of variance (ANOVA) Statistical technique designed to compare two or more groups of observations by which sets of measurements are investigated to find out to what extent they are determined by experimental influences or chance influences. *See also* variance.

analytic psychology Jung's system of psychology, characterized by a belief in the collective unconscious or objective psyche, the archetype, and the complex.

anamnesis Medical history of a patient, particularly used in connection with the patient's own recollections.

anancasm Repetitious or stereotyped behavior or thought usually used as a tension-relieving device. Alternate spelling, anankastic.

anankastic personality OBSESSIVE-COMPULSIVE PERSONALITY.

ancillary care Auxiliary, accessory, or secondary services provided in the treatment or rehabilitation of patients.

androgen Hormone that is produced in the testes and the adrenal glands and that stimulates activity of the accessory male sex organs, promotes development of male sex characteristics, or prevents changes in the latter following castration; for example, testosterone. *Compare* estrogen.

androgyny Combination of female and male characteristics in one person. *See also* bisexuality.

androphobia Irrational fear of men.

anergia Lack of energy.

anergic schizophrenic BURNED-OUT SCHIZOPHRENIC.

anesthesia Absence of sensation. *See also* hysterical anesthesia.

angel dust Street term for phencyclidine (PCP).

anhedonia State of being unable to experience pleasure. *Compare* hedonism.

anima According to jungian psychology, the person's inner self, as opposed to the self that one presents to the outside world (persona).

Anna O A patient of Joseph Breuer and Sigmund Freud whose arm became paralyzed whenever she was unconsciously reminded of her feelings for her father (she had been sitting at his bedside with her arm pressed against the side of his bed at the time of his death). The motor disturbance represented an action that was associated with an objectionable infantile sexual impulse. Later identified as Bertha Pappenheim who became a noted feminist, social pioneer, and founder of the Jewish Women's Association.

anniversary reaction Behavior, symptoms, or dreams that occur at the same time of year as a significant past event.

anomia Inability to recall the names of objects.

anomic aphasia AMNESTIC APHASIA.

anomie Term, popularized by Émile Durkheim as a major cause for suicide, referring to a sense of alienation and despair resulting from the loss or weakening of previously valued norms, ideals, or goals.

anorectic Drug that decreases appetite.

anorexia nervosa Serious and sometimes life-endangering condition characterized by a disturbed body image and self-imposed severe dietary limitation, usually resulting in serious malnutrition; appetite usually remains intact.

anorgasmia Inability to achieve orgasm in the female. *See also* frigidity.

anosognosia Unawareness or nonacceptance of a neurologic deficit.

anosomnia Loss of ability to smell.

ANOVA Analysis of variance.

antagonist In pharmacology, a substance that reduces or blocks the action of another; for example, naloxone (Narcan) is used in the treatment of opioid overdose because it is reverses the effects of narcotics. *Compar*e agonist.

anterograde amnesia Loss of memory for events subsequent to the onset of the amnesia. *Compare* retrograde amnesia.

anthropology Study of humans in relation to distribution, origin, classification, and relationship of races, physical characteristics, environmental and social relations, and culture.

antianxiety drug ANXIOLYTIC.

anticholinergic Drug that blocks the actions of atropine; primarily used in the treatment of medication-induced movement disorders. *See also* antiparkinsonism drug.

anticholinergic effect Effect, such as a dry mouth and blurred vision, due to a blockade of the cholinergic (especially the parasympathetic) nerves; often seen as a side effect of therapy with phenothiazine and tricyclic antidepressant drugs.

anticipation Term implying careful planning or worrying, and premature but realistic expectation of dire and potentially dreadful outcomes, in goal-directed inner discomfort.

anticonvulsant Any drug with the ability to prevent the excessive discharge of neurons in the brain and thus lower the threshold for convulsions or seizures. Also called antiepileptic drug.

antidepressant drug Any of a class of psychotherapeutic drugs used in the treatment of pathologic depression. Two main categories are the cyclic antidepressants and the monoamine oxidase inhibitors; more recently, serotonergic agents such as fluoxetine. Lithium carbonate also has antidepressant activity in certain patients.

antiepileptic drug ANTICONVULSANT.

antimanic drug Drug used to alleviate the symptoms of mania; lithium is particularly effective in preventing relapses in bipolar disorder.

antiparkinsonism drug Drug that relieves the symptoms of parkinsonism and is effective in countering the extrapyramidal side effects often induced by antipsychotic drugs by acting on the central nervous system to diminish skeletal muscle tone and involuntary movements. Levodopa, the agent of choice for Parkinson disease, does not antagonize the extrapyramidal effects of antipsychotic drugs. Anticholinergic agents, such as benztropine and trihexyphenidyl, effectively counteract drug-induced parkinsonism but are now secondary drugs in the treatment of Parkinson disease.

antipsychotic Any of a class of psychotherapeutic drugs used to treat psychosis, particularly schizophrenia. Also called neuroleptic; major tranquilizer; ataractic drug.

antisocial behavior Behavior pattern of such marginal criminal types as racketeers, prostitutes, and illegal gamblers. Formerly called sociopathic behavior. In lay usage often confused with asociality. *See also* antisocial personality disorder.

antisocial personality disorder Disorder characterized by an inability to get along with other members of society and by repeated conflicts with individual persons and groups; common attributes include impulsiveness, egocentricity, hedonism, low frustration tolerance, irresponsibility, inadequate conscience development, exploitation of others, and rejection of authority and discipline. *See also* antisocial behavior.

Anton syndrome Disorder caused by a lesion in the occipital lobe of the brain that causes blindness which the patient denies experiencing.

Anxiety Unpleasurable emotional state associated with psychophysiological changes in response to an intrapsychic conflict; in contrast to fear, the danger or threat in anxiety is unreal. Psychological changes consist of an uncomfortable feeling of impending danger, an overwhelming awareness of being powerless, inability to perceive the unreality of the threat, prolonged feeling of tension, and exhaustive readiness for the expected danger; physiologic changes consist of increased heart rate, disturbed breathing, trembling, sweating, and vasomotor changes. *Compare* fear.

anxiety disorder Disorder in which anxiety is the most prominent disturbance or in which patients experience anxiety if they resist giving in to their symptoms. In DSM, anxiety disorders include separation anxiety disorder, selective mutism, panic disorder with agoraphobia, agoraphobia without history of panic disorder, social phobia, simple phobia, and generalized anxiety disorder.

anxiety disorders of childhood or adolescence Disorders characterized by chronic anxiety, unrealistic fears, hypersensitive autonomic responses, or fear of leaving home; patients are usually immature, inhibited, timid, approval-seeking, and apprehensive in new situations or places.

anxiety hysteria PHOBIC DISORDER.

anxiety neurosis As initially described by Sigmund Freud, a syndrome characterized by general irritability, anxious expectation, pangs of conscience, anxiety attacks, and phobias. Called generalized anxiety disorder in DSM.

anxiety state Disorder characterized by panic and anxious overconcern; somatic symptoms are often prominent.

anxiolytic Any of a class of psychotherapeutic drugs used to treat pathologic anxiety and its related symptoms; includes the benzodiazepines and buspirone. Also called antianxiety drugs; minor tranquilizers.

APA American Psychiatric Association; American Psychoanalytic Association; American Psychological Association.

apathetic withdrawal *See* withdrawal.

apathy Want of feeling or affect, a lack of interest and emotional involvement in one’s surroundings; observed in certain types of schizophrenia and depression.

Apgar scores Measurements taken 1 and 5 minutes after birth to determine physical normality in the neonate; scores are based on color, respiratory rate, heartbeat, reflex action, and muscle tone.

aphasia Any disturbance in the comprehension or expression of language due to a brain lesion. For types of aphasia, see the specific term.

aphonia Loss of voice.

aphrodisiac Substance or drug that increases sexual desire.

apnea Temporary cessation of breathing. *See* obstructive sleep apnea syndrome.

apoplexy Old classical term for cerebrovascular accident.

apperception Awareness of the meaning and significance of a particular sensory stimulus as modified by one's own experiences, knowledge, thoughts, and emotions. *See also* perception.

appetitive phase First stage in the sexual response cycle, as described by Masters and Johnson, characterized by sexual fantasies and the desire to have sex.

appropriate affect Emotional tone in harmony with the accompanying idea, thought, or speech.

apraxia Inability to perform a voluntary purposeful motor activity; cannot be explained by paralysis or other motor or sensory impairment. In constructive apraxia, a patient cannot draw two- or three-dimensional forms.

aptitude Potential to acquire competence or skill if trained. *Compare* ability.

archetype Jung's term for the symbols used by individuals to express the collective experience of the human race. *See also* collective unconscious.

arithmetic disorder DEVELOPMENTAL ARITHMETIC DISORDER.

arousal The establishment and maintenance of a physiologic or psychological state of being awake, alert, or responsive to stimuli. *See also* hyperarousal.

arteriosclerosis Degenerative condition marked by loss of elasticity, hardening, and thickening of the arteries, with progressive narrowing of their lumina and gradual obstruction to blood flow; organic mental disorder may develop when cerebral arteries are involved in the degenerative process.

artificial intelligence Computer production of a pattern of output that would be considered intelligent if displayed by human beings.

art therapy Treatment procedure that uses the spontaneous creative work of the patient; for example, persons make and analyze drawings, which are often expressions of their underlying emotional problems.

asociality Indifference to social values or customs, withdrawal from society, as seen in a recluse or a regressed schizophrenic person.

Asperger syndrome Term previously used to describe a condition that develops in infancy and is associated with impaired social interaction and restricted and stereotyped patterns of behavior, but without significant delays in language or cognitive development. In DSM-5, the behavioral condition of Asperger syndrome is subsumed in Level I of the autism spectrum. *See also* autism spectrum disorder.

assertiveness training Behavior therapy that involves teaching frank and direct interpersonal expression of feelings and ideas.

assessment The evaluation of a patient that leads to diagnosis and treatment.

assimilation Subjective process that involves filtering the world through one’s own system of knowledge, that is, taking in of new experience through one’s established mental structure.

association Relation between ideas or feelings. *See* clang association; loosening of associations.

Association of American Medical Colleges (AAMC) Organization of American medical schools and teaching hospitals, founded in 1876, concerned with the advancement and improvement of medical education, biomedical research, and health care in the United States.

astasia-abasia Inability to stand or walk in a normal manner, even though normal leg movements can be performed in a sitting or lying down position; seen in conversion disorder.

astereognosis Inability to identify familiar objects by touch; cannot be explained by an impairment of elementary sensory pathways or a depressed sensorium. *See also* neurologic amnesia.

asthenic personality Disorder characterized by lack of enthusiasm, fatigability, lack of capacity for enjoyment, and low tolerance for stress. *See also* adynamia; anhedonia.

astraphobia, astrapophobia Abnormal fear of thunder and lightning.

asyndesis Disorder of language, commonly seen in schizophrenia, in which the patient combines unconnected ideas and images.

ataque de nervios A syndrome, usually found in Latin American cultures, characterized by uncontrollable shouting, crying, trembling, and physical or verbal aggression.

ataractic drug ANTIPSYCHOTIC.

ataxia Lack of coordination, either physical or mental. **1.** In neurology, refers to loss of muscular coordination. **2.** In psychiatry, the term "intrapsychic ataxia" refers to lack of coordination between feelings and thoughts; seen in schizophrenia and in severe obsessive-compulsive disorder.

athetosis Movement disorder characterized by involuntary, slow, writhing, worm-like movements of the fingers and extremities; a form of extrapyramidal dyskinesia.

atonia Lack of muscle tone. *See* waxy flexibility.

attachment According to John Bowlby, the affective tie between infant and mother or primary caregiver, the disruption of which heightens vulnerability to depression or anxiety. *See also* bonding.

attachment disorder of infancy Absence or disruption of behaviors that serve to anchor the infant to its mother or primary caregiver and that produce abnormal reactions in the infant's behavior.

attachment learning Theory that learning is enhanced in the presence of the person to whom one is attached or affiliated; applies especially to infant learning.

attention Concentration; the aspect of consciousness that relates to the amount of effort exerted in focusing on certain aspects of an experience, activity, or task.

attention-deficit hyperactivity disorder (ADHD) DSM-5 diagnosis for a childhood mental disorder characterized by a developmentally inappropriate short attention span, hyperactivity, and poor concentration that cause clinically significant impairment in functioning. The category subsumes abnormal behavior patterns that had historically been referred to by a variety of names, including hyperkinetic syndrome, hyperactive child syndrome, and minimal brain dysfunction. If hyperactivity is not present, the disorder is called attention-deficit disorder (ADD).

attitude Preparatory mental posture with which one receives stimuli and reacts to them.

attributable risk Absolute incidence of a disease in exposed individuals that can be attributed to the exposure.

atypical In psychiatry, denoting an unusual or rare variant of a known disorder.

atypical child Child with distorted personality development; often used in connection with brain-damaged or autistic children.

atypical depression Diagnosis used to describe depression characterized by weight gain and hypersomnia rather than weight loss and insomnia.

atypical psychosis Disorder whose psychotic features are related to a specific culture (e.g., koro); a certain time or event (e.g., postpartum psychosis); or a unique symptomatology (e.g., Capgras syndrome). In DSM, depending on the specific syndrome, may be diagnosed as Other Specified Schizophrenia Spectrum and Other Psychotic Disorder.

audit *See* medical audit.

auditory agnosia The inability to recognize specific sounds in the context of intact hearing.

auditory amnesia NEUROLOGIC AMNESIA.

auditory hallucination Hallucination primarily involving sound.

aura 1. Warning sensations that a person usually experiences just before an epileptic seizure. **2.** Sensory prodrome that precedes a classic migraine headache.

authority figure Real or projected person in a position of power; transferentially, a projected parent.

autism *See* autistic thinking; infantile autism.

autistic spectrum disorder A range of behavioral abnormalities that begin in infancy and are characterized by deficits in social communication; failure to communicate; lack of empathy; abnormalities in eye contact; repetitive patterns of activities; stereotyped or absent speech; and either hyper- or hypoarousal to the environment. In DSM-5, autism spectrum disorder is classified into three levels based on degrees of impairment in speech and social interaction. *See also* Asperger syndrome; Rett syndrome; childhood disintegrative disorder.

autistic thinking Thinking in which the thoughts are largely narcissistic and egocentric, with emphasis on subjectivity rather than objectivity, and without regard for reality; used interchangeably with autism and dereism.

autoerotic asphyxia Asphyxia caused by intentional strangulation while masturbating in order to intensify orgasm through reduction of oxygen blood flow to the brain. High risk of death.

autoerotism 1. Sexual arousal of oneself without the participation of another person; term coined by Havelock Ellis, which is presently used interchangeably with masturbation. **2.** In psychoanalysis, a primitive phase in object relationship development preceding the narcissistic stage; in narcissism, there is a love object, but there is no love object in autoerotism.

autogynephilia A gender identity disorder in which a male becomes sexually aroused by imagining themselves transformed into a woman.

automatic judgment Reflex performance of a way of acting based on a fixed set of beliefs or values.

automatic obedience Compliance to commands without exercising critical judgment; a feature of catatonia.

automatism Undirected behavior not controlled by conscious thought.

automysophobia Abnormal fear of smelling bad or of being unclean.

autonomic nervous system Part of the nervous system that functions outside of consciousness and that directs such visceral functions as respiration, heart rate, and digestion; subdivided into sympathetic (adrenergic) and parasympathetic (cholinergic) divisions.

autonomic side effect Disturbance of autonomic nervous function that may result from the use of antipsychotic or antidepressant drugs, particularly the phenothiazine and tricyclic derivatives. Common side effects of such agents include orthostatic hypotension, blurred vision, nasal congestion, dry mouth, and constipation; such anticholinergic effects are more prominent with tricyclic antidepressants than with phenothiazines.

autonomous ego function Operation present at birth, such as perception, intuition, thinking, speaking, language, and aspects of motor development; first described by Heinz Hartmann. *See also* ego psychology.

autophobia Abnormal fear of being alone or of self.

autoplasty Adaptation to stress by changing intrapsychic processes. *Compare* alloplasty.

autoscopic psychosis Rare disorder, not officially listed in DSM-5, in which the patient sees a phantom or specter of their own body; usually psychogenic in origin, but may also be caused by lesions in the temporoparietal lobe of the brain.

autotopagnosia Inability to recognize and identify the parts of one's own body. *See also* finger agnosia.

auxiliary therapist Co-therapist. *See also* co-therapy.

average Central value in a frequency distribution around which other values are distributed. Also called the mean.

aversive conditioning AVERSIVE THERAPY.

aversive therapy Behavior therapy that involves the repeated coupling of an unpleasant or painful stimulus, such as an electric shock, with an undesirable behavior pattern in an effort to eliminate the undesirable behavior. Also called aversive conditioning.

avoidance and escape learning In escape learning, an animal learns a response to get out of a place where it does not want to be; for example, it jumps off an electric grid whenever the grid is charged. Avoidance learning requires an additional response: the same animal on the grid will learn to avoid a shock if it quickly pushes a lever when a light signal goes on. Escape learning and avoidance learning are two forms of aversive control using negative reinforcement; behavior that terminates the source of aversive stimuli is strengthened and maintained.

avoidant attachment According to Mary Ainsworth, a form of attachment in which children avoid returning to their mothers after being separated. Also called shyness disorder. *Compare* secure attachment. *See also* resistant attachment.

avoidant disorder of childhood or adolescence Disorder of childhood or adolescence characterized by a persistent withdrawal or excessive shrinking from strangers.

avoidant personality disorder Personality disorder characterized by low self-esteem, hypersensitivity to rejection, and social withdrawal, but with a desire for affection and acceptance.

Avoidant/restrictive food intake disorder Disorder characterized by an avoidance or lack of interest in food, resulting in severe weight loss, nutritional deficiency, and failure to thrive; a DSM-5 diagnosis. Previously called feeding disorder of infancy or early childhood.

axon Single structure in a nerve cell that conducts nervous impulses away from the cell body and its dendrites.

B

barbiturate Any of a class of sedative-hypnotic drugs derived from barbituric acid that acts as a central nervous system depressant and readily produces psychic and physical dependence; for example, phenobarbital, secobarbital, thiopental sodium.

basal ganglia Collectively, the corpus striatum, substantia nigra, subthalamic nucleus, and substantia innominata, which are involved in disorders of movement, thought processes, affect, cognition, and symptoms of psychosis.

basic anxiety As conceptualized by Karen Horney, the main-spring from which neurotic trends get their intensity and pervasiveness, characterized by vague feelings of loneliness, helplessness, and fear of a potentially hostile world. *See also* anxiety.

basic benefits Insurance term that refers to minimum payments or coverage for medical or psychiatric care.

bathophobia Abnormal fear of depths.

battered child syndrome Physical damage to a child sustained as a result of repeated beatings, usually by a parent, parent surrogate, or relative. *See also* abuse.

battery tests Individual tests used together make up a psychological battery; for example, Halstead–Reitan Neuropsychological Battery.

A test battery can give more information about different areas of function than an individual test and can increase the level of confidence if there is a positive correlation between them.

BEAM Brain electrical activity mapping.

Beard, George M. (1839–1883) American psychiatrist who in 1869 introduced the term “neurasthenia.”

Beers, Clifford (1876–1943) Author of *A Mind That Found Itself* (1909) and who is generally considered to have founded the mental hygiene movement.

behavior Sum total of the psyche that includes impulses, motivation, wishes, drives, instincts, and cravings, as expressed by a person’s behavior or motor activity. Also called conation.

behavioral disorders of childhood *See* disruptive behavior disorders.

behavioral medicine Application of principles of behavior therapy to the prevention, diagnosis, treatment, and rehabilitation of medical disorders; applies to such areas as stress prevention and reduction, pediatric and adult patient management and compliance, pain control, and life-style modification. Biofeedback relaxation training, behavior therapy, and hypnosis are important treatment modalities.

behavioral neurology Field of study that attempts to combine psychiatry and neurology to explain abnormal behavior as a disorder of brain metabolism or as a result of structural alterations of the central nervous system.

behavioral science Scientific discipline concerned with any aspect of human behavior, including interpersonal relationships, development, values, experiences, and activities; the behavioral sciences include sociology, psychology, psychiatry, anthropology, and ethology.

behavior contract Therapeutic strategy that rewards a person for a desired behavior.

behaviorism School of psychological thought founded by John B. Watson that regards only measurable and observable behavior as the appropriate subject matter for human psychology; holds that human behavior can be described in terms of principles that do not require consideration of unobservable mental events, such as ideas and emotions. *See also* behavior therapy.

behavior modification *See* behavior therapy.

behavior therapy Psychiatric treatment modality that focuses on overt and objectively observable behavior and uses various conditioning techniques derived from learning theory to directly modify the patient’s behavior; aims exclusively at symptomatic improvement, without addressing psychodynamic causation. Behavior therapy techniques include assertiveness training, aversive therapy, flooding, implosion, reciprocal inhibition, and desensitization, shaping, and systematic desensitization. *See also* behaviorism; conditioning.

Bender-Gestalt Test Psychological test, designed by Lauretta Bender, which measures the subject’s ability to reproduce a set of geometric designs: is useful for measuring visuomotor coordination as a means of detecting brain damage. Also called visual motor gestalt test.

benzodiazepines Any of a group of drugs used mainly to treat anxiety; also used as sedatives, muscle relaxants, anticonvulsants, anesthetics, and hypnotics.

bereavement Feeling of grief or desolation, especially at the death or loss of a loved one. *See also* mourning.

bestiality Sexual deviation in which a person engages in sexual relations with an animal. *See also* zoophilia.

beta (β) error TYPE 2 ERROR.

beta waves As measured by an electroencephalogram (EEG), brainwaves that are between I4 and 30 Hz; altered in certain epileptic conditions.

bias In statistics, an error in the construction of the assessment instrument that favors one outcome over another.

binge eating disorder Eating disorder characterized by recurrent episodic, uncontrolled, compulsive, and rapid ingestions of large amounts of food over a short period of time (binge eating) followed by feelings of guilt; a DSM-5 diagnosis, differentiated from bulimia nervosa by an absence of compulsive compensatory behaviors to achieve weight loss.

binge eating, binge and purge *See* bulimia nervosa.

biodynamics System of psychoanalytic psychiatry, introduced by Jules H. Masserman, which holds that all behavior results from the organism's genetic potentialities, from the various paths and speed of maturation, and from its unique experiences.

biofeedback Provision of information to a person regarding one or more physiologic processes in an effort to enable the person to gain some element of voluntary control over bodily functions that normally operate outside consciousness. *See also* learned autonomic control.

biogenic amine Endogenous compound believed to play an important role in brain function as a central synaptic neurotransmitter substance; biogenic amines include the catecholamines (epinephrine, norepinephrine, dopamine) and serotonin, an indolamine.

biogenic amine hypothesis Theory of mood disorders which proposes that depression is associated with an absolute or relative deficiency of catecholamines, particularly norepinephrine, at functionally important receptor sites in the brain; conversely, elation may be associated with an excess of such amines.

biologic psychiatry School of psychiatric thought emphasizing biochemical, neurologic, and pharmacologic factors in the causation and treatment of psychiatric disorders.

biologic rhythm Periodic variation in physiologic and psychological functions. For example, a circadian rhythm shows aperiodicity of about 24 hours; an ultradian rhythm has a cycle shorter than 1 day; an infradian rhythm is longer than 1 day.

bipolar I Bipolar disorder in which a patient meets the criteria for a full manic episode, usually sufficiently severe to require hospitalization.

bipolar II Bipolar disorder in which a patient has at least one depressive episode and at least one hypomanic episode.

bipolar disorder A mood disorder in which the patient exhibits both manic and depressive episodes separated by various lengths of time. Previously referred to as manic–depressive illness.

birth trauma Otto Rank's term to describe the birth process, which he considered the basic source of anxiety in human beings.

bisexual Denoting or characterized by sexual attraction to and sexual contact with members of both sexes.

bisexuality Existence of the qualities of both sexes in the same person. Freud postulated that the sexes differentiated both biologically and psychologically from a common core; that differentiation between the two sexes was relative, rather than absolute; and that regression to the common core occurs to varying degrees in both normal and abnormal conditions. *Compare* bisexual; heterosexuality; homosexuality. *See also* androgyny.

bizarre delusion False belief that is patently absurd or fantastic.

blackout *See* alcoholic blackout.

black-patch syndrome Psychosis induced by sensory deprivation as a result of eye patches used after cataract surgery.

blank screen Neutral backdrop on which the patient projects a gamut of transferential irrationalities; passivity of analysts allows them to act as blank screens.

Bleuler, Eugen (1857–1939) Swiss psychiatrist known for his important studies in schizophrenia, a term he coined and preferred over the earlier term "dementia praecox."

blind spot In psychiatry, an area of a person's personality of which they are totally unaware; unperceived areas are repressed, since their recognition would arouse painful or unpleasant emotions. In the course of group or individual psychotherapy, such blind spots often appear obliquely as projected ideas, intentions, and emotions. *See also* projection; scotoma.

blocking Involuntary cessation or interruption of thought processes or speech because of unconscious emotional factors. Also called thought deprivation.

Blood–brain barrier Series of biologically active membranes that regulate the passage of substances into and out of the central nervous system; maintains a chemical homeostasis between the cerebrospinal fluid and the venous and arterial systems.

blood level In psychiatry, the measurement of the concentration of drugs in plasma or serum to determine their therapeutic effect.

blunted affect Disturbance of affect manifested by a severe reduction in the intensity of externalized feeling tone; one of the fundamental symptoms of schizophrenia, as outlined by Eugen Bleuler.

board-certified psychiatrist Psychiatrist who has passed the examinations of the American Board of Psychiatry and Neurology (ABPN). *See also* accreditation.

body dysmorphic disorder Disorder characterized by an unrealistic concern that part of the body is deformed. Included on the obsessive-compulsive spectrum in DSM-5. Also called dysmorphophobia.

body-focused repetitive behavior disorders Descriptive term for a group of behaviors consisting of compulsions that damage one's own body, such as skin biting, lip biting, or nail biting; not a diagnosis in DSM-5.

body image Conscious and unconscious perception of one's body at any particular time.

body language Expression of thoughts and feelings by body posture and movement.

bonding Nonsexual feelings of love, respect, and communality among individuals of the same sex. Specifically, a mother's feelings for her infant, but often used to refer to the general mother–infant relationship; fathers often bond to infants. *See also* attachment.

borderline intellectual functioning A term previously attributed to individuals with below-average cognitive ability (IQ in the 70–80 range) whose intellectual impairment did not meet criteria for intellectual disability (IQ below 70). This condition is not clearly differentiated from mild intellectual disability and may be the focus of clinical attention, however, no definitive criteria for diagnosis exist in DSM-5. *See also* intellectual disability.

borderline personality disorder Personality disorder marked by instability in various areas, impulsiveness, suicidal acts, selfmutilations, identity problems, and feelings of emptiness or boredom.

borderline psychosis BORDERLINE STATE.

borderline state State in which the symptoms are so unclear or transient that it is difficult to classify the patient as psychotic or nonpsychotic. Also called borderline psychosis.

boufée délirante A syndrome, observed in West Africa and Haiti, characterized by a sudden outburst of agitated and aggressive behavior, marked confusion, and psychomotor excitement. It is sometimes accompanied by visual or auditory hallucinations or paranoid ideations. May resemble an episode of brief psychotic disorder.

Bowlby, John (1907–1990) British psychoanalyst who formulated the theory of attachment to describing the mother–infant relationship. *See also* attachment; bonding.

bradykinesia Slowness of motor activity, with a decrease in normal spontaneous movement.

bradylalia Abnormally slow speech, common in depression.

bradylexia Inability to read at normal speed.

brain That part of the central nervous system contained within the cranium; includes the cerebrum, midbrain, cerebellum, pons, and medulla.

brain abscess Localized, suppurative, infectious process involving a part of the brain; may produce an organic mental disorder as well as focal neurologic signs.

brain electrical activity mapping (BEAM) Response of brain electrical activity to sensory stimuli which is recorded by computer averaging techniques that produce a topographic color map of brain activity.

brain imaging Radiologic techniques that allow assessment in vivo of structural and functional neuroanatomy; include computed tomography (CT), magnetic resonance imaging (MRI), regional cerebral blood flow (rCBF), positron emission tomography (PET), and brain electrical activity mapping (BEAM).

brain metabolism Chemical changes in the brain that convert small molecules into large ones (anabolism) and large molecules into small ones (catabolism), and that account for cellular function and repair.

brain stem Area of the brain comprising the mesencephalon, pons, and medulla oblongata, whose basic functions concern respiration, cardiovascular activity, sleep, and consciousness; also the site of the ascending biogenic amine (dopamine, noradrenalin, serotonin) pathways to higher brain areas.

brain syndrome *See* organic mental disorder.

brainwashing Any technique designed to manipulate human thought or action against the desire, will, or knowledge of the person involved; usually refers to systematic efforts to indoctrinate nonbelievers. In DSM-5 diagnosed as "identity disturbance due to prolonged and intense coercive persuasion." Also called coercive persuasion.

brain wave Rhythmic wave-like electrical pattern in an encephalogram.

Brawner decision *See* Model Penal Code.

breathing-related sleep disorder Characterized by sleep disruption leading to excessive sleepiness or insomnia caused by a sleep-related breathing disturbance; three types of breathing-related sleep disorders include obstructive sleep apnea, central sleep apnea, and sleep-related hypoventilation.

Breuer, Josef (1842–1925) Viennese physician who collaborated with Freud in studies of cathartic therapy, as reported in *Studies on Hysteria* (1895); he withdrew as Freud proceeded to introduce psychoanalysis, but left important imprints on that discipline, such as the concepts of the primary and secondary processes.

brief psychotherapy Psychotherapy in which the sessions are limited to 10 to 15 in number and during which time attempts to modify behavior occur; used in both individual and group settings.

brief psychotic disorder DSM-5 category for a psychosis of less than 4 weeks' duration, with sudden onset after a major stress. Also called brief reactive psychosis.

brief reactive psychosis *See* brief psychotic disorder.

Brigham, Amariah (1798–1849) One of the founders of the American Psychiatric Association and the first editor of its journal.

Brill, A. A. (1874–1948) First American analyst (1908) who translated several of Freud's more important works, was active in the formation of the New York Psychoanalytic Society (1911), and remained a foremost propagator of psychoanalysis as a lecturer and writer.

Briquet syndrome Alternate term for somatic symptom disorder.

broad affect Normal range (varies both within and among cultures) of the expression of a subjectively experienced feeling state (e.g., euphoria, anger, sadness); normal expression of affect involves variability in facial expression, pitch of voice, and hand and body movements.

Broca aphasia MOTOR APHASIA.

brooding compulsion INTELLECTUALIZATION.

bruxism Grinding or gnashing of the teeth, typically occurring during sleep.

bulimia nervosa Recurrent episodic, uncontrolled, compulsive, and rapid ingestion of large amounts of food over a short period of time (binge eating) followed by self-induced vomiting, use of laxatives or diuretics, fasting, or vigorous exercise in order to prevent weight gain (binge and purge); weight is not severely lowered.

burned-out schizophrenic Chronic schizophrenia characterized by a person who is apathetic and withdrawn, with minimal florid psychotic symptoms but with persistent and often severe schizophrenic thought processes. Also called anergic schizophrenia.

burnout Syndrome of physical or emotional exhaustion considered to be a stress reaction to unrelenting performance and emotional demands stemming from one's occupation; symptoms include impaired work performance, fatigue, insomnia, increased susceptibility to physical illness, and reliance on alcohol (or other substances) with a tendency to escalation into physiologic dependency or possibly suicide.

butyrophenone Any of a class of antipsychotic drugs of which haloperidol (Haldol) is a representative example.

C

C.A. Chronologic age.

cacodemonomania Condition in which patients think they are possessed by a devil or other evil spirit.

caffeine Alkaloid found in coffee, tea, and cola drinks which acts as a psychomotor stimulant; caffeine intoxication results from the excessive ingestion of caffeine.

Canadian Psychiatric Association (CPA) National organization of Canadian psychiatrists whose purpose is to further the development of psychiatric education, research, and practice in Canada.

cannabis *See* marijuana.

canonical correlation Multivariate technique used in statistics for simultaneously finding the relationship of linear combinations of two or more predictors and two or more outcomes.

Capgras syndrome Delusional disorder in which the patient believes that persons in the environment are not their real selves but are doubles imitating the patient or imposters imitating someone else.

capitation Prepayment fee based on a fixed number of persons that covers all health care services.

care and protection proceedings Intervention by a court on behalf of a child when the parents or caretakers provide inadequately for the child’s welfare.

carebaria Sensation of discomfort or pressure in the head.

caregiver 1. Any person involved in the identification, prevention, or treatment of illness or disability. **2.** In child psychiatry, a person who tends to the needs of an infant or child. Also called caretaker.

case-control study Study in which the illness is the dependent variable, already being present for the subject to be identified; such studies are therefore retrospective.

case history study Retrospective study that samples persons with a given disorder.

case register Medical records in central data banks.

castration anxiety Anxiety concerning a fantasied loss of or injury to the genitalia.

castration complex In psychoanalytic theory, a group of unconscious thoughts and motives that are referable to the fear of losing the genitalia, usually as punishment for forbidden sexual desires.

catalepsy Condition in which persons maintain the body position into which they are placed; observed in severe cases of catatonic schizophrenia. Also called waxy flexibility; cerea flexibilitas. *See also* command automatism.

cataphasia VERBIGERATION.

cataplexy Temporary sudden loss of muscle tone, causing weakness and immobilization; can be precipitated by a variety of emotional states, and it is often followed by sleep; commonly seen in narcolepsy.

catastrophic anxiety Anxiety associated with cognitive disorders when patients are aware of their defects in mentation; the anxiety may be overwhelming.

catathymia Situation in which elements in the unconscious are sufficiently affect-laden to produce changes in conscious functioning.

catathymic crisis Suddenly occurring, isolated, and nonrepetitive act that develops from a state of intolerable tension.

catatonia Clinical syndrome consisting of marked behavioral irregularities, which may include motoric excitement or immobility, negativism, stupor, echolalia, or echopraxia. In DSM-5, catatonia may be a stand-alone diagnosis; a feature of schizophrenia; associated with other mood/psychotic disorders; associated with medications or substance use; or due to an underlying medical condition.

catatonic excitement Excited, uncontrollable motor activity seen in catatonic schizophrenia. Patients in catatonic state may suddenly erupt into an excited state and may be violent.

catatonic posturing Voluntary assumption of an inappropriate or bizarre posture, generally maintained for long periods of time. May switch unexpectedly with catatonic excitement.

catatonic rigidity Fixed and sustained motoric position that is resistant to change.

catatonic schizophrenia A rare type of schizophrenia; no longer included in DSM-5. *See* catatonia.

catatonic state State characterized by muscular rigidity and immobility (catatonia), usually associated with schizophrenia.

catatonic stupor Stupor in which patients ordinarily are well aware of their surroundings.

catchment area In psychiatry, a circumscribed geographic area served by a particular community mental health center.

catecholamine Biogenic monoamine containing a catechol group. Norepinephrine, epinephrine, and dopamine are catecholamines with important effects on peripheral and central nervous system activity.

categorical attitude ABSTRACT THINKING.

categorical thought Term introduced by Jean Piaget to refer to abstract thinking. *See also* secondary process.

catharsis Release of ideas, thoughts, and repressed materials from the unconscious, accompanied by an affective emotional response and relief; commonly observed in the course of both individual and group psychotherapy, but can also occur outside therapy. *See also* abreaction; conversational catharsis.

cathexis In psychoanalysis, a conscious or unconscious investment of psychic energy in an idea, a concept, an object, or a person. *Compare* acathexis.

CAT scan Incorrect but commonly used variant of “CT scan.” *See also* computed tomography.

cat’s cry syndrome *See* cri-du-chat syndrome.

Cattell Infant Intelligence Scale Psychological test used to assess general motor and cognitive development in infants aged 3 months to 2 years.

causalgia Burning pain that may be either organic or psychic in origin.

cause–effect relationship Relationship in which one action causes a specific reaction.

causalgia Burning pain that may be organic or psychic in origin.

cenesthesia Change in the normal quality of feeling tone in a part of the body.

central nervous system (CNS) The spinal cord and the brain.

central sleep apnea Dyssomnia characterized by repeated episodes of apnea, shallow breathing, and/or irregular breathing that result from a periodic failure of the central nervous system mechanisms to stimulate breathing. There are three subtypes in DSM-5: idiopathic, Cheyne–Stokes breathing pattern, and comorbid with opioid use. *Compare* obstructive sleep apnea.

cephalalgia Headache.

cerea flexibilitas Condition of a person who can be molded into a position that is then maintained; when an examiner moves the person's limb, the limb feels as if it were made of wax. Also called catalepsy; waxy flexibility. Seen in schizophrenia.

cerebral arteriosclerosis Arteriosclerotic involvement (hardening) of arteries of the brain; may lead to organic mental disorder.

cerebrovascular accident Cerebral hemorrhage, thrombosis, or embolism. Also called stroke; apoplexy.

character Constellation of relatively fixed personality traits and attributes that govern a person's habitual modes of response.

character analysis Psychoanalytic treatment that concentrates on character defenses.

character defense Personality trait that serves an unconscious defensive purpose.

character disorder Pattern of personality characterized by maladaptive, inflexible behavior.

Charcot, Jean M. (1825–1893) French neurologist noted for describing hysteria and treating it by means of hypnosis; Freud based much of his early work on Charcot's pioneering studies in hysteria.

chemical dependence *See* drug dependence.

chemotherapy DRUG THERAPY.

child abuse *See* battered child syndrome.

child analysis Application of psychoanalytic principles to the treatment of the child.

childhood disintegrative disorder A condition defined by marked regression in several functional domains (e.g., intellectual, social, motor, and language skills) after at least 2 years of normal development; diagnosis is typically made in children who are between 3 and 4 years of age. In DSM-5, childhood disintegrative disorder is a subset of autism spectrum disorder.

childhood schizophrenia DSM term for pervasive developmental disorder. *See* autistic spectrum disorder.

child-onset fluency disorder *See* stuttering.

chi square Statistical technique in which variables are categorized in order to determine whether a distribution of scores is due to chance or to experimental factors.

chlorpromazine Aliphatic phenothiazine derivative used primarily as an antipsychotic agent and in the treatment of nausea and vomiting.

cholinergic Relating to nerve fibers, tracts, or pathways in which synaptic transmission is mediated by acetylcholine; for example, parasympathetic nerve fibers and somatic motoneurons; also refers to processes activated by acetylcholine.

chorea Movement disorder characterized by random and involuntary quick, jerky, purposeless movements. *See also* extrapyramidal dyskinesia; Huntington chorea.

chronic fatigue syndrome An illness characterized by 6 months or more of severe, debilitating fatigue, often accompanied by myalgia, headaches, pharyngitis, low-grade fever, cognitive complaints, gastrointestinal symptoms, and tender lymph nodes. The cause is unknown.

chronobiology Study of biologic rhythms that objectively explores and quantifies mechanisms of time structure in biologic systems.

chronologic age (C.A.) *See* intelligence quotient.

chronophobia Fear of time characterized by panic, anxiety, and claustrophobia. Sometimes called prison neurosis, since almost all prisoners are affected by it in some fashion.

circadian rhythm *See* biologic rhythm.

circadian rhythm sleep disorder A set of dyssomnias involving a misalignment between desired and actual sleep periods. Includes delayed sleep phase, jet lag, and work shift.

Circadian Rhythm Sleep–Wake Disorders A group of sleep disorders characterized by disruption in sleep that alters a patient's circadian rhythm and is associated with insomnia or excessive sleepiness. There are six types of circadian rhythm sleep–wake disorders in DSM-5: delayed sleep phase type, advanced sleep phase type, irregular sleep–wake type, non–24-hour sleep–wake type, shift work type, and unspecified type.

circumstantiality Disturbance in the associative thought and speech processes in which a patient digresses into unnecessary details and inappropriate thoughts before communicating the central idea; observed in schizophrenia, obsessional disturbances, and certain cases of dementia. *See also* tangentiality.

civil commitment In statutes governing hospitalization of the mentally ill, a legal term meaning a warrant for imprisonment; as recommended by the American Psychiatric Association and the American Bar Association, "commitment" has been replaced by "hospitalization."

CJD Creutzfeldt–Jakob disease.

claims review Method of peer review consisting of the examination of claims for the reimbursement of treatment after it has been rendered.

clairvoyance Extrasensory perception in which there is an alleged ability to perceive that which is out of sight or cannot be seen.

clang association, clanging Association or speech directed by the sound of a word rather than by its meaning; punning and rhyming may dominate the person's verbal behavior; seen most frequently in schizophrenia or mania.

classical conditioning *See* conditioning.

classic migraine *See* migraine.

claustrophobia Abnormal fear of closed or confining spaces.

client-centered psychotherapy Nondirective form of psychotherapy, originated by Carl Rogers, in which the therapeutic process focuses on the patient's own thinking and feeling which the therapist merely helps to clarify through understanding and empathy; developed as a reaction against the authoritativeness and interpretation of the more traditional psychotherapies.

climacteric Menopause and immediate postmenopausal period in women; sometimes used to refer to the same age period in men.

clinical psychologist Psychologist with additional training and experience in a clinical setting who specializes in evaluation and treatment of mental disorders.

clinical psychology Branch of psychology concerned with the application of psychological principles to the prevention, treatment, and understanding of psychopathology.

clonic convulsion An involuntary, violent muscular contraction or spasm in which the muscles alternately contract and relax. Characteristic phase in grand mal epileptic seizure.

clouding of consciousness Any disturbance of consciousness in which the person is not fully awake, alert, and oriented. Occurs in delirium, dementia, and cognitive disorder.

cluster suicides Suicides that occur together in geographic groups or in temporal association. *See also* copy-cat suicides.

cluttering Disturbance of fluency involving an abnormally rapid rate and erratic rhythm of speech that impedes intelligibility; the affected individual is usually unaware of communicative impairment.

CNS Central nervous system.

cocaine Alkaloid, obtained from leaves of the coca plant, whose systemic effects include striking central nervous system stimulation manifested by garrulousness, restlessness, excitement, and feelings of increased muscular strength and mental capacity; also an effective topical local anesthetic.

coefficient of correlation Statistical term referring to the relation between two sets of paired measurements, correlation coefficients are positive, negative, or curvilinear, depending on whether the variations are in the same direction, the opposite direction, or both directions. They can be computed in a variety of ways, the most common being the product–moment method (r); another method is rank correlation (p). Correlation coefficients are intended to show degree of relation, but they do not imply a causal relationship between variables.

coercive persuasion BRAINWASHING.

coexistent culture Alternative system of values, norms, and patterns for behavior; the group therapy experience often leads to an awareness of other systems as legitimate alternatives to one's own system.

cognition Mental process of knowing and becoming aware; an ego function closely associated with judgment.

cognitive-behavioral therapy (CBT) A form of psychotherapy in which the focus of treatment is on changing the thoughts and behaviors that are related to specific symptoms. The aim of treatment is on symptom reduction and improved functioning.

cognitive development Progressive acquisition of conscious thought and problem-solving abilities that begins in infancy and follows an orderly sequence of more or less discrete stages.

cognitive dissonance Incongruity in a person's beliefs, thoughts, or actions that produces an uncomfortable tension state that the person is motivated to change.

cognitive therapy Structured and directive form of therapy developed by Aaron T. Beck that emphasizes how patients think about themselves and the future; consists of reversing the cognitive triad of negative expression of self, negative expression of future, and negative view of the world.

cognitive triad Belief system described by the psychiatrist Aaron Beck that is present in depressed persons who have a negative view of the self, a negative interpretation of present and past experience, and a negative expectation of the future.

cohesion *See* group cohesion.

cohort Group chosen from a well-defined population; cohort studies provide direct estimates of risk associated with a suspected causal factor.

coitus interruptus Sexual intercourse that is interrupted before the man ejaculates.

cold turkey Abrupt withdrawal from opiates, or any addictive central nervous stimulant, without the benefit of methadone or other drugs; a term originated by drug addicts to describe their chills and consequent goose-flesh. Abstinence-oriented therapeutic communities use this type of detoxification.

collective experience Term coined by S. R. Slavson for the common emotional experiences of a group of people, which, in the setting of group psychotherapy, accelerates the therapeutic process. Identification, mutual support, reduction of ego defenses, sibling transferences, and empathy help integrate the individual member into the group.

collective unconscious Psychic contents outside the realm of awareness that are common to mankind in general. Carl Jung, who introduced the term, believed that the collective unconscious is inherited and derived from the collective experience of the species, that it transcends cultural differences, and that it explains the analogy between ancient mythologic ideas and the primitive projections observed in some patients who have never been exposed to those ideas. Also called objective psyche.

coma State of profound unconsciousness from which a person cannot be roused, with minimal or no detectable responsiveness to stimuli; seen in injury or disease of the brain, in such systemic conditions as diabetic ketoacidosis and uremia, and in intoxications with alcohol and other drugs. Coma may also be seen in severe catatonic states and in hysteria.

coma vigil Coma in which the eyes remain open; typically seen in acute organic mental syndromes associated with systemic infection.

combat fatigue Disabling physical and mental reaction to the stresses of military battle; now called posttraumatic stress disorder.

combined therapy Psychotherapy in which the patient is in both individual and group treatment with the same or two different therapists. In marriage therapy, it is the combination of married couples group therapy with either individual sessions with one spouse or conjoint sessions with the marital pair. *See also* co-therapy.

command automatism Condition closely associated with catalepsy in which suggestions are followed automatically.

command hallucination False perception of orders that a person may feel obliged to obey or unable to resist.

command negativism *See* negativism.

commitment Legal process by which a person is confined to a mental hospital, usually against their will.

common migraine *See* migraine.

communication disorder Term referring to a heterogeneous class of disorders resulting from impairment in the ability to receive, send, process, and comprehend concepts or verbal, nonverbal, and written symbolic systems. Communication disorders consist of language disorders, speech sound disorders, and social (pragmatic) communication disorder.

community *See* therapeutic community.

community mental health center Community or neighborhood mental health facility or a group of affiliated agencies that serve as a locus for the delivery of various services of community psychiatry. *See also* community psychiatry.

community psychiatry Psychiatry focusing on the detection, prevention, and early treatment of mental disorders and social deviance as they develop in the community, rather than as they are perceived and encountered at large centralized psychiatric facilities; particular emphasis is placed on the environmental factors that contribute to mental illness.

compensation Conscious or, usually, unconscious defense mechanism by which a person tries to make up (compensate) for an imagined or real physical or psychological deficiency.

compensation neurosis Neurosis induced by the desire for monetary compensation. It is believed that some persons, after sustaining an injury, develop neurotic symptoms in the hope of monetary gain.

competence Ability to perform or accomplish an action or task that another person of similar background and training, or any human being, could reasonably be expected to perform; mental capacity.

competency to stand trial Ability to be tried in a court of law. A person is competent to stand trial when, at the time of the trial, they (1) understand the nature of the charge and the potential consequences of conviction and (2) is able to assist their attorney in their defense.

competition Literally, the struggle for the possession or use of limited goods, concrete or abstract. In psychiatry, a rivalry that may exist between persons in which gratification for one person largely precludes gratification for the other.

complementarity of interaction Concept of bipersonal and multipersonal psychology in which behavior is viewed as a response to stimulation, and interaction replaces the concept of reaction; each person in an interactive situation plays both a provocative role and a responsive role.

complementary role *See* role.

complex Group of interrelated ideas, mainly unconscious, that have a common emotional tone and strongly influence the person’s attitudes and behavior; introduced by Jung, who called it a feeling-toned idea.

complex partial seizure Epilepsy characterized by recurrent behavioral disturbances. Complex hallucinations or illusions, frequently gustatory or olfactory, often herald the onset of the seizure, which typically involves a state of impaired consciousness resembling a dream (dreamy state) during which paramnestic phenomena, such as déjà vu and jamais vu, are experienced and the patient exhibits repetitive, automatic, or semipurposeful behavior; in rare instances, violent behavior may be prominent. Electroencephalography reveals a localized seizure focus in the temporal lobe. Also called psychomotor epilepsy; temporal lobe epilepsy.

complicated grief/bereavement *See* persistent complex bereavement disorder.

compulsion Uncontrollable, repetitive, and unwanted urge to perform an act; serves as a defense against unacceptable ideas and desires, and failure to perform the act leads to overt anxiety. *See also* obsession.

compulsive personality Personality characterized by rigidity, overconscientiousness, extreme inhibition, inability to relax, and the performance of repetitive patterns of behavior.

computed tomography (CT) Radiologic imaging technique for presenting anatomical features from a cross-sectional plane of the body, generated by computer synthesis of x-ray transmission data obtained from many different directions through the given plane; gray and white matter are indistinguishable on CT scans.

conation That part of a person's mental life concerned with cravings, strivings, motivation, drives, and wishes as expressed through behavior or motor activity.

concordance In studies of twins, the degree of similarity between the two members of each pair with respect to the presence or absence of a particular trait. *Compare* discordance.

concrete operations *See* Piaget, Jean.

concrete thinking Thinking characterized by actual things and events and immediate experience, rather than by abstractions; seen in young children, in those who have lost or never developed the ability to generalize (as in certain organic mental disorder), and in schizophrenic persons. *Compare* abstract thinking.

concussion Immediate and transient impairment of brain function resulting from a head injury; usually, an alteration of consciousness is involved.

condensation Mental process in which one symbol stands for a number of components.

conditioned reinforcer *See* reinforcement.

conditioned response *See* conditioning.

conditioned stimulus *See* conditioning.

conditioning Procedure resulting in the acquisition or learning of more or less permanent changes in behavior. There are two main types of conditioning: **1.** Classical conditioning, in which a neutral stimulus is repeatedly paired with a stimulus (unconditioned stimulus) that naturally elicits a response (unconditioned response) until the neutral stimulus (conditioned stimulus) comes to elicit that response (now the conditioned response) by itself. Also called Pavlovian conditioning and respondent conditioning. **2.** Operant conditioning, a procedure developed by B. F. Skinner in which a spontaneously emitted behavior (operant behavior) is either rewarded (reinforced) or punished and, as a result, then occurs with a frequency that is either increased (in the case of reinforcement) or decreased (in the case of punishment). Also called instrumental conditioning. *See also* behavior therapy.

conditioning therapy *See* behavior therapy.

conditions not attributable to a mental disorder DSM category given to a person who may or may not have a mental disorder but who has a problem which may require intervention by a psychiatrist. Included are marital problem, occupational problem, phase of life or other life circumstance problem, malingering, complicated bereavement, antisocial behavior, academic problem, other interpersonal problem, other specified family circumstance, and noncompliance with medical treatment.

conduct disorder A diagnosis used in child psychiatry to refer to children who show antisocial behavior. Such behavior may occur as part of a group (group type), individually (solitary aggressive type), or both.

confabulation Unconscious filling of gaps in memory by imagining experiences or events that have no basis in fact, commonly seen in organic amnestic syndromes; should be differentiated from lying. *See also* paramnesia.

confidence interval Range of values with a specified probability of including the population mean with a specified level of confidence; typically associated with a certain probability level, for example, the 95% confidence interval has a 95% chance of including the population mean.

confidentiality Ethical principle by which the physician is bound to hold secret all information given to them by the patient. Certain states do not legally recognize confidentiality and can require the physician to divulge such information if needed in a legal proceeding. *See also* privilege.

conflict Mental struggle arising from the clash of opposing impulses, drives, or demands. *See also* extrapsychic conflict; intrapsychic conflict.

conflict-free area Part of one's personality or ego that is well integrated and does not cause any conflicts, symptoms, or displeasures.

confrontation Act of letting a person know where one stands in relationship to that person, what one is experiencing, and how one is perceived. Used in a spirit of deep involvement, this technique is a powerful tool for changing relationships; used as an attempt to destroy another person, it can be harmful. In group and individual therapies, the value of confrontation is likely to be determined by the therapist.

confusion Disturbance of consciousness manifested by a disordered orientation in relation to time, place, or person.

congenital Referring to conditions present at birth, including hereditary conditions and those resulting from prenatal development or the process of birth itself.

conjoint therapy Type of marriage therapy in which a therapist sees the partners together in joint sessions. Also called triadic or triangular therapy (since two patients and one therapist work together). *See also* quadrangular therapy.

conjugal paranoia Paranoia characterized by delusions of jealousy; with no evidence or little evidence, a spouse or lover becomes convinced that their partner is being unfaithful. Also called Othello syndrome.

conscience The part of the self that judges one's own values and performance; often used synonymously with superego.

conscious **1.** One of three divisions of Freud's topographic theory of the mind (the others being the preconscious and the unconscious), referring to that portion of mental functioning within the realm of awareness at all times. **2.** More generally, the state having present knowledge of oneself, one's acts, and one's surroundings and thus a functioning sensorium. **3.** Denoting a state of awareness, with response to external stimuli.

consciousness State of awareness, with response to external stimuli. *See also* sensorium.

consensual validation Continuous comparison of the thoughts and feelings of group members toward one another that tends to modify and correct interpersonal distortions; introduced by Harry Stack Sullivan to refer to the dyadic therapeutic process between doctor and patient.

conservation According to Piaget, the crucial difference between preoperational and concrete operational thought, in which a child discovers which values remain invariant in the course of any kind of change or transformation.

conservator Legal guardian designated to take over and protect the interests of an individual.

constipation Inability to defecate or difficulty in defecating.

constitution One's intrinsic psychological or physical endowment. Broadly, it includes an aggregate of characteristics that have developed from the interaction of hereditary and environmental influences; more narrowly, it indicates characteristics that are purely hereditary or genetically determined.

constitutional types Categories of a person's traits based on physical and psychological characteristics (e.g., ectomorphic, endomorphic, mesomorphic).

constricted affect Reduction in intensity of feeling; tone less severe than that of blunted affect but clearly reduced.

constructional apraxia Inability to copy a drawing, such as a cube, clock, or pentagon, as a result of a brain lesion.

Consultation–liaison psychiatry Clinical psychiatry carried out in the medical or surgical wards of a general hospital; the psychiatrist collaborates closely with nonpsychiatric physicians in the total care of the patient.

contagion Force that operates in large groups or masses. When the level of psychological functioning has been lowered, some sudden upsurge of anxiety can spread through a group, accelerated by high degree of suggestibility, gradually mounting to panic, and the whole group may be simultaneously affected by a primitive emotional upheaval.

content thought disorder Disturbance in thinking in which a person exhibits delusions that may be multiple, fragmented, and bizarre.

continuous reinforcement A schedule of reinforcement in which reward is administered every time a response is emitted.

contract Explicit, bilateral commitment to a well-defined course of action; in group or individual therapy, the therapist–patient contract is to attain the treatment goal.

control group In an experimental design, the group in which a condition or factor being tested is deliberately omitted; for example, in a study measuring the effects of a new drug, the control group may be given a placebo instead of the drug. *Compare* experimental group.

controlled substance Drug that has the potential to be abused and is classified by the Drug Enforcement Administration (DEA) into one of five levels of control: I, high abuse potential and no accepted medical use (e.g., heroin); II, high abuse potential with medical use (e.g., amphetamine); III, moderate abuse potential (e.g., codeine); IV, low abuse potential, that is, preparations containing limited amounts of narcotic and nonnarcotic ingredients (e.g., cough medicines).

controlling Excessive attempt to manage or regulate events or objects in the environment to minimize anxiety and to resolve inner conflicts.

conversational catharsis Release of repressed or suppressed thoughts and feelings in group and individual psychotherapy as a result of verbal interchange. *See also* abreaction; catharsis.

conversion Unconscious defense mechanism by which the anxiety that stems from an intrapsychic conflict is converted and expressed in a symbolic somatic symptom; the hallmark of conversion disorder, but also seen in a variety of mental disorders.

conversion disorder Alternate term for functional neurologic symptom disorder.

conversion phenomena The development of symbolic physical symptoms and distortions involving the voluntary muscles or special sense organs; not under voluntary control and not explained by any physical disorder. Most common in conversion disorder, but also seen in a variety of mental disorders.

convulsion An involuntary, violent muscular contraction or spasm. *See also* clonic convulsion; tonic convulsion.

convulsive disorder *See* epilepsy.

convulsive therapy Psychiatric treatment involving administration of chemical substances or electric current to induce loss of consciousness or a convulsive reaction, to alter favorably the course of a mental disorder; for example, electroconvulsive therapy, insulin coma therapy.

cooperative therapy CO-THERAPY.

coping mechanism Unconscious or conscious way of dealing with stress.

coprolalia Involuntary use of vulgar or obscene language; observed in some cases of schizophrenia and in Tourette syndrome.

coprophagia Eating of filth or feces.

coprophilia Excessive interest in filth or feces. Sexual pleasure associated with feces is a perversion associated with fixations at the anal stage of psychosexual development.

coprophobia Abnormal fear of excreta.

copy-cat suicides Suicides that occur in persons who are geographically related; usually occurs among adolescents within a limited time period. *See also* cluster suicides.

correctional psychiatry Practice of psychiatry on persons incarcerated in jail or prison; presents special problems in relation to confidentiality and countertransference.

corrective emotional experience Re-exposure under favorable circumstances to an emotional situation that the patient could not handle in the past; as advocated by Franz Alexander, the therapist temporarily assumes a particular role to generate the experience and facilitate reality testing.

correlation Statistical index of the relationship among variables. The most common correlation coefficient is the Pearson Product

Moment Correlation, an index of bivariate association that varies between >1.0 and 1.0. *See also* coefficient of correlation.

Cotard syndrome *See* nihilistic delusional disorder.

co-therapy Psychotherapy in which more than one therapist treat the individual patient or the group. Also called cooperative, dual, multiple, or three-cornered therapy. *See also* combined therapy.

counseling Form of guidance in which a trained person offers advice or education about specific problems in living, for example, marriage, child rearing, mourning.

counterphobia State in which a phobic person shows a preference for the very situation that is feared.

countertransference Conscious or unconscious emotional response of the therapist to the patient; determined by the therapist's inner needs, rather than by the patient's needs, and may reinforce the patient's earlier traumatic history if not checked by the therapist. *See also* transference.

couples therapy *See* family therapy; marriage therapy.

covert sensitization Method of reducing the frequency of behavior by associating it with the imagination of an unpleasant consequence.

CPA Canadian Psychiatric Association.

crack Freebase form of cocaine prepared by heating a mixture of cocaine hydrochloride, water, and baking soda, which is then dried into a solid and broken into small pieces (rocks); it is inexpensive, has a rapid onset of action, and is highly addictive.

creativity Ability to produce something new. Silvano Arieti described creativity as the tertiary process, a balanced combination of primary and secondary processes, in which materials from the id are used in the service of the ego.

cretinism Mental and physical retardation resulting from severe thyroid deficiency in infancy and childhood.

Creutzfeldt–Jakob disease (CJD) A rapidly progressive, invariably fatal disorder that occurs mainly in middle-aged or older adults. It initially manifests with fatigue, flu-like symptoms, and cognitive impairment; as it progresses, focal neurologic findings such as aphasia and apraxia occur. Psychiatric symptoms include anxiety, euphoria, depression, delusions, hallucinations, or marked personality changes. Progresses over months and leads to dementia, akinetic mutism, coma, and death.

Cri-du-chat High-pitched catlike cry characteristic of the syndrome of mental and physical retardation caused by partial deletion of chromosome 5. Also called cat's cry syndrome.

criminal responsibility Legal term meaning the ability to formulate a criminal intent at the time of an alleged crime. A person cannot be convicted of a crime if it can be proved that they lacked criminal responsibility by reason of insanity.

crisis In psychiatry, a distressing sense of psychological disequilibrium that has an acute onset and that causes impaired functioning in a variety of interpersonal and vocational areas.

crisis intervention Brief therapeutic approach used in emergency rooms of general or psychiatric hospitals that is ameliorative, rather than curative, of acute psychiatric emergencies; treatment factors often focus on environmental modification, although interpersonal and intrapsychic factors are also considered. *See also* hot line.

critical period Specific period of early development during which an organism acquires certain patterns of behavior as a result of internal or external clues. *See also* imprinting.

critical ratio In a statistical study involving 30 or more subjects, a ratio constructed to determine whether differences found between two items are larger than could be expected from chance alone.

cross-cultural psychiatry Comparative study of mental illness and health among various societies around the world.

cryptographia A private written language.

cryptolalia A private spoken language.

CT scan Computed tomography.

cult System of beliefs and rituals, especially involving devoted attachment to or extravagant admiration of a person or principle, based on dogma or religious teachings; devoted adherents to such a system are called cultists.

cultural anthropology *See* anthropology.

cultural deprivation Restricted participation in the culture of the larger society.

cultural psychiatry Branch of psychiatry concerned with mental illness in relation to the cultural environment; symptoms regarded as psychopathologic in one cultural setting may be acceptable and normal in another.

culture shock Psychological response of a person who is suddenly thrust into an alien culture or has divided loyalties to two different cultures.

culture specific syndrome Disorder of behavior that occurs within or is limited to a specific culture; for example, amok, koro, latah.

cunnilingus Use of the mouth or tongue to stimulate the female genitalia.

Cushing syndrome Hyperadrenocorticism characterized by muscle wasting, obesity, osteoporosis, atrophy of the skin, and hypertension; emotional lability is common and frank psychoses are occasionally observed.

custody Term used to describe a parent's legal rights over a child; in cases of disputed custody, the child is awarded to the party who will best serve the interests of the child.

CVA Cerebral vascular accident.

cybernetics Norbert Wiener's term for the comparative study of computers and the human nervous system.

cybersex addiction A paraphilia characterized by dependence on the internet for sexual gratification (e.g., internet pornography, sex-chat rooms).

cyclazocine Narcotic antagonist that blocks the effects of morphine and heroin.

cycloplegia Paralysis of the muscles of accommodation in the eye; observed at times as an autonomic side effect (anticholinergic effect) of antipsychotic or antidepressant medication.

cyclothymia Mood disorder characterized by a nonpsychotic disturbance of at least 2 years length with numerous periods of both depressive and manic syndromes not severe enough to be classified as depressive or manic episodes.

cyclothymic disorder Mood disorder characterized by a nonpsychotic disturbance of at least 2 years' length with numerous periods of both depressive and hypomanic syndromes that are not severe enough to be classified as depressive or hypomanic episodes.

cynophobia Abnormal fear of dogs.

D

Da Costa syndrome Chest pain, palpitations, and lightheadedness occurring as a somatic manifestation of pathologic anxiety. Also called neurocirculatory asthenia; soldier's heart.

day hospital Form of partial hospitalization in which a patient spends the day in the hospital and returns home at night. *Compare* night hospital; weekend hospital.

day residue In Sigmund Freud's dream theory, the content of the dream imagery that is derived from events of the previous day.

DEA Drug Enforcement Administration.

death instinct AGGRESSIVE DRIVE.

decathexis ACATHEXIS.

decision tree algorithm.

de Clérambault syndrome *See* erotomania.

decoding Technical term for the process of understanding spoken language.

decompensation Deterioration of psychic functioning due to a breakdown of defense mechanisms.

decomposition In psychiatry, the division of a person into separate personalities or identities, as seen in the paranoid schizophrenic person who splits the persecutor into separate entities.

deep brain stimulation A wired electrode with multiple stimulating sites that is surgically implanted into the brain chronically administers electrical stimulation. Use to treat tremor in Parkinson disease.

defense mechanism Unconscious process acting to relieve conflict and anxiety arising from one's impulses and drives.

defensive emotion Strong feeling that serves as a screen for a less acceptable feeling, one that would cause a person to experience anxiety if it appeared. For example, expressing the emotion of anger is often more acceptable to a patient than expressing the fear that the anger covers up; in that instance, anger is defensive.

deficit In psychiatry, the absence of a function or capability that would normally be present, for example, judgment.

deinstitutionalization Discharge from the hospital to community, particularly of those chronically mentally ill patients who would otherwise be hospitalized for long periods of time.

déjà entendu Illusion that what one is hearing one has heard previously. *See also* paramnesia.

déjà pensè Condition in which a thought never entertained before is incorrectly regarded as a repetition of a previous thought.

déjà vu Illusion of visual recognition in which a new situation is incorrectly regarded as a repetition of a previous experience. *See also* paramnesia.

delayed ejaculation DSM-5 term for male orgasmic disorder.

delayed sleep phase circadian rhythm disorder A circadian rhythm sleep–wake disorder characterized by sleep–wake times that are later than conventional times; individuals with this disorder are referred to as “night owls.”

delirium Acute reversible organic mental disorder characterized by confusion and some impairment of consciousness; generally associated with emotional lability, hallucinations or illusions, and inappropriate, impulsive, irrational, or violent behavior.

delirium tremens (DTs) Acute and sometimes fatal reaction to withdrawal from alcohol, usually occurring 72 to 96 hours after the cessation of heavy drinking; distinctive characteristics are marked autonomic hyperactivity (tachycardia, fever, hyperhidrosis, dilated pupils), usually accompanied by tremulousness, hallucinations, illusions, and delusions. Called alcohol withdrawal delirium in DSM. *See also* formication.

delta sleep SLOW WAVE SLEEP.

delta waves As measured by an electroencephalogram (EEG), brain waves that are less than 4 Hz, indicative of deep sleep.

delusion False belief that is firmly held, despite objective and obvious contradictory proof or evidence and despite the fact that other members of the culture do not share the belief. Types of delusion include: **bizarre delusion,** one that is patently absurd or fantastic; **delusion of control,** that a person's thoughts, feelings, or actions are not one's own but are being imposed by some external force; **delusion of grandeur (grandiose delusion),** exaggerated concept of one's importance, power, knowledge, or identity; **delusion of jealousy (delusion of infidelity),** that one's lover is unfaithful; **delusion of persecution,** that one is being attacked, harassed, cheated, or conspired against; **delusion of poverty,** that one is or will be without material possessions; **delusion of reference,** that events, objects, or the behavior of others have a particular and unusual meaning specifically for oneself; **encapsulated delusion,** one without significant effect on behavior; fragmentary delusion, poorly elaborated delusion, often one of many with no apparent interconnection; **nihilistic delusion, delusion of negation** (Cotard syndrome), that the world and everything related to it have ceased to exist; **paranoid delusion,** combined delusion of grandeur and persecution; **religious delusion,** one involving the Deity or theologic themes; **sexual delusion,** one centering on sexual identity, appearances, practices, or ideas; **somatic delusion,** pertaining to the functioning of one's body; **systematized delusion,** a group of elaborate delusions related to a single event or theme.

delusional depression Depressive disorder characterized by false, fixed ideas, that is, delusions; listed in DSM as major depression, severe, with psychotic features.

delusional (paranoid) disorder Psychotic disorder in which there are persistent delusions (e.g., erotomanic, grandiose, jealous, persecutory, somatic, mixed, and unspecified); symptoms of a major mood disorder are absent, and the delusions lack the bizarre quality often seen in schizophrenia.

dementia Mental disorder characterized by general impairment in intellectual functioning, frequently characterized by failing memory, difficulty with calculations, distractibility, alterations in mood and affect, impairment in judgment and abstraction, reduced facility with language, and disturbance of orientation. Although generally irreversible because of underlying progressive, degenerative brain disease, dementia may be reversible if the cause can be treated. Called major neurocognitive disorder in DSM-5. *Compare* amentia.

dementia associated with alcoholism DSM term for alcoholic deterioration.

dementia of the Alzheimer type Term for a clinical disorder in an older adult patient characterized by a severe loss of intellectual function, for which no other cause is found, and that is progressive and insidious and marked by disorientation as to time, place, and person; senile onset is after age 65 (senile dementia); presenile onset is before age 65 (presenile dementia). Also called Alzheimer disease. *See also* Pick disease.

dementia praecox (precox) Obsolete term for schizophrenia.

demography Statistical study of populations, including births, marriages, mortality, health, geographic distribution, and population shifts.

demophobia Abnormal fear of crowds.

dendrite Projection from the nerve cell body, usually profusely branched and studded with small spikes (dendritic spines) that are the sites of synaptic connection; most dendrites receive neurotransmissions from an axon.

denial Defense mechanism in which the existence of unpleasant realities is disavowed; refers to a keeping out of conscious awareness any aspects of either internal or external reality that, if acknowledged, would produce anxiety.

dependence on therapy Pathologic need of a patient for therapy, created out of the belief that one cannot survive without it.

dependency State of reliance on another, as for security, love, protection, or mothering.

dependent personality disorder Personality disorder characterized by lack of self-confidence, a tendency to have others assume responsibility for significant areas of one's life, and a subordination of one's own needs and wishes to those of the others on whom one is dependent; solitude is extremely discomforting.

dependent variable In research methodology, those qualities that measure the influence of the independent variable or the outcome of the experiment; for example, measurement of a person's specific physiologic reactions to a drug. Also called outcome variable.

depersonalization Called depersonalization disorder in DSM-5, a disordered sensation of unreality concerning oneself, parts of oneself, or one's environment that occurs under extreme stress or fatigue. Seen also in schizophrenia and schizotypal disorder.

depersonalization disorder [Omit entry.]

depression Mental state characterized by feelings of sadness, loneliness, despair, low self-esteem, and self-reproach; accompanying signs include psychomotor retardation or at times agitation, withdrawal from interpersonal contact, and vegetative symptoms such as insomnia and anorexia. The term refers either to a mood that is so characterized or to a mood disorder.

depressive neurosis DYSTHYMIA.

depth psychology Psychology that focuses on unconscious processes, rather than conscious processes.

derailment Gradual or sudden deviation in train of thought without blocking; sometimes used synonymously with loosening of association.

derealization Called derealization disorder in DSM-5, disordered sensation of a changed reality or that one's surroundings have altered; also seen in schizophrenia.

dereism Mental activity that follows a totally subjective and idiosyncratic system of logic and fails to take the facts of reality or experience into consideration. *See also* autistic thinking.

dereistic thinking Mental activity not concordant with logic or experience.

descriptive psychiatry System of psychiatry focusing primarily on the study of observable symptoms and behavioral phenomena, rather than underlying psychodynamic processes; Emil Kraepelin's systematic description of mental illness was an early example. *See also* dynamic psychiatry.

descriptive statistics Methods used to summarize, organize, and describe observations; for example, the mean, standard deviation, and variance.

desensitization *See* systematic desensitization.

designer drug Any version of a regulated drug whose chemical structure has been modified so that it falls outside the restrictions placed on the regulated drug it imitates; usually a drug of abuse used to alter states of consciousness or to produce pleasurable feelings.

detachment A state of being in which the person is separated or divorced from emotional involvement with others, characterized by distant interpersonal relationships.

determinism Concept that nothing occurs by chance alone; instead, things result from specific causes or forces. This school of thought denies the notion of free will.

detoxification Any treatment aimed at the restoration of normal physiologic function after a disturbance by the use of some drug or chemical substance. *See also* cold turkey.

detumescence Subsiding of a penile erection; also used to describe the subsiding of genital engorgement in both sexes.

devaluation Defense mechanism in which a person attributes excessively negative qualities to self or others. Seen in depression and paranoid personality disorder.

developmental arithmetic disorder Academic skills disorder characterized by marked impairment in the development of arithmetic skills that interferes with academic achievement or other activities of daily living. Also called acalculia; Gerstmann–Straussler–Scheinker disease, congenital arithmetic disability; dyscalculia; arithmetic disorder.

developmental articulation disorder *See* speech sound disorder.

developmental coordination disorder Motor skills disorder of childhood with significant impairment in the development of fine or gross motor coordination that interferes with academic achievement or with activities of daily life.

developmental disorders Disorders originating in infancy and childhood that continue in adulthood and which interfere with the person’s ability to function normally. In DSM, called neurodevelopmental disorders and include intellectual developmental disorder (formally known as mental retardation) and autistic spectrum disorders.

developmental expressive language disorder Language and speech disorder in which there is marked impairment in the development of expressive language; interferes with academic achievement or with activity of daily living that require the expression of verbal (or sign) language.

developmental expressive writing disorder Academic skills disorder characterized by marked impairment in the development of expressive writing skills; significantly interferes with academic achievement or with activities of daily living that require expressive writing skills.

developmental reading disorder Academic skills disorder characterized by omissions, distortions, and substitutions of words and by slow, halting reading; there is marked impairment in the development of words recognition skills and reading comprehension. Also called specific reading disability and developmental word blindness.

developmental receptive language disorder Language and speech disorder in which there is marked impairment in the development of language comprehension which significantly interferes with academic achievement or with activities of daily living that require comprehension of verbal (or sign) language.

developmental word blindness DEVELOPMENTAL READING DISORDER.

deviation Departure from the average or norm.

Dexamethasone Suppression Test (DST) Test used to confirm a diagnostic impression of major depression. The patient is given 1 mg of dexamethasone (a long-acting synthetic glucocorticoid) by mouth at 11 PM, and plasma cortisol is measured at 8 AM, 4 PM, and

11 PM. Plasma cortisol above 5 μg/dL (nonsuppression) is considered abnormal, that is, positive; suppression of cortisol indicates that the hypothalamic–adrenal–pituitary axis is functioning properly.

***Diagnostic and Statistical Manual of Mental Disorders* (DSM)** Handbook for the classification of mental illnesses that is primarily descriptive and does not describe the theories or causes of disorders; first issued by the American Psychiatric Association in 1952 (DSM-I). The fourth edition (DSM-IV) was issued in 1994 and the text revision of fourth edition (DSM-IV-TR) was published in 2000. The DSM-5 was published in 2013 and its text revision (DSM-5-TR) was published in 2022.

dialectical behavior therapy (DBT) A form of therapy used for patients with borderline personality disorder and parasuicidal behavior that is eclectic using advice-giving, cognitive training, story-telling, and confrontation among many other techniques.

diathesis-stress paradigm In psychopathology, a view that abnormal behavior is a function of a series of stresses in predisposed individuals (due to hereditary, temperamental, and sociocultural factors).

differential diagnosis Determination of which of two or more diseases with similar symptoms is the one that afflicts the patient.

differential reinforcement Procedure in which desirable behavior is reinforced and problematic or less desirable behavior is extinguished or punished.

differentiation First of Margaret Mahler’s four subphases of the separation–individuation process, occurring from ages 5 to 10 months, when the child is able to distinguish between self and other objects.

2,5-dimethoxy-4-methylamphetamine (DOM) Synthetic (derivative) drug with hallucinogenic properties. Also colloquially called STP.

dimethyltryptamine (DMT) Tryptamine congener that acts as a psychedelic capable of producing hallucinations, other disorders of perception, and altered moods and states of consciousness.

diminished libido Decreased sexual interest and drive.

dipsomania An obsolete term that refers to the compulsion to drink alcoholic beverages.

disability Defined by the federal government as: "Inability to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment which can be expected to last or has lasted for a continuous period of not less than 12 full months."

disconnection syndrome Group of neurologic disorders produced after interruption of connection fibers from the cerebrum to other centers of the nervous system.

discontinuation syndrome Signs and symptoms that occur when a person stops taking a drug either suddenly or gradually when the drug previously has caused physical or psychological dependence to occur. Also called abstinence syndrome; withdrawal syndrome.

discordance In studies of twins, the degree of dissimilarity between the two members of each pair with respect to the presence or absence of a particular trait. *Compare* concordance.

discriminant analysis Multivariate method for finding the relationship between a single discrete outcome and a linear combination of two or more predictors.

discrimination Process of distinguishing between different but often similar stimuli and reacting appropriately; more specifically, the occurrence of operant behavior in environments in which the behavior has been reinforced but not in those in which it has not been reinforced. *See also* discriminative stimulus.

discriminative stimulus Environmental event correlating with the operation of a certain reinforcement matrix, such that behavior that has been reinforced in the presence of the stimulus but not in its absence tends to occur only when the stimulus is present.

disinhibition 1. Removal of an inhibitory effect, as in the reduction of the inhibitory function of the cerebral cortex by alcohol. **2.** In psychiatry, a greater freedom to act in accordance with inner drives or feelings and with less regard for restraints dictated by cultural norms or one's superego.

disinhibited social engagement disorder A disorder of markedly disturbed social relatedness in which a child approaches and interacts with unfamiliar adults in an overly familiar manner; presumed to result from grossly pathologic care-giving.

disintegration Psychic disorganization in which a person is unable to recognize reality and may develop delusions, hallucinations, or overwhelming anxiety and panic.

disorder of written expression A learning disorder characterized by writing skills that are significantly below the expected level for a child's age and intellectual capacity.

disorders of excessive somnolence Disorders characterized by excessive amounts of sleep or excessive daytime sleepiness. *See also* somnolence.

disorders of initiating and maintaining sleep Disorders characterized by inability to fall asleep or repeated awakening after failing asleep. *See also* insomnia disorder; dyssomnias.

disorganized schizophrenia Type of schizophrenia characterized by markedly disinhibited and aimless, nonconstructive behavior; no longer included in DSM-5.

disorientation Confusion; impairment of awareness of time, place, and person (the position of the self in relation to other persons); characteristic of cognitive disorders.

displacement Unconscious defense mechanism by which the emotional component of an unacceptable idea or object is transferred to a more acceptable one. Seen in phobias.

disposition Sum total of a person's inclinations as determined by their mood.

disruptive behavior disorders Disorders characterized by inattention, overagressiveness, delinquency, destructiveness, hostility, feelings of rejection, negativism, or impulsiveness. In DSM, they include attention-deficit hyperactivity disorder, conduct disorder, and oppositional defiant disorder.

disruptive mood dysregulation disorder A depressive disorder in individuals between 6 and 18 years of age that is characterized by chronic irritability, severe temper tantrums, and angry mood; a DSM-5 diagnosis.

dissociation Unconscious defense mechanism involving the segregation of any group of mental or behavioral processes from the rest of the person’s psychic activity; may entail the separation of an idea from its accompanying emotional tone, as seen in dissociative and conversion disorders.

dissociation syndrome Functional inability of the brain to process information from one area to another; first described by neurologist Norman Geschwind. *See also* disconnection syndrome.

dissociative amnesia Disorder characterized by memory loss of important autobiographical information that is usually stressful or traumatic in nature.

dissociative disorders Mental disorders characterized by a sudden temporary alteration in consciousness, identity, or motor behavior. In DSM, they include dissociative amnesia, dissociative fugue, dissociative identity disorder, and depersonalization disorder. *See also* hysteria.

dissociative identity disorder Dissociative reaction in which a person has two or more distinct personalities, most of them knowing nothing of the others; a DSM-5 diagnosis. Previously called multiple personality disorder.

dissociative trance disorder Disorder characterized by temporary alteration in the state of consciousness or personal identity without replacement by an alternate identity; may occur during the course of a psychotic disorder or result from substance use. Not a DSM diagnosis.

distortion 1. Misrepresentation of reality. **2.** In psychoanalysis, the process of modifying unacceptable unconscious mental elements so that they are allowed to enter consciousness in a more acceptable but disguised form.

distractibility Inability to focus one’s attention; the patient does not respond to the task at hand but attends to irrelevant phenomena in the environment.

distractible speech Pattern of speech in which a person changes subjects in response to nearby stimuli.

distributive analysis and synthesis Therapeutic application of Adolf Meyer’s psychobiologic school of psychiatry; systematic investigations of patients’ entire past experiences yield constructive syntheses of their assets and liabilities and lead to an effort to enable them to adapt to their environment.

disulfiram Drug used in the treatment of alcoholics. By altering the metabolism of alcohol, it produces unpleasant physical sensations that discourage the patient from further consumption of alcohol (trade name, Antabuse).

Dix, Dorothea Lynde (1802–1887) Former schoolteacher who became a prominent advocate for the indigent mentally ill. Her efforts helped create the first generation of American mental asylums in improving the humane treatment of the mentally ill.

dizygotic twins Twins (fraternal) who develop from two separately fertilized ova; in contrast, monozygotic (identical) twins develop from a single fertilized ovum.

DMT Dimethyltryptamine.

Doctor–patient relationship Human interchange that exists between the person who is sick and the person who is selected to heal because of training and experience.

DOES Disorders of excessive somnolence.

DOM 2,5-Dimethoxy-4-methylamphetamine.

dominance **1.** Disposition to play a commanding, controlling, or prevailing role over others. **2.** In neurology, the control of language function by one side of the cerebral cortex. **3.** In genetics, the ability of a gene (dominant gene) to be expressed in the phenotype of the bearer to the exclusion of another gene (recessive gene) with which it is paired.

dopamine Biogenic amine (catecholamine) that inhibits prolactin; functional integrity of the dopaminergic system is impaired in depression and in Parkinson disease.

dopamine hypothesis Theory that there may be an excess of dopamine in the brains of persons with schizophrenia.

dopamine-receptor agonists Drugs that activate dopamine receptors in the absence of dopamine. Commonly used in the treatment of Parkinson disease.

dopamine-receptor antagonists (DRAs) Drugs that block the receptors for dopamine in the brain. Also known as first-generation or typical antipsychotics. *Compare* serotonin-dopamine antagonists.

double bind As formulated by Gregory Bateson, a situation involving the simultaneous communication of conflicting messages from one person to another, in which the response to either message is met with disapproval or rejection; believed to be a characteristic mode of interaction in some families with a schizophrenic member.

double-blind study Study in which, for example, one or more drugs and a placebo are compared in such a way that neither the patient nor the persons directly or indirectly involved in the study know which is being given to the patient; the drugs being investigated and the placebo are coded for identification.

double depression A mood disorder that occurs when the criteria for major depressive disorder and dysthymia are met simultaneously.

double personality *See* multiple personality disorder.

Down syndrome Syndrome of mental retardation associated with a variety of abnormalities caused by replication of chromosome 21 three times. Also called trisomy 21 syndrome. Formerly called mongolism.

DRA Dopamine-receptor antagonist.

Draw-A-Person (DAP) Test Easily administered test used with both children and adults, usually with the instructions, “I’d like you to draw a picture of a person; draw the best person you can.” A general assumption is that the drawing of a person represents the expression of the self or of the body in the environment.

drawing test Any of a variety of psychological tests in which the subject is asked to draw certain familiar objects, such as people, trees, and houses; attitudes and feelings are often revealed in the way the subject depicts those objects.

dread Massive or pervasive anxiety, usually related to a specific danger.

dream Mental activity during sleep in which thoughts, emotions, and images are experienced as though real; associated with rapid eye movement (REM) sleep. Dreams were regarded by Freud as providing an outlet for the discharge in disguised form of unconscious and often unacceptable impulses and wishes.

dream anxiety disorder *See* nightmare disorder.

dreamy state Altered state of consciousness, likened to a dream situation, that develops suddenly and from which the patient usually recovers within a few minutes; accompanied by visual, auditory, and olfactory hallucinations and most commonly associated with temporal lobe lesions.

drive Hypothetical construct used to explain motivated behavior: a basic urge that produces a state of psychic tension that motivates the person into action to alleviate the tension; currently preferred over Freud’s term “instinct.” *See also* aggressive drive; sexual drive.

drowsiness State of impaired awareness associated with a desire or inclination to sleep.

drug abuse Use of any drug, usually by self-administration, in a manner that deviates from approved social or medical patterns.

drug dependence Habituation or addiction to the use of a drug or chemical substance, with or without physical dependence; physical dependence indicates an altered physiologic state due to repeated administrations of a drug, the cessation of which results in the appearance of a withdrawal or abstinence syndrome characteristic for the particular drug which is called discontinuation syndrome.

Drug Enforcement Administration (DEA) Agency of the federal government that regulates distribution of controlled substances such as opioids, barbiturates, and other drugs of potential abuse.

drug holiday Limited period of time during which the administration of a psychoactive drug is discontinued in order to control dosage and side effects or to evaluate baseline behavior.

drug-induced parkinsonism PSEUDOPARKINSONISM.

drug interaction Effect that one drug has on the actions of one or more other drugs given at the same time; effects of either drug may be increased, diminished, or inactivated as a result of the interaction.

drug levels Blood serum or plasma level of psychotropic drugs. *See also* pharmacokinetics.

drug therapy Use of chemical substances in the treatment of a disease or disorder. Also called chemotherapy.

drug tolerance *See* tolerance.

DSM *Diagnostic and Statistical Manual of Mental Disorders.*

DST Dexamethasone Suppression Test.

DTs Delirium tremens.

dual-sex therapy Specific form of psychotherapy developed by William Masters and Virginia Johnson in which treatment is focused on a particular sexual disorder. The crux of the program is the round-table session, in which both a male therapist and a female therapist are present with the patient couple; special exercises are prescribed for the couple, the overall goal being to diminish fears of sexual performance and to facilitate communication in sexual and nonsexual areas.

dual therapy CO-THERAPY.

dummy Uncommonly used term for placebo; for example, dummy drug.

Durham rule Federal court ruling in 1954 holding that “an accused is not criminally responsible if his unlawful act was the product of mental disease or mental defect.” It has since been replaced by the American Law Institute formulation of insanity, which states that “a person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to appreciate the wrongfulness of his conduct or to conform his conduct to the requirements of law.” *See also* M’Naghten rule.

dyad Pair of persons in an interactional situation, such as husband and wife, mother and father, co-therapists, or patient and therapist.

dyadic session Psychotherapeutic session involving only two persons, the therapist and the patient.

dynamic psychiatry System of psychiatry concerned with the mental mechanisms and emotional processes that govern and motivate human behavior, in contradistinction to descriptive psychiatry, which focuses on observable behavioral phenomena.

dynamic reasoning Franz Alexander's term for the formation of all the clinical evidence gained from free-associative anamnesis into a psychological reconstruction of the patient's development.

dynamics *See* psychodynamics.

dysarthria Impairment of articulation, the motor activity of shaping phonated sounds into speech.

dyscalculia Difficulty in performing calculations.

dysdiadochokinesis *See* adibdochokinesis.

dysgeusia Impairment of the sense of taste.

dysgraphia Difficulty in writing.

dyskinesia Any disturbance of movement, as in extrapyramidal dyskinesia.

dyslalia Faulty articulation due to structural abnormalities of the articulatory organs or impaired hearing.

dyslexia A specific learning disability syndrome involving an impairment of the ability to read; unrelated to the person's intelligence. *Compare* alexia.

dysmetria Impaired ability to gauge distance relative to bodily movements.

dysmnesia Impaired memory.

dysmorphophobia BODY DYSMORPHIC DISORDER.

dyspareunia Physical pain in sexual intercourse, usually emotionally caused and more commonly experienced by women; may also result from cystitis, urethritis, or other medical conditions. *See also* Genitopelvic pain/penetration disorder.

dysphagia Difficulty in swallowing.

dysphasia Difficulty in comprehending oral language (receptive dysphasia) or in trying to express verbal language (expressive dysphasia).

dysphonia Difficulty or pain in speaking.

dysphoria Feeling of unpleasantness or discomfort; a mood of general dissatisfaction, restlessness, depression, and anxiety.

dysphoric mood *See* dysphoria.

dysprosody Loss of normal speech melody (prosody).

dyssocial behavior Former term for antisocial behavior.

dyssomnias Disturbances in the amount, quality, or timing of sleep. *See* insomnia disorder, hypersomnia disorder, and circadian rhythm disorder. *Compare* parasomnias.

dysthymia Mood disorder characterized by chronic depressed mood for at least 2 years that is not severe enough to meet criteria for major depressive episode. Also called depressive neurosis. *See also* depression.

dystonia Extrapyramidal motor disturbance consisting of slow, sustained contractions of the axial or appendicular musculature; one movement often predominates, leading to relatively sustained postural deviations; acute dystonic reactions (facial grimacing, torticollis) are occasionally seen with the initiation of antipsychotic drug therapy.

E

early ejaculation DSM-5 term for premature ejaculation.

eating disorders Disorders characterized by a marked disturbance in eating behavior; they include anorexia nervosa, bulimia nervosa, pica, and rumination disorder of infancy.

echolalia Repetition of another person's words or phrases, as observed in certain cases of schizophrenia, particularly the catatonic types. The behavior is considered by some authors to be an attempt by the patient to maintain a continuity of thought processes.

echopraxia Repetition of another person's movements, as observed in some cases of catatonic schizophrenia.

ecology Study of the mutual relationships between living things and their environment; includes such studies as the differential incidence of mental disorders in various populations and the distribution of crime and delinquency within a specified geographical area.

ecopsychiatry Study of genetic data and hypotheses about the role of social factors in mental disease.

ecstasy **1.** State of rapturous delight. **2.** Slang term for methylenedioxymethamphetamine (MDMA).

ECT Electroconvulsive therapy.

ectomorphic Thin, characterized by a linear body musculature; one of Sheldon's constitutional types. *Compare* endomorphic; mesomorphic.

educable Capable of achieving a fourth- or fifth-grade academic level; describes the mildly mentally retarded (LQ of 50–55 to 70). *Compare* trainable.

educational psychology Area of psychology concerned with the application of psychological theories to teaching.

EEG Electroencephalogram.

ego One of the three components of the psychic apparatus in the Freudian structural framework, the others being the id and the superego. It occupies a position between the primal instincts and the demands of the outer world, therefore mediating between the person and external reality. In so doing, it performs the important functions of perceiving the needs of the self, both physical and psychological, the qualities and attitudes of the environment. It evaluates, coordinates, and integrates those perceptions so that internal demands can be adjusted to external requirements. It is also responsible for certain defensive functions to protect the person against the demands of the id and the superego. Although the ego has some conscious components, many of its operations are automatic; adaptation to reality is perhaps its most important function.

ego-alien Denoting aspects of a person's personality that are viewed as repugnant, unacceptable, or inconsistent with the rest of the personality. Also called ego-dystonic. *Compare* ego-syntonic.

ego analysis Psychoanalytic study of the manner in which the ego operates, especially in relation to intrapsychic conflicts. *See also* ego psychology.

ego boundaries Concept introduced by Paul Federn that refers to the ability of the intact ego to differentiate the real from the unreal. Ego boundaries prevent repressed unconscious material from overwhelming the ego; when that happens, a person experiences depersonalization and derealization. Boundaries are said to be weakened in schizophrenia and dissociative states.

egocentric Self-centered; selfishly preoccupied with one's own needs; lacking interest in others.

ego-coping skill Adaptive method or capacity developed by a person to deal with or overcome a psychological or social problem.

ego defense Defense mechanism to protect the person against the demands of the id and the superego.

ego-dystonic EGO-ALIEN.

ego-dystonic homosexuality Controversial term for the purported state of a person who has unwanted and distressful homosexual arousal and wishes to acquire or increase heterosexual arousal. Removed from DSM for the DSM-III-R revision of the book.

ego ideal Part of the personality representing a fusion of the ego and the superego; feelings of self-satisfaction are experienced when the person's thoughts, behavior, or impulses coincide with an internalized goal or ideal.

egomania Morbid self-preoccupation or self-centeredness. *See also* narcissism.

ego model Person after whom one patterns the ego.

ego psychology Study of the operations of ego in regard to control and regulation of instinctual drives, the relation to reality, and the development of object relations. *See also* ego analysis.

ego strength Degree to which the ego's functions, for example, perception, thinking, and speaking, are maintained.

ego-syntonic Denoting aspects of a person's personality that are viewed as acceptable and consistent with their total personality. *Compare* ego-alien.

eidetic image Unusually vivid or exact mental image seen or imagined event; may include color or movement.

ejaculation Expulsion of semen and seminal fluid from the urethra, as in male orgasm.

ejaculatory incompetence INHIBITED MALE ORGASM.

elaboration Unconscious process of embellishing the symbolic content of a dream.

elation Affect consisting of feelings of joy, euphoria, triumph, intense self-satisfaction, or optimism.

elective mutism Childhood mental disorder in which a child who is able and willing to speak to selected persons persistently refuses to speak in other social or school situations.

Electra complex Obsolete and controversial term for a supposed female analog of the Oedipus complex in the male; infrequently used term to describe unresolved developmental conflicts influencing a woman's relationships with men.

electroconvulsive therapy (ECT) Convulsive therapy, usually for depression, that involves the application of electric current to the brain for a fraction of a second through scalp electrodes to induce a convulsive reaction and unconsciousness.

electroencephalogram (EEG) Recording of the brain's electrical potentials obtained by scalp electrodes.

electromyogram (EMG) Recording of the electrical activity of muscles at specific sites.

electroshock therapy (EST) *See* electroconvulsive therapy.

electrostimulation In aversion therapy, a negative conditioning technique consisting of (usually painful) electric shocks.

elevated mood Air of confidence and enjoyment; a mood more cheerful than normal but not necessarily pathologic.

elimination disorders Inability as a result of physiologic or psychological immaturity to maintain bowel control (functional encopresis) or bladder control (functional enuresis).

elopement Escape; absenting oneself from a mental hospital without permission.

emancipated minor Legal term for a minor who exercises general control over their life and hence has the legal rights of an adult.

EMG Electromyogram.

emotion Complex feeling state with psychic, somatic, and behavioral components; external manifestation of emotion is affect.

emotional deprivation Relative lack of environmental or interpersonal experiences during the early developmental years. *See also* sensory deprivation.

emotional disorder Lay term for a mental illness or mental disorder; in common usage it does not specifically imply a mood disorder or schizophrenia.

emotional disturbance Emotional disorder or illness. *See also* mental disorder.

emotional illness General term for a behavioral or psychological syndrome or pattern associated with distress, disability, or increased risk of suffering.

emotional insight Deeper level of understanding or awareness that is more likely to lead to positive changes in personality and behavior.

emotional lability Excessive emotional responsiveness characterized by unstable and rapidly changing emotions.

emotional support Encouragement, hope, and inspiration given to one person by another.

empathy Intellectual and emotional awareness and understanding of another person's state of mind, involving the projection of oneself into another person's frame of reference; an important ability in a successful therapist or a helpful group member. *Compare* sympathy.

empty nest syndrome Depression that can occur in men and women when their youngest child is about to leave or has left home.

encapsulated delusion Delusion that has no significant effect on behavior.

encephalitis Inflammation of the brain.

encephalopathy Any disease of the brain.

encopresis Repeated passage of feces in inappropriate places, such as clothing or bed; diagnosis not made until after age 4.

encounter group Form of sensitivity training, introduced and developed by Jacob Moreno in 1914, that emphasizes the experiencing of individual relationships within the group and that minimizes intellectual and didactic input; the group focuses on the present, rather than concerning itself with the past or outside problems of its members. *See also* sensitivity training group.

endemic Belonging, native, restricted, or peculiar to a particular people, country, locality, or region.

endocrine disorder Dysfunction of any of the endocrine glands, frequently with psychiatric manifestations.

endogamy Marriage within a specific group. *Compare* exogamy.

endogenous depression Obsolete term for a type of depression that originates within the person and is not related to external events, with a cause that is biologic or somatic. Also called nonreactive depression.

endomorphic Obese, with prominence of the abdomen; one of Sheldon's constitutional types. *Compare* ectomorphic; mesomorphic.

endorphin Any of several endogenous opioid peptides (α-, β-, or γ-endorphin) found in highest concentrations in the pituitary gland and which has analgesic and behavioral effects in animals. *See also* enkephalin.

engram Memory trace, the neurophysiologic substrate of memory.

enkephalin Endogenous opioid peptide released by β-endorphin that has analgesic properties and many of the actions of morphine.

entomophobia Fear of insects.

enuresis Repeated incontinence of urine in inappropriate places, such as clothing or bed; diagnosis not made until after age 5.

epidemiology In psychiatry, the study of the incidence, prevalence, control, and distribution of mental disorders within a particular population.

epigenesas Term introduced by Erik Erikson to refer to the stages of ego and social development.

epilepsy Neurologic disorder resulting from a sudden, excessive, disorderly discharge of neurons in either a structurally normal or a diseased cerebral cortex; characterized by paroxysmal recurrence of short-lived disturbances of consciousness, involuntary convulsive muscle movements, psychic or sensory disturbances, or some combination thereof. Termed idiopathic epilepsy when there is no identifiable organic cause. For types of epilepsy, see the specific term.

epileptic dementia Epilepsy accompanied by progressive mental and intellectual impairment; some believe that the circulatory disturbances during epileptic attacks cause nerve cell degeneration and lead to dementia.

epileptic equivalent Epileptic phenomenon that is neither grand mal nor petit mal. *See also* complex partial seizure.

epileptic personality EXPLOSIVE PERSONALITY.

epinephrine Sympathomimetic catecholamine formed from norepinephrine and the major hormone secreted by the adrenal medulla; its release during states of fear or anxiety produces many of the physiologic changes associated with those emotions. Also called adrenaline.

episodic Denoting an illness in which there are sustained disturbances clearly distinguished from previous functioning.

epistemology Study or theory of the nature and grounds of knowledge, especially with reference to its limits and validity.

erectile dysfunction MALE ERECTILE DISORDER.

eremophobia Abnormal fear of being by oneself.

ergasia Adolf Meyer’s term for a person’s total activity, as opposed to the functioning of parts of the whole.

Erikson, Erik (1902–1990) Psychoanalyst noted for his theory of ego development and psychosocial development, which he conceptualized in terms of social adaptation as it relates to Sigmund Freud’s formulations.

erogenous zone Any area of the body, principally the genitals, anus, and mouth, capable of being stimulated sexually.

eros SEXUAL DRIVE.

erotic Invested with sexual feeling.

erotomania Delusional conviction that another person is in love with them; typically the object of the delusion is of higher rank or status. Also referred to as de Clérambault syndrome or psychose passionelle.

erythrophobia Abnormal fear of blushing.

escape learning *See* avoidance and escape learning.

ESP Extrasensory perception.

EST Electroshock therapy.

estrogen Hormone produced mainly by the ovaries and responsible for the development of the female reproductive organs and secondary sex characteristics: for example, estradiol, estriol, estrone. *Compare* androgen.

ethnocentrism Conviction that one's own group is superior to other groups; impairs one's ability to evaluate members of another group realistically or to communicate with them on an open, equal, and person-to-person basis.

ethnology Branch of anthropology dealing with the division of mankind into races and their origin, history, customs, and institutions.

ethology Study of animal behavior; often applied to the understanding of human behavior.

etiology Study of the causes of disease and their mode of operation.

euergasia Word used by Adolf Meyer to mean normal mental functioning.

euphoria Exaggerated feeling of well-being that is inappropriate to apparent events; may be induced by drugs such as opiates, amphetamines, and alcohol.

euphoric mood *See* euphoria.

euphorohallucinogen Substance capable of producing euphoric hallucinations.

euthymia Normal range of mood, implying absence of depressed or elevated mood.

euthymic Denoting a normal mood.

evasion Act of not facing up to or of strategically eluding something; consists of suppressing an idea that is next in a thought series and replacing it with another idea closely related to it. Also called paralogia; perverted logic.

event-related euphoria Morbid or abnormal sense of well-being associated with a specific situation or event.

evoked potential (EP) Following repetitive stimuli, the characteristic EEG waveform that remains after averaging out most of the brain electrical activity; clinically useful in evaluating the functional integrity of the somatosensory or special sensory pathways.

exaltation Feeling of intense elation and grandeur.

excited Agitated, purposeless motor activity uninfluenced by external stimuli.

excitement phase Stage in the human sexual response after sexual stimulation, as described by Masters and Johnson, characterized by an objective sense of pleasure and accompanying physical changes (penile erection and vaginal lubrication).

excitoxicity Increased simulation of neurons due to injury or chemical neurotransmitters that cause cell death.

excoriation disorder Disorder characterized by compulsive picking of one's own skin, resulting in lesions, despite repeated efforts to decrease or stop picking. Also called skin-picking disorder.

executive ego function Part of the ego that helps maintain or manage one's relationship to and sense of reality.

exhaustion state Final stage of the body's response to major stress, as in the general adaptation syndrome.

exhibitionism Paraphilia in which a man exposes his genitals to females in a socially inappropriate fashion; rarely occurs in women. Also called exhibitionistic disorder.

existentialism Philosophic theory that questions the purpose and nature of existence itself; a fundamental focus is on the meaning of life in face of death and the end of one's personal existence.

existential psychotherapy Type of therapy, based on existential philosophy, that puts the emphasis on here-and-now interaction and on feeling experiences, rather than on rational thinking; little attention is given to patient resistances, and the therapist is involved on the same level and to the same degree as the patient. *See also* phenomenology.

exogamy Marriage outside of a specific group; in a biologic sense, the union of gametes of unrelated parents. *Compare* endogamy.

exogenous In psychiatry, denoting disordered behavior produced by forces outside of the body, as in adjustment disorder.

exorcism Magical practice in which mystical incantations are invoked to remove demons that are alleged to have entered the mind.

expansive mood Expression of feelings without restraint, frequently with an overestimation of their significance or importance. Seen in mania and grandiose delusional disorder.

experiencing Feeling emotions and sensations, as opposed to thinking; being involved in what is happening, rather than standing back at a distance and theorizing.

experimental bias Error in construction of an experiment or study that favors one outcome over another.

experimental design Manner in which an experiment is logically organized in order to maximize the probability of obtaining real effects.

experimental group In an experimental design, the group that is exposed to the experimental manipulation under study. Compare control group.

experimental neurosis Abnormal behavior produced in a laboratory setting; used by Ivan Pavlov to describe symptoms of extreme and persistent agitation that he produced in dogs.

expert witness Legal concept that refers to the physician in the courtroom who is deemed sufficiently knowledgeable in a field to draw informed conclusions from data, for example, that a patient meets the required criteria for commitment or for an insanity defense.

explosive personality Personality type characterized by rigidity, egocentricity, selfishness, religiosity, seclusiveness, explosive outbursts of emotion, and extreme rage reactions when frustrated; enuresis is common.

expressive aphasia Disturbance of speech in which understanding remains, but ability to speak is grossly impaired; halting, laborious, and inaccurate speech. Also known as Broca's, nonfluent, and motor aphasias.

expressive dysphasia Difficulty in expressing verbal language; the ability to understand language is intact.

extended-family therapy Type of family therapy that involves family members, beyond the nuclear family, who are closely associated with it and affect it. *See also* social network therapy.

externalization More general term than projection that refers to the tendency to perceive in the external world and in external objects elements of one's own personality, including instinctual impulses, conflicts, moods, attitudes, and styles of thinking.

external validity Statistical term that refers to how well a test measures what it purports to measure and identifies what it is supposed to identify.

extinction Process by which an operant behavior occurs less frequently when its performance is no longer reinforced.

extrapsychic conflict Conflict that arises between a person and their environment. Also called interpersonal conflict. *Compare* intrapsychic conflict.

extrapyramidal dyskinesia Movement disorder related to dysfunction of the extrapyramidal motor system; signs include a variety of involuntary movements and postures, muscular rigidity, tremor, and gait disturbance, as seen in Parkinson disease, chorea, athetosis, and hemiballismus. Extrapyramidal dyskinetic syndromes may occur as a side effect of antipsychotic drugs; such drug-induced extrapyramidal effects include a parkinsonian syndrome, akathisia, acute dystonic reactions, and tardive dyskinesia.

extrapyramidal effect Bizarre, involuntary motor movement; a central nervous system side effect sometimes produced by antipsychotic drugs. *See also* extrapyramidal dyskinesia.

extrapyramidal motor system Polysynaptic neural pathways involving the basal ganglia and related subcortical nuclei that influence motor behavior.

extrapyramidal syndrome Abnormalities of movement related to injury of motor pathways other than the pyramidal tract; for example, parkinsonism, akathisia, dystonia, tardive dyskinesia.

extrasensory perception (ESP) Experiencing an external event by means other than the five senses, as by telepathy or clairvoyance.

extroversion, extraversion State of one's energies being directed outside oneself. *Compare* introversion.

F

factitious disorder Mental disorder characterized by the voluntary production of false or unreal physical or psychological symptoms; unlike malingering, there is no apparent goal or obvious benefit. Also called Munchausen syndrome. *Compare* somatic symptom disorder; somatoform disorders.

factitious disorder by proxy A disorder in which one person deceitfully creates false signs or symptoms of illness in another person who is under their care; most commonly this condition involves a mother presenting her child as ill. In some cases actual injury or illness may be inflicted on the dependent person. Diagnosed as "factitious disorder imposed on another" in DSM-5.

factitious disorder imposed on another *See* factitious disorder by proxy.

factor analysis Data reduction technique used to reduce a large number of variables to a smaller number of linear combinations of variables.

failure to thrive Delay in physical growth in infants and young children; often accompanied by impaired social and motor development; commonly associated with parental deprivation or neglect when no medical illness is detected. *See* reactive attachment disorder.

false memory A person's recollection and belief of an event that did not actually occur. In false memory syndrome, persons erroneously believe that they sustained an emotional or physical (e.g., sexual) trauma early in life.

false negative Diagnosis of a disorder as being absent in a person (tested negative) when in fact it is present.

false positive Diagnosis of a disorder as being present in a person (tested positive) when in fact it is not present.

falsifiable hypothesis Testable hypothesis that can be confirmed or disconfirmed.

family neurosis Mental disorder in which a person's psychopathology is unconsciously interrelated with that of other members of the family.

family therapy Treatment of more than one member of a family in the same session; family relationships and processes are explored as potential causes of mental disorder in one or more of the family members. *See also* extended-family therapy.

fantasy Daydream; fabricated mental picture of a situation or chain of events. A form of thinking dominated by unconscious material and primary processes that seeks wish fulfillment and immediate solutions to conflicts; may serve as the matrix for creativity or for neurotic distortions of reality.

fatal familial insomnia An inherited prion disease characterized by insomnia and autonomic nervous system dysfunction (e.g., fever, sweating, tachycardia).

father surrogate, father substitute In psychoanalysis, a person onto whom the patient projects their father image and to whom they respond unconsciously in an inappropriate and unrealistic manner, with the feelings and attitudes held toward the real father.

fatigue A feeling of weariness, sleepiness, or irritability after a period of mental or bodily activity. Seen in depression, anxiety, neurasthenia, and somatoform disorders.

fausse reconnaissance False recognition, a feature of paramnesia. Can occur in delusional disorders.

faux de mieux Literally, "for want of anything better," connoting in psychiatry a person choosing a homosexual relationship when no partner of the opposite sex is available.

FDA Food and Drug Administration.

fear Unpleasurable emotional state consisting of psychophysiologic changes in response to a realistic threat or danger. *Compare* anxiety.

feeble-mindedness Obsolete term for mental retardation, which is also called intellectual developmental disorder.

feedback 1. In a given system, the return, as input, of some form of information regarding the output; used as a regulatory mechanism. **2.** In psychiatry, the verbally or otherwise expressed response to a person's behavior by another person or a group.

feeding disorder of infancy or childhood *See* avoidant/restrictive food intake disorder.

fellatio Use of the mouth or tongue to stimulate the male genitalia.

femaleness Anatomic and physiologic features relating to female procreative functions.

female orgasmic disorder Disorder characterized by a woman's inability to achieve orgasm by coitus or masturbation.

feminine identity Inner sense of gender affiliation with females. *Compare* masculine identity.

Ferenczi, Sandor (1873–1933) Hungarian psychoanalyst who was one of Sigmund Freud's early followers and a brilliant contributor to all aspects of psychoanalysis; he came to favor more active and personal techniques than Freud's, to the point that his adherence to psychoanalysis during his last years was questioned.

fetal alcohol syndrome A syndrome that occurs when mothers consume alcohol during pregnancy resulting in fetuses born with birth defects, CNS dysfunction, and intellectual disability.

fetishism Paraphilia in which sexual excitement and gratification are achieved by substituting an inanimate object (such as a shoe, piece of underwear, or other article of clothing) for a human love object. Also called fetishistic disorder.

fibromyalgia A chronic pain disorder characterized by pain and stiffness of the soft tissues (e.g., muscles, ligaments, tendons).

field theory Conceptual approach to the study of personality formulated by Kurt Lewin that focuses on an operational analysis of the causal determinants of human behavior; it is expressed in geometric and mathematical terms borrowed from physics in an attempt to provide a framework suitable to the scientific study of human behavior. The person and their environment together constitute the life space, a complex field of forces acting on the personality and determining behavior.

finger agnosia Agnosia characterized by inability to name fingers; suggests a lesion in the dominant parietal lobe.

fixation 1. Arrest of psychosexual development at any stage before complete maturation. 2. Close and paralyzing attachment to another person, such as one's mother or father.

fixed-interval schedule *See* schedule of reinforcement.

fixed-ratio schedule *See* schedule of reinforcement.

flagellantism, flagellation Process by which sexual partners are aroused and gratified by whipping or being whipped.

flashback hallucinosis Spontaneous recurrence, after a drug-free period, of hallucinations similar to those experienced during an acute toxic episode; often associated with repeated ingestions of LSD. Hallucinations are predominantly visual and may last for months after the last use of a drug.

flat affect Absence or near absence of any signs of affective expression.

flexibilitas cerea CATALEPSY.

flight of ideas Rapid succession of fragmentary thoughts or speech in which content changes abruptly and speech may be incoherent, as seen in mania.

floccillation Aimless plucking or picking, usually at bed-clothes or clothing, commonly seen in senile psychosis and delirium.

flooding Behavior therapy designed to extinguish phobias or anxiety through the presentation of the aversive stimuli in intense forms until the emotional responses cease. Also called implosion therapy.

fluent aphasia Aphasia characterized by inability to understand the spoken word; fluent but incoherent speech is present. Also called Wernicke aphasia, sensory aphasia, and receptive aphasia.

folie à deux Mental illness shared by two persons, usually involving a common delusional system; if it involves three persons, it is referred to as folie à trios, and so forth. *See* shared psychotic disorder.

follow-up examination To the physician, the aftercare and monitoring of patients' responses to treatment and the state of their physical health or mental status.

Food and Drug Administration (FDA) Federal agency that standards for the sale and use of drugs, food substances, and cosmetics.

forensic psychiatry Branch of psychiatry concerned with the aspects of mental illness.

foreplay Sexual play that precedes sexual intercourse. Also called forepleasure.

forepleasure FOREPLAY.

formal operations Jean Piaget's label for completely developed logical thinking capacities. *See also* Piaget, Jean.

formal thought disorder Disturbance in the form of thought rather than the content of thought; thinking characterized by loosened associations, neologisms, and illogical constructs; thought process is disordered, and the person is defined as psychotic. Characteristic of schizophrenia.

formication Tactile hallucination involving the sensation that tiny insects are crawling over the skin; most commonly encountered in cocaine addiction and delirium tremens.

for-profit hospital Investor-owned or proprietary hospital; hospital owned and operated by physicians or corporations that earns a profit for its investors.

fragile X syndrome A syndrome resulting from a mutation on the X chromosome. Includes severe intellectual disability, a large head, long face, prominent ears, and, in males, large testicles.

fragmentary delusion Poorly elaborated delusion, often one of many with no apparent interconnection.

fraternal twins DIZYGOTIC TWINS.

free association Investigative psychoanalytic technique devised by Freud in which the patient seeks to verbalize without reservation or censorship the passing contents of their mind.

freebase Cocaine hydrochloride mixed with ammonia and an alkali solution, and then extracted with a solvent into a crystalline form. *See also* crack.

free-floating anxiety Severe, pervasive, generalized anxiety that is not attached to any particular idea, object, or event; observed particularly in anxiety disorders, although it may be seen in some cases of schizophrenia.

Fregoli syndrome The delusional belief that a known person is an imposter who assumes different appearances or disguises.

frequency Number of cases having a certain score or characteristic; relative frequency measures the number of persons in a specific group (e.g., sex or age) who have a disorder.

frequency distribution Statistical description of raw data in terms of the number of cases that fall into each interval within a set of data; often presented graphically in the form of a frequency histogram or frequency polygram.

Freud, Anna (1895–1982) Daughter of Sigmund Freud noted for her contributions to the developmental theory of psychoanalysis and the mechanisms of defense.

Freud, Sigmund (1856–1939) Austrian psychiatrist and the founder of psychoanalysis. With Josef Breuer, he explored the potentialities of cathartic therapy and then went on to develop the analytic technique and such fundamental concepts as mental sublimation, as well as superego, ego, and id formation and their applications throughout all spheres of human behavior. Among his publications are *The Interpretation of Dreams* (1899). *Studies on Hysteria* (1895, with Breuer), and *Three Essays on the Theory of Sexuality* (1905).

frigidity Outdated term referring, in the female, to lack of sexual response or feeling, ranging from complete lack of arousal to incomplete climax.

frontal lobe syndrome Mental disorder characterized by a marked change in personality, with development of certain relatively characteristic patterns of relating to the environment that have been associated with lesions of the frontal lobes; such patterns include emotional lability, impairment in impulse control and social judgment, and marked apathy and indifference.

frontotemporal dementia Rare form of presenile dementia caused by cerebral atrophy predominantly in the frontotemporal regions of the brain; early stages are characterized by personality and behavior changes. The following progressive cognitive deterioration is indistinguishable clinically from the far more common Alzheimer disease, in which the cerebral atrophy is diffuse. Also called Pick disease. *See also* dementia of the Alzheimer type.

frottage, frotteurism Sexual aberration in which arousal results from rubbing up against someone, usually without specific genital contact, as in a crowd; a person so afflicted is a frotteur. Also called frotteuristic disorder.

frustration Result of the thwarting of any goal-directed behavior; the ability to tolerate frustration and delay gratification is considered a sign of maturity and good ego strength.

fugue Called dissociative fugue in DSM-5, a subtype of dissociative amnesia characterized by a period of almost complete amnesia, during which a person actually flees from an immediate life situation and begins a different life pattern; apart from the amnesia, mental faculties and skills are usually unimpaired. *See also* psychogenic amnesia.

fulfillment Satisfaction of a need or wish.

functional Denoting changes in functioning not attributable to known organic alterations.

functional disorder Disorder that is not caused by an anatomical defect or any other identifiable or demonstrable cause. *Compare* organic disorder.

functional neurologic disorder *See* conversion disorder.

functional neurologic symptom disorder Mental disorder in which the patient experiences an involuntary limitation or alteration of physical function that is an expression of a psychological conflict or need, not of a physical disorder. Previously called conversion disorder in DSM-IV. *See also* hysteria.

functional speech disorder DEVELOPMENTAL ARTICULATION DISORDER.

fusion In psychoanalysis, the joining together of instincts which enables the person to function normally.

G

GABA γ-(gamma) aminobutyric acid.

GAF Scale GLOBAL ASSESSMENT OF FUNCTIONING SCALE.

galactorrhea Abnormal discharge of milk from the breast; may be a result of the endocrine influence of phenothiazine drugs.

galvanic skin response (GSR) Measurable change in the electrical resistance of the skin, used as an index of emotional response.

gambling disorder DSM-5 disorder characterized by persistent and recurrent maladaptive gambling that leads to economic problems and disturbances in personal, social, and occupational functioning. Also called pathologic gambling.

γ**-(gamma) aminobutyric acid (GABA)** Inhibitory amino acid neurotransmitter synthesized by the actions of glutamic acid decarboxylase; GABA receptors are located at postsynaptic sites.

gamophobia Abnormal fear of marriage.

Ganser syndrome Rare atypical disorder, seen in prisoners, characterized by a giving of approximate answers to questions and of talking past the point; to be distinguished from factitious disorder and malingering.

gargoylism Old name for the gargoyle-like facies and dwarfism seen in Hunter and Hurler syndromes: metabolic disorders of mucopolysaccharide metabolism characterized by severe developmental skeletal deformities and mental retardation.

GAS General adaptation syndrome.

gatekeeper In community psychiatry, a person who evaluates the degree of psychological or environmental stress sufficient to enable a patient to gain access to the mental health care system.

gateway drugs Substances of abuse, such as alcohol and marijuana, which provide major portals of entry into abuse of other drugs; considered particularly dangerous because they are believed by so many to be relatively harmless.

Gault decision U.S. Supreme Court 1967 decision mandating fair and accurate fact-finding procedures for juveniles when serious punishment could be inflicted should they be found guilty; rights guaranteed the juvenile include the right to notification of the specific charges, the right to counsel, the right against self-incrimination, and the right to confront and cross-examine the accusers.

gay 1. Preferred term for a homosexual, especially male. **2.** Denoting a homosexual or the lifestyle thereof.

gegenhalten Neurologic term (from the German) for involuntary active resistance to passive movement of the extremities, as may occur in cerebral cortical disorders.

gender dysphoria The DSM-5 term for gender identity disorder. May first appear in childhood, adolescence, or adulthood.

gender identity Culturally determined sets of attitudes, behavior patterns, and physical attributes usually associated with masculinity and femininity. *Compare* sexual identity.

gender identity disorder Psychosexual disorder in which a person feels discomfort with and inappropriateness of their biologic sex, with a marked preference for the clothing and activities of the opposite sex and/or repudiation of the sex organs. Also called gender dysphoria. *See also* transsexualism.

gender role Public declaration of gender; the image of maleness or femaleness that is communicated to others which may or may not coincide with gender identity.

general adaptation syndrome (GAS) Hans Selye's term for the responses of the body to major stress, passing through the alarm reaction, resistance and, finally, exhaustion.

generalization Process by which a behavior occurs in a setting in which it had not previously been reinforced.

generalized anxiety *See* free-floating anxiety.

generalized anxiety disorder An anxiety disorder characterized by chronic generalized anxiety consisting of feelings of tension, tremor, apprehension, and hyperarousal not attached to any particular idea, object, or event. *See* free-floating anxiety.

generalized tonic–clonic seizure Generalized onset of tonic–clonic movements of the limbs, tongue biting, and incontinence followed by slow, gradual recovery of consciousness and cognition. Also called grand mal seizure.

general paralysis, general paresis Mental disorder caused by chronic (tertiary) neurosyphilis.

general systems theory Theoretical framework asserting that the universe is composed of interacting components that are interrelated; in psychiatry, it emphasizes an integrated, holistic approach to understanding behavior.

genetic counseling Presentation and discussion, generally with a prospective parental couple, of factors involved in the inheritance of pathologic conditions as they relate to the couple's generic endowment.

genetic endowment Characteristics that are inherited, usually regarding potential capacities of intelligence, temperament, and general adaptability.

genetic marker Detectable sign of the presence of or vulnerability to inherited disease, such as a measurable biochemical abnormality or an identifiable defect in a particular gene.

genetic material In psychiatry and psychoanalysis, a term used to reflect the data from a patient's personal past history that are useful in developing an understanding of the psychodynamics of their present adaptation.

genital phase Final stage of psychosexual development, occurring during puberty, in which sexual gratification is achieved from genital-to-genital contact and the capacity exists for a mature, affectionate relationship with someone of the opposite sex.

genotype Genetic endowment of a person, consisting of a set of genes received from each parent at the time of conception. *Compare* phenotype.

geriatric psychiatry Branch of psychiatry concerned with the mental disorders that accompany old age. Also called geropsychiatry.

geriatrics Branch of medicine concerned with the aged and the problems of aging. *See also* gerontology.

genito–pelvic pain/penetration disorder Diagnostic category in DSM-5 referring to one or more of the following: genito–pelvic pain; difficulty having intercourse; fear of pain or penetration; and tension of the pelvic floor muscles; previously divided into the distinct categories of dyspareunia and vaginismus.

gerontology Scientific study of aging.

gerontophilia Sexual attraction for the old and the aged by a young person.

gerontophobia Fear of the older adults or of the aging process.

geropsychiatry GERIATRIC PSYCHIATRY.

Gerstmann–Straussler–Scheinker disease DEVELOPMENTAL ARITHMETIC DISORDER.

Gesell Developmental Schedules Developmental landmarks in infants and children that cover ages 8 weeks to 3.5 years. Data are obtained by direct observation of the child's responses to standard toys and other controlled stimulus objects and are supplemented by developmental information provided by the mother or primary caretaker; the schedules yield scores indicating the child's level of development in four separate areas: motor, adaptive, language, and personal and social.

gestalt psychology School of psychology primarily concerned with perceptual processes; a major tenet is that the whole is greater than the sum of its parts. Its development is associated particularly with Max Wertheimer, Wolfgang Koehler, and Kurt Koffka.

gestalt therapy Psychotherapy developed by Frederick S. Perls that emphasizes the treatment of the person as a whole: one's biologic component parts and their organic functioning, one's perceptual configuration, and one's interrelationships with the outside world. It can be used in either an individual or a group therapy setting; focuses on the sensory awareness of the person's here-and-now experiences, rather than on past recollections or future expectations; and uses role playing and other techniques to promote the patient's growth process and development to full potential.

Gilles de la Tourette syndrome *See* Tourette disorder.

global aphasia Combination of grossly nonfluent aphasia and severe fluent aphasia.

Global Assessment of Functioning Scale (GAF Scale) Scale used in DSM to describe the highest social, occupational, or educational level of functioning of a person in the 12 months preceding the current evaluation.

global developmental delay Preferred term for a condition of intellectual and adaptive impairment in children that fail to meet expected developmental milestones in multiple areas of intellectual functioning before 5 years of age. *See also* intellectual disability.

globus hystericus Hysterical symptom in which a person is disturbed by the sensation of a lump in the throat.

glossolalia Unintelligible jargon that has meaning to the speaker but not to the listener. Occurs in schizophrenia.

go-around Technique used in group therapy in which the therapist requires each member of the group to respond to another member, a theme, or an association to encourage the participation of all members in the group.

gonadal dysgenesis TURNER SYNDROME.

grandiose delusion DELUSION OF GRANDEUR. *See* delusion.

grandiosity Exaggerated feeling of one's importance, power, knowledge, or identity. Occurs in delusional disorder and manic states. *See also* delusion of grandeur.

grand mal epilepsy Major form of epilepsy in which gross tonic–clonic convulsive seizures are accompanied by loss of consciousness and, often, incontinence of stool or urine; frequently preceded by an aura.

gray-out syndrome Psychosis that occurs in pilots flying at high altitudes, out of sight of the horizon.

grief Alteration in mood and affect consisting of sadness appropriate to a real loss; normally, it is self-limited. *See also* depression; mourning.

group *See* therapeutic group.

group analytic psychotherapy Term applied by S. H. Foulkes to his pioneering group therapy procedure in which interventions dealt primarily with group, rather than individual, forces and processes, and the group was used as the principal therapeutic agent.

group cohesion Mutual bonds formed between members of a group as a result of their concerted effort for a common interest and purpose; until cohesiveness is achieved, the group cannot concentrate its full energy on a common task.

group dynamics Phenomena that occur in groups from their inception to their termination. Interactions and interrelations among members and between the members and the therapist create tension, which maintains a constantly changing group equilibrium; interactions and the tension created are highly influenced by individual members' psychological makeup, unconscious instinctual drives, motives, wishes, and fantasies. Understanding and effectively using group dynamics are essential in group treatment. Also called group process.

group mind Autonomous and unified mental life in an assemblage of people bound together by mutual interests; a concept used by group therapists who focus on the group as a unit rather than on the individual members.

group practice Delivery of medical services by three or more physicians formally organized to provide care, consultation, diagnosis, and treatment; they share the use of equipment and personnel, and income from the medical practice is distributed among the members of the group.

group pressure Demand by group members that individual members submit and conform to group standards, values, and behavior.

group process GROUP DYNAMICS.

group psychotherapy Application of psychotherapeutic techniques to a group of patients, using interpatient interactions to effect changes in the maladaptive behavior of the individual members.

group relations theory *See* Allport's group relations theory.

GSR Galvanic skin response.

guardianship Person appointed by the courts for the care and management of a person or the property of another, as of a minor or a person incapable of managing their affairs.

guilt Emotional state associated with self-reproach and the need for punishment. In psychoanalysis, refers to a neurotic feeling of culpability that stems from a conflict between the ego and the superego that begins developmentally with parental disapproval and becomes internalized as conscience in the course of superego formation. Guilt has normal psychological and social functions, but special intensity or absence of guilt characterizes many mental disorders, such as depression and antisocial personality disorder, respectively. Some psychiatrists distinguish shame as a less internalized form of guilt that relates more to others than to the self. *See also* shame.

gustatory hallucination Hallucination primarily involving taste.

gynecomastia Female-like development of the male breasts; may occur as an adverse effect of antipsychotic and antidepressant drugs due to increased prolactin levels or anabolic–androgenic steroid abuse.

gynephobia Abnormal fear of women.

gyrectomy Psychosurgery involving excision of a cerebral gyrus.

H

habeas corpus Legal term for the right to petition a court to decide whether confinement has been undertaken with due process of law.

habit Behavior pattern acquired by frequent repetition that leads to regularity or increased facility of performance.

habituation 1. Simple form of learning in which the response to a repeated stimulus lessens over time. **2.** Psychological dependence on the continued use of a drug or chemical substance to maintain a sense of well-being, which can result in addiction.

halfway house Facility for mental patients who no longer need the full facilities of a hospital but are not yet ready to return to their communities.

hallucination False sensory perception occurring in the absence of any relevant external stimulation of the sensory modality involved. For types of hallucination, see the specific term.

hallucinogenic drug, hallucinogen Chemical agent that produces hallucinations, such as LSD and mescaline. *See also* euphorohallucinogen; psychotomimetic drug.

hallucinosis State in which a person experiences hallucinations without any impairment of consciousness.

haloperidol Butyrophenone antipsychotic drug (trade name, Haldol).

Halstead–Reitan Scale Battery of neuropsychological tests used to determine the location and effects of specific brain lesions. A great variety of functions are tested: concept function, abstraction, and visual acuity; dexterity; spatial memory; auditory perception, attention, and concentration; finger-oscillation; speech-sounds perceptions; visual perception; time sense; and sensory stimulation, stereognosis, and tactile perception.

Hamilton Depression Scale Scale of depression based upon a clinical interview with the patient in which the examiner asks about guilt, suicide, sleep habits, weight change, and other symptoms of depression.

haphephobia Abnormal fear of being touched.

haplology Rapid speech in which syllables are omitted; seen in certain manias and in schizophrenic conditions.

haptic hallucination Hallucination of touch.

harm avoidance A term used to describe a character trait in which the person avoids novel situations and taking risks; may be an inborn and genetically determined.

hashish Dried resinous tops of the marijuana plant (*Cannabis sativa*): the active ingredient is tetrahydrocannabinol, which is a potent psychedelic drug.

Health and Human Services Federal organization composed of the following: Office of Human Development Services, Public Health Service (includes Centers for Disease Control; Food and Drug Administration; Health Resources and Services Administration; National Institutes of Health; Alcohol, Drug Abuse, and Mental Health Administration; Agency for Toxic Substances and Disease Registry; and Indian Health Service), Social Security Administration, Health Care Financing Administration, and Family Support Administration. Formerly called Department of Health, Education, and Welfare (HEW).

health insurance Insurance that reimburses one for the costs of medical care or lost income arising from medical disability; may include coverage for the care of mental conditions.

health maintenance organization (HMO) Form of group practice by physicians, often including psychiatrists, and allied health professionals to provide comprehensive health services to subscribers who pay a fixed premium.

Health Systems Agency (HSA) Nonprofit, federally mandated, statewide organizations that promote or limit the development of health services and faculties depending on the needs of the particular state or locale. HSAs develop long- and short-term goals and plans, approve health care proposals requesting federal funding, review existing facilities and services, and suggest future consideration and renovation projects based on their findings.

healthy identification Modeling of oneself, consciously or unconsciously, on another person who has a sound psychic makeup; such identification has constructive results. *See also* imitation.

hearsay Out-of-court statement offered to prove the truth of the matter asserted; usually inadmissible as testimony.

hebephrenia Complex of symptoms, considered a form of schizophrenia, characterized by wild or silly behavior or mannerisms, inappropriate affect, frequent hypochondriacal complaints, and delusions and hallucinations that are transient and unsystematized. *See also* schizophrenia.

hedonism Pleasure-seeking behavior. *Compare* anhedonia.

helplessness **1.** Learned: State of apathy that occurs as a type of experimental neurosis in animals who cannot escape an aversive stimulus; proposed as the animal model of human depression in which a patient feels helpless to control events. **2.** Psychic: State that occasions the first expressions of anxiety in humans, for example, the experience of birth.

hematophobia Abnormal fear of the sight of blood.

herd instinct Desire to belong to a group and to participate in social activities; used by Wilfred Trotter to indicate the presence of a hypothetical social instinct in humans. In psychoanalysis, herd instinct is viewed as a social phenomenon, rather than as an instinct.

hermaphrodite Person who has both female and male gonadal tissue, usually with one sex dominating.

heroin Diacetylmorphine; an illicit, addictive alkaloid opioid that requires subsequent larger and larger amounts to achieve the same narcotic and analgesic effects. Injected intravenously, it produces a more rapid onset of euphoria than does morphine.

heterogeneity Dissimilarity in the genotypal structure of individuals originating through sexual reproduction.

heterosexuality Sexual attraction or contact between opposite-sex persons. The capacity for heterosexual arousal is probably innate, biologically programmed, and triggered in very early life, perhaps in part by olfactory modalities as seen in lower animals. *Compare* bisexuality; homosexuality.

heterozygous Possessing different allelic forms of a gene at a given locus in homologous chromosomes.

heuristic 1. Involving or serving as an aid to learning, discovery, or problem-solving by experimental and especially trial-and-error methods. **2.** Of or relating to exploratory problem-solving techniques that utilize self-educating techniques.

5-HIAA 5-Hydroxyindoleacetic acid.

histrionic personality disorder Condition in which the patient, usually an immature and dependent person, exhibits unstable, overreactive, and excitable self-dramatizing behavior that is aimed at gaining attention and is at times seductive, although the person may not be aware of that aim.

HIV Human immunodeficiency virus.

HMO Health maintenance organization.

hoarding disorder Disorder characterized by persistent difficulty parting with or discarding possessions; delusional beliefs about the accumulated items may be present. Part of obsessive-compulsive disorder in DSM-5.

holism In psychiatry, the study of the total person as a distinctive entity, rather than as a collection of various characteristics.

holophrastic Using a single word to express a combination of ideas; schizophrenic persons may use holophrastic language.

homeless Persons who have no permanent place of residence, often sleeping in the street, many of whom require psychiatric treatment or hospitalization.

homeostasis Tendency to maintain a constancy and stability of bodily processes to ensure optimal functioning; the state of bodily equilibrium and the processes whereby such equilibrium is maintained.

homophobia Fear of homosexual persons.

homosexuality Sexual attraction or contact between same-sex persons. Some authors distinguish two types: overt and latent. *Compare* bisexuality; heterosexuality. When referring to people, homosexual is considered derogatory and the preferred terms are gay and lesbian.

homosexual panic Sudden, acute onset of severe anxiety precipitated by the unconscious fear or conflict that one may be a homosexual or act out homosexual impulses.

homovanillic acid (HVA) Phenol found in human urine; as the major metabolite of dopamine, it is most often measured in psychiatric research.

Horney, Karen (1885–1952) Psychiatrist and psychoanalyst whose theories of the genesis of neurosis emphasized environmental and cultural factors and thus departed from the biologic-instinctual framework of orthodox freudian thought.

hostile transference *See* transference.

hostility Conflict, opposition, or resistance in thought or principle.

hotline Telephone assistance service for crisis intervention.

HSA Health Systems Agency.

5-HT 5-Hydroxytryptamine.

human immunodeficiency virus (HIV) An infection caused by a retrovirus that incorporates itself into deoxyribonucleic acid (DNA). Symptoms include fever, myalgia, headache, fatigue, gastrointestinal symptoms, and rash. Can cause acquired immunodeficiency syndrome (AIDS).

humanistic theory Group of diverse theories that share a holistic conceptualization that emphasizes the uniqueness, value, dignity, and worth of each human being as an individual. Human behavior is considered understandable only in terms of the meaning of experiences for the individual person; it represents a complex interplay of numerous physical, psychological, and sociocultural factors.

humiliation Sense of disgrace, dishonor, and shame; often experienced by depressed patients.

humor Defense mechanism that allows the overt expression of feelings and thoughts without personal discomfort or immobilization and does not produce an unpleasant effect on others; allows the individual to tolerate and yet focus on what is too terrible to be borne. Differs from wit, a form of displacement that involves distraction from the affective issue.

Huntington disease (chorea) Hereditary and progressive central nervous system disease characterized by jerking motions and progressive mental deterioration: inherited as an autosomal dominant, with onset in adult life.

HVA Homovanillic acid.

hydrocephalus Excessive accumulation of cerebrospinal fluid which dilates the cerebral ventricles.

hydrophobia In psychiatry and psychology, an abnormal fear of water.

hydrotherapy Use of water by external application in the treatment of disease.

5-hydroxyindoleacetic acid (5-HIAA) Metabolite of serotonin, the concentration of which can be measured in cerebrospinal fluid and serum: determination of 5-HIAA concentrations is used in psychiatric research and may help guide clinical practice.

5-hydroxytryptamine (5-HT) SEROTONIN.

hyperactivity Increased muscular activity. The term is commonly used to describe a disturbance found in children that is manifested by constant restlessness, overactivity, distractibility, and difficulties in learning. Also called hyperkinesis. *See also* attention-deficit hyperactivity disorder; minimal brain dysfunction.

hyperalgesia Excessive sensitivity to pain. Seen in somatoform disorder.

hyperarousal A state of increased psychological and physiologic tension marked by such effects as reduced pain tolerance, anxiety, exaggeration of startle responses, insomnia, fatigue, and accentuation of personality traits.

hyperesthesia Increased sensitivity to tactile stimulation.

hyperkinesis HYPERACTIVITY.

hyperkinetic syndrome Disorder characterized by excessive muscular activity (hyperactivity) such as restlessness, aggressiveness, and destructive activity. *See also* attention-deficit hyperactivity disorder; minimal brain dysfunction.

hypermnesia Exaggerated degree of retention and recall. It can be elicited by hypnosis and may be seen in certain prodigies; also may be a feature of obsessive-compulsive disorder (OCD), some cases of schizophrenia, and manic episodes of bipolar I disorder.

hyperorexia Extreme appetite. *See also* bulimia nervosa.

hyperphagia Increase in appetite and intake of food.

hyperpragia Excessive thinking and mental activity, generally associated with the manic phase of bipolar I disorder.

hypersexuality Term used to describe excessive or compulsive interest in sex that may result in sex addiction.

hypersomnia Disorder characterized by excessive time spent asleep which may be nonorganic, organic, narcoleptic, idiopathic, part of Klein–Levin syndrome, or primary in nature. Also called hypersomnolence disorder.

hypersomnolence disorder *See* hypersomnia.

hypertensive crisis HYPERTENSIVE ENCEPHALOPATHY.

hypertensive encephalopathy Acute syndrome in which severe hypertension is associated with headache, nausea, vomiting, visual disturbances, convulsions, and alterations in consciousness characterized by confusion, somnolence, stupor, and coma. Prompt lowering of the blood pressure by pharmacologic means may reverse the picture in 24 to 48 hours; if the blood pressure is not controlled, intracerebral hemorrhage and a fatal outcome may result. Usually, the patient has entered the malignant phase of essential or secondary hypertension, with marked blood pressure elevation (generally in the range of 250/150 mm Hg) and evidence of cardiac decompensation, declining renal function, and advanced hypertensive retinopathy; also seen as a side effect of monoamine oxidase inhibitors when tyramine-containing foods are consumed. Also called hypertensive crisis.

hyperthyroidism Excessive activity of the thyroid gland; may be associated with dementia, anxiety disorders, and insomnia. *Compare* hypothyroidism.

hyperventilation Excessive breathing, generally associated with anxiety, which can cause a reduction in blood carbon dioxide that produces symptoms of lightheadedness, palpitations, numbness and tingling periorally and in the extremities, and occasionally syncope.

hypervigilance Excessive attention to and focus on all internal and external stimuli; usually seen in delusional or paranoid states.

hypesthesia Diminished sensitivity to tactile stimulation.

hypnagogic hallucination Hallucination occurring while falling asleep, not ordinarily considered pathologic.

hypnoanalysis Use of hypnosis in psychoanalysis to gain access to unconscious processes that the patient cannot reveal by means of ordinary therapeutic maneuvers.

hypnodrama Psychodrama under hypnotic trance; the patient is put into a hypnotic trance and encouraged to act out various past experiences.

hypnoid state Freud’s term describing an alteration of consciousness that occurs characteristically in hysteria during periods of emotional stress: characterized by heightened suggestibility and provides a basis for hysterical somatic symptom formation.

hypnophobia Abnormal fear of sleep.

hypnopompic hallucination Hallucination occurring while awakening from sleep, ordinarily not considered pathologic.

hypnosis Artificially induced alteration of consciousness characterized by increased suggestibility and receptivity to direction.

hypnotherapy Therapy that makes use of hypnosis.

hypnotic Drug used for the express purpose of producing sleep. *Compare* sedative.

hypoactive sexual desire disorder Sexual desire disorder characterized by deficiency or absence of sexual fantasies and desire for sexual activity. Also called sexual interest/arousal disorder in women.

hypoactivity Decreased motor and cognitive activity, as in psychomotor retardation; visible slowing of thought, speech, and movements. Also called hypokinesis.

hypochondria Exaggerated concern about health that is based not on real medical pathology, but on unrealistic interpretations of physical signs or sensations as abnormal.

hypochondriacal neurosis HYPOCHONDRIASIS.

hypochondriasis *See* somatic symptom disorder.

hypoglycemia Abnormally small concentration of glucose in the circulating blood; associated with anxiety and depression.

hypomania Mood abnormality with the qualitative characteristics of mania, but less intense; characterizes the hypomanic episode of bipolar II disorder when lasting 4 to 7 days.

hypothalamus Principal center in the forebrain for integration of visceral functions involving the autonomic nervous system. Functions include regulation of sexual activity, water, fat, carbohydrate metabolism, and temperature; also appears to play a role in the nervous mechanisms underlying moods and motivational states.

hypothyroidism Decreased activity of the thyroid gland; may be associated with dementia, depression, and sleep disorders. *Compare* hyperthyroidism.

hysteria A term previously used for a group of conditions involving a sudden impairment of function in response to emotional stress.

hysterical anesthesia Functional disorder characterized by the absence of tactile sensation in an area of the body; observed in certain cases of conversion disorder.

hysterical personality *See* histrionic personality disorder.

hysterics Expression of emotion often accompanied by crying, laughing, and screaming; a term generally used by laypersons rather than by psychiatrists.

I

iatrogenic Denoting an unfavorable response to a mode of treatment induced by the treatment itself. *Compare* idiopathic.

ICD *International Classification of Diseases.*

ice Pure form of methamphetamine that can be either inhaled or injected intravenously by abusers; its psychological effects last for hours and are particularly potent. Unlike the crack cocaine, which has to be imported, ice is a synthetic drug that can be manufactured in domestic illicit laboratories.

ICT Insulin coma therapy.

Id One of the three components of the psychic apparatus in the Freudian structural framework, the others being the ego and superego. It operates unconsciously; harbors the innate, biologic, instinctual drives; is the source of psychic energy (libido); follows the pleasure principle; seeks immediate reduction of drive tension without regard for external reality; and is under the influence of the primary-process mental activity that characterizes the unconscious.

idealization Mental mechanism whereby a person consciously or unconsciously overestimates an admired attribute or aspect of another person.

idea of reference Misinterpretation of incidents and events in the outside world as having a direct personal reference to oneself; occasionally observed in normal persons, but frequently seen in paranoid patients. If present with sufficient frequency or intensity or if organized and systematized, they constitute delusions of reference.

ideas of unreality Thought that events are artificial, illusory, unpredictable, or do not exist.

ideational shield Intellectual, rational defense against the anxiety that persons would feel if they became vulnerable to the criticisms and rejection of others; as a result of their fear of being rejected, they may feel threatened if they criticize another person, an act that is unacceptable to them. In both group and individual therapy, conditions are set up that allow the participants to lower the ideational shield.

idée fixe Fixed idea that is recurrent and most often associated with obsessional states.

identical twins MONOZYGOTIC TWINS.

identification Unconscious defense mechanism by which a person patterns themself after another person; in the process, the self is more or less permanently altered. *See also* healthy identification; imitation.

identification with the aggressor Unconscious process by which a person incorporates within themself the mental image of a person who represents a source of frustration from the outside world; a primitive defense that operates in the interest and service of the developing ego. A classic example of this defense occurs toward the end of the oedipal stage, when a boy, whose main source of love and gratification is his mother, identifies with his father who represents the source of frustration, being the powerful rival for the mother; the child cannot master or run away from his father, so he is obliged to identify with him.

identity One's global role in life and the perception of the sense of self; problems with identity are common during adolescence, in schizophrenia, and in the borderline and schizotypal personalities.

identity crisis Social role conflict as perceived by a person; loss of the sense of personal sameness and historical continuity, and/or inability to accept or adopt the role the person believes is expected by society. Identity crises are frequent in adolescence.

identity disorder A disorder characterized by a chaotic sense of self; a loss of the sense of personal sameness, usually involving a social role conflict as perceived by the individual. It is common in adolescence, when adolescents feel unwilling or unable to accept or adopt the roles they believe are expected of them by society; often manifested by isolation, withdrawal, rebelliousness, negativity, and extremism. Not a DSM-5 diagnosis.

idiopathic Without known or identifiable cause. *Compare* iatrogenic.

idiot Obsolete and pejorative term for a classification of mental retardation in which individuals have a mental age of less than 3 years.

idiotropic Egocentric or introspective.

idiot-savant Mentally retarded person who is able to perform unusual mental feats in sharply circumscribed intellectual areas, such as complicated calculations or puzzle solving.

I-It Philosopher Martin Buber's description of damaging interpersonal relationships. If a person treats themself or another person exclusively as an object, they prevent mutuality, trust, and growth. When pervasive in a group, I-It relationships prevent human warmth, destroy cohesiveness, and retard group process. *Compare* I-Thou.

illness anxiety disorder Preoccupation with being sick or developing a disease with few or no somatic symptoms and an absence of related pathologies found on medical examination. Differentiated from somatic symptom disorder (hypochondriasis) by the absence of somatic symptoms or complaints; patients with illness anxiety disorder are primarily concerned with "the idea" of being or becoming sick; a DSM-5 diagnosis. *Compare* somatic symptom disorder.

illogicality Pattern of speech or thinking in which conclusions that are reached do not follow logically; may take the form of unwarranted or faulty inferences.

illogic thinking Thinking containing erroneous conclusions or internal contradictions; psychopathologic only when it is marked and not caused by cultural values or intellectual deficit.

illuminism State of hallucinations in which a patient converses with imaginary supernatural creatures.

illusion Perceptual misinterpretation of a real external stimulus. *Compare* hallucination.

imago Jungian term referring to an idealized unconscious mental image of a key person in someone's early life.

imbecile Obsolete and pejorative term for a classification of mental retardation in which individuals have a mental age of 3 to 7 years.

imitation Duplication or mimicking of behaviors or portions of behaviors observed in a model. *See also* identification.

immediate memory Reproduction, recognition, or recall of perceived material within seconds after presentation. *Compare* long-term memory; short-term memory.

impaired insight Diminished ability to understand the objective reality of a situation.

impaired judgment Diminished ability to understand the objective reality of a situation. *See also* judgment.

implosion Behavior therapy technique in which anxiety-arousing stimuli are vividly presented in imagination; the patient repeatedly experiences intense anxiety in the absence of objective danger until the anxiety response is extinguished.

impotence A term used to refer to the inability to obtain an erection. In DSM, it is diagnosed as erectile disorder.

imprinting Particular kind of learning that occurs during an early critical developmental stage in animals, characterized by rapid acquisition and great resistance to extinction. It seems to represent a biologically predetermined proclivity to learn certain persistent species-specific behavior patterns, give exposure to the proper stimulus during the critical stage, but its precise role in human behavioral development has not been established.

improvisation In psychodrama, the acting out of situations without prior preparation.

impulse Psychic striving or urge to perform an action. *See also* drive; instinct.

impulse control Ability to resist an impulse, drive, or temptation to perform some action.

impulse-control disorders Mental disorders in which there is weak impulse control with impulsive behavior that is usually irresistible, pleasurable, and aimed at obtaining immediate gratification, without regard for the consequences of the behavior, for example, pathologic gambling, kleptomania, pyromania, intermittent explosive disorder, trichotillomania.

impulsion Blind following of internal drives without regard for social acceptance or pressure from the superego, normally seen in young children. In adults, it is common in those with weak defensive organizations; in such cases it tends to be a symbolic phenomenon.

inappropriate affect Emotional tone out of harmony with the idea, thought, or speech accompanying it. Seen in schizophrenia.

incest Sexual activity between close blood relatives. Common patterns are father–daughter, mother–son, and between siblings; may also be same sex.

incidence Research term for the number of cases of a disease whose onset occurs during a specific period of time. *See also* prevalence.

incidence rate Rate at which new cases of a disease or a condition are occurring. The incidence rate per 1,000 persons in the population is the number of new cases that occur within a defined unit of time, divided by the persons exposed or at risk during the same time unit, multiplied by 1,000.

incoherence Communication that is disconnected, disorganized, or incomprehensible. *See also* word salad.

incompetence Legal term indicating that thought processes are inadequate for sound judgment and may lead to maladaptive or abnormal behavior; legally, incompetent persons cannot be held responsible for their actions.

incorporation Primitive unconscious defense mechanism in which the psychic representation of another person or aspects of another person are assimilated into oneself through a figurative process of symbolic oral ingestion; represents a special form of introjection and is the earliest mechanism of identification.

increased libido Increase in sexual interest and drive. Often associated with mania.

independent variable In research methodology, those qualities that the experimenter systematically varies, for example, time, dosage.

individual psychology System of psychiatric theory developed by Alfred Adler that stresses compensation and overcompensation for feelings of inferiority in a person's strivings to adapt to the social milieu, and emphasizes the interpersonal nature of a person's problems; applied by some practitioners to group psychotherapy and counseling.

individual therapy Traditional dyadic therapeutic technique in which a psychotherapist treats only one patient during a given therapeutic session.

individuation Jungian term denoting a process whereby the person molds and develops a healthy integrated individual personality through maximum differentiation and development of each system of the personality. *See also* actualization.

Indoklon therapy Type of biologic therapy in which a convulsive seizure is induced by the inhalation of the drug Indoklon (hexafluorodiethyl ether).

indolamine Class of biogenic amines that contain an indole ring and are derived from the essential amino acid tryptophan; includes lysergic acid diethylamide (LSD) and serotonin. Monoamine oxidase inhibitors elevate the levels of indolamines by delaying their breakdown.

induced psychotic disorder *See* shared psychotic disorder.

industrial psychiatry Branch of psychiatry concerned with workers' adjustments to their jobs and with the effects of a business organization on its members.

ineffability Ecstatic state in which a person insists that their experience is inexpressible and indescribable, that it is impossible to convey what it is like to one who never experienced it.

infancy Childhood period of helplessness and marked dependency, generally the first year of life.

infantile articulation DEVELOPMENTAL ARTICULATION DISORDER.

infantile autism A term now subsumed under autism spectrum disorder that refers to a syndrome beginning in infancy and characterized by withdrawal and self-absorption, failure to develop attachment to a parental figure, ineffective communication and mutism, preoccupation with inanimate objects, and an obsessive demand for sameness in the environment. Also called Kanner syndrome; autistic spectrum disorder. *See also* autistic thinking.

infantile dynamics Psychodynamic integrations, such as the Oedipus complex, that are organized during childhood and continue to exert an unconscious influence on adult personality.

infantile perseveration DEVELOPMENTAL ARTICULATION DISORDER.

infantile sexuality Freudian concept regarding the erotic life of infants and children, that, from birth, infants are capable of erotic activities; encompasses the overlapping phases of psychosexual development during the first 5 years of life and includes the oral phase (birth to 18 months), the anal phase (ages 1 to 3), and the phallic phase (ages 2 to 6).

infant psychiatry That aspect of child psychiatry concerned with the diagnosis, treatment, and prevention of maladaptive psychological functioning in infants.

inferential statistics Methods used for drawing general conclusions about probabilities on the basis of a sample. *See also* statistical inference.

inferiority complex Concept, originated by Alfred Adler, that everyone is born with a feeling of inferiority or inadequacy secondary to real or fantasied organic or psychological deficits; how the inferiority or feeling of inferiority is handled determines a person's behavior in life. *See also* masculine protest.

informal admission As a general hospital model, a patient admitted to a psychiatric unit is free to enter and to leave, even against medical advice.

information theory System dealing with the transmission, reception, and distortion of communicated messages.

informed consent Consent given by a patient to a physician to perform a procedure legally requires that the patient be informed of the nature of the proposed procedure, its risks, its proposed benefits, and the available alternatives; the major problem in giving informed consent, legally and psychologically, is the extent or depth of disclosure.

infradian rhythm *See* biologic rhythm.

infusion Introduction of fluid other than blood, for example, medicines, into a vein.

inhalant Substance taken in by breathing. Production of psychoactive effects by the ingestion of such volatile substances as gasoline, airplane glue, aerosols (especially spray paints), or rubber cement, warrants the DSM diagnosis of inhalant dependence or inhalant abuse.

inhibited female orgasm Recurrent and persistent inhibition (absence) of the female orgasm, referring to the inability of a female to achieve orgasm by masturbation or coitus. Also called anorgasmia.

inhibited male orgasm Difficulty or inability of a male in achieving ejaculation during coitus. Also called ejaculatory incompetence; retarded ejaculation. *See also* primary retarded ejaculation; secondary retarded ejaculation.

inhibited sexual excitement Former term for female sexual arousal disorder.

inhibition Limitation or renunciation of ego functions occurring consciously, alone or in combinations, to evade anxiety arising out of conflict with instinctual impulses, the superego, or environmental forces or figures.

initial insomnia Falling asleep with difficulty; usually seen in anxiety disorder. *Compare* middle insomnia; terminal insomnia.

insane Of or pertaining to one who is of unsound mind. A legal rather than a psychiatric term.

insanity Legal concept denoting a mental disturbance, due to which a person lacks criminal responsibility for an alleged crime and hence cannot be convicted of the crime. *See also* American Law Institute formulation of insanity; Durham rule; M’Naghten rule.

insanity defense Criminal law defense plea that a person lacks criminal responsibility by reason of insanity.

insecurity Feelings of helplessness, unprotectedness, and inadequacy in the face of manifold anxieties arising from uncertainty regarding one’s goals, ideals, abilities, and relations to others.

insight Conscious recognition of one’s own condition. In psychiatry, it more specifically refers to the conscious awareness and understanding of one’s own psychodynamics and symptoms of maladaptive behavior; highly important in effecting changes in the personality and behavior of a person.

insight-oriented psychotherapy Psychotherapy based on the principles of psychoanalysis in which persons develop a conscious awareness and understanding of their own psychodynamics and symptoms of maladaptive behavior. There is greater emphasis on day-to-day reality issues and a lesser emphasis on the development of transference issues than in psychoanalysis.

insight therapy *See* insight-oriented psychotherapy.

insomnia disorder Difficulty in falling asleep or difficulty in staying asleep. It can be related to another mental disorder (nonorganic), can be related to a known organic factor (such as a physical disorder or a medication), or can be primary (not related to a known organic factor or to another mental disorder). *See also* initial insomnia; middle insomnia; terminal insomnia.

instinct Basic inborn urge or drive. Freud postulated the existence of two opposing primal instincts: a life instinct (eros) and a death instinct (thanatos). An extensive array of human instincts has been proposed, including possessive instinct, mastery instinct, and herd or social instinct. Because of its implication of a fixed, essentially unalterable, largely hereditary response or psychic tendency that does not involve learning or reason, the term has become ambiguous and controversial when applied to human behavior. Many modern psychoanalysts prefer the term “drive” for what Freud termed “instinct.” *See also* drive.

institutionalization Process of placing a patient in a facility for extended care, for example, a mental hospital or nursing home.

instrumental conditioning OPERANT CONDITIONING. *See* conditioning.

insulin coma therapy (ICT) Type of biologic therapy in which large amounts of insulin are given to a psychotic patient, usually a schizophrenic, producing profound hypoglycemia and resulting in a coma; introduced by Manfred Sakel in 1933, it declined in use drastically after the introduction of antipsychotic drugs. An occasionally used early stage of this treatment is called subcoma therapy.

intake Initial interview between a patient and a member of a psychiatric team, a term usually used in connection with admission to a mental health facility.

integration Constructive organizing, absorbing, and incorporating into the personality of data and experience; also refers to an amalgamation of separate parts into a functioning integral whole.

intellectual disability (ID) or intellectual developmental disorder Below average general intellectual functioning with concurrent impairment in adaptive behaviors, such as communication, self-care, and interpersonal skills, which manifests before adulthood. In DSM-5, the degree of ID is classified as mild, moderate, severe, or profound based on overall level of functioning. Previously called mental retardation.

intellectual insight Knowledge of the reality of a situation without the ability to successfully use that knowledge to effect an adaptive change in behavior. *Compare* true insight.

intellectualization Unconscious defense mechanism in which reasoning or logic is used in an attempt to avoid confrontation with an objectionable impulse and thus defend against anxiety. Also called brooding compulsion; thinking compulsion.

intellectual subaverage functioning IQ of more than two standard deviations below the test mean obtained on an intelligence test.

intelligence Capacity for learning and ability to recall, integrate constructively, and apply what one has learned; the capacity to understand and to think rationally.

intelligence quotient (IQ) Numerical measure of mental capability determined by dividing a mental age (MA) score achieved on a specific test, such as the Stanford–Binet Test, by the patient’s chronologic age (CA) and multiplying by 100.

intermission In psychiatry, the interval between attacks of a particular syndrome; when it is not certain that the symptoms will return, the interval is called a remission.

intermittent explosive disorder Condition in which a person has discrete episodes of losing control of aggressive impulses resulting in serious assault or destruction of property; often caused by a disorder of brain physiology, particularly in the limbic system.

intermittent reinforcement Any schedule of reinforcement in which the response is not always reinforced.

***International Classification of Diseases* (ICD)** World Health Organization's official list of disease categories subscribed to by all WHO member nations. ICD-CM is ICD's clinical modification, first published in 1919.

internship First year of clinical training following medical school, now called postgraduate year I (PGY-I), which progresses onto residency training in a specialty area.

interobserver reliability Degree of concurrence or disagreement among observers of a single event.

interpersonal conflict EXTRAPSYCHIC CONFLICT.

interpersonal psychiatry Dynamic-cultural system of psychiatry based on Harry Stack Sullivan's interpersonal theory, which proposes that each person must be viewed as an entity interacting with the sociocultural and interpersonal environment; emphasis is on interactive experiences.

interpersonal psychotherapy A type of brief psychotherapy used primarily for the treatment of depression. In a course of 10 to 12 sessions the therapist and patient focus of current problems with person in the environment which increases the patient's perception of self-defeating behaviors that contribute to low self-esteem and depression.

interpersonal skill Effectiveness of adaptive behavior in relation to other persons; ability to express feelings appropriately, to be socially responsible and responsive, and to work in harmony with others.

interpretation The formulation or description by the therapist to the patient of the therapist's understanding of the symbolic significance or meaning of any aspect of the patient's intrapsychic processes, behavior, or verbal productions; an attempt is made to construct into a more meaningful form the patient's resistances, defenses, transferences, and other symbolic activities, for example, dreams, fantasies.

Interpretation of Dreams The Sigmund Freud's major presentation (1899) not only of his theories about the meaning of dreams, a subject hitherto regarded as outside scientific interest, but also of his concept of a mental apparatus topographically divided into unconscious, preconscious, and conscious areas.

interview Person-to-person interaction in which information about a subject is collected. *See also* structured interview; unstructured interview.

intoxication Organic mental disorder due to the recent ingestion or presence in the body of an exogenous substance producing maladaptive behavior by virtue of its effects on the central nervous system. The most common psychiatric changes involve disturbances of perception, wakefulness, attention, thinking, judgment, emotional control, and psychomotor behavior; the specific clinical picture depends on the nature of the substance ingested. For types of intoxication, see the specific term.

intrapersonal conflict INTRAPSYCHIC CONFLICT.

intrapsychic Referring to what originates, takes place, or is situated within the psyche or mind.

intrapsychic ataxia *See* ataxia.

intrapsychic conflict State of tension arising from the clash of two or more incompatible or opposing forces; for example, wishes, needs, motives, thoughts operating within oneself. Also called intrapersonal conflict. *Compare* extrapsychic conflict.

intravenous Denoting introduction of a substance into a vein or veins. *See also* infusion.

introjection Unconscious, symbolic internalization of a psychic representation of a hated or loved external object with the goal of establishing closeness to and constant presence of the object; considered an immature defense mechanism. In the case of a loved object, anxiety consequent to separation or tension arising out of ambivalence toward the object is diminished; in the case of a feared or hated object, internalization of its malicious or aggressive characteristics serves to avoid anxiety by symbolically putting those characteristics under one's own control. *See also* incorporation.

intropunitive Turning anger inward toward oneself; commonly observed in depressed patients.

introspection Contemplating one's own mental processes so as to achieve insight.

introversion State in which a person's energies are directed inward toward the self, with little or no interest in the external world. *Compare* extroversion.

introverted personality disorder SCHIZOID PERSONALITY DISORDER.

inversion Sigmund Freud's term for homosexuality, in which he distinguished three types: absolute, amphigenous, and occasional.

involuntary admission Hospitalization of a person against their will. Following an application by a friend or relative for the admittance of a person who does not recognize the need for hospital care, certification by two physicians is required, which allows the patient to be hospitalized for 60 days. Patients have access to legal counsel and, if brought to court, a judge can order the patient's release. After 60 days, the case must be reviewed by a board (of psychiatrists, physicians, lawyers, and citizens not affiliated with the hospital). *See also* parens patriae.

involuntary discharge Termination of hospitalization against a patient's will.

involuntary treatment Care of a person, who has a mental disorder that constitutes a danger to self or others, against that person's will.

involutional melancholia Depression occurring in late middle-age in persons who generally have no history of previous mental illness. Characteristic manifestations include delusions of sin, guilt, or poverty; an obsession with death; preoccupation with somatic (particularly gastrointestinal) function; despair, dejection, agitation, anxiety, and insomnia; and, in some cases, paranoid ideation. The term is not used in DSM-5, but is replaced by the diagnosis of major depressive episode, melancholic type. Also called involutional psychosis.

involutional psychosis *See* involutional melancholia.

IQ Intelligence quotient.

irregular sleep phase circadian rhythm disorder A circadian rhythm sleep–wake disorder characterized by a fragmented, disorganized, or unpredictable sleep–wake pattern.

irrelevant answer Answer that is not responsive to the question.

irresistible impulse Concept that a person charged with a criminal offense is not responsible for their act if the act was committed under an impulse that the person was unable to resist because of mental disease. The courts have chosen to interpret this law in such a way that it has been called the "policeman-at-the-elbow" law; in other words, the court will grant the impulse to be irresistible only if it determines that the accused would have gone ahead with the act even if they had had a policeman at their elbow.

irritability Abnormal or excessive excitability, with easily triggered anger, annoyance, or impatience.

irritable mood State in which one is easily annoyed and provoked to anger. *See also* irritability.

isolated explosive disorder Disorder of impulse control in which a person has a single episode characterized by a failure to resist a violent impulse against others.

isolation In psychoanalysis, a defense mechanism involving the separation of an idea or memory from its attached feeling tone; unacceptable ideational content is thereby rendered free of its disturbing or unpleasant emotional charge. *Compare* relatedness. *See also* alienation.

isophilic Term used by Harry Stack Sullivan to mean liking or feeling affectionate toward people of the same sex, without the sexual or erotic aspects of homosexuality.

I-Thou Philosopher Martin Buber's conception that a person's identity develops from the true sharing by persons; basic trust can occur in a living partnership in which each member identifies the particular real personality of the other in their wholeness, unity, and uniqueness. In groups, I-Thou relationships promote warmth, cohesiveness, and constructive group process. *Compare* I-It.

J

jacksonian epilepsy Epilepsy characterized by recurrent episodes of focal motor seizures that usually begin with a highly localized tonic contraction or clonic movement which increases in severity and spreads progressively, often involving the entire body and terminating in a characteristic generalized convulsion with loss of consciousness; implies organic cerebral disease.

jamais vu Paramnestic phenomenon characterized by a false feeling of unfamiliarity with a real situation that one has experienced.

James–Lange theory of emotion Theory that the experience of emotion is preceded by a visceral or motor response, for example, first the person runs and then feels fear.

Janet, Pierre (1859–1947) Last great representative of the French school of psychiatry, known for his concept of psychological automatism and for his interest in cases of multiple personalities; first used the term “la belle indifférence.”

jargon aphasia Aphasia in which the words produced are neologistic, that is, nonsense words created by the patient.

JCAHO Joint Commission on Accreditation of Healthcare Organizations. Formerly JCAH, Joint Commission on Accreditation of Hospitals.

JCMHC Joint Commission on Mental Health of Children.

Joint Commission on Accreditation of Healthcare Organizations (JCAHO) Group of agencies that influence the standards and performance of hospitals and health organizations; they inspect hospitals every 2 years and determine requirements for hospital accreditation (on which hospital reimbursements from Medicare and Medicaid are contingent).

Joint Commission on Mental Health of Children (JCMHC) Multidisciplinary agency authorized by U.S. Congress in 1965 to monitor the treatment of children with emotional illness.

Joint Commission on Mental Illness and Health Multidisciplinary study group established in 1955 whose “Action for Mental Health”

(1961) was instrumental in the legislation and federal funding for community mental health centers; with the passage of the Community Mental Health Centers Act in 1963, deinstitutionalization became a reality.

Jones, Ernest (1879–1958) Welsh psychoanalyst, and one of Freud's early followers, who was an organizer of the American Psychoanalytic Association (1911) and of the British Psychoanalytical Society (1919) and a founder and long-time editor of the journal of the International Psychoanalytical Association; a prolific author most noted for his three-volume biography of Sigmund Freud.

judgment Mental act of comparing or evaluating choices within the framework of a given set of values for the purpose of electing a course of action. If the course of action chosen is consonant with reality or with mature adult standards of behavior, judgment is said to be intact or normal; judgment is said to be impaired if the chosen course of action is frankly maladaptive, results from impulsive decisions based on the need for immediate gratification, or is otherwise not consistent with reality as measured by mature adult standards.

Jung, Carl Gustav (1875–1961) Swiss psychiatrist and psychoanalyst, originally associated with Sigmund Freud, who later founded the school of analytic psychology.

justice In the context of mental health, a fair distribution and application of psychiatric procedures.

K

Kanner syndrome INFANTILE AUTISM.

kindling effect Electrophysiologic process in which repeated subthreshold stimulation of a neuron eventually generates an action potential; at the organ level, repeated subthreshold stimulation of an area of the brain results in the generation of a seizure.

kinesiology Study of body movement, especially in the context of its communicative function.

kinesthetic hallucination Hallucination primarily involving bodily movement.

kinesthetic sense Sense by which muscular motion, tension, position, and posture are perceived. Also called proprioception.

Kirkbride, Thomas S. (1809–1883) American psychiatrist, one of the 13 original founders of the American Psychiatric Association, noted for his 1854 manual advocating reform in the design of institutions for the mentally ill.

Klein, Melanie (1882–1960) British psychoanalyst and child analyst whose theories of early development departed from orthodox freudian thought.

Kleine–Levin syndrome Condition characterized by periodic episodes of hypersomnia and bulimia.

kleptomania Pathologic compulsion to steal.

Klinefelter syndrome Chromosomal anomaly in which there is an extra X chromosome (karyotype 47-XXY). Affected persons are male in development, with small firm testes, eunuchoid habitués, variable gynecomastia and other signs of androgen deficiency, and elevated gonadotropin levels; there is an increased frequency of mental retardation (about 25%) and other forms of psychopathology, particularly antisocial behavior and delinquency. *See also* Turner syndrome.

Klüver–Bucy syndrome Behavioral syndrome, first described in monkeys after bilateral temporal lobe ablation, characterized by increased oral, sexual, and aggressive behavior; loss of visual recognition: memory defect; and hyperreactivity to visual stimuli.

Kohut, Heinz (1913–1981) Important innovator in psychoanalysis, who originated the psychoanalytic school of self-psychology, which requires that patients become aware of their excessive needs for approval and self-gratification.

koro Acute delusional syndrome, seen in Southeast Asia and southern China, in which the patient suddenly becomes acutely anxious that his penis is shrinking and may disappear into his abdomen, in which case he will die.

Korsakoff psychosis Mental disorder seen in long-term alcoholics. Its major characteristic feature is profound memory impairment, particularly for recent events, for which the individual attempts to compensate by confabulation. Called alcohol amnestic disorder in DSM. *See also* Wernicke fluent encephalopathy.

Kraepelin, Emil (1865–1926) German psychiatrist noted for his pioneering work in psychiatric nosology and classification systems, and who differentiated between manic–depressive psychoses and dementia precox (schizophrenia). One of the last representatives of the predynamic school of psychiatry, he is often considered the father of descriptive psychiatry.

Kretschmer, Ernest (1888–1964) German psychiatrist noted for his theories of the relation of physique to character and personality.

kuru A prion disease, seen in Papua, New Guinea, caused by cannibalistic funeral rituals.

L

la belle indifférence Inappropriate attitude of calm or lack of concern about one’s disability; seen in patients with conversion disorder.

labile Unstable; characterized by rapidly changing emotions.

labile affect Affective expression characterized by repetitious and abrupt changes, unrelated to external stimuli.

labile mood Oscillations in mood between euphoria and depression or anxiety.

lability *See* emotional lability.

laconic speech Condition characterized by a reduction in the quantity of spontaneous speech; replies to questions are brief and unelaborated, and little or no unprompted additional information is provided. Occurs in major depression, schizophrenia, and organic mental disorders. Also called poverty of speech.

lacunar amnesia LOCALIZED AMNESIA.

lalophobia Abnormal fear of speaking. Also called glossophobia.

language disorder Persistent difficulties in the acquisition and use of language across modalities due to deficiencies in production (expressive) or comprehension (receptive-expressive) that include reduced vocabulary, difficulty with sentence formation, and impairments in discourse. Disorders lead to functional limitations in effective communication, social participation, and/or occupational performance. *See also* communication disorder.

lanugo Neonatal-like body hair; may be noted in anorexia nervosa.

lapsus linguae Slip of the tongue.

latah Culture-specific disorder found among Malaysians and characterized by either a sudden onset of unusual and inappropriate motor and verbal manifestations or by an echo reaction, in which the victim is compelled to imitate any words or actions to which they are exposed; in both forms affected persons cannot control or inhibit their behavior.

late language emergence DSM-5 disorder of early childhood characterized by delayed language onset in the absence of any developmental disabilities or delays.

late luteal phase dysphoric disorder (LLPDD) *See* premenstrual dysphoric disorder.

latency phase Period of psychosexual development after the phallic phase and succeeded by the genital phase, extending from about age 5 to the beginning of adolescence at age 12; during this phase there is an apparent cessation of sexual preoccupation and a blockade of libidinal impulses, and boys and girls are inclined to choose friends and join groups of their own sex.

latent content Hidden or unconscious meaning of symbolic representations, especially in fantasies and dreams. *See also* manifest content.

latent homosexuality Unexpressed conscious or unconscious homoerotic wishes that are held in check. Sigmund Freud’s theory of bisexuality postulated the existence of a constitutionally determined, although experientially influenced, instinctual masculine–feminine duality; normally, the opposite-sex component is dormant, but a breakdown in the defenses of repression and sublimation may activate latent instincts and result in overt homoeroticism.

latent schizophrenia Condition characterized by clear schizophrenic symptoms without a history of prior overtly psychotic schizophrenic episodes. In DSM the condition is subsumed under the category of schizotypal personality disorder.

learned autonomic control Learned regulation by a person of physiologic responses that are under autonomic nervous system control. *See also* biofeedback.

learned helplessness *See* helplessness.

learning disorders Disorders manifested in school-age children and characterized by specific difficulties in learning that are unrelated to intelligence. In DSM, disorders include reading disorder, dyscalculia, and disorder of written expression. *See also* dyslexia.

learning theory Approach to the understanding of human behavior that emphasizes the way in which learning comes about. Learning itself represents a change in behavior resulting from practice; through the application of certain laws of learning, learning theory attempts to explain the basic processes that are necessary for learning to occur.

lesbianism Female homosexuality. Also called sapphism.

Lesch–Nyhan syndrome A rare, X-linked disorder caused by a deficiency of an enzyme involved in purine metabolism. Characteristics include intellectual disability, seizures, self-mutilating behaviors (e.g., head banging, finger biting), microcephaly, choreoathetosis, and spasticity.

lethologica Momentary forgetting of a name or proper noun. *See* blocking.

leukotomy LOBOTOMY.

Lewin, Kurt (1890–1947) German psychologist who emigrated to United States in 1933. Two of his chief theoretical contributions are the field approach (field theory) and group dynamics, each of which has been useful in the experimental study of human behavior in a social situation.

Lewy body disease A type of dementia clinically similar to Alzheimer disease that is also accompanied by hallucinations and parkinsonian features. Often associated with Capgras syndrome. *See also* dementia of the Alzheimer type.

Liaison Committee on Medical Education (LCME) Group sponsored by the American Medical Association and the Association of American Medical Colleges that is principally involved in accrediting activities in undergraduate medical education.

liaison nursing Communication among nurses from various disciplines concerning the treatment of medical, surgical, and neuropsychiatric patients.

liaison psychiatry *See* consultation-liaison psychiatry.

libido In psychoanalysis, the psychic energy associated with the sexual drive (life instinct).

life instinct SEXUAL DRIVE.

lifetime expectancy Total probability of a person developing a disorder during a lifetime.

lifetime prevalence Measure at a point in time of the number of persons who had a given disorder at some time during their lives.

lilliputian hallucination Visual sensation that persons or objects are reduced in size, more properly regarded as an illusion. *See also* micropsia.

limbic system James Papez' proposal (1939) that a reverberating circuit or system, consisting of the hippocampus, hypothalamus, anterior thalamus, and cingulate gyrus, was the central nervous system localization for emotions; current models also include the amygdala and septal area, as well as other neuroanatomic areas. Various parts of the limbic system have been associated with emotions, sex drive, eating behavior, rage, violence, memory, and motivation.

lithium carbonate Lithium salt effective in the treatment of the manic phase of bipolar disorder and as a mood-stabilizing drug for chronic manic–depressive patients; may also be an effective antidepressant in some patients.

Little Albert Eleven-month-old boy who was conditioned to fear furry objects in experiments conducted by American psychologist John B. Watson (1920) to demonstrate that phobias are learned responses based on classical conditioning.

Little Hans Sigmund Freud's famous case history of a 5-year-old boy with a phobia. The boy's fear of horses, instead of his father, helped him to avoid hating his father, by whom he was threatened; and whom he also loved.

living will Legal document in which patients gives instructions to their physicians about withholding life support measures, for example, Do Not Resuscitate (DNR) if in extremis.

LCME Liaison Committee on Medical Education.

LLPDD Late luteal phase dysphoric disorder.

LNNB Luria-Nebraska Neuropsychological Battery.

lobotomy Neurosurgical procedure, introduced by Egas Moniz, in which one or more nerve tracts in a lobe of the cerebrum are severed. Prefrontal lobotomy is the ablation of one or more nerve tracts in the prefrontal area of the brain and has been used in the treatment of certain severe mental disorders that do not respond to other treatments. Also called leukotomy.

localized amnesia Partial loss of memory; amnesia restricted to specific or isolated experiences. Also called lacunar amnesia; patch amnesia.

locus ceruleus Nucleus (a collection of neuronal cell bodies) that is located in the pons (a region of the brain stem) and contains the largest population of noradrenergic neurons in a single identified brain region; these neurons project their axons to virtually every region of the brain and are thought to be involved in the regulation of mood, anxiety, and attention.

logorrhea Copious, pressured, coherent speech; uncontrollable, excessive talking; observed in manic episodes of bipolar disorder. Also called tachylogia; verbomania; volubility.

logotherapy Existential analysis based on spiritual values, rather than psychobiologic laws.

longitudinal study Examination or analysis carried out of a long period of time.

long-term memory Reproduction, recognition, or recall of experiences or information that were experienced in the distant past. Also called remote memory. *Compare* immediate memory; short-term memory.

loosening of associations Characteristic schizophrenic thinking or speech disturbance involving a disorder in the logical progression of thoughts, manifested as a failure to adequately verbally communicate; unrelated and unconnected ideas shift from one subject to another. *See also* tangentiality.

Lou Gehrig disease AMYOTROPHIC LATERAL SCLEROSIS.

LP Lumbar puncture.

LSD Lysergic acid diethylamide.

Lumbar puncture (LP) Use of a needle inserted into the lumbar spine intervertebral space to obtain cerebrospinal fluid for diagnostic purposes or to introduce local anesthetics or other drugs.

Luria-Nebraska Neuropsychological Battery (LNNB) Based on the work of the Russian psychologist A. R. Luria, the test assesses a wide range of cognitive functions: memory; motor functions; rhythm; tactile, auditory, and visual functions; receptive and expressive speech; writing; spelling; reading; and arithmetic. It is extremely sensitive for identifying specific types of problems (e.g., dyslexia, dyscalculia) rather than being limited to more global impressions of brain dysfunction; it also helps localize the various cortical zones that are involved in a particular function and is useful in establishing left or right cerebral dominance. The test is designed for persons at least 15 years of age; a children's version for use with 8 to 12 year olds is being developed.

lust dynamism Harry Stack Sullivan's term to describe clearly stated sexual desires and abilities.

lying Making statements that one knows consciously are false with intent to deceive; differentiated from confabulation, in which one is not conscious of lying.

lysergic acid diethylamide (LSD) A potent psychotogenic drug discovered in 1942 that produces psychotic-like symptoms and behavior changes, including hallucinations, delusions, and time-space distortions. Commonly called acid.

M

macropsia False perception that objects are larger than they really are. *Compare* micropsia.

magical thinking A form of dereistic thought; thinking similar to that of the preoperational phase in children (Jean Piaget), in which thoughts, words, or actions assume power (e.g., to cause or to prevent events). Common in dreams, certain mental disorders, and children.

magnetic resonance imaging (MRI) Diagnostic modality in which a patient's body is placed in a magnetic field and its hydrogen nuclei are excited by radiofrequency pulses at angles to the field's axis; resulting signals from the body's hydrogen ions are processed through a computer to produce an image; an excellent method for differentiating gray and white matter, visualizing cerebrospinal fluid, and estimating the size of ventricles.

Mahler, Margaret (1897–1985) Hungarian-born psychiatrist and child analyst who described the separation–individuation process in children.

maintenance drug therapy Stage in the course of chemotherapy at which, after the drug has reached its maximal efficacy, the dosage is reduced and sustained at the minimal therapeutic level that will prevent a relapse or exacerbation.

major affective disorder *See* major depression.

major depressive disorder Severe mood disorder characterized by one or more depressive episodes but no history of a manic episode; a DSM-5 diagnosis. *See also* mood disorder.

major depressive episode, melancholic type DSM diagnosis replacing involutional melancholia.

major epilepsy *See* grand mal epilepsy.

major motor seizure *See* grand mal epilepsy.

major neurocognitive disorder DSM-5 term for dementia.

major tranquilizer ANTIPSYCHOTIC.

maladaptive Referring to any mental activity or behavior that is dysfunctional or counterproductive with regard to a person's ability to cope effectively with the problems and stresses of life.

maladjustment Maladaptive reaction; poor, faulty, or inadequate coping ability.

male erectile disorder DSM sexual arousal disorder characterized by the recurrent and persistent partial or complete failure to obtain or maintain penile erection until the completion of the sex act. Also called erectile dysfunction. *See also* impotence.

male hypoactive sexual desire disorder Disorder characterized by an absence of or reduction in sexual fantasies and/or the desire for sexual activity in men

maleness Anatomic and physiologic features relating to the male's procreative functions.

malingering Feigning disease in order to achieve a specific goal, for example, to avoid an unpleasant responsibility. *Compare* factitious disorder.

malpractice Professional negligence including intentional or willful invasion of another's legally protected interests, such as battery or treatment without consent; must be demonstrated by the presence of *d*ereliction of a *d*uty *d*irectly causing *d*amages (the four D's of malpractice).

mania Mood abnormality characterized by elation; agitation; hyperactivity and hyperexcitability; accelerated thinking and speaking (flight of ideas); decreased need for sleep; and overinvolvement in pleasurable behavior. Characterizes the manic phase of bipolar I disorder when lasting more than a week or is severe enough to warrant hospitalization. *Compare* hypomania.

manic Denoting or characterized by mania. Sometimes used as a general description for a mentally disturbed person, often someone exhibiting violent or grossly irrational behavior.

manic–depressive illness Mood disorder characterized by severe alterations in mood that are usually episodic and recurrent. The unipolar type is characterized by either periodic episodes of mania (manic disorder) or periodic episodes of depression (depressive disorder). In the bipolar or circular type, the patient has at least one episode of each mood extreme and may alternate periodically between the two. Manic–depressive illness has been replaced by bipolar disorder in DSM.

manifest content That part of a dream or fantasy that a person remembers and reports; in a dream, it represents a disguised and symbolic expression of the latent content of the dream.

manipulation Exploitativeness; the maneuvering by patients to get their own way, characteristic of antisocial personalities.

mannerism Stereotyped gesture or expression that is peculiar to a given person.

mantra In transcendental meditation, a Sanskrit syllable or word that is repeated over and over in an effort to produce total relaxation and control over one's state of consciousness.

MAOI Monoamine oxidase inhibitor.

marasmus Progressive emaciation, especially in malnourished infants and children, in which weight is often below the third percentile and markedly below appropriate weight for height.

marathon Technique developed by George Bach and Frederick Stoller that incorporates a group meeting that usually lasts 8 to 72 hours (although some may last for a week), in which the sessions interrupted only for eating and sleeping; the leader works for the development of intimacy and the open expression of feelings. The time-extended group experience culminates in intense feelings of excitement and elation.

marijuana Dried leaves and flowers of *Cannabis sativa* (Indian hemp), which induce somatic and psychic changes in a person when smoked or ingested in sufficient quantity. Somatic changes include increased heart rate, rise in blood pressure, dryness of the mouth, and increased appetite; psychic changes include a dreamy-state level of consciousness, a sense of enhanced vividness of visual and auditory sensations, disturbances of time perception and memory-dependent, goal-directed behavior, disruptions of sequential thought processes, and euphoria and other alterations of mood. The compound believed to be responsible for most of its psychological effects is tetrahydrocannabinol (THC). In strong doses, marijuana can produce hallucinations and, at times, paranoid ideation. Also called pot; grass; weed; dope; Mary Jane.

marital counseling Process whereby a trained counselor helps married couples resolve problems that arise and trouble them in their relationship; the husband and the wife are seen by the same worker in separate and joint counseling sessions that focus on immediate family problems. The theory and techniques of this approach were first developed in social agencies as part of family case work.

marital problems Maladaptive behavior between spouses in various areas, for example, attitudes toward sex, contraception, child rearing, money handling.

marital therapy *See* marriage therapy.

marriage therapy Type of family therapy that involves the husband and the wife and focuses on the marital relationship as it affects the individual psychopathology of the partners. The rationale for the method is the assumption that psychopathologic processes within the family structure and in the social matrix of the marriage perpetuate individual pathologic personality structures, which find expression in the disturbed marriage and are aggravated by the feedback between partners.

masculine identity Inner sense of gender affiliation with males. *Compare* feminine identity.

masculine protest Adlerian doctrine that depicts a universal human tendency to move from a passive and feminine role to a masculine and active role; an extension of Alfred Adler's ideas about organic inferiority that became the prime motivational force in normal and neurotic behavior in his system. *See also* Adler, Alfred; inferiority complex.

masculinity–femininity scale Any scale on a psychological test that assesses the relative masculinity or femininity of the testee. Scales vary and may focus, for example, on basic identification with either sex or preference for a particular sex role and are strongly influenced by cultural definitions of masculinity and femininity.

masochism Paraphilia in which sexual gratification is derived from being physically or psychologically maltreated by a partner or oneself; first described by an Austrian novelist, Leopold von

Sacher-Masoch (1836–1895). *See also* sadism; sadomasochistic relationship.

masturbation Self-stimulation of the genitals for sexual pleasure. *See* autoerotism.

maternal deprivation Result of the premature loss or absence of the mother. In a broader sense, the lack of proper mothering may constitute a form of maternal deprivation which may lead to severe emotional disorders in infants and children.

maturational crises Normative stages of the life cycle that require adaptive coping or defense mechanisms.

maximum security unit That part of an institution for the mentally ill reserved for those who have committed crimes or who are considered dangerous to others.

MCE Medical care evaluation.

MDMA 3,4-Methylenedioxymethamphetamine.

mean Statistical measurement derived from adding a set of scores and then dividing by the number of scores. *See also* average.

mean deviation Statistical measure of the variability in a set of values defined as the sum of the absolute differences between the values and the mean divided by the number of values.

median Value in a set of values above and below which there are an equal number of values; for example, in the series 1, 3, 5, 9, 33 the value 5 is the median. *See also* average.

Medicaid Assistance program for certain needy and low-income persons, mandated by the federal government and financed by both federal and state governments; certain benefits, such as in- and outpatient hospital care, physicians' services, and laboratory tests, are required covered expenses, while other services may be provided at the state's option.

medical audit Retrospective assessment of the quality and utilization of care, including investigation of suspected problems, analysis of the problems identified, and a plan for corrective action. Also called medical care evaluation (MCE).

medical care evaluation (MCE) MEDICAL AUDIT.

medical ethics Moral set of principles that determine what is right or wrong and good or bad, as they pertain to professional responsibilities and opinions relating to the values of patients, patients' families, and society at large. For related issues, see the specific term.

medical record Written exposition describing a patient's diagnosis and treatment.

medical review Evaluation of the quality of care and assessment of medical care or services judged in terms of cost effectiveness.

Medicare Hospital and medical insurance system for the aged created in 1965 by amendments of the Social Security Act. Hospitalization in general hospitals for psychiatric conditions is covered in the same way as other conditions.

medication-induced movement disorders A group of disorders that include neuroleptic malignant syndrome, neuroleptic-induced parkinsonism, neuroleptic-induced acute akathisia, neuroleptic-induced tardive dyskinesia, neuroleptic-induced acute dystonia, and medication-induced postural tremor.

megalomania Morbid preoccupation with expansive delusions of power, importance, or wealth.

megavitamin therapy *See* orthomolecular treatment.

melancholia A severe depressive state characterized by anhedonia, early morning awakening, weight loss, psychomotor retardation, and profound feelings of guilt and hopelessness; often associated with suicidal ideation.

memory Process whereby what is experienced or learned is established as a record in the central nervous system (registration), where it persists with a variable degree of permanence (retention) and can be recollected or retrieved from storage at will (recall). For types of memory, see the specific term.

menarche Onset of menstrual function.

mendacity Pathologic lying.

mens rea Legal term meaning an intent to do harm.

mental aberration Pathologic deviation from normal thinking; unrelated to a person's intelligence.

mental age Measure of mental ability as determined by standard psychological tests.

mental deficiency Term often used interchangeably with mental retardation, applied to individuals with an IQ of less than 70. *See* intellectual developmental disorder.

mental disorder, mental disease Psychiatric illness or disease whose manifestations are primarily characterized by behavioral or psychological impairment of function, measured in terms of deviation from some normative concept.

mental health State of emotional well-being in which persons are able to function comfortably within their society and in which their personal achievements and characteristics are satisfactory to them.

mental hygiene 1. Conditions and practices conducive to the establishment and maintenance of mental health. **2.** Field dealing with the prevention and early treatment of mental illness. *See also* orthopsychiatry.

mental illness Mental disorder; any serious impairment of adjustment; any psychiatric disorder listed in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* or in the World Health Organization's *International Classification of Diseases.*

mental retardation Term previously assigned to the condition that DSM-5 now refers to as intellectual disability. Degrees of condition were historically classified according to IQ: mild (50–55 to 70), moderate (35–40 to 50–55), severe (20–25 to 30), and profound (below 20–25). *See* intellectual disability.

mental status General functional condition of mental and behavioral processes as determined by psychiatric assessment of a variety of areas of functioning, such as state of consciousness, mood and affect, thinking and speech, motor behavior, general knowledge, memory, calculation, judgment, abstraction, and insight.

mental status examination (MSE) Clinical evaluation, and the sum total of the examiner's observations and impressions, of a person's psychological and behavioral experiences. MSE includes specific references to such areas as attitude and general behavior during interview and to sensorium, mental capacity, and other areas of function.

merycism RUMINATION DISORDER OF INFANCY.

mescaline Hallucinogenic drug obtained from buttons of the peyote cactus that has psychotomimetic effects similar to those of LSD.

mesmerism Early term for hypnosis.

mesomorphic Muscular, with a well-proportioned physique; one of Sheldon's constitutional types. *Compare* ectomorphic; endomorphic.

meta-analysis General consensus of many studies that results in one common data base that helps to eliminate bias.

metapsychiatry Branch of psychiatry concerned with altered state of consciousness, holistic health, and parapsychology.

metapsychology Branch of theoretical or speculative psychology concerned with theories, hypotheses, or phenomena that are largely beyond the realm of empirical verification. *See also* parapsychology.

methadone Orally active synthetic narcotic agent that is an effective analgesic with pharmacologic actions qualitatively very similar to those of morphine. Its use in the maintenance treatment of heroin addicts is based on its extended duration of action in suppressing withdrawal symptoms in persons physically dependent on opiates.

methadone maintenance treatment *See* methadone.

3-methoxy-4-hydroxyphenylglycol (MHPG) Most significant metabolic product from central nervous system norepinephrine; excreted in urine.

3,4-methylenedioxymethamphetamine (MDMA) Hallucinogen more commonly known as ecstasy.

methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) Contaminant of an illicitly made synthetic heroin that induces irreversible parkinsonism.

metonymy Speech disturbance common in schizophrenia in which the affected person uses a word or phrase that is related to the proper one but is not the precise one that would ordinarily be used; for example, the patient speaks of consuming a "menu," rather than a “meal,” or refers to losing the “piece of string” of the conversation, rather than the “thread” of the conversation. *See also* paraphasia; word approximation.

Metrazol treatment Obsolete form of convulsive therapy in which a convulsive seizure is induced by injection of the drug Metrazol (pentylenetetrazol).

Meyer, Adolf (1866–1950) American psychiatrist known for his concept of psychobiology and “commonsense” psychiatry.

MHPG 3-Methoxy-4-hydroxyphenylglycol.

microcephaly Condition in which the head is unusually small as a result of defective brain development and premature ossification of the skull.

micropsia False perception that objects are smaller than they really are. Sometimes called lilliputian hallucination. *Compare* macropsia.

MID Multi-infarct dementia.

middle age, middle adulthood That period considered to be the prime of life, between the ages of approximately 40 to 65 years, following the termination of early adulthood.

middle insomnia Waking up after falling asleep without difficulty and then having difficulty in falling asleep again. *Compare* initial insomnia; terminal insomnia.

midlife crisis Syndrome that occurs in both sexes as they enter their early 40s or 50s, consisting of depression, substance abuse, concern about growing old, and fear of death; more common in adults with prior psychiatric illness than in normal adults.

migraine Severe unilateral, throbbing headache that appears periodically, usually beginning during the teenage years and continuing to recur with diminishing frequency during advancing years. Classic migraine is a syndrome ushered in by a transient derangement of neurologic function, often a visual disturbance, followed by hemicranial headache, nausea, and vomiting, all of which last for anywhere from several hours to 1 or 2 days. Common migraine has the same character and time course but develops without any prodrome. Patients with migraine are typically described as tense, meticulous, hard-driving persons who appear to have a high frequency of associated emotional conflicts.

mild mental retardation *See* mental retardation.

mild neurocognitive disorder Mild decline in cognitive function that is differentiated from the normal cognitive changes associated with aging; memory is impaired for age and education, however, general cognitive function and activities of daily living remain intact. In DSM-5 diagnosed as mild neurocognitive disorder due to multiple etiologies or unspecified neurocognitive disorder.

milieu therapy Treatment that emphasizes appropriate socioenvironmental manipulation for the benefit of the patient; the setting is usually the psychiatric hospital. *See also* moral treatment; therapeutic community.

mimicry Simple, imitative motion activity of childhood.

minimal brain dysfunction Behavioral syndrome of childhood, now called attention-deficit hyperactivity disorder, characterized by learning difficulties, decreased attention span, distractibility, hyperactivity, impulsiveness, emotional lability, and, often, disturbances in perceptuomotor and language development; its psychopathologic mechanisms have not been defined. The term implies neurologic causation, but in most cases there are no major unequivocal neurologic signs and there has been a tendency to apply the term as a convenient explanatory label to any child presenting with a specific learning difficulty or behavioral dysfunction. Although the syndrome has been diagnosed with increasing frequency over the past 20 years, its validity as a diagnostic entity has been questioned, and it is not listed as such in DSM.

Minnesota Multiphasic Personality Inventory (MMPI) Questionnaire type of psychological test for persons age 16 and over consisting of true–false statements that are coded in various scales that assess different dimensions of the person's personality structure and measure the closeness of fit with various psychiatric diagnostic categories.

minor epilepsy PETIT MAL EPILEPSY.

minor tranquilizer ANXIOLYTIC.

Mitchell, S. Weir (1830–1914) American neurologist known for his concept of rest treatment to cure emotional disorders.

mixoscopia Sexual perversion in which a person attains orgasm by watching their love object make love with another person.

MMPI Minnesota Multiphasic Personality Inventory.

M'Naghten rule Set of rules established in 1843 in England to determine criminal responsibility, based on the person's ability to distinguish right from wrong. *See also* Durham rule; American Law Institute formulation of insanity.

mode In a set of measurements, the value that appears most frequently. *See also* average.

modeling Learning of a new skill or behavior by observation and imitation of another person.

Model Penal Code American Law Institute's test of criminal responsibility: "A person is not responsible for criminal conduct if at the time of such conduct, as a result of mental disease or defect, he lacked substantial capacity to either appreciate the criminality (wrongfulness) of his conduct or to conform his conduct to the requirement of the law."

moderate mental retardation *See* mental retardation.

molecular psychiatry Etiologic view of psychiatry based on neurotransmitter chemicals as measured in molecular quantities.

mongolism Obsolete and culturally offensive term for Down syndrome.

monoamine Biogenic amine or class of neurotransmitters that consists of three catecholamines (dopamine, norepinephrine, and epinephrine), an indolamine (serotonin), a quaternary amine (acetylcholine), and an ethylamine (histamine).

monoamine oxidase (MAO) Enzyme that deaminates monoamines and that functions, in the nervous system by breaking down monoamine neurotransmitters (e.g., norepinephrine, epinephrine, dopamine, and serotonin); has been subdivided into Type A and Type B.

monoamine oxidase inhibitor (MAOI) Agent that inhibits the enzyme monoamine oxidase (MAO). MAO inhibitors are highly effective as antidepressants and are also used in the treatment of panic disorders and some eating disorders.

monomania Morbid mental state characterized by preoccupation with one subject. Also called partial insanity.

monosymptomatic hypochondriacal psychosis A delusional disorder consisting of somatic delusions; common delusions include infestation, dysmorphophobia, and foul body odors or halitosis.

monozygotic twins Twins who develop from a single fertilized ovum. In contrast, dizygotic or fraternal twins develop from two separately fertilized ova. Also called identical twins.

mood Pervasive and sustained feeling tone that is experienced internally and that, in the extreme, can markedly influence virtually all aspects of a person’s behavior and their perception of the world. Distinguished from affect, the external expression of the internal feeling tone. For types of mood, see the specific term.

mood-congruent delusion Delusion with content that is mood appropriate, for example, a depressed patient who believes they are responsible for the destruction of the world.

mood-congruent hallucination Hallucination with content that is consistent with either a depressed or a manic mood, for example, a depressed patient hearing voices telling the patient they are a bad person; a manic patient hearing voices telling the patient they have inflated worth, power, or knowledge.

mood-congruent psychotic features DSM term that refers to hallucinatory or delusional phenomena with contents that consistently reflect the mood of a manic or depressed patient.

mood disorders Qualitative terms used to describe the striking symptomatic features or timing patterns in various patients with moods disorders. Descriptive qualifiers include psychotic, melancholic, atypical, and catatonic features; patterns include postpartum onset, rapid cycling, and seasonal pattern.

mood-incongruent delusion Delusion based on incorrect reference about external reality, with content that has no association to mood or is mood-inappropriate, for example, a depressed patient who believes they are the new Messiah.

mood-incongruent hallucination Hallucination not associated with real external stimuli, with content that is not consistent with either depressed or manic mood, for example, in depression, hallucinations not involving such themes as guilt, deserved punishment, or inadequacy; in mania, not involving such themes as inflated worth or power.

mood-incongruent psychotic features DSM term that refers to hallucinatory or delusional phenomena with contents that do not consistently reflect the mood of a manic or depressed patient.

mood stabilizer Drug used to prevent mood swings in bipolar disorder.

mood swing Oscillation of a person’s emotional feeling tone between periods of elation and periods of depression.

moral treatment Philosophy and technique of treating mental patients, developed in the 18th century, that involved the establishment of a humane therapeutic environment designed to maximize the chances for recovery; considered the forerunner of current milieu therapy.

morbid jealousy Marked or severe jealousy. Also called pathologic jealousy.

morbid perplexity Condition seen in schizophrenia in association with loss of ego boundaries; the patient exhibits profound confusion about his own identity and the meaning of existence.

moron Obsolete and pejorative term for a classification of mental retardation in which an individual has a mental age of approximately 8 years.

morphine Major phenanthrene alkaloid opium derivative that produces a combination of central nervous system excitation and depression; repeated administration can lead to tolerance, physical dependence, and psychological dependence; used as an analgesic, sedative, and anxiolytic.

Moses and Monotheism Title of a book by Sigmund Freud (1939) in which he undertook a historical but frankly speculative reconstruction of the personality of Moses and examined the concept of monotheism and the abiding effect of the patriarch on the character of the Jews; one of Freud's last works, it bears the imprint of his latter-day outlook and problems.

mother surrogate In psychoanalysis, a mother substitute onto whom a patient projects their mother image and responds to unconsciously, in an inappropriate and unrealistic manner, with the feelings and attitudes that they had toward the real mother.

motivation Force or energy associated with an internal state that propels a person to engage in behavior to satisfy a need or desire.

motor aphasia Aphasia in which understanding is intact but the ability to speak is lost. Also called Broca's, expressive, or nonfluent aphasia.

motor disorders A class of neurodevelopmental disorders defined by a delay in meeting expected milestones for motor skills, based on age and intelligence, which significantly impairs daily functioning. The major types of motor disorders include developmental coordination disorder, stereotypic movement disorder, and tic disorder.

mourning Syndrome following loss of a loved one, consisting of preoccupation with the lost individual, weeping, sadness, and repeated reliving of memories. *See also* bereavement; grief.

MPD Multiple personality disorder.

MPTP Methyl-4-phenyl-1,2,3,6-tetrahydropyridine.

MRI Magnetic resonance imaging.

multi-infarct dementia *See* vascular dementia.

multiple delusions Concurrent delusions in which the delusions need not be interconnected.

multiple personality disorder *See* dissociative identity disorder.

multiple regression Multivariate analysis in which a scaled variable is correlated with a linear combination of independent or predictor variables.

multiple therapy CO-THERAPY.

multivariate analysis Methods for considering the relationship of three or more variables, including multiple regression, multivariate analysis of variance, discriminant analysis, canonical correlation, and factor analysis.

Munchausen syndrome Alternate term for factitious disorder.

muscarine Naturally occurring alkaloid found in certain species of poisonous mushrooms; symptoms of mushroom poisoning develop within 30 minutes of ingestion and consist of vomiting, diarrhea, bronchospasm, hypertension, delirium, and shock.

mutism Organic or functional absence of the faculty of speech. *See also* stupor.

mydriasis Dilation of the pupil; sometimes occurs as an autonomic (anticholinergic) side effect of some phenothiazines and tricyclic antidepressant drugs.

mysophobia Abnormal fear of dirt or contamination.

N

naloxone Opiate antagonist that, unlike nalorphine which it has largely replaced as a narcotic antagonist is essentially devoid of agonist effects. In opioid-dependent persons, small doses of naloxone precipitate moderate to severe withdrawal symptoms, an effect that can be used to diagnose physical dependence on narcotic drugs; however, the most important use of naloxone is in the treatment of narcotic overdosage. (Trade name, Narcan.)

NAMI National Alliance for the Mentally Ill.

narcissism Self-love linked to autoerotism but devoid of geniality. In psychoanalytic theory, it is divided into primary and secondary types: primary narcissism, the early infantile phase of object relationship development, when the child has not differentiated the self from the outside world and all sources of pleasure are unrealistically recognized as coming from within the self, giving the child a false sense of omnipotence; secondary narcissism, when the libido, once attached to external love objects, is redirected back to the self. *See also* autistic thinking.

narcissistic personality disorder Personality disorder characterized by a grandiose sense of self-importance, preoccupation with fantasies of unlimited success, exhibitionistic need for attention and admiration, exaggerated responses to criticism or other perceived threats to self-esteem, and disturbance in interpersonal relationships.

narcoanalysis NARCOTHERAPY.

narcolepsy Sleep disorder characterized by recurrent, brief, uncontrollable episodes of sleep; often associated with cataplexy. *See also* hypersomnia.

narcosis Drug-induced stupor.

narcosynthesis NARCOTHERAPY.

narcotherapy Psychotherapy conducted with the patient in a drug-induced stupor or semiconscious state, as by the administration of barbiturates; originally used in the treatment of acute disorders arising in the setting of military combat; also called narcoanalysis; narcosynthesis.

narcotic Any natural or synthetic drug producing morphine-like analgesia, drowsiness, mental clouding, and, with prolonged use, physical dependence. *See also* opioid.

narcotic blockade Inhibition of the euphoric effects of even high doses of intravenous opioids, such as heroin, by the use of other drugs, such as methadone.

National Alliance for the Mentally Ill (NAMI) Self-help group and influential national advocacy network comprised of and supported by the family members and parents of schizophrenic patients.

National Board of Medical Examiners (NBME) National board, founded in 1915, for testing medical students by three examinations given during different stages of their medical education to qualify for licensure to practice medicine in the United States; the diploma is accepted in 50 states for obtaining licensure without further examination by the state.

National Center for Complementary and Alternative Medicine (NCAAM) A government agency, part of the National Institute of Health that tests and evaluated alternative medicine therapies, for example, nutritional supplements, prayer, yoga, among many others.

National Institute of Alcohol Abuse and Alcoholism (NIAAA) *See* Alcohol, Drug Abuse, and Mental Health Administration.

National Institute of Drug Abuse (NIDA) *See* Alcohol, Drug Abuse, and Mental Health Administration.

National Institute of Mental Health (NIMH) *See* Alcohol, Drug Abuse, and Mental Health Administration.

National Mental Health Association (NMHA) Voluntary citizens' organization, founded in 1909 as the National Committee on Mental Hygiene, to promote mental health.

National Training Laboratories (NTL) Organization started in 1947 at Bethel, Maine, to train professionals who work with groups; interest in personal development eventually led to sensitivity training and encounter groups.

natural group Any group that tends to evolve spontaneously in human civilization, such as a kinship, tribal, or religious group in contrast to various contrived groups or aggregates of people who meet for a relatively brief time to achieve some goal.

NBME National Board of Medical Examiners.

NCAAM National Center for Complementary and Alternative Medicine.

necromania Pathologic preoccupation with dead bodies.

necrophilia Sexual attraction to or sexual contact with dead bodies.

necrophobia Abnormal fear of dead bodies.

need Ambiguous term referring to an internal state of unsatisfaction or tension related to a wish, urge, desire, or other endogenous behavioral stimulus; classically, it implied an innate or instinctual internal stimulus. *See also* drive; instinct; motivation.

needle phobia The persistent, intense, pathologic fear of receiving an injection.

negative reinforcement *See* reinforcement.

negative symptoms In schizophrenia: flat affect, alogia, abulia, and apathy.

negative transference *See* transference.

negativism Verbal or nonverbal opposition or resistance to outside suggestions and advice; commonly seen in schizophrenia in which the patient resists any effort to be moved or does the opposite of what is asked.

neologism New word or phrase whose derivation cannot be understood; often seen in schizophrenia. It has also been used to mean a word that has been incorrectly constructed but whose origins are nonetheless understandable (e.g., “headshoe” mean “hat”), but such constructions are more properly referred to as word approximations.

nervous breakdown Nonspecific lay term for mental disorder, often implying an acute decompensation.

neurasthenia Condition characterized by vague functional physical and mental fatigue; not used in DSM.

neurochemistry Branch of chemistry concerned with nervous system components, metabolism, and function.

neurocirculatory asthenia DA COSTA SYNDROME.

neuroendocrinology Subspecialty area concerned with the anatomical and functional relationships between the nervous system and the endocrine apparatus. *See also* psychoendocrinology.

neurohormone Chemical messengers formed by neurosecretory cells and released into the bloodstream.

neuroleptic drug ANTIPSYCHOTIC.

neuroleptic malignant syndrome (NMS) Life-threatening complication of antipsychotic treatment, including such symptoms as muscle rigidity and dystonia, akinesia, mutism, obtundation, and agitation.

neurologic amnesia **1.** Auditory amnesia: loss of ability to comprehend sounds or speech. *See also* Wernicke aphasia. **2.** Tactile amnesia: loss of ability to judge the shape of objects by touch. *See also* astereognosis. **3.** Verbal amnesia: loss of ability to remember words. **4.** Visual amnesia: loss of ability to recall or recognize familiar objects or printed words.

neurologic disorders Disorders involving the organization and function of nervous tissue, including diseases of the peripheral nerves, the spinal cord, and the brain, which are based on organic pathology.

neurologist Physician who specializes in diseases of the nervous system.

neurology Medical specialty concerned with organic diseases of the nervous system.

neuron Nerve cell, consisting of cell body, dendrites, and neuroaxon, which is the basic functional unit of the nervous system.

neuronal plasticity Ability of the central nervous system to alter its functional organization; a capacity for change that serves as a mechanism to shape neuronal circuitry during development, to change the way information is processed, and to compensate for injury.

neurophysiology Physiology of the nervous system.

neuropsychiatry Medical specialty that combines psychiatry and neurology.

neuroreceptors Neuronal and other nervous tissue sites that are stimulated or inhibited by neurotransmitters and/or psychotropic drugs.

neuroscience Branch of the life sciences concerned with the anatomy, physiology, biochemistry, or molecular biology of nerves and nervous tissue and, especially, with their relation to behavior and learning.

neurosis Mental or psychiatric disorder characterized primarily by anxiety, which may either be directly experienced and dominate the clinical picture or be unconsciously controlled and modified by various psychological mechanisms to produce other subjectively distressing and ego-alien symptoms. Although neuroses are not accompanied by gross distortion of reality or severe personality disorganization, normal functioning is impaired by the person’s symptoms; they have no organic basis, are relatively persistent, and are treatable. Also called psychoneurosis.

neurotic disorder *See* neurosis.

neurotic process Process in which unconscious conflict is expressed by neurotic symptoms or personality disturbance.

neurotoxin Agent, such as an antibody, causing destruction to nervous tissue.

neurotransmitter Any specific chemical agent released by a presynaptic cell that, upon excitation, crosses the synapse to stimulate or inhibit the postsynaptic cell.

NIAAA National Institute of Alcohol Abuse and Alcoholism.

nicotine dependence Inability to control the use of and dependence on tobacco. Nicotine dependence and withdrawal are classified as disorders in DSM.

NIDA National Institute of Drug Abuse.

night eating syndrome Eating disorder characterized by the consumption of large amounts of food after dinner; generally associated with insomnia and a lack of daytime appetite. In DSM-5, included in “Other Specified Feeding or Eating Disorders.”

night hospital Part-time hospital facility in which patients function in the outside world during the day but return to the hospital at night. *Compare* day hospital; weekend hospital.

nightmare Anxiety attack experienced while having terrifying or frightening dreams; characterized by mild anxiety, good recall of the dream, and mild autonomic reactions. *Compare* nightmare disorder; sleep terror disorder.

nightmare disorder Sleep disorder characterized by frequent nightmares and associated insomnia resulting from a fear of sleep.

night-terror SLEEP TERROR DISORDER.

nihilism Delusion of the nonexistence of the self or part of the self; also refers to an attitude of total rejection of established values or extreme skepticism regarding moral and value judgments.

nihilistic delusion Delusional disorder in which patients believe the world and everything related to it have ceased to exist.

NIMH National Institute of Mental Health.

nitrogen narcosis RAPTURE-OF-THE-DEEP SYNDROME.

NMS Neuroleptic malignant syndrome.

noesis Revelation in which immense illumination occurs in association with a sense that one has been chosen to lead and command. Can occur in manic or dissociative states.

nominal aphasia Aphasia characterized by difficulty in giving the correct name of an object. *See also* anomia; amnestic aphasia.

nonaffective hallucination Hallucination whose content is apparently unrelated to either depression or elation. *Compare* illusion.

noncomplementary role *See* role.

noncompliance with medical treatment DSM category used when a patient does not cooperate with the physician in their medical care; motivated by denial of illness, by religious beliefs, or by decisions based on personal value judgments about the advantages and disadvantages of the proposed treatment.

nondirective approach Technique in which the therapist follows the lead of the patient in the interview, rather than introducing their own theories and directing the course of the interview; applied in both individual and group therapy.

nonfluent aphasia MOTOR APHASIA.

nonparametric Denoting statistical methods that do not require restrictive assumptions about population distributions.

nonreactive depression ENDOGENOUS DEPRESSION.

nonsuicidal self-injury A DSM-5 disorder in which the patient intentionally inflicts damage to their body to induce bruises, bleeding, or pain without suicidal ideation. Also called self-mutilation.

non–24-hour sleep–wake circadian rhythm disorder A sleep–wake disorder characterized by a circadian period that is not congruent with the external environment; common in the blind, visually impaired, and those with traumatic brain injury.

noradrenaline NOREPINEPHRINE.

norepinephrine Catecholamine neurotransmitter substance liberated by adrenergic postganglionic neurons of the sympathetic nervous system; the precursor of epinephrine present in the adrenal medulla and also found in many areas of the brain. A disturbance in its metabolism at important brain sites has been proposed as an important factor in the causation of mood disorders. Also called noradrenaline.

normality In psychiatry, the absence of mental illness; that state of mental health about which there exists a scientific and professional consensus. The normal person has a sense of their own worth and maintains a confident and purposeful self-image.

nosology Science of the classification of diseases.

nosophobia PATHOPHOBIA.

not-for-profit hospital Any hospital not maintained for the purpose of monetary gain. *Compare* for-profit hospital.

not otherwise specified (NOS) Term used in previous editions of DSM to describe unusual or uncharacteristic variations of different mental disorders; for example, eating disorder NOS, tic disorder NOS, pervasive developmental disorder NOS.

NREM sleep Non-REM sleep consisting of four stages, based on EEG findings: stage 1: low amplitude, first frequency immediately after sleep begins; stage 2: cycles of 12 to 16 Hz called sleep spindles; stages 3 and 4: increased amplitude and slow frequency EEG changes. *Compare* REM sleep.

NTL National Training Laboratories.

nuclear family Immediate members of a family, including the parents and the children.

nuclear magnetic resonance (NMR) *See* magnetic resonance imaging.

null hypothesis Research term for the hypothetical assumption that any difference observed between two groups or conditions or between a particular group and the general population is purely accidental and due to chance alone. *See also* Type 1 error; Type 2 error.

nyctophobia Abnormal fear of darkness or of the night.

nymphomania Abnormal, excessive, insatiable desire in a female for sexual intercourse. *Compare* satyriasis.

O

objective tests Typically, pencil-and-paper tests based on specific items and questions and that yield numerical scores and profiles easily subjected to mathematical or statistical analysis; for example, Minnesota Multiphasic Personality Inventory (MMPI).

object permanence As described by Jean Piaget, the child's (birth to 2 years) ability to understand that objects have an existence independent of their involvement with the object; the child is able to maintain a mental image of a person or object even though it is not present and not visible.

object relation Emotional attachment formed between one person and another, as opposed to interest in and love for oneself.

OBS Organic brain syndrome.

OSA Obstructive sleep apnea syndrome.

obsession Persistent and recurrent idea, thought, or impulse that cannot be eliminated from consciousness by logic or reasoning; obsessions are involuntary and ego-dystonic. *See also* compulsion.

obsessional jealousy disorder Descriptive condition in which a person has persistent thoughts about infidelity in a partner or spouse; not a diagnosis in DSM-5.

obsessive-compulsive disorder (OCD) Neurotic disorder characterized by the persistent recurrence of obsessions and compulsions.

obsessive-compulsive personality Personality disorder characterized by perfectionism, overconscientiousness, and excessive inhibition with regard to self-expression and relaxation. Also called anancastic personality.

obstructive sleep apnea syndrome (OSA) A dyssomnia characterized by periods of functional obstruction of the upper airway during sleepy, resulting in degreases in oxygen, after which breathing (at least briefly) resumes normally.

occupational problem Maladaptive patterns related to work. For example, job dissatisfaction and uncertainties about career choices.

occupational psychiatry Area of psychiatry concerned with mental illness in industry, including the psychiatric aspects of absenteeism, vocational adjustment, operational fatigue, and accident proneness.

occupational therapy Form of therapy in which the patient engages in useful, purposeful activities in a therapeutic social setting involving interaction with other patients and hospital personnel.

OCD Obsessive-compulsive disorder.

oedipal phase PHALLIC PHASE. *See also* Oedipus complex.

Oedipus complex Constellation of feelings, impulses, and conflicts in the developing child that concern sexual impulses and attraction toward the opposite-sex parent and aggressive, hostile, or envious feelings toward the same-sex parent; real or fantasied threats from the same-sex parent result in the repression of those feelings. Development of the Oedipus complex coincides with the phallic phase of psychosexual development; as one of Freud's most important concepts, the term was originally applied only to the male. *Compare* Electra complex.

olfactory hallucination Hallucination primarily involving smell or odors; most common in medical disorders, especially in the temporal lobe.

Olfactory reference syndrome Alternate term for monosymptomatic hypochondriacal psychosis with delusions of foul body odor.

oligophrenia Old term for mental retardation.

OMD Organic mental disorder.

OMS Organic mental syndrome.

onanism Coitus interruptus. Sometimes used interchangeably with masturbation.

oneiroid state Dream-like state in which a patient is deeply perplexed and not fully oriented in time and place.

ontogenetic Pertaining to the development of an individual person. *Compare* phylogenetic.

open hospital Mental hospital without locked doors or physical restraints.

operant behavior *See* conditioning.

operant conditioning *See* conditioning.

operational definition Means of defining a word so that it is scientifically measurable.

opiate Morphine or other alkaloids derived from opium, an extractable product of the poppy plant *Papaver somniferum* or its variety, *P. album. See also* opioid.

opiate receptor Specific site on a neuronal cell surface that interacts with opiate drugs and mediates their pharmacologic actions; presumably, it also mediates the physiologic functions of endogenous opioids, such as the enkephalins and endorphins.

opioid Any chemical substance that has morphine-like pharmacologic activity; for example, narcotic analgesics, such as methadone and meperidine, and endogenous opioid peptides, the enkephalins and the endorphins. *See also* opiate.

opioid peptide Peptide that exerts morphine-like actions through interactions with opiate receptors; for example, endorphin, enkephalin.

oppositional defiant disorder Mental disorder of childhood characterized by pervasive negativism, continuous argumentativeness, and an unwillingness to comply with reasonable suggestions and persuasion.

oral dyskinesia *See* tardive dyskinesia.

oral phase Earliest stage in psychosexual development, which lasts through the first 18 months of life. During this period the oral zone is the center of the infant's needs, expression, and pleasurable erotic experiences; it has a strong influence on the organization and development of the child's psyche.

organic A term used in psychiatry to denote a structural change or a chemical imbalance in the brain to account for a mental disorder. The term is still in common use but is no longer part of the official DSM nomenclature. *See also* organic disorder.

organic anxiety syndrome Syndrome characterized by prominent, recurrent panic attacks or generalized anxiety attributable to some clearly defined organic factor; cognitive functioning may be adversely affected as a secondary phenomenon.

organic brain syndrome (OBS) ORGANIC MENTAL SYNDROME.

organic delusional disorder Disorder characterized by the presence of prominent delusions in a state of full wakefulness and alertness than can be attributed to some clearly defined organic factor.

organic disorder Illness that is caused by structural or biochemical pathology. The term is not used in DSM but remains in common use. *Compare* functional disorder.

organic hallucinosis Disorder characterized by prominent or persistent hallucinations in a state of full wakefulness that can be attributed to some specific organic factor.

organic mental disorder (OMD) Mental disorder caused by transient or permanent brain dysfunction attributable to specific organic factors. In DSM, the organic mental disorders include dementias arising in the senium and presenium, psychoactive substance-induced organic mental disorders, and organic disorders associated with Axis III physical disorders or conditions or whose etiology is unknown. Also called organic psychosis.

organic mental syndrome (OMS) Psychological or behavior abnormality associated with transient or permanent dysfunction of the brain of unknown or unspecified etiology. OMS is characterized by a combination of disorientation to time, place, or person; signs of psychosis, for example, hallucinations or delusions; and impaired judgment and impulse control. The organic mental syndromes include delirium, dementia, amnestic syndrome, organic delusional syndrome, organic hallucinosis, organic mood syndrome, organic anxiety syndrome, organic personality syndrome, intoxication, and withdrawal; if the cause is known, the condition is called organic mental disorder. Also called organic brain syndrome. Both terms are considered obsolete given their implied mind-brain dichotomy.

organic mood syndrome Obsolete disorder characterized by either a depressive or a manic mood attributed to a clearly organic factor.

organic personality syndrome Term for frontal lobe syndrome.

organic psychosis ORGANIC MENTAL DISORDER.

orgasm Sexual climax, the peak psychophysiologic reaction to sexual stimulation.

orgasm disorders One of the major categories of sexual dysfunctions. Characterized by the recurrent and persistent inability to achieve orgasm. Also called orgasmic dysfunction.

orgasmic phase Stage in the human sexual response cycle, as described by Masters and Johnson, in which orgasm occurs.

orgasmic platform Outer third of the vagina, which displays marked vasocongestion and contraction in the plateau phase of the female sexual response cycle, as described by Masters and Johnson.

orientation State of awareness of oneself and one's surroundings in terms of time, place, and person.

original thirteen *See* American Psychiatric Association.

orphan drug Any pharmacologic agent that is not economical to produce because the condition for which it is indicated affects relatively few persons.

Othello syndrome Eponym for conjugal paranoia.

orthomolecular psychiatry Theoretical school of thought proposing that psychiatric illnesses are due to biochemical abnormalities that result in increased requirements for specific substances, such as vitamins.

orthomolecular treatment Treatment that entails administration to psychiatric patients of large doses (megadoses) of vitamins (megavitamin therapy), trace elements, or other substances.

orthopsychiatry Interdisciplinary approach to the study and practice of maintaining or restoring mental health, involving principles derived from psychiatry, psychology, sociology, social work, medicine, and other fields; particular emphasis is placed on preventive techniques to promote healthy emotional development and growth, especially mental hygiene.

orthostatic hypotension Reduction in blood pressure brought about by a shift from a recumbent position to an upright position; observed as a side effect of several psychotropic drugs.

other-directed person Person who is readily influenced and guided by the attitudes and values of other people. *Compare* inner-directed person.

other interpersonal problems DSM category describing problems between one or more persons that cause sufficient strain to bring a person into contact with the mental health care system; the stress-inducing circumstances, coping mechanism, and symptoms that have brought someone to seek consultation or treatment must be individually evaluated.

other specified family circumstance DSM category used when the focus of attention or treatment is not due to a mental disorder or a parent–child or marital problem; may involve in-laws, grandparents, or other family members.

outcome variable DEPENDENT VARIABLE.

outpatient Ambulatory patient receiving medical care who does not require hospitalization.

ovarian dysgenesis TURNER SYNDROME.

overactivity Abnormality in motor behavior that can manifest itself as psychomotor agitation, hyperactivity (hyperkinesis), tics, sleepwalking, or compulsions.

overanxious disorder Anxiety disorder of childhood or adolescence characterized by excessive worrying and fearful behavior grossly disproportionate to the magnitude of the real environmental stress or threat.

overcompensation Conscious or unconscious exaggerated correction or compensation for a real or imagined physical or psychological deficit.

overdetermination Concept that phenomena such as dreams and neurotic symptoms reflect the operation of multiple causative factors, particularly with regard to symbolic meaning or significance.

overt homosexuality Behaviorally expressed homosexuality, as distinct from unexpressed conscious or unconscious homoerotic wishes that are held in check (latent homosexuality).

overvalued idea False or unreasonable belief or idea that is sustained beyond the bounds of reason. It is held with less intensity or duration than a delusion but is usually associated with mental illness.

P

pain disorder Disorder characterized by multiple, recurrent pain complaints with no apparent physical cause; symptoms are linked to psychological factors. Pain disorder is a variant of somatic symptom disorder in which pain symptoms are the predominant physical complaints. The field of pain medicine encompasses pain linked to both psychological and/or organic factors.

Pain–pleasure principle Psychoanalytic concept which states that, in psychic functioning, a person tends to seek pleasure and avoid pain. *See also* pleasure principle.

panic Acute, intense attack of anxiety associated with personality disorganization; the anxiety is overwhelming and accompanied by feelings of impending doom.

panic attack Episode of acute intense anxiety occurring in panic disorder, schizophrenia, major depression, and somatization disorder.

panic disorder DSM classification of anxiety disorder characterized by attacks of acute intense anxiety, with or without agoraphobia.

panphobia Overwhelming fear of everything.

pantomime Gesticulation; psychodrama without the use of words.

paradoxical sleep REM SLEEP.

paralogia EVASION.

paralysisagitans PARKINSONISM.

paralytic ileus Intestinal obstruction resulting from inhibition of bowel motility, a rare anticholinergic side effect of certain psychoactive drugs. Also called adynamic ileus.

parameter Way of measuring or describing an object or evaluating a subject.

parametric tests of significance Statistical methods used to estimate population parameters, such as the mean and standard deviation; use of these methods requires assumptions about the population characteristics, for example, that the variable under study is normally distributed within the population from which the sample is drawn.

paramnesia Disturbance of memory in which reality and fantasy are confused. It is observed in dreams and in certain types of schizophrenia and organic mental disorders; includes phenomena such as déjà vu and déjà entendu, which may occur occasionally in normal persons.

paranoia Rare psychiatric syndrome marked by the gradual development of a highly elaborate and complex delusional system, generally involving persecutory or grandiose delusions, with few other signs of personality disorganization or thought disorder.

paranoiac Person with paranoia.

paranoid delusion Includes persecutory delusions and delusions of reference, control, and grandeur.

paranoid disorder *See* delusional disorder.

paranoid ideation Thinking dominated by suspicious, persecutory, or grandiose content of less than delusional proportions.

paranoid personality disorder Personality disorder characterized by rigidity, hypersensitivity, unwarranted suspicion, jealousy, envy, an exaggerated sense of self-importance, and a tendency to blame and ascribe evil motives to others.

paranoid schizophrenia Type of schizophrenia characterized by the presence of persecutory or grandiose delusions, often accompanied by hallucinations; no longer included in DSM-5.

paraphasia Abnormal speech in which one word is substituted for another, the irrelevant word generally resembling the required one in its morphology, meaning, or phonetic composition; the inappropriate word may be either a legitimate one used incorrectly, such as “clover” instead of “hand,” or a bizarre nonsense expression, such as “treen” instead of “train.” Paraphasic speech may be seen in organic aphasias and in mental disorders, such as schizophrenia. *See also* metonymy; word approximation.

paraphilia Psychosexual disorder with a sexual dysfunction characterized by persistent and recurrent sexual fantasies, often of an unusual nature, without which imagery erotic arousal or orgasm is not attained; fantasies generally involve themes of suffering, humiliation, sexual activity with nonconsenting partners, or preference for a nonhuman object for sexual arousal. In DSM, the paraphilias include fetishistic disorder, transvestic disorder, frotteuristic disorder, pedophilic disorder, exhibitionistic disorder, voyeuristic disorder, sexual masochism disorder, and sexual sadism disorder. Also called sexual deviations.

paraphrenia Chronic condition characterized by fantastic and absurd delusions without significant personality deterioration and without the primary disturbances of thought and affect that characterize schizophrenia.

parapraxis Faulty act, such as a slip of the misplacement of an article. Freud ascribed parapraxes to unconscious motives.

paraprofessional Trained person who works in a mental hospital and may have a degree in the arts or from some professional school other than those serving the mental health group but has not obtained a degree in one of the usual mental health professions.

parapsychology Branch of psychology concerned with extranormal events and behavioral phenomena not accounted for or explained by the tenets and laws of contemporary science; for example, clairvoyance, telepathy, psychokinesis. *See also* metapsychology.

parasomnias A group of dysfunctions associated with sleep, sleep stages, or partial arousals characterized by unusual behavior, experiences, or physiologic events during sleep; composed of the three subtypes: non-REM movement sleep arousal disorders (incomplete awakening accompanied by sleepwalking, sleep-related eating, sleep-related sexual behavior, or sleep terror disorder); REM sleep behavior disorder (vocal or motor behavior during sleep); and nightmare disorder.

parasympathetic nervous system *See* autonomic nervous system.

parataxic distortion Harry Stack Sullivan’s term for distortions in judgment or attitude in interpersonal relations based on patterns set by earlier experience; previously developed ways of coping with significant people in a person’s life are applied by the person in later interpersonal integrations.

parens patriae Power of the state to involuntarily hospitalize mentally ill persons in need of care; also called “police power,” in that it prevents mentally ill persons from doing harm to themselves or to others.

Parent–child problem DSM category that refers to problems between a parent and a child; for example, in a family in which the parents are divorced, with either the custodial or the noncustodial parent, and in the remarriage of a divorced or widowed parent.

paresis Weakness or partial paralysis of organic origin.

paresthesia Abnormal spontaneous tactile sensation, such as a burning, tingling, or pins-and-needles sensation.

peripartum depression Depressed mood, anxiety, insomnia, and weight changes that occur near parturition or within 1 month of giving birth. Previously called postpartum depression.

parkinsonism Extrapyramidal dyskinesia characterized by resting tremor, rigidity, bradykinesia, and postural abnormalities. Parkinsonism is of unknown etiology and occurs in middle or late life, typically with a gradual progression and a prolonged course. Also called paralysis agitans; Parkinson disease; idiopathic parkinsonism. A parkinsonian syndrome may develop during the course of therapy with antipsychotic phenothiazine or butyrophenone drugs; such drug-induced parkinsonism is reversible when the causative drug is withdrawn or its dosage reduced.

Parkinson disease PARKINSONISM.

partial hospitalization System of treating mental illness in which the patient is hospitalized on a part-time basis. *See also* day hospital; night hospital; weekend hospital.

partial insanity MONOMANIA.

passive–aggressive personality disorder Personality disorder in which the patient manifests aggressive behavior in passive ways, such as obstructionism, pouting, stubbornness, and intentional inefficiency.

passive-dependent personality *See* dependent personality disorder.

pastoral counseling Use of psychotherapeutic principles by a clergyman in helping parishoners with emotional problems.

patch amnesia LOCALIZED AMNESIA.

pathognomonic Denoting a sign or symptom specifically diagnostic of a particular disease entity.

pathologic gambling Compulsion to gamble, one of the disorders of impulse control. *See also* gambling disorder.

pathologic intoxication ALCOHOL IDIOSYNCRATIC INTOXICATION.

pathophobia Abnormal fear of disease or suffering. Also called nosophobia.

paucity of speech Limited use of speech, seen in various pathologic states.

pavlovian conditioning CLASSICAL CONDITIONING. *See* conditioning.

Pavlov, Ivan Petrovich (1849–1936) Russian neurophysiologist famous for his work in conditioning.

pavor nocturnus SLEEP TERROR DISORDER.

PCP Phencyclidine.

peccatophobia Abnormal fear of sinning.

pecking order Hierarchy or sequence of authority in an organization or social group.

pederasty Anal intercourse between a man and a boy.

pedophilia Paraphilia in which a child is the sexual object.

pedophobia Abnormal fear of children.

peer review Review of physician services by a panel of other physicians.

Peer Review Organization (PRO) Federal review organization for hospitals receiving Medicare funds. In order to promote compliance with federal guidelines for health and hospital care, the PRO conducts independent utilization reviews and quality-of-care studies, validates diagnostic-related group (DRG) assignments, and reviews hospital admissions and readmissions. *See also* Professional Standards Review Organization; utilization review committee.

pellagra Deficiency of niacin (nicotinamide) characterized by diarrhea, dermatitis, and mental disturbance.

penis envy Sigmund Freud's concept that the woman envies the man for his possession of a penis; sometimes used to refer to a woman's generalized envy of men. A controversial concept given its heteronormative and patriarchal assumptions.

perception Conscious awareness of elements in the environment by the mental processing of sensory stimuli; sometimes used in a broader sense to refer to the mental process by which all kinds of data, intellectual, emotional, as well as sensory, are meaningfully organized. *See also* apperception.

perfectionism Practice of demanding of others or of oneself a higher quality of performance than is required by the situation.

period prevalence Research term for the total number of cases of a disease known to have existed during a specified time period. *Compare* point prevalence.

perseveration **1.** Pathologic repetition of the same response to different stimuli, as in a repetition of the same verbal response to different questions. **2.** Persistent repetition of specific words or concepts in the process of speaking. Seen in organic mental disorders, schizophrenia, and other mental illness. *See also* verbigeration.

persistence Term used to reflect a personality trait characterized by ambition, overachievement, resistance to frustration, and fatigue. Persistence may be enhanced by psychostimulants.

persistent (chronic) motor or vocal tic disorder A type of motor disorder consisting of either motor or vocal tics, but not both, that develops before adulthood. Compare Tourette disorder.

persistent complex bereavement disorder Disorder characterized by chronic and persistent grief about the death of someone close to the patient, accompanied by feelings of bitterness, anger, or ambivalence toward the deceased and social withdrawal. Also known as complicated grief or complicated bereavement.

persona In Jungian psychology, the outside personality that a person presents to the world, as opposed to their inner self (anima).

personality Characteristic configuration of behavior response patterns that each person evolves as a reflection of their individual adjustment to life.

personality disorder Mental disorder characterized by inflexible, deeply ingrained, maladaptive patterns of adjustment to life that cause either subjective distress or significant impairment of adaptive functioning; manifestations are generally recognizable in adolescence or earlier. Personality disorders listed in DSM include paranoid, schizoid, schizotypal, histrionic, narcissistic, antisocial, borderline, avoidant, dependent, and obsessive-compulsive.

personality traits Descriptive features of an individual's personality that are within normal limits.

persuasion Supportive psychotherapy in which the patient is encouraged to follow advice; persuasion is directive and limited in its goals.

pervasive developmental disorder Class of developmental disorders characterized by severe distortions in social skills, language, and contact with reality (e.g., autism spectrum disorder). Many psychological functions are deficient, and a child with this type of disorder displays abnormalities that are not normal for any stage of development. This condition was previously known as childhood schizophrenia. In DSM-5, Asperger syndrome, Rett syndrome, and childhood disintegrative disorder are types of pervasive developmental disorders. *Compare* specific learning disorder.

pervasive developmental disorder not otherwise specified A diagnostic category used in DSM-5 to refer to pervasive developmental disorders that may have contradictory features and/or lack sufficient findings to meet full criteria for one of the other pervasive developmental disorders (autism, Asperger syndrome, Rett syndrome, childhood disintegrative disorder).

perversion Deviation from the expected norm. In psychiatry it commonly signifies sexual perversion. *See also* psychosexual disorder.

perverted logic EVASION.

PET Positron emission tomography.

petit mal epilepsy Epileptic disorder characterized by recurrent minor, nonconvulsive, momentary lapses of consciousness, commonly beginning in childhood and having a self-limited course. A typical attack involves a sudden transient loss of contact with the environment, during which the patient develops a dazed expression and pauses in what he is doing; this type of seizure has a classical electroencephalographic representation consisting of synchronous, symmetrical spike-and-wave discharges at 3 per second, without cortical localization. Also called absence epilepsy; minor epilepsy.

phallic overbearance Domination of another person by aggressive means; generally associated with masculinity in its negative aspects.

phallic phase Third stage in psychosexual development, following the anal stage and lasting from about age 2 or 3 to about age 6; resolution of the Oedipus complex is the dominant developmental conflict during this stage. During this period, sexual interest, curiosity, and pleasurable experiences are centered on the penis in boys and the clitoris in girls. Also called oedipal phase.

phantasy Alternative spelling of fantasy.

phantom limb False sensation that an extremity that has been lost is in fact present.

pharmacokinetics Way in which the body processes a drug; its principal divisions are drug absorption, distribution, metabolism, and excretion.

pharmacotherapy PSYCHOPHARMACOLOGY.

phase of life problem DSM category that refers to problems not severe enough to warrant a psychiatric diagnosis but that interfere with functioning, as arising from marriage, work, change in job, culture shock.

phencyclidine (PCP) Substance of abuse with effects similar to hallucinogens. Also called angel dust.

phenomenology Study of events or happenings in their own right, rather than from the point of view of inferred causes; associated with existential psychiatry and reflects the theory that behavior is determined by the way the person perceives reality, rather than by external reality in objective terms.

phenothiazine One of a class of psychotropic drugs, the prototype of which is chlorpromazine, which are effective in the treatment of psychoses, especially schizophrenia; they exert a favorable influence with regard to the fundamental symptoms of schizophrenia as well as the secondary or accessory symptoms. Their mechanism of action may be related to the blockade of dopamine receptors at critical sites in the central nervous system.

phenotype Outward, observable expression of a person's genetic constitution. *Compare* genotype.

phenylketonuria (PKU) Genetic metabolic disorder characterized by the inability to convert the amino acid phenylalanine to tyrosine. If untreated, severe mental retardation results from the accumulation of large concentrations of toxic phenylketones; when detected early in infancy, the condition is treatable by the elimination of phenylalanine from the diet.

phenylpyruvic oligophrenia Former name for phenylketonuria.

pheromone Chemical signal that a person releases into the external environment and that affects the behavior or physiologic states of other persons.

phobia Persistent, pathologic, unrealistic, intense fear of an objector situation; the phobic person may realize that the fear is irrational but is, nonetheless, unable to dispel it. Some examples are: acarophobia (small objects), acrophobia (high places), agoraphobia (open places, leaving familiar setting of home), aichmophobia (pointed objects), ailurophobia (cats), algophobia (pain), androphobia (men), astraphobia or astrapophobia (thunder and lightning), autophobia (being alone, or of self), bathophobia (depths), claustrophobia (closed or confined places), coprophobia (excreta), cynophobia (dogs), demophobia (crowds), erythrophobia (blushing), gynophobia (women), haphephobia (being touched), hematophobia (sight of blood), hydrophobia (water), hypnophobia (sleep), lalophobia or glossophobia (speaking), mysophobia (dirt, contamination), necrophobia (dead bodies), nyctophobia (darkness, night), panphobia (everything), pathophobia or nosophobia (disease, suffering), peccatophobia (sinning), pedophobia (children), phonophobia (speaking aloud), photophobia (strong light), sitophobia (eating), taphophobia (being buried alive), thanatophobia (death), toxophobia (being poisoned), xenophobia (strangers), zoophobia (animals).

phobic disorder Frequent disorder characterized by a specific fear, which may represent a temptation, a punishment for the forbidden impulse, or a combination of both.

phobic neurosis PHOBIC DISORDER.

phonologic disorder *See* developmental articulation disorder.

phonophobia Abnormal fear of speaking aloud.

photophobia Abnormal fear of strong light.

phrenology Study of the bony conformation of the skull in the now discredited belief that it related to mental faculties and traits.

phylogenetic Pertaining to the development of the species. *Compare* ontogenetic.

physical dependence *See* drug dependence.

physical disorder Disorder of body function whose manifestations are not primarily behavioral or psychological.

Piaget, Jean (1896–1980) Swiss psychologist noted for his research on cognitive development in children; he divided the development of intelligence into three major periods: Sensorimotor (birth to 2 years), concrete operations (2 to 12 years), and formal operations (12 years through adult life).

piblokto Culture-specific syndrome seen in Eskimos, usually women, in which the affected person screams, cries, runs naked through the snow, sometimes with suicidal or homicidal tendencies.

pica Craving and eating of nonfood substances, such as paint and clay.

Pick disease *See* frontotemporal dementia.

pickwickian syndrome Condition characterized by obesity, hypoventilation, and hypersomnia.

Pinel, Phillipe (1746–1826) French reformer in the field of mental illness known for his work in abolishing physical restraints on hospitalized mental patients.

piperazine Phenothiazine derivative with antipsychotic affects; for example, fluphenazine (Prolixin).

piperidine Phenothiazine derivative with antipsychotic effects; for example, thioridazine (Mellaril).

PKU Phenylketonuria.

placebo 1. Inactive substance or preparation that is given as though it were a pharmacologically active medicine; primarily used in controlled studies to determine the efficacy of medicinal preparations; any beneficial or deleterious effects of the placebo may be ascribed to psychological factors. **2.** Any type of treatment that has no specific effects on the particular illness involved.

placebo effect Phenomenon in which a person exhibits a clinically significant response to a pill containing a therapeutically inert substance or a treatment without specific effect on the person's condition. Placebo effects are not limited to subjective reports; physiologic functions may be objectively influenced.

plaques In psychiatry, pathologic masses of axons, neurons, and dendrites with a central amyloid core; associated with Alzheimer disease.

plasma level *See* blood level.

plasticity *See* neuronal plasticity.

plateau phase Stage in the human sexual response cycle, as described by Masters and Johnson, characterized by increased genital blood flow and sense of sexual excitement; it precedes orgasm.

play therapy Type of therapy used with children in which they reveal their problems on a fantasy level with dolls, clay, and other toys; the therapist intervenes opportunely with helpful explanations about the patient's responses and behavior in language geared to the child's comprehension. *See also* activity group therapy.

pleasure principle In psychoanalysis, the principle by which the id seeks immediate tension reduction by either direct or fantasied gratification; developmentally, the pleasure principle antedates the reality principle. Also called the pleasure-pain principle. *See also* reality principle.

pleonexia Psychiatric disorder in which the patient has an excessive desire to acquire wealth or objects.

pluralism In psychiatry, the notion that multiple factors affect behavior.

PMDD Premenstrual dysphoric disorder.

point prevalence Research term for the total number of cases of a disease known to have existed at a given point in time. *Compare* period prevalence.

police power PARENS PATRIAE.

polygenic Involving the combined action or effects of several genes.

polymorphous perverse sexuality Psychoanalytic conceptualization of sexuality in the human infant, for whom sexual gratification can be achieved by homosexual, heterosexual, or inanimate stimulation of several body zones; all humans, therefore, have the potential to develop sexual perversions in adulthood, depending on how the sexual drives are channeled developmentally.

polyphagia Excessive eating. *See also* bulimia nervosa.

polysomnography Battery of tests that include an electroencephalogram, electrocardiogram, and electromyogram; often given with tests for penile tumescence, blood oxygen saturation, body movement, body temperature, galvanic skin response, and gastric acid. The tests are of assistance in a variety of conditions, such as insomnia, nocturnal myoclonus, sleep apnea, enuresis, somnambulism, seizure disorders, impotence, vascular headache, gastroesophageal reflux, and depression.

polysubstance dependence Disorder characterized by dependence on at least three categories of psychoactive substances.

polyuria Excessive excretion of urine.

population Research term for the entire collection of a set of objects.

porphyria Metabolic disorder characterized by excessive urinary excretion of porphyrins (pigments found in hemoglobin and certain cellular enzyme systems) or their precursors; mental symptoms are prominent in the form known as acute intermittent porphyria. Abdominal pain, constipation, and neuromuscular disturbances complete the clinical picture. Attacks may be precipitated by barbiturates.

positive reinforcement *See* reinforcement.

positive symptoms In schizophrenia: hallucinations, delusions, and thought disorder.

positive transference *See* transference.

positron emission tomography (PET) Imaging technique using short-lived positron-emitting elements of oxygen, carbon, and nitrogen used to label organic compounds (e.g., glucose, amino acids, and neurotransmitters), which are then introduced into the body by injection or inhalation; energy emitted is recorded by scintillation detectors and computer analyzed to construct and display images of transverse sections of tissues, such as the brain. Quantitative data can be assessed in specific areas for such functions as blood flow, oxygen or glucose metabolism, and other complex physiologic phenomena; alterations in metabolism associated with disease and the actions of medications can also be monitored in this way.

postpartum depression *See* peripartum depression.

postpartum psychosis A descriptive term for depressive or psychotic symptoms in women who have recently given birth; a subtype of bipolar disorder in DSM-5. Also called puerperal psychosis. *See also* peripartum depression.

posttraumatic stress disorder (PTSD) Anxiety disorder that occurs after and as a result of a major, disturbing, out of the ordinary event in a person’s life, such as war or rape. The prominent symptoms of this disorder include nightmares, flashbacks, withdrawal, and hypervigilance; may have a delayed onset. *See also* acute stress disorder.

posturing Strange, fixed, and bizarre bodily positions held by a patient for an extended time. *See also* catatonia.

potency Male ability to perform the sexual act; specifically, the capacity to achieve and maintain an erection during coitus.

poverty of content of speech Speech that is quantitatively adequate but qualitatively inadequate because it imparts minimal, vague, or repetitious information.

poverty of speech Restriction in the amount of speech used; replies may be monosyllabic. *See also* laconic speech.

PPO Preferred Provider Organization.

practice effect Research term referring to the improvement of performance in a task as a result of repeated trials.

Prader–Willi syndrome A developmental disability as a result of a small deletion involving chromosome 15. Symptoms include compulsive eating, obesity, intellectual disability, hypogonadism, small stature, hypotonia, and small hands and feet. Children also exhibit oppositional and defiant behavior.

preconscious One of three divisions of the psyche in Freud’s topographic theory of the mind (the others being the conscious and unconscious), including all mental contents that are not in immediate awareness but that can be consciously recalled with effort.

predictive value Proportion of true-positive or true-negative tests; indicates what percentage of test outcomes are expected to coincide with assigned diagnoses.

predictor variable Test or score used to assess or predict future performance.

Preferred Provider Organization (PPO) Alternative delivery organization that employs a prospective payment system in which a corporation or insurance company forms an agreement with a particular group of community hospitals and doctors to supply health services to PPO members at a previously determined lower rate; patients enrolled in a PPO select their physicians and hospitals from among the list of participants.

prefrontal lobotomy *See* lobotomy.

pregenital stages In psychoanalysis, the first stages of psychosexual development (the oral and anal stages) that occur before the genitals have begun to exert the predominant influence on sexual behavior.

prejudice Preconceived adverse judgment or opinion formed without just grounds or sufficient knowledge; elements of irrational suspicion or hatred are often involved, as in racial prejudice.

premature ejaculation Ejaculation occurring before or immediately after vaginal intromission during coitus; broader definitions take into consideration the ability of a male to delay ejaculation during coitus for a sufficient length of time to satisfy a normally responsive female partner.

premenstrual dysphoric disorder (PMDD) Depressive mood disorder characterized by irritability, emotional lability, headache, and anxiety or depression that occurs about 1 week before menses and resolves after the menstrual cycle is complete. Also called late luteal phase dysphoric disorder.

premorbid Occurring before the onset of disease.

preoccupation of thought Centering of thought content on a particular idea, associated with a strong affective tone, such as paranoid trend or a suicidal or homicidal preoccupation.

presenile dementia DEMENTIA OF THE ALZHEIMER TYPE; PRESENILE ONSET.

presenium Period preceding old age, generally ages 55 to 65.

pressure of speech, pressured speech Increase in the amount of spontaneous speech; rapid, loud, accelerated speech, as occurs in mania, schizophrenia, and cognitive disorders.

prevalence Total number of cases of a disease existing at a given point in time (point prevalence) or during a specified time period (period prevalence). *See also* incidence.

prevalence rate Number of diagnosed cases of a disease or a condition at a particular point in time; prevalence rate per 1,000 is defined as the cases at a specific point in time, divided by the number of persons in the population at a specific point in time, multiplied by 1,000.

preventive psychiatry Branch of preventive medicine concerned with mental disorders. Encompassed within its scope are measures to prevent mental disorders (primary prevention); measures to limit the severity of illness, as through early case finding and treatment (secondary prevention); and measures to reduce disability after a disorder (tertiary prevention).

primal father Hypothetical head of a tribe depicted by Sigmund Freud in *Totem and Taboo* as slain by his sons, who subsequently devour him in a cannibalistic rite; later, he is promoted to a god. The son who murders him is the prototype of the tragic hero, and the memory of the crime is perpetuated in the conscience of the individual and of the culture.

primal scene In psychoanalysis, the real or fantasied observation by a child of sexual intercourse, particularly between their parents.

primary care physician Physician who specializes in family medicine, internal medicine, pediatrics, and obstetrics and gynecology; many believe psychiatry should be designated a primary care specialty.

primary diagnosis Major emotional disorder for which the patient is treated. Also called principal diagnosis.

primary gain Reduction of anxiety achieved by a defense mechanism; relief from tension or conflict through neurotic illness. *Compare* secondary gain.

primary narcissism *See* narcissism.

primary parkinsonism *See* parkinsonism.

primary prevention *See* preventive psychiatry.

primary process thinking In psychoanalysis, the mental activity directly related to the functions of the id and characteristic of unconscious mental processes; marked by primitive, prelogic thinking and by the tendency to seek immediate discharge and gratification of instinctual demands. Includes thinking that is dereistic, illogic, and magical; normally found in dreams, abnormally in psychosis. *Compare* secondary process thinking.

primary reinforcer *See* reinforcement.

primary retarded ejaculation Lifelong inability to ejaculate during coitus. *Compare* secondary ejaculation.

primitive idealization Defense mechanism in which external objects that are viewed as either “all good” or “all bad” are unrealistically endowed with great power; most commonly, the “all good” object is seen as omnipotent, or ideal, while the badness of the "all bad" object is greatly inflated.

Prince, Morton (1854–1929) American neurologist and psychiatrist known for his study of multiple personalities.

principal diagnosis PRIMARY DIAGNOSIS.

prion disease A class of disorders caused by a transmissible infectious protein. Includes Creutzfeldt–Jakob disease (CJD), Gerstmann–Straussler–Scheinker disorder, kuru, and fatal familial insomnia.

prison neurosis *See* chronophobia.

prison psychosis Psychotic reaction to incarceration or to the prospect of incarceration.

privacy 1. Limiting access of others to one's body or mind, including dreams, fantasies, thoughts, or beliefs. **2.** Freedom from intrusion by the state or third persons. **3.** Restricted areas of personal decision, usually in such matters as personal association, abortion, or bodily integrity.

privilege Legal term referring to the legal right of patients to prevent their physicians from testifying about information obtained in the course of treatment; the legal formulation of the ethical principle of confidentiality. In some states, information obtained in the course of the doctor–patient relationship is considered privileged communication.

privileged communication *See* privilege.

PRO Peer Review Organization.

probability Quantitative statement of the likelihood that an event will occur. A probability of 0 means that the event is certain not to occur; a probability of 1.0 means the event will occur with certainty.

problem-oriented record Method for organizing a patient's clinical chart, consisting of five parts: data base, problem list, treatment plan, progress notes, and discharge summary.

problem-solving Capacity of a person to master a new or novel task.

process schizophrenia Unofficial term for schizophrenia attributed more to endogenous factors than to specific environmental influences. *Compare* reactive schizophrenia.

prodrome Early symptom of a disease, often serving as a warning or premonitory sign of the approach of a morbid condition; for example, the aura preceding an epileptic seizure or classic migraine.

Professional Standards Review Organization (PSRO) Organization of physicians in a given region established to review the quality of health care services, particularly with regard to whether the inpatient services rendered are medically necessary. Such an organization has been required by federal law since 1972 in relation to hospital care under Medicare and Medicaid. *See also* peer review organization; utilization review committee.

profound mental retardation *See* mental retardation.

prognosis 1. Opinion as to the probable future course, extent, and outcome of an illness. **2.** Specific goals of therapy.

projection Unconscious defense mechanism in which a person attributes to another those generally unconscious ideas, thoughts, feelings, and impulses that are in themself undesirable or unacceptable as a form of protection from anxiety arising from an inner conflict; by externalizing whatever is unacceptable, the person deals with it as a situation apart from themselves. *See also* blind spot.

projective identification Defense mechanism in which unwanted aspects of the self are projected onto another person such that the individual projecting feels at one with the object of the projection. It allows one to distance and make oneself understood by exerting pressure on another individual to experience feelings similar to one’s own.

projective test Psychological test with loosely structured test material that requires the subject to reveal their own feelings, personality, or psychopathology; for example, Rorschach Test, Thematic Apperception Test.

prolactin Anterior pituitary hormone that has been studied as a potential index of dopamine activity and receptor sensitivity in studies of central nervous system function in psychiatric patients.

proprioception KINESTHETIC SENSE.

prosopagnosia Inability to recognize familiar faces that is not due to impaired visual acuity or level of consciousness.

prospective study Study based on observing events as they occur.

prototaxic Harry Stack Sullivan's term for primitive processes. *See also* primary process.

pseudoaggression Neurotic defense in which h patient denies basic masochistic feelings and displays them, instead, as false aggression.

pseudoauthenticity False or copied expression of thoughts and feelings.

pseudocollusion Sense of closeness, relationship, or cooperation that is not real but is based on transference.

pseudocyesis Rare condition in which a nonpregnant patient has the signs and symptoms of pregnancy, such as abdominal distention, breast enlargement, pigmentation, cessation of menses, and morning sickness.

pseudodementia 1. Dementia-like disorder that can be reversed by appropriate treatment and is not caused by organic brain disease. The current preferred term is Depressive Cognitive Disorder. **2.** Condition in which a patient shows an exaggerated indifference to their surroundings in the absence of a mental disorder; also occurs in depression and factitious disorders.

pseudologia fantastica Disorder characterized by uncontrollable lying in which a patient elaborates extensive fantasies that he freely communicates.

pseudomotor agitation Physical and mental overactivity that is usually nonproductive and is associated with a feeling of inner turmoil, as seen in agitated depression.

pseudoparkinsonism Syndrome that resembles Parkinson disease, resulting from long-term use of antipsychotic drugs. Also called drug-induced parkinsonism.

psilocin *See* psilocybin.

psilocybin 4-Phosphoryloxy-*N*′,*N*′-dimethyltryptamine, the active ingredient of various Mexican hallucinogenic mushrooms, which acts as a psychedelic drug producing changes in perception and mood and alterations of consciousness. A variation is called psilocin.

PSRO Professional Standards Review Organization.

psychalgia PSYCHOGENIC PAIN DISORDER.

psychasthenia Obsolete term used by Pierre Janet to describe a syndrome characterized by fears and phobias.

psyche Mind, as divided into conscious, preconscious, and unconscious in Sigmund Freud’s topographic model.

psychedelic Denoting drugs that may induce hallucinations and psychotic states. LSD is the prototype; other psychedelic drugs include mescaline, psilocybin, and marijuana.

psychiatric emergency Disturbance in thoughts, feelings, or actions for which immediate treatment is necessary; for example, suicide intent.

psychiatric illness *See* mental disorder.

psychiatric nurse Clinical nurse specialist who functions as part of the mental health team, usually in an institutional setting; often carries out individual, family, and group psychotherapy.

psychiatric social worker Skilled professional, trained in social work, who works with psychiatrists, usually in an institutional setting; evaluates family, environmental, and social factors in the patient’s illness; may work in intake and reception with new patients; may follow-up and counsel after discharge, and also carry out individual, family, and group psychotherapy and participate in community organizations.

psychiatrist Medical doctor whose specialty is the study and treatment of mental disorders. After receiving the M.D. degree, a physician spends 4 years as a resident in a hospital setting. When they pass the written and oral examinations of the American Board of Psychiatry and Neurology, they become a diplomat in psychiatry and is said to be Board certified.

psychiatry Branch of medicine concerned with the prevention, cause, and treatment of mental disorders.

psychic determinism Freudian concept that all mental phenomena have specific antecedent causes, often operating on an unconscious level.

psychic energizer Rarely used term for any antidepressant or stimulant drug.

psychic helplessness *See* helplessness.

psychic trauma Psychologically upsetting event that may have lasting consequences on a person's thinking, feeling, or behavior.

psychoactive drug PSYCHOTHERAPEUTIC DRUG.

psychoactive substance abuse *See* substance use disorder.

psychoactive substance dependence DSM term for drug dependence.

psychoactive substance-induced mental disorder Mental disorder caused by the direct effects on the nervous system of a drug or other substance. In DSM, the substances listed are alcohol; sedatives, hypnotics, or anxiolytics; inhalants; opioids; cocaine; amphetamine or similarly acting sympathomimetics; phencyclidine (PCP) or similarly acting arylcyclohexylamine; hallucinogens; cannabis; nicotine; and caffeine.

psychoanalysis Theory of human mental phenomena and behavior, a method of psychic investigation and research, and a form of psychotherapy originally formulated by Sigmund Freud. As a technique for exploring the mental processes, psychoanalysts includes the use of free association and the analysis and interpretation of dreams, resistances, and transferences. As a form of psychotherapy, it uses the investigative technique, guided by Freud's libido and instinct theories and by ego psychology, to gain insight into a person's unconscious motivations, conflicts, and symbols and thus to effect a change in maladaptive behavior. Also called analysis in depth.

psychoanalyst Psychotherapist, usually a psychiatrist, who has had training in psychoanalysis and who uses its techniques in a treatment setting.

psychoanalytic group psychotherapy Method of group psychotherapy pioneered by Alexander Wolf and based on the operational principles of individual psychoanalytic therapy; analysis and interpretation of a patient's transferences, resistances, and defenses are modified to take place in a group setting.

psychobiology Adolf Meyer's term referring to the study of the human being as an integrated unit, incorporating psychological, social, and biologic functions.

psychodrama Psychotherapy method originated by Jacob Moreno in which personality makeup, interpersonal relationships, conflicts, and emotional problems are expressed and explored through dramatization. Therapeutic dramatization of emotional problems includes: (1) protagonist or patient, the person who presents and acts out their emotional problems with the help of (2) auxiliary egos, persons trained to act and dramatize the different aspects of the patient that are called for in a particular scene in order to help the patient express feelings, and (3) the director, leader, or therapist, who guides those involved in the drama for a fruitful and therapeutic session.

psychodynamics Science of mental forces and motivations that influence human behavior and mental activity; the role of unconscious motivation in the causation of human behavior is

psychoendocrinology Study of the role of hormones on thinking, behavior, and emotional disorder.

Psychose passionelle *See* erotomania.

psychogenic Produced or caused by mental factors, rather than organic factors; usually refers to a symptom or an illness.

psychogenic amnesia Dissociative disorder characterized by the sudden inability to recall information stored in memory in the absence of an underlying organic mental disorder; in DSM listed as dissociative amnesia.

psychogenic fugue Dissociative disorder characterized by periods of total amnesia in which one travels and assumes a new identity. In DSM listed as Dissociative amnesia, with dissociative fugue. *See also* fugue; psychogenic amnesia.

psychogenic pain disorder Condition in which there is a preoccupation with pain for at least 6 months in the absence of organic pathology to account for the pain. Also called psychalgia; somatoform pain disorder in previous versions of DSM.

psychohistory Approach to history that examines events within a psychological framework; attempts to connect individual and collective ideas and emotions with wider historical currents.

psychoimmunology Study of the relationship between the immune system and emotions.

psychokinesis Parapsychological phenomenon in which directed thought processes influence physical events.

psycholinguistics Exploration of the psychological factors involved in the development and use of language.

psychological defense system *See* defense mechanism.

psychological factor affecting physical condition Disorder characterized by physical symptoms partially caused or affected by psychological factors; physiologic and organic changes stem in part from a sustained emotional disturbance; for example, atopic dermatitis, migraine, ulcer. Also called psychophysiologic disorder; psychosomatic disorder.

psychologist Person trained in psychology, usually with a graduate degree (M.A., Ph.D.).

psychology Study and profession concerned with mental processes and behavior.

psychometry Science of testing and measuring mental and psychological functioning.

psychomotor Relating to combined physical and mental activity.

psychomotor agitation Physical and mental overactivity associated with a feeling of inner turmoil, as seen in agitated depression.

psychomotor epilepsy COMPLEX PARTIAL SEIZURE.

psychomotor retardation Slowing of mental and physical activity, common in depression.

psychomotor stimulant Drug that arouses the patient through its central excitatory properties; for example, amphetamine, methylphenidate (Ritalin). Also called analeptic.

psychoneuroimmunology Study of psychological factors as causes of diseases or their effect on the course of disease, particularly as it relates to the immune system. *See also* psychoimmunology.

psychoneurosis NEUROSIS.

psycho-oncology Study of the role of emotions and their relation to cancer in terms of prevention, diagnosis, and treatment.

psychopathic personality Former term for the personality characteristic of antisocial personality disorder.

psychopathologic Denoting abnormal, maladaptive behavioral or mental activity.

psychopathology **1.** Study of mental disorders. **2.** Abnormal, maladaptive behavioral or mental activity.

psychopharmacology Study of the mental and behavioral effects of drugs. *See also* pharmacokinetics.

psychophysiologic disorder PSYCHOLOGICAL FACTOR AFFECTING PHYSICAL CONDITION.

psychopolitics **1.** Psychological dimensions of political behavior, such as the reciprocal influence on persons of political environments and their societies. **2.** Specific tactics of politicians that are intended to yield benefits to them through the use of psychological strategies.

psychosexual development **1.** Broadly, the maturation and development of the sexuality of a person throughout the life cycle. **2.** More specifically, in psychoanalysis, series of dynamic and crucial developmental stages that each person passes through and that influence basic personality characteristics in later life; the stages or phases are oral, anal, phallic, latency, and genital.

psychosexual disorder Disorder of sexual functioning, desire, or performance that is caused, wholly or partly, by psychological factors.

psychosis Mental disorder in which a person’s thoughts, affective response, ability to recognize reality, and ability to communicate and relate to others are sufficiently impaired to grossly interfere with their capacity to deal with reality; the classical characteristics of psychosis are impaired reality testing, hallucinations, delusions, and illusions.

psychosocial deprivation Deprivation of social and intellectual stimulation; believed to be a causative factor in mental retardation and emotional disorders in children.

psychosocial development Developmental progress of a person with regard to social relations and social reality, as primarily described by Erik Erikson. Specific developmental tasks characterize successive chronologic periods from infancy through maturity: trust versus mistrust (infancy), autonomy versus doubt (toddler), initiative versus guilt (preschool), industry versus inferiority (school age), identity versus identity diffusion (adolescence), intimacy versus isolation (young adulthood), generativity versus self-absorption (adulthood), and integrity versus despair (mature age).

psychosocial dwarfism Syndrome usually first manifested in children 2 to 3 years of age, who typically are unusually short, have frequent growth hormone abnormalities, and have severe behavioral disturbances; all of these symptoms are the result of an inimical caretaker–child relationship and resolve without any medical or psychiatric treatment upon being placed in a more favorable domicile. Also called psychosocially determined short stature.

psychosocially determined short stature PSYCHOSOCIAL DWARFISM.

psychosocial stressor STRESS.

psychosomatic disorder PSYCHOLOGICAL FACTOR AFFECTING PHYSICAL CONDITION.

psychosurgery Neurosurgical intervention to treat a mental disorder that involves destruction or ablation of brain tissue.

psychotherapeutic drug Drug used to alter abnormal thinking, feelings, or behavior; traditionally divided into classes of antipsychotic, antidepressant, antimanic, and antianxiety (anxiolytic) drugs. Also called psychoactive drug; psychotropic drug.

psychotherapist Person trained to treat mental, emotional, and behavioral disorders. For types of therapists, see the specific term.

psychotherapy Treatment for mental illness and behavioral disturbances in which a trained person establishes a professional contract with the patient and through definite therapeutic communication, both verbal and nonverbal, attempts to alleviate the emotional disturbance, reverse or change maladaptive patterns of behavior, and encourage personality growth and development; distinguished from other forms of psychiatric treatment such as the somatic therapies, for example, drug and convulsive therapies.

psychotic 1. Person suffering from a psychosis. **2.** Denoting or characteristic of a psychosis.

psychotic depressive reaction Psychosis distinguished by depressed mood precipitated by some event, usually in a person who had not previously demonstrated severe depression. Also called major depressive disorder, single episode, with mood-congruent psychotic features, in DSM.

psychotomimetic drug Drug that produces psychic and behavioral changes that resemble psychosis; unlike other drugs that can produce organic psychosis as a toxic reaction (e.g., hallucinogenic and psychedelic drugs such as LSD, marijuana, mescaline), a psychotomimetic drug does not produce gross confusion, delirium, or overt memory impairment.

psychotropic drug PSYCHOTHERAPEUTIC DRUG.

PTSD Posttraumatic stress disorder.

puerperal psychosis *See* postpartum psychosis.

punishment Presentation of an aversive stimulus contingent on an operant response, resulting in a decreased probability of that response recurring. *See also* reinforcement.

Purging disorder Eating disorder characterized by recurrent episodes of self-induced vomiting, laxative and diuretic abuse, and/or the use of enemas after consuming small amounts of food; patients are normal weight and have distorted body image. In DSM-5, included in “Other Specified Feeding or Eating Disorders.”

pyromania Compulsion to set fires.

Q

q-sort Technique of personality assessment based on the degree to which a standard set of descriptive statements are believed to describe a particular person.

quadrangular therapy Marital therapy that involves four people, the married pair and each spouse's therapist.

quality assurance Measurement and evaluation of medical and other related services provided with regard to their effectiveness, efficiency, appropriateness, and acceptability; such evaluation will trigger corrective or remedial action for services not meeting the desired standards.

quantitative variable Experimental factor that can be measured.

R

random Statistical term that means occurring by chance or without attention to selection or planning.

random sample In a given population, a group of subjects selected in such a manner that each member of the population has an equal probability of being selected for the sample.

random variable Variable for which the variation is determined by chance.

range Statistical measure of the variability of a set of values, defined as the difference between the largest value and the smallest value.

Rank, Otto (1884–1939) Austrian psychoanalyst, long-time secretary and recorder of the minutes of the Vienna Psychoanalytic Society, and one of Sigmund Freud's earliest followers, who split with him on the significance of the birth trauma, which Rank used as a basis of brief psychotherapy. Rank's works include *The Trauma of Birth* and *The Myth of the Birth of the Hero.*

rape Sexual intercourse without the person's consent and against their will, chiefly by force; usually heterosexual but may be homosexual.

rapid eye movement *See* REM latency; REM sleep.

rapport Conscious feeling of harmonious accord, sympathy, and mutual responsiveness between two or more persons; contributes to an effective therapeutic process in both group and individual settings.

rapture-of-the-deep syndrome Psychosis, seen in scuba and deep-sea divers, characterized by disorientation and loss of judgment and skill, and associated with excessive blood nitrogen levels; sensory deprivation may also contribute to this acute, self-limited mental disorder. Also called nitrogen narcosis.

rationalization Ernest Jones's term for an unconscious defense mechanism in which irrational or unacceptable behavior, motives, or feelings are logically justified or made consciously tolerable by plausible means.

Ray, Isaac (1807–1881) One of the original 13 founders of the American Psychiatric Association, known for his *Treatise on Medical Jurisprudence of Insanity* (1837).

rCBF Regional cerebral blood flow.

reaction formation Unconscious defense mechanism in which a person develops a socialized attitude or interest that is the direct antithesis of some infantile wish or impulse that is harbored either consciously or unconsciously. One of the earliest and most unstable defense mechanisms, closely related to repression; both are defenses against impulses or urges that are unacceptable to the ego.

reactive attachment disorder of infancy or early childhood Severe impairment of the ability to relate, beginning before age 5; presumed to develop from grossly pathologic care-giving.

reactive depression Mental disorder characterized by depression developing in reaction to some precipitating life event or circumstance.

reactive disorder Mental disorder judged to be a reaction to one or more life events or circumstances, without which the disorder would not have occurred. *See also* brief reactive psychosis; reactive depression; reactive schizophrenia.

reactive schizophrenia Unofficial term for schizophrenia attributed primarily to predisposing or precipitating environmental factors. *Compare* process schizophrenia.

reading disorder Reading below the expected level for a child's age, education, and intelligence.

reality Totality of objective things and factual events; includes everything that is perceived by a person's special senses and is validated by other people.

reality principle In psychoanalysis, ego processes that modify the demands of the pleasure principle to meet the inescapable demands and requirements of the outside world; net result is behavior that gratifies id impulses within realistic limits. *See also* pleasure principle.

reality testing Fundamental ego function that consists of tentative actions that test and objectively evaluate the nature and the limits of the environment; includes the ability to differentiate between the external world and the internal world and to accurately judge the relation between the self and the environment.

recall Process of bringing stored memories into consciousness. *See* memory.

receiver operating characteristic (ROC) Special statistical method, recently used in psychiatry, which gives a balance between sensitivity and specificity for a diagnostic test, for example, Dexamethasone Suppression Test. The method was derived from signal detection theory as a way to determine how well an electronic receiver was able to distinguish signal from noise.

recent memory Recall of events over the past few days.

recent past memory Recall of events over the past few months.

receptive aphasia Organic loss of ability to comprehend the meaning of words; fluid and spontaneous, but incoherent and nonsensical, speech. *See also* fluent aphasia and sensory aphasia.

receptive dysphasia Difficulty in comprehending oral language; the impairment involves both comprehension and production of language.

receptors Proteins in the neuronal membrane that are exposed in part to the extracellular fluid and specifically recognize neuromessengers; stimulation of receptors generally causes a biologic response to occur.

reciprocal inhibition and desensitization Behavior therapy in which a person is conditioned to associate comfortable, supportive surroundings with anxiety-producing stimuli, thus decreasing the anxiety associated with those stimuli.

recovered memory syndrome Condition that occurs when a repressed memory of a traumatic experience is brought back to consciousness, usually during psychotherapy or under hypnosis. A variant of this condition, false memory syndrome, occurs if the recalled memory did not actually happen, but the person believes the memory to be true.

recurrent brief depressive disorder Multiple, brief (less than 2 weeks duration) episodes of depressive symptoms.

recurrent depressive episode Depression that lasts between 2 and 13 days and occurs at least once a month.

recurrent isolated sleep paralysis Sleep disorder characterized by the inability to make voluntary movements while partially conscious at sleep onset or awakening. *See also* parasomnias.

refractory period Temporary state of psychophysiologic resistance to sexual stimulation immediately after orgasm in the human sexual response cycle, as described by Masters and Johnson.

reframing Technique used in family therapy whereby a problem or conflict is viewed by the therapist in an entirely different and sometimes startling perspective, thus enabling the family members to achieve objectivity.

regional cerebral blood flow (rCBF) Test using xenon-133, a low-energy gamma ray-emitting isotope, to determine blood flow in the brain; useful for studying diseases that have a decrease in the amount of brain tissue, for example, dementia, ischemia, atrophy.

registration *See* memory.

regression Unconscious defense mechanism in which a person undergoes a partial or total return to earlier patterns of adaptation;

observed in many psychiatric conditions, particularly schizophrenia.

regressive-reconstructive approach Psychotherapeutic procedure in which regression is made an integral element of the treatment process; an original traumatic situation is reproduced to gain new insight and to effect personality change and emotional maturation.

rehabilitation All methods and techniques used in an attempt to achieve maximal function and optimal adjustment in a given patient; the physical, mental, social, and vocational preparation of a patient for the fullest possible life compatible with their abilities and disabilities. Inasmuch as the process aims also to prevent relapses or recurrences of the patient's condition, it is sometimes called tertiary prevention.

Reich, Wilhelm (1897–1957) Austrian psychoanalyst who emigrated to the United States in 1939. He coined the term "orgone," a kind of universally pervasive physical and biologic energy that was typified by the pent-up libidinal energy that sought free expression and release through orgasm.

Reik, Theodor (1888–1969) Psychoanalyst and early follower of Sigmund Freud, who considered Reik one of his most brilliant pupils; Freud's book, *The Question of Lay Analysis*, was written to defend Reik's ability to practice psychoanalysis without medical training. Reik made many valuable contributions to psychoanalysis on the subjects of religion, masochism, and therapeutic technique.

reinforcement *Positive reinforcement* refers to the process by which certain consequences of behavior increase the probability that the behavior will occur again. Positive reinforcers tend to be viewed as pleasant. *Negative reinforcement* describes the process by which behavior that leads to the removal of an unpleasant event strengthens that behavior. Negative reinforcers tend to be viewed as unpleasant. Stimuli that possess inherent reinforcing characteristics are called *primary reinforcers*, for example, food. Stimuli that acquire reinforcing characteristics by being paired with primary reinforcers are called *conditioned reinforcers*, for example, money. *See also* conditioning.

reinforcer Reinforcing stimulus. *See* reinforcement.

relatedness Sense of sympathy and empathy with regard to others; sense of oneness with others. *Compare* isolation; alienation.

relational problems Impaired functioning of a relational unit. In DSM, parent–child relational problem, partner relational problem, and sibling relational problem.

relative frequency *See* frequency.

relative risk Epidemiologic measure of the risk of developing a disorder in a specified subset of the population as compared with the total population or a different subset of it. Comparison is usually made between groups exposed and groups not exposed to a particular hereditary or environmental factor to gain information about the role of that factor in the disorder in question. Relative risk is then expressed as a ratio of the frequency of the disorder in exposed persons to its frequency in those not exposed.

Relative Value Scale Payment scale for medical services developed at Harvard University that reimburses physicians in terms of time spent with patients, office overhead expenses, years of training, and cognitive skills; under the program, surgeons would receive less money and internists would receive more money from third-party (insurance) carriers than under previous payment scale programs. Also called Resource-Based Relative Value Scale (RBRVS).

reliability Research term meaning replicability, that is, the degree to which a test result or experimental finding can be confirmed by repetition. *See also* validity.

religious delusion Delusion involving the Deity or theologic themes.

reminiscence Recall or recollection of past experiences, with alternations of remembering an apparently lost memory and forgetting one that was just there.

remission Significant improvement or recovery which may or may not be permanent. *Compare* intermission.

remote memory LONG-TERM MEMORY.

remotivation Group treatment technique used patients in mental hospitals to encourage them and vocational skills.

REM (rapid eye movement) latency First REM episode of the night, about 90 minutes after sleep onset; shortening of REM latency frequency occurs with disorders such as depression and narcolepsy.

REM (rapid eye movement) sleep Stage of sleep during which dreaming occurs and the sleeper exhibits coordinated rapid eye movement (REM). The electroencephalogram demonstrates a desynchronized pattern of cerebral activity. It accounts for one-fourth to one-fifth of total sleep time. Also called paradoxical sleep.

REM sleep behavior disorder Sleep disorder characterized by the absence of sleep paralysis (atonia) during the REM stage of sleep; consequently, patients will act out their dreams, putting themselves and/or their partners at risk of physical injury. *See also* parasomnias.

repetition compulsion Psychoanalytic concept of an impulse to repeat earlier experiences and situations.

repetitive pattern Continual attitude or mode of behavior characteristic of a person and performed mechanically or unconsciously.

repression Freud’s term for an unconscious defense mechanism in which unacceptable mental contents are banished or kept out of consciousness; important in both normal psychological development and in neurotic and psychotic symptom formation. Freud recognized two kinds of repression: (1) repression proper, in which the repressed material was once in the conscious domain; (2) primal repression, in which the repressed material was never in the conscious realm. *Compare* suppression.

repressive-inspirational group psychotherapy Group therapy in which discussion is intended to bolster the patients’ morale and help them avoid undesired feelings; used primarily with large groups of seriously regressed patients in institutional settings.

reserpine Alkaloid extracted from the *Rauwolfia serpentina* plant, formerly used as an antipsychotic agent because of its sedative effects but currently used primarily as an antihypertensive agent, although this use has declined partly because of its potential to produce a considerable incidence of nightmares and suicidal depression; that untoward effect may be related to reserpine's depletion of catecholamines at critical sites in the central nervous system.

resident **1.** In psychiatry, a patient living in a treatment center for a prescribed period of time. **2.** In medicine, a physician in postgraduate training after graduation from medical school; 4 years of residency training are necessary to become a psychiatrist.

residential treatment facility Center where patients live and receive treatment appropriate for their particular needs; a children's residential treatment facility ideally furnishes both educational and therapeutic experiences for the emotionally disturbed child. Also called a halfway house.

residual Denoting a condition in which a person's acute mental symptoms have subsided.

residual schizophrenia Schizophrenia in which the patient no longer is psychotic but does have some remaining signs of the illness.

resistance Conscious or unconscious opposition to the uncovering of unconscious material; linked to underlying psychological defense mechanisms against impulses from the id that are threatening to the ego.

resistant attachment As described by Mary Ainsworth, a child's simultaneous yet antithetical wish for and avoidance of their mother's approach. *Compare* avoidant attachment; secure attachment.

resolution phase Final stage in the human sexual response cycle, as described by Masters and Johnson, during which the sexual system returns to its normal, nonexcited state following orgasm.

respondent conditioning CLASSICAL CONDITIONING. *See* conditioning.

restless legs syndrome (RLS) A syndrome characterized by an uncomfortable, subjective sensation of the limbs, usually the legs. Tends to be worse at night.

restricted affect Reduction in intensity of feeling tone that is less severe than in blunted affect but clearly reduced. *See* constricted affect.

retardation Reduction or slowing down of mental and physical activity, as often observed in depression. *See also* mental retardation.

retarded ejaculation INHIBITED MALE ORGASM.

retention *See* memory.

retroflexion Psychoanalytic term used notably by Sandor Rado to describe the turning of rage onto oneself.

retrograde amnesia Loss of memory for events preceding the onset of the amnesia. *Compare* anterograde amnesia.

retrospective falsification Memory becomes unintentionally (unconsciously) distorted by being filtered through a person's present emotional, cognitive, and experiential state.

retrospective study Investigation based on past data or past events, such as a case control study or a case history study.

Rett disorder A progressive condition characterized by deceleration in head growth, loss of purposeful hand movements, development of stereotyped hand movements (e.g., hand wringing), psychomotor retardation, loss of social engagement, loss of acquired speech, and ataxia between the ages of 6 months to 2 years.

reversal In psychoanalysis, an impulse's change from passive to active or from active to passive, as when a destructive impulse changes from masochism to sadism or vice versa.

right to refuse treatment Legal doctrine that holds that a person cannot be forced to have treatment against their will unless it is a life-and-death emergency.

right to treatment Legal principle that a facility that has assumed the responsibility of providing treatment is legally obligated to provide adequate treatment for a patient.

rigidity In psychiatry, a person's resistance to change, a personality trait.

risk factor Factor that accounts for the increased risk of a disorder. *See also* attributable risk; relative risk.

ritual **1.** Formalized activity practiced by a person to reduce anxiety, as in obsessive-compulsive disorder. **2.** Ceremonial activity of cultural origin.

ritualistic behavior Automatic behavior of cultural or psychogenic origin.

RLS Restless legs syndrome.

ROC Receiver operating characteristics.

Rogers, Carl R. (1902–1987) American psychologist who developed person-centered or client-centered theory of personality and psychotherapy, utilizing the concepts of self-actualization and self-direction; emphasized subjective experiential life of people and considered the self as a primary entity in both adjustment and maladjustment in people.

role Pattern or type of behavior developed under the influence of significant people in the person's environment. When the behavior pattern conforms to the expectations and demands of other people, it is said to be a complementary role; if it does not, it is known as a noncomplementary role. *See also* alternating role.

role playing Psychodrama technique in which persons are trained to function more effectively in their real-life roles. In the therapeutic setting of psychodrama, protagonists are free to try and to fail in their roles, for they are given the opportunity to try again until they succeed; new approaches to feared situations can thus be learned and applied outside the therapeutic setting.

Romberg sign Markedly increased swaying of the body or falling when a person stands erect with heels together and closes the eyes; typically indicates dysfunction of the dorsal column proprioceptive system, but it may also be seen in hysteria.

Rorschach Test Projective test, developed by Hermann Rorschach, in which the person is asked for associations in response to a series of inkblot pictures.

rumination Constant preoccupation with and thinking about a single idea or theme, as in obsessive-compulsive disorder.

rumination disorder of infancy Rare eating disorder of early childhood characterized by recurrent regurgitation of food in the absence of nausea, retching, or associated gastrointestinal disorder; regurgitated food is either ejected from the mouth or rechewed and reswallowed. Also called merycism.

Rush, Benjamin (1745–1813) Father of American psychiatry whose book, *Medical Inquiries and Observations upon the Diseases of the Mind* (1872), was the only American textbook on psychiatry until the end of the 19th century.

S

SAD Seasonal affective disorder.

sadism Paraphilia in which sexual gratification is achieved by inflicting pain or humiliation on a partner; first described by the French writer the Marquis de Sade (1740–1814). *See also* masochism; sadomasochistic relationship.

sadistic personality disorder Disorder in which a person shows a pervasive pattern of cruel, demeaning, and aggressive behavior whereby physical cruelty or violence is used to inflict pain on others and not to achieve some goal; such a person is fascinated by violence, weapons, injury, or torture. Not a DSM diagnosis.

sadomasochistic relationship Relationship in which the enjoyment of suffering by one person (masochism) and the enjoyment of inflicting pain by the other person (sadism) are important and complementary attractions in their ongoing interaction.

SAMHSA Substance Abuse and Mental Health Services Administration. A branch of the U.S. Department of Health and Human Services charged with improving the quality and availability of treatments and services for substance abuse and mental illnesses.

sample Research term for a subset of observations selected from a population.

sanatorium Institution where patients are treated for chronic physical or mental diseases or where they are attended to during a period of recuperation.

satyriasis Morbid, insatiable sexual needs or desires in a male; may be caused by organic or psychiatric factors. *Compare* nymphomania.

schedule of reinforcement Scheme of intermittent reinforcement in which a response is not reinforced every time it occurs. Four broad classes of intermittent schedules are based on either the number of responses required before an opportunity for reinforcement is presented or the passage of a certain amount of time before a given performance is reinforced (I) *fixed-ratio schedule,* in which reinforcement is delivered after a specified number of responses have occurred, for example, after every tenth response; (2) *variable-ratio schedule,* in which reinforcement is delivered on a schedule randomly changing around a given ratio, for example, after the third, sixth, then second response, and so on; (3) *fixed-interval schedule,* in which reinforcement is delivered contingent on the first required response after the elapse of a specified interval of time, for example, after every 10 minutes; and (4) *variable-interval schedule,* in which reinforcement is delivered after variable intervals of time have elapsed, for example, after 3, 6, then 2 hours, and so on. *See also* reinforcement.

schizoaffective disorder Psychotic disorder with signs and symptoms compatible with both a mood disorder and schizophrenia. In contrast to mood disorder with psychotic features, schizoaffective disorder is characterized by periods of at least 2 weeks without prominent mood symptoms.

schizoid fantasy Through fantasy, indulgence by a person in autistic retreat to resolve conflict and to obtain gratification; interpersonal intimacy is avoided and eccentricity serves to repel others. The individual does not fully believe in or insist on acting out the fantasies.

schizoid personality disorder Diagnostic category in DSM for persons with defects in the capacity to form social relationships but without other striking communicative or behavioral eccentricities. Also called introverted personality disorder.

schizophrenia Term introduced by Eugen Bleuler for a chronic, psychotic mental disorder of unknown etiology characterized by disturbances in thinking, mood, and behavior. The thinking disturbance is manifested by a distortion of reality, sometimes with delusions and hallucinations, accompanied by a fragmentation of associations that results in characteristic disturbances of speech; the mood disturbance includes ambivalence and inappropriate or constricted affective responses; the behavior disturbance may be manifested by apathetic withdrawal or bizarre activity. Subtypes of schizophrenia (disorganized, catatonic, paranoid, undifferentiated, and residual) are no longer formally in the DSM nomenclature, however they remain as useful descriptors. Formerly called dementia praecox.

schizophreniform disorder In DSM-5 a disorder similar to schizophrenia except that it lasts between 1 and 6 months.

schizotypal personality disorder Common term for mood disorder subtype characterized by depression, psychomotor slowing, hypersomnia, and hyperphagia that develops in autumn or winter and improves in spring and summer; not an official DSM-5 diagnosis.

schneiderian first rank symptoms Symptoms that Kurt Schneider, a German psychiatrist (1887–1967), considered to be of great pragmatic value in making a diagnosis of schizophrenia, including audible thoughts, arguing voices, thought withdrawal and insertion, delusions, and flat affect.

school phobia Young child's sudden fear of and refusal to attend school, usually considered a manifestation of separation anxiety but also may be a symptom of depression.

Schreber case One of Sigmund Freud's cases. Daniel Paul Schreber's published autobiographical account, entitled *Memoirs of a Neurotic* (1903), was analyzed by Freud (1911) and allowed him to decipher the fundamental meaning of paranoid processes and ideas, especially the relationship between repressed homosexuality and projective defenses.

scopophilia Sexual deviation in which a person achieves sexual gratification by viewing sexual acts or the genitals of others. *Compare* voyeurism.

scotoma **1.** In psychiatry, a figurative blind spot in a person's psychological awareness. *See also* blind spot. **2.** In neurology, a localized visual field defect.

screening Initial patient evaluation that includes medical and psychiatric history, mental status evaluation, and diagnostic formulation to determine the patient's suitability for a particular treatment modality.

screen memory Memory that serves as a cover-up for an associated painful memory.

SCT Sentence Completion Test.

SD Standard deviation.

SDA Serotonin-dopamine antagonist.

SDS Zung Self-Rating Depression Scale.

seasonal affective disorder (SAD) Mood disorder subtype characterized by depression, psychomotor slowing, hypersomnia, and hyperphagia that develops in autumn or winter and improves in spring and summer. **seasonal pattern**. *See* seasonal affective disorder.

secondary gain Obvious advantage that a person gains from an illness, such as gifts, attention, and release from responsibility. *Compare* primary gain.

secondary narcissism *See* narcissism.

secondary prevention *See* preventive psychiatry.

secondary process thinking In psychoanalysis, the form of thinking that is logical, organized, reality-oriented, and influenced by the demands of the environment; characterizes the mental activity of the ego. *Compare* primary process.

secondary retarded ejaculation After previous normal sexual functioning, the inability to ejaculate during coitus. *Compare* retarded ejaculation.

secondary sexual characteristics Physical characteristics, other than the external sexual organs, that distinguish male from female.

secure attachment As described by Mary Ainsworth, the positive, happy responses of a child following the mother's appearance.

sedative Drug that produces a calming or relaxing effect through central nervous system depression; for example, barbiturates, chloral hydrate, paraldehyde, and benzodiazepines. *Compare* hypnotic.

seizure An attack or sudden onset of certain symptoms, such a convulsions, loss of consciousness, and psychic or sensory disturbances; seen in epilepsy and can be substance induced. For types of seizures, see the specific term.

selection bias Nonrepresentative sampling resulting in erroneous conclusions.

selective inattention Harry Stack Sullivan's term for inattentiveness to anxiety-generating areas.

selective mutism Persistent failure to speak in one or more specific social situations.

selective serotonin reuptake inhibitors (SSRIs) Drugs that increase the concentration of serotonin in the brain; commonly used for the treatment of depression and anxiety disorders.

self Psychophysical total of a person at any given moment, including conscious and unconscious attributes. *See also* identity.

self-actualization Abraham Maslow's term for the inborn capacity of people to direct themselves in the healthiest way, toward a level of completeness.

self-analysis Investigation of one's own psychic components, which plays a part in all analysis, although to a limited extent, since few people are capable of sustaining independent and detached attitudes to the degree necessary for this approach to be therapeutic.

self-awareness Sense of knowing oneself, particularly in terms of insight into one's own psychodynamics; a major goal of most psychotherapies.

self-defeating personality disorder Disorder in which persons avoid or undermine pleasurable experiences and are drawn to situations or relationships in which they will suffer, even when better options are available; positive personal events are responded to with guilt or depression, reasonable offers of assistance from others are rejected, and anger or rejecting responses from others are invited, leaving them feeling hurt or humiliated. Also called masochistic personality disorder.

self-discovery In psychoanalysis, the freeing of the repressed ego in a person who has been brought up to submit to the wishes of the significant others around him.

self-fulfilling prophecy Situation in which one predicts a type of behavior that eventually one carries out.

self-help groups Groups composed of persons who want to cope with a specific problem or life crisis; characterized by homogeneity and cohesion.

self-image One's conception of one's own identity, personality, and worth as a person.

self-reference Denoting a person's repeatedly referring the subject under discussion back to himself.

self-system Harry Stack Sullivan's term for a personality system designed to ward off anxiety and preserve a positive view of the self.

SEM Standard error of the mean.

senescence Aging, as determined by inherent or hereditary factors.

senile dementia Dementia secondary to diffuse cerebral atrophy associated with advancing age; onset is insidious, progression is slow and gradual, and no specific therapy is known.

senium Period of old age, generally after age 65.

sensate focus learning In the treatment of sexual disorders, exercises that focus on heightening sensory awareness to touch, sight, sound, and smell.

sensation Feeling or impression when the sensory nerve endings of any of the six senses (taste, touch, smell, sight, kinesthesia, and sound) are stimulated.

sense of self One's feeling of individuality, uniqueness, and self-direction.

sensitivity In research, the probability of a positive test result in one who has the disorder. *Compare* specificity.

sensitivity training group Group in which members seek to develop self-awareness and an understanding of group processes, rather than to gain relief from an emotional disturbance. *See also* encounter group; T-group.

sensorimotor period *See* Piaget, Jean.

sensorium Hypothetical sensory center in the brain that is involved with clarity of awareness about oneself and one's surroundings, including the ability to perceive and process ongoing events in light of past experiences, future options, and current circumstances; sometimes used interchangeably with consciousness.

sensory aphasia Organic loss of ability to comprehend the meaning of words; fluid and spontaneous, but incoherent and nonsensical, speech. *See also* fluent aphasia and receptive aphasia.

sensory deprivation Lack of external stimuli and the opportunity for the usual perceptions; may be produced experimentally or may occur in real-life contexts (e.g., deep-sea diving, solitary confinement, loss of hearing or eyesight), leading to hallucinations, panic, delusions, and disorganized thinking. *See also* emotional deprivation.

sensory extinction Neurologic sign operationally defined as failure to report one of two simultaneously presented sensory stimuli, despite the fact that either stimulus alone is correctly reported. Also called sensory inattention.

sensory inattention SENSORY EXTINCTION.

Sentence Completion Test (SCT) Test designed to tap patients' conscious associations to areas of functioning in which the clinician may be interested; it is composed of a series (usually 75 to 100) of sentence stems, such as "I like….," "Sometimes I wish…," that patients are asked to complete in their own words.

separation anxiety Reaction expressed in a child who is isolated or separated from the mother, such as tearfulness, irritability.

separation anxiety disorder In DSM, an anxiety disorder characterized by an infant's or child's fear and apprehension on being removed from the parent or parent figure.

separation–individuation As elaborated by Margaret Mahler, the phase during which the child perceives itself as distinct from the mother; as the child separates, it becomes more independent and aware of its physical and psychological autonomy and can endure separation from the mother.

serotonin 5-Hydroxytryptamine (5-HT), an endogenous indolamines synthesized from dietary tryptophan and found in the gastrointestinal tract, the platelets, and the central nervous system. Evidence indicates that this monoamine serves as a neurotransmitter substance in the central nervous system and may play an important role in such diverse functions as sleep, sexual behavior, aggressiveness, motor activity, perception (particularly pain), and mood; dysfunction in central serotonergic systems has been proposed as a cause or factor in various mental disorders, including schizophrenia and the mood disorders.

serotonin-dopamine antagonists (SDAs) Drugs that block the receptors for both serotonin and dopamine in the brain. Also known as second-generation or atypical antipsychotics. *Compare* dopamine-receptor antagonists.

serotonin-norepinephrine reuptake inhibitors (SNRIs) Drugs that inhibit the reuptake of both serotonin and norepinephrine in the brain.

serotonin syndrome An excess of serotonin in the brain that causes a syndrome of confusion, delirium, tremor, and hyperthermia which may be fatal. Caused by overdose of serotonergic drugs.

serum levels *See* blood serum levels.

severe mental retardation *See* mental retardation.

sexual aberration Deviation from usual sexual practices. For types of sexual aberrations, see the specific term.

sexual arousal disorders Sexual dysfunction characterized by an inability to achieve a degree of sexual excitement sufficient to engage in coitus. Formerly called inhibited sexual excitement.

sexual aversion disorder Sexual desire disorder characterized by a disgust and/or avoidance of genital contact with a sexual partner.

sexual delusion Delusion centering on sexual identity, appearances, practices, or ideas.

sexual desire disorder Dysfunction characterized by a diminished or hypoactive sexual drive. Formerly called inhibited sexual desire. *See also* hypoactive sexual desire disorder; sexual aversion disorder.

sexual deviation PARAPHILIA.

sexual drive One of the two primal drives or instincts (the other is the aggressive drive) in Freud's dual-instinct psychoanalytic theory. It operates under the pleasure–unpleasure (pleasure–pain) principle, with its main goal being the preservation and maintenance of life, in contrast to the aggressive drive that follows the repetition–compulsion principle; its unconscious psychic energy is known as libido. Also known as life instinct; Eros.

sexual dysfunction Class of sexual disorders characterized by inhibitions in sexual desire or the psychophysiologic changes that characterize the sexual response cycle.

sexual identity Biologically determined sexual state. *Compare* gender identity.

sexualization Endowment of an object or function with a sexual significance that it previously did not have, or possessed to a lesser degree, in order to ward off anxieties associated with prohibited impulses or their derivatives.

sexual masochism *See* masochism.

sexual orientation distress Term that refers to a disorder characterized by persistent and marked distress about one's sexual orientation; most often characterized by a desire to acquire heterosexual rather than homosexual arousal.

sexual pain disorders Sexual dysfunction characterized by pain during coitus, such as dyspareunia and vaginismus.

sexual reassignment *See* transsexualism.

sexual sadism *See* sadism.

shaman Healer in a primitive tribe who uses supernatural or spiritual powers to cure the sick.

shame Failure to live up to self-expectation; often associated with fantasy of how person will be seen by others. *See also* guilt.

shaping Operant conditioning technique, used in behavior therapy, in which a behavioral goal is reached in stepwise fashion through selective reinforcement of closer and closer approximations of the desired behavior.

shared paranoid disorder *See* shared psychotic disorder.

shared psychotic disorder Rare disorder in which the delusion system of the patient has developed out of a close relationship with another person who had a previously established, similar delusional system. In DSM-5, this condition is called “delusional symptoms in partner of individual with delusional disorder.” It has historically been referred to as shared paranoid disorder, induced psychotic disorder, folie à deux, and double insanity.

shell shock Obsolete World War I term for an acute mental disorder arising as a consequence of combat experience. *See* posttraumatic stress disorder.

shift work circadian rhythm disorder Characterized by sleep pattern disruptions that result from night shift work; associated with insomnia, excessive sleepiness, or both.

shock treatment Obsolete term for convulsive therapy.

short-duration depressive episode A DSM-5 diagnosis characterized by depressed mood lasting between 4 and 14 days that is nonrecurring.

short-term memory Reproduction, recognition, or recall of perceived material within seconds to minutes after the initial presentation. Compare immediate memory; long-term memory.

shyness disorder AVOIDANT DISORDER OF CHILDHOOD OR ADOLESCENCE.

sibling One of two or more individuals having one common parent.

sibling rivalry Competition among siblings for the attention, affection, and esteem of their parents or for other recognition or reward.

sick role Use of illness or complaints about bodily functioning as a means of escaping personal failure and the expectation to succeed, or as a way of avoiding other social responsibilities. *See also* hypochondriasis.

side effect Action of a drug other than the desired therapeutic effect.

sign Objective evidence of a disease or disorder. *Compare* symptom.

signal anxiety Sigmund Freud’s term describing a type of anxiety characteristic of the psychoneuroses, in which a child learns to anticipate and react to the possibility of a traumatic situation before it becomes traumatic.

significant differences *See* statistical significance.

simple phobia Phobic disorder characterized by fear and avoidance of an object or situation not included in agoraphobia or social phobia.

simultanagnosia Impairment in the perception or integration of visual stimuli appearing simultaneously.

sitophobia Abnormal fear of eating.

skin-picking disorder *See* excoriation disorder.

Skinner, B. F. (1904–1990) American behavioral psychologist and author noted for his work on operant conditioning.

sleep Recurrent temporary period of relative unconsciousness, characterized by a reversible cessation of waking. Sensorimotor activity is accompanied by characteristic electroencephalographic changes; divided into REM (rapid eye movement) and non-REM (NREM) sleep.

sleep disorder Abnormality of the sleep–wake cycle. Sleep disorders include dyssomnias (insomnia disorder, primary insomnia, hypersomnia disorder, sleep–wake schedule disorder) and parasomnias (dream anxiety disorder, sleepwalking disorder, sleep terror disorder).

sleep-related eating disorder Sleep disorder characterized by the inability to fall back asleep after awakening without having something to eat or drink; predominantly affects infants and children.

sleep-related hallucinations Hallucinations that occur at sleep onset (hypnagogic) or on awakening (hypnopompic); hallucinations are often visual and common in patients with narcolepsy.

sleep-related hypoventilation Dyssomnia characterized by recurrent episodes of shallow breathing during sleep associated with insomnia and/or frequent arousal during the breathing disturbances. *See also* breathing-related sleep disorders.

sleep terror disorder Sleep disorder in which there are repeated episodes of abrupt awakenings from sleep, usually beginning with a panicky scream; characterized by verbalizations, confusion, autonomic activity, and poor recall of the dream. A non-REM sleep arousal disorder in DSM-5; also called pavor nocturnus; night terror. *See also* parasomnias.

sleep–wake schedule disorder Sleep disorder in which persons cannot sleep when they wish to sleep, but are able to sleep at other times; correspondingly, they cannot be fully awake when they want to be fully awake, but they are able to be awake at other times. Term used in ICD.

sleepwalking disorder DSM-5 term for somnambulism; a non-REM sleep arousal disorder. *See also* parasomnias.

slow-wave sleep (SWS) Non-REM sleep stages 3 and 4; so termed because of the characteristic wave appearance on the electroencephalogram (EEG) record. Also called delta sleep.

SNRI Serotonin-norepinephrine reuptake inhibitor.

social adaptation Adjustment to the whole complex of interpersonal relationships; the ability to live and express oneself in accordance with society's restrictions and cultural demands. *See also* adaptational approach.

social anxiety disorder *See* social phobia.

social breakdown syndrome Signs and symptoms indicating that the long-term institutionalization of a mental patient has itself contributed to the patient's symptoms; for example, development of progressive social and vocational incompetence, excessive passivity, and learning the chronic sick role.

social control Influence exerted on a person or a group of persons to conform to the demands or expectations of a society or any of its representative institutions, agencies, or organizations.

social instinct *See* herd instinct.

socialization Process through which an individual becomes a competent member of society; the condition in which inner impulses, that is, instincts, are expressed in conformity with cultural demands of the environment.

social learning theory Approach to behavior therapy that incorporates the elements of respondent and operant conditioning, but views cognitive processes as important mediators in such learning; relies on role modeling, identification, and human interactions.

social network therapy Type of therapy in which the therapist assembles all persons (relatives, friends, social relations, work relations) who have emotional or functional significance in the patient's life to provide feedback and support to the patient; all or some of the social network may be assembled at any given time. *See also* extended-family therapy.

social phobia Any of a variety of phobic disorders characterized by a fear of being observed by others (e.g., fear of public-speaking, blushing, eating in public, writing in front of others, using public lavatories). Also called social anxiety disorder. DSM-5 includes a "performance only" diagnostic category if the fear is limited to speaking or performing in public.

social psychiatry Branch of psychiatry concerned with ecologic, sociologic, and cultural variables that engender, intensify, or complicate maladaptive patterns of behavior and their treatment.

social (pragmatic) communication disorder Persistent failure to use verbal and nonverbal communication for social purposes; differentiated from Autism Spectrum Disorder

social therapy Rehabilitation therapy used with psychiatric patients to improve social functioning; for example, occupational therapy, therapeutic community, recreational therapy, milieu therapy, attitude therapy.

sociobiology Systematic study of the biologic basis of social behavior, pioneered by E. O. Wilson.

sociology Scientific study of group behavior and social organization.

sociometry Study and measurement of the interpersonal psychological structure of a group or society.

sociopath Obsolete term for a person with an antisocial personality disorder.

sociopathic behavior Former term for antisocial behavior.

sociotherapy Any treatment modality whose primary emphasis is on socioenvironmental and interpersonal factors; for example, the therapeutic community.

sodomy Anal intercourse. Legally, the term may include other sexual deviations as well, such as bestiality.

soldier's heart DA COSTA SYNDROME.

soma The body.

somatic delusion Delusion pertaining to the functioning of one's body.

somatic hallucination Hallucination involving the perception of a physical experience localized within the body.

somatic therapy Treatment of psychiatric disorders with biologic, organic, or physical treatments.

somatic symptom disorder Disorder characterized by multiple, recurrent physical complaints with no apparent physical cause and a preoccupation with the false belief of having a serious disease; symptoms have no demonstrable organic findings, are often misinterpreted by the patient, and are apparently linked to psychological factors. Previously called somatization disorder; also known as hypochondriasis; Briquet syndrome.

somatization disorder Somatoform disorder characterized by recurrent and multiple physical complaints with apparent physical cause; now called somatic symptom disorder in DSM-5. Also called Briquet syndrome.

somatoform disorders Term previously used for a group of mental disorders characterized by a marked preoccupation with the body and fears of disease; production of the symptoms is linked to psychological factors or conflicts but is not under voluntary control. Previously in DSM-IV, somatoform disorders included somatization disorder, conversion disorder, body dysmorphic disorder, and somatoform pain disorder. This group of disorders is now referred to as somatic symptom and related disorders. *See also* factitious disorder; malingering.

somatopagnosia Inability to recognize a part of one's body as one's own (also called ignorance of the body and autotopagnosia).

somatotype Physical build and structure of a person; one's physique.

somnambulism Repeated episodes of arising from bed during sleep, usually during the first third of the night during non-REM (stages 3 and 4 sleep); on awakening, the person has amnesia for the episodes. More frequently seen in males than in females, most commonly in children, who generally outgrow it; in adults, often associated with other psychiatric disturbances, such as schizoid personality disorder and schizophrenia.

somnolence Pathologic sleepiness or drowsiness from which one can be aroused to a normal state of consciousness.

spatial agnosia Inability to recognize spatial relations.

speaking in tongues Expression of revelatory message through unintelligible words; not considered a disorder of thought if associated with practices of specific Pentecostal religions. *See also* glossolalia.

specific developmental disorder *See* specific learning disorder.

specific learning disorders A class of neurodevelopmental disorders beginning in early childhood that are characterized by persistent difficulty or delays in learning skills in reading, written expression, and/or mathematics, which are inconsistent with the overall intelligence of the child. Specific learning deficit is indicated by the phrase, "with impairment in," either reading (dyslexia), written expression, or mathematics (dyscalculia). Previously called specific developmental disorder. *Compare* pervasive developmental disorder.

specificity In research, the probability of a negative test result in one who does not have the disorder.

specific reading disability DEVELOPMENTAL READING DISORDER.

speech disturbance Any of a variety of language and communication disorders not due to impaired function of speech muscles or organs of articulation.

speech sound disorder Language or speech disorder in which there is a consistent failure to correctly articulate a speech sound at the developmentally appropriate age in the absence of any underlying anatomical, physiologic, auditory, or neurologic abnormalities. Also called developmental articulation disorder, functional speech disorder, infantile perseveration, infantile articulation disorder, and phonologic disorder.

splitting Defense mechanism in which external objects are divided into “all good” and “all bad,” accompanied by the abrupt shifting of an object from one extreme category to the other.

SSRI Selective serotonin reuptake inhibitor.

stammering STUTTERING.

standard deviation (SD) Statistical measure of variability within a set of values so defined that, for a normal distribution, about 68% of the values fall within one SD of the mean, and about 95% lie within two SDs of the mean; sometimes represented by the Greek letter sigma (σ).

standard error of the mean (SEM) Statistical measure of how variable, that is, how reliable, a mean value calculated on the basis of a given sample is as an estimate of the mean of the population from which the sample was selected; obtained by dividing the standard deviation of the population by the square root of the number of measures in the sample.

standardized score Deviation of a score from its group mean expressed in standard deviation units. Also called z-score.

Stanford–Binet Intelligence Scale Primarily a verbal test of intellectual functioning, administered individually to children and adults, which emphasizes problem solving; one of the most widely used IQ tests.

startle reaction Reflex response to a sudden intense stimulus, consisting of a diffuse motor response involving flexion movements of the trunk and extremities and associated with a sudden increase in alertness; occurs in normal persons and in anxiety disorders.

state-dependent learning or memory Facilitated recall of information when one is in the same internal state (e.g., drug state) or external environment that the information was first acquired.

statistical inference Research term used to make a general statement about a study based on a small group or to decide whether to attribute differences between two groups to chance or to nonchance factors.

statistical significance Research term referring to an experimental result, based on a sample of observations, which demonstrates an outcome or effect of sufficient magnitude such that the probability of that result being due to chance alone is less than 5%. *See also* Type I error.

status epilepticus Continuous epileptic seizures. *See also* epilepsy.

stereotypy Continuous mechanical repetition of speech or physical activities; observed in cases of catatonic schizophrenia.

stereotypy and habit disorder *See* stereotypic movement disorder.

stereotypic movement disorder Motor disorder consisting of intentional, repetitive, and typically rhythmic behaviors that are nonfunctional, that is, serve no constructive, socially acceptable purpose (e.g., hand flapping, body-rocking, head-banging, face-slapping, skin picking). Movements appear self-soothing or self-stimulating, but in certain cases may also result in self-injury.

stilted speech Excessively formal, stiff, stylized, or pompous speech; overly polite, distant, or antiquated speech.

stimulant Drug that stimulates the central nervous system to produce increased psychomotor activity; for example, methylphenidate (Ritalin), caffeine, amphetamines.

stimulus Any event, internal or external to the person, which elicits some kind of nervous system activity or response.

Stockholm syndrome Syndrome in which captives identify with, and have sympathy for, their captors on whom they are dependent for survival; seen in terrorist–hostage situations. The major defense mechanism, as described by Anna Freud, was known as identification with the aggressor.

STP Slang name for 2,5-dimethoxy-4-methylamphetamine.

street drugs Readily available, illegally acquired psychoactive substances; for example, marijuana, heroin, cocaine, crack.

strephosymbolia Reading and writing disability characterized by confusion between similarly formed but oppositely oriented letters (as in b and d, p and q), with tendency to reverse the order of letters and words.

stress Circumstance in a person's life that causes tension or dysphoria. Stressors may be mild (e.g., termination of a friendship) or catastrophic (e.g., suicide of a spouse). DSM contains a Severity of Psychosocial Stressor Scale for children, adolescents, and adults. Also called psychosocial stressor.

stress immunity Failure to react to emotional stress.

stress reaction Disturbed psychological or physiologic functioning encountered when strong, involuntary, often unconscious internal impulses press for action that conflict with one's conscious, reality-oriented behavior.

stroke CEREBROVASCULAR ACCIDENT.

structured interactional group psychotherapy Group therapy developed by Harold Kaplan and Benjamin Sadock, in which the therapist provides a structural matrix for the group's interactions, the most important of which is that a different member of the group is the focus of the interaction in each session.

structured interview 1. Interview in which the same questions are asked of all patients. **2.** Interview in which the interviewer asks specific questions about specific topics.

Studies on Hysteria Title of a book by Josef Breuer and Sigmund Freud (1895) that described the cathartic method of treatment and the beginnings of psychoanalysis; demonstrated the psychological origins of hysterical symptoms and the possibility of effecting a cure through psychotherapy.

stupor 1. State of decreased reactivity to stimuli and less than full awareness of one's surroundings; as a disturbance of consciousness, it indicates a condition of partial coma or semi coma. **2.** In psychiatry used synonymously with mutism and does not necessarily imply a disturbance of consciousness; in catatonic stupor, patients are ordinarily well aware of their surroundings.

stuttering Speech disorder characterized by repetitions or prolongations of sounds, syllables, or words, or by unusual hesitations and pauses that disrupt the normal flow of speech. May often be associated with habitual eye blinking, facial grimacing, head jerks, or abnormal body movements. Also called stammering; child-onset fluency disorder.

subcoma therapy *See* insulin coma therapy.

subconscious
Obsolete term for the preconscious and the unconscious.

subjectivity State in which evaluation and interpretation are influenced by one’s own feelings and thinking.

sublimation Unconscious defense mechanism in which the energy associated with unacceptable impulses or drives is diverted into personally and socially acceptable channels; unlike other defense mechanisms, it offers some minimal gratification of the instinctual drive or impulse.

substance abuse Called substance use disorder in DSM-5; diagnosed according to the substance used, for example, alcohol, cocaine, cannabis, hallucinogens, inhalants, opioids, sedatives, stimulants, or tobacco. *See also* drug abuse.

substance use disorder A maladaptive pattern of substance use leading to impairments in functioning characterized by the substance being used in larger amounts over time; unsuccessful efforts to cut down use; cravings; and symptoms of tolerance and withdrawals.

substance withdrawal Cessation or reduction of intake of a substance regularly used to induce a state of intoxication.

substitution Unconscious defense mechanism in which a person replaces an unacceptable wish, drive, emotion, or goal with one that is more acceptable.

succinylcholine Powerful skeletal muscle relaxant used in anesthesia and as a muscle relaxant in electroconvulsive treatment.

suggestibility State of uncritical compliance with influence or of uncritical acceptance of an idea, belief, or attitude; commonly observed among persons with hysterical traits.

suggestion Process of influencing a person to the point of uncritical acceptance of an idea, belief, or other cognitive process.

suicidal ideation Thoughts or acts of taking one’s own life.

suicide Act of taking one’s own life. Although the underlying factors that lead a person to take their own life may not necessarily be fully understood by that person, the act of suicide is considered to be both voluntary and intentional.

suicide cluster *See* cluster suicide.

Sullivan, Harry Stack (1892–1949) American psychiatrist and psychoanalyst best known for his interpersonal theory of psychiatry.

sundowner's syndrome SUNDOWNING.

sundowning Syndrome in the older adults that usually occurs at night and is characterized by drowsiness, confusion, ataxia, and falling as the result of being overly sedated with medications. Also called sundowner's syndrome.

superego One of the three components of the psychic apparatus in the freudian structural framework, the others being the ego and id. It represents the internalized values, ideals, and moral attitudes of society and has psychic functions that are expressed in guilt, self-criticism, and conscience. It develops through the child's identification with his parents, and the severity of its prohibitions or demands is said to be related to the intensity and extent of resolution of the Oedipus complex. *See also* conscience.

supportive psychotherapy Psychotherapy that seeks to strengthen patients' defenses and to provide them with reassurance, rather than to probe deeply into their conflicts.

suppression Conscious act of controlling and inhibiting an unacceptable impulse, emotion, or idea; differentiated from repression in that repression is an unconscious process.

surrogate One who takes the place of another; a substitute.

susto Culture-specific mental syndrome seen in Latin America; characterized by an intense fear of evil supernatural forces.

SWS Slow-wave sleep.

symbiosis Dependent, mutually reinforcing relationship between two persons. Generally a normal, constructive characteristic of the infant–mother relationship, it can also occur in a destructive context, as between two mentally ill persons who reinforce each other's pathology or in a mother–infant relationship that induces in the child an intense separation anxiety, autism, and severe regression (symbiotic psychosis).

symbiotic psychosis Disorder in children characterized by an inability to separate psychologically or physically from the mother; such children experience confusion, intense separation anxiety, and an inability to function.

symbolization Unconscious defense mechanism in which one idea or object comes to stand for another because of some common aspect or quality in both; based on similarity and association; the symbols formed protect the person from the anxiety that may be attached to the original idea or object.

sympathetic nervous system The part of the autonomic nervous system innervating the viscera by outflow from thoracolumbar segments of the spinal cord. Its physiologic effects (e.g., tachycardia, peripheral vasoconstriction, pupillary dilation, decreased bowel motility) are predominantly adrenergic, and they characterize the fight-or-flight reaction that occurs in response to dangerous or threatening situations.

sympathomimetic drug Drug whose effects mimic the responses to stimulation of the sympathetic nervous system; for example, amphetamine, epinephrine.

sympathy Compassion for or sharing of another person’s feelings, ideas, and experiences. *Compare* empathy.

symptom Any abnormality indicative of a mental or physical disorder experienced by the patient but not observable by a physician. *Compare* sign.

symptomatic psychosis Brain dysfunction as a result of some general physical disease, the most common feature of which is a delirious state; may occur in the course of an acute infectious disease, such as pneumonia, influenza, or pelvic infections following childbirth.

symptom formation Unconscious psychic process in which a repressed impulse is indirectly released and manifested through a symptom; may be regarded as a kind of compromise reflecting a partial satisfaction of both the forbidden impulse and the ego reaction against it.

symptom substitution Phenomenon in which a set of symptoms that are removed directly in therapy, without regard for the unconscious conflicts responsible for their formation, are replaced by new symptoms. It has constituted a theoretical objection to such modalities as behavior therapy and hypnotherapy.

synapse Specialized structure at which one neuron makes contacts with the receptor elements in the adjacent neuron.

syncretic thought Jean Piaget's term for the prelogic, egocentric, solipsistic thinking that characterizes early childhood mentation. *See also* primary process.

synesthesia Condition in which the stimulation of one sensory modality is perceived as sensation in a different modality, as when a sound produces a sensation of color.

syphilis A chronic bacterial disease caused by the spirochete *Treponema pallidum* (*T. pallidum*). Associated with delirium, dementia, mania, psychosis, personality change, and depression.

syntactical (central) aphasia Aphasia characterized by difficulty in understanding spoken speech, associated with gross disorder of thought and expression.

syntaxic mode of experience Harry Stack Sullivan's term denoting the highest level of interpersonal communication, which is based entirely on reality and which is considered the desirable outcome of healthy development. *See also* secondary process.

syntropy Adolf Meyer's term characterizing healthy or wholesome relationships.

systematic desensitization Form of behavior therapy developed by Joseph Wolpe and others in which anxiety-evoking stimuli are represented to a patient in a state of deep muscle relaxation in an attempt to weaken the bond between the stimuli and the anxiety; particularly useful in the treatment of phobias. *See also* reciprocal inhibition and desensitization.

systematized delusion Group of elaborate delusions related to a single event or theme.

T

taboo Restrictions or prohibitions on behavior imposed by a culture, usually involving a dangerous, forbidden, or unclean entity.

tachylogia LOGORRHEA.

tactile amnesia *See* neurologic amnesia.

tactile hallucination Hallucination primarily involving the sense of touch. Also called haptic hallucination.

talion Retaliation; the fear that all bad deeds, intentional or accidental, will be punished in kind; that is, an eye for an eye and a tooth for a tooth. In psychoanalysis, the fear of castration or punishment for oedipal strivings, as in the Oedipus complex.

tangentiality Disturbance in which the person replies to a question in an oblique, digressive, or even irrelevant manner and the central idea is not communicated; failure to communicate the central idea distinguishes tangentiality from circumstantiality, in which the goal idea is reached in a delayed or indirect manner. The term has also been used synonymously with loosening of associations and speech derailment.

taphophobia Abnormal fear of being buried alive.

Tarasoff decision California Supreme Court decision that a physician or psychotherapist who has reason to believe that a patient may injure or kill someone must notify the potential victim, their relatives or friends, or the authorities.

tardive dyskinesia Late-appearing extrapyramidal syndrome associated with antipsychotic drug use and characterized by stereotyped involuntary movements of the lips, jaw, and tongue (tardive oral dyskinesia) and by other involuntary dyskinetic movements; symptoms may remit eventually or persist indefinitely after discontinuation of the medication, and the condition does not respond to treatment with antiparkinsonism drugs.

TAT Thematic Apperception Test.

telepathy Extrasensory perception of the mental activities of another person.

telepsychiatry Psychiatric practice using the electronic transmission of images.

temperament Inherent, constitutional predisposition to react, in a certain way to stimuli; variations in temperament are evident very early in life.

temporal lobe epilepsy COMPLEX PARTIAL SEIZURE.

temporary admission Hospitalization in which a person is admitted on the recommendation of one physician (which must be confirmed by a psychiatrist, upon admittance), for no more than 15 days if against their will; applies to patients who are so senile or confused that they require hospitalization and are notable to make decisions on their own, or so acutely disturbed that they must be immediately admitted to a psychiatric hospital on an emergency basis.

tension Physiologic or psychic arousal, uneasiness, or pressure toward action; an unpleasurable alteration in mental or physical state that seeks relief through action.

terminal insomnia Early morning awakening or waking up at least 2 hours before planning to. *Compare* initial insomnia; middle insomnia.

terrorism The use or threat of violence to instill fear with the intent to intimidate or coerce governments or societies in the pursuit of political, religious, or idealogic goals.

tertiary prevention *See* preventive psychiatry.

testamentary capacity One's competency to make a will. Three psychological abilities are necessary to demonstrate this competency: (1) the nature and extent of their bounty (property); (2) that they are making a will; and (3) who their natural beneficiaries are, for example, their spouse, children, and other relatives.

testimonial privilege Right to maintain secrecy or confidentiality in the face of a subpoena. Privilege belongs to the patient, not to the physician.

test of significance Statistical test used to determine whether a difference in the data obtained is due to chance or is meaningfully related to some specific factor or variable (statistically significant).

tetrahydrocannabinol (THC) Chemical substance produced by the Indian hemp plant (*Cannabis sativa*), believed to be responsible for most of the characteristic psychological effects of marijuana. Street drugs alleged to be THC frequently contain phencyclidine

(PCP), a veterinary tranquilizer that may produce extreme excitement, visual disturbances, and delirium.

T (training)-group Group that emphasizes training in self-awareness and group dynamics.

thanatology Study of death and matters leading up to and following it.

thanatophobia Abnormal fear of death.

Thanatos A term used by Freud to refer to the death instinct. *Compare* eros.

THC Tetrahydrocannabinol.

Thematic Apperception Test (TAT) Projective psychological test in which subjects supply interpretations of a series of ambiguous life situation drawings, based on their own feelings and attitudes and reflecting their unconscious conflicts and defensive structures.

theory Plausible or scientifically acceptable general principle or body of principles offered to explain phenomena.

therapeutic agent Anything that promotes healing, such as a substance, person, or apparatus.

therapeutic alliance Conscious contractual relationship between therapist and patient in which each implicitly agrees that they need to work together to help the patient with their problems; involves a therapeutic splitting of the patient’s ego into observing and experiencing parts. A good therapeutic alliance is especially necessary for the continuation of treatment during phases of strong negative transference.

therapeutic atmosphere All therapeutic, maturational, and growth-supporting agents, that is, cultural, social, and medical.

therapeutic community Institutional treatment setting designed with an emphasis on the importance of socioenvironmental and interpersonal influences in the therapy, management, and rehabilitation of the hospitalized mental patient. *See also* milieu therapy.

therapeutic group Group of patients joined together under the leadership of a therapist for the purpose of working together for psychotherapeutic ends; specifically, for the treatment of each patient’s mental disorders.

therapeutic role Position in which one aims to treat, bring about an improvement, or provide alleviation of a distressing condition or state.

therapeutic window Range of blood plasma levels of a drug within which optimal therapeutic effects occur, based on such factors as age, sex, race, genetic makeup, and concurrent medications.

therapist *See* psychotherapist.

tetracyclic antidepressants One group of antidepressant drugs so named because of their four-ringed structure. *See also* tricyclic antidepressants.

thinking *See* cognition.

thinking compulsion INTELLECTUALIZATION.

thioxanthene derivative One of a class of antipsychotic drugs closely related, both structurally and pharmacologically, to the phenothiazines; for example, chlorprothixene (Taractan), thiothixene (Navane).

third ear Ability to make use of intuition, sensitivity, and awareness of subliminal cues to interpret clinical observations of individual and group patients; a term introduced by the German philosopher Friedrich Nietzsche and later used in analytic psychotherapy by Theodor Reik.

third-party payer Health or medical care insurer.

thought broadcasting Feeling that one's thoughts are being broadcast or projected into the environment. *See also* thought withdrawal.

thought control *See* delusion of control.

thought deprivation BLOCKING.

thought disorder Any disturbance of thinking that affects language, communication, or thought content; the hallmark feature of schizophrenia. Manifestations range from simple blocking and mild circumstantiality to profound loosening of associations, incoherence, and delusions; characterized by a failure to follow semantic and syntactic rules which is inconsistent with the person's education, intelligence, or cultural background.

thought insertion Delusion that thoughts are being implanted in one's mind by other people or forces.

thought latency The period of time between a thought and its verbal expression. Increased in schizophrenia (*see* blocking) and decreased in mania (*see* pressured speech).

thought process disorder Symptom of schizophrenia, involving the intellectual functions, manifested by irrelevance and incoherence of verbal productions; ranges from simple blocking and mild circumstantiality to total loosening of associations, as in word salad.

thought withdrawal Delusion that one's thoughts are being removed from one's mind by other people or forces. *See also* thought broadcasting.

three-cornered therapy CO-THERAPY.

Three Essays on the Theory of Sexuality Title of a book by Sigmund Freud (1905) in which the libido theory was applied to the successive phases of psychosexual maturation in the infant, child, and adolescent; made possible the integration of a vast diversity of clinical observations and promoted the direct observation of child development.

Tic Seemingly involuntary sudden, brief, and rapid motor movements or vocalizations stimulated by an irresistible internal urge. Common motor tics include eye blinking, head-jerking, and grimacing; common vocal tics include coughing, throat-clearing, and grunting. Tics may have symbolic meaning. *See also* Tourette disorder.

tic disorders Predominately psychogenic class of motor disorders consisting of involuntary, spasmodic, stereotyped movements of a small group of muscles or vocal sounds; seen prominently in moments of stress or anxiety, and rarely as a result of organic disease. The two types of tic disorder in DSM-5 include Tourette disorder and persistent chronic motor or vocal tic disorders.

timidity Inability to assert oneself for fear of some fancied reprisal, even in the absence of objective evidence of potential harm.

tinnitus Noises in one or both ears, such as ringing, buzzing, or clicking; a rare side effect of some tricyclic antidepressant drugs.

TMS Transcranial magnetic stimulation.

tobacco dependence *See* nicotine dependence.

toilet training Program of teaching a child to control bladder and bowel functions; attitudes of both parent and child during the period may have important psychological implications for the child's later development.

token economy Program applying principles of operant conditioning to the management of an institutional setting, such as a ward or a classroom, during which reinforcement of desirable behavior is provided in the form of tokens or credits (conditioned reinforcers) that may be exchanged for a variety of positive reinforcers, such as food, television time, or a weekend pass.

tolerance Phenomenon in which, after repeated administration, a given dose of a drug produces a decreased effect or increasingly larger doses must be administered to obtain the effect observed with the original dose. *Compare* drug dependence.

tomography *See* computed tomography; positron emission tomography.

tonic convulsion Convulsion in which the muscle contraction is sustained.

topectomy Psychosurgical procedure involving ablation of a small and specific area of the frontal cortex for the treatment of a chronic psychotic patient.

torture The intentional infliction of physical and psychological suffering with the intent to gain information or to intimidate individuals or communities.

Totem and Taboo Title of a book by Sigmund Freud (1913) in which he applied his psychoanalytic concepts to the data of anthropology. He was able to afford much insight into the meaning of tribal organizations and customs, especially by invoking the Oedipus complex and the characteristics of magical thought as discovered from his studies of the unconscious.

Tourette syndrome Rare motor disorder, first described by a Paris physician, Gilles de la Tourette, that has its onset in childhood and is characterized by multiple involuntary muscle movements (tics), motor incoordination, and the presence of at least one vocal tic, such as echolalia and coprolalia.

toxicity State of being poisonous.

toxic psychosis Psychosis caused by toxic substances produced by the body or introduced into it in the form of chemicals or drugs.

toxophobia Abnormal fear of being poisoned.

trailing phenomenon Perceptual abnormality associated with hallucinogenic drugs in which moving objects are seen as a series of discrete and discontinuous images.

trainable Capable of achieving a certain degree of self-care and social adjustment at home and vocational usefulness in a closed supervised setting; describes the moderately mentally retarded (IQ of 35–40 to 50–55). *Compare* educable.

training group *See* T-group.

trait Distinguishing quality as an observed or measured characteristic of an individual.

trance Sleep-like state of reduced consciousness and activity.

tranquilizer Psychotropic drug that induces tranquility by decreasing anxiety or agitation, usually without clouding consciousness; formerly classified as major and minor tranquilizers which are now called antipsychotic and antianxiety agents, respectively.

transactional analysis Eric Berne's system of psychodynamic psychotherapy that focuses on the interactions (transactions) between the patient and the therapist in the treatment session and between the patient and others in a social environment; used in both individual and group psychotherapy. The system includes four components: (1) structural analysis of intrapsychic phenomena; (2) transactional analysis proper, the determination of the currently dominant ego state (parent, child, or adult) of each participant; (3) game analysis, identification of the games played in their interactions and of the gratifications provided; and (4) script analysis, uncovering of the causes of the patient's emotional problems.

transcranial magnetic stimulation (TMS) A form of brain stimulation in which a rapidly changing magnetic field is applied to the scalp to induce small, focused electrical currents in superficial cortex. Used for treatment-resistant depression.

transcultural psychiatry Branch of psychiatry concerned with the mental health of members of different cultures, that is, what may be considered normal for one culture may be unacceptable to another.

transference Unconscious tendency of a person to assign to others in the present and immediate environment those feelings and attitudes originally linked with significant figures in the person's early life, for example, identification of the therapist with a parent; the transference may be negative (hostile) or positive (affectionate). Analysis of transference phenomena is used as a major therapeutic tool in both individual therapy and group therapy to help patients better understand and gain insight into their behavior and its origins. *See also* countertransference; transference neurosis.

transference neurosis Phenomenon occurring in psychoanalysis in which the patient develops a strong emotional attachment to the therapist as a symbolized nuclear familial figure; repetition and the depth of that misperception or symbolization characterize it as a transference neurosis.

transsexualism Gender identity disorder in which a person has a desire to be of the opposite sex. Some transsexuals, many of whom have adopted the role of the opposite sex since childhood, have successfully undergone sex-changing surgical procedures, accompanied by intensive hormonal therapy and psychotherapy. *Compare* transvestism.

transvestism, transvestitism Paraphilia characterized by dressing in the clothing of the opposite sex for the purpose of sexual arousal; usually seen in men with a strong desire to appear as women. Also called cross dressing. *Compare* transsexualism.

trauma In psychiatry, a significant, upsetting experience or event that may precipitate or aggravate a mental disorder.

traumatic brain injury Damage to the structure of the brain caused by physical force that affects cognitive functioning with associated changes in mood and behavior. Most commonly associated with war injuries.

traumatic neurosis *See* posttraumatic stress disorder.

treated prevalence Number of persons being treated for a disorder in a defined geographic area.

tremor Rhythmical alteration in movement, which is usually faster than one beat a second; typically, tremors decrease during periods of relaxation and sleep and increase during periods of anger and increased tension.

trend of thought Thinking that centers on a particular idea associated with an affective tone.

triad Interactional relationship among three persons. Its prototype, the father–mother–child relationship, may evolve projectively in group therapy.

triadic therapy, triangular therapy CONJOINT THERAPY.

trichotillomania Compulsion to pull out one's hair.

tricyclic antidepressants One of a group of antidepressant drugs so named because of their three-ringed structure; currently the most widely used drugs for the treatment of depression, although anticholinergic side effects are prominent; for example, imipramine (Tofranil), amitriptyline (Elavil).

triskaidekaphobia The fear of the number 13.

trisomy Chromosomal aberration in which there are three chromosomes of one type instead of the usual two in an otherwise diploid cell, as in Down syndrome.

trisomy 21 syndrome DOWN SYNDROME.

true insight Understanding of the objective reality of a situation coupled with the motivational and emotional impetus to master the situation or change behavior.

t-test Statistical procedure designed to compare two sets of observations.

Tuke, William (1732–1822) British pioneer in the treatment of mental patients without the use of physical restraints.

Turner syndrome Disorder secondary to a chromosomal anomaly in which affected females typically have an XO sex chromosome constitution (instead of the usual XX) and a total of 45 chromosomes (instead of the usual 46). Classical somatic features include short stature, webbed neck, distinctive facies, sexual infantilism, and cubitus valgus. The incidence of mental retardation is not increased and severe psychopathic manifestations are uncommon, in contrast to the increased frequency of those disorders in XXY males. Also called ovarian dysgenesis; gonadal dysgenesis. *See also* Klinefelter syndrome.

twilight state Disturbed consciousness with hallucinations.

twin studies Studies that are useful in separating genetic and environmental influences and in looking for protective and precipitating factors.

twirling Sign present in autistic children who continually rotate in the direction in which their head is turned.

Type 1 error Research term referring to the rejection of the null hypothesis when it should have been retained. When the statistically calculated probability (p value) of making a Type 1 error on the basis of an experimental result is less than 0.05, the result is considered statistically significant; that type of error and the arbitrarily chosen 0.05 level of significance are sometimes called alpha (α) error. *Compare* Type 2 error.

Type 2 error Research term referring to the failure to reject a false null hypothesis, that is, concluding that any difference or effect demonstrated in an experiment is no greater than that expected by chance alone when, in fact, there is an actual difference or effect. Also called beta (β) error. *Compare* Type 1 error.

Type A personality As defined by Meyer Friedman and Ray Rosenman, cardiologists, an action-oriented individual who struggles to achieve poorly identifiable goals by means of competitive hostility; has increased amounts of low-density lipoprotein, serum cholesterol, triglycerides, and 17-hydroxycorticosteroids; and tends to develop coronary heart disease.

Type B personality As defined by Meyer Friedman and Ray Rosenman, cardiologists, a relaxed and less-aggressive individual (than Type A) who tends to strive less vigorously to achieve goals.

tyramine Sympathomimetic amine that acts by displacing stored transmitter from adrenergic axonal terminals; its degradation is normally aided by monoamine oxidase (MAO). Ingestion of food containing tyramine (e.g., cheese, beer, wine, yeast) by a patient on MAO inhibitors can provoke a potentially fatal hypertensive crisis.

U

UCR Unconditioned response or reflex.

UCS Unconditioned stimulus.

ultradian rhythm *See* biologic rhythm.

ululation Incoherent crying of a psychotic or hysterical patient.

unconditioned response, unconditioned reflex (UCR) *See* conditioning.

unconditioned stimulus (UCS) *See* conditioning.

unconscious **1.** One of three divisions of Sigmund Freud's topographic theory of the mind (the others being the conscious and preconscious) in which the psychic material is not readily accessible to conscious awareness by ordinary means; its existence may be manifested in symptom formation, in dreams, or under the influence of drugs. **2.** In popular but more ambiguous usage, any mental material not in the immediate field of awareness. **3.** Denoting a state of unawareness, with lack of response to external stimuli, as in a coma.

underachiever One who fails to perform at the level of their known potential or capability.

undersocialized Characterized by the absence of adequate social bonds to others, as a type of conduct disorder.

undifferentiated schizophrenia Type of schizophrenia with prominent psychotic symptoms that do not fall into the other schizophrenic subtypes; no longer included in DSM.

undoing Unconscious primitive defense mechanism, repetitive in nature, by which a person symbolically acts out in reverse something unacceptable that has already been done or against which the ego must defend itself; a form of magical expiatory action, commonly observed in obsessive-compulsive disorder.

unio mystica Feeling of mystic unity with an infinite power.

unipolar psychosis Mood disorder characterized by recurrent episodes of depression or, much more rarely, recurrent manic states. *Compare* bipolar disorder.

universality Total effect produced when all group therapy members share specific symptoms or problems.

unstructured interview **1.** Interview in which the interviewer chooses questions based on clinical judgment. **2.** Interview in which questions are related to what the patient brings up spontaneously.

urolagnia Paraphilia in which sexual pleasure is derived from urinating on someone else, being urinated on, or drinking urine.

utilitarian principle In policy and ethical decisions, a choice that is determined by whatever promises the greatest advantage to the common good.

utilization review committee Hospital committee that reviews the quality of services rendered, with particular attention to the appropriate use of facilities.

V

vagina dentata Sigmund Freud’s theory that an unconscious fantasy exists in certain persons who are afraid of women and who believe that there are “biting teeth” in the vagina.

vaginismus Painful vaginal spasms, usually psychogenic, that occur during coitus, causing dyspareunia; may have its onset during preparation for intercourse, making penile insertion impossible. *See also* Genito–pelvic pain/penetration disorder.

validity Degree to which a test measures what it claims to measure; the degree to which an experimental design yields data truly applicable to the phenomenon under investigation. *See also* reliability.

variable In research, any characteristic or factor that may assume different values. The terms “independent variable” and “dependent variable" indicate a relationship of cause and of effect, respectively.

variable-interval schedule *See* schedule of reinforcement.

variable-ratio schedule *See* schedule of reinforcement.

variance Statistical measure of the variability within a set of observations, defined as the sum of the squared deviations from the mean divided by the number of observations; it equals the standard deviation squared. *See also* analysis of variance.

variation Statistical term referring to the manner in which persons in a population vary among themselves with respect to a given quality or trait; variability of persons also affects the precision of sample estimates. Three common measures of variation are the range, the variance, and the standard deviation. Knowledge of variation in a population is important because it indicates how useful the mean value is as a representative figure.

vascular dementia A type of dementia in which a succession of strokes has led to a stepwise deterioration of intellectual functioning with associated focal neurologic signs and symptoms. Previously called multi-infarct dementia and currently called Major or Mild Vascular Neurocognitive Disorder in DSM-5-TR.

vegetative **1.** Pertaining to autonomic functions that are largely physiologic or concerned with the growth, nutrition, or general physical health and homeostasis of the organism. **2.** In depression, denoting characteristic symptoms such as sleep disturbance (especially early morning awakening), decreased appetite, constipation, weight loss, and loss of sexual response.

vegetative nervous system Obsolete term for autonomic nervous system.

velocardiofacial syndrome A genetic disorder caused by a microdeletion in chromosome 22. Characterized by cleft palate, heart defects, immune dysfunction, mild intellectual disability, psychotic disorders, and mood disorders. Also called Shprintzen syndrome or DiGeorge syndrome.

verbal amnesia *See* neurologic amnesia.

verbigeration Meaningless and stereotyped repetition of words or phrases as seen in schizophrenia. Also called cataphasia. *See* *also* perseveration.

verbiomania LOGORRHEA.

vertigo Sensation that one or the world around one is spinning or revolving; a hallmark symptom of vestibular dysfunction, not to be confused with dizziness.

visceral learning *See* biofeedback; learned autonomic control.

visual agnosia Inability to recognize objects or persons.

visual amnesia *See* neurologic amnesia.

visual hallucination Hallucination primarily involving the sense of sight.

visual motor gestalt test BENDER-GESTALT TEST.

vitamin therapy *See* orthomolecular psychiatry.

voice disorder A communication disorder that is characterized by abnormalities in vocal quality, pitch, and volume.

volitional Self-initiated.

volubility LOGORRHEA.

voluntary admission Hospitalization in which patients apply in writing for admission to a psychiatric hospital and are admitted if psychiatric examination reveals such need.

voluntary control Deliberate bringing about of an activity or refraining from an activity. The person exercising the control is aware of the nature of the activities and of the end to be achieved.

volunteer bias Sampling bias of persons who volunteer to take part in an experiment; such persons are not truly representative of the population and may tend to skew the results of the study, for example, the subjects may be more cooperative.

von Domarus principle Theory proposed by Eilhard von Domarus (1946) to explain schizophrenic thinking, which is characterized by the idea that two things are identical merely because they have identical predicates or properties.

voyeurism Paraphilia in which sexual excitement, frequently with orgasm, is obtained by looking at others naked, disrobing, or engaging in sexual activity. Also called voyeuristic disorder. *Compare* scopophilia.

W

WAIS Wechsler Adult Intelligence Scale.

Watson, John B. (1878–1958) American psychologist who used Ivan Pavlov's theory of classical conditioning to explain human behavior. *See also* Little Albert.

waxy flexibility Condition in which a person maintains the body position into which they are placed. Also called catalepsy.

Wechsler Adult Intelligence Scale (WAIS) Intelligence test that assesses intellectual functioning in adults.

Wechsler Intelligence Scale for Children (WISC) Intelligence test that assess intellectual functioning in children aged 5 to 15.

Wechsler Preschool and Primary Scale of Intelligence (WPPSI) Intelligence test that assesses intellectual functioning in children aged 4 to 6.

Wednesday Evening Society Small group of Sigmund Freud's followers who in 1902 started meeting with him informally on Wednesday evenings to receive instructions in psychoanalysis; as the society grew in numbers and importance, it evolved in 1910 into the Vienna Psychoanalytic Society.

weekend hospital Partial hospitalization in which the patient spends only weekends in the hospital and functions in the outside world during the week. *Compare* day hospital; night hospital.

Wernicke–Korsakoff syndrome Thiamine deficiency seen in Korsakoff psychosis (alcohol amnestic syndrome). Also called alcohol dementia. *See also* Wernicke fluent encephalopathy; Korsakoff psychosis.

Wernicke aphasia FLUENT APHASIA.

Wernicke fluent encephalopathy Neurologic disorder caused by nutritional deficiency of thiamine (vitamin B) as seen in chronic alcoholism, with hallmark clinical features of ophthalmoplegia, ataxia, and confusion; some patients show disorders of short-term memory and other cognitive functions characteristic of Korsakoff psychosis (alcohol amnestic disorder), which has come to be regarded as a more advanced and chronic stage of the same disease process. Wernicke encephalopathy responds dramatically to the administration of thiamine, which may also prevent progression into the largely irreversible Korsakoff psychosis.

Weyer, Johann (1515–1588) Dutch physician considered by some to be the first psychiatrist, whose interest in human behavior led to his writing *De Praestigiis Daemonum* (1563), a landmark in the history of psychiatry.

white-out syndrome Psychosis that occurs in Arctic explorers and mountaineers who are exposed to a lack of diverse stimuli in the snow-clad environment.

WHO World Health Organization.

Williams syndrome A genetic disorder associated with a microdeletion on chromosome 7 and is characterized by “elfin-like” facial features, mild or moderate intellectual disability, cardiovascular disease, hyperacusis, hypercalcemia, and neuromusculoskeletal and renal abnormalities.

Wilson disease An autosomal recessive genetic disorder that causes excessive amounts of copper in the liver or brain. Patient may show psychological symptoms such as depression, suicidal ideation, and aggression. Is fatal without treatment.

windigo Culture-specific syndrome seen in Canadian Indians, characterized by a delusion of transformation into or possession by a windigo (wihtigo), a feared supernatural cannibalistic monster.

WISC Wechsler Intelligence Scale for Children.

withdrawal 1. Pathologic retreat from interpersonal contact and social involvement; extreme decrease of intellectual and emotional interest in the environment. It may be seen in schizophrenia and depression. **2.** *See* substance withdrawal.

withdrawal syndrome Stereotyped constellation of signs and symptoms that appears when a person stops taking a drug or chemical substance on which they have become physically dependent. It is more likely to appear immediately after abrupt discontinuation, but may also appear a few days later; specific clinical features are characteristic for the particular drug or substance. Also called abstinence syndrome; discontinuation syndrome.

word approximation Use of conventional words in an unconventional or inappropriate way (metonymy) or of new words that are developed by conventional rules of word formation (e.g., “handshoes” for gloves and “time measure” for clock); distinguished from a neologism, which is a new word whose derivation cannot be understood. *See also* paraphasia.

word association technique Technique devised by Carl Jung, who presented stimulus words to patients and had them respond with the first word that came to mind; it gives insight into persons’ unconscious or subconscious thinking.

word blindness Obsolete word for alexia.

word salad Incoherent, essentially incomprehensible mixture of words and phrases commonly seen in far-advanced cases of schizophrenia. *See also* incoherence.

working through In psychoanalysis, the repeated and varied examination of a conflict, neurotic trend, or other problem for the purpose of increasing insight into unconscious processes and effecting adaptive personality changes; constitutes a fundamental facet of the psychoanalytic process.

World Health Organization (WHO) Specialized agency of the United Nations, established in 1948, governed by the World Health Assembly. Technical committees and regional organizations further the purpose of the WHO: “the attainment by all people of the highest level of health.” It publishes the *International Classification of Diseases* (ICD).

World Psychiatric Association (WPA) International association, encompassing psychiatric associations from 75 countries throughout the world, whose objectives are to strengthen and promote international cooperation in psychiatry and among psychiatrists working in different fields and to encourage the exchange of information concerning psychiatric disorders.

WPA World Psychiatric Association.

WPPSI Wechsler Preschool and Primary Scale of Intelligence.

X-Y-Z

xenophobia Abnormal fear of strangers.

yohimbine Drug that blocks adrenergic α-receptors; because of its parasympathetic effect that causes vasodilatation, it has been used as an agent to help induce penile erection.

York retreat One of the earliest hospitals for the mentally ill, founded in 1792 in York, England, by William Tuke and other Quakers.

yttrium-90 Radioisotope that has been implanted experimentally in the brain to destroy certain tracts and areas in an effort to improve severe mental disorder.

zeitgeist German word for the general intellectual, moral, and cultural climate of an era.

zoophilia Paraphilia in which animals are used to produce sexual excitement. *See also* bestiality.

zoophobia Abnormal fear of animals.

z-score STANDARDIZED SCORE.

Zung Self-Rating Depression Scale (SDS) Twenty-item report in which the patient answers questions about symptoms of depressive disorders such as sleep patterns, appetite disturbance, libido, irritability, and hopelessness.

zygote Fertilized egg produced by the union of two cells (sperm and ovum) to form a diploid cell in sexual reproduction.

zygotic Pertaining to a zygote or to the individual developing from such a fertilized egg.

Appendix C

Comparative Diagnoses

The following tables summarize the key features of the most common psychiatric diagnoses.

The columns compare the diagnoses of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR) (2022), International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10), 2019 version which is still in use in the United States, and the International Statistical Classification of Diseases and Related Health Problems 11th Revision (ICD-11) (2022).

The rows list the domains common to most diagnostic systems: Diagnostic name, Duration (if required), Symptoms, Required number of symptoms, Psychosocial consequences (of the disorder), Exclusions (the symptoms are not better explained by____), Symptom specifiers, Course specifiers, Severity specifiers, and additional comments. In cases where no system commons on that domain, the row is omitted.

The tables are meant to compare diagnostic approaches and are not intended as a substitute for the actual criteria. The reader is referred to the source materials for the criteria.

Not all diagnoses are included here. For example, the three classification system includes secondary disorders (“due to a medical condition,” “due to a substance”), which should be used if the evaluation determines that the symptoms are better explained by a medical condition or the effects of a medication or substance. They also include “Other” and “Unspecified” categories to include cases of clinically significant symptoms that do not meet the full criteria, or there is insufficient information, or the clinician chooses not to specify why the criteria are not met. In both cases, those diagnoses are omitted here.

1. NEURODEVELOPMENTAL DISORDERS

Intellectual Disabilities

Table 1.

Intellectual Disability (Intellectual Developmental Disorder)/Disorders of Intellectual Development

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Intellectual disability (intellectual developmental disorder)	Mental retardation	Disorders of intellectual development
Symptoms	▸ Deficits in reasoning, abstraction, judgment, learning ▸ Deficits in adaptive functioning causing failure to meet developmental milestones for independence and social functioning ▸ Deficits arise during the developmental period	An arrest or incomplete development of ≥1 of the skills contributing to overall intelligence: cognitive, language, motor, and social abilities	▸ The presence of significant limitations in intellectual functioning across various domains such as perceptual reasoning, working memory, processing speed, and verbal comprehension. ▸ The presence of significant limitations in adaptive behavior, which refers to the set of conceptual, social, and practical skills that have been learned and are performed by people in their everyday lives ▸ Onset occurs during the developmental period.
Severity specifiers	**Mild** **Moderate** **Severe** **Profound**	**Mild** **Moderate** **Severe** **Profound**	**Mild** **Moderate** **Severe** **Profound**
Comments	**Unspecified diagnosis**. Unspecified intellectual developmental disorder (intellectual disability) for persons >5 years, and assessment is difficult or impossible due to sensory or other physical impairments		**Disorder of intellectual development, provisional.** Assigned when there is evidence of a disorder of intellectual development but the individual is an infant or child under the age of 4, making it difficult to ascertain whether the observed impairments represent a transient delay. Disorders of intellectual development,

unspecified.

Table 2.
Global Developmental Delay

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Global developmental delay	Disorder or developmental (global delay)	Disorders of intellectual development, unspecified
Symptoms	Failure to meet developmental milestones before the age of 5, with inability to perform full assessment of testing of intellectual functioning	Classified as an unspecified disorder	

Communication Disorders

Table 3.
Language Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Language disorder	Specific developmental disorders of speech and language	Developmental language disorder
Duration	Onset in the early developmental period	Onset in the early developmental period	Onset in the early developmental period
Symptoms	▸ Reduced vocabulary ▸ Limited sentence structure ▸ Impairments in discourse Difficulties are substantially below that expected for that age group	General category that includes the communication disorders listed below, including articulation, expressive, and receptive disorders	Persistent deficits in the acquisition, understanding, production, or use of language (spoken or signed) beyond that expected for the age. This includes deficits in the following: ▸ Phonologic awareness ▸ Syntax, morphology, or grammar ▸ Semantics ▸ Narrative or conversational discourse ▸ Pragmatics
Required number of symptoms	Any of the above		Any of the above
Psychosocial consequences	Limitations in communication affect social, academic, and occupational functioning		Limitations in communication affect social, academic, and occupational functioning
Exclusions	Sensory impairment, motor dysfunction, or medical or neurologic condition		Disorders of intellectual development, autism spectrum disorder, other neurodevelopmental disorders, Sensory impairment, motor dysfunction, or medical or neurologic condition
Symptom specifiers			With impairment of receptive and expressive language With impairment of mainly expressive language

			With impairment of mainly pragmatic language With other specified language impairment

Table 4.
Speech Sound Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Speech sound disorder	Expressive language disorder	Developmental speech sound disorder
Duration	Onset in the early developmental period		Onset in the early developmental period
Symptoms	Difficulty with speech sound production (articulation of phonemes).	The child's ability to use expressive spoken language is markedly below the appropriate level for its mental	Persistent errors of pronunciation, articulation, or phonology
Psychosocial consequences	Limitations in communication affect social, academic, and occupational functioning		Limitations in communication affect social, academic, and occupational functioning
Exclusions	Not due to congenital/acquired conditions affecting speech (cerebral palsy, cleft palate), sensory impairment, or other medical or neurologic condition		Not due to congenital/acquired conditions affecting speech (cerebral palsy, cleft palate), sensory impairment, or other medical or neurologic condition

Table 5.
Childhood Onset Fluency Disorder (Stuttering)

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Childhood onset fluency disorder (Stuttering)	Stuttering (stammering)	Developmental speech fluency disorder
Duration	Onset of symptoms during the developmental period		Onset of symptoms during the developmental period
Symptoms	Fluency and speech abnormalities as evidenced by: ▶ Repetition of sounds and/or syllables ▶ Prolonging sounds or consonants ▶ Using broken words ▶ Blocking or pausing ▶ Frequent word substitution (circumlocution) ▶ Excess physical tension during word production ▶ Repeating monosyllabic words	Speech is characterized by repetition or prolongation of sounds or words or by frequent pauses, resulting in disruption or normal flow or rhythm of speech.	Frequent or pervasive disruption of the normal rhythmic flow and rate of speech, including: ▶ repetitions and prolongations in sounds, syllables, words, and phrases ▶ as well as blocking (inaudible or silent fixations or inability to initiate sounds) ▶ word avoidance or substitutions
Psychosocial consequences	Causing severe anxiety and/or impairment in areas of functioning		Limitations in communication affect social, academic, and occupational functioning
Exclusions	Result of: ▶ Speech-motor or sensory issues ▶ Neurologic injury ▶ Another medical condition ▶ Another mental disorder Note: later-onset cases after the developmental period are coded as “adult-onset fluency disorder”	Result of: Tic disorders	Disorder of Intellectual Development Disease of the Nervous System Sensory impairment Structural abnormality

Table 6.
Social (Pragmatic) Communication Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Social (pragmatic) communication disorder	Developmental disorder of speech and language, unspecified	Listed as a specifier for developmental language disorder (with impairment of mainly pragmatic language)
Symptoms	Difficulties with using verbal or nonverbal communication in social situations, as evidenced by: ▶ Difficulty greeting others or sharing information with others ▶ Difficulties matching tone or context in a conversation ▶ Difficulty following social rules and expectations in conversations ▶ Difficulties understanding implicit meaning in conversations Onset of symptoms in the early developmental period		
Psychosocial consequences	Causing functional limitation		
Exclusions	Another medical or neurologic condition Autism spectrum disorder Intellectual disability or global developmental delay Another mental disorder		

Table 7.
Unspecified Communication Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Unspecified communication disorder	Developmental disorder of speech and language, unspecified	Developmental speech or language disorders, unspecified
Symptoms	Limitations in communication not meeting criteria for other disorders		
Psychosocial consequences	Limitations in communication affect social, academic, and occupational functioning		
Comments	Used for cases in which the clinician chooses not to specify reasons for the disorder or there is insufficient information.		

Autism Spectrum Disorder

Table 8.
Autism Spectrum Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Autism spectrum disorder	Childhood autism	Autism spectrum disorder
Symptoms	Symptoms of difficulties/deficits in social interaction, as evidenced by: ▶ Deficits in emotional reciprocity or failure to respond appropriately in social interactions ▶ Deficits in nonverbal communication, eye contact, body language/gesturing, and/or facial expression ▶ Deficits in forming, developing, or maintaining relationships or friendships Restricted and/or repetitive behaviors or interests, such as in: ▶ Stereotyped motor movements ▶ Inflexibility with regard to daily routines or rituals ▶ Restricted or fixed interests ▶ Hyper- or hyporeactivity to sensations or sensory stimuli Symptoms present in the early developmental period	Pervasive disorder of development involving ▶ Abnormal development present before the age of 3 ▶ Abnormal social functioning, communication, and repetitive/restricted behaviors	Persistent deficits in initiating and sustaining social communication and interactions outside the range for that age. May include limitations in: ▶ Understanding or responding to verbal or nonverbal communication ▶ Understanding or using language in a social context ▶ Social awareness ▶ Ability to imagine or respond to feelings ▶ Mutual sharing of interests ▶ Ability to make and sustain peer relationship Persistent restricted, repetitive, and inflexible patterns of behavior, interests, or activities that are inappropriate or excessive for the social context. These may include persistent: ▶ Lack of adaptability to new experiences or circumstances ▶ Inflexible adherence to routines ▶ Excess adherence to rules ▶ Excess and persistent ritualized patterns of behavior ▶ Repetitive and stereotyped motor movement

			▶ Preoccupation with ≥1 special interests, parts of objects, or stimuli ▶ Hyper or hyposensitivity to sensory stimuli or unusual interest in the stimuli
Required number of symptoms	All three of the above criteria met		
Psychosocial consequences	Marked impairment in functioning		Marked impairment in functioning
Exclusions	Result of: ▶ Intellectual disability ▶ Global developmental delay		Rett syndrome
Symptom specifiers	***With or without accompanying intellectual impairment*** ***With or without accompanying language impairment*** ***Associated with another neurodevelopmental, mental, or behavioral disorder*** ***With catatonia***		With or without accompanying intellectual impairment With or without accompanying language impairment ▶ With mild or no impairment of functional language ▶ With impaired functional language ▶ With complete, or almost complete, absence of functional language
Severity specifiers	For social impairment: **mild, moderate, severe** For behavior patterns: **mild, moderate, severe**		
		NOTE: atypical autism–symptoms do not meet all diagnostic criteria for childhood autism NOTE: Asperger syndrome—difficulties in social interaction similar to that found in autism, in addition to restricted interests;	

however, no deficits in language or cognition are seen

Attention-Deficit/Hyperactivity Disorder

Table 9.
Attention-Deficit/Hyperactivity Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Attention-deficit/hyperactivity disorder	Hyperkinetic disorders Disturbance of activity and attention	Attention-deficit hyperactivity disorder
Duration	6 months or longer		Persistent for ≥6 months, begins before age 12 (usually)
Symptoms	A persistent pattern of either or both of inattention, as evidenced by: ▶ Failure to give attention to details or making frequent mistakes ▶ Difficulty maintaining attention or focus ▶ Not seeming to listen when spoken to ▶ Failing to follow through with tasks or instructions ▶ Difficulty organizing tasks ▶ Avoiding or delaying tasks that require sustained mental effort or attention ▶ Frequently losing or misplacing items ▶ Easily distracted ▶ Forgetfulness Hyperactivity/impulsivity, as evidenced by: ▶ Fidgeting or restlessness ▶ Inability to remain in a seat ▶ Running around or climbing in inappropriate social situations ▶ Inability to engage in quiet play or activities ▶ Difficulty remaining still or inactive ▶ Talking excessively ▶ Blurting out answers without waiting for their turn	Early onset (usually before the age of 5) of inability to engage in activities requiring prolonged attention, in addition to disorganization of task completion and excessive activity. There may also be associated impulsivity, recklessness, and social disinhibition.	▶ Inattention symptoms and/or a combination of hyperactivity and impulsivity symptoms ▶ Outside the limits of normal variation expected for age and level of intellectual development ▶ Symptoms vary according to chronologic age and disorder severity ▶ Occur in a variety of settings Inattention ▶ Difficulty sustaining attention ▶ Easily distracted ▶ Loses things Hyperactivity ▶ Excess motor activity ▶ Difficulty engaging in activities quietly ▶ Blurts out answers in school ▶ Inappropriately acts out in response to stimuli

	▶ Difficulty waiting for one’s turn ▶ Frequently interrupting others Symptoms present before the age of 12, with symptoms present in ≥2 settings		
Required number of symptoms	≥6 symptoms of inattention (≥5 for adolescents/adults) and/or ≥6 of hyperactivity-impulsivity (≥5 for adolescents/adults)		
Psychosocial consequences	Impairment in functioning		Impairment in functioning
Exclusions	Result of: ▶ Other mental illness		Another mental disorder Substance or medication
Symptom specifiers	***Combined presentation*** ***Predominantly inattentive presentation*** ***Predominantly hyperactive/impulsive presentation***		Combined presentation Predominantly inattentive presentation Predominantly hyperactive/impulsive presentation
Course specifiers	***In partial remission (full/all criteria not met for prior 6 months, though symptoms still causing significant impairment)***		
Severity specifiers	**Mild** (few symptoms in excess of the minimum required for diagnosis, in addition to only minor impairment at most in functioning) **Moderate** (intermediate severity/impairment) **Severe** (several symptoms of high severity and resulting in marked impairment)		
	Other specified attention-deficit/hyperactivity	Hyperkinetic conduct disorder—referring to symptoms present in	**Other specified attention-**

	disorder reserved for cases where above symptoms cause distress or impairment but do not meet full criteria for a diagnosis **Unspecified attention-deficit hyperactivity** disorder is reserved for cases in which symptoms do not meet full criteria, and clinician chooses not to specify reasons for the disorder, or there is insufficient information	the context of conduct disorder	**deficit/hyperactivity disorder** **Unspecified attention-deficit hyperactivity**

Learning Disorders

Table 10.
Specific Learning Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Specific learning disorder	Specific developmental disorders of scholastic skills Specific reading disorder Specific spelling disorder Specific disorder of arithmetical skills Mixed disorder of scholastic skills	Developmental learning disorder
Duration	≥6 months		Onset in early school years
Symptoms	Difficulties with learning or academics, below what would be expected per individual's age or school level, as evidenced by: ▶ Inaccuracy or slowness with reading ▶ Difficulty understanding or comprehending what is being read ▶ Trouble with spelling ▶ Difficulty with written expression ▶ Difficulties with numbers or calculation ▶ Difficulties with mathematical reasoning Symptoms begin during school-age years (but may become obvious during the later part of academic career)		The presence of significant limitations in learning academic skills of reading, writing, or arithmetic, resulting in a skill level markedly below what would be expected for age
Required number of symptoms			
Psychosocial consequences	Impairment in functioning		Significant impairment to academic, occupational, or other areas of functioning
Exclusions	Result of:		External factors

	▸ Intellectual disability ▸ Visual or auditory impairment ▸ Another mental disorder ▸ Another medical disorder ▸ Psychosocial adversity ▸ Result of inadequate educational instruction	(economic, environmental disadvantage, lack of access to educational opportunities) Disorders of intellectual development, other neurodevelopmental disorders, and motor or sensory disorders
Symptom specifiers	***With impairment in reading*** ***With impairment in written expression*** ***With impairment in mathematics***	Impairment in reading Impairment in written expression Impairment in mathematics
Course specifiers		
Severity specifiers	**Mild** (minimal accommodations or support needed during school) **Moderate** (some accommodations needed) **Severe** (intensive individualized and specialized teaching needed for most of the schooling)	
Comments		Other specified impairments in learning Developmental learning disorder, unspecified

Motor Disorders

Table 11. Developmental Coordination Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Developmental coordination disorder	Specific developmental disorder of motor function	Developmental motor coordination disorder
Duration	Onset in the early developmental period		Onset in the early developmental period
Symptoms	The development of coordinated motor skills is far behind what would be expected for an individual's chronologic age	Significant impairment in motor coordination	Significant delay in the acquisition of gross or fine motor skills and impairment in the execution of coordinated motor skills. May manifest as ▶ Clumsiness ▶ Slowness ▶ Inaccuracy of motor performance. Coordinated motor skills are markedly below that expected on the basis of age.
Required number of symptoms			
Psychosocial consequences	Resulting in impaired ability to play, perform self-care, and function.		Limitations in activities of daily living, school work, vocation, leisure activities, or other important areas of functioning
Exclusions		General intellectual retardation Any specific congenital or neurologic disorder	Disease of the nervous system Disease of the musculoskeletal system Connective tissue disease Sensory impairment Disorder of intellectual development
Symptom specifiers	Result of: ***Another neurologic condition*** ***Intellectual disability*** ***Visual impairment***		

Table 12.
Stereotypic Movement Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Stereotypic movement disorder	Stereotyped movement disorders	Stereotyped movement disorder
Duration			Onset in the early developmental period. Persists ≥several months
Symptoms	Repetitive and seemingly purposeless movements with onset during the developmental period	The presence of voluntary, repetitive, stereotyped, and seemingly purposeless movements outside of any other psychiatric or neurologic condition; may include such things as repeatedly rocking, hand-flapping, or biting	Voluntary, repetitive, stereotyped, apparently purposeless, and often rhythmic movements
Required number of symptoms			
Psychosocial consequences	Marked interference in activities and functioning		Marked interference in activities and functioning or result in self-inflicted injury
Exclusions	Other medical condition Substance- or medication-induced, including withdrawal Other mental disorder	Tic disorder Trichotillomania Stereotypies that are part of a broader psychiatric condition Nail biting Nose-picking Movement disorders of organic origin	
Symptom specifiers	With self-injurious behavior Without self-injurious behavior Associate with a known medical or genetic condition, neurodevelopmental disorder, or environmental factor		With self-injury Without self-injury
Course specifiers			
Severity specifiers	**Mild** (symptoms easily suppressed by sensory stimulus or distraction)		

	Moderate (significant behavioral modification needed to control symptoms) **Severe** (continuous measures taken to prevent harm or injury)	
Comments		Stereotyped movement disorder, unspecified

Tic Disorders

Table 13.
Tourette Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Tourette disorder	Combined vocal and multiple motor tic disorder (de la Tourette)	Tourette syndrome
Duration	Despite waxing or waning, symptoms have been present for ≥1 year since the initial onset		≥1 year. Onset in developmental period.
Symptoms	Presence of both motor and vocal tics over the course of illness, with onset <age 18	Tic disorder involving multiple motor ticks and ≥1 more vocal tics; symptoms present in childhood or adolescence and tend to persist into adulthood.	The presence of both motor tic(s) and phonic tic(s) that may or may not manifest concurrently or continuously during the symptomatic course. Motor and phonic tics are defined as sudden, rapid, nonrhythmic, and recurrent movements or vocalizations, respectively.
Exclusions	Other medical condition Substance- or medication-induced, including withdrawal		Another medical condition (e.g., Huntington disease) Substance or medication on (e.g., amphetamine), including withdrawal effects (e.g., from benzodiazepines).
Comments	If symptoms of ONLY motor OR vocal tics, the diagnosis would be “Persistent (Chronic) Motor or Vocal Tic Disorder”		

Table 14.
Persistent (Chronic) Motor or Vocal Tic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Persistent (chronic) motor or vocal tic disorder	Tic disorder	▶ Chronic motor tic disorder ▶ Chronic phonic tic disorder
Duration	Despite waxing or waning, symptoms have been present >1 year since the initial onset		≥1 year
Symptoms	Presence of either motor or vocal tics over the course of illness, but not both. Onset before the age of 18	Presence of a tic, defined as an involuntary, fast, repeated, and nonrhythmic motor movement or vocal production that is sudden and seemingly purposeless. Such tics are experienced as urges that are strong, but that can be suppressed variably	Motor and vocal tics are defined as sudden, rapid, nonrhythmic, and recurrent movements or vocalizations, respectively
Required number of symptoms			
Psychosocial consequences			
Exclusions	Other medical condition Substance- or medication-induced, including withdrawal		
Symptom specifiers			
Course specifiers		Transient tic disorder (lasting <12 months) Chronic motor or vocal tic disorder (lasting longer than a year)	
Severity specifiers			
Comments	If symptoms are less than 1 year, the diagnosis would be **provisional tic disorder** **Other specified tic disorder** is reserved for cases in which symptoms do not meet full criteria, and the	Chronic motor or vocal tic disorder (lasting longer than a year)	

	clinician chooses not to specify reasons for the disorder, or there is insufficient information.	

2. SCHIZOPHRENIA SPECTRUM DISORDERS

Table 15.
Delusional Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Delusional disorder	Delusional disorder	
Duration	≥1 month	Does not specify	Typically ≥3 months
Symptoms	Defined by the presence of ≥1 delusions, characterized per specifiers listed below	Presence of single or multiple delusions that are persistent and may or may not be associated with hallucinations	▸ The presence of a delusion or set of related delusions ▸ Delusions vary in content but are stable within individuals, although they can evolve over time ▸ No hallucinations, disorganized thinking, or other evidence of a thought disorder ▸ Affect, speech, and behavior are typically unaffected
Psychosocial consequences	No marked impairment in psychosocial functioning, behavior, or obvious bizarre		
Exclusions	Having a history of meeting the criteria for schizophrenia Other medical condition Substance or medication including withdrawal Other mental illness	Personality disorder Psychosis, psychogenic reaction Schizophrenia	Other medical condition Substance or medication, including withdrawal Other mental illness
Symptom specifiers	**Erotomanic type**: delusion centered on the belief that another person is in love with the affected individual **Grandiose type**: delusions of grandeur prominent **Jealous type:** delusion centered on the belief that a partner is being unfaithful **Persecutory type:** delusions centered on the belief that one is being targeted		Positive symptoms Negative symptoms Depressive mood symptoms Manic mood symptoms Psychomotor symptoms Cognitive symptoms

	Somatic type: delusions focus on body and/or physical symptoms **Mixed type:** no predominant theme of delusions **Unspecified type** **With bizarre content:** if the content of delusions is felt to be wholly unrelated to reality or life experiences and/or completely impossible		
Course specifiers	**First episode, currently in acute episode** **First episode, currently in partial remission** (with partial remission refined as improvement in symptoms such that diagnostic criteria may no longer be fully met) **First episode, currently in full** remission (with full remission defined as lack of any symptoms) **Multiple episodes, currently in acute episode** **Multiple episodes, currently in partial remission** **Multiple episodes, currently in full remission** **Continuous** **Unspecified**		Currently symptomatic In partial remission In full remission

Table 16.
Brief Psychotic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Brief psychotic disorder	Acute and transient psychotic disorder	Acute and Transient psychotic disorder
Duration	Symptoms present 1 day to less than 1 month, followed by a return to baseline function		Occur suddenly and reach maximal severity within 2 weeks. Does not exceed 3 months.
Symptoms	Presence of: ▶ Delusions ▶ Hallucinations ▶ Disorganization of speech ▶ Disorganization of behavior and/or catatonia	Symptoms of schizophrenia, including: ▶ Thought echo ▶ Thought insertion or withdrawal ▶ Thought broadcasting ▶ Delusional perception ▶ Delusions or control, influence, or passivity ▶ Hallucinatory voices ▶ Disordered/disorganized thinking ▶ Negative symptoms May or may not be associated with polymorphic (unstable, frequently changing) delusions, hallucinations, and/or behavioral symptoms	Psychotic symptoms, which may include ▶ Delusions ▶ Hallucinations ▶ Disorganization of thought processes ▶ Perplexity or confusion ▶ Disturbances of affect and mood ▶ Catatonia-like disturbances
Required number of symptoms	≥1 of delusions, hallucinations, or disorganized speech must be present, with or without behavioral symptoms		
Psychosocial consequences			
Exclusions	Culturally sanctioned response/behavior Other medical condition Substance or medication, including withdrawal Other mental illness		Other medical condition Substance or medication, including withdrawal Other mental illness
Symptom specifiers	**With marked stressors** (presence of identifiable event triggering symptoms)	Acute polymorphic psychotic disorder without symptoms of schizophrenia	Positive symptoms Negative symptoms Depressive mood symptoms

	Without marked stressor **With catatonia** defined as the presence of ≥3 of the following: ▸ Reduced psychomotor activity/stupor ▸ Catalepsy (posture induced passively is held against gravity) ▸ Waxy flexibility (resistance to passive motion) ▸ Mutism ▸ Negativism ▸ Posturing ▸ Odd mannerisms ▸ Stereotypic behaviors ▸ Agitation ▸ Grimacing ▸ Echolalia (copying another's speech) ▸ Echopraxia (copying another's movements)	Acute polymorphic psychotic disorder with symptoms of schizophrenia **Acute schizophrenia-like psychotic disorder**–symptoms lasting less than 1 month, though if persistent, the diagnosis should be changed to schizophrenia **Other acute predominantly delusional psychotic disorders** **Other acute and transient psychotic disorder** **Acute and transient psychotic disorder, unspecified**	Manic mood symptoms Psychomotor symptoms Cognitive symptoms
Course specifiers	**With postpartum onset** (onset during pregnancy or within 4 weeks after giving birth)		First/multiple episode(s), currently symptomatic First/multiple episode(s), in partial remission First/multiple episode(s), in full remission First/multiple episode(s), unspecified

Table 17.
Schizophreniform Disorder/Acute and Transient Psychotic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Schizophreniform disorder	**Acute and transient psychotic disorder**	Acute and transient psychotic disorder
Duration	Presence of symptoms for ≥1 month but less than 6 months	Symptoms of a psychotic disorder present for less than 1 month on average. If symptoms persist, the diagnosis should be changed to schizophrenia	Occur suddenly and reach maximal severity within 2 weeks. Does not exceed 3 months.
Symptoms	Presence of: ▸ Delusions ▸ Hallucinations ▸ Disorganized speech ▸ Disorganized behavior and/or catatonia ▸ Negative symptoms (such as flattened affect, lack of motivation)	Symptoms of schizophrenia including: ▸ Thought echo ▸ Thought insertion or withdrawal ▸ Thought broadcasting ▸ Delusional perception ▸ Delusions or control, influence, or passivity ▸ Hallucinatory voices ▸ Disordered/disorganized thinking ▸ Negative symptoms May or may not be associated with polymorphic (unstable, frequently changing) delusions, hallucinations, and/or behavioral symptoms	Psychotic symptoms, which may include ▸ Delusions ▸ Hallucinations ▸ Disorganization of thought processes ▸ Perplexity or confusion ▸ Disturbances of affect and mood Catatonia-like disturbances
Required number of symptoms	≥2 of the above symptoms, among which ≥1 symptom must be delusions, hallucinations, or disorganized speech		
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness		Other medical condition Substance or medication, including withdrawal Other mental illness
Symptom specifiers	**With good prognostic features,** defined as the presence of ≥2 of the following: ▸ Psychotic symptoms occur within 4 weeks of initial behavioral changes		Positive symptoms Negative symptoms Depressive mood symptoms Manic mood symptoms Psychomotor symptoms Cognitive symptoms

	▶ Presence of confusion ▶ High presymptom functioning ▶ Absence of negative symptoms **Without good prognostic features** **With catatonia** defined as the presence of ≥3 of the following: ▶ Reduced psychomotor activity/stupor ▶ Catalepsy (posture induced passively is held against gravity) ▶ Waxy flexibility (resistance to passive motion) ▶ Mutism ▶ Negativism ▶ Posturing ▶ Odd mannerisms ▶ Stereotypic behaviors ▶ Agitation ▶ Grimacing ▶ Echolalia (copying another’s speech) ▶ Echopraxia (copying another’s movements)	
Course specifiers		First/multiple episode(s), currently symptomatic First/multiple episode(s), in partial remission First/multiple episode(s), in full remission First/multiple episode(s), unspecified

Table 18.
Schizophrenia

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Schizophrenia	Schizophrenia	Schizophrenia
Duration	Symptoms present continuously for ≥6 months		≥1 month
Symptoms	Presence of: ▶ Delusions ▶ Hallucinations ▶ Disorganized speech ▶ Disorganized behavior and/or catatonia ▶ Negative symptoms (such as flattened affect, lack of motivation)	Distortions in thinking and/or perception, in addition to negative changes in affect, frequently involving affective blunting. Symptoms may or may not be associated with cognitive dysfunction over time. Common symptoms include: ▶ Thought echo ▶ Thought insertion or withdrawal ▶ Thought broadcasting ▶ Delusional perception ▶ Delusions or control, influence, or passivity ▶ Hallucinatory voices ▶ Disordered/disorganized thinking ▶ Negative symptoms	Disturbance in multiple mental modalities, including ▶ Thinking (delusions, disorganized thought) ▶ Perception (hallucinations) ▶ Self-experience (of being externally controlled) ▶ Cognition (attention, verbal memory, social cognition) ▶ Volition (motivation) ▶ Affect (blunted) ▶ Behavior (bizarre, inappropriate, or purposeless) ▶ Catatonia may be present
Required number of symptoms	≥2 of the above symptoms, among which ≥1 symptom must be delusions, hallucinations, or disorganized speech		
Psychosocial consequences	Level of functioned decreased in ≥1 more major areas of function (e.g., work, interpersonal, self-care)		
Exclusions	Another medical condition Substance/Medication, including withdrawal Other mental illness (i.e., bipolar disorder, schizoaffective disorder if mood episode history is present)	Organic neurologic disease Schizoaffective disorder Epilepsy Psychoactive substances	Another medical condition Substance/medication, including withdrawal

Symptom specifiers	**With catatonia** defined as the presence of ≥3 of the following: ▶ Reduced psychomotor activity/stupor ▶ Catalepsy (posture induced passively is held against gravity) ▶ Waxy flexibility (resistance to passive motion) ▶ Mutism ▶ Negativism ▶ Posturing ▶ Odd mannerisms ▶ Stereotypic behaviors ▶ Agitation ▶ Grimacing ▶ Echolalia (copying another's speech) ▶ Echopraxia (copying another's movements)	**Paranoid schizophrenia** Hebephrenic schizophrenia prominent negative changes in affect, with inappropriate mood, social isolation, unpredictable behavior. Catatonic schizophrenia has prominent psychomotor changes associated with catatonia, such as posturing, odd mannerisms/affect, and stupor vs. agitation. **Undifferentiated schizophrenia** **Residual schizophrenia** chronic illness with evidence of cognitive changes as a result of long-standing disease **Simple schizophrenia** Slow progressive development of changes in behavior and functioning, with affective blunting developing often without clear preceding psychotic symptoms **Other schizophrenia** **Schizophrenia unspecified**	Positive symptoms Negative symptoms Depressive mood symptoms Manic mood symptoms Psychomotor symptoms Cognitive symptoms
Course specifiers	**First episode, currently in acute episode** First episode, currently in partial remission (with partial remission refined as improvement in symptoms such that diagnostic criteria may no longer be fully met) First episode, currently in full remission (with full remission defined as lack of any symptoms) Multiple episodes, currently in acute episode (at minimum		First/multiple episode(s), currently symptomatic First/multiple episode(s), in partial remission First/multiple episode(s), in full remission First/multiple episode(s), unspecified

has history of two episodes, separated by remission)
Multiple episodes, currently in partial remission
Multiple episodes, currently in full remission
Continuous
Unspecified

Table 19.
Schizoaffective Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Schizoaffective disorder	Schizoaffective disorders	Schizoaffective disorder
Duration	Mood episode symptoms must be present for more days than not during the total duration of the illness, though there must also be a period of ≥2 weeks during which delusions or hallucinations are present in the absence of mood episodes.		≥1 month
Symptoms	Presence of symptoms meeting criteria for a major depressive episode or manic episode while also exhibiting symptom criteria for Schizophrenia.	Episode presentation of symptoms of affective disturbance and schizophrenic symptoms that do not meet criteria for either schizophrenia or depressive/manic episodes.	Diagnostic requirements of schizophrenia and a manic, mixed, or moderate or severe depressive episode are met within the same episode of illness, either simultaneously or within a few days of each other.
Exclusions	Another medical condition Other mental illness A substance or medication, including withdrawal		Another medical condition Other mental illness A substance or medication, including withdrawal
Symptom specifiers	**Bipolar type:** manic episode present **Depressive type:** depressive episode present **With catatonia,** defined as the presence of ≥3 of the following: ▶ Reduced psychomotor activity/stupor ▶ Catalepsy (posture induced passively is held against gravity) ▶ Waxy flexibility (resistance to passive motion) ▶ Mutism	**Schizoaffective disorder manic type** Schizoaffective disorder, depressive type Schizoaffective disorder, mixed type Schizoaffective disorder, unspecified Other schizoaffective disorder	Positive symptoms Negative symptoms Depressive mood symptoms Manic mood symptoms Psychomotor symptoms Cognitive symptoms

	▶ Negativism ▶ Posturing ▶ Odd mannerisms ▶ Stereotypic behaviors ▶ Agitation ▶ Grimacing ▶ Echolalia (copying another's speech) ▶ Echopraxia (copying another's movements)
Course specifiers	**First episode, currently in acute episode** First episode, currently in partial remission (with partial remission refined as improvement in symptoms such that diagnostic criteria may no longer be fully met) First episode, currently in full remission (with full remission defined as lack of any symptoms) Multiple episodes, currently in acute episode (history of ≥2 episodes, separated by remission) Multiple episodes, currently in partial remission Multiple episodes, currently in full remission Continuous Unspecified

3. BIPOLAR AND RELATED DISORDERS

Table 20.
Bipolar 1 Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Bipolar 1 disorder	Bipolar affective disorder ***NOTE***: ICD-10 does not distinguish between Bipolar 1 and 2, requiring only a history of discrete episodes of mania, hypomania, and/or depression, with episodes demarcated by switches in mood/affect.	Bipolar 1 disorder
Duration	Manic episode: 1 week+ Hypomanic episode: 4 days+ Major depressive episode: 2 weeks+	Manic episode: 1 week+ Hypomanic episode: 4 days+ Major depressive episode: 2 weeks+	1 week+
Symptoms	**Manic or hypomanic episodes** ▶ abnormally ↑ or irritable mood (required) ▶ grandiose thoughts ▶ ↓ need for sleep ▶ pressured speech ▶ racing and expansive thoughts ▶ distractibility ▶ hyperactivity ▶ impulsivity/high-risk activities **Depressive episodes**—Similar to that for major depressive disorder	History of episodes of mania, hypomania, and/or depression, switches in mood/affect. **Mania** ▶ abnormally ↑ or irritable mood (required) ▶ Abnormally elevated or irritable mood ▶ ↑ activity ▶ ↑ talkativeness ▶ flight of ideas/racing thoughts ▶ social disinhibition ▶ ↓ need for sleep ▶ grandiose thoughts ▶ distractibility ▶ impulsivity/recklessness ▶ hypersexuality **Hypomania** ▶ abnormally ↑ mood (required) ▶ psychomotor agitation ▶ ↑ talkativeness ▶ Poor concentration/distractibility ▶ ↓ need for sleep ▶ hypersexuality ▶ impulsivity/ or ↑ spending ▶ over-familiarity **Depressive episode** ▶ Depressed mood ▶ Loss of interest or pleasure	**Manic episode** ▶ Extreme mood state (euphoria, irritability of expansiveness) ▶ Increased activity ▶ Talkativeness ▶ Flight of ideas ▶ Increased self-esteem ▶ Decreased need for sleep ▶ Distractibility ▶ Impulsivity ▶ Increased libido, sociability, or goal-directed activity **Mixed episode** Several prominent manic and several prominent depressive symptoms

		► Decreased energy ***Additional symptoms:*** ► Low self-esteem ► Excessive guilt or shame ► Recurrent thoughts of death or suicide ► Poor concentration ► Psychomotor changes ► Sleep disturbance ► Change in appetite and/or weight	
Required number of symptoms	≥1 manic episode Abnormally ↑ or irritable mood (required 3 of the other symptoms (4 if irritable mood)	**Bipolar affective disorder, current episode hypomanic** Hypomania (elevated mood and ≥3 symptoms History of a prior affective episode (manic, hypomanic, depressed, mixed) **Bipolar affective disorder, current episode manic** Mania (elevated or irritable mood and ≥ 3symptoms (4 if irritable mood) History of prior affective episodes (manic, hypomanic, depressed, mixed) **Bipolar Disorder, current episode mild/moderate/severe depression** Depressive episode (2/3 of the first 2 symptoms and 2 more of the additional symptoms) History of prior affective episodes (manic, hypomanic, depressed, mixed) **Bipolar disorder, current episode mixed** A mixture or rapid alternation of hypomanic, manic, and/or depressive symptoms (defined above) History of prior affective episodes (manic,	≥1 manic or mixed episode consisting of first 2 symptoms listed above, plus several of the other symptoms listed.

		hypomanic, depressed, mixed)	
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness	Psychoactive substance use Another mental disorder	Other medical condition Substance or medication, including withdrawal
Psychosocial consequences	***Manic episode:*** impaired functioning or needing hospitalization Hypomanic episode: no impairment or need for hospitalization Depressive episode: marked distress and/or psychosocial impairment		Impairment in ≥1 functional domain, or requiring intensive treatment (hospitalization), or includes hallucinations or delusion
Symptom specifiers	**Mild**—2 symptoms Moderate—3 symptoms Moderate–Severe—4–5 symptoms Severe—4–5 symptoms + motor agitation With mixed features ► Either a depressive or manic/hypomanic episode ► Additional symptoms of depressive or manic/hypomanic period (not full criteria) With rapid cycling ≥4 mood episodes in 1 year ≥2-month period of partial/full remission between episodes With melancholic features Similar to that for major depressive disorder With atypical features Similar to that for major depressive disorder With anxious distress:	**Current episode hypomanic** Current episode manic without psychotic symptoms Current episode manic with psychotic symptoms **Mood congruent** **Mood incongruent** **Current episode of depression** Current episode mixed Currently in remission—no symptoms. History of previous episodes. ***With somatic syndrome ≥4 of the following:*** ► ***Loss of interest or pleasure*** ► ***Lack of emotional reactivity*** ► ***Early morning awakening*** ► ***Depression gets worse in the mornings*** ► ***Psychomotor changes*** ► ***Loss of appetite*** ► ***Weight loss*** ► ***Loss of sexual drive***	**Current/Most recent episode:** Manic Hypomanic Depressive Mixed With/without psychotic symptoms With prominent anxiety symptoms With panic attacks With Seasonal pattern With rapid cycling Associated with pregnancy, childbirth, or the puerperium

	≥2 symptoms among the following: Feeling tense Restlessness Difficulty with concentration due to worrying XS fear without cause Fear of loss of control		
Course specifiers	**With peripartum onset:** episode occurs during pregnancy or within 4 weeks after delivery With seasonal pattern: pattern present for ≥2 years		
Severity specifiers		**For the current episode of depression** Mild: 4 symptoms Moderate: 6 symptoms, including ≥2 among loss of pleasure, depressed mood, and low energy Severe depression ≥8 symptoms, including all loss of pleasure, depressed mood, and low energy Without psychotic symptoms With psychotic symptoms Mood congruent Mood incongruent	**Mild** **Moderate** **Severe** **Unspecified severity**

Table 21.
Bipolar 2 Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Bipolar 2 disorder	Bipolar affective disorder **_*NOTE:_** ICD-10 does not distinguish between Bipolar 1 and 2, requiring only a history of discrete episodes of mania, hypomania, and/or depression, with episodes demarcated by switches in mood/affect.	Bipolar type II disorder
Duration	**Hypomanic episode:** lasting ≥4 days, with symptoms, present most of the day for most days **Major depressive episode**: symptoms lasting ≥2 weeks	Manic episode: lasting ≥1 week Hypomanic episode: lasting ≥4 days Major depressive episode: symptoms lasting ≥2 weeks	Several days
Symptoms	Hypomanic episodes involve abnormally elevated or irritable mood with: ▶ Grandiosity ▶ Decreased need for sleep ▶ Pressured speech ▶ Racing thoughts with expansiveness of thought ▶ Distractibility ▶ Goal-directed or non–goal-directed hyperactivity ▶ Engagement in high-risk activity or increased impulsivity Depressive episode involving depressed mood and/or decreased interest/pleasure with: ▶ Unintentional weight loss or change in appetite ▶ Disruption of sleep ▶ Psychomotor changes ▶ Loss of energy	History of discrete episodes of mania, hypomania, and/or depression, with episodes demarcated by switches in mood/affect. Episodes are defined as follows: **Mania:** abnormally elevated or irritable mood with ≥3 of the following (≥4 if mood irritable) ▶ Increased activity ▶ Increased talkativeness ▶ Flight of ideas/racing thoughts ▶ Social disinhibition ▶ Decreased need for sleep ▶ Grandiosity ▶ Distractibility ▶ Increased recklessness and impulsivity ▶ Hypersexuality **Hypomania:** abnormally elevated mood with ≥3	**Hypomanic Episode** Elevation of mood or irritability outside of the usual mood range Increased activity Talkativeness Flight of ideas Increased self-esteem Decreased need for sleep Distractibility Impulsivity Increased libido, sociability, or goal-directed activity Depressive episode ***See depressive disorders***

	▸ Difficulty with concentration and/or attention ▸ Excessive feelings of worthlessness and/or guilt ▸ Recurrent thoughts of death, including passive and/or active suicidal ideation	of the following: ▸ Psychomotor agitation ▸ Increased talkativeness ▸ Poor concentration or increased distractibility ▸ Decreased need for sleep ▸ Hypersexuality ▸ Impulsivity or increased spending ▸ Over-familiarity Depressive episode: 4 total symptoms from among the following: ≥2: ▸ Depressed mood ▸ Loss of interest or pleasure ▸ Decreased energy Additional symptoms: ▸ Low self-esteem ▸ Excessive guilt or shame ▸ Recurrent thoughts of death or suicide ▸ Poor concentration ▸ Psychomotor changes ▸ Sleep disturbance ▸ Change in appetite and/or weight	
Required number of symptoms	Lifetime history of a hypomanic episode involving elevated/irritable mood with ≥3 of the above symptoms (4 if mood irritable) Lifetime history of a depressive episode with ≥5 of the above symptoms, which must include ≥1 depressed mood or loss of pleasure	**Bipolar affective disorder, current episode hypomanic**—meets the above criteria for hypomania (elevated mood with ≥3 listed symptoms) and has a history of prior affective episodes (manic, hypomanic, depressed, mixed) **Bipolar affective disorder, current episode manic**—meets above criteria for mania (elevated or irritable mood with ≥3–4 listed symptoms) and	≥1 hypomanic episode ≥1 depressive episode

		has a history of prior affective episodes (manic, hypomanic, depressed, mixed) **Bipolar disorder, current episode mixed**—mixture or rapid alteration in hypomanic, manic, and/or depressive symptoms, with symptoms, present most of the time for ≥2 weeks, in addition to having a history of prior affective episodes (manic, hypomanic, depressed, mixed) **Bipolar disorder, current episode mild/moderate/severe depression**—current symptoms meet criteria for a depressive episode, in addition to having a history of prior affective episodes (manic, hypomanic, depressed, mixed)	
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness	Psychoactive substance use Another mental disorder	H/o manic episodes Other medical condition Effect of substance or medication
Psychosocial consequences	Hypomanic episode: NO marked impairment in functioning and NO hospitalization necessitated Depressive episode: marked distress and/or impairment in psychosocial functioning		Does not cause marked functional impairment
Symptom specifiers	**Current episode:** ***Depressed*** ***Hypomanic*** Additional specifiers to be added to the above:	**Current episode hypomanic** Current episode manic without psychotic symptoms	**Current/most recent episode** Hypomanic Depressive With/without psychotic symptoms

With anxious distress: including ≥2 symptoms among feeling tense, restlessness, difficulty with concentration due to worrying, excessive fear without an identifiable cause, fear of loss of control With mixed features—full criteria met for either depressive or manic/hypomanic episode, with additional symptoms of depressive or manic/hypomanic period without also meeting full criteria With rapid cycling—the presence of ≥4 mood episodes over the course of a 12-month period, with ≥2-month period of partial/full remission between mood episodes With melancholic features—involving loss of pleasure, reduced reactivity with ≥3 of the following: ▸ Depressed mood/despair ▸ Depressive symptoms worse in the morning ▸ Early morning awakening ▸ Psychomotor changes ▸ Loss of appetite/weight loss ▸ Excessive guilt **With atypical features:** during most days, an individual is experiencing mood reactivity with ≥2 symptoms of: ▸ Weight gain or increased appetite	Current episode manic with psychotic symptoms ▸ ***With mood congruent psychotic symptoms*** ▸ ***With mood incongruent psychotic symptoms*** **Current episode moderate or mild depression**—≥4 (mild) or 6 (moderate) symptoms present, including ≥2 among loss of pleasure, depressed mood, and low energy ***With somatic syndrome—≥4 of the following:*** ▸ ***Loss of interest or pleasure*** ▸ ***Lack of emotional reactivity*** ▸ ***Early morning awakening*** ▸ ***Depression gets worse in the mornings*** ▸ ***Psychomotor changes*** ▸ ***Loss of appetite*** ▸ ***Weight loss*** ▸ ***Loss of sexual drive*** ***Without somatic syndrome*** **Current episode severe depression without psychotic symptoms**—≥8 symptoms present, including all 3 of loss of pleasure, depressed mood, and low energy **Current episode severe depression with psychotic symptoms**—≥8 symptoms present, including all 3 of loss of pleasure, depressed mood, and	With prominent anxiety symptoms With panic attacks With seasonal pattern With rapid cycling Associated with pregnancy, childbirth, or the puerperium

	▸ Excessive sleeping ▸ Leaden paralysis ▸ A pattern of heightened interpersonal rejection sensitivity **With mood-congruent psychotic features**—mood-congruent delusions or hallucinations With mood-incongruent psychotic features—mood-incongruent delusions or hallucinations With catatonia—catatonic features present during most of the mood episode	low energy, occurring either ▸ ***With mood-congruent psychotic symptoms*** ▸ ***With mood incongruent psychotic symptoms*** **Current episode mixed** Currently in remission—does not currently meet criteria for an affective episode despite a prior history of a manic or hypomanic episode	
Course specifiers	**With peripartum onset**—full criteria for a mood episode are met during pregnancy or within 4 weeks of delivery With seasonal pattern (for ≥1 type of episode)—pattern present for ≥2 years In partial remission In full remission		
Severity specifiers	**Mild**—few symptoms in excess of minimum diagnostic criteria, with no to minimal impairment and/or distress **Moderate**—moderate symptomatology and impairment **Severe**—symptoms in marked excess of minimum diagnostic requirement, with marked distress and impairment in functioning	See the above criteria for mild, moderate, and severe depression	Mild Moderate Severe
Comments	**Note**: A diagnosis of		

Bipolar 2 cannot be made if there is a history of a manic episode, in which case a diagnosis of Bipolar 1 disorder would be made.

Table 22.
Cyclothymic Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Cyclothymic disorder	Cyclothymia	Cyclothymic disorder
Duration	Symptoms lasting ≥2 years (≥1 year in children), with depressive and hypomanic symptoms present for ≥half the time.	2-year period involving periods of depression and elevated mood, without ever meeting criteria for a depressive episode or manic episode.	≥2 years
Symptoms	Hypomanic episodes involve abnormally elevated or irritable mood with: ▶ Grandiosity ▶ Decreased need for sleep ▶ Pressured speech ▶ Racing thoughts with expansiveness of thought ▶ Distractibility ▶ Goal-directed or non–goal-directed hyperactivity ▶ Engagement in high-risk activity or increased impulsivity Depressive episode involving depressed mood and/or decreased interest/pleasure with: ▶ Unintentional weight loss or change in appetite ▶ Disruption of sleep ▶ Psychomotor changes ▶ Loss of energy ▶ Difficulty with concentration and/or attention ▶ Excessive feelings of worthlessness and/or guilt ▶ Recurrent thoughts of death, including passive and/or active suicidal ideation	Periods of depressed mood involving: ▶ Low energy ▶ Insomnia ▶ Low self-esteem ▶ Poor concentration ▶ Social withdrawal ▶ Loss of pleasure/enjoyment ▶ Reduced talkativeness ▶ Excessive pessimism Period of elevated mood involving: Increased energy ▶ Decreased need for sleep ▶ Inflated self-esteem ▶ Increased creativity ▶ More talkative ▶ More socially outgoing ▶ Hypersexuality and increased participation in pleasurable activities ▶ Grandiosity	Mood instability, with numerous hypomanic (see Bipolar type II disorder) and depressive periods (See depressive disorders). Children instead may show irritability. Hypomanic periods may or may not have been severe or long enough to meet criteria for hypomanic episode.
Required number	Symptoms of hypomania	≥3 symptoms during	

of symptoms	and depression present without ever meeting the full criteria for a depressive episode or manic episode	periods of mood elevation without meeting the full criteria for a manic episode ≥3 symptoms during periods of low mood without meeting the full criteria for a depressive episode	
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illnesses (Bipolar 1 or 2)	Bipolar affective disorder (though a history of bipolar affective disorder may be present)	Other medical condition Substance or medication, including withdrawal Other mental illness
Psychosocial consequences	Marked distress or impairment in areas of functioning		Marked distress or impairment in areas of functioning
Symptom specifiers	**With anxious distress:** including ≥2 symptoms among feeling tense, restlessness, difficulty with concentration due to worrying, excessive fear without an identifiable cause, fear of loss of control		
Course specifiers		**Early onset:** before 20s **Late onset:** onset between ages 30 and 50	
Severity specifiers	**Mild**—2 symptoms Moderate—3 symptoms Moderate–severe—4 or 5 symptoms Severe—4–5 symptoms with motor agitation		

4. DEPRESSIVE DISORDERS

Table 23.
Disruptive Mood Dysregulation Disorder/Other Persistent Mood [Affective] Disorders/Oppositional Defiant Disorder With Chronic Irritability-Anger, Unspecified

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Disruptive mood dysregulation disorder	Other persistent mood (affective) disorders	Oppositional defiant disorder with chronic irritability-anger, unspecified
Duration	Outbursts occurring ≥3 times per week for a duration of ≥12 months, during which there was no period >3 months during which all symptom criteria were not met.		
Symptoms	Recurrent verbal or physical outbursts out of proportion to the triggering event Not consistent with expected response per developmental level Mood between outbursts is irritable or mad most of the time		
Required number of symptoms	All of the above		
Exclusions	< age 6 > age 10 Other medical condition Substance or medication, including withdrawal Other mental illness Cannot co-occur with: ▸ Oppositional defiant disorder ▸ Intermittent explosive disorder ▸ Bipolar disorder		

Table 24.
Major Depressive Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Major depressive disorder	Major depressive episode	Major depressive disorder
Duration	2-week period	N/A	
Symptoms	Dysphoria/depressed mood Anhedonia Change in weight or appetite Change in sleep pattern Psychomotor disturbance: agitation or retardation Fatigue/decreased energy Depressive ruminations: worthlessness, guilt Poor concentration Recurrent thoughts of death, suicidal ideation, or a plan	Lowered mood Reduced energy Decreased activity Reduced capacity for enjoyment Reduced interest Reduced concentration Fatigue after even minimal effort Disturbed sleep/early morning awakening Disturbed appetite/weight loss Poor self-esteem Poor self-confidence Guilt or worthlessness Mood unreactive to circumstances Anhedonia Worse symptoms in the AM Psychomotor disturbance: agitation or retardation Loss of libido	**Affective cluster** ▸ Depressed mood ▸ Anhedonia **Cognitive-behavioral cluster** ▸ Poor concentration ▸ Low self-worth or inappropriate guilt ▸ Hopelessness ▸ Recurrent thoughts of death, suicidal ideation, or a plan **Neurovegetative cluster** ▸ Disrupted sleep ▸ Change in appetite ▸ Psychomotor agitation/retardation ▸ Fatigue/decreased energy
Required number of symptoms	5 (1 of these has to be symptom 1 or 2 listed above)	N/A	5 (≥1 from the affective cluster)
Psychosocial consequences	Distress or impaired functioning (social, occupational, or other significant areas)	Depends on severity	Significant impairment in ≥1 domain
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness History of: ▸ Manic episode ▸ Hypomanic episode	Adjustment disorder Conduct disorder Recurrent depressive disorder (which is considered a separate diagnosis)	Bereavement Mixed episode bipolar disorder Other medical consideration Effects of substance or medication
Symptom specifiers	**With anxious distress** ▸ 2+ symptoms of anxiety	Depressive reaction Psychogenic depression Reactive depression	Mild Moderate Severe

With mixed features

- 3+ manic/hypomanic symptoms occurring during the depressive episode (if occur independently, diagnose bipolar disorder)

With melancholic features

- Loss of pleasure or reactivity to pleasure
- 3+ of the following
- Severe depression/despair
- Mood worse in AM
- Early morning awakening
- Psychomotor disturbance
- Anorexia/weight loss
- Guilt

With atypical features

- Mood reactivity
- ≥2:
  - Increased appetite/weight
  - Hyposomnia
  - Leaden paralysis
  - Rejection sensitivity
  - With mood-congruent psychotic features
  - With mood-incongruent psychotic features

With psychotic features: symptoms occur only during a depressive episode. Also specify if delusions are mood congruent or incongruent.

With catatonia

- Must be present during most depressive episodes

With peripartum onset

- With seasonal pattern

Unspecified

With/without psychotic symptoms

With prominent anxiety symptoms

With panic attacks

Current depressive episode persistent

Current depressive episode with melancholia

With seasonal pattern

Associated with pregnancy, childbirth, or puerperium

	▶ Usually occurs during a specific season		
Course specifiers	**In partial remission:** full criteria no longer met **In full remission**: no symptoms of an episode for the past 2 months	**Recurrent depressive disorder** (coded as a separate disorder). Repeated episodes of the above symptoms. No mania.	**In partial remission** In full remission
Severity specifiers	**Mild:** minimum # symptoms **Moderate:** between mild and severe **Severe:** # of symptoms, the severity of symptoms, or dysfunction is much more than required for a diagnosis	**Mild** ▶ 2–3 symptoms ▶ Functions normally despite the distress **Moderate** ▶ 4+ symptoms. ▶ Difficulty with functioning **Severe** ▶ Several symptoms marked and distressing ▶ Loss of self-esteem/feels worthless and guilty ▶ Suicidal ideation/acts ▶ Somatic symptoms of depression **Severe with psychotic symptoms** ▶ As above but with psychosis Other ▶ Atypical depression ▶ Single episodes of “masked depression” Unspecified	

Table 25.
Persistent Depressive Disorder/Dythymia/Dysthymic Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Persistent depressive disorder	Dysthymia	Dysthymic disorder
Duration	Symptoms continuously present for ≥2 years No time during the affected period of no symptoms for >2 months		≥2 years
Symptoms	Depressed mood most of the time on a majority of days over a period of ≥2 years (NOTE: in children, irritability may be present, and duration must be for a period of ≥1 year) During the depressed period: ▶ Abnormal appetite (↑ or ↓) ▶ Abnormal sleep (↑ or ↓) ▶ ↓ energy ▶ ↓ self-esteem ▶ ↓ concentration or ↑ indecisiveness ▶ Hopelessness	Chronically depressed mood that does not meet criteria for a depressive episode, though criteria may have been met in the past.	Persistent depressed mood Additional symptoms of depression. May include: ▶ Anhedonia or decreased interest ▶ Poor concentration ▶ Low self-worth or inappropriate guilt ▶ Hopelessness ▶ Disturbed sleep ▶ Change in appetite ▶ Fatigue or low energy
Required number of symptoms	≥2 bulleted symptoms		
Psychosocial consequences	Marked distress and/or functional impairment		Marked distress and/or functional impairment
Exclusions	This cannot occur with a history of bipolar disorder Other medical condition Substance or medication, including withdrawal Other mental illness	Result of: ▶ Anxiety ▶ Depression ▶ Bereavement ▶ Schizophrenia	A depressive episode in the first 2 years of the disorder History of manic, mixed, or hypomanic episodes Other medical condition Substance or medication, including withdrawal
Symptom specifiers	**With pure dysthymic syndrome:** criteria for the depressive episode not met in the last 2 years		

With persistent major depressive episode: has met diagnostic criteria for a depressive episode for an entire 2-year period

With intermittent major depressive episode, with current episode: though may currently meet criteria for major depressive disorder, has had a period of ≥8 weeks during the prior 2-year period during which symptoms were below the diagnostic threshold for a depressive episode.

With intermittent major depressive episode, without current episode: does not currently meet criteria for a depressive episode, though has experienced ≥1 depressive episode over the course of the prior 2 years.

With anxious distress: including ≥2 symptoms among feeling tense, restlessness, difficulty with concentration due to worrying, excessive fear without an identifiable cause, fear of loss of control.

With mixed features—full criteria met for either depressive or manic/hypomanic episode, with additional symptoms of depressive or manic/hypomanic period without also meeting full criteria.

	With melancholic features—involving loss of pleasure, reduced reactivity with ≥3 of the following: ► Depressed mood/despair ► Depressive symptoms worse in the morning ► Early morning awakening ► Psychomotor changes ► Loss of appetite/weight loss ► Excessive guilt
	With atypical features: during most days, an individual is experiencing mood reactivity with ≥2 symptoms: ► Weight gain or increased appetite ► Excessive sleeping ► Leaden paralysis ► A pattern of heightened interpersonal rejection sensitivity
	With mood-congruent psychotic features – mood-congruent delusions or hallucinations
	With mood-incongruent psychotic features– mood-incongruent delusions or hallucinations
Course specifiers	**With peripartum onset**—full criteria for a mood episode are met during pregnancy or within 4 weeks of delivery
Severity specifiers	**Mild**—few symptoms in excess of minimum diagnostic criteria, with no to minimal

impairment and/or distress

Moderate—moderate symptomatology and impairment

Severe—symptoms in marked excess of minimum diagnostic requirement, with marked distress and impairment in functioning

Table 26.
Premenstrual Dysphoric Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Premenstrual dysphoric disorder		Premenstrual dysphoric disorder
Duration	Symptoms present in the week leading up to menses and improve with the onset of menses while resolving or nearly resolving in the week following menses. Symptoms should be present over the course of 2 menstrual cycles.		
Symptoms	Mood symptoms: ▶ Mood swings ▶ Irritability ▶ Depressed mood or low self-esteem ▶ Anxiety Other symptoms: ▶ Decreased interest in normally pleasurable activities ▶ Reduced concentration ▶ Fatigue ▶ Change in appetite ▶ Change in sleep (increased or decreased) ▶ Feeling of being overwhelmed ▶ Physical symptoms (bloating, musculoskeletal pain, breast discomfort)		Symptoms associated with most menstrual cyclesaffective symptoms ▶ Mood lability ▶ Irritability ▶ Depression ▶ Anxiety Somatic or cognitive symptoms: ▶ Lethargy ▶ Joint pain ▶ Overeating ▶ Hypersomnia ▶ Breast tenderness ▶ Extremity swelling ▶ Difficulty concentrating ▶ Forgetfulness
Required number of symptoms	≥1 Mood symptoms ≥1 Other symptom		≥ Affective symptom
Psychosocial consequences	Distress or functional impairment		Distress or functional impairment
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness		Other medical condition Substance or medication, including withdrawal Other mental illness

5. ANXIETY DISORDERS

Table 27.
Separation Anxiety Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Separation anxiety disorder	Separation anxiety disorder of childhood	Separation anxiety disorder
Duration	Symptoms persist over ≥4 weeks in children or ≥6 months in adults		≥ Several months
Symptoms	Recurrent distress on separation or anticipated separation from the attachment figure Worry of losing attachment figure or of their being harmed Worry about unexpected events that will result in forced separation from the attachment figure Hesitance to leave home or be away from the attachment figure Fear of being alone in a setting without an attachment figure Difficulty sleeping without being near attachment figure Nightmares related to fear of separation Physical symptoms occurring during separation or anticipation of separation	Fear of separation causes anxiety Anxiety arises during early childhood Severity is out of proportion and of longer duration than would be expected by normal developmental/age standards	Marked/ excessive fear/anxiety about separation from an attachment figure Symptoms vary by developmental level but may include: Persistent thoughts of being separated due to some harm/unexpected event Reluctance/refusal to go to school or work Excess distress related to a separation from an attachment figure Recurrent nightmares about separation Physical symptoms associated with separation (gastrointestinal, headaches)
Required number of symptoms	≥3 of the symptoms above meeting the required duration		
Psychosocial consequences	Distress or psychosocial impairment	Impaired functioning	Distress or psychosocial impairment
Exclusions	Another mental disorder		Another mental disorder
Comments		Note: classified under (Emotional disorders with onset specific to childhood)	

Table 28.
Selective Mutism

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Selective mutism	Elective mutism	Selective mutism
Duration	≥1 month (not exclusively during the first month of school)		≥1 month (not exclusively during the first month of school)
Symptoms	Persistent lack of speech in particular social situations despite being able to speak in other situations.	Presence of selectively speaking related to specific emotions, resulting in a child that has previously demonstrated the ability to speak not speaking in particular social situations. It may be associated with other symptoms of anxiety, social withdrawal, and sensitivity	Selectivity in speaking with language competence in some situations but not others.
Psychosocial consequences	Psychosocial impairment		Psychosocial impairment
Exclusions	Another mental disorder Communication disorder Lack of proficiency with the spoken language in a given situation	Pervasive developmental disorders Developmental disorders of speech and language Transient mutism as part of separation anxiety in young children	Another mental disorder
Comments		Note: classified as a disorder of social functioning with onset specific to childhood and adolescence	

Table 29.
Specific Phobia

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Specific phobia	Specific (isolated) phobias	Specific phobia
Duration	Persistent and lasting ≥6 months		Persistent and lasting ≥ several months
Symptoms	Fear and/or anxiety related to particular objects or situations Exposure to that object or situation often results in immediate fear and/or anxiety Avoidance of the feared object or situation Fear/anxiety response is out of proportion to the given context	Phobias are restricted to highly specific situations, objects, or activities. Exposure to such situations, objects, or activities can result in panic.	Marked and excessive fear due to the occurrence or anticipation of a specific object or situation Feared object/situation is avoided or endured with intense anxiety
Required number of symptoms	All of the above		
Psychosocial consequences	Distress or psychosocial impairment		Distress or psychosocial impairment
Exclusions	Other mental disorder		Other mental disorder
Symptom specifiers	Animals Natural environment Blood-injection-injury Fear of blood Fear of injections and transfusions Fear of other medical care Fear of injury Situational Other		With Panic attack
Comments		Other phobic anxiety disorder Phobic anxiety disorder, unspecified	

Table 30.
Social Anxiety Disorder/Social Phobias

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Social anxiety disorder (social phobia)	Social phobias	Social anxiety disorder
Duration	Persistent and lasting typically ≥6 months		Persistent and lasting ≥several months
Symptoms	Anxiety related to social situations, given the fear of being judged or scrutinized by others Fear of allowing anxiety to show and be noticed by others, be perceived negatively, or will lead to embarrassment or rejection Avoidance of feared social situations Anxiety and fear are out of proportion to the social context	Fear of being judged or scrutinized by other people Fear of judgment or embarrassment leads to avoidance of social situations It can be associated with low self-esteem Panic attacks may develop	Anxiety occurs in ≥1 social interaction or while performing Fear or concern of being negatively evaluated by others The feared situation is avoided or endured with intense anxiety
Required number of symptoms	All of the above		
Psychosocial consequences	Marked distress and/or impairment		Marked distress and/or impairment
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness		Other mental illness
Symptom specifiers	**Performance only** (anxiety related specifically to publicly performing or speaking)		

Table 31.
Panic Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Panic disorder	Panic disorder (episodic paroxysmal anxiety)	Panic disorder
Duration	≥1 panic attack must be followed by ≥1 month of worry symptoms as defined below		
Symptoms	Recurrent and unpredictable panic attacks, which can involve: ▶ Palpitations ▶ Sweating ▶ Tremulousness ▶ Feeling short of breath ▶ Chest discomfort ▶ Nausea or GI discomfort ▶ Dizziness ▶ Chills or feeling flushed ▶ Numbness or tingling ▶ Derealization or depersonalization ▶ Fear of losing control ▶ Fear of impending death Persistent symptoms between attacks, including: ▶ Persistent fear of having another attack and of its consequences ▶ Maladaptive behaviors to avoid situations perceived to be related to the development of a panic attack	Presence of recurrent “attacks” of severe anxiety and panic Panic attacks are not associated with a particular situation or setting and are thus unpredictable Prominent physiologic symptoms are present, such as chest pain, palpitations, difficulty breathing, sweating, dizziness Additional symptoms may include depersonalization, derealization, fear of losing control, fear of dying	Recurrent panic attacks ≥some of which are unexpected. Symptoms may include (but not be limited to): ▶ Palpitations ▶ Sweating ▶ Tremulousness ▶ Feeling short of breath ▶ Chest discomfort ▶ Nausea or GI discomfort ▶ Dizziness ▶ Chills or feeling flushed ▶ Numbness or tingling ▶ Derealization or depersonalization ▶ Fear of losing control ▶ Fear of impending death Attacks are followed by persistent concern or worry, typically for several weeks
Required number of symptoms	≥4 more symptoms of a panic attack present during a discrete, abrupt, and unpredictable episode, with ≥1 panic episode followed by >1 month of fear of the next		

	attack, avoidance, or both.		
Psychosocial consequences			Distress or psychosocial impairment
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness	Result of: ▶ Other mental illnesses presenting with panic-like symptoms	Other medical condition Substance or medication, including withdrawal Other mental illness
Comments	A separate diagnosis for “Panic Attack Specifier” is present and can be used for individuals experiencing discrete panic attacks (as defined by the above symptoms) occurring without additional criteria for panic disorder and occurring in the context of another psychiatric disorder		

Table 32.
Agoraphobia

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Agoraphobia	Agoraphobia	Agoraphobia
Duration	Persistent and lasting ≥6 months		Persistent and lasting ≥ several months
Symptoms	Fear or anxiety arising from: ▶ Public transportation ▶ Open spaces ▶ Confined spaces ▶ Being in a line or in a crowd ▶ Being alone outside of the home Avoidance of the above situations due to fear of developing panic-related symptoms or requiring companionship to endure said feared situations Fear and avoidance are out of proportion to the actual threat of the avoided situation	Fear of leaving home, going to public and/or crowded places or traveling alone Frequently associated with panic disorders Presence of a prominent pattern of avoidance of anxiety-provoking situations	Marked and excessive fear or anxiety occurring in (or in anticipation of) situations where escape would be difficult Fear of negative consequences of such situations (i.e., panic) Situations avoided, endured with anxiety, or only attempted under specific circumstances, such as with a companion
Psychosocial consequences	Distress or psychosocial impairment		Distress or psychosocial impairment
Exclusions	Other medical condition Other mental illness		Other mental illness
Comments	Diagnosis of agoraphobia can be made without the presence of a panic disorder.		

Table 33.
Generalized Anxiety Disorder

	DSM-5/DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Generalized anxiety disorder	Generalized anxiety disorder	Generalized anxiety disorder
Duration	≥6 months, more days than not		Persists for several months, more days than not
Symptoms	Excessive anxiety or worry about a wide range of situations Difficulty controlling or managing worry Symptoms of anxiety/worry, including: ▶ Restlessness ▶ Fatigue ▶ Poor concentration ▶ Irritability ▶ Muscle tension ▶ Sleep disturbance	Persistent and generalized anxiety Anxiety not associated with a particular object, event, or situation Predominant symptoms may include worry, shaking, muscle tension, sweating, lightheadedness, palpitations, and GI symptoms	Free-floating anxiety or excessive worry about negative events Anxiety is accompanied by symptoms such as: ▶ Motor restlessness, muscle tension ▶ Autonomic overactivity ▶ Feeling nervous, restless, or on edge ▶ Difficulty concentrating ▶ Irritability ▶ Insomnia
Required number of symptoms	Both excessive worry and difficulty controlling worry must be present, in addition to ≥3 of the associated symptoms of anxiety/worry		
Psychosocial consequences	Distress or psychosocial impairment		Distress or psychosocial impairment
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness		Other medical condition Substance or medication, including withdrawal Other mental illness

6. OBSESSIVE-COMPULSIVE AND RELATED DISORDERS

Table 34.
Obsessive-Compulsive Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Obsessive-compulsive disorder	Obsessive-compulsive disorder	Obsessive-compulsive disorder
Duration	Persistent		Persistent
Symptoms	Obsessions (intrusive thoughts that accompany behaviors and that individuals may attempt to ignore or counteract through action) Compulsions (repetitive actions, commonly driven by an obsessive thought, performed to relieve anxiety or fear)	Recurrent obsessional thoughts and/or compulsive acts ▸ Obsessions: ideas, images, or impulses occurring recurrently ▸ Compulsions: repeated, stereotyped behaviors meant to neutralize a thought/anxiety or prevent a perceived adverse event Attempts at resisting thoughts and behaviors are frequently unsuccessful and cause worsening anxiety	Obsessions: repetitive/persistent thoughts/images/impulses, experienced as intrusive and unwanted, associated with anxiety Compulsions: repetitive behaviors/rituals in response to obsessions
Required number of symptoms	Obsessive thoughts, compulsions, or both	Presence of obsessions, compulsions, or both	
Psychosocial consequences	Symptoms and behaviors consume ≥1 hour per day or cause distress or psychosocial impairment		Symptoms and behaviors consume ≥1 hour per day or cause distress or psychosocial impairment
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness	Obsessive-compulsive personality disorder	Other medical condition Substance or medication, including withdrawal
Symptom specifiers	Tic related: current or prior tic disorder With good or fair insight With poor insight With absent insight and/or delusional beliefs	Predominantly obsessional thoughts or ruminations Predominantly compulsive acts (obsessional rituals) Mixed obsessional thoughts and acts	Fair to good insight Poor to absent insight

Table 35.
Body Dysmorphic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Body dysmorphic disorder	Body dysmorphic disorder	Body dysmorphic disorder
Symptoms	Preoccupation with physical appearance and subjective flaws Presence of repetitive behaviors at some point during the disorder ▸ Checking ▸ Grooming ▸ Skin picking ▸ Recurrent thoughts/perseveration on comparing physical appearance to that of others	Excessive preoccupation with image or appearance	Persistent preoccupation with ≥1 perceived defects or flaws in appearance Excessive self-consciousness about the perceived defect(s)Accompanied by any of the following: ▸ Repetitive/excessive behaviors (repeatedly examining) ▸ Excessive attempts to hide or alter the defect ▸ Avoiding social or other situations
Required number of symptoms	Both of the above causing significant impairment and/or distress and excluding other causes of symptoms		
Psychosocial consequences	Distress or psychosocial impairment	Distress or psychosocial impairment	
Exclusions	An eating disorder (if the primary focus is body fat or weight)		Other medical illness Substance
Symptom specifiers	With muscle dysmorphia: preoccupation with perceived low muscle mass or small body build With good or fair insight With poor insight With absent insight and/or delusional beliefs		Fair to good insight Poor to absent insight

Table 36.
Hoarding Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Hoarding disorder	Hoarding disorder	Hoarding disorder
Symptoms	Excessive attachment to possessions Distress related to being separated from items Excessive accumulation of possessions Accumulation results in significant distress, impairment, and/or safety/health compromise	Difficulty parting with objects due to perceived need to save possessions Distress and functional impairment associated with being separated from possessions	Accumulation of possessions, resulting in unusable or unsafe living conditions due to ▶ Repetitive urges/behaviors to amass items ▶ Difficulty discarding possessions
Required number of symptoms	All of the above		
Psychosocial consequences	Symptoms and behaviors cause significant distress or impairment	Symptoms and behaviors cause significant distress or impairment	Symptoms and behaviors cause significant distress or impairment
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness		
Symptom specifiers	With excessive acquisition With good or fair insight With poor insight With absent insight and/or delusional beliefs		Fair to good insight Poor to absent insight
Severity specifiers	With or without insight (per above)		

Table 37.
Trichotillomania/Body-Focused Repetitive Behavior Disorders

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Trichotillomania (hair-pulling disorder)	Trichotillomania	Body-focused repetitive behavior disorders
Symptoms	Recurrent hair pulling causing hair loss Repeated attempts to change/end behaviors	Presence of noticeable hair loss due to recurrent hair pulling Hair-pulling behavior preceded by increased anxiety that is relieved by hair-pulling	Recurrent and habitual actions directed at the integument (e.g., hair-pulling, skin-picking, lip-biting) Typically accompanied by unsuccessful attempts to decrease or stop the behaviors
Required number of symptoms	All of the above		
Psychosocial consequences	Distress or psychosocial impairment		Distress or psychosocial impairment
Exclusions	Other medical condition Substance or medication, including withdrawal	Pre-existing inflammation of the skin Response to delusion or hallucination Stereotyped movement disorder	Cultural behaviors (i.e., self-flagellation as part of religious ceremonies)

Table 38.
Excoriation (Skin Picking) Disorder/Body-Focused Repetitive Behavior Disorders

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Excoriation (skin picking) disorder	Excoriation (skin picking) disorder	Body-focused repetitive behavior disorders
Symptoms	Recurrent skin picking causing lesions to the skin Ongoing skin picking despite repeated attempts to change/end behaviors Actions cause distress and/or impaired functioning	Included under obsessive-compulsive disorders	Recurrent and habitual actions directed at the integument (e.g., hair-pulling, skin-picking, lip-biting). Typically accompanied by unsuccessful attempts to decrease or stop the behaviors
Required number of symptoms	All of the above		
Psychosocial consequences	Distress or psychosocial impairment		Distress or psychosocial impairment
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness	Factitious dermatitis	Cultural behaviors (i.e., self-flagellation as part of religious ceremonies)

7. TRAUMA RELATED DISORDERS

Table 39.
Reactive Attachment Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Reactive attachment disorder	Reactive attachment disorder of childhood	Reactive attachment disorder
Symptoms	Behavioral pattern characterized by emotional withdrawal ▶ Not seeking or not responding to emotional comforting A pattern of social/emotional disturbance ▶ Lack of mood reactivity or positive affective response to stimuli ▶ Inappropriate mood lability/irritability/sadness History of exposure to severely insufficient caregiving ▶ Severe neglect ▶ Unstable caregiving environment with frequent changes in caregivers ▶ Being raised in unusual care settings Evidence for the disorder after the developmental age of 9 and before age 5	A pattern of abnormal social relationships involving emotional disturbance and behavioral changes ▶ i.e., fear, hypervigilance, poor socialization, aggression toward self and others, delayed growth in some instances Symptoms present before 5 years of age Likely occur as a direct result of neglect, abuse, or maltreatment during rearing	History of grossly insufficient care ▶ Disregard for the child's physical and emotional needs ▶ Frequent changes in primary caregivers ▶ Rearing in unusual settings ▶ Maltreatment Abnormal attachment behaviors ▶ Minimal seeking of comfort when distressed ▶ Minimal response to comfort when offered Insufficient care is presumed to be responsible for the child's behaviors Evidence for the disorder after the developmental age of 9 and before age 5
Required number of symptoms	All of the above categories, with behavioral symptoms, felt to be the direct result of exposure to inadequate caregiving or rearing environment		
Psychosocial consequences	Significant disturbance of behavior		
Exclusions	Autism spectrum disorder	Asperger syndrome Disinhibited attachment disorder of childhood Maltreatment syndromes Normal variation in the pattern of selective attachment Sexual or physical abuse in childhood,	Autism spectrum disorder Not confined to one relationship

		resulting in psychosocial problems
Symptom specifiers		Specifiers for: Failure to thrive [R62.51] Growth retardation
Course specifiers	**Persistent** (>12 months)	
Severity specifiers	**Severe:** all symptom criteria met, with symptoms presenting at a high level of severity as well	

Table 40.
Disinhibited Social Engagement Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Disinhibited social engagement disorder	Disinhibited attachment disorder of childhood	Disinhibited social engagement disorder
Symptoms	Childhood pattern of approaching unfamiliar adults ▶ Lack of inhibition in approaching unfamiliar adults ▶ Overly familiar behavior toward unfamiliar adults ▶ Reduced frequency of checking in with known adult/caregiver figure ▶ Minimal or no hesitation to venture forth with an unfamiliar adult History of exposure to insufficient care ▶ History of neglectful rearing environment ▶ A pattern of instability and frequent changes in caregiver ▶ Raised in an unusual setting for child rearing The developmental age of ≥9 months and before age 5 years	A pattern of nonselective attachment, attention-seeking, indiscriminate behaviors with peers and strangers Abnormal social interaction patterns arise before the age of 5 years It may or may not be associated with emotional or behavioral disturbance	History of grossly insufficient care ▶ Disregard for the child’s physical and emotional needs ▶ Frequent changes in primary caregivers ▶ Rearing in unusual settings ▶ Maltreatment Abnormal attachment behaviors ▶ Overly familiar behaviors with adult strangers ▶ Decreased checking back with caregiver ▶ Willing to go off with unfamiliar adults with little hesitation Evidence for the disorder after age 1 or developmental age of 9 months and before age 5
Required number of symptoms	Two symptoms of disinhibited behavior with an unfamiliar adult, in addition to a history of exposure to an abnormal child-rearing environment as above, are related to the development of abnormal behavior		
Psychosocial consequences	Significant disturbance of behavior resulting in impaired safety and/or functioning		
Exclusions	Symptoms limited to	Asperger syndrome	Asperger syndrome

	impulsivity without socially disinhibited behavior	Hospitalism in children Hyperkinetic disorders Reactive attachment disorder of childhood	Adjustment disorder Attention-deficit hyperactivity disorder Reactive attachment disorder of childhood
Course specifiers	**Persistent** Present for ≥12 months		
Severity specifiers	**Severe:** child exhibiting all symptoms, with all symptoms at relatively high levels		

Table 41.
Post-Traumatic Stress Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Post-traumatic stress disorder	Post-traumatic stress disorder	Post-traumatic stress disorder
Duration	>1 month	Symptoms arise in weeks to months following the traumatic event	
Symptoms	History of exposure to (directly experiencing, repeated exposure witnessing in person, learning of occurrence in close acquaintance) actual threatened death, severe injury, or sexual trauma Intrusive symptoms ▶ Involuntary intrusive memories ▶ In children <6 years may see a re-enactment of the event through play ▶ Recurrent nightmares/dreams of the event ▶ In children <6 years, frightening dreams without identifiable content may be present ▶ Dissociative responses or reliving of prior experience (i.e., flashbacks) ▶ In children <6 years may see a re-enactment of the event through play ▶ Psychological distress related to exposure to stimuli that are reminders of prior trauma	Symptoms arise as a delayed or prolonged response to a single or recurrent stressful experience that is life-threatening or catastrophic in severity The traumatic event would cause distress in most individuals Symptoms may include ▶ Flashbacks ▶ Dreams/nightmares ▶ Emotional blunting and subjective numbness ▶ Social withdrawal/detachment ▶ Anhedonia ▶ Avoidance ▶ Hyperarousal/hypervigilance ▶ Sleep disturbance May or may not be associated with co-occurring anxiety and/or depression	Exposure to an extremely threatening or horrifying event. Following the event develops symptoms in all three of the following categories: ▶ Re-experiencing the trauma ▶ Avoiding reminders of the trauma ▶ Persistent perception of heightened threat (e.g., hypervigilance)

	▸ Presence of physiologic response to stimuli that are reminders of prior trauma Avoidance of stimuli associated with prior experience of trauma ▸ Avoidance of memories related to trauma ▸ Avoidance of external reminders of the trauma Negative mood or cognitions related to trauma ▸ Impairment in memories related to the event ▸ Negative perceptions of self and others ▸ Cognitive distortions related to the event ▸ Excessive guilt, anger, or fear ▸ Diminished interest and social withdrawal ▸ Subjective detachment from others ▸ Difficulty experiencing positive feelings in response to previously pleasurable stimuli Altered level of arousal ▸ Irritability and/or anger ▸ Risk-taking ▸ Hypervigilance ▸ Increased startle response ▸ Difficulties with concentration ▸ Sleep disturbances	
Required number of symptoms	In addition to a history of exposure to	Any of the above

	trauma, they must have ▸ ≥1 symptom of intrusion ▸ ≥1 symptom of avoidance ▸ ≥2 symptoms of negative mood/cognition ▸ ≥2 symptoms of arousal alterations,		
Psychosocial consequences	Marked distress and impairment in functioning		Impairment in functioning
Exclusions	Exposure through media, electronics, movies, or photo Related to substance use Related to another medical condition		Acute stress reaction Complex post-traumatic stress disorder
Symptom specifiers	**With dissociative symptoms:** ▸ **Depersonalization:** perception of feeling outside one's own body ▸ **Derealization:** perception of the surrounding environment being unreal or distorted		
Course specifiers	**With delayed expression** ▸ All diagnostic criteria are not met until ≥6 months after an initial traumatic event		

Table 42.
Acute Stress Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Acute stress disorder	Acute stress reaction	Acute stress disorder
Duration	3 days to <1 month following exposure to trauma	Symptoms arise transiently in response to emotional or physical stress and are usually resolved within hours or days	Begin hours to days after stressor, end within a few days of removal of stressor, or if stressor is ongoing, reduced within 1 month.
Symptoms	Exposure to actual or threatened death/injury/sexual violence through ≥1 of the following ways: ▶ Direct experience ▶ Personally witnessing events occurring to others ▶ Learning of events happening to close family or friend ▶ Experiencing repeated/extreme exposure to aversive details of a trauma Intrusive symptoms ▶ Recurrent memories of the trauma ▶ Recurrent nightmares ▶ Dissociative reactions (e.g., flashbacks) ▶ Intense/prolonged distress or physiologic reactions in response to cues of the event Negative mood ▶ Inability to experience positive emotions Dissociate symptoms ▶ Altered sense of reality ▶ Dissociate amnesia regarding an event Avoidance symptoms ▶ Efforts to avoid thinking about the trauma ▶ Efforts to avoid reminders of the trauma	Cognitive “daze” Narrowed attention Difficulties with comprehension Disorientation Autonomic symptoms of panic May be followed by the further withdrawal from the situation or possible agitation/excitability May have partial or complete amnesia related to the event	Exposure to an extremely threatening or horrifying event Response to the stressor that is normal given the severity

	Arousal symptoms ► Insomnia ► Irritability ► Hypervigilance ► Decreased concentration ► Increased startle response	
Required number of symptoms	Nine symptoms from among the above five categories	
Psychosocial consequences	Marked distress and impairment in functioning	
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness	Post-traumatic stress disorder

Table 43.
Adjustment Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Adjustment disorder	Adjustment disorders	Adjustment disorder
Duration	Occurring within 3 months of an acute stressor and resolving within 6 months of the resolution of the stressor		Within 6 months of the resolution of the stressor
Symptoms	Emotional or behavioral changes Marked distress/impairment that is felt to be out of proportion to the stressor itself	Subjective distress/emotional disturbance Arises during the period of adaptation to a change or stress Symptoms of depressed mood or anxiety may arise, and conduct disturbance may be noted in adolescents	A maladaptive reaction to an identifiable psychosocial stressor, consisting of preoccupation with the stressor or its consequences
Required number of symptoms	All of the above		
Psychosocial consequences	Marked distress/functional impairment	Often result in impairment in social functioning	Functional impairment
Exclusions	Exacerbation of existing mental illness Normal bereavement Other psychiatric illness		Separation anxiety disorder of childhood Recurrent depressive disorder Single-episode depressive disorder Prolonged grief disorder Uncomplicated bereavement Burnout Acute stress reaction
Symptom specifiers	**With depressed mood:** ► Predominant low mood, hopelessness **With anxiety:** ► Predominant anxiety, worry **With mixed anxiety and depressed mood:** ► Both low mood and anxiety present **With mixed disturbance of emotions and**		

conduct:

- Mood and conduct disturbance present

Unspecified

Table 44.
Prolonged Grief Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Prolonged grief disorder	Adjustment disorder, grief reaction (see Adjustment disorder)	Prolonged grief disorder
Duration	Within 12 months (6 for child or adolescents) after a death of a person close to the individual. Symptoms last ≥1 month.		An atypically long time following a loss and ≥6 months
Symptoms	Persistent grief ▸ Longing for the deceased ▸ Preoccupation with thoughts of the deceased Accompanying symptoms relating to or resulting from the death: ▸ Identity disruption ▸ Disbelief about the death ▸ Avoiding reminders of the loss ▸ Intense emotional pain ▸ Difficulty resuming relationships/activities ▸ Emotional numbness ▸ Feeling life is meaningless ▸ Intense loneliness		Bereavement following the death of a close person Persistent/pervasive grief response
Required number of symptoms	Persistent grief: one or both Accompanying symptoms: ≥3		
Psychosocial consequences	Marked distress/functional impairment		Marked distress/functional impairment
Exclusions	Normal grief Other medical disorder Other mental disorder Substance use		

8. DISSOCIATIVE DISORDERS

Table 45.
Dissociative Identity Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Dissociative identity disorder	Dissociative (conversion) disorder, unspecified	Dissociative identity disorder
Symptoms	Disrupted sense of identity, which is related to the presence of ≥2 personality states, with associated changes in the sense of self-control, affect, behavior, memory, perception, or sensory-motor function Gaps in memory related to past events or personal information	Disruption of identity characterized by the presence of ≥2 distinct personality states ≥2 personalities recurrently take executive control of the individual's consciousness and functioning Changes in personality state are accompanied by alterations in sensation, perception, affect, cognition, memory, motor control, and behavior.	
Required number of symptoms	All of the above Symptoms part of an accepted belief system		
Exclusions	Normal forgetting Other medical condition Substance or medication, including withdrawal Fantasy play in children	Other mental disorder Substance or medication Other medical disorder	
Psychosocial consequences	Marked distress and/or impairment	Functional impairment	
Comments	Trance and possession disorders: temporary loss of the sense of personal identity and full awareness of the surroundings that are outside of culture-bound norms.		

Table 46.
Dissociative Amnesia

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Dissociative amnesia	Dissociative amnesia Dissociative fugue	Dissociative amnesia
Symptoms	Inability to remember information about self, often related to a prior traumatic experience; amnesia can be localized or specific to an event or generalized for life-long events/identity	Dissociative amnesia: loss of memory related to important, often traumatic or stressful, events Dissociative fugue: all the features of dissociative amnesia accompanied by travel, during which the patient may appear normal to outside observers despite there being amnesia during the period of the fugue	Inability to recall important autobiographical memories, typically recent and related to trauma
Exclusions	Normal forgetting Other mental illness Neurocognitive disorder	Nonalcoholic organic amnesic syndrome Substance-induced amnesic disorder Postictal amnesia in epilepsy Postictal fugue in epilepsy	Other dissociative disorders, trance disorder Other mental disorder Substance or medication Other medical condition
Psychosocial consequences	Marked distress and/or impairment		Functional impairment
Symptom specifiers	***With dissociative fugue*** (traveling or wandering that is accompanied by amnesia)		With or without dissociative fugue

Table 47.
Depersonalization/Derealization Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Depersonalization/derealization disorder	Other dissociative (conversion) disorders	Depersonalization/derealization disorder
Symptoms	Depersonalization (sense of being detached from self or being an outside observer of one's own experiences, sense of unreality, and distorted time/perception) Derealization (sense of unreality or detachment from one's surroundings) During depersonalization or derealization, reality testing remains intact	Includes: Ganser syndrome (intentionally acting like one has a physical or mental illness despite not being sick) Multiple personalities	Persistent experience of either or both: ▸ Depersonalization: experiencing oneself as strange or unreal ▸ Derealization: experiencing other persons/places/objects as strange or unreal During depersonalization or derealization, reality testing remains intact
Required number of symptoms	One or both depersonalization and/or realization, in addition to intact reality testing during episodes		
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness		Other medical condition Substance or medication, including withdrawal Other mental illness
Psychosocial consequences	Marked distress and/or impairment		Marked distress and/or impairment

9. SOMATIC SYMPTOMS AND RELATED DISORDERS

Table 48.
Somatic Symptom Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Somatic symptom disorder	Somatization disorder	Bodily distress disorder
Duration	Symptoms persistent for a period of usually ≥6 months	≥2 years	Persistent, last ≥ several months
Symptoms	Persistent and recurrent thoughts related to the potential severity and seriousness of one's symptoms Out-of-proportion anxiety about these symptoms and overall health Spending excessive amounts of time and energy on these symptoms	Multiple recurring and frequently changing physical symptoms, often accompanied by a long history of involvement in the medical system, frequent use of health care services, and workups not revealing the clear cause of symptoms.	≥1 distressing bodily symptoms Excessive attention to symptoms (preoccupation with symptoms, repeated health care contacts) Attention persists despite reassurance from a healthcare provider
Required number of symptoms	≥1 of the above		
Exclusions		Malingering	Other mental illness
Psychosocial consequences	Marked distress and/or impairment	Disruption of social, interpersonal, and family behavior.	Functional impairment
Symptom specifiers	***With predominant pain***		
Course specifiers	**Persistent:** severe symptoms and impairment for ≥6 months		
Severity specifiers	**Mild:** only 1 of the above symptom criteria is met **Moderate:** ≥2 of the above symptom criteria are met **Severe:** ≥2 of the above symptom criteria are met, in addition to having multiple somatic symptoms or 1 extremely severe somatic symptom	Undifferentiated somatoform disorder (<2 years of less severe symptoms) Somatoform autonomic dysfunction (particularly involving autonomic arousal symptoms, in addition to other less specific complaints Persistent somatoform pain disorder (pain symptoms are predominant among somatic complaints)	Mild Moderate Severe

Table 49.
Illness Anxiety Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Illness anxiety disorder	Hypochondriacal disorder	Hypochondriasis
Duration	Symptoms persistent for a period ≥6 months		
Symptoms	Being excessively preoccupied with having or potentially acquiring a serious illness without somatic symptoms being present (or if present, only mild) Intense anxiety that is out of proportion to any potential risk (i.e., family history) that may exist for a given illness Performance of excessive health-related behaviors and/or presence of maladaptive behaviors or avoidance	Persistent preoccupation with potentially having ≥1 serious medical illness, with anxiety potentially manifesting as somatic complaints, anxiety regarding physical appearance, and potential depression/distress)	Persistent preoccupation or fear about the possibility of having ≥1 serious, progressive, or life-threatening illness, accompanied by Repetitive and excessive health-related behaviors (e.g., checking for illness) Avoidance behaviors related to health
Required number of symptoms	All of the above		
Exclusions	Other mental illness		Body dysmorphic disorder Bodily distress disorder Fear of cancer
Psychosocial consequences			Marked distress and/or impairment
Symptom specifiers	***Care-seeking type*** Care-avoidant type		
Comments		Note: body dysmorphic disorder is classified here.	

Table 50.
Conversion Disorder (Functional Neurologic Symptom Disorder/Dissociative Neurologic Symptom Disorder)

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Conversion disorder (functional neurologic symptom disorder)	Dissociative (conversion) disorders	Dissociative neurologic symptom disorder
Duration			≥1 hour
Symptoms	≥1 symptom of a change in voluntary motor or sensory function; Clinical exam does not identify a correlation between symptoms and a recognized medical/neurologic illness		Abnormal integration of motor/sensory/cognitive functions Symptoms are not consistent with a recognized neurologic disorder
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness		Other dissociative disorder Other mental disorder Substance or medication Other medical disorder
Psychosocial consequences	Marked distress and/or impairment in functioning		Functional impairment
Symptom specifiers	***With weakness or paralysis*** With abnormal movement With swallowing symptoms With speech symptom With attacks or seizures With anesthesia or sensory loss With special sensory symptoms (i.e., vision, olfaction, hearing) With mixed symptoms	**Dissociative stupor** **Dissociative motor disorder** (can include limb, speech, or gait symptoms) **Dissociative convulsions** (mimic epileptic seizures, without characteristic features of epileptiform seizures) **Dissociative anesthesia and sensory loss** **Mixed dissociative (conversion) disorders**	With visual disturbance With auditory disturbance With vertigo or dizziness With other sensory disturbance With nonepileptic seizures With speech disturbance With paresis or weakness With gait disturbance With movement disturbance With cognitive symptoms With other specified symptoms With other unspecified symptoms
Course specifiers	**Acute episode** (<6 months) **Persistent** (≥6 months) **With psychological stressor** **Without psychological stressor**		

Table 51.
Factitious Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Factitious disorder	Intentional production or feigning of symptoms or disabilities, either physical or psychological (factitious disorder)	Factitious disorder
Symptoms	**Factitious disorder imposed on self:** ▸ Act of falsifying symptoms and clinical findings in one's self ▸ Presenting one's self to others as ill ▸ Deceptive behaviors present outside of identifiable external reward or motivation **Factitious disorder imposed on another:** ▸ Act of falsifying symptoms/injury/disease and clinical findings in another individual ▸ Presenting an individual in whom symptoms have been falsified to others as being ill ▸ Deceptive behaviors present outside of identifiable external reward or motivation	Feigning symptoms for no clear reason or motivation, with the aim being to take on the sick role.	Intentional feigning/falsifying/inducing/aggravating medical/psychological/behavioral signs and symptoms, or injury in self or others, associated with an identified deception.
Exclusions	Other mental illness	Factitious dermatitis Person feigning illness with obvious motivation	Malingering
Symptom specifiers			Factitious disorder imposed on self Factitious disorder imposed on another
Course specifiers	**Single episode** Recurrent episodes (≥2 discrete events)		
Comments	***Note: In factitious disorder imposed on another, the perpetrator receives the diagnosis, not the victim.***		

10. FEEDING AND EATING DISORDERS

Table 52.
Pica

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Pica	Pica in infancy and childhood Pica in adults	Pica
Duration	≥1 month	Persistent	Persistent, but can be episodic or variable, or chronic and continuous
Symptoms	Eating substances that are not foods and not nutritional Behavior is inappropriate for development level	Eating non-nutritive substances	Consuming non-nutritive substances Severe enough to require clinical attention Inappropriate for age
Psychosocial consequences			Risks to health/functional impairment
Exclusions	Other medical condition Other mental illness		Other medical conditions, such as a nutritional deficiency
Course specifiers			It can be episodic or variable, or chronic and continuous
Severity specifiers	In remission		

Table 53.
Rumination Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Rumination disorder	Feeding disorder of infancy and childhood: rumination disorder of infancy	Rumination-regurgitation disorder
Duration	≥1 month		≥ Several times per week Sustained for several weeks
Symptoms	Repeated regurgitation of food		Intentional regurgitation of food
Psychosocial consequences			Risks to health/functional impairment
Exclusions	Other eating disorder Other mental illness Other medical conditions, including gastrointestinal disorders		Age <2 Other medical condition Adult rumination syndrome Nausea or vomiting
Severity specifiers	In remission		

Table 54.
Avoidant/Restrictive Food Intake Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Avoidant/restrictive food intake disorder	Other eating disorder	Avoidant/restrictive food intake disorder
Symptoms	Eating/feeding disturbance associated with: ▶ Significant weight loss/lack of expected weight gain ▶ Significant nutritional deficiency ▶ Needing enteral feeding or oral supplements		Avoidance or restriction of food intake ▶ Insufficient intake for nutritional needs ▶ Nutritional deficiencies ▶ Dependence on supplements or tube feeding ▶ Other negative effects on health
Required number of symptoms			Any of the above or functional impairment
Psychosocial consequences	Functional impairment		Functional impairment
Exclusions	Food insecurity Culturally sanctioned practice Other eating disorder Other medical condition		Food insecurity Other medical condition Substance or medication, including withdrawal Other mental illness
Course specifiers	In remission		

Table 55.
Anorexia Nervosa

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Anorexia nervosa	Anorexia nervosa	Anorexia nervosa
Symptoms	Food restriction leading to abnormally low weight Fear of gaining weight and behaviors to prevent weight gain Distorted body image ▶ Misperceiving weight of shape ▶ Self-evaluation based on thinness, lack of recognition of the seriousness of one’s weight	Deliberate weight loss induced by the patient Fear of being fat or flabby Associated physiologic disturbances due to low weight Specific behaviors: ▶ Restricted diet ▶ Excess exercise ▶ Binging behavior ▶ Use of appetite suppressants and diuretics	Low body weight Typically BMI <18.5 kg/m^2 in adults, and under 5th percentile in children and adolescents Pattern or restrictive eating or other behavior for maintaining a low weight Usually associated with fear of weight gain Excess preoccupation with body weight or shape
Required number of symptoms	All		
Exclusions		Loss of appetite (physical or psychological)	Other medical condition Food insecurity
Symptom specifiers	Restricting type: primary behaviors are dieting, fasting, and exercise for the last 3 months Binge eating type: binging and purging behavior for the last 3 months. It can include the use of laxatives or other substances to aid in purging.		
Course specifiers	Partial remission: symptoms present but no longer an abnormally low weight Full remission: no symptoms for a sustained period		
Severity specifiers	Severity is measured by BMI level Mild: BMI ≥17 kg/m^2 Moderate: BMI ≥16, <17 kg/m^2 Severe: BMI ≥15, <16 kg/m^2 Extreme: BMI <15 kg/m^2		
Comments		If lacking core symptoms, the diagnosis of “Atypical	

		anorexia nervosa” is made.	

Table 56.
Bulimia Nervosa

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Bulimia nervosa	Bulimia nervosa	Bulimia nervosa
Duration	Occurs ≥1/wk for 3 months		Usually ≥1/wk for 1 month
Symptoms	Repeated binge-eating episodes ▶ Occurring within 2 hours period ▶ Loss of control during the episode Compensatory behaviors to prevent weight gain, such as ▶ Purging ▶ Laxatives ▶ Diuretics, other substances ▶ Excessive exercise ▶ Fasting Self-evaluation based on weight and shape	Binging Preoccupation with weight control Preoccupation with body shape and weight Associated physiologic disturbances due to repeated vomiting Specific compensatory behaviors: ▶ Vomiting ▶ Use of purgatives	Frequent episodes of binge eating Compensatory behaviors to prevent weight gain, most commonly purging Excess preoccupation with body weight or shape
Required number of symptoms	All	Not specified	
Psychosocial consequences			Distress or functional impairment
Exclusions	Does not occur only during episodes of anorexia nervosa		Anorexia nervosa
Course specifiers	Partial remission: full symptoms no longer met for a sustained time Full remission: no symptoms for a sustained period		
Severity specifiers	Severity is measured by the average number of compensatory behaviors in a week’s time Mild: 1–3 Moderate: 4–7 Severe: 8–13 Extreme: 14+		

Table 57.
Binge Eating Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Binge eating disorder	Atypical bulimia nervosa	Binge eating disorder
Duration	≥1/wk for 3 months		Usually ≥1/wk for 1 month
Symptoms	Repeated binge-eating episodes ▶ Occurring within 2 hours period ▶ Loss of control during the episode Binge-associated behaviors or reactions ▶ Eating quicker than normal ▶ Eating until uncomfortably full ▶ Eating despite not being hungry ▶ Eating alone to avoid embarrassment ▶ Feeling guilty or disgusted after eating	Similar to ICD-10 bulimia nervosa minus the specific compensatory behaviors	Frequent episodes of binge eating No compensatory behaviors to lose weight
Required number of symptoms	Binge behaviors + ≥3 of the associated behaviors or reactions		
Psychosocial consequences	Marked distress		Distress or functional impairment
Exclusions	No compensatory behavior as in bulimia nervosa		Bulimia nervosa Other medical condition Substance or medication, including withdrawal Other mental illness
Course specifiers	Partial remission: occurrences are less frequent (<1 episode/wk) Full remission: no symptoms for a sustained period		
Severity specifiers	Severity is measured by the average number of compensatory behaviors in a week's time Mild: 1–3 Moderate: 4–7 Severe: 8–13 Extreme: 14+		

11. ELIMINATION DISORDERS

Table 58.
Enuresis

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Enuresis	Nonorganic enuresis	Enuresis
Duration	Occurring ≥2/wk for a minimum of 3 consecutive months		Several times per week for several months
Symptoms	Urinating into bed or clothes either voluntary or unintentionally	Involuntary voiding of urine Abnormal for age	Repeated and persistent voiding of urine into bed or clothes Inappropriate for developmental age
Required number of symptoms			
Exclusions	Age <5 Substance use Other medical condition	Other medical condition	Other medical condition Substance or medication
Symptom specifiers	***Nocturnal only*** ***Diurnal only*** ***Nocturnal and diurnal***		

Table 59.
Encopresis

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Encopresis	Nonorganic encopresis	Encopresis
Duration	≥1/month for ≥3 consecutive months		≥1/month for several months
Symptoms	Passing stool into clothes or onto the floor	A pattern of voluntarily or involuntarily passing feces in places that are inappropriate within the sociocultural context.	Repeated and persistent passage of feces in inappropriate places Inappropriate for developmental age (4)
Exclusions	Age <4 Substance use Another medical condition other than once causing constipation	Organic causes of encopresis	Other medical condition Substance or medication, including withdrawal
Severity specifiers	With constipation and overflow incontinence Without constipation and overflow incontinence		With constipation and overflow incontinence Without constipation and overflow incontinence

12. SLEEP DISORDERS

Table 60.
Insomnia Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Insomnia disorder	Disorders of initiating and maintaining sleep (insomnias)	Insomnia disorders
Duration	≥3 nights/wk for ≥3 months		
Symptoms	Dissatisfaction with the quality of sleep due to: ▶ Difficulty falling asleep ▶ Difficulty maintaining sleep due to frequent awakening ▶ Experiencing early morning awakening	Disorders characterized by difficulties with falling asleep and/or staying asleep	Difficulty with initiating, maintaining, or consolidating sleep, or poor quality of sleep despite adequate opportunities to sleep
Required number of symptoms	Any of the above		
Exclusions	Medical disorder Substance or medication Other sleep–wake disorders	Nightmares Nonorganic sleep disorders Sleep terrors Sleepwalking	
Psychosocial consequences	Distress or functional impairment		Daytime impairment
Symptom specifiers	**With nonsleep disorder mental comorbidity** (including substance use disorder) **With other medical comorbidity** **With other sleep disorder**		
Severity specifiers			
Course specifiers	**Episodic** (1–3 months) **Persistent** (>3 months) **Recurrent** (≥2 episodes/year)		

Table 61.
Hypersomnolence Disorder/Idiopathic Hypersomnia

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Hypersomnolence disorder		Idiopathic hypersomnia
Duration	≥3 nights/wk for ≥3 months		Daily period for several months
Symptoms	Excessive sleepiness with Recurrent periods of sleep within the same day Nonrestorative sleep lasting more than 9 hours Difficulty becoming fully awake		Need to sleep Daytime lapses into sleep Mean sleep latency of ≤8 minutes or 24-hour sleep time ≥11 hours (determined on polysomnography) Prolonged and severe sleep inertia
Required number of symptoms	One of the above		
Exclusions	Other sleep disorder Substance or medication Other mental disorder Other medical disorder		Cataplexy Hypocretin deficiency Narcolepsy
Psychosocial consequences	Distress or functional impairment		
Symptom specifiers	With mental disorder With medical condition With another sleep disorder		
Severity specifiers	Mild Moderate Severe		
Course specifiers	Acute Subacute Persistent		

Table 62.
Narcolepsy

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Narcolepsy	Narcolepsy and cataplexy	Narcolepsy
Duration	Symptoms present ≥3 times per week for ≥3 months		
Symptoms	Excessive daytime sleepiness with Cataplexy defined as ▶ Brief losses in muscle tone precipitated by laughing or joking ▶ Abnormal jaw-opening or global hypotonia without a clear emotional trigger CSF hypocretin deficiency REM sleep latency <15 minutes on polysomnography OR mean sleep latency <8 minutes on multiple sleep latency testing with >sleep-onset REM-sleep periods	Episodes of daytime sleepiness occurring despite adequate sleep, often associated with cataplexy, and uncontrollable lapses in consciousness	Need to sleep Daytime lapses into sleep Mean sleep latency of <8 minutes ≥2 sleep-onset REM periods on multiple sleep latency testing or ≥1 sleep-onset REM periods on multiple sleep latency testing and 1 on preceding polysomnography. Sleep is disturbed, and brief naps are typically refreshing
Required number of symptoms	Presence of excessive daytime sleepiness in addition to ≥1 of the above		
Exclusions	Substance use disorder Another sleep–wake disorder	Nightmares Nonorganic sleep disorders Sleep terrors Sleepwalking	
Symptom specifiers	Narcolepsy with cataplexy or hypocretin deficiency Narcolepsy without cataplexy and without hypocretin deficiency or deficiency is unmeasured Narcolepsy with cataplexy or hypocretin deficiency due to a medical condition Narcolepsy with cataplexy and without hypocretin deficiency due to a medical condition		Type 1 with cataplexy and hypocretin deficiency Type 2 without cataplexy without hypocretin deficiency
Severity specifiers	**Mild:** infrequent cataplexy (less than 1/wk), with		

naps only 1–2 times daily, and only mildly or intermittently disturbed nighttime sleep

Moderate: cataplexy 1/day or every few days, in addition to disturbed nighttime sleep and need for multiple naps daily

Severe: drug-resistant with near-constant somnolence and multiple attacks per day, in addition to disturbed night-time sleep

Table 63.
Circadian Rhythm Sleep Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Circadian rhythm sleep disorder	Disorders of sleep–wake schedule **Delayed sleep phase syndrome** **Irregular sleep–wake pattern**	Circadian rhythm sleep-wake disorders
Duration	**Episodic** (lasting 1–3 months) **Persistent** (lasting >3 months) **Recurrent** (>2 episodes within 1 year)		
Symptoms	Disturbed sleep due to alteration in the circadian system Sleep disruption leading to excessive daytime sleepiness Sleep disturbance leads to impairment and/or distress	Persistent sleep disturbance (i.e., insomnia or excessive daytime sleepiness) due to discrepant circadian rhythm and environmental demands Disruption of the normal 24-hour circadian cycle	Disturbances of the sleep-wake cycle due to alterations of the circadian time-keeping system.
Required number of symptoms	All of the above	One of the above	
Exclusions	Other medical condition Substance or medication, including withdrawal	Nightmares Nonorganic sleep disorders Sleep terrors Sleepwalking	
Psychosocial consequences	Marked distress and/or impairment		
Symptom specifiers	Delayed sleep phase type Familial Overlapping with non–24-hour sleep–wake type Advanced sleep phase type Familial Irregular sleep–wake type Non–24-hour sleep-wake type Shift work type Unspecified type		Delayed sleep-wake phase disorder Advanced sleep-wake phase disorder Irregular sleep-wake rhythm disorder Non–24-hour sleep-wake rhythm disorder Circadian rhythm sleep-wake disorder, shift work type Circadian rhythm sleep-wake disorder, jet lag type Circadian rhythm sleep-wake disorders, unspecified
Course specifiers	**Episodic** (lasting 1–3		

months)
Persistent (lasting >3 months)
Recurrent (≥2 within 1 year)

Table 64.
REM Sleep Behavior Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	REM sleep behavior disorder	REM sleep behavior disorder	REM sleep behavior disorder
Symptoms	Repeated episodes of arousal during sleep with vocalization or complex motor behavior Behaviors occur during REM sleep No confusion or disorientation present on awakening Either of the following also present: ▸ REM sleep without atonia on polysomnography ▸ History consistent with REM sleep behaviors with an established diagnosis of synucleinopathy (i.e., Parkinson disease) Symptoms not adequately explained by coexisting mental and medical disorders	Recurrent episodes of sudden and often violent motor activity during REM sleep, at times, related to self-injury or harming others, in addition to difficulty awakening the individual during episodes; episodes often associated with dream enactment	Repeat episodes of either Sleep related vocalization Complex movements Document on polysomnography or presumed to occur during REM sleep
Required number of symptoms	All of the above		
Exclusions	Other medical condition Substance or medication, including withdrawal Other mental illness	-	
Psychosocial consequences	Marked distress and/or impairment in functioning		
Severity specifiers			Mild Moderate Severe

Table 65.
Non–REM-Sleep Behavior Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Non–REM-sleep behavior disorder	Nonorganic disorders of the sleep–wake schedule: **Sleepwalking (somnambulism)** **Sleep terrors (night terrors)** **Nightmares**	Disorders of arousal from non-REM sleep
Symptoms	Recurrent episodes of incomplete awakening from sleep, accompanied by either ▸ Sleepwalking ▸ Sleep terrors Little dream recall, if any Amnesia for episodes Symptoms no adequately explained by other mental or medical disorders	Sleepwalking: altered state of consciousness in which sleep and wakefulness are combined; episodes associated with low levels of awareness, with no recall of event on awakening Sleep terrors: nighttime episodes of terror and panic associated with motor movement and vocalization, in addition to high levels of autonomic tone, associated with little to no recall of the event Nightmares: dreams associated with fear and negative emotion, with the presence of intact recall of dream content on awakening. Not associated with vocalization and body movement	Experiences or behaviors related to non-REM sleep.
Symptom specifiers	**Sleep terror type** Sleepwalking type ▸ **With sleep-related eating** ▸ **With sleep-related sexual behavior (sexsomnia)**		Confusional arousals Sleepwalking disorder Sleep terrors Sleep-related eating disorder

Table 66.
Breathing-Related Sleep Disorders

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Breathing-related sleep disorders: ▶ **Obstructive Sleep apnea hypopnea** ▶ **Central sleep apnea** ▶ **Sleep-related hypoventilation**	Sleep apnea **Central** **Obstructive**	Sleep-related breathing disorders
Symptoms	**Obstructive sleep apnea hypopnea:** ▶ Evidence on polysomnography of ≥5 obstructive apneas or hypopneas per hour of sleep with either: ▶ Nocturnal breathing disturbance (i.e., snoring, breathing pauses) ▶ Daytime sleepiness, fatigue ▶ Evidence on polysomnography of 15 more obstructive apneas and/or hypopneas per hour of sleep, regardless of accompanying symptoms **Central sleep apnea:** ▶ ≥5 central apneas per hour of sleep on polysomnography **Sleep-related hypoventilation:** ▶ Decreased respiration causing increased CO_2 levels, as correlated by polysomnography		Abnormalities of respiration during sleep
Required number of symptoms	**Obstructive sleep apnea hypopnea**: one of the above criteria must be met **Central sleep apnea:** all criteria met **Sleep-related hypoventilation:** all criteria met		
Exclusions	Another sleep disorder Use of a substance or medication	Pickwickian syndrome Sleep apnea of newborn	Apnea of newborn

Symptom specifiers	Obstructive sleep apnea hypopnea Central sleep apnea: Idiopathic central sleep apnea Cheyne–Stokes breathing Central sleep apnea comorbid with opioid use Sleep-related hypoventilation Idiopathic hypoventilation Congenital central alveolar hypoventilation Comorbid sleep-related hypoventilation		Central sleep apnea Obstructive sleep apnea ≥15 obstructive events/hour or ≥5 events plus symptoms or distress or disorder results Sleep-related hypoventilation or hypoxemia Sleep-related Cheynes-Stokes respiration
Severity specifiers	Obstructive sleep apnea hypopnea **Mild:** apnea hypopnea index is <15 **Moderate**: apnea hypopnea index 15–30 **Severe**: apnea hypopnea index >30 **Central sleep apnea:** severity based on the frequency of breathing disturbance as well as the severity of oxygen desaturation/sleep fragmentation. **Sleep-related hypoventilation:** severity is based on the degree of hypoxemia/hypercarbia, in addition to the presence of end-organ impairment. Blood gas abnormality during wakefulness is an indicator of more severe illness.		

Table 67.
Nightmare Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Nightmare disorder	Nonorganic disorders of the sleep–wake schedule **Sleep terrors** **Nightmares**	Nightmare disorder
Symptoms	A pattern of experiencing long, dysphoric, and vividly remembered dreams that are often threatening in nature Rapid orientation and alertness on awakening Not adequately explained by a coexisting mental or medical disorder	Sleep terrors: nighttime episodes of terror and panic associated with motor movement and vocalization, in addition to high levels of autonomic tone, associated with little to no recall of the event Nightmares: dreams associated with fear and negative emotion, with the presence of intact recall of dream content on awakening. Not associated with vocalization and body movement	Recurrent, vivid, and dysphoric dreams
Required number of symptoms	All of the above		
Exclusions	Result of ▶ Effects of a substance or medication		
Psychosocial consequences	Marked distress and/or impairment in areas of functioning	Marked distress and/or impairment in areas of functioning	
Symptom specifiers	**During sleep onset** With associated nonsleep disorder (including substance use disorder) With associated other medical condition With associated other sleep disorder		
Severity specifiers	**Mild** (<1 episode/wk) **Moderate** (≥1 episodes per week, but less than nightly) **Severe** (nightly episodes)		
Course specifiers	**Acute** (duration <1 month) **Subacute** (1–6 months) **Persistent** (>6 months)		

Table 68.
Restless Legs Syndrome

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Restless legs syndrome	Other specified extrapyramidal and movement disorders Restless legs syndrome	Restless legs syndrome
Duration	Symptoms occur ≥3 times per week for ≥3 months		
Symptoms	Urge to move legs, including worsening urge at night, with urge relieved by movement	Experiencing a strong urge to move legs in order to relieve discomfort; often occurs while lying down or sitting and tends to be worse at night	Strong urge to move limbs
Required number of symptoms	All of the above		
Exclusions	Akathisia Effects of a substance or medication Another mental disorder or medical condition Behavioral condition (i.e., habitual movements)		
Psychosocial consequences	Marked distress and/or impairment	Marked distress and/or impairment	Marked distress and/or impairment

13. SEXUAL DISORDERS

Table 69.
Delayed Ejaculation

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Delayed ejaculation	Orgasmic dysfunction	Male delayed ejaculation
Duration	≥6 months duration of symptoms		
Symptoms	Delay in ejaculation Absence or reduced frequency of ejaculation	Marked delay or absence of orgasm	Inability to achieve or increased latency of ejaculation despite adequate stimulation
Required number of symptoms	Either or both of the above symptoms		
Exclusions	Nonsexual mental disorder Significant relationship distress Side effects of a medication Substance use Another medical condition	Organic causes of sexual dysfunction	
Psychosocial consequences	Marked distress		
Symptom specifiers	***Generalized vs. Situational***		Associated with a medical condition, injury, or the effects of surgery or radiation treatment Associated with psychological or behavioral factors, including mental disorders Associated with the use of psychoactive substances or medication Associated with a lack of knowledge or experience Associated with relationship factors Associated with cultural factors Other specific etiologic considerations in sexual dysfunctions and sexual pain disorders
Course specifiers	***Lifelong vs. Acquired***		
Severity specifiers	**Mild:** symptoms causing mild distress **Moderate:** symptoms causing moderate		

	distress **Severe:** symptoms causing severe distress	
Comments		Note: ICD-10 classifies orgasmic dysfunction as occurring in men and women, including delayed ejaculation in this category as per Table 16-5

Table 70.
Erectile Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Erectile disorder	Failure of genital response	Male erectile dysfunction
Duration	≥6 months duration of symptoms		
Symptoms	Difficulty achieving an erection Difficulty maintaining erection Reduction in erectile rigidity	Male: difficulty developing maintaining or developing an erection for satisfactory sexual intercourse Female: can present with vaginal dryness or lack of adequate lubrication for sexual intercourse	Decreased or no ability to attain/sustain penile erection despite adequate sexual desire and stimulation
Required number of symptoms	≥1 of the above present in all or most sexual encounters		
Exclusions	Nonsexual mental disorder Significant relationship distress Side effects of a medication Substance use Another medical condition		
Psychosocial consequences	Marked distress		
Symptom specifiers	***Generalized vs. Situational***		Associated with a medical condition, injury, or the effects of surgery or radiation treatment Associated with psychological or behavioral factors, including mental disorders Associated with the use of psychoactive substances or medication Associated with a lack of knowledge or experience Associated with relationship factors Associated with cultural factors Other specific etiologic considerations in sexual dysfunctions and sexual pain disorders

Course specifiers	***Lifelong vs. Acquired***	
Severity specifiers	**Mild:** symptoms causing mild distress **Moderate:** symptoms causing moderate distress **Severe:** symptoms causing severe distress	
Comments		Note: ICD-10 classifies this diagnosis as including male erectile disorder as well as female sexual arousal disorder

Table 71.
Female Orgasmic Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Female orgasmic disorder	Orgasmic dysfunction	Anorgasmia
Duration	≥6 months duration of symptoms		
Symptoms	Delay or reduced frequency of orgasm Reduced intensity of orgasm	Marked delay or absence of orgasm	Absence or infrequency of orgasm or diminished intensity. In women can include significant delay in orgasm.
Required number of symptoms	Either of the above is present for most or all sexual encounters		
Exclusions	Nonsexual mental disorder Significant relationship distress Side effects of a medication Substance use Another medical condition		
Psychosocial consequences	Marked distress		
Symptom specifiers	***Generalized vs. Situational***		Associated with a medical condition, injury, or the effects of surgery or radiation treatment Associated with psychological or behavioral factors, including mental disorders Associated with the use of psychoactive substances or medication Associated with a lack of knowledge or experience Associated with relationship factors Associated with cultural factors Other specific etiologic considerations in sexual dysfunctions and sexual pain disorders Associated with the use of psychoactive substances or medication

			Associated with a lack of knowledge or experience Associated with relationship factors Associated with cultural factors Other specific etiologic considerations in sexual dysfunctions and sexual pain disorders
Course specifiers	***Lifelong vs. Acquired*** Never experienced an orgasm under any conditions		
Severity specifiers	**Mild:** symptoms causing mild distress **Moderate:** symptoms causing moderate distress **Severe:** symptoms causing severe distress		
Comments		Note: ICD-10 classifies orgasmic dysfunction as occurring in men and women, including delayed ejaculation in this category as per Table 16-6	

Table 72.
Female Sexual Interest/Arousal Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Female sexual interest/arousal disorder	Lack or loss of sexual desire Sexual aversion and lack of sexual enjoyment	Female sexual arousal dysfunction
Duration	≥6 months duration of symptoms		
Symptoms	Symptoms of reduced sexual interest or arousal, such as in: ▶ Reduced interest in sex ▶ Reduced sexual thoughts or fantasies ▶ Reduced initiation of/receptiveness to sexual activity ▶ Absent or reduced pleasure from sexual activity ▶ Absent or reduced pleasure from internal or external sexual cues ▶ Absent or reduced genital sensations during sexual activity	Lack or loss of sexual desire not attributable to physiologic difficulties related to intercourse Sexual aversion due to anxiety and subsequent avoidance of sexual activity, or lack of enjoyment given that orgasm is not accompanied by appropriate pleasure.	Absence or reduction in response to sexual stimulation despite adequate stimulation. Areas that may be affected include: ▶ Genital response ▶ Nongenital response (e.g., nipples, skin, physiologic) ▶ Sexual arousal when stimulated
Required number of symptoms	≥3 of the above symptoms present during most or all sexual encounters		
Exclusions	Nonsexual mental disorder Significant relationship distress Side effects of a medication Substance use Another medical condition		
Psychosocial consequences	Marked distress		
Symptom specifiers	***Generalized vs. Situational***		Associated with medical condition, injury, or the effects of surgery or radiation treatment Associated with psychological or behavioral factors, including mental disorders Associated with the use of psychoactive substances or medication

			Associated with a lack of knowledge or experience Associated with relationship factors Associated with cultural factors Other specific etiologic considerations in sexual dysfunctions and sexual pain disorders
			Associated with the use of psychoactive substances or medication Associated with lack of knowledge or experience Associated with relationship factors Associated with cultural factors Other specific etiologic considerations in sexual dysfunctions and sexual pain disorders
Course specifiers	***Lifelong vs. Acquired***		
Severity specifiers	**Mild:** symptoms causing mild distress **Moderate:** symptoms causing moderate distress **Severe:** symptoms causing severe distress		

Table 73.
Genito-Pelvic Pain/Penetration Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	**Genito-pelvic pain/penetration disorder**	**Nonorganic vaginismus** Nonorganic dyspareunia	**Sexual pain-penetration disorder**
Duration	≥6 months duration of symptoms		
Symptoms	Difficulties with vaginal penetration Pain during intercourse/with penetration Fear or anxiety about pain during sexual encounters Tension or tightening of pelvic muscles during attempted penetration	***Vaginismus***: spasm of pelvic floor muscles around the vagina, causing pain and difficulties with penetration Dyspareunia: pain or discomfort related to penile penetration or intercourse	
Required number of symptoms	≥1 of the above		
Exclusions	Nonsexual mental disorder Significant relationship distress Side effects of a medication Substance use Another medical condition	Result of: Organic causes of vaginismus or dyspareunia	
Psychosocial consequences	Marked distress		
Symptom specifiers			Associated with a medical condition, injury, or the effects of surgery or radiation treatment Associated with psychological or behavioral factors, including mental disorders Associated with the use of psychoactive substances or medication Associated with lack of knowledge or experience Associated with relationship factors Associated with cultural factors Other specific etiologic considerations in

			sexual dysfunctions and sexual pain disorders Associated with the use of psychoactive substances or medication Associated with lack of knowledge or experience Associated with relationship factors Associated with cultural factors Other specific etiologic considerations in sexual dysfunctions and sexual pain disorders
Course specifiers	***Lifelong vs. Acquired***		
Severity specifiers	**Mild:** symptoms causing mild distress **Moderate:** symptoms causing moderate distress **Severe:** symptoms causing severe distress		
Comments		Note: ICD-10 classifies orgasmic dysfunction as occurring in men and women, including delayed ejaculation in this category as per Table 16-5	

Table 74.
Male Hypoactive Sexual Desire Disorder

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Male hypoactive sexual desire disorder	Lack or loss of sexual desire	Hypoactive sexual desire dysfunction
Duration	Persistent or recurrent for about 6 months		
Symptoms	Decreased sexual/erotic thoughts, fantasies, and sexual desire	Loss of sexual desire	Absent or decreased desire or motivation for sex: ► Reduced/absent desire ► Reduced/absent response to cues ► Inability to sustain desire or motivation once begun
Required number of symptoms			Any of the bulleted points
Exclusions	Other mental disorder Relationship distress Significant stressors Substance or medication Other medical condition	Other sexual difficulty	
Psychosocial consequences	Distress		
Symptom specifiers			Associated with a medical condition, injury, or the effects of surgery or radiation treatment Associated with psychological or behavioral factors, including mental disorders Associated with the use of psychoactive substances or medication Associated with a lack of knowledge or experience Associated with relationship factors Associated with cultural factors Other specific etiologic considerations in sexual dysfunctions and sexual pain disorders Associated with the use of psychoactive

			substances or medication Associated with a lack of knowledge or experience Associated with relationship factors Associated with cultural factors Other specific etiologic considerations in sexual dysfunctions and sexual pain disorders
Course specifiers	Lifelong vs. Acquired Generalized vs. Situational ***Never experienced an orgasm under any conditions***		
Severity specifiers	**Mild:** symptoms causing mild distress **Moderate:** symptoms causing moderate distress **Severe:** symptoms causing severe distress		
Comments	Requires clinical judgment to decide if not age or situationally appropriate		

Table 75.
Premature (Early) Ejaculation

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Premature (early) ejaculation	Premature ejaculation	Male early ejaculation
Duration	≥6 months duration of symptoms		Persistent or episodic for several months
Symptoms	A pattern of ejaculation within 1 minute of penetration and before the affected individual wishes for it to occur	Difficulty controlling ejaculation such that both partners cannot sufficiently enjoy sexual intercourse	Ejaculation before or shortly after penetration or stimulation.
Required number of symptoms	Above symptom present in all or most sexual encounters		
Exclusions			
Psychosocial consequences	Marked distress		
Symptom specifiers	***Generalized vs. Situational***		Associated with a medical condition, injury, or the effects of surgery or radiation treatment Associated with psychological or behavioral factors, including mental disorders Associated with the use of psychoactive substances or medication Associated with a lack of knowledge or experience Associated with relationship factors Associated with cultural factors Other specific etiologic considerations in sexual dysfunctions and sexual pain disorders Associated with the use of psychoactive substances or medication Associated with a lack of knowledge or experience Associated with relationship factors

			Associated with cultural factors Other specific etiologic considerations in sexual dysfunctions and sexual pain disorders
Course specifiers	***Lifelong vs. Acquired***		
Severity specifiers	**Mild:** ejaculation within 30 seconds to 1 minute of penetration **Moderate:** ejaculation within 15–30 seconds of penetration **Severe:** ejaculation within 15 seconds of penetration		

14. GENDER DYSPHORIA

Table 76.
Gender Dysphoria in Children/Gender Incongruence of Childhood

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Gender dysphoria in children	Gender identity disorder of childhood	Gender incongruence of childhood
Duration	≥6 months		2 years
Symptoms	Desire to be other gender or insistence one is that gender Preference for cross-dressing in boys, preference for typical masculine, and resistance to typical feminine clothing in girls. Preference for cross-gender roles during play Preference for toys/games/activities stereotypically used by another gender Preference for playmates of another gender Rejection of toys/games/activities stereotypically of one's assigned gender Dislike of one's sexual anatomy A desire for sex characteristics of one's experienced gender	Distress about assigned sex	Incongruence between experienced/expressed gender and assigned sex Strong desire to be a different gender Dislike of their sexual anatomy/anticipated secondary sex characteristic A desire for secondary sex characteristics to be of a different gender Activities and play typical of the experienced gender
Required number of symptoms	6		
Psychosocial consequences	Distress or psychosocial impairment		
Exclusions		Gender variant behavior or preferences	Gender variant behavior or preferences
Symptom specifiers	With a disorder/difference of sex development (e.g., congenital disorder)		

Table 77.
Gender Dysphoria in Adolescents and Adults/Gender Incongruence of Adolescence or Adulthood

Disorder	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Gender dysphoria in adolescents and adults	Transsexualism	Gender incongruence of adolescence or adulthood
Duration	≥6 months		
Symptoms	Incongruence between experienced/expressed gender and primary or secondary sex characteristics Desire to be rid of primary/secondary sex characteristics or prevent their development The desire for primary/secondary sex characteristics of another gender Desire to be a different gender Desire to be treated as a different gender A conviction that they feel and react like a different gender	Desire to live and be accepted as a member of the opposite sex Discomfort with one's anatomic sex or sense that it is inappropriate Desire to transition	Incongruence between experienced/expressed gender and assigned sex Desire to transition
Required number of	2		
Psychosocial consequences	Distress or psychosocial impairment		
Exclusions			Gender variant behavior or preferences
Symptom specifiers	With a disorder/difference of sex development (e.g., congenital disorder)		
Course specifiers	Post-transition: living as experienced gender and has undergone ≥ one gender-affirming medical procedure		

15. DISRUPTIVE, IMPULSE-CONTROL, AND CONDUCT DISORDERS

Table 78.
Oppositional Defiant Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Oppositional defiant disorder	Oppositional defiant disorder	Oppositional defiant disorder
Duration	≥6 months If <5 years, occurs most days If ≥5 years, occurs ≥1 time/wk		Extended period
Symptoms	Loses temper Sensitive/easily annoyed Angry/resentful Argues with authorities Disobeys authorities or rules Annoys others on purpose Blames others Spiteful or vindictive ≥2×/past 6 months	Conduct disorder, usually in younger children, with: ▶ Disobedience ▶ Defiance ▶ Disruptive behavior No delinquency or severe aggression or dissocial behaviors	Noncompliant, defiant, and disobedient behaviors inappropriate for age with: ▶ Difficulty getting along with others ▶ Provocative, spiteful or vindictive ▶ Extremely irritable or angry
Required number of symptoms	≥4		
Psychosocial consequences	Distress or psychosocial impairment		Distress or psychosocial impairment
Exclusions	Substance- or medication-induced, including withdrawal Other mental disorder		Relationship problems (i.e., response to overly harsh parents or teachers)
Symptom specifiers			With or without chronic irritability-anger With limited prosocial emotions With typical prosocial emotions
Severity specifiers	**Mild:** occurs in 1 setting **Moderate:** occurs in ≥2 settings **Severe:** occurs in ≥3 settings		

Table 79.
Intermittent Explosive Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Intermittent explosive disorder	Other habit and impulse disorders	Intermittent explosive disorder
Duration	See below		≤3 months or 1 year if outburst are severe
Symptoms	Age ≥6Recurrent spontaneous outbursts that are an overreaction to an event, with: ▸ Verbal/physical aggression 2×/wk for ≥3 months ▸ 3 episodes resulting in harm/destruction over 1 year period	Pattern of repeated maladaptive behaviors, not better explained by another disorder, associated with prodromal tension and release at time of act.	Age ≥6 Brief impulsive/reactive episodes of verbal/physical aggression. Out of proportion to the circumstance and developmental age
Required number of symptoms	All of the above		
Psychosocial consequences	Distress or psychosocial impairment		Distress or psychosocial impairment
Exclusions	Other medical condition Substance- or medication-induced, including withdrawal Other mental disorder		Other medical condition Substance- or medication-induced, including withdrawal Other mental disorder
Course specifiers	***Persistent:*** >12 months		
Severity specifiers	***Severe:*** if all symptom criteria met at relatively high level		

Table 80.
Conduct Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Conduct disorder	Conduct disorder	Conduct-dissocial disorder
Duration	≥1 symptom for ≥6 months Other symptoms ≥1 year	6 months or longer	≥12 months, persistent and recurrent behavior
Symptoms	Bullying others Initiating physical fights Use of a weapon causing serious harm Physical cruelty toward others Physical cruelty toward animals Stealing while physically confronting a victim Forcing sexual activity on another Deliberately setting a fire Deliberately destroying someone’s property Breaking into another’s property Lying to obtain goods or favors or to avoid obligations Stealing items of high value without necessarily confronting victim Staying out at night against parental restrictions ≤age 13 ≥2 episodes of running away (or 1× if prolonged) Frequent truancy ≤age 13	Repeated and pervasive pattern of antisocial behavior, aggression and defiance that are significantly deviant for age-appropriate expectations. Actions may involve cruelty, bullying, destructiveness, lying, running away/truancy, and anger.	Violating rights or others/laws/norms, including ▶ Aggression toward people/animals ▶ Destruction of property ▶ Deceit/theft ▶ Serious rule violations
Required number of symptoms	3		
Psychosocial consequences			Distress or psychosocial impairment
Symptom specifiers	**With limited prosocial emotions:** demonstrating at least two of the following over 12 months an in multiple contexts: ▶ ***Lack of remorse or guilt*** ▶ ***Callous—lack of empathy*** ▶ ***Unconcerned about performance***	**Conduct disorder confined to the family** context (within nuclear family) U**nsocialized conduct disorder** (predominant aggression, often toward other children) **Socialized conduct disorder** (despite aggression, there is	With limited prosocial emotions With typical prosocial emotions

	▶ ***Shallow or deficient affect***	good integration into their peer group)	
Course specifiers	**Childhood onset type** (symptoms present prior to age 10) **Adolescent onset type** (no symptoms prior to age 10)		Childhood onset Adolescent onset
Severity specifiers	**Mild:** few symptoms in excess of number needed to make diagnosis, with relatively minor resultant harm to others **Moderate:** intermediate symptoms and consequences **Severe:** many issues in excess of minimum needed for diagnosis, with resultant harm to others being considerable		

Table 81.
Pyromania

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Pyromania	Pyromania	Pyromania
Symptoms	Fire setting 1+ incidents Deliberate and purposeful Fascination/preoccupation with fire and related subjects Followed by a period of gratification or release	Fire setting Multiple acts or attempts No apparent motive Preoccupied by fire, burning, and related phenomena Preceded by a period of tension Followed by a period of excitement	Fire setting Strong impulse resulting in multiple acts or attempts No external motivation or gain Fascination/preoccupation with fire and related subjects Increased tension/affect before episodes Followed by a period of gratification or release
Required number of symptoms	All of the above	All of the above	
Exclusions	▶ Consult disorder ▶ Mania ▶ Antisocial personality disorder ▶ Psychosis ▶ Impaired cognition or judgment Motivated by: ▶ Desire for a perceived financial or social benefit ▶ Sociopolitical ideology ▶ Need to conceal criminal activity ▶ Anger ▶ Vengeance	▶ Dissocial personality disorder ▶ Alcohol of substance disorder ▶ Feigned illness ▶ Conduct disorder ▶ Organic mental disorder ▶ Schizophrenia	Disorder of intellectual development Substance- or medication-induced, including withdrawal Other mental disorder
Comments		Included among the habit and impulse disorders	

Table 82.
Kleptomania

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Kleptomania	Pathologic stealing (kleptomania)	Kleptomania
Symptoms	Recurrent stealing Failure to resist the impulse to steal Items stolen are not needed Preceded by a period of tension Followed by a period of gratification or release	Failure to resist impulse to steal Items stolen are not needed Usually preceded by period of tension Usually followed by a period of gratification or release	Failure to resist impulse to steal Items stolen are not needed Preceded by period of tension Followed by a period of gratification or release
Required number of symptoms	All of above		
Exclusions	Result of ▶ Anger ▶ Vengeance ▶ Psychosis Consult disorder Mania Antisocial personality disorder	Other mental disorder	Disorder of intellectual development Substance- or medication-induced, including withdrawal Other mental disorder
Comments		Included among the habit and impulse disorders	

16. SUBSTANCE USE AND ADDICTIVE DISORDERS

Table 83.
Substance Use Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Substance use disorder (indicate specific substance) ▶ Alcohol ▶ Cannabis ▶ Hallucinogens ▶ Inhalants ▶ Opioids ▶ Sedatives, Hypnotics, Anxiolytics ▶ Stimulants	Mental and behavioral disorders due to (indicate specific substance) Indicate if a dependence syndrome	Substance dependence (indicate specific substance)
Duration	Symptoms occur within 12 months		≥3 months
Symptoms	Impaired control ▶ Use of larger amounts/longer use than intended ▶ Persistent desire or unsuccessful efforts to reduce the use ▶ Significant time invested in obtaining, using, or recovering from a substance ▶ Cravings to use Social impairment ▶ Use interfering in the fulfillment of responsibilities ▶ Use despite social, personal, financial, health, and other consequences of substance use ▶ Reduction in other personal and interpersonal activities due to use Risky use ▶ Use in high-risk or dangerous situations ▶ Use despite recognition of potential physical or mental problems caused by substance use Pharmacologic criteria ▶ Tolerance ▶ Withdrawal (does not apply to inhalants or hallucinogens)	Behavioral, cognitive, and physiologic manifestations of repeated substance use Frequently associated with cravings to use the drug and difficulties controlling use Pattern of persistent use despite recognition of adverse consequences of use Prioritization of substance use over other social, financial, and personal obligations Frequently accompanied by tolerance and withdrawal states	Recurrent (episodic or continuous) use of a substance with associated impairment. ▶ Impaired control over the use ▶ Preference of substance to other obligations and activities ▶ Physiologic features, including tolerance or withdrawal

Required number of symptoms	≥2		≥2
Psychosocial consequences	Marked impairment and/or distress		
Exclusions		***Psychotic state*** Nonalcoholic Korsakov psychosis or syndrome	An episode of harmful use A harmful pattern of use
Symptom specifiers	**In a controlled environment:** currently in an environment where access to substances is restricted **Substance-specific specifiers:** Tobacco use disorder: On maintenance therapy Opioid use disorder: On maintenance therapy Hallucinogen use disorder: Specify the hallucinogen Inhalant use disorder: Specify the inhalant	***Associated specifiers:*** Psychotic disorder: psychotic symptoms following a psychoactive substance, not due to acute intoxication alone or due to withdrawal alone Amnesic syndrome: chronic impairment in memory due to the use	
Course specifiers	**In early remission:** no symptoms for 3–12 months (except cravings) In sustained remission: no symptoms >1 year (except cravings)		Current use Early full remission: 1–12 months Sustained partial remission: mostly abstinent for ≥12 months Sustained remission: fully abstinent for ≥12 months
Severity specifiers	**Mild**: 2–3 symptoms **Moderate**: 4–5 symptoms **Severe**: ≥6 symptoms		
Comments		Note: ICD additionally defines “Harmful use” as a pattern of use causing harm to health, both physical and/or mental. This classification is independent of the presence of a dependence syndrome.	Alternative diagnoses include: Harmful use of a substance: ≥1 month in which use of a substance causes physical or mental harm. A harmful pattern of use of a substance: regular or sporadic use of a substance that causes

physical or mental harm.

Table 84.
Substance Intoxication

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Substance intoxication ▶ Alcohol ▶ Caffeine ▶ Cannabis ▶ PCP, hallucinogens ▶ Inhalants ▶ Opioids ▶ Sedatives, hypnotics, anxiolytics ▶ Stimulants	Mental and behavioral disorders due to ▶ Use of alcohol ▶ Use of opioids ▶ Use of cannabinoids ▶ Use of sedatives or hypnotics ▶ Use of cocaine ▶ Use of other stimulants, including caffeine ▶ Use of hallucinogens ▶ Use of tobacco ▶ Use of volatile solvents Indicate if acute intoxication	Substance intoxication Disturbance in affect, behavior, perception, cognition, or consciousness after substance use Symptoms are known effects of the particular substance Symptoms are transient and improve as the substance is eliminated
Symptoms	Recent use of a substance Problematic behavior, mood, or cognition resulting ≥1 substance-specific symptoms of intoxication	Change in level of awareness, cognition, perception, behavior, and/or psychology following ingestion of a psychoactive substance Symptoms directly related to substance and resolve with time	
Required number of symptoms	The required minimum number of symptoms varies by substance: Alcohol: 1 Caffeine: 5 Cannabis: 2 PCP, other hallucinogens: 2 Inhalant: 2 Opioid: 2, 1 being pupillary constriction Sedative/Hypnotic/Anxiolytic: 1 Stimulant: 2		
Exclusions	Other medical condition Substance- or medication-induced, including withdrawal Other mental disorder	Intoxication due to poisoning	Other physical or mental disorders
Symptom specifiers	**Substance-specific specifiers:** Cannabis intoxication: With perceptual disturbance: hallucinations occurring in the setting of intact reality testing or hallucinations in the absence of delirium Opioid intoxication:		

	With perceptual disturbance: hallucinations occurring in the setting of intact reality testing or hallucinations in the absence of delirium	
Severity specifiers		Mild Moderate Severe

Table 85.
Substance Withdrawal

Diagnostic name	Substance withdrawal ► Alcohol ► Caffeine ► Cannabis ► Opioids ► Sedatives, hypnotics, anxiolytics ► Stimulants ► Tobacco	Mental and behavioral disorders due to the use of (indicate substance)	Substance withdrawal
Duration			Usually, 6–12 hours after last using the substance
Symptoms	Reduction or cessation of use following a pattern of prolonged and/or heavy use Symptoms of substance-specific withdrawal syndrome following cessation or reduction ► ≥2 symptoms of established sedative/hypnotic/anxiolytic withdrawal syndrome within hours to days ► ≥2 symptoms of established alcohol withdrawal syndrome ► Dysphoric mood in addition to ≥2 symptoms of stimulant withdrawal syndrome within hours to days ► ≥3 symptoms of opioid withdrawal syndrome minutes to days ► ≥3 symptoms of caffeine withdrawal within 24 hours ► ≥3 symptoms of cannabis withdrawal syndromes ► ≥4 symptoms of tobacco withdrawal syndrome within 24 hours Withdrawal symptoms cause distress and/or impairment in functioning	Group of symptoms occurring with cessation or reduction of a substance following a period of prolonged and/or regular use Symptoms are time limited and substance-specific	Presence of symptoms, behaviors or physiologic features associated with withdrawal from a particular substance (differs by substance).
Required number of symptoms	All of the above		
Psychosocial consequences	Distress and/or impairment in functioning		
Exclusions	Other medical condition Substance- or medication-induced, including		Other medical condition Other mental disorder

	withdrawal Other mental disorder		
Symptom specifiers	**Substance-specific specifiers:** Alcohol withdrawal: ***With perceptual disturbance:*** hallucinations occurring in the setting of intact reality testing or hallucinations in the absence of delirium Sedative/hypnotic/anxiolytic withdrawal: ***With perceptual disturbance:*** hallucinations occurring in the setting of intact reality testing or hallucinations in the absence of delirium	Withdrawal state with delirium	For alcohol and sedative, hypnotic, or anxiolytic: With perceptual disturbances With seizures

Table 86.
Hallucinogen Persisting Perception Disorder/Hallucinogen-Induced Psychotic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Hallucinogen persisting perception disorder	Posthallucinogen perception disorder	Hallucinogen-induced psychotic disorder
Symptoms	Re-experiencing, following cessation of use, perceptual symptoms that were present during intoxication Symptoms cause distress or functional impairment	Substance-induced changes in cognition, personality, affect, behavior persisting beyond generally accepted period, with onset of symptoms directly related to use of substance Flashbacks are episodic, short duration, and frequently identical or similar to perceptual distortions experienced during acute intoxication	Psychotic symptoms Psychotic symptoms are in excess of that expected for intoxication
Required number of symptoms	Both of the above		
Psychosocial consequences	Distress or psychosocial impairment		Distress or psychosocial impairment
Exclusions	Other medical condition Substance- or medication-induced, including withdrawal Other mental disorder	Alcohol- or psychoactive substance-induced Korsakov syndrome Alcohol- or psychoactive substance-induced psychotic state	Substance- or medication-induced, including withdrawal Other mental disorder

Table 87.
Gambling Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Gambling disorder	Pathologic gambling	Gambling disorder
Duration	Symptoms present within a 12-month period		Extended period
Symptoms	A pattern of problematic gambling behavior, as exhibited by the following: ▶ Need to gamble with higher stakes over time to experience the same level of excitement ▶ Irritability associated with trying to reduce gambling ▶ A pattern of unsuccessful attempts to stop or reduce gambling ▶ Preoccupation with thoughts related to gambling ▶ A pattern of gambling when feeling distressed ▶ A pattern of gambling to make up for prior losses ▶ A pattern of lying to conceal behaviors ▶ Loss of jobs, relationships, or other opportunities due to gambling behaviors ▶ Relies on others for money to fuel gambling	Pattern of gambling that is dominant in the patient's life and resulting in adverse consequences in social, occupational, family, and other areas of life.	Persistent gambling, with ▶ Loss of control over the behavior ▶ Gambling takes precedence over other activities ▶ Continues/escalates despite negative consequences
Required number of symptoms	≥4		All of the above
Psychosocial consequences	Marked impairment and/or distress		Psychosocial or other impairment
Exclusions	Manic episode	Excessive gambling by manic patients Gambling and betting NOS Gambling in dissocial personality disorder	Substance- or medication-induced, including withdrawal Other mental disorder
Symptom specifiers			Predominantly online Predominantly offline
Course specifiers	**Episodic**: intermittently meeting diagnostic criteria at more than one point over a period of		

	time, with periods of disordered gambling separated by ≥ several months **Persistent:** continuous symptoms and meeting diagnostic criteria for several years **In early remission:** no symptoms over a period of 3 months up to 12 months after the initial diagnosis **In sustained remission:** no symptoms over a period of 12 months or longer after the initial diagnosis
Severity specifiers	**Mild:** 4–5 criteria **Moderate:** 6–7criteria **Severe:** 8–9 criteria

Table 88.
Internet Gaming Disorder/Gaming Disorder

	DSM-TR	ICD-11
Diagnostic name	Internet gaming disorder	Gaming disorder
Duration	1 year	Extended period
Symptoms	Preoccupation with online games Withdrawal when games are taken away Tolerance to the activity (needing more time to play) Failed quit attempts Loss of interest in alternative activities Continued use despite insight into gaming as a problem Deceived others about the amount of gaming Use of games to avoid negative moods	Persistent gaming, with ▶ Loss of control over the behavior ▶ Gaming takes precedence over other activities ▶ Continues/escalates despite negative consequences
Required number of symptoms	≥5	All of the above
Psychosocial consequences	Psychosocial impairment: risked or lost relationships, jobs, or other opportunities due to gaming	Psychosocial or other impairment
Exclusions	Gambling games or pornography sites	Substance- or medication-induced, including withdrawal Other mental disorder
Symptom specifiers		Predominantly online Predominantly offline
Severity specifiers	Mild, moderate, or severe, depending on the number of symptoms and amount of psychosocial consequences	
Comment	Included as a proposed diagnosis under “Conditions for Further Study.”	

17. NEUROCOGNITIVE DISORDERS

Table 89.
Major and Mild Neurocognitive Disorders/Dementias

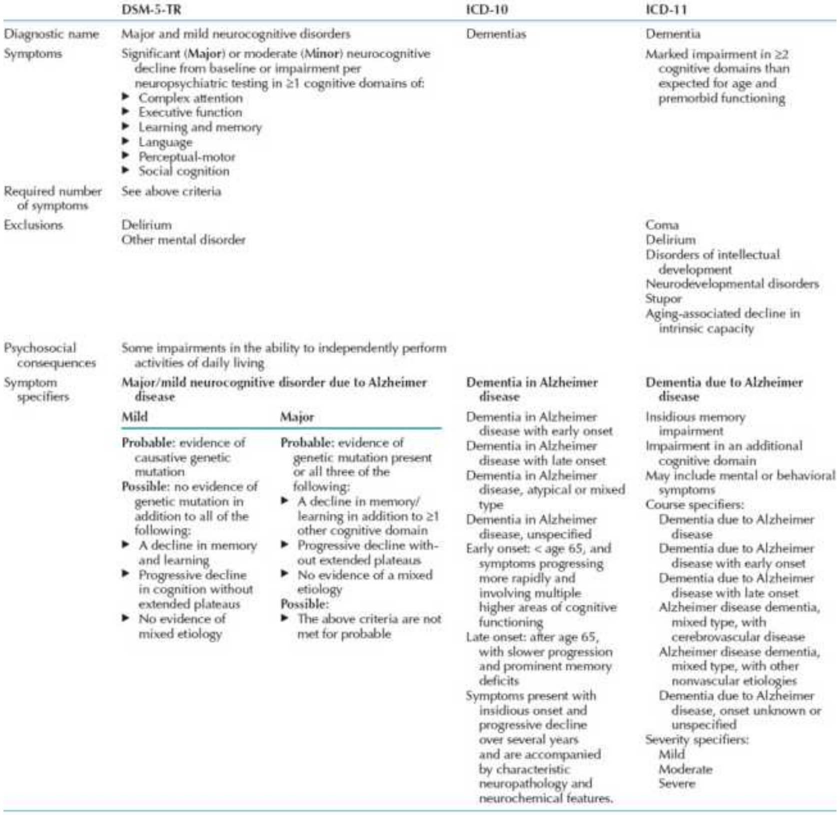

	DSM-5-TR		ICD-10	ICD-11
Diagnostic name	Major and mild neurocognitive disorders		Dementias	Dementia
Symptoms	Significant (**Major**) or moderate (**Minor**) neurocognitive decline from baseline or impairment per neuropsychiatric testing in ≥1 cognitive domains of: ▶ Complex attention ▶ Executive function ▶ Learning and memory ▶ Language ▶ Perceptual-motor ▶ Social cognition			Marked impairment in ≥2 cognitive domains than expected for age and premorbid functioning
Required number of symptoms	See above criteria			
Exclusions	Delirium Other mental disorder			Coma Delirium Disorders of intellectual development Neurodevelopmental disorders Stupor Aging-associated decline in intrinsic capacity
Psychosocial consequences	Some impairments in the ability to independently perform activities of daily living			
Symptom specifiers	**Major/mild neurocognitive disorder due to Alzheimer disease**		**Dementia in Alzheimer disease**	**Dementia due to Alzheimer disease**
	Mild	**Major**		
	Probable: evidence of causative genetic mutation **Possible:** no evidence of genetic mutation in addition to all of the following: ▶ A decline in memory and learning ▶ Progressive decline in cognition without extended plateaus ▶ No evidence of mixed etiology	**Probable:** evidence of genetic mutation present or all three of the following: ▶ A decline in memory/learning in addition to ≥1 other cognitive domain ▶ Progressive decline without extended plateaus ▶ No evidence of a mixed etiology **Possible:** ▶ The above criteria are not met for probable	Dementia in Alzheimer disease with early onset Dementia in Alzheimer disease with late onset Dementia in Alzheimer disease, atypical or mixed type Dementia in Alzheimer disease, unspecified Early onset: < age 65, and symptoms progressing more rapidly and involving multiple higher areas of cognitive functioning Late onset: after age 65, with slower progression and prominent memory deficits Symptoms present with insidious onset and progressive decline over several years and are accompanied by characteristic neuropathology and neurochemical features.	Insidious memory impairment Impairment in an additional cognitive domain May include mental or behavioral symptoms Course specifiers: Dementia due to Alzheimer disease Dementia due to Alzheimer disease with early onset Dementia due to Alzheimer disease with late onset Alzheimer disease dementia, mixed type, with cerebrovascular disease Alzheimer disease dementia, mixed type, with other nonvascular etiologies Dementia due to Alzheimer disease, onset unknown or unspecified Severity specifiers: Mild Moderate Severe

Major/mild frontotemporal neurocognitive disorder		**Dementia in Picks disease** Circumscribed brain atrophy	**Frontotemporal lobar degeneration**
Behavioral variant	**Language variant**		
Spared learning and memory, with a decline in social cognition and/or executive function occurring with ≥3 of the following: ► Disinhibited behavior ► Apathy ► Loss of sympathy/ empathy ► Perseveration or stereotypic behavior ► Hyperorality or dietary changes **Probable:** ► Evidence of possible genetic mutation ► Neuroimaging showing prominent frontal and/or temporal involvement **Possible:** above criteria not met	A decline in language prominent as evidenced by a decline in speech production, word finding, naming, grammar, and/ or comprehension, with relative sparing in learning and memory Probable: ► Evidence of possible genetic mutation ► Neuroimaging showing prominent frontal and/or temporal involvement Possible: The above criteria are not met	Progressive dementia, often beginning in middle age, involving changes in personality and social skills, with the eventual development of deficits in memory, mood, and neurologic function.	Frontotemporal dementia Personally changes (apathy, inappropriate social behavior) Relatively preserved neurocognitive functioning during early stages Variants: Behavioral variant: personality changes, usually apathy and inappropriate social behavior, predominate and usually precede other cognitive changes. Primary progressive aphasia: affects language skills first Motoric variant: progressive motor impairment, with attention, executive functioning, and visuospatial functioning usually affected before memory. Severity specifiers: Mild Moderate Severe
Major or mild neurocognitive disorder with Lewy bodies		**Dementia in other specified diseases classified elsewhere**	**Lewy body disease**
Mild	**Major**		
Core features: ► Fluctuation in cognition with significant variation in attention/alertness ► Recurrent well-formed visual hallucinations ► Parkinsonian symptoms presenting after cognitive decline Suggestive features: ► REM sleep behavior disorder ► Severe neuroleptic sensitivity **Probable:** the presence of two core features or one suggestive feature with ≥1 core feature **Possible:** one of the above three core features present			Dementia due to Lewy body disease Recurrent visual hallucinations Episodic confusion REM sleep behavioral disorder Parkinsonian symptoms Neurocognitive symptoms are progressive and usually affect visuospatial skills, attention, and executive functioning more than memory. Severity specifiers: Mild Moderate Severe
Major or mild vascular neurocognitive disorder		**Vascular dementia**	**Dementia due to cerebrovascular disease**
Mild	**Major**		
In addition to history, imaging, or physical findings consistent with cerebrovascular disease, there is cognitive decline with a pattern suggestive of vascular etiology, as evidenced by ≥1 of the following: ► Cognitive deficits temporally related to cerebrovascular events ► Decline predominantly in complex attention and frontal-executive function **Probable:** ≥1 of the following present ► Neuroimaging showing significant parenchymal injury due to cerebrovascular disease ► Neurocognitive symptoms correlate with documented cerebrovascular events ► Clinical and genetic evidence of cerebrovascular disease present **Possible:** ► No neuroimaging and no documented temporal relationship of cerebrovascular events to cognitive decline are available		**Vascular dementia of acute onset** (symptoms develop after multiple strokes or a single severe stroke) **Multi-infarct dementia** (progressive symptoms following multiple transient ischemic episodes and accumulating infarcts) **Subcortical vascular dementia** (associated with a history of hypertension and foci of injury in white matter, with relatively preserved cerebral cortex) **Mixed cortical and subcortical vascular dementia** **Other vascular dementia** **Vascular dementia, unspecified**	Deficits in: Information processing Complex attention Frontal-executive functioning Evidence of cerebrovascular disease

Major or mild neurocognitive disorder due to traumatic brain injury			**Dementia due to injury to the head**
Mild	**Major**		
Noncognitive symptoms exist in conjunction with a history of traumatic brain injury (symptoms occurring immediately after injury and persisting past the acute period), with ≥1 of the following present: ▶ Loss of consciousness ▶ Post-traumatic amnesia ▶ Disorientation and confusion ▶ Neurologic symptoms or injury on neuroimaging			Symptoms arrive immediately following head trauma or after the individual gains consciousness. Persists beyond acute postinjury period. Cognitive symptoms vary depending on the location of the injury.
Major or mild neurocognitive disorder due to HIV infection		**Dementia in human immunodeficiency virus (HIV)**	**Dementia due to diseases classified elsewhere**
Mild	**Major**		
Neurocognitive symptoms occurring in the presence of HIV diagnosis		Cognitive symptoms developing in the context of known HIV diagnosis	Dementia due to human immunodeficiency virus Confirmed HIV disease No other illness present to explain neurocognitive symptoms Typically subcortical symptoms predominate
Substance/medication-induced major or mild neurocognitive disorder		**Mental and behavioral disorders due to psychoactive substance use**	**Dementia due to psychoactive substances including medications**
Mild	**Major**		
Alcohol ▶ Inhalant ▶ Sedative, hypnotic, or anxiolytic ▶ Other (or unknown)	Alcohol, nonamnestic confabulatory type Alcohol, amnestic confabulatory type Inhalant Sedative, hypnotic, or anxiolytic Other (or known)	Amnesic syndrome Residual and late-onset psychotic disorder	Any dementia that is a direct consequence of substance use Exclusion: dementia due to heavy metals or other toxins
Symptoms of cognitive impairment, occurring not solely in the period of intoxication and/or withdrawal that remain present after a period of abstinence **Persistent:** symptoms remain present after an extended period of abstinence			
Major or mild neurocognitive disorder due to Prion disease		**Dementia in Creutzfeldt–Jakob disease**	**Dementia due to prion disease**
Mild	**Major**		
Symptoms of cognitive impairment, with insidious onset and rapid progression of deficits; symptoms accompanied by motor features (myoclonus, ataxia) or biochemical marker on testing.		Dementia with significant neurologic symptoms, often with a subacute course and progressive rapid decline.	Insidious and rapid onset of symptoms, including cognitive deficits, ataxia, and motor symptoms. Confirmed by brain imaging and spinal fluid or genetic testing. EEG may also support this.
Major or mild neurocognitive disorder due to Parkinson disease		**F02.3 Dementia in Parkinson disease**	
Mild	**Major**		
Neurocognitive symptoms occurring in the context of known Parkinson disease diagnosis, frequently involving insidious onset and gradual progression **Probable** (both of the following met) **Possible** (one of the following met) ▶ No evidence of mixed etiology ▶ Parkinson disease precedes the onset of neurocognitive disorder		Dementia occurring in the context of Parkinson disease.	

	Major or mild neurocognitive disorder due to Huntington disease		**Huntington disease**	**Dementia due to Huntington disease**
	Mild	Major		
	Neurocognitive symptoms occurring in the context of known Huntington disease diagnosis, frequently involving insidious onset and gradual progression		Dementia occurs as a result of diffuse degeneration in the brain related to Huntington disease, caused by an autosomal dominant gene; the disease has slow progression and often involves associated motor symptoms of chorea.	Insidious symptoms in the 30s or 40s, with gradual/slow progression. Typically executive impairments predominate over memory ≥ before motor symptoms (bradykinesia, chorea) begin.
	Major or mild neurocognitive disorder due to another medical condition			
	Major or mild neurocognitive disorder due to multiple etiologies			
	Unspecified neurocognitive disorder			
Severity specifiers	**Mild:** impairments in more complex activities of daily living (i.e., money management) **Moderate:** impairment in basic activities of daily living (i.e., eating, dressing) **Severe:** fully dependent on others for all daily activities			
				Other diagnoses included in ICD-11: Dementia due to Parkinson disease Dementia due to exposure to heavy metals and other toxins Dementia due to multiple sclerosis Dementia due to normal pressure hydrocephalus Dementia due to pellagra Dementia due to Down syndrome
Additional symptom specifiers (common to all)	**With or without behavioral disturbance**			

Table 90.
Delirium

	DSM-5-TR	**ICD-10**	**ICD-11**
Diagnostic name	Delirium	Delirium, not induced by alcohol and other psychoactive substances Delirium not superimposed on dementia Delirium superimposed on dementia Other delirium Delirium, unspecified	Delirium
Symptoms	Disturbance in attention and awareness Symptoms develop acutely, represent a change from baseline, and may fluctuate Additional disturbance in cognitive function, memory, language, orientation, and/or perception History, physical exam, labs, or imaging indicative of the physiologic cause of cognitive disturbance	Nonspecific organic cognitive syndrome with disturbed consciousness, awareness, memory, motor function, sleep, and emotion, with a variable duration and severity; it can be acute or subacute	Disturbance of attention, orientation, and awareness Develops quickly Change from baseline functioning It can present as confusion or global neurocognitive impairment May fluctuate Often includes behavioral and emotional disturbances May include disturbance of the sleep/wake cycle
Required number of symptoms	All of the above		
Exclusions	Neurocognitive disorder Reduced arousal (i.e., coma)		Result of: ▸ Neurocognitive disorder ▸ Another mental disorder ▸ Substance intoxication
Symptom specifiers	Substance intoxication delirium, with (mild, mod, or severe) or without use disorder Substance withdrawal delirium, with (mild, mod, or severe) or without use disorder Medication-induced delirium Delirium due to another medical condition Delirium due to multiple etiologies Hyperactive Hypoactive		Delirium due to disease classified elsewhere Delirium due to psychoactive substances including medications Delirium due to multiple etiologic factors Delirium, other specified cause Delirium, unspecified or unknown cause

	Mixed level of activity	
Course specifiers	**Acute** (lasting hours to days) **Persistent** (lasting weeks to months)	
Comments		If due to substance use, use “Mental and behavioral disorders due to psychoactive substances”

18. PERSONALITY DISORDERS

Table 91.
General Personality Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	General personality disorder	Disorders of adult personality and behavior	Personality disorder
Duration	Long duration of time, beginning ≥ in adolescence or early adulthood		≥2 years, tends to begin in childhood
Symptoms	A fixed and pervasive pattern of experience and behavior, as manifested in: ▶ Cognition and perception ▶ Affect ▶ Interpersonal functioning ▶ Impulsivity	Includes a variety of patterns of behavior that are persistent and involve a clinical disturbance in interpersonal relationships and views of self. These patterns of behavior and experience are fixed and result in inflexibility	Problems in function in aspects of the self or interpersonal dysfunction Patterns of cognition, emotional experience and expression, and maladaptive behaviors Disturbance occurs in a range of personal/social situations, although certain situations may provoke
Required number of symptoms	≥2 of the above being long-standing and present across a range of social situations		
Exclusions	Other medical condition Substance- or medication-induced, including withdrawal Other mental disorder		Other medical/nervous system conditions Substance- or medication-induced, including withdrawal Other mental disorder Developmentally appropriate behavior
Psychosocial consequences	Distress or psychosocial impairment	Frequently (but not always) associated with distress and difficulties with social performance	
Symptom specifiers			Categorized by traits or patterns: **Negative affectivity in personality disorder or personality difficulty:** tendency to experience a broad range of negative emotions. **Detachment in personality disorder or personality difficulty:** a tendency

to maintain interpersonal and emotional distance.

Dissociality in personality disorder or personality difficulty: disregard for rights and feelings of others. Self-centered and lacking empathy.

Disinhibition in personality disorder or personality difficulty: acts rashly and impulsively, without consideration of consequences.

Anankastia in personality disorder or personality difficulty: having a rigid focus, commonly manifested by perfectionism and emotional and behavioral constraint.

Borderline pattern: included as an additional specifier, which typically overlaps with the above specifiers. This is meant to indicate a pervasive pattern of instability in interpersonal relationships, self-image, emotions, and a tendency toward impulsivity.

Symptoms include ≥5 of the following:

- Avoidance of abandonment
- Unstable and intense interpersonal relationships
- Identity disturbance
- Acting rashly, often with risky activities (sexual behavior, driving, drinking, or eating)

			▶ Self-harm ▶ Emotional instability ▶ Feeling empty inside ▶ Difficulty controlling anger ▶ Transient dissociative episodes
Severity specifiers			Severity depends on the range of areas affected
Comments	Includes: **Cluster A:** paranoid personality disorder, schizoid personality disorder, schizotypal personality disorder **Cluster B:** antisocial personality disorder, borderline personality disorder, histrionic personality disorder, narcissistic personality disorder **Cluster C:** avoidant personality disorder, dependent personality disorder, obsessive-compulsive personality disorder	Includes: **Specific personality disorders:** Paranoid personality disorder, schizoid personality disorder, dissocial personality disorder, emotionally unstable personality disorder, histrionic personality disorder, anankastic personality disorder, anxious (avoidant) personality disorder, dependent personality disorder, other specific personality disorders (eccentric, haltlose, immature, narcissistic, passive-aggressive, psychoneurotic), personality disorder unspecified **Mixed and other personality disorders** **Enduring personality changes:** due to brain damage and disease, after a catastrophic experience, after psychiatric illness, chronic pain personality syndrome, enduring personality change unspecified	

Table 92.
Specific Symptoms and Exclusions for Individual Personality Disorders

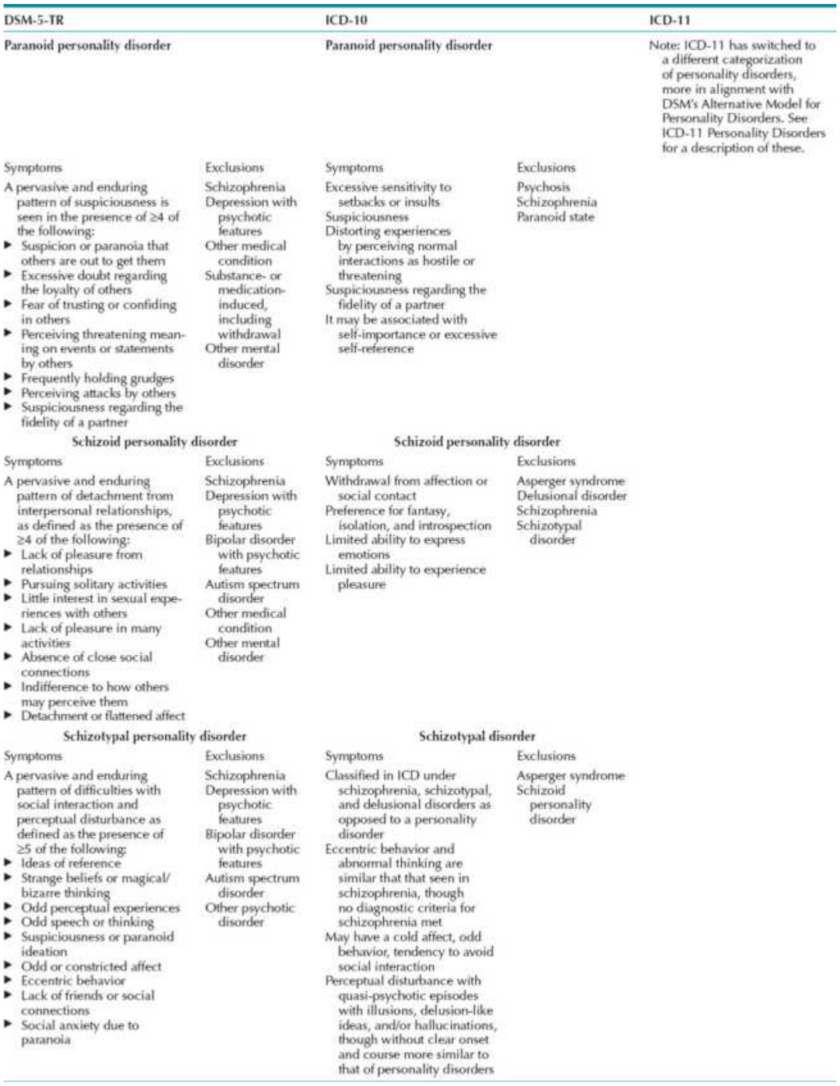

DSM-5-TR		ICD-10		ICD-11
Paranoid personality disorder		**Paranoid personality disorder**		Note: ICD-11 has switched to a different categorization of personality disorders, more in alignment with DSM's Alternative Model for Personality Disorders. See ICD-11 Personality Disorders for a description of these.
Symptoms	Exclusions	Symptoms	Exclusions	
A pervasive and enduring pattern of suspiciousness is seen in the presence of ≥4 of the following: ▶ Suspicion or paranoia that others are out to get them ▶ Excessive doubt regarding the loyalty of others ▶ Fear of trusting or confiding in others ▶ Perceiving threatening meaning on events or statements by others ▶ Frequently holding grudges ▶ Perceiving attacks by others ▶ Suspiciousness regarding the fidelity of a partner	Schizophrenia Depression with psychotic features Other medical condition Substance- or medication-induced, including withdrawal Other mental disorder	Excessive sensitivity to setbacks or insults Suspiciousness Distorting experiences by perceiving normal interactions as hostile or threatening Suspiciousness regarding the fidelity of a partner It may be associated with self-importance or excessive self-reference	Psychosis Schizophrenia Paranoid state	
Schizoid personality disorder		**Schizoid personality disorder**		
Symptoms	Exclusions	Symptoms	Exclusions	
A pervasive and enduring pattern of detachment from interpersonal relationships, as defined as the presence of ≥4 of the following: ▶ Lack of pleasure from relationships ▶ Pursuing solitary activities ▶ Little interest in sexual experiences with others ▶ Lack of pleasure in many activities ▶ Absence of close social connections ▶ Indifference to how others may perceive them ▶ Detachment or flattened affect	Schizophrenia Depression with psychotic features Bipolar disorder with psychotic features Autism spectrum disorder Other medical condition Other mental disorder	Withdrawal from affection or social contact Preference for fantasy, isolation, and introspection Limited ability to express emotions Limited ability to experience pleasure	Asperger syndrome Delusional disorder Schizophrenia Schizotypal disorder	
Schizotypal personality disorder		**Schizotypal disorder**		
Symptoms	Exclusions	Symptoms	Exclusions	
A pervasive and enduring pattern of difficulties with social interaction and perceptual disturbance as defined as the presence of ≥5 of the following: ▶ Ideas of reference ▶ Strange beliefs or magical/bizarre thinking ▶ Odd perceptual experiences ▶ Odd speech or thinking ▶ Suspiciousness or paranoid ideation ▶ Odd or constricted affect ▶ Eccentric behavior ▶ Lack of friends or social connections ▶ Social anxiety due to paranoia	Schizophrenia Depression with psychotic features Bipolar disorder with psychotic features Autism spectrum disorder Other psychotic disorder	Classified in ICD under schizophrenia, schizotypal, and delusional disorders as opposed to a personality disorder Eccentric behavior and abnormal thinking are similar that that seen in schizophrenia, though no diagnostic criteria for schizophrenia met May have a cold affect, odd behavior, tendency to avoid social interaction Perceptual disturbance with quasi-psychotic episodes with illusions, delusion-like ideas, and/or hallucinations, though without clear onset and course more similar to that of personality disorders	Asperger syndrome Schizoid personality disorder	

Antisocial personality disorder		**Dissocial personality disorder**	
Symptoms	Exclusions	Symptoms	Exclusions
The pervasive and enduring presence of ≥3 of the following, occurring since the age of 15 and diagnosed no earlier than the age of 18, with evidence of conduct disorder before the age of 15: ▶ Not following social norms with regard to the law ▶ A pattern of deceitfulness and lying ▶ Impulsivity ▶ Irritability and/or aggressiveness ▶ Blatant disregard for the safety of self or others ▶ Irresponsibility with regard to work, social, and personal obligations ▶ Lack of remorse	Schizophrenia Bipolar disorder	Disregard for social expectations or obligations Lack of concern for the feelings of others Behaviors starkly different from those expected per social norms Low distress/frustration tolerance Low threshold for aggression or violence Tendency to blame others	Conduct disorders Emotionally unstable personality disorder
Borderline personality disorder		**Emotionally unstable personality disorder**	
Symptoms		Symptoms	Exclusions
A pervasive and enduring pattern of interpersonal conflict and impulsivity as evidenced by ≥5 of the following: ▶ An excessive effort to avoid real or feared abandonment ▶ Unstable and frequently intense interpersonal relationships, with a pattern of alternating between idealization and devaluation ▶ Unstable image of self ▶ Impulsivity resulting in self-harm ▶ Pattern of suicidal behaviors, gestures, threats, or self-injury ▶ Affective lability ▶ Chronic feelings of emptiness ▶ Difficulties managing anger ▶ Intermittent stress-related paranoia and/or feelings of dissociation		Tendency to impulsiveness Unpredictable and labile mood Behavioral and/or emotional outbursts are common Frequent conflict with others Two distinct "Types" include: ▶ Impulsive type (emotional lability and lack of impulse control) ▶ Borderline type (disturbance of self-image, self-worth, and interpersonal relationships, with associated self-destructive behavior)	Dissocial personality disorder
Histrionic personality disorder		**Histrionic personality disorder**	
A pervasive and enduring pattern of needing attention and experiencing strong emotions as evidenced by ≥5 of the following: ▶ Discomfort in situations in which the individual is not the main object of attention ▶ Excessively sexual or provocative behavior in interpersonal interaction ▶ Labile emotions ▶ Use of physical appearance to draw attention ▶ Impressionistic style of speech ▶ Exaggerated expressions and speech ▶ High suggestibility ▶ Perceives relationships as being more intimate than they actually are		Shallow and labile affect Self-dramatization Exaggeration of speech and affect Suggestibility Egocentricity Lack of concern for others Attention seeking Easily offended or hurt emotionally	

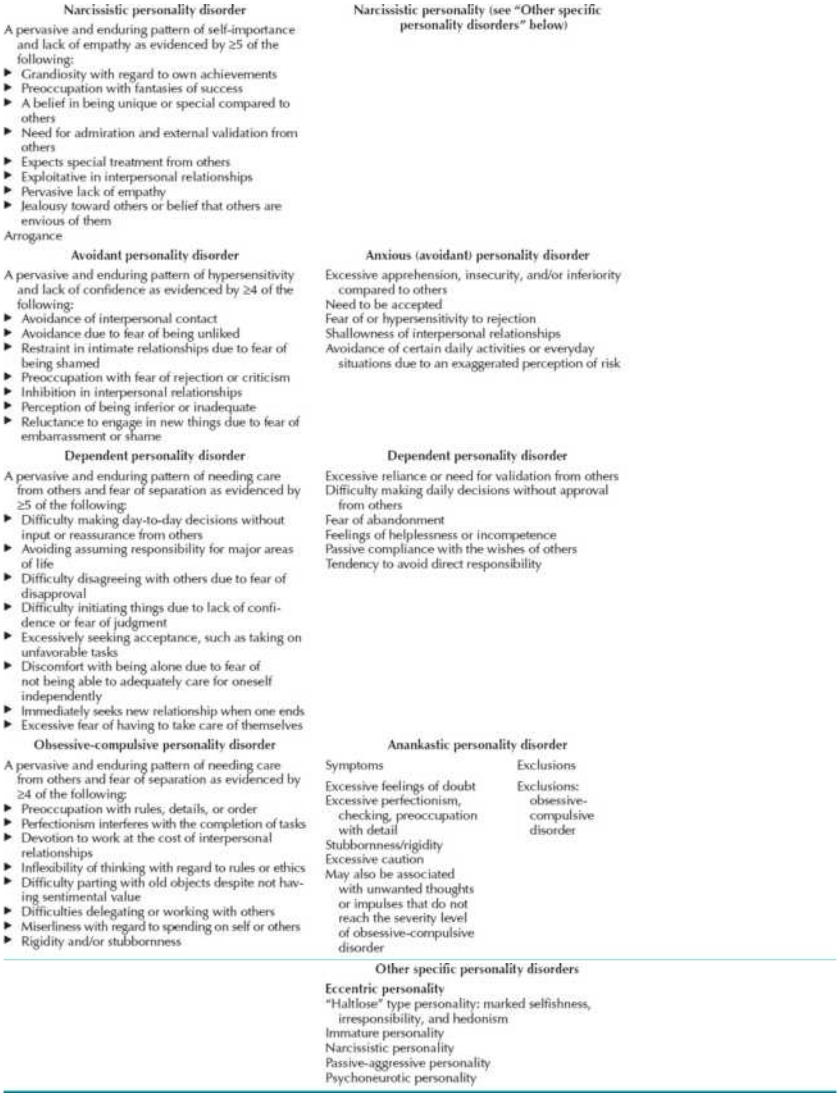

Narcissistic personality disorder A pervasive and enduring pattern of self-importance and lack of empathy as evidenced by ≥5 of the following: ▶ Grandiosity with regard to own achievements ▶ Preoccupation with fantasies of success ▶ A belief in being unique or special compared to others ▶ Need for admiration and external validation from others ▶ Expects special treatment from others ▶ Exploitative in interpersonal relationships ▶ Pervasive lack of empathy ▶ Jealousy toward others or belief that others are envious of them Arrogance	**Narcissistic personality (see "Other specific personality disorders" below)**
Avoidant personality disorder A pervasive and enduring pattern of hypersensitivity and lack of confidence as evidenced by ≥4 of the following: ▶ Avoidance of interpersonal contact ▶ Avoidance due to fear of being unliked ▶ Restraint in intimate relationships due to fear of being shamed ▶ Preoccupation with fear of rejection or criticism ▶ Inhibition in interpersonal relationships ▶ Perception of being inferior or inadequate ▶ Reluctance to engage in new things due to fear of embarrassment or shame	**Anxious (avoidant) personality disorder** Excessive apprehension, insecurity, and/or inferiority compared to others Need to be accepted Fear of or hypersensitivity to rejection Shallowness of interpersonal relationships Avoidance of certain daily activities or everyday situations due to an exaggerated perception of risk
Dependent personality disorder A pervasive and enduring pattern of needing care from others and fear of separation as evidenced by ≥5 of the following: ▶ Difficulty making day-to-day decisions without input or reassurance from others ▶ Avoiding assuming responsibility for major areas of life ▶ Difficulty disagreeing with others due to fear of disapproval ▶ Difficulty initiating things due to lack of confidence or fear of judgment ▶ Excessively seeking acceptance, such as taking on unfavorable tasks ▶ Discomfort with being alone due to fear of not being able to adequately care for oneself independently ▶ Immediately seeks new relationship when one ends ▶ Excessive fear of having to take care of themselves	**Dependent personality disorder** Excessive reliance or need for validation from others Difficulty making daily decisions without approval from others Fear of abandonment Feelings of helplessness or incompetence Passive compliance with the wishes of others Tendency to avoid direct responsibility
Obsessive-compulsive personality disorder A pervasive and enduring pattern of needing care from others and fear of separation as evidenced by ≥4 of the following: ▶ Preoccupation with rules, details, or order ▶ Perfectionism interferes with the completion of tasks ▶ Devotion to work at the cost of interpersonal relationships ▶ Inflexibility of thinking with regard to rules or ethics ▶ Difficulty parting with old objects despite not having sentimental value ▶ Difficulties delegating or working with others ▶ Miserliness with regard to spending on self or others ▶ Rigidity and/or stubbornness	**Anankastic personality disorder** Symptoms Excessive feelings of doubt Excessive perfectionism, checking, preoccupation with detail Stubbornness/rigidity Excessive caution May also be associated with unwanted thoughts or impulses that do not reach the severity level of obsessive-compulsive disorder Exclusions Exclusions: obsessive-compulsive disorder
	Other specific personality disorders **Eccentric personality** "Haltlose" type personality: marked selfishness, irresponsibility, and hedonism Immature personality Narcissistic personality Passive-aggressive personality Psychoneurotic personality

19. PARAPHILIC DISORDERS

Table 93.
Voyeuristic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Voyeuristic disorder	Voyeurism	Voyeuristic disorder
Duration	6 months or longer		
Symptoms	Sexual arousal, sexual fantasies, sexual urges resulting from observing an unknowing person undressing or engaging in sexual activity The individual must be 18 years or older at the time of diagnosis	A pattern of watching people engage in sexual behaviors or other intimate behaviors, such as undressing, with the individual experiencing sexual pleasure from the experience; those being observed are often not aware	Sustained focus and pattern or arousal from observing an unsuspecting person naked or disrobing Has acted on these thoughts or been distressed by them
Exclusions			Children or (usually) adolescents Single or limited instances
Psychosocial consequences	Distress or psychosocial impairment		
Symptom specifiers	***In a controlled environment (if the individual lives in an institution or other controlled environment)***		
Course specifiers	**In full remission** (no action on urges and no related distress or impairment over ≥5 years while in a noncontrolled environment)		

Table 94.
Exhibitionistic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Exhibitionistic disorder	Exhibitionism	Exhibitionistic disorder
Duration	6 months or longer		
Symptoms	Sexual arousal, sexual fantasies, sexual urges, and sexual behaviors resulting from the act of exposing one's genitals to unsuspecting individuals	A pattern of exposing genitals to strangers, often of the opposite sex, or exposing genitals in public places; such exposure is frequently associated with sexual excitement/arousal and may be followed by masturbation.	Sustained and focused pattern or arousal from exposing one's genitals to unsuspecting person in public Has acted on these thoughts or been distressed by them
Exclusions			Children or (usually) adolescents Single or limited instances
Psychosocial consequences	Distress or psychosocial impairment		
Symptom specifiers	**Sexually aroused by exposing genitals to prepubertal children** Sexually aroused by exposing genitals to physically mature individuals Sexually aroused by exposing genitals to prepubertal children and to physically mature individuals **In a controlled environment** (if the individual lives in an institution or other controlled environment)		
Course specifiers	**In full remission** (no action on urges and no related distress or impairment over ≥5 years while in a noncontrolled environment)		

Table 95.
Frotteuristic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Frotteuristic disorder	Other disorders of sexual preference ***Frotteurism***	Frotteuristic disorder
Duration	6 months or longer		
Symptoms	Sexual arousal, sexual fantasies, sexual urges, and sexual behaviors resulting from touching or rubbing against a nonconsenting individual	Sexual arousal through rubbing against people in public spaces	Sustained and focused pattern or arousal from touching or rubbing against a nonconsenting person Has acted on these thoughts or been distressed by them
Exclusions			Children or (usually) adolescents Single or limited instances
Psychosocial consequences	Distress or psychosocial impairment		
Symptom specifiers	**In a controlled environment** (if the individual lives in an institution or other controlled environment)		
Course specifiers	**In full remission** (no action on urges and no related distress or impairment over ≥5 years while in a noncontrolled environment)		
Comments		Also includes ***Necrophilia,*** in addition to behaviors such as making obscene phone calls, engaging in sexual activity with animals, and use of strangulation for sexual excitement	

Table 96.
Sexual Masochism Disorder/Sadomasochism

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Sexual masochism disorder	Sadomasochism	Paraphilic disorder involving solitary behavior or consenting individuals
Duration	6 months or longer		
Symptoms	Sexual arousal, sexual fantasies, and sexual urges resulting from being humiliated, hit, bound, or made to suffer.	Preference for sexual activity to include bondage or the infliction of pain or humiliation as the recipient (masochism) or as the one inflicting on another (sadism), with associated sexual excitement	General category for a pattern or atypical sexual arousal Causes distress and significant risk of injury
Psychosocial consequences	Distress or psychosocial impairment		Marked distress
Symptom specifiers	**With asphyxiophilia** (sexual arousal related to restriction of breathing) **In a controlled environment** (if the individual lives in an institution or other controlled environment)		
Course specifiers	**In full remission** (no action on urges and no related distress or impairment over ≥5 years while in a noncontrolled environment)		

Table 97.
Sexual Sadism Disorder/Sadomasochism

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Sexual sadism disorder	Sadomasochism	Coercive sexual sadism disorder
Duration	6 months or longer		
Symptoms	Sexual arousal, sexual fantasies, and sexual urges resulting from inflicting psychological suffering or humiliation on others.	Preference for sexual activity to include bondage or the infliction of pain or humiliation as the recipient (masochism) or as the one inflicting on another (sadism), with associated sexual excitement	Sustained and focused pattern or arousal from inflicting physical/psychological suffering on a nonconsenting person Has acted on these thoughts or been distressed by them
Exclusions			Children or (usually) adolescents Single or limited instances
Symptom specifiers	**In a controlled environment** (if the individual lives in an institution or other controlled environment)		
Course specifiers	**In full remission** (no action on urges and no related distress or impairment over ≥5 years while in a noncontrolled environment)		

Table 98.
Pedophilic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Pedophilic disorder	Pedophilia	Pedophilic disorder
Duration	6 months or longer		
Symptoms	Sexual arousal, sexual fantasies, sexual urges, and sexual behaviors involving sexual activity with prepubescent child or children Individual must be 16 years or older at the time of diagnosis and must also be a minimum of 5 years older than the child of sexual interest	Sexually preferring young boys and girls, often those of prepubertal or early pubertal age	Sustained and focused pattern or arousal from prepubertal children Has acted on these thoughts or been distressed by them
Required number of symptoms	Both of the above		
Exclusions	Individual of adolescent age in a relationship with another child aged 12 or 13		Children or (usually) adolescents Single or limited instances Cultural variations in acceptable behavior
Psychosocial consequences	Distress or psychosocial impairment		
Symptom specifiers	**Exclusive type:** attracted only to children Nonexclusive type Sexually attracted to males Sexually attracted to females Sexually attracted to both males and females Limited to incest		

Table 99.
Fetishistic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Fetishistic disorder	Fetishism	Paraphilic disorder involving solitary behavior or consenting individuals
Duration	≥6 months		
Symptoms	Sexual arousal, sexual fantasies, sexual urges, and sexual behaviors resulting from inanimate objects or a highly specific nongenital body part	Using nonliving objects as a source of sexual arousal or gratification or a way to enhance sexual excitement.	General category for a pattern or atypical sexual arousal. Causes distress and significant risk of injury.
Exclusions	Limited to: ▸ Clothing related to cross-dressing (transvestic disorder) ▸ Use of devices related to genital stimulation		
Psychosocial consequences			Marked distress
Symptom specifiers	Body part(s) Nonliving object(s) Other **In a controlled environment** (if the individual lives in an institution or other controlled environment)		
Course specifiers	**In full remission** (no action on urges and no related distress or impairment over ≥5 years while in a noncontrolled environment)		

Table 100.
Transvestic Disorder

	DSM-5-TR	ICD-10	ICD-11
Diagnostic name	Transvestic disorder	Dual-role transvestism Fetishistic transvestism	Paraphilic disorder involving solitary behavior or consenting individuals
Duration	6 months or longer		
Symptoms	Sexual arousal, sexual fantasies, and sexual urges resulting from cross-dressing	***Dual role:*** enjoyment associated with temporarily wearing clothes of the opposite sex, but without associated sexual excitement or associated desire for permanent sex change ***Fetishistic:*** wearing clothes of the opposite sex associated with sexual excitement and desire to remove clothing once orgasm occurs and arousal declines	General category for a pattern or atypical sexual arousal. Causes distress and significant risk of injury
Psychosocial consequences	Distress or psychosocial impairment		Marked distress
Symptom specifiers	**With fetishism:** sexual arousal from specific fabrics or garments) **With autogynephilia:** sexual arousal from imagining self as a female **In a controlled environment:** if the individual lives in an institution or other controlled environment		
Course specifiers	**In full remission:** no action on urges and no related distress/impairment ≥5 years while in a noncontrolled environment		

Index

Page numbers followed by *f* indicate figures; numbers followed by *t* indicate tables. Page numbers in **boldface** indicate major discussions.

A

- A2 noradrenergic neurons, 227
- AA. *See* Alcoholics Anonymous (AA)
- AACDP. *See* American Association of Chairmen of Departments of Psychiatry (AACDP)
- AACP. *See* American Academy of Child and Adolescent Psychiatry (AACAP)
- AADPRT. *See* American Association of Directors of Psychiatric Residency Training, Inc. (AADPRT)
- AAIDD. *See* American Association on Intellectual and Developmental Disabilities (AAIDD)
- AAMC. *See* Association of American Medical Colleges (AAMC)
- AAP. *See* American Academy of Pediatrics (AAP)
- AARP. *See* American Association of Retired People (AARP)
- Abdominal breathing, 2171
- Abecarnil, 3255
- Aberrant learning theory, 444
- Abilify. *See* Aripiprazole

- Ability, 4708
- Ablation, 4708
- Ablutomania, 4708
- ABMS. *See* American Board of Medical Specialties (ABMS)
- Abnormal, 4708
- Abnormal beliefs and experiences, 476
- Abnormal eating, 300
- Abnormal illness behaviors, 1193–1194
- Abnormal Involuntary Movement Scale (AIMS), 1098–1099, 1099*t*, 3813
- Abnormalities in agency, 476
- Abnormal reflexes, 470
- Abnormal social behavior, 476
- Abortion, 2589–2591, 2591*t*, 4708
- Above normal, 785–788
- ABPN. *See* American Board of Psychiatry and Neurology (ABPN)
- Abraham, Karl, 917, 1653, 2272
- Abrams, Richard, 4656, 4660
- Abreaction, 2959*t*, 4708
- Absence seizure, 4708
- Absolute refractory periods, 70
- Absolute threshold, 4708
- Abstinence, 4708
- Abstinence syndrome, 1420, 4708
- Abstinence threshold, 1386
- Abstraction, 4708
- Abstract thinking, 4708
- Abulia, 1183, 1458, 4708
- Abuse, 2521. *See also* Child abuse; Elder abuse; Maltreatment (child); Physical abuse; Sexual abuse
  - cognitive effects of, 2521
  - defined, 4708

  - psychological effects of, 2521
- Abusive behaviors, modeling of, 2521
- ACA. *See* Affordable Care Act (ACA)
- Academic or educational problem, 2537–2539. *See also* Educational system
  - clinical features, 2538
  - definition of, 2537–2538
  - differential diagnosis, 2538
  - etiology of, 2538
  - family therapy for, 3911*t*
  - identification of, 2538
  - treatment of, 2539
- Academy of Consultation-Liaison Psychiatry, 2278
- Acalculia, 4708
- Acamprosate (Campral), 92, **3372–3377**
  - absorption, 3373
  - for alcohol use disorders, 1310
  - blood concentrations, 3374
  - chemical structure, 3373*f*
  - chemistry, 3373
  - distribution, 3373
  - dosage and administration, 3377
  - drug interactions, 3377
  - drug studies, 3374–3376
  - elimination, 3373
  - history, 3372
  - mechanism of action, 3373–3374
  - metabolism, 3373
  - pharmacodynamics, 3373–3374
  - pharmacokinetics, 3373
  - pharmacologic actions, 3373–3374
  - precautions/adverse reactions, 3377

  - therapeutic indications, 3376–3377
- Acarophobia, 2432, 4708
- Acatalepsia, 4708
- Acataphasia, 4708
- Acathexis, 4708
- ACC. *See* Anterior cingulate cortex (ACC)
- Accelerated experiential dynamic psychotherapy (AEDP), 2772
- Accelerated interaction, 4708
- Accent method, 3709
- Acceptance, 2901, 2959*t*
- Acceptance and commitment therapy (ACT), 1827, 2311, 2819, **2900–2902**
  - acceptance, 2901
  - for anorexia nervosa, 2161–2162
  - being present, 2902
  - for bulimia nervosa, 2162
  - cognitive defusion, 2901–2902
  - committed action, 2902
  - efficacy and effectiveness of, 2990
  - evidence, 2900
  - examples, 2901*t*
  - functional contextualism, 2900
  - hexaflex, 2901*f*
  - for insomnia disorder, 2172
  - processes in, 2900–2901, 2901*t*
  - relational frame theory, 2900
  - self-as-context, 2902
  - values, 2902
- Acceptance-based interventions, 4365
- Accident proneness, 4708
- Accommodations, 745, 4026*t*, 4032*t*
- Accountable care organizations (ACOs), 4401–4402, 4500

- Accreditation, 4708
- Accreditation Council for Graduate Medical Education (ACGME), 2278, 2741, 4588
- Acculturation, 849
- Acenesthesia, 4708
- Aceruloplasminemia, 705–708, 708*f*
- ACEs. *See* Adverse childhood experiences (ACEs)
- Acetaldehyde, 1298
- Acetaminophen (Tylenol), 1060
  - for migraine, 651*t*
  - pain management in burn patients, 2470
- Acetylation, 236
- Acetylcholine, 4708
  - anatomy and function, 104
  - life cycle, 104
  - schematic model of synapse, 105*f*
  - schizophrenia pathology, 1506
  - synthesis, 105–106, 105*f*
- Acetylcholine receptors
  - muscarinic, 104–106
  - nicotinic, 106
- Acetyl-L-carnitine (ALC), 2634
- Achenbach System of Empirically Based Assessment (ASEBA), 3542
- Achievement testing, 3566–3567
  - assessment, 3566–3567
  - comprehensive, 3567
  - general, 3567
  - specific reading tests, 3567
- Achluophobia, 4708
- Acid, 4708
- Ackermann, Nathan, 2943, 2956, 4648

- Ackner, Brian, 1931
- ACLP. *See* Academy of Consultation-Liaison Psychiatry (ACLP)
- Acne, 2442–2443
  - clinical features, 2442
  - pathophysiology of, 2442
  - psychiatric comorbidity, 2443
- Aconitine, 73
- ACOs. *See* Accountable care organizations (ACOs)
- ACP. *See* American College of Physicians (ACP)
- Acquired immunodeficiency syndrome (AIDS). *See also* HIV (human immunodeficiency virus)
  - antiretroviral medications and drug classes for, 588*t*
  - child/adolescent
    - clinical features, 3995
    - epidemiology of, 3993–3994
    - etiology of, 3994
    - neurodevelopmental sequelae, 3995
    - treatment of, 3995–3997
  - defined, 4709
  - defining conditions, 583
  - description of, 583–585
  - epidemiology of, 585–587
  - in gay men, 2085
  - mania, 595
  - neuropsychiatric aspects of, 583–609
  - viral life cycle and treatment of, 587–589
- Acquired stuttering, 467*t*
- Acromegaly, 711, 4709
- Acrophobia, 4709
- ACT. *See* Acceptance and commitment therapy (ACT); Action and commitment therapy (ACT)
- ACTH secreting tumors, 711

- Actigraphy, 2189, 4083
- Acting out, 920*t*, 2748*t*, 4709
- Action and commitment therapy (ACT), 2898
- Action potentials, 69–70
  - axon conduction, 70
- Active chromatin, 277–278
- Active euthanasia, 2409
- Active imagination, 2984
- Active membrane properties, 69–70
- Active sleep, 200
- Activities of daily living (ADLs), 4171. *See also* Everyday functioning
  - driving, 4180
  - instrumental, 4179–4180
- Activity group psychotherapy, 4709
- Actualization, 4709
- Aculalia, 4709
- Acupuncture, 2647–2648, 4146
- Acute akathisia, medication-induced, 3095
- Acute and transient psychotic disorder, **1627–1631**
  - comparative nosology, 1627
  - course and prognosis, 1631
  - diagnosis and clinical features of, 1630
  - diagnostic and clinical features, 4782–4783*t*
  - differential diagnosis, 1630–1631
  - epidemiology of, 1629
  - etiology of, 1629–1630
  - history of, 1627, 1628*t*
- Acute dissociative reactions to stressful events, 2021–2024
  - acute mixed dissociative psychotic states, 2022–2023
  - definition, 2021–2022
  - diagnosis and clinical features, 2022

  - differential diagnosis, 2023
  - pathology and laboratory examination, 2023
  - treatment of, 2023–2024
- Acute dystonia, 3094–3095
  - anticholinergics and, 3327
  - clinical presentation, 3094
  - pathophysiology of, 3094
  - treatment options/implications, 3094–3095
- Acute intermittent porphyria (AIP), 691
- Acute mania, 1690–1691
  - carbamazepine for, 3286–3288
  - clinical recommendations, 1731–1733
  - management in an emergency setting, 1731–1732
  - pharmacological treatment of, 1732–1733, 1732*t*
  - treatment selection, 1728–1731
  - valproate for, 3301–3303, 3302*t*
- Acute pain disorders, hypnosis and, 2931
- Acute phase response proteins (ARP), 174–175, 174*f*
- Acute stress, 181
  - and burn injuries in children, 2477
- Acute stress disorder (ASD), 4709
  - in cancer patients, 2354
  - diagnosis of, 4802*t*
  - dissociative amnesia and, 2007, 2009
  - terrorism-induced trauma and, 4061
- Acute tryptophan depletion (ATD), 233
- AD. *See* Alzheimer disease (AD)
- ADA. *See* Americans with Disabilities Act (ADA)
- Adams, Joseph, 4685
- Adams, Raymond, 4658
- Adaptability, 2213
- Adaptation, 4709

- Adaptational approach, 4709
- Adapted child, 963
- Adaptive behavior
  - common measures of, 3635*t*
  - defined, 4709
- Adaptive functioning, 2528–2529, 3632–3633
- Adaptive immunity, 176–177
  - activation phase, 176
  - effector phase, 176
  - recognition phase, 176–177
  - schizophrenia and, 186
  - T cells, 176–177
- ADD. *See* Attention-deficit disorder (ADD)
- ADD-H Comprehensive Teacher Rating Scale (ACTeRS), 3546
- Addiction. *See also* Substance dependence
  - brain pathways in, 444–450
  - cellular and molecular substrates of, 450–454
    - epigenetics, 453–454
    - gene expression, 450–452
    - neuronal morphology, 452–453
    - synaptic plasticity, 452–453
  - defined, 4709
  - definition of, 2122–2123
  - endocannabinoids and, 135
  - group psychotherapy for, 3900–3901
  - ICD-11 criteria for, 1224
  - learning theory and, 438–439
  - theories of, 443–444
    - aberrant learning theory, 444
    - frontostriatal-dysfunction theory, 444
    - hedonic allostasis theory, 444
    - incentive sensitization theory, 444

    - psychomotor stimulant theory, 443–444
- Addiction Severity Index (ASI), 1111
- Addictive disorders
  - hypnosis for, 2937–2938
  - psychiatric examination and, 992–993
  - rating scales for, 1109–1111
  - after TBI, 552
- Addictive process, 2126–2133
  - adverse childhood events or conditions and, 2129–2130
  - affect regulation/shame, 2127–2128
  - behavioral inhibition, 2128
  - caregiving deficiencies during infancy and, 2128
  - character pathology, 2131
  - definition of, 2126–2127
  - development of, 2133
  - genetics of, 2127
  - healing of, 2139–2145
    - affect-regulating agents, 2139–2140
    - healthy lifestyle, 2144–2145
    - meditation, 2143–2144
    - mu-opioid receptor antagonists, 2140
    - psychiatric pharmacotherapy, 2139–2140
    - psychodynamic psychotherapy, 2140–2143
    - yoga, 2144
  - impaired ability to symbolize affects and, 2132
  - impaired self-regulation and, 2130–2131
  - maternal care deficiencies and, 2129
  - maternal gestational stress, 2128
  - maternal separation and, 2128–2129
  - motivation-reward, 2127
  - neurobiologic formulation, 2126–2127
  - phases of, 441–443

    - animal models of, 442*t*
    - binge/intoxication, 441–442
    - diagram of, 442*f*
    - preoccupation/anticipation, 443
    - withdrawal/negative affect, 442–443
  - primal fantasy and, 2133
  - psychological formulation, 2130–2133
  - self-protective processes and, 2131–2132
- Adenohypophysis, 710
- Adenosine
  - chemical structure, 1328*f*
  - schizophrenia pathology, 1506–1507
- Adenylyl cyclase, 3004, 3004*t*
- ADHD. *See* Attention-deficit hyperactivity disorder (ADHD)
- Adhesion molecules, 64
- Adiadochokinesis, 4709
- Adiposity, 226
- ADI-R (Autism Diagnostic Interview-Revised), 1120
- ADIS-C (Anxiety Disorders Interview Schedule for Children), 3540, 3541
- Adjustment
  - defined, 4709
  - in older adults, 4160–4162
- Adjustment disorder(s), **2201–2212**
  - in cancer patients, 2354
  - clinical features, 2208
  - comparative nosology, 2201–2202
  - course and prognosis, 2210
  - cultural perspectives, 2204–2205
  - definition of, 2202, 4709
  - diagnosis, 2208, 4803*t*
  - differential diagnosis, 2209

  - DSM-5 criteria for, 512*t*, 2201–2202
  - epidemiology of, 2205–2206
  - etiology of, 2206–2208
  - history, 2202
  - ICD-11 criteria for, 2203–2204
  - reliability, 2204
  - stressor criteria, 2203
  - subtypes, 2208, 2208*t*
  - symptoms, 2203
  - temporal course of, 2203
  - trazodone, 3199
  - treatment of, 2210–2212
  - validity, 2204
- Adler, Alfred, 947–949, 948*f*, 2943, 2955, 4687, 4709
  - paradoxical communication, 949
  - personality theory, 948
  - psychopathology theory, 948
  - psychotherapy, 948–949
  - reframing, 949
- Adlerian, 4709
- ADLs. *See* Activities of daily living (ADLs)
- Administrative psychiatry, 4709
- Adolescence/adolescents. *See also* Children/childhood
  - antecedents of homosexuality in, 2073
  - assessment of, 3527–3529
    - confidentiality, 3528
    - developmental issues, 3528
    - technical issues, 3528–3529
  - cannabis use and, 1343
  - definition of, 3505–3506, 4709
  - development, **3505–3515**
    - brain, 3512–3513

  - cognitive, 3510
  - moral, 3511
  - physical, 3509–3510
  - racial and ethnic identity, 3506–3507
  - social, 3510–3511
- developmental growth tasks of, 3529*t*
- diagnostic interviews for, 884–885
- disability assessment, 885
- discrimination and, 4104–4113
- disruptive behavior disorders in, **3747–3760**
  - oppositional defiant disorder, **3747–3751**
- early-onset bipolar disorder in, 3828–3836
- end-of-life care, 2406–2407
  - development during illness, 2407
  - interface of hospital and adolescents, 2407
- gender dysphoria in, 2107–2108, 2117
- growth tasks by developmental phase, 3529*t*
- HIV/AIDS in, **3993–3997**
- illness falsification, 2503
- Internet and social media use, 3514–3515
- LGBTQIA youth, 4100–4103
  - intersectional care for, 4102
  - mental health disparities, 4102
  - resilience, 4102–4103
  - sex and gender, 4100–4102
  - sex development, 4101–4102
  - sexual orientation, 4099–4100, 4100*t*
- migrants, 4113–4117
- mood disorders in, **3815–3836**
  - depressive disorders, 3815–3821
  - early-onset bipolar disorder, 3828–3836
  - suicidal behavior, 3821–3827

- prevention of psychiatric disorders in, **4040–4047**
- psychiatric disorders, onset of, 3489
- psychiatric emergencies in, **3571–3585**
- psychiatry, neuroimaging studies in, 3489–3490
- PTSD in, **3845–3850**
- racism and, 4104–4113, 4108–4109
- reproductive development, 2582
- response to hospitalization/.illnesses, 3988
- risk-taking behavior, 3513–3514
- self-esteem in, 3511–3512
- sexual behavior, 2036, 3513
- sleep disorders in, **4083–4089**
  - breathing, 4085
  - circadian rhythm sleep–wake disorders, 4087
  - insomnia, 4084–4085
  - narcolepsy, 4086
  - nocturnal enuresis, 4087
  - parasomnias, 4086–4087
  - somnambulism, 4086
- somatic symptom and related disorders in, 1892
- stages of, 3507–3509
  - early adolescence, 3507
  - late adolescence, 3508–3509
  - middle adolescence, 3508
- substance use disorder in, **3998–4003**
- suicide, 2710–2711
  - poor mental health, 2710
  - risk factors, 2710
  - treatment, 2710–2711
- surveys of mental disorders in, 891–892
- tobacco use disorder in, 1393
- treatment of, **3948–3955**

    - biologic interventions, 3952
    - controversies, 3949–3951
    - environmental and systemic treatments, 3954
    - group therapy, 3952
    - hospital and residential, 3954
    - indications for, 3951–3952
    - modality of psychiatric treatment, 3952–3954
    - psychosocial interventions, 3952–3954
  - typical, 3506
- Adolescent DES (A-DES), 1937*t*. *See also* Dissociative Experiences Scale (DES)
- Adolescent mania, 4709
- Adoption, **3962–3967**
  - closed, 3963
  - developmental issues for, 3964–3965
  - epidemiology of, 3963–3964
  - future directions, 3967
  - history, 3962–3963
  - international, 3965–3966
  - mental health risks and, 3965
  - overview, 3956
  - psychiatrist's role in, 3966–3967
- Adoption and Safe Families Act, 3957, 3960
- Adoption Assistance and Child Welfare Act, 3957
- Adoption studies
  - ADHD, 3729
  - schizophrenia, 1482–1483
- Adrenal disorders, 715–717, 2338–2341
  - adrenocortical insufficiency, 2340–2341
  - hyperaldosteronism, 716, 716*f*
  - hypercortisolism, 2338–2340
  - hyperparathyroidism, 715

  - hypoadrenalism, 715–716
  - pheochromocytoma, 716–717
- Adrenal tumors, 1835
- Adrenergic, 4709
- α-Adrenergic receptor agonists
  - absorption, 3220
  - for ADHD, 3221–3222
  - anxiety disorders, 3222
  - autism spectrum disorder, 3222
  - chemistry, 3220
  - cognition and neuroprotective effects of, 3222–3223
  - distribution, 3220
  - dosage and administration, 3223–3224
  - drug interactions, 3223
  - drug studies, 3220–3221
  - elimination, 3220
  - history, 3220
  - mechanism of action, 3220
  - metabolism, 3220
  - nicotine withdrawal, 3222
  - opiate withdrawal, 3222
  - organ/system effects of, 3221
  - pharmacodynamics, 3220–3221
  - pharmacokinetics, 3220
  - pharmacologic actions, 3220
  - precautions and adverse reactions, 3223
  - PSTD, 3222
  - sexual dysfunction and, 2047
  - therapeutic indications, 3221–3223
  - for tic disorders, 3221, 3784
  - for Tourette syndrome, 3221
- β-Adrenergic receptor antagonists, **3224–3228**

  - absorption, 3225
  - for aggression, 3650
  - for anxiety, 3227
  - for anxiety disorders, 1835
  - chemistry, 3224–3225
  - distribution, 3225
  - dosage and administration, 3228
  - drug interactions, 3228
  - drug studies, 3226
  - elimination, 3225
  - history, 3224
  - laboratory interferences, 3228
  - for lithium tremor, 3227
  - for major depressive disorder, 3227
  - mechanism of action, 3225–3226
  - neuroleptic-induced akathisia, 3227–3228
  - organ/system effects of, 3226
  - for performance anxiety, 3227
  - pharmacodynamics, 3225–3226
  - pharmacokinetics, 3225, 3225*t*
  - pharmacologic actions, 3225
  - precautions and adverse reactions, 3228
  - sexual dysfunction and, 2047
  - side effects and toxicity, 3226*t*
  - for substance use disorder, 3227
  - teratogenic effects, 3226–3227
  - therapeutic indications, 3227–3228
  - for violent behavior, 3227
- Adrenergic receptors, 99–100
- Adrenergic system, 4709
- Adrenocortical insufficiency, 2340–2341
  - course and prognosis, 2341

  - definition, 2340
  - diagnosis and clinical features, 2340
  - differential diagnosis, 2340–2341
  - epidemiology of, 2340
  - etiology of, 2340
  - pathology and laboratory examination, 2340
  - treatment of, 2341
- ADs. *See* Advanced directives (ADs)
- *Adult Attachment Interview* (AAI), 928
- Adult Attachment Inventory, 2927
- Adult Friend Finder, 1127
- Adult neuronal ceroid lipofuscinosis (ANCL), 689
- Adult psychological tests, Millon Clinical Multiaxial Inventory-III, 1042
- Adults/adulthood, physical violence on, **2623–2627**
- Advanced directives (ADs), 2401, 2402*t*, 4421–4422
- Advanced sleep phase circadian rhythm disorder, 4709
- Advanced sleep phase syndrome, 4212–4213
- Adventitious reinforcement, 4709
- Adverse childhood experiences (ACEs), 1948–1949, 1948*t*
- Adverse effects, 4709
- Adverse events
  - coping with, 1156–1157
  - medical documentation requirements for, 1156*t*
- Adverse life situations, 1714
- Adynamia, 4709
- Adynamic ileus, 4709
- Aerophagia, 4709
- Affect, 819, 4709
  - psychiatric examination and, 994–995
- Affect-constancy, 822
- Affect disorders, multiple sclerosis and, 581

- Affected sib pair (ASP) analysis, 255–256
- Affect-fantasy, 4709
- Affective blunting, 1459
- Affective disorders, 1649. *See also* Mood disorder(s)
  - concept of, 1652
  - in primary thyroid disorders, 3399
  - stress and, 2421
  - thyroid function in, 3399
- Affective dysregulation, 1658
- Affective hallucination, 4709
- Affective illness, chronic pain in, 2486
- Affective interaction, 4709
- Affective mentalizing, 2864
- Affective permeability, 2751
- Affective psychoses, signs and symptoms, 1622–1624
- Affective temperaments, 1651, 1670–1671
  - suicidal behavior and, 1713
- Affect-regulating agents, 2139–2140
- Affect(s), 1669. *See also* Mood
- Afferent projections, 194–195
- Affordable Care Act (ACA), 843, 2719–2720, 4561
- Africa
  - biologic therapies in, 3440–3441
  - electroconvulsive therapy in, 3441
  - pharmacoeconomics, 3441
  - psychotropic medications in, 3440–3441
  - transcranial magnetic stimulation in, 3441
- Aftercare, 4709
- Afterhyperpolarization, 70
- Ageism, 4709
- Agency, sense of, 429

- Agency for Healthcare Research and Quality (AHRQ), 1013, 1600, 4500
- Age progression, 2917
- Age regression, 2917
- Age(s), sleep and, 202–203
- Ages of onset
  - anxiety disorders, 1793
  - autism spectrum disorders, 3721, 3721*f*
- Age-structured homosexuality, 2072
- Ageusia, 4709
  - in neuro-Covid, 189
- Aggravated development crisis, 938, 943
- Aggression, 4709. *See also* Assault; Physical violence; Violence
  - animal models of, 302
  - child/adolescent, 3580–3581
    - assessment, 3580
    - management, 3580–3581
    - psychiatric illnesses and, 3580*t*
  - dementia and, 1245
  - environmental influences on, 2519–2520
  - epilepsy and, 527–528, 528*t*
  - group psychotherapy for, 3901
  - intellectual disabilities and, 3647, 3650
  - intrauterine and perinatal factors, 2519–2520
  - lithium and, 3323
  - management of, 2724–2726
  - in pediatric epilepsy, 679–680
  - pharmacotherapy, 3737–3738
    - antipsychotics, 3650
    - beta-adrenergic antagonists, 3650
    - mood stabilizers, 3650–3651
    - selective serotonin reuptake inhibitors, 3650

  - after TBI, 545
- Aggression-turned-inward model, 1653–1654
- Aggressive drive, 4709–4710
- Aging, 4710. *See also* Geriatric psychiatry
  - circadian rhythms and, 198
  - cognitive, 4129, 4163–4165
  - emotional, 4129
  - everyday functioning and, 4191
  - HIV and, **4458–4463**
    - comorbidity, 4461
    - death rates, 4458
    - epidemiology, 4458
    - everyday functioning, 4461–4462
    - mental health outcomes, 4459–4461
      - emotional functioning, 4459–4460
      - severe mental illness, 4460
      - substance use, 4460
    - multimorbidity, 4461
    - neurocognitive outcomes, 4458–4459
    - prevention, 4458
    - successful aging, 4460–4461
    - treatment and management of, 4462–4463
    - veterans, 4462
  - immune-brain interactions and, 190
  - menopause and, 2611
  - neurosteroids and, 139
  - normal, 4120–4121
  - positive psychiatry of, **4471–4476**
  - positive side of, 4129
  - psychological changes with, 4158–4166
    - adjustment and coping, 4160–4162
    - clinical practice implications, 4165–4166

    - design and sampling issues, 4159
    - resilience, 4161–4162
    - social functioning in late life, 4162–4163
    - stage theories of personality development, 4160
- sexuality and, 4438–4457
    - age-related changes in sexual function and behavior, 4440–4443, 4440*t*
    - cognitive capacity and sexual behaviors, 4454
    - frequency of sexual expressions in, 4441–4443
    - LGBTQ+, 4443–4445, 4453–4454
    - recommendations for PRC communities, 4454–4455
    - sexual behavior in long-term care, 4452–4455
    - sexual dysfunctions
        - arousal phase problems, 4448
        - assessment, 4446–4447
        - desire phase problems, 4447–4448
        - illnesses associated with, 4448, 4449–4451, 4449*t*
        - medications associated with, 4451–4452
        - orgasm phase problems, 4448
        - treatment, 4447–4452
    - sexual history, 4445–4446
- sleep and, 4210
- successful, 4165
- successful psychosocial, 4472
- technology and, 4464–4470
    - barriers and limitations, 4476
        - age-related impairments, 4476
        - billing, 4476
        - ethics, 4476
        - information overload, 4476
        - integration with health care system, 4476
        - lack of regulation and evidence, 4476

    - literacy, 4476
    - privacy, 4476
    - usability, 4476
  - future directions, 4469–4470
    - big data, 4469–4470
    - nanotechnology, 4469
    - robotics, 4469
  - specific applications of, 4465–4471
    - caregiver applications, 4469
    - games, 4468–4469
    - mobile health, 4465–4467
    - sensors/self-monitoring devices, 4467, 4467*t*
    - smart homes, 4468, 4468*t*
    - social media, 4469
    - tablets, 4469
    - telehealth, 4465
- Agitated depression, 4710
- Agitation, 4710
  - acute psychiatric illness and, 3577
  - in acute psychosis, 1579–1580
  - autism spectrum disorder and, 3577
  - benzodiazepines, 3247–3248
  - in bipolar disorder, 3057, 3058, 3059, 3101, 3103
  - buspirone for, 3172
  - delirium and, 3576
  - dementia and, 1245
  - developmental delays and, 3577
  - long-term care, 4405*t*
  - management of, 2724–2726
  - medications for, 3576*t*
  - psychomotor, 1672

  - psychotherapeutic and psychopharmacologic management of, 3575–3577, 3576*t*
  - in schizophrenia, 1464–1465, 3057, 3058, 3059, 3061, 3062, 3101, 3103
  - substance intoxication/withdrawal and, 3576–3577
  - after TBI, 545
- Agnosia, 4710
- Agnosias, 1175
- Agomelatine (Valdoxan)
  - for anxiety disorders, 1834
  - chemistry, 3391
  - dosage and administration, 3396
  - drug interactions, 3395–3396, 3396*t*
  - drug studies, 3393–3394
  - molecular structure, 3392*f*
  - organ/system effects of, 3394
  - pharmacodynamics, 3393
  - pharmacokinetics, 3391–3392, 3392*t*
  - precautions and adverse reactions, 3395
  - therapeutic indications, 3394–3395
- Agonist, 4710
- Agoraphobia
  - clinical features, 1784
  - cognitive behavioral therapy for, 1827–1828
  - comparative nosology, 1784
  - course, 1784–1785
  - defined, 4710
  - definition of, 1784
  - diagnosis, 1784, 4797*t*
  - differential diagnosis, 1784–1785
  - epidemiology of, 1784
  - etiology of, 1784–1785

  - history, 1784
  - ICD-11 criteria for, 1223
  - monoamine oxidase inhibitors, 3195–3196
  - in older adults, 4218
  - prevalence of, 1791–1792
  - treatment of, 1784–1785, 1837–1838
- Agranulocytosis, 3069
- Agraphia, 4710
- AHA. *See* American Hospital Association (AHA)
- AHRQ. *See* Agency for Healthcare Research and Quality (AHRQ)
- AI. *See* Artificial intelligence (AI)
- Aichmophobia, 4710
- Aides, 4028
- AIDS. *See* Acquired immunodeficiency syndrome (AIDS)
- AIDS dementia complex, 4710
- AIHQ. *See* Ambiguous Intentions and Hostility Questionnaire (AIHQ)
- Aiken, John, 4684
- Ailurophobia, 4710
- AIMS. *See* Abnormal Involuntary Movement Scale (AIMS)
- Ainsworth, Mary, 731, 3500, 3877
- AIP. *See* Acute intermittent porphyria (AIP)
- Akathisia, 469, 1195, 1580, 3046, 3104, 3780*t*, 4710
  - acute, 3095
  - β-adrenergic receptor antagonists for, 3227–3228
  - anticholinergics and, 3328
  - tardive, 3097
- A kinase anchoring proteins (AKAPs), 3005
- Akinesia, 468, 1459, 4710
- Akinetic mutism, 4710
- Akineton. *See* Biperiden (Akineton)
- Akiskal, Hagop, 1658, 4654

- Al-Anon, 4710
- Alateen, 4710
- Alcohol. *See also* Alcohol abuse; Alcohol dependence; Alcohol use disorders (AUDs); Alcohol withdrawal
  - blackout and, 1298–1299
  - cancer and, 1299
  - cardiovascular problems and, 1299
  - cerebellar degeneration and, 1299
  - CNS effects, 1299
  - conversion chart, 4280*t*
  - as depressant drug, 1298
  - EEG effects, 379–380
  - epidemiology, 1297*t*
  - fetal alcohol syndrome and, 1299–1300
  - gastrointestinal problems and, 1299
  - history, 1297
  - intoxication, 2726–2727
  - moderate use of, 4282
  - neurochemical effects, 1298
  - peripheral neuropathy and, 1299
  - prevalence of drinking, 1297
  - properties and metabolism of, 1298
  - severe use of, 4283
  - sexual dysfunction and, 2047
  - sleep impairment and, 1299
  - spirituality and, 2674
  - tolerance, 1298
- Alcohol abuse. *See also* Alcohol dependence; Alcohol use disorders (AUDs)
  - in adolescents, 3513–3514, 3998
  - bipolar II disorder and, 1693–1694
  - burn injuries and, 2467

  - homosexuality and, 2083–2084
  - insomnia comorbidity, 2169, 2169–2170
  - in older adults, 4133, 4279–4283
    - course and prognosis, 4282
    - differential diagnosis, 4281–4282
    - epidemiology, 4279–4280
    - etiology of, 4280
    - screening, 4280–4281
    - spectrum-of-use categories, 4279
  - recommended limits, 4280
- Alcohol abuse and dependence. *See* Alcohol use disorders (AUDs)
- Alcohol amnestic disorder, 4710
- Alcohol dehydrogenases (ADHs) enzyme, 290
- Alcohol dementia, 4710
- Alcohol dependence. *See also* Alcoholism
  - buspirone for, 3172
  - fMRI of, 345
  - MRI applications to, 341–342
  - MRS in, 349–350
- Alcoholic blackout, 4710
- Alcoholic cardiomyopathy, 1299
- Alcoholic deterioration, 4710
- Alcoholic hallucinosis, 4710
- Alcoholic hepatitis, 1299
- Alcoholic Korsakoff syndrome, 768
- Alcoholic neuropathies, 657
- Alcoholic psychosis, 4710
- Alcoholics Anonymous (AA), 2685, 3900, 4710
- Alcohol idiosyncratic intoxication, 4710
- Alcohol-induced anxiety disorders, 1304
- Alcohol-induced depressive disorder, 1303–1304

- Alcohol-induced major neurocognitive disorders, 1305
- Alcohol-induced psychiatric disorders, 1303
- Alcohol-induced psychotic disorders, 1304
- Alcohol intoxication, 1302, 4710
- Alcoholism, 4710. *See also* Alcohol dependence
  - amino acid neurotransmitters and, 92
  - pharmacokinetics and, 290
- Alcohol-related disorders, **1296–1305**
  - clinical features, 1301–1305
  - diagnosis, 1301–1305
  - ICD-11 criteria for, 1224
- *Alcohol Use Disorder and Associated Disabilities Interview Schedule* (AUDADIS), 884
- Alcohol use disorders (AUDs), **1296–1305**
  - biologic mediators, 1301
  - bipolar disorder and, 1743
  - clinical features, 1301–1305
  - comorbid conditions, 1297
  - course and prognosis, 1307, 1307*t*
  - definition of, 1296
  - dementia and, 1247
  - diagnosis, 1301–1305
  - differential diagnosis, 1306–1307
  - DSM-5 criteria for, 1296, 1302*t*
  - epidemiology of, 1297
  - etiology of, 1300–1301
  - gabapentin, 3267
  - genetic factors, 1300–1301, 1301*t*
  - history, 1297
  - ICD-11 criteria for, 1296, 1302*t*
  - identification of, 1305–1306, 1305*t*
  - medications for, 1310

  - mood disorders and, 1697–1698
  - neuroimaging, 3615
  - neuropeptide systems and, 116
  - in older adults, 4283
  - organ transplantation and, 2459
  - psychodynamic theories, 1300
  - psychological and sociocultural theories, 1300–1301
  - remission, 1296
  - screening for, 1305*t*
  - severity of, 1296
  - suicide and, 2709
  - treatment of, 1308–1311, 4283
    - acamprosate, 3376–3377
    - detoxification, 1308–1309
    - disulfiram, 3370–3371
    - ibogaine, 2907
    - intervention, 1308
    - LSD, 2907
    - nalmefene, 3359–3360
    - naltrexone, 3359–3360
    - psilocybin, 2907
    - rehabilitation, 1309–1311
    - topiramate, 3269
- Alcohol Use Disorders Identification Test (AUDIT), 1111, 4281, 4281*t*
- Alcohol withdrawal, 1303
  - benzodiazepines for, 3247
  - delirium, 1303
  - detoxification and, 1308–1309
  - mild or moderate, 1308
  - protracted, 1309
  - psychiatric emergencies, 2730

  - severe, 1308–1309
  - valproate for, 3307–3308
- Alcohol withdrawal syndrome, 4710
- Aldehyde dehydrogenase (ALDH) enzymes, 290, 1301
- Alemtuzumab, neurotoxicity of, 2465*t*
- Alertness, nonconscious *versus* unconscious, 807–808
- Alexander, Franz, 952–954, 952*f*, 2273–2274, 2741, 2742, 2752, 2765, 4710
  - personality theory, 953
  - psychopathology theory, 953–954
  - specificity hypothesis, 953–954
  - treatment by, 954
- Alexia, 4710
- Alexithymia, 4710
- Algophobia, 4710
- Algorithm, 4710
- Alien abductions, 2682
- Alienation, 957, 4710
- Alien hand, 476
- Alienist, 4710
- Alkyl nitrites, 1370*t*
- Allan Memorial Institute, 4650
- Allegiance effect, 3011
- Allen, Frederick, 3877
- Allen Brain Atlas, 250*t*
- Allergic jaundice, 4710
- Allocortex, 17
- Allodynia, 2480*t*
- Alloplasty, 4710
- Allopregnanolone, 163, 164–165, 1879
- Allostatic load, 4474
- Allostery, 3001

- Allport, Gordon, 972, 975, 2886
- Allport's group relations theory, 4710–4711
- Alogia, 1459, 4711
- Alopecia, 4711. *See also* Trichotillomania (hair-pulling disorder)
- Alopecia areata, 2440–2441
  - clinical features, 2441
  - pathology of, 2441
  - psychiatric comorbidity, 2441
  - treatment of, 2441
- Alpha (α) error, 4711
- α_2-Agonists, in pediatric anxiety disorders, 3927
- α-Adrenergic agonists, for ADHD, 3737
- α-Mannosidosis type II (AM), 689, 690*f*
- Alphabetic principle, 3656*t*
- Alpha-methylacyl-CoA racemase (AMACR) deficiency, 695
- Alpha-thalassemia/mental retardation syndrome, X-linked (ATR-X), 271*t*
- Alphaviridae, 616
- Alprazolam (Xanax)
  - for anxiety disorders in later life, 4224–4225
  - history, 3007, 4663
  - pharmacokinetics of, sex differences in, 3417
  - for premenstrual dysphoric disorder, 8
- Alprostadil, 2053–2054, 3434, 3434–3435
- ALS. *See* Amyotrophic lateral sclerosis (ALS)
- Altenberger, Helmut, 369
- Altered states of consciousness (ASC), 1178
- Alternating leg muscle activation, 2188
- Alternating psychosis, 525*f*
- Alternating role, 4711
- Alternative RNA splicing, 245
- Altman Self-Rating Mania Scale (ASRM), 1105–1106, 1107*t*

- Altruism, 793, 921*t*, 2674–2675, 2959*t*, 4711
- Alveus, 35
- Alzheimer, Alois, 1239, 3335, 4245, 4649, 4657
- Alzheimer disease (AD), **1252–1259**, **4245–4254**
  - biomarkers, 1255, 4248
  - clinical evaluation, 4247–4248
  - clinical features, 1255
  - comparative nosology, 1252–1253, 4245
  - course and prognosis, 1255–1256, 4250
  - definition of, 1252
  - dementia care management, 4252–4253
  - diagnosis, 1255
  - diagnostic criteria, 4246*t*
  - differential diagnosis, 4248–4250
  - early-onset, 1253
  - epidemiology, 1253, 4246–4247
  - etiology, 1253, 4246–4247
  - gene mapping, 259–260
  - genetics of, 1253–1254
  - history, 1252, 3335, 4245
  - imaging, 4248
  - laboratory examination, 4248
  - late-onset, 260–261, 1253–1254
  - lithium, 3320
  - in medically ill geriatric patients, 4202–4203
  - medications for, 4315*t*
  - neurodegenerative disorders due to, 4174–4175
  - neuropathology, 1256–1258
  - in older people, 4127
  - omega-3 fatty acids and, 2654
  - pharmacologic treatment for, 1258–1259
    - cholinesterase inhibitors, 1258

    - memantine, 1258
  - protective factors, 1255
  - psychopharmacotherapy, 4250–4252, 4251*t*
  - quantitative electroencephalography of, 1080–1081
  - risk factors, 1254–1255
  - sexual dysfunction and, 4451
  - treatment of, 4250–4253
    - cholinesterase inhibitors, **3335–3344**
      - donepezil, 3336–3338
      - galantamine, 3341–3342
      - rivastigmine, 3338–3341
      - tacrine, 3336
    - memantine, 3342–3344
- Alzheimer’s disease, sporadic, 260–261
- Alzheimer’s Disease Assessment Scale (ADAS-Cog), 4168
- AMA. *See* American Medical Association (AMA)
- *Amanita muscaria*, 89
- Amantadine (Symmetrel), 3107, **3331–3334**
  - chemistry, 3331
  - in children, 3333
  - clinical trials, 3331–3332
  - for concomitant therapy, 3334
  - deaths, 333
  - dosage and administration, 3114, 3114*t*, 3334
  - drug addiction, 3333
  - drug interactions, 3114, 3334
  - EPS, 3334
  - for extrapyramidal symptoms, 3111
  - extrapyramidal symptoms, 3111, 3332
  - molecular structure of, 3108*f*, 3331*f*
  - organ/system effects of, 3332
  - Parkinson disease and Parkinsonism, 3332

  - pathologic gambling, 3333
  - pharmacodynamics, 3109, 3331
  - pharmacokinetics, 3108
  - pharmacologic actions, 3331
  - precautions and adverse reactions, 3333, 3334*t*
  - side effects, 3113
  - therapeutic indications, 3332–3333
  - traumatic brain injury, 3332–3333
- Ambien. *See* Zolpidem (Ambien)
- Ambiguous Intentions and Hostility Questionnaire (AIHQ), 4551
- Ambivalence, 4711
- Ambivalent attachment, 3500
- Ambulatory schizophrenia, 4711
- Amentia, 1628*t*, 4711
- American Academy of Child and Adolescent Psychiatry (AACAP), 3942, 4711
- American Academy of Pediatrics (AAP), 1907
- American Association for Emergency Psychiatry, 2719
- American Association of Chairmen of Departments of Psychiatry (AACDP), 4711
- American Association of Directors of Psychiatric Residency Training, Inc. (AADPRT), 4711
- American Association of Retired People (AARP), 4138
- American Association on Intellectual and Developmental Disabilities (AAIDD), 3632
- American Board of Medical Specialties (ABMS), 4586, 4588*t*, 4590
- American Board of Psychiatry and Neurology (ABPN), 4586, 4588–4589, 4588*t*, 4687, 4711
- American College of Physicians (ACP), 1014, 2363, 4539, 4711
- American College of Psychiatry (ACP), 4588
- American Hospital Association (AHA), 4590
- American Indians/American Natives (AI/AN), mental health of, 844–845

- American Law Institute formulation of insanity, 4711
- American Medical Association (AMA), 4587, 4711
  - practice guidelines in psychiatry, 1013
- American Medical Colleges (AAMC), 4590
- American Medico-Psychological Association, 1199
- American Neurological Association (ANA), 4711
- American Osteopathic Association (AOA), 4596
- American Psychiatric Association (APA), 1199–1200, 4711
- American Psychoanalytic Association (APA), 4711
- American Psychological Association (APA), 4711
  - practice guidelines, 1014–1016
    - advantages, 1016–1017
    - dissemination and implementation strategies, 1016
    - format, 1015
    - future of, 1018
    - limitations, 1017–1018
    - recommendations, 1015
    - review and approval process, 1015–1016
    - updating, 1016
- American Recovery and Reinvestment Act (ARRA), 4501
- Americans with Disabilities Act (ADA), 3632, 3655, 4513–4516
- Amimia, 4711
- Amines, 4711
- Amino acid neurotransmitters, **83–92**
  - alcoholism and, 92
  - anxiety and, 92
  - depression and, 92
  - γ-aminobutyric acid (GABA), 87–90
  - glutamic acid, 83–87
  - glycine as neurotransmitter, 90
  - neuropsychiatric implications, 90–92
  - schizophrenia and, 90–92

- Amino acids, 2633–2634, 4711
  - inhibitory, 87–90
- Aminophenylpyridone, 4663
- Amisulpride, for tic disorders, 3785
- Amitriptyline (Elavil), 3152
  - for major depressive disorder, 1724
  - pharmacogenetics of, 294*t*
  - pharmacokinetics, sex differences in, 3416
  - for tension-type headaches, 651
- Amlodipine (Norvasc), 3260
- Amnesia, 472–474, 4711. *See also* Dissociative amnesia; Post-traumatic amnesia; Transient global amnesia
  - acute, evaluation of, 2006*t*
  - anterograde, 768–769, 1181, 1298–1299
  - declarative memory in, 767–772
  - in dissociative identity disorder, 1976–1977, 1976*t*
  - factitious, 2007
  - hypnosis and, 2917
  - infantile, 783
  - inter-identity, 1978–1979
  - malingering, 2007
  - malingering and, 2501
  - nonpathologic, 2005–2006
  - post-traumatic, 541
  - posttraumatic, 2007
  - psychogenic, 782
  - retrograde, 769–772, 1181
  - substance-related, 2007
  - transient global, 2007
  - traumatized populations, 1949–1950
- Amnesic aphasia, 4711
- Amnestic aphasia, 4711

- Amnestic disorder, 4711
- Amnestic disorders, **1282–1283**
  - assessment, 1283
  - causes of, 1282*t*
  - course and prognosis, 1283
  - definition of, 1281
  - epidemiology of, 1281–1282
  - etiology of, 1282
  - Korsakoff syndrome, 1282–1283
  - symptoms of, 1282–1283
  - treatment of, 1283
- Amnestic neuropeptide, **168**
- Amobarbital (Amytal), 1426*t*
- Amobarbital narcosynthesis, 2012
- Amoeba, 235
- Amoebiasis, 627–628
- Amok, 4711
- Amotivational syndrome, 1342
- Amoxapine (Asendin), 3152
  - pharmacogenetics of, 294*t*
- AMPA receptors (AMPARs), 84
- Amphetamine, 3114. *See also* Stimulant-related disorders
  - for ADHD, 3733–3736
  - adverse effects of, 3735–3736
  - defined, 4711
  - efficacy, 3735
  - history, 4704
  - lithium and, 3320
  - neural pathway, 446
  - pharmacogenetics of, 294*t*
  - pharmacology of, 3734
  - withdrawal, 2730

- Amphetamine-type stimulants, 1313
- Amplification, 242
- Amygdala, 10, 36, 2781
  - biologically relevant learning and, 1812
  - centromedial, 36
  - defined, 4712
  - depressive disorders/bipolar and related disorders, 3607–3608
  - emotional memory and, 1812
  - extended, 36
  - obsessive-compulsive disorders, 3611
  - output pathways, 38*f*
- Amyl, butyl, and isobutyl nitrite–related disorders, 1381
- Amyloid-beta peptides, 4324
- Amyloid-β protein, 1254
- Amyotrophic lateral sclerosis (ALS), 654–655, 4712
- Amytal interview, 4712
- ANA. *See* American Neurological Association (ANA)
- Anabolic-androgenic steroid-related disorders, 1432–1446
  - adverse medical effects, 1437–1438
  - anxiety disorders, 1444–1445
  - associated substance abuse or dependence, 1444–1445
  - chemistry of, 1436
  - commonly used anabolic steroids, 1433*t*
  - comparative nosology, 1435
  - dependence, 1439–1442, 1441*f*
  - endocrine tests, 1439
  - epidemiology of, 1435–1436
  - etiology of, 1439
  - history, 1433–1435, 1434*f*
  - laboratory findings, 1438–1439
  - medical uses, 1437
  - mood disorders and, 1442–1444

    - hypomanic or manic episodes, 1442–1443
    - major depressive disorder, 1443
    - mixed episodes, 1443–1444
    - violence, 1444
  - organ/system effects of, 1437
  - overview, 1432–1433
  - pharmacodynamics, 1436–1437
  - pharmacology of, 1436–1438
  - physical effects of anabolic steroid use, 1440*f*
  - psychotic disorders, 1444
- Anabolic steroid use, 1055
- Anaclitic, 4712
- Anaclitic depression, 3500, 3802
- Anaclitic therapy, 4712
- Anafranil. *See* Clomipramine (Anafranil)
- Anal character, 4712
- Anal erotism, 4712
- Analgesia, 2917, 4712
- Analgesics
  - nonopioid, 2389
  - opioid, 2389
- Analophilic orientation, 2058
- Anal phase, 4712
- Anal stage, 916
- Analysand, 4712
- Analysis in depth. *See* Psychoanalysis
- Analysis of variance (ANOVA), 4712
- Analytic psychology, 4712
- Analytic studies, 878
- Anamnesis, 4712
- Anancasm, 4712
- Anankastic personality, 4712

- Anankastic personality disorder, 4837*t*
- *Anatomy of Melancholy* (Burton), 1651–1652
- Anchin, Jack, 1707
- Ancillary care, 4712
- ANCL. *See* Adult neuronal ceroid lipofuscinosis (ANCL)
- Anderson, Carol, 2948–2949
- Anderson, Michael C., 1956
- Andreasen, Nancy, 1457
- Androgens, 4712. *See also* Antiandrogens; Complete androgen insensitivity syndrome (CAIS)
  - atypical exposure, 2070
  - for midlife depression, 3428–3429
- Androgyny, 4712. *See also* Bisexuality
- Androphilic orientation, 2058
- Androphobia, 4712
- Androstenedione, for midlife depression, 3428–3429
- Anergia, 4712
- Anergic schizophrenic, 4712
- Anesthesia, 2917, 4712
- Anesthetic agents, 3236
  - for chronic pain, 2488
  - pain management in burn patients, 2470
- Angel dust, 4712
- Angelman syndrome (AS), 271*t*, 3641*t*
- *Angels: God’s Secret Agents* (Graham), 2682
- Anger
  - in burn injuries, 2474
  - chronic pain in, 2486
  - response to, 1703
- Angioedema, 2445–2446
  - clinical features, 2445–2446
  - pathophysiology of, 2446

  - psychiatric comorbidity, 2446
  - treatment of, 2446
- Angst, Jules, 1652
- Anhedonia, 299, 1458–1459, 1672, 4712
  - striatal connectivity and, 415
- Anima, 4712
- Animal-facilitated therapy, 2642
- Animal models, **296–302**
  - of anxiety-like behavior, 298
  - chronic stress models, 299
  - of cognition, 301–302
  - of fear-like behavior, 297–298
  - importance and utility of, 304–305
  - of negative valence, 297–299
  - of panic-like behavior, 297
  - of positive valence, 299–301
  - of sleep and arousal, 302
  - of social anxiety-like behavior, 298–299
  - of social functioning, 302
- Anna O, 4712
- Anniversary reaction, 4712
- Anniversary reactions, 1670, 2656
  - bereavement depression and, 1695*t*
- Anoesis, 805
- Anomalous experiences, 2681
- Anomia, 4712
- Anomic aphasia, 4712
- Anomie, 4712
- Anorectic, 4712
- Anorexia nervosa, 1188, 1674, 2147, 4712
  - animal models of, 233
  - causes and treatment of, 2392*t*

  - course and prognosis, 2158–2159
  - diagnosis and clinical features, 2153
  - diagnostic criteria, 2148–2149, 4808–4809*t*
  - differential diagnosis, 2157
  - epidemiology of, 2150, 2150*t*
  - etiology of
    - developmental factors, 2151
    - environmental and social factors, 2152
    - genetic and biologic factors, 2151
    - psychological and behavioral factors, 2151–2152
  - family therapy for, 3920–3921
  - history, 2148–2149
  - pathology and laboratory examination, 2155*t*, 2156
  - subjective appetite in, 231, 232*f*
  - terminal, 2392
  - treatment of
    - nutritional rehabilitation, 2160–2161
    - psychopharmacology, 2161–2162
    - psychotherapy, 2161–2162
    - refeeding syndrome and, 2161, 2161*t*
    - weight restoration, 2160–2161
- Anorgasmia, 2042–2043, 4712
- Anosmia, in neuro-Covid, 189
- Anosognosia, 486, 1175, 4575, 4712
- Anosomnia, 4712
- ANOVA. *See* Analysis of variance (ANOVA)
- ANP. *See* Apparent normal personality (ANP)
- ANS. *See* Autonomic nervous system (ANS)
- Antabuse. *See* Disulfiram
- Antagonist, 4712
- Anterior capsulotomy, 3482–3483
- Anterior cingulate cortex (ACC), 459, 1756, 2917–2919

- Anterior cingulate gyrus, 798
- Anterior cingulotomy, 3482
- Anterior limb, 11
- Anterograde amnesia, 1181, 1298–1299, 4712
  - anatomy of, 768–769
- Anthropology, 4712
- Antiadrenergic agents, for post-traumatic stress disorder, 3931
- Antialiasing filter, 361
- Antiandrogens, 2055, 2098
- Antianxiety drugs
  - defined, 4712
  - for epilepsy, 531
  - sexual dysfunction and, 2047
- Anticholinergic deliriants, 1355
- Anticholinergic effect, 4712
- Anticholinergic(s), **3324–3329**
  - absorption, 3326
  - amantadine and, 3331–3334
  - atropine, 3325
  - chemistry, 3325–3326, 3325*f*
  - clinical studies, 3326–3327
  - defined, 4712
  - distribution, 3326
  - dosage and administration, 3329
  - drug interactions, 3329
  - extrapyramidal side effects, 3327
  - history, 3324
  - intoxication, 1237
  - long-term use of, 3328
  - organ/system effects of, 3327–3328
  - pharmacodynamics, 3326
  - pharmacologic actions, 3326

  - precautions and adverse reactions, 3328–3329, 3328*t*
  - scopolamine, 3329–3331
  - sexual dysfunction and, 2047
  - for stereotypic movement disorders, 3813
  - therapeutic indications, 3327
- Anticipation, 793, 921*t*, 2748*t*, 4712
- Anticonvulsants, 73, 4712
  - for acute bipolar depression, 1733–1734
  - for acute mania, 1729
  - anticonvulsant, 4301
  - for anxiety disorders, 1835
  - for bipolar disorder, 1737–1738, 3929–3930
  - for chronic pain, 2489–2490
  - for dementia, 1251
  - for early-onset bipolar disorder, 3835
  - for geriatric bipolar disorder, 4243
  - pharmacokinetics of, sex differences in, 3417
- Antidementia drugs, **4313–4324**
  - adjunctive and combination therapy, 4323
  - amyloid-beta peptides, 4324
  - for behavioral and psychological symptoms of dementia (BPSD), 4314–4315
  - FDA regulations, 4313–4314
  - future directions, 4324
  - mild cognitive impairment, 4323
  - overview, 4313
  - Parkinson disease dementia, 4323
  - treatment using, 4315–4316, 4315*t*, 4322
    - cholinergic hypothesis, 4315
    - cholinesterase inhibitors, 4316
      - adverse effects, 4318
      - clinical applications, 4318

            - contraindications, 4318
            - donepezil, 4316–4317
            - duration of treatment, 4318–4319
            - galantamine, 4317–4318
            - pharmacodynamics, 4316*t*
            - pharmacokinetics, 4316*t*
            - rivastigmine, 4317
            - tacrine, 4316–4317
        - docosahexaenoic acid, 4322
        - ergoloid mesylates, 4320
        - ginkgo biloba, 4321
        - HMG-CoA reductase inhibitors, 4321–4322
        - huperzine A, 4320
        - memantine, 4319
        - monoamine oxidase type B inhibitors, 4320
        - nootropics, 4320–4321
        - precursor loading and cholinergic stimulation, 4315–4316
        - souvenaid, 4322
        - vitamin B and homocysteine reduction, 4321
        - vitamin E, 4321
    - unproven medications
        - anti-inflammatory drugs, 4322
        - estrogens, 4322–4323
    - vascular dementia, 4323
- Antidepressant discontinuation syndrome, 3097–3098
    - clinical presentation, 3097–3098
    - pathophysiology of, 3098
    - treatment options/implications, 3098
- Antidepressant-induced rapid cycling, 1678
- Antidepressant(s), 1067. *See also specific Antidepressants*
    - acute bipolar depression, 1734
    - anxiety, 3650

  - anxiety disorders, 1832–1834
  - for anxiety disorders in later life, 4303
  - bipolar disorder, 1739
  - for bulimia nervosa, 2162
  - for cancer patients, 2348
  - for chronic pain, 2488
  - cutaneous adverse events, 2447
  - defined, 4712
  - for dementia with Lewy bodies, 1268–1269
  - for depression in later life, 4296–4301
  - dual-acting, 4297–4298
  - effect on cytokines, 185
  - epigenetic mechanisms of, 281
  - for epilepsy, 530–531, 532*t*
  - functional gastrointestinal disorders, 2301
  - history, 4691
  - inflammatory biomarkers, 183
  - major depressive disorder, 1720*t*, 1721–1725
  - migraine, 650
  - monoamine oxidase (MOA) inhibitors, 1067
  - for mood dysregulation in children and adolescents, 3819
  - novel, 1834, 2488
  - pharmacokinetics of, sex differences in, 3414–3416
  - for repolarization abnormalities, 2288–2289
  - sexual dysfunctions, 2047, 2054
  - for suicide, 2712
  - tetracyclic, 1067
  - tricyclic, 1067
  - weight gain and, 2308–2309, 2308*t*
- Antidiuretic hormone, 1055
- Antiepileptics, 530
  - behavioral effects of, 530*t*

  - for chronic pain, 2488
  - defined, 4712
  - pharmacokinetics of, sex differences in, 3417
  - pregnancy and, 2607–2608
- Antiestrogens, 2055
- Antigay/antilesbian violence, 2082
- Antihistamines, **3345–3349**. *See also* Histamine
  - absorption, 3345
  - for anxiety disorders in later life, 4305–4306
  - attenuation of iatrogenic weight gain and, 3348
  - clinical drug studies, 3347
  - cyproheptadine, 3348
  - diphenylhydramine, 3347
  - distribution, 3345–3346
  - dosage and administration, 3349, 3349*t*
  - drug interactions, 3349
  - elimination, 3346
  - history, 3345
  - hydroxyzine, 3347
  - laboratory interferences, 3349
  - metabolism, 3346
  - for neurocognitive disorders, 3348
  - pharmacodynamics, 3346–3347
  - pharmacokinetics, 3345, 3345*t*
  - pharmacologic actions, 3345
  - precautions and adverse reactions, 3348–3349
  - promethazine, 3347
  - sexual dysfunction and, 2047
  - therapeutic indications, 3348
- Anti-inflammatory drugs, 4322
  - for depressive disorders, 184–185
- Antimanic drug, 4713

- Antineoplastic hormones, 2367–2368
- Anti-NMDA receptor encephalitis (NMDARE), 81
- Antiparkinsonism drugs, 4713
  - for dementia with Lewy bodies, 1268–1269
- Antipsychotics, **4306–4312**
  - for acute mania, 1729–1730
  - adherence, 4309–4310
  - for aggression, 3650
  - for agitation, 2724
  - atypical, 99, 1729–1730, 1734, 1738–1739
  - for autism spectrum disorders (ASDs), 3727
  - for bipolar disorder, 1738–1739, 4308–4309
  - chemical structures of, 1547*f*
  - defined, 4713
  - for delirium, 1236–1237, 4308
    - in cancer patients, 2352, 2353*t*
  - for delusional disorder, 4307
  - for dementia, 1250–1251
  - for dementia with Lewy bodies, 1268–1269
  - for dementia with psychosis, 4307–4308
  - for depression, 4308
  - dosage forms, 4309
  - dosage formulations, 4310*t*
  - dosing recommendations, 4309, 4309*t*
  - drug interactions, 4310, 4310*t*
  - for early-onset bipolar disorder, 3834
  - for early-onset schizophrenia, 4307
  - for epilepsy, 530
  - for functional gastrointestinal disorders, 2302
  - history, 4663
  - for late-onset schizophrenia, 4307
  - long-acting injectable, 3053–3054

  - mortality rates, 4310
  - for neurocognitive disorders, 4411*t*
  - for Parkinson disease with psychosis, 4308
  - pharmacodynamics, 4306
  - pharmacokinetics of, 3038, 4306
    - sex differences in, 3417
  - precautions and adverse reactions, 4310–4312
    - anticholinergic side effects, 4311
    - cardiovascular effects, 4312
    - cerebrovascular events, 4310–4311
    - hypotension, 4312
    - metabolic effects, 4310–4311
    - sedation, 4311–4312
    - tardive dyskinesia, 4311
    - weight gain, 4311
  - pregnancy and, 2606–2607
  - receptor activity, 4306
  - for repolarization abnormalities, 2288
  - for schizophrenia, 1545–1549, 1576–1577*t*
  - selection of, 4309
  - for self-injurious behaviors, 3651
  - sexual dysfunction and, 2047
  - side effects, 99
  - substance/medication-induced, 2046–2047
  - for suicide, 2713
  - therapeutic indications, 4307–4309
  - weight gain and, 2308–2309, 2308*t*, 2332*t*
- Antipyretic analgesics, 2487
- Antiracism, 4112–4113
- Antisaccade tasks, 1570
- Antiseizure drugs, for migraine, 650
- Antisocial, narcissistic, and histrionic spectrum, 2248

- Antisocial behavior, 4713
- Antisocial personality disorder, 2246–2247
  - alcohol use disorder and, 1306
  - defined, 4713
  - diagnosis, 4836*t*
  - dissociative identity disorder and, 1975
  - malingering and, 2503–2504
  - mentalization-based treatment for, 2873
- Antisocial Process Screening Device (APSD), 3546
- Anton syndrome, 4713
- Anxiety, 1780–1781, 4713. *See also* Anxiety disorders; Generalized anxiety disorder (GAD); Hamilton Anxiety Rating Scale (HARS); Social anxiety disorder (SAD)
  - β-adrenergic receptor antagonists for, 3227
  - amino acid neurotransmitters and, 92
  - anatomical correlates of fear and, 1808–1813, 1809*f*
  - assessment, 1781
  - brain tumor patients and, 512–513
    - clinical features, 512
    - course and prognosis, 513
    - diagnosis of, 512
    - differential diagnosis, 512
    - epidemiology of, 512
    - etiology of, 512
    - laboratory investigations, 513
    - treatment of, 513
  - cannabis, 1342
  - cardiac disease and, 2292
  - cognition in, 738
  - cognitive behavioral therapy for, 4371
  - control, 2712
  - dementia and, 1245

- differential diagnosis, 2721*t*
- endocannabinoids in, 134–135
- epilepsy and, 527
- gene mapping, 266–267
- group psychotherapy for, 3902–3903
- history, 4656
- intellectual disabilities and, 3650–3651
- learning theory and, 439–440
- naltrexone for, 3364
- neurochemical basis of, 1796
- neuropsychological evaluation, 4171
- obesity and, 2308
- in older adult cancer patients, 2379
- omega-3 fatty acids and, 2653–2654
- Parkinson disease and, 559, 562
- in pediatric epilepsy, 679
- perception in, 738
- pharmacotherapy
  - antidepressants, 3650
  - anxiolytics, 3651
  - buspirone, 3651
  - mood stabilizers, 3650
- poststroke
  - assessment, 493
  - course and prognosis, 497
  - treatment of, 499–500
- in preterm adolescents, 673
- primary, 2757
- rating scales, 3544–3545
- in schizophrenia, 1463
- sensation in, 738
- in sexually abused children, 3971

  - signal, 2757
  - spirituality and, 2679–2681
  - after TBI, 551
  - in terminally ill children, 2406
  - in terminally ill patients, 2394–2395, 2394*t*
  - types of, 2757
- Anxiety disorders, 1185–1187, 1780–1839. *See also* Agoraphobia; Generalized anxiety disorder (GAD); Panic disorder; Separation anxiety; Social anxiety disorder (SAD)
  - age of onset, 1793
  - alcohol-induced, 1304
  - alcohol use disorder and, 1306
  - anabolic steroid-induced, 1444
  - arginine vasopressin, 1803
  - benzodiazepines and, 3244
  - bipolar disorder and, 1742
  - caffeine-induced, 1330–1332
  - in cancer patients, 2353–2354
    - symptoms, 2353
    - treatment of, 2354
  - categories, 1207*t*
  - changes in DSM, 1205–1206
  - child/adolescent, 3437, **3837–3866**, 4713
    - epidemiology of, 1792–1793
    - generalized anxiety disorder, **3855–3857**
    - obsessive-compulsive disorders, **3837–3844**
    - pharmacotherapy for, 3927–3928
    - PTSD, **3845–3850**
    - selective mutism, **3861–3866**
    - separation anxiety, **3851–3853**
    - sleep disorders and, 4089
    - social anxiety disorder, **3853–3855**

    - treatment of, 1839, 3857–3860
        - pharmacology, 3858–3860
        - psychotherapy, 3857–3858
- cholecystokinin and, 1804–1805
- chronic pain in, 2486
- classification, 2353–2354
- clinical features, **1782–1790**
- clonidine for, 3223
- cognitive behavioral therapy for, **1822–1831**, 2824–2825
    - acceptance-based techniques, 1827
    - anxiety management, 1826–1827
    - cognitive therapy, 1826–1827
    - efficacy, 1827–1830
    - emotional processing and, 1822–1824
    - exposure therapy, 1825–1826
    - interpersonal skills training, 1827
    - mindfulness-based training, 1827
    - stress inoculation training, 1826–1827
- comorbidities, 1794
- coronary artery disease and, 2281
- corticotropin-releasing hormone and, 1799
- course and persistence, 1795
- defined, 4713
- differential diagnosis, 1696, 3872–3873, 4794–4797*t*
- DSM-5 modifications, 1781
- due to another medical condition, 1789
- endocannabinoid system and, 1805–1806
- epidemiology of, **1790–1795**
- ethnicity and, 1793–1794
- exposure, 1794–1795
- familial factors, 1794
- family therapy for, 3911*t*

- future directions, 1795
- GABA and, 1801
- gabapentin, 3267
- galanin and, 1804–1805
- genetics of, 1794
- glutamate and, 1801–1802
- group psychotherapy for, 2965–2966
- guanfacine for, 3222
- HIV in children/adolescents and, 676
- HPA axis and, 1798–1799
- hypnosis and, 2928–2930
- ICD-11 criteria for, 1222–1223
- impact of, 1795
- inflammatory and immunologic factors, 1806
- inhalant-induced, 1379
- intellectual disabilities and, 3647
- Internet gaming disorder and, 4092
- in later life, 4302–4306
- in long-term care, 4414
- magnitude of, 1791–1793
- medical conditions associated with, 2353*t*
- medical symptoms/disorders and, 1794
- nefazodone, 3205–3206
- neuroanatomical circuits in, 1807–1813
- neurobiology and neuroscience, **1796–1806**
- neuroimaging, 1814–1821, 3612
- neuropeptide Y, 1803–1804
- neurosteroids and, 138
- noradrenergic system and, 1796–1797
- nosologic issues, 2757–2758
- in older adults, **4216–4227**
- in older people, 4127

- organ transplantation and, 2458
- oxytocin and, 1803–1804
- pediatric OCDs and, 3841
- perinatal, 2600
- phencyclidine- and other hallucinogen-induced, 1364
- pregabalin, 3275
- prevalence rates, 1792*t*
- psychodynamic therapy for, 2763
- rating scales for, 1106–1108
- serotonergic system and, 1800–1801
- short-term dynamic psychotherapy for, 2778
- social class and, 1793–1794
- somatic therapies for, 1832–1839
- specific learning disorder and, 3665
- stress and, 2421
- substance/medication-induced, 1789
- suicide and, 2709
- symptoms, 2353–2354
- temperament and, 3626
- temperament/personality and, 1794
- traumatic brain injury and, 683
- treatment of, 3738
  - anticonvulsants, 1835
  - antidepressants, 1832–1834
  - benzodiazepines, 1834–1835
  - beta-adrenergic antagonists, 1835
  - buspirone, 1835
  - pharmacologic options, 1832–1836
  - response and illness course, 1836–1837
  - trazodone, 3199
- treatment-resistant, 1835–1836

- Anxiety Disorders Interview Schedule for Children (ADIS-C), 3540, 3541
- Anxiety hysteria, 4713. *See also* Phobias
- Anxiety-like behavior, 298
- Anxiety management training, 1827
- Anxiety neurosis, 4713
- Anxiety-related hallucinations, symptoms in children and adolescents, 3868*t*
- Anxiety state, 4713
- Anxiolytic-related disorders. *See* Sedative-, hypnotic-, anxiolytic-related disorders
- Anxiolytics, 1420, 3651, 4713
- Anxious (avoidant) personality disorder, 4837*t*
- Anxious dysthymia, 1688. *See also* Dysthymia
- AOA. *See* American Osteopathic Association (AOA)
- Aortic stenosis, 2285
- Aortic valve disease, 2285
- APA. *See* American Psychiatric Association (APA); American Psychoanalytic Association (APA); American Psychological Association (APA)
- Apathetic social withdrawal, 1459
- Apathetic withdrawal, 4713
- Apathy, 475
  - defined, 4713
  - Parkinson disease and, 559, 562–563
  - poststroke, 488–489, 489t
    - assessment, 492–493
    - course and prognosis, 497
    - etiology, 491–492, 493t
    - treatment of, 500
  - after TBI, 548–549
- Apgar scores, 4713

- Aphasia, 471–472, 473*t*, 1020–1021, 1020*t*, 1182–1183. *See also* Language disorder
  - defined, 4713
  - fluent, 1182
  - nonfluent, 1182–1183
- Aphemia, 467*t*
- Aphonia, 4713
- Aphrodisiac, 4713
- Apnea, 4713
- Apokyn. *See* Apomorphine (Apokyn)
- Apollonian–Odyssean–Dionysian cluster, 2923
- Apology, 1157
- Apomorphine (Apokyn), 99
- Apoplexy, 4713. *See also* Stroke
- Apoprotein E receptor 2 (ApoER2), 63
- Apoptosis, 87
- Apoptotic cell death, 60
- APP. *See* Assimilative psychodynamic psychotherapy (APP)
- Apparent normal personality (ANP), 1967
- Appel, Kenneth, 4511
- Apperception, 4713
- Appetite. *See also* Feeding and eating disorders
  - basic science of, 219–234
  - brain mechanism of, 226–229
    - brainstem, 226–228
    - hypothalamus, 228, 228*f*
    - telencephalon, 228–229
  - disturbances of, 1188
  - as network function, 219–220
  - physiologic modulators of, 229–231
    - energy expenditures, 230
    - illness anorexia, 231

    - interactions, 230
    - learning, 229–230
    - reproductive axis, 230–231
  - postprandial satiety and, 224–226
  - psychiatric eating disorders and, 231–233
  - satiation and, 224–226
- Appetitive phase, 4713
- Appetitive symptoms, 465
- Apple ResearchKit, 3020
- Applied behavioral analysis, 756
- Appropriate affect, 4713
- Apps, 1121, 2703
- Apraxia, 4713
- Apraxia of speech, 467*t*
- Aprobarbital (Alurate), 1426*t*
- Aprosodias, 1182–1183
- APSD. *See* Antisocial Process Screening Device (APSD)
- Aptitude, 4713
- Aquachloral. *See* Chloral hydrate
- Archetype, 4713
- Archetypes, 950–951
- Archicortex, 17
- Arcuate fasciculus, 13, 461
- Arcuate nuclei, 37
- Arecoline, 105
- Aretaeus of Cappadocia, 1230
- ARFID. *See* Avoidant/restrictive food intake disorder
- Arginine, 2635
- Arginine vasopressin (AVP), **167–168**, 3792
  - amino acid sequence, 109*t*
  - anxiety disorders and, 1803
- Aricept. *See* Donepezil (Aricept)

- Arieti, Silvano, 1707
- Aripiprazole (Abilify), 99, 3100–3103
  - for acute mania, 3101
  - for bipolar disorder, 1730*t*, 1738, 3101–3103
  - chemistry, 3100
  - for delirium, in cancer patients, 2352
  - for dementia, 1250
  - dosage and administration, 3102–3103
  - drug-device combination, 3103
  - drug interactions, 3102
  - for early-onset bipolar disorder, 3834
  - effects on organs and systems, 3101
  - for geriatric bipolar disorder, 4243
  - for major depressive disorder, 1724
  - mechanism of action, 3100–3101
  - molecular structure, 3056*f*
  - pharmacodynamics, 3100–3101
  - pharmacogenetics of, 294*t*
  - pharmacokinetics, 3100
  - precautions and adverse reactions, 3102
  - for schizophrenia, 3101–3103
  - for stereotypic movement disorders, 3813
  - therapeutic indications, 3102
  - for tic disorders, 3785
- Aristotle, 434, 794, 968–969, 4683
- Arithmetic disorder, 4713
- Armodafinil (Nuvigil)
  - for acute bipolar depression, 1734
  - for bipolar depression, 3110
  - for depression, in cancer patients, 2357
  - dosage and administration, 3114, 3114*t*
  - drug interactions, 3114

  - for excessive daytime sleepiness, 3111
  - immunologic effects, 3112
  - for mood and fatigue related to PD, 3110
  - pharmacodynamics, 3109
  - pharmacokinetics, 3109
  - side effects, 3113
- Aromatherapy, 4146
- Arousal, 4713
- ARP. *See* Acute phase response proteins (ARP)
- ARRA. *See* American Recovery and Reinvestment Act (ARRA)
- Arsenic, impact on CNS and PNS, 655*t*, 658
- Artane. *See* Trihexyphenidyl (Artane)
- Arterial spin labeling (ASL), 338
- Arteriosclerosis, 4713
- Artificial intelligence, 4713
- Artificial intelligence (AI), 1553–1554, 2703
- Art therapy, 4713
- ASC. *See* Altered states of consciousness (ASC)
- Ascetism, 921*t*
- Asch, Stuart, 1703
- Ascites, 1299
- ASD. *See* Acute stress disorder (ASD); Autism spectrum disorders (ASDs)
- Asenapine (Saphris), 3084–3087
  - for bipolar disorder, 1730*t*, 3085
  - chemistry, 3085
  - dosage and administration, 3087
  - drug interactions, 3086–3087
  - molecular structure, 3056*f*
  - pharmacodynamics, 3085
  - pharmacokinetics, 3085
  - precautions and adverse reactions, 3085–3086

  - for schizophrenia, 3085
  - therapeutic indications, 3085
- Asendin. *See* Amoxapine (Asendin)
- Aserinsky, Eugene, 4652
- Asexuality, definition of, 2060
- Asher, Richard, 1906
- Ashley Madison, 1127
- Ashwagandha (*Withania somnifera*), 2637
- ASI (Addiction Severity Index), 1111
- Ask Suicide-Screening Questions (ASQ), 2711, 2716
- ASL. *See* Arterial spin labeling (ASL)
- Asociality, 4713
- Asperger, Hans, 3712, 4649
- Asperger syndrome, 4713. *See also* Autism spectrum disorders (ASDs)
- ASRM. *See* Altman Self-Rating Mania Scale (ASRM)
- Assault, **2623–2627**
  - domestic violence, 2624–2625
  - hospitalization, 2626
  - initial evaluation, 2626
  - intimate partner violence, 2624–2625
  - legal issues, 2626
  - postassault psychotherapy, 2626–2627
  - psychopharmacologic treatment, 2627
  - rape, 2624–2625
  - safety assessment, 2626
  - sequelae of, 2625–2626
  - treatment issues, 2626–2627
  - violent crime, 2624
  - workplace violence, 2625
- Assay for Transposase-Accessible Chromatin (ATAC), 44
- Assertive community treatment (ACT), 1607

- Assertiveness training, 4713
- Assessment, 4714
  - of adolescents, 3527–3529
  - culturally competent, 850–858
  - of infants and toddlers, 3529–3532
  - psychometric, 3547–3551
  - reliability, 3537
  - standardized assessment instruments, 3535–3547
    - diagnostic interviews, 3537–3541
    - rating scales, 3541–3547
- Assimilation, 745, 4714
- Assimilative psychodynamic psychotherapy (APP), 2739
- Assisted reproductive techniques (ART), 2586, 2588*t*
- Association, 4714
- Associational fibers, 13–15, 20
- Associationism, 434
- Association of American Medical Colleges (AAMC), 4714
- Association of Medical Officers of American Institutions for Idiotic and Feeble-Minded Persons, 3632
- Association of Medical Superintendents of American Institutions for the Insane, 1199
- Association relay nuclei, 23
- Association studies, 256
  - neurodegenerative disorders, late-life, 4173
- Astasia-abasia, 2272, 4714
- Astereognosis, 4714
- Asterixis, 1195
- Asthenic personality, 4714
- Asthma, 2317–2320, 2318
  - child/adolescent response to, 3989
  - comorbidities with psychiatric disorders, 2319–2320
  - historical perspective, 2317–2318

  - hypnosis for, 2936
  - lifestyle interventions in, 2320
  - manifestation of, 2318
  - pathophysiology, 2318
  - prevalence of, 2318
  - social burden of, 2318
  - stress and emotion in, 2318–2319
  - treatment of
    - adjunctive cognitive and behavioral, 2320
    - pharmacologic, 2318
- Astraphobia, 4714
- Astrapophobia, 4714
- Astrocytes, 4, 86
  - fibrous, 5*t*
  - protoplasmic, 5*t*
- Asylum seekers, 836
- Asymmetries, 3676
- Asyndesis, 4714
- ATAC. *See* Assay for Transposase-Accessible Chromatin (ATAC)
- ATAC-seq, 44, 3587*t*
- *Ataque de nervios* (attack of nerves), 2022, 2551, 2555–2559, 4714
  - clinical features, 2558
  - epidemiology of, 2556–2557, 2556*t*
  - phenomenology of, 2555–2559
  - precipitants/etiology, 2557
  - psychiatric diagnoses and, 2557–2558
  - specific cultural factors, 2558–2559
  - treatment of, 2559
- Ataractic drugs. *See* Antipsychotics, 4714
- Ataraxia, 3034
- Ataxia, 469, 481t, 4714

- Ataxic dysarthria, 467*t*
- ATD. *See* Acute tryptophan depletion (ATD)
- Atenolol (Tenormin), 3224–3228
- Athetosis, 3780*t*, 4714
- Ativan. *See* Lorazepam (Ativan)
- Atomoxetine (Strattera), 98, 3128–3129
  - for ADHD and disruptive behaviors, 3652, 3736–3737
  - drug interactions, 3128–3129
  - mechanism of action, 3128
  - pharmacogenetics of, 294*t*
  - pharmacologic properties of, 3128–3129
  - for substance use disorder, 4002*t*
  - for tic disorders, 3787
- Atonia, 4714
- Atopic dermatitis (AD), 2419, 2439*f*, 2443–2444
  - clinical features, 2443
  - pathophysiology of, 2443
  - psychiatric comorbidity, 2443–2444
  - treatment of, 2444
- Atropa belladonna, 105
- Atropine, 105, 3325
  - molecular structure, 3325*f*
- Attachment, 866–867, 4714. *See also* Bonding; Reactive attachment disorder (RAD)
  - reactive attachment disorder and, 3801–3802
  - social development and, 3500
  - type D, 1964
- Attachment disorder of infancy, 4714
- Attachment disturbance, 2927
- Attachment learning, 4714
- Attachment theory, 927–928, 3500
- Attention, 301, 729

  - assessment, 3568
  - capacity, 728*f*
  - focal, 730
  - PTSD and, 1878
  - schizophrenia and, 1557
  - selective, 729
  - sustained, 728*f*, 735
- Attentional disorders, 472
  - ADHD and, 3731
- Attention-deficit disorder (ADD), **3728–3746**
  - borderline intellectual functioning and, 2532
  - malingering and, 2498
  - neurosteroids and, 139
- Attention Deficit Disorders Evaluation Scale (ADDES), 3546
- Attention-deficit hyperactivity disorder (ADHD), **3728–3739**, 4714
  - adult, 3740–3746
    - clinical features, 3742–3743
    - comparative nosology, 3741
    - course and prognosis, 3744
    - definition, 3740–3746
    - diagnosis, 3742–3743
    - differential diagnosis, 3743
    - epidemiology of, 3741
    - etiology of, 3741–3742
    - historical background, 3740–3741
    - laboratory examination, 3743
    - treatment, 3744–3745
  - alcohol use disorder and, 1306–1307
  - bipolar disorder and, 1743
  - borderline intellectual functioning and, 2532
  - buspirone for, 3173
  - cardiac disease and, 2292

- in children, 673–674
- clinical features, 3731–3732
- clonidine for, 3221–3222
- cognition in, 735–736
- comorbidities, treatment of, 3737–3738
- comparative nosology, 3729
- course and prognosis, 3733
- definition, 3728–3729
- developmental coordination disorder and, 3674
- diagnosis, 3728–3729, 3731–3732, 4777*t*
- dietary influences, 3731
- differential diagnosis, 3732–3733
- disinhibited social engagement disorder and, 3805
- disruptive behaviors and, 3652
- DSM-5 criteria for, 3731–3732
- early-onset bipolar disorder and, 3831–3832
- EEGs in, 377
- encopresis and, 3798
- enuresis and, 3791
- etiology of, 3115–3119, 3729–3731
  - environmental factors and, 3731
  - genetics, 3729–3730
  - neuroanatomical aspects, 3730
  - neurotransmitters and, 3730–3731
- family therapy for, 3911*t*, 3915
- genetic overlap with other disorders, 267
- genetics of, 3119
- group psychotherapy for, 3901–3902
- guanfacine for, 3221–3222
- history, 3728
- intellectual disabilities and, 3647, 3652
- Internet gaming disorder and, 4092

  - laboratory examination, 3732
  - malingering and, 2498, 2502–2503
  - neurobiology of, 3119
  - neuroimaging, 3604–3606
    - basal ganglia, 3605, 3605*f*
    - cerebellum, 3605–3606
    - medication effects on brain function, 3606
    - prefrontal cortex, 3604–3605
  - omega-3 fatty acids and, 2654
  - parents of child with, 4080–4081
    - child characteristics, effects of, 4080–4081
    - clinical implications, 4081
    - parent characteristics, effects of, 4080–4081
  - pathogenesis of, 3115–3119
  - pediatric epilepsy and, 680
  - perception in, 735–736
  - pharmacotherapies for, 3931–3933
  - quantitative electroencephalography of, 1079–1080
  - rating scales, 3545–3546, 3732
  - secondary, 682–683
  - sensation in, 735–736
  - sleep disorders and, 4087–4088
  - specific learning disorder and, 3665
  - temperament and, 3625
  - tic disorders and, 3776, 3781–3782
  - treatment of, 3733–3739
    - nonstimulant medications, 3736–3737
    - psychosocial, 3738–3739
    - stimulants, 3122–3125, 3733–3736, 3734*t*, 3735*t*
  - vitamin D and, 2651
- Attenuated psychosis syndrome, 1514–1515, 1646–1647
- Attitude, 4714

- Attractor networks, 318
  - dynamical system theory, 802
  - metastability, 804
  - multistability, 802
  - Resting-state functional connectivity (RSFC), 803–804
- Attractors, 802
- Attractor states, 734
- Attributable fraction, 877
- Attributable risk, 4714
- Atypical, 4714
- Atypical antipsychotics
  - for acute bipolar depression, 1734
  - for acute mania, 1729–1730
  - for anorexia nervosa, 2162
  - EEG effects, 379
- Atypical child, 4714
- Atypical depression, 1695
  - defined, 4714
  - monoamine oxidase inhibitors, 3192
- Atypical psychosis, 4714
- AUDIT (Alcohol Use Disorders Identification Test), 1111
- Auditory agnosia, 4714
- Auditory amnesia, 4714
- Auditory hallucinations, 476, 1173, 4714
- Auditory network, 804
- Auditory paired-stimuli, 1570–1571
- Auditory radiation, 11
- AUDs. *See* Alcohol use disorders (AUDs)
- August, Ralph, 2931
- Augustine, St., 4683
- Auras, 521, 523*t*, 4714. *See also* Seizure(s)
- Australian Institute of Health and Welfare (AIHW), 1294

- Authority figure, 4714
- Autism, 4714
  - calcium regulation and signaling in, 262
  - gene mapping, 261
  - genetic overlap with other disorders, 267
  - immune-CNS interactions in, 188
  - irritability associated with, 3075
  - morphogenesis of, 262
  - normal, 922
  - psychiatric emergencies, 3582
  - role of common variants in, 262–263
  - synapse formation and maintenance in, 261–262
  - traumatic brain injury and, 683
- Autism Diagnostic Interview-Revised (ADI-R), 1120
- Autism Diagnostic Interview Revised (ADI-R), 3546
- Autism Diagnostic Observation Schedule (ADOS), 3690
- Autism Diagnostic Observation Schedule (ADOS- 2), 3546
- Autism spectrum disorders (ASDs), 3489, **3711–3727**, 4714. *See also* Asperger syndrome
  - age of onset, 3721, 3721*f*
  - agitation, 3577
  - assessment instruments, 3725*t*
  - associated features of, 3722–3724
  - biologic theories, 3714
  - borderline intellectual functioning and, 2533
  - brain development, 3715–3716, 3717*t*
  - case study, 1–2
  - clinical features, 3720–3724
  - clonidine for, 3222
  - cognition in, 736–737
  - comparative nosology, 3712–3713
  - connectome-based predictive model (CPM), 3719

- course and prognosis, 3725–3726
- culture and, 3714
- definition of, 3711
- developmental neuroanatomy, 3716–3717
- diagnosis, 3712*t*, 3720–3724, 4776*t*
- differential diagnosis, 3724–3725, 3872
- EEGs in, 377–378
- electroencephalography, 3724
- epidemiology of, 3713–3714
- epigenetic dysregulation in, 52*f*
- etiology of, 3714–3715
- fragile X syndrome and, 3639
- functional neuroimaging, 3717–3718, 3720*f*
- gender differences, 3713–3714
- genetic factors, 3714–3715
- guanfacine for, 3222
- historical background, 3711–3712
- immune theories of, 3720
- impairments in social communication, 3721–3722
- laboratory examination, 3724
- medical conditions and, 3715
- motor and sensory abnormalities in, 3723–3724
- neurochemistry, 3719–3720
- neuroimaging, 3616–3617
  - global brain abnormalities, 3616
  - regional brain alterations, 3617
- neuroimaging studies, 3715–3716, 3717–3718, 3720*f*, 3724
- omega-3 fatty acids and, 2654
- overview, 3711
- parents of child with ASDs, 4075–4078
  - coping, 4077
  - psychological responses, 4076–4077

    - quality of life, 4077
    - unique stressors, 4075–4076
  - pathology of, 3724
  - pediatric OCDs and, 3841
  - perception in, 736–737
  - pharmacotherapy for, 3933
  - physical characteristics of, 3724
  - polygenic disruptions in, 50
  - pre-, peri-, and neonatal factors, 3715
  - in preterm children, 674, 674t
  - psychosocial theories, 3714
  - reactive attachment disorder and, 3805
  - restricted activities and interests, 3722
  - seizures, 3724, 3724*f*
  - selective mutism and, 3864
  - sensation in, 736–737
  - severity levels, 3713*t*
  - sex ratio, 3713–3714
  - sleep disorders and, 4088
  - social (pragmatic) communication disorder and, 3691
  - social class and, 3714
  - temperament and, 3626
  - treatment of, 3726–3727
    - complementary and alternative, 3727
    - psychopharmacology, 3727
- Autistic phase, 922
- Autistic thinking, 4714
- Autobiographical memory, 2005–2006, 2501
- Autochthonous delusion, 1170
- Autoerotic asphyxia, 4714
- Autoerotic asphyxiation, 2049
- Autoerotism, 4714

- Autoerythrocyte sensitization syndrome, 2438
- Autogenic training, 2984
- Autogynephilia, 2110, 4714
- Autoimmune diseases, 1056–1057
  - encephalopathy, autoimmune or paraneoplastic, 1056–1057
  - stress and, 2419
  - systemic lupus erythematosus, 1056
- Autoimmune psychosis, 186
- Autoinhibition, 3014
- Automatic judgment, 4714
- Automatic mentalizing, 2864
- Automatic obedience, 4714
- Automatic writing and drawing techniques, 2923
- Automatism, 4714
- Automysophobia, 4714
- Autonoesis, 805
- Autonomic nervous system (ANS), 2416
  - defined, 4714
  - sleep and, 204
- Autonomic side effect, 4714–4715
- Autonomous ego function, 4715
- Autonomy, 938, 4012, 4424–4426
- Autophagic degeneration, 60
- Autophobia, 4715
- Autoplasty, 4715
- Autopoiesis, 419
- Autopsy, 2402
- Autoreceptors, 93
- Autoscopic hallucinations, 1174
- Autoscopic psychosis, 4715
- Autosuggestibility, 1178
- Autotopagnosia, 4715

- Auxiliary therapist, 4715
- Avanafil, 3433, 3434*t*
- Aventyl. *See* Nortriptyline (Aventyl, Pamelor)
- Average, 4715
- Aversion therapy, 2817
- Aversive conditioning, 4715
- Aversive stimuli, 299
- Aversive therapy, 4715
- Avery, David, 4655–4656
- Avery, Oswald, 236
- Avicenna, 3630, 4683
- Avoidance and escape learning, 4715
- Avoidance learning, 764–765
- Avoidant attachment, 3500, 4715
- Avoidant disorder of childhood or adolescence, 4715
- Avoidant personality disorder, 2249–2250, 4715, 4837*t*
- Avoidant/restrictive food intake disorder (ARFID), 2147–2148, 4715
  - diagnosis and clinical features, 2154
  - diagnostic criteria, 2149, 4808*t*
  - differential diagnosis, 2158
  - epidemiology of, 2151
  - feeding disorder of caregiver–infant reciprocity, 3768–3769, 3769–3770
  - feeding disorder of state regulation, 3767–3768
  - infantile anorexia, 3762–3764
  - sensory food aversions, 3764–3766
  - traumatic feeding disorder, 3766–3767
  - treatment of, 2163
- Avolition, 1458
- AVP. *See* Arginine vasopressin (AVP)
- Awareness, 807–808

- Awareness training, 2818
- Awl, William, 4510
- Axline, Virginia, 3877
- Axona, 4322
- Axon hillock, 3, 70
- Axons, 3, 4715
  - action potential conduction in, 70
- Axon terminals, 3
- Ayahuasca, 388, 1354, 2903
  - for treatment-resistant depression, 2906
- Ayllon, Ted, 3885
- Ayurvedic medicine, 4147
- Azathioprine, neurotoxicity of, 2465*t*
- Azilect. *See* Rasagiline (Azilect)
- Azrin, Nate, 3885

B

- Baars, Bernard, 418
- Babinski, Joseph, 486, 1928, 1929, 2272, 4650
- Babinski sign, 470, 2272
- Back propagation, 319
- Backward chaining, 2817
- *Bacopa monnieri*, 2640
- BACS (Brief Assessment of Cognition in Schizophrenia), 1569–1570
- Bacterial toxins, 3002
- Bagging, 907
- Bailey, Percival, 4652
- Baillarger, Jules, 1652, 4686
- Baker, Elgan, 2927, 2930
- Balanced Budget Refinement Act (BBRA), 4396
- Balancing selection, 243
- Balint, Michael, 924–925
- Balint syndrome, 474
- Band-pass filter, 360
- Bandura, Albert, 971, 3885
  - cognitive behavioral therapy for, 1828–1829
  - social learning theory, 2814
- *Banisteriopsis caapi*, 1350*f*, 1354
- Banjul Charter on Human and People’s Rights, 4680
- Barber Suggestibility Scale, 2920
- Barbiturates, **3229–3234**, 4715
  - absorption, 3229
  - anesthesia, 3231–3232
  - blood concentrations, 3230
  - cerebral lateralization, 3232
  - chemistry, 3229

  - delirium tremens, 3232
  - dependence, 1427, 3233–3234
  - detoxification, 1431–1432
  - distribution, 3229
  - dosage and administration, 3233*t*, 3234
  - drug interactions, 3234
  - drug studies, 3230–3231
  - EEG effects, 379
  - history, 3229
  - intoxication, 1426–1427
  - laboratory interferences, 3234
  - mechanism of action, 3230
  - metabolism/elimination, 3229–3230, 3230*t*
  - misuse of, 1426–1427
  - molecular structure of, 3229*f*
  - narcoanalysis, 3232
  - neuroprotection, 3232–3233
  - organ/system effects of, 3231, 3347
  - overdose, 1427
  - pharmacodynamics, 3230
  - pharmacokinetics, 3229–3230
  - pharmacologic actions, 3229–3230
  - precautions/adverse reactions, 3233–3234
  - psychomotor impairment, 3233–3234
  - seizure disorders, 3231
  - sexual dysfunction and, 2047
  - teratogenicity, 3233–3234
  - therapeutic indications, 3231–3233
  - tolerance, 1426
  - withdrawal, 1427, 1431–1432, 3232
- Bariatric surgery

  - laparoscopic adjustable gastric banding (LAGB), 2315–2316, 2315*f*
  - laparoscopic sleeve gastrectomy (LSG), 2316
  - psychological considerations for, 2316
- Barlow, David, 2929
- Barlow, Horace, 437
- Barnes Akathisia Scale (BAS), 3095
- Barnett, Octo, 2699
- Barrett, Albert, 4653
- Barrough, Philip, 1230
- Barthel Index, 4186
- BAS. *See* Barnes Akathisia Scale (BAS)
- Basal ganglia (BG), 10, **28–33**, 4715
  - ADHD and, 3605, 3605*f*
  - functional circuitry, 30
  - inputs into, 30–31, 31*f*
  - internal organization, 28–29
  - internal processing, 31–32, 32*f*
  - major structures, 28–30, 29*f*
    - caudate nucleus, 28
    - globus pallidus, 28
    - putamen, 28
    - substantia innominata, 28
    - substantia nigra, 28
    - subthalamic nucleus (of Luys), 28
    - ventral pallidum, 28
    - ventral striatum, 28
  - obsessive-compulsive disorders, 3608
  - outputs of, 32–34, 33*f*
  - patterning genes, 62–63
  - schizophrenia pathology, 1505
  - Tourette syndrome, 3613

- Base rates, 1024
- Basic anxiety, 4715
- Basic benefits, 4715
- Basic systems neuroscience, **304–314**, 307*f*
  - animal models, 304–305
  - discoordination of neural discharge and oscillations, 312–313
  - local field potentials, 308–312
  - neural oscillations, 308–312
- Basiliximab, neurotoxicity of, 2465*t*
- Basolateral complex, 35
- Bateson, Gregory, 2943
- Bathophobia, 4715
- Bath salts, 1314–1315
- Batrachotoxin, 73
- Batten disease, 689
- Battered child syndrome, 3802, 4715. *See also* Child abuse
- Battery approach, 1029
- Battery tests, 4715
- Battie, William, 4643, 4684
- Battle fatigue, 2658
- Bayes, Thomas, 435
- Bayesian inference, 323, 901–902
- Bayesianism, 435
- Bayesian network, 904
- Bayesian theorem, 901
- Bayley, Nancy, 784
- Bayley Scale of Infant and Toddler Development III (Bayley-III), 3548
- BBRA. *See* Balanced Budget Refinement Act (BBRA)
- BDAE. *See Boston Diagnostic Aphasia Examination* (BDAE)
- BDD. *See* Body dysmorphic disorder (BDD)
- BDI (Beck Depression Inventory), 1104

- BDNF. *See* Brain-derived neurotrophic factor (BDNF)
- Beach, Frank, 2062
- Beadle, George, 236
- BEAM. *See* Brain electrical activity mapping (BEAM)
- Beard, George, 1689, 2272, 2562, 4686, 4715
- Beauchamp, Tom, 2881, 4424
- Beck, Aaron, 423, 731, 821, 1654, 1706, 2799, 2822, 3885, 3886, 4648, 4696
- Beck Depression Inventory (BDI), 1104, 3543
- Beck Depression Inventory II, 1044*t*
- Beck Hopelessness Scale, 2712
- Beck Scale for Suicide Ideation (BSS), 2716
- BED. *See* Binge eating disorder (BED)
- Beers, Clifford, 4687, 4715
- Befloxatone (Synthelabo), 3185*t*
- Beggs, John, 804
- Behavior
  - defined, 4715
  - evolutionary explanations, 861
  - mathematical models of, 322–323
  - nitric oxide and, 126–127
  - obesity and, 2307
  - and payoff, 761–763
  - seasonal variations in, 196
- Behavioral analysis, 2846
- Behavioral assessment, 1047
- Behavioral changes
  - in brain tumor patients, 514–517
  - in epilepsy patients, 527–529
- Behavioral conditioning, 2213
- Behavioral disorders of childhood, 4715
- Behavioral disturbances, trazodone, 3201

- Behavioral ecology, 866–867
  - attachment, 866–867
  - life history traits, 866
  - reproductive effort, 866
- Behavioral exercises, 2050
- Behavioral health managed care organizations (BHMCOs), 4481
- Behavioral inhibition (BI), 3622
- Behavioral inhibition (BI) to the unfamiliar, 3852
- Behavioral insomnia of childhood, 2168
- Behavioral interventions
  - cognitive behavioral therapy and, 4362
  - for psychosis, 4367
- Behavioral management, for dementia, 1248
- Behavioral medicine, 4715
- Behavioral neurology, 1019, 4715
- Behavioral Pathology in Alzheimer's disease (BEHAVE-AD) rating scale, 4136
- Behavioral Risk Factor Surveillance System (BRFSS), 892
- Behavioral science, 4715
- Behavioral syndromes, in neuro-Covid, 189
- Behavioral therapy, for adolescents, 3953
- Behavioral tolerance, 1298
- Behavior and Symptom Identification Scale (BASIS), 1093, 1096–1098*t*
- Behavior Assessment System for Children (BASC), 2815, 3542, 3548
- Behavior contract, 4715
- Behaviorism, 4715. *See also* Behavior therapy
- Behavior therapy, **2813–2821**, 4715. *See also* Cognitive behavioral therapy (CBT); Dialectical behavior therapy (DBT)
  - assessment and conceptualization in, 2814
  - behavior assessment methods

  - behavioral observation, 2815
  - case conceptualization, 2815
  - clinician rating tools, 2815
  - interviews, 2814–2815
  - questionnaires, 2815
  - self-monitoring, 2815
- clinical issues, 2820–2821
- complications, 2821
- contraindications, 2821
- cultural considerations, 2820
- developmental considerations, 2820
- efficacy and effectiveness of, 2989
- for enuresis, 3794
- ethical issues in, 2821
- first-wave, 2815–2817
  - aversion-based approaches, 2817
  - contingency management, 2817
  - exposure therapies, 2816
  - operant strategies, 2816–2817
  - progressive muscle relaxation, 2816
  - skills training, 2817
- history of, 2813–2814
- implementation variables, 2820
- limitations, 2821
- principles of, 2814
- second-wave, 2817–2818
  - cognitive therapy, 2818
  - habit reversal training, 2818
  - problem-solving therapy, 2818
  - self-control approaches, 2818
- for selective mutism, 3865
- for sexual dysfunctions, 2051

  - strengths of, 2820–2821
  - techniques of, 2814–2820
  - third-wave, 2818–2820
    - acceptance and commitment therapy, 2819
    - behavioral activation, 2818–2819
    - cognitive-behavioral analysis system of psychotherapy, 2819
    - dialectical behavior therapy, 2819
    - functional analytic psychotherapy, 2819–2820
    - mindfulness-based therapy, 2819
- Belatacept, neurotoxicity of, 2465*t*
- Bell, Luther, 4510
- Bell Lysaker Emotion Recognition Task (BLERT), 4551
- Bell mania, 1676
- Bemporad, Jules, 1707
- Benactyzine (Suavitil), 4663
- Benadryl. *See* Diphenhydramine (Benadryl)
- Bender, Lauretta, 3808, 4664
- Bender-Gestalt Test, 4715–4716
- Beneficence, 4012
- Benign sleep myoclonus of infancy, 2187
- Benjamin, Harry, 2104
- Benson, Herbert, 2676
- Bentham, Jeremy, 4616–4617
- Benton, Arthur, 4166
- Benzamides, 3035
- Benzodiazepine receptor agonists (BZRAs), 2172
- Benzodiazepine(s), 4663
  - absorption, 3240
  - for agitation, 2724
  - agitation, 3247–3248
  - alcohol withdrawal, 3247
  - antagonists, 3240*t*

- for anxiety disorders, 1834–1835, 3244
- for anxiety disorders in later life, 4304–4305
- bipolar disorder, 3246
- blood concentrations, 3242
- cardiovascular disorders, 3247
- catatonia, 3246–3247
- chemistry, 3239–3240
- for chronic pain, 2489
- clinical drug studies, 3243–3244
- cognitive side effects, 3249–3250
- course and prognosis, 4287
- defined, 4716
- for delirium, 1237
- delirium, 3248, 3250
- for dementia, 1251
- dementia, 3250–3251
- for dementia with Lewy bodies, 1268–1269
- dependence, 3253
- depressive disorders, 3246
- designer, 3243
- detoxification, 1428–1431
  - adjunctive medications for, 1430
  - long-term outcome, 1431
  - polysubstance misuse and, 1429–1430
  - psychological treatment, 1430–1431
  - tapering, 1428–1429
- differential diagnosis, 4286–4287
- discontinuation syndrome, 3252–3253, 3252*t*
- disinhibition, 3251
- distribution, 3240–3241
- dosage and administration, 3254
- driving and, 3251

- drug interactions, 3254
- EEG effects, 379
- in the elderly, 3249
- electroconvulsive therapy and, 3248
- epidemiology, 4286
- etiology, 4286
- $GABA_A$ receptor binding, 89
- GABA receptor complex, 3237–3238, 3238*f*
- gastrointestinal disorders, 3255
- history, 4692
- insomnia, 3244–3246
- for insomnia disorder, 2172
- intoxication, 1425, 2727
- investigational drugs, 3255
- laboratory interferences, 3254
- mechanism of action, 3242
- metabolism, 3241–3242, 3241*f*
- misuse, 3252*t*, 3253
- misuse of, 1424–1426
- nonprescription, 3243
- organ/system effects of, 3244
- overdose, 1426
- pain management in burn patients, 2469
- for pediatric anxiety disorders, 3927
- pharmacodynamics, 3242
- pharmacokinetics of, sex differences in, 3417
- pharmacologic actions, 3240–3242
- precautions and adverse effects, 3249–3254
- predictors of good response to, 3244*t*
- for PTSD, 1885
- receptor agonists and antagonists, **3237–3255**
- receptor antagonists, 3254–3255

  - receptor ligands, 3238–3239
  - receptors, 1544
  - schizophrenia, 3246
  - screening, 4286
  - structure of, 3239*f*
  - substance use disorders and, 3248
  - suicide and, 3251–3252
  - therapeutic indications, 3244–3248
  - for tic disorders, 3786
  - tolerance, 3252
  - use in pregnancy, 3248–3249
  - withdrawal and treatment, 4287
  - withdrawal syndrome, 1425–1426, 1425*t*, 2729
- Benztropine (Cogentin), 106
  - for stereotypic movement disorders, 3813
- Benztropine mesylate, 3325, 3325*f*
- Berdache, 2072, 2102
- Bereavement, 1695–1696, 1695*t*, 4231. *See also* Grief; Mourning
  - defined, 4716
  - group psychotherapy for, 3903
  - spousal, 4161
  - treatment of, 4240
  - uncomplicated, 2547–2548
- Berg, Insoo Kim, 2852
- Berger, Hans, 4660
- Berkson, Gershon, 3808
- Bernard, Claude, 791
- Berne, Eric, 962–964, 962*f*, 2957, 2983
  - human psyche, 963
  - personality theory, 962–963
  - psychopathology theory, 963
  - treatment by, 964

- Bernheim, Hippolyte, 1928, 4647
- Bernoulli random variable, 906
- Bernoulli trial, 906
- Berrios, German E., 1933, 1954
- Bertalanffy, Ludwig von, 2943
- Bertillon, Jacques, 1200
- Bertillon Classification of Causes of Death, 1200
- Berze, Josef, 4649
- Bestiality, 4716
- Best Pharmaceuticals for Children Act, 3022
- Beta (β) error, 4716
- Beta blockers, for anxiety disorders in later life, 4305–4306
- Beta waves, 4716
- Bethlem Hospital, 4643
- Betrayal blindness, 1964
- Betrayal trauma, 1964
- Between-session extinction, 2816
- BG. *See* Basal ganglia (BG)
- Bhagavad Gita, 418
- *Bhang*, 2673
- Bias, 4564–4565, 4716
- Biased agonism, 3001
- Bibring, G. L., 2273
- Bieber, Irving, 2062
- Bierer, Joshua, 4648
- Big data, 4469–4470
- Bigender, 2102
- Billod, Ernest, 4645
- Binding assays, 2996, 2996*f*
- Bindover, 4008
- Binet, Alfred, 2526, 4687
- Binge, 441–442

- Binge eating, 238, 1188
  - animal models of, 233
- Binge eating disorder (BED), 2147, 4716
  - assessing, 2153*t*
  - course and prognosis, 2159
  - diagnosis and clinical features, 2154, 4809–4810*t*
  - differential diagnosis, 2158
  - DSM-5 criteria for, 2307*t*
  - epidemiology of, 2150–2151, 2150*t*
  - history, 2149
  - pathology and laboratory examination, 2156
  - topiramate for, 3269
  - treatment of
    - psychopharmacology, 2163
    - psychotherapy, 2163
    - weight loss treatment, 2163
- Bini, Lucio, 3444, 3463
- Binswanger, Ludwig, 2980
- Biodynamics, 4716
- Biofeedback, 4716
- Biofeedback therapy
  - for functional gastrointestinal disorders, 2303
  - for insomnia disorder, 2171
- Biogenic amine, 4716
- Biogenic amine hypothesis, 4716
- Biogenic amine neurotransmitters, **93–108**. *See also* Neurotransmitter(s)
  - acetylcholine, 104–106
  - dopamine, 94–100
  - histamine, 107–108
  - imbalance in, 1655–1657
  - serotonin, 100–104

  - trace amines, 107–108
- Biologic (efficacy) assays, 2996–2997
- Biologic agents, neuropsychiatric syndromes or symptoms in, 2663*t*
- Biological therapies, 2995–3487
  - dopamine antagonists, **3033–3054**
  - dopamine-serotonin antagonists, **3054–3089**
  - dopamine-serotonin partial agonists–antagonists (DSPAs), **3100–3106**
  - drug approval process and, 3017–3026
  - drug development and, 3017–3026
  - medication-induced movement disorders, **3090–3099**
  - neuroscience-based nomenclature and, 3026–3033
  - psychopharmacology and, **2995–3016**
- Biologic modifiers, 2367–2368
- Biologic psychiatry, 4716
- Biologic rhythm, 4716
- Biologic sex differences, 3603–3604
- Biologic stressors, 1660
- Biologic therapies
  - in Africa, 3440–3441
  - around the globe, 3438–3444
  - essential list of medicines, 3438–3439, 3439*t*
  - geriatric psychopharmacology, **3438**
  - in India, 3442–3443
  - in Latin America, 3443–3444
- Biomarkers, 3020
  - dementia in Huntington disease, 1270
  - dementia in Parkinson disease, 1269–1270
  - dementia with Lewy bodies, 1266–1267
  - drug development and, 3017
  - limbic predominant age-related TDP-43 encephalopathy, 1272

  - prion diseases, 1271
  - schizophrenia, 1618–1622
  - suicide, 284–285
  - traumatic brain injury, 538–539
- *Biomarkers Consortium*, 3026
- Biomedicine. *See* Complementary, alternative, and integrative medicine (CAIM)
- Bion, Wilfred, 924, 2496, 2956
- Biophysically based circuit models, 317–319
- Biosocial theory, of borderline personality disorder, 2842–2843
- Biperiden (Akineton), 3326
- Bipolar depression, 4300
  - acute
    - clinical recommendations, 1735–1737
    - pharmacological treatment of, 1736–1737, 1736*t*
    - treatment of, 1733–1735
    - treatment options, 1737–1739
  - monoamine oxidase inhibitors, 3194
  - nonconvulsive electrotherapy for, 3478
  - SSRIs for, 3131
  - treatment of, 3835
    - dopamine receptor agonists, 3110
    - ketamine, 3382
    - quetiapine, 3079
  - unipolar depression *versus*, 1686*t*
  - valproate for, 3304
- Bipolar disorder(s), 1649, 1653, 1690–1694
  - adjunctive psychotherapy for, 1750–1752
  - agitation in, 3057, 3059, 3101, 3103
  - antipsychotics for, 4308–4309
  - benzodiazepines, 3246
  - bipolar I disorder, 1690–1692

- bipolar II disorder, 1692–1694
- in cancer patients, 2358
- carbamazepine for, 3292–3293, 3294*t*
- changes in DSM, 1205
- child/adolescent, 3437
- circadian abnormalities and, 198
- cognitive behavioral therapy for, 2824
- cognitive theory of depression, 1707
- combination therapy for, 3407–3409
  - combining two mood stabilizers, 3407–3408
  - mood stabilizer–antidepressant combinations, 3408–3409
  - mood stabilizer–antipsychotic combinations, 3408
  - mood stabilizer–thyroid hormone combinations, 3409
- with comorbidity, 1742–1743
- cyclothymic disorder, 1692
- defined, 4716
- depression associated with, 3930
- epidemiology of, 1663
- gene mapping, 263–264
- genetic overlap with other disorders, 267
- history, 4657
- HIV infection and, 594–596
- immune system and bipolar psychopharmacology, 187
- insomnia comorbidity, 2168
- interpersonal theory of depression, 1707
- lamotrigine for, 3296, 3297
- late-life, 4230–4231
- in later life, 4300
- major depressive disorder *versus*, 1696*t*
- mania in, 3057
- mixed states in, 3057
- multiple sclerosis and, 580

- neuroimaging, 3606–3608
  - amygdala, 3607–3608
  - frontal lobe, 3606–3607
  - hippocampus, 3607–3608
  - striatum, 3606–3607
- parents of child with, 4078–4080
  - caregiver burden, 4079
  - lifespan issues, 4078–4079
  - unique stressors, 4078
- perinatal, 2599
- pharmacotherapies for, 3929–3930
- phencyclidine- and other hallucinogen-induced, 1363
- psychotherapy, 1747
- rating scales for, 1105–1106
- schizophrenia, 1489
- staging or neuroprogressive models, 187
- suicide and, 2709
- after TBI, 549–550
- topiramate for, 3269
- treatment of
  - acute episodes, 1727
  - adherence, 1727
  - anti-inflammatory effects, 188
  - asenapine for, 3085
  - baseline investigations, 1725–1726
  - cognitive impairment, 1727
  - comorbidities, 1727
  - diagnosis and early intervention, 1725
  - foundations of management, 1725–1728
  - lithium, 3315–3316, 3316*f*, 3320–3321
  - mood stabilizers, 3032
  - mood symptoms, 1727–1728

  - olanzapine, 3057
  - recommendations, 1726*f*
  - suicide risk, 1727–1728, 1728*t*
  - systems review, 1725–1726
- type I, 1690–1692
  - acute mania, 1690–1691
  - alcohol use disorder and, 1306
  - bipolar mixed phase, 1692
  - chronic mania, 1691
  - cognition in, 738
  - defined, 4716
  - depressive phase, 1692
  - diagnosis of, 4786–4787*t*
  - DSM-5 criteria for, 1681–1682*t*
  - ICD-11 criteria for, 1681–1682*t*
  - perception in, 738
  - secondary mania, 1691
  - sensation in, 738
  - treatment of
    - aripiprazole, 3102, 3103
    - ziprasidone, 3061, 3062
- type II, 1692–1694
  - with acute depression, 3058–3060
  - alcohol and substance abuse, 1693–1694
  - defined, 4716
  - depressive episodes associated with, 3063
  - diagnosis of, 4787–4789*t*
  - DSM-5 criteria for, 1683–1684*t*
  - hypomania, 1693, 1741
  - hysteroid dysphoria, 1694
  - ICD-11 criteria for, 1683–1684*t*
  - leadership and creativity, 1694

    - polarity of episodes, 1693
    - rapid-cycling, 1694
    - recurrent brief hypomania, 1694
    - seasonal patterns, 1693
    - temperament and, 1693
    - treatment of, 1739–1741, 1739*t*
      - bipolar II depression, 1741
      - first-line, 1740–1741
      - hypomania, 1741
      - lurasidone, 3063
      - olanzapine, 3058–3060
      - second-line, 1741
      - third-line, 1741
    - ultrarapid cycling, 1694
  - unipolar disorder *versus*, 187
  - valproate for, 3305*t*
  - zonisamide, 3273
- Bipolar disorder with psychotic features
  - differential diagnosis, 3871–3872
  - symptoms in children and adolescents, 3868*t*
- Bipolar mania
  - paliperidone for, 3071–3072, 3075
  - quetiapine for, 3079–3080
  - risperidone for, 3071–3072, 3075
- Bipolar-Schizophrenia Network for Intermediate Phenotypes (B-SNIP), 1567–1571
- Birchwood Social Functioning Scale (SFS), 4553
- Birth history, 464
- Birth trauma, 4716
- BIS. *See* Brief Impairment Scale (BIS)
- Bisexuality, 2058, 2059, 4716. *See also* Androgyny
- Bizarre delusions, 1456–1457, 4716, 4724

- Bizarre thought, 996*t*
- Black box warning, 3009
- Blackout, 1298–1299, 4716
- Black-patch syndrome, 4716
- Blackwell, Barry, 3184
- Bladder cancer, 2363
- Blank screen, 4716
- Blast injuries, 536–537
- Bleckwenn, William, 4656
- Blended families, family therapy for, 3922
- Blended sentences, 4009
- Blepharospasm, 1195, 3774
- Bleuler, Eugen, 486, 1449, 1468, 1556, 1889, 2786, 4650, 4657, 4687, 4716
- Bleuler, Manfred, 4650
- Blind spot, 4716
- Block, Felix, 326
- Blockages, 957
- Blocking, 759, 920*t*, 4716
- Blood–brain barrier (BBB), 15, 95, 4716
- Blood chemistry, in delirium, 1233
- Blood–CSF barrier, 15
- Blood flow, imaging of, 337–338
- Blood level, 4716
- Blood pressure
  - cannabis use and, 1342
  - naltrexone effects on, 3364
  - stress, 2413*f*
- Blood serum levels, 1330
- Blumberg, H., 3358
- Blunted affect, 1459, 4716
- Blurting, 467*t*

- Board-certified psychiatrist, 4716
- Board-eligible psychiatrist, 4716
- Body-based practices, 4146
- Body dysmorphic disorder (BDD), 1841*t*, **1845–1848**, 2436–2437
  - clinical features, 1847–1848, 1847*t*, 2437
  - comparative nosology, 1846
  - course and prognosis, 1848
  - defined, 4716
  - definition of, 1845–1846
  - diagnosis, 1847–1848, 2437, 4798*t*
  - differential diagnosis, 1848
  - DSM-5 criteria for, 2437*t*
  - epidemiology of, 1846
  - etiology of, 1846–1847
  - history, 1846
  - laboratory features, 1848
  - pathology of, 1848
  - treatment of, 1848, 2437
    - combined, 1848
    - pharmacotherapy, 1848
    - psychotherapy, 1848
- Body-focused repetitive behaviors, 2437–2438. *See also* Excoriation (skin-picking) disorder; Trichotillomania (hair-pulling disorder)
  - clinical features, 2437
  - defined, 4716
  - diagnosis, 2437, 4799*t*
  - epidemiology, 2437
  - treatment of, 2437–2438
- Body image, 4716
- Body image distortion, 1175
- Body image problems, 2049

- Body language, 4716
- Body mass index (BMI), 1063–1064, 2304
- Body modification and mortification, 2674
- Body ownership, 812
- Body schema, 812
- Boltzmann distribution, 328
- Bolwig, Tom, 4661
- Bonding, 4716. *See also* Attachment
- Bone marrow transplantation, 2369
- Bonhoeffer., Karl, 4649
- Boole’s inequality, 899
- Boon, Suzette, 1933
- Bootstrap aggregating approach, 907
- Bootzin, Richard, 2171
- Borbély, Alexander, 208
- Borderline intellectual functioning (BIF), 2502, 2525–2540
  - academic or educational problem, 2537–2539
  - adaptive functioning, 2528–2529
  - assessment of, 2535
  - clinical features, 2530*t*, 2531–2532
  - comparative nosology, 2530–2531
  - course and prognosis, 2534
  - defined, 4716
  - definition of, 2525, 2530
  - differential diagnosis, 2532–2534
  - epidemiology of, 2531
  - etiology of, 2531
  - historical background, 2530–2531
  - identification of, 2531–2532
  - treatment of, 2534–2537
    - accommodation, 2535
    - behavior modification, 2536

    - mindfulness, 2536–2537
    - modifications, 2535
    - motivational messages, 2536
    - psychosocial therapies, 2535–2536
    - school-based remediation, 2535
- Borderline personality disorder (BPD), 1650, 2248–2249
  - biosocial theory of, 2842–2843
    - emotional dysregulation, 2843
    - emotional vulnerability, 2843
    - invalidating environment, 2843
  - costs of, 2805–2806
  - course of, 2805–2806
  - defined, 4717
  - diagnosis, 4836*t*
  - dialectical behavior therapy for, 2716
  - differential diagnosis, 1697, 1697*t*
  - dimensional mentalizing profile and, 2864–2865
  - dissociative disorders and, 1975
  - dopamine receptor antagonists for, 3044
  - good psychiatric management, 2807*f*, 2810*t*
    - in inpatient settings, 2811*f*
    - principles of, 2808–2813, 2808*t*
  - good psychiatric management of, good psychiatric management, 2807–2813
  - group psychotherapy for, 3903–3904
  - mentalization-based treatment for, 2872–2873
  - mentalization-based treatment of, 2755–2756
  - mood disorders and, 1697, 1697*t*
  - nonmentalizing modes and, 2865–2866, 2865*t*
  - prevalence of, 2805–2806
  - psychosis and, 1625–1626
  - symptoms, 2805

  - treatment of, **2804–2813**
    - evidence-based, 2806–2807
  - valproate for, 3308–3309
- Borderline psychosis, 4717
- Borderline state, 4717
- Bose, Kartik Chandra, 3006
- Boss, Medard, 2980
- *Bostock v. Clayton Country*, 2079
- *Boston Diagnostic Aphasia Examination* (BDAE), 1032
- *Boston Naming Test-Revised*, 1032
- Boswell, John, 2060, 2684
- Botanicals, 2637–2640
- Botulinum toxin A, 104
- Boufée délirante, 1628*t*, 4717
- *Bowers v. Hardwick*, 2076
- Bowlby, John, 731, 866, 927, 1654, 1707, 2927, 3500, 3877, 4693, 4717
- Bowman, Carl M., 4654
- *Bowring v. Godwin*, 4628
- BPD. *See* Borderline personality disorder (BPD)
- BPRS (Brief Psychiatric Rating Scale), 1102, 1103*t*
- Braceland, Francis, 4652
- Bradley, Charles, 4660, 4664, 4688
- Bradycardia, 2286
- Bradykinesia, 4717
- Bradylalia, 4717
- Bradylexia, 4717
- Brahmyadiyoga, 2640
- Braid, James, 2912, 4685
- Brain, 4717. *See also* Blood-brain barrier (BBB); Brain stimulation; Brain tumors; Forebrain; Organic brain syndrome (OBS); Traumatic brain injury (TBI)

- aging, **4148–4158**, 4156–4158
  - cognitive changes, 4155
  - cognitive reserve, 4150–4151
  - diseases, 4155–4156, 4155*t*
  - genomics, 4150
  - healthy, 4156–4158
  - lesions, 4153–4154, 4153*t*
  - mechanisms, 4151, 4151*t*
  - motor abilities, 4155
  - neuroanatomy, 4150
  - neurochemical changes, 4154–4155
  - neuroimaging, 4150
  - oldest-old and centenarians, 4156
  - pathology thresholds, 4150
  - structural changes, 4151–4153
  - studying, 4149–4150
  - tissues, 4150
- analysis of
  - association *versus* directionality, 409–410
  - static *versus* dynamic assessments, 410
- architecture of, 5–6
- biopsy, 480
- cerebellum, 460
- cerebral cortex, 461
- circuitry, 1208
- connections, 6–7
- development
  - in adolescents, 3512–3513
  - in children, 3491–3498
  - critical periods, 3498
  - epigenetic regulation of, 49–50
  - functional genomics, **43–53**

  - plasticity, 3498
  - in preterm infants, 669
  - spatial–temporal transcriptional dynamics of, 44–46
- distinctiveness of, 7–9
- emotional processing and, 1755–1759, 1755*f*, 1756*f*, 1758*f*
- fronto-subcortical circuits, 459–460
- functional connectivity, 406
  - analysis of, 407–409
  - independent components analysis, 409
  - major depression, 415
  - modalities, 407–408
  - modalities for assessment of, 407–408
  - network approaches to, 410
  - networks, 406–407, 407*f*
  - psychopathology and, 410–415
  - resting-state, 406
  - ROI–ROI or network-based connectivity, 409
  - ROI/seed-based connectivity, 409
  - schizophrenia and, 411
  - study of, 406
- functional systems of, **17–43**
  - basal ganglia, 28–33
  - cerebral cortex, 17–23
  - limbic system, 34–42
  - thalamus, 23–28
- immune-brain communication, 177–181
- inflammatory activation, 179–180
- lateralization, 459
- limbic system, 460
- major structures of, 9–10
- mind and, 727
- modulators of brain states, 462

  - morality and, 2511
  - networks, 462
  - neuropsychiatric disorders (*See* Neuropsychiatric disorders)
  - noninvasive means of studying, 3470*f*
  - organization of, 3–9
  - psychotherapy and modulation of functions of, 2780
  - schematic representation of, 1797*f*
  - of schizophrenic patients, 1503–1504
  - self-organization, 801–805
  - structural components, **9–17**
  - structures for psychotherapy, 2781–2782
  - as a system, 250
  - white matter, 461–462
- Brain abscess, 4717
- Brain death, 2401–2403
- Brain-derived neurotrophic factor (BDNF), 141–145, 281, 451–452, 1657, 2997
- Brain dysfunction, violence and, 2522–2523
- Brain electrical activity mapping (BEAM), 4717
- Brain fag, 2551
- Brain imaging
  - defined, 4717
  - of PTSD, 1880
  - in schizophrenia, **1522–1534**
- *Brain Mechanisms and Intelligence* (Lashley), 418
- Brain metabolism, 4717
- BrainMind, 805
- Brain proteome, 249
- Brain Research Institute, 4654
- Brainstem, 40–42
  - appetite and, 227
  - defined, 4717

  - emotional, 811
  - medulla, 42
  - midbrain, 40–42, 41*f*
  - pons, 42
- Brainstem auditory-evoked potentials (BAEPs), 381
- Brain stimulation
  - brain development and, 3466
  - electrical, common pathway for, 3464–3465, 3465*f*
  - electroconvulsive therapy, **3444–3463**
  - historical roots, 3463
  - mechanism of action, 3464–3466
  - methods, 82
  - methods of, 3464*t*
  - neuroimaging and, 3467–3468
  - neurophysiology and, 3467–3468
  - neuroplasticity and, 3466
  - psychiatric applications of, 3467–3468
  - psychopharmacology, comparison with, 3465, 3466*t*
  - roles of, 3467
  - spatial resolution and invasiveness in, 3466–3467, 3467*f*
  - tools
    - cranial electrical stimulation, 3473–3474
    - implanted cortical stimulation, 3477–3478
    - magnetic seizure therapy, 3474–3476
    - nonconvulsive electrotherapy, 3478–3479
    - Stanford accelerated intelligent neuromodulation therapy, 3472–3473
    - transcranial direct current stimulation, 3473
    - transcranial magnetic stimulation, 3468–3472
    - trigeminal nerve stimulation, 3478
    - vagus nerve stimulation, 3476–3477
- Brain syndrome, 4717. *See also* Organic mental disorder (OMD)

- Brain tumors
  - in children, 517–518
  - epidemiology of, 501–502, 502*f*, 503*t*
  - management of, 502–504
  - neuropsychiatry of, **501–519**
  - prognosis of, 504
  - psychiatric classification, 501
  - in psychiatric patients, 504–506
    - investigations, 506
    - “red flag” symptoms and signs, 505–506
  - psychiatric symptoms in patients, 506, 2361–2362
    - neuropsychiatric syndromes, 508–518
      - anxiety disorders, 512–513
      - depressive disorder due to another medical condition, 508–512
      - personality, cognitive, and behavioral change, 514–517
      - steroid-induced psychotic disorder, 513–514
    - predisposing factors, 506–507
      - premorbid factors, 507
      - psychosocial factors, 507
      - tumor grade, 506–507
      - tumor location, 506
    - treatment of, 507–508
      - general approach to, 507
      - psychiatric medications, 507–508
- Brainwashing, 2019, 4717. *See also* Coercive persuasion
- Brain wave, 4717
- Brawner decision, 4717
- Breast cancer, 2364
- Breastfeeding, 2601–2602. *See also* Lactation
- Breast milk, 3059
- Breath holding spells, 534*t*

- Breathing exercises, 2171
- Breathing-related sleep disorder, 4717
- Breathing-related sleep disorders, 4814*t*
- Brenman, Margaret, 1930
- Brenner, Charles, 1704
- Breuer, Josef, 915, 2746, 2750, 4647, 4717
- Brexanolone, 2598, 3243
- Brexpiprazole, 3103–3105
  - clinical trials, 3103–3104
  - dosage and administration, 1579, 3105
  - drug interactions, 3104
  - history, 3103
  - for major depressive disorder, 1724
  - mechanism of action, 3104
  - molecular structure, 3056*f*
  - pharmacodynamics, 3104
  - pharmacokinetics, 3103–3104
  - precautions and adverse reactions, 3104
  - therapeutic indications, 3104
- Brief (time-limited) psychodynamic psychotherapy, 4353
- Brief-access eating, 222–223
- Brief Assessment of Cognition in Schizophrenia (BACS), 1569–1570, 4554
- Brief Cognitive Assessment Tool for Schizophrenia (B-CATS), 4551
- Brief Impairment Scale (BIS), 885, 3546
- Brief Negative Symptom Scale (BNSS), 1458
- Brief Psychiatric Rating Scale (BPRS), 1047, 1102, 1103*t*, 4136
- Brief psychodynamic psychotherapy. *See* Intensive short-term dynamic psychotherapies (ISTDPs)
- Brief psychotherapy, 4717
- Brief psychotic disorder, 2023
  - defined, 4717

  - diagnosis of, 4781–4782*t*
- Brief reactive psychosis, 2023, 4717
- *Brief Visuospatial Memory Test-Revised*, 1032
- Brigham, Amariah, 4717
- Brill, A. A., 4717
- Briquet, Paul, 1928
- Briquet, Pierre, 2272
- Briquet syndrome, 1890, 2272, 4717
- Broad affect, 4717
- Broad autism phenotype (BAP), 4077–4078
- Broca, Paul, 34, 417
- Broca aphasia, 4717
- Broca area, 7, 8*f*
- Broca’s aphasia, 1020–1021, 1020*t*
- Brocq, Louis-Anne-Jean, 1852
- Brocq, M. L., 2435–2436
- Brodmann, Korbinian, 21
- Brofaromine (Consonar), 3185*t*
- Bromocriptine (Parlodel), 3107
  - chemistry of, 3108
  - dosage and administration, 3114, 3114*t*
  - drug interactions, 3114
  - endocrine effects, 3112
  - molecular structure of, 3108*f*
  - pharmacodynamics, 3109
  - pharmacokinetics, 3109
  - side effects, 3113
- Bromo-DragonFLY, 1354
- Brooding compulsion, 4717
- Brooks, Peter, 2877
- Brown, Daniel, 2923, 2927, 2930
- Browne, William Alexander Francis, 4648

- Brown-Séquard, Charles-Édouard, 3421
- Bruch, Hilda, 2148, 2831, 3528
- Bruininks–Oseretsky Test of Motor Proficiency (BOTMP), 3676
- Bruner, Jerome, 732
- Brunner syndrome, 2518
- Bruxism, 3774, 3809
  - defined, 4717
  - sleep-related, 2186, 2188*f*
- B-SNIP (Bipolar-Schizophrenia Network for Intermediate Phenotypes), 1567–1571
- Buber, Martin, 4737
- Bucknill, John Charles, 4685
- Bucy, Paul, 34
- Budbill, David, 2878–2879
- Buddhism, 2680
- Bujard, Herman, 815
- Bulbar dysarthria, 467*t*
- Bulimia nervosa, 2147
  - binge eating and, 1188, 2153, 2153–2154
  - compensatory behaviors, 2154
  - course and prognosis, 2159
  - defined, 4717
  - diagnosis and clinical features, 2153–2154
  - diagnostic criteria, 2149, 4809*t*
  - differential diagnosis, 2157
  - eating in patients with, 231
  - epidemiology of, 2150, 2150*t*
  - etiology of
    - developmental factors, 2152
    - environmental and social factors, 2152
    - genetic and biologic factors, 2152
    - psychological factors, 2152

  - pathology and laboratory examination, 2155*t*, 2156
  - SSRIs for, 3131
  - topiramate, 3269
  - trazodone, 3200
  - treatment of
    - psychopharmacology, 2162–2163
    - psychotherapy, 2162
- Bunney, William, 1656
- Buprenex. *See* Buprenorphine (Buprenex)
- Buprenorphine
  - for opioid use disorders, 1411, 1416–1417
  - opioid use disorders, 3353–3355
  - for opioid withdrawal, 1413–1414
  - pain management in burn patients, 2469
  - for repolarization abnormalities, 2289
  - for substance use disorder, 4001*t*
- Buprenorphine (Buprenex), 3353–3355
  - adverse effects, 3354
  - chemistry, 3354
  - clinical studies, 3354
  - dosing, 3355
  - drug interactions, 3354–3355
  - history, 3353
  - molecular formula, 3354*f*
  - pharmacodynamics, 3354
  - pharmacokinetics, 3354
  - pharmacologic actions, 3354
- Bupropion (Wellbutrin, Zyban), 98, 1391*t*, **3163–3167**
  - for ADHD, 3737
  - for anxiety disorders, 1834
  - clinical uses, 3164–3166
  - for depression

    - in cancer patients, 2356
    - in later life, 4298
  - depressive disorders, 3164–3165
  - dosing, 3167
  - drug interactions, 3167
  - eating disorders, 3166
  - formulations, 3167
  - for geriatric depression, 4239
  - obesity, 3166
  - overdose, 3167
  - pharmacodynamics, 3164
  - pharmacokinetics, 3164
    - sex differences in, 3416
  - pregnancy and, 2606
  - side effects, 3166–3167
  - smoking cessation, 3165–3166
- Burlingame, C. Charles, 4652
- Burlingham, Dorothy, 922
- Burned-out schizophrenic, 4717
- Burning mouth syndrome (BMS), 2485
- Burn injuries, **2466–2479**
  - acute phase, 2471–2472
  - alcohol/substance abuse and, 2467
  - anger in, 2474
  - in children, 2476–2478
  - delirium and, 2468
  - end-of-life considerations, 2478–2479
  - ethical considerations, 2478–2479
  - etiology of, 2466
  - face grafts after, 2474–2475
  - family support for patients, 2475
  - grief in, 2473

  - hypnosis and, 2931
  - incidence of, 2466
  - intermediate phase, 2472
  - morbidity rates, 2466
  - mortality rates, 2466
  - pain management, 2469–2470
    - hypnosis, 2470
    - pharmacology, 2469–2470
    - psychotherapy, 2470
  - plastic surgery after, 2474–2475
  - psychopathology and, 2470–2471
  - psychotherapeutic interventions, 2471–2473
  - recovery phase, 2472–2473
  - regression, 2473
  - risk factors for, 2466–2467
  - self-inflicted, 2475–2476
  - staff support, 2478
  - survival rates, 2466
- Burnout, 2687–2688, 4717
  - consequences of, 2688
  - emotional distress and, 2687–2688
  - factors driving, 2687–2688
- Burnout syndrome, 2500
- *Buros Mental Measurement Yearbook*, 1036
- Burrow, Trigant, 2956
- Burton, Robert, 1651–1652, 1898, 4684
- Bush–Francis catatonia rating scale (BFCRS), 720
- Bush–Francis catatonia screening instrument (BFCSI), 720
- BuSpar. *See* Buspirone (BuSpar)
- Buspirone (Buspar), **3168–3174**
  - agitation, 3172
  - alcohol dependence, 3172

- for anxiety disorders, 1835, 3651, 3859
- for anxiety disorders in later life, 4305
- attention-deficit/hyperactivity disorder, 3173
- blood concentration, 3170
- cannabis dependence, 3172
- chemical structures of, 3168*f*
- cholinergic system, 3170
- dopaminergic system, 3169
- drug interactions, 3173–3174
- GABAergic system, 3169–3170
- generalized anxiety disorder, 3170
- major depressive disorder, 3172
- metabolism and elimination of, 3169
- norepinephrine system, 3169
- obsessive-compulsive disorder, 3170
- organ/system effects of, 3173
- panic disorder, 3170
- for pediatric anxiety disorders, 3927
- pharmacodynamics, 3169
- pharmacokinetics, 3168–3169
- pharmacologic actions, 3168
- post-traumatic stress disorder, 3170
- precautions/adverse reactions, 3173–3174
- premenstrual dysphoric disorder, 3173
- psychosis, 3172
- serotonin system, 3169
- sexual dysfunction, 3172–3173
- sleep apnea, 3173
- smoking cessation, 3172
- somatic dysfunction, 3172
- therapeutic indications, 3173
- withdrawal symptoms, 3173

- Buss–Durkee Hostility Index (BDHI), 3546
- Busse, Ewald, 4653
- Butabarbital (Butisol), 1426*t*
- Butalbital, 1426*t*
- Butyrophenon, 59*f*
- Butyrophenones, 3035
- B vitamins, 2632–2633

C

- Cabergoline (Dostinex, Caberlin), 3107
- Caberlin. *See* Cabergoline (Dostinex, Caberlin)
- Cachexia, 2392
- Cacodemonomania, 4717
- CAD. *See* Coronary artery disease (CAD)
- Cade, John, 3313
- Caffeine, 1324, 4717
  - chemical structure, 1328*f*
  - CNS effects, 1327–1328
  - in common foods and medications, 1326–1327*t*
  - EEG effects, 380
  - insomnia and, 2170
  - metabolism, 1328
  - modification guidelines, 1331*t*
- Caffeine-related disorders, **1324–1336**
  - age and, 1329
  - anxiety disorder, caffeine-induced, 1330–1332
  - caffeine use disorder
    - clinical features, 1335
    - comorbidity, 1334
    - course and prognosis, 1335
    - diagnosis, 1334–1335
    - differential diagnosis, 1335
    - epidemiology of, 1334
    - treatment of, 1335–1336
  - comparative nosology, 1325
  - definition of, 1324
  - DSM-5 criteria for, 1324*t*, 1325
  - epidemiology of, 1325–1327
  - etiology of, 1327–1329

  - genetics, 1329
  - history, 1324–1325
  - ICD-10 criteria for, 1324*t*, 1325
  - ICD-11 criteria for, 1324*t*, 1325
  - intoxication, 1329–1330
  - nonpsychiatric illnesses and, 1336
  - performance effects, 1328–1329
  - pharmacology of, 1327–1328
  - race and, 1329
  - reinforcer effects, 1328–1329
  - sex and, 1329
  - sleep disorder, caffeine-induced, 1332
  - special populations, 1329
  - subjective effects, 1328–1329
  - treatment, 1331–1332
  - withdrawal, 1332–1334
- CAGE, 1111
- CAIM. *See* Complementary, alternative, and integrative medicine (CAIM)
- Cajal–Retzius cells, 59*f*, 63
- Calcineurin inhibitors (CNIs), neurotoxicity of, 2463–2465
- Calcium (Ca^+) channels, 75–77, 3258–3259
  - high-voltage-activated, 76
  - low-voltage-activated, 76
  - L-type, 76
  - N-type, 76
  - P/Q-type, 76
  - R-type, 76
  - subtypes of, 3259*t*
- Calcium channel blockers
  - absorption, 3260
  - chemistry, 3259–3260, 3260*t*

  - for chronic pain, 2488, 2490
  - clinical drug studies, 3261–3262
  - distribution, 3260
  - dosage and administration, 3263
  - drug interactions, 3263
  - elimination, 3260
  - investigational drugs, 3263–3264
  - laboratory interferences, 3264
  - mechanism of action, 3260–3261
  - organ/system effects of, 3262
  - pharmacodynamics, 3260–3261
  - pharmacokinetics, 3260
  - precautions and adverse reactions, 3263
  - prediction of response, 3263–3264
  - structures of, 3259*f*
  - therapeutic indications, 3262–3263
- Calcium channel inhibitors, **3256–3264**
  - absorption, 3260
  - chemistry, 3259–3260
  - clinical drug studies, 3261–3262
  - distribution, 3260
  - dosage and administration, 3263
  - drug interactions, 3263
  - elimination, 3260
  - investigational drugs, 3263–3264
  - laboratory interferences, 3264
  - mechanism of action, 3260–3261
  - organ/system effects of, 3262
  - pharmacodynamics, 3260–3261
  - pharmacokinetics, 3260
  - precautions and adverse reactions, 3263
  - prediction of response, 3263–3264

  - structures of, 3259*f*
  - therapeutic indications, 3262–3263
- Calcium regulation and signaling, 262
- California Verbal Learning Test 3 (CVLT-3), 1032
- California Verbal Learning Test-II (CVLT-II), 4551
- CAM (Confusion Assessment Method), 1114
- Cambodian Somatic Symptom and Syndrome Inventory (CSSI), 2568
- Cambridge Depersonalization Scale (CDS), 1938*t*
- CAMD. *See* Childhood autobiographical memory disturbances (CAMD)
- Cameron, D. Ewen, 4648, 4650
- Campbell, Charles, 4652
- Campral. *See* Acamprosate (Campral)
- cAMP response element-binding protein (CREB), 450–451
- Canadian Psychiatric Association (CPA), 4717
- Cancer. *See also* Carcinogenesis
  - adaptation to, 2347
  - of affluence, 2381
  - alcohol and, 1299
  - biology of, 2380
  - breast, 2364
  - child/adolescent response to, 3990
  - CNS cancers, 2361–2362
  - depression in patients with, 4229
  - ethical issues in, 2373–2374
    - clinical trials, 2373–2374
    - informed consent, 2373
    - investigational treatment, 2373–2374
    - psychiatry–ethics interface, 2373
  - families of patients with
    - evaluation of, 2374

  - psychological problems of, 2374–2375
  - psychological reciprocity, 2374
- family-centered approaches, 2374–2375
- gastrointestinal (GI), 2362
- genetic risk for, 2372
- genitourinary, 2363–2364
- gynecologic, 2364
- head and neck cancers, 2362
- hematologic malignancies, 2364–2365
- incidence rates, 2347
- initiation, 2380
- lung, 2362–2363
- mortality rates, 2347
- pain in, 2369–2370
- paraneoplastic syndromes, 2365
- of poverty, 2381
- promotion, 2380
- psychiatric disorders in
  - anxiety disorders, 2353–2354
  - bipolar disorder, 2358
  - delirium, 2350–2353
  - depressive disorders, 2354–2358
  - diagnosis and management of, 2350–2361
  - epidemiology of, 2348
  - personality disorders, 2359–2360
  - schizophrenia, 2360
  - somatoform disorders, 2361
  - substance use disorders and, 2361
  - symptoms of, 2348*t*
  - trauma- and stressor-related disorders, 2354
- psychological concerns in patients, 2369–2372
- psychological interventions, 2381

- psychological responses to, 2347
- in psychotherapist, 2376
- risk of
  - environmental factors, 2380
  - internally mediated, 2381
  - psychosocial factors, 2380
- schizophrenia comorbidity, 1595
- screening for, 2372–2373
- sexual dysfunction and, 4449–4450
- sexuality in patients with, 2370–2371
- staff
  - clinical spectrum of responses, 2375
  - psychological issues for, 2375–2376
  - stress in, 2375
  - supportive interventions for, 2375–2376
- staff-family interface, 2375
- survival, 2380
  - factors, 2381
- tobacco use and, 1387
- treatment of, 2365–2369
  - alternative, 2381–2382
  - antineoplastic hormones, 2367–2368
  - biologic modifiers, 2367–2368
  - bone marrow transplantation, 2369
  - chemotherapy, 2366–2367
  - delayed consequences of, 2372
  - immune-mediated checkpoint inhibitor immunotherapy, 2369
  - investigational, 2373–2374
  - psychiatric problems due to, 2365
  - radiation therapy, 2365–2366
  - stem cell transplantation, 2369
  - surgery, 2369

- Cancer-related fatigue (CRF), 2358–2359
  - causes of, 2359
  - defined, 2358
  - ICD-10 criteria for, 2358
  - prevalence of, 2358
  - treatment of, 2359, 2360*t*
- Canino, Gloria, 2556
- Cannabidiol, 2636
- Cannabinoid hyperemesis syndrome (CHS), 1342
- Cannabinoids
  - effects in the periphery, 136
  - endogenous, 132*f*
  - intoxication, 2727–2728
  - nonpsychoactive, 136
  - PET imaging of, 1544–1545
  - plant-derived, 131*f*
  - receptors, 132–134
  - synthetic, 131*f*
- Cannabinol, 130
- Cannabis, 130, 2636, 4021, 4744
  - adolescents and, 3999
  - EEG effects, 380
  - epidemiology, 4285
  - etiology, 4285
  - history, 4646
  - older adults and, 4285–4286
  - pharmacologic treatment, 4285
  - risks, 4285–4286
  - screening, 4285
  - spirituality and, 2673
  - withdrawal, 4286
- Cannabis use and cannabis-related disorders, **1337–1347**

- adolescent development, effects on, 1343
- adverse effects of, 1343–1345
  - cancer, 1344
  - on cardiovascular system, 1344
  - cellular effects, 1344
  - on driving, 1343–1344
  - immunologic, 1344
  - on pregnancy, 1344–1345
  - reproductive, 1344
  - on respiratory system, 1344
- amotivational syndrome, 1342
- anxiety and, 1342
- attitudes and beliefs, 1338–1339
- blood pressure and, 1342
- buspirone for, 3172
- cannabinoid hyperemesis syndrome, 1342
- cognitive impairment, 1342–1343
- comparative nosology, 1338
- correlates of cannabis use
  - age, 1339
  - availability, 1339
  - ethnicity, 1339
  - gender, 1339
  - income, 1339
  - socioeconomic status, 1339
- diagnosis and clinical features, 1340–1343
- DSM-5 criteria for, 1340
- epidemiology, 1338–1339
- flashbacks, 1343
- heart rate and, 1342
- history, 1337
- intoxication, 1341

  - intoxication delirium, 1341
  - laboratory examination, 1345
  - methods of use, 1337–1338
  - neuroimaging, 3615
  - in older adults, 4285–4286
    - epidemiology, 4285
    - etiology, 4285
    - pharmacologic treatment, 4286
    - risks, 4285–4286
    - screening, 4285
    - withdrawal, 4286
  - patterns of use, 1339
  - pharmacology, 1339–1340
  - prevalence of, 1338
  - psychosis and, 1493–1494
  - schizophrenia and, 1341–1342, 1478–1479
  - sexual dysfunction and, 2047
  - therapeutic effects of cannabinoid, 1346
  - tolerance, 1342
  - treatment of, 1345–1346
  - withdrawal, 1342, 1345
- Cannabis withdrawal syndrome (CWS), 4285
- Cannon, Walter, 2412
- Canonical correlation, 4717
- Canonical link function, 908–909
- Canon of Creativity, 2971
- Cantwell, Dennis, 3489
- CAPA. *See* Child and Adolescent Psychiatric Assessment (CAPA)
- Capacitance, 69, 3445
- Capacity, 2733
- Capgras syndrome, 476, 1170*t*, 4717–4718
- Capitation, 4718

- CAPS (Clinician-Administered PTSD Scale), 1108
- Caracas Declaration, 4680
- Carbamazepine (Tegretol), 1065–1066, 3260*t*, 3278–3295
  - acute depression, 3288–3289
  - for acute mania, 1729
  - acute mania, 3286–3288
  - for benzodiazepine detoxification, 1430
  - for bipolar disorder, 1737–1738
  - bipolar disorder, 3292–3293, 3294*t*
  - chemistry, 3260*t*, 3278–3279
  - clinical and biochemical effects of, 3280*f*
  - clinical drug studies, 3286–3289
  - for dementia, 1251
  - drug interactions, 3293–3295
  - effects of, 3279*t*
  - for geriatric bipolar disorder, 4243
  - impulse control disorders, 3293
  - metabolism, 3294*f*
  - metabolism of, 3279*f*
  - organ/system effects of, 3289–3292
  - pharmacodynamics, 3280–3286
  - pharmacogenetics of, 294*t*
  - pharmacokinetics, 3279–3280
  - precautions and adverse reactions, 3293
  - pregnancy and, 2608
  - schizoaffective disorder, 3293
  - schizophrenia, 3293
  - side effects, 3291*t*
  - therapeutic indications, 3292–3293
  - tolerance, 3281*f*, 3283–3286, 3285*t*
- Carbolith. *See* Lithium
- Carbon disulfide, impact on CNS and PNS, 655*t*, 658–659

- Carbon monoxide, 128–130
  - enzymatic generation of, 128, 129*f*
  - molecular actions of, 128
  - neuroprotective roles of heme oxygenase pathway, 129–130
  - neurotransmission, 128–129
  - signaling roles of, 129
- Carcinogenesis. *See also* Cancer
  - nalmefene and, 3364
  - naltrexone and, 3364
- Carcinoid syndrome, 717
- Carcinomatous encephalomyelitis, 2365
- Cardiac implantable electronic devices (CIEDs), 2290
- Cardinal numbers, 747
- Cardiomyopathies, 2284–2285
  - clozapine-associated cardiomyopathy, 2284–2285
  - hypertrophic cardiomyopathy, 2284
  - Takotsubo syndrome, 2284
- Cardiomyopathy, 1065
  - clozapine and, 3070
- Cardiotoxicity, 2284–2285
- Cardiovascular disorders, **2279–2293**
  - benzodiazepines, 3247
  - coronary artery disease, 2280–2282
  - cytokine-brain link and, 2419
  - electrical conduction disease, 2286–2290
  - hypertension, 2280–2282
  - masqueraders, 2290–2291
  - myocardial disease, 2282–2284
  - psychiatric masqueraders of cardiac illness, 2291
  - psychiatric symptom management of patients, 2291–2292
    - ADHD, 2292
    - anxiety, 2292

    - depression, 2291–2292
    - PTSD, 2292
  - schizophrenia comorbidity, 1594
  - sexual dysfunction and, 4449
  - valvular heart disease, 2285–2286
- Cardiovascular medications, psychiatric side effects of, 2292–2293, 2293*t*
- Cardiovascular system
  - cannabis effects, 1344
  - electroconvulsive therapy and, 3460
  - esketamine effects on, 3383
  - ketamine effects on, 3383
  - sleep and, 204
- Care and protection proceedings, 4718
- Carebaria, 4718
- Career consolidation, 791
- Care ethics, 4616*t*
- Caregiver-fabricated illness in a child, 1907
- Caregivers
  - counseling and support for, **4389–4392**
    - clinical issues, 4390–4391
      - indications, 4390–4391
      - intervention goals, 4391
      - limitations, 4391
    - definition of, 4389
    - education and support groups, 4390
    - ethical issues, 4391
    - history, 4390
    - mental health consequences, 4389
    - physical health effects, 4389–4390
    - research and evaluation, 4391–4392
    - skill-training classes, 4390

  - cultural awareness of, 1139–1142
  - defined, 4718
  - dementia and, 1249
  - inadequacy of, 1704
  - multicomponent psychotherapeutic interventions for, 4354
  - physician-assisted suicide and, 2410
  - primary, 2400–2402
- Caretaker-Teacher Report Form (C-TRF), 3542
- Cariprazine
  - for acute bipolar depression, 1734
  - for bipolar disorder, 1730*t*
  - clinical trials, 3105–3106
  - dosage and administration, 1579, 3106
  - for geriatric bipolar disorder, 4243
  - history, 3105
  - mechanism of action, 3105
  - molecular structure, 3056*f*
  - pharmacodynamics, 3105
  - pharmacokinetics, 3105
  - precautions and adverse reactions, 3106
  - therapeutic indications, 3106
- Carisoprodol (Soma), 3236
- Carlson, Eve B., 1932
- Carlsson, Arvid, 1535, 3007, 3039, 3100, 4662
- Carothers, John, 1469
- Carpenter, William, 1457, 1567
- Carr, John, 2553
- Carraz, George, 4663
- Carroll, Bernard, 1657, 4659, 4660
- CASE. *See* Chronological Assessment of Suicide Events (CASE)
- Case-control studies, 878, 4718
- Case history study, 4718

- Case presentation, 1156
- Case register, 4718
- Cass, Vivian, 2073
- Cassel, Eric, 2874
- Castration, 2100
- Castration anxiety, 4718
- Castration complex, 4718
- Catalepsy, 1461, 2917, 4718
- Cataphasia, 4718
- Cataplexy, 2175, 4718
- Catapres. *See* Clonidine (Catapres)
- Cataracts, 3081
- Catastrophic anxiety, 4718
- Catastrophic reaction
  - assessment, 494
  - etiology, 492
- Catathymia, 4718
- Catathymic crisis, 4718
- Catatonia, **718–724**, 1194, 1461, **1644–1645**, 1676–1677
  - benzodiazepines, 3246–3247
  - causes of, 719*t*
  - course and prognosis, 1645
  - defined, 4718
  - diagnosis, 719–721
  - diagnostic and clinical features, 1644–1645
  - differential diagnosis, 721, 721*t*, 1645
  - EEGs in, 375–376
  - electroconvulsive therapy, 3450
  - epidemiology of, 719, 1644
  - etiology of, 719, 1644
  - features of, 719–721
  - history, 718, 1644

  - neuroimaging of, 721–723
  - neuropathophysiology, 720–722
  - nosology, 718
  - psychotherapeutic interventions for, 724
  - signs of, 720*t*
  - subtypes of, 719*t*
  - treatment of, 723–724, 724*t*
- Catatonic excitement, 4718
- Catatonic posturing, 4718
- Catatonic rigidity, 4718
- Catatonic schizophrenia, 4718
- Catatonic state, 4718
- Catatonic stupor, 4718
- Catatonic syndrome, 469
- Catchment area, 4718
- Catecholamines, 1879
  - defined, 4718
  - neurotransmission, 3119
- Catechol-o-methyltransferase (COMT), 98, 3614–3615
- Categorical attitude, 4718
- Categorical constructs, 1091
- Categorical diagnosis, 1199
- Categorical emotion, 726
- Categorical phenotypes, 257–258, 257*f*
- Categorical self-reports, 2254
- Categorical thought, 4718
- Catharsis, 2752, 2959*t*, 4718
- Cathexis, 917, 4718
- Cation channel blockers, 3032
- Catmint (*N. menthoides*), 2640
- CAT scans. *See* Computed tomography (CT), 4718
- Cat’s cry syndrome, 4718

- Cattell, Raymond, 975, 4659
- Cattell–Horn–Carroll (CHC) theory of intelligence, 2526
- Cattell Infant Intelligence Scale, 4718
- Caudal midbrain, 207
- Caudate nucleus, 28
- Causalgia, 4718
- Causal graph models, 904
- Causality, 902, 904
- Causal paths, 904
- Causation, 876
- Cause, 42
- Cause–effect relationship, 4718
- CBCL (Child Behavior Checklist), 1115, 1116–1119*t*
- CBD. *See* Corticobasal degeneration (CBD)
- CBT. *See* Cognitive behavioral therapy (CBT)
- CCK. *See* Cholecystokinin (CCK)
- CCRTs. *See* Core conflict relational themes (CCRTs)
- CDR (Clinical Dementia Rating Scale), 1114
- Celexa. *See* Citalopram (Celexa)
- Cell autonomous, 81
- Cell death, 59–60. *See also* Apoptosis
- Cell migration, 64–65
- Cells, brain organization and, 3–5
- Cellular electrophysiology, 68–71
  - active membrane properties, 69–70
  - passive membrane properties, 69
  - resting membrane potential, 67–70
- Celsus, 1230
- Cenesthesia, 4718
- Censored observations, 910
- *Centella asiatica*, 2640
- Center for Biologic Evaluation and Research (CBER), 3022

- Center for Devices and Radiologic Health (CDRH), 3022
- Center for Drug Evaluation and Research (CDER), 3022
- Center for Epidemiological Studies-Depression Scale (CES-D), 3543–3544, 4136
- Center for Epidemiologic Studies Depression Scale Revised (CESD-R-20), 3543–3544
- Center for Food Safety and Applied Nutrition (CFSAN), 3022
- Center for Mental Health Policy and Services Research (CMHPSR), 4653
- Center for Mental Health Services (CMHS), 3942
- Center for Veterinary Medicine (CVM), 3022
- Centers for Disease Control and Prevention (CDC), 3949
- Centers for Medicare and Medicaid Services, 1159
- Central executive network (CEN), 22, 803, 1755*f*
- Central Intelligence Agency (CIA), 2019–2020
- Centrally mediated abdominal pain syndrome (CAPS), 2297, 2297*t*
- Central median nucleus, 23
- Central nervous system (CNS)
  - defined, 4718
  - effect of inflammation on, 1776–1777
  - esketamine effects on, 3383
  - immune system, 2416
  - infections, 1052
  - neoplasms, 607–608
  - regulatory landscape of, 47–49
  - sexual behavior and, 2030–2031
  - toxic agents, impact of, 655*t*
  - transcriptional landscape, 44–46
  - transcriptional regulation, 48–49
- Central nervous system (CNS) cancers, 2361–2362
- Central nervous system (CNS) tumors, ketamine effects on, 3383
- Central poststroke pain, 2484

- Central sleep apnea, 4718
  - with Cheyne–Stokes breathing, 2179
  - diagnostic criteria, 2179
  - due to high altitude periodic breathing, 2179
  - due to medical disorders, 2179
  - due to medication or substance, 2179
  - of infancy, 2179
  - of prematurity, 2179
  - primary, 2179
  - subtypes of, 2179
  - treatment-emergent, 2179
- Centromedial amygdala, 36
- Cephalalgia. *See* Headache(s)
- CERAD Behavior Rating Scale, 4136
- Cerea flexibilitas, 1461, 4718
- Cerebellar disorders, 572–576
  - clinical features, 573–574
  - cognitive features, 574
  - course and prognosis, 574
  - diagnosis, 573–574
  - differential diagnosis, 574
  - diseases that cause, 573*t*
  - dystonias, 575
  - epidemiology of, 572
  - essential tremor, 575
  - etiology of, 572–573
  - historical literature on, 572
  - laboratory examination, 574
  - motor features, 573–574
  - pathology of, 574
  - psychiatric features, 574
  - treatment of, 574

- Cerebellar hemispheres, 28
- Cerebellothalamocortical systems, 28
- Cerebellum, 460, 3605–3606
- Cerebral aqueduct, 15, 17
- Cerebral arteriosclerosis, 4718
- Cerebral cortex, 461
  - claustrum, 22
  - corticocortical connections, 21
  - efferent projections, 18*f*
  - functional circuitry, 22
  - functional networks, **17–23**, 22*t*
  - regions, 9–10
- Cerebral creatine deficiency syndromes (CCDS), 695–696
- Cerebral folate deficiency (CFS), 697
- Cerebral lateralization, 3232
- Cerebral vasculitis, 582
- Cerebrospinal fluid (CSF), 15–17
  - delirium and, 1233
  - dementia and, 1246
  - examination of, 479
- Cerebrotendinous xanthomatosis (CTX), 693–694, 694*f*
- Cerebrovascular accident, 4718
- Cerebrovascular disorders
  - anosognosia, 490
  - anxiety and, 488
  - apathy and, 488–489, 489t
  - catastrophic reaction, 490
  - definition of, 485
  - depression and, 487–488
  - history, 486–487
  - mania and, 488
  - neuropsychiatric aspects of, 485–500

  - pseudobulbar affect, 490
  - psychosis and, 488
  - vascular major neurocognitive disorder and, 487
- Cerletti, Ugo, 3444, 3463, 4660
- Certification Commission for Health Information Technology (CCHIT), 1122
- Cervical caps, 2593*t*
- CES. *See* Cranioelectrotherapy stimulation (CES)
- CESD-R-20. *See* Center for Epidemiologic Studies Depression Scale Revised (CESD-R-20)
- Cessa, Gian Luigi, 1656
- Cessation of stuttering, 467*t*
- CFI. *See* Cultural Formulation Interview (CFI)
- CFS. *See* Chronic fatigue syndrome (CFS)
- cGMP. *See* Cyclic guanosine monophosphate (cGMP)
- Chaisson, Simon, 2931
- Chalmers, David, 2231
- Channelopathies, 81
- Channelrhodopsins, 815
- Chantix. *See* Varenicline
- Chaos theory, 733
- Chaplaincy services, 2350
- Character, 2225–2231, 4718
  - biotypes, 2229*f*
  - cognitive function and, 2227*t*
  - configurations, 2240*f*
  - definition of, 2225
  - descriptors, 2226*t*
  - dimensions of, 2226*t*
  - emotional state and, 2225*t*
  - genomic approaches, 2228–2230, 2228*f*
  - positive/negative reinforcement, 2225*t*

  - psychobiology of, 2225–2228
  - psychodynamic concept of, 2230–2231
  - temperament, identity and, 2228*t*
- Character analysis, 4718
- Character armor, 955
- Character defense, 4718
- Character disorder, 4718
- Characteristic behavior patterns, 746
- Character neurosis, 2231
- Characterologic depression, 1688
- Character pathology, 2131
- Charcot, Jean-Martin, 577, 1907, 1928, 2272, 2746, 2912, 4650, 4686, 4718
- Charismatic religious sects, 2020
- Charity, 2674–2675
- Charles Bonnet syndrome, 476, 1173
- Charney, Dennis, 1208
- Charon, Rita, 2874
- Charpentier, Paul, 3034, 3055–3056
- Charter schools, 4028
- Checkpoint inhibitors, 2369
- Chemical agents, neuropsychiatric effects of, 2663*t*
- Chemical aversion, 2817
- Chemical dependence, 4718
- Chemical shift, 330
- Chemical shift imaging (CSI), 346
- Chemoarchitecture, 6
- Chemotherapy, 2366–2367
  - CNS toxicity, 2366–2367
  - nausea and vomiting, 2366
  - neurologic complications, 2368*t*
- Cheney, Clarence O., 4659

- Chewing tobacco, 1385
- Chiarugi, Vincenzo, 4642, 4684
- Chibanda, Dixon, 2568
- Chikungunya, 616
- Child abuse. *See also* Battered child syndrome; Maltreatment (child)
  - cognitive effects of, 2521
  - forensic evaluations, 4006
  - forensic psychiatry and, 4615
  - medical, 3977
  - physical abuse, 3971
    - evaluation of, 3976
    - as maltreatment, 3968–3969
    - parent interviews, 3973–3974
    - physical examination, 3975
  - psychiatric emergencies, 3582–3583
  - psychiatric emergencies and, 2731
  - psychological effects of, 2521
  - sexual abuse
    - evaluation of, 3976
    - intrafamilial, 3979
    - as maltreatment, 3969, 3971–3972
    - parent interviews, 3973–3974
    - perpetrators, 3972
    - perpetrator substitution, 3977
    - physical examination, 3975–3976
  - vulnerability to, 871
- Child Abuse Prevention and Treatment Act, 3956
- Child/Adolescent Anxiety Multimodal Study (CAMS), 3857
- Child affected by parental relationship distress (CAPRD), 2544–2545
  - clinical features, 2545

  - comorbidity, 2545
  - definition of, 2544
  - diagnosis, 2545
  - differential diagnosis, 2545
  - etiology of, 2544–2545
- Child analysis, 4718
- Child and Adolescent Psychiatric Assessment (CAPA), 885, 3540, 3541
- Child and adolescent psychologists, 4112–4113
- Child and Adolescent Service System Program (CASSP), 3942
- Child Behavior Checklist (CBCL), 1115, 1116–1119*t*, 2815
- Child Behavior Checklist for Ages 1½-5 (CBCL/1½- 5), 3542
- Child Behavior Checklist for Ages 6 to 18 (CBCL/6 to 18), 3542
- Childbirth, lithium and, 3319–3320
- Child custody, 4614–4615
- Child Dissociative Checklist (CDC), 1937*t*
- Childhood, antecedents of homosexuality in, 2073
- *Childhood and Society* (Erikson), 937, 938–940
- Childhood Autism Rating Scale, 2nd Edition (CARS 2), 3546
- Childhood autobiographical memory disturbances (CAMD), 1948–1949
- Childhood disintegrative disorder, 4719
- Childhood disorders, 1206
  - rating scales for, 1115–1120
- Childhood disruptive behavioral disorders, trazodone, 3201
- Childhood-onset schizophrenia, 3614–3615
- Childhood psychopathology, 3516
- Childhood schizophrenia, 4719
- Childhood trauma, dissociative identity disorder and, 1973–1974
- Child interview, 3523–3527
  - assessment of functioning, 3526–3527
  - beginning the interview, 3526

  - closing the interview, 3527
  - components of, 3523
  - history taking, 3523
  - inquiry about psychiatric symptoms, 3527
  - interview techniques, 3524–3526
    - direct questioning, 3525–3526
    - play techniques, 3524–3525
    - projective techniques, 3525
  - in maltreatment investigation, 3974–3975
    - content of, 3974
    - extended forensic interview, 3975
    - process, 3974
    - protocol, 3974–3975
  - medical status examination, 3524
  - mental state examination, 3527
  - structure and sequence of, 3526–3527
- Child mental health services, **4048–4060**
  - access to care, 4057–4058
    - barriers to, 4057–4058
    - models, 4057
  - core values, 4056*t*
  - future research, 4060
  - historical context, 4049–4057, 4050–4056*t*
  - quality of care, 4058–4060
    - gaps between recommended and usual care, 4058–4059
    - measurement-based care, 4059–4060
    - models of, 4058
    - national quality measures, 4059, 4059*t*
    - psychotropic medication prescription trends, 4058
- Child molesters, 2095
- Child neurology, psychiatric aspects of, **666–684**
- Child-onset fluency disorder, 4719

- Child–parent psychotherapy (CPP), 3849
- Child psychiatric examination, **3516–3585**
  - clinical interview, 3516–3528
    - assessment, 3517
    - child interview, 3523–3527
    - clinician and system factors, 3527
    - context, 3517
    - multiple informants, 3517
    - parent interview, 3517–3523
    - preliminary contact, 3517
  - cognitive assessment, 3558–3561
  - cognitive domains of assessment, 3564–3567
    - achievement testing, 3566–3567
    - intelligence, 3564–3566
  - diagnostic formulation, 3552–3553
  - neuropsychological assessment, 3563–3564
  - neuropsychological domains, 3567–3570
    - attention, 3568
    - executive functioning, 3569
    - language, 3569
    - memory, 3568
    - sensory/motor functioning, 3570
    - visuoperceptual functioning, 3569–3570
  - preschooler children, **3553–3557**
  - psychoeducational assessment, 3562–3563
  - psychological and neuropsychological assessment, 3547–3552
  - recommendations, 3552–3553
  - special types of assessment, 3532–3535
- Child psychiatry, **3488–3515**
  - adoption, **3962–3967**
  - child-related concerns, 4013
  - collaboration with school system, **4026–4040**

- culture and, 4013–4014
- current and future issues, 4016
  - administration, 4023
  - conflict, violence, and advocacy, 4018–4019
  - funding and provision of patient care, 4020, 4022–4023
  - genomics, 4017–4018
  - information technologies and data science, 4016–4017
  - marijuana, 4021
  - pharmaceutical industry, 4021–4022
  - race, ethnicity, socioeconomic status, 4019–4020
  - research, neuroethics, and psychopharmacology, 4023–4024
  - sexual orientation and gender identity, 4021
  - social media, 4020–4021
  - telepsychiatry, 4020
- disorders
  - classification, 3488–3489
  - emergence of, 3489–3490
  - onset of, 3489
- ethical issues in, **4010–4025**
- forensic, **4003–4010**
- foster care, **3956–3962**
- genetic factors, 3490
- genetics in, **3586–3594**
  - diagnostic genetic testing, 3592
  - future directions, 3593–3594
  - health equity in research, 3594
  - pharmacogenomic testing, 3592–3593
  - precision psychiatry, 3594
  - training for mental health providers, 3594
- HIV and AIDS sequelae, **3993–3997**
- hospitalization, responses to, **3981–3992**
- illnesses, responses to, **3981–3992**

  - maltreatment, **3968–3980**
  - neuroimaging studies in, 3489–3490
  - pharmacogenomics in, 3591–3592
  - protective factors, 3490
  - provision of care, 4014–4016
    - assent and consent, 4014
    - boundaries, 4016
    - confidentiality, 4014–4015
    - fidelity, 4015–4016
    - mutual trust, 4015–4016
  - religion, 4013–4014
  - reporting violations, 4026–4040
  - risk factors, 3490
  - teaching ethics, 4024–4025
  - treatment options, **3876–3955**
    - for adolescents, **3948–3955**
    - cognitive behavioral therapy, **3884–3895**
    - community-based, **3941–3947**
    - family therapy, **3908–3924**
    - group psychotherapy, **3895–3907**
    - inpatient psychiatric care, **3936–3939**
    - intensive outpatient programs, 3939
    - partial hospital programs, 3939
    - psychodynamic psychotherapy, **3876–3883**
    - psychopharmacology, **3925–3935**
    - residential treatment centers, 3939–3940
  - updates in, 3488
- Children/childhood. *See also* Adolescence/adolescents; Child abuse
  - anxiety disorders in, **3837–3866**
  - attachment types, 2927
  - behavioral insomnia, 2168

- brain tumors in, 517–518
- burn injuries, 2476–2478
  - acute stress and, 2477
  - developmental considerations, 2478
  - etiologies, 2476
  - and physical abuse, signs of, 2476*t*
  - PSTD and, 2477
  - psychological sequelae, 2477
  - risk factors, 2476*t*
  - treatment of, 2477
- delirium in, 1234–1235
- development, **3491–3504**
  - brain, 3491–3498
  - cognitive, 3496*t*, 3499–3500
  - emotional, 3492*t*
  - gender identity, 3503
  - language, 3498–3499
  - moral, 3494–3496*t*, 3501–3502
  - physical, 3496–3497*t*
  - psychosexual, 3494*t*, 3504
  - racial and ethnic identity, 3503–3504
  - sensory, 3502–3503
  - social, 3492–3493*t*, 3500–3501
- diagnostic interviews for, 884–885
- disability assessment, 885
- discrimination and, 4104–4113
- disruptive behavior disorders in, **3747–3760**
  - oppositional defiant disorder, **3747–3751**
- early-onset bipolar disorder in, 3828–3836
- end-of-life care, 2403
  - communication with severely ill children, 2405
  - end-of-life decisions, 2408

    - assent, 2408
    - consent, 2408
    - neonatal and infant issues, 2408
  - healthy siblings
    - parental response, 2407
    - psychological issues of, 2407–2408
    - understanding of sibling’s terminal illness, 2407–2408
  - infants, 2404
  - psychiatric syndromes in terminally ill children, 2405–2406
    - anxiety, 2406
    - delirium, 2406
    - depression, 2406
    - pain management in, 2406
  - school-aged children, 2404
  - toddlers, 2404
  - tripartite connection, 2403–2406
- gender dysphoria in, 2107, 2117
- HIV/AIDS in, **3993–3997**
- illness falsification, 2503
- intellectual disabilities, 3644–3648
- lithium and, 3320
- maltreatment, **3968–3980**
- mental health services, **4048–4060**
- migrants, 4113–4117
- mood disorders in, **3815–3836**
  - depressive disorders, 3815–3821
  - early-onset bipolar disorder, 3828–3836
  - suicidal behavior, 3821–3827
- obsessive-compulsive personality disorder in, **3837–3844**
- prevention of psychiatric disorders in, **4040–4047**
- psychiatric emergencies in, **3571–3585**
- psychiatric examination, **3516–3585**

    - clinical interview, 3516–3528
        - assessment, 3517
        - child interview, 3523–3527
        - clinician and system factors, 3527
        - context, 3517
        - multiple informants, 3517
        - parent interview, 3517–3523
        - preliminary contact, 3517
    - cognitive assessment, 3558–3561
    - cognitive domains of assessment, 3564–3567
        - achievement testing, 3566–3567
        - intelligence, 3564–3566
    - diagnostic formulation, 3552–3553
    - neuropsychological assessment, 3563–3564
    - neuropsychological domains, 3567–3570
        - attention, 3568
        - executive functioning, 3569
        - language, 3569
        - memory, 3568
        - sensory/motor functioning, 3570
        - visuoperceptual functioning, 3569–3570
    - preschooler children, **3553–3557**
    - psychoeducational assessment, 3562–3563
    - psychological and neuropsychological assessment, 3547–3552
    - recommendations, 3552–3553
    - special types of assessment, 3532–3535
- PTSD in, **3845–3850**
- racism and, 4104–4113
- reaction to illness and hospitalization, 3980
- sleep disorders in, **4083–4089**
    - breathing, 4084–4085
    - circadian rhythm sleep–wake disorders, 4087

    - insomnia, 4085
    - narcolepsy, 4086
    - nocturnal enuresis, 4087
    - parasomnias, 4086–4087
    - somnambulism, 4086
  - somatic symptom and related disorders in, 1892
  - stereotypic movement disorders in, **3807–3813**
  - suicidal behavior in, 3821–3827
  - suicide, 2710–2711
    - poor mental health, 2710
    - risk factors, 2710
    - treatment, 2710–2711
  - surveys of mental disorders in, 891–892
  - terrorism and school shootings, impact of, 4060–4074
- Children’s Aggression Scale (CAS), 3546
- Children’s Appreciation Test (CAT), 3548
- Children’s Communication Checklist (CCC), 3690
- Children’s Depression Inventory (CDI), 3543
- Children’s Depression Rating Scale-Revised (CDRS, CDRSR), 3544
- Children’s Global Assessment Scale (CGAS), 885, 3546
- Children’s Health Insurance Authorization Act (CHIPRA), 4561
- Children’s Health Insurance Programs (CHIP), 4561
- Children’s Hypnotic Susceptibility Scale (CHSS), 2920
- Children’s Revised Impact of Event Scale (CRIES), 3546
- Children’s Yale-Brown Obsessive-Compulsive Scale (CY-BOCS), 3545, 3546
- Childress, James, 2881, 4424
- Child welfare, racism in, 4110
- Chinese medicine, 4146–4147
- ChIP-chip (chromatin immunoprecipitation on a chip), 248
- ChIP-seq, 44

- Chi square, 4719
- Chlamydia, 1053
- Chloral hydrate (Noctec), 1428, 3234–3235
  - carcinogenicity, 3235
  - chemistry, 3234
  - dosage/administration, 3235
  - dosage and administration, 1426*t*
  - drug interactions, 3235
  - drug studies, 3234
  - laboratory interferences, 3235
  - organ/system effects of, 3234–3235
  - pharmacodynamics, 3234
  - pharmacokinetics, 3234
  - precautions/adverse reactions, 3235
  - therapeutic indications, 3235
- Chlordiazepoxide (Librium), 1429*t*, 3007, 4663
- Chloride (Cl^-) channels, 77
- Chloroacetaldehyde, 2367
- Chlorpromazine (Thorazine), 2995, 4719
  - for agitation, 3576*t*
  - for delirium, in cancer patients, 2352
  - dosage, 1579
  - history, 3007, 3034, 3055–3056, 4662
  - for stereotypic movement disorders, 3813
- Chochinov, Harvey Max, 2976
- Chodorow, Nancy, 964–967, 964*f*
  - gender identity, 964–966
  - mothering theory, 966
  - parenting roles, 965–966
  - personality theory, 964–966
  - psychopathology theory, 966
  - treatment by, 966

- Cholecalciferol, 2633
- Cholecystokinin (CCK), 122, **169**
  - anxiety disorders and, 1804–1805
  - intestinal nutrient sensing, 224–225
- Cholesterol, structure of, 1436*f*
- Cholinergic, 4719
- Cholinergic receptors, 1543–1544
- Cholinesterase inhibitors, 1258, **3335–3344**, 4316
  - adverse effects, 4318
  - categories, 3336
  - clinical applications, 4318
  - contraindications, 4318
  - for dementia with Lewy bodies, 1268
  - donepezil, 3336–3338, 4317
  - duration of treatment, 4318–4319
  - galantamine, 3341–3342, 4317–4318
  - memantine and, 3342–3344
  - pharmacodynamics, 4316*t*
  - pharmacokinetics, 4316*t*
  - rivastigmine, 3338–3341, 4317
  - tacrine, 3336, 4316–4317
- Chomsky, Noam, 813, 820
- Chorea, 3780*t*, 4719
- Chorea-acanthocytosis (ChAc), 699, 700*f*
- Choreic movements, 469
- Choroid plexus, 15
- Christal, Raymond, 975
- Christian Science, 3978
- Chromatin, 236
  - active, 277–278
  - ATP-dependent remodeling, 275
  - constitutive, 278

  - domains, 277–278
  - modifications, 272*t*
  - null, 278
  - pericentric/constitutive heterochromatin, 273–274
  - polycomb-repressed, 278
- Chromatin immunoprecipitation (ChIP), 44
- Chromosomes, 235–236
- Chronic akathisia, 3095
- Chronic cough disorder, 2323–2324
- Chronic demoralization, 1689
- Chronic depression, 1687
- Chronic fatigue syndrome (CFS), 582, 1189
  - defined, 4719
  - mood disorders and, 1699–1700
- Chronic hyperventilation, 2323
- Chronic idiopathic pruritus, 2438–2440
  - diagnosis, 2440
  - pathophysiology of, 2439–2440
  - treatment of, 2440
- Chronic illness, adherence to medical regimens in, 3904
- Chronic inflammatory demyelinating polyneuropathy (CIDP), 657
- Chronic mania, 1691
- Chronic noncancer pain, 1346
- Chronic obstructive pulmonary disease (COPD), 2320–2321
  - adjunctive cognitive and behavioral treatment modalities for, 2321
  - comorbidity with psychiatric disorders, 2321
  - diagnosis of, 2320
  - lifestyle interventions in, 2321
  - in medically ill geriatric patients, 4203–4204
  - pharmacologic treatment for, 2320
- Chronic pain

  - hypnosis for, 2933
  - in long-term care, 4408
  - in medically ill geriatric patients, 4204–4205
  - naltrexone for, 3363
  - quantitative electroencephalography of, 1081
- Chronic pelvic pain (CPP), 2595–2596
- Chronic pruritus of unknown origin, 2438–2440
- Chronic renal failure, 3322
- Chronic schizophrenia, 1513–1514
- Chronic stress, 186
  - health consequences of, 181
- Chronic traumatic encephalopathy (CTE), 544
- Chronic wasting disease (CWD), 631–632
- Chronobiology, 194, 4719
- Chronologic age, 4719
- Chronological Assessment of Suicide Events (CASE), 2715
- Chronophobia, 4719
- Chronotherapy, 2181
- Churchland, Patricia, 419
- CIDI. *See Composite International Diagnostic Interview* (CIDI)
- Cigarette smoking. *See also* Nicotine dependence
  - in adolescents, 3999
  - in older adults, 4283–4284
  - schizophrenia association with, 1596
- Cimetidine (Tagamet), 107
- Cingulate gyrus, 35
- Cingulum, 13, 461
- Circadian clocks
  - pacemakers, 194–196
    - afferent projections, 194–195
    - anatomy, 194
    - efferent projections, 195

    - molecular regulation, 195–196, 195*f*
- Circadian rhythms, 194, 1658, 4719
  - advanced sleep phase type, 2180
  - delayed sleep phase type, 2180
  - dysregulation, 1675
  - of hunger, 220, 221*f*
  - irregular sleep–wake type, 2180
  - jet lag, 2180–2181
  - non–24-hour sleep–wake type, 2180
  - nonseasonal unipolar depression and, 197
  - pharmacotherapy and, 198–199
  - psychiatric disorders, 196–197
  - shift work type, 2180
  - skin and, 2446
  - sleep regulation and, 208–209
  - sleep-wage regulation, 196
  - sleep-wake regulation, 196*f*
  - treatment of, 2181
  - unspecified type, 2181
- Circadian rhythm sleep disorder, 4719
- Circadian rhythm sleep-wake disorders, 198, 2180–2181
  - advanced sleep phase syndrome, 4212–4213
  - in children/adolescents, 4087
  - defined, 4719
  - diagnostic criteria, 2180, 4812–4813*t*
  - in older adults, 4212–4213
  - subtypes, 2180–2181
  - treatment of, 4212–4213
- Circular insanity, 1652
- Circumcision, 2028, 3978
- Circumstantiality, 4719
- Cirrhosis, 1299

- CIS. *See* Columbia Impairment Scale (CIS)
- Cisgender, 2102, 4100*t*
- *Cis*-regulatory elements (CREs), 44, 247
- Citalopram (Celexa)
  - for anxiety disorders in later life, 4224–4225
  - chemistry, 3138
  - dosing and administration, 3141–3142
  - history, 4663
  - for mood dysregulation in children and adolescents, 3819
  - pharmacogenetics of, 294*t*
  - pharmacokinetics of, sex differences in, 3416
  - pharmacologic actions, 3138
  - for tic disorders, 3785–3786
- Civil commitment, 4719
- Civil Rights of Institutionalized Persons Act (CRIPA), 4513
- CJD. *See* Creutzfeldt–Jakob disease (CJD)
- Claim, 957
- Claims review, 4719
- Clairvoyance, 4719
- Clang association, 4719
- Clanging, 4719
- Clarke, J. Christopher, 2928
- Classical conditioning, 756, 3884*t*
  - biology of memory, 775–776
  - defined, 4719
- Classical European psychopathology, 1199
- Classical mechanics, 328–329
- Classic migraine. *See* Migraine
- Classification systems, 1198
  - alternatives to traditional classification, 1208–1211
  - in medicine, 1202
  - modern psychiatric classifications in North America, 1201–1202

  - U.S., 1199–1200
- Class inclusions, 747
- Claude, Henri, 4650
- Clausen, John, 786
- Claustrophobia, 4719
- Claustrum, 22
- Clayton, Paula, 1652
- Clearinghouses, 1121
- Cleckley, Hervey, 2519
- Cleghorn, Robert, 4650
- Client-centered psychotherapy, 2974–2976
  - defined, 4719
  - ethical issues in, 2976
  - history, 2974
  - indications for, 2976
  - research and evaluation, 2976
  - techniques, 2975
  - theoretical issues, 2974–2975
- Climacteric, 4719
- Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE), 3007–3008
- Clinical Assessment Interview for Negative Symptoms (CAINS), 1458
- Clinical Dementia Rating Scale (CDR), 1114
- Clinical epidemiology, 1488
- Clinical hold, 3024
- Clinical Institute Withdrawal Assessment for Alcoholism, revised (CIWA-Ar), 1308, 4261
- Clinical interventions, 3490
- Clinical interview, 2188
  - during child psychiatric examination, 3516–3528
    - assessment, 3517

    - child interview, 3523–3527
    - clinician and system factors, 3527
    - context, 3517
    - multiple informants, 3517
    - parent interview, 3517–3523
    - preliminary contact, 3517
- Clinical neuropsychologist, 1018–1019
- Clinical neuropsychology and assessment of adults, **1018–1034**
  - assessment, 1029–1030
    - battery approach, 1029
    - hypothesis testing, 1029
    - mental status examination, 1029
    - screening, 1029
    - teleneuropsychology, 1029–1030
  - behavioral neurology and neuropsychiatry *versus*, 1019
  - formal domains, 1030–1034
    - attention and information processing, 1031–1032
    - executive functions, 1033
    - intellectual functioning, 1031
    - interview, 1031
    - language, 1032
    - memory, 1032
    - performance and symptom validity, 1033–1034
    - psychological factors, 1033
    - sensory and motor functions, 1033
    - visuospatial functions, 1032–1033
  - history, 1019
  - neuroanatomical correlates, 1019–1023
    - aphasia, 1020–1021, 1020*t*
    - executive function, 1023
    - hemispheric dominance, 1020–1023, 1020*t*
    - language disorders, 1020–1021

    - limb apraxia, 1021
    - memory disorders, 1022–1023
      - encoding, 1022–1023
      - storage and retrieval, 1023
    - motor skills, 1023
    - neglect syndromes, 1022
    - neurodegenerative language disorders, 1021
    - visuospatial disorders, 1021–1022
  - psychometric issues, 1023–1026
    - base rates, 1024
    - diversity, 1025–1026
    - moderating factors, 1025–1026
    - normative data, 1024–1025
    - predictive utility, 1024
    - premorbid function, 1025, 1025*f*
    - reliability, 1024
    - validity, 1023–1024
  - referral issues, 1026–1028
    - change in functioning over time, 1027
    - decision-making capacity, 1027
    - differential diagnosis, 1026–1027
    - ethical considerations, 1027
    - forensic evaluation, 1027
    - level of functioning, 1026
    - presurgical assessment, 1027–1028
  - therapeutic discussion of results, 1034
  - *vs*. clinical psychology, 1019
- Clinical Opiate Withdrawal Scale (COWS), 1405, 1406*t*
- Clinical psychologist, 4719
- Clinical psychology, 1019, 4719
- Clinical significance, 883
- Clinical trials, 3017–3018

  - diagnosis and patient selection, 3017–3018
  - dose, 3017
  - effect size, 3017
  - expectation bias, 3018
  - failed, 3018–3019
  - health care professional interaction, 3018
  - medication nonadherence, 3018
  - negative, 3018–3019
  - positive, 3018–3019
  - regression to the mean, 3018
  - repeated measure, 3019
  - statistical considerations, 3018
  - study designs, 3017
  - subjective outcome measures, 3018
- Clinician-Administered Dissociative States Scale (CADSS), 1938*t*
- Clinician-Administered PTSD Scale (CAPS), 1108
- Clinicopathologic conference (CPC), 1156
- Clitoris, 2028–2029
- Clock genes, 198
- Clomiphene (Clomid), 2055
  - treatment of
    - mechanical approaches, 2055
    - surgical, 2055–2056
- Clomipramine (Anafranil), 3152
  - for body dysmorphic disorder, 1848
  - desmethyl clomipramine and, 290
  - for pediatric OCD, 3843
  - pharmacogenetics of, 294*t*
  - pharmacokinetics, sex differences in, 3416
  - for stereotypic movement disorders, 3813
- Clonazepam (Klonopin), 1429*t*
  - for dementia with Lewy bodies, 1268–1269

  - history, 4663
- Clonic convulsion, 4719
- Clonic tics, 3774
- Clonidine (Catapres), 99, **3219–3224**
  - absorption, 3220
  - for ADHD, 3221–3222
  - for agitation, 3576*t*
  - anxiety disorders, 3222
  - autism spectrum disorder, 3222
  - for benzodiazepine detoxification, 1430
  - chemistry, 3219
  - for child/adolescent PTSD, 3849
  - cognition and neuroprotective effects of, 3222–3223
  - distribution, 3220
  - dosage and administration, 3223
  - drug interactions, 3223
  - drug studies, 3220–3221
  - elimination, 3220
  - history, 3219
  - mechanism of action, 3220
  - metabolism, 3220
  - nicotine withdrawal, 3222
  - opiate withdrawal, 3222
  - organ/system effects of, 3221
  - pharmacodynamics, 3220
  - pharmacokinetics, 3219–3220
  - pharmacologic actions, 3219–3220
  - precautions and adverse reactions, 3223
  - PSTD, 3222
  - therapeutic indications, 3221–3223
  - for tic disorders, 3221, 3784
  - for Tourette syndrome, 3221

- Cloninger, C. Robert, 809, 2685
- Cloninger, Robert, 2216
- Clorazepate (Tranxene), 1429*t*
- Clorgyline, 3185
- Closed adoption, 3963
- Closed-ended questions, 989, 1001
- Closed head injuries, 373
- Closed testing procedure, 900
- Closeted individuals, 2074
- Clouding of consciousness, 4719
- Clouston, Thomas, 4658
- Clozapine (Clozaril), 99, 3064–3071
  - chemistry, 3065
  - desmethylclozapine and, 290
  - discontinuing, 3070
  - dosage and administration, 1579, 3071
  - drug interactions, 3070–3071
  - effects on organs and systems, 3068–3069
  - history, 3008, 3055, 3064–3065, 4698
  - molecular structure, 3056*f*
  - pharmacodynamics, 3065–3066
  - pharmacokinetics, 3065
  - pharmacokinetics of, sex differences in, 3417
  - for psychosis, 563
  - for schizophrenia, 2, 161, 3065–3069
  - side effects, 1581
  - therapeutic indications, 3069
- Clozapine-associated cardiomyopathy, 2284–2285
- Clozaril. *See* Clozapine (Clozaril)
- Club drugs, 1348
- Clumsy child syndrome, 3673. *See also* Developmental coordination disorder (DCD)

- Cluster A personality disorders, 2245–2246
  - in later life, 4274
  - paranoid personality disorder, 2245
  - schizoid personality disorder, 2245–2246
  - schizotypal personality disorder, 2246
- Cluster B personality disorders, 2246–2249
  - antisocial, narcissistic, and histrionic spectrum, 2248
  - antisocial personality disorder, 2246–2247
  - borderline personality disorder, 2248–2249
  - histrionic personality disorder, 2248
  - in later life, 4274–4275
  - narcissistic personality disorder, 2247–2248
- Cluster C personality disorders, 2249–2251
  - avoidant personality disorder, 2249–2250
  - dependent personality disorder, 2250
  - in later life, 4275
  - obsessive-compulsive personality disorder, 2250–2251
- Cluster headache, 646–647
  - lithium in, 3320
  - treatment of, 652
- Cluster suicides, 4719
- Cluttering, 4719
- CNPs. *See* Copy number polymorphisms (CNPs)
- CNS. *See* Central nervous system (CNS)
- CNVs. *See* Copy number variants (CNVs)
- Coarse tremors, 1195
- Cobb, Stanley, 4652, 4660
- Cocaine, 4719
  - history, 1312–1313
  - neural pathway, 446
  - use disorders, 3361
  - withdrawal, 2730

- Cocaine addiction
  - epigenetic mechanisms of, 281–282
  - lithium in, 3320
  - modafinil for, 3128
- Cocaine-related disorders, 1312–1313. *See also* Stimulant-related disorders
  - naltrexone for, 3361
  - psychedelics for, 2907
  - tiagabine, 3271
- Cochran–Mantel–Haenszel (CMH) statistic, 908
- Codeine
  - pharmacokinetics, 289
  - structure of, 1400*f*
- Coding genes, 237–239
- Codons, 237
- Coefficient alpha, 1092
- Coefficient of correlation, 4719
- Coercive persuasion, 2019–2021, 4719. *See also* Brainwashing
  - CIA and, 2019–2020
  - cult mind control, 2020
  - dissociative ego states, 2020–2021
  - treatment of, 2021
- Coexistent culture, 4719
- Coffin–Lowry syndrome, 271*t*
- Cogentin. *See* Benztropine (Cogentin)
- Cognex. *See* Tacrine (Cognex)
- Cognition, 808, 4719. *See also* Cognitive disorders; Mild cognitive impairment (MCI); Neurocognition; Neurocognitive disorders
  - animal models of, 301–302
  - assessments for (*See* Cognitive assessment)
  - attention, 301
  - cold, 726

  - declarative memory, 302
  - definition of, 725
  - delusions and, 1457
  - early-onset schizophrenia and, 4265
  - everyday functioning and, 4180–4181
  - hot, 726
  - intellectual disabilities and, 3652
  - narrative, 732
  - obesity and, 2307
  - in older adults, 4135
  - preattentive processing, 301
  - in psychiatric disorders, 734–743
  - in racial prejudice, 741–742
  - social, 732–733
  - in social phobia, 742
  - working memory, 302
- Cognitive aging, 4129, 4163–4165
  - attentional processes, 4163
  - executive functions, 4163–4164
  - language functions, 4163
  - learning, 4163
  - memory, 4163
  - models of, 4164–4165
  - processing speed, 4163
  - verbal abilities, 4163
  - visuoperceptual and visuoconstrictive abilities, 4163
- Cognitive assessment, 997, 3558–3561
  - cognitive domains and tests, 3560–3561*t*
  - cultural and linguistic considerations, 3559
  - measurements in testing, 3558–3559
- Cognitive-behavioral analysis system of psychotherapy (CBASP), 1707, 2819

- Cognitive-behavioral conjoint therapy, 2909
- Cognitive/behavioral family therapy, 2944–2945
- Cognitive Behavioral Intervention for Trauma in schools (CBITS), 4070
- Cognitive behavioral therapy (CBT), 2136, **2822–2830**
  - for anorexia nervosa, 2161
  - for antisocial characteristics, 4576
  - for anxiety, 4371
  - for anxiety disorders, **1822–1831**, 2824–2825
    - acceptance-based techniques, 1827
    - anxiety management, 1826–1827
    - cognitive therapy, 1826–1827
    - efficacy, 1827–1830
    - emotional processing and, 1822–1824
    - exposure therapy, 1825–1826
    - interpersonal skills training, 1827
    - mindfulness-based training, 1827
    - stress inoculation training, 1826–1827
  - for anxiety disorders in children, 3857–3858
  - assessment and initial treatment, 4360
  - assimilative approach, 2738–2739
  - assumptions of, 3886*t*
  - for asthma, 2320
  - for binge eating disorder, 2163
  - for bipolar disorder, 2824
  - for bulimia nervosa, 2162
  - child/adolescent, **3884–3895**, 3953
    - behavioral assessment, 3887–3889, 3888*t*
    - case examples, 3892–3894
    - clinical issues, 3892
    - cognitive interventions, 3890
    - complications, 3892

  - contradictions, 3892
  - ethical issues, 3894
  - exposure/behavior change techniques, 3890–3891
  - history, 3884–3886
  - limitations, 3892
  - parental involvement, 3891–3892
  - psychoeducation, 3889
  - relapse prevention, 3892
  - research and evaluation, 3894–3895
  - somatic management techniques, 3889–3890
  - techniques, 3887–3892
  - theoretical issues, 3886–3887
  - treatment goals, 3894
- for child/adolescent PTSD, 3848–3849
- for chronic pain, 2489
- clinical issues, 2829, 4369
- cognitive therapy *versus* cognitive remediation, 4371
- complications, 2829
- components of, 3893*t*
- contraindications, 2829
- contraindications for, 4370–4371
- coping strategies, 2823
- core beliefs, 2823
- defined, 4719
- definition of, 4358
- for depression, 4371
- for depressive disorders in children and adolescents, 3820
- for eating disorders, 2313–2314, 2825
- efficacy and effectiveness of, 2989–2990
- ethical issues in, 2829–2830
- for functional gastrointestinal disorders, 2302–2303
- generic cognitive model, 2824

- goal setting in, 4369–4370
- goals of, 2829
- history, 2822
- homework, 4370
- for illness anxiety disorder, 1899
- indications for, 2829, 4370
- individual *versus* group format, 4369
- for insomnia, 2170
- intermediate beliefs, 2823
- limitations, 2829
- for major depressive disorder, 1747–1750
- medications and, 4369
- middle stage interventions, 4361–4366
  - behavioral interventions, 4362
  - cognitive interventions, 4362–4363
  - mindfulness and acceptance-based interventions, 4365
  - psychoeducation, 4361–4362
  - schema change, 4365
  - Socratic questioning, 4364–4365
  - termination and relapse prevention, 4365
- modes, 2823–2824
- modifications for psychosis, 4365–4369
- for older adults, 4353, **4359–4371**
  - ageist beliefs, 4359
  - cognitive impairment, 4359
  - COVID-19 pandemic and, 4360
  - culture and age cohorts, 4360
  - interpersonal events, 4359–4360
  - medical problems, 4359
  - role changes, 4359–4360
  - transportation barriers, 4359
- for opioid use disorders, 1415

- for pediatric OCD, 3842–3843
- for personality disorders, 2825–2826
- for postpartum anxiety and related disorders, 2600
- for psychiatric rehabilitation, 4554–4555
- for psychosis, 2799–2802
  - A-B-C model, 2801
  - booster sessions, 2802
  - case study, 2803–2804
  - coping mechanisms, 2800–2801
  - delusional beliefs, 2801–2802
  - ethical issues, 2802
  - future directions, 2803
  - history, 2785–2786
  - normalizing and, 2801
  - patient engagement, 2799–2800
  - patient’s experience and core psychotic beliefs, 2801
  - patient’s story, 2800
  - personal life narratives, 2801–2802
  - research and evaluation, 2803
  - theoretical issues, 2786–2787
- for PTSD, 1884
- research and evaluation, 2830, 4371
- schemas, 2823–2824
- for schizophrenia, 1608, 2825, 4371
- for somatic symptom disorder, 1898
- stages of, 4360
- for stimulant-related disorders, 1321–1322
- strength-based cognitive conceptualization, 2824
- for substance use disorder, 2825
- for suicidal behaviors, 2716
- techniques, 2826–2828
  - behavioral, 2827–2828

    - biologic interventions, 2828
    - cognitive, 2827
    - emotional regulation, 2828
    - environmental intervention, 2828
    - exposure, 2828
    - interpersonal, 2828
    - motivational, 2828
    - structuring, 2826–2827
    - therapeutic relationship, 2826
  - theoretical issues, 2823–2826
  - third-wave, 2990
  - for trichotillomania, 2435
  - for weight loss, 2311
- Cognitive changes
  - age-related, 1026
  - in brain tumor patients, 514–517
  - stress-related, 1026
- Cognitive control deficits, 321
- Cognitive control network (CCN), 1755–1756
- Cognitive coping, 1430
- Cognitive defusion, 2901–2902
- Cognitive development, 733, 4719
  - in adolescents, 3510
  - in children, 3496*t*, 3499–3500
  - clinical correlate, 3499–3500
  - Piaget's theory of
    - assimilation and accommodation, 745
    - equilibration, 744–745
    - stages, 745–749, 746*t*, 3499
      - concrete operations, 747
      - formal operations, 748–749
      - preoperational subperiod and semiotic function, 746

    - sensorimotor period, 745–746
  - structuralism, 745
- Cognitive difficulties, in neuro-Covid, 189–190
- Cognitive disorders. *See also* Cognition
  - dissociative amnesia and, 2006–2007
  - due to another medical condition or substance, 1284–1292, 1286*t*, 1287*t*, 1289*t*
  - organ transplantation and, 2459–2460
  - psychotic disorders in later life *versus*, 4270
- Cognitive dissonance, 4719
- Cognitive distortions, 1677, 2818
- Cognitive disturbances, 1673
- Cognitive function. *See also* Intellectual functioning (IF), character and, 2227*t*
- Cognitive impairments. *See also* Mild cognitive impairment (MCI)
  - cannabis and, 1342–1343
  - dementia and, 1242–1243
  - multiple sclerosis and, 581–582
  - sleep deprivation and, 209–210
  - treatment of, 543–544
- Cognitive interventions
  - cognitive behavioral therapy and, 4362–4363
  - for psychosis, 4367–4369
- Cognitive mentalizing, 2864
- Cognitive model of depression, 1654–1655
- Cognitive neuroscience, 319
- Cognitive processing therapy, 2818
- Cognitive remediation, 4371, 4553–4554
- Cognitive restructuring, 2423–2424, 2424*t*
- Cognitive schema, 821
- Cognitive science, 725
- Cognitive screening, 4167–4168

- Cognitive symptoms, 465
- Cognitive testing, 4135
- Cognitive theory of depression, 1706–1707
  - dysthymic and bipolar disorders, 1707
  - psychodynamic theory *versus*, 1707
- Cognitive therapy, 1826–1827, 2171, 2818. *See also* Mindfulness-based cognitive therapy (MBCT)
  - cognitive remediation *versus*, 4371
  - defined, 4719–4720
  - efficacy and effectiveness of, 2988–2989
  - mindfulness-based, 2819
- Cognitive training, for dementia, 1248
- Cognitive triad, 1706, 2818, 4720
- Cohen, Mandel, 4660
- Cohen kappa statistic, 880
- Cohesinopathy, 3641*t*
- Cohesion, 2959*t*, 4720
- Cohort, 4720
- Cohort studies, 878
- Coining, 3978
- Coitus interruptus, 4720
- Cold cognition, 726
- Cold turkey, 4720
- Coles, Robert, 2880–2881
- Collaborative Assessment and Management of Suicidality (CAMS), 2716
- Collaborative care, 3983
- Collaborative networks, 3019
- Collaterals, 6
- Collective experience, 4720
- Collective unconscious, 4720
- Colliders, 904

- Colom, Francisco, 1707
- Colombo technique, 3528
- Columbia Impairment Scale (CIS), 3546
- Columbia Suicide Severity Rating Scale (C-SSRS), 2712, 2716
- Columbia University Scale for Psychopathology in Alzheimer's Disease (CUSPAD), 4136
- Coma, 541, 4720
- Coma therapy, 4660
- Coma vigil, 4720
- Combat and operational stress, 2658–2659
- Combat fatigue, 4720
- Combination therapy, 3403–3410
  - for bipolar disorder, 3407–3409
    - combining two mood stabilizers, 3407–3408
    - mood stabilizer–antidepressant combinations, 3408–3409
    - mood stabilizer–antipsychotic combinations, 3408
    - mood stabilizer–thyroid hormone combinations, 3409
  - for major depression, 3404–3407
    - antidepressant–antipsychotic combinations, 3404
    - esketamine/ketamine–oral antidepressant combination, 3406
    - lithium–antidepressant combinations, 3404–3405
    - sex steroid–antidepressant combinations, 3406
    - stimulants/dopamine agonist–antidepressant combinations, 3405–3406
    - thyroid supplement–antidepressant combinations, 3405
  - for schizophrenia, 3409–3410
    - antidepressant–antipsychotic combinations, 3409
    - combining two antipsychotics, 3409
    - mood stabilizer–antipsychotic combinations, 3409–3410
- Combined hormonal contraceptive (CHC), 2592
- Combined therapy, 4720
- Comedos, 2442

- Comfort zones, 806
- Coming out, 2058, 2059, 2073, 2076–2077
- Command auditory verbal hallucinations, 1454
- Command automatism, 4720
- Command hallucination, 4720
- Command hallucinations, 1173–1174
- Command negativism, 4720
- Commercial sexual exploitation of children (CSEC), 3978
- Commissural fibers, 12, 20
- Commitment, 4720
- Common factitious disorder, 1907
- Common factor model of temperament/psychopathology, 3624*t*
- Common migraine. *See* Migraine
- Common-mode rejection ratio (CMRR), 360
- Common Sense Media Family Media Toolkit, 4099
- Common variant, 3587*t*
- Communication
  - doctor-patient, 2386–2387
  - evolution of, 1122–1128
    - internet, 1126–1128
    - smartphones, 1123
    - spoken word recording, 1123
    - telemedicine/telepsychiatry, 1123–1125
    - telephone, 1122–1123
    - television, 1123
    - virtual reality, 1128
    - written words, 1122
  - with parents, 2405
  - with severely ill children, 2405
- Communication disorders
  - defined, 4720
  - language disorder, **3681–3687**

    - clinical features, 3683–3685
    - diagnosis, 3683–3685, 4774*t*
    - differential diagnosis, 3685–3686
    - epidemiology of, 3681
    - etiology of, 3682–3683
    - laboratory examination, 3685
    - pathology of, 3685
    - treatment of, 3686–3687
- social (pragmatic) communication disorder, **3688–3693**
    - course, 3692–3693
    - defined, 3688
    - development, 3692–3693
    - diagnosis, 3689–3690, 4775*t*
    - differential diagnosis, 3691
    - DSM-5-TR criteria for, 3689*t*, 3690
    - epidemiology of, 3692
    - etiology of, 3691–3692
    - historical background, 3688–3689
    - symptoms, 3689–3690
    - treatment, 3693
- speech sound disorder, **3694–3698**
    - clinical features, 3696
    - comparative nosology, 3695
    - course and prognosis, 3697
    - diagnosis, 3696, 4774*t*
    - differential diagnosis, 3696, 3696–3697
    - epidemiology of, 3695
    - etiology of, 3695–3696
    - laboratory examination, 3696
    - overview, 3694–3695
    - pathology of, 3696
    - treatment of, 3697–3698

  - stuttering, **3698–3704**
    - clinical features, 3702–3703
    - comparative nosology, 3699
    - course and prognosis, 3704
    - defined, 3698
    - diagnosis, 3699, 3702–3703, 4775*t*
    - differential diagnosis, 3703–3704
    - DSM-5 and DSM-5-Tr criteria for, 3699
    - epidemiology of, 3699–3700
    - etiology of, 3700–3702
    - historical background, 3699
    - ICD-10 and ICD-11 criteria for, 3699
    - ICD-11 criteria for, 3699
    - treatment of, 3704
  - unspecified communication disorder, **3705–3710**
    - clinical features, 3708
    - comparative nosology, 3706
    - course and prognosis, 3709
    - defined, 3705–3706
    - diagnosis, 3708, 4775*t*
    - differential diagnosis, 3709
    - epidemiology of, 3706
    - etiology of, 3706–3708
    - historical background, 3706
    - laboratory examination, 3708–3709
    - pathology of, 3708–3709
    - treatment of, 3709–3710
    - voice therapies for, 3709–3710
- Community-based treatment
  - child/adolescent, **3941–3947**
    - clinical issues, 3944–3945

    - collaborative and integrated behavioral health in primary care, 3947
    - community-based interventions, 3946
    - community system evaluations, 3946–3947
    - complications, 3944
    - contraindications, 3945
    - ethical issues, 3945
    - future research, 3947
    - history, 3942
    - indications, 3944
    - limitations, 3944
    - for minorities, 3947
    - overview, 3941–3942
    - psychiatrists and, 3944
    - research and evaluation, 3945–3946
    - techniques, 3943–3944
    - theoretical issues, 3942–3943
    - treatment goals, 3945
- Community Informant Detection Tool (CIDT), 2567
- Community mental health centers (CMHCs), 4480, 4720
- Community Mental Health Centers Act, 2719
- Community psychiatry
  - in 20th century U.S., 4479–4480
  - attitudes toward, 4486–4487
  - barrier to access, 4486
  - defined, 4720
  - description of, 4477
  - development, 4480–4481
  - development of, 4481–4482
  - early American psychiatry and, 4478–4479
  - health maintenance and, 4484–4485
  - history, 4477–4480, 4648

  - HMOs and, 4480–4481
  - inpatient to outpatient, 4483–4484
  - ongoing programs, 4489–4490
    - Harris County Community Behavioral Health Program (CBHP), 4489
    - Maternal Perinatal Addiction Treatment (MPAT) clinic, 4490
    - Multi-Visit Patient Program (MVP), 4489
  - organized systems of care and, 4480–4482
  - psychiatric training and, 4487–4488
  - psychiatrist's role in, 4482–4483
  - public sector and, 4485–4486
  - recipients of, 4482
  - recovery and, 4486
  - telepsychiatry in, 4488–4489
- Comorbidity, 42, 875–876
- Comparative effectiveness research (CER), 4500
- Comparative psychiatry, 1468–1469
- Compensation, 42, 4720
- Compensation neurosis, 2494, 4720
- Competence, 4720
- Competency, 2733
  - to stand trial, 4008, 4416, 4720
- Competing response training, 2818
- Competition, 4720
- Complementarity of interaction, 4720
- Complementary, alternative, and integrative medicine (CAIM), 2627–2649
  - definition of, 2627
  - ethical issues, 2628
  - for geriatric depression, 4239 (*See also* Geriatric psychiatry)
  - historical background, 2627–2628
  - legal issues, 2628

- modalities, 2630–2631, 2631*t*
- overview, 2628–2630
- pharmacologic and related treatments, 2631–2642
  - amino acids, 2633–2635
  - botanicals, 2637–2640
  - compound herbal formulas, 2640
  - dietary modification, 2631–2632
  - essential oils, 2640
  - hormones, 2641
  - minerals, 2633
  - multinutrient formulas, 2636–2637
  - omega-3 fatty acids, 2635–2636
  - phosphatidyl choline, 2636
  - phosphatidyl serine, 2636
  - probiotics, 2632
  - psychedelics, 2641–2642
  - vitamins, 2632–2633
- safety, 2628–2630, 2629*t*
- somatic therapies, 2642
- therapies, scientifically validated, 2643–2646
  - cranioelectrotherapy stimulation, 2646
  - electroencephalographic (EEG) biofeedback, 2644–2645
  - high-density negative ions, 2644
  - HRV biofeeback, 2644–2645
  - light, 2644
  - music and sound, 2643–2644
  - virtual reality graded exposure therapy, 2645–2646
- therapies not validated by current science, 2646–2649
  - acupuncture, 2647–2648
  - energy psychology, 2649
  - healing touch, 2648
  - homeopathy, 2648

    - qigong, 2648–2649
    - reiki, 2649
    - shamanic healing, 2649
    - somatoemotional release and craniosacral therapy, 2649
    - therapeutic touch, 2648
- Complementary and integrative medicine (CIM)
  - acupuncture, 4146
  - aromatherapy, 4146
  - for autism spectrum disorders (ASDs), 3727
  - Ayurvedic medicine, 4147
  - body-based practices, 4146
  - Chinese or oriental traditional medicine, 4146–4147
  - clinical suggestions, 4141*t*
  - conventional physical activity, 4141–4142
  - definition of, 4137
  - diet, 4142
  - digitization, 4147
  - epidemiology of, 4138–4147
  - future directions, 4147
  - in geriatric psychiatry, **4137–4148**
  - importance of, 4137–4138
  - mind-body therapies, 4139–4141
  - natural products, 4142–4146, 4145*t*, 4146*t*
  - persons with mental illnesses, 4139
  - types of, 4138*f*
- Complementary countertransference, 2751
- Complementary DNAs, 245
- Complementary role, 4720
- Complement factors, 175–176
- Complete abortion, 2590*t*
- Complete androgen insensitivity syndrome (CAIS), 2068
- Complex behavioral states, 79–80

- Complexes, 950, 4720
- Complex partial seizures (CPSs), 369, 4720. *See also* Epileptic equivalent
- Complex PTSD, 740–741
- Complex regional pain syndrome (CRPS), 2484
- Complicated grief/bereavement, 4720
- Comportment changes, 545–546
- Composite hypothesis, 898
- *Composite International Diagnostic Interview* (CIDI), 883
  - University of Michigan Version, 884
  - World Mental Health Surveys Version, 884
- Compound questions, 1001
- Comprehensive (omnibus) rating scales, 3541–3543
- Comprehensive achievement testing, 3567
- Comprehensive Assessment of At-Risk Mental State (CAARMS), 1480
- Comprehensive behavioral intervention for tics (CBIT), 3783
- Comprehensive Community Mental Health Services for Children and their Families Program, 3942
- Comprehensive school mental health programs (CSMHPs), 4037
- *Comprehensive Textbook in Psychiatry, The*, 1
- Compulsion, 3780*t*, 4720. *See also* Obsessive-compulsive disorders (OCDs)
- Compulsive behaviors, 300, 564, 1196–1197
- Compulsive character, 955
- Compulsive personality, 4720
- Compulsive sexual behavior, 2048, 2048*t*
- Compulsive use, 1316
- Computational modeling, 315–326
  - biophysically based circuit models, 317–319
  - connectionist models, 319–322
  - large-scale brain dynamics and, 324

  - mathematical models of behavior, 322–323
  - of neurophysiologic biomarkers, 323–324
  - phenotyping, 324–325
  - of treatments, 325–326
- Computed tomography (CT)
  - in aging adult (*See* Geriatric psychiatry)
  - brain imaging, 1061
  - defined, 4720–4721
  - neuroimaging and, 537–538
- Computers, **2699–2704**
  - apps and, 2703
  - artificial intelligence and, 2703
  - Health 2.0, 2701–2703
  - historical background, 2699
  - improving technologies, 2700–2701
  - internet and, 2699–2700
- Conation, 4721
- Conceptual priming, 774
- Concerta. *See* Methylphenidate
- Concordance, 4721
- Concordant transference, 2751
- Concrete operations, 747, 4721
- Concrete thinking, 4721
- Concurrent validity, 880, 1037, 1092
- Concussion, 4721
- Condensation, 4721
- Conditional probability and independence, 896
- Conditional response, 756
- Conditional stimulus, 756
- Conditional survival rate, 910
- Conditioned inhibitor, 759
- Conditioned reinforcers, 762, 4760. *See also* Reinforcement

- Conditioned response. *See* Conditioning
- Conditioned stimulus (CS). *See* Conditioning
- Conditioning, 792, 4721. *See also* Behavior therapy; Reinforcement
  - aversive, 4715
  - behavioral, 2213
  - classical, 756, 3884*t*, 4719
  - instrumental, 756, 4739
  - operant, 756, 1315, 4750
  - Pavlovian, 756, 4752
  - respondent, 1315
  - second-order conditioning, 760
- Conditioning therapy. *See* Behavior therapy
- Conditions not attributable to a mental disorder, 4721
- Condoms, 2593*t*
- Conductance, 69
- Conduct disorder, **3752–3760**. *See also* Oppositional defiant disorder (ODD)
  - biologic factors, 3755–3756
  - borderline intellectual functioning and, 2533
  - clinical features, 3757–3758
  - comorbid conditions, 3758
  - comparative nosology, 3754
  - course and prognosis, 3758–3759
  - defined, 4721
  - definition of, 3752–3753
  - diagnosis, 3757–3758, 4823*t*
  - differential diagnosis, 3758
  - DSM-5-TR criteria for, 3752*t*
  - epidemiology, 3754
  - etiology, 3754–3755
  - family factors, 3757

  - family therapy for, 3911*t*, 3915–3918
  - group psychotherapy for, 3901
  - heritable factors, 3755–3756
  - history, 3753–3754
  - intellectual disabilities and, 3647
  - neighborhood and cultural factors, 3757
  - neuroanatomic factors, 3756
  - neuroendocrinologic factors, 3756
  - neurotransmitter factors, 3756
  - peer factors, 3757
  - physiologic factors, 3756
  - prenatal and perinatal factors, 3756–3757
  - psychological factors, 3757
  - psychostimulants for, 3124–3125
  - risk factors, 3755*t*
  - sociologic factors, 3757
  - subtypes and specifiers, 3753
  - traumatic brain injury and, 683
  - treatment, 3737–3738, 3759–3760
- Conduction aphasia, 1020–1021, 1020*t*
- Confabulation, 1182, 4721. *See also* Paramnesia
- Confession, 2674
- Confidence intervals, 897–898, 4721
- Confidentiality, 2404, 4620–4621, 4721. *See also* Privilege
  - of adolescent assessments, 3528
  - in assessment of adolescents, 3528
  - in correctional psychiatry, 4628–4629
  - ethics of psychiatry, 4620–4621
  - in forensic psychiatry, 4609–4610, 4620–4621
  - group psychotherapy, 2962
  - interviews, 987
  - paraphilic disorders, 2100

  - psychiatric emergencies, 2732–2733
- Conflict, 4721
- Conflict-free area, 4721
- Conflicts of interest, 4622
- Confound, 42
- Confounding, 877
- Confrontation, 2744, 4721
- Confusion, 4721
- Confusional arousals, 2183
- Confusion Assessment Method (CAM), 1114, 4136
- Congenital, 4721
- Congruence, 2975. *See also* Client-centered psychotherapy
- Conjoint therapy, 4721. *See also* Quadrangular therapy
- Conjugal paranoia, 4721
- Connectionist models, 319–322. *See also* Parallel distributed processing (PDP) models
  - applications in psychiatry, 321
  - of cognitive control deficits, 321
  - of fear conditioning, 321
  - of hallucinations in schizophrenia, 320–321
- Connective tissue disorders, neuropathies of, 659
- Connectome-based predictive model (CPM), pivotal response treatment (PRT), 3719
- Conners, C. Keith, 3545
- Conners, Loney, and Milich (CLAM) Scale, 3545
- Conners Abbreviated Symptom Questionnaire, Hyperactivity Index (CASQ-HI), 3545
- Conners ADHD Index, 3545
- Conners Rating Scales, 1120, 3542
- Connolly. John, 4643
- Conn syndrome, 716
- Conolly, John, 4686

- Conscience, 4721. *See also* Superego
- Conscious, 4721
- Consciousness, 730–731, 807, 2669. *See also* Cognition; Executive functions; Trauma-related altered states of consciousness (TRASC)
  - binding problem and, 1087–1089
  - construction of, 1085
  - definition of, 1084
  - discontinuity of, 1084–1085
  - disturbances of, 731, 1175–1180
    - altered states and, 1178
    - dissociative phenomena and, 1178–1180
    - level of consciousness and, 1177
    - orientation and, 1176–1177
    - psychological and physiological factors of, 1176
    - suggestibility and, 1178
  - first-order theory of, 818–819
  - higher-order theory of, 818–819
  - post-traumatic disorders of, 540
  - resonance and, 1087–1089
  - revised higher-order theory of, 818–819
- Conscious system, 916
- Consensual validation, 2959*t*, 4721
- Consensus Cognitive Battery (MCCB), 4550–4551
- Consent, informed, 2387
  - defined, 4739
  - depressive disorders, 3819
  - ethical issues in cancer and, 2373
  - for experimental treatment, 2409
  - for forensic evaluations, 4004–4005
  - forensic geropsychiatry, 4417–4418
  - forensic psychiatry, 4621–4622

  - geriatric psychiatry, 4424–4425
  - narrative psychiatry, 2881
  - psychiatric emergencies, 2733–2734
  - psychopharmacology, 3011–3012
  - in terminally ill transplant candidates, 2454
- Consequence, 42
- Conservation, 4721
- Conservation of quantity, 747
- Conservation of weight and volume, 747
- Conservator, 4721
- Consolidated Appropriations Act of 2021, 1124
- Consolidation, 1878
- Constanitus Africanus, 4683
- Constipation, 2392–2393, 4721
- Constitution, 4721
- Constitutional types, 4721
- Constitutive chromatin, 278
- Constricted affect, 4721
- Constructional apraxia, 4721
- Constructive structuralism, 744
- Constructive use of ignorance, 3528
- Constructs
  - categorical, 1091
  - continuous, 1091
- Construct validity, 881, 1037, 1093, 3537
- Consultation–liaison psychiatry
  - access to, 2279
  - clinical practice of, 2275
  - definition of, 4722
  - evaluation process in, 2275–2276
  - history of research into, 2273–2274
  - patient care competencies, 2278*t*

  - treatments in, 2276
- Contagion, 2959*t*, 4722
- Content thought disorder, 4722
- Content validity, 880, 1092
- Contextual fear, role in contextual fear, 1811
- Contiguity, law of, 434
- Contingency management (CM), 762
  - for opioid use disorders, 1415
  - for stimulant-related disorders, 1321
- Contingency tables, 908
- Continuity of Care Health Record (CCHR), 1122
- Continuous constructs, 1091
- Continuous Performance Test, 735–736
- Continuous phenotypes, 257*f*, 258–259
- Continuous reinforcement, 4722
- Contraception, 2591–2594, 2593–2594*t*
- Contraceptive implants, 2593*t*
- Contract, 4722
- Contrast, law of, 434
- Control group, 4722
- Controllable pulse shape transcranial magnetic stimulation (TMS), 3464*t*
- Controlled mentalizing, 2864
- Controlled substance, 4722
- Controlled trials, 4659
- Controlling, 921*t*, 2748*t*, 4722
- Convergent systems, 7*f*
- Conversational catharsis, 4722
- Conversion, 4722
- Conversion disorder (functional neurologic symptom disorder), 484–485, 1900–1902, 2506
  - comparative nosology, 1900

  - defined, 4722
  - definition of, 1900
  - diagnosis of, 4806*t*
  - epidemiology of, 1900
  - etiology of, 1900–1901
  - history, 1900
  - treatment of, 1901
- Conversion experience, 3011
- Conversion paralysis, 2656
- Conversion phenomena, 4722
- Convulsion, 4722. *See also* Epilepsy; Seizure disorders
- Convulsive disorder. *See* Epilepsy
- Convulsive therapy, 4660, 4722. *See also* Electroconvulsive therapy (ECT)
- Conway, Martin, 809
- Cooley, Charles, 421
- Coping
  - age differences in, 4160–4162
  - in older adults, 4160–4162
  - parents of child with ASDs, 4077
  - self-efficacy, 4473
- Coping mechanism, 4722
- Coping with Depression Course (CWD), 4044
- Coppen, Alec, 1657
- Coprolalia, 1195, 3774, 4722
- Coprophagia, 4722
- Coprophilia, 4722
- Coprophobia, 4722
- Copropraxia, 3774
- Copy-cat suicides, 4722
- Copy number polymorphisms (CNPs), 235
- Copy number variants (CNVs), 255, 2615, 3586–3587

  - ADHD, 3730
  - definition of, 3587*t*
  - human genome and, 243, 243*f*
  - psychiatric disorders and, 3590*t*
- Copy number variations (CNVs), 1485–1486
- Core conflict relational themes (CCRTs), 2927–2928
- Corgard. *See* Nadolol (Corgard)
- Cornelia de Lange syndrome, 3641*t*
- Cornfield, Jerry, 895
- Corona radiata, 11
- Coronary artery bypass graft, 1232
- Coronary artery disease (CAD), 2280–2282
  - anxiety disorders and, 2281
  - behavioral risk factors in, 2281–2282
  - depression and, 2280–2281
  - epidemiology of, 2280
  - post-traumatic stress disorder and, 2281
  - severe mental illness and, 2282
  - stress factors in, 2281–2282
  - symptoms treatment, 2280
- Coronavirus Aid, Relief, and Economic Security (CARES) Act, 1124
- Corpora amylacea, 4153
- Corporal punishment, 3969
- Corpus callosum, 10*f*, 12, 3604*f*, 3613–3614
- Corpus luteum, 2583
- Correctional psychiatry, **4624–4641**
  - challenges in, 4626–4628
  - criminalization of mentally ill, 4627
  - defined, 4722
  - diversion and community-based interventions, 4639–4640
  - education, 4640

- electronic medical record, 4640–4641
- ethical principles, 4628–4629
  - collection of information, 4629
  - confidentiality, 4628–4629
  - doctor-patient relationship, 4628–4629
  - dual loyalty, 4628
  - medical autonomy, 4628
- future of, 4639–4641
- history of, in United States, 4625–4626
- landmark cases in, 4628
- mental illness in the correctional system, 4625
- overcrowding, 4626–4627, 4627*f*
- overview, 4624–4625
- psychiatric patients in correctional facilities, 4625
- psychiatric treatment in jails and prisons, 4628
  - access to mental health care, 4630–4631
  - accreditation and standards, 4630
  - barriers to inmates receiving services, 4630–4631, 4631*t*
  - consultations, 4631
  - co-occurring disorders, 4636
  - death row inmates, 4638
  - detoxification, 4636–4637
  - discharge planning, 4634–4635
  - disciplinary hearings, 4634
  - dying inmates, 4637
  - evidence-based treatment, 4631–4635
  - facility quality assurance plan, 4630
  - geriatric inmates, 4637
  - hunger strikes, 4639
  - information technology systems, 4630
  - institutional transfers, 4634
  - intellectual and developmental disabilities, 4637

    - involuntary treatment, 4632
    - LGBT and gender dysphoria, 4638–4639
    - life sentences, 4638
    - models of health care, 4629–4630
    - peer-to-peer programs, 4632
    - policy goal, 4629
    - purposes of, 4629, 4629*t*
    - rape and sexual assault victims, 4637–4638
    - reentry programs, 4635
    - referral, 4631
    - screening, 4631*t*
    - seclusion and restraint, 4633–4634, 4634*t*
    - self-harm monitoring and prevention, 4633
    - sick call and nonemergency requests, 4631
    - staffing, 4631
    - substance use disorders, 4636, 4636*t*
    - suicide prevention, 4632–4633, 4633*t*
    - therapeutic milieu, 4631, 4632*t*
    - treatment planning, 4632
    - women prisoners, 4635–4636
    - written policies and procedures, 4630
  - racial inequality, 4627–4628
  - shortage of psychiatrists, 4626
  - transinstitutionalization, 4627, 4627*f*
- Corrective emotional experience, 4722
- Corrective familial experience, 2959*t*
- Correlation, 897, 4722
- Corsellis, John Arthur Nicholas, 4658
- Cortex, 3612–3613, 3613*f*
- Cortical consolidation, 730
- Cortical plate stage, 59*f*
- Corticobasal degeneration (CBD), 564, 1263

  - biomarkers, 1263
  - clinical features, 1263
  - comparative nosology, 1263
  - course and prognosis, 1263
  - definition of, 1263
  - diagnosis, 1263
  - differential diagnosis, 1263
  - epidemiology of, 1263
  - epilepsy and, 1346
  - management of, 1263
  - neuropathology, 1263
- Corticolimbic disconnection hypothesis, 1954
- Corticosteroids
  - for cancer-related fatigue, 2359
  - neurotoxicity of, 2465*t*
- Cortico-striatal-thalamo-cortical (CSTC) circuit, 1757–1759, 1758*f*
- Corticotropin-releasing factor (CRF), 109*t*, 117–119, 118*f*
- Corticotropin-releasing hormone (CRH), 1759, 1799
- Cortisol, 4106–4107
  - structure of, 1436*f*
- Cosa Nostra, 2509
- Costa, J. L., 4662
- Cost-effectiveness ratio (CER), 912–913
- Cotard syndrome, 476, 4722, 4724. *See also* Nihilistic delusion
- Co-therapy, 4722
- Council for Exceptional Children (CEC), 3632
- Council on Consultation-Liaison Psychiatry, 2278
- Counseling, 4722
- Counseling on Access to Lethal Means (CALM), 2716
- Countercathexis, 916
- Counterconditioning, 760
- Counterconditioning model, 2816

- Counterphobia, 4722
- Countertransference, 857–858, 931–932, 1993, 2713, 2750–2751, 4014, 4356, 4385, 4722
- Couples therapy, **2942–2952**, 4722. *See also* Family therapy; Marriage therapy
  - areas of inquiry and observation in, 2951*t*
  - clinical applications, 2949–2952
  - clinical management issues, 2952*t*
  - contraindications for, 2950*t*
  - definition of, 2944
  - group, 2952–2953
  - indications for, 2950*t*
  - initial interview, 2951*t*
- Covariance, 897
- Covert sensitization, 4722
- COVID-19, 610–615, 660–661, 2507
  - and anxiety in late life, 4219
  - child/adolescent response to, 3990–3991
  - clinical presentation, 610–615
  - cognitive behavioral therapy and, 4360
  - delirium and, 4262
  - dementia and, 1252, 4409*t*
  - diagnosis, 615
  - epidemiologic studies, 894–895
  - epidemiology of, 610
  - health care workers’ reactions to, 2385–2386, 2385*t*
  - impact on health care workers, 2686–2687
  - management, 615
  - mucormycosis and, 615–616
  - neuropsychiatric sequelae of, 4206
  - older adult abuse and, 4437
  - pathophysiology of, 610

  - pediatric psychiatric services and, 3940
  - post-traumatic stress disorder and, 1874
  - social media use in youth and, 4097
  - technology-assisted psychiatric rehabilitation, 1592
- CPA. *See* Canadian Psychiatric Association (CPA)
- CPSs. *See* Complex partial seizures (CPSs)
- Crack, 4722. *See also* Cocaine
- Craig, Oscar, 4645
- Cranial electrical stimulation (CES), 2646, 3473–3474
  - clinical studies, 3474
  - definition, 3473
  - definition of, 3464*t*
  - dosing, 3474
  - history, 3473
  - mechanism of action, 3474
  - patient selection, 3474
  - side effects, 3474
  - in treatment algorithms, 3474
- Cranial nerve disorders, 657–658
- Craving, 448
- C-reactive protein (CRP), 183–184
- Creatinine, 3064
- Creativity, 4722
- CREB pathway, 1316
- Cretinism, 4722
- Creutzfeldt–Jakob disease (CJD), 4176
  - characteristics of subtypes, 629–631
  - defined, 4722
  - differential diagnosis, 642*t*
  - familial, 629
  - iatrogenic, 637
  - initial diagnostic tests for, 642*t*

    - sporadic (nonfamilial), 629, 637
    - universal precautions, 644
    - variant, 637
- CRH. *See* Corticotropin-releasing hormone (CRH)
- Crick, Francis, 236
- Cri-du-chat syndrome, 3640*t*, 4722
- Criminalization, **4571–4585**. *See also* Correctional psychiatry
    - causes of, 4572–4574
        - deinstitutionalization, 4572–4573
        - lack of adequate community support systems, 4573
        - restrictive civil commitment criteria, 4572–4573
        - role of police, 4573–4574
    - characteristics of persons with serious mental illness who are incarcerated, 4575–4577
    - definition of, 4572
    - mental health services after arrest, 4577–4578
    - mental illness in jails and prisons, 4571
    - mental illness/violence, 4574–4575
    - of mentally ill, 4627
    - treatment, 4579
        - after arrest, 4577–4578
        - case management and, 4582
        - criminal justice system and, 4580
        - decarceration of jails and prisons, 4583–4584
        - diversity, awareness, and sensitivity, 4583
        - goals, 4581
        - inpatient care, 4583
        - liaison between staff and the court, 4581
        - outpatient, 4579–4580
        - philosophy, 4581
        - pitfalls, 4583
        - to reverse/prevent, 4579–4585

    - strategies, 4584–4585
    - structure of, 4581–4582
    - therapeutic living arrangements, 4582–4583
    - violence management, 4582
    - working with family, 4583
- Criminal responsibility, 4722
- Crisis, 4723
- Crisis intervention, 4723
- Crisis intervention team (CIT), 1608
- Criteria validity, 1037
- Criterion validity, 880, 1092–1093
- Critical illnesses
  - adult survivors, 2427, 2428, 2429
  - families of survivors, 2428, 2429
  - pediatric survivors, 2427–2428, 2429
  - psychiatric morbidity after, **2427–2429**
  - stress of, 2427
  - survivors of
    - cognitive dysfunction in, 2428, 2429
    - interventions for psychiatric disorders in, 2428–2429
    - physical dysfunction in, 2428, 2429–2430
- Critical parent, 963
- Critical period, 4723
- Critical periods, 3498
- Critical ratio, 4723
- Crohn disease, 2294
- Cronbach’s coefficient alpha, 1037, 1092
- Cross, William, Jr., 848
- Cross-cultural psychiatry, 4723
- Cross-frequency coupling (CFC), 311
- Crossing over, 236, 237*f*
- Crosslinking, 247

- Crossover designs, 911
- Cross-tolerance tolerance, 1298
- Crow, T. J., 1457
- Cryptococcal meningitis, 607
- Cryptographia, 4723
- Cryptolalia, 4723
- Crystallized intelligence, 4164
- Crystallized knowledge, 4168
- CSF. *See* Cerebrospinal fluid (CSF)
- Csikszentmihalyi, Mihaly, 789, 2887
- CSTC. *See* Cortico-striatal-thalamo-cortical (CSTC) circuit
- CT scans. *See* Computed tomography (CT)
- Cullen, William, 4684
- Cult, 4723
- Cults, 2020
- Cultural anthropology. *See* Anthropology
- Cultural assessments
  - *DSM-IV* Outline for Cultural Formulation, 2571–2572
  - historical overview, 2569–2572
  - key domains in, 2572–2574
  - in psychiatry, **2569–2579**
- Cultural betrayal trauma, 1935–1936
- Cultural competency (CC), **1136–1150**
  - activities of, 1143–1150
    - clinical, 1144–1145
    - organizational, 1147, 1148*t*
    - program, 1147–1150, 1149*t*
  - in care delivery, 1139–1142
  - clinician self-assessment of, 1145
  - cultural activation of consumers, 1142, 1142*t*
  - curricula for clinicians, 1145
  - definitions of, 1137–1138

  - disparities and, 1138, 1140–1141*t*
  - diversity and, 1138
  - domains, 1146*f*
  - importance of, 1138–1139
  - legal requirements, 1142–1143
  - overview, 1136–1137
  - regulations, 1142–1143
  - standards, 1142–1143
  - structural competency, 1138
- Cultural deprivation, 4723
- Cultural dimensions
  - cultural values, 847–848
  - Hofstede, 846–847
  - individualism-collectivism, 847
- Cultural explanations, 852
- Cultural formulation, 851–853
- Cultural Formulation Interview (CFI), 853, 2574–2579, 2575–2576*t*
  - core, 2575–2577
  - future directions, 2579
  - informant, 2578
  - research on, 2579
  - supplementary modules, 2578
- Cultural group, 1137–1138
- Cultural identity, 851, 2572
- Cultural idioms of distress, 852, 2551, 2551*t*
- Culturally competent assessment, 850–858
  - conducting, 853–854
  - contextualized approach to, 853–854, 854*t*
  - cultural formulation, 851–853
  - integrated, 855*t*
  - interpreter, 850–851
  - social determinants of, 839

- Cultural mores, 2056
- Cultural psychiatry, 2552, 2569, 4723
- Cultural sensitivity, 2529–2530
- Cultural syndromes, 852, 2551, 2551*t*
- Cultural values, 847–848
- Culture
  - autism spectrum disorders (ASDs) and, 3714
  - characteristics of, 834
  - child psychiatry and, 4013–4014
  - definition of, 2574–2575
  - everyday functioning and, 4181
  - mental health and, 846–849
- Culture-bound syndromes, 1469, 2550
- Culture shock, 4723
- Culture specific syndrome, 4723
- Cummings, Jeffrey, 1023
- Cumulative distribution function, 896
- Cunnilingus, 4723
- Cupping, 3978
- Curcumin, 2639
- Curran, Desmond, 4254
- Current, 69
- Current (point and 1-month) prevalence, 874
- *Current Concepts of Positive Mental Health* (Jahoda), 2888–2889
- Current source density (CSD), 359
- Cushing, Harvey, 4660
- Cushing syndrome, 715, 2338–2340, 2365, 4723
- Custody, 4723
- Custody disputes, 4005–4006
- Cutaneous sensory disorders, 2431
- Cyberbullying, 1128, 3950, 4096–4097
- Cybernetics, 4723

- Cybersex addiction, 4723
- Cyclazocine, 4723
- Cyclic adenosine monophosphate (cAMP), 147–148, 151*f*, 1316, 3004–3006
- Cyclic guanosine monophosphate (cGMP), 125, 126*f*, 149, 149*f*
- Cycloid psychosis, 1628*t*
- Cycloplegia, 4723
- Cyclothymia, 1649–1650, 4723
- Cyclothymic disorder, 1651, 1692
  - defined, 4723
  - diagnosis of, 4790*t*
  - DSM-5 criteria for, 1681–1682*t*
  - ICD-11 criteria for, 1681–1682*t*
- Cylert. *See* Pemoline (Cylert)
- Cymbalta. *See* Duloxetine
- Cynophobia, 4723
- CYP2C19, 287
- CYP2D6, 287–288
  - autophenocopying, 288–289
  - phenocopying, 288–289
- CYP3A4, 287
- CYP3A4 isoenzyme system, 2284
- Cyproheptadine (Periactin), 3348
  - for tardive dyskinesia, 3096
- Cyproterone acetate (CPA; Androcur), 2098
- Cystic fibrosis, 2321–2323
  - child/adolescent response to, 3989
  - diagnosis of, 2322
  - in emerging adulthood/transition, 2322
  - fertility issues in, 2322
  - lifestyle interventions in in, 2323
  - manifestation of, 2321–2322

  - physiology of, 2321–2322
  - psychiatric comorbidities, 2322
  - psychological interventions in, 2323
  - psychosocial considerations, 2322
  - self-management, 2322
  - treatment of, 2322
- Cytoarchitecture, 5, 6*f*, 21*f*
- Cytokines, 174, 175*t*, 180*t*
  - antagonists, 187–188
  - brain link, 2419
  - influences on CNS and behavior, 2418–2419
  - inhibition, 184–185
  - innate and acquired immunity synchronization, 176
  - sleep, immunity, and, 2418
- Cytomegalovirus, 607, 619
- Cytomel. *See* Liothyronine (Cytomel)

D

- D_2 receptors, occupancy, 1545–1549
- Daclizumab, neurotoxicity of, 2465*t*
- Da Costa syndrome, 2500, 4723
- DailyStrength.org, 2700
- Daitch, Carolyn, 2923
- Dalai Lama, 2512
- Dalenberg, Constance, 1968
- Damadian, Raymond, 327
- Damasio, Antonio, 420, 726
- Dance therapy, 2642
- *Dando la luz el spirito* (giving light to the spirit), 2672
- Danger-associated molecular patterns (DAMPs), 172
- DAP. *See* Draw-A-Person (DAP) Test
- Darkness phobia, behavioral approach tasks for, 3888*t*
- Darwin, Erasmus, 1230
- Data breach, 1122
- Data breaches, 1131
- Data science, 4016–4017
- *Daubert v. Merrell Dow Pharmaceuticals, Inc*, 2505
- Davanloo, Habib, 2769–2772
- David, Anthony, 1933
- Davies, David Lewis, 4659
- Davis, John, 1656
- Davis, Kenneth, 1656
- Day hospital, 4723
- Day residue, 4723
- Dayton, H. B., 3358
- DBS. *See* Deep brain stimulation (DBS)
- DCD. *See* Developmental coordination disorder (DCD)
- DDAVP. *See* Desmopressin

- DEA. *See* Drug Enforcement Administration (DEA)
- Deafferentation pain, 2480*t*
- Deaf persons, psychotherapy for, 2978–2979
- Death/dying. *See also* Dying patients; End-of-life care
  - attended, 2400
  - awareness, 2383–2384
  - child’s cognitive understanding of, 2404, 2404t
  - desire for, 2410*t*
  - dignity laws, 2410*t*
  - physician aid in, 2409–2410
  - timely, 2383
  - trajectories, 2394
  - unattended, 2400
  - untimely, 2383
- Death instinct, 4723
- *Death with Dignity* laws, 2410
- Decathexis, 4723
- Decision making, sequential sampling and, 323
- Decision-making capacity, 4417–4418, 4424–4426
- Decision tree, 4723
- Decision trees, 907
- Declarative memory, 302, 730, 730*t*, 767–772
- De Clérambault syndrome, 1636, 4723
- Decoding, 4723
- Decompensation, 4723
- Decomposition, 4723
- Dedicated coding hypothesis, 305
- Deep brain stimulation (DBS), 81, 3484–3485
  - defined, 4723
  - definition of, 3464*t*
  - in depression, 35
  - major depression, 3485

  - for older adults, 4339–4340
- Deep sleep, 200
- Default mode network (DMN), 22, 462, 803, 820–821, 1755–1757, 1756*f*, 2904
- Defense mechanisms, 792, 4723
  - classification of, 920–921*t*, 2748*t*
  - conflicts and, 2748–2750
- Defensive behaviors, 297–299
  - anxiety-like behavior, 298
  - fear-like behavior, 297–298
  - panic-like behavior, 297
  - social anxiety-like behavior, 298–299
- Defensive emotion, 4723
- Defensiveness, 2493
- Deficit, 4723
- Dehospitalization, 4507
- Dehydroepiandrosterone (DHEA), **163**, 1879, 2641
  - for midlife depression, 3428–3429
- Deinstitutionalization, 3630–3631, 4507, 4572–4573, 4665, 4723
- Déjà entendu, 1182, 4723
- Déjà pensé, 1182
- Déjà pensè, 4723
- Déjà vu, 996*t*, 1182, 4723
- Dejerine, Jules-Joseph, 4650
- Delay, Jean, 1658, 3007, 3055–3056, 4650, 4690
- Delayed ejaculation, 2043, 4723, 4815–4816*t*
- Delayed recall, 739–740
- Delayed rectifier, 73*f*, 74
- Delayed sleep phase circadian rhythm disorder, 4723
- 22q11 Deletion syndrome, 3640*t*
- Delirious mania, 1676
- Delirium, 471, **4257–4264**

- agitation and, 3576
- alcohol withdrawal and, 1232, 1303
- anticholinergics intoxication and, 1237
- antipsychotics for, 4308
- benzodiazepines, 3248, 3250
- burn injuries and, 2468
- in cancer patients, 2350–2353
  - assessment and management, 2351–2352
  - causes of, 2351*t*
  - clinical features, 2351
  - diagnosis, 2351
  - epidemiology of, 2351
  - etiology of, 2351
  - treatment of, 2352–2353
- cardiac disease and, 2292
- in children, 1234–1235
- clinical features, 4257–4258
- comparative nosology, 1231, 4255
- coronary artery bypass graft and, 1232
- course and prognosis, 1235–1236, 4259–4260
- COVID-19 and, 4262
- defined, 4723
- definition of, 1229–1230, 4254
- dementia and, 1234, 1246–1247
- dementia *versus*, 4259*t*
- diagnosis, 1232–1233, 4257–4258, 4832–4833*t*
- differential diagnosis, 1233–1234, 4259, 4259*t*
- dopamine receptor antagonists for, 3045
- DSM-5-TR criteria for, 4254
- EEGs in, 377
- electroencephalogy, 1233
- end-of-life, 1231, 1237–1238

- epidemiology of, 4255–4256
- etiology of, 1231–1232, 4255*t*, 4256–4257
- future directions, 4262
- hip replacement and, 1232
- history, 1230, 4254
- in HIV, 589–590
- ICD-10 criteria for, 1230
- ICU, 1231–1232
- inhalant-induced, 1378
- laboratory examination, 1233, 4258–4259, 4258*t*
- in neuro-Covid, 189
- neuroimaging, 1233
- in older adult cancer patients, 2379
- opiate withdrawal and, 1232
- pathology of, 1233, 4258–4259
- postoperative, 1232
- prevention, 1236
- prevention of, 4260
- psychotic disorders in later life *versus*, 4270
- rating scales in, 1233
- risk factors, 4255*t*
- substance intoxication and, 1232, 1237
- substance withdrawal and, 1237–1238
- substance withdrawal-induced, 1232
- subtypes of, 2396*t*
- in terminally ill children, 2406
- in terminally ill patients, 1234, 2396–2397
  - causes and treatment of, 2397*t*
  - diagnosis, 2397–2398
  - management of, 2398
  - prevalence of, 2397
- treatment of, 1236–1238, 4260–4261, 4260*t*

    - nonpharmacologic, 1236–1238
    - pharmacologic, 1236–1237
      - antipsychotics, 1236–1237
      - benzodiazepines, 1237
      - electroconvulsive therapy and, 1237
      - melatonin, 1237
      - ramelteon, 1237
  - treatment of specific etiologies, 1237–1238
- *Delirium: Acute Confusional States* (Lipowski), 4254
- Delirium tremens (DTs)
  - barbiturates for, 3232
  - defined, 4723–4724
- Delis-Kaplan Executive Function System (D-KEFS), 1033, 4170, 4554
- Delivery System Reform Incentive Payment (DSRIP), 4484
- Delta sleep, 200, 4724
- Delta waves, 4724
- Delusional (paranoid) disorder, 4724
- Delusional atmosphere/mood, 1170
- Delusional disorder, **1634–1638**, 4656
  - antipsychotics for, 4307
  - comparative nosology, 1634
  - course and prognosis, 1638
  - diagnostic and clinical features, 1635–1636, 4781*t*
  - differential diagnosis, 1636–1638
  - dopamine receptor antagonists for, 3044
  - DSM-5 criteria for, 1635*t*
  - epidemiology of, 1634
  - etiology of, 1634–1635
  - history, 1634
  - psychotic disorders in later life *versus*, 4269–4270
  - types

    - erotomanic, 1636
    - grandiose, 1636
    - jealous, 1636
    - mixed, 1636
    - persecutory, 1635–1636
    - somatic, 1636
    - unspecified, 1636
- Delusional memory, 1170
- Delusional parasitosis, 2432–2434
  - clinical features, 2432
  - diagnosis of, 2433
  - pathogenesis of, 2433–2434
  - treatment of, 2434
- Delusional percept, 1170
- Delusional projection, 2748*t*
- Delusional separateness, 1974
- Delusional symptoms in partner of individual with delusional disorder, 1647
- Delusional thought, 996*t*
- Delusion of control, 4724
- Delusion of doubles, 1170
- Delusion of grandeur (grandiose delusion), 4724
- Delusion of jealousy (delusion of infidelity), 4724
- Delusion of negation, 4724
- Delusion of persecution, 4724
- Delusion of poverty, 4724
- Delusion of reference, 4724
- Delusions, 1169–1170
  - bereavement depression and, 1695*t*
  - bizarre, 1457
  - characteristics of, 1169*t*
  - cognition and, 1457

  - content of, 1456
  - defined, 4724
  - definition of, 1455
  - formation, 1677
  - of reference, 1456
  - in schizophrenia, 1455–1457
  - types of, 1170*t*
- Delvecchio Good, Mary-Jo, 2553
- Dement, William, 4654
- Dementia, 471, 480, 1182, **1238–1272**. *See also* Antidementia drugs; Pseudodementia
  - abuse and, 1249
  - alcohol use disorder and, 1247
  - Alzheimer disease, **1252–1259**
  - anxiety and, 4223–4224
  - behavioral and psychological symptoms of, 4314–4315
  - benzodiazepines, 3250–3251
  - biomarkers for diagnosis, 4136
  - caregivers, 1249, **4389–4392**
  - clinical features, 1241–1242
  - cognitive domain evaluation, 1243–1244, 1244*t*
  - cognitive impairment in, 1242–1243
  - comparative nosology, 1239–1240
  - corticobasal degeneration, 1263
  - course and prognosis of, 1247–1248
  - COVID-19 and, 1252, 4409*t*
  - CSF studies, 1246
  - defined, 4724
  - definition of, 1238–1239
  - delirium and, 1234, 1246–1247
  - delirium *versus*, 4259*t*
  - depression and, 1251, 4240

- depression *versus*, 1027
- detecting, 1027
- diabetes and, 2329
- diagnosis, 4829–4832*t*
- differential diagnosis, 1246–1247
- dopamine receptor antagonists for, 3045
- driving and, 1249
- DSM-5 criteria for, 1239–1240, 1242*t*
- EEGs in, 377
- end-of-life issues, 1249
- epidemiology of, 1240
- etiology of, 1241, 1241*t*
- evaluation for, 1243
- everyday functioning and, 4191–4192
- factitious disorder and, 1247
- fMRI of, 343
- frontotemporal, 564–565, 1259–1262
- functional impairment in, 1243
- health disparities in, 1252
- history, 1239, 4657–4658
- hypersexuality in, 4456*t*
- laboratory examination, 1245–1246, 1246*t*
- later-life degenerative, 544
- with Lewy bodies, 565, 1265–1269
- limbic predominant age-related TDP-43 encephalopathy, 1272
- in long-term care, 4408–4412
- management of
  - nonpharmacologic, 1248–1249
  - pharmacologic, 1249–1250
- medical conditions and, 1247
- medications and, 1247
- mild neurocognitive disorder and, 1247

  - molecular neuroimaging, 1246
  - mood disorders and, 1699
  - MRI applications to, 339–340
  - MRS in, 348
  - neglect and, 1249
  - neurodegenerative disorders and, 1247
  - neuropathology, 1246
  - neuropsychiatric manifestations, 1244–1245
    - treatment of, 1250–1251
  - neuropsychological testing, 1246
  - in older adult cancer patients, 2379–2380
  - in older adults, 4223–4224
  - in older people, 4127
  - in Parkinson disease, 1269–1271
  - with Parkinsonian signs, 564–565
  - persisting, 1305
  - prion diseases, 1271–1272
  - progressive supranuclear palsy, 1263–1265
  - with psychosis, 4307–4308
  - psychosis and, 1250
  - risk and protective factors, 1240–1241
  - schizophrenia and, 1247
  - screening, 1243
  - serious mental illness and, 1234
  - severity measures, 1244
  - sexual dysfunction and, 4450–4451
  - sleep disturbance in, 1251
  - structural neuroimaging, 1245–1246
  - vascular, 487
- Dementia associated with alcoholism, 4724
- Dementia of the Alzheimer type, 4724

- Dementia praecox, 1449, 1468, 1476, 4263, 4657, 4724. *See also* Schizophrenia
- Dementia Rating Scale-Second Edition (DRS-2), 4168
- Dementia-related psychosis, pimavanserin for, 3217*t*
- Dementia syndrome of depression, 4230
- Dementia with Lewy bodies, 1265–1269, 4175
  - biomarkers, 1266–1267
  - clinical features, 1266
  - comparative nosology, 1265
  - course and prognosis, 1268
  - definition of, 1265
  - diagnosis, 1266
  - differential diagnosis, 1267–1268
  - epidemiology of, 1265
  - etiology of, 1266
  - history, 1265
  - management of
    - nonpharmacologic, 1268
    - pharmacologic, 1268–1269
  - neuropathology, 1268
- Demisexuals, 2060
- Demography, 4724
- Demophobia, 4724
- Demyelinating disorders
  - classification of, 577*t*
  - neuropsychiatric aspects of, 567–583
- Dendrites, 3, 4724
- Dendritic spines, 17, 18*f*
- Dendy, Walter Cooper, 4685
- Dengue, 616–617
- Denial, 920*t*, 4724
- Deniker, Pierre, 3007, 3055–3056

- Dennett, Daniel, 419, 423
- Denny-Brow, Derek, 486
- Dentate gyrus, 35
  - hippocampal, 66–67, 66*f*
- Deontology, 4616*t*
- Deoxyribonucleic acid. *See* DNA
- Depacon. *See* Valproic acid (valproate, Depakene, Depakote)
- Depade. *See* Naltrexone
- Depakene. *See* Valproic acid (valproate, Depakene, Depakote)
- Dependence on therapy, 4724
- Dependency, 4724
- Dependent personality disorder, 2250, 4724, 4837*t*
- Dependent variable, 4724
- Depersonalization, 996*t*, 1179–1180, 4724
  - CNS response and, 1954
  - HPA axis function, 1953–1954
  - hypnosis and, 2917
  - neurobiology of, 1953–1955
  - neurochemical systems and, 1954–1955
- Depersonalization/derealization disorder, 1942–1943, **2012–2018**
  - cognitive theories, 2014
  - course and prognosis, 2016
  - cross-cultural aspects of, 2015
  - diagnosis and clinical features, 2014–2015, 4805*t*
  - differential diagnosis, 2015–2016
  - discrete behavioral states theory and, 2012–2013
  - epidemiology of, 1950
  - intellectual disabilities and, 3647–3648
  - life-threatening experiences and, 1950
  - psychodynamic model, 2013
  - symptoms, 1950
  - traumatic stress and, 2013–2014

    - treatment of, 2016–2018
        - pharmacotherapy, 2017–2018
        - psychotherapy, 2016–2017
- Depolarization, 69
- Depolarization abnormalities, 2290
- Depressed mood, 1671–1672
- Depression, 475, 1184–1185, 4240–4241. *See also* Beck Depression Inventory (BDI); Bipolar depression; Center for Epidemiologic Studies Depression Scale Revised (CESD-R-20); Hamilton Rating Scale for Depression (HRSD); Major depressive disorder (unipolar depression); Postpartum depression (PPD); Unipolar depression
    - amino acid neurotransmitters and, 92
    - anaclitic, 3802
    - animal models of, 1759
    - antipsychotics for, 4308
    - associated with medical illness and cognitive impairment, 4299
    - atypical, 1695
    - bipolar disorder-associated, 3930, 4300
    - in cancer patients, 4229
    - carbamazepine for, 3288–3289
    - cardiac disease and, 2291–2292
    - chronic, 1687
    - chronic pain, 2486
    - cognitive behavioral therapy for, 4371
    - cognitive theory of, 1706
    - contemporary models of, 1653–1658
        - aggression-turned-inward model, 1653–1654
        - biogenic amine imbalance, 1655–1657
        - cognitive model, 1654–1655
        - depression and reinforcement, 1655
        - learned helplessness model, 1655

  - loss of self-esteem and depression, 1654
  - neurophysiologic approaches, 1657–1658
  - object loss and depression, 1654
- coronary artery disease and, 2280–2281
- defined, 4724
- dementia and, 1251, 4240
- dementia *versus*, 1027
- diabetes and, 2326–2331
- differential diagnosis, 2721*t*
- disability and, 4230–4231
- dynamics of, 1705
- Endicott substitution criteria for, 2395*t*
- epigenetics of, 281
- epilepsy and, 522
- everyday functioning and, 4192
- family therapy for, 3911*t*, 3918–3919
- folate (vitamin B9) and, 2652–2653
- functional gastrointestinal disorders and, 2298–2299, 2299*t*
- group psychotherapy for, 3903
- heart failure and, 2283
- HIV in children/adolescents and, 676
- HIV infection and, 589–594
- hypnosis and, 2930
- inflammation and, 182–183
- interpersonal theory of, 1707–1708
- in later life, 4296–4301
  - patient management, 4299
  - pharmacological treatment of, 4296–4299
- long-term, 780
- malingering and, 2500–2501
- in medically ill patients, 4228–4229
- monoamines and, 1766

- naltrexone for, 3364
- neurosteroids and, 138
- obesity and, 2308
- in older adult cancer patients, 2378
- in older adults, 4127–4128, 4427–4428
- omega-3 fatty acids and, 2653
- Parkinson disease and, 559, 562
- in patients with dementia, 4229
- in pediatric epilepsy, 679
- pharmacologic treatment of, 4296–4297
- postpartum (*See* Postpartum depression (PPD))
- poststroke
  - assessment, 492–493
  - etiology, 490–491
- in preterm adolescents, 673
- as psychobiologic affective reaction types, 1653
- psychotic, 1687, 4230
- rating scales, 3543–3544
- with reversible dementia, 4230
- in schizophrenia, 1462–1463
- serotonin receptor availability in, 402–403
- in sexually abused children, 3971
- short-term dynamic psychotherapy for, 2779
- spirituality and, 2679–2681
- stress and, 2418
- stroke and, 487–488
- in terminally ill children, 2406
- in terminally ill patients, 2395–2396, 2396*t*
- treatment of
  - anti-inflammatory therapies, 184–185
  - monoamine enhancers, 3031, 3031*f*
- treatment-resistant, 4299–4300

  - typical courses of, 1683*f*
  - venlafaxine for, 3146
  - vitamin D and, 2651
- Depression-executive dysfunction, 4232
- Depressive behaviors, 298–299
- Depressive disorder due to another medical condition, 508–512
  - clinical features of, 510
  - course and prognosis, 510–511
  - diagnosis, 510
  - differential diagnosis, 510
  - epidemiology of, 509–510
  - etiology of, 510
  - laboratory investigations, 510
  - treatment of, 511–512
- Depressive disorders, 1678–1698, 3246. *See also* Depression; Major depressive disorder (unipolar depression)
  - alcohol-induced, 1303–1304
  - alcohol use disorder and, 1306
  - bupropion for, 3164–3165
  - in cancer patients, 2354–2358
    - diagnosis, 2355
    - prevalence of, 2355
    - risk factors, 2355, 2355*t*
    - suicide assessment and management, 2357–2358
    - treatment of
      - electroconvulsive therapy, 2357
      - pharmacologic, 2356–2357
      - psychotherapy, 2357
  - changes in DSM, 1205–1206
  - children and adolescents, 3437, 3815–3821
    - clinical features, 3817
    - course and prognosis, 3818

  - defined, 3815
  - diagnosis, 3817
  - differential diagnosis, 3818
  - epidemiology of, 3816
  - etiology of, 3816–3817
    - environmental, 3816–3817
    - genetic, 3816
    - neurobiology, 3816
  - history, 3815–3816
  - laboratory examinations, 3817–3818
  - pharmacologic treatment, 3818–3820
  - psychotherapy for, 3820–3821
  - treatment-resistant depression, 3821
- chronic demoralization, 1689
- cognition in, 737–738
- diagnosis of, 4791–4794*t*
- dopamine receptor antagonists for, 3044
- dysthymic disorder, 1687–1689
- hypnosis and, 2930
- in long-term care, 4412–4413
- major depressive disorder, 1678–1684
- minor depressive disorder, 1689
- neurasthenia, 1689–1690
- neuroimaging, 3606–3608
  - amygdala, 3607–3608
  - hippocampus, 3607–3608
  - pituitary gland, 3607–3608
  - striatum, 3606–3607
- perception in, 737–738
- phencyclidine- and other hallucinogen-induced, 1363–1364
- postpsychotic depressive disorder of schizophrenia, 1690
- preventive interventions, 4044–4045

  - rating scales for, 1104–1105
  - reactive depression, 1689
  - recurrent brief depressive disorder, 1689
  - sensation iu, 737–738
  - suicide, 2708
  - after TBI, 546–548
  - traumatic brain injury and, 683
- Depressive neurosis, 4724
- Depressive position, 821–822
- Depth psychology, 4724
- Derailment, 1462, 4724
- Derealization, 996*t*, 1179–1180
  - defined, 4724
  - neurochemical systems and, 1954–1955
- Dereism, 4724
- Dereistic thinking, 4724
- Dermatitis artefacta, 2438
  - clinical features, 2438
  - psychiatric comorbidities of, 2438
  - treatment of, 2438
- Derrida, Jacques, 969–970
- *Desarollo* (spiritual development), 2672
- Descriptive psychiatry, 4724. *See also* Dynamic psychiatry
- Descriptive statistics, 4724
- Descriptive studies, 878–879
- Desensitization. *See* Systematic desensitization
- Designer drug, 4724
- Desipramine (Norpramin, Pertofrane), 3152
  - pharmacogenetics of, 294*t*
  - for stereotypic movement disorders, 3813
- Desmethyl clomipramine, clomipramine and, 290
- Desmethylclozapine, clozapine and, 290

- Desmopressin (DDAVP), for enuresis, 3791–3793, 3794–3796
- Despair, 940
- Despert, Louise, 3837
- Despine, Antoine, 1927
- *Despojo* (exorcism), 2672
- DES-Taxon Scale (DES-T), 1937*t*
- Desvenlafaxine (Pristiq)
  - for depression in later life, 4297
  - side effects, 3819
- Desyrel. *See* Trazodone
- Detachment, 4724
- Determinism, 4725
- Detoxification, 1308–1309, 4725
- Detransitioning, 2105
- Detumescence, 4725
- Deuterium metabolic imaging, 350–351
- Deutetrabenazine (Austedo)
  - for stereotypic movement disorders, 3813
  - for tardive dyskinesia, 3096
- Deutsch, Felix, 2278
- Deutsch, Helene, 4689
- Devaluation, 2748*t*, 4725
- Developmental arithmetic disorder, 4725. *See also* Gerstmann–Sträussler–Scheinker disease
- Developmental articulation disorder, 4725
- Developmental coordination disorder (DCD), **3673–3680**
  - ADHD and, 3674
  - assessment instruments for, 3675–3676
  - clinical features, 3675–3678
  - clinical features of, 3676
  - comparative nosology, 3673–3674
  - course and prognosis, 3678–3679

  - defined, 4725
  - definition of, 3673
  - diagnosis of, 3675–3678, 3675*t*, 4778–4779*t*
  - differential diagnosis, 3677–3678
  - DSM-5-TR criteria for, 3675*t*
  - epidemiology of, 3674
  - etiology of, 3674
  - ICD-10 criteria for, 3677*t*
  - manifestations of, 3673*t*
  - motor signs and symptoms, 3677*t*
  - neurologic disorders and, 3677–3678
  - premature birth and, 3674–3675
  - psychiatric comorbidities of, 3674–3675
  - treatment of, 3679
    - physical and occupational therapy, 3680
    - process-oriented, 3679–3680
    - task-oriented, 3679
- Developmental Coordination Disorder Questionnaire (DCD-Q), 3675
- Developmental delays, agitation and, 3577
- Developmental disabilities, 483–484
  - in correctional facilities, 4637
  - racism and, 4110–4111
- Developmental disorders
  - defined, 4725
  - treatment of, 3738
- Developmental expressive language disorder, 4725
- Developmental expressive writing disorder, 4725
- Developmentally informed hypnotherapy, 2922
- Developmental model, 922–923
  - autistic phase, 922
  - separation-individuation, 922

  - symbiotic phase, 922
- Developmental reading disorder, 4725
- Developmental receptive language disorder, 4725
- Developmental stuttering, 467*t*
- Developmental trait personality models, 2238
- Developmental word blindness, 4725
- Development and Well Being Assessment (DAWBA), 3540
- Deviation, 4725
- Dewey, John, 789
- Dexamethasone Suppression Test (DST), 4725
- Dexedrine. *See* Dextroamphetamine (Dexedrine)
- Dexmedetomidine (Precedex), 3236
  - pain management in burn patients, 2470
- Dextroamphetamine (Dexedrine)
  - for cancer-related fatigue, 2359
  - for depression, in cancer patients, 2357
  - for geriatric depression, 4239
- Dextromethorphan
  - dextrorphan and, 290
  - self-injurious behaviors, 3651
- Dextrorphan, 290
- DHA. *See* Docosahexaenoic acid (DHA)
- DHEA. *See* Dehydroepiandrosterone (DHEA)
- *Dhyana*, 2670
- Diabetes mellitus, **2325–2333**, 3058, 3073, 3081
  - clinical features, 2342
  - clozapine and, 3070
  - depression and, 2326–2331
  - diagnosis, 2342
  - insulin resistance, 2342
  - laboratory examination, 2342
  - obesity epidemic and, 2326

  - patient–provider relationship in, 2327
  - psychiatric comorbidity, 2326–2328
  - psychiatric illness and mortality in patients with, 2329, 2342
  - psychiatric symptoms in patients with, 2327
  - schizophrenia comorbidity, 1594
  - sexual dysfunction and, 4449
  - treatment of, 2329–2331, 2342–2343
  - type 1, 2327–2328, 2341
  - type 2, 2331–2333, 2341–2342
- Diabetic neuropathy, 3273, 3275
- Diagnosis, 1199
- *Diagnostic and Statistical Manual of Mental Disorders* (DSM), 881–882, 4725
  - adjustment disorders, 2201–2202
  - advantages of DSM system, 1202
  - alcohol use disorders, 1296, 1302*t*
  - alternatives to traditional classification, 1208–1211
  - approving/disapproving changes or new diagnoses in, 1204
  - areas of concern, 1206–1208
  - binge eating disorder, 2307*t*
  - bipolar I disorder, 1681–1682*t*
  - bipolar II disorder, 1683–1684*t*
  - caffeine-related disorders, 1324*t*, 1325
  - cannabis use disorder, 1340
  - child maltreatment, 3970
  - in clinical practice, 1211
  - conduct disorder, 3752*t*
  - conflict of interest, 1208
  - controversies, 1224–1225
  - Cultural Formulation Interview in, 2574–2579, 2575–2576*t*
  - cyclothymic disorder, 1685*t*
  - delirium, 4254

- delusional disorder, 1635*t*
- dementia, 1239–1240, 1242*t*
- development of, 1203–1204
- dissociative disorders, 1939–1943
- encopresis, 3796, 3797*t*
- enuresis, 3789, 3790*t*
- evolution of, 1201*f*
- factitious disorder, 1906–1907, 1907*t*
- factitious disorder imposed on self, 1908–1909
- feeding disorder of infancy or early childhood, 3761
- female sexual interest/arousal disorder, 2041*t*
- field trials, 1204
- gambling disorder, 1858–1859
- generalized anxiety disorder, 3856*t*
- global issues, 1208
- harmonization with ICD, 1225–1227
- ICD *versus*, 1216–1217, 1217*t*
- increasing complexity of DSM process, 1203
- inhalant-related disorders, 1365–1366
- inhalant use disorder, 1376
- insomnia disorder, 2165–2166
- intermittent explosive disorder, 2194*t*
- Internet gaming disorder, 4090*t*
- key changes in, 1204–1206
- kleptomania, 2199*t*
- language disorder, 3681
- major depressive disorder, 1679–1680*t*
- major or mild neurocognitive disorder, 1285–1286*t*
- male erectile disorder, 2041*t*
- male hypoactive sexual desire disorder, 2040*t*
- malingering, 2494
- manual, 1204–1205

  - narcolepsy, 2175
  - obsessive-compulsive disorders, 1840–1841, 1843*t*
  - obsessive-compulsive personality disorder, pediatric, 3838
  - opioid-related disorders, 1397
  - oppositional defiant disorder, 3750*t*
  - Outline for Cultural Formulation, 2572–2574
  - paraphilic disorders, 2092
  - personality disorders, 1192–1193, 4272
  - post-traumatic stress disorder, 1870–1871, 1870*t*
  - psychosis, 1613–1614
  - research on diagnostic categories, 1212
  - restless legs syndrome, 2184*t*
  - schizoaffective disorder, 1639*t*
  - schizophrenia, 1450–1451
  - schizophreniform disorder, 1632*t*
  - separation anxiety, 3851
  - sexual addiction, 2123–2124
  - sexual dysfunction, 2038, 3430
  - shortcomings of DSM system, 1202–1203
  - sleep disorders, 4084–4087
  - social (pragmatic) communication disorder, 3689*t*
  - social anxiety disorder, 3854*t*
  - somatic symptom and related disorders, 1887, 1891*t*
  - somatic symptom disorder, 1896
  - specific learning disorder, 3659
  - speech sound disorder, 3695
  - stereotypic movement disorders, 3810–3811
  - stimulant-related disorders, 1312, 1313, 1313*t*
  - stuttering, 3699
  - tic disorders, 3775
  - updates on manual and diagnoses, 1211–1212
- Diagnostic classification, 1199

- Diagnostic criteria, 1199
- Diagnostic Interview for Anxiety, Mood and Obsessive Compulsive and Related Neuropsychiatric Disorders—Child and Adolescent Version (DIAMOND-KID), 3540, 3541
- Diagnostic Interview for Children and Adolescents (DICA), 884
- Diagnostic Interview for Personality Disorders (DIPD), 2254
- Diagnostic interviews, 3537–3541
  - interview-based interviews, 3538–3539, 3540–3541
  - limitations, 3547
  - respondent-based interviews, 3538–3540
- Diagnostic Interview Schedule (DIS), 883
- Diagnostic Interview Schedule for Children (DISC), 884, 1115–1120, 3540
- Diagnostic systems, biologically based, 42
- Dialectical behavior therapy (DBT), **2841–2852**, 4725
  - for adolescents, 3953–3954
  - and biosocial theory of BPD, 2842–2843
  - for borderline personality disorder, 2716
  - clinical issues, 2847–2849
    - agreements, 2848
    - assumptions, 2848, 2848*t*
    - levels of treatment and treatment targets, 2848–2849
  - definition of, 2841
  - dialectics, 2842
  - for dissociative identity disorder, 1999
  - efficacy and effectiveness of, 2990
  - ethical issues, 2850
  - history, 2841–2842
  - research and evaluation, 2850–2851
  - for sexual addiction, 2136–2137
  - skills modules
    - distress tolerance, 2844

    - emotion regulation, 2844–2845
    - interpersonal effectiveness, 2845
    - mindfulness, 2844
  - for suicide, 2716
  - techniques, 2843–2847
    - behavioral analysis, 2846
    - consultation team, 2847, 2847*t*
    - dialectical strategies, 2847, 2847*t*
    - diary cards, 2845–2846, 2845*f*
    - individual therapy, 2845
    - skill enhancement, 2846
    - skills training, 2844
    - telephone consultation, 2847
    - validation, 2846–2847
  - theoretical issues, 2842–2843
- Diaphragm (contraception method), 2593*t*
- Diarrhea, 2392–2393
- Diary cards, 2845–2846, 2845*f*
- Diathesis-stress paradigm, 4725
- Diazepam (Valium), 1429*t*, 3007
  - history, 4663
  - oxazepam and, 290
  - pharmacogenetics of, 294*t*
  - pharmacokinetics of, sex differences in, 3417
- Dibenzoxazepines, 3038
- DICA. *See* Diagnostic Interview for Children and Adolescents (DICA)
- Dick, Henry, 2949
- Diencephalon, 9, 9*f*
- Diethylstilbestrol (DES), 2030
- Differences of sex development (DSD), 4101–4102
- Differential diagnosis, 4725

- Differential psychotherapeutics, 2740–2741
- Differential reinforcement, 4725
- Differential relaxation, 2424
- Differentiation, 65, 4725
- Diffuse-projection nuclei, 23
- Diffusion magnetic resonance imaging, schizophrenia, 1517–1520
- Diffusion-tensor imaging (DTI), 337, 340
  - neuroimaging and, 538
- Diffusion-weighted imaging (DWI), 336–337
- DiGeorge syndrome, 3640*t*
- DiGeorge/velocardiofacial syndrome, 3587*t*
- Digital amplifiers, 360
- Digit Symbol (a.k.a., “Coding”) test, 4170
- Dignity therapy, 2976–2978
  - clinical issues, 2978
  - ethical issues in, 2978
  - history, 2976–2977
  - research and evaluation, 2978
  - techniques, 2977
  - theoretical issues, 2977
- Dihydrotestosterone (DHT), 2581
- Diltiazem (Cardizem)
  - chemistry of, 3259, 3260*t*
  - drug interactions, 3293
  - drug studies, 3261
  - pharmacodynamics, 3260–3261
  - precautions and adverse reactions, 3263
  - psychiatric side effects, 2293
  - therapeutic indications, 3262–3263
- Dimenhydrinate (Dramamine), 107
- Dimensional diagnosis, 1199
- Dimensions, 1090

- 2,5-Dimethoxy-4-methylamphetamine (DOM), 4725
- Dimethyltryptamine (DMT), 388, 1354, 2903
  - chemical structure of, 3386*f*
  - defined, 4725
  - for treatment-resistant depression, 2906
- Dimidjian, Susan, 2814
- Diminished libido, 4725
- Dim light melatonin onset (DMSO), 195
- Diphenhydramine (Benadryl)
  - for agitation, 3576*t*
  - for stereotypic movement disorders, 3813
- Diphenhydramine hydrochloride, 3325–3326
- Diphenylbutylpiperidine (Orap), 3034, 3035
- Diphenylhydramine, 3347
- Diprivan. *See* Propofol (Diprivan)
- Dipsomania, 4725
- Direct and Indirect Aggression Scale (DIAS), 3546
- Direct detection, 346
- Directed acyclic graphs (DAGs), 904
- Directive hypnotherapy, 2922
- Direct pathway, 31
- Direct questioning, in child interview, 3525–3526
- Disability, 4725
  - assessment instruments, 885
  - case identification, 882, 883*t*
  - depression and, 4230–4231
  - evaluating for, 4422
  - older adult applicants for, 4422
- Disability-adjusted life years (DALYs), 893
- Disabled 1 (Dab1) protein, 63
- Disaster psychiatry, **2655–2667**
  - combat and operational stress, 2658–2659

  - cultural issues in, 2666
  - emerging threats, 2665–2666
  - environments of care, 2656–2657
  - ethical challenges, 2666
  - etiology of, 2659–2661
  - historical background, 2656
  - management and care delivery, 2661–2663
    - general principles, 2661–2662, 2661*t*
    - psychological first aid, 2662*t*
    - therapeutic interventions, 2662–2663
  - medically unexplained physical symptoms, 2658–2659
  - military mental health care in peacetime, 2663–2665
  - mitigating or perpetuating factors, 2660
  - overview, 2655–2656
  - precipitating factors, 2660
  - predisposing factors, 2660
  - protective factors, 2660–2661
  - psychological and behavioral responses, 2658*f*, 2658*t*
  - psychological and behavioral syndromes, 2657–2658
  - PTSD and, 2657–2658
  - symptoms *versus* functioning, 2657
  - terrorism and, 2657–2658
  - traumatic stress responses and, 2657–2659
  - war and, 2656, 2657–2658
- DISC. *See* Diagnostic Interview Schedule for Children (DISC)
- DISC (Diagnostic Interview Schedule for Children), 1115–1120
- Disclosure, barriers to, 1157
- Disconnection syndrome, 4725
- Discontinuation syndrome, 1420, 3148
  - defined, 4725
  - nefazodone, 3209
  - vilazodone, 3178

  - vortioxetine, 3182–3183
- Discordance, 4725
- Discourse, 733
- Discrete behavioral states model, 1959–1960, 1971–1972
  - depersonalization/derealization disorder and, 2012–2013
  - dissociative amnesia and, 2001
- Discriminant analysis, 4725
- Discriminant validity, 1037, 1093
- Discrimination
  - child development and, 4107
  - defined, 4725
- Discriminative stimulus, 4725
- Disease, 1199
- Disidentification, 4601
- Disillusioning process, 957
- Disinhibited social engagement disorder (DSED), **3801–3807**
  - attachment, 3801–3802
  - clinical features, 3804
  - course and prognosis, 3805–3806
  - defined, 4726
  - diagnosis, 3804, 4800–4801*t*
  - differential diagnosis, 3805
  - environmental factors, 3802–3803
  - epidemiology of, 3804
  - etiology of, 3803–3804
  - nosology, 3803
  - treatment of, 3806–3807
- Disinhibition
  - benzodiazepines and, 3251
  - defined, 4725–4726
  - multiple sclerosis and, 580
- Disintegration, 4726

- Disk-over-water (DOW) apparatus, 211
- Disordered mood and emotion, 475
- Disorder of sexual development (DSD), 2582
- Disorder of written expression, 4726
- Disorder-related (domain specific) rating scales, 3543–3547
  - assessing adjustment and impairment, 3543–3546-2
  - for assessing externalizing symptoms—attention deficit/hyperactivity disorder with and without related symptoms, 3543–3546
  - for assessing internalizing symptomatology— depression, 3543–3545
  - for externalizing symptomatology, 3543–3546
  - for other symptom areas, 3543–3546
- Disorder(s)
  - definition of, 1198–1199
  - new, 1205
  - structure and grouping of, 1205
- Disorders/diseases, 1205
- Disorders of excessive somnolence, 4726. *See also* Somnolence
- Disorders of initiating and maintaining sleep, 4726
- Disorders of sexual development (DSDs), 2101
- Disorganization, 1460–1462
  - motor symptoms, 1461
  - thought disorder, 1461–1462
- Disorganized attachment, 2927
- Disorganized–disoriented attachment, 3500
- Disorganized schizophrenia, 4726
- Disorientation, 1176–1177, 4726
- Displacement, 921*t*, 932*t*, 2748*t*, 4726
- Disposition, 4726
- Disruption of family by separation or divorce, 2546–2547
  - clinical features, 2547

  - definition of, 2546
  - diagnosis, 2547
  - etiology of, 2546–2547
- Disruptive Behavior- Diagnostic Observational Schedule (DB-DOS), 3556
- Disruptive behavior disorders
  - borderline intellectual functioning and, 2533
  - in children and adolescents, **3747–3760**
  - conduct disorder, **3752–3760**
    - biologic factors, 3755–3756
    - clinical features, 3757–3758
    - comorbid conditions, 3758
    - comparative nosology, 3754
    - course and prognosis, 3758–3759
    - definition of, 3752–3753
    - diagnosis, 3757–3758
    - differential diagnosis, 3758
    - DSM-5-TR criteria for, 3752*t*
    - epidemiology, 3754
    - etiology, 3754–3755
    - family factors, 3757
    - heritable factors, 3755–3756
    - history, 3753–3754
    - neighborhood and cultural factors, 3757
    - neuroanatomic factors, 3756
    - neuroendocrinologic factors, 3756
    - neurotransmitter factors, 3756
    - peer factors, 3757
    - physiologic factors, 3756
    - prenatal and perinatal factors, 3756–3757
    - psychological factors, 3757
    - risk factors, 3755*t*

    - sociologic factors, 3757
    - subtypes and specifiers, 3753
    - treatment, 3759–3760
  - defined, 4726
  - family therapy for, 3911*t*
  - intellectual disabilities and, 3652
  - oppositional defiant disorder, **3747–3751**, **3761–3765**
    - biologic factors, 3749
    - clinical features, 3749
    - comorbid conditions, 3751
    - course and prognosis, 3751
    - definition of, 3747–3748
    - diagnosis, 3749
    - differential diagnosis, 3750–3751
    - DSM-5-TR criteria for, 3750*t*
    - epidemiology, 3748
    - etiology, 3748–3749
    - history, 3748
    - laboratory examination, 3750
    - pathology, 3750
    - psychological factors, 3749
    - sociologic factors, 3749
    - treatment, 3751
  - temperament and, 3625
- Disruptive mood dysregulation disorder (DMDD), 3747
  - defined, 4726
  - definition of, 3489
  - diagnosis of, 4791*t*
- Disseminated encephalomyelitis:, 582
- Dissimulation, 2492
- Dissocial personality disorder, 4836*t*
- Dissociation, 921*t*, 1178–1180

- borderline personality disorder and, 1974
- complex/dissociative PTSD and, 1962
- conceptualizations of, 1960*t*
- controversies, 1934–1935
- defined, 4726
- definition of, 1939
- depersonalization, 1179–1180
- derealization, 1179–1180
- in diverse cultural settings, 1962–1963
- etiology of, 1959
- fantasy model, 1968–1969, 1968*t*
- *FKBP5* and, 1952
- 4-D model of trauma-related, 1966
- Ganser syndrome, 1179
- genetics and, 1952
- hypnosis and, 2916, 2917
- hypnotic model, 1966
- iatrogenic model, 1967–1968
- identity alteration in children associated with armed groups, 2020–2021
- international and cross-cultural research on, 1932–1933
- media and, 1935
- in military survival school subjects, 1963
- mind control and, 2020
- neurobiology of, 1952–1959
- nonsuicidal self-injury and, 1961
- psychoanalytic models, 1966
- sleep model, 1967, 1969
- sociocognitive model, 1967
- structural, 1967
- suicide attempts and, 1961
- taxon model of, 1960*t*

  - theories of, 1959
  - as transdiagnostic process, 1960–1961
  - trauma model, 1968–1969, 1968*t*
  - as trauma response, 1962
- Dissociation syndrome, 4726
- Dissociative amnesia, 782, 2001–2012
  - adverse childhood experiences and, 1948–1949, 1948*t*
  - Childhood autobiographical memory disturbances and, 1948–1949, 1948*t*
  - classic presentation, 2002–2003
  - clinical course and prognosis, 2007
  - clinical presentation of, 2001–2002
  - conceptualizing, 2001
  - defined, 4726
  - definition of, 2001–2012
  - diagnosis and clinical features, 2002–2003, 4804*t*
  - differential diagnosis, 2005–2008
    - cognitive disorders, 2006–2007
    - nonpathologic amnesia, 2005–2006
    - posttraumatic amnesia, 2007
    - seizure disorders, 2007
    - substance-related amnesia, 2007
    - transient global amnesia, 2007
    - wandering behavior, 2007
  - discrete behavioral states theory and, 2001
  - dissociative fugue specifier, 2004
  - dissociative fugue subtype of, 1949, 2012
  - disturbed retrieval, 1956–1957
  - encoding and, 1955–1956
  - epidemiology of, 1947–1949
  - malingering and, 2501
  - neurobiology of, 1956–1957

  - nonclassic, 2010
  - nonclassic presentation, 2004
  - pathology and laboratory examination, 2005
  - peritraumatic, 1955–1956
  - traumatic experiences and, 1949*t*
  - treatment of, 2009–2012
    - cognitive therapy, 2011
    - free recall, 2010
    - group psychotherapy, 2012
    - hypnosis, 2011
    - phase-oriented, 2009
    - phasic, 2009–2010
    - recall of trauma material, 2011
    - somatic therapies, 2012
    - transference interpretations, 2010
  - types of, 1942*t*
  - validity of, 1944–1945
  - for violent/homicidal behavior, 2004
- Dissociative anesthetics, 1348, 1351
- Dissociative disorder(s), **1927–2025**. *See also* Dissociative amnesia
  - borderline personality disorder and, 1975
  - clinical populations, 1950–1952
    - emergency department, 1951
    - inpatient studies, 1951
    - non-Western/secular settings, 1951
    - outpatient studies, 1951
  - comparative nosology, 1936–1939
  - controversies, 1934–1935
  - defined, 4726
  - definition of, 1906
  - depersonalization/derealization disorder, **2012–2018**

- differential diagnosis, 4804–4805*t*
  - acute stress disorder, 2007
  - dissociative disorders, 2007
  - factitious amnesia, 2007
  - malingering amnesia, 2007
  - other psychiatric disorders, 2007
  - posttraumatic stress disorder, 2007
- dissociative amnesia and, 2007
- dissociative disorders and, 1952
- DSM-5 criteria for, 1939–1943
- epidemiology of, 1945
- etiology of, 1959, 2012–2014
  - betrayal trauma, 1964
  - culture, sociocognitive phenomena, 1963–1964
  - developmental mediation of psychopathology, 1965–1968
  - developmental model, 1965
  - developmental precursors and substrates, 1965
  - discrete behavioral states model, 1959
  - type D attachment, 1964
- general population studies, 1945–1950
- history, 1927–1933
  - in 19th and early 20th centuries, 1929
  - depersonalization/derealization disorder, 1931
  - DSM-III and beyond, 1932–1933
  - international and cross-cultural research, 1932–1933
  - Janet and Freud, 1930
  - later 20th century, 1931
  - trauma and La Salpêtrière patients, 1928, 1929*f*
  - World War II and, 1931
- institutional and cultural betrayal trauma, 1935–1936
- modern society and media, 1935
- neurobiology of, 1952–1959

  - other specified, 1943
  - other specified dissociative disorders, 2018
    - acute dissociative reactions to stressful events, 2021–2024
    - atypical presentations, 2018–2019
    - coercive persuasion, 2019–2021
    - dissociative trance disorder, 2024–2025
  - pathologic dissociation, 1945–1946
  - screening and interviews, 1937–1938*t*
  - in sexually abused children, 3971
  - substance use disorder and, 1951–1952
  - validity of, 1943–1945
- Dissociative Disorders Interview Schedule (DDIS), 1938*t*
- Dissociative Experiences Scale (DES), 1932, 1937*t*. *See also* Adolescent DES (A-DES)
- Dissociative fugue, 4732. *See also* Dissociative amnesia
- Dissociative identity disorder, 2559–2560
  - antisocial personality features, 1975
  - autohypnotic symptoms in, 1978*t*
  - borderline personality disorder and, 1974–1975
  - childhood trauma and, 1973–1974
  - comparison of DSM-III, IIIR, ICD-10, DSM-IV-TR, DSM-5 Diagnostic Criteria, 1941*t*
  - course and prognosis, 1984
  - dangerousness to self and others, 1970–1971, 1971*t*
  - defined, 4726
  - diagnosis and clinical features, 4804*t*
    - amnesia and memory symptoms in, 1976–1977, 1976*t*
    - child and adolescent presentations, 1982
    - differential diagnosis, 1982*t*, 1983
    - dissociative alterations in identity, 1980–1981
    - dissociative hallucinations, 1979–1980
    - dissociative thought disorder, 1980

  - inter-identity amnesia, 1978–1979
  - other associated symptoms, 1981–1982, 1982*t*
  - passive influence and interference symptoms, 1980
  - process symptoms, 1977*t*, 1979
  - spontaneous autohypnotic symptoms, 1981
- discrete behavioral states and, 1971–1972
- DSM-5 criteria for, 1939–1940
- etiology of, 1970
- factitious, imitative, and malingered, 1984–1985
- Four-Factor Theory of, 1973
- general population studies, 1946–1947
- ICD-11 criteria for, 1940–1942
- non-possession-form, 1978*t*, 1980
- partial, 1940–1942
- pathological possession form of, 1947
- pathology and laboratory examination, child and adolescent presentations, 1983
- possession-form, 1972–1973, 1978*t*, 1980–1981
- self-states and, 1972
- in sexually abused children, 3971
- shame and, 1973–1974
- treatment, 1985–2000
  - adjunctive, 1999–2000
  - affect regulation/shame, 1993
  - boundaries, 1991–1992
  - cost efficacy studies, 1986–1987
  - countertransference, 1993
  - dialectical behavior therapy, 1999
  - direct therapeutic engagement with self-states, 1990
  - electroconvulsive therapy, 1997
  - expressive therapy, 2000
  - eye movement desensitization and reprocessing, 1999–2000

    - family therapy, 1999
    - group therapy, 1999
    - hypnosis, 1995
    - informed consent, 1991
    - inpatient treatment, 1998
    - ketamine, 1997
    - longitudinal studies, 1987–1988
    - medications for mood disorder symptoms, 1997
    - medications for PTSD symptoms, 1996–1997
    - occupational therapy, 2000
    - outcome studies, 1986–1987
    - partial hospitalization, 1998–1999
    - phasic treatment, 1989–1990, 1989*t*
    - psychopharmacological interventions, 1995–1996, 1996*t*
    - repetitive transcranial magnetic stimulation, 1997
    - safety agreements, 1991
    - safety and stability interventions, 1992–2000
    - safety issues, 1990–1991
    - self-help groups, 2000
    - TOPDD Network Study, 1987–1988, 1988*t*
    - transference, 1993
    - trauma-focused cognitive-behavioral interventions, 1992–1993
    - trauma-informed inpatient care, 1998
    - traumatic memories, 1993–1994, 1994*t*
  - validity of, 1943–1944
- Dissociative psychosis, 2023
- Dissociatives, 1348, 1351
- Dissociative state, 2916
- Dissociative thought disorder, 1980
- Dissociative trance disorder (DTD), 2024–2025, 2559
  - defined, 4726
  - definition of, 2024–2025

  - diagnosis and clinical features, 2025
  - pathology and laboratory examination, 2025
  - treatment of, 2025
- Dissolvables, 1385
- Distance-dependent movement artifact, 410
- Distortion, 4726
- Distractibility, 4726
- Distractible speech, 4726
- Distress
  - cultural concepts of, 864, **2550–2569**, 2572–2573
    - clinical applications, 2566
      - collaborative care, 2567–2568
      - mental health services and interventions, 2566–2568
    - examples of, 2555–2565
      - *ataque de nervios* (attack of nerves), 2555–2559
      - possession syndrome, 2559–2562
      - *shenjing shuairuo*, 2562–2565
    - historical background, 2552–2554
    - psychiatric classification and, 2554–2555
    - research into, 2565–2566
    - terminology, 2550–2552
  - cultural idioms of, 2551
  - tolerance, 2844
- Distributive analysis and synthesis, 4726
- Distributive justice, 4022
- Disulfiram (Antabuse), 95, **3367–3371**
  - absorption, 3368
  - alcohol use disorder, 3370–3371
  - for alcohol use disorders, 1310
  - blood concentrations, 3369
  - chemical structure, 3368*f*
  - chemistry, 3368

  - for cocaine use disorder, 1323
  - defined, 4726
  - distribution, 3368
  - dosage and administration, 3372
  - drug interactions, 3372
  - drug studies, 3368–3370
  - elimination, 3368
  - history, 3366
  - mechanism of action, 3368
  - metabolism, 3368
  - organ/system effects of, 3370
  - pharmacodynamics, 3368
  - pharmacokinetics, 3368
  - precautions and adverse reactions, 3371–3372
  - therapeutic indications, 3370–3371
- Divalproex (Depakote)
  - for acute bipolar depression, 1733–1734
  - for acute mania, 1729
  - chemistry, 3260*t*
- Divergent system, 6, 7*f*
- Diversity, 1025–1026
- Divorce, family therapy for, 3922
- Dix, Dorothea Lynde, 4512, 4645, 4726
- Dizygotic twins, 253, 4726
- D-KEFS. *See* Delis-Kaplan Executive Function System (D-KEFS)
- DLPFC. *See* Dorsolateral prefrontal cortex (DLPFC)
- DMT. *See* Dimethyltryptamine (DMT)
- DNA
  - chromosomes and, 235–236
  - complementarity, 245
  - methylation, 275–277, 276*f*, 277*f*
  - repetitive, 240–242

- DNA microarray, 3587*t*
- Docosahexaenoic acid (DHA), 4322
  - for bipolar disorder, 188
  - for schizophrenia, 186
- Doctor–patient relationship, 4726
- Dollard, John, 971
- Dolophine. *See* Methadone (Dolophine)
- Domain-specific scales. *See* Disorder-related (domain specific) rating scales
- Domestic abuse or violence, 2624–2625. *See also* Intimate partner violence (IPV)
  - definition of, 2624
  - perpetrators, 2625
  - prevalence of, 2624
  - psychiatric emergencies and, 2730–2731
  - risk factors, 2625
- Dominance, 4726
- Donepezil (Aricept), 3335*t*, 3336–3338, 4317
  - for Alzheimer disease, 4315*t*
  - chemistry, 3337
  - clinical drug studies, 3337–3338
  - contraindications, 4318
  - dosage and administration, 3338, 4317
  - drug interactions, 3338
  - molecular structure of, 3337*f*
  - organ/system effects of, 3338
  - pharmacodynamics, 3337
  - pharmacogenetics of, 294*t*
  - pharmacokinetics, 3337
  - pharmacokinetics and pharmacodynamics of, 4315*t*, 4317
  - precautions and adverse reactions, 3338
  - therapeutic indications, 3338

- Don Juanism, 2049
- Dopamine, 94–100, 180
  - agonists, 1542
  - alcohol and, 1298
  - anatomy and functions, 94
  - anxiety disorders and, 1799–1800
  - appetite and, 227, 229*f*
  - D1 receptor availability, 400
  - D_2/D_3 receptor availability, 400
  - defined, 4726
  - depletion, 1542
  - extrastriatal dopamine D_2-like receptors, 1542
  - intrasynaptic, 1542
  - life cycle, 95–98
  - mesocorticolimbic dopamine pathway, 446
  - molecular imaging of, 1538–1542
  - neuroimaging studies, 1538–1542, 1539*t*, 1540*f*
  - plasma membrane transporter, 96–97
  - radiotracers, 1537*t*, 1539*f*
  - release, 401
  - schizophrenia and, 402
  - schizophrenia pathology, 1505–1507
  - stimulant-related disorders and, 1315
  - storage capacity, 401
  - substance dependence and, 402
  - subtypes, 1541
  - synthesis, 95–96, 95*f*, 401
- Dopamine agonists, for acute bipolar depression, 1734–1735
- Dopamine agonist withdrawal syndrome (DAWS), 560
- Dopamine antagonists
  - absorption, 3038

- adverse reactions, 3046–3047
  - acute extrapyramidal syndromes, 3046–3047
  - behavioral side effects, 3048–3049
  - blood, 3050
  - cardiovascular effects, 3049
  - chronic extrapyramidal syndromes, 3047–3048
  - endocrine effects, 3050
  - eye effects, 3050
  - gastrointestinal tract, 3049–3050
  - lactation, 3050
  - liver, 3050
  - neurologic side effects, 3046
  - orthostatic (postural) hypotension, 3049
  - overdose, 3050–3051
  - pregnancy, 3050
  - seizure threshold, 3050
  - sexual function, 3050
  - skin effects, 3050
  - sudden unexplained death, 3049
  - temperature regulation, 3050
- benzamides, 3035
- butyrophenones, 3035
- chemistry, 3034–3035
- clinical drug studies
  - design, 3041
  - interpretation, 3041–3042
  - maintenance therapy, 3042
- dependence, 3051
- dibenzoxazepines, 3038
- dihydroindoles, 3038
- diphenylbutylpiperidines, 3035
- distribution, 3038

  - dopamine receptor antagonists for, **3033–3054**
  - drug interactions, 3051
  - effects of
    - on motor systems, 3042–3043
    - neuroendocrine, 3043
  - elimination, 3038–3039
  - history, 3033–3034
  - indications, 3043–3046
  - laboratory interferences, 3051
  - maintenance treatment, 3053–3054
  - mechanism of action, 3039–3041
  - metabolism, 3038, 3051
  - molecular structures of, 3036–3037*f*
  - pharmacodynamics, 3039
  - pharmacokinetics, 3038, 3038*t*
  - pharmacologic actions, 3038–3039
  - phenothiazines, 3035
  - side-effect interactions, 3051
  - thioxanthenes, 3035–3038
  - tolerance, 3051
  - treatment of psychosis, 3051–3053
    - adjunctive medications, 3053
    - blood levels, 3053
    - choice of agent, 3052
    - clinical activity, 3053
    - dosage, 3052–3053
    - initiation of treatment, 3051
    - as-needed medication, 3053
    - route of administration, 3052
  - withdrawal symptoms, 3051
- Dopamine-β-hydroxylase (DBH), 95
- Dopamine blockers, 3027–3030

  - effects on dopamine system, 3040–3041
- Dopamine dysregulation syndrome, 560
- Dopamine enhancers, 3030–3031, **3106–3129**
  - dopamine receptor agonists, **3106–3114**
  - psychostimulants, **3115–3129**
- Dopamine hypothesis, 3039–3040, 4726
- Dopamine-noradrenaline reuptake inhibitors, **3163–3167**
- Dopamine receptor agonists, **3106–3114**
- Dopamine-receptor agonists, 4726
- Dopamine receptor agonists
  - adverse reactions, 3112–3113, 3113*t*
  - for bipolar depression, 3110
  - chemistry of, 3107–3108
  - dosage and administration, 3114
  - drug interactions, 3113–3114
  - effects on organs and systems, 3111–3112
    - cardiovascular system, 3112
    - central nervous system, 3111–3112
    - endocrine effects, 3112
    - immunologic effects, 3112
  - ergot class of, 3107
  - for extrapyramidal symptoms, 3111
  - for major depressive disorder, 3109–3110
  - for mood and fatigue related to PD, 3110
  - for periodic limb movements of sleep, 3110–3111
  - pharmacodynamics, 3109
  - pharmacokinetics, 3108–3109
  - precautions, 3112–3113
  - for restless leg syndrome, 3110–3111
  - for schizophrenia-associated cognitive dysfunction, 3111
  - therapeutic implications, 3112
- Dopamine-receptor antagonists (DRAs), 4726

- Dopamine receptors
  - function of, 98–99
  - occupancy, 1551–1553
  - subtypes of, 98–99
- Dopaminergic agents, for sexual dysfunctions, 2054
- Dopaminergic neurotransmission, 1549–1551
- Dopaminergic system, 402, 1762*f*, 1764–1766, 1765*f*, 1799–1800
  - neuroimaging, 400–402
- Dopamine-serotonin antagonists (DSAs), **3054–3089**
  - asenapine, 3084–3087
  - clozapine, 3064–3071
  - comparison of, 3055–3056
  - history, 3054–3055
  - iloperidone, 3082–3084
  - lumateperone, 3087–3089
  - lurasidone, 3062–3064
  - molecular structures of, 3056*f*
  - olanzapine, 3056–3058
  - paliperidone, 3071–3078
  - quetiapine, 3078–3082
  - risperidone, 3071–3078
  - ziprasidone, 3060–3062
- Dopamine-serotonin partial agonists–antagonists (DSPAs), **3100–3106**
  - aripiprazole, 3100–3103
  - brexpiprazole, 3103–3105
  - cariprazine, 3105–3106
- Dopamine serotonin receptor antagonists, 3931
- Doral. *See* Quazepam (Doral)
- Doriden. *See* Glutethimide (Doriden)
- Dormen, Lesley, 2878
- Dorsal FPN attention network, 803

- Dorsal root ganglia (DRG), 215
  - disorders of, 655
  - neuronopathies of, 654–655
- Dorsolateral prefrontal cortex (DLPFC), 9, 459, 2782, 2918
- Dostinex. *See* Cabergoline (Dostinex, Caberlin)
- Dostoyevsky, Feodor, 524
- Double bind, 4726
- Double-blind study, 4726
- Double-blind test, 4690
- Doublecortin (DCX), 66–67
- Double depression, 1688, 4727. *See also* Dysthymia
- Double dissociators, 1966
- Double helix, 236
- Double personality, 4727
- Double simultaneous stimulation, 1033
- Doubt, 938
- Douglas, Virginia, 736
- Downfield, 330
- Down syndrome (DS), 3640*t*, 3643
  - defined, 4727
  - genome-wide dysregulation in, 51*f*
  - polygenic disruptions in, 50
- Doxepin (Sinequan), 3152
  - pharmacogenetics of, 295*t*
- Doximity, 2702
- Draw-A-Person (DAP) Test, 3548, 4727
- Drawing, 746
- Drawing test, 4727
- DRD1-type receptors, 2518
- DRD2-type receptors, 2518
- Dread, 4727
- Dream anxiety disorder, 4727

- Dream Market, 1127
- Dream(s)
  - consciousness and, 212
  - defined, 4727
  - induced, 2917
  - sleep stages and, 211–212
  - theories of, 212–213
  - waking consciousness and, 211
- Dreamy state, 4727
- Dreikurs, Rudolph, 2955
- Drevets, Wayne, 1658
- Drift diffusion model (DDM), 323
- Drive, 2747
  - defined, 4727
  - neuroimaging and, 2782
- Driving, 4180
  - benzodiazepines and, 3251
- Dronabinol, 1346
- Drowsiness, 4727
- Drug abuse. *See also* Substance abuse; Substance use disorders (SUDs)
  - in adolescents, 3514
  - amantadine for, 3333
  - defined, 4727
  - laboratory testing, 1058–1059, 1059*t*
  - parasomnia and, 2185
- Drug abuse screen test for Adolescents (DASTA), 3546
- Drug-addicted parkinsonism. *See* Pseudoparkinsonism
- Drug and Alcohol Services Information System (DASIS), 1314
- Drug dependence, 4727
- Drug development, **3017–3026**
  - abuse liability assessment, 3024–3025

- biomarkers, 3017, 3020
- chemistry/product quality, 3025
- clinical hold, 3024
- clinical pharmacology review, 3025
- clinical trials, 3017–3018
  - diagnosis and patient selection, 3017–3018
  - dose, 3017
  - effect size, 3017
  - expectation bias, 3018
  - health care professional interaction, 3018
  - regression to the mean, 3018
  - repeated measure, 3019
  - study designs, 3017
  - subjective outcome measures, 3018
- collaborative networks, 3019
- efficacy review, 3024
- future directions, 3019–3021
- globalization, 3019–3020
- IND review process, 3023–3024
- marketing application submission and review, 3024
- nonclinical studies, 3023
- outcome measures, 3021
- patient centricity, 3020
- personalized medicine, 3020
- personalized technology, 3020
- pharmacology and toxicology, 3025
- phases of, 3023–3026
- post approval, 3025
- postmarketing safety, 3025–3026
- product labeling, 3025
- “real-world” data, 3020–3021
- regulation of, 3021–3022

    - Best Pharmaceuticals for Children Act, 3022
    - elixir sulfanilamide tragedy of 1937, 3021
    - Food, Drug, and Cosmetic Act of 1938, 3021
    - Food and Drug Administration Amendments Act of 2007, 3022
    - Food and Drug Administration Modernization Act of 199, 3022
    - Kefauver–Harris Amendments, 3022
    - Pediatric Research Equity Act, 3022
    - Prescription Drug User Fee Act, 3022
    - Public Health Security and Bioterrorism Preparedness Act of 2002, 3022
    - Pure Food and Drug Act of 1906, 3021
  - regulatory decision, 3025
  - safety review, 3024
  - scientific challenges, 3017
- Drug Enforcement Administration (DEA), 1127, 4727
- Drug holiday, 4727
- Drug-induced movement disorders, 1195–1196
  - akathisia, 1195
  - dystonic movements, 1195
  - neuroleptic malignant syndrome, 1196
  - rabbit syndrome, 1196
  - serotonin syndrome, 1196
  - tardive dyskinesia, 1195–1196
- Drug-induced parkinsonism, 4727
- Drug interaction, 4727
- Drug levels, 4727
- Drug(s). *See also* Medication(s)
  - allegiance effect, 3011
  - combinations, 3014
  - conversion experience, 3011
  - definition of, 3023
  - dosing, 3012

  - medical errors and, 1153, 1153*t*
  - past treatment response, 3011
  - potency, 3012
  - preference, 3011–3012
- Drug self-administration, 301
- Drug therapy, 4727. *See also* Medication(s)
- Drug tolerance. *See* Tolerance
- DS. *See* Down syndrome (DS)
- DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measures, 1100
- DST. *See* Dexamethasone Suppression Test (DST)
- Dual reuptake inhibitors. *See* Serotonin-norepinephrine reuptake inhibitors (SNRIs)
- Dual-sex therapy, 2050–2051, 4727
- Dual therapy, 4727
- Dubois, Paul, 2272, 4647
- DuBrul, Sascha, 2879
- Dudgard, Jaycee, 2020
- Duke University, 4653
- Duloxetine (Cymbalta), 98, 3148–3150
  - antidepressant efficacy of, 3149
  - for anxiety disorders in later life, 4304
  - for anxiety in older adults, 4224–4225
  - for depression in later life, 4297
  - pharmacokinetics of, sex differences in, 3416
  - pharmacology of, 3148–3149
  - side effects, 3819
  - tolerability of, 3149–3150
- Dummy, 4727
- Dunbar, Flanders, 2273
- Dunedin Multidisciplinary Health and Development study, 1495
- Durable power of attorney, 2401

- Durham rule, 4727
- Dutch approach, 2112
- Duty to disclose, 1156–1157
- Dyad, 4727
- Dyadic session, 4727
- Dying patients. *See also* Death/dying
  - anorexia and cachexia in, 2392
  - communication with doctor, 2386–2387
  - decision-making issues for, 2399
  - disease-specific considerations in, 2398
  - families of
    - coping, 2400–2401
    - final farewell after death, 2400
    - grief therapy for, 2399
    - involvement in decision making, 2403
    - last vigil, 2399
  - fatigue in, 2391
  - group therapies for, 2398
  - insomnia in, 2391–2392
  - pain assessment and treatment, 2391
  - psychological syndromes and symptoms, 2393–2397
  - spiritual concerns for, 2399
  - subjective experiences of, 2398
  - surrogates, 2400–2402
  - types of, 2388
- DynaCirc. *See* Isradipine (DynaCirc)
- Dynamical system theory, 802
- Dynamic interpersonal theory (DIT), 2754
- Dynamic psychiatry, 4727
- Dynamic reasoning, 4727
- Dynorphin, 447
- Dysarthria, 481t, 4727

- Dyscalculia, 4727
- Dysdiadochokinesia., 469
- Dysdiadochokinesis, 4727
- Dysesthesia, 2480*t*
- Dysesthetic vulvodynia, 2440
- Dysgeusia, 4727
- Dysgraphia, 4727
- Dyskinesia, 3780*t*, 4727
- Dyslalia, 4727
- Dyslexia, 4727
- Dyslipidemia, 1597–1598
- Dysmetria, 4727
- Dysmnesia, 4727
- Dysmorphophobia, 1175, 4727. *See also* Body dysmorphic disorder (BDD)
- Dysnarration, 475
- Dysnosognosia, 1193–1194
- Dyspareunia, 2044, 2045, 4727
- Dysphagia, 3086, 4727
- Dysphasia, 4727
- Dysphonia, 4727
- Dysphoria, 4728
- Dyspraxia, 3676
- Dysprosody, 4728
- Dyssocial behavior, 4728
- Dyssomnias, 4728
- Dysthymia (dysthymic disorder), 1649–1650, 1687–1689. *See also* Depression
  - anxious dysthymia, 1688
  - cognitive theory of depression, 1707
  - course, 1688
  - defined, 4728

  - diagnosis, 4793*t*
  - interpersonal theory of depression, 1707
  - monoamine oxidase inhibitors, 3194
  - premenstrual dysphoric disorder, 1689
  - psychoanalytic formulations of, 1704
  - social adjustment, 1688
  - variants, 1688
- Dystonia(s), 468, 575–576
  - acute, 3046, 3094–3095
  - defined, 4728
  - differential diagnosis, 3780*t*
  - forms of, 1580
- Dystonic movements, 1195
- Dystonic tic, 3774
- Dystrophinopathies, 662

E

- Early birds or larks, 2180
- Early ejaculation, 2043–2044, 4728
- Early memory, 730*t*
- Early-onset bipolar disorder, 3828–3836
  - assessment, 3830–3831
  - brain structure and function, 3829–3830
  - course, 3832–3834
  - diagnostic stability, 3833–3834
  - differential diagnosis, 3832
  - epidemiology, 3829
  - episode duration and recurrence, 3832–3833
  - etiology, 3829
  - family history, 3831
  - history, 3828–3829
  - mechanisms, 3829

  - percent time symptomatic, 3833
  - psychosocial functioning, 3833
  - relapse prevention, 3835
  - risk factors, 3829
  - suicidal thoughts and behaviors, 3833–3834
  - treatment, 3834–3836
    - pharmacotherapy, 3834–3835
    - psychosocial, 3835–3836
- Early-onset psychotic disorders, 3867–3874
  - comparative nosology, 3867–3868
  - course and prognosis, 3873
  - definition of, 3867
  - diagnosis and clinical features, 3869–3870
  - differential diagnosis, 3871–3873
  - epidemiology of, 3868–3869
  - etiology of, 3869
  - history, 3867
  - pathology and laboratory examination, 3870–3871
  - prevention of, 4045
  - symptoms in children and adolescents, 3868*t*
  - treatment of, 3873–3874
- Early-onset schizophrenia (EOS), 4254, 4264–4267
  - adaptive functioning and, 4265–4266
  - antipsychotics for, 4307
  - biologic features, 4266
  - cognition and, 4265
  - community integration and, 4265–4266
  - course and outcome of, 4264–4266
  - course and prognosis, 3873
  - definition of, 3867
  - depression and, 4263–4265
  - differential diagnosis, 3871–3873

  - dopamine receptor antagonists for, 3045
  - etiology of, 3869
  - features of, 4269*t*
  - late-onset, 4267–4268
  - medical comorbidities of, 4266
  - mortality rates, 4266
  - negative symptoms, 4265
  - organic etiologic agents of, 3873
  - pathology and laboratory examination, 3870–3871
  - positive symptoms, 4265
  - quality of life and, 4266
  - social functioning and, 4265–4266
  - symptom dimensions of outcome, 4264–4265
  - symptoms in children and adolescents, 3868*t*
  - treatment of, 3873–3874
    - community-based, 4267
    - nonpharmacologic, 4267
    - somatic, 4266–4267
- Early-onset schizophrenia spectrum disorders (EOSS), symptoms in children and adolescents, 3868*t*
- Early pregnancy loss, 2589
- Earth mother archetype, 951
- Eastern equine encephalitis (EEE), 616
- Eastern State Hospital, 4645
- Eating
  - abnormal, 300
  - in absence of hunger, 223, 223*t*
  - behavioral neuroscience of, 220
  - brief-access, 222–223
  - meal as functional unit of, 220
  - in patients with bulimia, 231
  - sham, 222–223

  - signals for meal initiation, 220–222
    - circadian rhythms, 220
    - endocrine, 220–221, 222*f*
    - hunger, 220
    - metabolic, 220
- Eating disorders (ED). *See* Feeding and eating disorders
- Eating Disorders Assessment-5 (EDA-5), 1112
- Eating Disorders Examination (EDE), 1112
- Eaton, James, 4586
- ECA. *See* Epidemiologic Catchment Area (ECA)
- *Echinopsis pachanoi*, 1350*f*
- Echoing approval, 467*t*
- Echolalia, 467*t*, 3774, 4728
- Echopraxia, 1461, 3774, 4728
- E-cigarettes, 1393
- Ecologic momentary assessment (EMA), 2423, 4186
- Ecology, 4728
- Economic disparities, 838–839
- Economo, Constantin von, 4649
- Ecopsychiatry, 4728
- Ecstasy, 4728
- ECT. *See* Electroconvulsive therapy (ECT)
- Ectomorphic, 4728
- Eczema, 2443–2444
  - clinical features, 2443
- ED. *See* Eating disorders (ED); Erectile dysfunction (ED)
- EDA-5 (Eating Disorders Assessment-5), 1112
- EDE (Eating Disorders Examination), 1112
- Edelman, Gerald, 730, 1176
- Edinburgh Royal Asylum, 4643
- EDS. *See* Excessive daytime somnolence (EDS)
- Educable, 4728

- Educational outcomes, 4026
- Educational psychology, 4728
- Educational system, **4026–4040**. *See also* Academic or educational problem
  - common terms, 4026, 4026*t*
  - disparities and, 4027
  - educational history and policies, 4027
  - educational outcomes and mental health, 4026
  - federal laws, 4027*t*
  - psychiatrist's role in, 4035–4039
    - communication through school letter, 4037
    - consulting to schools in a time of crisis, 4038–4039
    - coordinating care, 4036–4037
    - direct service model, 4037
    - school consultation, 4037–4038
  - roles of school personnel, 4028–4029
  - school resources and programming, 4029–4035
    - 504 Plan, 4031
    - individualized education program, 4031
    - multitiered systems of support, 4029–4031
    - private schools, 4035
    - special education, 4035
  - settings, 4027–4028
    - alternative programs/schools, 4028
    - charter schools, 4028
    - early education, 4028
    - homeschooling, 4028
    - private schools, 4028
    - public schools, 4028
  - social, emotional, and behavioral interventions by tier of support, 4030*t*

  - tiered systems of instructional and social, emotional, and behavioral supports, 4030*f*
- Education for All Handicapped Children Act, 3632
- Edwin Smith Surgical Papyrus, 1128
- EEG. *See* Electroencephalography (EEG)
- Effectiveness studies, 912
- Effect modification, 877
- Efferent projections, 195
- Effexor. *See* Venlafaxine
- Efficacy studies, 911–912
- Ego, 422, 917, 1702, 2749
  - Anna Freud's ego psychology, 919–922
  - defined, 4728
  - development of, 918–919
  - functions of, 918
  - Heinz Hartmann's ego psychology, 919
  - psychology of, 918–922
- Ego-alien, 4728
- Ego analysis, 4728
- *The Ego and the ID* (Freud), 917
- *The Ego and the Mechanisms of Defense* (Freud), 937
- Ego boundaries, 4728
- Egocentric, 4728
- Egocentrism, 749
- Ego-coping skill, 4728
- Ego defense, 4728
- Ego-dystonic, 4728
- Ego-dystonic homosexuality, 4728
- Ego ideal, 4728
- Egomania, 4728
- Ego model, 4728
- Ego psychologists, 2957

- Ego psychology, 4728
- Ego states, dissociative, 2020–2021
- Ego strength, 4728
- Ego-syntonic, 4728
- E/I balance, 80
- Eicosanoids, 136–137
  - adverse reactions, 137
  - chemistry, 136
  - effects on organs and systems, 136
  - precautions, 137
  - therapeutic indications, 136–137
- Eicosapentaenoic acid (EPA)
  - for bipolar disorder, 188
  - for schizophrenia, 186
- Eidetic image, 4728
- Ejaculation, 2028, 4728
- Ejaculation incompetence. *See* Inhibited male orgasm
- Ejaculatory incompetence, 4728
- Ekman, Paul, 798
- Elaboration, 4728
- Elated moods, 1185
- Elation, 1670, 4728
- Elavil. *See* Amitriptyline (Elavil)
- Elbert, Thomas, 1953
- Eldepryl. *See* Selegiline (Eldepryl)
- Elder abuse, **4432–4438**
  - assault and battery, 4433
  - court system and, 4435
  - COVID-19 and, 4437
  - definitions of, 4433
  - emotional/psychological, 4433
  - ethical analysis, 4436–4437

  - excuses and justifications in, 4436
  - factors in, 4434
  - family interventions and, 4379–4380
  - financial, 4433, 4435–4436
  - multidisciplinary teams (MDTs), 4434–4435
  - neglect, 4433
  - nosology, 4433–4434
  - profile of financial abuser, 4437–4438
  - profile of physical abuser, 4437
  - psychiatric emergencies and, 2731
  - red flags and indicators of, 4435
  - sexual assault, 4433
  - shoves *versus* nudges, 4437
  - threats *versus* offers, 4437
- Elderly persons. *See also* Geriatric psychiatry
  - benzodiazepines in, 3249–3254
  - frequency of sexual expressions in, 4441–4443
  - gender and sexual minorities, 4441–4443
  - psychotherapy of, **4341–4349**
  - retirement, 4161
  - role transitions, 4161
  - sexuality of, 4162–4163, 4438–4457
    - age-related changes in sexual function and behavior, 4440–4443, 4440*t*
    - cognitive capacity and sexual behaviors, 4454
    - frequency of sexual expressions in, 4440–4442
    - LGBTQ+, 4443–4445, 4453–4454
    - recommendations for PRC communities, 4454–4455
    - sexual behavior in long-term care, 4452–4455
    - sexual dysfunctions
      - arousal phase problems, 4448
      - assessment, 4446–4447

    - desire phase problems, 4447–4448
    - illnesses associated with, 4449–4451, 4449*t*
    - medications associated with, 4451–4452
    - orgasm phase problems, 4448
    - treatment, 4447–4451
  - sexual history, 4445–4446
- social functioning, 4162–4163
- substance abuse in, 4133, 4277–4290
  - barriers to detection, 4277*t*
  - benzodiazepines, 4286–4287
  - history, 4278
  - illicit drugs, 4288
  - nonpharmacologic therapies, 4288–4289
    - brief interventions, 4289
    - intensive outpatient programs, 4290
    - moderators and correlates of treatment, 4290
    - partial hospitalization, 4290
    - psychosocial interventions, 4289
    - twelve-step recovery programs, 4289–4290
  - opioid analgesics, 4287–4288
  - signs and symptoms, 4279*t*
  - tetrahydrocannabinol, 4285–4286
- suicide in, 2378, 2711, 4231
  - risk factors, 2711
  - treatment, 2711, 2718
- technology for, 4464–4470
  - aging in place and, 4465
  - barriers and limitations, 4476
    - age-related impairments, 4476
    - billing, 4476
    - ethics, 4476
    - information overload, 4476

      - integration with health care system, 4476
      - lack of regulation and evidence, 4476
      - literacy, 4476
      - privacy, 4476
      - usability, 4476
    - classification, 4464*t*
    - comparative nosology, 4464–4465
    - future directions, 4469–4470
      - big data, 4469–4470
      - nanotechnology, 4469
      - robotics, 4469
    - specific applications of, 4465–4471
      - caregiver applications, 4469
      - games, 4469
      - mobile health, 4465–4467
      - sensors/self-monitoring devices, 4467, 4467*t*
      - smart homes, 4468, 4468*t*
      - social media, 4469
      - tablets, 4469
      - telehealth, 4465
- Elective mutism, 4728
- Electra complex, 3504, 4728
- Electrical conduction disease, 2286–2290
  - atrial fibrillation, 2287–2288
  - cardiac implantable electronic devices for, 2290
  - depolarization abnormalities, 2290
  - ECG interpretation, 2286–2287
  - repolarization abnormalities, 2288–2290
  - sudden cardiac death, 2287
  - ventricular arrhythmia, 2287
- Electrocardiography, 1062–1063
  - axis, 2286

    - basics, 2286–2287
    - heart rate, 2286
    - intervals, 2286–2287
    - rhythms, 2287
    - segments, 2286–2287
    - terminology, 2286–2287
- Electroconvulsive therapy
    - for depression, in cancer patients, 2357
    - for schizophrenia, 1608
- Electroconvulsive therapy (ECT), 81, 1062, 1237, 3248, **3444–3463**, **4326–4340**. *See also* Convulsive therapy; Neurostimulation therapies
    - for acute bipolar depression, 1735
    - adverse effects of, 3460–3462, 4335–4338
        - cardiovascular system, 3460
        - dental, 3461
        - musculoskeletal, 3461
        - neurocognitive, 3461–3462, 3462*t*, 4336–4338
        - neurophysiologic, 4336
        - respiratory system, 3460
        - systemic, 4335–4336
    - in Africa, 3441
    - catatonia, 3450
    - clinical scenarios, 3477*t*
    - contraindications, 3460–3462
    - defined, 4728
    - definition of, 3464*t*
    - devices, 3447
    - for dissociative identity disorder, 1997
    - for early-onset schizophrenia, 4267
    - EEG profile ECT *vs*. MST, 4333*f*
    - efficacy, 4333–4335

- electrical field ECT *vs.* MST, 4333*f*
- future directions, 3462–3463
- for geriatric bipolar disorder, 4243
- for geriatric depression, 4239–4240
- historical overview, 3444–3445
- history, 4325–4326, 4660
- indications, 3449–3452, 3450*t*
- in Latin America, 3444
- major depression, 3449–3450
- mania, 3451
- mechanisms of action, 3447–3449
  - biochemical, 3448
  - genetic and epigenetic, 3449
  - neurophysiology, 3448–3449
  - neuroplasticity, 3449
- medication interactions with, 3459–3460
- milestones in, 3446*t*
- mixed states, 3451
- neurologic disorders, 3452
- for older adults, **4326–4340**
  - consent, 4329–4330, 4329*t*
  - indications for, 4326–4327, 4326*t*
  - medical evaluation for, 4327–4328, 4328*t*
  - neurocognitive assessment, 4328–4329, 4329*t*
  - for patients with comorbid medical diagnoses, 4332–4333
  - procedures in, 4330–4333
    - anesthesia, 4330
    - electrode placement sites, 4331, 4331*f*
    - treatment parameters, 4331–4332
  - recommendations, 4332–4333
- patient management, 3452–3459
  - anesthesia, 3455–3457

    - consent, 3453
    - continuation treatment, 3458–3459
    - electrode placement, 3454, 3454*f*
    - pretreatment evaluation, 3452–3453
    - seizure threshold and dosing, 3454–3455
    - stimulus parameters, 3453–3455
    - stimulus waveform, 3454
    - treatment course, 3457–3458
  - physiology, 3460–3462
  - psychiatric disorders, 3451–3452
  - rapid cycling, 3451
  - safety, 4333–4335
  - schizophrenia, 3450–3451
  - seizures induced by, 3445–3447
  - for suicide, 2713
- Electrocorticography, 369
- Electrodes, 357–358, 358*f*
- Electroencephalographic (EEG) biofeedback, 2644–2645
- Electroencephalography (EEG), **353–384**, 361, 477
  - activating procedures, 370–371
    - hyperventilation, 372
    - photic stimulation, 370, 371*f*
    - polysomnography, 371
    - sleep deprivation, 371
  - age and, 370
  - amplifiers, 360–361
  - apparatus, 357–361
    - electrodes, 357–358, 358*f*
    - montage selection, 358–360, 359*f*
  - artifacts, 363–365
    - electromyographic, 364*f*
    - electrooculographic, 363*f*

  - extraphysiologic, 364–365
  - ocular correction, 364*f*
  - physiologic, 363–364
  - sweat drift, 365*f*
- attention-deficit/hyperactivity disorder (ADHD), 377
- autism spectrum disorders (ASDs) and, 377–378
- basic principles, 354–355
- catatonia, 375–376
- characteristic features, 361–369
- clinical applications, 369–370, 384
- closed head injuries, 373
- defined, 4728
- delirium, 377
- dementia, 377
- digital recording, 361
- drug activation studies, 371–372
- endocrine disorders, 373
- error-related negativity, 383
- event-related potential, 381
- evoked potential, 381, 1062
- feature detection, 365–368
- history, 354
- homeostatic regulation of spectrum, 1070–1071
- interpretation of, 355–356
  - interpreter reliability, 356
- metabolic disorders, 373
- middle-latency components, 381
- mismatch negativity, 383
- mood disorders, 376–377
- N100 component, 381–382
- normal intrinsic frequencies, 361–362
  - alpha, 362

    - beta, 362
    - delta, 362
    - gamma, 362
    - theta, 362
  - normal *versus* abnormal, 369–370
  - obsessive-compulsive disorders, 376
  - in organic pathophysiology, 372–373
  - P50 component, 381–382
  - P300 component, 383
  - panic disorder, 376
  - pharmaco-EEG, 377
  - polysomnography, 1062
  - in psychiatry, 374–375
  - psychotropic medications and, 378
  - quantification, 365–368
  - recording, 356–361
  - research paradigms, 380–383
  - role in diagnosis, 1061–1062
  - schizophrenia, 375
  - seizure disorders, 372
  - sharp waves, 370
  - sleep patterns, 370
  - spike patterns, 370
  - structural lesions, 373
  - variants, 370
- Electroencephalography, quantitative (QEEG)
  - clinical applications, 369
  - defined, 365–366
  - methods of, 365–367
  - source analysis, 368
  - topographic analysis, 368
- Electrogenic pump, 69

- Electromyogram (EMG), 4728
- Electronic cigarettes, 1385
- Electronic health records (EHR), 1121–1122, 1129–1131, 1135–1136, 4501–4503
  - data breaches, 1131
  - interoperability, 1129–1130
  - malpractice, 1130
  - privacy, 1130–1131
  - problems in, 1129
  - ransomware and, 1131
  - security, 1130–1131
- Electronic mail (email), 2700
- Electronic media, **1121–1136**
  - clinical issues, 1131–1132
  - communication evolution, 1128
  - data capture, 1133
  - definitions, 1121
  - ethical issues in, 1132–1133
  - legislative history, 1121
  - medical practice and, 1133
  - medical records, 1129–1131
  - theoretical issues, 1133
- Electronic medical records (EMRs), 1004
  - in correctional psychiatry, 4640–4641
  - error prevention, 1154–1155
  - errors in, 1152
  - telepsychiatry, 4640–4641
- Electronic prescription, 1126–1127
  - error prevention, 1154–1155
- Electroshock therapy. *See* Electroconvulsive therapy (ECT)
- Electrosleep, 3463
- Electrostimulation, 4729

- Elevated mood, 4729
- Elimination disorders, **3789–3800**
  - defined, 4729
  - diagnosis, 4810*t*
  - encopresis, 3796–3800, 4810*t*
  - enuresis, 3789–3796, 4810*t*
  - intellectual disabilities and, 3647–3648
- Elixir sulfanilamide tragedy of 1937, 3021
- Elkes, Joel, 4650, 4661
- Elkins, David, 2675
- Ellis, Albert, 2676, 2677, 2818, 3885, 4705
- Ellis, Havelock, 2027, 2027*f*
- Ellson, Douglas, 437
- Elman network, 320
- Elopement, 4729
- Email journaling, 1132
- Emancipated minor, 4729
- Emde, Robert, 1959
- Emergency Medical Treatment and Labor Act (EMTALA), 2719–2720
- Emergency psychiatry. *See also* Psychiatric emergencies
  - aggression, management of, 2724–2726
  - agitation, management of, 2724–2726
  - alcohol intoxication, 2726–2727
  - benzodiazepine, sedative, or hypnotic intoxication, 2727
  - cannabis intoxication, 2727–2728
  - capacity, 2733
  - child abuse, 2731
  - competency, 2733
  - confidentiality, 2732–2733
  - difficult patients, 2723–2724
  - domestic abuse, 2730–2731

  - elder abuse, 2731
  - first priority, 2720
  - hallucinogen intoxication, 2728
  - high utilizers of service, 2732
  - history, 2719–2720
  - homelessness and, 2732
  - informed consent, 2733–2734
  - inhalant use, 2729
  - initial evaluation, 2722–2723
  - legal issues, 2732–2734
  - opioid intoxication, 2727–2728
  - pharmacologic management, 2724
  - phencyclidine intoxication, 2728
  - psychosocial issues, 2732
  - rape, 2731–2732
  - second priority, 2720
  - stimulant intoxication, 2726
  - substance withdrawal and, 2729–2730
    - amphetamines, 2730
    - benzodiazepines, 2729–2730
    - cocaine, 2730
    - hypnotics, 2729–2730
    - inhalants, 2730
    - opioids, 2730
    - sedatives, 2729–2730
  - suicide and, 2705–2719
  - third priority, 2720–2722
  - treatment principles, 2722–2726
- Emergency room, hypnosis in, 2931–2932
- Emetic aversion, 2817
- EMG. *See* Electromyogram (EMG)
- EMIC (Explanatory Model Interview Catalogue), 2570

- Emminghaus, Hermann, 4664, 4686
- Emotion, 726, 808, 819, 1757*f*, 4729
  - acknowledgement of, 999–1000
  - amygdala connectivity and, 413–414, 413*f*
  - in asthma, 2318–2319
  - categorical, 726
  - processing, 1811–1812
  - regulation of, 2846*t*
- Emotional aging, 4129
- Emotional brainstem, 811
- Emotional deprivation, 4729
- Emotional development, in children, 3492*t*
- Emotional disorders
  - defined, 4729
  - vulnerability to, 870
- Emotional disturbance, 4729
- Emotional dysregulation, 3519
- Emotional freedom technique, 2647
- Emotional illness, 4729
- Emotional insight, 4729
- Emotionalism, 476
- Emotional lability, 476, 4729
- Emotionally unstable personality disorder, 4836*t*
- Emotional memory, 1812
- Emotional personality (EP), 1967
- Emotional processing
  - anxiety disorders and, 1822–1824
  - brain and, 1758*f*
  - corrective information and conditions interfering with, 1823–1824
  - failures of, 1824
  - theory of, 1822–1824
- Emotional regulation, 3519

- Emotional support, 4729
- Emotions, 868–869
  - brain and, 1755–1759, 1755*f*, 1756*f*
  - etiology of, 1754–1755
  - positive *vs.* negative, 2891
- Empathic understanding, 2975
- Empathogen, 1348
- Empathy, 2959*t*
  - defined, 4729
  - psychotherapy integration and, 2736–2737
  - self-psychology, 926
- Empirism, 744
- Empty nest syndrome, 4729
- Enablers, 789
- Enactment, 2759
- Encapsulated delusion, 4724, 4729
- Encephalitis, 4729
- Encephalitis lethartica, 566
- Encephalopathy, 4729
- Encephalopathy, autoimmune or paraneoplastic, 1056–1057
- ENCODE Project, 240*f*, 250*t*
- Encopresis, 3796–3800
  - clinical features, 3798
  - comparative nosology, 3797
  - course and prognosis, 3799
  - defined, 4729
  - definition of, 3796
  - diagnosis, 3796, 3797*t*, 3798, 4810*t*
  - differential diagnosis, 3799
  - DSM-5-TR criteria for, 3797*t*
  - epidemiology of, 3797
  - etiology of, 3797–3798

  - history, 3796
  - ICD-11 criteria for, 3797*t*
  - laboratory examination, 3798–3799
  - pathology of, 3777
  - treatment of, 3799–3800
- Encounter group, 4729
- Encouragement, 999
- Endemic, 4729
- Endfeet, 4
- Endocannabinoids, 130–136
  - addiction and, 135
  - in anxiety and moods, 134–135
  - anxiety disorders and, 1805–1806
  - biosynthesis of, 131–132
  - discovery, 130
  - effects in the periphery, 136
  - effects on brain injury and pain, 135–136
  - inactivation of, 132
  - in psychosis, 135
  - retrograde transmission regulation, 133*f*, 134
  - synthetic, 136
- Endocrine and metabolic disorders, **2334–2345**
  - adrenal disorders, 2338–2341
  - defined, 4729
  - description of, 2334
  - diabetes mellitus and insulin resistance, 2341–2343
  - EEG findings, 373
  - hyperprolactinemia, 2344–2345
  - polycystic ovary syndrome, 2343–2344
  - thyroid disorders, 2334–2338
- Endocrine myopathies, 663
- Endocrine neuropathies, 658

- Endocrine system
  - characteristics of, 158
  - evaluation
    - adrenal disease, 1054–1055
    - anabolic steroid use, 1055
    - antidiuretic hormone, 1055
    - human chorionic gonadotropin, 1055
    - pancreatic function, 1056
    - parathyroid hormone, 1055
    - prolactin, 1055–1056
    - thyroid hormone, 1056
  - nalmefene effects on, 3364
  - naltrexone effects on, 3365
  - signal for meal initiation, 220–221, 222*f*
- Endocrinologic agents, 2138
- *The End of Faith* (Harris), 2683
- End-of-life care, **2383–2411**. *See also* Death/dying
  - adolescents and young adults, 2406–2407
  - anorexia and, 2392
  - cachexia and, 2392
  - caregiver responses, 2410
  - cross-cultural barriers, 2402–2403, 2402*t*
  - cultural, spiritual, and religious issues in, 2402–2403
  - decision-making issues in, 2399
  - disease-specific considerations in, 2398
  - doctor-patient communication, 2386–2387
  - ethical issues in, 2409
  - family interventions and, 4379
  - fatigue in, 2391
  - GI symptoms, 2392–2393
  - group therapies in, 2398
  - informed consent for experimental treatment, 2409

- insomnia and, 2391–2392
- legal issues, 2401–2402
- pain
  - assessment and treatment of, 2387–2391
  - incidence of, 2387
  - medications for, 2389
  - psychiatric symptoms and, 2389–2390
  - types, 2388*t*
  - undertreatment of, 2390
- pediatric, 2403
  - communication with severely ill children, 2405
  - end-of-life decisions, 2408
    - assent, 2408
    - consent, 2408
    - neonatal and infant issues, 2408
  - healthy siblings
    - parental response, 2407
    - psychological issues of, 2407–2408
    - understanding of sibling’s terminal illness, 2407–2408
  - infants, 2404
  - psychiatric syndromes in terminally ill children, 2405–2406
    - anxiety, 2406
    - delirium, 2406
    - depression, 2406
    - pain management in, 2406
  - school-aged children, 2404
  - toddlers, 2404
  - tripartite connection, 2403–2406
- physical and psychological symptoms, assessment, 2387–2388
- psychological accompaniment, 2397–2398
- psychological reactions of physicians, 2385
  - during epidemic or pandemic, 2385–2386

  - psychological syndromes and symptoms, 2393–2397
    - anxiety, 2394–2395, 2394*t*
    - confusional states, 2396–2397
    - depression, 2395–2396, 2396*t*
    - prevalence of, 2393
    - treatment of, 2393–2394
  - respiratory symptoms, 2393
  - societal implications, 2410–2411
  - spiritual concerns in, 2399
  - staff dynamics in, 2408
  - staff–family interface, 2403
  - voluntary refusal of food and water, 2410
- End-of-life delirium, 1231, 1238. *See also* Delirium
- Endogamy, 4729
- Endogenous affective disorders, 1652
- Endogenous depression, 4729. *See also* Major depressive disorder (unipolar depression)
- Endogenous opioids, 161–162
- Endogenous retroviruses, 241
- Endomorphic, 4729
- Endoreactive moods, 1671
- Endorphin, 4729
- Endothelial NOS (eNOS), 125
- Energy disturbances, 1188–1189
- Energy drinks, 1325
- Energy homeostasis, 226
- Energy psychology, 2647, 2649
- Engel, George, 1230, 2278, 4254
- Engle, George, 2874
- English Longitudinal Study of Ageing (ELSA), 4442
- Engram, 4729
- Engström, Georg, 4644

- Enkephalin, 4729
- Ensemble-coding hypothesis., 305
- Entactogen, 1348, 1354
- Entactogenic drugs, 431
- Entheogen, 1348
- Entomophobia, 4729
- Entorhinal cortex, 35
- Enuresis, 3789–3796, 4729
    - behavior therapy for, 3794
    - clinical features, 3793
    - comparative nosology, 3789–3790
    - course and prognosis, 3793–3794
    - definition of, 3789
    - diagnosis, 3793, 4810*t*
    - differential diagnosis, 3793–3794
    - diurnal, 3789
    - DSM-5-TR criteria for, 3790*t*
    - epidemiology of, 3790
    - etiology of, 3790–3793
    - historical background, 3789
    - ICD-11 criteria for, 3790*t*
    - laboratory examination, 3793
    - nocturnal, 3789
    - pathology, 3793
    - pharmacotherapy for, 3794–3796, 3934
    - primary, 3793
    - secondary, 3793
    - sleep-related, 2185
    - treatment of, 3794–3796
- Environment. *See also* Gene-environment interaction
    - everyday functioning and, 4181
    - psychiatric disorders and, 2615

  - role in mental health, 2613–2614
- Environmental neuropathies, 658
- Environmental tobacco smoke, 1387–1388. *See also* Tobacco use disorder (TUD)
- Environmental toxins, 1059–1060
- Enzyme deficiency disorders, 691–697
  - acute intermittent porphyria, 691
  - alpha-methylacyl-CoA racemase (AMACR) deficiency, 695
  - cerebral creatine deficiency syndromes, 695–696
  - cerebral folate deficiency, 697
  - cerebrotendinous xanthomatosis, 693–694, 694*f*
  - glycine encephalopathy, 696
  - hyperprolinemia, 695
  - maple syrup urine disease, 693
  - phenylketonuria, 691–692, 692*f*
  - propionic acidemia, 695
  - serine deficiency syndromes, 696–697
  - succinic semialdehyde dehydrogenase (SSADH) deficiency, 696, 696*f*
  - tetrahydrobiopterin deficiency, 694–695
  - urea cycle disorders, 692–693
- EOS. *See* Early-onset schizophrenia (EOS)
- Epac (exchange proteins activated by cAMP), 3005
- Ephebophiles, 2095
- Ephrins, 65
- Epidemic, psychological reactions of physicians during, 2385–2386
- Epidemiologic Catchment Area (ECA), 886–887. *See also* Geriatric psychiatry
- Epidemiology, 873, **873–895**
  - anorexia nervosa, 2150
  - application to psychiatry, 1791
  - assessment instruments, 883–885

    - for adult populations, 883–884
    - *Alcohol Use Disorder and Associated Disabilities Interview Schedule* (AUDADIS), 884
    - *Composite International Diagnostic Interview* (CIDI), 883
        - University of Michigan Version, 884
        - World Mental Health Surveys Version, 884
    - *Diagnostic Interview Schedule* (DIS), 883
    - diagnostic interviews for children and adolescents, 884–885
    - for disability and impairment, 885
    - *Structured Clinical Interview for DSM-IV* (SCID), 884
- case identification, 881–882
    - *Diagnostic and Statistical Manual of Mental Disorders*, 881–882
    - *International Classification of Diseases (ICD)*, 882
    - *International Classification of Functioning, Disability, and Health* (ICF), 882, 883*t*
- COVID-19 and, 894–895
- defined, 4729
- experimental studies, 878
- history, 4658
- major surveys, 886–895, 886*t*
    - *Epidemiologic Catchment Area Program* (ECA), 886–887
    - *National Comorbidity Survey* (NCS), 887–888
    - National Comorbidity Survey Replication Adolescent Supplement (NCS-A), 892
    - *National Epidemiologic Survey on Alcohol and Related Conditions* (NESARC), 890–891, 891*t*
    - *National Survey of Children's Health* (NSCH) survey, 892
    - *National Survey on Drug Use and Health* (NSDUH), 888–889, 889*t*
    - surveys of mental disorders in children and adolescents, 891–892
    - WHO Global Burden of Disease Study, 892–894, 893*t*
    - *World Mental Health* (WMH) *Initiative*, 889–890, 890*t*

    - Youth Risk Behavior Surveillance System (YRBSS), 892
  - measurement, 879–881
    - reliability, 879–880, 880*t*
    - validity, 880–881, 880*t*
  - measures of disease frequency
    - comorbidity, 875–876
    - incidence, 875
    - prevalence, 874–875
    - relationship between prevalence and incidence, 875
  - measures of effect and association
    - odd ratios, 876–877
    - relative risk, 876
    - risk, protection, and causation, 876
  - observational studies, 878–879
  - rumination disorder, 2151
  - study designs, 877–879
  - uses of, 873–874
- Epigenesis, 4729
- Epigenetic codes, genetic codes *versus*, 269*t*
- Epigenetic imprinting, 248
- Epigenetics, 248, 453–454, 1316, 2615
  - definition of, 269
  - machinery of, 270–278
  - positive mental health and, 4474
  - preclinical work, 278–282
  - of psychiatric illness, 282–285
  - psychiatry and, **269–285**
  - schizophrenia, 1508
  - of stress activity and maternal behavior, 278–279
- Epigenetic transmission, 159
- Epigenome, 248

- Epilepsy, 480–482. *See also* Myoclonic epilepsy with ragged red fibers
  - behavioral disorders in, 524*t*
  - behavioral disturbances in, 527
    - aggression, 527–528, 528*t*
    - anxiety disorders, 527
    - mood disorders, 527
    - sexuality, 528
    - suicide and self-harm, 528
  - cannabinoid for, 1346
  - course and prognosis, 529–530
  - defined, 4729
  - definition of, 520
  - diagnosis of, 524–529
  - differential diagnosis, 481t
  - epidemiology of, 521–522
  - etiology of, 522–523
  - historic literature on, 521
  - ictal features, 524, 524*t*
  - interictal features, 524–529, 524*t*
    - Gastaut–Geschwind syndrome, 526–527
    - interictal psychosis, 525–526, 526t
    - personality disorders, 526
  - interictal period, 520
  - neurodiagnostic tests, 529
  - neuropathology, 529
  - neuropsychiatric aspects of, 520–532
  - nosology, 521
  - pathology and laboratory examination, 529
  - pathophysiology of, 523
  - pediatric, 678–681
    - ADHD and, 680

    - adjustment issues, 681
    - aggression in, 679–680
    - anxiety in, 679
    - depression in, 679
    - epidemiology of, 678–679
    - interictal and postictal psychosis in, 679
    - psychogenic nonepileptic seizures, 680–681
    - seizures, 678, 678*t*
    - staring spells, 680
  - peri-ictal features, 524*t*, 525
  - peri-ictal period, 520
  - psychopathology of, 522–523
  - seizures, classification of, 521*t*
  - treatment of, 530–532
    - antianxiety drugs, 531
    - antidepressants, 530–531, 532*t*
    - antiepileptics, 530
    - antipsychotics, 530
    - drug interactions and, 531, 533*t*
    - psychotropic medications, 530–531
    - seizure management, 532
    - surgical, 531
- Epileptic dementia, 4729
- Epileptic equivalent, 4729. *See also* Complex partial seizures (CPSs)
- Epileptic personality, 4729. *See also* Explosive personality
- Epinephrine, 4729
  - anatomy and functions, 95
  - life cycle, 95–98
  - synthesis, 95–96, 95*f*
- Episodic, 4729
- Episodic memory, 730*t*, 777

  - object relations as, 814
  - optogenetics and, 815–816
  - recombinant DNA and, 815–816
  - schematization of, 817–819
- Epistemology, 4729
- Epitranscriptome, 249
- Epstein–Barr virus (EBV), 619
- Epworth Sleepiness Scale (ESS), 204, 2189
- Equanil. *See* Meprobamate (Equanil)
- Equilibration, 744–745, 754
- Equilibrium potential, 68
- Erectile dysfunction (ED). *See also* Male erectile disorder
  - avanafil, 3433, 3434*t*
  - definition of, 4729
  - diagnosis, 4816*t*
  - penile self-injection, 3434
  - phosphodiesterase-5, 3433
  - prevalence of, 3431
  - second-line therapies, 3434–3435
  - sildenafil, 3433, 3434*t*
  - surgical options, 3435
  - tadalafil, 3433, 3434*t*
  - testosterone, 3434
  - treatment of, 3433–3435
  - vacuum-assisted erection devices, 3434
  - vardenafil, 3433, 3434*t*
- Eremophobia, 4729
- Ergasia, 4729
- Ergoloid mesylates, 4320
- Ergot alkaloids, for migraine, 651*t*
- Erichson, John Eric, 4657
- Erickson, Milton, 2922, 2981

- Erikson, Erik, 421, 784, 790, **935–944**, 4690, 4729
  - application of concepts to clinical work, 943–944
  - biography of, 935–936
  - *Childhood and Society*, 937, 938–940
  - culture and, 942–943
  - *Identity: Youth and Crisis*, 938–939, 944
  - *Insight and Responsibility*, 938
  - life and work
    - as biographer of middle age, 941–942
    - as biographer of youth, 940–941
    - child psychoanalyst, 936
    - emigration to America, 936–937
  - life cycle stages, 937–940
    - autonomy *vs.* shame and doubt, 938
    - generativity *vs.* stagnation, 939–940
    - identity *vs.* role confusion, 939
    - industry *vs.* inferiority, 939
    - initiative *vs.* guilt, 938
    - integrity *vs.* despair, 940
    - intimacy *vs.* isolation, 939
    - trust *vs.* mistrust, 938
  - midlife transition, 937
  - *The Nature of Clinical Evidence*, 943–944
  - psychoanalytic anthropology, 937
  - *Young Man Luther*, 940–941
- Ernst, Richard, 327
- Erogenous zone, 4729
- Eros, 4729
- Erotic, 4729
- Eroticized children, 2036
- Erotomania, 4729
- Erotomanic thought, 996*t*

- ERPs. *See* Event-related potential (ERPs)
- Erythrophobia, 4729
- Escape learning, 4729
- Escitalopram (Lexapro)
  - for anxiety in older adults, 4224–4225
  - chemistry, 3138
  - dosing and administration, 3141–3142
  - pharmacogenetics of, 295*t*
  - pharmacologic actions, 3138
  - for tic disorders, 3785–3786
- Esdaile, James, 2912
- Eskalith. *See* Lithium
- Esketamine, **3378–3385**
  - dosage and administration, 3385, 3385*t*
  - drug interactions, 3384
  - drug studies, 3380–3382
  - major depressive disorder, 3380–3382
  - mechanism of action, 3381*f*
  - molecular structure, 3379*f*
  - organ/system effects of, 3382–3384
  - pharmacologic actions, 3379–3380
  - precautions and adverse reactions, 3383–3384, 3384*t*
  - suicidality associated with TRD, 3382
  - therapeutic implications, 3383
- Esmarch, Friedrich, 4644
- ESP. *See* Extrasensory perception (ESP)
- *Espiritismo*, 2672
- Esquirol, Jean Etienne Dominique, 1239, 4643, 4656, 4685
- Esquirol, Jean-Philippe, 1634, 1652, 3630
- Essentialism, 2059
- Essential oils, 2640
- Essential tremor, 575

- Estazolam (ProSom), 1429*t*
- *Estelle v. Gamble*, 4628
- Estimators, 897
- Estradiol, 163
  - for acute psychosis, 3428
  - for perimenopausal depression, 3426–3428
  - for postpartum depression, 3426–3428
- Estriol, 163
  - structure of, 1436*f*
- Estrogen, 163–164, 4730
- Estrogens, 4322–4323
  - for schizophrenia, 187
- Estrone, 163
- Ethanol, 4646
  - neurochemical effects of, 1298
- Ethchlorvynol (Placidyl), 1426*t*, 3236
- E-therapy, 1121
- Ethics
  - in child and adolescent psychiatry, **4010–4025**
    - codes of medical ethics, 4011
    - cultural and religious effects, 4013–4014
    - current and future issues, 4016–4024
    - dilemmas, 4012
    - principles, 4012
    - provision of care, 4014–4016
    - teaching, 402–405
  - in geriatric psychiatry, 4424–4426
    - autonomy, 4424–4426
    - core principles, 4424
    - decision-making capacity, 4424–4426
    - informed consent, 4424–4425
  - in psychiatry

    - additional information and expertise, 4618
    - basic principles, 4617, 4617*t*
    - confidentiality, 4620–4621
    - conflicts of interest, 4622
    - continuous learning, 4619
    - decision-making, 4619–4621
    - dilemmas, 4617–4619
    - informed consent, 4621–4622
    - issues and tensions, 4618
    - professional boundaries, 4620
    - psychiatric opinions, 4622–4623
    - risk and safeguards, 4618–4619
    - scope of practice, 4619
    - therapeutic relationship, 4620
    - third parties, 4622
    - values and perspectives, 4619
- Ethnic diversity, 835
- Ethnic identity
  - in adolescents, 3506–3507
  - in children, 848
  - development of, 3503–3504
- Ethnicity, 835
  - anxiety disorders and, 1793–1794
  - everyday functioning and, 4181
  - homosexuality and, 2077
  - mental disorders and, 839–841
  - mood disorders and, 1665
  - poverty and, 838–839, 838*t*
- Ethnic minorities, schizophrenia in, 1474–1475
- Ethnocentrism, 4730
- Ethnocultural betrayal trauma, 1935–1936
- Ethnology, 4730

- Ethnopsychiatry, 1468–1469
- Ethology, 4730
- Ethosuximide, 520
- Etiology, 4730
- Etomidate, 3236
- Euchromatin, 236
- Eudaimonia, 2893
- Euergasia, 4730
- Eunuch, 2102
- Euphoria
  - defined, 4730
  - multiple sclerosis and, 580
- Euphoric mood, 4730
- Euphorohallucinogen, 4730
- European Kleinian School, 2793
- European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), 1294
- Euthanasia, 2409
  - physician-assisted, 2409–2410
- Euthymia, 4730
- Euthymic, 4730
- Evasion, 4730
- Event recording, 2815
- Event-related euphoria, 4730
- Event-related potential (ERPs), 380–381
- Everolimus, neurotoxicity of, 2465*t*
- Everyday functioning, **4178–4194**. *See also* Activities of daily living (ADLs); Geriatric psychiatry
  - activities of daily living, 4178–4179
  - assessment of, 4181–4191
  - clinical interview of, 4184, 4184*t*
  - clinical populations and, 4191–4193

  - clinician rating of, **4048–4060**
  - cognition and, 4180–4181
  - collateral report of, 4186
  - culture/ethnicity and, 4181
  - definition of, 4178
  - dementia and, 4191–4192
  - depression and, 4192
  - direct observation of, 4189
  - domains of assessment, 4179–4180
  - environment and, 4181
  - factors influencing, 4180–4181
  - future instrument development, 4194
  - HIV in older adults and, 4461–4462
  - instrumental ADLs, 4179–4180
  - mood and, 4181
  - multidomain performance-based batteries of, 4182–4183*t*
  - normal aging and, 4189
  - of people living with HIV, 4193
  - performance-based measures of, 4187, 4188–4189*t*
  - physical disability and, 4181
  - quality of life, 4180, 4180*t*
  - of recently hospitalized older individuals, 4193
  - schizophrenia and, 4192–4193
  - self-report of, 4185–4186, 4185*t*
  - social functioning, 4180
  - virtual reality performance-based measures, 4188–4190, 4190*t*
- Evidence-based treatments (EBT), 4564
- Evoked potentials (EPs), 381, 1062, 4730
- Evolution, 864
  - and violence, 2514
- Evolutionary biology
  - core principles of, 863–866

    - evolution, 864
    - genotypes and phenotypes, 864–865
    - levels of selection, 865–866
    - natural selection, 863–864, 864
    - selection of systems, 864
    - sexual selection, 865
- Evolutionary medicine, 862–863
- Ewalt, Jack, 4486
- Exaltation, 4730
- Examination of psychiatric patient
  - affect, 994–995
  - appearance and behavior, 994
  - chief complaint, 991
  - cognitive assessment, 997
  - eye contact, 994
  - family history, 993
  - formulation, 998
  - history of present illness, 991
  - insight, 997
  - interviews, 986–1006
  - mental status examination, 994, 994*t*
  - mood, 994
  - motor activity, 994
  - past medical history, 993
  - past psychiatric history, 991–992, 992*t*
  - perception, 996, 996*t*
  - physical assessment, 998
  - privacy, 987
  - psychiatric report, **1007–1012**
  - review of systems in, 994
  - risk assessment, 996–997
  - social and developmental history, 993, 993*t*

  - speech, 995
  - substance use, abuse, and addictive behaviors, 992–993
  - suicide and violence risk history, 992
  - thought content, 995
  - thought process/form, 995, 995*t*
  - treatment recommendations, 998
- Excessive daytime sleepiness, 3111
- Excessive daytime somnolence (EDS), 4084–4085
- Excessive fragmentary myoclonus, 2188
- Excitatory (depolarizing) postsynaptic currents (EPSCs), 84
- Excited, 4730
- Excitement phase, 4730
- Excitotoxicity, 87, 4730
- Excoriation (skin-picking) disorder, 1841*t*, **1852–1854**, 2435–2436
  - clinical features, 1853–1854, 2436
  - comparative nosology, 1853
  - course and prognosis, 1854
  - defined, 4730
  - definition of, 1852
  - diagnosis of, 1853–1854, 1853*t*, 4799*t*
  - DSM-5 criteria for, 2436*t*
  - epidemiology of, 1853
  - etiology of, 1853
  - history, psychopharmacologic, 1852–1853
  - laboratory examination, 1854
  - pathology of, 1854
  - psychiatric comorbidities of, 2436
  - treatment of, 1854, 2436
- Executive cognitive dysfunction, 474–475
- Executive ego function, 4730
- Executive functions, 1023, 1033
  - assessment, 3569

  - cognitive aging, 4163–4164
  - neuropsychological evaluation, 4170
- Exelon. *See* Rivastigmine (Exelon)
- Exercise, 2642
  - appetite and, 230
  - for pain, 2490
- Exhaustion state, 4730
- Exhibitionism, 4730
- Exhibitionistic disorder, 2094, 4838*t*
- Existentialism, 4730
- Existential psychotherapy, 2979–2981
  - defined, 4730
  - ethical issues in, 2981
  - history, 2979–2980
  - indications for, 2981
  - research and evaluation, 2981
  - techniques, 2980–2981
  - theoretical issues, 2980
- Exner, John E., 1045
- Exner's Comprehensive System, 1045, 1046*t*
- Exogamy, 4730
- Exogenous, 4730
- Exon arrays, 44
- Exonic splicing enhancers (ESEs), 246
- Exonic splicing silencers (ESSs), 246
- Exorcism, 4730
- Exosomes, 242
- Expanding, interventions, 1000–1001
- Expansive mood, 4730
- Expected/maximization (EM) algorithm, 905
- Expected value, 897
- Experience, 806–807

- Experienced gender, 2101
- Experiencing, 4730
- Experiential avoidance, 2819
- Experiential/humanistic family therapy, 2948
- Experimental bias, 4730
- Experimental design, 4730
- Experimental group, 4730
- Experimental methods. *See* Quantitative and experimental methods
- Experimental neurosis, 4730
- Experimental studies, 877–879
  - sampling and sampling strategies, 879
- Expert witness, 4730
- Explanatory model, 2553
- Explanatory Model Interview Catalogue (EMIC), 2570
- Explicit memory, 730, 730*t*
- Exploding head syndrome, 2185
- Exploitation, family interventions and, 4379–4380
- Exploitative characters, 959
- *Explorations in Personality: A Clinical and Experimental Study of Fifty Men of College Age* (Murray), 937
- Exploratory hypnotherapy, 2922
- Explosive personality, 4730
- Exposure therapy, 760, 1825–1826
  - flooding, 2816
  - imaginal exposure, 1826
  - interoceptive exposure, 1826
  - between-session extinction, 2816
  - within-session extinction, 2816
  - systemic desensitization, 2816
  - virtual reality exposure, 2816
  - in vivo exposure, 1825

- Expressed emotion (EE) index, 1477
- Expressed gender, 2101–2102
- Expressed sequence tags (ESTs), 245
- Expressive aphasia, 4730
- Expressive aprosodia, 467*t*
- Expressive dysphasia, 4730
- Expressive language, 4168
- Expressive movement therapies, 2642
- Expressive therapy, for dissociative identity disorder, 2000
- Extended amygdala, 36
- Extended-family therapy, 4730. *See also* Social network therapy
- Extended forensic interview (EFI), 3975
- External capsule, 13
- External homophobia, 2075
- Externalization, 921*t*, 4730
- External loop, 2380
- External mentalizing, 2864
- External validators, 1093
- External validity, 4730
- Extinction, 757, 760, 1878, 4730
- Extinction model, 2816
- Extinction recall, 1878
- *Extracampine* hallucinations, 1453–1454
- Extrapsychic conflict, 4730
- Extrapyramidal dysarthria, 467*t*
- Extrapyramidal dyskinesia, 4730–4731
- Extrapyramidal effect, 4731
- Extrapyramidal motor system, 4731
- Extrapyramidal signs, 481t
- Extrapyramidal syndromes (EPS), 1580, 3064, 3070
  - acute, 3046–3047
  - akathisia, 3046

  - amantadine for, 3111, 3332, 3334
  - anticholinergics and, 3327
  - brexpiprazole and, 3104
  - chronic, 3047–3048
  - defined, 4731
  - dystonia, 3046
  - iloperidone and, 3083
  - lumateperone and, 3089
  - management of, 3046–3047
  - Parkinsonism, 3046
  - perioral tremor, 3048
  - tardive dyskinesia, 3047–3048
- Extrasensory perception (ESP), 4731
- Extraversion, 4731
- Extreme capsule, 13
- Extroversion, 4731
- Extroversion–introversion polarity, 951
- Ey, Henry, 1889–1890
- Eyberg Child Behavior Inventory (ECBI), 3546, 3791
- Eye contact, 994
- Eye fixation/closure, 2921. *See also* Hypnotic induction ceremony
- Eye movement desensitization and reprocessing, for dissociative identity disorder, 1999–2000
- Eye roll technique, 2921. *See also* Hypnotic induction ceremony
- Eyes
  - abnormal eye findings, 481t
  - neurologic examination, 467
  - physical examination, 466
- Eysenck, Hans, 971, 3885
- Eysenck Personality Questionnaire (EPQ), 601
- Ezriel, Henry, 2956

F

- *Fa’afafine*, 2072
- Fabry disease, 687–688
- Facebook, 1128, 2701–2702
- Face-to-face therapy, 1121
- Face validity, 1036–1037, 1092
- Facial movement, 467
- Facilitating interventions, 998–1000
- Factitious amnesia, 2007
- Factitious disorder by proxy. *See* Factitious disorder imposed on another
- Factitious disorder imposed on another
  - clinical features, 1917–1918
  - course and prognosis, 1923
  - defined, 4731
  - diagnosis of, 1917–1918, 1918*t*
  - differential diagnosis, 1921
  - epidemiology of, 1909–1910, 1911
  - etiology of, 1911–1912
  - treatment of, 1924–1925, 1925*t*
- Factitious disorder imposed on self
  - clinical features, 1912–1915*t*, 1912–1917
  - clues for, 1918*t*
  - course and prognosis, 1921–1922
  - diagnosis of, 1912–1917
  - differential diagnosis, 1919–1921
  - DSM-5 criteria for, 1907t
  - epidemiology of, 1910–1911
  - etiology of, 1910–1911
  - ICD-11 criteria for, 1907t
  - treatment of, 1923–1924

- Factitious disorder(s), 1902–1905, **1906–1925**
  - comparative nosology, 1908–1909
  - defined, 4731
  - definition of, 1906–1907, 2493
  - dementia and, 1247
  - dermatitis artefacta, 2438
  - diagnosis of, 4807*t*
  - differential diagnosis, 1919–1921
  - DSM-5 criteria for, 1906–1907, 1907t, 1907*t*
  - etiology of, 1910–1912
  - history, 1907–1908
  - ICD-11 criteria for, 1906–1907, 1907t, 1907*t*
  - imposed on self, 1909–1910, 2506
  - Munchausen syndrome, 1917
- Factor analysis, 975–976, 1037, 4659, 4731
- Factor validity, 1037
- Failure to thrive (FTT), 3762, 4731
- Fairbairn, Ronald, 924, 1930
- Fairbairn, W.R.D., 2751
- Faking bad, 2493
- Faking good, 2493
- Falling
  - fear of, 4218–4219
  - in medically ill patients, 4207
- Falloon, Ian, 2948–2949
- Falret, Jean Pierre, 1647, 1652, 4658, 4686
- False discovery rate, 901
- False imputation, 2504
- False memories, 2939–2940, 4731
- False memory syndrome, 1934
- False negatives, 880, 4731
- False positives, 880, 4731

- Falsifiable hypothesis, 4731
- Families
  - future directions, 3924
  - high expressed emotion level within, 2547
  - intellectual disabilities and, 3646
  - of LGBT people, 2080–2081
- Family Bereavement Program (FBP), 4043–4044
- Family-centered care, 854
- Family dynamics, 4474
- Family Educational Rights and Privacy Act of 1974 (FERPA), 4036, 4036*t*
- Family First Prevention Services Act, 3957
- Family interventions
  - research and evaluation, 4379–4380
  - and therapy with older adults, **4372–4380**
    - assessment, 4374
    - bonds of attachment and caregiving, 4374–4375
    - capacity and decision making, 4379
    - clinical applications, 4375
    - clinical issues, 4375–4379
    - communication patterns, 4374
    - community referral, 4373–4374
    - contraindications, 4378
    - cultural considerations, 4379
    - elder abuse, neglect, and exploitation, 4379–4380
    - end-of-life care, 4379
    - ethical issues, 4379
    - history, 4373
    - indications, 4378
    - inpatient medicine service, 4375–4376
    - intervention levels, 4374
    - limitations, 4378

    - long-term care, 4376–4377
    - outpatient treatment, 4375
    - palliative care, 4377–4378
    - problem solving, 4374
    - psychoeducation, 4374
    - strength-vulnerability model, 4373
    - techniques, 4373–4374
    - theoretical issues, 4373
    - treatment goals, 4378–4379
- Family Media Use Plan, 4099
- Family neurosis, 4731
- Family studies
  - ADHD, 3729
  - neurodegenerative disorders, late-life, 4172–4173
  - schizophrenia, 1482–1483, 1616–1617
- Family systems, 2947
- Family therapy, **2942–2952**, **3908–3924**, 4731. *See also* Couples therapy
  - for adolescents, 3954
  - assessment, 3910–3913
  - assessment and evaluation, 2950–2951
  - assumptions of, 3908*t*
  - care levels and, 3914
  - clinical applications, 2949–2952
  - clinical management issues, 2952*t*
  - cognitive/behavioral, 2944–2945
  - contraindications for, 2950
  - cultural differences, 3912*t*
  - definition of, 2944
  - for dissociative identity disorder, 1999
  - for divorcing and blended families, 3922
  - engagement, 3910

- experiential/humanistic, 2948
- for families from diverse backgrounds, 3922–3924
- family systems, 2947
- general clinical principles, 2949–2950
- historical trends, 2943–2944
- indications for, 2950, 2950*t*
- life cycle, 2947
- narrative, 2949
- overview, 3907–3910
- parent-child interactions during, 3913–3914
- process, 3910–3913
- psychodynamic, 2949
- psychoeducational, 2948–2949
- safety, 3910
- schools of, 2944–2949, 2945–2946*t*
- solution-focused, 2949
- for specific problems, 3915–3924
  - anorexia, 3920–3921
  - attention-deficit/hyperactivity disorder (ADHD), 3915
  - children with developmental or medical challenges, 3924
  - conduct disorder, 3915–3918
  - depression, 3918–3919
  - oppositional defiant disorder, 3915–3918
  - substance abuse, 3920–3921
  - traumatic stress in children, 3919–3920
- strategic, 2947–2948
- structural, 2947
- tasks, 3908*t*
- theoretical foundations of, 3910*t*
- treatment approaches, 3911*t*
- variations in family composition, 3912*t*
- varieties of, 3909*t*

- Familywise error rate (FWER), 899–901
- Famotidine (Zantac 360, Pepcid), 107
- Fanapt. *See* Iloperidone (Fanapt)
- Fantasies, 2749–2750
- Fantasy, 813, 2787, 4731
- Faradic aversion, 2817
- Farm Bill, 1124
- Farrington, David, 2518
- Fast Fourier transform (FFT), 311, 366–368
- Fasting blood sugar, 1064
- Fatal familial insomnia (FFI), 210, 629–631, 638–639, 4731
- Father surrogate/father substitute, 4731
- Fatigue, 1188–1189
  - cancer-related, 2358–2359
  - defined, 4731
  - end-of-life care and, 2391
  - in HIV infections, 608
  - in medically ill, 2391*t*
  - in neuro-Covid, 190
  - in older adult cancer patients, 2379
- Fatty acids, for schizophrenia, 186
- Fatty liver, 1299
- Faulkner, Larry, 4586
- Fausse reconnaissance, 4731
- Faux de mieux, 4731
- Fava, Giovanni, 2898
- Favazza, Armando, 2684
- FDA. *See* Food and Drug Administration (FDA)
- Fear, 1780–1781
  - anatomical correlates of anxiety and, 1808–1813, 1809*f*
  - chronic pain in, 2486
  - circuitry, 1808–1809

  - contextual, 1811
  - defined, 4731
  - endocrine and autonomic responses of, 1812–1813
  - expression, 1809–1810, 1810*f*
  - learning, 1807–1808
  - neurochemical basis of, 1796
  - neuroimaging, 1813
  - regulation, 1810–1811
- Fear circuitry, 1208
- Fear conditioning
  - CNS circuits involved in, 1877–1878
  - connectionist modeling of, 321–322
  - PDP-style model, 322*f*
  - PTSD and, 1877–1878
- Fear-like behavior, 297–298
- Fechner, Gustav Theodor, 4686
- Federal Financial Participation (FFP), 4512
- Federally qualified health centers (FQHCs), 4482
- Federal Patient Self-Determination Act, 2401
- Federation of State Medical Boards (FSMB), 4590
- Federn, Paul, 2020, 2786
- Feeble-mindedness, 4731
- Feedback, 801, 4731
- Feedback projection, 24
- Feedforward projection, 24
- Feeding and eating disorders, **2147–2164**
  - animal models of, 233, 239*t*
  - anorexia nervosa, 231, 2147, 2148–2149, 2150*t*, 2151–2152, 2153, 2156, 2156–2157, 2158–2159, 2160–2162, 4808–4809*t*
  - avoidant/restrictive food intake disorder, 2147–2148, 2149, 2151, 2154, 2158, 4808*t*

- binge eating disorder, 231, 2147, 2149, 2150–2151, 2150*t*, 2152, 2154, 2156, 2158, 2159, 2162–2163, 4809–4810*t*
- bulimia nervosa, 2147, 2149, 2150, 2152, 2153–2154, 2156, 2157, 2159, 2162–2163, 4809*t*
- bupropion for, 3166
- in children and adolescents, 3583–3584
- cognitive behavioral therapy for, 2825
- comparative nosology, 2149–2150
- complications, 2155*t*
- course and prognosis, 2158–2159
- defined, 4728
- diagnosis and clinical features, 2152–2154, 4807–4810*t*
- differential diagnosis, 2156–2158
- epidemiology of, 2150–2151, 2150*t*
- etiology of, 2151–2152, 2152
- family therapy for, 3911*t*
- group psychotherapy for, 3904
- history, 2148
- hypnosis for, 2938
- ICD-11 criteria for, 1224
- intellectual disabilities and, 3647–3648
- monoamine oxidase inhibitors, 3194–3195
- naltrexone for, 3362
- neuroimaging, 3615–3616
- neurosteroids and, 139
- obesity and, 2307–2308
- other specified, 2148, 2150*t*
- pathology and laboratory examination, 2154–2156
- pica, 2148, 2149, 2151, 2154, 2158, 2164, 4807*t*
- rating scales for, 1111–1112
- rumination disorder, 2148, 2149, 2154, 2158, 2164, 4808*t*
- serotonin and, 232–233

  - sleep-related, 2183–2185
  - in somatic illnesses, 1188
  - suicide and, 2709
  - treatment of, 2159–2160
    - psychological, 2313–2314
  - unspecified, 2148
  - vulnerability to, 871
  - zonisamide, 3273
- Feeding disorder of caregiver–infant reciprocity, 3768–3769, 3769–3770
- Feeding disorder of infancy or early childhood, **3761–3773**
  - comparative nosology, 3761–3762
  - defined, 4731
  - definition of, 3761
  - DSM-5-TR criteria for, 3761
  - failure to thrive (FTT), 3762
  - feeding disorder of caregiver–infant reciprocity, 3768–3769, 3769–3770
  - feeding disorder of state regulation, 3767–3768
  - history, 3761–3762
  - infantile anorexia, 3762–3764
  - pica, 3771–3773
  - post-traumatic feeding disorder, 3766–3767
  - rumination disorder, 3770–3771
  - sensory food aversions, 3764–3766
  - subtypes, 3762
- Feeding disorder of state regulation, 3767–3768
  - clinical features, 3767–3768
  - course, 3768
  - differential diagnosis, 3768
  - etiology, 3768
  - treatment, 3768

- Feelings, 819
- Feighner, John P., 1201
- Feigning, 2493
- Felix, Robert, 4661
- Fellatio, 4731
- Female genital mutilation (FGM), 3978
- Femaleness, 4731
- Female orgasmic disorder, 2042–2043, 4731, 4817*t*
- Female premature orgasm, 2049
- Female sexual interest/arousal disorder, 2040–2041, 2041*t*, 4817–4818*t*
  - treatment of, 3433
- Female sexual response cycle, 2033*t*, 2034*f*
- Feminine identity, 4731
- Fenfluramine (Ponderax, Pondimin), for stereotypic movement disorders, 3813
- Fenichel, Otto, 2121
- Fentanyl (Sublimaze, Actiq)
  - pain management in burn patients, 2469
  - prevalence, 1399
- Ferenczi, Sandor, 1930, 2272, 2751, 2764, 3877, 4731
- Fergusson, Robert, 4643
- Ferraro, Armando, 4651
- Fertility, 2585–2595
  - control of, 2589–2595
    - contraception, 2591–2594
    - sterilization, 2594–2595
    - voluntary termination of, 2589–2591
  - early pregnancy loss, 2589
  - gametogenesis and, 2585–2586
  - infertility and, 2586–2589
  - pseudocyesis and, 2589

  - risperidone and, 3074
- Fetal alcohol spectrum disorders (FAS), 92, 666–668
  - behavioral findings, 667
  - cognitive findings, 667–668
  - comorbidity, 668
  - diagnostic criteria, 666–667, 666*t*, 667*t*
- Fetal alcohol syndrome, 1299–1300, 3642*t*, 4731
- Fetal development, 2580–2581
- Fetal solvent syndrome, 1373–1374
- Fetishism, 4731
- Fetishistic disorder, 2094, 4841*t*
- FGAs. *See* First-generation antipsychotics (FGAs)
- Fibromyalgia, 664–665, 664*t*, 2484, 3275, 4731
- Ficompa. *See* Perampanel (Ficompa)
- Fidelity, 939, 4012
- Field theory, 4731
- Fifteen Item Test (FIT), 2505
- Fight–flight response, 863
- Figure drawing, 1045*t*
- Filtered back projection, 389–390, 389*f*, 390*f*
- Finances
  - of geriatric psychiatric care, 4392–4402
    - accountable care organizations, 4401–4402
    - demographics, 4393–4394
    - healthcare costs, 4393–4394
    - historical background, 4394
    - Medicaid system, 4398–4399
    - Medicare system, 4394–4399
  - in geriatric psychiatry
    - current and future directions, 4402
    - out-of-pocket costs, 4399–4400
    - Patient Protection and Affordable Care Act, 4400–4401

- Finger agnosia, 4731
- Fink, Max, 4656, 4660
- First-generation antipsychotics (FGAs), 3027, 4306
  - absorption, 3039
  - categories, 1578
  - dosage/administration, 1579, 1583–1584
  - drug interactions, 4310
  - elimination, 3039
  - formulation and dosing information, 1576–1577*t*
  - long-acting injectable, 1578
  - long-term effects of, 1578
  - maintenance treatment, 3079
  - mechanism of action, 3065
  - pharmacologic actions, receptor activity, 4306
  - precautions/adverse reactions
    - cerebrovascular adverse events, 1251, 4310
    - extrapyramidal side effects, 3047, 3070
    - metabolic effects, 4311
    - weight gain, 1063, 4311
  - pregnancy and, 2606, 2606–2607
  - side effects
    - akathisia, 3095
    - cardiovascular, 2291*t*
    - extrapyramidal side effects, 1580
    - metabolic effects, 1581–1582
    - sexual dysfunction, 3435
    - tardive dyskinesia, 1580–1581, 3096
  - therapeutic indications
    - acute mania, 1729, 1731, 1733
    - aggression in children with ASD, 3933
    - dementia with Lewy bodies, 1268

    - dementia with psychosis or severe behavioral disturbances, 4308
    - epilepsy, 530
    - non-affective psychosis, 2599
    - psychosis and agitation in dementia, 1250
    - schizophrenia, 1562, 1564
- First interstitial nucleus of the anterior hypothalamus (INAH1), 2069
- Fischer, Charles, 1930
- Fischer, Emil, 3184
- Fisher, Ronald, 900
- 504 Plan, 4031, 4032*t*
- Five-Factor Model, 975–976, 980, 2254
  - clinical applications, 983–984
- Fixation, 4731
- Fixed effects term, 909
- Fixed-interval schedule, 4762. *See also* Schedule of reinforcement
- Fixed-ratio schedule, 4762. *See also* Schedule of reinforcement
- *FKBP5* gene, 1952
- Flagellantism, 4731
- Flagellation, 4731
- Flashback hallucinosis, 4732
- Flashbacks, 1343
- Flashing, 2094
- Flat affect, 1459, 4732
- Flavor hedonics, 222–223
  - brief-access eating and, 222–223
  - innate, 222–224
  - learned, 222–224
  - psychophysics, 222
  - sham eating and, 222–223
- Flexibilitas cerea, 4732

- *Flexibilitas cerea* (waxy flexibility), 469
- Flexible Interview for ICD-11 (FLII-11-Epidemiology), 3540
- Flexner, Abraham, 4587
- Fliess, Wilhelm, 2272
- Flight of ideas, 1676, 4732
- Floccillation, 4732
- Flooding, 2816, 4732
- Florida Obsessive Compulsive Inventory (FOCI), 1108
- Flourens, Pierre, 417
- Fluency, schizophrenia and, 1557
- Fluent aphasia, 1182, 4732
- Fluid intelligence, 4164
- Flumazenil (Ramazicon), 3254–3255
  - for benzodiazepine detoxification, 1430
  - dosing and administration, 3255
  - precautions and adverse reactions, 3255
  - therapeutic indications, 3255
- Fluoxetine (Prozac), 3135–3136
  - absorption, 3135
  - for bulimia nervosa, 2162
  - chemistry, 3135
  - distribution, 3135–3136
  - dosing and administration, 3141
  - history, 3007, 4663
  - metabolism and elimination of, 3136
  - pharmacokinetics of, sex differences in, 3416
  - pharmacologic actions, 3135–3136
  - for social anxiety disorder, 3859
  - for substance use disorder, 4001*t*
  - for tic disorders, 3785–3786
- Fluphenazine (Prolixin), 3034
- Flurazepam (Dalmane), 1429*t*

- Fluvoxamine (Luvox)
  - for anxiety disorders in later life, 4224–4225
  - chemistry, 3137
  - dosing and administration, 3142
  - pharmacogenetics of, 295*t*
  - pharmacokinetics of, sex differences in, 3416
  - pharmacologic actions, 3137–3138
  - for tic disorders, 3785–3786
- Foa, Edna, 1822–1823
- Focal attention, 730
- Focal neurobehavioral syndromes, 470–480
  - abnormal beliefs and experiences, 476
  - abnormalities in agency, 476
  - abnormal social behavior, 476
  - amnesia, 472–474
  - aphasia, 471–472, 473*t*
  - attentional disorders, 472
  - delirium, 471
  - dementia, 471
  - disordered mood and emotion, 475
  - dysnarration, 475
  - electroencephalography, 477
  - executive cognitive dysfunction, 474–475
  - functional neuroimaging, 477
  - laboratory investigations, 478
  - neuroimaging, 476–477
  - neuropsychological assessment, 479–480
  - paraclinical investigations, 476
  - polysomnography, 477
  - visuospatial dysfunction, 474
- FOCI (Florida Obsessive Compulsive Inventory), 1108
- Fodéré, François, 4656

- Foerster, Otfried, 369
- Folate (vitamin B9), 2632, 2652–2653
  - depression and, 2652–2653
  - schizophrenia and, 2653
- Foley, Katherine, 2384
- Folia, 28
- Folie à deux, 1647, 4732
- *Folie à double forme*, 1652
- Folk illness, 2551, 2551*t*
- Follicle-stimulating hormone (FSH), 2343, 2582, 2583–2584
  - for paraphilic disorders, 2138
- Fölling, Ivor Asbjörn, 4644
- Follow-up examination, 4732
- Folstein Mini-Mental State Examination (MMSE), 4168
- Fonagy, Peter, 732, 928, 2930
- Food, Drug, and Cosmetic Act of 1938, 3021
- Food addiction, obesity and, 2306–2307
- Food and Drug Administration (FDA), 4732
  - future directions, 3026
  - structure of, 3022
- Food and Drug Administration Amendments Act of 2007, 3022
- Food and Drug Administration Modernization Act of 199, 3022
- Foramina of Monro, 15
- Forced desynchrony, 209
- Ford, Cleland, 2062
- Fordyce, Michael, 2897
- Forebrain
  - developing, 59*f*
  - embryonic, 61–62
    - neuron generation during cortical development, 62
- Foreign accent syndrome, 467*t*
- Forel, Auguste, 4650

- Forensic geropsychiatry
  - administrative aspects of, 4421–4422
  - advanced directives, 4421–4422
  - civil aspects of, 4417–4422
  - competence to stand trial restoration, 4416
  - competency to drive, 4421
  - competency to make a will, 4420–4421
  - competency to make decisions about death and dying, 4419–4420
  - competency to make financial decisions or contract, 4421
  - competency to make medical decisions, 4419
  - competency to stand trial, 4415–4416
  - consent to be a research subject, 4419
  - correctional aspects of, 4422–4423
  - criminal aspects of, 4415–4447
  - decision-making capacity, 4417–4418
  - informed consent, 4417–4418
  - mental health defenses, 4416
  - nosologic/theoretical issues, 4423
  - older adult applicants for disability benefits, 4422
  - older adult impaired physician, 4422
  - physician aid in dying, 4419–4420
  - restoration of sanity, 4416–4417
  - sentencing, 4417
  - testifying in court, 4423
  - undue influence, 4418–4419
- Forensic interviews, 1005–1006
- Forensic psychiatry
  - child/adolescent, **4003–4010**
    - civil litigation, 4006–4007
    - custody issues, 4005–4006
    - evaluations, 4004–4005

    - informed consent, 4004–4005
    - referrals, 4004–4005
    - written report, 4004–4005
  - future directions, 4009–4010
  - guidelines, 4004
  - juvenile justice cases, 4007–4009
  - overview, 4003–4004
- clinical-legal issues, **4607–4615**
  - child abuse, 4615
  - child custody, 4614–4615
  - children and the law, 4613–4615
  - confidentiality and privilege, 4609–4610
  - criminal competencies, 4612–4613
  - criminal responsibility and psychiatric illness, 4612–4613
  - doctor-patient relationship, 4607–4608
  - expert witness testimony, 4612
  - involuntary hospitalization and treatment, 4608–4609
  - juvenile justice, 4614
  - malpractice, 4610–4611
  - medical decision-making and competence, 4608
  - testamentary capacity, 4612
- defined, 4732
- ethics, 4615–4623
  - additional information and expertise, 4618
  - basic principles, 4617, 4617*t*
  - confidentiality, 4620–4621
  - conflicts of interest, 4622
  - continuous learning, 4619
  - decision-making, 4619–4621
  - dilemmas, 4617–4619
  - informed consent, 4621–4622
  - issues and tensions, 4618

    - professional boundaries, 4620
    - professionalism, 4623
    - psychiatric opinions, 4622–4623
    - risk and safeguards, 4618–4619
    - schools of thought, 4616–4617, 4616*t*
    - scope of practice, 4619
    - self-care and impairment, 4623
    - therapeutic relationship, 4620
    - third parties, 4622
    - values and perspectives, 4619
  - history, 4698
  - hypnosis in, 2938–2942
- Foreplay, 4732
- Forepleasure, 4732
- Fork cell neurons (FNs), 428
- Formal abstraction, 745
- Formal operations, 748–749, 4732
- Formal thought disorder, 1562, 4732
- Formication, 1455, 4732
- Fornix, 35
- For-profit hospital, 4732
- Forward chaining, 2816
- Fosha, Diana, 2772
- Fosshage, James, 3877
- Foster care, **3956–3962**
  - demographics, 3957–3959
  - foster parents, 3961–3962
  - history, 3956–3957
  - mental health problems and, 3961
  - outcomes of, 3959–3961
  - overview, 3956
  - placement reasons, 3959

  - psychiatrist’s role in, 3962
  - treatment foster care (TFC), 3961
- Foster Care Independence Act, 3957
- Fostering Connections to Success and Increasing Adoptions Act, 3957, 3960
- Foster parents, 3961–3962
- Foucault, Michel, 2060, 4691
- Foulkes, S. H., 2956
- Four-Factor Theory, 1973
- Fourier transform, 332–335, 334*f*, 335*f*, 366
- *Fou rire prodromique* (mad prodromal laughter), 475
- Four-miss rule, 2850
- Fourth ventricle, 15
- Fractional anisotropy, 337
- Fragile X syndrome, 271*t*, 3632, 3639, 3640*t*
  - autism spectrum disorder and, 3639
  - comorbidities, 3639
  - defined, 4732
  - epigenetic etiology of, 282, 283*f*
  - single gene mutations in, 50
- Fragmentary delusion, 4732
- Fragmentation, 2750
- Fraiberg, Selma, 3877
- Frailty, in medically ill patients, 4207
- Frank, Jerome, 4697
- Frankl, Viktor, 973, 2384, 2980
- Franklin, Benjamin, 4644
- Franks, Cyril, 3885
- Franzen, Göran, 4661
- Franzen, Jonathan, 2883
- Fraternal twins, 4732
- Fredrickson, Barbara, 795

- Free and appropriate education (FAPE), 4026*t*
- Free and Appropriate Public Education (FAPE), 4033
- Free association, 934, 2751, 4732
- Freebase, 4732
- Freedman, Daniel X., 2673, 4652
- Free-floating anxiety, 4732
- Free induction decay, 329
- Freeman, Walter, 419, 3479, 4690
- Free recall, 2010
- Frégoli phenomenon, 1170
- Fregoli syndrome, 4732
- French, John, 4654
- Frequency, 4732
- Frequency, law of, 434
- Frequency distribution, 4732
- Freud, Anna, 919–922, 4647, 4664, 4732
  - on armor plating of character, 2231
  - *The Ego and the Mechanisms of Defense*, 937
  - *Introduction to the Technique of Child Analysis*, 3877
- Freud, Sigmund, 4647, 4687, 4732
  - anal stage, 3504
  - on anxiety, 2757
  - on cocaine, 1312
  - on death anxiety, 2383
  - dissociative disorders, 1930
  - drive-structure doctrine, 810
  - *The Ego and the ID*, 917
  - ego psychology, 918–922
  - on free association, 2751
  - on God, 2678
  - *Group Psychology and Analysis of the Ego*, 1705
  - on group psychotherapy, 2955

- on homosexuality, 2061
- on humor, 793
- on hypnosis, 2912
- on hysteria, 915
- *Inhibitions, Symptoms, and Anxiety*, 2799
- intensive short-term psychodynamic psychotherapy, 2764
- *Interpretation of Dreams*, 212
- “Little Hans” case, 3877
- on male erectile disorder, 2042
- on mania, 1705
- mental health definition, 785
- *Moses and Monotheism*, 4747
- on motivation and drive, 2747
- *Mourning and Melancholia*, 918–919, 1703
- on narcissism, 2774
- on object loss, 1654
- oral phase, 3504
- on paraphilias, 2091
- personality theory, 422
- phallic phase, 3504
- primal phantasy, 812–813
- on psyche and soma, 2272
- psychoanalysis of, 930, 2746, 4647
- psychoanalytic frame, 2751
- psychological determinism, 914
- on psychosexual development, 3504
- on psychotherapy, 2875
- publications, 4687
- on pyromania, 2197
- on religion, 2677
- significance of, 4687
- on soma and psyche, 2272

  - structural model of mind, 917–918, 2748–2749
  - theory of psychic immaturity, 2061
  - *Three Essays on the Theory of Sexuality*, 3504, 4769
  - topographic theory, 916–917
    - conscious system, 916
    - preconscious system, 916
    - psychosexual development, 916–917
    - unconscious system, 916
  - *Totem and Taboo*, 4769
  - on transference, 2749
  - trauma hypothesis, 915–916
  - vagina dentata, 2040, 4771
- Frewen, Paul, 1966
- Freyd, Jennifer, 1933
- Freyhan, Fritz, 4662
- Fricchione, Gregory, 4656, 4660
- Friedman, M., 2273
- Friendship Bench intervention, 2568
- Frigidity, 4732
- Frisch, Michael, 2899
- Friston, Karl, 418, 419
- Fromm, Erich, 958–959, 958*f*
  - personality theory, 958
  - psychopathology theory, 958–959
  - treatment by, 959
- Fromm, Erika, 2923
- Fromm-Reichmann, Frieda, 2831, 4664
- Frontal lobe, 9, 3606–3607
- Frontal lobe syndrome, 4732
- Frontal personality changes, multiple sclerosis and, 580
- Frontoparietal network (FPN), 803
- Frontostriatal-dysfunction theory, 444

- Fronto-subcortical circuits, 459–460
- Frontotemporal dementia, 564–565, 1259–1262, 4175
  - biomarkers, 1260–1262
  - clinical features, 1260
  - comparative nosology, 1259
  - course and prognosis, 1262
  - defined, 4732
  - definition of, 1259
  - diagnosis of, 1260
  - differential diagnosis, 1262
  - epidemiology of, 1259–1260
  - etiology of, 1260
  - genetics of, 1260
  - history, 1259
  - management of, 1262
  - neuropathology, 1262
  - sexual dysfunction and, 4451
- Frontothalamic-striatal circuitry, 1208
- Frostbite, 658
- Frostig Movement Skills Test Battery, 3676
- Frottage, 4732
- Frotteurism, 4732
- Frotteuristic disorder, 2094–2095, 4732, 4839*t*
- Frustration, 4732
- Fuchs, Karl, 2931
- Fugue, 4732
- Fulfillment, 4732
- Full malingering, 2504
- Fulton, John, 3479
- Functional, 4732
- Functional abdominal pain (FAP), 2936
- Functional Activities Questionnaire (FAQ), 4185–4186

- Functional amnesia, 782
- Functional analytic psychotherapy (FAP), 2819–2820
- Functional assessment, 4552–4553
- Functional Assessment Measure (FAM), 4186–4187
- Functional brain connectivity, 406
  - analysis of, 407–409
  - independent components analysis, 409–410
  - major depression, 415
  - modalities for assessment of, 407–408
  - network approaches to, 410
  - networks, 406–407, 407*f*
  - psychopathology and, 410–415
  - resting-state, 406
  - ROI–ROI or network-based connectivity, 409
  - ROI/seed-based connectivity, 409
  - schizophrenia and, 411–413
  - study of, 406
- Functional capacity, 4461
- Functional chest pain, 2297–2298
- Functional disorder, 4732
- Functional genomics, 3590–3591
- Functional hypothalamic amenorrhea, 2584–2585
- Functional impairment, dementia and, 1243
- Functional Independence Measure (FIM), 4179, 4186–4187
- Functionalism, 2886
- Functional magnetic resonance imaging (fMRI)
  - alcohol dependence, 345
  - applications in psychiatry, 343–347
  - brain imaging, **1522–1534**
  - brain stimulation and, 3467–3468
  - cortical connectivity, low-frequency, 344
  - dementia, 343

    - extravascular dynamic dephasing, 342
    - intravascular dynamic dephasing, 342
    - methodology, 342–343
    - mood disorders, 344–345
    - neuroimaging and, 538
    - during right-hand movement, 342*f*
    - schematic of, 1522*f*
    - schizophrenia, 343–344, **1522–1534**
        - analytic methods, 1524–1526
            - connectivity analysis, 1525–1526
            - group-level random effects, 1524–1525
            - pattern analysis and classification, 1525
            - subject-level time series, 1524–1526
        - arterial spin labeling, 1524
        - blood oxygenation level-dependent imaging, 1522
        - cognition studies, 1527–1529, 1529*f*, 1530*f*
        - emotion studies, 1529–1532, 1531*f*
        - event-related MRI, 1523, 1524*f*
        - genetics studies, 1532–1533
        - pharmacology studies, 1532
        - of at-risk individuals, 1533, 1534*f*
        - special considerations, 1526–1527, 1527*f*
- Functional neuroimaging, 477
- Functional neurologic symptom disorder (conversion disorder), 484–485, 1900–1902, 2506, 4732–4733
    - comparative nosology, 1900
    - definition of, 1900
    - diagnosis of, 4806*t*
    - epidemiology of, 1900
    - etiology of, 1900–1901
    - history, 1900
    - treatment of, 1901

- Functional somatic syndromes, 1891
  - across medical specialties, 1892*t*
- Functional speech disorder. *See* Developmental articulation disorder
- Funding imbalance effect, 1481
- Fürstner; Otto, 4649

G

- $GABA_A$ receptor-positive allosteric modulators, 3417
- $GABA_A$ receptors, 3237–3238
  - antagonists, 89
  - distribution of, 89
  - schematic representation of, 88*f*
  - structure and function of, 2998*f*
- GABA-benzodiazepine receptor complex, 3237–3238, 3238*f*
- $GABA_B$ receptors
  - $GABA_A$ receptor and, 89–90
  - schematic representation of, 89*f*
  - variants, 89–90
- GABA enhancers
  - anesthetic agents, 3236
  - barbiturates, **3229–3234**
  - nonbarbiturate, nonbenzodiazepine sedative-hypnotic agents, 3234–3235
  - sedative-hypnotic agents, 3234–3236
- GABAergic systems, 88
- GABA neurons
  - chandelier class of, 18
  - double-bouquet, 18
  - stimulant-related disorders and, 1315
  - subclasses, 18*f*
  - subpopulations, 18

- Gabapentin (Neurontin), **3264–3267**
  - alcohol and substance use disorders, 3267
  - for alcohol use disorders, 1310
  - anxiety disorders, 3267
  - chemistry, 3266
  - for chronic pain, 2490
  - clinical drug studies, 3266–3267
  - dosage and administration, 3267–3268
  - drug interactions, 3267
  - history, 3265
  - mood disorders, 3267
  - pain management in burn patients, 2470
  - pain syndromes, 3266–3267
  - pharmacodynamics, 3266
  - pharmacokinetics, 3266
  - postherpetic neuralgia, 3266, 3267
  - precautions and adverse reactions, 3267
  - restless legs syndrome, 3266–3267, 3267
  - seizure disorders, 3267
  - structure of, 3267*f*
  - therapeutic indications, 3267
- Gabbard, Glen, 2753
- Gabitril. *See* Tiagabine (Gabitril)
- Gaedecke, Friedrich, 1312
- GAF Scale. *See* Global Assessment of Functioning Scale
- Gage, Phineas, 417–418, 418*f*, 425, 798
- Gaines, Atwood, 2554
- Gait, 468
- Gait disorder, 481t
- Galactorrhea, 4733
- Galanin, 122, **168–169**
  - amino acid sequence, 109*t*

  - anxiety disorders and, 1804–1805
- Galantamine (Razadyne), 106, 3335*t*, 3341–3342, 4317–4318
  - for Alzheimer disease, 4315*t*
  - chemistry, 3341
  - clinical drug studies, 3341–3342
  - contraindications, 4318
  - dosage and administration, 3342, 4318
  - drug interactions, 3342
  - organ/system effects of, 3342
  - pharmacogenetics of, 295*t*
  - pharmacokinetics and pharmacodynamics of, 3341, 4315*t*, 4317–4318
  - pharmacologic actions, 3341
  - therapeutic indications, 3342
- Galanter, Marc, 2020
- Galen of Pergamon, 4683
- Gall, Franz Joseph, 417, 4649, 4685
- Gallagher, R. J., 3808
- Gallagher, Shaun, 420
- Gallant, Donald, 4654
- Galloni, Antonio, 4642
- Galton, Francis, 781, 4659
- Galvanic skin response (GSR), 4733
- Galveston Orientation and Amnesia Test (GOAT), 541
- Gambling
  - in America, 1857–1858
  - evolutionary risk-taking and, 1864–1865
  - future of industry, 1866–1867
  - historical background, 1856–1857
  - laws on, 1866
- Gambling disorder, **1856–1868**. *See also* Pathological gambling
  - clinical features, 1859–1861

  - clinical vignettes, 1864
  - comorbidity, 1861–1862
  - defined, 4733
  - diagnosis, 1858–1859, 4828*t*
  - differential diagnosis, 1861–1862
  - DSM-5 criteria for, 1858–1859
  - epidemiology of, 1859
  - episodic, 1858
  - etiology of, 1862–1863
  - family history, 1859
  - imaging studies, 1863–1864
  - neurotransmitters and, 1864
  - persistent, 1858
  - personality characteristics, 1861
  - predisposing materialist factors, 1864
  - psychoanalytic theories, 1863
  - treatment of, 1865–1866
    - psychopharmacologic, 1865–1866
    - psychotherapy, 1865
    - self-help interventions, 1866
- Games, 4468–4469
- *Games People Play* (Berne), 963
- Gamete intrafallopian transfer (GIFT), 2587
- Gametogenesis, 2585–2586
- Gaming disorder, 1224, 4828*t*
- γ-Aminobutyric acid (GABA), 87, 1767–1768, 2518, 4733
  - anatomy of GABAergic systems, 88
  - anxiety disorders and, 1801
  - ethanol and, 1298
  - schizophrenia pathology, 1506
  - sedative-, hypnotic-, anxiolytic-related disorders, 1420–1421
- γ-Aminobutyric acid type A receptor ($GABA_AR$), 1420–1422

- γ-Hydroxybutyric acid (GHB), 90, 1427–1428
- Gamma ventral capsulotomy, 3483
- Gamophobia, 4733
- Ganser syndrome, 1179, 2502, 4733
- Ganzini, Linda, 2410
- Gaps, 908
- Gardner–Diamond syndrome, 2438
- Gargoylism, 4733
- Garrod, Archibald, 236
- Garside, Roger, 4659
- GAS. *See* General adaptation syndrome (GAS)
- Gas neurosis, 2656
- Gastaut–Geschwind syndrome, 526–527
- Gastric mechanoreception, 224
- Gastrin, **169–170**
- Gastrin-releasing peptide (GRP), **169–170**
- Gastroesophageal reflux disease (GERD), insomnia comorbidity, 2169
- Gastrointestinal (GI) cancer, 2362
- Gastrointestinal disorders, **2293–2303**
  - benzodiazepines, 3247
  - centrally mediated abdominal pain syndrome, 2297, 2297*t*
  - functional, 2296–2297
    - biopsychosocial model, 2299–2300
    - psychiatric approach to, 2300–2301
    - psychological therapies for, 2302–2303
      - biofeedback therapy, 2303
      - cognitive behavioral therapy, 2302–2303
      - hypnotherapy, 2303
      - interpersonal therapy, 2303
      - psychodynamic psychotherapy, 2303
      - relaxation therapy, 2302

    - psychopharmacological management of, 2301–2302
      - antidepressants, 2301
      - antipsychotics, 2302
      - mood stabilizers, 2302
      - psychotropic drugs, 2302
    - psychosocial factors, 2299–2300
    - somatic presentation of depression in, 2298–2299, 2299*t*
  - functional chest pain, 2297–2298
  - inflammatory bowel disease, 2294–2296
  - irritable bowel syndrome, 2296
- Gastrointestinal system
  - alcohol and, 1299
  - esketamine effects on, 3383
  - ketamine effects on, 3383
  - nalmefene effects on, 3364
  - naltrexone effects on, 3364
- Gatekeeper, 4733
- Gatekeeper tests, 900
- Gates, Bill, 1
- Gates Law, 1
- Gateway drugs, 4733
- Gaucher disease, 688
- Gault decision, 4733
- Gaustad Mental Hospital, 4644
- Gavin, Hector, 1907
- Gaw, Albert, 2552
- Gay, 4733
  - acquisition of gay identity, 2073
  - definition of, 2059
  - HIV/AIDS in, 2085
- Gay, Peter, 4647
- Gay baby boom, 2079, 2080

- Gay gene, 2071
- Gay panic defense, 2082
- Gay Rights Movement, 4444
- GD. *See* Gender dysphoria (GD)
- Gebhard, Paul, 2064
- Gegenhalten, 4733
- Geleerd, Elisabeth, 1930
- Gelenberg, Alan, 1908
- GENCODE annotation, 236*t*
- Gender, 2101, 2581–2582
  - definition of, 4100
  - legal change of, 2117
  - mood disorders and, 1660, 1663–1664
  - terminology, 4100*t*
  - violence and, 2516
- Gender affirmative approach, 2112–2113
- Gender-affirming surgery (GAS), 2102
- Gender assigned at birth, 2101
- Gender binary, 4100*t*
- Gender confirmation surgery, 2102
- Gender constancy, 2109
- Gender dysphoria (GD), 2060, 2101
  - in adolescents, 2107–2108, 2117, 4821*t*
  - in adults, 2107–2108, 2117, 4821*t*
  - in children, 2107, 2117, 4100, 4101*t*
  - clinical approaches, 2112–2117
  - clinical issues, 2116–2117
  - cultural issues, 2112
  - defined, 4733
  - diagnosis, 2107–2108, 4821*t*
  - DSM-5-TR diagnosis, 2108
  - ethical and legal issues, 2117–2119

  - ICD-11 diagnosis, 2108
  - origins of, 2108–2111
  - social issues, 2112
  - training and research needs, 2119–2120
  - treatment of
    - in adolescents, 2113–2114
    - in adults, 2114–2115
    - in individuals with DSDs, 2115–2116
    - in prepubescent children, 2112–2113
- Gender expression, 4100*t*
- Gender identity, 2032, 2060, 2101, 3503, 4021, 4100*t*
  - acquisition of, 2108–2110
  - defined, 4733
  - psychosocial and biologic factors, 2109–2110
- Gender identity conversion efforts, 2112
- Gender identity disorder (GID), 2101
  - defined, 4733
  - group psychotherapy for, 3904
- Gender incongruence (GI), 2101
  - of adolescence, 2108
  - of adults, 2108
  - of childhood, 2108
  - distress associated with, 2110–2111
- Gender minority stress, 2111
- Gender nonconforming (GNC), 2102
- Genderqueer, 2101
- Gender reassignment surgery (GRS), 2102
- Gender-reversed homosexuality, 2072
- Gender roles, 2032–2033, 2101, 4100*t*, 4733
- Gender stability, 2109
- Gender transitions, 4100–4101, 4101*t*
- Gender variance

    - disorder or normal, 2105–2106
    - frequency of, 2106–2107
    - theories of, 2103–2106
        - historical, 2103–2106
- Gene-environment interaction, 1494–1495. *See also* Environment
- Gene expression, 450–452
- Gene mapping
    - approaches to, *252f*, *253f*
        - genetic epidemiologic, 252–253
    - basic concepts of, 253–254
    - genetic markers, 255
    - linkage, 254
    - linkage disequilibrium, 254
    - phenotypes, 257–259
        - animal models, 259
        - categorical, 257–259, *257f*
        - continuous, *257f*, 258–259
    - of psychiatric disorders, **251–268**
    - recombination, 254
    - strategies, 255–257
        - association studies, 256
        - pedigree analysis, 255
        - sib pair analysis, 255–256
        - statistical considerations, 256–257
- General achievement testing, 3567
- General adaptation syndrome (GAS), 4733
- General anesthetics, EEG effects, 379
- General education, *4026t*
- Generalization, 4733
- Generalized anxiety. *See* Free-floating anxiety
- Generalized anxiety disorder (GAD)
    - behavioral approach tasks for, *3888t*

  - bipolar disorder and, 1742
  - buspirone for, 3170
  - child/adolescent, 1817–1818, **3855–3857**
  - clinical features, 1787, 3856
  - cognitive behavioral therapy for, 1829–1830, 1830
  - comparative nosology, 1787, 3855
  - course, 1787–1788
  - criteria for, 1787*t*
  - defined, 4733
  - definition of, 1787, 3855
  - diagnosis, 1787, 3856, 4797*t*
  - differential diagnosis, 1787–1788, 3857
  - DSM-5-TR criteria for, 3856*t*
  - epidemiology of, 1787, 3855–3856
  - etiology of, 1787–1788, 3856
  - history, 1787
  - hypnosis and, 2928–2930
  - ICD-11 criteria for, 1222–1223
  - insomnia and, 2169
  - neuroimaging, 1817–1818, 3612
  - neuropeptide systems and, 115
  - in older adults, 4221–4222, 4427–4428
  - prevalence of, 488, 1791
  - psychodynamic therapy for, 2760–2761
  - short-term dynamic psychotherapy for, 2778
  - sleep disorders and, 4089
  - SSRIs for, 3131
  - temperament and, 3626, 3626*f*
  - trazodone, 3199
  - treatment of, 1787–1788, 1837, 3859
  - venlafaxine for, 3146–3147
- Generalized Anxiety Disorder Scale (GAD-7), 1106, 3544–3545

- Generalized epilepsy with febrile seizures plus (GEFS+), 81
- Generalized estimating equations (GEE), 909
- Generalized tonic–clonic seizure, 4733
- General linear model, 908–909
- General paralysis, 4733
- General paresis, 4733
- General personality disorder, 4833–4834*t*
- General systems theory, 4733
- Generativity, 791, 939–940
- Genes, 236
  - coding, 237–239
  - noncoding, 239–240
  - violence and, 2516–2519
- Gene splicing, 237
- Genetic anticipation, 3489
- Genetic codes, epigenetic codes *versus*, 269*t*
- Genetic counseling, **2612–2622**, 4733
  - challenges to, 2620–2621
  - definition, 2613
  - ethical, legal, and social considerations, 2621–2622
  - focus of, 2614
  - outcomes of, 2613
  - priorities of, 2615–2620
    - complex etiology with unusual accumulation of risk factors, 2620
    - educating clients, 2616
    - risk assessment, 2618
      - based on family and medical history, 2618–2620, 2620*f*
    - risk reduction in children, 2616–2617
    - steps and process of, 2616*t*
  - for psychiatric disorders, 2613–2614
- Genetic disorders, 1057–1058

  - Huntington disease, 1058
  - Wilson disease, 1058
- Genetic endowment, 4733
- Genetic markers, 255, 4733
- Genetics
  - in child psychiatry, **3586–3594**
  - dissociation and, 1952
  - positive mental health and, 4474
- Gene transcription, 1316
  - regulation of, 47
- Geneva Declaration on Armed Violence and Development (GDAVD), 2512
- Gene X environment, 3663
- Genital phase, 4733
- Genital stage, 919*t*
- Genito–pelvic pain, 4733
- Genitopelvic pain/penetration disorder, 2044, 4818–4819*t*
- Genitourinary cancer, 2363–2364
- Genitourinary system, 3365
  - esketamine effects on, 3383
  - ketamine effects on, 3383
  - naltrexone effects on, 3365
- Genomes
  - dark matter of, 235
  - definition of, 235
  - function of, 243–244
  - online scientific resources, 250*t*
  - role in mental health, 2613–2614
  - sizes, 236*t*
- Genome sequencing, 3590
- Genome-wide association studies (GWAS), 256, 3588–3589
  - ADHD, 3730

  - animal models of, 234
  - definition of, 3587*t*
  - neurodegenerative disorders, late-life, 4173–4174
  - psychiatric disorders and, 3589*t*
  - schizophrenia, 1483–1485
  - timeline of publications, 3588*f*
- Genomics, 234, 4017–4018
- Genotypes, 864–865, 4733
- Genotype-Tissue Expression (GTEx) consortium, 3591
- Genu, 11
- Genuineness, 2975. *See also* Client-centered psychotherapy
- Geodon. *See* Ziprasidone
- George, Mark, 1660
- Gerard, Ralph, 4648, 4661
- Gergen, Thomas, 970
- Geriatric bipolar disorder
  - course and outcome, 4235–4236
  - electroconvulsive therapy for, 4243
  - evaluation and treatment of, 4242
  - lithium for, 4242–4243
  - pharmacotherapy for, 4242–4243
- Geriatric depression
  - care delivery, 4241–4242
  - complementarity and alternative medicines for, 4239
  - continuation and maintenance treatment, 4241
  - course and outcome, 4235–4236
  - electroconvulsive therapy for, 4239–4240
  - evaluation and treatment of, 4236–4243
  - MAO inhibitors, 4239
  - pharmacotherapy for, 4237–4239
  - psychotherapy for, 4237
  - selective serotonin reuptake inhibitors, 4237–4238

    - serotonin-norepinephrine reuptake inhibitors, 4238
    - somatic treatments, 4239–4240
    - transcranial magnetic stimulation for, 4240
    - tricyclic antidepressants, 4238
    - vagus nerve stimulation for, 4240
- Geriatric Depression Scale (GDS), 1044*t*, 4136, 4139
- Geriatric psychiatry, **4119–4489**. *See also* Aging
  - aging brain, **4148–4158**
  - anxiety disorders, **4216–4227**
    - agoraphobia, 4218
    - characteristics of, 4223*t*
    - clinical features, 4218–4219
    - comorbidity wit mood disorders, 4217–4218
    - coronavirus pandemic and anxiety in late life, 4219
    - course and prognosis, 4222–4224
    - dementia and, 4223–4224
    - diagnosis, 4218–4219
    - differential diagnosis, 4222, 4223*t*
    - epidemiology, 4216–4217
    - etiology, 4218
    - fear of falling, 4218–4219
    - generalized anxiety disorder, 4221–4222
    - genetic factors, 4222
    - hoarding, 4219
    - incidence, 4217
    - pathology and laboratory examination, 4219–4222
    - prevalence, 4216–4217
      - with special groups, 217, 4217*t*
    - PTSD and, 4219–4221, 4220*f*
    - risk factors, 4218
    - treatment, 4224–4227
      - pharmacotherapy, 4224–4225

    - psychotherapy, 4225–4226
    - transcranial magnetic stimulation, 4226–4227
- assessment, **4131–4137**
  - biomarkers for dementia diagnosis, 4136
  - competence, 4136
  - daily functioning, 4135
  - history taking, 4131–4133
    - alcohol and substance abuse, 4133
    - family interview, 4132–4133
    - medical history, 4133
    - personal history, 4133
    - present illness, 4131–4132
    - psychiatric history, 4132
  - medical examination, 4133–4135
  - mental status examination, 4134–4135
    - cognitive testing, 4135
    - homicidal ideation, 4134–4135
    - insight, 4135
    - language and thought disorder, 4134
    - mood and affect, 4134
    - observation of behavior, 4134
    - perception, 4134
    - sensorium and cognition, 4135
    - sleep, 4135
    - suicidal ideation, 4134–4135
  - neurologic evaluation, 4135
  - rating scales, 4135–4136
  - syndromes specific to geriatric patients, 4136
- cognitive behavioral therapy, 4359–4371
- complementary and integrative, **4137–4148**
  - acupuncture, 4146
  - aromatherapy, 4146

  - Ayurvedic medicine, 4147
  - body-based practices, 4146
  - Chinese or oriental traditional medicine, 4146–4147
  - clinical suggestions, 4141*t*
  - conventional physical activity, 4141–4142
  - definition of, 4137–4138
  - diet, 4142
  - digitization, 4147
  - epidemiology of, 4138–4147
  - future directions, 4147
  - importance of, 4137–4138
  - mind-body therapies, 4139–4141
  - natural products, 4142–4146, 4145*t*, 4146*t*
  - persons with mental illnesses, 4139
  - types of, 4138*f*
- defined, 4733
- dementia (*See* Antidementia drugs)
- demographics
  - of aging, 4119–4120, 4120*t*
  - mental illness, 4121–4125
  - normal aging, 4120–4121
  - positive side of aging, 4129
- elder abuse (*See* Elder abuse)
- electroconvulsive therapy, **4326–4340**
- ethical issues in, **4423–4426**
  - autonomy, 4424–4426
  - core principles, 4424
  - decision-making capacity, 4424–4426
  - informed consent, 4424–4425
- everyday functioning, **4178–4194**
- family intervention and therapy, **4372–4380**
- family role (*See* Family interventions)

- finances (*See* Finances)
- financial issues in, 4392–4402
  - accountable care organizations, 4401–4402
  - current and future directions, 4402
  - demographics, 4393–4394
  - healthcare costs, 4393–4394
  - historical background, 4394
  - Medicaid system, 4398–4399
  - Medicare system, 4394–4398
  - out-of-pocket costs, 4399–4400
  - Patient Protection and Affordable Care Act, 4400–4401
- gender issues, **4427–4432**
  - depression, 4428
  - major neurocognitive disorders, 4427–4428
  - pharmacodynamics, 4431
  - pharmacokinetics, 4430–4431
  - pharmacologic interventions, 4430–4431
  - psychosis, 4428–4429, 4429–4431
  - psychotherapeutic interventions, 4429–4430
  - treatment issues, 4429–4431
- group therapy (*See* Group therapy)
- HIV and aging, **4458–4463**
- individual psychotherapy, 4349–4358
  - brief (time-limited) psychodynamic psychotherapy, 4353
  - cognitive behavioral therapy, 4353
  - complications, 4356–4357
  - contraindications, 4357
  - ethical issues in, 4357
  - future directions, 4358
  - general techniques, 4351–4352
  - history, 4350–4351
  - indications for, 4355

  - insight-oriented psychotherapy, 4353–4354
  - interpersonal psychotherapy, 4352
  - life review or reminiscence therapy, 4352–4353
  - limitations, 4355–4356
  - problem-solving therapy, 4353
  - supportive psychotherapy, 4352
  - theoretical issues, 4351
  - treatment goals, 4357
- medically ill patients, **4196–4209**
  - Alzheimer disease, 4202–4203
  - approach to, 4196–4197
  - chronic obstructive respiratory disease, 4203–4204
  - chronic pain, 4204–4205
  - falling, 4206–4207
  - frailty, 4207
  - health systems, 4208–4209
  - key concepts in, 4196*t*
  - musculoskeletal disorders, 4204–4205
  - neuropsychiatric sequelae of COVID-19 syndrome, 4205–4206
  - overreliance on drugs, 4207–4208
  - Parkinson disease, 4202–4203
  - psychiatric syndromes in, 4197–4200
    - differential diagnosis, 4198–4199
    - drug and disease interactions, 4198, 4198*t*
    - evaluation of, 4199*t*
    - medications as cause of, 4197–4198
    - treatment of, 4200
  - urinary incontinence, 4207
  - urinary tract infections, 4207
  - vascular disease, 4200–4201
  - vision and hearing impairment, 4206
- mental illness in older people, 4127–4129

  - Alzheimer disease (AD), 4127
  - anxiety disorders, 4127
  - dementia, 4127
  - depression, 4127–4128
  - psychotic disorders, 4128
  - schizophrenia, 4128
  - substance use disorders and, 4128–4129
- mood disorders, **4227–4244**
  - bereavement, 4231
  - biologic dysfunction, 4232–4235
  - brain structure and function studies, 4235
  - course and outcome, 4235–4236
  - depression-executive dysfunction, 4232
  - epidemiology, 4228
  - genetic studies, 4234
  - inflammation hypothesis, 4232–4233
  - late-life depression, 4229–4231
  - late-onset hypothesis, 4232
  - suicide, 4231
  - treatment of, 4236–4244
  - vascular depression hypothesis, 4233–4234
- neurobiology of sexual function, 4439–4440
- neurostimulation therapies, 4338–4340
- normal aging, 4158–4166
- older adult abuse, **4432–4438**
- positive psychiatry, **4471–4476**
  - biology of, 4474–4475
  - biomarkers of positive mental health, 4474
    - allostatic load, 4474
    - genetics and epigenetics, 4474
    - inflammation and immune function, 4474
    - oxidative stress, 4474

    - telomere length, 4474
  - interventions, 4475–4476
    - biologic treatments, 4476
    - diet and nutrition, 4475–4476
    - lifestyle, 4475
    - physical activity, 4475
  - low perceived stress, 4472
  - neuroplasticity, 4474–4475
  - positive environmental factors, 4474
    - family dynamics, 4474
    - social engagement and support, 4474
  - positive mental health outcomes, 4472–4473
  - positive psychosocial factors, 4473–4474
  - positive psychosocial traits, 4472–4473
    - coping self-efficacy, 4473
    - meaning in life, 4473
    - optimism, 4473
    - personal mastery, 4473
    - religiosity, 4473–4474
    - resilience, 4473
    - spirituality, 4473
    - wisdom, 4473
  - post-traumatic growth, 4472
  - prevention, 4472–4473
  - recovery from serious mental illness, 4472
  - successful psychosocial aging, 4472
  - well-being, 4472
- sexuality and aging, 4438–4457
  - age-related changes in sexual function and behavior, 4440–4443, 4440*t*
  - cognitive capacity and sexual behaviors, 4454
  - frequency of sexual expressions in, 4441–4443

    - LGBTQ+, 4443–4445, 4453–4454
    - recommendations for PRC communities, 4454–4455
    - sexual behavior in long-term care, 4452–4455
    - sexual dysfunctions
        - arousal phase problems, 4448
        - assessment, 4446–4447
        - desire phase problems, 4447–4448
        - illnesses associated with, 4448, 4449–4451, 4449*t*
        - medications associated with, 4451–4452
        - orgasm phase problems, 4448
        - treatment, 4447–4452
    - sexual history, 4445–4446
- sleep disorders, **4210–4215**
- as subspecialty, 4125–4127
- technology for seniors, 4464–4470
    - aging in place and, 4465
    - barriers and limitations, 4476
        - age-related impairments, 4476
        - billing, 4476
        - ethics, 4476
        - information overload, 4476
        - integration with health care system, 4476
        - lack of regulation and evidence, 4476
        - literacy, 4476
        - privacy, 4476
        - usability, 4476
    - classification, 4464*t*
    - comparative nosology, 4464–4465
    - future directions, 4469–4470
        - big data, 4469–4470
        - nanotechnology, 4469
        - robotics, 4469

    - specific applications of, 4465–4471
        - caregiver applications, 4469
        - games, 4469
        - mobile health, 4465–4467
        - sensors/self-monitoring devices, 4467, 4467*t*
        - smart homes, 4468, 4468*t*
        - social media, 4469
        - tablets, 4469
        - telehealth, 4465
- Geriatric psychopharmacology, **3438**
    - absorption, 4292
    - antianxiety drugs, 4302–4306
    - antidepressants, 4296–4301
    - distribution, 4292
    - excretion, 4293–4294
    - future directions, 4295
    - general principles, 4291
    - metabolism, 4292–4293
    - pharmacodynamics, 4294–4295
    - pharmacokinetics, 4291–4294
- Geriatrics, 4733
- Gerontology, 4733
- Gerontophilia, 4733
- Gerontophobia, 4733
- Geropsychiatry. *See* Geriatric psychiatry
- Gerstmann–Sträussler–Scheinker disease, 631, 638, 4733
- Geschwind, Norman, 458
- Gesell Developmental Schedules, 4733–4734
- Gestalt psychology, 4734
- Gestalt therapy, 4734
- Ghrelin, **169**
- Gibbs, Fred, 369

- Gill, Merton, 1930
- Gilles de la Tourette, George, 3774, 3775
- Gilles de la Tourette syndrome. *See* Tourette disorder (TD)
- Gillham, Jane, 2898
- Gillin, J. Christian, 1656
- Ginkgo (*Ginkgo biloba*), 2637, 4144, 4321
- Ginseng (*Panax ginseng*), 2637
  - for sexual dysfunctions, 2054
- Glaucoma, cannabinoid for, 1346
- Glia-derived neurotrophic factor (GDNF), 1771
- Glial cells, 4
  - location and function of, *5t*
- Glial fibrillary acidic protein (GFAP), 538–539
- Global aphasia, 1020–1021, 1020*t*, 4734
- Global Assessment of Functioning Scale (GAF Scale), 4734
- Global Assessment Scale (GAS), 885
- Global developmental delay
  - defined, 4734
  - diagnosis of, 4774*t*
- Globalization, 3019–3020
- Globus hystericus, 4734
- Globus pallidus, 28
- Glossodynia, 2440
- Glossolalia, 4734
- Glucagon-like peptide (GLP-1), **169**
  - intestinal nutrient sensing, 225
- Glucose transporters (GLUTs), 1507
- Glutamate, 184–185
  - agonists, 1542
  - anxiety disorders and, 1801–1802
  - molecular imaging of, 1544
- Glutamate receptors, 2999*f*

- Glutamatergic modulators, for acute bipolar depression, 1735
- Glutamatergic neurotransmission, 86–87
- Glutamic acid, 83–87
  - astrocytes and, 86
  - excitotoxicity, 87
  - glutamatergic pathways in the brain, 84
  - ionotropic glutamate receptors, 84–86
  - metabotropic glutamate receptors, 86
  - plasticity and glutamatergic neurotransmission, 86–87
  - postsynaptic density, 86
  - synthesis, 83–84
- Glutamine, schizophrenia pathology, 1506
- Glutethimide (Doriden), 1426*t*, 3236
- Glycine, 2635
  - as neurotransmitter, 90
- Glycine encephalopathy, 696
- Glycine modulatory site (GMS), 85
- Glycogen synthase kinase 3, 153
- Glycolysis, 1507
- Glycosylated hemoglobin, 1064
- GM2 gangliosidosis (Tay–Sachs disease), 688–689, 688*f*
- Gmelin, Eberhardt, 1927
- Go-around, 4734
- GOAT. *See* Galveston Orientation and Amnesia Test (GOAT)
- God, 2677–2678
- Goddard, Henry, 3631
- Godding, W. W., 4511
- Godelius, Eric, 4644
- God spot, 2676
- Goldberg, Lewis R., 975–976
- Golden bouquet (*Galphimia glauca*), 2636–2637
- Goldmann/Hodgkin/Katz (GHK) equation, 69

- Goldstein, Kurt, 486
- Gómez, Jennifer M., 1936
- Gonadal dysgenesis. *See* Turner syndrome
- Gonadotropin-releasing hormone (GnRH), 2054, 2098, 2581, 2583
- Gonadotropin-releasing hormone (GnRH) analogs, for paraphilic disorders, 2138
- Gonorrhea, 1053
- Gooch, Robert, 4685
- Good, Byron, 2553
- Goodall, Jane, 2514
- Good Behavior Game (GBG), 4044
- Goodman, Paul, 2957
- Goodwin, Frederick, 1657, 1658, 4697
- Gossen, Manfred, 815
- Gottlieb, Jacques, 4653
- Goulding, Mary, 2957
- Goulding, Robert, 2957
- “Gourmand” syndrome, 465
- G protein–activated ion channels, 3004, 3005*f*
- G protein–coupled receptors (GPCRs), 79, 1401, 2518, 2997, 2997*f*, 2999*f*, 3000
  - internalization and biased agonism, 3001*f*
  - model of, 3000*f*
- G-protein-coupled receptor signaling, 147–151, 147*f*
  - cyclic adenosine monophosphate (cAMP) pathway, 147–148
  - cyclic guanosine monophosphate (cGMP) pathway, 149, 149*f*
  - ion channel regulation, 150
  - phosphatases and, 151
  - phosphatidylinositol pathway, 148–149, 148*f*
  - regulation of, 150–151, 150*f*
- G protein effectors, 3003, 3003*f*
- G proteins, 3002

- Gradient descent, 320–321
- Graduate psychiatric education, **4586–4598**
  - ABPN and, 4588–4589
  - ACGME establishment and, 4588
  - Flexner report influence, 4587–4588
  - history
    - current, 4591–4598
    - early, 4587–4590
  - liaison committee on, 4590
  - major organizations affecting, 4587*t*
  - medical education relation to, 4598
  - Next Accreditation System, 4596–4598
  - program requirements, 4592–4593
    - clinical experiences, 4592–4593, 4593*t*
    - didactic curriculum, 4593
    - faculty, 4592
    - institutionalism, 4592
    - program director, 4592
    - resident evaluation, 4593
  - psychiatry milestones, 4593–4596
    - authors of, 4594*t*
    - description, 4594*t*
    - examples of, 4595*f*, 4596*f*
    - format of, 4595*f*
    - subcompetencies, 4597*t*
    - subspecialty authors, 4598*t*
  - residency training assessment, 4596–4598
  - residential function funding, 4592
  - resident recruitment, 4591
- Graefenberg, Ernst, 2029
- Graham, Billy, 2682
- Graham, John R., 1041

- Grandiose delusion. *See* Delusions
- Grandiose self, 926
- Grandiose thought, 996*t*
- Grandiosity, 1677, 4734
- Grand mal epilepsy, 4734
- Granulomatous neuropathies, 659–661
- Gray matter, 3597, 3603
- Gray-out syndrome, 4734
- Greatest Generation, 4444
- Green, Richard, 2062, 2104
- Greenberg, Joanne, 2790–2791
- Green's Word Memory Test (WMT), 4551
- Gregory, John, 4617
- Grief, 1670, 4734. *See also* Bereavement
  - in burn injuries, 2474
- Grief therapy, family-focused, 2399
- Griesinger, Wilhelm, 1239, 4644, 4649, 4650, 4685
- Grinker, Roy, 787, 4652
- Groddeck, George, 2272
- Grof, Stanislav, 432
- Groleau, Danielle, 2554
- Group A beta-hemolytic streptococcus (GAS), 625–626
  - clinical presentation, 625–626
  - diagnosis, 626
  - epidemiology of, 625
  - management of, 626
  - pathophysiology of, 625
- Group analytic psychotherapy, 4734
- Group cohesion, 4734
- Group dynamics, 4734
- Group mind, 4734
- Group practice, 4734

- Group pressure, 4734
- Group process. *See* Group dynamics
- *Group Psychology and Analysis of the Ego* (Freud), 1705
- Group psychotherapy, **2954–2967**
  - anxiety disorders, 2965–2966
  - attendance rules, 2960–2961
  - billing procedures, 2961
  - checklist, 2960–2962
  - child/adolescent, **3895–3907**
    - choice of treatment, 3907
    - clinical issues, 3900–3907
    - complications, 3907
    - developmental factors, 3897
    - ethical issues, 3907
    - history, 3895–3896
    - indications, 3900–3906
    - limitations of, 3906
    - research and evaluation, 3907
    - techniques, 3897–3900
      - co-therapist relationship, 3899
      - family involvement, 3900
      - group members, 3898
      - group rules, 3899–3900
      - length of intervention, 3899
      - parents, 3898
      - patient exclusion, 3898
      - patient selection, 3897–3898
      - process in structured group, 3900
      - setting, 3898, 3898–3899
      - tools, 3899
    - theoretical issues, 3896–3897
    - training, 3906–3907

- clinical applications, 2964–2967
- confidentiality, 2962
- contract modifications, 2962
- coordination of other therapies, 2961
- cotherapy, 2964
- defined, 4734
- definition, 2955
- development stages, 2962–2963
  - adjourning/separation, 2963
  - norming/intimacy, 2963
  - performing/differentiation, 2963
  - power and control, 2963
  - preaffiliation, 2962–2963
- for dissociative amnesia, 2012
- extragroup socialization, 2961–2962
- fees, 2961
- future directions, 2967
- Gestalt approach, 2957
- goals, 2960
- historical roots, 2955–2956
- history, 4648
- individual therapy and, 2964–2965
- leadership issues, 2963–2964
- leader's role, 2960
- members
  - loss and addition of, 2960
  - selection of, 2958
- mood disorders, 2965
- observation of, 2960
- personality disorders, 2965
- phases of, 2955
- pregroup preparation, 2958–2959

  - psychotic disorders, 2966–2967, 2966*t*
  - purpose, 2960
  - questions and answers, 2962
  - recording, 2960
  - redecision, 2957
  - research issues, 2967
  - schizophrenia, 2966–2967, 2966*t*
  - self-help, 2958
  - setting arrangement, 2960
  - systems-oriented approach, 2957
  - theoretical orientations, 2956–2957, 2957*t*
  - therapeutic factors, 2959*t*
  - therapeutic potential, 2958
  - therapeutic power of group in, 3896–3897
  - time factors, 2960
  - transactional analysis, 2957
- Group relations theory. *See* Allport's group relations theory
- Group therapy, **4380–4388**
  - for adolescents, 3952
  - clinical issues, 4385–4386
  - contraindications, 4386
  - countertransference, 4385
  - for dissociative identity disorder, 1999
  - in end-of-life care, 2399
  - ethical issues in, 4386–4387
  - future directions, 4388
  - goals, 4386
  - history, 4380–4381
  - indications for, 4385–4386
  - interpersonal psychotherapy, 2836–2838
    - initial phase, 2837
    - intermediate phase, 2837

    - interventions, 2838
    - summaries, 2838
    - techniques, 2836–2837
    - termination phase, 2837–2838
    - therapeutic stance, 2837
    - timeline, 2836–2837
  - research and evaluation, 4387–4388
  - for sexual dysfunctions, 2051–2052
  - technical interventions, 4382–4385
    - activity and creativity, 4385
    - caregivers, 4384–4385
    - cognitive behavioral, 4383–4384
    - cognitive impairment, 4385
    - existential, 4384
    - life-review, 4384
  - theoretical issues, 4381–4382
- Growth cone, 64, 65*f*
- Growth factors, receptors for, 2998–3000
- Growth hormone (GH), **167**
  - mood disorders and, 1772
  - sleep and, 204–205
- Growth-hormone secreting tumors, 711
- Growth mixture modeling, 909–910
- Gruenberg, Ernest, 1481
- Gruhle, Hans, 4649
- Gruzelier, John, 2918
- G spot, 2029
- Guanfacine (Intuniv, Tenex), 99, **3219–3224**
  - absorption, 3220
  - for ADHD, 3221–3222
  - anxiety disorders, 3222
  - autism spectrum disorder, 3222

  - chemistry, 3219
  - for child/adolescent PTSD, 3849
  - cognition and neuroprotective effects of, 3222–3223
  - distribution, 3220
  - dosage and administration, 3223–3224
  - drug interactions, 3223
  - drug studies, 3220–3221
  - elimination, 3220
  - history, 3219
  - mechanism of action, 3220
  - metabolism, 3220
  - nicotine withdrawal, 3222
  - opiate withdrawal, 3222
  - organ/system effects of, 3221
  - pharmacodynamics, 3220
  - pharmacokinetics, 3219–3220
  - pharmacologic actions, 3220
  - precautions and adverse reactions, 3223
  - PSTD, 3222
  - therapeutic indications, 3221–3223
  - for tic disorders, 3221, 3784
  - for Tourette syndrome, 3221
- Guardian, 791
- Guardian angels, 2682
- Guardianship, 4734
- Guarnaccia, Peter, 2556
- Gudjonsson Suggestibility Scale (GSS), 1969, 2940
- Guided imagery, 2171
- Guided imagery therapy, 2984–2985
  - clinical issues, 2985
  - ethical issues in, 2985
  - history, 2984

  - techniques, 2984–2985
  - theoretical issues, 2984
- Guided mastery, 1828
- Guilford, J. P., 975
- Guillain–Barré syndrome, 608
- Guillain–Barré syndrome (GBS), 657
- Guillemin, Roger, 116
- Guilt, 938, 1703
  - bereavement depression and, 1695*t*
  - defined, 4734
- Gull, William, 2149
- Gunderson, John, 4652
- Guntrip, Harry, 1930
- Gustatory hallucination, 4734
- Gustatory hallucinations, 1455
- Gut-brain axis, 2416
  - schematic representation of, 2300*f*
- Guze, Samuel, 1201
- Gynecologic cancers, 2364
- Gynecologic disorders, neurosteroids and, 139
- Gynecology, hypnosis in, 2931
- Gynecomastia, 4734
- Gynephilic orientation, 2058
- Gynephobia, 4734
- Gyrectomy, 4734
- Gyromagnetic ratio, 327

H

- Habeas corpus, 4734
- Habit, 4734
- Habit cough, 2324
- Habit learning, 766

- Habit-reversal therapy
  - for tic disorders, 3783
  - for trichotillomania, 1852
- Habit reversal training (HRT), 2818
  - for trichotillomania, 2435
- Habit training, 947
- Habitual use, 1316
- Habituation, 4734
- HADS. *See* Hospital Anxiety and Depression Scale (HADS)
- *Halderman v. Pennhurst State School & Hospital* (1977), 4513
- Haldol. *See* Haloperidol
- Haley, Jay, 2943, 2947, 2981
- Half-life, 386
- Half-sibling studies, ADHD, 3729
- Halfway house, 4734
- Hall, Stuart, 2877
- Hallevorden–Spatz syndrome. *See* Neurodegeneration with brain iron (NBIA)
- Hallopeau, François, 1850
- Hallucinations, 996*t*, 1172–1175
  - auditory, 1173, 1453–1454
  - autoscopic, 1174
  - brain diseases and, 476
  - command, 1173–1174
  - command auditory verbal, 1454
  - defined, 4735
  - dissociative, 1979–1980
  - gustatory, 1455
  - haptic, 1174
  - hypnagogic, 1173
  - hypnopompic, 1173
  - hypnosis and, 2917

  - ictal, 1174
  - learning theory and, 437–438
  - negative, 1980
  - olfactory, 1174, 1455
  - pseudohallucination, 1174
  - in schizophrenia, 320–321, 1452–1455
  - sleep-related, 2185
  - tactile, 1455
  - visceral, 1455
  - visual, 1174, 1454–1455
- Hallucinogen-assisted psychotherapy, 2903, **2903–2911**
  - clinical issues, 2905–2906
  - complications, 2906
  - contraindications, 2906
  - definition of, 2903
  - history, 2904
  - indications for, 2905
  - ketamine, 2909–2910
  - limitations of, 2905
  - MDMA, 2908–2909
  - for OCD, 2907
  - for patients with life-threatening illness, 2906–2907
  - physiologic toxicity, 2906
  - principles and goals of, 2905
  - psychedelic therapy, 2905
  - psychological toxicity, 2906
  - psycholytic therapy, 2905
  - psychopharmacology, 2903–2904
  - for treatment of addiction, 2907
  - for treatment-resistant depression, 2906
- Hallucinogen-persisting perception disorder, 1360–1361, 4827*t*
- Hallucinogen-related disorders, **1347–1364**

  - comparative nosology, 1355–1356
  - hallucinogen-persisting perception disorder, 1360–1361
  - phencyclidine- and other hallucinogen-induced anxiety disorder, 1364
  - phencyclidine- and other hallucinogen-induced bipolar disorder, 1363
  - phencyclidine- and other hallucinogen-induced depressive disorder, 1363–1364
  - phencyclidine- and other hallucinogen-induced intoxication delirium, 1364
  - phencyclidine- and other hallucinogen-induced psychotic disorder, 1362–1363
  - phencyclidine and other hallucinogen intoxication, 1356–1359
  - phencyclidine use disorder and other hallucinogen use disorder, 1359–1360
  - prevalence of, 1352*t*
  - unspecified phencyclidine-related disorder and unspecified hallucinogen-related disorder, 1364
- Hallucinogens, **1347–1364**. *See also* Psychedelics
  - categories, 1348
  - classes of, 1349*t*
    - anticholinergic deliriants, 1355
    - iboga alkaloids, 1355
    - kappa-opioid agonists, 1355
    - NMDA receptor antagonists, 1351–1353
    - serotonergic hallucinogens, 1353–1355
  - classic, 1353–1354, 1358
  - defined, 4735
  - EEG effects, 380
  - epidemiologic overview, 1351
  - historical overview, 1348
  - intoxication, 2728
  - mechanism of, 2904

  - psychopharmacology, 2903–2904
  - sexual dysfunction and, 2047
  - subjective effects, 2904
- Hallucinosis, 1175, 4735
- Haloperidol (Haldol), 166
  - for acute mania, 1729
  - for agitation, 2724, 3576*t*
  - defined, 4735
  - for delirium, 1237
    - in cancer patients, 2352
  - history, 3007, 3034, 4692
  - pharmacokinetics of, sex differences in, 3417
  - for stereotypic movement disorders, 3813
  - for tic disorders, 3785
- Halstead, William, 4166–4167
- Halstead–Reitan Scale, 4735
- Ham, George, 4653
- Hamilton, Max, 3152
- Hamilton Anxiety Rating Scale (HARS), 1106, 3545
- Hamilton Rating Scale for Depression (HRSD), 1104, 3150, 3544, 4136, 4735
- Hampstead Index, 1703
- Handedness, 465
- Handoffs, 1155–1156
- Hansen, Helena, 1138
- Haphephobia, 4735
- Haplology, 4735
- Haplotypes, 243
- HapMap Project, 250*t*, 1483–1485
- Happiness, empirical findings, 2890–2893
- *The Happy Life—Voyages to Well Being* (Cloninger), 2685
- Haptic hallucinations, 1174, 4735

- Hare, Robert, 2519
- Hare's Psychopathy Checklist-Revised, 3546
- Harlow, Harry, 798
- Harm avoidance, 2219–2222, 4735
  - psychological findings, 2221
  - stimulus–response patterns, 2220*t*
- Harmon, Robert, 1959
- Harris, Sam, 2683
- Harrison Narcotics Act, 1312, 4512
- HARS. *See* Hamilton Anxiety Rating Scale (HARS)
- Hartland, John, 2922
- Hartley, David, 434
- Hartmann, Heinz, 422, 919
- Harvard Group Scale of Hypnotic Susceptibility (HGS), 2920
- Harvard University, 4652
- Hashimoto encephalopathy, 713–714
- Hashish, 4646, 4735
- Haslam, John, 4643, 4645
- Hathaway, Starke R., 4659
- Hawthorne, James, 4587
- Hawthorne effect, 2737
- Hayes, Steven, 2819, 2898
- Hazard rate, 910
- HCN channels, 75
- Head, physical examination, 466
- Headache(s). *See also* Migraine
  - attributable to head or neck trauma, 647
  - caffeine withdrawal, 1333
  - chronic daily, 2483–2484
  - clinical evaluation, 649–650
  - cluster, 646–647, 652
  - comorbidity, 648

  - definitions of, 645
  - differential diagnosis, 649–650
  - epidemiology and course of, 647–649
  - etiology of, 649
  - hypnosis for, 2936
  - migraine, 645–646
  - neuropsychiatric aspects of, **645–652**
  - post-traumatic, 652
  - prevalence of, 647–648
  - secondary causes of, 645*t*
  - tension-type, 646, 646*t*, 651–652
  - treatment of, 650–652
- Head and neck cancers, 2362
- Head injury, 464. *See also* Traumatic brain injury (TBI)
- Head trauma, 1057
  - headache attributable to, 647
- Healing After Trauma Skills (HATS), 4070
- Healing touch, 2648
- Health
  - definition of, 2894
  - strength-vulnerability model, 4373
- Health 2.0, 2701–2703
- Health and Human Services, 4735
- Health care proxy, 2401
- Health care reform, **4491–4503**
  - in the 21st century, 4495–4496
  - Affordable Care Act and, 4496–4498
  - care management, 4493–4494
  - Commonwealth Fund report, 4491–4492
  - electronic health records, 4501–4503
  - goals of, 4491
  - health information technology, 4501–4503

  - historical overview, 4492–4495
  - managed behavioral care, 4494–4495
  - managed care, 4493
  - Mental Health Parity and Addiction Equity Act and, 4498
  - models in, 4499–4501
    - accountable care organizations, 4500
    - behavioral health care services for children with serious emotional disorders, 4501
    - comparative effectiveness research, 4500
    - medical neighborhoods, 4498–4499
    - patient-centered medical homes, 4498–4499
    - patient-centered outcomes research, 4500
    - pay for performance, 4500–4501
    - public reporting of performance, 4500–4501
    - teaching health centers, 4500
  - selective closed panels and networks, 4493
  - Triple Aim and, 4496–4498
  - utilization review, 4493
  - waster in health care and, 4495–4496, 4496*t*
- Health care workers
  - impact of COVID-19 on, 2686–2687
  - reactions to COVID-19 pandemic, 2385–2386, 2385*t*
- Health information technology (HIT), 4501–4503
- Health insurance, 4735. *See also* Affordable Care Act (ACA); Health maintenance organizations (HMOs); Medicaid; Medicare
- Health Insurance Portability and Accountability Act (HIPAA), 1121–1122, 1157–1158, 2700–2701, 4036, 4036*t*
- Health maintenance organizations (HMOs), 4480–4481, 4735
  - chronic care initiatives in, 4498
- Health-related quality of life (HRQoL), 4180
- Health Systems Agency (HSA), 4735
- Healthy identification, 4735

- Healy, David, 4650, 4663
- Hearing voices network (HVN), 1609
- Hearsay, 4735
- Hearst, Patty, 2020
- Heart. *See also* Cardiovascular system
  - physical examination, 466
  - transplantation, 2461–2462
- Heart disease. *See also* Cardiovascular disorders, child/adolescent response to, 3989–3990
- Heart failure, 2282–2284. *See also* Cardiovascular disorders
  - classification, 2282*t*
  - depression and, 2283
  - epidemiology of, 2282–2283
  - heart transplant for, 2283
  - psychiatric evaluation for transplant or MCS, 2283–2284
  - psychological factors affecting, 2283
  - surgical management of, 2283
  - symptoms treatment, 2282–2283
  - ventricular assist devices, 2283
- Heart rate, cannabis use and, 1342
- Heart rate variability (HRV) biofeeback, 2644–2645
- Heart transplants, 2283
- Heath, Robert, 4654
- Hebb, Donald, 730t
- Hebbian plasticity, 81
- Hebephiles, 2095
- Hebephrenia, 4657, 4735
- Hebrew Home at Riverdale, 4454
- Hecker, Ewald, 4657, 4686
- Hedonic allostasis theory, 444
- Hedonism, 2893, 4735
- Heiddeger, Martin, 418

- Heinrichs- Carpenter Quality of Life Scale (QLS), 4553
- Heinroth, Johann Christian, 1634, 2271, 4685
- Helminths, 628
- Helms, Janet, 848
- HELP. *See* Hospital Elder Life Program (HELP)
- Helping Older People Experience Success (HOPES) program, 4243
- Helplessness, 4735
- Hematologic malignancies, 2364–2365
- Hematophobia, 4735
- Hemiagnosia, 1175
- Hemiasomatognosia, 1175
- Hemineglect, 472
- Hemispheric dominance, 1020–1023, 1020*t*
- Hemophilia, child/adolescent response to, 3990
- Henninger, George, 1656
- Hepatic encephalopathy, 2460–2461
- Hepatic system
  - esketamine effects on, 3383
  - ketamine effects on, 3383
- Hepatitis B, 1054
- Hepatitis C virus (HCV), 608–609, 1054
- Herbal formulas, 2640
- Herd instinct, 4735
- Herdt, George, 2072
- Heredity, 1659
- Heritability, 2071, 3587*t*
- Hermaphrodite, 4735
- Hermaphroditism, 2068, 2101
- Heroin, 4735. *See also* Opioid use
  - for opioid use disorders, 1418
  - structure of, 1400*f*

- Herpes simplex virus encephalitis (HSVE), 618–619
  - clinical presentations of, 618
  - diagnosis, 618
  - management of, 618–619
  - pathophysiology of, 618
- Herpes zoster neuropathy, 660
- Heterochromatin, 236
- Heterocyclic antidepressants (HCAs), anxiety disorders, 1834
- Heterogeneity, 4735
- Heteronuclear single quantum coherence spectroscopy (HSQC), 346
- Heteroreceptors, 93
- Heterosexism, 2076
- Heterosexual-Homosexual Rating Scale (Kinsey Scale), 2058
- Heterosexual identity, 2059
- Heterosexuality, 4735
  - definition of, 2058
- Heterotrimeric G proteins, 3002, 3002*f*, 3002*t*
- Heterozygous, 4735
- Heterozygous deletions, 242
- Heuristic, 4735
- Hexacarbons, impact on CNS and PNS, 655*t*, 659
- Hexobarbital, 3232
- HHV-6, 619
- Hi-C, 3587*t*
- Hidden observer phenomenon, 1178
- Hierarchical Taxonomy of Psychopathology (HiTOP) model, 1961
- High-density negative ions, 2644
- High expressed emotion level within family, 2547
  - clinical features, 2547
  - definition of, 2547
  - diagnosis, 2547

  - etiology of, 2547
- High-pass filter, 361
- High-risk behavior. *See* Risk-taking behavior
- Hijras, 2102
- Hilgard, Ernest, 1931, 2912, 2914. *See also* Hypnosis
- Hill, Bradford, 4659
- Hillside Hospital, 4651
- Hinduism, 2680
- Hinting Test, 4551
- HIPAA. *See* Health Insurance Portability and Accountability Act (HIPAA)
- HIPAA-covered entities, 1121
- Hippocampal dentate gyrus, 66–67, 66*f*
- Hippocampal formation, 35, 36*f*, 37*f*, 1504–1505
- Hippocampus, 10, 35
  - depressive disorders/bipolar and related disorders, 3607–3608
  - patterning genes, 62
  - role in contextual fear, 1811
- Hippocrates, 521, 1230, 2271, 3630, 4683
- Hip replacement, 1232
- Hirschfeld, Magnus, 2061, 2103
- Histamine H_1 receptors, 1544
- Histamines. *See also* Antihistamines
  - anatomy and functions, 106
  - antihistamines and, 3345–3349
  - life cycle, 106
  - receptors, 106–107
- Histone acetyltransferases (HATs), 454
- Histone deacetylases (HDACs), 454
- Histone(s)
  - acetylation, 272–273, 272*f*
  - methylation, 273–274, 274*f*

  - modifications, 271–274
  - phosphorylation, 274
- History taking, 3523
- Histrionic personality disorder, 2248, 4735, 4836*t*
- HITECH (Health Information Technology for Economic and Clinical Health) Act of 2009, 1121–1122, 1157–1158, 2701, 4501
- HIV (human immunodeficiency virus), 589, 1053. *See also* Acquired immunodeficiency syndrome (AIDS)
  - antidepressants in
    - antidepressants and doses, 594*t*
    - side effects, 593*t*
  - bipolar (manic-depressive) illnesses in, 594–596
  - child/adolescent, 674–678, **3993–3997**
    - behaviorally acquired, 675–677
    - clinical features, 3995
    - epidemiology of, 674–675, 3993–3994
    - etiology of, 3994
    - mental health and, 675
    - mental illness and, 676–677
    - neurocognitive impact of, 677
    - neurodevelopmental sequelae, 3995
    - perinatally acquired, 677
    - psychiatric disorders in, 675
    - substance abuse and, 676–677
    - treatment of, 677–678, 3995–3997
  - CNS neoplasms and, 607–608
  - course of, if left untreated, 586*f*
  - defined, 4736
  - delirium and, 589
  - dementia and, 589–590
  - depression and, 589–594
  - depression in

  - differential diagnosis, 591–592
  - side effects of antidepressant medications, 593*t*
  - treatment of, 592
- description of, 583–585
- epidemiology of, 585–587
- everyday functioning and, 4192–4193
- fatigue in, 608
- in gay men, 2085
- hepatitis C and, 608–609
- medication adherence in, 602
- neurologic complications of, 606–609
- neuropathies associated with, 659–660
- neuropsychiatric aspects of, 583–609
- in older adults, **4458–4463**
  - comorbidity, 4461
  - death rates, 4458
  - epidemiology, 4458
  - everyday functioning, 4461–4462
  - mental health outcomes, 4459–4461
    - emotional functioning, 4459–4460
    - severe mental illness, 4460
    - substance use, 4460
  - multimorbidity, 4461
  - neurocognitive outcomes, 4458–4459
  - prevention, 4458
  - successful aging, 4460–4461
  - treatment and management of, 4462–4463
  - veterans, 4462
- opportunistic infections
  - cryptococcal meningitis, 607
  - cytomegalovirus, 607
  - progressive multifocal leukoencephalopathy, 607

    - toxoplasmosis, 606–607
  - peripheral neuropathy in, 608
  - personality disorders in, 602
  - personality issues in, 600–603
    - dimensional traits and personality, 600
    - personality models, 600–601
    - risk behaviors, 600–602
    - sensation seeking, 601
    - treatment implications of, 602–603
  - prevalence of, 585
  - psychological problems in, 603–606
  - psychotherapeutic issues
    - adherence counseling, 606
    - capacity to consent/competence, 606
    - partner notification, 605–606
    - posttest counseling, 605
    - pretest/test counseling, 604–605
  - PTSD and, 604
  - schizophrenia and, 596
  - substance use disorders and, 596–600
  - viral life cycle and treatment of, 587–589, 587*f*
- HIV-associated dementia, 589–590
  - clinical course, 590
  - epidemiology of, 590
  - neuropathology, 589–590
  - risk factors, 590
  - treatment of, 590
- HMG-CoA reductase inhibitors, 4321–4322
- HMOs. *See* Health maintenance organizations (HMOs)
- Hoarders, 959
- Hoarding disorder, 1841*t*, **1848–1850**, 4735
  - clinical features, 1849–1850

  - comparative nosology, 1849
  - course and prognosis, 1850
  - definition of, 1849
  - diagnosis, 1849–1850, 1849*t*, 4799*t*
  - differential diagnosis, 1850
  - epidemiology of, 1849
  - etiology of, 1849
  - history, 1849
  - in older adults, 4219
  - treatment of, 1850
- Hobbyism, 560
- Hoch, August, 4652
- Hoch, Paul, 4651
- Hoff, Hans, 4663
- Hoffman, Irwin, 3877
- Hoffman sign, 470
- Hofmann, Albert, 431–432, 4662
- Hofstede cultural dimensions, 846–847
- Holding environment, 2661
- Holiday blues, 1670
- Holism, 4735
- Holmes, Thomas, 2273, 2414
- Holm procedure, 900
- Holophrastic, 4735
- Holsboer, Florian, 1657
- Holtzman Inkblot Technique (HIT), 1045*t*
- Holzman, Philip, 4653
- Homburger, August, 4649
- Homeless, 4735
- Homelessness, 1587
  - psychiatric emergencies and, 2732
- Homeopathy, 2648

- Homeostasis, 4735
- Homeostatic plasticity, 81
- Home parenteral nutrition (HPN), 2369
- Homeschooling, 4028
- Home sleep apnea testing (HSAT), 2189
- Homicidal behavior
  - dissociative amnesia for, 2004
  - in older adults, 4134–4135
- Homicide
  - epidemiology of, 2512–2513
  - rates, 2512–2513
  - victims, 2513
- Homocystinuria, 699–700
- Homophobia, 2075, 4735
- Homosexual gene, 2071
- Homosexuality, **2057–2089**, 4100
  - acquisition of gay and lesbian identities, 2073
  - adolescence and, 2078–2079
  - age-structured, 2072
  - alcohol abuse and, 2083–2084
  - antecedents in childhood and adolescence, 2073
  - anthropometric characteristics, 2070
  - antihomosexual attitudes and, 2075–2076
  - biologic models of, 2066
    - direct biologic effects, 2066
    - indirect biologic effects, 2066
  - childhood and, 2078
  - clinical approaches, 2081–2087
    - inclusive and special clinical settings, 2081–2082
    - sexual orientation of psychiatrist, 2082
    - talking about sexual problems, 2084
  - “coming out,” 2058, 2073, 2076–2077

- declassification as mental disorder, 2058, 2062–2064
- defined, 4735
- definitions, 2058–2060
- ethical and legal issues, 2087–2088
- families and, 2080–2081
- gender-reversed, 2072
- genetic research, 2070–2071
  - genome-wide association studies, 2071
  - genome-wide linkage studies, 2071
  - twin studies, 2070–2071
- historical views, 2060–2064
- HIV/AIDS and, 2085
- homosexual orientation constant, 2074*f*
- Internet resources, 2089*t*
- medical illness and, 2084–2085
- medicalization, 2060
- midlife and, 2079–2080
- moral condemnations of, 2076
- neuroanatomical sexual dimorphism, 2068–2069
- neurophysiological measures, 2069
- nongay identities and, 2073–2075
- normalization of, 2062
- in older adults, 2080
- origins of, 2065–2075
  - biologic models, 2066
  - biologic studies, 2065–2072, 2073
  - cross-cultural studies, 2072
  - neuroendocrine research, 2066–2067
  - prenatal hormonal hypothesis, 2067–2068
  - psychosocial studies, 2072–2075
- prevalence of, 2064–2065
- psychotherapy, 2085–2087

  - race/ethnicity and, 2077
  - religion and, 2077–2078
  - role-specialized, 2072
  - sexual orientation conversion efforts, 2086–2087
  - societal/domestic violence and, 2082–2083
  - substance abuse and, 2083–2084
  - suicide and, 2083
  - terminology, 2058
  - theories of, 2060–2064
  - training and research needs on, 2088–2089
  - young adults and, 2079
- Homosexually self-aware individuals, 2074–2075
- Homosexual panic, 4736
- Homotypical cortex, 21
- Homovanillic acid (HVA), 4736
- Homozygous deletions, 242
- Honesty, 2404
- Hooker, Evelyn, 2062
- Hoover, J, Edgar, 2509
- HOPES. *See* Helping Older People Experience Success (HOPES) program
- Hope therapy, 2898
- *Hopkins Verbal Learning Test-Revised* (HVLT-R), 1032
- Hormonal agents, 2098–2099
- Hormone replacement therapy (HRT), 2609–2610
  - pharmacokinetics/pharmacodynamics and, 3413–3414
- Hormone(s)
  - classification, 157*t*, 161
  - evolution, 161
  - modes of action, 158
  - regulating, 158
  - secretion, 161

  - structure, 161
  - synthesis, 161
  - tissue specificity, 158
  - violence and, 2516
- Hormone therapy, for sexual dysfunctions, 2054–2055
- Horn, Ernst, 4649
- Horney, Karen, 956–958, 956*f*, 2831, 4688, 4736
  - neurosis theory, 956–957
    - alienation, 957
    - claims, “shoulds,” and self-hatred, 957
    - idealized image, 957
    - neurotic pride and the pride system, 957
  - personality theory, 956
  - therapeutic process, 957
- Horwitz, Allan V., 1201
- *HORYZONS* system, 1591
- Hospital addicts, 1907
- Hospital Anxiety and Depression Scale (HADS), 2368
- Hospital Elder Life Program (HELP), 1236
- Hospital hoboes, 1907
- Hospitalism, 3500
- Hospitalist physicians, 4534
  - benefits of, 4537–4538
  - as clinician-educators, 4538
  - disadvantages of, 4538–4539
  - financial, political, and organizational issues, 4539
  - future directions, 4540
  - history, 4536–4537
  - professional organizations, 4538–4539
  - roles of, 4534–4535
  - specialty training, 4539
  - stages of hospital care, 4537

- Hospitalization
  - child/adolescent, **3981–3992**
    - at different ages, 3986–3988, 3987*t*
      - adolescents, 3988
      - infants, 3986–3987
      - preschool age, 3987
      - school age, 3987–3988
      - toddlers, 3987
    - historical background, 3981–3982
    - impact on children, 3983–3984
    - impact on families, 3984–3986
    - response at different ages, 3986–3988
    - social media and, 3992
    - transition from pediatric to adult medical care, 3991–3992, 3992*t*
  - collaborative care, 3983
  - integrated care, 3983
  - involuntary, 4608–4609
  - medical home, 3983
  - pediatric hospitals, 3982–3983
- Hospital medicine, 4539–4540
- Hospitals. *See* Psychiatric hospitals
- Hospital utilization review, 1159
- Hostile transference. *See* Transference
- Hostility, 4736
  - in schizophrenia, 1464–1465
- Hot cognition, 726
- Hot deck imputation, 905
- Hotline, 4736
- House-Tree-Person (HTP) test, 3548
- Housing
  - homelessness and, 1587

  - models, 1587–1588
  - supportive, 1587–1588
- Housing first (HF), 1607
- $5HT_1$ receptors, 102–103
- 5-HT2A/C receptors, 3008
- $5HT_2$ receptors, 103
- $5HT_3$ receptors, 103
- $5HT_4$ receptors, 103
- $5HT_5$ receptors, 103–104
- $5HT_6$ receptors, 103–104
- $5HT_7$ receptors, 104
- Huarte de San Juan, Juan, 4684
- Huber, Gerd, 4660
- Hug-Hellmuth, Hermine, 3877
- Hughes, Charles, 2552
- Hughlings-Jackson, John, 458, 460
- Hull, Clark, 2913
- Human chorionic gonadotropin (hCG), 1055, 2582
- Human Connectome Project, 4707
- Human genome, 234
  - common *versus* rare genetic variation, 243
  - organization and structure of, 235–242
  - sequence variation, 242
  - structural variation, 243
  - variation, 242–244
- Human Genome Project, 4017
- Human herpesviruses (HHV), 618–619
- Humanism, 2887
  - in medicine, 2384
- Humanistic psychology, 2887
- Humanistic theory, 4736

- Human nuclear genome, organization of, 44
- Human papillomavirus (HPV), 2595
  - hypnosis for, 2937
- Human Proteome Initiative (HPI), 249
- Human Proteome Organization (HUPO), 249
- Hume, David, 419
- Humiliation, 4736
- Humor, 793, 921*t*, 1000, 2748*t*, 4736
- Hunger, 220
  - circadian rhythm of, 221*f*
  - eating in absence of, 223, 223*t*
- Hunter, Edward, 2019
- Hunter, John, 1230
- Huntington, George, 566
- Huntington disease, 240, 566–571, 1058
  - ancillary tests, 569
  - basal ganglia disruption in, 32
  - clinical features, 568–569
  - course and prognosis, 569–570
  - defined, 4736
  - definition of, 566
  - diagnosis, 568–569
  - differential diagnosis, 569, 570*t*
  - dopamine receptor antagonists for, 3045
  - epidemiology of, 566
  - etiology of, 566–567
  - historical literature on, 566
  - laboratory evaluation of, 569
  - neuropathology, 568*f*
  - pathologic examination of, 569
  - premanifest testing of, 569
  - treatment of, 570–571

- Huntjens, Rafaele, 1979
- Huperzine A, 2639, 4320
- Husserl, Edmund, 418
- HVA. *See* Homovanillic acid (HVA)
- *Hwa-byung*, 2552–2553
- Hybrid therapy, 1121
- Hydrocephalus, 4736
- Hydrocodone, structure of, 1400*f*
- Hydrocortisone, for PTSD, 1884
- Hydrogen sulfide, 130
- Hydromorphone, structure of, 1400*f*
- Hydrophobia, 4736
- Hydrotherapy, 4736
- 5-Hydroxyindoleacetic acid (5-HIAA), 4736
- 5-Hydroxytryptophan, 2633–2634
- Hydroxyzine (Vistaril), 3347
- Hygienic voice therapy, 3709
- Hyperactive hallucinatory delirium, 2396
- Hyperactivity, 4736
- Hyperaldosteronism, 716, 716*f*
- Hyperalgesia, 2480*t*, 4736
- Hyperarousal, 4736
- Hypercholesterolemia, 3073, 3081
- Hypercortisolism, 2338–2340. *See also* Cushing syndrome
  - comparative nosology, 2338
  - definition, 2338
  - diagnosis and clinical features, 2338–2339
  - differential diagnosis, 2339–2340
  - epidemiology of, 2338
  - etiology of, 2338
  - pathology and laboratory examination, 2339
  - treatment and course, 2340

- Hypercortisolism (Cushing syndrome), 2365
- Hyperdynia, 2440
- Hyperesthesia, 2480*t*, 2917, 4736
- Hyperglycemia, 1598, 3058, 3073, 3081
- Hyperkinesis. *See* Hyperactivity
- Hyperkinetic disorders, 3092, 4736. *See also* Attention-deficit hyperactivity disorder (ADHD)
- Hyperlipidemia, 3058
- Hypermnesia, 1182, 2917, 2939, 4736
- Hyperorexia, 4736
- Hyperparathyroidism, 715
- Hyperpathia, 2480*t*
- Hyperphagia, 4736
- Hyperphosphorylation, 1254
- Hyperpolarization, 69
- Hyperpragia, 4736
- Hyperprolactinemia, 99, 2344–2345, 3064, 3074, 3084, 3086
  - clinical features, 2345
  - comparative nosology, 2344
  - definition of, 2344
  - diagnosis, 2345
  - epidemiology of, 2344–2345
  - etiology of, 2344–2345
  - laboratory examination, 2345
  - treatment of, 2345
- Hyperprolinemia (HP), 695
- Hyperschemazia, 1175
- Hypersexuality
  - defined, 4736
  - definition of, 2121
  - manic syndrome and, 1677
  - in patients with dementia, 4456*t*

- Hypersomnia
  - in children, 4088
  - defined, 4736
  - definition of, 1675
  - depressive syndrome and, 1188
- Hypersomnolence disorder, 2172–2174
  - characteristics of, 2173–2174
  - diagnostic criteria, 2172–2173, 4811*t*
  - sleep stage histogram, 2173*f*
  - subtypes, 2173–2174
  - treatment of, 2174
- Hypertension, 1597, 2280–2282
  - anxiety disorders and, 2281
  - behavioral risk factors in, 2281–2282
  - dementia and, 487
  - depression and, 2280–2281
  - epidemiology of, 2280
  - post-traumatic stress disorder and, 2281
  - severe mental illness and, 2282
  - stress factors in, 2281–2282
  - symptoms treatment, 2280
- Hypertensive crisis. *See also* Hypertensive encephalopathy, monoamine oxidase (MOA) inhibitors, 3191, 3191*t*
- Hypertensive encephalopathy, 4736
- Hyperthyroidism, 712–713, 2337–2338
  - comparative nosology, 2337
  - course and prognosis, 2337
  - defined, 4736
  - definition of, 2337
  - diagnosis and clinical features, 2337
  - differential diagnosis, 2337
  - epidemiology of, 2337

  - etiology of, 2337
  - pathology and laboratory examination, 2337
  - treatment of, 2337–2338
- Hypertonia, 3676
- Hypertriglyceridemia, 3081
- Hypertrophic cardiomyopathy (HCM), 2284
- Hyperventilation, 372, 4736
- Hypervigilance, 4736
- Hypesthesia, 4736
- Hypnagogic foot tremor, 2188
- Hypnagogic hallucinations, 1173, 4736
- Hypnic jerks, 2188
  - absorption, 2915
  - involuntariness, 2915
- Hypnoanalysis, 4736
- Hypnodrama, 4736
- Hypnoid state, 4736
- Hypnophobia, 4736
- Hypnopompic hallucinations, 1173, 4736
- Hypnoprojective techniques, 2923
- Hypnosis, 1178, **2911–2942.** *See also* Stanford Hypnotic Susceptibility Scale (SHSS)
  - addictive behaviors and, 2937–2938
  - anxiety disorders, 2928–2930
  - asthma, 2936
  - characteristics, in clinical context, 2917
  - in chronic pain disorders, 2933
  - clinical applications, 2923–2931
  - complex habits, 2938
  - defined, 4737
  - definition of, 2914–2915
  - depression, 2930

- dermatology and, 2937
- dimensions and phenomena of, 2914
- for dissociative amnesia, 2011
- dissociative disorders, 1995
- dissociative state, 2916
- eating disorders, 2938
- electroencephalographic (EEG) studies of, 2918
- in emergency room, 2931–2932
- experiences and behavior, 2914
- in forensic psychiatry, 2938–2942
- in headache management, 2936
- heightened suggestibility model of, 2913
- history, 2911–2912
- hypnotic induction ceremony, 2920–2921
  - eye fixation/closure, 2921
  - eye roll technique, 2921
  - holding on/letting go, 2921
  - progressive relaxation, 2921
  - purpose and preparation, 2920–2921
  - staircase/escalator, 2921
  - visual imagery techniques, 2921
- hypnotizability
  - assessment, 2919–2920
  - trait, 2912–2913
- immune-related disorders, 2936–2937
- indirect, 2922
- irritable bowel syndrome and, 2936
- neurobiology of, 2917–2919
- neuropsychology of, 2917–2919
- for pain alleviation in bone marrow aspirations, 2932
- pain and, 2931–2936
- pain management in burn patients, 2470

  - personality disorders, 2930–2931
  - PTSD, 2924–2926
  - as redistributed attentiveness, 2915
  - response expectancy model, 2916
  - sexual dysfunction, 2937
  - smoking cessation, 2937–2938
  - social influence models for, 2916
  - somatic symptom and related disorders, 2926–2927
  - state/nonstate controversy, 2916–2917
  - suggestibility and, 2913–2914
  - in surgical procedures, 2932
  - task motivational theory, 2916
  - trance characteristics, 2915–2916
  - trauma and, 2924–2926
  - in weight management, 2938
  - women’s issues and, 2931
- Hypnotherapy, 4737
  - automatic writing and drawing techniques, 2923
  - for conversion disorder, 1901
  - developmentally informed, 2922
  - directive, 2922
  - exploratory, 2922
  - for functional gastrointestinal disorders, 2303
  - hynoprojective techniques, 2923
  - ideomotor questioning technique, 2923
  - individualized approach, 2922
  - permissive, 2922
  - problem-solving, 2922
  - protocol-based approach, 2922
  - relational disturbances, 2927–2928
  - for sexual dysfunctions, 2051
  - supportive, 2922

  - techniques, 2921–2923
  - treatment planning, 2921–2923
- Hypnotic hypermnesia, 2939
- Hypnotic induction ceremony, 2920–2921
  - eye fixation/closure, 2921
  - eye roll technique, 2921
  - holding on/letting go, 2921
  - progressive relaxation, 2921
  - purpose and preparation, 2920–2921
  - staircase/escalator, 2921
  - visual imagery techniques, 2921
- Hypnotic induction plus manualized treatment protocol, 2922
- Hypnotic Induction Profile (HIP), 2920
- Hypnotic-related disorders. *See* Sedative-, hypnotic-, anxiolytic-related disorders
- Hypnotics, 1420. *See also* Sedatives
  - defined, 4737
  - history, 4646
  - intoxication, 2727
  - withdrawal, 2729
- Hypoactive delirium, 2396
- Hypoactive sexual desire disorder (HSDD), 3431
  - defined, 4737
  - treatment of, 3433
- Hypoactivity, 4737
- Hypoadrenalism, 715–716
- Hypochondria, 1888, 4737. *See also* Somatic symptom and related disorders
- Hypochondriacal neurosis. *See* Hypochondriasis
- Hypochondriasis, 920*t*, 1899, 2506–2507. *See also* Illness anxiety disorder; Somatic symptom disorder
  - defined, 4737

  - pediatric OCDs and, 3841–3842
- Hypoesthesia, 2480*t*
- Hypoglycemia, 4737
- Hypogonadotropic hypogonadism, 2584
- Hypokinetic disorders, 3092
- Hypomania, 1693, 4656
  - anabolic steroid-induced, 1443
  - defined, 4737
  - mania *versus*, 1677–1678
  - traumatic brain injury and, 683
  - treatment of, 1741
- Hyponatremia, 3291
- Hypoparathyroidism, 715
- Hypophysis, 710
- Hypopituitarism, 712
- Hypopnea, 4213–4214
- Hyposchemazia, 1175
- Hyposexuality, 465
- Hyposmia, 466
- Hyposomnia, 1677
- Hypotension, orthostatic, 3069
- Hypothalamic disorders, 709–710
- Hypothalamic lesions, 710, 711*f*
- Hypothalamic-pituitary-adrenal (HPA) axis, 159–161, 1798–1799
  - racism-based stress and, 4106–4107
  - sleep and, 204–205
- Hypothalamic-pituitary-gonadal (HPG) axis, GnRH administration and, 162
- Hypothalamic-pituitary-thyroid (HPT) axis, **165–166**
- Hypothalamus, 10, 36–39
  - appetite and, 228, 228*f*
  - defined, 4737

  - nuclei, 37*t*, 39*f*
- Hypothesis testing, 898–901, 1029. *See also* Statistics and experimental design
  - false discovery rate, 901
  - multiple hypotheses, familywise error rate, 899–901
  - power of test, 898–899
  - *p* values, 899
  - type I and type II errors, 898
- Hypothetical-deductive thought, 748
- Hypothyroidism, 713, 2335–2337
  - comparative nosology, 2335
  - course and prognosis, 2336
  - defined, 4737
  - definition, 2335
  - diagnosis and clinical features, 2335–2336
  - differential diagnosis, 2336
  - epidemiology of, 2335
  - etiology of, 2335
  - neuropsychiatric symptoms, 2336
  - pathology and lab examination, 2336
  - physical symptoms, 2335–2336
  - psychiatric symptoms of, 166
  - treatment of, 2336–2337
- Hypotonia, 3676
- Hypoventilation, sleep-related, 2179–2180
- Hypoxyphilia, 2049
- Hysteria, 915, 1888, 4737. *See also* Somatic symptom and related disorders
- Hysterical amnesia, 782
- Hysterical anesthesia, 4737
- Hysterical character, 955
- Hysterical personality. *See* Histrionic personality disorder

- Hysterical psychosis, 1628*t*, 2022–2023
- Hysterics, 4737
- Hysteroid dysphoria, 1694

I

- Iatrogenic, 4737
- Iboga alkaloids, 1355
- Ibogaine, 1355, 1358
  - for alcohol use disorder, 2907
  - chemical structure of, 3386*f*
  - intoxication, 2903
  - studies, 3388
- Ibuprofen, for tension-type headaches, 651
- Icarus Project, 2879
- ICD. *See International Classification of Diseases and Related Health Problems* (ICD-11)
- Ice, 4737
- ICF. *See International Classification of Functioning, Disability, and Health* (ICF)
- ICT. *See* Insulin coma therapy (ICT)
- Ictal hallucinations, 1174
- Ictus, 521
- Id, 422, 917, 1702, 2749, 4737
- Idealization, 2748*t*, 4737
- Idealized image, 957
- Idealized parental imago, 926, 2752
- Ideas of reference, 1456, 4737
- Ideas of unreality, 4737
- Ideational apraxia, 1021
- Ideational shield, 4737
- Idée fixe, 4737
- Ideler, Carl, 4649
- Identical by descent (IBD), 253, 256
- Identical twins. *See* Monozygotic twins
- Identification, 919

  - with the aggressors, 4737
  - definition of, 2959*t*, 4737
  - with predecessors, 4601
- Identity, 790–791, 809, 939, 2233–2234, 4737. *See also* Dissociative identity disorder
  - assessment of, 2233–2234
  - definition of, 2232
  - psyche and, 2231–2234
  - self-psychology, 926
  - temperament, character and, 2228*t*
- Identity crisis, 4737
- Identity disorder, 4737
- Identity thought, 996*t*
- *Identity: Youth and Crisis* (Erikson), 938–939
- Ideomotor activity, 2917
- Ideomotor exploration, 2917
- Ideomotor limb apraxia, 1021
- Ideomotor questioning technique, 2923
- Ideosensory characteristics, 2917
- Idioms of distress, 852, 2551, 2551*t*
- Idiopathic, 4737
- Idiopathic amenorrhea, 2584
- Idiopathic hypersomnia, 4811*t*
- Idiopathic insomnia, 2167
- Idiopathic Parkinson disease (IPD). *See* Parkinson disease (PD)
- Idiot, 4737
- Idiotropic, 4737
- Idiot-savant, 4737
- IDS (Inventory of Depressive Symptomatology), 1105
- IED. *See* Intermittent explosive disorder (IED)
- Ifenprodil, 84*f*
- Ifosfamide, 2367

- I-It, 4737
- Illicit drugs, 4288
- Illness anorexia, 231
- Illness anxiety disorder, 1898–1900
  - comparative nosology, 1899
  - definition of, 1898, 4737–4738
  - diagnosis of, 4806*t*
  - epidemiology of, 1899
  - history, 1898–1899
  - hypochondriasis, 2507
  - pediatric OCDs and, 3841–3842
  - treatment of, 1899–1900
    - pharmacologic, 1899
    - psychotherapy, 1899
    - reassurance, 1899
- Illnesses
  - child/adolescent response to, **3981–3992**
    - asthma, 3989
    - cancer, 3990
    - COVID-19, 3990–3991
    - cystic fibrosis, 3989
    - at different ages, 3986–3988, 3987*t*
      - adolescents, 3988
      - infants, 3986–3987
      - preschool age, 3987
      - school age, 3987–3988
      - toddlers, 3987
    - heart disease, 3989–3990
    - hemophilia, 3990
    - historical background, 3981–3982
    - impact on children, 3983–3984
    - impact on families, 3984–3986

    - inflammatory bowel disease, 3989
    - insulin dependent diabetes mellitus (IDDM), 3989
    - rheumatologic disorders, 3989
    - social media and, 3992*t*
    - surgical procedures, 3990
    - transition from pediatric to adult medical care, 3991–3992, 3992*t*
  - strength-vulnerability model, 4373
- Illness falsification, 2503
- Illness prototypes, 2554
- Illness/wellness management skill training, 4555–4556
- Illogicality, 4738
- Illogic thinking, 4738
- Illuminism, 4738
- Illusion, 996*t*, 4738
- Illusions, 1172
- Iloperidone (Fanapt), 3082–3084
  - chemistry, 3082
  - clinical trials, 3083
  - dosage and administration, 3084
  - drug interactions, 3084
  - molecular structure, 3056*f*
  - pharmacodynamics, 3083
  - pharmacokinetics, 3082–3083
  - precautions and adverse reactions, 3083–3084
  - therapeutic indications, 3083
- Imaginal exposure, 1826
- Imaginary companionship, 1965
- Imago, 4738
- Imbecile, 4738
- Imipramine (Tofranil), 3007, 3152
  - for anxiety disorders in later life, 4224–4225

  - for enuresis, 3794–3795
  - pharmacogenetics of, 295*t*
  - pharmacokinetics, sex differences in, 3416
- Imitation, 746–747, 2959*t*, 4738
- Immature defenses, 920*t*, 2748*t*
- Immediate memory, 775, 4738
  - schizophrenia and, 1557
- Immigrants, 835
  - mental disorders among, 841–843
  - psychotic disorders in, 842–843
  - schizophrenia in, 842–843
- Immigration, 835–836
- Immune-related disorders, hypnosis for, 2936–2937
- Immune system
  - adaptive immunity, 176–177
  - brain communication, 177–181
  - brain-computer interface (BCI), 191
  - central nervous system and, 2416
  - components of, 172–177
  - functions of, 171
  - immune-CNS interactions
    - autism, 193
    - bipolar disorder, 187–188
    - factors in
      - aging, 190
      - microbiome, 191–192
      - nutrition, 191
      - obesity, 191
      - physical activity, 191
      - sleep, 192–193
    - major depression, 181–185
    - neuro-Covid, 190

    - obsessive-compulsive disorders, 189
    - post-traumatic stress disorder, 189
  - innate immunity, 172–176
  - naltrexone for, 3364
  - neurochemical receptors, 177
  - schizophrenia and, 185–187
  - stress and, 181
- Immune therapies, for schizophrenia, 186–187
- Immunity
  - behavioral influences on, 2418
  - central modulation of, 2417
  - chronic stress, depression, and, 2418
  - psychological influences on, 2418
  - sleep, cytokines, and, 2418
  - stress and, 2418
- Immunodeficiency– centromeric instability– facial anomalies syndrome (ICF), 271*t*
- Immunosuppressants, neurotoxicity of, 2463–2465, 2465*t*
- Impaired insight, 4738
- Impaired judgment, 4738
- Impairments, 3624
- Impedance, 3445
- Impersistence, 3676
- Implanted cardioverter defibrillators, 2290
- Implanted cortical stimulation, 3477–3478, 3477*f*
- Implicit memory, 730, 730*t*
- Implosion, 4738
- Impostor syndrome, 1180
- Impotence, 2041, 4738. *See also* Male erectile disorder
- Imprinting, 4738
- Imprinting Prader–Willi syndrome (PWS), 3641*t*
- Improvisation, 4738

- Impulse, 4738
- Impulse control, 4738
- Impulse control disorders (ICDs), **2194–2200**
    - borderline intellectual functioning and, 2533
    - carbamazepine for, 3293
    - defined, 4738
    - dopamine receptor antagonists for, 3044–3045
    - ICD-11 criteria for, 1224
    - intellectual disabilities and, 3647
    - intermittent explosive disorder, **2194–2196**
    - kleptomania, **2198–2200**
    - Parkinson disease and, 559–560
    - pyromania, **2197–2198**
    - valproate for, 3308–3309
- Impulsion, 4738
- Impulsive actions, 299–300
- Impulsive behavior, 1676
- Impulsive choices, 300
- Impulsivity, 299–300
    - ADHD and, 3731
- Inactivation, 70
- Inadequate sleep hygiene, 2167–2168
- INAH3. *See* Third interstitial nucleus of the anterior hypothalamus (INAH3)
- Inappropriate affect, 4738
- Inattention/Overactivity with Aggression (IOWA) Conners Scale, 3545
- Inborn errors of metabolism (IEMs), diagnosis, 3638
- Incentive learning, 764
- Incentive sensitization theory, 444
- Incest, 4738
    - as maltreatment, 3979

- Incidence, 875
  - definition of, 4738
  - prevalence and, 875
- Incidence rate, 4738
- *Incilius alvarius*, 1350*f*
- Inclusion body myositis (IBM), 663
- Incoherence, 1462, 4738. *See also* Word salad
- Incompetence, 4738
- Incomplete abortion, 2590*t*
- Incomplete data, 904–905
- Incorporation, 918, 4738
- Increased libido, 4738
- Incremental cost-effectiveness ratio (ICER), 912–913
- Incubus, 2182
- Indecent exposure, 2094
- Indels, 242
- Independent Living Skills-Survey (ILSS), 4553
- Independent variable, 4738
- Inderal. *See* Propranolol (Inderal)
- India, biologic therapies in, 3442–3443
- Indigenous historical trauma (IHT), 844–845
- Indigenous Peoples, mental health of, 844–845
- Indiplon, 3255
- Indirect hypnosis, 2922
- Indirect pathway, 31
- Individualism-collectivism, 847
- Individualized education program (IEP), 4032–4035
  - assessments, 4032
  - definition of, 4026*t*
  - eligibility categories, 4034, 4034*t*
  - five504 Plan *versus*, 4032*t*
  - key parts of, 4035

  - meeting components, 4032–4033
  - meeting requirements, 4033
  - steps in initial process, 4033*t*
- Individual psychology, 4738
- Individuals with Disabilities Education Act, 4026*t*
- Individual therapy, 4738
- Individuation, 4738. *See also* Actualization
- Indoklon therapy, 4738
- Indolamine, 4738
- Induced abortion, 2590*t*
- Induced dreams, 2917
- Induced psychotic disorder. *See* Shared psychotic disorder
- Inducible NOS (iNOS), 125
- Industrial psychiatry, 4738
- Industry, 939
- Ineffability, 4738
- *I Never Promised You a Rose Garden* (Greenberg), 2790–2791
- Infancy, 4738
- Infantile amnesia, 783
- Infantile anorexia, 3762–3764
  - clinical features, 3762
  - course, 3762–3763
  - diagnosis, 3762
  - differential diagnosis, 3762
  - etiology, 3762–3763
  - treatment, 3763–3764
- Infantile articulation. *See* Developmental articulation disorder
- Infantile autism, 4738–4739
- Infantile dynamics, 4739
- Infantile perseveration. *See* Developmental articulation disorder
- Infantile sexuality, 4739
- Infantilism, 2091–2092

- Infant psychiatry, 4739
- Infants. *See also* Children/childhood; Neonates; Preterm infants
  - assessment of, 3529–3532
    - diagnostic formulation of, 3532
    - special types of, 3532–3535
  - attachment, 3500
  - disoriented/disorganized, 928
  - formal tests of development, 3533*t*
  - low birth weight, 668–669
  - mental status examination, 3534–3535*t*
  - negative affect states, 822
  - observation and examination of, 3531–3532
  - premature birth, 668–669
  - racism and, 4107
  - response to hospitalization/.illnesses, 3986–3987
  - structured assessment instruments, 3532
  - technical issues in the development of, 3530–3531
- Infectious diseases, 484, 1052–1054
  - Alphaviridae, 616
  - amoebiasis, 627–628
  - central nervous system, 1052
  - Chikungunya, 616
  - chlamydia, 1053
  - COVID-19, 610–615
  - cytokine-brain link and, 2419
  - cytomegalovirus, 619
  - dengue, 616–617
  - Eastern equine encephalitis, 616
  - Epstein–Barr virus (EBV), 619
  - gonorrhea, 1053
  - group A beta-hemolytic streptococcus (GAS), 625–626
  - helminths, 628

  - hepatitis B, 1054
  - hepatitis C, 1054
  - HHV-6, 619
  - HIV, 1053
  - human herpesviruses, 618–619
  - influenza, 620
  - intravenous drug use and infection, 1053
  - leptospirosis (Weil disease), 625
  - *Listeria*, 620
  - Lyme disease, 623–625
  - malaria, 626–627
  - measles, 619–620
  - neurocysticercosis, 628
  - with neuropsychiatric manifestations, 611–614*t*
  - pediatric acute onset neuropsychiatric symptoms, 626
  - Powassan virus disease, 617
  - protozoa, 626–628
  - rabies, 617–618
  - rubella, 620
  - schizophrenia comorbidity, 1595
  - sexually transmitted diseases, 1053
  - Spirochetes, 620–623
  - syphilis, 620–623, 1053–1054
  - toxoplasmosis, 627
  - trypanosomiasis, 627
  - viral hepatitis, 1053–1054
  - Western equine encephalitis, 616
  - West Nile encephalitis, 617
  - Zika virus, 620
- Infectious neuropathies, 659–661
- Infective endocarditis (IE), 2285
- Inferential statistics, 4739

- Inferiority, 939
- Inferiority complex, 4739
- Inferior longitudinal fasciculus, 13
- Inferior occipitofrontal fasciculus, 13
- Infertility, 2585–2595
  - lifestyle factors, 2586*t*
  - tests for, 2588
- Inflammasomes, 173–174
- Inflammation, depression and, 182–183
- Inflammation hypothesis, 4232–4233
- Inflammatory biomarkers, 183, 4474
- Inflammatory bowel disease, 2294–2296
  - child/adolescent response to, 3989
  - psychological and psychiatric interventions in, 2295
- Inflammatory diseases, 484, 1056–1057
- Inflammatory myopathies, 662–663
- Influence, 996*t*
- Influenza, 620
- Informal admission, 4739
- Information, 727–728
- Information processing, 728–729
  - models, 728*f*
- Information theory, 4739
- Informed consent, 2387
  - defined, 4739
  - depressive disorders, 3819
  - ethical issues in cancer and, 2373
  - for experimental treatment, 2409
  - for forensic evaluations, 4004–4005
  - forensic geropsychiatry, 4417–4418
  - forensic psychiatry, 4621–4622
  - geriatric psychiatry, 4424–4426

  - narrative psychiatry, 2881
  - psychiatric emergencies, 2733–2734
  - psychopharmacology, 3011–3012
  - in terminally ill transplant candidates, 2454
- Informed refusal, 2387
- Infradian rhythm. *See* Biologic rhythm
- Infusion, 4739
- Ingvar, David, 4661
- Inhalant intoxication, 1375
- Inhalant-related disorders, **1365–1382**
  - amyl, butyl, and isobutyl nitrite–related disorders, 1381
  - comorbidity, 1369
  - comparative nosology, 1365–1366
  - course and prognosis, 1375
  - definition of, 1365
  - development of, 1374–1375
  - diagnosis and clinical features, 1375–1376
  - differential diagnosis, 1375
  - DSM-5 criteria for, 1366
  - fetal solvent syndrome, 1373–1374
  - ICD-11 criteria for, 1366
  - inhalant-induced psychiatric disorders, 1378–1380
  - inhalant intoxication, 1375
  - inhalant-related deaths, 1368–1369
  - inhalant-related emergency department visits, 1367
  - inhalant-related treatment admissions, 1367–1368
  - inhalant use disorder, 1376–1378
  - laboratory examination, 1375
  - mechanism of action, 1370*t*
  - nitrous oxide-related disorders, 1380–1381
  - organ pathology, 1372–1373
  - patterns of inhalant use, 1374

  - pharmacology of, 1369–1372, 1370*t*
  - prevalence of, 1367
  - toxicology, 1369–1372
  - treatment of, 1376
  - unspecified, 1380
- Inhalants
  - commonly misused, 1365*t*
  - EEG effects, 380
  - psychiatric emergencies, 2729
  - withdrawal, 2730
- Inhalant use disorder, 1376
  - course and prognosis, 1377
  - differential diagnosis, 1376–1377
  - DSM-5 criteria for, 1376
  - laboratory examination, 1376
  - treatment of, 1377–1378
- Inhibited female orgasm, 2042–2043, 4739. *See also* Anorgasmia
- Inhibited male orgasm, 4739. *See also* Primary retarded ejaculation; Secondary retarded ejaculation
- Inhibited sexual excitement, 4739
- Inhibition, 921*t*, 4739
- Inhibitory interneurons, 62–63
- Initial insomnia, 4739
- Initiative, 938
- Initiator element (INR), 247
- Injections/doctor phobia, behavioral approach tasks for, 3888*t*
- Innate immunity, 172–176
  - acquired immunity *versus*, 172
  - acute phase response proteins (ARP), 174–175, 174*f*
  - complement factors, 175–176
  - cytokines, 174, 175*t*, 180*t*
  - inflammasomes, 173–174

  - intracellular signaling cascades, 173
  - schizophrenia and, 185
  - surface pattern recognition receptors, 172–173
  - surveillance system, 172
- Inositol, 2633, 4146*t*
- Inpatient psychiatric care, **3936–3939**
  - family interventions and, 4375–4376
  - legal considerations, 3938
  - milieu management, 3934
  - multidisciplinary approach, 3938
  - transitional age youth, 3939
  - treatment and discharge planning, 3938–3939
- Insane, 4739
- Insanity, 4008, 4739
- Insanity defense, 4739
- Insecurity, 4739
- Insel, Thomas, 1209
- Insight, 997, 2959*t*, 4739
  - lack of, in schizophrenia, 1465
  - in older adults, 4135
- *Insight and Responsibility* (Erikson), 938
- Insight-oriented psychotherapy, 4353–4354, 4739
- Insight therapy. *See* Insight-oriented psychotherapy
- Insomnia, 1187–1188, 1674–1675. *See also* Fatal familial insomnia (FFI)
  - animal models of, 303
  - antihistamines for, 3348
  - benzodiazepines, 3244–3246
  - in children/adolescents, 4085
  - comorbid with medical conditions, 2168–2169
  - comorbid with mental disorder, 2168–2169
  - end-of-life care and, 2391–2392

  - idiopathic, 2167
  - monoamine oxidase inhibitors (MAOIs), 3191
  - paradoxical, 2167
  - pharmacotherapy for, 3933–3934
  - psychophysiological, 2167
  - short sleeper, 2170
  - trazodone, 3201
  - treatment of, 2170–2172
- Insomnia disorder, 2165, **2165–2172**, 2170
  - defined, 4739
  - definition of, 2165
  - diagnostic criteria, 2165–2166, 4811*t*
  - ICD-3 criteria for, 2166*t*
  - in older adults, 4211–4212
  - subtypes of, 2166–2170, 2166*t*
    - behavioral insomnia of childhood, 2168
    - comorbid with medical conditions, 2169
    - comorbid with mental disorder, 2168–2169
    - idiopathic, 2167
    - inadequate sleep hygiene, 2167–2168
    - paradoxical, 2167
    - psychophysiological, 2167
    - short sleeper, 2170
    - substance-abuse related, 2169–2170
  - treatment of, 2170–2172, 4210–4211
    - behavioral, 4211
    - pharmacologic, 4211–4212, 4211*t*
- Insomnia Severity Index (ISI), 2189
- Inspiration, 2959*t*
- Instagram, 1128
- Instincts
  - defined, 4739

  - neuroimaging and, 2782
- Institute for Safe Medication Practices (ISMP), 1158
- Institute of Living, 4652
- Institute of Medicine (IOM), 1013
- Institutional betrayal trauma, 1935–1936
- Institutionalization, 4739
- Institutional racism, 4105*t*
- Institutions for Mental Diseases (IMDs), 4512–4513
- Instrumental action, 765–767
- Instrumental activities of daily living (IADLs), 4179–4180
- Instrumental conditioning, 756, 4739
- Insula, 798, 3614
- Insulin, **169**
- Insulin coma therapy (ICT), 4739
- Insulin dependent diabetes mellitus (IDDM), child/adolescent response to, 3989
- Insulin-like growth factor 1 (IGF-1), 64
- Insulinoma, 717
- Insulin resistance, 1598, 3058
- Insulin withholding, 2329
- Intake, 4739
- Integrated care, 3983
- Integrated sex therapy, 2052
- Integration, 4739
- Integrative cognitive-affective therapy (ICAT), for bulimia nervosa, 2162
- Integrative relational therapy, 2739
- Integrity, 791, 940
- Intellectual disabilities (ID), 3630–3655
  - adaptive functioning, 3632
  - common measures of adaptive behavior, 3635*t*
  - in correctional facilities, 4637

- definition of, 3632, 4739
- deinstitutionalization, 3630–3631
- developmental considerations, 3644–3648
  - child-related aspects of development, 3645
    - cross-domain relations, 3645
    - developmental rates, 3645
    - sequences, 3645
  - contextual aspects of development, 3645–3646
    - family reactions, 3646
    - mother-child interactions, 3645–3646
- diagnosis, 4773*t*
  - clinical history and examination, 3637
  - criteria, 3633*t*, 3634*t*
  - electroencephalogy, 3638
  - genetic, 3637, 3637*t*
  - instruments, 3648
  - metabolic testing, 3638, 3638*t*
  - neuroimaging, 3637–3638
  - rating scales, 3648
- disinhibited social engagement disorder and, 3805
- Down syndrome, 3643
- DSM-IV severity criteria, 3634*t*
- DSM-IV *vs.* DSM-5 criteria for, 3633*t*
- etiology of, 3636–3637, 3636*t*
- forensic issues, 3654
- fragile X syndrome, 3639
- genetic syndromes and phenotypes, 3638–3644
- historical background, 3630
- ICD-11 criteria for, 1223
- intellectual functioning, 3632
- legal issues in, 3654–3655
- legislation for, 3654–3655

  - maladaptive behavior and, 3648
  - malingering and, 2502
  - measures of assessing, 3633*t*
  - mild, 3635
  - moderate, 3635
  - organizations, 3632
  - Prader–Willi Syndrome, 3639–3643
  - prevalence of, 3635–3636
  - prevention, 3648–3649
  - profound, 3635
  - psychiatric emergencies, 3582
  - psychopathology in, 3646–3648
  - reactive attachment disorder and, 3805
  - research, 3631–3632
  - school and, 3631
  - services for, 3652–3654, 3653*t*
    - adult services, 3654
    - early intervention, 3652–3653
    - school, 3653–3654
  - severe, 3635
  - severity of, 3633–3635
  - Smith–Lemli–Opitz syndrome, 3644
  - Smith–Magenis syndrome, 3644
  - supports for, 3654–3655
  - treatment, 3648–3654
    - etiologic factors, 3649
    - pharmacotherapy, 3649–3652
    - psychotherapy, 3649
  - Williams syndrome, 3643–3644
- Intellectual functioning, 3632
- Intellectual functioning (IF), 2526–2528. *See also* Cognitive function

- Intellectual insight, 4739
- Intellectualization, 921*t*, 2748*t*, 4740
- Intellectual subaverage functioning, 4740
- Intelligence, 2527
  - assessment, 3564–3565
  - comprehensive intellectual tests, 3565
  - crystallized, 4164
  - defined, 4740
  - definition of, 3564
  - fluid, 4164
  - infant and child tests, 3565
  - nonverbal and language-free tests, 3565
  - Piaget's models of, 749–753
    - developmental levels, 753–754
    - emotional basis of, 751–752
    - extensions of theory, 749–750
    - post-Piaget era, 750–751
  - protection, 3564
  - quick tests, 3565
  - stability, 3564
- Intelligence quotient (IQ), 4740
- Intensive outpatient programs (IOPs), 3939
- Intensive short-term dynamic psychotherapies (ISTDPs), **2763–2783**. *See also* Short-term dynamic psychotherapies (STDPs)
  - clinical issues, 2770–2772
  - complications, 2771
  - contraindications, 2771
  - history, 2764–2765
  - limitations, 2770–2771
  - medications and, 2779
  - outcomes, 2771–2772
  - process-outcome correlations, 2779–2780

  - techniques, 2770
  - theoretical issues, 2769–2770
  - treatment goals, 2771
- Intentionality, 2213
- Intentionality Bias Task (IBT), 4551
- Interaction, 2959*t*
- Interactive apps, 1121
- Interictal dysphoric disorder of epilepsy, 527
- Interictal epileptiform discharge (IED), 372
- Interictal period, 520
- Interictal psychosis, 525–526, 526t, 679
- Interiorization, 747
- Interjudge reliability, 1092
- Interleukin-1 (IL-1), 184
- Interleukin-6 (IL-6), 183
  - sleep and, 192–193
- Interleukin-8 (IL-8), 183
- Intermal object relations, 2790*t*
- Intermediate denial, 2748*t*
- Intermission, 4740
- Intermittent explosive disorder (IED), **2194–2196**
  - definition of, 4740
  - diagnosis, 2194–2195, 4822*t*
  - DSM-5 criteria for, 2194*t*
  - epidemiology of, 2194–2195
  - etiology of, 2195
  - pharmacotherapy for, 2195–2196
  - psychotherapy for, 2196
  - treatment of, 2195–2196
- Intermittent photic stimulation (IPS), 370
- Intermittent reinforcement, 4740
- Intermittent theta-burst stimulation (iTBS), 3472

- Intermodal perception, 3503
- Internal capsule, 10–11
- Internal consistency, 1092, 3537
- Internal consistency reliability, 1037
- Internal homophobia, 2075
- Internalized homophobia, 2075
- Internalized racism, 836–837
- Internal loop, 2380
- Internal medullary lamina, 23
- Internal mentalizing, 2864
- Internal object(s), 820–821
- Internal working model, 927
- International adoption, 3965–3966
- International Classification of Causes of Death, 1200
- International Classification of Diseases (ICD), 1200, 1213–1217
  - development of, 1214–1215
  - DSM classification *versus*, 1216–1217, 1217*t*
  - ICD-9, 1215
  - ICD-10, 1215–1216
  - within WHO Family of International Classifications, 1214
- *International Classification of Diseases and Related Health Problems* (ICD-11), 882, **1213–1227**
  - addiction, 1224
  - adjustment disorders, 2203–2204
  - agoraphobia, 1223
  - alcohol-related disorders, 1224
  - alcohol use disorders, 1296, 1302*t*
  - anxiety disorders, 1222–1223
  - bipolar I disorder, 1681–1682*t*, 1683–1684*t*
  - caffeine-related disorders, 1324*t*, 1325
  - child maltreatment, 3970
  - in clinical practice, 1211

- controversies, 1224–1225
- cultural issues in, 1218
- cyclothymic disorder, 1685*t*
- defined, 4740
- definition of mental disorders in, 1218–1219
- delirium, 1230
- developing from ICD-10, 1217–1218
- developmental coordination disorder, 3677*t*
- developmental speech sound disorder, 3695
- diagnostic categories and clinical descriptions in, 1220, 1220*t*
- dimensional classification and, 1210–1211
- dimensional perspectives in, 1219
- dimensions and categories in, 1219
- dissociative identity disorder, 1940–1942
- encopresis, 3797*t*
- enuresis, 3790*t*
- factitious disorder, 1906–1907, 1907*t*
- factitious disorder imposed on self, 1908–1909
- feeding and eating disorders, 1224
- field trials, 1220–1221
- functional somatic syndromes, 1891
- gaming disorder, 1224
- generalized anxiety disorder, 1222–1223
- harmonization with DSM, 1225–1227
- hypochondriasis, 1899
- impulse control disorders, 1224
- intellectual disability/mental retardation, 1223
- Internet gaming disorder, 4090*t*
- language disorder, 3681
- major depressive disorder, 1679–1680*t*
- malingering, 2494
- mental and behavioral disorders, 1219–1220, 1220*t*

  - mood disorders, 1222
  - obsessive-compulsive disorders, 1223, 1840–1841
  - opioid-related disorders, 1397
  - panic disorder, 1223
  - personality disorders, 1193, 1224, 1225*t*
  - phobias, 1223
  - post-traumatic stress disorder, 1223
  - primary health care (ICD-11-PHC), 1221
  - schizoaffective disorder, 1221
  - schizophrenia, 1221
  - schizotypal disorder, 1221
  - sexual dysfunction, 2038, 3430
  - social phobia, 1223
  - somatic symptom and related disorders, 1891
  - somatic symptom disorder, 1896
  - specific disorders in, 1221–1224
  - specifiers in, 1219
  - stereotypic movement disorders, 3811–3812
  - stimulant-related disorders, 1313, 1313*t*
  - stuttering, 3699
  - tic disorders, 3775
- *International Classification of Functioning, Disability, and Health* (ICF), 882, 883*t*, 1229–1230, 1891*t*
- International Covenant on Civil and Political Rights (ICCPR), 1199
- International HapMap Consortium, 234
- International Pilot Study of Schizophrenia (IPSS), 1200, 1450
- International Society for Brain Electromagnetic Topography (ISBET), 368
- International Society for the Study of Trauma and Dissociation (ISSTD), 1932
- International Statistical Classification of Diseases, Injuries, and Causes of Death, 1200

- *International Statistical Classification of Diseases and Related Health Problems* (ICD-10)
  - child maltreatment, 3970
  - inhalant-related disorders, 1366
  - malingering, 2498
  - specific speech articulation disorder, 3695
  - stereotypic movement disorders, 3811–3812
  - stuttering, 3699
- International Study of Schizophrenia (ISoS), 1477, 1477*t*
- Internet, 1126–1128, 3514–3515. *See also* Social media; Technology
  - addiction, 1127–1128
  - dating apps, 1127
  - gaming, 1127–1128
  - historical background, 2699–2700
  - identity and, 1133–1134
  - information on disease, diagnosis, treatment, and therapists, 1126
  - online advice, 1126
  - pornography, 1127
  - prescription/illicit drug sales, 1126–1127
  - problematic use of, 1135
  - social media, 1122–1128
  - support and self-help groups, 1126
  - treatment of pathologic use, 1134
- Internet Archive Wayback Machine, 2699
- Internet gaming disorder, **4090–4093**
  - anxiety disorders and, 4092
  - attention-deficit/hyperactivity disorder and, 4092
  - biologic correlates of, 4091
  - clinical recommendations on gaming limits, 4093
  - comorbidity, 4091–4092

  - definition of, 4090
  - diagnosis, 4827*t*
  - DSM-5 criteria for, 4090*t*
  - future directions, 4093
  - ICD-11 criteria for, 4090*t*
  - major depressive disorder and, 4091–4092
  - neurobiology of, 4091
  - online video games, risks and benefits oif, 4091
  - parent recommendations on gaming limits, 4092–4093
  - prevalence of, 4090
  - public policy on, 4093
  - treatment of, 4092
- Interneurons, 3, 18
  - inhibitory, 62–63
- Interns, working hours of, 1155–1156
- Internship, 4740
- Interobserver reliability, 4740
- Interoceptive exposure, 1826, 2929
- Interoceptive sensation-feelings, 812
- Interpersonal conflict. *See* Extrapsychic conflict
- Interpersonal interpretive technique, 2794–2799
- Interpersonal psychiatry, 4740
- Interpersonal psychotherapy (IPT), 961–962, **2831–2840**, 4352
  - for anorexia nervosa, 2161–2162
  - for binge eating disorder, 2163
  - for bulimia nervosa, 2162
  - clinical issues, 2838–2840
  - complications, 2839
  - contraindications, 2839
  - defined, 4740
  - for depressive disorders in children and adolescents, 3820
  - for eating disorders, 2313–2314

- efficacy and effectiveness of, 2990
- ethical issues in, 2841
- goals of, 2840
- group, 2836–2838
- history, 1745, 2831–2832
- indications for, 2839
- initial phase
  - diagnosis, 2833
  - group, 2837
  - interpersonal inventory, 2833
  - problem areas, 2833–2834
  - sick role assignment, 2833
  - treatment goals, 2834
- intermediate phase
  - facilitate work on identified goals, 2835
  - group, 2837
  - problem areas, 2834–2835, 2834*t*
- limitations, 2839
- research and evaluation, 2841
- techniques, 2833–2836
  - group, 2836–2837
- termination phase, 2835–2836
  - group, 2837–2838
  - maintenance of therapeutic gains, 2836
  - as potential time of grief, 2835–2836
  - progress review, 2836
- therapeutic stance, 2833
  - group, 2837
- therapeutic techniques, general
  - clarification, 2836
  - communication analysis, 2836
  - encouraging effect, 2836

    - exploratory questions, 2836
    - therapeutic relationship and, 2836
  - time-limited, 2832
  - timeline, 2833–2836
    - group, 2836–2837
- Interpersonal racism, 836–837
- Interpersonal relationships
  - disturbances in, 1192–1194
    - dysnosognosia and, 1193–1194
    - interpersonal systems, 1193
    - personality trait and disorders, 1192–1193
- Interpersonal skill, 4740
- Interpersonal skills training, 1827
- Interpersonal systems, 1193
- Interpersonal theory of depression, 1707–1708
  - bipolar disorder, 1707
  - dysthymic disorder, 1707
- Interpersonal therapy
  - for adolescents, 3953
  - for functional gastrointestinal disorders, 2303
- Interpretation, 934–935, 2959*t*, 4740
- *Interpretation of Dreams, The* (Freud), 212, 4740
- Interpreter, 850–851
- Inter-rater reliability, 880, 3537
- Interrater reliability, 1092
- Interrogatory Suggestibility (IS), 1969
- Intersex conditions, 2101
- Intersex youth, 4101–4102
- Interstitial lung disease, 2321
- Intersubjective theory, 928–929
- Intersubjectivity, 2750–2751
- Interval enforcement schedules, 762

- Interval recording, 2815
- Interventions
  - expanding, 1000–1001
  - facilitating, 998–1000
  - obstructive, 1001–1002, 1001*t*
- Interview-based interviews (IBIs), 3538–3539, 3540–3541
- Interviews. *See also* Examination of psychiatric patient
  - in behavior therapy, 2814–2815
  - challenging scenarios
    - patients engaging in deception, 1004
    - patients with depression, 1003
    - patients with psychosis, 1002–1003
    - suicidal patients, 1003
    - violent patients, 1003–1004
  - closing, 998
  - comfort and safety, 987
  - confidentiality, 987
  - cultural variation in, 1004–1005
  - defined, 4740
  - disorder-based, 989–990
  - electronic medical record, 1004
  - engagement and relationship building in, 988–989
  - first encounter, setting stage for, 987
  - forensic, 1005–1006
  - interpreters, 1005
  - meaning and communication, 989
  - note taking, 1004
  - observed or recorded, 1005
  - opening, 990–991
  - open- versus closed-ended questions in, 989
  - person-centered, 989–990
  - pre-interview contacts and interactions, 986–987

  - process, 990
  - telepsychiatry and, 1006
  - third parties in, 987–988
  - time and number of sessions, 990
- Intestinal nutrient sensing, 224–226
- Intimacy, 791, 939, 2037
- Intimate partner violence (IPV), 2624–2625
  - child affected by parental relationship distress and, 2545
  - definition of, 2624
  - perpetrators, 2625
  - prevalence of, 2624
  - risk factors, 2625
- Intoxication, 441–442
  - agitation and, 3576–3577
  - caffeine, 1329–1330
  - cannabis, 1341
  - defined, 4740
  - diagnosis, 4826*t*
  - DSM-5 criteria for, 444*t*
  - inhalant, 1375
  - stimulant, 1318
    - treatment of, 1321
- Intoxication delirium, phencyclidine- and other hallucinogen-induced, 1364
- Intracellular cascades, 1771
- Intracellular signaling pathways, 3004–3006
- Intracranial self-stimulation (ICSS), 445
- Intracytoplasmic sperm injection (ICSI), 2586
- Intrafamilial child sexual abuse, 3979
- Intrahemispheric localization, 1020–1023
- Intralaminar nuclei, 23
- Intraneuronal signaling, **145–156**

  - complexes and scaffold, 155–156
  - functional selectivity, 155–156, 155*f*
  - G-protein-coupled receptor signaling, 147–151
  - remote control, 156
  - signaling pathways, 146–147, 146*f*
  - signal transduction, 145–147, 146*f*
  - synaptic plasticity and, 153–154, 154*f*
  - tyrosine kinase pathways, 151–153
  - Wnt signaling, 153–154, 153*f*
- Intrapersonal conflict. *See* Intrapsychic conflict
- Intrapsychic, 4740
- Intrapsychic ataxia. *See* Ataxia
- Intrapsychic conflict, 4740
- Intrasynaptic dopamine, 1542
- Intrauterine device (IUD), 2592, 2594*t*
- Intravenous, 4740. *See also* Infusion
- Intravenous drug use, 1053
- Introjection, 813–814, 918–919, 920*t*, 4740
- Introject self-state, 1973
- Introns, 237
- Intropunitive, 4740
- Introspection, 4740
- Introversion, 4740
- Introverted personality disorder. *See* Schizoid personality disorder
- Intuition, 726
- Intuniv. *See* Guanfacine; Guanfacine (Intuniv, Tenex)
- Inventory of Childhood Memories and Imaginings (ICMI), 2940
- Inventory of Depressive Symptomatology (IDS), 1105
- Inverse agonism, 3000–3001
- Inversion, 328, 4740
- Inversion recovery (IR), 336
- Inverted personality disorder. *See* Schizoid personality disorder

- In vitro fertilization - embryo transfer (IVT-ET), 2586–2587
- In vivo desensitization, 2816
- In vivo exposure, 1825
- Involuntary adaptive coping mechanisms, 792
- Involuntary admission, 4740
- Involuntary discharge, 4740
- Involuntary treatment, 4740
- Involutional melancholia, 1650, 1652, 4740. *See also* Major depressive disorder (unipolar depression)
- IOM. *See* Institute of Medicine (IOM)
- Ion channel disorders, 708–709
  - Timothy syndrome, 709
  - voltage-gated potassium channel, 709
- Ion channels, **70–77**
  - calcium (Ca^+) channels, 75–77
    - high-voltage-activated, 76
    - low-voltage-activated, 76
    - L-type, 76
    - N-type, 76
    - P/Q-type, 76
    - R-type, 76
  - chloride (Cl^-) channels, 77
  - clinical aspects of, 79–82
    - complex behavioral states, 79–80
    - E/I balance, 80
    - homeostatic plasticity and psychiatric therapeutics, 81
    - neural excitability and brain stimulation methods, 82
    - oscillatory neuronal firing, 79–80
  - neuropsychiatric disorders and, 80–81
  - neurotransmitters and, 77–79
  - potassium (K^+) channels, 73*f*, 74–76, 74*f*
    - calcium-activated, 74

    - delayed rectifier, 74
    - H-channels, 75
    - inward rectifiers, 75
    - M-channels, 74
    - S-channels, 74
    - tandem pore, 75
    - A-type, 74
  - sodium (Na$^+$) channels, 71–73
  - voltage-gated, 70–71, 72*f*
- Ionotropic glutamate receptors, 84–86
- Ionotropic receptors, 2997, 2997*f*
- Iproniazid (Marsilid), 3007, 3185*t*
- IPV. *See* Intimate partner violence (IPV)
- IQ test, 2526
- Irregular sleep phase circadian rhythm disorder, 4741
- Irrelevant answer, 4741
- Irresistible impulse, 4741
- Irritability, 1676, 3519, 4741
- Irritable bowel syndrome, 2296, 2296*f*
  - hypnosis and, 2296*t*, 2936
- Irritable mood, 4741
- Isaacs, Susan, 812
- Ischemic neuropathies, 658
- Isocarboxazid (Marplan), 3185*t*, 3188–3189
  - adverse effects, 3189
  - dosage, 3189
  - lactation and, 3189
  - for major depressive disorder, 1724
  - monitoring parameters, 3189
  - pharmacodynamics, 3188
  - pharmacokinetics, 3189
  - pregnancy and, 3189

- Isolated explosive disorder, 4741
- Isolated populations, schizophrenia in, 1475
- Isolation, 921*t*, 939, 4741
- Isolation of affect, 2748*t*
- Isoniazid (Nydrazid), 3185*t*
- Isophilic, 4741
- Isoptin. *See* Verapamil
- Isradipine (DynaCirc), 3260
- ISTDPs. *See* Intensive short-term dynamic psychotherapies (ISTDPs)
- Isthmus, 12
- Itard, Jean-Marc Gaspard, 3630, 3775
- Iterative reconstruction (IR) algorithm, 390–391
- I-Thou, 4741
- Ivor Asbjörn Fölling, 4644

J

- Jackson, Don, 2947
- Jackson, John Hughlings, 4687
- Jackson Hole Group, 4481
- Jacksonian epilepsy, 4741
- Jacobsen, Charles, 3479
- Jahoda, Marie, 787, 2888–2889
- Jamais entendu, 1182
- Jamais pensé, 1182
- Jamais vu, 996*t*, 1182, 4741
- James, William, 420, 430–431, 789, 2674, 2683, 4471
- James Anthony., E., 2956
- James–Lange theory of emotion, 4741
- Jamison, Kay, 1658, 4697
- Janet, Pierre, 1928, 1930, 3837, 4686, 4741
- Janowsky. David, 1656

- Janssen, Paul, 3007, 3034
- Jargon aphasia, 4741
- Jarvis, Edward, 4509–4510, 4512
- Jasper, Herbert, 357, 369
- Jaspers, Karl, 1199, 1455, 1889, 4254, 4649, 4691
- JCAHO. *See* Joint Commission on Accreditation of Healthcare Organizations (JCAHO)
- JCMHC, see Joint Commission on Mental Health of Children (JCMHC)
- J-coupling, 330
- Jernberg, Ann, 3877
- Jessen, Peter Willers, 4644
- Jet lag, 2180–2181
- Jiu Wei Zhen Xin Granula (JWZXG), 2640
- Jobes, David, 2716
- Joffe, Walter, 1703
- Johns Hopkins University, 4652
- Johnson, Susan, 2948
- Johnson, Virginia, 2027, 4693
- Joint, reliability, 1092
- Joint Commission on Accreditation of Healthcare Organizations (JCAHO), 1158, 4741
- Joint Commission on Mental Health of Children (JCMHC), 4741
- Joint Commission on Mental Illness and Mental Health, 4692, 4741
- Jones, Ernest, 4741
- Jones, Howard, 784
- Jones, Maxwell, 2956
- Jordans, Mark, 2567
- Jorgensen, Christine, 2103
- Joy, 1669–1670
- Judgment
  - defined, 4741

  - disturbances of, 1171–1172
- Judgmental questions, 1002
- Jung, Carl, 422–423, 789, 949, 949*f*, 2685, 2750, 4691, 4741
  - on active imagination, 2984
  - personality theory, 950
    - archetypes, 950–951
    - complexes, 950
    - personality structure, 951–952
    - therapeutic process, 952
    - unconscious, 951
  - psychic apparatus, 951*f*
  - on religion, 2677
- Justice, 4012, 4741
- Justice-based ethics, 4616*t*
- Just right phenomenon, 3775
- Juvenile court, 4007–4008
- Juvenile justice, 4007–4009
  - blended sentences, 4009
  - competency to stand trial, 4008
  - court, 4007–4008
  - criminal responsibility assessments, 4008
  - interventions, 4008–4009, 4009
  - preadjudication examinations, 4008
  - racism in, 4110
  - sentencing evaluations, 4008–4009
  - waiver, 4008

K

- Kabat–Zinn, Jon, 2671, 2898
- Kahana, R. J., 2273
- Kahlbaum, Karl, 1644, 4656, 4657, 4686
- Kahn, Eugen, 4652

- Kainic acid, 520
- Kaiser-Wilhelm-Institut für Psychiatrie, 4649
- Kalinowsky, Lothar, 4651
- Kallio, Sakari, 2915
- Kamin, Leon, 434
- *Kami-untan-to* (KUT), 2640
- Kandel, Eric, 4651, 4696
- Kanner, Leo, 797, 3711, 3837, 4689
- Kanner syndrome. *See* Infantile autism
- Kant, Immanuel, 969, 4616
- Kaplan, Edith, 1029
- Kaplan–Meier survival function estimation technique, 910
- Kappa-opioid agonists, 1355
- Kappa statistic, 880
- Kardec, Allan, 1928
- Karolinska Institute, 4644
- Kasanin, Jacob, 1638
- Katz ADL Index, 4185
- Kaufman, Milissa, 1958
- Kaufman Assessment Battery for Children-III (K-ABC-III), 3548
- Kava (*Piper methysticum*), 2637–2638, 2638*f*
- Kefauver–Harris Amendments, 3022
- Keillor, Garrison, 785
- Kelly, George, 972–973
- Kempe, C. Henry, 3802
- Kennedy, John, 4512
- Kennedy, John F., 3632
- Kenner syndrome. *See* Infantile autism
- Keppra. *See* Levetiracetam (Keppra)
- Kernberg, Otto, 422, 927, 2751, 2930, 4695
- Kernel trick, 908
- Kertbeny, Károli Mária, 2061

- Ketalar. *See* Ketamine
- Ketamine (Ketalar, special K), 1350, 2909–2910, 3009, **3378–3385**
  - for acute bipolar depression, 1735
  - bipolar depression, 3382
  - chemistry, 3379
  - clinical issues, 2910
  - complications, 2910
  - contraindications, 2910
  - depersonalization/derealization and, 1954–1955
  - for depressive disorders, 185
  - difference from psychedelics, 2909
  - for dissociative identity disorder, 1997
  - dosage and administration, 3385, 3385*t*
  - drug interactions, 3384
  - history, 3378–3379
  - indications, 2910
  - limitations, 2910
  - major depressive disorder, 3380–3382
  - mechanism, 2910
  - mechanism of action, 3381*f*
  - metabolism, 3380*f*
  - molecular structure, 3379*f*, 3386*f*
  - organ/system effects of, 3382–3384
  - pain management in burn patients, 2470
  - pharmacodynamics, 3379–3380
  - pharmacokinetics, 3379
    - sex differences in, 3417
  - pharmacologic actions, 3379–3380
  - precautions and adverse reactions, 3383–3384, 3384*t*
  - principles and goals of therapy, 2910
  - psychopharmacology, 2909–2910
  - subjective effects, 2910

  - suicidality associated with TRD, 3382
  - therapeutic implications, 3383
- Khat, 1314
- Kiddie-Schedule for Affective Disorders and Schizophrenia (K-SADS), 885
- Kiddie Schedule for Affective Disorders and Schizophrenia—Computerized version (KSADS-COMP), 3540–3541
- Kidney cancer, 2363
- Kiesler, Donald, 1707, 2831
- Kigen, Dogen, 418
- Kihlstrom, John, 2915
- Kiloh, Lesley, 4659
- Kindling effect, 4741
- Kinesiology, 4741
- Kinesthetic hallucination, 4741
- Kinesthetic sense, 4741
- Kinetic Family Drawing (KFD), 3548
- King, Martin Luther, 2669
- Kin selection, social behavior and, 862
- Kin selection hypothesis, 2071
- Kinsey, Alfred, 2027, 2029, 2042, 2062, 4444
  - survey of male sexuality, 2064–2065
- Kinsey Scale (Heterosexual-Homosexual Rating Scale), 2058
- Kirkbride, Thomas, 4505, 4507*f*, 4510, 4686, 4741
- Kirmayer, Laurence, 2554
- Kirsch, Irving, 2916
- Klein, Donald, 4660
- Klein, Melanie, 4741
  - child psychiatry and, 3877
  - depressive position, 821–822
  - European Kleinian School, 2793
  - focal points, 805–806

  - on guilt, 1703
  - object relations theory and, 422, 805–806, 923–924
  - phantasy and, 812
  - projective identification, 2496
  - The *Psycho-Analysis of Children*, 4647
  - on psychosis, 2787
  - work of, 2956, 4689
- Kleine–Levin syndrome, 3011, 4741
- Kleinman, Arthur, 852, 2553, 2570
- Kleist, Karl, 486, 4657
- Kleitman, Nathaniel, 209, 4652
- Kleptomania, **2198–2200**
  - defined, 4741
  - diagnosis, 2198–2199, 4824*t*
  - DSM-5 criteria for, 2199*t*
  - epidemiology of, 2198–2199
  - etiology of, 2199
  - pharmacotherapy, 2199–2200
  - psychotherapy, 2200
  - treatment of, 2199–2200
- Klerman, Gerald, 1708, 3953, 4648
- Klonopin. *See* Clonazepam (Klonopin)
- Kluft, Richard P., 1932–1933, 1973, 1993
- Klüver, Heinrich, 34
- Klüver–Bucy syndrome, 475, 1188, 4742
- Knitzer, Jane, 3942
- Kohlberg, Lawrence, 790
- Kohut, Heinz, 422, 926–927, 2750, 2752, 4693, 4742
- Kolb, Lawrence, 4651
- Koo, John, 2431
- Kopeloff, Nicholas, 4659
- Koro, 4742

- Korsakoff, Sergei, 1283, 4657
- Korsakoff psychosis, 4742
- Korsakoff syndrome, 1282–1283, 1304
- Kozak, Michael, 1822–1823
- Kraepelin, Emil, 425, 486, 521, 1199, 1239, 1283, 1449, 1457, 1468, 1556, 1634, 1638, 1652, 1678, 2569, 4263, 4649, 4655, 4656, 4657, 4658, 4659, 4687, 4742
- Krafft-Ebing, Richard von, 2027, 2036, 2061, 2091, 2096, 2103, 2121, 4655, 4686
- Kral, Vojtech Adalbert, 4650
- Kramer, Henry, 4683
- Kretchmer, Ernest, 2257
- Kretschmer, Ernest, 4742
- Kreyòl Distress Inventory, 2568
- Krystal, Jonathan, 1953
- Kübler-Ross, Elizabeth, 2384
- *Kufungisisa*, 2568
- Kuhn, Roland, 3152, 4663
- Kupfer, David, 1658
- Kurfs disease, 689
- Kuru, 631, 637–638, 638*f*, 4742. *See also* Prion diseases
- Kynurenine (KYN) pathway, 180, 184

L

- La belle indifférence, 4742
- Labile, 4742
- Labile affect, 4742
- Labile mood, 4742
- Lability, 1676, 4742
- Laboratory testing, 3551–3552. *See also* Medical assessment and laboratory testing
- Laborit, Henri, 3055
- Lacan, Jacques, 4693

- Laconic speech, 4742
- Lactation. *See also* Breastfeeding
  - asenapine and, 3086
  - brexpiprazole and, 3104
  - clozapine and, 3070
  - lithium and, 3319–3320
  - lumateperone and, 3089
  - lurasidone and, 3064
  - melatonin receptor agonists and, 3395
  - olanzapine and, 3059
  - risperidone and, 3074
  - SSRIs and, 3143–3144
  - ziprasidone and, 3062
- Lactational amenorrhea method, 2594*t*
- Lacunar amnesia. *See* Localized amnesia
- LAGB. *See* Laparoscopic adjustable gastric banding (LAGB)
- Lalophobia, 4742
- Lalor, Joseph, 4643
- Lambert–Eaton myasthenic syndrome (LEMS), 661
- Lambo, Thomas A., 1469
- Lam Employment Absence and Productivity Scale (LEAPS), 1721
- Lamictal. *See* Lamotrigine
- Lamotrigine (Lamictal), 1066, **3295–3299**
  - for acute bipolar depression, 1734
  - for bipolar disorder, 1738
  - bipolar disorders, 3296, 3297
  - chemistry, 3295
  - dosage and administration, 3298–3299
  - drug interactions, 3298
  - for geriatric bipolar disorder, 4243
  - history, 3295, 4663
  - major depressive disorder, 3296–3297

  - molecular structure of, 3295*f*
  - organ/system effects of, 3297
  - pain syndromes, 3296
  - pharmacodynamics, 3295–3296
  - pharmacokinetics, 3295
  - pharmacologic actions, 3295–3296
  - precautions and adverse reactions, 3297–3298
  - pregnancy and, 2607–2608
  - seizure disorders, 3296, 3297
  - therapeutic indications, 3297
- Lancet Commission on Global Mental Health and Sustainable Development, 4681
- Landreth, Gary, 3877
- Lange, Carl Georg, 4644, 4655
- Langley Porter Clinic, 4654
- *Langley v. Coughlin*, 4628
- Language, 731
  - assessment, 3569
  - expressive, 4168
  - as mechanism of reflection, 820–821
  - receptive, 4168
- Language development, in children, 3492–3499
- Language disorder, 1182–1183, **3681–3687**. *See also* Communication disorders
  - clinical features, 3683–3685
  - comparative nosology, 3681
  - defined, 4742
  - definition of, 3681
  - diagnosis, 3683–3685, 4774*t*
  - differential diagnosis, 3685–3686
  - epidemiology of, 3681
  - etiology of, 3682–3683

  - laboratory examination, 3685
  - in older adults, 4134
  - pathology of, 3685
  - treatment of, 3686–3687
- Language network, 804
- Language therapy, for selective mutism, 3866
- Lanius, Ruth A., 1966
- Lanugo, 4742
- Laparoscopic adjustable gastric banding (LAGB), 2315–2316, 2315*f*
- Laparoscopic sleeve gastrectomy (LSG), 2316
- Lapsus linguae, 4742
- Larmor frequency, 327, 328*f*
- Larrey, Dominique-Jean, 2656
- La Salpetrière, 1928, 1929*f*
- Laségue, Charles, 1647, 2149, 4650
- Laser interstitial thermotherapy, 3483
- Lashley, Karl, 417
- Latah, 4742
- Late language emergence, 4742
- Late-life depression
  - bipolar disorder, 4230–4231
  - in cancer patients, 4229
  - depression with reversible dementia, 4230
  - in medically ill patients, 4228–4229
  - model of, 4235
  - in patients with dementia, 4229
  - psychotic depression, 4230
- Late luteal phase dysphoric disorder (LLPDD). *See* Premenstrual dysphoric disorder
- Late memory, 730*t*
- Latency phase, 4742

- Latency stage, 916, 919*t*
- Latent class growth analysis, 909–910
- Latent content, 4742
- Latent homosexuality, 4742
- Latent inhibition, 760
- Latent learning, 763
- Latent schizophrenia, 4742
- Late-onset hypothesis, 4232
- Late-onset schizophrenia (LOS), 4267–4268
  - antipsychotics for, 4307
  - course and prognosis, 4268
  - epidemiology of, 4267–4268
  - features of, 4269*t*
  - treatment of, 4268
- Lateral FPN network, 803
- Laterality, 731–732
- Lateralization, 459
- Lateral orbitofrontal cortex, 459
- Lateral ventricles, 15
- Laterodorsal tegmental (LDT) nuclei, 201, 202*f*, 208
- Latin America
  - biologic therapies in, 3443–3444
  - drug treatment/response in, 3443–3444
  - electroconvulsive therapy (ECT) in, 3444
  - magnetic seizure therapy (MST) in, 3444
  - mental disorders in, 3443
  - repetitive transcranial magnetic stimulation (rTMS) in, 3444
  - transcranial direct current stimulation (tCDS) in, 3444
  - treatment gap in, 3443
- Latino/a Racial Identity Orientation Model, 848
- Latuda. *See* Lurasidone (Latuda)
- Lauterbur, Paul, 327, 1510–1521

- Lavender (*Lavandula officinalis*), 2638
- Law of effect, 757, 757*f*
- *Lawrence and Garner v. Texas*, 2076
- Lawton–Brody IADL scale, 4185
- Laycock, Thomas, 4686
- Lazabemide (Pakio), 3185*t*
- Lazarus, Arnold, 2735, 3885
- Lazell, Edward W., 2955
- LBD. *See* Lewy body disease (LBD)
- L-deprenyl (Eldepryl), 3185
- Lead, impact on CNS and PNS, 655*t*, 658
- Learned autonomic control, 4742
- Learned helplessness model, 765, 1655
- Learned mindfulness model, 2536–2537
- Learning, 801, 2959*t*
  - biologically relevant, 1812
  - cognitive aging, 4163
  - hierarchical model of, 2220*f*
  - interactive systems of, 2213
  - neuropsychological evaluation, 4169–4170
  - sleep and, 213
  - specific learning disorder (*See* Specific learning disorder (SLD))
  - stimulant-related disorders and, 1315–1317
- Learning disorders, 4742
- Learning theories, **434–440**, **755–767**
  - addiction and, 438–439
  - anxiety and, 439–440
  - basic concepts, 756–757
  - brief history, 434–439
  - defined, 4742
  - hallucinations and, 437–438
  - historical context, 756

  - negative symptoms and, 438
  - operant/instrumental learning, 761–764
    - behavior and payoff, 761–763
    - motivation, 763–764
    - reinforcement, 763
  - Pavlovian conditioning, 758–761
    - and behavior, 758–759
    - erasing of, 760–761
    - learning process, 759–760
  - Pavlovian plus operant, 764–767
    - avoidance learning, 764–765
    - instrumental action, 764–765
  - psychosis and, 435–437
- LEARN model, 854–855, 856*t*
- Leary, Timothy, 2676
- Least restrictive environment (LRE), 4026*t*
- Least significant difference (LSD) procedure, 900
- Leber hereditary optic neuropathy (LHON), 703
- LeBois, Lauren, 1958
- Lebovici, Serge, 3775
- LeCron, Leslie, 2923
- LeDoux, Joseph, 813
- Leg cramps, sleep-related, 2186
- Lehmann, Heinz, 4649, 4651, 4662
- Leibniz, Gottfried, 435
- Leighton, Alexander, 784, 787
- Leighton, Dorothea, 787
- Lennox-Gastaut syndrome, 3240
- Leonhard, Karl, 1652, 4657
- Leprosy, 660
- Leprous neuropathy, 660
- Leptin, **169**, 226

- Leptospirosis (Weil disease), 625
- Leriche syndrome, 2055
- Lesbian, gay male, bisexual, and transgender (LGBT), 2058
  - acquisition of gay and lesbian identities, 2073
  - alcohol/substance abuse ion, 2083–2084
  - in correctional facilities, 4639
  - LGBTQIA youth, 4100–4103
    - intersectional care for, 4102
    - mental health disparities, 4102
    - resilience, 4102–4103
    - sex and gender, 4100–4102
    - sex development, 4101–4102
    - sexual orientation, 4099–4100, 4100*t*
    - trauma, 4102
    - trauma-informed affirming care for, 4103
  - and long-term care, 4453–4454
  - medical illness, 2084–2085
  - mood disorders in, 1664
  - older adults, 4443–4445
  - as parents, 2080–2081
  - psychoanalytic approaches to gay and lesbian patients, 2086
  - same-sex marriage, 2080
  - sex trafficking and, 3978
  - suicide in, 2083
- Lesbian and gay studies, 2060
- Lesbianism, 4742
- Lesbian(s)
  - acquisition of lesbian identity, 2073
  - definition of, 2059
- Lesch–Nyhan syndrome, 1191, 3642*t*, 4742
- Lethal catatonia, 1476
- Lethologica, 4742

- Leukocytosis, 3322
- Leukodystrophies, 582
- Leukopenioa, 3086
- Leukotomy. *See* Lobotomy
- Leuner, Hanscar, 432, 2984
- Leuprolide (Eligard), 2098
- Level 2 Substance Abuse Module from the DSM-5, 1110–1111, 1110*t*
- Levetiracetam (Keppra), 3271–3273
  - chemistry, 3272
  - clinical drug studies, 3272
  - dosage and administration, 3273
  - drug interactions, 3272–3273
  - history, 3271–3272
  - neurologic disorders, 3272
  - pharmacodynamics, 3272
  - pharmacokinetics, 3272
  - precautions and adverse reactions, 3272
  - psychiatric disorders, 3272
  - seizure disorders, 3272
  - for tardive dyskinesia, 3096
  - therapeutic indications, 3272
- Levine, Ruth, 2552
- Levodopa, 561
- Levomilnacipran
  - characteristics of, 3150
  - for depression in later life, 4297
  - dosage and administration, 3150
- Levothyroxine (Levoxyl), 1153*t*, 1159*t*, 2336, 3397, 3409
- Levoxyl. *See* Levothyroxine (Levoxyl)
- Lewin, Kurt, 2887, 2956, 4742
- Lewinson, Peter, 1655

- Lewis, Aubrey, 1201, 4650, 4659
- Lewis, Judd, 3837
- Lewis, Nolan, 4651
- Lewy, Friedrich Heinrich, 4658
- Lewy body disease (LBD), 4658, 4742
  - dopamine receptor antagonists for, 3045
- Lexapro. *See* Escitalopram (Lexapro)
- Leyton Obsessional Inventory– Child Version (LOI-CV), 3546
- LGBT (lesbian, gay male, bisexual, and transgender), 2058
  - acquisition of gay and lesbian identities, 2073
  - alcohol/substance abuse ion, 2083–2084
  - in correctional facilities, 4639
  - LGBTQIA youth, 4100–4103
    - intersectional care for, 4102
    - mental health disparities, 4102
    - resilience, 4102–4103
    - sex and gender, 4100–4102
    - sex development, 4101–4102
    - sexual orientation, 4099–4100, 4100*t*
    - trauma, 4102
    - trauma-informed affirming care for, 4103
  - and long-term care, 4453–4454
  - medical illness, 2084–2085
  - mood disorders in, 1664
  - older adults, 4443–4445
  - as parents, 2080–2081
  - psychoanalytic approaches to gay and lesbian patients, 2086
  - same-sex marriage, 2080
  - sex trafficking and, 3978
  - suicide in, 2083
- Liability-threshold model, 3587*t*

- Liaison Committee for Graduate Medical Education (LCGME), 4588, 4590
- Liaison Committee on Medical Education (LCME), 4742
- Liaison nursing, 4743
- Liaison psychiatry. *See* Consultation-liaison psychiatry
- Liang, King-Yee, 909
- Libet, Benjamin, 429
- Libido, 4743
- Librium. *See* Chlordiazepoxide (Librium)
- Lichtenberg, Joseph, 3877
- Lichtheim, Ludwig, 458
- Lidz, Theodore, 2943–2944
- Lieberman, Alicia, 3877
- Liepmann, Hugo, 458
- Life cycle family therapy, 2947
- Life cycle stages, 937–940
  - autonomy *vs.* shame and doubt, 938
  - generativity *vs.* stagnation, 939–940
  - identity *vs.* role confusion, 939
  - industry *vs.* inferiority, 939
  - initiative *vs.* guilt, 938
  - integrity *vs.* despair, 940
  - intimacy *vs.* isolation, 939
  - trust *vs.* mistrust, 938
- Life events, 1163–1164, 1660
- Life instinct, 4743
- Life-review therapy, 4352–4353
- Life Skills Profile (LSP), 4553
- Life table methods, 910
- Lifetime expectancy, 4743
- Lifetime prevalence, 874–875, 4743
- Lifton, Robert J., 2019, 4657

- Ligand-gated receptors, 2997–2998, 2998*f*
- Light, 2644
- Light therapy, 2181
- Likely responder method of analysis, 903
- Likert scale, 1091
- Lilliputian hallucination, 4743
- Lilliputian hallucinations, 1455
- Limb apraxia, 1021
- Limbic cortex, 34–35
- Limbic encephalitis, 2365
- Limbic leucotomy, 3484
- Limbic predominant age-related TDP-43 encephalopathy, 1272
  - biomarkers, 1272
  - clinical features, 1272
  - course and prognosis, 1272
  - definition of, 1272
  - diagnosis, 1272
  - differential diagnosis, 1272
  - epidemiology of, 1272
  - genetics of, 1272
  - neuropathology, 1272
- Limbic system, 34–42, 460
  - definition of, 4743
  - functional neural circuitry, 39–40, 40*f*
  - major structures, 34–39, 34*f*
    - amygdala, 36, 38*f*
    - hippocampal formation, 35, 36*f*, 37*f*
    - hypothalamus, 36–39
    - limbic cortex, 34–35
    - septal area, 36
- Lindenmann, Erich, 4653
- Linehan, Marsha, 4648

- Linkage, 254
- Linkage disequilibrium (LD), 254, 3587*t*
- Link function, 908
- Linneaus, Carl, 3630
- Liothyronine (Cytomel), 1435, 3397, 3405
- Lipids, 1064
- Lipofuscin, 4153
- Lipowski, Zbigniew, 1890, 2274, 4255
- Lipton, Morris, 4653
- Liraglutide, for weight loss, 2315
- Lisdexamfetamine (Vyvanse)
  - for depression, in cancer patients, 2357
  - pharmacokinetics, 289
- *Listeria*, 620
- Lithium (Carbolith, Eskalith), 1066–1067, 2995, **3313–3324**
  - in acute antidepressant treatment, 3323
  - for acute bipolar depression, 1733
  - for acute mania, 1729
  - aggressivity and, 3323
  - Alzheimer disease, 3320
  - in amphetamine, methylphenidate, or cocaine abuse, 3320
  - animal models, 3318–3319
  - for bipolar disorder, 1737, 3032, 3315–3316, 3316*f*, 3320–3321, 3929–3930
  - chemistry, 3260*t*, 3314
  - childbirth and, 3319–3320
  - childhood use, 3320
  - in cluster headache, 3320
  - for early-onset bipolar disorder, 3834
  - EEG effects, 379
  - effects of, 3318
  - episode-prophylactic effect, 1717

- for geriatric bipolar disorder, 4242
- geriatric use, 3320
- history, 3314, 3318*t*, 4646
- lactation and, 3319–3320
- in later life, 4300–4301
- mechanisms of action, 3318–3319, 3318*t*
- for mood dysregulation in children and adolescents, 3819
- patient history and, 3318
- pharmacodynamics, 3314–3315
- pharmacokinetics, 3314–3315
- polypharmacy and administration of, 3316–3318
- precautions, 3309–3311
- pregnancy and, 2607, 3319–3320
- in schizoaffective disorder, 3323
- for self-injurious behaviors (SIBs), 3813
- sexual dysfunction and, 2047
- side effects, 3322
  - chronic renal failure, 3322
  - dental, 3322
  - dermatologic, 3322–3323
  - leukocytosis, 3322
  - parathyroid, 3322
  - polydipsia, 3322
  - polyuria, 3322
  - thyroid, 3321
  - tremor, 3322
- for substance use disorder, 4001*t*
- suicidality and, 3319
- for suicide, 2712
- for suicide prevention, 1717
- therapeutic indications, 3315–3316
- toxicity, 3321

  - in unipolar depression maintenance treatment, 3323
- Lithium carbonate, 4743
- Lithium tremor, β-adrenergic receptor antagonists for, 3227
- Little Albert, 4743
- Little Hans, 4743
- Liver
  - nalmefene effects on, 3364
  - naltrexone effects on, 3364
- Liver waves, 373
- Livi, Carlo, 4642
- Living donor, 2454–2455, 2454*t*
- Living will, 2401, 4743
- Llinás, Rodolfo, 419
- Lloyd, Edward, 1324
- Lobotomy, 4743
- Local circuit neurons, 3
- Local educational agency, 4026*t*
- Local field potentials, 308–312
- local field potentials (LFPs), 804
- Localized amnesia, 4743
- Local sleep, 210
- Lock, Margaret, 2553
- Locke, John, 434, 2512
- Locus ceruleus, 4743
- Locus coeruleus, 6
- Locus heterogeneity, 252
- Loeb, James, 4649
- Loevinger, Jane, 790
- Loewald, Hans, 2752
- Logistic regression, 907
- Log linear models, 908
- Logopenic aphasia, 1021*t*

- Logorrhea, 4743
- Logotherapy, 4743
- Lombroso, Cesare, 4687
- Long-acting injectables, 1578
- Long association fibers, 13
- Long interspersed elements (LINES), 241
- Longitudinal models, 909
- Longitudinal study, 4743
- Long non-coding RNAs (lncRNAs), 2229
- Long-term care (LTC)
  - family interventions and, 4376–4377
  - LGBTQI+ individuals in, 4453–4454
  - psychiatric aspects of, **4402–4414**
    - agitation, 4405*t*
    - anxiety disorders, 4414
    - chronic pain, 4408
    - delirium and comorbid medical conditions, 4404–4408, 4407*t*
    - dementias, 4408–4412
    - depressive disorders, 4412–4413
    - loneliness and trauma, 4409*t*
    - malnutrition, 4407–4408
    - management of disorders
      - action plan, 4404
      - safety, 4404
      - team-based evaluation, 4404
      - treatment response, 4404
    - palliative care, 4414
    - psychiatric symptoms, 4406*t*
    - psychopharmacology, 4407*t*
    - psychotic disorders, 4413
    - PTSD, 4414
    - sleep-wake cycle disorders, 4413–4414

  - psychiatric consultant, 4405*t*
  - psychiatric disorders in, 4403*t*
  - sexual behavior in, 4452–4455
  - training programs, 4410*t*
- Long-term depression (LTD), 780
- Long-term explicit memory network, 804
- Long terminal repeats (LTRs), 241
- Long-term memory, 4743
- Long-term potentiation (LTP), 779–780
- Looking-glass self, 421
- Loosening of associations, 1462, 4743. *See also* Tangentiality
- Lopez-Ibor, Juan José, 1889–1890, 4644
- Lopressor. *See* Metoprolol
- Lorazepam (Ativan)
  - for agitation, 2724, 3576*t*
  - for delirium, 4261
  - dosage, 1429*t*
  - for geriatric bipolar disorder, 4243
  - pharmacokinetics of, sex differences in, 3417
- Lord's Resistance Army (LRA), 2021
- LORETA. *See* Low-resolution brain electromagnetic tomography (LORETA)
- Loss, response to, 1703
- Lou Gehrig disease. *See* Amyotrophic lateral sclerosis (ALS)
- Louisville Behavior Checklist (LBC), 3543
- Lourie, Reginald, 3808
- Love, 2037
- Low back pain (LBP), 2482–2483
- Low birth weight, 668–669
- Low-pass filter, 361
- Low-resolution brain electromagnetic tomography (LORETA), 368
- Loyalty conflict, 2545

- LP. *See* Lumbar puncture (LP)
- LSD (lysergic acid diethylamide), 431–432, 2642, 2673, 2903
  - for alcohol use disorder, 2907
  - chemical structure of, 3386*f*
  - defined, 4743
  - effects of, 1353–1354
  - historical overview, 1348
  - history, 4662
  - microdosing, 3390
  - for patients with life-threatening illness, 2906–2907
  - spirituality and, 2673
  - tolerance, 1354
  - for treatment-resistant depression, 2906
- LSG. *See* Laparoscopic sleeve gastrectomy (LSG)
- LTC. *See* Long-term care (LTC)
- L-theanine, 2634
- L-tryptophan, 2633–2634
- Luborsky. Lester, 787
- Lubrano, Aldo, 4655–4656
- Ludiomil. *See* Maprotiline (Ludiomil)
- Luhmann, Niklas, 419
- Lumateperone, 3087–3089
  - for bipolar disorder, 1730*t*
  - chemistry, 3087
  - dosage and administration, 3088
  - mechanism of action, 3088
  - pharmacodynamics, 3088
  - pharmacokinetics, 3087
  - precautions and adverse reactions, 3088–3089
  - for schizophrenia, 3088
  - therapeutic indications, 3088
- Lumbar puncture (LP), 4743

- Luminal. *See* Phenobarbital (Luminal)
- Lumodil. *See* Maprotiline (Ludiomil)
- Lung cancer, 2362–2363
- Lurasidone (Latuda), 3062–3064
  - for bipolar disorder, 1730*t*, 1739
  - chemistry, 3062
  - dosage and administration, 3064
  - drug interactions, 3064
  - for geriatric bipolar disorder, 4243
  - molecular structure, 3056*f*
  - pharmacodynamics, 3062–3063
  - pharmacokinetics, 3062
  - precautions and adverse reactions, 3063–3064
  - for schizophrenia, 3062–3064
  - special populations, 3064
  - therapeutic indications, 3063
- Luria-Nebraska Neuropsychological Battery (LNNB), 4743
- Lust dynamism, 4743
- Luteinizing hormone (LH), 2068, 2138, 2583–2584
- Luteinizing hormone–releasing hormone (LHRH), 2054, 2098
- Luther, Martin, 2669, 2669*f*
- Luton, Frank, 4654
- Luvox. *See* Fluvoxamine (Luvox)
- Lying, 4743
- Lyme disease, 623–625, 660
  - chronic, 624
  - clinical presentation, 623–624
  - diagnosis, 624
  - epidemiology of, 623
  - management of, 624–625
  - pathophysiology of, 623
  - post-treatment Lyme syndrome, 624

- Lymphoma, 607
- Lypemania, 1652
- Lyrica. *See* Pregabalin (Lyrica)
- Lysergic acid diethylamide. *See* LSD (lysergic acid diethylamide)
- Lysosomal disorders, 686
- Lysosomal storage diseases, 3642*t*
- Lyssenko, Lisa, 1960–1961

M

- MacArthur Story-Stem Battery (MSSB), 3556
- MacFarlane, Jean, 784
- Machine learning (ML), 907–908, 1553–1554
- Macht, David, 4661–4662
- MacLean, Paul, 34
- Macropsia, 4743
- Madanes, Cloe, 2947
- Madonna-Putana complex, 2042
- Mad prodromal laughter (*fou rire prodromique*), 475
- Madrid Declaration, 4680–4681
- *Madrid v. Gomez*, 4628
- MADRS (Montgomery-Asberg Depression Rating Scale), 1104
- Mafia, 2509
- Magical thinking, 4743
- Magnan, Valentin, 4650, 4656, 4657
- Magnetic dipole, 327
- Magnetic diseases, 1927
- Magnetic resonance imaging (MRI). *See also* Functional magnetic resonance imaging (fMRI)
  - in aging adult (*See* Geriatric psychiatry)
  - alcohol dependence, 341–342
  - basic physics of, 327–330
  - of blood flow, 337–338
  - brain imaging, 1061
  - communication with patients, 351–352
  - contrast agents, 336
  - defined, 4743
  - dementia, 339–340
  - diffusion, 336–337
  - Fourier transform, 332–335, 334*f*, 335*f*

  - gradient-echo, 333
  - history, 4660–4661
  - image contrast, 333
  - methodology, 333–339
  - mood disorders and, 341
  - neuroimaging and, 538
  - safety issues, 351
  - schizophrenia, 340–341
    - structural, 1512–1520
  - spin-echo, 333
  - system hardware, 330–332
  - T_1 weighting, 336
  - T_2 weighting, 336
- Magnetic resonance spectroscopic imaging (MRSI), 346
- Magnetic resonance spectroscopy (MRS)
  - applications in psychiatry, 347–350
    - alcohol dependence, 349–350
    - dementia, 348
    - mood disorders, 348–349
    - schizophrenia, 348
  - basic physics of, 327–330
  - ^{13}C, 347, 347*f*
  - communication with patients, 351–352
  - ^{7}F, 350
  - ^{19}F, 350
  - Fourier transform, 332–335, 334*f*, 335*f*
  - future directions, 352–353
  - ^{1}H, 345–346
  - measurements, 330
  - neuroimaging and, 538
  - ^{31}P, 346–347

  - safety issues, 351
  - system hardware, 330–332
- Magnetic seizure therapy (MST), 3474–3476
  - for acute bipolar depression, 1735
  - clinical studies, 3475
  - current status and treatment algorithms, 3475
  - definition, 3474–3476
  - definition of, 3464*t*
  - future directions, 3475–3476
  - history, 3474
  - in Latin America, 3444
  - mechanism of action, 3474
  - for older adults, 4339
  - patient selection, 3475
  - side effects, 3475
- Magnetization, 327–328, 329*f*
- Magnetizers, 1927
- Mahler, Margaret, 422, 3500, 3775, 4693, 4743
  - developmental model, 922–923
- Main, Mary, 928, 3501, 3877
- Main, Thomas, 4648
- Maintaining factors, 2815
- Maintenance drug therapy, 4743
- Maintenance of Certification (MOC), 4590
- Maintenance of wakefulness test (MWT), 204, 2189
- Major, Herman, 4644
- Major affective disorder. *See* Major depression
- Major depression
  - deep brain stimulation, 3485
  - electroconvulsive therapy, 3449–3450
  - epidemiology of, 1663
  - functional connectivity abnormalities in, 415

    - amygdala connectivity, 413, 413*f*
    - default mode network connectivity, 414–415
    - striatal connectivity, 415
  - gene mapping, 265–266
  - genetic overlap with other disorders, 267
  - HIV infection and, 589–594
  - immune-CNS interactions in, 181–185
  - inflammation-depression association, 182–183
    - cross-sectional, 182
    - longitudinal, 182–183
  - nefazodone, 3205
  - psychoanalytic descriptions of, 1703–1704
  - translocator protein and, 404
- Major depression with psychotic features
  - differential diagnosis, 3871
  - symptoms in children and adolescents, 3868*t*
- Major depressive disorder (unipolar depression), 1649, 1678–1684. *See also* Bipolar disorder(s)
  - β-adrenergic receptor antagonists for, 3227
  - anabolic steroid-induced, 1443
  - atypical features, 1686–1687
  - bipolar depression *versus*, 1686*t*
  - for bipolar disorder, 3407–3409
  - bipolar disorder *versus*, 1696–1697, 1696*t*
  - buspirone for, 3172
  - chronic depression, 1687
  - combination therapy for, 3404–3407
    - antidepressant–antipsychotic combinations, 3404
    - esketamine/ketamine–oral antidepressant combination, 3406
    - lithium–antidepressant combinations, 3404–3405
    - sex steroid–antidepressant combinations, 3406

  - stimulants/dopamine agonist–antidepressant combinations, 3405–3406
  - thyroid supplement–antidepressant combinations, 3405
- comorbid physical disease, 1678–1680
- defined, 4743
- diagnosis, 4791–4792*t*
- diagnostic threshold, 1680–1686
- DSM-5 criteria for, 1679–1680*t*
- epidemiology of, 1663
- hypnosis and, 2930
- ICD-11 criteria for, 1679–1680*t*
- insomnia comorbidty, 2168
- Internet gaming disorder and, 4091–4092
- lamotrigine for, 3296–3297
- lithium for, 3323
- melancholic features, 1686
- mixed features, 1687
- monoamine oxidase inhibitors, 3193–3194
- neurotrophins and, 143–144
- pharmacotherapy for, 3928–3929
  - dopamine receptor agonists, 3109–3110
- psychodynamic psychotherapy for, 2754
- psychopharmacology of, 1719–1725
- psychotherapy
  - combined psychotherapy and pharmacotherapy, 1746–1747
  - group and couples, 1746
  - individual, 1745–1746
- psychotic features, 1687
- recurrent subtype, 1686
- single-episode subtype, 1686
- SSRIs for, 3130–3131
- suicide and, 2708

    - trazodone, 3198–3199
    - treatment of, 1719–1725
        - adherence evaluation, 1720
        - antidepressants, 1720*t*, 1721–1725
            - first-line, 1722
            - second-line, 1723
            - third-line, 1723, 1724
        - digital tools, 1721
        - foundations of management, 1720–1721
        - measurement-based, 1720
        - remission maintenance, 1725
        - sequencing and staging, 1721
    - treatment-resistant, 3058–3060
- Major depressive disorder with melancholic features, 1653
- Major depressive episode (MDE), 4744
    - uncomplicated bereavement and, 2548
- Major epilepsy. *See* Grand mal epilepsy
- Major histocompatibility complex (MHC), 90
- Major motor seizure. *See* Grand mal epilepsy
- Major neurocognitive disorders. *See also* Dementia
    - abuse and, 1249
    - alcohol use disorder and, 1247
    - caregiver support and, 1249
    - clinical features, 1241–1242
    - cognitive domain evaluation, 1243–1244, 1244*t*
    - cognitive impairment in, 1242–1243
    - comparative nosology, 1239–1240
    - corticobasal degeneration, 1263
    - course and prognosis of, 1247–1248
    - COVID-19 and, 1252
    - CSF studies, 1246
    - defined, 4744

- dementia in Parkinson disease, 1269–1271
- dementia with Lewy bodies, 1265–1269
- depression and, 1251
- diagnosis, 4829–4832*t*
- differential diagnosis, 1246–1247
- driving and, 1249
- DSM-5 criteria for, 1239–1240, 1242*t*
- end-of-life issues, 1249
- epidemiology of, 1240
- etiology of, 1241, 1241*t*
- evaluation for, 1243
- factitious disorder and, 1247
- frontotemporal dementia, 1259–1262
- functional impairment in, 1243
- health disparities in, 1252
- inhalant-induced, 1379, 1379–1380
- laboratory examination, 1245–1246, 1246*t*
- limbic predominant age-related TDP-43 encephalopathy, 1272
- management of
  - nonpharmacologic, 1248–1249
  - pharmacologic, 1249–1250
- medical conditions and, 1247
- medications and, 1247
- molecular neuroimaging, 1246
- neglect and, 1249
- neurodegenerative disorders and, 1247
- neuropathology, 1246
- neuropsychiatric manifestations, 1244–1245
- neuropsychiatric manifestations, treatment of, 1250–1251
- neuropsychological testing, 1246
- in older adults, 4427–4428
- progressive supranuclear palsy, 1263–1265

  - psychosis and, 1250
  - risk and protective factors, 1240–1241
  - schizophrenia and, 1247
  - screening, 1243
  - severity measures, 1244
  - sleep disturbance in, 1251
  - structural neuroimaging, 1245–1246
- Major tranquilizers. *See* Antipsychotics; Neuroleptics
- Make-a-Picture Story (MAPS), 1045*t*
- Maladaptive, 4744
- Maladaptive behavior, intellectual disabilities and, 3648
- Maladaptive styles, 979–980
- *Maladi bon Dieu*, 2551
- *Maladi dyab*, 2551
- Malan, David, 2767–2768
- Malaria, 626–627
- Male erectile disorder, 2041–2042, 4744. *See also* Erectile dysfunction (ED)
  - DSM-5 criteria for, 2041*t*
  - due to a general medical condition, 2044–2045, 2045*t*
- Male hypoactive sexual desire disorder, 2038–2040, 2040*t*, 4744, 4819–4820*t*
- Maleness, 4744
- Male prostheses, 2055–2056
- Male sexual response cycle, 2034*f*, 2034*t*
- Malingering, 2492–2508
  - antisocial personality disorder and, 2503–2504
  - avoiding responsibility, 2496–2497
  - clinical features, 2495–2496
  - clinical presentation
    - ADHD, 2502–2503
    - Ganser syndrome, 2502

- clinical presentations of, 2498–2504, 2498*t*
  - amnesia, 2501
  - borderline intellectual functioning, 2502
  - depression, 2500–2501
  - intellectual disabilities, 2502
  - neurocognitive presentations, 2499
  - pain, 2499
  - psychogenic nonepileptic seizures, 2499
  - psychological symptomatology, 2499–2503
  - psychosis, 2501–2502
  - PTSD, 2500
- coding of, 2504, 2504*t*
- collateral diagnostic data, 2505–2506
- comparative nosology, 2494
- course and prognosis, 2507
- criminal contexts, 2497
- defined, 4744
- definitions, 2492–2494
- diagnosis of, 2495–2496
- differential diagnoses, 2506, 2506*t*
- DSM-5 criteria for, 2494
- emergency facilities and, 2498
- epidemiology of, 2495
- etiology of, 2495
- forensic contexts, 2497
- history in literature, 2494–2495
- ICD-11 criteria for, 2494
- illness falsification and, 2503
- medicolegal contexts, 2497
- nonmalingerer’s dilemma and, 2503–2504
- outpatient settings, 2498
- projective identification, 2496

  - treatment and management of, 2507
  - truth and, 2496
- Malingering amnesia, 2007
- Malingering by proxy, 2503
- Malingering Probability Scale (MPS), 2940
- Malnutrition, in long-term care, 4407–4408
- Malpractice, 2717–2718, 4610–4611, 4744
- Maltreatment, 1759
- Maltreatment (child), **3968–3980**. *See also* Child abuse; Neglect; Psychological abuse
  - allegations of abuse, 3976–3977
  - child interviews, 3974–3975
    - content of, 3974
    - extended forensic interview, 3974–3975
    - process, 3974
    - protocol, 3974–3975
  - clinical features, 3971–3972
  - collateral information, 3974
  - comparative nosology, 3970
  - cultural variation in, 3978
  - definitions of, 3968–3969
  - denials of abuse, 3976
  - diagnosis, 3971–3972
  - differential diagnosis, 3976–3977
    - physical findings, 3976
    - psychiatric considerations, 3976–3977
  - epidemiology of, 3970–3971
  - evaluation of, 3973–3975
  - federal law definition, 3968
  - future directions, 3980
  - genetics, 3972
  - history, 3969

  - laboratory examination, 3976
  - long-term consequences of, 3972
  - multidisciplinary team, 3978
  - neglect, 3968
  - neurobiologic aspects of, 3972–3973
  - neuroimaging and, 3972–3973
  - parent interviews, 3973–3974
  - perpetrator substitution, 3977
  - physical abuse, 3968–3969, 3971
    - evaluation of, 3976
    - physical examination, 3975
  - physical examination, 3975–3976
  - prevention
    - primary, 3979–3980
    - secondary, 3980
  - psychological abuse, 3969
  - religious practices and, 3978
  - reporting of, 3978–3979
  - risk factors, 3971
  - sexual abuse, 3969, 3971–3972
    - evaluation of, 3976
    - physical examination, 3975–3976
  - special issues in, 3977–3978
  - state law definition, 3968
  - treatment of
    - of child, 3979
    - of parents, 3979
- Mammillary bodies, 38
- Manerix. *See* Moclobemide (Manerix)
- Mania, 475, 1650–1651. *See also* Somatic symptom and related disorders
  - acute, 1690–1691

        - clinical recommendations, 1731–1733
        - management in an emergency setting, 1731–1732
        - pharmacological treatment of, 1732–1733, 1732*t*
        - treatment options, 1728–1731
        - treatment selection, 1732
        - valproate for, 3301–3303, 3302*t*
    - animal models of, 299
    - bipolar, 3061
    - chronic, 1691
    - defined, 4744
    - differential diagnosis, 2722*t*
    - dopamine receptor antagonists for, 3044
    - electroconvulsive therapy, 3451
    - history, 4656
    - hypomania *versus*, 1677–1678
    - language disorders and, 1182–1183
    - multiple sclerosis and, 580
    - organ transplantation and, 2458
    - poststroke
        - assessment, 492–493
        - course and prognosis, 497
        - etiology, 491
        - treatment of, 499
    - psychoanalytic formulations of, 1705
    - secondary, 1691
    - traumatic brain injury and, 683
    - treatment-resistant, 3068
- Maniacal grief, 1670
- Manic, 4744
- Manic–depressive illness, 4744
- Manic depressive psychosis, 1449, 1649
- Manic syndrome, 1676–1678

- Manifest content, 4744
- Manipulation, 4744
- Mannerism, 4744
- Mansfield, Peter, 327, 1510–1521
- Mantra, 4744
- MAOIs. *See* Monoamine oxidase inhibitors (MAOIs)
- Maple syrup urine disease (MSUD), 693
- Mapother, Edward, 4650
- Maprotiline (Ludiomil), 98, 3152
- Marasmus, 4744
- Marathon, 4744
- Marginal distributions, 897
- Margins, 908
- Marijuana, 4021. *See also* Cannabis; Cannabis use and cannabis-related disorders
  - adolescents and, 3999
  - defined, 4744
  - EEG effects, 380
  - epidemiology, 4285
  - etiology, 4285
  - older adults and, 4285–4286
  - pharmacologic treatment, 4286
  - risks, 4285–4286
  - screening, 4285
  - withdrawal, 4286
- Marital counseling, 4744
- Marital problems, 4744
- *Marital Tensions* (Dick), 2949
- Marital therapy. *See* Marriage therapy
- Marketers, 959
- Markowitsch, Hans J., 1956, 2003
- Marmer, Steven, 1966

- Marmor, Judd, 2062
- Marplan. *See* Isocarboxazid
- Marriage therapy, 4744. *See also* Couples therapy
- Marsh, L. Cody, 2955
- Marsilid. *See* Iproniazid (Marsilid)
- Martin, Emily, 2883–2884
- Martland, Harrison Stanford, 4658
- Masculine identity, 4744
- Masculine protest, 4744
- Masculinity–femininity scale, 4744
- *The Mask of Sanity* (Cleckley), 2519
- Maslow, Abraham, 972, 2669, 2887
- Masochism, 4744
- Masochistic character, 955
- Massachusetts General Hospital Utility Multi-Programming System (MUMPs), 2699
- Massachusetts Mental Health Center, 4653
- Massage therapy, 2642
- Massively parallel sequencing. *See* Next-generation sequencing (NGS)
- Mass shootings, 2513–2514
- Masters, William, 2027, 4693
- Masterson, James, 4695
- Masturbation, 2035–2036, 4744
- Masturbatory pain, 2049–2050
- Mata y Fontanet, Pedro, 4644
- Material me, 420
- Maternal behavior, epigenetics of, 278–279
- Maternal deprivation, 4744
- Maternal gestational stress, 2128
- Maternally inherited Leigh syndrome (MILS), 703
- Maternal separation, 2128–2129

- Maternal transference, 2749
- Maternity blues, 1670
- Mathematical disability. *See* Specific learning disorder (SLD)
- Mathematical models, of behavior, 322–323
- Mattis Dementia Rating Scale-Second Edition (DRS-2), 4168
- Maturana, Humberto, 419
- Maturational crises, 4745
- Mature defenses, 921*t*, 2748*t*
- Maturity, 785, 790–791, 790*f*
- Maudsley, Henry, 1652, 4650
- Maudsley anorexia treatment for adults (MANTRA), 2161–2162
- Maudsley Hospital, 4643, 4650, 4659
- Maudsley method, 2161
- Maximization step (M step), 905–906
- Maximum security unit, 4745
- May, Rollo, 2980
- Mayer, Adolf, 2414–2415
- Mayer-Gross, Wilhelm, 1931, 4649
- McAdams, Dan, 973
- McConkey, Kevin, 2914
- McCulloch–Pitts networks, 435
- McCullough, James, 1707, 2819
- McGill Illness Narrative Interview (MINI), 2554, 2571
- McGill Pain Questionnaire (MPQ), 1044*t*
- McHugh, Paul, 1928
- McKinley, J. Charnley, 4659
- McKinney. William, 1658
- Mclean Hospital, 4652
- McNamara, Ashley, 2879
- MDD. *See* Major depressive disorder (unipolar depression)
- MDMA. *See* 3,4-Methylenedioxymethamphetamine (MDMA)
- Mead, Margaret, 798, 2512

- Meadow, Roy, 1908
- Meal(s). *See also* Nutrition
  - as functional unit of eating, 220
  - maintenance, 222–224
  - signals for initiation, 220–222
- Mean, 896–897, 4745
- Mean deviation, 4745
- Meaning in life, 4473
- Mean/median substitution, 905
- Meares, Russell, 2930
- Measles, 619–620
- Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS), 4550
- Measurement instruments, 879–881
  - reliability, 879–880, 880*t*
  - validity, 880–881, 880*t*
- MECA. *See* Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA)
- Meclizine (Antivert), 107
- Medial diencephalon, 768
- Medial forebrain bundle (MFB), 1761*f*
- Medial frontoparietal network (FPN), 803
- Medial orbitofrontal cortex, 459
- Median, 896–897, 4745
- Mediation, 877
- Medicaid, 4561, 4745. *See also* Finances; Medicare
- Medicaid and CHIP Payment and Access Commission (MACPAC), 4561
- Medicaid system, 4398–4399
- Medical anthropology, 2553
- Medical assessment and laboratory testing, **1049–1068**, 1054–1055

- brain imaging role in diagnosis, 1060–1061
  - computed tomography, 1061
  - magnetic resonance imaging, 1061
- diagnostic assessment, **1049–1068**
  - autoimmune and inflammatory disease, 1056–1057
  - endocrine evaluation, 1054–1056
  - history and physical examination, 1050–1051
  - infectious diseases, 1052–1054
  - liver function, 1051–1052
  - neurologic disorders, 1057
  - renal function, 1051
  - serum electrolyte, 1051
- electroencephalogram role in diagnosis, 1061–1062
  - evoked potential, 1062
  - polysomnography, 1062
- genetic disorders, 1057–1058
- head trauma, 1057
- intoxication/toxic exposure, 1058–1060
  - acetaminophen or salicylate overdose, 1060
  - alcohol, 1059
  - drug abuse, 1058–1059
  - environmental toxins, 1059–1060
  - tested substances, 1059
  - volatile substances, 1060
- neoplastic disorders, 1057
- nutritional deficiency, 1058
- physical health monitoring, 1049–1050
- treatment selection and monitoring
  - electrocardiography, 1062–1063
  - electroconvulsive therapy, 1062
  - metabolic monitoring, 1063–1064
  - psychotropic medications, 1063t, 1064–1067

- Medical audit, 4745
- Medical boards, 2692
- Medical care evaluation. *See* Medical audit
- Medical child abuse, 3977
- Medical errors, **1150–1159**
  - coping with, 1156–1157
  - definition of, 1151
  - in general medicine
    - causes of, 1151–1152
    - types of, 1151
  - overview, 1150–1151
  - population demographics, 1152–1153
  - prevention of, 1154–1159
  - in psychiatry
    - causes of, 1153–1154
    - types of, 1151
  - reporting structure for, 1155
  - restraint and seclusion, 1154
  - types of, 1151
- Medical ethics, 4745
- Medical home, 3983
- Medical illness, homosexuality and, 2084–2085
- *Medical Inquiries and Observations*, *upon Diseases of the Mind* (Rush), 4644
- Medically ill patients, **4196–4209**
  - Alzheimer disease, 4202–4203
  - approach to, 4196–4197
  - chronic obstructive respiratory disease, 4203–4204
  - chronic pain, 4204–4205
  - depression in, 4228–4229
  - falling, 4207
  - frailty, 4207

  - health systems, 4208–4209
  - key concepts in, 4196*t*
  - musculoskeletal disorders, 4204–4205
  - neuropsychiatric sequelae of COVID-19 syndrome, 4205–4206
  - overreliance on drugs, 4207–4208
  - Parkinson disease, 4202–4203
  - psychiatric syndromes in, 4197–4200
    - differential diagnosis, 4198–4199
    - drug and disease interactions, 4198
    - evaluation of, 4199*t*
    - medications as cause of, 4198
    - treatment of, 4200
  - urinary incontinence, 4207
  - urinary tract infections, 4207
  - vascular disease, 4200–4201
  - vision and hearing impairment, 4206
- Medical neighborhoods, 4498–4499
- Medical records
  - breaches, 1131
  - defined, 4745
  - electronic, 1129–1131
  - interoperability, 1129–1130
  - malpractice, 1130
  - privacy and security, 1130–1131
  - problems in, 1129
  - ransomware and, 1131
- Medical review, 4745
- Medical students
  - early career, 2696
  - late career, 2697
  - mental health, **2686–2697**
  - mid career, 2697

  - minority, 2689–2690
  - personality traits, 2688–2689
  - residency and fellowship training, 2694–2695
  - substance use, 2690–2691
  - wellness, 2694
  - women, 2689
- Medicare, 4394–4398. *See also* Finances; Medicaid
  - defined, 4745
  - insurance coverage supplementing, 4398
  - Medicare Part A, 4395–4396
  - Medicare Part B, 4396–4397
  - risk in managed care, 4397–4398
- Medication-induced movement disorders, **3090–3099**, 4745
  - acute akathisia, 3095
  - acute dystonia, 3094–3095
  - antidepressant discontinuation syndrome, 3097–3098
  - CNS control of movement, 3091–3092
  - diagnostic criteria for, 3091, 3091*t*
  - DSM-5 categories for, 3091*t*
  - history, 3091
  - neuroleptic malignant syndrome, 3092–3094
  - parkinsonism, 3092
  - tardive akathisia, 3097
  - tardive dyskinesia, 3096–3097
  - tardive dystonia, 3097
- Medication(s)
  - adherence to, 602
  - for agitation, 3576*t*
  - for alcohol use disorders (AUDs), 1310
  - for Alzheimer disease (AD), 4315*t*
  - as cause of psychiatric syndromes in medically ill patients, 4198
  - cognitive behavioral therapy (CBT) and, 4369

  - for dementia, 1247
  - electroconvulsive therapy (ECT) and, 3459–3460
  - electroencephalography (EEG) and, 378–379
  - for epilepsy, 530–531
  - intensive short-term dynamic psychotherapies (ISTDPs) and, 2779
  - for major neurocognitive disorders, 1247
  - for neurocognitive disorders, 3335*t*
  - for pain in end-of-life care, 2389
  - for personality disorders, 2265–2268, 2266*t*
  - psychotropic, 1063t, 1064–1067, 3440–3441
  - sexual dysfunctions and, 4451–4452
  - and substance use disorders in adolescents, 4001–4002*t*
  - transcranial magnetic stimulation (TMS) and, 3469–3470
  - transplant, 2283
  - weight gain and, 2308–2309, 2308*t*
- Medicine
  - evolutionary, 862–863
  - humanism in, 2384
- Meditation, 2643
  - for adolescents, 3954
  - mindfulness, 2671
  - for sexual addiction, 2143–2144
  - spirituality and, 2670
  - transcendental, 2671
- Medium spiny neurons, 28
- Medroxyprogesterone (MPA; Depo-Provera), 2055, 2098
  - for paraphilic disorders, 2138
- Medulla, 42
- Meduna, Ladislaus von, 4660
- Meeks, John, 3877
- Megace. *See* Megestrol acetate (Megace)

- Megalomania, 4745
- Megavitamin therapy. *See* Orthomolecular treatment
- Megestrol acetate (Megace), for cancer-related fatigue, 2359
- Meissner, William, 2677
- Melancholia, 1650, 4655, 4745. *See also* Somatic symptom and related disorders
- Melanin, 161
- Melanocortins, 161
- Melanocyte-stimulating hormone, 161
- MELAS. *See* Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS)
- Melatonin, 101, 161, 2641, **3391–3396**
  - chemistry, 3391
  - for delirium, 1237
  - for dementia with Lewy bodies, 1268–1269
  - dosage and administration, 3396
  - drug interactions, 3395–3396
  - drug studies, 3393–3394
  - history, 3396*t*
  - molecular structure, 3391*f*
  - organ/system effects of, 3394
  - pharmacodynamics, 3393
  - pharmacokinetics, 3391–3393, 3392*t*
  - pharmacologic actions, 3391–3393
  - precautions and adverse reactions, 3395
  - seasonality and levels of, 196
  - secretion of, 205
  - sleep-wake regulation, 196
  - synthesis, 100*f*
  - therapeutic indications, 3394–3395
- Melatonin receptor agonists, **3391–3396**
- Mellaril. *See* Thioridazine (Mellaril)

- Melzack, Ronald, 2933
- Memantine (Namenda), 3342–3344
  - adverse events, 4319–4320
  - for Alzheimer disease, 1258, 4315*t*, 4319
  - chemical structure of, 84*f*
  - chemistry, 3342–3343
  - clinical drug studies, 3343
  - for dementia with Lewy bodies, 1268
  - dosage/administration, 4319
  - drug interactions, 3343–3344, 4319
  - mechanism of action, 4318
  - molecular structure, 3343*f*
  - neurocognitive disorders, 3335*t*
  - organ/system effects of, 3343–3344
  - pharmacodynamics, 3343
  - pharmacokinetics, 3343, 4319
  - pharmacologic actions, 3343
  - precautions and adverse reactions, 3343–3344
  - therapeutic indications, 3343–3344
- Memantine and donepezil (Namzaric), 4315*t*
- Memorial Delirium Assessment Scale (MDAS), 2351
- Memory, 472–473. *See also* Childhood autobiographical memory disturbances (CAMD); False memories; Immediate memory; Metamemory; Pseudomemories; Recovered memory; Tests of Malingering and Memory (TOMM)
  - aging and, 198
  - assessment, 3568
  - autobiographical, 2005–2006
  - biology of, **767–784**
    - amnesia, 767–772, 768*f*, 769*f*, 770*f*, 770*t*, 771*f*, 772*f*
    - classical conditioning, 775–776
    - cortical organization, 774–777

    - functional assessment, 780–784
        - infantile amnesia, 783
        - memory distortion, 782–783
        - memory self-reports, 781–782
        - new learning capacity, 781
        - psychogenic amnesia, 782
        - remote memory, 781
        - and unconscious, 783–784
    - impairments in psychiatric illness, 780
    - long-term potentiation, 779–780
    - neuroimaging, 776–777, 777*f*, 778*f*
    - plasticity, 778–779
- circadian rhythms and, 198
- cognitive aging, 4163
- declarative, 302, 767–772
- defined, 4745
- delusional, 1170
- distortion, 782–783
- disturbances of, 1180–1182
    - in recall, 1181–1182
    - in registration, 1181
    - in retention, 1181
- emotional, 1812
- explicit, 730, 730*t*
- impairments
    - in psychiatric illness, 780
    - technologic impairments for, 783
- implicit, 730, 730*t*
- information processing model of, 1022*f*
- neurocognitive disorders and, 198
- neuropsychological evaluation, 4169–4170
- nondeclarative, 774

  - priming, 774, *774f*, *775f*
  - PTSD and, 1878
  - recovered, 1933
  - self-reports, 781–782
  - sleep and, 213, 772
  - stimulant-related disorders and, 1315–1317
  - subtypes of, *778f*
  - systems, 777–778
  - traumatic, 1993–1994, *1994t*
  - types of, 767–774
  - working, 302
- Memory disorders, neurosteroids and, 139
- Memory systems, 729–730, 730t
- Menarche, 2582, 4745
- Mendacity, 4745
- Mendel, Emanuel Ernst, 4656
- Mendel, Gregor, 251
- Mendels, Joseph, 1657
- Menninger, Karl, 1907, 2679, 3754, 4697
- Menninger, William, 790, 4655, 4697
- Menninger Institute Psychotherapy Project, 2743
- Menninger Psychotherapy Project, 787
- Menninger Psychotherapy Research Project, 2743
- Menopause, 2608–2611
  - aging and, 2611
  - cognitive changes, 2609–2611
  - definition of, 2608
  - depression and, 2609
  - hormone replacement therapy, 2609–2610
  - pharmacokinetics/pharmacodynamics and, 3414
  - physical symptoms, 2608–2609
  - treatment of menopausal symptoms, 2609–2611

- Mens rea, 4745
- Menstrual cycles, 2583–2585
  - functional hypothalamic amenorrhea and, 2584–2585
  - ovulatory dysfunction, 2584
  - pharmacokinetics/pharmacodynamics and, 3413
  - physiology, 2583–2584, 2583*f*
- Mental aberration, 4745
- Mental age, 4745
- Mental coprolalia, 1195
- Mental deficiency, 4745
- Mental disorders. *See also* Psychiatric diseases/disorders
  - defined, 4745
  - definition of, 1199
  - empirical risks for, 2618*t*
  - global burden of, 858–859
  - HIV in children/adolescents and, 676–677
  - in immigrants, 841–843
  - insomnia comorbidity, 2168–2169
  - in minorities, 839–841
  - neurodevelopmental models, 312–313
  - neurosteroids and, 138–139
  - opioid-induced, 1405–1406
  - in primary care, 2276–2277
  - proliferation of, 1206–1207
  - stigmatizing attitudes toward people, 4678–4679
- Mental function, 801
- Mental health
  - culture and, 846–849
  - defined, 4745
  - definitions of, 786*t*
  - educational outcomes and, 4026
  - global perspective, 858–860

  - of Indigenous Peoples, 844–845
  - intervention spectrum, 4041*f*
  - mental illness and, 2889*t*
  - normality of (*See* Normality)
  - over the life span, 4159–4160
  - police and, 4573–4574
  - racism and, 836–839, 838*t*
  - of refugees, 845–846
  - social determinants of, 839
- Mental Health America, 2700
- Mental health apps, 1123, 2703
- Mental health courts (MHC), 1608–1609, 4577–4578
- Mental health defenses, 4416
- Mental Health Literacy, 3515
- Mental Health Parity and Addiction Equity Act (MHPAEA), 843, 4498, 4561, 4562
- Mental health services
  - migrants and, 843–844
  - minorities and, 844–845
  - treatment disparities, 843–844
- Mental health services research, **4526–4532**
  - case reports, 4528–4532
    - services for families of persons with schizophrenia, 4529–4530
    - technology and implementation science, 4530–4532
    - treatment of depression in primary care settings, 4528–4529
  - scope of, 4526–4528
- Mental hygiene, 4745
- Mental illness
  - criminalization and, 4574–4575
  - defined, 4745
  - mental health and, 2889*t*
  - phases of severe, 4544–4550

    - multidimensional framework, 4545*f*
    - prodromal phase, 4546–4548
      - continuity of care, 4547–4548
      - early intervention and relapse prevention, 4546–4547
      - elapsed, 4546*t*
      - interventions, 4546*t*
    - recovery phase, 4550
    - stabilization phase, 4548–4549
    - stable phase, vocational rehabilitation in, 4549–4550
- Mental image, 746
- Mentalization, 476
- Mentalization-based treatment (MBT), **2863–2873**
  - for antisocial personality disorder, 2873
  - attachment, 2866
  - for borderline personality disorder, 2755–2756, 2872–2873
  - clinical issues, 2870–2871
    - complications, 2870
    - contraindications, 2870
    - indications, 2870, 2870–2871
    - limitations, 2870
    - treatment goals, 2871
  - clinical principles, 2867*t*
  - definition of, 2863
  - dimensional mentalizing profile and BPD, 2864–2865
  - ethical issues, 2871–2872
  - history, 2863
  - mentalizing
    - automatic *vs.* controlled, 2864
    - cognitive *vs*. affective, 2864
    - definition of, 2863
    - internal *vs*. external, 2864
    - multidimensional nature of, 2864

    - relationship, 2869–2870
    - role in therapy, 2863–2864
    - self *vs*. others, 2864
  - nonmentalizing modes and BPD, 2865–2866, 2865*t*
    - pretend mode, 2865*t*, 2866
    - psychic equivalence mode, 2865–2866, 2865*t*
    - teleologic mode, 2866
  - for personality disorders, 2930
  - research and evaluation, 2872–2873
  - techniques, 2866–2870
    - affects and affect identification, 2868–2869
    - alternative perspectives, 2869
    - clarification and exploration, 2868
    - clinical principles, 2867–2868, 2867*t*
    - empathic validation, 2868
    - mentalizing the relationship, 2869–2870
    - session trajectory, 2868–2870
    - therapeutic alliance, 2867, 2867*t*
  - theoretical issues, 2863–2866
- Mentalizing, definition of, 2863
- Mentally ill
  - criminalization of, 4627, 4627*f*
  - criminal justice system and, 4576
- Mental models, 731
- Mental processing speed, 4170–4171
- Mental retardation, 4745
- Mental Retardation Research Centers (MRRCs), 3632
- Mental state, 734
- Mental status, 4745
- Mental status examination (MSE), 994, 994*t*, 1029, 3524, 3527, 4745
- Men who have sex with men (MSM), 2060

- Meperidine hydrochloride (Demerol), pain management in burn patients, 2470
- Mephedrone, 1314
- Meprobamate (Equanil/Miltown), 1426*t*, 3236, 4662
- Mercury, impact on CNS and PNS, 655*t*, 658
- Meresa. *See* Sulpiride (Meresa)
- Merleau-Ponty, Maurice, 418
- MERRF. *See* Myoclonic epilepsy with ragged red fibers (MERRF)
- Merycism. *See* Rumination disorder of infancy
- Mescaline, 2903, 4745
- Mesencephalon, 9, 9*f*
- Mesmer, Franz Anton, 1927, 2911–2912, 4684
- Mesmerism, 4745
- Mesomorphic, 4745
- Messenger RNA (mRNA), 234
- Messud, Claire, 2880
- Meta-analysis, 4745
- Metabolic disorders
  - EEG screening, 373
  - psychiatric presentations, 479t
- Metabolic monitoring, 1063–1064
  - body mass index, 1063–1064
  - fasting blood sugar, 1064
  - glycosylated hemoglobin, 1064
  - lipids, 1064
  - weight, 1063
- Metabolic neuropathies, 658
- Metabolic Smith–Lemli–Opitz syndrome, 3641*t*
- Metabolic syndrome, 161, 187
  - definition of, 1597*t*
  - dyslipidemia, 1597–1598
  - hyperglycemia, 1598

  - hypertension, 1597
  - insulin resistance, 1598
  - obesity, 1597
  - overweight, 1597
  - schizophrenia and, 1596–1598
  - stress and, 2419
- Metabotropic glutamate receptors, 86
- Metachromatic leukodystrophy, 686–687, 687*f*
- Metacognition, 732
- Metacognitive monitoring, 732
- Metal metabolism disorders, 703–708
  - aceruloplasminemia, 705–708, 708*f*
  - neurodegeneration with brain iron, 704–708
  - Wilson disease, 703–704
- Metamemory, 781
- Metaphor, 2744
- Metaplasticity, 81
- Metapsychiatry, 4745
- Metapsychology, 4745
- Metastability, 804
- Metencephalon, 9, 9*f*
- Methadone (Dolophine)
  - adverse effects of, 3352
  - chemistry, 3351
  - clinical studies, 3351–3352
  - description of, 4745
  - dosing, 3353
  - drug interactions, 3352–3353
  - molecular formula, 3351*f*
  - opioid use disorders, 1415–1416, 1416*t*, 3351–3353
  - opioid withdrawal, 1413
  - pain management in burn patients, 2469

  - pharmacodynamics, 3352
  - pharmacokinetics, 3351–3352
  - pharmacologic actions, 3351–3352
  - repolarization abnormalities, 2289
- Methamphetamine
  - emergency department visits, 1367
  - gene transcription and, 1316
  - history, 1313
  - negative reinforcement mechanisms, 1316
  - neurocognitive disorders, 1291
  - oxidative stress and, 1316
  - prevalence of, 1294, 1314
  - psychotic symptoms, 1318–1319
  - routes of administration, 1317
  - toxicity and medical complications, 1319–1320
- Methamphetamine use disorder
  - naltrexone-bupropion for, 3361
  - prevalence of, 1314
  - treatment of, 1322–1323
- Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA), 880, 885
- 3-Methoxy-4-hydroxyphenylglycol (MHPG), 4745
- Methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), 4746
- Methylation, 236
- Methylation arrays, 44
- Methylene blue, 2367
- 3,4-Methylenedioxymethamphetamine (MDMA), 199, 431, 1314–1315, 1350*f*, 1354, 2904, 3386–3387
  - chemical structure, 3386*f*
  - clinical issues, 2908–2909
  - cognitive-behavioral conjoint therapy, 2909
  - complications, 2908–2909

  - contraindications, 2909
  - defined, 4745
  - history, 2908
  - indications, 2908
  - intoxication, 1357
  - limitations, 2908
  - mechanism of MDMA-assisted psychotherapy, 2908
  - for patients with life-threatening illness, 2909
  - principles and goals of therapy, 2908
  - psychopharmacology, 2908
  - for PTSD, 2909
  - SERT availability, 403
  - for social anxiety, 2909
  - subjective effects, 2908
  - trials, 3388–3389, 3389*t*
- Methylenedioxypyrovalerone (MDPV), 1314
- Methylone, 1314
- Methylphenidate (Ritalin or Concerta), 3114
  - for ADHD, 3606, 3733–3736
  - adverse effects of, 3735–3736
  - for depression, in cancer patients, 2357
  - for depression in later life, 4298
  - efficacy, 3735
  - formulations, 3734, 3734*t*, 3735, 3735*t*
  - for geriatric depression, 4239
  - lithium and, 3320
  - pharmacology of, 3734
- Metonymy, 4746
- Metoprolol (Lopressor, Toprol XL), 3224–3228
- Metrazol treatment, 4746
- Metzinger, Thomas, 418
- Metzl, Jonathan, 1138

- Meyer, Adolf, 486, 791, 946–947, 947*f*, 1200, 1653, 4479, 4651, 4652, 4688, 4746
  - habit training, 947
  - personality and psychopathology, 946–947
  - psychobiology of, 946
  - treatment by, 947
- Meyer, Jon, 2104–2105
- Meyer. Adolph, 789
- Meyer-Gross, Willy, 4661
- Meynert, Theodor, 2746, 4649
- MHPAEA. *See* Mental Health Parity and Addiction Equity Act (MHPAEA)
- Michigan Alcoholism Screening Test (MAST), 1044*t*, 3546
- Microbiome, 2416
  - immune-brain interactions and, 191–192
- Microcephaly, 4746
- Microdosing, 3390
- Microglia, 5, 5*t*
- Micronutrients, **2650–2654**
  - folate (vitamin B9), 2652–2653
  - omega-3 fatty acids, 2653–2654
  - vitamin B12, 2651–2652
  - vitamin D, 2650–2651
- Micropsia, 4746
- MicroRNAs (miRNAs), 239, 454, 1316, 2229
- Microsatellite markers, 255
- Microsleeps, 210
- Microtubule-associated proteins (MAPs), 64
- Microtubules (MTs), 64
- MID. *See* Multidimensional Inventory of Dissociation (MID)
- Midazolam (Versed)
  - absorption, 3240

  - agitation, 3247–3248, 4405*t*
  - delirium, 3250
  - distribution, 3241
  - drug interactions, 3064, 3254
  - drug studies, 3231, 3233
  - major depressive disorder, 3928
  - medication-induced movement disorders, 3097
  - metabolism, 3241
  - organ/system effects of, 3207
  - pharmacokinetics, 3140
  - treatment-resistant depression, 3821
  - withdrawal, 3249, 3252
- Midbrain, 40–42, 41*f*
- Middle adulthood, 4746
- Middle age, 4746
- Middle insomnia, 4746
- Midlife crisis, 4746
- Midlife depression, androgen administration in, 3428–3429
- Midline nuclei, 23
- Migraine, 645–646. *See also* Headache(s)
  - with aura, 646*t*
  - defined, 4746
  - differential diagnosis, 2483–2484
  - health seeking in, 648–649
  - impact of, 648–649
  - topiramate, 3269
  - treatment of
    - nonpharmacologic, 651
    - pharmacologic, 650–651, 650*t*, 651*t*
  - without aura, 646*t*
- Migraines, 534*t*
- Migranous hallucinations, 1174

- Migrant Integration Policy Index (MIPEX), 849
- Migrants, 835
  - child/adolescent, 4113–4117
    - client interview, 4114
    - contextual history, 4114–4115
    - diagnostic screening, 4115–4116
    - focused trauma history, 4115
    - mental status examination, 4116
    - past medical history, 4115
    - past mental health history, 4115
    - physical examination, 4115
    - psychological evaluation of, 4113–4117
    - social history, 4115
    - substance use history, 4115
  - mental healthcare utilization, 843
  - mental health prevalence in, 4114
  - psychological affidavit for, 4116–4117
- Migration, 849
- Mild cognitive impairment (MCI), 1027, **1273–1281**. *See also* Cognitive impairments
  - assessment, 1276–1277
  - biomarkers, 1277
  - clinical presentation, 1275–1276
  - conversion rate to Alzheimer disease, 1277
  - course and prognosis, 1280
  - criteria for, 1274*t*
  - definition of, 1273–1274
  - diagnosis of, 1274*f*, 1275*t*
  - differential diagnosis, 1278–1279
  - epidemiology of, 1274–1275
  - etiology of, 1274–1275
  - genetics of, 1277

  - historical perspective, 1274
  - neuroimaging, 1277–1278
  - neuropsychological assessment, 1276–1277
  - omega-3 fatty acids and, 2654
  - prevalence of, 1276*t*
  - subtypes of, 1276*f*
  - terminology, 1275*t*
  - treatment of, 1280, 1281*t*
- Mild mental retardation. *See* Mental retardation
- Mild neurocognitive disorder, 1379–1380
  - defined, 4746
  - dementia and, 1247
  - diagnosis, 4829–4832*t*
  - inhalant-induced, 1379–1380
- Mild stress, 181
- Milieu therapy, 4746
- Miller, Neal, 971
- Miller Forensic Assessment of Symptoms Test (M-FAST), 2505
- Millon, Theodore, 976
- Millon Adolescent Clinical Inventory (MACI), 3548
- Millon Adolescent Personality Inventory (MAPI), 3548
- Millon Clinical Multiaxial Inventory (MCMI), 2255, 2505
- Millspaugh, J. A., 4658
- Milnacipran (Savella), 3150
  - antidepressant efficacy of, 3150
  - for depression in later life, 4297
  - pharmacology of, 3150
  - tolerability of, 3150
- Milrod, David, 1704
- Mimicry, 4746
- Mind, 727, 801
  - development and differentiation, 808–810

  - object relations as basic building units of, 806–808
  - self-organization, 805–809
  - states of, 734
  - theory of, 732
- MindBody, 805
- Mind-body practices, 2642–2643
- Mind-body therapies (MBTs), 4139–4141
- Mind control, 2019–2020
- Mindful meditation, for adolescents, 3954
- Mindfulness, 2844
  - for borderline intellectual functioning, 2536–2537
  - cognitive behavioral therapy and, 4365
  - for weight loss, 2311
- Mindfulness-based cognitive therapy (MBCT), 1827, 2642–2643, 2819, 2898
- Mindfulness-based stress reduction (MBSR), 1827, 2819
- Mindfulness-based therapy, 2819
- Mindfulness meditation, 2671
- Mindfulness training, 2642–2643
- Mind Your Mind, 3515
- Minerals, 2633
- MINI (McGill Illness Narrative Interview), 2571
- MINI International Neuropsychiatric Interview (MINI), 1099–1100
- MINI International Neuropsychiatric Interview for Children and Adolescents (MINI-KID), 3540, 3541
- Minimal brain dysfunction, 4746
- Mini-Mental State Examination (MMSE), 997, 1029, 1112, 4168, 4454
- Mini-practice, 2424
- Mini Profile of Nonverbal Sensitivity (MiniPONS), 4551
- Minnesota Multiphasic Personality Inventory (MMPI), 1100–1101, 4659, 4746

- Minnesota Multiphasic Personality Inventory-2 (MMPI-2), 1036, 1038–1041, 2505
  - clinical scales, 1040
  - computer scoring, 1041
  - content scales, 1041
  - Harris & Lingoes scales, 1041
  - scale development, 1040*t*
  - scale elevations, 1040–1041
  - supplemental scales, 1041
  - validity scales, 1038–1040
    - backpage infrequency scale, 1040
    - F-scale, 1038–1040, 2505
    - K scale, 2505
    - Lie (L) scale, 1038
- Minnesota Multiphasic Personality Inventory-3 (MMPI-3), 1042
- Minnesota Multiphasic Personality Inventory-Adolescent (MMPI-A), 3548
- Minocycline
  - for bipolar disorder, 188
  - for depressive disorders, 185
  - for schizophrenia, 187
- Minor depressive disorder, 1689
- Minor epilepsy. *See* Petit mal epilepsy
- Minorities
  - mental disorders among, 839–841
  - mental healthcare utilization, 843
- Minuchin, Salvador, 2947
- Miosis, 3352
- *Miranda* rights, 4008
- Mirapex. *See* Pramipexole (Mirapex)
- Mirroring, 2752
- Mirtazapine, **3211–3214**

- Mirtazapine (Remeron)
  - antidepressant efficacy of, 3211–3212
  - for anxiety disorders, 1834
  - for cancer patients, 2348
  - chemical structure, 3211*f*
  - for depression, in cancer patients, 2357
  - for depression in later life, 4297
  - future directions, 3213–3214
  - for geriatric depression, 4239
  - indications for, 3211–3213
  - lactation and, 3213
  - pharmacokinetics, sex differences in, 3416
  - pharmacology of, 3211
  - pregnancy and, 2606, 3213
  - prophylactic therapy, 3213
  - safety, 3213–3214
  - side effects, 3819
  - SSRIs/SNRIs, combining with, 3212
  - tolerability, 3213–3214
  - toxicity, 3213–3214
- Missed abortion, 2590*t*
- Missing at random (MAR), 905
- Missing completely at random (MCAR), 905
- Missing data, 904–905
- Mistrust, 938
- Mitchell, Arthur, 4477
- Mitchell, Silas Wier, 2272, 2656, 4746
- Mitchell, Stephen, 2750, 3877
- Mitochondria, 1507
- Mitochondrial diseases, 700–703
  - with cerebral white matter disease, 703

  - mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes, 702
  - myoclonic epilepsy with ragged red fibers, 702–703
  - with progressive external ophthalmoplegia, 703
  - Wolfram disease type I and type II, 703, 703*f*
- Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS), 702
- Mitochondrial membrane protein-associated neurodegeneration (MPAN), 705
- Mitochondrial myopathies, 663
- Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE), 703
- Mitosis, 236
- Mitotoxin methylazoxymethanol acetate (MAM), 313
- Mitral valve prolapse, 2285
- Mixed anxiety-depressive disorder, 1694–1695
- Mixed delirium, 2398
- Mixed dopamine serotonin receptor antagonists, 3929
  - for autism spectrum disorders, 3933
- Mixed episodes, 1649, 1650
  - anabolic steroid-induced, 1443–1444
- Mixed states, electroconvulsive therapy for, 3451
- Mixoscopia, 4746
- MMPI-2-Restructured Form, 1041–1042
- MMPI for PD (MMPI PD) scale, 2254
- MMSE (Mini-Mental State Examination), 1112, 4168
- M’Naghten rule, 4746
- M’Naghton, Daniel, 4685
- *Mobile Assessment and Treatment for Schizophrenia (MATS)*, 1591
- Mobile health, 1591–1592, 4465–4467
- MoCA (Montreal Cognitive Assessment), 997, 1029, 1112–1114, 1113*t*, 4168

- Moclobemide (Aurorix/Manerix), 3185, 3185*t*
  - for major depressive disorder, 1724
- Modafinil (Provigil), 3127–3128
  - for acute bipolar depression, 1734
  - adverse effects of, 3128
  - for bipolar depression, 3110
  - for cancer-related fatigue, 2359
  - chemistry of, 3108
  - for depression, in cancer patients, 2357
  - dosage and administration, 3114, 3114*t*
  - drug interactions, 3114
  - for excessive daytime sleepiness, 3111
  - immunologic effects, 3112
  - for major depressive disorder, 3109–3110
  - mechanism of action, 3127
  - molecular structure of, 3108*f*
  - for pain, 2389
  - pharmacokinetics, 3109
  - pharmacologic properties of, 3127
  - side effects, 3113
  - uses of, 3127–3128
    - ADHD, 3128
    - cocaine addiction, 3128
    - cognitive enhancement, 3128
    - narcolepsy, 3127–3128
    - obstructive sleep apnea, 3128
- Mode, 4746
- Model fitting, 906
- Modeling, 4746
- Model Penal Code, 4746
- Moderate mental retardation. *See* Mental retardation
- Moderation, 877

- Modern gay movement, 2072
- Modes of processing, 731–732
- Modifications, 4026*t*, 4032*t*
- Modified Checklist for Autism in Toddlers (M-CHAT), 3546
- Modularity, 463–464
- Moeli, Carol, 4657
- Molecular genetic studies, ADHD, 3730
- Molecular imaging
  - radiotracers, 1537*t*
  - in schizophrenia, **1535–1555**
    - artificial intelligence, 1553–1554
    - drug development, 1545–1549
      - antipsychotics, 1545–1549
    - genetics, 1553
    - machine learning, 1553–1554
    - pathophysiology, 1538–1545
      - benzodiazepine receptors, 1544
      - cannabinoid, 1544–1545
      - cholinergic receptors, 1543–1544
      - dopamine, 1538–1542, 1539*t*
      - glutamate, 1544
      - histamine H_1 receptors, 1544
      - opiates, 1544–1545, 1545
      - serotonin, 1542–1543
    - theories of, 1535–1538
      - cerebral blood flow, 1536
      - cerebral metabolism, 1536
      - D_2-like receptors, 1536–1537
    - treatment, 1555
  - types of, 1538
- Molecular psychiatry, 4746

- Money, John, 2103
- Mongolism. *See* Down syndrome
- Moniz, Egas, 4690
- Monoamine, 4746
- Monoamine enhancers
  - dopamine-noradrenaline reuptake inhibitors, **3163–3167**
  - selective serotonin reuptake inhibitors, **3129–3144**
  - serotonin$_{1A}$ agonists, **3168–3174**
  - serotonin-norepinephrine reuptake inhibitors, **3144–3151**
  - tetracyclic antidepressants, **3151–3163**
  - for treatment of depression, 3031
  - tricyclic antidepressants, **3151–3163**
- Monoamine modulators
  - mirtazapine, **3211–3214**
  - nefazodone, **3204–3210**
  - trazodone, **3197–3203**
- Monoamine oxidase (MAO), 96, 98, 4746
- Monoamine oxidase A (MAOA), 2518
- Monoamine oxidase inhibitors (MAOIs), 1067, **3184–3196**
  - agoraphobia, 3195–3196
  - for anxiety disorders, 1834
  - for anxiety disorders in later life, 4304
  - application site reaction, 3192
  - atypical depression, 3192
  - for bulimia nervosa, 2163
  - “cheese effect,” 3184
  - chemistry, 3185–3187
  - classification of, 3185*t*, 3187
  - clinical drug studies, 3192–3196
  - contraindications, 3192
  - defined, 4746
  - for depression, in cancer patients, 2356

- for depression in later life, 4298
- development of, 3184–3185
- dosage and administration, 3188*t*
- drug-drug interactions, 3191
- drug level monitoring, 3190
- dysthymia, 3194
- food-drug interactions, 3191, 3191*t*
- forms of, 3185–3186
- future directions, 3196
- for geriatric depression, 4239
- hepatoxicity, 3192
- history, 3007, 3184–3185, 4662
- hypertensive crisis, 3191, 3191*t*
- indications for, 3192–3196
- insomnia, 3191
- investigational drugs, 3196
- irreversible, 3185, 3185*t*
- for major depressive disorder, 1724
- major depressive disorder, 3193–3194
- mechanism of action, 3187
- molecular structures of, 3186*f*
- organ/system effects of, 3190–3191, 3194
- orthostatic hypotension, 3191
- overdose, 3192
- panic disorder, 3195
- pharmacokinetics, sex differences in, 3416–3417
- pharmacologic actions, 3187–3190
- posttraumatic stress disorder, 3195
- pregnancy and, 2606
- psychotic depression, 3194
- reactions catalyzed by, 3186*f*
- reversible, 3185*t*

  - serotonin syndrome, 3191, 3191*t*
  - social anxiety disorder, 3195
  - treatment-resistant depression, 3193
  - warning and precautions, 3191–3192
  - weight gain, 3191
  - weight gain and, 2309
  - withdrawal, 3192
- Monoamine oxidase type B inhibitors, 4320
- Monoamines, 1759–1760
  - depression and, 1766
- Monoclonal antibodies, neurotoxicity of, 2465*t*
- Monomania, 4656, 4746
- Monosymptomatic hypochondriacal psychosis, 4746
- Monozygotic twins, 253, 4746
- Monro, Edward Thomas, 4643
- Monro, Henry, 4645
- Montage, 358–360, 359*f*
- Montgomery-Asberg Depression Rating Scale (MADRS), 1104, 4136
- Montreal Cognitive Assessment (MoCA), 997, 1029, 1112–1114, 1113*t*, 4168
- Montreal procedure, 369
- Mood, 819
  - brain circuitry and, 1208
  - defined, 4746–4747
  - dementia and, 1245
  - disturbances of, 1183–1185
    - depression, 1184–1185
    - elated moods, 1185
  - everyday functioning and, 4181
  - neuropsychological evaluation, 4171
  - in older adults, 4134

  - psychiatric examination and, 994
- Mood and Feelings Questionnaire (MFQ), 3544
- Mood-congruent delusion, 4747
- Mood-congruent hallucination, 4747
- Mood-congruent psychotic disturbances, 1673–1674
- Mood-congruent psychotic features, 4747
- Mood disorder(s), **1648–1779**, 1702–1709. *See also specific mood disorders*
  - affective temperaments, 1670–1671
  - affects, 1669
  - alterations of hormonal regulations and, 1771–1772
    - HPA activity, 1771
    - thyroid axis activity, 1771–1772
  - anabolic steroid-induced, 1442–1444
  - ancient Greek and Roman descriptions, 1650–1651
  - atypical depression, 1695
  - bipolar disorders, 1690–1694
  - in cancer patients, medical causes of,’ 2355*t*
  - chemotherapy agents and, 2368*t*
  - children and adolescents, **3815–3836**
    - depressive disorders, 3815–3821
    - early-onset bipolar disorder, 3828–3836
    - suicidal behavior, 3821–3827
  - clinical and public health scope of, 1648–1649
  - clinical features, **1668–1701**
  - clinical phenomenology, 1753–1754
  - cognition in, 737–738
  - comorbidity, 1666
  - concepts of, 1648
  - correlates of, 1663–1665
    - age, 1664
    - race and ethnicity, 1664–1665

  - sex/gender, 1663–1664
- cyclothymia, 1649–1650
- defined, 4747
- definitions of, 1649
- depressive disorders, 1678–1698
- depressive syndrome, 1671–1676
- diagnostic classification, 1678
- diagnostic trends, 1668–1669
- differential diagnosis, 1695–1701
  - alcohol and substance use disorders, 1697–1698
  - anxiety disorders, 1696
  - bereavement, normal, 1695–1696, 1695*t*
  - major depressive disorder *versus* bipolar disorder, 1696–1697
  - personality disorders, 1697, 1697*t*
  - physical disease, 1698–1700, 1698*t*
  - schizoaffective disorder, 1700–1701
  - schizophrenia, 1700, 1700*t*
- dissociative identity disorder and, 1997
- dysthymia, 1649–1650
- EEGs in, 376–377
- elation, 1670
- epidemiology of, 1648, **1662–1668**
- epilepsy and, 527
- etiology of emotions and, 1754–1755
- fMRI of, 344–345
- gabapentin, 3267
- genetics, 1754
- grief, 1670
- group psychotherapy for, 2965
- growth hormones, 1772
- historical trends, 1667–1668
- ICD-11 criteria for, 1222

- impact of, 1667
- inflammation and, 1774–1779
  - cross-sectional and simulation studies, 1774
  - depression, 1774–1775
  - mechanism, 1775–1777
  - treatment, 1777–1778
- inhalant-induced, 1379
- intellectual disabilities and, 3650–3651
- intracellular cascades, 1771
- lifetime prevalence of, 1664*f*
- major depressive disorder, 1649
- manic syndrome, 1676–1678
- mixed anxiety-depressive disorder, 1694–1695
- morbidity, 1648–1649
- morbid mood states, 1671
- mortality, 1648–1649, 1667
- MRI applications to, 341
- MRS in, 348–349
- nefazodone, 3205
- neurobiology, 1752–1779
- neuroplasticity and neurotrophic changes in, 1770–1771
- nitric oxide and, 127
- in older adults, **4227–4244**
- organ transplantation and, 2458
- perception in, 737–738
- pharmacological treatment of, 1719–1743
- prolactin and, 1772
- psychodynamic aspects of, 1702–1703
- psychopathology, 1671–1678
- psychosocial factors, 1665–1666
  - social stressors, 1665–1666
  - social support, 1666

- psychotherapy, **1744–1752**
- psychotic disorders in later life *versus*, 4269
- risk factors, 1754
- sadness and joy, 1669–1670
- second messengers, 1771
- sensation in, 737–738
- service patterns, 1666–1667
- sleep disorders and, 4088–4089
- sleep neurophysiology and, 1772
- social correlates of, 1665
  - geographic trends, 1665
  - marital status, 1665
  - residence, 1665
  - seasonal factors, 1665
  - socioeconomic factors, 1665
- specific learning disorder and, 3665
- spectrum of, 1669
- structural and functional brain imaging, 1772–1774
- subthreshold mood states, 1650
- suicidal behavior, **1709–1719**
  - adverse life situations, 1714
  - affective temperaments, 1713
  - bipolarity, 1716
  - clinical characteristics of mood episode, 1712
  - clinical utility of risk factors, 1714–1715, 1715*t*
  - comorbid disorders, 1712–1713
  - family history of, 1713–1714
  - negative life events, 1714
  - prevention of suicide, 1716–1719
    - nonpharmacologic, 1717
    - pharmacologic, 1717
    - in primary care, 1718–1719

    - previous suicide attempt, 1713
    - prior course of mood disorder, 1713
    - protective factors, 1712, 1715–1716
    - psychological autopsies, 1711
    - risk assessment and formulation, 1709–1711
    - risk factors, 1712–1716, 1715*t*
  - temperament and, 3626
  - terminology, 1668
  - theoretical synthesis, 1658–1662
    - integrative approaches, 1660
    - pathophysiologic understanding, 1658–1660
  - treatment of, 3738
  - unspecified, 1694–1695
- Mood-incongruent delusion, 4747
- Mood-incongruent hallucination, 4747
- Mood-incongruent psychosis, 1677
- Mood-incongruent psychotic disturbances, 1674
- Mood-incongruent psychotic features, 4747
- Mood(s). *See also* Affect(s); Mood disorder(s)
  - depressed, 1671–1672
  - disturbances, 1671, 1676
  - elevation, 1676
  - endocannabinoids in, 134–135
- Mood stabilizers, 1065–1067
  - for aggression, 3650
  - anticonvulsant, 4301
  - for anxiety, 3651
  - for bipolar disorder, 3032
  - calcium channel inhibitors, **3256–3264**
  - carbamazepine, 1065–1066
  - cutaneous adverse events, 2447
  - defined, 4747

  - for functional gastrointestinal disorders, 2302
  - history, 4663
  - lamotrigine, 1066
  - lithium, 1066–1067, 2712
  - oxcarbazepine, 1066
  - for repolarization abnormalities, 2289
  - for self-injurious behaviors, 3651
  - for suicide, 2712
  - trichotillomania, 2435
  - valproate, 1067, 3304–3307
  - weight gain and, 2308–2309, 2308*t*
- Mood swing, 4747
- Moral development
  - in adolescents, 3511
  - in children, 3494–3496*t*, 3501–3502
  - clinical correlate, 3502
- Moral injury, 2688
- Morality, 809
  - and brain, 2511
  - social development in children and, 3501
- *The Moral Judgment of the Child* (Piaget), 750
- Moral treatment, 4747
- Morbidity and mortality conference (MMC), 1156
- Morbid jealousy, 4747
- Morbid mood states, 1671
- Morbid perplexity, 4747
- Moreau, Jacques-Joseph, 4644, 4646, 4685
- Morel, Bénédict-Augustin, 4656, 4686
- Moreno, Jacob Levy, 2955, 2968–2973
- Moreno, Zerka Toeman, 2968–2971
- Moreno-Lewy, Jakob, 4648
- Mores, 2056, 2091–2101

- Morgan, Charles A., 1955
- Morgan, Thomas, 236
- Morgellons disease, 2434
  - clinical features of, 2434
  - diagnosis of, 2434
  - treatment of, 2434
- Morning larks, 196
- Moron, 4747
- Morphine, 1885
  - for child/adolescent PTSD, 3849
  - defined, 4747
  - for opioid use disorders, 1418
  - structure of, 1400*f*
- Morphogenetic cell death, 60
- Morris, Stuart L., 873
- Moser, Charles, 2110
- *Moses and Monotheism* (Freud), 4747
- Mother, schizophrenogenic, 2790–2791
- Mother-child interactions, 3645–3646
- Mother surrogate, 4747
- Motivation, 2224–2225, 2747–2748, 4747
  - intrinsic, 808
  - mechanism of pain and, 216–217
  - opponent process theory of, 442–443, 443*f*
- Motivational interviewing, 1608, 2821
  - for weight loss, 2311–2312, 2312*t*, 2313*t*
- Motor activity
  - disturbances in, 1194–1195
    - catatonia, 1194
    - mutism, 1194–1195
    - overactivity, 1194
  - psychiatric examination and, 994

- Motor aphasia, 4747
- Motor areas, 21
- Motor deficiency syndrome, 3673
- Motor disorders, 4747
- Motor disturbances and movement disorders, 1195
  - blepharospasm, 1195
  - tics, 1195
  - tremor, 1195
- Motor overflow, 3676
- Motor skill disorders. *See* Developmental coordination disorder (DCD)
- Mourning, 4747
- *Mourning and Melancholia* (Freud), 918–919, 1703
- Moustakas, Clark, 3877
- Mouth, physical examination, 466
- Movement, abnormalities of, 468–470
- Movement Assessment Battery for Children (MABC), 3676
- Movement disorders, 483
  - bedside assessment of, 555
  - cannabinoid for, 1346
  - cerebellar disorders, 572–576
  - classification, 554, 554*t*
  - comparative nosology, 554
  - corticobasal degeneration, 564
  - drug-induced, 1195–1196
    - akathisia, 1195
    - dystonic movements, 1195
    - movement disturbances, 1196–1197
    - neuroleptic malignant syndrome, 1196
    - rabbit syndrome, 1196
    - serotonin syndrome, 1196
    - tardive dyskinesia, 1195–1196

  - dystonias, 575–576
  - encephalitis lethartica, 566
  - essential tremor, 575
  - frontotemporal dementia, 565
  - Huntington disease, 566–571
  - multiple system atrophy, 565–566
  - neuropsychiatric aspects of, **553–576**
  - Parkinson disease, 555–565
  - progressive supranuclear palsy, 563–564
  - psychiatric disturbances associated with, 554–555
  - Wilson disease, 571–572
- Movement disturbances
  - compulsive behaviors, 1196–1197
  - seizure-like behaviors, 1196
- MPTP. *See* Methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)
- MRI. *See* Magnetic resonance imaging (MRI)
- MRS. *See* Magnetic resonance spectroscopy (MRS)
- MSA. *See* Multiple system atrophy (MSA)
- MSE. *See* Mental status examination (MSE)
- MSLT. *See* Multiple sleep latency test (MSLT)
- Mucormycosis, 615–616
- Mullen Scales of Early Learning, 3548
- Müllerian agenesis, 2582
- Multiaxial system, 1205
- Multidimensional Anxiety Scale for Children (MASC), 3544–3545
- Multidimensional Inventory of Dissociation (MID), 1938*t*
- Multidisciplinary Association for Psychedelic Science (MAPS), 432
- Multiethnic Placement Act, 3957
- Multi-infarct dementia. *See* Vascular dementia
- Multimodal association areas, 21
- Multinutrient formulas, 2636–2637
- Multiple delusions, 4747

- Multiple imputation, 905
- Multiple informants, 3555
- Multiple minority, 2077
- Multiple personality disorder. *See* Dissociative identity disorder
- Multiple regression, 4747
- Multiple sclerosis (MS)
  - associated focal neurobehavioral syndrome, 580*t*
  - clinical features, 577–578
  - course, 578
  - definition of, 577
  - diagnosis, 577–578, 579*t*
  - differential diagnosis, 582
  - epidemiology of, 577
  - etiology of, 578
  - neuropsychiatric aspects of, **567–583**
    - affect disorders, 581
    - anxiety, 579–580
    - bipolar disorder, 580
    - cerebellar involvement, 582
    - cognitive impairment, 582
    - depression, 579–580
    - disinhibition, 580
    - euphoria, 580
    - fatigue, 578–579
    - frontal personality changes, 580
    - mania, 580
    - psychosis, 581–582
    - sleep disorders, 582
  - treatment of, 579*t*, 582
  - white matter lesion location, 580*t*
- Multiple sleep latency test (MSLT), 204, 2189
- Multiple system atrophy (MSA), 565–566

- Multiple therapy. *See* Co-therapy
- Multi-scale Dissociation Inventory (MDI), 1938*t*
- Multistability, 802
- Multisystemic therapy (MST), for adolescents, 3954
- Multitasking, 3514–3515
- Multitiered systems of support (MTSS), 4029–4031
- Multitier Mental Health and Psychosocial Support (MHPSS) model, 846
- Multivariate analysis, 4747
- Mummification, 1695*t*
- Munchausen syndrome, 1194, 1906–1907, 1917, 2506, 4747
- Munchausen syndrome by proxy, 1908, 2506
- Mu-opioid receptor antagonists, 2140
- Murphy, Dominique, 1202
- Murphy, Henry, 1477, 4650
- Murray, Henry, 937, 4688
- Muscarine, 4747
- Muscarinic acetylcholine receptors (mAChR), 1543–1544
- Muscarinic receptors, 104–106
- Muscimol, 89
- Muscle diseases, 661–663
  - dystrophinopathies, 662
  - inflammatory myopathies, 662–663
  - mitochondrial myopathies, 663
  - myotonic dystrophy, 661–662
- Musculoskeletal disorders, in medically ill geriatric patients, 4204–4205
- Music therapy, 2643–2644
- Mutagenesis
  - nalmefene and, 3364
  - naltrexone and, 3364
- Mutations, 242

- Mutism, 1194–1195, 4747
- Mutual analysis, 2751
- Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), 665
- Myasthenia gravis, 661
- Mycophenolate mofetil, neurotoxicity of, 2465*t*
- Mydriasis, 4747
- Myelencephalon, 9, 9*f*
- Myelination, 1509
- Myelin sheath, 5
- Myeloarchitecture, 5
- Myers, Charles, 4657
- Myers, Isabel, 976
- Myerson sign, 470
- Myocardial disease, 2282–2284
  - cardiomyopathies, 2284–2285
  - heart failure, 2282–2284
  - sexual dysfunction and, 4449
- Myocarditis, 1065
  - clozapine and, 3070
- Myoclonic epilepsy with ragged red fibers (MERRF), 702–703
- Myoclonus, 469, 481t, 3780*t*
- Myotonic dystrophy, 661–662
- Mysophobia, 4747
- Mysticism, 2670–2671

N

- NA. *See* Narcotics Anonymous (NA)
- NAB. *See Neuropsychological Assessment Battery* (NAB)
- Nabilone, 1346
- Nabiximol, 1346
- N-acetylcysteine (NAC), 2634–2635
  - for acute bipolar depression, 1735
  - for bipolar disorder, 187
  - for schizophrenia, 187
  - for substance use disorder, 4001*t*
- Nadolol (Corgard), 3224–3228
- Nail biting, 3809
- Nalmefene (Revex), **3357–3367**
  - alcohol use disorders, 3359–3360
  - chemistry, 3358
  - dosage and administration, 3366–3367
  - drug interactions, 3366
  - drug studies, 3358
  - history, 3358
  - laboratory interferences, 3366
  - for opioid use disorders, 1411
  - organ/system effects of, 3364–3365
  - pharmacodynamics, 3358
  - pharmacokinetics, 3358
  - pharmacologic actions, 3358
  - precautions/adverse reactions, 3365–3366
  - therapeutic indications, 3359–3364, 3365
- Naloxone (Narcan)
  - defined, 4747
  - for opioid overdose, 1411–1412, 3361
  - for opioid use disorders, 1411

- Naltrexone (ReVia; Depade), **3357–3367**
  - for alcohol use disorders, 1310
  - alcohol use disorders, 3359–3360
  - anxiety disorders, 3364
  - chemistry, 3358
  - chronic pain, 3363
  - cocaine use disorder, 3361–3362
  - depression, 3364
  - dosage and administration, 3366–3367
  - drug interactions, 3366
  - drug studies, 3358
  - eating disorders, 3362
  - history, 3357
  - immune system, 3363
  - laboratory interferences, 3366
  - methamphetamine use disorder, 3361
  - molecular structure, 3358*f*
  - nicotine use disorders, 3361–3362
  - obesity, 3362
  - for opioid use disorders, 1411, 1417–1418
  - opioid use disorders, 3360–3361
  - for opioid withdrawal, 1414
  - opioid withdrawal, 3361
  - organ/system effects of, 3364–3365
  - pathologic gambling, 3362–3363
  - pharmacodynamics, 3358
  - pharmacokinetics, 3358
  - pharmacologic actions, 3358
  - precautions/adverse reactions, 3365–3366
  - self-injurious behavior, 3362
  - therapeutic indications, 3359–3364, 3365
- Namenda. *See* Memantine (Namenda)

- NAMI. *See* National Alliance for the Mentally Ill (NAMI)
- Nanotechnology, 4469
- NARC. *See* National Association for Retarded Citizens (NARC)
- Narcan. *See* Naloxone
- Narcissism, 4747
- Narcissistic personality disorder, 2247–2248, 4748
  - diagnosis, 4837*t*
  - short-term dynamic psychotherapy for, 2773–2778
- Narcissistic-psychotic defenses, 920*t*
- Narcoanalysis. *See* Narcotherapy
- Narcolepsy, 534*t*, 710
  - animal models of, 303
  - antihistamines for, 3348
  - cataplexy and, 2175
  - characteristics of, 2174–2175
  - in children/adolescents, 4086
  - defined, 4748
  - diagnosis, 4812*t*
  - diagnostic criteria, 2175, 2175*t*
  - polysomnographic tracing, 2174*f*
  - treatment of, 2175
- Narcosis, 4748
- Narcosynthesis. *See* Narcotherapy
- Narcotherapy, 4748
- Narcotic, 4748
- Narcotic blockade, 4748
- Narcotics Anonymous (NA), 3900
- Nardil. *See* Phenelzine; Phenelzine (Nardil)
- Narrative, 733
  - definition of, 2875–2876
- Narrative cognition, 732
- Narrative family therapy, 2949

- Narrative psychiatry, **2874–2885**
  - character, 2877
  - first-person narratives, 2877–2880
  - future directions, 2884–2885
  - informed consent and, 2881
  - literary example, 2880–2881
  - metaphor, 2876
  - metatheoretical rationale for, 2881
  - neurochemical self and, 2883
  - pharmaceutical person and, 2883
  - plot, 2876–2877
- NAS. *See* Next Accreditation System (NAS)
- Natal gender, 2101
- Nathanson, Donald, 1973–1974
- National Alliance for the Mentally Ill (NAMI), 4748
- National Association for Retarded Citizens (NARC), 3632
- National Association of Inpatient Physicians, 4539
- National Board of Medical Examiners (NBME), 4590
- National Center for Biotechnology Information, 250*t*
- National Center for Complementary and Alternative Medicine (NCAAM), 4748
- *National Comorbidity Survey* (NCS), 887–888
- National Comorbidity Survey Replication Adolescent Supplement (NCS-A), 892
- National Council on Problem Gambling (NCPG), 1867
- National Cyber Security Alliance (NCSA), 3515
- National Drug Early Warning System (NDEWS), 1314
- *National Epidemiologic Survey on Alcohol and Related Conditions* (NESARC), 890–891, 891*t*
- *National Federation of Independent Business (NFIB) v. Sebelius*, 4398
- *National Health and Nutrition Examination Survey* (NHANES), 892

- *National Health Interview Survey* (NHIS), 892
- National Household Survey on Drug Use and Health (NSDUH), 1314
- National Human Genome Research Institute, 250*t*
- National Institute of Mental Health (NIMH), 2887, 4480, 4512, 4651–4652
- National Latino and Asian American Study (NLAAS), 887–888
- National Mental Health Act, 4512
- National Mental Health Association (NMHA), 4748
- National Organ Transplant Act (NOTA), 2449
- National Resident Matching Program (NRMP), 4591
- National Standards for Cultural and Linguistically Appropriate Services (CLAS), 4562
- *National Survey of American Life* (NSAL), 887–888
- *National Survey of Children's Health* (NSCH) survey, 892
- *National Survey of the Diagnosis and Treatment of ADHD and Tourette Syndrome* (NS-DATA), 892
- *National Survey on Drug Use and Health* (NSDUH), 888–889, 889*t*
- National Training Laboratories (NTL), 4748
- Nativism, 744
- Natural child, 963
- Natural group, 4748
- Natural products, 4142–4146, 4145*t*, 4146*t*
- Natural selection, 863–864
- Nausea, 1346, 2392
- NBME. *See* National Board of Medical Examiners (NBME)
- NBOMes, 1354
- NCAAM. *See* National Center for Complementary and Alternative Medicine (NCAAM)
- NCCAM. *See* National Center for Complementary and Alternative Medicine (NCAAM)
- NCS-A. *See* National Comorbidity Survey Replication Adolescent Supplement (NCS-A)

- Near-death experiences, 2681–2682
- Neck trauma, headache attributable to, 647
- Necromania, 4748
- Necrophilia, 4748
- Necrophobia, 4748
- Need, 4748
- Needle phobia, 4748
- Nefazodone (Serzone), **3204–3210**, 3416
  - absorption, 3204
  - anxiety disorders, 3205–3206
  - chemistry of, 3204
  - for depression in later life, 4297
  - discontinuation syndrome, 3209
  - distribution, 3204
  - drug interactions, 3209–3210
  - elimination, 3204
  - lactation and, 3209
  - major depression, 3205
  - metabolism, 3204
  - molecular structure, 3204*f*
  - mood disorders, 3205
  - organ/system effects of, 3207–3208
    - cardiovascular system, 3207
    - central nervous system, 3207
    - gastrointestinal system, 3207
    - genitourinary system, 3208
  - overdose, 3208–3209
  - pain, 3206
  - pharmacodynamics, 3204–3205
  - pharmacokinetics, 3204
  - precautions and adverse reactions, 3208–3209
  - pregnancy and, 3209

  - psychotropic agents and, 3209–3210
  - substance abuse, 3206
  - therapeutic indications, 3205–3207
- Negative affect states, 822
- Negative allosteric modulator (NAM), 115
- Negative hallucinations, 1980
- Negative life events, 1714
- Negative predictive value, 881
- Negative reinforcement, 1316, 4748, 4760. *See also* Reinforcement
- Negative symptoms
  - categories of, 1458–1459
  - course and associated features, 1459–1460
  - defined, 4748
  - primary, 1460
  - of schizophrenia, 1457–1460
- Negative transference, 2750, 4748. *See also* Transference
- Negative valence, 297–299
- Negativism, 4748
- Neglect
  - child, 3582–3583, 3968
  - family interventions and, 4379–4380
  - forensic evaluations, 4006
  - psychiatric emergencies, 3582–3583
- Neglect syndromes, 1022
- Nembutal. *See* Pentobarbital (Nembutal)
- Nemeroff, Charles, 1657
- Neocortex
  - *cis*-regulation of transcription in, 47–48
  - fetal, transcriptional organization of, 46
  - layers of, 17
  - regions, 21

  - transcriptional patterns, 46
- Neodissociation theory, 1931
- Neologism, 4748
- Neologisms, 1462
- NEO model, 601
- Neonatal abstinence syndrome, 3977
- Neonatal ventral hippocampus lesion (NVHL), 313
- Neonates
  - body ownership, 812
  - body schema, 812
  - interoceptive sensation-feelings, 812
  - reproductive development, 2580–2581
- NEO Personality Inventory 3 (NEO-PI-3), 976
- Neoplastic disorders, 1057
- Neostigmine (Prostigmin), 106
- Nernst potential, 68
- Nerve blocks, 2491
- Nerve doctors, 4645
- Nerve growth factor (NGF), 141, 2998
- *Nervios*, 2551
- Nervous breakdown, 4748
- Nervous system. *See also* Autonomic nervous system (ANS); Central nervous system (CNS) tumors; Peripheral nervous system (PNS)
  - embryonic, 55–56
  - overview of morphologic development, 54–60
  - proliferative zones, 56–58
  - regional differentiation, 55–56, 55*f*
- NESARC. *See National Epidemiologic Survey on Alcohol and Related Conditions* (NESARC)
- Net magnetization ratio, 327
- Netrins, 64

- Neupro. *See* Rotigotine (Neupro)
- Neural codes, 305–308
- Neural development, **53–67**
- Neural electrical activity, 305–308
- Neural excitability, 82
- Neural net profile, 727
- Neural network, 726–727
- Neural oscillations, 308–312
  - analysis of, 311–312
  - discoordination of neural discharge and, 312–313
  - phase-amplitude coupling, 312
  - synchrony, 310–311
- Neural patterning, 60–63
  - basal ganglia, 62–63
  - embryonic forebrain, 61–62
  - hippocampus, 62
  - principles of functions, 60–61
- Neural plasticity, 778–779
- Neural plate, 54–55
- Neural sciences, 1–3
- Neural signaling, cellular and synaptic basis, **67–82**
- Neurasthenia, 1689–1690, 2562–2563, 4748
- Neurobiology, 2780–2781
  - “Beyond the Pleasure Principle,” 2780
  - “On Narcissism: An Introduction,” 2780
  - “Project for a Scientific Psychology,” 2780
- Neuroborreliosis, 660
- Neurochemical receptors, 177
- Neurochemistry, 4748
- Neurocircuitry
  - biologically relevant, 1812
  - contextual fear structures, 1811

  - emotional memory, 1812
  - emotion processing and, 1811–1812
  - fear circuitry, 1808–1809
  - fear expression and, 1809–1810, 1810*f*
  - fear learning, 1807–1808
  - fear regulation, 1810–1811
- Neurocirculatory asthenia. *See* Da Costa syndrome
- Neurocognition
  - degeneration in, 1560–1561
  - functional capacity and, 1564
  - longitudinal course of, 1562–1563
  - in older adults, 1561
  - in schizophrenia, **1556–1566**
- Neurocognitive disorders, 1205, **1228–1292**
  - Alzheimer disease and, 4174
  - antihistamines for, 3348
  - antipsychotics for, 4411*t*
  - circadian rhythms and, 198
  - Creutzfeldt–Jakob disease, 4176
  - delirium, **1229–1238**
  - dementia with Lewy bodies, 4174–4175
  - diagnosis, 4829–4832*t*
  - due to another medical condition, **1289–1291**
  - due to multiple etiologies, 1287*t*, 1292
  - inhalant-induced, 1379
  - in long-term care, 4408–4412
  - major (dementia), **1238–1272**
  - malingering and, 2499
  - medications for, 3335*t*
  - overview, 1228–1229
  - Parkinson disease and, 4174–4175
  - Pick disease and other frontotemporal dementias, 4175

  - psychosocial/environmental interventions for, 4410*t*
  - rating scales for, 1112–1114
  - sleep in, 4215
  - substance/medication-induced, 1287*t*, 1291–1292
  - trazodone, 3201
  - unspecified, 1288*t*, 1292
  - vascular dementia, 4175–4176
- Neurocognitive impairment
  - assessment of, 1565–1566
  - in children, 1559–1560
  - costs of, 1564
  - functioning and, 1563
  - genetics of, 1559
  - medical comorbidities, 1563–1564
  - natural history of, 1559
  - prevalence of, 1559
  - profile and magnitude of, 1556–1559
  - quality of life, 1563
  - relapse prevention, 1563
  - in schizophrenia, 1556–1559
    - formal thought disorder, 1562
    - negative symptoms, 1561–1562
    - positive symptoms, 1561
  - treatment of, 1564–1565
  - unemployment and, 1563
- Neuro-Covid, 190
  - in mentally ill, 194
  - neurologic symptoms, 189
  - neuropsychiatric symptoms, 190
  - treatment of, 190
- Neurocysticercosis, 628
- Neurodegeneration with brain iron (NBIA), 704–708

    - mitochondrial membrane protein-associated neurodegeneration, 705
    - pantothenate kinase-associated neurodegeneration, 704–705
    - PLA2G6-associated neurodegeneration, 705
- Neurodegenerative disorders
    - association studies, 4173
    - clinical implications, 4177
    - dementia and, 1247
    - family studies, 4172–4173
    - future clinical directions, 4176
    - future research, 4177
    - genetics of, 4172–4177
    - genome-wide association studies, 4173–4174
    - neurotrophins and, 143
    - segregation and linkage studies, 4172–4174
    - twin studies, 4173
- Neurodegenerative language disorders, 1021
- Neurodermatitis, hypnosis for, 2937
- Neurodevelopment, extracellular factors, 63–65
- Neurodevelopmental disorders
    - changes in DSM, 1205
    - chromatin-mediated, 271*t*
    - diagnosis of, 4773–4780*t*
    - noncoding mutations in, 51–52
    - transcriptional and epigenetic dysregulation in, 50–53
- Neuroendocrine axis, 2416
- Neuroendocrine disorders, 709
    - adrenal disorders, 715–717
        - hyperaldosteronism, 716, 716*f*
        - hyperparathyroidism, 715
        - hypoadrenalism, 715–716
        - pheochromocytoma, 716–717

  - hypopituitarism, 712
  - hypothalamic disorders, 709–710
  - hypothalamic lesions, 710, 711*f*
  - narcolepsy, 710
  - neuroendocrine tumors, 717
  - neuropsychiatry of, 685–686, 709–717
  - parathyroid disorders, 714–715
    - hyperparathyroidism, 715
    - hypoparathyroidism, 715
  - pituitary disorders, 710
  - pituitary tumors, 711–712
  - thyroid and parathyroid disorders, 712–714
    - Hashimoto encephalopathy, 713, 713–714
    - hyperthyroidism, 712–713
    - hypothyroidism, 713
    - subclinical hypothyroidism, 713
- Neuroendocrine tumors, 717
- Neuroendocrinology, 4748
- Neurofeedback, 2644
  - for PTSD, 1883
- Neurofibromatosis type, 3641*t*
- Neurogenesis, **53–67**, 1508
  - postnatal, 65–67
  - radial and tangential patterns of, 58–60, 59*f*
  - regulation of, 63–65
- Neurohormone, 4748
- Neuroimaging, 476–477. *See also* Forensic psychiatry
  - in aging adult (*See* Geriatric psychiatry)
  - anxiety disorders, 1814–1821
  - brain stimulation and, 3467–3468
  - cognitive and emotional activation, 1814
  - of core psychoanalytic constructs, 2782

- drives and, 2782–2783
- fear, 1813
- focal neurobehavioral syndromes, 476–477
- functional, 1813
- generalized anxiety disorder, 1817–1818
- history, 4660–4661, 4702
- instincts and, 2782–2783
- opioid use disorder, 1402–1403
- outcome prediction, 2783
- panic disorder, 1814–1815
- in pediatric neuropsychiatric disorders, 3596–3617
  - alcohol use disorder, 3615
  - anxiety disorders, 3612
  - attention-deficit/hyperactivity disorder, 3604–3606
  - autism spectrum disorders (ASDs) and, 3616–3617
  - bipolar disorder, 3606–3608
  - cannabis use disorder, 3615
  - chemistry, 3598
  - childhood-onset schizophrenia, 3614–3615
  - critical considerations, 3600–3602
    - clinical factors, 3602
    - cognitive factors, 3602
    - neuroanatomical factors, 3600–3602
    - technical factors, 3600
  - depressive disorders, 3606–3608
  - feeding and eating disorders, 3615–3616
  - function, 3598–3600
  - generalized anxiety disorder, 3612
  - morphology/structure, 3596–3598
  - normal development, 3602–3604
  - obsessive-compulsive disorders, 3608–3611

    - pediatric autoimmune neuropsychiatric disorders associated with group A β-hemolytic streptococcal infection, 3611–3612
    - PTSD, 3612
    - recommendations for, 3601*t*
    - specific learning disorder, 3617
    - techniques, 3596
    - Tourette syndrome, 3612–3614
  - and psychodynamic principles, 2782–2783
  - schizophrenia
    - functional, **1522–1534**
    - molecular, **1535–1555**
    - structural, **1510–1521**
  - social anxiety disorder, 1816–1817
  - specific phobias, 1815–1816
  - structural techniques, 1813
  - traumatic brain injury and, 537–538
- Neuroleptic malignant syndrome (NMS), 663, 1065, 1196, 3092–3094
  - antipsychotics, 3063
  - asenapine, 3086
  - clinical features, 3047
  - clinical presentation, 3092–3093
  - clozapine, 3070
  - defined, 4748
  - differential diagnosis, 3093*t*
  - FGAs, 1581
  - iloperidone, 3083
  - lumateperone, 3089
  - management of, 3047
  - paliperidone, 3073
  - pathophysiology of, 3093
  - quetiapine, 3081

  - risks for, 3093
  - risperidone, 3073
  - tardive dyskinesia and, 3058
  - treatment of, dopamine receptor agonists, 3111
  - treatment options/implications, 3092–3094, 3094*t*
- Neuroleptics
  - cutaneous adverse events, 2447
  - defined, 4748
  - EEG effects, 379
  - for tic disorders, 3784–3785
- Neurologic amnesia, 4748
- Neurologic disorders
  - defined, 4748
  - developmental coordination disorder and, 3677–3678
  - diagnostic assessment, 1057
  - electroconvulsive therapy for, 3452
- Neurologic examination
  - abnormal reflexes, 470
  - eyes, 466
  - facial movement, 467
  - head, 466
  - movement abnormalities, 468–470
  - olfaction, 466–467
  - sensation abnormalities, 468–470
  - soft signs, 470
  - speech, 467–468
- Neurologic soft signs (NSS), 1509
- Neurologic syndromes, self and, 425–426
- Neurologist, 4748
- Neurology, 4748
- Neuromelanin, 4153
- Neurometabolic disorders, 685–717, 706–707*t*

  - chorea-acanthocytosis, 699, 700*f*
  - cysteine and homocysteine disorders, 699–700, 701*f*
  - enzyme deficiency disorders, 691–697
  - Fabry disease, 687–688, 687*f*
  - Gaucher disease, 688
  - GM2 gangliosidosis (Tay–Sachs disease), 688–689, 688*f*
  - ion channel disorders, 708–709
  - lysosomal disorders, 686
  - α-mannosidosis type II (AM), 689, 690*f*
  - metachromatic leukodystrophy, 686–687, 687*f*
  - metal metabolism disorders, 703–708
  - mitochondrial diseases, 700–703
  - neuronal ceroid lipofuscinosis, 689
  - neuropsychiatry of, 685–709
  - Niemann–Pick type C disease, 697, 698*f*
  - Pelizaeus-Merzbacher disease, 698–699, 699*f*
  - peroxisomal disorders, 689–691
- Neuromodulators, 161
- Neuromuscular disorders
  - COVID-19-related, 660–661
  - fibromyalgia, 664–665, 664*t*
  - myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), 665
  - neuroleptic malignant syndrome (NMS), 663
  - neuropsychiatric aspects of, **652–665**
  - serotonin syndrome, 663–664
  - stiff-person syndrome, 663
- Neuromuscular junction disorders, 661
- Neuronal assemblies, 804
- Neuronal ceroid lipofuscinosis, 689
- Neuronal hyperexcitability, 1657–1658
- Neuronal migration, 1508

- Neuronal nitric oxide synthase (nNOS), 125*f*
- Neuronal plasticity, 4748
- Neuronal process outgrowth, 65
- Neurons, 3
  - classification of, 3
  - defined, 4748
  - GABA, 18, 18*f*
  - immune receptors on, 177
  - local circuit, 3
  - major features of, 4*f*
  - orexins, 39, 39*f*
  - projection, 3
  - pyramidal, 17, 17*f*
- Neurontin. *See* Gabapentin (Neurontin)
- Neuropathic pain, 2388*t*, 2480*t*
  - associated with spinal cord injury, 3275
- Neuropathies
  - alcoholic, 657
  - chronic inflammatory demyelinating polyneuropathy, 657
  - of connective tissue disorders, 659
  - diagnosis of, 656
  - endocrine, 658
  - environmental, 658
  - frostbite, 658
  - granulomatous, 659–661
  - Guillain–Barré syndrome (GBS), 657
  - herpes zoster, 660
  - HIV-associated, 659–660
  - infectious, 659–661
  - ischemic, 658
  - leprous, 660
  - Lyme disease, 660

  - metabolic, 658
  - nutritional, 657
  - peripheral, 656–661
  - peripheral nerve injuries (PNI), 658
  - sarcoid, 661
  - sensory or sensorimotor, 656
  - toxic, 658
  - treatment of, 656–657
  - vasculitic, 659
- Neuropeptides, **108–124**
  - biosynthesis of, 111–113, 113*f*
  - cholecystokinin, 119–121
  - corticotropin-releasing factor, 117–119, 118*f*
  - distribution of, 113
  - as drug targets, 115–116
  - function of, 111–112
  - galanin, 122
  - neuropeptide Y (NPY), 122
  - neurotensin, 121–122
  - orexins, 122–123
  - oxytocin, 119–121
  - peptidases and, 115
  - regulation of, 113
  - signaling, 113–114
  - structures, 109*t*
  - substance P, 122
  - thyrotropin-releasing hormone, 116–117
  - transmitters, 109*t*
  - urocortins, 117–119
  - vasopressin, 119–121
- Neuropeptide Y (NPY), 122, **168**
  - amino acid sequence, 109*t*

  - anxiety disorders and, 1803–1804
  - PTSD and, 1879
- Neurophenomenology, 1952–1959
- Neurophysiologic biomarkers, 323–324
- Neurophysiology, 4748
  - brain stimulation and, 3467–3468
- Neuroplasticity
  - brain stimulation and, 3466
  - electroconvulsive therapy and, 3449
  - mood disorders and, 1770–1771
- Neuroplasticity hypotheses, 1660
- Neuropsychiatric disorders
  - ion channels and, 81
  - pharmacogenomics in, 250
- Neuropsychiatric Institute, 4654
- Neuropsychiatric Inventory (NPI), 1114
- Neuropsychiatric neuroanatomy, **459–462**
  - cerebellum, 460
  - cerebral cortex, 461
  - fronto-subcortical circuits, 459–460
  - lateralization, 459
  - limbic system, 460
  - white matter, 461–462
- Neuropsychiatric physical examination
  - extremities, 466
  - general, 466
  - neurologic examination, 466–470
- Neuropsychiatric sequelae of COVID-19 syndrome, in medically ill geriatric patients, 4205–4206
- Neuropsychiatry, 458–459, 1019
  - clinical evaluation, 464–465
    - appetitive symptoms, 465

    - birth history and early development, 464
    - cognitive symptoms, 465
    - handedness, 465
    - head injury, 464
    - neuropsychiatric history, 464
    - paroxysmal disorders, 465
    - personality change, 465
  - defined, 4748
  - modularity and, 463–464
  - of self, **416–431**
- Neuropsychoanalysis, 929–930
- Neuropsychological assessment, 479–480, 3563–3564
  - applications of, 3563–3564
  - assessment of functioning, 3563
  - test instruments, 3563
- *Neuropsychological Assessment Battery* (NAB), 1029
- Neuropsychological assessments, of adults, **1018–1034**
- Neuropsychological domains, 3567–3570
  - attention, 3568
  - executive functioning, 3569
  - language, 3569
  - memory, 3568
  - sensory/motor functioning, 3570
  - visuoperceptual functioning, 3569–3570
- Neuropsychological evaluation, **4166–4172**
  - activities of daily living/everyday functioning, 4171
  - anxiety, 4171
  - assessment domains, 4167–4171
  - attention, 4168–4169
  - cognitive screening, 4167–4168
  - common uses of, 4167
  - crystallized knowledge, 4168

  - definition, 4166
  - diversity, 4171
  - executive functions, 4170
  - expressive language, 4168
  - historical development, 4166
  - issues in, 4171
  - learning, 4169–4170
  - memory, 4169–4170
  - mental processing speed/psychomotor speed, 4170–4171
  - mood, 4171
  - receptive language, 4168
  - sensory impairment, 4171
  - telehealth assessment, 4171
  - test result interpretation, 4171
  - vigilance, 4168–4169
  - visuospatial cognition, 4170
  - working memory, 4170
- Neuropsychologist, 1018–1019, 4166
- Neuropsychotic conditions, 482–483
  - conversion disorder (functional neurologic symptom disorder), 484–485
  - dementia, 480
  - developmental disabilities, 483–484
  - epilepsy, 480–482
  - infectious diseases, 484
  - inflammatory diseases, 484
  - movement disorders, 483
  - traumatic brain injury (TBI), 482–483
- Neuroreceptors, 4748
- Neuroscience, 4749
- Neuroscience-based nomenclature (NbN), **3027–3033**
  - domains, 3032*f*

  - dopamine blockers, 3027–3030
  - dopamine enhancers, 3031
  - glossary, 3028–3030*t*
  - mood stabilizers, 3032
  - principles of, 3027
- Neurosis
  - defined, 4749
  - history, 4658
- Neurosis theory, 956–957
  - alienation, 957
  - claims, “shoulds,” and self-hatred, 957
  - idealized image, 957
  - neurotic pride and the pride system, 957
- Neurosteroids, 137–139
  - aging and, 139
  - anxiety disorders and, 138, 1802–1803
  - attention-deficit disorder and, 139
  - biosynthetic pathway of, 138*f*
  - depression and, 138
  - eating disorders and, 139
  - gynecologic disorders and, 139
  - memory disorders and, 139
  - neurodevelopment, 138
  - neuroprotection, 138
  - psychotic disorders and, 139
  - PTSD and, 1879
  - role in mental illness, 138–139
  - substance abuse and, 139
- Neurostimulation therapies. *See also* Electroconvulsive therapy (ECT)
  - deep brain stimulation, 4339
  - magnetic seizure therapy, 4339

  - for older adults, 4338–4340
  - repetitive transcranial magnetic stimulation, 4338–4339
  - transcranial alternating current stimulation, 4338
  - transcranial direct current stimulation, 4338
  - vagus nerve stimulation, 4339
- Neurosurgical treatments, **3479–3486**
  - clinical studies, 3481–3482
  - contraindications, 3480–3481
  - current state of, 3485–3486
  - deep brain stimulation, 3484–3485
  - future directions, 3486
  - history, 3480
  - indications, 3480–3481
  - lesion procedures, 3482–3484
    - anterior capsulotomy, 3482–3483
    - anterior cingulotomy, 3482
    - gamma ventral capsulotomy, 3483
    - laser interstitial thermotherapy, 3483
    - limbic leucotomy, 3484
    - subcaudate tractotomy, 3483–3484
  - multidisciplinary case review, 3481
  - patient selection, 3480–3481
  - postoperative care, 3481
  - preoperative workout, 3481
  - research, 3486
- Neurotensin (NT), 109*t*, 121–122, **170**
- Neurotic character, 2231
- Neurotic defenses, 921*t*, 2748*t*
- Neurotic disorder. *See* Neurosis
- Neuroticism, 979
- Neurotic pride, 957
- Neurotic process, 4749

- Neurotoxin, 4749
- Neurotransmission
  - cannabinoid receptors, 134
  - carbon monoxide and, 128–129
  - nitric oxide and, 125
- Neurotransmitter(s). *See also* Biogenic amine neurotransmitters
  - classes of, 77–78
  - conductance mechanisms, 78
  - defined, 4749
  - disturbances, 1766–1770
  - dopamine, 179–180
  - eicosanoids, 136–137
  - endocannabinoids, 130–136
  - gases as, 124–130
    - carbon monoxide, 128–130
    - hydrogen sulfide, 130
    - nitric oxide, 124–128
  - glutamate, 184–185
  - immunologic effects of, 179*t*
  - kynurenine (KYN) pathway, 180
  - neurosteroids, 137–139
  - norepinephrine, 179
  - novel, **124–139**
  - schizophrenia pathology, 1505–1507
  - serotonin, 179
  - structure of receptors, 78–79
  - temperament and, 3623–3624
  - violence and, 2516–2519
- Neurotrophic factors, **140–145**
- Neurotrophin receptors, 2999*f*
- Neurotrophins, 140
  - metabolism and, 143

  - neurodegenerative disorders and, 143
  - psychiatric disorders and, 143–144
  - receptors, 140–141
  - synapse, 142*f*
  - synaptic plasticity and, 142–143
  - therapeutic potential of, 144–145
- Neurulation, 54–55, 54*f*
- Neutropenia, 3086
- New Age spirituality, 2675
- New Beginnings Program (NBP), 4043–4044
- New Freedom Commission on Mental Health, 4542
- Newmann-Keuls procedure, 900
- Newton, Robert Denis, 4658
- New York Hospital, 4645, 4651
- New York/Presbyterian Westchester Division, 4651
- New York State Pathological Institute, 4651
- Next Accreditation System (NAS), 4596–4598
- Next-generation sequencing (NGS), 234, 235*f*
- NGS. *See* Next-generation sequencing (NGS)
- NHIS. *See National Health Interview Survey* (NHIS)
- Nialamide (Niamid), 3185*t*
- Nichols, Charles, 4510
- NICHQ Vanderbilt Assessment Scales, 3545, 3732
- Nicotine, 1385–1386. *See also* Smoking; Tobacco use disorder (TUD)
  - EEG effects, 380
  - organ transplantation and, 2459
- Nicotine dependence, 4749
- Nicotine gum, 1391*t*
- Nicotine lozenge, 1391*t*
- Nicotine patch, 1391*t*
- Nicotine replacement therapies, 1390, 4001*t*

- Nicotine withdrawal
  - clonidine for, 3222
  - guanfacine for, 3222
- Nicotinic acetylcholine receptors (nAChRs), 1543–1544
- Nicotinic receptors, 106
- Niemann–Pick type C disease (NPC), 697, 698*f*
- Nifedipine (Procardia), chemistry of, 3260*t*
- Night eating syndrome, 4749
- Night hospital, 4749
- Nightmare, 4749
- Nightmare disorder
  - defined, 4749
  - diagnosis, 4814–4815*t*
  - parasomnias, 2183
- Night owls, 196, 2180
- Night terror, 1188, 2182, 4749
- Nigrescence theory, 848
- Nihilism, 4749
- Nihilistic delusion, 4724, 4749. *See also* Cotard syndrome
- Nijenhuis, Ellert, 1967
- NIMH. *See* National Institute of Mental Health (NIMH)
- Nin, Anaïs, 422
- 9q34 subtelomeric deletion syndrome, 271*t*
- Nitric oxide, 124–128
  - behavior and, 126–127
  - cyclic GMP pathway, 125, 126*f*
  - mechanism of action, 125
  - mood disorders and, 127
  - neuropathologic roles of, 127–128
  - neurotransmission, 125
  - schizophrenia and, 127
  - S-nitrosylation pathway, 125, 126*f*

  - synthesis, 125, 125*f*
- Nitric oxide enhancers, 2052–2053
- Nitrogen narcosis. *See* Rapture-of-the-deep syndrome
- Nitrous oxide, 1370*t*
  - pain management in burn patients, 2470
    - relaxation techniques, 2470
- Nitrous oxide-related disorders, 1380–1381
- NMDA antagonists, for self-injurious behaviors, 3651
- NMDA receptor antagonists, 1351–1353
- NMDA receptors, 84–86, 1767–1770
  - depersonalization/derealization and, 1954–1955
  - schematic representation of, 85*f*
- N-methyl-D-aspartate (NMDA), 1298
- NMS. *See* Neuroleptic malignant syndrome (NMS)
- Nociception, 2480*t*
- Nociceptive pain, 2388*t*, 2480*t*
- Noctec. *See* Chloral hydrate
- Nocturnal enuresis, 4087
- Nodes of Ranvier, 70
- Noesis, 805, 4749
- Noguchi, Hideo, 4660, 4687
- Nolen-Hoeksema, Susan, 1660
- Nolvadex. *See* Tamoxifen
- Nominal aphasia, 4749
- Non–24-hour sleep–wake circadian rhythm disorder, 4749
- Nonaffective hallucination, 4749
- Nonaffective psychosis, 2599
- Nonafluent aphasia. *See* Motor aphasia
- Nonbinary, 2102
- Nonbinary identities, 4100, 4100*t*
- Noncoding genes, 239–240
- Non-coding RNAs (ncRNAs), 2229–2230

- Noncompliance with medical treatment, 4749
- Nonconvulsive electrotherapy, 3478–3479
  - contraindications, 3478
  - definition of, 3478
  - future directions, 3478
  - mechanism of action, 3478
  - side effects, 3478
  - treatment-resistant depression, 3478
- Nondeclarative memory, 730, 730*t*, 774
- Nondirective approach, 4749
- Nonfluent aphasia, 1182–1183, 4749
- Nongay/non–lesbian-identified homosexual individuals, 2075
- Nonketotic hyperglycemia (NKH), 696
- Nonmaleficence, 4012
- Nonopioid analgesics, 2389
- Nonparametric, 4749
- Nonpathologic amnesia, 2005–2006
- Nonprobability sampling, 879
- Nonpyramidal cells, 18
- Non-rapid eye movement (NREM) sleep arousal disorders, 2182–2183, 4813*t*
- Nonreactive depression. *See* Endogenous depression
- Nonreflective behavior, 2751
- Nonseasonal unipolar depression, 197
- Nonself, 426
- non–small cell lung cancer (NSCLC), 2362–2363
- Nonsteroidal anti-inflammatory drugs (NSAIDs), 2389
  - for Alzheimer disease, 1254–1255
  - augmentation therapy, 184
  - for chronic pain, 2487
  - for depressive disorders, 184
  - monotherapy, 184

  - pain management in burn patients, 2470
  - for schizophrenia, 186
  - for tension-type headaches, 651
- Nonsuicidal self-injurious behavior (NSSI), 3821
- Nonsuicidal self-injury, 4749
- Nonverbal communication, 1000, 1002
- Nonverbal learning disabilities (NVLDs), borderline intellectual functioning and, 2533
- Nootropics, 4320–4321
- Noradrenaline. *See* Norepinephrine
- Noradrenaline modulators, **3219–3228**
  - α-adrenergic receptor agonists, **3219–3224**
  - β-adrenergic receptor antagonists, **3224–3228**
  - clonidine, **3219–3224**
  - guanfacine, **3219–3224**
- Noradrenergic systems, 1760–1762, 1761*f*, 1796–1797
- Norepinephrine, 179, 1796–1797
  - anatomy and functions, 94–95
  - brain functioning and, 42
  - defined, 4749
  - history, 4662
  - life cycle, 95–98
  - plasma membrane transporter, 97*f*, 98
  - synthesis, 95–96, 95*f*
- Norepinephrine-dopamine reuptake inhibitors (NDRIs), 98
- Norepinephrine reuptake inhibitors (NRIs), for ADHD, 3932–3933
- Norepinephrine transporters (NETs), 98
- Normal aging, 4120–4121
- Normal autism, 922
- Normal equations, 906
- Normality, 4749
  - mental health and, **784–800**

    - future directions, 798–800
    - healthy involuntary mental mechanisms, 793
    - model A: mental health as above normal, 785–788
    - model B: mental health as positive psychology, 788–790
    - model C: mental health as maturity, 790–791, 790*f*
    - model D: mental health as resilience, 791–793
    - model E: mental health as socioemotional intelligence, 794
    - model F: mental health as subjective well-being, 794–797, 796*f*
    - model G: mental health as positive or “spiritual” emotions, 797–798
- Norman, Warren, 975–976
- Normative data, 1024–1025
- Norpramin. *See* Desipramine
- North American Prodrome Longitudinal Study (NAPLS), 1514–1515
- Nortriptyline (Aventyl, Pamelor), 3152
  - pharmacogenetics of, 295*t*
  - pharmacokinetics, sex differences in, 3416
- Nosology, 4749
- Nosophobia, 4749
- NOTA. *See* National Organ Transplant Act (NOTA)
- Note taking, 1004
- Not-for-profit hospital, 4749
- Not otherwise specified (NOS), 4749
- Novelty seeking, 2222–2223, 2222*t*
- NPC. *See* Niemann–Pick type C disease (NPC)
- NPI (Neuropsychiatric Inventory), 1114
- NREM (nonrapid eye movement) sleep, 200–201
  - defined, 4749
  - dreaming, 211–212
  - neurobiology of, 206–207
- NRMP. *See* National Resident Matching Program (NRMP)

- NSUDH. *See National Survey on Drug Use and Health* (NSDUH)
- NTL. *See* National Training Laboratories (NTL)
- Nuclear DNA of mitochondrial origin (NUMTs), 242
- Nuclear family, 4749
- Nuclear magnetic resonance (NMR) imaging
  - basic physics of, 327–330
  - system hardware, 330–332
- Nuclear Overhauser effect (NOE), 348
- Nucleic acids, 235
- Nucleosome- free region (NFR), 275, 275*f*
- Nucleosomes, 236, 271
- Nucleotides, 235
- Nucleus intermedius, 2069
- Null chromatin, 278
- Null hypothesis, 4749
- Nurse practitioners, 4029
- Nursing homes. *See also* Long-term care (LTC), sleep in, 4215
- Nurturing, 2520
- Nurturing parent, 963
- Nutrition, immune-brain interactions and, 191
- Nutritional neuropathies, 657
- Nutritional supplements, 4142–4146
- Nuvigil. *See* Armodafinil (Nuvigil)
- Nyctophobia, 4749
- Nymphomania, 2049, 2121, 4749
- Nyquist frequency, 361
- Nyquist theorem, 361

O

- *Obergefell v. Hodges*, 2079
- Obesity, 1597, **2304–2316**
  - behaviors and, 2307

- biology of, 2305–2306
- bupropion for, 3166
- classification, 2304*t*
- clinical features, 2304
- cognitions and, 2307
- commercial diets for, 2309–2310, 2309*t*
- course and prognosis, 2305
- definition, 2304
- diagnosis, 2304
- environmental factors, 2306
- epidemic, 2326
- epidemiology of, 2304
- etiology of, 2305–2307
- food addiction and, 2306–2307
- genetics of, 2305–2306
- immune-brain interactions and, 191
- medical comorbidities of
  - in adolescents, 2306*f*
  - in adults, 2305*f*
- naltrexone for, 3362
- nosology, 2304
- pathology of, 2305
- personality dimensions and, 2307
- prevalence of, 2304, 2304*t*
- psychiatric comorbidities of, 2307–2309
  - anxiety, 2308
  - depression, 2308
  - eating disorders, 2307–2308
  - severe mental illness, 2308–2309
  - substance use disorders, 2309
- topiramate for, 3269
- treatment of, 2309–2316

    - medical, 2314–2315, 2314*t*
    - psychological, 2311, 2311–2314
        - cognitive behavioral therapy, 2311
        - motivational interviewing, 2311–2312, 2312*t*, 2313*t*
    - surgical, 2315–2316, 2315*f*
- Obesogenic environment, 2306
- Object constancy, 923
- Objective burden, 4079
- Objective personality tests, 1038–1043
    - Millon Clinical Multiaxial Inventory-III, 1042
    - Minnesota Multiphasic Personality Inventory-2 (MMPI-2), 1038–1041
    - Minnesota Multiphasic Personality Inventory-3 (MMPI-3), 1042
    - MMPI-2-Restructured Form, 1041–1042
    - Personality Assessment Inventory, 1043
    - Personality Inventory for DSM-5-Adult, 1042
    - Personality Inventory-Revised, 1043
    - rapid personality assessment instruments, 1043, 1044*t*
- Objective tests, 4749
- Object loss, 1654
- Object permanence, 4749–4750
    - schema of, 746
- Object personality tests, 1038–1043
- Object relation, 4750
- Object relations, 422
    - as basic building units of mind, 806–808
    - definition of, 810
    - dynamical systems theory and, 811
    - as episodic memory, 814
    - integrated, 817–819
    - neuroscience of, 811
    - partial, 814–816

  - prenatal somatosensory origins of, 811
  - therapeutic actions, 2752
- Object relations theory, 422, 923–926
  - basics of, 810–811
  - neuroscience of, **800–823**
- Object representation, 814, 817–818
- OBS. *See* Organic brain syndrome (OBS)
- Observational studies, 878–879
  - analytic studies, 878
  - case-control studies, 878
  - cohort studies, 878
  - descriptive studies, 878–879
- Obsession, 4750
- Obsessional jealousy disorder, 4750
- Obsessive-compulsive disorders (OCDs), **1840–1845**, 2250–2251
  - bipolar disorder and, 1742–1743
  - body dysmorphic disorder, 2436–2437
  - body-focused repetitive behaviors, 2437–2438
  - buspirone for, 3170
  - in children, **3837–3844**
    - behavioral approach tasks for, 3888*t*
    - clinical features, 3840–3841
    - cognitive factors, 3840
    - comorbidities, 3838, 3838*t*
    - comparative nosology, 3838
    - course and prognosis, 3842
    - definitions of, 3837
    - diagnosis, 3840–3841
    - differential diagnosis, 3841–3842
    - DSM-5 criteria for, 3838
    - epidemiology of, 3838
    - etiology of, 3838–3840

    - genetic factors, 3839–3840
    - group psychotherapy for, 3905
    - history, 3837–3838
    - neurobiological factors, 3838–3839
    - pharmacotherapy for, 3930
    - symptoms, 3838*t*
    - treatment of, 3842
        - cognitive behavioral therapy, 3842–3843
        - pharmacotherapy, 3843–3844
- clinical features, 1842–1844
- cognitive behavioral therapy for, 3894
- comparative nosology, 1840–1841
- course and prognosis, 1844
- defined, 4750
- definition of, 1840
- diagnosis, 1842–1844, 4798*t*, 4837*t*
- differential diagnosis, 1844
- DSM-5 criteria for, 1841*t*, 1843*t*
- due to another medical disorder, 1854
- EEGs in, 376
- epidemiology of, 1841*t*
- excoriation, 2435–2436
- history, 1840
- hypnosis and, 2929
- ICD-11 criteria for, 1223, 1841*t*
- immune-CNS interactions in, 189
- key points, 1846*f*
- laboratory examination, 1844
- neuroimaging
    - amygdala, 3611
    - basal ganglia, 3608
    - prefrontal cortex, 3608–3610

    - thalamus, 3610–3611
  - pathology of, 1844
  - perinatal, 2600
  - psychedelics for, 2907
  - psychodynamic therapy for, 2762–2763
  - rating scales for, 1108
  - rheumatic heart disease and, 2285–2286
  - in schizophrenia, 1463–1464
  - selective mutism and, 3864
  - SSRIs for, 3131
  - substance/medication-induced, 1854
  - tic disorders and, 3781
  - traumatic brain injury and, 683
  - trazodone, 3199
  - treatment of, 1844–1845, 3651
    - combined, 1845
    - pharmacotherapy, 1844–1845
    - psychotherapy, 1845
    - somatic therapies, 1845
  - trichotillomania (hair-pulling disorder), 2434–2435
- Obsessive-compulsive personality, defined, 4750
- Obstetrics, hypnosis in, 2931
- Obstructive interventions, 1001–1002, 1001*t*
- Obstructive sleep apnea hypopnea (OSAH), 2176–2179
  - comorbidities, 2176–2177
  - diagnostic criteria, 2176
  - pediatric diagnostic criteria, 2176
  - treatment of, 2177
    - oral appliance, 2178
    - positional therapy, 2178
    - positive airway pressure, 2177–2178
    - surgical, 2178–2179

- Obstructive sleep apnea syndrome (OSAS), 4084
  - in children, 4084–4085
  - defined, 4750
  - in older adults, 4213–4214
- Occasion setting, 766
- Occipital lobe, 9
- Occupational problem, 4750
- Occupational psychiatry, 4750
- Occupational therapist, 4029
- Occupational therapy
  - defined, 4750
  - for developmental coordination disorder, 3680
  - for dissociative identity disorder, 2000
- OCDs. *See* Obsessive-compulsive disorders (OCDs)
- OCF. *See* Outline for Cultural Formulation (OCF)
- *O'Connor v. Donaldson* (1975), 4513
- Odbert, Henry, 975
- ODD. *See* Oppositional defiant disorder (ODD)
- Oddball-evoked responses, 1570–1571
- Odds ratios in epidemiology, 876–877
  - attributable fraction, 877
  - confounding, 877
  - effect modification/moderation, 877
  - mediation, 877
- Oedipal phase. *See* Phallic phase
- Oedipus complex, 3504, 4750
- of anorexia nervosa, bulimia nervosa, 2150*t*
- Ogden, Thomas, 813
- Ohm's law, 3445
- Öhrström, Wilhelm, 4644
- Olanzapine (Zyprexa), 99, 3056–3058
  - for acute bipolar depression, 1734

  - for bipolar disorder, 1730*t*, 1738
  - for bipolar disorder with acute depression, 3058–3060
  - for bipolar II disorder, 1741
  - for cancer patients, 2348
  - chemistry, 3056
  - for delirium, in cancer patients, 2352
  - for dementia, 1250
  - dosage and administration, 3059
  - drug interactions, 3059
  - for early-onset bipolar disorder, 3834
  - effects on organs and systems, 3057
  - history, 3008, 4698
  - long-acting injectable, 3059–3060
  - for mania and mixed states in bipolar disorder, 3057
  - molecular structure, 3056*f*
  - pharmacodynamics, 3057
  - pharmacokinetics, 3056–3057
  - pharmacokinetics of, sex differences in, 3417
  - for schizophrenia, 3057–3058, 3059
  - therapeutic indications, 3058
  - for tic disorders, 3785
  - treatment-resistant major depressive disorder, 3058–3060
- Older adult abuse, **4432–4438**
  - assault and battery, 4433
  - court system and, 4435
  - COVID-19 and, 4437
  - definitions of, 4433
  - emotional/psychological, 4433
  - ethical analysis, 4436–4437
  - excuses and justifications in, 4436
  - factors in, 4434
  - financial, 4433, 4435–4436

  - multidisciplinary teams (MDTs), 4434–4435
  - neglect, 4433
  - nosology, 4433–4434
  - profile of financial abuser, 4437–4438
  - profile of physical abuser, 4437
  - red flags and indicators of, 4435
  - sexual assault, 4433
  - shoves *versus* nudges, 4437
  - threats *versus* offers, 4437
- Older adults
  - frequency of sexual expressions in, 4441–4443
  - gender and sexual minorities, 4441–4443
  - retirement, 4161
  - role transitions, 4161
  - sexuality of, 4162–4163, 4438–4457
    - age-related changes in sexual function and behavior, 4440–4443, 4440*t*
    - cognitive capacity and sexual behaviors, 4454
    - frequency of sexual expressions in, 4440–4442
    - LGBTQ+, 4443–4445, 4453–4454
    - recommendations for PRC communities, 4454–4455
    - sexual behavior in long-term care, 4452–4455
    - sexual dysfunctions
      - arousal phase problems, 4448
      - assessment, 4446–4447
      - desire phase problems, 4447–4448
      - illnesses associated with, 4449–4451, 4449*t*
      - medications associated with, 4451–4452
      - orgasm phase problems, 4448
      - treatment, 4447–4451
    - sexual history, 4445–4446
  - social functioning, 4162–4163

- spousal bereavement, 4161
- substance abuse in, 4133, 4277–4290
  - barriers to detection, 4277*t*
  - benzodiazepines, 4286–4287
  - history, 4278
  - illicit drugs, 4288
  - nonpharmacologic therapies, 4288–4289
    - brief interventions, 4289
    - intensive outpatient programs, 4290
    - moderators and correlates of treatment, 4290
    - partial hospitalization, 4290
    - psychosocial interventions, 4289
    - twelve-step recovery programs, 4289–4290
  - opioid analgesics, 4287–4288
  - signs and symptoms, 4279*t*
  - tetrahydrocannabinol, 4285–4286
- suicide in, 2378, 2711, 4231
  - risk factors, 2711
  - treatment, 2711, 2718
- technology for, 4464–4470
  - aging in place and, 4465
  - barriers and limitations, 4476
    - age-related impairments, 4476
    - billing, 4476
    - ethics, 4476
    - information overload, 4476
    - integration with health care system, 4476
    - lack of regulation and evidence, 4476
    - literacy, 4476
    - privacy, 4476
    - usability, 4476
  - classification, 4464*t*

        - comparative nosology, 4464–4465
        - future directions, 4469–4470
            - big data, 4469–4470
            - nanotechnology, 4469
            - robotics, 4469
        - specific applications of, 4465–4471
            - caregiver applications, 4469
            - games, 4469
            - mobile health, 4465–4467
            - sensors/self-monitoring devices, 4467, 4467*t*
            - smart homes, 4468, 4468*t*
            - social media, 4469
            - tablets, 4469
            - telehealth, 4465
- Olds, James, 798
- Olfaction, 466–467
- Olfactory aversion, 2817
- Olfactory delusions, 1170
- Olfactory hallucinations, 1174, 1455, 4750
- Olfactory reference syndrome, 4750
- Oligodendrocytes, 5, 5*t*
- Oligophrenia, 4750
- Olivopontocerebellary atrophy (OPCA). *See* Multiple system atrophy (MSA)
- *Olmstead v. L.C.* (1999), 4513–4516
- OMD. *See* Organic mental disorder (OMD)
- Omega-3 fatty acids, 2635–2636, 2653–2654, 4144–4145
    - for acute bipolar depression, 1735
    - ADHD and, 2654
    - Alzheimer disease and, 2654
    - anxiety and, 2653–2654
    - ASD and, 2654

  - for bipolar disorder, 188
  - depression and, 2653
  - mild cognitive impairment and, 2654
  - for schizophrenia, 186–187
- Omnibus Budget Reconciliation Act (OBRA), 4512
- Omnipotent control, 2748*t*
- OMS. *See* Organic mental syndrome (OMS)
- OnabotulinumtoxinA, 650
- Onanism, 2035, 4750. *See also* Masturbation
- Ondansetron (Zofran), for tardive dyskinesia, 3096
- 1-year prevalence, 874
- “One gene, one enzyme” hypothesis, 236
- Oneiroid state, 4750
- Oneirophrenia, 1628*t*
- Online pharmacies, 1126–1127
- Online video games. *See also* Internet gaming disorder
  - benefits of, 4091
  - risks of, 4091
- Onset-time, 3656*t*
- Ontogenetic, 4750
- Onychophagia, 3809
- Onychotillomania, 3812
- Open Dialogue, 1609
- Open-ended questions, 989
- Open hospital, 4750
- Open neurosensory fixation induction, 2934, 2935*f*
- Operant, 2816
- Operant behavior. *See* Conditioning
- Operant conditioning, 756, 1315, 4750. *See also* Conditioning
- Operant/instrumental learning, 761–764
  - behavior and payoff, 761–763
  - motivation, 763–764

  - reinforcement, 763
- Operational definition, 4750
- Operations, 745, 747
- Opiate, 4750
- Opiate receptor, 4750
- Opioid agonists
  - acute pain management, 3355–3356
  - chronic pain management, 3356
  - for opioid use disorders, 3351
- Opioid analgesics, 2389, 2487–2488, 2487*t*, 4287–4288
  - epidemiology, 4287
  - etiology, 4287
  - pharmacologic treatment, 4288
  - screening, 4287
  - withdrawal, 4287
- Opioid antagonists, **3357–3367**
  - nalmefene, **3357–3367**
  - naltrexone, **3357–3367**
  - for opioid use disorders, 1411
  - for self-injurious behaviors, 3651
- Opioid-induced mental disorders, 1405–1406
- Opioid intoxication
  - atypical, 1407
  - comparative nosology, 1397
  - differential diagnosis, 1407
  - DSM-5 criteria for, 1404, 1404*t*
- Opioid peptide, 4750
- Opioid receptors, 1399–1401
  - pharmacology of, 1401
  - structural elements of, 1401*f*
- Opioids (opiates), 1399–1401
  - chemical structures of, 1400*f*

  - defined, 4750
  - description of, 1395–1396
  - EEG effects, 380
  - endogenous, 161–162
  - history, 1396–1397
  - intoxication, 2727–2728
  - neural pathway, 446
  - nonmedical use of, 1399
  - overdose, 1398, 1399*f*, 1411–1412, 3352
  - pain management in burn patients, 2469
  - PET scanning of, 1545
  - routes of administration, 2390*t*
  - sexual dysfunction and, 2047
  - tolerance, 1404–1405
  - withdrawal, 2729–2730
- Opioid use, 1396–1397
  - crime and, 1408
  - epidemiology of, 1397–1398, 1398*f*
  - genomics of, 1403
  - molecular genetics of, 1403
  - overdose, 1398, 1399*f*
  - prevalence of, 1398
  - psychiatric disorders in, 1410*t*
  - relapse, 1402
  - worldwide trends, 1399, 1400*f*
- Opioid use disorders (OUD), **1395–1418**
  - comorbidities, 1409
  - comparative nosology, 1397
  - course and prognosis, 1407–1409
  - diagnosis, 1404–1407
  - differential diagnosis, 1407
  - DSM-5 criteria for, 1406, 1406*t*

  - etiology of, 1399
  - laboratory examination, 1406–1407
  - long-term prognosis, 1407–1408
  - morbidity, 1408–1409
  - mortality, 1408–1409
  - neural circuits in, 1401–1402
  - neurobiology of, 1399–1402
  - neuroimaging, 1402–1403
  - in older adults, 4287
  - opioid agonist therapies for, 3351
  - overdose, 1409
  - pathogenesis of, 1399
  - pathology of, 1406–1407
  - patterns of use, 1407–1408
  - physical dependence, 3350–3351
  - prevalence of, 1398
  - psychosocial factors, 1403–1404
  - relapses, 1407–1408
  - treatment of, 1409–1411, 1415–1418, **3350–3367**
    - buprenorphine, 1416–1417, 3353–3355
    - methadone, 1415–1416, 3351–3353
    - naltrexone, 1417–1418, 3360–3361
    - oral morphine, 1418
    - pharmacologic, 1415–1418, 1416*t*
    - prescription heroin, 1418
    - psychedelics, 2907
    - psychosocial interventions, 1415
    - in special populations, 1418
  - withdrawal, 3350–3351
  - worldwide trends, 1399
- Opioid withdrawal (detoxification), 1405
  - agonist-assisted, 1413

  - antagonist-assisted, 1414
  - clonidine for, 3222
  - comparative nosology, 1397
  - differential diagnosis, 1407
  - DSM-5 criteria for, 1405
  - guanfacine for, 3222
  - precipitated, 1405
  - relapse after, 1414–1415
  - spontaneous, 1405
  - trajectory of severity, 1412*f*
  - treatment of, 1412–1415
    - buprenorphine, 1413–1414
    - methadone, 1413
    - naltrexone, 3361
- Opium, 4646
- Oppenheim, Hermann, 486
- Opponent process theory, 442–443, 443*f*
- Oppositional defiant disorder (ODD), **3761–3765**. *See also* Conduct disorder
  - biologic factors, 3749
  - clinical features, 3749
  - comorbid conditions, 3751
  - course and prognosis, 3751
  - defined, 4750
  - definition of, 3747–3748
  - diagnosis, 3749, 4791*t*, 4822*t*
  - differential diagnosis, 3750–3751
  - DSM-5-TR criteria for, 3750*t*
  - epidemiology, 3748
  - etiology, 3748–3749
  - family therapy for, 3915–3918
  - history, 3748

  - intellectual disabilities and, 3647
  - laboratory examination, 3750
  - pathology, 3750
  - psychological factors, 3749
  - psychostimulants for, 3124–3125
  - sociologic factors, 3749
  - traumatic brain injury and, 683
  - treatment, 3737–3738, 3751
- Opsoclonus-myoclonus ataxia syndrome, 2365
- Optic expansion, 3502
- Optic radiation, 11
- Optimism, 4473
- Optogenetics, 815
- Oral contraceptives, pharmacokinetics/pharmacodynamics, 3413–3414
- Oral dyskinesia. *See* Tardive dyskinesia (TD)
- Oral hypoesthesia, 3086
- Oral phase, 4750
- Oral prostaglandin (Vasomax), 2052
- Oral stage, 916, 918*t*
- Ordered subset expectation maximization (OSEM), 390, 390*f*
- Ordinal classification, 1091
- Ordinary least squares (OLS), 906
- Orexin A, 122–123
- Orexin B, 109*t*, 122–123
- Orexrin neurons, 39, 39*f*
- Organ donation, 2402
- Organic, 4750
- Organic anxiety syndrome, 4750
- Organic brain syndrome (OBS), 4750
- Organic delusional disorder, 4750
- Organic disorders

  - defined, 4750
  - differential diagnosis, 3872–3873
- Organic hallucinosis, 4750
- Organic mental disorder (OMD), 4750
- Organic mental syndrome (OMS), 4750
- Organic mood syndrome, 4751
- Organic personality syndrome, 4751
- Organic psychosis. *See* Organic mental disorder (OMD)
- Organized crime, 2509
- Organized Crime Control Act of 1970, 2509
- Organophosphates, impact on CNS and PNS, 655*t*, 659
- Organ Procurement and Transplantation Network (OPTN), 2449
- Organ transplantation, **2449–2465**. *See also* Heart transplants
  - behavioral contracting, 2455
  - clinical evaluation for, 2450–2452
  - comparative nosology, 2449
  - course and prognosis, 2450–2457
  - definition of, 2449
  - dialysis and, 2461
  - diseases resulting in, 2451*t*
  - early posttransplant adaptation, 2457
  - epidemiology of, 2449
  - etiology of, 2450
  - living donor
    - evaluation of, 2454–2455, 2454*t*
    - outcomes, 2462
  - long-term outcome, 2457
  - neurotoxicity of immunosuppressant medications, 2463–2465
  - perioperative period, 2457
  - preparation, 2455
  - psychiatric disorders affecting, 2458–2461
    - alcohol use disorder, 2459

    - anxiety disorders, 2458
    - cognitive disorders, 2459–2460
    - hepatic encephalopathy, 2460–2461
    - mania, 2458
    - mood disorders, 2458
    - nicotine dependence, 2459
    - posttraumatic stress disorder, 2458
    - psychotic disorders, 2458
    - substance use disorders, 2459
    - uremic encephalopathy, 2461
- psychosocial criteria, 2452
    - evaluator’s role, 2452–2453
    - goals of pretransplant evaluation, 2452*t*
    - screening and content of evaluation, 2452*t*, 2453
        - emergency evaluation, 2454
        - informed consent in terminally ill transplant candidates, 2454
        - living donor, 2454–2455, 2454*t*
        - social support, 2453
        - substance use, 2453–2454
        - treatment adherence, 2453
- special issues in, 2461–2465
- substance use-disordered patient, 2456
- survival by organ type, 2450*f*
- timeline for, 2451*f*
- transplant-related illness behavior, 2457
- treatment issues
    - adherence to posttransplant regimen, 2462–2463
    - psychopharmacologic, 2463, 2464*t*
    - psychotherapeutic, 2462
- ventricular assist devices and, 2461–2462
- waiting list for, 2450*f*
    - candidate reassessment, 2456

  - waiting period, 2456–2457
- Orgasm, 4751
- Orgasmic disorders, 2042–2043, 4751
- Orgasmic phase, 4751
- Orgasmic platform, 4751
- Oribasius, 1239
- Oriental traditional medicine, 4146–4147
- Orientation, 4751
- Original thirteen. *See* American Psychiatric Association (APA)
- Orlistat, for weight loss, 2314
- Orne, Marin, 2020
- Ornithine transcarbamylase deficiency, 3642*t*
- Orofacial pain, 2484–2485
- Orphan drug, 4751
- Orthodox sleep, 200
- Orthographic transparency, 3656*t*
- Orthography, 3656*t*
- Orthomolecular psychiatry, 4751
- Orthomolecular treatment, 4751
- Orthopsychiatry, 4751
- Orthostatic hypotension, 3069, 3074, 3081, 3084, 3086, 3158
  - defined, 4751
  - monoamine oxidase inhibitors (MAOIs), 3191
- Orton, Samuel, 4653
- OSA. *See* Obstructive sleep apnea syndrome (OSAS)
- Oscillations, 324
- Oscillatory neuronal firing, 79–80
- Osmond, Humphrey, 432
- Osmotic-release methylphenidate (OROS-MPH), 4002*t*
- Othello syndrome, 4721, 4751
- Other-directed person, 4751
- Other interpersonal problems, 4751

- Other specified family circumstance, 4751
- Other specified feeding and eating disorders (OSFED), 2148
  - epidemiology of, 2150*t*
- Other specified personality disorder, 2251
- Other specified schizophrenia spectrum and other psychotic disorder, **1645–1646**
  - course and prognosis, 1646
  - diagnostic and clinical features, 1646
  - differential diagnosis, 1646
  - epidemiology of, 1646
  - etiology of, 1646
- Otsuka Long Evans Tokushima Fatty (OTELF) rats, 230
- Outcomes, 789
- Outcome variable. *See* Dependent variable
- Outer subventricular zone (OSVZ), 58
- Outline for Cultural Formulation (OCF), 2571–2572, 2572–2574
- Outpatient, 4751
- Outpatient services, family interventions and, 4375
- Ovarian dysgenesis. *See* Turner syndrome
- Overactivity, 1194, 4751
- Overanxious disorder, 4751
- Overcompensation, 4751
- Overdetermination, 4751
- Over-fitting, 908
- Overshoot, 70
- Overt Aggression Scale (OAS), 3546
- Overt homosexuality, 4751
- Overvalued idea, 4751
- Ovulatory dysfunction, 2584
- Oxazepam (Serax), 1429*t*
  - for delirium, 4261
  - diazepam and, 290

  - pharmacokinetics of, sex differences in, 3417
- Oxcarbazepine (Trileptal), 1066
  - for acute mania, 1729
  - for chronic pain, 2490
  - for dementia, 1251
  - for descending neuropathic pain, 218
  - for early-onset bipolar disorder, 3835
  - effects of, 3279*t*
  - metabolism of, 3279*f*
  - pharmacogenetics of, 294*t*
- Oxford Group, 2685
- Oxidative phosphorylation, 1507
- Oxidative stress, 4474
- Oxycodone, structure of, 1400*f*
- OxyContin, 2510
- Oxytocin (OT), 119–121, **168**
  - amino acid sequence, 109*t*
  - anxiety disorders and, 1803–1804
  - receptor distribution in brain, 110*f*

P

- *Paced Auditory Serial Addition Test* (PASAT), 1032
- Pacemakers, 2290
- PAI (Personality Assessment Inventory), 1101
- Pain
  - affective regulation, 217–218
  - assessment and treatment of, 2387–2391
  - beliefs, 2490
  - in cancer, 2369–2370
  - chronic
    - common conditions, 2483*t*
    - management of, **2481–2491**
  - clinical features, 2482
  - cognitive processing and, 218
  - comparative nosology, 2480
  - coping with, 2490
  - course and prognosis, 2485–2486
  - definition of, 2480
  - diagnosis of, 2482
  - differential diagnosis, 2482–2485
  - dysregulation, 217–218
  - endocannabinoids and, 135–136
  - epidemiology of, 2481
  - etiology of, 2481–2482
  - exercise for, 2490
  - hypnosis and, 2931–2936
  - imaging studies, 2482
  - incidence of, 2387
  - management of
    - burn injuries, 2469–2470
    - chronic, **2481–2491**

- medications for, 2389
- motivational mechanism and, 216–217
- multidisciplinary rehabilitation, 2491
- multiple convergent pain management, 2933–2936, 2934*f*
- nefazodone, 3206
- neuroanatomy of, 217*f*
- neuropathic, 2480*t*
- neuropathic and nociceptive pain conditions, 2481*t*
- nociceptive, 2480*t*
- pathology of, 2482
- pathophysiology of, 2481–2482
- perception of, 215–216
- physical therapy for, 2490
- placebo response, 2490
- psychiatric comorbidities of, 2485–2486
  - affective illness, 2486
  - anger, 2486
  - anxiety disorders, 2486
  - depression, 2486
  - fear, 2486
  - somatic symptom disorder, 2485
  - substance use disorder, 2485–2486
- psychiatric symptoms and, 2389–2390
- system, 215–216
- in terminally ill children, 2406
- trazodone, 3201
- treatment of, 2487–2491
  - invasive, 2490–2491
  - pharmacologic, 2486–2489
    - adjunctive drugs, 2489
    - anticonvulsants, 2488–2489
    - antidepressants, 2488

        - antipyretic analgesics, 2487
        - novel antidepressants, 2488
        - opioid analgesics, 2487–2488
        - selective serotonin reuptake inhibitors, 2488
        - serotonin-norepinephrine reuptake inhibitors, 2488
        - tricyclic antidepressants, 2488
    - psychological, 2489–2490
    - spinal cord stimulation, 2491
    - surgery, 2491
  - types, 2388*t*
  - types of experiences, 2480*t*
  - undertreatment of, 2390
- Pain disorder, 4751. *See also* Somatic symptom and related disorders
- Pain–pleasure principle, 4751
- Pain syndromes
  - gabapentin, 3267
  - lamotrigine for, 3296
  - pregabalin, 3275
  - tiagabine, 3271
- Paired associative stimulation, 3464*t*
- Palazzoli, Mara Selvini, 2947
- Paleocortex, 17
- Paleostriatum. *See* Globus pallidus
- Palilalia, 467*t*, 3774
- Palinacousis, 476
- Palinopsia, 476
- Paliperidone (Invega), 3071–3078. *See also* Infusion
  - for bipolar disorder, 1730*t*, 1739
  - for bipolar mania, 3071–3072, 3075
  - chemistry, 3071
  - dosage and administration, 1579, 3074–3075

  - drug interactions, 3074–3075, 3076*t*
  - effects on organs and systems, 3073
  - for geriatric bipolar disorder, 4243
  - geriatric use, 3074
  - molecular structure, 3056*f*
  - pediatric use, 3074
  - pharmacodynamics, 3071–3072
  - pharmacokinetics, 3071
  - precautions and adverse reactions, 3073
  - for schizophrenia and psychosis, 3072–3073, 3075
  - therapeutic indications, 3073
- Palliative care, 2384
  - family interventions and, 4377–4378
  - in long-term care, 4414
- Pallidum. *See* Globus pallidus
- Palm Pilot, 2699–2700
- Pamelor. *See* Nortriptyline (Aventyl, Pamelor)
- Pancreas, 1056
- Pandemics, psychological reactions of physicians during, 2385–2386
- Panic, 4751
- Panic attacks, 1186, 4751
- Panic disorder, 1783–1790
  - bipolar disorder and, 1742
  - buspirone for, 3170
  - comparative nosology, 1783–1784
  - course, 1784
  - defined, 4751
  - definition of, 1783
  - diagnosis, 4796*t*
  - differential diagnosis, 1784
  - DSM-5 criteria for, 1783, 1783*t*

  - EEGs in, 375
  - epidemiology of, 1784
  - etiology of, 1784
  - history, 1783–1784
  - ICD-11 criteria for, 1223
  - monoamine oxidase inhibitors, 3195
  - neuroimaging, 1814–1815
  - prevalence of, 1791
  - psychodynamic therapy for, 2760
  - short-term dynamic psychotherapy for, 2778
  - SSRIs for, 3131
  - trazodone, 3199
  - treatment of, 1784, 1837–1838
  - valproate for, 3308
- Panic Disorder Severity Scale (PDSS), 1106–1108
- Panic-focused psychodynamic psychotherapy (PFPP), 2754, 2760, 2772–2773
- Panic-like behavior, 297
- Panksepp, Jaak, 420, 806
- Panphobia, 4751
- PANS. *See* Pediatric acute onset neuropsychiatric symptoms (PANS)
- Pansexuality, 2059
- PANSS (Positive and Negative Syndrome Scale), 1102–1104
- Pantomime, 4751
- Pantothenate kinase-associated neurodegeneration, 704–705
- PAP. *See* Positive airway pressure (PAP)
- Papageno effect, 2714
- Papez, James, 34, 460
- Papez circuit, 460
- Paradiso, Sergio, 487
- Paradoxical communication, 949

- Paradoxical insomnia, 2167
- Paradoxical intention, 2172
- Paradoxical reward effect, 764
- Paradoxical sleep, 200, 4751
- Paradoxical therapy, 2981–2983
  - clinical issues, 2982–2983
  - ethical issues, 2982–2983
  - history, 2981
  - techniques, 2982
  - theoretical issues, 2981
- Parahippocampal gyrus, 35
- Paraldehyde, 3236
- Parallel distributed processing (PDP) models, 319–322. *See also* Connectionist models
- Parallel form reliability, 1037
- Paralogia. *See* Evasion
- Paralysis agitans, 4751. *See also* Parkinsonism
- Paralytic ileus, 4751
- Parameter, 4751
- Parametric exponential family, 908
- Parametric forms, 896
- Parametric tests of significance, 4751
- Paramnesia, 1182, 4751. *See also* Confabulation
- Paraneoplastic neurologic disorders, 2365
- Paraneoplastic syndromes, 2365
- Paranoia, 4751
- Paranoiac, 4751
- Paranoid delusion, 4724, 4751
- Paranoid disorder. *See* Delusional disorder
- Paranoid ideation, 4752
- Paranoid personality disorder, 1624–1625, 2245, 4752
- Paranoid schizophrenia, 4752

- Paraphasia, 4752
- Paraphilias, 2049, 2058, 2091, 4752
- Paraphilic disorders, **2090–2101**
  - behaviorist theory, 2093
  - comorbid psychiatric disorders, 2093
  - comparative nosology, 2091–2092
  - course and prognosis, 2097
  - crossing over, 2092–2093
  - definition of, 2090–2092
  - diagnosis and clinical features, 2093–2097, 4838–4841*t*
  - diagnostic tests, 2097
  - differential diagnosis, 2097
  - DSM-5 criteria for, 2092
  - epidemiology of, 2092–2093
  - ethical and legal considerations, 2099–2100
  - etiology of, 2093
  - exclusive and nonexclusive subtypes, 2094
  - exhibitionistic disorder, 2094, 4838*t*
  - fetishistic disorder, 2094, 4841*t*
  - frequency of paraphilic behavior, 2092
  - frotteuristic disorder, 2094–2095, 4839*t*
  - gender ratio, 2092
  - history of literature on, 2091
  - pathology and laboratory examination, 2097
  - pedophilic disorder, 2095, 4840*t*
  - populations, 2092
  - psychoanalytical theory, 2093
  - sexual masochism disorder, 2095–2096, 4839*t*
  - sexual sadism disorder, 2096, 4839*t*
  - transvestic disorder, 2096–2097, 4841*t*
  - treatment of
    - behavior modification, 2137–2138

    - biologic, 2098
    - confidentiality and privilege, 2100
    - endocrinologic agents, 2138
    - models of, 2099
    - pharmacologic, 2098–2099
    - psychological, 2099
    - surgical, 2098
  - voyeuristic disorder, 2097, 4838*t*
- Paraphrenia, 4752
- Parapraxis, 916, 4752
- Paraprofessional, 4752
- Paraprofessionals, 4028
- Parapsychology, 4752
- Parasitic foci, 320
- Parasomnias, 534*t*, 2181–2188
  - benign sleep myoclonus of infancy, 2187
  - in children/adolescents, 4086–4087
  - defined, 4752
  - description of, 2182
  - due to drug to substance abuse, 2185
  - due to medical disorders, 2185
  - exploding head syndrome, 2185
  - nightmare disorder, 2183
  - non-rapid eye movement (NREM) sleep arousal disorders, 2182–2183
  - periodic limb movement disorder, 2185–2186
  - propriospinal myoclonus at sleep onset, 2187
  - rapid eye movement sleep behavior disorder, 2183
  - restless legs syndrome, 2183
  - sleep enuresis, 2185
  - sleep paralysis, 2185
  - sleep-related bruxism, 2186, 2188*f*

  - sleep-related hallucinations, 2185
  - sleep-related leg cramps, 2186
  - sleep-related movement disorder due to a medical disorder, 2187
  - sleep-related movement disorder due to medication or substance, 2187
  - sleep-related rhythmic movement disorder, 2186–2187
- Parasympathetic nervous system. *See* Autonomic nervous system (ANS)
- Parataxic distortion, 4752
- Parathyroid disorders, 714–715
  - hyperparathyroidism, 715
  - hypoparathyroidism, 715
- Parathyroid gland, 3322
- Parathyroid hormone (PTH), **166–167**, 1055
- Paraventricular nuclei, 37
- Pareidolia, 1172
- *Parens patriae*, 4608–4609, 4752
- Parental alienation, 2545
- Parental relationship distress (PRD), 2545
- Parent–child interaction therapy (PCIT), 2817
  - for selective mutism, 3865
- Parent–child problem, 3970, 4752
- Parent–child relational problem, 2542–2543
  - clinical features, 2543
  - comorbidity, 2543
  - definition of, 2542
  - diagnosis, 2543
  - differential diagnosis, 2543
  - etiology of, 2542–2543
  - parenting styles diagram, 2542*t*
- Parent interview
  - during child psychiatric examination, 3517–3523

  - child’s difficulties, 3517–3518
  - community background, 3520–3521
  - developmental history of child, 3517–3518
    - academic development, 3518–3519
    - adverse experience, 3520
    - cognitive development, 3518–3519
    - conscience, 3519
    - educational development, 3519
    - family relationships, 3519
    - interests, hobbies, talents, and avocations, 3519–3520
    - medical history, 3519
    - peer relationships, 3519
    - physical development, 3519
    - temperament, 3519
    - values, 3519
  - expectations and anxieties, 3522–3523
  - family background, 3520
  - family medical and psychiatric history, 3521
  - informant differences, 3521
  - interview techniques, 3522
  - preparation of child, 3523
  - reasons for referral, 3517–3518
  - special issues in, 3521–3523
  - in suspected physical abuse, 3973–3974
  - in suspected sexual abuse, 3973–3974
- Parent of origin effects, 248
- Parent(s)
  - of child with ADHD, 4080–4081
    - child characteristics, effects of, 4080–4081
    - clinical implications, 4081
    - parent characteristics, effects of, 4080–4081
  - of child with ASDs, 4075–4078

    - coping, 4077
    - psychological responses, 4076–4077
    - quality of life, 4077
    - unique stressors, 4075–4076
  - of child with bipolar disorder, 4078–4080
    - caregiver burden, 4079
    - interventions, 4079–4080
    - lifespan issues, 4078–4079
    - unique stressors, 4078
  - gay and lesbian, 2080–2081
  - upbringing away from, 2544
- Paresis, 4752
- Paresthesia, 2480*t*, 4752
- Pargyline (Eutonyl), 3185*t*
- Parietal lobe, 9
- Parkinson, James, 556
- Parkinson disease (PD), **555–563**. *See also* Parkinsonism
  - amantadine for, 3332
  - anxiety and, 559, 562
  - apathy and, 559, 562–563
  - course and prognosis, 561
  - definition of, 555
  - dementia in, 1269–1271
  - depression and, 559
  - diagnosis
    - cognitive manifestations, 559
    - differential, 560–561
    - laboratory examination, 560
    - motor manifestations, 558–559
    - psychiatric manifestations, 559–560
  - epidemiology of, 556
  - etiology of, 556–558

    - cognitive manifestations, 557–558
    - motor manifestations, 557
    - psychiatric manifestations, 558
  - historical literature on, 556
  - impulse control disorders and, 559–560
  - in medically ill geriatric patients, 4202–4203
  - neurodegenerative disorders due to, 4174–4175
  - with psychosis, 4308
  - psychosis and, 559–560, 563
  - treatment of, 561–563
    - pharmacologic, 561
    - surgical, 561–563
- Parkinson disease dementia, 4323–4324
- Parkinson disease psychosis, pimavanserin for, 3216–3217
- Parkinsonism, 4752. *See also* Parkinson disease (PD)
  - amantadine for, 3332
  - anticholinergics and, 3328
  - drug-induced, 3046, 3092
- Parkinson psychosis, dopamine receptor antagonists for, 3045
- Parnate. *See* Tranylcypromine
- Paroxetine (Paxil), 3137
  - chemistry, 3137
  - dosing and administration, 3141
  - history, 4663
  - pharmacogenetics of, 295*t*
  - pharmacokinetics of, sex differences in, 3416
  - pharmacologic actions, 3137
  - for social anxiety disorder, 3859
  - for tic disorders, 3785–3786
- Paroxysmal disorders, 465
- Paroxysmal extreme pain disorder (PEPD), 81
- Paroxysmal movement disorder, 534*t*

- Paroxysmal photosensitive responses (PRRs), 371
- PARS. *See* Pediatric Anxiety Rating Scale (PARS)
- Pars compacta, 28
- Pars reticulata, 28
- Partial hospitalization, 4752
- Partial hospital programs (PHPs), 3939
- Partial insanity. *See* Monomania
- Partial malingering, 2504
- Partial object relations, 814–816
  - neuroscience of, 815–816
- Partial regression, 2917
- Partial reinforcement extinction effect (PREE), 764, 2224
- Participant modeling, 1828
- PASAT. *See Paced Auditory Serial Addition Test* (PASAT)
- Passion flower (*Passiflora incarnata*), 2640
- Passive aggression, 2748*t*
- Passive–aggressive behavior, 920*t*
- Passive–aggressive personality disorder, 4752
- Passive-dependent personality. *See* Dependent personality disorder
- Passive euthanasia, 2409
- Passive membrane properties, 69
- Passive social withdrawal, 1459
- Pastoral counseling, 4752
- Patch amnesia. *See* Localized amnesia
- Patch clamp, 69
- Paternal transference, 2749
- Pathogen-associated molecular patterns (PAMPs), 172
- Pathognomonic, 4752
- Pathological gambling, **1856–1868**. *See also* Gambling disorder
  - amantadine for, 3333
  - clinical features, 1859–1861

  - clinical vignettes, 1867–1868
  - comorbidity, 1861–1862
  - diagnosis, 1858–1859
  - differential diagnosis, 1861–1862
  - epidemiology of, 1859
  - etiology of, 1862–1863
  - family history, 1859
  - future for gamblers, 1867
  - imaging studies, 1863–1864
  - naltrexone for, 3363
  - neurotransmitters and, 1864
  - nosological considerations, 1858
  - personality characteristics, 1861
  - predisposing materialist factors, 1864
  - psychoanalytic theories, 1863
  - treatment of, 1865–1866
    - psychopharmacologic, 1865–1866
    - psychotherapy, 1865
    - self-help interventions, 1866
- Pathological Institute, 4479
- Pathological intoxication. *See* Alcohol idiosyncratic intoxication
- Pathologic dissociation, 1945–1946
- Pathologic gambling, 4752
- Pathologic intoxication. *See* Alcohol idiosyncratic intoxication
- Pathologic laughter and crying (PLAC), 475, 486, 494*f*, 548
- Pathologic moods, 1671
- Pathologic possession trance, 2559–2560
- Pathophobia, 4752
- Pathoplastic model of temperament/psychopathology, 3624*t*
- Pathway analysis, 250
- Patient-centered care, 854
- Patient-centered medical home (PCMH), 4498–4499

- Patient-centered outcomes research (PCOR), 4500
- Patient-doctor relationship, 1156
- Patient Health Questionnaire (PHQ), 1100
- Patient Health Questionnaire-9 (PHQ-9), 1104–1105, 1105*t*, 1720, 3543
- Patient Protection and Affordable Care Act, 4400–4401
- Patient-Reported Outcomes Measurement Information System (PROMIS) Emotional Distress—Anxiety Short-Form, 1106
- Paucity of speech, 4752
- Pavlov, Ivan, 4752
  - classical conditioning, 756
  - on hypnosis, 2912
  - learning theory, 434
  - work of, 4689
- Pavlovian conditioning, 756, 758–761
  - and behavior, 758–759
  - defined, 4752
  - erasing of, 760–761
  - learning process, 759–760
- Pavlovian-instrumental transfer, 758–759
- *Pavor nocturnus*, 2182, 4752
- Paxil. *See* Paroxetine (Paxil)
- Pay for performance (P4P), 4500–4501
- PBA. *See* Pseudobulbar affect (PBA)
- PCL-5 (PTSD Checklist for DSM-5), 1108, 1109*t*
- PCP. *See* Phencyclidine (PCP)
- PD. *See* Parkinson disease (PD)
- PDMR. *See* Progressive deep muscle relaxation (PDMR)
- PDQ (Personality Disorder Questionnaire), 1115
- PDSS (Panic Disorder Severity Scale), 1106–1108
- PDUFA. *See* Prescription Drug User Fee Act (PDUFA)
- Peace, 2512

- Peccatophobia, 4752
- Pecking order, 4752
- Pederasty, 2072, 4752
- Pediatric acute onset neuropsychiatric symptoms (PANS), 626
- Pediatric Anxiety Rating Scale (PARS), 3545
- Pediatric autoimmune neuropsychiatric syndromes associated with streptococcal infections (PANDAS), 189
  - neuroimaging, 3611–3612
  - tic disorders and, 3777
- Pediatric evaluation, 3551
- Pediatric pharmacology, **3437–3438**
- Pediatric psychiatry, history of, 4664–4665
- Pediatric Research Equity Act, 3022
- Pedigree analysis, 255
- Pedophilia, 4752
- Pedophilic disorder, 2095, 4840*t*
- Pedophobia, 4752
- Pedunculopontine tegmental (PPT) nuclei, 201, 202*f*, 208
- Peer group, 3628
- Peer review, 4752
- Peer review organization (PRO), 1159, 4752
- Peer support specialists (PSS), 1612–1613
- Pekala, Ronald, 2914
- Pelizaeus-Merzbacher disease (PMD), 698–699, 699*f*
- Pellagra, 4753
- Pelvic inflammatory disease (PID), 2595
- Pelvic pain, 2595–2596
- Pemoline (Cylert), 4002*t*
- Penance, 2674
- Penetration disorder, 4733
- Penfield, Wilder, 369
- Penicillin, 90

- Penile self-injection, alprostadil, intraurethral, 3434–3435
- Penis, 2028
- Penis envy, 4753
- Penn Emotion Recognition Test (ER-40), 4551
- Penn Resiliency Program (PRP), 2898
- Pentobarbital (Nembutal)
  - dose, 1426*t*
  - $GABA_A$ receptor and, 89
- Peptidases, 115
- Peptide tyrosine tyrosine (PYY), 225
- Perampanel (Ficompa), 3276–3277
  - chemistry, 3276
  - clinical drug studies, 3276
  - dosage and administration, 3277
  - drug interactions, 3277
  - history, 3276
  - pharmacodynamics, 3276
  - pharmacokinetics, 3276–3277
  - precautions and adverse reactions, 3276–3277
  - therapeutic indications, 3276
- Perception
  - abnormalities, 996
  - defined, 4753
  - definition of, 725
  - disturbances of, 1172–1175
    - body image distortion, 1175
    - hallucinations, 1172–1175
    - illusions, 1172
  - fractionated elements of, 1087
  - in older adults, 4134
  - in psychiatric disorders, 734–743
  - psychiatric examination and, 996

  - sensation and, 729
- Perceptual constancy, 3502
- Perceptual decoupling, 423
- Perceptual priming, 774
- Percival, Thomas, 4617
- Percutaneous endoscopic gastrostomy (PEG), 2369
- Percutaneous endoscopic jejunostomy (PEJ), 2369
- Peregrination, 1907
- Perfectionism, 4753
- Perforant path, 35
- Performance anxiety, β-adrenergic receptor antagonists for, 3227
- Performance-Based Skills Assessment, 4552
- Performance validity tests (PVTs), 1033–1034
- Pergolide (Permax), 3107
- Periactin. *See* Cyproheptadine (Periactin)
- Periaqueductal gray, 40–41
- Pericentric/constitutive heterochromatin, 273–274
- Perimenopausal depression, estradiol for, 3426–3428
- Perimenopause, pharmacokinetics/pharmacodynamics and, 3414
- Perinatal anxiety, 2600
- Perinatal psychiatric disorders, 2597–2608
  - anxiety disorders, 2600
  - bipolar disorder, 2599
  - breastfeeding and, 2601–2602
  - nonaffective psychosis, 2599
  - obsessive-compulsive disorder, 2600
  - pathogenesis of, 2600–2602
  - perinatal depression, 2597–2598
  - postpartum blues, 2597
  - postpartum psychosis, 2598–2599
  - screening for, 2602
  - treatment of, 2602–2608

    - nonsomatic, 2603–2604
    - psychopharmacology, 2604
- Periodic lateralizing epileptiform discharges, 524, 525*f*
- Periodic limb movement disorder (PLMD), 2185–2186
  - in older adults, 4214
  - treatment of, 4214
- Periodic movements of sleep, 3780*t*
- Period prevalence, 4753
- Peripartum depression, 4752
- Peripheral nervous system (PNS)
  - structure of, 653, 653*f*
  - toxic agents, impact of, 655*t*
- Peripheral neuropathic pain, 2483
- Peripheral neuropathies
  - alcohol and, 1299
  - chronic inflammatory demyelinating polyneuropathy, 657
  - cranial nerve disorders, 657–658
  - diabetic neuropathy, 657
  - diagnosis of, 656
  - differential diagnosis, 481t
  - frostbite, 658
  - Guillain–Barré syndrome, 657
  - herpes zoster neuropathy, 660
  - HIV-associated, 608, 659–660
  - infectious and granulomatous neuropathies, 659–661
  - leprous nephropathy, 660
  - lyme nephropathy, 660
  - metabolic, endocrine neuropathies, and ischemic neuropathies, 658
  - nutritional and alcoholic neuropathies, 657
  - peripheral nerve injuries, 658
  - sarcoid neuropathy, 661

  - sensory or sensorimotor, 656
  - treatment of, 656–657
- Peritraumatic Dissociative Experiences Questionnaire (PDEQ), 1937*t*
- Perls, Fritz, 2957
- Permanent pacemakers (PPMs), 2290
- Permax. *See* Pergolide (Permax)
- Permissive hypnotherapy, 2922
- Peroxisomal disorders, 692*f*
- Perphenazine (Trilafon), 3008, 3034
  - pharmacogenetics of, 294*t*
- Perris, Carlo, 1652
- Persecution, 996*t*
- Perseveration, 4753
- Persistence, 2224, 2224*t*, 4753
- Persistent (chronic) motor or vocal tic disorder, 4753
- Persistent complex bereavement disorder, 4753
- Persistent depressive disorder, 1687–1689, 4793*t*
- Persistent genital arousal disorder, 2048
- Persisting dementias, 1305
- Persona, 4753
- Personal Genome Project, 250*t*
- *Personal Happiness Program*, 2897
- Personal health records (PHRs), 1122
- Personal inventories, 4659
- Personality, 801
  - anxiety disorders and, 1794
  - defined, 4753
  - definition of, 2217–2218
  - Five-Factor Model of, 2254
  - gene types, 2230*f*
  - integrated view of, 2234–2235

    - over the life span, 4160
- Personality Assessment Inventory (PAI), 1043, 1101
- Personality assessments, 982, **1035–1048**
    - adult psychological tests, 1038–1048
        - behavioral assessment, 1047
        - Brief Psychiatric Rating Scale, 1047
        - Millon Clinical Multiaxial Inventory-III, 1042
        - Minnesota Multiphasic Personality Inventory-2, 1038–1041
        - Minnesota Multiphasic Personality Inventory-3 (MMPI-3), 1042
        - MMPI-2-Restructured Form, 1041–1042
        - objective personality tests, 1039*t*
        - object personality tests, 1038–1043
        - Personality Assessment Inventory, 1043
        - Personality Inventory for DSM-5-Adult, 1042
        - Personality Inventory-Revised, 1043
        - projective personality tests, 1043–1047, 1045*t*
        - rapid personality assessment instruments, 1043, 1044*t*
        - Rorschach test, 1044–1045, 1045*f*, 1046*t*
        - sentence completion test, 1047
        - Thematic Apperception Test, 1045–1047
    - in clinical practice, 2217, 2235–2237
    - in medicine, 2216–2217
    - psychological testing, purposes of, 1035–1036
    - psychometric properties of instruments
        - normative sample, 1036
        - reliability, 1037–1038
        - test characteristics, 1036
        - validity, 1036–1037
- Personality Assessment Test (PAI), 2505
- Personality change due to another medical condition, 2251
- Personality changes, 465
    - in brain tumor patients, 514–517

  - dementia and, 1245
  - multiple sclerosis and, 580
  - psychotherapy and, 984
- Personality development
  - creativity and, 2213–2214
  - fluidity and, 2213–2214
  - self-organizing biopsychosocial complex, 2238–2243
  - self-transcendence and, 2214–2216
  - stage theories of, 4160
- Personality Diagnostic Questionnaire (PDQ R), 2254
- Personality disorder, 4753
- Personality Disorder Examination (PDE), 2254
- Personality Disorder Questionnaire (PDQ), 1115
- Personality disorders, 1192–1193, 1206, 1207, **2213–2270**
  - alloplastic nature of, 2252
  - borderline core of, 2258
  - in cancer patients, 2359–2360
  - categorical approach to, 2252
  - character and, 2225–2231
  - classification of, 2243–2244
    - DSM-5, 2243–2244
    - ICD-11, 2243
  - clinical and psychometric issues, 2252–2257
  - cluster A disorders, 2245–2246
    - paranoid personality disorder, 2245
    - schizoid personality disorder, 2245–2246
    - schizotypal personality disorder, 2246
  - cluster B disorders, 2246–2249
    - antisocial, narcissistic, and histrionic spectrum, 2248
    - antisocial personality disorder, 2246–2247
    - borderline personality disorder, 2248–2249
    - histrionic personality disorder, 2248

  - narcissistic personality disorder, 2247–2248
- cluster C disorders, 2249–2251
  - avoidant personality disorder, 2249–2250
  - dependent personality disorder, 2250
  - obsessive-compulsive personality disorder, 2250–2251
- cognitive behavioral therapy for, 2825–2826
- definitions of, 4271–4272
- diagnosis of, 2256–2257, 4833–4838*t*
- differential diagnosis, 1697
- dimensional approach to, 2253
- dimensional assessment of, 2254–2255
- DSM-5, disadvantages of, 2251–2252
- DSM-5 criteria for, 4272
- DSM-5 revisions, 980–981, 981*t*
- egosyntonic nature of, 2252
- epilepsy and, 526
- Five-Factor Model descriptions of, 980, 981*t*
- general, 4833–4834*t*
- group psychotherapy for, 2965
- hypnosis and, 2930–2931
- ICD-11 criteria for, 1224
- identity and, 2225–2231-22
- individual, diagnostic guidelines, 2244
- measurement of
  - categorical tests, 2254
  - interviews, 2253
  - self-reports, 2253
  - state-trait effect, 2253–2254
  - temperament and character inventory, 2255, 2257*t*
- mood disorders and, 1697*t*
- motivation and, 2224–2225
- normal *versus* deviant personality, 2243

- in older adults, **4271–4276**
  - clinical features and course, 4273–4275
  - cluster A disorders, 4274
  - cluster B disorders, 4274–4275
  - cluster C disorders, 4275
  - diagnosis, 4272–4273
  - epidemiology of, 4272–4273
  - pharmacotherapy for, 4276
  - psychotherapy for, 4276
  - treatment of, 4275–4277
- other specified personality disorder, 2251
- pathogenesis of, 2257–2260
- pathology of, 2263
- personality change due to another medical condition, 2251
- personality traits and, 980
- predictions of from self-directedness scores, 2257*t*
- psyche and, 2225–2231-22
- psychobiological summary of, 2259–2260
- psychodynamic model of, 2257–2259
- psychosis and, 1624–1626
- quantifiable features of, 2251*t*
- rating scales for, 1114–1115
- secondary symptoms of, 2263
- short-term dynamic psychotherapy for, 2779
- social or clinical diagnosis, 2252
- spectrum disorder hypothesis, 2257
- subtypes of, 2253
- suicide and, 2709–2710
- symptoms and exclusions, 4835–4838*t*
- temperament and, 2218–2224
- transference-focused psychotherapy for, 2930
- treatment of, 2260–2270

    - general approach to, 2264–2265
    - general principles, 2261–2263
    - integrated approach, 2268–2270, 2269*t*
    - medications, 2265–2268, 2266*t*
    - preventable errors in, 2263–2264
    - stages in, 2264
    - and stages in the development of self-awareness and well-being, 2260–2261, 2268*t*
  - unspecified personality disorder, 2251
  - vulnerability to, 871
- Personality Inventory for Children (PIC), 3543, 3548
- Personality Inventory for DSM-5 (PID), 1115
- Personality Inventory for DSM-5-Adult, 1042
- Personality Inventory for DSM-5 Brief Form (PID-5-BF), 1115
- Personality Inventory-Revised, 1043
- Personality measures, 3548–3549
- Personality structure, 951–952, 975–976
- Personality theories, 950
  - Adler, Alfred, 948
  - Alexander, Franz, 953
  - Allport, Gordon, 972
  - application to therapies, 971
  - archetypes, 950–951
  - Berne, Eric, 962–963
  - Chodorow, Nancy, 964–966
  - complexes, 950
  - Frankl, Viktor, 973
  - Freud, Sigmund, 422
  - Fromm, Erich, 958
  - Horney, Karen, 956
  - Jung, Carl, 950
  - Kelly, George, 972–973

  - Maslow, Abraham, 972
  - McAdams, Dan, 973
  - personality structure, 951–952
    - individuation, 951
    - psychological types, 951
    - self, 951
  - radical behaviorism, 970
  - Reich, Wilhelm, 955
  - Rogers, Carl, 973
  - social learning theory, 970–971
  - Sullivan, Harry Stack, 959–960
  - therapeutic process, 952
  - trait and factor models, 974–984
  - unconscious, 951
- Personality traits, 1192–1193
  - characteristics of, 975
  - defined, 4753
  - definition of, 2237
  - developmental trait personality models, 2238
  - origin and development of, 976–978
  - personality disorders and, 980
  - person *versus* situation debate, 2237–2238
  - phenotypic trait personality models, 2238
  - profiles, 982
  - rating scales for, 1114–1115
  - stress and, 2421–2422
- Personalized medicine, 3020
- Personalized technology, 3020
- Personal mastery, 4473
- Personal Responsibility and Work Opportunity Reconciliation Act, 3960
- Person-centered approaches, 4542

- Persuasion, 4753
- Pertofrane. *See* Desipramine
- Perugi, Giulio, 1660
- Pervasive developmental disorder
  - defined, 4753
  - dopamine receptor antagonists for, 3045
  - not otherwise specified, 4753
- Perversions, 2091, 4753
- Perverted logic. *See* Evasion
- PET. *See* Positron emission tomography (PET)
- Petit mal epilepsy, 4753
- Petrarch, 4683
- Petty, Frederick, 1657
- PGC. *See* Psychiatric Genomics Consortium (PGC)
- Phallic-narcissistic character, 955
- Phallic overbearance, 4753
- Phallic phase, 4753
- Phallic stage, 916, 918–919*t*
- Phantasy, 812–813, 2787, 2790–2792, 4753
- Phantom limb, 4753
- Phantom limb pain, 2485
- Pharmacodynamics, 3010. *See also specific drugs*
  - pharmacogenetics of, 292
  - pharmacokinetics and, 286
  - reproductive periods and, 3413–3414
  - sex differences in, 3411–3412
- Pharmacodynamic tolerance, 1298
- Pharmacogenetics, 159, **286–296**
  - CYP2C19, 287
  - CYP2D6, 287–288
  - CYP3A4, 287
  - definition of, 3586, 3587*t*

  - drug entry into the brain, 290
  - of pharmacodynamics, 292
  - pharmacogenomics and, 286
  - of pharmacokinetics, 286–287
  - testing, 292–295
- Pharmacogenomics, 159
  - in child psychiatry, 3591–3592
  - definition of, 3586, 3587*t*
  - in neuropsychiatric disorders, 250
  - pharmacogenetics and, 286
- Pharmacokinetics, 3010–3011. *See also specific drugs*
  - alcoholism and, 290
  - codeine, 289
  - defined, 4753
  - drug entry into the brain, 290
  - ethnic variability, 291–292
  - geographic variability, 291–293
  - lisdexamphetamine, 289
  - multiple active compounds, 289–290
  - pediatric, 3925–3926
    - absorption, 3926
    - drug distribution, 3926
    - metabolism, 3926
  - pharmacodynamics and, 286
  - pharmacogenetics of, 286–287
  - reproductive periods and, 3413–3414
  - sex differences in, 3411–3412
    - anticonvulsants, 3417
    - antidepressants, 3414–3416
    - antiepileptics, 3417
    - antipsychotics, 3417
    - benzodiazepines, 3417

    - noradrenaline and dopamine reuptake inhibitors, 3416
    - selective serotonin reuptake inhibitors, 3416
    - serotonin-2 antagonists and reuptake inhibitors, 3416
    - serotonin-norepinephrine reuptake inhibitors, 3416
    - tetracyclic antidepressants, 3416
    - tricyclic antidepressants, 3416
  - tramadol, 289
- Pharmacokinetic tolerance, 1298
- Pharmacologic bridge, 1656
- Pharmacotherapy. *See also* Psychopharmacology
  - addictive process, 2139–2140
  - aggression, 3737–3738
    - antipsychotics, 3650
    - beta-adrenergic antagonists, 3650
    - mood stabilizers, 3650
    - selective serotonin reuptake inhibitors, 3650
  - Alzheimer disease (AD), 4250–4252, 4251*t*
  - anxiety
    - antidepressants, 3650
    - anxiolytics, 3651
    - buspirone, 3651
    - mood stabilizers, 3650
  - anxiety disorders, 3927–3928, 4224–4225
  - autism spectrum disorders (ASDs), 3933
  - body dysmorphic disorder (BDD), 1848
  - child/adolescent, 3437
  - circadian rhythms and, 198–199
  - depersonalization/derealization disorder, 2017–2018
  - early-onset bipolar disorder, 3834–3835
  - enuresis, 3794–3796, 3934
  - geriatric bipolar disorder, 4242–4243
  - geriatric depression, 4237–4239

  - insomnia, 3933–3934
  - intellectual disabilities (ID), 3649–3652
  - for intermittent explosive disorder (IED), 2195–2196
  - kleptomania, 2199–2200
  - major depressive disorder (unipolar depression), 3109–3110, 3928–3929
  - obsessive-compulsive disorders (OCDs) in, 1844–1845, 3843–3844, 3930
  - for personality disorders in older adults, 4276–4277
  - post-traumatic stress disorder (PTSD), 1882–1883
  - schizophrenia, 1607, 3931
  - self-injurious behaviors (SIBs), 3362, 3651, 3813
    - antipsychotics, 3651
    - mood stabilizers, 3651
    - NMDA antagonists, 3651
    - opioid antagonists, 3651
    - selective serotonin reuptake inhibitors, 3651
  - sexual dysfunction(s), 2052–2055
  - sociocultural factors, 856
  - somatic symptom disorder, 1897
  - suicidal behavior, 3928
  - suicide, 2712–2713
  - of tic disorders, 3786*t*
  - tic disorders, 3784–3787, 3934
- Phase-amplitude coupling (PAC), 312
- Phase of life problem, 4753
- Phase reversal, 360
- Phencyclidine (PCP)
  - defined, 4753
  - effects of, 1351–1352
  - historical overview, 1348–1349
  - structure of, 84*f*

- Phencyclidine- and other hallucinogen-induced anxiety disorder, 1364
- Phencyclidine- and other hallucinogen-induced bipolar disorder, 1363
- Phencyclidine- and other hallucinogen-induced depressive disorders, 1363
- Phencyclidine- and other hallucinogen-induced intoxication delirium, 1364
- Phencyclidine- and other hallucinogen-induced psychotic disorder, 1362–1363
- Phencyclidine and other hallucinogen intoxication, 1356–1359, 2728
  - course and prognosis, 1358
  - definition of, 1356
  - diagnosis and clinical features, 1357
  - differential diagnosis, 1358
  - epidemiology of, 1356–1357
  - etiology of, 1357
  - laboratory examination, 1358
  - pathology of, 1358
  - treatment of, 1358–1359
- Phenelzine (Nardil), 3185*t*, 3188
  - adverse effects, 3188
  - lactation and, 3188
  - for major depressive disorder, 1724
  - monitoring parameters, 3188
  - pharmacodynamics, 3188
  - pharmacokinetics, 3188
  - pregnancy and, 3188
- Phenergan. *See* Promethazine (Phenergan)
- Phenethylamines, 430
- Phenobarbital (Luminal)
  - for barbiturate detoxification, 1431–1432

  - $GABA_A$ receptor and, 89
- Phenocopies, 252
- Phenomenology, 4753
- Phenothiazine, 4753
- Phenothiazines, 161, 3035
- Phenotypes, 257–259, 864–865
  - animal models, 259
  - categorical, 257–258, 257*f*
  - continuous, 257*f*, 258–259
  - defined, 4753
- Phenotypic trait personality models, 2238
- Phentermine, for weight loss, 2314–2315
- Phentolamine, 2052
- Phenylethanolamine-N-methyltransferase (PNMT), 95–96
- Phenylethylamine, 107
- Phenylketonuria (PKU), 96, 691–692, 692*f*, 3642*t*, 4753
- Phenylpyruvic oligophrenia. *See* Phenylketonuria (PKU)
- Phenytoin, pharmacogenetics of, 294*t*
- Pheromones, 2055, 4754
- Philosophy
  - Aristotle, 968–969
  - Derrida, Jacques, 969–970
  - Gergen, Thomas, 970
  - Kant, Immanuel, 969
  - Plato, 968
  - postmodernism, 969–970
  - Schopenhauer, Arthur, 969
  - Shankara and Eastern philosophy, 969
  - Tzasz, Thomas, 970
- Phinney, Jean, 848
- Phipps, Henry, 4652
- Phipps Clinic, 4652

- Phobias
  - cognitive behavioral therapy for, 1827–1828
  - defined, 4754
  - definition of, 1187
  - diagnosis, 4795*t*
  - family therapy for, 3911*t*
  - hypnosis and, 2929
  - ICD-11 criteria for, 1223
  - neuroimaging, 1815–1816
  - prevalence of, 1792
  - psychotherapy for, 2759
  - specific, 1187*t*
    - clinical features, 1786
    - comparative nosology, 1786
    - course, 1786–1787
    - criteria for, 1786*t*
    - definition, 1786
    - diagnosis, 1786
    - differential diagnosis, 1786–1787
    - epidemiology of, 1786*t*
    - etiology of, 1786–1787
    - history, 1786
    - treatment of, 1786–1787
  - treatment of, 1837–1838
- Phobic disorder, 4754
- Phobic neurosis. *See* Phobias
- Phoneme, 3656*t*
- Phonemic awareness, 3656*t*
- Phonics, 3656*t*
- Phonological disorder. *See* Developmental articulation disorder
- Phonologic awareness, 3656*t*
- Phonologic coding, 3656*t*

- Phonologic disorder. *See* Developmental articulation disorder
- Phonophobia, 4754
- Phosphatases, 151
- Phosphatidyl choline, 2636
- Phosphatidylinositol, 148–149, 148*f*
- Phosphatidyl serine, 2636
- Phosphodiesterase-5 (PDE-5), 3433
- Phospholipases, 3004, 3004*f*
- Phosphorylation, 236
- Photic-driving, 370, 371*f*
- Photomyoclonic response (PMR), 371
- Photophobia, 4754
- PHQ (Patient Health Questionnaire), 1100
- PHQ-9 (Patient Health Questionnaire-9), 1104–1105, 1105*t*, 2700
- Phrenology, 4754
- Phylogenetic, 4754
- Phylogenetic cell death, 60
- Phylogeny of behavior. *See* Evolutionary biology
- Physical abuse
  - child, 3971
    - evaluation of, 3976
    - as maltreatment, 3968–3969
    - parent interviews, 3973–3974
    - physical examination, 3975
  - elder abuse, 4437
  - parent interviews in suspected, 3973–3974
- Physical activity, immune-brain interactions and, 191
- Physical assessment, 998
- Physical dependence. *See* Drug dependence
- Physical development
  - in adolescents, 3509–3510
  - in children, 3496–3497*t*

- Physical disability, everyday functioning and, 4181
- Physical disorder, 4754
- Physical examination, general. *See also* Medical assessment and laboratory testing
  - eyes, 466
  - general appearance, 466
  - heart and vessels, 466
  - mouth, 466
  - skin, 466
  - vital signs, 466
- Physical health monitoring, 1049–1050
- Physical illnesses, 1163
- Physical therapy
  - for conversion disorder, 1901
  - for developmental coordination disorder, 3680
  - for pain, 2490
- Physical violence, **2623–2627**
  - domestic violence, 2624–2625
  - hospitalization, 2626
  - initial evaluation, 2626
  - intimate partner violence, 2624–2625
  - legal issues, 2626
  - psychopharmacologic treatment, 2626
  - rape, 2624–2625
  - safety assessment, 2626
  - sequelae of, 2625–2626
  - treatment issues, 2626
  - violent crime, 2623
  - workplace violence, 2625
- Physicians
  - countertransference issues, 2692
  - doctor-patient communication, 2386–2387

  - early career, 2696
  - emotional distress, 2687–2688
  - impairment, 2692
  - late career, 2697
  - licensure, 2692
  - medical boards, 2692
  - mid career, 2697
  - moral burden of, 2411
  - moral injury/distress, 2688
  - psychological reactions of, 2385
    - during epidemic or pandemic, 2385–2386
  - rating sites, 2702–2703
  - risk factors in practice, 2691–2692
- *Physicians' Desk Reference* (PDR), 3009
- Physiologic tremors, 1195
- Physiologic voice therapy, 3709
- Piaget, Jean, **743–755**, 3499, 4693, 4754
  - biography of, 743–744
  - clinical method of, 744
  - constructive structuralism, 744
  - developmentally based psychotherapy, 754–755
  - egocentrism, 749
  - genetic epistemology, 744
  - models of intelligence, 749–753
    - developmental levels, 753–754
    - emotional basis of, 751–752
    - extensions of theory, 749–750
    - post-Piaget era, 750–751
  - *The Moral Judgment of the Child*, 750
  - *Play, Dreams, and Imitation in Childhood*, 746
  - theory of cognitive development
    - assimilation and accommodation, 745

    - equilibration, 744–745
    - stages of, 745–749, 746*t*
      - concrete operations, 747
      - formal operations, 748–749
      - preoperational subperiod and semiotic function, 746
      - sensorimotor period, 745–746
    - structuralism, 745
  - work of, 754
- Piaroa people, 2512
- Piblokto, 4754
- Pica, 2148, 3771–3773
  - clinical features, 3771–3772
  - course, 3772
  - defined, 4754
  - diagnosis and clinical features, 2154
  - diagnostic criteria, 2149, 4807*t*
  - differential diagnosis, 2158, 3772
  - epidemiology, 3772
  - epidemiology of, 2151
  - etiology, 3772
  - treatment, 3772
  - treatment of, 2164
- Pick, Arnold, 4658
- Pick disease, 4175, 4754
- Pickwickian syndrome, 4754
- Picrotoxin, 89
- PID (Personality Inventory for DSM-5), 1115
- PID-5-BF (Personality Inventory for DSM-5 Brief Form), 1115
- Pimavanserin, 3215–3218
  - absorption, 3215
  - chemistry, 3215
  - clinical drug studies, 3216–3218, 3216*t*, 3217*t*

  - for dementia, 1251
  - distribution, 3215
  - dosage and administration, 3218
  - drug interactions, 3218
  - elimination, 3215
  - history, 3215
  - mechanism of action, 3216
  - metabolism, 3215
  - molecular structure, 3215*f*
  - pharmacodynamics, 3216
  - pharmacologic actions, 3215–3216
  - precautions and adverse reactions, 3217–3218
- Pindolol (Visken)
  - absorption, 3225
  - for anxiety, 3227
  - chemistry, 3224–3225
  - distribution, 3225
  - dosage and administration, 3228
  - drug interactions, 3228
  - drug studies, 3226
  - elimination, 3225
  - history, 3224
  - laboratory interferences, 3228
  - for major depressive disorder, 3227
  - mechanism of action, 3225–3226
  - organ/system effects of, 3226
  - pharmacodynamics, 3225–3226
  - pharmacokinetics, 3225, 3225*t*
  - pharmacologic actions, 3225
  - precautions and adverse reactions, 3228
  - side effects and toxicity, 3226*t*
  - for substance use disorder, 3227

  - teratogenic effects, 3226–3227
  - therapeutic indications, 3227–3228
- Pinel, Philippe, 1468, 1652, 4643, 4684
- Pinel, Phillipe, 4754
- Pioglitazone, for bipolar disorder, 188
- Piperazine, 4754
- Piperidine, 4754
- Pirindole (Pirazidol), 3185*t*
- Pithiatism, 1928, 1929
- Pitolisant (Wakix), 107
- Pituitary adenoma, 711, 712*f*
- Pituitary disorders, 710
- Pituitary gland, 710, 3607–3608
- Pituitary tumors, 711–712
- Pivotal response treatment (PRT), 3719
- PKU. *See* Phenylketonuria (PKU)
- PLA2G6-associated neurodegeneration (PLAN), 705
- Placebo, 2491, 4754
- Placebo effect, 4754
- Placidyl. *See* Ethchlorvynol (Placidyl)
- *Planned Parenthood of Southeastern Pennsylvania v. Caseyi* (1992), 2591
- Planum temporale, 21
- Plaques, 4754
- Plasma level. *See* Blood level
- *Plasmodium*, 626
- Plasticity, 86–87, 778–779, 864, 3498, 4754
- Plasticity of plasticity, 81
- Plateau phase, 4754
- Plater, Felix, 4684
- Plato, 434, 968, 2060, 2271, 4683
- *Play, Dreams, and Imitation in Childhood* (Piaget), 746

- Play techniques, in child interview, 3524–3525
- Play therapy, 4754
- Pleasure principle, 916, 4754
- Pleiotropy, 161, 3587*t*
- Plenz, Dietmar, 804
- Pleonexia, 4754
- PLMD. *See* Periodic limb movement disorder (PLMD)
- P-loop, 71
- Pluralism, 2740–2741, 4754
- PMDD. *See* Premenstrual dysphoric disorder (PMDD)
- PNS. *See* Peripheral nervous system (PNS)
- Point estimators, 897
- Point prevalence, 4754
- Polarization transfer, 346, 348
- Police
  - criminalization and, 4574
  - mental health and, 4574
  - training of, 4577
- Police power. *See Parens patriae*
- Poliomyelitis, 654–655
- Pollan, Michael, 3008
- Polycomb-repressed chromatin, 278
- Polycystic ovary syndrome (PCOS), 2343–2344
  - clinical and laboratory examination, 2343
  - clinical features, 2343
  - definition of, 2343
  - differential diagnosis, 2343–2344
  - etiology of, 2343
  - psychiatric illness and, 2344
  - treatment of, 2344
- Polydipsia, 3322
- Polydipsia–hyponatremia syndrome, 3068

- Polygenetic threshold model, 1490
- Polygenic, 4754
- Polygenic inheritance, 252
- Polygenicity, 3587*t*
- Polygenic risk scores (PRS), 3589
- Polymorphisms, 242, 3586
- Polymorphous perverse sexuality, 4754
- Polyphagia, 4754
- Polysomnography, 371, 477, 1062, 2188–2189, 4083, 4754
- Polysubstance dependence, 4754
- Polyuria, 3322, 4754
- Pons, 42, 207
- Ponto-geniculo-occipital (PGO) waves, 208
- Popular illness category, 2551
- Population, 4754
- Population attributable fraction, 877
- Porphyria, 4754
- Porta, Giambattista, 4684
- Porter, Robert Langley, 4654
- Positive airway pressure (PAP), 2177–2178
- Positive allosteric modulator (PAM), 115
- Positive and Negative Syndrome Scale (PANSS), 1102–1104
- Positive behavioral support, 2898–2899
- Positive emotions, 784, 797–798
- Positive euthanasia, 2409
- Positive feecback, 2068
- Positive mental health, 2887–2889
  - biomarkers of, 4474
    - allostatic load, 4474
    - genetics and epigenetics, 4474
    - inflammation and immune function, 4474
    - oxidative stress, 4474

    - telomere length, 4474
- Positive predictive value, 881
- Positive psychiatry, **4471–4476**
  - of aging, **4471–4475**
  - biology of, 4474–4475
  - biomarkers of positive mental health, 4474
    - allostatic load, 4474
    - genetics and epigenetics, 4474
    - inflammation and immune function, 4474
    - oxidative stress, 4474
    - telomere length, 4474
  - definition of, 4471
  - history, 4471–4472
  - interventions, 4475–4476
    - biologic treatments, 4476
    - diet and nutrition, 4475–4476
    - lifestyle, 4475
    - physical activity, 4475
  - low perceived stress, 4472
  - neuroplasticity, 4474–4475
  - positive environmental factors, 4474
    - family dynamics, 4474
    - social engagement and support, 4474
  - positive mental health outcomes, 4472–4473
  - positive psychosocial factors, 4473
  - positive psychosocial traits, 4472–4473
    - coping self-efficacy, 4473
    - meaning in life, 4473
    - optimism, 4473
    - personal mastery, 4473
    - religiosity, 4473–4474
    - resilience, 4472

    - spirituality, 4473–4474
    - wisdom, 4473
  - post-traumatic growth, 4472
  - prevention, 4472–4473
  - recovery from serious mental illness, 4472
  - successful psychosocial aging, 4472
  - well-being, 4472
- Positive psychology, 785, 788–790, **2885–2899**
  - acceptance and commitment therapy, **2900–2902**
  - assessment, 2894–2895, 2895*t*
  - clinical work and, 2893–2899 (*See also* Client-centered psychotherapy)
  - concerns and criticisms, 2889–2890
  - defined, 2885
  - empirical findings, 2890–2893
  - history, 2886–2889
    - early ancestors, 2886–2887
    - humanistic psychology and, 2887
    - positive mental health and, 2887–2889
  - positive psychotherapies, 2897–2899
  - psychological health vision of, 2894
  - techniques, 2895–2897, 2896*t*
    - counting blessings, 2895–2896
    - savoring, 2895
    - strength-based, 2896–2897
- Positive psychotherapies, 2897–2899
  - definition of, 2897
  - examples of, 2897–2899
  - idealized sequence of, 2898*t*
  - traditional psychotherapy *versus*, 2888
- Positive reinforcement, 4760
  - defined, 4755

  - stimulant-related disorders and, 1315
- Positive symptoms, 4755
- Positive transference, 2750, 4755
- Positive valence, 299–301
- Positive-value blockages, 957
- Positron emission tomography (PET), 477
  - coincidence detection, 388, 388*f*
  - decay, 386–387
  - defined, 4755
  - gamma-ray detection, 387–388, 388*f*
  - physical principles of, 387–388
  - radionuclides, 386*t*, 387
  - random coincidences in, 392
  - spatial resolution, 392
- Possession syndrome, 2559–2562
  - clinical features, 2562
  - epidemiology of, 2560–2561
  - phenomenology of, 2560, 2561*t*
  - precipitants/etiology, 2561–2562
  - psychiatric diagnoses and, 2562
  - treatment of, 2562
- Possession trance, 2559
- Post, Robert, 1658
- Postcoital dysphoria, 2048
- Postcoital headache, 2049
- Posterior limb, 11
- Posterior nucleus, 38
- Postherpetic neuralgia (PHN), 2483
  - gabapentin, 3266, 3267
  - pregabalin, 3275
- Postictal psychosis, 679
- Postmodernism, 969–970

- Postpartum depression (PPD), 2597–2598
  - brexanolone for, 2598
  - estradiol for, 3426–3428
  - evidence-based treatments, 2597–2598
  - nonpharmacologic treatments for, 2598
- Postpartum disorders, 139
- Postpartum psychosis, 2598–2599, 4755
- Postpolio syndrome, 654
- Postprandial satiety, 224–226
- Postpsychotic depressive disorder of schizophrenia, 1690
- Poststroke depression (PSD), 487–488
  - course and prognosis, 495–498
  - etiology, 490–491
  - laboratory examination, 494–495
  - treatment of, 499
- Poststroke syndromes, 1027
- Postsynaptic density (PSD), 85*f*, 86
- Post-traumatic amnesia, 541, 2007
- Post-traumatic confusional state, 541
- Post-traumatic dysexecutive syndrome, 542
- Post-traumatic feeding disorder, 3766–3767
  - clinical features, 3766
  - course, 3766
  - differential diagnosis, 3766
  - etiology, 3766
  - treatment, 3766–3767
- Post-traumatic growth, 4472
- Post-traumatic stress disorder (PTSD), **1869–1885**
  - alcohol use disorder and, 1306
  - attention and, 1878
  - bipolar disorder and, 1742
  - borderline intellectual functioning and, 2533

- burn injuries and, 2470–2471
- buspirone for, 3170
- in cancer patients, 2354
- cardiac disease and, 2292
- child/adolescent, **3845–3850**
  - biologic factors, 3846
  - burn injuries and, 2477
  - civil litigation and, 4006
  - clinical features, 3846–3847
  - comparative nosology, 3846
  - course and prognosis, 3848
  - definition of, 3845
  - diagnosis, 3846–3847
  - differential diagnosis, 3848
  - epidemiology of, 3846
  - etiology of, 3846
  - group psychotherapy for, 3905
  - history, 3845–3846
  - laboratory examination, 3847–3848
  - pathology of, 3847–3848
  - psychological factors, 3846
  - sleep disorders and, 4089
  - social factors, 3846
  - telehealth forensic evaluations, 4009
  - treatment of, 3848–3849
    - biologic, 3849
    - child–parent psychotherapy, 3849
    - cognitive-behavioral therapies, 3848–3849
    - psychopharmacologic, 3849, 3931
    - psychotherapeutic, 3848–3849
- in children/adolescents, 3845–3850
- classification, 1873

- clinical features, 1875–1876
- clonidine for, 3222
- CNS mediation of clinical manifestations of, 1876–1878
- cognition in, 738–739
- cognitive control, 1878
- comparative nosology, 1873
- complex, 740–741
- context appraisal and update, 1878
- coronary artery disease and, 2281
- course and prognosis, 1881–1882
- defined, 4755
- definition of, 1869–1872, 2658
- delayed recall, 739–740
- diagnosis of, 1875–1876, 4801–4802*t*
- differential diagnosis, 1881, 3872–3873
- dissociation and, 1962
- dissociative amnesia and, 2007
- dissociative identity disorder and, 1996–1997
- DSM-5 criteria for, 1870–1871, 1870*t*, 2658
- early and preventive intervention for, 1883–1885
- emotion regulation and, 1878
- epidemiology of, 1873–1874
- epigenetics of, 281, 282–285, 283*f*, 284*f*, 1880
- etiology of, 1874–1875, 2659–2661
- fear conditioning theory of, 1877–1878
- functional brain imaging of, 1880
- gene-by-environment interaction modulates FKBP5 expression, 282–285, 283*f*, 284*f*
- genetic contribution to, 1879–1880
- genetics of, 1874–1875
- guanfacine for, 3222
- history, 1872

- HIV infection and, 604
- hypnosis for, 2924–2926
- ICD-11 criteria for, 1223
- immune-CNS interactions in, 189
- laboratory examination, 1876–1881
- malingering and, 2500
- MDMA-assisted therapy for, 2909
- memory and, 1878
- monoamine oxidase inhibitors, 3195
- neuroendocrine modulation, 1878
- neuroimaging, 3612
- neuropeptide systems and, 115
- neurosteroids and, 1879
- in older adults, 4219–4221, 4220*f*
- organ transplantation and, 2458
- pathology of, 1876–1881
- perception in, 738–739
- predisposing factors, 2660
- prevalence of
  - following mass trauma and disasters, 1874
  - in veterans, 1873–1874
- psychodynamic therapy for, 2761–2762
- psychophysiologic studies, 1876
- rating scales for, 1108
- in refugees, 845–846
- reliability and validity of syndrome, 1872
- risk factors, 1874
- in schizophrenia, 1463
- sensation in, 738–739
- sleep and, 1880–1881
- speechless terror, 739
- SSRIs for, 3131

  - stress and, 2421–2422
  - stroke and, 488
  - structural brain imaging of, 1880
  - substance use disorder and, 1883
  - subtyping, 1873
  - suggestibility, 740
  - symptoms in children and adolescents, 3868*t*
  - after TBI, 550–551
  - terrorism-induced trauma and, 4061–4062
  - threat detection and, 1878
  - topiramate for, 3269
  - translocator protein and, 404
  - traumatic brain injury and, 683
  - trazodone, 3199
  - treatment, 1882–1885
    - alternative, 1883
    - benzodiazepines, 1885
    - combined pharmacotherapy and psychological, 1883
    - complementary, 1883
    - early CBT, 1884
    - hydrocortisone, 1884
    - morphine, 1885
    - neurofeedback, 1883
    - nontraditional, 1883
    - pharmacotherapy, 1882–1883
    - propranolol, 1884–1885
    - psychological, 1882
    - psychological debriefing, 1884
    - psychotherapy, 1882
  - validity and social origin of, 1875–1876
  - valproate for, 3308
- Postural tremor, medication-induced, 3097

- Posturing, 4755. *See also* Catatonia
- Potassium (K^+) channels, 73*f*, 74–76, 74*f*
  - calcium-activated, 75
  - delayed rectifier, 74
  - H-channels, 75
  - inward rectifiers, 74
  - M-channels, 74
  - S-channels, 75
  - tandem pore, 76
  - A-type, 75
- Potency, 3012, 4755
- Potentially inappropriate medication (PIM), 4207–4208
- Pötzl, Otto, 4660
- Poverty, ethnoracial background and, 838–839, 838*t*
- Poverty of content of speech, 4755
- Poverty of speech, 4755
- Powassan virus disease, 617
- Power of test, 898–899
- PPACA. *See* Patient Protection and Affordable Care Act (PPACA)
- PPOs. *See* Preferred provider organizations (PPOs)
- Praag, Herman van, 1656
- Practice effect, 4755
- Practice guidelines, 1012–1018
  - AMA guidelines, 1013
  - APA guideline projects, 1014–1016
    - advantages, 1016–1017
    - dissemination and implementation, 1016
    - format, 1015
    - limitations, 1017–1018
    - recommendations, 1015
    - review and approval process, 1015–1016
    - updates, 1016

  - early efforts, 1012–1013
  - future of, 1018
  - Institute of Medicine reports, 1013
- Practicing, 922
- Prader–Willi syndrome, 271*t*, 3639–3643
  - defined, 4755
  - epigenetics of, 280, 280*f*
- Prados, Miguel, 4650
- Pramipexole (Mirapex), 99, 3107
  - for acute bipolar depression, 1735
  - for bipolar depression, 3110
  - dosage and administration, 3114, 3114*t*
  - drug interactions, 3113
  - for major depressive disorder, 3109
  - molecular structure of, 3107*f*
  - pharmacodynamics, 3109
  - pharmacokinetics, 3108
  - for restless leg syndrome, 3110–3111
- Prana, 2646–2649
- Prange, Arthur, 1656
- Pratt, Joseph, 2955
- Prayer, 2671
- PREA. *See* Prison Rape Elimination Act (PREA)
- Preadjudication examinations, 4008
- Preattentive processing, 301
- Prechtl, Heinz, 1959
- Precipitating factors, 2815
- Precision psychiatry, 3594
- Preconscious, 4755
- Preconscious system, 916
- Prediction error, 759
- Predictive utility, 1024

- Predictive validity, 880, 1037, 1092
- Predictive value, 881, 4755
- Predictor variable, 4755
- Predisposing factors, 2815
- PREE. *See* Partial reinforcement extinction effect (PREE)
- Preferred provider organizations (PPOs), 4755
- Prefrontal cortex, 9, 1504, 2781–2782, 3604–3605, 3608–3610, 3614
- Prefrontal lobotomy. *See* Lobotomy
- Pregabalin (Lyrica), 3274–3276
  - for anxiety disorders in later life, 4224–4225, 4305
  - chemistry, 3274
  - for chronic pain, 2490
  - clinical drug studies, 3275
  - dosage and administration, 3276
  - drug interactions, 3275–3276
  - history, 3274
  - pain management in burn patients, 2470
  - pharmacodynamics, 3275
  - pharmacokinetics, 3274–3275
  - precautions and adverse reactions, 3275
  - therapeutic indications, 3275
- Pregenital stages, 4755
- Pregnancy
  - asenapine and, 3086
  - benzodiazepines and, 3248–3249
  - brexpiprazole and, 3104
  - cariprazine and, 3106
  - clozapine and, 3070
  - early pregnancy loss, 2589
  - esketamine effects on, 3383
  - iloperidone and, 3084

  - ketamine effects on, 3383
  - lithium and, 3319–3320
  - lumateperone and, 3089
  - lurasidone and, 3064
  - melatonin receptor agonists and, 3395
  - nalmefene effects on, 3364–3365
  - naltrexone effects on, 3364–3365
  - olanzapine and, 3059
  - pharmacokinetics/pharmacodynamics and, 3414
  - pseudocyesis, 2589
  - psychiatric illness during, 2603
    - impact on mother, 2603
    - impact on offspring, 2603
  - quetiapine and, 3082
  - recurrent, 2589
  - recurrent pregnancy loss, 2589
  - risperidone and, 3074
  - SSRIs and, 3142–3143
  - stillbirth, 2589
  - tobacco use and, 1392–1393
  - voluntary termination of, 2589–2591
  - ziprasidone and, 3062
- Pregnanolone, 1879
- Pregnenolone, 163, 164–165
- Prejudice, 4755
- Premack, David, 763
- Premack principle, 761–763
- Premature birth, 668–669. *See also* Preterm infants
  - developmental coordination disorder and, 3674–3675
- Premature ejaculation, 3435, 4755, 4819–4820*t*
- Premenstrual dysphoric disorder (PMDD), 1689, 2596–2597
  - alprazolam for, 8

  - buspirone for, 3173
  - defined, 4755
  - diagnosis, 4794*t*
  - hormone suppression in, 3425–3426
  - SSRIs for, 3131
- Premenstrual mood syndromes, 2596–2597
- Premenstrual tension, 1670
- Premorbid, 4755
- Premorbid functioning, 1025
- Premorbid personality, 560
- Preoccupation of thought, 4755
- Preplate stage, 59*f*
- Preschool children. *See also* Children/childhood; Toddlers
  - psychiatric assessment of, 3547–3552, **3553–3557**
    - clinical formulation, 3557
    - multiaxial assessment, 3556–3557
    - multidisciplinary assessment, 3555
    - multiple informants, 3555
    - multiple modes of assessment, 3555–3556
    - multiple sessions, 3555
  - psychiatric symptoms and disorders in, 3554–3555
  - racism and, 4107–4108
  - response to hospitalization/.illnesses, 3987–3988
- Prescription Drug User Fee Act (PDUFA), 3022
- Presenile dementia, 4755
- Presenium, 4755
- Present State Examination (PSE), 1200
- Pressured speech, 4755
- Pressure of speech, 4755
- Preterm infants. *See also* Infants; Premature birth
  - brain development, 669
  - brain development in, 669

  - brain injury in, 669–672
    - cognitive outcomes, 672–673
    - imaging of, 672
    - neurodevelopmental outcome, 672
  - family impact of, 673
  - nomenclature for, 668*t*
  - psychopathologic outcome, 673–674
- Prevalence, 4755. *See also* Incidence
  - current (point and 1-month), 874
  - incidence and, 875
  - lifetime, 874–875
  - 1-year, 874
- Prevalence rate, 4755
- Preventive interventions
  - barriers, 4046–4047
  - categories of, 4041, 4041*f*
  - community-level, 4042–4046
  - definitions of, 4041–4042
  - depressive disorders, 4044–4045
  - early-onset psychosis, 4045
  - economic issues, 4046
  - family adversity, 4043–4044
  - future of, 4047
  - research, 4042
  - research issues, 4046–4047
  - substance abuse disorders, 4045
  - targeted development phase, 4042*f*
  - training issues, 4046
  - youth externalizing behaviors, 4044
- Preventive psychiatry, 4755
- Priapism, 3084
- Pride system, 957

- Primal fantasy, 2133
- Primal father, 4755
- Primal phantasy, 812–813
- Primal scene, 2036, 4755
- Primary anxiety, 2757
- Primary care, mental disorders in, 2276–2277
- Primary caregivers, 2400–2402
- Primary care physician, 4755
- Primary diagnosis, 4755
- Primary gain, 4755
- Primary interventions, 3490
- Primary narcissism, 926
- Primary parkinsonism. *See* Parkinsonism
- Primary prevention. *See* Preventive psychiatry
- Primary process thinking, 1166, 4756
- Primary progressive aphasia, 1021
- Primary psychiatric disorders, 2431
- Primary reinforcer. *See* Reinforcement
- Primary reinforcers, 4760
- Primary retarded ejaculation, 4756
- Primary sensory areas, 20–21
- Priming, 774, 774*f*, 775*f*
- Primitive idealization, 4756
- Prince, Morton, 1935, 4687, 4756
- Prince, Raymond, 2551, 2552
- Prince, Virginia, 2102
- Principal diagnosis. *See* Primary diagnosis
- Principals, 4029
- *Principles of Biomedical Ethics* (Beauchamp/Childress), 4424
- Principlism, 4424
- Prion diseases, 1271–1272
  - acquired, 631

- biomarkers, 1271
- clinical features, 1271
- comparative nosology, 1271
- course and prognosis, 642–643, 1272
- Creutzfeldt–Jakob disease (CJD)
  - characteristics of subtypes, 629–631
  - familial, 629
  - iatrogenic, 637
  - sporadic (nonfamilial), 629, 637
  - variant, 637
- defined, 4756
- definition of, 1271
- diagnosis, 1271
- diagnosis and clinical features, 636–639
- differential diagnosis, 642, 1271–1272
- epidemiology of, 632, 1271
- etiology of, 1271
- familial, 629–632
- familial CJD, 629
- fatal familial insomnia (FFI), 629–631
- genetic factors, 635
- Gerstmann–Sträussler–Scheinker disease, 631
- history, 1271
- history of, 632
- infectiousness, 644
- Kuru, 631, 637–638, 638*f*
- management of, 1272
- molecular pathogenesis, 632–633
- neurologic manifestations, 636–639
- neuropathology, 636, 1272
- neuropsychiatric aspects of, 629–644
  - clinical assessment, 641

    - mental status examination, 639–640
    - neurodiagnostic workup, 641–642
    - neuropsychological assessment, 640
    - psychological functioning, 640–641
  - nonfamilial, 629
  - nonhuman mammalian spongiform encephalopathies scrapie is a fatal disease, 631–632
  - preventive measures, 643–644
  - psychiatric manifestations, 639
  - transmissibility, 644
  - treatment of, 643
- Prion proteins (PrP)
  - cell (PrP^C)
    - function, 633–634
    - structure, 634, 635*t*
  - genetics of, 635
  - host protein interactions, 635
  - normal and abnormal forms of, 634*t*
  - strains, 635
- Prison neurosis. *See* Chronophobia
- Prison psychosis, 4756
- Prison Rape Elimination Act (PREA), 2118
- Prisons. *See* Correctional psychiatry
- Pristiq. *See* Desvenlafaxine (Pristiq)
- Pritchard, James Cowles, 4685
- Privacy, 4756
- Private schools, 4028, 4035
- Privilege, 4609–4610, 4756
- Privileged communication. *See* Privilege
- PRO. *See* Peer review organization (PRO)
- Probability, 4756
- Probability density functions, 896

- Probability sampling, 879
- Probability theory, 896
- Probe flow questions, 883
- Probiotics, 2631–2632
- Problem-oriented record, 4756
- Problem solving
  - defined, 4756
  - schizophrenia and, 1557
- Problem-solving hypnotherapy, 2922
- Problem-solving therapy, 2818, 4353
- Procarbazine (Matulane), 3185*t*
- Procedural memory, 730*t*
- Process C, 196, 208
- Process-oriented therapies, 3679–3680
- Process S, 196, 208
- Process schizophrenia, 4756
- Procyclidine hydrochloride, 3326, 3326*f*
- Prodrome, 1560, 4756
- Prodrome period, 1514
- Product labeling, 3009
- Professionalism, 4623
- Professional patients, 1907
  - disidentification, 4601
  - examination of, **4599–4606**
  - hypomanic-obsessive defensive style, 4602–4603
  - hypomanic-obsessive style, 4601–4602
  - identification with predecessors, 4601
  - professional crises in later years, 4600
  - psychological assessment, 4603–4606
  - psychotherapy for, 4604–4605
  - psychotherapy process group, 4605–4606
  - termination of treatment, 4606

  - traumatic abandonment, 4601
  - treatment, 4603–4606
- Professional Standards Review Organization (PSRO), 1159, 4756
- Profound mental retardation. *See* Mental retardation
- Progesterone, 163–164
- Progestin-only contraceptives, 2593*t*
- Prognosis, 4756
- Progressive aphasia, 1021*t*
- Progressive deep muscle relaxation (PDMR), 2816
- Progressive external ophthalmoplegia, 703
- Progressive multifocal leukoencephalopathy, 607
- Progressive muscle relaxation, 2171
- Progressive nonfluent aphasia, 1021*t*
- Progressive relaxation, 2921. *See also* Hypnotic induction ceremony
- Progressive supranuclear palsy (PSP), 563–564, 1263–1265
  - biomarkers, 1264
  - clinical features, 563, 1264
  - comparative nosology, 1264
  - course and prognosis, 564, 1264
  - definition of, 1263
  - diagnosis, 1264
  - differential diagnosis, 564, 1264
  - epidemiology of, 563, 1264
  - etiology of, 563, 1264
  - laboratory examination, 563–564
  - management of, 1265
  - neuropathology, 1264–1265
  - pathology of, 563–564
  - treatment of, 564
- Prohormones, 161
- Projection, 813–814, 920*t*, 932*t*, 2748*t*, 4756

- Projection fibers, 10–11
- Projection neurons, 3
- Projective Consciousness Model, 418
- Projective identification, 932*t*, 2494, 2496, 2748*t*, 4756
- Projective personality tests, 1043–1047, 1045*t*
  - Rorschach test, 1044–1045
  - sentence completion test, 1047
  - Thematic Apperception Test, 1045–1047
- Projective techniques, in child interview, 3525
- Projective test, 4756
- Project Liberty, 4068–4069
- Prolactin, 165, 1055–1056, 1772, 4756
- Prolactinomas, 711
- Prolixin. *See* Fluphenazine (Prolixin)
- Prolonged grief disorder, 4803*t*
- Promethazine (Phenergan), 3347
- Proopiomelanocortin, 161
- Propellant-fuels, 1370*t*
- Propensity scores, 903–904
- Propionic acidemia (PA), 695
- Propofol (Diprivan), 3232, 3236
  - pain management in burn patients, 2470
- Proportionality assumption, 910
- Propositional thought, 748
- Propranolol (Inderal). *See also* Beta blockers
  - absorption, 3225
  - for anxiety, 3227
  - chemistry, 3224–3225
  - for child/adolescent PTSD, 3849
  - distribution, 3225
  - dosage and administration, 3228
  - drug interactions, 3228

  - drug studies, 3226
  - elimination, 3225
  - history, 3224
  - laboratory interferences, 3228
  - for lithium tremor, 3227
  - mechanism of action, 3225–3226
  - organ/system effects of, 3226
  - for performance anxiety, 3227
  - pharmacodynamics, 3225–3226
  - pharmacogenetics of, 295*t*
  - pharmacokinetics, 3225, 3225*t*
  - pharmacologic actions, 3225
  - precautions and adverse reactions, 3228
  - for PTSD, 1884–1885
  - side effects and toxicity, 3226*t*
  - for substance use disorder, 3227
  - teratogenic effects, 3226–3227
  - therapeutic indications, 3227–3228
- Proprioception. *See* Kinesthetic sense
- Propriospinal myoclonus at sleep onset, 2187
- Prosaccade tasks, 1570
- Prosencephalon, 9, 9*f*
- Prosopagnosia, 4756
- Prospective study, 4756
- Prostaglandin E1 (PGE1), 3434
- Prostate cancer, 2363
- Prostigmin. *See* Neostigmine
- Prostitution, dissociative disorders and, 1952
- Protaxic mode, 960
- Protected health information (PHI), 1121
- Protection, 876
- Protective blockages, 957

- Protein chips, 249
- Protein Data Bank, 250*t*
- Protein-protein interactions (PPIs), 250
- Proteins, 236
  - regulation of, 249–250
  - structures, 239*t*
- Proteome, 234–235, 249–250
- Proteomics, 249
- Proton-observed/carbon-edited (POCE) spectroscopy, 346, 348
- Protoplasmic astrocytes, 5*t*
- Protoself, 420
- Prototaxic, 4756
- Protozoa, 626–628
- Protriptyline (Vivactil), 3152
  - pharmacogenetics of, 295*t*
- Provigil. *See* Modafinil (Provigil)
- Prozac. *See* Fluoxetine (Prozac)
- Pruning, 1509
- Prurigo nodularis, 2441–2442
  - clinical features, 2441
  - pathophysiology of, 2441
  - psychiatric comorbidity, 2441
  - treatment of, 2442
- Pruritus, 2438–2440
  - diagnosis, 2440
  - medical conditions associated with, 2440
  - neurologic conditions associated with, 2440
  - pathophysiology of, 2439–2440, 2439*f*
  - psychogenic, 2440
  - treatment of, 2440
- Pseudoaddiction, 1407
- Pseudoaggression, 4756

- Pseudoauthenticity, 4756
- Pseudobulbar affect (PBA), 486, 490, 492
  - assessment, 494
  - treatment of, 500
- Pseudocollusion, 4756
- Pseudocyesis, 2589, 4756
- Pseudodementia, 4230, 4756
- Pseudohallucination, 1174
- Pseudohermaphroditism, 2068, 2101
- Pseudologia fantastica, 1907, 4756
- pseudologia fantastica, 1180
- Pseudomemories, 2939–2940
- Pseudomotor agitation, 4756
- Pseudoparkinsonism, 4756
- Pseudospecificity, 1563
- Pseudounipolar, 1678
- Psilocin, 3008
- *Psilocybe cubensis*, 1350*f*
- Psilocybin, 1354, 2903, 2907, 3008
  - for alcohol use disorder, 2907
  - chemical structure of, 3386*f*
  - described, 4756
  - microdosing, 3390
  - studies, 3387–3388, 3387*t*
  - for treatment-resistant depression, 2906
- Psoriasis, 2444–2445
  - clinical features, 2444
  - neurogenic inflammation hypothesis, 2444
  - pathophysiology of, 2444
  - psychiatric comorbidity, 2445
  - treatment of, 2445
- PSP. *See* Progressive supranuclear palsy (PSP)

- PSRO. *See* Professional Standards Review Organization (PSRO)
- PSS. *See* Peer support specialists (PSS)
- PSTD. *See* Post-traumatic stress disorder (PTSD)
- Psychalgia. *See* Psychogenic pain disorder
- Psychasthenia, 4756
- Psyche, 801
  - assessment of, 2233–2234
  - defined, 4757
  - definition of, 2231
  - development of, 2233–2234
  - identity and, 2231–2234
  - psychobiology of, 2232–2233
- Psychedelics, 431–432, 2641–2642, 2673, 3386. *See also* Hallucinogens
  - antidepressant properties of, 3008–3009
  - defined, 4757
  - for OCD, 2907
  - organ/system effects of, 3386*f*
  - for patients with life-threatening illness, 2906–2907
  - for treatment of addiction, 2907
  - for treatment-resistant depression, 2906
- Psychedelic therapy, 2905
- Psychiatric diagnosis
  - brief history, 1199
  - heuristic formulations, 1201–1202
- Psychiatric diseases/disorders. *See also* Mental disorders
  - agitation and, 3577
  - alcohol-induced, 1303
  - broad autism phenotype, 4077–4078
  - cardiac masqueraders of, 2290
  - circadian rhythms in, 196–197
  - clinical manifestations of, **1161–1197**

    - anxiety disorders, 1185–1187
    - behavior, disturbances of, 1190–1192
    - challenges for, 1161–1164
    - consciousness, disturbances of, 1175–1180
    - drug-induced movement disorders, 1195–1196
    - interpersonal relationships, disturbances of, 1192–1194
    - language disorders, 1182–1183
    - memory, disturbances of, 1180–1182
    - mood, disturbances of, 1183–1185
    - motor activity, disturbances of, 1194–1195
    - motor disturbances and movement disorders, 1195
    - overview, 1161
    - perception, disturbances of, 1172–1175
        - body image distortion, 1175
        - hallucinations, 1172–1175
        - illusions, 1172
    - self, disturbances of, 1180
    - signs and symptoms features, 1164–1165
    - somatic disturbances, 1187–1190
        - appearance, 1189–1190
        - appetite and weight, 1188
        - energy, 1188–1189
        - sexual drive, 1189
        - sleep, 1187
    - thinking, disturbances of, 1165–1171
    - will disorders, 1183
- cognition in, 734–743
- due to a medical condition, dopamine receptor antagonists for, 3045
- E/I balance and, 80
- electroconvulsive therapy for, 3451–3452
- environmental contributions to, 2615

  - epigenetic mechanisms of, 282–285
  - gene mapping, 259–267
  - gene mapping investigations of, **251–268**
  - genetic contributions to, 2614–2615
    - copy number variants, 2615
    - epigenetics, 2615
    - single nucleotide variants, 2615
  - genetic overlap, 267
  - inhalant-induced, 1378–1380
  - judgment, disturbances of, 1171–1172
  - late life, **4178–4194**
  - memory impairments in, 780
  - neurotrophins and, 143–144
  - organ transplantation and, 2457–2461
  - perception in, 734–743
  - perinatal (*See* Perinatal psychiatric disorders)
  - pharmacogenetics of, 292
  - self and, 424–425
  - sensation in, 734–743
  - stress and, 2419–2422
  - suicide and, 2708–2710
  - transcriptional and epigenetic dysregulation in, 50–53
  - violence and, 2523–2524
- Psychiatric education
  - difficulties, 4600
  - examination of professional patients, **4599–4606**
    - disidentification, 4601
    - identification with predecessors, 4601
    - professional crises in later years, 4600
    - traumatic abandonment, 4601
  - graduate, **4586–4598**
  - residency, 4599–4600

- Psychiatric emergencies
  - aggression, management of, 2724–2726
  - agitation, management of, 2724–2726
  - alcohol intoxication, 2727
  - benzodiazepine, sedative, or hypnotic intoxication, 2727
  - cannabis intoxication, 2727–2728
  - capacity, 2733
  - child abuse, 2731
  - child/adolescent, **3571–3585**
    - assessment of developmental factors, 3577–3578
    - causes of, 3573*t*
    - comprehensive assessment, 3573–3574, 3574*t*
    - epidemiology, 3572
    - evaluation of, 3572–3573
    - family and social factors, 3578
    - psychotherapeutic and psychopharmacologic management of agitation, 3575–3577
    - specific
      - aggression, 3580–3581
      - autism, 3582
      - child abuse or neglect, 3582–3583
      - eating disorders, 3583–3584
      - intellectual disabilities, 3582
      - psychosis, 3581–3582
      - risky behaviors, 3581
      - substance abuse, 3581
      - suicidality and self-harm, 3578–3580
    - trauma-informed assessment and care, 3574–3575
  - competency, 2733
  - confidentiality, 2732–2733, 3584
  - defined, 4757
  - difficult patients, 2724

  - domestic abuse, 2730–2731
  - elder abuse, 2731
  - first priority, 2720
  - hallucinogen intoxication, 2728
  - high utilizers of service, 2732
  - history, 2719–2720
  - homelessness and, 2732
  - informed consent, 2733–2734
  - inhalant use, 2729
  - initial evaluation, 2722–2723
  - legal considerations, 3584
  - legal issues, 2732–2734
  - models of care and innovation, 3584
  - opioid intoxication, 2727–2728
  - phencyclidine intoxication, 2728, 2729
  - psychologic management, 2724
  - psychosocial issues, 2732
  - rape, 2731–2732
  - second priority, 2720
  - substance withdrawal and, 2729–2730
    - amphetamines, 2730
    - benzodiazepines, 2730
    - cocaine, 2730
    - hypnotics, 2729–2730
    - inhalants, 2730
    - opioids, 2730
    - sedatives, 2729–2730
  - suicide, **2705–2714**, 2705–2719
  - third priority, 2720–2722
  - treatment principles, 2722–2726
- Psychiatric Genomics Consortium (PGC), 1483
- Psychiatric hospitalist, **4533–4540**

  - benefits of, 4537–4538
  - as clinician educators, 4538
  - definition of, 4534
  - disadvantages of, 4538–4539
  - financial, political, and organizational issues, 4539
  - future directions, 4540
  - history of, 4536
  - as new specialty, 4539–4540
  - overview, 4535–4536
  - professional organizations, 4539
  - roles of, 4534–4535
  - specialty training, 4539
  - stages of hospital care, 4537
- Psychiatric hospitals
  - built since January 2000, 4506*t*
  - child/adolescent, **3935–3940**
    - COVID-19 pandemic-related challenges, 3940
    - inpatient psychiatric care, 3936–3939
    - levels of care, 3936, 3936*t*
    - reasons for hospitalization, 3935–3936
    - referrals, 3936
    - residential treatment centers, 3939–3940
  - future directions, 4524–4525
  - medical ethics in, 2384–2385
  - overview, 4504–4505
  - role of, 4504–4525
    - defining, 4509–4516
      - clinical rationale, 4509–4511
      - federal government and court actions, 4513–4514*t*
      - fiscal rationale, 4511–4512
      - governmental considerations, 4513–4516
      - social control rationale, 4509–4511

    - fulfilling, 4516–4520
      - assessment, 4516
      - discharge planning, 4519–4520
      - formation, 4516
      - treatment, 4517–4519
      - treatment planning, 4516–4517
    - historical evolution of, 4505–4509
    - other roles, 4523–4524
      - community education and patient/family support, 4524
      - consultation-liaison, 4524
      - emergency services, 4523
      - outpatient, 4524
      - step down/admission diversions services, 4523–4524
    - special populations, 4520–4523
      - children and adolescents, 4520
      - deaf individuals, 4523
      - developmental disabilities comorbidities, 4522
      - forensic patients, 4523
      - geriatrics, 4521–4522
      - medical comorbidities, 4522–4523
      - substance abuse comorbidities, 4521
      - veterans, 4520–4521
- Psychiatric illness. *See* Mental disorders
- Psychiatric nosology, 1200
- Psychiatric nurse, 4757
- Psychiatric rehabilitation, **1585–1592**, **4541–4558**
  - approaches to, 4553–4557
    - cognitive-behavioral therapy, 4554–4555
    - cognitive remediation, 4553–4554
    - illness/wellness management skill training, 4555–4556
    - rehabilitation of social cognition, 4554
    - sex education, 4556–4557

    - social learning, 4555
    - social skills training, 4555
    - supported education, 4556
  - assessment, 4550–4553
    - communication and social skills, 4551
    - functional, 4552–4553
    - neurocognitive symptoms, 4550–4551
    - social cognition, 4551
  - comorbid substance use disorders, 4557
  - concepts in, 4542
  - empirically based models of, 4543
  - evidence-based treatment and services, 4543–4544
  - integrative models of care, 4557–4558
  - in long-term care, 4414
  - in medical setting, 4541–4542
  - person-centered approach, 4542
  - and phases of severe mental illness, 4544–4550
    - multidimensional framework, 4545*f*
    - prodromal phase, 4546–4548
      - continuity of care, 4547–4548
      - early intervention and relapse prevention, 4546–4547
      - elapsed, 4546*t*
      - interventions, 4546*t*
    - recovery phase, 4550
    - stabilization phase, 4548–4549
    - stable phase, vocational rehabilitation in, 4549–4550
  - recovery, 4542–4543
- Psychiatric report, **1007–1012**
  - comprehensive treatment plan, 1012
  - diagnosis, 1012
  - further diagnostic studies, 1009
  - mental status, 1008–1009, 1010–1011*t*

  - prognosis, 1012
  - psychiatric history, 1007–1008, 1011–1012*t*
  - psychodynamic formulation, 1012
  - summary of findings, 1009
- Psychiatric research
  - animals models in, 296–297
  - evolutionary foundations for, **861–872**
- Psychiatric social worker, 4757
- Psychiatrist(s), 4757. *See also* Physicians
  - community-based treatment and, 3944
  - role in adoption, 3966–3967
  - role in foster care, 3962
  - role in psychiatric recovery, 1610–1612
  - sexual orientation of, 2081–2082
- Psychiatry
  - animal models in, 296–302
  - classification in, **1198–1227**
  - cultural assessments in, **2569–2579**
  - defined, 4757
  - electronic media in, **1121–1136**
  - epigenetics and, **269–285**
  - ethics
    - additional information and expertise, 4618
    - basic principles, 4617, 4617*t*
    - confidentiality, 4620–4621
    - conflicts of interest, 4622
    - continuous learning, 4619
    - decision-making, 4619–4621
    - dilemmas, 4617–4619
    - informed consent, 4621–4622
    - issues and tensions, 4618
    - professional boundaries, 4620

    - professionalism, 4623
    - psychiatric opinions, 4622–4623
    - risk and safeguards, 4618–4619
    - scope of practice, 4619
    - self-care and impairment, 4623
    - therapeutic relationship, 4620
    - third parties, 4622
    - values and perspectives, 4619
- evolutionary foundations for, **861–872**
- history of, **4642–4665**, **4683–4707**
    - Canada, 4650–4651
    - Denmark, 4644
    - diagnosis, 4654–4658
        - anxiety, 4656
        - bipolar disorder, 4657
        - brief reactive psychosis or brief psychotic disorder, 4657
        - catatonia, 4655–4656
        - delusional disorder, 4656
        - dementia, 4657–4658
        - mania, 4655–4656
        - melancholia, 4655
        - neurosis, 4658
        - psychotic depression, 4655–4656
        - schizophrenia, 4656–4657
        - trauma, 4657
    - early beginnings, 4642
    - first biologic psychiatry, 4645–4646
        - degeneration, 4645
        - early pharmacotherapy, 4645–4646
        - nerves, 4645
        - predisposition, 4645
        - psychotherapy, 4647–4648

  - sedatives and hypnotics, 4646
  - social psychiatry, 4647–4648
- France, 4643, 4650
- German-speaking Europe, 4643–4644, 4648–4650
  - Berlin, 4649
  - Heidelberg, 4649
  - Munich, 4649
  - Vienna, 4649
  - Zurich, 4650
- Germany, 4643–4644
- Italy, 4642
- new forms of care, 4665
- new techniques, 4658–4661
  - brain connectivity, 4661
  - controlled trials, 4659
  - convulsive and coma therapy, 4660
  - epidemiology, 4658
  - factor analysis, 4659
  - magnetic resonance imaging, 4661
  - neuroimaging, 4660–4661
  - personal inventories, 4659
  - quantitative EEG (QEEG) brain mapping, 4661
  - statistics, 4658–4659
  - tests in psychiatry, 4659
- Norway, 4644
- pediatric psychiatry, 4664–4665
- psychopharmacology, 4661–4662
  - antipsychotics, 4663
  - benzodiazepines, 4663
  - mood stabilizers, 4663
  - second-generation antipsychotics, 4663–4664
  - selective serotonin reuptake inhibitors, 4663

            - tranquilizers, 4662–4663
            - tricyclic antidepressants, 4663
        - Scandinavia, 4644
        - Spain, 4644
        - Sweden, 4644
        - United Kingdom, 4643, 4650
        - United States, 4644–4645, 4651–4654
    - immune-brain interactions in, 171–193
    - practice guidelines in, 1012–1018
    - sleep and, 213–214
    - stress and, **2412–2426**
    - subspecialties of, 4589–4590, 4589*t*
    - treatment gap, 859–860
    - world aspects of, **4667–4682**
- Psychic auras, 523*t*
- Psychic determinism, 4757
- Psychic energizer, 4757
- Psychic helplessness. *See* Helplessness
- Psychic immaturity theory, 2061
- Psychic trauma, 4757
- Psychoactive drug. *See* Psychotherapeutic drug
- Psychoactive substance abuse. *See* Substance use disorders (SUDs)
- Psychoactive substance dependence, 4757
- Psychoactive substance-induced mental disorder, 4757
- Psychoanalysis
    - defined, 4757
    - derivatives of, 2753–2756
    - history, 4647–4648, 4703
    - objects in, 810
    - psychoanalytic psychotherapy and, 2745–2756
    - relational turn in, 2750

  - therapeutic actions, 2752–2753
- Psychoanalyst, 4757
- Psychoanalytic frame, 2751–2752
- Psychoanalytic group psychotherapy, 4757
- Psychoanalytic psychotherapy, psychoanalysis and, 2745–2756
- Psychoanalytic theories, **914–937**
  - attachment theory, 927–928
  - developmental model, 922–923
  - ego psychology, 918–922
  - neuropsychoanalysis, 929–930
  - object relations theory, 923–926, 927
  - in practice, 930–934
  - relational theory, 928–929
  - self-psychology, 926
  - structural model, 917–918
  - topographic theory, 916–917
  - trauma hypothesis, 915–916
- *PsychoBible* (Favazza), 2684
- Psychobiology, 1653, 4757
- Psychocutaneous disorders, **2430–2448**
  - classification of, 2432*t*
  - delusional parasitosis, 2432–2434
  - factitious disorder, 2438
  - future directions, 2448
  - Koo classification, 2431
  - Morgellons disease, 2434
  - obsessive-compulsive and related disorders, 2434–2438
    - body dysmorphic disorder, 2436–2437
    - body-focused repetitive behaviors, 2437–2438
    - excoriation, 2435–2436
    - trichotillomania (hair-pulling disorder), 2434–2435

  - primary dermatologic disorders associated with psychiatric symptoms, 2440–2442
    - alopecia areata, 2440–2441
    - prurigo nodularis, 2441–2442
    - vitiligo, 2442
  - psychogenic purpura, 2438
  - psychophysiologic disorders, 2442–2446
    - acne, 2442–2443
    - angioedema, 2445–2446
    - atopic dermatitis, 2443–2444
    - psoriasis, 2444–2445
    - telogen effluvium, 2445
    - urticaria, 2445–2446
  - psychosocial impact of skin disorders, 2430–2431, 2432*f*
  - psychotropic medications for, 2433*t*
  - somatic symptom disorder with predominant pain, 2440
    - glossodynia, 2440
    - vulvodynia, 2440
  - unspecified somatic symptom and related disorder, 2438–2440
- Psychodrama, 2957, **2968–2973**
  - defined, 4757
  - Do–Undo–Redo, 2972
  - ethics, 2973
  - history, 2968–2969
  - methods, techniques, clinical issues, 2969–2970
  - research, 2972–2973
  - social and organic unity of humanity and, 2973
  - sociodrama, 2969–2970
  - sociometry, 2970
  - structure of, 2972
  - tele, 2970
  - theories, 2970–2972

    - Canon of Creativity, 2971
    - encounter, 2970–2971
    - human development, 2971
    - role theory, 2971
    - social atom, 2970
    - surplus reality, 2971
    - time, space, reality, and cosmos, 2972
  - as treatment of choice for trauma, 2972
- Psychodynamic family therapy, 2949
- Psychodynamic psychotherapy, 2140–2143, 2753–2754
  - child/adolescent, **3876–3883**, 3952–3953
    - clinical considerations, 3881
    - future directions, 3883
    - history, 3876–3877, 3877
    - practical matters, 3880–3881
    - research evaluation, 3882–3883
    - theoretical issues, 3877–3880
    - treatment phases, 3881–3882
  - efficacy and effectiveness of, 2990–2991
  - for functional gastrointestinal disorders, 2303
  - integration, 2141–2143
  - internalization, 2143
  - for major depressive disorder, 2754
  - panic-focused, 2754
  - for psychosis, **2783–2799**
    - case study, 2803–2804
    - ethical issues, 2802
    - future directions, 2803
    - history, 2785–2786
    - research and evaluation, 2802–2803
    - techniques, 2792–2802
    - theoretical issues, 2786–2787

  - short-term, **2763–2783**
  - understanding, 2140–2141
- Psychodynamics, 4757
- Psychodynamic theory, 1707
- Psychoeducation, 2744
  - cognitive behavioral therapy and, 4361–4362
  - for dementia, 1248
  - family interventions, 4374
  - for psychosis, 4366–4367
- Psychoeducational assessment, 3562–3563
- Psychoeducational family therapy, 2948–2949
- Psychoendocrinology, 4757
- Psychogenic, 4757
- Psychogenic amenorrhea, 2584
- Psychogenic amnesia, 782, 4757
- Psychogenic cough, 2324
- Psychogenic fugue, 1179, 4757
- Psychogenic nonepileptic seizures (PNES), **532–535**, 680–681
  - diagnosis of, 533–534
  - epileptic seizures *versus*, 533*t*, 534*t*
  - malingering and, 2499
  - pathophysiology of, 534
  - physiologic paroxysm and, 534*t*
  - treatment of, 534–535
- Psychogenic pain, 2388*t*
- Psychogenic pain disorder, 4757
- Psychogenic purpura
  - clinical features, 2438
  - pathophysiology of, 2438
  - treatment of, 2438
- Psychogenic voice therapy, 3709
- Psychohistory, 4757

- Psychoimmunology, 4757
- Psychokinesis, 4757
- Psycholinguistics, 4757
- Psychological abuse, child/adolescent, 3969
- Psychological defense system. *See* Defense mechanisms
- Psychological energies, 917
- Psychological factors affecting other medical conditions (PFAOMC), 1902
- Psychological factors affecting physical conditions, 4757
- Psychological first aid (PFA), 2425, 2661, 2662*t*
- Psychological testing
  - purposes of, 1035–1036
    - differential diagnosis, 1035
    - narrow-band assessment, 1036
    - psychotherapy aid, 1035–1036
- Psychologist, 4757
- Psychology
  - defined, 4757
  - positive, **2885–2899**
- Psycholytic therapy, 2905
- Psychometric assessment
  - interpretation, 3550–3551
  - personality measures, 3548–3549
  - referral for, 3547–3548
  - reliability, 3549
  - report, 3551
  - test batteries, 3548, 3549*t*
  - validity, 3549–3550
- Psychometry, 4757
- Psychomotor, 4757–4758
- Psychomotor acceleration, 1676
- Psychomotor agitation, 1672, 4758

- Psychomotor epilepsy. *See* Complex partial seizures (CPSs)
- Psychomotor retardation, 1672, 4758
  - bereavement depression and, 1695*t*
- Psychomotor seizures, 369. *See also* Complex partial seizures (CPSs)
- Psychomotor speed, 4170–4171
- Psychomotor stimulant, 4758
- Psychomotor stimulant theory, 443–444
- Psychonalysis, fundamental rule of, 2751
- Psychoneuroendocrinology, **157–170**
  - developmental, 159
  - human studies, 159
- Psychoneuroimmunology, 2416–2422
  - behavioral and psychological influences on immunity, 2418
  - biologic connections between CNS and immune system, 2416–2422
  - cancer and, 2419
  - central modulation of immunity, 2417
  - cytokine-brain link, 2418–2419
  - cytokines, 2418–2419
  - defined, 4758
  - neuroendocrine axis, 2416
- Psychoneurosis. *See* Neurosis
- Psycho-oncology, **2346–2382**
  - assessment, 2348
  - defined, 4758
  - geriatric, 2377–2379
    - anxiety, 2379
    - delirium, 2379
    - dementia, 2379–2380
    - depression, 2378
    - fatigue, 2379

  - historical background, 2346–2347
  - overview, 2346
  - pediatric, 2376–2377
    - diagnosis, 2376
    - family involvement, 2377
    - genetic testing in, 2377
    - pain management, 2376–2377
    - patients, 2376
    - problem magnitude, 2376
    - staff problems, 2377
    - team approach, 2377
    - treatment, 2376
  - treatment modalities, 2348–2350
    - chaplaincy services, 2350
    - psychopharmacology, 2348–2349
    - psychotherapy, 2349
    - social services, 2350
    - volunteerism, 2350
    - wellness programs, 2350
- Psychopathic personality, 4758
- Psychopathologic, 4758
- Psychopathology
  - defined, 4758
  - of epilepsy, 522
  - neuroticism and, 979
  - temperament and, 3624–3628, 3624*t*
- Psychopathology theories, 971, 973–974
  - Adler, Alfred, 948
  - Alexander, Franz, 953–954
  - application to therapies, 974
  - Berne, Eric, 963
  - Chodorow, Nancy, 966

  - Dollard, John, 971
  - Eysenck, Hans, 971
  - Fromm, Erich, 958–959
  - humanistic approaches, 971–973
  - Miller, Neal, 971
  - Sullivan, Harry Stack, 961
- Psychopathy, 2519
- Psychopharmacology. *See also* Medication(s)
  - acute *v.* prolonged effects, 3466, 3466*t*
  - brain stimulation, comparison with, 3465, 3466*t*
  - child/adolescent, **3925–3935**
    - anxiety disorders, 3927–3928
    - autism spectrum disorder, 3933
    - bipolar disorder, 3929–3930
    - developmental concerns, 3925–3926
    - enuresis, 3934
    - insomnia, 3933–3934
    - major depressive disorder, 3928–3929
    - obsessive-compulsive disorder, 3930
    - post-traumatic stress disorder, 3930–3931
    - schizophrenia, 3931
    - sleep-related disorders, 3933–3934
    - tic disorders, 3934
  - classification, 3008
  - clinical trials
    - medication nonadherence, 3018
    - statistical considerations, 3018
  - combination of drugs and, 3014
  - defined, 4758
  - dosing, 3012
  - drug factors and, 3010–3011
    - mechanisms, 3010

  - pharmacodynamics, 3010
  - pharmacokinetics, 3010–3011
- duration of treatment, 3012
- general issues in, 3006–3008
- general principles, 2995, **2995–3016**
- geriatric, **3438**
- history, 3008, 4661–4662
- informed consent, 3011–3012
- intracellular signaling pathways, 3004–3006
- monitoring, 3012
- neurotransmitter signal integration, 3006, 3006*f*
- patient education and, 3011–3012
- patient-related factors, 3011
  - concurrent medical or psychiatric disorders, 3011
  - diagnosis, 3011
  - drug preference, 3011
  - past treatment response, 3011
- pediatric, **3437–3438**
- pharmacogenetics, 286
- pharmacologic actions, 3009–3010
- pharmacotherapy for, 3933–3934
- receptor pharmacology, 2995–3004
- regulatory issues, 3009
  - black box warning, 3009
  - product labeling or information, 3009
- response to, 3013
- sensitization, 3013
- sex differences in, **3411–3421**
  - absorption, 3411–3412
  - distribution, 3412
  - efficacy, 3417–3420, 3418–3419*t*
  - elimination, 3412

    - gonadal steroid modulation and, 3412–3413
    - metabolism, 3412
    - pharmacodynamics, 3411–3412
    - pharmacokinetics, 3416
    - reproductive periods and, 3413–3414
    - safety, 3418–3419*t*, 3420
  - special populations, 3014–3016
    - children, 3014–3015
    - medically ill, 3015–3016
    - older adults, 3015
    - substance abuse, 3016
  - tolerance, 3013
  - treatment failure, 3013
  - treatment outcomes, 3013–3014
  - withdrawal, 3013–3014
- Psychophysiological insomnia, 2167
- Psychophysiologic disorders, 2431
- Psychopolitics, 4758
- *Psychose passionelle*, 1636, 4757
- Psychosexual development, 916–917
  - in children, 3494*t*, 3504
  - defined, 4758
  - stages of, 918–919*t*
- Psychosexual disorder, 4758
- Psychosexuality
  - definition of, 2032
  - factors in, 2032–2033
  - female sexual response cycle, 2033*t*
  - male sexual response cycle, 2034*t*
  - physiologic responses, 2033–2034, 2034*f*
- Psychosis
  - alternating, 525*f*

- biotypes, 1571–1572, 1573*f*
- buspirone for, 3172
- chemotherapy agents and, 2368*t*
- in children and adolescents, 3581–3582
- clinical high-risk for, 1479–1480
- cognitive behavioral therapy for, 2799–2802
  - A-B-C model, 2801
  - booster sessions, 2802
  - case study, 2803–2804
  - coping mechanisms, 2800–2801
  - delusional beliefs, 2801–2802
  - ethical issues, 2802
  - future directions, 2803
  - history, 2785–2786
  - normalizing and, 2801
  - in older adults, 4365–4369
  - patient engagement, 2799–2800
  - patient’s experience and core psychotic beliefs, 2801
  - patient’s story, 2800
  - personal life narratives, 2801–2802
  - research and evaluation, 2803
  - theoretical issues, 2786–2787
- defined, 4758
- definition of, 1613–1614
- dementia and, 1245, 1250, 4307–4308
- diagnosis
  - B-SNIP results, 1567–1569
  - phenotypes and, 1567–1568
- drugs for
  - comparisons of, 1578
  - dosage selection, 1578–1579
  - effectiveness of, 1575

  - selection of, 1575–1578
  - side effects, 1580–1582
  - starting, 1578–1579
- DSM-5 criteria for, 1613–1614
- endocannabinoids in, 135
- environmental risks, 1491*t*
- epilepsy and, 522
- estradiol for, 3428
- family therapy for, 3911*t*
- HIV in children/adolescents and, 676
- intellectual disabilities and, 3647–3648
- interictal, 526, 526t, 679
- learning theory and, 435–437
- malingering and, 2501–2502
- monoamine oxidase inhibitors, 3192
- multiple sclerosis and, 581–582
- NMDA receptor antagonist model of, 1352–1353
- nonaffective, 2599
- in older adults, 4427–4428
- Parkinson disease and, 559–560, 563, 4308
- personality disorders and, 1624–1626
- phenotypes of, **1566–1573**
- postictal, 679
- postpartum, 2598–2599
- poststroke
  - etiology, 491
  - treatment of, 500
- psychodynamic psychotherapy for, **2783–2799**
  - case study, 2803–2804
  - ethical issues, 2802
  - future directions, 2803
  - history, 2785–2786

    - research and evaluation, 2803
    - techniques, 2792–2802
    - theoretical issues, 2786–2787
  - psychoeducation for, 4366–4367
  - psychology of, 2787–2790
  - quantitative electroencephalography of, 1081–1082
  - schizophrenia, defining dimension of, **1613–1626**
    - biologic factors, 1616–1622
    - biomarkers, 1618–1622
    - clinical continuum, 1622–1626
    - cognitive dysfunction in, 1620
    - defined, 1613–1614
    - diagnostic approaches, 1614–1615, 1615*f*
    - genetics of, 1617–1618
  - schizophrenia and, 1479–1480
  - selective estrogen receptor modulators for, 3428
  - spirituality and, 2679–2681
  - syndromes, 1572–1573
  - threshold model of, 1623*f*
  - translocator protein and, 404
  - traumatic brain injury and, 683
  - treatment of, 3030
- Psychosis spectrum disorder
  - causal criteria for, 1490*t*
  - clinical epidemiology of, **1488–1502**
  - course and outcome of, 1498–1501, 1499*t*, 1500*t*
  - diagnosis, 1615–1616
  - early childhood adversity and, 1494
  - empirically based model of, 1501–1502
  - environmental risk factors, 1490–1491, 1491*t*
  - epidemiologic psychosis proneness-persistence-impairment model, 1502, 1502*f*

  - gene-environment interaction and, 1494–1495
  - genetic model of liability for, 1489–1490
  - genetic risk, 1488–1489
  - heritability, 1489
  - high-risk models, 1496–1498
  - integrated model of, 1502
  - neurodevelopmental model of, 1501–1502
  - phenotypes, 1489
  - premorbid indicators, 1495–1496
    - early developmental abnormalities, 1495
    - psychopathology, 1495–1496
  - risk and protective factors, 1491–1492
  - risk factors, 1488–1489
    - during childhood and adolescence, 1491–1492
      - cannabis use, 1493–1494
      - migration, 1493
      - minority position, 1493
      - tobacco use, 1493–1494
      - urban birth and upbringing, 1492–1493
  - stressful life events and, 1494
- Psychosocial deprivation, 4758
- Psychosocial development, 4758
- Psychosocial dwarfism, 4758
- Psychosocial factors
  - economic and environmental factors, 4346
  - ethnocultural issues, 4347–4348
  - future directions, 4349
  - positive, 4473–4474
  - in psychotherapy of elderly, **4341–4349**
  - relationships, 4341–4344
  - role changes, 4344–4346
  - societal attitudes and responses, 4345–4346

  - spiritual issues, 4348–4349
- Psychosocial identity, 421
- Psychosocially determined short stature. *See* Psychosocial dwarfism
- Psychosocial stressors. *See* Stress
- Psychosocial traits
  - positive, 4472–4473
    - coping self-efficacy, 4473
    - meaning in life, 4473
    - optimism, 4473
    - personal mastery, 4473
    - religiosity, 4473–4474
    - resilience, 4473–4474
    - spirituality, 4473–4474
    - wisdom, 4473
- Psychosomatic disorder. *See* Psychological factors affecting physical conditions
- Psychosomatic medicine
  - ancient history of, 2271
  - cardiovascular disorders, **2279–2293**
  - clinical practice of, 2275
  - clinical problems in, 2275*t*
  - current trends in, 2274–2275, 2274*t*
  - diabetes, **2325–2333**
  - endocrine and metabolic disorders, **2334–2345**
  - evaluation process in, 2275–2276
  - field of, 2278–2279
  - gastrointestinal disorders, **2293–2303**
  - history, 4703
  - history of research into, 2273–2274
  - modern history of, 2271–2272
  - obesity, **2304–2316**

  - research in, 2277–2278, 2277*t*
  - respiratory disorders, **2317–2325**
  - treatments in, 2276
- Psychostimulants, **3115–3129**. *See also* Stimulants
  - for acute bipolar depression, 1734–1735
  - for ADHD, 3122–3125
  - adverse events, 3125–3127, 3125*t*
  - atomoxetine, 3128–3129
  - for conduct disorder, 3124–3125
  - for depression, in cancer patients, 2357
  - EEG effects, 380
  - immediate-release formulations, 3120–3122
  - indications for, 3116–3118*t*
  - long-acting formulations, 3122
  - mechanisms of action, 3119–3120
  - modafinil, 3127–3128
  - for oppositional defiant disorder, 3124–3125
  - pharmacodynamics, 3122
  - pharmacokinetics, 3120–3122
  - pharmacologic properties of, 3121–3122*t*
- Psychosurgery, 4758
- Psychotherapeutic drug, 4758
- Psychotherapist
  - cancer in, 2376
  - defined, 4758
- Psychotherapy, 2735–2994. *See also* Psychoanalysis
  - for adolescents, 3952–3953
  - for anorexia nervosa, 2161–2162
  - behavior therapy, **2813–2821**
  - for binge eating disorder, 2163
  - for bipolar disorder, 1747, 1750–1752
  - for bulimia nervosa, 2162

- for cancer patients, 2349
- client-centered, 2974–2976
- cognitive behavioral therapy, **2822–2830**
- consumers of, 2987
- for conversion disorder, 1901
- deaf persons, 2978–2979
  - clinical issues, 2979
  - history, 2978
  - research, 2979
  - techniques, 2979
  - theoretical issues, 2978–2979
- defined, 4758
- for dementia, 1248
- for depression, in cancer patients, 2357
- for depressive disorders in children and adolescents, 3820–3821
- developmentally based, 754–755
- dialectical behavior therapy, **2841–2852**
- dignity therapy, 2976–2978
- effectiveness studies, 2988
- efficacy and effectiveness of, 2990–2991
- efficacy studies, 2987–2988
- ethical issues in, 2984
- evaluation of, **2987–2993**
- existential, 2979–2981
- family and couples therapy, **2942–2952**
- future directions, 2992–2993
- for gambling disorder, 1865
- goals, content, and process, 857
- group, **2954–2967**
- guided imagery therapy, 2984–2985
- hallucinogen-assisted, **2903–2911**
- history, 1744–1745, 4647–4648

- hypnosis, **2911–2942**
- for illness anxiety disorder, 1899
- integration, **2735–2742**
  - assimilative approach, 2738–2739
  - common factors approach, 2736–2737
  - differential psychotherapeutics, 2740–2741
  - future directions, 2741
  - pluralism, 2740–2741
  - stepped care, 2740–2741
  - technical eclecticism, 2737–2738
  - theoretical, 2739–2740
- for intellectual disabilities, 3649
- intermittent explosive disorder (IED), 2196
- interpersonal, **2831–2840**
- kleptomania, 2200
- mentalization-based treatment, **2863–2873**
- meta-analyses, 2988
- and modulation of brain functions, 2780
- for mood disorders, **1744–1752**
- narrative psychiatry, **2874–2885**
- neurobiology and, 2780–2781
- for older adults, 4349–4358
  - psychosocial factors
    - economic and environmental factors, 4346–4347
    - ethnocultural issues, 4347–4348
    - future directions, 4349
    - in psychotherapy of elderly, **4341–4349**
    - relationships, 4341–4344
    - role changes, 4344–4346
    - societal attitudes and responses, 4346
    - spiritual issues, 4348–4349
- paradoxical, 2981–2983

  - patient clinical characteristics and, 2992
  - patient cognitive characteristics and, 2992
  - patient demographics and, 2991–2992
  - patient-therapist matching and, 2992
  - patient-treatment matching and, 2992
  - patient variables related to outcome, 2991–2992
  - personality change and, 984
  - positive psychology, **2885–2899**
  - psychodynamic, **2783–2799**
  - for PTSD, 1882
  - research, 2987
  - single-session therapy, **2856–2862**
  - sociocultural factors, 857
  - solution-focused brief therapy, **2852–2862**
  - for somatic symptom disorder, 1898
  - strategic, 2981–2983
  - for suicide, 2716
  - supportive, **2742–2744**
  - synaptic plasticity and, 2781
  - systematic reviews, 2988
  - therapeutic alliance and, 2991
  - therapist characteristics and, 2991
  - for tic disorders, 3783
  - time, locus, and activity, 857
  - transactional, 2983–2984
  - transpersonal, 2985–2986
  - treatment of, cognitive behavioral therapy, 1899
  - treatment variables related to outcome, 2991
  - unipolar disorders, 1745–1747
- Psychotic, 4758
- Psychotic denial, 2748*t*
- Psychotic depression, 1687, 4230

  - history, 4655–4656
  - monoamine oxidase inhibitors, 3194
  - treatment of, 4240
- Psychotic depressive disorder. *See* Major depressive disorder (unipolar depression)
- Psychotic depressive reaction, 4758
- Psychotic disorders. *See also specific disorders*
  - acute and transient, **1627–1631**
  - alcohol-induced, 1304
  - anabolic steroid-induced, 1444
  - anatomical brain alterations in, 1620–1621
  - changes in DSM, 1205
  - cognitive dysfunction in, 1620
  - due to another medical condition, **1642–1644**, 1643*t*
    - comparative nosology, 1642–1643
    - course and prognosis, 1644
    - diagnostic and clinical features, 1643
    - differential diagnosis, 1643–1644
    - epidemiology of, 1643
    - etiology of, 1643
    - history, 1642
  - group psychotherapy for, 2966–2967, 2966*t*
  - in immigrants, 841–843
  - inhalant-induced, 1378–1379
  - in later life, 4268–4270
    - causes of, 4269*t*
    - differential diagnosis, 4268–4270
  - in long-term care, 4413
  - neurosteroids and, 139
  - oculomotor alterations in, 1618–1619
  - in older people, 4128
  - organ transplantation and, 2458

  - pediatric OCDs and, 3841–3842
  - phencyclidine- and other hallucinogen-induced, 1362–1363
  - in preterm adolescents, 674
  - rating scales for, 1102–1104
  - schizophrenia and, 1489
  - sensory and sensory-motor gating deficits, 1619–1620
  - steroid-induced, 513–514
    - clinical features, 513–514
    - course and prognosis, 514
    - diagnosis of, 513–514
    - differential diagnosis, 514
    - DSM-5 criteria for, 514*t*
    - epidemiology of, 513
    - etiology of, 513
    - laboratory investigations, 514
    - treatment of, 514
  - stimulant-induced, 1319
  - substance-induced, 4270
  - substance/medication-induced, **1640–1642**
  - after TBI, 551–552
- Psychotomimetic drug, 4758
- Psychotomimetic phencyclidine (PCP), 308
- *Psychotria viridis*, 1354
- Psychotropic drugs, 530–531, 1064–1067. *See also* Psychopharmacology
  - in Africa, 3440–3441
  - antianxiety drugs, 531
  - antidepressants, 530–531
  - antiepileptics, 530
  - antipsychotics, 530
  - for cancer-related fatigue, 2360*t*
  - cardiovascular effects of, 2291*t*

  - classification of, **3027–3033**
  - clozapine, 1064–1065
  - defined, 4758
  - dermatologic effects of, 2446–2448, 2447*t*
  - dosage and administration, 3210
  - EEG effects, 378
  - exposure, efficacy, and safety of, sex differences in, 3418–3419*t*
  - for functional gastrointestinal disorders, 2302
  - laboratory monitoring for, 1063t
  - nefazodone and, 3209–3210
  - for neuro-Covid, 190
  - pharmacokinetics of, sex differences in, 3415*t*
  - and physical health monitoring
    - antidepressants, 1064–1067, 1067
    - antipsychotic agents, 1064
    - mood stabilizers, 1065–1067
  - for repolarization abnormalities, 2288–2289
  - seizure threshold lowering effect of, 531
  - sexual dysfunction and, 3435–3436
- PTH. *See* Parathyroid hormone (PTH)
- PTSD Checklist for DSM-5 (PCL-5), 1108, 1109*t*
- Puberty, 2581
- Public Health Security and Bioterrorism Preparedness Act of 2002, 3022
- Public reporting of performance (PR), 4500–4501
- Public schools, 4028
- Puerperal psychosis. *See* Postpartum psychosis
- Pulmonary disease, 1595
- Pulmonary system, sleep and, 204–205
- Pulmonary thromboembolism (PTE), 2290
- Punch drink, 4658
- Punding, 560

- Punishment, 757, 4758
- Purcell, Edward, 326
- Pure Food and Drug Act of 1906, 3021
- Pure malingering, 2504
- Purging disorder, 4758
- Putamen, 28
- Putnam, Frank W., 1932, 1957
- *P* values, 899
- Pyramidal dysarthria, 467*t*
- Pyramidal neurons, 17–18, 17*f*
- Pyramidal signs, 481t
- Pyromania, **2197–2198**
  - defined, 4758
  - diagnosis, 2197, 4824*t*
  - epidemiology of, 2197
  - etiology of, 2197–2198
  - treatment of, 2198

Q

- QEEG. *See* Quantitative electroencephalography (qEEG)
- Qigong, 2649
- QIOs. *See* Quality improvement organizations (QIOs)
- Q-sort, 4758
- QTL. *See* Quantitative trait loci (QTL)
- QT syndrome, 2288
- Quadrangular therapy, 4758. *See also* Conjoint therapy
- Quality assurance, 4758
- Quality improvement organizations (QIOs), 1159
- Quality of life (QoL), 4180, 4180*t*
  - early-onset schizophrenia and, 4266
  - parents of child with ASDs, 4077
- Quality-of-life therapy, 2899

- Quantitative and experimental methods, 871
  - epidemiology, **873–895**
- Quantitative electroencephalography (qEEG), 1068–1090, 4661
  - anesthesia monitoring, 1082
  - biomarkers, 1076–1079
  - clinical applications, 1076–1082
    - ADHD, 1079–1080
    - Alzheimer disease (AD), 1080–1081
    - chronic pain, 1081
    - psychosis, 1081–1082
    - thalamocortical dysrhythmia, 1081
  - consciousness and, 1084
  - historical background, 1068–1070
  - homeostatic system generating and regulating EEG, 1072–1074
  - key-lock principle, 1083
  - modulated exogenous–endogenous interaction, 1085–1086
  - multivariate features, 1073–1074
  - neurofeedback, 1083–1084
  - neurometric, 1070–1071
  - neuron–interneuron interactions, 1086–1087
  - parallel multiplexed information processing channels, 1085
  - pharmacoEEG, 1082–1083
  - results reporting, 1072
  - selecting anatomical regions, 1075–1076
  - source localization, 1074–1076
  - steps in neurometric analysis, 1071–1072
- Quantitative law of effect, 762
- Quantitative trait loci (QTL), 255, 3587*t*
- Quantitative variable, 4758
- Quantum mechanics, 328–329
- Quazepam (Doral), 1429*t*, 3241
- Queen Square montage, 360, 360*f*

- Queer, 2060
- Queer theory, 2060
- Questioning, 2060
- Quetelet, Adolphe, 786
- Quetiapine (Seroquel), 99, 3078–3082
  - for agitation, 3576*t*
  - for anxiety disorders in later life, 4305
  - for bipolar depression, 3079
  - for bipolar disorder, 1730*t*, 1739
  - for bipolar mania, 3079–3080
  - for cancer patients, 2348
  - chemistry, 3078
  - for delirium, 4262
    - in cancer patients, 2352
  - for dementia, 1250
  - dosage and administration, 3082
  - drug interactions, 3082
  - for early-onset bipolar disorder, 3834
  - effects on organs and systems, 3080
  - for major depressive disorder, 1724
  - molecular structure, 3056*f*
  - overdose, 3082
  - pharmacodynamics, 3078
  - pharmacokinetics, 3078
  - precautions and adverse reactions, 3081
  - for schizophrenia, 3079
  - therapeutic indications, 3080
  - for tic disorders, 3785
- Quick Inventory of Depressive Symptomatology (QIDS), 1105, 1720, 3544
- QUIDS. *See* Quick Inventory of Depressive Symptomatology (QIDS)

- Quiet sleep, 200

R

- Rabbit syndrome, 1196
- Rabies, 617–618
- Race, 835
  - homosexuality and, 2077
  - mood disorders and, 1665
- Rachman, Stanley, 3885
- Racial/ Cultural Identity Development Model, 848
- Racial diversity, 835
- Racial identity
  - in adolescents, 3506–3507
  - in children, 848
  - development of, 3503–3504
- Racial prejudice, cognition in, 741–742
- Racism
  - child development and, 4107
    - early adolescence, 4108–4109
    - infant and toddlers, 4107
    - late-adolescence/transition age youth, 4109
    - preschoolers, 4107–4108
    - school-aged children, 4108
    - 10 to 12 years old, 4108
  - in child welfare, 4110
  - definition of, 4104–4105
  - developmental disabilities and, 4110–4111
  - in early care and early intervention, 4109
  - in educational system, 4109–4110
  - internalized, 836–837, 4105*t*
  - interpersonal, 836–837
  - interpersonal/personally mediated, 4105*t*
  - in juvenile justice system, 4110

  - levels of, 4105*t*
  - mental health and, 836–839, 838*t*
  - in mental health treatment, 4111–4112
  - neurobiology of, 4106
  - neuroplasticity of, 4106
  - role of child and adolescent psychiatrist, 4112–4113
  - in special education, 4110
  - stress and, 4105–4107
  - structural, 837–838
  - structural/institutional, 4105*t*
  - systemic, 837–838
  - in systems of care, 4109–4111
  - trauma, 4105
- Racker, Heinrich, 2751
- Radiation therapy, 2365–2366
- Radical behaviorism, 970
- Radioactive decay, 386–387, 386*f*
- Radioactivity, 386–387
- Radiotracer imaging with PET and SPECT, 385–405
  - basic physical principles, 386–387
  - emission tomography principles, 387–388
  - filtered back projection, 389–390, 389*f*, 390*f*
  - future directions, 404
  - general framework and compartment models, 394–397
    - irreversible models, 396–397
    - reversible models, 394–396, 394*f*
  - hybrid imaging systems, 392–393
  - image quality, 390–392
  - image quantification, 390–392
  - image reconstruction, 388–390
  - iterative reconstruction (IR) algorithm, 390
  - neurochemical imaging, 397–404

    - dopamine, 400–402
    - serotonin, 402–404
  - neuroinflammation imaging, 404
  - ordered subset expectation maximization (OSEM), 390, 390*f*
  - photon attenuation, 391, 391*f*
  - photon scatter, 391–392
  - radiotracers, 397, 398–399*t*
  - safety, 393
  - translocator protein, 404
- Rado, Sandor, 2061, 4647, 4651
- Rage, 1191
- Rahe, Richard, 2273, 2414
- Ramazicon. *See* Flumazenil (Ramazicon)
- Ramelteon (Rozerem)
  - chemistry, 3391
  - for delirium, 1237
  - for dementia with Lewy bodies, 1268–1269
  - dosage and administration, 3396
  - drug interactions, 3395–3396, 3396*t*
  - drug studies, 3393–3394
  - molecular structure, 3392*f*
  - organ/system effects of, 3394
  - pharmacodynamics, 3393
  - pharmacokinetics, 3391–3392, 3392*t*
  - precautions and adverse reactions, 3395
  - therapeutic indications, 3394–3395
- Ramón y Cajal, Santiago, 4644
- Random, 4759
- Random forests, 907
- Randomization tests, 903
- Randomized clinical trials (RCTs), 902–906
  - causality, 902, 904

  - likely responder method of analysis, 903
  - matching and randomization, 902–903
  - missing or incomplete data, 904–906
  - propensity scores, 903–904
  - randomization tests, 903
- Random sample, 4759
- Random variable, 4759
- Range, 4759
- Rank, Otto, 2764, 4687, 4759
- Ransomware, 1131
- Rape, 2624–2625. *See also* Sexual abuse
  - defined, 4759
  - legal definition of, 2624
  - perpetrators, 2624
  - prevalence of, 2624–2625
  - psychiatric emergencies and, 2731–2732
  - risk factors, 2624
- Raphe nuclei, 41
- Rapid cycling, electroconvulsive therapy for, 3451
- Rapid-cycling bipolar disorder, 1694
- Rapid eye movement. *See* REM (rapid eye movement) sleep
- Rapid eye movement sleep behavior disorder, 2183
- Rapid personality assessment instruments, 1043, 1044*t*
- Rappaport, David, 1930, 1966
- Rapport, 4759
- Rapprochement, 922–923
- Rapture-of-the-deep syndrome, 4759
- Rare variant, 3587*t*
- Rasagiline (Azilect), 3190
  - adverse effects, 3190
  - dosing, 3190
  - lactation and, 3190

  - monitoring parameters, 3190
  - pharmacodynamics, 3190
  - pharmacokinetics, 3190
  - pregnancy and, 3190
- Rashid, Tayyab, 2897
- RASopathy, 3641*t*
- Rastafari, 2673
- Rating scales, 1047–1048, **1090–1120**, 3541–3547
  - anxiety disorders, 1106–1108
  - assessment of, 1091–1093
  - benefits of, 1090
  - bipolar disorders, 1105–1106
  - childhood disorders, 1115–1120
  - comprehensive (omnibus), 3541–3543
  - definition of, 1090
  - depressive disorders, 1104–1105
  - disorder-related (domain specific), 3543–3547
  - eating disorders, 1111–1112
  - key references, 1094–1095*t*
  - limitations, 1090, 3547
  - measurement procedures, 1091
  - neurocognitive disorders, 1112–1114
  - obsessive-compulsive disorders, 1108
  - personality disorders and personality traits, 1114–1115
  - psychometric properties, 1091–1093
  - schizophrenia spectrum and other psychotic disorders, 1102–1104
  - selection of, 1093–1120
    - functional status, impairment, and general symptom severity, 1093–1098
    - psychiatric diagnosis and screening, 1099–1100
    - psychological functioning, 1100–1102

    - side effects, 1098–1099
  - substance-related and addictive disorders, 1109–1111
  - trauma and stressor-related disorders, 1108
  - types of, 1090–1091
- Ratio enforcement schedules, 762
- Rational emotive behavior therapy (REBT), 2818
- Rationalization, 921*t*, 2748*t*, 4759
- Rawls, John, 4617
- Ray, Isaac, 4685, 4759
- Raymon, Fulgence, 4650
- Rayner, Rosalie, 3884
- Raz, Amir, 2915, 2918
- Razadyne. *See* Galantamine (Razadyne)
- RBANS. *See* Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)
- RCBF. *See* Regional cerebral blood flow; Regional cerebral blood flow (rCBF)
- RDoC (Research Domain Criteria)
  - brief history, 1209–1210
  - constructs, 1209
  - recent updates, 1210
  - task paradigms, 1210*t*, 1211*t*
- Reaction formation, 921*t*, 2748*t*, 4759
- Reactive attachment disorder (RAD), **3801–3807**. *See also* Attachment
  - attachment, 3801–3802
  - clinical features, 3804
  - course and prognosis, 3805–3806
  - defined, 4759
  - diagnosis, 3804, 4800*t*
  - differential diagnosis, 3805
  - environmental factors, 3802–3803

  - epidemiology of, 3804
  - etiology of, 3803–3804
  - nosology, 3803
  - treatment of, 3806–3807
- Reactive depression, 1689, 4759
- Reactive disorder, 4759
- Reactive dissociative psychosis, 2023
- Reactive oxygen species (ROS), 242
- Reactive schizophrenia, 4759
- Reading disorder, 4759
- Reading the Mind in the Eyes Test, 4551
- Reality, 4759
- Reality principle, 4759
- Reality testing, 2959*t*, 4759
- Real-world functioning, 4461
- Reasoning, schizophrenia and, 1557
- Reassurance, 999
- Rebellious child, 963
- Recall, 472, 4759
- Receiver operating characteristic (ROC), 4759
- Recent memory, 4759
- Recent past memory, 4759
- Receptive aphasia, 4759. *See also* Fluent aphasia; Sensory aphasia
- Receptive character, 959
- Receptive dysphasia, 4759
- Receptive language, 4168
- Receptor pharmacology, **2995–3004**
  - adenylyl cyclase, 3004, 3004*t*
  - G protein–activated ion channels, 3004, 3005*f*
  - G protein effectors, 3003, 3003*f*
  - phospholipases, 3004, 3004*f*

  - plasma membrane microdomains and GPCR signaling, 3002–3003, 3002*f*
  - receptor analysis, 2996–2997
    - binding assays, 2996, 2996*f*
    - biologic (efficacy) assays, 2996–2997
    - G protein modulation of agonist binding, 2997, 2997*f*
  - receptor structure and function, 2997–3002
    - bacterial toxins, 3002
    - G proteins, 3002
    - heterotrimeric G proteins, 3002, 3002*f*, 3002*t*
    - ionotropic receptors, 2997–2998, 2997*f*
    - ligand-gated receptors, 2997–2998, 2998*f*
    - metabotropic (G Protein–coupled) receptors, 2999*f*, 3000
    - receptor conformation upon agonist activation, 3000
    - receptor desensitization, 3002
    - receptors for growth factors, 2998–3000, 2999*f*
    - RGS proteins, 3002
    - voltage gated receptors, 2997, 2997*f*
- Receptors, 77, 4759
- Reciprocal connections, 6
- Reciprocal inhibition and desensitization, 4759
- Reciprocity, 6, 7*f*
- Recognition memory, 472
- Recombination, 254
- Reconciliation, 2674
- Reconsolidation, 1878
- Recovered memory, 1933
- Recovered memory syndrome, 4759
- Recovery, 4542–4543. *See also* Rehabilitation
- Recovery model, schizophrenia and, **1601–1613**
  - defined, 1601
  - evidenced-based practices, 1607–1608

    - assertive community treatment, 1607
    - cognitive behavioral therapy, 1608
    - electroconvulsive therapy, 1608
    - housing first, 1607
    - motivational interviewing, 1608
    - pharmacotherapy, 1607
    - repetitive transcranial magnetic stimulation, 1608
    - supported employment, 1607
    - wellness recovery action planning, 1608
  - exemplary practices, 1608–1610
    - crisis intervention team, 1608
    - hearing voices network, 1609
    - mental health courts, 1608–1609
    - Open Dialogue, 1609
    - Soteria, 1609
    - trauma-informed approach, 1609–1610
  - history, 1596–1603
  - illness-centered vs. person-centered, 1603–1604, 1607*t*
  - journey of, 1596–1606
  - model, 1605–1606, 1606*t*
  - outcomes, 1606–1607
  - peer support, 1612–1613
  - psychiatrist’s role, 1610–1612
- Recurrent brief depressive disorder, 1689, 4759
- Recurrent depressive episode, 4759
- Recurrent isolated sleep paralysis, 4759
- Recurrent pregnancy loss, 2589
- Recurrent spontaneous abortion, 2590*t*
- Recursive partitioning, 907
- Refeeding syndrome, 2161, 2161*t*
- Reference, 996*t*
- Referrals, 4004–4005

- *Reflechi twòp*, 2568
- Reflection, 820, 999
- Reflective abstraction, 745
- Reflective functioning, 928
- Refractory period, 2033, 4759
- Reframing, 949, 4759
- Refugees, 836
  - mental health of, 845–846
  - premigration experiences, 853*t*
  - PTSD in, 845–846
- Reggio Emilia, 4642
- Regional cerebral blood flow (rCBF), 4759
- Regional health information organization (RHIO), 1121
- Registration. *See* Memory
- Regression, 921*t*
  - in burn patients, 2474
  - defined, 4759–4760
- Regression models, 906
- Regressive-reconstructive approach, 4760
- Regulating hormones, 158
- Regulators of G protein signaling (RGS) proteins, 3002
- Rehabilitation
  - defined, 4760
  - history, 1586
  - individual placement, 1588–1590
  - overview, 1587
  - psychiatric, **1585–1592**
  - supported employment, 1588–1590, 1590*f*
  - supportive housing, 1587–1588
  - technology tools for, 1590–1592
    - mobile health, 1591
    - virtual reality, 1592

    - web-based interventions, 1590–1591
- Reich, Theodore, 1652
- Reich, Wilhelm, 954–956, 955*f*, 4760
  - character armor, 955
  - compulsive character, 955
  - hysterical character, 955
  - masochistic character, 955
  - personality theory, 955
  - phallic-narcissistic character, 955
  - treatment by, 955
- Reik, Theodor, 4760
- Reiki, 2646, 2649
- Reil, Johann, 4684
- Reinders, A. T. S., 1957
- Reinforce devaluation effect, 766*f*
- Reinforcement, 763, 999, 2816, 4760. *See also* Conditioning
  - depression and, 1655
- Reinforcement learning
  - mathematical models, 322–323
  - model-based *versus* model-free, 323
  - temporal-difference, 322–323
- Reinforcers, 756, 1315, 4760
- Reivich, Karen, 2898
- Relapse
  - conditioned cue-induced, 449
  - drug-primed, 449
- Relatedness, 4760
- Relational frame theory (RFT), 2900
- Relational problems, **2540–2548**
  - child affected by parental relationship distress, 2544–2545
  - comparative nosology, 2541–2542
  - defined, 4760

  - definitions of, 2540–2541
  - disruption of family by separation or divorce, 2546–2547
  - high expressed emotion level within family, 2547
  - hypnotherapy, 2927–2928
  - parent–child, 2542–2543
  - relationship distress with spouse or intimate partner, 2545–2546
  - sibling, 2543–2544
  - terminology, 2542*t*
  - uncomplicated bereavement, 2547–2548
  - upbringing away from parents, 2544
- Relational theory, 928–929
- Relationship Across Domains (RADs), 4551
- Relationship distress with spouse or intimate partner (RSDP), 2545–2546
  - child affected by parental relationship distress and, 2545
  - clinical features, 2546
  - comorbidity, 2546
  - definition of, 2545
  - diagnosis, 2546
  - differential diagnosis, 2546
  - etiology of, 2546
- Relationships, 869–870
  - as factor in psychotherapy of elderly, 4341–4344
  - mental representation of, 931–934
  - real, 933
- Relative frequency. *See* Frequency
- Relative refractory periods, 70
- Relative risk, 876
- Relative Value Scale, 4760
- Relaxation, 329–330
- Relaxation techniques, 2642
- Relaxation therapy

  - for functional gastrointestinal disorders, 2302
  - for insomnia disorder, 2171
  - pain management in burn patients, 2470
- Relaxation training, 2424
- Relay nuclei, 23
- Reliability, 879–880, 880*t*, 1037–1038
  - Cohen kappa statistic, 880
  - defined, 4760
  - definition of, 1091
  - internal consistency, 1037, 1092, 3537
  - inter-rater, 880, 3537
  - interrater, 1092
  - parallel form, 1037
  - of personality tests, 1024
  - of psychometric assessment, 3549
  - standard error of measurement, 1037–1038
  - test-retest, 879, 1037, 1092, 3537
- Religion, 2675–2677. *See also* Spirituality
  - child maltreatment and, 3978
  - child psychiatry and, 4013–4014
  - homosexuality and, 2077–2078
  - psychiatrists' perceptions about, 2677–2678
  - psychiatry and, 2350
- Religiosity, 2675–2677
  - positive psychiatry and, 4473–4474
- Religious delusion, 4724, 4760
- Religious life, 2683–2684
- REM (rapid eye movement) latency, 4760
- REM (rapid eye movement) sleep
  - alcohol and, 1299
  - defined, 4759, 4760
  - description of, 200–201

  - dreaming, 211–212
  - EEG, 202*f*
  - inflammation and, 193
  - neurobiology of, 207–208
  - sexual function, 205
- Remen, Rachel, 2691
- Remeron. *See* Mirtazapine
- Reminiscence, 4760
- Reminiscence therapy, 4352–4353
- Remission, 4760
- Remote memory, 781, 4760
- Remote monitoring, 1124
- Remotivation, 4760
- REM sleep behavior disorder (RBD), 4214–4215, 4760, 4813*t*
- REM sleep latency, 203
- Renewal effect, 761
- Reparative therapies, 2112
- Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), 4168
- Repeatable Battery for the Neuropsychological Assessment of Schizophrenia (RBANS), 4554
- Repellor states, 734
- Repetition compulsion, 4760
- Repetitive DNA, 240–242
- Repetitive pattern, 4760
- Repetitive Thoughts and Behaviors Scale, 1108
- Repetitive transcranial magnetic stimulation (rTMS), 82
  - brain stimulation and, 3467
  - for dissociative identity disorder, 1997
  - for geriatric depression, 4240
  - in Latin America, 3444
  - for older adults, 4338–4339

  - safety guidelines, 3470*t*
  - for schizophrenia, 1608
- Repolarization abnormalities, 2288–2290
  - antidepressants for, 2288–2289
  - antipsychotics for, 2288
  - buprenorphine for, 2289
  - methadone for, 2289
  - mood stabilizers for, 2289
  - psychotropic drugs for, 2288–2289
  - QT syndrome, 2288
  - stimulant and stimulant-like medications for, 2289
  - Torsades de pointes, 2288
  - tricyclic antidepressants for, 2289
- Repression, 921*t*, 2748*t*, 4760
- Repressive-inspirational group psychotherapy, 4760
- Reproduction, 2585
- *The Reproduction of Mothering* (Chodorow), 966
- Reproductive development, 2580–2582
  - adolescence, 2582
  - early sexuality, 2582
  - fetal development, 2580–2581
  - gender, 2581–2582
  - neonatal development, 2580–2581
  - puberty, 2581
- Reproductive hormonal therapy, **3421–3429**
  - actions, 3422–3424, 3424*f*
  - future directions, 3430
  - history, 3421–3422
  - reasons for, 3424–3425
  - structure, 3422–3424
  - types of, in affective illness, 3425–3429
    - androgen administration in midlife depression, 3428–3429

    - estradiol administration in perimenopausal depression and postpartum depression, 3426–3428
    - estradiol and SERMS in acute psychosis, 3428
    - hormone suppression in premenstrual dysphoric disorder, 3425–3426
- Reproductive medicine
  - fertility and infertility, 2585, 2585–2595
  - menstrual cycles and, 2583–2585
  - pelvic pain, 2595–2596
  - perinatal psychiatric disorders, 2597–2608
  - premenstrual mood syndromes, 2596–2597
  - psychiatry and, **2580–2611**
  - reproduction, 2585
  - reproductive development, 2580–2582
  - reproductive senescence, 2608–2611
  - sexuality and sexual dysfunction, 2585
  - sexually transmitted infections, 2595
- ReputationDefender, 2703
- Requip. *See* Ropinirole (Requip)
- Research Diagnostic Criteria (RDC), 1200, 1201
- Research Diagnostic Criteria-Preschool Age (RDC-PA), 3554
- Research Domain Criteria (RDoC)
  - brief history, 1209–1210
  - constructs, 1209
  - recent updates, 1210
  - task paradigms, 1210*t*, 1211*t*
- Reserpine, 4760
- Resick, Patricia, 2818
- Residency, starting, 4599–4600
- Residential treatment centers, 3939–3940, 4760
- Residents
  - defined, 4760

  - working hours of, 1155–1156
- Residual, 4760
- Residual schizophrenia, 4760
- Resilience, 785, 791–793
  - in adolescents, 3954–3955
  - of aging self, 4161–4162
  - cultural features of, 852, 2573
  - positive psychiatry and, 4473
  - temperament and, 3628
- Resistance, 69, 933–934, 2751, 4761
- Resistant attachment, 4761
- Resolution phase, 4761
- Resonance frequency, 327
- Resonant voice therapy, 3709–3710
- Resource Specialist Program (SRP) teachers, 4028–4029
- Respiratory disorders, **2317–2325**
  - asthma, 2317–2320
  - chronic obstructive pulmonary disease, 2320–2321
  - comparative nosology, 2317
  - cystic fibrosis, 2321–2323
  - definition, 2317
  - interstitial lung disease, 2321
  - psychiatric symptoms associated with medications, 2324–2325
  - respiratory failure, 2323
  - somatoform, 2323–2324
    - chronic cough disorder, 2323–2324
    - chronic hyperventilation, 2323
    - sighing dyspnea, 2323–2324
    - vocal cord dysfunction, 2323–2324
- Respiratory failure, 2323
- Respiratory system
  - cannabis effects, 1344

  - electroconvulsive therapy and, 3460
  - nalmefene effects on, 3364
  - naltrexone effects on, 3364
- Respondent-based interviews (RBIs), 3538–3540
- Respondent conditioning, 1315, 4761. *See also* Conditioning
- Response, 3013
- Response elements (REs), 246
- Response style, 2492
- Resting membrane potential, 67–69
- Resting metabolic rate (RMR), 230
- Resting-state functional connectivity, 406
  - analysis of, 408–410
    - association *versus* directionality, 410
    - independent components analysis, 409–410
    - ROI–ROI or network-based connectivity, 409
    - ROI/seed-based connectivity, 409
    - static *versus* dynamic assessments, 410
  - modalities for assessment of, 407–408
  - network approaches to, 409
  - networks, 406–407, 407*f*
  - psychopathology and, 410–415
- Resting-state functional connectivity (RSFC), 803–804
- Restless legs syndrome (RLS), 2183, 3110–3111
  - in children/adolescents, 4086–4087
  - defined, 4761
  - diagnostic criteria, 2183, 2184*t*, 4815*t*
  - gabapentin, 3266–3267
  - in older adults, 4214
  - questions for, 4214*t*
  - treatment of, 4214
- Restoril. *See* Temazepam (Restoril)
- Restraints, errors in, 1154

- Restricted affect, 4761
- Resurgence, 761
- Retardation, 4761. *See also* Mental retardation
  - psychomotor, 1672–1673
- Retarded ejaculation, 2043, 4761
- Retention. *See* Memory
- Reter, Donna, 2104–2105
- Reticular nuclei, 23
- Retinohypothalamic tract (RHT), 194
- Retirement, 4161
- Retroflexion, 4761
- Retrograde amnesia, 769–772, 1181, 4761
- Retrolenticular limb, 11
- Retrospective falsification, 4761
- Retrospective study, 4761
- Rett disorder, 4761
- Rett syndrome, 271*t*, 3641*t*
- Reversal, 4761
- Reverse transference, 4356
- Reversible inhibitor of monoamine oxidase type A (RIMAs), for anxiety disorders, 1834
- Revex. *See* Nalmefene
- ReVia. *See* Naltrexone
- Revised Behavior Problem Checklist (RBPC), 3543
- Revised Child Anxiety and Depression Scale (RCADS, RCADS-25), 3543, 3544
- Revised Children’s Manifest Anxiety Scale (RCMAS), 3545
- Revivification, 2917
- Revonsuo, Antti, 2915
- Reward circuitry, 1208
- Reward dependence, 2223
- Reward prediction error (RPE), 322

- Rey Auditory Verbal Learning Test (RAVLT), 1032, 4554
- Reynolds, Mary, 1927
- Reynolds Adolescent Depression Scale (RADS, RADS-2), 3544
- Reynolds Child Depression Scale (RCDS, RCDS-2), 3544
- Rheinstadter, August, 2272
- Rheumatic heart disease (RHD), 2285–2286
- Rheumatologic disorders, child/adolescent response to, 3989
- RHIO. *See* Regional health information organization (RHIO)
- Rhombencephalon, 9, 9*f*
- Rhythms, 2287
- Ribonucleic acid (RNAs), 236
  - noncoding, 239–240
- Ribosomes, 237
- Ribot law, 768
- Richmond Lunatic Asylum, 4643
- Rickels, Karl, 4653, 4663
- Right to refuse treatment, 4761
- Right to treatment, 4761
- Rigidity, 4761
- Risk factors, 876, 4761
- Risk genes, 3588–3591, 4173
- Risk-taking behavior, in children and adolescents, 3513–3514, 3581
- Risk/vulnerability model of temperament/psychopathology, 3624, 3624*t*
- Risperdal. *See* Risperidone (Risperdal)
- Risperidone (Risperdal), 3071–3078
  - for agitation, 2724, 3576*t*
  - for bipolar disorder, 1730*t*, 1738–1739
  - for bipolar mania, 3071–3072, 3075
  - for cancer patients, 2348
  - chemistry, 3071

  - for child/adolescent PTSD, 3849
  - for delirium, 4262
    - in cancer patients, 2352
  - for dementia, 1250
  - dosage and administration, 1579, 3074–3075, 3077–3078*t*
  - drug interactions, 3074–3075, 3076*t*
  - for early-onset bipolar disorder, 3834
  - effects on organs and systems, 3073
  - geriatric use, 3074
  - history, 3055, 4698
  - molecular structure, 3056*f*
  - for mood dysregulation in children and adolescents, 3819
  - overdose, 3075, 3078
  - pediatric use, 3074
  - pharmacodynamics, 3071–3072
  - pharmacogenetics of, 295*t*
  - pharmacokinetics, 3071
  - precautions and adverse reactions, 3073
  - for schizophrenia and psychosis, 3072–3073, 3075
  - for stereotypic movement disorders, 3813
  - therapeutic indications, 3073
  - for tic disorders, 3784–3785
- Ritalin. *See* Methylphenidate
- Ritanserin, 3055
- Rituals, 3841, 4761
- Rituximab, neurotoxicity of, 2465*t*
- Rivastigmine (Exelon), 3335*t*, 3338–3341, 4317
  - for Alzheimer disease, 4315*t*
  - chemistry, 3339
  - clinical drug studies, 3339–3340
  - contraindications, 4318
  - dosage and administration, 3340–3341, 4317

  - drug interactions, 3340
  - organ/system effects of, 3340
  - pharmacokinetics and pharmacodynamics of, 3339, 4315*t*, 4317
  - pharmacologic actions, 3339
  - precautions and adverse reactions, 3340
  - therapeutic indications, 3340
- Rizzuto, Ana-Marie, 2677
- RLS. *See* Restless legs syndrome (RLS)
- RNA editing, 246–248
- RNAs, 236
  - noncoding, 239–240
- RNA sequencing (RNA-seq), 44, 3587*t*
- RNA splicing, 245
- Robert’s Thematic Apperception Test (RTAT), 3548
- Robins, Eli, 1201, 4654
- Robinson, George, 4686
- Robotics, 4469
- ROC. *See* Receiver operating characteristic (ROC)
- Rogers, Carl, 973, 2887, 2974, 4648, 4690, 4761
- Rogers, Richard, 2495, 2503
- Role, 4761
- Role changes, 4344–4346
- Role confusion, 939
- Role playing, 4761
- Role-specialized homosexuality, 2072
- Role theory, 2971
- Role transitions, 4161
- Roller, Christian F. W., 4644
- Romano, John, 1230, 2278, 4254
- Romberg, Moritz, 2272
- Romberg sign, 470, 4761
- Ropinirole (Requip), 99, 3107

  - for bipolar depression, 3110
  - dosage and administration, 3114, 3114*t*
  - drug interactions, 3113
  - for major depressive disorder, 3109
  - molecular structure of, 3108*f*
  - pharmacodynamics, 3109
  - pharmacokinetics, 3108
  - for restless leg syndrome, 3110–3111
- Rorschach, Herman, 1044, 4687
- Rorschach test, 1044–1045, 1045*f*, 1045*t*, 1046*t*, 3549, 4761
- Rose, Nicolas, 2883–2884
- Rosenman, R. H., 2273
- Rosenthal, Norman, 1658
- Roseroot (*Rhodiola rosea*), 2638
- Rostroventral medulla (RVM), 215
- Rostrum, 12
- Roth, Martin, 4659
- Rotigotine (Neupro), 3107–3108
  - dosage and administration, 3114, 3114*t*
  - drug interactions, 3113
  - molecular structure of, 3108*f*
  - pharmacodynamics, 3109
  - pharmacokinetics, 3108
  - side effects, 3113
- Rotter, Julian, 970
- Rousseau, Jean-Jacques, 2512
- Royal Edinburgh Hospital, 4643
- Rozerem. *See* Ramelteon (Rozerem)
- Rubber-hand illusion, 426
- Rubella, 620
- Rubinstein–Taybi syndrome, 271*t*
- Rüdin, Ernst, 4649

- Rudrauf, David, 418
- Rudy, Lester, 4586
- *Ruiz v. Estelle*, 4628
- Rumination, 4761
- Rumination disorder, 2148, 3770–3771
  - clinical features, 3770–3771
  - course, 3771
  - default mode network connectivity and, 414–415
  - defined, 4761
  - diagnosis and clinical features, 2154, 4808*t*
  - diagnostic criteria, 2149
  - differential diagnosis, 2158, 3771
  - epidemiology of, 2151, 3771
  - etiology, 3771
  - treatment of, 2164, 3771
- Rush, Benjamin, 1652, 1927, 4644, 4653, 4684, 4761
- Rush, William, 4646
- Russell, Gerald, 2149
- Ryff, Carol, 2888

S

- SAD. *See* Seasonal affective disorder (SAD)
- S-adenosyl-L-methionine (SAMe), 2634–2635, 4143
- Sadism, 1191, 4761
- Sadistic personality disorder, 4761
- Sadness, 1669–1670
- Sadomasochism, 2095–2096, 4761, 4839–4840*t*
- Saffron (*Crocus sativus*), 2638–2639, 4146*t*
- SAFTEE (Systematic Assessment of Treatment-Emergent Events), 1098
- SAINT. *See* Stanford accelerated intelligent neuromodulation therapy (SAINT)
- Sakel, Manfred, 4649, 4660
- Salicylates, 1060
- Salience network (SN), 462, 803–804, 1756–1757, 1757*f*
- Saltatory conduction, 70
- *Salvia divinorum*, 1355, 1358
- *Samadhi*, 2670
- Same-sex marriage, 2079, 2080
- SAMHSA. *See* Substance Abuse and Mental Health Services Administration (SAMHSA)
- Sample, 4761
- Sample space, 896
- Sampling
  - definition of, 879
  - nonprobability, 879
  - probability, 879
- Sanatorium, 4761
- Sandler, Joseph, 1703
- Sanity, restoration of, 4416–4417
- SANS (Scale for the Assessment of Negative Symptoms), 1104

- Saperstein, Jerome, 1931
- Saphris. *See* Asenapine (Saphris)
- SAPS (Scale for the Assessment of Positive Symptoms), 1104
- Sarcoidosis, 661
- Sartre, Jean Paul, 419
- Satellite oligodendrocytes, 5
- Satiation, 222–224
- Satir, Virginia, 2943, 2947, 2948
- Saturation, 329
- Satyriasis, 2049, 2121, 4761
- Savage, Charles, 432
- Savella. *See* Milnacipran (Savella)
- Saxitoxin (STX), 72–73
- Scale for Suicidal Ideation (SSI-C), 2712
- Scale for the Assessment of Negative Symptoms (SANS), 1104
- Scale for the Assessment of Positive Symptoms (SAPS), 1104
- Scales. *See* Rating scales
- Scar model of temperament/psychopathology, 3624*t*
- Schaeffer, Charles, 3877
- Schally, Andrew, 116
- Schatzberg, Alan, 4652, 4654
- Schauer, Maggie, 1953
- Schedule for Affective Disorders and Schizophrenia for Children (K-SADS), 3540–3541
- Schedule of reinforcement, 4761
- Scheiber, Steven, 4586
- Scheider, Kurt, 1888–1889
- Schein, Edgar, 2019
- Schema, 745
  - of object permanence, 746
- Schemata, 731
- Schema therapy, 2818, 2930

- Schilder, Paul, 1931, 2956, 3808, 4648, 4663
- Schildkraut, Joseph, 1656, 4662
- Schizoaffective disorder, 4762
- Schizoaffective disorder (SA), **1638–1640**
  - acute psychotic episodes in, 3043–3044
  - carbamazepine for, 3293
  - comparative nosology, 1638
  - course and prognosis, 1640
  - diagnostic and clinical features, 1639–1640, 4785*t*
  - differential diagnosis, 1640, 3871
  - dopamine receptor antagonists for, 3043–3044
  - DSM-5 criteria for, 1639*t*
  - epidemiology of, 1638–1639
  - etiology of, 1639
  - history, 1638
  - ICD-11 criteria for, 1221
  - lithium for, 3323
  - mood disorders and, 1700–1701
  - psychotic disorders in later life *versus*, 4269
  - symptoms in children and adolescents, 3868*t*
  - valproate for, 3307
- Schizoid fantasy, 921*t*, 2748*t*, 4762
- Schizoid personality disorder, 1625, 2245–2246, 4762
- Schizophrenia, **1447–1521**, 4263–4268
  - acute psychotic episodes in, 3043–3044
  - agitation in, 3057, 3059, 3061, 3062, 3101, 3103
  - alcohol use disorder and, 1306
  - amino acid transmitters and, 90–92
  - auditory hallucinations and, 1173–1174
  - benzodiazepines, 3246
  - bioenergetic dysfunction and, 1507
  - biologic classifications for, 1521

- bipolar disorders and, 1489
- in cancer patients, 2360
- cannabis and, 1341–1342
- carbamazepine for, 3293
- caregiver burden, 4081–4082
- changes in DSM, 1205
- chronic, 1513–1514, 1518
- clozapine treatment of, 2, 161
- cognition in, 735
- cognitive behavioral therapy for, 2825, 4371
- cognitive dysfunction in, 3111
- cognitive model of, 4366
- combination therapy for, 3409–3410
  - antidepressant–antipsychotic combinations, 3409
  - combining two antipsychotics, 3409
  - mood stabilizer–antipsychotic combinations, 3409–3410
- comorbidities, 1478–1479
  - physical disease, 1478
  - substance abuse, 1478–1479
- comparative nosology, 4263
- course and outcome, 1475–1478
  - secular trends, 1476
  - untreated schizophrenia, 1475–1476
  - variations in outcome, 1477–1478
  - in WHO studies, 1476–1477, 1476*t*
- criminal offending and, 1479
- D_2 receptor availability in, 400–401
- defined, 4762
- dementia and, 1247
- for depressive episodes associated with bipolar I disorder, 3063
- diagnostic criteria, 1449–1451, 4783–4784*t*
  - current, 1450–1451

  - historical perspective, 1449
  - modern, 1449–1450
  - subtypes, 1451
- differentiation from other disorders, 1469–1470
- dopamine hypothesis of, 1535
- dopamine receptor antagonists for, 3043–3044
- dopamine transmission in, 401–402
- drug development, 1584–1585
- DSM-5 criteria for, 1450–1451
- early-course, 1514, 1518–1519
- early-onset, 4264–4267, 4307
- economic costs of, 1480–1481
- EEGs in, 375
- electroconvulsive therapy, 3450–3451
- epidemiology of, 4264
- epigenetic dysregulation in, 52
- in ethnic minorities, 1474–1475
- everyday functioning and, 4192–4193
- family burden, 4081–4082
- first-episode, 1514, 1518–1519, 1575, 3057
- fMRI of, 343–344
- folate (vitamin B9) and, 2653
- functional brain imaging in, **1522–1534**
- functional connectivity abnormalities in, 411–413
  - altered default mode connectivity, 412
  - altered thalamic connectivity, 411, 411*f*
  - graph theory analyses, 412
- functional magnetic resonance imaging of, **1522–1534**
  - analytic methods, 1524–1526
    - connectivity analysis, 1525–1526
    - group-level random effects, 1524–1525
    - pattern analysis and classification, 1525

        - subject-level time series, 1524–1526
    - arterial spin labeling, 1524
    - blood oxygenation level-dependent imaging, 1522–1523
    - cognition studies, 1527–1529, 1529*f*, 1530*f*
    - emotion studies, 1529–1532, 1531*f*
    - event-related MRI, 1523, 1524*f*
    - genetics studies, 1532–1533
    - pharmacology studies, 1532
    - of at-risk individuals, 1533, 1534*f*
    - special considerations, 1526–1527, 1527*f*
- GABAergic neuropathology and, 91–92
- gene mapping, 264–265
- genetic overlap with other disorders, 267
- genetics of, 1482–1487
    - adoption studies, 1482–1483
    - architecture of, 1486–1487
    - biologic mechanisms, 1486
    - clinical implications, 1487
    - copy number variations (CNVs), 1485–1486
    - family studies, 1482–1483
    - genome-wide association (GWA) studies, 1483–1485
    - molecular, 1483–1486
    - neurodevelopmental spectrum, 1487
    - sequencing studies, 1486
    - twin studies, 1482–1483
- global burden of, 1467–1482
- glutamatergic synapse and, 91*f*
- group psychotherapy for, 2966–2967, 2966*t*
- hallucinations in, 320–321
- history, 4263, 4656
- HIV infection and, 596
- ICD-11 criteria for, 1221

- in immigrants, 841–843
- immune-CNS interactions in, 185–187
  - adaptive immune changes, 186
  - innate immune changes, 185
  - maternal immune activation, 186
- incidence of, 1470–1472, 1472*f*, 1472*t*, 1473*f*
- inflammation hypothesis, 1509
- insomnia comorbidity, 2168–2169
- integrative pathophysiology of, 1509
- in isolated populations, 1475
- late-onset, 4307
- as a lifespan disorder, 1521
- maintenance therapy, 1583–1584
- medical health in, **1592–1600**
  - cancer, 1595
  - cardiovascular disease, 1594
  - diabetes mellitus, 1594
  - infectious diseases, 1595
  - mortality rates, 1593
  - oral, 1595
  - overview, 1592–1593
  - prevention, 1598–1599
  - pulmonary disease, 1595
- in migrant populations, 1474
- molecular brain imaging in, **1535–1555**
- molecular imaging in, **1535–1555**
  - artificial intelligence, 1553–1554
  - drug development, 1545–1549
  - genetics, 1553
  - machine learning, 1553–1554
  - pathophysiology, 1538–1545
    - benzodiazepine receptors, 1544

        - cannabinoid, 1544–1545
        - cholinergic receptors, 1543–1544
        - dopamine, 1538–1542, 1539*t*
        - glutamate, 1544
        - histamine H_1 receptors, 1544
        - opiates, 1544–1545, 1545
        - serotonin, 1542–1543
    - theories of, 1535–1537
        - cerebral blood flow, 1536
        - cerebral metabolism, 1536
        - D_2-like receptors, 1536–1537
    - treatment, 1555
- mood disorders and, 1700, 1700*t*
- mortality rates, 1479
- motor disturbances of, 1196
- MRI applications to, 340–341
- MRS in, 348
- negative symptoms of, 438, 1582
- neural electrical activity and, 305–308
- neurocognition in, **1556–1566**
- as a neurodevelopmental disorder, 1508–1509
- neuroimaging
    - functional, **1522–1534**
    - molecular, **1535–1555**
    - structural, **1510–1521**
- neuroinflammation imaging, 1545
- neuropathology of, **1503–1509**
- neurotransmitters and, 1505–1507
- nitric oxide and, 127
- in older people, 4128
- overview, 1447–1448

- perception in, 735
- pharmacologic treatment, **1574–1585**, 1576–1577*t*
  - acute episodes, 1575–1580
  - dosage selection, 1579
  - history, 1574–1575
  - long-acting injectable, 1578
  - phases of, 1575
  - route of administration, 1578
  - side effects, 1580–1582
- phenomenology of, 1448–1467
- phenotypes, 1469–1470
- postpsychotic depressive disorder of, 1690
- precautions and adverse reactions, 3069, 3069–3070
- prevalence of, 1470
- psychiatric rehabilitation, **1585–1592**
- psychosis, as defining dimension, **1613–1626**
  - biologic factors, 1616–1622
  - biomarkers, 1618–1622
  - clinical continuum, 1622–1626
  - cognitive dysfunction in, 1620
  - defined, 1613–1614
  - diagnostic approaches, 1614–1615, 1615*f*
  - genetics of, 1617–1618
- psychosis and, 1479–1480
- psychosis spectrum disorder and, **1488–1502**
- psychosocial treatment, 1584
- psychotic disorders and, 1489
- rating scales for, 1102–1104
- recovery, **1601–1613**
  - defined, 1601
  - evidenced-based practices, 1607–1608
    - assertive community treatment, 1607

        - cognitive behavioral therapy, 1608
        - electroconvulsive therapy, 1608
        - housing first, 1607
        - motivational interviewing, 1608
        - pharmacotherapy, 1607
        - repetitive transcranial magnetic stimulation, 1608
        - supported employment, 1607
        - wellness recovery action planning, 1608
    - exemplary practices, 1608–1610
        - crisis intervention team, 1608
        - hearing voices network, 1609
        - mental health courts, 1608–1609
        - Open Dialogue, 1609
        - Soteria, 1609
        - trauma-informed approach, 1609–1610
    - history, 1596–1603
    - illness-centered vs. person-centered, 1603–1604, 1607*t*
    - journey of, 1596–1606
    - model, 1605–1606, 1606*t*
    - outcomes, 1606–1607
    - peer support, 1612–1613
    - psychiatrist’s role, 1610
- risk factors, modifiable, 1596–1598
- self and, 424–425
- sensation in, 735
- small area variations, 1474
- social costs of, 1480–1481
- with substance abuse, 3068
- suicidal behavior, 1582
- suicide and, 2709
- suicide risk and, 3067–3068
- symptoms, 1451–1465

  - agitation, 1464–1465
  - anxiety, 1463
  - compulsions, 1463–1464
  - delusions, 1455–1457
  - depression, 1462–1463
  - disorganization syndrome, 1460–1462
  - hallucinations, 1452–1455
  - hostility, 1464–1465
  - insight, lack of, 1465
  - negative, 1457–1460
  - obsessions, 1463–1464
  - premorbid and prodromal signs and, 1622–1624
  - psychotic, 1452–1457
  - suicide, 1465
  - trauma, 1463
- topiramate for, 3269
- translocator protein and, 404
- treatment of
  - asenapine, 3085
  - clozapine, 3065–3069
  - immune therapies, 186–187
  - lumateperone, 3088
  - lurasidone, 3062–3064
  - olanzapine, 3057–3058, 3059
  - paliperidone, 3072–3073, 3075
  - pharmacotherapy, 3931
  - quetiapine, 3079
  - risperidone, 3072–3073, 3075
  - trazodone (Desyrel), 3200
  - valproate, 3307
  - ziprasidone, 3060–3061
- treatment-resistant, 3069

  - untreated, 1475–1476
  - very-late onset, 4267–4268
  - vulnerability to, 870
  - WHO Ten-Country Study, 1473, 1473*t*
- Schizophrenia Cognition Rating Scale (SCoRS), 4551
- Schizophrenia Online Access to Resources (SOAR), 1591
- Schizophrenia spectrum personality disorders, 1624
- Schizophreniform disorder, **1631–1634**
  - comparative nosology, 1632
  - course and prognosis, 1634
  - defined, 4762
  - diagnostic and clinical features, 1633, 4782–4783*t*
  - differential diagnosis, 1633
  - DSM-5 criteria for, 1632*t*
  - epidemiology of, 1633
  - etiology of, 1633
  - history, 1628*t*, 1631–1632
  - symptoms in children and adolescents, 3868*t*
- Schizophreniform/schizophrenia like psychosis, 525–526
- Schizophrenogenic mother, 2790–2791
- Schizotypal personality disorder, 1221, 1624, 2246, 4762
- Schneider, Kurt, 1449, 1449*t*, 4649, 4691
- Schneiderian first rank symptoms, 4762
- School academic/guidance counselors, 4029
- School adjustment counselors, 4029
- School administrators, 4029
- School-aged children
  - racism and, 4108
  - response to hospitalization/.illnesses, 3987–3988
- School-based health centers (SBHCs), 4037
- Schooler, Jonathan, 423
- School interventions, 3954

- School mental health counselors, 4029
- School nurses, 4029
- School phobia, 4762
- School shootings, 4067–4069
  - children's reactions to, 4067
  - counseling interventions, 4068
  - media coverage, 4067–4068
  - Project Liberty and, 4068–4069
- Schopenhauer, Arthur, 969
- Schore, Allan, 726
- Schreber, Daniel Paul, 4762
- Schultz, Wolfram, 437
- Schultze, Johannes Heinrich, 2984
- Schwab, John, 2278
- Schwab, Sidney, 4653
- Schwann cells, 5, 5*t*
- SCID (Structured Clinical Interview for DSM), 1099
- SCID-5-PD (Structured Clinical Interview for DSM-5 Personality Disorders), 1114–1115
- Scientific Registry of Transplant Recipients (SRTR), 2449
- SCN. *See* Suprachiasmatic nucleus (SCN)
- Scopace. *See* Scopolamine (Scopace)
- Scopolamine (Scopace), 3329–3331
  - chemistry, 3329
  - clinical drug studies, 3330–3331
  - dosage and administration, 3329–3330
  - molecular structure of, 3329*f*
  - pharmacologic actions, 3329
  - precautions and adverse reactions, 3329–3330
  - therapeutic indications, 3329
- Scopophilia, 4762
- Scotoma, 4762

- Scott, William, 2888
- Screen for Child Anxiety-Related Emotional Disorders (SCARED), 3544
- Screening, 1029, 4762
- Script theory, 751
- SCT. *See* Sentence Completion Test (SCT)
- SD. *See* Standard deviation (SD)
- SDS. *See* Serine deficiency syndromes (SDS); Zung Self-Rating Depression Scale; Zung Self-Rating Depression Scale (SDS)
- Searles, Harold, 2831
- Seasonal affective disorder (SAD), 196–197
  - defined, 4762
  - sleep disorders in children with, 4088
- Seasonality, 1675
- Seasonal patterns. *See* Seasonal affective disorder (SAD)
- Seclusion, errors in, 1154
- Secobarbital (Seconal). *See* Seasonal affective disorder (SAD)
- Seconal. *See* Secobarbital (Seconal)
- Secondary gain, 4762
- Secondary mania, 1691
- Secondary narcissism. *See* Narcissism
- Secondary prevention. *See* Preventive psychiatry
- Secondary process thinking, 1166, 4762
- Secondary psychiatric disorders, 2431
- Secondary retarded ejaculation, 4762
- Secondary sexual characteristics, 4762
- Second-generation antipsychotics (SGAs), 3027, 4306. *See also specific drugs*
  - antidepressants combined with, 3404
  - dosage/administration, 1579, 1583–1584
  - drug interactions, 4310
  - formulation and dosing information, 1576–1577*t*

- history, 4663–4664
- long-acting injectable, 1578
- long-term effects of, 1578
- maintenance treatment, 3079
- mechanism of action, 3065
- mood stabilizers combined with, 3407–3409
- pharmacologic actions, receptor activity, 4306
- precautions/adverse reactions
  - metabolic effects, 4311
  - tardive dyskinesia, 4311
  - weight gain, 4311
- pregnancy and, 2606–2607
- side effects
  - cardiovascular, 2291*t*
  - cerebrovascular adverse events, 1251
  - extrapyramidal side effects, 1545–1547, 1580
  - metabolic effects, 1581–1582
  - metabolic syndrome, 1251
  - tardive dyskinesia, 1580–1581
- therapeutic indications
  - aggression, 3650, 3737
  - agitated delirium, 2468
  - agitation, 2724
  - AIDS mania, 595
  - Alzheimer disease, 4202, 4522
  - bipolar disorder, 3834
  - catatonia, 723
  - delirium, 4260
  - dementia with psychosis or severe behavioral disturbances, 4308
  - epidemiology of, 503
  - mood disorders, 1723

    - non-affective psychosis, 2599
    - psychosis and agitation in dementia, 1250
    - schizophrenia, 1535
    - steroid-induced psychosis, 1056
    - suicidal behavior, 2712
    - tic disorders, 3785
    - unipolar depression, 3404
- Second Life, 1127
- Second messengers, 1771
- Second-order conditioning, 760
- Secretan, Henri, 1907
- Secure attachment, 4762
- Securely attached attachment, 3500
- Sedation
  - antihistamines for, 3348
  - quetiapine and, 3081–3082
  - of terminally ill patients, 2393
- Sedative-, hypnotic-, anxiolytic-related disorders, **1419–1432**
  - animal studies, 1422
  - clinical features, 1424–1428
  - defined, 1419–1420
  - diagnosis of, 1424
  - epidemiology of, 1424
  - etiology of, 1420–1421
  - $GABA_A$ receptors and, 1420–1422
  - human studies, 1422–1423
  - misuse liability, 1421
  - neuropharmacology of, 1420–1421
  - overdose, 1432
  - patterns of misuse, 1424–1428
  - substances, 1420
    - barbiturates, 1426–1427, 1426*t*

    - benzodiazepines, 1424–1426, 1425*t*
    - carisoprodol, 1427
    - chloral hydrate, 1428
    - γ-hydroxybutyric acid (GHB), 1427–1428
    - meprobamate, 1427
  - treatment of, 1428–1432
- Sedatives, 1420
  - defined, 4762
  - history, 4646
  - intoxication, 2727
  - withdrawal, 2729
- Seduction theory, 1930
- Seeman, Philip, 3039, 4662
- Segal, Zindel, 2898
- Segmental duplications (SegDups), 241, 3586. *See also* Copy number variants (CNVs)
- Séguin, Édouard, 4664
- Seizure disorders
  - barbiturates for, 3231
  - dissociative amnesia and, 2007
  - EEG findings, 372
  - gabapentin, 3267
  - intellectual disabilities and, 3652
  - lamotrigine for, 3296, 3297
  - levetiracetam, 3272
  - perampanel, 3276
  - tiagabine, 3271
  - topiramate, 3269
  - treatment of, 3652, 3737–3738
  - zonisamide, 3273
- Seizure-like behaviors, 1196
- Seizure(s). *See also* Complex partial seizures (CPSs); Epilepsy

  - classification of, 521*t*
  - clozapine and, 3069–3070
  - defined, 4762
  - definition of, 520
  - focal onset, 521
  - generalized onset, 521
  - management of, 532
  - pregabalin, 3275
  - tonic-clonic, 521
- Selection bias, 4762
- Selective estrogen receptor modulators, for acute psychosis, 3428
- Selective inattention, defined, 4762
- Selective mutism, **3861–3866**
  - behavioral approach tasks for, 3888*t*
  - clinical features, 1788, 3862–3863
  - cognitive behavioral therapy for, 3893–3894
  - comparative nosology, 1788, 3861
  - course and prognosis, 1788, 3864–3865
  - defined, 4762
  - definition of, 1788, 3861
  - diagnosis, 1788, 3862–3863, 4795*t*
  - differential diagnosis, 1788, 3863–3864
  - DSM-5 criteria for, 3864
  - epidemiology of, 1788, 3861
  - etiology of, 1788, 3861–3862
  - historical background, 3861
  - history, 1788
  - treatment of, 1788, 1839, 3865–3866
    - behavioral, 3865
    - psychopharmacologic, 3866
    - psychosocial therapies, 3866
    - speech and language therapy, 3866

- Selective serotonin reuptake inhibitors (SSRIs), 2488, **3129–3144**, 4297
  - for aggression, 3650
  - for anxiety disorders, in children and adolescents, 3859–3860, 3927
  - anxiety disorders, 1832–1834
  - for anxiety disorders in later life, 4303–4304
  - for autism spectrum disorders (ASDs), 3727
  - for binge eating disorder, 2163
  - bipolar depression, 3131
  - bulimia nervosa, 3131
  - for cancer-related fatigue, 2359
  - citalopram, 3138
  - defined, 4762
  - delayed onset of therapeutic effects of, 3132
  - for dementia, 1250
  - for depression, in cancer patients, 2356
  - dosing and administration, 3140–3142
  - drug interactions, 3138–3140
  - EEG effects, 379
  - escitalopram, 3138
  - excoriation, 2436
  - fluoxetine, 3135–3136
  - fluvoxamine, 3137–3138
  - generalized anxiety disorder, 3131
  - for geriatric depression, 4238
  - history, 3007, 4663
  - lactation and, 3143–3144
  - for major depressive disorder, 1722–1723
  - major depressive disorder, 3130–3131
  - mirtazapine and, 3212
  - for mood dysregulation in children and adolescents, 3819

- for obsessive-compulsive disorders, 1844–1845
- obsessive-compulsive disorders, 3131
- panic disorder, 3131
- for paraphilic disorders, 2099
- paroxetine, 3137
- for pediatric OCD, 3843–3844
- pharmacodynamic interactions, 3139
- pharmacodynamics, 3131–3132
- pharmacokinetics of, 3139–3140
  - sex differences in, 3416
- for poststroke depression, 499
- for post-traumatic stress disorder, 3931
- post-traumatic stress disorder, 3131
- pregnancy and, 2604–2605, 3142–3143
- premenstrual dysphoric disorder, 3131
- for selective mutism, 3866
- for self-injurious behaviors, 3651
- sertraline, 3136–3137
- for sexual dysfunctions, 2054
- side effects, 3819
  - bleeding, 3133
  - cardiovascular effects, 3134
  - daytime somnolence, 3134
  - diabetes, 3134
  - discontinuation syndrome, 3133
  - emotional blunting, 3134
  - extrapyramidal symptoms, 3135
  - fractures, 3135
  - gastrointestinal effects, 3132–3133
  - hyponatremia, 3134
  - manic switch, 3135
  - sexual dysfunction, 3133

    - sleep disturbances, 3134
    - sweating, 3133
    - weight gain, 3134
  - social anxiety disorder, 3131
  - structures of, 3132*f*
  - suicidality and, 3142
  - therapeutic indications, 3130–3131, 3130*t*
  - for tic disorders, 3785–3786
  - tolerability of, 3132
  - transplant medication interactions, 2284
  - trichotillomania, 2435
  - use in children, 3144
- Selectivity, 69
- Selegiline (Eldepryl), 3185*t*, 3189
  - adverse effects, 3189
  - dosage, 3189
  - lactation and, 3189
  - pharmacodynamics, 3189
  - pharmacokinetics, 3189
  - pregnancy and, 3189
  - transdermal, 3189
- Self, 809
  - as agent, 429–430
  - clinical perspective, 422
  - defined, 4762
  - disturbances of, 1180
  - DMN as neutral substrate of, 820–821
  - evolutionary perspective, 420–421
  - experience of, 819–821
  - human developmental perspective, 421–422
  - mental state attribution, 427–428
  - multidimensionality of, 419–422

  - neuropsychiatry of, **416–431**
    - brain networks, 423–424
    - historical perspective, 418–419
    - neurologic syndromes, 425–426
    - psychedelics, 431–432
    - psychiatric disorders, 424–425
    - schizophrenia, 424–425
    - split-brain studies, 426
  - nondualistic notions of, 430–431
  - nonself *versus*, 426–427
  - other *versus*, 427–428
  - paleoneurologic perspective, 420–421
  - perspective-taking, 427–428
  - protoself, 420
  - self-representation, 423–424
  - self-transcendence, 430–431
  - self-transformation, 423–424
  - sense of agency, 429–430
  - voluntary action, 430
- Self-actualization, 2887, 4762
- Self-analysis, 4762–4763
- Self-as-object, 420, 421
- Self-awareness, 2213, 4763
- Self-defeating personality disorder, 4763
- Self-discovery, 4763
- Self-esteem
  - in adolescents, 3511–3512
  - impairment in regulation, 1704
  - loss of, 1654
- Self-fulfilling prophecy, 4763
- Self-harm
  - epilepsy and, 528

  - group psychotherapy for, 3903–3904
- Self-hatred, 957
- Self-help groups, 1126, 1310–1311, 2958, 4763
  - for dissociative identity disorder, 2000
- Self/identity-state dissociation, neurobiology of, 1957–1959
- Self-image, 4763
- Self-injurious behaviors (SIBs)
  - intellectual disabilities and, 3647, 3651
  - pharmacotherapy, 3651, 3813
    - antipsychotics, 3651
    - mood stabilizers, 3651
    - naltrexone, 3362
    - NMDA antagonists, 3651
    - opioid antagonists, 3651
    - selective serotonin reuptake inhibitors, 3651
  - prevalence of mental disorders, 3809
  - stereotypic movements and, 3809, 3812
- Self-Injurious Thoughts and Behaviors Interview (SITBI), 2716
- Self-injury, dissociation and, 1961
- Self-monitoring devices, 4467, 4467*t*
- Self-mutilation, 1191
- Self-object functions, 2752
- Self-objects, 422
- Self-observation, 2423, 2423*t*
- Self-organization
  - brain, 801
  - mind, 805–809
  - of open biologic systems, 801
- Self-organizational processes, 733–734
- Self-other configuration, 2751
- Self-protective processes, 2131–2132
- Self-psychology, 926–927

  - therapeutic actions, 2752
- Self-reference, 4763
- Self-reflective capacity, 732
- Self-representation, 423–424, 814, 817–818
- Self-schema, 819–820
- Self-states, 1972
  - direct therapeutic engagement with, 1990
  - substrates shaping, 1974
- Self-system, 4763
- Self-transcendence, 430–431, 2214–2216
- Self-transformation, 423–424
- Seligman, Martin E., 789, 795, 2887, 2897
- Selye, Hans, 2412
- SEM. *See* Standard error of the mean (SEM)
- Semai tribe, 2512
- Semantic dementia, 1021*t*
- Semantic memory, 777
- Semantic networks, 2553
- Semantic priming, 1562
- Semaphorins, 66
- SEMI (Short Explanatory Model Interview), 2570–2571
- Semiotic function, 746
- Semiparametric regression method, 909
- Sen, Ganpath, 3006
- Senescence, 4763
- Senile dementia, 4763
- Senium, 4763
- Sensate focus exercises, 2050
- Sensate focus learning, 4763
- Sensation
  - abnormalities of, 470
  - defined, 4763

  - definition of, 725
  - information processing and, 728–729
  - perception and, 729
  - in psychiatric disorders, 734–743
- Sensation–intuition polarity, 951
- Sense of self, 819–821
  - defined, 4763
  - loss of, 4600–4601
- Sensitivity, 881, 4763
- Sensitivity training group, 4763
- Sensitization, 2480*t*, 3013
- Sensorimotor period, 4763
- Sensorium, 4135, 4763
- Sensors, 4467, 4467*t*
- Sensory aphasia, 4763
- Sensory deprivation, 4763
- Sensory development, 3502–3503
- Sensory extinction, 4763
- Sensory food aversions, 3764–3766
  - clinical features, 3764
  - comorbidities, 3831–3832
  - crisis, 3765
  - diagnosis, 3764
  - differential diagnosis, 3764–3765
  - etiology, 3765
  - treatment, 3765
- Sensory inattention. *See* Sensory extinction
- Sensory integration therapy (SIT), 3679
- Sensory/motor functioning, assessment, 3570
- Sensory-motor network, 804
- Sensory preconditioning, 760
- Sentence completion test, 1045*t*, 1047

- Sentence Completion Test (SCT), 4763
- Sentencing, 4417
- Separation anxiety, 3851
  - clinical features, 3852
  - comparative nosology, 3851
  - course and prognosis, 3853
  - defined, 4763
  - definition of, 3851
  - diagnosis, 3852
  - differential diagnosis, 3853
  - DSM-5-TR criteria for, 3851*t*
  - epidemiology of, 3851
  - etiology of, 3851–3852
  - rapprochement, 922
- Separation anxiety disorder (SpAD)
  - behavioral approach tasks for, 3888*t*
  - comparative nosology, 1788
  - course, 1789
  - defined, 4763
  - diagnosis, 4794*t*
  - differential diagnosis, 1789
  - epidemiology of, 1789
  - etiology of, 1789
  - history, 1788
  - prevalence of, 1792
  - treatment of, 1789, 1838–1839
- Separation-individuation, 922–923, 4763
- Septal area, 36
- Septic abortion, 2590*t*
- Sequence variation, 242–243
- Sequential sampling, 323

- Sequential Treatment Alternatives to Relieve Depression (STAR*D) trial, 3007
- Serine deficiency syndromes (SDS), 696–697
- Serious mental illness
  - delirium and, 1234
  - recovery from, 4472
- Seroquel. *See* Quetiapine
- Serotonergic hallucinogens, 1353–1355
  - classic hallucinogens, 1353
  - intoxication, 1358
- Serotonergic system, 1800–1801
  - neuroimaging, 402–403
- Serotonin (5-HT), 100–104
  - 1A/2A receptors, 1543
  - aggression and, 2518
  - alcohol and, 1298
  - anatomy and functions, 100
  - anxiety disorders and, 1800–1801
  - appetite and, 228
  - chemical structure of, 3386*f*
  - defined, 4763
  - eating disorders and, 232–233
  - immunological effects of, 179
  - metabolism, 101
  - neuroimaging, 402–404
  - neuroimaging studies, 1542–1543
  - occupancy, 1548–1549
  - receptor availability, 402
  - release, 403
  - schematic model of synapse, 102*f*
  - storage capacity, 403
  - synthesis, 100–101, 100*f*, 403

  - transporters, 101
- Serotonin$_{1A}$ agonists, **3168–3174**
  - agitation, 3172
  - alcohol dependence, 3172
  - attention-deficit/hyperactivity disorder, 3173
  - blood concentration, 3170
  - cannabis dependence, 3172
  - chemical structures of, 3168*f*
  - cholinergic system, 3170
  - dopaminergic system, 3169
  - drug interactions, 3173–3174
  - GABAergic system, 3169–3170
  - generalized anxiety disorder, 3170
  - major depressive disorder, 3172
  - metabolism and elimination of, 3169
  - norepinephrine system, 3169
  - obsessive-compulsive disorder, 3170
  - organ/system effects of, 3173
  - panic disorder, 3170
  - pharmacodynamics, 3169
  - pharmacokinetics, 3168–3169
  - pharmacologic actions, 3168
  - post-traumatic stress disorder, 3170
  - precautions/adverse reactions, 3173–3174
  - premenstrual dysphoric disorder, 3173
  - psychosis, 3172
  - serotonin system, 3169
  - sexual dysfunction, 3172–3173
  - sleep apnea, 3173
  - smoking cessation, 3172
  - somatic dysfunction, 3172
  - therapeutic indications, 3173

  - withdrawal symptoms, 3173
- Serotonin-2 antagonists and reuptake inhibitors, 3416
- Serotonin antagonists/inverse agonists, **3215–3218**
- Serotonin-dopamine antagonists (SDAs), 4763. *See also* Second-generation antipsychotics (SGAs)
- Serotoninergic systems, 1762–1764, 1762*f*
- Serotonin-norepinephrine reuptake inhibitors (SNRIs), 97, **3144–3151**
  - for anxiety disorders, 3859
    - in children and adolescents, 3927
  - anxiety disorders, 1832–1834
  - for anxiety disorders in later life, 4304
  - for chronic pain, 2488
  - defined, 4763
  - for depression, in cancer patients, 2356
  - for descending neuropathic pain, 218
  - desvenlafaxine, 3145–3148
  - duloxetine, 3148–3150
  - for geriatric depression, 4238
  - levomilnacipran, 3150
  - for major depressive disorder, 1722–1723
  - milnacipran, 3150
  - mirtazapine and, 3212
  - pain management in burn patients, 2470
  - pharmacokinetics of, sex differences in, 3416
  - pregnancy and, 2604–2605
  - structures of, 3145*f*
  - venlafaxine, 3145–3148
  - weight gain and, 2309
- Serotonin reuptake transporter (SERT), 402–403
- Serotonin syndrome, 663–664, 1196
  - defined, 4763

  - monoamine oxidase (MOA) inhibitors, 3191, 3191*t*
  - vilazodone, 3178
  - vortioxetine, 3183
- Serotonin transporters, 1543
- SERT. *See* Serotonin reuptake transporter (SERT)
- Sertindole, 3056
- Sertraline (Zoloft), 3136–3137, 4663
  - for anxiety disorders in later life, 4224–4225
  - chemistry, 3136
  - for child/adolescent PTSD, 3849
  - dosing and administration, 3141
  - pharmacokinetics of, sex differences in, 3416
  - pharmacologic actions, 3136–3137
  - for tic disorders, 3785
- Serum level. *See* Blood level
- Serum levels. *See* Blood serum levels
- Serzone. *See* Nefazodone (Serzone)
- Seth, Anil, 418
- Setoperone, 3055
- Severe mental illness (SMI)
  - criminalization of, **4571–4585**
  - obesity and, 2308–2309
  - phases of, 4544–4550
    - multidimensional framework, 4545*f*
    - prodromal phase, 4546–4548
      - continuity of care, 4547–4548
      - early intervention and relapse prevention, 4546–4547
      - elapsed, 4546*t*
      - interventions, 4546*t*
    - recovery phase, 4550
    - stabilization phase, 4548–4549
    - stable phase, vocational rehabilitation, 4549–4550

  - substance use disorder and, 4557
- Severe mental retardation. *See* Mental retardation
- Severe stress, 181
- Sex assigned at birth, 2101, 4100
- Sex differences. *See also* Gender
  - in autism spectrum disorders (ASDs), 3713–3714
  - biological, 2032
  - in desire, 2034–2035
  - in erotic stimuli,’ 2034–2035
  - genetic, 2032
  - hormonal, 2032
  - in substance use disorders, 454–455
    - dopamine signaling and, 455
    - gonadal hormones and, 454–455
- Sex education, 4556–4557
- Sex history, 2038, 2039*t*
- Sex offenders, castration of, 2100
- Sex organs
  - female anatomy, 2028–2029
  - innervation of, 2029, 2029*t*
  - male anatomy, 2028
- Sex ratio, autism spectrum disorders (ASDs), 3713–3714
- Sex reassignment surgery (SRS), 2102
- Sexting, 3950
- Sex trafficking, 3978
- Sexual aberration, 4763
- Sexual abuse
  - child/adolescent
    - evaluation of, 3976
    - group psychotherapy for, 3905
    - intrafamilial, 3979
    - as maltreatment, 3969, 3971–3972

    - parent interviews, 3973–3974
    - perpetrators, 3972
    - perpetrator substitution, 3977
    - physical examination, 3975–3976
  - family therapy for, 3911*t*
  - older adults, 4433
- Sexual addiction, **2121–2145**
  - clinical examples, 2124–2125
  - comorbidities, 2126
  - course of illness, 2126
  - diagnosis of, 2048
  - differential diagnosis, 2125
  - DSM-5 criteria for, 2123–2124
  - epidemiology of, 2125–2126
  - familiar patterns, 2126
  - prognosis, 2145
  - signs of, 2048*t*
  - terminology, 2122
  - treatment of, 2134–2139
    - cognitive behavioral therapy, 2136
    - dialectical behavior therapy, 2136–2137
    - endocrinologic agents, 2138
    - support groups, 2138–2139
- Sexual arousal disorders, 4763
- Sexual aversion disorder, 4763
- Sexual behavior, 2036
  - adolescence, 2036, 3513
  - adulthood, 2036–2037
  - age-related changes in, 4440–4443
  - animal models of, 303
  - behavior mastery training, 2135–2136
  - cognitive capacity for decisions regarding, 4454

  - compulsive, 2048, 2048*t*
  - disturbances in, of sexually abused children, 3971
  - first coitus, 2036
  - in long-term care, 4452–4455
  - middle age, 2037
  - old age, 2037
  - risk management, 2135
  - selection of, 2133–2134
  - sexual learning and, 2036–2037
  - symptom containment, 2135–2136
  - treatment of, symptomatic behavior management, 2134–2139
  - urge handling, 2135
- Sexual delusion, 4724, 4763
- Sexual desire disorder, 4763
- Sexual deviation. *See* Paraphilias
- Sexual drive, 4763
  - disturbances of, 1189
- Sexual dysfunction(s), 1207, 1675–1676, **2027–2056**, 2585
  - assessment, 3432
    - in older patients, 4446–4447
  - buspirone for, 3172–3173
  - classification of, 3430*t*
  - defined, 4764
  - definition of, 2038
  - delayed ejaculation, 2043, 4815–4816*t*
  - diagnosis, 4815–4820*t*
  - DSM-5 criteria for, 2038, 3430
  - due to a general medical condition, 2044–2046
    - dyspareunia, 2045
    - hypoactive sexual desire and arousal/interest disorder, 2045–2046
    - male erectile disorder, 2044–2045, 2045*t*

  - other female sexual dysfunction, 2046
  - other male sexual dysfunction, 2046
- early ejaculation, 2043–2044
- effect of recreational/illegal substances on, 3432*t*
- erectile disorder, 3433–3435, 4816*t*
- erectile dysfunction, 3431
- female orgasmic disorder, 4817*t*
- female sexual arousal disorder, 3433
- female sexual interest/arousal disorder, 2040–2041, 2041*t*, 4817–4818*t*
- genitopelvic pain/penetration disorder, 2044, 4818–4819*t*
- hypnosis for, 2937
- hypoactive sexual desire disorder, 3431, 3433
- ICD-11 criteria for, 2038
- intellectual disabilities and, 3647–3648
- male erectile disorder, 2041*t*
- male hypoactive sexual desire disorder, 2038–2040, 2040*t*, 4819–4820*t*
- in older adults
  - arousal phase problems, 4448
  - assessment, 4446–4447
  - desire phase problems, 4447–4448
  - illnesses associated with, 4448, 4449–4451, 4449*t*
  - medications associated with, 4451
  - treatment, 4447–4452
- orgasmic disorders, 2042–2043
- other specified, 2047
- pathophysiology of, 3431–3432
- pharmacologic treatment of, **3430–3436**
- premature ejaculation, 3435, 4819–4820*t*
- prevalence of, 3431
- psychotropic-induced, 3435–3436

    - rating scales, 2038
    - substance/medication-induced, 2046–2047
    - treatment of, 2050–2056
        - behavioral exercises, 2050
        - behavior therapy, 2051
        - dual-sex therapy, 2050–2051
        - goals of, 2051
        - group therapy, 2051–2052
        - hormone therapy, 2054–2055
        - hypnotherapy, 2051
        - integrated sex therapy, 2052
        - pharmacotherapy, 2052–2055
    - unspecified, 2048–2050
    - vortioxetine, 3182
- Sexual fantasies, 2049, 2049*t*
- Sexual function
    - age-related changes in, 4440–4443, 4440*t*
    - neurobiology of, 4439–4440
    - strategies for enhancing in late life, 4442*t*
- Sexual history, obtaining, 4445–4446
- Sexual hyperaethesia, 2048
- Sexual identity, 2032, 2058, 4764
- Sexuality, 2585. *See also* Bisexuality; Heterosexuality; Homosexuality; Hypersexuality; Psychosexuality
    - in adults over the age of 65, 4162–4163
    - aging and, 4438–4457
        - age-related changes in sexual function and behavior, 4440–4443, 4440*t*
        - cognitive capacity and sexual behaviors, 4454
        - frequency of sexual expressions in, 4441–4443
        - LGBTQ+, 4443–4445, 4453–4454
        - recommendations for PRC communities, 4454–4455

        - sexual behavior in long-term care, 4452–4455
        - sexual dysfunctions
            - arousal phase problems, 4448
            - assessment, 4446–4447
            - desire phase problems, 4447–4448
            - illnesses associated with, 4448, 4449–4451, 4449*t*
            - medications associated with, 4451–4452
            - orgasm phase problems, 4448
            - treatment, 4447–4452
        - sexual history, 4445–4446
    - in cancer patients, 2370–2371
    - cultural mores and, 2056
    - endocrinology of, 2029–2030
    - epilepsy and, 528
    - female anatomy, 2028–2029
    - genetics of, 2030
    - law and, 2038
    - male anatomy, 2028
    - normal, **2027–2056**
    - sexual response cycle and, 4438–4439
    - survivors of, 2371–2372
- Sexualization, 921*t*, 4764
- Sexual learning, 2036
- Sexually dimorphic nucleus of the preoptic area (SDN-POA), 2069
- Sexually transmitted diseases (STDs), 1053
- Sexually transmitted infections (STIs), 2595
- Sexual masochism. *See* Masochism
- Sexual masochism disorder, 2095–2096, 4839*t*
- Sexual minorities, 2058
- Sexual minority stress, 2077
- Sexual orientation, 2033, 4021
    - definition of, 2058

  - interaction with stage of development, 2078–2080
- Sexual orientation conversion efforts (SOCE), 2087, 2112
- Sexual orientation distress, 4764
- Sexual pain disorders, 4764
- Sexual preference, 2058
- Sexual reassignment. *See* Transsexualism
- Sexual response cycle, 4438–4439
- Sexual sadism. *See* Sadism
- Sexual sadism disorder, 2096, 4840*t*
- Sexual selection, 865
- SGAs. *See* Second-generation antipsychotics (SGAs)
- Shaman, 2672, 4764
- Shamanic healing, 2648–2649
- Shame, 938
  - defined, 4764
  - dissociative identity disorder and, 1973–1974
- Sham eating, 222–223
- Shankara, 969
- Shaping, 2816, 4764
- Shapiro, Frances, 1134
- Shared paranoid disorder. *See* Shared psychotic disorder
- Shared psychotic disorder, 1645, 1647, 2020, 4764
- Shawcross, Arthur, 2522
- Shazer, Steve de, 2852
- Sheehan, Peter, 2914
- Sheehan Disability Scale (SDS), 1721
- Shell shock, 1929, 2500, 2656, 4657, 4764. *See also* Post-traumatic stress disorder (PTSD)
- *Shenjing shuairuo*, 2553, 2562–2565
  - clinical features, 2564–2565
  - cultural factors, 2565
  - epidemiology of, 2563–2564, 2563*t*

  - phenomenology of, 2563
  - precipitants/etiology, 2564
  - psychiatric diagnoses and, 2564
  - treatment of, 2565
- Shenk, Joshua Wolf, 2879
- Shepherd, Michael, 4659
- Shift work circadian rhythm disorder, 4764
- Shock treatment, 4764
- Shor, Ronald, 2914
- Short association fibers, 13
- Shorter, Edward, 2271
- Short Explanatory Model Interview (SEMI), 2570–2571
- Short-Form 36 (SF-36), 1098
- Short interfering RNAs (siRNAs), 2229
- Short interspersed elements (SINEs), 241
- Short Michigan Alcoholism Screening Test–Geriatric version (SMAST-G), 4281, 4281*t*
- Short Physical Performance Battery (SPPB), 4179
- Short sleeper, 2170
- Short tandem repeats (or STRs), 242
- Short-term anxiety-provoking psychotherapy (STAPP), 2765–2767
  - clinical issues, 2766
  - complications, 2766
  - contraindications, 2766
  - dynamic formulations, 2767
  - indications for, 2766
  - limitations of, 2766
  - outcomes of, 2767
  - phases of, 2766
  - process and goals of, 2767
  - selection criteria, 2766
  - techniques, 2765

  - theoretical issues, 2765
- Short-term dynamic psychotherapies (STDPs), **2763–2783**. *See also* Intensive short-term dynamic psychotherapies (ISTDPs)
  - accelerated experiential dynamic psychotherapy, 2772
  - clinical issues, 2768–2769
  - clinical spectrum of, 2765–2772
  - complications, 2769
  - indications for, 2768
  - limitations of, 2768–2769
  - of narcissistic and other self-disorders, 2773–2778
  - outcomes of, 2769
  - Panic-focused psychoanalytic psychotherapy, 2772–2773
  - short-term anxiety-provoking psychotherapy, 2765–2767
  - techniques, 2767–2769
  - theoretical issues, 2767–2768
- Short-term explicit memory network, 804
- Short-term memory, 4764
- Short-time Fourier transform (STFT), 311, 366
- Shorvonn, Harold, 1931
- SHSS. *See* Stanford Hypnotic Susceptibility Scale (SHSS)
- Shy–Drager syndrome. *See* Multiple system atrophy (MSA)
- Shyness disorder. *See* Avoidant disorder of childhood or adolescence
- Sialorrhea, 3070
- Sibling, 4764
- Sibling relational problem, 2543–2544
  - clinical features, 2543
  - comorbidity, 2543–2544
  - definition of, 2543
  - diagnosis, 2543
  - differential diagnosis, 2543–2544
  - etiology of, 2543

- Sibling rivalry, 4764
- Sibling studies, ADHD, 3729
- Sibling transference, 2749
- Sib pair analysis, 256
- Sickness history, 2554
- Sick role, 4764
- Side effect, 4764
- Sierra, Mauricio, 1933, 1954
- Siever, Larry, 1656
- Sifneos, Peter, 2765–2767
- Sighing dyspnea, 2323–2324
- Sign, 4764
- Signal anxiety, 2757, 4764
- Signal transduction, 145–147, 146*f*
- Signal transduction genes, 125
- Significant differences. *See* Statistical significance
- Sign tracking, 757, 757*f*
- Sigurdsson, Björn, 631
- Sildenafil (Viagra), 149, 3434*t*
  - adverse effects of, 2053
  - for erectile dysfunction, 2052–2053, 3433
  - use in women, 2054
- Silence, 1000
- Silent Generation, 4444
- Silent synapses, 86
- Silexan, 4144
- Silk Road, 1127
- Simeon, Daphne, 1933
- Simeon, Jovan, 4664
- Similarity, law of, 434
- Simmel, Ernst, 2912
- Simon, Daphne, 2016–2017

- Simons, Raymond, 2552
- Simple abstractions, 745
- Simple phobia, 4764
- Simple sequence length polymorphisms (SSLPs), 255
- Simple sequence repeats (SSRs), 241
- Simple tandem repeats (STRs), 255
- Simpson, George, 4651
- Sims, James, 1230
- Simulation, 2493
- Simulation-oriented therapies, for dementia, 1248
- Simultanagnosia, 474, 4764
- Sinequan. *See* Doxepin (Sinequan)
- Singer, Margaret T., 2020
- Single-nucleotide polymorphisms (SNPs), 242, 255, 1483, 2519, 3587*t*
- Single nucleotide variants, 2615
- Single-photon emission computed tomography (SPECT), 477
  - gamma-ray detection, 387–388, 388*f*
  - photon detection by collimation in, 388
  - physical principles of, 387–388
  - radionuclides, 386*t*, 387
  - spatial resolution, 392
- Single-session family consultation (SSFC), 2860–2861
- Single-session therapy (SST), **2856–2862**
  - assessment, 2861
  - clinical supervision, 2861
  - context of competence, 2858
  - couple/family conflict, 2860
  - early sessions, 2857
  - examples of, 2859–2861
  - future of, 2861–2862
  - history, 2855–2856

  - with individuals, 2859–2860
  - Kaiser study, 2857
  - one-at-a-time basics, 2858
  - ongoing medication management, 2860
  - practice ideas, 2859
  - risk management, 2861
  - serial (or sequential) single sessions, 2860
  - single-session family consultation, 2860–2861
  - single-session thinking, 2857–2858
  - temporal structure of sessions and therapy, 2858–2859, 2858*f*
  - unplanned, 2857
- Single-session thinking, 2856, 2857–2858
- Single session work (SSW), 2857
- siRNAs, 239
- Sirolimus, neurotoxicity of, 2465*t*
- Sitophobia, 4764
- Sjogren syndrome, 105
- Skin
  - circadian rhythms and, 2446
  - physical examination, 466
  - sleep and, 2446
- Skin disorders, psychosocial impact of, 2430–2431
- Skinner, B. F., 4764
  - radical behaviorism, 970
  - on reinforcement, 763
  - work of, 3885, 4691
- Skin-picking disorder. *See* Excoriation disorder
- Slavson, Samuel, 2956
- SLD. *See* Specific learning disorder (SLD)
- Sleep, **199–214**. *See also* Insomnia; NREM (nonrapid eye movement) sleep; REM (rapid eye movement) sleep; Slow-wave sleep (SWS)

- age and, 202–203
- aging and, 4210
- alcohol and, 1299
- animal models of, 302
- brain restitution and, 213
- cardinal rules of, 4211*t*
- circadian rhythms and, 196, 196*f*
- cytokines, immunity, and, 2418
- daytime sleepiness measurement, 203–204
- definition of, 199–200, 4764
- deprivation of
  - cognitive impairment and, 209–210
  - effects of, 209–210
- disruption, neurocognitive effects of, 4083
- disturbances of, 1187–1188
  - dementia and, 1245, 1251
- EEG patterns, 370
- evaluation methods, 4083–4084
- fragmentation, 1299
- functions of, 213
- homeostatic regulation of, 209–210
- hygiene, 4211*t*
- immune-brain interactions and, 192–193
- insomnia, 302
- learning and, 213
- memory and, 213, 772, 773*f*
- monitoring, 203–204
- mood disorders and, 1772
- narcolepsy, 302–303
- natural sleep, 302
- neurobiology of, 205–208
- in neurocognitive disorders, 4214–4215

  - in nursing homes, 4215
  - in older adults, 4135
  - organization of, 201
  - periodic limb movements of, 3110–3111
  - physiology of, 204–205
  - psychiatry and, 213–214
  - PTSD and, 1880–1881
  - regulation of, 208–209
  - skin and, 2446
  - slow-wave, 200, 772
  - stages of, 200–201, 200*t*
    - EEG, 201*f*
  - standardized visual scoring, 203
- Sleep apnea, buspirone for, 3173
- Sleep deprivation, 371
- Sleep diary, 2189
- Sleep-disordered breathing (SDB), 4084–4085
- Sleep disorders, **2165–2193**
  - breathing-related, 4814*t*
  - caffeine-induced, 1332
  - cases, 2189
  - child/adolescent, **4083–4089**
    - breathing, 4084–4085
    - circadian rhythm sleep–wake disorders, 4087
    - insomnia, 4085
    - narcolepsy, 4086
    - nocturnal enuresis, 4087
    - parasomnias, 4086–4087
    - somnambulism, 4086
  - in children with psychiatric disorders, 4087–4089
    - ADHD, 4087–4088
    - anxiety disorders, 4089

    - autism spectrum disorder, 4088
    - mood disorders, 4088–4089
  - circadian rhythm sleep–wake disorders, 2180–2181, 4812–4813*t*
  - classification of, 2165, 4084–4087
  - defined, 4764
  - DSM-5 criteria for, 4085*t*
  - hypersomnolence disorder, 2172–2174, 4811*t*
  - insomnia disorder, **2165–2172**, 4811*t*
  - intellectual disabilities and, 3651–3652
  - multiple sclerosis and, 581
  - narcolepsy, 2174–2175, 4812*t*
  - nightmare disorder, 4814–4815*t*
  - non-REM-sleep behavior disorder, 4813*t*
  - in older adults, **4210–4215**
    - circadian rhythm sleep–wake disorders, 4212–4213
    - consequence of poor sleep, 4210
    - insomnia disorder, 4211–4212
    - obstructive sleep apnea-hypopnea syndrome, 4213–4214
    - REM sleep behavior disorder, 4214–4215
    - restless legs syndrome/periodic limb movements in sleep, 4214
    - sleep-wake disorders, 4210–4211
  - overview, 2165
  - parasomnias, 2181–2188
  - REM sleep behavior disorder, 4813*t*
  - restless legs syndrome, 4815*t*
  - sleep-related breathing disorders, 2175–2181
  - substance/medication-induced, 2188
  - treatment of, 3652
- Sleep efficiency, 203
- Sleep enuresis, 2185
- Sleep homeostat, 196
- Sleep latency, 203

- Sleep medicine
  - tools in, 2188–2190
    - actigraphy, 2189
    - clinical interview, 2188
    - Epworth Sleepiness Scale, 2189
    - home sleep apnea testing, 2189
    - Insomnia Severity Index, 2189
    - maintenance of wakefulness test, 2189
    - multiple sleep latency test, 2189
    - polysomnography, 2188–2189
    - sleep diary, 2189
    - STOPBANG, 2189
- Sleep paralysis, 2185
- Sleep rebound, 209
- Sleep-related breathing disorders, 2175–2181
  - central sleep apnea, 2179
  - obstructive sleep apnea hypopnea, 2176–2179
  - overview, 2175–2176
  - sleep-related hypoventilation, 2179–2180
- Sleep-related bruxism, 2186, 2188*f*
- Sleep-related eating disorder, 2183–2185, 4764
- Sleep-related hallucinations, 2185, 4764
- Sleep-related hypoventilation, 2179–2180, 4764
- Sleep-related leg cramps, 2186
- Sleep-related movement disorder
  - due to a medical disorder, 2187
  - rhythmic, 2186–2187
  - unspecified, 2187
- Sleep restriction therapy, 2171
- Sleep starts (hypnic jerks), 2188
- Sleep talking, 2187
  - in children/adolescents, 4086

- Sleep terror disorder, 1188, 4764
- Sleep terrors, 2182
  - in children/adolescents, 4086
  - polysomnographic tracing, 2182*f*
- Sleep-wake disorders, 1206
  - comorbidities, 4210
  - in long-term care, 4413–4414
  - medications for, 4211
  - in older adults, 4210–4211
- Sleep–wake schedule disorder, 4764
- Sleepwalking, 1188, 2182
  - in children/adolescents, 4086
  - defined, 4764
  - diagnostic criteria, 2183
- Slit molecules, 65
- sLORETA. *See* Standardized low-resolution brain electromagnetic tomography (sLORETA)
- Slosson Drawing Coordination Test, 3676
- Slow euthanasia, 2409
- Slow-wave sleep (SWS), 200, 772, 4764
- Sluggish cognitive tempo (SCT), 2533–2534
- SLUMS. *See* St. Louis University Mental Status Examination (SLUMS)
- Small cell lung cancer (SCLC), 2363
- Small sequence repeats (SSRs), 242
- Smallwood, Jonathan, 423
- Smart homes, 4468, 4468*t*
- Smartphones, 1121, 1123
- Smith, Richard, 1897
- Smith–Lemli–Opitz syndrome, 3644
- Smith–Magenis syndrome, 3640*t*, 3644

- Smoking. *See also* Nicotine; Nicotine dependence; Tobacco use disorder (TUD)
  - in adolescents, 3999
  - in older adults, 4283–4284
  - schizophrenia association with, 1596
- Smoking cessation
  - bupropion for, 3165–3166
  - buspirone for, 3172
  - hypnosis and, 2937–2938
  - topiramate for, 3269
- Smooth muscles, naltrexone effects on, 3364
- S-nitrosylation pathway, 125, 126*f*
- Snoezelen, 2643
- Snowdon, David, 795
- SNRI. *See* Serotonin-norepinephrine reuptake inhibitors (SNRIs)
- Snuff, 1385
- Snus, 1385
- Snyder, C. Rick, 2898
- Snyder, Solomon, 3039, 4662
- Social (pragmatic) communication disorder, 3489, **3688–3693**. *See also* Communication disorders
  - course, 3692–3693
  - defined, 3688, 4765
  - development, 3692–3693
  - diagnosis, 3689–3690, 4775*t*
  - differential diagnosis, 3691
  - DSM-5-TR criteria for, 3689*t*, 3690
  - epidemiology of, 3692
  - etiology of, 3691–3692
  - historical background, 3688–3689
  - symptoms, 3689–3690
  - treatment, 3693

- Social adaptation, 4764
- Social anxiety disorder (SAD), **3853–3855**. *See also* Social phobias
  - assessment of, 3888*t*
  - behavioral approach tasks for, 3888*t*
  - behavioral inhibition and, 3852
  - bipolar disorder and, 1742
  - child/adolescent, 1820
  - clinical features, 1785, 3854
  - cognitive behavioral therapy for, 1829–1830, 3892–3893
  - comparative nosology, 1785, 3853
  - course, 1785
  - course and prognosis, 3855
  - definition of, 1785, 3853
  - diagnosis, 1785, 3854, 4795*t*
  - differential diagnosis, 1785, 3855
  - DSM-5-TR criteria for, 3854*t*
  - epidemiology of, 1785, 3853
  - etiology of, 1785, 3853–3854
  - history, 1785
  - MDMA-assisted therapy for, 2909
  - monoamine oxidase inhibitors, 3195
  - neuroimaging, 1816–1817
  - prevalence of, 1792
  - SSRIs for, 3131
  - treatment of, 1785–1786, 1837–1838, 3859
  - venlafaxine for, 3147
- Social anxiety-like behavior, 298–299
- Social atom, 2970
- Social attachment, 301
- Social Attribution Task-Multiple Choice (SAT-MC), 4551
- Social behaviors

  - evolutionary explanations, 861–862
  - kin selection and, 862
- Social Behavior Schedule (SBS), 4553
- Social breakdown syndrome, 1481, 4764
- Social capital, 4566
- Social class
  - anxiety disorders and, 1793–1794
  - autism spectrum disorders (ASDs) and, 3714
- Social cognition, 732–733, 4551
  - rehabilitation of, 4554
  - schizophrenia and, 1557–1558
- Social Cognition Psychometric Evaluation (SCOPE), 4551
- Social cohesion, 4566
- Social Communication Questionnaire (SCQ), 3546
- Social control, 4765
- Social development, 960
  - in adolescents, 3510–3511
  - attachment theories, 3500
  - in children, 3492–3493*t*, 3500–3501
  - morality, 3501
- Social dysfunction, trazodone, 3201
- Social engagement, 4474
- Social functioning, 4180
  - aggression, 302
  - animal models of, 302
  - social attachment, 301
  - social hierarchy, 302
- Social hierarchy, 302
- Social instinct. *See* Herd instinct
- Socialization
  - in adolescence, 3510–3511
  - defined, 4765

- Social justice, 4012
- Social learning, 4555
- Social learning theory, 970–971, 2814, 4765
- Social me, 420
- Social media, 2701–2703. *See also* Internet
  - adolescents and, 3514–3515, 3950
  - app engagement, 1128
  - child/adolescent illnesses and, 3992
  - and child and adolescent psychiatry, 4020–4021
  - older adults and, 4469
  - for youth, **4094–4099**
    - anxiety, 4096
    - benefits of, 4094–4095
    - body image distortion, 4096
    - clinical considerations, 4097
    - clinician guidance, 4099*t*
    - COVID-19, 4097
    - cyberbullying, 4096–4097
    - depressive symptoms, 4096
    - poor sleep, 4097
    - risks, 4095–4096
    - seeking mental health information, 4095
    - seeking support, 4095
    - self-harm, 4096
    - social connectedness, 4094
- Social Media Initial Questionnaire-revised (SMIQ-r), 4093, 4097, 4098*f*
- Social networks, 4566
- Social network therapy, 4765
- Social phobias, 1187. *See also* Social anxiety disorder (SAD)
  - cognition in, 742
  - defined, 4765

  - diagnosis, 4795*t*
  - ICD-11 criteria for, 1223
- Social psychiatry, 4648, 4765
- Social Security Act, 3956–3957
- Social services, 2350
- Social skills
  - group psychotherapy for, 3905–3906
  - training, 4555
- Social stressors, 1665–1666
- Social support, 1666, 4474
- Social therapy, 4765
- *Social Tolerance and Homosexuality* (Boswell), 2684
- Social withdrawal, 1459
- Social workers, 4029
- Sociatry, 2968
- Sociobiology, 4765
- Sociocultural framework, **4558–4570**
  - community system, 4565–4569
    - environmental level, 4565–4566
    - individual factors, 4566–4567
      - competing needs, 4567
      - general education, 4566
      - health literacy and numeracy, 4566–4567
      - negative provider stereotypes, 4567
    - social capital and social networks, 4566
    - social cohesion level, 4566
  - future directions, 4569–4570
  - health care systems
    - organization, 4563–4564
    - policy, 4561–4563
    - provider-level factors, 4564–4565
  - mechanisms producing inequities, 4567–4569

  - overview, 4558–4559
  - for service disparities, 4559–4561, 4560*f*
- Sociodrama, 2969–2970
- Socioemotional intelligence, 794
- Sociology, 4765
- Sociometry, 2968, 2970, 4765
- Sociopath, 4765
- Sociopathic behavior, 4765
- Sociopathy, vulnerability to, 871
- Sociotherapy, 4765
- Socratic questioning, 4364–4365. *See also* Cognitive behavioral therapy (CBT)
- Sodium (Na^+) channels, 71–73
- Sodium amytal, 1901
- Sodium spike, 70
- Sodomy, 4765
- Soft signs, 470
- Soldier's heart, 2659
- Soldier's heart. *See* Da Costa syndrome
- Solomon, Harry, 4511
- Solomon, Judith, 928
- Solution-focused brief therapy (SFBT), **2852–2862**
  - components of, 2853*f*
  - coping questions, 2854
  - core skills, 2852–2855
  - empirical evidence, 2852–2853
  - end of sessions, 2855
  - engagement compliments, 2853
  - goal negotiation, 2854–2856
  - positive differences, 2854
  - resource activation, 2853
  - scaling questions, 2855

  - VIPs, 2853
  - “yes-set”: “for you” statements, 2853–2855
- Solution-focused family therapy, 2949
- Solvents, 1370*t*
- Soma. *See* Carisoprodol (Soma)
- Soma (cell body), 3
- Soma, definition of, 4765
- Somatic cough disorder, 2324
- Somatic delusion, 4724, 4765
- Somatic hallucination, 4765
- Somatic markers, 726
- Somatic mutations, 52
- Somatic pain, 2388*t*
- Somatic symptom and related disorders, 1206, **1887–1905**
  - assessment instruments, 1905
  - attitudinal./behavioral factors, 1893
  - biologic factors, 1893
  - Briquet Syndrome, 1890
  - buspirone for, 3172
  - in children and adolescents, 1892
  - clinical features, 1893–1894
  - comorbidities, 1894
  - comparative nosology, 1890–1891
  - conversion disorder (functional neurologic symptom disorder), 1900–1902
  - course and prognosis, 1894–1895
  - definition of, 1887
  - diagnosis of, 1893–1894, 4805–4807*t*
  - diagnostic classification, 1905
  - differential diagnosis, 1894–1895
  - DSM-5 criteria for, 1887, 1891, 1891*t*
  - environmental and cultural factors, 1892–1893

  - epidemiology of, 1892*t*
  - etiology of, 1892–1893
  - factitious disorder, 1902–1905
  - functional somatic syndromes, 1891, 1892*t*
  - gender and, 1892
  - history, 1887–1888
  - hypnosis for, 2926–2927
  - hypochondria, 1888
  - hysteria, 1888
  - ICD-11 criteria for, 1891
  - illness anxiety disorder, 1898–1900
  - NIMH research domain criteria, 1905
  - psychological factors, 1893
  - psychological factors affecting other medical conditions, 1902
  - in sexually abused children, 3972
  - somatic presentations in classical psychopathology, 1888–1890
  - somatic symptom disorder, 1895–1898
  - somatization, 1890
  - somatoform, 1890
  - St. Louis criteria, 1890
  - suicide and, 2710
  - treatment of, 1895, 1904
  - unspecified and below-threshold, 1903
    - definition of, 1903
    - epidemiology of, 1903
    - treatment of, 1903–1904
- Somatic symptom disorder, 1895–1898, 2506
  - chronic pain, 2485
  - comparative nosology, 1896
  - defined, 4765
  - diagnosis of, 4805*t*
  - DSM-5 criteria for, 1896

  - epidemiology of, 1896–1897
  - etiology of, 1897
  - ICD-11 criteria for, 1896
  - with predominant pain, 2440
  - treatment of, 1897–1898
    - cognitive behavioral therapy, 1898
    - pharmacotherapy, 1897
    - psychotherapy, 1898
    - reassurance and supportive therapy, 1898
- Somatic therapies, 2642
  - animal-facilitated therapy, 2642
  - anxiety disorders, 1832–1839
  - defined, 4765
  - dissociative amnesia, 2012
  - exercise, 2642
  - expressive movement therapies, 2642
  - massage therapy, 2642
  - meditation, 2643
  - mind-body practices, 2642–2643
  - mindfulness training, 2642–2643
  - relaxation techniques, 2642
  - snoezelen, 2642
  - yoga, 2643
- Somatic thought, 996*t*
- Somatization, 921*t*
- Somatization disorder, 1896, 4765
- Somatoemotional release and craniosacral therapy, 2649
- Somatoform, 1890
- Somatoform disorders, 4765
  - in cancer patients, 2361
  - intellectual disabilities and, 3647–3648
- Somatoform Dissociation Questionnaire (SDQ-20), 1937*t*

- Somatoform Dissociation Questionnaire-5 (SDQ-5), 1937*t*
- Somatoform respiratory disorders, 2323–2324
  - chronic cough disorder, 2323–2324
  - chronic hyperventilation, 2323
  - sighing dyspnea, 2323–2324
  - vocal cord dysfunction, 2323–2324
- Somatomotor/sensorimotor network (SMN), 1757, 1758*f*
- Somatopagnosia, 4765
- Somatosensory-evoked potentials (SSEPs), 381
- Somatostatin, **167**
- Somatotropin, **167**
- Somatotype, 4765
- Somnambulism, 1188, 2917
  - in children/adolescents, 4086
  - defined, 4765
- Somnifen, 3231–3232
- Somniloquy, 4086
- Somnolence, 3081–3082, 4765
- Sonata. *See* Zaleplon (Sonata)
- Soteria, 1609
- Sound therapy, 2643–2644
- Sourkes, Theodore, 4662
- Southard, Elmer, 4653
- Southern California Sensory Integration Test, 3676
- Souvenaid, 4322
- Spasm, 3780*t*
- Spatial agnosia, 4765
- Spatz, Hugo, 4649
- Speaking in tongues, 4765
- Spearman, Charles, 2526, 4659
- Spearman-Brown formula, 1037
- Special education, 4035

  - definition of, 4026*t*
  - racism in, 4110
  - teachers, 4028
- Special K. *See* Ketamine
- Species-specific defensive reactions (SSDRs), 764–765
- Specific developmental disorder, see Specific learning disorders (SLD)
- Specificity, 881, 4765
- Specific learning disorder (SLD), 3656–3671
  - ADHD and, 3665
  - anxiety disorders and, 3665
  - borderline intellectual functioning and, 2532–2533
  - clinical features, 3663–3665
  - cognitive factors, 3661–3662
  - comparative nosology, 3659
  - course, 3670
  - defined, 4765
  - definition of, 3489, 3657–3658
  - diagnostic criteria, 3665–3669, 4778*t*
  - differential diagnosis, 3669–3670
  - DSM-5 criteria for, 3659, 3666*t*
  - emotional problems and, 3665
  - environmental factors
    - family, 3663
    - gene X, 3663
    - prenatal, 3662
  - epidemiology of, 3659–3660
  - etiology of, 3660–3663
  - genetics
    - behavioral, 3660–3661
    - molecular, 3661
  - history, 3658–3659

  - ICD-11 criteria for, 3659
  - in mathematics, 3661, 3664–3665
  - mood disorders and, 3665
  - neuroimaging, 3617
  - neurologic substrate, 3661
  - prognosis, 3670
  - in reading, 3661, 3664
  - risk factors, 3663*t*
  - significance for psychiatrists, 3656–3657
  - signs and symptoms, 3663*t*, 3664–3665
  - speech/language impairments and, 3665
  - standardized tests, 3668*t*
  - suicide risk and, 3665
  - terminology, 3656, 3656*t*
  - treatment of, 3671
  - in written expression, 3661, 3664
  - in written expressions, 3665
- Specific Levels of Functioning (SLOF), 4186, 4553
- Specific reading disability. *See* Developmental reading disorder
- Specific reading tests, 3567
- Specific relay nuclei, 23
- Specific speech articulation disorder, 3695
- Specific understanding, 2140–2141
- SPECT. *See* Single-photon emission computed tomography (SPECT)
- Spectatoring, 2050
- SPECTRA, 1101–1102, 1101*t*, 1102*f*
- Spectral editing, 346
- Spectrum/continuum model of temperament/psychopathology, 3624, 3624*t*
- Spectrum disorder hypothesis, 2257
- Speech

  - examination, 467–468
  - psychiatric examination and, 995
- Speech disturbance, 4765
- Speech-language therapists/pathologists (SLP), 4029
- Speechless terror, 739
- Speech sound disorder, **3694–3698**. *See also* Communication disorders
  - clinical features, 3696
  - comparative nosology, 3695
  - course and prognosis, 3697
  - defined, 4765
  - diagnosis, 3696, 4774*t*
  - differential diagnosis, 3696, 3696–3697
  - epidemiology of, 3695
  - etiology of, 3695–3696
  - laboratory examination, 3696
  - overview, 3694–3695
  - pathology of, 3696
  - treatment of, 3697–3698
- Speech syndromes, 467*t*
- Speech therapy, for selective mutism, 3866
- Sperry, Roger, 426, 4694
- Spiegel, David, 1932, 2915
- Spiegel, Eric, 2928
- Spiegel, John, 787
- Spielman, Arthur, 2171
- Spinal cord injury (SCI), 2484
- Spinal cord stimulation (SCS), 2491
- Spinal muscular atrophies (SMAs), 655
- Spinogenesis, 1509
- Spirit possession, 2671–2673
- Spiritual development, 430–431

- Spiritual emotions, 784, 797–798
- Spirituality, **2667–2685**. *See also* Religion
  - achieving, forms and methods of, 2669–2675
    - altruism, 2674–2675
    - body modification and mortification, 2674
    - confession, 2674
    - consciousness, 2669
    - drugs, 2673–2674
    - meditation, 2670
    - mysticism, 2670–2671
    - New Age spirituality, 2675
    - penance, 2674–2675
    - prayer, 2671
    - psychedelics, 2673
    - reconciliation, 2674
    - spirit possession, 2671–2673
    - spiritual site visits, 2674
    - yoga, 2675
  - assessment, 2678–2679
  - charity and, 2674–2675
  - in clinical practice, 2675–2685
    - problems with, 2679–2686
      - alien abductions, 2682
      - anomalous experiences, 2681
      - anxiety, 2679–2681
      - depression, 2679–2681
      - guardian angels, 2682
      - near-death experiences, 2681–2682
      - psychosis, 2679–2681
      - religious life, 2683–2684
      - specific spiritual therapies, 2683
      - spiritual life, 2682–2683

    - treatment issues, 2682
  - overview, 2667–2669
  - positive psychiatry and, 4473–4474
  - psychiatrists’ perceptions about, 2677–2678
  - religion, religiosity and, 2675–2677
  - spiritual assessment, 2678–2679
- Spiritual life, 2683
- Spiritual me, 420
- Spiritual site visits, 2674
- Spirochetes, 620–623
- Spitz, Rene, 3500, 4692, 4693
- Spitzer, Robert, 1201
- Splenium, 12
- Split-brain studies, 424–425
- Splitting, 2748*t*, 4765
- Spoken word recording, 1123
- Spontaneous abortion, 2590*t*
- Spontaneous recovery, 761
- Spousal bereavement, 4161
- Sprenger, James, 4683
- Spurzheim, J. C., 417, 4685
- Squeeze technique, 2050
- SSFC. *See* Single-session family consultation (SSFC)
- SSRIs. *See* Selective serotonin reuptake inhibitors (SSRIs)
- SST. *See* Single-session therapy (SST)
- Stagnation, 939–940
- Stahl, Georg Ernst, 4684
- Stammering, 1182, 4763
- Standard deviation (SD), 4765–4766
- Standard error of measurement (SEM), 1037–1038
- Standard error of the mean (SEM), 4766

- Standardized low-resolution brain electromagnetic tomography (sLORETA), 368, 368*f*
- Standardized score, 4766
- Stanford accelerated intelligent neuromodulation therapy (SAINT), 3472–3473
  - clinical studies, 3472
  - history, 3472
  - mechanism of action, 3472
  - side effects, 3472
- Stanford–Binet Intelligence Scale, 4766
- Stanford Hypnotic Susceptibility Scale (SHSS), 2912
- Stanford Sleepiness Scale (SSS), 204
- Stanford University, 4654
- Staniloiu, Angelica, 2003
- staring spells, 680
- Start codon, 237
- Startle reaction, 4766
- State-dependent learning, 761
- State-dependent learning or memory, 4766
- State of consciousness, 1959
- States of mind, 734
- State-Trait Anxiety Inventory for Children (STAIC), 3545
- Statistical inference, 4766
- Statistical significance, 4766
- Statistics, 4658–4659
- Statistics and experimental design, **895–913**
  - Bayesian approach
    - Bayesian interference, 901–902
    - Bayesian theorem, 901
  - conditional probability and independence, 896
  - cost-effectiveness, 912–913
  - covariance and correlation, 897

- effectiveness studies, 912
- efficacy studies, 911–912
- estimation of parameters, 897–898
  - confidence interval estimation, 897–898
  - confidence interval interpretation, 898
  - point estimators, 897
  - properties of estimators, 897
- hypothesis testing, 898–901
  - false discovery rate, 901
  - multiple hypotheses, familywise error rate, 899–901
  - power of test, 898–899
  - *p* values, 899
  - type I and type II errors, 898
- mean, median, and variance, 896–897
- multivariate random variables, 897
- parametric forms, 896
- probability density functions, 896
- probability theory, 896
- randomized clinical trials, 902–906
  - causality, 902, 904
  - likely responder method of analysis, 903
  - matching and randomization, 902–903
  - missing or incomplete data, 904–906
  - propensity scores, 903–904
  - randomization tests, 903
- specific statistical methods, 906–911
  - contingency tables, 908
  - general linear model, 908–909
  - growth mixture modeling, 909–910
  - latent class growth analysis, 909–910
  - logistic regression, 906–907
  - log linear models, 908

    - longitudinal models, 909
    - regression models, 906
    - supervised machine learning, 907–908
    - survival analysis, 910–911
  - study design, 911
    - crossover designs, 911
- Status epilepticus, 520, 4766
- Stauder, Karl Heinz, 4656
- Stein, Calvert, 2923
- Stekel, Wilhelm, 1890, 2091
- Stem cell transplantation, 2369
- Step-down procedure, 900
- Stepped care, 2740–2741
- Step-up procedure, 900
- Stereotypic movement disorders
  - in children, **3807–3813**
  - defined, 4766
  - diagnosis, 3812, 4779*t*
  - differential diagnosis, 3811
  - DSM criteria for, 3810–3811
  - epidemiology of, 3808–3810
  - etiology of, 3808
  - historical background, 3807–3808
  - ICD-10 and ICD-11 criteria for, 3811–3812
  - injurious behaviors, 3812
  - intellectual disabilities and, 3647
  - nosology, 3810–3811
  - treatment of, 3812–3813
- Stereotyping, 4564–4565
- Stereotypy, 3780*t*, 4766
- Stereotypy and habit disorder. *See* Stereotypic movement disorders

- Sterilization, 2594–2595, 2594*t*
- Stern, Daniel, 3877
- Stern, Karl, 4650
- Sternbach, Leo, 4663
- Steroid-induced psychotic disorder, 513–514
- Steroid(s). *See also* Anabolic-androgenic steroid-related disorders; Dihydrotestosterone (DHT); Estradiol; Estrogens; Hormone(s); Progesterone; Testosterone
  - actions, 3424*f*
  - neurosteroids, 137–139
  - psychotic disorders induced by, 513–514
  - structure, 3422–3424
  - synthetic pathways, 3423*f*
- Steven–Johnson syndrome (SJS), 2447
  - clinical features, 2447
  - diagnosis, 2447
  - treatment of, 2447
- Stevenson, Janine, 2930
- Stevenson, Robert Louis, 1935
- Stiff-person syndrome, 663
- Stillbirth, stillbirth, 2589
- Stilted speech, 4766
- Stimulant-related disorders, **1312–1323**
  - amphetamine-type stimulants, 1313
  - bath salts, 1314–1315
  - clinical features, 1317–1319
  - cocaine, 1312–1313
  - cognitive/executive dysfunction, 1316–1317
  - comorbidity, 1319
  - comparative nosology, 1313
  - compulsive use, 1316
  - course and prognosis, 1320–1321

- cyclic adenosine monophosphate (cAMP) and, 1316
- definitions of, 1312
- diagnosis, 1317–1319
- differential diagnosis, 1320
- DSM-5 criteria for, 1312, 1313, 1313*t*
- epidemiology of, 1314–1315
- epigenetics, 1316
- etiology of, 1315–1317
- gene transcription and, 1316
- history, 1312–1313
- ICD-11 criteria for, 1313, 1313*t*
- intoxication, 1318
- laboratory examination, 1320
- learning and, 1315–1317
- mechanism of action, 1315
- medical complications, 1319–1320
- memory and, 1315–1317
- metabolic factors, 1317
- microRNAs and, 1316
- negative reinforcement mechanisms, 1316
- neuroendocrine dysregulation and stress, 1316
- neurotransmitter systems, 1315
- pathology, 1320
- positive reinforcement mechanisms, 1315
- psychotic disorders and, 1319
- routes of administration, 1317
- stimulant use disorder, 1317–1318, 1318*t*
- toxicity, 1319–1320
- treatment, 1321–1323
  - immunization, 1323
  - intensity of, 1321
  - of intoxication and withdrawal, 1321

    - pharmacologic, 1322–1323
    - psychosocial, 1321–1322
  - unspecified, 1319
  - withdrawal, 1318–1319
- Stimulants, 3780*t*. *See also* Psychostimulants
  - for ADHD, 3733–3736, 3787, 3931–3932
  - adverse effects of, 3735–3736
  - agitation due to intoxication, 2724
  - amphetamine-type, 1313
  - for autism spectrum disorders (ASDs), 3727
  - behavioral effects of, 3735
  - cognitive effects of, 3735
  - defined, 4766
  - for depression in later life, 4298
  - efficacy, 3735
  - intoxication, 2726
  - pharmacology of, 3734
  - for repolarization abnormalities, 2289
  - for tic disorders, 3786–3787
- Stimulant use disorder, 1317–1318, 1318*t*
- Stimulus, 4766
- Stimulus control therapy, 2171, 2818
- St. John wort (*Hypericum perforatum*), 2639, 2639*f*, 4143
- St. Louis University Mental Status Examination (SLUMS), 1029
- St. Luke's Hospital (London), 4643
- Stockholm syndrome, 2020, 4766
- Stoller, Robert, 2062, 2101, 2104
- Stolorow, Robert, 2750, 3877
- Stone, Leo, 1704
- STOPBANG, 2189
- Stop codons, 237
- Stop signal task (SST), 1570

- Stop-start technique, 2050
- Storytelling, 2744
- Stott Test of Motor Impairment, 3676
- Stranger anxiety, 1186
- Strange situation, 928
- Strange situation paradigm, 2927
- Strategic family therapy, 2947–2948
- Strategic therapy, 2981–2983
  - clinical issues, 2982–2983
  - ethical issues, 2982–2983
  - history, 2981
  - techniques, 2982
  - theoretical issues, 2981
- Strattera. *See* Atomoxetine
- Strauss, John, 1457
- Street drugs, 4766
- Strengths, 789
- Strengths and Weaknesses of ADHD-Symptoms and Normal Behavior (SWAN) scale, 3546, 3732
- Strephosymbolia, 4766
- Stress. *See also* Post-traumatic stress disorder (PTSD)
  - acute, effects on immunity, 2418
  - affective disorders and, 2421
  - animal models of depression and, 1759
  - anxiety disorders and, 2421
  - in asthma, 2318–2319
  - autoimmune disorders, 2419
  - blood pressure and, 2413*f*
  - chronic, effects on immunity, 2418
  - of critical illnesses, 2427
  - defined, 4766
  - depression and, 1657, 2418

  - endocrine and autonomic responses of, 1812–1813
  - endocrine responses to, 2415–2416
  - epigenetics of, 278–279
  - history of literature on, 2412–2413
  - immunity and, 2418
  - measuring psychologically, 2414–2415
  - metabolic syndrome and, 2419
  - neurotransmitter responses to, 2415
  - personality traits and, 2421–2422
  - post-traumatic stress disorder (PTSD) and, 2421–2422
  - psychiatric illness and, 2419–2422
  - psychiatry and, **2412–2426**
  - psychoneuroimmunology, 2416–2422
  - racism and, 4105–4107
  - resilience to, 2413–2414, 2413*f*
  - severe/chronic, 181
  - vulnerability to, 2413–2414
- Stress immunity, 4766
- Stress inoculation training, 1827
- Stress management
  - cognitive-behavioral approaches to, 2422–2425
  - cognitive restructuring, 2423–2424, 2424*t*
  - pharmacologic approaches to, 2422
  - problem-solving, 2425
  - psychological first aid (PFA), 2425
  - relaxation training, 2424
  - self-observation, 2423, 2423*t*
  - time management, 2424–2425
  - training, 2422–2425
- Stressors
  - physiologic responses to, 2415–2416
  - psychological, 2573

  - psychosocial, 852
- Stress reaction, 4766
- Stria terminalis, 40
- Striatonigral degeneration. *See* Multiple system atrophy (MSA)
- Striatum, 3606–3607
- Striosomes, 29
- Stroke. *See also* Apoplexy
  - anxiety and, 488
    - assessment, 493
    - course and prognosis, 497
    - treatment of, 500
  - apathy and, 488–489
    - assessment, 492–493
    - etiology, 491, 493t
    - treatment of, 500
  - defined, 4766
  - definition of, 485
  - depression and, assessment, 492–493
  - incidence of, 486
  - mania and, 488
    - assessment, 492–493
    - course and prognosis, 497
    - etiology, 490–491, 491
    - treatment of, 500
  - in neuro-Covid, 189
  - psychosis and, 488
    - etiology, 491
    - treatment of, 500
- Strömgren, Eric, 4657, 4658
- Stroop Color-Word task, 4551
- Structural competency, 1138
- Structural dissociation theory, 1967

- Structural family therapy, 2947
- Structuralism, 745
- Structural magnetic resonance imaging (sMRI), schizophrenia, 1512–1516
- Structural model, 917–918
- Structural racism, 837–838, 4105*t*
- Structural variation, 243, 243*f*
- Structured Clinical Interview for DSM (SCID), 1099
- Structured Clinical Interview for DSM-5 (Kid-SCID), 3540
- Structured Clinical Interview for DSM-5 Personality Disorders (SCID-5-PD), 1114–1115
- *Structured Clinical Interview for DSM-IV* (SCID), 884
- Structured Clinical Interview for DSM-IV PDs (SCID II), 2254
- Structured Clinical Interview for DSMIV- TR Dissociative Disorders, Revised (SCID-D-R), 1938*t*
- Structured interactional group psychotherapy, 4766
- Structured interview, 4766
- Structured Interview of Reported Symptoms (SIRS), 2940
- Structured Inventory of Malingered Symptomatology (SIMS), 2940
- Structured Inventory of Reported Symptoms (SIRS), 2505
- Studentized range, 900
- *Studies on Hysteria* (Breuer/Freud), 4766
- Study design, 911
- Stunkard, Albert, 2149
- Stupor, 1673
  - defined, 4766
  - mood disorders and, 1699
- Stuttering, 1182, **3698–3704**. *See also* Communication disorders
  - clinical features, 3702–3703
  - comparative nosology, 3699
  - course and prognosis, 3704
  - defined, 3698, 4766

  - diagnosis, 3699, 3702–3703, 4775*t*
  - differential diagnosis, 3703–3704
  - DSM-5 and DSM-5-Tr criteria for, 3699
  - epidemiology of, 3699–3700
  - etiology of, 3700–3702
  - historical background, 3699
  - ICD-10 and ICD-11 criteria for, 3699
  - ICD-11 criteria for, 3699
  - treatment of, 3704
- Styron, William, 2878
- Subacute necrotizing encephalomyelopathy, 703
- Subbipolar dysthymia, 1678
- Subcaudate tractotomy, 3483–3484
- Subclinical clozapine-associated cardiotoxicity (SCAC), 2285
- Subclinical hypothyroidism, 713
- Subcoma therapy. *See* Insulin coma therapy (ICT)
- Subconscious, 4766
- Subcortical reward nodes, 229–230
- Subgenual anterior cingulate, 35
- Subgranular zone (SGZ), 57, 65
- Subicular complex, 35
- Subjective burden, 4079
- Subjective experience, 807
- Subjective well-being (SWB), 785, 794–797, 796*f*
- Subjectivity, 4766
- Sublenticular limb, 11
- Sublimation, 921*t*, 2748*t*, 4766
- Submissive immobility, 1953
- Subscales, 1090
- Substance abuse. *See also* Substance use disorders (SUDs)
  - amantadine for, 3333
  - bipolar II disorder and, 1693–1694

- burn injuries and, 2467
- in children and adolescents, 3581
- defined, 4766
- family therapy for, 3920–3921
- group psychotherapy for, 3900–3901
- HIV in children/adolescents and, 676–677
- homosexuality and, 2083–2084
- insomnia comorbidity, 2169–2170
- lithium for, 3320
- in medical students, 2690–2691
- nefazodone, 3206
- neurosteroids and, 139
- in older adults, 4133, 4277–4290
  - barriers to detection, 4277*t*
  - benzodiazepines, 4286–4287
  - history, 4278
  - illicit drugs, 4288
  - nonpharmacologic therapies, 4288–4289
    - brief interventions, 4289
    - intensive outpatient programs, 4290
    - moderators and correlates of treatment, 4290
    - partial hospitalization, 4290
    - psychosocial interventions, 4289
    - twelve-step recovery programs, 4289–4290
  - opioid analgesics, 4287–4288
  - signs and symptoms, 4279*t*
  - tetrahydrocannabinol, 4285–4286
- parasomnia and, 2185
- pharmacological treatments of, **3350–3377**
  - acamprosate, 3372–3377
  - buprenorphine, 3353–3355
  - disulfiram, 3368–3373

    - methadone, 3351–3353
    - nalmefene, 3357–3367
    - naltrexone, 3357–3367
  - schizophrenia and, 1478–1479, 1596
  - schizophrenia with, 3068
  - vulnerability to, 870–871
- Substance Abuse and Mental Health Services Administration (SAMHSA), 1294
- Substance dependence. *See also* Addiction
  - D_2 receptor availability in, 401
  - dopamine transmission in, 402
- Substance-induced psychosis
  - differential diagnosis, 3872
  - dopamine receptor antagonists for, 3044
  - symptoms in children and adolescents, 3868*t*
- Substance-induced psychotic disorders, 4270
- Substance/medication-induced anxiety disorder, 1789
- Substance/medication-induced psychotic disorder, **1640–1642**
  - comparative nosology, 1641
  - course and prognosis, 1642
  - diagnostic and clinical features, 1641
  - differential diagnosis, 1641–1642
  - epidemiology of, 1641
  - etiology of, 1641
  - history, 1640–1641
- Substance/medication-induced sleep disorder, 2188
- Substance P, 109*t*, 122
- Substance-related amnesia, 2007
- Substance use disorders (SUDs), **1293–1394**. *See also* Substance abuse
  - addiction process, phases of, 441–443
    - animal models of, 442*t*

  - diagram of, 442*f*
  - preoccupation/anticipation, 443
  - withdrawal/negative affect, 442–443
- addiction theories, 443–444
- ADHD and, 3736
- adolescent, **3998–4003**
  - clinical evaluation, 3999–4002
  - medications, 4001–4002*t*
  - prevalence, 3998–3999
  - prevention, 4000
  - screening, 3999–4000
  - treatment, 4000–4002
- β-adrenergic receptor antagonists for, 3227
- alcohol-use disorders, **1296–1305**
- anabolic-androgenic steroids and, 1444–1445
- animal models of, 300–301
- benzodiazepines, 3248
- bipolar disorder and, 1743
- caffeine-related disorders, **1324–1336**
- in cancer patients, 2361
- cannabis-related disorders, **1337–1347**
- changes in DSM, 1206
- chronic pain, 2486
- cognitive behavioral therapy for, 2825
- in correctional facilities, 4636, 4636*t*
- defined, 4766
- definition of, 1293–1294
- diagnosis, 4825*t*
- dissociative disorders and, 1951–1952
- drug reward, 300–301
- drug self-administration, 301
- DSM-5 criteria for, 441

  - extent of, 1294
  - family therapy for, 3911*t*
  - gabapentin, 3267
  - hallucinogen-related disorders, **1347–1364**
  - HIV infection and, 596–600
  - inhalant-related disorders, **1365–1382**
  - mood disorders and, 1697–1698
  - neural pathways in, 448–450
    - craving, 448
    - impaired control, 448
    - relapse, 448–450
  - neuroscience of, **441–456**
  - obesity and, 2309
  - in older people, 4128–4129
  - organ transplantation and, 2459
  - policy issues, 1295
  - prevention of, 4045–4046
  - psychedelics for, 2907
  - psychiatric examination and, 992–993
  - PTSD and, 1883
  - rating scales for, 1109–1111
  - severe mental illness and, 4557
  - sex differences in, 454
    - dopamine signaling and, 455
    - gonadal hormones and, 454–455
  - stimulant-related disorders, **1312–1323**
  - suicide and, 2709
  - after TBI, 551–552
  - tobacco-related disorders, **1382–1394**
  - treatment of, 455–456, 1294–1295
  - vulnerability to, 870–871
- Substance withdrawal, 4766

- Substantia innominata, 28
- Substantia nigra, 28
- Substitution, 4766
- Subthalamic nucleus (of Luys), 28
- Subventricular zone (SVZ), 57–58
  - regulation of, 65–67
- Succinic semialdehyde dehydrogenase (SSADH) deficiency, 696, 696*f*
- Succinylcholine, 4766
- Sudden sniffer’s death, 1368
- Sudden sniffing death, 1368
- SUDs. *See* Substance use disorders (SUDs)
- Suggestibility, 740, 1178, 2913–2914, 4766
- Suggestion, 4767
- Suicidal behavior
  - children and adolescents, 3821–3827
    - assessment, 3578–3579, 3823
    - biologic factors, 3824
    - course and prognosis, 3826–3827
    - environmental factors, 3824–3825
    - epidemiology, 3821–3823, 3822*f*, 3823*f*
    - etiology, 3823
    - familiar factors, 3824–3825
    - group psychotherapy for, 3903–3904
    - management, 3579–3580
    - medical conditions and, 3824
    - pharmacotherapy, 3928
    - psychiatric disorders and, 3823–3824
    - psychological factors, 3824
    - risk factors, 3823
    - treatment, 3826–3827
  - definitions of, 3821

  - early-onset bipolar disorder and, 3833–3834
  - esketamine for, 3382
  - exposure to, 3824–3825
  - ketamine for, 3382
  - lithium and, 3319
  - mood disorders and, **1709–1719**
    - adverse life situations, 1714
    - affective temperaments, 1713
    - bipolarity, 1716
    - clinical characteristics of mood episode, 1712
    - clinical utility of risk factors, 1714–1715, 1715*t*
    - comorbid disorders, 1712–1713
    - family history of, 1713–1714
    - negative life events, 1714
    - prevention of suicide, 1716–1719
      - nonpharmacologic, 1717
      - pharmacologic, 1717
      - in primary care, 1718–1719
    - previous suicide attempt, 1713
    - prior course of mood disorder, 1713
    - protective factors, 1712, 1715–1716
    - psychological autopsies, 1711
    - risk assessment and formulation, 1709–1711
    - risk factors, 1712–1716, 1715*t*
  - schizophrenia, 1582
  - specific learning disorder and, 3665
  - vilazodone, 3177–3178
- Suicidal ideation, 4767
  - bereavement depression and, 1695*t*
  - bipolar disorder and, 1727–1728
  - dissociation and, 1961
  - in older adults, 4134–4135

  - SSRIs and, 3142
- Suicide, **2705–2714**
  - in adolescents, 2710–2711, 2718
  - age-adjusted rates, 2707*t*
  - age-standardized suicide rate, 2705–2706, 2706*t*
  - alcohol use disorder and, 2709
  - antidepressants for, 2712
  - antipsychotics for, 2713
  - anxiety control, 2712
  - anxiety disorders and, 2709
  - attempts, 2708
  - benzodiazepines and, 3251–3252
  - biomarkers, 284–285
  - bipolar disorders and, 2709
  - in cancer patients
    - assessment and management, 2357–2358
    - risk factors, 2357*t*
  - children/adolescent, 2710–2711
    - poor mental health, 2710
    - risk factors, 2710
    - treatment, 2710–2711, 2718
  - countertransference, 2713
  - defined, 4767
  - depressive disorders and, 2708
  - eating disorders and, 2709
  - electroconvulsive therapy for, 2713
  - epidemiology of, 2705–2708
  - epigenetic signatures of, 284–285
  - epilepsy and, 528
  - homosexuality and, 2083
  - hospitalization, 2711–2712
  - hospital role, 2716–2717

- loss survivors, 2713–2714
- malpractice and, 2717–2718
- medical errors and, 1153–1154
- methods, 2706–2707
- mood stabilizers for, 2712
- in older adults, 2711, 4231
  - cancer patients, 2378
  - risk factors, 2711
  - treatment, 2711, 2718
- patient management, 2711–2712
- personality disorders and, 2709–2710
- pharmacotherapy, 2712–2713
- prevention
  - environment control, 2714
  - media’s role in, 2714
  - school-based, 2714
- protective factors, 2708
- psychiatric disorders and, 2708–2710
- psychological interventions, 2713
- psychological treatments
  - brief, 2713
  - long-term, 2713
- psychotherapy for, 2713
- public health perspective, 2714
- publicly shamed and, 2718
- rates by age and sex, 2706
- rates by age and sex`, 2706*t*
- risk, 2708
  - assessment, 2711, 2715–2716
    - clinical interview, 2711
    - psychometric scales, 2711
  - history, 992

    - during treatment, 2713
  - risk assessment, 3825–3826
  - schizophrenia and, 1465, 2709
  - school programs, 4038–4039
  - significant others, 2717–2718
  - socio-demographic factors, 2706
  - somatic disorders and, 2710
  - substance use disorder and, 2709
  - suicidal statements, 2713
  - supportive telephone calls and letters, 2713
  - survivors, 2717
  - transference, 2713
  - in transgender people, 2116
  - treatment, **2715–2718**
    - biologic, 2716
    - psychotherapy, 2716
  - treatment of acute suicidal crisis, 2712
  - trends, 2707–2708
  - victim-precipitated, 2718
  - in the ward, 2712
- Suicide Behavior Questionnaire-Revised (SBQ-R), 2716
- Suicide-by-cop, 2718
- Suicide cluster. *See* Cluster suicides
- Suicide Intent Scale (SIS), 2711
- Suicide Prevention Resource Center, 2716
- Suicide Status Form (SSF), 2716
- Sullivan, Harry Stack, 959–962, 959*f*, 2793–2794, 2831, 3877, 4690, 4767
  - developmental theories, 960–961, 961*t*
  - interpersonal psychotherapy, 961–962
  - personality theory, 959–960
  - psychopathology theory, 961

- Sulpiride (Meresa), for tic disorders, 3785
- Summarizing, 999
- Sundowner’s syndrome. *See* Sundowning
- Sundowning, 4767
- Superego, 422, 792, 917, 1702, 2748–2749, 4767
- Superintendents, 4029
- Superior longitudinal fasciculus, 13
- Superstitions, 3841
- Supervised machine learning, 907–908
- Supported education, 4556
- Supported employment, 1607
- Support groups, 1126
- Supportive housing, 1587–1588
- Supportive hypnotherapy, 2922
- Supportive psychotherapy, **2742–2744**, 4352, 4767
- Support vector machines (SVMs), 907–908
- Suppression, 793, 921*t*, 2748*t*, 4767
- Suprachiasmatic nucleus (SCN), 37, 194, 194*f*
  - afferent projections, 194–195
  - anatomy, 194, 194*f*
  - efferent projections, 195
  - molecular regulation, 195–196, 195*f*
- Supraoptic nuclei, 37
- Surgery, for epilepsy, 531
- Surgical procedures, child/adolescent response to, 3990
- Surmontil. *See* Trimipramine (Surmontil)
- Surplus reality, 2971
- Surrogate(s), 2400–2402
  - decision-maker, 2401
  - defined, 4767
  - legal issues, 2401–2402
- Survival analysis, 910–911

- Survival circuits, 813
- Survival functions, 813
- Survival instincts, 813
- Susceptibility genes, 4173
- Susto, defined, 4767
- Sutherland, Edwin, 2510
- Sutleffe, Edward, 4662
- Sutter–Eyberg Student Behavior Inventory (SESBI-R), 3546
- Sutton, Thomas, 4684
- Svrakic, Dragan, 809
- Swanson, Nolan, and Pelham (SNAP-IV) Rating Scale, 3546
- Swanson, Nolan, and Pelham (SNAP) scales, 3732
- Swept field NMR, 330
- SWS. *See* Slow-wave sleep (SWS)
- Sydenham, Thomas, 1898, 4646, 4684
- Sydenham chorea, 189
- Symbiosis, 4767
- Symbiotic phase, 922
- Symbiotic psychosis, 4767
- Symbolic play, 746
- Symbolic representations, 2751
- Symbolization, 4767
- Symmetrel. *See* Amantadine (Symmetrel)
- Sympathetic nervous system (SNS), 177–178, 4767
- Sympatholytics, 99
- Sympathomimetics, 99
  - defined, 4767
  - sexual dysfunction and, 2047
- Sympathy, 4767
- Symptom, 4767
- Symptomatic psychosis, 4767
- Symptomatic voice therapy, 3709

- Symptom formation, 4767
- Symptom substitution, 4767
- Synapse, 4767
- Synapse proteome, 249
- Synapses, 3
- Synaptic plasticity, 142–143, 153–154, 154*f*, 452–453, 2781
- Synaptic scaling, 81
- Syncope, 534*t*, 2290, 3074, 3081, 3084, 3086
- Syncretic thought, 4767
- Syndrome, 1198, 2550
- Synesthesia, 4767
- Synkinesia, 3676
- Syntactical (central) aphasia, 4767
- Syntactic mode, 960
- Syntaxic mode of experience, 4767
- Syntropy, 4767
- Syphilis, 620–623, 1053–1054, 4767
  - clinical presentation, 621–623
  - epidemiology of, 620–621
  - management of, 623
  - pathophysiology of, 621
- Systematic Assessment of Treatment-Emergent Events (SAFTEE), 1098
- Systematic desensitization, 760, 4767
- Systematic treatment selection (STS), 2738
- Systematized delusion, 4724, 4767
- System conscious, 916
- Systemic desensitization, 2816
- Systemic lupus erythematosus, 1056
- Systemic racism, 837–838
- System preconscious, 916
- Systems biology, 250

- System unconscious, 916
- Szasz, Thomas, 970, 2062

T

- T (training)-group, 4768
- T3, 3400–3402, 3401*t*
- T4, 3402, 3402*t*
- *Tabernanthe iboga*, 1355
- Tablets, 4469
- Taboos, 792, 4767
- Tachycardia, 2286, 3158–3159
  - types of rhythms with, 2287*t*
- Tachylogia. *See* Logorrhea
- Tacrine (Cognex), 3335*t*, 3336, 4316–4317
- Tactile amnesia. *See* Neurologic amnesia
- Tactile hallucinations, 1455, 4767–4768
- Tadalafil, 3433, 3434*t*
- Taft, William, 1312
- Talents, 789
- Talion, 4768
- Talmon, Moshe, 2857
- Tamburini, Augusto, 4642
- Tamoxifen (Nolvadex), 163, 2055
  - for acute mania, 1731
- Tangentiality, 4768. *See also* Loosening of associations
- Tantrums, 1190–1191
- Taphophobia, 4768
- Tarasoff decision, 4768
- Tardive akathisia, 3091, 3095, 3097
  - clinical presentation, 3097
  - pathology of, 3097
  - treatment options/implications, 3097
- Tardive dyskinesia (TD), 3096–3097
  - antipsychotic drug effects, 1195–1196, 1580–1581, 4311

  - asenapine, 3089
  - clinical presentation, 3096
  - clozapine, 3067
  - defined, 4768
  - differential diagnosis, 3047–3048
  - dopamine antagonists, 3047–3048
  - iloperidone, 3083, 3086
  - lurasidone, 3063
  - olanzapine, 3058
  - paliperidone, 3073
  - pathophysiology of, 3096
  - quetiapine, 3081
  - risperidone, 3073
- Tardive dystonia, 3097
  - clinical presentation, 3097
  - pathology of, 3097
  - treatment options/implications, 3097
- Targeted interventions, 3490
- Tarquinio, Cyril, 1933
- TARs. *See* Transcriptionally active regions (TARs)
- Tart, Charles, 2916
- Tasimelteon
  - chemistry, 3391
  - dosage and administration, 3396
  - drug interactions, 3395–3396, 3396*t*
  - drug studies, 3393–3394
  - molecular structure, 3392*f*
  - organ/system effects of, 3394
  - pharmacodynamics, 3393
  - pharmacokinetics, 3391–3392, 3392*t*
  - precautions and adverse reactions, 3395
  - therapeutic indications, 3394–3395

- TASIT. *See* The Awareness of Social Inferences Test (TASIT)
- Task motivational theory, 2916
- Task-oriented therapies, 3679
- TAT. *See* Thematic Apperception Test (TAT)
- TATA box, 247
- Tatum, Edward, 236
- Tauopathy, 563
- Tax Equity and Fiscal Responsibility Act (TEFRA), 4396, 4512–4513
- Taxon, 2552
- Taylor, Michael Alan, 4656, 4660
- Tay–Sachs disease (GM2 gangliosidosis), 688–689, 688*f*
- TBI. *See* Traumatic brain injury (TBI)
- TCAs. *See* Tricyclic antidepressants (TCAs)
- T cells, 176–177
- TD. *See* Tardive dyskinesia (TD)
- TDCS. *See* Transcranial direct current stimulation (tDCS)
- TDRL. *See* Temporal-difference reinforcement learning (TDRL)
- Teacher Report Form (TRF/6 to 18), 3542
- Teachers, 4028
- Teaching health centers, 4500
- Teasdale, John, 2898
- Technical eclectism, 2737–2738
- Technology. *See also* Internet
  - barriers and limitations, 4476
    - age-related impairments, 4476
    - billing, 4476
    - ethics, 4476
    - information overload, 4476
    - integration with health care system, 4476
    - lack of regulation and evidence, 4476
    - literacy, 4476

    - privacy, 4476
    - usability, 4476
  - future directions, 4469–4470
    - big data, 4469–4470
    - nanotechnology, 4469
    - robotics, 4469
  - for seniors, 4464–4470
  - specific applications of, 4465–4471
    - caregiver applications, 4468–4469
    - games, 4468–4469
    - mobile health, 4465–4467
    - sensors/self-monitoring devices, 4467, 4467*t*
    - smart homes, 4468, 4468*t*
    - social media, 4468–4469
    - tablets, 4468–4469
    - telehealth, 4465
- TEFRA. *See* Tax Equity and Fiscal Responsibility Act (TEFRA)
- Tegretol. *See* Carbamazepine (Tegretol)
- Tele, 2970
- Telehealth, 4465
- Telehealth and Improvement and Modernization Act of 2002, 1124
- Telehealth forensic evaluations, 4009
- Telemedicine, 1121, 1123–1125, 2700
  - cost benefit/effectiveness/outcomes, 1125
  - government support, 1124
  - history, 1123–1124
  - licensure issues, 1125
  - malpractice, 1125
  - reimbursement, 1124
  - remote monitoring, 1124
- Telemental health, 4037
- Telencephalon, 9, 9*f*, 56*f*

  - appetite and, 228–229
  - cortical reward nodes, 229
  - subcortical reward nodes, 228–229
- Teleneuropsychology, 1029–1030
- Telepathy, 4768
- Telephone, 1122–1123
- Telepsychiatry, 1006, 1121, 1123–1125, 1132, 1135, 2349, 4488–4489, 4768
  - advocacy, 4641
  - telepsychiatry, 4640–4641
- Television, 1123
- Telogen effluvium (TE), 2445
  - clinical features, 2445
  - pathophysiology of, 2445
  - psychiatric comorbidity, 2445
  - treatment of, 2445
- Telomere length, 4474
- Temazepam (Restoril), 1429*t*
  - pharmacokinetics of, sex differences in, 3417
- Temperament, 809, 1659–1660, 2218–2224, 4768
  - activity level, 3621
  - ADHD and, 3625
  - adverse events and, 3628
  - anxiety disorders and, 1794, 3626
  - autistic spectrum disorder and, 3626
  - behavioral inhibition and disinhibition, 3621
  - biotypes, 2229*f*
  - brain circuitry and, 3623–3624
  - categories, 3531*t*
  - character, identity and, 2228*t*
  - child, 3519
  - child psychiatric disorders and, **3619–3629**

- clinical implications, 3628
- cognitive factors, 3628
- configurations, 2239*f*
- continuity, 3622
- definition of, 3620, 3620*t*
- difficult child, 3622
- dimensions of, 3620–3621
- disruptive behavior disorders, 3625
- emotional regulation and constraint, 3621
- environmental factors, 3623
- features of, 3622
- genetic–environmental interplay, 3623
- genetics influences on, 3623
- genomic approaches, 2228–2230, 2228*f*
- goodness of fit, 3628
- harm avoidance, 2219–2222
- historical antecedents, 3619–3620
- mood disorders and, 3626
- negative emotions, 3621
- neurobiology and etiology of, 3622–3624
- neurotransmitters and, 3623–3624
- novelty seeking, 2222–2223, 2222*t*
- parenting and, 3627
- peers and, 3628
- persistence, 2224
- positive emotions, 3621
- positive/negative reinforcement, 2225*t*
- psychobiology of, 2218–2219
- psychopathology and, 3624–3627, 3624–3628, 3624*t*
- resiliency and, 3628
- reward dependence, 2223
- sex differences, 3622

  - social affiliation, 3621
  - Thomas and Chess classification of, 3619*t*
  - trauma and, 3628
  - types, 3621–3622
- Temperament and Character Inventory (TCI), 976, 2255, 2257*t*
- Temper tantrms, 1190–1191
- Temporal-difference reinforcement learning (TDRL), 322–323
- Temporal lobe, 3614
- Temporal lobe epilepsy. *See* Complex partial seizures (CPSs)
- Temporary admission, 4768
- Temporomandibular disorder (TMD), 2485
- Tend-and-befriend response, 2582
- Tenex. *See* Guanfacine
- Tenormin. *See* Atenolol (Tenormin)
- Tension, 4768
- Tension-type headache, 646, 646*t*
  - ICHD-3 criteria for, 646*t*
  - treatment of
    - nonpharmacologic, 651–652
    - pharmacologic, 651
- Terman, Lewis, 2526
- Terminal insomnia, 4768
- Terminally ill patients
  - anorexia and cachexia in, 2392
  - communication with doctor, 2386–2387
  - decision-making issues for, 2399
  - disease-specific considerations in, 2398
  - families of
    - coping, 2400–2401
    - final farewell after death, 2400
    - grief therapy for, 2399
    - involvement in decision making, 2403

    - last vigil, 2399
  - fatigue in, 2391
  - group therapies for, 2398
  - insomnia in, 2391–2392
  - pain assessment and treatment, 2391
  - psychological syndromes and symptoms, 2393–2397
  - spiritual concerns for, 2399
  - subjective experiences of, 2398
  - surrogates, 2400–2402
  - types of, 2388
- Terrorism, 4768
  - definition of, 4061
- Terrorism-induced trauma. *See also* Disaster psychiatry
  - in adolescents, 4062, 4064
  - assessment instruments, 4069
  - childhood development and, 4062
  - childhood reactions to, 4061–4062
  - clinical interventions and treatment
    - community-based interventions, 4069
    - individual and family interventions, 4070–4072
    - school-based interventions, 4069–4070
  - community factors, 4063
  - family factors and vulnerabilities, 4063
  - future directions, 4073–4074
  - history of terrorist events and, 4062
  - in latency-age children, 4062, 4064
  - literature review, 4064–4069
    - 1991 Persian Gulf War and Scud missile attacks, 4065
    - 1993 World Trade Center bombing, 4065
    - 1995 Oklahoma City bombing, 4065–4066
    - 1998 American Embassy bombing in Nairobi, Kenya, 4066
    - April 2013 Boston Marathon bombings, 4066

    - July 22, 2011, Utøya youth camp attack, 4066
    - school shootings, 4066–4069
    - September 11, 2001, Attacks, 4066
  - media coverage
    - constructive effects, 4064
    - destructive effects, 4064
    - guidelines, 4072*t*
  - media responses, 4063–4064
  - mental health resources, 4070*t*
  - neurobiologic consequences of, 4062–4063
  - predisposing risk factors, 4063
  - in preschool-age children, 4062, 4064
  - protective factors and resilience, 4063
- Terrorist attacks, group psychotherapy for, 3906
- Tertiary prevention. *See* Preventive psychiatry
- Testamentary capacity, 4768
- Testicular cancer, 2363
- Testimonial privilege, 4768
- Test of Adaptive Behaviors (TABS), 4553
- Test of Memory Malingering (TOMM), 2505
- Test of Pragmatic Language (TOPL), 3690
- Test of significance, 4768
- Testosterone, 162–163
  - for erectile dysfunction, 3434
  - in hormone therapy, 2054
  - structure of, 1436*f*
- Test-retest reliability, 880, 1037, 1092, 3537
- Tests of Malingering and Memory (TOMM), 4551
- Tetanus toxin, 104
- Tetracyclic antidepressants (TCAs), 1067, **3151–3163**. *See also specific drugs*
  - chemistry, 3152

  - defined, 4768
  - dosage/administration, 3162–3163
  - drug interactions, 3161–3162
  - drug preparations, 3163*t*
  - drug studies, 3157
  - history, 3151–3152
  - laboratory interferences, 3162
  - molecular structures of, 3152*f*
  - organ/system effects of, 3158–3159
  - pharmacodynamics, 3154–3156
  - pharmacokinetics, 3152–3154, 3155*t*
    - sex differences in, 3416
  - precautions/adverse reactions, 3159–3161
  - therapeutic indications, 3159
- Tetrahydrobiopterin deficiency (THD), 694–695
- Tetrahydrocannabinol (THC), 130, 4285–4286, 4768. *See also* Cannabis
- Tetrodotoxin (TTX), 72–73
- Text mining, 250
- Thalamocortical association systems, 25–27
- Thalamocortical dysrhythmia, quantitative electroencephalography of, 1081
- Thalamocortical motor systems, 25
- Thalamocortical sensory systems, 24–25
- Thalamus, 10, 13*f*, **22–27**
  - dorsal, 23
  - functional circuitry, 23–27
    - cerebellothalamocortical systems, 27
    - thalamocortical association systems, 25–27
    - thalamocortical motor systems, 25
    - thalamocortical sensory systems, 24–25
  - nuclei, 23*f*

    - classification, 23
    - connections of, 24*t*
  - obsessive-compulsive disorders, 3610–3611
  - in schizophrenic patients, 1505
- Thanatology, 4768
- Thanatophobia, 4768
- Thanatos, 4768
- THC. *See* Tetrahydrocannabinol (THC)
- The Awareness of Social Inferences Test (TASIT), 4551
- Thematic Apperception Test (TAT), 1045–1047, 1045*t*
- Theory, 4768
- Theory of mind, 476
  - schizophrenia and, 1557
- Theory theory, 750–751, 754
- Therapeutic actions, 2752–2753
- Therapeutic agent, 4768
- Therapeutic alliance, 933, 4768
- Therapeutic atmosphere, 4768
- Therapeutic community, 4768
- Therapeutic group, 4768
- Therapeutic privilege, 1156
- Therapeutic role, 4768
- Therapeutic touch, 2648
- Therapeutic window, 4768
- Therapist. *See* Psychotherapist
- Thermoregulation, 204
- Theta burst stimulation, 3464*t*
- Thiamin, 2632–2633
- Thiamine pyrophosphate (TPP), 1283
- Thinking. *See also* Cognition
  - content of, 1168–1171
  - continuity of, 1167–1168

  - disturbances of, 1165–1171
  - flow and form disturbances, 1166–1167
  - normal, 1165–1166
  - primary process, 1166
  - in psychiatric disorders, 1165–1171
  - secondary process, 1166
  - types of, 1166
- Thinking compulsion. *See* Intellectualization
- Thinking–feeling polarity, 951
- Thioridazine (Mellaril), pharmacogenetics of, 294*t*
- Thioxanthene derivative, 4768
- Thioxanthenes, 3035–3038
- Third ear, 4768
- Third interstitial nucleus of the anterior hypothalamus (INAH3), 2069
- Third parties, responsibilities to, 4623
- Third-party payer, 4768
- Third ventricle, 15
- Thomson, Judith Jarvis, 2511
- Thorazine. *See* Chlorpromazine (Thorazine)
- Thorndike, Edward, 756, 2813–2814, 3885
- Thought, 731
- Thought broadcasting, 996*t*, 1456, 4768
- Thought content, 995
- Thought control. *See* Delusion of control
- Thought deprivation. *See* Blocking
- Thought disorder, 1461–1462
  - defined, 4768–4769
  - differential diagnosis, 2722*t*
  - in older adults, 4134
- Thought field therapy, 2647
- Thought insertion, 996*t*, 1456, 4769

- Thought latency, 4769
- Thought process disorder, 4769
- Thought process/form, 995, 995*t*
- Thought reform, 2019
- Thought withdrawal, 4769
- Threatened abortion, 2590*t*
- Three-cornered therapy. *See* Co-therapy
- *Three Essays on the Theory of Sexuality* (Freud), 4769
- Threshold potential, 69
- Thuillier, Jean, 4661
- Thumb sucking, 3808
- Thyroid disorders, 2334–2338
  - affective disorders in, `, 3399
  - hyperthyroidism, 2337–2338
  - hypothyroidism, 2335–2337
- Thyroid hormones, 1056, **3397–3403**
  - biochemistry, 3397–3398
  - clinical studies, 3399–3402
  - dosage and administration, 3403
  - metabolism, 3398*f*
  - pharmacodynamics, 3399
  - pharmacokinetics, 3398–3399
  - precautions and adverse reactions, 3402–3403
  - therapeutic indications, 3402
  - thyroid function in primary affective disorders, 3399
  - use in treatment, 3397
- Thyroid-stimulating hormone (TSH), 3081
- Thyrotropin, 3400
- Thyrotropin-releasing hormone (TRH), 109*t*, 116–117, 3400
- Tiagabine (Gabitril), 3270–3271
  - chemistry, 3270
  - clinical drug studies, 3271

  - cocaine use disorder, 3271
  - dosage and administration, 3271
  - drug interactions, 3271
  - history, 3270
  - pain syndromes, 3271
  - pharmacodynamics, 3270–3271
  - pharmacokinetics, 3270
  - precautions and adverse reactions, 3271
  - seizure disorders, 3271
  - therapeutic indications, 3271
- Tic disorders, **3774–3788**. *See also* Tics
  - clinical features, 3779–3780
  - clonidine for, 3221
  - comorbidities, 3781–3782
    - suicidal behavior, 3781–3782
  - comparative nosology, 3775
  - course and prognosis, 3781
  - defined, 4769
  - definitions, 3774–3775
  - diagnosis, 3779–3780, 4780*t*
  - differential diagnosis, 3780–3781, 3780*t*
  - DSM-5-TR criteria for, 3775
  - epidemiology of, 3776
  - etiology of, 3776–3777
  - genetics of, 3779
  - guanfacine for, 3221
  - history, 3775
  - ICD-11 criteria for, 3775
  - laboratory examination, 3780
  - neuroimaging studies, 3777–3778
  - PANDAS and, 3777
  - pathology of, 3780

  - pathophysiology of, 3776–3777
  - pediatric OCDs and, 3841
  - rating scales, 3780*t*
  - treatment of, 3737–3738, 3782–3787
    - behavioral therapies, 3783
    - general guidelines, 3783
    - pharmacotherapy, 3784–3787, 3786*t*, 3934
    - psychoeducational approaches, 3783
    - psychotherapy, 3783
    - school and occupational interventions, 3783–3784
    - surgical, 3787
- Tic douloureux, 2484
- Tics, 469, 1195
  - classification, 3774
  - defined, 4769
  - definition of, 3774, 3780*t*
  - examples of, 3774*t*
- Time distortion, 2917
- Timed Up and Go Test (TUG), 4179
- Time-limited treatments. *See* Brief psychotherapy
- Time-of-flight (TOF) angiography, 338, 339*f*
- Timidity, 4769
- Timothy syndrome, 709
- Tinnitus, 4769
- Tissot, Samuel, 4645
- Tizard, Barbara, 3802
- TMS. *See* Transcranial magnetic stimulation (TMS)
- TNS. *See* Trigeminal nerve stimulation (TNS)
- Tobacco, EEG effects, 380
- Tobacco dependence. *See* Nicotine dependence
- Tobacco use, psychosis and, 1493–1494
- Tobacco use disorder (TUD), **1382–1394**

- abstinence threshold, 1386
- in adolescents, 1393
- cancer and, 1387
- cessation interventions, 1389*t*
- comparative nosology, 1384
- course and prognosis, 1387–1388
- diagnosis and clinical features, 1386–1387
- diagnostic features, 1384
- differential diagnosis, 1387
- environmental tobacco smoke and, 1387–1388
- epidemiology of, 1384
- etiology of, 1385–1386
- history, 1383–1384
- nicotine and, 1385–1386
- non-nicotine chemicals and, 1388
- in older adults, 4283–4285
  - course and prognosis, 4284
  - epidemiology of, 4283–4284
  - etiology of, 4284
  - pharmacologic treatment, 4284–4285
  - screening for, 4283–4284
- pregnancy and, 1392
- prevalence, 1384–1385
- prevention and policy initiatives, 1393–1394
- psychedelics for, 2907
- treatment, 1386–1392
  - combined psychosocial and pharmacologic therapy, 1391–1392
  - e-cigarettes or vaping for smoking cessation, 1393
  - emerging technologies, 1393
  - naltrexone, 3361–3362
  - nicotine replacement therapies, 1390

    - nonabstinence goals, 1392
    - non-nicotine prescription medications, 1390–1391
    - pharmacologic/interventional therapies, 1390, 1391*t*
    - of psychiatric patients, 1392
    - psychosocial therapies and community resources, 1389–1390
    - transcranial magnetic stimulation, 1391
  - withdrawal, 1386
- Todd, Eli, 4645
- Toddlers. *See also* Children/childhood
  - assessment of, 3529–3532
    - diagnostic formulation of, 3532
    - special types of, 3532–3535
  - formal tests fo development, 3533*t*
  - mental status examination, 3534–3535*t*
  - negative affect states, 822
  - observation and examination of, 3531–3532
  - response to hospitalization/.illnesses, 3987
  - structured assessment instruments, 3532
  - technical issues in the development of, 3530–3531
- Tofranil. *See* Imipramine (Tofranil)
- TogetherAll.org, 2700
- Toilet training, 4769
- Token economy, 4769
- Token economy programs, 2817
- Tolerance, 3013, 4769
  - alcohol, 1298
  - behavioral, 1298
  - cannabis, 1342
  - cellular, 1298
  - cross-tolerance, 1298
  - neural pathways in, 446*f*
  - pharmacodynamic, 1298

  - pharmacokinetic, 1298
- Toloxatone (Humoryl), 3185*t*
- Tomography. *See* Computed tomography (CT); Positron emission tomography (PET); Single-photon emission computed tomography (SPECT)
- Tonic convulsion, 4769
- Tonic tics, 3774
- Toothed clubmoss (*Huperzia serrata*), 2639
- Topamax. *See* Topiramate (Topamax)
- Topectomy, 4769
- Topiramate (Topamax), 3268–3270
  - alcohol use disorder, 3269
  - for alcohol use disorders, 1310
  - binge eating disorder, 3268
  - bipolar disorder, 3269
  - bulimia nervosa, 3268
  - chemistry, 3268
  - clinical drug studies, 3268–3269
  - dosage and administration, 3270
  - drug interactions, 3269–3270
  - for early-onset bipolar disorder, 3835
  - history, 3268
  - migraine, 3268
  - obesity, 3268
  - pharmacodynamics, 3268
  - pharmacokinetics, 3268
  - posttraumatic stress disorder, 3269
  - precautions and adverse reactions, 3269
  - schizophrenia, 3269
  - seizure disorders, 3268
  - smoking cessation, 3269
  - therapeutic indications, 3269

  - weight loss, 3268
- Topologically associating domains (TADs), 277
- Toprol XL. *See* Metoprolol
- Torrey, E. Fuller, 4697
- Torsades de pointes, 2288, 2290
- Tort, 4006
- Torticollis, 3774
- Torture, 4769
- Total parenteral nutrition (TPN), 2392
- Total sleep time, 203
- *Totem and Taboo* (Freud), 4769
- Tourette, Georges Gilles de la, 4686
- Tourette Association of America (TAA), 3775
- Tourette disorder (TD), 3774
  - aripiprazole for, 3102, 3103
  - diagnosis, 4780*t*
  - intellectual disabilities and, 3647–3648
  - pediatric OCDs and, 3841
  - pharmacotherapies for, 3786*t*
- Tourette syndrome, 3774
  - clonidine for, 3221
  - defined, 4769
  - dopamine receptor antagonists for, 3045
  - guanfacine for, 3221
  - neuroimaging, 3612–3614
    - basal ganglia, 3613
    - corpus callosum, 3613
    - cortex, 3612–3613, 3613*f*
- Tower of Fools, 4643
- Toxic epidermal necrolysis (TEN), 2447
  - clinical features, 2447
  - diagnosis, 2447

  - treatment of, 2447
- Toxicity, 4769
- Toxic neuropathies, 658
- Toxic psychosis, 4769
- Toxophobia, 4769
- Toxoplasmosis, 606–607, 627
- *Trabajando la causa* (working the cause), 2672
- Trace amines, 107–108
- Traditional psychotherapy, 2888–2889
- Trailing phenomenon, 4769
- Trail Making Test, 4170–4171
- Trainable, 4769
- Training group. *See* T (training)-group
- Trait, 4769
- Trait anxiety, 1186
- Tramadol, pharmacokinetics, 289
- Trance, 4769
  - trance characteristics, 2915–2916
- Tranquilizers, 3008, 4662–4663, 4769
- Transactional analysis, 4769
- Transactional psychotherapy, 2983–2984
  - ethical issues in, 2984
  - history, 2983
  - research, 2984
  - techniques, 2983–2984
  - theoretical issues, 2983
- Transaminases, 3081
- Transcendental meditation, 2671
- Transcranial alternating current stimulation, for older adults, 4338
- Transcranial direct current stimulation (tDCS), 81, 3473
  - for acute bipolar depression, 1735
  - clinical studies, 3473

  - definition, 3473
  - definition of, 3464*t*
  - future directions, 3473
  - history, 3473
  - in Latin America, 3444
  - mechanism of action, 3473
  - for older adults, 4338
  - side effects, 3473
  - in treatment algorithms, 3473
- Transcranial electrical stimulation, 3464*t*
- Transcranial magnetic stimulation (TMS), 1391, 3468–3472, 3821
  - for acute bipolar depression, 1735
  - in Africa, 3441
  - for anxiety disorders, 4226–4227
  - basic studies, 370
  - brain stimulation and, 3467–3468
  - clinical studies, 3471
  - defined, 4769
  - definition of, 3464*t*, 3468–3469
  - dosing, 3471, 3471*t*, 3472*f*
  - future directions, 3471–3472
  - for geriatric depression, 4240
  - history, 3469
  - mechanisms of action, 3469
  - medication interactions with, 3469–3470
  - patient selection, 3471
  - safety guidelines for rTMS, 3470*t*
  - side effects, 3469–3470
  - treatment algorithms, 3471
- Transcription, regulation of, 245–249
- Transcriptionally active regions (TARs), 240
- Transcriptional start site (TSS), 247

- Transcriptome, 234
  - regulation of, 245–249
- Transcripts of unknown function (TUFs), 239–240
- Transcultural psychiatry, **834–850**, 4769
- Transdermal selegiline, 3189
  - adverse effects, 3189
  - dosage, 3189
  - lactation and, 3189
  - pharmacodynamics, 3189
  - pharmacokinetics, 3189
  - pregnancy and, 3189
- Transference, 857–858, 931, 932*t*, 952, 1993, 2713, 2749, 2749–2750, 2751, 2959*t*, 4769
- Transference-focused psychotherapy (TFP), 2754–2755, 2930
- Transference neurosis, 4769
- Transgender, 2102, 4100*t*
  - antidiscrimination protections, 2118
  - definition of, 2058
  - gatekeeping and informed consent models, 2118
  - incarcerated individuals, 2118
  - Internet resources, 2120*t*
  - medical illness, 2117
  - mental health, 2116–2117
  - suicide in, 2116
  - violence against, 2116
- Transient global amnesia, 534*t*, 2007
- Transient ischemic attack, 534*t*
- Transinstitutionalization, 4627, 4627*f*
- Transition, 4100*t*
- Transitions, 1001, 1002
- Translation parameter families, 896
- Translocator protein (TSPO), 404, 3238, 3255

- Transmembrane AMPA receptor regulatory proteins (TARPs), 86
- Transmissible spongiform encephalopathies (TMEs), 631
- Trans panic defense, 2082
- Transpersonal psychotherapy, 2985–2986
  - history, 2985
  - research and evaluation, 2986
  - techniques, 2986
  - theoretical issues, 2985–2986
- Transplantation, **2449–2465**. *See also* Heart transplants
  - behavioral contracting, 2455
  - clinical evaluation for, 2452
  - comparative nosology, 2449
  - course and prognosis, 2450–2457
  - definition of, 2449
  - dialysis and, 2461
  - diseases resulting in, 2451*t*
  - early posttransplant adaptation, 2457
  - epidemiology of, 2449
  - etiology of, 2450
  - living donor
    - evaluation of, 2454–2455
    - outcomes, 2462
  - long-term outcome, 2457
  - neurotoxicity of immunosuppressant medications, 2463–2465, 2465*t*
  - perioperative period, 2457
  - preparation, 2455
  - psychiatric disorders affecting, 2457–2461
    - alcohol use disorder, 2459
    - anxiety disorders, 2458
    - cognitive disorders, 2460–2461
    - hepatic encephalopathy, 2460–2461

    - mania, 2458
    - mood disorders, 2458
    - nicotine dependence, 2459
    - posttraumatic stress disorder, 2458
    - psychotic disorders, 2458
    - substance use disorders, 2459
    - uremic encephalopathy, 2461
  - psychosocial criteria, 2452
    - evaluator’s role, 2452–2453
    - goals of pretransplant evaluation, 2452*t*
    - screening and content of evaluation, 2452*t*, 2453
      - emergency evaluation, 2454
      - informed consent in terminally ill transplant candidates, 2454
      - living donor, 2454–2455
      - social support, 2453
      - substance use, 2453–2454
      - treatment adherence, 2453
  - special issues in, 2461–2465
  - substance use-disordered patient, 2456
  - survival by organ type, 2450*f*
  - timeline for, 2451*f*
  - transplant-related illness behavior, 2457
  - treatment issues
    - adherence to posttransplant regimen, 2462–2463
    - psychopharmacologic, 2463, 2464*t*
    - psychotherapeutic, 2462
  - ventricular assist devices and, 2461–2462
  - waiting list for, 2450*f*
    - candidate reassessment, 2456
  - waiting period, 2456–2457
- Transplant medications, 2283
- Transposons, 241

- Trans-regulatory elements, 247
- Transsexualism, 1180, 2101, 4770
- Transtheoretical Model (TTM), 2312, 2740
- Transverse magnetization, 328
- Transvestic disorder, 2060, 2096–2097, 2103, 4841*t*
- Transvestic fetishism, 2060, 2091, 2103
- Transvestism, 2060, 2102–2103, 4770
- Tranylcypromine, for major depressive disorder, 1724
- Tranylcypromine (Parnate), 3185*t*, 3187–3188
  - adverse effects, 3188
  - dosage and administration, 3187–3188
  - lactation, 3188
  - monitoring parameters, 3188
  - pharmacodynamics, 3187
  - pharmacokinetics, 3187
  - pregnancy, 3188
- Trauma. *See also* Traumatic brain injury (TBI)
  - alcohol use disorder and, 1306
  - betrayal, 1964
  - defined, 4770
  - dissociation as response to, 1962–1963
  - history, 4657
  - hypnosis and, 2924–2926
  - psychodrama as treatment of choice for, 2972
  - racism and, 4105
  - rating scales for, 1108
  - in schizophrenia, 1463
  - temperament and, 3628
- Trauma and stressor-related disorders, 1206
- Trauma hypothesis, 915–916
- Trauma-informed approach, 1609–1610
- Trauma-related altered states of consciousness (TRASC)

  - 4-D model of, 1966
  - neurobiology of, 1953–1955
- Traumatic abandonment
  - examination of professional patients
    - hypomanic-obsessive defensive style, 4602–4603
    - hypomanic-obsessive style, 4601–4602
    - treatment, 4603–4606
- Traumatic brain injury (TBI)
  - amantadine for, 3332–3333
  - biomarkers, 538–539
  - blast injuries and, 536–537
  - in children/adolescents, 681–684
    - case formulation, 683
    - intervention for, 683–684
    - pathophysiology of, 681
    - psychiatric disorders and, 681–683
      - postinjury, 682
      - preinjury, 682
      - specific, 682–683
    - treatment of, 683–684
  - chronic traumatic encephalopathy and, 544
  - cognitive impairments, treatment of, 543–544
  - cognitive manifestations and sequelae of, 540–542
    - acute traumatic encephalopathy, 540
    - post-traumatic amnesia, 541
    - post-traumatic confusional state, 541
    - post-traumatic disorders of consciousness, 540
    - post-traumatic dysexecutive syndrome, 542
  - defined, 4770
  - differential diagnosis, 1026–1027
  - electrophysiologic studies, 539
  - endocannabinoids and, 135–136

  - epidemiology of, 535–536
  - later-life neurodegenerative dementia and, 544
  - major or mild neurocognitive disorder due to, 542
  - malingering and, 2499
  - mechanisms of, 536–537
  - mild, 542–543
  - moderate-severe, 542
  - neuroendocrine testing, 539
  - neuroimaging, 537–538
  - neuropsychiatric consequences of, **535–553**
  - neuropsychiatric disturbances, 544–553
    - aggression, 545
    - agitation, 545
    - anxiety disorders, 551
    - apathy, 548–549
    - bipolar and related disorders, 549–550
    - comportment changes, 545–546
    - depressive disorders, 546–548
    - pathologic laughing and crying, 548
    - psychotic disorders, 552
    - substance-related and addictive disorders, 551–552
    - trauma- and stressor-related disorders, 550–551
  - persistent postconcussive symptoms, 543–544
  - personality change due to, 682
  - severity criteria, 1026*t*
  - severity of, 539–540, 540*t*
    - grading of, 540*t*
- Traumatic memories, 1993–1994, 1994*t*
- Traumatic neurosis. *See* Post-traumatic stress disorder (PTSD)
- Traumatic stress
  - depersonalization/derealization disorder and, 2013–2014
  - family therapy for, 3919–3920

- Trazodone (Desyrel), **3197–3203**
  - adjustment disorders, 3199
  - adverse reactions, 3202–3203
  - anxiety disorders, 3199
  - behavioral disturbances, 3201
  - bulimia nervosa, 3200
  - for cancer patients, 2348
  - chemistry, 3197
  - childhood disruptive behavioral disorders, 3201
  - clinical drug studies, 3198–3201
  - for depression, in cancer patients, 2356
  - for depression in later life, 4297
  - dosage and administration, 3203
  - drug interactions, 3203
  - generalized anxiety disorder, 3199
  - history, 3197
  - insomnia, 3200
  - laboratory interference, 3203
  - major depressive disorder, 3198–3199
  - molecular structure of, 3197*f*
  - neurocognitive disorders, 3201
  - obsessive-compulsive disorder, 3199
  - organ/system effects of, 3201–3202
    - cardiovascular system, 3201–3202
    - central nervous system, 3202
    - genitourinary system, 3202
  - pain, 3200–3201
  - panic disorder, 3199
  - pharmacodynamics, 3198
  - pharmacokinetics, 3197–3198
    - sex differences in, 3416
  - posttraumatic stress disorder, 3199

  - precautions, 3202–3203
  - schizophrenia, 3200
  - social dysfunction, 3201
  - therapeutic indications, 3202
- Treated prevalence, 4770
- *Treatise on Human Nature* (Hume), 419
- Treatment Episode Data Set (TEDS), 1314
- Treatment foster care (TFC), 3961
- Treatment-resistant depression, 4240–4241, 4299–4300
  - monoamine oxidase inhibitors, 3193
  - nonconvulsive electrotherapy for, 3478
- Treatment(s)
  - failure, 3013
  - outcomes, 3013–3014
    - failure, 3013
    - response and remission, 3013
    - sensitization, 3013
    - tolerance, 3013
    - withdrawal, 3013–3014
- Tremor, 468–469, 1195, 3322, 3676, 3780*t*, 4770
- Trend of thought, 4770
- TRH. *See* Thyrotropin-releasing hormone (TRH)
- Triad, 4770
- Triadic therapy. *See* Conjoint therapy; Co-therapy
- Triangular therapy. *See* Conjoint therapy; Co-therapy
- Triazolam (Halcion), 1429*t*, 4663
  - pharmacokinetics of, sex differences in, 3417
- Tricarboxylic acid (TCA) cycle, 1507
- Trichotillomania (hair-pulling disorder), 1191, 1841*t*, **1850–1852**, 2434–2435
  - automatic pulling, 2434–2435
  - clinical features, 2434

  - comparative nosology, 1851
  - course and prognosis, 1852
  - defined, 4770
  - definition of, 1851
  - diagnosis, 1851–1852, 1851*t*, 4799*t*
  - differential diagnosis, 1852
  - DSM-5 criteria for, 2434–2435
  - epidemiology of, 1851
  - etiology of, 1851
  - focused pulling, 2434
  - laboratory examination, 1852
  - pathology of, 1852
  - pathophysiology of, 2435
  - psychoanalytic theories, 1851–1852
  - subtypes, 2434–2435
  - treatment of, 1852, 2435
- Tricyclic antidepressants (TCAs), 1067, **3151–3163**. *See also specific drugs*
  - anxiety disorders, 1834
  - for anxiety disorders in later life, 4304
  - for binge eating disorder, 2163
  - for bulimia nervosa, 2163
  - chemistry, 3152
  - for chronic pain, 2488
  - defined, 4770
  - for depression, in cancer patients, 2356
  - for depression in later life, 4298
  - for descending neuropathic pain, 218
  - dosage/administration, 3162–3163
  - drug interactions, 3161–3162
  - drug preparations, 3163*t*
  - drug studies, 3157

  - EEG effects, 379
  - for geriatric depression, 4238
  - history, 3151–3152, 4663
  - laboratory interferences, 3162
  - for major depressive disorder, 1724
  - molecular structures of, 3152*f*
  - norepinephrine transporters and, 97
  - organ/system effects of, 3158–3159
  - pain management in burn patients, 2470
  - for pediatric anxiety disorders, 3927
  - pharmacodynamics, 3154–3156
  - pharmacokinetics, 3152–3154, 3155*t*
    - sex differences in, 3416
  - precautions/adverse reactions, 3159–3161
  - pregnancy and, 2605
  - for repolarization abnormalities, 2289
  - therapeutic indications, 3159
  - for tic disorders, 3786
  - weight gain and, 2309
- Tridimensional Personality Questionnaire (TPQ), 601
- Trigeminal nerve stimulation (TNS), 2484, 3478
  - clinical studies, 3478
  - contraindications, 3478
  - definition, 3478
  - definition of, 3464*t*
  - future directions, 3478
  - mechanism of action, 3478
  - side effects, 3478
- Trihexyphenidyl (Artane), 106, 3325
- Trihexyphenidyl hydrochloride, 3325, 3325*f*
- Triiodothyronine (T3), for major depressive disorder, 1724
- Trilafon. *See* Perphenazine (Trilafon)

- Trileptal. *See* Oxcarbazepine (Trileptal)
- Trimipramine (Surmontil), 3152
  - pharmacogenetics of, 295*t*
- Trintellix. *See* Vortioxetine (Trintellix)
- Triphasic waves, 373
- Triptans, for migraine, 651*t*
- Triptorelin (Trelstar), 2098
- Triskaidekaphobia, 4770
- Trisomy, 4770
- Trisomy 21 syndrome. *See* Down syndrome (DS)
- Tristimania, 1652
- True insight, 4770
- True negatives, 880
- True positives, 880
- Truman, Harry, 4512
- Trust, 938, 2404
- Trustworthiness Test, 4551
- Truth-telling, 2403
- Trypanosomiasis, 627
- Tryptamine, 107
- Tryptamines, 430
- Tryptophan, 100–101
- Tsai, Shih-Jen, 1657
- TSH. *See* Thyroid-stimulating hormone (TSH)
- TSS. *See* Transcriptional start site (TSS)
- Tsuang, Ming, 4653
- T-test, 4770
- Tuber cinereum, 36
- Tuberomammillary nucleus (TMN), 201, 202*f*
- Tuberous sclerosis complex, 3641*t*
- TUFs. *See* Transcripts of unknown function (TUFs)
- Tuke, Daniel Hack, 4656, 4685

- Tuke, Samuel, 4685
- Tuke, William, 4685, 4770
- Tukey–Kramer test, 900
- Tulving, Endel, 419, 2231
- Tumor necrosis factor (TFN), 183
- Tupes, Ernest, 975
- Turing test, 2699
- Turmeric (*Curcuma longa*), 2639
- Turner syndrome, 4770
- Twilight state, 4770
- Twin studies, 4770
  - ADHD, 3729
  - dizygotic twins, 253
  - homosexuality, 2070–2071
  - monozygotic twins, 253
  - neurodegenerative disorders, late-life, 4173
  - schizophrenia, 1482–1483, 1617
  - tic disorders, 3779
- Twirling, 4770
- Twitter, 1128
- Two-alternative forcedchoice (2AFC) tasks, 323
- Two-factor theory, 764
- Two-process theory, 764
- Two-spirit, 2102
- Tylenol. *See* Acetaminophen (Tylenol)
- Type 1 error, 4770
- Type 2 error, 4770
- Type A personality, 4770
- Type B personality, 4770
- Type D attachment, 1964
- Tyramine, 107, 4770
- Tyrer, Peter, 1688

- Tyrosine, 95–96
- Tyrosine kinase pathways, 151–153
  - mammalian target of rapamycin, 152–153, 153*f*
  - mitogen-activated protein kinase cascades, 152
  - phosphoinositide 3-kinase pathway, 152, 152*f*

U

- UCLA Post-traumatic Stress Disorder Reaction Index (UCLS-PTSD-RI), 3546
- UCS. *See* Unconditioned stimulus
- UCSC Genome Bioinformatics Center, 250*t*
- U fibers, 13
- Ulcerative colitis, 2294
- Ulrich, Karl, 2061, 2103
- Ultradian rhythm. *See* Biologic rhythm
- Ululation, 4770
- Uncinate fasciculus, 13, 461
- *Unclaimed Children* (Knitzer), 3942
- Uncomplicated bereavement, 2547–2548
  - clinical features, 2548
  - comorbidity, 2548
  - definition of, 2547–2548
  - diagnosis, 2548
  - differential diagnosis, 2548
- Unconditional response, 756
- Unconditioned stimulus (US), 756, 775, 1315, 1807, 3884, 4721, 4770
- Unconscious, 951
  - defined, 4770
  - memories and, 783–784
  - psychotherapeutic technique and, 2758–2759
- Unconscious system, 916

- Unconsummated marriage, 2048–2049
- Underachiever, 4770
- Undershoot, 70
- Undersocialized, 4770
- Undifferentiated schizophrenia, 4770
- Undoing, 2748*t*, 4770
- Undue influence, 4418–4419
- Unhelpful thinking, types of, 4363*t* (*See also* Cognitive behavioral therapy (CBT))
- Unified Protocol for Transdiagnostic Treatment of Emotional Disorders (UPTTED), 3904
- Uniform Data System for Medical Rehabilitation (UDS), 4186
- Unimodal association regions, 21
- Unio mystica, 4770
- Unipolar depression, nonseasonal, 197
- Unipolar disorder, bipolar disorder *versus*, 187
- Unipolar psychosis, 4770
- Universal interventions, 3490
- Universality, 4770
- Universalization, 2959*t*
- Universal sleep hygiene, 2170–2171
- University Centers for Excellence in Developmental Disabilities Education, Research, and Service (UCEDDs), 3632
- University of Iowa, 4653
- University of North Carolina (UNC), 4653
- University of Pennsylvania, 4653
- Unmet online therapy, 1121
- Unspecified communication disorder, **3705–3710**. *See also* Communication disorders
  - clinical features, 3708
  - comparative nosology, 3706
  - course and prognosis, 3709

  - defined, 3705–3706
  - diagnosis, 3708, 4775*t*
  - differential diagnosis, 3709
  - epidemiology of, 3706
  - etiology of, 3706–3708
  - historical background, 3706
  - laboratory examination, 3708–3709
  - pathology of, 3708–3709
  - treatment of, 3709–3710
  - voice therapies for, 3709–3710
- Unspecified feeding and eating disorders (USFED), 2148
- Unspecified personality disorder, 2251
- Unspecified schizophrenia spectrum and other psychotic disorder, **1645–1646**
  - course and prognosis, 1646
  - diagnostic and clinical features, 1646
  - differential diagnosis, 1646
  - epidemiology of, 1646
  - etiology of, 1646
- Unspecified somatic symptom and related disorder, 2438–2440
- Unstructured interview, 4771
- Upbringing away from parents, 2544
  - clinical features, 2544
  - definition of, 2544
  - diagnosis, 2544
  - etiology of, 2544
- Upfield, 330
- Urea cycle disorders (UCDs), 692–693
- Uremic encephalopathy, 2461
- Urethral stage, 918*t*
- Urinalysis, in delirium, 1233
- Urinary incontinence, in medically ill patients, 4207

- Urinary tract infections, in medically ill patients, 4207
- Urocortins, 117–119
- Urolagnia, 4771
- Urticaria, 2445–2446
  - acute, 2446
  - chronic, 2446
  - clinical features, 2445–2446
  - pathophysiology of, 2446
  - psychiatric comorbidity, 2446
  - treatment of, 2446
- Utica Hospital, 4645
- Utilitarianism, 2893, 4616*t*
- Utilitarian principle, 4771
- Utilization review committee, 4771
- 5’ UTR, 237
- 3’ UTR, 237

V

- VACAA. *See* Veterans Access, Choice, and Accountability Act of 2014 (VACAA)
- Vacuolar myelopathy, 608
- Vacuum-assisted erection devices, 3434
- Vacuum pumps, 2055
- Vagal afferent signaling, 225–226, 226*f*, 226*t*
- Vagina, 2029
- Vagina dentata, 2040, 4771
- Vaginismus, 4771
- Vagus nerve stimulation (VNS), 82, 3476–3477
  - clinical scenarios, 3477*t*
  - clinical studies, 3476
  - contraindications, 3476
  - definition, 3476

  - definition of, 3464*t*
  - dosing, 3477
  - evaluation for, 3476*t*
  - future directions, 3477
  - for geriatric depression, 4240
  - history, 3476
  - mechanism of action, 3476
  - for older adults, 4339
  - patient selection, 3476–3477
  - side effects, 3476
  - in treatment algorithms, 3476
- Vaillant, George, 4271
- Valbenazine (Ingrezza)
  - for stereotypic movement disorders, 3813
  - for tardive dyskinesia, 3096
- Valdoxan. *See* Agomelatine (Valdoxan)
- Valerian (*Valeriana officinalis*), 2639–2640, 4143–4144
- Validity, 880–881, 880*t*. *See also* Reliability
  - concurrent, 880, 1037
  - construct, 881, 1037, 1093, 3537
  - content, 880, 1092
  - criteria, 1037
  - criterion, 880, 1092–1093
  - definition of, 1092
  - face, 1036–1037, 1092
  - factor, 1037
  - of personality tests, 1023–1024, 1036–1037
  - predictive, 880, 1037
  - of psychometric assessment, 3549–3550
- Validity Indicator Profile (VIP), 2505
- Valproic acid (valproate, Depakene, Depakote), 1067, **3300–3313**
  - for acute bipolar depression, 1733–1734

  - acute bipolar depression, 3304
  - acute mania, 3301–3303, 3302*t*
  - adverse reactions, 3309–3311, 3311*t*
  - alcohol withdrawal and relapse prevention, 3307–3308
  - for bipolar disorder, 1738
  - bipolar disorder, 3305*t*
  - black box warnings, 3310*t*
  - borderline personality disorder, 3308–3309
  - chemistry of, 3300
  - clinical drug studies, 3301–3304
  - for dementia, 1251
  - dosage and administration, 3312–3313
  - drug interactions, 3312
  - history, 3300
  - impulse control disorder, 3308–3309
  - laboratory interference, 3312
  - for mood stabilization, 3304–3307
  - panic disorder, 3308
  - pharmacodynamics, 3300–3301
  - pharmacokinetics, 3300
  - pharmacologic actions, 3300–3301
  - post-traumatic stress disorder, 3308
  - pregnancy and, 2608
  - schizoaffective disorder, 3307
  - schizophrenia, 3307
  - therapeutic indications, 3309
- Values, cultural, 847–848
- Valvular heart disease, 2285–2286
  - aortic valve disease, 2285
  - infective endocarditis, 2285
  - mitral valve prolapse, 2285
  - rheumatic heart disease, 2285–2286

- Vancouver Coastal Health Authority (VCHA), 4454
- Van der Hart, Onno, 1967
- Vaping, 1393, 3998–3999
- Vardenafil, 3433, 3434*t*
- Varela, Francisco, 418, 419
- Varenicline (Chantix), 106, 1391*t*
- Variable-interval schedule, 4762
- Variable-ratio schedule, 4762
- Variables, 4771
- Variance, 896–897, 4771
- Variation, 4771
- *The Varieties of Religious Experiences* (James), 430–431, 2683
- Vascular assist devices (VADs), 2461–2462
- Vascular dementia, 487, 4175–4176, 4323, 4771
- Vascular depression hypothesis, 4233–4234
- Vascular disease, in medically ill geriatric patients, 4200–4201
- Vascular neurocognitive disorders, major and minor, 4175–4176
  - clinical features, 492–494
  - course and prognosis, 495–498
  - diagnosis, 492–494
  - epidemiology of, 487–490
  - etiology of, 490–492
  - pathology and laboratory examination, 494–495
  - treatment of, 498–500
- Vasculitic neuropathies, 659
- Vasopressin, 119–121
  - receptor distribution in brain, 110*f*
- VCD. *See* Vocal cord dyskinesia (VCD)
- Vegetative, 4771
- Vegetative disturbances, 1674, 1677
- Vegetative nervous system, 4771
- Velocardiofacial (VCF) syndrome, 3640*t*, 4771

- Venlafaxine (Effexor), 98, 3145–3148
  - for anxiety disorders in later life, 4224–4225, 4304
  - for depression in later life, 4297
  - discontinuation syndrome and, 3148
  - drug interactions, 3148
  - hypertension and, 3148
  - overdose, 3148
  - pharmacogenetics of, 294*t*
  - pharmacokinetics of, sex differences in, 3416
  - pharmacology of, 3145–3146
  - side effects, 3819
  - for social anxiety disorder, 3859
  - therapeutic indications, 3146–3147
    - depression, 3146
    - generalized anxiety disorder, 3146–3147
    - social anxiety disorder, 3147
  - tolerability of, 3147–3148
- Ventilation, 2959*t*
- Ventral amygdalofugal pathway, 40
- Ventral pallidum, 28
- Ventral striatum, 28
- Ventral tegmental area (VTA), 444–446, 446*f*
- Ventricular assist devices, 2283
- Ventricular system, 15–17
- Ventricular zones (VZ), 55–58, 57–58, 57*f*
- Ventromedial nuclei, 37
- Ventromedial prefrontal cortex, 726
- Verapamil (Isoptin)
  - for acute mania, 1731
  - chemistry of, 3259, 3260*t*
  - for chronic pain, 2490
  - dosage and administration, 3263

  - drug interactions, 3064, 3174, 3263
  - drug studies, 3261–3262
  - organ/system effects of, 3262
  - pharmacodynamics, 3260–3261
  - pharmacokinetics, 3259
  - precautions and adverse reactions, 3263
  - therapeutic indications, 3262–3263
- Verbal amnesia. *See* Neurologic amnesia
- Verbal aversion, 2817
- Verbal evocation of events, 746
- Verbal fluency, schizophrenia and, 1557
- Verbal learning, schizophrenia and, 1557
- Verbal memory, schizophrenia and, 1557
- Verbal therapies, 2780
- Verbigeration, 4771
- Verbiomania. *See* Logorrhea
- Verghese, Abraham, 2875
- Verhaeghe, Louis, 2272
- Vermis, 28
- Verny, Thomas, 3501
- Vertigo, 4771
- Very-late onset schizophrenia (VLOS), 4267–4268, 4269*t*
- Very–low-density lipoprotein receptor (VLDLR), 63
- Vesicular monoamine transporter (VMAT), 96
- Vestibular disorder, 534*t*
- Veterans Access, Choice, and Accountability Act of 2014 (VACAA), 4521
- Veterans Health Administration (VHA), 4520–4521
- VGKC. *See* Voltage-gated potassium channel (VGKC)
- Viagra. *See* Sildenafil (Viagra)
- Vice principals, 4029
- Vieta, Eduard, 1707

- Vigilance, schizophrenia and, 735, 1557
- Viibyrd. *See* Vilazodone (Viibyrd)
- Vilazodone (Viibyrd), **3175–3179**
  - for anxiety disorders, 1834
  - chemical structures of, 3175*f*
  - chemistry, 3175
  - clinical drug studies, 3176, 3176*t*
  - for depression, in cancer patients, 2356–2357
  - for depression in later life, 4298
  - dosage and administration, 3178–3179
  - drug interactions, 3178
  - history, 3175
  - for major depressive disorder, 1722
  - mechanisms of action, 3175
  - pharmacokinetics, 3175
  - pharmacologic actions, 3175
  - precautions and adverse reactions, 3176–3178
  - side effects, 3819
  - therapeutic indications, 3175
- Viloxazine, pharmacogenetics of, 295*t*
- Vineland Adaptive Behavior Scales 3rd Edition (Vineland-3), 3546–3547
- Violence, 2511–2512, **2623–2627**. *See also* Aggression; Assault; Physical abuse; Sexual abuse
  - anabolic steroid-induced, 1444
  - brain dysfunction and, 2522–2523
  - child abuse and, 2521
  - criminalization and, 4574–4575
  - differential diagnosis, 2524
  - domestic violence, 2624–2625
  - epidemiology of, 2512
  - ethical implications of, 2524–2525

  - evolution and, 2514
  - gender and, 2516
  - genes and, 2516–2519
  - hormones and, 2516
  - hospitalization, 2626
  - initial evaluation, 2626
  - intimate partner violence, 2624–2625
  - legal issues, 2626
  - by the medical profession, 2524
  - neurotransmitters and, 2516–2519
  - postassault psychotherapy, 2626–2627
  - psychopharmacologic treatment, 2627
  - rape, 2624–2625
  - risk history, 992
  - safety assessment, 2627
  - in schizophrenia, 1464–1465
  - school programs, 4038–4039
  - sequelae of, 2626
  - severe psychiatric disorders and, 2523–2524
  - social factors and, 2521–2522
  - treatment issues, 2626–2627
  - violent crime, 2624
  - white-collar, 2510
  - workplace violence, 2625
- Violent behavior
  - β-adrenergic receptor antagonists for, 3227
  - differential diagnosis, 2725*t*
  - dissociative amnesia for, 2004
- Violent crime
  - definition of, 2623
  - perpetrators, 2624
  - prevalence of, 2623

  - risk factors of, 2623–2624
- Viral hepatitis, 1053–1054
- Virtual Kitchen Challenge (VKC), 4189
- Virtual reality, 1121, 1128, 1592
- Virtual reality exposure, 2816
- Virtual Reality Functional Capacity Assessment Tool (VRFCAT), 4553
- Virtual reality graded exposure therapy (VRGT), 2645–2646
- Virtual reality multiple errands test (VMET), 4189
- Virtual reality therapy (VRT), 1131, 1132, 1135
- Virtue, 2892
- Virtue-based ethics, 4616*t*
- Viruses
  - cytomegalovirus, 607, 619
  - endogenous retroviruses, 241
  - Epstein–Barr virus, 619
  - hepatitis C virus, 608–609, 1054
  - herpes simplex virus encephalitis, 618–619
  - HIV (*See* HIV (human immunodeficiency virus))
  - human herpesviruses, 618–619
  - human papillomavirus, 2595, 2937
  - Powassan virus disease, 617
  - Zika virus, 620
- Visceral hallucinations, 1455
- Visceral learning. *See* Biofeedback; Learned autonomic control
- Visceral pain, 2388*t*
- Vision impairment, in medically ill geriatric patients, 4206
- Vistaril. *See* Hydroxyzine (Vistaril)
- Visual agnosia, 4771
- Visual amnesia. *See* Neurologic amnesia
- Visual-evoked potentials (VEPs), 381
- Visual hallucinations, 476, 1174, 1454–1455, 4771

- Visual imagery techniques, 2921. *See also* Hypnotic induction ceremony
- Visual learning, schizophrenia and, 1557
- Visual memory, schizophrenia and, 1557
- Visual motor gestalt test. *See* Bender-Gestalt Test
- Visual network, 804
- Visuoperceptual functioning, assessment, 3569–3570
- Visuospatial cognition, 4170
- Visuospatial disorders, 1021–1022
- Visuospatial dysfunction, 474
- Vitaliano, Peter, 2553
- Vital signs, 466
- Vitamin B12, vitamin B12, 2651–2652
- Vitamin B and homocysteine reduction, 4321
- Vitamin D
  - attention-deficit/hyperactivity disorder and, 2651
  - deficiency, treatment of, 2651
  - depression and, 2651
  - late-life depression, 4146*t*
- Vitamin D_3, 2633
- Vitamin E, 4321
- Vitamins, 2632–2633, 4144–4145
- Vitamin therapy. *See* Orthomolecular psychiatry
- Vitiello Aggression Questionnaire (VAQ), 3546
- Vitiligo, 2442
  - clinical features, 2442
  - pathophysiology of, 2442
  - psychiatric comorbidity, 2442
  - treatment of, 2442
- Vivactil. *See* Protriptyline (Vivactil)
- Vives, Julian Luis, 4683
- VMAT. *See* Vesicular monoamine transporter (VMAT)

- VNS. *See* Vagus nerve stimulation (VNS)
- Vocal cord dysfunction, 2323–2324
- Vocal cord dyskinesia (VCD), 2929
- Vocal function exercises, 3709
- Vocal tic disorders
  - defined, 4753
  - diagnosis of, 4780*t*
- Vogt, Cécile, 4649
- Vogt, Oskar, 4649
- Voice disorder, 4771. *See also* Communication disorders
- Voice therapies, 3709–3710
- Voisin, Auguste, 4646, 4664
- Volatile substances, 1060
- Volition, disturbances of, 1183
- Volitional, 4771
- Voltage, 3445
- Voltage clamp, 69
- Voltage-gated potassium channel (VGKC), 709
- Voltage gated receptors, 2997, 2997*f*
- Volubility. *See* Logorrhea
- Voluminous mental state, 460
- Voluntary actions, 430
- Voluntary admission, 4771
- Voluntary control, 4771
- Voluntary termination of pregnancy (VTOP), 2589–2591, 2591*t*
- Volunteer bias, 4771
- Volunteerism, 2350
- Vomiting, 1346, 2392
- Von Domarus principle, 4771
- Von Economo neurons (VENs), 428
- von Feuchtersleben, Ernst, 4649
- von Jauregg, Julius Wagner, 4649

- von Krafft-Ebing, Richard, 4649
- von Linne, Carl, 4644
- von Meduna, Laszlo, 4652
- Vortioxetine (Trintellix), **3179–3183**
  - for anxiety disorders, 1834
  - chemistry, 3179
  - clinical drug studies, 3179–3181, 3180*t*
  - for depression
    - in cancer patients, 2357
    - in later life, 4298
  - dosage and administration, 3183
  - drug interactions, 3183
  - history, 3179
  - for major depressive disorder, 1722
  - mechanisms of action, 3179
  - pharmacogenetics of, 294*t*
  - pharmacokinetics, 3179
  - pharmacologic actions, 3179
  - precautions and adverse reactions, 3181–3183
  - therapeutic indications, 3179
- Voyeurism, 4771
- Voyeuristic disorder, 2097, 4838*t*
- VTA. *See* Ventral tegmental area (VTA)
- Vulnerability
  - cultural features of, 852, 2573
  - evolutionary explanations for, 862–863
  - to specific disorders, 870–871
    - child abuse, 871
    - eating disorders, 871
    - emotional disorders, 870
    - personality disorders, 871
    - schizophrenia, 870

    - sociopathy, 871
    - substance abuse, 870–871
- Vulva, 2028
- Vulvodynia, 2440
- Vyvanse. *See* Lisdexamfetamine (Vyvanse)
- VZ. *See* Ventricular zones (VZ)

W

- Waelsch, Heinrich, 4651
- Wagner-Jauregg, Julius von, 4687
- Wain, Harold, 2923
- WAIS. *See* Wechsler Adult Intelligence Scale (WAIS); *Wechsler Adult Intelligence Scale IV* (WAIS-IV)
- Waiver, 4008
- Wakefield, Jerome, 1201
- Wakefulness
  - maintenance of wakefulness test (MWT), 204
  - neurobiology of, 205–206
- Wakefulness-promoting agents, 2357
- Waking consciousness, 211
- Wakix. *See* Pitolisant
- Walkabout, 560
- Wall, Patrick, 2933
- Wandering behavior, 2007
- War neurosis, 2500. *See also* Dissociative disorder(s); Post-traumatic stress disorder (PTSD)
- Warren, Stafford, 4654
- Washington University, 4653–4654
- Wasserman, August, 4659
- Wasting syndrome, 1346
- Waterloo-Stanford Group Scale-Form C, 2920
- Water pipe, 1385

- Watkins, John, 2020, 2923
- Watson, James, 236
- Watson, John B., 756, 3884, 4771
- Watts, James, 3479
- Watzlawick, Paul, 2947
- Waxy flexibility, 4771
- WCST. *See* Wisconsin Card Sorting Test (WCST)
- Weakland, John, 2947, 2981
- Wechsler, David, 4167
- Wechsler Adult Intelligence Scale (WAIS), 4771
- Wechsler Adult Intelligence Scale (WAIS-III), 4170
- Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) Vocabulary subtest, 4168
- Wechsler Adult Intelligence Scale IV (WAIS-IV), 1031
- Wechsler Individual Achievement Test—Third Edition (WIAT-III), 3548
- Wechsler Intelligence Scale for Children (WISC), 4771
- Wechsler Intelligence Scale for Children–Third Edition (WISC-III), 3791
- Wechsler Intelligence Scale for Children-V (WISC-V), 3548
- Wechsler Intelligence Scales, 2505
- Wechsler Memory Scale, 4167
- Wechsler Preschool and Primary Scale of Intelligence (WPPSI), 4771
- Wednesday Evening Society, 4771
- Weekend hospital, 4771
- Wehr, Thomas, 1658
- Weight, 1063
  - disturbances of, 1188
- Weighted gene co-expression network analysis (WGCNA), 1508
- Weight gain, 1674
  - clozapine and, 3070

  - monoamine oxidase (MOA) inhibitors, 3191
  - psychiatric medications and, 2308–2309, 2308*t*
- Weight loss, 1674
  - behavioral, 2310
  - benefits of, 2309
  - maintenance, 2310–2311
  - topiramate, 3269
  - treatment of, commercial diets, 2309–2310, 2309*t*
- Weight management, hypnosis in, 2938
- Weil disease (leptospirosis), 625
- Weiss, Mitchell, 2570
- Weiss Functional Impairment Rating Scale, 3732
- Weissman, Myna, 1708
- Weitzenhoffer, Andre, 2912, 2916
- Weizenbaum, Joseph, 2699
- Well-being
  - of adolescents, 3954–3955
  - positive psychiatry and, 4472
- Well-being therapy, 2898
- Wellbutrin. *See* Bupropion (Wellbutrin, Zyban)
- Wellcome Trust Genome Research Center, 250*t*
- Wellness programs, 2350
- Wellness recovery action planning (WRAP), 1608
- Wellstone-Domenici Parity Act. *See* Mental Health Parity and Addiction Equity Act (MHPAEA)
- Welt, Leonore, 486
- Wender, Louis, 2956, 4648
- Wender, Paul, 4700
- Wernicke, Carl, 417, 458, 1199, 1283
- Wernicke aphasia. *See* Fluent aphasia
- Wernicke area, 21
- Wernicke encephalopathy, 1283, 1304

- Wernicke fluent encephalopathy, 4771
- Wernicke-Korsakoff syndrome, 92, 1283, 1304–1305, 4771
- Wernicke's aphasia, 1020–1021, 1020*t*
- West, Louis J., 2020, 4654
- Westen, Drew, 2753
- *Western Aphasia Battery* (WAB), 1032
- Western equine encephalitis (WEE), 616
- West Nile encephalitis, 617
- Westphal, Alexander, 4646
- Westphal, Carl, 2061, 2103, 4649
- Wetware, 463
- Weyer, Johann, 4683, 4772
- WhatsApp, 1123
- Whitaker, Carl, 2943
- White, Hayden, 2875
- White-collar crime, 2509–2510
  - definition of, 2509
  - economic costs of, 2509
  - epidemiology of, 2510
  - etiology of, 2510–2511
  - violence and, 2510
- White matter, 461–462, 3597–3598, 3603, 3614
  - associational fibers, 13–15, 16*f*
  - commissural fibers, 12
  - projection fibers, 10–11
- White-out syndrome, 4772
- White Racial Identity Model, 848
- WHO. *See* World Health Organization (WHO)
- WHO Disability Assessment Schedule (WHO-DAS), 885, 3546
- WHO Family of International Classifications, 1214, 1214*t*
- WHO Global Burden of Disease Study, 892–894, 893*t*
- Whole exome sequencing (WES), 3587*t*

- Whole genome bisulfite sequencing (WGBS), 44
- Whole genome sequencing (WGS), 3587*t*
- Whybrow, Peter, 1657
- Why questions, 1002
- Wide Range Achievement Test-Fifth Edition Reading subtest, 4168
- Widrow–Hoff error correction, 435
- Wilbur, Cornelia, 1931–1932
- Will disorders, 1183
- Williams, Mark, 2898
- Williams syndrome, 3640*t*, 3643–3644, 4772
- Willis, Thomas, 1898, 4645, 4684
- Wilmanns, Carl, 4649
- Wilson, Bill, 3386
- Wilson, Kinnear, 571
- Wilson, Woodrow, 4520
- Wilson disease, 571–572, 703–704, 1058, 4772
  - clinical features, 571
  - course and prognosis, 572
  - definition of, 571
  - diagnosis, 571–572
  - differential diagnosis, 572
  - epidemiology of, 571
  - etiology of, 571
  - historical literature on, 571
  - treatment of, 572
- Wimmer, August, 4657
- Windigo, 4772
- Wing, John, 1200
- Winnicott, Donald, 925–926, 3525, 4647, 4690
- Winokur, George, 1652, 4653
- Winslow, Forbes, 4686
- Winter depression, 197

- Winternitz, Milton C., 4652
- WISC. *See* Wechsler Intelligence Scale for Children (WISC)
- Wisconsin Card-Sorting Task, Trails, 4551
- Wisconsin Card Sorting Test (WCST), 1033, 4170, 4554
- Wisdom, 4473
- Withdrawal, 3013–3014, 4772
  - agitation and, 3576–3577
  - neural pathways in, 447
  - opponent process theory of motivation and, 442–443, 443*f*
- Withdrawal akathisia, 3095
- Withdrawal syndromes
  - alcohol withdrawal, 1232
  - buspirone for, 3173
  - caffeine withdrawal, 1332–1334
    - caffeine dose and, 1333
    - clinical features, 1332–1333, 1333*t*
    - comorbidity, 1332
    - course and prognosis, 1333
    - diagnosis, 1332–1333
    - differential diagnosis, 1333
    - epidemiology, 1332
    - headache, 1333
    - severity of, 1333
    - treatment of, 1333–1334
  - cannabis withdrawal, 1342, 1345
  - defined, 4772
  - delirium and, 1232
  - diagnosis, 4826–4827*t*
  - opiate withdrawal, 1232
  - sedative-hypnotic agents, 1420
  - stimulant withdrawal, 1318–1319
    - treatment of, 1321

  - tobacco withdrawal, 1386
- Within-session extinction, 2816
- Witmer. Lightner, 2886
- Wittkower, Eric, 4650
- Wnt signaling, 153–154, 153*f*
- Wolf, Alexander, 2956
- Wolf, Stewart, 2273
- Wolff, Christian von, 805
- Wolff, Harold, 2273
- Wolff, Peter, 1959
- Wolfram disease type I and type II, 703, 703*f*
- Wolpe, Joseph, 971, 3885
- Wolper, Daniel, 430
- Woodard, John, 4658
- Woodcock–Johnson Psychoeducational Battery—Fourth Edition (W/JPB-IV), 3548
- Woodward, Samuel, 4510, 4645
- Worcester Recovery Center and Hospital (WRCH), 4509*f*, 4510*f*
- Word approximation, 4772
- Word association technique, 4772
- Word blindness, 4772
- Word salad, 4772. *See also* Incoherence
- Working memory, 301, 775
  - attractor network models, 318
  - neural circuit basis of, 318
  - neuropsychological evaluation, 4170
  - schizophrenia and, 1557
- Working through, 4772
- Workplace violence, 2625
  - perpetrators, 2625
  - prevalence of, 2625
  - risk factors, 2625

- World Congress of Psychiatry, 4681
- World Federation for Mental Health (WFMH), 4681
- World Health Organization (WHO), 4681, 4772
- World Health Organization Disability Assessment Schedule (WHODAS 2.0), 1093
- *World Mental Health* (WMH) *Initiative*, 889–890, 890*t*
- World Psychiatric Association (WPA), 4678, 4681, 4772
- World psychiatry, **4667–4682**
  - barriers to change, 4678
  - burden/prevalence of mental disorders, 4667–4672, 4668–4672*t*
  - children/adolescent, 4670–4671
  - disability-adjusted life years, 4669–4670
  - ethical issues in, 4680–4681
  - future directions, 4682
  - international organizations, 4681–4683
  - overview, 4667
  - physical health comorbidities, 4679–4680
  - primary care, 4669
  - program development principles, 4677–4678
  - resources for, 4675–4677, 4676–4677*t*
  - stigmatizing attitudes toward people with mental disorders, 4678–4679
  - suicide, 4671–4672
  - treatment gap, 4673–4675*t*
- World salad, 1462
- WPA. *See* World Psychiatric Association (WPA)
- Writing expression disability. *See* Specific learning disorder (SLD)
- Wynne, Lyman, 2943–2944

X

- Xanax. *See* Alprazolam (Xanax)
- Xenophobia, 4772

- X-linked adrenoleukodystrophy (X-ALD), 689–691, 690*f*

Y

- Yaden, David, 431
- Yale-Brown Obsessive-Compulsive Scale (YBOCS), 1108
- Yale Food Addiction Scale (YFAS), 2306
- Yale University, 4652
- Yalom, Irwin, 2980
- Yap, Pow Meng, 1469, 2552
- YBOCS (Yale-Brown Obsessive-Compulsive Scale), 1108
- YMRS (Young Mania Rating Scale), 1105–1106
- Yocon. *See* Yohimbine (Yocon)
- Yoga, 2144, 2643, 2675
- Yohimbine (Yocon)
  - described, 4772
  - for sexual dysfunctions, 2054
- York retreat, 4772
- Young, Allan, 2554
- Young, Jeffrey, 2818, 2930
- Young Mania Rating Scale (YMRS), 1105–1106
- *Young Man Luther* (Erikson), 940–941
- Youth. *See* Adolescence/adolescents; Children/childhood
- Youth Aware of Mental Health (YAM), 2714
- Youth Risk Behavior Surveillance System (YRBSS), 892, 3950
- Youth Risk Behavior Survey (YRBS), 3949
- Youth Self Report (YSR/11 to 18), 3542
- YouTube, 1123
- YRBSS. *See* Youth Risk Behavior Surveillance System (YRBSS)
- Yttrium-90, 4772
- Yudofsky, Stuart, 4652

Z

- Zacchia, Paolo, 4684
- Zaleplon (Sonata), 1429*t*, 3242–3243, 3242*f*
- *Zar*, 2672
- Zarontin. *See* Ethosuximide
- Z-drugs, 3240*t*, 3254
- Zeger, Scottt, 909
- Zeitgebers, 194
- Zeitgeist, 4772
- Zeller, Ernst Albert, 3184, 4644
- Zen Buddhism, 2670–2671
- Ziconotide, for chronic pain, 2490
- Zimelidine, 3007
- Ziprasidone (Geodon), 99, 3060–3062
  - for agitation, 2724
  - for bipolar disorder, 1730*t*, 1739
  - for bipolar I disorder, 3061, 3062
  - chemistry, 3060
  - clinical trials, 3060–3061
  - dosage and administration, 1579, 3062
  - drug interactions, 3061–3062
  - effects on organs and systems, 3061
  - molecular structure, 3056*f*
  - pharmacodynamics, 3060
  - pharmacokinetics, 3060
  - precautions and adverse reactions, 3061–3062
  - for schizophrenia, 3060–3061
  - for tic disorders, 3785
- Zolpidem (Ambien), 1429*t*, 3242–3243, 3242*f*
- Zonegran. *See* Zonisamide (Zonegran)
- Zone of optimal performance, 806
- Zonisamide (Zonegran), 3273–3274
  - bipolar disorder, 3273

  - chemistry, 3273
  - clinical drug studies, 3273
  - for dementia with Lewy bodies, 1268–1269
  - diabetic neuropathy, 3273
  - dosage and administration, 3274
  - drug interactions, 3274
  - eating disorders, 3273
  - history, 3273
  - pharmacodynamics, 3273
  - pharmacokinetics, 3273–3274
  - precautions and adverse reactions, 3273–3274
  - seizure disorders, 3273
- Zoophilia, 4772
- Zoophobia, 4772
- Z-score, 4772
- Zuckerman Kuhlman Personality Questionnaire (ZKPQ), 601
- Zung Self-Rating Depression Scale (SDS), 4772
- Zwir, Igor, 2216
- Zygote, 4772
- Zygotic, 4772
- Zyprexa. *See* Olanzapine